МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

Пономаренко В. С.
Малярець Л. М.

БАГАТОВИМІРНИЙ АНАЛІЗ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

Навчальний посібник

Харків. Вид. ХНЕУ, 2009

УДК 330.42:330.342+316.323(075.8)
ББК 65.013я73
П56

Рецензенти: докт. екон. наук, професор, зав. кафедри економічної кібернетики Прикарпатського національного університету імені Василя Стефаника *Благун І. С.;* докт. екон. наук, професор, зав. кафедри економічної кібернетики і маркетингового менеджменту Національного технічного університету "Харківський політехнічний інститут" *Заруба В. Я.;* докт. екон. наук, професор, зав. кафедри економічної кібернетики Київського національного університету імені Тараса Шевченка *Черняк О. І.*

Затверджено на засіданні вченої ради Харківського національного економічного університету.
Протокол №5 від 02.12.2008 р.

Рекомендовано Міністерством освіти і науки України як навчальний посібник для студентів вищих навчальних закладів (лист №14/18-Г-2490 від 03.12.2008 р.)

Пономаренко В. С.
П56 Багатовимірний аналіз соціально-економічних систем : навчальний посібник / В. С. Пономаренко, Л. М. Малярець. – Харків : Вид. ХНЕУ, 2009. – 384 с. (Укр. мов.)

За організацією і змістом матеріалу посібника передбачається формування знань та вмінь багатовимірного аналізу соціально-економічних систем, здатності за допомогою математичних методів та описового моделювання аналітично обґрунтовувати ухвалення рішень в управлінні соціально-економічними системами різного рівня, на основі методів і моделей аналізу даних аналітично забезпечувати функції управління соціально-економічними системами: діагностику, оцінку, планування та прогнозування, контроль і контролінг.

Рекомендовано для студентів при неперервній математичній підготовці, магістрів, аспірантів для проведення наукових досліджень за допомогою інструментів аналізу даних, економістів-практиків для обґрунтування та ухвалення управлінських рішень.

ISBN 978-966-676-347-4

УДК 330.42:330.342+316.323(075.8)
ББК 65.013я73

Вступ

Очевидним фактом є те, що ухвалені управлінські рішення в економіці фундаментально обґрунтовуються використаними математичними методами та моделями. У свою чергу, розвиток інструментів економіко-математичного описового моделювання спрямовується нагальною практичною потребою проведення достовірного, повномасштабного, комплексного аналізу даних соціально-економічних систем за їх ознаками, що вимірюються на різних шкалах. Саме в процесі багатовимірного аналізу даних адекватно описуються соціально-економічні системи, формується достовірна модель реальних об’єктів в економіці, генеруються нові знання про них. Це ініціювало авторів багато років проводити дослідження проблем аналізу даних в економіці, удосконалювати методи, методики та технології обробки економічної інформації, але й написати посібник, в якому були б викладені теоретичні та практичні проблеми багатовимірного аналізу соціально-економічних систем для підвищення рівня та якості навчання студентів економічних спеціальностей з метою формування професійних компетенцій економістів, менеджерів, спеціалістів галузі знань "Комп’ютерні науки".

За організацією та змістом матеріалу посібника передбачається накопичення знань та вмінь багатовимірного аналізу соціально-економічних систем, здатності за допомогою математичних методів та описового моделювання аналітично обґрунтовувати ухвалення рішень в управлінні соціально-економічними системами різного рівня, на основі методів та моделей аналізу даних аналітично забезпечувати функції управління соціально-економічними системами: діагностику, оцінку, планування та прогнозування, контроль та контролінг.

Матеріал кожного розділу в посібнику структурований відповідно до формування п’яти рівнів компетентності у вирішенні завдань: стандартних та передбачуваних; складних та нестандартних; тих, що супроводжуються формуванням нових знань, на основі чого буде здійснюватися контроль та управління; складних технічних завдань із значною відповідальністю та розподілом ресурсів; тих, що передбачають застосування фундаментальних принципів та комплексних методик; тих, що мають особисту відповідальність за проведення аналізу та виявлення причин відхилень, розробку, планування, виконання й оцінку. Кожний розділ посібника складається з огляду сучасних розробок відомих вчених близького та дальнього зарубіжжя з даного питання; узагальнень та рекомендацій авторських результатів вирішення даного питання; викладені методики проведення багатовимірного аналізу соціально-економічних систем різного рівня управління; практичні рекомендації щодо здійснення аналізу у відповідному програмному середовищі: Statgraphics Plus, Excel; наведені типології розв'язування реальних задач в економіці; подані задачі для самостійних занять та запитання для самоперевірки. Для набуття навичок аналізу даних важливо вміти розв'язувати типові задачі в спеціальних програмних середовищах, точніше знати теоретичні основи багатовимірного аналізу, які синтезуються в уміння практично вирішувати завдання за допомогою відповідних програмних засобів.

Посібник складається із семи розділів: методологічні засади багатовимірного аналізу соціально-економічних систем; методи багатовимірного статистичного аналізу метричних ознак соціально-економічних систем; багатовимірний аналіз метричних ознак соціально-економічних систем; аналіз соціально-економічних систем за неметричними ознаками; використання вимірників в аналізі соціально-економічних систем; аналіз соціально-економічних систем у скороченому просторі різних ознак та вимірників; базисні описові моделі багатовимірного аналізу соціально-економічних систем. Логіка цих розділів відповідає системному аналізу об'єктів в економіці та технології визначення величин їх ознак, що вимірюються на різних шкалах.

При розробці форми та змісту посібника автори мали за мету написати навчально-методичну роботу, яка б допомогла сформувати високі професійні компетенції майбутнім економістам, менеджерам, аналітикам, інженерам-економістам та була доступна у вивченні інструментів багатовимірного аналізу соціально-економічних систем студентам при неперервній математичній підготовці, магістрам у поглибленні своїх знань, аспірантам для проведення наукових досліджень за допомогою інструментів аналізу даних, економістам-практикам для обґрунтування та ухвалення управлінських рішень.

Розділ 1
Методологічні засади багатовимірного аналізу соціально-економічних систем

1.1. Генезис аналізу даних в економіці

У сучасних умовах науково-технічного розвитку в усіх сферах діяльності людини стало аксіомою прийняття рішення на основі аналізу даних. Способи, методи отримання інформації з даних та вироблення нових знань є супроводом у системах підтримки ухвалення управлінських рішень в управлінні об’єктами різної природи. Метою аналізу даних є вивчення властивостей об’єктів, явищ та процесів, отримання нових знань про них для більшого підпорядкування.

Сучасний аналіз даних обумовлюється способами отримання величин, методами їх обробки й залежить від розвитку математичних методів і моделювання. Це твердження доводять теорія і практика. Методи аналізу даних спрямовані на формулювання й уточнення основних закономірностей систем на перших стадіях їх вивчення в умовах невизначеності або часткової визначеності на основі емпіричних даних [187; 270; 274; 290; 306]. Дана ситуація є типовою для всіх сфер діяльності людини, наприклад, директор фірми на основі даних про діяльність підрозділів намагається скласти об’єктивне уявлення про їх функціонування; працівники економічного управління намагаються вивчити основні тенденції економічного і соціального розвитку регіону на основі системи показників протягом встановленого періоду; спеціалістам науково-експертного управління країни потрібно вивчити й достовірно порівняти економічний та соціальний стан областей. Перелік прикладів можна продовжити до нескінченності, але всі вони потребують використання методів аналізу даних для впорядкування наявної інформації, подання її в лаконічній, узагальненій, стислій, очевидній формі, яка полегшує процедуру формування управлінського рішення за виявленими тенденціями, закономірностями, вилученими новими знаннями [281; 316; 320; 323].

Відомий факт, що людина в своїй свідомій діяльності завжди виконувала аналіз даних, але як наукова проблема аналіз даних сформувався тільки в епоху інформатизації суспільства, тобто зовсім недавно. Прорив у теорії і практиці аналізу даних відбувся з появою обчислювальних машин, а точніше з появою персональних комп’ютерів (ПК) та статистичних програмних пакетів. З появою потужних та зручних пакетів для аналізу даних на ПК розширилося та змінилося коло споживачів методів аналізу даних [275, с. 5]. Раніше методи аналізу даних розглядалися як інструмент наукових досліджень, але починаючи з середини 80-х років користувачами методів стали не наукові організації, а структури бізнесу, комерційні та різні організації. Спочатку в країнах далекого зарубіжжя, а тепер і в Україні методи аналізу даних, їх реалізації в статистичних пакетах стали загальновживаним інструментом планових, аналітичних, маркетингових відділів виробничих і торговельних корпорацій, банків і страхових компаній, урядових і медичних закладів. Таке поширення пояснюється швидкістю отримання результатів обчислень, навіть за умов великих масивів вихідних даних, а також можливостями візуалізації, наочного подання результатів обробки даних та інтерактивного діалогового процесу обробки даних. Звичайно, перелік переваг аналізу даних на ПК можна продовжити, вони є загальновизнаними [240; 264; 322; 395].

В умовах трансформаційної економіки швидко відбуваються докорінні зміни, структурні зрушення, що відстежуються в змінах характеристик СЕС. Це має оперативно фіксуватися, аналізуватися й узгоджуватися на мікро-, мезо- й макрорівнях управління.

Таку функцію виконують інформаційні системи, які передбачають аналіз даних. Останнім часом інтенсивно розробляються інформаційні технології Business Intelligence (BI), які об’єднують збір даних, аналіз моделювання і прогнозування, при цьому для збільшення матеріальних доходів використовуються нематеріальні активи: знання персоналу, цінність клієнтів, мережі бізнес-партнерів і т. д. Інструменти BI включають технології аналітичної обробки даних On-line Analytical Processing (OLAP), здобування даних (Data Mining), різні програмні продукти у вигляді статистичних пакетів. OLAP – це технологія оперативної аналітичної обробки даних, що використовує методи і засоби для збору, збереження та аналізу багатовимірних даних з метою підтримання процесів ухвалення рішень [182, с. 53]. Технологією OLAP передбачається аналіз інформації за допомогою найзручнішого способу її подання – багатовимірної моделі або гіперкуба, ребрами якого є результати вимірювання, що дозволяє аналізувати дані відразу за декількома вимірюваннями, тобто виконувати багатовимірний аналіз.

Наразі методи аналізу даних мають підґрунтям методи математичної статистики, еволюційне моделювання та методи машинного навчання. Сучасний розвиток методів математичної статистики відображається в удосконаленні оцінювання параметрів розподілу величин, перевірці статистичних гіпотез, дисперсійному аналізі, кореляційному аналізі, регресійному аналізі, аналізі часових рядів, багатовимірному статистичному аналізі (БСА). Еволюційне моделювання передбачає використання генетичних алгоритмів, штучних нейронних мереж (ART-мереж, мереж зворотного розповсюдження, мереж зустрічного розповсюдження, мереж Хеммінга, мереж Хопфілда, мереж Кохонена, RBF-мереж). Машинне навчання будується на деревах рішень і використанні ентропійної міри.

Для здійснення автоматичного аналізу даних розроблені сучасні методики, наприклад Data Mining (здобування даних). Методики Data Mining сформовані як синтез методів статистики, теорії інформації, машинного навчання, теорії баз даних. Добування даних визначається як процес аналітичного дослідження великих масивів інформації з метою виявлення закономірностей і систематизації взаємозв'язків між змінними, які потім можна використати для нових даних. Здобування даних як процес включає три основних етапи: дослідження, побудови моделі та її перевірки, що є традиційними для економіко-математичного моделювання. Вважається, що методики Data Mining ґрунтуються на класичних принципах розвідувального аналізу даних (РАД) й побудови моделей, але, на відміну від них, вони спрямовані на кінцеве практичне застосування отриманих результатів, а не на дослідження природи даного явища. При цьому зосереджується увага на пошуці рішень для подальшої побудови прогнозів розвитку явищ та процесів на основі нейронних мереж без конкретизації залежностей, на яких ґрунтується прогноз.

За допомогою методів Data Mining вирішуються такі завдання: класифікації, регресії, кластеризації, асоціації, встановлення послідовних шаблонів, аналізу відхилень у результаті виявлених нехарактерних шаблонів.

Нейронні мережі – один з методів, на основі якого організовується Data Mining. Сучасна інтелектуалізація систем управління здійснюється за двома напрямами формалізації: знань людини про об'єкт управління і способів мислення. Перший напрям формують застосування інструментів нечіткої логіки (fuzzy logic), другий – обумовлений розвитком генетичних алгоритмів та штучних нейронних мереж (ШНМ). Другий напрям формалізації знань є альтернативним класичній парадигмі й відносно новим напрямом у теорії автоматичного управління, за яким пропонується інший спосіб відображення й перетворення дійсності [29, с. 40]. Для класичної парадигми характерне те, що синтез математичної моделі об'єкта обов'язково випереджається фазою аналізу, протягом якого об'єкт на основі мислення декомпозується на елементарні, кожне з яких детально досліджується, і, як результат, пропонується проста, частіше за все лінійна, модель зі сталим коефіцієнтом. Відмінність мережного підходу полягає в тому, що він спочатку не передбачає фази аналізу в процесі побудови моделі. Нейронні мережі – суто синтетичний, а не аналітичний підхід. Тут залишається поняття шаблону і синтез моделі зводиться до параметричної оптимізації шаблону, сконструйованого за базисом активаційних функцій нейронів. Задача формулюється для всієї мережі в цілому. Тому й об'єкт досліджується відразу повністю і в не очищених ситуаціях, а в тих режимах, що цікавлять дослідника з практичної точки зору [29; 90].

Нейронні мережі визначаються як потужний математичний апарат, який може бути застосований в якості універсального відтворювача складних нелінійних функціональних залежностей і дозволяє виявити головні тенденції зміни фінансового показника за експериментальними даними попередніх періодів та відповідно до них робити прогноз зміни даного показника в майбутньому на встановлену кількість кроків вперед [174, с. 66 – 67]. Крім того, даний метод позбавлений таких недоліків прогнозування, як монотонність чи періодичність майбутнього курсу, властивих чисельним методам екстраполяції, або усереднення прогнозованого курсу, як у методах найменших квадратів, середнього плинного чи в регресійних моделях. Як головну перевагу виділяють їх здатність до навчання, що не вимагає ніякої апріорної інформації про структуру шуканої функціональної залежності, необхідною є лише навчальна вибірка у вигляді експериментальних пар "входи-виходи".

Поряд із перевагами нейронних мереж існують недоліки, а саме кінцеве рішення залежить від початкових установок мережі і його практично неможливо інтерпретувати в традиційних аналітичних термінах, що змушує розглядати їх як "чорний ящик". У методі зосереджена увага виключно на практичному результаті, а не на суті механізмів, що лежать в основі явища, або відповідності отриманих результатів існуючим теоріям в економіці [182, с. 88].

Серед переваг дерева рішень в аналізі даних виокремлюються порівняння з нейронними мережами: для навчання дерева рішень потребується менше часу, ніж на навчання в нейронних мережах; проблеми вимірності простіше вирішуються в методі дерева рішень, на відміну від мереж, що обумовлює й так достатню тривалість процесу навчання.

Другий метод еволюційного моделювання, який пропонується в Data Mining, – нечітка логіка. Виникнення невизначеності даних має такі причини: невідо-мість, неповнота (недостатність, неадекватність) і недостовірність, яка поділяється на фізичну і лінгвістичну [182, с. 83 – 84]. Фізична невизначеність буває у формі неточності й випадковості; лінгвістична невизначеність – невизначеність значень слів і невизначеність змісту фраз. Саме лінгвістична невизначеність обумовила розвиток теорії і методів нечіткої логіки. Виокремлюються три відмінні особливості нечіткої логіки: 1) правила ухвалення рішення є умовними висловлюваннями на зразок "якщо …, то …" і реалізуються за допомогою механізму логічного висновку; 2) замість одного чіткого узагальненого правила нечітка логіка оперує з множиною окремих правил. При цьому для кожної локальної області розподіленого інформаційного простору, для кожної величини, що регулюється, для кожної мети управління задаються свої правила, що дозволяє не використовувати трудомісткий процес згортки цілей і отримання узагальненого цільового критерію, а надає змогу оперувати навіть з протилежними цілями; 3) правила у вигляді "якщо …, то …" дозволяють вирішувати завдання класифікації в режимі діалогу з оператором, що обумовлює підвищення якості класифікатора вже в процесі експлуатації [19; 182]. Пропонується в аналізі даних використовувати нейрочіткі системи, що складаються з методів нечіткої логіки, нейронних мереж, генетичних алгоритмів і експертних систем. Ці інтелектуальні гібридні системи розбудовують для того, щоб усунути обмеження кожного методу окремо [112].

Нейрочіткі системи призначені для розвитку автоматичних систем управління. Дискусійність застосування даних систем для повномасштабного опису СЕС обумовлена існуванням "чорних ящиків" в їх реалізації, що складно приживається в економіко-математичному моделюванні, де кожна обчислювальна процедура має свій зміст в економіці. Лінгвістична нечіткість оперує з величинами, виміряними на неметричних шкалах, а рух в обчислювальних процедурах від метричних шкал до неметричних не є прогресивним процесом, проте допустимим у випадку необхідності. Практична цінність нейрочітких систем в управлінні технічними системами визнана, але для розбудови описових моделей стану СЕС процедури "чорного ящика" мало підходять, тим паче, що існують саме для даних величин розроблені методи математичної статистики. Так, вони потребують модифікацій, удосконалень, але їх процедури, проміжні величини мають інтерпретацію в економіці.

У цілому ж методи Data Mining узаконюють "фізичний підхід", основою якого є організація процедур аналізу даних, подібних до встановлення фізичних законів: збір експериментальних даних, подання їх у вигляді таблиць й пошук схем міркувань, що обумовлюють очевидність отриманих результатів і надають змогу розробляти прогнози.

Проте серед методів аналізу даних пріоритетне місце залишається за методами статистичного аналізу, оскільки вони універсальні, тобто можуть застосовуватися в різних сферах діяльності людини. Незважаючи на те, що в статистичних пакетах акумулюються найновітніші наукові розробки в галузі статистичного аналізу, визначення складних характеристик СЕС, їх моделювання залишаються методологічною проблемою.

Соціально-економічні системи визначаються сукупністю ознак, які вимірюються на різних шкалах. Соціальна підсистема, як правило, характеризується якісними ознаками, величини яких отримуються за допомогою порядкових чи номінальних шкал. Повний опис соціально-економічних систем можливий тільки завдяки врахуванню кількісних і якісних ознак, величини яких вимірюються на різних шкалах. Тут маємо проблеми багатовимірності та їх наслідки – проблеми сумісної обробки різних типів даних. Дані проблеми вирішуються за допомогою методів багатовимірного статистичного аналізу (БСА), але проблеми сумісної обробки різних величин залишаються невирішеними до цього часу. Ефективність управлінських рішень знижується у зв'язку з нехтуванням сумісним розглядом метричних і неметричних ознак або окремою їх обробкою, що обумовлено особливостями наявного математичного інструментарію. Так, за даними інформаційно-аналітичної системи соціально-економічних показників (ІАССЕП) ЦЕМІ РАН [339], соціально-економічний стан країн світу елементарно характеризується 16 основними соціально-економічними ознаками, виміряними на метричних і порядкових шкалах і вираженими 5 показниками, що описують стан населення, праці, суспільства (IMD); 4 показниками стану науки й технології (IMD); 5 показниками стану макроекономіки (IMD); 2 показниками стану фінансів (IMD). Величини наведених показників вимірюються в різних шкалах. Для прикладу аналізу стану населення, праці, суспільства Німеччини на рис. 1.1 наведена динаміка трьох порядкових ознак (x_1– якість життя; x_2– економічна грамотність; x_3 – "відплив умів") і двох метричних ознак (x_4– частка безробіття в загальній робочій силі (%%); x_5 – частка зайнятих у загальному населенні (%%)), а також динаміка вимірників.

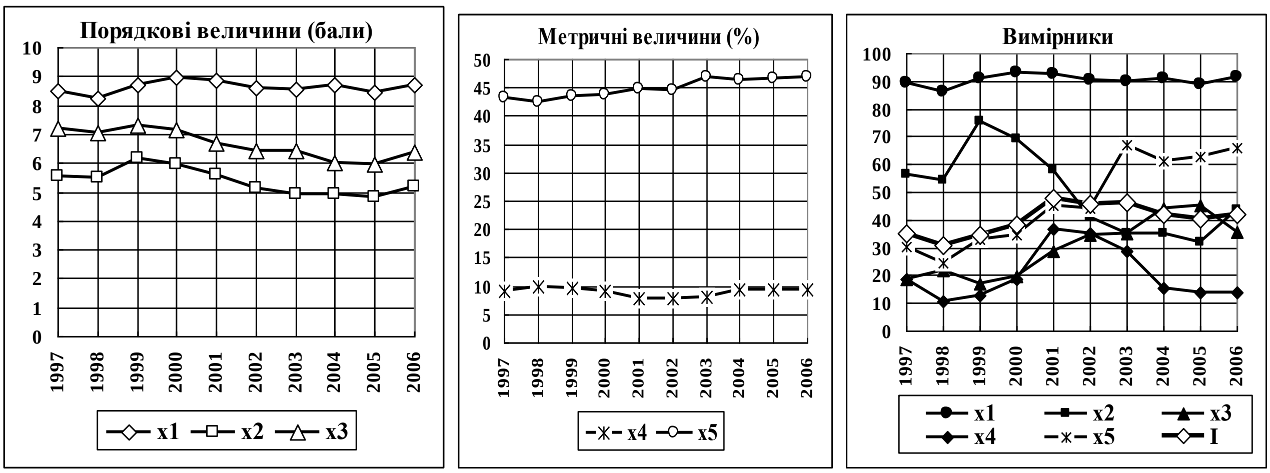


Рис. 1.1. **Динаміка показників і вимірників стану населення, праці, суспільства Німеччини**

Окремий порівняльний аналіз показників однакової розмірності в статиці й динаміці можливий, але якщо показники мають різну розмірність, більше того, ознаки, виміряні на різних шкалах, то перевага використання вимірників для порівняльного аналізу очевидна.

У практичних застосуваннях статистичний аналіз найчастіше є багатовимірним, що виокремився в інший напрямок розвитку математичних методів на межі XIX і XX ст. Вважається, що й сьогодні основна частина методів активно доопрацьовується [23; 104; 107; 115].

Хронологія розвитку методів БСА розпочинається ще з III ст. до н. е., коли Аристотель запропонував багатовимірний підхід до класифікації предметів за їх подібністю та відмінностями. З XVIII до початку XX ст. особливо поширеним був багатопараметричний опис об'єктів. Розвиток багатопараметричного опису пов'язується з іменами наступних учених: ботаніка М. Адамсона (60-ті роки XVIII ст.), англійського дослідника природи Ч. Дарвіна, хіміка Д. І. Менделеєва.

Становлення багатовимірної статистики як окремого розділу в математичних методах аналізу даних розпочато з появою в 1901 – 1904 роках наукових статей англійських вчених К. Пірсона і Ч. Спірмена, які були присвячені теорії факторного аналізу. Великий внесок у розвиток факторного аналізу зробили праці вчених Г. Кайзера, Л. Л. Терстоуна, К. Холзінгера, Д. Максвела, С. Р. Рао, Г. Хармана; кластерного аналізу – Р. Тріонона, Р. Льюїса, Р. Сокала, У. Уіл'ямса, М. Жамбю; багатовимірного шкалювання − Дж. Б. Краскала, Р. В. Хемінга, Л. Гутмана; дискримінантного аналізу – Р. Фішера, Т. В. Хейка, В. Р. Клека; загальних основ багатовимірної математичної статистики – Л. Гурмана, О. Андерсена, П. Махаланобіса, С. Уїлкса. Відомі американські школи відпрацьовували факторний аналіз, багатовимірне шкалювання, сучасні статис-тичні теорії – нечітких множин, шляхового аналізу, англійські школи – факторний аналіз, дискримінантний аналіз, багатовимірний кореляційно-регресійний аналіз, багатовимірну математичну статистику, французька школа – кластерний аналіз. Теоретичне доопрацювання багатовимірного статистичного аналізу міститься в працях відомих російських учених А. Я. Боярського, С. А. Айвазяна, П. Ф. Андруковича, А. М. Диброва, А. А. Френкеля, І. І. Єлисеєвої, А. І. Орлова, І. С. Єнюкова, Б. Г. Міркіна, І. Д. Манделя, Л. Д. Мешалкіна, В. М. Бухштабера, В. С. Мхітаряна та ін. [191, с. 15 – 22]. В Україні вирішенню проблем багатовимірного статистичного аналізу присвячені роботи відомих аналітиків В. С. Пономаренка, Т. С. Клебанової, В. В. Вітлінського, О. І.Черняка та ін. [24; 62; 95; 304].

Аналіз даних СЕС для вивчення їх визначальних характеристик – ознак – і отримання нових знань про них обумовлюється способами одержання величин, а вірніше, їх вимірювання. Тому вирішення проблем аналізу даних залежить від рівня вирішення проблем вимірювання величин в економіці. Аналіз стану теорії та практики вимірювання в економіці свідчить, що проблеми вимірювання були й залишаються центральними в методології економічної науки [18; 173; 209; 215]. Філософи науки визнають, що сучасна наука, у тому числі й економіка, виросла з вимірювання, без якого вона немислима, і затвердила себе тільки завдяки вимірюванню [12; 84; 85; 93; 206; 236]. Теорія вимірювання визнається задовільною, якщо вона формалізована на базі достатньо глибоких концепцій та має абсолютне застосування [17, с. 3].

Для систематизації знань про визначення величин в економіці був проведений аналіз ідей загальної теорії вимірювання та її проекцій у різних науках [124; 134; 149].

Як наукову теорію вимірювання почали сприймати тільки в XIX ст., при цьому розглядали тільки фізичні величини. Але джерела, що сформували теорію вимірювання, виникли набагато раніше. Вимірювання зводилось до рахунку перервних і вимірювання в одиницях міри неперервних величин. У цілому за гносеологічним критерієм погляди вчених різних часів розділяються на дві групи, що реалізують два різних підходи в розумінні сутності вимірювання та на синтезі яких автор розробляв концептуальний підхід в аналізі даних, що має підґрунтям вимірювання в економіці. Для розуміння передумов формування даного концептуального підходу коротко розглянемо зміст кожного підходу. Перший підхід, його можна назвати традиційним, має в якості джерел розробки античну науку. У більш повній формі він представлений концептуальними положеннями Б. Расела, котрий вимірювання бачив як числове подання величин [84]. Головним у цьому підході під час трактування поняття "вимірювання" є те, що воно є вимірюванням величин. Передбачається, що існує щось, що можна виміряти, а саме величина. У центрі цього підходу знаходиться поняття величини; серед усіх об’єктів дійсності вимірювання відбирає для себе один клас об’єктів. Є клас величин, і все, що не є величиною, не піддається й не підлягає вимірюванню. Вимірювання обмежується вимірюванням величин. Цей підхід іноді умовно називають дескриптивним, оскільки вимірюваний об’єкт – величина, а завдання вимірювання – опис, визначення існуючої величини.

Другий підхід, який найбільше представлений у працях Н. Кемпбела, визначав вимірювання як приписування чисел для подання властивостей об’єктів відповідно до законів науки [325; 326]. Даний підхід передбачає, що вимірювані об'єкти не мають ніяких числових властивостей і в процесі вимірювання цим об'єктам надаються числові властивості та приписуються числа. Цей підхід іноді називають конструктивним, оскільки числові властивості створюються, конструюються в процесі вимірювання. Дане розуміння вимірювання дуже абстрактне, воно зводиться до арифметизації (тобто до зіставлення чисел і об'єктів). Основна проблема вимірювання, згідно з поглядами другого напряму, полягає в тому, щоб показати, що дана емпірична область виявляється тією ж самою структурою, що й певна арифметична система чисел, а якщо ідентифіковано загальну структуру, то можна говорити, що арифметична система ізоморфна емпіричній області. Після того як ізоморфізм установлений, питання відносно емпіричної області можуть бути віднесені до арифметичної системи й до розрахунків, які зроблені в ній, а потім результати перетворені зворотньо й інтерпретовані.

Формальна концепція теорії вимірювання представлена в працях Д. Скота й П. Суппеса, Дж. Зінеса, Д. Кранца, Р. Льюїса, С. Кангера [227; 265; 328; 331; 332; 336; 337]. У роботах останніх десятиліть [1; 2; 190; 199; 203; 210; 230; 233] доводиться, що суть вимірювання полягає не в числовому поданні самому по собі, а в тому, що в процесі цього числового подання властивості виділяються, зіставляються, впорядковуються, підпорядковуються відношенням порядку. Число повинне існувати не заради числа, а як інструмент упорядкування, зіставлення.

Два, на перший погляд, абсолютно різних підходи, на думку авторів, разом відображають існуючі дві нерозривні сторони процесу вимірювання: з одного боку – число, величину, інтенсивність, а з іншого – порівняння, зіставлення, упорядкування, порядок. Така двоєдиність має утворити потужну методологічну базу теорії вимірювання в економіці, якщо фундуватися на основних ідеях даних двох напрямів розвитку теорії вимірювання.

Проблема вимірювання, як і будь-яка наукова проблема, повинна розглядатися в трьох аспектах – філософсько-гносеологічному, теоретико-методологічному і практичному. Всі ці три проблеми паралельно вирішувалися відомими вченими. Слід зазначити, що в загальній теорії вимірювання донині ігноруються завдання та обґрунтування теоретичних засад вимірювання нефізичних величин. Основна перепона у визнанні вимірювання нефізичних величин математиками вбачається в неможливості виконання адитивних операцій з даними величинами. Але щораз вища значущість вимірювання нефізичних величин спонукає вважати вимірюваннями будь-які операції присвоєння об'єктам чисел, що підпорядковуються певним правилам [119; 244]. Дане розуміння є основою формування сучасної репрезентаційної теорії (існують пропозиції називати її замість репрезентаційної теорії вимірювання репрезентаційною теорією даних, що важливо для ідентифікації ознак об'єкта). Вона включає в себе декілька основних розділів, що містять базові поняття, системи аксіом, теореми подання та єдності, теорію адекватності. Одна з основних ідей, що реалізуються в репрезентаційній теорії, стверджує можливість існування різних видів шкал, котрі відрізняються тими наборами відношень емпіричних об'єктів, які зберігаються в процесі відображення в числову множину. Слід звернути увагу на особливий розділ репрезентаційної теорії – теорію адекватності, яка передбачає проблеми усвідомлення операцій, які можна виконувати над одержуваними даними [38; 41; 75; 79; 184; 249]. Отже, репрезентаційна теорія містить багато положень, які розкривають сутність вимірювання в економіці, а отже, є підґрунтям його теорії в економіці.

В аналізі даних соціально-економічних систем величини найчастіше є нефізичними. На відміну від фізичного вимірювання нефізичне вимірювання концептуально й операційно пов'язане з людиною, має відбиток суб'єктивної складової. За виконанням процедур вимірювання в економіці доцільно розрізняти первинні і вторинні – похідні (створені) – вимірювання. Фундаментальне первинне вимірювання величини найчастіше інтерпретується як вимірювання, що не включає в себе ніяких попередніх вимірювань якої-небудь іншої величини. Похідним (створеним) вважається таке вимірювання деякої величини, яке залежить від фундаментальних вимірювань інших величин. Особливо заслуговує на увагу проблематика похідного вимірювання – системи правил одержання узагальненої та інтегральної характеристик складної якості на основі вимірювання величини елементарних ознак. Виділяють серед визначальних ознак СЕС, як мінімум, два види ознак – елементарну й складну ознаки.

Отже, існує теоретичне підґрунтя для формування цілісних концептуальних основ вимірювання в економіці, яке проявляється в наявності таких процедур в економіці, як квантифікація, формалізація, моделювання, шкалювання, оцінка, зміст яких цілісно пов'язується зі змістом процесу вимірювання [134]. Вимірювання в економіці повинне мати свою методологію, яка має відображатися в принципах та положеннях. На рис. 1.2 наведена рекомендована схема логічного взаємозв'язку методологічних принципів вимірювання в економіці.

Рис. 1.2. **Основа принципів вимірювання в економіці**

Методологічні принципи метрології збагачують загальні принципи пізнання об'єкта конкретними визначеннями процесу вимірювання в різних сферах діяльності людини, але тільки фізичних величин [37].

Отже, конструкцію концепції визначення величин ознак доцільно зводити на фундаменті триєдиної основи: концепції величин в економіці, умов їх отримання й системи вимірників за допомогою адекватних математичних методів та моделей; схематично це наведено на рис. 1.3 [134]. Названі фундаментальні складові можна розглядати як ті, що утворюють базис теорії вимірювання в економіці сьогодні.

Вирішення проблем аналізу даних на підґрунті концепції визначення величин ознак забезпечує достовірність, об'єктивність та якість результатів аналізу. Незважаючи на існування численних методів, способів, методик, концепцій, теорій аналізу та оцінки в економіці, всі вони прямо чи опосередковано передбачають вимірювання ознак об'єкта в економіці; їх можна розділити на дві групи: перша оперує з натуральними (фізичними) величинами ознак, друга – з нефізичними (вартісними) величинами ознак об'єкта та величинами некількісних ознак [238; 241; 250; 251; 261].

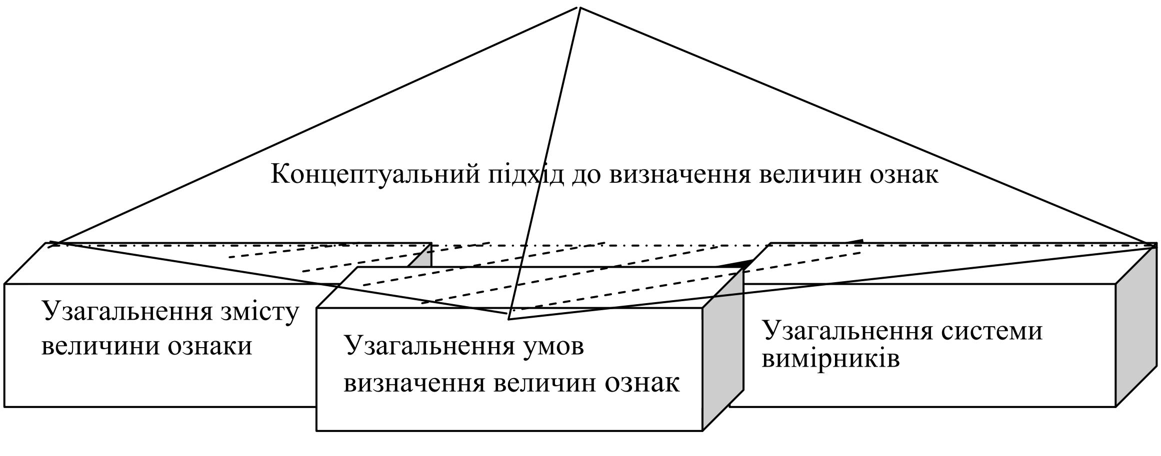


Рис. 1.3. **Базис концепції визначення величин в економіці**

Початково процедуру вимірювання слід розглядати на двох рівнях: перший – коли формуються первинні дані в економіці, і другий – коли обчислюються похідні показники. У вузькому розумінні процедурою вимірювання можна вважати процедуру тільки першого рівня, але специфіка суспільних наук (економіки, соціології) передбачає віднесення процедури і другого рівня до вимірювання. Об'єктом вимірювання першого рівня є величини затрат та випуску. Ці величини трактуються як параметри виробничих процесів (в операціях хронометражу, контролю розмірів виробів і т. д.) та реєструються приладами. Завдання вимірювання на першому рівні полягають у досягненні точності вимірювальних пристроїв та об'єктивності виконання вимірювальних процедур. Вимірювання вартісних величин ознак об'єкта зводиться до реєстрації окремих значень вартісних показників у процесах обміну та розподілу в масових і одиничних актах купівлі-продажу, з чого складається господарська діяльність суб'єктів бізнесу сьогодні [266; 267; 279]. Порівняння вартісних величин ознак з нормативними, що регламентуються державою, ринковими умовами, також є процедурою вимірювання в економіці. Доцільно розглядати дві процедури вимірювання (фізичних і нефізичних величин елементарних ознак об'єкта) як процедури первинного вимірювання в економіці, що є предметом різних видів обліку й аналізу господарської діяльності суб'єктів економіки. Результати первинного вимірювання складають базу даних для виконання процедур вторинного вимірювання з метою подальшого дослідження об'єкта. До процедури вторинного вимірювання відносять: обчислення характеристик сукупностей даних первинного вимірювання, обчислення, що виконує статистика, опосередковані (непрямі) обчислення, вимірювання складних ознак за допомогою математичних методів та моделей.

Встановлення тенденцій змін, вивчення кількісних взаємозв'язків ознак в економічних і соціальних процесах опираються на їх моделювання. У процесі моделювання, особливо на регіональному або державному рівнях, виникає взаємодія "первинних" і "вторинних" показників. Будь-яка математична модель в економіці спирається на певну систему визначених величин, що є, в основному, показниками (продукції, ресурсів, елементів і т. д.). Водночас одним із важливих результатів макроекономічного моделювання є одержання нових (вторинних) показників економічно обґрунтованих цін на продукцію різних галузей, оцінок ефективності різноякісних природних ресурсів, показників суспільної корисності продукції [13; 14; 35; 314]. Однак ці показники можуть зазнавати впливу від недостатньо обґрунтованих первинних показників, що змушує розробляти особливу методику формування первинних показників, а саме перетворення їх величин в об'єктивні вимірники. Маємо проблему перетворення величин первинного вимірювання у величини вторинного вимірювання.

Багато вчених вважають, що протягом тривалого часу головним гальмом практичного застосування математичного моделювання в економіці є проблема наповнення розроблених моделей конкретною та якісною інформацією [7; 22; 226; 252; 253]. Точність і повнота первинної інформації в первинних вимірниках, реальні можливості її збору й обробки багато в чому визначають вибір типів прикладних моделей. З іншого боку, дослідження за допомогою моделювання висувають нові вимоги до системи інформації, що містять вимірники. Методи економічних спостережень і використання результатів цих спостережень розробляються економічною статистикою. Тому слід відзначити тільки існування специфічних проблем економічних спостережень, пов'язаних із моделюванням економічних процесів [54; 73; 263; 268; 293; 321]. У фізиці в процесі вимірювання ваги чи температури існує можливість перевірки та повторної експериментальної перевірки результатів випробувань. Таким чином, питання обґрунтованості даних втрачає свою проблематику. У вимірюваннях об'єктів в економіці обґрунтованість даних – одна з найважливіших проблем, оскільки у вимірюваннях об'єктів у фізиці мають справу із самими величинами властивостей, а у вимірюваннях в економіці – з показниками ознак з огляду на неекспериментальний характер даних.

У теорії вимірювання визначальним фундаментальним поняттям, на основі якого вибудовується весь методологічний та методичний каркас, є величина. Основні положення концепції величини ознак об'єктів в економіці такі.

Основною відмінністю величин в економіці є існування великого класу нефізичних величин і, як наслідок, екстенсивних та інтенсивних величин. Для формалізації системи екстенсивних та інтенсивних величин економіки автори простежили концептуалізацію зв'язку величини з кількістю, якістю і числом. Розглянемо основні висновки проведеного аналізу.

Пізнання об'єкта здійснюється за допомогою абстрагування його особливостей, що значимі для вирішення практичного завдання. У результаті такого абстрагування у свідомості людини формується модель об'єкта, що відтворює структуру та властивості останнього, а також взаємозв'язки між його елементами [120; 287]. Для того щоб оцінити об'єкт, потрібно виміряти його якість, що акумулюється у властивостях, які він має. Отже, об'єкт вимірювання знаходиться скрізь, де знаходиться об'єкт управління в економіці, тобто на всіх ієрархічних рівнях системи управління в економіці. Слід виміряти величину властивості, а точніше, її формалізованого виду – ознаки. Ознаки характеризують речі, явища та об'єкти в економіці, які можуть бути простими чи складними, тобто вичерпно характеризуватися однією елементарною ознакою чи системою ознак, інакше кажучи, бути одновимірними або багатовимірними. Багатовимірність збільшується й ускладнюється зі зростанням рівня ієрархії системи управління.

Адекватно рівню управління в економіці виділяють безпосереднє вимірювання величини елементарної ознаки (пряме, первинне чи безпосереднє, фізичне вимірювання) та непряме вимірювання (створене, похідне, опосередковане), що використовує результати безпосереднього вимірювання, при цьому реалізується технологія саме процесу вимірювання. Звідси випливає, що зазвичай об’єктом первинного вимірювання математики в економіці є величина елементарної ознаки об’єкта, а об’єктом створеного вимірювання – величина складної ознаки об’єкта, який може належати різним рівням ієрархічної системи управління. Отже, об’єктом первинного вимірювання в економіці є величини кількісних та якісних елементарних ознак. Первинно вимірюється фізична величина кількісної ознаки – це екстенсивна величина елементарної ознаки. Кількісні ознаки відрізняються від якісних зміною значень величини та своєю мірою. Для якісних ознак не виконується жодна з умов існування кількісних ознак. Величини якісних елементарних ознак та кількісних складних ознак (що можна розглядати за ступенем накопичення кількості) утворюють вид інтенсивних величин в економіці. Ознаки можуть характеризувати або один об’єкт, або множину, і звідси випливає відмінність індивідуальних та групових, сукупних ознак.

Форма існування величини в економіці – показник. Щоб надати форму у вигляді показника, необхідно враховувати тип величини, її види і специфікацію в даному конкретному випадку об’єктивізації ознаки. Отже, щоб формалізувати ознаку об’єкта в економіці, потрібно навести описову (атрибутивну) характеристику властивості об’єкта, метричну або неметричну величину. У величині числа подають кількісну визначеність: у неметричних величин – порядок, упорядкованість; у метричних – ще й величину в розумінні "розмір". Поняття іменування, номінації є необхідною описовою складовою будь-якого визначення величини. Якісна визначеність метричних та неметричних величин описується у формі іменування. Числова складова неметричних величин утворює послідовність іменованих порядкових чисел, їх можна трактувати як одиниці впорядкування окремих ступенів їх якісних елементів.

Отже, кількісну визначеність величин знаходять на основі відомих процедур порівняння однорідних властивостей об’єктів, тобто введенням необхідної дискрети (одиниці) та певної шкали величини, що узгоджує різні ступені прояву властивості з визначеними числами. Для адекватного відображення величин різних ознак існують різні типи шкал. Розробка узгоджених систем загальноприйнятих шкал є однією з перших ознак переходу до наукового періоду розвитку суспільства [230, с. 13]. Виміряти все, що вимірюється, і зробити вимірюваним усе, що таким ще не є, – відома програма точного природознавства, окреслена Г. Галілеєм ще у XVII ст. За шкалами визначають величини ознак об'єкта, вони є інструментом, за допомогою якого досягається точність вимірювання [98].

Будь-яке вимірювання відносне. Воно варіюється за видом і ступенем, за типом і точністю. Побудовою шкали величини завершується важливий етап процесу пізнання, що пов'язаний з виділенням окремих властивостей та встановленням їх особливостей у різних ситуаціях, – так звана арифметизація простору [289]. У техніці на роль фундаменту вимірювання претендували теорія інформації, теорія сигналів, теорія зворотних перетворювачів та інші, але більшість учених визнали саме сучасну теорію шкал основою для загального вимірювання [99]. В. Г. Кноррінг, визнаний спеціаліст з проблем теорії шкал, вважає причиною те, що вимірювання не є передачею інформації, зміною форми сигналу або перетворення енергії, точніше, не зводиться до цих процесів, його сутністю є перехід від світу фізичних реальностей до системи знаків, які відображають реальність [99]. Для вимірювання важливо встановити взаємно однозначне відображення однієї системи з відношеннями в іншу, при цьому виконання певного відношення між елементами однієї системи спричиняє виконання відповідного відношення між елементами іншої системи, і навпаки; у математиці, як відомо, даний факт називається ізоморфізмом. Так, у процесі формування даних в економіці завжди намагаються, щоб абстрактна система, яка їх утворює, відображала реальну (емпіричну) систему ізоморфно.

Перехід від змістовного опису об'єктів дослідження та їх властивостей до моделей об'єктів і величин супроводжується формалізацією даного опису із залученням математичного апарату, таким чином, будь-яка величина може бути подана як певний параметр математичної моделі. Часто в економіці величина є модельним поняттям, а виділення окремих груп ознак – поведінкових, соціальних, системних, фізичних – передбачає додаткову характеризацію параметрів, підтвердження розкриття специфіки об'єкта дослідження чи його моделі. Фізичні величини відображають об'єктивні властивості об'єкта дослідження, а показники якості об'єкта – їх суспільну значимість у конкретних умовах [119, с. 21].

Елементи методології визначення величин складного об'єкта в економіці з точки зору системи його функціональних якостей, ознак можна бачити в прикладному математичному розділі – кваліметрії. Цей розділ математики вивчає найзагальніші принципи та способи комплексної оцінки якостей продукції та ефективності діяльності в цілому. Принципи комплексного вимірювання якості, що ґрунтуються на знаходженні та додаванні вихідних відносних оцінок, розроблені кваліметрією для вимірювання якості продукції. Вчені, які працюють над проблемами кваліметрії, вважають, що за допомогою таких підходів можна інтегрувати будь-які якості. Без жодного сумніву, принципи кваліметрії доцільно враховувати для побудови узагальнюючого показника.

Викладені в посібнику рекомендації щодо вимірювання величин стандартизують процес визначення величин ознак в економіці. Основною частиною даного процесу є процедури економіко-математичного моделювання. Практика моделювання в процесі розв'язання реальних задач економіки спонукає рекомендувати дванадцять основних етапів моделювання [126; 145; 157].

У наступних розділах посібника будуть викладені розробки щодо визначення величин ознак соціально-економічної системи. Показники та критерій ефективності функціонування соціально-економічної системи мають відображати загальносистемні ознаки, структурні ознаки та функціональні ознаки, які визначаються процесом та цілями функціонування системи. Перераховані ознаки є, в основному, складними ознаками системи.

Визначення величин ознак СЕС спрямоване на отримання їх фізичних параметрів, а саме: динамічних, контрольних, прогнозних, оцінювальних, які слугують, відповідно, основою побудови поведінкової моделі.

Для реалізації процесу визначення величин потрібні принципи, постулати його здійснення. Сформована система таких принципів, що методологічно забезпечує технологію вимірювання ознак об'єктів в економіці та фундує теорію, наведена в табл. 1.1. Дієвість наведених принципів залежить від дотримання постулатів у вимірюванні в економіці. У процесі вимірювання величин ознак можуть виникнути похибки, класифікація яких згідно з етапами технології визначення величин в економіці наведена в роботі [138].

Наведені постулати та принципи надають об'єктивності та фундаментальності методології моделювання соціально-економічних систем, забезпечують обґрунтованість концепцій та формують змістовність отриманих практичних результатів визначення величин в аналізі даних.

Таблиця 1. 1

Принципи та постулати вимірювання в економіці, на основі яких розроблено загальну технологію визначення величин різних ознак СЕС

Принципи Вимірювання	Постулати вимірювання в економіці	Процедури технології визначення величин ознак
1. Філософський розгляд об'єкта та його властивостей у процесі вимірювання в економіці. 2. Збереження та дотримання наукових основ (методологічних та методичних) економіки у вимірюванні ознак об'єктів. 3. Кібернетичність, системнологічність у розробці моделей та використанні математичних методів у технології визначення величин в економіці. 4. Метрологічність у реалізації операції та процесу вимірювання в економіці	Прийнята концептуальна модель об'єкта обумовлює процес вимірювання величин елементарних та складних ознак; багатовимірність об'єктів в економіці передбачає множину вимірюваних величин та їх істинних значень; існують показники в економіці, що є формою величин ознак об'єктів; складна ознака вимірюється сумісно й виражається синтезованою величиною; вимірювання складних ознак здійснюється за допомогою системи вимірників; якщо враховується дія законів і закономірностей в економіці й виконуються загальні положення теорії вимірювання, то вимірювання ознак об'єктивне та істинне; вимірюються детерміновані величини або умовно детерміновані; в економіці існують фізичні та нефізичні величини, метричні й неметричні величини; дослідження неметричних ознак об'єкта, визначення нефізичних величин розширює пізнання об'єкта в економіці; множина нефізичних величин залежить від ступеня врахування складності ознак, їх взаємозв'язків; рівень нефізичної величини в економіці має суспільну визначеність і отримується статистично; в економіці є великий клас статистичних величин, для визначення яких існують спеціальні методи; існує неточність первинного вимірювання елементарної ознаки, що залежить від прийнятої шкали; існує неточність вимірювання складної ознаки, що залежить від її моделі; істинне значення величини встановлюється з певною невизначеністю та потребує повторного вимірювання; вимірювання величин залежить від людини, тому потрібні правила, норми, нормативи і закони, які б регламентували процес та інструменти вимірювання в економіці; вимірювання в економіці евристичне; загальноприйняті правила побудови моделей об'єктів та прийнята технологія визначення наближають величини ознак до істинних; вимірювання в економіці методологічно має бути спрямованим на створення системи основних величин, яка уможливлює узгоджену дію механізмів управління на різних його рівнях	1. Процедура постановки. 2. Процедура підготовки. 3. Первинне вимірювання. 4. Вторинне вимірювання. 5. Процедура контролю за похибками, які розподіляються на: методологічні (пов'язані з парадигмою, принципами вимірювання в економіці та моделлю об'єкта), методичні (пов'язані з методиками визначення ознак, організацією технології, її етапами реалізації, процедурами вимірювання, з аналітичними, математичними методами), технічні (пов'язані з похибками окремих приладів, технічними системами та технічними засобами), особистісні похибки (обумовлені індивідуальностями людини, експерта під час встановлення еталонів мір в економіці)

1.2. Основи методології опису соціально-економічних систем у багатовимірному просторі їх ознак

Опис СЕС за допомогою системи елементарних і складних, метричних і неметричних ознак є повномасштабним та достовірним. Складні ознаки СЕС моделюються. Побудова кожної моделі спирається на існуючу теорію об'єкта. Ця теорія вміщується спочатку в методологічні межі дескриптивного подання соціально-економічної системи, на основі якого далі розбудовуються різні прескриптивні моделі. Необхідною умовою в переході від об'єкта, що досліджується, до дослідження його моделі та подальшого перенесення результатів на об'єкт дослідження є вимога адекватності моделі й об'єкта [195, с. 56]. Адекватність передбачає відтворення в моделі з необхідною повнотою всіх характеристик об'єкта для його подальшого дослідження. Оцінка ступеня схожості є оцінкою відмінності моделі від реального об'єкта. Поняття адекватності ґрунтується на математичних поняттях ізоморфізму та гомоморфізму. Саме ці вимоги є основними у визначенні величин ознак, у використанні шкал для вимірювання. Ідентичність моделі й об'єкта в техніці перевіряється мінімальною різницею їх вихідних сигналів, якщо подаються однакові вихідні впливи на об'єкт та його модель [76; 239; 273]. В економіці адекватність описових моделей обумовлюється логікою визначення величин, ідентифікацією складних ознак, встановленням конструкції взаємозв'язків між ознаками й перевіркою її за статистичними критеріями. Отже, спрямованість методології описового моделювання СЕС на визначення системи ознак максимально наближає модель до реального об'єкта в економіці.

У техніці ідентифікація та опис об'єкта залежать від обсягу інформації про структуру об'єкта і від обсягу інформації, що вимірюється. Обидва види інформації необхідні для синтезу моделі, що буде розбудовуватися, та мають різний ступінь важливості. Апріорні знання про структуру об'єкта допомагають визначити структуру моделі. Цю процедуру називають ідентифікацією в широкому розумінні, або структурною ідентифікацією. Наявність виміряних величин параметрів є підґрунтям для параметричної ідентифікації [40; 44; 237; 255; 312].

Загальні ідеї ідентифікації в описі об'єктів у техніці підтверджують правомірність використання концепцій вимірювання в економіці для обґрунтування базисних моделей опису соціально-економічних систем. Отже, розвиток описового моделювання соціально-економічних систем за новою методологією спрямований на відтворення структури системи, її елементів, взаємозв'язку між ними, складових на основі визначення та виявлення ознак, величини яких мають вимірюватися, перетворюватися та бути апріорною інформацією для подальшого математичного моделювання СЕС.

Основні положення методології модельного базису опису соціально-економічних систем мають такий зміст [221; 223]:

1. Методологія розбудована за окремим понятійним базисом, який складається з основних означень і утворює її теоретико-понятійний каркас. У новій методології моделювання введені наступні визначення.

Елементарна ознака соціально-економічної системи – це характерна властивість системи, яка первинно вимірюється і яка структурно визначається не більше, ніж двома ознаками, що елементарно не визначаються іншими.

Елементарна структура соціально-економічної системи – це структура, що утворена елементарними ознаками та їх парними взаємозв’язками.

Складна ознака соціально-економічної системи – це характерна властивість системи, яка вторинно вимірюється і яка структурно визначається елементарними або іншими складними ознаками, тобто є багатовимірною й зазнає сукупного впливу інших ознак.

Структура зі складних ознак соціально-економічної системи – це структура, яка утворена складними ознаками та їх взаємозв’язками. Структура ознак СЕС утворюється елементарними, складними ознаками та їх взаємозв’язками (рис. 1.4).

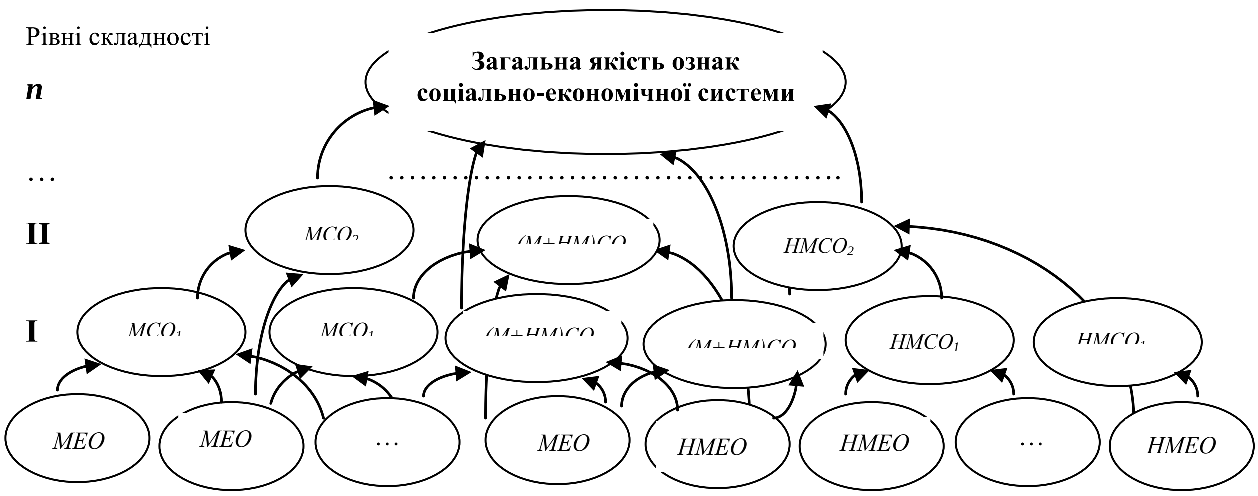


Рис. 1.4. **Схематичне подання структури ознак соціально-економічної системи:** MEO – метрична елементарна ознака; $HMEO$ – неметрична елементарна ознака; MCO_i – метрична складна ознака i-го рівня складності; $HMCO_i$ – неметрична складна ознака i-го рівня складності; $(M+HM)CO_i$ – сумісна складна ознака i-го рівня складності

Нульовий рівень складності структури ознак – це рівень елементарних ознак у структурі.

Перший рівень складності структури ознак – це рівень складних ознак, що утворений елементарними ознаками; n-й рівень складності структури ознак – це рівень складних ознак, що утворений різними ознаками, у тому числі й ознаками $n-1$-го рівня складності.

Технологія визначення є процесом практичної реалізації вимірювання величин ознак, за яким передбачається побудова математичних моделей складних ознак соціально-економічних систем.

Відображення соціально-економічної системи за її ознаками – це вираження реальної структури соціально-економічної системи структурою її ознак, що утворена елементарними ознаками, які мають взаємозв’язки й формуються у складні ознаки першого рівня складності, та складними ознаками, що мають парні й непарні взаємозв’язки і формуються у складні ознаки вищого рівня складності, ніж перший.

Простір елементарних ознак опису соціально-економічної системи $\boldsymbol{EO} = \{EO_j\}, j = \overline{1,m}$ – це кінцевовимірний простір, осями якого є елементарні ознаки системи.

Ознаковий простір, або простір ознак опису соціально-економічної системи $\boldsymbol{O} = \{O_j\}, j = M, M > m$ – це кінцевовимірний простір, осями якого є різні ознаки системи.

Величина ознаки соціально-економічної системи – це метрична або неметрична величина, яка вимірюється відповідно в метричних або в неметричних шкалах.

Метрична ознака – це ознака, величина якої вимірюється в метричних шкалах.

Неметрична ознака – це ознака, величина якої вимірюється в неметричних шкалах.

Показник – це форма метричної величини ознаки соціально-економічної системи. Звідси показники є окремими, що виражають елементарну ознаку, та узагальнюючими, що виражають складну ознаку.

Вимірник – це перетворена змодельована величина ознаки соціально-економічної системи за допомогою неперервних та табличних функцій перетворення.

Математична модель ознак соціально-економічних систем – це математичне подання, вираження соціально-економічних систем за допомогою їх ознак та величин ознак для аналізу даних.

Математична модель метричних складних ознак соціально-економічної системи ($M_{СМО}$) – це математичне подання системи, відображення за допомогою математичних знаків зв'язків елементарних метричних ознак соціально-економічної системи чи результатів їх взаємозв'язків, а також параметрів, що виражають форму зв'язків.

Математична модель неметричних складних ознак соціально-економічної системи ($M_{СНМО}$) – це математичне подання системи, відображення за допомогою математичних знаків зв'язків елементарних неметричних ознак соціально-економічної системи чи результатів їх взаємозв'язків, а також параметрів, що виражають форму зв'язків.

Математична модель складних ознак соціально-економічної системи ($M_{СО}$) – це математичне подання системи, відображення за допомогою математичних знаків зв'язків елементарних метричних та неметричних ознак, їх змішаних систем чи результатів взаємозв'язків у системах ознак, а також параметрів, що виражають форму зв'язків.

Модельний базис опису соціально-економічних систем – це типові узагальнені математичні моделі, що описують соціально-економічні системи за їх ознаками та величинами ознак і є підґрунтям для продовження моделювання систем.

2. У вимірюванні величин в економіці реалізується ізоморфно-гомоморфне відображення ознак реального об'єкта в числові форми цього об'єкта.

Математичні методи та моделі широко застосовують у вимірюваннях величин у техніці, оскільки вони необхідні для коректного виконання підготовчих етапів вимірювання, обробки експериментальних даних та формалізованого опису вимірювальної процедури в цілому [119], але у вимірюваннях в економіці вони, поряд з наведеним призначенням, формують умови проведення всіх процедур визначення величин з урахуванням неекспериментального характеру отримання величин. На основі знакового формального моделювання мовою математики, яким є математичне моделювання, попередньо виявляються ознаки, їх системи, які характеризують об'єкт, і таким чином створюються умови для проведення обчислень величин і надаються нові знання про СЕС. У загальній технології визначення величин ознак в економіці реалізуються методологічні принципи математичного моделювання: адекватність, динамізм, заміщення та евристичність. Дані принципи реалізуються у розбудові моделей різних типів. Дотримання принципу адекватності (об'єктивної відповідності моделі оригіналу) сприяє об'єктивній істинності отриманого знання про ознаки та об'єкт у цілому, а отже, істинності результатів вимірювання. Принцип динамізму стверджує релятивність у пізнанні об'єкта за допомогою його моделі та вказує на змінність усіх основних компонентів модельно-пізнавальної ситуації. У зв'язку з цим модель вимірників для визначення величини ознак передбачає змінність окремих параметрів, обумовлених розвитком його ознак. Саме блочно-модульна структура побудови моделі забезпечує адаптивність та гнучкість у моделюванні. Як наслідок, випливає необхідність системного моделювання об'єкта, а точніше, системологічного моделювання. Як реалізацію принципу заміщення можна розглядати проведення модельних експериментів за допомогою процедур визначення величин, а саме програвання ситуацій у досягненні окремих рівнів величин ознак об'єктів, що є, на думку авторів, великим здобутком у запобіганні розвитку негативних процесів в економіці. Це сполучається з принципом евристичності, а саме з отриманням нового знання про ознаку й об'єкт. Розширене відтворення знання за допомогою моделі фундує майбутні джерела його отримання.

Дотримання основних методологічних принципів математичного моделювання у відповідних процедурах визначення величин ознак обґрунтовує сам процес визначення величин в економіці і тим самим підтверджується нерозривність визначення ознак з математичним моделюванням. Визначені величини ознак реального об'єкта ідентифікують ознаки соціально-економічних систем.

Виокремлення первинного і вторинного етапів вимірювання в економіці, а саме вимірювання величин елементарних метричних та неметричних ознак і визначення величин складних ознак різних рівнів складності об'єкта, передбачає виявлення та визначення структури складних ознак на основі розбудови моделей самих ознак, таким чином повномасштабно визначаючи структуру соціально-економічної системи.

Існують особливості вимірювання в економіці. Математичні особливості пов'язані з відображенням реальної системи в числову. Першим етапом є встановлення емпіричної системи з відношеннями, тобто виділення множини об'єктів, що вивчаються, та властивостей їх відношень. Потім підбирається числова система з відношеннями і гомоморфізм емпіричної системи в числову, що встановлює адекватну відповідність між відношеннями емпіричних об'єктів та відношеннями числових символів із числової системи. У суспільних науках спочатку постулюються вид та властивості числової системи, потім визначається клас допустимих перетворень вимірювання, а потім саме вимірювання, тобто функція, що зіставляє числа з ознаками вимірювання. Проблема адекватності вимірювання спочатку зводиться до проблеми несуперечливості отриманих результатів. Саме тому вимірювання в суспільних науках не стали до цього часу фундаментальними. Визначальним у вимірюванні в техніці, як було вже сказано, є побудова шкали величини, що вимірюється, у вимірюванні ж в економіці спочатку говорять про метод вимірювання, його формалізацію. Формальне визначення методу вимірювання в економіці маємо у роботах [17; 18].

Визначимо метод вимірювання ознак, такий як:

$$\langle \mathscr{H}, O, f, \Phi \rangle,$$

де $\mathscr{H} = \langle H, G_H \rangle$ − числова система;

H − область системи, яка складається з набору числових символів;

G_H − сукупність відношень, що виконуються для елементів H;

O − ознаки СЕС; $f: O \to H$ функція вимірювання;

Φ − клас допустимих перетворень $\varphi: H \to H$.

Багато величин в економіці отримуються за допомогою моделювання.

Дану формалізацію підтверджує моделювання вимірника окремої ознаки. Тут у якості H розглядаються значення показників, що є формою величини ознаки; G_H – відношення між ознаками; O – множина ознак, точніше, їх моделі. Функція f – це окрема функція перетворення, якщо ознака елементарна, або узагальнена функція перетворення, якщо ознака складна. Множина Φ, що постулюється, висуває вимоги до вимірювання елементів множини O. Тривіально розглядати елементарні ознаки, що виміряні за допомогою метричних шкал. Якщо наявні якісні ознаки, що виміряні за допомогою неметричних шкал, а точніше, ординальних, то вони також задовольняють умови описаного методу вимірювання, оскільки величину, яка виміряна метричними шкалами можна розглядати в економіці як накопичення ступенів інтенсивності, а це не що інше як ранги. Проблема полягає в доведенні методу вимірювання для атрибутивних та видових величин.

Відомо, що коли заданий метод вимірювання $\langle \mathscr{H}, O, f, \Phi \rangle$, то на множині початкових об'єктів O виникає система відношень G_O, яка відповідає числовій системі G_H. А саме, будь-яке n-не числове відношення $P_H \in G_H$ породжує n-не відношення P_O на множині O, що визначається умовою:

$$(o_1, o_2, \ldots, o_n) \in P_O \text{ тоді, і лише тоді, коли } (f(o_1), f(o_2), \ldots, f(o_n)) \in P_H.$$

У прикладі з вимірюванням ознаки на основі моделювання вимірника ознаки дана умова виконується завдяки властивостям функції перетворення (для елементарної ознаки – одностороння зростаюча чи спадна або двостороння функція перетворення, для складної ознаки – властивості середнього геометричного, за допомогою якого будується узагальнена функція перетворення). Ф. М. Бородкін і Б. Г. Міркін [17] дієвість формалізації методу вимірювання продемонстрували на методі вимірювання зв’язку змінних за допомогою коефіцієнта детермінації, який визначається як квадрат коефіцієнта кореляції.

Посилання методів вимірювання в суспільних науках відмінні від посилань теорії шкал у технічних науках. Визначальною перевагою теорії методів порівняно з теорією шкал є те, що в теорії методів вимірювання системи відношень S_A на ознаках, що вимірюються, розглядаються як під час первинних, так і під час вторинних вимірювань, і вони рівноправні, на відміну від умов РТВ. Даний факт надзвичайно важливий для ствердження загальних основ визначення величин ознак в економіці, їх застосувань у методології описового моделювання. У визначенні величин ознак в економіці передбачається ідентифікація й опис ознак соціально-економічної системи.

3. Формування концептуальної моделі об’єкта в економіці для визначення його величин ознак є концептуальним моделюванням соціально-економічної системи.

Дане положення пов’язане з першим, але інакше за змістом. Для визначення величин ознак об’єкта потрібно його концептуалізувати. Вчені, що досліджують проблеми управління в економіці, рекомендують суб’єкти бізнесу (у нас – об’єкти), ознаки яких визначаються, розглядати як соціально-економічні системи, визначаючи відразу два типи ознак: соціальні $\boldsymbol{O} = \{O_j\}$, $\boldsymbol{O} = \{O_j\}, j = M_1,\ M_1 \in M$ й економічні $\boldsymbol{O} = \{O_j\}, j = M_2,\ M_2 \in M$. Узагальнено, що поняття "система" в основному застосовується в декількох випадках: 1) це система, що вивчається чи є керованою; 2) це система, яка здійснює вивчення чи керує; 3) це система, яка підтримує керівну систему, а в окремих випадках вона може бути й складовою частиною керівної системи [8; 39; 106; 208; 217; 220; 288]. Будемо розуміти систему в першому значенні. У роботі [134, с. 60 – 61] наведені основні типи визначень системи, що мали принципове значення для системологічного подання об’єкта в економіці.

Наведене розуміння системи в економіці з метою визначення величин її ознак узгоджується із сучасним системологічним підходом у науці. З точки зору теорії організацій системи в економіці – це соціально-економічні системи, які в ракурсі інженерного підходу до бізнесу називаються бізнес-системи. Моделюючи об’єкт системно, в зміні значень величин його ознак слід розпізнати форму прояву різних типів загальних законів в економіці: руху, структури, казуальності, динамічності, статичності і т. д. [32; 88; 92; 96]. Такий підхід гарантує уникнення спрощеного бачення об'єкта в економіці та виключає некоректність у моделюванні, що випливає з автоматичного перенесення формалізованих схем об'єктів у техніці на об'єкти в економіці.

Окремо слід виділити рекурентні відношення структурних елементів складних систем в економіці, що формалізують існування в них багатофакторності [11, с. 368 – 381]. Цілісність системи в економіці, рівень системності залежать від рушійної сили факторів, їх ролі в механізмі функціонування й розвитку як окремих частин цілого, так і системи в цілому. З яких би частин з різнорідними структурними відношеннями факторів не складалась система, її органічна побудова як цілого враховує за допомогою матриці рекурентних відношень структурних елементів системи індивідуальні особливості та пропорції частин, що входять до системи. Насправді фактори в економіці виражаються натуральними та вартісними показниками. Рух від відтворювальної структури частин до органічної побудови цілого зазвичай починається з натуральних чи вартісних показників елементарних ознак, які розкривають внутрішню логіку досліджуваного об'єкта [27; 34].

Аналітичну роботу визначення величин ознак в економіці розпочинають зі встановлення меж досліджуваного об'єкта; функціонального призначення об'єкта як організації, її структурних елементів; критерію та принципів функціонування самих механізмів розвитку об'єкта; видів і різновидів розподілу праці, спеціалізації, умов кооперації структурних підрозділів; вхідних керівних сигналів та потоків інформації на кожному ієрархічному рівні управління; системи вхідних матеріальних потоків і регуляторів їх перетворення; вихідних матеріальних потоків; комунікації прямих та зворотних зв'язків між фізичними складовими об'єкта управління. Для структурного аналізу складних систем встановлюється склад мажорант, вхідних, керівних та вихідних елементів, типи і види їх зв'язків, джерел та споживачів сигналів, визначається елементи прямого управління та підпорядкування [11, с. 368 – 397].

Сучасні методології побудови системи управління організацією на основі процесного підходу, функціонально-ієрархічного управління, методики виділення процесів в організації, способи побудови сітки бізнес-процесів, методики моделювання процесів у нотаціях IDEF0, IDEF3, DFD і ARIS залишають основними елементами об'єкт, його ознаки та їх зв'язки [263]. Моделювання їх виконується за допомогою опису атрибутів об'єктів, які відображають окремі характеристики реального об’єкта. Визнається також факт, що кожне підприємство повинне створювати таку систему з урахуванням специфіки своїх процесів [263].

Спираючись на викладені загальні правила формалізованого подання об’єкта в економіці, на рекомендації фахівців-управлінців [28; 39; 46; 77; 94; 100; 218; 219; 228; 229; 299], СЕС формально подається виразом [158; 166; 168]:

$$CEC = \left(S_C, S_E, E\left(S_i \right) F_S, A_S \right),$$

де S_C – соціальна підсистема;

S_E – економічна підсистема;

$E\left(S_i \right)$ – структурні елементи підсистем;

F_S – закони функціонування підсистем та системи в цілому;

A_S – механізми функціонування підсистем та системи в цілому;

A^S – алгоритми (механізми) функціонування підсистем.

Для детальнішої формалізації підсистем рекомендується виокремити елементи згідно з "методологічним каркасом" за їх ознаками з точки зору ухвалення рішень в економіці [33; 185; 287; 311]:

$$O = \left(O_e, O_S, O_o, O_K, O_p, O_M \right),$$

де O_e – ознаки матеріальних і нематеріальних елементів;

O_S – ознаки структур;

O_o – ознаки організацій;

O_K – ознаки конструкцій;

O_p – ознаки логічно спрямованих процесів;

O_M – ознаки механізмів у системі.

Окремий показник, що виражає елементарну ознаку в процесі первинного вимірювання, подається кортежем:

$$x_i = \left\langle \begin{array}{l} \textit{назва}, \textit{число}, \textit{одиниці вимірювання}, \\ \textit{область допустимих значень}, \textit{реперні значення} \end{array} \right\rangle .$$

Залежно від типу показника можуть бути відсутні деякі елементи кортежу.

Функціональні здатності об’єкта також проявляються через його елементарні та складні ознаки. Розвиток ознак можна подати у вигляді закону функціонування елемента F^s:

$$O_j(t) = F^s(x, n, u, h, t),$$

де $x(t)$ – вхідний сигнал;

$u(t)$ – керівний сигнал;

$n(t)$ – сигнал впливу зовнішнього середовища;

$h(t)$ – власні характеристики підсистеми [8; 183; 319].

4. Об’єкти в економіці мають складні ознаки, що моделюються.

Складні системи в економіці мають характеристики, які потрібно враховувати при їх описовому моделюванні, у протилежному разі некоректно говорити про адекватність розбудованої математичної моделі. Виокремлюються наступні визначальні характеристики СЕС – робастність (поняття не математичне), наявність неоднорідних зв’язків, емерджентність, масовий характер економічних явищ і процесів, випадковість і невизначеність у розвитку об’єктів в економіці. Робастність обумовлюється здатністю системи зберігати часткову працездатність (ефективність) у разі відмови окремих елементів чи підсистем. Друга визначальна характеристика відображає складність механізмів взаємодії внутрішнього середовища системи та обмінних процесів із зовнішнім середовищем. Третя характеристика розкриває ефекти інтегративності, цілісності, саморозвитку, синергетичні явища в системах в економіці. Закономірності економічних процесів не виявляються на невеликих сукупностях спостережень, тому моделювання спирається на масові спостереження, а точніше, систематичні визначення величин у просторі та часі. П’ятою характеристикою є випадковість і невизначеність у розвитку об’єктів, які є наслідком великої кількості умов, факторів функціонування об’єкта, тому вивчати ознаки потрібно за допомогою математичних методів на основі інструментів теорії ймовірностей та математичної статистики. Наведені характеристики повністю враховуються в описовому моделюванні за його елементарними та складними ознаками, зокрема їх дослідженні в статиці та динаміці [21; 63; 65; 71; 214]. Складними ознаками підприємства є причини, характеристики функціонування, розвитку, які мають багатовимірність. Наприклад, конкурентоспроможність, техніко-економічний рівень, інтелектуальне забезпечення діяльності тощо. Соціально-економічні системи описуються як відомими складними ознаками, так і прихованими латентними ознаками [272; 278; 282].

У моделюванні враховуються визначальні, основні ознаки і відхиляються не основні. Але факт урахування елементарних ознак і складних ознак залишається. Складні ознаки можуть оцінюватися опосередковано (це широко розповсюджена практика в економічному аналізі) та визначатися за допомогою математичного моделювання, вимірюватися за допомогою математичних методів.

Отже, до модельного базису опису соціально-економічних систем мають входити базисні моделі складних ознак, що перелічені в кортежі:

$$БМСО = \left\langle M_{СМО}, M_{СНМО}, M_{С(М+НМ)О} \right\rangle,$$

де $БМСО$ – базисні моделі складних ознак СЕС;

$M_{СМО}$ – математичні моделі складних метричних ознак (моделі І типу);

$M_{СНМО}$ – математичні моделі складних неметричних ознак (моделі ІІ типу);

$M_{С(М+НМ)О}$ – математичні моделі складних метрично-неметричних (сумісних систем) ознак (моделі ІІІ типу).

5. Виявлення, дослідження складних ознак можливі завдяки застосуванню математичних методів багатовимірного статистичного аналізу (БСА); таким чином формуються умови для визначення величин складних ознак.

Складні ознаки СЕС можуть бути встановлені на основі апріорного теоретико-логічного аналізу, економічного аналізу. Але латентні складні ознаки ідентифікуються лише за допомогою математичних методів. У математичній статистиці для вирішення різних проблем, пов’язаних із багатовимірністю, існує спеціальний розділ, що називається багатовимірним статистичним аналізом, основні методи якого викладені в другому розділі даного посібника.

Безпосереднє визначення величин складних ознак здійснюється як реалізація вторинного вимірювання в такій послідовності: визначення структури складних ознак за їх елементарними ознаками; оцінювання взаємозв’язку елементарних ознак, що становлять складну багатовимірну ознаку, та оцінювання взаємозв’язку складних ознак у системі, яка визначає соціально-економічну систему; статистичний опис величин ознак, встановлення закону розподілу величини окремої ознаки для побудови адекватної шкали та функції перетворення.

Методи БСА добре відпрацьовані для метричних величин, але залишаються погано адаптованими до неметричних величин. Важливі вдосконалення та розробка методів моделювання складних ознак для соціально-економічної системи, яка здебільшого визначається неметричними ознаками або сумісними системами метричних і неметричних ознак. Розвиненість методів обробки неметричних величин сприяє закріпленню основ ідентифікації СЕС.

6. Опис соціально-економічних систем завдяки метричним і неметричним величинам складних ознак соціально-економічних систем достовірний на основі визначених їх величин.

Визначення величин складних ознак об'єкта здійснюється після того як вони виявлені, ідентифіковані, а отже, розроблені їх моделі. Математичні моделі складних ознак відображають структуру елементарних ознак, що утворюють складну. Величина складної ознаки визначається за допомогою її моделі. Цьому сприяє моделювання вимірників, основною процедурою якого є перетворення величин елементарних ознак на основі функцій перетворення, що враховують закони й закономірності в економіці, існуючі теорії. Таким чином, перетворені величини синтезують величину складної ознаки. Перетворення виконуються згідно з діями, які входять до групи допустимих перетворень у процедурах вимірювання.

7. Система перетворених величин (якісно і кількісно) визначає соціально-економічну систему. Розроблені вимірники є вихідною інформацією для подальшої розбудови різних математичних моделей соціально-економічних систем.

За наведеним вище визначенням, показник є формою величини ознаки в економіці. У зв'язку з цим автори зауважують, що мало абсолютних і вартісних показників чинно обчислюються, для більшої частини відсутні методики або науково обґрунтовані рекомендації, що ускладнює їх використання. Для достовірного визначення величин елементарних ознак потрібно мати науково обґрунтовані шкали величин, уніфікувати їх для співвимірності. Функції перетворень конструюються з урахуванням концепцій, парадигм в економіці. Далі в посібнику викладені пропоновані функції перетворень різних величин метричних і неметричних ознак. Отримані після перетворення величини є вимірниками, їх можна порівнювати статично й динамічно. Вимірники можна розглядати як достовірні індикатори соціально-економічних систем. Перетворені величини є інформаційною основою для розбудови описових, оптимізаційних, прогнозувальних математичних моделей та моделей управління соціально-економічних систем. Система вимірників адекватно визначає соціально-економічну систему за її метричними й неметричними ознаками. Описана логіка ідентифікації узагальнена на схемі на рис. 1.5.

На рис. 1.6 наведена структурно-логічна побудова методології модельного базису опису соціально-економічних систем на основі концепції визначення величин ознак в економіці. При розробці нової методології були вдосконалені математичні методи, які оперують з неметричними величинами та сумісними системами. Ці вдосконалення наведені в наступних розділах посібника.

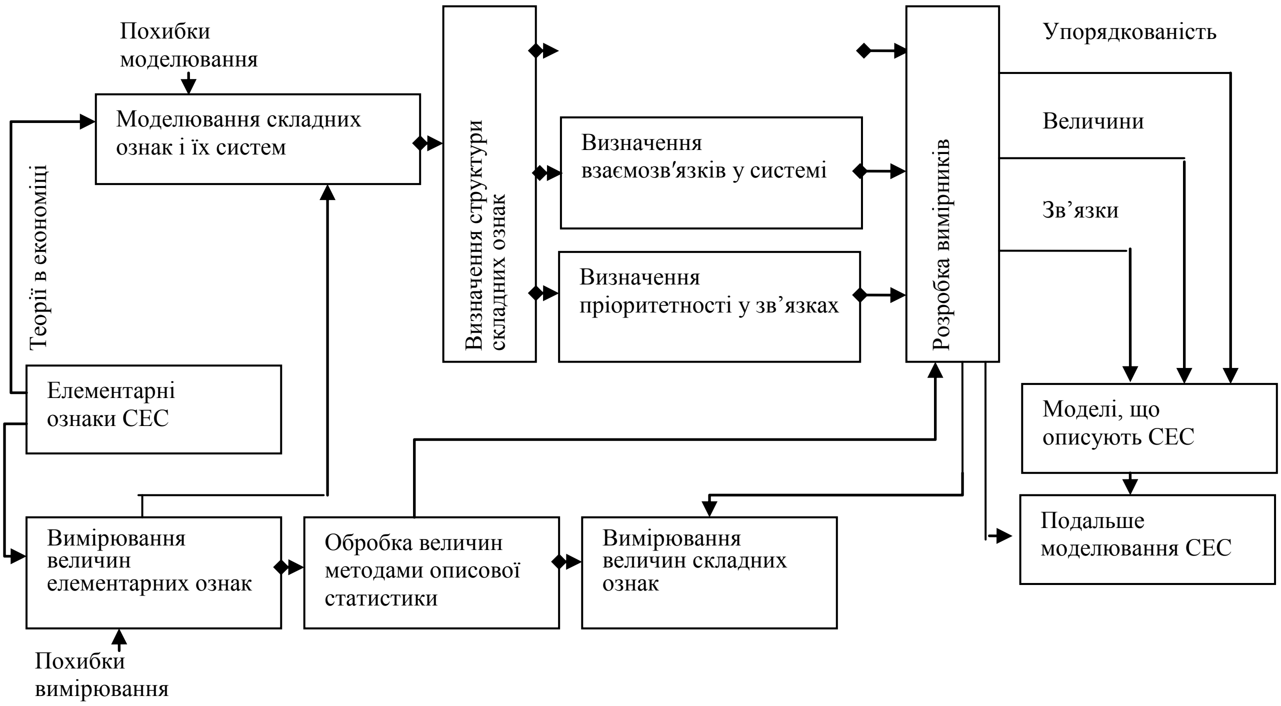


Рис. 1.5. **Загальна схема ідентифікації й опису СЕС**

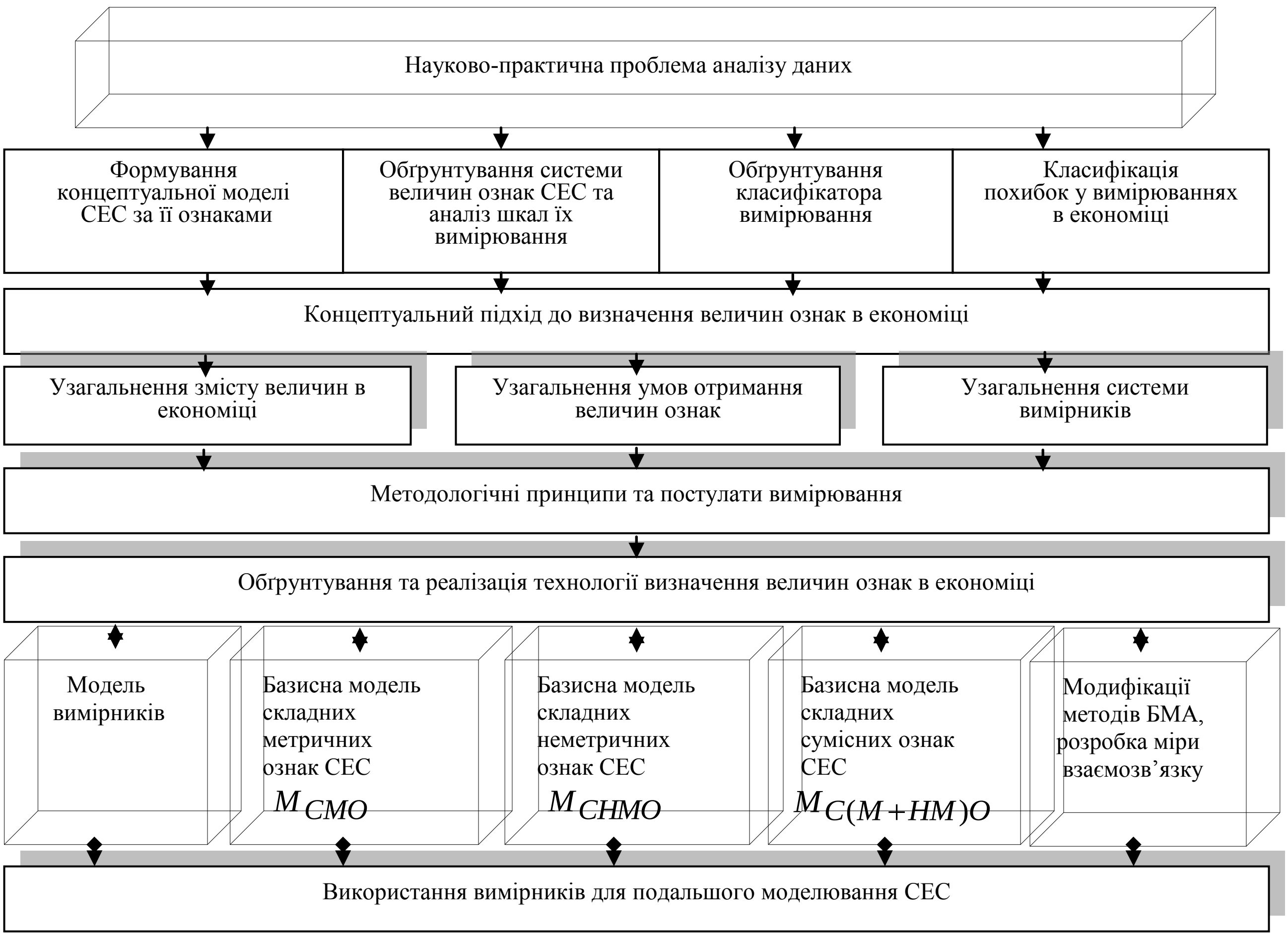


Рис. 1.6. **Структурно-логічна побудова методології модельного базису опису СЕС**

Викладені основні положення нової методології відкривають нові можливості моделювання, посилюють загальні методологічні принципи моделювання, а отже, підвищують якість розбудованих моделей та ефективність, дієвість управлінських рішень, що формуються на основі результатів моделювання.

Запитання для самоперевірки

1. Яка мета сучасного аналізу даних в економіці?
2. Які існують основні концепції аналізу даних в економіці?
3. Яке значення має вимірювання у визначенні величин ознак в економіці та аналізі даних?
4. Які основні етапи розвитку виокремлюються в багатовимірному статистичному аналізі?
5. Назвати основні напрямки розвитку теорії і практики вимірювання.
6. Сформулювати основні принципи та постулати вимірювання величин ознак в економіці.
7. Які особливості виокремлюються у вимірюванні в економіці?
8. Навести структуру системи ознак, за допомогою якої визначаються соціально-економічні системи.
9. Які типи моделей складних ознак рекомендуються для опису соціально-економічних систем?
10. Чим особливе ізоморфно-гомоморфне відображення ознак реального об'єкта в числові форми цього об'єкта в економіці?
11. Навести загальну формалізацію соціально-економічних систем.
12. Навести приклади складних ознак соціально-економічних систем.
13. Які моделі входять до модельного базису опису соціально-економічних систем?
14. Дати визначення вимірника в економіці.
15. Назвати основні елементи загальної схеми ідентифікації й опису соціально-економічних систем.
16. Сформулювати особливості структурно-логічної побудови методології модельного базису опису соціально-економічних систем.

Розділ 2
Методи багатовимірного статистичного аналізу метричних ознак соціально-економічних систем

2.1. Специфікація моделей складних ознак для аналізу соціально-економічних систем

Адекватність моделі залежить від правильності сукупності гіпотез, сформульованих на етапах концептуальної й математичної постановок і від точності отриманих результатів, які в економіці обумовлюються стійкістю розв'язків математичних методів, що застосовувалися. Гіпотези, що фундують модель, знаходять своє відображення в математичних методах, завдяки яким розбудовують моделі. Виходячи з того, що складні ознаки багатовимірні, для розробки їх моделей слід використовувати, перш за все, багатовимірний статистичний аналіз.

За визначенням економіко-математичного енциклопедичного словника "багатовимірний статистичний аналіз – розділ математичної статистики, який присвячений математичним методам побудови оптимальних планів збору, систематизації, обробки та інтерпретації багатовимірних статистичних даних, які спрямовані, перш за все, на виявлення характеру та структури взаємозв'язків між компонентами багатовимірної ознаки та призначені для одержання наукових і практичних висновків" [334, с. 298]. Отже, підкреслюється спрямованість даних методів на визначення структури складних ознак, оцінки взаємозв'язку елементарних ознак, що складають складну багатовимірну ознаку, та оцінки взаємозв'язку складних ознак у системі, що описує багатовимірний об'єкт.

Соціально-економічна система визначається системою елементарних метричних і неметричних ознак. Метричні ознаки у формі показників ($x_1, \ldots, x_k$) скалярно вимірюють у визначеній шкалі ступінь прояву властивостей об'єкта – фізичної форми соціально-економічної системи. Порядкові, або ординальні ознаки ($y_1, \ldots, y_p$), дозволяють упорядкувати об'єкти, що досліджуються, за ступенем прояву в них некількісної властивості. Класифікаційні, або номінальні ознаки ($z_1, \ldots, z_q$), дозволяють відрізняти об'єкти за властивостями, що формалізуються лише за назвами.

Багатовимірна ознака, або багатовимірний об'єкт, визначається n-вимірним вектором елементарних ознак, які виміряні за різними шкалами:

$$\boldsymbol{X} = (x_1,\dots,x_k, y_1,\dots,y_p, z_1,\dots,z_q), \text{ або } \boldsymbol{X} = (x_i, y_j, z_k), \; k+p+q=n.$$

Результати вимірювань наведених ознак у формі показників на кожному з m об'єктів подаються у вигляді:

$$(x_{ij}), \tag{2.1}$$

де $i=\overline{1,n}$, $j=\overline{1,m}$;

n – кількість показників;

m – кількість об'єктів.

Вони утворюють послідовність багатовимірних спостережень або початковий масив багатовимірних даних для проведення багатовимірного статистичного аналізу. Багато методів БСА концептуально розроблені у ході інтерпретації ознаки як багатовимірної випадкової величини, де послідовність спостережень вважається за вибіркою з генеральної сукупності. У цьому випадку вибір методів обробки початкових статистичних даних та аналіз їх властивостей виконуються на основі тих чи інших припущень відносно природи багатовимірного сумісного закону розподілу ймовірностей $P(x)$ [64; 248; 317].

Відомий математик, спеціаліст з проблем БСА Т. Андерсон рекомендує статистичні методи багатовимірного аналізу розподіляти на п'ять груп [6, с. 11 – 13].

Перша група методів призначена для оцінювання ступеня залежності між двома випадковими величинами в сукупності й у вибірці. Поняття коефіцієнта кореляції розповсюджується на визначення залежності між однією випадковою величиною і множиною випадкових величин; між двома множинами випадкових величин. Окремий парний коефіцієнт кореляції оцінює тісноту зв'язку між двома випадковими величинами, коли дія інших випадкових величин виключена. Множинний коефіцієнт кореляції оцінює тісноту зв'язку між однією випадковою величиною (результативною ознакою) і множиною так званих "пояснювальних ознак". Канонічний коефіцієнт кореляції оцінює тісноту зв'язку між двома групами ознак. Різні вибіркові коефіцієнти кореляції використовуються для оцінки відповідних параметрів розподілу і для перевірки таких гіпотез, як гіпотеза незалежності.

Друга група методів БСА, призначена для вирішення проблем, аналогічна до одновимірних статистичних методів. Наприклад, в одновимірному випадку перевіряють гіпотезу про рівність нулю математичних сподівань скалярних випадкових величин; у багатовимірному випадку також перевіряють гіпотезу про те, що математичні сподівання векторних випадкових величин дорівнюють нулю. Аналогією одновимірного t-критерію Стьюдента для багатовимірного випадку є узагальнений T^2-критерій. Для вирішення більшості таких проблем вибір системи координат не відіграє ролі, оскільки лінійні перетворення змінних не впливають на результати аналізу. Зокрема, у багатьох випадках перевірки гіпотез лінійні перетворення не змінюють гіпотези або процесу їх перевірки.

Третя група методів БСА пов’язана з проблемою вибору системи координат, щоб випадкові величини мали бажані статистичні властивості. Ці проблеми пов’язані з алгебраїчними проблемами подання матриць у канонічній формі. Наприклад, відшукується така нормалізована лінійна комбінація випадкових величин, щоб її дисперсія була максимальною (наприклад, знаходження головних компонент); це рівноцінно пошуку такого повороту осей, що приводить коваріаційну матрицю до діагональної форми. Інший приклад – характеристика залежності між двома множинами випадкових величин (знаходження канонічних кореляцій). Для вирішення цих проблем потрібно знаходити характеристичні числа і характеристичні вектори різних матриць.

Четверта група методів БСА вирішує деталізовані проблеми, у більшості яких множини випадкових величин розбиваються на підмножини та перевіряються гіпотези про незалежність цих підмножин або про симетрію між підмножинами чи всередині підмножин. До цієї групи методів Т. Андерсон відносить факторний аналіз.

П’ята група методів БСА (ще детально не відпрацьованих) пов’язана з проблемами аналізу залежності випадкових величин, що послідовні в часі. Ці проблеми передбачають вивчення внутрішньорядних кореляцій (автокореляцій) стохастичними різницевими рівняннями.

З точки зору практики виявлення та аналізу ознак об’єктів в економіці сучасні методи БСА призначені для розв’язання трьох типів проблем. Перша проблема пов’язана зі статистичним дослідженням залежностей між ознаками, що описують об’єкт. Передбачається, що сформовану систему показників можна подати у вигляді вектора $x^{(1)}$ передбачуваних (залежних) змінних і підвектора $x^{(2)}$ незалежних змінних, що впливають на інші змінні. Проблема полягає у визначенні на основі вибірки такої векторної функції $f(x^{(2)})$ з класу допустимих рішень F, яка б забезпечувала найкращу апроксимацію поведінки підвектора показників $x^{(1)}$. У даній ситуації залежно від конкретного вигляду функціонала в якості апроксимації та змісту ознак використовують спектр схем множинної регресії, дисперсійного, коваріаційного чи конфлюентного аналізу.

Вирішенням другої проблеми БСА є класифікація об’єктів за ознаками або ознак. Спрощене математичне розуміння проблеми складається з того факту, що всю сукупність значень показників, подану у вигляді матриці $\|x_{ij}\|, i=\overline{1,n}, j=\overline{1,m}$, можна розділити на порівняно невелику кількість однорідних груп. Залежно від змісту ознак, які подаються у формі показників, та від вигляду функціонала, що задає критерій якості класифікації, використовують ті чи інші схеми дискримінантного і кластерного аналізів. Третя проблема БСА передбачає розв'язання задач зниження розмірності простору ознак та відбору найінформативніших показників. Якщо стисло формулювати дану проблему, то потрібно визначити набір порівняно невеликої кількості показників $z = (z_1, ..., z_r)$, які знайдені допустимими перетвореннями $Z(x)$ початкових показників $\boldsymbol{X} = (x_1, ..., x_k, x_{k+1}, ..., x_n)$, на якому досягається верхня межа (грань) певної екзогенно заданої міри інформативності r-вимірної системи ознак. Конкретизація функціонала, що задає міру автоінформативності, спрямованого на максимальне збереження інформації, яка міститься в початковому масиві даних, відносно самих початкових ознак, приводить до різних схем факторного аналізу та багатовимірного шкалювання. Функціонали, які задають міру зовнішньої інформативності (тобто спрямовані на отримання максимальної інформації відносно деяких інших ознак, що безпосередньо не містяться в x ознаках), приводять до різних методів відбору найінформативніших показників у схемах статистичного дослідження залежностей та дискримінантного аналізу.

Така практична спрямованість БСА узгоджується із загальноприйнятими визначеннями сучасних математиків, які розподіляють методи БСА на три великих групи: 1) БСА багатовимірних розподілів та їх основних характеристик; 2) БСА характеру і структури взаємозв'язків між компонентами багатовимірної ознаки; 3) БСА геометричної структури сукупності багатовимірних спостережень [317, с. 298 – 300].

Перша група БСА враховує ситуації, в яких спостереження мають імовірнісну природу, тобто інтерпретуються як вибірка з відповідної генеральної сукупності. До основних завдань цієї групи БСА належать: статистичне оцінювання багатовимірних розподілів, оцінка їх головних числових характеристик і параметрів; дослідження властивостей статистичних оцінок. За допомогою цього дослідження розподілів імовірностей для ряду статистик будуються статистичні критерії перевірки гіпотез про ймовірнісну природу багатовимірних даних. Вважається, що багатовимірна вихідна ознака x, що досліджується, підпорядкована багатовимірному нормальному закону розподілу $N_p(\boldsymbol{\mu}, \boldsymbol{V})$, функція щільності якого $f(x|\boldsymbol{\mu}, \boldsymbol{V})$ задається співвідношенням:

$$f(x|\mu,V)=\frac{1}{(2\pi)^{\frac{n}{2}}|V|^{\frac{1}{2}}}\cdot e^{-\frac{1}{2}(x-\mu)'V^{-1}(x-\mu)}, \qquad (2.2)$$

де $\mu=(\mu_1,...,\mu_n)'$ – вектор математичних сподівань компонент випадкової величини x (тобто $\mu_i=Ex_i$, $i=\overline{1,n}$), а коваріаційна матриця випадкового вектора x $V=(\sigma_{ij})$, $i,j=\overline{1,n}$, де $\sigma_{ij}=M(x_i-\mu_i)(x_j-\mu_j)$ – коваріації компонент вектора x (розглядається невироджений випадок, коли ранг V дорівнює n; у протилежному випадку, тобто коли ранг $p<n$, усі результати залишаються правильними, але стосовно підпростору меншої вимірності p, в якому, виявляється, міститься розподіл імовірностей випадкового вектора x).

Якщо (2.1) – послідовність незалежних спостережень, що утворюють випадкову вибірку з $N_p(\mu,V)$, то оцінками максимальної правдоподібності для параметрів μ і V, що містяться в (2.2), є відповідно статистики:

$$\overline{x_i}=\hat{\mu}_i=\frac{1}{n}\sum_{k=1}^{n}x_{ik} \quad \text{і} \quad \hat{V}=(s_{ij}), \qquad (2.3)$$

де

$$s_{ij}=\frac{1}{n}\sum_{p=1}^{n}\sum_{q=1}^{n}(x_{ip}-\hat{\mu}_i)(x_{jq}-\hat{\mu}_j)'. \qquad (2.4)$$

Часто в математичній літературі V позначають ще як $\hat{\Sigma}$.

До речі, випадковий вектор $\hat{\mu}$ підпорядковується p-вимірному нормальному закону $N_p(\mu,\frac{1}{n}V)$ і не залежить від $\hat{V}$, а сумісний розподіл елементів матриці $\hat{Q}=n\hat{V}$ описується так званим розподілом Уїшарта, щільність якого

$$\varpi(Q|V;n)=\begin{cases}\dfrac{|\hat{Q}|^{\frac{n-p-2}{2}}e^{-\frac{1}{2}tr(V^{-1}\hat{Q})}}{2^{\frac{(n-1)p}{2}}\pi^{\frac{p(p-1)}{4}}|V|^{\frac{(n-1)}{2}}\prod\limits_{j=1}^{p}\Gamma\left\{\dfrac{n-j}{2}\right\}} & \text{якщо } \hat{Q}\text{ – додатно визначена;}\\ 0 & \text{– у протилежно му разі.}\end{cases} \qquad (2.5)$$

У межах цієї ж схеми досліджені розподіли та моменти таких вибіркових характеристик багатовимірної випадкової величини, як коефіцієнти парної, часткової та множинної кореляції, узагальнена дисперсія (статистика $\left|\hat{V}\right|$), узагальнена T^2 – статистика Хоттелінга.

Зокрема, якщо визначити в якості вибіркової коваріаційної матриці S_n виправлену "на незміщеність" оцінку $\hat{V}$, а саме

$$S_n = \frac{n}{n-1}\hat{V},$$

то розподіл випадкової величини $\sqrt{n}\left(\frac{|S_n|}{|V|}-1\right)$ наближається до $N_1(0, 2p)$ при $n \to \infty$, а випадкові величини

$$\frac{n-p}{p(n-1)}T^2 = \frac{n-p}{p(n-1)}n(\hat{\mu}-\mu)' S_n^{-1}(\hat{\mu}-\mu) \qquad (2.6)$$

та

$$\frac{n_1+n_2-p-1}{(n_1+n_2-2)p}\tilde{T}^2 = \frac{n_1+n_2-p-1}{(n_1+n_2-2)p}\frac{n_1 n_2}{n_1+n_2}(\hat{\mu}_{n_1}-\mu_{n_2})' S_{n_1+n_2}^{-1}(\hat{\mu}_{n_1}-\mu_{n_2}) \qquad (2.7)$$

підпорядковуються F-розподілам з числами степенів свободи відповідно $(p, n-p)$ і (p, n_1+n_2-p-1). У співвідношенні (2.7) n_1 і n_2 – обсяги двох незалежних вибірок вигляду (2.1), що відбираються з однієї й тієї ж генеральної сукупності $N_p(\mu, V)$; μ_{n_i} і S_{n_i} – оцінки вигляду (2.3) і (2.4), (2.5), що побудовані за даними i-ї вибірки, а

$$S_{n_1+n_2} = \frac{1}{n_1+n_2-2}\left((n_1-1)S_{n_1} + (n_2-1)S'_{n_2}\right)$$

– загальна вибіркова коваріаційна матриця, що побудована за оцінками S_{n_1} і S_{n_2}.

Результати вирішення проблем методів БСА розділяються за двома основними типами [317].

Перший тип проблем спрямований на побудову найкращих статистичних оцінок для параметрів даних моделей та аналізу властивостей. Наприклад, нехай багатовимірна ознака x інтерпретується як векторна випадкова величина, що підпорядкована p-вимірному нормальному розподілу $N_p(\mu, V)$, розділена на два вектори-стовпчики $\boldsymbol{x}^{(1)}$ і $\boldsymbol{x}^{(2)}$ розмірності q і $p-q$ відповідно. Це визначає і відповідне розділення вектора математичних сподівань $\boldsymbol{\mu}$ теоретичної і вибіркової коваріаційної матриць V і $\hat{V}$, а саме:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \quad \boldsymbol{V} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \quad \text{і} \quad \hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\mu}^{(1)} \\ \hat{\mu}^{(2)} \end{pmatrix}, \quad \hat{\boldsymbol{V}} = \begin{pmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{pmatrix}.$$

Тоді умовний розподіл підвектора $\boldsymbol{x}^{(1)}$ (за умови, що другий підвектор набув фіксованого значення $\boldsymbol{x}^{(2)}$), буде також нормальним $N_q\left(\mu^{(1)} + B\left(x^{(2)} - \mu^{(2)}\right), \Sigma\right)$. При цьому оцінками максимальної правдоподібності $\hat{B}$ і $\hat{\Sigma}$ для матриць регресійних коефіцієнтів B і коваріацій Σ цієї класичної багатовимірної моделі множинної регресії

$$M\left(x^{(1)} \mid x^{(2)}\right) = \mu^{(1)} + B\left(x^{(2)} - \mu^{(2)}\right) \tag{2.8}$$

будуть взаємно незалежні статистики відповідно $\hat{\boldsymbol{B}} = \hat{\boldsymbol{V}}_{12}\hat{\boldsymbol{V}}_{22}^{-1}$ і $\hat{\Sigma} = \hat{\boldsymbol{V}}_{11} - \hat{\boldsymbol{V}}_{12}\hat{\boldsymbol{V}}_{22}^{-1}\hat{\boldsymbol{V}}_{21}$,

де розподіл оцінки $\hat{B}$ підпорядкований нормальному закону $N_{q(p-q)}(B, V_B)$, а оцінки $n\hat{\Sigma}$ – закону Уїшарта з параметрами Σ і $n-(p-q)$ (елементи коваріаційної матриці $\boldsymbol{V}_b$ виражаються в термінах елементів матриці $\boldsymbol{V}$).

Основні результати побудови оцінок параметрів та дослідження їх властивостей у моделях факторного аналізу, головних компонент та канонічних кореляцій стосуються аналізу властивостей характеристичних значень і векторів різних вибіркових коваріаційних матриць.

У схемах, що відходять від рамок класичної нормальної моделі, і тим паче від рамок певної ймовірносної моделі, основні результати стосуються побудови алгоритмів та дослідження властивостей оцінок параметрів, найкращих відносно певного екзогенно заданого функціонала якості (чи адекватності) моделі.

Другий тип проблем спрямований на побудову статистичних критеріїв для перевірки різних гіпотез про структуру досліджуваних взаємозв'язків. У рамках багатовимірної нормальної моделі, де послідовності спостережень вигляду (2.1) інтерпретуються як випадкові вибірки з відповідних багатовимірних нормальних генеральних сукупностей, побудовані, наприклад, статистичні критерії для перевірки наступних гіпотез.

1. Гіпотеза $\boldsymbol{\mu} = \boldsymbol{\mu}^*$ про рівність векторів математичних сподівань показників, де $\boldsymbol{\mu}^*$ − відомий конкретний вектор; ця гіпотеза перевіряється за допомогою T^2-статистики Хоттелінга з підстановкою у формулу (2.6) $\boldsymbol{\mu} = \boldsymbol{\mu}^*$.

2. Гіпотеза $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$ про рівність векторів математичних сподівань у двох генеральних сукупностях (з однаковими, але невідомими коваріаційними матрицями), що представлені двома вибірками; ця гіпотеза перевіряється за допомогою статистики T^2.

3. Гіпотеза $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = ... = \boldsymbol{\mu}^{(k)} = \boldsymbol{\mu}_0$ про рівність векторів математичних сподівань у декількох генеральних сукупностях (з однаковими, але невідомими коваріаційними матрицями), що представлені своїми вибірками; ця гіпотеза перевіряється за допомогою статистики:

$$U_{p,\,k-1,\,n-k} = \frac{\left|\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{nj}\left(x_{.i}^{(j)} - \hat{\boldsymbol{\mu}}^{(j)}\right)\left(x_{.i}^{(j)} - \hat{\boldsymbol{\mu}}^{(j)}\right)^T\right|}{\left|\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{nj}\left(x_{.i}^{(j)} - \hat{\mu}\right)\left(x_{.i}^{(j)} - \hat{\mu}\right)^T\right|}, \tag{2.9}$$

в якій $x_{.i}^{(j)}$ є i-м p-вимірним спостереженням у вибірці обсягу n_j, що представляє j-ту сукупність, а $\hat{\boldsymbol{\mu}}^{(j)}$ і $\hat{\boldsymbol{\mu}}$ – оцінки вигляду (2.3), побудовані відповідно окремо за кожною із вибірок та за об'єднаною вибіркою обсягу $n = n_1 + ... + n_k$.

4. Гіпотези $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = ... = \boldsymbol{\mu}^{(k)} = \boldsymbol{\mu}$ та $\boldsymbol{V}_1 = ... = \boldsymbol{V}_k = \boldsymbol{V}$ про еквівалентність декількох нормальних генеральних сукупностей, що представлені своїми вибірками $\left\{x_i^{(j)}\right\}_{i=1}^{n_j}$, $j = \overline{1,k}$; ця гіпотеза перевіряється за допомогою статистики:

$$\lambda = \frac{\prod\limits_{j=1}^{k}\left|n_j \hat{\boldsymbol{V}}_j\right|^{\frac{n_j - 1}{2}}}{\left|\sum\limits_{j=1}^{k}\sum\limits_{i=j}^{n_j}\left(x_{.i}^{(j)} - \boldsymbol{\mu}\right)\right|^{\frac{n-k}{2}}}, \tag{2.10}$$

в якій $\hat{\boldsymbol{V}}_j$ – оцінка виду (2.4), що побудована окремо за спостереженнями j-ї вибірки, $j=\overline{1,k}$.

5. Гіпотеза про взаємну незалежність підвекторів-стовпчиків $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, ..., \boldsymbol{x}^{(m)}$

розмірності відповідно $p_1, p_2, ..., p_m$, на які розділений початковий p-вимірний вектор ознак $\boldsymbol{x}$, $p_1+p_2+...+p_m=p$; ця гіпотеза перевіряється за допомогою статистики

$$\Psi = \frac{\left|n\hat{\boldsymbol{V}}\right|}{\prod_{i=1}^{m}\left|n\hat{\boldsymbol{V}}_{\boldsymbol{i}}\right|}, \tag{2.11}$$

в якій $\boldsymbol{V}$ і $\hat{\boldsymbol{V}}_i$ − вибіркові коваріаційні матриці вигляду (2.4) для всього вектора $\boldsymbol{x}$ та для його підвектора $\boldsymbol{x}^{(i)}$ відповідно [317].

Багатовимірний статистичний аналіз геометричної старуктури багатовимірних спостережень створений на основі концепцій дискримінантного аналізу, суміші ймовірнісних розподілів, кластерного аналізу та таксономії, багатовимірного шкалювання. Ключовим для всіх схем даних методів є поняття відстані (міри близькості, міри схожості) між ознаками. Причому аналізуватися можуть реальні об’єкти, що фіксують значення ознак $\boldsymbol{x}$, тоді геометричним образом

i-го об’єкта буде точка $\boldsymbol{x}_i=(x_{1i},...,x_{pi})$ у відповідному p-вимірному просторі; аналогічно ознаці $x_{l.}$, $l=\overline{1,p}$ буде відповідати точка $\boldsymbol{x}_{l.}=(x_{l1},x_{l2},...,x_{\ln})$ у відповідному n-вимірному просторі.

Цілі багатовимірних методів, таких, як дискримінантний і кластерний аналізи, багатовимірне шкалювання, аналіз сумішей коротко характеризуються так.

У задачах дискримінантного аналізу необхідно сформулювати класифікаційні правила. Ці класифікаційні правила дозволяють приписати деякий новий елемент (спостереження X) до своєї підсукупності, якщо наперед не відомо, до якої із сукупностей цей елемент належить. Зазвичай під класифікаційним правилом розуміють послідовність дій, спрямованих на обчислення скалярної функції від показників ознак, за значеннями яких ухвалюються рішення про віднесення елемента до одного з класів, на впорядкування самих показників за ступенем їх інформативності з точки зору правильного віднесення елементів до класів, на обчислення відповідних імовірностей помилкової класифікації.

Задача аналізу суміші розподілів імовірностей частіше за все виникає у зв'язку з дослідженням "геометричної структури" сукупності. При цьому поняття r-го однорідного класу формалізується за допомогою генеральної сукупності, що описується певним унімодальним законом розподілу $P(X|\theta_r)$, так що розподіл загальної генеральної сукупності, з якої вилучається вибірка (2.1), описується сумішшю розподілів вигляду

$$P(X)=\sum_{r=1}^{k}\pi_r P(X|\theta_r),$$

де π_r – апріорна ймовірність (частка елементів) r-го класу в загальній генеральній сукупності.

Задача складається з достатнього статистичного оцінювання (за вибіркою $\{X_i\}_1^m$) невідомих параметрів θ_r, π_r, а іноді й k. Це, зокрема, дозволяє звести задачу класифікації елементів до схеми кластерного аналізу.

У кластерному аналізі геометрична структура сукупності елементів, що аналізуються, задана або координатами відповідних точок, тобто матрицею $\|x_{ij}\|$, $i=\overline{1,n}$, $j=\overline{1,m}$, або набором геометричних характеристик їх взаємного розміщення, наприклад матрицею попарних відстаней $\|d_{ij}\|_{i,j=1}^{n}$. Задачами методу передбачається розділення сукупності елементів на порівняно невелику кількість класів так, щоб елементи одного класу знаходилися на відносно невеликій відстані один від одного так, щоб різні класи були у міру можливості достатньо взаємно віддаленими один від одного і не розділялися б на такі віддалені частини.

Задачами багатовимірного шкалювання частіше всього передбачаються ситуації, коли немає первинної метричної міри для оцінок відстані між об'єктами і потрібно ввести таку міру для метричних даних. За відомої метричної матриці відстаней багатовимірне шкалювання дуже подібне до факторного аналізу і розв'язує ті самі задачі, що і кластерний аналіз. Слід зауважити, що основні методи кластерного аналізу та багатовимірного шкалювання розвиваються, зазвичай, без будь-яких припущень про ймовірнісну природу початкових даних.

На рис. 2.1 наведений узагальнений зміст методів БСА з точки зору формування умов визначення величин в економіці для аналізу соціально-економічних систем.

Результат узагальнення практики застосувань методів викладений у вигляді переліку аналітичних можливостей рекомендованих методів для проведення багатовимірного аналізу соціально-економічних систем та практичних задач в економіці, що можна розв’язати завдяки цим методам.

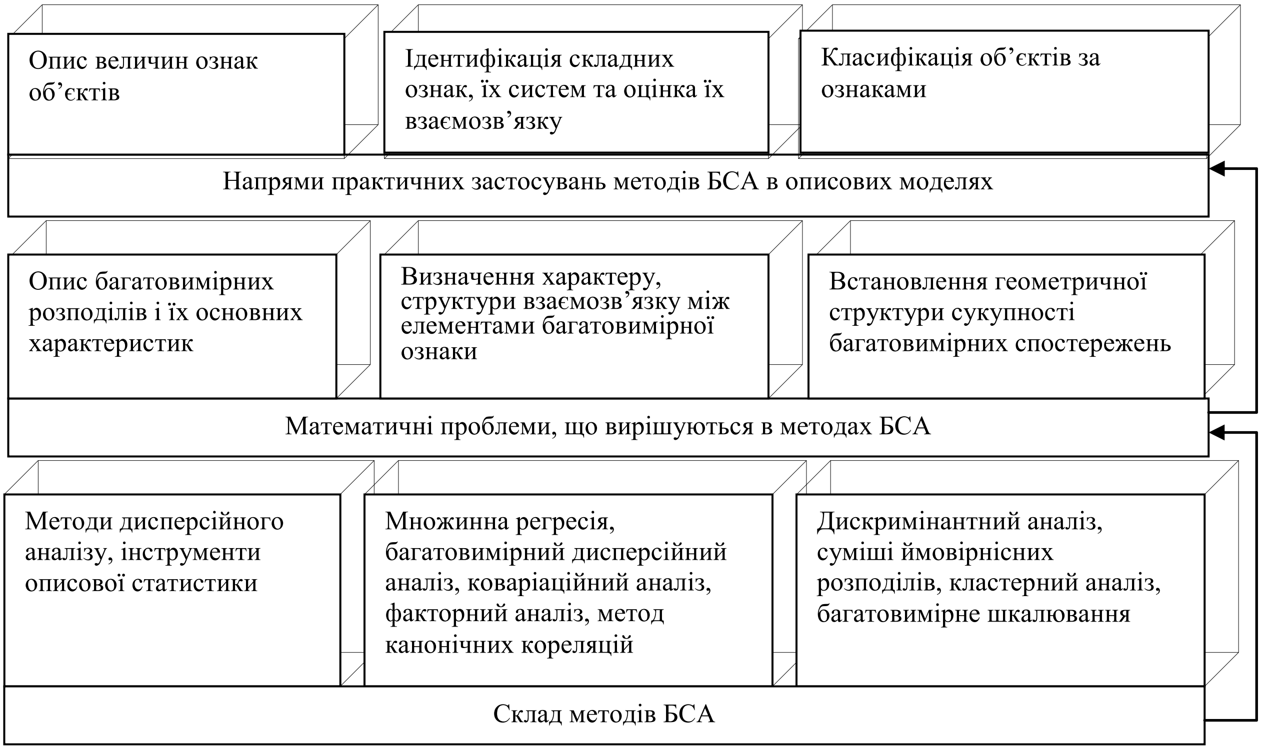


Рис. 2.1. **Зміст БСА**

Аналітичні можливості факторного аналізу:

а) скорочення багатовимірного простору ознак об’єктів;

б) статистична оцінка взаємозв’язку елементарних ознак і нових виділених складних ознак;

в) статистична оцінка значущості виділених складних ознак;

г) вимірювання значень виокремлених ознак об’єкта;

д) побудова систем складних і елементарних ознак, що описують об’єкт;

е) розробка досконалого обчислювального алгоритму з використанням спеціальних програмних середовищ.

Перелік практичних задач, які можна розв’язувати за допомогою факторного аналізу: визначення внутрішніх неявних (латентних) складних факторів розвитку підприємства; визначення внутрішніх неявних складних факторів розвитку підприємств у регіоні; визначення внутрішніх неявних складних факторів розвитку об'єктів і суб'єктів економіки в країні; діагностика ступеня інформативності показників, що визначають фактори розвитку підприємств; оцінка ступеня інформативності системи основних показників діяльності підприємства; наукова обґрунтованість виявлених факторів розвитку підприємств; визначення типів розвитку підприємств за критерієм системи факторів, що формують даний розвиток; визначення та оцінка ієрархічної структури факторів, що обумовлюють розвиток підприємств, регіону, країни; розробка збалансованої системи показників, що забезпечує дієвість системи управління на різних рівнях управління; розробка комплексного економічного аналізу з урахуванням різних рівнів управління.

Аналітичні можливості кластерного аналізу:

а) багатовимірна класифікація об'єктів "без учителя" чи без навчальної вибірки, тобто для наперед не відомої структури сукупності;

б) багатовимірна класифікація об'єктів, коли наперед відомі структура сукупності і її окремі еталонні об'єкти;

в) визначення структури простору ознак об'єктів.

Перелік практичних задач, які можна розв'язувати за допомогою кластерного аналізу: визначення типів підприємств за критерієм встановленої множини їх кількісних характеристик; визначення типів розвитку підприємств у регіоні за встановленим комплексним критерієм; визначення типів розвитку підприємств у країні за встановленими критеріями; визначення груп однорідних підприємств у сукупності за заданими характеристиками, що описують еталонний стан підприємства; визначення однорідних підприємств, що об'єднуються за значеннями показників економіки згідно із заданими нормативами чи еталонами в регіоні; визначення однорідних підприємств у країні згідно з їх рівнями розвитку; визначення характеристик складних явищ, процесів в економіці підприємств, у регіоні та країні.

Аналітичні можливості дискримінантного аналізу:

а) перевірка наявної структури сукупності об'єктів;

б) визначення класифікаційних функцій;

в) статистична оцінка початкових дискримінантних функцій;

г) визначення класифікаційних правил (інформантів), на основі яких об'єкти відносяться до відповідної підсукупності;

д) багатовимірна класифікація об'єктів, коли наперед відомі структура сукупності та її окремі еталонні об'єкти.

Перелік практичних задач, які можна розв'язувати за допомогою дискримінантного аналізу: перевірка усталеності класифікації груп підприємств у регіоні, країні; визначення критеріїв кластеризації підприємств у регіоні, країні; аналіз інформативності показників, що забезпечують існування кластерів розвитку підприємств; розробка методики комплексного аналізу характеристик підприємств, що обумовлюють віднесення його до окремого кластера розвитку; визначення кластерів однорідних підприємств у сукупності за заданими характеристиками, що описують еталонний стан підприємства; визначення кластерів однорідних підприємств, що об'єднуються за значеннями показників економіки згідно із заданими нормативами чи еталонами; визначення однорідних підприємств у країні згідно з рівнями їх розвитку.

Аналітичні можливості методу канонічних кореляцій:

а) скорочення багатовимірного простору ознак об'єктів до системи пар найбільш корельованих складних ознак;

б) статистична оцінка взаємозв'язку елементарних ознак і нових виділених ознак та взаємозв'язку в системі складних ознак;

в) статистична оцінка значущості виділених пар складних ознак;

г) визначення значень виокремлених складних ознак у системі;

д) побудова ієрархічної системи складних і елементарних ознак, що комплексно описують об'єкт;

е) розробка досконалого обчислювального алгоритму з використанням спеціальних програмних середовищ.

Перелік практичних задач, які можна розв'язувати за допомогою методу канонічних кореляцій: визначення внутрішніх латентних факторів соціального та економічного розвитку підприємства; визначення неявних складних факторів розвитку підприємства в розрізі "витрати – результати"; визначення внутрішніх неявних факторів економічного розвитку підприємств у регіоні; визначення внутрішніх неявних факторів економічного розвитку об'єктів і суб'єктів економіки в країні; діагностика ступеня інформативності показників, що визначають фактори розвитку підприємств; оцінка ступеня інформативності підсистем основних показників діяльності підприємств; наукова обґрунтованість виявлених економічних факторів розвитку підприємств; визначення типів економічного розвитку підприємств за критерієм виявленої системи факторів, що формують даний розвиток; визначення та оцінка виявленої структури факторів, що обумовлюють економічний розвиток підприємств, регіону, країни; розробка збалансованої системи показників, що забезпечує дієвість системи управління на різних рівнях; розробка комплексного економічного аналізу з урахуванням різних рівнів управління.

Аналітичні можливості задачі розділення суміші:

а) виокремлення підсукупностей у сукупності об'єктів, що мають різні закони розподілу значень ознак;

б) обчислення статистичних характеристик окремих підсукупностей, що складають одну сукупність;

в) дослідження величини ознак об'єкта;

г) визначення рівнів значень величини ознак об'єкта.

Перелік практичних задач, які можна розв'язувати за допомогою задачі розділення суміші: діагностика процесів розвитку підприємств у регіоні, країні; визначення статистичних показників процесів розвитку підприємств у регіоні, країні; побудова об'єктивних шкал значень показників в економіці; наукова обґрунтованість нормативів, планових значень показників в управлінні, узгодженість нормативів на різних рівнях управління.

Виходячи зі змісту кожного методу, система математичних методів для проведення багатовимірного аналізу соціально-економічних систем за метричними складними ознаками схематично подається кортежем:

$$СММ = \langle ФА, КА, ДА, МКК, ЗРС \rangle,$$

де *СММ* – система математичних методів, завдяки яким розбудовується модель метричних складних ознак;

ФА – факторний аналіз (мета застосування: ідентифікація загальних метричних складних ознак у сукупності об'єктів та в кластерах – однорідних групах);

КА – кластерний аналіз (мета застосування: кластеризація сукупності об'єктів за однорідними групами – типами);

ДА – дискримінантний аналіз (мета застосування: перевірка стійкості кластеризації);

МКК – метод канонічних кореляцій (мета застосування: ідентифікація системи метричних складних ознак);

ЗРС – задача розділення суміші (мета застосування: дослідження величини елементарної метричної ознаки на предмет її розвитку).

Методи БСА відпрацьовані в основному для метричних величин і погано адаптовані до неметричних величин. Соціально-економічні системи ідентифікуються як за метричними ознаками, так і за неметричними, а також за їх сумісними системами. Потрібні математичні методи, що передбачали б різні величини.

Математичні моделі ідентифікації неметричних складних ознак передбачають застосування таких математичних методів, що висувають вимоги до особливого виду зв'язків між елементарними ознаками, виміряними в порядкових та номінальних шкалах. Для визначення взаємозв'язку в системі неметричних ознак існують коефіцієнти рангової кореляції, коефіцієнти асоціації і контингенції, їх аналіз визначає доцільність використання для розвинення математичних методів метричних величин на випадок неметричних величин та їх сумісних систем. Існує декілька шляхів вирішення даної проблеми: розробка нових методів, що працюють сумісно з метричними і неметричними величинами; доопрацювання відомих математичних методів, що розроблялися для метричних величин з урахуванням, що вихідні дані будуть різними величинами.

Окремо слід розглянути проблему можливого порушення нормальності розподілу і його наслідків у ході практичних досліджень [56; 58 – 60]. Існує думка, що коректне використання коефіцієнтів кореляції Пірсона вимагає обов'язкового виконання нормального (багатовимірного) закону і це стосується майже всіх багатовимірних методів, зокрема, регресійного, факторного, дисперсійного аналізів [256]. Це не так. Почнемо з того, що метод головних компонент, факторний аналіз і канонічна кореляція – це цілком детерміновані процедури, які не враховують жодних умов щодо розподілів вихідних даних. Те ж саме стосується кластерного аналізу.

У дискримінантному аналізі вирішальні правила виведені дійсно на основі існування багатовимірного нормального розподілу даних, що привело до такої методики віднесення об'єктів до тих класів, до яких об'єкти найближчі за мірою відстані Махаланобіса. Тобто гіпотеза про нормальність обумовила конкретний вид відстані – міру Махаланобіса. Якщо будуть порушення нормальності розподілу даних, міра Махаланобіса не буде найкращою (з точки зору принципу максимуму правдоподібності), але ця міра не буде помилковою. Сучасний дискримінантний аналіз передбачає попередній перехід до так званих "дискримінантних функцій", що є новими координатами, у просторі яких найчіткіше видно всі відмінності між навчальними підсукупностями. У просторі дискримінантних функцій міра Махаланобіса переходить у звичайну евклідову відстань, яка застосовується всюди незалежно від закону розподілу. Таким чином, і для дискримінантного аналізу не будуть великою проблемою навіть значні порушення нормальності розподілу. Що стосується регресійного аналізу, то сучасний регресійний аналіз базується на гіпотезах Гауса-Маркова, які відомий фахівець Е. Маленво розмістив у наступному порядку (за зменшенням важливості):

1) $y_i = \hat{y}(x_i, b) + \varepsilon_i$ − усі випадкові похибки стосуються лише однієї результативної ознаки (пояснювальні змінні x_i вважаються не випадковими);

2) $M(\varepsilon_i) = 0$ − систематичних похибок немає;

3) $M(\varepsilon_i^2) = \sigma_\varepsilon^2$ − усі спостереження рівноточні (гомоскедастичні);

4) $M(\varepsilon_i \varepsilon_j) = 0$ − спостереження незалежні;

5) $N(\varepsilon_i; 0; \sigma_\varepsilon)$ − похибки розподілені нормально.

Отже, гіпотеза про нормальність розподілу (до речі, розподілу похибок ε_i, але не розподілу вихідних даних $Y, X_1, X_2, \ldots$) є найменш важливою серед усіх інших. З гіпотези про нормальність розподілу похибок ε_i випливає, що метод найменших квадратів (МНК) є найкращою процедурою оцінювання параметрів моделей. Але істотної визнаної альтернативи МНК поки ще не запропоновано, тому жодних змін у методах регресійного аналізу в разі порушення нормальності похибок ε_i немає.

Порушення нормальності не відображається на критерії Стьюдента, оскільки вибіркові середні розподілені асимптотично нормально на основі центральної граничної теореми.

Коефіцієнт парної кореляції Пірсона як міра тісноти лінійного зв'язку випливає з двовимірного нормального розподілу. Але це не означає, що лінійні зв'язки і коефіцієнт Пірсона не можуть з'являтися за інших умов. Слід зауважити, що застосування коефіцієнта Пірсона як міри лінійного зв'язку не вимагає взагалі ніяких припущень про розподіл ознак.

Таким чином, багато статистичних критеріїв, які виведені для випадку нормального розподілу, правильні у разі деяких порушень цієї нормальності.

2.2. Методи факторного аналізу

Факторний аналіз належить до складу методів багатовимірного статистичного аналізу, які найчастіше застосовуються для вирішення практичних задач в економіці.

