

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
Київський національний університет будівництва і архітектури

ЕКОНОМЕТРІЯ

Методичні вказівки
до виконання лабораторних і практичних робіт
для студентів спеціальності 6.030601 «Менеджмент»
усіх форм навчання

Київ 2013

УДК 519.862.6

E45

Укладачі: Н.Д. Федоренко, канд. техн. наук, професор
С.В. Білощицька, канд. техн. наук, доцент
А.О. Білощицький, канд. техн. наук, доцент
О.В. Доля, канд. фіз.-мат. наук, доцент
С.А. Теренчук, канд. фіз.-мат. наук, доцент

Рецензент В.М. Михайленко, д-р техн. наук, професор

Відповідальний за випуск В.В. Демченко, канд. техн. наук,
доцент

*Затверджено на засіданні кафедри прикладної математики,
протокол № 17 від 17 травня 2012 р.*

Видається в авторській редакції.

Економетрія: методичні вказівки до лабораторних і практичних
E45 робіт. /уклад.: Н.Д. Федоренко, С.В. Білощицька,
А.О. Білощицький. – К.: КНУБА, 2013. – 40 с.

Містять завдання для лабораторних і практичних робіт з
дисципліни «Економетрія», методику та приклади їх виконання,
основні формули та поняття, контрольні запитання та список
рекомендованої літератури.

Призначені для студентів спеціальності 6.030601 «Менеджмент»
усіх форм навчання.

Загальні положення

Процес прийняття науково обґрунтованих рішень в економіці тісно пов'язаний з визначенням *кількісних* співвідношень між економічними показниками. Економетрія – галузь економічної науки, яка вивчає методи кількісного вимірювання взаємозв'язків між економічними показниками, а також розглядає основні напрямки застосування економетричних моделей в економічних дослідженнях. Економетрія є однією з найважливіших дисциплін фундаментальної підготовки сучасних економістів.

Основне завдання економетрії полягає в оцінюванні параметрів і перевірці значущості економетричної моделі з урахуванням особливостей вихідної інформації.

Мета вивчення дисципліни полягає в тому, щоб навчити студентів кількісно оцінювати взаємозв'язки економічних показників для різних масивів економічної інформації, вдаючись до тестування останньої щодо відповідності її певним передумовам, а також до визначення методів кількісного вимірювання зв'язків, які доцільно застосовувати в кожному конкретному випадку згідно з особливостями економічної інформації.

Економетрія є синтезною дисципліною, яка поєднує в собі економічну теорію, математичну економіку, економічну та математичну статистику. Економетрія використовує методи розв'язання задач з багатьох розділів математики, наприклад, теорії ймовірностей, лінійної та матричної алгебри, диференціального числення та ін. Економетрія з огляду на громіздкість обчислень та вимоги до точності результатів вивчається за допомогою комп'ютера. Вона надає додаткові можливості оволодіти обчислювальною технікою, розвиває аналітичні навички та є основою економічних досліджень.

Знання, здобуті студентами під час вивчення економетрії, широко застосовуються в менеджменті, маркетингу, фінансовій справі, податковому менеджменті та ін.

Методичні вказівки містять завдання для лабораторних робіт, рекомендації та приклади їх виконання. Перед виконанням роботи студентам рекомендується ознайомитися зі списком рекомендованої літератури.

Контрольні запитання використовуються викладачем для перевірки знань студентів під час захисту лабораторної роботи.

У результаті вивчення курсу студенти повинні

знати:

- сутність економетричного моделювання та його етапи;
- методи тестування економічної інформації;
- методи оцінювання параметрів економетричної моделі з урахуванням особливостей конкретної економічної інформації;
- методи оцінювання достовірності моделі та її параметрів;
- методи оцінювання прогнозних властивостей моделі;
- методи економетричного прогнозування з урахуванням особливостей економетричних моделей;

вміти:

- ідентифікувати змінні моделі;
- специфікувати модель;
- оцінювати параметри економетричної моделі в разі:
 - нормально розподілених залишків моделі;
 - мультиколінеарності незалежних змінних;
 - наявності гетероскедастичності залишків;
- визначати прогнозні властивості моделі;
- перевіряти достовірність моделі та її параметрів;
- виконувати точковий та інтервальний прогнози на основі економетричних моделей;
- визначати основні економічні характеристики взаємозв'язку та правильно їх тлумачити;
- опанувати методи побудови та реалізації економетричних моделей за допомогою персонального комп'ютера;
- застосовувати економетричні моделі в економічних дослідженнях;
- самостійно поглиблювати теоретичні знання в галузі математичного моделювання економічних процесів і явищ.

ЛАБОРАТОРНА РОБОТА 1



Тема: «Побудова деяких економетричних моделей за допомогою засобів діалогового вікна **Линия тренда** програми **Excel**»

Мета роботи: Навчитися будувати економетричні моделі та виконувати прогноз за допомогою програми **Excel**.

Завдання: Для заданого набору пар значень незалежної змінної x та залежної змінної y (табл. 1-3) визначити найкраще наближення у вигляді лінії тренда, за допомогою якого розрахувати прогнозне значення.

Методика виконання

1. Запустіть програму **Exel** (**Пуск – Программы – Microsoft Excel**) та створіть робочу книгу **ЕКОН_ПБ.xls**. Активізуйте **Лист 1** та перейменуйте його у **Лабораторна робота 1**, введіть вихідні дані.

2. За вихідними даними побудуйте поле кореляції. Для цього виділіть вихідні дані лівою кнопкою миші та на панелі інструментів клацніть на кнопці  (**Мастер диаграмм**). У діалоговому вікні **Мастер диаграмм** на вкладці **Стандартные** виберіть в секції **Тип:** – **Точечная**, в секції **Вид:** – . Натисніть на кнопці **Далее**, та ще раз натисніть на кнопці **Далее**, на вкладці **Заголовки** заповніть поля вводу **Ось X**, **Ось Y**, на вкладці **Линии сетки** встановіть у секції **Ось X** прапорець **Основные линии**, на вкладці **Легенда** приберіть прапорець **Добавить легенду**, натисніть на кнопці **Готово**. Зробіть ще 5 копій рисунків.

3. Для побудови лінії тренда клацніть правою кнопкою миші по будь-якій точці першого графіка вихідних даних та в контекстному меню виберіть пункт **Добавить линию тренда...** На екрані з'явиться діалогове вікно **Линия тренда** (рис. 1).

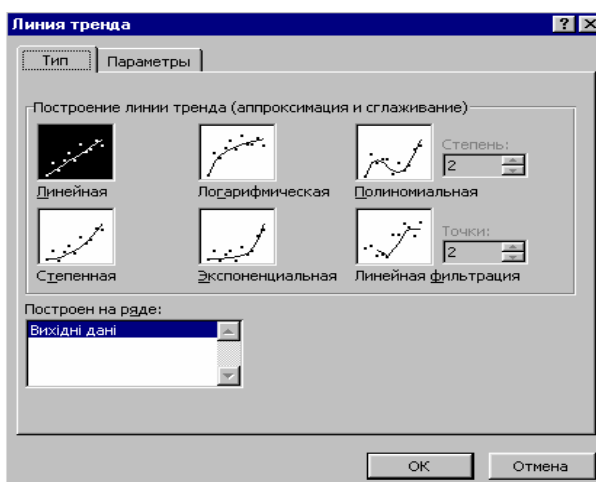


Рис. 1. Діалогове вікно **Линия тренда**, вкладка **Тип**

У вікні, що з'явилося, виберіть тип тренда **Линейная**, на вкладці **Параметры** встановіть прапорці **Показывать уравнение на диаграмме** та **Поместить на диаграмму величину достоверности аппроксимации (R^2)** (рис. 2).

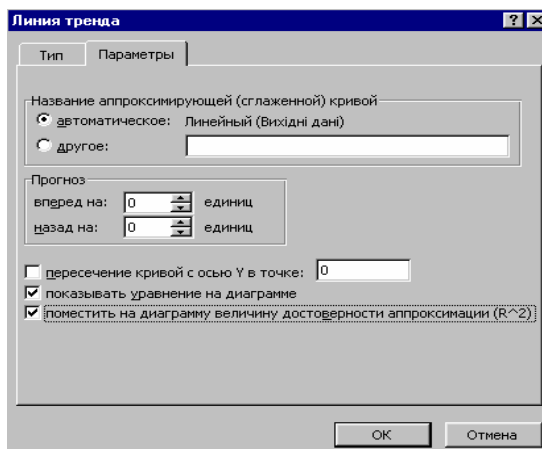


Рис. 2. Діалогове вікно Линия тренда, вкладка Параметры

Натисніть на кнопки **ОК**. У результаті на полі кореляції з'являться графік вибраної лінії тренда, її рівняння та значення показника якості рівняння – коефіцієнта детермінації R^2 (на рисунках він позначається як R^2).

Коефіцієнт детермінації змінюється в межах: $0 \leq R^2 \leq 1$. Чим менше експериментальні точки відхиляються від лінії тренда, тим ближче до 1 коефіцієнт детермінації, та навпаки, чим більше точки відхиляються від лінії тренда, тим ближче R^2 до нуля.

4. Виконайте ще декілька разів дії, перелічені у п. 3 для побудови логарифмічної, поліноміальної (2, 3 степенів), ступеневої, експоненціальної лінії тренда, обираючи кожен раз у діалоговому вікні **Линия тренда** (див. рис. 1) відповідний тип тренда (**Логарифмическая**, **Полиномиальная (Степень: 2)**, **Полиномиальная (Степень: 3)**, **Степенная**, **Экспоненциальная**). В результаті повинні з'явитися 6 рисунків, на яких зображені вихідні дані, лінії тренда, рівняння регресії та коефіцієнти детермінації R^2 .

5. Визначте найкраще наближення, якому відповідає найбільше значення коефіцієнта детермінації R^2 .

6. За рівнянням регресії, яке відповідає обраній у п.5 лінії тренда, розрахуйте прогнозне значення у для прогнозного значення $x_p = 1,2x_{max}$.

7. Відобразіть прогнозне значення на графіку. Для цього клацніть правою клав'яшею миші по осі x , у контекстному меню, що з'явилося, виберіть пункт **Формат оси**, на вкладці **Шкала** встановіть у полі **максимальное значение** відповідне значення, натисніть на **ОК**. Якщо потрібно – так само змініть формат осі y . Потім клацніть правою клав'яшею миші по лінії тренда та у контекстному меню, що з'явилося, виберіть

пункт **Формат линии тренда...** На вкладці **Параметры** встановіть у секції **Прогноз** у полі **вперед на:** необхідну кількість одиниць ($0,2x_{max}$). Натисніть на кнопці **ОК**. Зверніть увагу на те, що лінія тренда продовжилася на відповідну величину.

