

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ,
МОЛОДІ ТА СПОРТУ УКРАЇНИ**

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

**Лабораторний практикум
з модуля 2 "Методи і моделі статистичного
дослідження рядів динаміки" навчальної дисципліни**

"СТАТИСТИКА"

**для студентів напряму підготовки
6.030506 "Прикладна статистика"
денної форми навчання**

Харків. Вид. ХНЕУ, 2012

Затверджено на засіданні кафедри статистики та економічного прогнозування.

Протокол № 5 від 12.12.2011 р.

Укладачі: Раєвнєва О. В.

Горохова О. І.

Чанкіна І. В.

Шаповалова В. О.

Дмитрусенко К. О.

Бобкова О. Ю.

Л12 Лабораторний практикум з модуля 2 "Методи і моделі статистичного дослідження рядів динаміки" навчальної дисципліни "Статистика" для студентів напряму підготовки 6.030506 "Прикладна статистика" денної форми навчання / укл. Раєвнєва О. В., Горохова О. І., Чанкіна І. В. та ін. – Х. : Вид. ХНЕУ, 2012. – 64 с. (Укр. мов.)

Подано лабораторні роботи, метою яких є закріплення й поглиблення знань теоретичного та практичного матеріалу, придбання навичок аналізу різних типів даних за допомогою системи STATISTICA.

Рекомендовано для студентів напряму підготовки 6.030506 "Прикладна статистика".

Вступ

Лабораторні роботи призначені для закріплення теоретичного й практичного матеріалу, набуття навичок роботи сучасним пакетом прикладних програм, що забезпечують побудову й дослідження різних типів моделей, а також розширюють знання студентів у галузі застосування прикладних пакетів для економічних розрахунків та прогнозування.

Для виконання лабораторних робіт пропонується використовувати пакети прикладних програм STATISTICA та Excel. Пакети містять множини статистичних методів, що підтримують вирішення різних економіко-статистичних задач, що складають зміст дисципліни "Статистика".

Важливою характеристикою системи STATISTICA є те, що вона є повноцінною прикладною програмою Windows.

Пакет Excel містить множини статистичних методів, що підтримують вирішення різних статистичних та економіко-математичних задач. При розробці лабораторних робіт передбачається, що студент знайомий з основними принципами й прийомами роботи в середовищі Windows.

Кожна лабораторна робота розглянута на прикладі вирішення конкретної задачі й забезпечена докладними коментарями й рисунками. Лабораторні роботи стосуються основних тем дисципліни й ґрунтуються на теоретичному матеріалі відповідної теми, а також попередніх тем. Кожна робота містить мету й завдання та методичні рекомендації щодо їх виконання.

Оцінка за виконання роботи ставиться за результатами виконання й захисту лабораторної роботи. Певна увага приділяється знанню студентами теоретичного матеріалу, правильності висновків і повноті економічної інтерпретації отриманих результатів.

Кваліфікаційні вимоги в галузі статистика до студентів

Необхідна навчальна база перед початком вивчення дисципліни

З метою кращого засвоєння навчального матеріалу дисципліни студенти повинні до його початку опанувати знаннями та навичками в галузі загальної теорії статистики.

У свою чергу знання з даної дисципліни забезпечують успішне засвоєння виконання курсових проектів і дипломних проектів.

У результаті вивчення дисципліни студенту необхідно набути таких компетенцій (табл. 1).

Таблиця 1

Вимоги до компетенцій, які повинен здобути студент, вивчаючи навчальну дисципліну

Назва компетенцій	Зміст компетенцій	Вміння студента щодо даної компетенції
1	2	3
1. Обліково-статистична	1.1. Здатність проводити статистичні розрахунки	1.1.1. Проводити розрахунок показників для перевірки статистичних гіпотез. 1.1.2. Проводити розрахунок показників динамічних рядів
2. Аналітична	2.1. Здатність щодо проведення статистичного аналізу	2.1.1. Визначати чинники, що впливають на зміну показників. 2.1.2. Прогнозувати найімовірніший перебіг процесів на перспективу та аналізувати й оцінювати можливі наслідки зміни явища
3. Організаційно-методологічна	3.1. Здатність впроваджувати статистичну методологію	3.1.1. Володіти методами прогнозування. 3.1.2. Володіти методиками дослідження видів тенденцій у зміні економічних та соціальних явищ. 3.1.3. Володіти методами оцінки загальної тенденції явища.

Закінчення табл. 1

1	2	3
		3.1.4. Володіти методами оцінки стохастичних явищ. 3.1.5. Володіти методами прогнозу екстраполяції на основі трендових моделей 3.1.6. Володіти методами коливань та сталості динамічних рядів 3.1.9. Володіти методами побудови однофакторних та багатофакторних регресійних моделей. 3.1.10. Володіти методом побудови автокореляції
4. Контрольна	4.1. Здатність будувати однофакторні та багатофакторні моделі та перевіряти їх на адекватність	4.1.1. Готувати інформацію, вибирати тип моделей, здійснювати розрахунки їх параметрів та перевіряти на адекватність. 4.1.2. Використовувати відповідні критерії для аналізу рівня достовірності прогнозних оцінок
5. Інформаційна	5.1. Здатність використовувати при вирішенні економічних завдань сучасні засоби обробки інформації	5.1.1. Використовувати різноманітні можливості пакета прикладних програм STATISTICA та Excel для аналізу соціально-економічних процесів чи явищ з метою підвищення якості управлінських рішень

Лабораторна робота № 8 "Перевірка статистичних гіпотез"

Мета – закріплення теоретичного та практичного матеріалу, придбання навичок перевірки статистичних гіпотез у модулі t-test, single sample, t-test, independent sample, Distribution Fitting.

Завдання – необхідно підтвердити гіпотези про рівність середніх та про нормальність розподілу за допомогою критерію Колмогорова-Смирнова.

Методичні рекомендації

Гіпотеза – один із важливих чинників руху науки до досягнення прогресу. Виникаючи як результат спостереження за явищами (фактами), гіпотеза приймає форму теоретичного припущення. Звернення до фактів допускає можливість перевірки цього припущення. При цьому факти, якими перевіряється гіпотеза, повинні бути науково обґрунтованими, тобто бути результатом спостереження, що базується на наукових принципах.

1. Перевірка параметричних гіпотез щодо рівності генеральної середньої нормально розподіленої генеральної сукупності встановленому значенню.

У системі "STATISTICA" створіть файл розмірності 2 x 30 (2 змінні, 30 спостережень). Змінним надайте імена й характеристики відповідно до табл. 2.

Таблиця 2

Імена й характеристики змінних

Змінна	Ім'я	Формат відображення	Тип	Кількість десяткових знаків	Дані для поля "Довге ім'я або формула"
Var1	Вибірка 1	Числовий	Double	1	=VNormal(rnd(1);5;0,5)
Var2	Вибірка 2	Числовий	Double	1	=VNormal(rnd(1);4;0,5)

Виконані дії дозволять згенерувати дві вибірки, що вибрані з нормально розподілених генеральних сукупностей з різними генеральними середніми значеннями ($a_1 = 5$ і $a_2 = 4$) і однаковим генеральним середньоквадратичним відхиленням ($\sigma = 0,5$). Збережіть файл.

У пункті меню "Statistics" виберіть пункт "Basic statistics/Tables" (Основні статистики/Таблиці). У вікні, що відкрилося, виберіть пункт "Descriptive statistics" (Описові статистики) і відкрийте вікно розрахунку описових статистик. Перейдіть від вкладки "Quick" (Швидкий) до вкладки "Advanced" (Розширений). На даній вкладці відзначте показники, як показано на рис. 1.

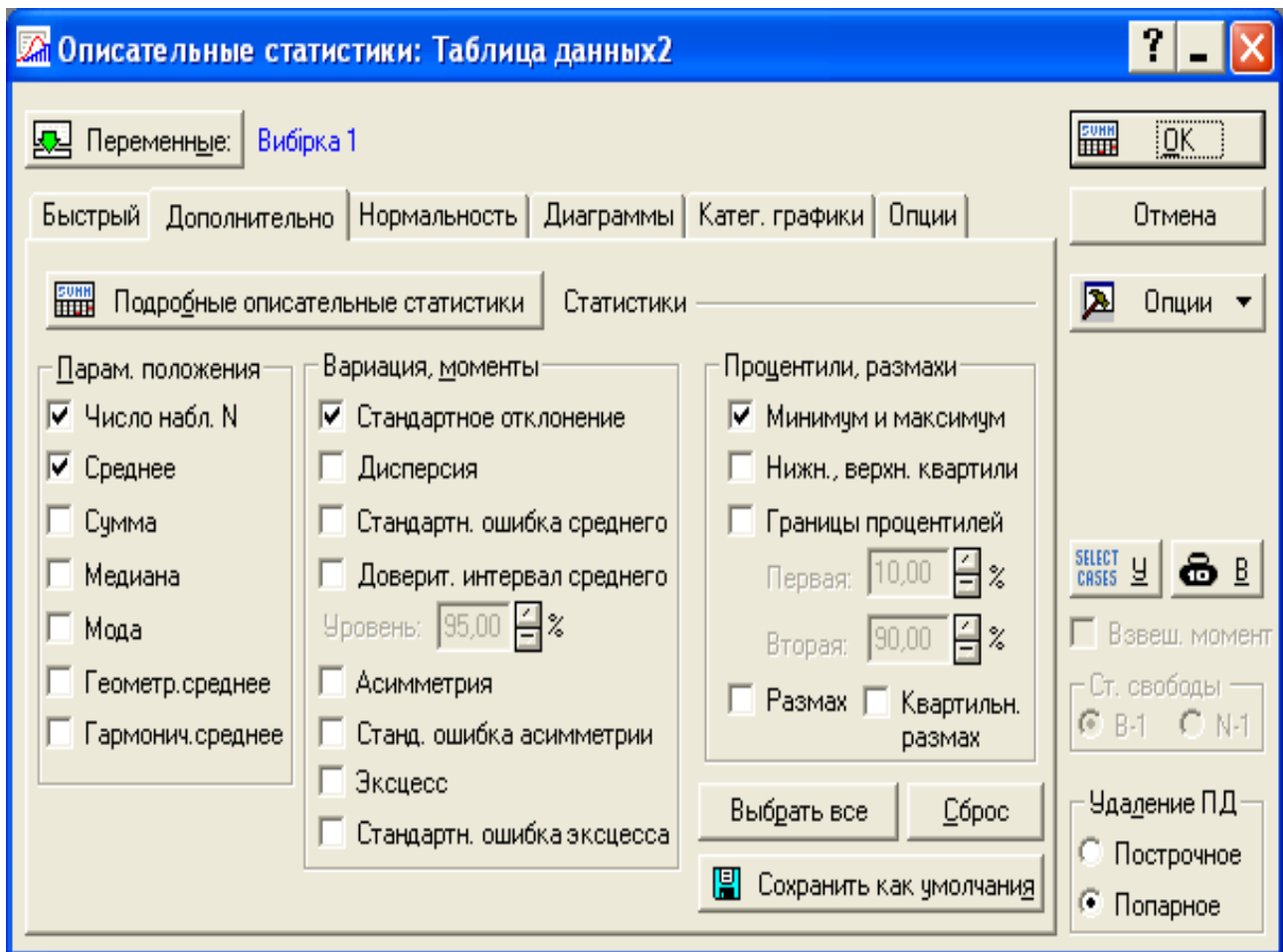


Рис. 1. Вікно розрахунку описових статистик. Вкладка "Advanced (Додатково)"

Установивши прапорці так, як показано на рис. 1, система зробить розрахунок кількості спостережень (Valid N), вибіркового середнього арифметичного (Mean) і вибіркового середньоквадратичного відхилення (Standard Deviation).

Для перевірки гіпотез про рівність генеральних середніх значенням $H_0 : a_1 = a_0 = 5$ і $H_0 : a_2 = a_0 = 4$ відповідно закрийте вікно описових статистик і у вікні "Basic statistics/Tables (Основні статистики/Таблиці)" виберіть пункт "t-test, single sample" (рис. 2).

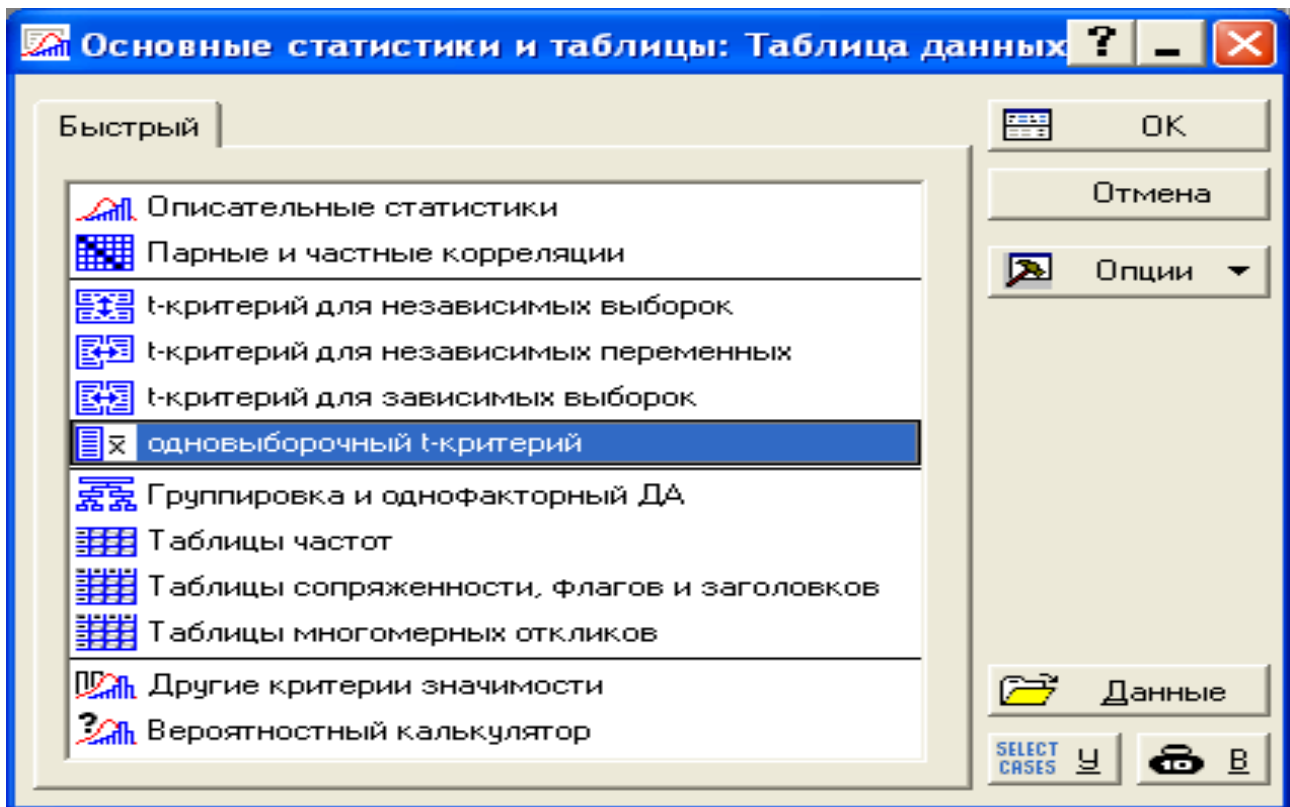


Рис. 2. Початкове вікно розрахунку описових статистик

У результаті буде відкрите вікно перевірки простих параметричних гіпотез щодо генеральних середніх (рис. 3).

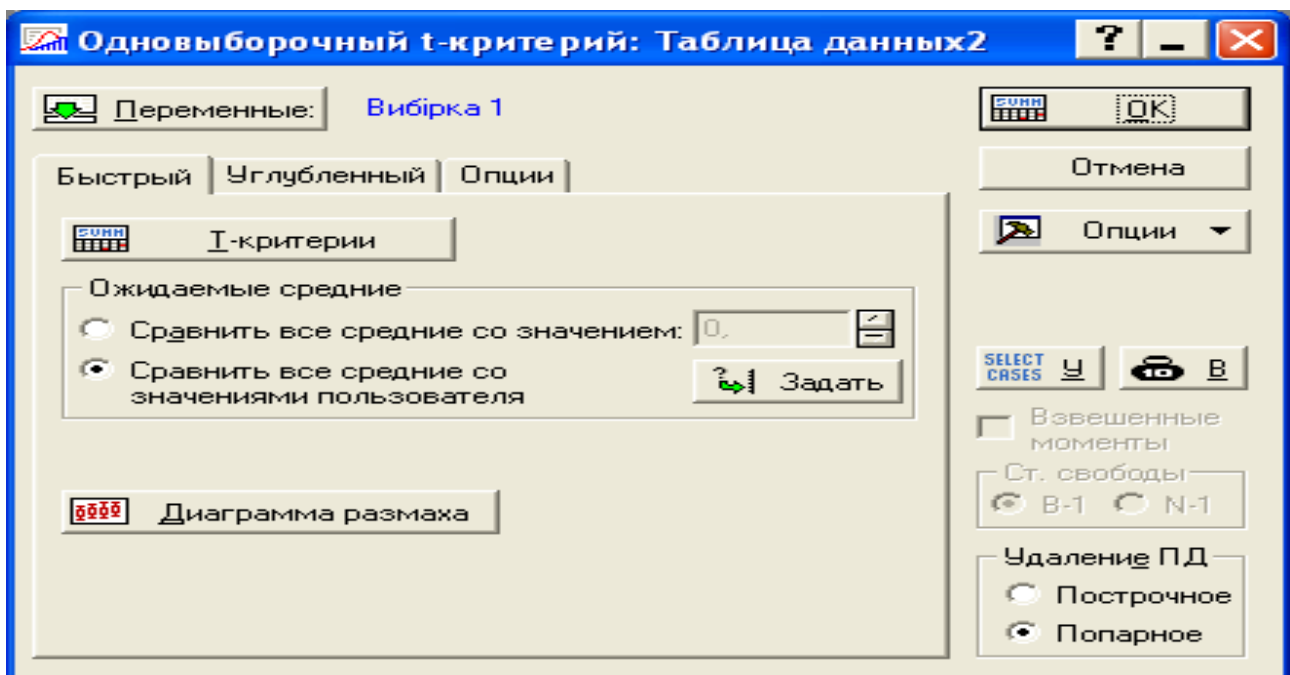


Рис. 3. Вікно перевірки параметричних гіпотез щодо середнього значення генеральної сукупності

Виберіть змінну "Вибірка 1" (натиснувши на кнопку Variables) і відзначте пункт "test means against different user-defined constant" (що можна перевести як "гіпотеза для середнього проти нерівності встановленої користувачем константі"). У цьому випадку ми перевіряємо гіпотезу $H_0 : a_1 = a_0$ при $H_1 : a_1 \neq a_0$.

Перейдіть на вкладки "Advanced" ("Розширене") і переконайтесь, що в поле "p-level for highlighting" (рис. 4) встановлене значення рівня значимості при перевірці гіпотези рівне 0,05. При такій установці значення t-статистики Стюдента й вихідні за двосторонні критичні границі розподілу Стюдента з n-1 ступенем свободи, що відповідають встановленому рівню значимості, будуть виділені червоним кольором.

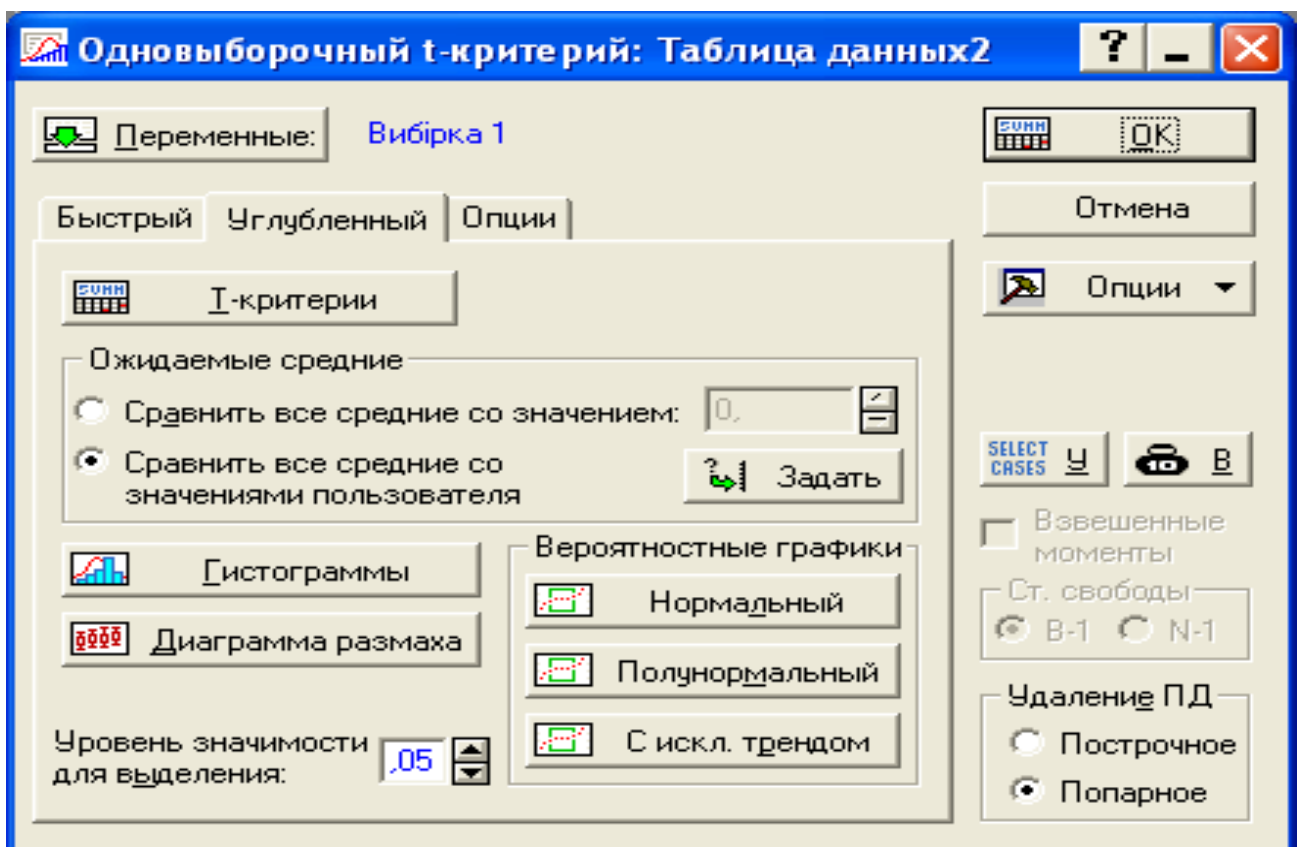


Рис. 4. Вікно установки додаткових параметрів при перевірці гіпотези щодо середнього значення

Натисніть кнопку Specify і встановити у вікні, що відкрилося, значення генеральної середньої в першій сукупності $a_0 = 5$ (рис. 5).

