МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»

МЕТОДИЧНІ ВКАЗІВКИ

до розрахунково-графічного завдання «Парний кореляційно-регресійний аналіз» за розділом «Економетрія» дисципліни «Економіко-математичне моделювання» для студентів заочної форми навчання спеціальностей 8.030601 «Менеджмент організацій», 8.050106 «Облік та аудит», 8.050107 «Маркетинг»

Затверджено
редакційно-видавничою
радою університету
протокол № 1 від 20.06.2012 р.

Харків
НТУ «ХПІ»
2013

Методичні вказівки до розрахунково-графічного завдання «Парний кореляційно-регресійний аналіз» за розділом «Економетрія» дисципліни «Економіко-математичне моделювання» для студентів заочної форми навчання спеціальностей 8.030601 «Менеджмент організацій», 8.050106 «Облік та аудит», 8.050107 «Маркетинг» / Уклад. О.Є. Скворчевський, В.Л. Товажнянський, Р.О. Побережний. – Х.: НТУ «ХПІ», 2013. – 52 с.

Укладачі: О.Є. Скворчевський
В.Л. Товажнянський
Р.О. Побережний

Рецензент проф. Л.М. Любчик

Кафедра організації виробництва та управління персоналом

ВСТУП

Сучасна економіка є значною мірою формалізованою наукою, в якій широко використовуються кількісні методи. Застосування математичних методів в економіці має багато розгалужень і сформувало декілька окремих дисциплін, а саме: дослідження операцій, математичне програмування, теорія ігор тощо. Серед цих дисциплін однією із найважливіших є економетрія, яка входить до циклу базової економічної підготовки не тільки в Україні, але і в Європі, Сполучених Штатах Америки, Російській Федерації та інших провідних країнах. Економетрія є наукою, що в основному застосовує методи математичної статистики для вивчення закономірностей протікання економічних явищ та процесів.

Однією із базових тем економетрії є парний кореляційно-регресійний аналіз. Із цієї теми починається майже кожний підручник із економетрії незалежно від країни його походження. Розуміння положень парного кореляційно-регресійного аналізу дозволяє осягнути більш складні розділи економетрії. Саме тому ця тема була обрана для виконання розрахунково-графічного завдання для студентів заочної форми навчання.

Метою виконання цього завдання є краще засвоєння студентами теоретичних положень та набуття практичних навичок проведення парного кореляційно-регресійного аналізу взаємозв'язку двох економічних величин. У свою чергу, знання та практичні навички парного кореляційно-регресійного аналізу стануть для студентів ключем для вивчення більш складних тем економетрії.

У першому розділі викладено теоретичні аспекти парного кореляційно-регресійного аналізу, а в другому – приклад його проведення «вручну». Однак сучасні економіко-математичні, зокрема економетричні, дослідження неможливо уявити без застосування комп'ютерної техніки. Тому в третьому розд. показано приклад виконання розрахунково-графічного завдання засобами Microsoft Excel 2010. У четвертому розділі наведено варіанти для виконання розрахунково-графічного завдання. При виконанні парного кореляційно-регресійного аналізу за своїм варіантом студент може виконати розрахунково-графічне завдання вручну (розд. 1) або за допомогою Microsoft Excel 2010 (розд. 2) за власним вибором. Також студент може комбінувати розрахунки вручну та розрахунки за допомогою Microsoft Excel. Обрання тієї чи іншої форми виконання завдання здійснюється персонально студентом, виходячи із рівня володіння ним комп'ютером.

1. ОСНОВНІ ТЕОРЕТИЧНІ ПОЛОЖЕННЯ ПАРНОГО КОРЕЛЯЦІЙНО-РЕГРЕСІЙНОГО АНАЛІЗУ

1.1. Парний коефіцієнт кореляції

Парний коефіцієнт кореляції r – міра тісноти *лінійного* зв'язку двох випадкових величин X та Y, поданих у вигляді двох вибірок однакового обсягу n:

$$X = \begin{pmatrix} x_1 \\ \dots \\ x_i \\ \dots \\ x_n \end{pmatrix}, \tag{1.1}$$

$$Y = \begin{pmatrix} y_1 \\ \dots \\ y_i \\ \dots \\ y_n \end{pmatrix}. \tag{1.2}$$

Парний коефіцієнт кореляції часто називають вибірковим або коефіцієнтом кореляції К. Пірсона. Як повну назву величини r можна використовувати таку: вибірковий лінійний парний коефіцієнт кореляції К. Пірсона, однак у літературі, як правило, користуються більш короткими назвами [1–6]. Парний коефіцієнт кореляції може бути розрахований за формулою [1, 2]:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2) \cdot (\overline{y^2} - \overline{y}^2)}}, \tag{1.3}$$

де $\overline{x}$, $\overline{y}$ – середнє значення незалежної та залежної змінних відповідно; $\overline{xy}$ – середнє значення добутку незалежної та залежної змінних; $\overline{x^2}$, $\overline{y^2}$ – середнє значення квадратів незалежної та залежної змінних відповідно.

Відповідні середні визначаються за формулами:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \tag{1.4}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}; \tag{1.5}$$

$$\overline{xy} = \frac{\sum_{i=1}^{n} (x_i \cdot y_i)}{n}; \tag{1.6}$$

$$\overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}. \tag{1.7}$$

Коефіцієнт кореляції може набувати значення від –1 до 1. За знаком та абсолютним значенням коефіцієнта кореляції можна зробити такі висновки:

1. Чим ближче абсолютне значення $|r|$ коефіцієнта кореляції до 1, тим тісніший зв'язок існує між двома змінними x та y (рис. 1.1 *а*), а чим ближче $|r|$ до нуля тим зв'язок слабкіший (рис. 1.1 *б*). Якщо парний коефіцієнт кореляції дорівнює 1, то між двома змінними існує лінійна функціональна залежність. При цьому усі спостереження на кореляційному полі розташовуються вздовж прямої лінії. Якщо парний коефіцієнт кореляції дорівнює 0, то між двома змінними залежність відсутня.

Для перетворення кількісної характеристики тісноти лінійного зв’язку між двома випадковими у якісну може бути використана шкала англійського статистика Чеддока (табл. 1.1).

Таблиця 1.1 – Шкала Чеддока

Величина абсолютного значення парного коефіцієнта кореляції	Характеристика лінійного зв’язку між двома випадковими величинами
до 0,3	практично відсутній
0,31–0,5	слабкий
0,51–0,7	помітний
0,71–0,9	сильний
0,91–0,99	дуже сильний

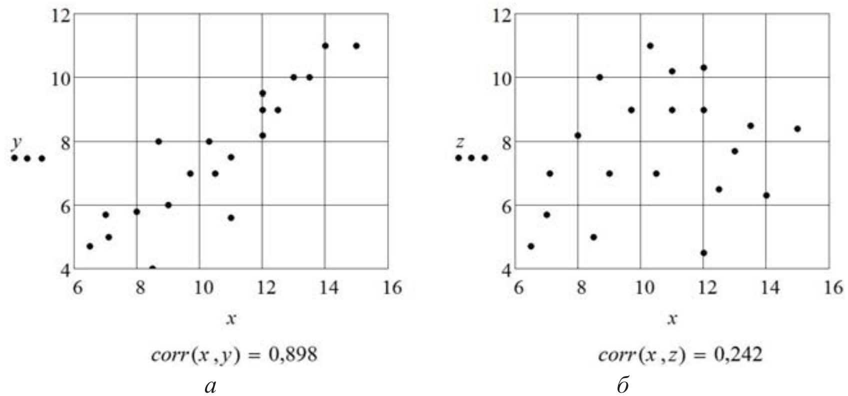


Рисунок 1.1 – Кореляційні поля: *а* – для вибірки сильнокорельованих даних; *б* – для вибірки слабокорельованих даних

2. Якщо коефіцієнт кореляції позитивний $r > 0$, то зв'язок між змінними прямий, тобто при збільшенні незалежної змінної x залежна змінна y теж збільшується (рис. 1.2 *а*). При $r < 0$ зв'язок між змінними зворотний, тобто при збільшенні незалежної змінної x залежна змінна y зменшується (рис. 1.2 *б*).

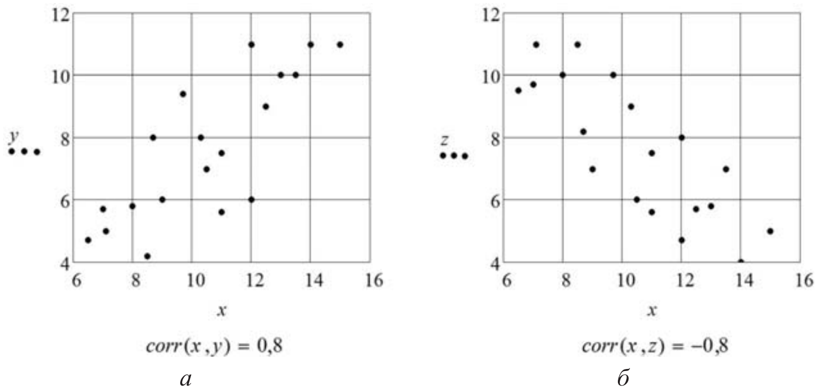


Рисунок 1.2 – Кореляційні поля: *а* – для вибірки даних з прямим зв’язком; *б* – для вибірки даних зі зворотним зв’язком

Парний коефіцієнт кореляції є випадковою величиною, оскільки обчислюється для випадкових величин. Для нього необхідно висувати і перевіряти гіпотезу про те, чи статистично значуще він відрізняється від нуля (тобто чи є взаємозв'язки між величинами). Ця гіпотеза перевіряється за допомогою *t*-критерію Ст'юдента, фактичне значення якого визначається за формулою:

$$t_{\text{факт}} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}. \tag{1.8}$$

Фактичне значення критерію Ст'юдента порівнюється із критичним $t_{p;k}$, отриманим за відповідною таблицею (додаток А) для степенів свободи $k = n-2$ та вірогідності p, яка в загальному випадку може обиратися довільно. В багатьох джерелах як критерій вибору пропонується рівень значущості $\alpha = 1-p$. На сучасному етапі розвитку економетрії та інших наук, пов'язаних із математичною статистикою, критичне значення критерію Ст'юдента можна визначити за допомогою функції СТЬЮДРАСПОБР Microsoft Excel 2003-2007 та функції СТЬЮДЕНТ.ОБР.2Х Microsoft Excel 2010. При використанні цих функцій у рядку «Вірогідність» діалогового вікна потрібно зазначати не вірогідність p, а рівень значущості α. Така невідповідність, скоріше за все, пов'язана із неточністю перекладу інтерфейсу статистичних функції із англійської мови на російську. В MathCAD критичне значення *t*-критерію Ст'юдента можна отримати за допомогою оператора $qt(p,k)$.

Якщо $|t_{\text{факт}}| > t_{p;k}$, то коефіцієнт кореляції статистично значуще відрізняється від нуля та залежність є достовірною. У протилежному випадку коефіцієнт кореляції статистично не значущий та кореляційний зв'язок між змінними відсутній.

Очевидно, що різним рівням вірогідності будуть відповідати різні значення $t_{p;k}$. Тобто при меншій вірогідності коефіцієнт кореляції може бути статистично значущим, а при більшій – не значущим. Тому обов'язково необхідно вказувати при якій вірогідності було обрано *t*-критерій Ст'юдента. Це можна зробити, відзначивши, що *коефіцієнт кореляції статистично значущий (або не значущий) при вірогідності, наприклад,* 0,9. Вказати вірогідність p (чи рівень значущості α) та степінь свободи, при якому обиралося табличне значення *t*-критерію можна в його індексі – $t_{p;k}$.

Зі збільшенням вірогідності p критичне значення *t*-критерію Ст'юдента буде зростати. Тому, якщо парний коефіцієнт кореляції виявився статистично не значущим для високої вірогідності (наприклад 0,95–0,99), доцільно перевірити його статистичну значущість при меншому рівні вірогідності (наприклад 0,8–0,85).

При проведенні кореляційного аналізу потрібно пам'ятати, що парний коефіцієнт кореляції є мірою тісноти *лінійного* зв'язку між двома випадковими величинами. Тому мале абсолютне значення коефіцієнта кореляції $|r|$ і (або) його статистична незначущість може свідчити лише про відсутність чи слабкість *лінійного* зв'язку, в той час як між двома випадковими змінними *може існувати тісний нелінійний зв'язок*. Одним із способів виявити такий зв'язок є аналіз відповідних кореляційних полів (див. тему «Нелінійна регресія»).

Коефіцієнт кореляції лише констатує факт присутності чи відсутності лінійного зв'язку між двома величинами X та Y, але він не вказує на причино-наслідкові зв'язки між ними. Таким чином визначити яка змінна є причиною, а яка наслідком можна, виходячи із економічного змісту задачі. Також в економіці зустрічаються випадки, коли тісний кореляційний зв'язок між двома величинами X та Y пояснюється дією третього фактора Z, що одночасно діє і на X, і на Y. Причому причинно-наслідкові зв'язки безпосередньо між X та Y відсутні.