У сучасній математичній статистиці під факторним аналізом розуміють сукупність методів, які на основі реально існуючих зв'язків ознак об'єкта дозволяють виявляти латентні узагальнюючі характеристики. Спочатку є сукупність елементарних ознак об'єкта x_j, взаємодія яких передбачає наявність визначених причин, тобто наявність деяких латентних факторів. Останні встановлюються в результаті узагальнення елементарних ознак і виступають як інтегровані характеристики чи ознаки, але складніші, вищого рівня. Слід зазначити, що корелювати можуть не тільки тривіальні ознаки x_j, але й самі об'єкти N_i, що спостерігаються, тому пошук латентних факторів теоретично можливий як за ознаками, так і за об'єктними даними. Залежно від того, який з розглянутих вище тип кореляційного зв'язку – елементарних ознак чи об'єктів, що спостерігаємо – досліджується у факторному аналізі, розрізняють R- і Q-техніки.

Назву R-техніки носить процедура аналізу m ознак, реалізуючи яку можна отримати r лінійних комбінацій ознак $F_r = f(X_j), r = \overline{1, m}$. Процедуру аналізу близькості (зв'язку) n об'єктів, що спостерігаються, називають Q-технікою, в результаті виконання якої отримуємо r лінійних комбінацій (груп) об'єктів $F_i = f(N_i), i = \overline{1, N}$. Практично більше 90% задач в економіці розв'язують за допомогою R-техніки.

Велику роль у розумінні факторного аналізу відіграє фундаментальна факторна теорема.

Фундаментальна факторна теорема. Вихідні дані в економіці, як правило, мають вигляд матриці, рядки якої відповідають різним об'єктам (спостереженням), а стовпчики – різним ознакам, що описують ці об'єкти. Позначимо матрицю вихідних даних через $\boldsymbol{X}$, розмір матриці $n \times m$, де n– кількість об'єктів, m– кількість ознак. Якщо стовпчики ознак стандартизувати, то нову матрицю будемо позначати $\boldsymbol{Z}$; стандартизовані значення обчислюються за формулою:

$$z_{ij} = \frac{x_{ij} - \bar{x}_i}{s_{x_i}}, \tag{2.12}$$

де $\bar{x}_i$ – середні;

s_{x_i} – стандартні відхилення $s_{x_i} = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)^2}$.

Взаємозв'язки між ознаками характеризуються кореляційною матрицею $\boldsymbol{R}$ – матрицею коефіцієнтів парної кореляції:

$$r_{x_i x_j} = \frac{s_{x_i x_j}}{s_{x_i} s_{x_j}}, \tag{2.13}$$

де $s_{x_i x_j}$ – коваріації $s_{x_i x_j} = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$.

Стандартизація не змінює коефіцієнтів кореляції. У матричній формі кореляційну матрицю можна записати у вигляді матричного добутку:

$$\boldsymbol{R} = \frac{1}{n-1}\boldsymbol{Z}'\boldsymbol{Z}, \tag{2.14}$$

де $\boldsymbol{Z}'$ − транспонована матриця $\boldsymbol{Z}$.

Розмір кореляційної матриці − $m \times m$.

Теорема 1 (теорема про факторизацію кореляційної матриці). Кореляційну матрицю можна подати у такому вигляді:

$$\boldsymbol{R} = \lambda_1 \boldsymbol{U}_1 \boldsymbol{U}_1' + \lambda_1 \boldsymbol{U}_2 \boldsymbol{U}_2' + ... + \lambda_m \boldsymbol{U}_m \boldsymbol{U}_m',$$

де λ_i – власні числа кореляційної матриці;

$\boldsymbol{U}_i$ – власні вектори, що відповідають кожному власному числу λ_i.

Оскільки норма кожного власного вектора дорівнює одиниці $\|\boldsymbol{U}_i\|^2 = \boldsymbol{U}_i'\boldsymbol{U}_i = 1$, внесок кожного доданка $\lambda \boldsymbol{U}_i \boldsymbol{U}_i'$ пропорційний власному числу λ_i. Отже, основний внесок у формування матриці $\boldsymbol{R}$ роблять кілька членів з найбільшими власними числами, якщо знехтувати доданками з малими власними числами, втрачається небагато інформації. Саме в цьому полягає можливість скорочення багатовимірного простору до меншої розмірності $k < m$.

Для доведення теореми 1 слід попередньо розглянути лему про властивості власних чисел і векторів [271].

Лема 1. Власні числа кореляційної матриці – дійсні й невід’ємні. Власні вектори, що відповідають різним власним числам, – взаємно ортогональні.

Маючи на увазі, що власні числа і вектори матриці визначаються з матричного рівняння $\boldsymbol{RU} = \lambda \boldsymbol{U}$:

а) у загальному випадку з квадратною матрицею $\boldsymbol{A}$ можливі комплексні власні числа і власні вектори: $\lambda = \alpha + i\beta$; $\boldsymbol{U} = \boldsymbol{V} + i\boldsymbol{W}$, де i – уявна одиниця. Але якщо матриця $\boldsymbol{A}$ – симетрична $\boldsymbol{A}' = \boldsymbol{A}$, цього не може бути. Доведемо це твердження. Перетворюємо матричний вираз $A\boldsymbol{U} = \lambda \boldsymbol{U}$ у випадку комплексних власних чисел і векторів:

$$\boldsymbol{A}(\boldsymbol{V} + i\boldsymbol{W}) = (\alpha + i\beta)(\boldsymbol{V} + i\boldsymbol{W}), \tag{2.15}$$

звідки

$$\begin{cases} \boldsymbol{AV} = \alpha \boldsymbol{V} - \beta \boldsymbol{W} \\ \boldsymbol{AV} = \alpha \boldsymbol{V} + \beta \boldsymbol{W} \end{cases}. \tag{2.16}$$

Позначимо $\boldsymbol{A} - \alpha \boldsymbol{E} = \boldsymbol{B}$ ($\boldsymbol{E}$ – одинична матриця). Якщо матриця $\boldsymbol{A}$ – симетрична, то матриця $\boldsymbol{B}$ буде також симетричною $\boldsymbol{B}' = \boldsymbol{B}$. Отже, уявні частини власних чисел потрібно знаходити із системи дійсних матричних рівнянь:

$$\begin{cases} \boldsymbol{BV} = -\beta \boldsymbol{W} \\ \boldsymbol{B}'\boldsymbol{W} = \beta \boldsymbol{V} \end{cases}. \tag{2.17}$$

Помножимо обидві частини першого рівняння на $\boldsymbol{V}'\boldsymbol{B}'$ ліворуч і врахуємо, що

$\boldsymbol{B}'\boldsymbol{W} = \beta \boldsymbol{V}$ − друге рівняння:

$$\boldsymbol{V}'\boldsymbol{B}'\boldsymbol{BV} = -\beta \boldsymbol{V}'\boldsymbol{B}'\boldsymbol{W} = -\beta^2 \boldsymbol{V}'\boldsymbol{V} = -\beta^2 \|\boldsymbol{V}\|^2. \tag{2.18}$$

Пригадаємо правило транспонування добутку матриць: $(\boldsymbol{BV})' = \boldsymbol{V}'\boldsymbol{B}'$.

Отже, маємо

$$\begin{aligned} -\beta^2 \|\boldsymbol{V}\|^2 &= (\boldsymbol{V}'\boldsymbol{B}')(\boldsymbol{BV}) = (\boldsymbol{BV})'(\boldsymbol{BV}) = (-\beta \boldsymbol{W}')(-\beta \boldsymbol{W}) = \\ &= \beta^2 \boldsymbol{W}'\boldsymbol{W} = \beta^2 \|\boldsymbol{W}\|^2. \end{aligned} \tag{2.19}$$

Звідси випливає, що $\beta^2 (\|\boldsymbol{V}\|^2 + \|\boldsymbol{W}\|^2) = 0$, тобто $\beta = 0$.

Отже, уявна частина власного числа $\lambda = \alpha + i\beta$ дорівнює нулю $\beta = 0$, тобто власні числа симетричної матриці $\boldsymbol{A}$ не можуть бути комплексними;

б) кореляційна матриця $\boldsymbol{R}$ не лише симетрична, вона має специфічну структуру:

$$\boldsymbol{R} = \frac{1}{n-1}\boldsymbol{Z'Z}\,. \tag{2.20}$$

Позначимо $\boldsymbol{ZU} = \boldsymbol{V}$; цей добуток можливий, тому що розміри матриць-добутків узгоджені: $[n \times m]\cdot[m \times 1] = [n \times 1]$. Помножимо матричну рівність $\boldsymbol{RU} = \lambda\boldsymbol{U}$ на $\boldsymbol{U'}$ ліворуч:

$$\frac{1}{n-1}\boldsymbol{U'Z'}\cdot\boldsymbol{ZU} = \lambda\boldsymbol{U'U}\text{, тобто } \frac{1}{n-1}\boldsymbol{V'V} = \lambda\boldsymbol{U'U}\,, \tag{2.21}$$

або $\frac{1}{n-1}\|\boldsymbol{V}\|^2 = \lambda\|\boldsymbol{U}\|^2 = \lambda$, звідки одержуємо $\lambda \geq 0$.

Отже, доведено, що власні числа кореляційної матриці не можуть бути від'ємними;

в) доведемо, що власні вектори $\boldsymbol{U}_i, \boldsymbol{U}_j$, які відповідають різним власним числам $\lambda_i \neq \lambda_j$, взаємно ортогональні, тобто $\boldsymbol{U}_i'\boldsymbol{U}_j = 0$, де $\boldsymbol{RU}_i = \lambda i\boldsymbol{U}_i$ і $\boldsymbol{RU}_j = \lambda_j\boldsymbol{U}_j$. Помножимо першу матричну рівність ліворуч на $\boldsymbol{U}_j'$, а другу рівність – $\boldsymbol{U}_i'$. Одержимо $\boldsymbol{U}_j'\boldsymbol{RU}_i = \lambda_i\boldsymbol{U}_j\boldsymbol{U}_i$ і $\boldsymbol{U}_i'\boldsymbol{RU}_j = \lambda_j\boldsymbol{U}_i'\boldsymbol{U}_j$. Обидві частини цих рівнянь – числа (вони не змінюються в ході транспонування), тому

$$\boldsymbol{U}_i'\boldsymbol{U}_j = \left(\boldsymbol{U}_i'\boldsymbol{U}_j\right)' = \boldsymbol{U}_j'\boldsymbol{U}_i\,, \tag{2.22}$$

$$\boldsymbol{U}_j'\boldsymbol{RU}_i = \left(\boldsymbol{U}_j'\boldsymbol{RU}_i\right)' = \boldsymbol{U}_i'\boldsymbol{R'U}_j = \boldsymbol{U}_i'\boldsymbol{RU}_j \text{ (оскільки } \boldsymbol{R'} = \boldsymbol{R}\text{)}. \tag{2.23}$$

Отже, маємо

$$\begin{cases} \boldsymbol{U}_i'\boldsymbol{RU}_j = \lambda_i\boldsymbol{U}_i'\boldsymbol{U}_j \\ \boldsymbol{U}_i'\boldsymbol{RU}_j = \lambda_j\boldsymbol{U}_i'\boldsymbol{U}_j \end{cases}. \tag{2.24}$$

$$0 = \left(\lambda_i - \lambda_j\right)\boldsymbol{U}_i'\boldsymbol{U}_j\,.$$

Оскільки $\lambda_i \neq \lambda_j$, то маємо $\boldsymbol{U}_i'\boldsymbol{U}_j = 0$, тобто всі твердження леми доведені.

Зауваження. Власні вектори визначаються з точністю до постійного множника. Домовимось задавати цей множник так, щоб норма вектора дорівнювала одиниці.

Доведення теореми 1. Систему матричних рівнянь

$$\boldsymbol{RU}_1 = \lambda_1\boldsymbol{U}_1,\ \boldsymbol{RU}_2 = \lambda_2\boldsymbol{U}_2,\ \boldsymbol{RU}_m = \lambda_m\boldsymbol{U}_m \tag{2.25}$$

можна записати у вигляді одного матричного рівняння

$$\boldsymbol{RU} = \boldsymbol{UD}_\lambda, \tag{2.26}$$

де $\boldsymbol{U}$ – матриця власних векторів $\boldsymbol{U} = [\boldsymbol{U}_1, \boldsymbol{U}_2, ..., \boldsymbol{U}_m]$;

$\boldsymbol{D}_\lambda$ – діагональна матриця, в якій відмінні від нуля лише елементи на діагоналі, і ці елементи дорівнюють власним числам λ_i.

Оскільки власні вектори кореляційної матриці – взаємно ортогональні

$$\boldsymbol{U}_i'\boldsymbol{U}_j = \begin{cases} 1, & \text{якщо } i = j \\ 0, & \text{якщо } i \neq j \end{cases}, \tag{2.27}$$

то цей факт можна записати у вигляді $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{E}$ ($\boldsymbol{E}$ – одинична матриця).

Отже, матриці $\boldsymbol{U}$ і $\boldsymbol{U}'$ є взаємно оберненими, тобто $\boldsymbol{U}' = \boldsymbol{U}^{-1}$ (добуток взаємно обернених матриць дорівнює одиничній матриці). Для взаємно обернених матриць справджується вираз $\boldsymbol{U}^{-1}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^{-1} = \boldsymbol{E}$. Помножимо матрицю $\boldsymbol{RU} = \boldsymbol{UD}_\lambda$ на $\boldsymbol{U}'$ праворуч і врахуємо, що $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{E}$. Одержимо $\boldsymbol{R} = \boldsymbol{UD}_\lambda\boldsymbol{U}'$, що можна розписати на кілька доданків з множниками λ_i:

$$R = \lambda_1\boldsymbol{U}_1\boldsymbol{U}_1' + \lambda_2\boldsymbol{U}_2\boldsymbol{U}_2' + ... + \lambda_m\boldsymbol{U}_m\boldsymbol{U}_m', \tag{2.28}$$

що і треба було довести. ■

Оскільки всі власні числа кореляційної матриці – невід'ємні, то симетричну діагональну матрицю $\boldsymbol{D}_\lambda$ можна записати як:

$$\boldsymbol{D}_\lambda = \boldsymbol{D}_{\sqrt{\lambda}}\boldsymbol{D}_{\sqrt{\lambda}}'. \tag{2.29}$$

Тоді вищерозглянутий матричний вираз $\boldsymbol{R} = \boldsymbol{UD}_\lambda\boldsymbol{U}'$ можна переписати у вигляді

$$\boldsymbol{R} = \boldsymbol{UD}_\lambda \cdot \boldsymbol{D}_\lambda'\boldsymbol{U}' = \boldsymbol{AA}', \tag{2.30}$$

де $\boldsymbol{A} = \boldsymbol{UD}_{\sqrt{\lambda}}$ – матриця факторних навантажень $\boldsymbol{A} = [\boldsymbol{A}_1\ \boldsymbol{A}_2 ..\boldsymbol{A}_m]$.

Кожний власний вектор $\boldsymbol{U}_i$ множиться на свій співмножник $\sqrt{\lambda_i}$: $\boldsymbol{A}_i = \boldsymbol{U}_i\sqrt{\lambda_i}$. Норми цих нових векторів тепер не однакові, а дорівнюють власним числам. Якщо деякі $\lambda_i = 0$, то дорівнюють нулю всі компоненти відповідних векторів $\boldsymbol{A}_i$. Нульові вектори і вектори з малою нормою вилучаються з матриці факторних навантажень, тобто матриця $\boldsymbol{A}$ має менший розмір $[n \times k]$, де $k < m$.

Розклад $\boldsymbol{R} = \boldsymbol{A}\boldsymbol{A}'$ називається фундаментальною факторною теоремою.

Відразу зауважимо, що такий розклад – не єдиний.

Якщо $\boldsymbol{T}$ – деяка ортогональна матриця: $\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{E}$, то $\boldsymbol{R} = \boldsymbol{A}\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{A}' = \boldsymbol{B}\boldsymbol{B}'$, де $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{T}$. Отже, маємо інший розклад кореляційної матриці й іншу матрицю факторних навантажень, причому зберігаються всі переваги початкового розкладу. Вдалим підбором матриці $\boldsymbol{T}$ намагаються покращити змістовну інтерпретацію результатів факторного аналізу. До речі, ортогональні перетворення з матрицею $\boldsymbol{T}$ ліворуч не змінюють ні суми квадратів елементів у рядках факторної матриці, ні суми добутків елементів будь-яких двох рядків. Такі перетворення є перетвореннями повороту, тому сама матриця $\boldsymbol{T}$ називається матрицею повороту.

Латентні фактори – компоненти. Коли декілька ознак пов'язані між собою кореляційними співвідношеннями, то це означає, що деякі ознаки визначаються через інші або вся така група ознак є проявами (наслідками) спільної для них причини, природа якої нам невідома, тобто не вивчена. Факторний аналіз базується саме на останньому припущенні. Ці латентні фактори, які знаходяться за лаштунками нашого уявлення про процес, безпосередньо нами не вимірюються, але саме вони визначають всі ознаки, що ми спостерігаємо. Знаючи прояви (наслідки) латентних факторів, можна побудувати модель для вивчення цих латентних факторів, кількість яких значно менша від загальної кількості ознак. До речі, якщо можна коротко описати великий масив даних, то це означає, що знайдена певна об'єктивна закономірність, яка обумовила можливість цього короткого опису, тобто стиснення інформації.

Початкова система факторів (компонент) складається за екстремальним принципом – перша компонента має пояснювати максимум всієї змінності всіх ознак; друга компонента, що незалежна від першої, має пояснювати максимум залишкової змінності ознак і т. д. Невелика кількість таких компонент спроможна відновити майже всю загальну змінність ознак.

Компоненту складаємо у вигляді лінійної комбінації

$$\boldsymbol{V} = u_1\boldsymbol{Z}_1 + u_2\boldsymbol{Z}_2 + \ldots + u_m\boldsymbol{Z}_m = \boldsymbol{Z}\boldsymbol{U}, \tag{2.31}$$

де коефіцієнти u_k треба знайти з умови максимуму дисперсії

$$S_V^2 = \frac{1}{n-1}\boldsymbol{V}'\boldsymbol{V} = \frac{1}{n-1}(\boldsymbol{Z}\boldsymbol{U})'(\boldsymbol{Z}\boldsymbol{U}) = \frac{1}{n-1}\boldsymbol{U}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{U} = \boldsymbol{U}'\boldsymbol{R}\boldsymbol{U}. \tag{2.32}$$

Тут враховано, що компоненти – центровані (як лінійна комбінація стандартизованих змінних $\boldsymbol{Z}_k$).

На коефіцієнти u_k потрібно накласти умову нормування $\boldsymbol{U'U}=1$, у протилежному разі, просто збільшуючи ці коефіцієнти, можна зробити дисперсію лінійної комбінації як завгодно великою.

Отже, маємо задачу про умовний екстремум:

потрібно знайти максимум квадратичної форми

$$S_V^2=\boldsymbol{U'RU}=\sum_{p=1}^{m}\sum_{q=1}^{m}r_{pq}u_pu_q\to\max \tag{2.33}$$

за умови нормування $\boldsymbol{U'U}=\sum_{p=1}^{m}u_p^2=1$.

Цю задачу розв’язуємо методом Лагранжа. Складаємо функцію Лагранжа:

$$F=\boldsymbol{U'RU}-\lambda(\boldsymbol{U'U})=\sum_{p=1}^{m}\sum_{q=1}^{m}r_{pq}u_pu_q-\lambda\left(\sum_{p=1}^{m}u_p^2-1\right), \tag{2.34}$$

де λ – множники Лагранжа.

Прирівнюємо до нуля часткові похідні функції Лагранжа $\dfrac{\partial F}{\partial u_{ik}}=0$ і одержуємо систему:

$$\begin{cases}2\sum_{p=1}^{m}r_{pk}u_p-2\lambda u_k=0\\ (k=\overline{1,m})\end{cases}, \tag{2.35}$$

яку в матричній формі можна записати як $\boldsymbol{RU}=\lambda\boldsymbol{U}$.

Отже, коефіцієнти u_k лінійної комбінації є компонентами власного вектора $\boldsymbol{U}$ кореляційної матриці, а множник Лагранжа – власним числом, що відповідає вектору $\boldsymbol{U}$. Таким чином, можна скласти m лінійних комбінацій (компонент) $\boldsymbol{V}_i=\boldsymbol{ZU}_i$ (за кількістю власних векторів).

Теорема 2. Компоненти – взаємно ортогональні, їх дисперсії дорівнюють власним числам кореляційної матриці.

Доведення:

1. Розглянемо скалярний добуток двох різних компонент:

$$\boldsymbol{V}_i'\boldsymbol{V}_j=(\boldsymbol{ZU}_i)'(\boldsymbol{ZU}_j)=\boldsymbol{U}_i'\boldsymbol{Z'ZU}_j=(n-1)\boldsymbol{U}_i'\boldsymbol{RU}_j=(n-1)\lambda_j\boldsymbol{U}_i'\boldsymbol{U}_j=0.$$

У ході перетворення використано:

правило транспонування добутку матриць $(\boldsymbol{Z}\boldsymbol{U}_i)' = \boldsymbol{U}_i'\boldsymbol{Z}'$;

вираз кореляційної матриці $\boldsymbol{R} = \dfrac{1}{n-1}\boldsymbol{Z}'\boldsymbol{Z}$;

визначення власних векторів $\boldsymbol{R}\boldsymbol{U}_j = \lambda_j \boldsymbol{U}_j$;

умову ортогональності власних векторів кореляційної матриці $\boldsymbol{U}_i'\boldsymbol{U}_j = 0$.

Отже, всі компоненти – взаємно ортогональні $\boldsymbol{V}'\boldsymbol{V} = 0$.

2. Дисперсію центрованої змінної (компоненти) обчислюємо так:

$$S_V^2 = \frac{1}{n-1}\boldsymbol{V}_i'\boldsymbol{V}_i = \frac{1}{n-1}\boldsymbol{U}_i'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{U}_i = \boldsymbol{U}_i'\boldsymbol{R}\boldsymbol{U}_i = \lambda_i \boldsymbol{U}_i'\boldsymbol{U}_i = \lambda_i .$$

Тут використана ще умова нормування власних векторів $\boldsymbol{U}_i'\boldsymbol{U}_i = 1$.

Отже, дисперсії компонент дорівнюють власним числам кореляційної матриці, які (доведено вище) не можуть бути від’ємними. ■

Теорема 3. Сума дисперсії компонент дорівнює сумі дисперсії ознак.

Доведення. У коваріаційній матриці $\boldsymbol{S} = (s_{x_i x_j})$ дисперсії ознак $s_{x_i}^2 = s_{x_i x_j}$ розташовуються по діагоналі:

$$\boldsymbol{S} = \begin{bmatrix} s_{x_1x_1} & s_{x_1x_2} & \dots & s_{x_1x_m} \\ s_{x_2x_1} & s_{x_2x_2} & \dots & s_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ s_{x_mx_1} & s_{x_mx_2} & \dots & s_{x_mx_m} \end{bmatrix}.$$

Якщо ознаки попередньо стандартизовані, всі дисперсії будуть дорівнювати одиниці $s_{z_i}^2 = 1$, а коваріаційна матриця буде кореляційною $\boldsymbol{R} = (r_{x_i x_j})$. Отже, сума дисперсій стандартизованих ознак дорівнює m (кількості ознак).

Розглянемо ще раз визначення власних чисел

$$\boldsymbol{RU} = \lambda \boldsymbol{U},$$

звідки маємо

$$(\boldsymbol{R} - \lambda \boldsymbol{E})\boldsymbol{U} = 0.$$

Це є системою однорідних рівнянь для визначення компонент власного вектора $\boldsymbol{U}$. Відомо, що система однорідних рівнянь має нетривіальні рішення, якщо її визначник дорівнює нулю:

$$\begin{vmatrix} 1-\lambda & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & 1-\lambda & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_mx_1} & \dots & 1-\lambda \end{vmatrix} = 0.$$

Це становить характеристичне рівняння, яке визначає всі власні числа. Можна розкласти характеристичне рівняння за степенями λ і звести його до більш звичного вигляду. Наприклад, для $m = 3$ одержимо:

$$\lambda^3 - (1+1+1)\lambda^2 + \left(\begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_2x_1} & 1 \end{vmatrix} + \begin{vmatrix} 1 & r_{x_1x_3} \\ r_{x_3x_1} & 1 \end{vmatrix} + \begin{vmatrix} 1 & r_{x_2x_3} \\ r_{x_3x_2} & 1 \end{vmatrix} \right)\lambda - \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & 1 & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & 1 \end{vmatrix} = 0.$$

Але за теоремою Вієта маємо:

$$\lambda^3 - (\lambda_1 + \lambda_2 + \lambda_3)\lambda^2 + (\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)\lambda - \lambda_1\lambda_2\lambda_3 = 0,$$

тобто $\lambda_1 + \lambda_2 + \lambda_3 = 1+1+1 = 3$. Отже, у загальному випадку буде $\sum \lambda_i = m$, де m – сума дисперсій стандартизованих змінних. ■

Кілька перших (головних) компонент з найбільшими власними числами пояснюють майже всю загальну змінність системи ознак, а рештою компонент з нульовими і малими значеннями власних чисел можна знехтувати.

Існують декілька рекомендацій щодо того, які власні числа вважати "малими".

Наведемо дві найпростіші й найпоширеніші рекомендації.

Розташуємо власні числа в спадному порядку:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_m .$$

У модель включаємо стільки компонент, щоб пояснити близько 80% загальної змінності: $\frac{1}{m}\sum_{i=1}^{k}\lambda_i \approx 0{,}8$.

Відкидаємо всі компоненти, для яких $\lambda_i < 1$(якщо дисперсія компоненти менша за дисперсію стандартизованої ознаки).

Існують складніші рекомендації.

Розглянемо повну систему компонент, яку можна записати так:

$$\boldsymbol{V} = \boldsymbol{ZU}, \tag{2.36}$$

де $\boldsymbol{V}$ – матриця розміру $(n \times m)$, розмір матриці $\boldsymbol{Z}$ – $(n \times m)$, розмір $\boldsymbol{U}$ – $(m \times m)$. Матриця $\boldsymbol{U}$ – ортогональна, тобто $\boldsymbol{UU'} = \boldsymbol{E}$.

Помножимо матричну рівність $\boldsymbol{V} = \boldsymbol{ZU}$ на $\boldsymbol{U'}$ ліворуч і одержимо вирази вихідних (стандартизованих) ознак через компоненти:

$$\boldsymbol{Z} = \boldsymbol{VU'}. \tag{2.37}$$

Це точні рівняння. А тепер врахуємо, що декілька останніх стовпчиків матриці $\boldsymbol{V}$ або нульові (якщо для них $\lambda = 0$) або близькі до нульових (якщо їх $\lambda \approx 0$). Якщо в матриці $\boldsymbol{V}$ залишити тільки k головних компонент, для яких $\lambda \geq 1$, то одержимо кореляційні рівняння:

$$\hat{\boldsymbol{Z}} = \boldsymbol{VU'}, \tag{2.38}$$

де тепер розмір матриці $\boldsymbol{V}$ – $(n \times k)$; в матриці $\boldsymbol{U'}$ залишено тільки перші k рядків (перші k власних векторів), тобто її розмір $(k \times m)$; $\hat{\boldsymbol{Z}}$ – розрахункові значення матриці $\boldsymbol{Z}$.

У кореляційних рівняннях $\boldsymbol{Z}_i$ – стандартизовані, а $\boldsymbol{V}_i$ – лише центровані з дисперсіями $\lambda_i \geq 1$. Нормує їх

$$\boldsymbol{\Phi} = \boldsymbol{VD}_{1/\sqrt{\lambda}}. \tag{2.39}$$

Тут $\boldsymbol{D}_{1/\sqrt{\lambda}}$ – діагональна матриця розміру $(k \times k)$, де по діагоналі розташовані величини $\dfrac{1}{\sqrt{\lambda_i}}$. Оскільки $\boldsymbol{D}_{1/\sqrt{\lambda}}\,\boldsymbol{D}_{\lambda} = \boldsymbol{E}$, то

$$\hat{\boldsymbol{Z}} = \boldsymbol{V}\boldsymbol{U}' = \boldsymbol{V}\boldsymbol{D}_{1/\sqrt{\lambda}}\boldsymbol{D}_{\sqrt{\lambda}}\boldsymbol{U}' = \boldsymbol{\Phi}\boldsymbol{A}', \tag{2.40}$$

де $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda}}$ – матриця факторних навантажень, яка вже була розглянута вище.

Транспонуємо останній вираз $\hat{\boldsymbol{Z}} = \boldsymbol{A}\boldsymbol{\Phi}'$ і перепишемо в координатній формі:

$$\begin{array}{ccccc}
 & \varphi_1 & \varphi_2 & \cdots & \varphi_k \\
\hat{Z}_1 & a_{11} & a_{21} & \dots & a_{k1} \\
\hat{Z}_2 & a_{12} & a_{22} & \dots & a_{k2} \\
\hat{Z}_3 & a_{13} & a_{23} & \dots & a_{k3} \\
\vdots & \vdots & \vdots & & \vdots \\
\hat{Z}_m & a_{1m} & a_{2m} & \dots & a_{km}
\end{array},$$

тобто $\hat{Z}_i = a_{1i}\varphi_1 + a_{2i}\varphi_2 + \ldots + a_{ki}\varphi_k$ для $i = \overline{1,m}$.

Це є рівнянням регресії у стандартизованих змінних, тому факторні навантаження (рядки матриці $\boldsymbol{A}$) є β-коефіцієнтами. Оскільки окремі члени цієї моделі − взаємно ортогональні:

$$\frac{1}{n-1}\boldsymbol{\Phi}'\boldsymbol{\Phi} = \frac{1}{n-1}\left(\boldsymbol{V}\boldsymbol{D}_{1/\sqrt{\lambda}}\right)'\left(\boldsymbol{V}\boldsymbol{D}_{1/\sqrt{\lambda}}\right) = \boldsymbol{D}_{1/\sqrt{\lambda}}\left(\frac{1}{n-1}\boldsymbol{V}'\boldsymbol{V}\right)\boldsymbol{D}_{1/\sqrt{\lambda}} = \boldsymbol{D}_{1/\sqrt{\lambda}} \cdot \boldsymbol{E} \cdot \boldsymbol{D}_{1/\sqrt{\lambda}} = \boldsymbol{E},$$

то одночасно факторні навантаження є коефіцієнтами парної кореляції між вихідними ознаками і компонентами:

$$a_{ij} = r_{x_i\varphi_j} = \beta_{ij}.$$

Обчислюємо коефіцієнти детермінації R_i^2, які показують частку загальної змінності ознаки x_i, що відновлюється знайденою кореляційною залежністю:

$$R_i^2 = \sum_{j=1}^{k}\beta_{ij} \cdot r_{x_i\varphi_j} = \sum_{j=1}^{k}a_{ij}^2 .$$

У факторному аналізі ці величини називаються спільностями (суми квадратів факторних навантажень у рядках матриці $\boldsymbol{A}$ – спільності, суми квадратів факторних навантажень у стовпчиках – власні числа λ_i).

Отже, практично вся корисна інформація міститься в матриці факторних навантажень розміру $(n \times k)$; вона дозволяє відновити кореляційну матрицю $\boldsymbol{R}$, її коефіцієнти.

Факторні навантаження є β-коефіцієнтами регресії в розкладі вихідних ознак за системою факторів-компонент; спільності показують якість апроксимації вихідних ознак дією латентних факторів-компонент; сума врахованих власних чисел – частку загальної змінності, що відновлена моделлю факторного аналізу.

Найкращий ортогональний факторний розв’язок. Як уже було сказано вище, початковий факторний розв’язок у вигляді головних компонент не є єдиним. За допомогою ортогонального перетворення з матрицею $\boldsymbol{T}$ одержують нову матрицю факторних навантажень $\boldsymbol{B} = \boldsymbol{AT}$. Перетворимо кореляційні рівняння $\hat{\boldsymbol{Z}} = \boldsymbol{A\Phi}'$, враховуючи умову ортогональності $\boldsymbol{TT}' = \boldsymbol{E}$:

$$\hat{\boldsymbol{Z}} = \boldsymbol{A\Phi}' = \boldsymbol{AT} \cdot \boldsymbol{T}'\boldsymbol{\Phi}' = \boldsymbol{BF}', \qquad (2.41)$$

де $\boldsymbol{F} = \boldsymbol{\Phi T}$ – нові фактори.

Розглянемо, що дає перехід до нової системи факторів. Нижче схематично зображена матриця факторних навантажень, де великі навантаження (що близькі за модулем до одиниці), позначені зірками, а малі навантаження взагалі не позначені (рис. 2.2). Введемо наступну класифікацію факторів. Фактор називається генеральним, якщо всі його навантаження є великими (на схемі це фактор f_1). Фактор називається спільним, якщо він має два або більше великих навантажень (на рис. 2.2 це фактор f_2).

	φ_1	φ_2	φ_3	R^2
Z_1	*	*	*	R_1^2
Z_2	*	*		R_2^2
Z_3	*	*		R_3^2
Z_4	*			R_4^2
Z_5	*			R_5^2

Рис. 2.2. **Схема факторних навантажень**

Фактор називається характерним, якщо в нього є лише одне велике навантаження. Цей фактор відтворює лише одну ознаку, і краще буде виділити цю ознаку окремо. У факторному аналізі заслуговують на особливу увагу спільні фактори.

Введемо класифікацію змінних за кількістю великих факторних навантажень: на рис. 2.2 змінні Z_4, Z_5 мають складність 1; змінні Z_2, Z_3– складність 2; змінна Z_1– складність 3. Змінні складності 2 і вище важко інтерпретувати. Також незручно мати справу з генеральними факторами.

Початковий факторний розв'язок у вигляді головних компонент призводить до появи генеральних факторів, і тому це рішення не є задовільним. Ортогональними перетвореннями ми намагаємося перейти до еквівалентної факторної матриці, де складність всіх змінних дорівнює одиниці (рис. 2.3).

$$
\begin{array}{ccc c ccc}
 & \varphi_1 & \varphi_2 & & & f_1 & f_2 \\
Z_1 & * & * & & Z_1 & * & \\
Z_2 & * & * & \rightarrow & Z_2 & & * \\
Z_3 & * & & & Z_3 & * & \\
Z_4 & * & * & & Z_4 & & *
\end{array}
$$

Рис. 2.3. **Еквівалентні факторні матриці**

Після перетворення фактор f_1 визначає змінні (Z_1, Z_3), а фактор f_2 – (Z_2, Z_4). Склад змінних, що визначаються кожним фактором, допомагає дати змістовну інтерпретацію факторам і назвати їх.

Для визначення матриці перетворення $\boldsymbol{T}$ запропоновано кілька методів, найпростішим із яких є метод *VARIMAX*. За цим методом потрібно визначити матрицю повороту $\boldsymbol{T}$, щоб максимізувати суму четвертих степенів факторних навантажень. Оскільки сума квадратів факторних навантажень незмінна (вона дорівнює сумі врахованих власних чисел), то максимум суми четвертих степенів приводить до максимуму дисперсії (*VARIANCE*) квадратів факторних навантажень, що й відображено у назві методу. Метод *VARIMAX* реалізовано у будь-якій програмі факторного аналізу на комп'ютері. Немає сенсу описувати деталі цього алгоритму, як і немає сенсу описувати алгоритм визначення власних чисел і власних векторів – усе це реалізовано в сучасному програмному забезпеченні на комп'ютері.

Проблеми факторного аналізу. Факторний аналіз пройшов довгий шлях у своєму розвитку з того часу, коли ще взагалі не було сучасної електронної обчислювальної техніки. Тоді дійсно кожний етап факторного аналізу був проблемою і для їх вирішення було запропоновано багато штучних прийомів. Зараз все це не є предметом для обговорення, і наразі, проблеми факторного аналізу слід розуміти як етапи його реалізації, проте, дотримуючись класичних понять, наведемо оцінку кожної проблеми сьогодні.

Проблема факторів полягає у визначенні найменшої кількості факторів, яких досить для відновлення матриці кореляцій m змінних. Наразі ми не маємо даної проблеми – потрібно обчислити всі власні числа кореляційної матриці і, згідно з прийнятими рекомендаціями, визначити кількість головних компонент. Пригадаємо дві найпростіші й найпоширеніші рекомендації: 1) сума перших власних чисел має бути близькою до $0{,}8 \cdot m$; 2) усі враховані в моделі власні числа мають бути більшими за одиницю: $x_i \geq 1$. Але в минулому обчислення вручну власних чисел і власних векторів матриць великого розміру дійсно було проблемою і навіть автор методу головних компонент Хоттелінг пропонував почекати, доки не з’явиться автоматична обчислювальна техніка.

Проблема спільностей. Кореляційна матриця відтворюється точно, якщо врахувати всі компоненти, і з деякими похибками, якщо врахувати лише головні компоненти $\hat{\boldsymbol{R}} = \boldsymbol{A}\boldsymbol{A}'$. На діагоналі відтвореної матриці тоді будуть не одиниці, а спільності $R_i^2 = \sum_{j=1}^{k} a_{ij}^2$. Було запропоновано редукувати кореляційну матрицю, заміняючи одиниці на діагоналі на оцінки спільностей. У якості оцінки спільностей рекомендовано брати максимальний коефіцієнт кореляції в рядку чи у стовпчику або коефіцієнт детермінації після множинного регресійного аналізу (є й інші рекомендації). Очікувалось, що попередня редукція кореляційної матриці приведе до меншої кількості головних компонент або до більш чіткої структури факторного розв’язку після перетворення повороту. Але чисельні розрахунки на комп’ютері не підтвердили цих очікувань – ніякої переваги від редукції кореляційної матриці не спостерігалося. Водночас були певні негаразди – після редукції окремі власні числа виявилися від’ємними, а отже, вони не можуть бути дисперсіями. Якщо дослідник впевнений у необхідності редукції, то наразі немає ніяких проблем у попередній оцінці спільностей: для цього потрібно провести на комп’ютері факторний аналіз двічі: перший раз − для визначення спільностей, і другий раз – з редукованою матрицею.

Проблема повороту. Матриця повороту $\boldsymbol{T}$ конструюється як сукупність поворотів у площині кожних двох факторів. Наприклад, для $k=3$ перший цикл перетворень буде $\boldsymbol{T}_{11}\boldsymbol{T}_{13}\boldsymbol{T}_{23}$, де $\boldsymbol{T}_{ij}$ – матриця повороту у площині факторів f_i, f_j. Ці цикли перетворень повторюються, доки збільшується сума четвертих степенів факторних навантажень. У минулому під час розрахунків вручну це, можливо, й було проблемою, але наразі всі ітераційні перетворення виконує комп'ютер і це не є вже проблемою.

Проблема вимірювання факторів полягає в обчисленні значень факторів f_i для кожного об'єкта (кожного спостереження). У минулому ЕОМ мали невеликий обсяг оперативної пам'яті і тому не зберігали всі проміжні матриці. Потрібно було, знаючи лише кінцеву матрицю факторних навантажень $\boldsymbol{B}$, виразити $\boldsymbol{F}$ через $\boldsymbol{Z}$, а ми маємо навпаки $\boldsymbol{Z}=\boldsymbol{F}\boldsymbol{B}'$. Це було вирішено у вигляді:

$$\boldsymbol{F}=\boldsymbol{Z}\boldsymbol{C}, \text{ де } \boldsymbol{C}=\boldsymbol{B}\left(\boldsymbol{B}'\boldsymbol{B}\right)^{-1}. \tag{2.42}$$

Сучасні комп'ютери мають практично неосяжний обсяг оперативної пам'яті і тому не існує більше проблеми вимірювання факторів. Якщо переглянути попередній ланцюг формул, то відразу одержимо відповідь у вигляді:

$$\boldsymbol{F}=\boldsymbol{Z}\boldsymbol{C}, \text{ де } \boldsymbol{C}=\boldsymbol{U}\boldsymbol{D}_{1/\sqrt{\lambda}}\boldsymbol{T}. \tag{2.43}$$

Для цих розрахунків потрібно було зберігати матрицю власних векторів $\boldsymbol{U}$ і матрицю повороту $\boldsymbol{T}$.

Порівняємо формули у двох колонках:

$\boldsymbol{F}=\boldsymbol{\Phi}\boldsymbol{T}$,

$\Phi=\boldsymbol{V}\boldsymbol{D}_{1/\sqrt{\lambda}}$

$\boldsymbol{V}=\boldsymbol{Z}\boldsymbol{U}$,

звідки

$\boldsymbol{F}=\boldsymbol{Z}\boldsymbol{U}\cdot\boldsymbol{D}_{1/\sqrt{\lambda}}\cdot\boldsymbol{T}$.

Розміри матриць:

$(n\times k)=(n\times m)(m\times k)(k\times k)(k\times k)$.

Можна добавити в ланцюг формул ще $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda}}$, тоді $\boldsymbol{C}=\boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda}}\boldsymbol{T}$.

Якщо відома лише матриця $\boldsymbol{B}$

$\boldsymbol{Z}=\boldsymbol{F}\boldsymbol{B}'$

$(n\times m)=(n\times k)(k\times m)$

Помножимо матричне рівняння на $\boldsymbol{B}$ праворуч $\boldsymbol{Z}\boldsymbol{B}=\boldsymbol{F}\left(\boldsymbol{B}'\boldsymbol{B}\right)$.

Матриця $\boldsymbol{B}'\boldsymbol{B}$ – квадратна, невироджена, тому вона має обернену матрицю $\left(\boldsymbol{B}'\boldsymbol{B}\right)^{-1}$ розміру $(k\times k)$. Помножимо останній матричний вираз на $\left(\boldsymbol{B}'\boldsymbol{B}\right)^{-1}$ праворуч і одержимо $\boldsymbol{F} = \boldsymbol{ZB}\left(\boldsymbol{B}'\boldsymbol{B}\right)^{-1}$.

Розміри матриць:

$\left(n \times k\right) = \left(n \times m\right)\left(m \times k\right)\left(k \times k\right)$.

Ці формули – еквівалентні, тобто матриця $\boldsymbol{C}$ однакова за обох способів її обчислення.

Коректність вирішення задач за допомогою факторного аналізу потребує підтвердження значущості початкової матриці парних кореляцій (коваріацій) та достатності числа узагальнених факторних ознак в аналізі.

Перевірка значущості матриці парних кореляцій здійснюється за допомогою критерію Уілкса, статистика якого оцінюється за формулою:

$$\chi^2 = -\left(n - \frac{1}{6}\left(2m + 5\right)\right) ln|\boldsymbol{R}|, \qquad (2.44)$$

де n, m – число об'єктів і число початкових ознак у задачі;

$|\boldsymbol{R}|$ – обчислюють за допомогою формули $|\boldsymbol{R}| = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_m$;

$\boldsymbol{R}$ – матриця парних кореляцій.

Обчислене значення критерію Уілкса порівнюється з табличним значенням χ^2-розподіл при заданому рівні параметра α і числу степенів свободи $v = \frac{m\left(m-1\right)}{2}$. Значущість кореляційної матриці підтверджується при $\chi^2_p > \chi^2_{\alpha, v}$.

Критерій χ^2 іншої конструкції використовується в оцінці достатності числа виділених загальних ознак – факторів. Для методу головних компонент обчислення χ^2_p -критерію здійснюється за формулою Бартлета:

$$\chi^2_p = -\left(n - \frac{1}{6}\left(2m + 5\right) - \frac{2}{3} r\right) ln\, \boldsymbol{R}_{m-r}, \qquad (2.45)$$

де r – число залишених в аналізі головних компонент F_r;

$\boldsymbol{R}_{m-r}$ обчислюється за формулою:

$$R_{m-r} = \frac{|R|}{\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_r \left(\frac{m - \lambda_1 - \lambda_2 - \ldots - \lambda_r}{m-r} \right)^{m-r}}.$$

Число степенів свободи для χ^2 буде $\nu = \frac{(m-r)^2 - m - r - 1}{2}$, χ_p^2 порівнюється з табличним $\chi^2_{\alpha,\nu}$ і якщо $\chi_p^2 < \chi^2_{\alpha,\nu}$, то приймається гіпотеза про те, що виділені r головних компонент достатньо повно представляють дисперсію m елементарних ознак і решта головних компонент $F_{r+1}, F_{r+2}, \ldots, F_m$ можуть в аналізі не розглядатися через незначний рівень їх інформативності. Якщо $\chi_p^2 > \chi^2_{\alpha,\nu}$, то для достовірності аналізу повинні бути введені додатково інші головні компоненти.

У факторному аналізі для перевірки гіпотези про достатність числа узагальнюючих ознак – факторів – використовується χ^2-критерій Ловлі, який має аналогічне, як і в попередньому критерії, змістовне навантаження:

$$\chi_p^2 = (n-1) \ln \frac{|R^+|}{|R|}, \tag{2.46}$$

де $|R^+|$ і $|R|$ – визначники матриці парних кореляцій, відтвореної ($R^+ = AA'$) та початкової.

Критичні значення для χ^2-критерію знаходять за таблицями при заданому рівні значущості α і числі степенів свободи $\nu = \frac{(m-r)^2 - m - r}{2}$. Гіпотеза про достатність числа узагальнених факторів приймається за умови, якщо $\chi_p^2 < \chi^2_{\alpha,\nu}$.

Існують критерії перевірки схожості факторів, що знайдені різними методами. Також у спеціальній літературі є розробки з оцінки адекватності факторної моделі в цілому, проте ці критерії не знайшли загального визнання через недостатню обґрунтованість основних своїх положень [46; 64; 139].

2.3. Методи кластерного аналізу

Об’єкти в економіці, як правило, мають складну структуру, а отже, сукупності значень ознак, тобто значень показників, перш ніж обробляти, потрібно завжди перевіряти на однорідність, тоді результати обчислень будуть надійними. За допомогою методів автоматичної класифікації можна здійснювати ефективні перевірки. Різні застосування кластерного аналізу зводяться до чотирьох основних завдань: розробки типології чи класифікації; дослідження корисних концептуальних схем групування об’єктів; породження гіпотез на основі дослідження даних; перевірки гіпотез або дослідження для визначення, чи дійсно типи (групи), виділені тим або іншим способом, присутні в наявних даних [4; 23; 50; 201; 215; 231; 280].

Існування різних обчислювальних алгоритмів кластерного аналізу пояснюється вимогами завдань, що вирішуються на практиці, а також тим, що теорія кластерного аналізу ще далека від завершення. Фахівці з даного методу застерігають, що, по-перше, багато методів кластерного аналізу не мають достатнього статистичного обґрунтування; по-друге, дані методи розроблялися для багатьох наукових дисциплін і тому несуть у собі їх відбиток; по-третє, різні кластерні методи можуть породжувати й породжують різні рішення для одних і тих же даних; по-четверте, сама процедура кластеризації може привнести певну структуру, якої насправді немає [15; 49; 51; 191].

Сучасними різновидами кластерного аналізу є детерміновані процедури, проте глибоке вивчення можливостей розшарування на кластери приводить до висновку, що даний метод має бути заснований на імовірнісних підходах з обчисленням ймовірностей належності об’єкта до того чи іншого класу. Техніка кластерного аналізу базується на поняттях схожості об’єктів (типів об’єктів). Підбором найподібніших об’єктів виконується розподіл сукупності на групи (кластери, класи, таксони). На відміну від звичайних групувань за однією ознакою кластерний аналіз виявляє природне розшарування величин на групи з урахуванням всіх ознак одночасно. Проте строго аналітично не визначаються чіткі межі кожної групи (це право залишається за ОУР – особою, що ухвалює рішення). У комп’ютерних реалізаціях методу також не визначається, на яку кількість груп розподіляється сукупність, – це також має зробити ОУР. Такий стан у кластерному аналізі свідчить про незавершеність теорії і є його недоліком, що не заважає його широкому застосуванню [3; 15; 51; 62; 68; 171; 191].

Умови задачі в економіці, яка так чи інакше зводиться до задачі кластеризації, визначають перелік проблем, які необхідно вирішити для її розв’язання. Незважаючи на те що існує багато обчислювальних процедур, у результаті яких отримуються різні види дендрограм, тобто декілька різних розшарувань об'єктів, наразі існують уже надійні методи кластерного аналізу, що пройшли численні перевірки [142; 147; 156; 167]. Однією з таких процедур, що заслуговує на довіру, є метод Уорда. Звичайно, результати класифікації, у тому числі і за методом Уорда, потребують незалежної перевірки. Перевірка усталеності отриманої класифікації за методом Уорда за допомогою дискримінантного аналізу, заснованого на відстані Махаланобіса (це буде обговорюватися нижче), підтверджує якість класифікації.

Обчислювальні процедури кластерного аналізу передбачають вирішення наступних математичних проблем.

Проблема вибору типу кластерного аналізу. Методи кластерного аналізу підрозділяються на сім великих груп: ієрархічні агломеративні методи; ієрархічні дивізимні методи; ітеративні методи групувань; методи пошуку модальних значень щільності; факторні методи; методи згущень; методи, що використовують теорію графів. Але на практиці, особливо в процесі розв'язування задач в економіці, найчастіше використовують дві великі групи: агломеративні (об'єднавчі) та дивізимні (розподільчі). Агломеративні методи послідовно об'єднують окремі об'єкти в групи (кластери), а дивізимні послідовно розділяють великі групи на менші – однорідніші групи. У свою чергу, кожен метод як об'єднавчого, так і розподільчого типу може бути реалізований за допомогою різних алгоритмів. Більшість із цих алгоритмів була перевірена на відомих класифікаціях. Слід урахувати існуючу в аналітиків думку про те, що ніякою кількістю успішних прикладів не можна підтвердити математичну методику, але одного невдалого прикладу достатньо, щоб відбраковувати неправильний алгоритм. Ці пробні обчислення були використані для відбраковування явно неефективних методів кластерного аналізу. У цілому вважається, що ієрархічний агломеративний кластерний аналіз найдоцільніший та найдоступніший у використанні для великих сукупностей значень величин ознак [167; 187; 280].

Ієрархічні процедури майже не допускають помилок під час початкового об'єднання (перші ітерації під час об'єднання початкових об'єктів), і тільки далі зі збільшенням кількості ітерацій можуть об'єднуватися не досить схожі об'єкти.

Основна ідея ієрархічних алгоритмів полягає в наступному. На першому кроці кожен об'єкт вважається окремим кластером. На наступному кроці об'єднуються два найближчі об'єкти, які утворюють новий кластер. Далі ця процедура повторюється, доки всі об'єкти не об'єднаються в один кластер.

Результати роботи ієрархічної процедури зазвичай оформляються у вигляді дендрограми, в якій наведені номери об'єднаних об'єктів та значення міри схожості (S_{ij}), за якою ці об'єкти були об'єднані. Об'єкти об'єднуються в кластери за мірою схожості між ними. Протилежне поняття до міри схожості – це відстань між об'єктами (d_{ij}). Міра схожості корисна на заключній стадії для кращої інтерпретації результатів. Однакові об'єкти мають схожість $S = 100$ (100%), абсолютно різнорідні об'єкти – схожість $S = 0$. На якомусь критичному рівні схожості дендрограма розрізається та розпадається на окремі кластери, виявлення яких є метою будь-якого економічного аналізу. Найчастіше дендрограму розрізають за рівнем схожості $S = (50 - 60)\%$. Отже, якщо реалізовано метод з урахуванням міри схожості, автоматично розв'язується проблема визначення кількості кластерів.

Проблема нормування значень показників ознак об'єкта. Оцінка відстані між об'єктами надзвичайно залежить від абсолютного значення величини ознаки та від ступеня її варіації в сукупності значень. Щоб ліквідувати подібний вплив на процедуру класифікації, вихідні значення показників нормуються. Найпопулярнішою є стандартизація всіх показників. У процесі стандартизації як норма приймається стандартне відхилення (середньоквадратичне відхилення S_x):

$$z_i = \frac{x_i - \bar{x}_i}{S_{x_i}}, \qquad (2.47)$$

де $\overline{x_i}$ – середнє значення показника.

Звичайно, даний спосіб нормування не позбавлений недоліків: за стандартизації всі обчислення прив'язуються до конкретної вибірки спостережень, яка може бути будь-якою, у тому числі такою, що викликає сумніви щодо своєї репрезентативності. Різні показники по-різному впливають на класифікацію. Зокрема, маловаріабельні величини ознак практично не впливають на процес, але після стандартизації всі показники стають рівноправними за дисперсією, що дорівнює одиниці, у тому числі й маловаріабельні ознаки. Тому стандартизацію необхідно доповнювати процедурою попереднього відбраковування, що відноситься до перевірки робастності величини ознак. Відбраковувати потрібно ті ознаки, які мають коефіцієнт варіації величини, менший за 2% чи навіть 5% (про це йшла мова в розділі 1). У деяких випадках стандартизацію проводять з урахуванням вагових коефіцієнтів, які задаються пропорційно до ступеня важливості ознак, а отже, їх показників.

Альтернативне нормування, що враховує відмінності у варіації окремих показників:

$$z_i = \frac{x_i}{\bar{x}_i}, \tag{2.48}$$

позбавлене багатьох недоліків попередньої, але воно залишилось прив'язаним до конкретної вибірки. Модифікацією альтернативного нормування є:

$$z_j = \frac{x_j}{x_{j\,max} - x_{j\,min}}. \tag{2.49}$$

У деяких умовах управління діяльністю підприємств має сенс ще один вид нормування:

$$z_i = \frac{x_i}{a_i}, \tag{2.50}$$

де a_i – визначене (економічно обґрунтоване) еталонне значення величини даної ознаки (якщо це значення відоме).

Проблема усталеності класифікації (за ознаками). Результати класифікації в скороченому просторі усталеніші та надійніші, ніж у багатовимірному; мала кількість параметрів легше піддається змістовному сприйняттю та подальшому аналізу, ніж велика кількість; у випадку скорочення простору до розмірності 1 – 3 дані стають візуально спостережуваними, а наочність корисна в усіх відношеннях; скорочення кількості ознак приводить до спрощення обчислювальних процедур класифікації.

Для скорочення простору ознак доцільно спочатку використовувати факторний аналіз, який на основі об'єктивно існуючих залежностей між показниками ознак дозволяє виконувати частину редукції простору ознак.

Проблема відстані між об'єктами. Вибір метрики є вузловим моментом дослідження, від якого залежить кінцевий варіант розподілу сукупності об'єктів на класи. Схожість чи відмінність між об'єктами, що класифікуються, встановлюється залежно від метричної відстані між ними. Якщо об'єкт сукупності описується m показниками, то він може бути представлений точкою в m-вимірному просторі і схожість з іншими об'єктами буде визначатися відстанню до інших об'єктів сукупності. У кластерному аналізі використовуються різні міри відстаней між об'єктами:

1) евклідова відстань:

$$d_{ij} = \sqrt{\sum_{k=1}^{m} (z_{ik} - z_{jk})^2}, \; i, j = \overline{1,n}, \; k = \overline{1,m}; \quad (2.51)$$

2) зважена евклідова відстань:

$$d_{ij} = \sqrt{\sum_{k=1}^{m} w_k (z_{ik} - z_{jk})^2}; \quad (2.52)$$

3) відстань city-block:

$$d_{ij} = \sum_{k=1}^{m} \left| z_{ik} - z_{jk} \right|; \quad (2.53)$$

4) узагальнена відстань Мінковського:

$$d_{ij} = (\sum_{k=1}^{n} \left| z_{ik} - z_{jk} \right|^p)^{1/p}; \quad (2.54)$$

5) відстань Махаланобіса:

$$d_{ij}^2 = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \times \boldsymbol{S}^{-1} \times (\boldsymbol{x}_i - \boldsymbol{x}_j), \quad (2.55)$$

де d_{ij} – відстань між i-м та j-м об’єктами;

w_k – вагові коефіцієнти, щодо k-го показника;

x_{ki}, x_{kj} – значення k-го показника відповідно для i-го та j-го об’єктів;

z_{ki}, z_{kj} – те ж для стандартизованих величин;

$\boldsymbol{x}_i, \boldsymbol{x}_j$ – вектори значень показників для i-го та j-го об’єктів;

$\boldsymbol{S}$ – загальна коваріаційна матриця.

Щодо відстані Махаланобіса, яка, на думку деяких фахівців, є найбільш доцільною в реалізації процедур кластерного аналізу, то для її застосування потрібно обчислити матрицю коваріацій за всіма показниками, знайти обернену матрицю і скласти квадратну форму з цією матрицею. Як на позитивну якість відстані Махаланобіса вказують на її інваріантність відносно геометричних лінійних перетворень. За відсутності кореляції між показниками відстань Махаланобіса зводиться до евклідової відстані з попередньою стандартизацією всіх показників. Проте ця відстань має багато негативних якостей, у тому числі всі негативи, що обумовлені стандартизацією (про що було сказано раніше). Крім того, у разі зростання кореляції ці ефекти нормування не зникають, а, навпаки, набувають нових, ускладнених форм. Якщо кореляції близькі до одиниці, визначник коваріаційної матриці наближається до нуля, а сама матриця є близькою до виродженої. Урахування коваріацій обумовлює залежність відстані між двома точками від відстаней між іншими точками. Наявність таких негативних рис може спотворити всю матрицю цих відстаней. Тому громіздка міра Махаланобіса, на думку автора, не має жодних переваг і не повинна рекомендуватися до широкого застосування.

Однією з найкращих мір відстаней між окремими об'єктами є так звана "дивергенція":

$$d_{ij} = \frac{1}{m}\sum_{k=1}^{m}\left|\frac{x_{ki} - x_{kj}}{x_{ki} + x_{kj}}\right|. \tag{2.56}$$

Дана міра не потребує ніякого нормування та враховує відносні внески окремих членів. За умов однакової різниці $\left|x_{ki} - x_{kj}\right|$ внесок тим менший, чим більші абсолютні значення $\left|x_{ki} + x_{kj}\right|$.

Проблема відстаней між кластерами. Одним з основних моментів в обчисленнях кластерного аналізу є вибір відстані між групами об'єктів (між кластерами). Відомо близько дванадцяти різних методів, чотири з яких найпоширеніші: близького, далекого, середнього зв'язків, центроїдного зв'язку та метод Уорда. Відмінності між цими відстанями випливають із самих назв принципів. Наприклад, принцип центроїдного зв'язку свідчить, що відстань між групами вважається такою, що дорівнює відстані між їх центрами. Вибір виду відстані між кластерами обумовлюється умовами конкретної задачі.

Принцип Уорда (Ward) слід вважати найкращим принципом відстані між кластерами, згідно з яким два кластери будуть найближчими, якщо в разі їх об'єднання мінімізується приріст загальної дисперсії. Цільова функція подається як середньогрупова сума квадратів чи сума квадратів відхилень (СКВ), що обчислюються за формулою:

$$CKB = x_j^2 - \frac{\left(\sum_j x_j\right)^2}{n},$$

де x_j – значення величини ознаки j-го об'єкта.

Спочатку, коли кожен кластер складається з одного об'єкта, СКВ дорівнює 0. За методом Уорда об'єднуються ті групи чи об'єкти, для яких СКВ отримує мінімальний приріст. Метод передбачає знаходження кластерів майже однакових розмірів, які мають гіперсферичну форму. Необхідно вважати це перевагою використання саме методу Уорда для класифікації об'єктів у вимірюванні величини його ознак.

Відстань між об’єктами в методі Уорда передбачається середньою евклідовою:

$$d_{ij} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(z_{ki} - z_{kj})^2}. \quad (2.57)$$

Використання цієї відстані теоретично обґрунтоване у випадках, коли:

спостереження вибирають з генеральних сукупностей, що мають багатовимірний нормальний розподіл з коваріаційною матрицею виду $\sigma^2 E_k$, тобто складові X взаємно незалежні та мають одну й ту ж дисперсію;

складові вектора спостережень X однорідні за фізичним змістом (якщо різнорідні, то вони нормуються) та однаково важливі для класифікації.

Проблема міри схожості. Часто відстань між об’єктами має шкалу від нуля до нескінченності, тому міра схожості є визначеним функціональним перетворенням відстані, наприклад, у вигляді так званої потенційної функції:

$$S_{ij} = \frac{100}{1 + d_{ij}^2} \text{ або } S_{ij} = \frac{100}{1 + \frac{1}{2}d_{ij}^2}$$

(при $d_{ij} = 0$ отримується $S_{ij} = 100$, при $d_{ij} \to \infty$ – $S_{ij} \to 0$).

Міра відстані між об’єктами та міра відстані між кластерами – основні поняття в кластерному аналізі. Відстань між об’єктами на сукупності ознак є багатокритеріальною функцією мети, зведеним показником якості, нехай не в абсолютних, а тільки у відносних одиницях. Проблеми обґрунтованої побудови багатокритеріальної функції мети виділені в окрему задачу, і з цього приводу щодо різних запропонованих конструкцій існують суперечливі думки.

Проблема стійкості класифікації. Для перевірки достовірності рішень кластерного аналізу рекомендується використовувати: тести значущості ознак, повторну вибірку, тести значущості для незалежних ознак, метод Монте-Карло [4; 280]. Проте умови задачі класифікації й обґрунтовують використання того чи іншого методу перевірки. Дану проблему у вирішенні економічних задач можна розглядати неоднозначно. По-перше, рекомендується виконувати тест на робастність величин ознак; по-друге, класифікацію об’єктів раціонально здійснювати у скороченому просторі ознак, тобто попередньо виконати редукцію ознак об’єкта; по-третє, слід перевіряти усталеність класифікації в динаміці, аналізуючи зміни розшарування їх у кожен момент часу протягом зазначеного періоду; по-четверте, отриману класифікацію об’єктів необхідно діагностувати за допомогою дискримінантного аналізу, який передбачає наявність навчальних вибірок, тобто проводить "класифікацію з учителем".

2.4. Методи дискримінантного аналізу

Дискримінантний аналіз призначений для з’ясування того, до якого з відомих класів потрібно віднести новий об’єкт, що характеризується певною множиною значень кількісних ознак. Передбачається, що для кожного класу вже складені характерні вибірки об’єктів (навчальні підсукупності). Ці вибірки зазвичай складаються кваліфікованими спеціалістами-експертами, які для діагностики (віднесення об’єкта до того чи іншого класу) можуть використовувати всі доступні їм знання, у тому числі й неформалізовані [304].

Розглянемо основні проблеми дискримінантного аналізу.

Проблема процедури класифікації. Класичний дискримінантний аналіз складається з формулювання деяких правил, критеріїв класифікації, "інформантів", "дискримінантних функцій", які дозволяють оцінити ймовірність належності нового об’єкта до кожного класу. Зазвичай у якості таких критеріїв класифікації вибирають відстань Махаланобіса від об’єкта до центрів кожного класу. За цим критерієм об’єкт відноситься до найближчого класу. Цю стандартну методику (ядро, основу дискримінантного аналізу) можна доповнити. По-перше, деякі класи (групи) об’єктів більш поширені, розповсюджені, ніж інші, і за однакових умов мають деякі переваги. Якщо є апріорна інформація про кількісне співвідношення (про обсяг) класів, тобто про ймовірності q_i належності об’єктів до кожного класу ($q_1 + q_2 + \ldots + q_m = 1$), то цю додаткову інформацію слід врахувати у правилах класифікації (інформантах). По-друге, можна оцінити наслідки від можливих помилок класифікації. Якщо експерти складуть матрицю збитків c_{ij} (від неправильного віднесення об’єкта з класу i до класу j), то й цю додаткову інформацію можна врахувати в класифікаційних правилах, які тепер будуть оцінювати величину можливих збитків від віднесення об’єкта до того чи іншого класу. Об’єкт потрібно відносити до того класу, для якого ризик (величина можливих збитків) буде найменшим.

Теоретично можливо було б урахувати особливості багатовимірного розподілу об’єктів у кожному класі та прийняти замість відстані Махаланобіса обґрунтованішу міру. Проте ця можливість залишається практично не реалізованою, оскільки з багатовимірних розподілів достатньо добре вивчений лише багатовимірний нормальний закон, що і визначає вибір в якості міри саме відстань Махаланобіса.

Відстань Махаланобіса є квадратичним інформантом – квадратичною функцією виміряних ознак об'єктів. Проте на практиці вводять додаткове спростовувальне припущення про рівність всіх коваріаційних матриць для кожного класу (класи тепер відрізняються тільки центрами групувань). Зазвичай навчальні підсукупності (характерні вибірки) надто малі, щоб для кожної вибірки достатньо надійно можна було визначити дисперсії та коваріації ознак – саме тому і доводиться використовувати загальну (об'єднану) коваріаційну матрицю. Вимушене прийняття припущення про рівність коваріаційних матриць для всіх класів істотно спростовує (та полегшує) класифікаційні правила – тепер замість квадратичних інформантів одержали зручніші лінійні інформанти. Викладене вище в загальних рисах описує методи класичного дискримінантного аналізу.

Останнім часом класичний дискримінантний аналіз доповнений теорією "канонічних дискримінантних функцій", або просто "дискримінантних функцій". У зв'язку з цим пропонується за класифікаційними правилами залишити назву "інформанти" та не використовувати для них назву "дискримінантні функції", що тепер будуть позначати дещо інше.

Дискримінантними функціями за новим змістом є деякі узагальнюючі (агреговані) змінні, у просторі яких найчіткіше видно відмінності між класами (відомими підсукупностями). Ці узагальнюючі змінні складаються у вигляді лінійних комбінацій початкових ознак, а коефіцієнти цих лінійних комбінацій визначаються з умови максимуму певного функціонала, що характеризує відмінності між класами. Декілька перших (головних) дискримінантних функцій вичерпно описують практично всі відмінності між класами.

Попередній перехід до скороченого простору головних дискримінантних функцій усуває багато проблем класичного дискримінантного аналізу – результати стають стійкішими й надійнішими. Справа в тому, що відстань Махаланобіса враховує кореляції між ознаками, що вважається позитивною властивістю даної міри. Але, з іншого боку, за високих кореляцій ця міра стає нестійкою, а за точної мультиколінеарності ознак – виродженою. Слід зазначити, що спеціалісти не зобов'язані перевіряти (та й не перевіряють) незалежність ознак у системі, що визначає об'єкт. У скороченому просторі (головних) дискримінантних функцій виродженість виключається, оскільки дискримінантні функції ортогональні. Крім того, часто для опису відмінностей між класами достатньо всього двох перших дискримінантних функцій і тоді з'являється можливість візуального відображення об'єктів кожного класу (і нового спірного об'єкта) на площині, що є надзвичайно цінним надбанням.

Теорія канонічних дискримінантних функцій дуже схожа за методами і цілями на теорію головних компонент факторного аналізу.

Проблема наочності (візуалізації). Впровадження канонічних дискримінантних функцій здійснюється через створення упорядкованих за значущістю функцій, аналогічних до головних компонентів факторного аналізу. Перша дискримінантна функція обирається так, щоб її середні значення для різних класів відрізнялись найістотніше. Друга дискримінантна функція добирається за ідентичних умов, але вона не повинна корелювати з першою. Третя функція не повинна корелювати з упровадженими раніше функціями. Тоді кілька перших дискримінантних функцій (головних) несуть усю інформацію про відмінність між класами.

Після завершення аналізу можна виконувати зворотний перехід у початковий простір реальних змінних.

Позначимо k – кількість класів; m – кількість дискримінантних ознак; h_i – кількість об'єктів (спостережень) класу p; n – загальна кількість об'єктів усіх класів. У дискримінантному аналізі приймають наступні математичні допущення:

1) існує два чи більше класів $k \geq 2$;

2) у кожному класі маємо принаймні два об'єкти $h_p \geq 2$;

3) кількість дискримінантних ознак необмежена, але не перевищує загальну кількість об'єктів мінус два, $0 < m < (n-2)$;

4) вимірювання дискримінантних ознак повинне бути здійснене за інтервальною шкалою;

5) дискримінантні ознаки лінійно незалежні;

6) передбачається приблизна рівність коваріаційних матриць для кожного класу (якщо не використовуються спеціальні прийоми);

7) приймається гіпотеза про багатовимірну нормальність закону розподілу об'єктів для кожного класу [280].

Зазвичай відмінності між класами за кількісними значеннями однієї ознаки оцінюються за допомогою дисперсійного аналізу. Робиться це таким чином. Нехай y – деяка змінна, наприклад, одна з m початкових ознак x_j. Необхідно знаходити її загальне середнє μ і середньогрупові u_p, тобто їх середні значення в кожному класі (k – кількість класів, p – номер класу, h_p – кількість об’єктів у кожному класі).

Слід обчислити суми квадратів відхилень: загальну SS_y, міжгрупову SS_u, внутрішньогрупову SS_e, між якими існує простий зв’язок $SS_y = SS_e + SS_u$, тобто

$$\sum_{p=1}^{k}\sum_{q=1}^{h_p}(y_{pq}-\mu)^2 = \sum_{p=1}^{k}\sum_{q=1}^{h_p}(y_{pq}-u_p)^2 + \sum_{p=1}^{k}\sum_{q=1}^{h_p}(u_p-\mu)^2. \quad (2.58)$$

У такому записі цей вираз виглядає дещо громіздким, тому Гаус запропонував позначати квадратними дужками підсумовування за всіма n спостереженнями і не вказувати індексів підсумовування у квадратних дужках. Тоді вищенаведений громіздкий вираз можна записати так:

$$[(y-\mu)^2] = [(y-u)^2] + [(u-\mu)^2]. \quad (2.59)$$

Далі обчислюються наступні статистики (чисельні характеристики):

$$\eta^2 = \frac{SS_u}{SS_y},\ \lambda = \frac{SS_u}{SS_e},\quad F = \lambda\frac{n-k}{k-1}, \quad (2.60)$$

де η^2 – кореляційне відношення;

F – дисперсійне відношення Фішера.

Чим більші величини цих статистик, тим більш потужним дискримінатором є ознака y.

Для останньої статистики – дисперсійного відношення Фішера – існують таблиці, де для кожної пари числа ступенів свободи наведені квантілі $F_{0,05}$ і $F_{0,01}$. Якщо обчислене значення $F < F_{0,05}(k-1, n-k)$, то ознака y є поганим дискримінатором, за її значеннями не можна зробити надійних висновків про відмінності між класами. Ознаку y можна вважати добрим дискримінатором, якщо $F < F_{0,01}(k-1, n-k)$.

Перша статистика $\eta^2 = \dfrac{\lambda}{1+\lambda}$ (кореляційне відношення) змінюється від 0 (ознака y не є дискримінатором) до 1 (ознака y – ідеальний найпотужніший дискримінатор).

Поширимо цю методику на багатовимірний простір.

Якщо є кілька ознак-дискримінаторів, можна ввести комплексний дискримінатор, найкращий за критерієм максимуму λ. Таку "дискримінантну функцію" слід складати як лінійну комбінацію початкових ознак

$$y = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m = \sum_{j=1}^{m} \beta_j x_j \,, \tag{2.61}$$

де коефіцієнти β_j слід визначити за умови максимуму λ:

$$\lambda = \frac{SS_u}{SS_e} \to \max \,. \tag{2.62}$$

Для початкових ознак x_j потрібно знайти їх загальні середні μ_j і середньогрупові u_{jp} (середні значення ознак у кожному класі).

Обчислюємо суму квадратів SS_e для синтезованої ознаки y:

$$\begin{aligned} SS_e = \left[(y-u)^2\right] = \left[\left(\sum_{i=1}^{m}\beta_i (x_i - u_i)\right)^2\right] = \left[\sum_{i=1}^{m}\sum_{j=1}^{m}\beta_i\beta_j (x_i - u_i)(x_j - u_j)\right] = \\ = \sum_{i=1}^{m}\sum_{j=1}^{m}\beta_i\beta_j \left[(x_i - u_i)(x_j - u_j)\right] = \sum_{i=1}^{m}\sum_{j=1}^{m}\beta_i\beta_j w_{ij} = \boldsymbol{\beta}^T \boldsymbol{W}\boldsymbol{\beta}, \end{aligned} \tag{2.63}$$

де $\boldsymbol{\beta}^T = (\beta_1, \beta_2, ..., \beta_m)$ – коефіцієнти дискримінантної функції;

$\boldsymbol{W}$ – матриця з елементами

$$w_{ij} = \left[(x_i - u_i)(x_j - u_j)\right] = \sum_{p=1}^{k}\sum_{q=1}^{h_p}(x_{ipq} - u_{ip})(x_{ipq} - u_{jp}). \tag{2.64}$$

Елементи матриці $\boldsymbol{W}$ наведені в коротких позначеннях Гауса і в розгорнутому вигляді.

Аналогічним чином можна перетворити суму квадратів SS_u:

$$SS_u = E\left[(u-\mu)^2\right] = E\left[\left(\sum_{i=1}^{m}\beta_i (x_i-\mu_i)\right)^2\right] = E\left[\sum_{i=1}^{m}\sum_{j=1}^{m}\beta_i\beta_j (x_i-\mu_i)(x_j-\mu_j)\right] =$$
$$= \sum_{i=1}^{m}\sum_{j=1}^{m}\beta_i\beta_j E\left[(x_i-\mu_i)(x_j-\mu_{ji})\right] = \sum_{i=1}^{m}\sum_{j=1}^{m}\beta_i\beta_j b_{ij} = \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{B\beta}, \qquad (2.65)$$

де $\boldsymbol{B}$ – матриця з елементами

$$b_{ij} = E\left[(x_1-\mu_i)(x_j-\mu_j)\right] = \sum_{p=1}^{k}\sum_{q=1}^{h_p}(x_{ip}-\mu_i)(x_{jp}-\mu_j).$$

Отже, коефіцієнти β_j слід визначати за умови максимуму відношення

$$\lambda = \frac{\boldsymbol{\beta}^T\boldsymbol{B\beta}}{\boldsymbol{\beta}^T\boldsymbol{W\beta}} \to max. \qquad (2.66)$$

Можна накласти на коефіцієнти $\boldsymbol{\beta}$ умову нормування $\boldsymbol{\beta}^T\boldsymbol{W\beta} = const$, де константа дорівнює або одиниці, або кількості степенів свободи $\nu_e = n-k$. Тоді буде отримано таку задачу на умовний екстремум:

знайти максимум квадратичної форми (міжгрупової суми квадратів)

$$\boldsymbol{\beta}^T\boldsymbol{B}\,\boldsymbol{\beta} \to max \qquad (2.67)$$

за умови нормування $\boldsymbol{\beta}^T\boldsymbol{W\beta} = const$.

Для розв'язання цієї задачі на умовний екстремум складемо функцію Лагранжа (де γ – множник Лагранжа):

$$L = \boldsymbol{\beta}^T\boldsymbol{B}\,\boldsymbol{\beta} - \gamma\,\boldsymbol{\beta}^T\boldsymbol{W}\,\boldsymbol{\beta}, \qquad (2.68)$$

і прирівняти до нуля її частинні похідні за компонентами вектора $\boldsymbol{\beta}$. Наприкінці буде отримано таку систему рівнянь: $\boldsymbol{B\beta} = \gamma\boldsymbol{W\beta}$, де $\boldsymbol{B}, \boldsymbol{W}$ – симетричні напівдодатні означені матриці, причому матриця $\boldsymbol{W}$ не вироджена тобто, додатно визначена. Це так звана узагальнена проблема власних значень і власних векторів матриці $\boldsymbol{B}$. Множник Лагранжа γ виявився власним числом матриці $\boldsymbol{W'}\boldsymbol{B}$.

Теорема. Для симетричних додатно означених матриць $\boldsymbol{B}, \boldsymbol{W}$ всі узагальнені власні числа γ – дійсні, невід'ємні, а власні вектори – $\boldsymbol{W}$ –

ортонормовані, тобто $\boldsymbol{B}^{T}\boldsymbol{W}\boldsymbol{\beta}^{*}=0$ (але $\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}=const$). Тут з $\boldsymbol{\beta}$ і $\boldsymbol{\beta}^{*}$ визначені власні вектори, що відповідають власним числам $\gamma \neq \gamma^{*}$.

Доведення (від супротивного). Нехай $\gamma=\omega+i\omega^{*}, \beta=\boldsymbol{X}+i\boldsymbol{Y}$. Тоді з матричної рівності $B(\boldsymbol{X}+i\boldsymbol{Y})=(\omega+i\omega^{*})\boldsymbol{W}(\boldsymbol{X}+i\boldsymbol{Y})$ маємо систему

$$\begin{cases}\boldsymbol{BX}=\omega\boldsymbol{WX}-\tilde{\omega}\boldsymbol{WY}\\ \boldsymbol{BY}=\omega\boldsymbol{WY}-\tilde{\omega}\boldsymbol{WX}\end{cases} \text{ або } \begin{cases}\boldsymbol{CX}=-\tilde{\omega}\boldsymbol{WY}\\ \boldsymbol{CY}=-\tilde{\omega}\boldsymbol{WX}\end{cases},$$

де через $\boldsymbol{C}$ позначена матриця $\boldsymbol{C}=\boldsymbol{B}-\omega\,\boldsymbol{W}$.

Помножимо першу матрицю рівності ліворуч на $\boldsymbol{Y}^{T}$, а другу – ліворуч на $\boldsymbol{X}^{T}$:

$$\begin{cases}\boldsymbol{Y}^{T}\boldsymbol{CX}=-\tilde{\omega}\,\boldsymbol{Y}^{T}\boldsymbol{WY}\\ \boldsymbol{X}^{T}\boldsymbol{CY}=-\tilde{\omega}\,\boldsymbol{X}^{T}\boldsymbol{WX}\end{cases}.$$

Враховуючи симетрично матриці $\boldsymbol{C}$, переконаємося, що $\boldsymbol{Y}^{T}\boldsymbol{CX}=\boldsymbol{X}^{T}\boldsymbol{CY}$, звідки одержуємо $O=\tilde{\omega}\,(\boldsymbol{X}^{T}\boldsymbol{CX}+\boldsymbol{Y}^{T}\boldsymbol{CY})$.

Матриця $\boldsymbol{W}$ – додатно визначена, тому $\boldsymbol{X}^{T}\boldsymbol{CX}>0$, $\boldsymbol{Y}^{T}\boldsymbol{CY}>0$ і з попередньої матричної рівності одержуємо $\tilde{\omega}=0$. Отже, всі узагальнені власні числа γ – дійсні.

Тепер розглянемо два узагальнені вектори $\beta,\ \beta^{*}$, що відповідають різним власним числам $\gamma \neq \gamma^{*}$:

$$\begin{cases}\boldsymbol{B\beta}=\gamma\boldsymbol{W\beta}\\ \boldsymbol{B\beta}^{*}=\gamma^{*}\boldsymbol{W\beta}^{*}\end{cases}.$$

Помножимо першу матричну рівність ліворуч на $\boldsymbol{\beta}^{*T}$, а другу – на $\boldsymbol{\beta}^{T}$:

$$\begin{cases}\boldsymbol{\beta}^{*T}\boldsymbol{B\beta}=\gamma\,\boldsymbol{\beta}^{*T}\boldsymbol{W\beta}\\ \boldsymbol{\beta}^{T}\boldsymbol{B\beta}^{*}=\gamma^{*}\boldsymbol{\beta}^{T}\boldsymbol{W\beta}^{*}\end{cases}.$$

Унаслідок симетрії матриць $\boldsymbol{B}, \boldsymbol{W}$ маємо:

$$\boldsymbol{\beta}^{*T}\boldsymbol{B}\boldsymbol{\beta}=\boldsymbol{\beta}^{T}\boldsymbol{B}\boldsymbol{\beta}^{*},\quad \boldsymbol{\beta}^{*T}\boldsymbol{W}\boldsymbol{\beta}=\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}^{*},$$

звідки $O=(\gamma-\gamma^{*})\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}^{*}$.

Оскільки $\gamma\neq\gamma^{*}$, то з останньої рівності випливає умова $\boldsymbol{W}$– ортогональності векторів $\boldsymbol{\beta}$ і $\boldsymbol{\beta}^{*}$:

$$\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}^{*}=0.$$

Наприкінці доведемо, що $\gamma\geq 0$. Врахуємо, що матриця $\boldsymbol{W}$– додатно визначена, тобто $\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}^{*}=\|\beta\|^{2}>0$; матриця $\boldsymbol{B}$– напівдодатно визначена, тобто $\boldsymbol{\beta}^{T}\boldsymbol{B}\,\boldsymbol{\beta}^{*}\geq 0$. Додатна і напівдодатна визначеність випливає зі структури цих матриць $\boldsymbol{B}=\boldsymbol{Z}'\boldsymbol{Z}$ (обидві ці матриці є кореляційними і тому мають таку структуру). Як було доведено вище, власні числа таких матриць – невід'ємні, звідки випливає, що будь-яка квадратична форма з такими матрицями ніколи не може набувати від'ємних значень.

Помножимо матричну рівність $\boldsymbol{B}\,\boldsymbol{\beta}=\gamma\boldsymbol{W}\boldsymbol{\beta}$ ліворуч на $\boldsymbol{\beta}^{T}$ і врахуємо, що $\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}\neq 0$. Одержимо $\gamma=\dfrac{\boldsymbol{\beta}^{T}\boldsymbol{B}\,\boldsymbol{\beta}}{\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}}\geq 0$, що і треба було довести. ■

Теорема. Характеристика λ, яка пропорційна дисперсійному відношенню Фішера $F=\lambda\dfrac{n-k}{k-1}$, співпадає з множником Лагранжа γ і є узагальненим власним числом матриці $\boldsymbol{B}$.

Доведення. Дійсно $\lambda=\dfrac{\boldsymbol{\beta}^{T}\boldsymbol{B}\boldsymbol{\beta}}{\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}}=\dfrac{\boldsymbol{\beta}^{T}(\gamma\boldsymbol{W}\boldsymbol{\beta})}{\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}}=\gamma\dfrac{\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}}{\boldsymbol{\beta}^{T}\boldsymbol{W}\boldsymbol{\beta}}=\gamma$, тобто максимальне значення λ дорівнює максимальному (узагальненому) власному числу $\boldsymbol{B}$. ■

Кількість ненульових узагальнених власних чисел матриці $\boldsymbol{B}$ не перевищує меншого з чисел $(k-1; m)$. Необхідно знайти власні вектори $\boldsymbol{\beta}$, що відповідають цим ненульовим числам, і скласти систему канонічних дискримінантних функцій. Перша функція з найбільшим власним числом є найпотужнішим дискримінатором, а функції з малими і нульовими власними числами вже нічого не розрізняють. Внесок інших дискримінантних функцій потрібно оцінити.

Проблема оцінки дискримінантних функцій та їх послідовного відбору. Дискримінантним методом передбачається обчислення таких числових характеристик:

1. Процентний вміст $\frac{\lambda_j}{\sum_{j=1}^{m}\lambda_j}\cdot 100\%$ показує, наскільки важливіша та чи інша дискримінантна функція.

2. Квадрат коефіцієнта канонічної кореляції (кореляційного відношення) $\eta_i^2 = \frac{\lambda_i}{1+\lambda_i}$ показує, яка частка повної мінливості дискримінантної функції пояснюється відмінністю груп.

3. Критерій Фішера $F = \lambda \frac{n-k}{k-1}$, який можна порівняти з табличними значеннями $F_{0,05}(k-1, n-k)$ і $F_{0,01}(k-1, n-k)$ та визначити, чи є значущою одержана дискримінантна функція.

4. Статистика Уїлкса. Слід розглянути її детально. Нехай $l = min(k-1, m)$ – загальна кількість дискримінантних функцій з ненульовими λ_i. Тоді

$$\Delta_0 = \frac{1}{1+\lambda_1}\cdot\frac{1}{1+\lambda_2}\cdot\frac{1}{1+\lambda_3}\cdots\frac{1}{1+\lambda_l} \tag{2.69}$$

є мірою остаточної мінливості, якщо врахувати всі дискримінантні функції; тобто Δ_0 оцінює дискримінантну здатність усієї системи функцій. Тепер потрібно оцінити дискримінантну здатність системи без першої, найважливішої функції:

$$\Delta_1 = \frac{1}{1+\lambda_2}\cdot\frac{1}{1+\lambda_3}\cdots\frac{1}{1+\lambda_l}. \tag{2.70}$$

Ця величина вже перевищує Δ_0. Чим ближче Δ_1 до одиниці, тим слабшою є дискримінантна здатність інших функцій системи, які можна враховувати.

Потім слід обчислити $\Delta_2 = \frac{1}{1+\lambda_3} \dots \frac{1}{1+\lambda_l}$, Δ_z, … до Δ_{l-1}. Значущість послідовних значень Δ_j оцінюється за допомогою критерію Пірсона:

$$\chi_j^2 = -\left[n - \frac{m+k}{2} - 1\right] \ln \Delta_j , \tag{2.71}$$

який слід порівнювати з табличними значеннями $\chi^2_{0,05}(v_j)$ і $\chi^2_{0,01}(v_j)$, де $v_j = (m - j)(k - j - 1)$ – число ступенів свободи.

На певному етапі система функцій, що залишилась, відкидається, і можна отримати систему інформативних показників – систему головних дискримінантних функцій.

Як було наведено вище, дискримінантні функції корисні для візуальної наочності розподілу об'єктів за класами, а також для одержання стабільних інформантів, які слід складати з дискримінантних функцій, а не вихідних ознак. Отже, переходимо до створення класифікаційних функцій (інформантів), за чисельними значеннями яких можна визначати, до якого класу найімовірніше віднести той чи інший об'єкт.

Значення m показників ознак x_i утворюють вектор (матрицю – стовпчик) $\boldsymbol{X} = (x_1, x_2, ..., x_m)'$. Конкретний набір показників для об'єкта, що потрібно ідентифікувати, позначимо як вектор $\boldsymbol{U} = (u_1, u_2, ..., u_m)'$. Об'єкт має належати одному з k класів $\omega_1, \omega_2, ..., \omega_k$, що не перетинаються. Розподіл імовірностей у кожному класі описується диференційними функціями розподілу $f_j(X)$. Відносно цих класів відомі (за даними навчальних вибірок) середні значення й коваріаційні матриці:

$$\boldsymbol{M}_j = (\mu_1(j); \mu_2(j); ..., \mu_m(j)), \quad \boldsymbol{S}_j = \left[s_{\alpha\beta}(j) \right]_{\alpha, \beta=1}^{m} . \tag{2.72}$$

Дуже часто відомі апріорні ймовірності належності об'єкта до кожного з класів $q_j = P(X \in \omega_j)$. Ці ймовірності іноді можна вважати пропорційними обсягам навчальних вибірок; у сумі ймовірності дорівнюють одиниці. Якщо апріорні ймовірності не визначені, їх вважають однаковими, тобто $q_i = \frac{1}{k}$.

Можливі наслідки від неправильної класифікації характеризуються матрицею втрат $C(j|i)$ – це збитки від помилкового віднесення об'єкта, який належить класу ω_i, до іншого класу ω_j. Якщо цю матрицю попередньо не складено, всі збитки вважають однаковими $C(j|i)=1$ $(i \neq j)$; для $j=i$ збитки дорівнюють нулю $C(j|j)=0$. Потрібно скласти таке правило віднесення нових об'єктів U до відповідних класів, щоб зменшити можливі наслідки від неправильної класифікації.

Знаходимо за теоремою Бейєса ймовірність того, що новий об'єкт належить до класу ω_i. За теоремою Бейєса оцінюються внески окремих членів у формулу повної ймовірності:

$$f(X)=q_1 f_1(X)+q_2 f_2(X)+...+q_k f_k(X)=\sum_{\alpha=1}^{k} q_\alpha f_\alpha(X). \tag{2.73}$$

Звідси
$$P(U \in \omega_i \mid X=U)=\frac{q_i f_i(U)}{\sum_\alpha q_\alpha f_\alpha(U)}. \tag{2.74}$$

Отже, тепер можемо обчислити математичне очікування втрат від неправильності віднесення об'єкта U до класу ω_j як суму добутків:

$$\sum_{i=1}^{k} C(j|i)P(U \in \omega_i \mid X=U)=\frac{C(j|1)q_1 f_1(U)+C(j|2)q_2 f_2(U)+...+C(j|k)q_k f_k(U)}{\sum_\alpha q_\alpha f_\alpha(U)}.$$

Можна обчислити k таких виразів, послідовно відносячи об'єкт до кожного з класів $(j=\overline{1,k})$. Об'єкт потрібно віднести до того класу, де втрати від помилкової класифікації будуть найменшими:

$$\begin{cases} \sum\limits_{i=1}^{k} C(j|i)q_i f_i(U) \to \min \\ j=\overline{1,k} \end{cases} \tag{2.75}$$

(сталий знаменник вилучено).

Якщо втрати однакові − $C(j|i)=1$ $(j \neq i)$ (для $j=i$ завжди $C(j|j)=0$), останній вираз спрощується:

$$\sum_{i=1}^{k} C(j|i)q_i f_i(U)=\sum_{i=1}^{k} q_i f_i(U)-q_j f_j(U)=Const-q_j f_j(U) \to \min. \tag{2.76}$$

Звідси випливає правило віднесення об'єкта до того класу, де вираз $q_j f_j(U) \to \max$ набуває максимального значення.

Для багатовимірного нормального розподілу

$$f_j(U) = \frac{1}{(2\pi)^{\frac{m}{2}} \left|S_j\right|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2} D_j^2 (U, M_j)}, \tag{2.77}$$

де $D_j^2(U, M_j) = (U - M_j)' S_j^{-1} (U - M_j)$ – відстань Махаланобіса об'єкта U від центра класу ω_j.

Після логарифмування одержимо так звані квадратичні інформанти:

$$Q_j = Const + \ln q_j - \frac{1}{2} \ln\left|S_j\right| - \frac{1}{2} D_j^2 (U, M_j). \tag{2.78}$$

Об'єкт слід віднести до класу з найбільшим інформантом $Q_j \to \max$.

Якщо обсяги навчальних вибірок не дозволяють надійно оцінити всі коваріаційні матриці S_j, вважають, що для всіх класів коваріаційні матриці однакові й дорівнюють об'єднаній матриці коваріацій $S_j = S$. Тоді перетворюємо вираз для відстані Махаланобіса:

$$D_j^2(U, M_j) = (U - M_j)' S_j^{-1} (U - M_j) = U'S^{-1}U - 2U'S^{-1}M_j + M_j{}' S^{-1} M_j. \tag{2.79}$$

Від усіх виразів Q_j можна відняти сталу $Const - \frac{1}{2}\ln|S| - \frac{1}{2}U'S^{-1}U$ і одержати більш прості лінійні інформанти (лінійні відносно U):

$$L_j = \ln q_j - \frac{1}{2}\ln\left|S_j\right| - \frac{1}{2} M_j{}' S^{-1} M_j + U' S^{-1} M_j \to \max. \tag{2.80}$$

Вираз L_j можна подати у вигляді:

$$L_j = l_0 + l_1 u_1 + l_2 u_2 + \ldots + l_m u_m \to \max, \tag{2.81}$$

$$l_0 = \ln q - \frac{1}{2}\left(l_1 \mu_1^{(j)} + l_2 \mu_2^{(j)} + \ldots + l_m \mu_m^{(j)}\right). \tag{2.82}$$

Нарешті, коли невідомі апріорні ймовірності q_j, їх вважають однаковими і в інформантах (квадратичних і лінійних) можна відкинути сталі члени $\ln q_j$.

2.5. Методи канонічного аналізу

У математиці існує один метод, який установлює зв'язок між двома групами величин ознак, – це метод канонічних кореляцій.

Сферу застосування канонічних кореляцій поки що однозначно не визначено, найпоширеніше пояснення цього стосується факту, що дослідники практично не ознайомлені з цим математичним апаратом.

Метод канонічних кореляцій дозволяє знаходити максимальні кореляційні зв'язки між двома групами випадкових величин. Ця проблема вирішується за допомогою нових змінних – канонічних функцій, визначених як лінійні комбінації початкових ознак. Попередньо всю сукупність елементарних ознак необхідно розділити на дві групи: пояснювальні ознаки і результативні ознаки, що акумулюють наслідки впливу перших. Оскільки в економіці причинно-наслідкові взаємозв'язки приводять у дію механізми функціонування та розвитку об'єктів, то розділити елементарні ознаки на причинні та наслідкові можна, хоча і з певною умовністю. Отже, система елементарних ознак, що характеризує ресурсно-ринкову здатність об'єкта (пояснювальні змінні) протистоїть системі ознак, що виражають результати його функціонування (результативні змінні). З двох наборів змінних – результативних ознак $\{y_j\}$ і елементарних ознак, що впливають на результативні ознаки $\{x_i\}$, – складають дві (канонічні) функції – дві нові змінні, що однозначно визначаються даною множиною величин ознак. Канонічні змінні завжди можна трактувати як комплексні показники якості двох груп ознак, тому вони потребують найпильнішого вивчення [132].

Концепція методу канонічної кореляції започаткована наступними міркуваннями. Множинна кореляція є мірою зв'язку між однією випадковою величиною і множиною інших випадкових величин. Насправді ж множинна кореляція є максимальною кореляцією між однією випадковою величиною й лінійною функцією інших. Ця концепція була узагальнена Хоттелінгом (1936 р.) на випадок зв'язку між двома множинами випадкових величин [15, с. 522 – 523].

Нові канонічні величини вибираються таким чином, щоб нові координати безпосередньо вказували значення кореляції. Окремі показники не характеризують повністю групу, до якої вони належать. За характерним представником груп складають дві лінійні комбінації з показників кожної групи. Коефіцієнти цих лінійних комбінацій визначають із умови максимуму коефіцієнта кореляції між комбінаціями. Знайдена таким чином пара лінійних комбінацій утворює першу пару "канонічних функцій", яка описує певні властивості обох груп первинних показників. Але групи показників мають ще інші властивості, для концентрованого опису яких знаходять другу пару канонічних функцій, яка незалежна (некорельована) з першою парою. Виділяють стільки пар канонічних функцій, поки не буде практично повністю вичерпано всі особливості груп показників. Систему канонічних функцій можна вважати новою координатною системою, яка найкращим чином пристосована для опису сукупності кореляційних зв’язків між показниками різних груп.

У канонічному аналізі ознак об’єкта матриця величин ознак розбивається на дві частини, як це показано в табл. 2.1.

Таблиця 2.1

Матриця значень елементарних ознак

Номер спостереження	Перша група ознак				Друга група ознак			
	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	...	$\boldsymbol{X}_q$	$\boldsymbol{Y}_1$	$\boldsymbol{Y}_2$	...	$\boldsymbol{Y}_p$
1	x_{11}	x_{12}	...	x_{1q}	y_{11}	y_{12}	...	y_{1p}
2	x_{21}	x_{22}	...	x_{2q}	y_{21}	y_{22}	...	y_{2p}
3	x_{31}	x_{32}	...	x_{3q}	y_{31}	y_{32}	...	y_{3p}
…	...	...	...	...	...	...	...	...
n	x_{n1}	x_{n2}	...	x_{nq}	y_{n1}	y_{n2}	...	y_{np}

Тут $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_q$ − величини ознак у формі значень показників, що описують, наприклад, ресурсно-ринковий стан об’єкта; $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ..., $\boldsymbol{Y}_p$ − величини результативних ознак, що характеризують функціонування об’єкта. Практика свідчить, що кількість елементарних ознак ресурсно-ринкового стану об’єкта значно перевищує кількість результативних ознак його функціонування, тому $p \le q$.

Канонічна кореляція – це кореляція між новими компонентами (канонічними змінними) U і V [191]:

$$\begin{aligned} \boldsymbol{U} &= a_1\boldsymbol{X}_1 + a_2\boldsymbol{X}_2 + ... + a_q\boldsymbol{X}_q; \\ \boldsymbol{V} &= b_1\boldsymbol{Y}_1 + b_2\boldsymbol{Y}_2 + ... + b_p\boldsymbol{Y}_p. \end{aligned} \tag{2.83}$$

За аналогією з парною кореляцією тіснота зв'язку між канонічними змінними буде визначатися "канонічним" коефіцієнтом кореляції r:

$$r = \frac{cov(\boldsymbol{U},\boldsymbol{V})}{\sqrt{\mathrm{var}(\boldsymbol{U})\,\mathrm{var}(\boldsymbol{V})}}. \tag{2.84}$$

Залежно від того, яких значень набувають коефіцієнти a_i і b_j $(i = \overline{1,q},\ j = \overline{1,p})$, будуть змінюватися значення канонічних змінних та канонічного коефіцієнта кореляції. Одна з основних задач, що розв'язується у процесі аналізу канонічних кореляцій, полягає в знаходженні такої пари значень канонічних змінних, якій відповідає максимальне значення канонічного коефіцієнта кореляції.

Для обчислення канонічних коефіцієнтів кореляції необхідно, перш за все, визначити матриці коваріацій початкових змінних:

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{pmatrix}, \tag{2.85}$$

де $\boldsymbol{S}_{11}$ – коваріаційна матриця початкових змінних першої групи $\boldsymbol{X}_1, \ldots \boldsymbol{X}_2, ..., \boldsymbol{X}_q$ розміром $q \times q$;

$\boldsymbol{S}_{22}$ – коваріаційна матриця початкових змінних другої групи $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_p$ розміром $p \times p$;

$\mathbf{S}_{12}, \mathbf{S}_{21}$ – коваріаційні матриці початкових змінних обох груп $\boldsymbol{X}_1,\ \boldsymbol{X}_2, ...,\ \boldsymbol{X}_q$ і $\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_p$; розміри цих матриць дорівнюють відповідно $q \times q$ та $p \times p$.

Матриця $\boldsymbol{S}_{21}$ є результатом транспонування матриці $\boldsymbol{S}_{12}$. У матрицях $\boldsymbol{S}_{11}$ та $\boldsymbol{S}_{22}$ елементи, які розміщені на головній діагоналі, є дисперсіями відповідних змінних.

Елементи матриці $\boldsymbol{S}$ є величиною коваріації кожної пари змінних. Так $s_{x_1 y_1}$ – коваріація змінних x_1 та y_1, що визначається за формулою:

$$s_{x_1 y_1} = \frac{\sum_{i=1}^{n} \left(x_{1i} - \overline{x_1}\right)\left(y_{1i} - \overline{y_1}\right)}{n} = \operatorname{cov}\left(\boldsymbol{X}_1, \boldsymbol{Y}_1\right). \qquad (2.86)$$

Вирази (2.33) зображуються в матричному вигляді так:

$$\begin{cases} \boldsymbol{U} = \boldsymbol{XA} \\ \boldsymbol{V} = \boldsymbol{YB} \end{cases}, \qquad (2.87)$$

де $\boldsymbol{U}, \boldsymbol{V}$ – вектори канонічних змінних;

$\boldsymbol{X}, \boldsymbol{Y}$ – матриці початкових значень змінних;

$\boldsymbol{A}, \boldsymbol{B}$ – вектори коефіцієнтів.

У матричній формі формулу для обчислення канонічних коефіцієнтів кореляції можна записати таким чином:

$$r = \frac{cov\left(\boldsymbol{U}, \boldsymbol{V}\right)}{\sqrt{\operatorname{var}\left(\boldsymbol{U}\right)\operatorname{var}\left(\boldsymbol{V}\right)}} = \frac{cov\left(\boldsymbol{XA}, \boldsymbol{YB}\right)}{\sqrt{\operatorname{var}\left(\boldsymbol{XA}\right)\operatorname{var}\left(\boldsymbol{YB}\right)}} = \frac{\boldsymbol{A}^T \boldsymbol{S}_{12} \boldsymbol{B}}{\sqrt{\boldsymbol{A}^T \boldsymbol{S}_{11} \boldsymbol{A}} \cdot \sqrt{\boldsymbol{B}^T \boldsymbol{S}_{22} \boldsymbol{B}}}. \qquad (2.88)$$

Для спрощення процедури знаходження оптимальних коефіцієнтів канонічних змінних припускають, що кожна з цих змінних має одиничну дисперсію та нульове математичне сподівання. Тоді маємо задачу на умовний екстремум:

потрібно знайти максимум $\boldsymbol{A}^T \boldsymbol{S}_{12} \boldsymbol{B} \to \max$ за умови:

$$\begin{cases} \boldsymbol{A}^T \boldsymbol{S}_{11} \boldsymbol{A} = 1 \\ \boldsymbol{B}^T \boldsymbol{S}_{22} \boldsymbol{B} = 1 \end{cases}. \qquad (2.89)$$

Тоді $\operatorname{var}\left(XA\right) = \boldsymbol{A}^T \boldsymbol{S}_{11} \boldsymbol{A} = 1$ і $\operatorname{var}\left(YB\right) = \boldsymbol{B}^T \boldsymbol{S}_{22} \boldsymbol{B} = 1$.

Отже, вираз (2.88) набуде вигляду: $r = \boldsymbol{A}^T \boldsymbol{S}_{12} \boldsymbol{B}$. Цю задачу розв’язуємо методом множників Лагранжа. Складаємо функцію Лагранжа:

$$F = \boldsymbol{A}^T \boldsymbol{S}_{12} \boldsymbol{B} - \lambda_1 \boldsymbol{A}^T \boldsymbol{S}_{11} \boldsymbol{A} - \lambda_2 \boldsymbol{B}^T \boldsymbol{S}_{22} \boldsymbol{B}. \qquad (2.90)$$

Якщо продиференціювати функцію Лагранжа за компонентами векторів $\boldsymbol{A}$ та $\boldsymbol{B}$ і прирівняти їх до нуля, можна отримати вираз:

$$\boldsymbol{S}_{12}\boldsymbol{B}-\lambda_1\boldsymbol{S}_{11}\boldsymbol{A}=0; \tag{2.91}$$

$$\boldsymbol{S}_{21}\boldsymbol{A}-\lambda_2\boldsymbol{S}_{22}\boldsymbol{B}=0. \tag{2.92}$$

де λ_1,λ_2 – множники Лагранжа.

Теорема. Множники Лагранжа λ_1,λ_2 однакові й дорівнюють r коефіцієнту канонічної кореляції.

Доведення. Покажемо, що $\lambda_1=\lambda_2=r$. Дійсно, помножимо 1-ше матричне рівняння (2.91) на транспонований вектор $\boldsymbol{A}^T$, а 2-ге (2.81) – на транспонований вектор $\boldsymbol{B}^T$:

$$\boldsymbol{A}^T\boldsymbol{S}_{12}\boldsymbol{B}=\lambda_1\boldsymbol{A}^T\boldsymbol{S}_{11}\boldsymbol{A}, \tag{2.93}$$

$$\boldsymbol{B}^T\boldsymbol{S}_{21}\boldsymbol{A}=\lambda_2\boldsymbol{B}^T\boldsymbol{S}_{22}\boldsymbol{B}. \tag{2.94}$$

Оскільки $\boldsymbol{A}^T\boldsymbol{S}_{11}\boldsymbol{A}=\boldsymbol{B}^T\boldsymbol{S}_{22}\boldsymbol{B}=1,$

отримаємо $\boldsymbol{A}^T\boldsymbol{S}_{12}\boldsymbol{B}=\boldsymbol{B}^T\boldsymbol{S}_{21}\boldsymbol{A},$

тобто $r=\boldsymbol{A}^T\boldsymbol{S}_{12}\boldsymbol{B}=\lambda_1$ і $r=\boldsymbol{B}^T\boldsymbol{S}_{21}\boldsymbol{A}=\lambda_2,$

звідки $\lambda_1=\lambda_2(=\lambda)=r$. ■

Тепер помножимо вираз (2.91) на λ, а вираз (2.92) – на обернену матрицю $\boldsymbol{S}_{22}^{-1}$:

$$\boldsymbol{S}_{12}\lambda\boldsymbol{B}=\lambda^2\boldsymbol{S}_{11}\boldsymbol{A}, \tag{2.95}$$

$$\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\boldsymbol{A}=\lambda\boldsymbol{B}. \tag{2.96}$$

Звідси отримуємо:

$$\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\boldsymbol{A}=\lambda^2\boldsymbol{S}_{11}\boldsymbol{A}. \tag{2.97}$$

Якщо помножити обидві частини рівняння (2.94) на $\boldsymbol{S}_{11}^{-1}$, можна отримати вирази:

$$\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\boldsymbol{A} = \lambda^2\boldsymbol{A} \quad \text{чи} \quad \boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21} - \lambda^2\boldsymbol{E} = 0, \tag{2.98}$$

тобто λ^2 є найбільшим власним числом матриці $\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}$, а $\boldsymbol{A}$ − власним вектором цієї матриці.

Аналогічні перетворення приводять ще до одного виразу:

$$\left(\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12} - \lambda^2\boldsymbol{E}\right)\boldsymbol{B} = 0. \tag{2.99}$$

Для того щоб розв’язати ці рівняння, потрібно знайти власні корені та власні вектори. З умови $p \le q$ випливає, що розмірність вектора $\boldsymbol{B}$ менша від розмірності вектора $\boldsymbol{A}$. Отже, простіше знайти власні корені та вектори матриці (2.99).

Для того щоб знайти компоненти вектора $\boldsymbol{A}$, необхідно спочатку визначити вектори $\boldsymbol{B}$ та λ.

Усі значення λ^2 знаходяться як власні значення матриці:

$$\boldsymbol{C} = \boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}. \tag{2.100}$$

Вимірність цієї матриці дорівнює $(p \times p)(p \times q)(q \times q)(q \times p)$. Таким чином, можна знайти p власних значень λ^2 та p відповідних їм власних векторів $\boldsymbol{B}$.
Вектор $\boldsymbol{A}$ визначимо з виразу (2.91):

$$\boldsymbol{A} = \frac{\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}\boldsymbol{B}}{\lambda}. \tag{2.101}$$

Отже, якщо ранжувати власні значення λ^2 таким чином, щоб $\lambda_1^2 \ge \lambda_2^2 \ge ... \ge \lambda_p^2$, то λ_1^2 буде відповідати максимальний канонічний коефіцієнт кореляції. У процесі визначення вектора $\boldsymbol{A}$ та канонічних коефіцієнтів кореляції знак при λ вибирається, виходячи з економічного змісту множин X та Y, що аналізуються.

Канонічні коефіцієнти кореляції можна обчислювати на основі вибіркової кореляційної матриці $\boldsymbol{R}$:

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \tag{2.102}$$

При цьому всі викладки будуть аналогічні тим, що наведені для коваріаційної матриці [191]. Коефіцієнти a_i та b_j $\left(i = \overline{1, q},\ j = \overline{1, p}\right)$ для канонічних змінних U та V тепер будуть безрозмірними, оскільки вони відносяться до стандартизованих значень початкових змінних x_i' та y_j':

$$x_i' = \frac{x_i - \overline{x}}{\sigma_{x_i}}, \quad y_j' = \frac{y_j - \overline{y}}{\sigma_{y_j}}. \tag{2.103}$$

Донедавна складність описаної обчислювальної процедури канонічного аналізу була поясненням її обмеженого застосування в аналізі ознак об’єктів в економіці, але сучасний розвиток програмного та технічного забезпечення розкриває широкі можливості даного математичного методу, залишається лише питання формування системи ознак, що описують об’єкт, та виявлення механізму їх причино-наслідкової взаємодії.

Для перевірки значущості коефіцієнтів канонічної кореляції доцільно використовувати критерій Бартлета. Якщо позначити як W_0 добуток p множників $(1-\lambda_i^2)$, то можна перевірити нульову гіпотезу за допомогою χ^2-критерію:

$$\chi^2 = -\left[n - 1 - 0{,}5\left(p + q + 1\right)\right]\ln W_0 \tag{2.104}$$

з числом ступенів свободи, що дорівнює $p \times q$. Якщо обчислене χ^2 більше від табличного значення критерію за вибраним рівнем значущості, то можна стверджувати, що принаймні $r_1 = \sqrt{\lambda_1^2}$ буде істотно відрізнятися від нуля. Щоб перевірити значущість другого коефіцієнта канонічної кореляції, необхідно обчислити величину $W_1 = \prod_{i=2}^{p}\left(1 - \lambda_i^2\right)$ і визначити χ^2 за формулою:

$$\chi^2 = -\left[n - 2 - 0{,}5\left(p + q + 1\right)\right]\ln W_1 \tag{2.105}$$

для $\left(p - 1\right)\left(q - 1\right)$ ступенів свободи.

Запитання для самоперевірки

1. На які групи ділиться багатовимірний статистичний аналіз?
2. Які основні проблеми вирішуються за допомогою методів багатовимірного статистичного аналізу?
3. Які відмінності методів багатовимірного статистичного аналізу від методів класичної статистики?
4. Який узагальнений зміст методів багатовимірного статистичного аналізу?
5. Сформулювати основні аналітичні можливості множинного регресійного аналізу, факторного аналізу, кластерного аналізу, дискримінантного аналізу, методу канонічних кореляцій.
6. Назвати перелік практичних задач, які можна вирішувати за допомогою множинного регресійного аналізу, факторного аналізу, кластерного аналізу, дискримінантного аналізу, методу канонічних кореляцій.
7. Дати визначення факторного аналізу.
8. Які переваги переходу від простору елементарних ознак до простору факторів?
9. Що стверджує фундаментальна факторна теорема?
10. Як інтерпретувати виявлені фактори?
11. Назвати основні математичні проблеми факторного аналізу.
12. Які основні види кластерного аналізу?
13. Які основні математичні проблеми вирішуються в кластерному аналізі?
14. Які міри схожості рекомендуються для багатовимірної класифікації?
15. Які існують рекомендації у вирішенні проблем стійкості кластеризації?
16. Назвати відмінності дискримінантного аналізу від інших методів багатовимірної класифікації.
17. Сформулювати основні математичні проблеми дискримінантного аналізу.
18. В яких випадках складно провести правильну класифікацію нових об’єктів?
19. Чим відмінний метод канонічних кореляцій від інших методів багатовимірного статистичного аналізу?
20. Визначити основну концепцію канонічного аналізу.
21. Які основні математичні проблеми вирішуються в канонічному аналізі?

Розділ 3
Багатовимірний аналіз метричних ознак соціально-економічних систем

3.1. Загальні рекомендації щодо початку роботи з меню пакета Statgraphics Pius for Windows

Успішне вирішення наведених у розділі 2 математичних проблем багатовимірного статистичного аналізу обумовлюється наявністю відповідного програмного забезпечення. Задовільні результати розв'язання багатьох практичних економічних задач підприємств спонукають рекомендувати для обчислень окремих блоків у багатовимірному аналізі соціально-економічних систем спеціальний математичний пакет Statgraphics *Plus* V5.1 International Professional. Даний програмний продукт виграшно відрізняється від інших пакетів зручностями інтерфейсу та своїм складом процедур обробки даних, вдалим сполученням наукових методів обробки даних із сучасною інтерактивною графікою. Слід відмітити, що початкові версії надають змогу гнучко реалізувати вирішення обчислювальних проблем математичних методів, оскільки дозволяють інтерактивно змінювати логіку обчислення математичних методів, вносити свої корективи.

Починаючи з 1994 року корпорація Manugistics, тобто коли вийшла перша версія універсальної статистичної графічної системи Statgraphics, періодично удосконалює даний програмний продукт. Але залишається ділення системи на базові процедури та доповнення до неї. Базову систему складають процедури Опис даних (Describe), Порівняння даних (Compare) та Відношення даних (Relate).

До процедури Спеціальні (Special) входять Контроль якості (Quality Control), Планування експерименту (Experimental Design), Аналіз часових рядів (Time-Series Analysis), Багатовимірні методи (Multivariate Methods), Розширений регресійний аналіз (Advanced Regression).

До переваг пакета Statgraphics Pius for Windows відносяться гнучкий імпорт / експорт даних, широкі можливості маніпуляції даних, інтегрована графіка, створення власного статистичного пакета за допомогою процедури StatFolio, повномасштабна статистична консультація за допомогою StatAdvisor можливості комбінування пакета і графіків для складання статистичних звітів на основі інструмента StatGallery.

У цілому статистичний пакет надає широкі можливості проведення глибокого, наочного аналізу даних соціально-економічних систем, які описуються різними ознаками, виміряними на метричних і неметричних шкалах.

Але залишається ділення системи на базові процедури та доповнення до неї. Як уже було зазначено, базову систему складають процедури Describe (Опис даних), Compare (Порівняння даних), Relate (Відношення даних). На рис. 3.1. наведений вигляд екрана на початку роботи.

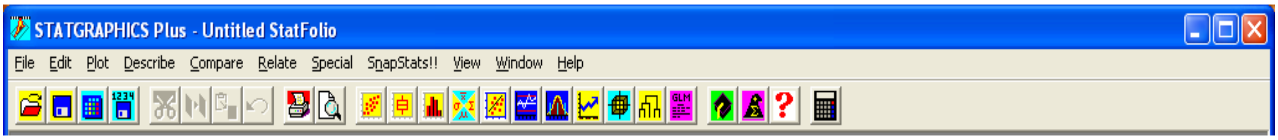


Рис. 3.1. **Панель меню**

У меню File (Файл) розкривають вікна відкриття файлів, збереження, вихід на друк (рис. 3.2).

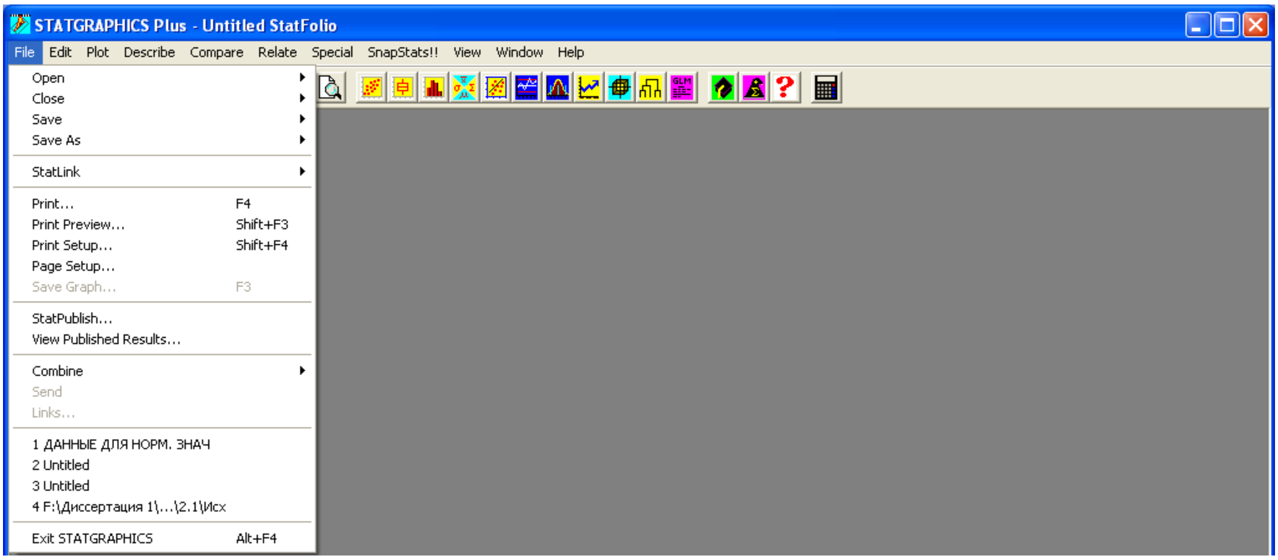


Рис. 3.2. **Меню File**

У меню Plot виконується побудова всіх графіків на замовлення (рис. 3.3).

Рис. 3.3. **Меню Plot**

Процедурою Describe (рис. 3.4) передбачається статистичний аналіз однієї та декількох змінних, підбір розподілів, табуляція та кростабуляція.

Рис. 3.4. **Меню Describe**

У свою чергу, в аналізі однієї змінної (One-Variable Analysis) можна обчислити сумарні статистичні (середні, медіану, середнє геометричне, дисперсію, стандартне відхилення, стандартну похибку, мінімум, максимум, розмах, нижній квартиль, міжквартальний розмах, коефіцієнт асиметрії, нормальний коефіцієнт асиметрії, коефіцієнт ексцесу, нормальний коефіцієнт ексцесу), процентилі, таблиці частот графіків "дерево з листками", довірчі інтервали, перевірку гіпотез (про середню і медіану, Т-тест, знаковий тест, знаковий ранговий тест), діаграму розсіювання, графік "ящик з вусами", гістограму, квантильний графік, графік нормального розподілу, графік щільності, симетричний графік. У меню аналізу багатьох змінних (Multiple-Variable Analysis) обчислюються сумарні статистики, довірчі інтервали, кореляції (рангові Спірмена, окремі кореляції), коваріації, діаграми розсіювання, графік "зірка", графік "сонячні промені". У меню Distributions (Підбір розподілу) обчислюються вбудовані розподіли (експоненціальний, екстремальних значень, лог-нормальний, нормальний, Вейбула), перевірка на нормальність (скоректований χ^2, тест Шапіро – Уілкса, тести для малих вибірок), тести узгодженості (χ^2, Колмогорова – Смірнова), площі залишків, критичні значення, щільності, асиметричні графіки, графік нормального розподілу, графік розподілу Вейбула, частотна гістограма, функції розподілу (щільність, кумулята, функція виживання, логарифм функції виживання, функція ризику). У меню Табуляція (Tabulation) передбачаються табуляція частот (відношення і кумуляти), прямокутні діаграми, кругові діаграми. У меню Кростабуляція (Crosstabulation) виконується табуляція частот, критерій χ^2, вимірювання зв’язку (лямбда, коефіцієнти невизначеності, R-Пірсона, Д. Сомера, Е-та, коефіцієнт контенгенції, V-Крамера, умовний гама, τ-Кандела), прямокутні діаграми, мозаїчні відображення (горизонтальні і вертикальні), трьохвимірні діаграми (частот або процентів).

Для порівняння даних використовується меню Compare (рис. 3.5), яке надає можливості порівняти дві вибірки (Two Samples), декілька вибірок (Multiple Samples), дисперсійний аналіз (Analysis of Variance).

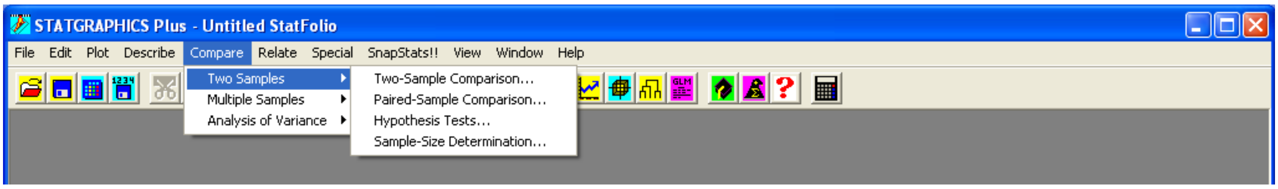


Рис. 3.5. **Меню Compare**

Вікно Two Samples надає змогу активізувати сумарні статистики, порівняння середніх: T-тест, довірчі інтервали; порівняння стандартних відхилень: відношення дисперсій, F-тест, довірчі інтервали; порівняння медіан: тест Манна – Уітнея (Вілкоксона); тест Колмогорова – Смірнова; гістограми частот; щільності розподілів; порівняльні графіки "ящика з вусами"; графіки квантилів; графіки "Квантиль – Квантиль".

За допомогою вікна Multiple Samples можна порівнювати декілька вибірок на основі сумарних статистик; таблиці дисперсійного аналізу: суми квадратів, середнього квадрату, F-відношення; таблиці і графіка середніх: стандартної похибки, довірчих інтервалів, найменших значущих відмінностей (LSD), Т'юкки HSD, Шеффе, Бонфероні; множинних рангових тестів: LSD, Т'юкки HSD, Шеффе, Бонфероні, Н'юмена – Кеулса, Дункана; відповідність дисперсій: тест Кокрена, тест Бартлета, тест Хартлея; тест Краскала – Уоліса; діаграми розсіювання; для порівняння графіки "ящик з вусами"; залишки для вибірок; залишки для прогнозів; залишки для спостереження.

За допомогою активізації вікна Analysis of Variance можна обчислити: сумарні статистики; таблиці дисперсійного аналізу; таблиці і графіки середніх; множинні рангові тести; аналіз дисперсії; тест Краскала – Уоліса; діаграму розсіювання; графік "ящик з вусами"; залишки і рівні фактора; залишки й описи; залишки і номер рядка. Для багатофакторного дисперсійного аналізу надається змога обчислити: таблиці дисперсійного аналізу (суму квадратів, тип I; суму квадратів, тип III); таблицю середніх; множинні рангові тести; діаграму розсіювання; графіки середніх; графіки взаємодій; залишки і рівні факторів; залишки й опис; залишки і номер рядків.

Для обчислення простої і множинної регресій слід активізувати меню Relate (Відношення даних) (рис. 3.6).

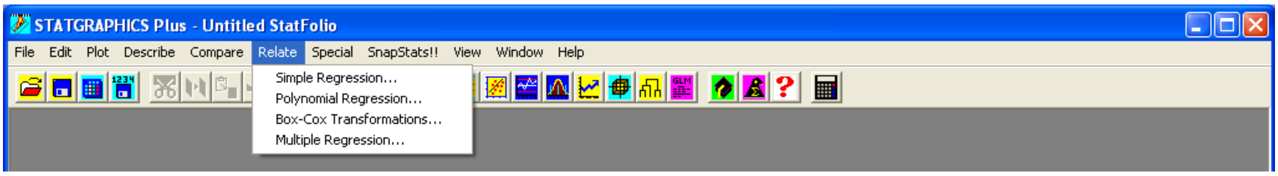


Рис. 3.6. **Меню Relate**

Активізація вікна Simple Regression дозволяє побудувати моделі простої регресії: лінійну, експоненціальну, зворотню Y, двічі зворотню, логарифм X, мультиплікативну, квадратний корінь X, квадратний корінь Y, S-криву, логістичну, логарифм імовірності, а також обчислити t-статистики, аналіз дисперсії: коефіцієнт кореляції, R-квадрат, стандартну похибку оцінки; прогнози; порівняння альтернативних моделей; незвичайні залишки; точки впливу; графік підібраної моделі: опис і довірчі інтервали, спостереження й опис, залишки й X: залишки, стьюдентизовані залишки; залишки й опис; залишки і номер рядка.

Активізація вікна Multiple Regression дозволяє побудувати модель множинної регресії, при цьому обчислити коефіцієнти моделі; t-статистики; аналіз дисперсії: R-квадрат, стандартну похибку, середню абсолютну похибку, статистику Дарбіна – Ватсона; умовну суму квадратів: суму квадратів, середній квадрат, F-відношення; довірчі інтервали; кореляційну матрицю; звіти: Y, що спостерігається, Y, що підганяється, залишки, стьюдентизовані залишки, стандартні похибки і прогнози, довірчі межі; незвичайні залишки; точки впливу; компонентні ефекти; спостереження і опис; залишки і X; залишки й опис; залишки і номер рядка; інтервальні графіки: величини, що описуються, середні, прогнози, прогнози середніх.

До меню Special (Спеціальні), як уже було сказано, входять вікна Quality Control (Контроль Якості), Experimental Design (Планування експерименту), Time-Series Analysis (Аналіз часових рядів), Multivariate Methods (Багатовимірні методи), Advanced Regression (Розширений регресійний аналіз) (рис. 3.7).

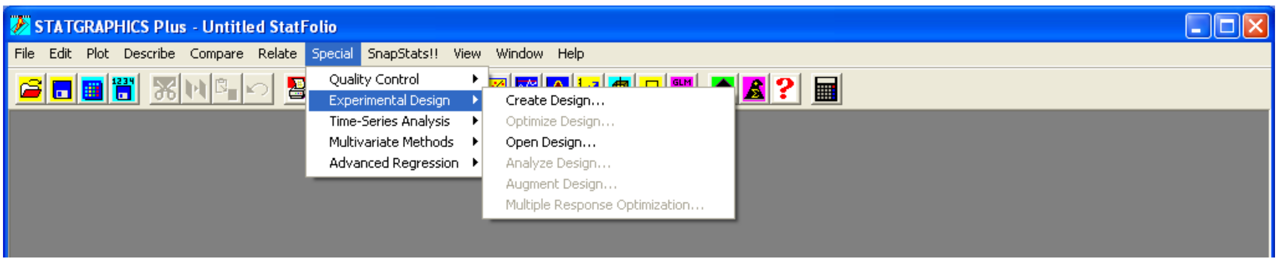


Рис. 3.7. **Меню Special**

У меню ShapStats!! надається змога активізувати набір графічно-документального супроводу, який найчастіше використовується у вирішенні задач за допомогою Statgraphics (рис. 3.8).

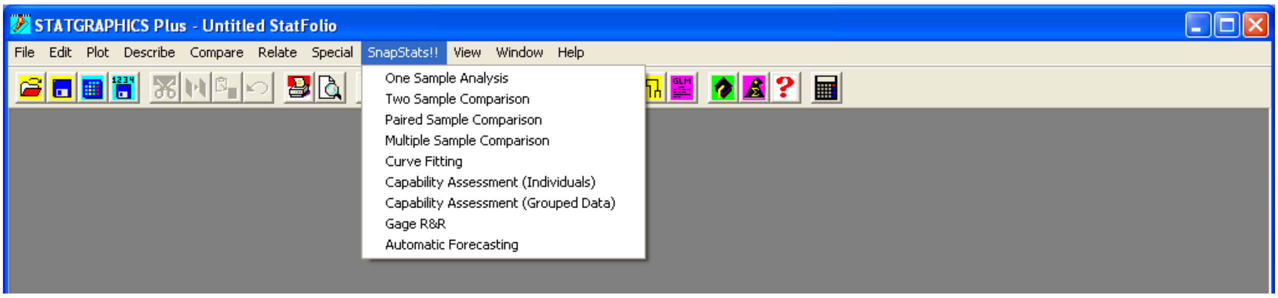


Рис. 3.8. **Меню ShapStats!!**

Управління оформленням робочого столу Statgraphics здійснюється за допомогою меню View (рис. 3.9).

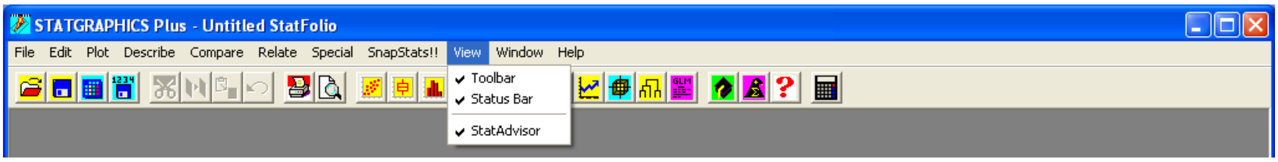


Рис. 3.9. **Меню View**

Активізуючи вікна меню Window, користувач пакета може здійснювати розподіл вікон, змінювати їх положення (рис. 3.10).

