

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

Єгоршин О. О.

Малярець Л. М.

**Лабораторний практикум
з навчальної дисципліни
"Економіко–математичні
методи та моделі: Економетрика"**

Навчально–практичний посібник

Харків. Вид. ХНЕУ, 2011

УДК 330.43(076.5)

ББК 6566я73

Є 30

Рецензенти: докт. екон. наук, проф., завідувач кафедри математики і математичних методів економіки Донецького національного університету Христіановський В. В., д.е.н., проф., завідувач кафедри статистики та прогнозування Харківського національного економічного університету Раєвнева О.В.

Рекомендовано до видання рішенням Вченої ради Харківського національного економічного університету як навчально-практичний посібник
Протокол № 6 від 28.02.2011 р.

Єгоршин О. О.

Є 72 Лабораторний практикум з економетрики навчальної дисципліни "Економіко–математичні методи та моделі". Начально–практичний посібник / О. О. Єгоршин, Л. М. Малярець. – Харків: Вид. ХНЕУ, 2011 – 148 с. (Укр. мов.)

Подано лабораторні роботи з даної навчальної дисципліни, що передбачені галузевим стандартом вищої освіти України. Всі розрахунки реалізовані в середовищі Excel, запропоновані шаблони для розв'язання типових задач. У додатку наведені реальні економічні дані України з офіційних джерел для завдань самостійної роботи.

Рекомендовано для студентів галузі знань «Економіка підприємства» всіх форм навчання

УДК

ББК

© Харківський національний економічний університет, 2011 р.

© Єгоршин О. О., Малярець Л. М., 2011 р.

Вступ

Електронна таблиця Excel є універсальним обчислювальним інструментом, знання якого обов'язкове для студентів всіх напрямів підготовки галузі знань 0305 "Економіка підприємства". Вивчення проблем економетрії в середовищі Excel має певні переваги порівнянно з використанням спеціалізованих статистичних пакетів, де усі розрахунки відбуваються автоматично й користувач одержує відразу готовий результат.

Метою написання навчально-практичного посібника було сформувати навички, які мають здобути студенти у процесі вивчення навчальної дисципліни "Економіко-математичні методи і моделі. Економетрика", а саме: визначати ознаковий простір для опису реальних об'єктів і процесів в економіці; формувати репрезентативні сукупності спостережень ознак економічних процесів, перевіряти їх однорідність; моделювати залежності результативних ознак від пояснюючих факторів (тобто складати структурні економетричні моделі); аналізувати отримані результати; вміти визначати стандартні похибки параметрів моделі і перевіряти їх значущість; вміти застосовувати статистичні критерії для перевірки якості моделі у цілому; знати стандартні прийоми подолання небажаних наслідків мультиколінеаності факторів, гетероскедастичності і автокореляції залишків моделі; знати, як враховувати у моделі якісні змінні, коли серед факторів присутні як кількісні, так і якісні змінні; вміти користуватися інструментальними і заміщаючими змінними, моделювати вплив запізнюючої дії факторів на результативні ознаки; вміти складати моделі з розподіленими лагами і авто-регресійні моделі; вміти аналізувати такі моделі і визначати як короткочасні, так і довгострокові прогнози; вміти аналізувати часові ряди спостережень для прогнозування розвитку об'єктів, явищ і процесів в економіці.

У процесі користування посібником студенти засвоюють особливості обчислювальних алгоритмів і доводять теоретичні міркування до практичних розрахунків. Усе це передбачає подальший змістовний аналіз проблеми, інтерпретацію отриманих результатів і економічні висновки.

З огляду на таку мету, в додатку наведені реальні економічні дані з офіційних джерел інформації України.

Будь-яка робота (а деякі лабораторні роботи з економетрії являють собою невелике наукове дослідження) повинне завершуватися повним і зрозумілим звітом. При використанні Excel сам робочий лист може бути таким звітом, якщо там додані відповідні пояснення. Проте у середовищі Excel дуже незручно набирати тексти математичних формул, які необхідно навести у звіті. І хоча, у принципі, є можливість при наборі формул використовувати грецькі букви, а також верхні або нижні індекси, по-символьне редагування тексту в Excel вимагає невиправдано великих працездат. Лише в особливих випадках слід форматувати тексти з урахуванням зазначених можливостей, але робити це треба як можна рідше, коли обійтися без цього не можна. Тому трохи спростимо стандартні позначення математичної статистики. Так, замість S_y цілком припустимо використовувати Sy , замість $r_{xy} - R_{xy}$, вираження R^2 іноді записують як $R2$. Позначення для коваріації S_{xy} цілком припустимо замінити на Sxy . Дисперсія є частковим випадком коваріації, тому замість S_y^2 пропонується позначення Syy . Середнє \bar{y} завжди можна записати як Y_{cp} , тому відомі формули для коваріації й дисперсій приймають цілком знайомий вигляд: $Sxy = (XY)_{cp} - (X_{cp}) \cdot (Y_{cp})$; $Syy = (YY)_{cp} - (Y_{cp})^2$. Зручного позначення для кореня квадратного в Excel немає, тому будемо використовувати відповідне ім'я вбудованої функції Excel $Sy = \text{КОРЕНЬ}(Syy)$. В аналогічних випадках також будемо використовувати імена вбудованих функцій Excel, наприклад, $SSE = \text{СУММКВ}(E)$.

Якщо обчислення на робочому листі Excel організовані раціонально, то цей робочий лист можна використовувати як готовий шаблон для виконання зазначених обчислень із будь-якими іншими даними. Бажане із самого початку передбачити можливість додавання нових і видалення зайвих даних, так щоб при цьому не було потрібно вносити в розрахунки ніяких змін. Тоді виконана робота не буде одноразовою й за необхідністю може бути використана надалі.

Тема 1. Парна лінійна регресія

Лабораторна робота № 1. Лінійна кореляційна модель

Мета роботи. За даними спостережень (X,Y) потрібно:

а) оцінити параметри (коефіцієнти регресії) лінійної моделі

$$Y_p = b_0 + b_1 * X$$

за відомими формулами:

$$b_1 = R_{xy} * S_y / S_x,$$
$$b_0 = Y_{cp} - b_1 * X_{cp},$$

де R_{xy} – коефіцієнт парної кореляції

$$R_{xy} = S_{xy} / (S_x * S_y);$$

S_{xy} , S_x , S_y – коваріація і стандартні відхилення:

$$S_{xy} = (XY)_{cp} - (X_{cp}) * (Y_{cp}),$$
$$S_{xx} = (XX)_{cp} - (X_{cp})^2,$$
$$S_{yy} = (YY)_{cp} - (Y_{cp})^2,$$
$$S_x = \text{КОРЕНЬ}(S_{xx}), \quad S_y = \text{КОРЕНЬ}(S_{yy});$$

б) побудувати графік знайденої залежності разом з емпіричними точками;

в) дати інтерпретацію отриманим результатам;


г) оформити роботу у вигляді звіту на робочому листі Excel.

Зразок звіту показаний на рис. 1, де використані наступні прийоми.



Тексти формул вирівняні *по правому краю*, а в сусідніх клітинках обчислені результати і вирівняні *по лівому краю*, тоді одержано зрозумілі записи типу: $R_{xy} = S_{xy} / (S_x * S_y) = 0,99019$.



Напівжирний шрифт можна включати / виключати кнопкою , або клавіатурною комбінацією **Ctrl+B**.

Числові дані в таблиці оформлені зверху і знизу рядком заголовка з іменами стовпців. Будь-яка таблиця завжди має рядок заголовка зверху. Якщо таблиця дуже велика, такий же рядок заголовка знизу допомагає визначити, де який обчислений підсумок. Додавання текстового рядка наприкінці таблиці має ще інше призначення. Наш аркуш Excel – це шаблон, що дозволяє робити зазначені розрахунки для *будь-яких* даних. Якщо нових даних менше, зайві рядки просто віддаляються і таблиця скорочується. Якщо нових даних буде більше,

потрібно додавати в таблицю нові рядки. Саме для можливості розширення діапазонів (без корегування формул) у діапазон доцільно включити також текстовий рядок заголовків знизу.

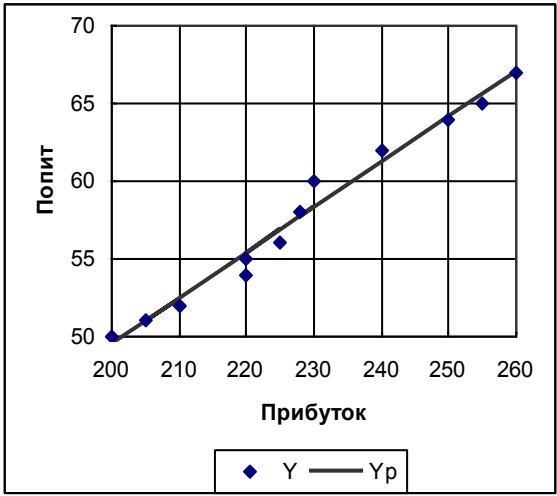
| | A | B | C | D | E | F | G | H | | | | |
|----|--|----------------|----------------|----------------|---|-------------|----------------|---|-----------|-----------|-----------|--|
| 1 | Робота № 1. Для заданих значень X, Y розрахувати за відомими формулами | | | | | | | | | | | |
| 2 | параметри лінійної моделі і побудувати графік залежності: Y – попит, X – прибуток | | | | | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | Лінійна модель | | | |  | | | | | | | |
| 5 | $Y_p = b_0 + b_1 \cdot X$ | | | | | | | | | | | |
| 6 | Коваріація і дисперсії | | | | | | | | | | | |
| 7 | $S_{xy} = (XY)_{cp} - (X_{cp}) \cdot (Y_{cp}) =$ 101,431 | | | | | | | | | | | |
| 8 | $S_{xx} = (XX)_{cp} - (X_{cp})^2 =$ 346,243 | | | | | | | | | | | |
| 9 | $S_{yy} = (YY)_{cp} - (Y_{cp})^2 =$ 30,3056 | | | | | | | | | | | |
| 10 | Середні квадратичні відхилення | | | | | | | | | | | |
| 11 | $S_x = \text{КОРЕНЬ}(S_{xx}) =$ 18,6076 | | | | | | | | | | | |
| 12 | $S_y = \text{КОРЕНЬ}(S_{yy}) =$ 5,50505 | | | | | | | | | | | |
| 13 | Коефіцієнт кореляції | | | | | | | | | | | |
| 14 | $R_{xy} = S_{xy} / (S_x \cdot S_y) =$ 0,99019 | | | | | | | | | | | |
| 15 | Коефіцієнти регресії | | | | | | | | | | | |
| 16 | $b_1 = R_{xy} \cdot S_y / S_x =$ 0,29295 | | | | | | | | | | | |
| 17 | $b_0 = Y_{cp} - b_1 \cdot X_{cp} =$ -9,1293 | | | | | | | | | | | |
| 18 | | | | | | | | | | | | |
| 19 | № | X | Y | XX | | | | | XY | YY | Yp | |
| 20 | 1 | 200 | 50 | 40000 | | | | | 10000 | 2500 | 49,46 | |
| 21 | 2 | 210 | 52 | 44100 | 10920 | 2704 | 52,3894 | | | | | |
| 22 | 3 | 205 | 51 | 42025 | 10455 | 2601 | 50,9247 | | | | | |
| 23 | 4 | 220 | 55 | 48400 | 12100 | 3025 | 55,3189 | | | | | |
| 24 | 5 | 225 | 56 | 50625 | 12600 | 3136 | 56,7836 | | | | | |
| 25 | 6 | 220 | 54 | 48400 | 11880 | 2916 | 55,3189 | | | | | |
| 26 | 7 | 230 | 60 | 52900 | 13800 | 3600 | 58,2483 | | | | | |
| 27 | 8 | 228 | 58 | 51984 | 13224 | 3364 | 57,6624 | | | | | |
| 28 | 9 | 240 | 62 | 57600 | 14880 | 3844 | 61,1778 | | | | | |
| 29 | 10 | 250 | 64 | 62500 | 16000 | 4096 | 64,1073 | | | | | |
| 30 | 11 | 255 | 65 | 65025 | 16575 | 4225 | 65,572 | | | | | |
| 31 | 12 | 260 | 67 | 67600 | 17420 | 4489 | 67,0367 | | | | | |
| 32 | № | X | Y | XX | XY | YY | Yp | | | | | |
| 33 | Середні | 228,583 | 57,8333 | 52596,6 | 13321,2 | 3375 | 57,8333 | | | | | |

Рис. 1. Зразок звіту до лабораторної роботи № 1

У підсумковому рядку таблиці обчислені середні за допомогою функції СРЗНАЧ, в якій задані діапазони разом текстовими рядками. Так, для обчислення X_{cp} у клітинці В33 набрана формула =СРЗНАЧ(В19:В32), яка потім скопійована в сусідні клітинки. Після розрахунку відповідних середніх проводиться розрахунок параметрів лінійної моделі за наведеними вище формулами. Так, у клі-

тинці D7 набрана формула =E33-B33*C33, що відповідає тексту формули ліворуч (у клітинці C7) для підрахунку S_{xy} .

Не треба самому набирати адреси операндів формули, після набору знака рівності (=) потрібно вказувати мишкою на відповідні клітинки із проміжними значеннями.

У відповідальних випадках можна дати змістовні імена деяким ключовим клітинкам, наприклад, $X_{\text{ср}}$, $X_{\text{ср}}$, $U_{\text{ср}}$. Тоді усі формули приймуть зрозумілий вигляд (типу = $X_{\text{ср}}$ - $X_{\text{ср}}$ * $U_{\text{ср}}$), що дуже зручно для перевірки.


Формула для обчислення розрахункових значень Y_p набрана лише для першого спостереження (у клітинці G20 набираємо: =\$D\$17+\$D\$16*B20); ця формула скопійована в розташовані нижче клітинки стовпця G. Адреси параметрів моделі b0, b1 повинні бути абсолютними, щоб вони не змінювалися при копіюванні.

Фіксатори (\$) не слід набирати вручну. При редагуванні формули треба встановити курсор ліворуч або праворуч адреси операнда і натиснути клавішу **F4**. При послідовних натисканнях цієї клавіші циклічно фіксується вся адреса або його частина: D17 → \$D\$17 → D\$17 → \$D17 → D17.



Для побудови графіка треба виділити значення X , Y і Y_p (разом із заголовками зверху) і викликати Майстер Діаграм. Для виділення несуміжних діапазонів треба натиснути і утримувати клавішу **Ctrl** (спочатку мишкою виділяються суміжні діапазони X і Y , потім при натиснутій клавіші **Ctrl** виділяється діапазон Y_p). Якщо виділяти тільки числа, то імена рядів даних на діаграмі прийдеться задавати додатково (за замовчуванням будуть задані невиразні імена Ряд 1, Ряд 2 і т. д.). До речі, нові дані, а також імена стовпців можна додати пізніше.

Відповідно до запиту Майстра Діаграм (кроки 1, 2) треба вказати (російською мовою): *Тип диаграммы – Точечная, Вид диаграммы – Точечная диаграмма, на которой значения соединены отрезками*. Далі (крок 3) варто вказати

місце розміщення легенди – *внизу* і включити (поставити прапорець)  *Основные линии сетки по оси X* (по осі У вони включені за замовчуванням).

Нагадуємо, що всі повідомлення *русифікованої версії* Excel виділені в тексті курсивом, або великими літерами.

При побудові графіків варто дотримуватися деяких вимог (ДСТУ). Графіки повинні будуватися разом з координатною сіткою, причому *Цена основных делений* по осях з точністю до множника $10^{\pm m}$ може дорівнювати 1, 2, або 5 (наприклад, 10, 20, 50, або 100, 200, 500 і ін.). Мітки по осях повинні бути кратні кроку. У нашому випадку треба задати наступний формат осі Х: *Минимальное значение – 200, Максимальное значение – 260, Цена основных делений – 10, Цена промежуточных делений – 5, Штрихи основных делений – внутрь*. Формат осі В: *Минимальное значение – 50, Максимальное значение – 70, Цена основных делений – 5, Цена промежуточных делений – 2, Штрихи основных делений – внутрь*.

Для того, щоб викликати панель формату осі, треба клацнути правою кнопкою мишки по будь-якому місці на осі, тоді у контекстному меню з'явиться пункт *Формат оси*.

Змінимо фон діаграми на білий, для чого клацнемо правою кнопкою мишки по області побудови діаграми, тоді у контекстному меню з'явиться пункт *Формат области построения*.

На нашій діаграмі є два графіки з маркерами. Клацнемо правою кнопкою мишки по ряду даних У, у контекстному меню виберемо пункт *Формат рядов данных*, де задамо: *Линия – отсутствует, Цвет маркера – черный*. Потім форматуємо інший ряд – ряд Ур. Виключаємо маркери і залишаємо лінію (товщину збільшуємо, задаємо чорний колір лінії).

Нарешті, у вікні *Параметры диаграммы* треба задати заголовки осей: по осі Х – Прибуток, по осі У – Попит.

Змінимо розмір діаграми і перенесемо її в бажане місце звіту.

Отримано рівняння регресії

$$Y_p = -9,1293 + 0,29395 * X$$

з коефіцієнтом кореляції $R_{xy} = 0,99019$.

Попит на товар на $R^2 = (R_{xy})^2 = 0,99019^2 = 0,98047 = 98\%$ визначається величиною особистого доходу (лінійна модель пояснює $R^2 = 98\%$ мінливості попиту на товар).

Коефіцієнт регресії $b_1 = 0,29295$ показує, що збільшення доходу на 10 одиниць призводить до збільшення попиту на 2,9.

Контрольне запитання. Як інтерпретувати вільний член $b_0 = -9,1293$? Чи не має непорозумінь внаслідок від'ємного знака цього коефіцієнта?

| № | X | Y |
|----|------|-------|
| 1 | 2310 | 142 |
| 2 | 2333 | 144 |
| 3 | 2356 | 151 |
| 4 | 2379 | 150 |
| 5 | 2402 | 139 |
| 6 | 2425 | 169 |
| 7 | 2448 | 126 |
| 8 | 2471 | 142,9 |
| 9 | 2494 | 163 |
| 10 | 2517 | 169 |
| 11 | 2540 | 149 |

Збережемо робочу книгу. Щоб не зіпсувати виконане, створимо копію робочого листа і у цій копії змінимо вихідні дані (рис. 2). Даних тепер менше, тому один рядок попередньої таблиці видаляємо. Нові значення X і Y копіюємо на місце старих значень. Усі параметри одразу перераховуються. Діаграма ж виявляється порожньою. Справа в тому, що новий графік розміщується в іншій області варіювання змінних.

Рис. 2. Нові дані

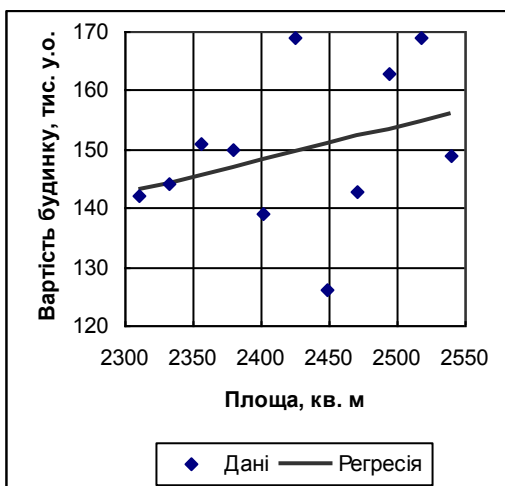


Рис. 3. Новий графік

Змінюємо формати осей. На осі X: *Минимальное значение – 2300, Максимальное значение – 2550, Цена основных делений – 50*; На осі Y: *Минимальное значение – 120, Максимальное значение – 170, Цена основных делений – 10*. З'являється новий графік (рис. 3), у якому залишилося замінити заголовки: по осі X – Площа, м², по осі Y – Вартість будинку, тис. у. о.

Контрольні запитання

1. Чому в якості міри мінливості використовується середнє *квадратичне* відхилення, а не просто середнє відхилення? Обчисліть середню величину відхилень $(Y - Y_{cp})$. Поясніть отриманий результат.

2. Дисперсія визначається як середній квадрат відхилень $S_{yy} = SSY / n$, де $SSY = \sum((Y - Y_{cp})^2)$. При виконанні цієї роботи використана інша формула: $S_{yy} = (YY)_{cp} - (Y_{cp})^2$. Доведіть еквівалентність цих двох формул.

3. У списку функцій Excel для обчислення дисперсії є функції ДИСПР і ДИСП. Функція ДИСП дає більші значення. Яка різниця між цими функціями?

4. Що таке “коваріація”, приведіть її еквівалентні розрахункові формули. Який висновок виходить з рівності коваріації нулю?

5. Регресійна модель має вигляд $Y = Y_p + E$. Покажіть, що суми квадратів і дисперсії зв'язані таким же співвідношенням: $SSY = SSR + SSE$ або (після ділення всіх членів цієї рівності на n) $S_{yy} = S_{rr} + S_{ee}$.

6. Обчисліть коефіцієнт детермінації за двома еквівалентними формулами: $R^2 = SSR / SSY = 1 - SSE / SSY$ (або $R^2 = S_{rr} / S_{yy} = 1 - S_{ee} / S_{yy}$). Що показує коефіцієнт детермінації, згідно з уведеним визначенням? Покажіть, що для нашої лінійної моделі $\sqrt{R^2} = |R_{xy}|$.

7. Згідно з МНК, параметри моделі визначаються з системи нормальних рівнянь. Зміст “системи *нормальних* рівнянь” – залишки моделі (e) повинні бути ортогональними (нормальними) до кожного члена моделі й до розрахункових значень. Переконайтеся, що $(E)_{cp} = 0$, $(EX)_{cp} = 0$, $(EY_p)_{cp} = 0$. Чому дорівнює коваріація між Y_p і E ? Дайте інтерпретацію отриманому результату.

Лабораторна робота № 2. Функції Excel для обробки даних

Мета роботи. Потрібно освоїти функції Excel для обробки даних:

- а) графічні опції *Добавить тренд*, *Показать уравнение* і *R2*;
- б) функції КОРРЕЛ, НАКЛОН, ОТРЕЗОК, ПРЕДСКАЗ;
- в) функцію діапазону ЛИНЕЙН;
- г) вивчити статистичні показники функції ЛИНЕЙН.

Для обробки одновимірних залежностей (лише одна пояснююча змінна) Excel має кілька вбудованих функцій: Функція КОРРЕЛ повертає коефіцієнт кореляції R_{xy} , функція НАКЛОН – коефіцієнт регресії b_1 , функція ОТРЕЗОК – вільний член лінійної моделі b_0 , функція ПРЕДСКАЗ обчислює розрахункові значення за лінійною моделлю $Y_p = b_0 + b_1 \cdot x$. У всіх цих функціях потрібно вказувати діапазони даних X і Y , причому для функції ПРЕДСКАЗ ці адреси повинні бути абсолютними ($\$$), щоб функція правильно працювала при копіюванні формули. Першим аргументом функції ПРЕДСКАЗ є адреса на поточне значення x (ця адреса не повинна бути абсолютною, інакше при копіюванні функція буде повертати однакове значення). Функцію ПРЕДСКАЗ можна застосовувати також як функцію діапазону, для чого необхідно попередньо виділити діапазон результату, першим аргументом функції задати не одне значення x , а весь діапазон X , введення завершити клавіатурною комбінацією **Ctrl+Shift+OK**.



У графічних опціях Майстра Діаграм є цікава опція *Добавить линию тренда*. Розглянемо її докладно. Якщо діапазони даних X , Y – суміжні і відділені від інших чисел порожніми клітинками, то досить поставити табличний курсор на будь-яку клітинку цього діапазону XY і викликати Майстер Діаграм, де треба вказати *Тип* і *Вид* діаграми (*точечная*, 1-й вид). За замовчуванням буде побудований графік розкиду точок (X , Y), що вимагає мінімуму форматування – в основному буде потрібно уточнювати лише формати осей. Клацнемо правою кнопкою миші по ряду даних на графіку і одержимо доступ до опції *Добавить линию тренда*. У діалоговому вікні *Линия тренда* (рис. 4), де пропонується широкий вибір нелінійних форм зв'язку, виберемо Тип – *Линейная*, у

вкладці **Параметры** поставимо прапорці Показывать уравнение на диаграмме і Поместить на диаграмму величину достоверности аппроксимации (R^2).

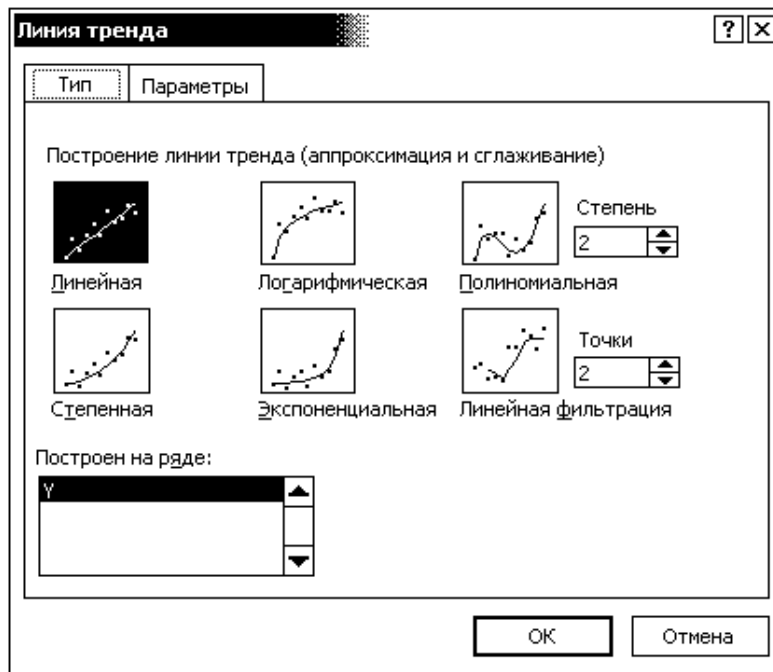


Рис. 4. Діалогове вікно Лінія тренда

На графіку з'явиться лінія тренда, рівняння лінійної регресії і коефіцієнт детермінації. Ці написи можна перенести до заголовка діаграми. Порівняємо параметри рівняння тренду з коефіцієнтами регресії, обчисленими за допомогою функцій НАКЛОН і ОТРЕЗОК. Корінь квадратний з коефіцієнта детермінації має дорівнювати абсолютній величині коефіцієнта кореляції.

На рис. 5 наведено зразок звіту до лабораторної роботи № 2.

Переходимо до функції діапазону ЛИНЕЙН. За допомогою цієї функції можна докладно аналізувати як одновимірні, так і багатовимірні залежності, що буде потрібно надалі. На жаль, інтерфейс функції залишає бажати кращого. Варто попередньо вивчити довідкові відомості і розібратися, що і на якому місці буде виведено в результаті роботи функції ЛИНЕЙН. Як для будь-якої функції діапазону тут потрібно попередньо виділити діапазон результатів, що буде завжди мати 5 рядків, а число стовпців дорівнюватиме числу параметрів моделі (два параметри b_0 й b_1). Не знімаючи виділення, набираємо знак рівності (=),

викликаємо функцію ЛИНЕЙН і заповнюємо її поля введення: *Известные значения y* – вказуємо мишкою на діапазон Y, *Известные значения x* – вказуємо на діапазон X, *Конст* – 1 (відповідає логічному значенню ІСТИНА), *Статистика* – 1 (те ж саме). Введення завершуємо клавіатурною комбінацією **Ctrl+Shift+OK**.

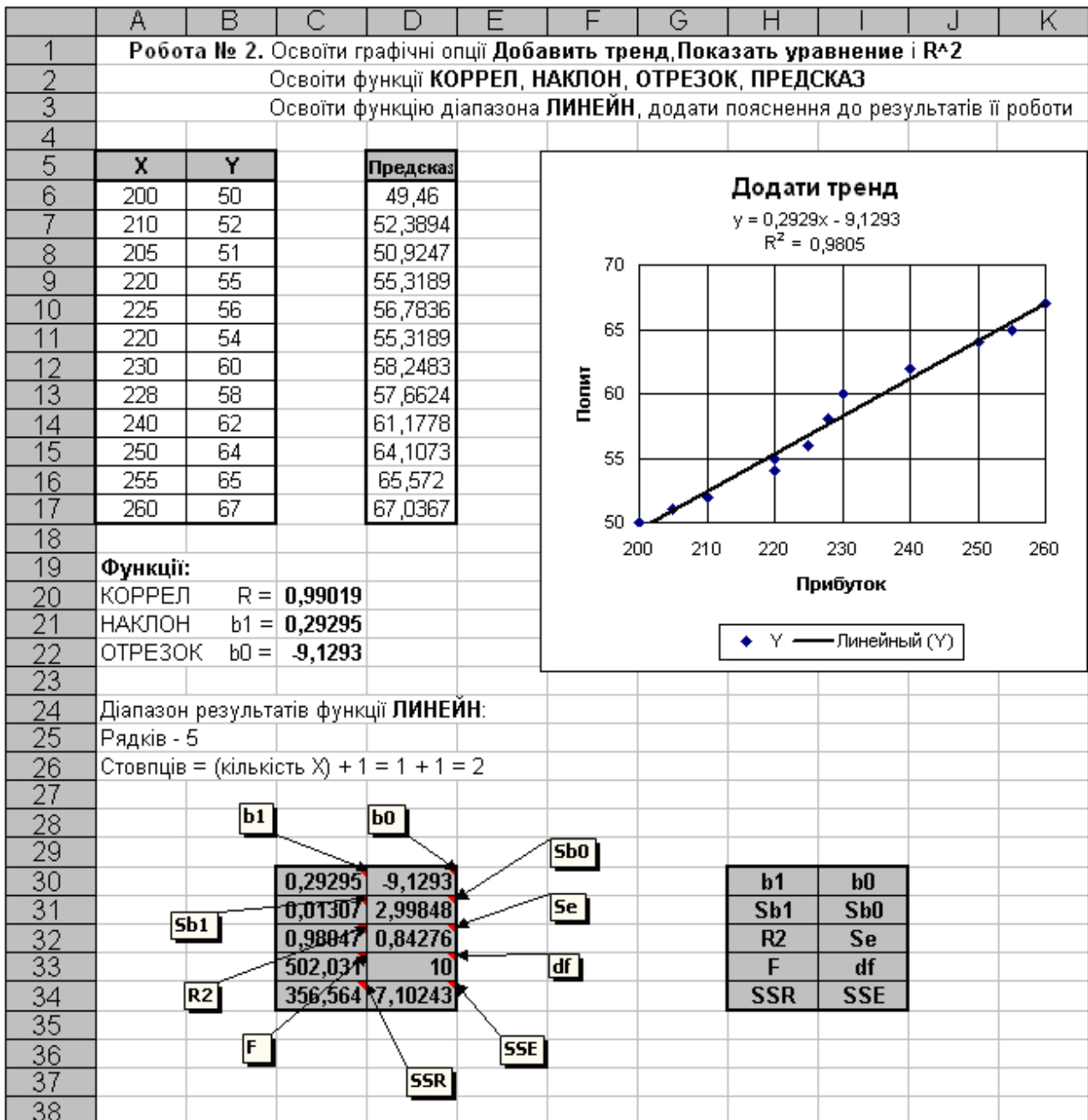


Рис. 5. Зразок звіту до лабораторної роботи № 2

У перших двох рядках діапазону результату функції ЛИНЕЙН виводяться коефіцієнти регресії і стандартні похибки цих коефіцієнтів у зворотному по-

рядку. В інших рядках перших двох стовпців приводяться наступні характеристики: коефіцієнт детермінації R^2 , стандартна похибка Se (незміщена оцінка), дисперсійне відношення Фішера F , залишкове число ступенів свободи df , суми квадратів розрахункових значень SSR і залишків моделі SSE . Рекомендується виділити діапазон результату рамкою або кольорами; поруч у діапазоні такого ж розміру записати, що виведено в кожній клітинці. Є також можливість додати примітки до кожної клітинки діапазону результату, для чого в контекстному меню слід вибрати *Добавить примечание* і занести в нього відповідне пояснення. Якщо в цьому ж меню встановити *Скрыть примечания*, то примітка буде з'являтися, коли на клітинку вказують мишкою (клітинка із приміткою відзначена червоним трикутничком у правому верхньому куті). Якщо ж у контекстному меню вибрати *Отображать примечания*, то примітка буде завжди залишатиметься на екрані (тоді варто перенести її в бажану позицію, щоб вона не перекривала інші результати розрахунків). Сховати всі примітки разом можна через меню *Сервис, Параметры, Вид, ☒ Только индикатор*; відобразити відразу всі примітки – у тому ж меню *☒ Примечание и индикатор*. Щоб примітки вивести на принтер, треба в меню *Файл, Параметры страницы, Лист* установити *Примечания: Как на листе*.

Розглянемо докладно усі числові характеристики (статистики), які обчислюються функцією ЛИНЕЙН.

SSR – сума квадратів відхилень розрахункових значень від $Y_{ср}$. Оскільки розрахункові значення Y_p вже обчислені функцією ПРЕДСКАЗ, то для перевірки SSR можна одержати як добуток $SSR = n * ДИСПР(Y_p)$. У даному прикладі $n = 12$, у загальному випадку n треба обчислювати: $n = СЧЕТ(Y_p)$.

SSE – залишкова сума квадратів, її можна обчислити за формулою:

$$SSE = СУММКВРАЗН(Y; Y_p).$$

Можна обчислити суми квадратів SSY , SSR , SSE і порівняти останні дві суми з розрахунками функції ЛИНЕЙН. Перевірте, що $SSY = SSR + SSE$.

R^2 – коефіцієнт детермінації, що показує, яка частка повної мінливості Y визначається регресійною моделлю: $R^2 = SSR / SSY$ або $R^2 = 1 - SSE / SSY$.

df – залишкове число ступенів свободи: $dfE = n - 1 - m$, де m – число пояснюючих змінних (у даному прикладі $m = 1$).

Se – незміщена оцінка стандартного відхилення: $Se = \sqrt{MSE}$.

Тут MSE (Mean Squares Error) – середній квадрат похибки – незміщена оцінка дисперсії залишків моделі: $MSE = SSE / dfE$.

Корисно заповнити таблицю дисперсійного аналізу (табл. 1).

Таблиця 1

Таблиця дисперсійного аналізу

| Джерело мінливості | | Суми квадратів | | Числа ступенів свободи | | Середні квадрати | | Дисперсійне відношення | |
|--------------------|---|----------------|---------|------------------------|----|------------------|---------|------------------------|---------|
| Регресія | R | SSR= | 356,564 | dfR= | 1 | MSR= | 356,564 | F= | 502,031 |
| Залишки | E | SSE= | 7,10243 | dfE= | 10 | MSE= | 0,71024 | | |
| Загальна | Y | SSY= | 363,667 | dfY= | 11 | | | | |

Суми квадратів SSR , SSE вже обчислені. Для чисел ступенів свободи має місце співвідношення $dfY = dfR + dfE$, аналогічне розкладу $SSY = SSR + SSE$.

Загальне ЧСС дорівнює $dfY = n - 1$ (Чому?), залишкове ЧСС обчислено за формулою $dfE = n - 1 - m$ (Чому так?), останнє ЧСС одержуємо як різницю $dfR = dfY - dfE = m$.

Середні квадрати – незміщені оцінки дисперсій – дорівнюють відношенням сум квадратів до ЧСС ($MS = SS / df$).

Дисперсійне відношення Фішера F показує, у скільки разів мінливість розрахункових значень перевищує залишкову мінливість: $F = MSR / MSE$ (порівняйте отримане значення з розрахунком функції ЛИНЕЙН).

У першому рядку діапазону результатів функції ЛИНЕЙН наведені коефіцієнти регресії b_1 і b_0 у зворотному порядку (вони також отримані функціями НАКЛОН і ОТРЕЗОК). У другому рядку наведені стандартні похибки коефіцієнтів регресії S_{b_1} і S_{b_0} (також у зворотному порядку). Формули для обчислення цих характеристик будуть наведені далі. Це дає можливість оцінити значимість окремих членів моделі за критерієм Ст'юдента $t_b = b / S_b$:

$$t_{b0} = b_0 / S_{b0} = -9,129 / 2,998 = -3,04; \quad t_{b1} = b_1 / S_{b1} = 0,293 / 0,0131 = 22,41.$$

Статистики Ст'юдента t_b показують у скільки разів коефіцієнти регресії перевищують свої стандартні відхилення.

Завдання для самостійної роботи

Розрахувати параметри лінійної однофакторної регресії за допомогою вбудованих функцій Excel. Провести дисперсійний аналіз для оцінки значущості регресійної моделі.

Для порівняння взяти той самий варіант індивідуального завдання, який був прийнятий в попередніх роботах.

Контрольні запитання

1. Сформулюйте мету і предумови дисперсійного аналізу.
2. Опишіть структуру стандартної таблиці дисперсійного аналізу.
3. Що таке "сума квадратів" (відхилень), як вона розраховується?
4. Що таке "числа ступенів свободи", як вони визначаються?
5. Що таке "середні квадрати"?
6. Що показує дисперсійне відношення Фішера?
7. Що таке "рівень значущості"? Як його визначити для отриманого дисперсійного відношення Фішера?
8. Які висновки можуть бути в результаті проведення дисперсійного аналізу?
9. Як визначається коефіцієнт детермінації, що він показує?
10. Виразіть через коефіцієнт детермінації незміщену оцінку залишкової дисперсії и дисперсійне відношення Фішера.
11. Опишіть правила роботи з функцією ЛИНЕЙН. Які характеристики розраховує ця функція і в якій послідовності розташовує їх у блоці результатів розрахунків?
12. Опишіть класи нелінійних регресійних моделей, які передбачені у графічній опції ДОБАВИТЬ ТРЕНД.
13. Як розраховувати параметри нелінійних моделей, які не передбачені у графічній опції ДОБАВИТЬ ТРЕНД?

Тема 2. Множинні регресійні моделі

Лабораторна робота № 3. Регресійний аналіз у матричній формі

Мета роботи. За даними спостережень потрібно:

а) освоїти техніку роботи з матричними функціями Excel;

б) оцінити параметри лінійної моделі за допомогою матричних перетворень $V = (X^T X)^{-1} X^T$, де Y – матриця-стовпець значень y_i , X^T означає транспонування, матриця $X = [X_0, X_1, \dots, X_m]$ складається зі стовпців значень кожного показника-аргументу X_1, X_2, \dots, X_m , перший стовець X_0 складається з одних одиниць;

в) оформити роботу у вигляді звіту, зразок якого наведено далі (рис. 6).

Теоретична частина

Позначимо матриці-стовпці значень показників і залишків моделі:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, X_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, X_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ \vdots \\ x_{1n} \end{bmatrix}, X_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ x_{23} \\ \vdots \\ x_{2n} \end{bmatrix}, \dots, X_m = \begin{bmatrix} x_{m1} \\ x_{m2} \\ x_{m3} \\ \vdots \\ x_{mn} \end{bmatrix}, E = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_n \end{bmatrix}.$$

Тоді лінійну модель для всіх спостережень можна записати у вигляді

$$Y = b_0 X_0 + b_1 X_1 + b_2 X_2 + \dots + b_m X_m + E,$$

або, у матричній формі, якщо ввести матрицю $X = [X_0, X_1, X_2, \dots, X_m]$ і вектор параметрів моделі (коефіцієнтів регресії) $V^T = (b_0, b_1, b_2, \dots, b_m)$:

$$Y = X V + E.$$

Далі операцію транспонування записуємо в один рядок без застосування верхніх індексів: $V^T = V^T$.

Обов'язкові умови ортогональності залишків до кожного члена моделі

$$\sum_{j=1}^n e_j x_{ij} = 0$$

у матричній формі можна записати як $X^T E = 0$ (перевірте це).

Помножимо матричну рівність $Y = X V + E$ ліворуч на X^T і врахуємо, що $X^T E = 0$. Одержимо таку систему рівнянь щодо параметрів:

$$(X^T X)B = X^T Y.$$

Позначимо обернену матрицю цієї системи через $C = \text{МОБР}(X^T X)$, тоді

$$B = C(X^T Y).$$

Перевірте, що добуток взаємно обернених матриць дорівнює одиничній матриці, де відмінними від нуля є лише діагональні елементи, що дорівнюють одиниці: $C(X^T X) = I$. (Оскільки через E позначено стовпець залишків моделі, для одиничній матриці використовуємо еквівалентне позначення I).

Розрахункові значення можна обчислити у вигляді матричного добутку:

$$Y_p = XB.$$

Правила роботи з функціями Excel, що повертають діапазони результатів (далі просто *функції діапазонів*). Зокрема, такими функціями є всі матричні функції, наприклад, ТРАНСП, МУМНОЖ, МОБР.

Насамперед, слід розрахувати розмір діапазону результату і виділити на робочому листі Excel блок необхідного розміру. Далі (не знімаючи виділення) треба набрати знак рівності (=), викликати функцію діапазону і заповнити її поля введення. Уведення завершити натисканням одночасно трьох клавіш **Ctrl+Shift+Enter**, або **Ctrl+Shift+OK**. Згодом можна видалити весь блок результатів, але спроби редагувати частину результатів будуть заблоковані з відповідним попередженням. Розблокувати аркуш можна декількома натисканнями клавіші **Esc**.

На рис. 6 наведений зразок виконання роботи, де за вихідні дані взяті дані попереднього прикладу (для порівняння результатів). Для розрахунків у матричній формі в таблицю доданий одиничний стовпець X_0 .

Рекомендуються усі матриці виділяти рамкою і кольорами.

Розмір матриці $X = [X_0, X_1] - 12 \times 2$ (12 рядків, 2 стовпця). Розмір транспонованої матриці X^T буде 2×12 (2 рядка, 12 стовпців). Виділяємо на вільному місці робочого аркуша блок необхідного розміру, набираємо знак рівності (=) і викликаємо функцію =ТРАНСП(*Діапазон матриці X*). Введення завершуємо одночасним натисканням клавіш **Ctrl+Shift+OK** або **Ctrl+Shift+Enter**.

| Робота № 3. Для відомих X, Y обчислити в матричній формі параметри й розрахункові значення лінійної моделі, використовуючи функції діапазонів ТРАНСП, МУМНОЖ, МОБР | | | | | | | | | | | |
|--|-----|-----|-----|-----|--|-----------|------------|----------|------------------------|----------|----------|
| Модель у матричній формі $Y=b_0 \cdot X_0 + b_1 \cdot X_1 + \dots + b_m \cdot X_m + E$ | | | | | Правила роботи з функціями діапазонів: | | | | | | |
| Матриці-стовпці: Y, Yp, E, X0=1, X1, ..., Xm | | | | | 1. Виділити діапазон результату. | | | | | | |
| Матриця X = [X0, X1, ..., Xm] | | | | | 2. Не знімаючи виділення, викликати функцію. | | | | | | |
| Транспоновані матриці: XT, BT = (b0, b1, ..., bm) | | | | | 3. Заповнити поля введення функції. | | | | | | |
| Система нормальних рівнянь (XT)Y = (XT)XB ; (XT)E = 0 | | | | | 4. Уведення завершити комбінацією Ctrl+Shift+Enter | | | | | | |
| Обернена матриця: C = МОБР(XTX) | | | | | № | X0 | X1 | Y | Yp | E | |
| Коефіцієнти регресії B = C(XTX)B = C(XT)Y | | | | | 1 | 1 | 200 | 50 | 49,46 | 0,54 | |
| Розрахункові значення: Yp=XB | | | | | 2 | 1 | 210 | 52 | 52,389 | -0,389 | |
| Залишки і суми квадратів E = Y-Yp; SSE = СУММКВ(E) | | | | | 3 | 1 | 205 | 51 | 50,925 | 0,0753 | |
| SSY = СУММКВ(Y-Уср) | | | | | 4 | 1 | 220 | 55 | 55,319 | -0,319 | |
| SSR = СУММКВ(Yp-Уср) | | | | | 5 | 1 | 225 | 56 | 56,784 | -0,784 | |
| Коефіцієнт детермінації R2 = SSR/SSY = 1 – SSE/SSY | | | | | 6 | 1 | 220 | 54 | 55,319 | -1,319 | |
| Коеф. множинної кореляції: R = КОРЕНЬ(R2) | | | | | 7 | 1 | 230 | 60 | 58,248 | 1,7517 | |
| | | | | | 8 | 1 | 228 | 58 | 57,662 | 0,3376 | |
| | | | | | 9 | 1 | 240 | 62 | 61,178 | 0,8222 | |
| | | | | | 10 | 1 | 250 | 64 | 64,107 | -0,107 | |
| | | | | | 11 | 1 | 255 | 65 | 65,572 | -0,572 | |
| | | | | | 12 | 1 | 260 | 67 | 67,037 | -0,037 | |
| | | | | | СРЗНАЧ | 1 | 228,58 | 57,833 | 57,833 | -1E-12 | |
| | | | | | ДИСПР | 0 | 346,24 | 30,306 | 29,714 | 0,5919 | |
| | | | | | SS | 0 | 4154,9 | 363,67 | 356,56 | 7,1024 | |
| XT (Функція ТРАНСП) | | | | | | | | | | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 200 | 210 | 205 | 220 | 225 | 220 | 230 | 228 | 240 | 250 | 255 | 260 |
| | | | | | XTX | | XTY | | C | | B |
| Обчислення матриць XTX, XTY | | | | | 12 | 2743 | 694 | | 12,659 | -0,055 | -9,129 |
| з використанням матриці XT | | | | | 2743 | 631159 | 159854 | | -0,055 | 0,0002 | 0,2929 |
| | | | | | Обернена матриця | | | | | | |
| Обчислення матриць XTX, XTY | | | | | 12 | 2743 | 694 | | (Функцією МОБР) | | |
| без використання матриці XT | | | | | 2743 | 631159 | 159854 | | C(XTX) = I | | |

Рис. 6. Зразок звіту до лабораторної роботи № 3

Якщо вихідним матрицям дати імена X, Y, то формула для транспонування прийме вигляд: =ТРАНСП(X).

Розмір добутку матриць (XT)X – 2×2 (число рядків дорівнює числу рядків матриці ліворуч, число стовпців дорівнює числу стовпців матриці праворуч). Виділяємо на вільному місці блок цього розміру і за допомогою функції МУМНОЖ обчислюємо добуток матриць XTX.

Розмір добутку матриць $(ХТ)Y$ – 2×1 (матриця праворуч – один стовпець Y). Обчислюємо матричний добуток $ХTY$. Обертаємо квадратну матрицю $ХТХ$ за допомогою функції МОБР. Розмір зворотної матриці $С$ – такий же (2×2). Перевіряємо, що добуток взаємно зворотних матриць $С$ і $ХТХ$ дорівнює одиничній матриці I .

Нарешті, обчислюємо стовпець параметрів B як добуток матриць $С$ і $ХTY$ (розмір 2×1). У цій лінійній моделі врахована тільки одна пояснююча змінна X_1 , тому можна порівняти результати розрахунків у матричній формі з результатами розрахунків за готовими формулами (Лабораторна робота № 1).

Розрахункові значення можна обчислити блоком у вигляді матричного добутку $Y_p = XB$, для чого в таблиці виділяємо стовпець Y_p (тільки числа, без рядка заголовка і підсумкових рядків), далі обчислюємо XB функцією МУМНОЖ.

Стовпець залишків моделі E також можна обчислити блоком, для чого в таблиці виділяємо стовпець E , набираємо знак рівності, указуємо мишкою діапазон значень Y , набираємо знак мінус, указуємо діапазон Y_p і натискаємо клавіатурну комбінацію **Ctrl+Shift+Enter**.

Якщо числовим стовпцям дати імена Y , Y_p , то формула для обчислення значень E прийме зрозумілий вигляд: $=Y-Y_p$.

У підсумкових рядках таблиці обчислені середні, дисперсії (звичайні оцінки) і суми квадратів відхилень. Звичайні оцінки дисперсії розраховуються функцією ДИСПР; суми квадратів пропорційні цим значенням $SS = \text{ДИСПР} * n$, де число спостережень (n) варто визначати функцією СЧЕТ (для гарантії правильності обчислень при заміні даних).

Помітимо, що при використанні блокових операцій для обчислення Y_p , E і матриці $ХТ$ у подальшому неможливо ні видалення зайвих, ні додавання нових рядків даних. Бажано завжди так організовувати обчислення, щоб оформлений лист Excel був шаблоном і його можна було застосовувати для наступних розрахунків з іншими даними.

Збережемо виконану роботу і зробимо копію робочого листа, у якій відмовимося від блокових операцій і обчислимо матричні добутки $ХТХ$, $ХTY$ без

попереднього формування матриці ХТ. Для цього застосовуємо вкладені функції =МУМНОЖ(ТРАНСП(Х);Х) і =МУМНОЖ(ТРАНСП(Х);У). Якщо не прийняті імена для матриць Х і У, у формулах замість них треба вказувати адреси відповідних діапазонів.

Складні формули можна набирати за допомогою Майстра Функцій. При наборі у клітинці знака рівності (=) на панелі формул активізується список вбудованих функцій. Вибираємо в цьому списку функцію МУМНОЖ і на екрані з'являється панель цієї функції (рис. 7). У поле *Массив1* треба викликати другу матричну функцію ТРАНСП.

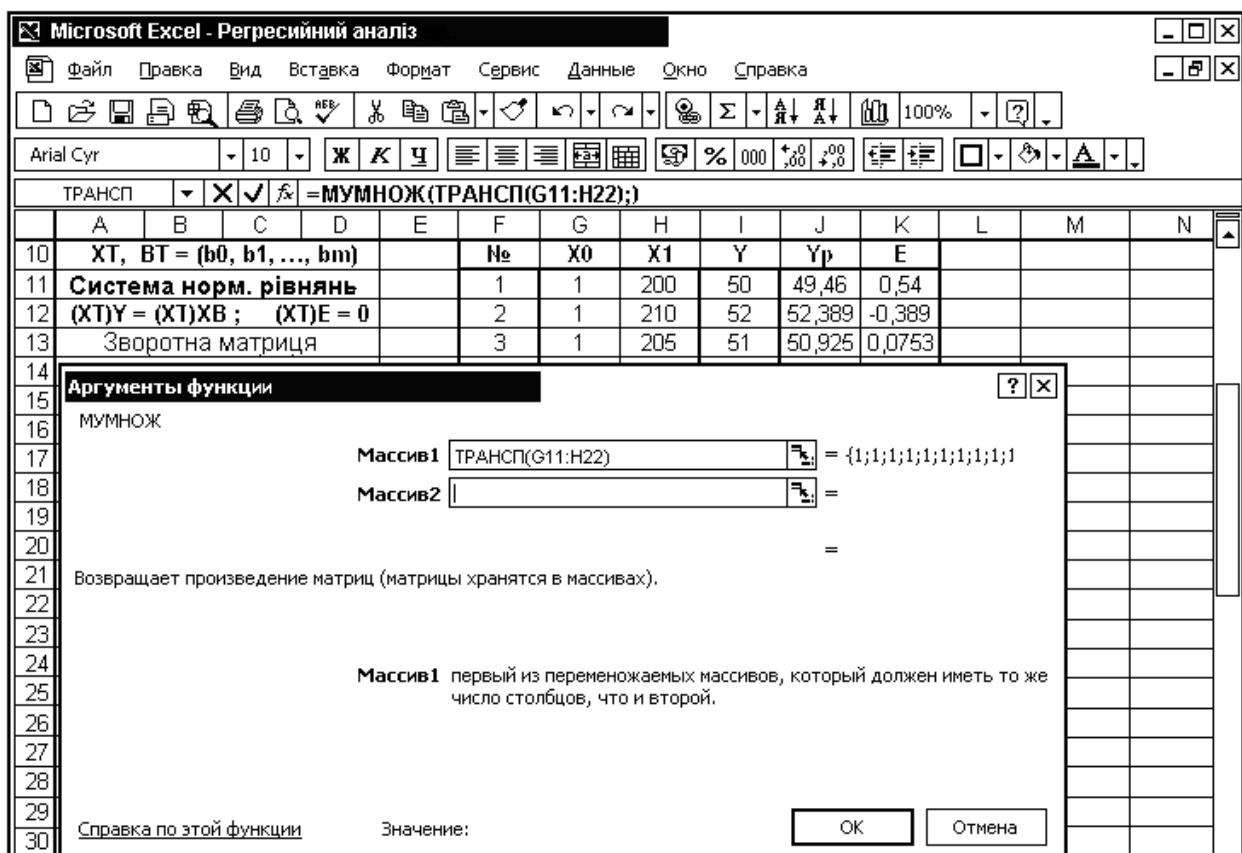


Рис. 7. Панель функції МУМНОЖ

Якщо ми бажаємо скористатися допомогою Майстра Функцій, треба натиснути на кнопку f_x , вибрати у списку нову функцію (при цьому активізується новий Майстер Функцій), заповнити її поля введення (**ОК** не натискати!) і на панелі формул, де набирається текст формули, після 1-го операнда поставити крапку з комою (;). Після цього можна заповнювати поле *Массив2* на панелі

функції МУМНОЖ. Введення треба завершити клавіатурною комбінацією **Ctrl+Shift+Enter**.



Якщо панель Майстра перекриває потрібний діапазон, треба натиснути кнопку праворуч від поля введення (із червоною стрілкою по діагоналі), панель відразу буде згорнута, залишивши тільки поле введення. Повторне натискання на цю ж кнопку відновлює панель Майстра Функцій.

Використання вкладених функцій дозволяє не формувати транспоновану матрицю ХТ, яка досить незручно розташовується на робочому листі. Відмовимося також від блокових операцій для обчислення Y_r і E . Позначимо через U – рядок матриці X (рядок аргументів x_0, x_1, \dots, x_m). Тоді розрахункові значення для кожного рядка спостережень можна обчислювати як добутки $Y_r = UB$.

Так, для першого спостереження у клітинці J11 потрібно набрати `=МУМНОЖ(G11:H11;L33:L34)`, де замість імен операндів зазначені адреси відповідних матриць. Адреси діапазону матриці B повинні бути абсолютними. Цю формулу треба скопіювати в усі клітинки діапазону Y_r . Стовпець E формуємо звичайним чином.

Якщо діапазону `L33:L34` дати ім'я B , а діапазону `$G11:$H11` – ім'я U , то формула у клітинці J11 прийме простий вигляд: `=МУМНОЖ(U;B)`. За замовчуванням при створенні нового імені використовується абсолютна адреса, але при створенні імені U слід фіксувати (\$) лише стовпці адресу діапазону рядка `$G11:$H11`.

Щоб привласнити рядку матриці X ім'я U , треба виділити 1-й рядок матриці і через меню *Вставка, Ім'я, Присвоїть* викликати панель, де в поле *Ім'я* набрати ім'я U , а в поле *Формула* поставити курсор наприкінці формули і два рази натиснути клавішу **F4** для того, щоб зняти зайві фіксатори \$.

Панель імен діапазонів наведена на рис. 8.

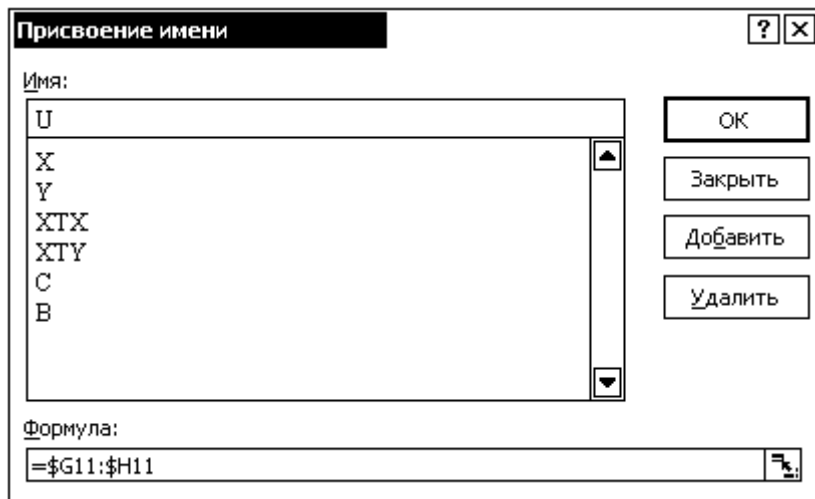


Рис. 8. Панель імен діапазонів

| № | X | Y |
|----|------|-------|
| 1 | 2310 | 142 |
| 2 | 2333 | 144 |
| 3 | 2356 | 151 |
| 4 | 2379 | 150 |
| 5 | 2402 | 139 |
| 6 | 2425 | 169 |
| 7 | 2448 | 126 |
| 8 | 2471 | 142,9 |
| 9 | 2494 | 163 |
| 10 | 2517 | 169 |
| 11 | 2540 | 149 |

Рис. 9. Нові дані

Тепер (після відмови використати матрицю ХТ і блокових операцій для обчислення Y_p , Е) з'являється можливість видаляти зайві рядки таблиці й додавати нові рядки в таблицю даних. Змінюємо вихідні дані на відомості про вартість будинку залежно від його площі (рис. 9). Нових даних менше, один рядок треба буде видалити. Усі матриці ХТХ, ХТУ, С, В, розрахункові значення Y_p і залишки моделі Е повинні перераховуватися автоматично.

Можна повернутися назад до вихідних даних. Тепер потрібно *додавати* рядок даних. Новий рядок можна додати в будь-яке місце масиву, але не перед першим і не після останнього рядка. Якщо буде потрібно перенести дані першого або останнього рядків на нове місце, робити це можна тільки копіюванням, але не вирізанням.

Обчислюємо коефіцієнт детермінації двома способами:

$$R^2 = SSR / SSY \quad \text{і} \quad R^2 = 1 - SSE / SSY.$$

Ці формули еквівалентні лише при оцінці параметрів моделі методом найменших квадратів (МНК), тому що при цьому дотримуються умови:

$$SSY = SSR + SSE,$$

$$Re, y_p = 0.$$

Корінь квадратний з коефіцієнта детермінації $R = \sqrt{R^2}$ називається коефіцієнтом множинної кореляції; він дорівнює коефіцієнту парної кореляції між експериментальними Y і розрахунковими Y_r значеннями. Для лінійної моделі з однією пояснюючою змінною X коефіцієнт множинної кореляції R дорівнює абсолютній величині коефіцієнта парної кореляції $|R_{xy}|$.

Завдання для самостійної роботи

Розрахувати параметри лінійної однофакторної регресії двома способами:

а) за відомими формулами лінійного регресійного аналізу; при розрахунках замість заповнення таблиць проміжних результатів можна використовувати функції Excel СЧЕТ, СУММ, СУММПРОИЗВ, СУММКВ, СУММКВРАЗН;

б) в матричній формі, використовуючи вкладені функції Excel без формування і запису на робочому листі транспонованої матриці ХТ.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Запишіть формули лінійного регресійного аналізу в матричній формі.
2. Як у матричній формі записати умови ортогональності залишків до кожного члена регресійної моделі?
3. Що таке "обернена матриця"?
4. Перелічіть правила роботи з "функціями діапазонів" Excel.
5. Чи можна за допомогою вкладених функцій записати кінцеві формули відразу для коефіцієнтів регресії без формування і запису на робочому листі Excel будь-яких проміжних результатів?
6. Що таке "коефіцієнт детермінації", які його властивості?
7. Коефіцієнт детермінації можна обчислити двома способами. Опишіть реалізацію кожного способу в Excel.
8. Що таке "коефіцієнт множинної кореляції"?
9. Що таке "коефіцієнт парної кореляції"? Чи можна за його допомогою оцінювати щільність однофакторного, але нелінійного зв'язку?

Лабораторна робота № 4. Багатовимірна лінійна модель

Мета роботи. За даними спостережень (X_1, X_2, X_3, Y) потрібно:

- розрахувати параметри трифакторної моделі в матричній формі;
- розрахувати ці самі параметри за допомогою функції ЛИНЕЙН;
- заповнити таблицю дисперсійного аналізу для $m \leq 3$;
- виготовити шаблон для розрахунку параметрів багатовимірної моделі.