Формула (1.3) оцінки коефіцієнта кореляції рекомендується до застосування при великій кількості спостережень та якщо r не близьке до $|1|$. Якщо величина коефіцієнта кореляції близька до 1, то розподіл його оцінок відрізняється від нормального або розподілу Ст`юдента, оскільки величина коефіцієнта кореляції обмежена значеннями від – 1 до + 1. Щоб обійти це ускладнення для оцінки істотності коефіцієнта кореляції вводиться допоміжна величина z [2].

Математичні вирази, що знаходяться в чисельнику та знаменнику формули (1.3), мають самостійний статистичний зміст. Так величина

$$Cov(X,Y) = \overline{xy} - \bar{x} \cdot \bar{y} \tag{1.9}$$

називається *вибірковим кореляційним моментом* або *вибірковою коваріацією.* Ця величина є мірою тісноти лінійного зв'язку двох випадкових величин X та

Y, однак, на відміну від коефіцієнта кореляції вибіркова коваріація має розмірність, що ускладнює її практичне застосування [1, 3]. Знак вибіркової коваріації має таку ж інтерпретацію, як і знак коефіцієнта кореляції.

У знаменнику формули (1.3) знаходиться корінь квадратний із добутку вибіркових дисперсій двох випадкових величин X та Y:

$$Var(X) = \overline{x^2} - \bar{x}^2; \tag{1.10}$$

$$Var(Y) = \overline{y^2} - \bar{y}^2. \tag{1.11}$$

Вибіркова дисперсія (*варіація*) – оцінка дисперсії випадкової величини за вибіркою. Під дисперсією випадкової величини тут потрібно розуміти міру розсіювання випадкової величини, тобто її відхилення від математичного очікування. Вибіркова дисперсія X позначається, як $Var(X)$ від англ. variance, також можуть бути такі позначення $D[x]$, σ_x^2. Дисперсія $Var(X)$ вимірюється у квадраті одиниці вимірювання випадкової величини X, що є незручним. Квадратний корінь із дисперсії називається *середньоквадратичним відхиленням* або *стандартним відхиленням.* Стандартне відхилення вимірюється у тих же величинах, що і сама випадкова величина.

Враховуючи формули (1.9)–(1.11), коефіцієнт кореляції (1.3) може бути розрахований так:

$$r = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}. \tag{1.12}$$

На сучасному етапі розвитку наук, пов’язаних із математичною статистикою, зокрема економетрії, немає необхідності здійснювати громіздкі розрахунки за формулами (1.3) чи (1.12). У Microsoft Excel існує функція «КОРРЕЛ», яка дозволяє розрахувати парний коефіцієнт кореляції на основі двох вибірок статистичних даних, а в MathCAD – функція *corr(X,Y)*.

1.2. Сутність парного лінійного регресійного аналізу

Регресійний аналіз – статистичний метод дослідження залежності випадкової змінної Y від однієї X або декількох незалежних змінних

$X_1,...,X_m$. Завданнями регресійного аналізу є встановлення форми залежності між змінними, оцінка функції регресії, прогноз значень залежної змінної. Будь-яке економетричне дослідження починається зі *специфікації моделі*, тобто із вибору структури моделі та визначення набору пояснюючих змінних. Вибір структури моделі може здійснюватися на основі економічної теорії, попередніх результатів, апріорних знань або спеціальних математичних процедур. Парна регресія достатня, якщо існує домінуючий фактор, який і використовується як незалежна змінна. Оскільки найбільш простою формою залежності в математиці є лінійна, то в регресійному аналізі найбільш популярні лінійні моделі. Саме їм приділяється основна увага у базовому курсі економетрії. Таким чином, у рамках цього пункту розглянемо парний лінійний регресійний аналіз, тобто статистичний метод дослідження лінійної залежності випадкової змінної Y від однієї X незалежної змінної. Тісноту лінійного зв’язку між ними можна визначити за парним коефіцієнтом кореляції, а причинно-наслідкові зв’язки – виходячи із економічного змісту задачі.

Парна лінійна регресія – причинна модель статистичного лінійного зв'язку між двома кількісними змінними Y та X, подана рівнянням:

$$\hat{Y} = \beta_0 + \beta_1 \cdot X + \varepsilon, \tag{1.13}$$

де X – незалежна змінна; Y – залежна змінна; β_0, β_1 – параметри моделі; ε – випадкова величина (збурення).

Випадкова величина ε включає вплив не врахованих у моделі факторів, випадкових похибок та особливостей вимірювання. Її присутність у моделі викликана трьома причинами: специфікацією моделі, вибірковим характером вихідних даних, особливостями вимірювання змінних [1].

В основі оцінки рівняння парної лінійної регресії лежать вибірки значень двох змінних – незалежної X (1.1) та залежної Y (1.2). Геометричною інтерпретацією таких вибірок розміром n буде сукупність n точок на площині (рис. 1.3 *а*), яка називається *діаграмою розсіювання* чи *кореляційним полем*.

Задачею парного лінійного регресійного аналізу буде знаходження такої прямої, яка найкращим чином аналітично характеризує кореляційне поле. Тобто потрібно мінімізувати деяку «міру» відхилень $g(\varepsilon_i)$ від лінії регресії (рис. 1.3 *б*).

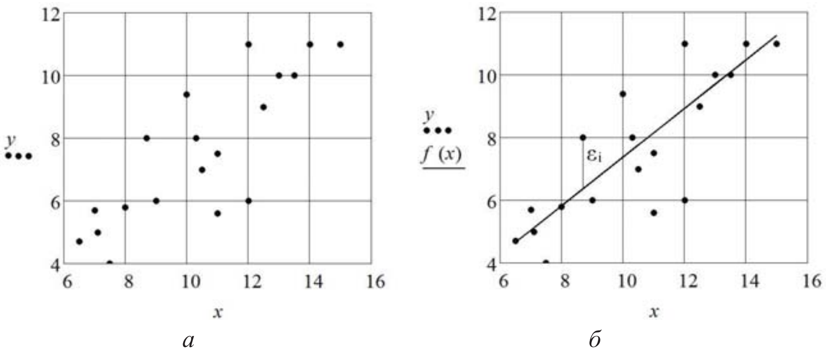


Рисунок 1.3 – Геометрична інтерпретація двох вибірок статистичних даних:
а – кореляційне поле; *б* – кореляційне поле із лінією регресії

Як міру відхилення функції $y_i = \beta_0 + \beta_1 \cdot x_i$ від точок набору спостережень можна взяти [4]:

1) суму квадратів відхилень $F = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \cdot x)]^2$;

2) суму модулів відхилень $F = \sum_{i=1}^{n} |\varepsilon_i| = \sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 \cdot x)|$;

3) функцію Хубера, яка при малих відхиленнях квадратична, а при великих – лінійна.

Із ряду причин, які будуть розглядатися нижче, найбільшого поширення як міра відхилення набула сума квадратів відхилень. Метод знаходження оцінок параметрів регресії за допомогою цієї функції дістав назву методу найменших квадратів.

1.3. Метод найменших квадратів

Згідно з методом найменших квадратів (МНК) оцінки невідомих параметрів β_0, β_1 обираються таким чином, щоб сума квадратів відхилень значень, знайдених за рівнянням $\hat{y}_i = \beta_0 + \beta_1 \cdot x_i$ від емпіричних значень y_i, була мінімальною [1–7]:

$$F = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \cdot x)]^2 \to \min. \qquad (1.14)$$

Функціонал (1.14) також іноді називають *функцією незв'язності*.

На підставі необхідної умови екстремуму функції двох змінних β_0, β_1 прирівнюємо до нуля її частинні похідні:

$$\begin{cases} \dfrac{\partial F}{\partial \beta_0} = -2 \cdot \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \cdot x_i)] = 0; \\ \dfrac{\partial F}{\partial \beta_1} = -2 \cdot \sum_{i=1}^{n} x_i \cdot [y_i - (\beta_0 + \beta_1 \cdot x_i)] = 0. \end{cases} \qquad (1.15)$$

Для подальших математичних викладок систему (1.15) доцільно записати у вигляді:

$$\begin{cases} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 \cdot x_i)] = 0; \\ \sum_{i=1}^{n} x_i \cdot [y_i - (\beta_0 + \beta_1 \cdot x_i)] = 0. \end{cases} \qquad (1.16)$$

Розкриємо дужки та отримаємо стандартну форму нормальних рівнянь:

$$\begin{cases} \beta_0 \cdot n + \beta_1 \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ \beta_0 \cdot \sum_{i=1}^{n} x_i + \beta_1 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i \cdot y_i. \end{cases} \qquad (1.17)$$

Потім, розділивши обидві частини системи на n, отримаємо систему із двох алгебричних рівнянь із двома невідомими β_0, β_1:

$$\begin{cases} \beta_0 + \beta_1 \cdot \bar{x} = \bar{y}; \\ \beta_0 \cdot \bar{x} + \beta_1 \cdot \overline{x^2} = \overline{xy}; \end{cases} \qquad (1.18)$$

Розв'язуючи систему (1.18) алгебричних рівнянь відносно β_0, β_1, одержимо:

$$\beta_0 = \bar{y} - \beta_1 \cdot \bar{x}\,; \tag{1.19}$$

$$\beta_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{Cov(X,Y)}{Var(X)}\,. \tag{1.20}$$

Потрібно відзначити, що (1.19) та (1.20) не є параметрами парної лінійної регресії, а лише оцінками цих параметрів, отриманими за МНК. Реальні значення параметрів невідомі, оскільки на практиці працюють не із генеральними сукупностями, а із вибірками (1.1), (1.2) з них.

Обрання як міри відхилення ліній регресії від емпіричних точок обґрунтовується *теоремою Гаусса-Маркова* [1, 4]: при виконанні ряду вимог до вихідних вибірок статистичних спостережень, МНК-оцінки параметрів лінійної регресії є найбільш *ефективними*, тобто мають найменшу дисперсію в класі усіх лінійних незміщених оцінок.

Вимоги до вихідних вибірок статистичних спостережень будуть розглянуті в темі «Множинний лінійний кореляційно-регресійний аналіз», оскільки він є узагальненням парного лінійного кореляційно-регресійного аналізу на моделі із довільним числом незалежних змінних.

Величина $\hat{\beta}_j$ називається *незміщеною оцінкою* параметра β_j, якщо її вибіркове математичне очікування дорівнює параметру β_j генеральної сукупності $M[\hat{\beta}_j] = \beta_j$. Змістовно незміщеність оцінки означає, що вона не має систематичної похибки [4, 7].

Серед характеристик оцінок параметрів також потрібно відзначити *обґрунтованість* – властивість оцінки $\hat{\beta}_j$ параметра β_j наближатися до його реального значення при збільшенні обсягу n вибірки, за якою проводилася оцінка $\lim\limits_{n \to \infty} \hat{\beta}_j = \beta_j$ [4].

Оскільки МНК при виконанні ряду вимог до вихідних вибірок спостережень дає найбільш ефективні оцінки параметрів парної лінійної регресії, його алгоритм реалізований багатьма комп'ютерними програмами. Так у Microsoft Excel існує статистична функція «ЛИНЕЙН» та інструмент «Регресія» надбудови «Аналіз даних». У MathCAD – оператор *line(X,Y)*.

1.4. Економічний зміст параметрів парної лінійної регресії

Економічний зміст параметрів парної лінійної регресії визначається сутністю залежної X та незалежною Y змінними. Розглянемо декілька прикладів.

У підручнику [2] вивчається потреба підприємства в електроенергії Y залежно від обсягу продукції X, що випускається. Якщо потреба в електроенергії може бути оцінена рівнянням парної лінійної регресії $\hat{Y} = \beta_0 + \beta_1 \cdot X$, то все споживання електроенергії $\hat{Y}$ можна розділити на дві частини – безпосередньо не пов'язане із виробництвом продукції β_0 та безпосередньо пов'язане з обсягом продукції, що випускається $(\beta_1 \cdot x)$, яке пропорційно зростає із збільшенням обсягу випуску X. Причому β_1 показує оцінку безпосередніх витрат електроенергії на кожну одиницю виробленої продукції.

За аналогією із цим прикладом можна запропонувати такий. Залежність витрат підприємства від обсягу виробництва взагалі характеризується нелінійною залежністю. Однак на певному невеликому відрізку вони можуть оцінюватися рівнянням регресії $\hat{Y} = \beta_0 + \beta_1 \cdot X$. Тоді витрати підприємства можуть бути розділені на умовно-постійні β_0, такі, що не змінюються із зміною обсягу виробництва (орендна плата, утримання адміністрації та ін.), і на умовно-змінні $(\beta_1 \cdot X)$, такі, що змінюються пропорційно зміні обсягу продукції (витрата матеріалу, заробітна плата працівників зайнятих безпосередньо на виробництві, та ін.). Причому тут β_1 – оцінка умовно-змінних витрат на одиницю виробленої продукції.

Наведені вище приклади ілюструють пряму залежність Y від X. Розглянемо зворотну залежність кількості бракованої продукції Y (%) від кількості робітників із спеціальною підготовкою X (%). Якщо така залежність буде описуватися рівнянням $\hat{Y} = \beta_0 - \beta_1 \cdot X$, то β_0 – оцінка кількості бракованої продукції при відсутності робітників із спеціальною підготовкою $x = 0$, β_1 – оцінка зменшення кількості (%) бракованої продукції при збільшенні кількості робітників із спеціальною підготовкою на 1 %.