Завдання до лабораторної роботи №1

Таблиця 1

Середня урожайність зернових культур (y_i) за останні 10 років

Варіант	x_i	1	2	3	4	5	6	7	8	9	10
1	y_i	27,66	26,76	25,71	25,37	26,64	27,03	28,36	30,39	32,77	33,05
2	y_i	15,23	16,37	18,34	17,95	19,45	20,43	18,65	20,56	22,65	30,08
3	y_i	25,98	27,06	31,04	28,85	29,04	31,83	32,67	34,01	36,25	35,04
4	y_i	17,56	19,32	20,67	21,44	24,63	25,08	23,79	22,56	24,12	25,07
5	y_i	28,07	31,7	32,61	33,09	31,12	33,83	42,6	44,09	46	49,89
6	y_i	19,65	20,99	34,15	35,87	33,05	32,77	35,14	36,47	38,94	39,15
7	y_i	20,13	21,99	23,15	24,97	25,67	28,11	29,43	27,26	28,15	29,11
8	y_i	25,16	26,76	28,46	29,96	31,06	32,77	33,63	35,27	36,06	37,09
9	y_i	16,94	17,25	19,15	20,61	19,33	18,12	21,34	22,11	19,99	20,7
10	y_i	23,9	25,8	26,56	28,15	29,31	30,6	33,5	34,96	36,02	35,57

Таблиця 2

Динаміка зміни середньоденного вироблення продукції на 1 робітника (y_i , ум. од.) за останні 10 років

Варіант	x_i	1	2	3	4	5	6	7	8	9	10
11	y_i	27,31	27,99	28,67	29,15	30,05	29,45	28,14	30,11	31,16	32,45
12	y_i	24,11	25,87	26,73	27,73	26,99	27,45	29,08	30,05	31,67	32,11
13	y_i	16,76	17,85	18,34	19,28	20,09	21,39	22,94	23,17	24,42	25,03
14	y_i	10,43	11,06	12,45	11,38	12,01	13,85	14,95	15,67	17,32	19,31
15	y_i	36,61	37,11	38,54	39,03	40,15	40,34	40,51	41,73	42,78	43,12
16	y_i	26,98	26,00	27,71	28,61	29,55	30,16	30,45	31,88	32,15	33,04
17	y_i	27,03	28,16	29,52	29,99	30,45	30,94	31,51	32,88	33,92	34,6
18	y_i	10,6	11,74	12,84	13,09	14,73	13,96	14,37	15,45	17,21	18,5
19	y_i	31,87	32,15	33,5	34,16	35,61	36,9	36,8	37,51	38,71	37,02
20	y_i	36,08	37,18	38,11	37,41	38,49	39,41	40,11	41,6	43,8	44,84

Таблиця 3

Залежність попиту на товар (y_i , шт.) від його ціни (x_i , ум. од.)

Варіант	x_i	8,5	8,1	7,7	7,3	6,5	5,7	4,9	3,2	1,3	0,9
21	y_i	3	5	7	8	8	10	15	17	19	22
22	y_i	6	10	13	17	20	22	27	31	39	43
23	y_i	10	12	15	18	21	25	28	29	31	33
24	y_i	8	9	12	14	15	16	19	23	29	34
25	y_i	25	27	31	35	37	38	42	44	47	49
26	y_i	11	13	14	17	18	21	25	27	31	32
27	y_i	6	8	9	13	15	16	19	21	23	24
28	y_i	14	15	19	21	24	26	29	30	34	38
29	y_i	4	7	9	13	15	18	20	21	25	27
30	y_i	20	22	25	26	29	31	33	36	40	41

Приклад виконання лабораторної роботи №1

Завдання. Для заданого набору пар значень незалежної змінної x та залежної змінної y наведених у табл. 4 визначити найкраще наближення у вигляді лінії тренда, за допомогою якого розрахувати прогнозне значення.

Таблиця 4

Рік (x_i)	1	2	3	4	5	6	7	8	9	10
Зростання прибутку, % (y_i)	0,8256	1,5674	2,9891	3,5582	4,6754	5,2426	7,4742	9,3446	10,2341	11,2205

За даними табл. 4 побудуємо поле кореляції у 6 екземплярах (задані точки на рисунках зображені ромбовидними маркерами). Потім на кожному з рисунків за допомогою діалогового вікна **Линия тренда** програми *Excel* побудуємо відповідну лінію тренда (рис. 3).

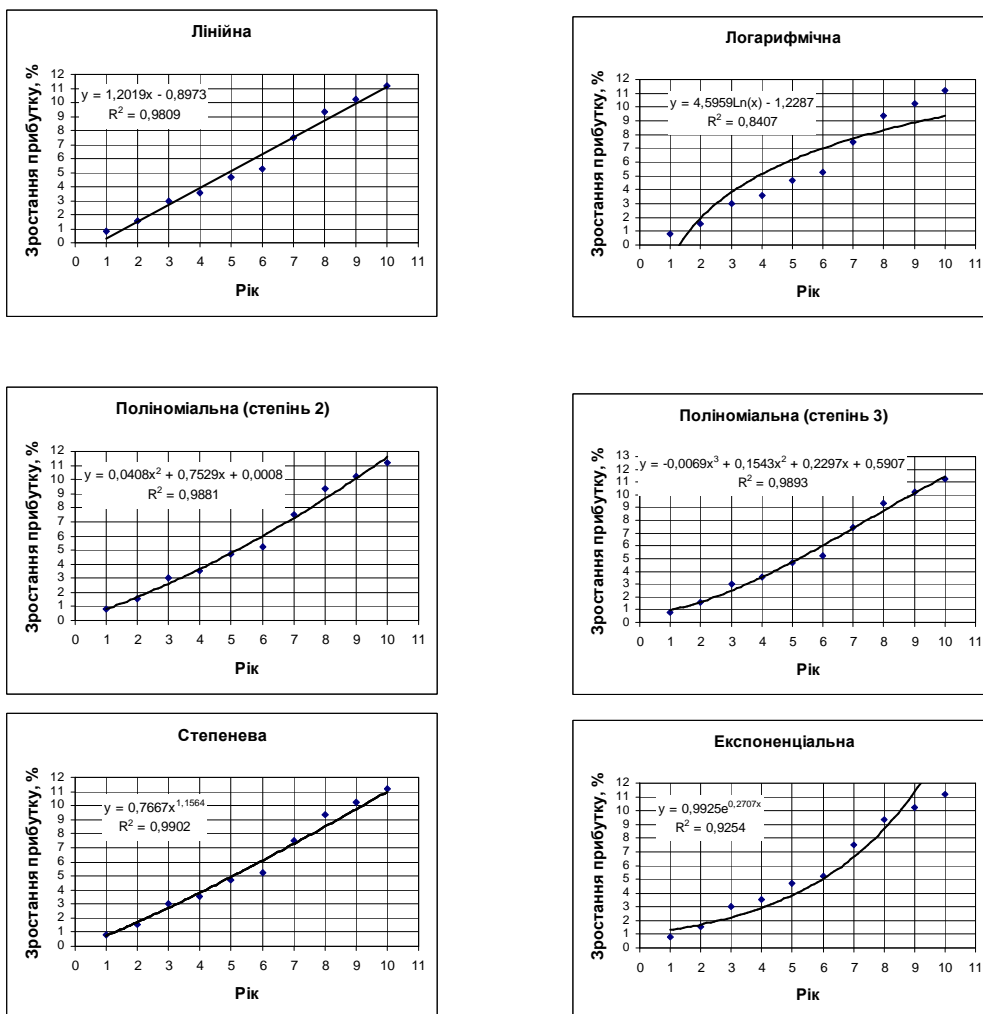


Рис.3. Моделювання динаміки зростання прибутку за допомогою стандартних ліній тренда програми Excel

Висновок: оскільки степенева лінія тренда має найвищий коефіцієнт детермінації ($R^2=0,9902$) – вона є найкращим наближенням до вихідних даних.

Для прогнозного значення аргументу $x_p = 1,2x_{max}=1,2 \cdot 10 = 12$ за рівнянням степеневі лінії тренда розрахуємо прогнозне значення прибутку:

$$y_p=0,7667x_p^{1,1564}=0,7667 \cdot 12^{1,1564} = 13,5703.$$

Продовжимо графік вперед на $x_p - x_{max} = 2$ одиниці і відобразимо прогнозне значення на графіку степеневі залежності (зображено круглим маркером) (рис.4).

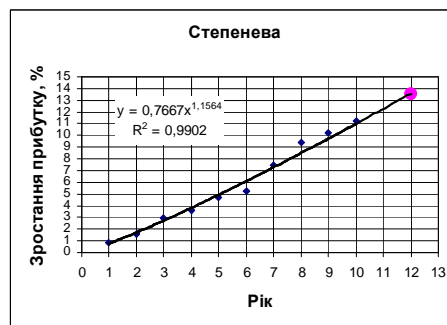


Рис.4 Прогнозне значення прибутку для прогнозного значення аргументу x_p

Контрольні запитання

1. Що є об'єктом, предметом та метою економетрії? Яке основне завдання економетричних досліджень?
2. Наведіть формули та графіки для наступних видів парних залежностей: а) лінійної; б) логарифмічної; в) поліноміальної; г) степеневі; г) експоненціальної.
3. З якою метою розробляються математичні моделі?
4. За допомогою якого показника можна обрати найкраще рівняння регресії?
5. Що таке випадкова величина (ВВ)? Яки види ВВ Вам відомі? Наведіть приклади дискретних та неперервних ВВ з економіки.
6. Дайте означення закону розподілу ВВ. Яким чином можна його задати?
7. Дайте означення функції розподілу ВВ.
8. Дайте означення функції щільності ймовірності ВВ.
9. Перелічіть основні числові характеристики ВВ.
10. Дайте означення математичному сподіванню ВВ. Перелічіть його основні властивості.

11. Дайте означення *дисперсії* ВВ.
12. Дайте означення *середньому квадратичному відхиленню* ВВ.
13. Дайте означення *коваріації*.
14. Як визначається і для чого використовується *коефіцієнт кореляції*?
15. Що таке генеральна сукупність, вибірка?
16. Як за результатами вибірки визначаються: вибіркоче середнє, вибіркоче дисперсія, вибіркоче середнє квадратичне відхилення, вибіркочі коефіцієнти коваріації та кореляції?

ЛАБОРАТОРНА РОБОТА 2

Тема: «Побудова парної лінійної регресії та аналіз її якості»

Мета роботи: Навчитися будувати рівняння парної лінійної регресії та оцінювати його якість.

Завдання: На підставі даних табл. 5 побудувати парну лінійну модель залежності витрат на 1 грн виробленої продукції (y) від поточного періоду (x) та проаналізувати її якість.

Для виконання завдання потрібно:

1. Розрахувати коефіцієнт кореляції r_{xy} та зробити висновок про зв'язок між X і Y (прямий або зворотний, тісний або ні). Перевірити статистичну значущість r_{xy} для рівня надійності $P = 0,95$.
2. Методом найменших квадратів (МНК) розрахувати коефіцієнти b_0, b_1 парної лінійної регресії:

$$\hat{y}(x) = b_0 + b_1x, \quad (1)$$

намалювати поле кореляції та графік лінії регресії (1).