Нижче на рис. 10 наведено зразок виконання 1-ї частини роботи.

| | A | B | C | D | E | F | G | H | I | J | K | | | | | | |
|----|--|---------------|--------|--------|------------|--|--------------------|-----------|-----------|-----------|----------|-----------|-----------|-----------|-----------|----------|-----------|
| 1 | Робота № 4. Обчислити параметри лінійної моделі в матричній формі | | | | | | | | | | | | | | | | |
| 2 | Порівняти розрахунки з результатами роботи функції ЛИНЕЙН | | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | | | |
| 4 | Модель у матричній формі $Y = b_0 \cdot X_0 + b_1 \cdot X_1 + \dots + b_m \cdot X_m + E$ $Y = XB + E$ Матриці-стовпці: $Y, Y_p, E, X_0=1, X_1, \dots, X_m$ Матриця $X = [X_0, X_1, \dots, X_m]$ Транспоновані матриці: $XT, BT = (b_0, b_1, \dots, b_m)$ Система нормальних рівнянь $(XT)Y = (XT)XB; (XT)E = 0$ Обернена матриця: $C = MOBR(XTX)$ Коефіцієнти регресії $B = C(XTX)Y = C(XT)Y$ Розрах. Значення: $Y_p = XB$ Залишки і суми квадратів $E = Y - Y_p; SSE = \sum MKB(E)$ $SSY = \sum MKB(Y - Y_{cp})$ $SSR = \sum MKB(Y_p - Y_{cp})$ Коефіцієнт детермінації $R^2 = SSR/SSY = 1 - SSE/SSY$ | | | | | Y - попит на товар X_1 - прибуток на 1 людину X_2 - ціна одиниці товару X_3 - ціна одиниці взаємозамінних товарів | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | № | X0 | X1 | X2 | X3 | Y | Yp |
| 10 | | | | | | | | | | | 1 | 1 | 200 | 22 | 2 | 50 | 49,858 |
| 11 | | | | | | | | | | | 2 | 1 | 210 | 21 | 3 | 52 | 52,123 |
| 12 | | | | | | | | | | | 3 | 1 | 205 | 21,5 | 2,5 | 51 | 50,99 |
| 13 | | | | | | | | | | | 4 | 1 | 220 | 19 | 4 | 55 | 54,862 |
| 14 | | | | | | | | | | | 5 | 1 | 225 | 18 | 5 | 56 | 56,344 |
| 15 | | | | | | | | | | | 6 | 1 | 220 | 18,5 | 4,5 | 54 | 55,212 |
| 16 | | | | | | 7 | 1 | 230 | 16 | 6 | 60 | 58,301 | | | | | |
| 17 | | | | | | 8 | 1 | 228 | 17 | 9 | 58 | 58,187 | | | | | |
| 18 | | | | | | 9 | 1 | 240 | 15 | 10 | 62 | 61,24 | | | | | |
| 19 | | | | | | 10 | 1 | 250 | 12 | 11 | 64 | 64,454 | | | | | |
| 20 | | | | | | 11 | 1 | 255 | 11 | 11 | 65 | 65,711 | | | | | |
| 21 | | | | | | 12 | 1 | 260 | 11 | 12 | 67 | 66,718 | | | | | |
| 22 | | | | | | СРЗНАЧ | 1 | 228,58 | 16,833 | 6,6667 | 57,833 | 57,833 | | | | | |
| 23 | | | | | | ДИСПР | 0 | 346,24 | 14,181 | 12,514 | 30,306 | 29,811 | | | | | |
| 24 | | | | | | ss | 0 | 4154,9 | 170,17 | 150,17 | 363,67 | 357,73 | | | | | |
| 25 | | | | | | | | | | | | | | | | | |
| 26 | | XTX | | | XTY | | C=MOBR(XTX) | | | | B | | | | | | |
| 27 | | 12 | 2743 | 202 | 80 | 694 | 1842,7 | -6,07 | -27,91 | 2,2165 | 28,558 | | | | | | |
| 28 | | 2743 | 631159 | 45339 | 19045 | 159854 | -6,07 | 0,0204 | 0,0885 | -0,013 | 0,1565 | | | | | | |
| 29 | | 202 | 45339 | 3570,5 | 1194 | 11437 | -27,91 | 0,0885 | 0,4509 | 0,0116 | -0,475 | | | | | | |
| 30 | | 80 | 19045 | 1194 | 683,5 | 4851,5 | 2,2165 | -0,013 | 0,0116 | 0,084 | 0,2247 | | | | | | |
| 31 | | | | | | | | | | | | | | | | | |
| 32 | | ЛИНЕЙН | | | | | Пояснення | | | | | | | | | | |
| 33 | | 0,2247 | -0,475 | 0,1565 | 28,558 | | b3 | b2 | b1 | b0 | | | | | | | |
| 34 | | 0,2496 | 0,5783 | 0,123 | 36,965 | | Sb3 | Sb2 | Sb1 | Sb0 | | | | | | | |
| 35 | | 0,9837 | 0,8611 | #Н/Д | #Н/Д | | R2 | Se | | | | | | | | | |
| 36 | | 120,61 | 8 | #Н/Д | #Н/Д | | F | df | | | | | | | | | |
| 37 | | 357,73 | 5,9322 | #Н/Д | #Н/Д | | SSR | SSE | | | | | | | | | |

Рис. 10. Зразок звіту до лабораторної роботи № 4

До матриці X доданий одиничний стовпець X0. Матриці ХТХ і ХТУ обчислені без проміжного формування транспонованої матриці ХТ за допомогою ланцюжка вкладених функцій $ХТХ = \text{МУМНОЖ}(\text{ТРАНСП}(Х);Х)$ і $ХТ = \text{МУМНОЖ}(\text{ТРАНСП}(Х);У)$. Розмір матриці ХТХ дорівнює $(m+1) \times (m+1) = 4 \times 4$, розмір матриці ХТ дорівнює $(m+1) \times 1 = 4 \times 1$ (тому що $m = 3$). Такі ж самі розміри мають матриці С і В.

Рекомендується всі матриці на робочому листі виділяти кольорами і рамкою. Бажано дати імена матрицям X, Y, C, B. Крім цього корисно ввести ім'я U для окремих рядків матриці X (як це зробити?).

Розрахункові значення для 1-го рядка матриці X обчислені за формулою $Ур = \text{МУМНОЖ}(U;B)$. Ця формула далі була скопійована для всіх інших спостережень. У підсумкових рядках таблиці підраховані середні, дисперсії і суми квадратів SS.

Одержимо ті самі результати за допомогою функції ЛИНЕЙН. Розмір діапазону висновків функції ЛИНЕЙН дорівнює $5 \times (m+1)$, тобто 5×4 . Заповнюємо поля функції так: *Известные_значения_y* – Y, *Известные_значения_x* – X, *Конст* – 0 (відповідає логічному значенню НЕПРАВДА – додаткової константи не потрібно, тому що в матриці X вже врахований стовпець X0), *Статистика* – 1 (статистика необхідна). Зайві клітинки в діапазоні висновку функції ЛИНЕЙН заповнені повідомленням #Н/Д (у діапазоні результатів використовуються тільки перші два рядки і перші два стовпці). Щоб повідомлення #Н/Д не відволікали увагу, рекомендується дати їм кольори фону клітинки, тоді вони не будуть видними. Рекомендується поруч із діапазоном висновків функції ЛИНЕЙН у блоці такого ж розміру записати всі пояснення.

Заповнюємо таблицю дисперсійного аналізу для $n = 12$, $m = 3$ (табл. 2).

Таблиця 2

Таблиця дисперсійного аналізу

| Мінливість | Суми квадратів | ЧСС | Середні квадрати | Дисп. відношен. |
|------------|----------------|---------|------------------|-----------------|
| Регресія R | SSR = 357,734 | dfR= 3 | MSR = 119,245 | F = 160,81 |
| Залишки E | SSE = 5,93219 | dfE= 8 | MSE = 0,74152 | |
| Загальна Y | SSY = 363,667 | dfY= 11 | | |

Число спостережень n підраховано функцією =СЧЕТ(У), а число пояснюючих змінних m – функцією =СЧЕТЗ(F9:I9)-1. Функція СЧЕТ підраховує кількість числових клітинок у діапазоні, а функція СЧЕТЗ – кількість непустих клітинок, діапазон цієї функції є текстовий діапазон заголовків матриці X .

Усі інші елементи таблиці дисперсійного аналізу розраховуються за відомими формулами, причому більша частина цих елементів правильно обчислена функцією ЛИНЕЙН за винятком статистики Фішера F . Дисперсійне відношення $F = 120,61$ неправильно обчислене і занижене в $(m+1)/m = 4/3$ рази (правильне значення $F = 160,81$). Це відбулося внаслідок того, що ми взяли за *Известные значения x* – всю матрицю X разом з одиничним стовпцем X_0 і тому задали *Конст* – 0 (додаткового вільного члена не потрібно). Слід пам'ятати, що при такому способі заповнення полів введення функції ЛИНЕЙН обчислену статистику Фішера потрібно збільшувати в $(m+1)/m$ раз. Надалі іноді будемо так заповнювати поля введення функції ЛИНЕЙН, оскільки при цьому з'являється можливість зміни кількості *пояснюючих змінних*.

Поки виготовлений шаблон допускає лише видалення зайвих і додавання нових рядків даних (але не в початок і кінець таблиці). Збережемо виконану роботу і зробимо копію робочого листа, у який внесемо деякі зміни. Чим менше на листі формується проміжних матриць, тим більш універсальним стає шаблон. Спробуйте набрати ланцюжок вкладених функцій для безпосереднього обчислення зворотної матриці S , коефіцієнтів регресії B і розрахункових значень Y_p , використовуючи лише вихідні матриці X і Y :

$$\begin{aligned} S &= \text{МОБР}(\text{МУМНОЖ}(\text{ТРАНСП}(X);X)); \\ B &= \text{МУМНОЖ}(\text{МОБР}(\text{МУМНОЖ}(\text{ТРАНСП}(X);X));\text{МУМНОЖ}(\text{ТРАНСП}(X);Y)); \\ Y_p &= \text{МУМНОЖ}(U;\text{МУМНОЖ}(\text{МОБР}(\text{МУМНОЖ}(\text{ТРАНСП}(X);X));\text{МУМНОЖ}(\text{ТРАНСП}(X);Y))). \end{aligned}$$

При наборі цих формул слід копіювати вже набрані фрагменти (вони виділені напівжирним шрифтом). Введення треба завершувати комбінацією **Ctrl+Shift+Enter**. Використання цих громіздких виражень окупається тим, що при розрахунку параметрів моделі в матричній формі з'являється можливість додавати (і видаляти) не лише рядки, але і *стовпці* матриці X . Варто мати на

увазі, що стовпці матриці X, які видаляють чи додають, не повинні перетинати діапазони матриці C і результатів функції ЛИНЕЙН, для чого перекомпонуємо розташування матеріалу на робочому листі.

На рис. 11 наведено зразок розрахунків після видалення стовпців X2 і X3.

| | | | | | |
|---|--|-----------|-----------|----------|-----------|
| Модель у матричній формі | Y - попит | | | | |
| $Y=b_0 \cdot X_0 + b_1 \cdot X_1 + \dots + b_m \cdot X_m + E$ | X1 - прибуток на 1 людину | | | | |
| $Y = XB + E$ | X2 - ціна одиниці товару | | | | |
| Матриці-стовпці: | X3 - ціна одиниці взаємозамінних товарів | | | | |
| Y, Yp, E, X0=1, X1, ..., Xm | | | | | |
| Матриця X = [X0, X1, ..., Xm] | № | X0 | X1 | Y | Yp |
| Транспоновані матриці: | 1 | 1 | 200 | 50 | 49,46 |
| $X^T, B^T = (b_0, b_1, \dots, b_m)$ | 2 | 1 | 210 | 52 | 52,3894 |
| Система нормальних рівнянь | 3 | 1 | 205 | 51 | 50,9247 |
| $(X^T)Y = (X^T)XB; (X^T)E = 0$ | 4 | 1 | 220 | 55 | 55,3189 |
| Обернена матриця: | 5 | 1 | 225 | 56 | 56,7836 |
| $C = \text{ОБРАТН}(X^T X)$ | 6 | 1 | 220 | 54 | 55,3189 |
| Коефіцієнти регресії | 7 | 1 | 230 | 60 | 58,2483 |
| $B = C(X^T X)B = C(X^T)Y$ | 8 | 1 | 228 | 58 | 57,6624 |
| Розрахунк. значення: Yp = XB | 9 | 1 | 240 | 62 | 61,1778 |
| Залишки і суми квадратів | 10 | 1 | 250 | 64 | 64,1073 |
| $E = Y - Y_p; SSE = \text{СУММКВ}(E)$ | 11 | 1 | 255 | 65 | 65,572 |
| $SSY = \text{СУММКВ}(Y - Y_{\text{ср}})$ | 12 | 1 | 260 | 67 | 67,0367 |
| $SSR = \text{СУММКВ}(Y_p - Y_{\text{ср}})$ | СРЗНАЧ | 1 | 228,583 | 57,8333 | 57,8333 |
| Коефіцієнт детермінації | ДИСПР | 0 | 346,243 | 30,3056 | 29,7137 |
| $R^2 = SSR/SSY = 1 - SSE/SSY$ | ss | 0 | 4154,92 | 363,667 | 356,564 |
| C | | | | | |
| 12,6589 | -0,055 | #Н/Д | #Н/Д | -9,1293 | |
| -0,055 | 0,00024 | #Н/Д | #Н/Д | 0,29295 | |
| #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д | |
| #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д | |
| Функція ЛИНЕЙН | | | | | |
| 0,29295 | -9,1293 | #Н/Д | #Н/Д | | |
| 0,01307 | 2,99848 | #Н/Д | #Н/Д | | |
| 0,98047 | 0,84276 | #Н/Д | #Н/Д | | |
| 251,016 | 10 | #Н/Д | #Н/Д | | |
| 356,564 | 7,10243 | #Н/Д | #Н/Д | | |

Рис. 11. Результат вилучення незначущих факторів X2 і X3

Після видалення стовпців усе автоматично перераховується (перераховується правильно, за винятком дисперсійного відношення F, що тепер занижене у 2/1=2 рази); у зайвих клітинках буде повідомлення #Н/Д.

Одержано універсальний шаблон для розрахунку параметрів багатомірної регресії до 3-го порядку включно.

Для того, щоб повернутися до вихідних даних, треба між стовпцями матриці X вставити два нових стовпці (додаються нові стовпці на робочому листі), далі треба скопіювати заздалегідь підготовлену матрицю X (разом із заголовками) на виділене для неї місце. Автоматично має правильно перераховуватися: обернена матриця C , стовпець коефіцієнтів регресії B , розрахункові значення Y_r , таблиця дисперсійного аналізу і практично всі результати функції ЛИНЕЙН (крім статистики Фішера F).

Якщо потрібно, щоб статистика Фішера завжди обчислювалася правильно, треба за *Известные значения x* функції ЛИНЕЙН брати діапазон X_1, X_2, \dots, X_m (без стовпця X_0) і задавати *Конст* – 1 (константа потрібна). Але тоді допускається лише видалення стовпців даних (іноді цього досить), а при додаванні нових стовпців у матрицю X треба буде заново заповнювати поля введення функції ЛИНЕЙН.

Блок результатів функції ЛИНЕЙН досить інформативний, сама функція універсальна, допускає розташування даних не лише за стовпцями, але і за рядками (вид розташування даних розпізнається автоматично за розмірністю діапазону *Известные значения y*).

Коефіцієнти регресії і їх стандартні відхилення виводяться функцією ЛИНЕЙН у зворотному порядку. Для задач невеликого розміру їх можна переставити вручну, але в задачах великого розміру бажане це робити автоматично, для чого можна скористатися функцією ГПР (горизонтальний перегляд).

Синтаксис цієї функції:

ГПР(*Значення*; *Таблиця*; *Номер_рядка*; *Інтервальний_перегляд*).

До блоку результатів функції ЛИНЕЙН (“Таблиця”) слід додати 1-м рядком заголовки b_3, b_2, b_1, b_0 ; “Значення” – ті самі заголовки, але в потрібному порядку. “Номер_рядка” – з якого рядка таблиці зчитуватиме дані, що відповідають заданим *Значенням*; “Інтервальний_перегляд” – 0 (потрібна точна відповідність заголовкам у таблиці).

Для роботи з таблицями крім функції ГПР в Excel є дуже схожа функція ВПР (вертикальний перегляд) з розташуванням значень показників у вигляді стовпців.

Завдання для самостійної роботи

Розрахувати параметри лінійної багатовимірної регресії у матричній формі і за допомогою функції ЛИНЕЙН. Провести дисперсійний аналіз для оцінки значущості багатовимірної регресії.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Опишіть послідовність розрахунків параметрів багатовимірної лінійної моделі в матричній формі.
2. Чи можна організувати всі розрахунки без запису проміжних результатів на робочому листі Excel? Які переваги очікуються при такій модифікації методу?
3. Опишіть таблицю дисперсійного аналізу для перевірки значущості багатовимірної регресійної моделі.
4. Що таке "квантілі розподілу Фішера", як їх знайти і де вони використовуються?
5. Які висновки можливо зробити з порівняння дисперсійного відношення з квантілями розподілу Фішера? Яка є альтернативна форма критерію Фішера?
6. Опишіть особливості роботи функції ЛИНЕЙН для аналізу статистичних характеристик в багатовимірному випадку.
7. Чи можна додати в блок висновків функції ЛИНЕЙН значення статистик Ст'юдента для коефіцієнтів регресії?
8. Як оцінити значущість окремих членів регресійної моделі?
9. Опишіть правила роботи з таблицями, зокрема правила роботи з функціями ГПР і ВПР. Ознайомтеся також з функцією ИНДЕКС.

Лабораторна робота № 5. Статистичний аналіз даних

Мета роботи. За даними спостережень (X_1, X_2, X_3, X_4, Y) потрібно:

- а) зробити розрахунок параметрів чотирифакторної моделі;
- б) обчислити розрахункові значення Y_p , варіюючи по черзі кожен пояснюючу змінну X_k при фіксованих значеннях інших аргументів;
- в) побудувати графіки розрахункових значень за кожним аргументом;
- г) перевірити стійкість обчислених оцінок пробним вибракуванням сумнівних даних (викидів або незначущих за Ст'юдентом членів);
- д) установити факт наявності або відсутності мультиколінеарності;
- е) записати рівняння регресії і дати інтерпретацію кожному параметру;
- ж) привести значення коефіцієнта детермінації і пояснити його зміст;
- з) оцінити значимість моделі в цілому за критерієм Фішера;
- і) оцінити значимість кожного члена моделі за критерієм Ст'юдента;
- к) побудувати довірчі інтервали на коефіцієнти регресії.

Зразок виконання роботи наведено на рис. 12, де позначено: Y – вартість будинку в тис. у. о., x_1 – загальна площа у m^2 , x_2 – кількість офісів, x_3 – кількість входів (піввходу 0,5 відповідає вхід лише для доставки кореспонденції), x_4 - вік (час експлуатації) будинку.

Лінійна модель може бути представлена в однієї з трьох форм:

$$\begin{aligned} Y &= b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4; \\ Y &= Y_{cp} + b_1*(x_1-x_{1cp}) + b_2*(x_2-x_{2cp}) + b_3*(x_3-x_{3cp}) + b_4*(x_4-x_{4cp}); \\ Y &= Y^\circ + b_1*(x_1-x_1^\circ) + b_2*(x_2-x_2^\circ) + b_3*(x_3-x_3^\circ) + b_4*(x_4-x_4^\circ). \end{aligned}$$

Перша форма – найпростіша, але в неї важко інтерпретувати вільний член b_0 , особливо коли він від'ємний. Звичайна інтерпретація вільного члена – значення результативної ознаки при нульових значеннях аргументів $b_0 = Y_p(0,0,0,0)$. Але як в даному прикладі розуміти вартість будинку нульової площі, без офісів і без входів? Тому самою природною є друга форма, окремі члени якої називаються компонентними ефектами. Проте для нашого прикладу і у цій формі є непорозуміння, тому що x_2 і x_3 (число офісів і число входів) не можуть бути дробовими числами ($x_{2cp} = 2,64$; $x_{3cp} = 2,36$). Третя форма прив'я-

зана до типового (еталонного) варіанта (реальний варіант, найбільш близький до середнього), тобто $x1^{\circ} = 2400$, $x2^{\circ} = 3$, $x3^{\circ} = 2$, $x4^{\circ} = 40$ (площа 2400 м², три офіси, два входи, 40 років експлуатації); вартість будинку для цього варіанта виявилася $Y^{\circ} = 151,98$ тис. у. о.

| | A | B | C | D | E | F | G | H | I | J | K |
|----|---|---------------|-------------|---------------|---|---------------|----------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1 | Робота № 5. Зробити повний аналіз лінійної залежності вартості будинку | | | | | | | | | | |
| 2 | від площі, кількості офісів, входів, а також від віку будинку | | | | | | | | | | |
| 3 | У-вартість будинку, тис. у. о. | | | | $U_p = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + b_4 \cdot x_4$ | | | | | | |
| 4 | x1 - площа, кв. м. | | | | $U_p(x_2) = U_{cp} + b_2 \cdot (x_2 - x_{2cp})$ | | | | | | |
| 5 | x2 - кількість офісів | | | | $U_p(x_2) = U_{cp} + b_2 \cdot (x_2 - x_{2cp})$ | | | | | | |
| 6 | x3 - кількість входів | | | | $U_p(x_3) = U_{cp} + b_3 \cdot (x_3 - x_{3cp})$ | | | | | | |
| 7 | x4 - вік будинку, рік | | | | $U_p(x_4) = U_{cp} + b_4 \cdot (x_4 - x_{4cp})$ | | | | | | |
| 8 | № | У | x1 | x2 | x3 | x4 | U_p | U_p(x1) | U_p(x2) | U_p(x3) | U_p(x4) |
| 9 | 1 | 142 | 2310 | 2 | 2 | 20 | 141,65 | 146,36 | 141,56 | 148,61 | 153,73 |
| 10 | 2 | 144 | 2333 | 2 | 2 | 12 | 144,16 | 146,99 | 141,56 | 148,61 | 155,61 |
| 11 | 3 | 151 | 2356 | 3 | 1,5 | 33 | 151,13 | 147,63 | 154,09 | 147,33 | 150,69 |
| 12 | 4 | 150 | 2379 | 3 | 2 | 43 | 150,7 | 148,26 | 154,09 | 148,61 | 148,34 |
| 13 | 5 | 139 | 2402 | 2 | 3 | 53 | 139,02 | 148,90 | 141,56 | 151,16 | 146,00 |
| 14 | 6 | 169 | 2425 | 4 | 2 | 23 | 169,19 | 149,54 | 166,62 | 148,61 | 153,03 |
| 15 | 7 | 126 | 2448 | 2 | 1,5 | 99 | 125,68 | 150,17 | 141,56 | 147,33 | 135,23 |
| 16 | 8 | 142,9 | 2471 | 2 | 2 | 34 | 142,82 | 150,81 | 141,56 | 148,61 | 150,45 |
| 17 | 9 | 163 | 2494 | 3 | 3 | 23 | 161,12 | 151,44 | 154,09 | 151,16 | 153,03 |
| 18 | 10 | 169 | 2517 | 4 | 4 | 55 | 169,34 | 152,08 | 166,62 | 153,71 | 145,53 |
| 19 | 11 | 149 | 2540 | 2 | 3 | 22 | 150,09 | 152,72 | 141,56 | 151,16 | 153,26 |
| 20 | Серед | 149,54 | 2425 | 2,6364 | 2,3636 | 37,909 | 149,54 | 149,54 | 149,54 | 149,54 | 149,54 |
| 21 | | Еталон | 2400 | 3 | 2 | 40 | 151,98 | | | | |
| 22 | | b0 | b1 | b2 | b3 | b4 | | | | | |
| 23 | | 52,318 | 0,0276 | 12,53 | 2,5532 | -0,234 | | | | | |
| 24 | | | | | | | | | | | |
| 25 | Функція ЛИНЕЙН | | | | | | b4 | b3 | b2 | b1 | b0 |
| 26 | tb4 | tb3 | tb2 | tb1 | tb0 | | -17,65 | 4,8113 | 31,319 | 5,0911 | 4,2753 |
| 27 | b4 | b3 | b2 | b1 | b0 | | -0,234 | 2,5532 | 12,53 | 0,0276 | 52,318 |
| 28 | Sb4 | Sb3 | Sb2 | Sb1 | Sb0 | | 0,0133 | 0,5307 | 0,4001 | 0,0054 | 12,237 |
| 29 | R2 | Se | | | | | 0,9967 | 0,9706 | #Н/Д | #Н/Д | #Н/Д |
| 30 | F | df | | | | | 459,75 | 6 | #Н/Д | #Н/Д | #Н/Д |
| 31 | SSR | SSE | | | | | 1732,4 | 5,6521 | #Н/Д | #Н/Д | #Н/Д |

Рис. 12. Зразок звіту до лабораторної роботи № 5

Розрахунок параметрів моделі та статистичних характеристик зроблений функцією ЛИНЕЙН. Діапазон результатів цієї функції доповнений двома рядками – рядком статистик Ст'юдента $tb = b / Sb$ і рядком заголовків. Ліворуч від цього блоку розміщений блок пояснень.

Під таблицею даних сформований рядок заголовків у потрібному порядку (для задачі великого розміру треба використати автозаповнення) і нижче функцією =ГПР(B\$22;\$G25:\$K\$28;3;0) знайдені значення коефіцієнтів регресії для кожного зазначеного заголовка. Тут формула наведена лише для клітинки B23, для інших клітинок рядка 23 формула копіюється.

Значення U_r для першого спостереження (у клітинці G9) обчислено за формулою =B\$23+СУММПРОИЗВ(\$C\$23:\$F\$23;\$C9:\$F9), для інших спостережень і еталонного значення у клітинці G21 результат отриманий копіюванням.

У сусідніх стовпцях обчислені розрахункові значення при варіюванні тільки однієї пояснюючої змінної при середніх значеннях інших (тобто компонентні ефекти + $U_{ср}^{\circ}$). За цими даними побудовані графіки, на яких крім розрахункових компонентних ефектів наведені вихідні дані (Y) і пунктиром додані лінії трендів за кожною змінною (рис. 13).

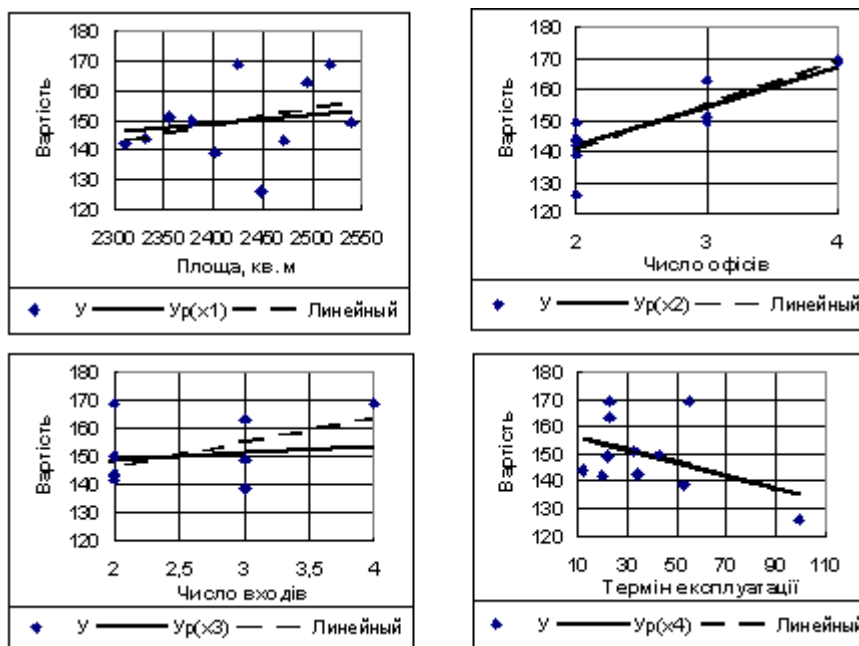


Рис. 13. Компонентні ефекти за кожним фактором

Для незалежних (некорельованих) змінних лінії тренда повинні збігатися із графіками компонентних ефектів. Тут найбільша розбіжність вийшла для аргументу x_3 (число входів). Сформуємо функцією КОРРЕЛ кореляційну матрицю між всіма змінними (рис. 14):

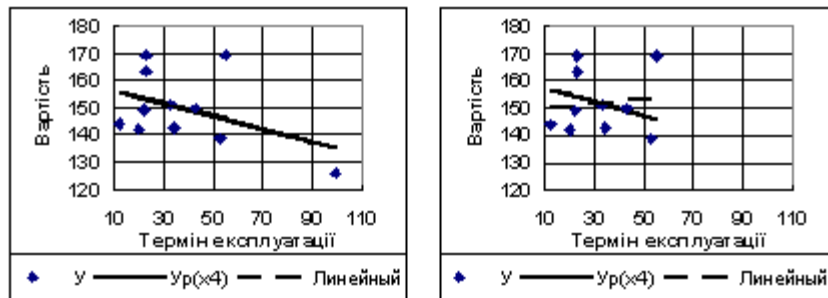
| | x1 | x2 | x3 | x4 | y |
|----|--------|--------|--------|--------|--------|
| x1 | 1 | 0,2236 | 0,6205 | 0,2214 | 0,3289 |
| x2 | 0,2236 | 1 | 0,3107 | -0,052 | 0,8742 |
| x3 | 0,6205 | 0,3107 | 1 | -0,051 | 0,5108 |
| x4 | 0,2214 | -0,052 | -0,051 | 1 | -0,448 |
| y | 0,3289 | 0,8742 | 0,5108 | -0,448 | 1 |

Рис. 14. Кореляційна матриця

Дійсно, найбільший коефіцієнт кореляції між *пояснюючими змінними* виявився для x3 і x1: $R(x1,x3) = 0,6205$. Це ще небагато. Тісних мультиколінеарних зв'язків немає.

Розкид емпіричних точок біля графіків компонентних ефектів показує вплив інших змінних. Судячи з цієї ознаки, найбільш важливою змінною є x2 (число офісів), до речі, ця змінна більш тісно пов'язана з Y: $R(y,x2) = 0,8742$).

Останній графік наводить на деякі міркування. У наших даних є лише один старий будинок – 99 років. А чи не вийде так, що після вибракування цього спостереження залежність по цьому показнику докорінно зміниться? Перенесемо цей графік ближче до таблиці і вилучимо рядок із цим спостереженням. На рис. 15 наведені порівняльні графіки компонентного ефекту x4 до та після вилучення сумнівного спостереження:



| b0 | b1 | b2 | b3 | b4 | b0 | b1 | b2 | b3 | b4 |
|--------|--------|-------|--------|--------|--------|--------|-------|--------|--------|
| 52,318 | 0,0276 | 12,53 | 2,5532 | -0,234 | 56,496 | 0,0257 | 12,64 | 2,9087 | -0,254 |

Рис. 15. Компонентний ефект фактору x4

Лінійний тренд дійсно змінив напрям на протилежне, але компонентний ефект залишився практично тим же. Для порівняння під графіками наведені коефіцієнти регресії до та після вибракування сумнівного спостереження.



Спостерігайте за зміною графіків, натискаючи комбінацію клавіш **Ctrl+Z** (скасування вибракування) і **Ctrl+Y** (повторити вибракування). Робимо висновок про відсутність викидів, що впливають на результати розрахунку.

Отримано наступне рівняння регресії:

$$U_p = 52,318 + 0,0276 * x_1 + 12,530 * x_2 + 2,553 * x_3 - 0,234 * x_4.$$

Інтерпретація його коефіцієнтів:

$b_1 = 0,0276$ показує, що кожні додаткові 100 м^2 площі приводять до збільшення вартості будинку на 2,76 тис. у.о.;

$b_2 = 12,530$ показує, що кожен додатковий офіс збільшує вартість на 12,53 тис. у. о.;

$b_3 = 2,553$ показує, що кожен додатковий вхід збільшує вартість на 2,553 тис. у. о.;

$b_4 = -0,234$ показує, що через кожні 10 років експлуатації вартість будинку знижується на 2,34 тис. у. о.

Вільний член $b_0 = 52,318$ потрібний для того, щоб вартість варіанта ($x_{1\text{ср}}$, $x_{2\text{ср}}$, $x_{3\text{ср}}$, $x_{4\text{ср}}$) дорівнювала $U_{\text{ср}}$. Для еталонного варіанта (2400 м^2 , 3 офіси, 2 входи, 40 років експлуатації) вартість будинку дорівнює 151,95 тис. у. о.:

$$U_p = 151,95 + 0,0276*(x_1-2400) + 12,530*(x_2-3) + 2,553*(x_3-2) - 0,234*(x_4-40).$$

Коефіцієнт детермінації $R^2 = 0,9967$ показує, що цією чотирифакторною моделлю пояснюється 99,67% загальної мінливості вартості, іншими словами, вартість будинку практично повністю визначається врахованими показниками.

Дисперсійне відношення Фішера $F = 459,75$ зараз обчислено правильно, тому що ми правильно заповнили поля введення функції ЛИНЕЙН. Воно показує, що мінливість розрахункових значень майже в 450 разів перевищує мінливість перешкоди (остатків моделі). Цього вже достатньо, щоб стверджувати, що регресійна модель значима в цілому. У загальному випадку варто порівнювати обчислене F з табличними значеннями F_{05} і F_{01} , які можна одержати функцією $=\text{FRASPOBR}(\text{Імовірність}; \text{ЧСС1}; \text{ЧСС2})$, де $\text{ЧСС1} = \text{dfR} = m$; $\text{ЧСС2} = \text{dfE} = n - 1 - m$; $\text{Імовірність} = 0,05$ або $0,01$. Якщо вийде, що $F < F_{05}$, зробимо висновок про незначимість моделі (“нуль-гіпотеза” не може бути відкинута). Модель вважа-

ється значимою, якщо $F > F_{01}$. Якщо $F_{05} < F < F_{01}$, то модель вважається значимою із застереженням: “модель значима при рівні значимості 5%”.

Набираючи формули $=F_{\text{РАСПОБР}}(0,05;4;6)$ і $=F_{\text{РАСПОБР}}(0,01;4;6)$, визначаємо $F_{05} = 4,53$ і $F_{01} = 9,15$. Оскільки обчислене значення $F = 459,75 > F_{01}$, робимо висновок про значимість моделі у цілому.

Для оцінки значимості можна також скористатися іншою функцією $=F_{\text{РАСП}}(X;ЧСС1;ЧСС2)$, що обчислює імовірність справедливості нуль-гіпотези при відомих значеннях F , df_R , df_E . Ця імовірність часто позначається грецькою літерою α . Якщо виявиться, що $\alpha > 0,05$, нуль-гіпотеза не може бути відкинута, модель незначима. Модель значима, якщо $\alpha < 0,01$. Функцією $=F_{\text{РАСП}}(459,75;4;6)$ знаходимо, що $\alpha = 1,372E-07 = 0,0000001372$, тобто імовірність нуль-гіпотези практично дорівнює нулю (менше 0,01). 0-гіпотеза відкидається і приймається альтернативна гіпотеза про значимість моделі у цілому.

Значимість окремих членів моделі оцінюється за критерієм Ст'юдента. Діапазон висновку функції ЛИНЕЙН доповнений нами рядком відношень $t_b = b / S_b$, які показують, у скільки разів обчислені значення параметрів перевершують свої стандартні відхилення. Ці “статистики Ст'юдента” треба порівнювати з табличними значеннями t_{05} і t_{01} , які можна одержати функцією $=СТ'ЮДРАСПОБР(\text{Імовірність};ЧСС)$. Тут $\text{Імовірність} = 0,05$; $ЧСС = df = n - 1 - m$. Якщо вийде, що $t < t_{05}$, робиться висновок про незначимість даного параметра і розглядається питання про доцільність збереження в моделі даного члена. Член зізнається значимим, якщо $t_b > t_{05}$. У нашому прикладі вийшло $t_{05}(6) = 2,447$; $t_{01}(6) = 3,707$. Оскільки найменше значення статистики Ст'юдента $t_{b0} = 4,28$ виявилось більшим ніж $t_{05} = 2,45$, робимо висновок про значимість всіх членів моделі (модель значима у цілому і складається лише зі значимих членів).

Для оцінки значимості окремих членів моделі можна скористатися також іншою функцією $=СТ'ЮДРАСП(X;ЧСС;Хвости)$, що обчислює імовірність нуль-гіпотези при даних значеннях t_b , df , $\text{Хвости}=2$ (двосторонній критерій Ст'юдента, допускаються помилки в обидва боки). Ця імовірність також позначається грецькою літерою α . Якщо виявиться, що $\alpha > 0,05$, то цей член моделі

зізнається незначущим. Функцією =СТЬЮДРАСП(4,28;6;2) знаходимо, що для найменшого $t_{b0} = 4,28$ вийшло $\text{Alpha} = 0,005 < 0,01$, отже, всі члени моделі зізнаються значимими.

За допомогою критерію Ст'юдента перевіряється справедливість нуль-гіпотези, яка затверджує, що генеральне значення параметра дорівнює нулю $\beta = M(b) = 0$, а обчислене значення (вибіркова оцінка b) є випадковим відхиленням від нуля. Для перевірки більше загального твердження, про те, що генеральне значення параметра дорівнює відомому значенню β° , варто побудувати довірчий 95%-й інтервал на математичне очікування коефіцієнта регресії $\beta = M(b)$. Якщо цей інтервал перекриє задане відоме значення β° , узагальнена "нуль-гіпотеза" не може бути відкинута.

За висловленням К. Доугерті, коефіцієнти регресії – це "інструменти економічного впливу", тому цікаво визначити найменшу і найбільшу межі варіювання цих показників, тобто межі 95%-х інтервалів.

Границі таких 95%-х інтервалів розраховуються за формулами:

$$b - t_{05} * S_b < \beta < b + t_{05} * S_b.$$

Величина $t_{05} = 2,447$ для $df_E = 6$ вже обчислена функцією СТЬЮДРАСП-ОБР, значення коефіцієнтів регресії b і їх стандартних відхилень S_b підраховані функцією ЛИНЕЙН, звідки одержуємо:

$$\begin{aligned} 52,318 - 2,447 * 12,237 < \beta_0 < 52,318 + 2,447 * 12,237, \\ 0,0276 - 2,447 * 0,0054 < \beta_1 < 0,0276 + 2,447 * 0,0054, \\ 12,530 - 2,447 * 0,4001 < \beta_2 < 12,530 + 2,447 * 0,4001, \\ 2,5532 - 2,447 * 0,5307 < \beta_3 < 2,5532 + 2,447 * 0,5307, \\ -0,234 - 2,447 * 0,0133 < \beta_4 < -0,234 + 2,447 * 0,0133, \end{aligned}$$

остаточно:

$$\begin{aligned} 22,374 < \beta_0 < 82,262, \\ 0,0144 < \beta_1 < 0,0408, \\ 11,551 < \beta_2 < 13,509, \\ 1,2545 < \beta_3 < 3,8518, \\ -0,266 < \beta_4 < -0,201. \end{aligned}$$

Жоден із цих інтервалів не перекриває 0, отже, гіпотеза про рівність нулю коефіцієнтів регресії відкидається (з рівнем довіри 95%) – всі параметри моделі значимо відрізняються від нуля.

Завдання для самостійної роботи

Розрахувати параметри лінійної чотирифакторної регресії у матричній формі і за допомогою функції ЛИНЕЙН. Побудувати графіки розрахункових значень при варіюванні кожної пояснюючої змінної окремо при середніх фіксованих рівнях решти аргументів. Перевірити щільність і значущість регресійної моделі за допомогою дисперсійного аналізу. Перевірити значущість окремих членів моделі за допомогою критерія Ст'юдента. Знайти інтервальні оцінки параметрів моделі. Дати змістовну інтерпретацію коефіцієнтам регресії.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Запишіть рівняння регресії у центрованих змінних. Які переваги цієї форми представлення моделі? Яку інтерпретацію має вільний член лінійної регресійної моделі?
2. Як перевірити значущість моделі у цілому? За яким критерієм?
3. Опишіть таблицю дисперсійного аналізу для перевірки значущості моделі. Наведіть готові формули для розрахунку дисперсійного відношення і незміщеної оцінки залишкової дисперсії через коефіцієнт детермінації.
4. Як розрахувати незміщені оцінки дисперсій параметрів моделі? Як обчислити статистики Ст'юдента для коефіцієнтів регресії? Як перевірити значущість окремих членів моделі?
5. Чи може значима у цілому модель складатися з незначимих членів?
6. Опишіть правила роботи з функцією КОРРЕЛ.
6. Що таке "мультиколінеарність"? Які її наслідки?
7. Як побудувати довірчі інтервали на коефіцієнти регресії? Які переваги мають інтервальні оцінки параметрів порівняно з точковим оцінками?
8. Яку змістовну інтерпретацію мають коефіцієнти регресії?
9. Яку роль відіграє вільний член лінійного рівняння регресії?

Лабораторна робота № 6. Мультиколінеарність

Мета роботи. На прикладі трифакторної залежності попиту на товар від особистого доходу, ціни на товар і ціни на взаємозамінні товари вивчити ефект мультиколінеарності та засвоїти стандартні дії для подолання небажаних наслідків цього ефекту.

Роботу виконуємо в наступній послідовності:

| у | x1 | x2 | x3 |
|----|-----|------|-----|
| 50 | 200 | 22 | 2 |
| 52 | 210 | 21 | 3 |
| 51 | 205 | 21,5 | 2,5 |
| 55 | 220 | 19 | 4 |
| 56 | 225 | 18 | 5 |
| 54 | 220 | 18,5 | 4,5 |
| 60 | 230 | 16 | 6 |
| 58 | 228 | 17 | 9 |
| 62 | 240 | 15 | 10 |
| 64 | 250 | 12 | 11 |
| 65 | 255 | 11 | 11 |
| 67 | 260 | 11 | 12 |

Рис. 16. Вихідні дані

а) у попередній роботі замінюємо вихідні дані (нові дані наведені на рис. 16). Видаляємо стовпець F, що містить значення змінної x_4 ; додаємо ще один рядок у середину таблиці (додається новий рядок робочого аркуша); на підготовлене місце копіюємо нові дані. Пересуваємо на одну позицію рядок заголовків b_3, b_2, b_1, b_0 перед висновком функції ЛИНЕЙН. Усе одразу перераховується, за винятком стовпця $U_r(x_4)$ – цієї змінної більше немає. Останній стовпчик висновку функції ЛИНЕЙН зайвий та заповнений повідомленнями #Н/Д;

б) змінюємо на порожніх графіках заголовки осей і для кожної осі вказуємо нові мінімальні і максимальні значення, а також нові ціни основних ділень. З'являються три графіки заданих залежностей. Четвертий графік залишається порожнім (немає змінної x_4);

в) обчислюємо кореляційну матрицю і переконуємося в існуванні тісних кореляційних зв'язків між пояснюючими змінними;

г) робимо повний статистичний аналіз залежності, інтерпретуємо оцінки параметрів, оцінюємо значимість моделі в цілому і значимість її окремих членів;

д) для демонстрації апроксимаційних властивостей багатовимірної моделі будуємо графік Спостереження – Розрахункові значення ($Y - U_r$);

е) робимо пробні вибракування незначущих членів, домагаючись мінімуму незміщеної оцінки залишкової дисперсії (MSE).

На рис. 17 наведено зразок звіту до лабораторної роботи № 6.

| | A | B | C | D | E | F | G | H | I |
|----|-----------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 7 | | | | | | | | | |
| 8 | № | Y | X1 | X2 | X3 | Ур | Ур(x1) | Ур(x2) | Ур(x3) |
| 9 | 1 | 50 | 200 | 22 | 2 | 49,858 | 53,36 | 55,38 | 56,785 |
| 10 | 2 | 52 | 210 | 21 | 3 | 52,123 | 54,925 | 55,855 | 57,009 |
| 11 | 3 | 51 | 205 | 21,5 | 2,5 | 50,99 | 54,143 | 55,617 | 56,897 |
| 12 | 4 | 55 | 220 | 19 | 4 | 54,862 | 56,49 | 56,804 | 57,234 |
| 13 | 5 | 56 | 225 | 18 | 5 | 56,344 | 57,273 | 57,279 | 57,459 |
| 14 | 6 | 54 | 220 | 18,5 | 4,5 | 55,212 | 56,49 | 57,042 | 57,346 |
| 15 | 7 | 60 | 230 | 16 | 6 | 58,301 | 58,055 | 58,229 | 57,684 |
| 16 | 8 | 58 | 228 | 17 | 9 | 58,187 | 57,742 | 57,754 | 58,358 |
| 17 | 9 | 62 | 240 | 15 | 10 | 61,24 | 59,62 | 58,704 | 58,582 |
| 18 | 10 | 64 | 250 | 12 | 11 | 64,454 | 61,185 | 60,129 | 58,807 |
| 19 | 11 | 65 | 255 | 11 | 11 | 65,711 | 61,967 | 60,603 | 58,807 |
| 20 | 12 | 67 | 260 | 11 | 12 | 66,718 | 62,75 | 60,603 | 59,032 |
| 21 | Середн. | 57,833 | 228,58 | 16,833 | 6,6667 | 57,833 | 57,833 | 57,833 | 57,833 |
| 22 | | Еталон | 230 | 17 | 7 | 58,051 | | | |
| 23 | | b0 | b1 | b2 | b3 | | | | |
| 24 | | 28,558 | 0,1565 | -0,475 | 0,2247 | | | | |
| 25 | | | | | | | | | |
| 26 | Функція ЛИНЕЙН | | | | | b3 | b2 | b1 | b0 |
| 27 | tb3 | tb2 | tb1 | tb0 | | 0,9002 | -0,821 | 1,2719 | 0,7726 |
| 28 | b3 | b2 | b1 | b0 | | 0,2247 | -0,475 | 0,1565 | 28,558 |
| 29 | Sb3 | Sb2 | Sb1 | Sb0 | | 0,2496 | 0,5783 | 0,1230 | 36,965 |
| 30 | R2 | Se | | | | 0,9837 | 0,8611 | #Н/Д | #Н/Д |
| 31 | F | df | | | | 160,81 | 8 | #Н/Д | #Н/Д |
| 32 | SSR | SSE | | | | 357,73 | 5,9322 | #Н/Д | #Н/Д |

Рис. 17. Зразок звіту до лабораторної роботи № 6

На рис. 18 наведені графіки компонентних ефектів і лінійних трендів.

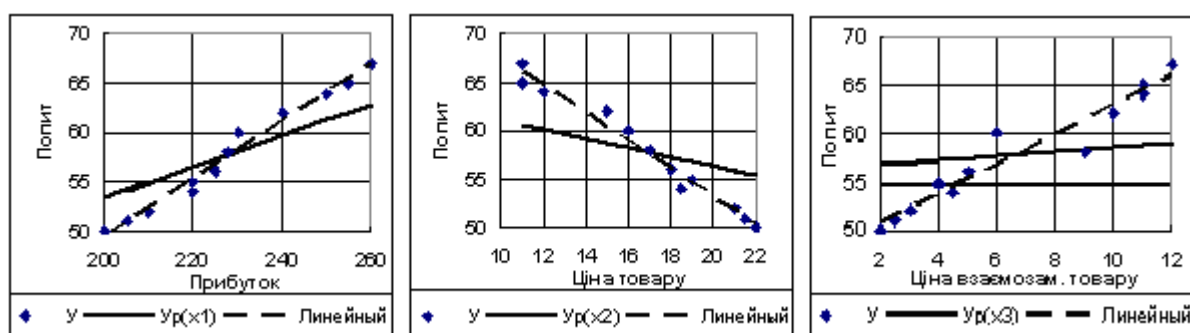


Рис 18. Компонентні ефекти і лінійні тренди

Насамперед відмітимо істотну різницю між графіками компонентних ефектів і відповідних лінійних одновимірних трендів. Це свідчить про наявність тісних кореляційних зв'язків між аргументами (пояснюючими змінними).

Дійсно, кореляційна матриця (рис. 19) показує, що між всіма змінними мають місце майже функціональні лінійні зв'язки (всі коефіцієнти кореляції за абсолютною величиною близькі до одиниці), причому між аргументами x_1 і x_2 зв'язок більш тісний, ніж з результативною ознакою Y . Детермінант такої кореляційної матриці близький до нуля, система нормальних рівнянь для визначення параметрів моделі практично вироджена, рішення такої системи вкрай відчутно до неминучих обчислювальних погрешностей.

| | x_1 | x_2 | x_3 | Y |
|-------|--------|--------|--------|--------|
| x_1 | 1 | -0,993 | 0,9594 | 0,9902 |
| x_2 | -0,993 | 1 | -0,955 | -0,988 |
| x_3 | 0,9594 | -0,955 | 1 | 0,9621 |
| Y | 0,9902 | -0,988 | 0,9621 | 1 |
| Y_p | 0 | 0 | 0 | 0,9918 |

Рис. 19. Кореляційна матриця

Проте, зараз отримане цілком розумне рівняння:

$$Y_p = 28,558 + 0,1565 * x_1 - 0,475 * x_2 + 0,2247 * x_3$$

(збільшення доходу на 100 грн призводить до збільшення попиту на 15,65 одиниць; підвищення ціни товару на 10 грн знижує попит на 4,75; підвищення ціни взаємозамінних товарів на 10 грн підвищує попит на 2,25 одиниць).

Рівняння регресії значиме в цілому за критерієм Фішера – обчислене значення $F = 160,81$ значно перевищує табличне $F_{01}(3,8) = 7,59$ (табличне значення обчислене функцією FРАСПОБР). Коефіцієнт детермінації $R^2 = 0,9837$ показує, що лінійною трифакторною моделлю пояснюється більше 98% мінливості попиту (попит практично повністю визначається врахованими показниками). Ця модель правильно відтворює спостереження. Для демонстрації якості апроксимації розглянемо графік $Y-U_p$ (рис. 20). Усі точки цього графіка дуже тісно групуються навколо діагоналі координатного кута.

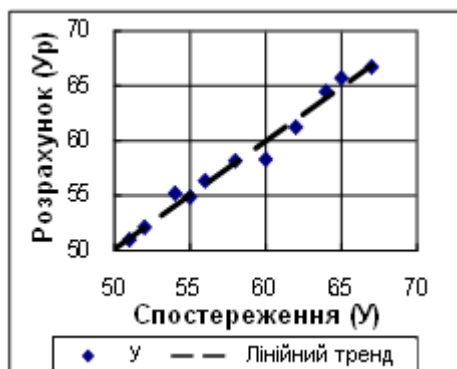


Рис. 20. Графік $Y - Y_p$

Коефіцієнт кореляції між спостережуваними і розрахунковими даними (коефіцієнт множинної кореляції) дорівнює $R = R_{y,yp} = 0,9918$.

Таким чином, отримана добра апроксимаційна модель. Але усі члени цієї моделі виявилися незначущими за критерієм Ст'юдента – навіть найбільше значення $\max(\mathbf{tb}) = \mathbf{tb}_1 = 1,2719$ виявилось менше табличного $t_{05}(8) = 2,306$ (обчислено функцією СТЬЮДРАСПОБР), а більшість коефіцієнтів регресії виявилися менше своїх стандартних похибок. Довіряти таким оцінкам не можна. Якщо мета дослідження – не просто апроксимація (опис залежності досить простим рівнянням), а оцінка параметрів моделі для подальшого їх використання (оцінка “інструментів економічного впливу”), то ця мета не досягнута.

Отже, відзначимо найбільш частий наслідок мультиколінеарності – значима у цілому модель може складатися з незначущих членів. У нашому прикладі виявилися далеко не всі неприємні наслідки мультиколінеарності – оцінки коефіцієнтів регресії могли виявитися незрозумілими, навіть мати протилежні знаки (і при цьому бути значимими).

Донедавна страховкою від мультиколінеарності вважався метод покрокової регресії (його основні різновиди – метод послідовного підключення і метод послідовного вилучення). В Excel досить просто реалізується метод послідовного вилучення. Відповідно до цього методу, треба вибракувати з моделі найбільш незначущий член. Якщо при цьому залишкова дисперсія (незміщена) зменшиться, вибракування зізнається корисним і робиться пробне вибракування наступного найбільш незначущого члена. Процес припиняється, коли в результаті чергового вибракування відбудеться збільшення незміщеної оцінки залишкової дисперсії (MSE).

У нашому прикладі найбільш незначимим є вільний член $\mathbf{tb}_0 = 0,773$, але вільний член не підлягає вибракуванню; наступний за “незначимістю” параметр \mathbf{b}_2 – для нього $\mathbf{tb}_2 = -0,812$ (тобто оцінка \mathbf{b}_2 менше своєї стандартної похибки). Виключаємо з таблиці стовпець D, що містить дані x_2 . Для того, щоб правильно відпрацювала функція ГПР, пересуваємо рядок заголовків $\mathbf{b}_1, \mathbf{b}_0$ перед блоком результатів функції ЛИНЕЙН на одну позицію ліворуч.

Усе, що потрібно, перераховується. Нові результати функції ЛИНЕЙН наведені на рис. 21.

| b3 | b1 | b0 | | |
|---------------|---------------|---------------|------|------|
| 0,9684 | 5,3698 | -0,092 | #Н/Д | #Н/Д |
| 0,2369 | 0,2497 | -0,832 | #Н/Д | #Н/Д |
| 0,2446 | 0,0465 | 9,0811 | #Н/Д | #Н/Д |
| 0,9823 | 0,8454 | #Н/Д | #Н/Д | #Н/Д |
| 249,92 | 9 | #Н/Д | #Н/Д | #Н/Д |
| 357,23 | 6,4323 | #Н/Д | #Н/Д | #Н/Д |

Рис. 21. Висновки функції ЛИНЕЙН після вилучення x_4

Значення Se зменшилося з 0,8611 до 0,8454, отже, вибракування незначущого члена було корисним. Значимість членів, що залишилися в моделі, зростає, проте ще $tb_3 = 0,968 < t_{05}$ (tb_3 навіть менше одиниці).

Робимо чергове вибракування – видаляємо стовпець, що містить дані x_3 , і пересуваємо рядок заголовків b_1, b_0 ще на одну позицію ліворуч (рис. 22):

| b1 | b0 | | | |
|---------------|---------------|------|------|------|
| 22,406 | -3,045 | #Н/Д | #Н/Д | #Н/Д |
| 0,2929 | -9,129 | #Н/Д | #Н/Д | #Н/Д |
| 0,0131 | 2,9985 | #Н/Д | #Н/Д | #Н/Д |
| 0,9805 | 0,8428 | #Н/Д | #Н/Д | #Н/Д |
| 502,03 | 10 | #Н/Д | #Н/Д | #Н/Д |
| 356,56 | 7,1024 | #Н/Д | #Н/Д | #Н/Д |

Рис. 22. Висновки функції ЛИНЕЙН після вилучення x_4 і x_3

Значення Se знову зменшилося з 0,8454 до 0,8428, отже, друге вибракування також було корисним. Члени, що залишилися в моделі, тепер значимі за критерієм Ст'юдента $tb_1=22,406$; $tb_0=-3,045$ при $t_{05}(10)=2,228$.

Ми одержали необхідні результати. Трифакторна модель пояснює 98,37% мінливості попиту, а лише одна змінна x_1 (особистий дохід) пояснює 98,05%. На частку x_2 і x_3 доводиться усього $98,37-98,05=0,32\%$ – чи варто враховувати в моделі ці члени? Згідно з “принципом економії” Л. Клейна, незначущі члени не лише марні, вони спотворюють внески інших показників і тому має бути вибраковані.

Таким чином, отримана одновимірна модель:

$$U_p = -9,129 + 0,2929 * x_1,$$

$$U_p = 57,833 + 0,2929*(x_1-228,58),$$

яка значима у цілому і складається зі значимих членів.

Завдання для самостійної роботи

Проаналізувати дані з сильною мультиколінеарністю пояснюючих змінних. Обчислити за допомогою функції КОРРЕЛ кореляційну матрицю між аргументами моделі. Розрахувати параметри лінійної багатофакторної регресії у матричній формі і за допомогою функції ЛИНЕЙН. Побудувати графіки розрахункових значень при варіюванні кожної пояснюючої змінної окремо при середніх фіксованих рвнях решти аргументів. Перевірити щільність і значущість регресійної моделі за допомогою дисперсійного аналізу. Перевірити значущість окремих членів моделі за допомогою критерія Ст'юдента. Провести процес послідовного вилучення найнезначимих членів до досягнення мінімуму незміщеної залишкової дисперсії.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Які звичайні наслідки мультиколінеарності факторів?
2. Які висновки аналізу будуть сумнівними при наявності мультиколінеарності?
3. Які є стандартні прийоми уникнення мультиколінеарності?
4. Що таке "принцип економії" Л. Клейна?
5. Опишіть методи покрокової регресії – послідовного підключення, послідовного вилучення, методу підключення-вилучення.
6. Опишіть детально реалізацію методу послідовного вилучення в Excel.
7. Який є загальноприйнятий критерій найкращої регресійної моделі?
8. Як функцією ГПР зробить вибірку статистик з блоку результатів функції ЛИНЕЙН у потрібному порядку?

Лабораторна робота № 7. Довірчі границі на лінію регресії

Мета роботи. За даними спостережень (X, Y) потрібно побудувати лінію регресії разом із 95%-и довірчими смугами на розрахункові значення Y_p і на очікуваний розкид даних навколо лінії регресії (прогнози).

Роботу виконуємо в наступному порядку:

- а) робимо розрахунок параметрів лінійної моделі в матричній формі;
- б) для кожного спостереження X обчислюємо розрахункові значення Y_p ;
- в) обчислюємо залишкову дисперсію MSE;
- г) для кожного спостереження підраховуємо дисперсію розрахункового значення S_{pp} за відомою формулою для одновимірного випадку ($m=1$);
- д) обчислюємо дисперсії S_{pp} за загальною формулою в матричній формі;
- е) розраховуємо границі 95%-х довірчих інтервалів на розрахункові значення (95%p) і на прогнози (95%q);
- ж) будуємо графік залежності разом з довірчими границями.

Теоретична частина

Приймаємо передумови регресійного аналізу (Гауса – Маркова):

- 1) $Y = X\beta + \varepsilon \rightarrow$ випадкові похибки ε відносяться лише до Y ;
- 2) $M(\varepsilon) = 0 \rightarrow$ систематичних похибок немає;
- 3) $M(\varepsilon\varepsilon^T) = \sigma^2 I \rightarrow$ спостереження незалежні $M(\varepsilon_i \varepsilon_j) = 0$ і рівноточні $M(\varepsilon_i^2) = \sigma^2$;
- 4) $\varepsilon \sim N(0; \sigma^2 I) \rightarrow$ випадкові похибки спостережень розподілені нормально.

З останньої передумови випливає, що найкращою процедурою оцінювання буде метод найменших квадратів (МНК), відповідно до якого оцінки \mathbf{b} параметрів β слід обчислювати за формулою:

$$\mathbf{b} = (X^T X)^{-1} (X^T Y) = C(X^T Y),$$

а найкращою оцінкою σ^2 буде MSE.

Перетворимо: $\mathbf{b} = C(X^T Y) = \beta + C(X^T \varepsilon)$, звідки виходить, що оцінки параметрів \mathbf{b} є випадковими величинами з математичними очікуваннями $M(\mathbf{b}) = \beta$ (тобто оцінки \mathbf{b} – незміщені) і з коваріаційною матрицею:

$$S_{bb} = M(\mathbf{b} - \beta)(\mathbf{b} - \beta)^T = M(CX^T \varepsilon)(CX^T \varepsilon)^T = CX^T M(\varepsilon \varepsilon^T) X C^T = \sigma^2 C(X^T X) C^T = \sigma^2 C.$$

Тут при перетвореннях використане правило транспонування добутку матриць $(AB)^T = B^T A^T$, третя гіпотеза Гауса – Маркова $M(\epsilon\epsilon^T) = \sigma^2 I$, співвідношення між взаємно оберненими матрицями $C(X^T X) = I$, симетрія матриці $C^T = C$.

Розрахункові значення $Y_p = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m = U\mathbf{b}$ є лінійними комбінаціями випадкових величин \mathbf{b} з відомою коваріаційною матрицею.

Оскільки дисперсія суми дорівнює сумі дисперсій плюс подвоєні коваріації, то $S_{pp} = \mathbf{M}(U\mathbf{b} - U\beta)(U\mathbf{b} - U\beta)^T = U\mathbf{S}_{bb}U^T = \sigma^2 UCU^T$; незміщену оцінку дисперсії розрахункових значень будемо позначати так само $S_{pp} = \text{MSE} \cdot UCU^T$. Тут $C = (X^T X)^{-1}$ – обернена матриця, $U = (1, x_1, x_2, \dots, x_m)$ – рядок значень пояснюючих змінних. Для одновимірної регресії ($m = 1$) формула для S_{pp} спрощується: $S_{pp} = (\text{MSE}/n) * (1 + (X - X_{cp})^2 / S_{xx})$, де $S_{xx} = \text{ДИСПР}(X)$.

Спостереження групуються навколо розрахункових значень (лінії регресії), тому дисперсія прогнозних значень дорівнює: $S_{qq} = S_{pp} + \text{MSE}$.

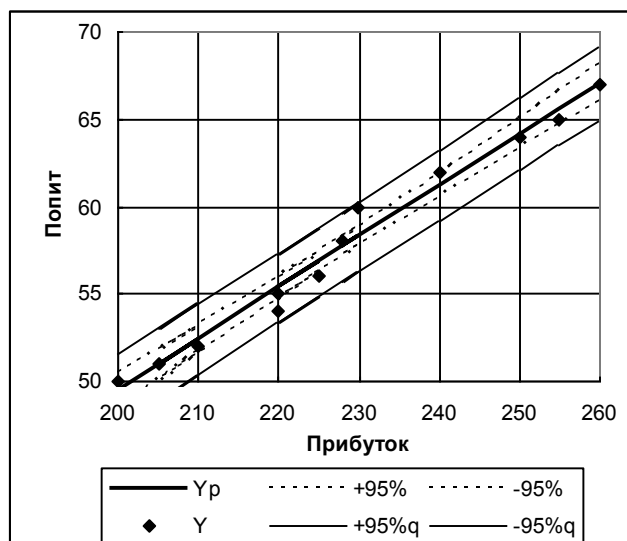
Довірчі 95%-і інтервали на розрахункові значення і на прогнози обчислюються за формулами: $Y_p(x) \pm t_{05} * S_p$, $Y_p(x) \pm t_{05} * S_q$, де S_p , S_q – корені квадратні з відповідних дисперсій (стандартні відхилення).

Зразок виконання роботи наведено на рис. 23.