Параметр β_0 не завжди буде мати економічний зміст, зокрема він не має економічного змісту, якщо він від'ємний. У [3] розглядається приклад, коли параметр β_0 не має економічного змісту, навіть в тому випадку, коли він позитивний.

1.5. Перевірка статистичної значущості та довірчі інтервали оцінок параметрів регресії

Параметри регресії оцінюються на основі вибірки емпіричних величин, у значеннях яких суттєву роль відіграють випадкові фактори. Це призводить до наявності суттєвої випадкової складової в оцінці кожного із параметрів, тому вони потребують перевірки статистичної значущості та побудови довірчих інтервалів. З цією метою за кожним з параметрів визначається його стандартна помилка – m_{β_0}, m_{β_1} [1–7]. Стандартна помилка оцінки параметра регресії β_1 визначається за формулою:

$$m_{\beta_1} = \sqrt{\frac{S^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}, \tag{1.21}$$

де S^2 – залишкова дисперсія на один степінь свободи.

У разі нормального розподілу величин відхилень ε вона розраховується за формулою:

$$S^2 = \frac{\sum_{i=1}^{n}\varepsilon_i^2}{n-2} = \frac{\sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 \cdot x_i)]^2}{n-2}. \tag{1.22}$$

Далі визначається фактичне значення *t*-критерію Ст'юдента:

$$t_{\beta_1} = \frac{\beta_1}{m_{\beta_1}}, \tag{1.23}$$

яке потім порівнюється критичним (табличним), обраним при певному рівні вірогідності p, або значущості α, і числі степенів свободи $k = n - 2$.

Якщо фактичне значення *t*-критерію перевищує табличне:

$$t_{\beta_1} > t_{p,k}, \tag{1.24}$$

то гіпотезу про статистичну незначущість оцінки параметра β_1 регресії можна відхилити.

Довірчий інтервал оцінки параметра β_1 регресії визначається:

$$\beta_1 - t_{k,p} \cdot m_{\beta_1} \le b_1 \le \beta_1 + t_{k,p} \cdot m_{\beta_1}, \tag{1.25}$$

де b_1 – дійсне значення параметра β_1 регресії, яке могло б бути отримане при оцінці за генеральною сукупністю.

Стандартна помилка параметра β_0 визначається за формулою:

$$m_{\beta_0} = \sqrt{S^2 \cdot \frac{\sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2}}. \tag{1.26}$$

В іншому послідовність оцінювання статистичної значущості оцінки параметра β_0 і побудови її довірчого інтервалу не відрізняється від розглянутої вище для β_1.

Оскільки оцінка параметра регресії в економетричних дослідженнях має чітку економічну інтерпретацію, то довірчі межі інтервалу для неї не повинні містити суперечливих результатів, наприклад $-10 \le b_j \le 40$. Запис , що дійсне значення параметра регресії може приймати позитивні, від'ємні і навіть нульове значення, чого, скоріше за все, не може бути з економічної точки зору.

При перевірці статистичної значущості та побудові довірчих інтервалів параметрів регресії обов'язково необхідно вказувати вірогідність *p*, для якої визначалося $t_{p,k}$. Це пов'язано із тим, що для одного рівня вірогідності оцінка параметра регресії може бути статистично значущою, в той же час, як для вищого рівня вірогідності вона буде статистично незначущою. Межі довірчих інтервалів також будуть відрізнятися для різних значень вірогідності *p*.

1.6. Коефіцієнт детермінації

Окрім перевірки статистичної значущості оцінок параметрів парної лінійної регресії, важливо також визначити якість моделі в цілому. Однією з найбільш ефективних оцінок адекватності регресійної моделі, мірою якості рівняння регресії, характеристикою її прогностичної сили є *коефіцієнт* *детермінації*. Цей коефіцієнт показує частину дисперсії, пояснену рівнянням регресії [3]:

$$R^2 = \frac{Var(\hat{Y})}{Var(Y)}. \tag{1.27}$$

Іншою формою подання формули (1.27) може бути така:

$$R^2 = 1 - \frac{Var(\varepsilon)}{Var(Y)}. \tag{11.28}$$

У разі парної лінійної регресійної моделі коефіцієнт детермінації рівний квадрату коефіцієнта кореляції:

$$R^2 = r^2. \tag{1.29}$$

Чим ближче R^2 до одиниці, тим краще регресія апроксимує емпіричні дані, тим тісніше точки спостереження примикають до лінії регресії. Якщо $R^2 = 1$, то емпіричні точки, що характеризують незалежну X та залежну Y змінні, лежать на лінії регресії і між змінними існує лінійна функціональна залежність. Якщо $R^2 = 0$, то варіація залежної змінної повністю обумовлена дією неврахованих у моделі змінних, і лінія регресії паралельна осі абсцис. Зазначимо, що коефіцієнт R^2 є сенс розглядати тільки за наявності параметра β_0 в рівнянні регресії.

Якщо коефіцієнт детермінації відомий, то критерій його статистичної значущості (F-критерій Фішера) для парної лінійної регресії може бути записаний у вигляді:

$$F = \frac{R^2 \cdot (n-2)}{1-R^2}. \tag{1.30}$$

Для заданого рівня значущості α, степенів свободи k_1 та k_2 знаходиться табличне (додаток В) значення критерію Фішера (у деяких джерелах він

називається критерієм Фішера-Снедекора). Причому $k_1 = m-1$, де m – кількість змінних у моделі (для парної регресії $k_1 = 1$), $k_2 = n-2$. На сучасному етапі розвитку економетрії та інших наук, пов'язаних із математичною статистикою, критичне значення критерію Фішера можна визначити за допомогою статистичних функцій F.ОБР та F.ОБР.ПХ Microsoft Excel 2010. При використанні цієї функції F.ОБР.ПХ у рядку «Вірогідність» діалогового вікна потрібно зазначати не вірогідність p, а рівень значущості α. Така невідповідність довідки за функцією, скоріше за все, пов'язана із неточністю перекладу інтерфейсу статистичних функції із англійської мови на російську. В більш старих версіях Microsoft Excel використовується функція FРАСПОБР, у рядку «Вірогідність» діалогового вікна якої необхідно вказувати не вірогідність p, а рівень значущості α.

У MathCAD критичне значення F-критерію Фішера можна отримати за допомогою оператора $qF(p, k_1, k_2)$.

Якщо $F > F_{\text{кр}}$, то для вірогідності p коефіцієнт детермінації статистично значущий.

Після побудови та дослідження якості моделі парної лінійної регресії можна перейти до прогнозування розвитку економічного явища чи процесу, який описує ця модель.

1.7. Прогнозування за лінійним рівнянням регресії

У прогнозних розрахунках за рівнянням регресії визначається прогнозне значення залежної змінної, при деякому прогнозному значенні незалежної змінної $x_{\text{пр}}$. Прогноз поділяють на *точковий* (коли оцінкою прогнозної величини є число) та *інтервальний* (коли оцінкою є інтервал).

Точковою оцінкою залежної змінної вважають її умовне математичне очікування $M_{x_{\text{пр}}}(y)$, яке ототожнюють [1] із груповою середньою $\hat{y}_{x_{\text{пр}}}$, яку знаходять за рівнянням регресії:

$$\hat{y}_{x_{\text{пр}}} = \beta_0 + \beta_1 \cdot x_{\text{пр}}. \tag{1.31}$$

Дисперсію групової середньої оцінюють за формулою:

$$s^2_{\hat{y}_{x_{\text{пр}}}} = s \cdot \left(\frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right). \qquad (1.32)$$

Тоді стандартна помилка групової середньої:

$$s_{\hat{y}_{x_{\text{пр}}}} = \sqrt{s^2_{\hat{y}_{x_{\text{пр}}}}}. \qquad (1.32)$$

Тоді p %-й довірчий інтервал для умовного математичного очікування становить:

$$\hat{y}_{x_{\text{пр}}} - s_{\hat{y}_{x_{\text{пр}}}} \cdot t_{p;k} \le M_{x_{\text{пр}}}(y) \le \hat{y}_{x_{\text{пр}}} + s_{\hat{y}_{x_{\text{пр}}}} \cdot t_{p;k}. \qquad (1.33)$$

Побудована довірча область для $M_{x_{\text{пр}}}(\text{y})$ умовного математичного очікування, але не окремих можливих прогнозних значень залежної змінної, які відхиляються від середньої. Тому при визначенні довірчого інтервалу для індивідуальних значень $y^*_{x_{\text{пр}}}$ залежної змінної необхідно враховувати іще одне джерело дисперсії – розсіювання навколо лінії регресії. Це реалізується шляхом включення в оцінку дисперсії $s^2_{\hat{y}_{x_{\text{пр}}}}$ величини s^2. У результаті оцінка дисперсії індивідуальних значень $y^*_{x_{\text{пр}}}$ дорівнює:

$$s^2_{y^*_{x_{\text{пр}}}} = s^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right). \qquad (1.34)$$

Тоді стандартна помилка індивідуальних значень становить:

$$s_{y^*_{x_{\text{пр}}}} = \sqrt{s^2_{y^*_{x_{\text{пр}}}}}. \qquad (1.35)$$

Для прогнозів індивідуальних значень p %-й довірчий інтервал $y^*_{x_{\text{пр}}}$ буде становити:

$$\hat{y}_{x_{\text{пр}}} - s_{y^*_{x_{\text{пр}}}} \cdot t_{p;k} \le y^*_{x_{\text{пр}}} \le \hat{y}_{x_{\text{пр}}} + s_{y^*_{x_{\text{пр}}}} \cdot t_{p;k}. \qquad (1.36)$$

Чим далі $x_{\text{пр}}$ відхиляється від середнього за вибіркою значення незалежної змінної $\bar{x}$, тим менш точним буде прогноз і ширшими довірчі інтервали для умовного математичного очікування $M_{x_{\text{пр}}}(\text{y})$ та індивідуальних значень $y^*_{x_{\text{пр}}}$. Тому *екстраполяція* регресії, тобто її використання поза межами дослідженого діапазону значень незалежної змінної, може призвести до значної похибки. Навіть у тому випадку, якщо така екстраполяція виправдана для розглянутої змінної, виходячи із економічного змісту задачі, що розв'язується.

При прогнозуванні розвитку економічних явищ чи процесів на основі економетричних моделей взагалі та парних лінійних регресійних моделей зокрема потрібно пам'ятати про складність та імовірнісний характер цих явищ та процесів. Тому отримані прогнози рекомендується використовувати лише як важливий допоміжний матеріал при прийнятті управлінського рішення, пов'язаного із прогнозом розвитку економічного явища чи процесу.

На економічні та соціально-економічні явища та процеси сукупно та одночасно діє значна кількість факторів різної природи, а модель парної регресії враховує лише один найбільш важливий, на думку дослідника, фактор. Більш ефективним інструментом опису вказаних вище явищ та процесів можуть служити множинні регресійні моделі, що одночасно враховують декілька факторів.

2. ПРИКЛАД ВИКОНАННЯ РОЗРАХУНКОВО-ГРАФІЧНОГО ЗАВДАННЯ

Досліджується залежність витрат підприємств (тис. грн) від кількості виробленої продукції (т) на основі статистичних даних (табл. 2.1), отриманих по 12-ти підприємствах однієї галузі.

У вихідних даних присутні лише дві величини, тому будемо будувати модель парної регресії. Зазвичай витрати підприємства визначаються в основному кількістю виробленої продукції, причому витрати поділяються на умовно-постійні (такі, що не пов'язані безпосередньо із виробництвом) та умовно-змінні (безпосередньо пов'язані із виробництвом). Таким чином незалежною змінною буде кількість виробленої продукції, а залежною – витрати підприємства.

Таблиця 2.1 – Статистичні дані кількості виробленої продукції та витрат підприємств однієї галузі

Номер спостереження	Кількість виробленої продукції, т	Витрати підприємств, тис. грн
1	2,3	165
2	4,2	122
3	4,9	250
4	6,1	167
5	7	220
6	9,3	275
7	11,5	150
8	12,6	330
9	17	270
10	18	285
11	18,7	384
12	23	344

Побудуємо допоміжну таблицю, яка спростить подальші розрахунки.

Таблиця 2.2 – Допоміжна розрахункова таблиця

№	x_i	y_i	$x_i{\cdot}y_i$	x_i^2	y_i^2
1	2,3	165	379,5	5,29	27225
2	4,2	122	512,4	17,64	14884
3	4,9	250	1225	24,01	62500
4	6,1	167	1018,7	37,21	27889
5	7	220	1540	49	48400
6	9,3	275	2557,5	86,49	75625
7	11,5	150	1725	132,25	22500
8	12,6	265	3339	158,76	70225
9	17	270	4590	289	72900
10	18	285	5130	324	81225
11	18,7	384	7180,8	349,69	147456
12	23	344	7912	529	118336
Сума	134,6	2897	37109,9	2002,34	769165
Середнє	11,217	241,417	3092,492	166,862	64097,1

Парний коефіцієнт кореляції розраховується за формулою:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{(\overline{x^2} - (\overline{x})^2) \cdot (\overline{y^2} - (\overline{y})^2)}},$$

де $\overline{x}$, $\overline{y}$ – середнє значення незалежної та залежної змінних відповідно; $\overline{xy}$ – середнє значення добутку незалежної та залежної змінних; $\overline{x^2}$, $\overline{y^2}$ – середнє значення квадратів незалежної та залежної змінних відповідно.