3. Розрахувати коефіцієнт детермінації R^2 .
4. Перевірити адекватність моделі за F -критерієм Фішера.
5. Перевірити результати розрахунків за допомогою програми **Анализ данных**.

Методика виконання:

1. Розрахуйте коефіцієнт кореляції r_{xy} .

Коефіцієнт кореляції є характеристикою *лінійного* взаємозв'язку між двома випадковими величинами (ВВ) X і Y та розраховується за формулою:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}}, \quad (2)$$

де $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, $\overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i$, $\overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$ – середнє значення;

n – кількість спостережень.

Коефіцієнт кореляції набуває значення в інтервалі $-1 \leq r_{xy} \leq 1$.

Додатне значення коефіцієнта кореляції свідчить про прямий зв'язок між X і Y (із зростанням однієї ВВ зростає середнє значення іншої), від'ємне – про зворотний зв'язок (із зростанням однієї ВВ середнє значення іншої убыває). Якщо $r_{xy} \rightarrow \pm 1$ – зв'язок тісний, якщо $r_{xy} \rightarrow 0$ – лінійного зв'язку немає.

Для перевірки статистичної значущості коефіцієнта кореляції за t -критерієм Стьюдента необхідно:

1) розрахувати t -відношення: $t_r = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$; (3)

2) з таблиць критичних точок розподілу Стьюдента знайти $t_{\alpha/2, n-2}$, де α – рівень значущості, зв'язаний з рівнем надійності P співвідношенням: $\alpha = 1 - P$;

3) якщо $|t_r| > t_{\alpha/2, n-2}$ – коефіцієнт r_{xy} статистично значимо відрізняється від нуля.

2. Методом найменших квадратів розрахуйте коефіцієнти b_0, b_1 парної лінійної регресії (1).

Згідно з МНК параметри рівняння (1) визначаються так, щоб мінімізувати суму квадратів відхилень фактичних значень y_i від теоретичних \hat{y}_i :

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 = f(b_0, b_1) \rightarrow \min. \quad (4)$$

Прирівнявши до нуля частинні похідні першого порядку від функції $f(b_0, b_1)$ за параметрами b_0, b_1 , одержимо систему лінійних рівнянь:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i; \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i, \end{cases} \Rightarrow \begin{cases} b_0 + b_1 \bar{x} = \bar{y} \\ b_0 \bar{x} + b_1 \overline{x^2} = \overline{xy}. \end{cases} \quad (5)$$

Із системи (5) знаходимо:

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\text{cov}(X, Y)}{\text{var}(X)}; \quad b_0 = \bar{y} - b_1 \bar{x}. \quad (6)$$

3. Розрахуйте коефіцієнт детермінації R^2 .

Коефіцієнт детермінації є сумарною мірою якості рівняння регресії та розраховується за формулою:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}, \quad (7)$$

де $SST = \sum_{i=1}^n (y_i - \bar{y})^2 = n \text{var}(Y) = n(\overline{y^2} - \bar{y}^2)$ - загальна сума квадратів;

$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$ - сума квадратів помилок;

$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ - сума квадратів, що пояснюється регресією

Коефіцієнт детермінації показує, на скільки відсотків варіація залежної змінної визначається варіацією незалежної змінної. Наприклад, значення коефіцієнта детермінації $R^2 = 0,95$ свідчить про те, що 95 % загальної дисперсії величини Y пояснюється рівнянням регресії.

4. **Перевірка адекватності моделі за F -критерієм Фішера** здійснюється у такій послідовності:

1) розраховується F -відношення:

$$F = \frac{SSR}{SSE/(n-2)} = \frac{R^2}{(1-R^2)/(n-2)}. \quad (8)$$

2) для рівня значущості α та ступенів вільності $k_1 = 1$, $k_2 = n - 2$ із статистичних таблиць F -розподілу Фішера знаходиться критичне значення $F_{кр} = F_{\alpha; 1, n-2}$;

3) якщо розраховане значення $F > F_{кр}$, то побудована регресійна модель адекватно апроксимує дані спостережень.

5. Перевірте результати розрахунків за допомогою програми **Анализ данных**. Виконайте команди **Сервис-Анализ данных**. У діалоговому вікні, що з'явилось, **Анализ данных** виберіть пункт **Регрессия**, натисніть на **ОК**. На екрані з'явиться діалогове вікно **Регрессия** (рис. 5), в якому необхідно заповнити поля **Входной интервал Y:**, **Входной интервал X:** відповідними інтервалами вихідних даних. Натисніть на кнопку **ОК**. На екрані з'явиться додатковий лист **ВЫВОД ИТОГОВ**, дані якого необхідно порівняти з розрахованими раніше відповідними значеннями. Значення розрахованого r_{xy} повинно співпадати зі значенням *Множественный R* таблиці *Регрессионная статистика*, значення коефіцієнта детермінації R^2 – зі значенням *R-квадрат* тієї ж таблиці, значення розрахованих SSR , SST , SSE , F -критерію – з відповідними значеннями таблиці *Дисперсионный анализ*, значення коефіцієнтів b_0 , b_1 – зі значеннями стовпчика *Коэффициенты* останньої таблиці. Якщо дані співпадають – розрахунки проведені правильно.

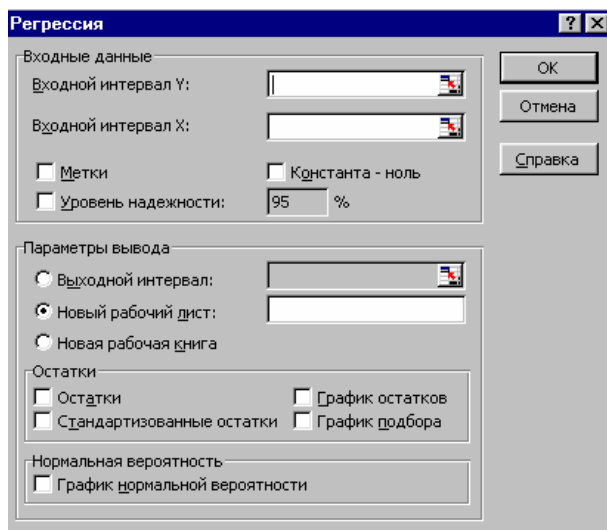


Рис. 5. Диалоговое окно Регрессия

Завдання до лабораторної роботи №2

Таблиця 5

Варіант	i	1	2	3	4	5	6	7	8	9	10
1	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	0,142	1,346	2,831	1,934	2,153	3,149	4,231	5,218	6,002	7,301
2	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	2,945	3,637	4,001	3,951	5,484	6,584	5,394	7,001	8,184	7,991
3	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	6,849	7,582	9,341	8,128	7,993	9,003	10,151	11,834	10,995	11,731
4	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	2,429	2,995	3,869	4,312	3,741	4,342	5,736	6,377	7,511	8,004
5	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	6,219	5,993	6,483	7,841	8,626	9,128	8,413	9,489	10,063	11,381
6	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	3,404	4,319	5,429	6,494	7,365	9,418	10,001	11,74	12,056	11,994
7	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	0,561	1,747	2,741	3,567	4,861	5,957	6,051	7,672	7,992	8,827
8	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	5,847	6,957	7,639	6,736	8,003	9,629	10,341	11,748	11,116	12,538
9	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	27,351	27,929	28,677	29,154	30,205	29,145	28,714	30,211	31,816	32,045
10	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	24,611	25,827	26,173	27,633	26,919	27,845	29,408	30,325	31,467	32,161
11	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	16,776	17,285	18,134	19,628	20,709	20,139	21,964	22,917	24,142	23,033
12	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	10,453	11,076	12,145	11,938	12,401	13,485	14,395	15,167	17,332	19,381
13	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	36,651	37,171	38,524	39,093	40,152	40,344	40,581	41,732	42,710	43,105
14	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	26,918	26,006	27,718	28,621	29,555	30,196	30,453	31,886	32,151	33,041
15	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	27,073	28,162	29,526	29,997	30,453	30,945	31,517	32,858	33,912	34,688
16	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	10,642	11,743	12,841	13,094	14,736	13,996	14,371	15,454	17,213	18,537

Закінчення табл. 5

17	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	31,817	32,152	33,534	34,165	35,621	36,952	36,847	37,513	38,771	37,022
18	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	36,308	37,118	38,131	37,411	38,491	39,414	40,191	41,696	43,865	44,843
19	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	27,631	27,969	28,167	29,615	30,025	29,445	29,144	30,112	31,316	32,485
20	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	23,181	25,187	25,973	27,073	26,499	27,465	28,058	30,025	31,647	32,119
21	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	15,764	16,185	18,034	19,228	20,809	21,239	22,494	23,179	24,422	25,703
22	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	10,643	11,036	12,845	11,378	12,701	12,854	14,956	15,671	17,324	19,361
23	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	26,998	26,002	27,716	28,611	29,554	30,162	30,415	31,885	32,153	33,074
24	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	27,053	28,178	29,527	29,989	30,495	30,904	31,551	32,088	33,945	34,621
25	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	10,366	11,724	12,846	13,098	14,723	13,956	14,372	15,451	17,214	18,596
26	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	32,876	32,135	33,675	34,116	35,461	36,259	36,368	37,511	38,716	37,032
27	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	36,508	37,128	38,115	37,413	38,459	39,741	40,911	41,246	43,842	44,184
28	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	24,151	25,827	26,733	27,753	26,991	27,455	29,038	30,505	31,767	32,121
29	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	10,473	11,206	12,445	11,538	12,061	13,825	14,975	15,617	17,332	19,317
30	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	26,198	26,748	27,271	27,614	29,575	30,136	30,458	31,882	32,115	33,048

Приклад виконання лабораторної роботи №2

Завдання. На підставі даних, що містяться у табл. 6, побудувати парну лінійну модель залежності витрат на 1 грн виробленої продукції (y) від поточного періоду (x) та проаналізувати її якість.