Розрахунок параметрів моделі зроблений у матричній формі: $X^T X = \text{МУМНОЖ}(\text{ТРАНСП}(X); X)$, $X^T Y = \text{МУМНОЖ}(\text{ТРАНСП}(X); Y)$, $C = \text{МОБР}(X^T X)$, $B = \text{МУМНОЖ}(C; X^T Y)$, $Y_p = \text{МУМНОЖ}(U; B)$. Варто задати імена для матриць, тоді всі формули на робочому листі будуть мати саме такий вигляд (інакше треба вказувати адреси відповідних діапазонів). Середні (X_{cp} , Y_{cp}) і дисперсії (S_{xx} , S_{yy} , S_{rr}) змінних X , Y , Y_p обчислені в підсумкових рядках таблиці даних. Далі підрахована величина залишкової дисперсії $\text{MSE} = (S_{yy} - S_{rr}) * n / dfE$ (цю формулу можна розмістити на вільному місці робочого листа над таблицею даних). Для можливості заміни даних кількість спостережень визначено функцією $n = \text{СЧЕТ}(Y)$, після чого обчислено число ступенів свободи dfE і табличне значення t_{05} (функцією СТЬЮДРАСПОБР).

Робота № 7. Побудувати 95%-і довірчі границі на розрахункові значення і на прогнози.
Зробити розрахунки за загальною методикою і за формулою для одновимірної регресії.

| |
|---|
| Довірчі 95%-і інтервали на розрахункові значення і на прогнози |
| $Y_p(x) - t_{05} * S_p < Y_T < Y_p(x) + t_{05} * S_p$ |
| $Y_p(x) - t_{05} * S_q < B < Y_p(x) + t_{05} * S_q$ |
| $S_{pp} = (S_p)^2$ - дисперсія випадкової мінливості розрахункових значень |
| $S_{qq} = (S_q)^2$ - дисперсія прогнозів |
| $X = [X_0, X_1, \dots, X_m]$ - матриця, де $X_0=1, X_1, \dots, X_m$ - матриці-стовпці |
| $C = \text{МОБР}(X^T X)$ - Обернена матриця x, x_1, \dots, x_m - змінні значення |
| $U = (1, x_1, \dots, x_m)$ - матриця-рядок |
| U^T - транспонована матриця U |
| $S_{pp}(x) = \text{MSE} * (UCUT)$ |
| $S_{qq}(x) = S_{pp}(x) + \text{MSE}$ |
| Для $m=1$ (один аргумент x) |
| $S_{pp} = (\text{MSE}/n) * (1 + ((X - X_{cp})^2) / S_{xx})$ |



$$m = 1; \quad n = 12$$

$$dfE = n - 1 - m = 10$$

$$\text{MSE} = (S_{yy} - S_{rr}) * n / df = 0,7102$$

$$t_{05}(10) = 2,2281$$

| № | Прибуток | | Попит | Y | Yp | Spp | UCUT | m>1 | | | |
|--------|------------------|--------|------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| | X0 | X1 | | | | | | Spp | +95%p | -95%p | +95%q |
| 1 | 1 | 200 | 50 | 49,46 | 0,1988 | 0,28 | 0,1988 | 50,454 | 48,466 | 51,584 | 47,336 |
| 2 | 1 | 210 | 52 | 52,389 | 0,1182 | 0,1664 | 0,1182 | 53,156 | 51,623 | 54,417 | 50,361 |
| 3 | 1 | 205 | 51 | 50,925 | 0,1543 | 0,2172 | 0,1543 | 51,8 | 50,050 | 52,996 | 48,853 |
| 4 | 1 | 220 | 55 | 55,319 | 0,0718 | 0,1011 | 0,0718 | 55,916 | 54,722 | 57,289 | 53,348 |
| 5 | 1 | 225 | 56 | 56,784 | 0,0614 | 0,0864 | 0,0614 | 57,336 | 56,232 | 58,741 | 54,826 |
| 6 | 1 | 220 | 54 | 55,319 | 0,0718 | 0,1011 | 0,0718 | 55,916 | 54,722 | 57,289 | 53,348 |
| 7 | 1 | 230 | 60 | 58,248 | 0,0595 | 0,0838 | 0,0595 | 58,792 | 57,705 | 60,203 | 56,293 |
| 8 | 1 | 228 | 58 | 57,662 | 0,0592 | 0,0834 | 0,0592 | 58,205 | 57,120 | 59,617 | 55,708 |
| 9 | 1 | 240 | 62 | 61,178 | 0,0815 | 0,1147 | 0,0815 | 61,814 | 60,542 | 63,16 | 59,195 |
| 10 | 1 | 250 | 64 | 64,107 | 0,1376 | 0,1937 | 0,1376 | 64,934 | 63,281 | 66,159 | 62,056 |
| 11 | 1 | 255 | 65 | 65,572 | 0,1785 | 0,2513 | 0,1785 | 66,513 | 64,631 | 67,673 | 63,471 |
| 12 | 1 | 260 | 67 | 67,037 | 0,2279 | 0,3209 | 0,2279 | 68,1 | 65,973 | 69,195 | 64,879 |
| СРЗНАЧ | 1 | 228,58 | 57,833 | 57,833 | | | | | | | |
| ДИСПР | 0 | 346,24 | 30,306 | 29,714 | | | | | | | |
| | X ^T X | | X ^T Y | | C | | B | | | | |
| | 12 | 2743 | 694 | 12,659 | -0,055 | -9,129 | | | | | |
| | 2743 | 631159 | 159854 | -0,055 | 0,0002 | 0,2929 | | | | | |

Рис. 23. Зразок звіту до лабораторної роботи № 7

Після стовпця Y_r у таблиці обчислені дисперсії розрахункових значень S_{pp} за відомою формулою для $m = 1$ і у загальному вигляді за допомогою матричних операцій $S_{pp} = \text{MSE} * (\text{UCUT})$. Добуток трьох матриць реалізовано за допомогою вкладених функцій: $\text{UCUT} = \text{МУМНОЖ}(\text{МУМНОЖ}(U;C); \text{ТРАНСП}(U))$.

Переконаємося, що розрахунки за відомою формулою і за загальною методикою в матричній формі дають однакові результати. У наступних стовпцях таблиці обчислені довірчі границі на розрахункові значення $\pm 95\%r$ і на прогнози (на розкид точок навколо лінії регресії) $\pm 95\%q$.

За результатами розрахунку побудовано графік лінії регресії разом із границями 95%-х довірчих смуг на розрахункові значення (пунктир) і на прогнозні значення (зовнішня смуга).

Наша лінія регресії Y_r побудована за даними малої вибірки ($n = 12$). Але точно така вибірка спостережень більше ніколи не зустрінеться. Саме наявність довірчих смуг дозволяє за відносно малою вибіркою спостережень зробити висновки, що справедливі для сукупності в цілому. З гарантією 95% можна затверджувати, що графік залежності, розрахований за даними всієї сукупності (якби вони були нам відомі), буде цілком розташований в межах внутрішньої 95%-ї довірчої смуги (тобто лише в 5-и випадках із 100 буде виходити за ці границі). Наявність довірчих смуг ставить межі застосування моделі у цілому – модель корисна, поки вона описує досліджувану залежність із припустимою похибкою. Довірча смуга найбільш вузька в області варіювання вихідних даних (похибки інтерполяції невеликі), але швидко збільшується при екстраполяціях (при виході за межі області варіювання). Саме тому не можна за регресійними моделями робити довгострокові прогнози на віддалене майбутнє, оскільки швидке розширення довірчої смуги (збільшення похибок розрахункових значень) знецінює такі прогнози.

Контрольне запитання. Вільний член лінійної моделі для нашого прикладу виявився від'ємним: $b_0 = -9,1293$ (тобто попит на товар за відсутністю прибутку від'ємний). Пояснить, як це розуміти, чи немає тут будь-якого дефекта моделі? Врахуйте, що попит на товар дорівнює нулю не лише за повною від-

сутністю прибутку, але також за малим прибутком (що є меншим за прожитковий мінімум). Придивитися до графіків цієї залежності, побудованих на рис. 24 для різних діапазонів варіювання пояснюючої змінної x_1 .

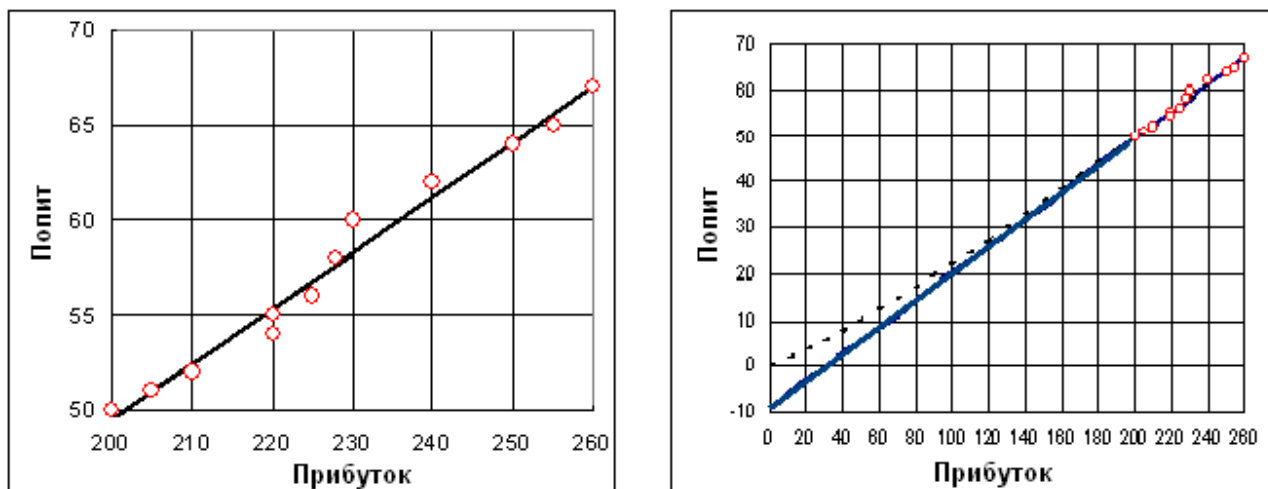


Рис. 24. Графіки залежності для різних діапазонів варіювання x_1

Які висновки можна зробити з цих графіків?

Завдання для самостійної роботи

Побудувати графік лінії регресії разом з границями 95%-ї довірчої смугою на розрахункові значення і на 95%-ю смугою на прогностні значення (на розкид даних навколо лінії регресії. Дисперсії розрахункових значень обчислити за готовою формулою і за загальною методикою.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Перелічить гіпотези Гауса – Маркова і пояснить сенс кожної з них.
2. Як за прийнятими гіпотезами визначаються випадкові похибки параметрів моделі? Як обчислюються їх дисперсії та коваріації?
3. У блоці висновків функції ЛИНЕЙН для кожного члена моделі обчислені характеристики S_b . Що це таке? Де вони використовуються?
4. Як обчислюються дисперсії розрахункових значень за лінійним рівнянням регресії?
5. Як визначаються дисперсії прогностних значень?
6. Як визначаються границі 95%-х довірчих інтервалів?

Лабораторна робота № 8. Довірчі границі в загальному випадку

Мета роботи. За даними спостережень (X_1, X_2, X_3, Y) потрібно побудувати графіки компонентних ефектів разом з 95%-и довірчими смугами на розрахункові значення Y_r і на очікуваний розкид даних навколо лінії регресії (прогнози).

Роботу виконуємо в наступному порядку:

- а) робимо розрахунок параметрів трифакторної моделі в матричній формі;
- б) для кожного спостереження X обчислюємо розрахункові значення Y_r ;
- в) обчислюємо залишкову дисперсію MSE;
- г) варіюємо кожен показник окремо при середніх значеннях інших;
- д) для кожного варіанта підраховуємо розрахункові значення і дисперсію розрахункового значення S_{rr} за загальною формулою в матричній формі;
- е) розраховуємо границі 95%-х довірчих інтервалів на розрахункові значення (95%p) і на прогнози (95%q);
- ж) будуємо графіки залежності від кожної змінної разом з довірчими границями.

Оскільки потрібно побудувати довірчі інтервали на розрахункові значення, прийдеться обчислювати елементи оберненої матриці $C = (X^T X)^{-1}$. Отже, усі розрахунки треба виконувати в матричній формі. До матриці X додаємо одиничний стовпець $X_0 = 1$, виділяємо цю матрицю рамкою і кольорами. Корисно ввести імена для матриці X і матриці-стовпця Y . Розмір матриці $X^T X$ (і зворотної матриці C) тут буде дорівнювати 4×4 , розмір матриць $X^T Y$ і B – 4×1 .

Обчислюємо матричні добутки:

$$X^T X = \text{МУМНОЖ}(\text{ТРАНСП}(X); X), \quad X^T Y = \text{МУМНОЖ}(\text{ТРАНСП}(X); Y),$$

$$C = \text{МОБР}(X^T X), \quad B = \text{МУМНОЖ}(C; X^T Y).$$

Розрахункові значення можна обчислити блоком $Y_r = \text{МУМНОЖ}(X; B)$. Ці значення потрібні для підрахунку $MSE = (S_{yy} - S_{rr}) * n / df_E$, де для даного прикладу $n = 12$, $df_E = n - 1 - m = 12 - 1 - 3 = 8$. Дисперсію вихідних даних S_{yy} і дисперсію розрахункових значень S_{rr} за всіма спостереженнями (не змішувати S_{rr}

з S_{pp}) одержуємо функцією ДИСПР. Якщо в майбутньому передбачається заміна даних, не треба обчислювати розрахункові значення Y_p блоком.

Зразок звіту до лабораторної роботи № 8 наведено на рис. 25.

| Робота № 8. Побудувати 95%-і довірчі границі на розрахункові значення і на прогнози для трифакторної лінійної моделі, де Y - попит на товар; X_1 – особистий прибуток; X_2 - ціна одиниці товару; X_3 - ціна одиниці взаємозамінних товарів | | | | | | | | | |
|---|--------|--------|---------|--------|--|--------|--------|--------|--------|
| Довірчі 95%-і інтервали на розрахункові значення і прогнози | № | X_0 | X_1 | X_2 | X_3 | Y | Y_p | | |
| | 1 | 1 | 200 | 22 | 2 | 50 | 49,858 | | |
| | 2 | 1 | 210 | 21 | 3 | 52 | 52,123 | | |
| $Y_p(x) - t_{05} * S_p < Y_t < Y_p(x) + t_{05} * S_p$ | 3 | 1 | 205 | 21,5 | 2,5 | 51 | 50,990 | | |
| $Y_p(x) - t_{05} * S_q < Y < Y_p(x) + t_{05} * S_q$ | 4 | 1 | 220 | 19 | 4 | 55 | 54,862 | | |
| | 5 | 1 | 225 | 18 | 5 | 56 | 56,344 | | |
| $S_{pp} = (S_p)^2$ - дисперсія випадкової мінливості розрахункових значень | 6 | 1 | 220 | 18,5 | 4,5 | 54 | 55,212 | | |
| $S_{qq} = (S_q)^2$ - дисперсія прогнозів | 7 | 1 | 230 | 16 | 6 | 60 | 58,301 | | |
| | 8 | 1 | 228 | 17 | 9 | 58 | 58,187 | | |
| | 9 | 1 | 240 | 15 | 10 | 62 | 61,240 | | |
| $X = [X_0, X_1, \dots, X_m]$ – матриця, де $X_0=1, X_1, \dots, X_m$ - матриці-стовпці | 10 | 1 | 250 | 12 | 11 | 64 | 64,454 | | |
| | 11 | 1 | 255 | 11 | 11 | 65 | 65,711 | | |
| $C = \text{МОБР}(X^T X)$ - Обернена матриця x_0, x_1, \dots, x_m - змінні значення | 12 | 1 | 260 | 11 | 12 | 67 | 66,718 | | |
| $U = (1, x_1, \dots, x_m)$ - матриця-рядок | СРЗНАЧ | 1 | 228,58 | 16,833 | 6,6667 | 57,833 | 57,833 | | |
| U^T - транспонована матриця U | ДИСПР | 0 | 346,24 | 14,181 | 12,514 | 30,306 | 29,811 | | |
| $S_{pp}(x) = \text{MSE} * (UCUT)$ | | | | | $m = 3; \quad n = 12$ | | | | |
| $S_{qq}(x) = S_{pp}(x) + \text{MSE}$ | | | | | $df = n - 1 - m = 8$ | | | | |
| | | | | | $\text{MSE} = (S_{yy} - S_{rr}) * n / df = 0,7415$ | | | | |
| | | | | | $t_{05}(8) = 2,306$ | | | | |
| $X^T X$ | | | $X^T Y$ | | C | | B | | |
| 12 | 2743 | 202 | 80 | 694 | 1842,7 | -6,07 | -27,91 | 2,2165 | 28,558 |
| 2743 | 631159 | 45339 | 19045 | 159854 | -6,07 | 0,0204 | 0,0885 | -0,013 | 0,1565 |
| 202 | 45339 | 3570,5 | 1194 | 11437 | -27,91 | 0,0885 | 0,4509 | 0,0116 | -0,475 |
| 80 | 19045 | 1194 | 683,5 | 4851,5 | 2,2165 | -0,013 | 0,0116 | 0,084 | 0,2247 |

Рис. 25. Зразок звіту до лабораторної роботи № 8

Створимо три таблиці, де кожна змінна варіюється окремо при середніх значеннях інших (рис. 26). При заповненні цих таблиць можна використати автозаповнення. Рядку (1, x_1, x_2, x_3) таблиці привласнимо ім'я U . Для кожного варіанта обчислюємо розрахункове значення $Y_p = UB$, матричний добуток $UCUT$, дисперсію випадкової похибки розрахункового значення $S_{pp} = \text{MSE} * (UCUT)$, границі довірчих інтервалів на розрахункові значення $Y_p \pm t_{05} * \text{КОРЕНЬ}(S_{pp})$ і на розкид даних навколо лінії регресії (прогнози) $Y_p \pm t_{05} * \text{КОРЕНЬ}(S_{pp} + \text{MSE})$. Табличне значення t_{05} варто визначити попередньо функцією СТЬЮДРАСПОБР.

| Прибуток | | | | | | | | | | |
|------------------------|--------|--------|--------|--------|---------|---------|--------|--------|--------|--------|
| X0 | X1 | X2 | X3 | Ур | UCUT | Spp | +95%p | -95%p | +95%q | -95%q |
| 1 | 200 | 16,833 | 6,6667 | 53,360 | 16,762 | 12,429 | 61,490 | 45,230 | 61,729 | 44,991 |
| 1 | 205 | 16,833 | 6,6667 | 54,143 | 11,437 | 8,4810 | 60,858 | 47,427 | 61,146 | 47,140 |
| 1 | 210 | 16,833 | 6,6667 | 54,925 | 7,1332 | 5,2895 | 60,229 | 49,622 | 60,588 | 49,262 |
| 1 | 215 | 16,833 | 6,6667 | 55,708 | 3,8499 | 2,8548 | 59,604 | 51,811 | 60,081 | 51,335 |
| 1 | 220 | 16,833 | 6,6667 | 56,490 | 1,5873 | 1,1770 | 58,992 | 53,988 | 59,684 | 53,296 |
| 1 | 225 | 16,833 | 6,6667 | 57,273 | 0,3455 | 0,2562 | 58,440 | 56,105 | 59,576 | 54,969 |
| 1 | 230 | 16,833 | 6,6667 | 58,055 | 0,1243 | 0,0922 | 58,755 | 57,355 | 60,161 | 55,949 |
| 1 | 235 | 16,833 | 6,6667 | 58,837 | 0,9239 | 0,6851 | 60,746 | 56,929 | 61,592 | 56,083 |
| 1 | 240 | 16,833 | 6,6667 | 59,620 | 2,7441 | 2,0348 | 62,909 | 56,330 | 63,462 | 55,778 |
| 1 | 245 | 16,833 | 6,6667 | 60,402 | 5,5851 | 4,1415 | 65,095 | 55,710 | 65,498 | 55,307 |
| 1 | 260 | 16,833 | 6,6667 | 62,750 | 20,232 | 15,003 | 71,682 | 53,818 | 71,900 | 53,600 |
| 1 | 270 | 16,833 | 6,6667 | 64,315 | 35,101 | 26,028 | 76,079 | 52,550 | 76,246 | 52,384 |
| Ціна | | | | | | | | | | |
| X0 | X1 | X2 | X3 | Ур | UCUT | Spp | +95%p | -95%p | +95%q | -95%q |
| 1 | 228,58 | 11 | 6,6667 | 60,603 | 15,428 | 11,4400 | 68,403 | 52,804 | 68,652 | 52,555 |
| 1 | 228,58 | 12 | 6,6667 | 60,129 | 10,618 | 7,8734 | 66,599 | 53,658 | 66,897 | 53,360 |
| 1 | 228,58 | 13 | 6,6667 | 59,654 | 6,7097 | 4,9754 | 64,797 | 54,510 | 65,167 | 54,140 |
| 1 | 228,58 | 14 | 6,6667 | 59,179 | 3,7034 | 2,7462 | 63,000 | 55,357 | 63,485 | 54,872 |
| 1 | 228,58 | 15 | 6,6667 | 58,704 | 1,5990 | 1,1857 | 61,215 | 56,193 | 61,905 | 55,503 |
| 1 | 228,58 | 16 | 6,6667 | 58,229 | 0,3965 | 0,294 | 59,479 | 56,979 | 60,576 | 55,882 |
| 1 | 228,58 | 17 | 6,6667 | 57,754 | 0,0959 | 0,0711 | 58,369 | 57,139 | 59,833 | 55,675 |
| 1 | 228,58 | 18 | 6,6667 | 57,279 | 0,6971 | 0,5169 | 58,937 | 55,621 | 59,866 | 54,692 |
| 1 | 228,58 | 19 | 6,6667 | 56,804 | 2,2003 | 1,6315 | 59,750 | 53,859 | 60,357 | 53,252 |
| 1 | 228,58 | 20 | 6,6667 | 56,330 | 4,6053 | 3,4149 | 60,591 | 52,068 | 61,031 | 51,628 |
| 1 | 228,58 | 21 | 6,6667 | 55,855 | 7,9122 | 5,8671 | 61,440 | 50,269 | 61,783 | 49,927 |
| 1 | 228,58 | 22 | 6,6667 | 55,380 | 12,1210 | 8,9880 | 62,293 | 48,466 | 62,573 | 48,187 |
| Ціна взаємозам. товару | | | | | | | | | | |
| X0 | X1 | X2 | X3 | Ур | UCUT | Srr | +95%p | -95%p | +95%q | -95%q |
| 1 | 228,58 | 16,833 | 2 | 56,785 | 1,9134 | 1,4188 | 59,532 | 54,038 | 60,174 | 53,395 |
| 1 | 228,58 | 16,833 | 3 | 57,009 | 1,2131 | 0,8996 | 59,197 | 54,822 | 59,964 | 54,055 |
| 1 | 228,58 | 16,833 | 4 | 57,234 | 0,6809 | 0,5049 | 58,873 | 55,596 | 59,809 | 54,660 |
| 1 | 228,58 | 16,833 | 5 | 57,459 | 0,3168 | 0,2349 | 58,576 | 56,341 | 59,737 | 55,180 |
| 1 | 228,58 | 16,833 | 6 | 57,684 | 0,1207 | 0,0895 | 58,373 | 56,994 | 59,786 | 55,581 |
| 1 | 228,58 | 16,833 | 7 | 57,908 | 0,0927 | 0,0687 | 58,513 | 57,304 | 59,984 | 55,833 |
| 1 | 228,58 | 16,833 | 8 | 58,133 | 0,2327 | 0,1726 | 59,091 | 57,175 | 60,338 | 55,928 |
| 1 | 228,58 | 16,833 | 9 | 58,358 | 0,5409 | 0,4011 | 59,818 | 56,897 | 60,823 | 55,893 |
| 1 | 228,58 | 16,833 | 10 | 58,582 | 1,0170 | 0,7542 | 60,585 | 56,580 | 61,403 | 55,762 |
| 1 | 228,58 | 16,833 | 11 | 58,807 | 1,6613 | 1,2319 | 61,367 | 56,248 | 62,046 | 55,568 |
| 1 | 228,58 | 16,833 | 12 | 59,032 | 2,4736 | 1,8343 | 62,155 | 55,909 | 62,733 | 55,331 |
| 1 | 228,58 | 16,833 | 13 | 59,256 | 3,4540 | 2,5613 | 62,947 | 55,566 | 63,447 | 55,066 |

Рис. 26. Розрахунки давірчих інтервалів на компонентні ефекти

За даними цих таблиць будемо графіки компонентних ефектів (залежності попиту, коли варіюється кожна пояснююча змінна по черзі при середніх значеннях інших) разом з довірчими смугами на розрахункові значення і на розкид спостережень навколо ліній регресії (рис. 27).

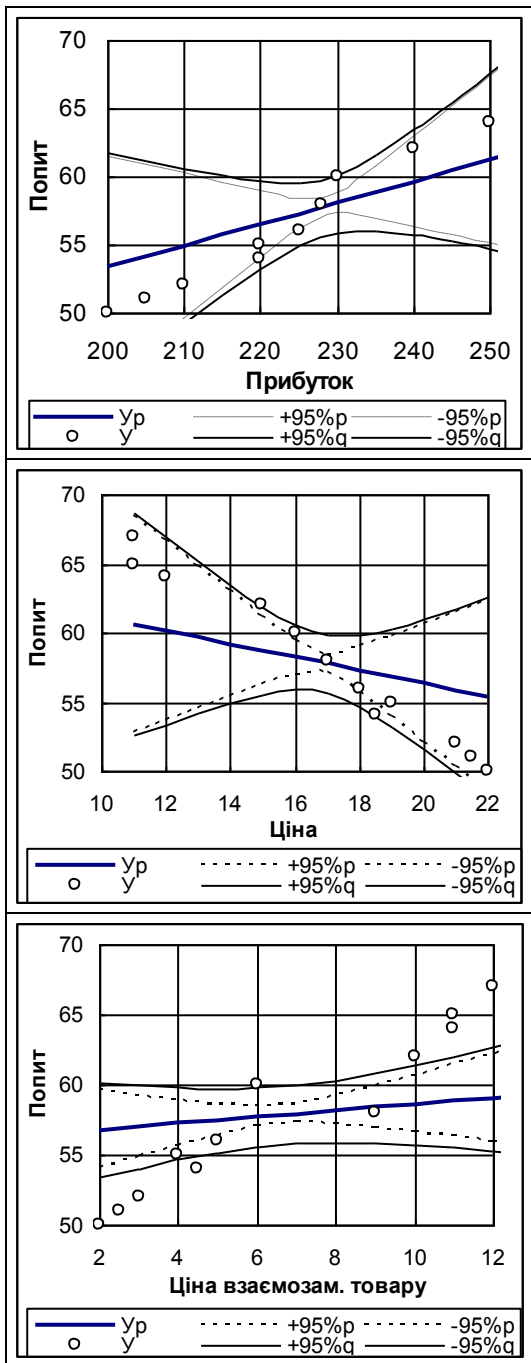


Рис. 27. Довірчі смуги на компонентні ефекти

А тепер звернемо увагу на не зовсім зрозуміле розташування емпіричних точок (Y) на кожному графіку. Оскільки $R^2 = 0,9837$ (майже 100%), то спостережувані точки не можуть мати таких великих систематичних відхилень від розрахункових значень. Очікується, що 95% всіх спостережень повинні розташовуватися у межах зовнішньої довірчої смуги, а на третьому графіку (залежність попиту від ціни на взаємозамінні товари) навіть це не дотримується.

Зверніть увагу на те, що в цьому прикладі довірчі смуги дуже широкі і в їх границях цілком розміщується горизонтальна пряма $Y = Y_{ср}$. Отже, нуль-гіпотеза про незначимість кожного компонентного ефекту не може бути відкинута. Це не суперечить, а ілюструє з іншого боку раніше зроблений висновок про незначимість кожного члена моделі за критерієм Ст'юдента. Нагадаємо, що дисперсії коефіцієнтів регресії обчислюються за формулами $S_{bb_i} = MSE * c_{ii}$, де c_{ii} – діагональні елементи оберненої матриці C . Значимість параметрів моделі оцінювалася за допомогою статистик Ст'юдента $t_{bi} = b_i / S_{bi}$, де $S_{bi} = \text{КОРЕНЬ}(S_{bb_i})$. Якщо $t_{bi} < t_{05}$, відповідний член моделі визнавався незначущим.

Обчислимо коефіцієнт детермінації за формулою $R^2 = S_{rr} / S_{yy} = 0,9837$ і переконаємось в його значимості за критерієм Фішера. Цей приклад показує наслідки мультиколінеарності – модель значима в цілому, але складається лише з незначущих членів.

Виявляється, що ми неправильно порівнювали дані. У таблиці вихідних даних варіюють всі пояснюючі змінні, а на кожному графіку компонентних ефектів змінюється лише один аргумент (решта змінних зафіксована на своїх середніх рівнях). Тому треба попередньо скоректувати вихідні дані на середні рівні зафіксованих факторів за допомогою рівняння регресії (тотожності):

$$Y = Y_{\text{ср}} + b_1 \cdot (x_1 - x_{1\text{ср}}) + b_2 \cdot (x_2 - x_{2\text{ср}}) + b_3 \cdot (x_3 - x_{3\text{ср}}) + E, \text{ де } E = Y - Y_{\text{ср}}.$$

Для першого графіка приймаємо $x_2 = x_{2\text{ср}}$, $x_3 = x_{3\text{ср}}$, звідки $Y(x_1) = Y_{\text{ср}} + b_1 \cdot (x_1 - x_{1\text{ср}}) + E$.

$$\text{Аналогічно, } Y(x_2) = Y_{\text{ср}} + b_2 \cdot (x_2 - x_{2\text{ср}}) + E, \quad Y(x_3) = Y_{\text{ср}} + b_3 \cdot (x_3 - x_{3\text{ср}}) + E.$$

Скоректовані значення $Y(x_1)$, $Y(x_2)$, $Y(x_3)$ обчислені в таблиці (рис. 28).

| № | X1 | X2 | X3 | Y | E | Y(x1) | Y(x2) | Y(x3) |
|---------------|---------------|---------------|---------------|---------------|----------|----------|----------|----------|
| 1 | 200 | 22 | 2 | 50 | 0,142 | 53,50223 | 55,52068 | 56,92639 |
| 2 | 210 | 21 | 3 | 52 | -0,123 | 54,80223 | 55,73068 | 56,88609 |
| 3 | 205 | 21,5 | 2,5 | 51 | 0,01 | 54,15273 | 55,62618 | 56,90674 |
| 4 | 220 | 19 | 4 | 55 | 0,138 | 56,62823 | 56,94168 | 57,37179 |
| 5 | 225 | 18 | 5 | 56 | -0,344 | 56,92873 | 56,93468 | 57,11449 |
| 6 | 220 | 18,5 | 4,5 | 54 | -1,212 | 55,27823 | 55,82918 | 56,13414 |
| 7 | 230 | 16 | 6 | 60 | 1,699 | 59,75423 | 59,92768 | 59,38219 |
| 8 | 228 | 17 | 9 | 58 | -0,187 | 57,55523 | 57,56668 | 58,17029 |
| 9 | 240 | 15 | 10 | 62 | 0,76 | 60,38023 | 59,46368 | 59,34199 |
| 10 | 250 | 12 | 11 | 64 | -0,454 | 60,73123 | 59,67468 | 58,35269 |
| 11 | 255 | 11 | 11 | 65 | -0,711 | 61,25673 | 59,89268 | 58,09569 |
| 12 | 260 | 11 | 12 | 67 | 0,282 | 63,03223 | 60,88568 | 59,31339 |
| СРЗНАЧ | 228,58 | 16,833 | 6,6667 | 57,833 | 0 | | | |

Рис. 28. Корегування вихідних даних на середні рівні фіксованих факторів

На графіках рис. 29 емпіричні точки замінені на скоректовані $Y(x_1)$, $Y(x_2)$, $Y(x_3)$.

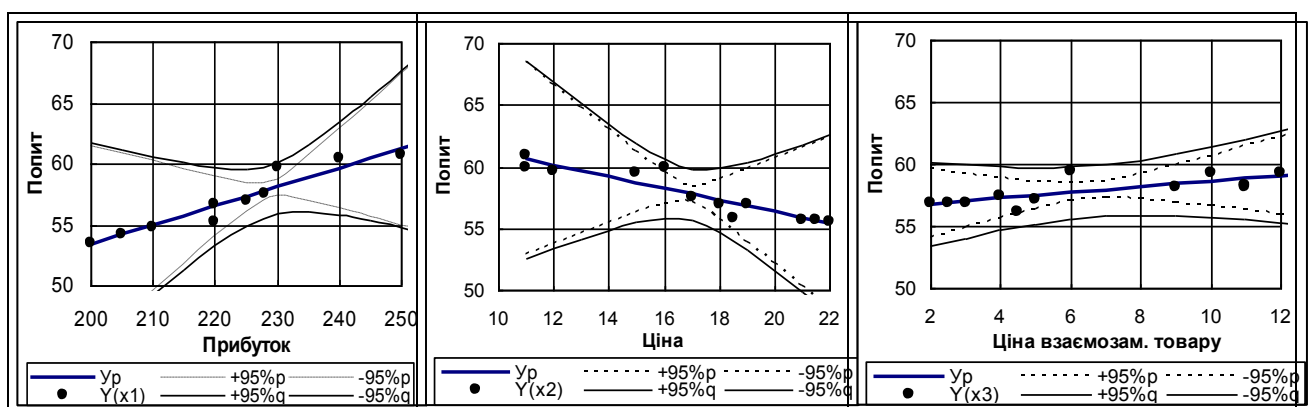


Рис. 29. Довірчі смуги і розкид даних навколо компонентних ефектів

Ось тепер на кожному графіку скоректовані дані близько групуються навколо відповідних ліній регресії (компонентних ефектів).

Завдання для самостійної роботи

Розрахувати параметри чотирифакторної регресійної моделі. Побудувати графіки компонентних ефектів разом з границями 95%-х довірчих смуг на розрахункові значення. Показати, як треба коррегувати вихідні дані на середні значення всіх аргументів крім одного.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Опишіть структуру блока висновків функції ЛИНЕЙН для багатовимірної моделі.
2. Напишіть ланцюжок функцій для розрахунку дисперсій і коваріацій коефіцієнтів регресії. Як скласти комбіновану функцію для розрахунку дисперсій і коваріацій коефіцієнтів регресії без запису проміжних результатів?
3. Що таке "компонентні ефекти"?
4. Чому емпіричні точки не дуже добре відповідають розрахунковим значенням за компонентними ефектами? Як треба скоректувати вихідні дані, щоб усунути це непорозуміння?
5. Яку інформацію додатково можна одержати з графіків довірчих смуг на лінії регресії? Чи можна вважати еквівалентною заміну графіків довірчих смуг на графіки розкиду скорегованих емпіричних точок?
6. Які наслідки сильної мультиколінеарності даних?
7. Опишіть стандартну процедуру послідовної вибраковки незначимих членів моделі.
8. Як модифікується блок висновків функції ЛИНЕЙН при вибраковках незначимих членів моделі?
9. За яким критерієм окремий член моделі визнається незначимим? Коли завершується процес послідовних вибравок?

Тема 3. Узагальнені схеми регресійного аналізу

Лабораторна робота № 9. Ридж–регресія

Мета роботи. Освоїти основний спосіб подолання сильної мультиколінеарності – метод гребінної (ридж-) регресії. Побудувати графіки гребінного сліду для коефіцієнтів регресії й коефіцієнтів детермінації. Познайомитися з особливостями ридж-оцінок.

Теоретична частина

Дотепер ми зустрічалися з добropорядною мультиколінеарністю, коли незважаючи на тісні зв'язки між пояснюючими змінними всі оцінки параметрів моделі виходили цілком правдоподібними, а мультиколінеарність проявлялася у зниженні значимості цих оцінок (значима в цілому модель могла складатися лише з незначущих за Ст'юдентом членів). Методом покрокової регресії вдавалося відкинути незначущі члени й одержати модель, що значима у цілому й складалася, в основному, зі значимих членів. Тому метод покрокової регресії вважався гарним засобом для подолання наслідків мультиколінеарності. У дійсності все виявилось значно складнішим. У розглянутому нижче прикладі мультиколінеарність привела до зовсім абсурдного результату:

$$Y_p = -253,61 - 3,027x_1 + 2,181x_2 + 1,184x_3; \quad R^2 = 0,993; \\ (tb) \quad (5,8) \quad (5,1) \quad (6,0) \quad F = 1556.$$

Тут Y – працємісткість обробки, x_1 – кількість операцій, x_2 – кількість оброблюваних поверхонь, x_3 – кількість деталей. Вийшло, що зі збільшенням кількості операцій трудомісткість обробки зменшується. Від'ємний коефіцієнт перед x_1 виявився значимим за критерієм Ст'юдента й не вибракoвується з моделі. Звичайно, ми знаємо причину цього непорозуміння – для цього досить розглянути кореляційну матрицю між усіма змінними цієї задачі (рис. 30):

| | x1 | x2 | x3 | Y |
|----|--------|--------|--------|--------|
| x1 | 1 | 0,9967 | 0,9939 | 0,9813 |
| x2 | 0,9967 | 1 | 0,9955 | 0,9901 |
| x3 | 0,9939 | 0,9955 | 1 | 0,9925 |
| Y | 0,9813 | 0,9901 | 0,9925 | 1 |

Рис. 30. Кореляційна матриця

Найбільш тісно з Y зв'язаний показник x_3 – кількість деталей. Повна три-факторна модель визначила $0,993 = 99,3\%$ загальної мінливості Y , а облік в моделі одного показника x_3 дозволяє пояснити $(0,9925)^2 = 0,985 = 98,5\%$, тобто на долю всіх інших факторів доводиться всього $0,8\%$ мінливості Y . Показник x_3 також тісно пов'язаний з іншими змінними, так що облік x_3 вичерпує мінливість усіх змінних, від яких залишаються малі випадкові залишки. Проте на наступному кроці в модель підключається x_1 (точніше малі залишки від цієї змінної) і вже виходить абсурдна модель: $Y=254,9-1,174*x_1+1,738*x_3$. Далі підключається змінна x_2 . Таким чином, метод покрокової регресії не виправдав покладених на нього сподівань, як страховки від підключення в модель тісно зв'язаних груп показників.

Найбільш радикальним засобом усунення наслідків мультиколінеарності зараз вважається метод гребінної (ридж-) регресії, у якій до діагональних елементів кореляційної матриці R_{xx} (між пояснюючими змінними) додається мале число (ридж-параметр k). Додавання цього малого параметра приводить до невеликої систематичної похибки в оцінках коефіцієнтів регресії, але суттєво знижує їх випадкову мінливість (іншими словами, оцінки стають трохи зміщеними, але більш ефективними, а ефективність оцінок означає їх більшу стабільність). Як відомо, дисперсії випадкових похибок коефіцієнтів регресії пропорційні діагональним елементам оберненої матриці $C = (R_{xx})^{-1} = \text{МОБР}(R_{xx})$. При мультиколінеарності визначник кореляційної матриці може виявитися близьким до нуля (виродження), тоді елементи оберненої матриці будуть дуже великими. Уведення навіть дуже малого ридж-параметра стримує непомірне зростання елементів оберненої матриці. На рис. 31 для нашого прикладу наведено дві обернені матриці – вихідна й з урахуванням ридж-параметра $k = 0,05$.

| k = 0 | | | k = 0,05 | | |
|--------------|---------|--------|-----------------|-------|-------|
| 157,15 | -127,65 | -29,11 | 12,28 | -6,29 | -5,66 |
| -127,65 | 215,87 | -88,05 | -6,29 | 12,65 | -6,04 |
| -29,11 | -88,05 | 117,59 | -5,66 | -6,04 | 12,03 |

Рис. 31. Обернені матриці

Як бачимо, введення малого ридж-параметра $k = 0,05$ знизило елементи оберненої матриці більш ніж в 10 разів. Множник пропорційності MSE при цьому також змінився, але не так суттєво.

Всі розрахунки робимо у стандартизованих змінних:

$$y = \frac{Y - \bar{Y}}{s_Y}; \quad x_j = \frac{X_j - \bar{X}_j}{s_{X_j}}.$$

Система нормальних рівнянь для цих змінних прийме вигляд:

$$\mathbf{Rxx} \beta = \mathbf{Ryx},$$

де \mathbf{Rxx} – кореляційна матриця між пояснюючими змінними, \mathbf{Ryx} – стовпець коефіцієнтів кореляції з Y , β – стовпець стандартизованих коефіцієнтів регресії. Зворотний перехід від стандартизованих до нестандартизованих коефіцієнтів регресії робиться за формулами:

$$b_i = \beta_i \cdot \frac{s_Y}{s_{X_i}};$$

$$b_0 = \bar{Y} - \sum_{i=1}^m \beta_i \bar{X}_i.$$

Для коефіцієнта детермінації маємо просту формулу:

$$R^2 = \sum_{i=1}^m \beta_i \cdot r_{yx_i}$$

Статистики Ст'юдента для оцінки значимості коефіцієнтів регресії знаходимо за формулами:

$$t_{\beta_i} = \frac{\beta_i}{\sqrt{c_{ii} \cdot \frac{1-R^2}{dfE}}},$$

де c_{ii} – діагональні елементи матриці C , що обернена до кореляційної $C = R_{xx}^{-1}$.

При введенні ридж-параметра k з'являється кореляція розрахункових значень Y_r із залишками моделі $Re, y_r \neq 0$, порушується умова $S_{yy} = S_{rr} + S_{ee}$, тому при обчисленні коефіцієнта детермінації раніше еквівалентні формули $RR = S_{rr}/S_{yy}$ і $R^2 = 1 - S_{ee}/S_{yy}$ тепер будуть давати різні значення.

На рис. 32 наведено зразок виконання роботи.

| Робота № 9. Обчислити коефіцієнти гребінної регресії з ридж-параметром k | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|-------------|-----------|-----------|-----------|---|----------------------|----------|--------|---|--------|---------|--------|---------|--------|--------|--------|--------|--------|------|--------|--------|--------|------|--------|--------|--------|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|-----------|-----------|-----------|----------------------|-----------|--------------|
| X1 – кількість операцій; X2 – кількість оброблюємих поверхонь; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| X3 – кількість деталей; Y – працємісткість, нормо-години. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Система нормальних рівнянь в стандартизованих змінних з урахуванням ридж-параметра k : $(R_{xx}+k*I) \beta = R_{yx}$ R_{xx} – кореляційна матриця R_{yx} – коеф. кореляції з Y I – одинична матриця Обернена матриця C_k = МОБР(R_{xx}+k*I) Коефіцієнти регресії: $\beta = C_k * R_{yx}$ $b_i = \beta_i * S_y / S_{x_i}$ $b_0 = Y_{cp} - \sum(b_i * X_{i,cp})$ Коефіцієнти детермінації $RR = S_{rr} / S_{yy}$; $R^2 = 1 - S_{ee} / S_{yy}$ $dfE = n - m - 1$ t - статистики для b_i : $t_{b_i} = \beta_i / \text{КОРЕНЬ}(c_{ii} * (1 - R^2) / dfE)$ c_{ii} - елементи матриці C _k | № | X1 | X2 | X3 | Y | Y_p | e | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 1 | 239 | 449 | 52 | 72,4 | 40,86 | 31,54 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 2 | 240 | 450 | 50 | 72,7 | 41,14 | 31,56 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 3 | 251 | 465 | 53 | 73,4 | 59,86 | 13,54 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 4 | 256 | 470 | 65 | 75,5 | 72,50 | 3,00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 5 | 258 | 475 | 67 | 78,5 | 78,27 | 0,23 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 6 | 263 | 478 | 74 | 88,0 | 86,93 | 1,07 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 7 | 264 | 479 | 80 | 91,5 | 91,30 | 0,20 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 8 | 267 | 482 | 85 | 93,0 | 97,76 | -4,76 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 9 | 270 | 485 | 89 | 95,5 | 103,71 | -8,21 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 10 | 272 | 487 | 97 | 98,8 | 110,40 | -11,60 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 11 | 274 | 492 | 99 | 105,8 | 116,16 | -10,36 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 12 | 275 | 493 | 108 | 107,2 | 122,06 | -14,86 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 13 | 276 | 494 | 111 | 107,9 | 124,90 | -17,00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 14 | 278 | 498 | 117 | 114,8 | 131,99 | -17,19 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 15 | 283 | 502 | 128 | 122,7 | 143,41 | -20,71 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 16 | 286 | 504 | 134 | 134,6 | 149,67 | -15,07 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 17 | 290 | 509 | 137 | 147,5 | 157,12 | -9,62 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 18 | 291 | 512 | 144 | 150,3 | 163,43 | -13,13 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 19 | 293 | 520 | 153 | 170,7 | 174,90 | -4,20 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 20 | 295 | 525 | 168 | 192,1 | 187,31 | 4,79 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 21 | 298 | 528 | 172 | 194,6 | 193,26 | 1,34 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 22 | 300 | 532 | 173 | 198,8 | 197,80 | 1,00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 23 | 302 | 537 | 175 | 202,7 | 203,57 | -0,87 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 24 | 304 | 539 | 178 | 206,1 | 207,70 | -1,60 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 25 | 305 | 541 | 181 | 208,8 | 211,25 | -2,45 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 26 | 308 | 546 | 185 | 216,6 | 218,63 | -2,03 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 27 | 311 | 549 | 189 | 220,7 | 224,58 | -3,88 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 28 | 318 | 554 | 193 | 229,0 | 234,32 | -5,32 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 29 | 319 | 557 | 205 | 239,0 | 243,17 | -4,17 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 30 | 322 | 560 | 215 | 254,6 | 252,18 | 2,42 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 31 | 325 | 564 | 224 | 267,9 | 261,40 | 6,50 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 32 | 327 | 569 | 231 | 275,0 | 269,72 | 5,28 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 33 | 328 | 573 | 238 | 275,4 | 276,74 | -1,34 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 34 | 332 | 574 | 244 | 290,5 | 282,87 | 7,63 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 35 | 335 | 580 | 248 | 295,7 | 290,96 | 4,74 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 36 | 338 | 584 | 256 | 311,2 | 299,67 | 11,53 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 37 | 340 | 590 | 265 | 322,9 | 309,72 | 13,18 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 38 | 344 | 598 | 274 | 336,5 | 322,38 | 14,12 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 39 | 346 | 605 | 281 | 346,8 | 332,12 | 14,68 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Средн. | 295,46 | 524,33 | 157,38 | 181,68 | 181,68 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ДИСПР | 872,92 | 1803,40 | 4644,39 | 7163,32 | 6785,87 | 134,35 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Корреляційна матриця R_{xx} x1 x2 x3 <table border="1"> <tr><td>1</td><td>0,9967</td><td>0,9939</td></tr> <tr><td>0,9967</td><td>1</td><td>0,9955</td></tr> <tr><td>0,9939</td><td>0,9955</td><td>1</td></tr> </table> Обернена. матриця C=МОБР(R_{xx}) <table border="1"> <tr><td>157,15</td><td>-127,65</td><td>-29,11</td></tr> <tr><td>-127,65</td><td>215,87</td><td>-88,05</td></tr> <tr><td>-29,11</td><td>-88,05</td><td>117,59</td></tr> </table> k =0,05 R_{xx} + k*I <table border="1"> <tr><td>1,05</td><td>0,9967</td><td>0,9939</td></tr> <tr><td>0,9967</td><td>1,05</td><td>0,9955</td></tr> <tr><td>0,9939</td><td>0,9955</td><td>1,05</td></tr> </table> C_k = МОБР(R_{xx}+k*I) <table border="1"> <tr><td>12,280</td><td>-6,290</td><td>-5,659</td></tr> <tr><td>-6,290</td><td>12,647</td><td>-6,037</td></tr> <tr><td>-5,659</td><td>-6,037</td><td>12,032</td></tr> </table> R_{yx} <table border="1"> <tr><td>0,9813</td><td>0,9901</td><td>0,9925</td></tr> </table> | 1 | 0,9967 | 0,9939 | 0,9967 | 1 | 0,9955 | 0,9939 | 0,9955 | 1 | 157,15 | -127,65 | -29,11 | -127,65 | 215,87 | -88,05 | -29,11 | -88,05 | 117,59 | 1,05 | 0,9967 | 0,9939 | 0,9967 | 1,05 | 0,9955 | 0,9939 | 0,9955 | 1,05 | 12,280 | -6,290 | -5,659 | -6,290 | 12,647 | -6,037 | -5,659 | -6,037 | 12,032 | 0,9813 | 0,9901 | 0,9925 | k | β1 | β2 | β3 | R² | RR | Re,yp |
| | 1 | 0,9967 | 0,9939 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0,9967 | 1 | 0,9955 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0,9939 | 0,9955 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 157,15 | -127,65 | -29,11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -127,65 | 215,87 | -88,05 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -29,11 | -88,05 | 117,59 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1,05 | 0,9967 | 0,9939 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0,9967 | 1,05 | 0,9955 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0,9939 | 0,9955 | 1,05 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 12,280 | -6,290 | -5,659 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -6,290 | 12,647 | -6,037 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -5,659 | -6,037 | 12,032 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0,9813 | 0,9901 | 0,9925 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0,05 | 0,206 | 0,358 | 0,411 | 0,981 | 0,947 | 0,127 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | b0 | b1 | b2 | b3 | R² = ∑ β*R_{yx} =0,9643 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | -446,80 | 0,590 | 0,713 | 0,511 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Рис. 32. Зразок звіту до лабораторної роботи № 9

Корреляційна матриця R_{xx} і коефіцієнти кореляції R_{yx} розраховані за допомогою функції КОРРЕЛ. Через симетрію матриці C_k бета-коефіцієнти обчис-

лені відразу у вигляді матриці-рядка: $\beta^T = R_{yx} \cdot C_k$. Далі зроблений зворотний перехід до b -коефіцієнтів, обчислені значення $Y_p = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3$ і залишки моделі $e = Y - Y_p$. Функцією ДИСПР знайдені дисперсії S_{yy} , S_{rr} , S_{ee} , далі обчислені коефіцієнти детермінації R^2 і RR (вони дорівнюють між собою лише для $k = 0$). Наприкінці, функцією КОРРЕЛ розрахована величина $Re,_{yp}$.

Робочий лист Excel є шаблоном зазначених розрахунків для конкретного значення ридж-параметра k . Задаємо послідовно значення цього параметра з кроком 0,01 і копіюємо результати розрахунків в окрему таблицю (рис. 33). Копіювати треба тільки значення за допомогою меню *Правка, Спеціальна вставка, \odot Вставити значення* (можна також скористатися контекстним меню правого клацання мишкою).

| k | β_1 | β_2 | β_3 | R^2 | RR | $Re,_{yp}$ |
|------|-----------|-----------|-----------|--------|--------|------------|
| 0 | -1,0567 | 1,0941 | 0,9535 | 0,9926 | 0,9926 | 0,0000 |
| 0,01 | -0,1049 | 0,4759 | 0,6167 | 0,9865 | 0,9742 | 0,0540 |
| 0,02 | 0,0674 | 0,4070 | 0,5100 | 0,9840 | 0,9667 | 0,0691 |
| 0,03 | 0,1405 | 0,3808 | 0,4599 | 0,9826 | 0,9601 | 0,0875 |
| 0,04 | 0,1808 | 0,3667 | 0,4305 | 0,9818 | 0,9536 | 0,1071 |
| 0,05 | 0,2060 | 0,3577 | 0,4111 | 0,9812 | 0,9473 | 0,1273 |
| 0,06 | 0,2232 | 0,3513 | 0,3971 | 0,9808 | 0,9411 | 0,1476 |
| 0,07 | 0,2355 | 0,3464 | 0,3865 | 0,9804 | 0,9349 | 0,1679 |
| 0,08 | 0,2447 | 0,3424 | 0,3781 | 0,9800 | 0,9288 | 0,1881 |
| 0,09 | 0,2517 | 0,3391 | 0,3713 | 0,9797 | 0,9228 | 0,2079 |
| 0,10 | 0,2572 | 0,3363 | 0,3655 | 0,9794 | 0,9168 | 0,2275 |

Рис. 33. Значення бета-коефіцієнтів залежно від параметра k

Одержавши таку таблицю, будуюмо графіки гребінного сліду для бета-коефіцієнтів, коефіцієнтів детермінації й коефіцієнта кореляції розрахункових значень із залишками моделі (рис. 34).

Насамперед помічаємо, що оцінка коефіцієнта детермінації за формулою $R^2 = 1 - S_{ee}/S_{yy}$ зі зростанням ридж-параметра k зменшується не так інтенсивно, як оцінка за формулою $RR = S_{rr}/S_{yy}$. Тому в подальшому в якості оцінки коефіцієнта детермінації будемо використовувати оцінку $R^2 = 1 - S_{ee}/S_{yy}$. Зі зростанням ридж-параметра k зростає кореляція Y_p із залишками моделі (це ефект систематичної похибки, що з'являється внаслідок введення ридж-параметра).

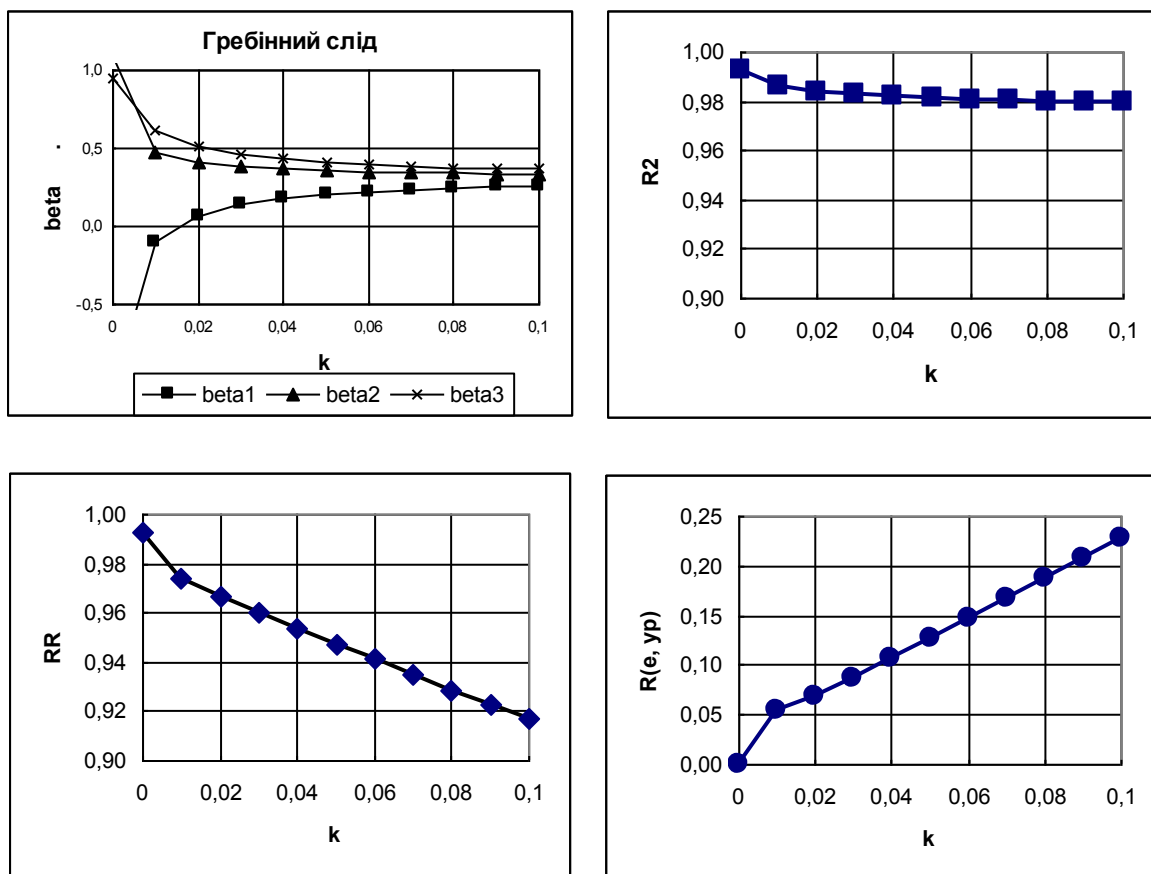


Рис. 34. Графіки залежностей характеристик моделі від параметра k

Самий цікавий графік – це графік гребінного сліду для коефіцієнтів регресії. Цей графік демонструє, наскільки нестабільними були вихідні оцінки і як значно вони змінюються при зовсім незначних значеннях ридж-параметра. При $k = 0,02$ раніше від'ємна оцінка β_1 уже стала позитивною, а при $k = 0,04$ усі оцінки вже стабілізувалися; коефіцієнт детермінації R^2 практично не змінився (він знизився з 0,993 усього до 0,982); між розрахунковими значеннями й залишками з'явилася невелика кореляція $R_{e,yp} = 0,11$.

При $k > 0,05$ всі β -коефіцієнти стають близькими між собою (приймають майже однакове значення) – такий самий результат можна одержати, використовуючи метод головних компонентів.

Отже, маємо три альтернативні моделі (у стандартизованих змінних):

$$Y_p = 0,993X_3; R = 0,993 \text{ – однофакторна модель;}$$

$$Y_p = 0,206X_1 + 0,358X_2 + 0,411X_3; R = 0,991 \text{ – гребінна регресія; (k = 0,05).}$$

$$Y_p = 0,330X_1 + 0,331X_2 + 0,330X_3; R = 0,990 \text{ – за методом головних компонент.}$$

Завдання для самостійної роботи

Розрахувати ридж-оцінки параметрів багатофакторної регресійної моделі при суттєвій мультиколінеарності факторів. Побудувати графіки гребінного сліду для оцінок параметрів і коефіцієнта детермінації.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Які можуть бути наслідки суттєвої мультиколінеарності факторів?
2. Які існують методи подолання мультиколінеарності?
3. Що таке "спроможність", "незміщеність" і "ефективність" статистичних оцінок?
4. У чому полягає метод "гребінної регресії"?
5. Як побудувати графіки гребінного сліду для параметрів моделі?
6. Що порушується при уведенні ридж-параметру?:
7. Як треба обчислювати коефіцієнт детермінації при уведенні ридж-параметра?
8. Які міркування обмежують величину ридж-параметру?
9. Які недоліки (порівнянно з методом гребінної регресії) має метод покрокової регресії?
10. Чи завжди застосування метода покрокової регресії гарантує подолання всіх небажаних наслідків суттєвої мультиколінеарності?
11. У чому полягає ідея метода головних компонентів?

Лабораторна робота № 10. Dummy-змінні

Мета роботи. Освоїти загальну методику введення dummy-змінних для обліку в регресійній моделі якісних факторів.

Dummy у перекладі – макет, і деякі автори вживають словосполучення “макетні змінні”, але частіше використовується неправильна термінологія “фіктивні змінні”. Фіктивною є “змінна” $X_0 \equiv 1$, яку вводять, щоб у матричній формі врахувати в моделі наявність вільного члена. На відміну від X_0 , dummy-змінні дійсно варіюють і тому не повинні називатися “фіктивними”. Найбільш близьким за змістом є назва “індикаторні змінні”.

Dummy-змінні призначені для обліку в моделі якісних ознак, які вимірюються в так званій шкалі імен і не можуть бути охарактеризовані одним числом. Різні рівні якісної ознаки прийнято називати “категоріями”. Приклади: стать – чоловіча або жіноча, стан економіки – до нафтової кризи або пізніше, сезон – зимовий, весняний, літній, осінній, тощо.

Уявимо, треба врахувати якісну ознаку, що має кілька різних категорій. Вводимо таку ж кількість dummy-змінних, кожна з яких дорівнює одиниці для спостережень конкретної категорії і нулю – для всіх інших категорій. У сумі всі dummy-змінні утворюють $X_0 \equiv 1$, тому не можна включати в модель одразу всі dummy-змінні (або ж не треба включати в модель вільний член). Звичайно одну з категорій вибирають як еталон і у модель *не включають* еталонну dummy-змінну. У результаті одержують модель для еталонної категорії з виправленнями на усі інші категорії. Значимості цих виправлень за критерієм Ст’юдента є значимостями відмінностей кожної категорії від еталонної.

У розглянутому нижче прикладі (витрати на газ і електроенергію в США за 1977 – 1982 р.) вивчається типова ситуація, коли, бажаючи збільшити обсяг вибірки, замість середньорічних використали квартальні дані. Хоча вибірка при цьому збільшилася в чотири рази, значимість моделі різко знизилася, а коефіцієнт детермінації впав майже до нуля. Причина виявилася в тому, що при пе-

реході до квартальних даних були привнесені сезонні коливання, на інтенсивному фоні яких зовсім втратився досліджуваний ефект – часовий лінійний тренд. Тому для обліку різниці за кварталами варто ввести чотири додаткові “індикаторні” змінні: $Z_1 = 1$ – для 1-го кварталу і $Z_1 = 0$ – для всіх інших кварталів; $Z_2 = 1$ – для 2-го кварталу і $Z_2 = 0$ – для всіх інших кварталів; $Z_3 = 1$ – для 3-го кварталу і $Z_3 = 0$ – для всіх інших кварталів; $Z_4 = 1$ – для 4-го кварталу і $Z_4 = 0$ – для всіх інших кварталів. Якщо за еталон прийняти 1-й квартал, то змінну Z_1 не слід включати в модель. У результаті розрахунків буде отримано рівняння регресії для еталонної категорії з виправленнями до вільного члена для інших категорій. Можна підключити в модель одразу всі dummy-змінні, але тоді з моделі треба виключити вільний член. У результаті таких розрахунків будуть отримані рівняння регресії для кожної категорії окремо, будуть правильно підраховані всі статистичні характеристики, за винятком статистики Фішера, яка буде занижена в $(m+1)/m = 5/4$ разів. При цих різних підходах за критерієм Ст’юдента оцінюються різні ефекти якісного фактору. За другою методикою оцінюється значимість ефекту кожної категорії, а по першою (що рекомендується) – значимість розходжень кожної категорії від еталонної.