Зазначені величини візьмемо із табл. 2.2 та підставимо у формулу для парного коефіцієнта кореляції:

$$r = \frac{3092{,}492 - 11{,}217 \cdot 241{,}417}{\sqrt{(166{,}862 - 11{,}217^2) \cdot (64097{,}1 - 241{,}417^2)}};$$

$$r = 0{,}787.$$

Фактичне значення критерію Ст`юдента для парного коефіцієнта кореляції:

$$t_{\text{факт}} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}; \qquad t_{\text{факт}} = \frac{0{,}787 \cdot \sqrt{12-2}}{\sqrt{1-0{,}787^2}}; \qquad t_{\text{факт}} = 4{,}034.$$

Табличне значення визначене для степенів свободи $n-2=10$ та вірогідності 0,9 дорівнює $t_{0,9;10} = 1{,}812$.

Оскільки фактичне значення критерію Ст`юдента більше ніж табличне $|t_{\text{факт}}| > t_{0,9;10}$ (4,034>1,812), то парний коефіцієнт кореляції статистично значуще відрізняється від 0 при вірогідності 0,9.

Парний коефіцієнт кореляції між витратами підприємств та кількістю виробленої ними продукції дорівнює $r = 0{,}787$. Це свідчить про те, що між цими економічними величинами існує тісний лінійний зв'язок, причому цей зв'язок прямий.

У цьому випадку потрібно будувати модель парної лінійної регресії залежності витрат підприємства від кількості виробленої продукції.

Припустимо, що вихідні статистичні дані відповідають вимогам теореми Гаусса-Маркова, тому параметри регресії оцінимо згідно з МНК за формулами:

$$\beta_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}; \qquad \beta_1 = \frac{3092{,}492 - 11{,}217 \cdot 241{,}417}{166{,}862 - 11{,}217^2}; \qquad \beta_1 = 9{,}369.$$

$$\beta_0 = \bar{y} - \beta_1 \cdot \bar{x}; \qquad \beta_0 = 241{,}417 - 9{,}369 \cdot 11{,}217; \qquad \beta_0 = 136{,}322.$$

Модель парної лінійної регресійної залежності витрат підприємств від кількості виробленої ними продукції має вигляд:

$$y = 136{,}322 + 9{,}369 \cdot x + \varepsilon.$$

Економічним змістом параметра β_0 будуть умовно-постійні витрати підприємства за рівнянням регресії без урахування випадкових та неврахованих у моделі чинників. Таким чином у нашому випадку приблизні умовно-постійні витрати будуть становити $\beta_0 = 136{,}322$ тис. грн.

Економічним змістом параметра β_1 будуть умовно-змінні витрати підприємства на одиницю виробленої продукції без урахування випадкових та неврахованих у моделі чинників. Таким чином у нашому випадку приблизні умовно-змінні витрати на одиницю виробленої продукції будуть становити $\beta_1 = 9{,}369$ тис. грн.

Кореляційне поле, що відповідає статистичним даним (табл. 2.1), та лінія регресії, що його описує, наведені на рис. 2.1.

Стандартну помилку регресії знайдемо за формулою:

$$s = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i}{n-2}}.$$

Для зручності знаходження стандартної помилки регресії та подальших розрахунків обчислимо ще одну допоміжну таблицю для рівняння регресії (табл. 2.3).

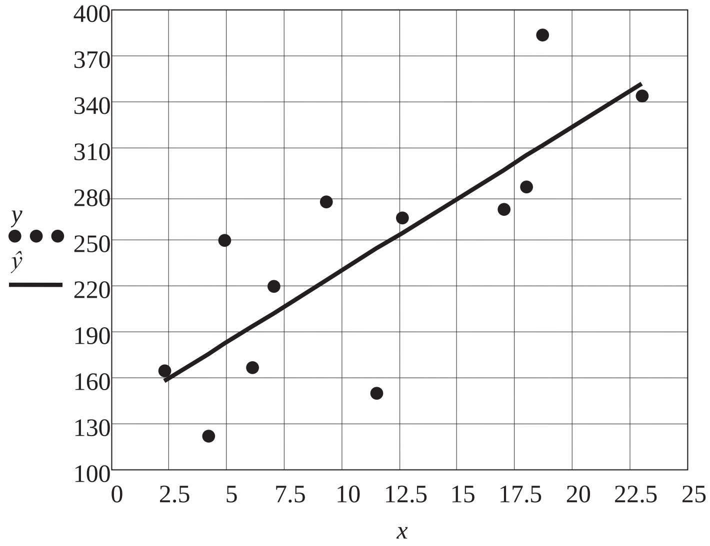


Рисунок 2.1 – Кореляційне поле та лінія регресії, що його описує

Тоді стандартна помилка регресії буде дорівнювати:

$$s = \sqrt{\frac{26538{,}416}{12-2}}; \qquad s = 51{,}515.$$

Стандартна помилка параметра β_0 визначається, як:

$$m_{\beta_0} = s \cdot \sqrt{\frac{\overline{x^2}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \qquad m_{\beta_0} = 51{,}515 \cdot \sqrt{\frac{166{,}862}{492{,}577}}; \qquad m_{\beta_0} = 29{,}983.$$

Стандартна помилка параметра β_1 визначається, як:

$$m_{\beta_1} = s \cdot \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}; \qquad m_{\beta_1} = 51{,}515 \cdot \sqrt{\frac{1}{492{,}577}}; \qquad m_{\beta_1} = 2{,}321.$$

Таблиця 2.3 – Допоміжна розрахункова таблиця для рівняння регресії

№	x_i	y_i	$\hat{y}_i$	ε_i	ε_i^2	$(x_i - \bar{x})^2$
1	2,3	165	157,872	7,128	50,814	79,507
2	4,2	122	175,674	-53,674	2880,869	49,234
3	4,9	250	182,232	67,768	4592,447	39,900
4	6,1	167	193,476	-26,476	700,971	26,180
5	7,0	220	201,908	18,092	327,305	17,780
6	9,3	275	223,458	51,542	2656,538	3,674
7	11,5	150	244,071	-94,071	8849,423	0,080
8	12,6	265	254,378	10,622	112,830	1,914
9	17,0	270	295,604	-25,604	655,557	33,447
10	18,0	285	304,973	-19,973	398,936	46,014
11	18,7	384	311,532	72,468	5251,603	56,000
12	23,0	344	351,821	-7,821	61,169	138,847
Сума					26538,461	492,577

Фактичні значення критерію Ст`юдента для перевірки статистичної значущості параметрів регресії дорівнюють:

$$t_{\beta_0} = \frac{\beta_0}{m_{\beta_0}}; \qquad t_{\beta_0} = \frac{136{,}322}{29{,}983}; \qquad t_{\beta_0} = 4{,}55;$$

$$t_{\beta_1} = \frac{\beta_1}{m_{\beta_1}}; \qquad t_{\beta_1} = \frac{9{,}369}{2{,}321}; \qquad t_{\beta_1} = 4{,}037.$$

Оскільки фактичне значення критерію Ст`юдента для параметра β_0 більше ніж табличне значення, обране для $(n-2)$ степенів свободи при вірогідності 0,9 $t_{\beta_0} > t_{0,9;10}$ (4,55>1,812), то параметр β_0 статистично значуще відрізняється від 0 при цій вірогідності.

Оскільки фактичне значення критерію Ст`юдента для параметра β_1 більше ніж табличне значення, обране для $(n-2)$ степенів свободи при вірогідності 0,9 $t_{\beta_1} > t_{0,9;10}$ (4,037>1,812), то параметр β_1 статистично значуще відрізняється від 0 при цій вірогідності.

Довірчі інтервали для параметрів регресії при вірогідності 0,9 будуть становити:

$$(\beta_0 - m_{\beta_0} \cdot t_{0,9;10}) \le \beta_0 \le (\beta_0 + m_{\beta_0} \cdot t_{0,9;10});$$

$$(136{,}322 - 29{,}983 \cdot 1{,}812) \le \beta_0 \le (136{,}322 + 29{,}983 \cdot 1{,}812);$$

$$82 \le \beta_0 \le 190{,}651;$$

$$(\beta_1 - m_{\beta_1} \cdot t_{0,9;10}) \le \beta_1 \le (\beta_1 + m_{\beta_1} \cdot t_{0,9;10});$$

$$(9{,}369 - 2{,}321 \cdot 1{,}812) \le \beta_1 \le (9{,}369 + 2{,}321 \cdot 1{,}812);$$

$$5{,}163 \le \beta_1 \le 13{,}575.$$

Якість побудованої моделі оцінимо за коефіцієнтом детермінації:

$$R^2 = r^2; \qquad R^2 = 0{,}787^2; \qquad R^2 = 0{,}62.$$

Фактичне значення критерію Фішера для R^2 дорівнює:

$$F_{\text{факт}} = \frac{R^2 \cdot (n-2)}{1-R^2}; \qquad F_{\text{факт}} = \frac{0{,}62 \cdot (12-2)}{1-0{,}62}; \qquad F_{\text{факт}} = 16{,}3.$$

Табличне значення для рівня значущості $\alpha = 0{,}1$ степенів свободи $k_1 = 1$, $k_2 = (n-2) = 10$ складає $F_{0,1;1;10} = 3{,}285$.

Проведені вище розрахунки дозволяють зробити висновки про те, що модель відносно якісна, оскільки $R^2 = 0{,}62$. Коефіцієнт детермінації дорівнює $0{,}62$, отже, 62 % варіації залежної змінної пояснюється отриманим рівнянням регресії, а 48 % припадають на випадкові та невраховані в моделі фактори.

Оскільки $F_{\text{факт}} > F_{0,1;1;10}$ (16,3>3,285), коефіцієнт детермінації статистично значущий при $\alpha = 0{,}1$.

Визначимо точкову оцінку умовного математичного очікування $M_x(y)$ для $x_{пр}$=15 т. Вибірковою оцінкою $M_{x=15}(y)$ є групова середня $\hat{y}_{x=15}$, яку знайдемо за рівнянням регресії:

$$\hat{y}_{x=15}=136{,}322+9{,}369\cdot x_{пр}; \qquad \hat{y}_{x=15}=276{,}856 \text{ тис. грн.}$$

Оцінка дисперсії групової середньої:

$$s^2_{\hat{y}_{x=15}} = s^2\cdot\left(\frac{1}{n}+\frac{(x_{np}-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right);$$

$$s^2_{\hat{y}_{x=15}} = 51{,}515^2\cdot\left(\frac{1}{12}+\frac{(15-11{,}217)^2}{492{,}577}\right); \qquad s^2_{\hat{y}_{x=15}}=298{,}252.$$

Стандартна помилка групової середньої:

$$s_{\hat{y}_{x=15}}=\sqrt{s^2_{\hat{y}=15}}; \qquad s_{\hat{y}_{x=15}}=\sqrt{298{,}252}; \qquad s_{\hat{y}_{x=15}}=17{,}27.$$

Тоді 90 %-й довірчий інтервал для умовного математичного очікування становить:

$$\hat{y}_{x=15}-s_{\hat{y}_{x=15}}\cdot t_{0,9;10}\le M_{x=15}(y)\le \hat{y}_{x=15}+s_{\hat{y}_{x=15}}\cdot t_{0,9;10};$$

$$276{,}856-17{,}27\cdot 1{,}812\le M_{x=15}(y)\le 276{,}856+17{,}27\cdot 1{,}812;$$

$$245{,}563\le M_{x=15}(y)\le 308{,}149.$$

Побудована довірча область для $M_{x=15}(y)$ умовного математичного очікування, але не окремих можливих значень залежної змінної, які відхиляються від середньої. Тому при визначенні довірчого інтервалу для індивідуальних значень $y^*_{x_0}$ залежної змінної необхідно враховувати ще одне

джерело варіації – розсіювання навколо лінії регресії. Це реалізується шляхом включення в оцінку дисперсії $s^2_{\hat{y}}$ величини s^2. У результаті оцінка дисперсії індивідуальних значень $y^*_{x_0}$ дорівнює:

$$s^2_{y_{x_0}}=s^2\cdot\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right).$$

Для нашого випадку:

$$s^2_{y_{x_0=15}}=51{,}515^2\cdot\left(1+\frac{1}{12}+\frac{(15-11{,}217)^2}{492{,}577}\right); \qquad s^2_{y_{x_0=15}}=2952.$$

Тоді стандартна помилка індивідуального значення $y_{x_0=15}$ становить:

$$s_{y_{x_0=15}}=\sqrt{s^2_{y_{x_0=15}}}; \qquad s_{y_{x_0=15}}=\sqrt{2952}; \qquad s_{y_{x_0=15}}=54{,}332.$$

Для прогнозів індивідуальних значень $y^*_{x_0}$ 90 %-й довірчий інтервал буде становити:

$$\hat{y}_{x=15}-s_{y_{x_0=15}}\cdot t_{0,9;10}\le y^*_{x_0}\le \hat{y}_{x=15}+s_{y_{x_0=15}}\cdot t_{0,9;10};$$

$$276{,}856-54{,}332\cdot 1{,}812\le y^*_{x_0}\le 276{,}856+54{,}332\cdot 1{,}812;$$

$$178{,}406\le y^*_{x_0}\le 375{,}306.$$

У результаті виконання цього розрахунково-графічного завдання була побудована модель парної лінійної регресійної залежності витрат підприємств від кількості виробленої продукції, оцінена якість рівняння регресії та здійснено прогнозування на основі отриманої моделі.