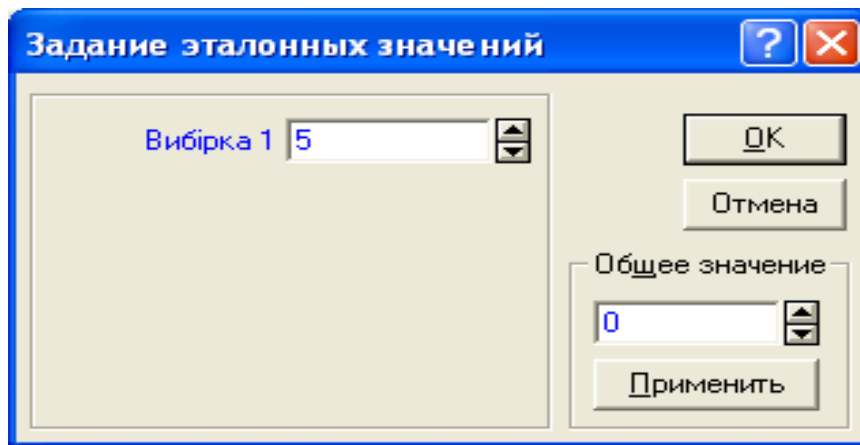


Рис. 5. Вікно установки значення параметра a_0

Після натискання кнопки "OK" натисніть кнопку Summary: T-tests. Система зробить розрахунок необхідних значень і представить результати так, як показано на рис. 6 (зверніть увагу, що результати, які вийшли у вас, можуть відрізнятись від представлених).

Сравнение выборочных средних с константой (Таблица данных2)								
Переменная	Среднее	Стд. от.	N	Стд. ош.	Константа Constant	t-знач.	Ст. св.	p
Вибірка 1	5,041098	0,435018	30	0,079423	5,000000	0,517457	29	0,608762

Рис. 6. Результати перевірки статистичної гіпотези щодо рівності генеральної середньої, що перевіряється по "Вибірці 1", значенню $a_0 = 5$

Проаналізуємо результати, представлені в таблиці на рис. 6. Це: вибіркове середнє значення (Mean) – 5,041098;

вибіркове середньоквадратичне відхилення (Std.dev.) – 0,435018;

кількість спостережень (N) – 30;

середньоквадратичне відхилення середньої арифметичної (Std.err) – 0,079423;

значення константи, щодо якої перевіряється гіпотеза (Reference Constant) – 5;

розрахункове значення статистики Стьюдента (t-value) – 0,517457;

число ступенів волі статистики Стьюдента (df = N - 1) – 29;

імовірність того, що значення статистики (по модулю) можуть

прийняти такі ж або ще більші значення при справедливості нуль-гіпотези (p) – 0,608762.

Зверніть увагу на те, що значення ймовірності p , представлено в останньому стовпчику таблиці, істотно більше встановленого нами рівня значимості $\alpha = 0,05$. Останнє означає, що при справедливості нуль-гіпотези таке значення статистики може бути обумовлено випадковими причинами. А це, у свою чергу, говорить про те, що у нас немає підстав не прийняти нуль-гіпотезу.

2. Перевірка параметричної гіпотези щодо рівності двох генеральних середніх двох нормально розподілених генеральних сукупностей.

У цьому випадку ми намагаємося перевірити статистичну гіпотезу такого змісту: $H_0: a_1 = a_2$ при $H_1: a_1 \neq a_2$. Зазначена гіпотеза перевіряється в припущенні, що в обох генеральних сукупностях дисперсії є рівними. Іншими словами, вихідна гіпотеза повинна перевірятися разом з гіпотезою $H_0: \sigma_1^2 = \sigma_2^2$ при $H_1: \sigma_1^2 > \sigma_2^2$.

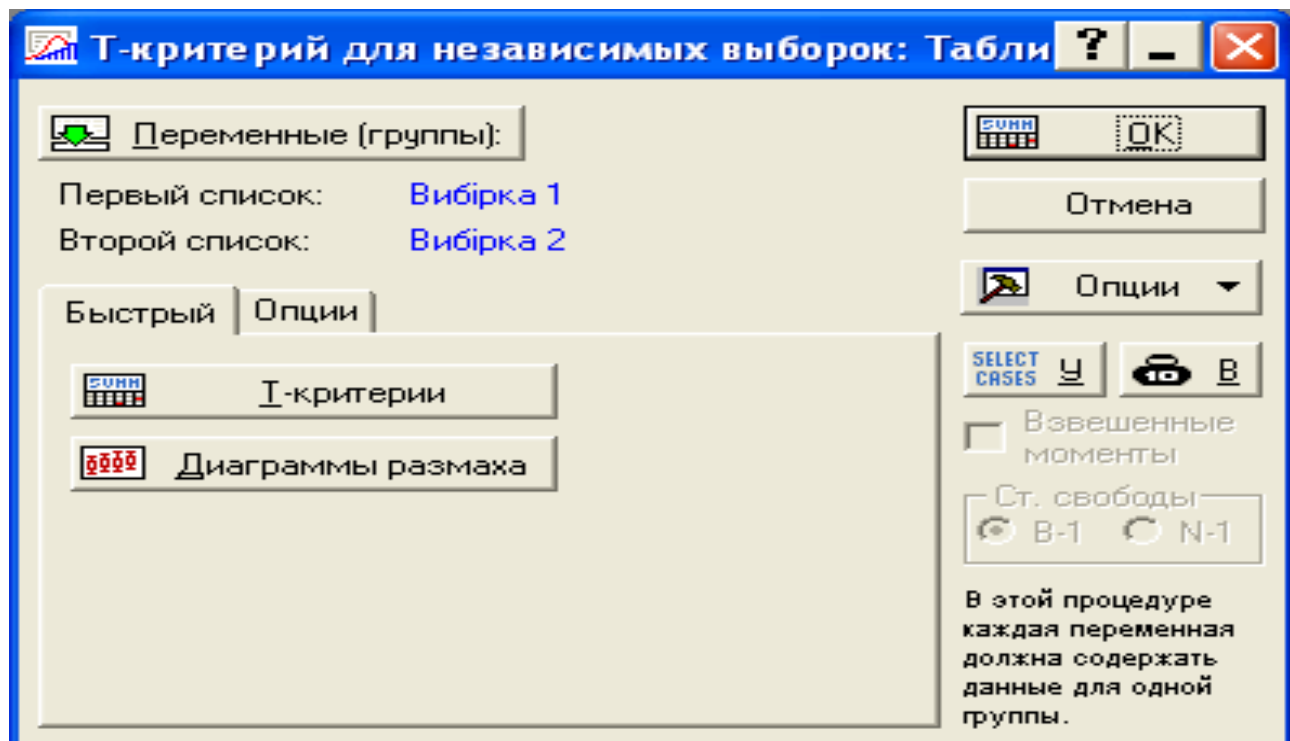


Рис. 7. Вікно перевірки статистичної гіпотези про рівність двох генеральних середніх двох нормально розподілених генеральних сукупностей

Для перевірки цих гіпотез поверніться до вікна "Basic statistics/Tables (Основні статистики/Таблиці)", виберіть пункт "t-test, independent sample (Т-критерій для незалежних вибірок)" (див. рис. 2). Натисніть ОК. У вікні, що відкрилося, встановіть параметри так, як показано на рис. 7. Переконайтеся, що на вкладці "Options" у поле "p-level for highlighting" встановлене значення 0,05.

Натисніть кнопку Summary: T-tests. Система здійснить розрахунок t-статистики (статистики Стьюдента), F-статистики (статистики Фішера), зробить ряд додаткових обчислень і представить результати, як показано на рис. 8.

		Т-критерий независимых выборок (Таблица данных2)										
		Замечание: Переменные рассм. как независимые выборки										
Группа 1 и Группа 2		Среднее	Среднее	t-знач.	ст. св.	р	N набл.	N набл.	Ст. откл.	Ст. откл.	F-отн.	р
		Группа 1	Группа 2				Группа 1	Группа 2	Группа 1	Группа 2	Дисперсии	Дисперсии
Вибірка 1 vs. Вибірка 2		5,041098	3,995574	9,642634	58	0,000000	30	30	0,435018	0,404293	1,157768	0,695878

Рис. 8. Результати перевірки гіпотези про рівність генеральних середніх двох нормально розподілених генеральних сукупностей

Проаналізуємо отримані результати. У таблиці розраховані: вибіркові середні значення (Mean Group 1 і Mean Group 2) у першій і другій вибірках – 5,04, 3,995 відповідно;

значення середньоквадратичних відхилень (Std.Dev. Group 1 і Std.Dev. Group 2) у першій і другій вибірках – 0,435, 0,404 відповідно;

кількість спостережень в обох вибірках (Valid N Group 1 і Valid N Group 2) – 30;

значення t-статистики (статистики Стьюдента – t) – 9,6426;

число ступенів волі t-статистики ($df = n1 + n2 - 2$) – 58;

імовірність того, що при справедливості нуль-гіпотези щодо середніх, значення t-статистики (по модулю) можуть прийняти такі ж або ще більші значення (p) – 0,00000;

значення F-статистики (статистики, що перевіряє гіпотезу про рівність дисперсій у першій і в другій вибірці) – 1,1577;

імовірність того, що при справедливості нуль-гіпотези щодо дисперсій, значення F-статистики можуть прийняти такі ж або ще більші значення (p) – 0,695878.

Зверніть увагу на те, що значення ймовірності p , представлене в останньому стовпчику таблиці, більше встановленого нами рівня значимості $\alpha = 0,05$. Це свідчить про те, що при гіпотезі $H_0 : \sigma_1^2 = \sigma_2^2$ імовірність одержати такі ж або більші значення F-статистики становить 0,695878. Ця ймовірність досить велика, і, отже, отримана розбіжність може бути пояснена випадковими причинами. А це, у свою чергу, говорить про те, що в нас немає підстав відкинути нуль-гіпотезу про рівність дисперсій у двох генеральних сукупностях.

Тепер зверніть увагу на те, що значення ймовірності p , представлене в п'ятому стовпчику таблиці, істотно менше встановленого нами рівня значимості $\alpha = 0,05$. Це говорить про те, що при справедливості нуль-гіпотези $H_0 : \mu_1 = \mu_2$ імовірність одержати такі ж або більші значення t-статистики практично дорівнює нулю, тобто отримана розбіжність не може бути пояснена випадковими причинами. А це, у свою чергу, свідчить про те, що нуль-гіпотезу ми повинні відхилити й вважати, що генеральні середні в першій і другий генеральних сукупностях розрізняються значимо.

3. З непараметричних гіпотез щодо законів розподілу за критерієм згоди Пірсона χ^2 .

Створіть файл із даними розмірності 5 x 100 (5 змінних 100 спостережень). Змінним привласніть імена й характеристики відповідно до табл. 3.

Таблиця 3

Імена та характеристики змінних

Змінна	Ім'я	Формат відображення	Тип	Кількість десяткових знаків	Рядок в полі "Довге ім'я або формула"
Var1	Рівномірний розподіл	Числовий	Цілочисельний (Integer)	0	=Rnd(5)
Var2	Нормальний розподіл	Числовий	Double	1	=VNormal (rnd(1);2;0,5)
Var3	Показовий розподіл	Числовий	Double	1	=VExpon (rnd(1); 0,2)
Var4	Розподіл Релея	Числовий	Double	1	=VRayl(rnd (1);0,5)
Var5	Розподіл Пуассона	Числовий	Цілочисельний (Integer)	0	=RndPoisson (5)

Після натискання кнопки "ОК" уведені характеристики дозволяють згенерувати дані, розподілені по деяких теоретичних законах розподілу випадкових.

У меню "Статистика" виберіть пункт Distribution Fitting (підгонка розподілів). У результаті вибору відкриється початкова панель перевірки гіпотез щодо законів розподілу з використанням критеріїв згоди (рис. 9). У цьому вікні необхідно вибрати тип розподілу для аналізованої змінної.

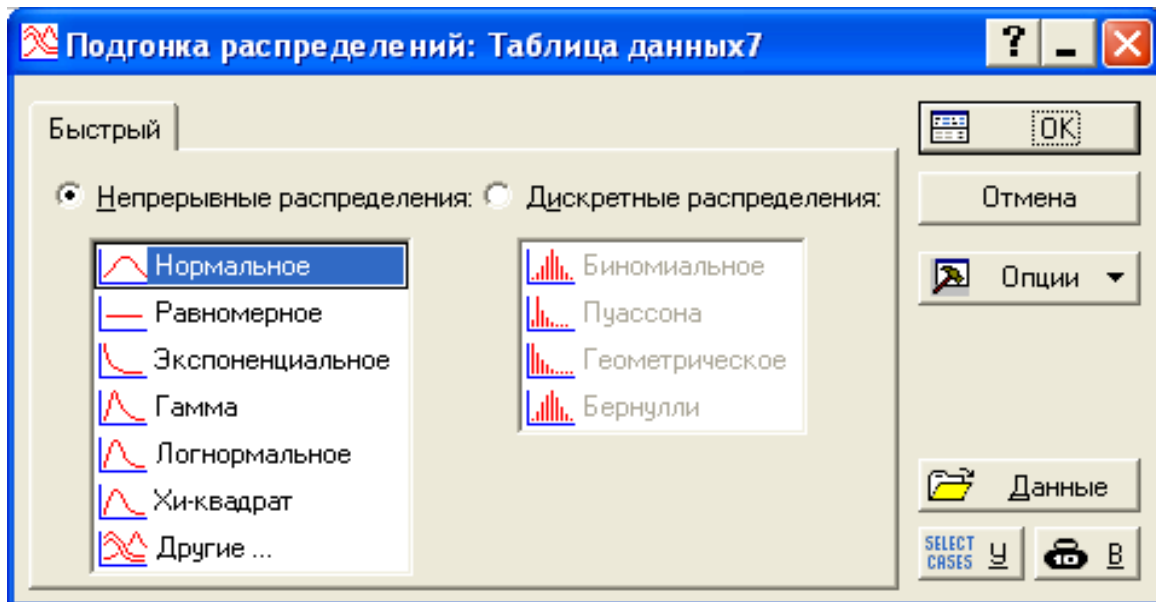


Рис. 9. Панель вибору розподілу при перевірці гіпотез щодо закону розподілу

Так як перша змінна файла вихідних даних повинна підкорятися закону рівної ймовірності, то в початковому вікні необхідно вибрати тип розподілу Rectangular. Іншими словами, ми хочемо перевірити гіпотезу H_0 : вибірка витягнута з генеральної сукупності, що підкоряється закону розподілу рівної ймовірності. Натисніть "ОК". У результаті відкриється вікно вибору змінних і умов перевірки гіпотези щодо закону розподілу вибірки за критеріями згоди.

Виберіть змінну "Рівномірний розподіл" і перейдіть на вкладку "Параметри". На цій вкладці задаються умови перевірки гіпотези H_0 . Змініть значення параметрів аналізу так, як показано на рис. 10, і перейдіть назад на вкладку "Швидке (Quick)".

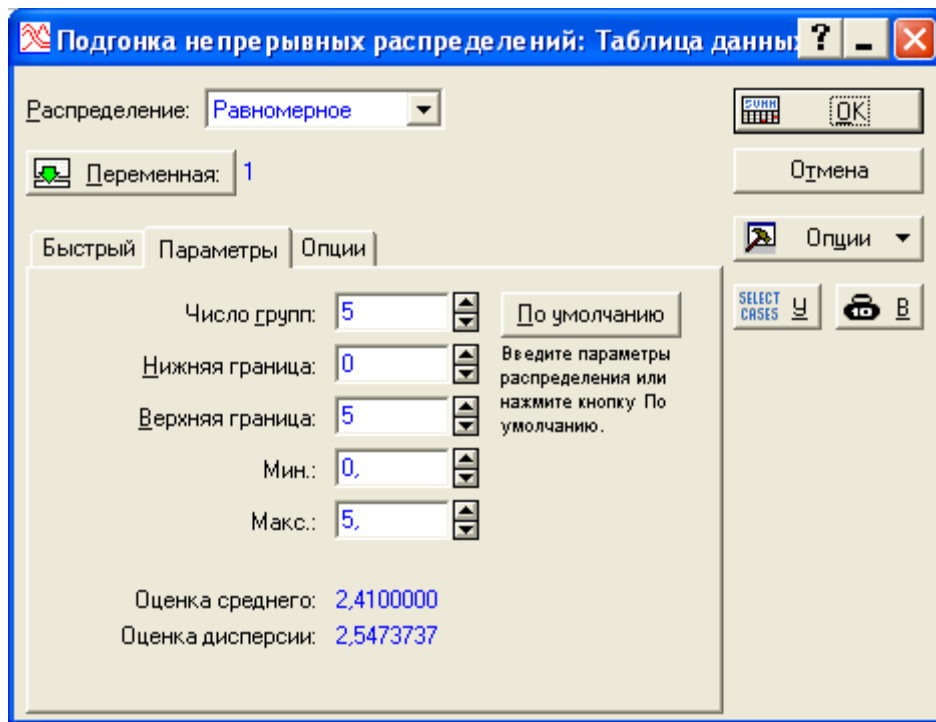



Рис. 10. Установлені параметри перевірки гіпотези H_0

Натиснете на кнопку  Summary: Observed and expected distribution (підсумок, спостережувані й розраховані частоти). У результаті відкриється вікно з результатами розрахунку критерію згоди Пірсона χ^2 (рис. 11).

Верхняя граница	Наблюд. частота	Кумул. Наблюд.	Процент Наблюд.	Кумул. % Наблюд.	Ожидаем. частота	Кумул. Ожидаем.	Процент Ожидаем.	Кумул. % Ожидаем.	Наблюд.- Ожидаем.
<= 1,00000	33	33	33,00000	33,0000	20,00000	20,0000	20,00000	20,0000	13,00000
2,00000	20	53	20,00000	53,0000	20,00000	40,0000	20,00000	40,0000	0,00000
3,00000	18	71	18,00000	71,0000	20,00000	60,0000	20,00000	60,0000	-2,00000
4,00000	17	88	17,00000	88,0000	20,00000	80,0000	20,00000	80,0000	-3,00000
< бесконеч.	12	100	12,00000	100,0000	20,00000	100,0000	20,00000	100,0000	-8,00000

Рис. 11. Результати перевірки гіпотези H_0

Проаналізуємо отримані результати.

У першому стовпчику таблиці показані інтервали групування значень змінної ($[-\infty, 1)$; $[1,2)$; ...; $[4, +\infty)$). Далі по стовпчиках представлені:

- спостережувані частоти (Observed Frequency);
- накопичені спостережувані частоти (Cumulative Observed);

- частки(Percent Observed);
- накопичені частки(Cumulative % Observed);
- розрахункові частоти (Expected Frequency);
- накопичені розрахункові частоти (Cumulative Expected);
- розрахункові частки (Percent Expected). Розрахункові частки – це аналог імовірності влучення значень досліджуваної ознаки у відповідний інтервал при справедливості гіпотези, що перевіряється;
- накопичені розрахункові частки (Cumulative % Expected);
- різниця накопичених і розрахункових частот (Observed–Expected). У верхньому рядку представлені такі результати:
 - змінна, щодо якої перевіряється гіпотеза H_0 (рівномірний розподіл);
 - вид закону розподілу (Distribution: Rectangular (закон рівної ймовірності) та найменування файла з вихідними даними;
 - розрахункове значення статистики χ^2 ;
 - число ступенів волі статистики χ^2 (df);
 - значення ймовірності p , що відповідає розрахунковому значенню статистики χ^2 .

Властиво трьом останнім значенням і дозволяє зробити висновок про те, чи можна ні приймати гіпотезу H_0 . Припустимо, що ми хочемо перевірити гіпотезу H_0 на рівні значимості $\alpha = 0,05$.

Зверніть увагу на те, що значення ймовірності $p = 0,00213$, представлене у верхньому рядку таблиці результатів, менше встановленого нами рівня значимості $\alpha = 0,05$. Останнє означає, що при справедливості гіпотези H_0 таке значення статистики не може бути обумовлено випадковими причинами. А це, у свою чергу, говорить про те, що в нас немає підстав прийняти гіпотезу H_0 (іншими словами, не можна вважати, що вибірка підкоряється закону рівної ймовірності).

Закрийте вікно результатів і поверніться у вікно аналізу. Натисніть на вкладці "Quick" (Швидке) кнопку "Plot of observed and expected distribution". У результаті буде побудована гістограма розподілу вибірки з накладеною на неї функцією щільності закону рівної ймовірності (рис. 12).