Таблиця 6

i	1	2	3	4	5	6	7	8	9	10
y_i	2,2	4,2	5,7	6,8	5,9	7,6	9,5	8,4	10,1	12,3
x_i	1,4	2,2	3,3	2,6	3,2	4,5	5,1	6,7	7,3	8,9

1. Для спрощення розрахунків і знаходження коефіцієнтів економетричної моделі побудуємо таблицю:

Таблиця 7

i	y_i	x_i	$x_i y_i$	x_i^2	y_i^2	$Y_i = b_0 + b_1 x_i$	$e_i = y_i - Y_i$	e_i^2
1	2,2	1,4	3,08	1,96	4,84	3,7354	-1,5354	2,3576
2	4,2	2,2	9,24	4,84	17,64	4,6417	-0,4417	0,1951
3	5,7	3,3	18,81	10,89	32,49	5,8879	-0,1879	0,0353
4	6,8	2,6	17,68	6,76	46,24	5,0949	1,7051	2,9074
5	5,9	3,2	18,88	10,24	34,81	5,7746	0,1254	0,0157
6	7,6	4,5	34,2	20,25	57,76	7,2473	0,3527	0,1244

7	9,5	5,1	48,45	26,01	90,25	7,9271	1,5729	2,4741
8	8,4	6,7	56,28	44,89	70,56	9,7397	-1,3397	1,7947
9	10,1	7,3	73,73	53,29	102,01	10,4194	-0,3194	0,102
10	12,3	8,9	109,47	79,21	151,29	12,2320	0,0680	0,0046
Σ	72,7	45,2	389,82	258,34	607,89			10,011
	Y	X	XY	X ²	Y ²			SSE
Σ/n	7,27	4,52	38,982	25,834	60,789			

2. Розрахуємо коефіцієнт кореляції:

$$r_{xy} = (XY - X*Y)/\text{КОРЕНЬ}((X^2 - X^2)(Y^2 - Y^2)) = \\ = (38,982 - 4,52*7,27)/\text{КОРЕНЬ}((25,834 - 4,52^2)(60,789 - 7,27^2)) = 0,9348.$$

Отже зв'язок прямий, тісний.

Для перевірки статистичної значущості коефіцієнта кореляції обчислимо t -статистику:

$$t_r = r_{xy} \text{КОРЕНЬ}((n - 2)/(1 - r_{xy}^2)) = \\ = 0,9348 * \text{КОРЕНЬ}((10 - 2)/(1 - 0,9348^2)) = 7,4443,$$

де $n = 10$ – кількість спостережень.

Порівняємо значення t_r з критичним значенням, визначеним за таблицею критичних точок розподілу Стьюдента для рівня значущості

$$\alpha = 1 - P = 1 - 0,95 = 0,05, \quad t_{кр}(\alpha/2; n - 2) = t_{кр}(0,025; 8) = 2,306.$$

Оскільки $|t_r| > t_{кр}$ – коефіцієнт кореляції є статистично значущим.

3. Розрахуємо коефіцієнти рівняння лінійної регресії:

$$b_1 = (XY - X*Y)/(X^2 - X^2) = \\ = (38,982 - 4,52*7,27)/(25,834 - 4,52^2) = 1,1328; \\ b_0 = Y - b_1*X = 7,27 - 1,1328*4,52 = 2,1497.$$

Рівняння лінійної регресії:

$$y = b_1x + b_0 = 1,13x + 2,15.$$

На рис. 6 наведено поле кореляції та графік лінійної регресії.

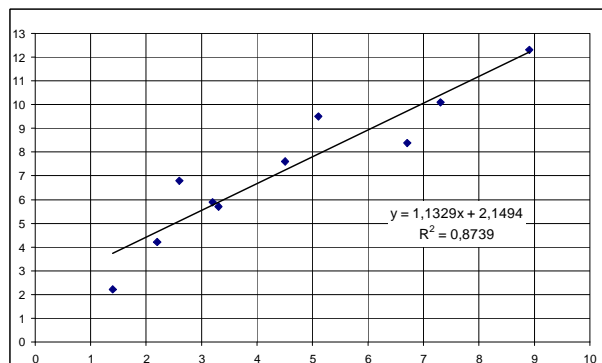


Рис. 6. Графік парної лінійної регресії

4. Розрахуємо коефіцієнт детермінації:

$$R^2 = 1 - SSE/SST = 1 - \sum e_i^2 / (n(Y^2 - Y^{\wedge 2})) = \\ = 1 - 10,011 / (10 * (60,789 - 7,27^2)) = 0,8739,$$

де $SSE = \sum (y_i - Y_i)^2 = \sum e_i^2$ – сума квадратів помилок;

$SST = \sum (y_i - Y)^2 = n(Y^2 - Y^{\wedge 2})$ – загальна сума квадратів.

Оскільки $R^2 = 0,8739$, то 87,39 % дисперсії величини у пояснюється рівнянням регресії.

5. Перевірка адекватності моделі за F -критерієм Фішера.

Розрахуємо F -відношення:

$$F = F(1, n - 2) = (n - 2)R^2 / (1 - R^2) = \\ = (10 - 2) * 0,8739 / (1 - 0,8739) = 55,4417.$$

За таблицею критичних точок розподілу Фішера для рівня значущості $\alpha = 0,05$ знайдемо: $F_{кр}(\alpha; 1; n - 2) = F_{кр}(0,05; 1; 8) = 5,32$.

Оскільки $F > F_{кр}$ – побудована регресійна модель адекватно апроксимує дані спостережень.

Правильність розрахунків перевіримо за допомогою програми **Анализ данных**, результати роботи якої наведені нижче.

ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>		
Множественный R	0,93480233	Коефіцієнт кореляції
R-квадрат	0,873855396	Коефіцієнт детермінації
Нормированный R-квадрат	0,85808732	
Стандартная ошибка	1,118646613	S
Наблюдения	10	

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	1	SSR =69,35003805	69,35003805	55,41928009	7,30326E-05
Остаток	8	SSE =10,01096195	1,251370244		
Итого	9	SST =79,361			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	b₀ =2,149407802	S_{b0} =0,773476535	t_{b0} =2,77889206	0,023965499	0,36767716	3,933047889
Переменная X 1	b₁ =1,13287438	S_{b1} =0,152177805	t_{b1} =7,444412676	7,30326E-05	0,781951733	1,4837797027

Контрольні запитання

1. Сформулюйте означення *функції регресії*.
2. Сформулюйте означення парної лінійної регресії.
3. У чому складається розходження між *теоретичним* і *емпіричним* рівняннями регресії?
4. У чому суть методу найменших квадратів (МНК)?
5. Наведіть формули для розрахунків коефіцієнтів емпіричного парного лінійного рівняння регресії за МНК.
6. Для чого призначені, у яких діапазонах змінюються, за якими формулами розраховуються *коефіцієнти кореляції, детермінації*?
7. Сформулюйте означення та наведіть формули для розрахунків SSR, SSE, SST.
8. Чому дорівнюють ступені вільності величин SSR, SSE, SST?
9. Що називається *середнім квадратом ВВ*?
10. Опишіть процес перевірки адекватності моделі за *F*-критерієм Фішера.

ЛАБОРАТОРНА РОБОТА 3

Тема: «Аналіз якості коефіцієнтів рівняння парної лінійної регресії. Прогнозування залежної змінної»

Мета роботи: За допомогою *t*-критерію Стюдента навчитися аналізувати точність визначення оцінок коефіцієнтів регресії, будувати інтервали довіри для параметрів теоретичної лінійної регресії, визначати точковий і інтервальний прогнози для залежної змінної.

Завдання: Для вихідних даних лабораторної роботи 2 зробіть наступне:

1. За допомогою *t*-критерію Стюдента перевірте статистичну значущість коефіцієнтів b_0 та b_1 лінійної регресії, які визначені в лабораторній роботі 2.
2. Розрахуйте інтервали довіри для параметрів β_0 , β_1 теоретичної лінійної регресії.
3. Зробіть точковий та інтервальний прогнози для залежної змінної, якщо значення незалежної змінної збільшиться на 20 % від її максимального рівня.
4. Розрахуйте середній частинний коефіцієнт еластичності і оцініть силу впливу фактора на результат.

Методика виконання

1. Для перевірки значущості параметрів b_0, b_1 лінійного рівняння регресії за допомогою t -тесту Стьюдента необхідно:

1) розрахувати t -відношення: $t_{b_1} = \frac{b_1}{S_{b_1}}, t_{b_0} = \frac{b_0}{S_{b_0}},$ (9)

де $S_{b_1}^2 = \frac{S^2}{n(x^2 - \bar{x}^2)} = \frac{S^2}{n \text{var}(X)}, S_{b_0}^2 = \bar{x}^2 S_{b_1}^2$ – дисперсії параметрів b_1, b_0 ;

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2 \text{ – незміщена оцінка дисперсії залишків.}$$

Значення $\bar{x}^2, \bar{x}, \sum e_i^2$ можна знайти в лабораторній роботі 2.

Розраховані значення t -відношень порівняйте з відповідними значеннями, розташованими на листі ВИВОД ИТОГОВ;

2) з таблиць критичних точок розподілу Стьюдента знайти $t_{\alpha / 2, n-2}$ (α – рівень значущості, $(n - 2)$ – кількість степенів вільності);

3) якщо $|t_{b_i}| > t_{\alpha/2, n-2}$ – коефіцієнт b_i є статистично значущим.

2. Інтервали довіри для параметрів β_0, β_1 теоретичної регресії:

$$\bar{y}(x) = M(Y | X = x) = \beta_0 + \beta_1 x$$

визначаються за формулами:

$$b_0 - \Delta b_0 < \beta_0 < b_0 + \Delta b_0, \quad b_1 - \Delta b_1 < \beta_1 < b_1 + \Delta b_1, \quad (10)$$

де $\Delta b_0 = t_{\alpha/2, n-2} S_{b_0}, \Delta b_1 = t_{\alpha/2, n-2} S_{b_1}$ – граничні відхилення параметрів $b_0, b_1,$
 α – рівень значущості, зв'язаний з рівнем надійності P формулою: $\alpha = 1 - P$.

Розраховані інтервали довіри порівняйте з відповідними інтервалами на листі ВИВОД ИТОГОВ.

3. Точковий прогноз залежної змінної для значення $x_p = 1,2x_{max}$ розраховується за допомогою емпіричного рівняння регресії:

$$\hat{y}_p = b_0 + b_1 x_p. \quad (11)$$

Інтервальний прогноз для індивідуального значення залежної змінної визначається за формулою:

$$\hat{y}_p - \Delta y < y_p < \hat{y}_p + \Delta y, \quad (12)$$

де $\Delta y = t_{\alpha/2, n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \text{var}(X)}}$ – гранична похибка прогнозу.

4. **Середній частинний коефіцієнт еластичності** вказує, на скільки відсотків зміниться середнє значення показника Y , якщо середнє значення фактора X зміниться на один відсоток, і обчислюється за формулою:

$$\varepsilon = b_1 \bar{x} / \bar{y}. \quad (13)$$

Приклад виконання лабораторної роботи №3

1. За допомогою t -критерію Стьюдента перевіримо статистичну значущість параметрів b_0 і b_1 лінійної регресії, які визначені в лабораторній роботі 2.