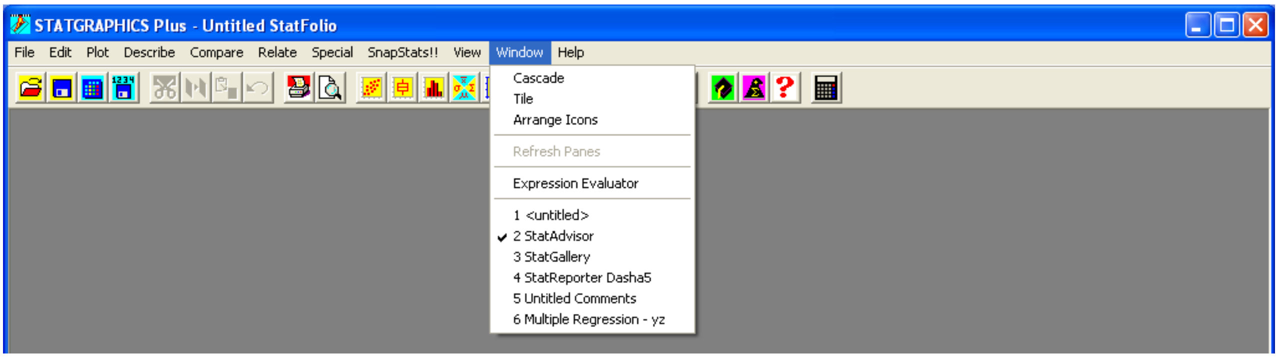


Рис. 3.10. **Меню Window**

Різного роду підказки маємо в меню Help, які демонструють широкі можливості пакета Statgraphics (рис. 3.11)

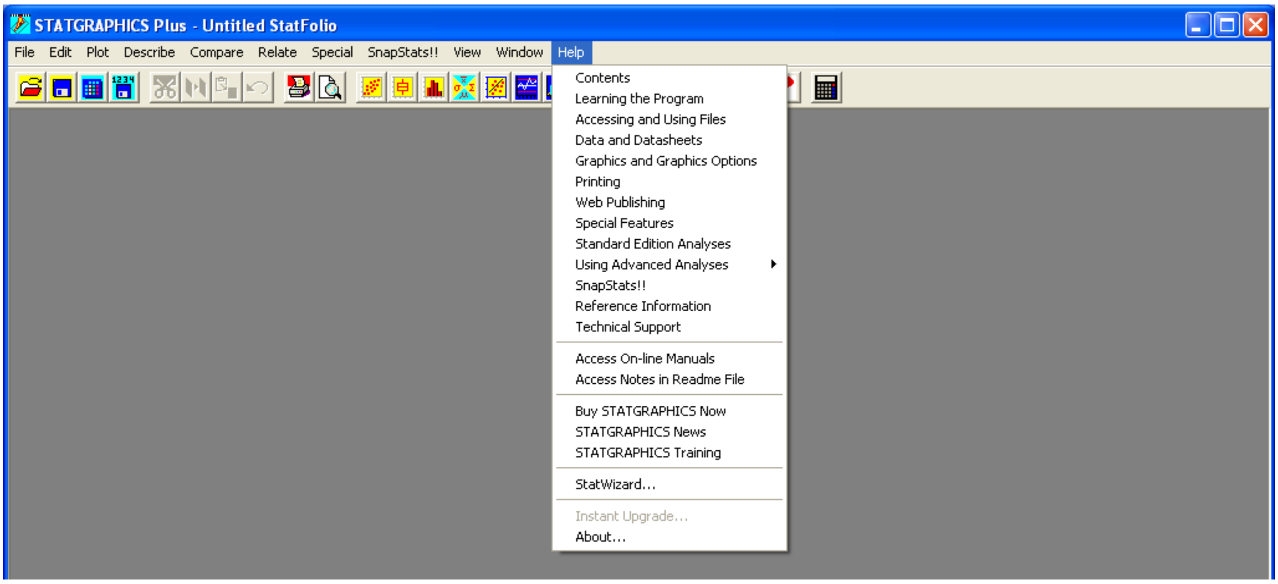
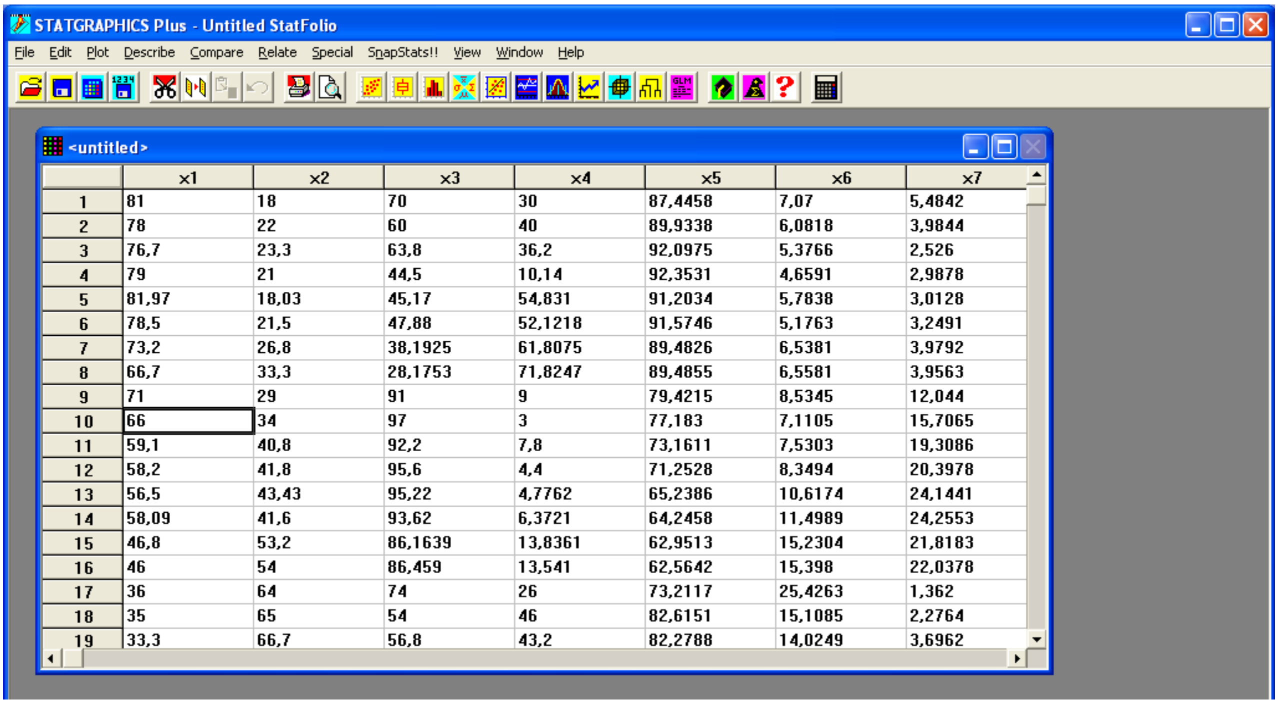


Рис. 3.11. **Меню Help**

Введення даних здійснюється за допомогою ініціалізації електронної таблиці, що відповідає вікну Untitled, як це показано на рис. 3.12.



	x1	x2	x3	x4	x5	x6	x7
1	81	18	70	30	87,4458	7,07	5,4842
2	78	22	60	40	89,9338	6,0818	3,9844
3	76,7	23,3	63,8	36,2	92,0975	5,3766	2,526
4	79	21	44,5	10,14	92,3531	4,6591	2,9878
5	81,97	18,03	45,17	54,831	91,2034	5,7838	3,0128
6	78,5	21,5	47,88	52,1218	91,5746	5,1763	3,2491
7	73,2	26,8	38,1925	61,8075	89,4826	6,5381	3,9792
8	66,7	33,3	28,1753	71,8247	89,4855	6,5581	3,9563
9	71	29	91	9	79,4215	8,5345	12,044
10	66	34	97	3	77,183	7,1105	15,7065
11	59,1	40,8	92,2	7,8	73,1611	7,5303	19,3086
12	58,2	41,8	95,6	4,4	71,2528	8,3494	20,3978
13	56,5	43,43	95,22	4,7762	65,2386	10,6174	24,1441
14	58,09	41,6	93,62	6,3721	64,2458	11,4989	24,2553
15	46,8	53,2	86,1639	13,8361	62,9513	15,2304	21,8183
16	46	54	86,459	13,541	62,5642	15,398	22,0378
17	36	64	74	26	73,2117	25,4263	1,362
18	35	65	54	46	82,6151	15,1085	2,2764
19	33,3	66,7	56,8	43,2	82,2788	14,0249	3,6962

Рис. 3.12. **Приклад вигляду вікна з даними**

3.2. Описова статистика величин елементарних ознак соціально-економічних систем

Необхідним початковим етапом вивчення природи функціонування та розвитку об'єкта в економіці є аналіз спостережень чи даних його елементарних ознак, що передбачає аналіз закономірної та випадкової мінливості величини ознаки, сформованої в показник, виявлення закономірної мінливості на фоні випадкової. Сучасні методи багатовимірного статистичного аналізу та їх комп'ютерна реалізація у вигляді спеціальних пакетів дозволяють комплексно провести дослідження, оцінити параметри наявних закономірностей, перевірити різні гіпотези про ці закономірності.

Для того щоб об'єктивно дослідити величину елементарної ознаки об'єкта, рівні її значень, потрібно проаналізувати сукупність значень у динаміці, по-перше, на даному об'єкті, по-друге, в однорідній групі об'єктів, по-третє, у всій сукупності об'єктів, якщо це можливо і доцільно. Існує переконання, що дане дослідження величини ознаки можливе, якщо використати засоби описової статистики [120; 122; 186]. Корисність методів описової статистики ґрунтується на тому факті, що декілька простих та досить інформативних статистичних показників і візуальних засобів виключають монотонний перегляд усіх значень величини ознаки.

До інструменту описової статистики входять:

1. Показники положення чи стану описують положення значень величини ознаки на числовій осі, а саме: вибіркове середнє, вибіркова медіана, мода, найбільше та найменше значення величини ознаки, перцентилі, зокрема, квартилі.

2. Показники розкиду описують ступінь розкиду значень величини відносно свого центру та характеризують ступінь їх мінливості. До цих показників відносяться дисперсія сукупності значень (вибіркової чи генеральної), стандартне відхилення, розмах сукупності, міжквартильний розмах. Ці показники характеризують купчастість значень величини відносно свого центру.

3. Показники форми розподілу значень величини ознаки – коефіцієнти асиметрії та ексцесу. Відомо, що до показників асиметрії відносяться такі: коефіцієнт асиметрії, стандартизований коефіцієнт асиметрії, положення вибіркової медіани відносно вибіркового середнього, положення вибіркових квартилів і т. д. Важливість коефіцієнта асиметрії обумовлена тим, що він описує симетрію розподілу значень величини відносно свого центру. Плосковершинність (гольчатість) розподілу значень величини ознаки відображає коефіцієнт ексцесу. Якщо значення стандартизованого коефіцієнта ексцесу менше двох, то асиметрії немає.

4. Графічні засоби, що описують закон розподілу значень величини ознаки. Наведені засоби дають уявлення про закон розподілу значень. До їх складу відносяться гістограми, кумуляти, блокові діаграми, таблиці частот.

Слід вважати, що комплекс перелічених засобів описової статистики є основним, достатнім для комплексного дослідження різних величин ознак об’єкта в економіці. Звичайно, даний комплекс можна розширяти, але одержання принципової інформації про величину здійснюється завдяки наведеному комплексу засобів. На практиці традиційно використовується вибіркове середнє, медіана, дисперсія і стандартне відхилення. Проте для одержання точніших та достовірніших висновків настійно рекомендується уважно вивчати й інші характеристики, а також звертати увагу на умови одержання вибіркових сукупностей [22; 190; 219].

Особливу увагу слід також звертати на наявність у сукупності значень викидів – грубих (помилкових), надто відмінних від основового масиву значень. Саме вони спотворюють вибіркові характеристики, зокрема, середнє, дисперсію, стандартне відхилення, коефіцієнти асиметрії та ексцесу. Викиди виявляють або за допомогою правила 3-х сигм, або за допомогою гістограми з достатньо великою кількістю інтервалів групування, або за допомогою детальної блокової діаграми. Присутність викидів також призводить до багатьох помилкових висновків, тому потрібно ретельно вивчати причину викидів. Традиційно статистичні характеристики чуттєві до відхилень від умов застосування методу. Інтенсивно розвиваються сучасні статистичні методи, які стійкі до викидів, але вони ще не одержали широкого розповсюдження на практиці, за винятком рангових процедур для найстандартніших задач. Причина тут полягає в обчислювальній складності цих методів, що утруднює їх застосування, у відсутності спеціальних комп’ютерних програм.

Описова статистика формує перше уявлення про теоретичні та вибіркові характеристики випадкових величин елементарних ознак, дозволяє досліджувати тенденції змін значень величини ознаки з урахуванням реальних умов функціонування об’єкта. У складних ситуаціях вона становить ефективні засоби "побачити всю картину". Детальне вивчення кожного окремого випадку само по собі не є завданням статистики, але вимірювання величин ознак у розв’язуванні практичних задач управління об’єктом в економіці потребує виявлення та ідентифікації особливостей, які в цілому характерні для випадків, що розглядаються, та складають цілі описової статистики ознак. Досягнення мети статистики – описи інформації, що містять великі сукупності значень величин невеликим числом показників, які виражають найфундаментальніші властивості сукупностей, можливо здійснити за допомогою раціонально побудованого комплексу засобів описової статистики чи, як їх ще іноді називають – методів статистичного аналізу. Отже, потрібно детально проаналізувати кожен засіб чи інструмент описової статистики, щоб відмітити важливість його включення до комплексу та створення єдиного інструменту дослідження значень величини ознаки об'єкта.

Характеристику положення значень величини надають середнє, медіана, мода – це різні способи вибору єдиного числа, яке найкраще описує всі значення величини в сукупності. Такий представленний одним числом показник називають типовим значенням, або центром, мірою центральної тенденції зміни значень величини. Визначення демонструє бузсумнівну важливість даного показника.

Середнє найкорисніше в якості узагальнюючого показника положення при відсутності викидів, коли набір значень подається порівняно однорідною групою з елементами випадковості.

Середньозважене найкраще інтерпретувати як середнє, що використовується в ситуаціях, коли розставляється приорітетність серед рівнів значень. Найважливіші рівні значень величини ознаки можуть вносити більший внесок в обчислення значення середнього зваженого. Зважені середні використовують для того, щоб скоректувати недоліки репрезентативності вибірки відносно генеральної сукупності.

Медіана розміщується в центрі сукупності значень величини і дає уявлення про всю послідовність значень. Щоб знайти медіану, значення розміщують у зростаючому порядку, а потім визначають центральне значення. У половини всіх спостережень значення величини ознаки більше медіани, у половини – менше. Якщо в сукупності даних немає одного центрального значення, то усереднюють ті два значення, які знаходяться посередені послідовності значень величини. Медіану можна визначити в термінах рангу (коли це необхідно для метричної величини, і як єдиний спосіб – для неметричної величини). Ранги позв'язують числа $1, 2, \ldots\ n$ із значеннями величини таким чином, щоб найменше значення мало ранг 1, наступне значення величини – ранг 2 і т. д. до найбільшого значення, що має ранг n. Якщо існує сукупність значень метричної величини, то медіана є середньою значень. Якщо сукупність значень величини виміряна за допомогою порядкових шкал, тобто значення виражаються ординальними числами, то значення, розміщене в середині ряду значень, є медіаною кількісних значень чи предикатних значень величини ознаки. Медіана – типове, усталене до наявності викидів значення метричної та неметричної величин ознаки.

Якщо значення величини в сукупності розподілені симетрично, то значення медіани і середнього близькі між собою, оскільки симетричні розподіли мають чітко виражену середню точку. Але навіть при "практично нормальному" розподілі середнє та медіана можуть дещо відрізнятися, оскільки кожна з цих характеристик визначається по-своєму, і, крім того, у реальних значеннях величини ознаки майже завжди присутня випадковість. Наявність викидів різко відображується на відмінності між середнім і медіаною. Якщо розподіл значень величини асиметричний, то медіана й середнє дуже відрізняються. Зазвичай середнє відносно медіани здвинуте в напрямку довшого хвоста чи в напрямку викидів, оскільки середнє реально враховує значення таких екстремальних спостережень, у той час як для медіани важливо лише те, з якої сторони від неї лежить те чи інше значення величини ознаки.

Мода визначається найпоширенішим значенням величини ознаки, тобто значенням, що найчастіше зустрічається в сукупності. Це єдина характеристика, яку можна визначити для неметричної величини, що вимірюється за домогою шкал назв чи номінацій, оскільки неупорядковані категорії якісних описових ознак об'єкта не можуть бути просумовані, як це потрібно для середнього, і не можуть бути проранжовані, як це потрібно для медіани. Моду знаходять і для величин, що шкальовані порядково, якщо ігнорувати впорядкованість категорій та виконати всі дії так, як для сукупності величин номінальних ознак з неупорядкованими категоріями. Мода визначається для метричних величин за найвищою точкою на гістограмі, хоча при цьому має місце деяка невизначеність, наприклад, на гістограмі можуть бути два високих стовпці або визначення можуть залежити від того, яким чином побудована діаграма: зміна ширини стовпців приводить до помірних змін форми розподілу значень, у результаті чого мода може змінитися. Мода для сукупності значень метричних величин є дуже неусталеною характеристикою. Вона зосереджує увагу лише на найважливішому значенні величини ознаки. При аналізі неметричних величин ознак акцентуємо увагу на модальній категорії ознаки.

Отже, рекомендації щодо вибору характеристик сукупності значень величин ознак об'єкта залежно від типу ознак можна сформулювати так. У випадку ознак, що мають метричну величину, яка дозволяє обчислювати всі три характеристики, якщо розподіл значень величини близький до нормального, то різниця між трьома характеристиками буде невелика. У випадку асиметричного розподілу значень величини ці три характеристики можуть суттєво відрізнятися. Середнє слід використовувати, коли сукупність значень величини має розподіл нормальний чи близький до нормального, оскільки в цьому випадку середнє є найпростішою характеристикою. Медіану необхідно використовувати при асиметричному розподілі через те, що на нього не впливає невелика кількість значень високого рівня. У випадку великої асиметрії розподілу медіана значно краще середнього характеризує більшість значень величини. Застосування медіани надзвичайно доцільне наразі в сукупності значень величини ознаки викидів, оскільки вона усталена до їх впливів. Медіану корисно обчислювати для сукупності значень неметричних величин номінальних ознак, а також ознак, виміряних за порядковими шкалами, тому що важливо знати найпоширенішу категорію.

Перцентилі узагальнюють інформацію про сукупність значень величин ознак, характеризуючи значення, що досягається заданим процентом загальної кількості спостережень, після того як значення проранжуються за зростанням. Перцентилі – це характеристики сукупності значень величини ознак, які виражають ранги значень у вигляді процентів від 0 до 100%, а не у вигляді чисел від 1 до n, таким чином, найменшому значенню відповідає нульовий перцентиль, найбільшому – 100-й перцентиль, медіані – 50% перцентиль і т. д. Перцентилі іноді розглядають як показники, що характеризують розділення сукупності значень різних величин на окремі частини. Перцентилі використовують у двох цілях: щоб показати рівень величини ознаки при заданому перцентильному ранзі та перцентильний ранг даного значення в сукупності. Перцентилі відіграють важливу роль опорних характеристик. Для того щоб узагальнити основні риси розподілу значень величини, достатньо мати декілька значень перцентилів. Так, 50-й перцентиль – це медіана, оскільки він знаходиться посередині між найбільшим і найменшим значеннями величини ознаки. Становлять інтерес і екстремуми величини ознаки – найбільші та найменші значення. Доповнюють набір базових характеристик сукупності значень величини квартилей, що визначаються як 25 і 75-й перцентилі.

Існує декілька методів визначення вибіркових значень квартилів, різних за алгоритмом обчислення [189; 190; 219]. Квартилі – це значення ранжованого ряду значень величини, що знаходяться на відстані однієї чверті в послідовності від найменшого до найбільшого значення. Очевидний факт, що дане визначення не вказує точно на те, як обчислити квартилі. Джон Тьюки, один із засновників практичного аналізу даних, визначає квартилі таким чином: обчислює ранг медіани за формулою $\frac{1+n}{2}$ і відкидає дробову частину; прибавляє до одержаного значення 1 та ділить на 2 [189]. Отримане значення становить ранг нижнього квартиля; віднімає отримане значення від $(n+1)$, а результатом буде ранг верхнього квартиля. Значення квартилів обчислюється, виходячи з цих рангів. Таким чином, наведена схема обчислення квартилів реалізується формулами:

$$\text{ранг нижнього квартиля} = \frac{1 + int\left[\frac{1+n}{2}\right]}{2};$$

$$\text{ранг верхнього квартиля} = n + 1 - \text{ранг нижнього квартиля},$$

де int – функція цілого значення величини.

У статистичному аналізі відомий потужний засіб проведення порівняльного аналізу за різними вибірками однієї і тієї ж ознаки – це блокова діаграма Тьюки, або, як її ще називають, "ящик з вусами". Блокова діаграма зображує значення п’яти базових статистичних показників, що характеризують положення значень величини ознаки. Вона містить п’ять базових показників множини значень величини однієї ознаки, що дозволяє швидко визначити характер розподілу. Блокова діаграма, як і гістограма, візуально демонструє розподіл, але при цьому є іншим способом графічного відображення. Блокова діаграма не обтяжена деталями і тому охоплює статистичний опис множини значень величини ознаки в цілому та дозволяє порівнювати декілька груп ознаки, що мають різні значення величини, не вдаючись до кількісного опису кожної групи. Слід зазначити, що аналіз форми розподілу краще виконувати на гістограмі. Якщо відсутні викиди, то до п’яти базових показників входять: найменше значення величини (0-й перцентиль), нижній квартиль (25-й перцентиль, на чверть відстані від найменшого значення), медіана (50-й перцентиль, середина), верхній квартиль (75-й перцентиль, на три чверті відстані від найменшого значення або на чверть відстані від найбільшого значення), найбільше значення (100-й перцентиль). Якщо ж множина значень величини ознаки має викиди, то вони зображуються світлими точками, тоді межі вусів встановлюють на 1,5 міжквартильного розмаху. Разом ці перераховані статистичні характеристики дають достатньо чітке уявлення про особливості ще не обробленої множини значень величини ознаки. Два значення – найбільше і найменше – характеризують розмах (діапазон) зміни значень величини, медіана показує центр, два квартилі визначають межі "розміщеної в центрі половини множини значень величини", а положення медіани відносно квартилей і середнього дає грубе уявлення про наявність чи відсутність асиметрії. Велике значення величини розглядається як викид, якщо воно перевищує верхній квартиль плюс 1,5 міжквартильного розмаху. Мале значення величини – як викид, якщо воно менше, ніж нижній квартиль мінус 1,5 міжквартильного розмаху. Так визначаються викиди за Д. Тьюки [189].

Другий тип інструментів описової статистики направлений на вивчення різноманіття значень величини ознаки, тобто на мінливість значень величини. Очевидний факт, що коли б значення величини ознаки не змінювались, то не було б сенсу звертатися до методів статистичного аналізу. Ситуація, в якій присутня мінливість, часто містить ризик, оскільки навіть використання всієї доступної інформації не дозволяє точно передбачити, що відбудеться в майбутньому. Для об'єктивного аналізу ризику потрібно розуміти його природу і вміти виміряти мінливість, що є його наслідком.

Мінливість – міра відмінності окремих значень величини ознаки. У той час, коли статистичні характеристики центру – середнє значення, медіана, мода – вказують на типове для величини значення, мінливість показує, наскільки близько до цього центру зазвичай розміщені окремі значення в множині. Якщо всі значення величини однакові, то мінливість дорівнює нулю. Чим більше розкид значень величини, тим більша її мінливість та мінливість ознаки об'єкта.

Мінливість визначається як ступінь відмінностей між окремими значеннями величини ознаки. Подібний сенс мають також такі поняття, як різноманіття, невизначеність, розсіяння та розкид. Існують чотири основні характеристики опису степеня мінливості множини значень величини:

1. Розмах є простим в обчисленні та надає дещо поверхневе уявлення про мінливість значень величини і має обмежене використання. Ця характеристика описує межі змін значень величини в множині та є відстанню між найменшим і найбільшим значеннями. Великі відхилення рідкісні. Тому у вибірку значень ознаки найменше й найбільше значення можуть не потрапити, звідси розмах буде обчислений за іншими числами та буде заниженим, а отже розмах – неусталена та малонадійна статистична характеристика.

2. Стандартне відхилення використовують найчастіше. Цей статистичний показник описує, наскільки сильно значення величини ознаки, що спостерігається, зазвичай відрізняється від середнього значення. При піднесенні стандартного відхилення до квадрату отримують дисперсію. Стандартне відхилення відображує типову відстань між середнім значенням і окремими значеннями величини в множині, а також показує степінь випадковості в розміщенні окремих значень відносно їх загального середнього.

Відхилення – це відстань між кожним значенням і середнім у множині значень величини ознаки. Стандартне відхилення є деяким середнім з цих відхилень – "середнім квадратичним відхиленням", оскільки звичайне середнє арифметичне завжди тут дорівнює нулю через різні знаки окремих відхилень. Стандартне відхилення є надзвичайно важливою характеристикою в статистиці, тому що воно становить основний інструмент визначення степеня випадковості в дослідженні величини ознаки. Порівняння середньоквадратичних відхилень значень величини однієї й тієї ж ознаки в різних розподілах величини дозволяє встановити, де варіація ознаки більша. Воно є також показником однорідності даної сукупності значень величини ознаки. Невелике стандартне відхилення говорить про те, що середнє добре відображує в собі сукупність значень величини, що подається. Тому кожне обчислення середнього в розподілі значень повинне супроводжуватися обчисленням стандартного відхилення. Коли сукупність значень величини ознаки має приблизно нормальний розподіл, стандартне відхилення набуває особливого сенсу. Імовірність того, що випадкові відхилення значень величини від свого середнього будуть менші, ніж 2σ, дорівнює 0,954 і є правилом "двох сигм" при нормальному розподілі.

При будь-яких інших розподілах використовується менше визначене відхилення випадкової величини від свого середнього, що за абсолютною величиною практично не перевищує потроєного середнього квадратичного відхилення. На цьому правилі ґрунтується така наближена оцінка середнього квадратичного відхилення: із сукупності значень величини ознаки вибирають максимальне і мінімальне значення та їх різницю ділять на 6. Одержане число приймають як грубу оцінку середнього квадратичного відхилення. Слід зазначити, що існує безліч скошених (чи інших, що відрізняються від нормального) розподілів, для яких неможливо вказати рівень довіри.

У таких випадках використовують співвідношення, яке називається нерівністю Чебишева. Відповідно до цього правила принаймні $1-\frac{1}{t^2}$ значень попадає у проміжок, що лежить у межах t стандартних відхилень від середнього значення. Наприклад, для $t=3$ буде отримано, що принаймні 88,9% значень сукупності знаходитимуться в межах потроєного стандартного відхилення від середнього значення, тобто для довільного розподілу з гарантією 89% і вище справедливе відоме правило 3-х сигм. З огляду на чуттєвість до граничних значень розмах виявляється не зовсім корисним в якості тієї статистичної міри розкиду, яка характеризує сукупність значень величини ознаки в цілому. Розмах не відображає типову мінливість значень, а скоріше фокусується всього лише на двох значеннях. Стандартне ж відхилення обчислюється за всіма значеннями сукупності, завдяки чому дана характеристика є об’єктивнішою та усталенішою й дає чітке уявлення про всю картину значень у сукупності.

3. Коефіцієнт варіації вибирається як відносна (протилежно абсолютній) міра мінливості. Цей показник використовується достатньо часто та показує, наскільки сильно зазвичай відрізняється результат окремого спостереження ознаки від середнього значення у процентному відношенні до середнього, при цьому застосовується відношення стандартного відхилення від середнього значення величини ознаки.

Коефіцієнт варіації – безрозмірна статистична величина. Його доцільно використовувати для порівняння мінливості сукупностей значень величин ознак, що подані в різних одиницях вимірювання. Слід зазначити, що коефіцієнт варіації може дорівнювати і навіть бути більшим за 100%. Зокрема, при показниковому законі розподілу теоретичний коефіцієнт варіації дорівнює 100%, оскільки для цього розподілу $M(x) = \sigma$. Коефіцієнт варіації використовується для відбракування маловаріабельних величин ознак, коли значення коефіцієнта менше 5%.

4. Міжквартильний розмах є "робастною" (усталеною до наявності викидів) мірою мінливості.

Таким чином, виконавши теоретичний аналіз існуючих методів та засобів процесу проведення статистичного аналізу опису значень величин елементарних ознак об’єктів в економіці, одержані узагальнення слід відобразити у вигляді їх концептуальної схеми, що наведена на рис. 3.13.

Рекомендована концептуальна схема була реалізована для статистичного опису значень величин 49 ознак, що характеризують розвиток економіки 28 промислових підприємств Харківського регіону протягом 8 років з використанням спеціального статистичного пакета Statgraphics *Plus* V5.1 International Professional. Початкові елементарні ознаки сформовані завдяки теоретико-логічному аналізу виробничо-господарської діяльності підприємств [45; 47; 55; 70; 80; 86; 212]. Умовне позначення показників подано в табл. 3.1.

Таблиця 3. 1

Система показників, що описує виробничо-господарську діяльність підприємства

№ з/п	Назва показника	Одиниця виміру	Умовне позначення
1	2	3	4
1	Питома вага основних засобів та позаоборотних активів у балансі	%	x_1
2	Питома вага оборотних засобів у балансі	%	x_2
3	Питома вага матеріальних засобів в оборотних засобах	%	x_3
4	Питома вага грошових коштів і виробничих активів в оборотних засобах	%	x_4
5	Питома вага оборотних засобів у виробничому потенціалі	%	x_5
6	Частка виробничих запасів у виробничому потенціалі	%	x_6
7	Частка незавершеного виробництва у виробничому потенціалі	%	x_7
8	Частка виробничого потенціалу в загальній сумі господарських засобів підприємства	%	x_8
9	Питома вага активної частини в оборотних засобах	%	x_9
10	Коефіцієнт зносу основних засобів	–	x_{10}
11	Коефіцієнт поновлення основних засобів	–	x_{11}
12	Коефіцієнт вибуття основних засобів	–	x_{12}
13	Коефіцієнт покриття	–	x_{13}
14	Питома вага власних засобів у балансі	%	x_{14}
15	Питома вага власних оборотних засобів у загальній сумі власних засобів	%	x_{15}
16	Коефіцієнт автономії	–	x_{16}
17	Коефіцієнт позичених засобів	–	x_{17}
18	Коефіцієнт покриття інвестицій	–	x_{18}
19	Коефіцієнт забезпеченості поточних активів власними оборотними засобами	–	x_{19}
20	Коефіцієнт забезпеченості матеріальних оборотних засобів власними оборотними засобами	–	x_{20}
21	Коефіцієнт покриття матеріальних оборотних засобів	–	x_{21}
22	Коефіцієнт маневрування власного капіталу	–	x_{22}
23	Коефіцієнт маневрування капіталу, який функціонує	–	x_{23}
24	Коефіцієнт абсолютної ліквідності	–	x_{24}

Закінчення табл. 3.1

1	2	3	4
25	Коефіцієнт швидкої ліквідності	–	x_{25}
26	Коефіцієнт забезпеченості власними засобами	–	x_{26}
27	Чисельність виробничо-промислового персоналу	осіб	x_{27}
28	Виручка від реалізації	тис. грн	x_{28}
29	Чистий прибуток	тис. грн	x_{29}
30	Оборотність основного капіталу (активів)	разів	x_{30}
31	Оборотність власного капіталу	разів	x_{31}
32	Оборотність інвестованого капіталу	разів	x_{32}
33	Оборотність основних фондів	разів	x_{33}
34	Оборотність матеріальних оборотних засобів	разів	x_{34}
35	Термін зберігання матеріальних оборотних засобів	днів	x_{35}
36	Оборотність кредитної заборгованості	разів	x_{36}
37	Термін обороту кредитної заборгованості	днів	x_{37}
38	Оборотність дебіторської заборгованості	разів	x_{38}
39	Період погашення дебіторської заборгованості	днів	x_{39}
40	Рентабельність активів	–	x_{40}
41	Рентабельність власного капіталу	–	x_{41}
42	Загальна рентабельність	–	x_{42}
43	Рентабельність продажів	–	x_{43}
44	Рентабельність основної діяльності	–	x_{44}
45	Прибутковість власного капіталу	–	x_{45}
46	Продуктивність праці	тис. грн/особу	x_{46}
47	Прибуток на одного працівника	тис. грн	x_{47}
48	Фондоозброєність	тис. грн/особу	x_{48}
49	Тривалість операційного циклу	днів	x_{49}

Роботу в пакеті для отримання відповідних обчислень слуд організувати так. Ввести вихідні дані (вікно вводу подано на рис. 3.12) та відкрити файл, який містить дані для аналізу. У головному меню активізувати Describe/Numeric Data.

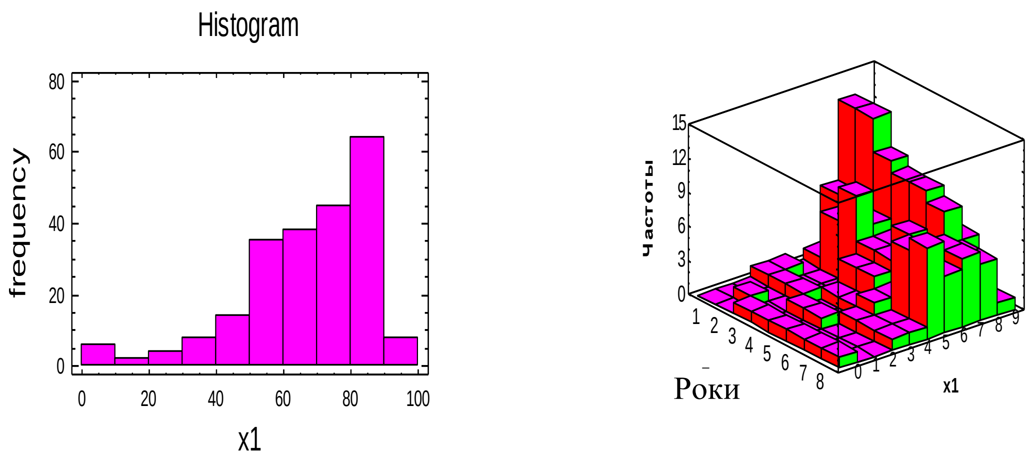
За викладеними рекомендаціями описової статистики з системи показників спочатку слід виключити ті показники, в яких коефіцієнт варіації менше 5%, це виявились $x_{11}, x_{12}, x_{15}, x_{49}$.

Інструменти описової статистики	Тип ознаки		
	Кількісна	**Порядкова**	**Номінальна**
1. Показники, що описують положення значень величини ознаки			
середнє	+		
медіана	+	+	
мода	+	+	+
найбільше значення	+	+	
найменше значення	+	+	
нижній квартиль	+	+	
верхній квартиль	+	+	
2. Показники, що описують розкид значень величини ознаки			
дисперсія	+		
стандартне відхилення	+		
розмах сукупності	+		
коефіцієнт варіації	+		
3. Показники, що описують форму розподілу значень величини ознаки			
коефіцієнт асиметрії	+		
коефіцієнт ексцесу	+		
4. Графічні засоби, що описують закон розподілу значень величини ознаки			
гістограми, кутуляти	+		
блокові діаграми	+		
таблиці частот	+	+	+

Рис. 3.13. **Схема інструментарію описової статистики елементарних ознак об’єктів**

Отже, тепер у системі 45 показників. Згідно з першим етапом схеми був виконаний візуальний аналіз розподілу значень величини ознак за допомогою гістограм – двохвимірних та тривимірних, що враховують динаміку та типи розвитку в однорідних групах підприємств, отриманих за допомогою кластерного аналізу, окремий розгляд якого здійснено в наступному підрозділі.

Пріоритет візуального зображення підтверджується змістом рис. 3.14, що демонструє елементарну наглядність та можливість відразу виділити існуючі відхилення в значеннях величини ознак. Детальна демонстрація реалізації етапів концептуальної схеми виконана для дослідження ознаки x_1 – частки основних засобів та позаоборотних активів у балансі підприємства.



а) б)

Рис. 3.14. **Двовимірна (а) та тривимірна (б) гістограми частот значень величини ознаки** x_1

На рис. 3.14 (а) зображена гістограма частот елементарної ознаки у всій сукупності підприємств протягом усього аналізованого періоду (роки замінені порядковим номером). Слід відмітити асиметрію в розподілі значень ознаки в напрямку високих значень. Тривимірну гістограму (рис. 3.14. (б)) можна представити як набір гістограм ознаки за кожний рік (усього 8 гістограм, що демонструють динаміку мінливості значень величини ознаки в часі). Оскільки просторовий графік достатньо складний для сприйняття, доцільно мати детальний аналіз розподілу значень ознаки в перетині кожного року (рис. 3.15).

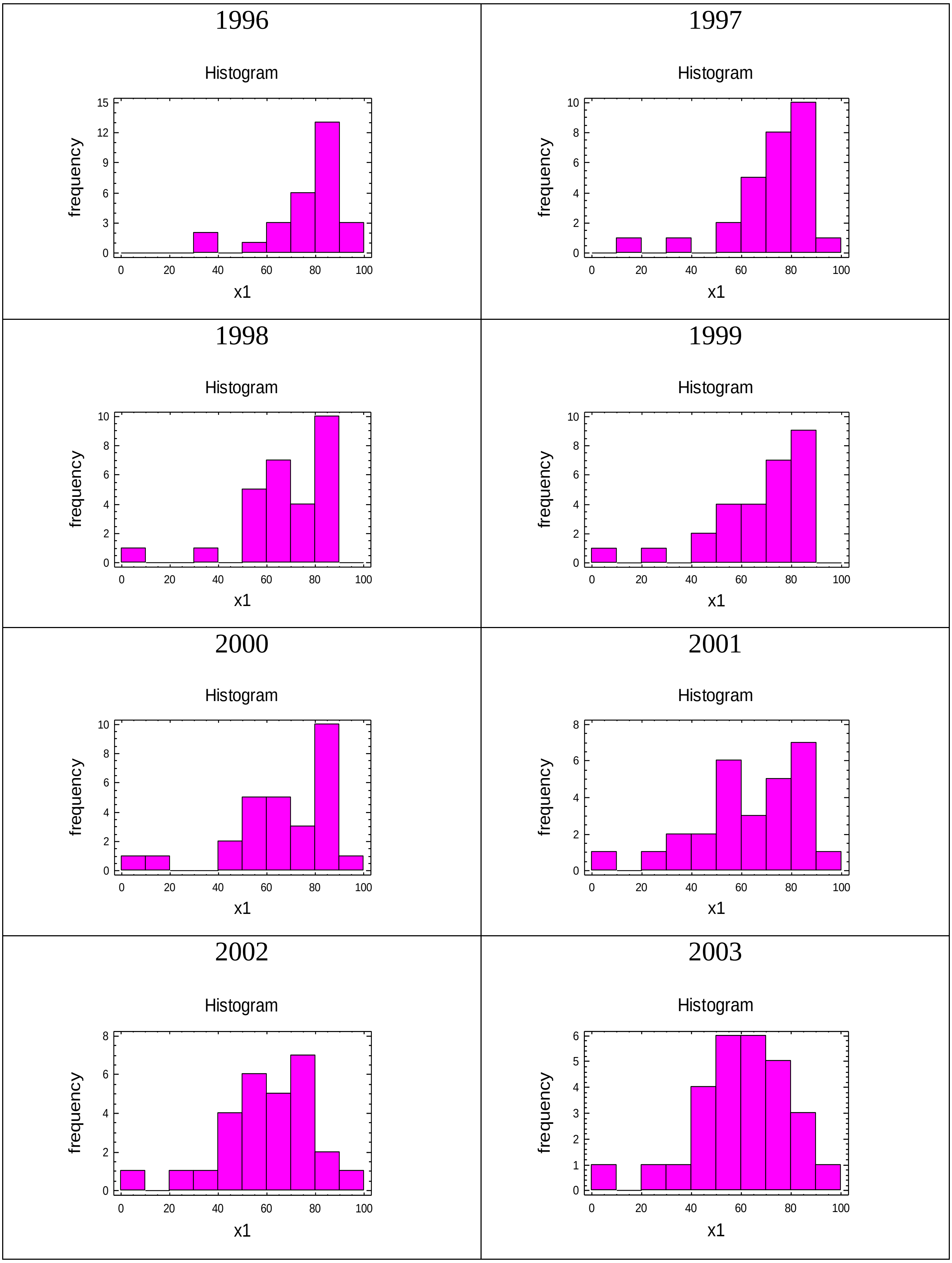


Рис. 3.15. **Гістограми розподілу значень частки основних засобів та позаоборотних активів у балансі підприємств щорічно**

Порівняльний аналіз даних гістограм засвідчує асиметрію в розподілі значень ознаки в перших шести роках, що дає змогу говорити про хаотичність розвитку підприємств, і тільки в останні два роки спостерігається деяка узгодженість у розподілі.

За другим етапом статистичного аналізу потрібно обчислити показники, що характеризують положення значень величини ознаки на числовій осі в кожній з 8-ми груп підприємств та в цілому в сукупності за 2003 рік (табл. 3.2).

Таблиця 3.2

Сумарна статистика для ознаки x_1

Статистичний показник	Номер групи підприємств (2003 р.)								
	1	2	3	4	5	6	7	8	Загальне за сукупністю
Кількість об'єктів у групі	9	9	2	1	1	3	1	2	28
Середнє	75,04	51,45	52,22	64,88	94,02	45,77	1,85	65,76	59,80
Медіана	75,79	51,8	53,22	64,88	94,02	47,25	1,85	65,76	61,034
Мода			53,2	64,88	94,02		1,85	60,3	
Стандартне відхилення	5,49	8,30	0,02	0,00	0,00	16,79	0,00	7,71	18,70
Мінімум	66,70	37,65	53,20	64,88	94,02	28,28	1,85	60,30	1,85
Максимум	82,62	67,00	53,23	64,88	94,02	61,77	1,85	71,21	94,02
Нижній квартиль	71,23	46,00	53,20	64,88	94,02	28,28	1,85	60,30	50,45
Верхній квартиль	80,13	55,48	53,23	64,88	94,02	61,77	1,85	71,21	72,51
Міжквартильний розмах	8,90	9,48	0,03	0,00	0,00	33,49	0,00	10,91	22,06
Коефіцієнт варіації	7,3%	16,1%	0,0%	0,0%	0,0%	36,7%	0,0%	11,7%	31,3%

Порівнюючи значення середнього, медіани та моди в таблиці, можна зробити висновок про практично нормальний закон розподілу значень ознаки в кожній групі. Згідно із значеннями величини стандартного відхилення в табл. 3.2 маємо, що ознака x_1 мінлива у всій сукупності підприємств, а також у групах: 6-й та 2-й. Перша група має найменший міжквартильний розмах, отже спостерігається скупченість значень величини; така ж ситуація і в 2-й групі. 6-та група є розмитою за значеннями величини ознаки і, звичайно, тут маємо найбільший коефіцієнт варіації. Для детального вивчення різних джерел варіації виконують дисперсійний аналіз (скорочено дана стандартна обчислювальна процедура позначається ANOVA). Табл. 3.3 містить результати обчислень дисперсійного аналізу величини ознаки x_1: сум квадратів, степенів свободи, середніх квадратів, F-відношень та p-значень.

Таблиця 3.3

Таблиця дисперсійного аналізу

Вид мінливості (джерело)	Сума квадратів	Степені свободи	Середні квадрати	F-відно-шення	Рівень значущості
Між групами	8 023,01	7	1 146,14	16,20	0,0000
Усередині груп	1 415,2	20	70,7599		
Загальна	9 438,21	27			

F-статистика для однофакторного дисперсійного аналізу становить відношення дисперсії з двох джерел: міжгрупову варіацію та всередині груп. *F*-статистика показує, у скільки разів мінливість між групами перевищує мінливість усередині груп, яка вважається випадковою. З табл. 3.3 випливає, що міжгрупова дисперсія (за групами підприємств) в 16,2 раза перевищує варіацію всередині груп (випадкову). Таким чином, варіація значень величини ознаки за групами в 16,2 раза більше, ніж можливо було б чекати, виходячи тільки з варіації окремих значень ознаки на конкретному підприємстві. В останній колонці табл. 3.2. наведена ймовірність отримання вказаного чи більшого значення F-відношення (рівень значущості F-статистики). При справедливості нульової гіпотези ця ймовірність (рівень значущості) практично дорівнює нулю, тому між виділеними групами об’єктів наявні значимі відмінності. Отже, потрібно з’ясувати, між якими групами спостерігаються істотні відмінності. Для цього корисно обчислити межі довірливих 95-процентних інтервалів на середні в кожній групі та порівняти їх. Результати цих обчислень наведені в табл. 3.4.

Таблиця 3.4

Вихідні дані для графічного зображення середніх значень величин

Номер групи	Кількість об’єктів у групі	Середнє	Похибка середнього	Нижня межа	Верхня Межа
1	9	75,0444	2,80396	70,9086	79,1803
2	9	51,4467	2,80396	47,3108	55,5825
3	2	53,215	5,94811	44,4415	61,9885
4	1	64,88	8,41189	52,4724	77,2876
5	1	94,02	8,41189	81,6124	106,428
6	3	45,7667	4,85661	38,6032	52,9302
7	1	1,85	8,41189	-10,5576	14,2576
8	2	65,755	5,94811	56,9815	74,5285
Загальне	28	59,8004			

На рис. 3.16 зображені середні в кожному кластері з 95-процентним довірливим інтервалом.

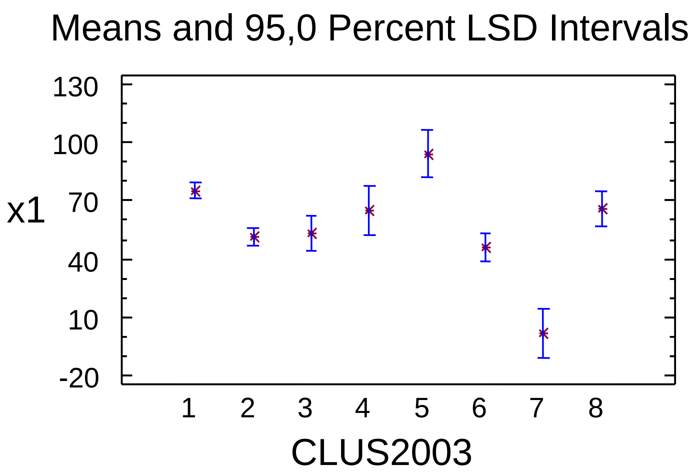


Рис. 3.16. **Середні значення величини ознаки x_1 у кожному кластері (8 кластерів) з 95-процентними довірливими інтервалами**

Чим більше об’єктів у групі, тим надійніше визначене середнє в групі, тим вужче довірливий інтервал (ширина якого обернено пропорційна кореню квадратному з числа об’єктів). Найширші інтервали (полоси невизначеності) для 4, 5, 7 груп, які містять по одному об’єкту (існують найбільші розсіяння значень величини ознаки). Ширина довірливих інтервалів у 1-й та 2-й групах у 3 рази вужче, оскільки ці групи містять по 9 об’єктів. Довірливі інтервали для деяких груп перекриваються, що відповідає відсутності значущих відмінностей між цими групами.

Необхідно врахувати переваги блокових діаграм, що заключаються в можливості сконцентрувати увагу на основних особливостях декількох множин значень величини ознаки одночасно, не відволікаючись на деталі. Отже, є вісім самостійних наборів значень величини ознаки (кластерів). Для кожної групи підприємств доцільно обчислити п'ять базових статистичних величин та побудувана блокова діаграма. Розмістивши на одному рисунку побудовані в одному маштабі блокові діаграми, можна легко порівняти між собою особливості розподілу значень величини ознаки x_1. На рис. 3.17 зображені "ящики з вусами", побудовані за групами підприємств (часовий зріз: 2003 рік).

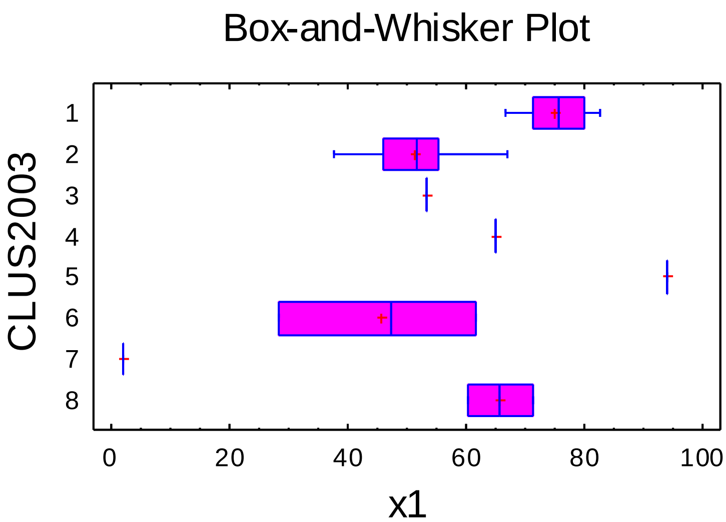


Рис. 3.17. **Блокова діаграма значень величини ознаки x_1 за групами підприємств (у 2003 році)**

Блокові діаграми 6-ї та 8-ї груп підприємств на мають "вусів", оскільки в цих групах мало об'єктів. Блокові діаграми 3-ї, 4-ї, 5-ї та 7-ї груп виявились виродженими, оскільки ці групи складаються з одного чи двох об'єктів. Результати аналізу множинних порівнянь за групами подані в табл. 3.5.

У стовпці "Однорідні групи" вертикальними стовпцями зірок виділені можливі однорідні групи рівнів обробки величини ознаки. Маємо декілька наборів однорідних груп: до першого набору увійшли групи 6-та, 2-га, 3-тя та 4-та; до другого набору – 3-тя, 4-та, 8-ма; до третього – 4-та, 8-ма, 1-ша. Отже, для даних груп об'єктів можна застосовувати однакові коефіцієнти в їх моделях, розробляти та використовувати однакові методи управління економіки. Існують групи, що в сукупності не мають однорідних, це 7-ма та 5-та, вони нерепрезентативні для виявлення загальних закономірностей досліджуваних об’єктів, але потребують детального вивчення, оскільки вносять особливе в сукупність об’єктів.

Таблиця 3.5

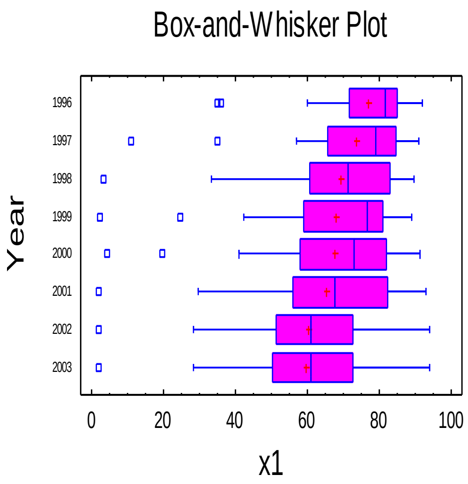
Порівняння груп об’єктів

Групи	Кількість об’єктів у групі	Середнє	Однорідні групи
1	2	3	4
7	1	1.85	X
6	3	45,7667	X
2	9	51,4467	X
3	2	53,215	XX
4	1	64,88	XXX
8	2	65,755	XX
1	9	75,0444	X
5	1	94,02	X
Парне порівняння	Різниця між середніми	Найменша суттєва різниця	
1-2	*23,5978	8,2717	
1-3	*21,8294	13,7171	
1-4	10,1644	18,4961	
1-5	*-18,9756	18,4961	
1-6	*29,2778	11,698	
1-7	*73,1944	18,4961	
1-8	9,28944	13,7171	
2-3	-1,76833	13,7171	
2-4	-13,4333	18,4961	
2-5	*-42,5733	18,4961	
2-6	5,68	11,698	
2-7	*49,5967	18,4961	
2-8	*-14,3083	13,7171	
3-4	-11,665	21,4905	
3-5	*-40,805	21,4905	
3-6	7,44833	16,0181	
3-7	*51,365	21,4905	
3-8	-12,54	17,5469	
4-5	*-29,14	24,8151	
4-6	19,1133	20,2615	

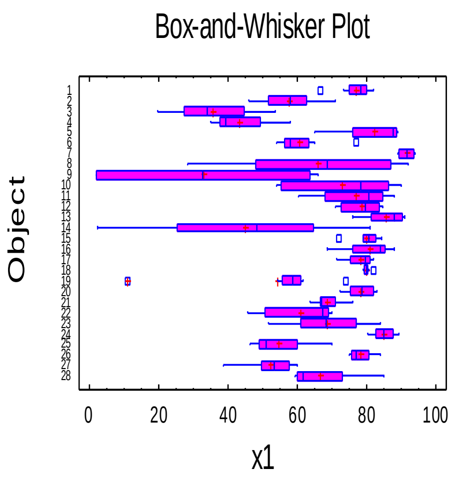
Закінчення табл. 3.5

1	2	3	4
4-7	*63,03	24,8151	
4-8	-0,875	21,4905	
5-6	*48,2533	20,2615	
5-7	*92,17	24,8151	
5-8	*28,265	21,4905	
6-7	*43,9167	20,2615	
6-8	*-19,9883	16,0181	
7-8	*-63,905	21,4905	

Обґрунтування тенденції зміни величини ознаки в динаміці випливає зі змісту рис. 3.18, що демонструє напрямок зміни – спадну односторонню, що свідчить про правильну стратегію системи управління на підприємствах відносно даної ознаки їх розвитку.



а



б

Рис. 3.18. **Блокові діаграми значень величини ознаки в динаміці всієї сукупності підприємств (а) та на кожному підприємстві за весь період дослідження (б)**

На рис. 3.19 наведені блокові діаграми значень величини ознаки в статиці та динаміці у всіх можливих зрізах.

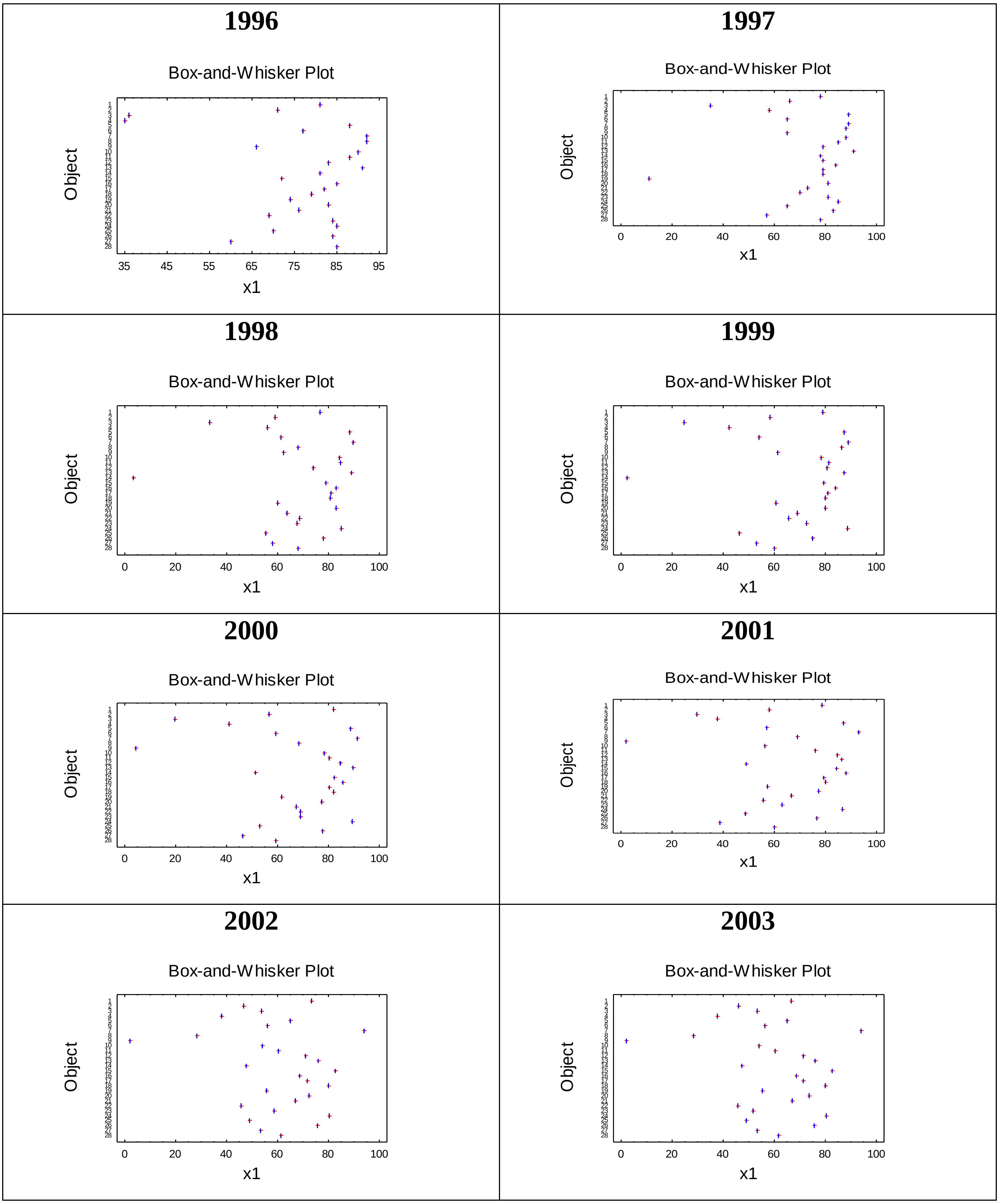


Рис. 3.19. **Блокові діаграми значень величини ознаки в статиці та динаміці у всіх можливих зрізах**

На рис. 3.19 спостерігається розмитість значень величини, яка не дозволяє виявити яку-небудь закономірність у змінах значень ознаки об'єктів сукупності, що підтверджує об'єктивність групового підходу до визначення статистичних величин ознак об'єктів.

Результати застосувань інструментів описової статистики для дослідження величин 45 елементарних ознак виробничо-господарської діяльності 28 промислових підприємств Харківського регіону зведені в табл. 3.6,

де *ОСНВ* – одностороння, спадна тенденція до змін величини, викидів немає;

ОСВ – одностороння, спадна тенденція до змін величини, викиди є;

ОЗНВ – одностороння, зростаюча тенденція до змін величини, викидів немає;

ОЗВ – одностороння, зростаюча тенденція до змін величини, викиди є;

ДНВ – двостороння тенденція до змін величини, викидів немає;

ДВ – двостороння тенденція до змін величини, викиди є;

НЗР – нормальний закон розподілу;

БНЗР – близький до нормального закону;

НСР – несиметричний закон розподілу;

ЗРАП – закон розподілу абсолютних показників.

Таблиця 3.6

Типи змін величин елементарних ознак

Закон розподілу	Тенденція до змін та існування викидів					
	ОСНВ	*ОСВ*	*ОЗНВ*	*ОЗВ*	*ДНВ*	*ДВ*
НЗР	x_1, x_{17}	x_{10}^*	x_{19}	x_9^*, x_{18}^*		x_{16}, x_{26}
БНЗР		x_6^*, x_7^*		x_5^*, x_8^*		x_{22}
НСР				x_4, x_{20}^*, x_{21}^*	x_{13}^*, x_{23}	x_{14}, x_{25}^*, x_{24}
ЗРАП			$x_{27} \div x_{48}$			

Ознаки, що позначені символом "*****", у табл. 3.6 мають величини, тенденція змін значень яких у даному періоді протилежна закономірній тенденції функціонування підприємства. Органам управління в регіоні слід ретельно проаналізувати причину стійких негативних тенденцій і стратегічні плани розробляти з урахуванням зміни напрямків розвитку, що склалися. Для виявлених однорідних груп підприємств можна встановлювати однакові нормативи, еталони, оскільки для них, як було визначено, наявні однакові реперні значення величин на шкалі значень ознак. Для виявлених однорідних груп підприємств можна встановлювати однакові нормативи, еталони, оскільки для них, як було визначено, спостерігаються однакові реперні значення величин на шкалі значень ознак.

Багато практичних задач в управлінні різного рівня соціально-економічними системами передбачає побудову шкал величини окремої ознаки. Для побудови шкали потрібно статистично дослідити величину ознаки, зміну її значень, керуючись наступною схемою: визначити тип ряду спостережень: просторовий, динамічний, просторово-динамічний; перевірити сукупність значень величини, сукупність об’єктів на однорідність, використавши при цьому кластерний і дисперсійний аналізи; визначити показники, що описують положення означень величини в сукупності однорідних об’єктів; визначити показники, що характеризують мінливість ознаки; визначити показники асиметрії значень величини ознаки; використати графічні засоби, що описують закон розподілу значень величини ознаки; якщо сукупність значень показників не однорідна, то всі засоби описової статистики слід повторити в кожній з груп; сформувати основні статистичні закономірності змін значень величини по шкалі за критеріями: закону розподілу, існування викидів, тенденції змін.

Загальний вигляд шкали окремої ознаки, розбудованої за допомогою статистичних характеристик величини, зображений на рис. 3.20,
де $x'_{найм}$, $x'_{найб}$ – найменше і найбільше значення ознаки в сукупності;

Me – медіана значень ознаки в сукупності;

Me_i – медіани значень ознаки в i-й групі (отриманій завдяки кластерному аналізу) сукупності.

Наведена шкала є попередньою для побудови шкали величин ознак об’єкта та шкали якості ознак об’єкта в економіці в цілому.

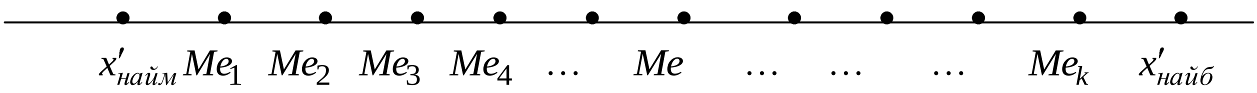


Рис. 3.20. **Узагальнений вигляд шкали окремої ознаки об’єктів у сукупності**

Розглянута логіка проведення статистичного аналізу величин ознак рекомендується для проведення будь-якого економічного аналізу, а також, першим необхідним етапом у побудові економіко-математичної моделі об’єкта в економіці.

3.3. Типові задачі багатовимірного аналізу соціально-економічних систем за метричними ознаками

Задача 3.1. Ідентифікувати складні ознаки виробничо-господарської діяльності 28 промислових підприємств Харківського регіону за 45 елементарними метричними ознаками, перелік яких наведений у табл. 3.1.

Розв'язання задачі. Першим етапом розбудови моделей I типу є ідентифікація загальних складних ознак СЕС. У підрозділі 2.4 обговорювалось питання про те, що початковий факторний розв'язок доцільно одержувати за допомогою методу головних компонент. У статистичному пакеті Statgraphics початковий факторний розв'язок отримує за допомогою активації Special/Multivariate Methods/Principal Components. Існує декілька рекомендацій щодо визначення кількості головних компонент. На рис. 3.21 зображений графік послідовності власних чисел, який має характерну точку, що відділяє гілку швидкого спаду значень власних чисел від гілки майже сталих значень.

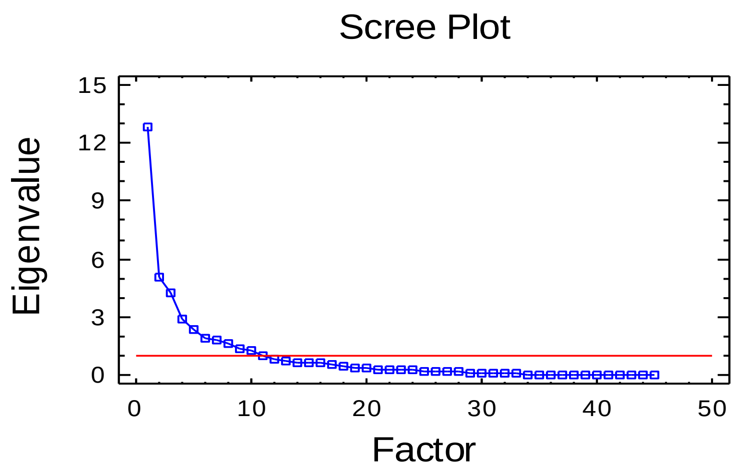


Рис. 3.21. **Графік власних значень кореляційної матриці**

Рекомендується кількість головних компонент визначати за цією точкою. Згідно з даною рекомендацією можна одержати 6 головних компонент. За іншою рекомендацією – $\lambda_i \geq 1$ – потрібно 11 компонент, за третьою (накопичена сума повинна бути більшою від 75%) – необхідно залишити 9 компонент. Зміст прикладної задачі та початкові результати її вирішення обумовлюють кількість головних компонент. Рис. 3.21 демонструє найбільший внесок першого фактора, який пояснює більше 30% всієї мінливості, у два рази менший процент дисперсії для другого фактора. Починаючи з 6-го фактора дисперсії, внески їх не дуже відрізняються між собою.

У табл. 3.7 наведені значення факторних навантажень до процедури повертання 11-ти факторів (тобто розклад за головними компонентами). З даної таблиці видно, що 11-ти виділених факторів достатньо для вичерпного пояснення системи ознак, крім x_{23}. Ця ознака тільки на 51,62% пояснюється виділеними факторами, отже, вона не так тісно взаємозв'язана в системі ознак, як усі інші. У стовпчиках даної таблиці наведений розклад компонент за системою елементарних ознак, а в рядках – розклад кожної ознаки за компонентами. Вирішивши проблему повертання методом $VARIMAX$, маємо розв'язок простішої структури в описанні елементарних ознак об'єкта. Матриця факторних навантажень після повертання $VARIMAX$ подана в табл. 3.8.

Порівняння змісту табл. 3.7 і табл. 3.8 демонструє явне спрощення структури факторів, що виражається в тому, що кожну елементарну ознаку обумовлює один фактор. З точки зору економіки виявляються ознаки-причини, що є визначальними для функціонування об'єкта на елементарному рівні.

Модель загальних складних метричних ознак виробничо-господарської діяльності промислових підприємств наведена в табл. 3.9.

Задача розпізнавання головних ознак і визначення для них назви є важливою та ґрунтується на рівнях вагових коефіцієнтів a_{jr} з матриці відображення $\boldsymbol{A}$. Існує спеціальна методика визначення критичної кількості ознак для пояснення головних факторів [48; 49].

Таблиця 3.7

Факторні навантаження до обертання

№	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	Спільності
1	2	3	4	5	6	7	8	9	10	11	12	13
x1	0,86649	0,123975	0,274603	-0,18052	0,053842	-0,08437	0,039929	-0,05956	-0,04028	0,110381	0,139472	0,922593
x2	-0,866341	-0,12514	-0,27466	0,180308	-0,05391	0,084548	-0,04213	0,060453	0,040037	-0,10944	-0,14005	0,92284
x3	0,459972	0,29727	-0,21403	0,729227	-0,15886	0,067468	-0,03489	0,07394	-0,12790	0,058659	0,064113	0,937912
x4	-0,46809	-0,29781	0,208791	-0,71472	0,151546	-0,11616	0,035464	-0,04905	0,118751	-0,03572	-0,07799	0,923806
x5	0,704558	0,126711	0,438985	-0,08484	0,243071	-0,04625	-0,16658	-0,15230	0,00754	0,042050	-0,10446	0,837272
x6	-0,334932	0,557655	-0,37308	0,156756	0,015291	-0,18495	0,047781	0,074007	-0,09416	-0,20890	-0,08957	0,689655
x7	-0,609214	-0,49163	-0,27792	0,001370	-0,29108	0,167958	0,163465	0,130699	0,049488	0,080452	0,176423	0,886874
x8	0,646641	0,150592	0,137771	-0,00534	0,127062	-0,12254	-0,09091	-0,24158	-0,00077	0,232546	-0,20455	0,653543
x9	0,0033251	-0,14828	0,278992	0,14426	-0,03377	0,481633	-0,38372	-0,20796	0,378965	-0,23215	0,034102	0,742925
x10	-0,241596	-0,02170	-0,34140	0,37825	0,496111	-0,04897	0,045158	0,15491	-0,28234	-0,13888	-0,09694	0,701439
x13	-0,748834	-0,42070	0,019277	0,001468	-0,11136	0,031507	-0,05941	0,006766	-0,32638	0,143471	0,089747	0,890261
x14	-0,682145	-0,18809	-0,22158	-0,08825	-0,28003	0,050094	-0,39686	0,092837	-0,14290	0,139854	-0,22647	0,895906
x16	-0,485653	-0,08646	0,723395	0,283245	0,073590	-0,04379	0,064156	-0,05475	-0,08477	0,111792	-0,05702	0,884246
x17	0,465662	0,072846	-0,64621	-0,13182	-0,06211	0,120594	-0,02959	-0,14353	0,099930	-0,24775	0,15114	0,791215
x18	-0,484688	-0,06765	0,643691	0,213658	0,129005	-0,04783	0,182884	-0,15624	-0,19389	-0,03233	-0,02380	0,815483
x19	-0,624135	0,051930	-0,09545	0,371002	0,453036	-0,10591	0,122638	-0,16836	0,186797	0,077727	0,106869	0,851197
x20	-0,688248	-0,19755	-0,09282	0,246317	0,42414	0,038807	0,071177	-0,25032	0,14158	0,109521	0,256238	0,928827
x21	-0,584855	-0,32955	-0,04589	0,132058	0,373515	0,204488	-0,00224	-0,28196	0,150214	0,080797	0,335078	0,872412
x22	-0,410344	0,038852	-0,72948	-0,03232	0,216465	-0,26276	-0,13645	-0,05619	0,124619	0,20847	-0,04870	0,902122
x23	-0,550828	-0,22201	-0,05764	0,053177	0,110802	0,103565	-0,21497	-0,19083	0,126978	-0,12804	-0,13844	0,516171
x24	-0,674837	-0,57632	-0,05049	-0,18883	-0,13533	0,088272	-0,04358	-0,00295	-0,27260	0,077362	0,110695	0,946329

Закінчення табл. 3.7

1	2	3	4	5	6	7	8	9	10	11	12	13
x25	-0,734228	-0,16147	0,15033	-0,07642	-0,04197	-0,07265	-0,09239	-0,02591	-0,16935	0,093920	-0,23853	0,704256
x26	-0,7252	-0,02096	0,237479	0,306463	0,14666	-0,00927	-0,17805	-0,02502	0,130877	0,035288	-0,30159	0,839926
x27	0,0820509	0,098962	-0,28266	-0,11384	-0,06826	0,518589	0,465187	-0,30308	-0,12270	-0,13242	-0,07253	0,729093
x28	-0,121588	0,369077	-0,24052	-0,06829	-0,08923	0,293302	0,509005	-0,32169	-0,10567	0,283146	-0,22089	0,810204
x29	-0,204398	0,031288	0,57771	0,097017	-0,14048	-0,01321	-0,26738	0,369757	0,096308	0,206307	0,292414	0,751388
x30	-0,496935	0,717223	0,045821	-0,08557	-0,09744	0,045047	-0,26858	-0,15104	-0,10464	-0,06937	0,030359	0,893934
x31	-0,530247	0,69352	0,061868	-0,12922	-0,13413	0,078814	-0,24210	-0,14913	-0,10338	-0,11231	-0,00451	0,911039
x32	-0,614238	0,539188	-0,22074	-0,19370	-0,13785	0,12226	-0,07924	0,026638	0,299245	0,058939	-0,02573	0,888885
x33	-0,693639	-0,52106	-0,14698	-0,14059	-0,25215	0,129745	0,017103	0,040448	-0,17605	0,077295	0,175452	0,944113
x34	-0,085319	0,017448	0,336615	-0,61497	0,439405	0,063525	0,005870	-0,01687	0,074583	0,22135	0,047864	0,753372
x35	0,204862	-0,22587	0,004678	0,654972	-0,15677	-0,24114	-0,18782	-0,19669	-0,05275	0,258632	-0,04196	0,750129
x36	-0,71537	0,319284	0,17938	-0,21759	-0,02347	-0,01629	-0,20204	-0,14591	-0,21814	0,061926	0,052121	0,810291
x37	0,222962	-0,15648	-0,10633	0,054943	-0,52243	-0,25680	-0,07915	-0,18411	0,328239	0,2806	0,145648	0,67526
x38	-0,199851	0,604985	0,064474	0,106406	0,010623	0,270179	-0,00027	0,366698	-0,07257	0,389575	0,183334	0,819651
x39	0,128706	-0,59827	-0,04455	-0,03108	-0,27504	-0,24657	-0,05254	-0,38029	-0,05538	-0,33652	-0,16372	0,804406
x40	-0,651624	0,354199	0,012482	0,090989	-0,10262	-0,06283	-0,11041	-0,29525	0,081160	-0,13863	0,043686	0,700068
x41	-0,517527	0,15394	0,357627	0,152089	-0,29268	0,067607	0,262377	-0,01861	0,388649	0,142218	-0,17624	0,804319
x42	-0,573156	0,106703	0,020564	0,013045	-0,22665	-0,28414	0,347455	0,108489	0,271862	0,027934	-0,01068	0,679892
x43	-0,661874	0,16898	0,101388	-0,03083	-0,22009	-0,28510	0,303215	0,198905	0,17846	-0,10287	-0,03238	0,78257
x44	-0,500374	0,098085	-0,08413	-0,23704	0,151642	-0,56130	0,133812	0,047612	-0,01307	-0,01424	0,014574	0,682088
x45	-0,195993	-0,01897	0,665758	0,219002	-0,19745	0,111961	0,41808	-0,03251	-0,10085	-0,26203	0,036383	0,837497
x46	-0,459882	0,715353	0,130213	-0,11730	-0,20000	0,073010	-0,13051	-0,19980	-0,16349	0,060197	0,125273	0,902273
x47	-0,3286	0,383336	-0,07378	0,001394	0,054750	-0,42286	0,121388	-0,22693	-0,14110	-0,24533	0,324665	0,693926
x48	0,56588	-0,07721	0,112141	-0,02388	-0,37473	-0,23005	-0,01954	-0,40904	-0,12943	0,193206	0,175729	0,785346

Таблиця 3.8

Матриця факторних навантажень після обертання *VARIMAX*

№	F'_1	F'_2	F'_3	F'_4	F'_5	F'_6	F'_7	F'_8	F'_9	F'_{10}	F'_{11}	Спільності
1	2	3	4	5	6	7	8	9	10	11	12	13
x1	-0,70406	-0,27489	-0,02671	-0,30892	0,070599	-0,33701	0,313207	0,063471	-0,01596	0,035887	-0,18151	0,922593
x2	0,704947	0,27433	0,026150	0,30835	-0,07044	0,335886	-0,31304	-0,06316	0,014515	-0,03651	0,18318	0,92284
x3	-0,30091	-0,05140	-0,09566	-0,06066	-0,87933	-0,11580	0,007976	0,20203	0,046663	-0,02194	-0,04126	0,937912
x4	0,306475	0,044611	0,099516	0,054507	0,862758	0,132486	-0,00294	-0,19709	-0,07081	0,068049	0,067905	0,923806
x5	-0,75292	-0,15927	0,130792	-0,17330	0,121945	-0,35992	0,123868	-0,04485	-0,10655	-0,13202	0,085463	0,837272
x6	0,006164	0,526943	-0,20069	0,062909	-0,26190	0,270401	-0,35458	0,045959	0,070489	0,30516	-0,00848	0,689655
x7	0,868545	-0,14128	-0,02744	0,161901	0,020657	0,249818	0,075737	0,023546	0,079874	-0,03572	-0,09377	0,886874
x8	-0,64274	-0,14099	-0,04177	-0,11639	-0,04185	-0,27749	0,231794	-0,04490	0,082975	0,030029	0,250916	0,653543
x9	-0,02282	0,047293	0,093745	0,13827	0,002936	-0,05049	0,037683	-0,11868	-0,13242	-0,82254	-0,00881	0,742925
x10	0,13507	-0,05037	-0,05472	0,356166	-0,26542	-0,11455	-0,62222	0,012683	0,002487	0,272228	0,076164	0,701439
x13	0,839003	0,097061	0,302286	0,190762	0,103053	-0,08630	-0,00837	-0,02845	-0,06036	0,125317	0,104169	0,890261
x14	0,76176	0,308994	-0,07332	-0,04481	0,041623	0,042926	-0,04186	-0,02862	-0,12242	-0,03281	0,436524	0,895906
x16	0,128942	0,106859	0,860322	0,224862	0,059351	0,085702	-0,01024	0,054003	-0,16743	-0,04045	0,148068	0,884246
x17	-0,16202	-0,05406	-0,73932	-0,07957	-0,15838	-0,11938	0,021769	-0,17635	0,225255	-0,08822	-0,28226	0,791215
x18	0,12202	0,149112	0,83201	0,241598	0,097409	0,033458	-0,09658	-0,06600	-0,03797	0,044334	0,006932	0,815483
x19	0,153538	0,169872	0,150339	0,80121	-0,05490	0,230636	-0,22357	0,074577	0,003567	0,126433	0,080453	0,851197
x20	0,380957	0,088245	0,148343	0,852712	0,080490	0,071164	-0,11535	0,036266	0,021787	0,011794	-0,00298	0,928827
x21	0,42437	0,008552	0,116938	0,778708	0,172443	-0,05898	-0,03785	0,014989	0,026157	-0,17169	-0,08455	0,872412
x22	0,297529	0,178371	-0,58440	0,447634	-0,00288	0,142341	-0,09303	0,003572	0,023538	0,34011	0,307752	0,902122
x23	0,38921	0,178127	0,048954	0,35164	0,115383	0,084273	-0,17214	-0,23302	-0,04802	-0,23058	0,216988	0,516171

Закінчення табл. 3.8

1	2	3	4	5	6	7	8	9	10	11	12	13
x24	0,895779	-0,02123	0,17359	0,126147	0,255736	-0,10719	0,02508	-0,12195	-0,02511	0,058460	0,031101	0,946329
x25	0,543671	0,263284	0,342893	0,110322	0,218873	0,104166	-0,11033	-0,08754	-0,05399	0,106534	0,341694	0,704256
x26	0,295423	0,271897	0,406473	0,374942	-0,00333	0,268645	-0,25137	-0,01614	-0,16667	-0,13766	0,43653	0,839926
x27	0,035684	0,012804	-0,10737	-0,05925	-0,00841	-0,04205	-0,11158	-0,04380	0,781652	-0,15241	-0,24939	0,729093
x28	-0,01320	0,21477	0,002033	0,028329	-0,01892	0,123184	0,049538	0,188328	0,827489	0,130045	0,089404	0,810204
x29	0,080230	0,149131	0,405279	-0,12713	0,034291	0,1184	-0,01535	0,174654	-0,61580	-0,25075	-0,23296	0,751388
x30	0,025268	0,919888	0,045262	0,022208	0,027562	0,063084	-0,08872	0,172436	0,002542	-0,00826	0,0462	0,893934
x31	0,066763	0,92376	0,066125	-0,02116	0,067596	0,094408	-0,10788	0,139205	0,033602	-0,03848	0,03631	0,911039
x32	0,212854	0,660321	-0,22297	0,113008	0,161787	0,45185	-0,05147	0,28921	0,102756	-0,08239	0,10528	0,888885
x33	0,939266	-0,00403	0,091566	0,118282	0,165821	0,023228	0,074606	-0,05580	0,015344	0,023584	-0,04465	0,944113
x34	-0,15020	0,007402	0,109491	0,108597	0,796478	-0,14940	0,002282	0,21603	-0,01437	0,020620	0,054489	0,753372
x35	-0,04603	-0,20708	0,14302	0,182857	-0,63131	-0,14353	0,315628	-0,15001	-0,15633	0,039583	0,28972	0,750129
x36	0,332501	0,712145	0,252949	0,122002	0,276678	-0,01915	-0,03635	0,104447	-0,04703	0,103926	0,107486	0,810291
x37	0,013569	-0,11825	-0,20380	-0,05622	-0,20328	0,188728	0,716209	-0,11748	-0,09124	-0,02368	0,061704	0,67526
x38	-0,01394	0,347082	0,075439	-0,03400	-0,11239	0,052215	-0,06855	0,818576	0,029585	0,033501	0,004601	0,819651
x39	0,18058	-0,24045	0,005541	-0,09575	-0,04105	-0,07741	0,19916	-0,80957	-0,02047	-0,03863	0,010941	0,804406
x40	0,203252	0,689835	0,132898	0,298952	-0,02661	0,253977	-0,01825	-0,09343	0,008710	-0,03088	0,023323	0,700068
x41	0,146001	0,207548	0,453643	0,094755	0,030029	0,638491	0,154869	0,147219	0,117239	-0,18191	0,155205	0,804319
x42	0,26956	0,185945	0,170955	0,129022	0,066869	0,675324	0,072314	0,038325	0,001148	0,242962	-0,02272	0,679892
x43	0,304915	0,294651	0,246681	0,038093	0,111265	0,670173	-0,06772	0,029289	-0,07799	0,252041	-0,06253	0,78257
x44	0,158989	0,253317	0,003986	0,207236	0,301281	0,298075	-0,08795	-0,10592	-0,16875	0,56684	0,036300	0,682088
x45	0,019693	0,021117	0,803149	-0,10241	-0,04777	0,214034	-0,03881	-0,08001	0,090745	-0,12243	-0,31921	0,837497
x46	0,039089	0,887662	0,153155	-0,01903	0,040271	0,037702	0,081181	0,252902	0,111409	0,043664	-0,03272	0,902273
x47	-0,03783	0,528096	0,034778	0,260128	-0,02692	0,108172	-0,02134	-0,15713	-0,06296	0,434452	-0,33829	0,693926
x48	-0,27812	-0,09819	0,011595	-0,20863	-0,17409	-0,29694	0,675316	-0,23825	0,068584	0,104743	-0,08781	0,785346

Таблиця 3.9

Модель загальних складних метричних ознак виробничо-господарської діяльності промислових підприємств у регіоні

Власні числа λ_i	Рівняння моделі складних ознак
1	2
$\lambda_1 = 12{,}804$	$F_1 = -0{,}7x_1 + 0{,}7x_2 - 0{,}3x_3 + 0{,}31x_4 - 0{,}75x_5 + 0{,}01x_6 + 0{,}87x_7 - 0{,}64x_8 -$ $- 0{,}02x_9 + 0{,}14x_{10} + 0{,}84x_{13} + 0{,}76x_{14} + 0{,}13x_{16} - 0{,}16x_{17} + 0{,}12x_{18} + 0{,}15x_{19} +$ $+ 038x_{20} + 0{,}42x_{21} + 0{,}3x_{22} + 0{,}39x_{23} + 0{,}9x_{24} + 0{,}54x_{25} + 0{,}3x_{26} + 0{,}04x_{27} -$ $- 0{,}01x_{28} + 0{,}08x_{29} + 0{,}03x_{30} + 0{,}07x_{31} + 0{,}21x_{32} + 0{,}94x_{33} - 0{,}15x_{34} - 0{,}05x_{35} +$ $+ 0{,}33x_{36} + 0{,}01x_{37} - 0{,}02x_{38} + 0{,}18x_{39} + 0{,}2x_{40} + 0{,}15x_{41} + 0{,}3x_{42} + 0{,}3x_{43} +$ $+ 0{,}16x_{44} + 0{,}02x_{45} + 0{,}04x_{46} - 0{,}04x_{47} - 0{,}28x_{48}$
$\lambda_2 = 5{,}08149$	$F_2 = -0{,}27x_1 + 0{,}27x_2 - 0{,}05x_3 + 0{,}04x_4 - 0{,}16x_5 + 0{,}53x_6 - 0{,}14x_7 - 0{,}14x_8 +$ $+ 0{,}05x_9 - 0{,}05x_{10} + 0{,}1x_{13} + 0{,}31x_{14} + 0{,}11x_{16} - 0{,}05x_{17} + 0{,}15x_{18} + 0{,}17x_{19} +$ $+ 0{,}09x_{20} + 0{,}01x_{21} + 0{,}18x_{22} + 0{,}18x_{23} - 0{,}02x_{24} + 0{,}26x_{25} + 0{,}27x_{26} + 0{,}01x_{27} +$ $+ 0{,}21x_{28} + 0{,}15x_{29} + 0{,}92x_{30} + 0{,}92x_{31} + 0{,}66x_{32} - 0{,}004x_{33} + 0{,}01x_{34} - 0{,}21x_{35} +$ $+ 0{,}71x_{36} - 0{,}12x_{37} + 0{,}35x_{38} - 0{,}24x_{39} + 0{,}69x_{40} + 0{,}21x_{41} + 0{,}19x_{42} + 0{,}29x_{43} +$ $+ 0{,}25x_{44} + 0{,}02x_{45} + 0{,}89x_{46} + 0{,}53x_{47} - 0{,}1x_{48}$
$\lambda_3 = 4{,}23909$	$F_3 = -0{,}03x_1 + 0{,}03x_2 - 0{,}1x_3 + 0{,}1x_4 + 0{,}13x_5 - 0{,}2x_6 - 0{,}03x_7 - 0{,}04x_8 +$ $+ 0{,}09x_9 - 0{,}05x_{10} + 0{,}3x_{13} - 0{,}07x_{14} + 0{,}86x_{16} - 0{,}74x_{17} + 0{,}83x_{18} + 0{,}15x_{19} +$ $+ 0{,}15x_{20} + 0{,}12x_{21} - 0{,}58x_{22} + 0{,}05x_{23} + 0{,}17x_{24} + 0{,}34x_{25} + 0{,}41x_{26} - 0{,}11x_{27} +$ $+ 0{,}002x_{28} + 0{,}41x_{29} + 0{,}05x_{30} + 0{,}07x_{31} - 0{,}22x_{32} + 0{,}09x_{33} + 0{,}11x_{34} + 0{,}14x_{35} +$ $+ 0{,}25x_{36} - 0{,}2x_{37} + 0{,}07x_{38} + 0{,}01x_{39} + 0{,}13x_{40} + 0{,}45x_{41} + 0{,}17x_{42} + 0{,}25x_{43} +$ $+ 0{,}003x_{44} + 0{,}8x_{45} + 0{,}15x_{46} + 0{,}03x_{47} + 0{,}01x_{48}$
$\lambda_4 = 2{,}93866$	$F_4 = -0{,}31x_1 + 0{,}31x_2 - 0{,}06x_3 + 0{,}05x_4 - 0{,}17x_5 + 0{,}06x_6 + 0{,}16x_7 - 0{,}12x_8 +$ $+ 0{,}14x_9 + 0{,}36x_{10} + 0{,}19x_{13} - 0{,}04x_{14} + 0{,}22x_{16} - 0{,}08x_{17} + 0{,}24x_{18} + 0{,}8x_{19} +$ $+ 0{,}85x_{20} + 0{,}78x_{21} + 0{,}45x_{22} + 0{,}35x_{23} + 0{,}13x_{24} + 0{,}11x_{25} + 0{,}37x_{26} - 0{,}06x_{27} +$ $+ 0{,}03x_{28} - 0{,}13x_{29} + 0{,}02x_{30} - 0{,}02x_{31} + 0{,}11x_{32} + 0{,}12x_{33} + 0{,}11x_{34} + 0{,}18x_{35} +$ $+ 0{,}12x_{36} - 0{,}06x_{37} - 0{,}03x_{38} - 0{,}1x_{39} + 0{,}3x_{40} + 0{,}1x_{41} + 0{,}13x_{42} + 0{,}04x_{43} +$ $+ 0{,}21x_{44} - 0{,}1x_{45} - 0{,}02x_{46} + 0{,}26x_{47} - 0{,}21x_{48}$
$\lambda_5 = 2{,}35433$	$F_5 = 0{,}07x_1 - 0{,}07x_2 + 0{,}88x_3 + 0{,}86x_4 + 0{,}12x_5 - 0{,}26x_6 + 0{,}02x_7 - 0{,}04x_8 +$ $+ 0{,}003x_9 - 0{,}26x_{10} + 0{,}1x_{13} + 0{,}04x_{14} + 0{,}06x_{16} - 0{,}16x_{17} + 0{,}1x_{18} - 0{,}05x_{19} -$ $- 0{,}08x_{20} + 0{,}17x_{21} - 0{,}003x_{22} + 0{,}12x_{23} + 0{,}26x_{24} + 0{,}22x_{25} - 0{,}003x_{26} - 0{,}01x_{27} -$ $- 0{,}02x_{28} + 0{,}03x_{29} + 0{,}03x_{30} + 0{,}07x_{31} + 0{,}16x_{32} + 0{,}16x_{33} + 0{,}8x_{34} - 0{,}63x_{35} +$ $+ 0{,}28x_{36} - 0{,}2x_{37} - 0{,}11x_{38} - 0{,}04x_{39} - 0{,}03x_{40} + 0{,}03x_{41} + 0{,}07x_{42} + 0{,}11x_{43} +$ $+ 0{,}3x_{44} - 0{,}05x_{45} + 0{,}04x_{46} - 0{,}03x_{47} - 0{,}17x_{48}$

Закінчення табл. 3.9

1	2
$\lambda_6 = 1{,}90854$	$F_6 = -0{,}34x_1 + 0{,}34x_2 - 0{,}12x_3 + 0{,}13x_4 - 0{,}36x_5 + 0{,}27x_6 + 0{,}25x_7 - 0{,}28x_8 -$ $- 0{,}05x_9 - 0{,}11x_{10} - 0{,}09x_{13} + 0{,}04x_{14} + 0{,}09x_{16} - 0{,}12x_{17} + 0{,}03x_{18} + 0{,}23x_{19} +$ $+ 0{,}07x_{20} - 0{,}06x_{21} + 0{,}14x_{22} + 0{,}08x_{23} - 0{,}1x_{24} + 0{,}1x_{25} + 0{,}27x_{26} - 0{,}04x_{27} +$ $+ 0{,}12x_{28} + 0{,}12x_{29} + 0{,}06x_{30} + 0{,}09x_{31} + 0{,}45x_{32} + 0{,}02x_{33} - 0{,}15x_{34} - 0{,}14x_{35} -$ $- 0{,}02x_{36} + 0{,}19x_{37} + 0{,}05x_{38} - 0{,}08x_{39} + 0{,}25x_{40} + 0{,}64x_{41} + 0{,}68x_{42} + 0{,}67x_{43} +$ $+ 0{,}3x_{44} + 0{,}21x_{45} + 0{,}04x_{46} + 0{,}11x_{47} - 0{,}3x_{48}$
$\lambda_7 = 1{,}83471$	$F_7 = 0{,}31x_1 - 0{,}31x_2 + 0{,}01x_3 - 0{,}003x_4 + 0{,}12x_5 - 0{,}35x_6 + 0{,}08x_7 + 0{,}23x_8 +$ $+ 0{,}04x_9 - 0{,}62x_{10} - 0{,}01x_{13} - 0{,}04x_{14} - 0{,}01x_{16} + 0{,}02x_{17} - 0{,}1x_{18} - 0{,}22x_{19} -$ $- 0{,}12x_{20} - 0{,}04x_{21} - 0{,}09x_{22} - 0{,}17x_{23} + 0{,}03x_{24} - 0{,}11x_{25} - 0{,}25x_{26} - 0{,}11x_{27} +$ $+ 0{,}05x_{28} - 0{,}02x_{29} - 0{,}09x_{30} - 0{,}11x_{31} - 0{,}05x_{32} + 0{,}07x_{33} + 0{,}002x_{34} + 0{,}32x_{35} -$ $- 0{,}04x_{36} + 0{,}72x_{37} - 0{,}07x_{38} + 0{,}2x_{39} - 0{,}02x_{40} + 0{,}15x_{41} + 0{,}07x_{42} - 0{,}07x_{43} -$ $- 0{,}09x_{44} - 0{,}04x_{45} + 0{,}08x_{46} - 0{,}02x_{47} + 0{,}67x_{48}$
$\lambda_8 = 1{,}60816$	$F_8 = 0{,}06x_1 - 0{,}06x_2 + 0{,}2x_3 - 0{,}2x_4 - 0{,}04x_5 + 0{,}05x_6 + 0{,}02x_7 - 0{,}04x_8 -$ $- 0{,}12x_9 + 0{,}01x_{10} - 0{,}03x_{13} - 0{,}03x_{14} + 0{,}05x_{16} - 0{,}18x_{17} - 0{,}07x_{18} + 0{,}07x_{19} +$ $+ 0{,}04x_{20} + 0{,}01x_{21} + 0{,}004x_{22} - 0{,}23x_{23} - 0{,}12x_{24} - 0{,}09x_{25} - 0{,}02x_{26} - 0{,}04x_{27} +$ $+ 0{,}18x_{28} + 0{,}17x_{29} + 0{,}17x_{30} + 0{,}14x_{31} + 0{,}29x_{32} - 0{,}06x_{33} + 0{,}22x_{34} - 0{,}15x_{35} +$ $+ 0{,}1x_{36} - 0{,}12x_{37} + 0{,}82x_{38} - 0{,}81x_{39} - 0{,}09x_{40} + 0{,}15x_{41} + 0{,}04x_{42} + 0{,}03x_{43} -$ $- 0{,}11x_{44} - 0{,}08x_{45} + 0{,}25x_{46} - 0{,}16x_{47} - 0{,}24x_{48}$
$\lambda_9 = 1{,}36621$	$F_9 = -0{,}02x_1 + 0{,}01x_2 + 0{,}05x_3 - 0{,}07x_4 - 0{,}11x_5 + 0{,}07x_6 + 0{,}08x_7 + 0{,}08x_8 -$ $- 0{,}13x_9 + 0{,}002x_{10} - 0{,}06x_{13} - 0{,}12x_{14} - 0{,}17x_{16} + 0{,}23x_{17} - 0{,}04x_{18} + 0{,}004x_{19} +$ $+ 0{,}02x_{20} + 0{,}03x_{21} + 0{,}02x_{22} - 0{,}05x_{23} - 0{,}03x_{24} - 0{,}05x_{25} - 0{,}17x_{26} + 0{,}78x_{27} +$ $+ 0{,}82x_{28} - 0{,}62x_{29} + 0{,}002x_{30} + 0{,}03x_{31} + 0{,}1x_{32} + 0{,}01x_{33} - 0{,}01x_{34} - 0{,}15x_{35} -$ $- 0{,}05x_{36} - 0{,}09x_{37} + 0{,}03x_{38} - 0{,}02x_{39} + 0{,}01x_{40} + 0{,}12x_{41} + 0{,}001x_{42} - 0{,}08x_{43} -$ $- 0{,}17x_{44} + 0{,}09x_{45} + 0{,}11x_{46} - 0{,}06x_{47} + 0{,}07x_{48}$
$\lambda_{10} = 1{,}275$	$F_{10} = 0{,}04x_1 - 0{,}04x_2 - 0{,}02x_3 + 0{,}07x_4 - 0{,}13x_5 + 0{,}3x_6 - 0{,}03x_7 + 0{,}03x_8 -$ $- 0{,}82x_9 + 0{,}27x_{10} + 0{,}13x_{13} - 0{,}03x_{14} - 0{,}04x_{16} - 0{,}08x_{17} + 0{,}44x_{18} + 0{,}13x_{19} +$ $+ 0{,}01x_{20} - 0{,}17x_{21} + 0{,}34x_{22} - 0{,}23x_{23} + 0{,}06x_{24} + 0{,}11x_{25} - 0{,}14x_{26} - 0{,}15x_{27} +$ $+ 0{,}13x_{28} - 0{,}25x_{29} + 0{,}01x_{30} - 0{,}04x_{31} - 0{,}08x_{32} + 0{,}02x_{33} + 0{,}02x_{34} + 0{,}04x_{35} +$ $+ 0{,}1x_{36} - 0{,}02x_{37} + 0{,}03x_{38} - 0{,}04x_{39} - 0{,}03x_{40} - 0{,}18x_{41} + 0{,}24x_{42} + 0{,}25x_{43} +$ $+ 0{,}57x_{44} - 0{,}12x_{45} + 0{,}04x_{46} + 0{,}43x_{47} + 0{,}1x_{48}$
$\lambda_{11} = 1{,}05674$	$F_{11} = -0{,}18x_1 + 0{,}18x_2 - 0{,}04x_3 + 0{,}07x_4 + 0{,}09x_5 - 0{,}01x_6 - 0{,}09x_7 + 0{,}25x_8 -$ $- 0{,}01x_9 + 0{,}08x_{10} + 0{,}1x_{13} + 0{,}44x_{14} + 0{,}15x_{16} - 0{,}28x_{17} + 0{,}01x_{18} + 0{,}08x_{19} -$ $- 0{,}003x_{20} - 0{,}08x_{21} + 0{,}31x_{22} + 0{,}22x_{23} + 0{,}03x_{24} + 0{,}34x_{25} + 0{,}44x_{26} - 0{,}25x_{27} +$ $+ 0{,}09x_{28} - 0{,}23x_{29} + 0{,}05x_{30} + 0{,}04x_{31} + 0{,}11x_{32} - 0{,}04x_{33} + 0{,}05x_{34} + 0{,}29x_{35} +$ $+ 0{,}11x_{36} + 0{,}06x_{37} + 0{,}004x_{38} + 0{,}01x_{39} + 0{,}02x_{40} + 0{,}16x_{41} - 0{,}02x_{42} - 0{,}06x_{43} +$ $+ 0{,}04x_{44} - 0{,}32x_{45} - 0{,}03x_{46} - 0{,}34x_{47} - 0{,}09x_{48}$

Для кожного головного фактора F множина значень a_{jr} умовно розбивається на чотири підмножини з нечіткими межами: W_1 – підмножина незначущих вагових коефіцієнтів; W_2 – підмножина значущих вагових коефіцієнтів; W_3 – підмножина значущих вагових коефіцієнтів, що не беруть участі у формуванні назви головного фактора; W_2-W_3 – підмножина значущих вагових коефіцієнтів, що беруть участь у формуванні назви головного фактора.

Додаткове виділення підмножини W_3 пояснюється прагненням до простішої структури головної компоненти, що завжди легко піддається інтерпретації. На своїх межах підмножина W_3 має критичні значення: в $a_{кр1}$ – максимальна кількість ознак, що пояснюють головну компоненту, в $a_{кр2}$ – мінімальна кількість пояснювальних ознак. У процесі ретельного аналізу критичні межі W_3 можуть встановлюватися за статистичним критерієм t, як у випадку перевірки значущості коефіцієнтів кореляції. Загальний склад множини вагових коефіцієнтів поданий на рис. 3.22.

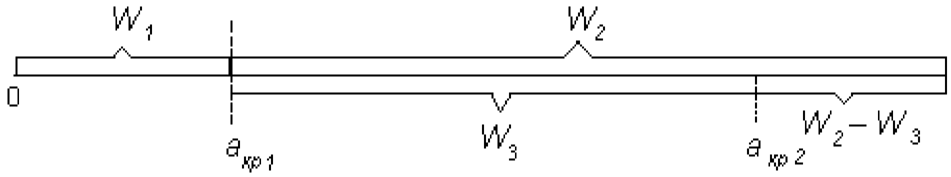


Рис. 3.22. **Склад множини вагових коефіцієнтів** $a_{jr},\ j = \overline{1, m}$ **для** r**-ї головної компоненти**

Підтвердження значущості ознак, що брали участь у формуванні назви головних факторів, визначають шляхом обчислення коефіцієнтів інформативності:

$$K_u = \frac{\sum_{j=1}^{n} a_{jr}^2 \left(W_2 - W_3\right)}{\sum_{j=1}^{n} a_{jr}^2}.$$

Набір пояснювальних ознак вважається задовільним, якщо значення K_u знаходяться у межах 0,75 – 0,95.

Граничні значення для підмножини W_3 будуть такими: $a_{кр1}$= 0,2 і $a_{кр2}$= = 0,4. За наведеними рекомендаціями були виявлені складні ознаки виробничо-господарської діяльності підприємств та визначені їх коефіцієнти інформативності, вони показані в табл. 3.10.

Таблиця 3.10

Обґрунтування назви загальних складних метричних ознак виробничо-господарської діяльності підприємств у Харківському регіоні

Складні ознаки, що характеризують виробничо-господарську діяльність промислових підприємств	Позначення елементарних ознак, що визначають складну ознаку	Коефіцієнт інформативності K_u
F_1' – забезпеченість підприємств оборотними засобами	$x_1, x_2, x_5, x_7, x_8, x_{13}, x_{14}, x_{21}, x_{24}, x_{25}, x_{33}$	0,824
F_2' – ефективність використання всього капіталу підприємств	$x_6, x_{30}, x_{31}, x_{32}, x_{36}, x_{40}, x_{46}, x_{47}$	0,795
F_3' – достатність та прибутковість власного капіталу підприємств	$x_{16}, x_{17}, x_{18}, x_{22}, x_{26}, x_{29}, x_{41}, x_{45}$	0,824
F_4' – покриття активів власними оборотними засобами	$x_{19}, x_{20}, x_{21}, x_{22}$	0,632
F_5' – достатність найліквідніших елементів оборотних активів	x_4, x_{34}, x_{35}	0,78
F_6' – прибутковість діяльності підприємств	$x_{32}, x_{41}, x_{42}, x_{43}$	0,55
F_7' – фактор технічного стану основних засобів	x_{10}, x_{37}, x_{48}	0,621
F_8' – своєчасність проведення розрахунків з дебіторами	x_{38}, x_{39}	0,646
F_9' – здатність підприємств отримувати дохід	x_{27}, x_{28}, x_{29}	0,842
F_{10}' – здатність підприємств генерувати прибуток від основної діяльності	x_9, x_{44}, x_{47}	0,507
F_{11}' – забезпеченість підприємств власними засобами	x_{14}, x_{26}	0,268

Для визначення однорідності сукупності об’єктів рекомендовано застосовувати кластерний аналіз за загальними складними ознаками СЕС, виміряними в кожному спостереженні. Для обчислень кластерного аналізу в ППП

Statgraphics потрібно активізувати меню Special/Multivariate Methods/Cluster Analysis.

У результаті чого маємо розшарування підприємств на групи за системою загальних складних ознак, що відображає дендрограма на рис. 3.23.

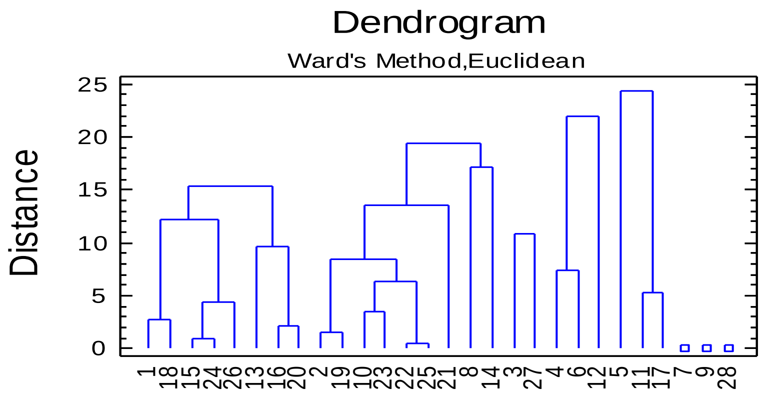


Рис. 3.23. **Дендрограма кластеризації 28 промислових підприємств Харківського регіону за критерієм загальних латентних складних ознак виробничо-господарської діяльності у 2003 році**

"Будь-яку модель, що рекомендується до практичного застосування, необхідно дослідити на стійкість відносно коливань передумов та початкових даних" [209, с. 3]. І далі – найрадикальнішим засобом перевірки стійкості розв'язків вважається побудова іншої моделі об'єкта. Якщо в новій моделі розв'язки виявляються близькими до розв'язків початкової, то найбільш вірогідно, що обидві вони достатньо адекватно відображають реальність, ніж у процесі розгляду тільки однієї моделі. На основі цієї думки був застосований дискримінантний аналіз для перевірки класифікації об'єктів за 8-ма групами, одержаної за допомогою кластерного аналізу в скороченому просторі загальних складних ознак. Для обчислень дискримінантного аналізу в ППП Statgraphics потрібно активізувати меню Special/Multivariate Methods/Discriminant Analysis. Для досліджуваних 8 груп (класів) розраховані дискримінантні функції, наведені в табл. 3.11; у ній же наведені характеристики перших 7 функцій.

Таблиця 3.11

Результати обчислення дискримінантних функцій (ДФ) і їх статистичний аналіз

№ ДФ	Узагаль-нені власні числа	Власні числа у процентах	Кореляційні відношення	Статистика Уїлкса	Критерій Пірсона	Число ступенів свободи	Рівень значущості α
1	106,134	61,03	0,99532	4,56328E-8	295,7962	77	0,0000
2	37,8139	21,74	0,98703	0,00000488	213,9998	60	0,0000
3	10,7827	6,20	0,95662	0,00018975	149,9712	45	0,0000
4	10,0363	5,77	0,95362	0,00223581	106,8051	32	0,0000
5	6,34244	3,65	0,92941	0,024675	64,7844	21	0,0000
6	1,86797	1,07	0,80704	0,181175	29,8952	12	0,0029
7	0,924548	0,53	0,69311	0,519602	11,4571	5	0,0430

Виокремлено сім (головних) дискримінантних функцій (ДФ), що впорядковані послідовно за значеннями узагальнених власних чисел. Ця послідовність збігається і з власними числами у відсотках від їх загальної суми та з даними в наступній колонці таблиці, де обчислені кореляційні відношення. При піднесенні кореляційних відношень до квадрату можна отримати індекси детермінації, які також підтверджують, що перша дискримінантна функція вагоміша для розділення сукупності об’єктів на класи. Статистики критерію Уїлкса демонструють відносний внесок перешкоди в загальну мінливість з урахуванням усіх функцій і при почерговому виключенні кожної наступної. За допомогою критерію Пірсона можна оцінити значущість статистик Уїлкса. Усі функції значущі, крім останньої, хоча для неї $\alpha > 0{,}01$, але її слід також вважати значущою, якщо прийняти рівень значущості $\alpha = 0{,}05$.

Таким чином, ознаковий 11-вимірний простір ознак скоротився до 7-вимірного, а можливо, і до 2-вимірного, оскільки перші 2 функції визначають загальний внесок (майже 83%). Коефіцієнти дискримінантних функцій обчислені у стандартизованій (табл. 3.12) і нестандартизованій формі.

Коефіцієнти в нестандартизованій формі несуть інформацію про абсолютний внесок даної складної ознаки в значення дискримінантної функції, а коефіцієнти в стандартизованій формі – про відносний внесок ознаки. За даними табл. 3.12 визначається перша дискримінантна функція від стандартизованих змінних (складних ознак):

$$L_1 = -2{,}50Z_{F_1} - 0{,}80Z_{F_2} - 0{,}52Z_{F_3} - 0{,}09Z_{F_4} + 0{,}73Z_{F_5} + 2{,}26Z_{F_6} + \\ + 0{,}17Z_{F_7} - 0{,}84Z_{F_8} + 0{,}39Z_{F_9} - 0{,}22Z_{F_{10}} - 0{,}09Z_{F_{11}}.$$

Таблиця 3.12

Значення стандартизованих β-коефіцієнтів дискримінантних функцій

Ознаки (фактори)	Стандартизовані β-коефіцієнти						
F1	-2,50037	-0,407685	1,08752	0,526164	0,152259	1,86148	0,40472
F2	-0,795044	1,23193	0,720436	-1,01413	-0,0530502	0,214707	-0,036864
F3	-0,518632	1,41863	0,181346	0,82298	-0,268686	-0,066616	0,248815
F4	-0,0897036	-0,932923	-0,510947	-0,0129002	0,0475044	-1,28796	-0,432817
F5	0,722934	-0,229954	0,154992	-0,840602	-0,165299	-0,114081	0,861639
F6	2,25938	0,582309	-1,06868	0,540365	0,159043	-1,18266	-0,195323
F7	0,170045	-0,512609	1,13815	0,273807	0,479308	-0,498877	-0,140826
F8	-0,841862	-0,967958	-0,157215	0,408536	0,979613	0,266041	-0,244616
F9	0,392519	0,846252	0,500343	0,308441	-1,53069	-0,149921	0,10317
F10	-0,201835	-1,27124	0,733416	-0,305875	0,360367	-0,375734	0,223119
F11	-0,0875112	0,785047	0,020938	-0,148257	-0,675002	-0,288981	-0,404306

Таким же чином можна записати і всі 7 дискримінантних функцій. Порівнюючи внески кожної змінної F у формування дискримінантних функцій, слід зазначити, що перша функція утворена завдяки F_1 та F_6; друга – F_2 і F_3, F_{10}; третя – F_1, F_6, F_7; четверта – F_2; п’ята – F_9 і F_8; шоста – F_1, F_4,F_6; сьома – можна вважати, що це є F_5. Таким чином, ознаки F_1 та F_6 є найважливішими у формуванні всіх 7 дискримінантних функцій, а отже, вони основні дискримінантні ознаки даної класифікації.

Таблиця 3.13

Коефіцієнти розкладу класифікаційних функцій

Змінні (фактори)	Групи об'єктів							
	1	2	3	4	5	6	7	8
F1	-4,1232	0,84100	9,1987	-1,3873	-2,0522	-6,14791	37,8466	-6,2356
F2	-2,6389	-0,7706	15,282	-1,5999	-5,9659	-6,58575	7,27225	0,34547
F3	-1,5873	0,00743	5,5177	-2,6632	-8,8959	-3,49504	2,16902	0,78797
F4	3,33387	6,99643	-8,368	15,1025	17,2596	5,84979	19,4026	1,10874
F5	1,16822	-1,0786	-1,749	0,43047	2,55944	1,56529	-7,5846	1,73652
F6	3,93159	-0,8955	-11,28	-0,6938	-1,5882	6,81424	-38,455	4,2031
F7	2,67939	-0,4132	-1,973	2,18639	6,30752	19,29	4,53485	1,39896
F8	-0,8482	2,47989	-0,032	3,20124	4,32736	0,160481	23,6132	-8,7713
F9	-0,0316	-1,7449	0,5804	-4,3748	-6,2920	0,320612	-12,331	11,969
F10	2,27373	1,07965	-2,507	6,8641	15,1353	11,7921	21,2033	-1,2756
F11	-1,3003	-0,0140	6,0165	-1,6775	-5,7874	-4,55526	-4,0508	5,64396
Const	-13,147	-13,010	-156,6	-35,182	-104,56	-156,187	-838,78	-79,641

Таким же чином можна записати і всі 7 дискримінантних функцій. Порівнюючи внески кожної змінної F у формування дискримінантних функцій, слід зазначити, що перша функція утворена завдяки F_1 та F_6; друга – F_2 і F_3, F_{10}; третя – F_1, F_6, F_7; четверта – F_2; п'ята – F_9 і F_8; шоста – F_1, F_4,F_6; сьома – можна вважати, що це є F_5. Таким чином, ознаки F_1 та F_6 є найважливішими у формуванні всіх 7 дискримінантних функцій, а отже, вони основні дискримінантні ознаки даної класифікації.

Далі визначені класифікаційні функції, коефіцієнти яких наведені в табл. 3.13. Так, класифікаційна функція для першої групи об'єктів має вигляд:

$$y_1 = -13{,}15 - 4{,}12F_1 - 2{,}64F_2 - 1{,}59F_3 + 3{,}33F_4 + 1{,}17F_5 + 3{,}93F_6 + 2{,}68F_7 - \\ - 0{,}85F_8 - 0{,}03F_9 + 2{,}27F_{10} - 1{,}3F_{11}.$$

Аналогічно записуються всі інші класифікаційні функції – інформанти. Знайдені функції дозволяють побудувати класифікаційну таблицю об'єктів, яка показує ступінь підтвердження того початкового розшарування в сукупності, що отримано за допомогою кластерного аналізу. Склад кожного кластера відомий.

Результат зіставлення складу груп, отриманих за кластерним аналізом (КА), зі складом груп, одержаних за дискримінтним аналізом (ДА), відмічений процентом співпадання (табл. 3.14).

Зміст табл. 3.14 переконливо свідчить, що наявна цілком стійка та достовірна класифікація об'єктів за системою загальних складних ознак, оскільки в кожній групі дискримінантний аналіз підтвердив склад груп на 100%. Отже, як безпосередня міра точності передбачення цей процентний зміст підтверджує правильність складання кластерів за методом Уорда. У результаті обчислень за допомогою дискримінантного аналізу отримано інформацію про окремі об'єкти, відмінності між класами, про здатність наявної системи загальних складних ознак як цілого точно відрізняти класи [142].

Можна продовжити аналіз вирішенням проблеми послідовного відбору дискримінантних ознак, але це доцільно робити, коли є система елементарних ознак об'єкта. У випадку коли дискримінантні ознаки – латентні складні ознаки, такий аналіз підтверджує значущість складних ознак, але з позиції їх відмінностей за об'єктами сукупності, тобто з метою класифікації [142].

Таблиця 3.14

Класифікаційна таблиця об'єктів

Групи за КА	Групи за ДА							
	1	2	3	4	5	6	7	8
1	8 100,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%
2	0 0,00%	9 100,0%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%
3	0 0,00%	0 0,00%	2 100%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%
4	0 0,00%	0 0,00%	0 0,00%	3 100,0%	0 0,00%	0 0,00%	0 0,00%	0 0,00%
5	0 0,00%	0 0,00%	0 0,00%	0 0,00%	3 100,0%	0 0,00%	0 0,00%	0 0,00%
6	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	1 100,0%	0 0,00%	0 0,00%
7	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	1 100,0%	0 0,00%
8	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	1 100,0%

Моделі загальних складних ознак діяльності підприємств доповнюються моделями систем складних ознак, які відображають взаємозв'язки між різними станами підприємства, наприклад, взаємозв'язок ресурсно-ринкового стану підприємств та результативності їх діяльності. Моделі систем складних ознак отримуємо за допомогою канонічного аналізу. Для цього слід активізувати меню Special/Multivariate Methods/Canonical Correlations. Маємо системи загальних ознак виробничо-господарської діяльності підприємств:

$r_{u_1 v_1} = 0{,}95$

$p = 0{,}00$

$$\begin{cases} u_1 = 0{,}56x_1 - 0{,}23x_2 + 0{,}14x_3 + 0{,}21x_4 - 0{,}12x_5 - 0{,}01x_6 + 0{,}01x_8 + 0{,}14x_9 + 0{,}07x_{10} + 0{,}23x_{13} + \\ + 0{,}09x_{14} - 0{,}12x_{16} - 0{,}01x_{17} - 0{,}11x_{18} - 0{,}06x_{19} + 0{,}06x_{20} - 0{,}09x_{21} + 1{,}03x_{22} - 0{,}03x_{23} + 0{,}46x_{24} - \\ - 0{,}09x_{25} + 0{,}06x_{26} + 0{,}04x_{27} - 0{,}18x_{30} + 0{,}39x_{31} + 0{,}12x_{32} - 0{,}42x_{33} - 0{,}06x_{34} + \\ + 0{,}03x_{35} - 0{,}21x_{36} - 0{,}04x_{37} - 0{,}05x_{38} + 0{,}01x_{39} + 0{,}14x_{46} - 0{,}1x_{48} \\ v_1 = 0{,}17x_{28} - 0{,}06x_{29} + 0{,}24x_{40} - 0{,}23x_{41} + 0{,}2x_{42} + 0{,}05x_{43} + 0{,}05x_{44} - 0{,}88x_{45} + 0{,}06x_{47} \end{cases}$$

u_1 – швидкість обігу власних коштів підприємств;

v_1 – дохідність власного капіталу;

$r_{u_2 v_2} = 0{,}92$

$p = 0{,}00$

$$\begin{cases} u_2 = -0{,}65x_1 - 0{,}89x_2 - 0{,}09x_3 + 0{,}13x_4 + 0{,}08x_5 + 0{,}09x_6 - 0{,}18x_8 - 0{,}16x_9 - 0{,}06x_{10} - 0{,}51x_{13} - \\ - 0{,}13x_{14} + 0{,}21x_{16} + 0{,}02x_{17} + 0{,}09x_{18} - 0{,}13x_{19} + 0{,}69x_{20} - 0{,}52x_{21} - 0{,}28x_{22} + 0{,}11x_{23} + \\ + 1{,}21x_{24} + 0{,}04x_{25} + 0{,}19x_{26} + 0{,}07x_{27} + 0{,}72x_{30} - 0{,}98x_{31} + 0{,}9x_{32} - 0{,}59x_{33} - 0{,}52x_{34} + \\ + 0{,}08x_{35} + 0{,}08x_{36} + 0{,}1x_{37} - 0{,}1x_{38} - 0{,}1x_{39} + 0{,}28x_{46} - 0{,}14x_{48} \\ v_2 = 0{,}21x_{28} + 0{,}19x_{29} + 0{,}53x_{40} + 0{,}15x_{41} - 0{,}08x_{42} + 0{,}51x_{43} + 0{,}03x_{44} - 0{,}05x_{45} - 0{,}06x_{47} \end{cases}$$

u_2 – швидкість обігу усього капіталу підприємства;

v_2 – рентабельність виробленої продукції;

$r_{u_3 v_3} = 0{,}77$

$p = 0{,}00$

$$\begin{cases} u_3 = 1{,}11x_1 + 1{,}36x_2 + 0{,}35x_3 + 0{,}04x_4 + 0{,}25x_5 + 0{,}09x_6 - 0{,}05x_8 + 0{,}14x_9 + 0{,}29x_{10} + 0{,}19x_{13} - \\ + 0{,}41x_{14} - 0{,}59x_{16} - 0{,}39x_{17} + 0{,}09x_{18} + 0{,}2x_{19} - 0{,}34x_{20} + 0{,}01x_{21} + 0{,}09x_{22} + 0{,}06x_{23} + \\ + 0{,}02x_{24} + 0{,}14x_{25} + 0{,}01x_{26} + +0{,}57x_{27} + 0{,}42x_{30} - 0{,}88x_{31} - 0{,}28x_{32} + 0{,}09x_{33} + 0{,}51x_{34} + \\ + 0{,}001x_{35} - 0{,}22x_{36} + 0{,}03x_{37} + 0{,}01x_{38} + 0{,}18x_{39} + 1{,}28x_{46} + 0{,}13x_{48} \\ v_3 = 0{,}92x_{28} + 0{,}29x_{29} + 0{,}09x_{40} - 0{,}18x_{41} - 0{,}29x_{42} + 0{,}51x_{43} + 0{,}24x_{44} + 0{,}27x_{45} + 0{,}29x_{47} \end{cases}$$

u_3 – структура капіталу підприємства;

v_3 – отриманий дохід;

$r_{u_4 v_4} = 0{,}73$

$p = 0{,}00$

$$\begin{cases} u_4 = 0{,}45x_1 + 0{,}22x_2 + 0{,}24x_3 + 0{,}28x_4 - 0{,}74x_5 - 0{,}35x_6 + 0{,}09x_8 - 0{,}38x_9 - 0{,}3x_{10} + 0{,}78x_{13} - \\ - 0{,}62x_{14} - 0{,}46x_{16} - 0{,}4x_{17} - 0{,}22x_{18} + 0{,}12x_{19} - 0{,}01x_{20} - 0{,}05x_{21} - 0{,}22x_{22} - 0{,}003x_{23} - \\ - 0{,}32x_{24} + 0{,}34x_{25} - 0{,}03x_{26} + 0{,}27x_{27} - 0{,}09x_{30} + 0{,}07x_{31} + 0{,}09x_{32} - 0{,}56x_{33} - 0{,}45x_{34} + \\ + 0{,}08x_{35} - 0{,}08x_{36} + 0{,}04x_{37} + 0{,}7x_{38} + 0{,}41x_{39} - 0{,}24x_{46} - 0{,}39x_{48} \\ v_4 = 0{,}47x_{28} + 0{,}13x_{29} - 0{,}74x_{40} - 0{,}31x_{41} + 0{,}81x_{42} + 0{,}31x_{43} - 0{,}17x_{44} + 0{,}04x_{45} - 0{,}02x_{47} \end{cases}$$

u_4 – ефективність використання матеріальних засобів та грошових коштів;

v_4 – прибутковість діяльності підприємств;

$r_{u_5 v_5} = 0,71$

$p = 0,00$

$$\begin{cases} u_5 = -8,87x_1 - 9,94x_2 - 0,17x_3 - 0,27x_4 - 0,41x_5 - 0,74x_6 + 0,001x_8 + 0,17x_9 + 0,16x_{10} + \\ + 0,22x_{13} + 0,93x_{14} + 0,47x_{16} + 0,2x_{17} - 0,31x_{18} - 0,02x_{19} - 0,1x_{20} + 0,2x_{21} - 0,17x_{22} - \\ - 0,1x_{23} - 0,8x_{24} + 0,004x_{25} + 0,21x_{26} + 0,07x_{27} + 0,76x_{30} - 1,16x_{31} + 0,62x_{32} + 0,15x_{33} - \\ - 0,17x_{34} - 0,11x_{35} - 0,44x_{36} - 0,13x_{37} + 0,14x_{38} + 0,14x_{39} + 0,53x_{46} - 0,3x_{48} \\ v_5 = 0,47x_{28} + 0,75x_{29} - 0,03x_{40} + 0,31x_{41} - 0,09x_{42} + 0,01x_{43} - 0,51x_{44} - 0,61x_{45} - 0,3x_{47} \end{cases}$$

u_5 – стан виробничого потенціалу підприємства;

v_5 – прибутковість основної діяльності підприємства;

$r_{u_6 v_6} = 0,59$

$p = 0,00$

$$\begin{cases} u_6 = -3,26x_1 - 3,96x_2 + 0,32x_3 - 0,12x_4 + 0,04x_5 + 0,36x_6 - 0,62x_8 + 0,16x_9 - 0,18x_{10} - \\ - 0,18x_{13} + 0,25x_{14} - 0,38x_{16} + 0,01x_{17} + 0,25x_{18} + 0,23x_{19} + 0,07x_{20} - 0,4x_{21} + 0,12x_{22} + \\ + 0,32x_{23} - 0,71x_{24} - 0,17x_{25} + 0,43x_{26} - 0,18x_{27} + 0,6x_{30} - 0,18x_{31} - 0,97x_{32} + 1,17x_{33} - \\ - 0,1x_{34} - 0,68x_{35} - 0,34x_{36} - 0,07x_{37} - 0,12x_{38} - 0,04x_{39} + 0,25x_{46} + 0,12x_{48} \\ v_6 = -0,13x_{28} - 0,01x_{29} + 0,04x_{40} - 0,67x_{41} - 0,52x_{42} + 1,32x_{43} - 0,93x_{44} - 0,04x_{45} + 0,41x_{47} \end{cases}$$

u_6 – платоспроможність підприємства та його майновий стан;

v_6 – рентабельність усього вкладеного капіталу;

$r_{u_7 v_7} = 0,56$

$p = 0,04$

$$\begin{cases} u_7 = -9,51x_1 - 9,92x_2 - 1,12x_3 - 0,66x_4 - 0,28x_5 - 0,22x_6 - 0,13x_8 - 0,15x_9 + 0,1x_{10} - 0,13x_{13} + \\ + 0,04x_{14} - 0,17x_{16} + 0,39x_{17} + 0,3x_{18} + 0,3x_{19} - 0,69x_{20} + 0,46x_{21} + 0,18x_{22} - 0,17x_{23} - \\ - 0,42x_{24} + 0,27x_{25} + +0,27x_{26} - 0,72x_{27} - 0,07x_{30} + 0,35x_{31} - 0,71x_{32} + 0,15x_{33} + 0,15x_{34} + \\ + 0,02x_{35} - 0,18x_{36} + 0,14x_{37} + 0,53x_{38} + 0,13x_{39} + 0,36x_{46} - 0,18x_{48} \\ v_7 = 0,02x_{28} + 0,82x_{29} - 0,59x_{40} + 0,37x_{41} - 0,21x_{42} - 0,33x_{43} + 0,53x_{44} - 0,36x_{45} + 0,63x_{47} \end{cases}$$

u_7 – достатність обігових коштів підприємств;

v_7 – здатність підприємств генерувати прибуток.

Наведені складні ознаки мають великий коефіцієнт канонічної кореляції, який дорівнює 0,95478. Структуру утворює пара $x_{22} - x_{45}$, тобто прибутковість власного капіталу повністю залежить від його маневрування. За величиною вагових коефіцієнтів перед змінними, що визначають елементарні ознаки, конкретизуємо сформовані складні ознаки, які характеризують два визначальних аспекти виробничо-господарської діяльності підприємств: u_1 – це швидкість обігу власних коштів підприємств; v_1 – дохідність власного капіталу.

Результати обчислень статистичних показників у розбудованих моделях систем складних ознак наведено в табл. 3.15. Перші сім пар канонічних функцій виокремлені за наявним тісним взаємозв'язком, про що свідчать значення коефіцієнтів канонічної кореляції в табл. 3.15, а також підтверджується їх істотність за критерієм Пірсона.

Рівень величини взаємозв'язку елементарної та складної ознак свідчить про рівень корегованості чи керованості ознак у наявній структурі ознак об'єкта. За допомогою канонічних кореляцій можна спостерігати деформації в структурах усіх ознак і елементарних, і складних СЕС. Для цього слід застосувати даний математичний метод на множині значень показників, що виражають елементарні ознаки в динаміці.