Змінна: Рівномірний розподіл, Розподіл: Рівномірний
Критерій Хі-квадрат = 12,30000,
число ступенів свободи = 2,
вірогідність p = 0,00213

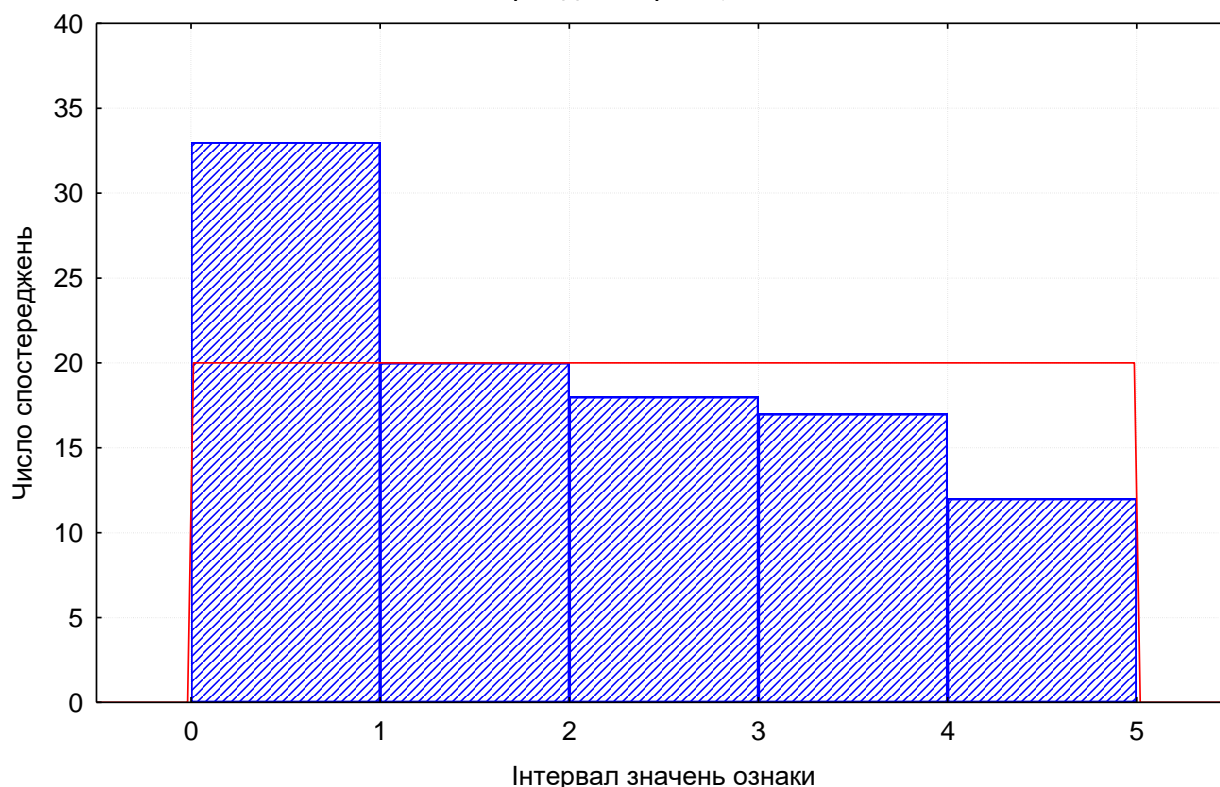


Рис. 12. Гістограма розподілу змінної з накладеною функцією щільності закону рівної ймовірності й результатами розрахунку

Аналізуючи графік, представлений на рис. 12 і дані заголовка графіка, також можна зробити висновок про те, що вихідні дані погано узгоджуються із законом рівної ймовірності, незважаючи на те, що вибірка була отримана шляхом генерації даних, що підкоряються саме цьому закону. Останнє може бути обумовлене тим, що критерій згоди Пірсона є чутливим до обсягу спостережень і характеру групування одиниць вибірки по інтервалах (або групах).

Для перевірки такої чутливості досліджуйте поведінку статистики χ^2 при зміні числа інтервалів групування з 5 на яке-небудь інше значення. Зробіть відповідні висновки.

Для встановлення нового числа груп поверніться у вікно аналізу на вкладку "Parameters" (Параметри) і в поле Number of categories установіть необхідне вам число груп.

Завдання. Для змінних: Нормальний розподіл, Експонентний розподіл і розподіл Пуассона – перевірте гіпотези про відповідність розподілів вибірових значень змінних зазначеним законам розподілу випадкових величин. Вибір кількості інтервалів надайте зробити системі STATISTICA.

Лабораторна робота № 9 "Побудова однофакторної регресійної моделі"

Мета – закріплення теоретичного й практичного матеріалу, придбання навичок побудови однофакторних регресійних моделей.

Завдання – необхідно перевірити наявність лінійного зв'язку між відповідними показниками в модулі Statistics – Multiple Regression.

Методичні рекомендації

Регресійний аналіз застосовується для визначення виду зв'язку і дає можливість для прогнозування значення однієї (залежної) змінної відштовхуючись від значення іншої (незалежної) змінної.

Рівняння $y = \beta_1x + \beta_0 + \varepsilon$ називається рівнянням простої лінійної регресії, та має такі складові:

x – незалежна змінна;

y – залежна змінна;

β_1, β_0 – невідомі параметри, що описують пряму, яку потрібно оцінити за результатами спостережень $(x_i, y_i), i = 1, n$;

ε – незалежна випадкова помилка з нульовим середнім, яка інтерпретується як помилка спостережень.

Потрібно побудувати рівняння простої регресії за результатами спостереження (x_i, y_i) . У цьому випадку одержати точні значення параметрів β_1, β_0 неможливо, оскільки потрібна нескінченна вибірка, тому отримують їх оцінки b_1, b_0 , і рівняння регресії записується у вигляді $y = b_1x + b_0$. Побудову цього рівняння з допомогою системи STATISTICA

розглянемо на прикладі даних (табл. 4), що становлять оптові ціни за одну пляшку на марочні вина(y) залежно від року виготовлення вина (x).

Після запуску системи необхідно вибрати модуль Multiple Regression – множинна регресія, після запуску якого з'являється "Стартова панель" модуля, що містить основні операції, доступні в модулі, і різні параметри аналізу.

Таблиця 4

Оптові ціни за одну пляшку на марочні вина (y) залежно від року закладки вина (x)

Рік (x)	Ціна (y)
1890	50,00
1900	35,00
1920	25,00
1931	11,98
1934	15,00
1935	13,00
1940	6,98
1941	10,00
1944	5,99
1948	8,98
1950	6,98
1952	4,99
1955	5,98
1960	4,98

На першому кроці необхідно створити електронну таблицю з вихідними даними, в якості змінних у розглянутому прикладі виступають "Рік" закладки вина і його "Ціна" на аукціоні

Для початку проведення статистичного аналізу викликається "Стартова панель" модуля, в якому задаються різні параметри аналізу. Вибирається залежна змінна "Ціна" і незалежна змінна "Вік". Для завдання змінних використовують кнопку Variables – Змінні, при натисканні на яку відкривається вікно Select dependent and independent variable list – Вибір залежної змінної і списку незалежних змінних, в якому вибираються відповідні змінні (Ціна, вік) (рис. 13). Натисканням на кнопку ОК повертаємося в "Стартову панель" модуля (рис. 14).

Для виведення результатів і їх аналізу на стартовій панелі необхідно натиснути на кнопку ОК. Система робить обчислення, і через секунду вікно результатів з'явиться на екрані. Вікно результатів аналізу має таку просту структуру: верхня частина вікна – інформаційна, нижня містить функціональні кнопки, що дозволяють всебічно ознайомитись з результатами аналізу (рис. 15).

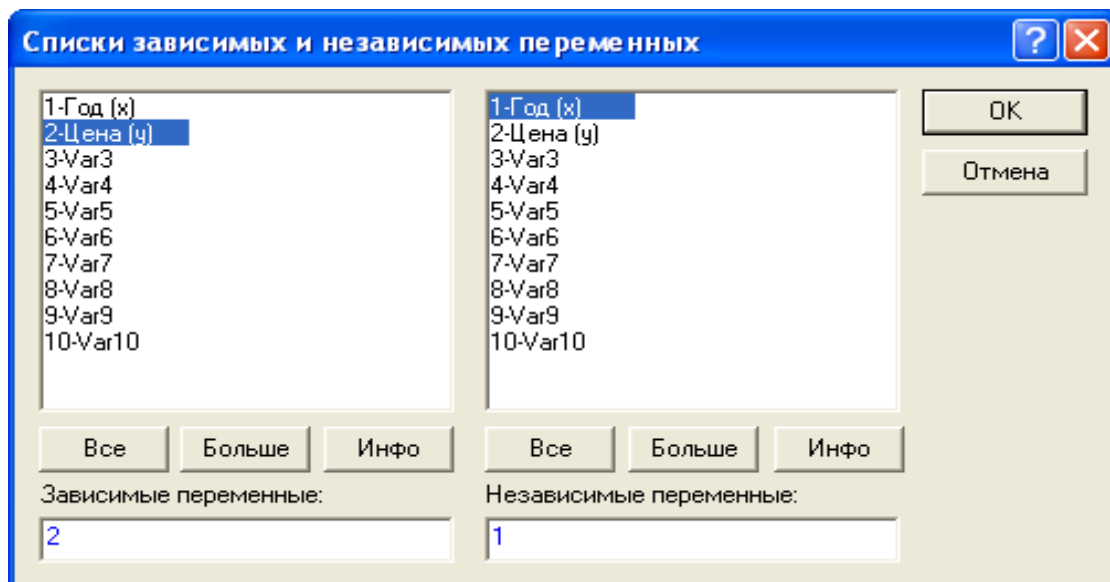


Рис. 13. Вибір залежної змінної і списку незалежних змінних

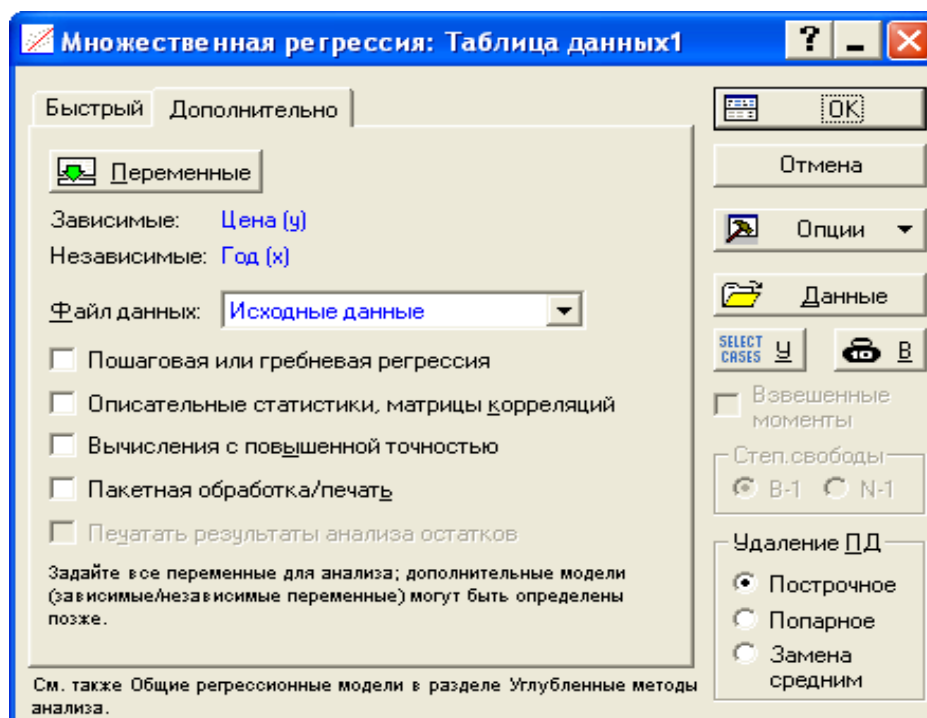


Рис. 14. Стартову панель модуля "Множинна регресія"

В інформаційній частині вікна міститься коротка інформація про проведений аналіз, а саме:

Dep. Var. – ім'я залежної змінної: Вартість;

No. Of Cases – число випадків, по яких побудована регресія: 14;

Multiple R – коефіцієнт множинної кореляції;

R-square – квадрат коефіцієнта множинної кореляції;

Adjusted R-square – скоригований коефіцієнт кореляції, який визначається як $\text{Adjusted R-square} = 1 - (1 - R\text{-Square}) (n / (n - p))$, де n – число спостережень в моделі; p – число параметрів моделі;

Std. error of estimate – стандартна помилка оцінки. Ця статистика є мірою розсіювання спостережуваних значень щодо прямої регресії;

Intercept – оцінка вільного члена регресії. Значення коефіцієнта b_0 у рівнянні регресії;

Std. Error – стандартна помилка оцінки вільного члена. Стандартна помилка коефіцієнта b_0 у рівнянні регресії;

t(df) and p-value – значення критерію Стюдента та рівня помилки. Критерій Стюдента використовується для перевірки гіпотези про рівність нулю вільного члена регресії.

F – значення критерію Фішера.

df – Число ступенів свободи F-критерію.

p – рівень значущості.

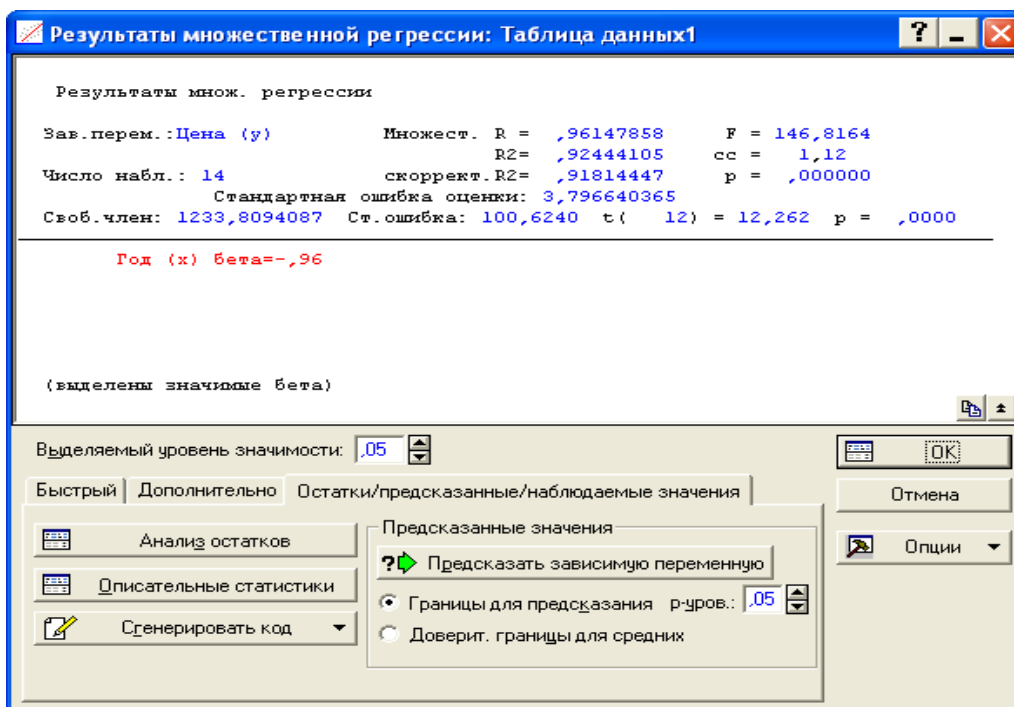


Рис. 16. Вікно результатів аналізу множинної регресії

У розглянутому прикладі значення коефіцієнта детермінації $R^2 = 0,924$, що означає, що побудована регресія пояснює 92,4 % розкиду значень ціни щодо середнього.

Для перевірки значимості регресії використовують F-критерій. У даному випадку цей критерій використовується для перевірки гіпотези, яка стверджує, що між залежною змінною і незалежною змінною немає лінійної залежності, тобто $b_1 = 0$ проти альтернативи $b_1 \neq 0$. Велике значення F-критерію: $F = 146,82$ і рівень значимості $p = 0,0000$ показує, що побудована регресія високо значима.

При натисканні на кнопку Regression summary – Короткі результати регресії на екрані з'являється електронна таблиця з результатами аналізу (рис. 17)..

Итоги регрессии для зависимой переменной: Ціна (Таблица 4)						
R= ,96147858 R2= ,92444105 Скорректир. R2= ,91814447						
F(1,12)=146,82 p<,000000 Станд. ошибка оценки: 3,7966						
N=14	БЕТА	Стд.Ош. БЕТА	В	Стд.Ош. В	t(12)	p-уров.
Св.член			1233,809	100,6240	12,2616	0,000000
Рік	-0,961479	0,079351	-0,630	0,0520	-12,1168	0,000000

Рис. 17. Короткі результати регресії

У третьому стовпці цієї таблиці розташовані оцінки невідомих параметрів моделі:

$$b_0 = 1233,809,$$

$$b_1 = -0,630,$$

тобто модель має вигляд

$$\text{ЦІНА} = -0,63 * \text{ВІК} + 1233,809.$$

Параметри моделі визначають вид зв'язку. В даному випадку зв'язок зворотний, тобто зі зростанням року виготовлення падає ціна на вино. Такий висновок підтверджується параметром $b_1 = -0,630$, а саме знаком "-" перед значенням.

Лабораторна робота № 10 "Побудова багатофакторної регресійної моделі"

Мета – закріплення теоретичного та практичного матеріалу за темою множинна регресія, придбання навичок побудови й аналізу багатофакторних моделей в модулі Statistics – Multiple Regression.

Завдання – необхідно перевірити наявність лінійного множинного зв'язку між відповідними показниками. Побудувати лінійну багатофакторну модель, зробити висновки щодо адекватності побудованої моделі, дати економічну інтерпретацію моделі.

Методичні рекомендації

У разі множинного регресійного аналізу необхідно оцінити коефіцієнти рівняння

$$y = b_1x_1 + b_2x_2 + \dots + b_nx_n + a,$$

де n – кількість незалежних змінних, позначених як x_1 та x_n ,

y – залежна змінна,

a – константа.

Незалежні змінні, можуть корелювати між собою, тобто між ними може існувати мультиколінійність. Цей факт необхідно обов'язково враховувати при визначенні коефіцієнтів рівняння регресії для того, щоб уникнути помилкових кореляцій.

Також, як і у випадку лінійної регресії, насамперед створюється електронна таблиця з вихідними даними, які у разі необхідності може бути перетворена до бажаного виду.

Розглянемо модель залежності роздрібного товарообігу від обсягу виробництва товарів народного споживання, об'єму імпорту товарів, середньомісячної заробітної плати. Вихідні дані для аналізу наведені у табл. 5.

Вихідні дані по регіонах України

Регіон	Виробництво товарів народного споживання, (млн грн)	Обсяги імпорту товарів, (млн дол США)	Середньомісячна заробітна плата (грн)	Роздрібний товарообіг на душу населення (грн)
Автономна Республіка Крим	384	148,4	143	322
Вінницька	657,7	117	115	301
Волинська	236,6	107,9	105	273
Дніпропетровська	893,3	1047,9	189	411
Донецька	1145,8	891,9	195	378
Житомирська	420,6	93,5	118	328
Закарпатська	194,7	150,4	108	306
Запорізька	639,7	611,9	183	468
Івано-Франківська	234,4	169,8	120	287
Київська	736,1	398,9	161	377
Кіровоградська	218,6	37,3	119	232
Луганська	468,9	218,7	163	298
Львівська	788,3	345,2	132	385
Миколаївська	243,5	201,3	145	304
Одеська	776,3	357,1	146	279
Полтавська	537,5	223,7	150	362
Рівненська	166,1	113,5	120	253
Сумська	569,9	158,4	130	359
Тернопільська	347,2	47,8	104	206
Харківська	1116,6	793,2	159	419
Херсонська	255,6	54,1	125	302
Хмельницька	568,9	79,8	114	258
Черкаська	651	189,3	127	312
Чернівецька	195,9	45,4	106	315
Чернігівська	582,9	131,3	122	387

Необхідно знайти прогнозне значення роздрібного товарообігу на душу населення в регіоні, якщо виробництво товарів народного споживання буде складати 380 млн грн, обсяги об'єму імпорту товарів будуть складати 60 млн дол, середньомісячна заробітна плата буде складати 140 грн.

Перш за все необхідно перевірити незалежні зміни на мультиколінійність. Для цього в меню Statistics – Basic statistics and

Tables вибираємо меню Correlation Matrices (Парні та часткові кореляції) (рис. 18).

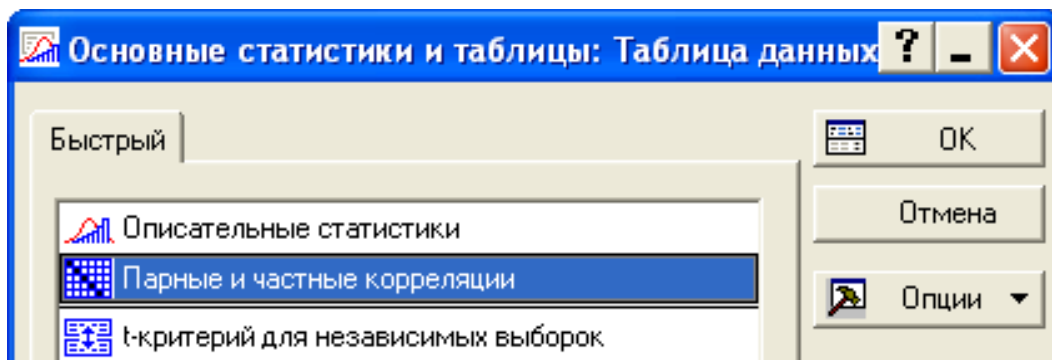


Рис. 18. Модуль "Парні та часткові кореляції"

У даному модулі необхідно, натиснувши на кнопці "Квадратна матриця" вибрати всі факторні змінні. Потім натиснути ОК (рис. 19).