Зараз неявно передбачається, що для кожної сукупності даних (для кожної категорії) зберігаються незмінними усі закономірності в залежностях від кількісних змінних, а вплив якісної ознаки проявляється тільки у виправленнях вільного члена. Бувають і більш складні ситуації, коли для різних категорій виявляються різними ефекти кількісних факторів. Тоді в моделі слід враховувати також члени із взаємодією.

Роботу виконуємо в наступному порядку.

а) до таблиці вихідних даних (стовпці Т та Y) додаємо стовпці індикаторних змінних Z_1, Z_2, Z_3, Z_4 ;

б) за допомогою функції ЛИНЕЙН розраховуємо параметри моделі

$$Y_p = (b_0 + b_2 \cdot Z_2 + b_3 \cdot Z_3 + b_4 \cdot Z_4) + b_5 \cdot T \text{ (1-й квартал – еталон);}$$

г) обчислюємо розрахункові значення (стовпець Y_p);

д) будуємо графіки залежності спостережуваних і розрахункових значень;

е) відбіраємо з таблиці вибірки, що відносяться до різних кварталів;

ж) додаємо на графіку лінії регресії для кожного кварталу.

Зразок звіту до лабораторної роботи № 10 наведено на рис. 35.

| | A | B | C | D | E | F | G | H | I | J | K |
|----|--|---------------|---------------|---------------|---------------|----------|-----------|--------------------|-----------|--------------------|-----------|
| 1 | Робота № 10. Dummy – змінні | | | | | | | | | | |
| 2 | T – Номер кварталу, починаючи з 1977 по 1982 включно | | | | | | | | | | |
| 3 | Y – витрати на газ і електроенергію в США, млрд доларів | | | | | | | | | | |
| 4 | Z1=1 для кварталу I, і Z1=0 для інших кварталів (еталон) | | | | | | | | | | |
| 5 | Z2=1 для кварталу II, і Z2=0 для інших кварталів | | | | | | | | | | |
| 6 | Z3=1 для кварталу III, і Z3=0 для інших кварталів | | | | | | | | | | |
| 7 | Z4=1 для кварталу IV, і Z4=0 для інших кварталів | | | | | | | | | | |
| | Критерії відбору | | | | | | | | | | |
| 8 | Z1 | Z2 | Z3 | Z4 | T | Y | Yp | Z1 | Z2 | Z3 | Z4 |
| 9 | 1 | 0 | 0 | 0 | 1 | 7,33 | 7,5141 | 1 | 1 | 1 | 1 |
| 10 | 0 | 1 | 0 | 0 | 2 | 4,7 | 4,7691 | | | | |
| 11 | 0 | 0 | 1 | 0 | 3 | 5,1 | 4,9991 | 1-й квартал | | 2-й квартал | |
| 12 | 0 | 0 | 0 | 1 | 4 | 5,46 | 5,4158 | T | Yp | T | Yp |
| 13 | 1 | 0 | 0 | 0 | 5 | 7,65 | 7,6405 | 1 | 7,5141 | 2 | 4,7691 |
| 14 | 0 | 1 | 0 | 0 | 6 | 4,92 | 4,8955 | 5 | 7,6405 | 6 | 4,8955 |
| 15 | 0 | 0 | 1 | 0 | 7 | 5,15 | 5,1255 | 9 | 7,7668 | 10 | 5,0218 |
| 16 | 0 | 0 | 0 | 1 | 8 | 5,55 | 5,5421 | 13 | 7,8932 | 14 | 5,1482 |
| 17 | 1 | 0 | 0 | 0 | 9 | 7,96 | 7,7668 | 17 | 8,0195 | 18 | 5,2745 |
| 18 | 0 | 1 | 0 | 0 | 10 | 5,01 | 5,0218 | 21 | 8,1459 | 22 | 5,4009 |
| 19 | 0 | 0 | 1 | 0 | 11 | 5,05 | 5,2518 | 3-й квартал | | 4-й квартал | |
| 20 | 0 | 0 | 0 | 1 | 12 | 5,59 | 5,6685 | T | Yp | T | Yp |
| 21 | 1 | 0 | 0 | 0 | 13 | 7,74 | 7,8932 | 3 | 4,9991 | 4 | 5,4158 |
| 22 | 0 | 1 | 0 | 0 | 14 | 5,1 | 5,1482 | 7 | 5,1255 | 8 | 5,5421 |
| 23 | 0 | 0 | 1 | 0 | 15 | 5,67 | 5,3782 | 11 | 5,2518 | 12 | 5,6685 |
| 24 | 0 | 0 | 0 | 1 | 16 | 5,92 | 5,7948 | 15 | 5,3782 | 16 | 5,7948 |
| 25 | 1 | 0 | 0 | 0 | 17 | 8,04 | 8,0195 | 19 | 5,5045 | 20 | 5,9212 |
| 26 | 0 | 1 | 0 | 0 | 18 | 5,27 | 5,2745 | 23 | 5,6309 | 24 | 6,0476 |
| 27 | 0 | 0 | 1 | 0 | 19 | 5,51 | 5,5045 | | | | |
| 28 | 0 | 0 | 0 | 1 | 20 | 6,04 | 5,9212 | | | | |
| 29 | 1 | 0 | 0 | 0 | 21 | 8,26 | 8,1459 | | | | |
| 30 | 0 | 1 | 0 | 0 | 22 | 5,51 | 5,4009 | | | | |
| 31 | 0 | 0 | 1 | 0 | 23 | 5,41 | 5,6309 | | | | |
| 32 | 0 | 0 | 0 | 1 | 24 | 5,83 | 6,0476 | | | | |
| 33 | b0 | b2 | b3 | b4 | b5 | | | | | | |
| 34 | 7,4825 | -2,777 | -2,578 | -2,193 | 0,0316 | | | | | | |
| 35 | | | | | | | | | | | |
| 36 | b5 | b4 | b3 | b2 | b0 | | T | Z4 | Z3 | Z2 | 1 |
| 37 | 7,2975 | -25,91 | -30,66 | -33,15 | 98,546 | | tb4 | tb3 | tb2 | tb1 | tb0 |
| 38 | 0,0316 | -2,193 | -2,578 | -2,777 | 7,4825 | | b4 | b3 | b2 | b1 | b0 |
| 39 | 0,0043 | 0,0846 | 0,0841 | 0,0838 | 0,0759 | | Sb4 | Sb3 | Sb2 | Sb1 | Sb0 |
| 40 | 0,9867 | 0,1449 | #Н/Д | #Н/Д | #Н/Д | | R2 | Se | | | |
| 41 | 351,16 | 19 | #Н/Д | #Н/Д | #Н/Д | | F | df | | | |
| 42 | 29,480 | 0,3988 | #Н/Д | #Н/Д | #Н/Д | | SSR | SSE | | | |

Рис. 35. Зразок звіту до лабораторної роботи № 10

Усі розрахунки зроблені за допомогою функції ЛИНЕЙН, блок результатів якої доповнений зверху рядком статистик Ст'юдента і рядком заголовків. Поруч у блоці такого ж розміру наведені пояснення до результатів функції ЛИНЕЙН. Одразу під таблицею даних у клітинках А33 – Е33 розташований рядок заголовків b0, b2, b3, b4, b5 у потрібному порядку і нижче у клітинках А34 – Е34 функцією ГПР наведені відповідні коефіцієнти регресії. У клітинці А34 записана формула =ГПР(А33;\$А\$36:\$Е\$39;3;0), яка далі скопійована у сусідні клітинки В34, С34, D34, Е34. Тут \$А\$36:\$Е\$39 – блок результатів функції ЛИНЕЙН, з якого витягується значення коефіцієнта регресії з ім'ям, записаним в А33; у діапазон цього блоку необхідно обов'язково включати рядок заголовків; адреси діапазону повинні бути абсолютними (з фіксаторами \$), щоб ці адреси не змінювалися при копіюванні; числове значення витягується з 3-го рядка блоку (1-й рядок – заголовки, 2-й рядок – статистики Ст'юдента, 3-й рядок – коефіцієнти регресії, 4-й рядок – стандартні похибки коефіцієнтів регресії); нарешті, останній операнд формули ГПР прийнятий рівним нулю (НЕПРАВДА), що означає вимогу *точної* відповідності чисел зазначеному заголовку.

Розрахункові значення Y_p обчислені за допомогою функції СУММПРОИЗВ. Так, для 1-го спостереження у клітинці G9 записана формула:

$$= \$A\$34 + \text{СУММПРОИЗВ}(\$B\$34:\$E\$34; B9:E9).$$

Праворуч від таблиці даних зроблені вибірки розрахункових значень витрат на газ і електроенергію за кожним кварталом. Вибірки з таблиці можна зробити різними способами, наприклад, за допомогою розширеного фільтру. Першим рядком бази даних (таблиці, з якої витягаються дані) завжди повинен бути рядок заголовків стовпців. У блоці критеріїв відбору (у нас таких чотири блоки з адресами H8:H9, I8:I9, J8:J9, K8:K9) у першому рядку повинен бути заголовок хоча б одного стовпця бази даних, у наступних рядках блоку критеріїв записуються умови відбору (в 1-му блоці задана умова $Z_1=1$). Діапазон обраних записів повинен також починатися рядком заголовків стовпців таблиці, що нас цікавлять. Усі чотири діапазони відбору записів (H12:I18, J12:K18, H20:I26, J20:K26) містять однакові заголовки T і Y_p . Встановлюємо табличний курсор у

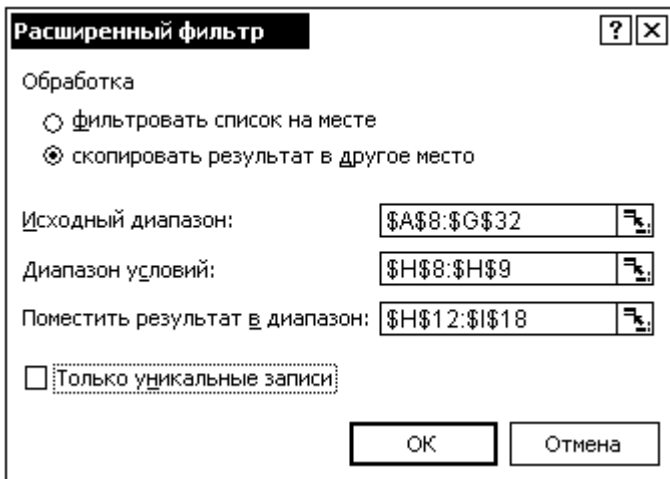


Рис. 36. Панель розширеного фільтру

будь-яке місце таблиці і через меню *Данные, Фильтр, Расширенный фильтр* викликаємо панель розширеного фільтра, на якій поле *Исходный диапазон* вже заповнено за замовчуванням (рис. 36). Включаємо умову \oplus *скопировать результат на другое место*, заповнюємо поля *Диапазон условий* і

Поместить результат в диапазон; нарешті натискаємо кнопку **ОК**.

Попередження. У кожному із цих діапазонів має бути включений рядок заголовків. Не припустимі ніякі помилки при завданні діапазону умов (блоку критеріїв). Якщо виділити діапазон із зайвим рядком (порожнім), буде обрана вся таблиця, а якщо захопити зайвий стовпець – не буде обраний жодний запис. Справа в тому, що умови в декількох рядках поєднуються логічною операцією АБО, а в декількох стовпцях – операцією І.

Далі будуємо діаграму із графіками досліджуваної залежності (рис. 37). За допомогою графічних опцій *Добавить тренд, Показать уравнение* і *R2* додаємо на графіку лінійний тренд, його рівняння і коефіцієнт детермінації.

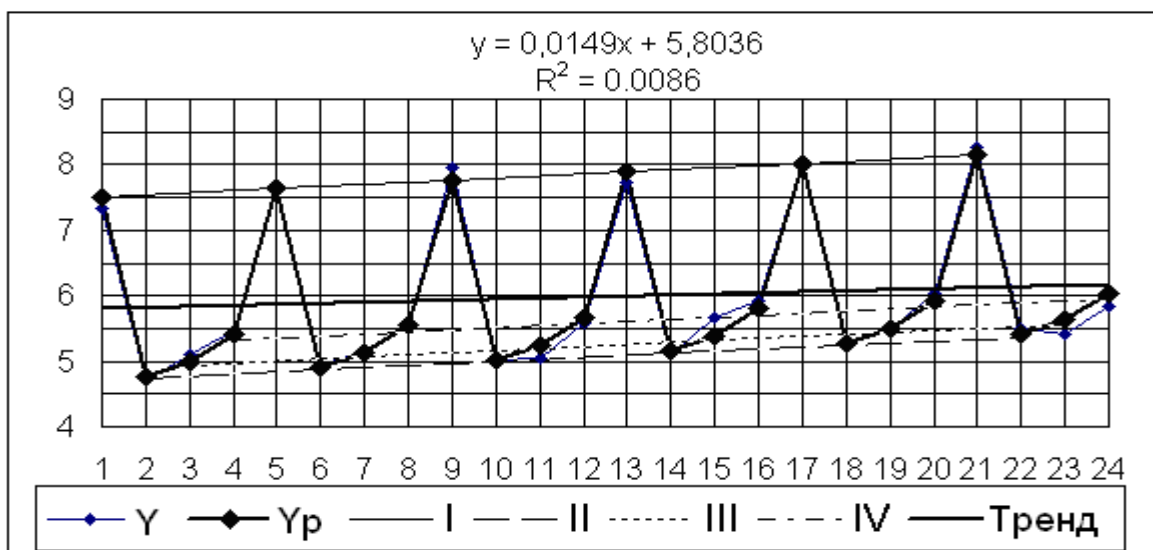


Рис. 37. Поквартальна динаміка витрат

Варто виділити в таблиці стовпці Т, У, Ур разом із заголовками і викликати Майстер Діаграм, тоді в легенді автоматично будуть зазначені назви рядів У і Ур. Після цього можна додати графіки для кожного кварталу наступним чином. Клацаємо правою кнопкою по області діаграми і у контекстному меню вибираємо *Исходные данные*, **Ряд**, **Добавить**.

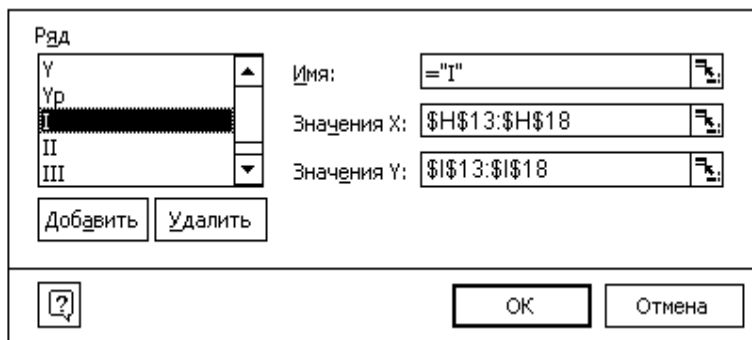


Рис. 38. Панель додавання ряду даних

На панелі додавання даних (рис. 38) у поле *Имя* заносимо I (1-й квартал); це ім'я автоматично замінюється на =I". У поле *Значения X* вказуємо діапазон Т і в поле *Значения Y* – діапазон Уr для 1-го кварталу. Саме так додаємо

графіки для 2-го, 3-го і 4-го кварталів. Нарешті, на діаграмі до ряду У додаємо лінійний тренд, його рівняння і коефіцієнт детермінації R².

На фоні інтенсивних сезонних коливань значимість досліджуваного лінійного тренда виявилася заниженою практично до нуля (R² = 0,0086), оцінка кутового коефіцієнта занижена більш ніж у два рази.

Контрольне запитання. Якої величини були коефіцієнт регресії і коефіцієнт детермінації за 6-ма середньорічними спостереженнями?

За допомогою всього трьох додаткових dummy-змінних Z₂, Z₃, Z₄ дуже добре описана вся залежність разом із сезонними коливаннями. Помітимо, що спроби описати ці коливання аналітичним вираженням у вигляді трьох гармонік (з періодами рік, півроку, квартал) будуть неощадливими і зажадають 6-и додаткових параметрів (по два параметра на кожну гармоніку).

З блоку результатів функції ЛИНЕЙН виписуємо рівняння регресії:

$$Y_r = 7,4825 - 2,777 \cdot Z_2 - 2,278 \cdot Z_3 - 2,193 \cdot Z_4 + 0,0316 \cdot T;$$

(tb) (98,5) (33,2) (30,7) (25,9) (7,3).

Усі члени цієї моделі значимі за критерієм Ст'юдента (всі tb > t₀₁ = 2,9), модель значима в цілому за критерієм Фішера F = 351,2 (F > F₀₁ = 4,5); коефіці-

ент детермінації дорівнює $R^2 = 0,9867$, тобто модель із *dummy*-змінними пояснює практично всю мінливість даних.

Значимості коефіцієнтів регресії перед Z_2, Z_3, Z_4 у цій моделі показують, що витрати на газ і електроенергію в еталонному 1-му кварталі істотно відрізняються від витрат в інших кварталах.

Наводимо рівняння регресії для кожного кварталу:

$$\begin{aligned} Y_p(I) &= 7,4825 + 0,0316 \cdot T & (Z_1 = 1), \\ Y_p(II) &= 7,4825 - 2,777 + 0,0316 \cdot T = 4,7059 + 0,0316 \cdot T & (Z_2 = 1), \\ Y_p(III) &= 7,4825 - 2,278 + 0,0316 \cdot T = 4,9034 + 0,0316 \cdot T & (Z_3 = 1), \\ Y_p(IV) &= 7,4825 - 2,193 + 0,0316 \cdot T = 5,2894 + 0,0316 \cdot T & (Z_4 = 1). \end{aligned}$$

Вище вже вказувалося, що можна в модель включити одразу всі *dummy*-змінні, але тоді модель не повинна містити вільного члена. За допомогою функції ЛИНЕЙН розраховуємо параметри такої лінійної моделі (рис. 39):

$$Y_p = (b_1 \cdot Z_1 + b_2 \cdot Z_2 + b_3 \cdot Z_3 + b_4 \cdot Z_4) + b_5 \cdot T \quad (\text{Константа} - \text{відсутня}).$$

| T | Z4 | Z3 | Z2 | Z1 |
|---------------|---------------|---------------|---------------|---------------|
| 7,2975 | 62,464 | 60,075 | 59,784 | 98,546 |
| 0,0316 | 5,2894 | 4,9043 | 4,7059 | 7,4825 |
| 0,0043 | 0,0847 | 0,0816 | 0,0787 | 0,0759 |
| 0,9867 | 0,1449 | #Н/Д | #Н/Д | #Н/Д |
| 280,93 | 19 | #Н/Д | #Н/Д | #Н/Д |
| 29,4795 | 0,3988 | #Н/Д | #Н/Д | #Н/Д |

Рис. 39. Блок функції ЛИНЕЙН

У результаті будуть отримані одразу всі рівняння за кожним кварталом – не потрібно перераховувати вільний член. Але тепер статистики Ст'юдента будуть оцінювати відхилення середнього рівня витрат у кожному кварталі

від нуля, а не від обраного еталона. Усі характеристики (крім F) вийдуть тими ж самими, лише статистику Фішера треба буде збільшити в $(m+1)/m = 5/4$ разів: $F = 280,93 \cdot 5/4 = 351,16$.

Дотепер вважалося, що нахил лінії регресії є однаковим для кожної категорії, а вплив якісної ознаки проявляється лише у виправленнях до вільного члена в рівнянні регресії. Проте бувають більш складні ситуації, коли для різних категорій виявляються різними ефекти кількісних факторів. Для обліку виправлень на кутові коефіцієнти вводяться *dummy*-змінні із взаємодіями з кількісними факторами. Розглянемо випадок, коли якісна ознака призначена для опису відмінностей двох категорій, причому перша категорія прийнята як еталонна. Тоді досить увести одну *dummy*-змінну: $Z = 0$ – для 1-й категорії й $Z = 1$ –

для 2-й категорії. У рівнянні регресії (яке при $Z = 0$ описує залежність для еталонної 1-й категорії) треба для 2-ї категорії врахувати виправлення як на вільний член, так і на кутовий коефіцієнт:

$$Y_p = (a + c*Z) + (b + d*Z)*X.$$

Тут c – виправлення на вільний член, d – виправлення на кутовий коефіцієнт. Розкриваємо дужки й одержуємо лінійну модель відносно X , Z і $Z*X$:

$$Y_p = a + b*x + c*Z + d*Z*X.$$

Таким чином, для обліку виправлення на кутовий коефіцієнт треба в моделі увести член із взаємодією $Z*X$.

Розглянемо приклад обробки даних (рис. 40), де вивчаються наслідки нафтової кризи 1974 р. Спочатку функцією ЛИНЕЙН оцінюємо параметри лінійної моделі окремо для даних до 1974 р. (блок A30:B35), після 1973 р. (блок D30:E35) і за всім даними (блок G30:H35). Випишуємо із цих блоків рівня регресії:

До нафтової кризи $X < 1974$ р.: $Y_p = -61,87 + 1,280*X$; $R^2 = 0,804$;
 Після нафтової кризи $X \geq 1974$ р.: $Y_p = -0,306 + 0,392*X$; $R^2 = 0,097$;
 За всім даними 1963– 1982 р.р.: $Y_p = -17,83 + 0,625*X$; $R^2 = 0,581$.

Нижче побудовані графіки цих залежностей.

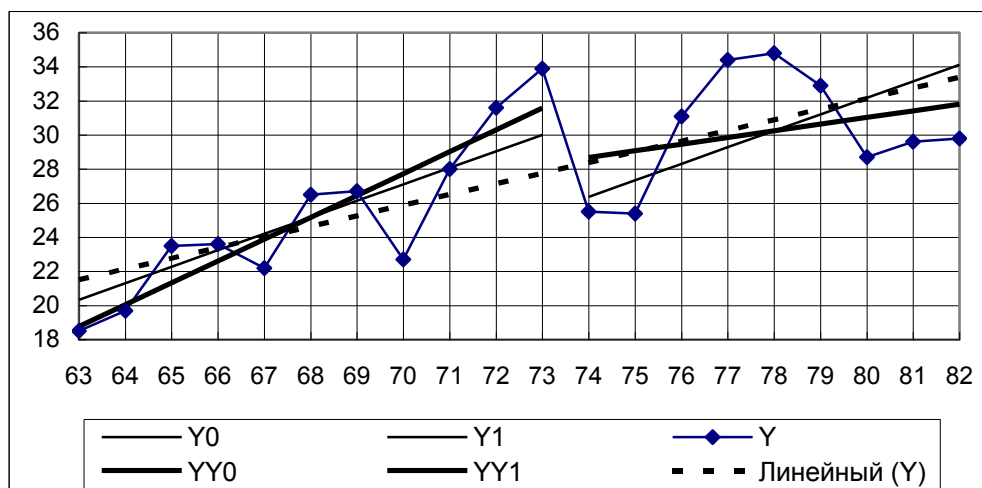


Рис. 40. Графіки залежності витрат до та після нафтової кризи 1974 р.

Тут пунктиром зображений лінійний тренд за всією сукупністю даних і напівжирними відрізками – тренди залежності до та після 1974 року. Візуально відзначаємо різницю між усіма цими залежностями, яка нам видається істот-

ною. Але цей майже очевидний висновок має бути підтверженим більш обгрунтованими міркуваннями з урахуванням думту-змінних.

Ці розрахунки наведено нижче на рис. 41.

| | A | B | C | D | E | F | G | H |
|----|--|---------------|---------------------------|---------------|---------------|---------------|---------------|---------------|
| 1 | Думту-змінні із взаємодією | | | | | | | |
| 2 | У - витрати на автомобілі (США), млрд. доларів | | | | | | | |
| 3 | X = рік – 1900; | | В 1974 р. - нафтова криза | | | | | |
| 4 | Z=0 для X<74; | | Z=1 для X>73 | | | | | |
| 5 | Yp – з обліком Z; YYp - з обліком Z і ZX | | | | | | | |
| 6 | P – ціни на бензин стосовно 1972 р. | | | | | | | |
| 7 | № | X | Z | ZX | У | Yp | YYp | P |
| 8 | 1 | 63 | 0 | 0 | 18,5 | 20,34 | 18,773 | 84,5 |
| 9 | 2 | 64 | 0 | 0 | 19,7 | 21,307 | 20,053 | 84,4 |
| 10 | 3 | 65 | 0 | 0 | 23,5 | 22,273 | 21,333 | 87,5 |
| 11 | 4 | 66 | 0 | 0 | 23,6 | 23,24 | 22,613 | 89,5 |
| 12 | 5 | 67 | 0 | 0 | 22,2 | 24,206 | 23,893 | 92,4 |
| 13 | 6 | 68 | 0 | 0 | 26,5 | 25,173 | 25,173 | 93,8 |
| 14 | 7 | 69 | 0 | 0 | 26,7 | 26,139 | 26,453 | 97,0 |
| 15 | 8 | 70 | 0 | 0 | 22,7 | 27,106 | 27,733 | 97,9 |
| 16 | 9 | 71 | 0 | 0 | 28 | 28,072 | 29,013 | 98,7 |
| 17 | 10 | 72 | 0 | 0 | 31,6 | 29,039 | 30,293 | 100 |
| 18 | 11 | 73 | 0 | 0 | 33,9 | 30,005 | 31,573 | 109,4 |
| 19 | 12 | 74 | 1 | 74 | 25,5 | 26,379 | 28,678 | 147,7 |
| 20 | 13 | 75 | 1 | 75 | 25,4 | 27,345 | 29,069 | 157,7 |
| 21 | 14 | 76 | 1 | 76 | 31,1 | 28,312 | 29,461 | 164,3 |
| 22 | 15 | 77 | 1 | 77 | 34,4 | 29,278 | 29,853 | 173,7 |
| 23 | 16 | 78 | 1 | 78 | 34,8 | 30,244 | 30,244 | 181,3 |
| 24 | 17 | 79 | 1 | 79 | 32,9 | 31,211 | 30,636 | 243,2 |
| 25 | 18 | 80 | 1 | 80 | 28,7 | 32,177 | 31,028 | 337,9 |
| 26 | 19 | 81 | 1 | 81 | 29,6 | 33,144 | 31,419 | 376,4 |
| 27 | 20 | 82 | 1 | 82 | 29,8 | 34,11 | 31,811 | 356,6 |
| 28 | | | | | | | | |
| 29 | 1963 -1973 | | 1974 -1982 | | | 1963 -1982 | | |
| 30 | 6,069 | -4,309 | | 0,8677 | -0,009 | | 4,9978 | -1,961 |
| 31 | 1,28 | -61,87 | | 0,3917 | -0,306 | | 0,6246 | -17,83 |
| 32 | 0,2109 | 14,357 | | 0,4514 | 35,228 | | 0,125 | 9,0891 |
| 33 | 0,8036 | 2,212 | | 0,0971 | 3,4965 | | 0,5812 | 3,2227 |
| 34 | 36,832 | 9 | | 0,7529 | 7 | | 24,978 | 18 |
| 35 | 180,22 | 44,038 | | 9,2042 | 85,578 | | 259,42 | 186,95 |
| 36 | Z | X | 1 | | ZX | Z | X | 1 |
| 37 | -1,683 | 4,1043 | -2,528 | | -1,945 | 1,8047 | 4,7167 | -3,349 |
| 38 | -4,593 | 0,9665 | -40,55 | | -0,888 | 61,562 | 1,28 | -61,87 |
| 39 | 2,7294 | 0,2355 | 16,039 | | 0,4568 | 34,112 | 0,2714 | 18,474 |
| 40 | 0,641 | 3,0703 | #Н/Д | | 0,7096 | 2,8462 | #Н/Д | #Н/Д |
| 41 | 15,176 | 17 | #Н/Д | | 13,034 | 16 | #Н/Д | #Н/Д |
| 42 | 286,12 | 160,25 | #Н/Д | | 316,75 | 129,62 | #Н/Д | #Н/Д |

Рис. 41. Розрахунки з думту-змінними і їх взаємодіями

Тут крім аргументу $X = \text{Рік} - 1900$ була уведена *dummy*-змінна Z , яка дорівнює нулю для $X < 1974$ і одиниці для $X \geq 1974$ (до і після нафтової кризи). Для врахування взаємодії була додана ще змінна $ZX = Z * X$ (добуток Z і X).

Оцінюємо (функцією ЛИНЕЙН) параметри залежності з *dummy*-змінними $Y_p = b_0 + b_1 * X + b_2 * Z$ і $Y_p = b_0 + b_1 * X + b_2 * Z + b_3 * ZX$ (блоки А37:С42 і Е37:Н42 відповідно). Графіки цих залежностей також наведені на діаграмі рис. 40 (тонкі й напівжирні відрізки прямих).

Кожний блок висновку функції ЛИНЕЙН доповнений нами рядком статистик Ст'юдента $t_b = b/S_b$, звідки з блоку Е37:Н42 виходить висновок про незначимість обох *dummy*-ефектів ($t_{b2} = 1,80 < t_{05}$; $t_{b3} = 1,94 < t_{05}$). Спочатку вилучаємо незначиму взаємодію ZX і переходимо до блоку А38:С43, де зроблені розрахунки (функцією ЛИНЕЙН) параметрів моделі без взаємодії $Z * X$. Внесок, що визначається залишеною *dummy*-змінною Z також виявився незначимим за критерієм Ст'юдента ($t_{b2} = 1,68 < t_{05}$).

Виявилось, що на нас неявно впливає психологічний фактор – адже ми знаємо, що в 1974 р. стабільні до цього ціни на бензин у США підвищилися відразу в півтора рази й далі почали зростати. Природно, ми свідомо очікували, що наслідки нафтової кризи повинні позначитися на статті витрат на автомобілі. Але цей статистично очікуваний ефект раніше не був доведений, а подальший аналіз показав, що на фоні досить великої мінливості даних нуль-гіпотеза про відсутність відмінностей між двома вибірками *не може бути відкинута*.

Уважно розглядаючи діаграму на рис. 40, відзначимо, що після 1973 р. суттєво збільшився розкид даних. Підозрюємо, що саме це і є причина того, що ефекти *dummy*-змінних виявилися незначущими. Більша мінливість даних після 1973 р. пояснюється нестабільністю цін на бензин після цього періоду, тому облік у моделі змінної P (останній стовпчик таблиці на рис. 41) має суттєво знизити величину залишкової дисперсії. Нижче на рис. 42 наведені розрахунки (функцією ЛИНЕЙН) з урахуванням цін на бензин, звідки виходить, що тепер ефекти всіх змінних значимі за критерієм Ст'юдента:

| P | ZX | Z | X | 1 |
|----------------|---------------|----------------|---------------|----------------|
| -3,7618 | 2,0610 | -2,1335 | 7,0505 | -4,825 |
| -0,0795 | 1,4668 | -112,47 | 1,4549 | -66,277 |
| 0,0211 | 0,7117 | 52,7144 | 0,20636 | 13,7362 |
| 0,8506 | 2,1086 | #Н/Д | #Н/Д | #Н/Д |
| 21,3479 | 15 | #Н/Д | #Н/Д | #Н/Д |
| 379,675 | 66,6943 | #Н/Д | #Н/Д | #Н/Д |

| P | ZX | Z | X | 1 |
|-----------|-----------|-----------|-----------|-----------|
| tb4 | tb3 | tb2 | tb1 | tb0 |
| b4 | b3 | b2 | b1 | b0 |
| Sb4 | Sb3 | Sb2 | Sb1 | Sb0 |
| R2 | Se | | | |
| F | dfe | | | |
| SSR | SSE | | | |

Рис. 42. Розрахунки з урахування цін на пальне

Випишемо загальне рівняння регресії й за кожним періодом окремо:

$$Y_p = (-66,277 - 112,470 \cdot Z) + (1,4549 + 1,4668 \cdot Z) \cdot X - 0,0795 \cdot P$$

До 1974 р.: $Y_p = -66,277 + 1,4549 \cdot X - 0,0795 \cdot P;$

Після 1973 р.: $Y_p = -178,75 + 2,9217 \cdot X - 0,0795 \cdot P.$

Позначимо через $W = Y - b_4 \cdot P = Y + 0,0795 \cdot P$ – додаткові витрати на автомобілі крім витрат на пальне. Побудуємо на рис. 43 графіки залежностей W за всіма спостереженнями, до та після нафтової кризи.

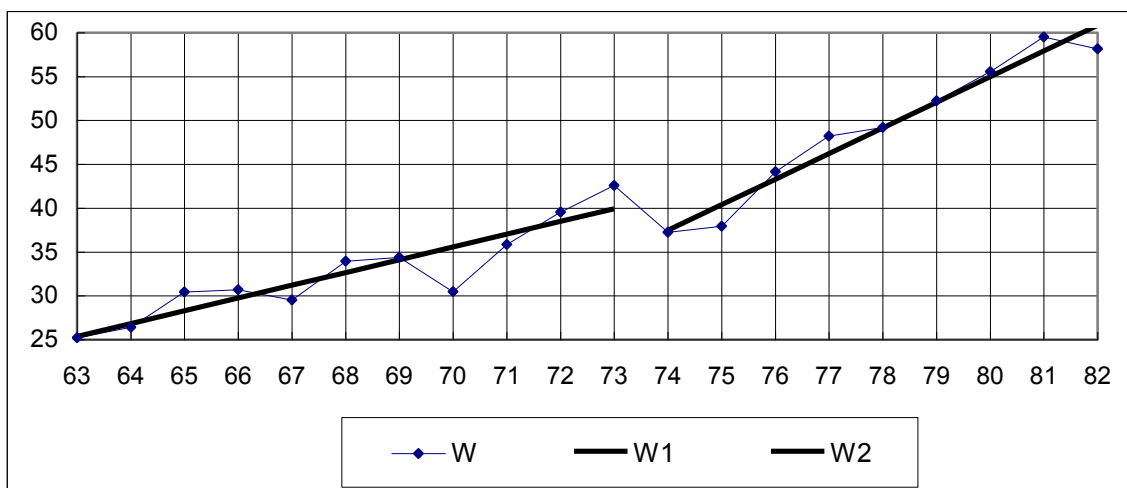


Рис. 43. Динаміка додаткових витрат на автомобілі крім витрат на пальне

Які тепер можна зробити висновки? Чи змінилися вони?

Завдання для самостійної роботи

Виконати математичну обробку даних, узятих з різних сукупностей (тобто в моделі врахувати якісну змінну, що визначає, до якої сукупності належать окремі спостереження). Оцінити суттєвість різниць між сукупностями. Чи не можна їх просто об'єднати в одну сукупність? Отримати рівняння регресії для кожної сукупності та разом для об'єднаних даних.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Що таке "dummy-змінні"? З якою метою вони застосовуються? Які є пропозиції щодо найближчого перекладу назви цих змінних на українську мову?
2. У "коваріаційному аналізі", який є подальшою модифікацією дисперсійного аналізу, крім середніх за групами спостережень оцінюється також лінійний ефект кількісної змінної. Чи не можна вважати модель коваріаційного аналізу еквівалентним частинним випадком регресійного аналізу з dummy-змінними? Для яких умов?
3. Яку роль відіграють члени взаємодій кількісних змінних з dummy-змінними?
4. В якому випадку аналіз з dummy-змінними буде еквівалентним серії звичайних регресійних аналізів окремо для кожної категорії якісної змінної?
5. Як оцінити значущість членів моделі з dummy-змінними?

Лабораторна робота № 11. Виробничі функції Коба – Дугласа

Мета роботи. Освоїти обробку даних за загальною степеневою моделлю й моделлю Коба – Дугласа, оцінити еластичності виробничих факторів і ефект масштабу виробництва, увести в модель експоненціальний часовий тренд для оцінки ефекту науково–технічного прогресу.

Теоретична частина

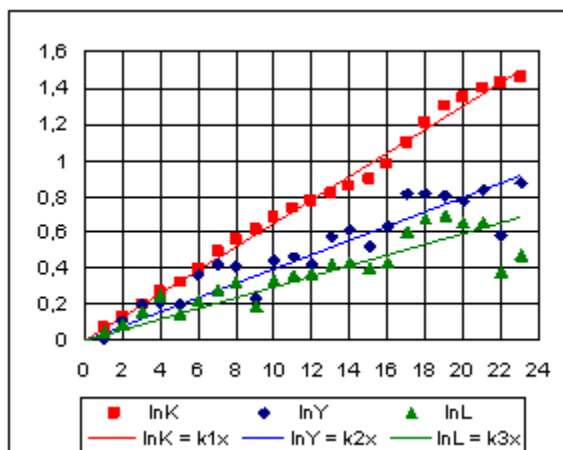


Рис. 44. Часові ряди $\ln K$, $\ln Y$, $\ln L$

У 1927 р. економіст Пол Дуглас, розглядаючи діаграми логарифмів капітальних витрат (K), обсягів випуску продукції (Y) і витрат труда (L), виявив, що відстані від точок графіка $\ln Y$ до точок графіків $\ln K$, $\ln L$ складають однакову пропорцію для всіх спостережень з 1900 по 1922 рр. На підставі цього факту математик Чарльз Коб показав, що така особливість має місце для залежності $Y = K^b L^{1-b}$, яка зараз має

назву "модель Коба – Дугласа". Цікаво, яким шляхом йде думка винахідника. Своє відкриття П. Дуглас зробив лише тому, що для 1989 року він прийняв значення всіх показників K , Y , L за 1 (100 %), тобто задав фіксований вузол, де перетинаються графіки залежностей $\ln K$, $\ln Y$, $\ln L$ за часом ($x = t - 1989$). Фактично він помітив лінійність трендів логарифмів показників (рис. 44), і цього факту має бути досить, що зробити всі наступні висновки.

Відтворимо хід міркувань математика Ч. Коба. Запишемо умову пропорційності відрізків $(\ln K - \ln Y) = \lambda (\ln Y - \ln L)$, що помічена П. Дугласом, і розв'яжемо цей вираз відносно $\ln Y$: $\ln Y = b \cdot \ln K + c \cdot \ln L$, де $b = 1/(1+\lambda)$, $c = \lambda/(1+\lambda)$. Виявилось, що в сумі $(b + c) = 1$, звідки виходить залежність $Y = K^b L^{1-b}$.

Пом'якшимо вимогу пропорційності й замінимо її на більш загальну умову лінійності:

$$\ln Y = a + b \cdot \ln K + c \cdot \ln L.$$

Після потенціювання одержуємо загальну степеневу функцію:

$$Y = A K^b L^c.$$

Параметри b і c у степеневій залежності є коефіцієнтами еластичності випуску за витратами капіталу й праці. Іншими словами, загальна степенева функція описує залежності зі сталою еластичністю.

Коефіцієнт еластичності за капіталом $E_{Y/K}$ показує, на скільки відсотків збільшується обсяг випуску при збільшенні витрат капіталу на 1%. Аналогічно визначається коефіцієнт еластичності за працею $E_{Y/L}$.

Покажіть, що для степеневі залежності $Y = A K^b L^c$ виходить:

$$E_{Y/K} = \frac{\partial Y/Y}{\partial K/K} = \frac{\partial Y}{\partial K} \cdot \frac{K}{Y} = b; \quad E_{Y/L} = \frac{\partial Y/Y}{\partial L/L} = \frac{\partial Y}{\partial L} \cdot \frac{L}{Y} = c;$$

тобто $E_{Y/K} = b$, $E_{Y/L} = c$.

Можна також показати, що параметри b і c у степеневі залежності характеризують прогнозовані частки виробничих факторів, зокрема, b є частка прибутку на одиницю витрат капіталу, а c – на одиницю витрат праці.

Тепер обговоримо умову $(b + c) = 1$. Зовсім очевидно, що $b, c > 0$, тому що збільшення витрат виробничих факторів повинне викликати збільшення випуску продукції. Розумно припустити, що $b, c < 1$, оскільки пропорційне збільшення витрат на виробництво приводить до усе більш повільного росту випуску продукції. Далі, якщо обидва виробничих фактора (K і L) збільшити вдвічі, то обсяг випуску (згідно зі степеневою залежністю) збільшиться в 2^{b+c} раз і при $(b + c) > 1$ вийде зростаючий ефект від масштабу виробництва (що не спостерігається у дійсності). Спочатку Коб і Дуглас припускали сталу віддачу від масштабу виробництва, тобто постулювали умову $(b + c) = 1$. Згодом вони послабили це допущення, бажаючи оцінювати ступінь віддачі від масштабу виробництва.

Після логарифмування модель стає лінійною щодо параметрів і може бути оцінена методом найменших квадратів. При додатковій умові $(b + c) = 1$ модель перетворюють до виду:

$$\ln \frac{Y}{L} = a + b \cdot \ln \frac{K}{L}.$$

(Зробіть ці перетворення).

Для оцінки значимості фактора масштабу виробництва можна прийняти $(b + c) = 1 + d$ (звідки $c = 1 - b + d$) і перевірити за критерієм Ст'юдента значимість параметра d у моделі:

$$\ln \frac{Y}{L} = a + b \cdot \ln \frac{K}{L} + d \cdot \ln L.$$

Ця модель еквівалентна загальній степеневій моделі, але по-різному застосовуються статистичні критерії.

Зростання населення, а також зростаючий внесок науки й техніки у наступні роки викликав необхідність обліку в моделі цих ефектів. Але ця якісна ознака не вимірюється одним числом, тому в якості так званої *заміщаючої змінної* в модель був доданий час t :

$$Y = A K^b L^c e^{gt}.$$

Проте уведення заміщаючої змінної не виправдало свого призначення, тому що виявилось, що всі виробничі фактори мають часовий тренд, і включення часу в число пояснюючих змінних приводить до мультиколінеарності, через це всі оцінки параметрів будуть ненадійними.

На рис. 45 наведено зразок виконання роботи. Стівці в таблиці даних розташовані так, щоб єдиним блоком можна було задавати діапазон *Известные значения X* функції ЛИНЕЙН для різних варіантів розрахунку. У перших 4-х стівцях задані вихідні дані: T – рік, Y , K , L – у відсотках стосовно до 1899 г. Далі обчислені логарифми $\ln Y$, $\ln K$, $\ln L$, але перед стівцями $\ln K$, $\ln L$ вставлений стівпець $t = T - 1990$ для можливості включення в модель заміщаючої змінної t . Поруч зі стівцем $\ln L$ обчислені значення $\ln(K/L)$, щоб при бажанні можна було включити в модель обидві ці змінні. Нарешті, в останньому стівці обчислені значення $\ln(Y/L)$.

Далі оцініні коефіцієнти еластичності випуску за витратами капіталу й праці $b = E_{Y/K}$, $c = E_{Y/L}$ (на думку Коба і Дугласа ці величини має бути близькими до $\beta = 0,25$ і $\gamma = 0,75$ відповідно).

| | A | B | C | D | E | F | G | H | I | J |
|----|-------------------------------------|----------|---|-------------------------------|------------|---------------|---|------------|----------------|----------------|
| 1 | Робота 11. | | Виробничі функції Коба – Дугласа | | | | | | | |
| 2 | | | | | | | | | | |
| 3 | | | Y – обсяг випуску в % до 1899 г. | | | | | | | |
| 4 | | | K – витрати капіталу в % до 1899 г. | | | | | | | |
| 5 | | | L – витрати праці в % до 1899 г. | | | | | | | |
| 6 | | | | | | | | | | |
| 7 | $Y_p = A * (Kb)^c * (Lc)$ | | | При $b + c = 1$ | | | При $b + c = 1 + d$ | | | |
| 8 | $\ln y = a + b * \ln k + c * \ln l$ | | | $\ln(Y/L) = a + b * \ln(K/L)$ | | | $\ln(Y/L) = a + b * \ln(K/L) + d * \ln l$ | | | |
| 9 | | | | | | | | | | |
| 10 | T | Y | K | L | lnY | T-1900 | lnK | lnL | ln(K/L) | ln(Y/L) |
| 11 | 1900 | 101 | 107 | 105 | 4,61512 | 0 | 4,67283 | 4,65396 | 0,01887 | -0,0388 |
| 12 | 1901 | 112 | 114 | 110 | 4,71850 | 1 | 4,73620 | 4,70048 | 0,03572 | 0,01802 |
| 13 | 1902 | 122 | 122 | 118 | 4,80402 | 2 | 4,80402 | 4,77068 | 0,03334 | 0,03334 |
| 14 | 1903 | 124 | 131 | 129 | 4,82028 | 3 | 4,87520 | 4,85981 | 0,01538 | -0,0395 |
| 15 | 1904 | 122 | 138 | 116 | 4,80402 | 4 | 4,92725 | 4,75359 | 0,17366 | 0,05043 |
| 16 | 1905 | 143 | 149 | 125 | 4,96284 | 5 | 5,00395 | 4,82831 | 0,17563 | 0,13453 |
| 17 | 1906 | 152 | 163 | 133 | 5,02388 | 6 | 5,09375 | 4,89035 | 0,20340 | 0,13353 |
| 18 | 1907 | 151 | 176 | 138 | 5,01728 | 7 | 5,17048 | 4,92725 | 0,24323 | 0,09003 |
| 19 | 1908 | 126 | 185 | 121 | 4,83628 | 8 | 5,22036 | 4,79579 | 0,42457 | 0,04049 |
| 20 | 1909 | 155 | 198 | 140 | 5,04343 | 9 | 5,28827 | 4,94164 | 0,34662 | 0,10178 |
| 21 | 1910 | 159 | 208 | 144 | 5,06890 | 10 | 5,33754 | 4,96981 | 0,36772 | 0,09909 |
| 22 | 1911 | 153 | 216 | 145 | 5,03044 | 11 | 5,37528 | 4,97673 | 0,39854 | 0,05370 |
| 23 | 1912 | 177 | 226 | 152 | 5,17615 | 12 | 5,42053 | 5,02388 | 0,39665 | 0,15227 |
| 24 | 1913 | 184 | 236 | 154 | 5,21494 | 13 | 5,46383 | 5,03695 | 0,42688 | 0,17798 |
| 25 | 1914 | 169 | 244 | 149 | 5,12990 | 14 | 5,49717 | 5,00395 | 0,49322 | 0,12595 |
| 26 | 1915 | 189 | 266 | 154 | 5,24175 | 15 | 5,58350 | 5,03695 | 0,54654 | 0,20479 |
| 27 | 1916 | 225 | 298 | 182 | 5,41610 | 16 | 5,69709 | 5,20401 | 0,49309 | 0,21209 |
| 28 | 1917 | 227 | 335 | 196 | 5,42495 | 17 | 5,81413 | 5,27811 | 0,53602 | 0,14684 |
| 29 | 1918 | 223 | 366 | 200 | 5,40717 | 18 | 5,90263 | 5,29832 | 0,60432 | 0,10885 |
| 30 | 1919 | 218 | 387 | 193 | 5,38450 | 19 | 5,95842 | 5,26269 | 0,69573 | 0,12180 |
| 31 | 1920 | 231 | 407 | 193 | 5,44242 | 20 | 6,00881 | 5,26269 | 0,74612 | 0,17973 |
| 32 | 1921 | 179 | 417 | 147 | 5,18739 | 21 | 6,03309 | 4,99043 | 1,04265 | 0,19695 |
| 33 | 1922 | 240 | 431 | 161 | 5,48064 | 22 | 6,06611 | 5,08140 | 0,98470 | 0,39923 |

Рис. 45. Зразок звіту до лабораторної роботи № 11

Заповнюємо поля введення функції ЛИНЕЙН: *Известные значения Y* – стовпець lnY; *Известные значения X* – стовпці lnK, lnL; *Константа* – 1, *Статистика* – 1. До блоку результатів функції ЛИНЕЙН, як звичайно, доданий рядок статистик Ст'юдента. Поруч наведений блок пояснень (рис. 46).

| c | b | a | | c | b | a |
|----------------|----------------|---------------|--|-----------|-----------|-----------|
| 5,14785 | 3,81268 | -0,2363 | | tb2 | tb1 | tb0 |
| 0,77845 | 0,24715 | -0,111 | | b2 | b1 | b0 |
| 0,15122 | 0,06482 | 0,46978 | | Sb2 | Sb1 | Sb0 |
| 0,94844 | 0,06082 | #Н/Д | | R2 | Se | |
| 183,936 | 20 | #Н/Д | | F | df | |
| 1,3606 | 0,07397 | #Н/Д | | SSR | SSE | |

Рис. 46. Блок висновків функції ЛИНЕЙН

Одержано $b = 0,247$ і $c = 0,778$ – дійсно близькі значення до $\beta = 0,25$ і $\gamma = 0,75$. Значення b і c значимо відрізняються від нуля, тому що $t_b = 3,81 > t_{05}(20)$ і $t_c = 5,14 > t_{05}(20)$, де $t_{05}(20) = 2,08$; $t_{01}(20) = 2,84$.

Перевіримо справедливість гіпотези $\beta = M(b) = 0,25$ і $\gamma = M(c) = 0,75$, для чого обчислюємо границі 95%-х довірчих інтервалів на β і γ :

| | |
|--|---|
| $b - t_{05} * S_b < \beta < b + t_{05} * S_b;$ | $c - t_{05} * S_c < \gamma < c + t_{05} * S_c;$ |
| $0,247 - 2,08 * 0,065 < \beta < 0,247 + 2,08 * 0,065;$ | $0,778 - 2,08 * 0,151 < \gamma < 0,778 + 2,08 * 0,151;$ |
| $0,112 < \beta < 0,382;$ | $0,463 < \gamma < 1,094.$ |

Ці інтервали накривають очікувані значення, тому нуль-гіпотеза про рівність $\beta = 0,25$ і $\gamma = 0,75$ не може бути відкинута. Сума $(b + c) = 1,0256$ дуже близька до одиниці.

Вільний член моделі $a = -0,111$ не значимий за Ст'юдентом ($t_a = 0,24 < 1$) і може бути прирівняний до нуля.

Таким чином, вихідні дані добре описуються моделлю $Y = K^{0,25} L^{0,75}$, або $\ln Y = 0,25 * \ln K + 0,75 * \ln L$.

Приймаємо зв'язок $(b + c) = 1$ між коефіцієнтами еластичностей випуску за витратами капіталу й праці і приводимо загальну степеневу модель $\ln Y = a + b * \ln K + c * \ln L$ до вигляду $\ln(Y/L) = a + b * \ln(K/L)$.

Якщо ми бажаємо перевірити значимість ефекту масштабу виробництва, то слід урахувати виправлення $b + c = 1 + d$, для чого додатково включаємо в модель член $\ln L$: $\ln(Y/L) = a + d * \ln L + b * \ln(K/L)$. Нижче функцією ЛИНЕЙН обчислені параметри моделі Коба – Дугласа з виправленням і без неї (рис. 47).

| b | d | a | | b | a |
|----------------|---------------|---------------|--|----------------|----------------|
| 3,81268 | 0,26118 | -0,2363 | | 5,83319 | 0,52534 |
| 0,24715 | 0,0256 | -0,111 | | 0,25922 | 0,01155 |
| 0,06482 | 0,09801 | 0,46978 | | 0,04444 | 0,02199 |
| 0,61966 | 0,06082 | #Н/Д | | 0,61836 | 0,05945 |
| 16,2923 | 20 | #Н/Д | | 34,0261 | 21 |
| 0,12052 | 0,07397 | #Н/Д | | 0,12026 | 0,07422 |

Рис. 47. Оцінка значущості ефекту масштабу виробництва

Модель із виправленням еквівалентна повної степеневій моделі й дає ті самі оцінки параметрів. Проте тепер за Ст'юдентом перевіряється значимість

виправлення d (замість значимості параметра c). Оскільки $td = 0,26 < 1$, нуль-гіпотеза про рівність виправлення $d = 0$ не може бути відкинута.

У сусідньому блоці функцією ЛИНЕЙН оцінені параметри моделі після вибракування незначущого виправлення: $\ln(Y/L) = 0,012 + 0,259 \cdot \ln(K/L)$.

Відзначаємо, що залишкова дисперсія MSE у результаті вибракування зменшалася з 0,0608 до 0,0594, отже, вибракування зізнається корисним. Відзначаємо, що оцінка еластичності $b = 0,259$ дуже близька до $\beta = 0,25$.

Але з'явилася деяка дивина – коефіцієнт детермінації для повної степеневі моделі дорівнює $R^2 = 0,948$, а для еквівалентної моделі з виправленням цей показник знизився до $R^2 = 0,620$. У чому тут справа, поясніть! (Натяк: уявіть собі, що для моделі відносно $\ln(Y/L)$ отримане $R^2=0$. Який зв'язок тоді буде між $\ln Y$ і $\ln L$, яка тіснота цього зв'язку?).

Питання. Чи можна оцінити значимість виправлення $d = (b + c) - 1$ безпосередньо за даними для вихідної степеневі моделі? Як побудувати 95%-й довірчий інтервал на суму $(b + c)$? Якщо цей інтервал накріє очікуване значення 1, гіпотеза про рівність $(b + c) = 1$ не може бути відхилена.

Порада. Дисперсія суми двох випадкових величин дорівнює сумі їх дисперсій плюс подвоєна коваріація. Дисперсії й коваріації оцінок параметрів моделі пропорційні елементам оберненої матриці $C = \text{МОБР}(XTX)$. З яким коефіцієнтом пропорційності? Матриця X складається зі стовпців $X_0, \ln K, \ln L$. Стовпець $X_0=1$ треба сформулювати на місці стовпця $t = T - 1900$ (стовпці X_0 і t легко формуються за допомогою автозаповнення).

Обчисліть дисперсії й коваріації оцінок параметрів моделі й перевірте правильність розрахунків $S_a = 0,4798$, $S_b = 0,0648$, $S_c = 0,1512$ функцією ЛИНЕЙН. Обчисліть дисперсію суми $(b + c)$ і побудуйте 95%-й довірчий інтервал на цю суму. Перевірте гіпотезу про рівність $(b + c) = 1$.

Включіть в модель Коба – Дугласа час як заміщаючу змінну для обліку технічного прогресу. Проаналізуйте результат і спробуйте його пояснити.

Порада. Обчисліть (функцією КОРРЕЛ) коефіцієнти кореляції між пояснюючими змінними й переконаєтеся в наявності тісних мультиколінеарних зв'язків (найбільш сильно корелюють t і $\ln K$).

Завдання для самостійної роботи

Скласти модель Коба – Дугласа. Оцінити значущість масштабного ефекту і часового тренда. Оцінити еластичності виробничих факторів

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Сформулюйте область застосування моделі Коба – Дугласа.
2. Чим модель Коба – Дугласа відрізняється від звичайної двофакторної степеневі моделі?
3. Дайте змістовну інтерпретацію параметрам виробничої функції Коба – Дугласа.
4. Що таке "коефіцієнт еластичності"?
5. Що таке "масштабний ефект виробництва"?
6. Що таке "заміщаючі змінні"? Чому їх треба враховувати?
7. Чому для повної степеневі моделі коефіцієнт детермінації виявився значно більшим, ніж для майже еквівалентній моделі Коба – Дугласа? Як інтерпретувати цей ефект? Натяк: при розрахунку параметрів степеневі моделі за результативну ознаку прийнято $\ln Y$, а для моделі Коба – Дугласа прийнято $\ln(Y/L)$.
8. Як побудувати 95%-й довірчий інтервал на суму $(b+c)$? Яке теоретичне значення очікується і що це має означати?
9. Чому врахування часового тренда, як заміщаючої змінної, виявилось в моделі Коба – Дугласа неефективним?

Тема 4. Системи економетричних рівнянь

Лабораторна робота № 12. Двокроковий метод найменших квадратів

Мета роботи: Оцінити параметри моделі Л. Клейна 1МНК і 2МНК.

Теоретична частина

Економетричні моделі макроекономіки складаються у вигляді системи декількох регресійних рівнянь, тотожностей і нерівностей. Модель у такій формі називається структурної й відбиває основні закони економіки. Для прикладу розглянемо першу економетричну модель Л. Клейна для опису економіки США в період між двома світовими війнами (1921 – 1941 рр.). Модель Клейна складається з трьох регресійних рівнянь і трьох тотожностей:

$$\begin{aligned}C &= a_0 + a_1 \cdot P + a_2 \cdot \text{PLAG} + a_3 \cdot (W + WP) + u; \\I &= b_0 + b_1 \cdot P + b_2 \cdot \text{PLAG} + b_3 \cdot \text{KLAG} + v; \\W &= c_0 + c_1 \cdot X + c_2 \cdot \text{XLAG} + c_3 \cdot t + w; \\X &= C + I + G; \\P &= X - W - T; \\K &= \text{KLAG} + I.\end{aligned}$$

Ці шість рівностей визначають шість результативних ознак, які називаються *ендогенними* (внутрисистемними) змінними:

C – (consumption) – обсяг споживання; **I** – (investment) – інвестиції;

W – (private wage bill) – особистий дохід у приватному секторі;

X – (private production) – обсяг виробництва приватних фірм;

P – (profits) – прибуток; **K** – (capital stock) – акціонерний капітал.

Серед пояснюючих змінних є ті самі ендogenous змінні й ендogenous змінні із запізнюванням на один період (тобто дані минулого року): XLAG, PLAG, KLAG. У науковій літературі звичайно такі змінні позначають X_{t-1} , P_{t-1} , K_{t-1} , інші змінні мають індекс t : C_t , I_t , W_t , X_t , P_t , K_t . Далі серед пояснюючих змінних є так звані *екзогенні* (позасистемні) змінні, значення яких задаються з зовні:

t – (year) – час; WP – (government wage bill) – зарплата в держсекторі;

G – (government demand) – державні закупівлі; T – (taxes) – податки.

Лагові ендогенні й екзогенні змінні називаються *передвизначеними*, їх значення відомі на цей момент часу.

Члени u , v , w описують випадкові похибки в кожному рівнянні.

Контрольне запитання. Дайте економічну інтерпретацію кожній рівності (рівнянню й тотожності) економетричній моделі Клейна.

Помітимо, що в системі *одночасних рівнянь* деякі змінні в одних рівняннях виступають у ролі результативних ознак, а в інших – у якості пояснюючих змінних (ендогенні пояснюючі змінні). Це приводить до певних утруднень в оцінюванні параметрів моделі. Так, оцінювання звичайним МНК кожного регресійного рівняння окремо приводить до наступних результатів (отримайте ці результати):

$$\begin{aligned}C &= 16,24 + 0,193 \cdot P + 0,090 \cdot \text{PLAG} + 0,796 \cdot (W + WP) + u; \\I &= 10,13 + 0,480 \cdot P + 0,333 \cdot \text{PLAG} - 0,110 \cdot \text{KLAG} + v; \\W &= 1,497 + 0,439 \cdot X + 0,146 \cdot \text{XLAG} + 0,130 \cdot t + w.\end{aligned}$$

Оскільки всі змінні виражені в одних одиницях (млрд дол. США за курсом 1931 р.), то коефіцієнти регресії при P і PLAG можна трактувати як норми відрахувань прибутку в споживання й в інвестиції. Але тоді виходить, що відрахування в споживання становлять 19% прибутку поточного року й лише 9% прибутку минулого року; аналогічно, в інвестиції відраховується 48% прибутку поточного року й 33,3% прибутку минулого року. Це зовсім не відповідає дійсності. Лише невелика частка прибутку поточного року може бути освоєна в цьому ж році й тільки тому, що фінансовий рік не збігається з календарним.

Звичайні МНК-оцінки параметрів моделі виявилися зміщеними.

Причина настільки великого зсуву полягає в порушенні першої (основний) передумови регресійного аналізу, згідно з якою випадкові похибки відносяться лише до результативних ознак і не повинні корелювати з пояснюючими змінними. Проте у системі одночасних рівнянь через систему тотожностей випадкові похибки аддитивно переносяться на ендогенні пояснюючі змінні. Так, з першого рівняння виходить, що випадкова похибка u додається до C , далі згідно з першою тотожністю $X = C + I + G$ ця похибка переноситься на X і, нарешті, згідно із другою тотожністю $P = X - W - T$ похибка u переноситься на P – пояс-

ноючу змінну цього ж рівняння: $u \rightarrow \mathbf{C} \rightarrow \mathbf{X} \rightarrow \mathbf{P}$. Випадкова похибка другого рівняння v також тісно корелює з \mathbf{P} – пояснюючої змінної другого рівняння: $v \rightarrow \mathbf{I} \rightarrow \mathbf{X} \rightarrow \mathbf{P}$ (покажіть це). Із третім рівнянням усе в порядку і його параметри оцінені без зсувів.

Коли правильно визначена причина хвороби, знаходять способи її лікування. Одним з таких способів є заміна в моделі *пояснюючої* ендогенної змінної на так звану *інструментальну змінну*, яка тісно корелює із заміненою змінною, але не корелює з випадковим членом. Інструментальні змінні **CR**, **IR**, **WR** складаються у вигляді лінійних комбінацій передвизначених змінних, на які випадкові похибки не впливають.

Перенесемо всі ендогенні змінні в ліву частину рівностей і запишемо структурну модель у такий матричний формі:

$$\mathbf{AY} = \mathbf{BZ} + \mathbf{E},$$

де $\mathbf{Y} = (\mathbf{C}, \mathbf{I}, \mathbf{W}, \mathbf{X}, \mathbf{P}, \mathbf{K})$ – матриця значень ендогенних змінних;

$\mathbf{Z} = (\mathbf{X}_0, \mathbf{XLAG}, \mathbf{PLAG}, \mathbf{KLAG}, \mathbf{WP}, \mathbf{G}, \mathbf{T}, t)$ – матриця значень передвизначених змінних; $\mathbf{X}_0 = 1$;

$\mathbf{E} = (u, v, w, 0, 0, 0)$ – випадкові члени кожної рівності;

\mathbf{A} – матриця коефіцієнтів перед ендогенними змінними [6×6];

\mathbf{B} – коефіцієнти перед передвизначеними змінними [6×8].