Розрахуємо дисперсію S^2 та середнє квадратичне відхилення S залишків:

$$S^2 = \sum e_i^2 / (n - 2) = 10,011 / (10 - 2) = 1,2514,$$

$$\square S = \text{КОРЕНЬ}(S^2) = 1,1186.$$

Обчислимо дисперсії коефіцієнтів:

$$S_{b_1}^2 = S^2 / (n(X^2 - X^{\wedge 2})) = 1,2514 / (10 * (25,834 - 4,52^{\wedge 2})) = 0,0232,$$

$$S_{b_0}^2 = S_{b_1}^2 \sum x_i^2 / n = S_{b_1}^2 X^2 = 0,0232 * 25,834 = 0,5993.$$

Стандартні похибки коефіцієнтів:

$$S_{b_1} = \text{КОРЕНЬ}(S_{b_1}^2) = 0,1523,$$

$$S_{b_0} = \text{КОРЕНЬ}(S_{b_0}^2) = 0,7741.$$

Обчислимо t -статистики коефіцієнтів:

$$t_{b_1} = b_1 / S_{b_1} = 1,13 / 0,1523 = 7,4196;$$

$$t_{b_0} = b_0 / S_{b_0} = 2,15 / 0,7741 = 2,7774.$$

За таблицею t -розподілу Стьюдента для рівня значущості $\alpha = 0,05$ знайдемо критичне значення $t_{кр} = t(\alpha/2; n - 2) = t(0,025; 8) = 2,306$.

Оскільки $|t_{b_1}| > t_{кр}$, $|t_{b_0}| > t_{кр}$ – параметри b_1 , b_0 є статистично значущими.

2. Інтервали довіри для параметрів теоретичної лінійної регресії $\tilde{y} = \beta_0 + \beta_1 x$ визначаються за формулами:

$$b_0 - t_{кр} S_{b_0} < \beta_0 < b_0 + t_{кр} S_{b_0},$$

$$2,15 - 2,306 * 0,7741 < \beta_0 < 2,15 + 2,306 * 0,7741$$

$$0,36 < \beta_0 < 3,94.$$

$$b_1 - t_{кр} S_{b_1} < \beta_1 < b_1 + t_{кр} S_{b_1},$$

$$1,13 - 2,306 * 0,1523 < \beta_1 < 1,13 + 2,306 * 0,1523$$

$$0,7788 < \beta_1 < 1,48.$$

3. Точковий прогноз залежної змінної для $x_p = 1,2x_{max} = 1,2*10 = 12$:

$$Y_p = b_0 + b_1x_p = 2,15 + 1,13*12 = 15,71.$$

Інтервальний прогноз для індивідуального значення залежної змінної визначимо за формулою:

$$Y_p - \Delta Y < y_p < Y_p + \Delta Y,$$

де $\Delta Y = t_{кр} S \text{КОРЕНЬ}(1 + 1/n + (x_p - X)^2/(n(X^2 - X^2))) =$

$$= 2,306*1,1186* \text{КОРЕНЬ}(1 + 1/10 + (12 - 4,52)^2/(10*(25,834 - 4,52^2))) = \\ = 3,7694 - \text{гранична похибка прогнозу.}$$

87,39 % інтервал довіри для індивідуального значення залежної змінної:

$$11,9406 < y_p < 19,4794.$$

4. Середній частинний коефіцієнт еластичності визначимо за формулою:

$$\varepsilon = b_1X/Y = 1,13*4,52/7,27 = 0,7026.$$

Розраховане значення коефіцієнта еластичності свідчить про те, що зміна середнього значення поточного періоду на 1% призведе до зміни середнього значення витрат на 1 грн. виробленої продукції на 0,7 %.

Контрольні запитання

1. Перелічіть передумови МНК (умови Гаусса-Маркова), сформулюйте теорему Гаусса-Маркова.
2. Наведіть формули для оцінок дисперсії залишків та коефіцієнтів парної регресії.
3. Опишіть процес перевірки статистичної значущості коефіцієнтів b_0 та b_1 лінійної регресії за допомогою t -тесту Стьюдента.
4. Як визначаються *інтервали довіри* для параметрів β_0 , β_1 теоретичної регресії?
5. Як визначається точковий прогноз для залежної змінної?
6. Як побудувати *інтервал довіри* для прогнозованого значення залежної змінної?

ЛАБОРАТОРНА РОБОТА 4

Тема: «Побудова та аналіз якості багатofакторної лінійної регресії»

Мета роботи: за методом найменших квадратів навчитися будувати рівняння багатofакторної лінійної регресії та оцінювати його якість.

Завдання: За даними спостережень величин прибутку (y), фондівдачі (x_1), продуктивності праці (x_2) та інвестицій (x_3) для 11 однотипних підприємств необхідно:

1. Визначити параметри лінійної моделі залежності прибутку (y) від рівня фондівдачі (x_1), продуктивності праці (x_2) та інвестицій (x_3):

$$\hat{y}(x) = b_0 + b_1x_1 + b_2x_2 + b_3x_3. \quad (14)$$

2. Розрахувати коефіцієнт детермінації та оцінити його статистичну значущість за F -критерієм Фішера.
3. За критерієм Дарбіна-Уотсона перевірити наявність автокореляції залишків.
4. За алгоритмом Фаррара-Глобера перевірити наявність мультиколінеарності пояснюючих змінних.
5. Зробити точковий та інтервальний прогнози для залежної змінної, якщо прогнозні значення факторів складають 110 % від їхніх максимальних значень.

Методика виконання

1. Методом найменших квадратів розрахуйте параметри рівняння (14). Згідно з МНК параметри b_0, b_1, b_2, b_3 рівняння (14) визначаються так, щоб мінімізувати суму квадратів відхилень фактичних значень y_i від теоретичних \hat{y}_i :

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (b_0 + \sum_{j=1}^3 b_j x_{ji}))^2 = f(b_0, b_1, b_2, b_3) \rightarrow \min \quad (15)$$

Прирівнявши до нуля частинні похідні функції (15) за параметрами b_0, b_1, b_2, b_3 , дістанемо систему рівнянь:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^n x_{1i} + b_2 \sum_{i=1}^n x_{2i} + b_3 \sum_{i=1}^n x_{3i} = \sum_{i=1}^n y_i \\ b_0 \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i}x_{2i} + b_3 \sum_{i=1}^n x_{1i}x_{3i} = \sum_{i=1}^n x_{1i}y_i \\ b_0 \sum_{i=1}^n x_{2i} + b_1 \sum_{i=1}^n x_{1i}x_{2i} + b_2 \sum_{i=1}^n x_{2i}^2 + b_3 \sum_{i=1}^n x_{2i}x_{3i} = \sum_{i=1}^n x_{2i}y_i \\ b_0 \sum_{i=1}^n x_{3i} + b_1 \sum_{i=1}^n x_{1i}x_{3i} + b_2 \sum_{i=1}^n x_{2i}x_{3i} + b_3 \sum_{i=1}^n x_{3i}^2 = \sum_{i=1}^n x_{3i}y_i \end{cases} \quad (16)$$

У матричній формі система (16) набуде вигляду:

$$ZB = C, \quad (17)$$

де Z – матриця коефіцієнтів при невідомих;

C – матриця вільних членів рівнянь системи;

B – матриця невідомих коефіцієнтів рівняння (14):

$$Z = \begin{pmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{3i} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & \sum_{i=1}^n x_{1i}x_{3i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{2i}x_{1i} & \sum_{i=1}^n x_{2i}^2 & \sum_{i=1}^n x_{2i}x_{3i} \\ \sum_{i=1}^n x_{3i} & \sum_{i=1}^n x_{3i}x_{1i} & \sum_{i=1}^n x_{3i}x_{2i} & \sum_{i=1}^n x_{3i}^2 \end{pmatrix}, \quad C = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{1i} \\ \sum_{i=1}^n y_i x_{2i} \\ \sum_{i=1}^n y_i x_{3i} \end{pmatrix}, \quad B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

Зауваження. Матриці Z і C можна обчислити за формулами:

$$Z = X^T X, \quad C = X^T Y, \quad (18)$$

$$\text{де } X = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ x_{31} & x_{32} & \dots & x_{3n} \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Якщо визначник матриці Z не дорівнює нулю, з рівняння (17) можна визначити матрицю B :

$$B = Z^{-1} C, \quad (19)$$

де Z^{-1} – матриця, обернена до матриці Z .

Матрицю Z^{-1} можна обчислити за допомогою функції **=МОБР(масив)** програми *Excel*. Для цього необхідно:

- виділити діапазон клітинок 4×4 для розміщення матриці Z^{-1} ;
- виконати команди **Вставка-Функція...**, у списку категорій функцій вибрати **Математические**, у списку функцій – **МОБР**;
- у поле діалогового вікна функції **МОБР** ввести адресу діапазону клітинок, що містять елементи матриці Z ;
- натиснути комбінацію клавіш **Ctrl+Shift** та, утримуючи ці клавіші, натиснути на **Enter**. Після цього функція **МОБР** буде взята у фігурні дужки (*наприклад*: **{=МОБР(B40:E43)}**);
- натиснути на **ОК**.

Помножити матриці Z^{-1} та C можна за допомогою функції **=МУМНОЖ(масив1; масив2)**.

Для цього необхідно:

- виділити діапазон клітинок для розміщення елементів матриці B ;
- виконати команди **Вставка-Функція...**, у списку категорій функцій вибрати **Математические**, у списку функцій – **МУМНОЖ**;

- у поля діалогового вікна функції **МУМНОЖ** ввести адреси діапазонів клітинок, що містять елементи матриць Z^{-1} та C відповідно;
- натиснути комбінацію клавіш **Ctrl+Shift** та, утримуючи ці клавіші, натиснути на **Enter**, потім – на **OK**.

2. Коефіцієнт детермінації розраховується за формулою (7).

Аналіз статистичної значущості коефіцієнта детермінації

1) розрахуйте F -статистику:

$$F = \frac{R^2/m}{(1-R^2)/(n-m-1)}, \quad (20)$$

де $m = 3$ – кількість незалежних змінних;

- 2) для заданого рівня значущості α , кількості незалежних змінних $m = 3$ із таблиць критичних точок розподілу Фішера знайдіть $F_{кр} = F_{\alpha; m, n-m-1}$;
- 3) якщо $F > F_{кр}$, то коефіцієнт детермінації R^2 є статистично значущим, рівняння якісно описує зв'язок між залежною і незалежними змінними.

3. Перевірка наявності автокореляції залишків за критерієм Дарбіна-Уотсона.

Важливою передумовою побудови якісної регресійної моделі за МНК є незалежність значень випадкових відхилень ε_i від значень відхилень у всіх інших спостереженнях ε_j , тобто відсутність автокореляції залишків ($\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$). Для перевірки наявності автокореляції 1-го порядку (кореляції між сусідніми залишками) за критерієм Дарбіна-Уотсона необхідно:

- 1) визначити значення відхилень $e_i = y_i - \hat{y}_i$ для $i = 1, 2, \dots, n$;
- 2) розрахувати статистику Дарбіна-Уотсона:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})}{\sum_{i=1}^n e_i^2} \quad (21)$$

3) за таблицею критичних точок критерію Дарбіна-Уотсона визначити нижню d_l та верхню d_u критичні межі для рівня значущості α , кількості спостережень n , та пояснювальних змінних $m = 3$ і здійснити висновки за правилом:

$0 \leq DW < d_l$ – існує додатна автокореляція;

$d_l \leq DW < d_u$ – висновок про наявність автокореляції не визначений;

$d_u < DW < 4 - d_u$ – автокореляція відсутня;

$4 - d_u \leq DW \leq 4 - d_l$ – висновок про наявність автокореляції не визначений;
 $4 - d_l < DW \leq 4 - d_u$ – існує від’ємна автокореляція.