3. ПРОВЕДЕННЯ СТАТИСТИЧНИХ РОЗРАХУНКІВ У MICROSOFT EXCEL 2010

Для проведення статистичних розрахунків у Microsoft Excel 2010 перш за все внесемо вихідні дані на робочий лист (рис. 3.1). Парний коефіцієнт кореляції розраховується за допомогою функції КОРРЕЛ. Вона є достатньо простою у використанні та у якості аргументів вимагає введення двох векторів статистичних даних, розмірність яких має бути однакова (рис. 3.1).

Фактичне значення критерію Ст`юдента для перевірки статистичної значущості парного коефіцієнта кореляції знайдемо за формулою (8.1). Реалізація цих розрахунків у Microsoft Excel показана на рис. 3.2.

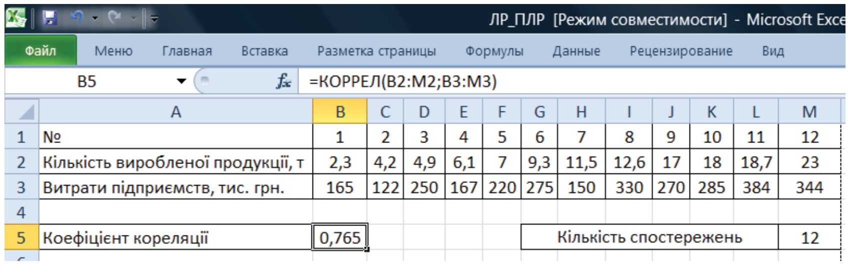


Рисунок 3.1 – Визначення парного коефіцієнта кореляції

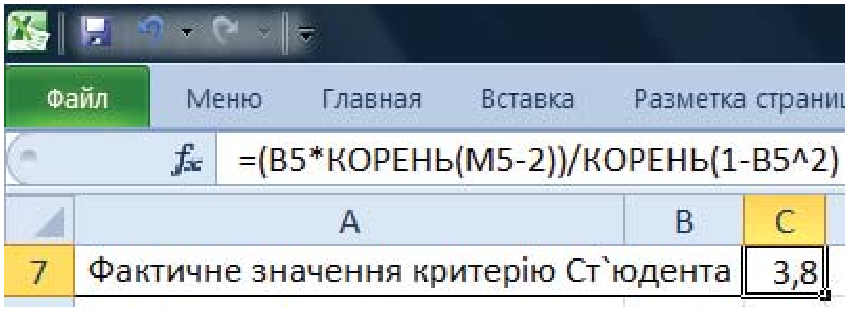


Рисунок 3.2 – Визначення фактичного значення критерію Ст`юдента

Критичне значення двостороннього розподілення Ст`юдента знайдемо за допомогою статистичної функції СТЬЮДЕНТ.ОБР.2Х (рис. 3.3). Як аргументи цієї функції використовуються рівень значущості α та степені свободи $(n-2)$.

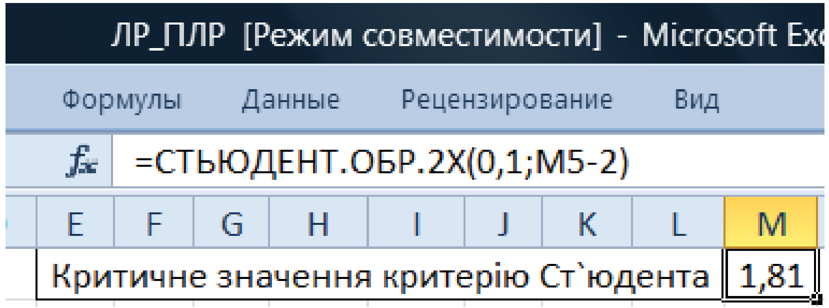


Рисунок 3.3 – Визначення фактичного значення критерію Ст`юдента

При використанні цієї функції необхідно зауважити, що у рядку «Вірогідність» необхідно вказувати не вірогідність p, а рівень значущості α. Перехід від рівня значущості α до вірогідності p здійснюється за формулою:

$$\alpha = 1 - p. \tag{3.1}$$

Аналогом функції СТЬЮДЕНТ.ОБР.2Х у Microsoft Excel 2007 та більш ранніх версіях є СТЬЮДРАСПОБР.

Параметри лінійної регресії методом найменших квадратів дозволяє отримати функція ЛИНЕЙН (категорія статистичні), також ця функція може видавати додаткову інформацію, що характеризує якість отриманої моделі. Розглянемо порядок використання функції ЛИНЕЙН на базі схеми діалогового вікна (рис. 3.4).

У рядку «Известные_значения_y» вводиться діапазон комірок із вибіркою спостережень залежної змінної, а в рядку «Известные_значения_x» – діапазон комірок із вибіркою незалежної змінної. При використанні зазначеної функції вибірки, для яких будується та досліджується модель, можуть бути записані, як у стовпчики, так і у рядки. Причому, якщо вибірка спостережень залежної змінної записана в рядок (стовпчик), то і вибірка незалежної змінної має бути записана в рядок (стовпчик), інакше функція повертає значення помилки. Очевидно, що для використання функції ЛИНЕЙН кількість спостережень у кожній із вибірок даних має співпадати.

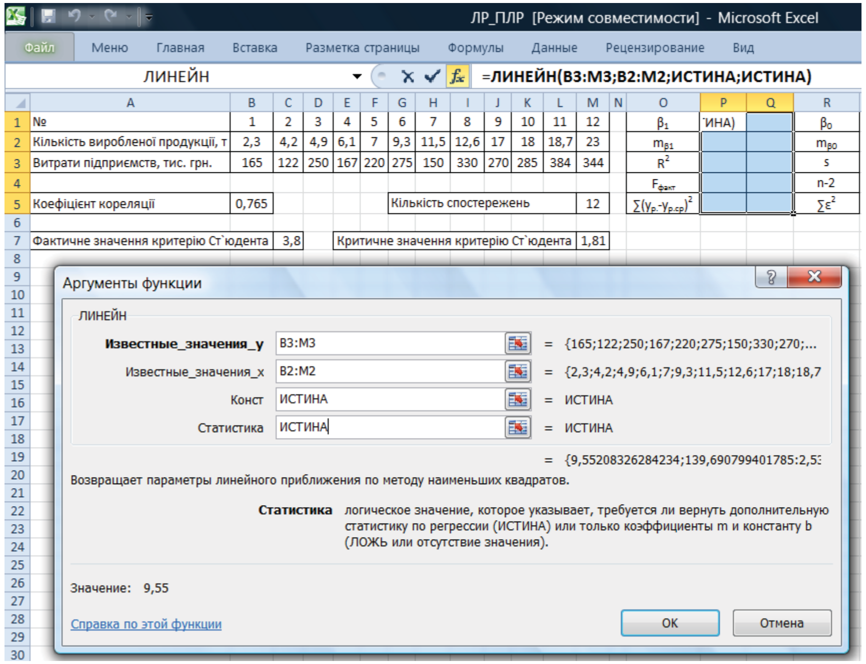


Рисунок 3.4 – Схема діалогового вікна функції ЛИНЕЙН

Якщо в рядку «Конст» записано «ИСТИНА», чи нічого не вказано, функція повертає обидва параметри моделі парної лінійної регресії β_0 та β_1. Якщо вказано «ЛОЖЬ» – повертає лише один параметр β_1, при цьому передбачається, що $\beta_0 = 0$, а модель будується у вигляді $y = \beta_1 \cdot x + \varepsilon$.

Якщо в рядку «Статистика» записано «ИСТИНА», то функція повертає додаткову інформацію, що характеризує якість моделі. Якщо в цьому рядку нічого не вказано, чи записано «ЛОЖЬ», функція не повертає такої інформації.

Найбільш повне використання можливостей функції ЛИНЕЙН для парної лінійної регресії дозволяє отримати таблицю із такими величинами (табл. 3.1).

Таблиця 3.1 – Схематичне відображення результатів використання функції ЛИНЕЙН

Параметр β_1	Параметр β_0
Стандартна помилка m_{β_1} параметра β_1	Стандартна помилка m_{β_0} параметра β_0
Коефіцієнт детермінації R^2	Стандартна помилка регресії s
Фактичне значення критерію Фішера $F_{факт}$	Число степенів свободи $k = n - 2$
Регресійна сума квадратів, $\sum_{i=1}^{n}(\hat{y}_i - \hat{y}_{ср})$, де $\hat{y}_i = \beta_0 + \beta_1 \cdot x_i$, $\hat{y}_{cp.} = \frac{\sum_{i=1}^{n} \hat{y}_i}{n}$	Сума квадратів залишків $\sum_{i=1}^{n} \varepsilon^2$

Якщо при використанні ЛИНЕЙН одночасно заповнюється декілька комірок, то перед запуском зазначеної функції потрібно виділити відповідний діапазон комірок, в який функція поверне результат розрахунків. Після введення аргументів функції ЛИНЕЙН клавішу Enter чи значок OK можна натискати лише при одночасно натиснутих клавішах Shift та Ctrl.

Фактичні значення критеріїв Ст`юдента для перевірки статистичної значущості параметрів регресії отримаємо за формулою (1.23), (рис. 3.5).

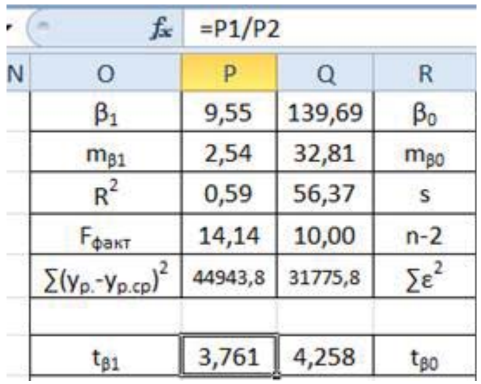


Рисунок 3.5 – Визначення фактичних значень критеріїв Ст`юдента для перевірки статистичної значущості параметрів регресії

Побудова довірчих інтервалів параметрів регресії здійснюється за формулою (1.25), що реалізується в Microsoft Excel, як показано на рис. 3.6.

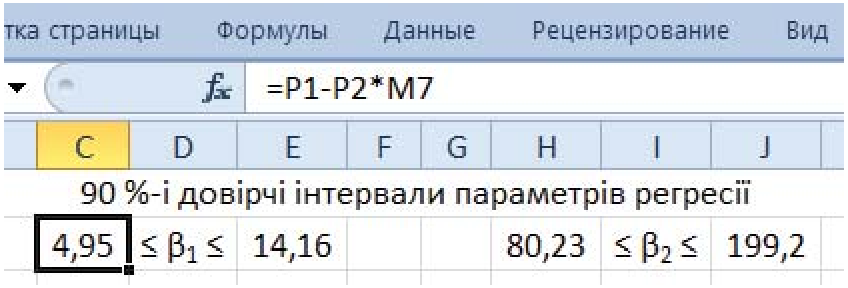


Рисунок 3.6 – Побудова довірчих інтервалів параметрів регресії

Для перевірки статистичної значущості коефіцієнта детермінації фактичне значення критерію Фішера потрібно порівняти із критичним значенням *F*-критерію Фішера, яке можна отримати за допомогою статистичної функції F.ОБР.ПХ (рис. 3.7). Аргументом рядка «Вірогідність» є рівень значущості α. Степені свободи 1 розраховуються як $k_1 = m - 1$ (m – кількість змінних в моделі). Тобто для парної лінійної регресії $k_1 = 1$. Степені свободи 2 розраховуються як $k_2 = n - 1$.

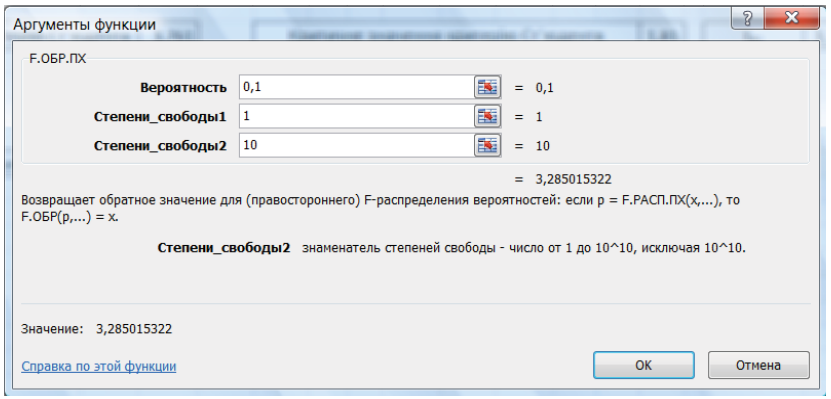


Рисунок 3.7 – Діалогове вікно функції F.ОБР.ПХ

Для прогнозування на базі моделі парної лінійної регресії перш за все внесемо в електронну таблицю Microsoft Excel прогнозне значення незалежної змінної (рис. 3.8). У цьому випадку це кількість виробленої продукції в обсязі 15 т.