Таблиця 3.15

Статистичні характеристики канонічних кореляцій

№ пар складних ознак	Власні числа	Коефіцієнти канонічних кореляцій	Статистика Уїлкса	Критерій Пірсона	Число ступенів свободи	Рівень значущості α
1	0,911604	0,95478	0,0003138	939,761	315	0,0000
2	0,848407	0,92109	0,0035504	657,14	272	0,0000
3	0,598735	0,77378	0,0234209	437,356	231	0,0000
4	0,531898	0,739313	0,0583676	330,976	192	0,0000
5	0,501735	0,708332	0,12469	242,544	155	0,0000
6	0,346449	0,588599	0,250248	161,388	120	0,0070
7	0,309634	0,556448	0,382905	111,836	87	0,0377
8	0,281878	0,530922	0,55464	68,6693	56	0,1192
9	0,227652	0,477129	0,772348	30,0943	27	0,3099

У результаті обчислених моделей за 1999 – 2002 рр., узагальнення містяться в табл. 3.16. Системи складних ознак слід формувати за рівнем значущості взаємозв'язків, не більшим ніж 0,05.

За змістом табл. 3.16 можна зробити висновок, що результативність виробничо-господарської діяльності підприємств даної сукупності усталено (протягом п'яти років) визначається дохідністю власного капіталу та рентабельністю виробленої продукції, які залежать від оборотності власних коштів підприємств, мобільності наявних ресурсів. Зміна стану систем складних ознак обумовлюється зміною рівнів величин елементарних ознак, а отже, підприємствам потрібно оперативно, тактично та стратегічно планувати свою діяльність, що має виражатися в рівнях показників, які відображають елементарні ознаки основних факторів діяльності.

Якщо закон розподілу значень величини ознаки несиметричний, потрібно детально дослідити величину; рекомендуємо в даному випадку розв'язати задачу розділення суміші. Задача розділення суміші завжди була непростою математичною задачею. Рекомендується її розв'язувати наступним чином.

Таблиця 3.16

Зміст змодельованих систем складних ознак виробничо-господарської діяльності підприємств, періодично фіксованих в часі

Ознаки, що характеризують ресурсно–ринковий стан підприємств (складні ознаки U)	Ознаки, що характеризують результативність діяльності підприємств (складні ознаки V)
1	2
1999 – 2000 рр.	
u_1 – мобільність ресурсів підприємств; u_2 – структура капіталу підприємств; u_3 – стан, достатність і оборотність активів; u_4 – забезпеченість підприємств власними коштами	v_1 – отриманий дохід; v_2 – прибутковість основної діяльності підприємств; v_3 – отриманий дохід; v_4 – прибутковість діяльності
1999 – 2001 рр.	
u_1 – ліквідність активів підприємств; u_2 – мобільність ресурсів підприємств; u_3 – структура виробничого потенціалу та його гнучкість; u_4 – стан оборотних засобів та швидкість їх руху	v_1 – прибутковість коштів, вкладених в активи; v_2 – загальна рентабельність підприємств; v_3 – здатність підприємств генерувати чистий прибуток; v_4 – прибутковість діяльності

1	2
1999 – 2002 рр.	
u_1 – оборотність коштів підприємств; u_2 – мобільність ресурсів підприємств; u_3 – структура капіталу підприємств; u_4 – ефективність використання обігових коштів	v_1 – віддача активів та капіталу; v_2 – рентабельність виробленої продукції; v_3 – отриманий дохід; v_4 – прибутковість діяльності підприємств
1999 – 2003 рр.	
u_1 – швидкість обігу власних коштів підприємств; u_2 – швидкість обігу всього капіталу підприємств; u_3 – структура капіталу підприємств; u_4 – ефективність використання матеріальних засобів та грошових коштів; u_5 – стан виробничого потенціалу підприємств; u_6 – платоспроможність підприємств та їх майновий стан; u_7 – стан і достатність обігових коштів	v_1 – дохідність власного капіталу; v_2 – рентабельність виробленої продукції; v_3 – отриманий дохід; v_4 – прибутковість діяльності підприємств; v_5 – прибутковість основної діяльності підприємств; v_6 – рентабельність усього вкладеного капіталу; v_7 – здатність підприємств генерувати прибуток

Задача розділення суміші. Розв’язання задачі за К. Пірсоном. Ідея про подання асиметричних розподілів суміші двома нормальними законами була реалізована К. Пірсоном ще в 1894 році [333]. Невідомими параметрами, які слід визначити, тут є ймовірності p_1, p_2, з якими змішуються нормально розподілені компоненти, математичні сподівання a_1, a_2 і середні квадратичні відхилення σ_1, σ_2 окремих розподілів – усього 5 параметрів, оскільки $p_1 + p_2 = 1$.

Диференціальна та інтегральна функції розподілу суміші з двох компонент виражаються у вигляді лінійної комбінації нормальних розподілів:

$$f(x) = p_1 \cdot f_1(x) + p_2 \cdot f_2(x), \tag{3.1}$$

$$F(x) = p_1 \cdot F_1(x) + p_2 \cdot F_2(x), \tag{3.2}$$

де $f_i(x) = \dfrac{1}{\sqrt{2\pi} \cdot \sigma_i} \exp\left\{ -\dfrac{(x - a_i)^2}{2 \cdot \sigma_i^2} \right\}$; $F_i(x) = \int\limits_{-\infty}^{x} f_i(t) dt$.

Одним із методів визначення параметрів окремих розподілів є метод моментів – теоретичні моменти розподілу дорівнюють відповідним вибірковим моментам m_1, m_2, m_3, m_4 і отримується система нелінійних алгебраїчних рівнянь відносно параметрів. Для того щоб не використовувати моменти вище 4-го порядку, К. Пірсон прийняв додаткове припущення про рівність дисперсій двох підсукупностей: $\sigma = \sigma_1 = \sigma_2$. При цьому додатковому обмеженні нелінійна система 5-ти рівнянь може бути розв'язана (і тільки в цьому значенні потрібно розуміти твердження К. Пірсона про "однозначне" поновлення компонент за розподілом суміші):

$$\begin{cases} 1 = p_1 + p_2, \\ m_1 = p_1 a_1 + p_2 a_2, \\ m_2 = p_1 a_1^2 + p_2 a_2^2 + \sigma^2, \\ m_3 = p_1 a_1^3 + p_2 a_2^3 + 3\sigma^2 (p_1 a_1 + p_2 a_2), \\ m_4 = p_1 a_1^4 + p_2 a_2^4 + 6\sigma^2 (p_1 a_1^2 + p_2 a_2^2) + 3\sigma^4. \end{cases} \tag{3.3}$$

У результаті послідовності перетворень наведеної нелінійної системи випливає, що проблема зводиться до вирішення бікубічного рівняння відносно дисперсії σ^2:

$$\begin{aligned} 2m_1\sigma^6 - 6m_1 (m_2 - m_1^2)\sigma^4 + (3m_2^2 m_1 - 4m_3 m_1^2 + m_4 m_1)\sigma^2 + \\ + (m_3^2 m_1 + m_2^3 m_1 - 2m_3 m_2 m_1^2 - m_4 m_2 m_1 + m_4 m_1^3) = 0. \end{aligned} \tag{3.4}$$

Доводиться, що знаки коефіцієнтів цього рівняння чергуються, отже, від'ємних коренів немає; відносно дисперсії σ^2 рівняння має один чи три дійсних корені, з яких потрібно брати найменший.

Після того як знайдена оцінка σ, обчислюються добуток і сума параметрів розподілів:

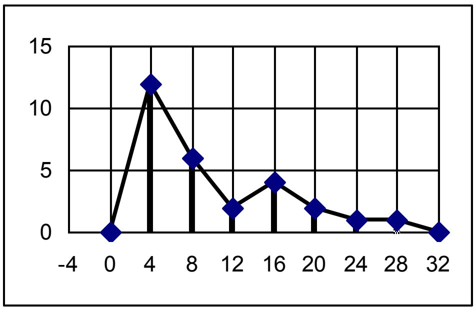
$$\begin{gathered} a_1 a_2 = \frac{m_3 m_1 - 3\sigma^2 m_1^2 - m_2^2 + 2\sigma^2 m_2 - \sigma^4}{m_2 - m_1^2 - \sigma^2}, \\ a_1 + a_2 = \frac{m_2 - \sigma^2 + a_1 a_2}{m_1}, \end{gathered} \tag{3.5}$$

і далі вже відносно просто остаточно обчислюються a_1, a_2, p_1, p_2.

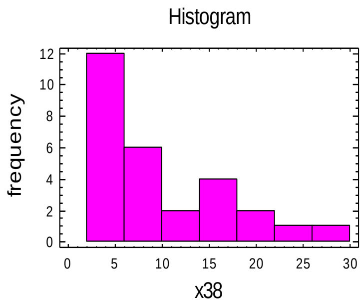
Отже, за значеннями 4-х вибіркових моментів m_1, m_2, m_3, m_4 суміші можна встановити параметри двох компонент, що мають нормальні розподіли. Але слід мати на увазі, що моменти високого порядку чутливі до викидів (до наявності великих відхилень у вибірці) і це відбивається на усталеності оцінок параметрів, нелінійна система може мати декілька різних розв'язків і потрібні додаткові критерії для вибору найкращого з них; наприкінці не слід забувати, що в апроксимації К. Пірсона постульована рівність дисперсій окремих розподілів.

Методику К. Пірсона розділення суміші двох нормальних розподілів з однаковою дисперсією продемонструємо на дослідженні величини ознаки x_{38} (оборотність дебіторської заборгованості підприємства), яка має явно не симетричний закон розподілу. На рис. 3.24 подані полігон і гістограма відносних частот (ймовірності).

Спостерігається явна асиметрія в розподілі, яка пояснюється наявністю суміші двох нормальних компонент.



а



б

Рис. 3.24. **Полігон (а) і гістограма (б) відносних частот значень величини ознаки x_{38} у 2003 році за всією досліджуваною сукупністю підприємств**

Формула для обчислення перших чотирьох моментів розподілу має такий вигляд:

$$m_k = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^k . \qquad (3.6)$$

Було отримане $m_1 = 9{,}989368, m_2 = 154{,}8052, m_3 = 3080{,}182, m_4 = 69\,989{,}51$.

Складене бікубічне рівняння відносно σ:

$$\begin{aligned} 2m_1\sigma^6 - 6m_1\left(m_2 - m_1^2\right)\sigma^4 + \left(3m_2^2 m_1 - 4m_3 m_1^2 + m_4 m_1\right)\sigma^2 + \\ + \left(m_3^2 m_1 + m_2^3 m_1 - 2m_3 m_2 m_1^2 - m_4 m_2 m_1 + m_4 m_1^3\right) = 0, \end{aligned} \qquad (3.7)$$

яке можна переписати як кубічне рівняння відносно $D = \sigma^2$:

$$c_0 \cdot D^3 + c_1 \cdot D^2 + c_2 \cdot D + c_3 = 0,$$

де $c_0 = 19{,}97874, c_1 = -3297{,}56, c_2 = 187871{,}8, c_3 = -1795235$.

Таким чином, $y(D)=19{,}9787D^3-3297{,}56D^2+187871D-1795235=0$. Знайдемо найменший корінь $y(1{,}8395)=-1{,}26$. Звідси $D=\sigma^2=11{,}8395$, $\sigma=3{,}4408$. Далі обчислено добуток та суму математичних сподівань для двох вибірок:

$$a_1a_2=\frac{m_3m_1-3\sigma^2m_1^2-m_2^2+2\sigma^2m_2-\sigma^4}{m_2-m_1^2-\sigma^2}=157{,}1527\,;$$

$$a_1+a_2=\frac{m_2-\sigma^2+a_1a_2}{m_1}=30{,}04379.$$

З цього випливає $a_1=6{,}7451$, $a_2=23{,}2986$. Тепер слід визначити процентний склад вибірок зі співвідношення: $\begin{cases}1=p_1+p_2\,,\\ m_1=p_1a_1+p_2a_2\,,\end{cases}$ $p_1=0{,}8040$, $p_2=0{,}1960$.

Якість отриманої апроксимації перевірена за критерієм К. Пірсона (зміст табл. 3.17).

Таблиця 3.17

Обчислення за методикою К. Пірсона

Інтервали		Центри інтер-валів	Час-тоти	Функція для a_1,σ	Функція для a_2,σ	Частоти $\Delta F_1\cdot n$	Частоти $\Delta F_2\cdot n$	Теоре-тичні частоти	Крите-рій
x_p	x_k	x	m	F_1	F_2	m_1	m_2	m_{teor}	χ^2
-6	-2	-4	–	0,000106	–	0,121834	–	–	–
-2	2	0	0	0,005518	–	1,765434	–	–	–
2	6	4	12	0,083938	3,02193E-10	7,436718	1,36454E-06	9,323988	0,768023
6	10	8	6	0,414276	2,48963E-07	9,311928	0,000300365	9,312232	1,178115
10	14	12	2	0,827911	5,55836E-05	3.480256	0,01858256	3,498839	0,642075
14	18	16	4	0,982504	0,00344189	0,381816	0,32018920	0,702006	15,49385
18	22	20	2	0,999464	0,06179012	0,011962	1,59765584	1,609618	0,09468
22	26	24	1	0,999995	0,35293169	0,000104	2,36440162	2,364506	0,673893
26	30	28	1	1	0,78379770	2,47E-07	1,04521488	1,18129	–
30	34	32	0	1	0,97426770	–	0,13607527	–	–
34	38	36	–	1	0,99906476	–	–	–	–
–	–	$\sum$	28	–	$\sum$	22,51005	5,48242440	27,99248	18,85063

Одержано значення $\chi^2=18{,}85$. На рис. 3.25. наведені порівняльні полігони частот при групуванні даних на 5 інтервалах (жирною лінією з квадратними маркерами показаний полігон для вихідних даних).

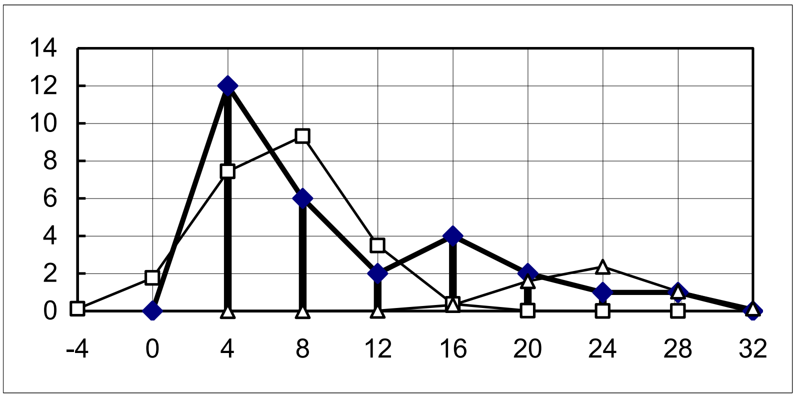


Рис. 3.25. **Полігони частот значень величини ознаки x_{38} вихідних даних та для суміші двох нормальних розподілів за методикою К. Пірсона**

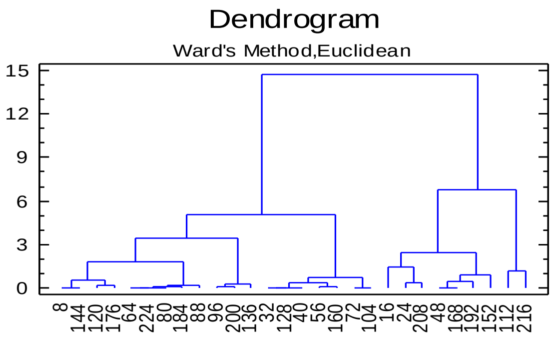
($\chi^2 = 18{,}85$)

$$(p_1 = 0{,}804,\ a_1 = 6{,}745,\ \sigma_1 = 3{,}441),\ (p_2 = 0{,}196,\ a_2 = 23{,}299,\ \sigma_2 = 3{,}441)$$

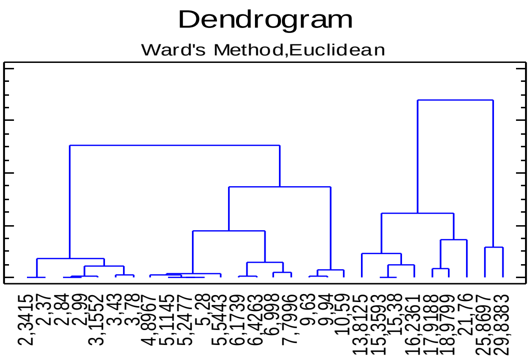
Розв’язування задачі за методикою К. Пірсона є громіздким і не дуже якісним.

Рекомендоване розв’язування задачі розділення суміші. Пропонується використовувати на першому етапі кластерний аналіз, який дозволяє за даними суміші визначити кількість компонент та їх склад. Основою цієї пропозиції є факт, що протягом останніх десятиліть у кластерному аналізі були досягнуті істотні успіхи, і деякі сучасні різновиди аналізу зараз заслуговують на довіру щодо об’єктивності класифікації (про це достатньо було сказано в розділі 2). Однією з таких надійних процедур вважається метод Уорда.

Потрібно виконати кластерний аналіз для отримання кількості і складу однорідних підсукупностей. На рис. 3.26 наведені результати обчислень процедури Уорда у вигляді дендрограми. Наочно спостерігається існування двох основних класів. У дендрограмі на рис. 3.26б вже враховано, що кластерів 2, та замість номерів об’єктів на горизонтальній осі проставлені значення показника x_{38}, тому що тут є одна ознака.



а



б

Рис. 3.26. **Дендрограми розшарування значень величини ознаки** x_{38}

Звідси видно, що величина ознаки в 1-й групі $x_{38} < 11$, а в 2-й – $x_{38} \geq 11$.

Характеристики обох кластерів теж обчислюються за допомогою пакета Statgraphics *Plus* V5.1 International Professional. У табл. 3.18 подані статистичні характеристики цих двох вибірок.

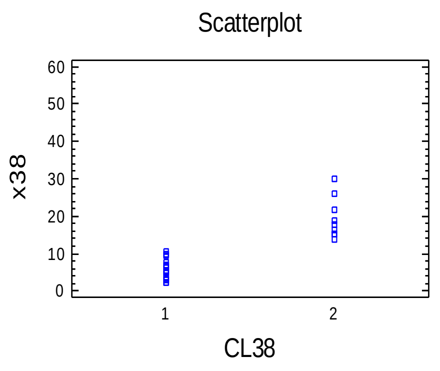
Таблиця 3.18

Характеристики двох кластерів (вибірок)

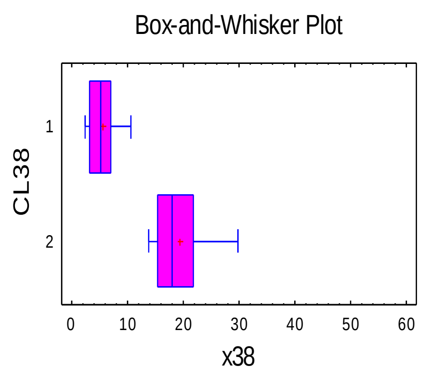
Характеристики	1-й кластер	2-й кластер
Кількість	19	9
Мінімальне значення	2,3415	13,8125
Максимальне значення	10,59	29,8384
Середнє	5,50251	19,4616
Медіана	5,2477	17,9188
Середнє квадратичне відхилення	2,56081	5,39026
Нижній квартиль	3,1552	15,38
Верхній квартиль	6,998	21,76
Міжквартильний розмах	3,8428	6,38

До першої вибірки (1-й кластер) потрапили всі спостереження, що мають значення величини менші ніж $x_0 = 11$, решта значень віднесені до другої вибірки.

На рис. 3.27а графічно зображений розкид значень величини за двома вибірками, який демонструє радикальні їх відмінності. На рис. 3.27б зображена блочна діаграма, що містить п’ять базових характеристик одновимірного набору величини ознаки для кожної вибірки.



а



б

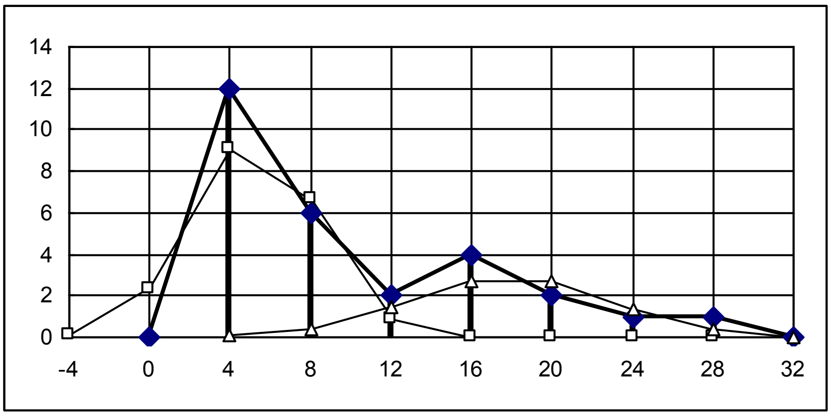
Рис. 3.27. **Графіки розкиду значень величини ознаки за двома вибірками у вигляді стовпчиків (а) і блочна діаграма одновимірного набору даних для кожної з двох вибірок (б)**

Другий етап запропонованої процедури – визначення параметрів двох підсукупностей. З табл. 3.18 випливає, що $p_1 = \dfrac{19}{28} = 0{,}679$ та $p_2 = \dfrac{9}{28} = 0{,}321$.

Слід вважати, що вибіркові середні та середні квадратичні відхилення є зміщеними оцінками для відповідних параметрів. Справа в тому, що в першій вибірці відкинуті максимальні значення величини (якщо вони перевершують x_0, то помилково віднесені до другого кластера); з іншого боку, до першої вибірки включена певна кількість даних з другої підсукупності, або вони були меншими ніж x_0. Така ж ситуація спостерігається і в другій вибірці. Отже, обидва вибіркових розподіли є скороченими і їх характеристики ($a_1 = 5{,}248$, $\sigma_1 = \dfrac{3{,}84}{1{,}35} = 2{,}844$, $a_2 = 17{,}919$, $\sigma_2 = \dfrac{6{,}38}{1{,}35} = 4{,}726$) є зміщеними оцінками параметрів нормальних розподілів.

Усталенішими оцінками до наявності спостережень з великими відхиленнями (чи до їх викидів) є медіана (як оцінка центра розподілу $Me_1 = 5{,}2477$, $Me_2 = 17{,}9188$) і міжквартильний розмах для оцінки мінливості $x_{0{,}25} = 3{,}8428$ і $x_{0{,}75} = 6{,}38$. В інтервалі між верхнім і нижнім квартилем потрапляє 50% спостережень; для нормального закону ширина цього інтервалу дорівнює $1{,}35\sigma$, звідки отримуються усталеніші оцінки $\sigma_1 = \dfrac{3{,}84}{1{,}35} = 2{,}844$; $\sigma_2 = \dfrac{6{,}38}{1{,}35} = 4{,}726$.

На рис. 3.28 наведено полігон суміші двох нормальних розподілів з усталеними оцінками характеристик, що засновані на квантилях ($a_1 = 5{,}248$, $\sigma_1 = 2{,}844$, $a_2 = 17{,}919$, $\sigma_2 = 4{,}726$). Як видно з цього графіка, відповідність дуже добра за критерієм узгодженості $\chi^2 = 1{,}04$, значення якого на порядок менше, ніж за методикою К. Пірсона ($\chi^2 = 18{,}85$). Порівняння рис. 3.24 та рис. 3.28 підтверджує висновок, що розв’язання К. Пірсона потребує значних великих витрат праці, воно нестійке до викидів і значно гірше (χ^2 за методикою К. Пірсона є більше майже в 15 разів). Слід зазначити, що цей приклад показує, що декларованої К. Пірсоном однозначності рішення немає, крім того, в методиці К. Пірсона ньому прийняте обтяжливе припущення про рівність дисперсій компонент.



(p_1 = 0,679, a_1 = 5,248, σ_1 = 2,844); (p_2 = 0,321, a_2 = 17,919, σ_2 = 4,726);

Рис. 3.28. **Полігони частот для даних і для суміші двох нормальних розподілів, χ^2 = 1,04**

Застосування кластерного аналізу дозволяє просто (за наявності відповідного програмного забезпечення) визначити кількість компонент (вона може бути більшою від двох), їх склад та характеристики нормально розподілених підсукупностей, причому цей розв'язок стійкий до наявності викидів.

Рекомендується використання даної задачі у вирішенні проблеми визначення етапів процесів розвитку об'єктів в економіці, а саме визначення моменту, коли зароджується новий ступінь розвитку, а саме новий етап розвитку одновимірного процесу та рівень його прогресивності [156]. Так, процес оборотності дебіторської заборгованості підприємств у даній сукупності демонстраційного прикладу має явно два розподіли, причому середнє однієї вибірки більше від середнього другої, тому одержане обґрунтоване розділення підприємств за даною ознакою.

Виконані практичні перевірки підтверджують припущення, що ґрунтується на теоретичних уявленнях про граничний закон розподілу випадкових величин, отже, будемо вважати, що для побудови якісної шкали потрібно враховувати закон розподілу ознаки з параметрами, що дорівнюють значенням характеристик досліджуваної сукупності.

Таким чином, розбудовані моделі складних метричних ознак 28 промислових підприємств Харківського регіону описують економічний стан промисловості в регіоні і є аналітичною основою економічного аналізу. На основі даних моделей визначаються конкретні напрями ухвалення управлінських рішень для подальших змін у функціонуванні та розвитку окремих підприємств і всього регіону.

Задача 3.2. Провести багатовимірний аналіз даних для формування й ухвалення управлінських рішень в управлінні регіонами на основі ідентифікації складних ознак соціально-економічного стану 24 областей України за 18 елементарними ознаками, які описують дані системи (зміст та значення соціально-економічних показників взяті зі збірника [257]).

Розв'язання задачі. Позначимо x_1 – територія (тис. км^2); x_1 – кількість наявного населення (тис. осіб); x_3 – кількість зайнятих економічною діяльністю (тис. осіб); x_4 – наявні доходи населення у розрахунку на одну особу (грн); x_5 – витрати населення у розрахунку на одну особу (грн); x_6 – середньомісячна заробітна плата (грн); x_7 – індекс споживчих цін (до грудня попереднього року, %); x_8 – валова додана вартість (у фактичних цінах, млн грн); x_9 – основні засоби (у фактичних цінах, на кінець року, млн грн); x_{10} – обсяг реалізованої промислової продукції, робіт, послуг (у фактичних цінах, млн грн); x_{11} – продукція сільського господарства (у порівнянних цінах, млн грн); x_{12}– введення в експлуатацію загальної площі житла (тис. м^2); x_{13} – роздрібний товарообіг підприємств (у фактичних цінах, млрд грн); x_{14} – обсяг вироблених послуг (у фактичних цінах, млн грн); x_{15} – експорт товарів і послуг (млрд дол. США); x_{16} – імпорт товарів і послуг (млрд дол. США); x_{17} – фінансовий результат (сальдо) від звичайної діяльності до оподаткування (млрд грн); x_{18} – інвестиції в основний капітал (у фактичних цінах, млрд грн). Таким чином, маємо соціально-економічні системи макрорівня, що елементарно описуються в просторі метричних ознак. Визначення складних соціально-економічних ознак здійснимо за допомогою розбудови базисної моделі І типу.

Згідно з моделлю, спочатку слід визначити загальні складні соціально-економічні ознаки, що описують області, за допомогою факторного аналізу. У табл. 3.19 наведені факторні навантаження після процедури $VARIMAX$. За значеннями коефіцієнтів факторних навантажень a_{ij} на i-ті показники встановлені складні соціально-економічні ознаки розвитку областей за пріоритетністю, а саме: перша загальна ознака – фінансова результативність операційної діяльності підприємств в областях (F_1), друга – географічне розміщення регіону (F_2), третя – ступінь сприятливості життєвих умов (F_3).

Таблиця 3.19

Факторна матриця після обертання *VARIMAX*
(за даними сукупності областей України)

F	F_1	F_2	F_3	R^2
x_1	0,130011	0,88021	0,0408273	0,793339
x_2	0,876834	0,34725	0,224001	0,939597
x_3	0,877279	0,375815	0,201302	0,951378
x_4	0,594228	0,711426	-0,150779	0,881969
x_5	0,396598	0,524399	0,405245	0,596508
x_6	0,835779	0,393418	-0,123033	0,868442
x_7	0,117958	0,0790707	-0,866056	0,77022
x_8	0,91449	0,354762	0,0770841	0,968089
x_9	0,815388	0,0131021	0,133149	0,682758
x_{10}	0,936127	0,253519	-0,0949882	0,949628
x_{11}	0,318809	0,825225	0,125582	0,798406
x_{12}	0,288237	0,369778	0,500902	0,470719
x_{13}	0,812347	0,385886	0,403093	0,9713
x_{14}	0,707174	0,380133	0,434812	0,833657
x_{15}	0,946401	0,185609	-0,0929652	0,938768
x_{16}	0,848552	0,392645	0,220896	0,923005
x_{17}	0,904294	0,270609	0,0762984	0,896798
x_{18}	0,846654	0,464312	0,172575	0,96219

За виявленими трьома загальними складними соціально-економічними ознаками країн встановимо однорідність даної сукупності соціально-економічних систем. На рис. 3.29 наведена дендрограма сукупності областей України, отримана завдяки реалізації кластерного аналізу за методом Уорда. Візуальне подання свідчить, що за даною системою складних соціально-економічних ознак потрібно розглядати 5 кластерів областей України.

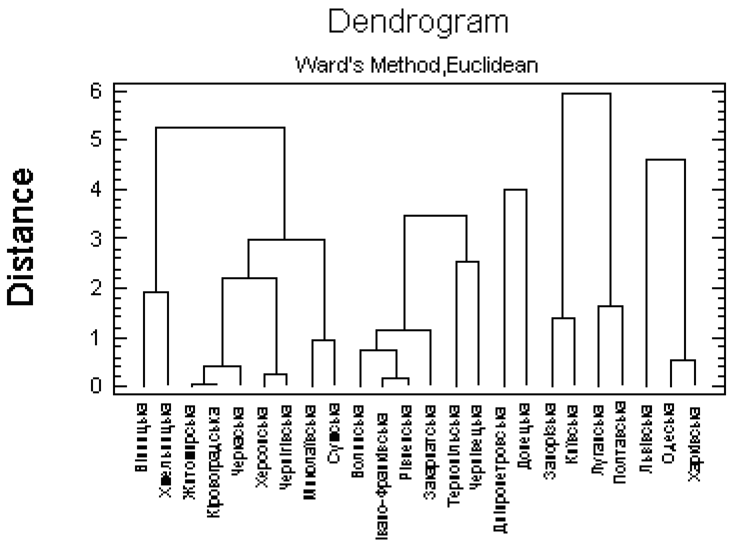


Рис. 3.29. **Дендрограма сукупності 24 областей України за їх загальними складними соціально-економічними ознаками**

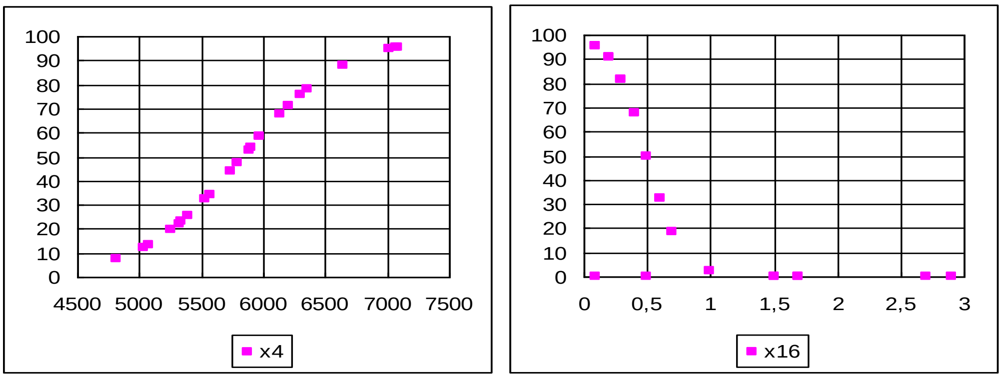
У табл. 3.20 наведений склад кожного кластера.

Таблиця 3.20

Склад кластерів однорідних областей України за критерієм їх загальних складних соціально-економічних ознак

Область	Кластер	Область	Кластер	Область	Кластер
Вінницька	1	Чернігівська	1	Донецька	3
Житомирська	1	Волинська	2	Запорізька	4
Кіровоградська	1	Закарпатська	2	Київська	4
Миколаївська	1	Івано-Франківська	2	Луганська	4
Сумська	1	Рівненська	2	Полтавська	4
Херсонська	1	Тернопільська	2	Львівська	5
Хмельницька	1	Чернівецька	2	Одеська	5
Черкаська	1	Дніпропетровська	3	Харківська	5

Далі визначимо значення загальних складних соціально-економічних ознак кожної з областей. Для цього потрібно перетворити значення 18 елементарних соціально-економічних ознак за функціями перетворення й обчислити вимірники кожної з областей. Як приклад на рис. 3.30 показані функції перетворення ознак x_4 і x_{16}.



а б

Рис. 3.30. **Функції перетворення показників x_4 – наявних доходів населення у розрахунку на одну особу (грн) (а) й x_{16} – імпорту товарів і послуг (млрд дол. США) (б)**

Були обчислені значення всіх 18 вимірників, що визначають соціально-економічні ознаки стану розвитку кожної області України. За сформованими окремими функціями перетворення соціально-економічних ознак всіх областей були побудовані графіки динаміки значень складних соціально-економічних ознак стану областей (рис. 3.31).

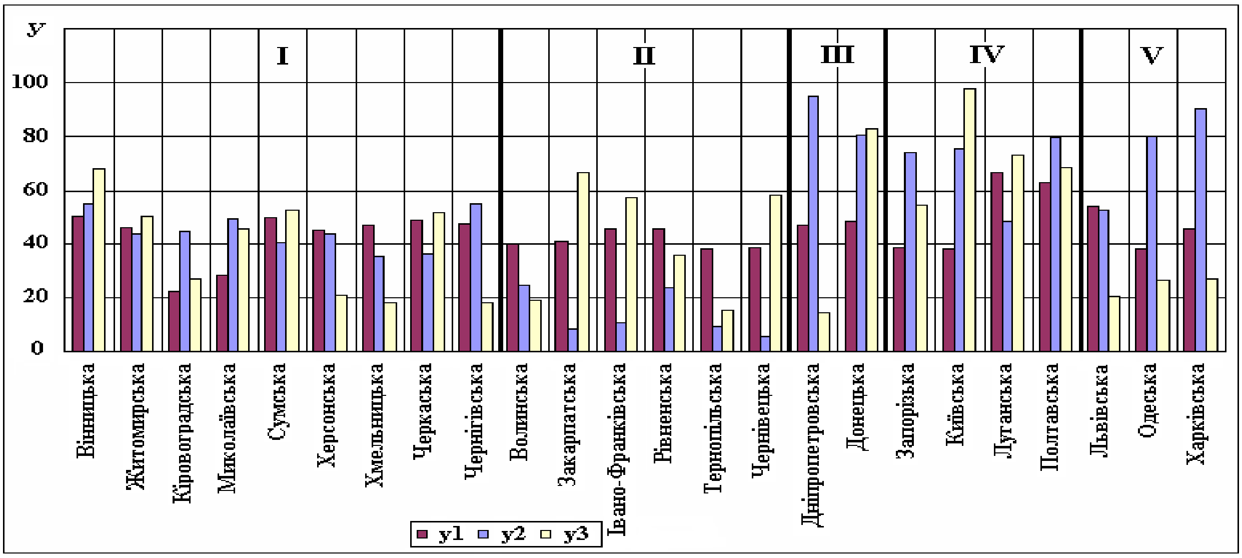


Рис. 3.31. **Діаграма соціально-економічного стану областей України за 5 кластерами за складними ознаками**

Тут складні ознаки: y_1 – фінансова результативність операційної діяльності підприємств в областях; y_2 – географічне розміщення регіону; y_3 – ступінь сприятливості життєвих умов.

Наочне зображення свідчить про низький рівень складних соціально-економічних ознак областей першого та другого кластерів. Три ж кластери (третій, четвертий та п’ятий) мають достатні рівні значень даних складних ознак, але вони різні за змістом. Так, третій кластер (Дніпропетровська та Донецька області) відзначається високим рівнем першої складної ознаки y_1 – фінансової результативності операційної діяльності підприємств в областях, четвертий кластер (Запорізька, Київська, Луганська та Полтавська області) мають достатньо високий рівень y_3 – сприятливості життєвих умов. Області, що об’єдналися в п’ятий кластер (Львівська, Одеська та Харківська області), визначаються за всіма трьома складними ознаками. Таким чином, встановлено, що тільки в областях п’ятого кластера маємо відносно узгоджений, достатній за рівнем соціально-економічний стан за 18 елементарними ознаками, що змістовно сформульовані Держкомстатом України.

Отже, змодельовані рівні розвитку трьох складних ознак є обґрунтованими критеріями кластеризації областей України і спрямовують порівняльний й подальший детальний економічний аналіз на усунення структурних деформацій та диспропорцій соціально-економічного розвитку регіонів, а також на вдосконалення процедур узгодження цілей та пріоритетів економічного розвитку між центром і місцевими органами влади.

Задача 3.3. Провести багатовимірний аналіз даних для формування та ухвалення управлінського рішення в управлінні адміністративними районами області (на прикладі Харківської області).

Розв’язання задачі. Для аналізу та ухвалення управлінських рішень соціально-економічного розвитку 27 районів Харківської області слід реалізувати модель I типу.

Розглядаючи райони Харківської області як СЕС макрорівня, що описуються 13 елементарними ознаками, позначимо: x_1 – реалізація зернових культур сільськогосподарськими підприємствами (тис. ц); x_2 – інвестиції в основний капітал районів області (млн грн); x_3 – введення в експлуатацію житлових будинків (тис. м^2); x_4 – вантажообіг автомобільного транспорту в районі (млн т/км); x_5 – роздрібний товаробіг підприємств, включаючи ресторанне господарство (млн грн); x_6 – основні засоби в економіці (млн грн); x_7 – ступінь зносу основних засобів (%); x_8 – споживання електроенергії (млн кВт/год.); x_9 – середньорічна кількість найманих працівників (тис. осіб); x_{10} – підготовка та підвищення кваліфікації кадрів (% від загальної кількості працівників); x_{11} – потреба підприємств у працівниках на заміщення вільних робочих місць (осіб); x_{12} – працевлаштування не зайнятих трудовою діяльністю громадян (осіб); x_{13} – середньомісячна номінальна заробітна плата найманих працівників (грн). Дані взято зі статистичного збірника [292]. Маємо соціально-економічні системи, що описуються в просторі метричних ознак, а отже, складні ознаки слід визначати на основі реалізації базисної моделі І типу.

За базисною моделлю складних метричних ознак соціально-економічних систем потрібно визначити загальні складні соціально-економічні ознаки, що характеризують райони Харківської області. У табл. 3.21 наведено факторні навантаження після процедури $VARIMAX$. За значеннями коефіцієнтів факторних навантажень a_{ij} на i-ті показники встановлені загальні складні соціально-економічні ознаки стану районів Харківської області за пріоритетністю, а саме: перша загальна ознака – рівень економічного розвитку району (F_1), друга – рівень доходів населення (F_2), третя – технічний стан основних засобів (F_3), четверта – ефективність соціальної політики районів (F_4).

Таблиця 3.21

Факторна матриця після обертання *VARIMAX*
(за даними сукупності районів Харківської області)

F	F_1	F_2	F_3	F_4	R^2
x_1	-0,0599218	0,622776	-0,0190637	-0,138699	0,411041
x_2	0,590285	0,614271	0,347294	0,0532212	0,849211
x_3	0,8859	0,120405	0,268904	0,0078774	0,871687
x_4	0,905776	-0,0688385	-0,312214	-0,0630123	0,926616
x_5	0,932407	0,158742	-0,106654	0,118496	0,919999
x_6	0,71112	0,485544	0,396464	0,132623	0,916218
x_7	-0,0826876	-0,0565731	0,907145	0,016108	0,833208
x_8	0,747728	0,405243	-0,15619	0,167428	0,775746
x_9	0,906926	0,327172	0,0861803	0,139818	0,956533
x_{10}	-0,015718	0,362634	0,384868	0,559238	0,592621
x_{11}	0,747536	-0,316845	-0,0954031	0,0884396	0,676124
x_{12}	-0,538	0,128944	0,06536	-0,880709	0,814898
x_{13}	0,366769	0,743562	-0,0838691	0,319004	0,796201

За виявленими чотирма загальними складними соціально-економічними ознаками районів встановимо однорідність даної сукупності соціально-економічних систем. На рис. 3.32 зображена кластеризація сукупності районів Харківської області за складними ознаками на основі методу Уорда.

Візуальне подання рис. 3.32 свідчить, що за даною системою складних соціально-економічних ознак потрібно розглядати 5 кластерів районів Харківської області. У табл. 3.22 наведений склад кожного кластера.

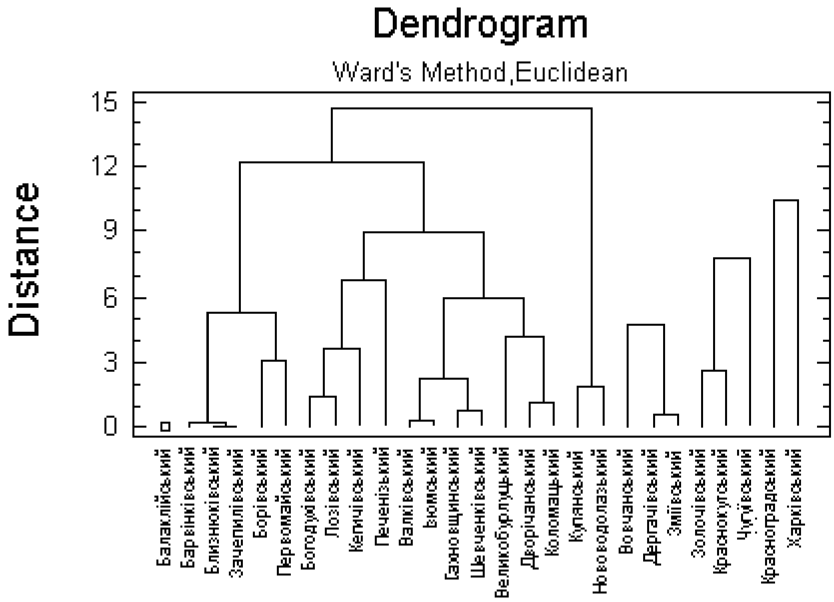


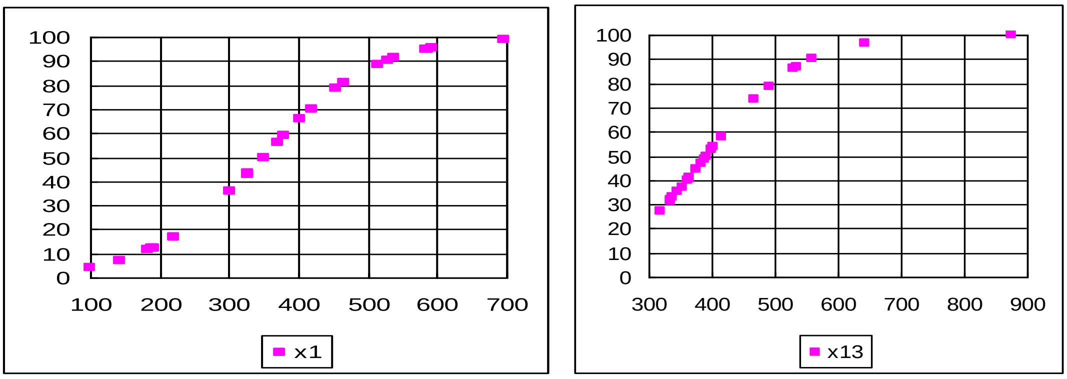
Рис. 3.32. **Дендрограма сукупності 27 районів Харківської області за їх загальними складними соціально-економічними ознаками**

Таблиця 3.22

Склад кластерів однорідних районів Харківської області за критерієм їх загальних складних соціально-економічних ознак

Район	Кластер	Район	Кластер	Район	Кластер
Балаклійський	1	Борівський	4	Коломацький	4
Золочівський	2	Зачепилівський	4	Лозівський	4
Краснокутський	2	Первомайський	4	Печенізький	4
Чугуївський	2	Богодухівський	4	Сахновщанський	4
Вовчанський	3	Валківський	4	Шевченківський	4
Дергачівський	3	Великобурлуцький	4	Куп’янський	4
Зміївський	3	Дворічанський	4	Нововодолазький	4
Барвінківський	4	Ізюмський	4	Красноградський	5
Близнюківський	4	Кегичівський	4	Харківський	5

За базисною моделлю I типу слід визначити значення загальних складних соціально-економічних ознак кожного з районів області, перетворивши значення 13 елементарних соціально-економічних ознак за функціями перетворення й обчисливши вимірники кожного з районів Харківської області. Як приклад на рис. 3.33 наведені функції перетворення ознак x_1 і x_{13}.



а б

Рис. 3.33. **Функції перетворення показників** x_1 **– реалізації зернових культур сільськогосподарськими підприємствами (а) й** x_{13} **– середньомісячної номінальної заробітної плати найманих працівників (б)**

Були обчислені значення всіх 13 вимірників, що визначають соціально-економічні ознаки порівняльного стану розвитку кожного району Харківської області.

Отже, у результаті реалізації базисної моделі І типу отримана діаграма соціально-економічного стану районів (рис. 3.34).

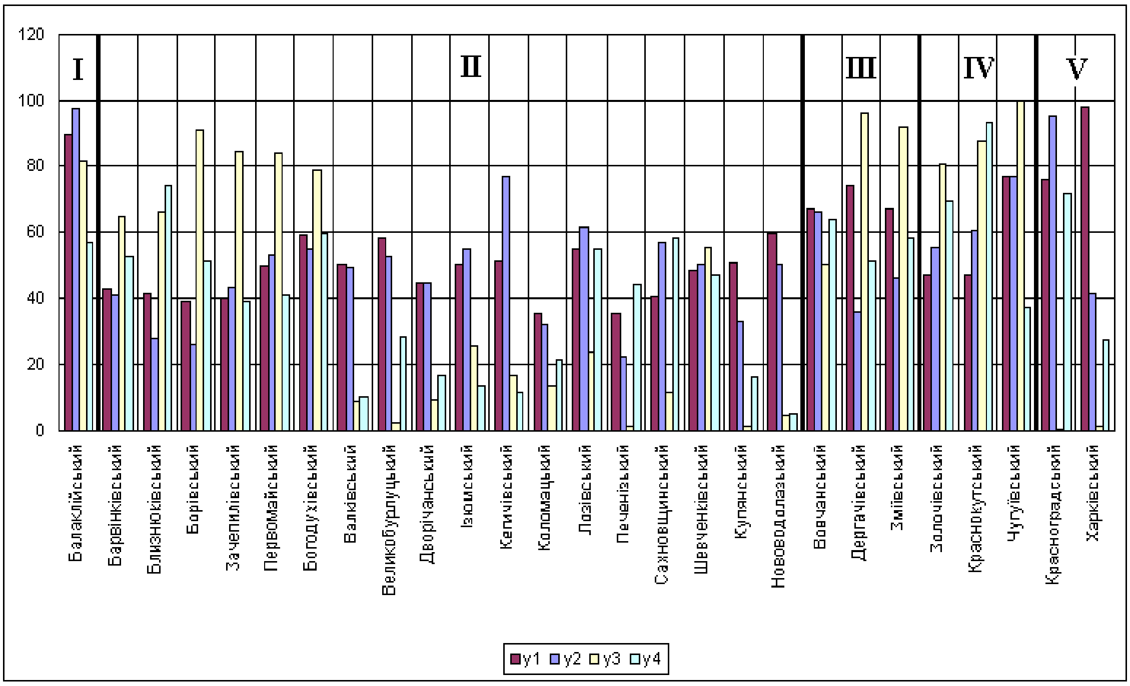


Рис. 3.34. **Діаграма соціально-економічного стану районів Харківської області за їх кластерами за складними ознаками** y_1 **– рівня економічного розвитку району;** y_2 **– рівня доходів населення;** y_3 **– технічного стану основних засобів;** y_4 **– ефективності соціальної політики районів**

За змістом рис. 3.34 видно, що найсприятливіший соціально-економічний стан спостерігається в Балаклійському районі, оскільки саме він один має значні величини всіх чотирьох складних ознак. Протилежний стан маємо в четвертому кластері, в який об'єдналися райони з низьким розвитком всіх чотирьох складних ознак. У третьому кластері, що утворили Вовчанський, Дергачівський та Зміївський райони, визначається добрий технічний стан основних засобів, а також достатній рівень економічного розвитку. Райони ж другого кластера відзначаються підвищеним рівнем доходів населення, а також добрим станом основних засобів. Порівняно високий рівень економічного розвитку мають Красноградський та Харківський райони, який, проте, супроводжується низькою ефективністю соціальної політики, що проводиться в даних районах. Таким чином, майже всі райони кожного кластера, крім першого, потребують цілеспрямованих управлінських заходів щодо покращення їх складних визначальних характеристик, а отже, органам управління в районах потрібно відпрацьовувати програми, які б цілісно підвищували рівні всіх складових соціально-економічного розвитку. Отже, за результатами реалізації базисної моделі I типу для розробки цілеспрямованого управлінського рішення отримане наочне подання задовільних та критичних станів окремих районів з одночасним розкриттям основних причин, що їх обумовили; визначена диференціація між районами передбачає управління, спрямоване на комплексне вирішення виявлених соціально-економічних проблем, специфічних для кожного району.

Розглянуті рішення практичних завдань управління СЕС демонструють очевидність прийняття гіпотези, що моделювання, залишаючись загальнонауковим методом пізнання об'єктів управління в економіці, набуває нового значення – виробництво нових знань в умовах, коли формуються нові парадигми в економіці та з'являються нові математичні методи, наявні потужні програмні середовища для аналізу даних. А також маємо підтвердження констатації, що математичне моделювання соціально-економічних систем, яке ґрунтується на концепціях визначення величин ознак в економіці, спроможне на вищому науковому рівні методологічно організувати аналіз даних в економіці, а отже, забезпечити якісною інформацією системи розробки та ухвалення управлінських рішень.

Задачі до розділу 3

За даними, що характеризують виробничо-господаську діяльність 15 промислових підприємств протягом 5 років (додаток А), розв'язати задачі багатовимірного аналізу соціально-економічних систем (умовне позначення табл. 3.1):

1. За допомогою інструментів описової статистики описати елементарні ознаки виробничо-господарської діяльності підприємств.

2. За допомогою факторного аналізу виявити загальні латентні складні ознаки виробничо-господарської діяльності підприємств.

3. За допомогою кластерного аналізу визначити кластери підприємств за критерієм їх загальних складних ознак виробничо-господарської діяльності.

4. Використавши дискримінантний аналіз, перевірити стійкість отриманої кластеризації підприємств.

5. Визначити латентні складні ознаки виробничо-господарської діяльності підприємств в окремих кластерах та за допомогою канонічного аналізу виявити системи складних ознак діяльності.

6. Провести економічний аналіз отриманих результатів моделювання, синтезувати висновки для ухвалення рішень в управлінні даною сукупністю підприємств.

Запитання для самоперевірки

1. Перелічити процедури, що входять до базової системи ППП Statgraphics.

2. Назвати переваги використання ППП Statgraphics у вирішенні реальних задач аналізу соціально-економічних систем.

3. Назвати основні можливості процедур, що входять до базової системи ППП Statgraphics.

4. Яке основне призначення інструментів описової статистики?

5. Який склад інструментів описової статистики?

6. Назвати основні призначення кожного з інструментів описової статистики.

7. Назвати основні етапи аналізу результатів використання інструментів описової статистики.

8. Який загальний вигляд шкали окремої ознаки об’єкта в економіці, розбудованої за допомогою статистичних характеристик величини ознаки?

9. Яка логіка етапів розробки описових моделей соціально-економічних систем за метричними ознаками (І тип моделей)?

10. Як ідентифікуються латентні фактори розвитку соціально-економічних систем?

11. Які існують методи перевірки стійкості груп, отриманих за допомогою кластерного аналізу?

12. Яка основна ідея вирішення задачі розділення суміші і її значення в економіці?

13. Який метод слід використовувати у моделюванні взаємозв’язку в складних системах ознак, що типізуються за змістом у дві групи?

14. Як інтерпретувати зміст діаграми, що візуально представляє результати розробки описових моделей складних метричних ознак соціально-економічних систем?

Розділ 4
Аналіз соціально-економічних систем за неметричними ознаками

4.1. Початковий аналіз соціально-економічних систем за порядковими ознаками

Вимірювання якісних ознак об'єкта здійснюється за допомогою номінальних та порядкових шкал. Ці ознаки, як уже було сказано, неметричні, і їх кількість велика в описі СЕС. За допомогою відповідей на запитання в анкетах, як правило, визначають якісні характеристики явища чи процесу, що аналізується. Якщо експерт дає відповідь за бальною системою, то наявна неметрична величина ознак, яка виміряна в порядкових шкалах. Використання таких величин має місце, коли властивості об'єкта погано вивчені, тобто проводиться вимірювання в умовах невизначеності. Результатом такого вимірювання є порядок, в якому вибудовуються інтенсивності прояву даної ознаки в часі чи в просторі [160].

Наявні рекомендації розбудови моделей у порядкових шкалах [254; 317]. Вони спираються на властивість порядкових шкал бути інваріантними відносно взаємних однозначних монотонних їх перетворень, звідки випливає, що властивості залежності між порядковими змінними повинні бути інваріантними відносно взаємно однозначних монотонних перетворень й головною проблемою є визначення напрямку зв'язку між змінними, який може бути прямим ("додатним") або зворотним ("від'ємним"). Якщо потрібно будувати функції декількох порядкових змінних, то необхідно як визначити напрям впливу кожної з незалежних змінних на залежну, так і впорядкувати всі можливі комбінації напрямів змін незалежних змінних за значенням залежної змінної. Розглядається простір змін незалежних змінних, який утворює частково впорядковану множину, на якій один елемент $a = (a_1, a_2, ..., a_n)$ не більший (не менший) від іншого елемента $b = (b_1, b_2, ..., b_n)$, якщо всі його значення $a_i \le (\ge) b_i, i = \overline{1, n}$. У протилежному разі елементи непорівнянні.

Завдання функції на частково впорядкованій множині полягає в упорядкуванні цієї множини відповідно до змін значень залежної змінної. При цьому має виконуватися умова монотонності залежності: якщо y зростає зі збільшенням $x_1, x_2, ..., x_n$, то для двох елементів a і b, таких, що $a_i \le b_i, i = \overline{1, n}$, справджується $y(a) \le y(b)$.

Простір змін незалежних порядкових змінних утворює решітку, тобто множину, в якій для будь-яких двох елементів a і b завжди існує нижня грань $a \wedge b$, причому $a \wedge b \leq a$ і $a \wedge b \leq b$, а також двійкова їй верхня грань $a \vee b$: $a \vee b \geq a,\ a \vee b \geq b$.

У разі задавання функції порядкової змінної на множині типу решітки спочатку необхідно задати розбиття простору зміни незалежних змінних на підмножини $S_1 = \left(s_1^1, \ldots, s_1^{m_1}\right), \ldots, S_k = \left(s_k^1, \ldots, s_k^{m_k}\right)$, такі, що $S_1, \ldots, S_k$ повністю покривають початкову множину, а будь-які два елементи s_e^i і s_e^j з однієї підмножини S_e непорівнянні, а отже, верхня і нижня грані не належать множині S_e, при цьому граничний випадок: кожна підмножина S_e містить один елемент s_e^1.

Наступним етапом задавання функції порядкових змінних є впорядкування підмножин $S_1, \ldots, S_k$ відповідно до змін значень залежної змінної з виконанням умови монотонності, тобто визначення ланцюга $S_{p_1}, \ldots, S_{p_k}$ такого, що $y\left(s_{p_a}^i\right) \leq y\left(s_{p_b}^i\right)$ для будь-яких елементів $s_{p_a}^i \in S_{p_a}, s_{p_b}^i \in S_{p_b}, p_a \leq p_b$; у випадку якщо знайдуться елементи $s_{p_a}^i$ і $s_{p_b}^j$ такі, що $y\left(s_{p_a}^i\right) \leq y\left(s_{p_b}^j\right)$, то $p_a \leq p_b$. При цьому елементи з однієї підмножини S_{p_a} вважаються однаковими за значенням змінної y [317, с. 307].

Таким чином, процедура ідентифікації функції порядкових змінних включає такі три етапи: визначення напряму впливу кожної з незалежних змінних на залежну; розбиття простору зміни незалежних змінних на підмножини непорівнянних елементів; упорядкування цих підмножин відповідності до зміни значень залежної змінної. При цьому відзначаються жорсткі вимоги до наявної вибірки, яка має достатньо повно подавати все поле змін незалежних змінних; у протилежному разі неможливо ідентифікувати залежність. Перспективним напрямом вважається побудова функцій порядкових змінних за допомогою експертних процедур.

Також вважається, що надійним у процесі побудови моделей у порядкових шкалах є використання для відображення залежностей порядкових функцій, коли до змінних у моделі висуваються наступні вимоги: вони повинні мати однакову кількість градацій змін, а зі змістовної позиції – бути порівнянними одна з одною за своїми значеннями. Порядкова функція при цьому може мати такий вигляд. Якщо $x_1, x_2, \ldots, x_n, y$ – порядкові змінні і y

зростає зі збільшенням $x_1, x_2, ..., \cdots x_n$, тоді y може набувати значень незалежної змінної, яка стоїть на m-му місці проранжованого ряду значень усіх незалежних змінних. У цьому випадку можуть вибиратися екстремальні значення – мінімальний $(m=1)$ і максимальний $(m=n)$ елементи, різні квантилі: децилі, квартилі, медіана і т. д. Від вибору номера елемента m, значення якого визначає величину змінної y, залежить характер функції, що будується.

Звичайно, рекомендації побудови моделей у порядкових шкалах теоретично обґрунтовані, але не знайшли свого поширення в моделюванні СЕС у процесі вирішення реальних завдань в економіці, оскільки вихідні дані подаються матрицями великих розмірів, що ускладнює процедури порівняння. Для математичного моделювання неметричних порядкових ознак мають значення не стільки результати їх вимірювання (числа, що оцінюють інтенсивності прояву ознаки) $X_1, X_2, ..., X_n$ для об’єктів $O(1), O(2), ..., O(n)$, скільки ранги $r_1, r_2, ..., r_n$ чисел $X_1, X_2, ..., X_n$, де r_i – ранг X_i серед чисел $X_1, X_2, ..., X_n$. Аналіз зв’язків між порядковими ознаками зводиться до аналізу різних упорядкувань (ранжувань) однієї й тієї ж скінченної множини об’єктів та здійснюється за допомогою методів рангової кореляції. Упорядкування ступеня інтенсивності прояву величини ознак здійснюється або за допомогою експертів, або формалізовано – за допомогою переходу від початкового ряду спостережень опосередкованої (похідної) кількісної ознаки до відповідного варіаційного ряду [317, с. 226]. Тут слід зазначити, що будь-яку метричну величину також можна вимірювати за допомогою порядкових шкал, якщо присвоїти бали інтенсивностям метричної величини. Цей факт надзвичайно важливий в обробці даних в економіці, коли не завжди можливо перевірити закон розподілу значень показників [103; 108; 162; 164].

До основних задач теорії та практики рангової кореляції належать: аналіз структури досліджуваної сукупності упорядкувань (задача А); аналіз інтегральної (сукупної) узгодженості змінних і їх умовне ранжування за критерієм ступеня тісноти зв’язку кожної з них із рештою змінними (задача В); побудова єдиного групового впорядкування об’єктів на основі наявної сукупності узгоджених упорядкувань (задача С) [317, с. 226]. Спираючись на це, розбудову моделі системи порядкових ознак слід починати з визначення парних і множинних зв’язків у системі, а отже, з обчислення коефіцієнтів рангової кореляції. Аналіз взаємозв’язків порядкових змінних будується на основі різних варіантів моделей імовірнісного простору, в якому роль простору елементарних результатів (наслідків) виконує множина всіх можливих сполучень з n елементів (n – кількість статистично досліджуваних об’єктів).

Методи рангової кореляції є єдиним шляхом узагальнення експертних оцінок, коли ознаки об’єкта в економіці не вдається описати метричною величиною. Ранг вказує, яке місце займає об’єкт у ряді об’єктів, упорядкованих за певною ознакою [89].

Традиційно в якості основних характеристик парного статистичного зв’язку між упорядкуваннями використовуються коефіцієнти рангової кореляції Спірмена або Кендела. Як відомо, першим запропонував розв’язання задачі перевірки гіпотези про незалежність порядкових ознак психолог Ч. Спірмен у 1900 році. Другий за популярністю коефіцієнт рангової кореляції – коефіцієнт М. Кендела – в якості міри схожості між двома ранжуваннями використовує мінімальну кількість перестановок сусідніх об’єктів, яку потрібно зробити, щоб одне впорядкування об’єктів перетворити в інше [275, с. 300].

Для побудови складних порядкових ознак СЕС слід проаналізувати процедури обчислень коефіцієнтів рангової кореляції й уточнити необхідні теореми [164].

Нехай p_k, q_k – ранги величин ознак X і Y. Спостереження завжди можна відсортувати в порядку зростання рангів однієї ознаки $p_k = 1, 2, 3, \ldots, n$. Ранги q_k – це ті ж самі числа, але в іншому порядку.

Мірою тісноти зв’язку між ознаками X і Y може бути сума квадратів різниць рангів:

$$S = \sum_{k=1}^{n} (p_k - q_k)^2 . \tag{4.1}$$

Якщо ранги двох ознак співпадають $p_k = q_k$, то $S = 0$, і це відповідає найтіснішому додатному зв’язку.

Якщо порядок слідування v_k протилежний порядку u_k, то $S = S_{max}$, що відповідає найтіснішому від’ємному зв’язку. Слід обчислити S_{max}. Для цього випадку маємо $p_k + q_k = n + 1$, $q_k = n + 1 - p_k$, $p_k - q_k = 2p_k - (n+1)$, де $p_k = k$. Звідси:

$$\begin{aligned} S &= \Sigma (p_k - q_k)^2 = \Sigma (2k - (n+1))^2 = 4\Sigma k^2 - 4(n+1)\Sigma k + (n+1)^2 n = \\ &= 4\frac{n(n+1)(2n+1)}{6} - 4(n+1)\frac{n(n+1)}{2} + (n+1)^2 n = \frac{(n-1)n(n+1)}{3} . \end{aligned} \tag{4.2}$$

У процесі перетворень використані відомі формули:

$$\sum_{k=1}^{n} k^2 = 1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}; \qquad (4.3)$$

$$\sum_{k=1}^{n} k = 1 + 2 + 3 + \ldots + n = \frac{n(n+1)}{2}. \qquad (4.4)$$

Замість міри зв'язку S вводимо міру

$$\rho = 1 - 2\frac{S}{S_{max}}, \qquad (4.5)$$

яка дорівнює $\rho = 1$ для $S = 0$ (для найтіснішого додатного зв'язку), і $\rho = -1$ для $S = S_{max}$ (для найтіснішого від'ємного зв'язку).

Ця міра обчислюється за формулою:

$$\rho = 1 - \frac{6\sum (p_k - q_k)^2}{(n-1)n(n+1)}, \qquad (4.6)$$

і має назву коефіцієнта рангової кореляції Спірмена.

Теорема. Коефіцієнт рангової кореляції Спірмена співпадає зі звичайним коефіцієнтом парної кореляції Пірсона, обчисленим за рангами:

$$\rho = r_{pq} = \frac{S_{pq}}{S_p \cdot S_q}. \qquad (4.7)$$

Доведення.

Оскільки $p_k = k$, а q_k – ті ж числа в іншому порядку, то для X і Y будуть однакові середні $\bar{p} = \bar{q} = \frac{n+1}{2}$ і дисперсії $S_p^2 = S_q^2 = \overline{p^2} - (\bar{p})^2$. Перетворимо вираз $S = \sum_{k=1}^{n} (p_k - q_k)^2$ (чисельник у формулі Спірмена $\rho = 1 - \frac{2S}{S_{\max}}$):

$$\begin{aligned} S &= \sum (p_k^2 + q_k^2 - 2p_k q_k) = n(\overline{p^2} + \overline{q^2} - 2\overline{pq}) = n(S_p^2 + S_q^2 - 2S_{pq}) = \\ &= 2n(S_p^2 - S_{pq}) = 2nS_p^2(1 - r_{pq}). \end{aligned} \qquad (4.8)$$

Знайдемо S_{max}, що буде при $p_k + q_k = n + 1$.

$$S_{\max} = \sum \left(2k - (n+1)\right)^2 = 4\sum \left(k - \frac{n+1}{2} \right)^2 = 4n \cdot S_p^2 . \tag{4.9}$$

Звідси:

$$\rho = 1 - 2\frac{S}{S_{\max}} = 1 - \frac{2 \cdot 2nS_p^2 (1 - r_{pq})}{4nS_p^2} = 1 - (1 - r_{pq}) = r_{pq} \tag{4.10}$$

звідки дійсне:
$$\rho = r_{pq} = \frac{S_{pq}}{S_p \cdot S_q}, \tag{4.11}$$

де $S_p^2 = \overline{p^2} - (\bar{p})^2 = \frac{1}{n}\sum (p_k - \bar{p})^2$;

$S_q^2 = \overline{q^2} - (\bar{q})^2 = \frac{1}{n}\sum (q_k - \bar{q})^2$;

$S_{pq} = \overline{pq} - \bar{p} \cdot \bar{q} = \frac{1}{n}\sum (p_k - \bar{p})(q_k - \bar{q})$;

$p_k = k$; $\bar{p} = \bar{q} = \frac{n+1}{2}$; $S_p^2 = S_q^2$.

Отже, коефіцієнт рангової кореляції Спірмена дорівнює звичайному коефіцієнту парної кореляції Пірсона, якщо за змінні прийняти ранги p_k, q_k. Таким чином, для розв'язання практичних задач з порядковими ознаками можна використовувати програми, що написані для обчислення коефіцієнтів парної кореляції Пірсона. ■

Окремо слід розглянути випадок наявності груп однакових рангів. Якщо існує декілька однакових значень p_k чи q_k, яким присвоєно середній ранг для цієї зв'язки, то формула коефіцієнта рангової кореляції Спірмена ускладнюється, тобто слід ввести поправки на наявність груп однакових рангів.

Сума рангів завжди дорівнює $(1 + 2 + ... + n) = \frac{n(n+1)}{2}$, тобто середній ранг завжди буде дорівнювати $\frac{n+1}{2}$.

Сума квадратів незв'язаних, тобто різних рангів дорівнює $\frac{n(n+1)(2n+1)}{6}$, їх середній квадрат – $\frac{(n+1)(2n+1)}{6}$ і дисперсія різних рангів дорівнює

$$S_p^2 = \overline{p^2} - (\bar{p})^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}. \qquad (4.12)$$

Якщо t рангів зв'язані, сума рангів не змінюється, але зменшується сума квадратів і дисперсія. Нехай ранги $p_k+1, p_k+2, \dots, p_k+t$ – зв'язані і замінені на $\left(p_k + \frac{t+1}{2}\right)$. Тоді сума їх квадратів буде дорівнювати $t\left(p_k + \frac{t+1}{2}\right)^2$, а загальна сума квадратів скорочується на

$$(p_k+1)^2 + (p_k+2)^2 + \dots + (p_k+t)^2 - t\left(p_k + \frac{t+1}{2}\right)^2 = tp_k^2 + 2p_k(1+2+\dots+t) +$$

$$+ (1^2 + 2^2 + \dots + t^2) - tp_k^2 - p_k t(t+1) - \frac{t(t+1)^2}{4} = \frac{t(t+1)(2t+1)}{6} - \frac{t(t+1)^2}{4} =$$

$$= \frac{t(t^2-1)}{12}.$$

Якщо маємо кілька груп зв'язок, то сума квадратів рангів скорочується на

$$T = \frac{1}{12}\sum (t^3 - t), \qquad (4.13)$$

а дисперсія рангів – на $\frac{1}{n}T$. Виправлена дисперсія рангів обчислюється за формулою:

$$S_p^2 = \frac{n^2-1}{12} - \frac{1}{12n}\sum (t^3 - t) = \frac{n^2-1}{12}(1 - A), \qquad (4.14)$$

де $A = \frac{\sum (t-1)t(t+1)}{(n-1)n(n+1)}$ – поправка на зв'язаність рангів.

Для дисперсії S_q^2 поправку на зв'язаність рангів q_k позначимо через B:

$$S_q^2 = \frac{n^2-1}{12}(1 - B). \qquad (4.15)$$

Перетворюємо вираз коваріації S_{pq}. Оскільки

$$S=\sum (p_k-q_k)^2=\sum p_k^2+\sum q_k^2-2\sum p_k q_k=n(S_p^2+S_q^2-2S_{pq}), \quad (4.16)$$

то

$$\begin{aligned} S_{pq}&=\frac{S_p^2+S_q^2}{2}-\frac{1}{2n}S=\frac{n^2-1}{12}\cdot\frac{(1-A)+(1-B)}{2}-\frac{1}{2n}S= \\ &=\frac{n^2-1}{12}\left(1-\frac{A+B}{2}-\frac{6S}{n^3-n}\right)=\frac{n^2-1}{12}\left(\rho-\frac{A+B}{2}\right), \end{aligned} \quad (4.17)$$

де $\rho=\left(1-\frac{6S}{n^3-n}\right)$ – коефіцієнт Спірмена для незв’язаних послідовностей рангів.

Обчислюємо виправлене значення ρ_s – коефіцієнта Спірмена з урахуванням однакових груп рангів у послідовностях p_k і q_k:

$$\rho_s=r_{pq}=\frac{S_{pq}}{S_p\cdot S_q}=\frac{\frac{n^2+1}{2}\left(\rho-\frac{A+B}{2}\right)}{\sqrt{\frac{n^2+1}{2}(1-A)}\sqrt{\frac{n^2+1}{2}(1-B)}}=\frac{\rho-\frac{A+B}{2}}{\sqrt{(1-A)}\sqrt{(1-B)}}. \quad (4.18)$$

Отже, одержуємо
$$\rho_s=\frac{\rho-\frac{A+B}{2}}{\sqrt{(1-A)}\sqrt{(1-B)}}. \quad (4.19)$$

Найчастіше цей коефіцієнт записують у більш громіздкій і дуже незручній формі:

$$\rho_s=\frac{\frac{n^3-n}{6}-\sum (p_k-q_k)^2-A^*-B^*}{\sqrt{\frac{n^3-n}{6}-2A^*}\cdot\sqrt{\frac{n^3-n}{6}-2B^*}}, \quad (4.20)$$

де A^* і B^* − поправки для зв’язків у послідовностях рангів p і q, що обчислюють за формулами:

$$A^*=\frac{1}{12}\sum (A_i^3-A_i); \quad (4.21)$$

$$B^*=\frac{1}{12}\sum (B_i^3-B_i). \quad (4.22)$$

Тут A_i і B_i – кількість однакових значень ряду рангів у групах ряду p_k і q_k.

У коефіцієнті рангової кореляції Кендела не використовуються числові значення рангів p_k і q_k.

Порівняємо кожну пару рангів $(p_i - p_j)$ і $(q_i - q_j)$. Якщо ці пари мають зворотний порядок, тобто при $p_i < p_j$ буде $q_i > q_j$ (або, навпаки, при $p_i > p_j$ буде $q_i < q_j$), то Кендел пропонує присвоїти цим парам значення -1. Якщо порядок в обох парах однаковий, тобто при $p_i < p_j$ буде також $q_i < q_j$ (або навпаки: при $p_i > p_j$ буде $q_i > q_j$), Кендел пропонує присвоїти цим парам значення +1.

Обчислюємо суму привласнених значень ± 1 за всіма парами рангів, що можна записати у вигляді формули:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn\left[(p_i - p_j)(q_i - q_j)\right], \tag{4.23}$$

де $sgn(x) = \begin{cases} +1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$ – функція виділення знака.

Найтісніший зв'язок спостерігається, коли послідовності рангів співпадають, тоді всі присвоєні значення дорівнюють 1 і буде

$$S = S_{\max} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}\left[(p_i - q_j)^2\right] = \frac{n(n-1)}{2}. \tag{4.24}$$

Позначимо

$$\tau = \frac{S}{S_{max}} = \frac{2S}{n(n-1)}. \tag{4.25}$$

Ця величина називається коефіцієнтом рангової кореляції Кендела. Значення коефіцієнта Кендела змінюється в межах $-1 \leq \tau \leq 1$.

Розглянемо проблему зв'язаних рангів. Якщо є групи однакових значень рангів, то в межах цих груп різниці $(p_i - p_j)$ чи $(q_i - q_j)$ дорівнюють нулю. Це вже враховано у визначенні функції $sgn(x)$, і величина суми S буде автоматично скоректована. Але за наявності зв'язаних рангів також має змінюватися величина S_{max}. Отже, існують дві форми коефіцієнта Кендела τ_a і τ_b (аналогічно до двох форм коефіцієнта Спірмена ρ і ρ_S):

Якщо прийняти в якості $S_{max} = \frac{n(n-1)}{2}$ як для незв'язаної форми, то буде отримано коефіцієнт, який позначено τ_a:

$$\tau_a = \frac{2S}{n(n-1)}. \tag{4.26}$$

Наявність груп рівних рангів зменшує лише чисельник S у цій формулі.

Коефіцієнт τ_a не є доброякісною мірою узгодженості, оскільки в разі повного збігу послідовностей, але за наявності зв'язаних рангів ніколи не виконується умова $\tau_a = 1$. Слід зазначити, що може мати місце випадок, коли навіть замість $\tau_a \approx 1$ буде отримане $\tau_a \approx 0$. Наприклад, цей випадок виникає, якщо перші $(n-1)$ членів однакові і дорівнюють середньому рангу $p_k = q_k = \frac{n}{2}$, а останній член дорівнює $p_n = q_n = n$. Тоді замість $\tau_a = 1$ одержимо $\tau_a = \frac{2}{n} \to 0$.

Для правильного врахування груп рівних рангів спочатку розглянемо узагальнений коефіцієнт кореляції у вигляді формули:

$$R = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}}. \tag{4.27}$$

Теорема. З узагальненої форми R за різних виборів a_{ij}, b_{ij} можна отримати всі відомі коефіцієнти кореляції: при $a_{ij} = x_i - x_j$ і $b_{ij} = y_i - y_j$ з узагальненої форми випливає коефіцієнт парної кореляції Пірсона; при $a_{ij} = p_i - p_j$ і $b_{ij} = q_i - q_j$ – коефіцієнт рангової кореляції Спірмена; при $a_{ij} = \mathrm{sgn}(p_i - p_j)$ і $b_{ij} = \mathrm{sgn}(q_i - q_j)$ – коефіцієнт рангової кореляції Кендела.

Доведення. А. Покажемо, що при $a_{ij} = x_i - x_j$, $b_{ij} = y_i - y_j$ з узагальненого коефіцієнта кореляції випливає коефіцієнт парної кореляції Пірсона.

Дійсно,

$$\sum\sum a_{ij}^2 = \sum\sum \left(x_i - x_j\right)^2 = \sum\sum \left[\left(x_i - \overline{x}\right) - \left(x_j - \overline{x}\right)\right]^2 =$$
$$= \sum\sum \left(x_i - \overline{x}\right)^2 + \sum\sum \left(x_j - \overline{x}\right)^2 - 2\sum\sum \left(x_i - \overline{x}\right)\left(x_j - \overline{x}\right) = 2n^2 S_x^2, \tag{4.28}$$

аналогічно

$$\sum\sum b_{ij}^2 = n^2 S_y^2, \tag{4.29}$$

$$\sum\sum a_{ij} b_{ij} = \sum\sum \left(x_i - x_j\right)\left(y_i - y_j\right) = \sum\sum \left[\left(x_i - \overline{x}\right) - \left(x_j - \overline{x}\right)\right]\left[\left(y_i - \overline{y}\right) - \left(y_j - \overline{y}\right)\right] =$$
$$= \sum\sum \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right) + \sum\sum \left(x_j - \overline{x}\right)\left(y_j - \overline{y}\right) - \sum\sum \left(x_i - \overline{x}\right)\left(y_j - \overline{y}\right) -$$
$$- \sum\sum \left(x_j - \overline{x}\right)\left(y_i - \overline{y}\right) = 2n^2 S_{xy}.$$

Звідси дійсно одержуємо коефіцієнт парної кореляції Пірсона:

$$R = \frac{2n^2 S_{xy}}{\sqrt{2n^2 S_x^2 \cdot 2n^2 S_y^2}} = \frac{S_{xy}}{S_x S_y} = r_{xy}. \tag{4.30}$$

Б. Якщо $a_{ij} = p_i - p_j$ і $b_{ij} = q_i - q_j$, то з узагальненої форми одержуємо $R = r_{pq}$, що (як було доведено вище) дорівнює коефіцієнту рангової кореляції Спірмена.

В. Якщо $a_{ij} = \pm 1, b_{ij} = \pm 1$, тобто

$$a_{ij} = \begin{cases} 1 & \text{при} \quad p_i < p_j \\ -1 & \text{при} \quad p_i > p_j \end{cases}; \quad b_{ij} = \begin{cases} 1 & \text{при} \quad q_i < q_j \\ -1 & \text{при} \quad q_i > q_j \end{cases}, \tag{4.31}$$

то із узагальненої форми випливає коефіцієнт τ.

Дійсно, маємо

$$\sum_{i,j} a_{ij}^2 = \sum_{i,j} 1 = n(n-1); \ \sum_{i,j} b_{ij}^2 = \sum_{i,j} 1 = n(n-1). \tag{4.32}$$

Таким чином, знаменник узагальненої форми за відсутності зв'язаних рангів дорівнює $n(n-1)$. Чисельник у формулі (4.27) $\sum a_{ij} b_{ij}$ дорівнює $2S$ (подвоєння буде, оскільки під час додавання кожна пара зустрічається двічі: один раз – у сполученні ij, другий – у сполученні ji). Звідси

$$R = \frac{2S}{n(n-1)} = \tau. \tag{4.33}$$

■

В узагальненій формі можна легко враховувати наявність зв'язаних рангів на величину S_{max} у формулі τ.

Якщо існує зв'язок t послідовних членів, то всі оцінки a_{ij}, що стосується будь-якої вибраної з них пари, дорівнюють нулю:

$$a_{ij} = 0 \text{ при } \ p_i = p_j .$$

Таких пар нараховується $t(t-1)$, тому суму $\sum a_{ij}^2$ можна скоротити на цю величину і вона буде дорівнювати:

$$\sum a_{ij}^2 = n(n-1) - \sum t(t-1) = n(n-1)[1 - A], \tag{4.34}$$

де $A = \dfrac{\sum t(t-1)}{n(n-1)}$ – поправка на зв'язність у ряді p_k.

Аналогічно можна одержати

$$\sum b_{ij}^2 = n(n-1)[1 - B], \tag{4.35}$$

де B – поправка на зв'язки в ряді q_k.

Звідси маємо:

$$\tau_b = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum a_{ij} \sum b_{ij}}} = \frac{2S}{\sqrt{n(n-1)(1-A)}\sqrt{n(n-1)(1-B)}} = \frac{\tau_a}{\sqrt{1-A}\sqrt{1-B}} . \tag{4.36}$$

Коефіцієнт τ_b є доброякісною мірою узгодженості.

Для моделювання системи порядкових ознак потрібно ввести коефіцієнт конкордації (W) як міру узгодженості оцінок різних експертів.

Нехай m експертів (кожний окремо) проранжували n об'єктів. Ранги y_{ij} для кожного експерта B_j є натуральними числами від 1 до m, що розміщуються в різному порядку. Усі ці відомості зведені в табл. 4.1, де U_i – суми рангів усіх експертів для кожного об'єкта.

Таблиця 4.1

Величини рангів об'єктів за різними експертами

Об'єкти	Експерти				Суми рангів	Середні
	B_1	B_2	...	B_m		
1	y_{11}	y_{12}	...	y_{1m}	U_1	u_1
2	y_{21}	y_{22}	...	y_{2m}	U_2	u_2
3	y_{31}	y_{32}	...	y_{3m}	U_3	u_3
...	...	...	...	...	...	...
n	y_{n1}	y_{n2}	...	y_{nm}	U_n	u_n

Міру узгодженості оцінок двох експертів можна оцінити за допомогою коефіцієнта рангової кореляції Спірмена ρ_S. Таких порівнянь буде $\frac{m(m-1)}{2}$. Але потрібно знайти загальну міру узгодженості всієї групи експертів. Така міра (коефіцієнт конкордації) була запропонована Кенделом у вигляді

$$W = \frac{12 \cdot S}{m^2(n^3 - n)}, \tag{4.37}$$

де $S = \sum \left(U_i - m \cdot \frac{n+1}{2} \right)^2$.

Теорема. Коефіцієнт конкордації співпадає з індексом детермінації, обчисленим за таблицею рангів y_{ij}.

Доведення. У табл. 4.1 введені такі позначення: $U_i = \sum_{j=1}^{n} y_{ij}$, $u_i = \frac{1}{m} \sum_{j=1}^{m} y_{ij} = \frac{1}{m} U_i$ (середні за об'єктами). Потрібно обчислити коефіцієнт детермінації рангів, для чого обчислимо загальну дисперсію рангів S_y^2 та дисперсію середніх за об'єктами S_u^2. Доведемо, що індекс детермінації рангів $R^2 = \frac{S_u^2}{S_y^2}$ (для метричних величин – квадрат кореляційного відношення $\eta^2 = \frac{S_u^2}{S_y^2}$) дійсно дорівнює коефіцієнту конкордації W.

Оскільки ранги y_{ij} у стовпчиках табл. 4.1 є натуральними числами в різних послідовностях, то їх суми у стовпчиках однакові й дорівнюють $\frac{n(n+1)}{2}$, також однакові середні, що дорівнюють $\frac{n+1}{2}$. Загальна сума рангів дорівнює $m\frac{n(n+1)}{2}$, загальна середня – $\bar{y} = \frac{1}{m \cdot n} \cdot \frac{mn(n+1)}{2} = \frac{n+1}{2}$.

Суми квадратів у стовпчиках за відсутності зв'язок обчислюються за формулою $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$. Тому загальна сума квадратів буде $\sum_i \sum_j y_{ij}^2 = m\frac{n(n+1)(2n+1)}{6}$ і середній квадрат – $\overline{y^2} = \frac{(n+1)(2n+1)}{6}$. Звідси можна одержати загальну дисперсію рангів (за відсутності зв'язок):

$$S_y^2 = \overline{y^2} - (\bar{y})^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}. \tag{4.38}$$

Дисперсія S_u^2 (між групами або між об'єктами) $S_u^2 = \frac{1}{n}\sum_{i=1}^{n}(u_i - \bar{y})^2$ пропорційна величині

$$S = \sum\left(U_i - m\frac{n+1}{2}\right)^2, \tag{4.39}$$

де U_i – сума рангів експертів для кожного об'єкта, $U_i = mu_i$;

$\frac{n+1}{2}$ – середній ранг;

$m\frac{n+1}{2}$ – середня суми рангів експертів;

S – сума квадратів відхилень U_i від своєї середньої.

Виконаємо перетворення:

$$\begin{aligned} S &= \sum_{i=1}^{n}\left(U_i - m\frac{n+1}{2}\right)^2 = \sum_{i=1}^{n}\left(mu_i - m\frac{n+1}{2}\right)^2 = m^2\sum_{i=1}^{n}\left(u_i - \frac{n+1}{2}\right)^2 = \\ &= m^2\sum_{i=1}^{n}\left(u_i - \bar{y}\right)^2 = m^2 \cdot n \cdot S_u^2. \end{aligned} \tag{4.40}$$

Звідси

$$S_u^2 = \frac{1}{m^2 n} S\,. \tag{4.41}$$

Таким чином, одержуємо індекс детермінації $R^2 = \frac{S_u^2}{S_y^2} = \frac{12S}{m^2 n\left(n^2 - 1\right)}$, що дійсно дорівнює коефіцієнту конкордації W.

Тоді одержуємо виправлену формулу коефіцієнта конкордації:

$$R^2 = \frac{\frac{1}{m^2 n} S}{\frac{n^2 - 1}{12}\left(1 - \frac{A}{m}\right)} = \frac{12 \cdot S}{m^2\left(n^3 - n\right)\left(1 - \frac{A}{m}\right)} = W_*\,. \tag{4.42}$$

■

Якщо дійсно $W = R^2$, то спрощується проблема врахування зв’язаних рангів під час обчислення коефіцієнта конкордації.

За наявності груп зв’язаних (однакових) рангів їх сума та середня залишаються одними й тими ж.

За наявності груп зв’язаних рангів сума квадратів скорочується на величину

$$T = \frac{1}{12} \sum_t \left(t^3 - t\right), \tag{4.43}$$

де t – кількість однакових рангів у зв’язці.

Середній квадрат і дисперсія скорочуються на величину $\frac{1}{m \cdot n} T$, звідки

$$S_y^2 = \frac{n^2 - 1}{12} - \frac{1}{12 m \cdot n} \sum \left(t^3 - t\right) = \frac{n^2 - 1}{12}\left(1 - \frac{A}{m}\right), \tag{4.44}$$

де $A = \frac{\sum (t-1)t(t+1)}{(n-1)n(n+1)}$ (поправка, аналогічна до тієї, що у формулі Спірмена).

За умови відсутності зв’язку $(A = 0)$ із W_* одержуємо звичайну форму W:

$$W = \frac{12 \cdot S}{m^2\left(n^3 - n\right)}. \tag{4.45}$$

Значущість коефіцієнта детермінації (і коефіцієнта конкордації) можна оцінити у такий спосіб.

Для $n \le 7$ (малої кількості об'єктів) слід обчислити статистику

$$z = \frac{1}{2} ln \frac{(n-1)W}{1-W} \tag{4.46}$$

та порівняти її з табличними значеннями Z-розподілу Фішера з числом ступенів свободи $\nu_1 = n-1-\frac{2}{m}$; $\nu_2 = (m-1)\nu_1$.

Для $n > 7$ можна використати таблиці розподілу Пірсона χ^2, для чого слід обчислити статистику

$$\chi^2 = m(n-1)W \tag{4.47}$$

та порівняти цю величину з табличними значеннями $\chi^2_{0,05}$ і $\chi^2_{0,01}$ з $\nu = n-1$ степенями свободи.

Як уже було сказано вище, іншою мірою узгодженості оцінок m експертів може бути середнє значення коефіцієнтів рангової кореляції Спірмена $\overline{\rho}$, обчислених для кожної пари експертів (таких пар буде $\frac{m(m-1)}{2}$):

$$\overline{\rho} = \frac{1}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \rho_{ij} = 1 - \frac{12}{(m^2-m)(n^3-n)} \sum_{k=1}^{n} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} (y_{ik} - y_{jk})^2 . \tag{4.48}$$

Теорема. Між середнім коефіцієнтом рангової кореляції Спірмена і коефіцієнтом конкордації Кендела існує простий зв'язок: $\overline{\rho} = \frac{mW-1}{m-1}$.

Доведення. Перетворимо спочатку формулу для коефіцієнта конкордації:

$$W = \frac{12 \cdot S}{m^2(n^3-n)},$$

де $$S = \sum_{k=1}^{n} \left(U_k - \overline{U_k}\right)^2 = \sum_{k=1}^{n} U_k^2 - n\left(\overline{U_k}\right)^2 = \sum_{k=1}^{n} \left(\sum_{i=1}^{m} y_{ik} \right)^2 - n \left[\frac{m(n+1)}{2} \right]^2 ,$$

звідки $$W + \frac{3(n+1)}{n-1} = \frac{12}{m^2(n^2-n)} \sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} y_{ik} y_{jk} .$$

Тепер перетворюємо формулу для середнього коефіцієнта рангової кореляції Спірмена:

$$\bar{\rho} = 1 - \frac{12}{(m^2 - m)(n^3 - n)} \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=i+1}^{m} (y_{ik} - y_{jk})^2 = 1 - \frac{6}{(m^2 - m)(n^3 - n)} \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} (y_{ik} - y_{jk})^2 =$$

$$= 1 - \frac{6}{(m^2 - m)(n^3 - n)} \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} (y_{ik}^2 + y_{jk}^2) + \frac{12}{(m^2 - m)(n^3 - n)} \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} y_{ik} y_{jk} =$$

$$= 1 - \frac{6 \cdot 2}{(m^2 - m)(n^3 - n)} \left(m^2 \frac{n(n+1)(2n+1)}{6} \right) + \frac{12}{(m^2 - m)(n^3 - n)} \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} y_{ik} y_{jk} =$$

$$= \frac{-3mn - 3m - n + 1}{(m-1)(n-1)} + \frac{12}{(m^2 - m)(n^3 - n)} \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} y_{ik} y_{jk}.$$

Звідси

$$(m-1)\bar{\rho} + \frac{3mn + 3m + n - 1}{n-1} = \frac{12}{m(n^3 - 1)} \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} y_{ik} y_{jk} .$$

Порівнюючи знайдені залежності W і $\bar{\rho}$ від потрійної суми $\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m} y_{ik} y_{jk}$, встановлюємо зв'язок між $\bar{\rho}$ і W:

$$(m-1)\bar{\rho} + \frac{3m(n+1) + n - 1}{n-1} = m\left[W + \frac{3(n+1)}{n-1}\right],$$

або
$$(m-1)\bar{\rho} + 1 = mW,$$

звідки
$$\bar{\rho} = \frac{mW - 1}{m-1},$$

що і треба було довести. ■

Задача 4.1. Провести початковий етап багатовимірного аналізу даних діяльності харківського підприємства АО ХарП за порядковими ознаками.

Розв'язання задачі. Діяльність підприємства визначалась неметричними (порядковими) ознаками поданими у формі балів анкетного опитування експертів. Експертами були керівники відділу маркетингу чи економічного відділу підприємств, яким постачається продукція АО ХарПа, всього 50 осіб. У табл. 4.2 містяться 14 питань, що ставилися експертам для визначення якісних ознак діяльності підприємства АО ХарП.

Оцінки виставлялися експертами за п’ятибальною шкалою, тобто значення відповідей (порядкові ознаки) виражені ординальними числами. Практика свідчить, що в разі наявності неповних оцінок відповідей такі дані необхідно вилучити з матриці. У даному випадку кінцева матриця оцінок має дещо менші розміри 24×14(тобто є оцінки 14 якісних ознак діяльності виробника з точки зору 24 споживачів виробленої продукції).

Для обчислення скоректованих коефіцієнтів Кендела, Спірмена і конкордації спочатку слід визначити ранги значень оцінок, оскільки саме ранги значень оцінок дозволяють далі виконувати їх обчислення як з метричними величинами кількісних ознак [164].

Таблиця 4.2

Перелік питань анкетування оцінки якості виробничо-господарської діяльності підприємства

Питання	Позначення	Експерти				
		y_1	y_2	y_3	...	y_{50}
1	2	3	4	5	6	7
Оцінювання діяльності підприємства						
1. Якість продукції	x_1					
2. Асортимент продукції	x_2					
3. Освоєння нових типів продукції	x_3					
4. Своєчасність постачання	x_4					
5. Ціна на продукцію	x_5					
Оцінювання рівня обслуговування						
6. Виконання спеціальних замовлень	x_6					
7. Технічний супровід	x_7					
8. Надання інформації про можливості та хід виконання замовлення	x_8					
9. Оперативність вирішення питань	x_9					
10. Взаємодія під час вирішення проблем	x_{10}					
Рекламна підтримка						
11. Забезпечення рекламно-інформаційними матеріалами	x_{11}					

Закінчення табл. 4.2

1	2	3	4	5	6	7
12. Проведення семінарів, конференцій, зустрічей	x_{12}					
13. Оцінка web-сайта	x_{13}					
14. Оцінка авторитету підприємства	x_{14}					

У табл. 4.3 і 4.4 наведені матриці обчислених значень скоректованих коефіцієнтів рангової кореляції Спірмена (ρ_s), Кендела (τ_b) і конкордації (W_*), математичні формули яких наведені вище. Обчислення виконані для двох зрізів анкетного опитування експертів. Згідно з оцінками експертів спостерігалася тісна статистична залежність оперативності вирішення питань із взаємодією в процесі їх вирішення (коефіцієнт Спірмена $\rho_{x_9 x_{10}} = 0{,}917$), а також залежність виконання спеціальних замовлень від технічного супроводу продукції ($\rho_{x_6 x_7} = 0{,}748$) (табл. 4.3). Взаємодія працівників під час вирішення проблем надзвичайно залежить від надання інформації про можливості та хід виконання замовлення (коефіцієнт Кендела $\tau_{x_8 x_{10}} = 0{,}917$, табл. 4.4), а ціна на продукцію взаємопозв'язана з технічним супроводом ($\tau_{x_6 x_7} = 0{,}715$). Відзначається низька узгодженість питань опитування в анкеті, про що свідчить невелике значення коефіцієнта конкордації $W = 0{,}310068$. Якщо дане значення розглядати як значення коефіцієнта детермінації, як це було вже доведено, то сукупність питань, наведених в анкеті, мало відображає якісну характеристику підприємства, яка оцінюється за допомогою даної анкети.

Для аналізу відповідей експертів слід транспонувати матрицю балів та обчислити ті ж самі коефіцієнти, що й для порядкових ознак. Згідно зі значеннями коефіцієнта Спірмена (табл. 4.5) відзначено тісний кореляційний зв'язок оцінок експертів y_{37} з y_{10} ($\rho_{y_{37} y_{10}} = 0{,}8987$), y_{49} з y_{24} ($\rho_{y_{49} y_{24}} = 0{,}8594$); y_{49} з y_{26} ($\rho_{y_{49} y_{26}} = 0{,}8411$); y_{49} з y_1 ($\rho_{y_{49} y_1} = 0{,}8339$).

Матриця коефіцієнтів Кендела τ_b наведена в табл. 4.6. Значуща узгодженість величин оцінок експертів за структурною логікою значна для y_1 з y_{24} ($\tau_{y_1 y_{24}} = 0{,}88$), а також y_{10} з y_{37} ($\tau_{y_{10} y_{37}} = 0{,}876$).

За величиною коефіцієнта W можна зробити висновок про узгодженість думок експертів в оцінці вдоволеності діяльністю підприємства як виробника. Значення $W = 0{,}412$ свідчить про достатню узгодженість думок експертів в оцінці даної діяльності підприємства.

Дослідження якості діяльності підприємства з точки зору споживачів були продовжені і наступного року. Характерним залишився взаємозв'язок оперативності вирішення питань із взаємодією в процесі їх вирішення (коефіцієнт Спірмена $\rho_{x_9 x_{10}} = 0{,}8445$), яка, у свою чергу, залежить від проведення семінарів, конференцій, зустрічей, що надзвичайно природно ($\rho_{x_{10} x_{12}} = 0{,}7061$). Максимальне значення коефіцієнта Кендела для ознак по підприємству дорівнювала $\tau_{x_8 x_{10}} = 0{,}8183$, що характеризує таку ж саму структуру послідовності величин ознак, що і для підприємства, $W = 0{,}300873$.

Співпадали думки щодо діяльності підприємства експертів y_{37} та y_{45} ($\rho_{y_{37} y_{45}} = 0{,}899$), а також y_{44} та y_{46} ($\rho_{y_{44} y_{46}} = 0{,}835$). Для названих експертів y_{37} та y_{45} співпадає і структурна послідовність значень балів, про що свідчить значення коефіцієнта Кендела ($\tau_{y_{37} y_{45}} = 0{,}876$), для y_{44} та y_{46} ($\tau_{y_{44} y_{46}} = 0{,}791$). У цілому ж думка експертів в якісній характеристиці виробника їх продукції була погано узгодженою $W = 0{,}161138$, тобто наявні найрізноманітніші думки з даного приводу.

Таблиця 4.3

Матриця коефіцієнтів Спірмена
(перший рік досліджень)

ρ	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_1	1	0,24	0,35	0,262	0,411	0,275	0,187	0,221	0,324	0,41	0,268	0,159	0,02	0,26
x_2		1	0,637	0,529	0,447	0,161	0,196	0,377	0,218	0,143	0,114	0,267	0,061	0,222
x_3			1	0,15	0,391	0,368	0,306	0,136	0,368	0,354	0,414	0,362	0,273	0,149
x_4				1	0,507	-0	-0,04	0,24	0,346	0,302	0,227	0,035	-0,16	0,214
x_5					1	-0	0,161	0,146	0,228	0,192	0,074	0,08	-0,12	0,331
x_6						1	0,748	0,05	-0,15	-0,01	0,248	0,085	0,333	0,155
x_7							1	0,274	-0,05	0,068	0,117	0,156	0,185	0,418

Закінчення табл. 4.3

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_8								1	0,298	0,447	-0,02	0,337	0,296	0,364
x_9									1	**0,917**	0,313	0,572	0,206	0,288
x_{10}										1	0,428	0,535	0,303	0,323
x_{11}											1	0,432	0,316	0,223
x_{12}												1	0,636	0,353
x_{13}													1	0,329
x_{14}														1
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}

Таблиця 4.4

Матриця коефіцієнтів Кендела
(перший рік досліджень)

τ	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_1	1	0,24	0,34	0,26	0,41	0,26	0,18	0,21	0,32	0,41	0,26	0,15	0,02	0,25
x_2		1	0,61	0,53	0,45	0,15	0,19	0,37	0,22	0,14	0,11	0,26	0,06	0,21
x_3			1	0,14	0,37	0,34	0,28	0,13	0,35	0,34	0,39	0,33	0,25	0,14
x_4				1	0,51	0	-0,03	0,23	0,35	0,30	0,22	0,03	-0,15	0,21
x_5					1	0	0,15	0,14	0,23	0,19	0,07	0,08	-0,12	0,32
x_6						1	0,72	0,05	-0,14	-0,01	0,22	0,08	0,31	0,15
x_7							1	0,26	-0,05	0,06	0,11	0,14	0,16	0,40
x_8								1	0,29	0,43	-0,02	0,30	0,28	0,34
x_9									1	**0,92**	0,30	0,55	0,20	0,28
x_{10}										1	0,42	0,51	0,29	0,31
x_{11}											1	0,40	0,29	0,22
x_{12}												1	0,60	0,33
x_{13}													1	0,30
x_{14}														1
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}

Таблиця 4.5

Матриця коефіцієнтів Спірмена (перший рік досліджень)

ρ	y_1	y_2	y_3	y_5	y_7	y_8	y_9	y_{10}	y_{11}	y_{15}	y_{16}	y_{22}	y_{23}	y_{24}	y_{26}	y_{28}	y_{35}	y_{37}	y_{40}	y_{44}	y_{45}	y_{46}	y_{49}	y_{50}
y_1	1	0,47	0,6	0,26	0,57	0,64	0,6	0,75	0,65	0,78	0,59	0,29	0,8	**0,92**	**0,85**	0,59	0,6	0,61	0,37	0,28	0,35	0,56	**0,83**	0,23
y_2		1	0,78	-0,3	0,14	0,4	0,28	0,35	0,21	0,22	0,39	0,41	0,54	0,31	0,48	0,06	0,28	0,29	0,36	-0,3	0,17	0,55	0,29	0,26
y_3			1	-0	0,34	0,39	0,15	0,45	0,39	0,28	0,35	0,17	0,69	0,47	0,28	0,08	0,58	0,19	0,14	-0	0,21	0,34	0,46	0,33
y_5				1	0,4	0,21	-0	0,04	0,49	0,08	-0,2	-0,1	0,45	0,45	0,13	0,35	0,34	0,07	-0	0,54	-0,1	0,38	0,49	0,14
y_7					1	0,49	0,34	0,41	0,38	0,52	0,43	0,48	0,5	0,64	0,43	0,66	0,02	0,36	0,55	0,55	0,23	0,13	0,63	0,07
y_8						1	0,63	0,72	0,43	0,34	0,54	0,2	0,55	0,7	0,58	0,75	0,39	0,7	0,72	0,1	0,23	0,42	0,4	0,09
y_9							1	0,45	0,39	0,66	0,35	0,52	0,27	0,64	0,74	0,63	0,15	**0,9**	0,7	0,1	0,21	0,34	0,3	-0,1
y_{10}								1	0,44	0,47	0,63	-0	0,46	0,65	0,64	0,63	0,45	0,52	0,49	0,04	0,47	0,34	0,53	-0,2
y_{11}									1	0,58	0,31	0,26	0,56	0,69	0,51	0,61	0,62	0,49	0,19	0,62	0,3	0,49	0,55	-0,1
y_{15}										1	0,65	0,62	0,54	0,71	0,75	0,57	0,28	0,68	0,29	0,41	0,44	0,32	0,69	-0
y_{16}											1	0,35	0,44	0,38	0,54	0,6	0,08	0,49	0,27	0	0,3	0,35	0,39	-0,1
y_{22}												1	0,28	0,33	0,46	0,35	-0,2	0,56	0,56	0,28	0,41	0,15	0,34	-0
y_{23}													1	0,78	0,58	0,31	0,69	0,37	0,18	0,22	0,21	0,63	**0,86**	0,54
y_{24}														1	0,77	0,64	0,64	0,65	0,49	0,49	0,39	0,45	**0,84**	0,26
y_{26}															1	0,63	0,28	0,71	0,55	0,09	0,3	0,7	0,69	0,05
y_{28}																1	0,08	0,73	0,66	0,47	0,3	0,35	0,39	-0,3
y_{35}																	1	0,19	-0,1	0,24	0,21	0,34	0,55	0,33
y_{37}																		1	0,7	0,18	0,48	0,37	0,38	-0,1
y_{40}																			1	0,1	0,39	0,18	0,23	-0,2
y_{44}																				1	0,34	-0,2	0,41	-0,1
y_{45}																					1	-0,1	0,43	-0,3
y_{46}																						1	0,49	0,14
y_{49}																							1	0,31
y_{50}																								1
	y_1	y_2	y_3	y_5	y_7	y_8	y_9	y_{10}	y_{11}	y_{15}	y_{16}	y_{22}	y_{23}	y_{24}	y_{26}	y_{28}	y_{35}	y_{37}	y_{40}	y_{44}	y_{45}	y_{46}	y_{49}	y_{50}

Таблиця 4.6

Матриця коефіцієнтів Кендела (перший рік досліджень)

τ	y_1	y_2	y_3	y_5	y_7	y_8	y_9	y_{10}	y_{11}	y_{15}	y_{16}	y_{22}	y_{23}	y_{24}	y_{26}	y_{28}	y_{35}	y_{37}	y_{40}	y_{44}	y_{45}	y_{46}	y_{49}	y_{50}
y_1		0,47	0,6	0,26	0,56	0,62	0,6	0,75	0,61	0,75	0,57	0,29	0,75	**0,88**	**0,82**	0,57	0,6	0,59	0,35	0,27	0,35	0,56	0,78	0,23
y_2		1	0,78	-0,3	0,14	0,38	0,28	0,35	0,2	0,21	0,37	0,41	0,51	0,29	0,47	0,06	0,28	0,28	0,34	-0,3	0,17	0,55	0,27	0,26
y_3			1	-0	0,33	0,37	0,15	0,45	0,37	0,27	0,34	0,17	0,65	0,45	0,27	0,07	0,58	0,19	0,13	-0	0,21	0,34	0,43	0,33
y_5				1	0,39	0,2	-0	0,04	0,46	0,08	-0,1	-0,1	0,43	0,43	0,13	0,33	0,34	0,07	-0	0,51	-0,1	0,38	0,46	0,14
y_7					1	0,46	0,33	0,4	0,34	0,5	0,42	0,47	0,46	0,59	0,41	0,63	0,02	0,34	0,51	0,51	0,22	0,12	0,57	0,07
y_8						1	0,6	0,69	0,4	0,31	0,51	0,19	0,5	0,66	0,53	0,71	0,37	0,67	0,68	0,09	0,22	0,4	0,33	0,08
y_9							1	0,45	0,37	0,63	0,34	0,52	0,26	0,61	0,72	0,61	0,15	**0,88**	0,67	0,09	0,21	0,34	0,28	-0,1
y_{10}								1	0,41	0,45	0,61	0	0,43	0,62	0,62	0,61	0,45	0,51	0,46	0,04	0,47	0,34	0,49	-0,2
y_{11}									1	0,51	0,26	0,25	0,52	0,66	0,48	0,56	0,58	0,45	0,19	0,56	0,28	0,46	0,5	-0,1
y_{15}										1	0,61	0,59	0,45	0,66	0,72	0,54	0,27	0,65	0,25	0,36	0,42	0,31	0,6	0
y_{16}											1	0,34	0,4	0,33	0,5	0,57	0,07	0,47	0,25	0	0,29	0,33	0,34	-0,1
y_{22}												1	0,27	0,32	0,45	0,34	-0,2	0,54	0,53	0,26	0,41	0,15	0,31	0
y_{23}													1	0,71	0,51	0,26	0,65	0,34	0,16	0,2	0,2	0,59	**0,8**	0,51
y_{24}														1	0,73	0,6	0,61	0,62	0,47	0,43	0,37	0,43	0,76	0,25
y_{26}															1	0,6	0,27	0,68	0,51	0,09	0,29	0,68	0,63	0,05
y_{28}																1	0,07	0,7	0,63	0,41	0,29	0,33	0,34	-0,3
y_{35}																	1	0,19	-0,1	0,23	0,21	0,34	0,51	0,33
y_{37}																		1	0,66	0,16	0,47	0,36	0,34	-0,1
y_{40}																			1	0,1	0,37	0,17	0,2	-0,2
y_{44}																				1	0,32	-0,2	0,36	-0,1
y_{45}																					1	-0,1	0,4	-0,3
y_{46}																						1	0,46	0,14
y_{49}																							1	0,29
y_{50}																								1
	Y_1	y_2	y_3	y_5	y_7	y_8	y_9	y_{10}	y_{11}	y_{15}	y_{16}	y_{22}	y_{23}	y_{24}	y_{26}	y_{28}	y_{35}	y_{37}	y_{40}	y_{44}	y_{45}	y_{46}	y_{49}	y_{50}

Аналіз системи елементарних порядкових ознак СЕС слід продовжити дисперсійним аналізом.

Зазвичай "дисперсійний аналіз передбачає, що експеримент організований відповідно до плану, який дозволяє при порівняно малій кількості вимірювань незалежно оцінити вплив кожного з факторів на вимірюваний кількісний показник. Застосування дисперсійного аналізу у ході пасивних спостережень, наприклад, в економіці, передбачає належний відбір груп даних з більш значної сукупності даних" [275, с. 131]. У класичному дисперсійному аналізі вважається, що вихідні дані вимірюються за метричними шкалами та є кількісними. Це – найпоширеніший і добре відпрацьований апарат дисперсійного аналізу.

Проте, незважаючи на широке практичне застосування в економіці, існує істотний недолік класичного дисперсійного аналізу – це обтяжливе припущення про нормальний розподіл значень кількісного показника в кожній групі (вибірці) спостережень. Рангові методи значною мірою позбавлені цих недоліків, вони добре пристосовані для роботи з малими вибірками, розподілів яких ми не знаємо. Зрозуміло, що перехід від самих спостережень (x) до їх рангів (u) супроводжується певною втратою інформації, але відразу слід зазначити, що ця втрата невелика (цей оптимістичний висновок випливає з порівняння рангових правил з точними, коли відомий закон розподілу похибок). Після ранжування (переходу до рангів) статистичні висновки стають більш загальними та надійними.

Розглянемо однофакторну схему дисперсійного аналізу Краскала − Уолліса, яка є узагальненням звичайного дисперсійного аналізу на рангових змінних [162; 294]. У дисперсійному аналізі порівнюються k вибірок спостережень та перевіряється "нуль-гіпотеза" про те, що всі дані були відібрані з однієї сукупності. Вибірки сформовані так, щоб вони відповідали різним станам певної якісної ознаки. Якщо "нуль-гіпотеза" відхиляється (між вибірками є істотні відмінності), то далі слід встановити, між якими вибірками є найістотніші відмінності, тобто потрібно виявити причину значущих відмінностей.

Якщо дані у вибірці були початково виміряні в метричних шкалах, то потрібно ранжувати множини всіх даних як єдиної вибірки обсягу $N = \sum_{j=1}^{k} n_j$.

Упорядковані дані мають номери від 1 до N. Загальна сума рангів обчислюється за формулою:

$$\sum_{i=1}^{N} u_i = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}. \tag{4.49}$$

Сума квадратів рангів обчислюється за відомою формулою:

$$\sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}. \tag{4.50}$$

Далі потрібно буде обчислити ще подвоєну суму добутків рангів $2\sum_{i<j} u_i u_j$.

Розглянемо вираз:

$$\left(\sum u_i\right)^2 = \sum u_i^2 + 2\sum_{i<j} u_i u_j , \tag{4.51}$$

звідки

$$\begin{aligned} 2\sum_{i<j} u_i u_j &= \left(\sum u\right)^2 - \sum u^2 = \left[\frac{N(N+1)}{2}\right]^2 - \frac{N(N+1)(2N+1)}{6} = \\ &= \frac{N(N+1)}{12}\left[3N(N+1) - 2(2N+1)\right] = \\ &= \frac{N(N+1)}{12}\left(3N^2 - N - 2\right) = \frac{(N-1)N(N+1)}{12}(3N+2). \end{aligned} \tag{4.52}$$

Відомо, що "нуль-гіпотеза" полягає у припущенні, що в кожну вибірку елементи відбираються випадковим чином.

Обчислимо математичне сподівання та дисперсію суми рангів об’єктів у вибірці обсягом n (за наявності "нуль-гіпотези").

Розглянемо випадкову величину – суму рангів об’єктів, що відібрані у вибірку випадковим чином:

$$z = \sum_{i=1}^{N} \alpha_i u_i , \tag{4.53}$$

де $\alpha_i = 1$, якщо x_i потрапило до вибірки, і $\alpha_i = 0$, якщо x_i не потрапило до вибірки.

Імовірність того, що елемент випадковим чином буде відібраний до вибірки, дорівнює $p_i = \frac{n}{N}$, а два елементи u_i, u_j будуть відібрані до вибірки з імовірністю $\frac{n}{N} \cdot \frac{n-1}{N-1}$. Інакше кажучи, нам відомі розподіли α_i та $\alpha_i \alpha_j$, що зображені рядами розподілу (рис. 4.1).

α_i	0	1
p_i	$1-\frac{n}{N}$	$\frac{n}{N}$

$\alpha_i \alpha_j$	0	1
p_{ij}	$1-\frac{n}{N} \cdot \frac{n-1}{N-1}$	$\frac{n}{N} \cdot \frac{n-1}{N-1}$

Рис. 4.1. **Закони розподілу α_i та $\alpha_i \alpha_j$**

Звідси маємо:

$$M(\alpha_i) = 0 \cdot \left(1-\frac{n}{N}\right) + 1 \cdot \frac{n}{N} = \frac{n}{N}; \tag{4.54}$$

$$M(\alpha_i^2) = 0^2 \cdot \left(1-\frac{n}{N}\right) + 1^2 \cdot \frac{n}{N} = \frac{n}{N}; \tag{4.55}$$

$$D(\alpha_i) = M(\alpha_i^2) - M^2(\alpha_i) = \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \frac{n}{N}\left(1-\frac{n}{N}\right); \tag{4.56}$$

$$M(\alpha_i \alpha_j) = 0 \cdot \left(1-\frac{n}{N} \cdot \frac{n-1}{N-1}\right) + 1 \cdot \frac{n}{N} \cdot \frac{n-1}{N-1} = \frac{n}{N} \cdot \frac{n-1}{N-1}; \tag{4.57}$$

$$\begin{aligned} cov(\alpha_i \alpha_j) &= M(\alpha_i \alpha_j) - M(\alpha_i) M(\alpha_j) = \frac{n}{N} \cdot \frac{n-1}{N-1} - \left(\frac{n}{N}\right)^2 = \\ &= \frac{n}{N}\left(\frac{n-1}{N-1} - \frac{n}{N}\right) = -\frac{n}{N}\left(1-\frac{n}{N}\right) \cdot \frac{1}{N-1} \ . \end{aligned} \tag{4.58}$$

Скориставшись властивостями математичного сподівання, обчислюємо:

$$M(z)=M\left(\sum\alpha_i u_i\right)=\sum_{i=1}^{N}u_i M(\alpha_i)=\sum_{i=1}^{N}u_i\cdot\frac{n}{N}=\frac{n}{N}\cdot\sum_{i=1}^{N}u_i=$$
$$=\frac{n}{N}\cdot\frac{N(N+1)}{2}=\frac{n(N+1)}{2}\ . \tag{4.59}$$

Відомо, що дисперсія суми випадкової величини дорівнює сумі їх дисперсій плюс подвоєні коваріації:

$$D(z)=D\left(\sum\alpha_i u_i\right)=\sum_{i=1}^{N}u_i^2 D(\alpha_i)+2\sum_{1\le i<j\le N}u_i u_j \operatorname{cov}(\alpha_i,\alpha_j)=$$
$$=\frac{n}{N}\left(1-\frac{n}{N}\right)\sum_{i=1}^{N}u_i^2-\frac{n}{N}\left(1-\frac{n}{N}\right)\frac{1}{N-1}\cdot 2\sum_{i<j}u_i u_j=$$
$$=\frac{n}{N}\left(1-\frac{n}{N}\right)\cdot\frac{N(N+1)(2N+1)}{6}-\frac{n}{N}\left(1-\frac{n}{N}\right)\frac{1}{N-1}\cdot\frac{(N-1)N(N+1)(3N+2)}{12}=$$
$$=\frac{n}{N}\left(1-\frac{n}{N}\right)\frac{N(N+1)}{12}\left[(4N+2)-(3N+2)\right]=n\left(1-\frac{n}{N}\right)\frac{N(N+1)}{12}.$$

Отже,

$$D(z)=n\left(1-\frac{n}{N}\right)\frac{N(N+1)}{12}. \tag{4.60}$$

Для кожної з k вибірок обсягами $n_1,n_2,\ldots,n_k$ можна обчислити математичні сподівання та дисперсії середніх рангів (за наявності "нуль-гіпотези"):

$$M(\bar{z}_j)=\frac{1}{n_j}M(z_j)=\frac{1}{n_j}\cdot\frac{n_j(N+1)}{2}=\frac{N+1}{2}, \tag{4.61}$$

$$D(\bar{z}_j)=\frac{1}{n_j^2}D(z_j)=\left(1-\frac{n_j}{N}\right)\cdot\frac{N(N+1)}{12n_j}, \tag{4.62}$$

де j – номер вибірки.

Відомо, що сума квадратів стандартизованих нормально розподілених величин w_j має розподіл χ^2:

$$\chi^2=\sum_{j=1}^{k}\left[\frac{w_j-M(w_j)}{\sqrt{D(w_j)}}\right]^2. \tag{4.63}$$

Проте середні ранги за групами $\overline{z_j}$ мають дещо інший розподіл. Краскал і Уолліс довели, що аналогічна статистика H_i (з поправковими коефіцієнтами $1-\dfrac{n_j}{N}$)

$$H=\sum_{j=1}^{k}\left(1-\frac{n_j}{N}\right)\left[\frac{\overline{z_j}-M(\overline{z_j})}{\sqrt{D(\overline{z_j})}}\right]^2 \qquad (4.64)$$

має розподіл $\chi^2(k-1)$.

Число ступенів свободи (ЧСС) дорівнює $k-1$, оскільки склад однієї з вибірок завжди відомий, у сумі ранги елементів усіх вибірок складають множину номерів $1,\ 2,\ ...,\ N$.

Необхідно перетворити статистику Краскала − Уолліса H до простішого вигляду:

$$H=\sum_{j=1}^{k}\left(1-\frac{n_j}{N}\right)\frac{\left(\overline{z_j}-\dfrac{N+1}{2}\right)^2}{\left(1-\dfrac{n_j}{N}\right)\dfrac{N(N+1)}{12n_j}}=\frac{12}{N(N+1)}\sum_{j=1}^{k}n_j\left[\overline{z_j}-\frac{N+1}{2}\right]^2, \qquad (4.65)$$

де $\sum\limits_{i=1}^{k}z_i=\sum\limits_{i=1}^{N}i=\dfrac{N(N+1)}{2}$.

Виконуємо подальші перетворення:

$$H=\frac{12}{N(N+1)}\sum_{j=1}^{k}n_j\left[\overline{z_j}-\frac{N+1}{2}\right]^2=\frac{12}{N(N+1)}\left\{\sum\frac{z_j^2}{n_j}-2\frac{N+1}{2}\sum z_j+\left(\frac{N+1}{2}\right)^2\sum n_j\right\}=$$

$$=\frac{12}{N(N+1)}\left\{\sum\frac{z_j^2}{n_j}-2\frac{N(N+1)^2}{4}+\frac{N(N+1)^2}{4}\right\}=\frac{12}{N(N+1)}\sum\frac{z_j^2}{n_j}-$$

$$-\frac{12}{N(N+1)}\frac{N(N+1)^2}{4}=\frac{12}{N(N+1)}\sum\frac{z_j^2}{n_j}-3(N+1).$$

Отже, одержуємо

$$H = \frac{12}{N(N+1)} \sum \frac{z_j^2}{n_j} - 3(N+1), \tag{4.66}$$

де z_j – сума рангів у групі j.

Якщо виявиться, що обчислене значення H буде більшим від табличного $H > \chi^2_{0,01}(k-1)$, то "нуль-гіпотеза" відхиляється і потрібно виконати попарне порівняння всіх вибірок для визначення причини значущості. Усього потрібно виконати $\frac{k(k-1)}{2}$ попарних порівнянь середніх рангів за групами.

Стандартизовану величину різниці середніх рангів двох груп, обчислену за формулою:

$$t = \frac{\left|\overline{z_j} - \overline{z_i}\right|}{\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}, \tag{4.67}$$

потрібно порівняти з квантилями нормального розподілу $t_{0,05} = 1{,}96 \approx 2$ та $t_{0,01} = 2{,}58$. Різниці вважають значущими, якщо буде $t > t_{0,05}$ (ЧСС $n_i + n_j - 2$).

У дисперсійному аналізі Фрідмена оцінки експертів ранжуються для кожного експерта окремо, а не для всього масиву спостережень, як за схемою Краскала − Уолліса. Нехай n експертів оцінюють k об’єктів. Після ранжування оцінок кожного експерта одержуємо табл. 4.7, де ранги R_{ij} у кожному рядку є різними перестановками чисел 1, 2, …, k. У сумарному рядку табл. 4.7 обчислені суми рангів у кожному стовпчику (для кожного об’єкта) $R_{0j} = \sum_{i=1}^{n} R_{ij}$.

Таблиця 4.7

Ранги оцінок кожного експерта

Експерти	Об’єкти				
	1	2	3	…	k
1	R_{11}	R_{12}	R_{13}		R_{1k}
2	R_{21}	R_{22}	R_{23}		R_{2k}

3	R_{31}	R_{32}	R_{33}		R_{3k}
…					
n	R_{n1}	R_{n2}	R_{n3}		R_{nk}
Суми	R_{01}	R_{02}	R_{03}		R_{0k}

Розглянемо статистичні характеристики рангів за нуль-гіпотези (тобто коли немає відмінності між об'єктами). Оскільки для кожного експерта ранги $U_j = R_{ij}$ набувають послідовних значень від 1 до k і всі перестановки цих рангів рівноймовірні, то всі $p(R_{ij}) = \frac{1}{k}$. Тому

$$M(R_{ij}) = 1 \cdot \frac{1}{k} + 2 \cdot \frac{1}{k} + \ldots + k \cdot \frac{1}{k} = \frac{1}{k}\sum_{j=1}^{k} j = \frac{k(k+1)}{2k} = \frac{k+1}{2}, \qquad (4.68)$$

$$M(R_{ij}^2) = 1 \cdot \frac{1}{k} + 2^2 \cdot \frac{1}{k} + \ldots + k^2 \cdot \frac{1}{k} = \frac{1}{k}\sum_{j=1}^{k} j^2 = \frac{k(k+1)(2k+1)}{6k} = \frac{(k+1)(2k+1)}{6}, \qquad (4.69)$$

$$D(R_{ij}^2) = \frac{(k+1)(2k+1)}{6} - \left[\frac{k+1}{2}\right]^2 = \frac{k+1}{12}\left[2(2k+1) - 3(k+1)\right] = \frac{k^2-1}{12}. \qquad (4.70)$$

Але ранги в кожному рядку не є незалежними і потрібно обчислити їх коваріацію $\operatorname{cov}(R_{mi}, R_{mj}) = M(R_{mi}, R_{mj}) - M(R_{mi})M(R_{mj})$. У випадку парних порівнянь маємо $k(k-1)$ рівноймовірних пар, тому $p(R_{mi}, R_{mj}) = \frac{1}{k(k-1)}$.

Обчислюємо суму добутків:

$$M(R_{mi}, R_{mj}) = 2\sum_{i<j} R_{mi}, R_{mj} \cdot \frac{1}{k(k-1)} = \frac{k(k+1)(k-1)(3k+2)}{12 \cdot k(k-1)} = \frac{(k+1)(3k+2)}{12}. \qquad (4.71)$$

(Сума $2\sum_{i<j} u_i u_j$, де $u_i, u_j \in \{1, 2, \ldots, k\}$ була обчислена раніше).

Отже,

$$\operatorname{cov}(R_{mi}, R_{mj}) = \frac{(k+1)(3k+2)}{12} - \left[\frac{k+1}{2}\right]^2 = \frac{k+1}{2}\left[3k+2 - 3(k+1)\right] = -\frac{k+1}{12}. \qquad (4.72)$$

Для суми рангів R_{0j} маємо

$$M(R_{0j}) = \frac{k+1}{2},\ D(R_{0j}) = n\frac{k^2-1}{12},\ \mathrm{cov}(R_{mi}, R_{mj}) = -n\frac{k+1}{12}. \tag{4.73}$$

$$D(R_{0j} - R_{0i}) = D(R_{0j}) + D(R_{0j}) - 2\mathrm{cov}(R_{0j}, R_{oi}) =$$

$$= n\frac{k^2-1}{12} + n\frac{k^2-1}{12} + 2n\frac{k+1}{12} =$$

$$= n\frac{k^2-1}{6} + n\frac{k+1}{6} = \frac{n(k+1)}{6}[(k-1)+1] = \frac{nk(k+1)}{6}.$$

$$D(R_{0j} - R_{0i}) = n\frac{k(k+1)}{6}, \tag{4.74}$$

тому різниця між двома групами визнається значущою, якщо

$$\left|R_{0j} - R_{0i}\right| \geq Z_{\frac{\alpha}{2}}\sqrt{\frac{nk(k+1)}{6}}, \tag{4.75}$$

де $Z_{\frac{\alpha}{2}}$ − квантиль нормального розподілу (для $\alpha = 0{,}01$ $Z = 2{,}58$).

Далі (як і за схемою Краскала − Уолліса) складаємо статистику:

$$K = \sum_{j=1}^{k} C_j^2 \left\{\frac{R_{0j} - M(R_{0j})}{\sqrt{D(R_{0j})}}\right\}^2 = \sum_{j=1}^{k} C_j^2 \frac{12}{n(k^2-1)}\left[R_{0j} - \frac{n(k+1)}{2}\right]^2, \tag{4.76}$$

де вагові коефіцієнти C_j^2 підбираються так, щоб K було асимптотично розподілене як $\chi^2(k-1)$. Фрідмен показав, що для цього потрібно задавати $C_j^2 = 1 - \frac{1}{k}$ (за схемою Краскала − Уолліса $C_j^2 = 1 - \frac{n_j}{N}$, що для однакових $n_j = n$ і $N = nk$ приводить до того самого виразу $C_j^2 = 1 - \frac{1}{k}$).

Найчастіше перевірка узгодженості оцінок експертів здійснюється за допомогою обчислення коефіцієнта конкордації W. Якщо виявиться, що сукупність оцінок різних експертів не узгоджена, то слід перевірити $\frac{m(m-1)}{2}$ пар експертів і знайти групи експертів, що мають різні точки зору. Для цього слід застосувати процедуру – тест Краскала − Уолліса, яка реалізується в ППП Statgraphics за допомогою активації меню Describe/Categorical Data/Contingency

Tables. Як приклад, у табл. 4.8 наведені обчислення для всіх ознак (перший рік дослідження діяльності підприємства з позиції споживача).

Таблиця 4.8

Тест Краскала − Уоліса

Ознака	Групи експертів	Кількість експертів	Середній ранг	Значення H
x_1	1	11	15,6667	6,13333
	2	12	9,33333	$p-value = 0,0132634$
x_2	1	11	15,3182	4,41618
	2	13	10,1154	$p-value = 0,0355963$
x_3	1	11	16,4545	8,5431
	2	13	9,15385	$p-value = 0,00346722$
x_4	1	11	14,5	3,89231
	2	13	10,8077	$p-value = 0,0485042$
x_5	1	11	14,0	2,78022
	2	13	11,2308	$p-value = 0,0954314$
x_6	1	11	15,2273	4,32643
	2	13	10,1923	$p-value = 0,0375215$
x_7	1	11	15,4091	4,20285
	2	13	10,0385	$p-value = 0,0403527$
x_8	1	11	14,0455	1,26771 $p-value = 0,260195$
	2	13	11,1923	
x_9	1	11	15,9091	6,70163
	2	13	9,61538	$p-value = 0,00963052$
x_{10}	1	11	16,4091	8,4969
	2	13	9,19231	$p-value = 0,00355637$
x_{11}	1	11	16,5214	7,8795
	2	13	9,18564	$p-value = 0,0053453$
x_{12}	1	11	16,5	7,78462
	2	13	9,11538	$p-value = 0,00526787$
x_{13}	1	11	14,9545	2,97782
	2	13	10,4231	$p - value = 0,0844091$
x_{14}	1	11	16,4545	7,72947
	2	13	9,15385	$p-value = 0,00543121$

За обчисленим значенням Краскала − Уоліса $H = 6{,}13333$ для визначення узгодженості думок експертів у розрізі якості продукції (x_1), для якого p-значення дорівнює 0,0132634 < 0,05, "нуль-гіпотеза" відхиляється (між вибірками є істотні відмінності), тому потрібно виявити причину значущих відмінностей.