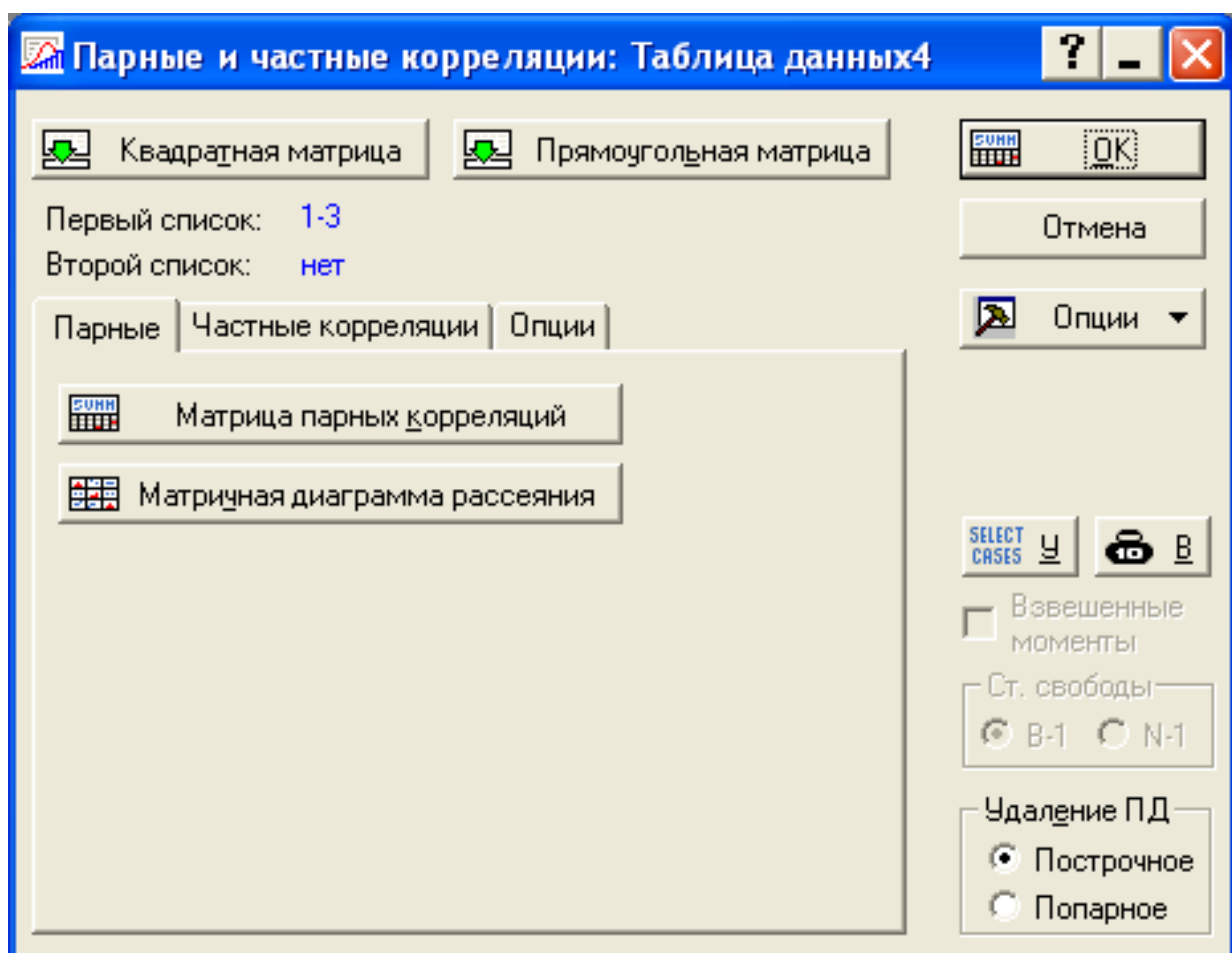


Рис. 19. Вибір параметрів парної кореляції

Результат формується у вигляді квадратної матриці (рис. 20).

Корреляции (Таблица данных4)			
Отмеченные корреляции значимы на y $N=25$ (Построчное удаление ПД)			
Переменная	Производство товаров	Объемы импорта	Зарплата
Производство товаров	1,00	0,80	0,69
Объемы импорта	0,80	1,00	0,86
Зарплата	0,69	0,86	1,00

Рис. 20. Корреляционная матрица

З корреляционной матрицы видно, что наиболее коррелируют фактор "зарплата" та "объемы импорта" та имеют высокую, но допустимую корреляцию факторы "зарплата" та "производство товаров". С целью избавления от большой мультиколлинеарности избавимся от фактора "объемы импорта".

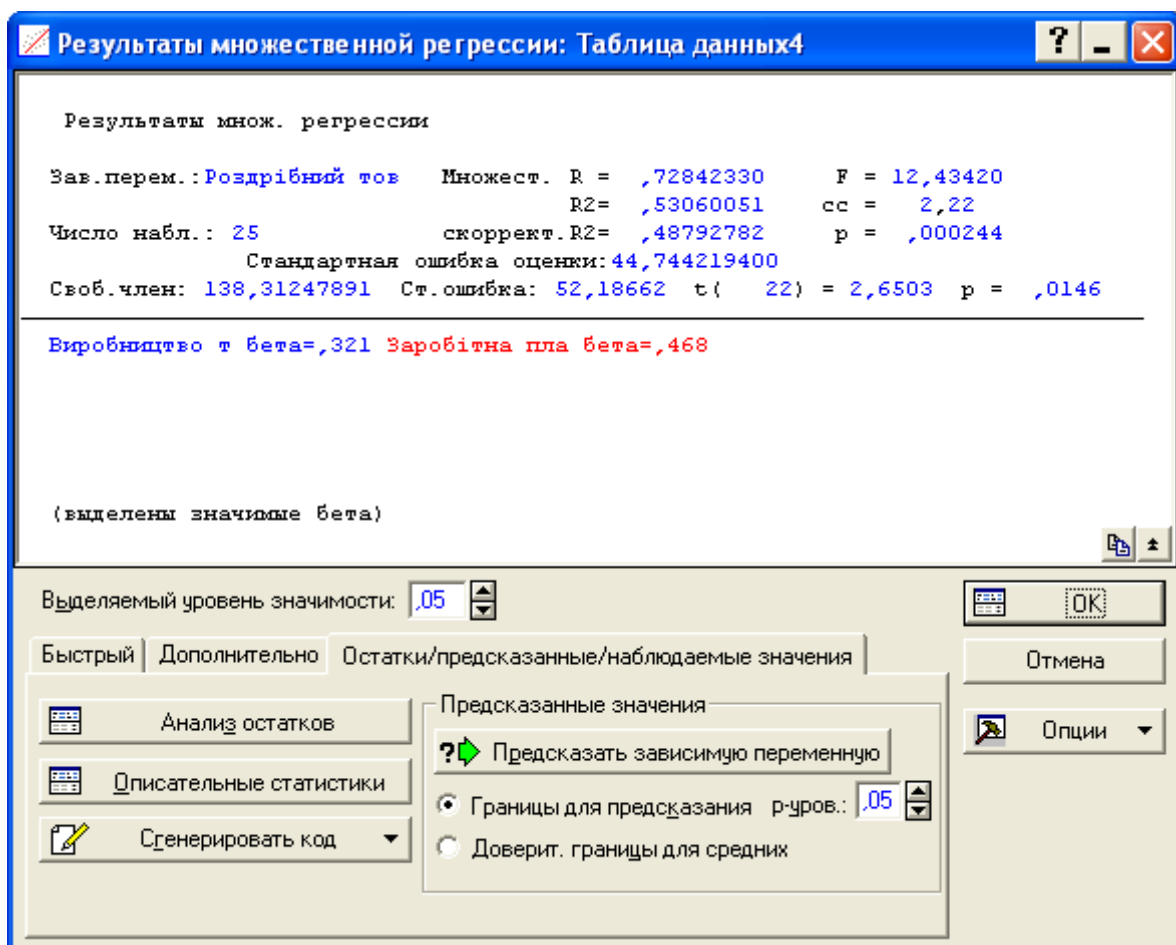


Рис. 21. Окно результатов множественной регрессии

Задля здійснення регресійної оцінки запустимо програму **Statistica** і сформуємо файл даних. На панелі інструментів (або в меню **Statistics**) вибираємо модуль **Multiple Regression**. У стартовому вікні модуля, натиснувши кнопку **Variables**, вибираємо залежну (**Dependent var.**) і незалежну (**Independent var.**) змінні. На закладці **Advanced** можна задати додаткові параметри побудови регресійної моделі (рис. 21). За командою виконання програми з'явиться вікно результатів аналізу (рис. 22).

У функціональній частині вікна містяться кнопки та опції, що дозволяють всебічно розглянути результати регресійного аналізу. Так, на закладці **Quick** є кнопка **Summary: Regression Results – Результати регресії**, яка виводить таблицю результатів побудови регресії (рис. 23). У цій таблиці наведені такі результати побудови регресії: Ві-коефіцієнти (**Beta**) і коефіцієнти регресії B_i з стандартними похибками, значення t-критерію та фактичні рівні істотності **p-level**.

		Итоги регрессии для зависимой переменной: Роздрийний тс R= ,72842330 R2= ,53060051 Скорректир. R2= ,48792782 F(2,22)=12,434 p<,00024 Станд. ошибка оценки: 44,744					
N=25		БЕТА	Стд. Ош. БЕТА	В	Стд. Ош. В	t(22)	p-уров.
Св.член				138,3125	52,18662	2,650344	0,014615
Виробництво товарів		0,321289	0,202471	0,0709	0,04468	1,586842	0,126818
Заробітна плата		0,468074	0,202471	1,1004	0,47600	2,311812	0,030536

Рис. 24. Результати регресії

Виходячи з таблиці, можемо зробити висновок, що значущим є лише фактор середньомісячної заробітної плати.

Наступним кроком є аналіз адекватності побудованої моделі. Про адекватність моделі можна судити за значеннями коефіцієнтів множинної кореляції та детермінації, за значеннями критеріїв Стьюдента та Фішера. Окрім того, слід провести аналіз залишків моделі. Для цього призначена кнопка **Perform residual analysis**, яка знаходиться на закладці **Residuals/assumptions/prediction** у вікні **Результатів множинної регресії**. Натиснувши цю кнопку, переходимо у вікно **Аналіз залишків** (рис. 25).

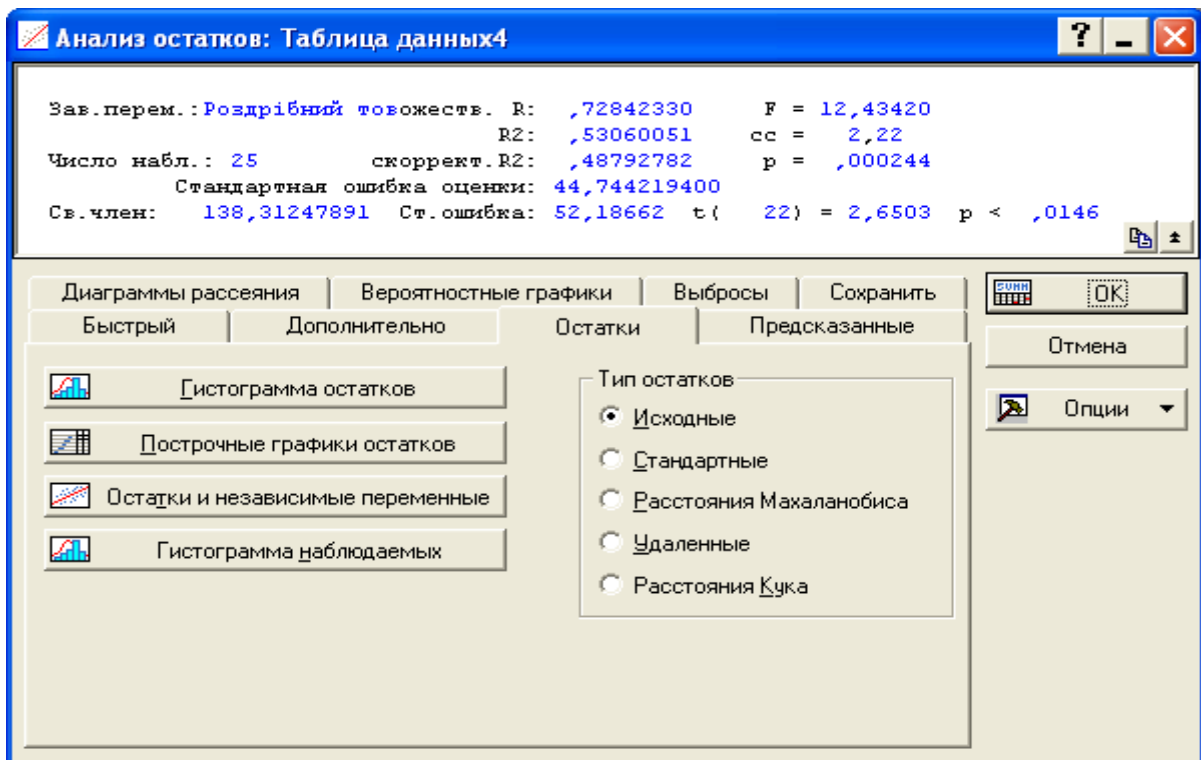


Рис. 25. Вікно аналізу залишків

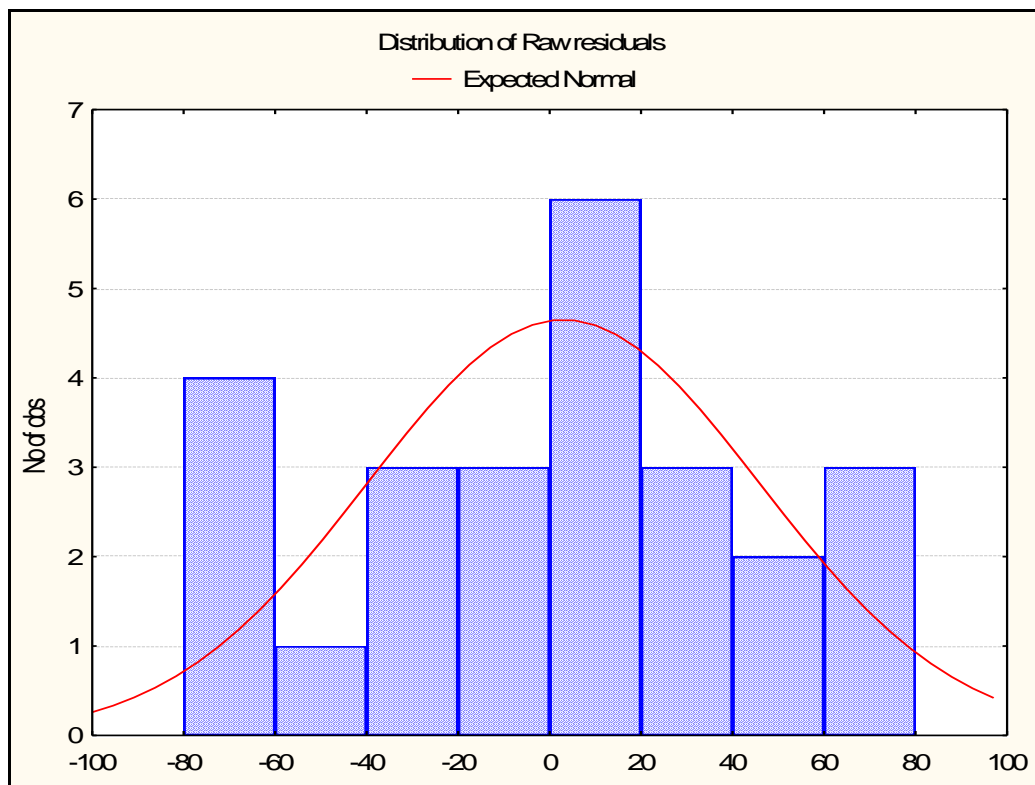


Рис. 26. Гістограма розподілу залишків

Наглядними і найважливішими є гістограма розподілу залишків (зкладка **Residuals**, кнопка **Histogram of residuals**) і графік залишків на

нормальному ймовірнісному папері (закладна **Probability plots**, кнопка **Normal plot of residuals**). Відповідні графіки наведено на рис. 26 та рис. 27.

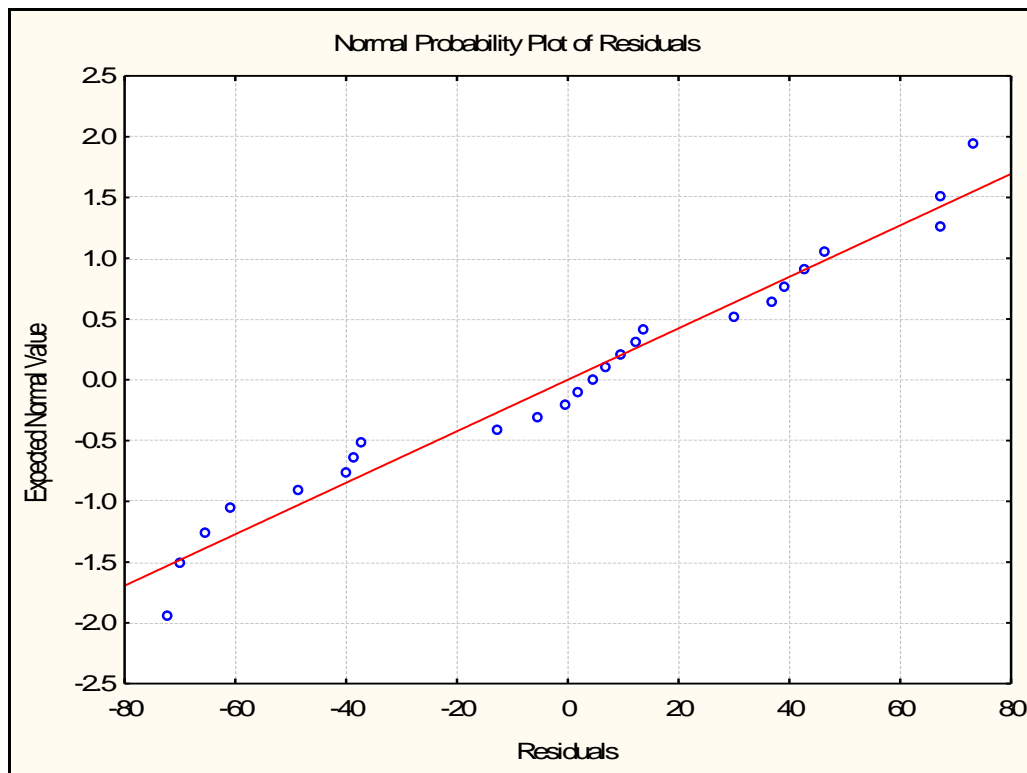


Рис. 27. Графік залишків на нормальному ймовірнісному папері

Ми бачимо, що залишки розподілені за нормальним законом розподілу (гістограма залишків) і добре лягають на пряму (графік залишків на нормальному ймовірнісному папері), тобто це свідчить про адекватність побудованої моделі.

У модулі Множинної регресії можна знайти прогнозне значення залежної змінної. Для цього у вікні результатів необхідно перейти на закладку **Residuals/assumptions/prediction** і натиснути кнопку **Predict dependent variable** – **Прогнозне значення залежної змінної**. У вікні, що з'явилося, потрібно задати значення незалежних змінних, при яких слід знайти прогнозне значення залежної змінної. Наприклад, задамо такі значення (рис. 28).

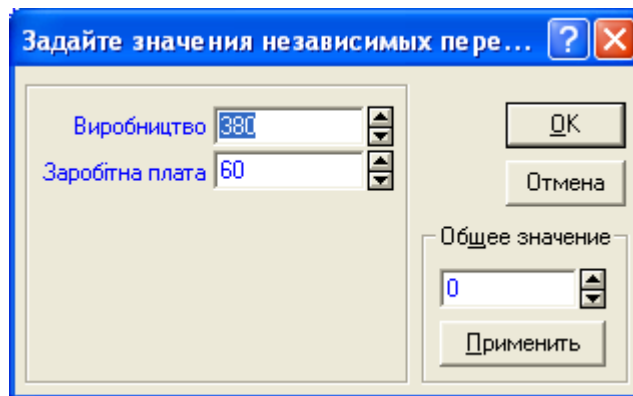


Рис. 28. Значення незалежних змінних для обчислення прогнозного значення залежної змінної

Після виконання команди, отримаємо наступну таблицю результатів (рис. 29).

Переменная	Предск.значения для (Таблица д перемен.: Роздрібний товарообіг		
	В-Веса	Значение	В-Веса * знач.
Производство товаров	0,070900	380,0000	26,9419
Зарплата плата	1,100428	60,0000	66,0257
Св.член			138,3125
Предсказ.			231,2801
-95,0%ДП			162,1453
+95,0%ДП			300,4149

Рис. 29. Таблиця результатів прогнозування

Висновки. Отже, якщо виробництво товарів народного споживання, (млн грн) буде складати 380, а середньомісячна заробітна плата буде складати 140 грн, то значення роздрібного товарообігу на душу населення (грн) складатиме 231,28, та може коливатися в межах від 162,14 грн до 300,41 грн.

Лабораторна робота № 11 "Аналіз рядів динаміки в Excel"

Мета – закріплення теоретичного та практичного матеріалу за темою аналіз рядів динаміки, придбання навичок визначення середніх показників динаміки в **Excel**.

Завдання – необхідно провести аналіз рядів динаміки за допомогою середніх показників, надати економічну інтерпретацію розрахованим показникам.

Методичні рекомендації

Ряди динаміки – статистичні дані, що відображають розвиток у часі досліджуваного явища. Їх також називають динамічними рядами, часовими рядами. В кожному ряді динаміки є два основні елементи:

- 1) показник часу t ;
- 2) відповідні їм рівні розвитку досліджуваного явища y .

В якості показників часу в рядах динаміки виступають або певні дати (моменти), або окремі періоди (роки, квартали, місяці, доба). Рівні рядів динаміки відображають кількісну оцінку (міру) розвитку в часі досліджуваного явища. Вони можуть виражатися абсолютними, відносними або середніми величинами.