Для виконання подальших розрахунків знайдемо середнє значення незалежної змінної за допомогою функції СРЗНАЧ (рис. 3.8). Точкову оцінку умовного математичного очікування визначимо шляхом підстановки прогнозного значення незалежної змінної в рівняння регресії (рис. 3.9).

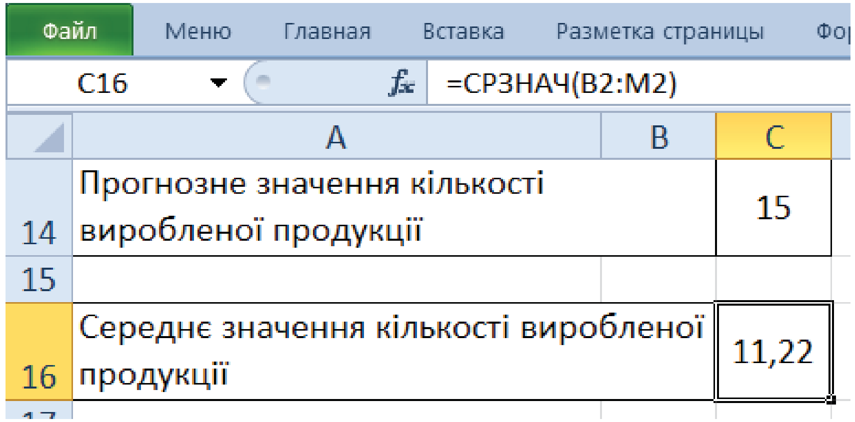


Рисунок 3.8 – Прогнозне та середнє значення кількості виробленої продукції

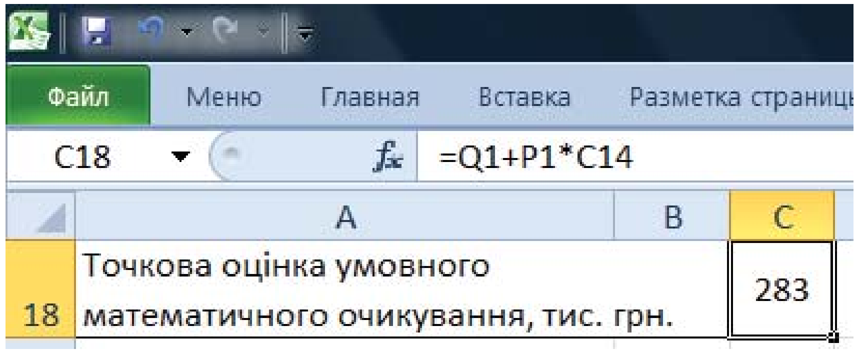


Рисунок 3.9 – Точкова оцінки умовного математичного очікування

Точкового прогнозу, як правило, недостатньо, тому він доповнюється інтервальним прогнозом (1.33) для умовного математичного очікування (рис. 3.10). Для побудови вказаного інтервалу знайдемо стандартну помилку групової середньої (1.32), для чого визначимо величину $\sum_{i=1}^{n}(x_i-\bar{x})^2$. Для цього в першій комірці таблиці (рис. 3.10) від першого статистичного значення незалежної змінної віднімемо середнє, причому посилання на комірку із середнім значенням має бути абсолютним.

Рисунок 3.10 – Визначення величини $\sum_{i=1}^{n}(x_i-\bar{x})^2$

Далі за допомогою функції СУММКВ (категорія математичні) знайдемо суму квадратів різниць $(x_i-\bar{x})$ (рис. 3.10, комірка R12).

Реалізація розрахунку стандартної помилки групової середньої показана на рис. 3.11.

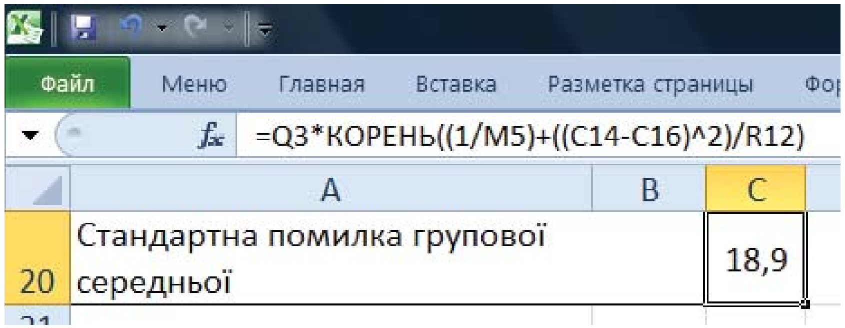


Рисунок 3.11 – Розрахунок стандартної помилки групової середньої

На рис. 3.12 наведено приклад побудови довірчого інтервалу для умовного математичного очікування.

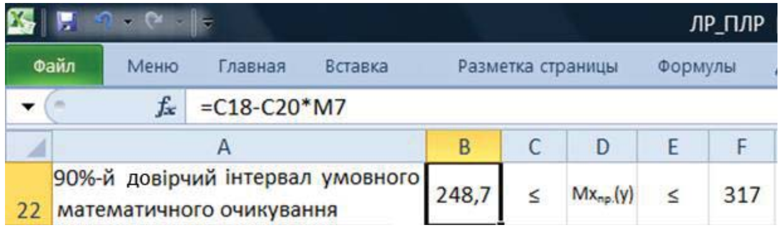


Рисунок 3.12 – Довірчий інтервал для умовного математичного очікування

Нагадаємо, що знаки «≤» в MS Excel вводяться за допомогою закладок «Вставка» → «Символ».

Принцип розрахунку стандартної помилки (1.34), (1.35) та побудови довірчого інтервалу індивідуальних значень (1.36) аналогічний цим розрахункам для умовного математичного очікування та наведений на рис. 3.13 та рис. 3.14 відповідно.

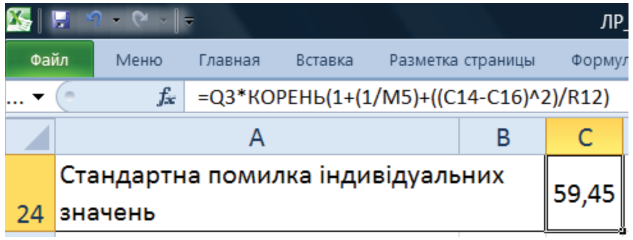


Рисунок 3.13 – Розрахунок стандартної помилки індивідуальних значень $y^*_{x_0}$

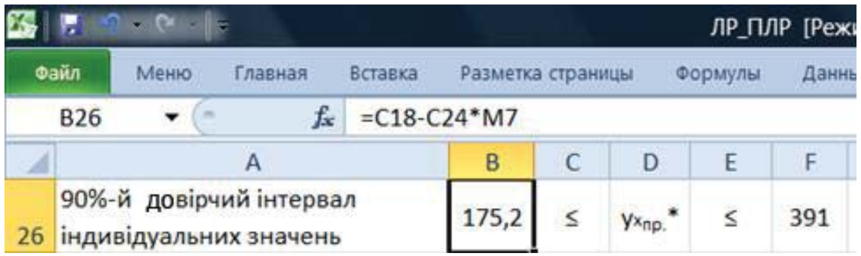


Рисунок 3.14 – Довірчий інтервал для індивідуальних значень $y^*_{x_0}$

Провівши вказані вище розрахунки в MS Excel, висновки за результатами аналізу зробимо як і в розд. 2.

Останнім етапом проведення парного кореляційно-регресійного аналізу засобами MS Excel буде побудова геометричної інтерпретації рівняння парної лінійної регресії – лінії регресії.

Для цього необхідно визначити діапазон по осі абсцис, в якому будується графік. За допомогою статистичної функції МИН знайдемо найменше число із вибірки незалежної змінної $x_{\min}$, а за допомогою функції МАКС – найбільше $x_{\max}$ (рис. 3.15).

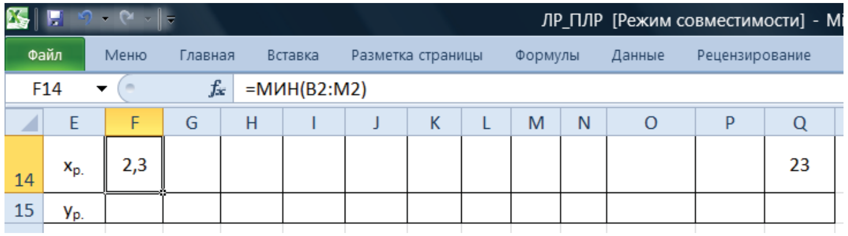


Рисунок 3.15 – Визначення найменшого $x_{\min}$ та найбільшого $x_{\max}$ значення із вибірки незалежної змінної

Далі розіб'ємо отриманий діапазон на рівні частини за формулою:

$$x_i = x_{(i-1)} + \frac{x_{\max} - x_{\min}}{n-1}. \quad (3.1)$$

При реалізації розрахунків за формулою (3.1) потрібно правильно обирати тип посилання на комірки (відносне чи абсолютне), рис. 3.16.

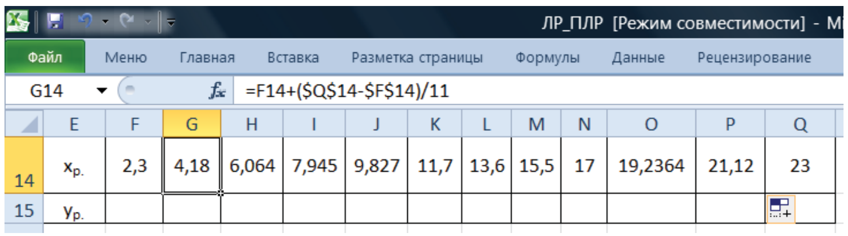


Рисунок 3.16 – Поділ діапазону незалежної змінної на рівні частини

Для отриманого та розбитого на рівні частини діапазону незалежної змінної одержимо відповідні значення залежної за рівнянням парної лінійної регресії (рис. 3.17).

Рисунок 3.17 – Визначення значень залежної змінної для лінії регресії

Після підготовки даних перейдемо до безпосередньої побудови графіка. Для цього виділимо колонку розрахованих значень залежної змінної, оберемо закладку «Вставки», графік та визначимо потрібний нам тип графіка (рис. 3.18).

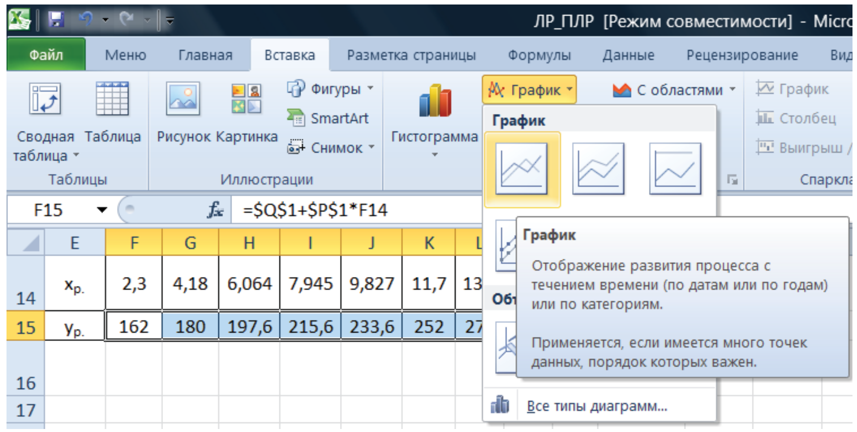


Рисунок 3.18 – Перший етап побудови двовимірного графіка в MS Excel 2010

Для того щоб змінити значення в отриманому графіку по осі абсцис, потрібно натиснути на графік мишею та обрати «Вибрати дані» (рис. 3.19).

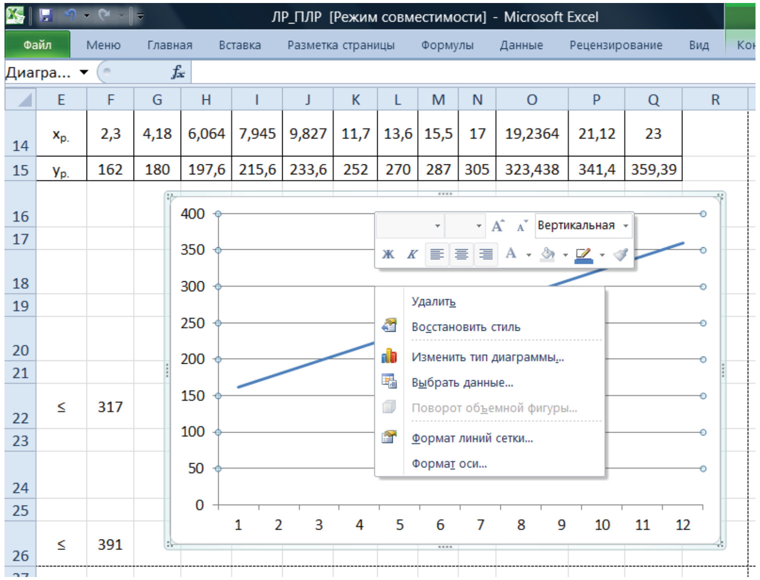


Рисунок 3.19 – Вибір даних незалежної змінної для побудови графіка

У діалоговому вікні (рис. 3.20), що з'явилося, необхідно змінити підписи горизонтальної осі (категорій). Для цього в рядок «Діапазон підписів осей» потрібно ввести розрахункові значення незалежної змінної (рис. 3.21).