4. Перевірка наявності мультиколінеарності за алгоритмом Фаррара-Глобера.

Однією з проблем при побудові моделі багатофакторної регресії за методом найменших квадратів є *мультиколінеарність – існування тісної лінійної залежності, або сильної кореляції, між двома чи більше незалежними змінними*. За наявності мультиколінеарності знижується точність оцінювання параметрів моделі, оскільки детермінант матриці Z наближається до нуля.

Найповніше дослідити мультиколінеарність можна за допомогою алгоритму Фаррара-Глобера, згідно з яким перевіряється мультиколінеарність:

- 1) усього масиву пояснюючих змінних (за критерієм χ^2);
- 2) кожної пояснюючої змінної з усіма іншими (за F -критерієм);
- 3) кожної пари пояснюючих змінних (за t -критерієм).

4.1. Перевірка мультиколінеарності усього масиву пояснюючих змінних.

Розрахуйте матрицю парних коефіцієнтів кореляції (кореляційну матрицю), яка у випадку трьох незалежних факторів має вигляд:

$$r = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix}, \quad (22)$$

де $r_{jk} = \frac{\overline{x_j x_k} - \bar{x}_j \bar{x}_k}{S_j S_k}$ – коефіцієнт кореляції між змінними x_j, x_k ,

$S_j = \sqrt{\overline{x_j^2} - \bar{x}_j^2}$ – стандартні відхилення незалежних факторів,

$\overline{x_j x_k} = \frac{1}{n} \sum_{i=1}^n x_{ij} x_{ik}$, $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ji}$, $\overline{x_j^2} = \frac{1}{n} \sum_{i=1}^n x_{ij}^2$ – середні значення.

Визначник матриці r позначимо $|r|$ ($|r| \in [0, 1]$). Якщо існує повна мультиколінеарність ($r_{jk} = 1$), то $|r| = 0$. Якщо мультиколінеарність відсутня ($r_{jk} = 0$), то $|r| = 1$. Чим ближче $|r|$ до нуля, тим з більшою впевненістю можна стверджувати, що між пояснювальними змінними існує мультиколінеарність.

Для перевірки відсутності мультиколінеарності усього масиву пояснюючих змінних (гіпотези $|r| = 1$) необхідно:

- розрахувати статистику χ^2 (“хі”-квадрат):

$$\chi^2 = -[n-1-(2m+5)/6]\ln\|r\|, \quad (23)$$

де $|r|$ – визначник кореляційної матриці r , який можна обчислити за допомогою функції =МОПРЕД(*масив*) програми *Excel*;

- з таблиць критичних точок розподілу χ^2 знайти $\chi^2_{кр} = \chi^2_{\alpha\nu}$, де α – рівень значущості, $\nu = m(m-1)/2$ – кількість степенів вільності;
- якщо розраховане значення $\chi^2 < \chi^2_{кр}$, то в масиві пояснювальних змінних існує мультиколінеарність.

4.2. Перевірка мультиколінеарності між однією пояснювальною змінною та іншими.

Розрахуйте матрицю r^* , обернену до матриці r (це можна зробити за допомогою функції =МОБР(*масив*) програми *Excel*):

$$r^* = r^{-1} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}. \quad (24)$$

Для кожної пояснюючої змінної x_j розрахуйте частинний коефіцієнт детермінації R_j^2 (показник тісноти лінійного зв’язку x_j з іншими змінними):

$$R_j^2 = 1 - 1/c_{jj}, \quad (25)$$

де c_{jj} – діагональні елементи матриці r^* .

Перевірте статистичну значущість кожного частинного коефіцієнта детермінації R_j^2 за F -критерієм. Для цього:

- розрахуйте F_j -статистику:

$$F_j = \frac{R_j^2/(m-1)}{(1-R_j^2)/(n-m)} = (c_{jj} - 1) \frac{n-m}{m-1}, \quad (26)$$

де n – кількість спостережень;

m – кількість пояснюючих змінних у початковому рівнянні регресії;

- з таблиць критичних точок F -розподілу знайдіть $F_{кр} = F(\alpha; m-1; n-m)$, де α – рівень значущості, $(m-1)$, $(n-m)$ – ступені вільності;
- якщо $F_j > F_{кр}$, то змінна x_j мультиколінеарна з іншими пояснюючими змінними.

4.3. Перевірка тісноти парної кореляції пояснюючих факторів.

Для перевірки колінеарності факторів x_i , x_j ($i \neq j$) аналізується статистична значущість частинних коефіцієнтів кореляції, які

характеризують тісноту зв'язку між двома пояснюючими змінними за умови, що інші змінні не впливають на цей зв'язок.

Розрахуйте частинні коефіцієнти кореляції між змінними x_i та x_j :

$$r_{ij.s} = \frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}} \Rightarrow r_{12.3} = \frac{-c_{12}}{\sqrt{c_{11}c_{22}}}, \quad r_{13.2} = \frac{-c_{13}}{\sqrt{c_{11}c_{33}}}, \quad r_{23.1} = \frac{-c_{23}}{\sqrt{c_{22}c_{33}}}, \quad (27)$$

де c_{ij} – елементи матриці r^* (24), оберненої до матриці r .

У випадку 3-х незалежних факторів частинні коефіцієнти кореляції можна обчислити за формулами:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}, \quad r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}}, \quad r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}}, \quad (28)$$

Якщо значення $|r_{ij.s}|$ близьке до 1, це вказує на колінеарність факторів x_i, x_j . Для перевірки наявності колінеарності факторів x_i, x_j необхідно:

- розрахувати t -відношення: $t_{ij} = \frac{r_{ij.s} \sqrt{n-k}}{\sqrt{1-r_{ij.s}^2}}$, де $k = 4$ – кількість параметрів (b_0, b_1, b_2, b_3);
- з таблиць критичних точок розподілу Стюдента знайти $t_{\alpha/2, n-k}$ (α – заданий рівень значущості, $n - k$ – кількість ступенів вільності);
- якщо $|t_{ij}| > t_{\alpha/2, n-k}$ – коефіцієнт $r_{ij.s}$ є статистично значущим, між пояснюючими змінними x_i, x_j існує колінеарність.

5. Точковий та інтервальний прогнози для залежної змінної.

Розрахуйте прогнозні значення факторів, які складають 110 % від їхніх максимальних значень: $x_{p1} = 1,1x_{1max}$, $x_{p2} = 1,1x_{2max}$, $x_{p3} = 1,1x_{3max}$.

Точковий прогноз розраховується підстановкою прогнозних значень факторів, заданих матрицею $X_p = (1, x_{p1}, x_{p2}, x_{p3})$, в рівняння регресії:

$$\hat{y}_p = b + b_1 x_{p1} + b_2 x_{p2} + b_3 x_{p3} \Leftrightarrow \hat{y}_p = X_p \quad (29)$$

Дисперсія точкового прогнозу має вигляд:

$$D(\hat{y}_p) \approx S_{\hat{y}_p}^2 = S^2 X_p^T (X^T X)^{-1} X_p, \quad (30)$$

де $S^2 = \frac{1}{n-k} \sum_{i=1}^n e_i^2$ – незміщена оцінка дисперсії залишків ($k = 4$ – кількість

параметрів); $X_p^T = \begin{pmatrix} 1 \\ x_{p1} \\ x_{p2} \\ x_{p3} \end{pmatrix}$ – матриця, транспонована до матриці X_p .

Інтервальний прогноз для індивідуального значення y_p залежної змінної розраховується за формулою:

$$\hat{y}_p - \Delta y < y_p < \hat{y}_p + \Delta y, \quad (31)$$

де $\Delta y = t_{\alpha/2, n-m-1} S \sqrt{1 + X_p (X^T X)^{-1} X_p^T}$ - гранична похибка прогнозу; матриця $(X^T X)^{-1} = Z^{-1}$ знайдена раніше в п. 1.

Завдання до лабораторної роботи №4

Вибір завдання: К – остання цифра, N – передостання цифра варіанта.

Таблиця 8

i	y	x_1	x_2	x_3
1	17,08 + К	11,08 + К	9,71 + N	13,02 + N
2	15,96 + К	13,15 + К	9,62 + N	12,66 + N
3	17,44 + К	13,30 + К	9,86 + N	14,68 + N
4	20,54 + К	12,29 + К	10,20 + N	17,48 + N
5	23,98 + К	13,11 + К	10,93 + N	22,50 + N
6	23,42 + К	12,71 + К	11,02 + N	21,86 + N
7	25,98 + К	13,11 + К	9,34 + N	24,22 + N
8	26,06 + К	13,33 + К	12,10 + N	23,86 + N
9	27,42 + К	13,38 + К	12,53 + N	27,94 + N
10	31,82 + К	15,70 + К	11,55 + N	27,68 + N
11	33,62 + К	15,79 + К	11,97 + N	28,50 + N

Приклад виконання лабораторної роботи №4

1. Визначимо параметри лінійної моделі залежності прибутку (y) від рівня фондівдачі (x_1), продуктивності праці (x_2) та інвестицій (x_3):

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$$

на підставі даних 11 спостережень, наведених у табл. 9.