Описані вище процедури аналізу були виконані для всіх порядкових ознак діяльності підприємства АО ХарП. За результатами дисперсійного аналізу порядкових ознак маємо однорідність суджень експертів за якістю продукції (x_1), асортиментом продукцію (x_2), освоєння нових типів продукції (x_3), ціни на продукції (x_5), виконання спеціальних замовлень (x_6). На дві групи розділилися експерти за своїми позиціями щодо своєчасності постачання (x_4), технічного супроводження (x_7), надання інформації про можливості та хід виконання замовлення (x_8), оперативності вирішення питань (x_9), взаємодії під час вирішення проблем (x_{10}), забезпечення рекламно-інформаційними матеріалами (x_{11}), проведення семінарів, конференцій, зустрічей (x_{12}), оцінки web-сайта (x_{13}), оцінки авторитету підприємства (x_{14}).

Чотири ознаки, за якими сукупність експертів однорідна, є визначальними та задовільними в оцінці діяльності підприємства; за рештою ознак думки експертів не співпадають. Це потрібно врахувати у процесі формування маркетингової стратегії підприємства.

Аналіз елементарних порядкових ознак СЕС, їх взаємозв’язку є першим етапом в моделюванні складних порядкових ознак систем.

4.2. Початковий аналіз соціально-економічних систем за номінальними ознаками

Незважаючи на те що багато ознак соціально-економічних систем вимірюються на номінальних шкалах, розбудова моделей, які передбачають вихідні дані категорії чи назви, є великою невирішеною проблемою в економіко-математичному моделюванні, оскільки номінальні шкали передбачають лише операції $x = y$ або $x \neq y$. Відомі спеціалісти, які розробляють дану проблему, Ю. Н. Тюрін і А. А. Макаров рекомендують дослідження ознак розпочинати з визначення типу даних, тобто їх віднесення до тієї чи іншої шкали вимірювання, а на другому етапі − перевіряти гіпотезу про відсутність зв’язку (незалежності) між ознаками [276, с. 243 – 257]. Якщо гіпотеза про незалежність ознак відхиляється, то, зазвичай, встановлюють ступінь сили зв’язку ознак за допомогою різних коефіцієнтів. Якщо модуль міри зв’язку належить інтервалу від 0,8 до 1, то це свідчить про сильний зв’язок ознак, якщо він знаходиться в інтервалі [0,3, 0,7] – про нечітко виражений зв’язок; міра зв’язку, що близька до 0, свідчить про відсутність залежності або дуже слабку залежність ознак. Отже, у розбудові моделей за номінальними ознаками теж спочатку слід встановити незалежність ознак та визначити міру їх зв’язку.

В аналізі номінальних даних сформована на основі теоретико-логічного аналізу система елементарних ознак СЕС, що має, перш за все, досліджуватися на предмет визначення взаємозв’язків, і лише за таких умов можна говорити про існування адекватної системи номінальних ознак у характеристиці системи. Якщо наявне співвідношення між ознаками, які якісно варіюються, то існують їх асоціації, контингенції. Точніше, коли ознаки об’єкта приймають номінацію у двох категоріях, наприклад, "так – ні", то за наявності взаємозв’язків між такими якісними ознаками говорять про їх асоціації, іноді використовують також термін "кореляція". Якщо ж ознаки об’єкта можуть набувати номінацію в більш ніж двох категоріях, то зв’язок між ними називають "контингенцією" (сполученістю).

Для розбудови моделей складних номінальних ознак соціально-економічної системи був проведений аналіз методів, що чисельно оцінюють зв’язок між ознаками даного типу [9; 30; 31; 113; 137; 188; 291]. Основні висновки такі.

Нехай перша ознака соціально-економічної системи може набувати p різних категорій, а друга ознака – q різних категорій. У таблиці контингенції записують числа m_{ij}, які характеризують частоту появи пари ознак $A_i - B_j$. Позначимо суми частот у стовпчиках таблиці через k_i $(k_i = m_{i0})$, а в рядках –

через l_j $(l_j = m_{0j})$; позначимо суми частот через n $(n = m_{00})$, як це показано в табл. 4.9.

Таблиця 4.9

Умовне подання категорій двох номінальних ознак

	A_1	A_2	...	A_p	$\sum$
B_1	m_{11}	m_{21}	...	m_{p1}	$l_1 = m_{01}$
B_2	m_{12}	m_{22}	...	m_{p2}	$l_2 = m_{02}$
...	...	...	...	...	...
B_q	m_{1q}	m_{2q}	...	m_{pq}	$l_q = m_{0q}$
$\sum$	$k_1 = m_{10}$	$k_2 = m_{20}$	...	$k_p = m_{p0}$	$n = m_{00}$

Зауваження: $k_i = \sum_j m_{ij}$, $l_j = \sum_i m_{ij}$, $n = \sum_i \sum_j m_{ij}$.

Відносні частоти $\frac{k_i}{n}$ і $\frac{l_j}{n}$ є оцінками ймовірностей появи категорій ознак A_i і B_j. За незалежності ознак імовірність їх сумісної появи дорівнює добутку ймовірностей, звідки можна отримати очікувані частоти $\hat{m}_{ij} = n \cdot \frac{k_i}{n} \cdot \frac{l_j}{n} = \frac{k_i l_j}{n} = \frac{m_{i0} m_{0j}}{n}$ (при правильності гіпотези про відсутність зв'язку між ознаками A і B).

За допомогою критерію χ^2 можна оцінити значущість відмінностей між частотами, що спостерігаються, та очікуваними частотами:

$$\chi^2 = \sum_i \sum_j \frac{(m_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \quad \text{або} \quad \chi^2 = \sum_i \sum_j \frac{m_{ij}^2}{\hat{m}_{ij}} - n. \qquad (4.77)$$

Ця статистика (характеристика) має χ^2-розподіл з числом степенів свободи, що дорівнює $\nu = (p-1)(q-1)$.

Якщо $\chi^2 < \chi^2_{0,05}(\nu)$, то "нуль-гіпотеза" про відсутність зв'язку між ознаками не може бути відхилена. Якщо $\chi^2 > \chi^2_{0,01}(\nu)$, то "нуль-гіпотеза" відхиляється, тобто між ознаками A і B є значущий зв'язок. Потрібно оцінити тісноту цього зв'язку. Слід зауважити, що тіснота і значущість – різні речі.

Максимального значення χ^2 набуде при "функціональній" залежності, коли кожній категорії A відповідає лише одна категорія B, тобто коли в кожному рядку і в кожному стовпчику таблиці контингенції наявна лише одна клітинка з $m_{ij} \neq 0$. Завжди можна переставити категорії A_i і B_j так, щоб заповнені клітинки розмістилися блоками по діагоналі таблиці: $m_{ii} \neq 0$ і $m_{ij} = 0$ для $i \neq j$, $p = q$.

Для $p \neq q$ маємо $d = \min(p, q)$. В окремих роботах як, наприклад, Е. Ферстера, Б. Ренца [284], рекомендується користуватися середнім для обчислення $d = \dfrac{p+q}{2}$. Це не правильно. Обґрунтуємо вибір $d = \min(p, q)$ на прикладі $p = 4, q = 3$. Тоді $p > q$, при найтіснішому зв'язку кожному значенню x_i (кожній категорії цього показника) буде відповідати єдине значення y_j (лише одна категорія іншого показника). Оскільки для неметричних ознак категорії можна переставляти місцями, таблицю спряженості 4×3 можна звести до блочно-діагонального вигляду (табл. 4.10).

Таблиця 4.10

Найтісніший зв'язок

	X_1	x_2	x_3	x_4	l_j
y_1	m_{11}	m_{12}			$m_{11}+m_{12}$
y_2			m_{22}		m_{22}
y_3				m_{33}	m_{33}
k_i	m_{11}	m_{12}	m_{22}	m_{33}	N

Обчислимо для цього випадку статистику Пірсона χ^2:

$$\chi^2_{\max} = \sum_i \sum_j \frac{m_{ij}^2}{\hat{m}_{ij}} - n = \sum_i \sum_j \frac{m_{ij}^2}{k_i l_j} n - n =$$

$$= n\left[\frac{m_{11}^2}{m_{11}(m_{11}+m_{12})} + \frac{m_{12}^2}{m_{12}(m_{11}+m_{12})} + \frac{m_{22}^2}{m_{22} \cdot m_{22}} + \frac{m_{33}^2}{m_{33} \cdot m_{33}}\right] - n =$$

$$= n\left[\frac{m_{11}}{m_{11}+m_{12}} + \frac{m_{12}}{m_{11}+m_{12}} + 1 + 1\right] - n = n(3-1) =$$

$$= n(q-1). \tag{4.78}$$

У загальному вигляді

$$\chi^2_{max} = n(d-1), \tag{4.79}$$

де $d = min(p, q)$ – найменша кількість рядків або кількість стовпчиків таблиці.

Отже, встановлюємо, що для $p \neq q$ потрібно завжди набувати d як мінімальне значення з p, q.

Зазвичай використовують дві міри для коефіцієнта контингенції:
міру Крамера:

$$C = \sqrt{\frac{\chi^2}{\chi^2_{max}}} = \sqrt{\frac{\chi^2}{n(d-1)}}, \quad 0 \leq C \leq 1; \tag{4.80}$$

міру спряженості Кендела :

$$K = \sqrt{\frac{\chi^2}{\chi^2 + n}}, \quad 0 \leq K \leq \sqrt{\frac{d-1}{d}} < 1. \tag{4.81}$$

Максимальне значення статистики K (при "функціональній", однозначній залежності) менше від одиниці, і це є недоліком міри Кендела.

Якщо обидві номінальні ознаки мають тільки альтернативну варіацію ("так – ні"), то тісноту зв’язку обчислюють за допомогою коефіцієнта асоціації Пірсона:

$$\Phi = \frac{\begin{vmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{vmatrix}}{\sqrt{k_1 k_2 l_1 l_2}}. \tag{4.82}$$

Теорема. Коефіцієнт асоціації Пірсона $C = \Phi$ дорівнює коефіцієнту контингенції Крамера при $p = q = 2$.

У табл. 4.11 наведені умови даної задачі.

Таблиця 4.11

Умовне подання альтернативних категорій двох ознак

	A_1 (ні)	A_2 (так)	$\sum$
B_1 (ні)	m_{11}	m_{21}	l_1
B_2 (так)	m_{21}	m_{22}	l_2
$\sum$	k_1	k_2	

Доведення. У цьому випадку $d = 2$ і $C = \sqrt{\dfrac{\chi^2}{n}}$. Перетворюємо:

$$\frac{1}{n}\chi^2 = \sum_i \sum_j \frac{m_{ij}^2}{\hat{m}_{ij} \cdot n} - 1 = \sum_i \sum_j \frac{m_{ij}^2}{k_i \cdot l_j} - 1 = \frac{m_{11}^2}{k_1 l_1} + \frac{m_{12}^2}{k_1 l_2} + \frac{m_{21}^2}{k_2 l_1} + \frac{m_{22}^2}{k_2 l_2} - 1 =$$

$$= \frac{1}{k_1 k_2 l_1 l_2}[m_{11}^2 (m_{21} + m_{22})(m_{12} + m_{22}) + m_{12}^2 (m_{21} + m_{22})(m_{11} + m_{21}) +$$
$$+ m_{21}^2 (m_{11} + m_{12})(m_{12} + m_{22}) + m_{22}^2 (m_{11} + m_{12})(m_{11} + m_{21}) -$$
$$- (m_{11} + m_{12})(m_{21} + m_{22})(m_{11} + m_{21})(m_{12} + m_{22})], \tag{4.83}$$

де $\hat{m}_{ij} = \dfrac{k_i l_j}{n}$, $k_1 = m_{11} + m_{21}, k_2 = m_{21} + m_{22}, l_1 = m_{11} + m_{21}, k_2 = m_{21} + m_{22}$.

У виразі (4.83) слід розкрити дужки та звести подібні члени:

$$\frac{1}{n}\chi^2 = \frac{1}{k_1 k_2 l_1 l_2}(m_{11}^2 m_{22}^2 + m_{12}^2 m_{21}^2 - 2 m_{11} m_{22} m_{21} m_{12} =$$
$$= \frac{1}{k_1 k_2 l_1 l_2}(m_{11} m_{22} - m_{12} m_{21})^2 = \frac{1}{k_1 k_2 l_1 l_2} \begin{vmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{vmatrix}^2 . \tag{4.84}$$

Отже, при $p = q = 2$ коефіцієнт Крамера C дійсно переходить у коефіцієнт асоціації Пірсона. ■

Теорема. Коефіцієнт асоціації Пірсона Φ дорівнює коефіцієнту парної кореляції $|r_{uv}|$, якщо категорії "ні" присвоювати значення 0, а альтернативній категорії "так" − значення 1 (табл. 4.12).

Таблиця 4.12

Позначення частот

v \ u	0	1	$\sum$
0	m_{11}	m_{21}	l_1
1	m_{21}	m_{22}	l_2
$\sum$	k_1	k_2	n

Доведення. Для табл. 4.12 обчислюємо коефіцієнт парної кореляції:

$$r_{uv} = \frac{s_{uv}}{s_u \cdot s_v},$$

де

$$\bar{u} = \frac{(0 \cdot k_1 + 1 \cdot k_2)}{n} = \frac{k_2}{n}; \quad \overline{u^2} = \frac{(0^2 \cdot k_1 + 1^2 \cdot k_2)}{n} = \frac{k_2}{n}; \tag{4.85}$$

$$\bar{v} = \frac{(0 \cdot l_1 + 1 \cdot l_2)}{n} = \frac{l_2}{n}; \quad \overline{l^2} = \frac{(0^2 \cdot l_1 + 1^2 \cdot l_2)}{n} = \frac{l_2}{n}; \tag{4.86}$$

$$s_u^2 = \overline{u^2} - (\bar{u})^2 = \frac{k_2}{n} - \left(\frac{k_2}{n}\right)^2 = \frac{k_2}{n}\left(1 - \frac{k_2}{n}\right) = \frac{k_2}{n} \cdot \frac{k_1}{n}; \tag{4.87}$$

$$s_v^2 = \overline{v^2} - (\bar{v})^2 = \frac{l_2}{n} - \left(\frac{l_2}{n}\right)^2 = \frac{l_2}{n}\left(1 - \frac{l_2}{n}\right) = \frac{l_2}{n} \cdot \frac{k_1}{n}, \tag{4.88}$$

за умови, що $k_1 + k_2 = n$ та $l_1 + l_2 = n$.

Далі

$$\overline{uv} = \frac{m_{22}}{n}; \quad s_{uv} = \overline{uv} - \bar{u} \cdot \bar{v} = \frac{m_{22}}{n} - \frac{k_2}{n} \cdot \frac{l_2}{n} = \frac{1}{n^2}(m_{22}n - k_2 l_2).$$

Отже, маємо

$$r_{uv} = \frac{s_{uv}}{s_u \cdot s_v} = \frac{m_{22}n - k_2 l_2}{\sqrt{k_1 k_2 l_1 l_2}}. \tag{4.89}$$

Перетворюємо чисельник:

$$\begin{aligned} m_{22}n - k_2 l_2 &= m_{22}(m_{11} + m_{12} + m_{21} + m_{22}) - (m_{21} + m_{22})(m_{12} + m_{22}) = \\ &= (m_{11}m_{22} + m_{12}m_{22} + m_{21}m_{22} + m_{22}^2) - \\ &- (m_{12}m_{21} + m_{21}m_{22} + m_{12}m_{22} + m_{22}^2) = \\ &= m_{11}m_{22} - m_{12}m_{21} = \begin{vmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{vmatrix}. \end{aligned} \tag{4.90}$$

Звідки

$$r_{uv} = \frac{s_{uv}}{s_u \cdot s_v} = \frac{\begin{vmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{vmatrix}}{\sqrt{k_1 k_2 l_1 l_2}} = \Phi, \tag{4.91}$$

тобто коефіцієнт асоціації Пірсона дійсно є коефіцієнтом парної кореляції "на рангах" неметричних величин. Коефіцієнт асоціації вимірює зв’язок дихотомічних номінальних ознак. ■

Вважають, що статистики χ^2 є доброякісною мірою, якщо очікувані частоти – більші ніж 5, а обсяг вибірки – не менший ніж 40. Якщо не виконуються ці умови, то для більшої відповідності χ^2-розподілу вводять так звану поправку на неперервність, або коефіцієнт Йєйтса [9]:

$$\chi^2 = \frac{n\left(m_{11}m_{22} - m_{12}m_{21} - \frac{n}{2}\right)^2}{k_1 k_2 l_1 l_2}. \tag{4.92}$$

Існують також інші показники асоціації, наприклад, тетрахоричний коефіцієнт $r_{тет}$. Його теж обчислюють для вимірювання зв’язку дихотомічних значень номінальних ознак. Тетрахоричний коефіцієнт обчислюється за формулою:

$$r_{тет} = \cos\left(\frac{180^0}{1+\sqrt{\frac{m_{11}m_{22}}{m_{12}m_{21}}}}\right) = \cos\frac{180^0 \cdot \sqrt{m_{12}m_{21}}}{\sqrt{m_{11}m_{22}} + \sqrt{m_{12}m_{21}}}. \tag{4.93}$$

Значення $r_{тет}$ знаходиться в межах $r_{тет} \in [-1;1]$. Якщо розподіл частот $l_1,\ l_2,\ k_1,\ k_2$ нерівномірний, то $r_{тет}$ стає ненадійним показником зв’язку.

Відомий також коефіцієнт асоціації Q, запропонований Г. У. Юлом у 1900 році:

$$Q = \frac{m_{11}m_{22} - m_{12}m_{21}}{m_{11}m_{22} + m_{12}m_{21}}. \tag{4.94}$$

Значення коефіцієнта Q знаходиться в інтервалі $Q \in [-1;1]$.

Л. Гудмен і Є. Краскал також вважали за потрібне, щоб міри зв’язку набували значень на всьому відрізку $[-1;1]$. Вчені запропонували ще кілька мір, які мають імовірнісну інтерпретацію на відрізку $[-1;1]$. Найвідомішими серед них є міри Л. Гутмана $\lambda_a, \lambda_b, \lambda$ [276]. Статистика λ_b заснована на порівнянні двох ситуацій: для взятого навмання індивіда, який належить сукупності, намагаються розпізнати його B-категорію, коли: a) не має ніякої додаткової інформації; б) відома його A-категорія. Якщо категорії A і B зовсім не зв'язані, то не буде в одній ситуації краще, ніж в іншій, але в протилежному разі буде відчуватися певне покращення. Міра λ_b характеризує це покращення чисельно в термінах відносного приросту ймовірності помилки передбачення B-категорії у процесі переходу від однієї ситуації до іншої, коли припускається, що передбачення забезпечує в кожному випадку ймовірну з B-категорій:

$$\lambda_b = \frac{\sum_{i=1}^{2} \tilde{m}_{i0} - k_0}{n - k_0}, \tag{4.95}$$

де $\tilde{m}_{i0}$ – найбільша частота в i-му рядку табл. 4.9;

k_0 – найбільше значення із сум у стовпчиках.

Якщо A і B поміняти місцями, тобто поцікавитися приростом ефективності передбачення A-категорії за відомої B-категорії, то одержимо наступну статистику:

$$\lambda_a = \frac{\sum_{j=1}^{2} \tilde{m}_{0j} - l_0}{n - l_0}, \tag{4.96}$$

де $\tilde{m}_{0j}$ – найбільша частота в j-му стовпчику табл. 4.9;

l_0 – найбільше значення із сум у рядках.

Якщо припустити, що нас рівнозначно цікавлять A і B, то пропонується міра:

$$\lambda = \frac{\left(\sum_{i=1}^{2} \tilde{m}_{i0} - k_0\right) + \left(\sum_{j=1}^{2} \tilde{m}_{0j} - l_0\right)}{2n - k_0 - l_0}. \tag{4.97}$$

Відома також статистика Р. Сомерса, яка визначається за формулою:

$$d = \frac{S - D}{S + D + T_b}, \tag{4.98}$$

де S – загальна кількість пар спостережень (для A, B маємо пару рангів (i, j) і для A', B' – (i', j')), для яких одночасно $i > i'$ і $j > j'$ або, навпаки, $i < i'$ і $j < j'$;

D – загальна кількість пар спостережень, для яких або $i > i'$, але $j < j'$, або $i < i'$, але $j > j'$;

T_b – загальна кількість пар, для яких $j = j'$ (ще T_a – загальна кількість пар, для яких $i = i'$).

Статистику Р. Сомерса можна розглядати як різницю між імовірностями отримати правильний і неправильний порядок при вилученні із сукупності двох спостережень випадковим чином, коли змінна A не має рангів, що співпадають.

У своїй роботі [9] Г. Аптон довів, що різні міри зв’язку акцентують увагу на різних аспектах взаємних відношень між змінними. На додачу до цього слід навести думку В. Краскала про те, що немає сенсу використовувати якусь одну міру зв’язку; декілька різних мір дають багатоаспектну інформацію, а це потрібно для розуміння того, з якою мірою потрібно працювати [9, с. 26]. А також "жодній із цих мір не слід приписувати роль, більшу, ніж роль засобів початкового припущення перед систематичним аналізом" [9, с. 40]. Отже, слід обчислювати декілька мір взаємозв’язку номінальних ознак, які в цілому становлять попередній етап комплексного аналізу якісних ознак об’єкта.

Таким чином, розбудова моделі системи номінальних ознак об’єкта має розпочинатися з теоретико-логічного аналізу та математичної формалізації взаємозв’язків у системі, що ґрунтуються на обчисленнях описаної системи показників тісноти зв’язку, кожний із яких робить свій внесок у загальну характеристику структури системи ознак. Визначення складних номінальних ознак можливе на основі модифікації математичних методів, що розроблені для метричних ознак.

Задача 4.2. Провести початковий аналіз інтелектуального забезпечення 7 промислових підприємств за 10 номінальними ознаками, а саме: A – якість програмного забезпечення; B – наявність чітко розробленої та документально оформленої філософії бізнесу; C – рівень здоров’я працівників; D – рівень користування об’єктами соціальної інфраструктури; E – ступінь довіри працівників до вищого керівництва; F – рівень залежності працівників від підприємства; G – наявність постійних клієнтів (за критерієм постачання або замовлень понад 5); H – наявність рекламацій із боку споживачів; K – наявність персональних комп’ютерів та програмного забезпечення у відділі з роботи із клієнтами;

L – ведення уніфікованої бази даних постачальників та споживачів. У таблиці 4.13 наведені номінації ознак, а в табл. 4.14 та табл. 4.15 – умовні позначення.

Таблиця 4.13

Умовні позначення номінальних ознак у задачі

Позначення номінальної ознаки	Номінації (назви), яких набувають ознаки	Позначення номінації
A	Високе	$A1$
	Середнє	$A2$
	Низьке	$A3$
B	Так	$B1$
	Ні	$B0$
C	Високий	$C1$
	Середній	$C2$
	Низький	$C3$
D	Високий	$D1$
	Середній	$D2$
	Низький	$D3$
E	Високий	$E1$
	Середній	$E2$
	Низький	$E3$
F	Високий	$F1$
	Середній	$F2$
	Низький	$F3$
G	Є	$G1$
	Немає	$G0$
H	Є	$H1$
	Немає	$H0$
K	Є	$K1$
	Немає	$K0$
L	Так	$L1$
	Ні	$L0$

Таблиця 4.14

Номінації ознак

Рік	Умовне позначення ознак									
	A	B	C	D	E	F	G	H	K	L
АТЗТ "ХЗЕВ – 1" (1)										
2001	низьке	ні	високий	високий	високий	високий	є	немає	немає	Так
2002	середнє	ні	високий	високий	високий	високий	немає	є	є	Так
2003	середнє	ні	середній	середній	високий	високий	є	є	є	Так
ВАТ "Автрамат" (2)										
2001	середнє	ні	високий	високий	середній	середній	є	є	немає	Так
2002	середнє	ні	середній	високий	середній	високий	є	є	немає	Так
2003	середнє	так	середній	середній	середній	високий	є	є	є	Так
ВАТ "Електромашина"(3)										
2001	низьке	ні	високий	середній	середній	середній	є	є	немає	Ні
2002	низьке	ні	середній	середній	низький	середній	є	є	немає	Ні
2003	середнє	ні	середній	середній	низький	середній	є	є	є	Ні
ВАТ "Укрелектромаш" (4)										
2001	низьке	ні	середній	середній	середній	середній	є	є	є	Ні
2002	низьке	ні	середній	середній	середній	середній	є	є	є	Ні
2003	середнє	ні	середній	середній	середній	середній	є	є	є	Так
ДП "завод "Електроважмаш"" (5)										
2001	середнє	ні	середній	середній	низький	високий	немає	є	є	Ні
2002	середнє	ні	середній	низький	низький	низький	є	немає	є	Ні
2003	середнє	ні	середній	низький	середній	низький	є	є	є	Так
ДП "ХЕМЗ" (6)										
2001	середнє	ні	низький	високий	середній	низький	немає	є	немає	Ні
2002	середнє	ні	середній	високий	середній	низький	є	немає	є	Ні
2003	середнє	ні	середній	середній	середній	середній	є	є	є	Ні
ЗАТ "Завод "Південкабель"" (7)										
2001	низьке	ні	низький	низький	середній	середній	є	є	немає	Так
2002	середнє	так	середній	середній	середній	середній	є	є	немає	Так
2003	середнє	так	середній	низький	середній	високий	є	є	немає	Так

Таблиця 4.15

Номінації ознак в умовних позначеннях

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*K*	*L*
A1	B0	C3	D3	E3	F3	G1	H0	K0	L1
A2	B0	C3	D3	E3	F3	G0	H1	K1	L1
A2	B0	C2	D2	E3	F3	G1	H1	K1	L1
A2	B0	C3	D3	E2	F2	G1	H1	K0	L1
A2	B0	C2	D3	E2	F2	G1	H1	K0	L1
A2	B1	C2	D2	E2	F3	G1	H1	K1	L1
A1	B0	C3	D2	E2	F3	G1	H1	K0	L0
A1	B0	C2	D2	E1	F2	G1	H1	K0	L0
A2	B0	C2	D2	E1	F2	G1	H1	K1	L0
A1	B0	C2	D2	E2	F2	G1	H1	K1	L0
A1	B0	C2	D2	E2	F2	G1	H1	K1	L0
A2	B0	C2	D2	E2	F2	G1	H1	K1	L1
A2	B0	C2	D2	E1	F3	G0	H1	K1	L0
A2	B0	C2	D1	E1	F1	G1	H0	K1	L0
A2	B0	C2	D1	E2	F1	G1	H1	K1	L1
A2	B0	C1	D3	E2	F1	G0	H1	K0	L0
A2	B0	C2	D3	E2	F1	G1	H0	K1	L0
A2	B0	C2	D2	E2	F2	G1	H1	K1	L0
A1	B0	C1	D1	E2	F2	G1	H1	K0	L1
A2	B1	C2	D2	E2	F2	G1	H1	K0	L1
A2	B1	C2	D1	E2	F3	G1	H1	K0	L1

Розв'язання задачі. Дві якісні ознаки є незалежними, якщо значення однієї змінної не допомагає прогнозувати значення іншої [269, с. 888]. Оскільки кожна ознака характеризується власними значеннями процентів, які становлять імовірності появи кожної категорії (номінації, назви), то дослідження номінальних ознак слід починати саме з визначення цих імовірностей [137]. Умовні проценти в сукупності є ймовірностями появи категорій однієї ознаки за умов обмеженого розгляду тільки однієї категорії з іншою ознакою. За допомогою ППП Statgraphics побудуємо таблицю парних спряженостей номінацій якісних ознак, що характеризують інтелектуальне забезпечення діяльності підприємств за обчисленими процентними частотами номінальних ознак (початок кростабуляції номінальних ознак), активізуючи Describe/Categorical Data/Crosstabulation.

Як приклад, у табл. 4.16 наведена кростабуляція двох номінальних ознак A і B, яка свідчить про найбільшу спряженість номінацій $A2$ і $B0$ – 57,14%.

Таблиця 4.16

Кростабуляція спряженості двох номінальних ознак інтелектуального забезпечення діяльності підприємств

Номінації ознак	$B0$	$B1$	Загалом для рядків
$A1$	6 28,57%	0 0,00%	6 28,57%
$A2$	12 57,14%	3 14,29%	15 71,43%
Загалом для стовпчиків	18 85,71%	3 14,29%	21 100,00%

Далі потрібно обчислити значення статистики χ^2, рівень її значущості, число степенів свободи для перевірки гіпотези про незалежність ознак (активізація меню Describe/Categorical Data/Contingency Tables ППП Statgraphics). Обчислення наведені в табл. 4.17.

Таблиця 4.17

X^2-тест

Значення χ^2	Число степенів свободи	Рівень значущості p
1,40	1	0,2367
0,24	1	0,6220

Рівні значущості статистики χ^2 свідчать, що "нуль-гіпотезу" про незалежність ознак не можна відхилити. Тест Фішера також свідчить про відсутність спряженості номінальних ознак ($p = 0{,}342105$ і $p = 0{,}526316$), оскільки обчислені рівні значущості перевищують 0,05.

Для переконання у правильності висновків про відсутність спряженості ознак за χ^2-тестом слід обчислити низку коефіцієнтів: коефіцієнт невизначеності λ, коефіцієнт Сомерса d, коефіцієнт η, коефіцієнт контингенції K, коефіцієнт Крамера C, коефіцієнт гамма, коефіцієнт Пірсона, коефіцієнт Кендела τ_b (табл. 4.18) [137]. Дана процедура виконується за допомогою ППП Statgraphics з активізацією меню Describe/Categorical Data/Contingency Tables/Tabular Options/Summary Statistics.

Таблиця 4.18

Таблиця сумарних статистик для номінальних ознак

Статистики	Значення	Залежність номінацій у рядках	Залежність номінацій у колонках
λ	0,0000	0,0000	0,0000
Сомерса d	0,2500	0,3333	0,2000
η	–	0,2582	0,2582
Коефіцієнт контингенції K	0,2500	–	–
Коефіцієнт Крамера C	0,2582	–	–
Коефіцієнт гамма	1,0000	–	–
Коефіцієнт Спірмена ρ	0,2582	0,1292	19
Коефіцієнт Кендела τ_b	0,2582	0,2482	–

Коефіцієнт асоціації Пірсона обчислюється наступним чином:

$$\Phi = \frac{\begin{vmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{vmatrix}}{\sqrt{k_1 k_2 l_1 l_2}} = \frac{\begin{vmatrix} 6 & 0 \\ 12 & 3 \end{vmatrix}}{\sqrt{18 \cdot 3 \cdot 6 \cdot 15}} = \frac{18}{18\sqrt{15}} = 0{,}2582.$$

Як було доведено вище значення даного коефіцієнта співпадає зі значенням коефіцієнта контингенції K.

Обчислимо величину χ^2:

$$\chi^2 = \frac{n}{k_1 k_2 l_1 l_2} \begin{vmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{vmatrix}^2 = n \cdot \Phi^2 = 21 \cdot 0{,}2582^2 = 1{,}4.$$

При $d = 2$ маємо коефіцієнт контингенції C:

$$C = \sqrt{\frac{\chi^2}{n\ \ 2-1}} = \sqrt{\frac{1{,}4}{21}} = 0{,}2582,$$

і коефіцієнт контингенції K:

$$K = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{1{,}4}{1{,}4 + 21}} = 0{,}25.$$

Значення коефіцієнта кореляції r_{uv} теж співпадає з мірою C і дорівнює 0,2582.

Якщо перенумерувати номінальні категорії, то можна обчислити кореляційні відношення η, обчислені на рангах категорій. Дана міра заслуговує окремого розгляду.

Коефіцієнт Юла:

$$Q = \frac{m_{11}m_{22} - m_{12}m_{21}}{m_{11}m_{22} + m_{12}m_{21}} = \frac{6 \cdot 3 - 0 \cdot 12}{6 \cdot 3 + 0 \cdot 12} = 1$$

(маємо незадовільне значення коефіцієнта).

Тетрахоричний коефіцієнт:

$$r_{тет} = cos \frac{180^\circ \sqrt{m_{12}m_{21}}}{\sqrt{m_{11}m_{22}} + \sqrt{m_{12}m_{21}}} = cos\, 0^\circ = 1$$

(маємо незадовільне значення коефіцієнта).

Міри Гудмена:

$$\lambda_b = \frac{\sum_i m_{ik}^\circ - k_0}{n - k_0} = \frac{(2+6) - 18}{21 - 15} = 0$$

(маємо незадовільне значення коефіцієнта);

$$\lambda_a = \frac{\sum_j m_{kj}^\circ - l_0}{n - l_0} = \frac{(2+3) - 15}{21 - 15} = 0$$

(маємо незадовільне значення коефіцієнта);

$$\lambda = \frac{\left(\sum_i m_{ik}^\circ - k_0\right) + \sum_j m_{kj}^\circ - l_0}{2n - k_0 - l_0} = 0$$

(маємо незадовільне значення коефіцієнта).

Значення коефіцієнта Сомерса дорівнює $d = 0{,}25$.

У підрозділі 3.1 розглядалися коефіцієнти Кендела, їх обчислення за формулами та наводився приклад їх використання у розв'язанні практичних задач, коли початкова сукупність представлена значеннями величин ординальних ознак.

Теорема. Для дихотомічних ознак коефіцієнт рангової кореляції Кендела співпадає з коефіцієнтом асоціації Пірсона Ф.

Доведення. Розглянемо формули обчислення коефіцієнта Кендела τ_b для дихотомічних ознак. Як було доведено вище, коефіцієнт рангової кореляції Кендела можна одержати з загальної формули:

$$\tau_b = \frac{\sum\sum a_{ij} b_{ij}}{\sqrt{\sum\sum a_{ij}^2}\sqrt{\sum\sum b_{ij}^2}}, \tag{4.99}$$

де

$$a_{ij} = sgn(u_i - u_j), \quad b_{ij} = sgn(v_i - v_j), \tag{4.100}$$

u_i, v_i – ранги.

У чисельнику дорівнюють $+1$ тільки $m_{11}m_{22}$ комбінацій, причому потрібно враховувати, що в сумі кожна пара зустрічається двічі: один раз – коли індекс ij і другий – коли індекс ji, тому в сумі $+1$ буде $2m_{11}m_{22}$. Аналогічно для -1 буде $2m_{12}m_{21}$. Таким чином, чисельник буде дорівнювати $2(m_{11}m_{22} - m_{12}m_{21})$. Розглянемо $\sqrt{\sum\sum a_{ij}^2}$, де

$$a_{ij}^2 = \begin{cases} 1 \ – \text{для комбінацій } k_1 k_2, \\ 0 \ – \text{для рівних значень } u_i = u_j. \end{cases}$$

Отримаємо: $\sqrt{\sum\sum a_{ij}^2} = 2k_1k_2$.

Для $\sqrt{\sum\sum b_{ij}^2}$ маємо:

$$b_{ij}^2 = \begin{cases} 1 \ – \text{для комбінацій } l_1 l_2, \\ 0 \ – \text{для рівних значень } v_i = v_j. \end{cases} \tag{4.101}$$

Отримаємо: $\sqrt{\sum\sum b_{ij}^2} = 2l_1l_2$.

Таким чином, для визначення тісноти зв’язку номінальних дихотомічних ознак коефіцієнт Кендела має обчислюватися за формулою:

$$\tau_b = \frac{m_{11}m_{22} - m_{12}m_{21}}{\sqrt{k_1 k_2 l_1 l_2}} = \frac{\begin{vmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{vmatrix}}{\sqrt{k_1 k_2 l_1 l_2}}. \qquad (4.102)$$

Отже, для дихотомічних ознак коефіцієнт рангової кореляції Кендела дійсно співпадає з коефіцієнтом асоціації Пірсона. ■

Значення коефіцієнта Кендела для оцінки тісноти зв’язку двох дихотомічних номінальних ознак A і B (якості програмного забезпечення і наявності чітко розробленої та документально оформленої філософії бізнесу) дорівнює $\tau_b = 0{,}2582$ і демонструє незначний взаємозв’язок цих двох ознак на досліджуваних підприємствах, що свідчить про погану організацію інтелектуального забезпечення виробничо-господарської діяльності підприємств.

Отже, порівняння значень наведених коефіцієнтів для оцінки тісноти зв’язку дихотомічних номінальних ознак свідчить про те, що коефіцієнт парної кореляції Пірсона співпадає з коефіцієнтом Кендела τ_b і з коефіцієнтом асоціації Φ [137] (табл. 4.18) де η – кореляційні відношення, обчислені на рангах рядків і стовпчиків. Зміст даної таблиці також підтверджує факт, що гіпотеза про незалежність ознак повинна бути прийнятою. Оскільки сукупність номінальних ознак, що описують інтелектуальне забезпечення діяльності підприємств, налічує 10 якісних ознак, то описана процедура була виконана $C_{10}^2 = 45$ разів.

Таким чином, у наведеному комплексі різних мір тісноти зв’язку коефіцієнти не узгоджені між собою, а отже, використовувати їх потрібно вибірково. Найбільш обґрунтованими серед них вважаємо міри C, K, η (хоча вони теж дають дещо різні значення). Отже, потрібна об’єктивна, універсальна міра взаємозвязку різних ознак соціально-економічних систем. Ця міра має бути об’єктивною, тобто вона повинна узгоджуватися з метричними мірами, коли їх можна обчислити. Обґрунтування розробки такої міри викладемо на прикладі розв’язання однієї реальної задачі.

Задача 4.3. Маємо 10 елементарних ознак, серед них 6 метричних (позначені $x_{16}, x_{17}, x_{18}, x_{34}, x_{35}, x_{38}$ у табл. 3.1), які кількісно характеризують виробничо-господарську діяльність 7 підприємств за 3 роки (перелік підприємств у табл. 4.14) та 4 неметричних порядкових (позначені C, D, E, F), які якісно характеризують діяльність цих підприємств. Потрібно оцінити взаємозв'язок між елементарними ознаками в системі, припускаючи, що вони можуть бути виміряними на різних шкалах.

Розв'язання задачі. Величини метричних ознак, які виміряні на шкалі відношень, завжди можна перетворити в ранги на основі згрупованих значень величин за інтервалами рівної ширини. У табл. 4.19 перші 6 колонок містять значення метричних ознак, які згруповані за 6-ма рівновеликими інтервалами, останні 4 колонки містять порядкові ознаки, що мають 3 рівні варіації. Для кожної пари ознак складені таблиці спряженості (всього 45 таблиць). У табл. 4.20 наведена спряженість ознак X і Y. Кожна з цих ознак має 6 категорій, які пронумеровані цілими числами від 1 до 6.

Таблиця 4.19

Початкові дані

№	X (X_{16})	Y (X_{17})	Z (X_{18})	U (X_{34})	V (X_{35})	W (X_{38})	C	D	E	F
1	6	1	6	2	3	3	3	3	3	3
2	5	2	5	2	4	3	3	3	3	3
3	5	2	5	2	4	3	2	2	3	3
4	4	3	4	3	2	1	3	3	2	2
5	3	3	3	3	2	2	2	3	2	3
6	3	3	3	3	2	2	2	2	2	3
7	4	3	4	2	3	2	3	2	2	2
8	3	3	3	1	5	2	2	2	1	2
9	3	3	3	1	5	2	2	2	1	2
10	2	5	2	5	1	3	2	2	2	2
11	2	4	3	6	1	4	2	2	2	2
12	2	4	3	6	1	4	2	2	2	2
13	2	6	2	2	3	3	2	2	1	3
14	1	5	1	1	4	4	2	1	1	1
15	1	5	1	1	4	4	2	1	2	1
16	2	6	3	1	4	1	1	3	2	1
17	2	4	3	1	6	1	2	3	2	1
18	1	5	2	1	6	1	2	2	2	2
19	5	2	6	3	2	5	1	1	2	2
20	5	2	5	4	2	6	2	2	2	2
21	5	2	5	4	2	6	2	1	2	3

Для обчислення коефіцієнтів контингенції не має значення, як пронумеровані категорії, оскільки їх числові значення не беруть участі в обчисленнях. Зміст останнього рядка і останнього стовпчика табл. 4.20 розглянемо окремо.

Таблиця 4.20

Таблиця спряженості ознак X і Y

Y \ X	1	2	3	4	5	6	Суми l_j	Середні групові
6	0	2	0	0	0	0	2	2
5	3	1	0	0	0	0	4	1,25
4	0	3	0	0	0	0	3	2
3	0	0	4	2	0	0	6	3,3333
2	0	0	0	0	5	0	5	5
1	0	0	0	0	0	1	1	6
Суми k_i	3	6	4	2	5	1	21	3,1429
Середні групові	5	4,8333	3	3	2	1	3,4762	

За відсутності зв'язку між ознаками X, Y ("нуль-гіпотеза") частоти m_{ij}, що спостерігаються, повинні бути близькими до частот $\hat{m}_{ij} = \frac{k_i l_j}{n}$, що очікуються. Для порівняння частот обчислюється статистика Пірсона $\chi^2 = n \cdot \left(\sum_{i=1}^{p} \sum_{j=1}^{q} \frac{m_{ij}^2}{k_i l_j} - 1 \right)$, яка за даними задачі дорівнює

$$\chi^2 = 21 \cdot \left(\frac{2^2}{6 \cdot 2} + \frac{3^2}{3 \cdot 4} + \frac{1^2}{6 \cdot 4} + \frac{3^2}{6 \cdot 3} + \frac{4^2}{4 \cdot 6} + \frac{2^2}{2 \cdot 6} + \frac{5^2}{5 \cdot 5} + \frac{1^2}{1 \cdot 1} - 1 \right) = 76{,}125.$$

Для оцінки тісноти зв'язку (при $\chi^2 > \chi^2_{0,05}$) застосовується або міра $C = \sqrt{\frac{\chi^2}{n(d-1)}} = 0{,}851$, або міра $K = \sqrt{\frac{\chi^2}{\chi^2 + n}} = 0{,}885$ (коефіцієнти контингенції), де $d = \min\{p, q\}$. Остання міра змінюється від 0 до $K_{\max} = \sqrt{\frac{d-1}{d}}$, що менше від одиниці. Тому введемо скоректовану міру $CK = \sqrt{\frac{\chi^2}{\chi^2 + n} \cdot \frac{d}{d-1}} = 0{,}970$, необхідність якої обговорювалась у підрозділі 3.1. Скоректована міра CK, як і міра C, змінюється від 0 до 1. Обчислимо: $CCK = \sqrt{C \cdot CK} = \sqrt{0{,}851 \cdot 0{,}970} = 0{,}909$.

Доводиться, що скоректований коефіцієнт $CK = \sqrt{\frac{\chi^2}{\chi^2 + n} \cdot \frac{d}{d-1}}$ і коефіцієнт контингенції Крамера $C = \sqrt{\frac{\chi^2}{n(d-1)}}$ співпадають лише для граничних ситуацій ($CK = C = 0$, якщо немає жодного зв'язку; $CK = C = 1$, якщо є функційне співвідношення, коли кожній категорії X відповідає одна конкретна категорія Y). Для всіх проміжних ситуацій коефіцієнт CK дає систематично більше значення, ніж коефіцієнт C $(CK > C)$. Міри C і CK зв'язані функціональною залежністю $CK = \sqrt{\frac{C^2 d}{C^2 (d-1) + 1}}$. Ця залежність продемонстрована на рис. 4.2, де точками позначені всі 45 пар C і CK, обчислених за даними табл. 4.20, для $d = 6$ (15 точок) і $d = 3$ (30 точок).

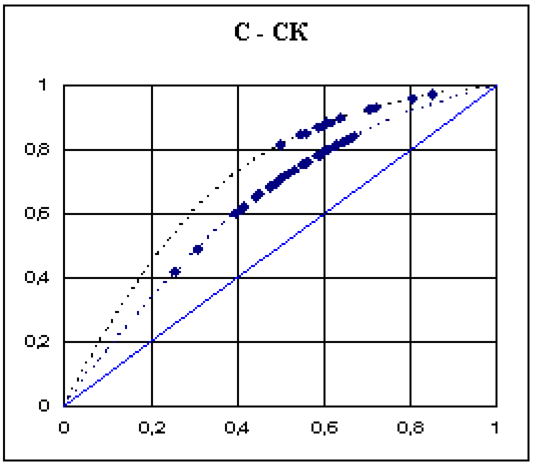


Рис. 4.2. **Зв'язок між коефіцієнтами контингенції *C* і *CK***

Як видно з рис. 4.2, коефіцієнт CK завжди більший від коефіцієнта C.

Рекомендація. Для оцінки тісноти взаємозв'язку ознак слід визначати коефіцієнт *CCK*, який обчислюємо як середнє геометричне двох мір C і CK:

$$CCK = \sqrt{C \cdot CK},$$

ця міра є об'єктивною й універсальною [161].

Практична перевірка рекомендації. Універсальна міра може бути обчислена для будь-яких метричних і неметричних величин, тому коефіцієнти контингенції *C*, *K*, *CK* – універсальні. Універсальна міра *CCK* добре узгоджується з метричними мірами в тих випадках, коли метричні міри можуть бути обчислені. Якщо ознаки – метричні (або хоча б порядкові), є можливість оцінити тісноту лінійного зв'язку за допомогою коефіцієнта кореляції R_{xy} та тісноту довільної форми (не обов'язково лінійної) за допомогою кореляційних відношень $\eta_{y/x}, \eta_{x/y}$. У процесі обчислення цих характеристик важливо, щоб імена категорій були замінені порядковими числами. Проте в загальному випадку неметричних ознак категорії можна переставляти, оскільки порядок розташування категорій не має значення. Оскільки перенумерація категорій змінює величини коефіцієнтів кореляції, тобто коефіцієнти кореляції не є універсальною мірою, тісноту зв'язку потрібно оцінювати за допомогою максимального кореляційного відношення $Eta_0 = \max\{\eta_{y/x}, \eta_{x/y}\}$, яке стійке до перенумерації категорій. У табл. 4.19 імена категорій ознак X і Y зашифровані порядковими номерами; в останньому рядку таблиці обчислені середні значення ознаки Y для кожної категорії X (середні для стовпчиків таблиці) $\bar{Y}_{X_i} = \frac{1}{k_i}\sum_{j=1}^{q} m_{ij} Y_j$; аналогічно, в останньому стовпчику таблиці обчислені середні значення ознаки X для кожної категорії Y (середні для рядків таблиці) $\bar{X}_{Y_j} = \frac{1}{l_j}\sum_{i=1}^{p} m_{ij} X_i$. Так, для 2-го стовпчика

$$\bar{Y}_{X_2} = \frac{1}{k_2}\sum_{j=1}^{q} m_{2j} Y_j = \frac{6\cdot2+5\cdot1+4\cdot3}{6} = 4,833,$$

а для 4-го рядка

$$\bar{X}_{Y_4} = \frac{1}{l_4}\sum_{i=1}^{p} m_{i4} X_i = \frac{3\cdot4+4\cdot2}{6} = 3,333.$$

У цій же таблиці наведені загальні середні:

$$\bar{Y} = \frac{1}{n}\sum_{j=1}^{q} l_i Y_j = 3,476; \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{p} k_i X_i = 3,143.$$

Для обчислення кореляційних відношень підраховуємо суми квадратів:

$$SSY = \sum_{j=1}^{q} l_j Y_j^2 - n\cdot(\bar{Y})^2 = 41,238; \qquad SSX = \sum_{i=1}^{p} k_i X_i^2 - n\cdot(\bar{X})^2 = 48,571;$$

$$SS\,y/x = \sum_{i=1}^{p} k_i \left(\bar{Y}_{X_i}\right)^2 - n\cdot(\bar{Y})^2 = 36,405; \quad SS\,x/y = \sum_{j=1}^{q} l_j \left(\bar{X}_{Y_j}\right)^2 - n\cdot(\bar{X})^2 = 46,488.$$

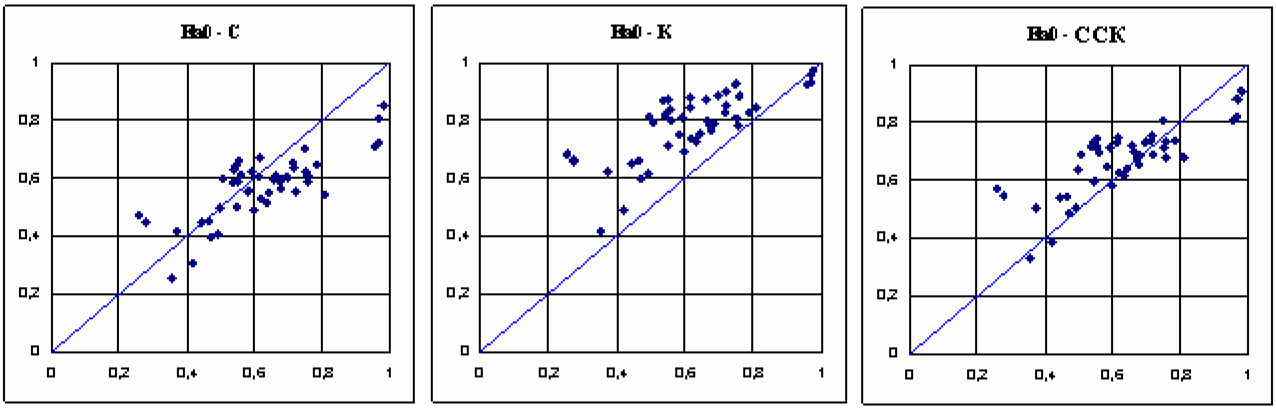
Обчислимо кореляційні відношення та їх максимальне значення:

$$\eta_{y/x} = \sqrt{\frac{SS\,y/x}{SSY}} = 0,940; \quad \eta_{x/y} = \sqrt{\frac{SS\,x/y}{SSX}} = 0,978; \quad Eta_0 = \max\{\eta_{y/x}, \eta_{x/y}\} = 0,978.$$

Доведення об'єктивності міри ССК. Об'єктивна міра тісноти зв'язку повинна узгоджуватися з іншими мірами тісноти зв'язку, коли надається можливість їх обчислення. Порівняємо отримані неметричні міри тісноти зв’язку C, CK, CCK з метричною мірою Eta_0 на прикладі $\frac{10 \cdot 9}{2} = 45$ пар ознак. На рис. 4.3 наведені відповідні графіки.

Зауваження. Таке порівняння можливе, якщо всі ознаки можна вважати метричними.

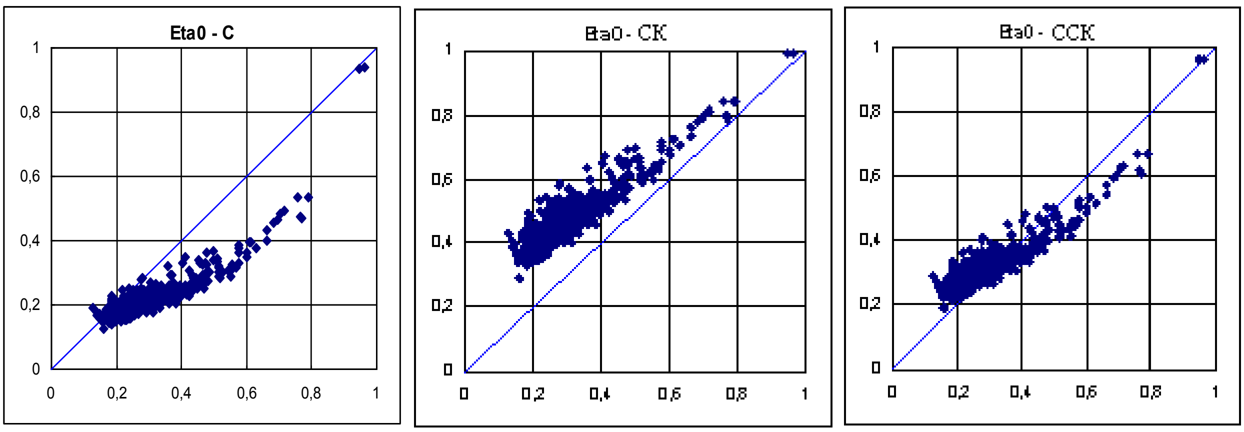
З рис. 4.3а видно, що коефіцієнт контингенції C систематично менший від метричної міри Eta_0, а з рис. 4.3б – що скоректований коефіцієнт CK істотно перевищує Eta_0. Саме тому була рекомендована міра CCK як середнє геометричне систематично заниженої міри C та систематично завищеної міри CK. На рис. 4.3в чітко підтверджується відповідність між рекомендованою мірою CCK та метричною мірою Eta_0.



а б в

Рис. 4.3. **Залежність між мірами тісноти зв’язку Eta_0, C, CK, CCK**

Були побудовані залежності між коефіцієнтами Eta_0, C, CK, CCK, що визначають взаємозв’язок між неметричними ознаками. Графіки залежностей наведені на рис. 4.4.



а б в

Рис. 4.4. **Залежність між мірами тісноти зв’язку Eta_0, C, CK, CCK**

Факт положення величин коефіцієнтів чітко демонструється на рис. 4.4а та б. Отже, у випадках, коли можливі метричні міри тісноти зв'язку, об'єктивна міра CCK добре з ними узгоджується. Це й треба було довести.

Оскільки міра CCK (і її складові міри C і CK) ґрунтуються на величині χ^2, то проблема оцінки значущості цієї міри зводиться до порівняння обчисленого χ^2 з критичними табличними значеннями $\chi^2_{0,05}$, $\chi^2_{0,01}$.

Доведення універсальності міри CCK. Міра CCK не використовує ніяких числових і порядкових характеристик категорій, вона може бути обчислена для будь-яких ознак, як метричних, так і неметричних. Отже, тому вона є універсальною мірою тісноти зв'язку. Характеристика Eta_0 є доброякісною альтернативою міри CKK, що рекомендується. Ці дві міри дуже тісно зв'язані між собою, хоча обчислюються по-різному. Цей зв'язок чітко видно на рис. 4.4в.

Таким чином, коефіцієнт CCK є об'єктивною, універсальною мірою для всіх типів ознак, незалежно від шкали їх вимірювання.

Продовжуючи розв'язання поставленої задачі, були обчислені коефіцієнти парної кореляції Пірсона (R_{xy}) і після ранжування значень величин всіх ознак – коефіцієнти рангової кореляції Спірмена (ρ_s). Для порядкових ознак ці коефіцієнти співпадають, це теоретично доводилось у підрозділі 3.1; у поставленій задачі наведені коефіцієнти дуже близькі між собою (рис. 4.5).

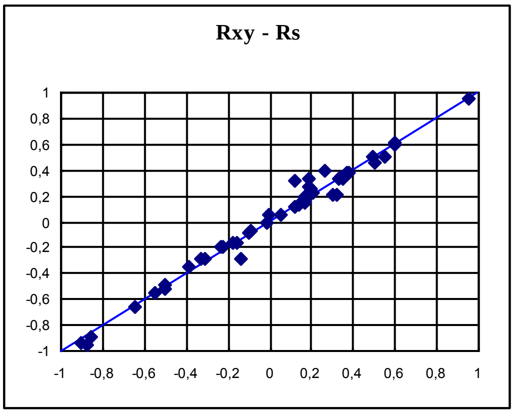


Рис. 4.5. **Залежність між коефіцієнтами R_{xy} і ρ_s**

У табл. 4.21 наведені обчислені значення R_{xy}, Eta_0 та CCK для всіх 45 пар ознак.

Таблиця 4.21

Значення мір тісноти зв’язку між ознаками

		R_{xy}	ρ_s	Eta_0	CCK			R_{xy}	ρ_s	Eta_0	CCK
X	*Y*	-0,90333	-0,93776	0,978319	0,908717	*U*	*V*	-0,88010	-0,94912	0,969425	0,818187
X	*Z*	0,954945	0,947458	0,967005	0,87742	*U*	*W*	0,491903	0,496439	0,748359	0,806083
X	*U*	0,123721	0,317255	0,808684	0,679336	*U*	*C*	-0,00813	0,058171	0,549217	0,597131
X	*V*	-0,23529	-0,20102	0,616123	0,728059	*U*	*D*	-0,09385	-0,06835	0,37228	0,508863
X	*W*	0,317094	0,212615	0,759547	0,735144	*U*	*E*	0,185803	0,276657	0,541434	0,715745
X	*C*	0,340174	0,346462	0,715748	0,734926	*U*	*F*	0,265195	0,401481	0,787572	0,7328
X	*D*	0,170161	0,192837	0,619074	0,625359	*V*	*W*	-0,49957	-0,48578	0,551597	0,714547
X	*E*	0,551954	0,504487	0,683965	0,685991	*V*	*C*	-0,01701	-0,00084	0,443533	0,538171
X	*F*	0,598919	0,609516	0,67883	0,653715	*V*	*D*	0,124349	0,116308	0,255996	0,568892
Y	*Z*	-0,86026	-0,89898	0,956966	0,80954	*V*	*E*	-0,15548	-0,16300	0,556407	0,744812
Y	*U*	-0,14343	-0,28645	0,660193	0,721903	*V*	*F*	-0,32783	-0,28391	0,541935	0,728049
Y	*V*	0,169176	0,141257	0,537293	0,714547	*W*	*C*	-0,17588	-0,15955	0,5581	0,69609
Y	*W*	-0,30734	-0,28654	0,698666	0,73125	*W*	*D*	-0,6479	-0,65175	0,752826	0,711662
Y	*C*	-0,38457	-0,35400	0,583236	0,647144	*W*	*E*	0,052713	0,061975	0,509163	0,688034
Y	*D*	-0,09707	-0,08218	0,598136	0,580818	*W*	*F*	0,141068	0,127569	0,757842	0,679097
Y	*E*	-0,50341	-0,51166	0,675146	0,6756	*C*	*D*	0,372168	0,382809	0,465562	0,544867
Y	*F*	-0,54736	-0,55761	0,64305	0,643094	*C*	*E*	0,329683	0,337677	0,472533	0,486169
Z	*U*	0,184977	0,332068	0,722395	0,688655	*C*	*F*	0,346353	0,336517	0,355776	0,328509
Z	*V*	-0,22469	-0,20146	0,496678	0,637658	*D*	*E*	0,374803	0,380812	0,421075	0,387383
Z	*W*	0,298364	0,208724	0,719639	0,755238	*D*	*F*	0,168342	0,166423	0,276169	0,545136
Z	*C*	0,198518	0,254903	0,613208	0,749299	*E*	*F*	0,366599	0,369922	0,493197	0,502058
Z	*D*	0,201382	0,221481	0,591711	0,71031						
Z	*E*	0,602148	0,590847	0,632997	0,611863						
Z	*F*	0,501697	0,459758	0,663563	0,698955						

Для візуалізації та подальшого аналізу дані табл. 4.21 можна подати у вигляді матриць, як це зображено в табл. 4.22.

Таблиця 4.22

Матриці R_{xy}, ρ_s, CCK

R_{xy}	*X*	*Y*	*Z*	*U*	*V*	*W*	*C*	*D*	*E*	*F*
1	2	3	4	5	6	7	8	9	10	11
X	1	-0,90333	0,954945	0,123721	-0,23529	0,317094	0,340174	0,170161	0,551954	0,598919
Y	-0,90333	1	-0,86026	-0,14343	0,169176	-0,30734	-0,38457	-0,09707	-0,50341	-0,54736
Z	0,954945	-0,86026	1	0,184977	-0,22469	0,298364	0,198518	0,201382	0,602148	0,501697
U	0,123721	-0,14343	0,184977	1	-0,8801	0,491903	-0,00813	-0,09385	0,185803	0,265195
V	-0,23529	0,169176	-0,22469	-0,8801	1	-0,49957	-0,01701	0,124349	-0,15548	-0,32783
W	0,317094	-0,30734	0,298364	0,491903	-0,49957	1	-0,17588	-0,6479	0,052713	0,141068
C	0,340174	-0,38457	0,198518	-0,00813	-0,01701	-0,17588	1	0,372168	0,329683	0,346353
D	0,170161	-0,09707	0,201382	-0,09385	0,124349	-0,6479	0,372168	1	0,374803	0,168342
E	0,551954	-0,50341	0,602148	0,185803	-0,15548	0,052713	0,329683	0,374803	1	0,366599
F	0,598919	-0,54736	0,501697	0,265195	-0,32783	0,141068	0,346353	0,168342	0,366599	1
ρ_s										
X	1	-0,93776	0,947458	0,317254	-0,20102	0,212615	0,346461	0,192837	0,504487	0,609516
Y	-0,93776	1	-0,89899	-0,28645	0,141257	-0,28654	-0,35400	-0,08218	-0,51166	-0,55761
Z	0,947458	-0,89899	1	0,332068	-0,20146	0,208724	0,254903	0,221481	0,590847	0,459758
U	0,317254	-0,28645	0,332068	1	-0,94912	0,496439	0,058171	-0,06835	0,276657	0,401481
V	-0,20102	0,141257	-0,20146	-0,94912	1	-0,48578	-0,00084	0,116308	-0,16390	-0,28391
W	0,212615	-0,28654	0,208724	0,496439	-0,48578	1	-0,15955	-0,65175	0,061975	0,127569

Закінчення табл. 4.22

1	2	3	4	5	6	7	8	9	10	11
C	0,346461	-0,35400	0,254903	0,058171	-0,00084	-0,15955	1	0,382809	0,337677	0,336517
D	0,192837	-0,08218	0,221481	-0,06835	0,116308	-0,65175	0,382809	1	0,380812	0,166423
E	0,504487	-0,51166	0,590847	0,276657	-0,16390	0,061975	0,337677	0,380812	1	0,369921
F	0,609516	-0,55761	0,459758	0,401481	-0,28391	0,127569	0,336517	0,166423	0,369921	1
CCK										
X	1	0,908717	0,87742	0,87742	0,728059	0,735144	0,734926	0,625359	0,685991	0,653715
Y	0,908717	1	0,80954	0,721903	0,714547	0,73125	0,647144	0,580818	0,6756	0,643094
Z	0,87742	0,80954	1	0,688655	0,637658	0,755238	0,91696	0,814682	0,77083	0,763474
U	0,87742	0,721903	0,688655	1	0,818187	0,806083	0,79795	0,745509	0,88758	0,967951
V	0,728059	0,714547	0,637658	0,818187	1	0,714547	0,70597	0,768419	0,843782	0,941349
W	0,735144	0,73125	0,755238	0,806083	0,714547	1	0,738435	0,848156	0,844966	0,854687
C	0,734926	0,647144	0,91696	0,79795	0,70597	0,738435	1	0,544867	0,486169	0,328509
D	0,625359	0,580818	0,814682	0,745509	0,768419	0,848156	0,544867	1	0,387383	0,545136
E	0,685991	0,6756	0,77083	0,88758	0,843782	0,844966	0,486169	0,387383	1	0,502058
F	0,653715	0,643094	0,763474	0,967951	0,941349	0,854687	0,328509	0,545136	0,502058	1

Наведені матриці є початковими даними для розробки модифікацій математичних методів та їх доведень. Таким чином, розроблена об’єктивна, універсальна міра тісноти взаємозв’язку між ознаками, яка надає широкі можливості для модифікації математичних методів, зокрема методів БСА для всіх ознак соціально-економічних систем.

Запитання для самоперевірки

1. Сформулювати загальні рекомендації щодо розбудови описових моделей у порядкових шкалах.
2. Які основні задачі теорії рангової кореляції?
3. Які основні інструменти рангової кореляції?
4. Який взаємозв’язок між коефіцієнтами рангової кореляції та коефіцієнтом кореляції Пірсона?
5. З якої форми можна отримати всі відомі коефіцієнти кореляції?
6. Який зміст початкового аналізу соціально-економічних систем за порядковими ознаками?
7. Яке призначення теста χ^2 в багатовимірному аналізі за порядковими ознаками?
8. Який зв'язок між коефіцієнтом асоціації Пірсона та коефіцієнтом контингенції Крамера?
9. Які коефіцієнти взаємозв’язку між номінальними ознаками рекомендується обчислювати перш за все та який їх зміст?
10. Яка міра взаємозв’язку між різними ознаками об’єктивна та універсальна?
11. Яке значення має коефіцієнт *CCK* для математичних методів в економіці?

Розділ 5
Використання вимірників в аналізі соціально-економічних систем

5.1. Математичні методи побудови узагальнюючих показників в економіці

Характеризуючи системи різної природи, завжди переходять від елементарних ознак їх опису до узагальнення, синтезу, що виражається окремою величиною. В основному це величина, що визначає складну ознаку. Початковим етапом визначення величини складної ознаки є її розгляд як агрегації (синтезу) елементарних. Перехід до агрегованих ознак не тільки доповнює опис системи, але має математичну доцільність. Агрегований опис містить порівняно з початковим дещо менше інформації, при цьому корисна інформація залишається, а надмірна – звужується. Найчастішими прийомами агрегації числової інформації є перехід до суми, середньозваженого: якщо потрібно агрегувати показники (скаляри, вектори, матриці однієї розмірності і т. д.) $a_1, a_2, ..., a_n$, то за агрегат A відповідно приймають $A = a_1 + a_2 + ... + a_n$ або $A = \frac{a_1 + a_2 + ... + a_n}{n}$, або $A = k_1 a_1 + k_2 a_2 + ... + k_n a_n$, де коефіцієнти ваги повинні задовольняти умовам $k_1, k_2, ..., k_n \geq 0$ і $k_1 + k_2 + ... + k_n = 1$. Слід зауважити, що аналогічно агрегуються рівняння, нерівності, інші математичні об'єкти, а також графічні моделі, при цьому метод агрегації визначається залежно від способу опису системи та конкретної задачі (цілі) його дослідження [308, с. 10 – 11].

У математиці відомі дві форми згортки сукупності показників: адитивна і мультиплікативна [123; 172; 176; 186]. Адитивна згортка здійснюється за формулою:

$$I_{\sum} = \sum_{i=1}^{n} \lambda_i x_i ,$$

де x_i – значення i-го показника, що виміряний на шкалі інтервалів або відношень;

λ_i – коефіцієнт значущості (важливості) показника.

Мультиплікативна згортка здійснюється за формулою:

$$I_{\prod} = \prod_{i=1}^{n} x_i^{\lambda_i} .$$

На практиці найчастіше застосовується адитивна згортка в модифікованому вигляді:

$$I_{\sum} = \left(\sum_{i=1}^{n} \lambda_i \left| x_i - \alpha_i \right|^k \right)^{\frac{1}{k}},$$

де α_i – еталонне значення показника;

k – параметр.

Часто в якості параметра використовується $k = 2$, що приводить до евклідової відстані між вектором показників та вектором еталонів. Збільшення k спричиняє зростання ролі найбільшого відхилення, і навпаки. При низьких значеннях $k \approx 0$ модифікована згортка наближається до мультиплікативної згортки.

Результат агрегації в економіко-математичному моделюванні задається відображенням з простору $X \times A$ змінних $x \in X$ і параметрів $a \in A$ початкової моделі в простір $Z \times B$ змінних $z \in Z$ і параметрів $b \in B$ агрегованої моделі: $X \times A \to Z \times B$, це у випадку агрегації змінних і параметрів [317, с. 11]. Зазвичай це відображення визначається набором функцій від змінних x, параметрів a та від додатково заданих параметрів агрегування λ: $z = f(a, x, \lambda)$, $b = g(a, x, \lambda)$. У якості параметрів агрегації λ можуть використовуватися значення спеціальним чином підібраних функцій $\lambda = h(x^1, x^2, \ldots)$ від одного або декількох наближених розв'язків початкової моделі, тобто від елементів $x^1, x^2, \ldots$ простору її змінних *X*. Клас відображень, що задають агрегування, вибирається з урахуванням встановлених цілей, виду і властивостей початкової моделі.

У роботі [296] викладені ідеї побудови "формули порівняльної оцінки" проектів лінкорів відомим адміралом О. М. Криловим, який продемонстрував зразок методу згортки багатьох оцінок в одну, що надзвичайно широко використовується для побудови зведених (глобальних, інтегральних, узагальнених, генеральних, синтетичних тощо) показників, що синтезують окремі (локальні, диференціальні, маргінальні, аналітичні) показники, які характеризують якість (ефективність, надійність, корисність, надання переваги і т. д.) будь-яких багатопараметричних об'єктів. Н. В. Хованов у тій же роботі рекомендує будувати спрощену схему розробки зведеного показника Q окремого об'єкта:

1. Формується вектор $\boldsymbol{X} = (x_1, \ldots, x_n)$ вихідних характеристик, кожна з яких необхідна, а всі разом – достатні для повного, всебічного оцінювання об'єкта.

2. Формується вектор $\boldsymbol{Q} = (q_1, \ldots, q_n)$ окремих показників, що подаються функціями $q_i(X)$, $i = \overline{1, m}$, вектора початкових показників $\boldsymbol{X} = (x_1, \ldots, x_n)$ та оцінюють різні аспекти об'єкта з використанням m різних критеріїв. У спрощеному випадку кожний окремий показник $q_i(X)$ є функцією однієї початкової характеристики x_i: $q_i = q_i(x_i), i = \overline{1, m = n}$.

3. Вибирається вид функції $Q(q)$, яка синтезує і порівнює з вектором окремих показників $\vec{q}(q_1, \ldots, q_m)$ зведену оцінку Q (значення зведеного показника $Q = Q(q)$), що характеризує об'єкт у цілому. Зазвичай передбачається, що синтезувальна функція $Q(q)$ залежить від вектора $W = (\omega_1, \ldots, \omega_m)$ додатних параметрів $\omega_1, \ldots, \omega_m$, які визначають значимість окремих показників q_1, ..., q_m для зведеної оцінки Q відповідно: $Q = Q(q) = Q(q; \omega)$.

4. Визначається значення вектора ω параметрів $\omega_1, \ldots, \omega_m$, $\omega_i \geq 0$, які інтерпретуються як вагові коефіцієнти (вага), що задають ступені впливу окремих показників $q_1, \ldots, q_m$ на зведену оцінку Q. Часто використовується додаткова умова нормування ($\omega_1 + \ldots + \omega_m = 1$), яка дозволяє розглядати ω_i як оцінку відносної ваги окремого показника q_i.

Цей метод належить до типу кваліметричних методів, які є найпоширенішими у побудові зведених показників. Основою другого методу є теорія функцій корисності. Сьогодні накопичено дуже багато математичних моделей, що дозволяють будувати функції, які оцінюють корисність у цілому різних наборів господарських благ. При цьому такі зведені оцінки корисності складних наборів благ можуть мати як числову ("кардинальну корисність"), так і нечислову ("ординальну корисність") [296]. Третє джерело методу зведених показників криється в теорії економічних індексів, що оцінюють єдиним числом багатопараметричні об'єкти і явища. Для здійснення наведених етапів Н. В. Хованов рекомендує дотримуватися наступних принципів: а) принципу лінеаризації, який дозволяє переходити від окремого упорядкування до лінійного впорядкування множини зведених оцінок цих об'єктів; б) принципу арифметизації, який дозволяє отримати числові оцінки для початкової нечислової інформації, що лежить в основі побудови показників і визначення вагових коефіцієнтів; в) принципу рандомізації, що дозволяє моделювати дефіцит інформації, який зазвичай існує на всіх етапах синтезу зведених оцінок складних багатопараметричних об'єктів.

В економіці широко використовується синтезована інформація, частіше всього у вигляді узагальнюючого показника, що виражає якість властивостей соціально-економічної системи або всієї системи.

В економічному аналізі передбачається побудова узагальнюючих показників. Тут найпоширенішими методами є бальна або рейтингова оцінка окремих коефіцієнтів, які характеризують стан підприємства, оцінка ефективності виробництва, що побудована на додаванні ефективності його ресурсів – в адитивній чи мультиплікативній формі, за допомогою кратних моделей, змішаних моделей та метод порівняння з еталонним підприємством за значущими економічними параметрами [26; 36; 114; 117; 121]. Узагалі зауважимо, що існують два основних підходи до оцінки функціонування та розвитку об’єкта: на основі використання системи елементарних економічних показників та одного економічного показника. У другому підході акцентується економічність функціонування об’єкта на основі порівняння вартості та ефективності або витрат та прибутку, при цьому узагальнений економічний показник є відносним. Проте більшість економістів дотримуються спільної думки, що складні явища, процеси неправомірно характеризувати одним економічним показником. Цей показник може бути абсолютним чи відносним й передбачатися статистичною звітністю або обчислюватися як відношення показника, що свідчить про результат, до показника, який свідчить про загальні витрати, які забезпечили результат [193; 194; 196; 201; 234; 246]. Отже, аналітично проблема існування узагальнюючих показників в економіці зводиться до проблеми їх побудови, а точніше, до того, який підхід використовується – економічний (як відношення результату до затрат) чи математичний (за допомогою спеціального математичного методу). Узагальнені показники, що побудовані за допомогою математичного методу, приймають значення від 0 до 1. Саме за допомогою математичного методу здійснюється механізм взаємозв’язку "аналіз − синтез". Система окремих показників дозволяє поглиблено вивчити елементарні ознаки об’єкта, а відновлення єдності за допомогою математичного методу синтезує загальну характеристику соціально-економічної системи.

Згортку величин ознак можна здійснити за допомогою факторного аналізу, канонічних кореляцій, багатовимірного шкалювання, кластерного та дискримінантного аналізів, тобто методів багатовимірного статистичного аналізу, а також за допомогою евристичних методів зниження розмірності [50; 52; 66; 74; 235]. Але перелічені методи виконують неповну редукцію ознак; їхнє використання дозволяє значно скоротити кількість ознак, але в сукупності завжди залишається більше однієї ознаки.

Найбільш поширені методики обчислення узагальнюючого показника в економіці спираються на формулу [1]:

$$Y_i = \sum_{j=1}^{n} \left(\frac{y_{ij} - y_{jo}}{y_{jo}} \right)^2 , \qquad (5.1)$$

де y_{jo} – найкраще (ідеальне) еталонне значення j-го показника.

Якщо в i-й момент часу значення показників ознак збігаються з еталонним, то Y_i дорівнює нулю. Це і є те значення, до якого необхідно прагнути (чим ближче до нуля, тим краще).

Основні недоліки таких методів зводяться до нівелювання окремих показників ознак. На практиці різні показники бувають неоднаково значущими чи важливими. Усувають цей недолік уведенням певної ваги α_j:

$$Y_i = \sum_{j=1}^{n} \alpha_j \left(\frac{y_{ij} - y_{jo}}{y_{jo}} \right)^2 , \qquad (5.2)$$

причому $\sum_{j=1}^{n} \alpha_j = 1, \alpha_j > 0$. Ранжування за ступенем важливості показників та визначення їх ваги виконують за допомогою експертних оцінок. Якщо зміна значень показника ознаки відбувається у двох межах, тобто $y_{min} \le y \le y_{max}$, то пропонується у формулі (5.2) задавати y_{j0} як середнє:

$$y_{j0} = \frac{y_{j\,max} + y_{j\,min}}{2} .$$

Рекомендують також розрізняти статичні згортки показників і динамічні. При статичній згортці показники повинні вимірюватися в один момент часу. Динамічна згортка характеризується тим, що значення показників вимірюється протягом періодів часу та згортка їх здійснюється в послідовні періоди часу, таким чином відображаючи історію змін показників, тобто процес [105 , с. 43].

Для побудови соціальних індикаторів Ф. М. Бородкін і С. А. Айвазян рекомендують використовувати уніфіковані шкали, які надають змогу перетворювати окремі індикатори в безрозмірну шкалу [18, с. 341 – 355]. Вони стверджують, що уніфікація шкали – це таке її перетворення (перенесення початку відліку і зміна масштабу), у результаті якого область можливих значень вимірювання завжди обмежується відрізком $[0; N]$, де число N призначається і вказує розмах нової шкали. При цьому нульове значення (початок відліку) перетвореного показника має відповідати найменшій якості даної властивості, а значення, що дорівнює N, – найвищій якості. Вони також вважають, що конкретний вибір уніфікованого перетворення залежить від того, до якого з трьох типів належить показник, що аналізується. Окремий показник може бути зв'язаний з якістю, що синтезується, монотонно зростаючою залежністю, монотонно спадною або немонотонною залежністю. Ф. М. Бородкін і С. А. Айвазян рекомендують: якщо залежність монотонно зростає, значення відповідної уніфікованої змінної $\tilde{x}$ визначати за формулою:

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \cdot N ,$$

де $x_{\min}$, $x_{\max}$ – відповідно найменше і найбільше значення показника;

якщо залежність монотонно спадна, значення відповідної уніфікованої змінної $\tilde{x}$ визначати за формулою:

$$\tilde{x} = \frac{x_{\max} - x}{x_{\max} - x_{\min}} \cdot N ;$$

якщо залежність немонотонна, значення відповідної уніфікованої змінної $\tilde{x}$ визначати за формулою:

$$\tilde{x} = \left(1 - \frac{|x - x_{onm}|}{max\,(x_{max} - x_{onm}; x_{onm} - x_{min})} \right) \cdot N .$$

Складність перетворень пояснюється визначенням $x_{min}, x_{max}, x_{onm}$, яке можливе завдяки змістовним обґрунтуванням, за моделлю, за встановленими нормативами.

Далі для редукції простих індикаторів у складний рекомендуються наступні етапи формування інформаційної бази:

1) визначення початкової апріорної системи показників;

2) реалізації методики відбору з апріорної системи окремих показників меншої кількості, тобто розв'язати задачі формування апостеріорної системи показників. На даному етапі розв'язуються такі підзадачі: а) перехід від початкових показників до відповідних уніфікованих їх значень; б) з метою аналізу мультиколінеарності обчислення парних коефіцієнтів кореляції уніфікованих показників, обчислення коефіцієнтів детермінації і на цій основі виявлення

тісноти зв’язку між показниками. З груп тісно зв’язаних показників слід відібрати по одному представнику;

3) розв’язання задачі зведення багатокритеріальної схеми порівняльного аналізу об’єктів, що ґрунтується на окремих показниках, до однокритеріальної. При цьому синтетичний чи інтегральний індикатор $\tilde{y}$ може мати форму лінійної або нелінійної функції від окремих показників (окремих індикаторів).

Якщо $\dfrac{\lambda_1}{\sum\limits_{j=1}^{p} \lambda_j} \ge 0{,}55$, де $\lambda_1, \lambda_2, ..., \lambda_p$ – власні значення кореляційної матриці апостеріорної системи показників, які розміщуються в порядку спадання, то рекомендується інтегральний індикатор визначати у формі лінійної згортки як модифікацію першої головної компоненти. Якщо інтегральний індикатор є нелінійною функцією від окремих, то рекомендується спочатку розбити апостеріорну систему показників на групи відповідно до встановлених вимог: характеризувати який-небудь один аспект якості, що синтезується; мати відносно високий рівень взаємного коригування; бути мало корельованими між собою. Далі в кожній групі слід визначити модифіковану першу головну компоненту. Значення зведеного інтегрального індикатора обчислюється за груповими інтегральними індикаторами на основі зваженої евклідової відстані до еталону $E = (N; N; ...; N)$. Описаний метод згортки індикаторів можна розглядати як моделювання інтегрального індикатора, оскільки, перш ніж обчислити величину, потрібно аналітично відібрати найупливовіші прості індикатори, а потім вибрати форму функції згортки.

У цілому ж усі відомі математичні методи побудови узагальнюючих показників в економіці слід розглядати як дві великі групи. До першої групи належать методи, що передбачають існування еталону досягнення значень показників (еталонних рівнів величини ознаки) та знаходження відстані від реального рівня до еталонного. Існує багато способів уведення метрики, що задає "близькість до ідеалу". Формули знаходження відстані вже були описані в підрозділі 2.2 посібника. До цієї групи належить також метод побудови таксономічного показника розвитку В. Плюти [216], коротко розглянемо його алгоритм.

З. Хельвиг одним із перших став використовувати спеціальну дослідницьку методику узагальнення ознак [216]. Ним був запропонований так званий таксономічний показник, що становить синтетичну величину, утворену на основі системи показників ознак. Існує думка, що слід залишати в системі лише ті показники, які мають значну інформаційну цінність, розраховану за показниками їх абсолютної та відносної інформаційної цінності. На думку В. Плюти, результати, що найбільше підходять для подальшого аналізу даних, дає метод Я. Чекановського [216], який передбачає аналіз форми сліду багатовимірного розподілу ознак об’єкта.

Дослідження починається із застосування придатного методу угруповання, наприклад, методу куль чи побудови дендритів. Для економетричних досліджень слід виділяти підмножини, що складаються з елементів з однорідним рівнем чи з однорідною структурою значень ознак. Однорідні підмножини одержують при застосуванні таксономічних процедур. Сукупність операцій щодо виділення підмножин даних, однорідних за рівнем значень ознак, називають ізотонічним аналізом, а самі одержані підмножини – ізотонічними підмножинами. Розбивання сукупності на однорідні підмножини, елементи яких у багатовимірному просторі властивостей утворюють хмару точок у формі еліпсоїда, дозволяє об’єктам, що входять до складу підмножин, мати близькі структури значень ознак. Підмножини, об’єкти яких мають такі властивості, називають ізоморфічними. Щоб отримати ізоморфічні підмножини, потрібно перетворити реалізації ознак, у яких міститься інформація про дві компоненти, що характеризують рівень та структуру. Про структуру значень ознак-об’єктів можна робити висновок за величиною кута між векторами-об’єктами, що розглядаються, а про рівень значень ознак-об’єктів – за довжиною кожного вектора-об’єкта.

Виділяють ще ізотропні підмножини, що ґрунтуються на використанні методу куль. При такому способі розбиття існує потенційна можливість розділити дійсно однорідні об’єкти при розумінні однорідності як структурної подібності об’єктів, тобто подібності за структурою значень ознак, що описують об’єкти. Таке небажане розбиття виникає внаслідок того, що в значеннях ознак присутні обидві компоненти (структури і потенціалу). Тому обґрунтованішими є пропозиції використовувати той спосіб, що приводить до виділення підмножин ізоморфічно однорідних.

Побудова дендрита на ізотропних підмножинах дозволяє не тільки попередньо визначити форму сліду розподілу, й виділити зв'язні підмножини, тобто такі підмножини об'єктів, де реалізації ознак міняються досить плавно. З погляду вимог економетрії, ця обставина дуже істотна, тому що тим самим виключаються всі значні пропуски, стрибки і розриви, що можуть зустрітися у початковій сукупності реалізацій ознак. При виділенні декількох ізольованих зв'язних підмножин подальший аналіз форми сліду розподілу проводиться окремо для кожного з них.

Можна скоротити число розглянутих ознак, але так, щоб, з одного боку, число діагностичних ознак було невелике, а з іншого – не відбулося значної втрати інформації [216]. Виконання обох вимог можливе, коли діагностичні ознаки взаємонезалежні, а ознаки, що не входять у число діагностичних, від них залежать. Іноді висуваються додаткові вимоги, щоб діагностичні ознаки володіли високою поділяючою здатністю і не були залежні від зовнішніх впливів.

Однак дві перші умови є найважливішими, тому методам одержання діагностичних ознак із заданими властивостями приділяється багато уваги (С. Бартосевич, З. Хельвиг, В. Плюта). Неповну редукцію можна робити двома способами. Перший спосіб приводить до одержання так званих індивідуальних діагностичних ознак, якими є деякі з вихідних ознак. Сукупність Φ скорочується до сукупності:

$$\Phi' = \{\varphi_1, \varphi_2, ..., \varphi_P\},$$

причому $P \le N$.

Другий – полягає в побудові деяких синтетичних величин. Тоді сукупність Φ зменшується, а її елементи перетворюються, і в результаті можна одержати сукупність:

$$\Phi'' = \{\varphi'_1, \varphi'_2, ..., \varphi'_P\},$$

причому $P \le N$.

Ознаки φ'_p називаються синтетичними діагностичними ознаками.

Залежно від того, яким способом були отримані синтетичні ознаки, розрізняють: таксономічні показники, загальні фактори, головні компоненти й агреговані діагностичні ознаки.

При одержанні діагностичних ознак обох видів – як індивідуальних, так і синтетичних (виключаючи фактори й компоненти) – необхідно виконати дві операції. Одна з них полягає в розбивці сукупності вихідних ознак Φ на підмножини Φ_p $(p = \overline{1, P})$ однорідних елементів. До складу однорідних підмножин входять не тільки істотно корельовані ознаки, але й ознаки, однорідні з погляду якісних зв'язків, а це означає, що вони відносяться до однієї й тієї ж області досліджуваних процесів та явищ. Таким чином, сукупність Φ поділяється на непорожні підмножини Φ_p:

$$\Phi = \{\Phi_1, \Phi_2, ..., \Phi_P\}$$

з наступними властивостями:

$$\bigcup_p \Phi_P = \Phi,\ \Phi_m \bigcap \Phi_P = \varnothing,\ m \neq p,\quad m = \overline{1, P},\ p = \overline{1, P},$$

де $\Phi = \{\varphi_1, \varphi_2, ..., \varphi_{n_p}\}$.

Отже, сукупність ознак Φ виявляється розбитою на непорожні та непересічні підмножини, причому всі ознаки φ_n сукупності Φ входять у підмножини Φ_p і кожна з ознак φ_n може ввійти до складу тільки однієї з підмножин Φ_p. Елементи кожної підмножини розглядаються як характеристики визначеної сторони явища. Розбивку сукупності можна виконати різними методами (В. Букетинский, К. Флерек, Я. Коленко та ін.) [216].

Наступна операція – це визначення репрезентантів (ознак-представників) виділених підмножин ознак:

$$\varphi_p^* = f(\varphi_1, \varphi_2, ..., \varphi_{n_p}) \text{ для } p = \overline{1, P},$$

де φ_p^* – репрезентант p-ї однорідної підмножини ознак;

n_p – число ознак у p-й однорідній підмножині ознак.

Якщо репрезентант є ознакою, що входить до підмножини, то його називають індивідуальною діагностичною ознакою (В. Плюта). Якщо ж як репрезентант вибирається результуюча всіх ознак однієї підмножини, то він називається синтетичною діагностичною ознакою. У цій якості часто використовують таксономічний показник, що одержується за допомогою відповідної таксономічної процедури (З. Хельвиг).

Описана методика побудови індивідуальних і синтетичних діагностичних ознак має деякі недоліки. Так, при використанні індивідуальних діагностичних ознак немає цілковитої впевненості в тому, що саме ця ознака визначальна. Пояснення тут таке. Величини, на підставі яких вибираються репрезентанти підмножин ознак, не завжди помітно відрізняються одна від одної. Це значить, що роль індивідуальної діагностичної ознаки можуть виконувати декілька ознак тієї ж підмножини, та коли дисперсія ознаки буде зростати, ознака виявиться за межами тієї підмножини, що існує зараз.

Слід зазначити, що автори теж згодні, що негативною рисою синтетичних діагностичних ознак є суб'єктивний поділ ознак на стимулятори і дестимулятори та номінатори. У загальному вигляді вони визначаються так:

стимулятори $(x_{sn} \geq x_{rn}) \Rightarrow (\omega_s \succ \omega_r)$; (5.3)

дестимулятори $(x_{sn} \geq x_{rn}) \Rightarrow (\omega_s \prec \omega_r)$; (5.4)

номінатори $(x_{pn} \geq x_{sn} \geq x_{rn}) \Rightarrow (\omega_s \succ \omega_r)$, (5.5)

$(x_{sn} \geq x_{rn} > x_{pn}) \Rightarrow (\omega_s \prec \omega_r)$,

або $(x_{pn} \geq x_{sn} \geq x_{rn}) \Rightarrow (\omega_s \prec \omega_r)$, (5.6)

$(x_{sn} \geq x_{rn} > x_{pn}) \Rightarrow (\omega_s \succ \omega_r)$.

Для обчислень перетворюють дестимулятори в стимулятори, оскільки будуть використовуватися ізоморфічні перетворення ознак, які дають позитивні результати тільки тоді, коли досліджувана сукупність ознак складається зі стимуляторів.

Неправильна класифікація ознак призводить до одержання необґрунтованих значень таксономічного показника. Тому поміж теоретичного суттєвого визначення, яким є показник для досліджуваного явища в економіці, слід використовувати інструмент описової статистики. Особливо описова статистика корисна у випадку показника-номінатора, він має двосторонній розподіл значень. Статистичний аналіз демонструє, яка тенденція склалася за даний період, а яку б хотілося мати згідно з управлінським рішенням щодо регулювання поточного процесу в економіці, установлює вибір категорії показника (стимулятор чи дестимулятор).

Вважається, що згадані недоліки можна значною мірою пом'якшити, якщо застосовувати факторний аналіз чи канонічний разом із таксономічними методами, це дійсно підтверджується практично. Багатовимірні статистичні методи використовуються для розбивки сукупності вихідних ознак і одержання підмножин з відносно однорідними елементами. У кожній виділеній групі щодо однорідних ознак за допомогою факторного аналізу безпосередньо будується своя агрегатна діагностична ознака. Агреговані діагностичні ознаки слабо або некорельовані між собою.

При побудові таксономічного показника за методом В. Плюти вирішують обчислювальні проблеми, логіка яких наведена на рис. 5.1.

Визначення стимуляторів, дестимуляторів, номінаторів серед показників ознак соціально-економічної системи:
$X = (x_{ij}), i = \overline{1,m}, j = \overline{1,n}$, де i -та ознака на j -й період або об'єкт

⇩

Формування еталона:
1) за критерієм МініМакс; 2) еталонні значення встановлюються

⇩

Нормування або стандартизація показників:

$$Z = (z_{ij}); \; z_{ij} = \frac{x_{ij} - \overline{x_i}}{\sigma_i}, \; \overline{x_i} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, \sigma_i = \sqrt{\frac{\sum_{j=1}^{n}(x_{ij} - \overline{x_i})^2}{n}}$$

⇩

Обчислення значень узагальнюючого показника:

$$d_j = \left(\sum_{i=1}^{m}(z_{ij} - z_{i0})^2\right)^{\frac{1}{2}}; \; \overline{d} = \frac{1}{n}\sum_{j=1}^{n} d_j; \; \delta = \overline{d} \text{ або } \delta = M_e;$$

$$s_d = \left(\frac{1}{n}\sum_{j=1}^{n}(d_j - \delta)^2\right)^{\frac{1}{2}}; \; d = \delta + as_d; \; a = 3; \; d = \delta + as_d;$$

$$I_j = \frac{d_j}{d}; \; I_j^* = 1 - I_j$$

Рис. 5.1. **Логіка побудови таксономічного показника розвитку**

Наведені обчислення виконують для ознак-стимуляторів. Подана схема зображує основні моменти реалізації математичного методу побудови таксономічного показника в розв'язуванні різних задач. Перш за все, слід відмітити відмінності методів формування еталона. При використанні у формуванні еталона критерію МініМакс рівні узагальнюючих показників порівняльні локально в даній вибірці, зрівнювати її з іншими не об'єктивно. Коли ж формуємо еталон, встановлюючи значення показників і спираючись на нормативні значення чи директивно-управлінські, заплановані, експертні, маємо оцінку в глобальному порівнянні та можна порівнювати об'єкти з різних груп. У розв'язанні економічних задач це може інтерпретуватись так: порівняльний аналіз відносно еталона за критерієм МініМакс дозволяє ставити та досягати вирішення локальних задач оперативного управління підприємством, а відносно еталона, що установлюється, – глобальні задачі стратегічного управління підприємством.

Проблеми обчислення величин при розрахунку таксономічного показника складаються з обчислення величин a та δ. Величина a – кількість середньоквадратичних відхилень у частках σ, яке може дорівнювати 2, якщо розподіл ознаки симетричний, або 3 – у загальному випадку. Частіше всього a приймають рівним 3. Безсумнівно, якщо в задачі потрібно досягти визначеної точності, то всі показники слід діагностувати на симетричність.

Характерною властивістю інтегрального показника I_j є те, що його значення знаходиться в інтервалі від 0 до 1. Відповідно до обчислень інтерпретація таксономічного показника не узгоджується з інтуїтивними представленнями (таксономічний показник зростає з віддаленням значень показників від еталонного й спадає з наближенням їх до нього). Тому таксономічний показник привели до виду $I_j^* = 1 - I_j$. Інтерпретація даного показника наступна: він приймає високі значення при близьких значеннях показників у системі до еталона й низькі значення при далеких [216].

Існує безліч різних варіантів побудови таксономічного показника. Згідно з одним із них процес побудови модифікованого варіанта таксономічного показника починається зі стандартизації ознак, а також, за потребою, із уведення коефіцієнтів ієрархії та перетворення ознак до форми стимуляторів. Потім варто задати координати нижнього полюса. Нехай це буде точка N-вимірного простору. Установлюють її координати на одному рівні, рівному $(-a)$ стандартних відхилень, тобто координати нижнього полюса записуються в такий спосіб:

$$R_0 = (-a, -a, \ldots, -a),$$

де a – довільне позитивне число, відповідно до правила "трьох (двох) сигм", тобто $a = 3$ чи $a = 2$.

Є пропозиція, що для розв’язання практичних задач значення a доцільніше встановлювати на низькому рівні, але вона не зовсім обґрунтована, оскільки для групи об’єктів сукупність значень може виявитися надто розрідженою.

Можна взяти верхній полюс: $P = (a, a, \ldots, a)$, але тоді знову виникає складність у інтерпретації узагальнюючого показника. При такому положенні в аналізі варто очікувати, що значення ознак, які відносяться до більш пізніх періодів часу, будуть у цілому вище, ніж значення ознак, що відносяться до ранніх періодів часу. У результаті зростуть максимальні значення нормованих ознак, тому що їхнє перетворення проводилося з використанням середньої арифметичної і стандартного відхилення, що відносяться до попереднього періоду. Таким чином, найбільше значення P_0, прийняте для одного періоду, може виявитися недостатньо великим для іншого, і виникне необхідність у перерахуванні координат точки P_0, а отже, й у перерахуванні раніше обчислених значень таксономічного показника.

У даному модифікованому варіанті пропонується по-іншому розраховувати норму d. У попередньому випадку ця величина обчислювалася як сума середнього арифметичного і a стандартного відхилення. Тепер, d залежить від числа досліджуваних ознак. Це автором пояснюється в такий спосіб. Ознаки стандартизовані і майже всі їх значення лежать в інтервалі $(-a, +a)$, таким чином, реалізації багатовимірної ознаки знаходяться всередині багатовимірного куба, кожне ребро якого має довжину $2a$. Довжина діагоналі куба $d = b\sqrt{n}$ для $b = 2a$, що і дорівнює шуканій нормі. Однією з вершин куба є нижній полюс, тому відстань кожної точки (її координати представлені елементами рядка матриці даних) від нижнього полюса не може перевищувати довжину діагоналі даного куба. Тим самим виконується вимога, щоб значення таксономічного показника знаходились в інтервалі від 0 до 1.

Знаючи координати нижнього полюса і норму, модифікований таксономічний показник розраховується за формулою:

$$g_i = \frac{d_i}{d} ,$$

де $d_i = \left(\sum_{j=1}^{n} (z_{ij} - a)^2 \right)^{\frac{1}{2}}$ для $i = \overline{1,t}$;

$$d = b\sqrt{n} \ ; \quad b = |-2a|,$$

a – координата нижнього полюса;

n – число ознак;

t – число об’єктів.

Модифікований таксономічний показник інтерпретується таким чином: високі значення показника свідчать про високі значення ознак, що враховуються, а низькі значення – про низькі значення цих ознак.

Задача 5.1. Маючи однорідні, перевірені на усталеність групи об’єктів (промислових підприємств Харківського регіону), одержані за критерієм факторів їх розвитку, потрібно оцінити рівень розвитку факторів у кожній групі, реалізуючи два підходи у формуванні еталона: за критерієм МініМакс та еталона, що встановлюється емпірично (на основі нормативних, директивних розпоряджень та рішень регіональних управлінських органів).

Розв’язання задачі. Такі підходи у формуванні еталона мають різний економічний зміст і отримані результати обчислень інтерпретують різні порівнянності, що доповнюють одна одну та утворюють комплекс у дослідженні. У табл. 5.1 відображені значення узагальнюючих показників, побудованих відносно еталона МініМакс, що є основою порівняльного аналізу підприємств у групі, та можна прослідкувати їх динаміку змін.

Узагальнюючі показники оцінки факторів кожного об’єкта в групах не порівняльні між групами, а порівняльні тільки всередині кожної групи. Цей факт є результатом способу побудови еталонних значень відносно еталона МініМакс, його жорсткої прив’язки до вибіркових характеристик групи. Але в даній ситуації можна обчислити узагальнюючий показник у цілому в групі (за середнім геометричним) та порівняти рівні показників усіх факторів.

Таблиця 5.1

Значення таксономічних показників розвитку факторів за групами об’єктів

Групи, об’єкти		Узагальнюючі показники факторів										
		I_{F_1}	I_{F_2}	I_{F_3}	I_{F_4}	I_{F_5}	F_6	F_7	F_8	F_9	F_{10}	F_{11}
1	2	3	4	5	6	7	8	9	10	11	12	13
I	1	0,48	0,53	0,74	0,86	0,47	0,67	0,39	0,08	0,63	0,59	0,59
	13	0,41	0,21	0,58	0,53	0,31	0,49	0,59	0,15	0,34	0,38	0,42
	15	0,20	0,45	0,39	0,46	0,39	0,64	0,25	0,08	0,50	0,39	0,72
	16	0,26	0,26	0,23	0,51	0,20	0,28	0,34	0,15	0,27	0,16	0,41
	18	0,38	0,34	0,65	0,31	0,33	0,54	0,44	0,08	0,34	0,49	0,55
	20	0,27	0,42	0,54	0,58	0,30	0,31	0,52	0,13	0,54	0,37	0,36
	24	0,28	0,50	0,57	0,43	0,46	0,70	0,36	0,09	0,59	0,37	0,28
	26	0,23	0,33	0,45	0,47	0,21	0,62	0,25	0,08	0,65	0,40	0,39
		0,30	0,36	0,49	0,50	0,32	0,50	0,38	0,10	0,46	0,37	0,45

Закінчення табл. 5.1

1	2	3	4	5	6	7	8	9	10	11	12	13
II	2	0,26	0,39	0,54	0,60	0,21	0,64	0,24	0,22	0,67	0,48	0,47
	8	0,20	0,38	0,53	0,85	0,40	0,96	0,55	0,11	0,63	0,75	0,69
	10	0,48	0,38	0,60	0,46	0,36	0,61	0,37	0,10	0,63	0,58	0,95
	14	0,27	0,36	0,55	0,56	0,32	0,88	0,29	0,21	1,00	0,45	0,39
	19	0,31	0,32	0,58	0,36	0,30	0,54	0,37	0,16	0,64	0,56	0,80
	21	0,29	0,13	0,48	0,33	0,12	0,38	0,50	0,18	0,60	0,42	0,78
	22	0,41	0,24	0,63	0,73	0,27	0,43	0,23	0,13	0,66	0,19	0,77
	23	0,20	0,30	0,48	0,54	0,34	0,60	0,33	0,10	0,62	0,54	0,40
	25	0,33	0,48	0,12	0,69	0,24	0,62	0,31	0,14	0,24	0,61	0,64
		0,29	0,31	0,46	0,54	0,27	0,60	0,34	0,14	0,60	0,48	0,63
III	3	1,00	0,00	0,90	0,96	0,93	0,96	0,90	0,00	0,97	0,93	1,00
	27	0,97	0,00	0,88	0,93	0,90	0,93	0,93	0,00	1,00	0,90	0,97
		0,98	0,00	0,89	0,94	0,92	0,94	0,92	0,00	0,98	0,92	0,98
IV	4	0,51	0,59	0,63	0,50	0,50	0,98	0,73	0,25	0,73	0,53	0,84
	6	0,52	0,64	0,57	0,39	0,47	0,87	0,93	0,31	0,64	0,65	0,66
	12	0,64	0,50	0,73	0,41	0,59	0,77	0,82	0,31	0,83	0,68	0,75
		0,55	0,57	0,64	0,43	0,52	0,87	0,82	0,29	0,73	0,62	0,74
V	5	0,38	0,50	0,52	0,76	0,60	0,74	0,78	0,28	0,82	0,73	0,79
	11	0,45	0,56	0,49	0,94	0,77	0,84	0,88	0,34	0,68	0,59	0,84
	17	0,36	0,64	0,41	0,77	0,68	0,95	1,00	0,27	0,66	0,60	1,00
		0,40	0,57	0,47	0,82	0,68	0,84	0,88	0,30	0,72	0,64	0,87

У першій групі промислових підприємств найбільший рівень розвитку був 0,5 і тільки для трьох факторів: F_4, F_6, F_3, занадто низький рівень значення 0,1 узагальнюючого показника існує для фактора F_8.

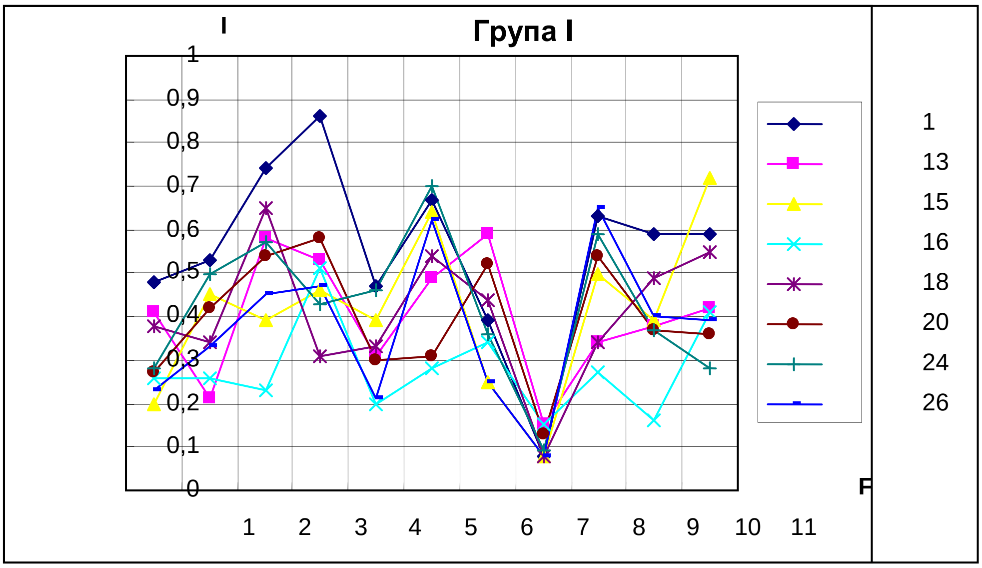
Сформувавши еталон з урахуванням рекомендацій експертів, нормативних документів та управлінських рішень, є змога виконати порівняльний аналіз між групами об’єктів відносно глобальної мети та стратегічних планів. У табл. 5.2 наведені значення узагальнюючих показників тих же факторів тих же груп об’єктів, але відносно еталона глобальних значень показників ознак факторів.

Таблиця 5.2

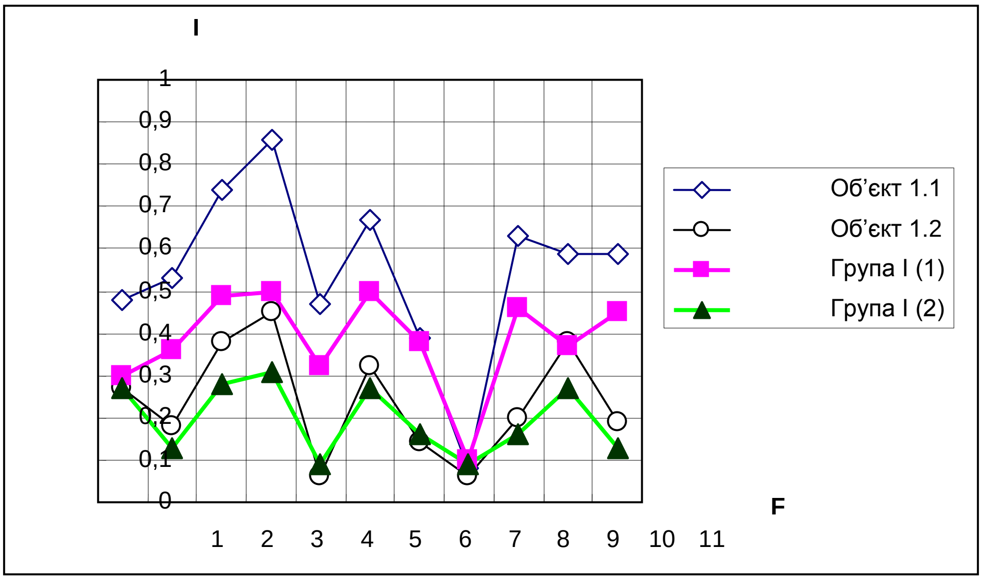
Значення таксономічних показників розвитку факторів за групами об'єктів, обчислених відносно еталона, що встановлюється

Групи, об'єкти		Узагальнюючі показники факторів										
		F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}
1		2	3	4	5	6	7	8	9	10	11	12
I	1	0,27	0,18	0,38	0,45	0,06	0,32	0,14	0,06	0,20	0,38	0,19
	13	0,31	0,07	0,35	0,31	0,12	0,23	0,19	0,14	0,11	0,37	0,12
	15	0,12	0,13	0,31	0,24	0,06	0,33	0,10	0,06	0,16	0,30	0,17
	16	0,33	0,18	0,10	0,46	0,12	0,18	0,23	0,11	0,12	0,10	0,10
	18	0,41	0,12	0,38	0,20	0,09	0,28	0,16	0,06	0,11	0,30	0,19
	20	0,30	0,16	0,31	0,36	0,13	0,19	0,23	0,09	0,16	0,27	0,11
	24	0,23	0,17	0,31	0,23	0,06	0,45	0,14	0,09	0,20	0,27	0,09
	26	0,28	0,10	0,25	0,32	0,09	0,31	0,12	0,10	0,25	0,27	0,10
		0,27	0,13	0,28	0,31	0,09	0,27	0,16	0,09	0,16	0,27	0,13
II	2	0,55	0,18	0,11	0,78	0,08	0,58	0,11	0,08	0,22	0,22	0,09
	8	0,16	0,11	0,09	0,65	0,16	0,86	0,10	0,15	0,20	0,26	0,13
	10	0,35	0,15	0,10	0,50	0,10	0,56	0,12	0,16	0,20	0,26	0,18
	14	1,00	0,16	0,08	0,77	0,14	0,79	0,19	0,08	0,35	0,24	0,07
	19	0,50	0,16	0,10	0,40	0,11	0,48	0,14	0,09	0,21	0,27	0,16
	21	0,36	0,10	0,07	0,24	0,06	0,32	0,22	0,08	0,19	0,19	0,18
	22	0,35	0,07	0,14	0,47	0,08	0,36	0,11	0,12	0,22	0,05	0,14
	23	0,43	0,13	0,07	0,65	0,10	0,53	0,13	0,16	0,20	0,29	0,08
	25	0,43	0,22	0,17	0,53	0,09	0,54	0,10	0,10	0,09	0,28	0,12
		0,42	0,13	0,10	0,52	0,10	0,53	0,13	0,11	0,20	0,21	0,12
III	3	0,15	0,56	0,06	0,07	0,37	0,88	0,04	0,25	0,75	0,30	0,34
	27	0,15	0,57	0,06	0,06	0,36	0,86	0,04	0,24	0,78	0,29	0,33
		0,15	0,56	0,06	0,06	0,37	0,87	0,04	0,24	0,77	0,29	0,34
IV	4	0,67	0,43	0,33	0,95	0,34	0,76	0,10	0,14	0,01	0,50	0,13
	6	0,83	0,47	0,31	0,86	0,39	0,66	0,12	0,14	0,01	0,60	0,12
	12	0,69	0,37	0,39	0,74	0,30	0,60	0,10	0,11	0,02	0,65	0,15
		0,72	0,42	0,34	0,85	0,34	0,67	0,11	0,13	0,01	0,58	0,13
V	5	0,18	0,23	0,57	0,42	0,04	0,33	0,27	0,10	0,12	0,45	0,37
	11	0,15	0,21	0,48	0,45	0,05	0,38	0,33	0,09	0,13	0,40	0,37
	17	0,15	0,26	0,45	0,35	0,05	0,43	0,34	0,11	0,10	0,51	0,45
		0,16	0,23	0,50	0,41	0,05	0,38	0,31	0,10	0,12	0,45	0,40

Порівнюючи дані табл. 5.1 та 5.2, можна спостерігати існуюче співвідношення рівнів факторів. Для наочності співвідношення значення узагальнюючих показників складних ознак об'єктів першої групи сукупності, а також співвідношення рівнів узагальнюючих показників, обчислених за еталоном, побудованого двома способами на одних і тих же об’єктах, слід зобразити графічно на рис. 5.2а, б. Власне можна провести різний аналіз залежно від зрізу на площині XOY: об’єкти – складні ознаки в групі, фактори – об’єкти у групі, об’єкти – фактори за групами, фактори-об’єкти за групами, а також можливо проаналізувати динаміку величини складної ознаки на об’єкті, динаміку величини складної ознаки об’єктів за групами, динаміку величини складної ознаки всіх об’єктів сукупності в цілому.



а



б

Рис. 5.2. **Значення узагальнюючих таксономічних показників розвитку складних ознак об’єктів першої групи (а); значення узагальнюючих показників, обчислених за еталонами, побудованими двома способами (б)**

На рис. 5.5а спостерігається явна синхронність зміни значень величин складних ознак на об’єктах, яка свідчить про існування розвинутіших ознак порівняно з іншими та розвинутіших об’єктів у групі за виділеними складними ознаками. Рис. 5.2б демонструє відносність рівнів порівняно з еталоном: якщо еталонні значення формувалися за критерієм МініМакс, то завжди буде отримано завищені величини узагальнюючого показника, порівняння відносне тільки для даної сукупності і говорити про інші немає об’єктивних підстав; якщо ж еталонні значення формуються теоретично на основі наявних умов, нормативів і т. д., то такі величини порівняні навіть тоді, коли в обчисленнях використовуються статистичні характеристики окремих сукупностей.

Даний метод широко використовується економістами у визначенні оцінок складних явищ та процесів, але він має значні вади з точки зору вимірювання величини. Коли еталонні значення окремих показників формуються за критерієм МініМакс, то завжди буде отримане завищене або занижене значення величини узагальнюючого показника відносно загальноприйнятих еталонів, оскільки порівняння відбувається у межах даної сукупності і тому особливості окремої сукупності об’єктів відображаються у величині узагальнюючого показника. Якщо ж еталонні значення формуються теоретично з урахуванням наявних умов, нормативів і т. д., то згідно з обчислювальною схемою даного математичного методу все одно залишається прив’язка до числових характеристик сукупності об’єктів. Це є великим недоліком даного методу.

Другу групу методів побудови узагальнюючого показника утворюють методи, які передбачають перетворення значень показника за допомогою уніфікованої шкали, що дозволяє вловлювати навіть найнезначніші зміни значень ознаки. Ключовим моментом у процесі перетворення є рівень кваліфікації, досвід експерта, чиєю авторитетною думкою керується аналітик у процесі побудові узагальнюючого показника. Але, щоб цей досвід розумно використати в рамках формальних процедур, його теж необхідно формалізувати. Найбільш природний шлях такої формалізації – введення системи переваг значень величини окремого показника на всій множині значень кожного окремого показника в економіці, одержання стандартної шкали, задавання основних оцінок на ній, а потім згортки перетворених величин. До цієї ж групи належить метод побудови узагальнюючого показника якості Харрінгтона [329].

Наведені дві групи методів мають принципову відмінність, але аналіз наукових робіт, де міститься опис методів побудови узагальнюючих показників та їх застосувань у різних сферах діяльності людини, дозволяє виділити спільні етапи у побудові: вирішення проблеми визначення елементарних ознак СЕС; вирішення проблеми визначення показників ознак; вирішення проблеми метричного нормування; вирішення проблеми вибору правил поєднання – згортки безрозмірних величин [1; 2; 45; 87; 213; 216; 309].

Згідно з положеннями математичного моделювання опису соціально-економічних систем, визначати величини складних ознак потрібно з мінімумом похибок, тобто висуваються жорсткі вимоги у моделюванні вимірників. Отже, з різних причин жоден з уже існуючих методів не забезпечує такої вимоги. Але перед тим як обговорювати питання побудови моделей вимірників, слід згадати питання допустимих перетворень величин в економіці, які аксіоматизовані в теорії вимірювання [91; 209; 227; 230; 231; 295]. Основне призначення аксіоматичних основ теорії вимірювання − довести на основі множини аксіом, які постулюються відносно деякої емпіричної структури, дві теореми – теорему задавання і теорему єдності, що таким чином показують існування ізоморфного і гомоморфного відображення емпіричної реляційної системи в числову. Дані аксіоматичні основи дозволяють фундаментально спиратися у розробках методологічних основ вимірювання, зокрема в економіці. Перша теорема дозволяє відповісти на запитання, чи існує числова система з відношеннями $\mathscr{H}$, в яку можна гомоморфно відобразити емпіричну систему з відношеннями O. Друга теорема описує всі гомоморфні відображення емпіричної системи в числову.

І. Пфанцагль збагатив аксіоматику вимірювань аксіомами та теоремами, пов’язаними з вимірюваннями некількісних ознак [230]. Складною є проблема одночасного вимірювання декількох властивостей. У своїх теоремах І. Пфанцагль запропонував обґрунтування правомірності таких дій у вимірюванні. Систематизуючи практичні рекомендації вимірювання неметричних ознак, що не передбачають безпосередньої числової оцінки, та теоретично обґрунтовуючи єдиний підхід, що базується на концепції універсальної шкали вимірювання якості і на методі арифметизації ординальних шкал, російський вчений Н. В. Хованов [295] теж розширив та збагатив аксіоматичні основи теорії вимірювання. Питання про існування універсальної шкали вимірювання загальної якості об’єкта вчений розглядає як питання про існування універсального подання довільної реляційної системи (системи з відношеннями). Вирішення питання здійснюється за допомогою сукупності теорем, де доводиться, що довільна реляційна система ізоморфна системі підмножин упорядкованої множини, які зв’язані відношеннями "ланцюгового мажорування". Отриманий результат є максимально загальним серед теорем про подання різних окремих реляційних систем. На думку вченого, наявність такої універсальної шкали вимірювання якості дозволяє редукціювати вимірювання якості на будь-якій шкалі до сукупності вимірювань на ординальних шкалах, що демонструє фундаментальність останніх. Розглядається також проблема арифметизації ординальних шкал. Вирішення проблеми ґрунтується на додатковій інформації про зв’язок арифметизованих градацій з числовими значеннями якості, що вимірюється. Для градації величин якості визначені комутативна, асоціативна операція додавання і віднімання, множення і ділення; операції додавання і множення зв’язані законом дистрибутивності; визначені рефлексивне, антисиметричне, транзитивне відношення нерівності "$\geq$", яке зв’язане з відношенням ">" строгого порядку та з операціями додавання і множення, а також для множини градацій величини якості виконується постулат Архімеда − Евдокса та принцип неперервності Кантора. У всіх теоремах про задавання мова йде про можливості встановлення ізоморфізму між реляційною системою, що взята з певного обмеженого класу реляційних систем, і сукупністю конкретних математичних об’єктів, що виступають в якості "подання" абстрактної системи, яка вивчається. Тому Н. В. Хованов висуває і доводить твердження про існування універсального подання, яке допускає встановлення ізоморфізму з довільною реляційною системою. На думку вченого, отримана універсальна шкала (універсальна форма шкали) вимірювання якості дозволяє вказати і загальну основу для процедур вимірювання якості, яка б виділяла окремі (маргінальні) якості із загальної якості, що вимірюється, та лінійно упорядковувала їх градації за ступенем прояву відповідної маргінальної якості. Вважається, що введення відношення ланцюгового мажорування та безпосереднього ланцюгового мажорування може не мати жодної наперед заданої властивості на зразок рефлективності, симетричності, транзитивності тощо, але здатне отримати всі ці властивості за рахунок перебудови упорядкованої множини, підмножини якої утворюють їх носій. У цьому розумінні вказане відношення є універсальним відношенням. Отже, загальний огляд робіт, що містять аксіоматичні основи теорії вимірювання, підтверджує обґрунтованість сумісного визначення різних величин ознак.

5.2. Вимірники метричних ознак в аналізі соціально-економічних систем

Визначати метричні й неметричні ознаки соціально-економічної системи і визначати її якість в цілому – послідовні процедури. Як уже було сказано раніше, визначення величин ознак спрямоване на визначення окремих якостей соціально-економічної системи. Згідно з основами теорії систем, є дві основні характеристики, що визначають систему, – це її якість і ефективність її функціонування [8, с. 333]. Обидві характеристики обумовлюються елементарними ознаками системи. У роботі [8] наведений взаємозв'язок та відмінності понять якості та ефективності систем, що подано в табл. 5.3.

Таблиця 5.3

Співвідношення понять якості та ефективності систем

Параметр	Якість	Ефективність
Визначення поняття	Властивість або сукупність істотних властивостей системи, що обумовлюють її придатність (відповідність) для використання за призначенням	Комплексна операційна властивість (якість) процесу функціонування системи, що характеризує його пристосування до досягнення мети операції (виконання задачі системи)
Область застосування	Об'єкти будь-якої природи, у тому числі елементи систем	Тільки цілеспрямовані операції, що проводяться системою
Основна характеристика	Сукупність атрибутивних властивостей системи, істотних для її використання за призначенням	Ступінь відповідності результатів операції її мети
Фактор структурного аналізу	Побудова системи (склад та властивості складових частин, структура, організація)	Алгоритм функціонування, якість системи, що реалізує алгоритм, дія зовнішнього середовища
Вимірність	Показник якості – вектор показників істотних властивостей	Показники результативності, ресурсоємності і оперативності щодо закінчення операції і якості "алгоритму", що забезпечує отримання результатів
Спосіб оцінювання	Критерії придатності, оптимальності, перевершення	Критерії придатності або оптимальності, що визначаються залежно від типу операції, що проводиться (детермінована, імовірнісна або невизначена)

Отже, кожна окрема якість відображається в окремій ознаці системи, величина якої характеризує міру (інтенсивність) даної якості, формою величини є окремий показник системи. Узагальнена величина якості ознак системи подається згорткою величин складних ознак, як це було сказано в підрозділі 4.1 посібника. Згортку можна виконати різними методами.

Аналіз робіт науковців, що використовували різні типи показників у своїх дослідженнях, дозволив сформулювати п’ять основних проблем у моделюванні вимірників. Вони наведені на схемі рис. 5.3 [1; 2; 57; 130; 131; 133; 139; 154].

1. Проблема узагальнення теоретичних та практичних знань щодо даної ознаки системи, проведення описової статистики ознак

2. Проблема побудови шкали перетворень значень показників ознак СЕС

3. Проблема визначення основних точок фазових змін значень показників ознак СЕС

4. Проблема визначення окремих функцій перетворень

5. Проблема визначення узагальненої функції перетворення

Рис. 5.3. **Логіка моделювання вимірників**

Розглянемо пропозиції вирішення кожної проблеми окремо.

Перша проблема – це проблема узагальнення теоретичних та практичних знань за даною ознакою, проведення статистичного опису змін значень показника ознаки, що демонструє наявний розвиток ознаки. Проблема вирішується на основі знань експертів, осіб, що ухвалюють рішення, як синтез науково-практичних досягнень у пізнанні даної ознаки та зіставленні з результатами використання інструментів описової статистики.

Другою проблемою є проблема побудови шкали перетворень. За допомогою цієї шкали відбувається перетворення значень окремих економічних показників, що мають свої одиниці вимірювання, у безрозмірну шкалу. Вирішення другої проблеми обумовлене результатами вирішення першої проблеми та є етапом продовження вимірювання величин в економіці, оскільки завдяки перетворенню нові величини можна порівнювати між собою. Від того, наскільки об'єктивно буде розроблена шкала перетворення залежить вся процедура вторинного вимірювання в економіці.

На сьогодні існують достатні обґрунтування таблиць відповідностей між відношеннями переваги в емпіричній і числовій системах. Завжди до побудови шкали перетворень рекомендується підходити ітераційно. Попередньою стадією обчислень є параметричний аналіз показників ознак.

Незважаючи на деякі розбіжності, майже всі типи шкал мають спільність: значення $y_j = 0$ відповідає абсолютно неприйнятному рівню даної ознаки, а значення $y_j = 1$ – найкращому значенню ознаки. Поняттям "дуже добре", "надзвичайно добре" чи "чудово" відповідають значення на шкалі перетворення $0{,}8 < y_j < 1{,}00$, а поняттям "дуже погано", "неприпустимо", "неприйнятно" – $0{,}00 < y_j < 0{,}20$. За методикою Е. Харрінгтона є ще два характерних значення: $0{,}63 \approx 1 - \frac{1}{e}$, $0{,}37 \approx \frac{1}{e}$. Функція перетворення за методикою Е. Харрінгтона має вигляд $y = e^{-e^{-x}}$ (рис. 5.4) або $y = exp(-exp(-x))$, де $exp-$ прийняте позначення експоненти.

Третя проблема – це проблема визначення основних точок фазових змін значень показників та, по суті, є підпроблемою другої проблеми моделювання вимірників, але її надзвичайна важливість обумовила виділити її окремо. Багато відносних показників в економіці мають значення від 0 до 1. Але багато в економіці й абсолютних показників, межі яких невизначені. Виникає принципове запитання, на якій підставі встановлюються межі допустимих значень для таких показників ознак. Спочатку слід дослідити тенденцію до змін величини ознаки. З огляду на економічну сутність показників ознак (за тенденціями до зміни стосовно еталонних значень) розрізняють односторонні обмеження і двосторонні.

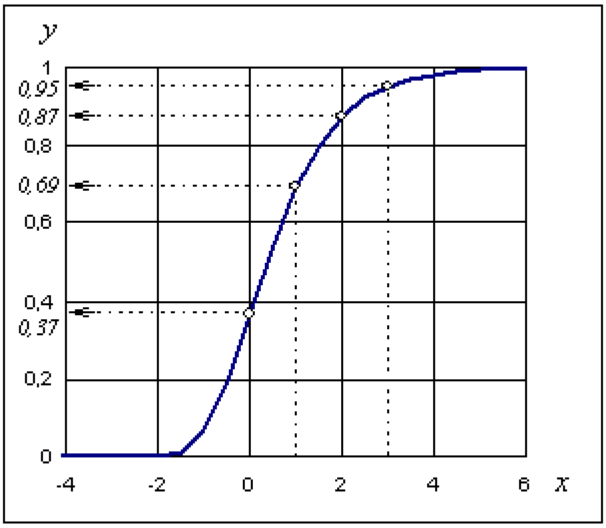


Рис. 5.4. **Графік функції перетворення в загальному вигляді**

Односторонні обмеження подаються у вигляді $x_j \le x_{\max}$ чи $x_j \ge x_{\min}$, а двосторонні – $x_{\min} \le x_j \le x_{\max}$. Тут можливі дві ситуації. Перша – найсприятливіша, коли існує інформація у вигляді розпоряджень, інструкцій, нормативних документів про еталонні значення, в яких чітко сформульовані вимоги до всіх фазових значень показників ознак, тобто існує специфікація з одним чи двома обмежувальними значеннями. Тоді оцінка на шкалі перетворення $d_j = 0$, наприклад, відповідає $x_{\min}$, якщо є однобічне обмеження $x_j \ge x_{\min}$; обмеження $d_j = 1{,}00$ відповідає $x_{\max}$ для обмеження $x_j \le x_{\max}$. Слід зазначити, що для випадку двостороннього обмеження ці міркування непридатні. У другій ситуації специфікація відсутня, тоді обмеження на шкалі встановлюються на основі науково-практичного знання про дану ознаку та з посиланнями на результати описової статистики. У даній ситуації доцільно використовувати статистичні залежності та взаємозв'язки ознак об'єктів в економіці (про це детально буде йти мова під час вирішення наступної проблеми).

Четвертою проблемою є визначення виду окремої функції перетворення кожної ознаки системи.

Перетворення величин ознак у їхні стандартні аналоги на шкалі здійснюється за складними законами. Складність перетворення полягає в установленні адекватної відповідності значень таблиці перетворення і значень функції перетворення, яку необхідно підібрати, виходячи із суті зміни ознаки та користуючись шкалою відношень.

Значення функції перетворення мають демонструвати рівень розвитку ознак у системі, тобто дана функція є моделлю розвитку ознаки. Моделі розвитку на сьогодні найширше використовуються у фізиці, хімії, біології й економіці [178; 189; 192; 205; 285]. У процесі розвитку відбувається певна стандартизація, уніфікація перетворень структури і функцій системи, тобто для розвитку характерний ізоморфізм. З урахуванням ізоморфізму розвитку моделі розвитку мають досить багато загальних рис, проте зберігають специфічні особливості, властиві кожній із наук [109; 116; 122]. Методи побудови математичних моделей у різних науках мають особливості, що базуються на тих законах, які вивчаються кожною з них.

Раймонд Перл (американський біолог, демограф і економіст) для розвитку популяції запропонував логістичну функцію, що описується рівнянням виду [109]:

$$Y(t) = \frac{Y_0}{1 + ae^{-bt}}, \tag{5.7}$$

де $Y(t)$ – чисельність в одиниці обсягу популяції в момент часу t;

Y_0 – початкова чисельність популяції;

a, b – константи.

На графіку (рис. 5.5) видно, що логістична крива починається в точці $\left(\frac{Y_0}{1+a}\right)$, вона має точку згину з координатами $t_k = \frac{ln\,a}{b}$; $y_k = \frac{Y_0}{2}$.