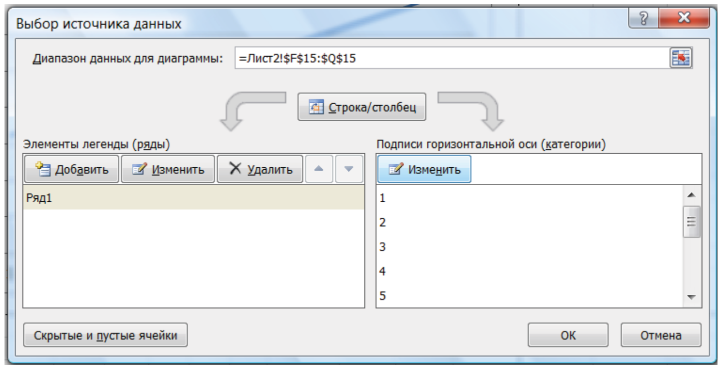


Рисунок 3.20 – Діалогове вікно «Вибір джерела даних»

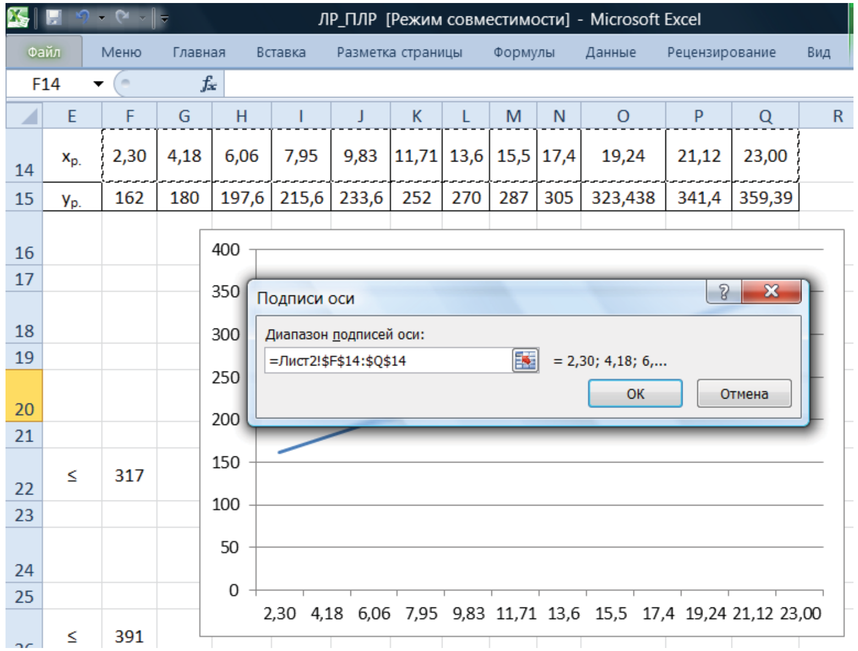


Рисунок 3.21 – Зміна підписів графіка по вісі абсцис

У результаті виконаної роботи отримаємо потрібний графік лінії регресії (рис. 3.21).

4. ВАРІАНТИ ДЛЯ ВИКОНАННЯ РОЗРАХУНКОВО-ГРАФІЧНОГО ЗАВДАННЯ

У цьому РГЗ необхідно провести парний кореляційно-регресійний аналіз залежності певних економічних величин, наведених за варіантами. Етапами виконання завдання будуть такі:

1. Провести специфікацію моделі, визначитися із залежною та незалежною змінними.
2. Визначити тісноту лінійного зв'язку між економічними величинами, наведеними за варіантами. Оцінити статистичну значущість та зробити висновки щодо парного коефіцієнта кореляції.

3. Оцінити параметри парної лінійної регресії за МНК, побудувати модель. Дати економічну інтерпретацію параметрів регресії.

4. Побудувати кореляційне поле за статистичними даними та лінію регресії, що його описує.

5. Перевірити статистичну значущість та побудувати довірчі інтервали оцінок параметрів регресії.

6. Розрахувати коефіцієнт детермінації, перевірити його статистичну значущість та оцінити якість побудованої регресійної моделі.

7. Здійснити прогнозування на базі отриманої моделі для прогнозного значення незалежної змінної $x_{пр}$, наведеного за варіантами.

Нижче наведені двадцять варіантів для виконання РГЗ. Номер варіанта відповідає номеру прізвища студента у журналі. Якщо в групі більше ніж 20 студентів, то студент, прізвище якого знаходиться під номером 21, виконує 1-й варіант, під номером 22 – 2-й варіант і т.д.

Варіант 1

Дослідити залежність споживання підприємствами однієї галузі електроенергії y від кількості виробленої продукції x.

Таблиця 4.1 – Статистичні дані споживання електроенергії (МВт/год) та кількості виробленої продукції (т)

x	4	4,7	5,5	6,8	8	11	12,5	14	19	20,2	21	25,5
y	7,42	5,5	11,3	7,5	10	12,4	6,75	12	12,2	12,8	17,3	15,5

Прогнозне значення кількості виробленої продукції $x_{пр}$ = 17 т.

Варіант 2

Дослідити залежність кількості бракованої продукції y від кількості робітників x зі спеціальною підготовкою за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.2 – Статистичні дані кількості бракованої продукції (%) та кількості робітників (%) зі спеціальною підготовкою

x	18,2	21,8	28,7	35,4	39,5	50,8	57,7	62,5	71,8	78,3	86,6	92,6
y	18	22,5	12,7	11,7	14,5	12,8	11	12,8	8,7	7,1	4	4

Прогнозне значення кількості робітників зі спеціальною підготовкою $x_{пр}$ = 60 %.

Варіант 3

Дослідити залежність обсягів продажів товару y від витрат на його рекламу x за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.3 – Статистичні дані обсягів продажів товару (тис. грн) від витрат на його рекламу (тис. грн)

x	12,5	18,8	19,2	22	23,4	29,2	28,3	28	35,2	27	35,7	43
y	520	560	600	620	650	680	720	730	740	800	850	800

Прогнозне значення витрат на рекламу $x_{пр}$ = 30 тис. грн.

Варіант 4

Дослідити залежність обсягів продажів товару y від його ціни x за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.4 – Статистичні дані обсягів продажів (тис. грн) товару від його ціни (грн)

x	96,3	90	70	76,2	62,5	69	54,7	42	40,5	42,5	38,7	24
y	110	105	112	280	235	480	280	327	596	549	570	480

Прогнозне значення ціни на товар $x_{пр}$ = 62 тис. грн.

Варіант 5

Дослідити залежність обсягів продажів підприємства y від витрат на розширення дилерської мережі x за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.5 – Статистичні дані обсягів продажів підприємства (тис. грн) від витрат на розширення дилерської мережі (тис. грн)

x	14	20,2	19,2	25	23,4	29,2	30,2	30	35,2	27	40	45
y	600	700	750	640	680	700	720	732	695	796	850	864

Прогнозні витрати на розширення дилерської мережі $x_{пр}$ = 25 тис. грн.

Варіант 6

Дослідити залежність прибутку підприємства *y* від кількості бракованої продукції *x* за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.6 – Статистичні дані прибутку підприємства (тис. грн) та кількості бракованої продукції (%)

x	20	22,5	12,7	11,7	14,5	15	18,5	12,8	8,7	7,1	4	7
y	180	218	290	355	400	510	600	625	718	783	866	926

Прогнозне значення кількості бракованої продукції $x_{пр}$ = 14 %.

Варіант 7

Дослідити залежність споживання води *y* від кількості виробленої продукції *x* дванадцятьма підприємствами однієї галузі.

Таблиця 4.7 – Статистичні дані споживання води (м^3) підприємствами та кількості виробленої ними продукції (т)

x	4	4,7	5,5	6,8	8	11	12,5	14	19	20,2	21	25,5
y	105	98	150	90	200	300	253	310	400	280	375	350

Прогнозне значення кількості виробленої продукції $x_{пр}$ = 10 т.

Варіант 8

Дослідити залежність кількості бракованої продукції *y* від інвестицій у модернізацію обладнання *x* за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.8 – Статистичні дані кількості бракованої продукції (%) та інвестицій у модернізацію обладнання (тис. грн)

x	190	220	287	355	400	510	577	652	720	782	865	1000
y	18	22,5	10	11,7	15	15	11	12,8	8,7	7,1	3,2	5

Прогнозні інвестицій у модернізацію обладнання $x_{пр}$ = 600 тис. грн.

Варіант 9

Дослідити залежність витрат на логістичне обслуговування *y* від рівня обслуговування *x* за даними дванадцяти підприємств.

Таблиця 4.9 – Статистичні дані витрат на логістичне обслуговування (тис. грн) від рівня обслуговування (%)

x	12	15	17	20	23	26	30	32	35	37	40	42
y	20	50	35	60	60	85	68	76	100	80	80	120

Прогнозне значення рівня обслуговування $x_{пр}$ = 25 %.

Варіант 10

Дослідити залежність втрат, викликаних погіршенням обслуговування *y* від рівня обслуговування *x* за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.10 – Статистичні дані втрат, викликаних погіршенням обслуговування (тис. грн) від рівня обслуговування (%)

x	27	29	30,5	33	35	39	41	42,5	44	44,5	48	50
y	30	35	30,7	35	42	40	49,7	43	52	45,5	60,1	51

Прогнозне значення рівня обслуговування $x_{пр}$ = 32 %.

Варіант 11

Дослідити залежність продуктивності праці *y* від рівня механізації робіт *x* за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.11 – Статистичні дані продуктивності праці (т/год) та рівня механізації робіт (%)

x	32,2	30	36	40,5	41	47	56	54	60	55	61,3	67
y	20	24	28,5	30	31	33	34,3	37	38	40,7	42	43

Прогнозне значення рівня механізації робіт $x_{пр}$ = 50 %.

Варіант 12

Дослідити залежність виробничого травматизму *y* від рівня автоматизації небезпечних для працівників робіт *x* за статистичними даними дванадцяти підприємств однієї галузі.

Таблиця 4.12 – Статистичні дані виробничого травматизму (кількість випадків на рік) та рівня автоматизації небезпечних робіт (%)

x	64,5	65	67	70	72	74	77	83	85	87	95	97
y	22	20	17	15	14	12	10	7	4	5	4	2

Прогнозне значення рівня автоматизації небезпечних робіт $x_{пр}$ = 96 %.

Варіант 13

Дослідити залежність витрат (*y*) на промислові товари від доходів на одного члена родини (*x*) за статистичними даними дванадцяти країн.

Таблиця 4.13 – Статистичні дані витрат (у.о.) на промислові товари та про доходи на одного члена родини (у.о.)

x	90	100	120	150	190	220	250	270	300	310	350	400
y	25	50	55	67	67	93	76	114	85	125	90	170

Прогнозне значення доходів на одного члена родини $x_{пр}$ = 200 у.о.

Варіант 14

Дослідити залежність витрати сировини *y* на одиницю продукції від кількості *x* випущеної продукції за даними дванадцяти підприємств однієї галузі.

Таблиця 4.14 – Статистичні дані витрат сировини (кг) на одиницю продукції та кількості (шт.) випущеної продукції

x	1050	1170	1300	1400	1650	1700	1800	2050	2100	2300	2370	2500
y	7,2	7,6	8	8,2	6	7,5	6,7	6,5	6,1	6	6	5,7

Прогнозне значення кількості випущеної продукції $x_{пр}$ = 2000 шт.

Варіант 15

Дослідити залежність рівня рентабельності *y* від частки продовольчих товарів *x* за дванадцяти однотипних магазинів.

Таблиця 4.15 – Статистичні дані рівня рентабельності (%) та частки продовольчих товарів (%) дванадцяти однотипних магазинів

x	55	61	70	74,2	75	77	84,3	67,3	70,1	83,1	85	90
y	2,2	3	3,4	3,3	3	3,4	3,5	3	3,1	3,5	3,6	3,8

Прогнозне значення частки продовольчих товарів $x_{пр}$ = 80 %.

Варіант 16

Дослідити залежність рівня витрат у товарообігу *y* від розміру товарообігу *x* за даними дванадцяти однотипних торговельних підприємств.

Таблиця 4.16 – Статистичні дані рівня витрат у товарообігу (%) та розміру товарообігу (тис. грн)

x	7	10	15	20	30	45	60	65	70	84	100	120
y	10	9	7,5	6	6,3	5,8	5,4	5,3	5	4,8	3,8	5

Прогнозне значення товарообігу $x_{пр}$ = 50 тис. грн.

Варіант 17

Дослідити залежність товарних запасів *y* від розміру товарообігу *x* за даними дванадцяти однотипних торговельних підприємств.

Таблиця 4.17 – Статистичні дані товарних запасів (т) та розміру товарообігу (тис. грн)

x	40	47	55	60	64	75	82	88	94	100	104	110
y	1,8	3,7	4,4	4,5	3,9	5	4,8	5,5	5,2	5,9	5,3	6,4

Прогнозне значення товарообігу $x_{пр}$ = 78 тис. грн.

Варіант 18

Дослідити залежність споживання електроенергії *y* від інвестицій в енергозбереження *x* дванадцятьма однотипними підприємствами.