Таблиця 9

i	y	x_1	x_2	x_3
1	17,08	11,08	18,71	22,02
2	15,96	13,15	18,62	21,66
3	17,44	13,3	18,86	23,68
4	20,54	12,29	19,2	26,48
5	23,98	13,11	19,93	31,5
6	23,42	12,71	20,02	30,86
7	25,98	13,11	18,34	33,22
8	26,06	13,33	21,1	32,86
9	27,42	13,38	21,53	36,94
10	31,82	15,7	20,55	36,68
11	33,62	15,79	20,97	37,5
Σ	263,32	146,95	217,83	333,4
Σ/n	23,9382	13,3591	19,8027	30,3091

Коефіцієнти рівняння регресії визначаються із системи рівнянь (16) представленої в матричній формі (17). Коефіцієнти регресії знайдемо за формулою (19).

$$X = \begin{pmatrix} 1 & 11,08 & 18,71 & 22,02 \\ 1 & 13,15 & 18,62 & 21,66 \\ 1 & 13,3 & 18,86 & 23,68 \\ 1 & 12,29 & 19,2 & 26,48 \\ 1 & 13,11 & 19,93 & 31,5 \\ 1 & 12,71 & 20,02 & 30,86 \\ 1 & 13,11 & 18,34 & 33,22 \\ 1 & 13,33 & 21,1 & 32,86 \\ 1 & 13,38 & 21,53 & 36,94 \\ 1 & 15,7 & 20,55 & 36,68 \\ 1 & 15,79 & 20,97 & 37,5 \end{pmatrix} \quad Y = \begin{pmatrix} 17,08 \\ 15,96 \\ 17,44 \\ 20,54 \\ 23,98 \\ 23,42 \\ 25,98 \\ 26,06 \\ 27,42 \\ 31,82 \\ 33,62 \end{pmatrix}$$

Транспонуємо матрицю X за допомогою програми **ТРАНСП**(масив) пакета *Excel*:

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 11,08 & 13,15 & 13,3 & 12,29 & 13,11 & 12,71 & 13,11 & 13,33 & 13,38 & 15,7 & 15,79 \\ 18,71 & 18,62 & 18,86 & 19,2 & 19,93 & 20,02 & 18,34 & 21,1 & 21,53 & 20,55 & 20,97 \\ 22,02 & 21,66 & 23,68 & 26,48 & 31,5 & 30,86 & 33,22 & 32,86 & 36,94 & 36,68 & 37,5 \end{pmatrix}$$

За допомогою стандартної програми **МУМНОЖ**(масив1; масив2) пакета *Excel* знайдемо матриці Z та C :

$$Z = X^T X = \begin{pmatrix} 11 & 146,95 & 217,83 & 333,4 \\ 146,95 & 1981,439 & 2918,225 & 4510,186 \\ 217,83 & 2918,225 & 4326,263 & 6654,004 \\ 333,4 & 4510,186 & 6654,004 & 10460,14 \end{pmatrix}.$$

$$C = X^T Y = \begin{pmatrix} 263,32 \\ 3580,846 \\ 5262,422 \\ 8316,974 \end{pmatrix}.$$

Обернену матрицю Z^{-1} знайдемо за допомогою стандартної програми **МОБР**(масив) пакета *Excel*:

$$Z^{-1} = \begin{pmatrix} 56,93297 & -0,90631 & -3,02274 & 0,49899 \\ -0,90631 & 0,106326 & -0,00023 & -0,01681 \\ -3,02274 & -0,00023 & 0,196619 & -0,02863 \\ 0,49899 & -0,01681 & -0,02863 & 0,009655 \end{pmatrix}.$$

За формулою (19) за допомогою стандартної програми *МУМНОЖ*(массив1; массив2) пакета *Excel* знайдемо матрицю *B*:

$$B = Z^{-1}C = \begin{pmatrix} -10,6202 \\ 1,052112 \\ -0,20297 \\ 0,809077 \end{pmatrix}.$$

Лінійна модель залежності величини прибутку y від фондівіддачі x_1 , продуктивності праці x_2 та інвестицій x_3 набуде вигляду:

$$y = -10,62 + 1,05x_1 - 0,2x_2 + 0,81x_3.$$

2. Для визначення коефіцієнта детермінації і перевірки відсутності автокореляції залишків наведемо таблицю розрахунків.

Таблиця 10

i	y_i	x_{1i}	x_{2i}	x_{3i}	$y_i - Y$	$(y_i - Y)^2$	$Y_i = b_0 + \sum b_j x_{ji}$	$e_i = y_i - Y_i$	e_i^2	$e_i - e_{i-1}$	$(e_i - e_{i-1})^2$
1	17,08	11,08	18,71	22,02	-6,8582	47,0347	15,0556	2,0244	4,0982		
2	15,96	13,15	18,62	21,66	-7,9782	63,6514	16,9605	-1,0005	1,0009	-3,0249	9,1498
3	17,44	13,3	18,86	23,68	-6,4982	42,2264	18,7039	-1,2639	1,5975	-0,2634	0,0694
4	20,54	12,29	19,2	26,48	-3,3982	11,5476	19,8377	0,7023	0,4933	1,9662	3,8660
5	23,98	13,11	19,93	31,5	0,0418	0,0017	24,6138	-0,6338	0,4017	-1,3361	1,7853
6	23,42	12,71	20,02	30,86	-0,5182	0,2685	23,6569	-0,2369	0,0561	0,3969	0,1575
7	25,98	13,11	18,34	33,22	2,0418	4,1690	26,3281	-0,3481	0,1212	-0,1113	0,0124
8	26,06	13,33	21,1	32,86	2,1218	4,5021	25,7082	0,3518	0,1238	0,7000	0,4900
9	27,42	13,38	21,53	36,94	3,4818	12,1231	28,9745	-1,5545	2,4165	-1,9064	3,6342
10	31,82	15,7	20,55	36,68	7,8818	62,1231	31,4040	0,4160	0,1731	1,9706	3,8831
11	33,62	15,79	20,97	37,5	9,6818	93,7376	32,0769	1,5431	2,3813	1,1271	1,2704
Σ	263,32	146,95	217,83	333,4	SST	341,3852		SSE	12,8636		24,3181
	Y	X1	X2	X3							
Σ/n	23,938	13,359	19,803	30,309							

Коефіцієнт детермінації: $R^2 = 1 - SSE/SST = 0,9623$ – 96 % варіації величини y пояснюється варіацією змінних x_1, x_2, x_3 .

Для перевірки статистичної значущості коефіцієнта детермінації обчислимо F -статистику Фішера:

$$F = R^2(n - m - 1) / ((1 - R^2)m) = 59,59,$$

де $n = 11$ – кількість спостережень;

$m = 3$ – кількість незалежних змінних.

З таблиці критичних точок F -розподілу для рівня значущості $\alpha = 0,05$ та кількості степенів вільності $k_1 = m = 3$, $k_2 = n - m - 1 = 7$ знайдемо критичне значення $F_{кр} = F_{кр}(\alpha; k_1; k_2) = F_{кр}(0,05; 3; 7) = 4,35$.

Оскільки $F > F_{кр}$, то з надійністю $1 - 0,05 = 0,95$ можна стверджувати, що коефіцієнт детермінації є статистично значущим.

3. Для перевірки відсутності автокореляції залишків за критерієм Дарбіна-Уотсона обчислимо DW -статистику:

$$DW = \Sigma(e_i - e_{i-1})^2 / \Sigma e_i^2 = 1,89.$$

З таблиць критичних точок критерію Дарбіна-Уотсона для рівня значущості $\alpha = 0,05$, кількості спостережень $n = 11$ та пояснюючих змінних $m=3$ визначимо нижню $d_l = 0,585$ і верхню $d_u = 1,928$ критичні межі.

Оскільки $d_l < DW < d_u$ – висновок про наявність автокореляції залишків не визначений.

4. Перевіримо наявність мультиколінеарності факторів. Розрахуємо парні коефіцієнти кореляції (табл):

$$r_{12} = \Sigma(x_{1i} - X_1)(x_{2i} - X_2) / \text{КОРЕНЬ}(\Sigma(x_{1i} - X_1)^2 \Sigma(x_{2i} - X_2)^2) = 0,5399,$$

$$r_{13} = \Sigma(x_{1i} - X_1)(x_{3i} - X_3) / \text{КОРЕНЬ}(\Sigma(x_{1i} - X_1)^2 \Sigma(x_{3i} - X_3)^2) = 0,6976,$$

$$r_{23} = \Sigma(x_{2i} - X_2)(x_{3i} - X_3) / \text{КОРЕНЬ}(\Sigma(x_{2i} - X_2)^2 \Sigma(x_{3i} - X_3)^2) = 0,7730.$$

Матриця парних коефіцієнтів кореляції:

$$r = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0,5399 & 0,6976 \\ 0,5399 & 1 & 0,7730 \\ 0,6976 & 0,7730 & 1 \end{pmatrix}.$$

Таблиця 11

$x_{1i} - X_1$	$(x_{1i} - X_1)^2$	$X_{2i} - X_2$	$(x_{2i} - X_2)^2$	$X_{3i} - X_3$	$(x_{3i} - X_3)^2$	$(x_{1i} - X_1) \times (x_{2i} - X_2)$	$(x_{1i} - X_1) \times (x_{3i} - X_3)$	$(x_{2i} - X_2) \times (x_{3i} - X_3)$
-2,2791	5,1943	-1,0930	1,1946	-8,2890	68,7075	2,4910	18,8914	9,0599
-0,2091	0,0437	-1,1830	1,3995	-8,6490	74,8052	0,2474	1,8084	10,2318
-0,0591	0,0035	-0,9430	0,8892	-6,6290	43,9436	0,0557	0,3917	6,2511
-1,0691	1,1430	-0,6030	0,3636	-3,8290	14,6612	0,6447	4,0935	2,3089
-0,2491	0,0620	0,1270	0,0161	1,1910	1,4185	-0,0316	-0,2967	0,1513
-0,6491	0,4213	0,2170	0,0471	0,5510	0,3036	-0,1409	-0,3576	0,1196
-0,2490	0,0620	-1,4630	2,1404	2,9110	8,4739	0,3643	-0,7248	-4,2588
-0,0290	0,0008	1,2970	1,6822	2,5510	6,5076	-0,0376	-0,0740	3,3086
0,0210	0,0004	1,7270	2,9825	6,6310	43,9702	0,0363	0,1393	11,4517
2,3410	5,4803	0,7470	0,5580	6,3710	40,5896	1,7487	14,9145	4,7591
2,4310	5,9098	1,1670	1,3619	7,1910	51,7105	2,8370	17,4813	8,3919
0,0005	18,3211	-0,0030	12,6352	0,0010	355,0915	8,2149	56,2670	51,7751

Зауваження. В останньому рядку табл. 11 містяться суми величин, розташованих у відповідних стовпцях. Середні значення X_1 , X_2 , X_3 розраховані в останньому рядку табл. 10.

За допомогою функції **МОПРЕД**(массив) знайдемо визначник кореляційної матриці: $|r| = 0,2066$.

4.1. Перевірка мультиколінеарності усього масиву пояснюючих змінних.

Розрахуємо “хі”-квадрат статистику:

$$\chi^2 = -[n - 1 - (2m + 5)/6] \ln ||r|| = 12,8769.$$

З таблиці критичних точок розподілу χ^2 для рівня значущості $\alpha=0,05$ та кількості степенів вільності $\nu = m(m - 1)/2 = 3(3 - 1)/2 = 3$ знайдемо $\chi^2_{кр}=7,8$.

Оскільки розраховане значення $\chi^2 > \chi^2_{кр}$, то в масиві пояснювальних змінних існує мультиколінеарність.

4.2. Перевірка мультиколінеарності між однією пояснювальною змінною та іншими.