Середній рівень ряду розраховується за такими формулами (табл. 6).

Таблиця 6

Формули розрахунку середнього рівня ряду

	При рівних проміжках часу	При нерівних проміжках часу
Для інтервальних рядів	середня арифметична проста: $\bar{y} = \frac{\sum y}{n}$	середня арифметична зважена: $\bar{y} = \frac{\sum yt}{\sum t}$
Для моментних рядів	середня гармонійна: $\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}$	$\bar{y} = \frac{\frac{y_{t_{i-1}} + y_{t_i}}{2} \times t_{i-1;i} + \frac{y_{t_i} + y_{t_{i+1}}}{2} \times t_{i;i+1} + \dots + \frac{y_{t_{n-1}} + y_{t_n}}{2} \times t_{n-1;n}}{t_{i-n}}$

Для розрахунку основних характеристик часових рядів застосовуються такі формули (табл. 7).

Основні характеристик часових рядів

Показник	Формула	
	Базисний спосіб	Ланцюговий спосіб
Абсолютний приріст (Δy)	$y_i - y_0$	$y_i - y_{i-1}$
Темп зростання (T_p)	$\frac{y_i}{y_0} \times 100$	$\frac{y_i}{y_{i-1}} \times 100$
Темп приросту (T_{np})	$T_p - 1(100)$	
Абсолютне значення одного проценту приросту (A1%)	$\frac{Y_{i-1}}{100}$	
Середній абсолютний приріст	$\frac{y_i - y_0}{n}$	
Середній темп зростання	$\sqrt[n-1]{\frac{y_i}{y_0}}$	$\sqrt[n]{T_1 \cdot T_2 \cdot \dots \cdot T_n}$ $\sum^t \sqrt{(T_{PI})^{t1} * (T_{PII})^{t2} \dots (T_{PIN})^{tn}}$
Середній темп приросту	$\bar{T}_{np} = \bar{T}_p - 1(100)$	

Створимо файл "Аналіз рядів динаміки". На Листі 1 стовпець А та В будуть стовпцями вихідних даних ("Роки" та "Обсяг інвестицій в галузь, тис грн" відповідно) (рис. 30).

	А	В
1	Годы	Объем производства, тис. грн.
2		
3	2000	125,00
4	2005	135,00
5	2006	140,00
6	2007	144,00
7	2008	149,00
8	2009	145,00

Рис. 30. Вихідні дані в Excel

Необхідно розрахувати всі показники динамічного ряду (база порівняння – 2000 р.), середньорічні темпи зростання й приросту за періоди:

1. 2000 – 2005 рр.
2. 2005 – 2009 рр.
3. 2000 – 2009 рр.

1. Для визначення абсолютних приростів:

у клітинці С4 вводимо формулу =B4-\$B\$3. Розтягуємо цю формулу на весь стовпець;

у клітинці D5 (для 2005 р. абсолютний приріст неможливо розрахувати ланцюговим способом, тому що немає попереднього періоду – 2004 р.) вводимо формулу =B5-B4 і також розтягуємо цю формулу на весь стовпець.

Результати розрахунку представлені на рис. 31.

	А	В	С	Д
1	Годы	Объем производства, тис. грн.	Δ у, тис. грн. (-)	
2			базисний	ланцюговий
3	2000	125.00	-	-
4	2005	135.00	10.00	-
5	2006	140.00	15.00	5.00
6	2007	144.00	19.00	4.00
7	2008	149.00	24.00	5.00
8	2009	145.00	20.00	-4.00

Рис. 31. Результати розрахунку абсолютних приростів

2. Визначення темпів зростання:

у клітинці Е4 вводимо формулу =B4/\$B\$3*100. Розтягуємо цю формулу на весь стовпець;

у клітинці F5 (для 2005 р. темп зростання, як і абсолютний приріст, неможливо розрахувати ланцюговим способом, тому що немає попереднього періоду – 2004 р.) вводимо формулу =B5/B4*100 і також розтягуємо цю формулу на весь стовпець.

Визначення темпів приросту:

у клітинці G4 вводимо формулу =E4-100. Розтягуємо цю формулу на весь стовпець;

у клітинці Н5 вводимо формулу =F5–100 і також розтягуємо цю формулу на весь стовпець.

Визначення абсолютного значення 1 % приросту:

у клітинці І5 вводимо формулу =B4/100. Розтягуємо цю формулу на весь стовпець.

Результати розрахунків темпів росту та приросту представлені на рис. 32.

	А	В	Е		Г		Н	
1	Годы	Объем производства, тис. грн.	Тр. %		Тпр. %			
2			базисний	ланцюговий	базисний	ланцюговий		
3	2000	125.00	-	-	-	-		
4	2005	135.00	108.00	-	8.00	-		
5	2006	140.00	112.00	103.70	12.00	3.70		
6	2007	144.00	115.20	102.86	15.20	2.86		
7	2008	149.00	119.20	103.47	19.20	3.47		
8	2009	145.00	116.00	97.32	16.00	-2.68		

Рис. 32 . Результат розрахунку темпів зростання та приросту

3. Визначення середніх показників динаміки.

3.1. Визначення середнього рівня ряду. Через те, що ми маємо пропущені рівні ряду, то середній рівень ряду доцільно розраховувати за період 2005 – 2009 рр. – у клітинці В10 вибираємо вбудовану статистичну функцію СРЗНАЧ.

3.2. Середній абсолютний приріст – у клітинці В11 вводимо формулу =(B8–B3)/8.

3.3. Середній темп зростання.

1. 2000 – 2005 рр. – у клітинці В14 вводимо формулу =(B4/B3)^(1/3).

2. 2005 – 2009 рр.

базисним способом – у клітинці В16 вводимо формулу =(B8/B4)^(1/4);

ланцюговим способом – у клітинці В17 вводимо формулу =((F5/100)*(F6/100)*(F7/100)*(F8/100))^(1/4);

3. 2000 – 2009 рр.

базисним способом – у клітинці В19 вводимо формулу =(B8/B3)^(1/7);

ланцюговим способом за формулою середньої геометричної зваженої: у клітинці В20 вводимо формулу =((B14)^3*(B16)^4)^(1/7).

3.4. Середній темп приросту

1. 2002 – 2005 рр.: у клітинці С14 вводимо формулу =(B14–1)*100;

2. 2005 – 2009 рр.: у клітинці С16 вводимо формулу $=(B16-1)*100$;
 3. 2002 – 2009 рр.: у клітинці С20 вводимо формулу $=(B20-1)*100$.
- Результати розрахунків середніх величин представлені на рис. 33.

	А	В
10	середній рівень ряду	139,67
11	середній абсолютний приріст	2
12	середній темп росту	
13	1. 2000-2005	
14	базисний спосіб	1,015511278
15	2. 2005-2009	
16	базисний спосіб	1,01802527
17	ланцюговий спосіб	1,01802527
18	3. 2000-2009	
19	базисний спосіб	1,016627841
20	ланцюговий спосіб за формулою середньої геометричної зваженої	1,016627841

Рис. 33. Результати визначення середніх показників динаміки

Таким чином, протягом 2000 – 2009 рр. спостерігалось постійне зростання об'єму виробництва, лише у 2009 р. об'єм виробництва знизився порівняно з 2008 р. на 4 тис грн (порівняно з 2000 р. збільшився на 20 тис грн). Темп зростання в 2009 році, розрахований базисним способом, становить 116,00 %, що означає збільшення обсягу виробництва на 2,68 %. Ланцюговий темп приросту показав зменшення обсягу виробництва порівняно з 2008 р. на 2,68 %. В 1 % приросту в 2009 році містилося 1,49 тис грн. У період 2000 – 2009 рр. об'єм виробництва щорічно збільшувався в середньому на 2 тис грн. Середній об'єм виробництва за період 2000 – 2009 рр. знаходився на рівні 139,67 тис грн. У період з 2000 по 2005 рр. об'єм виробництва щорічно в середньому збільшується на 1,55 %.

Лабораторна робота №12 "Екстраполяція та інтерполяція в Excel"

Мета – закріплення теоретичного та практичного матеріалу за темою аналіз тенденцій розвитку, придбання навичок використання екстраполяції та інтерполяції в рядах динаміки.

Завдання – необхідно знайти рівні динамічного ряду за допомогою екстраполяції та інтерполяції. Побудувати прогноз та знайти, в яких межах буде знаходитись прогнозне значення, зробити висновки щодо прогнозу, дати економічну інтерпретацію.

Методичні рекомендації

Екстраполяція й інтерполяція відносяться до статистичних методів прогнозування. Прогнозування – це визначення орієнтованих розмірів явища в майбутньому, поширення виявленої закономірності на інші періоди часу. При прогнозуванні явища вирішуються такі проблеми:

- 1) визначення довжини базисного періоду, закономірність якого поширюють на майбутнє;
- 2) визначення довжини майбутнього періоду, на який поширюють виявлену закономірність.

Екстраполяція – це метод прогнозування, який передбачає, що закономірність розвитку, що діяла в минулому, збережеться і в прогнозованому майбутньому.

Існують такі методи екстраполяції даних:

- 1) якщо для вимірювання основної тенденції проводилося аналітичне вирівнювання ряду динаміки, то для екстраполяції використовується рівняння тренда, в яке підставляється значення t у майбутні періоди. Оскільки в базисному періоді рівень коливається близько тренда, то коливання ймовірні і в майбутньому. Отже, фактичний рівень в майбутньому не обов'язково точно збігається з екстрапольованими по тренду;

- 2) екстраполяція на майбутнє за допомогою середньої абсолютної або відносної швидкості зміни рівня. Цей метод заснований на припущенні про рівномірний зміну рівня:

- а) якщо в базисному періоді ланцюгові показники динаміки не мали різкі коливання;

- б) якщо ланцюгові показники динаміки в базисному періоді сильно коливалися і щорічний приріст (або темпи зростання і приросту) обчислювалися на базі середніх рівнів, розрахованих за рівні періоди.

Інтерполяція – це визначення невідомого рівня всередині динамічного ряду.

При інтерполяції припускають, що виявлена тенденція та її характер, істотно не змінювалися в тому проміжку часу, рівень якого невідомий. Існують такі методи інтерполяції даних:

1) на основі середнього абсолютного приросту, середнього темпу зростання;

2) якщо, окрім кінцевого і базисного рівнів, відомі також деякі проміжні рівні, інтерполяції можна здійснити на основі рівняння тренда.

Лінії тренда можна проводити на гістограмах, графіках, лінійчатих і ХУ-точкових діаграмах. Не можна будувати лінії тренда для рядів даних на об'ємних, кругових і кільцевих діаграмах. Якщо змінити тип подання групи рядів на один з перерахованих, то відповідні цим рядом даних лінії тренда будуть втрачені.

В Excel можна вибрати один з п'яти типів екстраполяції за допомогою діалогового вікна, що зображена на рис. 34:

- лінійний,
- поліноміальний,
- логарифмічний,
- експонентний,
- ступеневий.

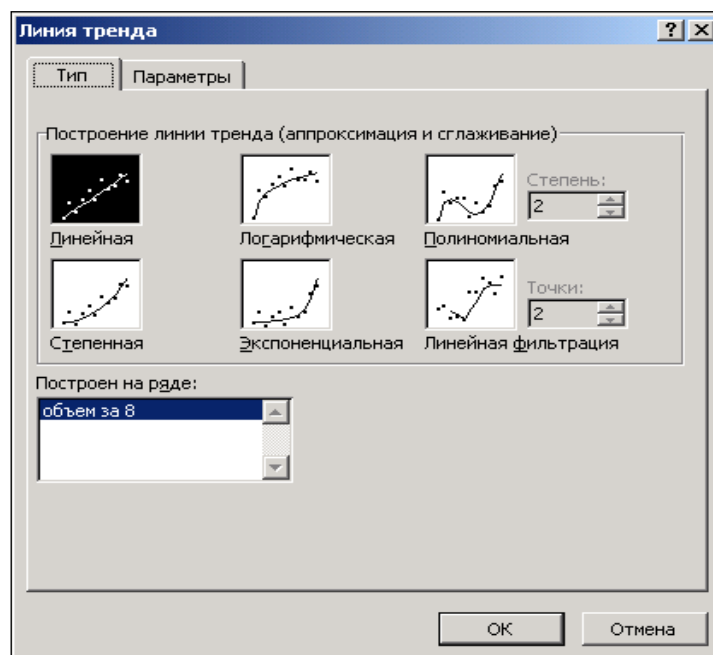


Рис. 34. Діалогове вікно "Лінія тренда"

Тип вибраної екстраполяції задає спосіб обчислень лінії тренда. Залежно від використовуваних даних, деякі типи екстраполяції можуть виявитися надійніше інших з точки зору оцінки результатів зроблених прогнозів. Найчастіше використовується лінійний тренд.

Маємо дані про валовий внутрішній продукт (ВВП) за 1998 – 2009 рр. (рис. 35). Необхідно визначити за допомогою аналітичного вирівнювання прогнозне значення обсягу прибутку у 2010 р.

	А	В
1	Роки	ВВП, млрд. грн.
2		
3	1998	190
4	1999	210
5	2000	200
6	2001	220
7	2002	240
8	2003	230
9	2004	220
10	2005	240
11	2006	260
12	2007	260
13	2008	270
14	2009	280
15	сума	2820

Рис. 35. Вхідні дані



Рис. 36. Меню для побудови тренда

Для побудови лінії тренда необхідно виділити часовий ряд та вибрати в контекстному меню команду "Добавити лінію тренда" (рис. 35). Після вибору команди відкриється діалогове вікно (рис. 36), в якому вибираємо тип, параметри побудови тренда.

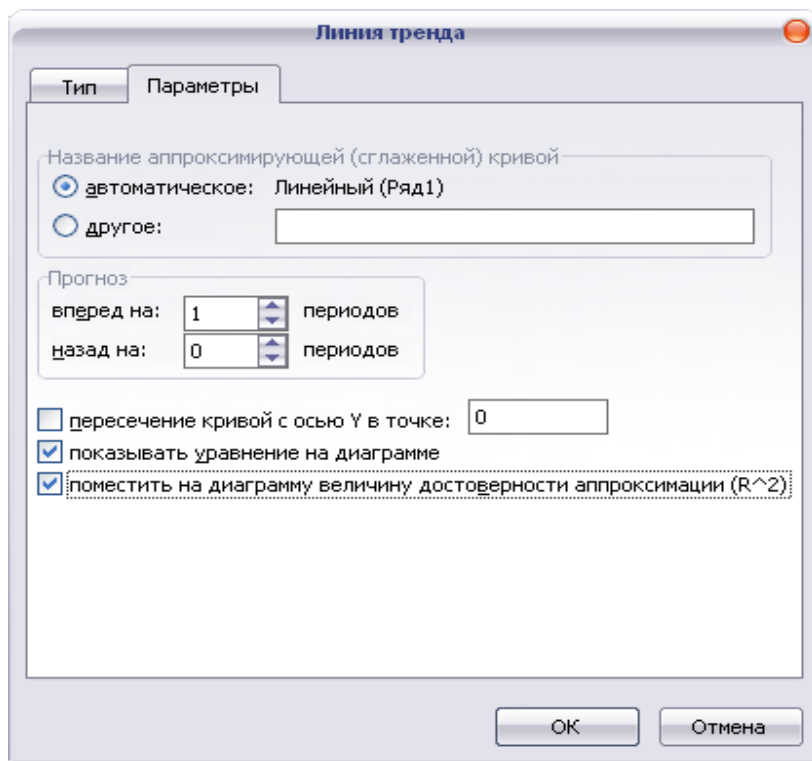


Рис. 36. Діалогове вікно: лінія тренда

Результат побудови лінійного тренда представлений на рис. 37.

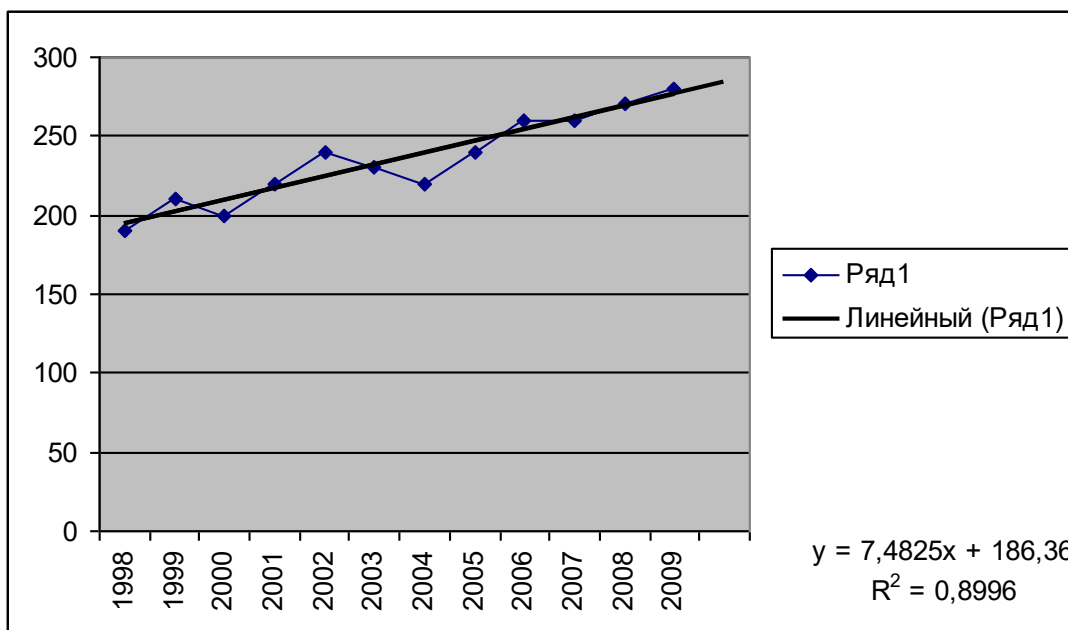


Рис. 37. Результати розрахунку параметрів тренда

Таким чином, рівняння прямої має вигляд:

$$Y = 186,36 + 7,4825 \times t.$$

Розрахувати обсяг ВВП у 2010 році можна:

1. Точковий прогноз. Використовуючи рівняння прямої, де t – це умовний показник часу, за яким буде будуватись прогноз. Тобто обсяг ВВП в 2010 році очікується на рівні:

$$Y = 186,36 + 7,4825 \times t, \quad t = 13$$

$$Y = 283,63 \text{ млрд грн.}$$

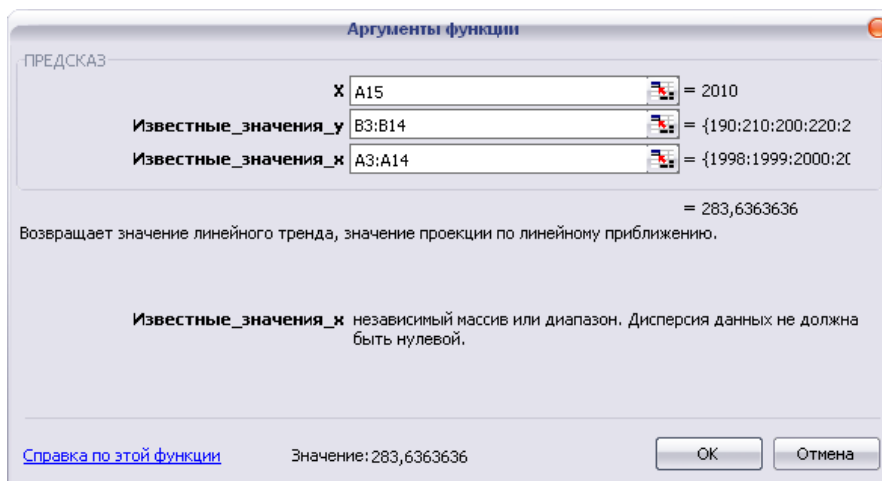


Рис. 38. **Діалогове вікно: аргументи функції ПРЕДСКАЗ**

2. Використовуючи стандартну функцію ПРЕДСКАЗ. Після вибору даної функції відкривається діалогове вікно (рис. 38), у якому задаємо параметри.

Результат функції ПРЕДСКАЗ представлений на рис. 39.

	A	B	C	D	E
1	Роки	ВВП, млрд. грн.			
2					
3	1998	190			
4	1999	210			
5	2000	200			
6	2001	220			
7	2002	240			
8	2003	230			
9	2004	220			
10	2005	240			
11	2006	260			
12	2007	260			
13	2008	270			
14	2009	280			
15	2010	283,6363636			

Рис. 39. **Результат функції ПРЕДСКАЗ**

3. Використовуючи стандартну функцію ТЕНДЕНЦИЯ. Після вибору даної функції відкривається діалогове вікно (рис. 40), в якому задаємо параметри.