Таблиця 4.18 – Статистичні дані споживання електроенергії (МВт/год) та інвестицій в енергозбереження (тис. грн)

x	40	47	55	60	64	75	82	88	94	98	104	110
y	15,5	17,3	12,8	12,15	11,95	0,67	12,4	10	0,75	11,25	0,55	0,75

Прогнозне значення інвестицій в енергозбереження $x_{пр}$ = 100 тис. грн.

Варіант 19

Дослідити залежність рівня рентабельності *y* від кількості ліквідної продукції *x* в асортименті дванадцяти однотипних підприємств.

Таблиця 4.19 – Статистичні дані рівня рентабельності (%) та кількості ліквідної продукції (%) в асортименті підприємств

x	5	7	8	11	14	17	19	19	22	26	28	30
y	2,4	2	2,5	2,3	2,5	3,2	3	2,8	2,3	2,8	3	3,8

Прогнозне значення кількості ліквідної продукції $x_{\text{пр}}$ = 18 %.

Варіант 20

Дослідити залежність рівня збитковості y від кількості неліквідної продукції x в асортименті дванадцяти однотипних підприємств.

Таблиця 4.20 – Статистичні дані рівня збитковості (%) та кількості неліквідної продукції (%) в асортименті підприємств

x	5	7	8	11	14	17	19	19	22	26	28	30
y	46	49	45	45	44,5	39	40	42	37	41,5	40	32

Прогнозне значення кількості неліквідної продукції $x_{\text{пр}}$ = 20 %.

СПИСОК ЛІТЕРАТУРИ

1. Кремер Н.Ш. Эконометрика : учебник для вузов / Н.Ш. Кремер, Б.А. Путко ; под ред. Н.Ш. Кремера. – М. : ЮНИТИ-ДАНА, 2007. – 311 с.
2. Елисеева И.И. Эконометрика : учеб. / И.И. Елисеева ; под ред. И.И. Елисеевой. – М. : Финансы и статистика, 2004. – 344 с.
3. Доугерти К. Введение в эконометрику / К. Доугерти ; пер. с англ. – М. : ИНФРА-М, 1999. – 402 с.
4. Магнус Я.Р. Эконометрика. Начальный курс : учеб. / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – 6-е изд., перераб. и доп. – М. : Дело, 2004. – 576 с.
5. Айвазян С.А. Прикладная статистика и основы эконометрики / С.А. Айвазян, В.С. Мхитарян. – М. : ЮНИТИ, 1998. – 432 с.
6. Джонстон Дж. Эконометрические методы / Дж. Джонстон ; пер. с англ. – М.: Статистика, 1980. – 444 с.
7. Лещинський О.Л. Економетрія : навч. посіб. для студ. вищ. навч. закл. / О.Л. Лещинський, В.В. Рязанцева, О.О. Юнькова. – К. : МАУП, 2000. – 208 с.
8. Microsoft Excel 2010 Interactive Guide RUS – Microsoft Corporation, 2010.
9. Сингаевская Г.И. Функции в Microsoft Office Excel 2010 / Г.И. Сингаевская. – М. : Изд. дом. «Вильямс», 2011. – 1094 с.
10. Долженков В.А. Microsoft Office Excel / В.А. Долженков, А.С. Струченков. – СПб. : БХВ-Петербург, 2011. – 816 с.
11. Курбатова Е.А. Microsoft Office Excel 2010. Самоучитель / Е.А. Курбатова. – М. : Диалектика, 2010. – 416 с.

ДОДАТКИ

Додаток А

Таблиця значень $t_{p;k}$.-критерію Ст`юдента

Степені свободи k	Вірогідність p						
	0,7	0,75	0,8	0,85	0,9	0,95	0,99
1	1,963	2,414	3,078	4,165	6,314	12,706	63,657
2	1,386	1,604	1,886	2,282	2,920	4,303	9,925
3	1,250	1,423	1,638	1,924	2,353	3,182	5,841
4	1,190	1,344	1,533	1,778	2,132	2,776	4,604
5	1,156	1,301	1,476	1,699	2,015	2,571	4,032
6	1,134	1,273	1,440	1,650	1,943	2,447	3,707
7	1,119	1,254	1,415	1,617	1,895	2,365	3,499
8	1,108	1,240	1,397	1,592	1,860	2,306	3,355
9	1,100	1,230	1,383	1,574	1,833	2,262	3,250
10	1,093	1,221	1,372	1,559	1,812	2,228	3,169
11	1,088	1,214	1,363	1,548	1,796	2,201	3,106
12	1,083	1,209	1,356	1,538	1,782	2,179	3,055
13	1,079	1,204	1,350	1,530	1,771	2,160	3,012
14	1,076	1,200	1,345	1,523	1,761	2,145	2,977
15	1,074	1,197	1,341	1,517	1,753	2,131	2,947
16	1,071	1,194	1,337	1,512	1,746	2,120	2,921
17	1,069	1,191	1,333	1,508	1,740	2,110	2,898
18	1,067	1,189	1,330	1,504	1,734	2,101	2,878
19	1,066	1,187	1,328	1,500	1,729	2,093	2,861
20	1,064	1,185	1,325	1,497	1,725	2,086	2,845
21	1,063	1,183	1,323	1,494	1,721	2,080	2,831
22	1,061	1,182	1,321	1,492	1,717	2,074	2,819
23	1,060	1,180	1,319	1,489	1,714	2,069	2,807
24	1,059	1,179	1,318	1,487	1,711	2,064	2,797
25	1,058	1,178	1,316	1,485	1,708	2,060	2,787
26	1,058	1,177	1,315	1,483	1,706	2,056	2,779
27	1,057	1,176	1,314	1,482	1,703	2,052	2,771
28	1,056	1,175	1,313	1,480	1,701	2,048	2,763
29	1,055	1,174	1,311	1,479	1,699	2,045	2,756
30	1,055	1,173	1,310	1,477	1,697	2,042	2,750

Додаток В

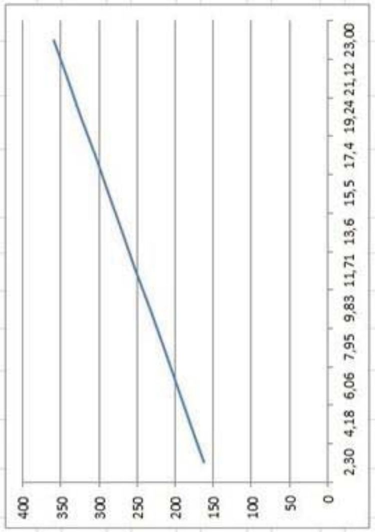
Таблиці значень F_{α,k_1,k_2}-критерію Фішера

k_2	α = 0,05					α = 0,1				
	k_1					k_1				
	1	2	3	4	5	1	2	3	4	5
1	161,4	199,5	215,7	224,6	230,2	4052	5000	5403	5625	5764
2	18,51	19,00	19,16	19,25	19,30	98,50	99,00	99,17	99,25	99,30
3	10,13	9,55	9,28	9,12	9,01	34,12	30,82	29,46	28,71	28,24
4	7,71	6,94	6,59	6,39	6,26	21,20	18,00	16,69	15,98	15,52
5	6,61	5,79	5,41	5,19	5,05	16,26	13,27	12,06	11,39	10,97
6	5,987	5,143	4,757	4,534	4,387	13,75	10,92	9,78	9,15	8,75
7	5,591	4,737	4,347	4,120	3,972	12,25	9,55	8,45	7,85	7,46
8	5,318	4,459	4,066	3,838	3,687	11,26	8,65	7,59	7,01	6,63
9	5,117	4,256	3,863	3,633	3,482	10,56	8,02	6,99	6,42	6,06
10	4,965	4,103	3,708	3,478	3,326	10,04	7,56	6,55	5,99	5,64
11	4,844	3,982	3,587	3,357	3,204	9,646	7,206	6,217	5,668	5,316
12	4,747	3,885	3,490	3,259	3,106	9,330	6,927	5,953	5,412	5,064
13	4,667	3,806	3,411	3,179	3,025	9,074	6,701	5,739	5,205	4,862
14	4,600	3,739	3,344	3,112	2,958	8,862	6,515	5,564	5,035	4,695
15	4,543	3,682	3,287	3,056	2,901	8,683	6,359	5,417	4,893	4,556
16	4,494	3,634	3,239	3,007	2,852	8,531	6,226	5,292	4,773	4,437
17	4,451	3,592	3,197	2,965	2,810	8,400	6,112	5,185	4,669	4,336
18	4,414	3,555	3,160	2,928	2,773	8,285	6,013	5,092	4,579	4,248
19	4,381	3,522	3,127	2,895	2,740	8,185	5,926	5,010	4,500	4,171
20	4,351	3,493	3,098	2,866	2,711	8,096	5,849	4,938	4,431	4,103
21	4,325	3,467	3,072	2,840	2,685	8,017	5,780	4,874	4,369	4,042
22	4,301	3,443	3,049	2,817	2,661	7,945	5,719	4,817	4,313	3,988
23	4,279	3,422	3,028	2,796	2,640	7,881	5,664	4,765	4,264	3,939
24	4,260	3,403	3,009	2,776	2,621	7,823	5,614	4,718	4,218	3,895
25	4,242	3,385	2,991	2,759	2,603	7,770	5,568	4,675	4,177	3,855
26	4,225	3,369	2,975	2,743	2,587	7,721	5,526	4,637	4,140	3,818
27	4,210	3,354	2,960	2,728	2,572	7,677	5,488	4,601	4,106	3,785
28	4,196	3,340	2,947	2,714	2,558	7,636	5,453	4,568	4,074	3,754
29	4,183	3,328	2,934	2,701	2,545	7,598	5,420	4,538	4,045	3,725
30	4,171	3,316	2,922	2,690	2,534	7,562	5,390	4,510	4,018	3,699

Додаток С

Приклад оформлення розрахунків у Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	№	1	2	3	4	5	6	7	8	9	10	11	12		β_1	9,55	139,69	β_0
2	Кількість виробленої продукції, т	2,3	4,2	4,9	6,1	7	9,3	11,5	12,6	17	18	18,7	23		$m_{\beta 1}$	2,54	32,81	$m_{\beta 0}$
3	Витрати підприємств, тис. грн	165	122	250	167	220	275	150	330	270	285	384	344		R^2	0,59	56,37	s
4															$F_{факт}$	14,14	10,00	n-2
5	Коефіцієнт кореляції	0,765					Кількість спостережень						12		$\sum(y_p-y_{ср})^2$	44943,8	31775,8	$\sum\varepsilon^2$
6																		
7	Фактичне значення критерію Ст'юдента		3,76		Критичне значення критерію Ст'юдента								1,81		$t_{\beta 1}$	3,761	4,258	$t_{\beta 0}$
8																		
9				90 %-і довірчі інтервали параметрів регресії														
10			4,95	$\le \beta_1 \le$	14,16			80,23	$\le \beta_2 \le$	199,2								
11																		$\sum(x_i-x_{ср})^2$
12	Критичне значення критерію Фішера	3,29			$x_i-x_{ср}$	-8,92	-7,02	-6,32	-5,12	-4,22	-1,9	0,28	1,38	5,8	6,78	7,48	11,78	492,577
13																		
14	Прогнозне значення кількості виробленої продукції		15		x_p	2,30	4,18	6,06	7,95	9,83	11,71	13,6	15,5	17,4	19,24	21,12	23,00	
15					y_p	161,7	179,6	197,6	215,6	233,6	252	270	287	305	323,438	341,41	359,39	
16	Середнє значення кількості виробленої продукції		11,22															
17																		
18	Точкова оцінка умовного математичного очікування, тис. грн		283															
19																		
20	Стандартна помилка групової середньої		18,90															
21																		
22	90%-й довірчий інтервал умовного математичного очікування	248,7	≤	$M_{xp}(y)$	≤	317,2												
23																		
24	Стандартна помилка індивідуальних значень		59,45															
25																		
26	90%-й довірчий інтервал індивідуальних значень	175,2	≤	y_{xp}*	≤	390,7												



Навчальне видання

МЕТОДИЧНІ ВКАЗІВКИ

до виконання розрахунково-графічного завдання «Парний кореляційно-регресійний аналіз» за розділом «Економетрія» дисципліни «Економіко-математичне моделювання» для студентів заочної форми навчання спеціальностей 8.030601 «Менеджмент організацій», 8.050106 «Облік та аудит», 8.050107 «Маркетинг»

Укладачі: СКВОРЧЕВСЬКИЙ Олександр Євгенович
ТОВАЖНЯНСЬКИЙ В'ячеслав Леонідович
ПОБЕРЕЖНИЙ Роман Олегович

Відповідальний за випуск О.Д. Матросов
Роботу до видання рекомендував М.І. Погорєлов

Редактор Л.А. Пустовойтова

План 2013 р., поз. 179

Підп. до друку __.__.__. Формат 60х84 1/16. Папір офсетний.
Друк – ризографія. Гарнітура Times New Roman. Ум. друк. арк. 3,12.
Обл.-вид. арк. 2,02. Наклад 100 пр. Зам. № _____. Ціна договірна.

Видавничий центр НТУ "ХПІ".
Свідоцтво про державну реєстрацію ДК №
61002, Харків, вул. Фрунзе, 21.

Друкарня НТУ "ХПІ".
61002, Харків, вул. Фрунзе, 21.