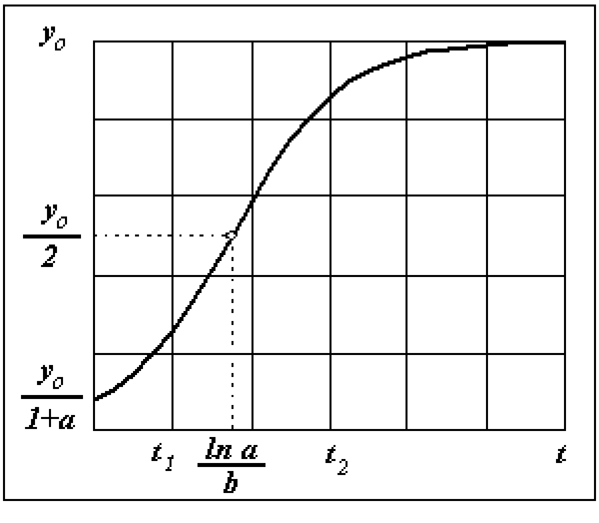


Рис. 5.5. **Графік логістичної функції**

Константа a визначає положення логістичної кривої в часі (зрушення ліворуч, праворуч), константа b – нахил кривої. Ці константи обчислюються за формулами:

$$a = \frac{Y_0}{Y_{t=0}} - 1; \qquad b = \frac{(1+a)^2}{a y_0}\left(\frac{dy}{dt}\right)_{t=0}.$$

При відомому часі подвоєння біологічних популяцій $t_{\frac{1}{2}}$ константу b можна визначити за наближеною формулою $b \cong t_{\frac{1}{2}} ln\frac{2a}{a-1}$. Логістична функція має одну чудову властивість: на ній можна чітко виділити три основних періоди в розвитку системи:

1) період $(0 \div t_1)$ – початок розвитку, "молодість" (прогресивна ділянка);

2) період $(t_1 \div t_2)$ – інтенсивний розвиток, "зрілість" (рівномірна ділянка);

3) період $(t_2 \div \infty)$ – екстенсивний розвиток, "старість" (регресивна ділянка).

Знаючи, на якій із ділянок розвитку знаходиться система у момент спостереження, можна використовувати спрощені, наближені формули, що описують її розвиток. Так, за початкових низьких значень t розвиток може бути описаний експоненціальною функцією $y(t) \cong \frac{Y_0}{a} e^{bt}$, при високих значеннях t у момент "старості" $y(t) \cong Y_0$, тобто система асимптотично наближається до максимуму. Отриману експериментально логістичну функцію зростання можна використовувати для розкриття внутрішньої структури системи та з'ясування механізму її розвитку [180]. Такий якісний перехід до вдосконалення аналізу об'єкта може бути здійснений шляхом переходу до диференціального опису системи (феноменології). Знаючи рівняння логістичної функції, можна скласти диференціальне рівняння, що породжує цю функцію як рішення.

У випадку односторонніх $x_j \le x_{\max}$ чи обмежень $x_j \ge x_{\min}$ на рис. 5.6 подана окрема функція перетворення ознаки, що визначає властивість системи і має обмеження з однієї сторони.

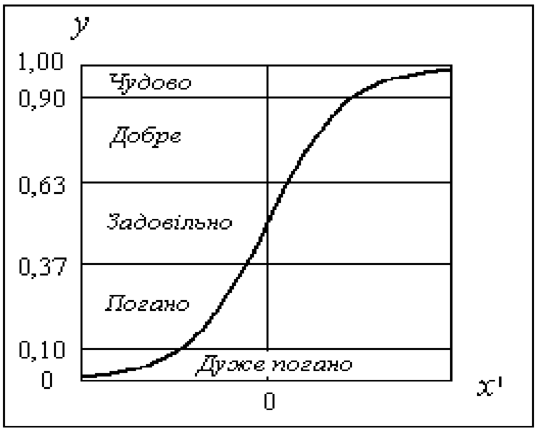


Рис. 5.6. **Функція перетворення для ознаки, допустимі значення якої обмежені з однієї сторони**

До ознак, що підпорядковані одностороннім обмеженням, належить більшість ознак соціально-економічних систем, хоча існують ознаки з двосторонніми обмеженнями, вони, як правило, виражаються відносними показниками.

Рідше в економіці зустрічаються показники ознаки, що характеризують властивості соціально-економічних систем із двосторонніми обмеженнями. Для двостороннього обмеження $x_{\min} \le x_j \le x_{\max}$ графік функції перетворення наведений на рис. 5.7.

Випадок двостороннього обмеження, як уже було сказано, зустрічається не так часто, як одностороннього, і він складніший у побудові функції перетворення. У підрозділі 3.2 посібника в табл. 3.6 наведені результати виконаного статистичного аналізу ознак підприємств. Для визначення тенденції до змін значень показників рекомендується застосовувати інструменти описової статистики. У табл. 3.6 містяться показники, які мають односторонню та двосторонню обмеженість у змінах своїх значень. Побудова функції перетворення детально розглядається в роботах [57; 153].

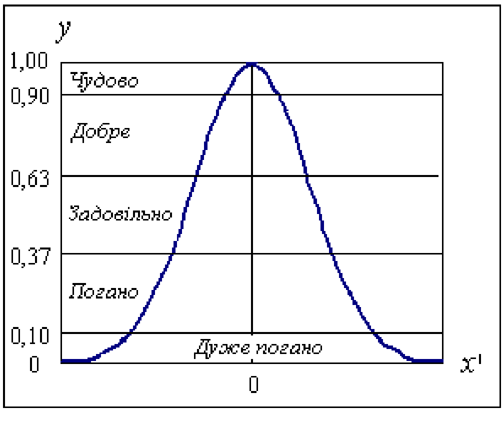


Рис. 5.7. **Функція перетворення для ознаки, допустимі значення якої обмежені з двох сторін**

Методику побудови узагальнюючого показника Е. Харрінгтона успішно використовував Ф. С. Новік для комплексної оцінки якості сталі та сплавів [203; 204] і Ю. П. Адлер для оцінки якості хімічних складів речовин [1].

Наведені форми функції перетворення не є єдино можливими. Залежно від постановки задачі можна вибирати й інші види функції перетворення. У роботі [203] для вибору сплаву, який має одночасно з високою твердістю і достатню за технічним завданням електропровідність, використовували функцію $ln\,y_i = x_i'$.

Зазначимо, що всім методам побудови узагальнюючого показника якості властивий ще один загальний недолік − використання одного виду залежності для всіх показників.

У роботі [213], присвяченій комплексній оцінці якості ґрунтів, запропоновані більш прості й досить гнучкі функції перетворення:

– симетричні двосторонні:

$$Y_i = exp\left\{-k \cdot \left(\frac{x_i - a_i}{b_i - a_i}\right)^2\right\}; \tag{5.8}$$

– односторонні:

$$Y_i = \frac{1}{1 + exp\left\{-k \cdot \left(\frac{x_i - c_i}{a_i - c_i}\right)\right\}}. \tag{5.9}$$

Тут a_i – оптимальне значення ознаки x_i, при якому двостороння функція перетворення дорівнює 1 (100%), а однобічна – не менше ніж 0,95; b_i – значення ознаки, при якому якість низька, менше ніж 0,05 (5%); c_i – рівень, при якому отримується 50% якості (0,5). Параметр k керує формою кривої. Слід зазначити, що відповідно до оцінок економістів найкраще значення цього параметра практично у всіх розглянутих нижче економічних задачах дорівнює $k = 3$ як для односторонніх, так і для двосторонніх залежностей.

На основі узагальнення результатів проведення модельних експериментів побудови функцій перетворень для розв’язання реальних задач в економіці, автором одержані найтиповіші функції перетворення ознак соціально-економічних систем, які мають одну визначальну характеристику – гнучкість [57; 153; 155].

Для двосторонніх несиметричних тенденцій розвитку ознаки системи мають бути наступні функції перетворення:

$$y_{ij} = \begin{cases} 100 \cdot e^{-3\left(\frac{x_{ij}-a_i}{b_i-a_i}\right)^2}, \text{для } x_{ij} \le a_i,\ b_i < a_i, \\ 100 \cdot e^{-3\left(\frac{x_{ij}-a_i}{c_i-a_i}\right)^2}, \text{для } x_{ij} \ge a_i,\ c_i > a_i, \end{cases} \quad (5.10)$$

де a_i, b_i, c_i – реперні значення: a_i – найкраще значення показника x_{ij}, за якого функція перетворення досягає найбільшого значення 1 (100%);

$b_i,\ c_i\ (b_i < c_i)$ – незадовільне значення показника x_{ij} (по різні сторони від найкращого), за якого функція перетворення набуває значення, не більшого ніж 0,05 (5%).

За умови симетричних тенденцій розвитку ознак функція перетворення набуває значення 1 (100%) при $a_i = \frac{b_i + c_i}{2}$. Вигляд функції спрощується:

$$y_{ij} = 100 \cdot e^{-3\left(\frac{x_{ij}-a_i}{b_i-a_i}\right)^2}, \quad (5.11)$$

або (це еквівалентно):

$$y_{ij} = 100 \cdot e^{-3\left(\frac{x_{ij}-a_i}{c_i-a_i}\right)^2}. \quad (5.12)$$

Для односторонніх типів розвитку ознак соціально-економічних систем рекомендуємо монотонні функції перетворення типу логістичної функції:

$$y_{ij} = \frac{100}{1 + e^{-\frac{x_{ij}-p_i}{q_i-p_i}}}, \quad (5.13)$$

де q_i – значення показника x_{ij}, при якому функція перетворення набуває значення, не меншого ніж 0,95 (95%);

p_i– значення показника x_{ij}, при якому функція перетворення набуває значення 0,5 (50%).

Зауваження. Наведені функції, як уже було сказано, пройшли багаторічну перевірку у розв’язанні реальних задач на промислових підприємствах. Проте маємо зауваження щодо обґрунтування шкали перетворень. Проблема пов'язана зі специфікою величин в економіці, а конкретніше, з їх формою – показниками. Кількісні показники в економіці можуть бути відносними або абсолютними. Якщо соціально-економічна система характеризується ознаками, величини яких виражаються відносними показниками, то побудова вимірників здійснюється чітко за схемою, що на рис. 5.3. Виходячи із суті тенденцій змін величини ознаки, конкретна функція перетворення відносного показника може бути односторонньою (зростаючою чи спадною) та двосторонньою (симетричною чи несиметричною). Очевидно, для абсолютних показників теж існують такі типи функції, але їх потрібно конкретизувати. Складна ознака об'єкта синтезується елементарними ознаками та виражається системою окремих показників. Можливі три випадки в системі метричних ознак, що характеризують об'єкт. Перший випадок: система метричних ознак виражається тільки відносними показниками; другий випадок – відносними й абсолютними показниками; третій випадок – тільки абсолютними показниками. Таким чином, виникають три задачі в моделюванні вимірників складних ознак СЕС, які випливають із структури груп показників, що їх виражають. У табл. 5.4 наведений зміст кожної із задач.

Таблиця 5.4

Типи задач побудови окремих функцій перетворень

Умови задачі	Параметри задачі – репери величин	Спосіб розв'язання
Задача 1. Складна ознака синтезується елементарними ознаками, вираженими тільки відносними показниками	Найчастіше визначені за існуючими парадигмами в економіці	Калібрування елементарних метричних ознак за допомогою функцій перетворень
Задача 2. Складна ознака синтезується елементарними ознаками, вираженими системою відносних і абсолютних показників	Визначені для відносних показників і погано визначені для абсолютних показників	На основі гіпотези, що в межах системи показників існує однакова усталена статистична характеристика, завдяки якій і будуються функції перетворень абсолютних показників
Задача 3. Складна ознака синтезується елементарними ознаками, вираженими тільки абсолютними показниками	Невідомі або погано визначені для абсолютних показників	На основі гіпотези ототожнення функції перетворення з інтегральною функцією розподілу абсолютного показника

Перша задача моделювання вимірників ознак системи, що виражаються відносними показниками, розв'язується простіше, оскільки існують рекомендації щодо визначення реперних точок змін значень показників. Якщо до складу групи показників, які описують ознаку, входять крім відносних і абсолютні показники, то ця (друга) задача складніша від попередньої. Третя задача найскладніша, оскільки відсутні будь-які обґрунтовані міркування щодо визначення реперних точок шкали перетворень абсолютних показників.

Розглянемо другу задачу. Системи показників, що виражають складну ознаку, виявлені на основі математичних методів (факторного або канонічного аналізів) чи на основі теоретико-логічного аналізу. Показники в системі суттєво однорідні і, як правило, тісно корелюють між собою. Висувається гіпотеза, що в межах груп показників існує однакова усталена статистична характеристика. Раціонально було б зв'язати реперні точки значень показників, що вказали експерти в межах конкретної групи, з усталеними статистичними характеристиками показників ознак. Дослідження показали, що дійсно в межах кожної однорідної групи показників, які виражають ознаки, практично є сталою величина λ, що обчислюється за формулою:

$$\lambda = \frac{x_{f=0,95} - x_{f=0,05}}{x_{0,25} - x_{0,75}}, \qquad (5.14)$$

де $x_{f=0,95}$ – реперне значення показника, для якого функція перетворення досягає значення 0,95 (95%) на шкалі перетворення;

$x_{f=0,05}$ – реперне значення показника, для якого функція перетворення досягає значення 0,05 (5%) на шкалі перетворення;

$x_{0,25}$ – значення верхнього квартиля;

$x_{0,75}$ – значення нижнього квартиля;

$x_{0,25} - x_{0,75}$ – міжквартильний розмах (усталена (робастна) статистична характеристика).

Задача 5.2. У задачі аналізу соціально-економічного стану промислових підприємств Харківського регіону, що утворили 11 однорідних груп показників (приклад підрозділу 3.3), за допомогою факторного аналізу виділені складні ознаки їх стану. Визначити вимірники ознак соціально-економічного стану підприємств.

Розв'язання задачі. Серед цих груп розрізняються групи трьох типів, для кожного з яких розв'язується своя задача побудови окремих функцій перетворень (як це вказано в табл. 5.4). Величина λ була обчислена для всіх відносних показників у кожній з 11 однорідних груп показників. Ця величина за модулем майже однакова для всіх показників у групи. У тих випадках, коли для будь-якого показника значення λ відрізнялось у групі, то зміна значень даного показника ще раз досліджувалась і реперні точки ще раз обговорювалися з експертами. Зазвичай експерти були згодні з неточністю початкового задавання реперних точок.

Оскільки в групи входять і абсолютні показники, для яких відсутня експертна оцінка реперних значень, то величину λ рекомендується вважати сталою і для абсолютних показників. Отже, для абсолютних показників були прийняті реперні точки, значення яких обчислені за формулами:

$$x_{f=0,95} = M_e + \frac{\lambda}{2}(x_{0,25} - x_{0,75}); \quad (5.15)$$

$$x_{f=0,05} = M_e - \frac{\lambda}{2}(x_{0,25} - x_{0,75}). \quad (5.16)$$

У табл. 5.5 наведені обчислення величини λ для всіх показників 45 елементарних ознак підприємств, що досліджувалися. Згідно з даними табл. 5.5 спостерігається усталеність величини λ за групами показників. Це підтверджує гіпотезу, що величина λ стала як для елементарних, так і для складних метричних ознак в економіці. У табл. 5.5 наведені значення величини λ для однорідних груп показників, що виражають складні ознаки стану промислових підприємств.

До складу системи із 45 показників, що виражають елементарні ознаки стану даних підприємств за 11 складними ознаками, входять 14 абсолютних показників $(x_{27}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36}, x_{37}, x_{38}, x_{39}, x_{47})$ (вони виділені в табл. 5.5).

Спочатку вдалось ідентифікувати 9 показників, що увійшли до груп з відносними показниками (розв’язувалася друга задача моделювання вимірників). Для решти 5 показників $(x_{27}, x_{28}, x_{29}, x_{38}, x_{39})$ розв’язувалась третя задача, в умові якої відомо тільки те, що тип зміни значень абсолютних показників односторонній зростаючий.

Для розв’язання третьої задачі була прийнята гіпотеза ототожнення функції перетворення з інтегральною функцією розподілу показника. Точкам $x_{f=0,95}$ і $x_{f=0,05}$ інтегральної функції нормального закону відповідає стандартизоване значення $t = \pm 1{,}65$, а у міжквартильному розмаху міститься $1{,}35\sigma$. Звідси для таких показників λ обчислювалась так: $\lambda = \frac{t}{1{,}35} = \frac{1{,}65}{1{,}35} \approx 1{,}222$.

Таблиця 5.5

Значення λ за ознаками

Основні ознаки	Показники елементарних ознак, що визначають ознаку	Значення величини λ
O_1 – ознака забезпеченості підприємств оборотними засобами	$x_1, x_2, x_5, x_7, x_8, x_{13}, x_{14}, x_{24}, x_{25}, \boldsymbol{x_{33}}$	1,12
O_2 – ознака ефективності зростання всього капіталу підприємства	$x_6, \boldsymbol{x_{30}}, \boldsymbol{x_{31}}, \boldsymbol{x_{32}}, \boldsymbol{x_{36}}, x_{40}, x_{46}, \boldsymbol{x_{47}}$	1,9
O_3 – ознака достатності та прибутковості власного капіталу підприємств	$x_{16}, x_{17}, x_{18}, x_{22}, x_{45}$	3,0
O_4 – ознака покриття активів власними оборотними засобами	x_{19}, x_{20}, x_{21}	1,0
O_5 – ознака достатності найліквідніших елементів оборотних активів	$x_3, x_4, \boldsymbol{x_{34}}, \boldsymbol{x_{35}}$	2,15
O_6 – ознака прибутковості діяльності підприємств	x_{41}, x_{42}, x_{43}	0,4
O_7 – ознака технічного стану основних засобів	$x_{10}, \boldsymbol{x_{37}}, x_{48}$	1,25
O_8 – ознака своєчасності проведення розрахунків із дебіторами	$\boldsymbol{x_{38}}, \boldsymbol{x_{39}}$	1,22
O_9 – ознака здатності підприємств отримувати дохід	$\boldsymbol{x_{27}}, \boldsymbol{x_{28}}, \boldsymbol{x_{29}}$	1,22
O_{10} – ознака здатності підприємств генерувати прибуток від основної діяльності	x_9, x_{44}	1,2
O_{11} – ознака забезпеченості підприємств власними засобами	x_{26}	

Саме таким способом були остаточно побудовані функції перетворення показників $\{x_{38}, x_{39}\}$ в групі O_8 та $\{x_{27}, x_{28}, x_{29}\}$ у групі O_9. Ознаки O_8 і O_9, визначаються тільки абсолютними показниками. У разі появи додаткової інформації функції цих показників можуть уточнюватися.

Отже, для показників, у яких відсутня інформація про значення реперних точок (група показників O_8 та O_9), слід встановити значення λ за нормальною ймовірнісною функцією.

Для визначення відомих одинадцяти складних ознак за допомогою вимірників були обчислені параметри й побудовані окремі функції перетворення для кожної елементарної метричної ознаки (табл. 5.6 та рис. 5.8).

Таблиця 5.6

Параметри для моделювання вимірників

	Тип	$x_{f=0,05}$	$x_{f=0,5}$	$x_{f=0,95}$	Min	Max	M_e	$x_{0,25}$	$x_{0,75}$	λ
1	2	3	4	5	6	7	8	9	10	11
x_1	1	90,000	75,000	60,000	1,850	94,030	67,100	80,018	53,443	-1,129
x_2	1	10,000	25,000	40,000	5,970	98,100	32,900	46,670	19,988	1,124
x_3	1	90,000	70,000	50,000	10,301	98,534	2,750	9,425	60,945	-2,165
x_4	1	10,000	30,000	50,000	1,466	89,699	26,866	38,861	20,227	2,147
x_5	1	60,000	70,000	80,000	18,380	99,708	83,221	91,088	71,257	1,008
x_6	1	25,000	15,000	5,000	0,000	49,710	10,595	16,562	5,822	-1,862
x_7	1	10,000	5,500	1,000	0,000	80,156	3,769	10,095	1,472	-1,044
x_8	1	50,000	65,000	80,000	7,823	97,689	72,120	83,223	57,313	1,158
x_9	1	30,000	60,000	90,000	0,970	100,000	45,859	82,708	32,760	1,201
x_{10}	1	50,000	40,000	30,000	21,880	87,200	59,670	70,320	54,000	-1,225
x_{13}	2	0,000	1,500	3,000	0,155	40,286	1,582	2,835	0,740	
x_{14}	3	5,000	60,000	90,000	0,008	95,711	20,550	32,812	8,697	
x_{16}	3	0,200	0,700	0,900	0,023	0,976	0,735	0,866	0,638	
x_{17}	1	1,000	0,600	0,200	0,024	0,900	0,297	0,433	0,135	-2,693
x_{18}	1	0,300	0,650	1,000	0,096	0,976	0,762	0,873	0,679	3,611
x_{19}	1	0,200	0,400	0,600	-4,892	0,981	0,356	0,534	0,143	1,022
x_{20}	1	0,200	0,500	0,800	-12,880	7,906	0,400	0,743	0,147	1,007
x_{21}	1	0,000	0,400	0,800	-11,097	8,138	0,891	1,350	0,543	0,991

Закінчення табл. 5.6

1	2	3	4	5	6	7	8	9	10	11
x_{22}	3	0,000	0,300	1,000	-0,702	2,884	0,295	0,528	0,058	
x_{23}	2	-0,100	0,400	0,900	-0,914	0,710	0,002	0,041	0,000	
x_{24}	3	0,000	0,200	0,500	0,000	10,233	0,002	0,029	0,000	
x_{25}	2	0,200	0,600	1,000	0,012	10,440	0,551	1,382	0,168	
x_{26}	2	0,500	0,750	1,000	-1,479	2,740	0,234	0,599	-0,278	
x_{27}	1	-328,138	549,500	1427,18	35,000	13204,0	549,500	1634,20	195,500	1,220
x_{28}	1	-21116,9	23263,4	67643,8	194,200	2836880	23263,4	77705,2	4950,47	1,220
x_{29}	1	-5229,67	-92,450	5044,77	-539456,	190940,	-92,450	1267,72	-7153,95	1,220
x_{30}	1	-0,140	0,282	0,704	0,020	3,231	0,282	0,601	0,157	1,900
x_{31}	1	-0,090	0,451	0,993	0,030	3,560	0,451	0,798	0,227	1,900
x_{32}	1	-0,075	0,503	1,080	0,027	2,516	0,503	0,874	0,267	1,900
x_{33}	1	0,001	0,550	1,099	0,015	39,064	0,550	1,196	0,215	1,120
x_{34}	1	0,164	1,466	2,768	0,000	20,772	1,466	2,244	1,039	2,160
x_{35}	1	47,485	244,045	440,605	0,000	2216,00	244,045	338,500	156,500	2,160
x_{36}	1	-0,992	1,580	4,151	0,000	21,555	1,580	3,506	0,799	1,900
x_{37}	1	7,352	215,121	422,889	0,000	8378,18	215,121	430,921	101,130	1,260
x_{38}	1	1,242	6,714	12,186	1,188	35,781	6,714	12,817	3,846	1,220
x_{39}	1	11,767	52,132	92,497	1,594	302,917	52,132	93,600	27,428	1,220
x_{40}	1	0,000	5,000	10,000	-34,907	55,032	0,203	2,139	-2,938	1,970
x_{41}	1	0,000	5,000	10,000	-179,796	65,827	2,966	17,524	-7,260	0,403
x_{42}	1	0,000	5,000	10,000	-15,503	45,244	4,161	17,237	-2,181	0,515
x_{43}	1	0,000	5,000	10,000	-19,336	45,917	6,661	23,091	-4,154	0,367
x_{44}	1	-13,214	11,503	36,220	-86,332	173,209	11,503	34,937	-6,258	1,200
x_{45}	1	0,000	0,250	0,500	-6,470	0,500	-0,003	0,059	-0,092	3,307
x_{46}	1	0,000	12,000	24,000	0,100	176,458	12,649	19,994	7,716	1,955
x_{47}	1	-3,634	0,098	3,830	-16,626	52,287	0,098	2,473	-1,456	1,900
x_{48}	1	0,000	15,000	30,000	0,100	176,680	27,428	37,131	14,109	1,303

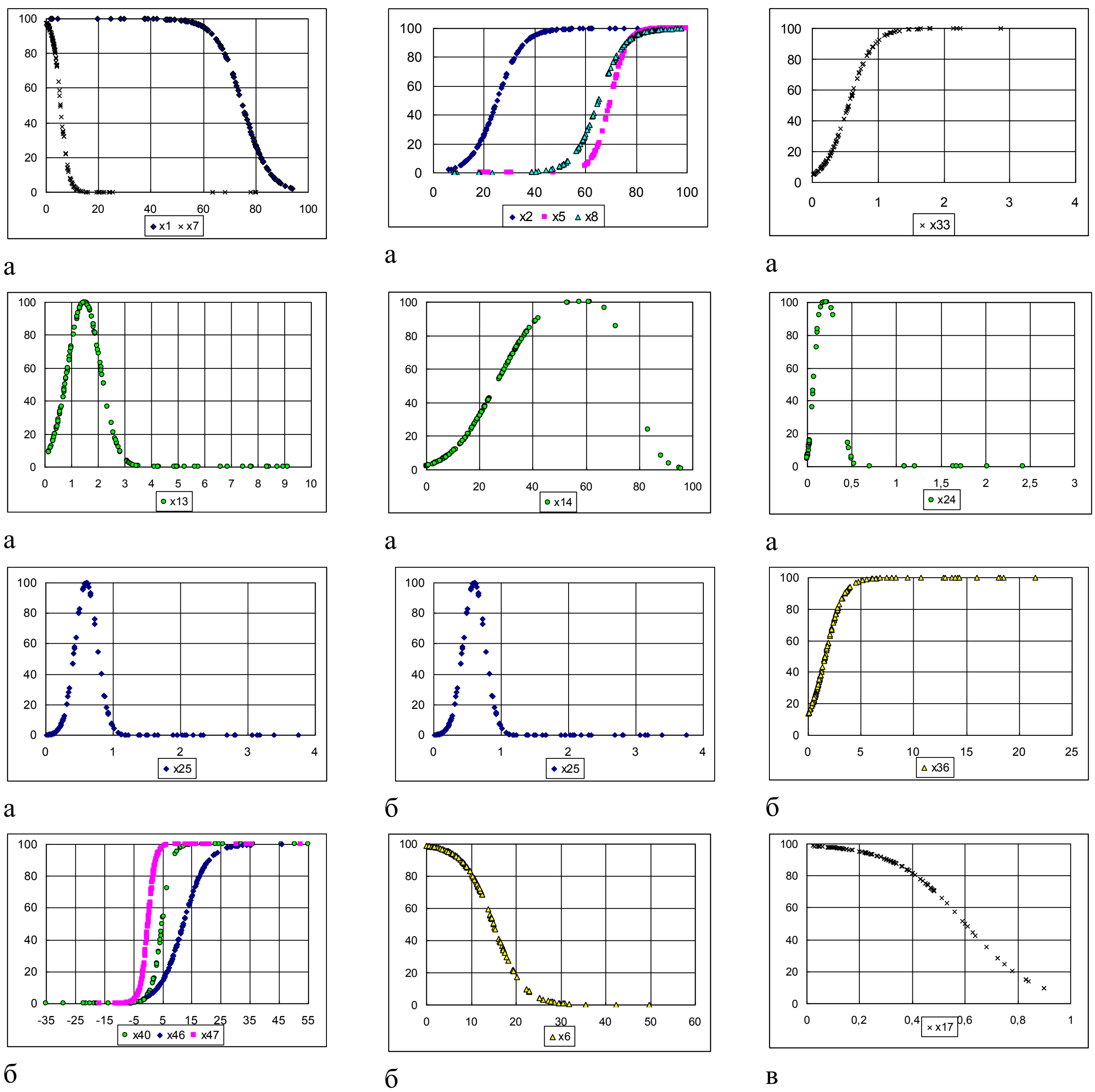
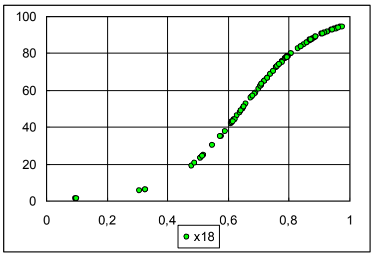
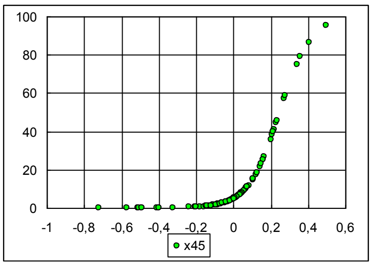


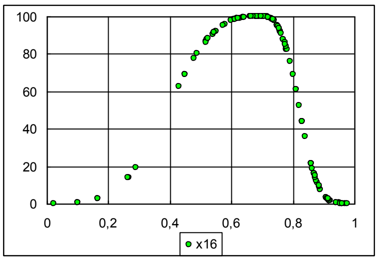
Рис. 5.8. **Окремі функції перетворення показників, що виражають складні метричні ознаки виробничо-господарської діяльності промислових підприємств: а – ознака O_1; б – ознака O_2; в – ознака O_3; г – ознака O_4; д – ознака O_5; е – ознака O_6; є – ознака O_7; ж – ознака O_8; з – ознака O_9; и – ознака O_{10}; і – ознака O_{11}**



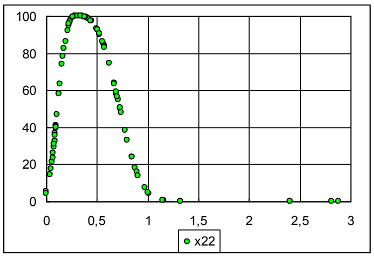
в



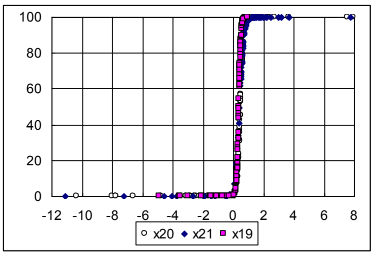
в



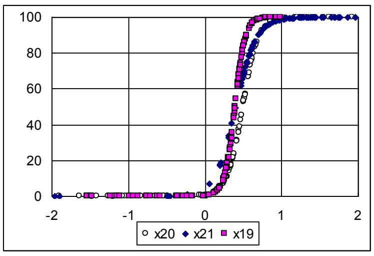
в



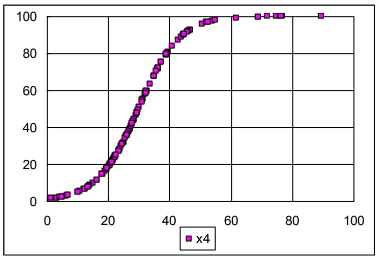
в



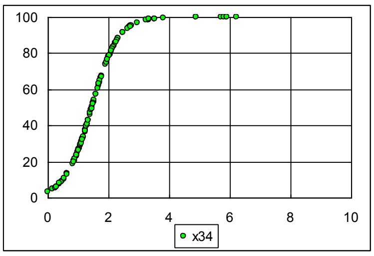
г



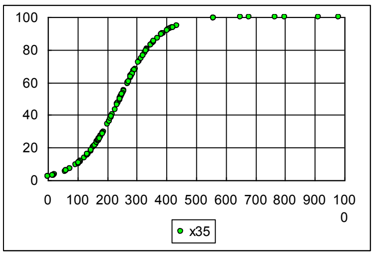
г



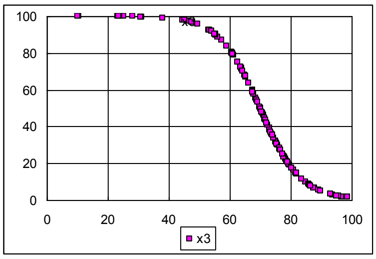
д



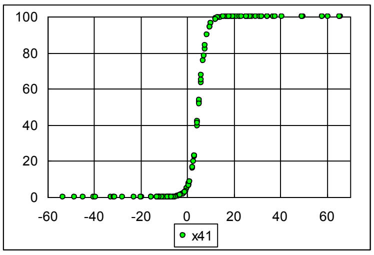
д



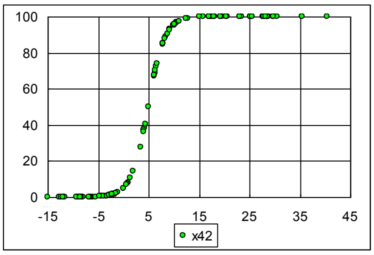
д



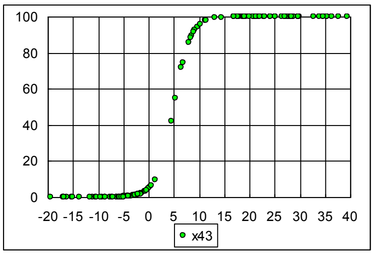
д



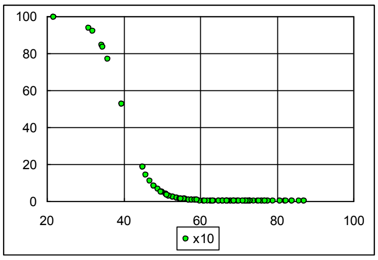
е



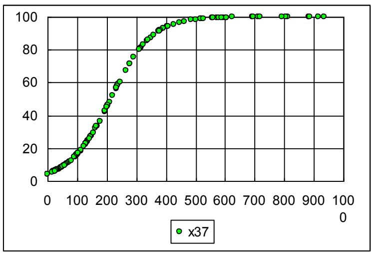
е



е

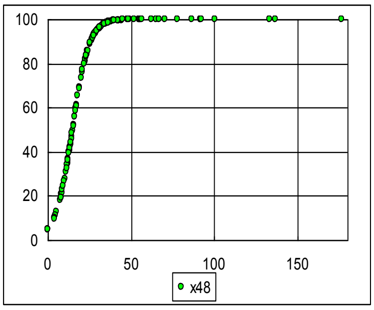


є

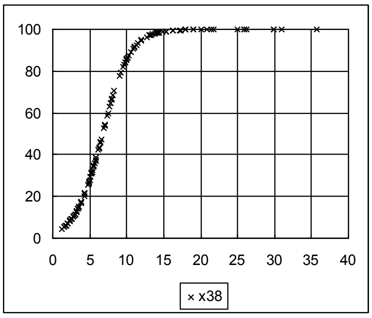


є

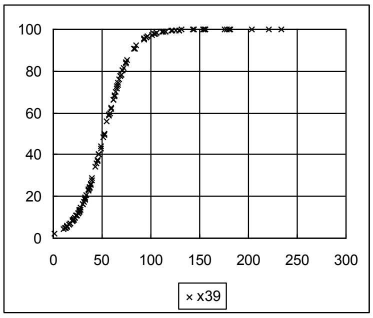
Продовження рис. 5.8



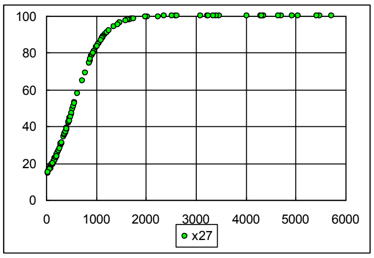
є



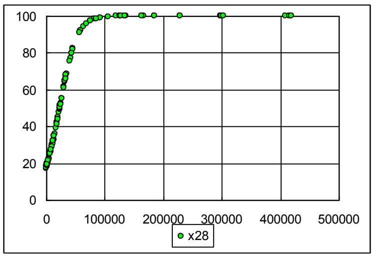
ж



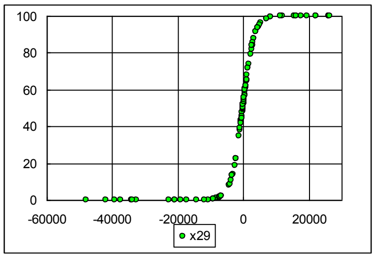
ж



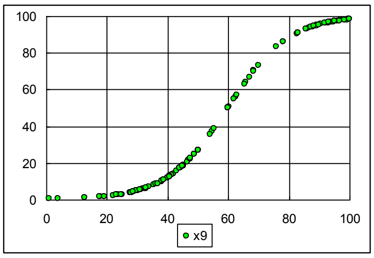
з



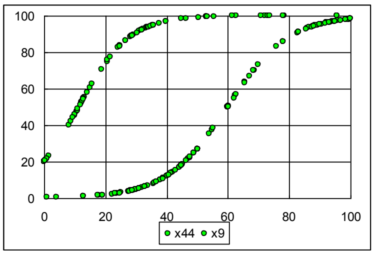
з



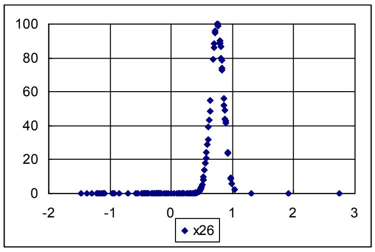
з



и



и



і

Закінчення рис. 5.8

Функції перетворень величин елементарних метричних ознак соціально-економічних систем є тими твірними елементами, за допомогою яких можна вимірювати сумісно чи сукупно складні ознаки різного рівня опису системи. При цьому функції перетворень не залежать від способу виявлення складних ознак: чи за допомогою теоретико-логічного аналізу суттєвого змісту соціально-економічної системи, чи аналітично вимірників за допомогою математичних методів. Перетворені значення економічних показників порівнянні між собою в статиці та динаміці, без прив’язок до будь-яких сукупностей чи вибірок.

П’ятою проблемою в побудові є визначення узагальненої функції. Вибір мультиплікативної форми узагальнюючої функції визначення складної ознаки пояснюється наступним. Окремі показники $y_j(x_j)$ становлять функції розподілу одновимірних випадкових величин, тому їх добуток є функцією розподілу m-вимірної випадкової величини з незалежними компонентами. У цьому випадку значення зведеного показника $Y(y_j)$ в точці $y^0 = (y_1^0, ..., y_m^0)$ може інтерпретуватися як частка об’єктів, що мають за всіма окремими показниками значення, менші, ніж значення $y_1^0, ..., y_m^0$ [315, с. 99 – 103].

Одним з перших, хто запропонував розробку узагальнюючого показника якості мультиплікативного типу, був Льюїс Керрол (Ч. Л. Доджсон), який розглянув проблему комплексної оцінки якості у вигляді цікавих задач [110]. В одній із цих задач потрібно оцінити роботу трьох майстринь, які в’язали шарфи, за трьома ознаками: продуктивності, ваги продукції та здатності шарфа зберігати тепло. Л. Керрол припустив, що для одержання зведеної оцінки всі отримані відносні оцінки якості за кожною ознакою треба перемножити. Цю пропозицію він обґрунтовує наступним прикладом: як оцінити роботу декількох грабарів, що викопали котловани різних розмірів? Поза будь-яким сумнівом, це можна зробити, за обсягом вийнятого ґрунту, для цього три розміри котловану потрібно перемножити. Зі свого боку, слід додати, що всі відомі комплексні показники у фізиці складені саме за цим принципом (кВт/год., м/с^2).

Відомо, що мультиплікативна синтезувальна функція називається консервативною функцією, оскільки вона надає оцінку $Y(y)$ об’єкту $y = (y_1, ..., y_m)$ в цілому і значення якої не перевищують "найгіршої" (мінімальної) з оцінок $y_1, ..., y_m$, що отримують за окремими критеріями. Достатньо за одним з критеріїв мати оцінку $y_i = 0$ для того, щоб отримати зведену оцінку $Y(y) = 0$. Оптимістичнішою вважається оцінка, отримана за допомогою зведеного показника $Y_*(y)$, що є середнім геометричним величин $y_1, ..., y_m$, та задовольняє нерівності:

$$\min_s y_s \le Y_*(y) = \sqrt[m]{\prod_{s=1}^{m} y_s} \le \max_s y_s. \qquad (5.17)$$

Введення середнього геометричного $Y_*(y)$, як уже було сказано, є одним із найбільш популярних узагальнених середніх величин, добре відпрацьованих з точки зору автоарифметизації числових шкал. Узагальнення поняття автоарифметизації шкали дійсних чисел, що отримується шляхом застосування ізоморфізмів між множиною всіх дійсних чисел і різними його підмножинами, дозволяє розширити клас неперервних строго монотонних відображень φ, які використовуються для породження різних видів узагальненого середнього $Y_+^{\varphi}(y)$ [296, с. 95 – 106].

Отже, після того, як визначена шкала перетворення та описані ознаки окремими функціями перетворення, має обчислюватися узагальнений вимірник Y як середнє геометричне окремих функцій перетворення ознак:

$$Y = \sqrt[n]{\prod_{i=1}^{n} y_i}\ . \tag{5.18}$$

Вибір формули середнього геометричного для розрахунку узагальнюючого вимірника також цілком обґрунтований практичним розв’язанням задач в економіці: під час визначення рівня конкурентоспроможності підприємства, рівня конкурентних переваг, конкурентного статусу за виділеною системою ознак у заданих умовах, моделювання вимірників у проведенні контролінгу банківської діяльності [145; 150; 154; 159; 163]. Якщо хоча б одна з ознак, що входять у систему параметрів конкурентоспроможності, не задовольняє вимогам закономірності свого розвитку, що закладені в її функції перетворення, то, якими б достатніми не були інші ознаки, підприємство не можна вважати конкурентоспроможним. Дійсно, спосіб задавання узагальненої функції перетворення такий, що коли хоча б одна окрема перетворена ознака $y_i = 0$, то узагальнена функція теж буде дорівнювати нулю, з іншого боку, оскільки $0 < y_i \le 1$, то і $0 < Y \le 1$. Узагальнена функція перетворення дуже чутлива до низьких перетворених значень окремих ознак. Узагальнена функція перетворення є кількісним, однозначним, єдиним та універсальним показником визначення ознак соціально-економічних систем, і якщо додати ще такі властивості, як адекватність, ефективність і статистична чутливість, то стає очевидним, що її можна використовувати як критерій оптимізації у процесі розв’язання оптимізаційних задач функціонування та розвитку соціально-економічних систем.

Спосіб задавання граничних позначок на шкалі перетворення узагальненої функції якості ознак СЕС є таким самим, як і для окремих функцій перетворень ознак. В узагальнену функцію перетворення можуть входити різноманітні за своєю суттю показники ознак (і відносні, і абсолютні), що мають свої окремі функції перетворення з різними типами обмежень. Таким чином поєднуються різні види метричних величин та вимірюються елементарні й складні метричні ознаки системи.

У табл. 5.7 наведені значення вимірників кожної з 11 складних ознак кожного з 28 підприємств у динаміці за п’ять років.

Таблиця 5.7

Значення вимірників складних ознак виробничо-господарської діяльності підприємств

Період	Об'єкт	I_{F_1}	I_{F_2}	I_{F_3}	I_{F_4}	I_{F_5}	I_{F_6}	I_{F_7}	I_{F_8}	I_{F_9}	$I_{F_{10}}$	$I_{F_{11}}$	I_{1-10}	I_{1-11}
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	33,38	19,49	8,28	0,00	23,02	0,00	33,78	49,92	22,09	25,45	0,00	0,00	0,00
2	1	10,80	31,74	90,28	75,67	8,23	8,60	5,88	41,48	32,95	47,35	0,99	23,99	17,96
3	1	0,00	57,67	36,08	99,98	64,04	100,00	0,00	51,36	33,76	98,24	0,00	0,49	0,00
4	1	22,39	15,83	32,06	88,70	7,54	0,28	0,00	50,38	25,65	7,05	17,68	1,38	1,74
5	1	17,20	21,37	7,58	0,00	34,44	0,09	20,79	44,59	0,00	23,25	0,00	0,01	0,00
1	2	47,76	73,14	55,38	82,18	64,85	86,95	4,56	40,54	42,51	3,88	17,68	34,96	32,86
2	2	14,59	6,31	0,90	0,00	44,12	0,00	32,88	23,53	0,00	0,87	0,00	0,00	0,00
3	2	10,57	11,84	24,73	1,99	38,22	0,14	21,57	40,54	25,43	0,00	0,00	0,31	0,01
4	2	23,40	22,88	32,07	16,34	62,36	6,72	8,28	24,97	48,65	97,22	0,01	25,47	12,87
5	2	5,32	42,80	14,73	99,62	25,24	60,35	15,05	47,76	28,65	76,96	5,88	30,95	26,61
1	3	19,53	16,54	0,01	0,00	33,41	0,42	7,26	49,45	31,29	11,69	0,00	0,05	0,00
2	3	43,39	46,06	36,55	75,97	50,20	0,34	4,20	49,07	28,76	22,32	0,00	20,43	4,32
3	3	14,82	12,19	7,10	0,00	25,03	0,12	97,63	32,75	24,57	1,69	0,00	0,09	0,00
4	3	0,00	68,36	35,01	99,96	69,10	0,00	2,26	47,03	20,50	18,99	0,00	0,04	0,00
5	3	27,48	0,00	0,00	0,00	44,09	0,00	41,36	41,67	14,90	11,01	0,00	0,00	0,00
1	4	16,56	20,54	6,36	0,00	37,70	0,00	33,79	51,20	23,42	15,34	0,00	0,01	0,00
2	4	19,50	18,54	0,25	0,00	16,94	0,35	28,22	29,83	37,22	12,83	0,00	0,08	0,00
3	4	31,93	15,46	25,54	74,15	37,83	0,15	32,33	35,75	25,48	6,62	9,80	15,86	15,18
4	4	0,85	78,16	36,27	99,84	55,25	100,00	2,56	49,92	50,63	94,73	0,00	30,97	9,20
5	4	14,45	18,02	30,38	9,65	28,79	4,41	19,62	50,49	34,47	78,97	0,00	21,96	0,38
1	5	17,66	18,61	32,77	94,06	7,06	0,00	8,31	26,49	32,46	23,06	79,04	2,06	2,87

Продовження табл.. 5.7

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	5	21,09	14,96	13,16	89,65	7,20	0,42	2,01	23,89	32,04	9,17	39,03	10,48	11,81
3	5	0,02	26,27	0,00	23,41	42,47	0,15	2,72	40,54	26,72	25,52	0,01	1,64	1,03
4	5	13,74	7,10	0,03	0,00	18,22	0,00	43,23	41,81	17,72	4,64	0,00	0,01	0,00
5	5	63,58	83,80	61,49	46,46	63,86	99,97	9,66	44,30	82,15	90,10	7,89	56,08	46,92
1	6	18,93	16,49	13,54	0,00	21,84	0,00	19,36	46,50	21,21	16,54	0,00	0,11	0,00
2	6	17,57	73,24	47,61	99,71	41,06	99,96	8,14	27,07	87,96	88,99	2,32	46,09	35,12
3	6	42,15	74,82	50,99	92,30	21,12	76,41	13,71	51,21	100,00	65,57	0,00	50,61	5,97
4	6	33,45	39,37	32,75	53,33	56,89	5,36	27,41	44,00	60,75	58,95	0,00	35,33	1,85
5	6	0,02	18,18	37,06	42,70	8,57	0,69	5,57	34,59	31,56	19,80	0,01	7,21	3,83
1	7	0,00	35,94	41,92	99,98	65,00	99,99	0,00	45,17	40,92	83,75	41,37	0,48	0,71
2	7	38,96	10,40	36,50	91,65	46,34	0,10	0,00	32,20	25,56	28,46	2,44	1,65	1,71
3	7	16,54	0,00	6,30	0,00	28,20	0,00	19,86	37,43	0,00	12,59	0,00	0,00	0,00
4	7	55,04	17,72	28,37	38,89	63,31	0,00	5,47	30,21	32,87	56,97	0,00	2,54	0,69
5	7	12,71	0,00	0,10	0,00	38,59	0,00	36,20	26,68	0,00	1,43	0,00	0,00	0,00
1	8	24,17	16,69	25,12	1,91	6,71	4,74	12,66	24,11	25,23	43,69	0,00	13,54	1,41
2	8	0,00	32,92	22,07	99,99	70,89	99,94	3,31	26,39	48,02	96,88	24,13	0,19	0,29
3	8	0,07	34,52	13,83	99,91	66,93	19,07	10,64	35,71	28,28	26,32	88,69	16,25	18,96
4	8	22,90	9,57	2,60	0,00	31,08	0,00	5,85	50,90	13,22	39,03	0,00	0,00	0,00
5	8	14,28	0,01	12,85	93,51	42,03	0,00	2,60	50,77	27,26	2,99	5,17	0,06	0,08
1	9	15,51	23,45	8,88	0,00	53,25	52,93	96,70	40,59	25,25	15,02	0,00	0,02	0,00
2	9	22,11	75,36	42,77	20,79	7,98	74,75	1,05	3,61	98,64	76,40	0,00	21,88	8,90
3	9	20,07	0,00	0,00	0,00	57,02	0,00	43,11	39,93	10,72	3,28	0,00	0,00	0,00
4	9	20,83	9,62	14,85	0,00	44,76	0,00	28,70	51,30	5,70	8,72	0,00	0,01	0,00
5	9	25,09	0,00	0,11	0,00	41,38	0,00	26,81	36,35	15,07	3,32	0,00	0,00	0,00
1	10	34,34	20,19	20,69	65,17	50,04	0,35	31,84	35,33	24,91	15,17	0,09	19,23	11,83
2	10	0,00	43,05	16,82	99,90	40,61	54,05	1,94	45,22	43,06	43,47	43,88	9,18	10,58
3	10	25,11	7,52	9,39	0,00	46,76	0,00	11,94	50,43	10,55	6,08	0,00	0,00	0,00
4	10	22,99	8,27	27,99	92,32	6,99	0,15	7,63	44,22	27,50	2,82	43,11	10,33	11,77

Продовження табл.. 5.7

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	10	26,15	22,35	37,32	60,60	8,63	0,38	3,19	23,34	30,57	42,01	1,44	14,49	11,74
1	11	8,86	23,36	28,31	6,96	16,76	0,00	2,87	49,78	26,41	72,96	0,00	1,47	0,32
2	11	18,39	0,00	0,01	0,00	43,36	0,00	25,83	29,92	7,85	13,66	0,00	0,00	0,00
3	11	47,26	68,73	51,02	79,76	32,31	48,82	9,02	49,48	64,35	43,91	2,31	43,85	33,55
4	11	8,85	21,05	5,62	0,00	20,37	0,25	18,15	36,61	24,23	30,89	0,00	0,09	0,00
5	11	16,94	57,87	56,71	99,46	26,78	94,82	7,73	20,50	92,31	45,72	99,02	39,56	43,00
1	12	46,14	25,86	25,06	82,12	50,49	0,11	8,68	50,70	0,00	11,93	0,00	1,66	0,09
2	12	38,94	72,33	36,57	49,01	56,76	76,57	14,70	49,11	81,60	47,03	0,00	47,68	13,46
3	12	0,02	62,86	55,72	66,77	11,06	78,81	4,07	26,39	34,28	33,61	0,84	14,55	11,23
4	12	0,00	55,93	61,33	99,95	59,99	99,68	4,13	34,16	46,09	81,66	89,03	8,44	10,45
5	12	21,22	29,32	55,05	99,19	50,71	11,16	0,00	40,84	26,71	33,59	54,65	3,01	3,91
1	13	19,83	12,12	4,85	0,00	49,14	0,00	15,50	28,99	0,00	9,89	0,00	0,00	0,00
2	13	32,63	72,16	63,65	98,65	41,86	99,90	3,20	49,93	43,01	73,62	13,73	44,62	40,09
3	13	11,17	11,57	0,10	0,00	24,91	0,00	36,89	21,48	28,47	2,76	0,00	0,00	0,00
4	13	23,15	17,74	36,67	14,69	10,28	4,74	47,78	26,85	24,94	43,76	0,00	20,79	6,76
5	13	0,00	32,55	18,49	99,99	70,87	99,23	1,29	28,77	47,35	97,04	23,58	0,12	0,19
1	14	0,38	40,13	26,82	99,89	63,34	17,31	2,75	37,40	28,59	36,92	90,34	18,57	21,44
2	14	25,66	49,03	35,87	2,08	31,25	33,66	4,81	44,52	76,46	41,54	0,00	24,14	0,21
3	14	4,61	49,38	30,18	96,22	42,63	3,83	3,64	46,86	28,53	8,74	0,00	18,46	5,54
4	14	19,23	0,00	11,10	0,00	56,56	0,00	91,56	37,57	24,04	10,85	0,00	0,00	0,00
5	14	23,26	89,00	61,45	98,44	29,64	86,90	1,51	34,07	99,96	59,26	0,08	39,73	22,72
1	15	18,98	54,59	0,30	0,00	59,11	40,49	29,61	47,43	63,96	32,61	0,00	0,02	0,00
2	15	16,64	32,28	21,13	0,00	37,05	12,45	24,04	43,24	73,07	2,35	0,00	1,46	0,01
3	15	24,02	9,54	0,17	0,00	39,28	0,00	56,66	41,95	27,89	1,96	0,00	0,01	0,00
4	15	43,87	21,71	27,33	72,25	58,30	2,56	22,91	50,68	25,27	51,69	0,06	29,04	16,42
5	15	0,01	32,97	19,02	99,87	40,40	16,98	1,61	39,25	41,27	43,61	49,18	11,64	13,27
1	16	26,15	27,87	20,15	0,00	49,63	2,02	10,01	38,37	54,41	8,32	0,00	1,13	0,01
2	16	1,70	27,10	19,73	98,68	8,00	6,48	5,57	40,40	29,44	6,87	78,51	13,57	15,92
3	16	44,09	0,00	12,65	84,56	18,23	0,00	0,00	21,65	20,93	34,62	2,81	0,07	0,09
4	16	27,69	75,19	46,26	25,19	43,12	93,33	2,58	35,76	27,45	67,86	0,00	33,30	9,39

Продовження табл.. 5.7

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	16	22,58	77,63	10,53	0,00	43,32	97,40	19,08	36,06	77,05	25,02	0,00	0,03	0,00
1	17	30,54	92,93	58,89	99,01	44,95	97,51	4,53	46,01	77,52	48,86	48,83	47,40	47,53
2	17	9,97	69,91	24,59	0,15	33,20	99,97	16,14	39,10	97,18	33,42	0,00	21,07	0,24
3	17	0,00	63,22	42,91	99,93	42,77	98,77	6,23	25,92	92,81	50,26	56,30	15,65	17,58
4	17	34,17	64,34	41,41	44,13	31,21	16,57	8,62	50,17	100,00	14,03	0,00	32,36	2,92
5	17	34,05	53,15	45,33	45,81	61,54	99,87	16,59	51,76	0,00	42,05	0,00	4,20	0,98
1	18	0,02	51,28	43,46	82,93	19,47	100,00	2,09	31,50	69,69	13,11	0,31	14,81	10,43
2	18	0,00	66,43	29,44	66,78	63,61	100,00	20,60	28,21	61,91	79,86	72,87	6,60	8,21
3	18	4,02	32,16	32,70	52,60	59,48	100,00	0,00	32,08	31,84	9,40	20,90	2,41	2,94
4	18	32,58	16,92	1,49	24,85	35,69	0,00	0,00	35,14	0,00	12,05	0,00	0,01	0,00
5	18	18,84	38,60	38,87	41,28	30,99	99,97	3,65	32,82	18,31	21,45	3,58	26,49	22,08
1	19	10,44	22,52	31,06	5,13	8,41	2,36	21,95	37,32	0,00	49,34	0,00	1,45	0,00
2	19	0,00	31,68	16,37	99,48	61,42	100,00	97,29	49,28	37,48	68,46	2,36	12,27	10,56
3	19	0,00	46,24	7,58	99,99	70,36	100,00	0,83	30,33	53,18	20,06	9,51	0,35	0,47
4	19	0,00	45,52	26,55	67,74	61,67	99,98	3,07	48,70	41,89	23,42	86,44	14,05	16,57
5	19	32,79	45,14	0,00	33,37	49,02	0,00	0,00	50,47	0,00	12,80	0,00	0,00	0,00
1	20	0,00	35,99	28,54	54,12	41,89	2,04	3,99	45,00	44,63	32,85	85,93	5,51	7,07
2	20	39,37	31,14	52,28	23,98	54,71	80,73	79,89	28,86	30,01	51,15	0,00	43,46	0,75
3	20	2,97	32,89	34,66	64,07	33,61	100,00	0,00	22,13	100,00	10,01	0,00	2,72	1,00
4	20	24,36	49,06	48,58	13,62	58,89	98,60	8,63	51,91	91,09	19,20	0,00	35,94	0,91
5	20	33,42	30,69	12,24	15,48	25,63	0,00	0,00	34,05	0,00	5,19	0,00	0,02	0,00
1	21	28,59	26,81	0,00	17,83	40,76	0,00	21,09	44,68	0,00	41,81	0,00	0,01	0,00
2	21	47,96	34,14	32,65	8,86	60,64	25,06	20,89	51,93	23,18	43,12	0,01	30,85	14,47
3	21	27,25	27,87	35,43	14,84	34,30	58,00	1,91	35,31	64,95	19,67	95,88	24,19	27,41
4	21	32,14	30,59	43,18	29,35	50,08	4,06	5,48	40,67	0,00	8,51	0,00	1,99	0,17
5	21	0,00	21,52	13,84	37,05	5,83	5,89	5,35	34,07	20,74	7,82	52,01	4,27	5,35
1	22	0,44	22,84	52,10	96,39	31,32	2,83	0,00	50,91	60,83	5,59	31,99	1,55	2,04
2	22	24,79	52,00	42,35	67,37	52,59	94,83	0,00	47,57	30,54	16,26	0,00	3,76	1,21
3	22	35,86	53,35	51,06	18,79	48,32	95,96	9,45	31,47	80,30	26,75	0,00	37,46	0,09
4	22	1,55	49,40	63,60	94,18	23,60	98,51	4,95	45,71	0,00	27,00	8,07	2,58	2,86

Закінчення табл.. 5.7

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	22	0,00	0,00	33,07	14,50	29,91	97,95	7,32	29,29	0,00	5,85	0,00	0,20	0,00
1	23	0,01	73,70	49,15	57,64	36,66	99,17	8,63	23,65	94,06	41,48	99,85	18,62	21,69
2	23	36,01	62,88	22,50	33,84	37,31	71,38	9,17	49,16	0,00	16,10	0,00	3,06	0,06
3	23	23,94	50,10	46,28	63,81	61,79	99,76	17,74	51,72	0,00	42,14	0,00	4,22	1,94
4	23	0,02	47,56	41,87	84,30	18,90	100,00	1,72	25,99	80,02	13,11	0,23	14,23	9,77
5	23	0,00	66,50	29,75	68,04	63,73	100,00	20,92	28,23	64,39	79,43	73,78	6,65	8,28
1	24	3,20	32,40	32,74	54,69	59,44	100,00	0,00	32,63	31,97	9,40	24,21	2,37	2,93
2	24	32,51	15,53	1,45	24,90	35,70	0,00	0,00	35,12	0,00	11,87	0,00	0,01	0,00
3	24	17,60	51,73	38,82	41,88	30,59	99,97	2,49	32,76	14,68	21,45	4,10	25,50	21,60
4	24	10,44	19,68	31,06	5,13	8,42	2,36	32,10	37,32	0,00	49,34	0,00	1,47	0,00
5	24	0,00	40,54	16,69	99,47	61,31	100,00	53,04	49,27	37,41	68,41	2,55	11,85	10,30
1	25	0,00	46,71	7,11	99,99	70,26	100,00	0,56	29,08	53,21	19,75	8,65	0,33	0,44
2	25	0,00	47,57	24,87	67,50	61,67	99,98	3,01	48,69	41,75	22,63	79,68	13,20	15,54
3	25	32,80	29,65	0,00	33,43	49,17	0,00	0,00	50,34	0,00	12,80	0,00	0,00	0,00
4	25	0,00	35,46	27,95	53,89	41,81	2,06	4,19	45,76	44,04	32,75	88,69	5,50	7,08
5	25	39,21	26,61	52,31	23,88	54,79	80,11	79,98	28,87	30,18	49,40	0,00	42,60	0,73
1	26	3,19	39,51	33,62	65,55	33,17	100,00	0,00	22,14	100,00	9,88	0,00	2,78	1,00
2	26	24,48	41,97	48,58	13,71	59,03	98,60	8,77	51,93	90,60	19,20	0,00	35,47	0,91
3	26	33,47	34,78	12,22	15,56	25,69	0,00	0,00	33,81	0,00	5,21	0,00	0,02	0,00
4	26	28,92	23,46	0,00	18,16	41,60	0,00	21,53	42,93	0,00	41,81	0,00	0,01	0,00
5	26	46,89	34,58	31,75	8,59	59,83	25,89	20,83	51,95	23,01	42,99	0,01	30,66	15,25
1	27	28,30	24,12	35,94	14,96	34,66	54,75	2,68	35,54	49,60	20,02	94,85	24,09	27,29
2	27	30,21	42,53	41,81	28,01	50,39	4,00	6,70	40,14	0,00	9,58	0,00	2,10	0,03
3	27	0,00	20,42	12,75	36,61	5,81	5,55	5,39	33,02	20,52	7,73	42,56	3,22	4,07
4	27	0,53	22,75	52,08	96,47	31,43	3,26	0,00	50,92	59,19	5,52	28,82	1,60	2,08
5	27	19,47	37,32	41,68	75,07	52,13	95,66	0,00	47,63	30,53	16,38	0,09	3,58	2,57
1	28	35,79	50,33	51,07	18,73	48,30	97,01	12,68	31,45	80,22	26,28	0,00	38,30	0,09
2	28	1,59	64,15	62,42	94,21	23,75	97,54	4,08	46,43	0,00	26,13	7,88	2,59	2,86
3	28	29,13	36,16	34,67	14,52	29,99	97,95	8,01	29,33	0,00	34,22	0,00	2,77	0,00
4	28	0,01	73,71	49,39	57,82	36,86	98,55	8,48	23,65	95,38	41,51	99,90	18,51	21,58
5	28	36,09	63,55	22,81	33,05	35,25	66,31	9,12	49,35	0,00	16,16	0,00	3,02	0,06

При графічному зображенні значень вимірників 11 метричних складних ознак підприємств, що увійшли до першої групи, маємо наочну демонстрацію узгодженості змін значень вимірників та підтвердження існування підприємств (об’єкти 24, 13 на рис. 5.9), які мають високі рівні величин вимірників усіх ознак своєї виробничо-господарської діяльності.

Існування нульових значень вимірників окремих складних метричних ознак на кожному підприємстві (рис. 5.9) – результат жорстких умов побудови даного показника: достатньо мати нульове значення величини елементарної ознаки, щоб одержати нульове значення величини всієї складної ознаки, до якої входить елементарна. Ця вимога передбачає структуру складної ознаки формувати обґрунтовано: якщо аналітично, то адекватність перевірити за статистичними критеріями, якщо теоретично, то структурно-логічно та довести дією законів і закономірностей в економіці.

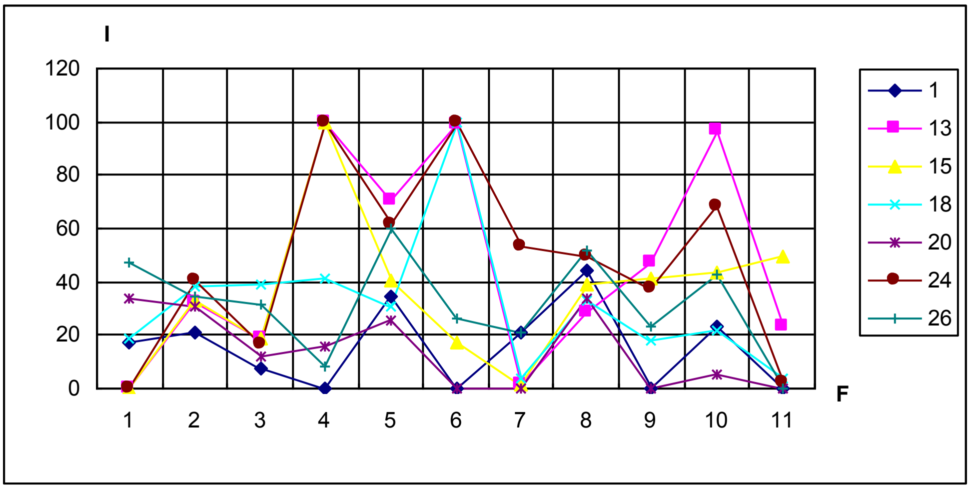


Рис. 5.9. **Значення вимірників складних метричних ознак підприємств першої групи (період – перший рік)**

Для порівняння результатів моделювання вимірників з іншими моделями, що передбачають побудову узагальнюючого показника, були обчислені значення таксономічного показника розвитку $\mathbf{(}_{p}$ метричної складної ознаки забезпеченості підприємства ОАО "Автрамат" оборотними засобами в динаміці за 5 років. Графік значень узагальнюючого показника $\mathbf{(}_{p}$ наведений на рис. 5.10, де також зображений графік значень вимірника $\mathbf{(}_{V}$ за тими ж даними.

Оскільки в обчисленні таксономічного показника розвитку здійснюються порівняння з еталоном за критерієм МінМах в межах вибірки, то всі негативні наслідки, пов’язані з неякісними статистичними характеристиками окремих вибірок, відображаються в рівнях значень узагальнюючих показників, що за ними обчислюються, це є очевидним фактом на рис. 5.10.

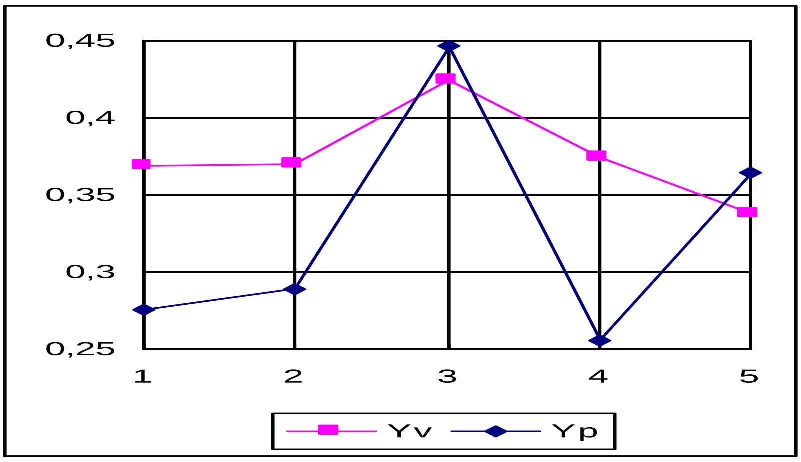


Рис. 5.10. **Графіки значень таксономічного показника розвитку (Y_p) і вимірника (Y_V)**

Низькі значення показника Y_p маємо порівняно з третім роком спостережень. Порівняння маємо в межах тільки даного підприємства. Реально йде спад – це також підтверджує графік Y_V рис. 5.11, але не такий критичний, як на 4-му році спостереження, й незростання за 5-й рік, як за графіком Y_p.

Отримані функції перетворень демонструють закономірності розвитку метричних елементарних і складних ознак, які характеризують стан промислових підприємств протягом періоду, що розглядається. Перетворені величини є вимірниками та тією початковою базою, з якої можна конструювати складні моделі для управлінського аналізу в економіці. Таким чином, отримана система величин формує систему вимірників, що можуть бути індикаторами соціально-економічних систем (у статиці й динаміці).

Отже, вказані переваги моделей вимірників та їх практичні застосування підтверджують доцільність впровадження його в систему управління економікою і свідчать, що вимірники посилюють аналітичну підтримку бухгалтерського, фінансового, управлінського обліків та статистичної звітності в країні.

Формалізація у моделюванні вимірників складних ознак соціально-економічної системи виконується через поєднання чітких математичних викладок та знання закономірностей розвитку ознак системи. Виявилося, що для шкал перетворень статистична чутливість і ефективність окремих та узагальненої функцій перетворення не нижчі, ніж для будь-якого технологічного показника в техніці [87].

Таким чином, у процесі розв'язання практичних задач в економіці, що передбачають визначення метричних ознак соціально-економічних систем за допомогою побудови їх вимірників, слід дотримуватися наступних рекомендацій:

розробляти функцію перетворення для кожної елементарної ознаки об'єкта окремо чи спільно (за типами) залежно від знань про дану ознаку та результатів її статистично-математичного опису;

враховувати, що не існує єдиної функції перетворення для всіх ознак системи з огляду на складність її розвитку;

необхідно знайти оптимальну кількість функцій перетворень ознак для конкретної системи, які є моделями розвитку окремих ознак соціально-економічної системи;

пам'ятати про існування однотипних груп ознак СЕС, що передбачають існування типів функцій перетворень;

базуватись на дослідженні, що виокремлюються такі функції перетворень ознак СЕС: односторонні симетричні (зростаючі та спадні), двосторонні симетричні і двосторонні несиметричні;

знати, що кожна функція перетворень ознаки описується аналітично;

враховувати, що тип функції перетворень ознак адаптується для окремої ознаки окремої системи за рахунок зміни та підбору параметрів функції.

5.3. Вимірники неметричних ознак та загальної якості ознак соціально-економічної системи

Загальну якість ознак соціально-економічної системи можна визначити за умови моделювання характеристик, які є складними ознаками стану, процесів функціонування та розвитку й визначаються метричними і неметричними елементарними ознаками. Ідентифікація моделі об'єкта – це вже процес його вимірювання, але, з іншого боку, модель будується, щоб визначити ознаки об'єкта для глибшого його пізнання з метою управління ним: "процес ідентифікації моделі на основі даних про реальний об'єкт ... можна розглядати як процес вимірювання різних характеристик об'єкта; інструментом вимірювання при цьому є сама ідентифікована модель" [97, с. 7]. У підрозділі 5.2 викладені обґрунтування розбудови моделей вимірників на основі метричних ознак. Але оцінка соціально-економічної системи тільки кількісними характеристиками обмежує її, а в окремих випадках призводить до похибок. Слід ще раз нагадати про важливість вирішення наукової проблеми сумісного та сукупного вимірювання різних величин, які характеризують одну й ту ж систему; вона залишається актуальною в загальній теорії вимірювання й сьогодні [99]. У спеціальній літературі майже відсутні рекомендації щодо моделювання вимірників неметричних ознак системи. Комплексно об'єкт в економіці характеризується системою ознак, що розподіляються на кількісні та якісні, екстенсивні та інтенсивні, метричні та неметричні, й прояв системи властивостей об'єкта на елементарному рівні акумулюється в загальній якості його ознак [160].

Моделювання вимірників неметричних ознак рекомендується здійснювати за логікою, наведеною на рис. 5.3, але для неметричних ознак існує своє вирішення проблем 2, 3 та 4 даної схеми.

Перетворення неметричних величин має відбуватися за допомогою функцій перетворень ознак, що вимірюються за допомогою номінальних та порядкових шкал.

Функції перетворення величин неметричних ознак мають бути дискретними функціями і задаватися таблицею.

Вигляд функції обумовлений специфікою неметричних величин. Якісна ознака може бути представлена множиною рівнозначних номінацій, які є переліком можливих назв якості, наприклад, перелік функцій управління: організація, планування, прогнозування, аналіз, оцінка, контроль, регулювання. Окремій номінації якісної ознаки ставиться у відповідність кількісне значення функції перетворення залежно від шкали бажаності окремих номінацій в їх системі адекватно до дій закономірностей функціонування та розвитку соціально-економічних систем.

Задача 5.3. Провести аналіз даних інтелектуальної діяльності 7 промислових підприємств Харківського регіону на основі моделювання вимірників неметричних ознак соціально-економічної системи.

Розв'язання задачі. Спираючись на умовні позначення номінальних ознак інтелектуального забезпечення діяльності підприємств (табл. 5.8), задамо окремі функції перетворення неметричних величин кожної ознаки. Так, ознака A (якість програмного забезпечення) може набувати три номінації: низьку (A_1), середню (A_2) та високу (A_3), при цьому функція перетворення набуває значень 0,3; 0,7 та 1,0. Значення функції перетворення залишається на проміжку $[0;1]$. Присвоєння значень відбувається експертним шляхом залежно від важливості окремої номінації властивості в описі даного об'єкта та різниці між окремими номінаціями. Якщо неметрична ознака набуває тільки альтернативних значень: "так" чи "ні" або "є" чи "немає", то значення функції встановлюється залежно від важливості даних категорій чи номінацій. Наприклад, ознака "наявність чітко розробленої та документально оформленої філософії бізнесу" набуває таких альтернативних значень. Можлива відсутність даної ознаки оцінюється у 30% ($y = 0{,}3$), а її бажана наявність – у 90% ($y = 0{,}9$). У табл. 5.8 задана функція перетворення для системи номінальних ознак, що є шкалами вимірювань якісних показників інтелектуального забезпечення виробничо-господарської діяльності промислових підприємств.

Таблиця 5.8

Окремі функції перетворення неметричних номінальних ознак інтелектуального забезпечення діяльності підприємств

Позначення ознак	Умовні позначення номінацій ознак	Функція перетворення
A – якість програмного забезпечення	$A1$	1,0
	$A2$	0,7
	$A3$	0,3
B – наявність чітко розробленої та документально оформленої філософії бізнесу	$B1$	0,9
	$B0$	0,3
C – рівень здоров’я працівників	$C1$	1,0
	$C2$	0,8
	$C3$	0,5
D – рівень користування об’єктами соціальної інфраструктури	$D1$	0,9
	$D2$	0,6
	$D3$	0,2
E – ступінь довіри працівників до вищого керівництва	$E1$	0,9
	$E2$	0,7
	$E3$	0,5
F – рівень залежності працівників від підприємства	$F1$	0,9
	$F2$	0,7
	$F3$	0,6
G – наявність постійних клієнтів (за критерієм постачання або замовлень понад 5)	$G1$	0,8
	$G0$	0,2
H – наявність рекламацій із боку споживачів	$H1$	0,8
	$H0$	0,5

K – наявність персональних комп'ютерів та програмного забезпечення у відділі з роботи з клієнтами	$K1$	0,8
	$K0$	0,5
L – ведення уніфікованої бази даних постачальників та споживачів	$L1$	0,8
	$L0$	0,6

У табл. 5.9 наведені відкалібровані значення неметричних номінальних ознак інтелектуального забезпечення виробничо-господарської діяльності 7 промислових підприємств Харківського регіону в динаміці за 3 роки.

Таблиця 5.9

Значення окремих функцій перетворення номінальних ознак та їх узагальнюючого показника вимірювання *Y*

Підпри-ємство	Роки	Ознаки										Показник
		A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*K*	*L*	*Y*
1	2001	0,3	0,3	1,0	0,9	0,9	0,9	0,8	0,8	0,5	0,8	0,66
	2002	0,7	0,3	1,0	0,9	0,9	0,9	0,2	0,5	0,8	0,8	0,63
	2003	0,7	0,3	0,8	0,6	0,9	0,9	0,8	0,8	0,8	0,8	0,68
2	2001	0,7	0,3	1,0	0,9	0,7	0,7	0,8	0,5	0,5	0,8	0,66
	2002	0,7	0,3	0,8	0,9	0,7	0,9	0,8	0,5	0,5	0,8	0,66
	2003	0,7	0,9	0,8	0,6	0,7	0,9	0,8	0,5	0,8	0,8	0,74
3	2001	0,3	0,3	1,0	0,6	0,7	0,7	0,8	0,5	0,5	0,6	0,56
	2002	0,3	0,3	0,8	0,6	0,5	0,7	0,8	0,5	0,5	0,6	0,53
	2003	0,7	0,3	0,8	0,6	0,5	0,7	0,8	0,5	0,8	0,6	0,61
4	2001	0,3	0,3	0,8	0,6	0,7	0,7	0,8	0,5	0,8	0,6	0,58
	2002	0,3	0,3	0,8	0,6	0,7	0,7	0,8	0,5	0,8	0,6	0,58
	2003	0,7	0,3	0,8	0,6	0,7	0,7	0,8	0,5	0,8	0,8	0,65
5	2001	0,7	0,3	0,8	0,6	0,5	0,9	0,2	0,5	0,8	0,6	0,54
	2002	0,7	0,3	0,8	0,2	0,5	0,6	0,8	0,8	0,8	0,6	0,66
	2003	0,7	0,3	0,8	0,2	0,7	0,6	0,8	0,5	0,6	0,8	0,57
6	2001	0,7	0,3	0,5	0,9	0,7	0,6	0,2	0,5	0,5	0,6	0,51
	2002	0,7	0,3	0,8	0,9	0,7	0,6	0,8	0,8	0,8	0,6	0,67
	2003	0,7	0,3	0,8	0,6	0,7	0,7	0,8	0,5	0,8	0,6	0,63
7	2001	0,3	0,3	0,5	0,2	0,7	0,7	0,8	0,5	0,5	0,8	0,48
	2002	0,7	0,9	0,8	0,6	0,7	0,7	0,8	0,5	0,5	0,8	0,69
	2003	0,7	0,9	0,8	0,2	0,7	0,9	0,8	0,5	0,5	0,8	0,63

Остання колонка табл. 5.9 демонструє рівні комплексного розвитку якості інтелектуального забезпечення діяльності промислових підприємств, які досліджувалися, вираженого за допомогою неметричних величин за наведеною шкалою перетворення. На основі порівняльної оцінки найгірший рівень розвитку якісних ознак виявлено на підприємств ЗАТ "Завод "Південкабель" (7) у 2001 році ($Y = 0{,}48$), а найвищий рівень ($Y = 0{,}74$) був зафіксований у 2003 році на

підприємстві ВАТ "Автрамат" (2). У цілому ж динаміка рівнів узагальнюючого показника вимірювання свідчить про розвиток якості інтелектуального забезпечення на кожному із семи промислових підприємств.

Розглянемо моделювання вимірників неметричних порядкових ознак виробничо-господарської діяльності промислового підприємства. Некількісна властивість об'єкта може виражатися за допомогою своїх ступенів інтенсивностей прояву. Залежно від рівнів ступенів прояву встановлюються їх оцінки – бали, і таким чином здійснюється вимірювання якості за ординальними шкалами. Встановлені бали (наприклад, 1, 2, 3, 4, 5) – це вираз номінацій якісної властивості об'єкта (у прикладі: дуже низька, низька, помірна, сильна, дуже сильна). Для встановлення балів для вираження якісної ознаки психологи, соціологи, економісти розробили різні методики. Але майже всі вони ґрунтуються на відомих дослідженнях Е. Вебера, Г. Фехнера, С. Стівена. У 1846 році Е. Вебер сформулював закон, який пов'язаний зі стимулом вимірюваної величини s [245, с. 67 – 69]. Закон Вебера стверджує, що зміна сприймання відзначається під час збільшення стимулу на сталу частку самого стимулу. Цей закон діє, коли Δs мале порівняно з s, але практично перестає діяти, коли s або дуже мале, або дуже велике. Т. Сааті вважає, що синтезування чи декомпозиція стимулів, що необхідно в кластерах або рівнях ієрархії, є ефективним засобом розширення застосування даного закону [245, с. 68]. У табл. 5.10 наведений опис шкали Т. Сааті.

Г. Фехнер встановив, що стимули із помітними відмінностями розміщуються в геометричній прогресії, а відповідні сприйняття становлять арифметичну прогресію в дискретних точках, де спостерігаються ледь помітні відмінності. На практиці якісні відмінності в реакціях на стимули не чисельні. Їх існує близько п'яти з додатковими, які становлять компроміси між сусідніми реакціями. Це збільшує кількість відмінностей до дев'яти, що узгоджується з передбаченнями про порядок величини [245, с. 69].

Таблиця 5.10

Шкала величин за Т. Сааті

Ступінь важливості	Визначення	Пояснення
1	2	3
1	Однакова значущість	Дві дії роблять однаковий внесок у досягнення мети
3	Певна перевага значущості однієї дії порівняно з іншою (слабка значущість)	Досвід і судження дають незначну перевагу однієї дії над іншою
5	Істотна або сильна значущість	Досвід і судження дають значну перевагу однієї дії над іншою
7	Дуже сильна або очевидна значущість	Перевага однієї дії над іншою дуже значна. Її зверхність практично явна
9	Абсолютна значущість	Свідчення на користь переваги однієї дії над іншою вищою мірою переконливі
2, 4, 6, 8	Проміжні значення між сусідніми значеннями шкали	Ситуація, коли необхідне компромісне рішення
Обернені величини наведених вище чисел	Якщо дії i порівняно з дією j приписується одне з наведених вище чисел, то дії j у процесі порівняння з i приписується обернене значення	Обґрунтоване припущення
Раціональні значення	Відношення, які виникають в заданій шкалі	Якщо постулювати узгодженість, то для отримання матриці потребується n числових значень

Для вимірювання порядкових ознак необхідно встановлені бали калібрувати за допомогою функції перетворення. Приклад калібрування ординальних ознак виконаємо для системи ознак (приклад у підрозділі 4.1). Балам потрібно встановити відповідне значення функції перетворення залежно від їх важливості та порівняльних відмінностей між собою. У табл. 5.11 наведені різні види функції.

Таблиця 5.11

Функція перетворення якісних порядкових ознак соціально-економічної системи

Значення ординальних величин (бали)	Види функції перетворення		
	1	2	3
1	0,05	0,00	0,00
2	0,15	0,20	0,25
3	0,50	0,50	0,50
4	0,85	0,80	0,75
5	0,95	1,00	1,00

Перелік видів функції перетворення можна продовжити. Присвоєння значень функції перетворення, як уже було сказано, відбувається на основі накопичених науково-практичних знань про дану ознаку. Перетворені значення неметричних порядкових ознак діяльності підприємства АО "ХарП" наведено в табл. 5.12 і табл. 5.13.

Таким чином, маємо динаміку вимірника якості ознак соціально-економічної системи. За змістом даних таблиць практично всі експерти вважають якість діяльності даного підприємства високою (найнижчий рівень якості $Y_{49} = 0{,}7030$), у цілому вимірник якості першого року дорівнює $Y = 0{,}8281$, у наступному $Y = 0{,}7447$, отже, є певне зниження рівня якості. Підприємству потрібно завчасно переглянути функціонування сфер його діяльності, щоб негативна тенденція до змін якісних ознак не переросла в аналогічну тенденцію до змін кількісних ознак діяльності.

Визначити якість ознак СЕС означає визначити сумісний прояв усіх його властивостей (характеристик), при цьому форма математичної моделі загальної якості соціально-економічної системи має бути мультиплікативною.

Отже, сумісний прояв кількісних і некількісних властивостей має виражатися мультиплікативно (узагальнено) метричною величиною початкових метричних і неметричних величин ознак, зведених до порівняльного вигляду шляхом перетворень. Перетворення різних величин до однієї міри здійснюється за допомогою функцій перетворень (метричних і неметричних), при цьому допускається різний вигляд функцій перетворення.

Узагальнюючу якість ознак соціально-економічної системи слід визначати за формулою:

$$I_y = \sqrt[3]{I_м \cdot I_{орд} \cdot I_{ном}}, \qquad (5.19)$$

де $I_м$ – вимірник метричних ознак;

$I_{орд}$ – вимірник ординальних ознак;

$I_{ном}$ – вимірник номінальних ознак;

або

$$I_у = \sqrt[N]{I_м^{n_1} \cdot I_{орд}^{n_2} \cdot I_{ном}^{n_3}}\ , \qquad (5.20)$$

де $N = n_1 + n_2 + n_3$, n_1, n_2, n_3 – вагові коефіцієнти важливості ознак (з цього приводу слід зауважити, що занижена роль неметричних ознак у характеристиці об'єкта пояснюється їх непізнанністю на даному етапі його вивчення та невідпрацьованістю математичного інструментарію для їх визначення).

Таблиця 5.12

Функції перетворення порядкових ознак діяльності АО "ХарП" (перший рік спостережень)

Експерти	Окремі функції перетворення														Y_{B_i}
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	
B_1	0,95	0,95	0,85	0,95	0,95	0,85	0,85	0,95	0,95	0,95	0,85	0,85	0,85	0,95	0,9058
B_2	0,85	0,85	0,50	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,50	0,85	0,85	0,95	0,7870
B_3	0,95	0,95	0,85	0,95	0,95	0,95	0,85	0,95	0,95	0,95	0,85	0,95	0,95	0,95	0,9276
B_5	0,85	0,95	0,95	0,95	0,95	0,85	0,85	0,95	0,95	0,95	0,95	0,95	0,85	0,85	0,9130
B_7	0,85	0,85	0,50	0,95	0,85	0,85	0,50	0,85	0,85	0,85	0,85	0,50	0,50	0,50	0,7089
$B8$	0,85	0,95	0,85	0,95	0,95	0,85	0,85	0,95	0,85	0,85	0,50	0,50	0,85	0,85	0,7089
B_9	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,85	0,85	0,85	0,95	0,9276
B_{10}	0,95	0,95	0,85	0,95	0,95	0,85	0,85	0,95	0,85	0,85	0,85	0,85	0,85	0,95	0,8915
B_{11}	0,85	0,95	0,85	0,95	0,95	0,50	0,50	0,50	0,95	0,85	0,50	0,85	0,50	0,85	0,7260
B_{15}	0,95	0,85	0,85	0,95	0,95	0,85	0,85	0,85	0,95	0,95	0,85	0,50	0,50	0,95	0,8264
B_{16}	0,85	0,85	0,50	0,95	0,95	0,85	0,85	0,95	0,85	0,85	0,85	0,50	0,85	0,95	0,8070
B_{22}	0,95	0,85	0,85	0,95	0,95	0,95	0,95	0,85	0,95	0,95	0,85	0,85	0,85	0,85	0,8986
B_{23}	0,85	0,85	0,50	0,95	0,95	0,50	0,50	0,95	0,95	0,95	0,50	0,85	0,85	0,95	0,7600
B_{24}	0,95	0,95	0,85	0,95	0,95	0,50	0,50	0,95	0,95	0,95	0,50	0,50	0,50	0,85	0,7435
B_{26}	0,95	0,95	0,85	0,95	0,95	0,85	0,95	0,95	0,95	0,95	0,85	0,85	0,50	0,95	0,8790
B_{28}	0,85	0,95	0,85	0,95	0,95	0,85	0,85	0,85	0,85	0,85	0,85	0,50	0,50	0,85	0,8070
B_{35}	0,95	0,95	0,95	0,95	0,95	0,85	0,85	0,95	0,95	0,95	0,85	0,95	0,95	0,95	0,9276
B_{37}	0,85	0,85	0,85	0,85	0,95	0,85	0,85	0,85	0,85	0,85	0,50	0,50	0,50	0,85	0,7647
B_{40}	0,95	0,95	0,85	0,95	0,95	0,95	0,95	0,95	0,85	0,85	0,50	0,50	0,50	0,50	0,7722
B_{44}	0,95	0,95	0,95	0,95	0,95	0,85	0,50	0,50	0,95	0,95	0,95	0,85	0,85	0,50	0,8084
B_{45}	0,95	0,85	0,85	0,85	0,95	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,8636
B_{46}	0,95	0,85	0,85	0,85	0,95	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,8636
B_{49}	0,95	0,85	0,50	0,95	0,95	0,15	0,50	0,95	0,95	0,95	0,85	0,85	0,50	0,85	0,7030
B_{50}	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,95	0,95	0,95	0,85	0,85	0,95	0,85	0,8774

Таблиця 5.13

Функції перетворення порядкових ознак діяльності АО "ХарП" (другий рік спостережень)

Експерти	Окремі функції перетворення														Узагальнююча функція Y_{B_i}
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	
B_1	0,50	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,95	0,95	0,95	0,95	0,85	0,85	0,8448
B_2	0,85	0,85	0,95	0,95	0,85	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,9276
B_3	0,15	0,85	0,50	0,50	0,15	0,85	0,85	0,50	0,85	0,85	0,85	0,85	0,85	0,85	0,5921
B_5	0,85	0,95	0,85	0,85	0,85	0,95	0,85	0,95	0,85	0,95	0,95	0,85	0,95	0,95	0,8986
B_7	0,85	0,95	0,95	0,50	0,95	0,15	0,85	0,15	0,95	0,95	0,95	0,95	0,95	0,95	0,6861
$B8$	0,85	0,95	0,85	0,50	0,50	0,85	0,50	0,50	0,50	0,85	0,85	0,85	0,85	0,85	0,7089
B_9	0,85	0,50	0,85	0,85	0,85	0,85	0,50	0,85	0,85	0,85	0,85	0,95	0,85	0,50	0,7647
B_{10}	0,85	0,95	0,95	0,95	0,85	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,9350
B_{11}	0,85	0,85	0,50	0,50	0,85	0,85	0,85	0,50	0,85	0,85	0,85	0,95	0,85	0,85	0,7647
B_{15}	0,50	0,85	0,85	0,85	0,50	0,85	0,85	0,85	0,95	0,95	0,85	0,95	0,85	0,85	0,8070
B_{16}	0,85	0,95	0,50	0,95	0,95	0,85	0,85	0,50	0,95	0,95	0,95	0,95	0,95	0,95	0,8464
B_{22}	0,85	0,85	0,50	0,50	0,50	0,85	0,85	0,85	0,50	0,85	0,50	0,85	0,85	0,85	0,7033
B_{23}	0,85	0,85	0,85	0,85	0,50	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,8184
B_{24}	0,85	0,85	0,85	0,85	0,95	0,85	0,85	0,85	0,95	0,95	0,50	0,50	0,50	0,85	0,7769
B_{26}	0,85	0,50	0,05	0,85	0,85	0,50	0,85	0,85	0,85	0,95	0,50	0,50	0,85	0,85	0,6014
B_{28}	0,85	0,50	0,85	0,85	0,50	0,50	0,50	0,50	0,85	0,85	0,50	0,50	0,85	0,95	0,6571
B_{35}	0,85	0,85	0,85	0,85	0,50	0,85	0,85	0,95	0,95	0,95	0,05	0,95	0,95	0,95	0,7011
B_{37}	0,95	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,50	0,50	0,50	0,85	0,7647
B_{40}	0,95	0,50	0,50	0,85	0,05	0,95	0,95	0,95	0,95	0,95	0,85	0,85	0,85	0,85	0,6750
B_{44}	0,85	0,50	0,50	0,85	0,15	0,85	0,85	0,95	0,85	0,85	0,05	0,05	0,85	0,50	0,4507
B_{45}	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,50	0,50	0,50	0,85	0,7586
B_{46}	0,85	0,85	0,85	0,95	0,50	0,85	0,95	0,95	0,95	0,95	0,50	0,50	0,85	0,85	0,7894
B_{49}	0,95	0,50	0,50	0,95	0,50	0,85	0,85	0,95	0,95	0,95	0,50	0,50	0,85	0,95	0,7376
B_{50}	0,85	0,85	0,85	0,85	0,85	0,85	0,85	0,95	0,95	0,95	0,85	0,85	0,95	0,85	0,8774

Окремо обчислений вимірник складних номінальних ознак поєднується з вимірником складних порядкових ознак та з вимірником метричних ознак. Форма поєднання мультиплікативна.

За допомогою моделей вимірників можна архітектонічно визначати всі ознаки, що є основними властивостями об'єкта в економіці: метричні й неметричні. Відмінність моделювання вимірників ознак забезпечується не тільки поєднанням, комплексністю визначення метричних і неметричних величин, але й можливістю просторово-часового порівняння. Існуючі спроби побудови оцінки якості соціально-економічних систем в економіці ґрунтуються на поєднанні величин, що одержані різними шляхами. Звичайно, розглядаючи такі пропозиції, як перші наближення в дослідженні системи, принципово відхиляти їх недоцільно. Але всі методики обмежуються синтезом метричних або в окремих випадках, порядкових ознак, не враховуючи номінальних ознак об'єкта.

Визначення якісних ознак у системі з кількісними інформаційно доповнює моделі СЕС, забезпечуючи повномасштабність в описі, а моделі вимірників номінальних ознак соціально-економічної системи забезпечують фундаментальну аналітичну основу аналізу в управлінні якісними характеристиками систем.

У свою чергу, запропонована система вимірників спроможна адекватно визначати величини в економіці, відображати зміни величин основних ознак, властивостей та комплексно, цілісно діагностувати, оцінювати, контролювати зміну самої структури властивостей СЕС мікрорівнів і їх узгодженого підпорядкування СЕС макрорівнів.

Задачі до розділу 5

За даними, що характеризують виробничо-господаську діяльність 15 промислових підприємств протягом 5 років (додаток А), розв'язати задачі багатовимірного аналізу соціально-економічних систем (умовне позначення табл. 3.6) 1) побудувати вимірники для елементарних ознак виробничо-господарської діяльності підприємств; 2) побудувати вимірники для складних ознак виробничо-господарської діяльності підприємств; 3) визначити рівень загальної якості ознак виробничо-господарської діяльності підприємств; 4) провести аналіз вимірників виробничо-господарської діяльності підприємств та сформулювати пропозиції щодо підвищення рівня розвитку складних ознак діяльності.

Запитання для самоперевірки

1. Які бувають види узагальнюючих показників в економіці?
2. Які принципові математичні відмінності в логіці обчислення узагальнюючих показників в економіці?
3. Як обчислювальні проблеми вирішуються при побудові таксономічного показника розвитку?
4. Які основні проблеми вирішуються в побудові вимірників метричних ознак соціально-економічних систем?
5. Які типи функцій перетворень доцільно використовувати при побудові вимірників метричних ознак?
6. Які типи задач розрізняються в побудові окремих функцій перетворень?
7. Навести основні етапи аналізу складних ознак соціально-економічних систем на основі вимірників.
8. Які рекомендується використовувати функції перетворення для неметричних ознак соціально-економічних систем?
9. Навести основні етапи аналізу складних неметричних ознак соціально-економічних систем на основі вимірників.
10. Як моделюється загальна якість ознак соціально-економічних систем?

Розділ 6
Аналіз соціально-економічних систем у скороченому просторі різних ознак та вимірників

6.1. Аналіз складних сумісних ознак соціально-економічних систем на основі факторного аналізу

Факторний аналіз успішно застосовується для розв'язання багатьох практичних задач функціонування та розвитку соціально-економічних систем, зокрема підприємств, при цьому початковими даними є значення метричних ознак або, інакше кажучи, значення кількісних показників. Проте в цьому полягає обмеження застосувань факторного аналізу. Для адекватного моделювання складних соціально-економічних систем передбачається їх опис за допомогою різних ознак: метричних і неметричних, які сумісно вимірюються на номінальних та порядкових шкалах. Проблема застосування стандартного факторного аналізу в даній ситуації полягає в тому, що початкові дані представлені матрицею різних величин, а отже, коефіцієнт кореляції, на основі якого відбуваються всі обчислення методу, некоректно використовувати [125; 137; 150; 169].

Рекомендується модифікувати факторний аналіз за рахунок вихідної матриці розроблених коефіцієнтів *ССК.* Обґрунтування вирішення проблеми модифікації факторного аналізу для ознак, величини яких виміряні на різних шкалах, розглянемо на прикладі розв'язання практичних економічних задач.

Задача 6.1. Провести аналіз даних діяльності 7 підприємств за 3 роки, що елементарно характеризуються 6 метричними ознаками виробничо-господарської діяльності ($x_{16}, x_{17}, x_{18}, x_{34}, x_{35}, x_{38}$ в табл. 4.19) і 4 неметричними ознаками інтелектуальної діяльності (C, D, E, F в табл. 4.19).

Розв'язання задачі. Задачу можна розв'язати за допомогою модифікації факторного аналізу [140].

Розглянемо декілька варіантів реалізації факторного аналізу згідно з початковими даними задачі.

Факторний аналіз з матрицею коефіцієнтів кореляції R_{xy}. Виконана процедура факторного аналізу, що ґрунтується на матриці R_{xy}. Проаналізуємо всі проміжні обчислення. У табл. 6.1 у 1-й колонці наведені власні числа кореляційної матриці R_{xy} (у порядку спадання), у 2-й колонці – накопичена сума цих чисел у процентах. Сума перших 3-х власних чисел становить 78,28%, тобто 3 головні компоненти описують майже всю мінливість 10-ти початкових ознак. У колонках табл. 6.1 наведена матриця факторних навантажень.

Таблиця 6.1

Власні числа, факторна матриця, спільності

EVALS	*Cum%*	*Vars*	*FMAT*	*FMAT*	*FMAT*	R^2
4,14244	41,42%	*X*	0,925647	-0,13600	-0,27654	0,95179
2,37754	65,20%	*Y*	-0,88210	0,127287	0,304413	0,88697
1,30791	78,28%	*Z*	0,898138	-0,11097	-0,26494	0,889162
0,761996	85,90%	*U*	0,411023	*0,676666*	*0,549488*	0,928755
0,581035	91,71%	*V*	-0,45900	*-0,67606*	*-0,50283*	0,920564
0,429219	96,00%	*W*	0,382099	0,78409	-0,32851	0,868713
0,164547	97,65%	*C*	0,417273	-0,45354	0,299608	0,469578
0,147228	99,12%	*D*	0,194087	*-0,72050*	*0,521984*	0,829253
0,077372	99,89%	*E*	0,67903	-0,27056	0,148674	0,556386
0,010722	100,00%	*F*	0,708786	-0,05275	0,146792	0,526708
Усього: 10		СумКв=	4,142436	2,377538	1,307905	

У підрозділі 2.4 доводилось, що факторні навантаження – це β-коефіцієнти розкладу стандартизованих ознак ($X, Y, Z, U, V, W, C, D, E, F$) за системою (стандартизованих) головних компонент ($\varPhi_1, \varPhi_2, \varPhi_3, \varPhi_4$). Суми квадратів факторних навантажень у стовпчиках матриці *FMAT* дорівнюють власним числам та показують внесок кожної компоненти в загальну мінливість. Суми квадратів факторних навантажень у рядках матриці *FMAT* дорівнюють спільностям – коефіцієнтам детермінації R^2, які показують частину мінливості кожної ознаки, що пояснюється 3-ма головними компонентами.

Від головних компонент можна перейти до системи головних факторів, які зберігають усі позитивні риси компонент, але мають чіткішу матрицю факторних навантажень: ортогональним перетворенням (обертанням) досягається стан, коли складність кожної ознаки дорівнює одиниці, щоб у кожному рядку було тільки одне високе навантаження. У матриці *FMAT* три ознаки мають більшу складність.

У табл. 6.2 наведені розклади початкових ознак (стандартизованих) за системою (стандартизованих) головних факторів (F_1, F_2, F_3).

Таблиця 6.2

Факторна матриця після обертання

Vars	*RFMAT*	*RFMAT*	*RFMAT*	R^2
X	0,97297	0,065731	0,028272	0,95179
Y	-0,94104	-0,03751	0,002279	0,88697
Z	0,939449	0,079935	0,01441	0,889162
U	0,077649	0,957896	-0,07184	0,928755
V	-0,13827	-0,94449	0,09684	0,920564
W	0,338793	0,446065	-0,74496	0,868713
C	0,355164	0,045561	0,584261	0,469578
D	0,115426	-0,06806	0,900721	0,829253
E	0,619392	0,163961	0,381911	0,556386
F	0,611899	0,318394	0,225639	0,526708
СумКв=	3,752252	2,15569	1,919937	
Усього	7,827879			

Нова матриця факторних навантажень *RFMAT* отримана перетворенням *FMAT* за допомогою матриці обертання *TRMAT*, яка наведена в табл. 6.3.

Таблиця 6.3

Матриця обертання

TRMAT	*TRMAT*	*TRMAT*
0,923176	0,363604	0,124649
-0,16353	0,665013	-0,72871
-0,34785	0,652342	0,673385

Виділені наступні групи ознак: (X, Y, Z), (U, V), (W, D). Усі ці ознаки відтворюються трьома факторами практично точно. Ознаки C, E, F, що залишились у системі, відтворюються не настільки точно, тому в разі зміни кількості компонент висновки про їх належність до конкретної групи можуть змінитися. Ознаки (E, F) віднесені до першої компоненти, а ознака (C) – до третьої.

Факторний аналіз з матрицею коефіцієнтів рангової кореляції ρ_s. У підрозділі 3.1 доводилось, що коефіцієнти рангової кореляції Спірмена (з урахуванням зв'язаних рангів) дорівнюють звичайним коефіцієнтам парної кореляції, обчисленим за ранжованими величинами, та мало відрізняються від коефіцієнтів парної кореляції R_{xy}, обчислених за початковими даними табл. 4.19, тому й результати факторного аналізу з матрицею ρ_s будуть мало відрізнятися від факторного аналізу з матрицею R_{xy}. Результати цього аналізу повністю співпадають з факторним аналізом з ранжованими початковими даними.

Зауваження. Ранги змінюються від 1 до 21. У табл. 6.4 у першому стовпчику наведені власні числа кореляційної матриці ρ_s (у порядку спадання), у другому стовпчику – накопичена сума цих чисел у процентах. Сума перших 3-х власних чисел становить 78,76%, тобто 3 головні компоненти описують майже всю мінливість 10-ти ранжованих ознак (порівнюючи з величиною 78,28% під час обчислень з матрицею R_{xy}). У стовпчиках *FMAT* табл. 6.4 наведена матриця факторних навантажень.

Зміст табл. 6.4 мало відрізняється від змісту табл. 6.2. У табл. 6.5 наведений розклад ранжованих величин ознак (стандартизованих) за системою головних факторів (F_1, F_2, F_3) після обертання за методом *VARIMAX*.

Таблиця 6.4

Власні числа, факторна матриця, спільності

EVALS	*Cum%*	*Vars*	*FMAT*	*FMAT*	*FMAT*	R^2
4,25628	42,56%	*X*	0,915314	0,159406	-0,27703	0,939955
2,35027	66,07%	*Y*	-0,88774	-0,12257	0,373798	0,942821
1,26969	78,76%	*Z*	0,889751	0,15239	-0,26943	0,887473
0,737332	86,14%	*U*	0,589654	-0,6197	0,482	0,964041
0,611748	92,25%	*V*	-0,45939	0,682188	-0,52656	0,953681
0,475823	97,01%	*W*	0,33163	-0,79003	-0,30978	0,830097
0,172525	98,74%	*C*	0,424637	0,452609	0,260863	0,45322
0,069406	99,43%	*D*	0,206352	0,73394	0,520305	0,851965
0,042007	99,85%	*E*	0,670567	0,259564	0,149297	0,539323
0,014915	100,00%	*F*	0,705326	0,046495	0,118395	0,513665
Усього: 10		*СумКв=*	4,256284	2,350272	1,269689	
		Усього	7,876245			

Виділені наступні групи ознак: (X, Y, Z, E, F), (U, V), (W, C, D). Ці результати повністю повторюють висновки факторного аналізу з кореляційною матрицею R_{xy}.

Таблиця 6.5

Факторна матриця після обертання

Vars	*RFMAT*	*RFMAT*	*RFMAT*	R^2
X	0,962031	0,105687	0,057288	0,939955
Y	-0,96869	-0,0555	0,037151	0,942821
Z	0,934727	0,104289	0,053691	0,887473
U	0,218671	0,954805	-0,06761	0,964041
V	-0,07209	-0,96854	0,102012	0,953681
W	0,269251	0,446607	-0,74709	0,830097
C	0,364116	0,056678	0,563407	0,45322
D	0,119046	-0,0488	0,914009	0,851965
E	0,592133	0,212249	0,379015	0,539323
F	0,594754	0,34347	0,204843	0,513665
СумКв=	3,714194	2,242929	1,919119	
Усього	7,876243			

Факторний аналіз з матрицею коефіцієнтів CCK. У підрозділі 3.3 доводилось, що коефіцієнти CCK є універсальною мірою тісноти зв'язку та можуть обчислюватися як для метричних, так і для неметричних ознак, а також їх сумісних систем [161]. Обговоримо результати факторного аналізу, якщо на вході задати матрицю коефіцієнтів CCK. У табл. 6.6 наведені власні числа цієї матриці, факторна матриця та спільності.

Коефіцієнти CCK систематично більші від відповідних коефіцієнтів кореляції (для випадків, коли можна обчислити R_{xy}), через це найменше власне число виявилось від'ємним. Проте малі власні числа, як відомо, необхідно відкинути, тому слід продовжити аналіз. Одна перша головна компонента пояснює 71,95% всієї мінливості. Цього було б достатньо, проте перша головна компонента є генеральним фактором, вона містить високі факторні навантаження для всіх ознак. Слід виділити перші дві компоненти (матриця *FMAT*).

Таблиця 6.6

Власні числа, факторна матриця, спільності

EVALS	*Cum%*	*Var*	*FMAT1*	*FMAT2*	*FMAT3*	R^2
7,585934	71,95%	*X*	0,901958	-0,19952	0,151206	0,876201
0,837547	79,90%	*Y*	0,855424	-0,16021	0,182764	0,790821
0,656160	86,12%	*Z*	0,919051	-0,30639	-0,07086	0,943554
0,458137	90,46%	*U*	0,953681	0,18007	0,072283	0,947157
0,365343	93,93%	*V*	0,90164	0,280394	-0,00927	0,891661
0,261537	96,41%	*W*	0,920069	0,085135	-0,11873	0,867872
0,205944	98,36%	*C*	*0,795513*	*-0,51872*	-0,01827	0,902241
0,156764	99,85%	*D*	*0,789575*	-0,04617	*-0,57656*	0,957986
0,015717	100,00%	*E*	0,820139	0,134296	0,477628	0,918792
-0,54308	100,00%	*F*	*0,835531*	*0,520255*	-0,12073	0,983355
СумКв=			7,585934	0,837547	0,65616	
		Усього	9,079641			

Суми квадратів факторних навантажень у стовпчиках матриці *FMAT* дорівнюють власним числам, відповідно, 7,586 та 0,838, 0,656, тобто перша компонента пояснює 71,95% повної мінливості, а друга і третя – всього 14,17%. Суми квадратів факторних навантажень у рядках матриці *FMAT* дорівнюють спільностям R^2. У табл. 6.7 наведена факторна матриця *RFMAT* після обертання.

Таблиця 6.7

Факторна матриця після обертання

Var	*RFMAT*	*RFMAT*	*RFMAT*	R^2
X	***0,746844***	***0,51253***	0,236091	0,876201
Y	***0,695012***	***0,520545***	0,191868	0,790821
Z	0,789997	0,366285	0,430458	0,943554
U	0,477792	0,762804	0,370138	0,947157
V	0,352509	0,762542	0,431193	0,891661
W	0,486909	*0,603802*	*0,515961*	0,867872
C	0,883052	0,168445	0,306734	0,902241
D	0,405358	0,253267	0,854123	0,957986
E	*0,515451*	*0,806259*	-0,05522	0,918792
F	0,107411	***0,832536***	***0,527922***	0,983355
СумКв	3,475864	3,631423	1,972354	
Усього	9,079641			

За найбільшими факторними навантаженнями маємо наступні групи ознак: (X, Y, Z, C), (U, V, W, E, F), (D). Порівнюючи їх з групами факторного аналізу з матрицею R_{xy}: (X, Y, Z), (U, V), (W, D), маємо майже ті ж результати, але ознака W перенесена з третьої групи в другу, хоча її можна було б віднести до третьої групи (табл. 6.8). Отже, ці висновки не суперечать висновкам факторного аналізу з кореляційною матрицею. Ще раз підкреслюємо, що коефіцієнти CCK є універсальною мірою тісноти зв'язку між різними типами ознак у системі: метричними, неметричними, метрично-неметричними, тоді коли коефіцієнт R_{xy} обчислити не можливо.

Завдяки модифікованому факторному аналізу можна розбудувати модель складних ознак, вихідними даними якої є величини, виміряні на метричних і порядкових шкалах [140].

Задача 6.2. Для аналізу соціально-економічного стану 48 країн світу необхідно визначити складні ознаки соціально-економічних систем, що елементарно характеризуються 16 соціально-економічними показниками, а саме 5 показниками, що описують стан населення, праці, суспільства (IMD); 4 показниками стану науки й технології (IMD); 5 показниками стану макроекономіки (IMD); 2 показниками стану фінансів (IMD) (дані IACCEП ЦЕМІ РАН [260]).

Розв'язання задачі. Позначимо x_1 – якість життя (порядкові величини за 10-бальною шкалою); x_2 – економічна грамотність (порядкові величини за 10-бальною шкалою); x_3 – "відплив умів" (порядкові величини за 10-бальною шкалою); x_4 – частка безробіття в загальній робочій силі (метричні величини, %%); x_5 – частка зайнятих у загальному населенні (метричні величини, %%); x_6 – капіталізація фондового ринку (метричні величини, млн дол.); x_7 – банківські збереження (метричні величини, %% від ВВП); x_8 – випуск валового продукту (ВВП) на одну особу населення (метричні величини, дол. на особу); x_9 – частка промисловості в ВВП (метричні величини, %% від ВВП); x_{10} – частка сільськогосподарської продукції (метричні величини, %% до ВВП); x_{11} – зростання валових внутрішніх збережень (метричні величини, %% від реальних змін); x_{12} – порівняльна вартість життя (метричні величини, %% порівняно з Нью-Йорком); x_{13} – фінансування розвитку технологій (порядкові величини за 10-бальною шкалою); x_{14} – фундаментальні дослідження (порядкові величини за 10-бальною шкалою); x_{15} – зацікавленість молоді в науці і технології (порядкові величини за 10-бальною шкалою); x_{16} – загальні витрати на НДР (метричні величини, млн дол.). За допомогою модифікованого факторного аналізу з матрицею коефіцієнтів ρ_s отримана матриця факторних навантажень, яка після процедури *VARIMAX* має значення, що містяться в табл. 6.8.

За значеннями коефіцієнтів факторних навантажень a_{ij} на i-ті показники встановлені складні ознаки за пріоритетністю, а саме: перша загальна ознака рівня розвитку економіки й технології (O_1), друга – стан ринку фінансових послуг (O_2), третя – сприятливість умов життя (O_3), четверта – інтенсивність руху капіталу у фінансовій формі (O_4), п'ята – наявність умов для розвитку інновацій (O_5).

Таблиця 6.8

Факторна матриця після обертання *VARIMAX* (вся сукупність країн)

	O_1	O_2	O_3	O_4	O_5	R^2
x_1	0,876683531	0,07381311	-0,222768	-0,00226214	-0,057526402	0,826962
x_2	0,868110052	-0,1721374	0,041581	-0,041379797	0,212415459	0,831808
x_3	0,762777989	-0,0147734	-0,073721	0,139259004	0,19553366	0,645110
x_4	-0,58519002	0,02747812	-0,361541	0,418159911	0,446263693	0,847924
x_5	0,606642043	0,07040975	0,224017	-0,292834642	-0,523159677	0,782604
x_6	0,36782316	0,89442292	-0,033771	-0,085615768	-0,071656683	0,948891
x_7	0,715501937	0,2321059	0,119164	-0,271192453	-0,066241976	0,657950
x_8	0,862450256	0,09777186	-0,041668	0,284499651	-0,249994133	0,898553
x_9	-0,421933618	-0,0648467	0,647047	-0,238470553	0,188609872	0,693344
x_{10}	-0,721171519	-0,2014603	-0,07233	-0,419771114	0,129309138	0,758835
x_{11}	-0,281510584	0,01407981	-0,046537	-0,843087312	0,016632943	0,792685
x_{12}	0,458206153	0,20060847	0,707795	0,246278278	-0,081553666	0,818475
x_{13}	0,897131952	0,14741147	-0,026241	-0,029290488	0,248700624	0,889974
x_{14}	0,802222085	0,30445476	0,108456	0,040560001	0,235887632	0,805304
x_{15}	0,302043834	-0,138531	0,104071	-0,103110329	0,811730761	0,790791
x_{16}	0,387033269	0,85047976	0,11081	0,107817241	-0,049059436	0,899421

За виявленими п'ятьма загальними складними соціально-економічними ознаками країн слід встановити однорідність даної сукупності. На рис. 6.1 наведена дендрограма сукупності країн світу, отримана завдяки реалізації кластерного аналізу за методом Уорда.

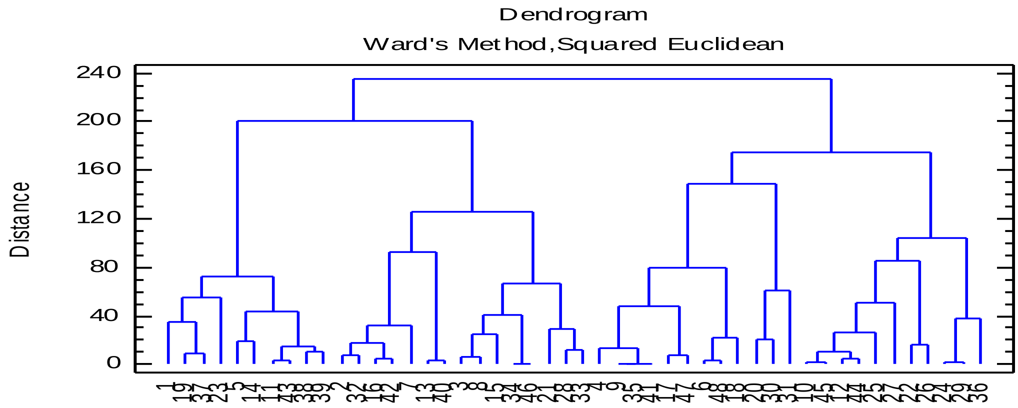


Рис. 6.1 . **Дендрограма сукупності країн світу за складними соціально-економічними ознаками їх стану**

За визначеними за моделлю 5 складними ознаками розрізняються 12 груп однорідних за своїм соціально-економічним станом країн. У табл. 6.9 наведений рейтинг груп країн за визначеними складними соціально-економічними ознаками.

Таблиця 6. 9

Склад груп однорідних країн за критерієм загальних складних соціально-економічних ознак

Країна	Група	Країна	Група	Країна	Група
Австрія	1	Швеція	4	Індія	9
Данія	1	Угорщина	5	Туреччина	9
Гонконг	1	Ізраїль	5	Бразилія	10
Швейцарія	1	Фінляндія	5	Мексика	10
Бельгія	2	Ірландія	6	Іспанія	10
Канада	2	Чехія	6	ПАР	10
Німеччина	2	Норвегія	6	Словенія	11
США	2	Люксембург	7	Естонія	11
Франція	2	Нова Зеландія	7	Філіппіни	11
Малайзія	3	Португалія	7	Чилі	11
Сінгапур	3	Таїланд	7	Аргентина	12
Тайвань	3	Китай	8	Венесуела	12
Великобританія	4	Корея	8	Індонезія	12
Японія	4	Росія	8	Колумбія	12
Італія	4	Австралія	9	Словаччина	12
Нідерланди	4	Греція	9	Польща	12

Далі дослідження складних соціально-економічних ознак слід продовжити визначенням їх у кожній однорідній групі країн за допомогою модифікованого факторного аналізу, коли вхідними даними є метричні та неметричні величини. У табл. 6.10 наведені виявлені складні ознаки в найбільшій за чисельністю групі країн – групі 2.

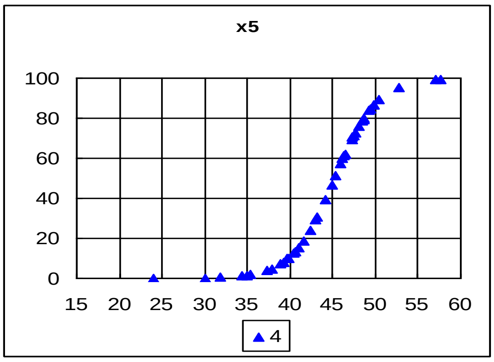
Таблиця 6. 10

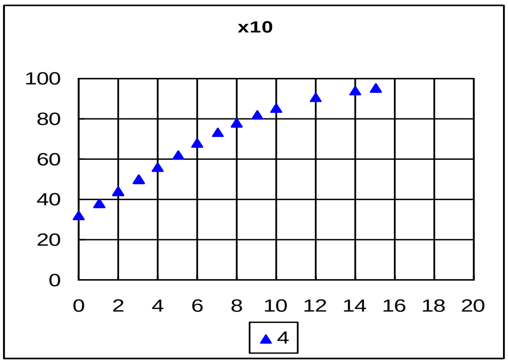
Факторна матриця після обертання *VARIMAX* (країни групи 2)

	O_1	O_2	O_3	R^2
x_1	0,92958	-0,1727	-0,1768	0,9252
x_2	0,91406	-0,1062	-0,2002	0,88687
x_3	0,9104	0,14608	0,05817	0,85355
x_4	-0,6957	*0,5577*	-0,1314	0,81225
x_5	*0,5814*	*-0,6781*	-0,0252	0,79844
x_6	0,72877	0,20106	0,37459	0,71184
x_7	0,91011	-0,2002	0,07304	0,87372
x_8	0,89712	-0,1328	-0,0097	0,82255
x_9	-0,006	-0,2376	0,93429	0,92937
x_{10}	-0,8129	0,06076	0,02617	0,66512
x_{11}	-0,1568	-0,8308	0,3322	0,82512
x_{12}	0,79056	-0,0158	0,31111	0,72202
x_{13}	0,9447	0,1736	0,00204	0,92259
x_{14}	0,95222	0,16838	-0,0576	0,93839
x_{15}	0,75448	-0,0096	-0,3702	0,70636
x_{16}	0,86155	0,23702	0,14462	0,81936
λ	9,93568	1,79673	1,48037	

У країнах кластера 2 система складних соціально-економічних ознак дещо змінилася. Згідно з факторними навантаженнями на показники виявлено, що перша ознака – ефективність соціально-економічних процесів у країнах (O_1), друга – несприятливість умов для формування збережень (O_2), третя – рівень розвитку промисловості (O_3).

Дотримуючись рекомендацій у розбудові складних неметричних ознак соціально-економічних систем, значення кожного із 16-ти соціально-економічних показників були перетворені у вимірники за функціями перетворення з урахуванням закономірностей розвитку окремої елементарної ознаки, що виражає даний показник. На рис. 6.2 зображена функція перетворення x_5 – частки зайнятих у загальному населенні (а) й x_{10} – частки сільськогосподарської продукції (б).





а б

Рис. 6.2. **Функції перетворення показників частки зайнятих у загальному населенні (x_5) (а) й частки сільськогосподарської продукції у процентах від ВВП (x_{10}) (б)**

Були обчислені значення вимірників, що визначають соціально-економічні ознаки розвитку кожної країни в динаміці. За визначеними окремими функціями перетворення соціально-економічних ознак усіх країн були побудовані графіки динаміки значень вимірників за чотирма складовими соціально-економічного розвитку Росії, що є її складними ознаками, сформованими за змістом (рис. 6.3): населення, праці, суспільства (а); науки й технологій (б); макроекономіки (в); фінансів (г).

Отже, робимо наступний висновок: у Росії надзвичайно низький рівень життя (x_1) порівняно з іншими країнами світу, динаміка ж "відпливу умів" (x_3) незначна (закономірна тенденція даного вимірника спадна, а отже, високі значення свідчать про відмінний стан даної ознаки). Виокремлюється тенденція збільшення зайнятих у загальному населенні (x_5). За такого стану соціально-економічних ознак значення складної ознаки населення, праці, суспільства в Росії не перевищує рівня 0,2 за шкалою вимірювання даних величин ознак у світі. Зі змісту рис. 6.3б бачимо, що в Росії низький рівень фінансування розвитку технологій (x_{13}).

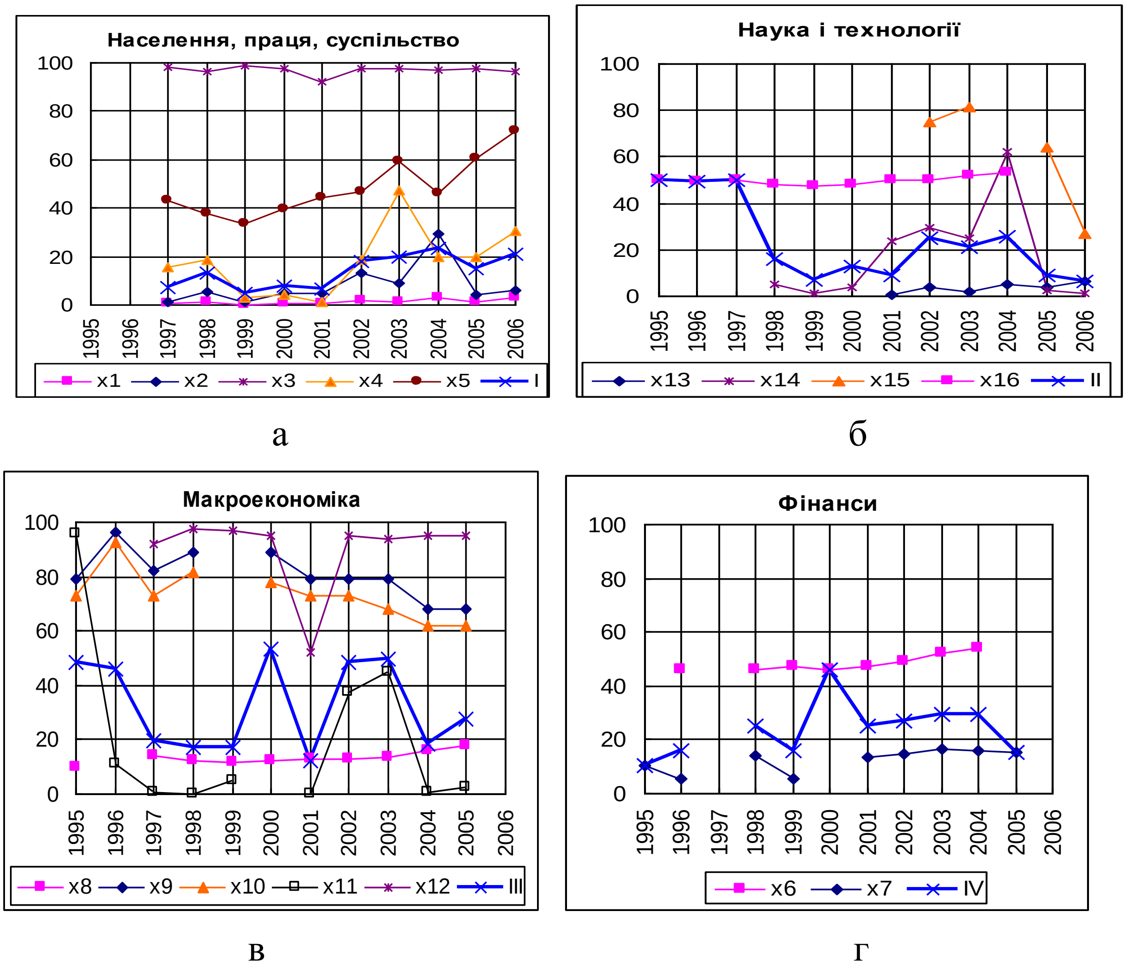


Рис. 6.3. **Динаміка вимірників за чотирма складними ознаками: населення, праці, суспільства (а); науки й технологій (б); макроекономіки (в); фінансів (г), що описують соціально-економічний розвиток Росії**

Слід зауважити, що позитивні зрушення у фундаментальних дослідженнях (x_{14}) у 2004 році мають різкий спад в наступні роки, такі негаразди протистоять добрій динаміці загальних витрат на НДР (x_{16}). У цілому ж складна ознака науки і техніки набагато більше розвинута, ніж ознака населення, праці, суспільства. Складна ознака макроекономіки (див. рис. 6.3в) характеризується різкими змінами: зростання частки промисловості ВВП (x_9) супроводжується зростанням порівняльної вартості життя (x_{12}), при цьому наявні низькі значення випуску ВВП на одну особу (x_8). Величини вимірників фінансової складної ознаки соціально-економічного стану Росії за останні роки надзвичайно погіршилися навіть на фоні зростання капіталізації фондового ринку (x_6). На рис. 6.4 зображена динаміка складових соціально-економічного розвитку Росії.

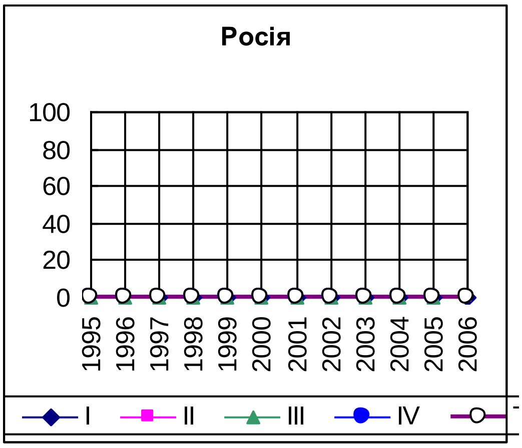
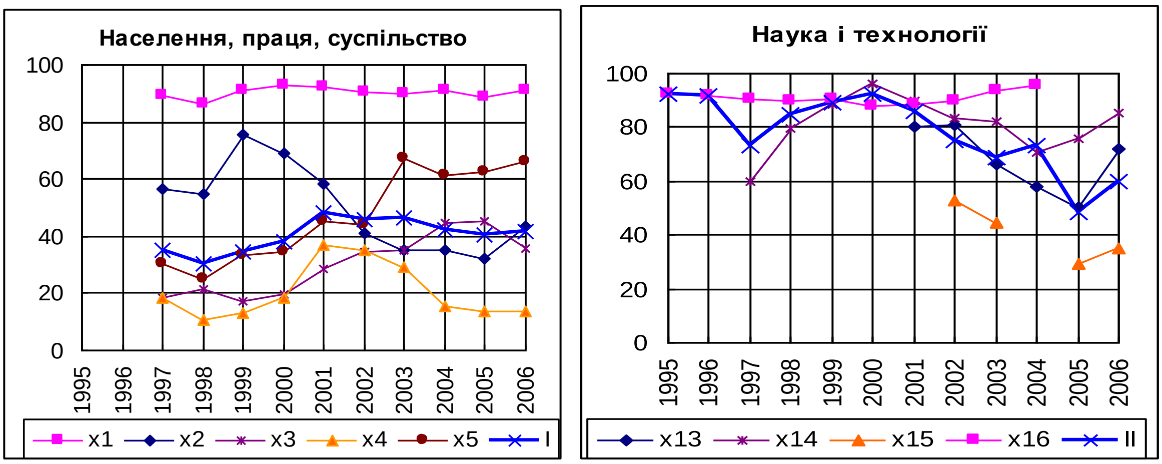


Рис. 6.4. **Динаміка вимірників чотирьох складних ознак: населення, праці, суспільства (I); науки й технологій (II); макроекономіки (III); фінансів (IV) та їх загального стану (Total), що описують соціально-економічний розвиток Росії**

З рис. 6.4 видно, що спостерігається низький рівень розвитку складної ознаки населення, праці, суспільства (I), хоча стан розвитку макроекономіки (III) дещо покращувався, але загальний соціально-економічний стан Росії потребує явного покращення.