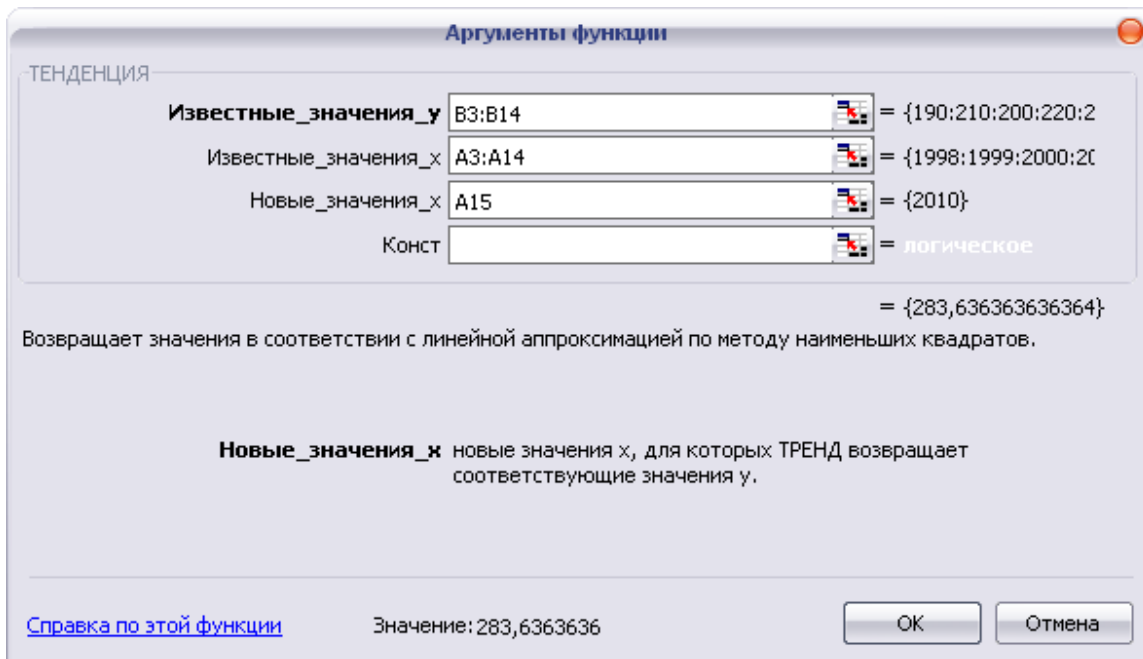


Рис. 40. Діалогове вікно: аргумент функції ТЕНДЕНЦИЯ

Результат функції ТЕНДЕНЦИЯ представлен на рис. 41.

	B15	fx =ТЕНДЕНЦИЯ(B3:B14;A3:A14;A15)			
	A	B	C	D	E
1		ВВП, млрд.			
2	Роки	грн.			
3	1998	190			
4	1999	210			
5	2000	200			
6	2001	220			
7	2002	240			
8	2003	230			
9	2004	220			
10	2005	240			
11	2006	260			
12	2007	260			
13	2008	270			
14	2009	280			
15	2010	283,6363636			

Рис. 41. Результат функції ТЕНДЕНЦИЯ

Визначимо межі обсягу ВВП у 2010 р.: $y_t \pm t_2 S_{yt}$

Розрахунок критерія Стюдента за допомогою вбудованої функції СТЬЮДРАСПОБР (рис. 42) при довірчій імовірності 0,95.

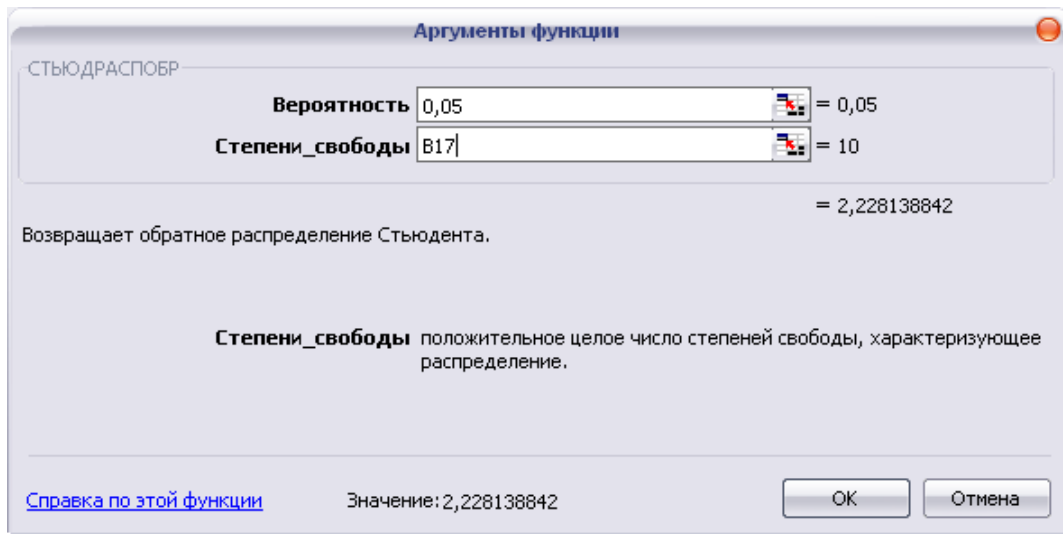


Рис. 42. Діалогове вікно: аргументи функції СТЬЮДРАСПОБР

Результат функції: $t_2 = 2,228$

Розрахуємо S_{yt} (рис. 43):

у клітинці G3 вводимо формулу =F3–B3. Розтягуємо цю формулу на весь стовпець;

у клітинці H3 вводимо формулу =G3^2. Розтягуємо цю формулу на весь стовпець. У ячейці H15 вводимо формулу =СУММ(H3:H14);

у клітинці B18 вводимо формулу =КОРЕНЬ(H15/10).

B20		=КОРЕНЬ(H15/10)							
	A	B	C	D			E		
1	Роки	ВВП, млрд. грн.	Умовний показник часу, t	t^2	y^*t	y_t	$y_i - y_t$	$(y_i - y_t)^2$	$(y_i - \bar{y})^2$
2									
3	1998	190	-5,5	30,25	-1045	193,85	3,85	14,79	2025,00
4	1999	210	-4,5	20,25	-945	201,33	-8,67	75,19	625,00
5	2000	200	-3,5	12,25	-700	208,81	8,81	77,64	1225,00
6	2001	220	-2,5	6,25	-550	216,29	-3,71	13,74	225,00
7	2002	240	-1,5	2,25	-360	223,78	-16,22	263,21	25,00
8	2003	230	-0,5	0,25	-115	231,26	1,26	1,58	25,00
9	2004	220	0,5	0,25	110	238,74	18,74	351,23	225,00
10	2005	240	1,5	2,25	360	246,22	6,22	38,74	25,00
11	2006	260	2,5	6,25	650	253,71	-6,29	39,61	625,00
12	2007	260	3,5	12,25	910	261,19	1,19	1,41	625,00
13	2008	270	4,5	20,25	1215	268,67	-1,33	1,77	1225,00
14	2009	280	5,5	30,25	1540	276,15	-3,85	14,79	2025,00
15	сума	2820		143	1070			893,706294	8900
16									
20	S	9,45							

Рис. 43. Результати розрахунку F-критерія

Кількість спостережень дорівнює 12, тобто $n = 12$, кількість параметрів моделі – 2, тобто $m = 2$, звідси число ступенів свободи $12 - 2 = 10$.

Розрахунок меж прогнозного значення обсягу ВВП у 2010 р. буде знаходитися у межах:

$$283,64 - 2,23 \times 9,45 \leq y_{\text{пр}} \leq 283,64 + 2,23 \times 9,45$$
$$262,55 \leq y_{\text{пр}} \leq 304,72 \text{ млрд грн.}$$

Прогнозне значення обсягу ВВП у 2010 р. буде знаходитись у межах від 262,55 до 304,72 млрд грн. Це можна стверджувати з імовірністю 95 %.

Лабораторна робота № 13 "Пошук характерних закономірностей в динамічних рядах"

Мета – закріплення теоретичного та практичного матеріалу за темою "Аналіз коливань та сталості динамічних рядів", придбання навичок побудови й аналізу моделей в Excel.

Завдання – необхідно побудувати різні моделі динаміки за даними часового ряду в Excel. Привести економічну інтерпретацію за отриманими результатами.

Методичні рекомендації

Тенденція – це об'єктивно існуюча властивість того чи іншого процесу, яка лише приблизно описується трендом певного виду. Трендом називають і саму основну тенденцію розвитку, і конкретний її опис за допомогою рівняння регресії.

Для виявлення та аналізу загальної тенденції розвитку досліджуваного явища необхідно абстрагуватися від впливу нетрендових факторів. Досягти цього, певною мірою, дозволяють прийоми згладжування або вирівнювання часового ряду.

Розрізняють механічне та аналітичне вирівнювання. Останнє дозволяє формалізувати тенденцію, представити її у вигляді конкретної математичної функції.

Суть різних прийомів, за допомогою яких здійснюється згладжування, зводиться до заміни фактичних рівнів динамічного ряду розрахунковими, що мають значно менше коливання, ніж вихідні дані. Зменшення коливання рівнів дозволяє тенденції розвитку проявитися більш чітко. У ряді випадків згладжування ряду може розглядатися як важливий допоміжний засіб, що полегшує застосування інших методів і, зокрема, більш суворих методів виділення тенденції.

Наприклад, для того щоб провести згладжування динамічного ряду середньої заробітної плати на одного працюючого з 2002 по 2011 рік за місяцями (всього 118 точок) 3-х і 5-членних ковзких середніх, застосовуємо дані з табл. 7.

Таблиця 7

Середня заробітна плата в розрахунку на одного штатного працівника, грн

Місяць	Рік									
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Січень	320,76	400,59	499,66	640,86	864,91	1112	1521	1665	1916	2297
Лютий	328,7	391,2	510,06	666,76	904,62	1142	1633	1723	1955	2338
Березень	354,81	415,49	545,07	722,01	986,65	1230	1702	1818	2109	2531
Квітень	355,78	422,58	547,9	733,73	984,29	1224	1735	1845	2107	2533
Травень	358,88	439,26	554,96	764,29	1002,88	1277	1774	1851	2201	2573
Червень	377,41	476,16	601,45	823,1	1063,59	1368	1883	1980	2373	2708
Липень	398,1	489,5	607,97	837,45	1078,86	1421	1930	2008	2367	2749
Серпень	390,07	479,16	604,19	831,06	1073,1	1398	1872	1919	2280	2694
Вересень	391,14	498,27	630,82	856,02	1086,58	1426	1916	1964	2349	2737
Жовтень	397,49	498,34	636,21	882,1	1088,07	1475	1917	1950	2322	2729
Листопад	395,7	489,52	644,27	896,58	1103,9	1485	1823	1955	2353	
Грудень	442,91	550,92	703,77	1019,71	1277,06	1675	2001	2233	2629	

Дані необхідно занести в таблицю Statistica послідовно в стовпчик, тобто в даному випадку буде одна змінна і 118 випадків.

Для початку роботи скористаємося меню Statistics / Advanced Linear / Nonlinear Models / Time Series / Forecasting. У вікні вказуємо змінну, з якою будемо працювати, натискаємо на кнопку ОК (transformations, autocorrelations, crosscorrelations, plots) і потім вибираємо закладку Smoothing (рис. 44).

На даній закладці ми вибираємо опцію N-точкової ковзкої середньої N-pts Moving Average. Для отримання 3-х членної ковзкої середньої в поле інтервалу згладжування N ставимо цифру 3 (N = 3), а для 5-тичленної – відповідно, цифру 5.

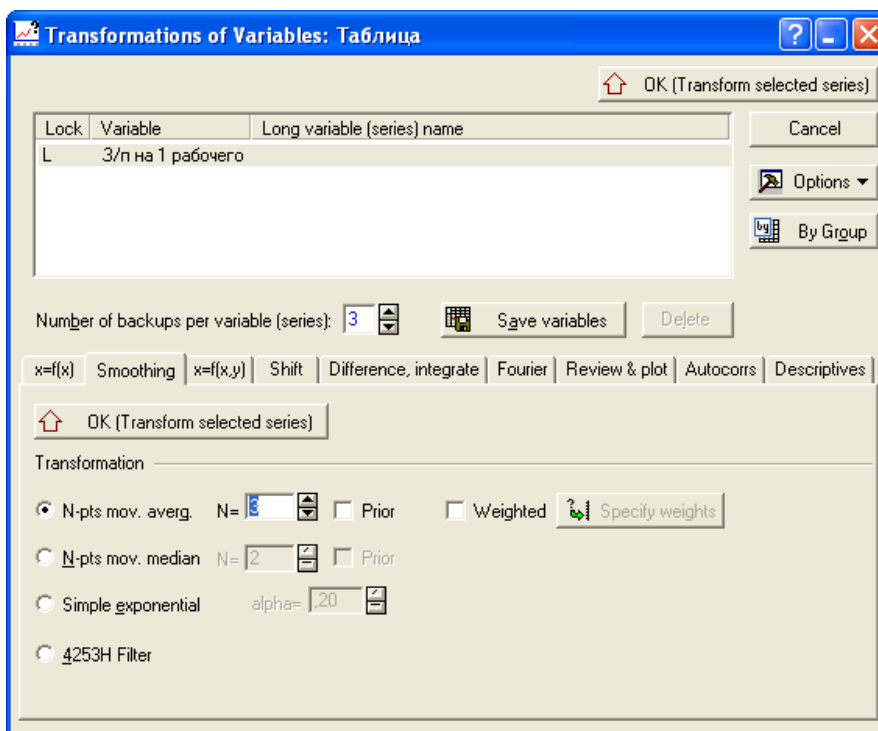


Рис. 44. Вид закладки Smoothing процедури Transformation of Variables

Далі натискаємо на кнопку ОК (Transform selected series), система автоматично буде графічне зображення ковзкої середньої з заданими параметрами (рис. 45).

В інформаційне поле додаються відповідні змінні, які необхідно зберегти за допомогою кнопки Save variables (рис. 46).

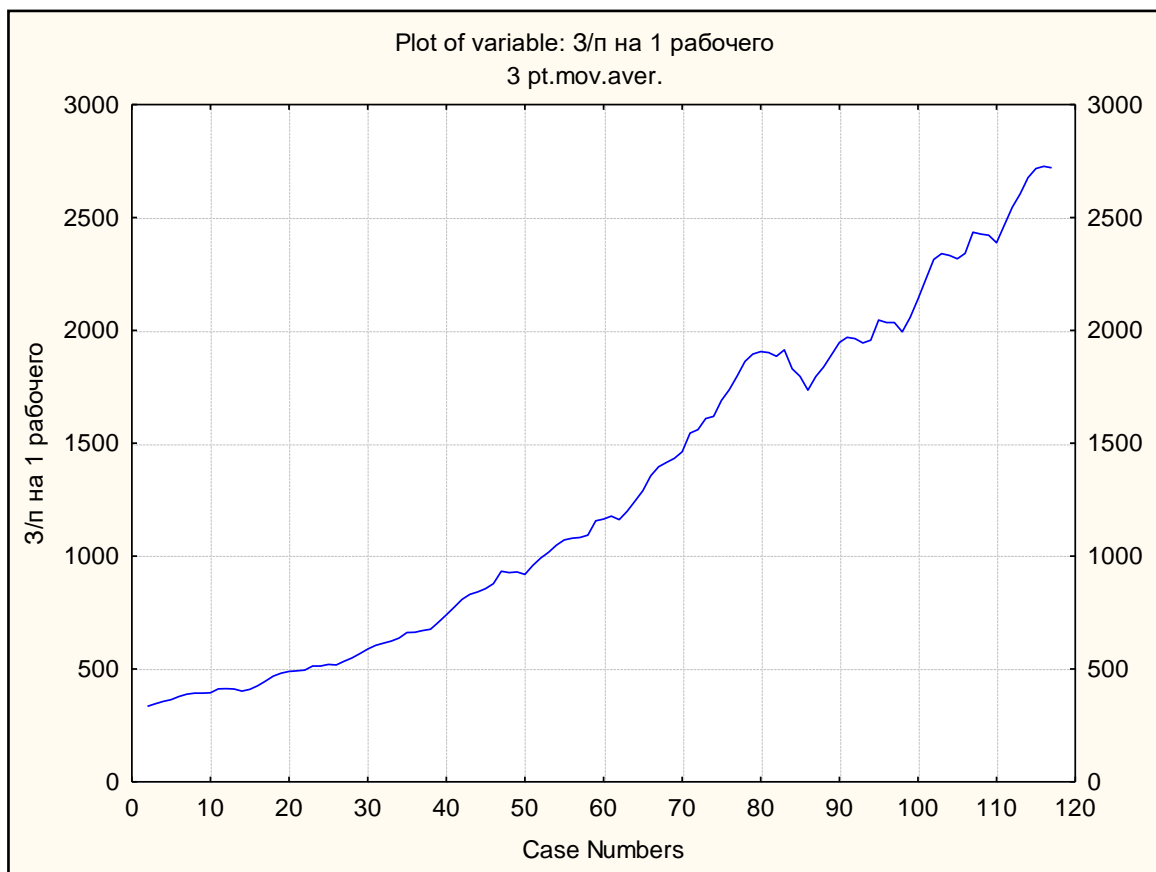


Рис. 45. Графічного зображення ковзної середньої (лаг = 3)

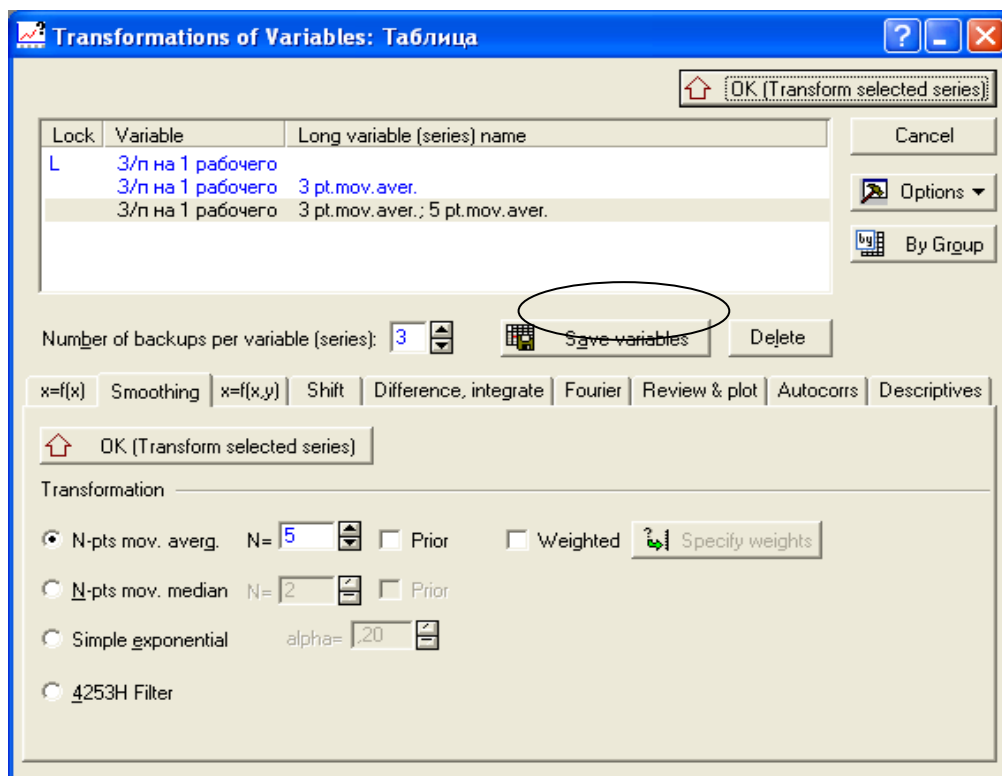


Рис. 46. Інформаційне поле з 3-х і 5-тичленною ковзною середньою

Далі будемо графіки за допомогою вже відомої процедури Graphs/2D Graphs / Line Plots (Variables). На одному графіку доцільно уявити вихідний динамічний ряд і ряд, отриманий у результаті вирівнювання по одній з ковзних середніх (тобто вибрати тип графіка Multiple) (рис. 47, 48).

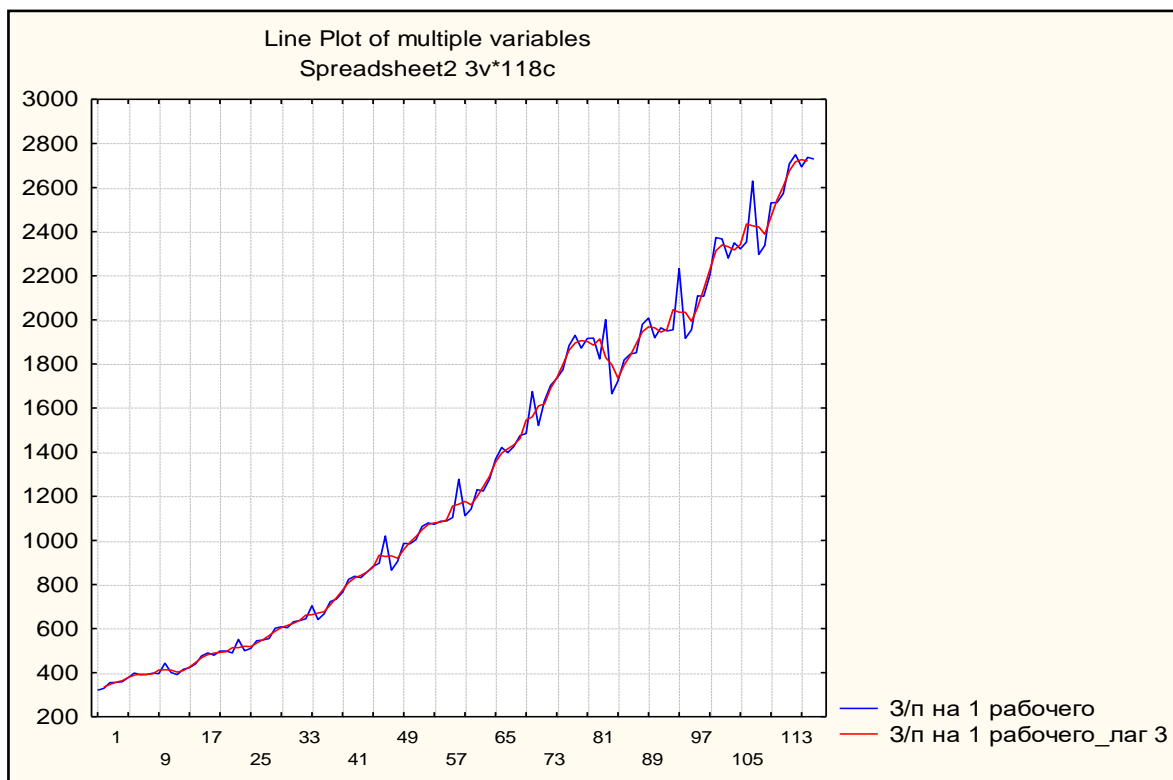


Рис. 47. Динамічний ряд, згладжений 3-х членів ковзаючими середніми

На представлених графіках чітко видно, що вирівнювання з періодом ковзання п'ять років, дає кращі результати.