Виразимо всі ендогенні змінні через передвизначені, для чого помножимо записану вище матричну рівність на \mathbf{A}^{-1} і одержимо модель у так званій *наведеній формі*:

$$\mathbf{CR} = a_0 + a_1 \cdot \mathbf{PLAG} + a_2 \cdot \mathbf{XLAG} + a_3 \cdot \mathbf{KLAG} + a_4 \cdot \mathbf{WP} + a_5 \cdot \mathbf{G} + a_6 \cdot \mathbf{T} + a_7 \cdot t;$$

$$\mathbf{IR} = b_0 + b_1 \cdot \mathbf{PLAG} + b_2 \cdot \mathbf{XLAG} + b_3 \cdot \mathbf{KLAG} + b_4 \cdot \mathbf{WP} + b_5 \cdot \mathbf{G} + b_6 \cdot \mathbf{T} + b_7 \cdot t;$$

$$\mathbf{WR} = c_0 + c_1 \cdot \mathbf{PLAG} + c_2 \cdot \mathbf{XLAG} + c_3 \cdot \mathbf{KLAG} + c_4 \cdot \mathbf{WP} + c_5 \cdot \mathbf{G} + c_6 \cdot \mathbf{T} + c_7 \cdot t.$$

Тут через **CR**, **IR**, **WR** позначені розрахункові значення відповідних ендогенних змінних. Розрахункові значення інших ендогенних змінних визначаються системою тотожностей:

$$\mathbf{XR} = \mathbf{CR} + \mathbf{IR} + \mathbf{G};$$

$$\mathbf{PR} = \mathbf{XR} - \mathbf{WR} - \mathbf{T};$$

$$\mathbf{KR} = \mathbf{KLAG} + \mathbf{IR}.$$

Розрахункові значення найбільше тісно пов'язані з відповідними результативними ознаками **C, I, W, X, P, K** (із коефіцієнтами детермінації 0,938; 0,847; 0,950; 0,918; 0,826; 0,977 – отримайте ці оцінки) і виражені лише через передвизначені змінні. Тому вони є найкращими інструментальними змінними.

Процедура оцінювання тепер природно розпадається на два етапи (2МНК-процедура). На першому етапі оцінюються (звичайним МНК) регресійні рівняння моделі в наведеній формі й обчислюються розрахункові значення всіх ендогенних змінних. На другому етапі оцінюється кожне рівняння структурної моделі з попередньою заміною ендогенних пояснюючих змінних на відповідні розрахункові значення.

Для моделі Клейна на другому кроці 2МНК отримані рівняння:

$$\begin{aligned} C &= 16,55 + 0,017*PR + 0,216*PLAG + 0,810*(WR + WP) + u; \\ I &= 20,28 + 0,150*PR + 0,616*PLAG - 0,160*KLAG + v; \\ W &= 1,500 + 0,439*XR + 0,147*XLAG + 0,130*t + w. \end{aligned}$$

(Зробіть ці розрахунки!)

Третє рівняння практично не змінилося, так і має бути. Характер же відрахувань прибутку в споживання й інвестиції докорінне змінилося – основні відрахуванні тепер припадають на прибуток минулого року. Цікаво, що в сумі всі відрахування склали 100% ($0,017 + 0,216 + 0,150 + 0,616 = 1$).

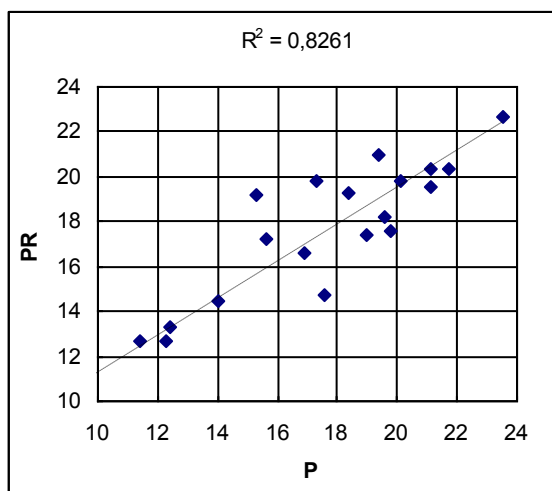


Рис. 48. Графік P – PR

Істотне коректування параметрів структурної моделі відбулося в основному за рахунок заміни *пояснюючої* ендогенної змінної **P** на свою інструментальну змінну **PR** (див. рис. 48). Ці виправлення, строго кажучи, незначні. Дивно, що такі малі виправлення в даних змогли запобігти великим зсувам в оцінках параметрів моделі.

На рис. 49 наведені дані й зразок виконання роботи.

Усі змінні виміряні в млрд дол. за курсом 1931 р.

| Робота № 12. Двоступовий метод найменших квадратів | | | | | | | | | | | | | |
|---|--------------|---|-------------|-------------|-------------|---|-----------------------------|--------------|-------------|-----------|----------|----------|----------|
| Ендогенні змінні: | | C - споживання; I - інвестиції; W – дохід у приватному секторі P - прибуток; X – обсяг виробництва; K - основний капітал | | | | | | | | | | | |
| Лагові ендогенні змінні: | | PLAG, XLAG, KLAG | | | | | | | | | | | |
| Екзогенні змінні: | | WP - зарплата в держсекторі; G - держ.закупівлі; T – податки; t | | | | | | | | | | | |
| Випадкові члени: u, v, w | | WS = W+WP; t = Рік – 1931. | | | | | | | | | | | |
| Модель у структурній формі: | | | | | | Двоступовий МНК | | | | | | | |
| C = a0 + a1*P + a2*PLAG + a3*WS + u | | | | | | C = a0 + a1*PR + a2*PLAG + a3*WRS | | | | | | | |
| I = b0 + b1*P + b2*PLAG + b3*KLAG + v | | | | | | I = b0 + b1*PR + b2*PLAG + b3*KLAG | | | | | | | |
| W = c0 + c1*X + c2*XLAG + c3*t + w | | | | | | W = c0 + c1*XR + c2*XLAG + c3*t | | | | | | | |
| X = C + I + G | | | | | | WRS = WR + WP | | | | | | | |
| P = X - W - T | | | | | | | | | | | | | |
| K = KLAG + I | | | | | | | | | | | | | |
| Модель у наведеній формі: | | | | | | | | | | | | | |
| CR = a0 + a1*PLAG + a2*XLAG + a3*KLAG + a4*WP + a5*G + a6*T + a7*t | | | | | | | | | | | | | |
| IR = b0 + b1*PLAG + b2*XLAG + b3*KLAG + b4*WP + b5*G + b6*T + b7*t | | | | | | | | | | | | | |
| WR = c0 + c1*PLAG + c2*XLAG + c3*KLAG + c4*WP + c5*G + c6*T + c7*t | | | | | | | | | | | | | |
| XR = CR + IR + G | | | | | | | | | | | | | |
| PR = XR - WR - T | | | | | | | | | | | | | |
| KR = KLAG + IR | | | | | | | | | | | | | |
| Ендогенні змінні | | | | | | | Предвизначені змінні | | | | | | |
| Рік | C | I | W | P | X | K | PLAG | XLAG | KLAG | WP | G | T | t |
| 1921 | 41,9 | -0,2 | 25,5 | 12,4 | 45,6 | 182,6 | 12,7 | 44,9 | 182,8 | 2,7 | 3,9 | 7,7 | -10 |
| 1922 | 45,0 | 1,9 | 29,3 | 16,9 | 50,1 | 184,5 | 12,4 | 45,6 | 182,6 | 2,9 | 3,2 | 3,9 | -9 |
| 1923 | 49,2 | 5,2 | 34,1 | 18,4 | 57,2 | 189,7 | 16,9 | 50,1 | 184,5 | 2,9 | 2,8 | 4,7 | -8 |
| 1924 | 50,6 | 3,0 | 33,9 | 19,4 | 57,1 | 192,7 | 18,4 | 57,2 | 189,7 | 3,1 | 3,5 | 3,8 | -7 |
| 1925 | 52,6 | 5,1 | 35,4 | 20,1 | 61,0 | 197,8 | 19,4 | 57,1 | 192,7 | 3,2 | 3,3 | 5,5 | -6 |
| 1926 | 55,1 | 5,6 | 37,4 | 19,6 | 64,0 | 203,4 | 20,1 | 61,0 | 197,8 | 3,3 | 3,3 | 7,0 | -5 |
| 1927 | 56,2 | 4,2 | 37,9 | 19,8 | 64,4 | 207,6 | 19,6 | 64,0 | 203,4 | 3,6 | 4,0 | 6,7 | -4 |
| 1928 | 57,3 | 3,0 | 39,2 | 21,1 | 64,5 | 210,6 | 19,8 | 64,4 | 207,6 | 3,7 | 4,2 | 4,2 | -3 |
| 1929 | 57,8 | 5,1 | 41,3 | 21,7 | 67,0 | 215,7 | 21,1 | 64,5 | 210,6 | 4,0 | 4,1 | 4,0 | -2 |
| 1930 | 55,0 | 1,0 | 37,9 | 15,6 | 61,2 | 216,7 | 21,7 | 67,0 | 215,7 | 4,2 | 5,2 | 7,7 | -1 |
| 1931 | 50,9 | -3,4 | 34,5 | 11,4 | 53,4 | 213,3 | 15,6 | 61,2 | 216,7 | 4,8 | 5,9 | 7,5 | 0 |
| 1932 | 45,6 | -6,2 | 29,0 | 7,0 | 44,3 | 207,1 | 11,4 | 53,4 | 213,3 | 5,3 | 4,9 | 8,3 | 1 |
| 1933 | 46,5 | -5,1 | 28,5 | 11,2 | 45,1 | 202,0 | 7,0 | 44,3 | 207,1 | 5,6 | 3,7 | 5,4 | 2 |
| 1934 | 48,7 | -3,0 | 30,6 | 12,3 | 49,7 | 199,0 | 11,2 | 45,1 | 202,0 | 6,0 | 4,0 | 6,8 | 3 |
| 1935 | 51,3 | -1,3 | 33,2 | 14,0 | 54,4 | 197,7 | 12,3 | 49,7 | 199,0 | 6,1 | 4,4 | 7,2 | 4 |
| 1936 | 57,7 | 2,1 | 36,8 | 17,6 | 62,7 | 199,8 | 14,0 | 54,4 | 197,7 | 7,4 | 2,9 | 8,3 | 5 |
| 1937 | 58,7 | 2,0 | 41,0 | 17,3 | 65,0 | 201,8 | 17,6 | 62,7 | 199,8 | 6,7 | 4,3 | 6,7 | 6 |
| 1938 | 57,5 | -1,9 | 38,2 | 15,3 | 60,9 | 199,9 | 17,3 | 65,0 | 201,8 | 7,7 | 5,3 | 7,4 | 7 |
| 1939 | 61,6 | 1,3 | 41,6 | 19,0 | 69,5 | 201,2 | 15,3 | 60,9 | 199,9 | 7,8 | 6,6 | 8,9 | 8 |
| 1940 | 65,0 | 3,3 | 45,0 | 21,1 | 75,7 | 204,5 | 19,0 | 69,5 | 201,2 | 8,0 | 7,4 | 9,6 | 9 |
| 1941 | 69,7 | 4,9 | 53,3 | 23,5 | 88,4 | 209,4 | 21,1 | 75,7 | 204,5 | 8,5 | 13,8 | 11,6 | 10 |
| | t | T | G | WP | KLAG | XLAG | PLAG | Const | | | | | |
| CR : | 0,701 | -0,37 | 0,205 | 0,193 | -0,15 | 0,23 | 0,748 | 58,3 | | | | | |
| IR : | 0,332 | -0,16 | 0,1 | -0,72 | -0,19 | -0,11 | 0,926 | 35,52 | | | | | |
| WR : | 0,714 | -0,6 | 0,866 | -0,44 | -0,12 | 0,095 | 0,872 | 43,44 | | | | | |
| | Const | PLAG | XLAG | KLAG | WP | G | T | t | | | | | |
| CR : | 58,3 | 0,748 | 0,23 | -0,15 | 0,193 | 0,205 | -0,37 | 0,701 | | | | | |
| IR : | 35,52 | 0,926 | -0,11 | -0,19 | -0,72 | 0,1 | -0,16 | 0,332 | | | | | |
| WR : | 43,44 | 0,872 | 0,095 | -0,12 | -0,44 | 0,866 | -0,6 | 0,714 | | | | | |

Рис. 49. Зразок звіту до лабораторної роботи № 12

Відразу під таблицею даних функцією ЛИНЕЙН обчислені параметри моделі в наведеній формі, де в полі *Известные значения y* – послідовно задані стовпці C, I; W; у полі *Известные значения x* – стовпці передвизначених змінних (PLAG, XLAG, KLAG, WP, G, T, t); у полі *Константа* – 1 (константа потрібна); у полі *Статистика* – 0 (статистика не потрібна). Без статистики висновки функції ЛИНЕЙН займає один рядок (8 клітинок). Нагадуємо, що коефіцієнти регресії функцією ЛИНЕЙН виводяться у зворотному порядку. Для зручності наступних розрахунків нижче вони переставлені в правильному порядку (функцією ГПР). Значення інструментальних змінних CR, IR, WR можна обчислити так:

$$\text{Розрахунок Значення} = \text{Const} + \text{СУММПРОИЗВ}(\text{Параметры}; \text{Известные Значения } x)$$

Обчислені значення CR, IR, WR наведені в таблиці (рис. 50).

| Рік | CR | IR | WR | KR | XR | XLAG | t | WRS | PR | PLAG | KLAG |
|------|-------|-------|-------|-------|-------|------|-----|-------|-------|------|-------|
| 1921 | 42,84 | 0,923 | 26,71 | 183,7 | 47,66 | 44,9 | -10 | 29,41 | 13,26 | 12,7 | 182,8 |
| 1922 | 44,79 | 1,337 | 28,85 | 183,9 | 49,33 | 45,6 | -9 | 31,75 | 16,58 | 12,4 | 182,6 |
| 1923 | 49,24 | 4,795 | 32,85 | 189,3 | 56,84 | 50,1 | -8 | 35,75 | 19,28 | 16,9 | 184,5 |
| 1924 | 52,45 | 4,787 | 35,97 | 194,5 | 60,73 | 57,2 | -7 | 39,07 | 20,96 | 18,4 | 189,7 |
| 1925 | 52,79 | 5,113 | 35,94 | 197,8 | 61,20 | 57,1 | -6 | 39,14 | 19,77 | 19,4 | 192,7 |
| 1926 | 53,63 | 4,358 | 36,05 | 202,2 | 61,29 | 61,0 | -5 | 39,35 | 18,24 | 20,1 | 197,8 |
| 1927 | 54,14 | 2,714 | 36,58 | 206,1 | 60,86 | 64,0 | -4 | 40,18 | 17,57 | 19,6 | 203,4 |
| 1928 | 55,44 | 2,730 | 38,63 | 210,3 | 62,37 | 64,4 | -3 | 42,33 | 19,54 | 19,8 | 207,6 |
| 1929 | 56,81 | 3,484 | 40,02 | 214,1 | 64,40 | 64,5 | -2 | 44,02 | 20,38 | 21,1 | 210,6 |
| 1930 | 56,70 | 2,478 | 39,5 | 218,2 | 64,38 | 67,0 | -1 | 43,70 | 17,18 | 21,7 | 215,7 |
| 1931 | 51,69 | -2,71 | 34,68 | 214,0 | 54,88 | 61,2 | 0 | 39,48 | 12,71 | 15,6 | 216,7 |
| 1932 | 47,55 | -5,32 | 29,83 | 208,0 | 47,13 | 53,4 | 1 | 35,13 | 9,00 | 11,4 | 213,3 |
| 1933 | 44,65 | -6,71 | 27,18 | 200,4 | 41,64 | 44,3 | 2 | 32,78 | 9,05 | 7,0 | 207,1 |
| 1934 | 49,05 | -2,08 | 31,50 | 199,9 | 50,97 | 45,1 | 3 | 37,50 | 12,67 | 11,2 | 202,0 |
| 1935 | 52,03 | -0,77 | 34,04 | 198,2 | 55,66 | 49,7 | 4 | 40,14 | 14,42 | 12,3 | 199,0 |
| 1936 | 54,81 | -0,40 | 34,30 | 197,3 | 57,31 | 54,4 | 5 | 41,70 | 14,71 | 14,0 | 197,7 |
| 1937 | 60,55 | 2,829 | 41,18 | 202,6 | 67,67 | 62,7 | 6 | 47,88 | 19,80 | 17,6 | 199,8 |
| 1938 | 61,40 | 1,509 | 41,60 | 203,3 | 68,21 | 65,0 | 7 | 49,30 | 19,21 | 17,3 | 201,8 |
| 1939 | 59,68 | 0,632 | 40,59 | 200,5 | 66,91 | 60,9 | 8 | 48,39 | 17,42 | 15,3 | 199,9 |
| 1940 | 64,88 | 2,996 | 45,37 | 204,2 | 75,28 | 69,5 | 9 | 53,37 | 20,31 | 19,0 | 201,2 |
| 1941 | 68,77 | 3,899 | 52,22 | 208,4 | 86,47 | 75,7 | 10 | 60,72 | 22,66 | 21,1 | 204,5 |

Рис. 50. Значення інструментальних змінних

Розрахункові значення інших показників обчислені за допомогою тотожностей: $KR = KLAG + IR$, $XR = CR + IR + G$, $PR = XR - WR - T$. Ці розрахунки становлять перший крок 2МНК процедури.

Переходимо до другого кроку 2МНК процедури. Функцією ЛИНЕЙН треба оцінити кожне рівняння структурної моделі, замінюючи в них ендогенні

пояснюючі змінні відповідними розрахунковими значеннями. Оскільки функція ЛИНЕЙН вимагає, щоб стовпці аргументів розташовувалися поруч, після стовпця XR вставлені стовпці XLAG, t, WRS = WR + WP і лише після цього розташовані стовпці PR, PLAG, KLAG. Тепер, задаючи в різних комбінаціях сусідні стовпці змінних, можна функцією ЛИНЕЙН оцінювати кожне рівняння моделі в структурній формі. Для порівняння результатів розрахунку зі звичайною процедурою 1МНК нижче складена аналогічна таблиця з вихідними (не інструментальними) значеннями ендогенних змінних (рис. 51).

| Год | C | I | W | K | X | XLAG | t | WS | P | PLAG | KLAG |
|------|------|------|------|-------|------|------|-----|------|------|------|-------|
| 1921 | 41,9 | -0,2 | 25,5 | 182,6 | 45,6 | 44,9 | -10 | 28,2 | 12,4 | 12,7 | 182,8 |
| 1922 | 45,0 | 1,9 | 29,3 | 184,5 | 50,1 | 45,6 | -9 | 32,2 | 16,9 | 12,4 | 182,6 |
| 1923 | 49,2 | 5,2 | 34,1 | 189,7 | 57,2 | 50,1 | -8 | 37,0 | 18,4 | 16,9 | 184,5 |
| 1924 | 50,6 | 3,0 | 33,9 | 192,7 | 57,1 | 57,2 | -7 | 37,0 | 19,4 | 18,4 | 189,7 |
| 1925 | 52,6 | 5,1 | 35,4 | 197,8 | 61,0 | 57,1 | -6 | 38,6 | 20,1 | 19,4 | 192,7 |
| 1926 | 55,1 | 5,6 | 37,4 | 203,4 | 64,0 | 61,0 | -5 | 40,7 | 19,6 | 20,1 | 197,8 |
| 1927 | 56,2 | 4,2 | 37,9 | 207,6 | 64,4 | 64,0 | -4 | 41,5 | 19,8 | 19,6 | 203,4 |
| 1928 | 57,3 | 3,0 | 39,2 | 210,6 | 64,5 | 64,4 | -3 | 42,9 | 21,1 | 19,8 | 207,6 |
| 1929 | 57,8 | 5,1 | 41,3 | 215,7 | 67,0 | 64,5 | -2 | 45,3 | 21,7 | 21,1 | 210,6 |
| 1930 | 55,0 | 1,0 | 37,9 | 216,7 | 61,2 | 67,0 | -1 | 42,1 | 15,6 | 21,7 | 215,7 |
| 1931 | 50,9 | -3,4 | 34,5 | 213,3 | 53,4 | 61,2 | 0 | 39,3 | 11,4 | 15,6 | 216,7 |
| 1932 | 45,6 | -6,2 | 29,0 | 207,1 | 44,3 | 53,4 | 1 | 34,3 | 7,0 | 11,4 | 213,3 |
| 1933 | 46,5 | -5,1 | 28,5 | 202,0 | 45,1 | 44,3 | 2 | 34,1 | 11,2 | 7,0 | 207,1 |
| 1934 | 48,7 | -3,0 | 30,6 | 199,0 | 49,7 | 45,1 | 3 | 36,6 | 12,3 | 11,2 | 202,0 |
| 1935 | 51,3 | -1,3 | 33,2 | 197,7 | 54,4 | 49,7 | 4 | 39,3 | 14,0 | 12,3 | 199,0 |
| 1936 | 57,7 | 2,1 | 36,8 | 199,8 | 62,7 | 54,4 | 5 | 44,2 | 17,6 | 14,0 | 197,7 |
| 1937 | 58,7 | 2,0 | 41,0 | 201,8 | 65,0 | 62,7 | 6 | 47,7 | 17,3 | 17,6 | 199,8 |
| 1938 | 57,5 | -1,9 | 38,2 | 199,9 | 60,9 | 65,0 | 7 | 45,9 | 15,3 | 17,3 | 201,8 |
| 1939 | 61,6 | 1,3 | 41,6 | 201,2 | 69,5 | 60,9 | 8 | 49,4 | 19,0 | 15,3 | 199,9 |
| 1940 | 65,0 | 3,3 | 45,0 | 204,5 | 75,7 | 69,5 | 9 | 53,0 | 21,1 | 19,0 | 201,2 |
| 1941 | 69,7 | 4,9 | 53,3 | 209,4 | 88,4 | 75,7 | 10 | 61,8 | 23,5 | 21,1 | 204,5 |

| $C = a_0 + a_1 \cdot P + a_2 \cdot PLAG + a_3 \cdot WS + u$ | | | | $I = b_0 + b_1 \cdot P + b_2 \cdot PLAG + b_3 \cdot KLAG + v$ | | | | | | |
|---|-------|-------|-------|---|--|------|-------|-------|-------|-------|
| | PLAG | P | WS | Const | | KLAG | PLAG | P | Const | |
| 1МНК | 0,090 | 0,193 | 0,796 | 16,24 | | 1МНК | -0,11 | 0,333 | 0,48 | 10,13 |
| 2МНК | 0,216 | 0,017 | 0,810 | 16,55 | | 2МНК | -0,16 | 0,616 | 0,15 | 20,28 |
| $W = c_0 + c_1 \cdot X + c_2 \cdot XLAG + c_3 \cdot t + w$ | | | | | | | | | | |
| | t | XLAG | X | Const | | | | | | |
| 1МНК | 0,13 | 0,146 | 0,439 | 1,497 | | | | | | |
| 2МНК | 0,13 | 0,147 | 0,439 | 1,500 | | | | | | |

Рис. 51. Значення ендогенних змінних

Блоки розрахунків за обома методиками (1МНК і 2МНК) розташовані поруч, звідки випикуємо (рис. 52):

| 1МНК | 2МНК |
|---|---|
| $C = 16,24 + 0,193*P + 0,090*PLAG + 0,796*WS$ | $C = 16,55 + 0,017*P + 0,216*PLAG + 0,810*WS$ |
| $I = 10,13 + 0,48*P + 0,333*PLAG - 0,11*KLAG$ | $I = 20,28 + 0,15*P + 0,616*PLAG - 0,16*KLAG$ |
| $W = 1,497 + 0,439*X + 0,146*XLAG + 0,13*t$ | $W = 1,500 + 0,439*X + 0,147*XLAG + 0,13*t$ |

Рис. 52. Порівняльні моделі 1МНК і 2МНК

Усі рівняння значимі в цілому й складаються лише зі значимих членів.

На відміну від стандартних, 2 МНК-оцінки відповідають змісту задачі.

Зверніть увагу, як суттєво змінилися оцінки параметрів перших двох рівнянь, що з'явилося наслідком зовсім незначних виправлень у пояснюючих ендогенних змінних (рис. 53).

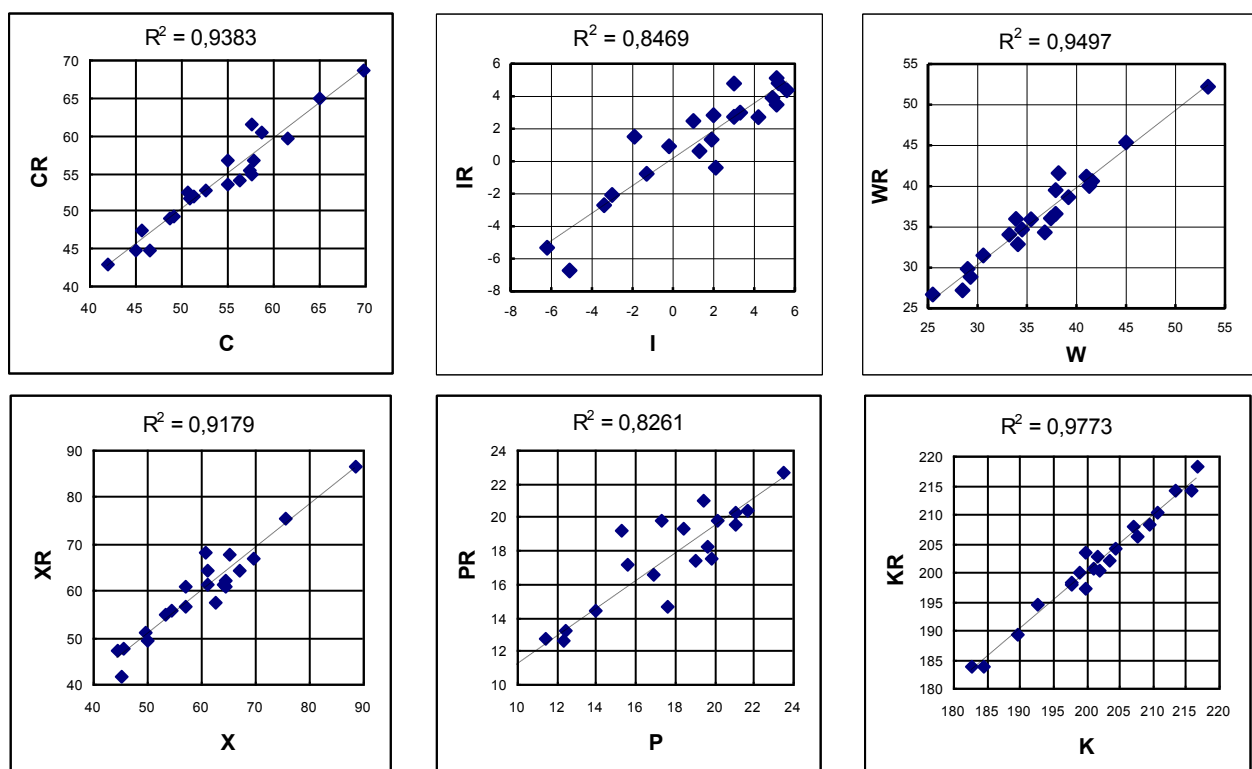


Рис. 53. Графіки спостережуваних і розрахункових значень

Доречі, досить було ввести лише одну інструментальну змінну PR замість ендогенної змінної P, щоб одержати правильні оцінки параметрів структурної моделі.

Завдання для самостійної роботи

Скласти багатофакторну економетричну модель у структурній формі і в наведеній формі. Методом 1МНК оцінити параметри моделі в наведеній формі і отримати вирази ендогенних змінних через передвизначені змінні. Методом 2МНК оцінити параметри економетричної моделі в структурній формі. Порівняти 2МНК-оцінки з 1МНК-оцінками параметрів структурної моделі. Дати змістовну інтерпретацію отриманим результатам.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Що таке "ендогенні" і "екзогенні" змінні?
2. Що таке "лагові змінні"? Що таке "передвизначені змінні"?
3. Опишіть структурну форму економетричної моделі. Звідки з'являються екзогенні змінні?
4. Як складається економетрична модель у наведеній формі? Що можна отримати після оцінки її параметрів?
5. Чому для оцінки параметрів структурної моделі звичайний метод найменших квадратів часто приводить до суттєвого зміщення оцінок? Чи можна передбачити появу цього небажаного ефекту?
6. Що таке "інструментальні змінні"? Чому заміна деяких ендогенних змінних на їх інструментальні змінні усуває зміщення МНК-оцінок параметрів у структурній формі економетричної моделі?
7. Як у 2МНК-методі складаються інструментальні змінні? Чи завжди цей прийом приводить до успіху?
8. Дайте поняття "невизначеної", "визначеної" і "надвизначеної" структурної моделі. Який тип економетричної моделі не можна проаналізувати за допомогою 2МНК-методу?

Лабораторна робота № 13. Гетероскедастичність.

Мета роботи. Ознайомитися з ефектом порушення передумови Гауса – Маркова про рівноточність спостережень $M(\hat{\beta}^2) = \sigma_{\varepsilon}^2$. Освоїти метод зважених найменших квадратів для одержання стабільних оцінок параметрів моделі при порушенні передумови про гомоскедастичність спостережень.

Теоретична частина

Звичайно вважається, що всі спостереження однаково надійні, що ймовірності випадкових похибок однакові для всіх спостережень. Ця властивість називається гомоскедастичністю (у перекладі – “однаковий розкид”). Якщо ж різні спостереження мають різну мінливість, то така властивість називається гетероскедастичністю (у перекладі – “неоднаковий розкид”).

Ця передумова не використовувалася при доказі незміщенності МНК-оцінок параметрів моделі, тому вважається, що при відсутності гомоскедастичності неправильно обчислюються лише стандартні похибки коефіцієнтів регресії, статистики Ст'юдента й довірчі границі на розрахункові значення. Проте теоретична незміщенність ще не означає, що коефіцієнти регресії оцінюються правильно, тому що МНК не робить різниці між спостереженнями, надаючи їм однакову вагу, незалежно від того, чи є спостереження надійним чи ні.

Іноді появу гетероскедастичності можна передбачити заздалегідь. Так при функціональних перетвореннях залежної змінної (результативної ознаки Y) нелінійним чином перетворюється вся система випадкових похибок різних спостережень, що може привести до небажаних наслідків.

Розглянемо умовний приклад. Нехай результативна ознака породжується нелінійною залежністю $Y_t = 1 / (1 + X_t)$, де $X = 0, 1, 2, \dots, 9$. Нам відомі значення Y з невеликими похибками $e = \pm 0,05$, знаки яких чергуються ($Y = Y_t + e$). Чи можливо за цими даними методом найменших квадратів відновити (з достатнім ступенем точності) породжувальну залежність? Форму зв'язку часто вдається вгадати. Якщо на вихідному графіку емпіричні точки не групуються навколо деякої прямої, то роблять пробні функціональні перетворення змінних (найчас-

тіше застосовується логарифмування або перехід до зворотних величин). У нашому прикладі в координатах $(X, 1/Y)$ емпіричні точки явно групуються навколо прямої, тому приймаємо в цих координатах лінійну модель $1/Y = b_0 + b_1 \cdot X$, звідки одержуємо правильну форму зв'язку $Y = 1 / (b_0 + b_1 \cdot X)$. Залишилося визначити оцінки коефіцієнтів регресії, які в породжувальній залежності дорівнюють $\beta_0 = \beta_1 = 1$.

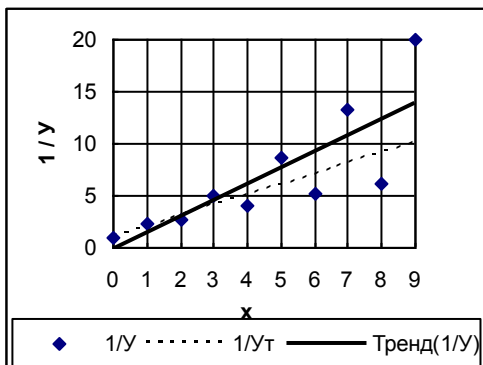


Рис. 54. Залежність у функціональних масштабах

ки одержуємо правильну форму зв'язку $Y = 1 / (b_0 + b_1 \cdot X)$. Залишилося визначити оцінки коефіцієнтів регресії, які в породжувальній залежності дорівнюють $\beta_0 = \beta_1 = 1$. На рис. 54 у функціональних масштабах $(X, 1/Y)$ графік породжувальної залежності зображений пунктиром.

Але з наведеного графіка ясно видно, що після переходу до зворотних величин $1/Y$ з'явився ефект “розбовтування”: розкид точок тепер збільшується зі збільшенням X , причому остання точка навіть схожа на викид. І як усякий викид, ця точка перетягнула до себе всю лінію регресії. Рівняння лінійного тренда виявилось таким: $Y_p = -0,23 + 1,56 \cdot X$. Отримані оцінки $b_0 = -0,23$ і $b_1 = 1,56$ зовсім не схожі на очікувані одиничні значення $\beta_0 = \beta_1 = 1$.

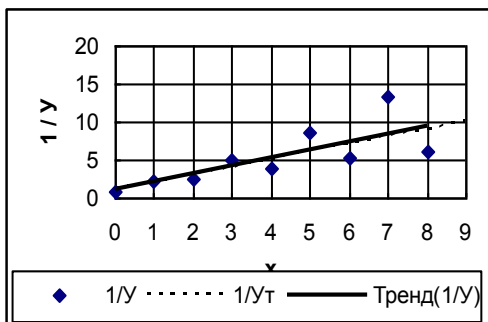


Рис. 55. Зміна графіка після видалення однієї точки

Спробуємо вилучити останню сумнівну точку (рис. 55). Виявляється, після видалення найбільш сумнівної точки роль викиду буде грати передостаннє спостереження. Положення лінії регресії помітно змінилося. Результати звичайного МНК виявилися нестабільними – при послідовних видаленнях даних лінія регресії “виляє хвостом”.

І хоча МНК-оцінки теоретично незміщені, довіряти їм не можна.

Ідея методу зважених найменших квадратів полягає у введенні вагової функції, яка надає більшу вагу надійним спостереженням і меншу – малонадійним. Наприклад, іноді вдається апроксимувати стандартну похибку розкиду вираженням типу $\sigma_i = k \cdot Z_i$ (знайдена якась величина Z_i , яка пропорційна σ_i). Тоді

все рівняння регресії $Y = b_0 + b_1 \cdot X$ слід розділити на Z_i : $Y/Z = b_0/Z + b_1 \cdot X/Z + u$. У цій моделі (без вільного члена) випадковий член u вже не схильний до гетероскедастичності.

У нашому випадку вагова функція визначається наступними міркуваннями. В нелінійній моделі $Y = 1 / (b_0 + b_1 \cdot X) + u$ усі спостереження були рівноточними. Для можливості оцінки параметрів моделі стандартною процедурою МНК потрібно, щоб модель була лінійною *щодо параметрів*; тому переходимо до зворотних величин $1/Y = b_0 + b_1 \cdot X + e$. Помітимо, що тепер випадковий член (e) відноситься не до вихідної змінної Y , а до перетвореної величини $1/Y$. Проте, згідно з основною передумовою регресійного аналізу, випадкова похибка повинна відноситися адитивно саме до вихідної залежної змінної Y і ні до якої іншої. Питання: чи не можна перетворити модель так, щоб вона була лінійною *щодо параметрів* і в той же час була розв'язана відносно Y ? Для цього помножимо обидві частини перетвореного рівняння на $Y \cdot Y$, і в результаті одержимо: $Y = b_0 \cdot Y \cdot Y + b_1 \cdot X \cdot Y \cdot Y + u$. У цій формі моделі задоволені відразу обидві умови – модель розв'язана відносно Y і лінійна *щодо параметрів* b_0, b_1 . Оцінюємо параметри цієї моделі за всіма 10-ю спостереженнями $Y = 0,963 \cdot Y \cdot Y + 0,868 \cdot X \cdot Y \cdot Y$ і за 9-ю спостереженнями $Y = 0,963 \cdot Y \cdot Y + 0,866 \cdot X \cdot Y \cdot Y$ (без сумнівної останньої точки). Як бачимо, тепер результати виявилися стійкими до вибракування декількох сумнівних спостережень. Отримані оцінки коефіцієнтів регресії дуже близькі до очікуваних одиничних значень породжувальній залежності $\beta_0 = \beta_1 = 1$.

У загальному випадку нелінійну модель намагаються функціональними перетвореннями змінних привести до вигляду

$$F(x,y) = b_0 + b_1 \cdot \Phi(x,y) + e$$

так, щоб вона була лінійною *щодо параметрів*. При цьому часто з'являється гетероскедастичність (якщо її не було раніше). Виявляється, випадкові похибки (e) будуть пропорційні величині $z = |F'_y|$. Для нейтралізації гетероскедастичності слід розділити обидві частини отриманого лінійного рівняння (у координатах (F, Φ)) на z . У нашому прикладі $F(x,y) = 1/y$. Похідна тут дорівнює

$F'(y) = -1/y^2$, звідки $1/z = y \cdot y$. Цей результат ми одержали вище з інших міркувань.

У спеціалізованих статистичних пакетах передбачений облік вагової функції, яка визначається як $g = 1/z^2$. Для нашого прикладу на запит про вагову функцію слід указати $g = y^4$. Тоді всі формули для розрахунку параметрів моделі формально залишаються тими самими, але при обчисленні середніх треба використовувати не середні арифметичні, а “середні зважені”, наприклад, $Y_{\text{ср}}$ слід обчислювати за формулою: $Y_{\text{ср}} = \text{СУММ}(g \cdot Y) / \text{СУММ}(g)$. При $g = 1$ виходить звичайне середнє: $Y_{\text{ср}} = \text{СУММ}(Y) / n$.

На рис. 56 наведений зразок виконання роботи.

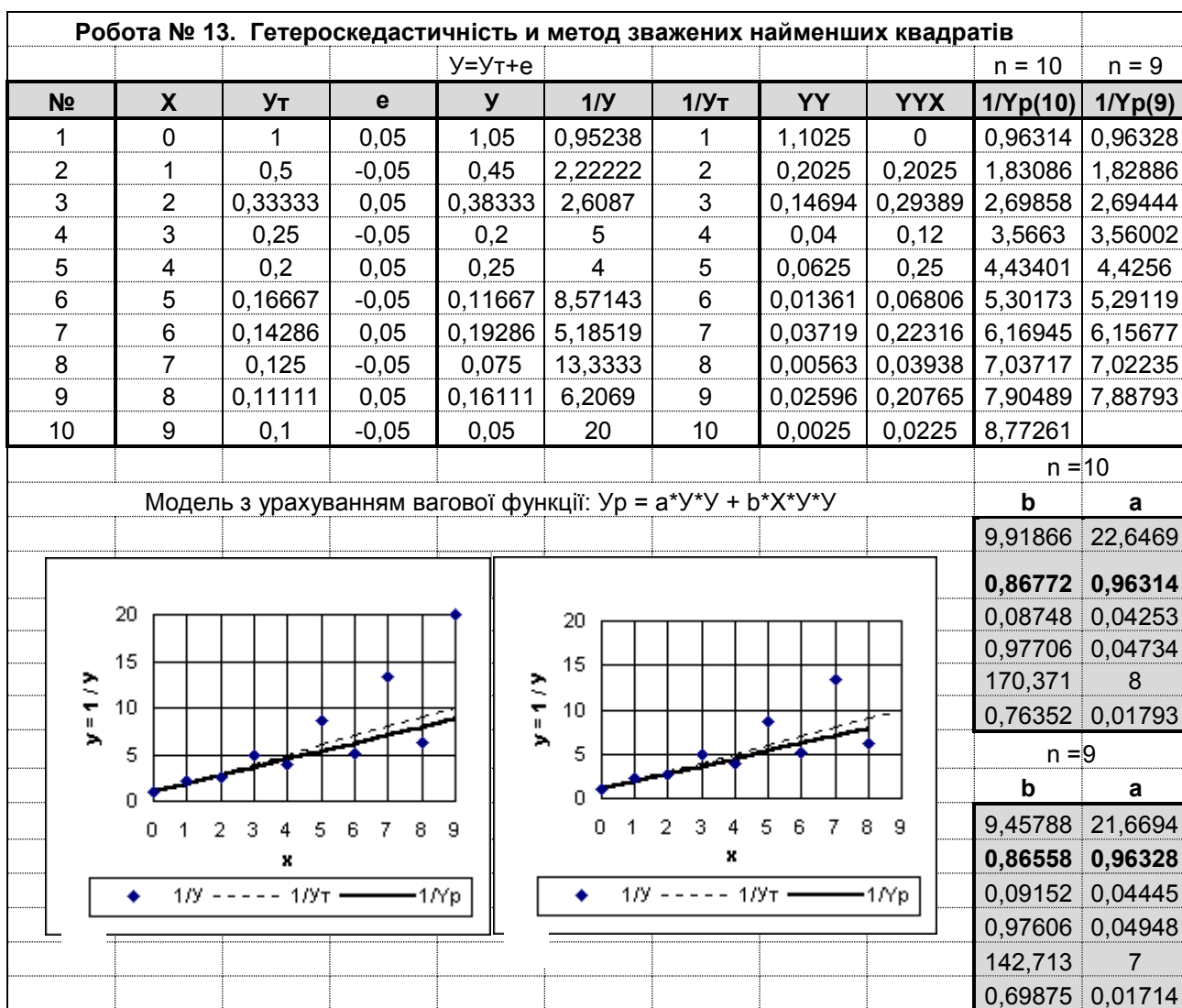


Рис. 56. Зразок звіту до лабораторної роботи № 13

Завдання для самостійної роботи

Для даних з явною гетероскедастичністю оцініть параметри регресійної моделі звичайним МНК і зваженим МНК. Порівняйте результати розрахунків і представте графічно особливості розкиду даних.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Перелічіть основні передумови регресійного аналізу (гіпотези Гауса – Маркова. Опишіть можливі наслідки порушення кожної з них.
2. Що таке "гомоскедастичність" і "гетероскедастичність"?
3. Що таке "вагова функція", яке її призначення?
4. У чому полягає "метод зважених найменших квадратів"?
5. Що таке "зважені середні"?
6. Чому при функціональних перетвореннях результативної ознаки часто з'являється гетероскедастичність перетворених даних? Які передумови аналізу при цьому порушуються?
7. Як визначається вагова функція при функціональних перетвореннях результативної ознаки?

Лабораторна робота № 14. Автокореляція залишків моделі.

Мета роботи: Ознайомитися з ефектом автокореляції й стандартними прийомами нейтралізації шкідливих наслідків порушення гіпотези Гауса – Маркова про некорельованість залишків моделі $M(\varepsilon_i \varepsilon_j) = 0$.

Теоретична частина

У переліку гіпотез Гауса – Маркова, розташованих у порядку убутання важливості, передумова про некорельованість окремих спостережень займає передостаннє місце. Тому порушення даної гіпотези звичайно не викликає особливих побоювань; вважається, що в цьому випадку всі коефіцієнти регресії залишаються незміщеними, неправильно оцінюються лише їх стандартні відхилення (і довірча смуга на лінію регресії).

На жаль, це не зовсім так. Дуже часто автокореляція залишків з'являється як наслідок порушення інших більш важливих передумов аналізу й може служити індикатором серйозних дефектів моделі. Так, якщо в авторегресійній моделі для опису витрат на житло Y від рівня прибутку X і відносної ціни P

$$\ln Y = 0,50 + 0,15 \ln X - 0,16 \ln P + 0,84 \ln Y_{LAG} + u; \quad R^2 = 0,9996$$

(tb) (1,3) (3,1) (2,4) (22,4)

пропустити член $\ln Y_{LAG}$ (Y_{LAG} – прибуток минулого року), то буде одержано таке рівняння регресії:

$$\ln Y = -1,51 + 1,18 \ln X - 0,35 \ln P + e; \quad R^2 = 0,9908$$

(tb) (0,8) (21,0) (1,1)

(Отримайте ці результати).

Останнє рівняння регресії значиме у цілому й пояснює 99 % мінливості $\ln Y$; незначущий за Ст'юдентом член $-0,35 \ln P$ з моделі не вибраковується, тому що при цьому збільшується MSE. На першій погляд тут усе нормально. Проте відомо, що пропуск важливого члена в моделі – дуже серйозна помилка. Дійсно, порівнюючи обидві моделі, відзначаємо істотні зсуви коефіцієнтів регресії в другому рівнянні. Сумнів викликає також коефіцієнт еластичності за прибутком, оцінка якого вийшла більше одиниці (1,18). Але як помітити дефекти моделі на практиці, якщо ми навіть не підозрюємо про її неправильну специфіка-

цію? Саме тому рекомендується в обов'язковому порядку завершувати дослідження аналізом залишків моделі $e = \ln Y - \ln Y_p$.

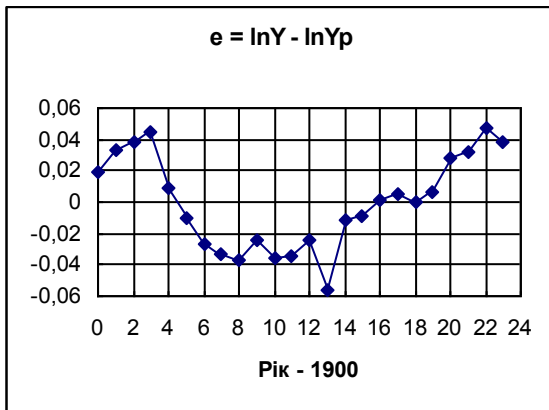


Рис. 57. Графік залишків моделі з пропущеним членом

На графіку цих залишків (рис. 57) явно видна відсутність випадковості в розкиду залишків і спостерігається ефект, дуже схожий на наявність сезонних коливань. Обчислимо (функцією КОРРЕЛ) коефіцієнт автокореляції $\rho = R(e, eLAG) = 0,83$. Тут і далі LAG означає зсув назад на один крок. У спеціалізованих статистичних пакетах замість ρ звичайно обчислюється

статистика Дарбина – Уотсона, яка пов'язана з коефіцієнтом автокореляції простим співвідношенням: $DW \approx 2 \cdot (1 - \rho) = 0,33$. Відсутності автокореляції відповідають значення $DW \approx 2$. Проте у середовищі Excel легше безпосередньо визначити ρ .

Дуже високий коефіцієнт автокореляції $\rho = 0,83$ сигналізує, що з моделлю не все в порядку й треба приймати якісь заходи. Найкраще, звичайно, виправити модель, включивши в неї пропущені члени (якщо ми маємо достатні знання в досліджуваній проблемі). Якщо ж ми не знаємо правильної специфікації, то залишається тільки застосовувати стандартні прийоми для нейтралізації небажаних наслідків автокореляції (тобто будемо лікувати симптоми, а не саму хворобу).

Звичайно передбачається автокореляція 1-го порядку:

$$e = \rho \cdot eLAG + u,$$

де залишки u вже не схильні до автокореляції.

Запишемо модель

$$\ln Y = a + b \cdot \ln X + c \cdot \ln P + e$$

для попереднього періоду

$$\ln YLAG = a + b \cdot \ln XLAG + c \cdot \ln PLAG + eLAG.$$

Помножимо цю другу рівність на ρ і віднімемо з першої. Одержимо:

$$\ln Y - \rho \ln YLAG = a(1 - \rho) + b(\ln X - \rho \ln XLAG) + c(\ln P - \rho \ln PLAG) + u.$$

У цьому запису моделі з'явився ще один невизначений параметр ρ , який на відміну від параметрів b і c присутній у моделі нелінійним чином. Рекомендується оцінку параметрів отриманої нелінійної моделі проводити або послідовними наближеннями, або ж перебором значень параметра ρ з якимсь кроком з інтервалу $[-1; 1]$. Обидва ці прийоми досить працемісткі і немає гарантії, що наприкінці буде одержане стабільне рішення.

Тому в перетвореній нелінійній моделі розкриємо дужки й перепишемо рівняння регресії у такій загальній формі:

$$\ln Y = a_1 + b \ln X + b_1 \ln XLAG + c \ln P + c_1 \ln PLAG + \rho \ln YLAG + u,$$

де коефіцієнти b, b_1 і c, c_1 зв'язані один з одним і параметром ρ :

$$b_1 = \rho b; \quad c_1 = \rho c.$$

Не будемо враховувати цих зв'язків. Тоді формально одержимо лінійну узагальнену модель (оцініть її параметри):

$$\ln Y_p = 0,74 + 0,22 \ln X - 0,11 \ln XLAG - 0,20 \ln P + 0,01 \ln PLAG + 0,87 \ln YLAG.$$

(tb) (1,5) (2,5) (0,9) (1,4) (0,1) (14,0)

Послідовно вибраковуємо незначущі за Ст'юдентом члени й одержуємо правильно специфіковану авторегресійну модель:

$$\ln Y = 0,50 + 0,15 \ln X - 0,16 \ln P + 0,84 \ln YLAG + u; \quad R^2 = 0,9996$$

(tb) (1,3) (3,1) (2,4) (22,4)



Рис. 58. Графік залишків авторегресійної моделі

Автокореляції залишків тепер немає: $\rho = R(u, uLAG) = 0,03$. Побудуйте новий графік залишків (рис. 58).

Успіх у даному прикладі був досягнутий тому, що в узагальненій лінійній моделі з'явився пропущений раніше член $\ln YLAG$.

Зразок виконання роботи наведено на рис. 59.

| Робота №14. Автокореляція залишків моделі | | | | | | | | | | | |
|--|---------------|---------------|---------------|---|---------------|--|---------------|-----------------------------|---------------|--------------|-------|
| Y - витрати на житлові послуги в млрд долл. США | | | | | | | | | | | |
| X - особистий дохід в млрд долл. США | | | | | | | | | | | |
| P - індекс реальних цін на житло в США; Для 1972 г індекс цін P=100 | | | | | | | | | | | |
| Рік | Y | X | P | lnY | lnYLAG | lnX | lnP | lnXLAG | lnPLAG | lnYp | e |
| 1960 | 64,0 | 489,7 | 104,5 | 4,159 | 4,109 | 6,194 | 4,649 | 6,173 | 4,649 | 4,140 | 0,02 |
| 1961 | 67,0 | 503,8 | 105,1 | 4,205 | 4,159 | 6,222 | 4,655 | 6,194 | 4,649 | 4,171 | 0,03 |
| 1962 | 70,7 | 524,9 | 105,0 | 4,258 | 4,205 | 6,263 | 4,654 | 6,222 | 4,655 | 4,220 | 0,04 |
| 1963 | 74,0 | 542,3 | 104,8 | 4,304 | 4,258 | 6,296 | 4,652 | 6,263 | 4,654 | 4,259 | 0,04 |
| 1964 | 77,4 | 580,8 | 104,5 | 4,349 | 4,304 | 6,364 | 4,649 | 6,296 | 4,652 | 4,341 | 0,01 |
| 1965 | 81,6 | 616,3 | 104,0 | 4,402 | 4,349 | 6,424 | 4,644 | 6,364 | 4,649 | 4,412 | -0,01 |
| 1966 | 85,3 | 646,8 | 102,6 | 4,446 | 4,402 | 6,472 | 4,631 | 6,424 | 4,644 | 4,474 | -0,03 |
| 1967 | 89,1 | 673,5 | 102,2 | 4,490 | 4,446 | 6,512 | 4,627 | 6,472 | 4,631 | 4,523 | -0,03 |
| 1968 | 93,5 | 701,3 | 100,9 | 4,538 | 4,490 | 6,553 | 4,614 | 6,512 | 4,627 | 4,575 | -0,04 |
| 1969 | 98,4 | 722,5 | 100,0 | 4,589 | 4,538 | 6,583 | 4,605 | 6,553 | 4,614 | 4,613 | -0,02 |
| 1970 | 102,0 | 751,6 | 99,6 | 4,625 | 4,589 | 6,622 | 4,601 | 6,583 | 4,605 | 4,661 | -0,04 |
| 1971 | 106,4 | 779,2 | 100,0 | 4,667 | 4,625 | 6,658 | 4,605 | 6,622 | 4,601 | 4,702 | -0,04 |
| 1972 | 112,5 | 810,3 | 100,0 | 4,723 | 4,667 | 6,697 | 4,605 | 6,658 | 4,605 | 4,748 | -0,02 |
| 1973 | 118,2 | 865,3 | 99,1 | 4,772 | 4,723 | 6,763 | 4,596 | 6,697 | 4,605 | 4,828 | -0,06 |
| 1974 | 124,2 | 858,4 | 95,1 | 4,822 | 4,772 | 6,755 | 4,555 | 6,763 | 4,596 | 4,833 | -0,01 |
| 1975 | 128,3 | 875,8 | 93,3 | 4,854 | 4,822 | 6,775 | 4,536 | 6,755 | 4,555 | 4,864 | -0,01 |
| 1976 | 134,9 | 906,8 | 93,7 | 4,905 | 4,854 | 6,810 | 4,540 | 6,775 | 4,536 | 4,903 | 0,00 |
| 1977 | 141,3 | 942,9 | 94,5 | 4,951 | 4,905 | 6,849 | 4,549 | 6,810 | 4,540 | 4,946 | 0,01 |
| 1978 | 148,5 | 988,8 | 94,7 | 5,001 | 4,951 | 6,896 | 4,551 | 6,849 | 4,549 | 5,001 | -0,00 |
| 1979 | 154,8 | 1015,5 | 93,8 | 5,042 | 5,001 | 6,923 | 4,541 | 6,896 | 4,551 | 5,036 | 0,01 |
| 1980 | 159,8 | 1021,6 | 93,0 | 5,074 | 5,042 | 6,929 | 4,533 | 6,923 | 4,541 | 5,046 | 0,02 |
| 1981 | 164,8 | 1049,3 | 94,2 | 5,105 | 5,074 | 6,956 | 4,545 | 6,929 | 4,533 | 5,073 | 0,03 |
| 1982 | 167,5 | 1058,3 | 96,7 | 5,121 | 5,105 | 6,964 | 4,572 | 6,956 | 4,545 | 5,074 | 0,05 |
| 1983 | 171,3 | 1095,4 | 99,2 | 5,143 | 5,121 | 6,999 | 4,597 | 6,964 | 4,572 | 5,105 | 0,04 |
| $\ln Y = a + r \cdot \ln YLAG + b \cdot \ln X + c \cdot \ln P$ | | | | $\ln Y = a + b \cdot \ln X + c \cdot \ln P$ | | | | $\ln Y = a + b \cdot \ln X$ | | | |
| lnP | lnX | lnYLAG | Const | lnP | lnX | Const | lnX | Const | | | |
| -2,371 | 3,1385 | 22,363 | 1,3121 | -1,072 | 20,975 | -0,816 | 47,528 | -20,24 | | | |
| -0,158 | 0,1485 | 0,8451 | 0,5015 | -0,352 | 1,1763 | -1,509 | 1,2297 | -3,482 | | | |
| 0,0666 | 0,0473 | 0,0378 | 0,3822 | 0,3285 | 0,0561 | 1,8488 | 0,0259 | 0,172 | | | |
| 0,9996 | 0,0063 | #Н/Д | #Н/Д | 0,9909 | 0,0316 | #Н/Д | 0,9904 | 0,0317 | | | |
| 18951 | 20 | #Н/Д | #Н/Д | 1137,6 | 21 | #Н/Д | 2258,9 | 22 | | | |
| 2,2865 | 0,0008 | #Н/Д | #Н/Д | 2,2664 | 0,0209 | #Н/Д | 2,2652 | 0,0221 | | | |
| $\ln Y = a + r \cdot \ln YLAG + b \cdot \ln X + c \cdot \ln P + b1 \cdot \ln XLAG + c1 \cdot \ln PLAG$ | | | | | | $\ln Y = a + r \cdot \ln YLAG + b \cdot \ln X + c \cdot \ln P + b1 \cdot \ln XLAG$ | | | | | |
| lnPLAG | lnXLAG | lnP | lnX | lnYLAG | Const | lnXLAG | lnP | lnX | lnYLAG | Const | |
| 0,0501 | -0,932 | -1,433 | 2,5394 | 13,995 | 1,5233 | -1,034 | -2,59 | 2,6103 | 19,944 | 1,6631 | |
| 0,0083 | -0,106 | -0,196 | 0,2216 | 0,8697 | 0,741 | -0,104 | -0,191 | 0,2214 | 0,8675 | 0,749 | |
| 0,1658 | 0,1138 | 0,137 | 0,0873 | 0,0621 | 0,487 | 0,1002 | 0,0736 | 0,0848 | 0,0435 | 0,450 | |
| 0,9997 | 0,0065 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | 0,9997 | 0,0063 | #Н/Д | #Н/Д | #Н/Д | |
| 10811 | 18 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | 14263 | 19 | #Н/Д | #Н/Д | #Н/Д | |
| 2,2865 | 0,0008 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | 2,2865 | 0,0008 | #Н/Д | #Н/Д | #Н/Д | |

Рис. 59. Зразок звіту до лабораторної роботи № 14

Функцією ЛИНЕЙН зроблені розрахунки параметрів і статистик вихідної авторегресійної моделі, із пропущеним авторегресійним членом, з вибракун-

ням незначущого за Ст'юдентом члена $\ln P$, узагальненої лінійної моделі й узагальненої моделі після вибракування найменш значимого члена $PLAG$ (вibraкування ще одного незначущого члена $XLAG$ приводить до авторегресійної моделі, параметри якої вже розраховані). Усі блоки висновків функції ЛИНЕЙН доповнені зверху рядком статистик Ст'юдента tb .

Ми розглянули одну з можливих причин появи позитивної автокореляції залишків – помилки специфікації моделі у вигляді пропуску важливих пояснюючих змінних.

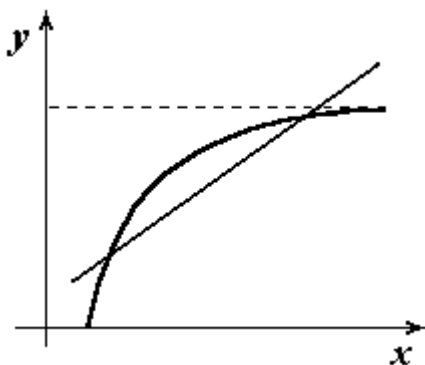


Рис. 60. Лінійаризація нелінійної залежності

Позитивна автокореляція залишків звичайно з'являється при неправильному виборі форми зв'язку, наприклад, при спробі описати нелінійну залежність більш простою лінійною моделлю. На рис. 60 зображена така апроксимація нелінійної залежності $Y = a + b/X$. У цьому випадку знаки залишків будуть однаковими для декількох послідовних спостережень, далі змінюються на протилежні для ряду декількох наступних спостережень, після чого знову змінюються.

Таку поведінку залишків не можна визнати випадковою. До речі, вибір неправильної форми зв'язку також є помилкою специфікації моделі.

Нарешті, при використанні часових рядів як вихідних даних зі зменшенням періоду відбору спостережень часто з'являється ефект сезонних коливань (див., наприклад, лабораторну роботу № 10 "Димту-змінні", коли при переході від середньорічних даних до квартальних з'явилися істотні сезонні коливання). Якщо ефектом сезонних коливань знехтувати не можна, то цей ефект треба врахувати в моделі, іншими словами, знову-таки слід уточнити специфікацію моделі.

Таким чином, можна сказати, що загальною причиною появи позитивної автокореляції залишків є ті або інші помилки специфікації моделі. Що стосується від'ємної автокореляції, то вона проявляється в тенденції до строгого чер-

гування знаків залишків. Нічого страшного в цьому випадку немає, хоча буде сумнівним доказ формул для розрахунку дисперсій і коваріацій коефіцієнтів регресії.

Завдання для самостійної роботи

Розглянути регресійну модель з суттєвою автокореляцією залишків. Проаналізувати щільність і значущість автокореляційного зв'язку і застосувати прийоми усунення автокореляції. Порівняти одержані результати.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Перелічіть основні передумови регресійного аналізу (гіпотези Гауса – Маркова. Опишіть можливі наслідки порушення кожної з них.
2. Що таке "автокореляція", як вимірюється її щільність і значущість?
3. Сформулюйте критерій Дарбіна – Уотсона.
4. Які наслідки від'ємної автокореляції, у чому вона проявляється?
5. Які наслідки позитивної автокореляції? Яка найвірогідніша причина появи позитивної автокореляції залишків моделі?
6. Які висновки регресійного аналізу будуть сумнівними при наявності суттєвої позитивної автокореляції залишків моделі?
7. Які є рекомендації для усунення автокореляції 1-го порядку?
8. Опишіть особливості модифікації цих рекомендацій, що запропоновані в цій лабораторній роботі.

Тема 5. Динамічні економетричні моделі

Лабораторна робота № 15. Моделі з лаговими змінними.

Мета роботи. Скласти модель із розподіленими лагами для опису витрат на житло залежно від рівня доходів і відносних цін поточного й декількох попередніх періодів; перетворити цю модель в авторегресійну методом Койка, оцінити її параметри й зробити висновки відносно коротко й довгострокового впливу пояснюючих змінних.

Теоретична частина

Моделі, що враховують лагові *ендогенні* змінні, прийнято називати *авторегресійними*; якщо ж вони включають лише лагові *екзогенні* змінні, їх іменують як *моделі з розподіленими лагами*. Цей вид моделей застосовується для опису економічних процесів, що мають тривалий характер.