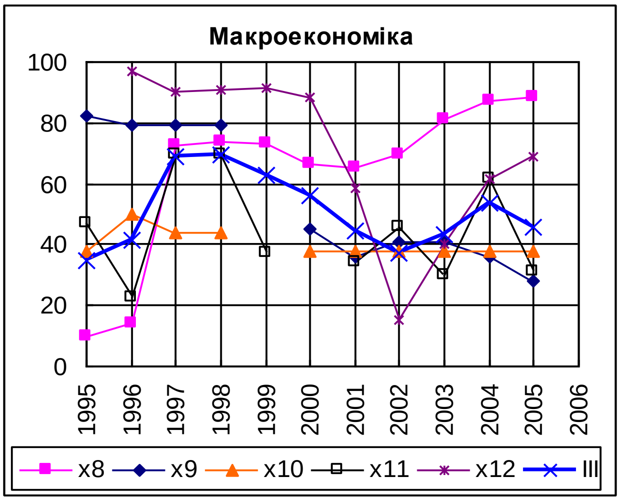
На рис. 6.5 наведена динаміка вимірників соціально-економічних ознак Німеччини.

Зі змісту рис. 6.5 випливає висновок, що стан соціально-економічних ознак Німеччини добрий, незважаючи на окремі різкі погіршення деяких ознак, як, наприклад, зростання частки безробітних у загальній робочій силі в 2001 році (x_3) (див. рис. 5.5а) або різке падіння порівняльної вартості життя (у %% порівняно з Нью-Йорком) у 2002 (див. рис. 6.5в).

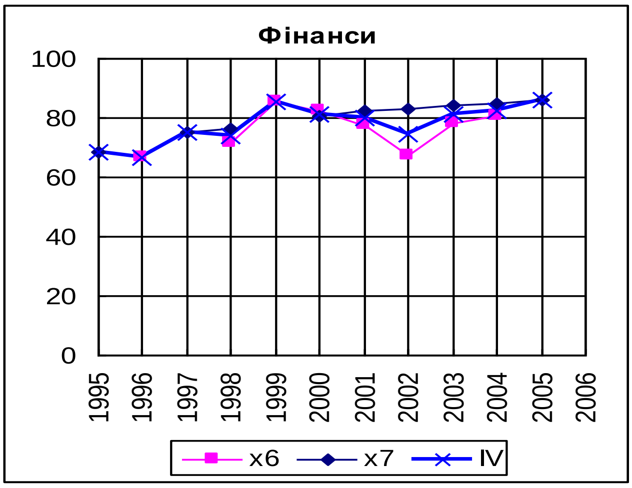


а б

Рис. 6.5. **Динаміка вимірників чотирьох складних ознак: населення, праці, суспільства (а); науки й технологій (б); макроекономіки (в); фінансів (г), що описують соціально-економічний розвиток Німеччини**



в



г

Закінчення рис. 6.5.

У цілому ж динаміка вимірників складних ознак соціально-економічного розвитку Німеччини (рис. 6.6) свідчить про закономірність достатньо високого рівня, при цьому найрозвиненішою є складна ознака науки й технологій (II) на фоні не дуже високого розвитку складної ознаки населення, праці, суспільства (I).

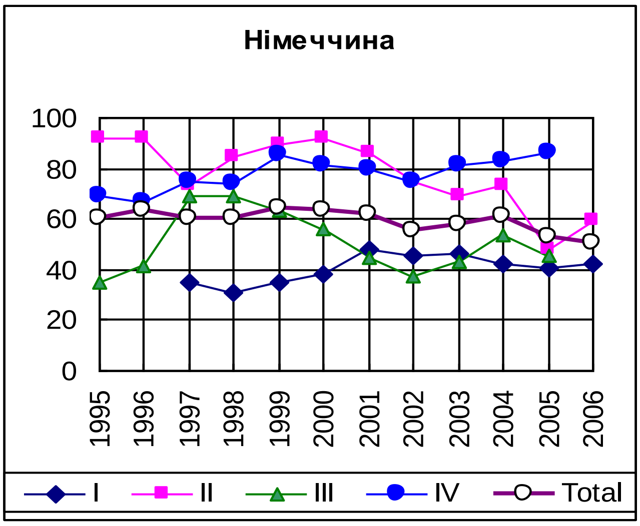
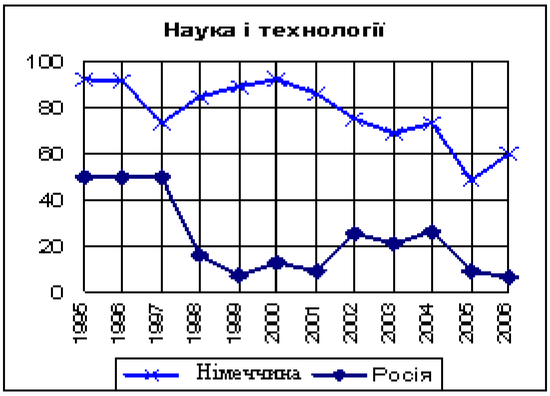


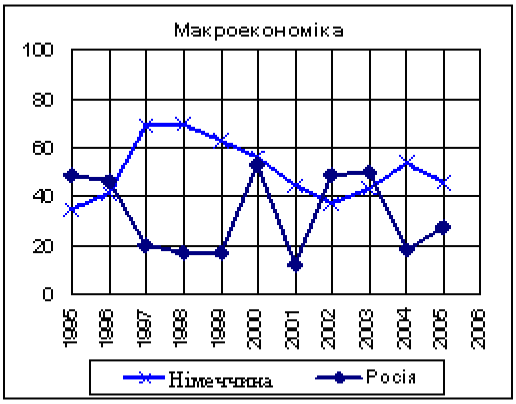
Рис. 6.6. **Динаміка вимірників чотирьох складних ознак: населення, праці, суспільства (I); науки й технологій (II); макроекономіки (III); фінансів (IV) та їх загального стану (Total), що описують соціально-економічний розвиток Німеччини**

Для порівняльного аналізу країн у динаміці на рис. 6.7 наведені графіки вимірників складних ознак соціально-економічного розвитку Росії та Німеччини.

а

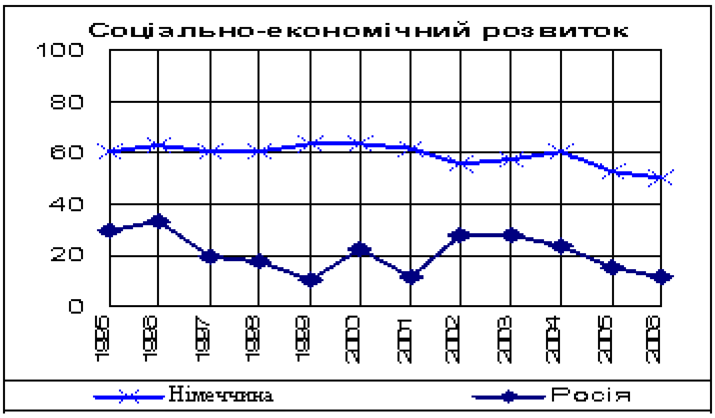


б



в

г



д

Рис. 6.7. **Динаміка вимірників чотирьох складних ознак: населення, праці, суспільства (а); науки й технології (б); макроекономіки (в); фінансів (г) та їх загального стану (д), що описують соціально-економічний розвиток Росії та Німеччини**

Зображення на рис. 6.7 наочно демонструє стан розвитку складних соціально-економічних ознак двох країн.

Задача 6.3 (продовження задачі 4.2). Для аналізу даних інтелектуального забезпечення діяльності 7 промислових підприємств Харківського регіону слід визначити складні ознаки.

Розв'язання задачі. Рекомендується застосувати факторний аналіз з вихідною матрицею коефіцієнтів CCK (табл. 6.11).

Таблиця 6.11

Коефіцієнти *ССК*

CCK	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*K*	*L*
A	1	0,302	0,349	0,226	0,063	0,527	0,302	0,051	0,354	0,283
B	0,302	1	0,302	0,314	0,336	0,357	0,197	0,197	0,231	0,447
C	0,349	0,302	1	0,545	0,486	0,329	0,446	0,218	0,595	0,259
D	0,226	0,314	0,545	1	0,387	0,545	0,413	0,496	0,404	0,395
E	0,063	0,336	0,486	0,387	1	0,502	0,345	0,345	0,247	0,649
F	0,527	0,357	0,329	0,545	0,502	1	0,449	0,590	0,220	0,547
G	0,302	0,197	0,446	0,413	0,345	0,449	1	0,197	0,093	0,184
H	0,051	0,197	0,218	0,496	0,345	0,590	0,197	1	0,093	0,184
K	0,354	0,231	0,595	0,404	0,247	0,220	0,093	0,093	1	0,290
L	0,283	0,447	0,259	0,395	0,649	0,547	0,184	0,184	0,290	1

Для матриці коефіцієнтів CCK обчислені власні вектори і власні числа (табл. 6.12).

Таблиця 6.12

Власні вектори U і власні числа λ матриці *ССК*

CCK	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
1	2	3	4	5	6	7	8	9	10	11
A	0,255	0,404	0,012	0,650	-0,145	-0,206	-0,145	0,265	-0,093	-0,430
B	0,277	0,035	-0,431	0,159	-0,069	0,826	-0,090	-0,080	0,095	0,014
C	0,348	0,364	0,224	-0,325	0,165	0,095	-0,308	0,477	-0,247	0,414
D	0,370	-0,063	0,282	-0,198	-0,170	0,137	0,710	0,306	0,202	-0,234
E	0,345	-0,244	-0,275	-0,333	0,344	-0,220	-0,356	0,105	0,369	-0,441

Закінчення табл. 6.12

1	2	3	4	5	6	7	8	9	10	11
F	0,394	-0,252	0,058	0,356	-0,146	-0,249	-0,054	-0,070	0,476	0,577
G	0,277	-0,014	0,413	0,221	0,668	0,124	0,105	-0.449	-0,159	-0,065
H	0,260	-0,508	0,326	-0,069	-0,467	0,080	-0,338	-0,136	-0,417	-0,179
K	0,264	0,550	-0,012	-0,341	-0,330	-0,152	-0,017	-0,603	0,115	-0,048
L	0,333	-0,126	-0,573	-0,007	0,072	-0,303	0,343	-0,044	-0,553	0,151
λ	4,146	1,258	1,058	0,995	0,835	0,656	0,448	0,293	0,170	0,139
%	41,46	12,58	10,58	9,95	8,35	6,56	4,48	2,93	1,70	1,39
$\sum$	41,46	54,05	64,63	74,58	82,93	89,49	93,97	96,90	98,61	100,00

Усі власні вектори ортонормовані. Для отримання початкової матриці факторних навантажень слід нормувати власні вектори так, щоб сума квадратів їх елементів дорівнювала власним числам: $V_k = U_k \sqrt{\lambda_k}$. За рекомендаціями, що викладені в розділі 2.2, слід залишити 4 перших (головних) компоненти, дисперсії яких (власні числа) не менші одиниці і які пояснюють 74,6% загальної мінливості. Матриця факторних навантажень після обертання *VARIMAX* наведена в табл. 6.13.

Таблиця 6.13

Факторна матриця після обертання *VARIMAX*

	O_1	O_2	O_3	O_4	R^2
A	0,908	-0,131	0,092	-0,218	0,897
B	0,316	-0,645	0,077	-0,136	0,541
C	0,131	-0,130	0,294	-0,841	0,828
D	0,046	-0,204	0,638	-0,495	0,696
E	-0,185	-0,693	0,400	-0,293	0,760
F	0,431	-0,428	0,695	-0,043	0,854
G	0,368	0,047	0,594	-0,243	0,549
H	-0,131	-0,148	0,827	0,003	0,724
K	0,148	-0,187	-0,064	-0,850	0,784
L	0,131	-0,876	0,158	-0,134	0,827
СумКв	1,354	1,981	2,218	1,906	
%	13,5	19,8	22,2	19,1	
$\sum$	13,5	33,4	55,5	74,6	

Перша складна ознака (O_1) виявляється за впливом її на якість програмного забезпечення ($a_{11} = 0{,}908$) і рівень залежності працівників від підприємства ($a_{61} = 0{,}431$), це – застосування передових наукоємних технологій. Друга складна ознака (O_2) за її впливом на елементарні ознаки ідентифікується як ефективність керівництва, третя (O_3) – як ефективність виробництва, четверта (O_4) – як стан здоров'я працівників.

Таким чином, застосування модифікованого факторного аналізу розширює можливості розв'язання задач моделювання соціально-економічних систем, особливо в умовах відносної визначеності, коли ознаки визначаються неметричними величинами – порядковими, назвами або категоріями.

6.2. Аналіз складних сумісних ознак соціально-економічних систем на основі методу багатовимірного шкалування

Мета багатовимірного шкалування – визначення взаємного розпізнавання об'єктів у певному теоретичному просторі невеликої розмірності, що адекватно відображає реальність. За вихідними значеннями певного набору ознак важко безпосередньо сформувати образ об'єкта та адекватно його ідентифікувати. Таким чином, мета багатовимірного шкалування фактично співпадає з метою кластерного аналізу – з'ясування структури взаємного розташування об'єктів, виявлення кількості та складу груп близьких об'єктів (кластерів), оцінка за певною шкалою відносних відстаней між об'єктами. Проте дана задача розв'язується цими двома багатовимірними статистичними методами зовсім по-різному.

Слід зауважити, що методи кластерного аналізу дотепер не мають жодного теоретичного обґрунтування, хоча більш ніж піввіковою практикою були відсіяні явно невдалі техніки, що призводили до неправильної класифікації об'єктів, і наразі відпрацьовані відносно надійні процедури, які витримали ряд перевірок на прикладах аналізу сукупності об'єктів з відомими класифікаціями. У підрозділі 2.2 зазначалось, що серед ієрархічних процедур кластерного аналізу сьогодні найперспективнішою вважається процедура Уорда.

Методи багатовимірного шкалування в цілому також ще далекі від теоретичної завершеності. Розрізняють метричне і неметричне багатовимірне шкалування. Найбільш обґрунтованим серед них є один із перших алгоритмів багатовимірного шкалування – метричний метод Торгенсона. Елементи цього методу використовуються в інших неметричних методах багатовимірного шкалування [52].

З обчислювальної точки зору метод Торгенсона надзвичайно схожий на метод головних компонент факторного аналізу, проте між цими методами наявні істотні відмінності. Порівняємо ці методи, оскільки кінцеві результати, що отримані за ними, дуже схожі.

У методі головних компонент за матрицею початкових метричних даних $\boldsymbol{X}$ розміру $n \times m$ (n – об’єктів, m – ознак) складають кореляційну матрицю $\boldsymbol{R}$ розміру $m \times m$, знаходять всі її власні числа λ і власні вектори $\boldsymbol{U}$. На цьому етапі визначають кількість головних компонент ($p < m$), достатню для адекватного подання початкового простору ознак. Далі обчислюються елементи матриці факторних навантажень $\boldsymbol{A}$ розміру $m \times p$ і елементи матриці значень головних компонент $\boldsymbol{\Phi}$ розміру $n \times p$. У скороченому просторі головних компонент (латентних факторів) чіткіше проявляються всі основні властивості системи "ознаки – об’єкти", що приховані в таблиці початкових даних. Якщо кількість головних компонент дорівнює двом ($p = 2$), то розподіл об’єктів можна встановити візуально, зобразивши графік розкиду точок (для кожного об’єкта) у просторі головних компонент. У факторному аналізі (вже з іншою метою) додатково застосовується перетворення обертання, що переводить факторну матрицю $\boldsymbol{A}$ в нову матрицю факторних навантажень $\boldsymbol{B}$, а матрицю значень головних компонент $\boldsymbol{\Phi}$ – у нову матрицю значень головних факторів $\boldsymbol{F}$ (розміри матриць не змінюються). На взаємне розташування об’єктів у факторному просторі ця операція обертання не впливає.

У методі Торгенсона (основному методі багатовимірного шкалування) за матрицею початкових даних $\boldsymbol{X}$ розміру $n \times m$ складають матрицю квадратів евклідових відстаней $\boldsymbol{D^2}$ розміру $n \times n$. Елементи цієї матриці обчислюються за відомою формулою:

$$d_{ij}^2 = \sum_{k=1}^{m} \left(x_{ik} - x_{jk}\right)^2 . \tag{6.1}$$

Зауваження. Значення всіх ознак повинні бути заздалегідь нормовані.

Для матриці $\boldsymbol{D}^2$ обчислюють середні у рядках $d_{i\bullet}^2$, у стовпчиках $d_{\bullet j}^2$ і загальне середнє $d_{\bullet\bullet}^2$; далі за допомогою подвійного центрування переходять до матриці ***D****, елементи якої визначають за формулою:

$$d_{ij}^* = -\frac{1}{2}\left(d_{ij}^2 - d_{i\bullet}^2 - d_{\bullet j}^2 + d_{\bullet\bullet}^2\right). \qquad (6.2)$$

Подальші викладки майже повністю повторюють етапи методу головних компонент. Для матриці ***D**** знаходять власні числа μ і власні вектори ***V***. Матриця ***D**** має істотно більші розміри, ніж кореляційна матриця ***R***, але кількість ненульових власних чисел у цих матриць однакова і не перевищує кількості ознак m. Головні компоненти для матриці ***D**** називаються "стимулами" і їх кількість виявляється такою ж, як і кількість головних компонент із матрицею ***R*** (для визначення кількості головних компонент використовують евристичні правила, тому можливі незначні відхилення).

Матриця стимулів ***S*** складається аналогічно до матриці факторних навантажень ***A***, але за розмірами $n \times p$ і змістом вона еквівалентна матриці головних компонент ***Ф***. У скороченому просторі стимулів (латентних змінних) чіткіше проявляються всі властивості системи "ознаки – об’єкти", що приховані в таблиці початкових даних. Якщо число стимулів дорівнює двом ($p = 2$), то розподіл об’єктів можна встановити візуально, зобразивши графік розкиду точок (для кожного об’єкта) у просторі стимулів. Як і у факторному аналізі, допустимі перетворення матриці стимулів методом обертання, але для задачі розпізнавання образів ця операція вже нічого не дає.

Для обґрунтування методу Торгенсона наведемо доведення наступної теореми.

Теорема. Елементи матриці ***D**** можуть бути подані у вигляді скалярних добутків рядків початкової нормованої матриці ***X***:

$$d_{ij}^* = \sum_{k=1}^{m} x_{ik} x_{jk} . \qquad (6.3)$$

Доведення. Без втрати спільності можна вважати, що значення кожної ознаки центровані, тобто для будь-якого k справджується $\sum_{i=1}^{n} x_{ik} = \sum_{j=1}^{n} x_{jk} = 0$.

Операція центрування не змінює величин евклідових відстаней між об’єктами.

Виразимо середні $d_{i\bullet}^2$, $d_{\bullet j}^2$, $d_{\bullet\bullet}^2$ безпосередньо через початкові дані (нормовані й центровані).

Перетворимо середні за рядками матриці $\boldsymbol{D}^2$:

$$d_{i\bullet}^2 = \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\left(x_{ik} - x_{jk}\right)^2 =$$
$$= \frac{1}{n}\sum_{k=1}^{m}\sum_{j=1}^{n} x_{ik}^2 + \frac{1}{n}\sum_{k=1}^{m}\sum_{j=1}^{n} x_{jk}^2 - \frac{2}{n}\sum_{k=1}^{m}\sum_{j=1}^{n} x_{ik}x_{jk} = \tag{6.4}$$
$$= \sum_{k=1}^{m} x_{ik}^2 + \sum_{k=1}^{m} x_{\bullet k}^2 - 0,$$

де через $x_{\bullet k}^2$ позначені середні квадрати значень кожної ознаки: $x_{\bullet k}^2 = \frac{1}{n}\sum_{j=1}^{n} x_{jk}^2$.

Останній доданок у формулі $d_{i\bullet}^2$ дорівнює нулю $\sum_{k=1}^{m}\sum_{j=1}^{n} x_{ik}x_{jk} = 0$, оскільки прийнято $\sum_{j=1}^{n} x_{jk} = 0$. Аналогічно перетворюються середні за стовпчиками матриці $\boldsymbol{D}^2$:

$$d_{\bullet j}^2 = \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\left(x_{ik} - x_{jk}\right)^2 =$$
$$= \frac{1}{n}\sum_{k=1}^{m}\sum_{i=1}^{n} x_{ik}^2 + \frac{1}{n}\sum_{k=1}^{m}\sum_{i=1}^{n} x_{jk}^2 - \frac{2}{n}\sum_{k=1}^{m}\sum_{i=1}^{n} x_{ik}x_{jk} = \tag{6.5}$$
$$= \sum_{k=1}^{m} x_{\bullet k}^2 + \sum_{k=1}^{m} x_{jk}^2 - 0.$$

Загальне середнє $d_{\bullet\bullet}^2$ можна обчислити за однією з наведених нижче еквівалентних формул:

$$d_{\bullet\bullet}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\ \sum_{i=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n} d_{i\bullet}^2 = \frac{1}{n}\sum_{j=1}^{n} d_{\bullet j}^2 . \tag{6.6}$$

Перетворення будь-якої з цих формул приводить до виразу:

$$d_{\bullet\bullet}^2 = \sum_{k=1}^{m} x_{\bullet k}^2 + \sum_{k=1}^{m} x_{\bullet k}^2 = 2\sum_{k=1}^{m} x_{\bullet k}^2 . \tag{6.7}$$

Нарешті перетворимо елементи матриці ***D**** (після подвійного центрування матриці $\boldsymbol{D}^2$):

$$d_{ij}^{*} = -\frac{1}{2}\left(d_{ij}^{2} - d_{i\bullet}^{2} - d_{\bullet j}^{2} + d_{\bullet\bullet}^{2}\right) =$$
$$= -\frac{1}{2}\left[\left(\sum_{k=1}^{m} x_{ik}^{2} + \sum_{k=1}^{m} x_{jk}^{2} - 2\sum_{k=1}^{m} x_{ik}x_{jk}\right) - \left(\sum_{k=1}^{m} x_{ik}^{2} + \sum_{k=1}^{m} x_{\bullet k}^{2}\right) - \left(\sum_{k=1}^{m} x_{\bullet k}^{2} + \sum_{k=1}^{m} x_{jk}^{2}\right) + 2\sum_{k=1}^{m} x_{\bullet k}^{2}\right] =$$
$$= \sum_{k=1}^{m} x_{ik}x_{jk}.$$

Дійсно, елементи матриці ***D**** виявилися такими, що дорівнюють скалярним добуткам рядків матриці ***X*** (центрованої та нормованої), тому матрицю ***D**** часто називають матрицею скалярних добутків. ■

Можна сформулювати наступні наслідки, які випливають з доведеної теореми.

Наслідки. Як і кореляційна матриця ***R***, матриця ***D**** може бути представлена у вигляді матричного добутку взаємно транспонованих матриць: ***D* = XX'***.

Порівняємо ***D**** з кореляційною матрицею $\boldsymbol{R} = \frac{1}{n}\boldsymbol{Z}'\boldsymbol{Z}$, де ***Z*** – матриця стандартизованих даних. У матриці ***X*** дані також центровані й нормовані, але в якості коефіцієнтів нормування можуть бути прийняті не тільки середні квадратичні відхилення ознак, але також будь-які інші міри, наприклад, середні значення ознак. Розглянемо найбільш поширений варіант нормування матриці ***X*** у вигляді стандартизації: ***X*** = ***Z***. Тоді розклад матриці ***D****= ***ZZ'*** буде дуже схожим на розклад кореляційної матриці.

Раніше (підрозділ 2.2) була доведена лема, що у матриць типу ***ZZ'*** всі власні числа μ – дійсні, невід’ємні, а всі власні вектори ***V*** – взаємно ортогональні. Ранг цієї матриці дорівнює рангу одного з матричних множників і не перевищує кількості однак *m*; отже, кількість ненульових власних чисел матриці ***ZZ'*** не перевищує *m* (решта власних чисел дорівнює нулю).

Таким чином, матриця ***D****= ***ZZ'*** може бути розкладена на таку суму матричних добутків:

$$\boldsymbol{D}^* = \mu_1 \boldsymbol{V_1 V_1}' + \mu_2 \boldsymbol{V_2 V_2}' + \ldots + \mu_m \boldsymbol{V_m V_m}', \qquad (6.8)$$

де внесок кожного члена пропорційний відповідному власному числу (оскільки норма кожного власного вектора дорівнює одиниці). Отже, декілька перших доданків з найбільшими власними числами відтворюють достатньо точно всю матрицю $\boldsymbol{D^*}$, а доданки з малими власними числами можуть бути відкинуті. Згідно з найпоширеною рекомендацією сума залишених у моделі власних чисел повинна бути близькою до 70% від їх загальної суми. Допустимо користуватися й іншими евристичними рекомендаціями, що прийняті у факторному аналізі.

Отже, доведено, що метод Торгенсона зі стандартизованою вихідною матрицею даних має багато спільного з методом головних компонент, але обсяг обчислювальної роботи за методом Торгенсона істотно перевищує аналогічний обсяг роботи за методом головних компонент [62]. У разі великої кількості об'єктів це може виявитися перепоною для застосування методу Торгенсона. Проте цей недолік можна усунути, якщо довести наступну теорему.

Теорема. Нехай λ_i – ненульові власні числа і $\boldsymbol{U_i}$ – відповідні власні вектори матриці $\boldsymbol{Z'Z}$ (розміру $m\times m$). Тоді для матриці $\boldsymbol{ZZ'}$ (розміру $n\times n$) всі ненульові власні числа будуть тими ж самими, а відповідні їм власні вектори можна отримати за допомогою матричного добутку $\boldsymbol{V} = \boldsymbol{ZUD}_{1/\sqrt{\lambda}}$. Тут $\boldsymbol{D}_{1/\sqrt{\lambda}}$ – діагональна матриця (розміру $p\times p$), на головній діагоналі якої розміщені числа, обернені кореням квадратним з ненульових власних чисел λ_i; $\boldsymbol{U}$ – матриця власних векторів розміром $m\times p$, що відповідні ненульовим власним числам; $\boldsymbol{Z}$ – початкова стандартизована матриця розміром $n\times m$.

Доведення. Скористаємось сингулярним розкладом Форсайта для довільної прямокутної матриці. Сингулярні числа σ_i прямокутної матриці $\boldsymbol{Z}$ розміру $n\times m$ визначаються наступною системою матричних рівностей:

$$\begin{cases} \boldsymbol{ZU} = \boldsymbol{VD}_\sigma \\ \boldsymbol{Z'V} = \boldsymbol{UD}'_\sigma \end{cases}. \tag{6.9}$$

Розміри цих матриць узгоджені:

$$[\boldsymbol{Z}] = (n\times m);\ [\boldsymbol{U}] = (m\times m);\ [\boldsymbol{V}] = (n\times n);\ [\boldsymbol{D}_\sigma] = (n\times m);$$
$$[\boldsymbol{Z'}] = (m\times n);\ [\boldsymbol{V}] = (n\times n);\ [\boldsymbol{U}] = (m\times m);\ [\boldsymbol{D'}_\sigma] = (m\times n),$$

де $\boldsymbol{D}_\sigma$ – прямокутна матриця розміру $n\times m$, в якій відмінні від нуля елементи (сингулярні числа σ_i) розміщуються на головній діагоналі (решта елементів дорівнюють нулю).

Помножимо перше матричне рівняння системи для визначення сингулярних чисел ліворуч на $\boldsymbol{Z'}$ і врахуємо друге матричне рівняння системи:

$$\boldsymbol{Z'ZU} = \boldsymbol{Z'VD}_{\sigma} = \left(\boldsymbol{UD}'_{\sigma}\right)\boldsymbol{D}_{\sigma} = \boldsymbol{UD}_{\lambda}, \tag{6.10}$$

де $\boldsymbol{D}_{\lambda}$ – діагональна матриця розміру $m \times m$, в якій на головній діагоналі розміщені квадрати сингулярних чисел $\lambda_i = \sigma_i^2$.

З отриманого рівняння $(\boldsymbol{Z'Z})\boldsymbol{U} = \boldsymbol{UD}_{\lambda}$ випливає, що λ_i – власні числа, а $\boldsymbol{U_i}$ – власні вектори симетричної додатної визначеної матриці $\boldsymbol{Z'Z}$ (отже, всі власні числа λ_i – дійсні невід'ємні, а власні вектори $\boldsymbol{U_i}$ – взаємно ортогональні).

Тепер помножимо друге матричне рівняння системи для визначення сингулярних чисел ліворуч на $\boldsymbol{Z}$ і врахуємо перше матричне рівняння системи:

$$\boldsymbol{ZZ'V} = \boldsymbol{Z'UD}'_{\sigma} = \left(\boldsymbol{VD}_{\sigma}\right)\boldsymbol{D}'_{\sigma} = \boldsymbol{VD}_{\lambda}^{*}, \tag{6.11}$$

де $\boldsymbol{D}_{\lambda}^{*}$ – діагональна матриця розміру $n \times n$, в якій на головній діагоналі містяться m ненульових елементів $\lambda_i = \sigma_i^2$ (решта елементів дорівнюють нулю).

З отриманого рівняння

$$\boldsymbol{ZZ'V} = \boldsymbol{VD}_{\lambda}^{*}$$

випливає, що λ_i – власні числа, а $\boldsymbol{V_i}$ – власні вектори симетричної додатної напіввизначеної матриці $\boldsymbol{ZZ'}$. Ранг матриці $\boldsymbol{ZZ'}$ дорівнює рангу матриці $\boldsymbol{Z'Z}$ і кількості ненульових власних чисел $\lambda_i = \sigma_i^2$. Отже, ненульові власні числа $\lambda_i = \sigma_i^2$ співпадають для обох матриць.

Перша частина теореми доведена – матриці $\boldsymbol{Z'Z}$ і $\boldsymbol{ZZ'}$ мають одні й ті ж ненульові власні числа λ_i. Таким чином, можна не розглядати проблему власних чисел для матриці $\boldsymbol{ZZ'}$ (дуже великого розміру), досить знайти власні числа і власні вектори матриці $\boldsymbol{Z'Z}$ (істотно меншого розміру).

Залишилось визначити власні вектори матриці $\boldsymbol{ZZ'}$. Проте нам потрібні не всі власні вектори цієї матриці, а тільки ті, які відповідають ненульовим власним числам $\lambda_i > 0$. Ці вектори можна знайти з першої матричної рівності системи

$$\boldsymbol{ZU} = \boldsymbol{VD}_{\sigma},$$

що можна записати у вигляді $\boldsymbol{ZU}_i = \boldsymbol{V}_i\sqrt{\lambda_i}$, звідки випливає, що

$$\boldsymbol{V}_i = \boldsymbol{ZU}_i \frac{1}{\sqrt{\lambda_i}}.$$

У матричній формі сукупність власних векторів $\boldsymbol{V}_i$ (для $\lambda_i \neq 0$) можна записати у вигляді $\boldsymbol{V} = \boldsymbol{ZUD}_{1/\sqrt{\lambda}}$, що й потрібно було довести. ■

Теорема. Якщо початкова матриця $\boldsymbol{Z}$ – стандартизована, то класифікації повинні точно співпадати у просторі головних компонент і в просторі нормованих стимулів.

Доведення. Нормовані координати стимулів співпадають з власними векторами $\boldsymbol{V}$, які (як було доведено вище) виражаються через власні вектори $\boldsymbol{U}$: $\boldsymbol{V} = \boldsymbol{ZUD}_{1/\sqrt{\lambda}}$ – а це є координати головних компонент $\boldsymbol{\Phi} = \boldsymbol{ZUD}_{1/\sqrt{\lambda}}$. ■

На заключному етапі багатовимірного шкалування потрібно в координатному просторі стимулів візуально визначити структуру відносного розміщення об'єктів. Це можна здійснити, якщо кількість головних стимулів (головних компонент) дорівнює двом; іноді можна (з певними труднощами) зробити це також у тривимірному просторі стимулів.

Отже, для ідентифікації соціально-економічних систем на завершальному візуальному етапі багатовимірного шкалування рекомендується застосовувати кластерний аналіз за методом Уорда в скороченому просторі стимулів. У скороченому просторі (головних компонент) метод Уорда працює надійно і подає результати аналізу у зручній формі.

Таким чином, доведено еквівалентність результатів двох процедур: методу головних компонент факторного аналізу і метричного методу Торгенсона багатовимірного шкалування [61].

Задача 6.4. Провести аналіз даних діяльності 7 підприємств за 3 роки, що елементарно характеризуються 6 метричними ознаками виробничо-господарської діяльності ($x_{16}, x_{17}, x_{18}, x_{34}, x_{35}, x_{38}$ у табл. 4.19) і 4 неметричними ознаками інтелектуальної діяльності (C, D, E, F у табл. 4.19) за допомогою моделі складних ознак, ідентифікованих метричним методом Торгенсона багатовимірного шкалування.

Розв'язання задачі. У табл. 6.14 наведена стандартизована матриця ознак розміру 21×10.

Таблиця 6.14

Стандартизовані значення ознак (матриця Z)

Об'єкти	x_{16}	x_{17}	x_{18}	x_{34}	x_{35}	x_{38}	C	D	E	F
1	0,633	-0,352	0,677	0,527	-0,872	-1,104	1,501	1,226	1,821	1,338
2	-0,108	-0,152	-0,199	0,178	-0,663	-0,967	-0,075	-0,040	-1,655	-0,418
3	-0,233	-0,048	-0,342	0,174	-0,660	-0,973	-0,075	-1,729	0,083	-1,296
4	1,856	-1,717	1,776	-0,303	-0,245	-0,096	1,501	1,226	1,821	1,338
5	1,127	-1,120	0,975	-0,643	0,348	0,202	-0,075	-0,040	-1,655	-0,418
6	1,101	-1,099	0,947	-0,648	0,359	0,184	-2,439	1,226	0,083	-1,296
7	-0,979	1,742	-0,666	-0,823	0,808	-1,368	-0,075	-0,040	1,821	1,338
8	-0,731	0,401	-0,545	-1,061	1,987	-1,325	-0,075	-0,040	0,083	-0,418
9	-1,192	0,779	-1,078	-1,061	1,993	-1,335	-0,075	1,226	0,083	-1,296
10	-0,590	1,378	-0,910	1,566	-1,198	0,114	1,501	1,226	0,083	-0,418
11	-0,585	0,282	0,004	2,228	-1,006	0,548	-0,075	-0,040	0,083	-0,418
12	-0,579	0,276	0,012	2,235	-1,307	0,551	-0,075	-0,040	0,083	-0,418
13	0,411	-0,170	0,555	-0,359	-0,174	-0,469	-0,075	1,226	0,083	1,338
14	-0,185	-0,046	-0,292	-0,884	1,069	-0,407	-0,075	-0,040	0,083	-0,418
15	-0,194	-0,039	-0,298	-0,885	1,071	-0,450	-2,439	-1,729	0,083	-0,418
16	-0,812	1,651	-1,032	-0,624	0,276	-0,087	-0,075	-0,040	0,083	1,338
17	-1,474	1,009	-1,730	-0,749	0,617	0,811	-0,075	-0,040	-1,655	1,338
18	-1,477	1,012	-1,734	-0,750	0,619	0,805	-0,075	-0,040	0,083	-0,418
19	1,523	-1,353	1,532	0,548	-0,882	1,349	1,501	-0,040	0,083	-0,418
20	1,248	-1,219	1,179	0,667	-0,921	2,008	-0,075	-1,729	-1,655	-1,296
21	1,240	-1,213	1,171	0,666	-0,920	2,009	-0,075	-1,729	0,083	1,338

У табл. 6.15 обчислені елементи матриці $\mathbf{Z'Z}$, що пропорційна кореляційній матриці (її елементи більше відповідних коефіцієнтів кореляції в n = 21 раз).

Таблиця 6.15

Матриця $\mathbf{Z'Z}$

$\mathbf{Z'Z}$	1	2	3	4	5	6	7	8	9	10
1	21	-19,379	20,374	3,000	-8,545	7,784	3,249	-0,670	1,247	0,727
2	-19,379	21	-19,041	-2,324	6,529	-8,275	-0,532	2,757	2,008	3,062
3	20,374	-19,041	21	5,336	-9,358	7,026	3,313	-0,397	2,716	0,700
4	3,00	-2,324	5,336	21	-18,309	8,623	7,309	-1,404	-0,091	-2,162
5	-8,545	6,529	-9,358	-18,309	21	-10,314	-8,421	2,239	0,537	-1,573
6	7,784	-8,275	7,026	8,623	-10,314	21	1,044	-7,811	-8,032	-0,433
7	3,249	-0,532	3,313	7,309	-8,421	1,044	21	6,924	5,609	6,953
8	-0,670	2,757	-0,397	-1,404	2,239	-7,811	6,924	21	7,408	4,096
9	1,247	2,008	2,716	-0,091	0,537	-8,032	5,609	7,408	21	8,357
10	0,727	3,062	0,700	-2,162	-1,573	-0,433	6,953	4,096	8,357	21

У табл. 6.16 наведені власні вектори матриці $\boldsymbol{Z'Z}$ (і кореляційної матриці R), в останніх рядках таблиці наведені відповідні власні числа λ_i матриці $\boldsymbol{Z'Z}$, їх величини у відсотках від загальної суми ($n\times m$ = 210) і накопичені суми власних чисел у відсотках (власні числа кореляційної матриці R будуть у 21 раз менші).

Таблиця 6.16

Власні вектори $\boldsymbol{U}$ матриці $\boldsymbol{Z'Z}$

Показ-ники	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}
1	2	3	4	5	6	7	8	9	10	11
1	0,460	0,038	0,347	-0,007	0,035	0,024	-0,068	-0,313	0,334	0,671
2	-0,440	0,106	-0,362	-0,071	-0,064	-0,073	0,020	-0,270	0,759	-0,014
3	0,467	0,071	0,311	0,049	-0,090	-0,021	-0,091	0,228	0,476	-0,618
4	0,293	0,086	-0,549	0,262	-0,254	-0,162	-0,069	0,577	0,096	0,315
5	-0,389	-0,114	0,434	-0,058	0,152	0,124	0,323	0,623	0,234	0,242
6	0,330	-0,248	-0,250	-0,370	0,272	-0,285	0,684	-0,052	0,030	-0,061
7	0,149	0,469	-0,238	0,028	0,422	0,705	0,150	0,032	0,020	-0,031
8	-0,087	0,465	0,126	0,450	0,479	-0,566	0,071	-0,033	-0,041	-0,019
9	-0,033	0,524	0,164	-0,013	-0,642	-0,003	0,510	-0,104	-0,120	0,004
10	-0,008	0,436	0,013	-0,762	0,078	-0,231	-0,349	0,202	-0,065	0,062
λ	75,99	44,06	36,54	19,13	14,52	9,62	6,98	1,78	1,20	0,81
%	36,2%	21,0%	17,4%	9,1%	6,9%	4,6%	3,3%	0,8%	0,6%	0,1%
Σ	36,2%	57,2%	74,6%	83,7%	90,6%	95,2%	98,5%	99,3%	99,9%	100,0%

Перші три власних числа становлять близько 75% від загальної суми, тобто досить виділити три головних компоненти (табл. 6.16). Їх нормовані значення для всіх об’єктів (7 підприємств за 3 роки) обчислені в табл. 6.17 за формулою:

$$\boldsymbol{\Phi} = \boldsymbol{ZUD}_{1/\sqrt{\lambda}} .$$

Суми квадратів елементів нормованих компонент $\boldsymbol{\Phi}$ однакові і дорівнюють одиниці, проте внески головних компонент у загальну мінливість суттєво відрізняються і дорівнюють власним числам λ . Тому для визначення розшарування об’єктів на споріднені групи (кластери) подальший аналіз слід проводити з ненормованими компонентами

$$\boldsymbol{F}_i = \boldsymbol{\Phi}_i \sqrt{\lambda_i} ,$$

суми квадратів елементів яких дорівнюють відповідним власним числам. Ненормовані компоненти $\boldsymbol{F}_i$ також наведені в табл. 6.17.

Таблиця 6.17

Матриця головних компонент

Об'єкти	Φ_1	Φ_2	Φ_3	F_1	F_2	F_3
1	0,108	0,492	0,046	0,937	3,264	-0,276
2	-0,004	-0,122	-0,075	-0,035	-0,808	0,451
3	-0,013	-0,161	-0,085	-0,111	-1,067	0,511
4	0,282	0,430	0,333	2,459	2,851	-2,012
5	0,144	-0,189	0,213	1,259	-1,256	-1,290
6	0,080	-0,188	0,376	0,701	-1,245	-2,273
7	-0,300	0,265	0,049	-2,612	1,760	-0,295
8	-0,263	-0,031	0,203	-2,295	-0,205	-1,229
9	-0,348	-0,002	0,152	-3,032	-0,013	-0,920
10	-0,026	0,216	-0,428	-0,223	1,437	2,590
11	0,108	0,003	-0,366	0,942	0,019	2,210
12	0,110	0,003	-0,365	0,955	0,019	2,208
13	0,023	0,196	0,135	0,199	1,303	-0,818
14	-0,117	-0,048	0,154	-1,017	-0,322	-0,934
15	-0,143	-0,332	0,213	-1,247	-2,205	-1,290
16	-0,220	0,087	-0,111	-1,922	0,580	0,672
17	-0,239	-0,112	-0,195	-2,086	-0,745	1,178
18	-0,245	-0,090	-0,152	-2,136	-0,599	0,917
19	0,366	0,058	0,019	3,187	0,383	-0,118
20	0,354	-0,392	-0,085	3,082	-2,605	0,514
21	0,343	-0,082	-0,034	2,993	-0,547	0,203
СумКв	1	1	1	75,993	44,058	36,538

Для порівняння з методом Торгенсона (зі стандартизованою вихідною матрицею $\boldsymbol{Z}$) в табл. 6.18 обчислені елементи матриці $\boldsymbol{ZZ'}$, яка за методом Торгенсона співпадає з матрицею $\boldsymbol{D^*}$ (після подвійного центрування). Обчислення елементів матриці $\boldsymbol{D^*}$ за допомогою матричного добутку $\boldsymbol{ZZ'}$ є значно простішою процедурою, ніж розрахунок матриці квадратів евклідових відстаней між кожною парою об'єктів і подальшого подвійного центрування елементів цієї матриці за формулами (6.1) і (6.2).

Таблиця 6.18

Матриця ZZ' (матриця D^*)

D^*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	12,104	-2,146	-2,437	12,005	-2,833	-2,874	3,631	-2,374	-2,774	3,619	0,671	0,683	4,508	-1,818	-7,407	-0,488	-5,673	-4,860	3,842	-6,027	0,648
2	-2,146	4,402	1,990	-3,827	2,236	-0,133	-2,952	-0,053	0,368	0,873	0,797	0,795	-0,364	-0,341	-0,056	-0,857	1,213	-0,926	-0,959	1,959	-2,019
3	-2,437	1,990	6,268	-4,572	-0,601	-1,307	-0,482	0,754	-0,019	-0,349	1,462	1,458	-3,611	0,306	3,447	-1,395	-2,235	0,190	-0,943	2,673	-0,595
4	12,005	-3,827	-4,572	18,570	2,101	1,961	-0,864	-3,622	-5,696	-1,919	-2,540	-2,508	5,568	-1,309	-6,966	-4,265	-8,927	-8,130	9,585	-0,648	6,003
5	-2,833	2,236	-0,601	2,101	6,972	4,512	-6,733	-0,651	-1,798	-4,620	-2,703	-2,687	0,530	0,462	0,676	-3,979	-1,429	-3,582	4,262	6,932	2,932
6	-2,874	-0,133	-1,307	1,961	4,512	13,041	-4,500	0,078	1,341	-6,057	-2,078	-2,062	1,208	1,134	4,800	-4,792	-5,252	-2,841	1,027	3,042	-0,250
7	3,631	-2,952	-0,482	-0,864	-6,733	-4,500	12,748	5,666	5,924	0,598	-2,979	-3,011	1,627	2,042	2,374	7,161	3,142	3,977	-8,396	-12,84	-6,130
8	-2,374	-0,053	0,754	-3,622	-0,651	0,078	5,666	8,011	9,131	-2,695	-4,957	-4,981	-0,610	4,066	4,386	2,599	2,682	3,578	-6,541	-6,761	-7,705
9	-2,774	0,368	-0,019	-5,696	-1,793	1,341	5,924	9,131	13,268	0,495	-4,281	-4,314	-0,777	4,616	2,808	2,925	3,435	5,878	-8,275	-9,504	-12,75
10	3,619	0,873	-0,349	-1,919	-4,620	-6,057	0,598	-2,695	0,495	10,914	5,865	5,860	-0,550	-2,380	-7,988	1,659	1,155	2,037	0,301	-2,938	-3,878
11	0,671	0,797	1,462	-2,540	-2,703	-2,078	-2,979	-4,957	-4,281	5,865	7,579	7,592	-1,711	-3,305	-3,080	-1,409	-1,580	-0,708	1,918	3,200	2,247
12	0,683	0,795	1,458	-2,508	-2,687	-2,062	-3,011	-4,981	-4,314	5,860	7,592	7,606	-1,707	-3,316	-3,092	-1,436	-1,612	-0,740	1,955	3,235	2,281
13	4,508	-0,364	-3,611	5,568	0,530	1,208	1,627	-0,610	-0,777	-0,550	-1,711	-1,707	4,191	-0,504	-2,386	0,784	-0,346	-2,553	0,314	-3,634	0,026
14	-1,848	-0,341	0,306	-1,309	0,462	1,134	2,042	4,066	4,616	-2,380	-3,305	-3,316	-0,504	2,400	2,669	0,713	1,034	1,919	-2,573	-2,480	-3,384
15	-7,407	-0,056	3,447	-6,966	0,676	4,800	2,374	4,386	2,808	-7,988	-3,080	-3,092	-2,386	2,669	11,382	0,989	1,278	2,164	-6,147	0,552	-0,401
16	-0,488	-0,857	-1,395	-4,265	-3,979	-4,792	7,161	2,599	2,925	1,659	-1,409	-1,436	0,784	0,713	0,989	6,727	6,875	4,682	-6,418	-6,884	-3,191
17	-6,673	1,213	-2,235	-8,927	-1,429	-5,252	3,142	2,682	3,435	1,155	-1,580	-1,612	-0,346	1,034	1,278	6,875	12,319	7,104	-6,927	-3,467	-2,787
18	-4,860	-0,926	0,190	-8,130	-3,582	-2,841	3,977	3,578	5,878	2,037	-0,708	-0,740	-2,553	1,919	2,164	4,682	7,104	7,995	-6,074	-4,093	-5,018
19	3,842	-0,959	-0,943	9,585	4,262	1,027	-8,396	-6,541	-8,275	0,301	1,918	1,955	0,314	-2,573	-6,147	-6,418	-6,927	-6,074	11,830	9,604	8,615
20	-6,027	1,959	2,673	-0,648	6,932	3,042	-12,84	-6,761	-9,504	-2,938	3,200	3,235	-3,634	-2,430	0,552	-6,884	-3,467	-4,093	9,604	17,174	10,858
21	0,648	-2,019	-0,595	6,003	2,932	-0,250	-6,130	-7,705	-12,75	-3,878	2,247	2,281	0,026	-3,384	-0,401	-3,191	-2,787	-5,018	8,615	10,858	14,501

У табл. 6.19 наведені значення перших 10 власних векторів, що відповідають ненульовим власним числам матриці $\boldsymbol{ZZ'}$. Ранг матриці $\boldsymbol{ZZ'}$ дорівнює $m=10$ (кількості ознак), тому лише перші 10 її власних чисел відмінні від нуля. Матриці типу $\boldsymbol{Z'Z}$ і $\boldsymbol{ZZ'}$ додатньо напіввизначені, тому їх власні числа – невід’ємні.

Таблиця 6.19

Власні вектори *V* матриці *ZZ'*

Об’єкти	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
1	0,108	0,492	-0,046	-0,046	-0,112	-0,064	0,155	0,033	0,224	-0,122
2	-0,004	-0,122	0,046	-0,175	0,156	-0,080	0,597	0,107	0,172	0,053
3	-0,013	-0,161	0,085	-0,146	-0,367	-0,450	0,184	0,351	0,274	0,081
4	0,282	0,430	-0,333	0,064	0,049	-0,074	-0,125	-0,038	0,232	-0,098
5	0,144	-0,189	-0,213	-0,040	0,334	-0,080	0,229	-0,070	-0,219	-0,022
6	0,080	-0,188	-0,376	-0,304	-0,080	0,623	-0,121	0,285	0,013	-0,033
7	-0,300	0,265	-0,049	0,222	-0,326	-0,046	0,004	0,102	-0,404	0,368
8	-0,263	-0,031	-0,203	-0,078	0,019	-0,267	-0,034	-0,446	-0,130	-0,111
9	-0,348	-0,002	-0,152	-0,351	0,161	-0,093	-0215,	-0,225	-0,024	0,035
10	-0,026	0,216	0,429	-0,277	0,147	0,023	-0,089	0,246	-0,260	-0,585
11	0,108	0,003	0,366	-0,169	-0,207	0,209	-0,007	-0,344	0,034	0,164
12	0,110	0,003	0,365	-0,169	-0,208	0,210	-0,007	-0,347	0,032	0,164
13	0,023	0,196	-0,135	0,079	0,136	0,282	0,296	0,021	-0,031	0,307
14	-0,117	-0,048	-0,154	-0,033	0,042	-0,149	-0,128	-0,064	0,060	-0,095
15	-0,143	-0,332	-0,213	0,153	-0,435	0,076	0,062	-0,052	0,039	-0,348
16	-0,220	0,087	0,111	0,309	0,036	0,096	0,070	0,261	-0,362	-0,076
17	-0,239	-0,112	0,195	0,388	0,436	0,142	0,092	-0,110	0,309	-0,026
18	-0,245	-0,090	0,152	0,087	0,107	0,012	-0,475	0,291	0,396	0,213
19	0,366	0,058	-0,019	-0,046	0,164	-0,225	-0,281	0,083	-0,132	0,241
20	0,354	-0,393	0,085	0,034	0,093	-0,170	-0,109	0,089	-0,285	0,144
21	0,343	-0,082	0,034	0,498	-0,146	0,027	-0,098	-0,173	0,061	-0,254
λ	75,99	44,06	36,54	19,13	14,52	9,62	6,98	1,78	1,20	0,18
%	36,2%	21,0%	17,4%	9,1%	6,9%	4,6%	3,3%	0,8%	0,6%	0,1%
Σ	36,2%	57,2%	74,6%	83,7%	90,6%	95,2%	98,5%	99,3%	99,9%	100%

Порівнюючи табл. 6.19 з табл. 6.17, бачимо, що власні вектори $\boldsymbol{V_i}$ матриці $\boldsymbol{ZZ'}$ (які відповідають її ненульовим власним числам) дійсно співпадають з нормованими компонентами $\boldsymbol{\Phi_i}$ матриці $\boldsymbol{Z'Z}$ (у табл. 6.17 наведені значення лише перших 3-х головних компонент). У підсумкових рядках табл. 6.19 і табл. 6.16 наведені власні числа матриць $\boldsymbol{ZZ'}$ і $\boldsymbol{Z'Z}$, звідки бачимо, що ненульові власні числа обох цих матриць дійсно співпадають. Як і в методі головних компонент, у методі Торгенсона зі стандартизованою вихідною матрицею $\boldsymbol{Z}$ виділяються 3 латентних ознаки – 3 головних стимули, координати яких обчислені в табл. 6.20 за формулою:

$$\boldsymbol{S}_i = \boldsymbol{V}_i \sqrt{\lambda_i} \ .$$

Таблиця 6.20

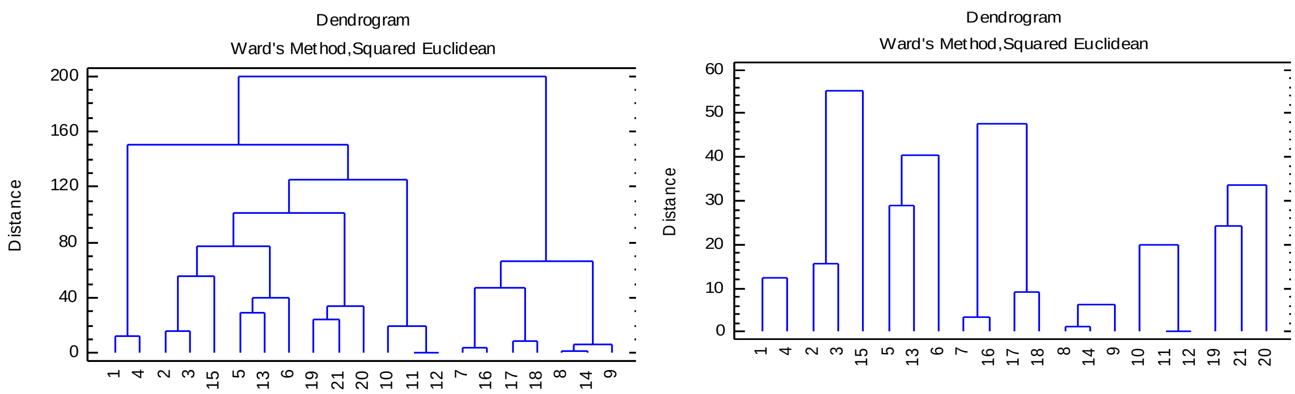
Матриця координат стимулів

Об'єкти	S_1	S_2	S_3
1	0,937	3,264	-0,276
2	-0,035	-0,808	0,451
3	-0,111	-1,067	0,511
4	2,459	2,851	-2,012
5	1,259	-1,256	-1,290
6	0,701	-1,245	-2,273
7	-2,612	1,760	-0,295
8	-2,295	-0,205	-1,229
9	-3,032	-0,013	-0,920
10	-0,223	1,437	2,590
11	0,942	0,019	2,210
12	0,955	0,019	2,208
13	0,199	1,303	-0,818
14	-1,017	-0,322	-0,934
15	-1,247	-2,205	-1,290
16	-1,922	0,580	0,672
17	-2,086	-0,745	1,178
18	-2,136	-0,599	0,917
19	3,187	0,383	-0,118
20	3,082	-2,605	0,514
21	2,993	-0,547	0,203
СумКв	75,993	44,058	36,538

Отже, на відміну від головних компонент, стимули не нормовані – суми їх квадратів дорівнюють не одиниці, а власним числам λ_i.

Переходимо до заключного (візуального) етапу багатовимірного шкалування. Як було вже сказано вище, на цьому етапі замість візуального аналізу будемо користуватися кластерним аналізом за методом Уорда.

Безпосередньо за 10-ма стандартизованими даними табл. 5.8 методом Уорда була побудована дендрограма і виділено 7 основних кластерів (рис. 6.8). Внизу під дендрограмами наведена таблиця, в якій показано, до якого кластера віднесено кожний об'єкт. Дані наведені за 7-ма основними об'єктами за 3 роки підряд, тому очікувалося об'єднання трьох послідовних об'єктів (рядків матриці даних) у споріднені групи (кластери), що є реальним щорічним станом СЕС.

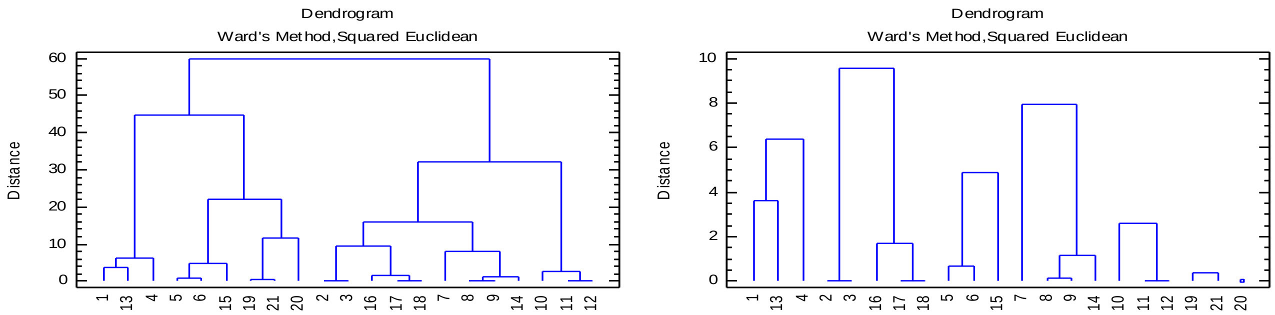


Кластери	1	1	2	2	2	3	3	3	4	4	4	4	5	5	5	6	6	6	7	7	7
Об'єкти	1	4	2	3	15	5	6	13	7	16	17	18	8	9	14	10	11	12	19	20	21

Рис. 6.8. **Кластерний аналіз на вихідних даних (10 показників)**

За дендрограмами рис. 6.8 дійсно виявилися близькими об'єкти (2, 3); (5, 6); (8, 9); (10, 11, 12); (16, 17, 18); (19, 20, 21). Об'єкти 1, 4, 7, 13, 14, 15 (28,6%) віднесені до інших кластерів. Такий результат ще раз підтверджує добру якість методу Уорда.

Далі методом головних компонент було виділено 3 перших компоненти, що в сумі пояснюють 74,6% всієї значимості початкових даних, і знову проведений кластерний аналіз вже у скороченому просторі перших 3-х головних компонент (рис. 6.9). Наведені дендрограми (рис. 6.9) більше впорядковані, ніж у попередніх (рис. 6.8). За дендрограмами виявилися близькими об'єкти (2, 3); (5, 6); (7, 8, 9); (10, 11, 12); (16, 17, 18); (19, 20, 21). Решта об'єктів 1, 4, 13, 14, 15 (23,8%) були віднесені до інших кластерів. До речі, один тільки візуальний аналіз лише за двома компонентами ($p = 2$) є дещо суб'єктивним і може не виявити таку добру узгодженість висновків двох порівняльних аналізів.



Кластери	1	1	1	2	2	2	2	2	3	3	3	4	4	4	4	5	5	5	6	6	7
Об'єкти	1	4	13	2	3	16	17	18	5	6	15	7	8	9	14	10	11	12	19	21	20

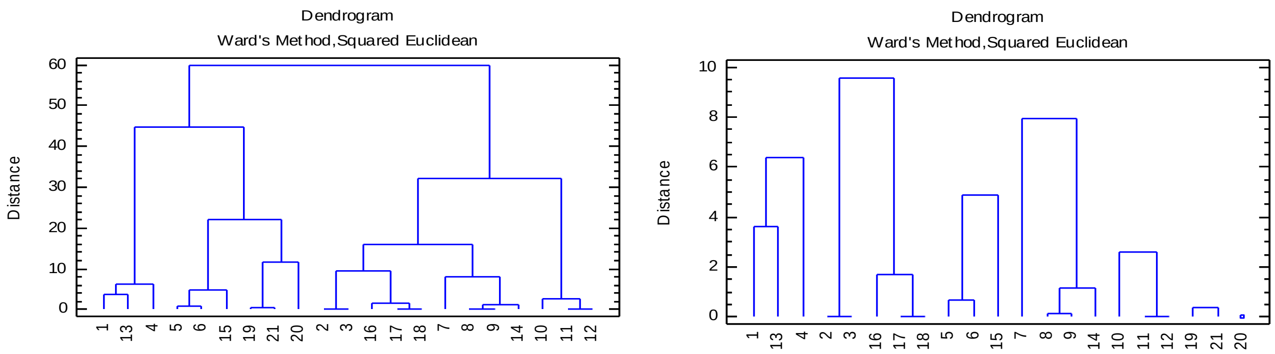
Рис. 6.9. **Кластерний аналіз на головних компонентах (3 компоненти)**

За методом Торгенсона були виділені 3 стимули, що так само, як і в методі головних компонент пояснюють 74,6% мінливості початкових даних, і далі за цими стимулами побудовані дендрограми рис. 6.10.

Ці дендрограми співпадають з дендрограмами рис. 6.9, що побудовані на головних компонентах. З одного боку, це природно, оскільки стимули виявилися пропорційними компонентам.

З іншого боку, тут може проявитися ефект різних масштабів за різними координатами, коли в разі збільшення відстаней за однією координатою (зміни масштабів) деякі об'єкти можуть виявитися близькими до інших угруповань.

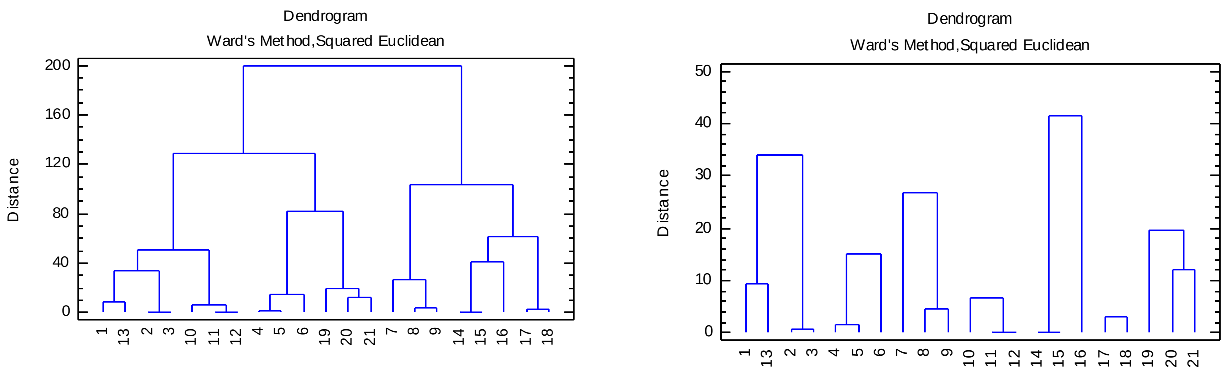
Далі ці самі початкові дані були попередньо перетворені в систему вимірників за функціями перетворення й увесь попередній аналіз був проведений знову.



Кластери	1	1	1	2	2	2	2	2	3	3	3	4	4	4	4	5	5	5	6	6	7
Об’єкти	1	4	13	2	3	16	17	18	5	6	15	7	8	9	14	10	11	12	19	21	20

Рис. 6.10. **Кластерний аналіз на стимулах (3 стимули)**

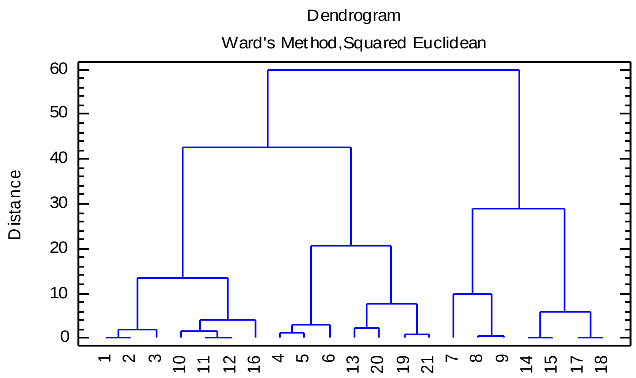
На цей раз навіть початкова дендрограма (рис. 6.11) виявилася достатньо об’єктивною. У ній визнані близькими об’єкти (1, 2, 3); (4, 5, 6); (7, 8, 9); (10, 11, 12); (14, 15); (17, 18); (19, 20, 21) – як бачимо, маємо надзвичайно добру відповідність наявній структурі. Лише два об’єкти 13, 16 (9,5%) віднесено до інших кластерів.

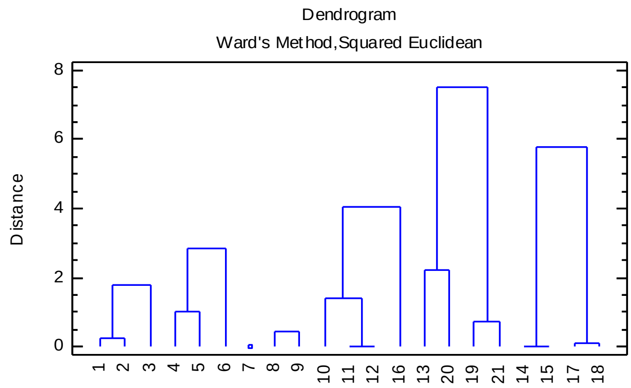


Кластери	1	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6	7	7	7
Об’єкти	1	2	3	13	4	5	6	7	8	9	10	11	12	14	15	16	17	18	19	20	21

Рис. 6.11. **Класифікація об’єктів за каліброваними даними (10 показників)**

Побудована на 3-х головних компонентах дендрограма рис. 6.12 виявила наступні групи споріднених об’єктів: (1, 2, 3); (4, 5, 6); (7, 8, 9); (10, 11, 12); (14, 15); (17, 18); (19, 20, 21). Лише два об’єкти 13, 16 виявилися віднесеними до інших кластерів. Таку ж добру відповідність демонструє дендрограма, що побудована за 3-ма стимулами.





Кластери	1	1	1	2	2	2	3	4	4	5	5	5	5	6	6	6	6	7	7	7	7
Об’єкти	1	2	3	4	5	6	7	8	9	10	11	12	16	13	19	20	21	14	15	17	18

Рис. 6.12. **Класифікація об’єктів за перетвореними даними у скорочених просторах**

Отже, використання перетворених величин за допомогою функцій перетворення забезпечує навіть за початковими даними відразу близький до правильного розподіл об’єктів за кластерами. Однією з причин збільшення стійкості класифікації слід визнати те, що у процесі перетворень вилучаються всі викиди, які, як відомо, надзвичайно спотворюють результати.

Добрі результати отримано під час класифікації у скороченому просторі факторів або стимулів. Більший обсяг обчислювальної роботи за методом Торгенсона (порівняно з методом головних компонент) за наявності потужної обчислювальної техніки вже не є недоліком. Обидва методи скорочення початкового простору слід визнати рівноцінними. Двовимірні діаграми (для суто візуального аналізу) у просторі перших 2-х факторів або у просторі перших 2-х стимулів виявилися малоінформативними і незручними. Нагадуємо, що розмірність простору штучно скорочується до двовимірного, щоб можна було візуально помітити розшарування об’єктів за кластерами. Проте двох узагальнених координат часто недостатньо для адекватного відображення реальної ситуації. Саме тому запропоновано доповнити завершальний (візуальний) етап багатовимірного шкалування кластерним аналізом у скороченому просторі. У скороченому просторі ефективніше працюють всі методи багатовимірного статичного аналізу, зокрема метод Уорда кластерного аналізу. Застосування сучасних програм кластерного аналізу не лише забезпечує наочність розшарування об'єктів за кластерами у вигляді дендрограм, а й супроводжується наданням всієї необхідної інформації про склад і характеристики кожного кластера.

Таким чином, усі багатовимірні методи краще працюють у скороченому просторі складних (латентних) ознак; найадекватнішу реальну класифікацію об'єктів маємо, якщо значення ознак, за якими виконується кластеризація, попередньо перетворити за рекомендованими в розділі 4 функціями перетворення, тобто виконувати кластеризацію на системі вимірників.

Переходимо до неметричного шкалування. Якщо вихідні дані є порядковими, класичний метод Торгенсона не може бути застосований, оскільки для початку аналізу за цим методом потрібно мати матрицю евклідових відстаней між об'єктами, а відстані між рангами не мають значення (ранги визначають лише порядок розташування даних). Але за порядковими даними можна обчислити коефіцієнти кореляції, причому якщо дані ранжовані, то ці коефіцієнти є коефіцієнтами рангової кореляції Спірмена. Метод головних компонент з кореляційною матрицею між ознаками також неможливий, оскільки для обчислення координат головних компонент, за якими виявляється взаємне розташування об'єктів у скорочених просторах, потрібно мати метричну матрицю ознак, а не порядкову.

Як було сказано вище, у методі Торгенсона зі стандартизованими даними матриця квадратів евклідових відстаней після подвійного центрування виявилася еквівалентною кореляційній матриці між об'єктами. Таку матрицю можна одержати і для порядкових даних, що потрібно для обчислення координат головних стимулів.

Продемонструємо цей новий метод неметричного шкалування за даними задачі 6.1 (для порівняння результатів аналізу).

Вихідні метричні дані ранжуємо по кожному рядку, тобто збережемо для кожного об'єкта лише порядок ознак, а не їх кількісні значення. За отриманими порядковими ознаками встановимо, які з об'єктів є близькими, і визначимо угруповання об'єктів у споріднені кластери. У табл. 6.21 за даними попередньої таблиці обчислені коефіцієнти кореляції Спірмена (ρ) між рядками (об'єктами).

Таблиця 6.21

Матриця коефіцієнтів рангової кореляції ρ між об’єктами

ρ_{ij}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	1	0,07	0,10	0,64	-0,35	-0,24	0,44	0,19	0,06	0,10	-0,20	-0,20	0,67	0,13	-0,22	0,065	-0,36	-0,37	0,18	-0,54	-0,12
2	0,067	1	0,152	-0,19	0,042	0,030	-0,31	-0,20	0,152	0,576	0,079	0,079	0,103	-0,27	-0,64	-0,30	-0,12	-0,42	0,164	0,006	-0,35
3	0,103	0,152	1	-0,04	-0,49	-0,39	0,115	0,018	0,127	0,418	0,442	0,442	-0,43	-0,03	0,188	-0,12	-0,44	-0,06	0,115	0,127	-0,12
4	0,636	-0,19	-0,04	1	0,333	0,261	-0,09	-0,18	-0,32	-0,42	-0,41	-0,41	0,564	0,067	0,042	-0,44	-0,62	-0,56	0,648	0,091	0,345
5	-0,35	0,042	-0,49	0,333	1	0,648	-0,65	-0,31	-0,24	-0,61	-0,52	-0,52	0,115	-0,04	0,006	-0,59	-0,22	-0,41	0,527	0,588	0,285
6	-0,24	0,030	-0,39	0,261	0,648	1	-0,33	-0,08	0,139	-0,49	-0,20	-0,20	0,200	0,273	0,176	-0,44	-0,39	-0,20	0,248	0,115	-0,08
7	0,442	-0,31	0,115	-0,09	-0,65	-0,33	1	0,842	0,576	-0,06	-0,15	-0,15	0,321	0,564	0,455	0,818	0,224	0,394	-0,71	-0,86	-0,53
8	0,188	-0,20	0,018	-0,18	-0,31	-0,08	0,842	1	0,855	-0,18	-0,38	-0,38	0,152	0,855	0,503	0,733	0,188	0,491	-0,69	-0,72	-0,78
9	0,055	0,152	0,127	-0,32	-0,24	0,139	0,576	0,855	1	0,055	-0,19	-0,19	-0,04	0,818	0,273	0,455	-0,02	0,394	-0,55	-0,62	-0,95
10	0,103	0,576	0,418	-0,42	-0,61	-0,49	-0,06	-0,18	0,055	1	0,636	0,636	-0,32	-0,27	-0,65	0,067	0,152	0,188	-0,08	-0,12	-0,26
11	-0,20	0,079	0,442	-0,41	-0,52	-0,20	-0,15	-0,38	-0,19	0,636	1	1,00	-0,48	-0,39	-0,25	-0,15	-0,07	0,212	0,055	0,152	0,091
12	-0,20	0,079	0,442	-0,41	-0,52	-0,20	-0,15	-0,38	-0,19	0,636	1,00	1	-0,48	-0,39	-0,25	-0,15	-0,07	0,212	0,055	0,152	0,091
13	0,673	0,103	-0,43	0,564	0,115	0,200	0,321	0,152	-0,04	-0,32	-0,48	-0,48	1	0,042	-0,08	0,055	-0,12	-0,49	0,067	-0,48	-0,03
14	0,127	-0,27	-0,03	0,067	-0,04	0,273	0,564	0,855	0,818	-0,27	-0,39	-0,39	0,042	1	0,491	0,430	-0,08	0,491	-0,39	-0,52	-0,72
15	-0,22	-0,64	0,188	0,042	0,006	0,176	0,455	0,503	0,273	-0,65	-0,25	-0,25	-0,08	0,491	1	0,333	-0,06	0,273	-0,36	-0,09	-0,06
16	0,055	-0,30	-0,12	-0,44	-0,59	-0,44	0,818	0,733	0,455	0,067	-0,15	-0,15	0,055	0,430	0,333	1	0,721	0,697	-0,93	-0,71	-0,43
17	-0,36	-0,12	-0,44	-0,62	-0,22	-0,39	0,224	0,188	-0,02	0,152	-0,07	-0,07	-0,12	-0,08	-0,06	0,721	1	0,600	-0,71	-0,22	-0,03
18	-0,37	-0,42	-0,06	-0,56	-0,41	-0,20	0,394	0,491	0,394	0,188	0,212	0,212	-0,49	0,491	0,273	0,697	0,600	1	-0,69	-0,29	-0,37
19	0,176	0,164	0,115	0,648	0,527	0,248	-0,71	-0,69	-0,55	-0,08	0,055	0,055	0,067	-0,39	-0,36	-0,93	-0,71	-0,69	1	0,673	0,503
20	-0,54	0,006	0,127	0,091	0,588	0,115	-0,86	-0,72	-0,62	-0,12	0,152	0,152	-0,48	-0,52	-0,09	-0,71	-0,22	-0,29	0,673	1	0,636
21	-0,12	-0,35	-0,12	0,345	0,285	-0,08	-0,53	-0,78	-0,95	-0,26	0,091	0,091	-0,03	-0,72	-0,06	-0,43	-0,03	-0,37	0,503	0,636	1

Оскільки коефіцієнти парної кореляції Пірсона r_{xy}, що обчислені на рангах, автоматично співпадають з коефіцієнтами рангової кореляції Спірмена ρ_s. Цей факт слід використати, коли в таблиці порядкових даних є групи однакових рангів, що дуже ускладнює обчислення за формулами Спірмена з урахуванням зв'язок (груп однакових рангів); у такому випадку доцільно використовувати звичайні формули Пірсона, які реалізовані в будь-якому програмному забезпеченні.

Далі були обчислені всі власні вектори і власні числа одержаної кореляційної матриці розміру 21×21. Оскільки ранг кореляційної матриці не може перевищувати кількість ознак, кількість ненульових власних чисел цієї матриці буде меншим 10. Тому в табл. 6.22 наведені значення лише перших 10 власних векторів і відповідних їм власних чисел (останні три рядки табл. 6.22).

Таблиця 6.22

Перші десять власних векторів матриці ρ

U	1	2	3	4	5	6	7	8	9	10
1	0,044	-0,147	0,495	-0,187	-0,128	-0,045	-0,219	-0,057	-0,148	0,028
2	-0,078	0,105	0,282	0,171	0,485	0,221	0,264	0,289	0,150	0,144
3	0,002	0,207	0,210	0,280	-0,375	0,406	0,172	0,278	0,054	-0,017
4	-0,151	-0,313	0,227	-0,064	-0,247	-0,051	-0,287	0,383	0,107	-0,086
5	-0,211	-0,277	-0,242	0,134	0,215	0,066	-0,055	-0,146	0,253	-0,054
6	-0,090	-0,244	-0,131	0,330	0,148	-0,517	0,170	0,394	-0,166	0,159
7	0,335	-0,056	0,150	-0,121	-0,194	-0,037	0,157	-0,154	0,155	-0,141
8	0,349	-0,133	0,026	0,122	-0,024	0,109	-0,027	-0,228	0,247	0,797
9	0,292	-0,057	0,077	0,372	0,139	0,038	0,010	-0,226	-0,205	-0,302
10	0,018	0,377	0,261	0,028	0,180	0,045	-0,225	0,327	0,051	0,082
11	-0,061	0,391	0,058	0,104	-0,158	-0,399	0,082	-0,128	0,195	0,019
12	-0,061	0,391	0,058	0,104	-0,158	-0,399	0,082	-0,128	0,195	0,019
13	0,017	-0,299	0,283	-0,273	0,115	-0,244	0,266	0,057	0,505	-0,086
14	0,268	-0,200	-0,023	0,302	-0,055	-0,058	-0,368	0,129	0,060	0,002
15	0,153	-0,162	-0,249	0,128	-0,421	0,078	0,389	0,249	0,151	-0,092
16	0,344	0,049	-0,076	-0,247	-0,001	0,058	0,013	0,157	0,076	-0,098
17	0,168	0,141	-0,270	-0,385	0,262	0,069	-0,038	0,173	0,181	-0,088
18	0,256	0,171	-0,270	0,024	-0,045	-0,158	-0,451	0,239	0,190	-0,006
19	-0,335	-0,087	0,141	0,121	-0,092	0,024	-0,274	-0,190	0,277	-0,043
20	-0,307	0,063	-0,248	0,114	-0,074	0,271	-0,098	-0,045	0,376	-0,050
21	-0,275	0,005	-0,156	-0,342	-0,244	0,000	0,042	0,119	-0,266	0,388
□	6,962	4,668	3,002	2,253	2,056	1,002	0,697	0,211	0,148	0,000
%	33,15%	22,23%	14,30%	10,73%	9,79%	4,77%	3,32%	1,01%	0,71%	0,00%
Cum %	33,15%	55,38%	69,68%	80,41%	90,20%	94,97%	98,29%	99,29%	100,00%	100,00%

Перші три власних числа кореляційної матриці (табл. 6.22) становлять близько 70% від загальної суми, отже можна виділити три головних стимули, координати яких обчислені в табл. 6.23 за формулою $s_{ij} = u_{ij}\sqrt{\lambda_j}$.

Таблиця 6.23

Координатний простір стимулів

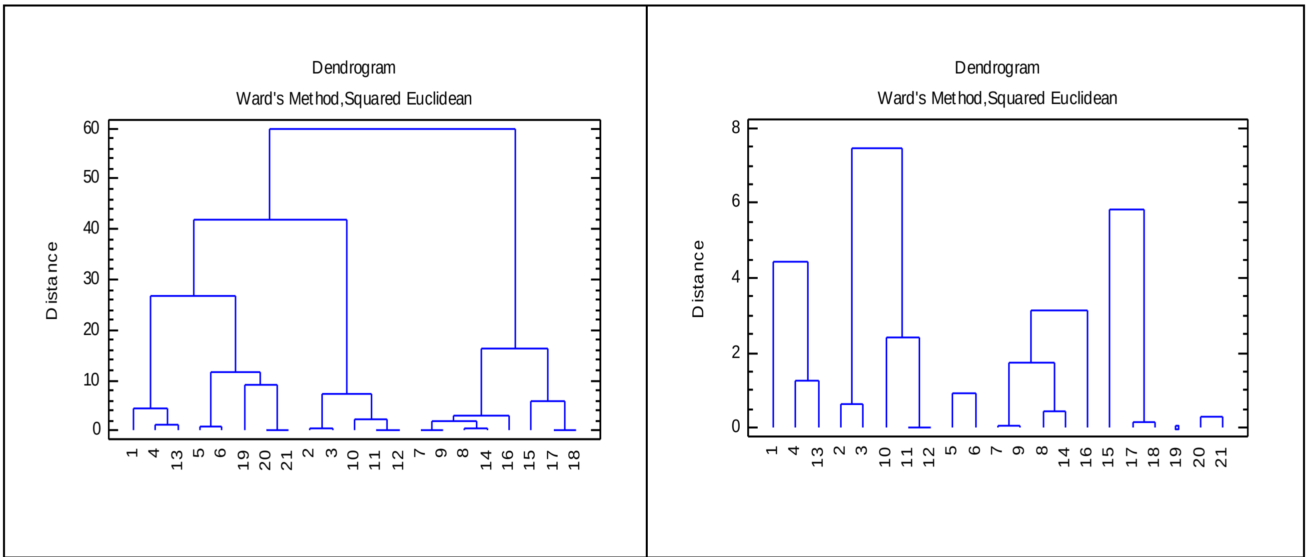
Стимули	S_1	S_2	S_3
1	0,116	-0,318	0,857
2	-0,207	0,226	0,488
3	0,006	0,446	0,365
4	-0,399	-0,677	0,393
5	-0,557	-0,598	-0,420
6	-0,237	-0,526	-0,228
7	0,883	-0,122	0,260
8	0,921	-0,287	0,044
9	0,772	-0,124	0,134
10	0,048	0,815	0,453
11	-0,161	0,845	0,101
12	-0,161	0,845	0,101
13	0,044	-0,645	0,490
14	0,706	-0,431	-0,041
15	0,403	-0,350	-0,431
16	0,908	0,107	-0,132
17	0,443	0,304	-0,469
18	0,675	0,369	-0,468
19	-0,885	-0,188	0,243
20	-0,809	0,137	-0,430
21	-0,726	0,012	-0,271
СумКв	6,962	4,668	3,002

Зазвичай нові власні числа (після операції ранжування) відрізняються від власних чисел, обчислених за метричними даними, але в обох випадках виділена однакова кількість латентних змінних, яка достатня для адекватного опису розшарування об'єктів.

Отже, незважаючи на те, що початкові дані не мали метрики для кількісного порівняння, ми одержали кількісні координати кожного об'єкта в певному метричному просторі стимулів. З обчислювальної точки зору запропонований простий метод неметричного шкалування практично співпадає з модифікованим методом Торгенсона. Відомо, що для метричних даних метод Торгенсона вважається найобґрунтованішим методом багатовимірного шкалування.

Подальший візуальний аналіз (як і для попередніх прикладів) замінимо кластерним аналізом у скороченому просторі головних стимулів. Слід зауважити, що безпосередній кластерний аналіз на вихідних порядкових даних неможливий, оскільки за порядковими даними неможливо обчислити евклідові відстані між об’єктами. На рис. 6.13 наведені побудовані у скороченому метричному просторі дендрограми і наведений склад виділених 7-ми кластерів.

Як видно з рис. 6.13, аналіз правильно визначив розташування понад 70% об’єктів за спорідненими групами: (2, 3); (5, 6); (7, 8, 9); (10, 11, 12); (17, 18); (19, 20, 21). Лише шість об’єктів 1, 4, 13, 14, 15, 16 (28,6%) були віднесені до інших кластерів. Це відмінний результат – метричний метод (рис. 6.8), коли була відома вся кількісна інформація, а не лише порядкова, визначив практично ті самі кластери і розпізнав лише на один об’єкт більше, ніж запропонований нами неметричний метод багатовимірного шкалування.



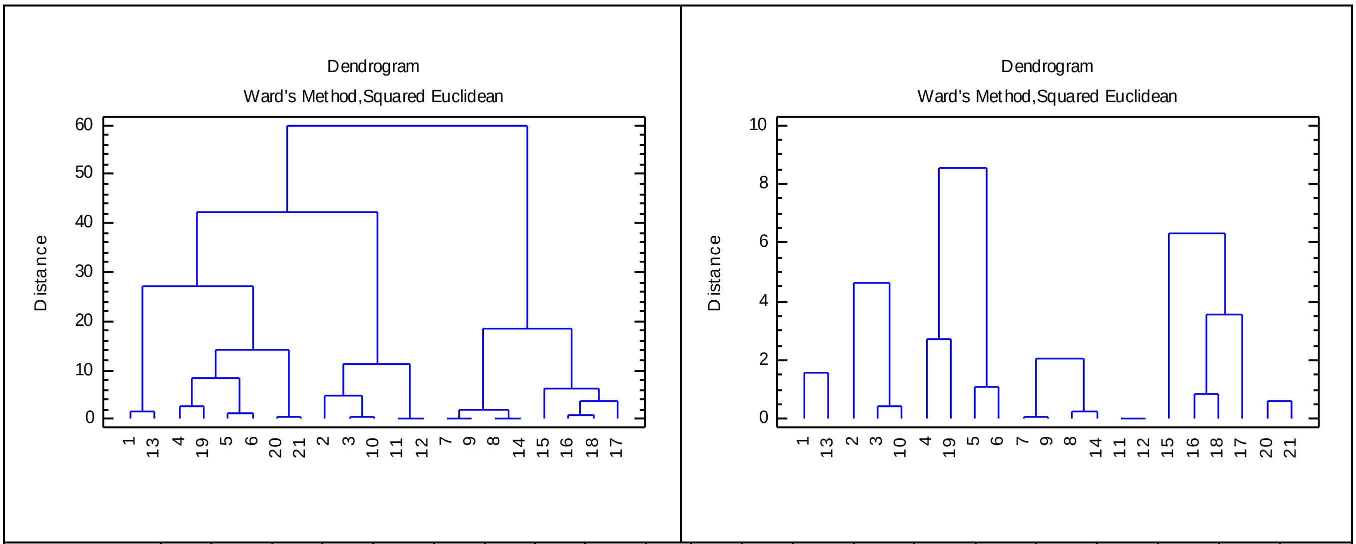
Кла-стер	1	1	1	2	2	2	2	2	3	3	4	4	4	4	4	5	5	5	6	7	7
Об’є-кти	1	4	13	2	3	10	11	12	5	6	7	8	9	14	16	15	17	18	19	20	21

Рис. 6.13. **Кластерний аналіз у скороченому просторі 3-х головних стимулів**

Дотепер у неметричному шкалуванні визнаним методом вважається ітераційний метод Краскала – Уолліса з неметричними етапами Гутмана – Лінгоса. Цей метод практично не обґрунтований, крім того, він надзвичайно громіздкий, вимагає багато нециклічних обчислень. Існують також інші різновиди методу Краскала – Уолліса з подібними недоліками. У програмному пакеті статистичних програм *Statistica* реалізовані два варіанти методу Краскала – Уолліса з кореляційною вихідною матрицею і з початковою матрицею відстаней, які обчислюються за формулою:

$$d_{ij} = \sqrt{1 - r_{ij}}\ .$$

За допомогою пакета *Statistica* були перевірені обидва варіанти цього методу розрахунками даних попереднього прикладу. У варіанті з кореляційною вихідною матрицею було зроблено 12 ітерацій і наприкінці одержано фінальну конфігурацію, за якою методом Уорда побудовані заключні дендрограми зі складом кожного кластера (рис. 6.14).



Кластер	1	1	2	2	2	3	3	3	3	4	4	4	4	5	5	6	6	6	6	7	7
Об'єкт	1	13	2	3	10	4	5	6	19	7	8	9	14	11	12	15	16	17	18	20	21

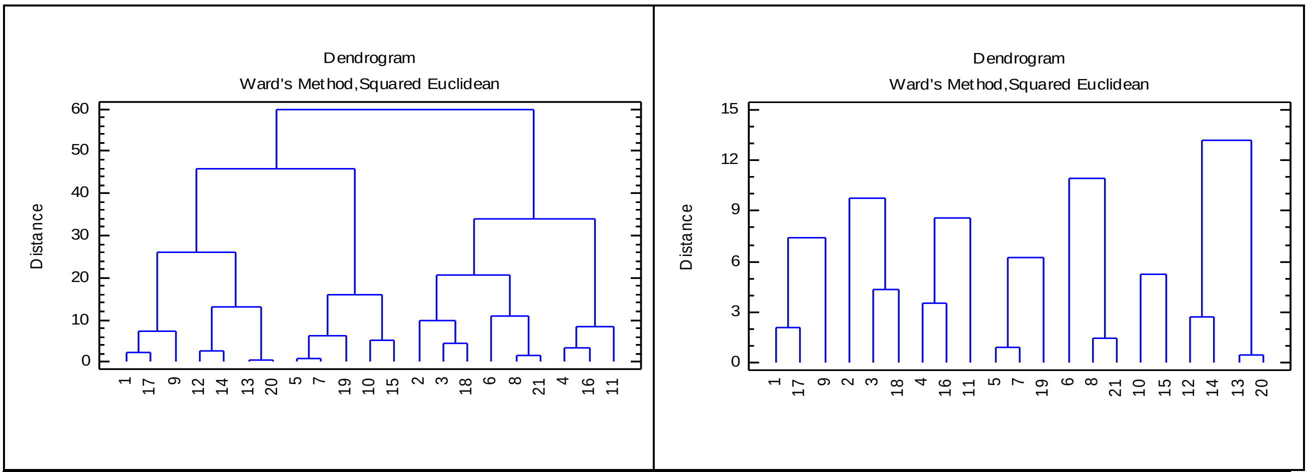
Рис. 6.14. **Розшарування об'єктів за першим варіантом методу Краскала – Віша**

Розглядаючи склад одержаних кластерів, зауважимо, що аналіз правильно визначив групи: (2, 3); (4, 5, 6); (7, 8, 9); (11, 12); 16, 17, 18); (20, 21). Шість об'єктів 1, 10, 13, 14, 15, 19 (28,6%) віднесені до інших класів. Цей результат не гірший, але й не кращий, ніж попередній, але є відносно простим методом, що пропонується як універсальний.

Другий варіант неметричного методу Краскала – Віша з початковою матрицею відстаней показав значно гірші результати (рис. 6.15).

Тепер процес закінчився через 35 ітерацій. Розглядаючи склад одержаних кластерів, підкреслимо погану відповідність результатів цього аналізу – правильно визначені лише дві пари об'єктів: (2,3) і (13,14) – лише 19% від 21%.

У методі Краскала – Віша немає стабільності, він може призводити до неправильних класифікацій. Отже, за методом Краскала – Віша можна отримати неправильну класифікацію.



Кластер	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6	7	7	7	7
Об'єкти	1	9	17	2	3	18	4	11	16	5	7	19	6	8	21	10	15	12	13	14	20

Рис. 6.15. **Розшарування об’єктів за другим варіантом методу Краскала – Віша**

Отже, викладені обґрунтування модифікації методу багатовимірного шкалування, які можна назвати універсальним методом багатовимірного шкалування.

Запитання для самоперевірки

1. Які відомі математичні методи можна використовувати для сумісної обробки величин ознак соціально-економічних систем, виміряних на різних шкалах?
2. У чому полягає суть модифікації факторного аналізу?
3. Які аналітичні можливості надає модифікований факторний аналіз?
4. Яке значення має модифікований факторний аналіз у багатовимірному аналізі соціально-економічних систем?
5. Які методи слід використовувати для класифікації об’єктів в економіці, що характеризуються ознаками, виміряними на різних шкалах?
6. Яке призначення кожного з алгоритмів методу багатовимірного шкалування?
7. У чому особливості методу метричного багатовимірного шкалування?
8. У чому особливості методу неметричного багатовимірного шкалування?
9. Які переваги має модифікація методу багатовимірного шкалування (універсальний метод багатовимірного шкалювання)?
10. Які можливості надає універсальний метод багатовимірного шкалування у проведенні багатовимірного аналізу соціально-економічних систем?

Розділ 7
Базисні описові моделі для аналізу соціально-економічних систем

7.1. Базисні моделі складних метричних ознак для аналізу соціально-економічних систем

Для проведення багатовимірного аналізу даних соціально-економічних систем слід мати формалізацію моделей складних ознак систем в операторній формі. У підрозділі 1.3 посібника мова йшла про типи моделей складних ознак, а саме: M_{CMO} – математичні моделі складних метричних ознак (моделі I типу); M_{CHMO} – математичні моделі складних неметричних ознак (моделі II типу); $M_{C(M+HM)O}$ – математичні моделі складних метрично-неметричних (сумісних систем) ознак (моделі III типу).

Моделі складних метричних ознак соціально-економічних систем M_{CMO} фундуються на вирішенні таких проблем [124; 128; 129; 134; 143; 149; 157]:

1) однорідності величини кожної ознаки;
2) ідентифікації загальних латентних складних метричних ознак всієї сукупності об’єктів, що є соціально-економічними системами;
3) однорідності сукупності об’єктів;
4) стійкості кластеризації об’єктів;
5) ідентифікації латентних складних ознак у кожному кластері об’єктів;
6) ідентифікації попарних систем складних ознак та їх зв’язків;
7) вимірювання складних метричних ознак.

Складні ознаки СЕС можуть бути відомими за допомогою теоретико-логічного аналізу, тоді моделювання виключає вирішення проблем ідентифікації, залишаючи решту. Проте актуальним залишається встановлення конструкції взаємозв’язків між ознаками.

Для формалізації моделей введемо наступні позначення.

Простір елементарних метричних ознак позначимо через X:

$$\boldsymbol{X} = \{x_{ij}\}, \quad i=\overline{1,n}, \quad j=\overline{1,m},$$

де n – кількість об’єктів; m – кількість метричних ознак;

або $$\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_j, ..., \boldsymbol{X}_m),$$

де $\boldsymbol{X}_j$ – величини метричної ознаки j для всіх об’єктів (стовпчики матриці $\boldsymbol{X}$).

Першою проблемою в розбудові математичних моделей складних метричних ознак є визначення однорідності величини елементарної ознаки. Якщо розподіл величини ознаки двомодальний, то дана ознака підрозділяється на дві різні ознаки – маємо класичну математичну задачу розділення суміші на компоненти, яка була розв’язана в підрозділі 2.3. Запропоноване окреме розв’язання
математичної задачі розділення суміші (ЗРС) має важливе значення для аналізу монотонності розподілу величин ознак. Згідно з даним розв’язанням на першому етапі передбачається використання методу Уорда кластерного аналізу для визначення кількості та складу однорідних підсукупностей. На другому етапі – визначення числових характеристик отриманих підсукупностей.

Отже, початкова кількість елементарних ознак СЕС може бути збільшена за рахунок введення одномодальних ознак. Позначимо оператор розділення суміші як $\mathbb{S}$, а множину одномодальних ознак $-$ $\boldsymbol{X}_0$ (рис. 7.1).

$$\boldsymbol{X} \longrightarrow \boxed{\;\mathbb{S}\;} \longrightarrow \boldsymbol{X}_0$$

Рис. 7.1. **Оператор розділення суміші**

Раніше обговорювалося твердження, що класифікації об’єктів стійкіші у скороченому просторі. Отже, другою проблемою у розбудові базисних моделей складних метричних ознак соціально-економічних систем є перетворення елементарних метричних ознак у фактори за допомогою факторного аналізу, що в операторній формі подається як:

$$\boldsymbol{X}_0 \xrightarrow{\mathscr{A}} \boldsymbol{F}^0,$$

де $\mathscr{A}$ – лінійний оператор перетворення елементарних ознак $\boldsymbol{X}_0$ у фактори $\boldsymbol{F}^0$:

$$\boldsymbol{F}^0 = \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \\ \ldots \\ \boldsymbol{U}_i \\ \ldots \\ \boldsymbol{U}_n \end{pmatrix},$$

де $\boldsymbol{U}_i$ – множина об’єктів (для метричних ознак – це значення всіх факторів для i -го об’єкта), рядки матриці $\boldsymbol{F}^0 \; [n \times p],\ p \le m$.

У підрозділі 2.2 наведено оцінку й рекомендації щодо вирішення класичних проблем факторного аналізу: проблеми факторів, проблеми спільностей, проблеми повороту, проблеми вимірювання факторів у сучасних умовах розвитку математики, комп’ютерної техніки та програмних засобів.

Третя проблема – проблема виділення однорідних підсукупностей об’єктів − вирішується за допомогою кластерного аналізу, який, як уже було зазначено, застосовується для розробки типології чи класифікації, дослідження раціональних обчислювальних методик групування об’єктів, формування гіпотез на основі дослідження даних, перевірки гіпотез чи підтвердження дійсного існування кластерів (груп) об’єктів у сукупності. Виконаний аналіз різних реалізацій кластерного аналізу дозволив рекомендувати лише окремі процедури даного математичного методу в моделюванні складних метричних ознак СЕС.

Введемо позначення для множини кластерів – множини однорідних груп об’єктів $\boldsymbol{G}$:

$$\boldsymbol{G} = \{\boldsymbol{G}_1, \boldsymbol{G}_2, \ldots, \boldsymbol{G}_k, \ldots, \boldsymbol{G}_g\},\ k = \overline{1, g},$$

де g – кількість кластерів.

Склад кластера k:

$$\boldsymbol{G}_k = \{\boldsymbol{U}_i \mid \boldsymbol{U}_i \in \omega_k\},$$

де ω_k – множина об’єктів $\boldsymbol{U}_i$, що належать кластеру $\boldsymbol{G}_k$.

У кластерному аналізі за загальними складними ознаками $\boldsymbol{F}$ визначають структуру сукупності об’єктів $\boldsymbol{G}$, тобто кількість і склад кластерів. В операторній формі це записується так (рис. 7.2):

$$\boldsymbol{F}^0 \longrightarrow \boxed{\mathscr{Y}} \longrightarrow \boldsymbol{G},$$

де $\mathscr{Y}$ – нелінійний оператор розпізнавання образів (без "учителя"):

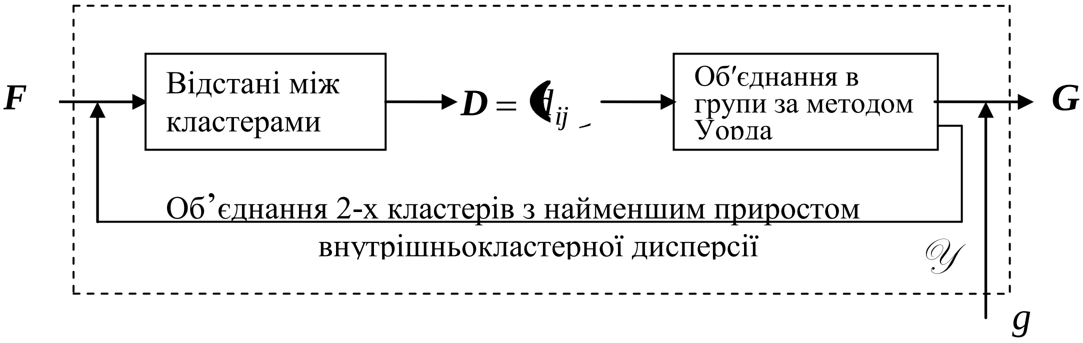


Рис. 7.2. **Операторна форма кластерного аналізу розшарування об’єктів у скороченому просторі**

Четверту проблему перевірки стійкості класифікацій рекомендується вирішувати за допомогою дискримінантного аналізу, який призначений для отримання правил класифікації об’єктів, спостережень в один з декількох встановлених кластерів чи груп. За допомогою дискримінантного аналізу підтверджується (або уточнюється) прийнята кількість кластерів g. Позначимо через $\boldsymbol{L}$ множину інформантів, яка отримується за допомогою дискримінантного аналізу (рис. 6.3), де $\mathscr{L}$ – оператор розпізнавання образів з "учителем".

де

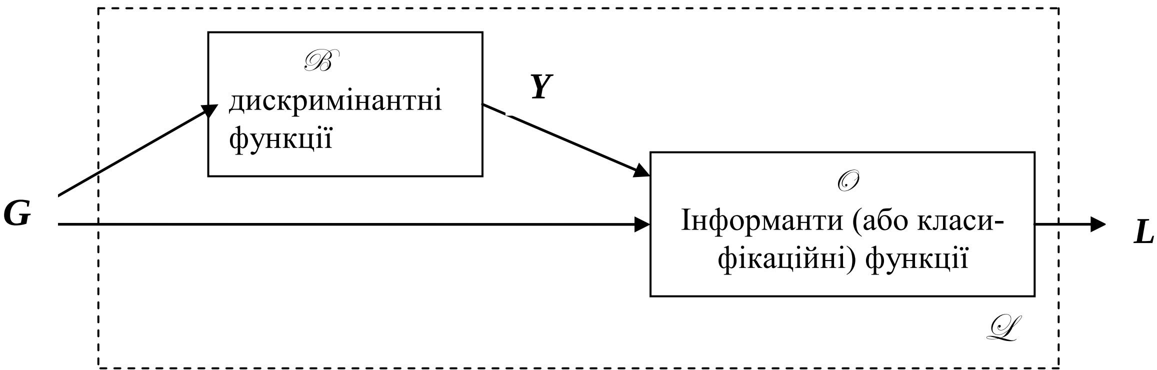


Рис. 7.3. **Операторна форма дискримінантного аналізу**

У більш детальній схемі слід позначити оператор визначення дискримінантних функцій

$$\boldsymbol{Y} = \boldsymbol{BX} \ ,$$

далі або за цими дискримінантними функціями обчислюються інформанти

$$\boldsymbol{L} = \boldsymbol{CY} \, ,$$

або безпосередньо через вихідні елементарні ознаки

$$\boldsymbol{L} = \boldsymbol{C}^{*}\boldsymbol{F} \, .$$

Наприкінці встановлюється належність кожного об’єкта кожній підсукупності за максимальною величиною інформантів (рис. 6.4):

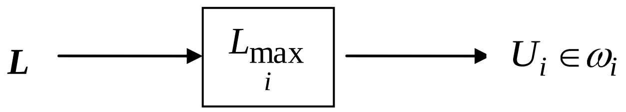


Рис. 7.4. **Операторна форма розпізнавання образів у дискримінантному аналізі**

Якщо у процесі вирішення проблеми перевірки стійкості прийнятої класифікації буде виявлена на дуже добра узгодженість, корегується кількість кластерів g .

У результаті вирішення п’ятої проблеми в кожному кластері виявляються латентні складні ознаки ***F***, які доповнюють простір ознак опису соціально-економічних систем та які спроможні за меншої кількості ознак СЕС досить об’єктивно подати великий масив інформації. На рис. 7.5 наведений оператор виявлення латентних складних метричних ознак у кожному кластері.

$$\boldsymbol{F} = \boldsymbol{AX}, \ \boldsymbol{X} \in \omega_k$$

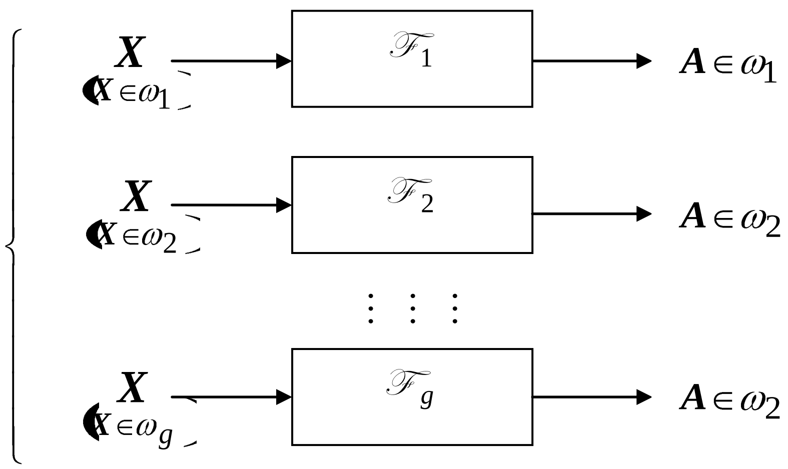


Рис. 7.5. **Операторна форма визначення структури латентних складних метричних ознак A у кожному кластері**

Формування моделі складних метричних ознак відбувається технологічно за етапами, а саме через доповнення новими складними ознаками, які призначені для встановлення існуючого механізму зв’язків між системами ознак СЕС.

У моделі наявні три види складних латентних ознак, що призначені для розв’язання суттєво різних проблем. Дискримінантні функції $\boldsymbol{Y}$ призначені для чіткішого зображення існуючого розшарування; латентні складні ознаки $\boldsymbol{A}$, що ідентифікуються факторним аналізом, призначені для скорочення простору елементарних ознак у кластеризації об’єктів і опису соціально-економічних систем у кожному кластері; латентні парні системи складних ознак $(\boldsymbol{U},\boldsymbol{V})$, що ідентифікуються методом канонічних кореляцій, призначені для виявлення механізму взаємозв’язку між ознаками СЕС. У підрозділі 2.1 для встановлення взаємозв’язків між системами ознак соціально-економічних систем був рекомендований метод канонічних кореляцій. У підрозділі 2.2 викладені аналітичні обґрунтування визначення залежності між двома системами випадкових величин, яка виражається за допомогою нових аргументів – канонічних величин, обчислених як лінійні комбінації початкових величин. Нові канонічні величини вибираються таким чином, щоб нові координати безпосередньо вказували значення кореляції [72, с. 270]. Нові канонічні величини визначають складні ознаки СЕС.

На рис. 7.6 наведений оператор взаємозв’язку пари складних ознак. Оскільки таких пар декілька, то це є система.

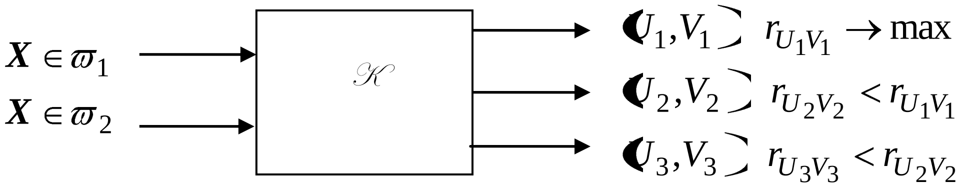


Рис. 7.6. **Операторна форма аналізу канонічних кореляцій (**ϖ_1, ϖ_2 **– взаємозалежні підсукупності метричних ознак)**

Загальна операторна схема ідентифікації складних ознак та визначення конструкції їх взаємозв’язків у розбудові базисних моделей складних метричних ознак соціально-економічних систем наведена на рис. 7.7.

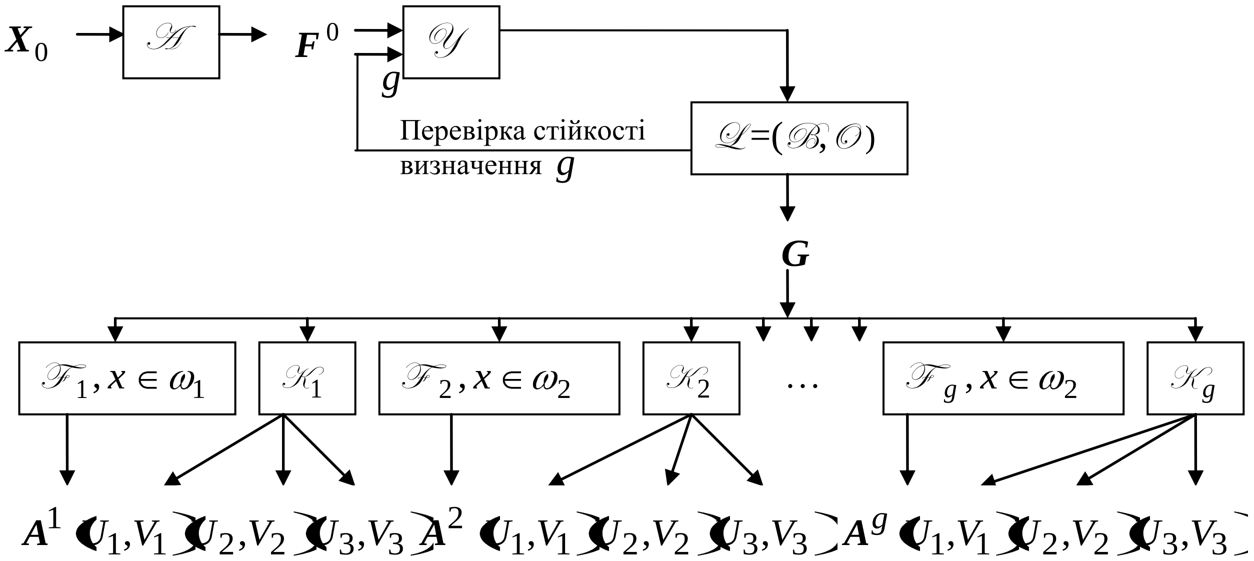


Рис. 7.7. **Загальна операторна форма ідентифікації складних ознак та визначення механізму їх взаємозв'язків**

Вирішення сьомої проблеми у розбудові складних метричних ознак пов'язане з їх вимірюванням за допомогою розроблених моделей вимірників, які викладені в підрозділі 4.2. Однією з основних процедур моделювання вимірників є функціональне перетворення величин елементарних метричних ознак, оператор якого позначимо *§,* та їх редукція, оператор якої, відповідно, позначимо 3. Слід ще раз вказати, що у процесі перетворень метричних величин існує проблема, яка обумовлена особливостями величин в економіці, а саме їх формою - показником. Вирішення даної проблеми зводиться до розв'язання однієї з трьох типів задач. Редукцію перетворених величин рекомендовано здійснювати за допомогою узагальнюючої функції мультиплікативного виду.

Перетворені значення елементарних метричних ознак позначимо $\boldsymbol{Z}$, а виміряні величини складних метричних ознак – $\boldsymbol{Z}^k$. Отже, безпосередню операцію вимірювання складних метричних ознак соціально-економічних систем відобразимо в операторі виду, що наведений на рис. 7.8.

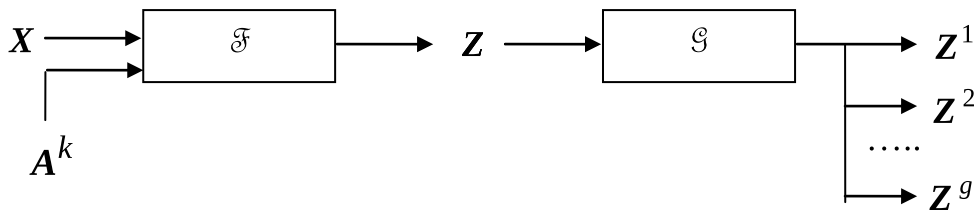


Рис. 7.8. **Оператор формування вимірників складних ознак**

Таким чином, для опису СЕС отримали наступну базисну математичну модель складних метричних ознак в операторній формі (рис. 7.9),

де $\boldsymbol{X} = \{x_{ij}\}$, $i=\overline{1,n}$, $j=\overline{1,m}$, – елементарні метричні ознаки (n – кількість об'єктів; m – кількість метричних ознак) або $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_j, ..., \boldsymbol{X}_m)$;

$\boldsymbol{X}_j$ – величини метричної ознаки j для всіх об'єктів (стовпчики матриці $\boldsymbol{X}$);

$\boldsymbol{X}_0$ – множина одномодально розподілених метричних ознак;

$\boldsymbol{A}$ – матриця факторних навантажень;

$\boldsymbol{F}^0$ – латентні фактори;

$\boldsymbol{G}$ – множина однорідних груп об'єктів $\boldsymbol{G}$: $\boldsymbol{G} = \{\boldsymbol{G}_1, \boldsymbol{G}_2, \ldots, \boldsymbol{G}_k, \ldots, \boldsymbol{G}_g\}$, $k=\overline{1,g}$, g – кількість кластерів. Склад кластера k: $\boldsymbol{G}_k = \{\boldsymbol{U}_i \mid \boldsymbol{U}_i \in \omega_k\}$, де ω_k – множина об'єктів $\boldsymbol{U}_i$, що належать кластеру $\boldsymbol{G}_k$;

$\boldsymbol{A}^k$ – матриці факторних навантажень у кожному кластері об'єктів;

$(\boldsymbol{U}^k, \boldsymbol{V}^k)$ – парні системи складних метричних ознак у кожному кластері об'єктів;

$\boldsymbol{Z}$ – перетворені значення елементарних метричних ознак;

$\boldsymbol{Z}^k$ – виміряні величини складних метричних ознак;

$\mathbb{S}$ – оператор розділення суміші на одномодальні підсукупності;

$\mathscr{F}$ – оператор ідентифікації загальних складних ознак;

$\mathscr{A}_0$ – оператор перетворення елементарних метричних ознак на фактори;

$\mathscr{Y}$ – оператор розпізнавання образів без "учителя";

$\mathscr{L}$ – оператор розпізнавання образів з "учителем";

$\mathscr{F}_k$ – оператор ідентифікації складних ознак у кожному кластері об'єктів;

$\mathscr{K}_k$ – оператор ідентифікації систем складних латентних ознак $(\boldsymbol{U}^k, \boldsymbol{V}^k)$ у кожному кластері об'єктів;

$\mathbb{F}$ – оператор нелінійного функціонального перетворення елементарних метричних ознак;

$\mathbb{G}$ – оператор узагальнення.

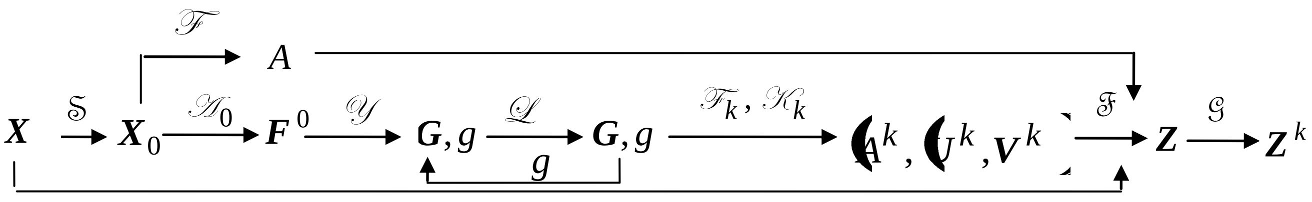


Рис. 7.9. **Базисна математична модель M_{CMO} складних метричних ознак соціально-економічних систем в операторній формі**

Розроблена модель дозволяє визначати різні складні ознаки соціально-економічних систем, які формуються з метричних елементарних ознак. За допомогою даної моделі генеруються нові знання в аналізі даних СЕС, що є фундаментальною основою системи підтримки управлінських рішень.

7.2. Базисні моделі складних неметричних (порядкових) ознак

З огляду на відсутність багатовимірних статистичних моделей, що розбудовані на неметричних величинах, і спираючись на уточнені теореми, що доведені в розділі 3 посібника, розроблені модифікації методів БСА – факторного аналізу та багатовимірного шкалування для моделювання СЕС. Моделі складних неметричних ознак для багатовимірного аналізу соціально-економічних систем слід формувати на вирішенні таких проблем [61; 124; 129; 140; 223]:

1) визначення міри парного взаємозв'язку неметричних ознак об'єкта в економіці, що є соціально-економічною системою;

2) ідентифікації загальних латентних складних неметричних ознак;

3) однорідності об'єктів;

4) ідентифікації існуючих латентних складних ознак у різних кластерах об'єктів;

5) вимірювання складних неметричних ознак.

Для моделювання використаємо позначення, що наведені в підрозділі 7.1.

Оскільки ідентифікувати загальні складні неметричні ознаки рекомендується за допомогою факторного аналізу, то необхідно обчислити матрицю коефіцієнтів парних взаємозв'язків між порядковими елементарними ознаками й на цій основі модифікувати сам математичний метод. У підрозділі 4.1 були доведені такі теореми:

1. Коефіцієнт рангової кореляції Спірмена (ρ) співпадає зі звичайним коефіцієнтом парної кореляції Пірсона (r_{pq}), обчисленим за рангами:

$$\rho = r_{pq} = \frac{S_{pq}}{S_p \cdot S_q},$$

де S_{pq} – коваріація рангів;

S_p – середньоквадратичне відхилення рангів ознаки p;

S_q – середньоквадратичне відхилення рангів ознаки q.

2. З узагальненої форми

$$R = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}}$$

за різних виборів a_{ij}, b_{ij} можна отримати всі відомі коефіцієнти кореляції:

− при $a_{ij} = x_i - x_j$ і $b_{ij} = y_i - y_j$ з узагальненої форми випливає коефіцієнт парної кореляції Пірсона;

− при $a_{ij} = p_i - p_j$ і $b_{ij} = q_i - q_j$ з узагальненої форми випливає коефіцієнт рангової кореляції Спірмена;

− при $a_{ij} = \mathrm{sgn}(p_i - p_j)$ і $b_{ij} = \mathrm{sgn}(q_i - q_j)$ з узагальненої форми випливає коефіцієнт рангової кореляції Кендела (τ).

3. Коефіцієнт конкордації співпадає з індексом детермінації, обчисленим за таблицею рангів y_{ij}.

4. Між середнім коефіцієнтом рангової кореляції Спірмена ($\overline{\rho}$) і коефіцієнтом конкордації Кендела (W) існує простий зв'язок:

$$\overline{\rho} = \frac{mW - 1}{m - 1},$$

де m − кількість спостережень.

5. Коефіцієнт асоціації Пірсона (Φ) дорівнює коефіцієнту контингенції Крамера (C) $\Phi = C$ при $p = q = 2$.

6. Коефіцієнт асоціації Пірсона (Φ) дорівнює коефіцієнту парної кореляції $|r_{uv}|$ $\Phi = |r_{uv}|$, якщо категорії "ні" присвоювати значення 0, а альтернативній категорії "так" − значення 1.

7. Для дихотомічних ознак коефіцієнт рангової кореляції Кендела (τ) співпадає з коефіцієнтом асоціації Пірсона (Φ) $\tau = \Phi$.

Звідси випливає, що коефіцієнт рангової кореляції Спірмена співпадає зі звичайним коефіцієнтом парної кореляції Пірсона, обчисленим за рангами. У підрозділі 5.1 викладений модифікований факторний аналіз, коли вихідними даними є матриця коефіцієнтів рангової кореляції Спірмена. Таким чином першим оператором у розбудові моделей складних неметричних (порядкових) ознак є нелінійний оператор $\mathscr{K}_x$ встановлення попарного взаємозв′язку між елементарними неметричними (порядковими) ознаками соціально-економічної системи ***X***, як це показано на рис. 7.10.

$$\boldsymbol{X} \longrightarrow \boxed{\mathscr{K}_x} \longrightarrow \boldsymbol{RO}_x$$

Рис. 7.10. **Оператор $\mathscr{K}_x$ оцінки тісноти взаємозв'язків між неметричними елементарними ознаками**

Матриця коефіцієнтів $\boldsymbol{RO_x}$ є вихідною інформацією для оператора ідентифікації загальних складних неметричних ознак $\mathscr{F}$ за допомогою модифікованого факторного аналізу з матрицею $\boldsymbol{RO_x}$ (рис. 7.11), де $\boldsymbol{A}$ – матриця факторних навантажень після обертання *VARIMAX,* яка призначена для визначення структури неметричних складних ознак.

$$\boldsymbol{RO_x} \longrightarrow \boxed{\mathscr{F}} \longrightarrow \boldsymbol{A}$$

Рис. 7.11. **Нелінійний оператор $\mathscr{F}$ утворення загальних складних неметричних ознак за початковою матрицею $\boldsymbol{RO_x}$**

Для вирішення проблеми однорідності сукупності об’єктів потрібно обчислити коефіцієнти рангової кореляції Спірмена між об’єктами, даний нелінійний оператор позначимо $\mathscr{X}_u$ (рис. 7.12).

$$\boldsymbol{X} \longrightarrow \boxed{\mathscr{X}_u} \longrightarrow \boldsymbol{RO_u}$$

Рис. 7.12. **Нелінійний оператор $\mathscr{X}_u$ оцінки тісноти взаємозв’язків між об’єктами**

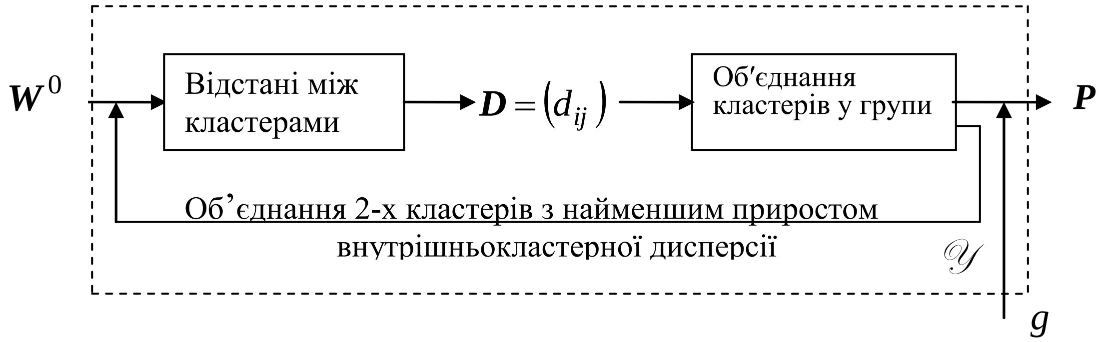
Матриця $\boldsymbol{RO_u}$ є вихідною для модифікованого методу багатовимірного шкалування, в операторній формі це наведено на рис. 7.13. Зазначимо, що матриці $\boldsymbol{RO_x}$ і $\boldsymbol{RO_u}$ є різними матрицями, і тому результати модифікованого факторного аналізу з даними матрицями будуть різними.

$$\boldsymbol{RO_u} \longrightarrow \boxed{\mathfrak{M}} \longrightarrow \boldsymbol{W}^0$$

Рис. 7.13. **Нелінійний оператор $\mathfrak{M}$ утворення стимулів $\boldsymbol{W}^0$ за початковою матрицею $\boldsymbol{RO_u}$**

У кластерному аналізі за стимулами $\boldsymbol{W}^0$ визначають структуру сукупності об’єктів $\boldsymbol{G}$, тобто кількість і склад кластерів. В операторній формі це записується таким нелінійним оператором (рис 7.14а), де $\mathscr{Y}$ – нелінійний оператор розпізнавання образів (без "учителя"). Детальніше подання кластеризації об’єктів в операторній формі наведене на рис. 7.14б:

а



б

Рис. 7.14. **Операторна форма кластерного аналізу розшарування об’єктів у скороченому стимульному просторі**

Для ідентифікації складних неметричних ознак у кожному кластері слід обчислити g матриць коефіцієнтів $\boldsymbol{RO}_x$, які є вихідними для оператора ідентифікації $\mathscr{F}_k$ за допомогою модифікованого факторного аналізу з матрицею $\boldsymbol{RO}_x$ (рис. 7.15). Тут $\boldsymbol{A}_k$ – матриця факторних навантажень після обертання *VARIMAX*, яка призначена для ідентифікації неметричних складних ознак у кожному кластері. У результаті вирішення четвертої проблеми в кожному кластері виявляються латентні складні ознаки A_k, які доповнюють простір ознак опису соціально-економічних систем та які за меншої їх кількості досить об’єктивно описують великий масив інформації. На рис. 7.15 наведений оператор виявлення латентних складних неметричних ознак у кожному кластері за вихідними неметричними елементарними ознаками.

Вирішення п’ятої проблеми у розбудові складних неметричних ознак пов’язане з їх вимірюванням за допомогою розробленої моделі вимірників неметричних ознак, яка викладена в підрозділі 4.3. Однією з основних процедур у даному моделюванні є функціональне перетворення елементарних неметричних ознак (оператор якого позначений $\mathfrak{F}$) та узагальнення або редукція (оператор якого відповідно позначимо $\mathfrak{G}$). Перетворені значення елементарних неметричних ознак позначимо $\boldsymbol{Z}$, а виміряні величини складних неметричних ознак − $\boldsymbol{Z}_k$.

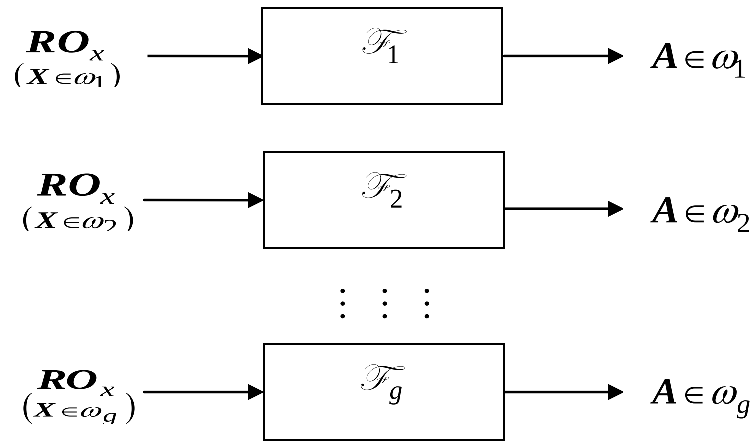


Рис. 7.15. **Операторна форма визначення латентних складних неметричних ознак A в кожному кластері**

Отже, безпосередню операцію вимірювання складних неметричних ознак соціально-економічних систем відобразимо в операторі виду, що зображений на рис. 7.16.

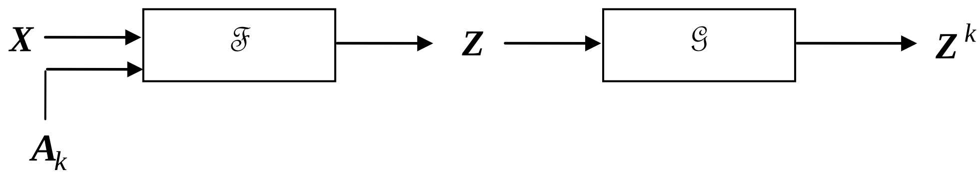


Рис. 7.16. **Оператор вимірювання складних неметричних ознак**

Таким чином, базисна математична модель складних неметричних ознак соціально-економічних систем в операторній формі наведена на рис. 7.17.

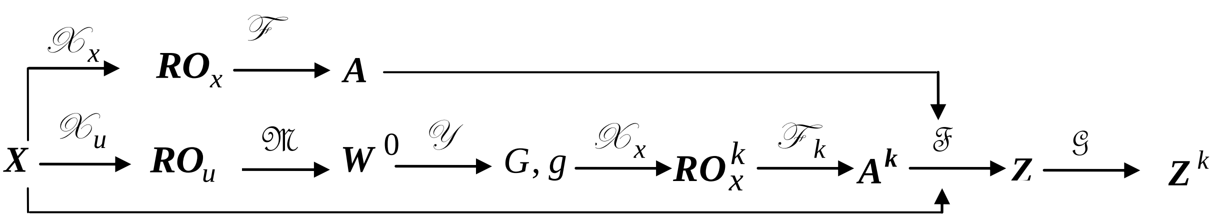


Рис. 7.17. **Базисна математична модель** M_{CHMO} **складних неметричних (порядкових) ознак соціально-економічних систем в операторній формі**

Зауваження. Де $\boldsymbol{X} = \{x_{ij}\}$, $i = \overline{1,n}$, $j = \overline{1,m}$, − елементарні неметричні (порядкові) ознаки (n – кількість об'єктів; m – кількість порядкових ознак);

$\boldsymbol{RO}_x$ – матриця коефіцієнтів рангової кореляції Спірмена між порядковими ознаками;

$\boldsymbol{RO}_u$ – матриця коефіцієнтів рангової кореляції Спірмена між об'єктами;

$\boldsymbol{A}$ – матриця факторних навантажень;

$\boldsymbol{W}^0$ – латентні складні стимули (аналоги метричних факторів);

$\boldsymbol{RO}_x^k$ – матриці коефіцієнтів рангової кореляції Спірмена в кожному кластері об'єктів;

$\boldsymbol{A}^k$ – матриці факторних навантажень у кожному кластері;

$\boldsymbol{Z}$ – перетворені значення елементарних неметричних ознак;

$\boldsymbol{Z}^k$ – виміряні величини складних неметричних ознак;

$\mathscr{K}_x$ – оператор оцінки тісноти взаємозв'язків між порядковими ознаками;

$\mathscr{K}_u$ – оператор оцінки тісноти взаємозв'язків між об'єктами;

$\mathscr{F}$ – оператор ідентифікації загальних складних ознак за допомогою модифікованого факторного аналізу;

$\mathfrak{M}$ – оператор перетворення неметричних ознак у метричні стимули за допомогою модифікованого методу багатовимірного шкалування;

$\mathscr{Y}$ – оператор розпізнавання образів без "учителя";

$\mathscr{F}_k$ – оператор ідентифікації складних ознак за допомогою модифікованого факторного аналізу в кожному кластері об'єктів;

$\mathfrak{F}$ – оператор функціонального перетворення елементарних неметричних ознак;

$\mathfrak{G}$ – оператор узагальнення.