Прості ковзкі середні – грубий статистичний прийом виявлення тенденції. В деяких випадках після згладжування дрібні хвилі або взагалі зникають, або міняють свій знак, тобто замість опуклої ділянки на кривій отримують увігнуту ділянку, і навпаки.

Більш тонкий прийом, який базується на тій же ідеї, що і прості ковзкі середні, полягає в застосуванні зважених ковзких середніх, експоненціального згладжування. Оскільки процедура експоненціального згладжування частіше використовується при вирішенні завдань прогнозування, розглянемо даний вид згладжування.

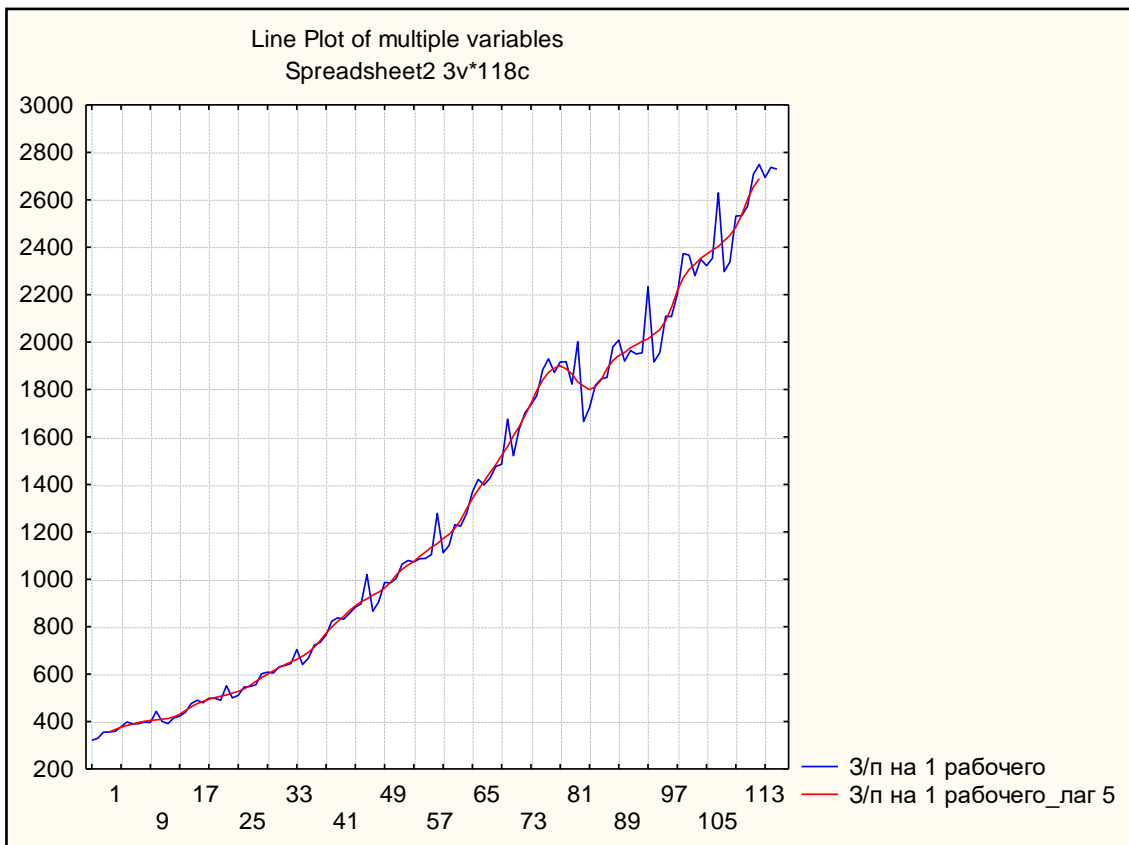


Рис. 48. Динамічний ряд, згладжений 5-ти членів ковзаючими середніми

Для вирішення поставленого завдання з аналітичного згладжування динамічних рядів у системі STATISTICA нам потрібно створити кілька нових додаткових змінних.

Отже, нам необхідно побудувати рівняння тренда, яке є рівнянням регресії, в якому як фактор виступає "час". Перш за все, ми створимо змінну "Т", що містить моменти часу, що відповідають місяцям протягом 2002 – 2011 років. Це період, що включає в себе 118 місяців, і змінна "Т" складатиметься з натуральних чисел від 1 до 118, відповідно до місяців року.

У підсумку виходить такий робочий лист (рис.49)

	1	2
	З/п на 1 рабочего	Т
1	320,76	1
2	328,7	2
3	354,81	3
4	355,78	4
5	358,88	5
6	377,41	6

Рис. 49. Робочий лист із створеними змінними часу

Далі розглянемо процедуру, що дозволяє будувати регресійні моделі як лінійного, так і нелінійного типу. Для цього вибираємо: Statistics / Advanced Linear / Nonlinear Models / Nonlinear Estimation.

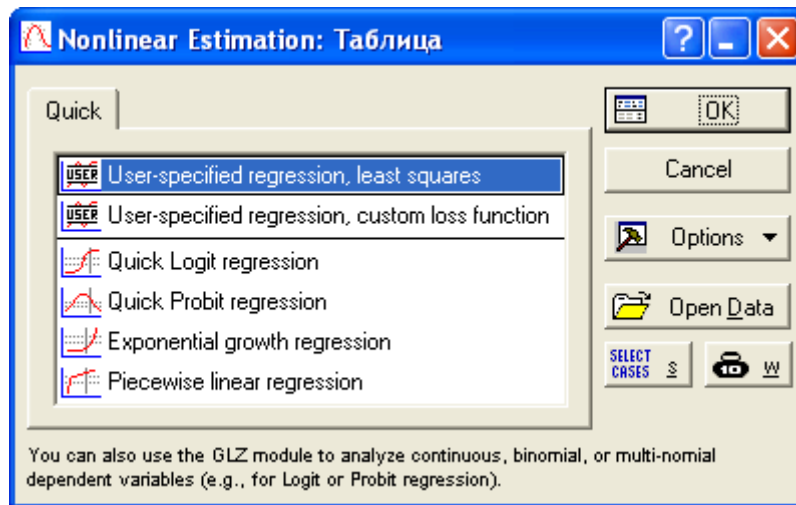


Рис. 13.7. Вікно процедури Nonlinear Estimation

У вікні (рис. 50) вибираємо функцію User-specified Regression, Least Squares (побудова моделей регресії користувачем вручну, параметри рівняння знаходяться за МНК).

У наступному діалоговому вікні натискаємо на кнопку Function to be estimated, щоб потрапити на екран для завдання моделі вручну (рис. 51).

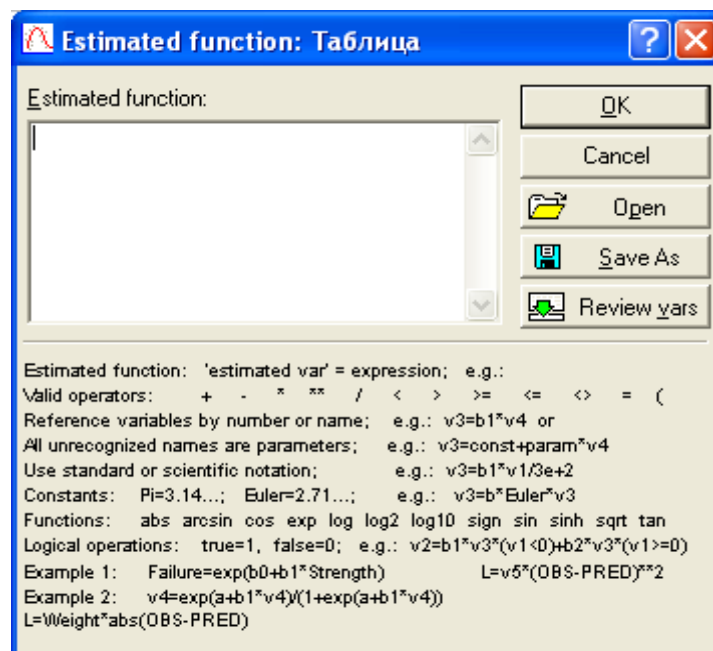


Рис. 51. Вікно для реалізації процедури завдання рівняння тренда вручну

У верхній частині екрану знаходиться поле для введення функції, в нижній частині розташовуються приклади введення функцій для різних ситуацій.

Перш ніж сформувані моделі, необхідно пояснити деякі умовні позначення. Змінні рівнянь задаються у форматі "v №", де "v" позначає змінну (від англ. "Variable"), а "№" – номер стовпця, в якому вона розташована в таблиці на робочому аркуші з вихідними даними. Якщо змінних дуже багато, то справа знаходиться кнопка Review vars, що дозволяє вибирати їх зі списку за назвами й переглядати їх параметри (рис.52).

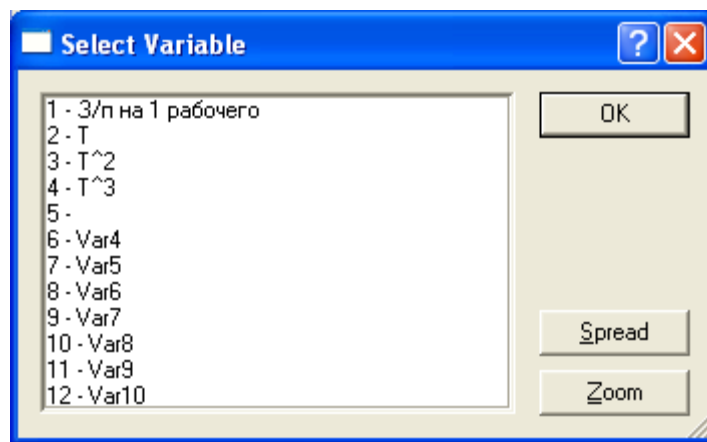


Рис. 52. Вікно вибору змінної за допомогою кнопки Review vars

Для спрощення роботи пропонується позначати параметри рівняння так, як в описі рівнянь тренда – латинською літерою "a", послідовно привласнюючи їм порядкові номери. Знаки математичних дій (віднімання, додавання, множення тощо) задаються в звичайному для Windows-додатків форматі. Розглянемо першу модель тренда – лінійну:

$$\bar{y}_t = a_0 + a_1 t.$$

Отже, після набору вона буде виглядати таким чином:

$$v_1 = a_0 + a_1 \times v_2,$$

де v_1 – це стовець на аркуші з вихідними даними, в якому знаходяться значення вихідного динамічного ряду;

a_0 і a_1 – параметри рівняння;

v_2 – стовець на аркуші з вихідними даними, в якому знаходяться значення моментів часу (змінна T).

Після цього двічі натискаємо кнопку ОК (рис. 53).

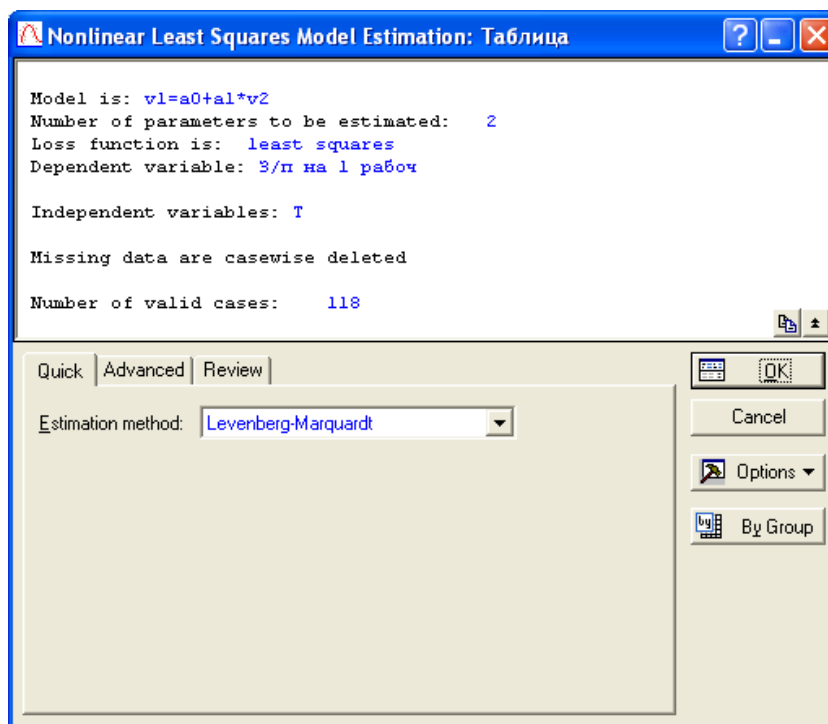


Рис. 53. Закладка Quick процедури оцінки рівняння тренда

У вікні (рис. 53) можна вибрати метод оцінки параметрів рівняння регресії (Estimation method), якщо це необхідно. У нашому випадку потрібно перейти до закладки Advanced і натиснути на кнопку Start values . У цьому діалозі задаються стартові значення параметрів рівняння для їх знаходження по МНК, тобто їх мінімальні значення.

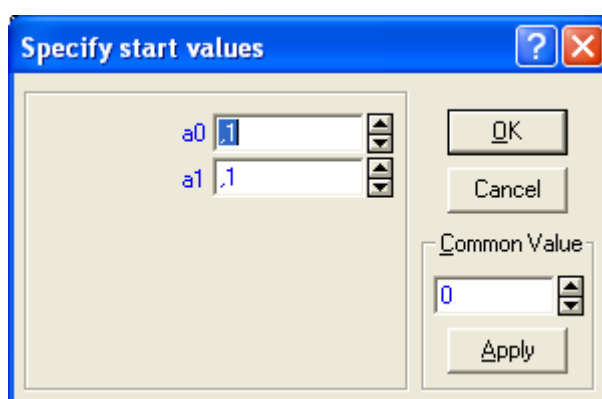


Рис. 54. Вікно завдання стартових значень параметрів рівняння тренда

Спочатку вони задані як 0,1 для всіх параметрів. У нашому випадку можна залишити ці значення в тому ж вигляді, але якщо значення в

наших вихідних даних менше одиниці, то необхідно задати їх у вигляді 0,001 для всіх параметрів рівняння тренда (рис. 54). Натискаємо кнопку ОК.

На закладці Quick (рис. 55) дуже важливим є значення рядка Proportion of variance accounted for, яке відповідає коефіцієнту детермінації.

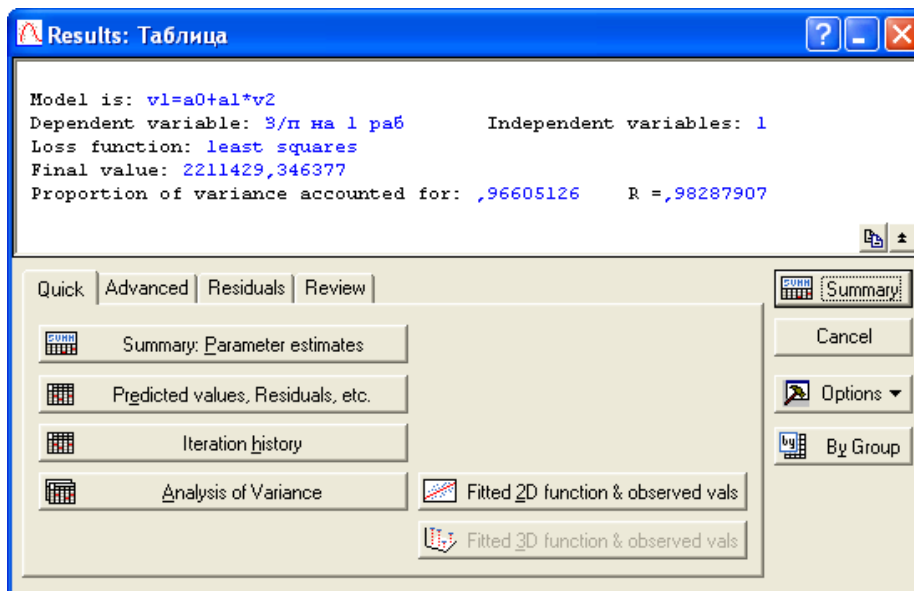


Рис. 55. Закладка Quick вікна результатів регресійного аналізу

Далі натискаємо кнопку Summary: Parameter estimates для отримання даних про параметри лінійного рівняння тренда.

	Estimate	Standard error	t-value df = 116	p-level	Lo. Conf Limit	Up. Conf Limit
a0	21,30360	25,58366	0,83270	0,406723	-29,3681	71,97527
a1	21,43917	0,37316	57,45362	0,000000	20,7001	22,17825

Рис. 56. Результати розрахунку параметрів лінійної моделі тренда

У таблиці (рис. 56) стовпець Estimate – числові значення параметрів рівняння; Standard error – стандартна помилка параметра; t-value – розрахункові значення t-критерію; df – число ступенів свободи (n-2); p-level – розрахунковий рівень значимості; Lo. Conf. Limit и Up. Conf. Limit – відповідно нижня и верхня межа довірчих інтервалів для параметрів рівняння з встановлення ймовірністю (вказана як Level of Confidence у верхньому полі таблиці).

Відповідно рівняння моделі регресії має вигляд:

$$\bar{y} = 21,3 + 21,43 \times t.$$

Після цього повертаємося до аналізу й натискаємо на кнопку Analysis of Variance (дисперсійний аналіз) на тій же закладці Quick (див. рис. 55).

Наступним кроком з'являється таблиця (рис. 57).

Model is: v1=a0+a1*v2 (Таблиця) Dep. Var. : 3/n на 1 рабочего					
Effect	1 Sum of Squares	2 DF	3 Mean Squares	4 F-value	5 p-value
Regression	261409379	2,0000	130704690	6856,083	0,00
Residual	2211429	116,0000	19064		
Total	263620809	118,0000			
Corrected Total	65140250	117,0000			
Regression vs. Corrected Total	261409379	2,0000	130704690	234,762	0,00

Рис. 57. Результати дисперсійного аналізу лінійної моделі тренда

У верхньому рядку таблиці знаходяться п'ять оцінок:

Sum of Squares – сума квадратів відхилень;

df – число ступенів свободи;

Mean Squares – середній квадрат;

F-value – критерій Фішера;

p-value – розрахунковий рівень значимості F-критерію.

У лівому стовпчику вказується джерело варіації:

Regression – квадрати теоретичних (отриманих з тренда) значень ознаки;

Residual – відхилення фактичних значень від теоретичних (отриманих за рівнянням тренда);

Total – відхилення фактичних значень у від їх середньої величини.

Model is: v1=a0+a1*v2 (Таблиця) Dep. Var. : 3/n на 1 рабочего			
	Observed	Predicted	Residuals
1	320,760	42,743	278,017
2	328,700	64,182	264,518
3	354,810	85,621	269,189
4	355,780	107,060	248,720
5	358,880	128,499	230,381
6	377,410	149,939	227,471
7	398,100	171,378	226,722
8	390,070	192,817	197,253
9	391,140	214,256	176,884
10	397,490	235,695	161,795

Рис. 58. Таблиця спостережуваних, прогнозних значень і залишків для лінійної моделі тренда

Далі доцільно побудувати графічне зображення, на якому лінія лінійного тренда буде накладена на вихідний динамічний ряд – це дозволить візуально оцінити ступінь відповідності. Для цього на робочому аркуші з прогнозними значеннями (рис. 58) виділяємо стовпці Observed і Predicted, клацаємо по них правою кнопкою миші і вибираємо функції Graphs of Block Data / Line Plot: Entire Columns (рис. 59).

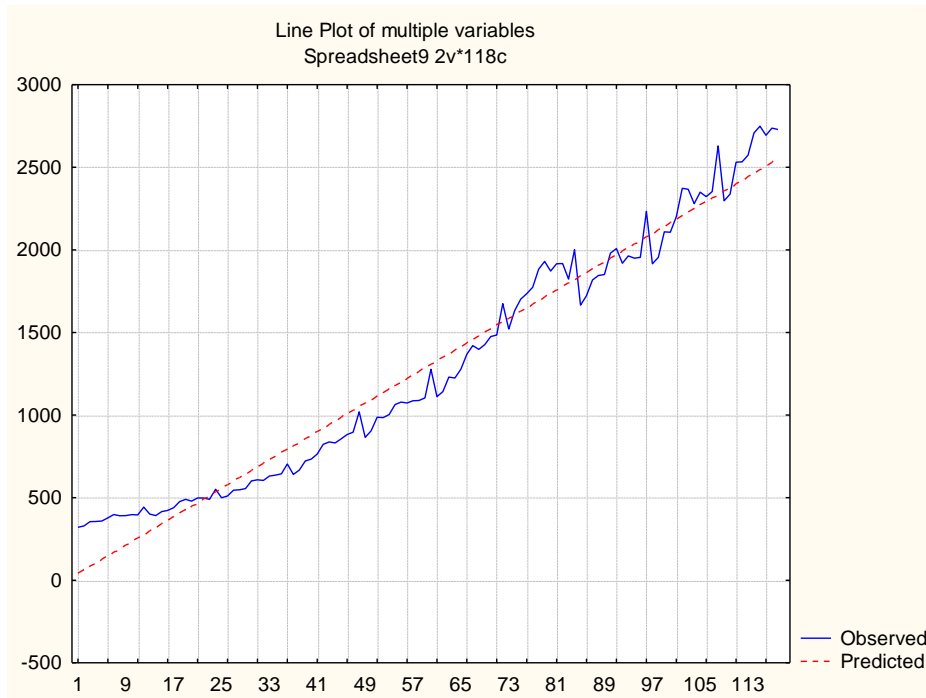


Рис. 59. Оригінальний динамічний ряд і лінійний тренд

Один з найбільш поширених методів прогнозування полягає в екстраполяції, тобто в продовженні в майбутнє тенденції, що спостерігалася в минулому. Екстраполяція тенденцій динамічних рядів порівняно широко застосовується в практичних дослідженнях в силу її простоти, можливості здійснення на основі відносно невеликого обсягу інформації, нарешті, ясності прийнятих припущень. Відсутність іншої інформації крім окремо розглянутого динамічного ряду часто виявляється вирішальним аргументом при виборі цього методу прогнозування.