У попередньому завданні була розглянута авторегресійна модель для опису витрат на житло Y залежно від доходу X і відносної ціни P . До цієї моделі привели наступні міркування. Очевидно, що попит на деякий товар визначається не лише поточним доходом, але також доходом за попередні роки. За даними витрат на житло в США (дані наведені нижче) можна скласти наступні моделі (оцініть параметри цих моделей):

$$\begin{array}{ll} \text{a)} & \ln Y_p = -1,51 + 1,18 \cdot \ln X - 0,35 \cdot \ln P ; & R^2 = 0,991 \\ & (t_b) \quad (0,8) \quad (21,0) \quad (1,1) \\ \text{a1)} & \ln Y_p = 0,86 + 1,09 \cdot \ln X_{LAG} - 0,73 \cdot \ln P_{LAG} ; & R^2 = 0,993 \\ & (t_b) \quad (0,5) \quad (19,4) \quad (2,2) \\ \text{a2)} & \ln Y_p = 2,31 + 1,02 \cdot \ln X_{LAG2} - 0,95 \cdot \ln P_{LAG2} ; & R^2 = 0,996 \\ & (t_b) \quad (1,4) \quad (20,6) \quad (3,3) \end{array}$$

Тут LAG і $LAG2$ означають зсуви назад на один і два періоди.

Відзначимо, що між рівняннями a, a1, a2 практично немає особливих відмінностей; значення еластичностей витрат за доходом майже однакові; усі рівняння значимі в цілому й мають майже однакові коефіцієнти детермінації. Немає ніяких підстав для переваги однієї із цих моделей. Це не випадково, тому що лагові змінні тісно корелюють: $R(X, X_{LAG}) = 0,997$; $R(P, P_{LAG}) = 0,951$. (Одержіть ці значення). Тому коефіцієнти регресії в кожному із цих рівнянь фа-

критично відбивають сумарні ефекти за кілька часових періодів. Правильна модель повинна враховувати ефекти кожного періоду окремо, таким чином, ми приходимо до моделі з розподіленими лагами.

Будуємо послідовно моделі з обліком екзогенних лагових змінних за попередні періоди (зробіть ці розрахунки):

$$\text{a) } \ln Y_p = -1,51 + 1,18 \cdot \ln X - 0,35 \cdot \ln P$$

(tb) (0,8) (21,0) (1,1)

$$\text{б) } \ln Y_p = 0,17 + 0,22 \cdot \ln X + 1,00 \cdot \ln P + 0,88 \cdot \ln XLAG - 1,60 \cdot \ln PLAG$$

(tb) (0,1) (0,8) (2,8) (2,9) (4,0)

$$\text{в) } \ln Y_p = 2,14 + 0,31 \cdot \ln X - 0,16 \cdot \ln P + 0,18 \cdot \ln XLAG - 0,65 \cdot \ln PLAG + 0,55 \cdot \ln XLAG^2 - 1,44 \cdot \ln PLAG^2$$

(tb) (1,2) (1,2) (0,3) (0,4) (0,7) (1,6) (2,5)

Усі три моделі значимі в цілому й мають приблизно однаковий коефіцієнт детермінації $R^2 = 0,991 \div 0,997$. Проте зі збільшенням числа членів коефіцієнти регресії стають усе менш значимими й усе більш нестабільними. Цей неприємний ефект пояснюється мультиколінеарністю – тісними кореляційними зв'язками між лаговими змінними.

Для того, щоб скоротити число пояснюючих змінних і тим самим усунути (або хоча б зменшити) ефект мультиколінеарності, Л. Койк запропонував прийняти гіпотезу про експоненціальне убавання коефіцієнтів регресії при послідовних лагових змінних – коефіцієнти регресії (ваги) має убавати в геометричній прогресії.

Запишемо модель

$$\ln Y = a + (b \cdot \ln X + c \cdot \ln P) + (b_1 \cdot \ln XLAG + c_1 \cdot \ln PLAG) + (b_2 \cdot \ln XLAG^2 + c_2 \cdot \ln PLAG^2) + \dots + e$$

для попереднього періоду

$$\ln YLAG = a + (b \cdot \ln XLAG + c \cdot \ln PLAG) + (b_1 \cdot \ln XLAG^2 + c_1 \cdot \ln PLAG^2) + \dots + eLAG$$

і врахуємо, що послідовні ваги убавують у геометричній прогресії:

$$\begin{aligned} b_1 &= b \cdot q; & b_2 &= b_1 \cdot q; & b_3 &= b_2 \cdot q; & \dots \\ c_1 &= c \cdot q; & c_2 &= c_1 \cdot q; & c_3 &= c_2 \cdot q; & \dots \end{aligned}$$

Помножимо 2-у рівність на q і віднімемо з 1-ї.

У результаті одержимо економну авторегресійну модель:

$$\ln Y = a_1 + b \cdot \ln X + c \cdot \ln P + q \cdot \ln YLAG + u,$$

де $a_1 = a \cdot (1 - q)$; $u = e - q \cdot eLAG$ (u – новий випадковий член).

Нижче на рис. 61 наведений зразок виконання цієї роботи.

| Робота № 15. Моделі з розподіленими лагами | | | | | | | | | | | |
|--|---------------|---------------|---------------|--------|---------------|---------------|---------------|--------|---------------|---------------|---------------|
| Y - витрати на житлові послуги в млрд долл. США | | | | | | | | | | | |
| X - особистий дохід в млрд долл. США | | | | | | | | | | | |
| P - індекс реальних цін на житло в США; Для 1972 р індекс цін P=100 | | | | | | | | | | | |
| Рік | Y | X | P | lnY | lnYLAG | lnX | lnP | lnXLAG | lnPLAG | lnXLAG2 | lnPLAG2 |
| 1960 | 64,0 | 489,7 | 104,5 | 4,1589 | 4,1092 | 6,1938 | 4,6492 | 6,1732 | 4,6492 | 6,1327 | 4,6501 |
| 1961 | 67,0 | 503,8 | 105,1 | 4,2047 | 4,1589 | 6,2222 | 4,6549 | 6,1938 | 4,6492 | 6,1732 | 4,6492 |
| 1962 | 70,7 | 524,9 | 105,0 | 4,2584 | 4,2047 | 6,2632 | 4,6540 | 6,2222 | 4,6549 | 6,1938 | 4,6492 |
| 1963 | 74,0 | 542,3 | 104,8 | 4,3041 | 4,2584 | 6,2958 | 4,6521 | 6,2632 | 4,6540 | 6,2222 | 4,6549 |
| 1964 | 77,4 | 580,8 | 104,5 | 4,3490 | 4,3041 | 6,3644 | 4,6492 | 6,2958 | 4,6521 | 6,2632 | 4,6540 |
| 1965 | 81,6 | 616,3 | 104,0 | 4,4018 | 4,3490 | 6,4237 | 4,6444 | 6,3644 | 4,6492 | 6,2958 | 4,6521 |
| 1966 | 85,3 | 646,8 | 102,6 | 4,4462 | 4,4018 | 6,4720 | 4,6308 | 6,4237 | 4,6444 | 6,3644 | 4,6492 |
| 1967 | 89,1 | 673,5 | 102,2 | 4,4898 | 4,4462 | 6,5125 | 4,6269 | 6,4720 | 4,6308 | 6,4237 | 4,6444 |
| 1968 | 93,5 | 701,3 | 100,9 | 4,5380 | 4,4898 | 6,5529 | 4,6141 | 6,5125 | 4,6269 | 6,4720 | 4,6308 |
| 1969 | 98,4 | 722,5 | 100,0 | 4,5890 | 4,5380 | 6,5827 | 4,6052 | 6,5529 | 4,6141 | 6,5125 | 4,6269 |
| 1970 | 102,0 | 751,6 | 99,6 | 4,6250 | 4,5890 | 6,6222 | 4,6012 | 6,5827 | 4,6052 | 6,5529 | 4,6141 |
| 1971 | 106,4 | 779,2 | 100,0 | 4,6672 | 4,6250 | 6,6583 | 4,6052 | 6,6222 | 4,6012 | 6,5827 | 4,6052 |
| 1972 | 112,5 | 810,3 | 100,0 | 4,7230 | 4,6672 | 6,6974 | 4,6052 | 6,6583 | 4,6052 | 6,6222 | 4,6012 |
| 1973 | 118,2 | 865,3 | 99,1 | 4,7724 | 4,7230 | 6,7631 | 4,5961 | 6,6974 | 4,6052 | 6,6583 | 4,6052 |
| 1974 | 124,2 | 858,4 | 95,1 | 4,8219 | 4,7724 | 6,7551 | 4,5549 | 6,7631 | 4,5961 | 6,6974 | 4,6052 |
| 1975 | 128,3 | 875,8 | 93,3 | 4,8544 | 4,8219 | 6,7751 | 4,5358 | 6,7551 | 4,5549 | 6,7631 | 4,5961 |
| 1976 | 134,9 | 906,8 | 93,7 | 4,9045 | 4,8544 | 6,8099 | 4,5401 | 6,7751 | 4,5358 | 6,7551 | 4,5549 |
| 1977 | 141,3 | 942,9 | 94,5 | 4,9509 | 4,9045 | 6,8490 | 4,5486 | 6,8099 | 4,5401 | 6,7751 | 4,5358 |
| 1978 | 148,5 | 988,8 | 94,7 | 5,0006 | 4,9509 | 6,8965 | 4,5507 | 6,8490 | 4,5486 | 6,8099 | 4,5401 |
| 1979 | 154,8 | 1015,5 | 93,8 | 5,0421 | 5,0006 | 6,9231 | 4,5412 | 6,8965 | 4,5507 | 6,8490 | 4,5486 |
| 1980 | 159,8 | 1021,6 | 93,0 | 5,0739 | 5,0421 | 6,9291 | 4,5326 | 6,9231 | 4,5412 | 6,8965 | 4,5507 |
| 1981 | 164,8 | 1049,3 | 94,2 | 5,1047 | 5,0739 | 6,9559 | 4,5454 | 6,9291 | 4,5326 | 6,9231 | 4,5412 |
| 1982 | 167,5 | 1058,3 | 96,7 | 5,1210 | 5,1047 | 6,9644 | 4,5716 | 6,9559 | 4,5454 | 6,9291 | 4,5326 |
| 1983 | 171,3 | 1095,4 | 99,2 | 5,1434 | 5,1210 | 6,9989 | 4,5971 | 6,9644 | 4,5716 | 6,9559 | 4,5454 |
| Год | Y | X | P | lnY | lnYLAG | lnX | lnP | lnXLAG | lnPLAG | lnXLAG2 | lnPLAG2 |
| | lnP | lnX | Const | | lnPLAG | lnXLAG | Const | | lnPLAG2 | lnXLAG2 | Const |
| | -1,072 | 20,975 | -0,816 | | -2,235 | 19,366 | 0,4633 | | -3,266 | 20,644 | 1,4046 |
| | -0,352 | 1,1763 | -1,509 | | -0,729 | 1,0869 | 0,8574 | | -0,947 | 1,0245 | 2,3093 |
| | 0,3285 | 0,0561 | 1,8488 | | 0,3263 | 0,0561 | 1,8505 | | 0,2899 | 0,0496 | 1,6441 |
| | 0,9909 | 0,0316 | #Н/Д | | 0,9939 | 0,0258 | #Н/Д | | 0,9955 | 0,0221 | #Н/Д |
| | 1137,6 | 21 | #Н/Д | | 1708,7 | 21 | #Н/Д | | 2322 | 21 | #Н/Д |
| | 2,2664 | 0,0209 | #Н/Д | | 2,2733 | 0,014 | #Н/Д | | 2,277 | 0,0103 | #Н/Д |

Рис. 61. Зразок звіту до лабораторної роботи № 15

Під таблицею даних функцією ЛИНЕЙН обчислені параметри моделей а, а1, а2 разом з усіма статистичними характеристиками. Кожний блок висновку функції ЛИНЕЙН доповнений зверху рядком статистик Стьюдента $t_b = b / S_b$. Рядок заголовків над кожним блоком дозволяє зрозуміти, яка модель мається на увазі.

Параметри моделей б) і в) також обчислюємо функцією ЛИНЕЙН (рис. 62).

| lnPLAG | lnXLAG | lnP | lnX | Const | lnPLAG2 | lnXLAG2 | lnPLAG | lnXLAG | lnP | lnX | Const |
|---------------|---------------|----------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| -4,01 | 2,9426 | 2,7833 | 0,7532 | 0,1073 | -2,511 | 1,6176 | 0,6798 | 0,3823 | -0,282 | 1,1782 | 1,1708 |
| -1,604 | 0,8818 | 1 | 0,2206 | 0,1745 | -1,438 | 0,5531 | 0,6508 | 0,1784 | -0,158 | 0,3136 | 2,1372 |
| 0,4 | 0,2997 | 0,3593 | 0,2928 | 1,6274 | 0,5726 | 0,3419 | 0,9573 | 0,4665 | 0,5591 | 0,2662 | 1,8254 |
| 0,996 | 0,0218 | #Н/Д | #Н/Д | #Н/Д | 0,9971 | 0,0196 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| 1196,2 | 19 | #Н/Д | #Н/Д | #Н/Д | 988,71 | 17 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| 2,2782 | 0,009 | #Н/Д | #Н/Д | #Н/Д | 2,2808 | 0,0065 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д |

Рис. 62. Блоки висновків функції ЛИНЕЙН

Ще раз звертаємо увагу на те, що зі збільшенням числа врахованих в моделі лагових змінних коефіцієнти регресії стають менш значимими за критерієм Ст'юдента, тобто вони стають більш нестабільними. Прийняття гіпотези Койка про експоненційне убавання коефіцієнтів регресії дозволило описати дані більш економною авторегресійною моделлю, яка містить усього 4 члени:

$$\ln Y_p = a_1 + b \cdot \ln X + c \cdot \ln P + q \cdot \ln YLAG.$$

Функцією ЛИНЕЙН обчислюємо параметри цієї авторегресійної моделі:

$$\ln Y_p = 0,50 + 0,15 \cdot \ln X - 0,16 \cdot \ln P + 0,845 \cdot \ln YLAG ; \quad R^2 = 0,9996$$

(tb) (1,3) (3,1) (2,4) (22,4)

Тепер ми можемо оцінити як коротко так і довгострокові ефекти. У короткостроковому аспекті (для поточного періоду) значення YLAG потрібно розглядати як фіксоване, тоді еластичності витрат за доходом і ціною будуть дорівнюють коефіцієнтам регресії: $b = 0,15$ і $c = -0,16$. У довгостроковому аспекті, коли X, P, Y наближаються до своїх рівноважним значенням (тобто YLAG = Y), виявляється, що довгостроковий вплив X на Y дорівнює величині $b / (1 - q) = 0,15 / (1 - 0,845) = 0,96$, а довгостроковий вплив P на Y дорівнює $c / (1 - q) = -0,16 / (1 - 0,845) = -1,02$. Ці цифри близькі до значимих коефіцієнтів регресій a, a_1, a_2 .

Перетворення Койка засноване на трохі обтяжливому припущенні, що коефіцієнти регресії при лагових пояснюючих змінних експоненціально убавують, починаючи з першого члена. Іноді розумно припустити, що зміна залежної змінної у відповідь на зміну пояснюючої змінної спочатку невелика, далі зростає і лише згодом починає зменшуватися. Щоб урахувати цей можливий ефект, приймемо гіпотезу про експоненціальне убавання ваги не з 1-го, а з 2-го члена. Тоді перетворення Койка приведе до такої моделі (одержить її):

$$\ln Y = a + (b \cdot \ln X + c \cdot \ln P) + (b_0 \cdot \ln X_{LAG} + c_0 \cdot \ln P_{LAG}) + q \cdot \ln Y_{LAG} + u,$$

де $b_0 = b - b \cdot q$; $c_0 = c - c \cdot q$.

Оцініть (функцією ЛИНЕЙН) параметри цієї моделі і обчисліть послідовні ваги для доходу попередніх років:

$$b_1 = b \cdot q + b_0;$$

$$b_2 = b_1 \cdot q;$$

$$b_3 = b_2 \cdot q;$$

.....

Ваги для цін минулих років обчислюються аналогічно.

Параметри b_0 , c_0 уточнюють особливу поведінку 1-го члена (b_1 , c_1).

Завдання для самостійної роботи

Скласти модель з розподіленими лагами. Методом Койка перетворити цю модель до економної авторегресійної моделі. Оцінити параметри моделі і обчислити короткостокі і довгострокові ефекти.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Що таке "лагові змінні"?
2. Чим модель з розподіленими лагами відрізняється від авторегресійної моделі?
3. Що заважає оцінюванню параметрів моделі з розподіленими лагами звичаним методом найменших квадратів?
4. Опишіть ідею Койка перетворення моделі з розподіленими лагами до економної авторегресійної моделі.
5. У стандартному алгоритмі Койка вважається, що внески кожної пояснюючої змінної при збільшенні часового періоду спадають за геометричною прогресією з однаковим знаменником (тобто з однаковим темпом спаду). Чи не можна модифікувати цей алгоритм так, щоб врахувати різні темпи спаду для різних пояснюючих змінних?
6. Що таке "короткострокові" і "довгострокові" ефекти, як вони визначаються?

Лабораторна робота № 16. Використання надбудови “Пакет анализа”

Мета роботи. Ознайомитися з можливостями надбудови “*Пакет анализа*” і освоїти процедуру “*Регрессия*” цієї надбудови. Порівняти результати роботи процедури “*Регрессия*” с висновками функции ЛИНЕЙН.

Ще раз нагадаємо, що службові слова та повідомлення *русифікованої* версії Excel виділяємо курсивом, не перекладаючи їх на українську мову.

Електронна таблиця Excel має ряд надбудов (вбудованих програм), що значно розширюють її можливості. Якщо надбудови встановлені, у меню *Сервис* з'являються відповідні пункти. Так, при встановленій надбудові “*Пакет анализа*” в меню *Сервис* має з'явитися пункт *Анализ данных* (рис. 63). Можливо, що надбудова вже встановлена, але виключена (не активізована). Тоді її треба включити: *Сервис, Надстройки, [√] Пакет анализа* (рис. 64).

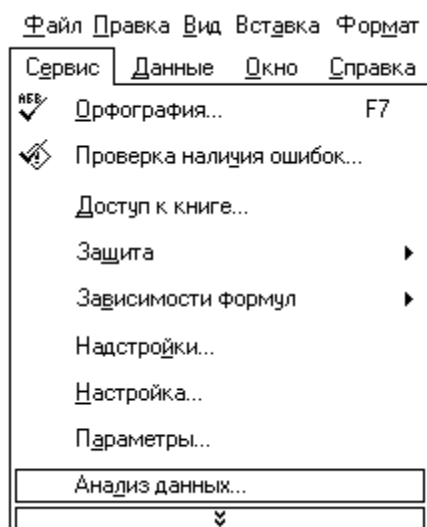


Рис. 63. Меню *Сервис*

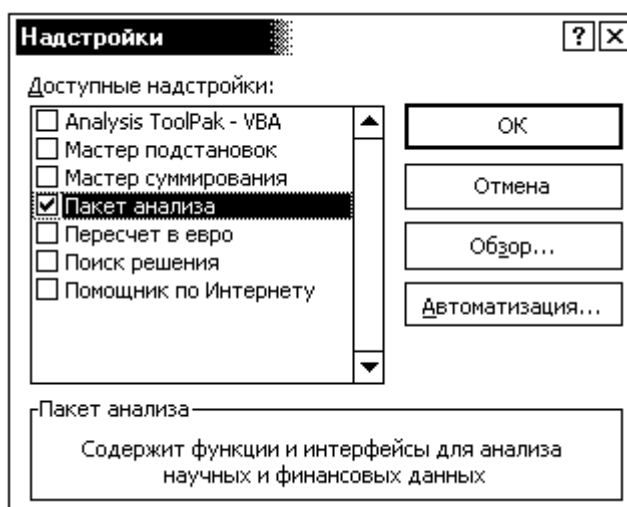


Рис. 64. Перелік встановлених надбудов

Отже, викликаємо надбудову “*Пакет анализа*”: *Сервис, Анализ данных*. У меню, що з'явиться (рис. 65), пропонується ряд статистичних процедур, з яких вибираємо процедуру “*Регрессия*”. У вікні цієї процедури (рис. 66) заповнюємо поля введення. Якщо поставити прапорець *Метки*, то вважається, що в першому рядку діапазонів *Входной интервал Y* і *Входной интервал X* записані імена відповідних стовпців даних. У розрахунках автоматично враховується рівень

надійності 95 %, але можна додатково задати ще один *Уровень надежности* 99 % . Результати аналізу можна вивести на новий робочий лист, у нову робочу книгу, або вказати координати верхнього лівого кута *Выходного интервала*.

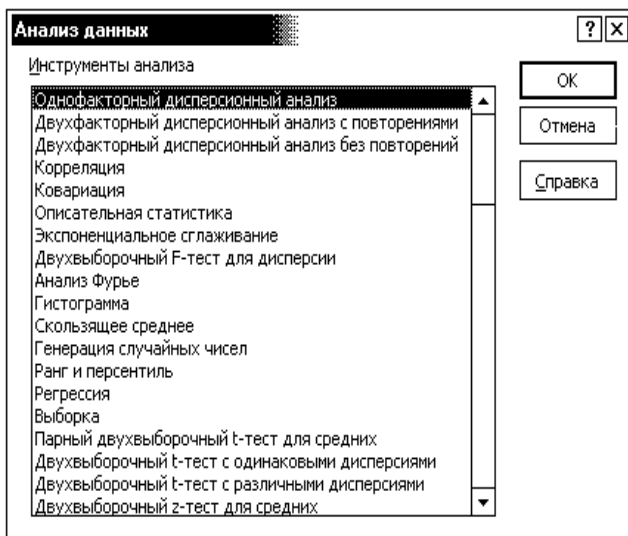


Рис. 65. Процедури Пакета Анализа

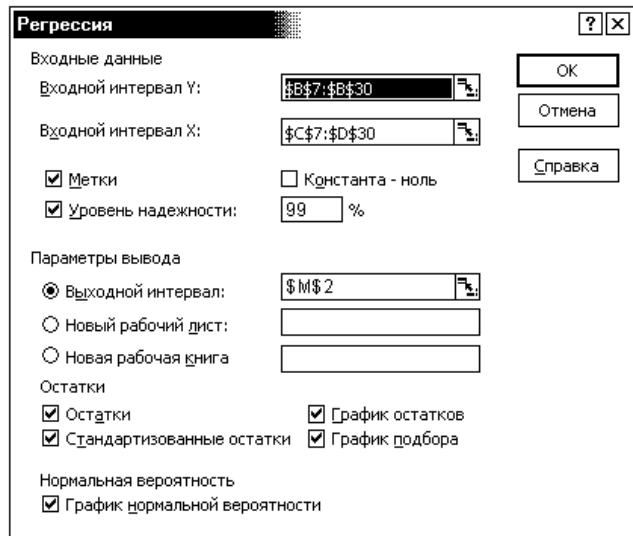


Рис. 66. Вікно процедури Регрессия

Будемо розглядати дані Коба – Дугласа й вивчати лінійну залежність після попереднього логарифмування всіх змінних: $\ln Y = a + b \cdot \ln K + c \cdot \ln L + e$, де Y – обсяг випуску продукції, K – витрати капіталу, L – витрати праці (у відсотках до рівня 1899 р.). Імена стовпців $\ln Y$, $\ln K$, $\ln L$ записані у першому рядку діапазону даних (див. лабораторну роботу № 10).

Результати роботи процедури представлені у вигляді обов'язкових трьох таблиць статистичного аналізу; крім того, на панелі процедури додатково пропонується розрахувати залишки моделі й побудувати деякі графіки. Замовляємо все (ставимо всі прапорці).

Нижче на рис. 67 наведений зразок виконання роботи. Для зручності друку результатів на принтері висновок процедури “Регрессия” довелося перекомпонувати й злегка відредагувати. Перша таблиця містить вихідні дані. Праворуч від цієї таблиці даних (після нашої перекомпонувки) розташовані ще дві таблиці: “ВЫВОД ОСТАТКА” й “ВЫВОД ВЕРОЯТНОСТИ”. У таблиці “ВЫВОД ОСТАТКА” наведені розрахункові значення $\ln Y_p$, залишки моделі $e = \ln Y - \ln Y_p$ і стандартизовані залишки. Таблиця “ВЫВОД ВЕРОЯТНОСТИ” необхідна для перевірки нор-

мальності розподілу $\ln Y$ (для побудови замовленого графіка нормальної імовірнісної кривої).

| Робота № 16. Аналіз даних за допомогою надбудови “Пакет аналіза” | | | | | | | | | | |
|---|---------------------------|---------------------------|---------------------------|---|-----------------------------|----------------------------|--------------------------|---------------------------|----------------|----------------|
| Y – обсяги випуску в % к 1899 г. | | | | | | | | | | |
| K – витрати капіталу в % к 1899 г. | | | | | | | | | | |
| L – витрати праці в % к 1899 г. | | | | ВЫВОД ОСТАТКА | | | ВЫВОД ВЕРОЯТНОСТИ | | | |
| $\ln Y = a + b \cdot \ln K + c \cdot \ln L + e$ | | | | Предсказанное $\ln Y$ | | Стандартные остатки | | $\ln Y$ | | |
| Рік | $\ln Y$ | $\ln K$ | $\ln L$ | Наблюдения | ↓ | Остатки | ↓ | Перцентиль | ↓ | |
| 1900 | 4,61512 | 4,67283 | 4,65396 | 1 | 4,66676 | -0,05160 | -0,89050 | 2,17391 | 4,61512 | |
| 1901 | 4,71850 | 4,73620 | 4,70048 | 2 | 4,71863 | -0,00010 | -0,00230 | 6,52174 | 4,71850 | |
| 1902 | 4,80402 | 4,80402 | 4,77068 | 3 | 4,79004 | 0,01398 | 0,24103 | 10,8696 | 4,80402 | |
| 1903 | 4,82028 | 4,8752 | 4,85981 | 4 | 4,87702 | -0,05670 | -0,97840 | 15,2174 | 4,80402 | |
| 1904 | 4,80402 | 4,92725 | 4,75359 | 5 | 4,80719 | -0,00320 | -0,05470 | 19,5652 | 4,82028 | |
| 1905 | 4,96284 | 5,00395 | 4,82831 | 6 | 4,88432 | 0,07853 | 1,35425 | 23,9130 | 4,83628 | |
| 1906 | 5,02388 | 5,09375 | 4,89035 | 7 | 4,95480 | 0,06908 | 1,19127 | 28,2609 | 4,96284 | |
| 1907 | 5,01728 | 5,17048 | 4,92725 | 8 | 5,00250 | 0,01478 | 0,25494 | 32,6087 | 5,01728 | |
| 1908 | 4,83628 | 5,22036 | 4,79579 | 9 | 4,91249 | -0,07620 | -1,31420 | 36,9565 | 5,02388 | |
| 1909 | 5,04343 | 5,28827 | 4,94164 | 10 | 5,04281 | 0,00062 | 0,01065 | 41,3043 | 5,03044 | |
| 1910 | 5,0689 | 5,33754 | 4,96981 | 11 | 5,07691 | -0,00800 | -0,13810 | 45,6522 | 5,04343 | |
| 1911 | 5,03044 | 5,37528 | 4,97673 | 12 | 5,09163 | -0,06120 | -1,05530 | 50,0000 | 5,06890 | |
| 1912 | 5,17615 | 5,42053 | 5,02388 | 13 | 5,13952 | 0,03663 | 0,63177 | 54,3478 | 5,12990 | |
| 1913 | 5,21494 | 5,46383 | 5,03695 | 14 | 5,16039 | 0,05454 | 0,94063 | 58,6957 | 5,17615 | |
| 1914 | 5,12990 | 5,49717 | 5,00395 | 15 | 5,14294 | -0,01300 | -0,22490 | 63,0435 | 5,18739 | |
| 1915 | 5,24175 | 5,5835 | 5,03695 | 16 | 5,18997 | 0,05178 | 0,89297 | 67,3913 | 5,21494 | |
| 1916 | 5,4161 | 5,69709 | 5,20401 | 17 | 5,34809 | 0,06801 | 1,17295 | 71,7391 | 5,24175 | |
| 1917 | 5,42495 | 5,81413 | 5,27811 | 18 | 5,43470 | -0,00980 | -0,16820 | 76,0870 | 5,38450 | |
| 1918 | 5,40717 | 5,90263 | 5,29832 | 19 | 5,47230 | -0,06510 | -1,12320 | 80,4348 | 5,40717 | |
| 1919 | 5,38450 | 5,95842 | 5,26269 | 20 | 5,45836 | -0,07390 | -1,27380 | 84,7826 | 5,4161 | |
| 1920 | 5,44242 | 6,00881 | 5,26269 | 21 | 5,47081 | -0,02840 | -0,48960 | 89,1304 | 5,42495 | |
| 1921 | 5,18739 | 6,03309 | 4,99043 | 22 | 5,26487 | -0,07750 | -1,33630 | 93,4783 | 5,44242 | |
| 1922 | 5,48064 | 6,06611 | 5,08140 | 23 | 5,34385 | 0,13679 | 2,35905 | 97,8261 | 5,48064 | |
| Min | 4,61512 | 4,67283 | 4,65396 | Min | 4,66676 | -0,07750 | -1,33630 | 2,17391 | 4,61512 | |
| Max | 5,48064 | 6,06611 | 5,29832 | Max | 5,47230 | 0,13679 | 2,35905 | 97,8261 | 5,48064 | |
| ВЫВОД ИТОГОВ | | | | | | | | | | |
| <i>Регрессион. статистика</i> | | | | | <i>Дисперсионный анализ</i> | | | | | |
| Множественный R | | 0,97388 | | | | df | SS | MS | F | Значим. |
| R-квадрат | | 0,94844 | | | Модель | 2 | 1,3606 | 0,6803 | 183,936 | 1,3E-13 |
| Нормир. R-квадрат | | 0,94328 | | | Остаток | 20 | 0,07397 | 0,0037 | | |
| Стандартная ошибка | | 0,06082 | | | Итого | 22 | 1,43457 | | | |
| Наблюдения | | 23 | | | | | | | | |
| | | | | | | | | | | |
| | Коэфф-т | Ст. ошиб. | t – стат. | P-Знач. | Нижн. 95% | Верхн. 95% | Нижн. 99% | Верхн. 99% | | |
| Y-перес. | -0,1110 | 0,46978 | -0,2363 | 0,81561 | -1,0910 | 0,86894 | -1,44770 | 1,22569 | | |
| $\ln K$ | 0,24715 | 0,06482 | 3,81268 | 0,00109 | 0,11193 | 0,38237 | 0,06271 | 0,43159 | | |
| $\ln L$ | 0,77845 | 0,15122 | 5,14785 | 4,9E-05 | 0,46301 | 1,09389 | 0,34818 | 1,20872 | | |

Рис. 67. Зразок звіту до лабораторної роботи № 15

На рис. 67 під таблицею даних наведено три обов'язкові таблиці “ВЫВОД ИТОГОВ”: “Регрессионная статистика”, “Дисперсионный анализ” і таблиця оцінок коефіцієнтів регресії з їх статистичними характеристиками.

Розглянемо детально зміст трьох таблиць “ВЫВОД ИТОГОВ”.

Зміст першої таблиці “Регрессионная статистика” зрозумілий, але замість поняття “Нормированный R-квадрат” у вітчизняній літературі використовується термін *Скоректований R-квадрат*. Максимальному значенню цього показника відповідає мінімальне значення MSE. Як видно із цієї таблиці, моделлю Коба – Дугласа пояснюється 94,8% мінливості результативної ознаки $\ln Y$.

Друга таблиця – таблиця дисперсійного аналізу – має стандартний вигляд; у графі “Значимость F” записаний рівень значимості F. Оскільки рівень значимості виявився менше за 0,01, робимо висновок про значимість прийнятої моделі у цілому.

У третій таблиці наведені коефіцієнти регресії (a; b; c), їх стандартні відхилення (S_a ; S_b ; S_c), статистики Ст'юдента (t_a ; t_b ; t_c), їх рівні значимості (“P-значения”), нижні й верхні границі 95%-х довірчих інтервалів на коефіцієнти регресії й (додатково) нижні й верхні границі 99%-х довірчих інтервалів. Звернемо увагу на те, що вільний член моделі (параметр a) тут іменується “Y-пересечение”. Коефіцієнти регресії перед пояснюючими змінними $\ln K$, $\ln L$ значимі як за критерієм Ст'юдента (t_b , $t_c > 3$), так і за рівнями значимості (вони менше за 0,01). Вільний член незначимий за обома критеріями. “Нуль-гіпотеза” про рівність вільного члена нулю не може бути відкинута. Цей же висновок виходить також з розгляду границь довірчих інтервалів – в інтервали для вільного члена потрапляє нуль. Випишемо отримане рівняння регресії:

$$\ln Y = -0,111 + 0,247 * \ln K + 0,778 * \ln L .$$

Практично всі ці результати є в блоці висновків функції ЛИНЕЙН, але без пояснень. Для порівняння на рис. 68 додатково наведений блок функції ЛИНЕЙН і поруч з ним складено блок пояснень. Пропонується порівняти висновки процедури “Регрессия” і висновки функції ЛИНЕЙН.

| InL | InK | Const | | InL | InK | Const |
|----------------|----------------|----------------|--|----------|----------|----------|
| 5,14785 | 3,81268 | -0,2363 | | tc | tb | ta |
| 0,77845 | 0,24715 | -0,1110 | | c | b | a |
| 0,15122 | 0,06482 | 0,46978 | | Sc | Sb | Sa |
| 0,94844 | 0,06082 | #Н/Д | | R2 | Se | |
| 183,936 | 20 | #Н/Д | | F | dfE | |
| 1,36060 | 0,07397 | #Н/Д | | SSR | SSE | |

Рис. 68. Блок висновків функції ЛИНЕЙН

Крім числових характеристик були замовлені графіки, які розташовуються каскадом. Усі ці графіки вимагають додаткового редагування (рис. 69).

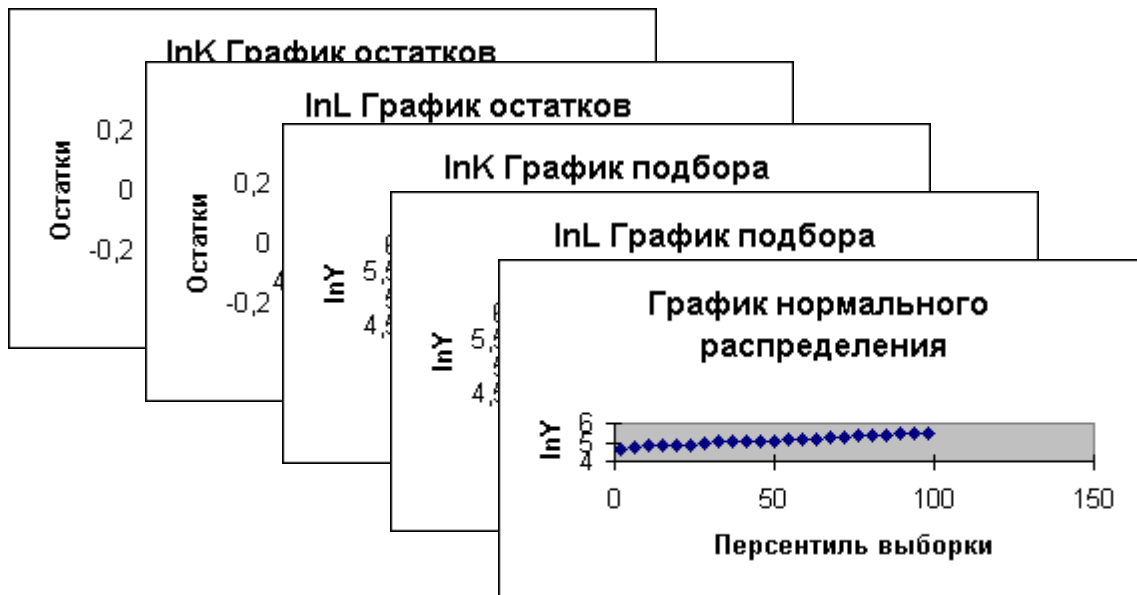


Рис. 69. Початковий вигляд замовлених графіків

Насамперед у кожному графіку слід указати мінімальні й максимальні значення за кожною змінною. Саме для цього таблиця вихідних даних, таблиця залишків і таблиця “ВЫВОД ВЕРОЯТНОСТЕЙ” були доповнені нами підсумковими рядками Min і Max. Вже після однієї цієї операції графіками можна користуватися (на відміну від вихідних заготовок). Повністю же всі вимоги до якісного оформлення графіків викладені в лабораторній роботі № 1 (порівняйте вигляд графіка нормального розподілу до й після редагування на рис. 69 і рис.70).

На останньому графіку (рис. 69 і рис. 70) зображена так звана “нормальна імовірнісна крива” для $\ln Y$. Будується вона у такий спосіб. Спочатку обчислюються ординати кумуляти (графіка накопичених ймовірностей). Далі отримана кумулята моделюється інтегральною функцією нормального закону розподілу й

за ординатами кумуляти (ймовірностями) обчислюються квантілі (персентілі) нормованого нормального розподілу. Якщо розподіл дійсний нормальний, то точки нормальної імовірнісної кривої лягають на пряму, що й спостерігається в побудованому графіку.

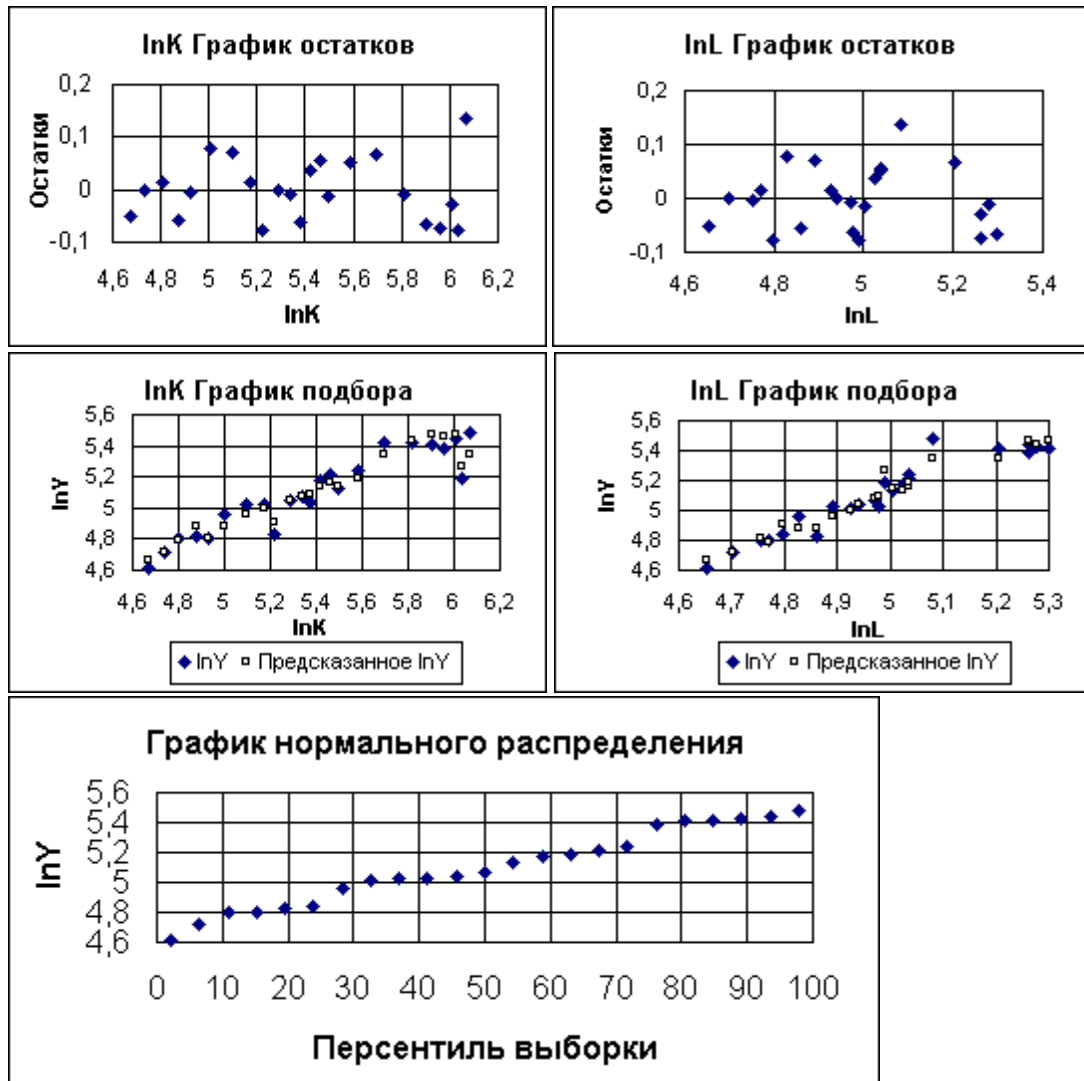


Рис. 70. Вигляд замовлених графіків після форматування

Розглядаючи на рис. 70 “Графики остатков”, не помічаємо появи гетероскедастичності, яка звичайно очікується при функціональних перетвореннях залежної змінної (lnY). Проте трохи хвилеподібний характер розташування залишків наводить на підозру про існування автокореляції залишків. Тому обчислюємо (функцією КОРРЕЛ) коефіцієнт автокореляції 1-го порядку: $\rho = 0,09$. Цей коефіцієнт виявився дуже малим, отже, вважаємо, що автокореляції немає. Запропонована модель – доброякісна.

На “Графіках підбора” порівняні спостережувані й розрахункові значення, звідки видна дуже гарна відповідність. Дійсно, коефіцієнт кореляції між спостережуваними й розрахунковими значеннями дорівнює коефіцієнту множинної кореляції $R = 0,974$, який у цьому завданні близький до одиниці.

Контрольне запитання. Рівняння регресії щодо логарифмів – лінійне. Чому ж на “графіках підбора” розрахункові точки не лягають строго на пряму? Чи є “графіки підбора” графіками компонентних ефектів?

Завдання для самостійної роботи

За допомогою надбудови “Пакет аналіза” зробить математичну обробку тих самих даних, що були прийняті як завдання до лабораторної роботи № 10. Порівняйте отримані результати.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Як активізувати надбудову Excel?
2. Які документи створюються процедурою “Регресія” надбудови “Пакет аналіза”?
3. Що таке “гомоскедастичність” і “гетероскедастичність” залишків моделі? Як перевіряється гомоскедастичність залишків моделі?
4. Як перевірити передумову про відсутність автокореляції залишків моделі? Як розрахувати коефіцієнт автокореляції 1-го порядку? Для чого призначений критерій Дарбіна – Уотсона?
5. Що таке “скорегований коефіцієнт детермінації”, які його властивості, де він використовується?:
6. Як оцінюється значущість моделі у цілому і значущість її окремих членів?
7. Як побудувати інтервальні оцінки параметрів моделі? Що це дає додатково в порівнянні з точковими оцінками?
8. Що таке “нормальна імовірнісна крива”, як вона будується і де використовується?

Лабораторна робота № 17. Дисперсійний аналіз (ANOVA)

Мета роботи. Познайомитися з процедурами дисперсійних аналізів надбудови “Пакет аналіза”, вивчити застосування однофакторного і двофакторного дисперсійного аналізу. У складі регресійного аналізу також є декілька специфічний блок дисперсійного аналізу для перевірки значимості регресійної моделі. Коли є можливість застосувати обидві різновиди аналізу, перевіряють також адекватність прийнятої моделі, тобто адекватність прийнятої форми зв’язку.

Теоретична частина

Дисперсійний аналіз (ANOVA – **A**nalysis of **V**ariance) призначений для встановлення розбіжностей між групами (класами, вибірками) спостережень за даними кількісного показника Y (результативної ознаки). Кожна група характеризується певним значенням класифікаційного фактору X (аналог пояснюючої змінної). На відміну від регресійного аналізу, у якому всі змінні має бути кількісними, у дисперсійному аналізі класифікаційний фактор може бути якісним, вимірюваний у найбільш загальній шкалі імен. Якщо фактор X – кількісний, то для всіх об’єктів однієї групи його значення повинні бути однаковими. Аналіз однофакторний, коли маємо угруповання за однією класифікаційною змінною. У двофакторному аналізі є угруповання за двома класифікаційними факторами, проте вимагається, щоб для кожної комбінації категорій двох факторів була однакова кількість спостережень.

Розглянемо спочатку алгоритм однофакторного дисперсійного аналізу.

Позначимо через n – число категорій (груп, класів, вибірок, варіантів), m_i – число об’єктів (спостережень) у кожній групі, $N = \sum m_i$ – загальне число спостережень, Y_{ij} – значення показника Y , U_{sr} – загальне середнє, Z_i – середні Y для кожної групи, ε_{ij} – відхилення Y_{ij} від Z_i (відхилення кожного спостереження від середнього в цій групі). Тоді маємо співвідношення:

$$\begin{aligned} Y &= Z + \varepsilon, \\ SSY &= SSZ + SS\varepsilon, \\ dfY &= dfZ + df\varepsilon, \end{aligned}$$

де
$$SSY = \sum\sum(Y_{ij} - Y_{cp})^2; \quad SSZ = \sum\sum(Z_i - Y_{cp})^2; \quad SS\varepsilon = \sum\sum(Y_{ij} - Z_i)^2,$$

$$dfY = N - 1; \quad dfZ = n - 1; \quad df\varepsilon = N - n.$$

Далі обчислюємо середні квадрати (незміщені оцінки дисперсій):

$$MSZ = SSZ / dfZ; \quad MS\varepsilon = SS\varepsilon / df\varepsilon.$$

Дисперсійне відношення Фішера $F = MSZ / MS\varepsilon$ показує, у скільки разів мінливість даних між групами спостережень перевищує випадкову мінливість (мінливість усередині груп). Якщо вийде, що обчислене значення менше табличного $F < F_{05}(dfZ, df\varepsilon)$, то робиться висновок про відсутність значимих розходжень між групами (нуль-гіпотеза не може бути відкинута); у цьому випадку вважаємо, що фактор X не має значимого впливу на Y . Розходження між групами визнаються значимими, якщо $F > F_{01}(dfZ, df\varepsilon)$.

Відношення $\eta^2 = SSZ / SSY$ називається *індексом детермінації*, воно показує, яка частина повної мінливості Y є наслідком різниць між групами (якщо фактор X – кількісний, говорять ще так: “яка частина повної мінливості Y пояснюється наявністю кореляційного зв'язку між X і Y ”).

У регресійному аналізі є блок дисперсійного аналізу для перевірки значимості регресійної моделі, де прийняте декілька інше розкладання “повного сигналу Y ” на “корисний сигнал Y_p ” і похибку:

$$\begin{aligned} Y &= Y_p + e, \\ SSY &= SSR + SSE, \\ dfY &= dfR + dfE; \\ MSR &= SSR / dfR; \quad MSE = SSE / dfE, \\ F &= MSR / MSE, \\ R^2 &= SSR / SSY. \end{aligned}$$

На відміну від індексу детермінації $\eta^2 = SSZ / SSY$ коефіцієнт детермінації $R^2 = SSR / SSY$ показує, яка частина повної мінливості Y пояснюється *регресійною моделлю*, тобто кореляційною залежністю заданої форми зв'язку (лінійною, якщо рівняння регресії лінійне). Природно, завжди $R^2 < \eta^2$. Залишок моделі прийнято позначати латинською буквою E (від слова ERROR), грецькою буквою ε позначається випадкова похибка (за випадкову мінливість приймається мінливість Y усередині груп при фіксованих значеннях класифікаційного фактора X). На відміну від дисперсійного аналізу в залишки регресійної моделі

крім випадкових похибок включаються також помилки специфікації моделі, помилки вибору невірної форми зв'язку, які не є випадковими. Якщо у нас є міра чисто випадкової мінливості, то можна перевірити *адекватність* прийнятої моделі, тобто встановити, чи є внесок систематичної помилки значимим, чи ні; чи можна знехтувати систематичною помилкою невірної специфікації моделі (адже найчастіше правильну специфікацію ми не знаємо).

Позначимо можливу систематичну помилку через W . Тоді маємо:

$$\begin{aligned} e &= W + \varepsilon, \\ SSe &= SSW + SS\varepsilon, \\ dfe &= dfW + df\varepsilon; \\ MSW &= SSW / dfW; \quad MS\varepsilon = SS\varepsilon / df\varepsilon, \\ F &= MSW / MS\varepsilon. \end{aligned}$$

Якщо обчислене значення $F = MSW / MS\varepsilon$ виявиться меншим за табличне значення $F_{05}(dfW, df\varepsilon)$, модель визнається адекватною (внесок систематичної помилки невірної специфікації моделі не значимий, ним можна знехтувати).

Для можливості перевірки адекватності регресійної моделі часто робиться попереднє угруповання даних за невеликими інтервалами варіювання пояснюючої змінної X . Усі проміжні викладення дисперсійного аналізу зручно записуються у стандартній табличній формі. Розглянемо дані на рис. 71.

| | A | B | C | D | E | F | G | H | I |
|----|----------|----------------|-----------|-----------|-----------|----------|----------|--------------|-------------|
| 5 | № | Добрива | y1 | y2 | y3 | m | Z | ДИСПР | ДИСП |
| 6 | 1 | Контроль | 21,14 | 22,80 | 20,40 | 3 | 21,45 | 1,01 | 1,51 |
| 7 | 2 | N180 | 27,40 | 39,20 | 45,06 | 3 | 37,22 | 53,94 | 80,91 |
| 8 | 3 | P180 | 21,10 | 21,80 | 25,00 | 3 | 22,63 | 2,88 | 4,32 |
| 9 | 4 | N180P180 | 51,00 | 49,60 | 52,20 | 3 | 50,93 | 1,13 | 1,69 |
| 10 | 5 | H40 | 31,20 | 25,20 | 33,20 | 3 | 29,87 | 11,56 | 17,33 |
| 11 | 6 | N180H40 | 38,40 | 51,40 | 44,20 | 3 | 44,67 | 28,28 | 42,41 |
| 12 | 7 | P180H40 | 27,88 | 33,80 | 25,50 | 3 | 29,06 | 12,18 | 18,27 |
| 13 | 8 | N180P180H40 | 40,38 | 48,80 | 49,00 | 3 | 46,06 | 16,14 | 24,21 |
| 14 | 9 | P90H20 | 27,32 | 18,40 | 29,84 | 3 | 25,19 | 24,09 | 36,13 |
| 15 | 10 | N180P90H20 | 45,48 | 51,34 | 44,00 | 3 | 46,94 | 10,05 | 15,07 |
| 16 | 11 | N90H20 | 39,80 | 49,66 | 41,42 | 3 | 43,63 | 18,64 | 27,96 |
| 17 | 12 | N90P180H20 | 48,30 | 55,50 | 46,40 | 3 | 50,07 | 15,36 | 23,04 |
| 18 | 13 | N90P90 | 35,40 | 47,20 | 47,40 | 3 | 43,33 | 31,48 | 47,21 |
| 19 | 14 | N90P90H40 | 54,40 | 46,40 | 55,40 | 3 | 52,07 | 16,22 | 24,33 |

Рис. 71. Дані польового дослід з добривами

Кожний рядок таблиці представляє певний варіант внесення добрив ($n = 14$), який реалізовано у трьох повтореннях ($m = 3$). Кожному варіанту можна дати змістовне ім'я, що і зроблено в колонці Добрива. Ми бажаємо з'ясувати, чи є значимим вплив добрив на врожайність культури.

Як вже було сказано раніше, електронна таблиця Excel має ряд надбудов (вбудованих програм), що значно розширюють можливості Excel. Якщо надбудови встановлені, у меню *Сервис* з'являються відповідні пункти (див. рис. 63). Так, при встановленій надбудові “*Пакет анализа*” у меню *Сервис* повинен з'явитися пункт *Анализ данных*. Можливо, що ця надбудова вже встановлена, але виключена (не активізована). Тоді треба її включити: *Сервис, Настройки, Пакет анализа* (рис. 64).

Отже, викликаємо надбудову “*Пакет анализа*”: *Сервис, Анализ данных*.

У меню, що з'являється (див. рис. 65), пропонується ряд статистичних процедур, з яких вибираємо першу процедуру “*Однофакторный дисперсионный анализ*”. З'являється панель цієї процедури (рис. 72), на якій заповнюємо поля введення.

Рис. 72. Вікно процедури дисперсійного аналізу

У полі *Входной интервал* вказуємо діапазон даних (без заголовків) разом зі стовпчиком Добрива і ставимо прапорець *Метки в первом столбце*. Встановлюємо перемикач *Группирование* *по строкам*. Рівень значимості *Альфа*

для табличних значень F05 або F01 приймаємо рівним 0,01. У полі *Выходной интервал* указуємо лівий верхній кут на вільному місці робочого аркуша. У результаті одержуємо дві таблиці (рис. 73 і рис. 74). Заголовки цих таблиць не перекладаємо на українську мову – саме такий вигляд має висновок Excel.

| Итоги | | | | |
|---------------|-------------|--------------|----------------|------------------|
| Группы | Счет | Суммы | Средние | Дисперсии |
| Контроль | 3 | 64,34 | 21,447 | 1,5105 |
| N180 | 3 | 111,7 | 37,220 | 80,909 |
| P180 | 3 | 67,90 | 22,633 | 4,3233 |
| N180P180 | 3 | 152,8 | 50,933 | 1,6933 |
| H40 | 3 | 89,60 | 29,867 | 17,333 |
| N180H40 | 3 | 134,0 | 44,667 | 42,413 |
| P180H40 | 3 | 87,18 | 29,060 | 18,267 |
| N180P180H40 | 3 | 138,2 | 46,060 | 24,207 |
| P90H20 | 3 | 75,56 | 25,187 | 36,132 |
| N180P90H20 | 3 | 140,8 | 46,940 | 15,068 |
| N90H20 | 3 | 130,9 | 43,627 | 27,957 |
| N90P180H20 | 3 | 150,2 | 50,067 | 23,043 |
| N90P90 | 3 | 130,0 | 43,333 | 47,213 |
| N90P90H40 | 3 | 156,2 | 52,067 | 24,333 |

Рис. 73. Таблица *Итоги*

У першій таблиці (*Итоги*) для кожного рядка (кожної групи) підраховане число спостережень m_i , середнє Z_i і дисперсія. Зараз у всіх групах по 3 спостереження, але допускаються також пропуски даних. У загальному випадку число спостережень визначається функцією СЧЕТ, середні – функцією СРЗНАЧ; дисперсії у цій таблиці

знайдені функцією ДИСП (не ДИСПР!).

У другій таблиці (*Дисперсионный анализ*) крім стандартних граф наведено ще рівень значимості (графа *P-знач*) для обчисленого значення $F = 13,9$ і лише одне табличне значення $F01 = 2,84$ (оскільки було замовлено *Альфа* = 0,01).

| Дисперсионный анализ | | | | | | |
|-----------------------------|-----------|-----------|-----------|----------|----------------|-----------------|
| Изменчивость | SS | df | MS | F | P-Знач. | F критич |
| Между группами | 4717,1 | 13 | 362,86 | 13,941 | 6E-09 | 2,8418 |
| Внутри групп | 728,81 | 28 | 26,029 | | | |
| Общая | 5446 | 41 | | | | |

Рис. 74. Таблица дисперсійного аналізу

Обчислимо всі необхідні характеристики самостійно. До таблиці вихідних даних нами вже додані графи m , Z , ДИСПР і ДИСП. Загальне число спостережень N , загальне середнє $U_{ср}$ і загальну суму квадратів SSY можна обчислити функціями:

$$\begin{aligned}
 N &= \text{СЧЕТ}(36:\text{E}19) = 42, \\
 U_{ср} &= \text{СРЗНАЧ}(36:\text{E}19) = 38,79, \\
 SSY &= \text{ДИСПР}(36:\text{E}19)*N = 5446,
 \end{aligned}$$

де С6:Е19 – діапазон даних (лише числа, без заголовків).

Суму квадратів SS_{ε} (усередині груп) можна обчислити або за допомогою функції ДИСПР: $SS_{\varepsilon} = \text{СУММПРОИЗВ}(F6:F19;H6:H19) = 728,81$, або функції ДИСП: $SS_{\varepsilon} = \text{СУММПРОИЗВ}(F6:F19-1;I6:I19) = 728,81$.

Остання сума квадратів SSZ обчислюється як різниця:

$$SSZ = SSY - SS_{\varepsilon} = 5446 - 729 = 4717 .$$

Далі обчислюємо числа ступенів свободи $dfY = N - 1 = 41$; $dfZ = n - 1 = 13$; $df_{\varepsilon} = N - n = 28$; середні квадрати $MSZ = SSZ / dfZ = 362,86$; $MS_{\varepsilon} = SS_{\varepsilon} / df_{\varepsilon} = 26,03$ і дисперсійне відношення $F = MSZ / MS_{\varepsilon} = 13,94$.

Табличні значення $F_{05}(13,28) = 2,09$; $F_{01}(13,28) = 2,84$; $F_{\text{РАСП}}(13,94;13;28) = 6,12E-09$ (ця величина записана в колонку *P-знач*). Оскільки *P-знач* $< 0,01$ (або $F > F_{01}$, що еквівалентно), робимо висновок про наявність значимих розходжень між групами (варіантами різного внесення добрив), тобто між Y і N, P, H має місце значима кореляційна залежність.

Надалі потрібні наступні характеристики:

$$SS_{\varepsilon} = 728,81 ; df_{\varepsilon} = 28 ; MS_{\varepsilon} = 26,03 ; S_{\varepsilon} = \text{КОРЕНЬ}(MS_{\varepsilon}) = 5,10 .$$

Додатково обчислюємо індекс детермінації $\eta^2 = SSZ / SSY = 0,866$, який показує, що 86,6% мінливості Y визначається різницею між групами (тобто впливом добрив).

Дані цього трифакторного (N, P, H) дослідження були оброблені за програмою множинного регресійного аналізу на повну квадратичну модель (10 членів). Одержане наступне рівняння регресії:

$$Y_p = 47,27 + 0,108*(N - 90) + 0,0244*(P - 90) + 0,131*(H - 20) - 0,0015*(N - 90)^2 + 0,00023*(N - 90)*(P - 90) - 0,00085*(N - 90)*(H - 20) - 0,00099*(P - 90)*(H - 20),$$

де три останні члени (взаємодії добрив) не значимі за критерієм Ст'юдента.

Нижче на рис. 75 наведено фінальний блок результатів роботи функції ЛИНЕЙН разом з блоком пояснень, який крім коефіцієнтів регресії містить статистичні характеристики якості моделі.

Модель значима у цілому за критерієм Фішера і пояснює $R^2 = 0,850$ (85%) мінливості Y .

Фінальна модель містить декілька членів, не значимих за критерієм Ст'юдента, але їх не можна відбракувати з моделі.

| b9 | b8 | b7 | b4 | b3 | b2 | b1 | b0 |
|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| -1,791 | -1,536 | 1,8425 | -7,101 | 2,9267 | 2,4537 | 10,925 | 33,458 |
| -1E-03 | -9E-04 | 0,0002 | -0,001 | 0,1308 | 0,0244 | 0,1085 | 47,273 |
| 0,0006 | 0,0006 | 0,0001 | 0,0002 | 0,0447 | 0,0099 | 0,0099 | 1,4129 |
| 0,8504 | 4,8945 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| 27,619 | 34 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| 4631,5 | 814,49 | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| tb9 | tb8 | tb7 | tb4 | tb3 | tb2 | tb1 | tb0 |
| b9 | b8 | b7 | b4 | b3 | b2 | b1 | b0 |
| Sb9 | Sb8 | Sb7 | Sb4 | Sb3 | Sb2 | Sb1 | Sb0 |
| R2 | Se | | | | | | |
| F | df | | | | | | |
| SSR | SSE | | | | | | |

Рис. 75. Блок висновків функції ЛИНЕЙН разом з блоком пояснень

Різниця $\eta^2 - R^2 = 86,6\% - 85,0\% = 1,6\%$ характеризує помилку специфікації моделі. Для перевірки адекватності моделі заповнюємо (вручну) таку таблицю дисперсійного аналізу (рис. 76):

Дисперсійний аналіз для перевірки адекватності

| <i>Мінливість</i> | <i>SS</i> | <i>df</i> | <i>MS</i> | <i>F</i> | <i>P-Знач.</i> |
|-------------------|-----------|-----------|-----------|----------|----------------|
| Неадекватність | 85,68 | 6 | 14,28 | 0,5486 | 0,7668 |
| Усередині груп | 728,81 | 28 | 26,029 | | |
| Залишки моделі | 814,49 | 34 | | | |

Рис. 76. Перевірка адекватності моделі

Суму квадратів $SSE = 814,49$ і число ступенів свободи залишку $dfe = 34$ взято із блоку результатів функції ЛИНЕЙН. Весь рядок *Усередині груп* (випадкова помилка) переписано з попередньої таблиці дисперсійного аналізу. Залишилося обчислити рядок *Неадекватність* (помилка специфікації):

$$SSW = SSE - SS_{\varepsilon} = 814,49 - 728,81 = 85,68; \quad dfW = dfe - df_{\varepsilon} = 34 - 28 = 6;$$

$$MSW = SSW / dfW = 85,68/6 = 14,28; \quad FW = MSW / MS_{\varepsilon} = 0,549.$$

Дисперсійне значення FW виявилося менше одиниці, тобто внесок систематичної помилки – незначимий, прийнята квадратична модель – адекватна (такий же висновок впливає з порівняння $P\text{-Знач} = 0,77 > 0,05$).