Моделі складних неметричних ознак соціально-економічних систем, що налаштовуються з даної базисної моделі, розширюють можливості аналізу даних СЕС за рахунок опису її якісними характеристиками, які виражаються в неметричних величинах, виміряних у порядкових шкалах.

7.3. Базисні моделі складних сумісних ознак для аналізу соціально-економічних систем

У науковій літературі майже відсутні пропозиції щодо багатовимірного статистичного моделювання соціально-економічних систем, коли вони сумісно описуються метричними і неметричними ознаками. Для розбудови даного типу базисних моделей була розроблена та викладена в посібнику міра тісноти взаємозв'язку між різними ознаками CCK (описана в підрозділі 2.3), за допомогою якої були модифіковані факторний аналіз та метод багатовимірного шкалування. Моделі складних сумісних ознак слід формувати на вирішенні таких проблем [61; 124; 129; 140; 223]:

1) визначення міри парного взаємозв'язку різних елементарних ознак об'єкта в економіці, що є соціально-економічною системою;
2) ідентифікації загальних латентних складних сумісних ознак;
3) однорідності об'єктів;
4) ідентифікації латентних складних сумісних ознак у кожному кластері об'єктів;
5) вимірювання складних сумісних ознак.

Для моделювання використаємо позначення, наведені в підрозділах 7.1 та 7.2.

З проведеного аналізу відомих числових мір тісноти взаємозв'язку номінальних ознак СЕС у розділі 3 випливає висновок, що в наведеному існуючому комплексі ці різні міри тісноти зв'язку не узгоджені між собою. Було визначено, що найбільш обґрунтованими серед них є коефіцієнт контингенції Крамера C і коефіцієнт контингенції Кендела $\boldsymbol{K}$, хоча вони дають дещо різні значення. Доводилось, що скоректований коефіцієнт $CK=\sqrt{\frac{\chi^2}{\chi^2+n}\cdot\frac{d}{d-1}}$ (де $d=\min(p,q)$, n – сума всіх частот) і коефіцієнт контингенції Крамера $C=\sqrt{\frac{\chi^2}{n(d-1)}}$ співпадають лише для граничних ситуацій ($CK=C=0$, якщо немає жодного зв'язку; $CK=C=1$, якщо є функційне співвідношення, коли кожній категорії X відповідає одна конкретна категорія Y). Для всіх проміжних ситуацій коефіцієнт CK дає систематично більше значення, ніж коефіцієнт C $(CK>C)$. Ці коефіцієнти зв'язані функціональною залежністю $CK=\sqrt{\frac{C^2 d}{C^2(d-1)+1}}$.

Було запропоновано для оцінки тісноти взаємозв'язку ознак визначати коефіцієнт CCK, що є середнім геометричним двох мір C і CK:

$$CCK = \sqrt{C \cdot CK}\,.$$

Розроблена об'єктивна, універсальна міра CCK тісноти взаємозв'язку ознак, що виміряні на будь-яких шкалах, надає унікальні можливості виявити та оцінити всю конструкцію зв'язків різних ознак СЕС. Універсальна міра може бути обчислена для будь-яких метричних і неметричних величин. Об'єктивна універсальна міра тісноти взаємозв'язку між ознаками CCK має принципове значення, оскільки надає широкі можливості для модифікації математичних методів, зокрема методів БСА, коли вихідними даними є різні величини, а отже продовження розбудови моделей складних ознак СЕС із сумісним складом метричних і неметричних елементарних ознак.

Таким чином, першим оператором у розбудові моделей складних неметричних ознак є нелінійний оператор встановлення попарного взаємозв'язку між різними елементарними ознаками соціально-економічної системи ***X***, як це подано на рис. 7.18.

$$\boldsymbol{X} \longrightarrow \boxed{\mathscr{R}_x} \longrightarrow \boldsymbol{CCK_X}$$

Рис. 7.18. **Нелінійний оператор $\mathscr{R}_x$ оцінки тісноти взаємозв'язків між різними елементарними ознаками**

Оскільки коефіцієнти CCK використовують для оцінки тісноти взаємозв'язку різних типів ознак (виміряних на метричних чи неметричних шкалах), то матриця даних коефіцієнтів є вихідною інформацією для оператора ідентифікації складних сумісних ознак за допомогою модифікованого факторного аналізу (рис. 7.19). Тут ***A*** – матриця факторних навантажень після обертання *VARIMAX*, яка призначена для ідентифікації структури складної ознаки.

$$\boldsymbol{CCK_x} \longrightarrow \boxed{\mathscr{F}} \longrightarrow \boldsymbol{A}$$

Рис. 7.19. **Нелінійний оператор $\mathscr{F}$ утворення латентних складних ознак за початковою узагальненою матрицею $\boldsymbol{CCK_x}$**

Для вирішення проблеми однорідності об’єктів за системою різних за типом величин ознак рекомендується модифікація методу багатовимірного шкалування, викладена в підрозділі 5.3, яка бере за вихідні дані матрицю $\boldsymbol{CCK_u}$ − коефіцієнтів тісноти взаємозв’язку між об’єктами $\boldsymbol{U}$. На рис. 7.20 наведений оператор $\mathscr{R}_u$ оцінки тісноти взаємозв’язків між об’єктами.

$$\boldsymbol{X} \longrightarrow \boxed{\mathscr{R}_u} \longrightarrow \boldsymbol{CCK_u}$$

Рис. 7.20. **Нелінійний оператор $\mathscr{R}_u$ оцінки тісноти взаємозв’язків між об’єктами**

За допомогою розробленої модифікації методу неметричного багатовимірного шкалування з матрицею $\boldsymbol{CCK_u}$ отримуються латентні складні ознаки, які в методі шкалування традиційно називаються стимулами $\boldsymbol{S}$ (рис. 7.21). Отримані таким чином ознаки вже є метричними (це і є результат оператора шкалування $\mathfrak{M}$).

$$\boldsymbol{CCK_u} \longrightarrow \boxed{\mathfrak{M}} \longrightarrow \boldsymbol{S}$$

Рис. 7.21. **Оператор $\mathfrak{M}$ утворення складних ознак (стимулів) з елементарних ознак незалежно від типу їх величин**

Тут $\boldsymbol{S}$ – матриця латентних складних ознак (стимулів).

Проблема однорідності об’єктів за різними елементарними ознаками вирішується за допомогою кластерного аналізу в скороченому метричному стимульному просторі $\boldsymbol{S}$ (рис. 7.22), де $\boldsymbol{Q}$ – множина латентних складних метричних ознак (стимулів).

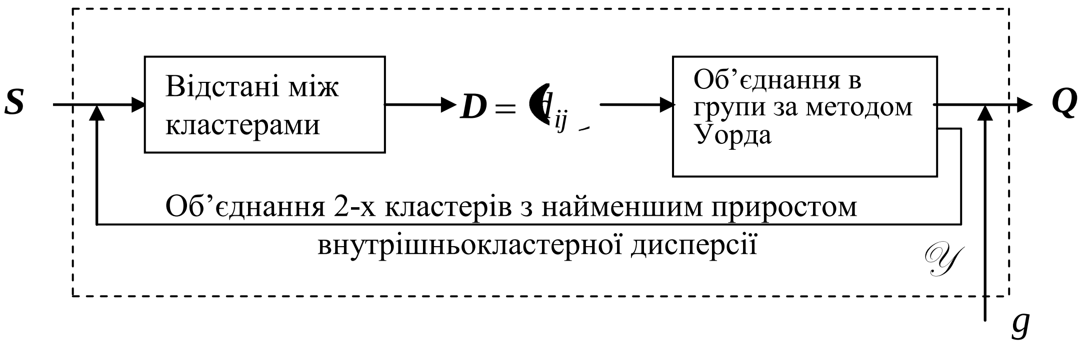


Рис. 7.22. **Операторна форма кластерного аналізу розшарування об'єктів у скороченому стимульному просторі**

Проблема ідентифікації існуючих латентних складних ознак в однорідних групах об'єктів (кластерах) вирішується за допомогою модифікованого факторного аналізу в кожному з виділених кластерів. В операторній формі це наведено на рис. 7.23.

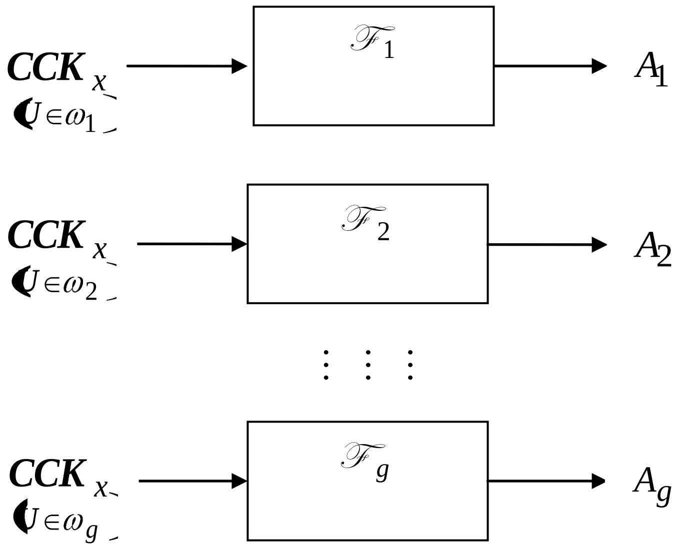


Рис. 7.23. **Операторна форма модифікованого факторного аналізу з матрицею CCK_x у кожному кластері однорідних груп об'єктів $U \in \omega$**

П'ята проблема вимірювання складних ознак, що утворені з різних елементарних ознак, вирішується за допомогою моделювання вимірників, яке описане в підрозділі 4.3. Функціональне перетворення елементарних ознак рекомендується здійснювати за допомогою функцій перетворення. Оператор функціонального перетворення елементарних ознак позначимо $\mathscr{F}$, а оператор узагальнення $-\ \mathscr{G}$, виміряні величини складних ознак $-\ Z^k$ (рис. 7.24)

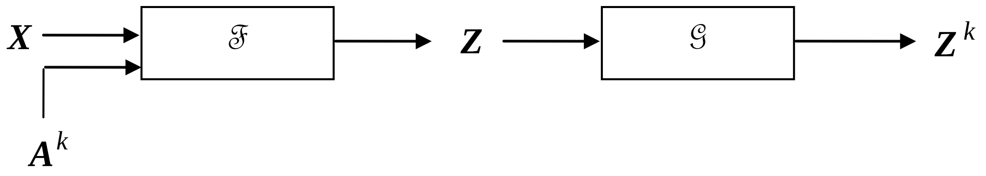


Рис. 7.24. **Оператор вимірювання складних ознак**

Таким чином, узагальнюючи всі оператори, що відображали вирішення проблем у процесі розбудови математичної моделі складних ознак, що утворені з різних метричних і неметричних величин, маємо базисну модель, яка в операторній формі наведена на рис. 7.25.

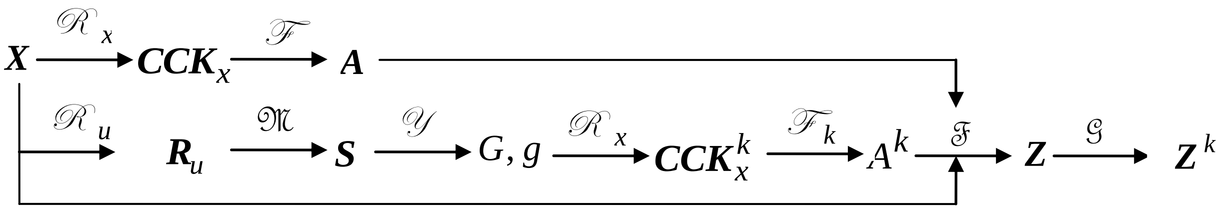


Рис. 7.25. **Базисна математична модель $M_{С\ М+НМ\ О}$ складних сумісних ознак соціально-економічних систем в операторній формі**

Зауваження. Тут:

$\boldsymbol{CCK}_x$ – матриця коефіцієнтів кореляції CCK між різними ознаками;

$\boldsymbol{A}$ – матриця факторних навантажень;

$\boldsymbol{R}_u$ – матриця коефіцієнтів узагальненого зв’язку між різними об’єктами;

$\boldsymbol{S}$ – латентні складні стимули;

$\boldsymbol{CCK}_x^k$ – матриці CCK у кожному кластері об’єктів;

$\boldsymbol{A}^k$ – матриці факторних навантажень у кожному кластері;

$\mathscr{R}_x$ – оператор оцінки тісноти взаємозв’язків між різними елементарними ознаками;

$\mathscr{F}$ – оператор ідентифікації складних латентних ознак за допомогою модифікованого факторного аналізу з матрицею $\boldsymbol{CCK}_x$;

$\mathscr{R}_u$ – оператор оцінки тісноти взаємозв’язків між об’єктами;

$\mathfrak{M}$ – оператор перетворення елементарних ознак у метричні стимули за допомогою модифікованого методу багатовимірного шкалування;

$\mathscr{F}_k$ – оператор ідентифікації складних сумісних ознак за допомогою модифікованого факторного аналізу в кожному кластері;

$\mathfrak{F}$ – оператор функціонального перетворення різних елементарних ознак;

$\mathfrak{G}$ – оператор узагальнення.

Моделі складних ознак соціально-економічних систем, що налаштовуються з даної базисної моделі, цілісно сумісно описують систему за її кількісними та якісними характеристиками, що виражаються в метричних і неметричних величинах, при цьому надається виключна можливість генерації знань на основі дійсно повномасштабного опису СЕС.

Розглянемо стисло стійкість розбудованих математичних моделей складних ознак соціально-економічних систем, детальніше ця проблема викладена в роботі [166]. Основним критерієм досягнення стійкості в рекомендованому моделюванні соціально-економічних систем є повторна перевірка застосувань математичних методів за рахунок дублювання. Посиленням основного критерію є стійкість окремих математичних методів, що використовувалися в розбудові моделей.

Для отримання стійкої класифікації об'єктів кластеризацію слід проводити за складними ознаками, виявленими за допомогою факторного аналізу. Отже, тут проблеми стійкості пов'язані з проблемою стійкості у факторному аналізі.

Стійкість факторного розв'язку забезпечується, по-перше, критерієм значущості, який застосовується для кожного факторного розв'язання, як початкового, так і проміжного та кінцевого; по-друге, визначенням коефіцієнта надійності методом максимальної правдоподібності, запропонованим Такером і Левісом [214]. Даний підхід базується на використанні окремих коефіцієнтів кореляції. Обчислюється коефіцієнт надійності за формулою:

$$rho = \frac{E_o - E_k}{E_o - 1},$$

де E_o – математичне сподівання статистики χ^2 за відсутності впливу факторів, яке ділиться на $\frac{1}{2}n(n-1)$;

E_k – математичне сподівання χ^2 для кінцевого факторного розв'язку, яке ділиться на $\frac{1}{2}\left[(n-r)^2-(n+r)\right]$.

Значення коефіцієнта rho змінюється від 0 до 1, якщо маємо 0, то наявне найгірше узгодження моделі з даними, а якщо 1 – то найкраще.

Рекомендується також перевіряти стійкість факторного розв'язання емпіричними підтвердженнями, а саме вибірковою адекватністю. Емпіричні підтвердження зростають, якщо: 1) зростає кількість змінних; 2) зменшується кількість загальних факторів; 3) зменшуються окремі коефіцієнти кореляції; 4) збільшується коефіцієнт детермінації. Перші дві умови гарантують збільшення емпіричних обмежень, що висуваються факторною моделлю на експериментальні дані. Чим більший обсяг вибірки, тим точніша χ^2-апроксимація. Лоулі і Максвелл вважають, що цей критерій застосовується, коли вибірка містить, принаймні, на 51 спостереження більше, ніж кількість змінних, тобто ця умова наступна: $N - n - 1 \geq 50$, де N – обсяг вибірки, а n – кількість змінних.

Терстоун вважає, що на один фактор повинно припадати, принаймні, три змінні. Загальна ж думка, що кількість змінних має бути вдвічі більшою, ніж кількість факторів.

Третя умова відноситься до оцінювання близькості коефіцієнтів кореляції величин ознак, що спостерігаються, з величинами, що відтворюються. Четверта умова полягає в тому, що збільшується частка спільності і дисперсії кожної ознаки, що спостерігається.

Емпіричний критерій вибіркової адекватності був запропонований Кайзером, який назвав його "мірою вибіркової адекватності" (МВА). Значення МВА обчислюється за формулою:

$$MBA = \frac{\sum\limits_{j \neq k}\sum r_{jk}}{\sum\limits_{j \neq k}\sum r_{jk}^2 + \sum\limits_{j \neq k}\sum q_{ik}^2},$$

де r_{ij} – коефіцієнти кореляції, що спостерігаються;

q_{ij} – елементи кореляційної матриці антиобразів, яка задається виразом

$$Q = SR^{-1}S,$$

де R^{-1}– обернена кореляційна матриця,

$$S = \left(\mathrm{diag}\, R^{-1}\right)^{-\frac{1}{2}}.$$

MBA набуває значень від 0 до 1. Значення $MBA = 1$ маємо, коли всі недіагональні елементи матриці, оберненої до кореляційної Q, – нульові, що змістовно пояснює факт – кожна змінна може бути виражена без помилок через решту змінних. Порогові змінні MBA за Кайзером такі: більші, ніж 0,9, – відмінно; більші, ніж 0,8, – добре; більші, ніж 0,7, – середньо; більші, ніж 0,6, – посередньо; більші, ніж 0,5, – погано; нижчі, ніж 0,5, – неприйнятно. Кайзер довів, що MBA збільшується у випадках: 1) зростання кількості змінних; 2) зменшення числа загальних факторів; 3) збільшення обсягу статистики; 4) збільшення середнього значення коефіцієнтів кореляції.

Проте кінцевий висновок про ступінь емпіричного підтвердження факторної моделі дослідник робить залежно від адекватності розв'язку змісту практичної задачі.

Стійкість розв'язків кластерного аналізу забезпечується завдяки методів перевірки достовірності (обґрунтованості) розв'язків. У підрозділі 2.2 обговорювалися проблеми стійкості класифікацій. Чотири фактори впливають на методи класифікації: характеристики кластерної структури; наявність викидів і ступінь повноти класифікації; ступінь перекриття кластерів; вибір міри схожості. Найважливішими характеристиками кластерної структури, що впливають на якість кластеризації, є форма кластерів, розміри кластерів і кількість кластерів. На кластерізацію знову-таки впливають викиди. Але викиди потрібно аналізувати та не поспішати усувати. Для цього в дослідженні використовували інструменти описової статистики для дослідження величин ознак. На кластеризацію впливає перекриття кластерів, що є наслідком існування об'єктів, що містяться на межі кластерів. Як відомо, в даному випадку метод Уорда працює краще більшості інших методів кластеризації. На роботу методів кластеризації впливає вибір міри схожості. У методі Уорда обчислюється евклідова відстань.

Для перевірки обґрунтованості результатів кластерного аналізу рекомендують застосовувати п'ять основних методів: кофенетичну кореляцію, тести значущості для ознак, за якими виконується кластеризація, повторну вибірку, тести значущості для незалежних ознак, методи Монте-Карло [214]. Перший метод застосовується тільки разом з ієрархічним агломеративним методом. Кофенетична кореляція має істотні недоліки: застосування змішаного моменту кореляції передбачає, що нормально розподілені значення у двох матрицях корельовані. Це передбачення зазвичай не виконується для значень вторинної матриці схожості, оскільки кластерні методи значною мірою визначають розподіл значень схожості в цій матриці. Тому застосування коефіцієнта кореляції для оцінки ступеня схожості між значеннями двох матриць не є оптимальним. Другим недоліком є те, що кількість різних значень у другій матриці схожості менша, ніж у початковій, отже, кількість інформації, що міститься в кожній із двох матриць, надзвичайно різна.

Другим методом для перевірки обґрунтованості розв’язків кластерного аналізу є тести значущості для ознак, що необхідні в процесі створення кластерів. Застосування багатовимірного дисперсійного аналізу (MANOVA) має на меті перевірку гіпотези однорідності, наскільки значуще розбиття даних на кластери. Даний метод перевірки можна застосовувати для різних методів кластеризації. Застосування дискримінантного аналізу в розбудові моделі метричних складних ознак має за мету виявлення стійкості класів об’єктів, отриманих у результаті використання кластерного аналізу на основі складних ознак – факторів, що виявили факторним аналізом.

Третій метод перевірки достовірності розв’язків кластерного аналізу – повторна вибірка – дозволяє оцінити ступінь повторення кластерного розв’язку в серії наборів даних. Якщо для різних вибірок з однієї й тієї ж генеральної сукупності отримується однаковий кластерний розв’язок, то можна робити висновок, що розв’язок наявний для всієї сукупності, але те, що кластери виявляються однаковими кластерними методами в різних підмножинах, не доводить обґрунтованості розв’язків.

Тести значущості для зовнішніх ознак передбачають порівняння кластерів за ознаками, що не застосовувалися у процесі отримання кластерного рішення. Дані тести погано відпрацьовані.

Підхід до обґрунтування достовірності розв’язків кластерного аналізу, що базується на процедурах Монте-Карло, передбачає генератори випадкових чисел для створення наборів даних з основними характеристиками, які відповідають характеристикам реальних даних без кластерів. Методи кластеризації використовуються для реальних даних і штучних, результати порівнюються.

Кожний із наведених методів має свої недоліки, але тести багатовимірного статистичного аналізу, зокрема дискримінантного аналізу, у сучасних умовах розвитку програмних засобів найдоцільніші у застосуванні для перевірки достовірності кластерного розв’язку.

Як і більшість евристичних методів багатовимірного статистичного аналізу, дискримінантний аналіз передбачає стійкі розв’язки, якщо чітко виконуються його гіпотези, але реальні дані в економіці мають багато вад з точки зору статистичних вимог, і тому окремі гіпотези порушуються, навіть коли початкові дані перевірені інструментами описової статистики (виявлені вади початкових даних не завжди підлягають коректуванню в економіці). Схемою обчислень дискримінантного аналізу передбачаються тести та перевірки значущості отриманих результатів, як це було описано в підрозділах 2.2 та 2.3.

Однією з основних вимог дискримінантного аналізу є вимога про нормальність багатовимірного розподілу дискримінантних змінних та рівності коваріаційних матриць класів. Але, як уже говорилось раніше, припущення про нормальність відіграє важливу роль у класифікації, що базується на використанні ймовірності належності до класу. Ці ймовірності обчислюються, виходячи з розподілу χ^2, що виправдано лише тоді, коли дискримінантні змінні мають багатовимірний нормальний розподіл. Якщо це припущення не виконується, імовірності, що обчислюються, будуть неточними.

Якщо коваріаційні матриці класів не однакові, виникає спотворення дискримінантних функцій і рівнянь класифікації. Похибки пов’язані з обчисленням внутрішньогрупової коваріаційної матриці. Внутрішньогрупова коваріаційна матриця є оцінкою загальної коваріаційної матриці класів для генеральної сукупності, що утворена вибірками з декількох класів.

Тести значущості раціонально застосовувати для малих вибірок. Вважається, що велика кількість відсутніх даних, сильно корельовані ознаки, ознаки з нульовим стандартним відхиленням усередині одного чи декількох класів, значні відмінності в розмірах класів і викиди – все це негативно впливає на ефективність переоцінки класифікацій. Але частина негативних впливів наразі відпадає, оскільки дискримінантний аналіз виконується на факторах.

На рис. 7.26 наведений комплекс математичних методів, які окремо й цілісно підтримують стійкість розбудованих моделей базису опису соціально-економічних систем. Сама логіка використання наведених математичних методів забезпечує подвійний контроль за стійкістю розв’язків [180].

Виконання всіх передумов розбудови моделей складних ознак соціально-економічних систем проблемне, але їх знання та коректна реалізація математичних методів забезпечує стійкість та адекватність моделей. Звідси випливає, що ухвалені управлінські рішення на основі такого багатовимірного аналізу є дійсно науково обґрунтованими.

Рис. 7.26. **Основні складові в перевірці стійкості моделей базису опису соціально-економічних систем**

Задачі до розділу 7

За даними, що характеризують виробничо-господарську діяльність 15 промислових підприємств протягом 5 років (додаток А), розв'язати задачі багатовимірного аналізу соціально-економічних систем (умовне позначення табл. 3.6).

1. Побудувати модель складних метричних ознак виробничо-господарської діяльності сукупності підприємств.

2. Побудувати моделі складних метричних ознак виробничо-господарської діяльності підприємств у кластерах, які отримали при вирішенні задач розділу 3.

3. Сформувати візуальне подання отриманих результатів багатовимірного аналізу виробничо-господарської діяльності підприємств та виокремити нові знання про діяльність підприємств.

Запитання для самоперевірки

1. На вирішенні яких проблем формуються моделі складних метричних ознак соціально-економічних систем (M_{CMO})?

2. Розкрити зміст кожної з проблем формування моделей складних метричних ознак соціально-економічних систем (M_{CMO}).

3. Які переваги мають моделі складних метричних ознак соціально-економічних систем?

4. На вирішенні яких проблем формуються моделі складних неметричних (порядкових) ознак соціально-економічних систем (M_{CHMO})?

5. Розкрити зміст кожної з проблем формування моделей складних метричних ознак соціально-економічних систем (M_{CHMO}).

6. Які переваги мають моделі складних метричних ознак соціально-економічних систем?

7. На вирішенні яких проблем формуються моделі складних сумісних ознак соціально-економічних систем ($M_{C(M+HM)O}$)?

8. Розкрити зміст кожної з проблем формування моделей складних сумісних ознак соціально-економічних систем ($M_{C(M+HM)O}$).

9. Які переваги мають моделі складних сумісних ознак соціально-економічних систем?

10. Як налаштовуються базисні моделі на вирішення реальних задач багатовимірного аналізу соціально-економічних систем?

Використана література

1. Адлер Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер, Е. В. Маркова, Ю. В. Грановский. – М. : Наука, 1976. – 280 с.
2. Азгальдов Г. Г. О квалиметрии / Г. Г. Азгальдов, Э. П. Райхман. – М. : Издательство стандартов, 1973. – 172 с.
3. Айвазян С. А. Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. – М. : Статистика, 1974. – 240 с.
4. Айвазян С. А. Прикладная статистика: Основы моделирования и первичной обработки данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1983. – 472 с.
5. Айвазян С. А. Прикладная статистика: Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин ; под ред. С. А. Айвазяна. – М. : Финансы и статистика, 1985. – 488 с.
6. Андерсон Т. Введение в многомерный статистический анализ / Т. Андерсон ; пер. с англ. Ю. Ф. Кичатова. – М. : Физматгиз, 1963. – 500 с.
7. Андрейчиков А. В. Анализ, синтез, планирование решений в экономике / А. В. Андрейчиков, О. Н. Андрейчикова. – М. : Финансы и статистика, 2000. – 368 с.
8. Анфилатов В. С. Системный анализ в управлении : [учеб. пособие] / В. С. Анфилатов, А. А. Емельянов, А. А. Кукушкин ; под ред. А. А. Емельянова. – М. : Финансы и статистика, 2002. – 368 с.
9. Аптон Г. Анализ таблиц сопряженности / Г. Аптон ; пер. с англ. и предисл. Ю. П. Адлера. – М. : Финансы и статистика, 1982. – 144 с.
10. Балашов Е. П. Эволюционный синтез систем / Е. П. Балашов. – М. : Радио и связь, 1985. – 328 с.
11. Баранников А. Ф. Теория организации : [учебник для вузов] / А. Ф. Баранников. – М. : ЮНИТИ-ДАНА, 2004. – 700 с.
12. Белов В. А. Ценностное измерение науки / В. А. Белов. – М. : Идея-Пресс, 2001. – 284 с.
13. Беляев М. Механизм управления факторами развития современных экономических систем / М. Беляев, О. Максимчук // Проблемы теории и практики управления. – 2006. – № 11. – С. 19–25.
14. Бережная Е. В. Математические методы и моделирование экономических систем: [учеб. пособие] / Е. В. Бережная, В. И. Бережной. – М. : Финансы и статистика, 2001. – 368 с.

15. Болч Б. У. Многомерные статистические методы для экономики / Б. У. Болч, К. Д. Хуань ; пер. с англ. А. Д. Плитмана. – М. : Статистика, 1979. – 316 с.

16. Бондаренко М. Ф. Моделирование и проектирование бизнес-систем: методы, стандарты, технологи : [учеб. пособие] / М. Ф. Бондаренко, С. И. Маторин, Е. А. Соловьева. – Харьков : Компания СМИТ, 2004. – 272 с.

17. Бородкин Ф. М. Некоторые проблемы систем измерения социологической информации / Ф. М. Бородкин, Б. Г. Миркин. – Новосибирск : Академия наук СССР, 1970. – 14 с.

18. Бородкин Ф. М. Социальные индикаторы : [учебник для студ. вузов] / Ф. М. Бородкин, С. А. Айвазян. – М. : ЮНИТИ-ДАНА, 2006. – 608 с.

19. Бочарников В. П. Fuzzy-технология: математические основы. Практика моделирования в экономике / В. П. Бочарников. – СПб. : Наука РАН, 2001. – 328 с.

20. Браун Марк Г. Сбалансированная система показателей: на маршруте внедрения / Марк Г. Браун ; пер. с англ. – М. : Альпина Бизнес Букс, 2005. – 228 с.

21. Бро Г. Г. Математические методы экономического анализа на предприятии / Г. Г. Бро, Л. М. Шнайдман. – М. : Экономика, 1976. – 184 с.

22. Бутник О. М. Економіко-математичне моделювання динамічних закономірностей розвитку економічних систем / О. М. Бутник. – Харків : Вид. Дім "ІНЖЕК", 2003. – 224 с.

23. Введение в математическое моделирование : [учеб. пособие] / под ред. П. В. Трусова. – М. : Логос, 2004. – 440 с.

24. Вітлінський В. В. Економічний ризик і методи його вимірювання : [підручник] / В. В. Вітлінський, С. І. Наконечний, О. Д. Шарапов. – К. : ІЗМН, 1996. – 400 с.

25. Винер Н. Кибернетика и общество / Н. Винер ; пер. Е. Г. Панфилова. – М. : Изд. иностр. лит., 1958. – 200 с.

26. Вольф Г.-П. Сравнительный анализ на комбинатах и предприятиях / Г. П. Вольф, У. Кюкк, Г. Форбриг ; пер. с нем. М. И. Пугачева. – М. : Финансы и статистика, 1987. – 160 с.

27. Геец В. М. Прогнозирование показателей развития экономики / В. М. Геец. – К. : Наукова думка, 1975. – 108 с.

28. Геєць В. М. Деякі порівняльні ознаки трансформаційних моделей економіки України та Росії / В. М. Геєць // Економіка України. − 2005. – № 5. – С. 4–17.

29. Генетические алгоритмы: искусственные нейронные сети и проблемы виртуальной реальности / [Г. К. Вороновский, К. В. Махотило, С. Н. Петрашев, С. А. Сергеев]. – Харків : Основа, 1997. – 112 с.

30. Гласс Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стэнли ; пер. с англ. Л. И. Хайдусовой. – М. : Прогресс, 1976. – 496 с.

31. Глинский В. В. Статистический анализ : [учеб. пособие] / В. В. Глинский, В. Г. Ионин. − [2-е изд., перераб. и доп.]. – М. : Информ.-издат. дом "Филинъ", 1998. – 264 с.

32. Глухов В. В. Математические методы и модели для менеджмента / В. В. Глухов, М. Д. Медников, С. Б. Коробко. – СПб. : Лань, 2000. – 480 с.

33. Глущенко В. В. Системы управления: интеллектуализация поддержки принятия решений / В. В. Глущенко. – СПб. : Судостроение, 2004. – 322 с.

34. Голов С. Ф. Управлінський облік : [підручник] / С. Ф. Голов. – К. : Лібра, 2003. – 704 с.

35. Голоденберг А. И. Многофакторная модель стимулирования хозяйственной деятельности / А. И. Голоденберг // Экономика и математические методы. – 2007. – № 1. – С. 97–112.

36. Гранберг А. Г. Моделирование социалистической экономики : [учебник для студ. экон. вузов] / А. Г. Гранберг. – М. : Экономика, 1988. – 488 с.

37. Грановский В. А. Содержание принципов теории измерений. Анализ и формализация измерительного эксперимента / В. А. Грановский, Л. М. Гунтер ; под ред. Ю. В. Тарбеева. – Л. : 1986. – 68 с.

38. Грановский В. А. Методы обработки экспериментальных данных при измерениях / В. А. Грановский, Т. Н. Сирая. – Л. : Энергоатомиздат. Ленингр. отд., 1990. – 288 с.

39. Гриньова В. М. Проблеми управління трудовими ресурсами підприємства / В. М. Гриньова, О. М. Ястремська. – Харків : ХНЕУ, 2006. – 191 с.

40. Гроп Д. Методы идентификации систем / Д. Гроп ; пер. с англ. В. А. Васильева, В. И. Лопатина. – М. : Мир, 1979. – 304 с.

41. Гутнер Л. М. Измерение в структуре теоретических отношений / Л. М. Гутнер. – Л. : Изд. Ленинг. ун-та, 1985. – 104 с.

42. Державний класифікатор України. Класифікатор системи позначень одиниць вимірювання та обліку. ДК 011-96 [електронний ресурс] – Режим доступу до докум. :http:/www.kmu.gov.ua.

43. Дик В. В. Методология формирования решений в экономических системах и инструментальные средства их поддержки / В. В. Дик. – М. : Финансы и статистика, 2001. – 300 с.

44. Дисперсионная идентификация / [под ред. Н. С. Райбмана]. – М. : Наука, 1981. – 336 с.

45. Доля В. Т. Экономический анализ: теория и практические методики : [учеб. пособие] / В. Т. Доля. – К. : Кондор, 2003. – 208 с.

46. Дороніна М. С. Управління економічними та соціальними процесами підприємства / М. С. Дороніна. – Харків : Вид. ХДЕУ, 2002. – 432 с.

47. Дорохина Е. Ф. Моделирование микроэкономики : [учеб. пособие для вузов] / Е. Ф. Дорохина, М. А. Халиков. – М. : Изд. "Экзамен", 2003. – 224 с.

48. Дубров А. М. Компонентный анализ и эффективность в экономике : [учеб. пособие] / А. М. Дубров. – М. : Финансы и статистика, 2002. – 352 с.

49. Дубров А. М. Многомерные статистические методы : [учебник] / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. – М. : Финансы и статистика, 2000. – 352 с.

50. Дюк В. Обработка данных на ПК в примерах / В. Дюк. – СПб. : Питер, 1997. –240 с.

51. Дюран Б. Кластерный анализ / Б. Дюран, П. Оделл ; пер. с англ. Е. З. Демиденко. – М. : Статистика, 1977. – 128 с.

52. Дэйвисон М. Многомерное шкалирование: Методы наглядного представления данных / М. Дэйвисон ; пер. с англ. В. С. Каменского. – М. : Финансы и статистика, 1988. – 254 с.

53. Евклид. Начала Евклида / Евклид; пер. с греч. и коммент. Д. Д. Мордухай-Болтовского. – М. : Гос. изд. техн.-теорет. лит., 1950. – 332 с.

54. Економічний аналіз : [навч. посібник] / за ред. М. Г. Чумаченка. – [2-ге вид., оновл.]. – К. : КНЕУ, 2003. – 556 с.

55. Економічний ризик: ігрові моделі: Навч. посібник / В. В. Вітлінський, П. І. Верченко, А. В. Сігал, Я. С. Наконечний ; [за ред. докт. екон. наук, професора В. В. Вітлінського]. – К. : КНЕУ, 2002. – 448 с.

56. Егоршин А. А. Корреляционно-регрессионный аналіз : [учеб. пособие для студ. экон. спец. вузов] / А. А. Егоршин, Л. М. Малярец. – Харьков : Основа, 1998. – 208 с.

57. Егоршин А. А. Моделирование интегрального показателя конкурентного статуса предприятия / А. А. Егоршин, Л. М. Малярец // Коммунальное хозяйство городов : науч.-техн. сб. – К. : "Техніка", 2003. – С. 54–65. – (Серия: Экономические науки; вып. 50).

58. Егоршин А. А. Проблемы эконометрического оценивания / А. А. Егоршин, Л. М. Малярец // Коммунальное хозяйство городов : науч.-техн. сб. – К. : "Техніка", 2005. – С. 267–273. – (Серия: Экономические науки; вып. 61).

59. Егоршин А. А. Практикум по эконометрии в Excel : [пособие для студ. высш. учебн. завед.] / А. А. Егоршин, Л. М. Малярец. – Харьков : ИД "ИНЖЭК", 2005. – 100 с.

60. Єгоршин О. О. Математичне програмування : [підручик для студ. вищ. навч. закл.] / О. О. Єгоршин, Л. М. Малярець. – Харків : ВД "ІНЖЕК", 2006. – 384 с.

61. Єгоршин О. О. Універсальний метод багатовимірного шкалювання / О. О. Єгоршин, Л. М. Малярець // Економіка розвитку. – 2007. – № 4 (44). – С. 43–47.

62. Єгоршин О. О. Методи багатовимірного статистичного аналізу : навч. посібник / О. О. Єгоршин, А. М. Зосімов, В. С. Пономаренко. – К. : ІЗМН, 1998. – 208 с.

63. Елисеева И. И. Общая теория статистики : [учебник] / И. И. Елисеева, М. М. Юзбашев. – [4-е изд., перераб. и доп.]. – М. : Финансы и статистика, 2003. – 480 с.

64. Елисеева И. М. Группировка, корреляция, распознавание образов: Статистические методы классификации и измерения связей / И. М. Елисеева, В. О. Рукавишников. – М. : Статистика, 1977. – 144 с.

65. Емельянов А. А. Имитационное моделирование экономических процессов : [учеб. пособие] / А. А. Емельянов, Е. А. Власова, Р. В. Дума. – М. : Финансы и статистика, 2004. – 368 с.

66. Енюков И. С. Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА / И. С. Енюков. – М. : Финансы и статистика, 1986. – 232 с.

67. Єріна А. М. Статистичне моделювання та прогнозування : [навч. посібник] / А. М. Єріна. – К. : КНЕУ, 2001. – 170 с.

68. Жамбю М. Иерархический кластер-анализ и соответствия / М. Жамбю ; пер. с фр. – М. : Финансы и статистика, 1988. – 342 с.

69. Жданов С. А. Экономические модели и методы в управлении / С. А. Жданов. – М. : Изд. "Дело и Сервис", 1998. – 176 с.

70. Жданов С. А. Эталоны нормального и кризисного функционирования предприятий / С. А. Жданов. – М. : Эдиториал УРСС, 2001. – 216 с.

71. Замков О. О. Математические методы в экономике : [учебник] / О. О. Замков, А. В. Толстопятенко, Ю. Н. Черемных. – М. : МГУ им. М. В. Ломоносова, Изд. "ДИС", 1997. – 368 с.

72. Зарубин В. С. Математическое моделирование в технике: [учебник для вузов] / В. С. Зарубин. – М. : Изд. МГТУ им. Н. Э. Баумана, 2001. – 496 с.

73. Захожай В. Б. Статистичне забезпечення управління якістю : [навч. посібик] / В. Б. Захожай, А. Ю. Чорний. – К. : Центр навчальної літератури, 2005. – 340 с.

74. Иберла К. Факторный анализ / К. Иберла ; пер. с нем. В. М. Ивановой. – М. : Статистика, 1980. – 398 с.

75. Измерения в промышленности : [справ. изд. : в 3-х кн.]. – М. : Металлургия, 1990– . –

Кн. 1. Теоретические основы ; пер. с нем. – [2-е изд., перераб. и доп.]. – 1990. – 492 с.

76. Ириков В. А. Распределенные системы принятия решений. Теория и приложения / В. А. Ириков, В. Н. Тренев. – М. : Наука. Физматлит., 1999. – 288 с.

77. Исследование систем управления : [учеб. пособие] / под ред. д. э. н., проф. Э. М. Короткова. – М. : ИНФРА-М, 2003. – 176 с.

78. История математики с древнейших времен до начала нового времени. – М. : Наука, 1970– . –

Т. 1. – 1970. – 352 с.

79. Кавалеров Г. И. Введение в информационную теорию измерений / Г. И. Кавалеров, С. М. Мандельштам. – М. : Энергия, 1974. – 376 с.

80. Каверина О. Д. Управленческий учет: системы, методы, процедуры / О. Д. Каверина. – М. : Финансы и статистика, 2003. – 352 с.

81. Камке Д. Физические основы единиц измерения / Д. Камке, К. Кремер ; пер. с нем. В. Е. Марковича и д. ф.-м. н. Н. В. Мицкевича. – М. : Изд. "Мир", 1980. – 208 с.

82. Каныгин Ю. М. Основы теоретической информатики / Ю. М. Каныгин, Г. И. Калитич. – К. : Наук. думка, 1990. – 232 с.

83. Каплан Роберт С. Сбалансированная система показателей. От стратегии к действию / Роберт С. Каплан, Дэйвид П. Нортон ; пер с англ. – М. : ЗАО "Олимп – Бизнес", 2004. – 320 с.

84. Карел Берка. Измерения. Понятия, теории, проблемы / Берка Карел ; пер. с чешск. К. Н. Иванова. – М. : Прогресс, 1987. – 320 с.

85. Карнап Р. Философские основания физики: Введение в философию науки / Р. Карнап ; пер. с англ., предисл. и коммент. Г. И. Рузавин. – М. : Эдиториал УРСС, 2003. – 388 с.

86. Карпова Т. П. Управленческий учет : [учебник для вузов] / Т. П. Карпова. – М. : ЮНИТИ, 2001. – 352 с.

87. Карташова Т. М. Применение совмещенных планов для исследования и оптимизации процессов переработки смесей полимеров / Т. М. Карташова, Б. П. Штаркман, А. М. Шаргородский // Пластические массы. – 1969. – № 9. – С. 16–22.

88. Кедровский О. И. Методологические проблемы развития математического познания / О. И. Кедровский. – К. : Вища школа, 1977. – 232 с.

89. Кенделл М. Ранговые корреляции / М. Кенделл ; пер. с англ. – М. : Статистика, 1975. – 216 с.

90. Кизим Н. А. Нейронные сети: теория и практика применения / Н. А. Кизим, Е. Н. Ястремская, В. Ф. Сенчуков. – Харьков : ИНЖЭК, 2006. – 234 с.

91. Кини Р. Принятие решений при многих критериях: предпочтения и замещения / Р. Кини, Х. Райфа ; пер. с англ. – М. : Радио и связь, 1981. – 560 с.

92. Кігель В. Р. Методи і моделі підтримки прийняття рішень у ринковій економіці / В. Р. Кігель. – К. : ЦУЛ, 2003. – 204 с.

93. Клайн М. Математика. Поиск истины / М. Клайн ; пер. с англ. – М. : Мир, 1988. – 296 с.

94. Клебанова Т. С. Механизмы и модели управления кризисными ситуациями / Т. С. Клебанова, Л. С. Гурьянова, А. Т. Рогович ; под ред. Т. С. Клебановой. – Харьков : ИД "ИНЖЭК", 2007. – 200 с.

95. Клебанова Т. С. Модели и методы координации в крупномасштабных экономических системах / Т. С. Клебанова, Е. В. Молдавская, Чанг Хонгвен. – Харьков : Бизнес Информ, 2002. – 148 с.

96. Клейнер Г. Б. Эконометрические зависимости: принципы и методы построения / Г. Б. Клейнер, С. А. Смоляк. – М. : Наука, 2000. – 104 с.

97. Клейнер Г. Б. Производственные функции: Теория, методы, применение / Г. Б. Клейнер. – М. : Финансы и статистика, 1986. – 240 с.

98. Клигер С. А. Шкалирование при сборе и анализе социологической информации / С. А. Клигер, М. С. Косолапов, Ю. Н. Толстова. – М. : Наука, 1978. – 112 с.

99. Кнорринг В. Г. Теоретические основы информационно-измерительной техники. Основные понятия теории шкал / В. Г. Кнорринг. – Л. : ЛПИ, 1983. – 44 с.

100. Ковалев А. И. Маркетинговый анализ / А. И. Ковалев, В. В. Войленко. – М. : Центр экономики и маркетинга, 2000. – 266 с.

101. Колмогоров А. Н. Математическая логика / А. Н. Колмогоров. – М. : Наука, 2004. – 368 с.

102. Колмогоров А. Н. Математика в ее историческом развитии / А. Н. Колмогоров ; под ред. В. А. Успенского. – М. : Наука, 1991. – 224 с.

103. Колпаков В. М. Теория и практика принятия управленческих решений : [учеб. пособие] / В. М. Колпаков. – К. : МАУП, 2000. – 256 с.

104. Компьютерное моделирование. Экономика / [под ред. д. т. н., проф. Г. А. Угольницкого]. – М. : Вузовская книга, 2002. – 100 с.

105. Кугаенко А. А. Основы теории и практики динамического моделирования социально-экономических объектов и прогнозирования их развития / А. А. Кугаенко. – М. : Вузовская книга, 1998. – 392 с.

106. Кузин Б. Методы и модели управления фирмой / Б. Кузин, В. Юрьев, Г. Шахдинаров. – СПб. : Питер, 2001. – 432 с.

107. Кундышева Е. С. Математическое моделирование в экономике : [учеб. пособие] / Е. С. Кундышева ; под науч. ред. проф. Б. А. Суслакова. – М. : Издательско-торговая корпорация "Дашков и К°", 2004. – 352 с.

108. Курбатов В. И. Математические методы социальных технологий : [учеб. пособие] / В. И. Курбатов, Г. А. Угольницкий. – М. : Вузовская книга, 1998. – 256 с.

109. Кучин Б. Л. Управление развитием экономических систем: технический прогресс, устойчивость / Б. Л. Кучин, Е. В. Якушева. – М. : Экономика, 1990. – 160 с.

110. Кэрролл Л. История с узелками / Л. Кэрролл ; пер. с англ. Ю. А. Данилова. – М. : Мир, 2000. – 400 с.

111. Лебег А. Об измерении величин / А. Лебег ; пер. с фр. О. И. Кисловской-Карской ; [под ред. И. М. Яглома]. – М. : Гос. учебно-педагогическое издательство Министерства просвещения РСФСР, 1960. – 204 с.

112. Леоненков А. В. Нечеткое моделирование в среде MATLAB и fuzzyTECH / А. В. Леоненков. – СПб. : БХВ-Петербург, 2003. – 726 с.

113. Лимер Э. Статистический анализ неэкспериментальных данных: Выбор формы связи / Э. Лимер ; пер. с англ. О. В. Ивановой, Ю. П. Федоровского ; [под ред. и с предисл. А. А. Рывкина]. – М. : Финансы и статистика, 1983. – 384 с.

114. Литвак Б. Г. Разработка управленческого решения : [учебник] / Б. Г. Литвак. – М. : Дело, 2003. – 392 с.

115. Лихтенштейн В. Е. Экономико-математическое моделирование : [учеб. пособие] / В. Е. Лихтенштейн, В. И. Павлов. – М. : Изд. "ПРИОР", 2001. – 448 с.

116. Лысенко Ю. Г. Экономическая динамика : [учеб. пособие] / Ю. Г. Лысенко, В. Л. Петренко, В. Н. Тимохин. – Донецк : ДонГУ, 2000. – 176 с.

117. Любушин Н. П. Анализ финансово-экономической деятельности предприятия : [учеб. пособие для вузов] / Н. П. Любушин, В. Б. Лещева, В. Г. Дьякова ; под ред. проф. Н. П. Любушина. – М. : ЮНИТИ-ДАНА, 1999. – 472 с.

118. Лямець В. І. Системний аналіз. Вступний курс / В. І. Лямець, А. Д. Тевяшев. – Харків : ХНУРЕ, 2004. – 448 с.

119. Лячнев В. В. Основы теории измерений физических величин : [учеб. пособие] / В. В. Лячнев, Т. Н. Сирая, Л. И. Довбеха ; под ред. В. В. Лячнева. – СПб. : Изд. СПбГЭТУ "ЛЭТИ", 2004. – 310 с.

120. Майминас Е. З. Процессы планирования в экономике: информационный аспект / Е. З. Майминас. – М. : Экономика, 1971. – 390 с.

121. Макроекономічна політика в Україні: проблеми науки та практики / [В. С. Пономаренко, В. А. Зінченко, М. О. Кизим та ін.] . – Харків : ВД "ІНЖЕК", 2007. – 352 с.

122. Малинецкий А. А. Тектология. Теория систем. Теоретическая биология / А. А. Малинецкий. – М. : Эдиториал УРСС, 2000. – 448 с.

123. Малыхин В. И. Математическое моделирование экономики : [учеб.-практ. пособие] / В. И. Малыхин. – М. : Изд. УРАО, 1998. – 160 с.

124. Малярец Л. М. Анализ теоретических проблем измерения экономических объектов / Л. М. Малярец // Економіка: проблеми теорії та практики : зб. наук. праць. – Вип. 190; в 4 т. – Дніпропетровськ : ДНУ, 2004 – . –

Т. 1. – 2004. – С. 281–288.

125. Малярець Л. М. Аналітичне визначення латентних факторів організаційно-економічної діяльності підприємства / Л. М. Малярець, Г. А. Іващенко // Вісник Національного технічного університету "ХПІ" : зб. наук. праць. Тематичний випуск: Технічний прогрес і ефективність виробництва. – Вип. 4. – Харків : НТУ "ХПІ", 2005. – С. 19–28.

126. Малярец Л. М. Аналитическая оценка оптимального соотношения в структуре источников формирования оборотного капитала / Л. М. Малярец, Е. В. Авраменко // Вісник ХДЕУ. – 1999. – № 4(12). – С. 104–108.

127. Малярец Л. М. Аналитические методы решения задач управления конкурентоспособностью предприятия / Л. М. Малярец // Спецвипуск. Вісник ХДЕУ. – 2001. – № 2(18). – С. 86–89.

128. Малярец Л. М. Аналитическое обоснование оценки процессов социально-экономического развития / Л. М. Малярец // Социальная экономика. – 2003. – № 1. – С. 239–243.

129. Малярець Л. М. Базис описових моделей складних ознак у сучасному аналізі соціально-економічних систем макрорівня / Л. М. Малярець // Економіка розвитку. – 2008. – № 1 (45). – С. 57–63.

130. Малярець Л. М. Визначення оптимальної структури капіталу банку при розрахунку узагальнюючого показника його діяльності / Л. М. Малярець, І. П. Отенко // Банківська справа. – 2000. – № 2(32). – С. 17–19.

131. Малярець Л. М. Визначення рівнів інвестиційної привабливості суб'єктів господарювання та напрямків їх зміни / Л. М. Малярець, В. Ф. Колесніченко // Економіка розвитку. – 2005. – № 2(34). – С. 8–12.

132. Малярець Л. М. Визначення системи складних ознак виробничо-господарської діяльності підприємства для її оцінки / Л. М. Малярець // Економіка: проблеми теорії та практики : зб. наук. праць. – Вип. 210; в 4 т. Дніпропетровськ : ДНУ, 2005. – . –

Том IV. – 2005. – С . 878–886.

133. Малярець Л. М. Визначення рівнів узагальнюючого показника ефективності управління капіталом підприємства / Л. М. Малярець, І. М. Чмутова, Р. А. Косінський // Вісник Хмельницького національного університету. – 2005. – № 3. – С. 35–39.

134. Малярець Л. М. Вимірювання ознак об'єктів в економіці: методологія та практика / Л. М. Малярець. – Харків : Вид. ХНЕУ, 2006. – 384 с.

135. Малярець Л. М. Діагностика етапів розвитку кризи в діяльності комерційного банку / Л. М. Малярець, В. Я. Вовк // Розвиток підприємницької діяльності в Україні: історія та сьогодення : Міжнарод. наук. конф., 10–11 червня 2004 р. : тези допов. ─ Тернопіль : Вид. Стародубець В. О., 2004. ─ С. 171–172.

136. Малярец Л. М. Экономико-математическая модель оценки премии за риск при обосновании инновационных проектов / Л. М. Малярец, Е. Б. Жихор // Вісник ХДУ ім. В. Н. Каразіна. – 2001. – № 535. – С. 358–365.

137. Малярець Л. М. Застосування методів статистичної обробки якісних номінальних ознак об'єкта в економіці / Л. М. Малярець // Економіка, менеджмент, підприємництво : зб. наук. праць. – Вип. 16. – Луганськ : Вид. СНУ ім. В. Даля, 2006. – С. 17–30.

138. Малярец Л. М. Классификация погрешностей в измерении признаков объекта в экономике / Л. М. Малярец // Бизнес Информ. – 2006. – № 10. – С. 56–62.

139. Малярець Л. М. Методичний підхід до комплексної оцінки організаційно-технічного рівня виробництва / Л. М. Малярець, К. В. Ларіна // Вісник Харківського національного університету ім. В. Н. Каразіна. – 2006. – № 719. – С. 82–90.

140. Малярец Л. М. Модификация факторного анализа / Л. М. Малярец // Бизнес Информ. – 2007. – № 8. – С. 130–135.

141. Малярец Л. М. Социально-экономическая оценка выручки рекламной кампании / Л. М. Малярец, Я. А. Полякова // Вестник Донецкого университета. – 2003. – № 1. – С. 81–91. – (Серия В. Экономика и право).

142. Малярець Л. М. Обґрунтування усталеності кластерів економічного розвитку підприємств в регіоні / Л. М. Малярець // Сучасні проблеми фінансово-господарського контролю : Всеукр. наук.-практ. конф., 30 березня 2005 р. : тези допов. – Кривий Ріг : Криворізький економічний інститут КНЕУ, 2005. – С. 188–190.

143. Малярец Л. М. Обоснование системы измерителей для комплексной оценки деятельности предприятия / Л. М. Малярец // Труды Одесского политехнического университета. – 2004. – Т. 3, № 1. – С. 73–76.

144. Малярец Л. М. О математических моделях в структурном анализе капитала предприятия / Л. М. Малярец // Технічний прогрес та ефективність виробництва : Вісник Харківського державного політехнічного університету : зб. наук. праць. – Вип. 122. В 4 ч. – Харків : ХДПУ, 2000. – . –
Ч. 2. – 2000. – С. 244–248.

145. Малярец Л. М. О постановке управленческих задач формирования конкурентоспособности предприятия / Л. М. Малярец // Економіка: проблеми теорії та практики : зб. наук. праць. – Вип. 85. – Дніпропетровськ : ДНУ, 2001. – С. 100–106.

146. Малярец Л. М. Определение механизма взаимосвязей в сбалансированной системе показателей / Л. М. Малярец, А. В. Штереверя // Бизнес Информ. – 2007. – № 6. – С. 71–89.

147. Малярец Л. М. Определение типов развития промышленных предприятий в регионе: аналитический аспект / Л. М. Малярец // Фінансово-економічні проблеми розвитку регіонів України : Всеукр. наук.-практ. конф., 26 жовтня 2004 р. : тези допов. – Дніпропетровськ : Наука і освіта, 2004. – С. 97–101.

148. Малярець Л. М. Основи технології аналізу інформаційності показників та внутрішніх факторів в управлінні фінансовим станом комерційного банку / Л. М. Малярець, В. Я. Вовк // Облік, контроль і аналіз в

управлінні підприємницькою діяльністю : Міжнар. наук.-практ. конф., 24–26 березня 2004 р. : тези допов. ─ Черкаси : ЧДТУ, 2004. – С. 99–101.

149. Малярец Л. М. Особенности теоретических основ измерения экономических объектов / Л. М. Малярец // Інвестиційні стратегії сталого розвитку : Всеукр. наук.-практ. конф., 27–28 лютого 2004 р. : тези допов. – Дніпропетровськ : Наука і освіта, 2004. – С. 78–80.

150. Малярец Л. М. Оценка факторов формирования конкурентоспособности предприятия: основы построения и исследование ее уровня / Л. М. Малярец, А. А. Иващенко // Коммунальное хозяйство городов : науч.-техн. сб.– К. : "Техніка", 2004. – С. 200–209. – (Серия: Экономические науки; вып. 59).

151. Малярець Л. М. Оцінка виробничо-господарської діяльності підприємства за системою складних ознак / Л. М. Малярець // Інвестиційні стратегії підприємств України на міжнародних товарних та фінансових ринках : Всеукр. наук.-практ. конф., 3–4 лютого 2006 р. : тези допов. – Дніпропетровськ : Наука і освіта, 2006. – С. 83–85.

152. Малярец Л. М. Параметрический анализ экономических показателей рекламной деятельности предприятия / Л. М. Малярец, Я. А. Полякова // Економіка: проблеми теорії та практики : зб. наук. праць. – Вип. 182; в 3 т. – Дніпропетровськ : ДНУ, 2003. – . –

Т. III. – 2003. – С. 671–679.

153. Малярець Л. М. Побудова окремих функцій бажаності для вимірювання ознак об’єктів в економіці / Л. М. Малярець // Вісник Харківського національного університету ім. В. Н. Каразіна. – 2005. – № 650. – С. 209–216.

154. Малярец Л. М. Построение обобщающих показателей в оценке конкурентных преимуществ предприятия / Л. М. Малярец // Економічний вісник ‘ 2004(1) Національного технічного університету України "КПІ" : зб. наук. праць. – К. : Вид. ПП "Екмо", 2004. – С. 432–438.

155. Малярец Л. М. Построение табло величин признаков деятельности предприятия для ее оценки / Л. М. Малярец, А. В. Штереверя // Бизнес Информ. – 2007. – № 11. – С. 147–155.

156. Малярец Л. М. Применение кластерного анализа к проблеме разделения смесей для решения задач развития экономических процессов / Л. М. Малярец, И. А. Никольский // Вісник Національного технічного університету "ХПІ" : зб. наук. праць. – Тематичний випуск : Технічний прогрес і ефективність виробництва. – Харків : НТУ "ХПІ", 2003. – № 22. – С. 122–124.

157. Малярець Л. М. Проблеми концептуального аналізу та економіко-математичного моделювання підприємства / Л. М. Малярець // Українська наука: минуле, сучасне, майбутнє. – 2003. – № 6. – С. 205–209.

158. Малярец Л. М. Проблемы методологии анализа стратегических возможностей предприятия / Л. М. Малярець, И. П. Отенко // Наука і освіта : Міжнар. конф., 1–15 лютого 2001 р. : тези допов. – Дніпропетровськ : Наука і освіта, 2001. – С. 10–11.

159. Малярец Л. М. Проведение комплексной оценки процессов формирования и развития конкурентных преимуществ предприятия / Л. М. Малярец, Е. А. Полтавская // Економіка: проблеми теорії та практики : зб. наук. праць. Вип. 185; в 4 т. – Дніпропетровськ : ДНУ, 2003. –.–

Т. ІІІ. – 2003. – С. 689–696.

160. Малярец Л. М. Разработка показателя качества неметрических признаков объекта в экономике / Л. М. Малярец // Бизнес Информ. – 2006. – № 8. – С. 77–85.

161. Малярец Л. М. Разработка универсальной, объективной меры тесноты связи признаков объектов в экономике / Л. М. Малярец, И. А. Никольский

// Бизнес Информ. – 2007. – № 9 (2). – С. 114–118.

162. Малярець Л. М. Розвиток дисперсійного аналізу ознак (виміряних на порядкових шкалах) об'єкта в економіці / Л. М. Малярець // Економіка розвитку. – 2006. – № 3(39). – С. 35–39.

163. Малярец Л. М. Система измерителей в построении контроллинга в коммерческом банке / Л. М. Малярец // Економіка розвитку. – 2004. – № 3(31). – С. 10–15.

164. Малярец Л. М. Статистические методы описания взаимосвязи качественных (порядковых) признаков объекта в экономике / Л. М. Малярец // Бизнес Информ. – 2006. – № 5. – С. 72–82.

165. Малярець Л. М. Статистичний аналіз елементарних ознак розвитку підприємства / Л. М. Малярець // Економіка розвитку. – 2004. – № 4(32). – С. 71–79.

166. Малярець Л. М. Стійкість економіко-математичного моделювання у вимірюванні ознак об'єктів в економіці / Л. М. Малярець // Економіка розвитку. – 2007. – № 1(41). – С. 33–38.

167. Малярець Л. М. Структурний аспект в аналізі діяльності промислових підприємств в регіоні / Л. М. Малярець, І. П. Отенко, Ю. Ф. Ярошенко // Регіональні перспективи. – 2000. – № 2–3 (9–10). – С. 288–291.

168. Малярець Л. М. Сучасні особливості концептуального аналізу та економіко-математичного моделювання підприємства / Л. М. Малярець // Розвиток підприємницької діяльності в Україні: історія та сьогодення : Міжнарод. наук. конф., 24 березня 2003 р. : тези допов. – Тернопіль : Вид. СТАРОДУБЕЦЬ, 2003. – С. 56.

169. Малярець Л. М. Факторний аналіз якісних ознак у діагностиці конкурентного статусу підприємства / Л. М. Малярець, Л. О. Норік // Коммунальное хозяйство городов : науч.-техн. сб. Вып. 75. – К. : "Техніка", 2007. – С. 307–315. – (Серия : Экономические науки: вып. 75).

170. Малярець Л. М. Шкали у вимірюванні величин ознак об'єктів в економіці / Л. М. Малярець // Розвиток економіки в трансформаційний період: глобальний та національний аспекти : Міжнар. наук.-практ. конф. 20 квітня 2005 р. : тези допов. – Дніпропетровськ : Наука і освіта, 2005. – С. 35–35.

171. Мандель И. Д. Кластерный анализ / И. Д. Мандель. – М. : Финансы и статистика, 1988. – 176 с.

172. Маркова Е. В. Применение латинского куба второго порядка при разработке рецептуры нового полимерного материала / Е. В. Маркова, Т. М. Карташова, Ю. М. Бусыгина // Заводская лаборатория. – 1969. – № 7. – С. 35.

173. Маркс Карл. Сочинения: в 30 т. / Карл Маркс, Фридрих Энгельс. – М. : Госполитиздат, 1960– . –

Т. 23. – 1960. – 907 с.

174. Матвійчук А. В. Аналіз та прогнозування розвитку фінансово-економічних систем із використанням теорії нечіткої логіки / А. В. Матвійчук. – К. : Центр навчальної літератури, 2005. – 206 с.

175. Математичні моделі та інформаційні технології в сучасній економіці / [Т. С. Клебанова, Л. С. Гур'янова, Н. Богоніколос. та ін.] ; за ред. А. О. Єпіфанова. – Суми : УАБС НБУ, 2007. – 246 с.

176. Математическая энциклопедия : [в 5 т.] / глав. ред. И. М. Виноградов. – М. : Советская Энциклопедия, 1977– . –

Т. 1. – 1977. – 1152 с.

177. Математическая энциклопедия : [в 5 т.] / глав. ред. И. М. Виноградов. – М. : Советская Энциклопедия, 1977– . –

Т. 4. – 1984. – 1 216 с.

178. Математические модели трансформационной экономики : [учеб. пособие] / [Т. С. Клебанова, Е. В. Раевнева, К. А. Стрижиченко и др.]. – Харьков : ИД "ИНЖЭК", 2006. – 280 с.

179. Мейер Маршал В. Оценка эффективности бизнеса / Маршал В. Мейер ; пер. с англ. А. О. Корсунский. – М. : ООО "Вершина", 2004. – 272 с.

180. Мельник Л. Г. Фундаментальные основы развития / Л. Г. Мельник. – Сумы : ИТД "Университетская книга", 2003. – 288 с.

181. Мельников Г. П. Системология и языковые аспекты кибернетики / Г. П. Мельников. – М. : Сов. радио, 1978. – 368 с.

182. Методы и модели анализа данных: OLAP и Data Mining / [А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод]. − СПб. : БХВ-Петербург, 2004. − 336 с.

183. Методы исследований и организация экспериментов / [под ред. проф. К. П. Власова]. – Харьков : Издательство "Гуманитарный Центр", 2002. – 256 с.

184. Миленький А. В. Классификация сигналов в условиях неопределенности / А. В. Миленький. – М. : Советское радио, 1975. – 328 с.

185. Минцберг Г. Структура в кулаке: создание эффективной организации / Г. Минцберг ; пер. с анг. под ред. Ю. Н. Каптуревского. – СПб. : Питер, 2001. – 512 с.

186. Минюк С. А. Математические методы и модели в экономике : [учеб. пособие] / С. А. Минюк, Е. А. Ровба, К. К. Кузьмич. – Мн. : ТетраСистемс, 2002. – 432 с.

187. Миркин Б. Г. Группировки в социально-экономических исследованиях: Методы построения и анализа / Б. Г. Миркин. – М. : Финансы и статистика, 1985. – 223 с.

188. Миркин Б. Г. Анализ качественных признаков и структур / Б. Г. Миркин. – М. : Статистика, 1980. – 320 с.

189. Михалевич В. С. Вычислительные методы исследования и проектирования сложных систем / В. С. Михалевич, В. Л. Волкович. – М. : Наука, 1982. – 288 с.

190. Михеев В. И. Моделирование и методы теории измерений в педагогике : [науч.-метод. пособие] / В. И. Михеев. – М. : Высшая школа, 1987. – 200 с.

191. Многомерный статистический анализ в экономике : [учеб. пособие для вузов] / Л. А. Сошникова, В. Н. Тамашевич, Г. Уебе, М. Шефер ; под ред. проф. В. Н. Тамашевича. – М. : ЮНИТИ-ДАНА, 1999. – 600 с.

192. Моделирование сложных систем / под ред. Н. П. Бусленко. – М. : Наука, 1978. – 396 с.

193. Моделирование народнохозяйственных процессов: [учеб. пособие для экон. вузов и фак.] / под ред. В. С. Дадаяна. – М. : Экономика, 1973. – 480 с.

194. Моделирование финансовых потоков предприятия в условиях неопределенности / [Т. С. Клебанова, Л. С. Гурьянова, Н. Богониколос и др.]. – Харьков : ИД "ИНЖЭК", 2006. – 312 с.

195. Молчанов А. А. Моделирование и проектирование сложных систем / А. А. Молчанов. – К. : Выща школа, 1988. – 360 с.

196. Монахов А. В. Математические методы анализа экономики / А. В. Монахов. – СПб. : Питер, 2002. – 176 с.

197. Морено Я. Л. Социометрия: Экспериментальный метод и наука об обществе / Я. Л. Морено ; пер. с англ. – М. : Академический Проект, 2001. – 384 с.

198. Національний класифікатор України. Класифікація видів економічної діяльності. ДК 009:2005 [Електронний ресурс]. – Режим доступу до докум. : http:/www.kmu.gov.ua.

199. Нейман Дж. Теория игр и экономическое поведение / Дж. Нейман, О. Моргенштерн ; пер. с англ. ; [под ред. Н. Н. Воробьева]. – М. : Наука, 1970. – 708 с.

200. Нивен Пол Р. Сбалансированная система показателей – шаг за шагом: Максимальное повышение эффективности и закрепление полученных результатов / Пол Р. Нивен ; пер. с англ. – Днепропетровск : Баланс-Клуб, 2003. – 328 с.

201. Ниворожкина Л. И. Многомерные статистические методы в экономике: учебник / Л. И. Ниворожкина, С. В. Арженовский. – М. : Изд.-торг. корпор. "Дашков и К°"; Ростов н/Д. : Наука-Спектр, 2008. – 224 с.

202. Нили Энди. Призма эффективности: Карта сбалансированных показателей для измерения успеха в бизнесе и управления им / Энди Нили, Крис Адамс, Майк Кеннерли ; пер. с англ. – Днепропетровск : Баланс-Клуб, 2003. – 400 с.

203. Новик Ф. С. Математические методы планирования экпериментов в металловедении : [учебн. пособие] / Ф. С. Новик. – М. : Наука, 1979– . –

Ч. 1. – 1979. – 96 с.

204. Новик Ф. С. Оптимизация процессов технологии металлов методами планирования экспериментов / Ф. С. Новик, Я. Б. Арсов. – М. : Наука, 1980. – 440 с.

205. Нуреев Р. М. Экономика развития: модели становления рыночной экономики : [учеб. пособие] / Р. М. Нуреев. – М. : ИНФРА-М, 2001. – 240 с.

206. Ньютон И. Математические начала натуральной философии / И. Ньютон ; пер. с латин. и коммент. А. Н. Крылова. – М. : Наука, 1989. – 688 с.

207. Ольве Нильс-Горан. Оценка эффективности деятельности компании. Практическое руководство по использованию сбалансированной системы показателей / Нильс-Горан Ольве, Жан Рой, Магнус Ветер ; пер. с англ. – М. : Изд. дом "Вильямс", 2003. – 304 с.

208. Омельяновский М. Э. Диалектика в современной физике / М. Э. Омельяновский. – М. : Наука, 1973. – 228 с.

209. Орлов А. И. Устойчивость в социально-экономических моделях / А. И. Орлов. – М. : Наука, 1979. – 296 с.

210. Осипов Г. В. Методы измерения в социологии / Г. В. Осипов, Э. П. Андреев. – М. : Наука, 1977. – 184 с.

211. Оскольский В. В. Во всемирной системе стандартов префикс "482" – визитная Украины / В. В. Оскольский. – К. : Изд. Дом "Світ знань", 2006. – 172 с.

212. Основы менеджмента : [учеб. пособие для вузов] / науч. ред. А. А. Радугин. – М. : Центр, 1997. – 432 с.

213. Оценка эволюции плодородия на основе сводного показателя качества почв / [Т. А. Гринченко, Е. И. Григорьев, А. А. Егоршин и др.] // Агрохимия. – 1991. – № 5. – С. 52–60.

214. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка / Э. Петерс ; пер. с англ. – М. : Мир, 2000. – 336 с.

215. Петти Вильям. Классика экономической мысли : сочинения / [Вильям Пети, Адам Смит, Давид Риккардо и др.]. – М. : ЭКСМО-Пресс, 2000. – 896 с. (Антология мысли).

216. Плюта В. Сравнительный многомерный анализ в эконометрическом моделировании / В. Плюта ; пер. с польск. В. В. Иванова. – М. : Финансы и статистика, 1989. – 176 с.

217. Пономаренко В. С. Стратегічне управління підприємством / В. С. Пономаренко. – Харків : Основа, 1999. – 620 с.

218. Пономаренко В. С. Стратегічне управління розвитком підприємства. Навч. посібник / В. С. Пономаренко, О. І. Пушкар, О. М. Тридід. – Харків : Вид. ХДЕУ, 2002. – 640 с.

219. Пономаренко В. С. Стратегія розвитку підприємства в умовах кризи: Монографія / В. С. Пономаренко, О. М. Тридід, М. О. Кизим. – Харків : ВД "ІНЖЕК", 2003. – 328 с.

220. Пономаренко В. С. Моделювання поведінки інвестора на фондовому ринку : Монографія / В. С. Пономаренко, О. В. Раєвнєва, К. А. Стрижиченко. – Харків : ВД "ІНЖЕК", 2004. – 260 с.

221. Пономаренко В. С. Парадигма вимірювання в економіці / В. С. Пономаренко, Л. М. Малярець // Економіка розвитку. – 2006. – № 1(37). – С. 81–87.

222. Пономаренко В. С. Розробка класифікатора вимірювань в економіці / В. С. Пономаренко, Л. М. Малярець // Економіка розвитку. – 2006. – № 4(40). – С. 37–42.

223. Пономаренко В. С. Основні положення методології математичного моделювання ідентифікації соціально-економічних систем / В. С. Пономаренко, Л. М. Малярець // Економіка розвитку. – 2007. – № 3(43). – С. 43–47.

224. Пономаренко О. І. Системні методи в економіці, менеджменті та бізнесі : [навч. посібник] / О. І. Пономаренко, В. О. Пономаренко. – К. : Либідь, 1995. – 240 с.

225. Прангишвили И. Об эффективности управления сложными социально-экономическими системами / И. Прангишвили // Проблемы теории и практики управления. – 2006. – № 2. – С. 24–31.

226. Прикладная статистика: классификация и снижение размерности / [С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин]. – М. : Финансы и статистика, 1989. – 606 с.

227. Психологические измерения : [сборник] / пер. с англ. Е. Ю. Артемьевой ; под ред. Л. Д. Мешалкина. – М. : Мир, 1967. – 196 с.

228. Пушкарь А. И. Модели управления развитием производственно-экономических систем / А. И. Пушкарь. – Харьков : ХГЭУ, 1997. – 268 с.

229. Пушкарь А. И. Антикризисное управление: модели, стратегии, механизмы / А. И. Пушкарь, А. Н. Тридед, А. Л. Колос. – Харьков : ООО "Модель Вселенной", 2001. – 452 с.

230. Пфанцагль И. Теория измерений / И. Пфанцагль ; пер. с англ. В. Б. Кузьмина. – М. : Мир, 1976. – 248 с.

231. Пытьев Ю. П. Методы математического моделирования измерительно-вычислительных систем / Ю. П. Пытьев. – М. : ФИЗМАТЛИТ, 2004. – 400 с.

232. Радченко С. Г. Анализ экспериментальных данных на основе использования многофакторных статистических математических моделей / С. Г. Радченко // Математичні машини і системи. – 2005. – № 3. – С. 102–115.

233. Райфа Г. Прикладная теория статистических решений / Г. Райфа, Р. Шлейфе ; пер. с англ. А. К. Звонкина, З. Г. Маймина и Б. Л. Розовского. – М. : Статистика, 1977. – 360 с.

234. Рамперсад К. Хьюберт. Универсальная система показателей деятельности: Как достигать результатов, сохраняя целостность / К. Хьюберт Рамперсад ; пер. с англ. – М. : Альпина Бизнес Букс, 2004. – 352 с.

235. Рао С. Р. Линейные статистические методы и их применение / С. Р. Рао ; пер. с англ. А. М. Кагана ; [под ред. акад. Ю. Ф. Линника]. – М. : Наука, 1968. – 548 с.

236. Рассел Бертран. Философия логического атомизма / Бертран Рассел ; пер. с англ. – Томск : Водолей, 1999. – 192 с. (Библиотека аналитической философии).

237. Растригин Л. А. Введение в идентификацию объектов управления / Л. А. Растригин, Н. Е. Маджаров. – М. : Энергия, 1977. – 216 с.

238. Ременников В. Б. Разработка управленческого решения : [учеб. пособие для вузов] / В. Б. Ременников. – М. : ЮНИТИ-ДАНА, 2000. – 140 с.

239. Репин В. В. Процессный подход к управлению. Моделирование бизнес-процессов / В. В. Репин, В. Г. Елиферов. – М. : РИА "Стандарты и качество", 2005. – 408 с.

240. Рогальский Ф. Б. Математические методы анализа экономических систем : в 2 кн. / Ф. Б. Рогальский, А. А. Цокаренко. – К. : Наукова думка. – 2001– . –

Кн. 2 : Методы и алгоритмы решения трудноформализируемых задач. – 2001. – 424 с.

241. Розенталь О. М. Модель управления производственным процессом с учетом его перенастройки / О. М. Розенталь, Е. Д. Копионова // Экономика и математические методы. – 2007. – Т. 43. – № 1. – С. 129–132.

242. Розин Б. Б. Конструирование экономико-статистических моделей с заданными свойствами / Б. Б. Розин, М. А. Ягольницер. – Новосибирск : Наука, 1981. – 176 с.

243. Рыбников К. А. История математики : [учебник] / К. А. Рыбников. – М. : Изд. МГУ, 1994. – 496 с.

244. Рыбников К. А. Введение в методологию математики / К. А. Рыбников. – М. : Изд. МГУ, 1979. – 128 с.

245. Саати Т. Принятие решений. Метод анализа иерархий / Т. Саати ; пер. с англ. – М. : Радио и связь, 1993. – 320 с.

246. Савицкая Г. В. Анализ хозяйственной деятельности предприятия : [учеб. пособие] / Г. В. Савицкая. – [6-е изд.]. – Мн. : Новое знание, 2001. – 704 с.

247. Селиванов М. Н. Качество измерений: Метрологическая справочная книга / М. Н. Селиванов, А. Э. Фридман, Ж. Ф. Кудряшова. – Л. : Лениздат, 1987. – 296 с.

248. Симчера В. М. Методы многомерного анализа статистических данных : учеб. пособие / В. М. Симчера. – М. : Финансы и статистика, 2008. – 400 с.

249. Системные исследования в метрологии : сб. науч. трудов ; [под ред. Ю. В. Тарбеева]. – Л. : Энергоатомиздат, 1985. – 80 с.

250. Скоун Т. Управленческий учет / Т. Скоун ; пер. с англ. – М. : Аудит, ЮНИТИ, 1997. – 180 с.

251. Смирнов А. Д. Лекции по макроэкономическому моделированию : [учеб. пособие для вузов] / А. Д. Смирнов. – М. : ГУ ВШЭ, 2000. – 352 с.

252. Смирнов Н. Н. Стратегический менеджмент / Н. Н. Смирнов. – СПб. : Питер, 2002. – 128 с.

253. Смирнов Э. А. Управленческие решения / Э. А. Смирнов. – М. : ИНФРА-М, 2001. – 264 с.

254. Смоляк С. А. Интерполяция функций нескольких нечисловых переменных / С. А. Смоляк // Экономика и математические методы. – 2006. – Т. 42, – № 3. – С. 105–121.

255. Современные методы идентификации систем : пер. с англ. / под ред. П. Эйкхоффа. – М. : Мир, 1983. – 400 с.

256. Статистический анализ многомерных объектов произвольной природы / [Васильев В. И., Красильщиков В. В., Плаксий С. И., Тягунова Т. Н.]. – М. : Издательство "ИКАР", 2004. – 382 с.

257. Статистичний збірник "Регіони України" : [в 2 ч.] – К. : Деркомстат України, 2007– . – (статистичний щорічник)

Ч. 1. – 2007. – 348 с.

258. Стахов А. П. Алгоритмическая теория измерения / А. П. Стахов. – М. : Знание, 1979. – 64 с.

259. Стивенс С. С. Экспериментальная психология / С. С. Стивенс ; пер. с англ. под ред. П. К. Анохина, В. А. Артемова. – М. : Изд. иностр. лит., 1960. – 90 с.

260. Страны мира. Социально-экономические показатели. Информационно-аналитическая система социально-экономических показателей (ИАССЭП ЦЭМИ РАН) [Электронный ресурс]. – Режим доступа докум. : http://data.cemi.rssi.ru/isepweb/coun.htm.

261. Стратегічне управління організаційними перетвореннями на промислових підприємствах / [В. С. Пономаренко, А. М. Золотарьов, О. М. Ястремська та ін.]. – Харків : ХНЕУ, 2005. – 452 с.

262. Стратегічні виклики XXI століття суспільству та економіці України : [у 3 т.] ; за ред. акад. НАН України В. М. Геєця, акад. НАН України В. П. Семиноженка, чл.-кор. НАН України Б. Є. Кваснюка. – К. : Фенікс, 2007 −.−

Т. 1 : Економіка знань – модернізаційний проект України. – 2007. – 544 с.

263. Суслов И. П. Теория статистических показателей / И. П. Суслов. – М. : Статистика, 1975. – 264 с.

264. Системи підтримки прийняття рішень : [навч. посібник / за ред. О. І. Пушкаря]. – Харків : ВД "ІНЖЕК", 2005. – 302 с.

265. Тарский А. Введение в логику и методологию дедуктивных наук / А. Тарский ; пер. с англ. О. Н. Дынник ; ред. проф. С. А. Яновской. – М. : Гос. изд. иностранной литературы, 1948. – 328 с.

266. Теория статистики : [учебник ; под ред. Р. А. Шмойловой]. − [4-е изд.]. – М. : Финансы и статистика, 2005. – 656 с.

267. Теорія бухгалтерського обліку : [навч. посібник] / Ю. Д. Маляревський, В. І. Отенко, В. Д. Понікаров, Т. М. Сєрікова. – Харків : Вид. ХДЕУ, 2001. – 372 с.

268. Теорія статистики : [навч. посібник / під ред. П. Г. Вашків]. – К. : Либідь, 2001. – 320 с.

269. Тихомиров Н. П. Эконометрика : учебник [для студ. высш. учебн. зав.] / Н. П. Тихомиров, Е. Ф. Дорохина. – М. : Экзамен, 2003. – 512 с.

270. Томас Ричард. Количественные методы анализа хозяйственной деятельности / Ричард Томас; пер. с англ. – М. : Дело и Сервис, 1999. – 432 с.

271. Травкін Ю. І. Основи лінійної алгебри і її застосування : [посібник для студ. екон. спец. вищ. навч. закл.] / Ю. І. Травкін, Л. М. Малярець. – Харків : Основа, 2001. – 376 с.

272. Трояновский В. М. Математическое моделирование в менеджменте : учебн. пособие [для студ. высш. учебн. завед.] / В. М. Трояновский. – М. : Русская деловая литература, 1999. – 240 с.

273. Тридід О. М. Організаційно-економічний механізм стратегічного розвитку підприємства / О. М. Тридід. – Харків : Вид. ХДЕУ, 2002. – 364 с.

274. Тьюки Дж. Анализ результатов наблюдений. Разведочный анализ / Дж. Тьюки ; пер. с англ. – М. : Мир, 1981. – 696 с.

275. Тюрин Ю. Н. Статистический анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; [под ред. В. Э. Фигурнова]. – М. : ИНФРА-М, 1998. – 528 с.

276. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров. – [4-е изд.]. – М. : ИД "ФОРУМ", 2008. – 368 с.

277. Україна у вимірі економіки знань / [В. М. Геєць, В. П. Александрова, Ю. М . Бажал та ін.] ; за ред. акад. НАН України В. М. Геєця. – К. : Основа, 2006. – 592 с.

278. Уотшем Т. Дж. Количественные методы в финансах : [учеб. пособие для вузов] / Т. Дж. Уотшем, К. Паррамоу ; пер. с англ. – М. : Финансы, ЮНИ-ТИ, 1999. – 528 с.

279. Управление затратами на предприятии : [учеб. пособие / под общ. ред. Г. А. Краюхина]. − [2-е изд.]. – СПб. : ИД "Бизнес-пресса", 2003. – 256 с.

280. Факторный, дискриминантный и кластерный анализ : [сборн.] / пер. с англ. А. М. Хотинского, С. Б. Королева ; под ред. И. С. Енюкова. – М. : Финансы и статистика, 1989. – 216 с.

281. Фатхутдинов Р. А. Управленческие решения : [учебник] / Р. А. Фат-хутдинов. − [4-е изд.]. – М. : ИНФРА-М, – 2001. – 283 с.

282. Федосеев В. В. Экономико-математические методы и модели в мар-кетинге : [учеб. пособие для вузов] / В. В. Федосеев, Н. Д. Эриашвили. – [2-е изд.]. – М. : ЮНИТИ-ДАНА, 2001. – 160 с.

283. Фелпс Боб. Умные бизнес-показатели: Система измерений эффективности как важный елемент менеджмента / Боб Фелпс ; пер. с англ. – Днепропетровск : Баланс Бизнес Букс, 2004. – 312 с.

284. Ферстер Э. Методы корреляционного и регрессионного анализа: Ру-ководство для экономистов / Э. Ферстер, Б. Ренц ; пер. с нем. и предисл. В. М. Ивановой. – М. : Финансы и статистика, 1983. – 302 с.

285. Филипенко А. С. Экономическое развитие: цивилизационный подход / А. С. Филипенко. – М. : ЗАО "Экономика", 2002. – 260 с.

286. Фишберн П. Теория полезности для принятия решений / П. Фишберн ; пер. с англ. В. Н. Воробьйовой и А. Я. Кируты. – М. : Наука, 1978. – 352 с.

287. Філософський словник / [ред.-упоряд. В. І. Шинкарук]. – [2-е вид., оновл.] – К. : Голов. ред. УРЕ, 1986. – 800 с.

288. Форрестер Дж. Мировая динамика / Дж. Форрестер ; пер. с англ. А. Н. Ворощука, С. А. Иегова ; [под ред. Д. М. Гвишиани, Н. Н. Моисеева]. – М. : Наука, 1978. – 168 с.

289. Фридман А. А. Мир как пространство и время / А. А. Фридман. – [2-е изд.]. – М. : Наука, 1965. – 112 с.

290. Хазанова Л. Э. Математическое моделирование в экономике : [учеб. пособие] / Л. Э. Хазанова. – М. : БЕК, 1998. – 141 с.

291. Халафян А. А. STATISTICA 6. Статистический анализ данных : [учебник] / А. А. Халафян. – [3-е изд.]. – М. : ООО "Бином-Пресс". – 512 с.

292. Харківська область у 2005 році. – Харків : Держкомстат України, 2006. – 606 с. (статистичний щорічник).

293. Хачатрян С. Р. Прикладные методы математического моделирования экономических систем : [науч.-метод. пособие] / С. Р. Хачатрян. – М. : Экзамен, 2002. – 192 с.

294. Хеттманспергер Т. Статистические выводы, основанные на рангах / Т. Хеттманспергер ; пер. с англ. – М. : Финансы и статистика, 1987. – 334 с.

295. Хованов Н. В. Математические основы теории шкал измерения качества / Н. В. Хованов. – Л. : Изд. Ленингр. университета, 1982. – 188 с.

296. Хованов Н. В. Анализ и синтез показателей при информационном дефиците / Н. В. Хованов. – СПб. : Издательство С.-Петерб. универ., 1996. – 196 с.

297. Христиановский В. В. Процессы дестабилизации в производственных системах: аксиоматика и анализ / В. В. Христиановский. – Донецк : ИЭП НАН Украины; ДонГУ, 1998. – 212 с.

298. Христиановский В. В. Прикладная эконометрия : [учебник для экон. вузов] / В. В. Христиановский, Н. Г. Гузь, О. Г. Кривенчуг. – Донецк : ДонГУ, 1998. – 172 с.

299. Христиановский В. В. Функция полезности : теория и анализ: [учеб. пособие] / В. В. Христиановский, В. П. Щербина. – Харьков : ИД "ИНЖЭК", 2006. – 120 с.

300. Циба В. Т. Математичні основи соціологічних досліджень: кваліметричний підхід : [навч. посібник для студ. вищ. навч. закл.] / В. Т. Циба. – К. : МАУП, 2002. – 248 с.

301. Чавкин А. М. Методы и модели рационального управления в рыночной экономике: разработка управленческих решений : [учеб. пособие] / А. М. Чавкин. – М. : Финансы и статистика, 2001. – 320 с.

302. Черноруцкий И. Г. Методы оптимизации и принятия решений : [учеб. пособие] / И. Г. Черноруцкий. – СПб. : "Лань", 2001. – 384 с.

303. Черняк О. І. Виявлення ознак неплатоспроможності підприємства та можливого його банкрутства / О. І. Черняк // Статистика України. − 2003. − № 4. − С. 87−94.

304. Черняк О. І. Застосування дискримінантного аналізу для оцінки платоспроможності підприємств України / О. І. Черняк // Вісник Київського

національного університету імені Тараса Шевченка. Економіка. – 2006. − № 86−87. − С. 36−39.

305. Шелобаев С. И. Математические методы и модели в экономике, финансах, бизнесе : [учеб. пособие для вузов] / С. И. Шелобаев. – М. : ЮНИТИ-ДАНА, 2000. – 368 с.

306. Шикин Е. В. Математические методы и модели в управлении : [учеб. пособие] / Е. В. Шикин, А. Г. Чхартишвили. – М. : Дело, 2000. – 440 с.

307. Шлеер С. Объектно-ориентированный анализ: моделирование мира в состояниях / С. Шлеер, С. Меллор ; пер. с англ. – К. : Диалектика, 1993. – 240 с.

308. Шрейдер Ю. А. Системы и модели / Ю. А. Шрейдер, А. А. Шаров. – М. : Радио и связь, 1982 . − 152 с.

309. Штаркман Б. П. Оптимизация рецептуры и режима желатинизации пластизолей / Б. П. Штаркман, Т. М. Карташова, Э. А. Середа // Пластические массы. – 1969. – № 2. – С. 18–23.

310. Шурыгин А. М. Прикладная стохастика: робастность, оценивание, прогноз / А. М. Шурыгин. – М. : Финансы и статистика, 2000. – 224 с.

311. Щедровицкий Г. П. Избранные труды / Г. П. Щедровицкий. – М. : Шк. культ. полит., 1995. – 800 с.

312. Эйкхофф П. Основы идентификации систем управления. Оценивание параметров и состояния / П. Эйкхофф ; пер. с англ. В. А. Лотоцкого, А. С. Менделя. – М. : Мир, 1975. – 684 с.

313. Эконометрика : [учебник] / под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2004. – 344 с.

314. Экономика и статистика фирм : [учебник] / В. Е. Адамов, С. Д. Ильенкова, Т. П. Сиротина, С. А. Смирнов ; под ред. С. Д. Ильенковой. – [3-е изд. перераб. и доп.]. – М. : Финансы и статистика, 2000. – 288 с.

315. Экономико-математические методы и модели : [учеб. пособие] / под общ. ред. А. В. Кузнецова. – Мн. : БГЭУ, 1999. – 416 с.

316. Экономико-математические методы и прикладные модели : Учеб. пособие для вузов / под ред. В. В. Федосеева. – М. : ЮНИТИ, 1999. – 392 с.

317. Экономико-математический энциклопедический словарь / [гл. ред. В. И. Данилов-Данильян]. – М. : Большая Российская энциклопедия : Изд. дом "ИНФРА-М", 2003. – 688 с.

318. Экономико-математический словарь. Словарь современной экономической науки / [авт.-состав. Л. И. Лопатников] – М. : Изд. "ABF", 1996. – 704 с.

319. Экономическая кибернетика : [учеб. пособие в 4 ч.]. – Л. : ЛФЭИ, 1974. – . –

Ч. 4 : Основы системного анализа / подгот. Т. Г. Попова, И. М. Сыроежин, Ю. Н. Эйснерр. – 1976. – 136 с.

320. Экономический потенциал региона: анализ, оценка, диагностика / А. Н. Тищенко, Н. А. Кизим, А. И. Кубах, Е. В. Давыскиба. − Харьков : ИД "ИНЖЭК", 2005. – 176 с.

321. Економічна статистика : [навч. посібник] / за наук. ред. докт. екон. наук Р. М. Моторина. – К. : КНЕУ, 2005. – 362 с.

322. Эндрю Ф. Сигел. Практическая бизнес-статистика / Ф. Сигел Ендрю ; пер. с англ. – М. : Изд. дом "Вильямс", 2002. – 1 056 с.

323. Юкаева В. С. Управленческие решения : [учеб. пособие] / В. С. Юкаева. – М. : Издательский дом "Дашков и К°", 1999. – 292 с.

324. Bernoulli D. Exposition of New Theory on the Measurement of Risk // Econometrica. – 1954. – P. 23–36.

325. Campbell N. R. What is Science? – New York : Dover, 1952. – P. 109–134.

326. Coombs C. H. A Theory of Data. – New York : John Wiley, 1964. – 143 p.

327. Ellis B. Basic Concepts of Measurement. – London : Cambridge University Press, – 1966. – 41 p.

328. Foundations of measurement / [Krantz D. H., Luce R. D., Suppes P., Tversky A.]. – New York : Acad. Press, 1971 – 1990. – 1 – 3 v.

329. Harrington E.C.Jr. The desirability Function // Industr. Quality Control. - 1965. – № 10. – P. 494–498.

330. Kanger S. Measurement: An essay in philosophy of science // Theoria. – 1972. – № 38. − P. 44.

331. Luce R. D. Semionders and a theory of utility discrimination // Econometrics. – 1956. − № 24. – P. 178–191.

332. Pearson K. Contributions to mathematical theory of evolution. I. Dissection of frequency curves // Philosophical Transactions of the Royal Society of London. Ser. A. – 1894. – V. 185. – P. 71–110.

333. Recommendation of BIPM Working Group. Assignment of experimental uncertainties // INC. – Paris. − 1980. – P. 17–34.

334. Russell B. Principles of Mathematics. – London : Allen and Unwin, 1964. – 134 p.

335. Sccott D. Foundational aspects of theories of measurement / Sccott D., Suppes P. // Symbol Logic. – 1958. – № 23. − P. 113–128.

336. Suppes P. Measurement, empirical meaningfulness and threevalued logic // Measurement: definition and theories. – 1959. – P. 129–143.

337. Torgerson W. S. Theory and Method of Scaling. – New York : Willey, 1958. – 455 p.

338. Kraft O. A maximin linear estimator for linear parameters under restrictions in form of inequalities // Statistics. – 1986. – V. 17. – № 1. – P. 3–8.

339. http: //data.cemi.rssi.ru/isepweb/coun.htm

Додаток А

Основні показники виробничо-господарської діяльності 15 промислових підприємств за 5 років

№	*x*1	*x*2	*x*3	*x*4	*x*5	*x*6	*x*7	*x*8	*x*9	*x*10	*x*13	*x*14	*x*16	*x*17	*x*18	*x*19	*x*20	*x*21	*x*22	*x*24
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	79,00	21,00	44,50	10,14	92,35	4,66	2,99	65,19	86,00	46,00	0,73	6,90	0,72	0,41	0,72	-0,37	-0,84	2,24	-0,11	0,00
2	58,20	41,80	95,60	4,40	71,25	8,35	20,40	81,52	97,00	67,00	1,98	20,70	0,76	0,32	0,79	0,49	0,52	1,05	0,27	0,00
3	24,70	75,30	46,10	53,90	59,77	29,62	10,61	38,76	100,00	79,00	8,96	66,90	0,92	0,09	0,92	0,89	1,92	2,17	0,73	1,73
4	42,30	57,70	95,00	5,00	80,22	14,26	5,52	52,69	99,60	84,00	2,37	33,30	0,73	0,36	0,76	0,58	0,61	1,05	0,45	0,00
5	87,50	12,40	73,00	27,00	92,81	4,82	2,37	84,50	94,00	53,00	0,57	9,00	0,77	0,30	0,78	-0,76	-1,04	1,37	-0,12	0,00
6	54,00	47,00	59,00	41,00	89,90	8,79	1,31	58,02	0,97	65,00	1,76	20,50	0,68	0,47	0,73	0,43	0,73	1,70	0,30	0,00
7	88,90	11,10	73,00	27,00	94,19	5,00	0,81	93,18	93,00	51,00	0,40	17,10	0,69	0,45	0,72	-1,54	-2,11	1,37	-0,25	0,00
8	86,30	13,70	75,40	24,60	99,71	0,04	0,26	38,79	96,00	55,00	0,95	0,01	0,84	0,17	0,86	0,00	-0,08	1,32	-0,01	0,00
9	61,26	38,70	54,00	46,00	78,15	6,60	15,25	73,97	89,40	58,00	1,21	6,69	0,68	0,32	0,68	0,17	0,32	1,86	0,10	0,00
10	78,20	21,80	80,30	19,70	89,71	2,07	8,21	87,20	98,80	54,00	4,93	17,35	0,96	0,05	0,96	0,80	0,99	1,25	0,18	0,00
11	81,30	18,70	78,36	21,64	92,37	2,66	4,97	80,52	86,60	66,00	0,46	21,96	0,02	0,75	0,59	-1,17	-1,50	1,28	-0,38	0,00
12	80,60	19,40	49,40	50,60	95,61	3,54	0,85	84,34	96,80	65,00	1,64	7,54	0,82	0,22	0,88	0,39	0,79	2,02	0,09	0,14
13	87,30	12,70	81,80	18,20	98,57	0,96	0,46	88,13	99,50	31,00	0,53	11,06	0,75	0,33	0,76	-0,87	-1,07	1,22	-0,15	0,00
14	2,33	97,67	31,20	68,80	99,61	0,00	0,39	69,00	100,00	72,00	6,86	83,40	0,86	0,17	0,86	0,85	2,74	3,20	0,97	2,42
15	79,40	20,60	60,60	39,40	85,81	11,18	3,01	59,69	93,40	45,00	0,40	31,70	0,45	0,12	0,48	-1,54	-2,54	1,65	-0,70	0,00
16	81,97	18,03	45,17	54,83	91,20	5,78	3,01	71,41	54,75	51,00	1,42	5,34	0,80	0,24	0,87	0,30	0,66	1,35	0,05	0,00
17	56,50	43,43	95,22	4,78	65,24	10,62	24,14	86,02	47,01	67,00	1,57	15,84	0,70	0,43	0,72	0,36	0,38	0,79	0,16	0,00
18	19,53	80,47	53,15	46,85	47,26	42,29	10,45	38,74	68,56	74,00	8,74	71,26	0,91	0,10	0,91	0,89	1,67	1,69	0,71	2,03
19	41,00	59,00	67,43	32,57	64,82	26,10	9,08	63,26	95,00	84,00	2,03	29,99	0,69	0,45	0,71	0,51	0,75	1,06	0,30	0,00
20	88,76	11,23	77,70	22,30	91,10	6,39	2,50	86,32	45,18	55,00	0,52	10,52	0,78	0,28	0,78	-0,94	-1,91	0,60	-0,11	0,00
21	59,20	38,00	60,93	39,07	83,06	14,43	2,52	70,98	62,48	66,00	1,40	10,90	0,64	0,51	0,70	0,29	0,47	0,64	0,11	0,00
22	91,30	8,60	75,20	24,79	94,60	4,88	0,51	95,27	30,24	50,00	0,22	30,78	0,55	0,83	0,61	-3,54	-6,62	-2,64	-0,31	0,00
23	68,26	31,70	80,98	19,02	95,92	4,08	0,00	62,74	92,30	50,00	1,00	0,05	0,68	0,46	0,68	0,00	0,00	0,35	0,00	0,00

Продовження додатка А

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
24	4,37	95,60	24,94	75,06	30,83	0,76	68,41	7,82	87,40	58,00	12,83	88,14	0,92	0,08	0,92	0,92	3,70	3,70	0,88	5,08
25	78,18	21,80	56,16	43,84	90,17	3,42	6,41	86,70	44,39	56,00	5,00	17,40	0,96	0,05	0,96	0,80	1,42	1,55	0,17	0,03
26	80,20	19,80	79,50	20,45	85,59	10,57	3,84	86,00	82,61	68,00	0,58	14,90	0,54	0,84	0,65	-0,75	-1,64	-1,48	-0,15	0,00
27	84,70	15,30	30,95	69,05	98,01	1,02	0,98	86,45	60,19	66,00	2,01	7,66	0,92	0,08	0,92	0,50	1,62	2,49	0,08	0,11
28	89,52	10,48	67,60	32,38	98,76	1,12	0,13	90,17	23,44	32,00	0,45	12,98	0,75	0,33	0,76	-1,24	-12,9	-11,1	-0,13	0,00
29	51,20	48,70	89,06	10,94	68,99	17,21	13,81	69,89	68,41	73,00	1,36	12,77	0,64	0,56	0,64	0,26	0,29	0,55	0,13	0,01
30	82,20	17,80	55,40	44,60	86,71	11,56	1,73	59,41	55,31	48,00	0,25	53,30	0,29	0,25	0,31	-2,99	-7,75	-4,58	-0,53	0,00
31	78,50	21,50	47,88	52,12	91,57	5,18	3,25	69,11	54,00	53,00	1,40	6,18	0,83	0,21	0,85	0,29	0,60	1,44	0,06	0,00
32	58,09	41,60	93,62	6,37	64,25	11,50	24,26	75,63	50,00	70,00	1,83	19,02	0,76	0,31	0,77	0,45	0,48	0,88	0,19	0,00
33	29,82	69,67	53,42	46,58	63,30	30,46	6,24	44,51	67,00	57,00	5,78	57,62	0,86	0,16	0,87	0,83	1,55	1,74	0,57	1,21
34	37,53	62,47	65,27	29,56	63,45	28,14	8,41	59,16	94,80	86,00	3,09	42,24	0,74	0,34	0,80	0,68	1,04	1,11	0,42	0,47
35	87,07	12,92	63,82	36,18	91,96	5,72	2,33	83,22	45,00	58,00	0,47	14,78	0,72	0,38	0,72	-1,14	-3,11	0,65	-0,15	0,00
36	57,10	42,80	74,08	25,92	79,13	17,03	3,84	71,82	62,00	68,00	3,10	20,90	0,81	0,24	0,86	0,68	0,91	1,02	0,29	0,00
37	93,15	6,85	78,87	21,13	95,31	4,49	0,21	96,41	32,00	49,80	0,17	33,50	0,54	0,78	0,55	-4,89	-7,75	-4,08	-0,34	0,00
38	69,10	30,90	89,84	10,16	99,32	0,68	0,00	56,80	93,20	47,00	1,23	6,53	0,76	0,32	0,76	0,21	0,24	0,62	0,07	0,00
39	1,88	98,10	25,12	74,88	20,60	1,36	78,05	9,11	88,24	66,00	13,94	91,05	0,92	0,08	0,93	0,93	3,70	3,70	0,91	7,18
40	56,40	43,43	60,91	39,09	75,50	12,16	12,35	74,65	43,00	69,00	4,98	34,70	0,91	0,10	0,91	0,80	1,31	1,44	0,35	0,20
41	76,00	24,00	79,34	20,66	83,78	11,57	4,65	82,38	83,00	70,00	0,98	0,37	0,69	0,46	0,76	-0,02	-0,04	1,26	0,00	0,00
42	84,50	15,50	23,13	76,87	98,05	1,39	0,56	85,71	60,00	63,00	2,16	8,32	0,87	0,14	0,93	0,54	2,32	2,47	0,08	0,46
43	86,30	13,70	66,30	33,68	98,89	1,09	0,02	86,20	50,00	36,00	0,58	9,74	0,73	0,36	0,76	-0,71	-10,4	-3,72	-0,10	0,00
44	48,93	50,88	79,60	20,39	69,46	14,56	15,99	64,92	70,00	73,00	3,15	34,74	0,68	0,48	0,84	0,68	0,86	1,10	0,35	0,02
45	84,30	15,70	60,95	39,05	86,96	11,94	1,10	57,16	60,00	52,00	0,33	32,45	0,45	0,12	0,52	-2,07	-4,75	-1,90	-0,32	0,00
46	73,20	26,80	38,19	61,81	89,48	6,54	3,98	64,56	44,73	55,27	1,24	4,67	0,76	0,23	0,78	0,27	0,64	1,15	0,35	0,00
47	46,80	53,20	86,16	13,84	62,95	15,23	21,82	61,43	27,53	72,47	1,70	21,91	0,68	0,32	0,69	0,53	0,57	0,90	0,78	0,00
48	53,50	46,50	55,10	44,90	82,32	14,95	2,73	61,03	65,66	34,34	8,35	40,98	0,92	0,08	0,94	0,46	0,51	0,57	0,50	1,10
49	37,98	62,02	63,69	36,32	60,03	28,49	11,48	58,74	12,80	87,20	2,59	38,05	0,73	0,27	0,76	0,62	0,37	0,51	0,85	0,01

Продовження додатка А

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
50	64,94	35,05	76,57	23,43	68,23	25,17	6,61	56,86	17,83	82,17	1,01	0,65	0,27	0,72	0,64	0,35	0,24	0,49	1,32	0,01
51	56,14	43,68	78,78	21,22	78,58	18,26	3,16	70,90	29,68	70,32	2,11	23,21	0,77	0,23	0,79	0,44	0,34	0,50	0,56	0,00
52	94,03	5,97	93,20	6,80	94,90	4,92	0,18	97,69	48,75	51,25	0,15	32,65	0,61	0,39	0,61	0,06	0,12	0,93	0,10	0,00
53	28,26	71,69	53,98	46,02	21,86	14,51	63,63	49,53	78,12	21,88	1,89	33,74	0,62	0,38	0,62	0,72	1,00	1,53	1,16	0,03
54	1,93	98,00	23,85	76,15	18,38	1,46	80,16	8,47	28,28	71,72	34,12	95,20	0,97	0,03	0,97	0,98	7,91	8,14	1,01	9,63
55	53,97	45,94	62,69	37,31	74,48	12,42	13,10	72,28	32,76	67,24	5,25	37,27	0,91	0,09	0,91	0,46	0,51	0,61	0,51	0,49
56	60,32	39,68	70,70	29,30	71,26	19,39	9,36	78,36	19,12	80,88	0,59	27,58	0,17	0,83	0,33	0,40	0,26	0,70	2,40	0,00
57	71,13	28,87	10,56	89,44	97,56	1,77	0,67	71,95	38,88	61,12	3,27	20,03	0,91	0,09	0,91	0,29	2,26	2,95	0,32	1,65
58	76,04	23,96	65,11	34,89	91,59	8,41	0,00	75,09	60,04	39,68	0,82	5,34	0,71	0,29	0,71	0,24	0,35	0,77	0,34	0,03
59	47,62	51,75	77,29	22,71	60,87	19,29	19,85	59,92	22,12	77,88	2,77	23,56	0,58	0,32	0,71	0,52	0,45	0,61	0,89	0,00
60	66,70	33,30	28,18	71,82	89,49	6,56	3,96	58,99	44,78	55,22	1,45	9,95	0,75	0,24	0,77	0,33	0,86	1,45	0,44	0,00
61	46,00	54,00	86,46	13,54	62,56	15,40	22,04	60,67	27,53	72,47	1,69	22,12	0,67	0,33	0,68	0,54	0,58	0,92	0,80	0,00
62	53,20	46,80	55,30	44,70	82,39	14,89	2,73	60,73	65,36	34,64	8,35	41,19	0,92	0,08	0,94	0,47	0,52	0,58	0,51	1,09
63	37,65	62,35	63,68	36,32	60,57	27,75	11,68	58,82	12,80	87,20	2,65	38,81	0,74	0,26	0,76	0,62	0,38	0,53	0,85	0,01
64	64,88	35,11	76,56	23,44	68,16	25,21	6,62	56,82	17,52	82,48	1,02	0,56	0,27	0,72	0,64	0,35	0,25	0,49	1,32	0,01
65	56,40	43,43	79,03	20,97	78,92	17,87	3,21	70,94	29,68	70,32	2,14	23,28	0,78	0,22	0,80	0,43	0,34	0,50	0,56	0,00
66	94,02	5,97	93,18	6,82	94,90	4,93	0,18	97,69	48,75	51,25	0,15	32,65	0,61	0,39	0,61	0,06	0,12	0,93	0,10	0,00
67	28,28	71,68	54,09	45,91	21,91	14,49	63,60	49,65	78,12	21,88	1,90	33,94	0,62	0,38	0,62	0,72	1,00	1,52	1,15	0,03
68	1,85	98,08	23,68	76,32	18,71	1,58	79,71	8,21	27,97	72,03	40,29	95,71	0,98	0,02	0,98	0,98	7,56	7,74	1,01	10,23
69	54,08	45,83	62,71	37,30	74,58	12,37	13,05	72,34	32,76	67,24	5,65	37,80	0,92	0,08	0,92	0,46	0,51	0,60	0,50	0,53
70	60,30	39,70	70,63	29,37	71,26	19,39	9,36	78,31	19,12	80,88	0,59	27,56	0,17	0,83	0,33	0,40	0,26	0,70	2,41	0,00
71	71,23	28,77	10,30	89,70	97,62	1,74	0,64	71,96	38,81	61,19	3,33	20,14	0,91	0,09	0,91	0,29	2,30	2,99	0,31	1,68
72	76,13	23,87	65,12	34,88	91,63	8,37	0,00	75,09	60,04	39,68	0,82	5,36	0,71	0,29	0,71	0,24	0,35	0,77	0,34	0,03
73	47,25	52,12	77,62	22,38	60,80	19,27	19,93	59,47	22,11	77,89	2,70	23,44	0,57	0,33	0,71	0,52	0,45	0,62	0,91	0,00
74	82,62	17,37	55,10	44,90	87,06	12,02	0,91	56,36	44,11	55,89	0,36	30,93	0,52	0,48	0,52	0,17	0,26	0,97	0,34	0,00
75	68,60	31,35	81,87	18,13	74,80	22,36	2,84	82,75	23,80	76,20	0,74	10,98	0,54	0,46	0,58	0,31	0,17	0,40	0,58	0,00

Продовження додатка А

№	*x*25	*x*26	*x*27	*x*28	*x*29	*x*31	*x*32	*x*33	*x*34	*x*36	*x*37	*x*38	*x*40	*x*41	*x*42	*x*43	*x*45	*x*46	*x*47	*x*48
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0,40	-0,21	1708,00	46017,30	-7460,40	0,75	0,74	0,61	5,70	2,43	148,00	5,56	-9,20	-12,10	-15,5	-9,09	-0,09	26,94	-4,37	37,08
2	0,09	0,44	358,00	3092,30	0,00	0,33	0,32	0,14	0,53	0,95	380,00	11,11	0,40	0,50	1,43	1,26	0,50	8,64	0,00	20,06
3	4,83	1,92	157,00	7173,50	1669,00	1,67	1,67	1,20	2,33	13,60	27,00	6,52	45,90	49,40	28,81	27,14	0,34	45,69	10,63	8,31
4	0,11	0,56	126,00	207,50	-23,30	0,13	0,12	0,05	0,28	0,65	557,00	7,69	-1,10	-1,40	-8,99	-3,92	-0,18	1,65	-0,18	8,48
5	0,15	-1,17	4726,00	105429,20	-17392,50	0,45	0,44	0,22	3,80	1,78	202,00	10,00	-5,90	-7,40	-1,63	-11,1	-0,08	22,31	-3,68	54,99
6	0,72	0,56	540,00	7552,10	656,80	0,86	0,80	0,39	1,52	1,87	193,00	3,85	5,26	7,50	7,91	9,71	0,07	13,99	1,22	13,11
7	0,11	-0,25	3479,00	12637,10	-18983,90	0,07	0,07	0,03	1,53	0,50	712,00	1,54	-7,40	-10,3	-12,8	-16,2	-0,11	3,63	-5,46	65,29
8	0,23	-0,24	144,00	831,10	-389,90	0,17	0,16	0,09	1,04	0,92	393,00	3,85	-6,30	-7,90	-3,84	-6,22	0,05	5,77	-2,71	35,32
9	0,56	0,32	4352,00	1853,10	634,20	0,67	0,67	0,24	2,33	1,77	203,00	1,77	-6,16	-8,50	40,50	32,75	-0,09	0,43	0,15	10,19
10	0,97	0,99	189,00	1569,20	288,80	0,16	0,16	0,09	0,65	2,73	132,00	4,99	5,30	2,94	17,17	11,25	0,03	8,30	1,53	44,19
11	0,10	-0,17	2615,00	16551,80	-4299,40	0,28	0,27	0,09	1,45	0,63	573,00	5,41	-4,76	-7,22	-2,55	-2,53	-0,09	6,33	-1,64	28,66
12	0,82	0,13	209,00	2511,80	-46,40	0,68	0,63	0,24	4,90	2,80	129,00	8,33	-1,01	-1,25	-1,61	-10,4	-0,01	12,02	-0,22	16,99
13	0,10	-1,20	90,00	194,20	-118,80	0,06	0,06	0,04	0,86	0,39	915,00	2,86	-2,93	-3,84	-4,32	-10,3	0,04	2,16	-1,32	39,75
14	4,72	2,74	4352,00	64405,50	-8115,20	0,71	0,71	9,96	1,07	2,63	137,00	4,84	31,72	37,17	-12,3	-15,3	0,22	14,80	-1,86	19,06
15	0,16	-0,19	1682,00	31988,40	-9071,80	1,15	1,10	0,71	5,81	1,76	205,00	11,33	-18,50	-32,93	-2,72	-2,69	-0,41	19,02	-5,39	23,17
16	0,78	-0,09	1485,00	33982,40	-1179,90	0,57	0,43	0,53	3,33	1,51	237,74	4,33	-1,43	-1,88	-2,71	20,91	-0,02	22,88	-0,79	42,71
17	0,08	0,30	388,00	3097,50	-628,70	0,36	0,29	0,48	0,56	0,92	392,13	14,09	-5,37	-7,38	-1,68	-0,70	-0,08	7,98	-1,62	17,28
18	4,10	0,89	173,00	14811,50	3273,00	2,28	2,05	8,50	2,03	18,49	19,47	9,80	55,03	60,41	19,57	37,75	0,40	85,62	18,92	8,85
19	0,66	0,47	127,00	456,40	-119,00	0,39	0,32	0,68	0,65	0,95	379,71	2,79	-7,20	-10,13	-1,89	-7,54	-0,11	3,59	-0,94	5,42
20	0,12	-0,94	4318,00	82361,80	-48008,40	0,39	0,33	0,36	3,33	1,29	278,25	12,72	-17,65	-22,76	-4,30	-7,25	-0,24	19,07	-11,1	55,52

Продовження додатка А

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
21	0,55	0,14	460,00	4564,60	-930,00	0,58	0,45	0,72	1,34	1,07	336,99	2,52	-7,82	-11,76	-1,92	10,29	-0,13	9,92	-2,02	14,36
22	0,05	-0,42	2520,00	12678,70	-41897,80	0,08	0,06	0,06	1,04	0,23	1562,56	2,03	-17,33	-27,98	-2,44	-4,20	-0,32	5,03	-16,6	86,40
23	0,19	0,00	105,00	251,30	0,00	0,13	0,10	0,12	0,00	0,00	0,00	1,63	0,00	0,00	0,00	0,00	0,00	2,39	0,00	49,26
24	9,37	0,92	3415,00	2202,30	394,20	0,72	0,68	14,30	1,46	3,48	103,48	1,99	0,88	19,34	15,76	39,47	0,12	0,10	0,00	0,10
25	2,19	0,80	186,00	2115,30	63,50	0,22	0,19	0,27	1,40	3,83	93,91	3,23	0,62	0,65	1,99	6,55	0,01	11,37	0,34	42,86
26	0,12	-1,30	2631,00	18681,40	-9002,00	0,40	0,28	0,29	1,24	0,52	693,44	7,35	-10,86	-19,47	-3,99	-1,42	-0,20	7,10	-3,42	25,44
27	1,39	0,50	211,00	2375,10	-558,20	0,66	0,55	0,74	18,60	8,28	43,48	7,44	-13,53	-15,61	-14,9	-5,71	-0,16	11,26	-2,65	16,13
28	0,14	-1,37	95,00	449,70	66,00	0,15	0,13	0,13	1,08	0,31	1143,41	3,85	1,74	2,32	8,45	18,83	0,02	4,73	0,69	36,34
29	0,15	0,26	5054,00	119835,90	5405,40	1,59	1,59	1,59	1,59	1,59	1,59	1,59	9,46	7,76	4,98	17,73	0,06	23,71	1,07	15,40
30	0,11	-0,30	1194,00	9792,80	-9239,20	0,56	0,44	0,27	2,94	0,50	717,41	3,77	-19,86	-53,24	-8,08	-4,87	-0,72	8,20	-7,74	31,47
31	0,73	0,20	1226,00	34509,50	3261,30	0,58	0,46	0,60	3,30	3,59	100,39	5,31	4,65	5,91	6,61	11,43	0,05	28,15	2,66	51,73
32	0,01	0,43	301,00	4974,80	582,00	0,58	0,50	0,77	0,88	1,36	264,68	20,85	4,97	7,11	7,81	6,77	0,07	16,53	1,93	22,28
33	2,69	0,80	196,00	23055,90	2980,20	3,55	2,09	6,08	2,64	14,31	25,16	14,33	50,35	37,88	12,75	33,90	0,27	117,63	15,21	7,82
34	1,07	0,64	153,00	673,20	49,80	0,57	0,46	1,14	0,94	1,52	236,70	3,88	4,30	6,22	6,18	-4,63	0,04	4,40	0,33	4,50
35	0,17	-1,14	4310,00	69783,20	-22664,80	0,33	0,31	0,32	3,25	1,11	324,35	17,98	-8,33	-11,95	-2,72	-11,5	-0,13	16,19	-5,26	55,62
36	1,01	0,55	509,00	7139,60	1193,30	0,90	0,72	1,19	1,22	2,04	176,58	7,93	11,20	17,34	15,11	26,44	0,14	14,03	2,34	12,97
37	0,04	-0,58	1633,00	4877,50	-3750,40	0,03	0,03	0,02	0,45	0,08	4243,89	1,19	-1,55	-3,04	-12,0	-2,63	-0,03	2,99	-2,30	133,32
38	0,13	0,21	92,00	284,00	0,00	0,14	0,11	0,14	0,16	0,18	1997,45	2,05	0,00	0,00	0,00	0,00	0,00	3,09	0,00	56,22
39	10,44	0,92	3357,00	2281,80	213,20	0,75	0,65	32,69	1,47	5,15	69,97	2,33	0,61	12,34	12,50	36,33	0,06	0,10	0,00	0,10
40	1,95	0,80	186,00	2890,70	30,30	0,30	0,35	0,91	1,34	4,48	80,30	3,44	0,38	0,53	0,96	19,38	0,01	15,54	0,16	42,86
41	0,20	-0,31	2370,00	20365,30	8338,40	0,44	0,32	0,35	1,18	0,81	446,73	10,03	10,06	17,23	18,08	-0,38	0,16	8,59	3,52	28,24

Продовження додатка А

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
42	1,66	0,19	199,00	2056,70	45,10	0,58	0,47	0,60	15,73	7,90	45,56	9,19	0,64	0,75	0,74	-2,54	0,01	10,34	0,23	17,11
43	0,20	-0,96	61,00	467,10	-114,40	0,16	0,13	0,14	1,34	0,52	696,61	3,47	-34,91	-48,64	-15,3	-13,6	-0,04	7,66	-1,88	56,60
44	0,64	0,37	5487,00	166505,00	11340,40	1,82	1,14	2,14	2,01	3,82	94,19	14,37	18,47	13,66	6,72	8,58	0,10	30,35	2,07	14,19
45	0,13	-0,25	1117,00	10994,60	7084,80	0,63	0,48	0,31	2,18	0,44	811,31	8,99	14,16	40,57	27,63	0,59	0,36	9,84	6,34	33,64
46	0,92	0,12	1050,00	299458,00	-20792,00	0,51	0,50	0,60	2,77	1,81	201,94	6,21	0,19	14,29	27,81	27,81	-0,04	28,52	19,80	42,01
47	0,97	0,40	408,00	76146,00	3918,00	0,75	0,86	1,32	0,99	1,88	193,93	16,20	0,35	19,89	23,09	23,09	0,04	18,66	0,00	12,27
48	6,42	0,83	208,00	228969,00	22046,00	3,49	1,07	1,78	2,16	18,16	20,10	19,00	0,91	49,66	45,24	45,24	0,11	110,08	14,49	7,52
49	1,61	0,57	180,00	14002,00	-642,00	0,59	1,22	2,20	1,76	3,88	94,16	2,77	0,66	31,86	25,13	25,13	-0,06	7,78	3,57	3,95
50	0,49	-1,10	3249,00	56947,00	-109714,0	0,37	0,09	1,18	1,98	1,16	315,00	3,16	-1,23	-179,2	-8,57	-8,57	-0,40	1,75	33,77	22,25
51	1,38	0,49	409,00	45483,00	-6788,00	0,79	0,57	0,68	0,94	0,18	2057,27	15,32	0,23	16,82	28,70	28,70	-0,09	11,12	16,60	13,66
52	0,03	-0,55	416,00	32965,00	-80134,00	0,03	0,03	0,02	0,24	2,24	162,82	3,43	0,00	0,57	-2,07	-2,07	-0,07	7,92	1,93	136,67
53	0,87	0,47	189,00	17952,00	370,00	0,14	2,21	10,53	2,48	0,04	8378,18	5,25	3,77	65,83	29,83	29,83	0,05	9,50	0,00	27,55
54	3,17	0,97	375,00	23473,00	2761,00	0,15	0,64	38,34	1,71	3,64	100,39	2,54	14,11	22,60	35,46	35,46	0,08	6,26	0,00	0,16
55	3,14	0,81	176,00	29718,00	286,00	0,41	0,58	0,86	1,28	5,54	66,00	5,12	0,29	17,23	29,80	29,80	0,01	16,89	0,00	37,29
56	0,26	-1,10	1366,00	126021,00	-275145,0	0,71	0,94	0,48	1,28	6,49	56,24	5,59	-0,09	-31,36	-1,68	-16,8	-4,08	9,23	2,01	16,71
57	3,07	0,69	209,00	24539,00	1493,00	0,16	0,66	0,77	20,16	7,46	49,00	10,25	-0,01	-1,09	-1,64	-1,64	0,04	11,74	0,00	13,63
58	0,58	-0,22	77,00	7118,00	956,00	0,41	0,25	0,20	0,90	0,75	487,00	2,34	0,05	5,18	20,71	18,82	0,03	9,24	0,00	39,41
59	1,51	0,20	5427,0	1760316,0	16127,00	0,15	1,68	2,87	2,14	0,60	605,19	29,86	0,92	58,19	28,21	28,21	0,02	32,44	1,99	9,92
60	1,18	0,26	1433,0	299436,00	-20799,00	0,45	0,47	0,60	2,76	1,56	233,30	6,17	0,19	13,25	27,81	27,81	-0,03	20,90	14,51	30,79
61	0,98	0,39	640,00	76153,00	3939,00	0,75	0,85	1,32	0,95	1,81	201,58	21,76	0,35	19,83	23,09	23,09	0,04	11,90	0,00	7,82
62	6,44	0,83	254,00	228954,00	22038,00	3,56	1,07	1,79	2,13	18,00	20,28	18,98	0,91	49,56	45,24	45,24	0,11	90,14	15,01	7,88

Закінчення додатка А

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
63	1,66	0,58	186,00	14000,00	-654,00	0,49	1,23	2,19	1,77	3,99	91,56	2,84	0,66	32,01	25,14	25,14	-0,06	7,53	3,52	3,86
64	0,49	-1,09	3089,0	56936,00	-109722,0	0,25	0,09	1,19	1,98	1,16	315,00	3,16	-1,23	-179,8	-8,57	-8,57	-0,40	1,84	35,52	22,84
65	1,40	0,49	129,00	45479,00	-6745,00	0,82	0,57	0,68	0,95	2,20	166,25	15,36	0,23	16,82	28,70	28,70	-0,09	35,26	52,29	13,32
66	0,03	-0,55	315,00	32969,00	-80131,00	0,04	0,03	0,02	0,24	0,04	8377,43	3,43	0,00	0,57	-2,07	-2,07	-0,07	10,47	2,54	176,68
67	0,87	0,47	187,00	17942,00	369,00	0,15	2,21	10,52	2,49	3,60	101,38	5,25	3,75	65,60	29,72	29,72	0,05	9,59	0,00	27,51
68	3,75	0,98	376,00	23471,00	2756,00	0,16	0,64	39,06	1,73	21,55	16,93	2,37	14,36	22,62	35,49	35,49	0,07	6,24	0,00	0,15
69	3,38	0,82	173,00	29705,00	278,00	0,63	0,57	0,85	1,29	6,01	60,76	5,11	0,29	17,08	29,71	29,71	0,01	17,17	0,00	37,59
70	0,26	-1,10	390,00	126078,00	-275103,0	0,16	0,94	0,48	1,28	0,46	792,54	5,54	-0,09	-31,29	-1,68	-16,7	-4,07	32,33	2,45	28,52
71	3,13	0,70	194,00	24529,00	1491,00	0,12	0,66	0,77	20,77	6,84	53,37	9,94	-0,01	-1,07	-1,62	-1,62	0,04	12,64	0,00	14,66
72	0,58	-0,22	83,00	7104,00	957,00	0,27	0,25	0,20	0,90	0,69	529,00	2,34	0,05	5,10	20,43	17,63	0,03	8,56	0,00	36,56
73	1,49	0,19	5729,00	1760402,0	16226,00	0,36	1,67	2,87	2,10	0,60	604,33	29,84	0,92	58,16	28,22	28,22	0,02	30,73	2,04	9,40
74	0,21	-0,18	862,00	127734,00	26387,00	0,22	0,61	0,55	2,73	6,13	59,54	7,00	0,11	10,31	16,95	16,95	0,13	14,82	6,28	23,13
75	0,25	-0,46	465,00	92469,00	-140906,0	0,96	0,30	0,23	0,95	0,65	561,01	2,99	-0,11	-12,84	10,07	-4,00	-0,49	19,89	2,85	70,54

Зміст

НАВЧАЛЬНЕ ВИДАННЯ

Пономаренко Володимир Степанович
Малярець Людмила Михайлівна

БАГАТОВИМІРНИЙ АНАЛІЗ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

Навчальний посібник

Відповідальний за випуск **Золотарьова І. О.**

Відповідальний редактор **Сєдова Л. М.**

Редактор **Грицай І. М.**

Коректор **Муштай Т. О.**

План 2009 р. Поз. №140-П.

Підп. до друку Формат 60 × 90 1/16. Папір MultiCopy. Друк Riso.

Ум.-друк. арк. 24,0. Обл.-вид. арк. 30,0. Тираж прим. Зам. №

Видавець і виготівник — видавництво ХНЕУ, 61001, м. Харків, пр. Леніна, 9а

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи
Дк №481 від 13.06.2001 р.

Пономаренко В. С.
Малярець Л. М.

БАГАТОВИМІРНИЙ АНАЛІЗ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

Навчальний посібник