Проведемо прогнозування на основі екстраполяції кращою лінійного тренда, отриманого вище

$$\bar{y} = 21,3 + 21,43 \times t.$$

Відповідно прогноз середньомісячної зарплати на листопад ($t = 119$) і грудень ($t = 120$) 2011 року складе:

$$\bar{y} = 21,3 + 21,43 \times 119 = 2571,47,$$
$$\bar{y} = 21,3 + 21,43 \times 120 = 2592,9.$$

Екстраполяція дає можливість отримати точкове значення прогнозу, що може бути визнано задовільним тільки при наявності функціональної залежності. Однак для економічних явищ характерна кореляційна залежність і змінні, як правило, є безперервними. Таким чином, прогноз має бути представлений у вигляді інтервалу значень, тобто необхідно визначення довірчого інтервалу прогнозу.

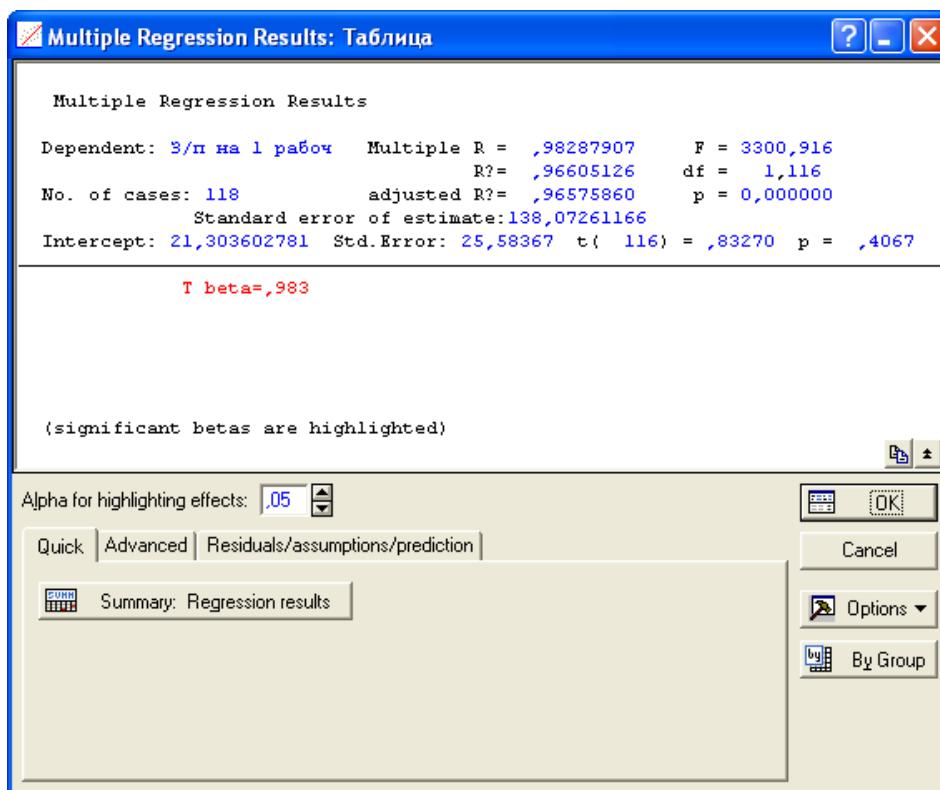


Рис. 60. Результати розрахунку лінійного тренда

У STATISTICA можливо використовувати кореляційно-регресійний аналіз для лінійного тренда. Для цього будуємо просту лінійну модель, як це було показано в лабораторній роботі № 10. Результати розрахунку лінійного тренда представлені на рис. 60.

У вкладці Residuals / assumptions / prediction використовуйте кнопку Predict dependent values та у вікні, що відкрилося необхідно вести показник наступного рівня ряду (рис. 61).

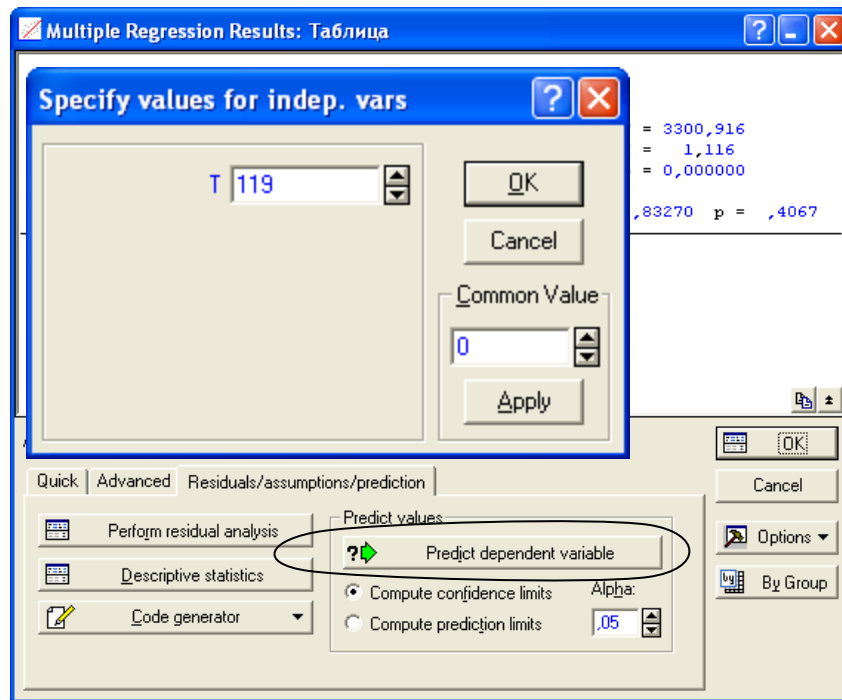


Рис. 61. Вікно введення прогнозного рівня ряду динаміки

Результати прогнозу представлені на рис. 62.

Predicting Values for (Таблица) variable: 3/п на 1 рабочего			
Variable	B-Weight	Value	B-Weight * Value
T	21,43917	119,0000	2551,261
Intercept			21,304
Predicted			2572,565
-95,0%CL			2521,893
+95,0%CL			2623,237

Рис. 62. Таблица результатів прогнозу та довірчі інтервали

Відповідно до рис. 62 прогнозний рівень заробітної плати на одного працюючого на листопад 2011 року буде складати 2 572,56 грн, довірчі інтервали прогнозу з 2 521,89 до 2 623,24грн.

Лабораторна робота № 14 "Автокореляція"

Мета – закріплення теоретичного та практичного матеріалу за темою "Кореляція динамічних рядів", придбання навичок побудови кореляції та автокореляції в Excel.

Завдання – необхідно перевірити наявність кореляційного зв'язку

між відповідними показниками в ППП Excel. Побудувати автокореляцію, зробити висновки, надати економічну інтерпретацію отриманим результатам.

Методичні рекомендації

Автокореляція – це взаємозв'язок послідовних елементів часового чи просторового ряду даних.

В економічних дослідженнях часто виникають і такі ситуації, коли дисперсія залишків стала, але спостерігаються їх коваріації. Це явище називають автокореляцією залишків. Автокореляція залишків найчастіше спостерігається тоді, коли економетрична модель будується на основі часових рядів. Якщо існує кореляція між послідовними значеннями деякої незалежної змінної, то спостерігається і кореляція послідовних значень залишків.

Автокореляція залишків на прикладі економетричної моделі з двома змінними має такий вигляд:

$$y_t = \alpha_0 + \alpha_1 x_t + u_t,$$

де залишки u_t задовольняють схему авторегресії першого порядку, тобто залежать тільки від залишків попереднього періоду.

Для перевірки наявності автокореляції залишків:

1) найчастіше застосовується критерій Дарвіна – Уотсона (DW):

$$DW = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2};$$

2) критерій фон Неймана:

$$Q = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\frac{\sum_{t=1}^n u_t^2}{n}}.$$

Фактичне значення критерію фон Неймана порівнюється з табличним для вибраного рівня значущості й заданого числа спостережень. Якщо $Q_{\text{табл}} > Q_{\text{факт}}$, то існує додатна автокореляція.

Побудуємо автокореляцію на основі табл. 8.

Таблиця 8.

Вихідні дані для аналізу

t	y	X ₁	X ₂	t	y	X ₁	X ₂
1	42,08	143,53	16,67	11	52,47	20,77	19,33
2	41,49	15,30	16,81	12	50,68	21,17	17,04
3	39,06	15,92	19,50	13	51,64	21,34	16,74
4	45,09	17,41	22,12	14	56,19	22,91	19,81
5	51,67	18,37	22,34	15	66,22	22,96	31,92
6	51,18	18,83	17,47	16	63,23	23,69	26,31
7	54,78	18,84	20,24	17	68,96	24,82	25,93
8	60,33	19,71	20,37	18	64,26	25,54	21,96
9	49,76	20,01	12,71	19	63,75	25,63	24,05
10	55,46	20,26	22,98	20	69,68	28,73	25,66

У цій таблиці через t позначено час (місяць), y – вартість однієї пластмасової чашки для чаю, x₁ – вартість матеріальних ресурсів, використаних на виробництво однієї чашки, x₂ – вартість робочої сили (заробітна плата), витрачена на виробництво однієї чашки. Крім цього, фірма-конкурент веде роботу по зниженню собівартості випускаємої продукції та величини її витрат на матеріальні та людські ресурси, які необхідні для виробництва однієї чашки. Необхідно побудувати модель, яка буде відображати залежність вартості чашки від відповідних факторів, з метою отримання можливості прогнозувати ціни.

Розв'язання:

1. Введення даних з включенням у модель додаткової змінної x₀, яка приймає тільки одне значення та дорівнює 1.

2. Знаходження параметрів регресії МНК з використанням матричних функцій Excel.

2.1. Оцінка вектора коефіцієнтів регресії $\hat{b} = (X'X)^{-1} X'y$ за допомогою функції ТРАНСП, МУМНОЖ та МОБР.

2.1.1. Формування оберненої матриці до матриці системи нормальних рівнянь $(X'X)^{-1}$.

1,9071	-0,0603	-0,0286
-0,0603	0,0055	-0,0026
-0,0286	-0,0026	0,0039

2.1.2. Отримані вектори оцінок коефіцієнтів регресії.

3,8470
1,8108
0,6343

2.2. Розрахунок залишків $\hat{e} = y - \hat{y}$.

2.2.1. Знаходження вектора розрахункових значень $\hat{y} = X \hat{b}$.

2.2.2. Обчислення різниць $\hat{e} = y - \hat{y}$ та оформлення проміжних результатів у вигляді табл. 9.

Таблиця 9

Таблиця розрахунку помилок та розрахункових значень функції

t	y	\hat{y}	\hat{e}	t	y	\hat{y}	\hat{e}
1	42,08	40,77	1,31	11	52,47	53,72	-1,25
2	41,49	42,21	-0,73	12	50,68	52,99	-2,31
3	39,06	45,04	-5,99	13	51,64	53,11	-1,46
4	45,09	49,40	-4,31	14	56,19	57,90	-1,71
5	51,67	51,28	0,39	15	66,22	65,67	0,55
6	51,18	49,02	2,16	16	63,23	63,43	-0,21
7	54,78	50,80	3,98	17	68,96	65,24	3,73
8	60,33	52,46	7,88	18	64,26	64,02	0,24
9	49,76	48,14	1,61	19	63,75	65,51	-1,76
10	55,46	55,11	0,35	20	69,68	72,15	-2,46

2.3. Знаходження оцінки коваріаційної матриці Σ для вектора \hat{b} за формулою $\Sigma = \frac{1}{n-1} \hat{e}' \hat{e}$.

2.3.1. Знаходження добутку векторів $\hat{e}' \hat{e}$ за допомогою функції ТРАНСП та МУМНОЖ $\hat{e}' \hat{e} = 177,3741$.

2.3.2. Обчислення величини $\hat{\sigma}^2 = \frac{\hat{e}' \hat{e}}{n-1}$:

$$\hat{\sigma}^2 = 177,3741 / 17 = 10,4348.$$

2.3.3. Знаходження Σ як добуток $\hat{\sigma}^2$ та оберненої матриці $(X'X)^{-1}$.

19,8979	-0,6292	-0,2983
-0,6292	0,0574	-0,0270
-0,2983	-0,0270	0,0410

3. Перевірка гіпотези про наявність автокореляції з використанням критерію Дарбіна – Уотсона.

3.1. Обчислення фактичного значення критерію Дарбіна – Уотсона з використанням матричних операцій Excel

$$d = \frac{\hat{e}'\hat{e} - 1,537}{\hat{e}'\hat{e}} = 0,95$$

3.2. Формулювання гіпотез:

H_0 – у залишках немає автокореляції;

H_1 – у залишках є позитивна автокореляція;

H_1^* – у залишках є негативна автокореляція.

3.3. Визначення по таблицях значень критерія Дарбіна – Уотсона критичних значень $d_L^* = 1,10$ та $d_U^* = 1,537$ для $n = 20$, $m = 3$ при ймовірності $\alpha = 0,05$.

3.4. Порівняння фактичного та критичного значень критерія $\bar{d} < d_L^*$ свідчить про те, що з ймовірністю 95 % можна відхилити нуль-гіпотезу та зробити висновок про наявність позитивної автокореляції.

Наявність автокореляції означає, що $\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t$, тобто не виконуються припущення класичного регресійного аналізу, тому можна знайти більш ефективну оцінку ніж \hat{b} . Крім того, обчислення в пункті 2.3 коваріаційна матриця Σ є зміщеною оцінкою коваріаційної матриці \hat{b} та її неможливо використовувати для отримання стандартної помилки оцінок коефіцієнтів регресії.

4. Перетворення вхідних даних з метою усунення автокореляції в залишках.

4.1. Оцінка параметра ρ авто регресії першого порядку $e_t = \rho e_{t-1} + \delta_t$.

4.1.1. Обчислення $e_t e_{t-1}$ та e_t^2 . Оформлення результатів розрахунку в табл. 10.

Проміжні розрахунки для визначення коефіцієнта автокореляції

t	e_t	$e_t e_{t-1}$	e_t^2	t	e_t	$e_t e_{t-1}$	e_t^2
1	1,31		1,72	11	-1,25	-0,44	1,56
2	-0,73	-0,95	0,53	12	-2,31	2,89	5,35
3	-5,99	4,35	35,83	13	-1,46	3,39	2,14
4	-4,31	25,81	18,60	14	-1,71	2,90	2,92
5	0,39	-1,68	0,15	15	0,55	-0,94	0,30
6	2,16	0,84	4,67	16	-0,21	-0,11	0,04
7	3,98	8,59	15,83	17	3,73	-0,77	13,89
8	7,88	31,34	62,05	18	0,24	0,88	0,06
9	1,61	12,71	2,60	19	-1,76	-0,41	3,09
10	0,35	0,57	0,12	20	-2,46	4,33	6,07
Сума						92,91	177,37

4.1.2. Обчислення коефіцієнта автокореляції

$$\hat{\rho} = \frac{\sum_{t=1}^n \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^n \hat{e}_t^2} = 0,54$$

4.2. Перетворення вхідних даних за формулами

$$y = \hat{P}y, x_0 = \hat{P}x_0, x_1 = \hat{P}x_1, \text{ та } x_2 = \hat{P}x_2;$$

$$y: y_1 = (1 - \hat{\rho}^2)^{1/2} y_1 = (1 - 0,29)^{1/2} 42,08 = 35,37;$$

$$y_2 = y_2 - \hat{\rho}y_1 = 41,49 - 0,54 \times 42,08 = 18,68;$$

$$y_3 = y_3 - \hat{\rho}y_2 = 39,06 - 0,54 \times 41,49 = 16,58;$$

$$x_0: x_{01} = (1 - \hat{\rho}^2)^{1/2} x_{01} = (1 - 0,29)^{1/2} 1,00 = 0,84;$$

$$x_{02} = x_{02} - \hat{\rho}x_{01} = 1,00 - 0,54 \times 1,00 = 0,46;$$

$$x_{03} = x_{03} - \hat{\rho}x_{02} = 1,00 - 0,54 \times 1,00 = 0,46;$$

$$x_1: x_{11} = (1 - \hat{\rho}^2)^{1/2} x_{11} = (1 - 0,29)^{1/2} 14,53 = 12,21;$$

$$x_{12} = x_{12} - \hat{\rho}x_{11} = 15,30 - 0,54 \times 14,53 = 7,47;$$

$$x_{13} = x_{13} - \hat{\rho}x_{12} = 15,92 - 0,54 \times 15,30 = 7,63;$$

$$x_2: x_{21} = (1 - \hat{\rho}^2)^{1/2} x_{21} = (1 - 0,29)^{1/2} 16,74 = 14,07;$$

$$x_{22} = x_{22} - \hat{\rho}x_{21} = 16,81 - 0,54 \times 16,74 = 7,74;$$

$$x_{23} = x_{23} - \hat{\rho}x_{22} = 19,50 - 0,54 \times 16,81 = 10,39.$$

Результати розрахунків

t	y	X ₀	X ₁	X ₂	t	y	X ₀	X ₁	X ₂
1	35,37	0,84	12,23	14,07	11	22,41	0,46	9,79	6,88
2	18,68	0,46	7,43	7,74	12	22,24	0,46	9,91	6,57
3	16,58	0,46	7,63	10,39	13	24,18	0,46	9,87	7,51
4	23,93	0,46	8,78	11,55	14	28,20	0,46	11,35	10,74
5	27,24	0,46	8,94	10,35	15	35,77	0,46	10,55	21,19
6	23,18	0,46	8,88	5,36	16	27,35	0,46	11,25	9,01
7	27,04	0,46	8,64	10,77	17	34,70	0,46	11,98	11,67
8	30,65	0,46	9,50	9,40	18	26,89	0,46	12,09	7,91
9	17,06	0,46	9,33	1,67	19	28,93	0,46	11,79	12,15
10	28,50	0,46	9,42	16,09	20	35,14	0,46	14,84	12,63

5. Знаходження параметрів регресії $\hat{y} = X'b + e$ за допомогою МНК, що еквівалентно застосуванню загального МНК до вихідних даних з використанням матричних функцій Excel (аналогічно п. 2).

5.1. Оцінка вектора коефіцієнтів регресії $\hat{b} = (X'X)^{-1}X'y$ за допомогою рівняння:

4,0352
1,6725
0,7593

5.2. Знаходження коваріаційної матриці $\hat{\Sigma} = \hat{\sigma}^2(X'X)^{-1}$ для \hat{b} .

37,57	-1,51	-0,21
-1,51	0,09	-0,01
-0,21	-0,01	0,02

6. Отримання прогнозного значення y, який задається $x_{\tau+1} = (1 \ 2 \ 0)$.

6.1. Знаходження прогновної оцінки загальним МНК, ігноруючи той факт, що $\hat{a}_{\tau+1}$ корельовано з попереднім значенням у вибірковому періоді

$$\hat{y}_{\tau+1} = x'_{\tau+1} \bar{b} = 58,69.$$

6.2. Знаходження прогновної оцінки загальним методом МНК з урахуванням того, що $\hat{a}_{\tau+1}$ корельовано з попереднім значенням в вибірковому періоді $\hat{y}_{\tau+1} = x'_{\tau+1} \bar{b} + \hat{\rho}(y_{\tau} - x'_{\tau} \bar{b})$.

Рекомендована література

1. Бабешко Л. О. Основы эконометрического моделирования : учебн. пособ. / Л. О. Бабешко. Изд. 3-е, стереотип. – М. : КомКнига, 2007. – 432 с.
2. Вуколов Э. А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL / Э. А. Вуколов. – М. : Форум: ИНФРА-М, 2004. – 464 с.
3. Джессен Р. Методы статистических обследований / Р. Джессен; пер. с англ. Ю. П. Лукашина, Я. Ш. Паппэ; под ред. Е. М. Четыркина. – М. : Финансы и статистика, 1985. – 318 с.
4. Кендэл М. Временные ряды / М. Кендэл; пер. с англ. и предисл. Ю. П. Лукашина. – М. : Финансы и статистика, 1981. – 358 с.
5. Статистическое моделирование и прогнозирование / учеб. пособ. / под. ред. А. Г. Гранберга. – М. : Финансы и статистика, 2000, – 384 с.
6. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров; под ред. В. Э. Фигурнова – 3-е изд. – М. : ИНФРА-М, 2003. – 544 с.

НАВЧАЛЬНЕ ВИДАННЯ

Лабораторний практикум
з модуля 2 "Методи і моделі статистичного
дослідження рядів динаміки" навчальної дисципліни
"СТАТИСТИКА"
для студентів напряму підготовки
6.030506 "Прикладна статистика"
денної форми навчання

Укладачі: **Раєвнєва** Олена Валентинівна
Горохова Ольга Іванівна
Чанкіна Ірина Володимирівна та ін.

Відповідальний за випуск **Раєвнєва О. В.**

Редактор **Пушкар І. П.**

Коректор **Бриль В. О.**

План 2012 р. Поз. № 419.

Підп. до друку. Формат 60x90 1/16. Папір MultiCopy. Друк Riso.

Ум.-друк. арк. 4,0. Обл.-вид. арк. 5,0. Тираж прим. Зам. №

Видавець і виготівник – видавництво ХНЕУ, 61166, м. Харків, пр. Леніна, 9а

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи

Дк № 481 від 13.06.2001 р.

**Лабораторний практикум
з Модуля 2 "Методи і моделі
статистичного дослідження рядів
динаміки" навчальної дисципліни**

"СТАТИСТИКА"

**для студентів напряму підготовки
6.030506 "Прикладна статистика"
денної форми навчання**