Експериментальні дані згруповані не лише за варіантами внесення добрив, але й за повтореннями, які при обробці дослідів часто об'єднуються. Проте повторення дослідів можуть розташовуватися на суттєво різних ділянках з різним рівнем природної родючості, забезпеченості вологою, поживними речовинами тощо. Тому часто виникають проблеми – чи можна поєднувати дані різних повторень? Якщо ці групи спостережень істотно різняться, у модель треба вводити додаткові *dummy*-змінні, які призначені для урахування суттєвих різниць між групами даних.

Викликаємо процедуру однофакторного дисперсійного аналізу і по іншому заповнюємо поля введення: У полі *Входной интервал* вказуємо діапазон даних C5:E19 разом з рядком заголовків *y1*, *y2*, *y3* і ставимо прапорець *Метки в первой строке*. Встановлюємо перемикач *Группирование* *по столбцам*. Рівень значимості *Альфа* для табличних значень F05 або F01 приймаємо рівним 0,05. У полі *Выходной интервал* вказуємо лівий верхній кут на вільному місці робочого листа. Одержуємо такі дві таблиці (рис. 77):

| Итоги | | | | | Дисперсионный анализ | | | | | | |
|--------|------|-------|-------|-------|----------------------|-------|----|--------|--------|--------|--------|
| Группы | Счет | Суммы | Средн | Дисп. | Изменчивость | SS | df | MS | F | P-Знач | F крит |
| y1 | 14 | 509,2 | 36,37 | 117,2 | Между групп | 123,3 | 2 | 61,666 | 0,4518 | 0,6397 | 3,2381 |
| y2 | 14 | 561,1 | 40,08 | 169,7 | Внутри групп | 5323 | 39 | 136,48 | | | |
| y3 | 14 | 559,0 | 39,93 | 122,5 | Общая | 5446 | 41 | | | | |

Рис. 77. Результати роботи процедури дисперсійного аналізу

З таблиці дисперсійного аналізу знаходимо $F = 0,452 < 1$, звідки робимо висновок про відсутність значимих розходжень між повтореннями (цей же висновок випливає з $P\text{-Знач} = 0,6397 > 0,05$ і з $F < F_{05} = 3,24$). Повторення можна поєднувати і не потрібно вводити в модель додаткові *dummy*-змінні.

Тепер ознайомимся з процедурами двофакторного дисперсійного аналізу, що реалізовані у надбудові “*Пакет анализа*”. Наші дані згруповані по рядках і стовпцях, яким відповідають різні значення класифікаційних факторів x_1 і x_2 . На перетину кожного рядка і кожного стовпця розташовані дані спостережень цього варіанту. Якщо для кожного варіанта є кілька спостережень (повторення дослідів), ці дані треба записувати в кілька рядків. У процедурах двофакторного

аналізу надбудови “*Пакет аналіза*” вважається, що повторення для всіх варіантів однакові. У нас усі $14 \times 3 = 42$ варіанти реалізовані в одному повторенні. Пропуски даних тепер не допускаються.

Викликаємо процедуру “*Двухфакторный дисперсионный анализ без повторений*”. Якщо у полі *Входной интервал* вказувати лише числовий масив, то не треба ставити прапорець у полі *Метки*, інакше в діапазон даних треба буде включати рядок і стовпець заголовків. Інші поля введення заповнюються як звичайно. Нова таблиця *Итоги* просто поєднає попередню інформацію з рядків і стовпців. Розглянемо нову таблицю дисперсійного аналізу:

| Двухфакторный дисперсионный анализ без повторений | | | | | | |
|--|-----------|-----------|-----------|----------|---------------|-----------------|
| <i>Изменчивость</i> | <i>SS</i> | <i>df</i> | <i>MS</i> | <i>F</i> | <i>P-Знач</i> | <i>F критич</i> |
| Строки | 4717,1 | 13 | 362,86 | 15,582 | 4E-09 | 2,1192 |
| Столбцы | 123,33 | 2 | 61,666 | 2,6481 | 0,0898 | 3,369 |
| Погрешность | 605,47 | 26 | 23,287 | | | |
| Общая | 5446 | 41 | | | | |

Рис. 78. Таблиця двофакторного дисперсійного аналізу

Статистичні висновки практично не змінилися – є значимі розбіжності між рядками (різними варіантами внесення добрив), тому що для них $P\text{-знач} = 4E-09 < 0,01$; між стовпцями (повтореннями досліду) значимих розходжень немає, тому що для них $P\text{-знач} = 0,0898 > 0,05$. Проте міра випадкової мінливості тепер дорівнює $MS_{\varepsilon} = 23,29$ (порівняйте з попереднім значенням $MS_{\varepsilon} = 26,029$, коли враховувалися розходження лише за рядками).

У процедурі “*Двухфакторный дисперсионный анализ с повторениями*” потрібно ще вказувати *Число строк для выборки* – число повторень, яке для наших даних дорівнює 1. У діапазон *Входной интервал* необхідно включати рядок і стовпець заголовків (обов'язково), тому поле *Метки* відсутнє. У цієї процедури крім головних ефектів додатково оцінюється ефект взаємодії між класифікаційними факторами.

Завдання для самостійної роботи

За допомогою однофакторного дисперсійного аналізу перевірити значущість різниць між групами спостережень (за різні роки, чи дані різних підприємств).

емств). За допомогою двофакторного дисперсійного аналізу оцініть головні ефекти А, В і ефект взаємодії АВ двох якісних (класифікаційних) факторів.

Варіанти індивідуальних завдань наведені у додатку А.

Контрольні запитання

1. Сформулюйте мету і основні передумови дисперсійного аналізу.

2. Опишіть стандартну таблицю дисперсійного аналізу і пояснить її окремі елементи: суми квадратів (відхилень), число ступенів свободи, середні квадрати, дисперсійне відношення Фішера.

3. Що таке "індекс детермінації" і "кореляційне відношення"? Виразіть дисперсійне відношення і незміщену оцінку випадкової дисперсії через індекс детермінації.

4. Якщо пояснююча змінна в регресійному аналізі – дискретна, можна перевірити адекватність прийнятої форми зв'язку. Як це можна зробити? Якщо пояснююча змінна – одна, але неперервна, то як застосувати методику перевірки адекватності моделі?

5. Яка різниця між "індексом детермінації" і "коефіцієнтом детермінації"? Що показують ці характеристики?

6. Опишіть критерій Фішера для оцінки значущості дисперсійного відношення. Що таке "рівень значущості", як його знайти, які висновки можна зробити за його величиною? Що таке "квантілі розподілу Фішера", як їх знайти і які висновки можна зробити, порівнюючи їх з дисперсійним відношенням?

7. Якщо встановлено, що є значимі різниці між групами спостережень, то як визначити, між саме якими групами дійсно є значимі різниці? Опишіть методику Ст'юдента для порівняння кожної пари груп. Які висновки можна зробити, розглядаючи графік середніх групових разом з границями 95%-х довірчих інтервалів?

Використана література

1. Додж М. Эффективная работа с Excel 7.0 для Windows 95 / М. Додж, К. Кината, К. Стинсон. – СПб: Питер, 1996. –1040 с.
2. Доугерти К. Введение в эконометрику / К. Доугерти. Пер. с англ. – М. : ИНФРА-М, 1999. –XIV, 402 с.
3. Егоршин А. Корреляционно–регрессионный анализ. Курс лекций и лабораторных работ: учебное пособие для ВУЗов / А.А. Егоршин, Л.М. Малярец. – Харьков : Основа, 1998. –208 с.
4. Егоршин А. Практикум по эконометрии в Excel: учебное пособие для экономических вузов / А.А. Егоршин, Л.М. Малярец. Х. : ИД "ИНЖЭК", 2005. – 100 с.

Додатки

Дані для індивідуальних завдань

Додаток А

Перелік показників (змінних)

| | |
|-----|---|
| P1 | Коеф. поточної ліквідності |
| P2 | Коеф. абсолютної ліквідності |
| P3 | Коеф. автономії |
| P4 | Коеф. достатності фінансування |
| P5 | Частка витрат на оплату праці в загальних витратах |
| P6 | Співвідношен. темпів зростання продуктивн. праці та заробітн. плати |
| P7 | Частка управлінського персоналу у загальній чисельності |
| P8 | Коеф. оборотності кредиторської заборгованості |
| P9 | Коеф. оборотності дебіторської заборгованості |
| P10 | Частка фінансових інвестицій у загальній сумі активів |
| P11 | Частка матеріальних витрат в загальних витратах |
| P12 | Коеф. рентабельності продукції |
| P13 | Коеф. рентабельності активів |
| P14 | Коеф. рентабельності власного капіталу |
| P15 | Плинність кадрів |
| P16 | Податкове навантаження |
| P17 | Ефективність збуту |
| P18 | Частка нереалізованої продукції в запасах |

Значення показників фінансової діяльності за 2004 рік

| Змінні | Підприємства | | | | | | | | |
|--------|--------------|--------|--------|-------|--------|--------|--------|-------|-------|
| | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 |
| P1 | 1,544 | 27,646 | 0,86 | 1,13 | 1,127 | 53,286 | 38,563 | 1,006 | 1,214 |
| P2 | 0,167 | 0,231 | 0,07 | 0,01 | 0,124 | 2,229 | 1,375 | 0,119 | 0,192 |
| P3 | 0,563 | 0,964 | 0,29 | 0,116 | 0,136 | 0,981 | 0,974 | 0,288 | 0,279 |
| P4 | 3,684 | | 3,979 | 1,193 | 2,286 | 1,554 | | 1,863 | 2,041 |
| P5 | 8,595 | | 5,898 | 1,504 | 2,647 | 144,9 | | 3,653 | 2,996 |
| P6 | 16,241 | | 33,529 | 1,358 | 2,639 | 1,622 | | 3,971 | 3,589 |
| P7 | 0,16 | -0,001 | 0,05 | 5,515 | 0,006 | 0,007 | | 0,096 | 0,036 |
| P8 | 0,005 | | 0,03 | 0,001 | 0,026 | | | 0,003 | 0,029 |
| P9 | 0,165 | 0,18 | 0,15 | 0,238 | 0,169 | 0,162 | 0,144 | 0,183 | 0,124 |
| P10 | 0,982 | 0,825 | 0,862 | 0,969 | 0,857 | 1,08 | 0,921 | 1,057 | 1,224 |
| P11 | 0,533 | 0 | 0,82 | 0,015 | 0,624 | 0,017 | 0 | 0,638 | 0,591 |
| P12 | 0,124 | 0,027 | 0,112 | 0,08 | 0,019 | 0,025 | 0,014 | 0,085 | 0,046 |
| P13 | 0,269 | 0,19 | 0,257 | 0,16 | 0,188 | 0,215 | 0,274 | 0,142 | 0,281 |
| P14 | 0,284 | 0,121 | 0,229 | | | | | 0,497 | 0,123 |
| P15 | 0,68 | | 0,612 | 1,263 | -4,256 | | | 0,499 | 4,533 |
| P16 | 2,443 | | 2,262 | 1,676 | 1,05 | 1,103 | 0 | 1,486 | 1,539 |
| P17 | 0,193 | 0,28 | 0,176 | 0,186 | 0,181 | 0,1 | 0,2 | 0,193 | 0,156 |
| P18 | 0,75 | -0,041 | 0,592 | 0,945 | 0,799 | 0,969 | -0,015 | 0,58 | 0,104 |

Значення показників фінансової діяльності за 2005 рік

| Змінні | Підприємства | | | | | | | | |
|--------|--------------|--------|-------|--------|-------|-------|--------|--------|--------|
| | I | II | III | IV | V | VI | VII | VIII | IX |
| P1 | 0,997 | 1,081 | 1,037 | 1,045 | 1,402 | 1,945 | 1,381 | 1,152 | 0,992 |
| P2 | 0,175 | 0,089 | 0,01 | 0,02 | 0,051 | 0,003 | 0,003 | 0,013 | 0,069 |
| P3 | 0,029 | 0,432 | 0,045 | 0,262 | 0,31 | 0,49 | 0,302 | 0,53 | 0,485 |
| P4 | 0,83 | 0,184 | 0,893 | 0,189 | 0,954 | 1,797 | 0,227 | 3,106 | 3,481 |
| P5 | 4,974 | 0,319 | 5,076 | 0,386 | 3,744 | 3,524 | 0,325 | 9,666 | 6,815 |
| P6 | 1,01 | 3,671 | 1,031 | 0,65 | 4,787 | 3,122 | 0,388 | 14,594 | 20,561 |
| P7 | 0,262 | 0,018 | 0,27 | | 0,189 | 0,029 | 0,019 | 0,057 | 0,033 |
| P8 | 0,032 | 0,004 | 0,033 | | 0,041 | 0,047 | 0,004 | 0,132 | 0,088 |
| P8 | | 0,001 | | 0,029 | 0,002 | 0,005 | 0,005 | | |
| P9 | 0,188 | 0,253 | 0,194 | 0,18 | 0,151 | 0,238 | 0,261 | 0,163 | 0,085 |
| P10 | 0,97 | 1,063 | 1 | 1,141 | 1,073 | 1,057 | 1,096 | 1,321 | 1,067 |
| P11 | 0,678 | | 0,699 | 0,377 | 0,442 | | | 0,57 | 0,533 |
| P12 | 0,015 | 0,065 | 0,016 | 0,056 | 0,086 | 0,021 | 0,066 | 0,087 | 0,065 |
| P13 | 0,239 | 0,172 | 0,244 | 0,2 | 0,198 | 0,176 | 0,176 | 0,243 | 0,214 |
| P14 | 0,084 | | 0,086 | 0,002 | 0,111 | | | 0,492 | 0,346 |
| P15 | 0,906 | 28,474 | 0,925 | 0,347 | 0,092 | 2,118 | 29,055 | 2,315 | 4,191 |
| P16 | 1,882 | 8,63 | 1,92 | 14,688 | 2,903 | 1,113 | 8,806 | 2,897 | 2,585 |
| P17 | 0,181 | 0,168 | 0,185 | 0,177 | 0,198 | 0,175 | 0,172 | 0,169 | 0,166 |
| P18 | 0,703 | 0,013 | 0,725 | | 0,133 | 0,096 | 0,014 | 0,249 | 0,181 |

Значення показників фінансової діяльності за 2006 рік

| Змінні | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 |
|--------|--------|-------|-------|-------|-------|--------|--------|-------|-------|
| P1 | 1,578 | 1,19 | 1,091 | 2,007 | 1,287 | 2,075 | 1,358 | 1,353 | 1,06 |
| P2 | 0,25 | 0,188 | 0,118 | 0,01 | 0,006 | 0,02 | 0,027 | 0,149 | 0,053 |
| P3 | 0,362 | 0,273 | 0,389 | 1,394 | 0,753 | 0,249 | 0,244 | 0,163 | 0,053 |
| P4 | 2,653 | 2 | 0,164 | 2,77 | 2,827 | 0,172 | 0,168 | 2,744 | 0,623 |
| P5 | 3,894 | 2,936 | 0,309 | 9,189 | 9,473 | 0,378 | 0,367 | 3,176 | 4,825 |
| P6 | 4,666 | 3,517 | 0,442 | 0,451 | 0,625 | 6,943 | 4,551 | 3,167 | 0,933 |
| P7 | 0,182 | 0,06 | 0,016 | 0,05 | 0,055 | 0 | 0 | 0,007 | 0,239 |
| P8 | 0,422 | -0,04 | 0,172 | 0,11 | 0,109 | 0,951 | -0,015 | 0,167 | 0,154 |
| P9 | 0,308 | 0,036 | | | | 0,003 | 0,003 | 0,034 | |
| P10 | 0,62 | 0,18 | 0,248 | 0,155 | 0,158 | 0,174 | 0,171 | 0,22 | 0,184 |
| P11 | 10,026 | 1,035 | 1,042 | 1,256 | 1,281 | 1,107 | 1,085 | 1,114 | 0,951 |
| P12 | 7,302 | 0,983 | | 0,542 | 0,553 | 0,365 | 0,358 | 0,811 | 0,665 |
| P13 | 0,229 | 0,135 | 0,059 | 0,078 | 0,086 | 0,055 | 0,05 | 0,025 | 0,014 |
| P14 | 2,198 | 0,308 | 0,157 | 0,217 | 0,238 | 0,196 | 0,178 | 0,244 | 0,218 |
| P15 | 0,988 | 0,275 | | 0,439 | 0,482 | 0,002 | 0,002 | 0,11 | 0,076 |
| P16 | 22,666 | 0,734 | 27,62 | 2,2 | 2,268 | 0,34 | 0,329 | 5,533 | 0,879 |
| P17 | 12,281 | 2,714 | 8,371 | 2,754 | 2,839 | 14,394 | 13,963 | 1,365 | 1,825 |
| P18 | 1,248 | 0,211 | 0,163 | 0,16 | 0,165 | 0,173 | 0,168 | 0,139 | 0,175 |
| P19 | 0,518 | 0,711 | 0,012 | 0,22 | 0,241 | 0,015 | 0,014 | 1,039 | 0,64 |

Значення показників фінансової діяльності за 2007 рік

| Змінні | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 |
|--------|-------|--------|--------|-------|-------|--------|-------|-------|-------|
| P1 | 2,334 | 1,823 | 1,307 | 1,488 | 1,527 | 1,797 | 0,985 | 1,265 | 1,356 |
| P2 | 0,003 | 0,066 | 0,155 | 0,177 | 0,242 | 0,285 | 0,117 | 0,15 | 0,012 |
| P3 | 0,588 | 0,403 | 0,374 | 0,425 | 0,35 | 0,412 | 0,282 | 0,362 | 0,14 |
| P4 | 2,156 | 1,24 | 2,422 | 2,757 | 2,567 | 3,02 | 1,826 | 2,344 | 1,432 |
| P5 | 4,229 | 4,867 | 4,748 | 5,405 | 3,768 | 4,433 | 3,58 | 4,594 | 1,805 |
| P6 | 3,746 | 6,224 | 5,163 | 5,877 | 4,515 | 5,311 | 3,892 | 4,995 | 1,629 |
| P7 | 0,038 | 0,338 | 64,521 | 9,32 | 0,007 | 0,01 | 0,192 | 7,026 | 7,169 |
| P8 | 1,274 | 0,206 | 0,004 | 0,116 | 0,128 | 0 | 0 | 0,142 | 0,029 |
| P9 | 0,006 | 0,055 | 0,014 | 0,002 | 0,034 | 0,044 | 0,006 | 0,002 | 0,002 |
| P10 | 0,309 | 2,782 | 2,786 | 0,402 | 0,215 | 0,286 | 0,198 | 0,303 | 0,31 |
| P11 | 1,374 | 12,365 | 11,334 | 1,637 | 1,092 | 1,448 | 1,179 | 1,234 | 1,259 |
| P12 | 0,523 | 4,704 | 0,178 | 0,026 | 0,795 | 1,055 | 0,639 | 0,019 | 0,02 |
| P13 | 0,028 | 0,249 | 0,937 | 0,135 | 0,024 | 0,032 | 0,148 | 0,102 | 0,104 |
| P14 | 0,228 | 2,055 | 1,871 | 0,27 | 0,239 | 0,318 | 0,322 | 0,204 | 0,208 |
| P15 | 0,098 | 0,886 | 3,981 | 0,575 | 0,108 | 0,143 | 0,34 | 0,433 | 0,442 |
| P16 | 2,753 | 24,779 | 14,774 | 2,134 | 5,422 | -7,192 | 0,816 | 1,609 | 1,642 |
| P17 | 1,447 | 13,023 | 19,609 | 2,832 | 1,337 | 1,774 | 2,932 | 2,135 | 2,179 |
| P18 | 0,176 | 1,582 | 1,544 | 0,223 | 0,136 | 0,18 | 0,232 | 0,168 | 0,172 |
| P19 | 0,125 | 1,122 | 11,055 | 1,597 | 1,018 | 1,351 | 0,9 | 1,204 | 1,228 |

Значення показників фінансової діяльності за 2008 рік

| Змінні | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 |
|--------|--------|--------|--------|-------|-------|-------|---------|--------|-------|
| P1 | 1,314 | 1,297 | 1,657 | 1,024 | 2,075 | 1,374 | 1,358 | 1,546 | 1,764 |
| P2 | 0,026 | 0,107 | 0,003 | 0,02 | 0,075 | 0,05 | 0,027 | 0,03 | 0,064 |
| P3 | 0,33 | 0,519 | 0,362 | 0,257 | 0,459 | 0,304 | 0,341 | 0,388 | 0,39 |
| P4 | 0,237 | 0,221 | 0,273 | 0,185 | 1,412 | 0,935 | 0,245 | 0,279 | 1,2 |
| P5 | 0,485 | 0,383 | 0,39 | 0,378 | 5,54 | 3,669 | 0,501 | 0,571 | 4,709 |
| P6 | 0,818 | 0,456 | 0,465 | 0,637 | 7,084 | 4,692 | 0,845 | 0,962 | 6,022 |
| P7 | 0,024 | 0,023 | 0,024 | 0,068 | 0,049 | 0,039 | 0,218 | 0,031 | 0,037 |
| P8 | 0,061 | 0,55 | 1,287 | 0,186 | 0,139 | 0,184 | 0,507 | 0,14 | 0,143 |
| P9 | 0,007 | 0,001 | 0,007 | 0,034 | 0,008 | 0,031 | 0,061 | 0,009 | 0,006 |
| P10 | 0,332 | 0,329 | 0,339 | 0,195 | 0,402 | 0,102 | 3,053 | 0,441 | 0,303 |
| P11 | 1,396 | 1,382 | 1,425 | 1,585 | 1,786 | 1,281 | 12,821 | 1,852 | 1,346 |
| P12 | 0,547 | 0,542 | 0,559 | 0,684 | 0,679 | 0,64 | 5,027 | 0,726 | 0,512 |
| P13 | 0,084 | 0,084 | 0,086 | 0,105 | 0,036 | 0,078 | 0,775 | 0,112 | 0,027 |
| P14 | 0,224 | 0,224 | 0,229 | 0,292 | 0,297 | 0,256 | 2,057 | 0,297 | 0,224 |
| P15 | 0,002 | 0,002 | 0,002 | 0,59 | 0,128 | 0,415 | 0,017 | 0,002 | 0,096 |
| P16 | 37,016 | 37,016 | 37,772 | 2,778 | 3,579 | 5,029 | 339,944 | 49,103 | 2,698 |
| P17 | 11,219 | 11,219 | 11,448 | 3,477 | 1,881 | 3,102 | 103,034 | 14,883 | 1,418 |
| P18 | 0,154 | 0,154 | 0,158 | 0,202 | 0,229 | 0,199 | 1,418 | 0,205 | 0,172 |
| P19 | 0,017 | 0,017 | 0,018 | 0,299 | 0,162 | 0,217 | 0,159 | 0,023 | 0,122 |

Значення показників фінансової діяльності за 2009 рік

| Змінні | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 |
|--------|--------|--------|--------|--------|-------|---------|--------|-------|--------|
| P1 | 1,527 | 1,298 | 1,314 | 1,342 | 1,024 | 1,358 | 1,309 | 1,012 | 1,546 |
| P2 | 0,017 | 0,014 | 0,026 | 0,015 | 0,02 | 0,027 | 0,118 | 0,011 | 0,03 |
| P3 | 0,362 | 0,307 | 0,314 | 0,318 | 0,244 | 0,324 | 0,467 | 0,239 | 0,369 |
| P4 | 0,249 | 0,211 | 0,216 | 0,219 | 0,168 | 0,223 | 0,196 | 0,165 | 0,254 |
| P5 | 0,543 | 0,461 | 0,475 | 0,477 | 0,37 | 0,491 | 0,371 | 0,359 | 0,559 |
| P6 | 0,915 | 0,778 | 0,802 | 0,804 | 0,625 | 0,828 | 0,442 | 0,606 | 0,943 |
| P7 | 0,028 | 0,021 | 0,023 | 0,192 | 0,066 | 0,211 | 0,021 | 0,06 | 0,03 |
| P8 | 0,005 | 0,005 | 0,005 | 0,158 | 0,079 | 0,105 | 0,048 | 0,007 | 0,06 |
| P9 | 0,001 | 0,001 | 0,001 | 0,006 | 0,003 | 0,006 | 0,001 | 0,003 | 0,001 |
| P10 | 0,419 | 0,316 | 0,322 | 2,902 | 0,189 | 2,961 | 0,322 | 0,186 | 0,428 |
| P11 | 1,76 | 1,327 | 1,354 | 12,188 | 1,538 | 12,436 | 1,354 | 1,507 | 1,796 |
| P12 | 0,69 | 0,52 | 0,531 | 4,778 | 0,663 | 4,876 | 0,531 | 0,65 | 0,704 |
| P13 | 0,1 | 0,075 | 0,083 | 0,691 | 0,103 | 0,759 | 0,077 | 0,093 | 0,11 |
| P14 | 0,265 | 0,2 | 0,219 | 1,834 | 0,286 | 2,016 | 0,204 | 0,26 | 0,291 |
| P15 | 0,002 | 0,002 | 0,002 | 0,015 | 0,578 | 0,016 | 0,002 | 0,526 | 0,002 |
| P16 | 46,677 | 35,187 | 36,276 | 323,15 | 2,722 | 333,145 | 35,906 | 2,64 | 48,121 |
| P17 | 14,147 | 10,665 | 10,995 | 97,944 | 3,407 | 100,973 | 10,883 | 3,305 | 14,585 |
| P18 | 0,195 | 0,147 | 0,151 | 1,348 | 0,198 | 1,389 | 0,15 | 0,192 | 0,201 |
| P19 | 0,02 | 0,015 | 0,017 | 0,14 | 0,29 | 0,154 | 0,016 | 0,264 | 0,022 |

Макроекономічні показники регіонів України

| | |
|-----|---|
| y1 | Кількість промислових підприємств, що займалися інноваційною діяльністю, од. |
| y2 | Кількість інноваційно активних підприємств у промисловості, од. |
| y3 | Освоєння нових видів продукції у промисловості, найменувань |
| y4 | Впровадження прогресивних технологічних процесів у промисловості, процесів |
| y5 | Подано заявок на видачу охоронних документів на об'єкти промислової власності (винаходи, корисні моделі, промислові зразки) – Державний департамент інтелектуальної власності України |
| y6 | Подано заявок на видачу охоронних документів на об'єкти промислової власності (винаходи, корисні моделі, промислові зразки) – Патентні відомства іноземних держав |
| y7 | Отримано охоронних документів на об'єкти промислової власності (винаходи, корисні моделі, промислові зразки) – Державний департамент інтелектуальної власності України |
| y8 | Отримано охоронних документів на об'єкти промислової власності (винаходи, корисні моделі, промислові зразки) – Патентні відомства іноземних держав |
| y9 | Чисельність фахівців, які виконують накові та науково-технічні роботи, осіб, з них мають науковий ступінь доктора наук |
| y10 | Чисельність фахівців, які виконують накові та науково-технічні роботи, осіб, з них мають науковий ступінь кандидата наук |
| y11 | Фахівців вищої кваліфікації, які зайняті в економіці України, осіб, з них мають науковий ступінь доктора наук |
| y12 | Фахівців вищої кваліфікації, які зайняті в економіці України, осіб, з них мають науковий ступінь кандидата наук |
| y13 | Кількість виїздів наукових працівників за межі України |
| y14 | Кількість вищих навчальних закладів 1 – 4 рівнів акредитації |
| y15 | Кількість студентів вищих навчальних закладів 1 – 4 рівнів акредитації у розрахунку на 10000 населення |
| y16 | Економічно активне населення у віці 15-70 років, тис. осіб |
| x1 | Інвестицій в основний капітал, млн грн |
| x2 | Обсяг інноваційної продукції, млн грн |
| x3 | Обсяг реалізованої інноваційної продукції поставленої на експорт, млн грн |
| x4 | Розподіл загального обсягу витрат на інноваційної діяльності, млн грн |
| x5 | Розподіл обсягу витрат на внутрішні науково-дослідні розробки, млн.грн. |
| x6 | Розподіл обсягу витрат на придбання науково-дослідні розробки, млн грн |
| x7 | Розподіл обсягу витрат на придбання машин, обладнання та програмного забезпечення |
| x8 | Розподіл обсягу витрат на придбання інших зовнішніх знань, млн грн |
| x9 | Розподіл обсягу витрат на інші витрати (маркетинг, реклама), млн грн |
| x10 | Валовий регіональний продукт, млн грн |
| x11 | Внутрішні поточні витрати на наукові та науково-технічні роботи, виконані власними силами наукових організацій, у тому числі: фундаментальні дослідження, млн грн |
| x12 | Внутрішні поточні витрати на наукові та науково-технічні роботи, виконані власними силами наукових організацій, у тому числі: прикладні дослідження, млн грн |
| x13 | Внутрішні поточні витрати на наукові та науково-технічні роботи, виконані власними силами наукових організацій, у тому числі: науково-технічні розробки, млн грн |
| x14 | Внутрішні поточні витрати на наукові та науково-технічні роботи, виконані власними силами наукових організацій, у тому числі: науково-технічні послуги, млн грн |
| x15 | Обсяг наукових та науково-технічних робіт, виконаних власними силами наукових організацій, млн грн |
| x16 | Фінансування наукових та науково-технічних робіт за регіонами, млн грн |
| z1 | Загальні обсяги імпорту послуг, млн дол. США |
| z2 | Загальні обсяги експорту послуг, млн дол. США |
| z3 | Загальні обсяги експорту товарів, млн дол. США |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | y1 | | | | | | | y2 | | | | | | |
|-------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 1430 | 1487 | 1505 | 1510 | 1452 | 1404 | 1378 | 1705 | 1808 | 1359 | 1193 | 1118 | 1472 | 1397 |
| Авт. Респ. Крим | 42 | 43 | 44 | 47 | 47 | 52 | 49 | 31 | 41 | 56 | 33 | 24 | 33 | 40 |
| Вінницька | 31 | 31 | 30 | 28 | 24 | 24 | 25 | 84 | 99 | 34 | 32 | 39 | 75 | 50 |
| Волинська | 17 | 19 | 16 | 16 | 19 | 16 | 13 | 29 | 56 | 15 | 14 | 13 | 36 | 24 |
| Дніпропетровська | 94 | 107 | 108 | 109 | 105 | 93 | 88 | 130 | 108 | 75 | 75 | 66 | 70 | 63 |
| Донецька | 101 | 86 | 83 | 81 | 77 | 74 | 69 | 109 | 112 | 92 | 77 | 61 | 82 | 83 |
| Житомирська | 14 | 11 | 12 | 12 | 9 | 11 | 10 | 80 | 79 | 36 | 42 | 41 | 37 | 41 |
| Закарпатська | 18 | 19 | 21 | 21 | 20 | 21 | 20 | 27 | 29 | 37 | 32 | 31 | 28 | 31 |
| Запорізька | 45 | 40 | 37 | 38 | 35 | 35 | 35 | 59 | 67 | 34 | 37 | 32 | 43 | 36 |
| Івано-Франківська | 18 | 21 | 12 | 21 | 21 | 23 | 27 | 34 | 45 | 44 | 40 | 38 | 92 | 92 |
| Київська | 38 | 39 | 37 | 36 | 37 | 36 | 35 | 82 | 67 | 77 | 41 | 43 | 52 | 62 |
| Кіровоградська | 13 | 12 | 11 | 11 | 11 | 12 | 12 | 19 | 22 | 65 | 37 | 28 | 39 | 39 |
| Луганська | 48 | 44 | 52 | 49 | 52 | 51 | 52 | 111 | 84 | 52 | 37 | 45 | 67 | 59 |
| Львівська | 90 | 95 | 90 | 87 | 82 | 84 | 79 | 97 | 88 | 58 | 44 | 62 | 127 | 101 |
| Миколаївська | 46 | 38 | 40 | 41 | 54 | 54 | 51 | 65 | 66 | 44 | 31 | 27 | 26 | 55 |
| Одеська | 72 | 73 | 73 | 74 | 66 | 65 | 67 | 19 | 57 | 42 | 49 | 54 | 46 | 49 |
| Полтавська | 28 | 29 | 26 | 32 | 30 | 25 | 24 | 79 | 97 | 41 | 40 | 30 | 64 | 47 |
| Рівненська | 18 | 16 | 17 | 17 | 15 | 13 | 14 | 64 | 53 | 25 | 10 | 8 | 22 | 29 |
| Сумська | 25 | 24 | 27 | 29 | 27 | 25 | 24 | 61 | 49 | 17 | 12 | 12 | 14 | 30 |
| Тернопільська | 15 | 19 | 17 | 16 | 16 | 14 | 14 | 40 | 44 | 33 | 29 | 21 | 36 | 43 |
| Харківська | 217 | 241 | 233 | 227 | 228 | 218 | 214 | 128 | 171 | 120 | 133 | 143 | 142 | 110 |
| Херсонська | 26 | 21 | 23 | 28 | 26 | 25 | 23 | 26 | 36 | 28 | 25 | 25 | 29 | 28 |
| Хмельницька | 6 | 6 | 7 | 6 | 6 | 7 | 6 | 47 | 48 | 23 | 21 | 22 | 25 | 16 |
| Черкаська | 33 | 32 | 32 | 30 | 29 | 31 | 30 | 29 | 48 | 29 | 29 | 23 | 27 | 37 |
| Чернівецька | 17 | 40 | 23 | 24 | 21 | 22 | 22 | 69 | 64 | 33 | 29 | 26 | 32 | 30 |
| Чернігівська | 25 | 23 | 22 | 22 | 22 | 22 | 22 | 73 | 88 | 29 | 24 | 25 | 41 | 42 |
| м. Київ | 375 | 364 | 389 | 393 | 360 | 340 | 339 | 98 | 78 | 215 | 213 | 171 | 180 | 155 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | у3 | | | | | | | у4 | | | | | | |
|-------------------|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 15323 | 7416 | 3978 | 3152 | 2408 | 2526 | 2446 | 1403 | 1482 | 1727 | 1808 | 1145 | 1419 | 1647 |
| Авт. Респ. Крим | 504 | 316 | 356 | 63 | 54 | 55 | 48 | 68 | 60 | 90 | 62 | 11 | 24 | 73 |
| Вінницька | 1008 | 313 | 34 | 30 | 37 | 124 | 91 | 32 | 50 | 33 | 14 | 24 | 79 | 58 |
| Волинська | 473 | 159 | 12 | 14 | 8 | 4 | 10 | 88 | 32 | 12 | 4 | 8 | 24 | 16 |
| Дніпропетровська | 500 | 89 | 92 | 88 | 99 | 151 | 136 | 149 | 80 | 62 | 79 | 65 | 53 | 58 |
| Донецька | 866 | 321 | 151 | 172 | 132 | 80 | 150 | 158 | 117 | 115 | 90 | 54 | 52 | 114 |
| Житомирська | 452 | 94 | 89 | 90 | 38 | 35 | 24 | 40 | 32 | 31 | 25 | 14 | 31 | 30 |
| Закарпатська | 110 | 62 | 57 | 33 | 34 | 54 | 22 | 1 | 14 | 21 | 10 | 4 | 11 | 13 |
| Запорізька | 230 | 135 | 101 | 86 | 108 | 109 | 148 | 130 | 86 | 210 | 351 | 252 | 234 | 257 |
| Івано-Франківська | 536 | 308 | 118 | 107 | 61 | 69 | 56 | 57 | 21 | 16 | 36 | 12 | 47 | 35 |
| Київська | 649 | 147 | 117 | 90 | 73 | 85 | 122 | 33 | 21 | 45 | 41 | 23 | 52 | 34 |
| Кіровоградська | 99 | 111 | 56 | 33 | 66 | 56 | 63 | 2 | 44 | 40 | 42 | 31 | 52 | 31 |
| Луганська | 258 | 92 | 34 | 42 | 33 | 34 | 51 | 40 | 49 | 43 | 35 | 20 | 56 | 45 |
| Львівська | 1242 | 1984 | 182 | 183 | 102 | 190 | 138 | 82 | 83 | 73 | 55 | 8 | 54 | 54 |
| Миколаївська | 713 | 254 | 38 | 85 | 8 | 12 | 26 | 24 | 16 | 13 | 18 | 14 | 9 | 20 |
| Одеська | 140 | 311 | 245 | 214 | 300 | 30 | 50 | 14 | 66 | 29 | 40 | 30 | 51 | 33 |
| Полтавська | 800 | 88 | 108 | 110 | 160 | 142 | 90 | 62 | 30 | 28 | 65 | 29 | 35 | 26 |
| Рівненська | 253 | 181 | 84 | 78 | 43 | 59 | 16 | 11 | 5 | 9 | 9 | 1 | 4 | 20 |
| Сумська | 353 | 333 | 205 | 143 | 166 | 165 | 169 | 82 | 35 | 57 | 55 | 67 | 56 | 77 |
| Тернопільська | 574 | 554 | 403 | 223 | 19 | 65 | 89 | 26 | 19 | 13 | 14 | 2 | 36 | 30 |
| Харківська | 810 | 219 | 191 | 161 | 137 | 175 | 167 | 126 | 172 | 108 | 112 | 116 | 96 | 95 |
| Херсонська | 280 | 114 | 72 | 62 | 27 | 33 | 95 | 20 | 16 | 12 | 10 | 5 | 22 | 61 |
| Хмельницька | 405 | 72 | 31 | 19 | 20 | 7 | 8 | 26 | 55 | 12 | 13 | 4 | 8 | 16 |
| Черкаська | 301 | 15 | 52 | 22 | 22 | 37 | 57 | 4 | 2 | 13 | 8 | 3 | 13 | 23 |
| Чернівецька | 696 | 167 | 75 | 103 | 76 | 111 | 79 | 23 | 20 | 8 | 8 | 6 | 15 | 11 |
| Чернігівська | 935 | 267 | 59 | 47 | 41 | 38 | 58 | 40 | 30 | 37 | 31 | 31 | 32 | 34 |
| м. Київ | 2052 | 612 | 998 | 833 | 538 | 597 | 477 | 64 | 326 | 596 | 581 | 311 | 273 | 381 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | у5 | | | | | | | у6 | | | | | | |
|-------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 4238 | 7703 | 7627 | 7639 | 7761 | 8564 | 8446 | 194 | 336 | 222 | 257 | 412 | 345 | 168 |
| Авт. Респ. Крим | 68 | 181 | 391 | 163 | 190 | 171 | 172 | 2 | 6 | 3 | 6 | 7 | 6 | 4 |
| Вінницька | 140 | 176 | 219 | 249 | 301 | 309 | 348 | - | 3 | 1 | 1 | 2 | - | - |
| Волинська | 17 | 39 | 29 | 50 | 40 | 56 | 61 | 2 | - | - | - | - | - | - |
| Дніпропетровська | 350 | 840 | 716 | 685 | 597 | 625 | 592 | 2 | 25 | 22 | 51 | 42 | 29 | 19 |
| Донецька | 326 | 714 | 714 | 870 | 907 | 839 | 774 | 40 | 51 | 50 | 37 | 85 | 44 | 22 |
| Житомирська | 15 | 38 | 35 | 31 | 29 | 23 | 14 | - | - | - | - | - | - | - |
| Закарпатська | 35 | 75 | 53 | 38 | 58 | 69 | 58 | - | - | - | - | - | - | - |
| Запорізька | 194 | 372 | 356 | 315 | 318 | 307 | 297 | 4 | 13 | 18 | 23 | 23 | 7 | 12 |
| Івано-Франківська | 126 | 245 | 129 | 148 | 209 | 222 | 234 | 4 | 4 | 8 | 4 | 3 | 6 | 3 |
| Київська | 29 | 100 | 91 | 99 | 81 | 120 | 77 | - | 3 | 1 | 3 | 4 | 5 | 6 |
| Кіровоградська | 30 | 67 | 48 | 64 | 63 | 104 | 74 | - | 11 | 5 | - | 1 | 5 | 5 |
| Луганська | 127 | 279 | 312 | 254 | 228 | 223 | 309 | 6 | 2 | 2 | 3 | 1 | - | 0 |
| Львівська | 222 | 298 | 276 | 311 | 263 | 361 | 396 | 4 | 2 | 2 | - | - | - | 2 |
| Миколаївська | 62 | 122 | 86 | 155 | 145 | 184 | 169 | 1 | 1 | - | 2 | 2 | - | - |
| Одеська | 199 | 330 | 314 | 307 | 340 | 399 | 403 | 3 | 14 | - | 1 | 1 | - | 2 |
| Полтавська | 99 | 180 | 152 | 203 | 218 | 208 | 230 | 17 | 28 | 21 | 36 | 7 | 17 | 22 |
| Рівненська | 19 | 65 | 57 | 72 | 68 | 64 | 79 | 8 | 7 | 5 | 6 | 1 | 5 | 3 |
| Сумська | 77 | 71 | 85 | 64 | 84 | 76 | 89 | 3 | 8 | 10 | 11 | 3 | 8 | 6 |
| Тернопільська | 90 | 131 | 150 | 106 | 114 | 126 | 191 | - | - | - | 2 | - | - | - |
| Харківська | 636 | 1091 | 1138 | 1131 | 1007 | 1140 | 1096 | 41 | 26 | 44 | 35 | 146 | 87 | 31 |
| Херсонська | 95 | 92 | 132 | 131 | 117 | 108 | 119 | - | - | - | 1 | 1 | - | 1 |
| Хмельницька | 41 | 45 | 44 | 128 | 94 | 80 | 119 | 1 | - | - | - | - | - | - |
| Черкаська | 67 | 84 | 77 | 201 | 79 | 52 | 39 | 5 | 1 | 3 | - | - | 9 | - |
| Чернівецька | 50 | 104 | 88 | 127 | 114 | 126 | 108 | - | - | 3 | 5 | 14 | 7 | 6 |
| Чернігівська | 27 | 29 | 22 | 27 | 33 | 51 | 55 | - | - | - | - | 2 | - | - |
| м. Київ | 1064 | 1879 | 1033 | 1639 | 1981 | 2455 | 2300 | 42 | 127 | 22 | 21 | 49 | 105 | 18 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | у7 | | | | | | | у8 | | | | | | |
|-------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 3256 | 7312 | 7627 | 7337 | 7632 | 8488 | 8243 | 132 | 193 | 222 | 245 | 264 | 281 | 152 |
| Авт. Респ. Крим | 46 | 160 | 391 | 148 | 169 | 144 | 128 | - | 3 | 3 | 3 | 1 | 1 | 1 |
| Вінницька | 76 | 226 | 219 | 309 | 311 | 256 | 362 | - | - | 1 | - | 2 | - | - |
| Волинська | 20 | 34 | 29 | 42 | 42 | 43 | 41 | - | - | - | - | 2 | 21 | 1 |
| Дніпропетровська | 244 | 679 | 716 | 605 | 620 | 616 | 616 | 6 | 15 | 22 | 37 | 22 | 74 | 27 |
| Донецька | 191 | 690 | 714 | 798 | 925 | 871 | 869 | 38 | 54 | 50 | 58 | 56 | - | 26 |
| Житомирська | 10 | 33 | 35 | 15 | 27 | 22 | 27 | - | - | - | - | - | - | - |
| Закарпатська | 39 | 63 | 53 | 46 | 49 | 55 | 59 | - | - | - | - | 26 | 25 | - |
| Запорізька | 144 | 292 | 363 | 316 | 330 | 359 | 277 | 3 | 4 | 18 | 23 | 14 | 4 | 11 |
| Івано-Франківська | 93 | 210 | 129 | 128 | 163 | 224 | 246 | - | 1 | 8 | - | 16 | 7 | - |
| Київська | 29 | 116 | 91 | 100 | 99 | 128 | 84 | 2 | 1 | 1 | 2 | 3 | 6 | 3 |
| Кіровоградська | 31 | 66 | 48 | 102 | 43 | 90 | 78 | - | 2 | 5 | 2 | 6 | - | - |
| Луганська | 80 | 283 | 312 | 320 | 312 | 280 | 346 | 1 | 3 | 2 | 8 | 3 | 2 | - |
| Львівська | 148 | 292 | 276 | 307 | 280 | 369 | 319 | 8 | 1 | 2 | 2 | 2 | - | 4 |
| Миколаївська | 43 | 117 | 86 | 131 | 133 | 167 | 155 | 2 | 2 | - | 1 | 1 | 1 | - |
| Одеська | 163 | 290 | 314 | 401 | 320 | 338 | 311 | 1 | 2 | - | 1 | 22 | 9 | - |
| Полтавська | 103 | 154 | 152 | 216 | 208 | 236 | 197 | 4 | 31 | 21 | 27 | 8 | 4 | 11 |
| Рівненська | 35 | 77 | 57 | 72 | 52 | 54 | 66 | 3 | 6 | 5 | 6 | 6 | 7 | 1 |
| Сумська | 37 | 96 | 85 | 61 | 71 | 76 | 79 | 1 | 6 | 10 | 10 | - | - | 7 |
| Тернопільська | 91 | 195 | 150 | 162 | 96 | 138 | 170 | - | - | - | 2 | 32 | 57 | - |
| Харківська | 474 | 1150 | 1138 | 1016 | 994 | 1068 | 1027 | 32 | 27 | 44 | 27 | 1 | - | 38 |
| Херсонська | 62 | 72 | 132 | 124 | 104 | 97 | 112 | - | - | - | - | - | - | 1 |
| Хмельницька | 25 | 45 | 44 | 84 | 73 | 67 | 107 | 1 | - | - | - | - | - | - |
| Черкаська | 100 | 76 | 77 | 78 | 83 | 61 | 47 | - | 2 | 3 | 3 | - | 8 | - |
| Чернівецька | 40 | 47 | 88 | 114 | 107 | 107 | 121 | 1 | - | 3 | 3 | 5 | - | 1 |
| Чернігівська | 20 | 24 | 82 | 26 | 30 | 44 | 45 | - | - | - | - | - | 48 | - |
| м. Київ | 907 | 1790 | 1833 | 1584 | 1931 | 2503 | 2281 | 24 | 30 | 22 | 23 | 30 | 7 | 15 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | y9 | | | | | | | y10 | | | | | | |
|--------------------------|------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 4103 | 4005 | 4114 | 4180 | 4315 | 4389 | 4453 | 17916 | 16756 | 16983 | 16983 | 16968 | 16952 | 17087 |
| Авт. Респ. Крим | 76 | 64 | 71 | 69 | 71 | 75 | 87 | 420 | 362 | 359 | 360 | 373 | 381 | 384 |
| Вінницька | 13 | 8 | 7 | 7 | 8 | 9 | 8 | 73 | 72 | 73 | 68 | 70 | 71 | 74 |
| Волинська | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 21 | 14 | 14 | 13 | 15 | 18 | 16 |
| Дніпропетровська | 139 | 142 | 150 | 138 | 141 | 146 | 141 | 932 | 895 | 903 | 859 | 832 | 801 | 783 |
| Донецька | 180 | 141 | 148 | 151 | 163 | 171 | 167 | 956 | 800 | 774 | 750 | 755 | 764 | 746 |
| Житомирська | 7 | 4 | 4 | 4 | 4 | 4 | 4 | 56 | 51 | 49 | 39 | 43 | 48 | 46 |
| Закарпатська | 20 | 24 | 25 | 25 | 27 | 31 | 30 | 158 | 154 | 144 | 135 | 147 | 154 | 145 |
| Запорізька | 14 | 10 | 10 | 11 | 11 | 12 | 13 | 186 | 155 | 136 | 123 | 122 | 122 | 115 |
| Івано-Франківська | 16 | 18 | 8 | 3 | 3 | 4 | 3 | 104 | 109 | 90 | 61 | 57 | 51 | 53 |
| Київська | 62 | 67 | 68 | 64 | 67 | 70 | 74 | 371 | 368 | 361 | 343 | 351 | 359 | 352 |
| Кіровоградська | 17 | - | - | 1 | 1 | 2 | 2 | 38 | 11 | 9 | 12 | 17 | 19 | 22 |
| Луганська | 11 | 10 | 6 | 7 | 7 | 7 | 10 | 163 | 124 | 123 | 102 | 97 | 86 | 90 |
| Львівська | 174 | 188 | 188 | 196 | 201 | 209 | 215 | 1003 | 1024 | 1035 | 1017 | 1021 | 1027 | 1028 |
| Миколаївська | 5 | 9 | 9 | 7 | 12 | 16 | 14 | 110 | 77 | 77 | 75 | 87 | 93 | 91 |
| Одеська | 134 | 152 | 153 | 142 | 143 | 144 | 142 | 632 | 642 | 636 | 605 | 605 | 603 | 599 |
| Полтавська | 7 | 9 | 8 | 8 | 15 | 20 | 17 | 115 | 96 | 93 | 89 | 115 | 139 | 163 |
| Рівненська | 4 | 1 | 1 | 1 | 2 | 2 | 2 | 24 | 19 | 17 | 14 | 15 | 16 | 17 |
| Сумська | 9 | 15 | 22 | 24 | 21 | 18 | 17 | 130 | 105 | 267 | 259 | 157 | 108 | 119 |
| Тернопільська | 1 | 1 | 2 | 3 | 4 | 4 | 1 | 33 | 30 | 32 | 29 | 29 | 28 | 25 |
| Харківська | 555 | 539 | 554 | 575 | 591 | 602 | 619 | 2746 | 2636 | 2628 | 2764 | 2758 | 2739 | 2759 |
| Херсонська | 16 | 10 | 10 | 18 | 18 | 17 | 21 | 130 | 116 | 115 | 124 | 119 | 111 | 101 |
| Хмельницька | - | - | 1 | 2 | 1 | - | - | 11 | 8 | 13 | 13 | 14 | 14 | 10 |
| Черкаська | 4 | 7 | 7 | 7 | 7 | 7 | 6 | 52 | 39 | 45 | 50 | 58 | 65 | 67 |
| Чернівецька | 14 | 20 | 21 | 22 | 24 | 25 | 31 | 102 | 111 | 121 | 122 | 138 | 147 | 182 |
| Чернігівська | 21 | 11 | 9 | 9 | 11 | 12 | 11 | 77 | 89 | 69 | 68 | 65 | 62 | 70 |
| м. Київ | 2559 | 2511 | 2588 | 2638 | 2712 | 2730 | 2766 | 9044 | 8465 | 8620 | 8699 | 8711 | 8723 | 8826 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | y11 | | | | | | | y12 | | | | | | |
|--------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 10339 | 11259 | 11573 | 12014 | 12234 | 12845 | 13423 | 58741 | 64372 | 65839 | 68291 | 72124 | 74191 | 77763 |
| Авт. Респ. Крим | 267 | 283 | 303 | 312 | 325 | 350 | 350 | 1433 | 1548 | 1613 | 11656 | 2671 | 2591 | 2065 |
| Вінницька | 151 | 166 | 166 | 168 | 170 | 181 | 183 | 983 | 1081 | 1121 | 1150 | 1157 | 1232 | 1285 |
| Волинська | 44 | 44 | 44 | 46 | 50 | 58 | 65 | 514 | 562 | 589 | 588 | 680 | 716 | 753 |
| Дніпропетровська | 740 | 756 | 782 | 807 | 815 | 830 | 848 | 4281 | 4489 | 4558 | 4682 | 4900 | 4956 | 5051 |
| Донецька | 587 | 627 | 646 | 686 | 701 | 734 | 760 | 4142 | 4298 | 4355 | 4449 | 4598 | 4820 | 5009 |
| Житомирська | 44 | 49 | 51 | 54 | 57 | 60 | 63 | 461 | 590 | 640 | 685 | 700 | 775 | 818 |
| Закарпатська | 110 | 122 | 112 | 128 | 130 | 134 | 146 | 541 | 582 | 613 | 643 | 644 | 693 | 719 |
| Запорізька | 215 | 233 | 244 | 263 | 270 | 287 | 307 | 1636 | 1823 | 1879 | 1931 | 1936 | 2165 | 2234 |
| Івано-Франківська | 134 | 156 | 166 | 174 | 181 | 195 | 201 | 802 | 953 | 1058 | 1120 | 1197 | 1273 | 1356 |
| Київська | 109 | 125 | 133 | 134 | 140 | 149 | 162 | 769 | 846 | 863 | 863 | 987 | 1054 | 1132 |
| Кіровоградська | 31 | 39 | 44 | 41 | 40 | 39 | 43 | 449 | 512 | 546 | 566 | 598 | 606 | 627 |
| Луганська | 167 | 199 | 205 | 221 | 229 | 237 | 256 | 1459 | 1577 | 1611 | 1672 | 1681 | 1875 | 1935 |
| Львівська | 652 | 702 | 723 | 761 | 810 | 829 | 859 | 4235 | 4524 | 4623 | 4788 | 5028 | 5218 | 5523 |
| Миколаївська | 61 | 83 | 90 | 100 | 105 | 108 | 113 | 630 | 740 | 799 | 823 | 826 | 913 | 977 |
| Одеська | 693 | 700 | 717 | 740 | 744 | 748 | 758 | 3887 | 3984 | 3847 | 4015 | 4020 | 4335 | 4432 |
| Полтавська | 134 | 162 | 178 | 182 | 200 | 207 | 206 | 1129 | 1321 | 1342 | 1386 | 1432 | 1505 | 1575 |
| Рівненська | 45 | 52 | 58 | 60 | 68 | 72 | 77 | 668 | 750 | 787 | 179 | 674 | 886 | 958 |
| Сумська | 89 | 95 | 99 | 101 | 105 | 110 | 115 | 826 | 912 | 943 | 976 | 1056 | 1103 | 1177 |
| Тернопільська | 146 | 155 | 162 | 172 | 178 | 180 | 198 | 984 | 1156 | 1220 | 1325 | 1352 | 1384 | 1440 |
| Харківська | 1468 | 1627 | 1652 | 1742 | 1831 | 1858 | 1923 | 8700 | 10019 | 10108 | 10582 | 11089 | 11125 | 11334 |
| Херсонська | 77 | 81 | 84 | 83 | 85 | 86 | 95 | 651 | 692 | 704 | 738 | 794 | 838 | 872 |
| Хмельницька | 55 | 59 | 60 | 69 | 74 | 81 | 92 | 729 | 865 | 880 | 962 | 1003 | 1072 | 1138 |
| Черкаська | 64 | 76 | 78 | 90 | 100 | 110 | 118 | 584 | 665 | 710 | 791 | 863 | 902 | 963 |
| Чернівецька | 121 | 151 | 153 | 159 | 167 | 175 | 198 | 894 | 965 | 968 | 999 | 1034 | 1089 | 1133 |
| Чернігівська | 48 | 50 | 53 | 56 | 60 | 64 | 62 | 505 | 553 | 584 | 598 | 659 | 688 | 711 |
| м. Київ | 3996 | 4351 | 4451 | 4552 | 4735 | 4837 | 5103 | 16265 | 17674 | 18184 | 18750 | 19765 | 20787 | 21728 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | у13 | | | | | | | у14 | | | | | | |
|--------------------------|-------|-------|-------|-------|------|-------|-------|------|------|------|------|------|------|------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 15085 | 18198 | 19100 | 19900 | 8922 | 10983 | 12796 | 979 | 1009 | 966 | 951 | 920 | 904 | 881 |
| Авт. Респ. Крим | 260 | 250 | 227 | 289 | 129 | 155 | 157 | 40 | 37 | 35 | 35 | 35 | 34 | 33 |
| Вінницька | 125 | 130 | 148 | 134 | 42 | 53 | 74 | 31 | 33 | 24 | 25 | 24 | 24 | 23 |
| Волинська | 128 | 124 | 90 | 80 | 41 | 43 | 54 | 18 | 18 | 19 | 18 | 17 | 17 | 16 |
| Дніпропетровська | 631 | 998 | 1077 | 1291 | 791 | 1203 | 1325 | 63 | 63 | 61 | 57 | 60 | 59 | 59 |
| Донецька | 668 | 745 | 841 | 1101 | 446 | 480 | 571 | 94 | 93 | 97 | 90 | 90 | 89 | 87 |
| Житомирська | 145 | 143 | 140 | 108 | 45 | 56 | 74 | 25 | 25 | 25 | 25 | 25 | 23 | 22 |
| Закарпатська | 169 | 160 | 155 | 199 | 82 | 120 | 135 | 19 | 19 | 18 | 19 | 20 | 20 | 18 |
| Запорізька | 385 | 574 | 692 | 694 | 338 | 408 | 451 | 35 | 36 | 36 | 35 | 35 | 37 | 31 |
| Івано-Франківська | 230 | 200 | 167 | 112 | 41 | 151 | 251 | 31 | 30 | 30 | 31 | 30 | 29 | 29 |
| Київська | 79 | 124 | 137 | 135 | 76 | 105 | 125 | 21 | 23 | 27 | 25 | 25 | 24 | 24 |
| Кіровоградська | 47 | 49 | 57 | 61 | 37 | 18 | 64 | 24 | 23 | 23 | 23 | 21 | 22 | 22 |
| Луганська | 218 | 301 | 352 | 319 | 56 | 174 | 354 | 37 | 41 | 43 | 42 | 38 | 38 | 39 |
| Львівська | 1803 | 1987 | 2014 | 2170 | 1079 | 1047 | 1254 | 57 | 63 | 62 | 61 | 56 | 53 | 53 |
| Миколаївська | 132 | 135 | 138 | 193 | 74 | 106 | 157 | 20 | 21 | 17 | 17 | 17 | 17 | 16 |
| Одеська | 485 | 578 | 714 | 697 | 222 | 302 | 365 | 52 | 54 | 43 | 45 | 43 | 42 | 41 |
| Полтавська | 76 | 140 | 150 | 164 | 51 | 119 | 168 | 32 | 31 | 28 | 26 | 25 | 23 | 23 |
| Рівненська | 66 | 70 | 73 | 78 | 28 | 27 | 35 | 16 | 22 | 18 | 17 | 17 | 17 | 17 |
| Сумська | 111 | 203 | 290 | 276 | 51 | 247 | 546 | 27 | 23 | 17 | 17 | 15 | 15 | 15 |
| Тернопільська | 152 | 214 | 260 | 229 | 60 | 82 | 85 | 22 | 23 | 23 | 23 | 22 | 20 | 20 |
| Харківська | 2443 | 2968 | 3009 | 3104 | 1171 | 1668 | 1777 | 75 | 78 | 78 | 75 | 75 | 74 | 73 |
| Херсонська | 32 | 45 | 51 | 108 | 57 | 111 | 165 | 24 | 24 | 21 | 21 | 22 | 22 | 21 |
| Хмельницька | 54 | 67 | 121 | 135 | 81 | 61 | 75 | 25 | 25 | 25 | 25 | 25 | 21 | 20 |
| Черкаська | 57 | 100 | 132 | 99 | 100 | 170 | 185 | 27 | 27 | 25 | 27 | 19 | 18 | 18 |
| Чернівецька | 168 | 187 | 217 | 147 | 77 | 101 | 241 | 20 | 22 | 20 | 20 | 19 | 17 | 17 |
| Чернігівська | 194 | 152 | 114 | 168 | 24 | 44 | 65 | 24 | 26 | 26 | 26 | 23 | 22 | 22 |
| м. Київ | 6073 | 7254 | 7355 | 7616 | 3575 | 3862 | 3974 | 112 | 116 | 115 | 116 | 113 | 118 | 113 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | у15 | | | | | | | у16 | | | | | | |
|-------------------|------|------|------|------|------|------|------|---------|---------|---------|---------|---------|---------|---------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 395 | 512 | 545 | 577 | 597 | 606 | 599 | 22830,8 | 22171,3 | 22202,4 | 22280,8 | 22245,4 | 22322,3 | 22397,4 |
| Авт. Респ. Крим | 276 | 359 | 339 | 325 | 355 | 360 | 344 | 989,8 | 989,9 | 966,4 | 958,7 | 957,2 | 960,4 | 964,4 |
| Вінницька | 270 | 339 | 321 | 338 | 344 | 353 | 336 | 919,1 | 869,2 | 783,2 | 776,4 | 773,3 | 773,9 | 775,3 |
| Волинська | 271 | 315 | 310 | 334 | 342 | 357 | 347 | 540 | 513,2 | 475,1 | 471,8 | 471,7 | 472,3 | 478,2 |
| Дніпропетровська | 399 | 509 | 523 | 548 | 563 | 569 | 559 | 1735,8 | 1711,1 | 1659,1 | 1665,7 | 1667,4 | 1664,9 | 1664,1 |
| Донецька | 350 | 441 | 461 | 470 | 475 | 473 | 450 | 2383,5 | 2253,9 | 2250,4 | 2265,1 | 2261,6 | 2260,7 | 2267,3 |
| Житомирська | 267 | 361 | 370 | 386 | 394 | 382 | 366 | 624,8 | 575,6 | 632,4 | 632,5 | 625,1 | 623,9 | 622,1 |
| Закарпатська | 180 | 216 | 206 | 217 | 228 | 234 | 225 | 569,6 | 592,9 | 586 | 592,3 | 584,6 | 584,6 | 590 |
| Запорізька | 359 | 466 | 502 | 550 | 561 | 621 | 613 | 924,4 | 884,2 | 896,1 | 908,3 | 908,9 | 904,7 | 904,3 |
| Івано-Франківська | 309 | 408 | 417 | 435 | 447 | 443 | 442 | 653,3 | 537,8 | 571,9 | 572,8 | 569,3 | 582,5 | 588,4 |
| Київська | 170 | 224 | 239 | 228 | 233 | 238 | 240 | 847,2 | 819,4 | 833,5 | 846,9 | 843,7 | 844,7 | 845,3 |
| Кіровоградська | 252 | 300 | 300 | 311 | 312 | 317 | 313 | 511,1 | 504,4 | 499,3 | 501,5 | 500,2 | 503,2 | 499,2 |
| Луганська | 299 | 414 | 446 | 475 | 508 | 505 | 502 | 1135,3 | 1078,4 | 1122,9 | 1143,2 | 1137 | 1142,3 | 1144 |
| Львівська | 442 | 553 | 572 | 615 | 641 | 647 | 656 | 1249,5 | 1192,4 | 1174,5 | 1167 | 1166,9 | 1167,7 | 1182,2 |
| Миколаївська | 272 | 364 | 345 | 360 | 370 | 378 | 367 | 586,4 | 580,9 | 598,5 | 595 | 595,7 | 597,6 | 598,2 |
| Одеська | 466 | 542 | 585 | 601 | 614 | 627 | 630 | 1151,3 | 1111,3 | 1117 | 1086,4 | 1095,2 | 1108,2 | 1117,8 |
| Полтавська | 342 | 434 | 445 | 470 | 499 | 494 | 488 | 748,2 | 722,8 | 723,6 | 744,5 | 744,9 | 744 | 739,3 |
| Рівненська | 303 | 409 | 343 | 466 | 485 | 501 | 489 | 491,7 | 483,4 | 500,1 | 511,1 | 513,4 | 518,5 | 524,3 |
| Сумська | 315 | 481 | 523 | 537 | 542 | 518 | 500 | 603,5 | 630,8 | 597 | 591,2 | 588 | 587,4 | 587,4 |
| Тернопільська | 476 | 617 | 604 | 602 | 580 | 568 | 578 | 475,1 | 430,2 | 440,3 | 467,4 | 453,5 | 463,3 | 466 |
| Харківська | 708 | 896 | 975 | 1012 | 1022 | 1043 | 1018 | 1463,5 | 1438,8 | 1396,3 | 1388,8 | 1391,3 | 1385,7 | 1386,5 |
| Херсонська | 332 | 412 | 389 | 397 | 403 | 390 | 381 | 556,4 | 517,4 | 535,1 | 548,9 | 553 | 551,3 | 553,6 |
| Хмельницька | 265 | 325 | 343 | 366 | 386 | 409 | 409 | 652,7 | 623,1 | 655,9 | 641,9 | 640,2 | 614,1 | 656 |
| Черкаська | 285 | 400 | 397 | 426 | 437 | 452 | 444 | 605,9 | 604,6 | 633,7 | 641,6 | 637,7 | 638 | 635 |
| Чернівецька | 336 | 434 | 414 | 433 | 437 | 425 | 430 | 354 | 359,8 | 402,5 | 401,2 | 400,7 | 411,2 | 415,9 |
| Чернігівська | 278 | 338 | 353 | 360 | 352 | 334 | 321 | 586,9 | 582,3 | 559,4 | 562,3 | 547,8 | 546,2 | 542,9 |
| м. Київ | 1167 | 1574 | 1887 | 2119 | 2244 | 2290 | 2268 | 1333,3 | 1403 | 1416,4 | 1412,6 | 1425,7 | 1450,3 | 1465,7 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | x1 | | | | | | | x2 | | | | | | |
|--------------------------|-------|-------|-------|-------|--------|--------|--------|-------|-------|-------|-------|-------|-------|-------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 23629 | 51011 | 75714 | 93096 | 125254 | 188486 | 233081 | 12148 | 12882 | 18784 | 24995 | 30893 | 40188 | 45830 |
| Авт. Респ. Крим | 784 | 1913 | 2740 | 3375 | 4684 | 7007 | 8655 | 79 | 345 | 625 | 932 | 558 | 624 | 280 |
| Вінницька | 365 | 857 | 1155 | 1693 | 2494 | 4136 | 4752 | 58 | 95 | 58 | 54 | 77 | 616 | 120 |
| Волинська | 278 | 727 | 1064 | 1186 | 1839 | 2867 | 3489 | 45 | 135 | 417 | 207 | 1209 | 3042 | 2139 |
| Дніпропетровська | 2252 | 4246 | 5906 | 7792 | 10737 | 14874 | 18063 | 837 | 1872 | 2256 | 2650 | 4278 | 4516 | 3229 |
| Донецька | 2901 | 4846 | 7239 | 9598 | 11728 | 16898 | 22136 | 592 | 3778 | 3143 | 4936 | 5206 | 5601 | 7412 |
| Житомирська | 256 | 609 | 854 | 1135 | 1546 | 2442 | 3646 | 203 | 57 | 129 | 148 | 230 | 226 | 202 |
| Закарпатська | 332 | 1021 | 1113 | 1115 | 2019 | 2642 | 3303 | 52 | 403 | 677 | 100 | 154 | 115 | 152 |
| Запорізька | 1190 | 2150 | 2745 | 3299 | 3896 | 6407 | 7459 | 550 | 1051 | 2946 | 3201 | 3545 | 3432 | 5345 |
| Івано-Франківська | 605 | 1130 | 1589 | 1683 | 2620 | 4006 | 5950 | 27 | 79 | 161 | 200 | 156 | 438 | 598 |
| Київська | 947 | 2918 | 3547 | 4519 | 7032 | 12658 | 16921 | 202 | 238 | 284 | 309 | 408 | 516 | 618 |
| Кіровоградська | 271 | 699 | 1357 | 1285 | 1777 | 2482 | 3048 | 12 | 140 | 249 | 220 | 476 | 485 | 427 |
| Луганська | 1030 | 1901 | 2941 | 4363 | 5538 | 9754 | 9000 | 296 | 572 | 1278 | 1450 | 3849 | 77203 | 12393 |
| Львівська | 1061 | 2559 | 3634 | 4682 | 5875 | 8287 | 10799 | 152 | 394 | 331 | 343 | 424 | 1064 | 627 |
| Миколаївська | 460 | 1312 | 1963 | 2534 | 3118 | 3699 | 4333 | 96 | 290 | 400 | 720 | 419 | 61 | 174 |
| Одеська | 1348 | 3309 | 5137 | 5154 | 7339 | 10500 | 12484 | 26 | 481 | 541 | 2177 | 2197 | 2287 | 383 |
| Полтавська | 1438 | 2158 | 2887 | 3709 | 4760 | 6672 | 8840 | 674 | 190 | 246 | 350 | 407 | 1077 | 256 |
| Рівненська | 365 | 1078 | 1937 | 1184 | 2144 | 2924 | 3672 | 107 | 37 | 80 | 55 | 23 | 88 | 109 |
| Сумська | 597 | 931 | 1102 | 1441 | 1508 | 2435 | 3015 | 580 | 550 | 621 | 1244 | 917 | 820 | 1079 |
| Тернопільська | 199 | 456 | 632 | 890 | 1214 | 1979 | 2843 | 26 | 34 | 38 | 44 | 12 | 180 | 262 |
| Харківська | 1373 | 3554 | 5055 | 5775 | 7957 | 12051 | 11819 | 1011 | 648 | 1450 | 1903 | 1953 | 1882 | 2885 |
| Херсонська | 282 | 549 | 886 | 10770 | 1503 | 2322 | 3929 | 210 | 129 | 155 | 128 | 159 | 201 | 258 |
| Хмельницька | 455 | 901 | 1745 | 1461 | 2075 | 3135 | 4849 | 111 | 84 | 142 | 190 | 111 | 97 | 59 |
| Черкаська | 330 | 967 | 2408 | 2046 | 3304 | 4236 | 5128 | 172 | 19 | 67 | 103 | 162 | 511 | 2666 |
| Чернівецька | 141 | 520 | 656 | 755 | 1451 | 2332 | 3733 | 57 | 103 | 86 | 78 | 88 | 304 | 220 |
| Чернігівська | 431 | 825 | 1140 | 1333 | 1520 | 2422 | 2053 | 249 | 69 | 190 | 253 | 173 | 399 | 394 |
| м. Київ | 3809 | 8481 | 13859 | 19468 | 24858 | 38230 | 46574 | 385 | 1086 | 2211 | 2986 | 3688 | 3870 | 3530 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | х3 | | | | | | | х4 | | | | | | |
|-------------------|---------|---------|---------|----------|----------|----------|----------|---------|---------|---------|---------|---------|----------|----------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 3023,6 | 4776,18 | 7984,45 | 12494,82 | 12797,02 | 14666,64 | 23633,07 | 1760,06 | 3059,85 | 4534,62 | 5751,56 | 6159,95 | 10821,02 | 11994,22 |
| Авт. Респ. Крим | 50,18 | 32,93 | 123,08 | 353,32 | 133,51 | 232,14 | 102,39 | 11,38 | 22,45 | 80,35 | 93,61 | 90,25 | 185,04 | 225,61 |
| Вінницька | 23,82 | 16,33 | 10,44 | 12,73 | 8,88 | 169,41 | 13,82 | 50,2 | 28,78 | 15,4 | 55,7 | 99,82 | 355,34 | 133,75 |
| Волинська | 23,4 | 1,03 | 0,06 | 3,72 | 1,42 | 1,34 | 77,15 | 9,1 | 8,75 | 117,38 | 9,32 | 52,78 | 169,77 | 232,86 |
| Дніпропетровська | 200,26 | 407,05 | 639,92 | 844,09 | 1196,49 | 1810,06 | 1703,46 | 214,61 | 315,31 | 144,26 | 385,47 | 685,22 | 1179,6 | 1355,47 |
| Донецька | 1679,82 | 2471,11 | 2213,67 | 3323,37 | 3389,27 | 2998,12 | 4356,66 | 467,6 | 807,66 | 996,94 | 1955,09 | 1785,34 | 2639,91 | 2740,37 |
| Житомирська | 33,41 | 9,53 | 18,16 | 25,29 | 19,98 | 37,42 | 54,97 | 27,66 | 24,7 | 16,69 | 46,8 | 28,96 | 61,37 | 48,56 |
| Закарпатська | 18,41 | 23,68 | 105,85 | 79,31 | 102,38 | 85,17 | 76,14 | 40,29 | 9,45 | 8,54 | 6,03 | 14,03 | 40,7 | 14,02 |
| Запорізька | 156,11 | 353,69 | 2673,58 | 2723,17 | 1277,38 | 2152,06 | 4522,39 | 155,93 | 152,06 | 206,98 | 158,38 | 91,54 | 160,33 | 202,49 |
| Івано-Франківська | 6,71 | 35,69 | 58,63 | 53,57 | 44,69 | 62,85 | 118,58 | 5,71 | 26,04 | 50,51 | 56,24 | 225,52 | 264,86 | 501,02 |
| Київська | 66,81 | 108,99 | 43,18 | 43,47 | 71,78 | 142,8 | 203,62 | 11,82 | 81,01 | 122,15 | 30,67 | 54,34 | 306,62 | 133 |
| Кіровоградська | 0,77 | 54,8 | 29,21 | 31,65 | 394,57 | 390,32 | 116,02 | 11,8 | 42,12 | 35,11 | 36,54 | 52,62 | 107,29 | 40,65 |
| Луганська | 88,15 | 67,6 | 267,69 | 200,4 | 2257,32 | 3513 | 9102,66 | 110,27 | 123,25 | 97,14 | 334,42 | 96,91 | 1343,37 | 2178,97 |
| Львівська | 15,44 | 148,71 | 136,06 | 62,16 | 69,4 | 155,16 | 17,51 | 41,45 | 69,68 | 64,42 | 75,86 | 72,46 | 272,51 | 296,66 |
| Миколаївська | 23,25 | 128,17 | 192,79 | 515,48 | 230,33 | 29,97 | 89,65 | 103,08 | 287,9 | 268,05 | 255,47 | 239,7 | 240,87 | 341,99 |
| Одеська | ! | 75,17 | 93,73 | 1688,55 | 1757,11 | 49,62 | 6,93 | 3,1 | 31,06 | 173,52 | 110,68 | 128,99 | 741,15 | 272,49 |
| Полтавська | 23,93 | 91,86 | 133,29 | 243,82 | 288,52 | 604,52 | 127,81 | 39,08 | 73,13 | 80,55 | 138,5 | 122,95 | 99,65 | 180,4 |
| Рівненська | 2,07 | 6,48 | 40,12 | 14,64 | 3,9 | 16,67 | 21,37 | 6,46 | 11,79 | 56,41 | 32,74 | 74,98 | 24,28 | 198,58 |
| Сумська | 206,44 | 404,88 | 453,41 | 1058,95 | 704,48 | 631,42 | 861,27 | 35,44 | 82,62 | 72,35 | 109,3 | 127,28 | 149,52 | 97,2 |
| Тернопільська | 1,82 | 14,04 | 7,73 | 4,36 | 0,36 | 50,67 | 196,37 | 3,59 | 13,82 | 3,25 | 5,68 | 6,72 | 19,26 | 62,73 |
| Харківська | 209,26 | 124,9 | 275 | 368,7 | 183,09 | 289,21 | 728,97 | 188,1 | 397,98 | 776,4 | 592,04 | 347,5 | 277,47 | 274,84 |
| Херсонська | 11,05 | 47,88 | 60,91 | 57,91 | 56,32 | 46,57 | 37,04 | 14,82 | 17,15 | 30,57 | 41,45 | 63,31 | 138,77 | 79,37 |
| Хмельницька | 45,12 | 26,45 | 42,13 | 36,76 | 2,95 | 17,33 | 22,76 | 7,87 | 16,88 | 10,81 | 18,07 | 136,25 | 91,95 | 179,49 |
| Черкаська | 37,27 | 3,44 | 4,02 | 5,2 | 81,8 | 355,71 | 17,67 | 4,93 | 12,5 | 108,2 | 53,9 | 16,09 | 42,81 | 15,53 |
| Чернівецька | 9,28 | 10,86 | 17,82 | 12,32 | 12,27 | 43,75 | 86,48 | 11,67 | 22,7 | 24,65 | 13,84 | 16,51 | 16,92 | 31,86 |
| Чернігівська | 48,87 | 23,54 | 67,87 | 73,58 | 100,6 | 180,41 | 184,28 | 103,24 | 96,59 | 125,8 | 122,45 | 80,8 | 225,44 | 385,72 |
| м. Київ | 41,97 | 87,38 | 276,11 | 658,04 | 407,6 | 600,82 | 782,3 | 79,68 | 283,32 | 847,98 | 1009,46 | 1447,84 | 1666,02 | 1769,48 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | x5 | | | | | | | x6 | | | | | | |
|-------------------|--------|--------|---------|---------|---------|--------|--------|-------|-------|-------|--------|--------|--------|--------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 430,1 | 839,71 | 1253,75 | 1603,99 | 1947,54 | 793,55 | 958,75 | 72,79 | 95,86 | 143,5 | 243,37 | 159,49 | 192,89 | 284,83 |
| Авт. Респ. Крим | 2,39 | 4,18 | 13,02 | 5,4 | 8,13 | 9,4 | 12,19 | 0,62 | 0,32 | 0,33 | 0,65 | 0,02 | 0,06 | 0,16 |
| Вінницька | 2,52 | 1,48 | 3,08 | 3,36 | 1,65 | 6,34 | 16,91 | 0,1 | 0,67 | 0,35 | 12,17 | 16,31 | 2,15 | 1,92 |
| Волинська | 0,55 | 1,56 | 1,08 | 1,74 | 1,15 | 2,73 | 1,56 | | | 0,03 | 0,08 | 0,09 | 0,1 | 7,43 |
| Дніпропетровська | 108,82 | 71,74 | 34,71 | 198,2 | 377,13 | 56,72 | 169 | 25,86 | 5,87 | 0,82 | 1,53 | 1,17 | 14,37 | 30,69 |
| Донецька | 41,28 | 121,63 | 111,84 | 140,12 | 192,64 | 83,63 | 88,19 | 0,96 | 5,65 | 11,52 | 90,26 | 46,55 | 77,2 | 72,8 |
| Житомирська | 2,04 | 1,41 | 1,14 | 2,47 | 1,12 | | 0,17 | 0,31 | 0,01 | 0 | 0,29 | 0,13 | 0,22 | |
| Закарпатська | 2,98 | 0,54 | 1,75 | 0,06 | 0,66 | 6,42 | 0,02 | 0,12 | | | | | 0,01 | 0,03 |
| Запорізька | 100,67 | 78,93 | 142,36 | 74,54 | 40,3 | 37,3 | 24,98 | 0,57 | 9,92 | 5,67 | 1,6 | 0,85 | 5,57 | 29,74 |
| Івано-Франківська | 0,91 | 2,62 | 6,93 | 2,53 | 70,84 | 119,45 | 231,95 | 0,02 | 0,04 | 0,29 | 3,94 | 17,57 | 9,86 | 25 |
| Київська | 2,9 | 3,68 | 4,08 | 15,73 | 7,19 | 0,59 | 2,27 | 2,13 | 0,2 | 35,69 | 0,05 | 0,77 | 1,99 | 1,99 |
| Кіровоградська | 2,67 | 33,77 | 15,59 | 20,53 | 12,61 | 18,1 | 17,64 | 0 | 0,62 | 0,36 | 0,38 | 0,18 | 1,33 | 1,1 |
| Луганська | 19,68 | 14,59 | 25,22 | 293,5 | 37,38 | 10,23 | 11,36 | 1,56 | 26,11 | 16,29 | 0,74 | | 9,17 | 22,46 |
| Львівська | 5,4 | 8,55 | 9,29 | 9,23 | 12,73 | 17,1 | 10,2 | 0,05 | 4,74 | 1,5 | 2,48 | 0,69 | 5,09 | 2,3 |
| Миколаївська | 94,53 | 151 | 126,98 | 153,92 | 189,86 | 87,38 | 138,29 | 0,03 | 0,3 | | 0,68 | 0,61 | ! | 1,47 |
| Одеська | 1,7 | 2,81 | 7,32 | 30,67 | 50,8 | 5,27 | 12,76 | | 1,3 | 4,28 | 3 | 4,3 | 0,48 | 9,88 |
| Полтавська | 11,74 | 16,2 | 25,27 | 45,08 | 46,66 | 5,88 | 14,57 | | 3,07 | 22,05 | 49,31 | | 6,94 | 9,52 |
| Рівненська | 0,61 | 2,58 | 2,55 | 4,19 | 1,97 | 3,09 | 1,66 | 1,14 | 8,5 | 0,65 | ! | 0 | 0,08 | 1,18 |
| Сумська | 19,65 | 31,97 | 44,09 | 59,14 | 59,01 | 11,09 | 11,9 | 0,03 | 0,89 | 0,28 | 0,21 | ! | 10,57 | 20,68 |
| Тернопільська | 0,39 | 2,72 | 1,79 | 2,05 | 0,75 | 1,65 | 0 | 0,09 | 0,01 | 0,01 | | | | |
| Харківська | 61,09 | 225,88 | 307,44 | 111,13 | 102,91 | 90,63 | 69,91 | 37,81 | 16,5 | 34,57 | 57,72 | 30,11 | 11,59 | 11,65 |
| Херсонська | 3,66 | 4,91 | 12,11 | 14,22 | 25,22 | 7,1 | 5,72 | 0,32 | 0,02 | 0 | ! | 0 | 3,65 | 2,48 |
| Хмельницька | 3,2 | 4,43 | 3,3 | 5,75 | 10,82 | 0 | 0 | | 0,14 | 0,08 | 0,21 | 7,41 | 0,03 | ! |
| Черкаська | 0,84 | 4,04 | 3,97 | 7,47 | 3,33 | 3,69 | 3,93 | 0,3 | 0,09 | | | | 0,92 | 0,1 |
| Чернівецька | 0,89 | 1,13 | 3,16 | 2,58 | 1,05 | 0,23 | 0,34 | 0 | 0,04 | | 0,06 | 1,22 | 2,4 | 1,07 |
| Чернігівська | 6,75 | 4,54 | 4,69 | 3,02 | 7,82 | 17,36 | 4,75 | 0,25 | 1,08 | 0,8 | 0,01 | 1,64 | 0,01 | 8,91 |
| м. Київ | 10,18 | 42,74 | 340,93 | 397,32 | 690,06 | 192,15 | 108,45 | 0,52 | 9,58 | 7,93 | 17,99 | 29,86 | 29,1 | 22,29 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | x7 | | | | | | | x8 | | | | | | |
|-------------------|---------|---------|---------|---------|---------|---------|---------|-------|-------|-------|------|-------|--------|--------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 1074,46 | 1873,71 | 2717,46 | 3149,56 | 3489,24 | 7441,27 | 7664,76 | 43,74 | 27,97 | 52,57 | 68,9 | 55,21 | 328,42 | 421,82 |
| Авт. Респ. Крим | 5,99 | 11,23 | 24,15 | 29,78 | 27,14 | 65,74 | 90,3 | 0,61 | ! | 0,06 | | | 13,17 | 0,37 |
| Вінницька | 47,14 | 24,21 | 11,36 | 35,1 | 79,4 | 340,97 | 111,66 | 0,1 | 0,11 | 0,14 | 0,48 | 0,04 | 0,38 | 0,33 |
| Волинська | 8,4 | 6,78 | 7,89 | 5,67 | 49,9 | 166,9 | 222,99 | | | 0,03 | 0,08 | 0,09 | ! | 0,06 |
| Дніпропетровська | 68,01 | 232,28 | 103,94 | 179,99 | 301,53 | 610,7 | 480,81 | 0,56 | 0,34 | 0,54 | 0,73 | 0,9 | 8,15 | 0,07 |
| Донецька | 418,79 | 669,67 | 861,28 | 1712,67 | 1533,8 | 2148,99 | 2239,54 | 0,09 | 0,71 | 0,9 | 0,15 | 21,23 | 237,19 | 317,69 |
| Житомирська | 24,39 | 18,96 | 11,11 | 39,56 | 25,54 | 56,72 | 46,89 | ! | | | | | | |
| Закарпатська | 30,36 | 8,58 | 5,62 | 2,02 | 9,89 | 31,28 | 13,3 | 0,04 | | | | | 0 | 0,13 |
| Запорізька | 30,74 | 57,3 | 48,64 | 62,58 | 31,98 | 84,59 | 101,96 | 0,43 | 8,55 | 4,31 | 1,6 | 0,3 | 0,28 | 0,18 |
| Івано-Франківська | 1,43 | 22,65 | 41,77 | 20,3 | 131,19 | 116,37 | 194,09 | 0,01 | 0 | 0 | 3,87 | 2,14 | 12,04 | 28,77 |
| Київська | 5,58 | 60,92 | 55,2 | 5,63 | 31,27 | 276,23 | 118,39 | 2,13 | 0 | 0,01 | 0,01 | 0 | 5,38 | 0,3 |
| Кіровоградська | 8,16 | 6,46 | 17,35 | 13,29 | 27,46 | 58,33 | 18,18 | 0 | 0,03 | 0,31 | 0,26 | | 0,84 | 0,17 |
| Луганська | 85,75 | 79,34 | 52,82 | 37,83 | 56,56 | 1278,32 | 1160,16 | 1,07 | 0,56 | 1,71 | 0,6 | | 3,16 | 1,28 |
| Львівська | 31,09 | 49,96 | 50,88 | 60,39 | 57,99 | 229,41 | 221,17 | 0,02 | 0,35 | 0,59 | 0,83 | 0,53 | 0,63 | 61,56 |
| Миколаївська | 74,37 | 120,7 | 134,13 | 81,37 | 45,56 | 80,35 | 186,36 | | | | 0,61 | 0,61 | ! | 0,02 |
| Одеська | 1,18 | 19,65 | 140,84 | 58,16 | 45,29 | 704,37 | 246,54 | | 0,41 | 4,28 | 2,74 | 1,81 | 0,1 | ! |
| Полтавська | 23,05 | 48,05 | 30,25 | 36 | 65,3 | 63,22 | 139,62 | | 0,01 | ! | 0,18 | | 0,58 | 1,08 |
| Рівненська | 1,07 | 0,49 | 52,95 | 27,86 | 67,12 | 7,2 | 163,53 | | | 0,03 | | | 9,53 | ! |
| Сумська | 14,39 | 32,06 | 26,93 | 47,5 | 61,77 | 80,83 | 44,16 | 0,01 | 0,85 | 0,28 | 0,21 | | 2,04 | 0,05 |
| Тернопільська | 3,05 | 10,83 | 1,33 | 3,45 | 5,85 | 16,59 | 61,63 | 0,05 | | | | | 0 | 0,09 |
| Харківська | 41,46 | 135,34 | 348,99 | 198,05 | 185,62 | 132,72 | 149,87 | 37,73 | 10,38 | 33,24 | 55 | 27,01 | 3,21 | 0,41 |
| Херсонська | 9,04 | 4,64 | 13,12 | 13,26 | 2,29 | 65,73 | 43,62 | 0,32 | 0 | | | | 0,05 | 0,08 |
| Хмельницька | 3,92 | 11,18 | 7,1 | 10,51 | 123,22 | 64,31 | 47,11 | | | | | | 0 | |
| Черкаська | 3,38 | 6,61 | 100,19 | 42,55 | 12,56 | 25,29 | 7,91 | 0,03 | 0,07 | | | | 8,81 | 0,12 |
| Чернівецька | 6,65 | 17,49 | 20,15 | 8,3 | 12,18 | 10,25 | 26,95 | 0 | | | | | | 0,18 |
| Чернігівська | 75,12 | 80,82 | 116,22 | 114,97 | 65,19 | 189,48 | 335,44 | 0,24 | | 0 | | | 0,59 | 0,06 |
| м. Київ | 50,88 | 137,13 | 433,14 | 298,96 | 432,5 | 536,22 | 1192,52 | 0,3 | 5,61 | 6,13 | 1,55 | 0,55 | 22,28 | 7,93 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | x9 | | | | | | | x10 | | | | | | |
|-------------------|-------|--------|--------|--------|--------|---------|---------|--------|-------|--------|--------|--------|--------|------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 82,13 | 168,96 | 297,45 | 376,68 | 359,88 | 2064,89 | 2664,06 | 137,99 | 244,5 | 345,11 | 441,45 | 544,15 | 720,73 | |
| Авт. Респ. Крим | 2,07 | 3,19 | 42,52 | 5,65 | 1,18 | 96,66 | 122,6 | 4,09 | 7,31 | 9,9 | 12,85 | 16,04 | 20,87 | |
| Вінницька | 0,26 | 2,33 | 0,53 | 4,84 | 1,73 | 5,49 | 2,93 | 3,8 | 6,17 | 8,12 | 10,21 | 12,41 | 15,38 | |
| Волинська | 0,14 | 0,38 | 108,36 | 1,81 | 1,63 | 0,04 | 0,83 | 2,2 | 3,51 | 4,99 | 6,55 | 7,69 | 10,07 | |
| Дніпропетровська | 4,73 | 4,08 | 4,62 | 3,46 | 3,85 | 489,65 | 674,9 | 13,16 | 21,31 | 30,04 | 41,23 | 52,35 | 71,17 | |
| Донецька | 4,2 | 10,11 | 11,86 | 10,53 | 12,29 | 92,9 | 22,15 | 17,28 | 30,32 | 45,62 | 58,04 | 72,36 | 92,09 | |
| Житомирська | 0,35 | 1,75 | 2,47 | 2,65 | 0,48 | 4,42 | 1,51 | 2,84 | 4,2 | 5,95 | 7,43 | 8,78 | 11,13 | |
| Закарпатська | 0,33 | 0,12 | 0,06 | 0,04 | 0,01 | 2,99 | 0,55 | 2,15 | 4,06 | 5,3 | 6,7 | 8,19 | 10,51 | |
| Запорізька | 5,8 | 4,53 | 4,59 | 7,54 | 8,25 | 32,59 | 45,62 | 7,57 | 10,18 | 15,26 | 19,97 | 24,79 | 33,16 | |
| Івано-Франківська | 0,13 | 0,31 | 0,59 | 0,08 | 0,69 | 7,14 | 21,21 | 3,12 | 5,54 | 7,31 | 9,62 | 11,32 | 13,92 | |
| Київська | 0,87 | 6,18 | 26,59 | 7,48 | 13,97 | 22,43 | 10,05 | 5,93 | 8,13 | 11,88 | 15,36 | 19,19 | 26,22 | |
| Кіровоградська | 0,96 | 0,93 | 0,29 | 0,03 | 0,03 | 28,7 | 3,57 | 2,16 | 4,02 | 5,59 | 6,88 | 8,19 | 9,99 | |
| Луганська | 1,04 | 1,44 | 1,28 | 1,08 | 1,5 | 42,49 | 983,71 | 6,4 | 9,95 | 14,67 | 19,72 | 24,16 | 32,28 | |
| Львівська | 0,65 | 5,16 | 2,72 | 3,72 | 1,05 | 20,28 | 1,44 | 5,85 | 10,55 | 13,99 | 17,19 | 21,49 | 27,99 | |
| Миколаївська | 8,36 | 13,86 | 4,29 | 1,81 | 0,84 | 73,14 | 15,85 | 3,31 | 5,34 | 7,93 | 9,55 | 11,88 | 14,77 | |
| Одеська | 0,22 | 5,46 | 5,67 | 5,39 | 5,79 | 30,94 | 3,32 | 7,07 | 12,79 | 17,03 | 20,76 | 24,9 | 33,12 | |
| Полтавська | 2,02 | 2,66 | 3 | 3,58 | 10,99 | 23,04 | 15,62 | 5,71 | 8,85 | 13,98 | 18,1 | 22,18 | 28,36 | |
| Рівненська | 0,07 | 0,16 | 0,1 | 0,18 | 4,87 | 4,38 | 32,21 | 2,51 | 4,05 | 5,6 | 7,26 | 8,92 | 11,18 | |
| Сумська | 0,75 | 16,91 | 0,98 | 2,45 | 0,57 | 44,99 | 20,42 | 3,5 | 4,82 | 6,28 | 8,03 | 9,57 | 12,34 | |
| Тернопільська | 0,03 | 0,09 | 0,09 | 0,06 | 0,11 | 1,02 | 1,01 | 1,85 | 3,1 | 3,95 | 5,14 | 6,45 | 8,28 | |
| Харківська | 18,98 | 11,52 | 28,44 | 214,47 | 27,38 | 39,31 | 43 | 8,27 | 14,52 | 20,52 | 25,62 | 32,02 | 43,87 | |
| Херсонська | 0,69 | 5,34 | 3,96 | 11,56 | 35,62 | 62,24 | 27,48 | 2,35 | 3,8 | 5,2 | 6,47 | 7,57 | 9,03 | |
| Хмельницька | 0,31 | 0,26 | 0,23 | 1,47 | 0,53 | 27,61 | 132,38 | 2,95 | 4,7 | 6,34 | 7,96 | 9,6 | 12,34 | |
| Черкаська | 0,18 | 0,87 | 0,2 | 0,33 | 0,14 | 4,09 | 3,48 | 3,18 | 4,57 | 6,62 | 9,01 | 10,96 | 13,66 | |
| Чернівецька | 0,58 | 0,53 | 0,52 | 0,34 | 1,3 | 4,04 | 3,31 | 1,31 | 2,54 | 3,28 | 4,23 | 5,13 | 6,67 | |
| Чернігівська | 17,66 | 1,35 | 3,45 | 2,07 | 3,65 | 18 | 36,57 | 3,07 | 4,58 | 6,18 | 7,63 | 8,95 | 11,53 | |
| м. Київ | 10,69 | 69,41 | 40,03 | 84,05 | 221,38 | 886,28 | 438,29 | 15,72 | 43,92 | 61,36 | 77,12 | 95,27 | 135,9 | |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | x11 | | | | | | | x12 | | | | | | |
|-------------------|--------|--------|--------|--------|---------|---------|---------|--------|--------|--------|--------|--------|---------|---------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 258,9 | 473,73 | 613,95 | 886,08 | 1107,94 | 1462,29 | 1895,07 | 375 | 383,69 | 518,15 | 655,79 | 781,27 | 1058,47 | 1447,47 |
| Авт. Респ. Крим | 5,87 | 10,86 | 13,08 | 21,42 | 26,18 | 35,03 | 42,47 | 7,35 | 9,3 | 11,55 | 14,51 | 19,15 | 25,67 | 32,49 |
| Вінницька | 1,58 | 2,3 | 2,54 | 3,82 | 4,92 | 6,27 | 7,51 | 2,49 | 2,71 | 3,5 | 5,05 | 5,48 | 6,93 | 9,85 |
| Волинська | 0,26 | 0,34 | 0,53 | 0,86 | 1,14 | 1,45 | 1,55 | 1,36 | 1,46 | 1,51 | 3,01 | 2,93 | 3,93 | 3,55 |
| Дніпропетровська | 10,84 | 18,5 | 24 | 35,18 | 46,26 | 61,4 | 73,37 | 24,1 | 21,92 | 29,28 | 48,94 | 38,14 | 53,36 | 63,23 |
| Донецька | 7,8 | 15,57 | 21,17 | 30,57 | 39,58 | 53,67 | 66,21 | 20,15 | 26,73 | 33,82 | 34,15 | 32,37 | 39,14 | 61,75 |
| Житомирська | 1,08 | 1,41 | 1,72 | 2,59 | 3 | 4,28 | 3,98 | 0,8 | 2,2 | 2,63 | 3,66 | 4,27 | 5,81 | 9,28 |
| Закарпатська | 2,07 | 3,49 | 6,47 | 5,3 | 7,31 | 9,07 | 9,17 | 1,85 | 3,35 | 1,33 | 4,55 | 7,34 | 4,58 | 8,67 |
| Запорізька | 1,68 | 2,74 | 4,53 | 5,65 | 7,72 | 9,55 | 11,55 | 5,08 | 7,85 | 10,37 | 12,58 | 12,26 | 16,09 | 20,06 |
| Івано-Франківська | 0,67 | 0,76 | 0,97 | 1,55 | 2,14 | 2,87 | 3,05 | 1,43 | 2,18 | 1,98 | 3,05 | 3,81 | 4,19 | 5,65 |
| Київська | 7,42 | 10,76 | 13,01 | 19,79 | 24,08 | 39,78 | 35,26 | 16,78 | 19,74 | 26,75 | 32,66 | 36,75 | 50,23 | 72,45 |
| Кіровоградська | 0,12 | 0,18 | 0,2 | 0,31 | 0,51 | 0,67 | 0,76 | 0,46 | 1,06 | 1,26 | 1,42 | 1,67 | 2,64 | 2,96 |
| Луганська | 1,01 | 2,02 | 2,04 | 2,85 | 3,92 | 5,25 | 5,52 | 5,55 | 9,78 | 12,07 | 12,55 | 11,98 | 15,5 | 18,92 |
| Львівська | 11,74 | 22,59 | 29,81 | 43,98 | 56,07 | 73,45 | 94,95 | 21,91 | 18,36 | 20,23 | 28,17 | 36,35 | 41,54 | 51,11 |
| Миколаївська | 1,21 | 2,64 | 3,17 | 4,76 | 7,37 | 9,54 | 11,66 | 3,19 | 3,31 | 4,76 | 5,7 | 6,9 | 8,79 | 11,58 |
| Одеська | 7,99 | 13,49 | 15,82 | 23,59 | 30,65 | 41,11 | 51,61 | 13,54 | 12,19 | 16,09 | 20,65 | 31,63 | 39,31 | 52,85 |
| Полтавська | 1,42 | 1,83 | 2,36 | 3,83 | 4,47 | 6,83 | 9,11 | 2,97 | 3,2 | 3,71 | 4,98 | 5,69 | 5,75 | 10,54 |
| Рівненська | 0,27 | 0,51 | 0,57 | 1,01 | 1,29 | 1,72 | 1,37 | 2,14 | 1,13 | 1,63 | 3,06 | 3,41 | 4,37 | 6,59 |
| Сумська | 1,36 | 1,65 | 2,2 | 3,49 | 4,92 | 6,7 | 8,18 | 3,01 | 1,82 | 2,27 | 3,88 | 4,29 | 5,83 | 8,86 |
| Тернопільська | 0,45 | 0,69 | 0,81 | 1,32 | 1,45 | 1,8 | 1,65 | 1,28 | 1,47 | 1,58 | 2,07 | 2,55 | 3,36 | 7,14 |
| Харківська | 52,5 | 74,74 | 97,3 | 132,43 | 164,97 | 221,01 | 286 | 49,86 | 49,9 | 63,25 | 83,7 | 99,78 | 160,4 | 222,16 |
| Херсонська | 1,79 | 2,82 | 3,82 | 6,53 | 7,6 | 9,97 | 11,23 | 1,32 | 2,23 | 3,13 | 4,24 | 5,68 | 8,82 | 14,55 |
| Хмельницька | 0,1 | 0,17 | 0,29 | 0,51 | 0,68 | 1,02 | 1,04 | 0,48 | 1,4 | 1,91 | 2,96 | 4,17 | 4,49 | 5,11 |
| Черкаська | 2,25 | 6,15 | 7,88 | 9,79 | 7,75 | 9,54 | 13,29 | 3,72 | 2,87 | 3,35 | 5,46 | 10,73 | 13,86 | 18,9 |
| Чернівецька | 1,52 | 3,69 | 2,98 | 4,26 | 5,61 | 7,68 | 9,56 | 1,47 | 4,89 | 4,77 | 4,94 | 7,6 | 13,32 | 14,6 |
| Чернігівська | 0,77 | 1,45 | 1,82 | 2,78 | 3,2 | 4,14 | 4,85 | 3,39 | 3,51 | 4,99 | 6,47 | 10,21 | 14,17 | 15,94 |
| м. Київ | 131,58 | 266,27 | 346,76 | 506,38 | 630,8 | 819,43 | 1104,75 | 178,32 | 164,26 | 246,95 | 299,23 | 366,19 | 496,87 | 684,18 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | x13 | | | | | | | x14 | | | | | | |
|--------------------------|--------|---------|--------|---------|--------|---------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 854,86 | 1544,28 | 1882,7 | 2195,84 | 2539,3 | 2793,52 | 3570,8 | 147,51 | 422,29 | 523,62 | 648,56 | 532,53 | 594,11 | 809,81 |
| Авт. Респ. Крим | 7,99 | 21,97 | 27,34 | 42,86 | 39,84 | 46,26 | 45,23 | 1,28 | 8,25 | 9,65 | 11,07 | 9,2 | 13,13 | 15,37 |
| Вінницька | 7,48 | 22,24 | 23,72 | 28,77 | 17,28 | 24,2 | 20,01 | 2,52 | 1,06 | 1,21 | 1,75 | 0,22 | 1,19 | 12,67 |
| Волинська | 2,42 | 4,25 | 2,96 | 4,17 | 7,5 | 4,81 | 7,6 | 0,38 | 2,36 | 2,82 | 2,95 | 0,63 | 0,8 | 2,54 |
| Дніпропетровська | 81,36 | 140,73 | 205,67 | 246,24 | 359,44 | 427,55 | 535,07 | 6,8 | 16,42 | 23,43 | 33,66 | 40,25 | 40,65 | 39,45 |
| Донецька | 55,72 | 89,5 | 110,83 | 131,58 | 174,34 | 209,9 | 241,45 | 14,11 | 33,89 | 46,95 | 54,2 | 49,53 | 53,33 | 65,73 |
| Житомирська | 0,21 | 0,03 | 0,54 | 0,28 | | 2,18 | 3,91 | 1 | 1,54 | 2,21 | 3,93 | 2,77 | 3,32 | 3,63 |
| Закарпатська | 0,44 | 0,86 | 1,48 | 2 | 2,13 | 3,07 | 4,27 | 0,28 | 1,1 | 1,42 | 2,83 | 0,41 | 0,35 | 0,77 |
| Запорізька | 79,31 | 147,4 | 180,79 | 209,25 | 229,19 | 259,67 | 355,46 | 5,69 | 15,6 | 11,12 | 19,05 | 8,31 | 14,91 | 17,01 |
| Івано-Франківська | 3,16 | 4,39 | 6,79 | 8,25 | 12,82 | 16,53 | 22,31 | 5,21 | 8,96 | 11,52 | 14,46 | 12,96 | 13,4 | 23,79 |
| Київська | 3,95 | 11,72 | 15,64 | 19,38 | 30,16 | 28,52 | 42,84 | 2,41 | 7,98 | 10,59 | 9,11 | 4,85 | 5,65 | 11,26 |
| Кіровоградська | 2,62 | 3,5 | 5,65 | 8,31 | 11,96 | 15,86 | 18,29 | 0,22 | 3,33 | 0,39 | 4,39 | 0,93 | 0,6 | 1,09 |
| Луганська | 25,47 | 97,13 | 32,75 | 39,21 | 48,39 | 54,92 | 84,24 | 2,29 | 5,91 | 6,38 | 6,28 | 3,87 | 4,63 | 5,3 |
| Львівська | 31,31 | 50,18 | 57,07 | 65,99 | 71,65 | 96,01 | 104,48 | 2,81 | 11,21 | 15,7 | 18,64 | 17,07 | 20,5 | 27,58 |
| Миколаївська | 66,6 | 97,27 | 104,78 | 142,34 | 155,54 | 135,85 | 177,3 | 0,82 | 3,12 | 5,78 | 5,17 | 3,45 | 7,54 | 20,16 |
| Одеська | 14,37 | 26,26 | 28,05 | 39,39 | 42,78 | 45,41 | 45,85 | 5,77 | 20,88 | 26,01 | 22,8 | 15,25 | 14,52 | 29,63 |
| Полтавська | 4,72 | 7,64 | 10,22 | 22,58 | 15,02 | 15,02 | 17,75 | 5,04 | 5,42 | 5,43 | 2,8 | 7,11 | 10,22 | 10,77 |
| Рівненська | 1,5 | 1,79 | 3,69 | 3,66 | 3,01 | 3,73 | 3,75 | 0,85 | 1,58 | 1,17 | 1,6 | 3,43 | 1,69 | 3,37 |
| Сумська | 18,54 | 44,99 | 59,55 | 60,97 | 69,27 | 81,78 | 102,32 | 2,4 | 6,24 | 12,75 | 15,48 | 4,88 | 3,51 | 5,68 |
| Тернопільська | 3,2 | 2,38 | 1,81 | 1,82 | 5,61 | 6,38 | 4,41 | 0,73 | 3,18 | 4,01 | 6,78 | 2,19 | 0,08 | 0,12 |
| Харківська | 160,24 | 274,54 | 345,07 | 387,16 | 452,59 | 546,86 | 686,02 | 19,02 | 51,58 | 65,86 | 79,48 | 87,92 | 75,51 | 90,1 |
| Херсонська | 4,19 | 4,19 | 3,5 | 5,18 | 6,44 | 8,37 | 8,11 | 0,28 | 1,08 | 1,64 | 1,24 | 2,16 | 0,95 | 1,03 |
| Хмельницька | 0,69 | 0,89 | 1,21 | 0,5 | 0,67 | 1,93 | 1,46 | 0,63 | 1,2 | 0,06 | 0,08 | 0,08 | 0,04 | 0,06 |
| Черкаська | 6,17 | 7,97 | 8,72 | 7,06 | 13,48 | 17,79 | 21,75 | 0,66 | 1,72 | 2 | 4,81 | 0,43 | 1,39 | 1,47 |
| Чернівецька | 2,32 | 0,65 | 0,81 | 1,44 | 2,59 | 1,67 | 2,01 | 0,52 | 2,14 | 2,35 | 3,2 | 0,15 | 0,48 | 0,94 |
| Чернігівська | 3,77 | 16 | 5,13 | 9,28 | 5,18 | 8,47 | 13,24 | 1,58 | 6,88 | 7,64 | 2,66 | 5,45 | 2,96 | 4,75 |
| м. Київ | 250,95 | 449,41 | 613,1 | 670,39 | 724,55 | 682,78 | 952,12 | 64,1 | 193,92 | 243,66 | 318,28 | 246,12 | 292,36 | 399,05 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | x15 | | | | | | | x16 | | | | | | |
|--------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 1978,41 | 3085,49 | 4112,38 | 4818,58 | 5354,57 | 6700,71 | 8538,94 | 2046,34 | 3521,48 | 4251,74 | 5160,4 | 5164,43 | 6149,23 | 8024,76 |
| Авт. Респ. Крим | 25,25 | 53,21 | 71,97 | 89,37 | 107,41 | 131,41 | 143,8 | 24,12 | 67,85 | 70,32 | 87,36 | 103,95 | 125,73 | 140,97 |
| Вінницька | 15,67 | 25,99 | 31,76 | 41,54 | 28,23 | 41,75 | 52,67 | 19,15 | 23,42 | 33,1 | 43,92 | 28,97 | 40,59 | 50,62 |
| Волинська | 4,54 | 5,46 | 7,71 | 11,85 | 14,06 | 12,4 | 17,18 | 5,14 | 6,46 | 8,21 | 12,06 | 13,36 | 11,74 | 16,77 |
| Дніпропетровська | 177,73 | 232,65 | 337,34 | 398,78 | 525,96 | 634,56 | 790,14 | 168,4 | 314,56 | 422,06 | 518,49 | 493,8 | 594,91 | 723,98 |
| Донецька | 104,61 | 202,31 | 224,62 | 261,9 | 322,29 | 395,41 | 478,55 | 111,31 | 198,77 | 224,84 | 264,5 | 301,47 | 365,87 | 447,75 |
| Житомирська | 3,25 | 5,41 | 6,94 | 10,65 | 10,43 | 16,81 | 22,15 | 3,28 | 5,48 | 7,19 | 10,65 | 10,15 | 15,98 | 21,35 |
| Закарпатська | 4,67 | 6,12 | 7,66 | 15,04 | 15,54 | 15,21 | 20,63 | 4,78 | 8,98 | 11,17 | 16,14 | 17,44 | 17,18 | 23,13 |
| Запорізька | 149,15 | 200,12 | 271,38 | 305,28 | 304,85 | 351,59 | 459,35 | 166,94 | 211,11 | 218,54 | 230,01 | 266,84 | 311,93 | 422,38 |
| Івано-Франківська | 11,49 | 19,87 | 23,84 | 31,14 | 36,02 | 44,73 | 50,9 | 12,9 | 19 | 24,73 | 30,65 | 35,96 | 45,11 | 56,3 |
| Київська | 31,74 | 57,85 | 68,57 | 85,91 | 102,55 | 136,37 | 181,19 | 32,63 | 57,85 | 72,57 | 89,61 | 99,51 | 134,03 | 180,35 |
| Кіровоградська | 3,34 | 5,22 | 7,2 | 14,64 | 15,56 | 20,36 | 24,17 | 3,41 | 8,77 | 11,76 | 19,49 | 15,46 | 20,51 | 23,89 |
| Луганська | 43,36 | 50,49 | 55,57 | 69,05 | 70,32 | 85,27 | 120,86 | 66,43 | 59,74 | 56,75 | 71,33 | 69,68 | 84,7 | 118,01 |
| Львівська | 69,99 | 124,36 | 135,35 | 179,18 | 199,16 | 252,89 | 288,67 | 68,29 | 102,54 | 138,39 | 157,6 | 189,06 | 240,25 | 283,84 |
| Миколаївська | 85,36 | 142,65 | 164,95 | 212,16 | 202,54 | 209,71 | 272,18 | 71,01 | 121,57 | 146,99 | 220,91 | 175,6 | 167,75 | 224,53 |
| Одеська | 42,1 | 67,84 | 94,47 | 110,34 | 125,64 | 154,05 | 182,34 | 46,21 | 75,42 | 96,26 | 119,33 | 123,44 | 145,67 | 186,91 |
| Полтавська | 15,24 | 23,11 | 25,28 | 37,39 | 35,45 | 41,58 | 54,6 | 16,39 | 20,55 | 27,03 | 42,21 | 33,15 | 39,08 | 51,79 |
| Рівненська | 4,96 | 5,41 | 7,29 | 9,38 | 11,98 | 12,47 | 15,44 | 5,14 | 6,74 | 8,15 | 10,19 | 11,33 | 11,7 | 15,35 |
| Сумська | 27,69 | 56,74 | 81,1 | 95,38 | 87,37 | 101,17 | 128,3 | 32,32 | 67,85 | 93,85 | 107,03 | 88,27 | 102,42 | 128,75 |
| Тернопільська | 5,69 | 6,75 | 8,98 | 11,91 | 12,25 | 12,89 | 14,67 | 3,23 | 5,64 | 8,59 | 12,38 | 11,8 | 11,62 | 13,65 |
| Харківська | 329,46 | 457,46 | 640,94 | 772,49 | 901,36 | 1130,48 | 1418,96 | 381,55 | 489,65 | 678,68 | 812,57 | 832,54 | 1036,77 | 1343,66 |
| Херсонська | 6,98 | 10,55 | 12,73 | 19,58 | 24,01 | 39,57 | 38,32 | 7,27 | 12,58 | 13,08 | 20,08 | 22,34 | 28,3 | 35,5 |
| Хмельницька | 1,93 | 2,57 | 3,74 | 4,14 | 5,89 | 7,78 | 8,27 | 1,89 | 2,65 | 3,68 | 4,14 | 5,89 | 7,63 | 8,29 |
| Черкаська | 14,18 | 20,46 | 24,03 | 28,56 | 33,72 | 46,76 | 55,97 | 13,35 | 19,43 | 22,56 | 27,45 | 32,91 | 44 | 56,6 |
| Чернівецька | 5,87 | 9,65 | 11,51 | 14,85 | 19,59 | 24,52 | 29,8 | 6,98 | 10,11 | 11,23 | 15,17 | 18,26 | 25,4 | 27,79 |
| Чернігівська | 10,96 | 16,55 | 20,32 | 22,53 | 23,46 | 28,9 | 38,21 | 12,75 | 17,69 | 25,94 | 32,08 | 24,37 | 30,13 | 39,4 |
| м. Київ | 756,35 | 1241,24 | 1724,49 | 1894,01 | 2042 | 2640,69 | 3498,12 | 722,79 | 1547,32 | 1766,15 | 2109,73 | 2067,88 | 2400,04 | 3266,11 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | z1 | | | | | | | z2 | | | | | | |
|-------------------|--------|------|--------|-------|--------|--------|--------|--------|-------|--------|--------|--------|--------|---------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 1150,9 | 1458 | 2058,6 | 2935 | 3719,4 | 4980,6 | 6468 | 3486,8 | 4248 | 5307,9 | 6134,7 | 7505,5 | 9038,9 | 11741,3 |
| Авт. Респ. Крим | 12 | 5,3 | 12,2 | 9,2 | 10,7 | 13,7 | 25 | 76 | 131 | 153,2 | 146,5 | 155,1 | 174,2 | 251,3 |
| Вінницька | 0,2 | 0,5 | 70,9 | 104,4 | 61,3 | 67,6 | 65,6 | 1,8 | 2,5 | 10,5 | 13,8 | 19,5 | 26,1 | 41,3 |
| Волинська | 2 | 5,7 | 10,5 | 17,3 | 16,7 | 26,4 | 57,6 | 3,8 | 4,4 | 5,7 | 10,5 | 18,3 | 25,4 | 40,9 |
| Дніпропетровська | 66 | 58,7 | 102,3 | 136,7 | 170,9 | 316,5 | 494,5 | 59,9 | 102,1 | 122,1 | 154,2 | 119,9 | 243,4 | 459,1 |
| Донецька | 38,9 | 25,5 | 66 | 135,7 | 218,5 | 375,9 | 399,1 | 55,9 | 66,2 | 205 | 256,5 | 306,8 | 399,3 | 453,6 |
| Житомирська | 0,01 | 0,1 | 0,6 | 1,9 | 2,4 | 5,9 | 7,2 | 1,5 | 3,7 | 3,6 | 4,2 | 5,9 | 8,7 | 11,1 |
| Закарпатська | 4,6 | 6,4 | 13,9 | 27,2 | 21 | 25,1 | 46,1 | 11,8 | 15,9 | 23,6 | 31,3 | 27,7 | 38,6 | 64,7 |
| Запорізька | 40,1 | 40,5 | 27,5 | 41,3 | 68,8 | 104,9 | 269,2 | 52,3 | 70,8 | 120,2 | 102,8 | 122,5 | 156 | 202 |
| Івано-Франківська | 1,1 | 7,3 | 6,2 | 7,2 | 20,6 | 15,5 | 23,8 | 11,9 | 18,3 | 15,9 | 13 | 17,1 | 20,3 | 25,8 |
| Київська | 10,4 | 15,7 | 73,9 | 34,4 | 66,9 | 130,1 | 174,9 | 20,9 | 40,9 | 76,4 | 101,7 | 104,1 | 118,9 | 148,5 |
| Кіровоградська | 0,9 | 3,2 | 4,5 | 4,6 | 3,4 | 5,1 | 5,1 | 5,1 | 9,5 | 11,4 | 11,7 | 15,4 | 16,3 | 16,4 |
| Луганська | 8,5 | 17 | 17,1 | 54 | 41,7 | 94,5 | 81,7 | 20,9 | 17 | 18,8 | 22,9 | 28 | 34,3 | 46,4 |
| Львівська | 10 | 21,3 | 26,9 | 33,6 | 43,6 | 58,5 | 105,9 | 104,9 | 37,6 | 44,7 | 57,6 | 72,8 | 93,1 | 129,9 |
| Миколаївська | 10,2 | 19,3 | 35 | 37,1 | 31,9 | 36 | 46,5 | 51,2 | 53,3 | 83 | 111,4 | 100,5 | 131,6 | 186,8 |
| Одеська | 43,8 | 36,1 | 79,5 | 124,3 | 130,7 | 168,5 | 242,1 | 386,9 | 505 | 657,2 | 830,5 | 879,8 | 1063,9 | 1548,7 |
| Полтавська | 13 | 24,6 | 49,8 | 67,4 | 91,2 | 99,8 | 138 | 135,1 | 10,3 | 17,6 | 26,9 | 23,6 | 29,9 | 41,3 |
| Рівненська | 1,3 | 4,4 | 7,3 | 8,7 | 11,9 | 19,4 | 19,4 | 16,2 | 24,1 | 30 | 31,9 | 50,1 | 55,4 | 51,3 |
| Сумська | 5,1 | 63,3 | 55,2 | 102,4 | 108,9 | 103,8 | 159,4 | 6,4 | 18 | 41,5 | 30,4 | 49,4 | 50,1 | 41,4 |
| Тернопільська | 0,5 | 1,1 | 1,8 | 1,6 | 2,4 | 3,2 | 7,1 | 1 | 2,5 | 3,8 | 4,4 | 6,5 | 8 | 10,1 |
| Харківська | 5,1 | 14,8 | 46,2 | 96,8 | 153,3 | 268 | 428,1 | 24,5 | 34,3 | 48,3 | 58,9 | 97,1 | 241,4 | 234 |
| Херсонська | 0,1 | 0,4 | 3,1 | 6,5 | 8 | 1,8 | 6,2 | 11,5 | 22,8 | 47,5 | 57,9 | 48,4 | 45,9 | 71,3 |
| Хмельницька | 0,8 | 2,5 | 7,3 | 8,4 | 6,9 | 6,4 | 10,5 | 2,8 | 4,9 | 7,2 | 8,7 | 10,2 | 14,2 | 23,7 |
| Черкаська | 3,6 | 6,7 | 11,2 | 12 | 14,7 | 15,2 | 29,3 | 1,4 | 2,9 | 4,5 | 5,6 | 6,6 | 10 | 20,4 |
| Чернівецька | 0,3 | 0,6 | 0,6 | 0,5 | 6,1 | 3,2 | 3,9 | 1,5 | 1,2 | 1,1 | 1,4 | 1,9 | 3,3 | 3,3 |
| Чернігівська | 4,1 | 10,8 | 9,2 | 13,5 | 12,7 | 13,6 | 19,1 | 2,7 | 3,4 | 4,3 | 2,9 | 3,8 | 4,1 | 6,7 |
| м. Київ | 196,6 | 287 | 464,3 | 887,6 | 1364,1 | 1774,9 | 2434,1 | 357,8 | 526 | 762,8 | 1068,7 | 1612,9 | 2255 | 3216,3 |

Значення макроекономічних показників за 2000 – 2009 рр

| Регіон | z3 | | | | | | |
|--------------------------|---------|---------|---------|---------|--------|---------|---------|
| | 2000 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
| Україна | 14572,5 | 23066,8 | 32666,1 | 34228,4 | 38368 | 49296,1 | 66954,4 |
| Авт. Респ. Крим | 173,5 | 317,1 | 274,9 | 355,1 | 400,1 | 463 | 586,5 |
| Вінницька | 199,1 | 333,5 | 421,7 | 424,5 | 403,7 | 485,7 | 482,8 |
| Волинська | 147,7 | 204,8 | 272,8 | 280,9 | 336,3 | 422,4 | 465 |
| Дніпропетровська | 2890,5 | 3588,9 | 5409,9 | 5946,2 | 7118,2 | 9819,2 | 13163,5 |
| Донецька | 2960 | 4960,3 | 8343,6 | 8198,2 | 8776,6 | 10654,9 | 14409,5 |
| Житомирська | 136,7 | 219,1 | 262,2 | 266,8 | 314,3 | 405,1 | 521,2 |
| Закарпатська | 215,6 | 414,6 | 605,5 | 552,7 | 676,9 | 1148 | 1220,6 |
| Запорізька | 1380,7 | 1719,6 | 2369,2 | 2653,3 | 3125,3 | 4324,8 | 5370,1 |
| Івано-Франківська | 191,5 | 680,7 | 867,3 | 813,5 | 885,8 | 1005,4 | 721,7 |
| Київська | 241,2 | 345,2 | 473,9 | 496 | 577,9 | 775,6 | 943,1 |
| Кіровоградська | 51,9 | 100 | 200 | 185,2 | 189 | 242,5 | 391,6 |
| Луганська | 680,7 | 1351,5 | 1899,8 | 2166,9 | 3004,6 | 3963,7 | 6361 |
| Львівська | 303,1 | 495,6 | 639,7 | 621 | 804,8 | 1036,7 | 998,5 |
| Миколаївська | 523,7 | 544 | 654,4 | 881,3 | 1059,7 | 1104,5 | 1649,8 |
| Одеська | 527,8 | 1048 | 1053,8 | 1019,2 | 766,4 | 1098,5 | 2341,9 |
| Полтавська | 440,7 | 1022,9 | 1381,1 | 1883,3 | 1874,7 | 2143,1 | 2574,6 |
| Рівненська | 85,1 | 158 | 202,6 | 235 | 291,4 | 384,7 | 452,1 |
| Сумська | 197,5 | 352,1 | 472,1 | 576,9 | 513,7 | 682 | 913,4 |
| Тернопільська | 42,8 | 79,8 | 92,9 | 76,1 | 81,6 | 124,8 | 144,3 |
| Харківська | 327,7 | 563,3 | 764,1 | 721 | 888,6 | 1143 | 1551,5 |
| Херсонська | 103,4 | 145,3 | 223,5 | 234,6 | 245,6 | 323,2 | 400,3 |
| Хмельницька | 87,5 | 157,8 | 181,2 | 184,8 | 221,1 | 336,7 | 418,1 |
| Черкаська | 224,6 | 291,1 | 412,2 | 339,5 | 482,7 | 690,6 | 867 |
| Чернівецька | 57,9 | 72,2 | 85,7 | 102,3 | 112,1 | 157,5 | 166,3 |
| Чернігівська | 131,8 | 243,2 | 245,5 | 250,4 | 313,1 | 362,5 | 418,6 |
| м. Київ | 1976,6 | 2846,3 | 4148 | 4462,4 | 4426,8 | 5337,3 | 8653,5 |

Зміст

| | |
|---|-----|
| Вступ..... | 3 |
| Тема 1. Парна лінійна регресія..... | 7 |
| Лабораторна робота № 1. Лінійна кореляційна модель | 7 |
| Лабораторна робота № 2. Функції Excel для обробки даних | 13 |
| Тема 2. Множинні регресійні моделі | 19 |
| Лабораторна робота № 3. Регресійний аналіз у матричній формі | 19 |
| Лабораторна робота № 4. Багатовимірні лінійні моделі | 27 |
| Лабораторна робота № 5. Статистичний аналіз даних..... | 33 |
| Лабораторна робота № 6. Мультиколінеарність..... | 41 |
| Лабораторна робота № 7. Довірчі границі на лінію регресії..... | 47 |
| Лабораторна робота № 8. Довірчі границі в загальному випадку | 52 |
| Тема 3. Узагальнені схеми регресійного аналізу | 58 |
| Лабораторна робота № 9. Ридж–регресія | 58 |
| Лабораторна робота № 10. Dummy-змінні | 65 |
| Лабораторна робота № 11. Виробничі функції Коба – Дугласа..... | 77 |
| Тема 4. Системи економетричних рівнянь | 84 |
| Лабораторна робота № 12. Двокроковий метод найменших квадратів..... | 84 |
| Лабораторна робота № 13. Гетероскедастичність. | 93 |
| Лабораторна робота № 14. Автокореляція залишків моделі. | 98 |
| Тема 5. Динамічні економетричні моделі | 104 |
| Лабораторна робота № 15. Моделі з лаговими змінними..... | 104 |
| Лабораторна робота № 16. Використання надбудови “Пакет аналіза” | 109 |
| Лабораторна робота № 17. Дисперсійний аналіз (ANOVA)..... | 116 |
| Використана література | 126 |
| Додатки..... | 127 |

НАВЧАЛЬНЕ ВИДАННЯ

Єгоршин Олександр Олександрович

Малярець Людмила Михайлівна

**Лабораторний практикум з економетрики навчальної
дисципліни "Економіко–математичні методи та моделі"**

Начально–практичний посібник

Відповідальний за випуск **Малярець Л. М.**

Відповідальний редактор

Редактор

Коректор

План 2010 р. Поз. №30-П.

Підп. до друку

Формат 60 × 90 1/16. Папір MultiCopy. Друк Riso.

Ум.-друк. арк. ___. Обл.-вид. арк. ___. Тираж прим. Зам. №

Видавець і виготівник — видавництво ХНЕУ, 61001, м. Харків, пр. Леніна, 9а