Обчислимо матрицю r^* , обернену до матриці r , за допомогою функції **МОБР**(массив):

$$r^* = r^{-1} = \begin{pmatrix} 1,9479 & -0,0034 & -1,3562 \\ -0,0034 & 2,4843 & -1,9179 \\ -1,3562 & -1,9179 & 3,4285 \end{pmatrix}.$$

Розрахуємо частинні коефіцієнти детермінації $R_j^2 = 1 - 1/c_{jj}$, та F_j -статистику для кожної пояснюючої змінної (c_{jj} – діагональні елементи матриці r^*):

$$R_1^2 = 1 - 1/c_{11} = 1 - 1/1,9479 = 0,4866;$$

$$F_1 = (c_{11} - 1)(n - m)/(m - 1) = 3,7917;$$

$$R_2^2 = 1 - 1/c_{22} = 1 - 1/2,4843 = 0,5975;$$

$$F_2 = (c_{22} - 1)(n - m)/(m - 1) = 5,9373;$$

$$R_3^2 = 1 - 1/c_{33} = 1 - 1/3,4285 = 0,7083;$$

$$F_3 = (c_{33} - 1)(n - m)/(m - 1) = 9,7141.$$

З таблиці критичних точок F -розподілу для рівня значущості $\alpha = 0,05$ та кількості степенів вільності $k_1 = m - 1 = 2$, $k_2 = n - m = 8$ знайдемо критичне значення $F_{кр}(\alpha; k_1; k_2) = F_{кр}(0,05; 2; 8) = 4,46$.

З надійністю $\gamma = 1 - 0,05 = 0,95$ можна стверджувати таке:

а) $F_1 < F_{кр}$ – коефіцієнт детермінації R_1^2 не є статистично значущим, змінна x_1 не мультиколінеарна з x_2 , x_3 ;

б) $F_2 > F_{кр}$ – коефіцієнт детермінації R_2^2 є статистично значущим, змінна x_2 мультиколінеарна з x_1 , x_3 ;

в) $F_3 > F_{кр}$ – коефіцієнт детермінації R_3^2 є статистично значущим, змінна x_3 мультиколінеарна з x_1 , x_2 .

4.3. Перевірка тісноти парної кореляції пояснюючих факторів.

Розрахуємо частинні коефіцієнти кореляції (c_{ij} – елементи матриці r^*):

$$r_{12.3} = -c_{12}/\text{КОРЕНЬ}(c_{11}c_{22}) = 0,0016;$$

$$r_{13.2} = -c_{13}/\text{КОРЕНЬ}(c_{11}c_{33}) = 0,5248;$$

$$r_{23.1} = -c_{23}/\text{КОРЕНЬ}(c_{22}c_{33}) = 0,6572.$$

Перевіримо статистичну значущість частинних коефіцієнтів. Обчислимо t -статистики:

$$t_{12} = r_{12.3}\text{КОРЕНЬ}((n - k)/(1 - r_{12.32})) = 0,0041;$$

$$t_{13} = r_{13.2}\text{КОРЕНЬ}((n - k)/(1 - r_{13.22})) = 1,6311;$$

$$t_{23} = r_{23.1}\text{КОРЕНЬ}((n - k)/(1 - r_{23.12})) = 2,3067,$$

де $k = 4$ – кількість параметрів.

З таблиць критичних точок розподілу Стьюдента знайдемо:

$$t_{\text{кр}}(\alpha/2; n - k) = t_{\text{кр}}(0,025; 7) = 2,365.$$

Оскільки $|t_{ij}| < t_{\text{кр}}$ – коефіцієнти $r_{ij,S}$ не є статистично значущими, тобто між незалежними змінними немає колінеарності.

Висновок: змінну x_3 слід вилучити з переліку змінних, або оцінювати параметри моделі на основі нормалізованих даних, або за методом головних компонентів.

5. Точковий та інтервальний прогнози для залежної змінної.

Розрахуємо прогнозні значення факторів, що складають 110% їхніх максимальних значень:

$$x_{p1} = 1,1x_{1\text{max}} = 17,37; x_{p2} = 1,1x_{2\text{max}} = 23,68; x_{p3} = 1,1x_{3\text{max}} = 41,25; \square$$

$$X_p = (1 \ x_{p1} \ x_{p2} \ x_{p3}) = (1 \ 17,369 \ 23,683 \ 41,25).$$

Точковий прогноз залежної змінної розрахуємо за допомогою рівняння регресії:

$$Y_p = b_0 + b_1x_{p1} + b_2x_{p2} + b_3x_{p3} = 36,2215.$$

$$\text{Оцінка дисперсії залишків: } S^2 = \text{SSE}/(n - k) = 1,608 \quad \square \quad S = 1,268.$$

За допомогою стандартної програми **МУМНОЖ**(массив1; массив2) пакета *Excel* знайдемо добуток матриць:

$$\begin{aligned} X_p(X^T X)^{-1} X_p^T &= X_p Z^1 X_p^t = \\ &= (1 \ 17,369 \ 23,683 \ 41,25) \begin{pmatrix} 56,93297 & -0,90631 & -3,02274 & 0,49899 \\ -0,90631 & 0,106326 & -0,00023 & -0,01681 \\ -3,02274 & -0,00023 & 0,196619 & -0,02863 \\ 0,49899 & -0,01681 & -0,02863 & 0,009655 \end{pmatrix} \begin{pmatrix} 1 \\ 17,369 \\ 23,683 \\ 41,25 \end{pmatrix} = \\ &= (-0,813 \ 0,2415 \ 0,4488 \ -0,0729) \begin{pmatrix} 1 \\ 17,369 \\ 23,683 \\ 41,25 \end{pmatrix} = 2,0032. \end{aligned}$$

З таблиць критичних точок розподілу Стьюдента знайдемо:

$$t_{кр}(\alpha/2; n - k) = t_{кр}(0,025; 7) = 2,365.$$

Інтервальний прогноз для індивідуального значення залежної змінної розрахуємо за формулою:

$$Y_p - \Delta Y < y_p < Y_p + \Delta Y,$$

де $\Delta Y = t_{кр}S \cdot \text{КОРЕНЬ}(1 + X_p(X^T X)^{-1} X_p^T) = 5,1971$ – гранична похибка прогнозу.

Інтервал довіри для індивідуального значення залежної змінної:

$$31,024 < y_p < 41,419.$$

Контрольні запитання

1. Сформулюйте означення багатofакторної лінійної регресії.
2. Напишіть теоретичне та емпіричне рівняння багатofакторної лінійної регресії. Поясніть різницю між ними.
3. Як визначити параметри багатofакторної лінійної регресії за методом найменших квадратів?
4. За допомогою яких показників можна оцінити якість багатofакторної лінійної регресії?
5. Для чого призначений, у якому діапазоні змінюється, за якою формулою розраховується коефіцієнт детермінації?
6. Як оцінити статистичну значущість коефіцієнта детермінації?
7. Як провести тестування наявності автокореляції за критерієм Дарбіна-Уотсона?
8. Сформулюйте означення мультиколінеарності та поясніть її вплив на оцінки параметрів моделі.
9. Які статистичні критерії використовуються для виявлення мультиколінеарності?
10. Як провести тестування наявності мультиколінеарності:
 - усього масиву пояснюючих змінних;
 - між однією пояснювальною змінною та іншими;
 - кожної пари пояснюючих змінних.
11. Як визначається точковий прогноз для залежної змінної?
12. Як побудувати *інтервал довіри* для прогнозованого значення залежної змінної?

Список літератури:

1. *Економетрія: Методичні вказівки щодо проведення лабораторно-практичних занять / Уклад.: В.М. Долгіх, Я.В. Долгіх. – Суми: УАБС НБУ, 2006. – 42 с.*
2. *Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: Підручник. – К.: КНЕУ, 2004. – 520 с.*
3. *Єлейко В.І. Основи економетрії: У 2 ч. – Львів: ТОВ «Марка ЛТД», 1995.*
4. *Лук'яненко І.Г., Краснікова Л. І. Економетрика: практикум з використанням комп'ютера. – К.: «Знання» КОО, 1998.*
5. *Лугінін О.Є. Економетрія: Навч. посіб. – К.: Центр учбової літератури, 2008. – 278 с.*
6. *Долгіх Я.В. Економетрія: методичні вказівки щодо проведення лабораторно-практичних занять. – Суми: УАБС НБУ, 2006. – 42 с.*
7. *Лециньський О.Л., Рязанцева В.В., Юнькова О.О. Економетрія: Навч. посіб. – К.: МАУП, 2003. – 208 с.*

ДОДАТКИ

Додаток 1

Критичні точки розподілення Стьюдента

		Рівень значущості α (двостоння критична область)						
		0,40	0,25	0,10	0,05	0,025	0,01	0,005
Число степеней вільності	1	0,325	1,000	3,078	6,314	12,706	31,821	63,657
	2	0,289	0,816	1,886	2,920	4,303	6,965	9,925
	3	0,277	0,765	1,638	2,353	3,182	4,541	5,841
	4	0,271	0,741	1,533	2,132	2,776	3,747	4,604
	5	0,267	0,727	1,476	2,015	2,571	3,365	4,032
	6	0,265	0,718	1,440	1,943	2,447	3,143	3,707
	7	0,263	0,711	1,415	1,895	2,365	2,998	3,499
	8	0,262	0,706	1,397	1,860	2,306	2,896	3,355
	9	0,261	0,703	1,383	1,833	2,262	2,821	3,250
	10	0,260	0,700	1,372	1,812	2,228	2,764	3,169
	11	0,260	0,697	1,363	1,796	2,201	2,718	3,106
	12	0,259	0,695	1,356	1,782	2,179	2,681	3,055
	13	0,259	0,694	1,350	1,771	2,160	2,650	3,012
	14	0,258	0,692	1,345	1,761	2,145	2,624	2,977
	15	0,258	0,691	1,341	1,753	2,131	2,602	2,947
	16	0,258	0,690	1,337	1,746	2,120	2,583	2,921
	17	0,257	0,689	1,333	1,740	2,110	2,567	2,898
	18	0,257	0,688	1,330	1,734	2,101	2,552	2,878
	19	0,257	0,688	1,328	1,729	2,093	2,539	2,861
	20	0,257	0,687	1,325	1,725	2,086	2,528	2,845
	21	0,257	0,686	1,323	1,721	2,080	2,518	2,831
	22	0,256	0,686	1,321	1,717	2,074	2,508	2,819
	23	0,256	0,685	1,319	1,714	2,069	2,500	2,807
	24	0,256	0,685	1,318	1,711	2,064	2,492	2,797
	25	0,256	0,684	1,316	1,708	2,060	2,485	2,787
	26	0,256	0,684	1,315	1,706	2,056	2,479	2,779
	27	0,256	0,684	1,314	1,703	2,052	2,473	2,771
	28	0,256	0,683	1,313	1,701	2,048	2,467	2,763
	29	0,256	0,683	1,311	1,699	2,045	2,462	2,756
	30	0,256	0,683	1,310	1,697	2,042	2,457	2,750
40	0,255	0,681	1,303	1,684	2,021	2,423	2,704	
50	0,255	0,680	1,296	1,676	2,009	2,403	2,678	
60	0,255	0,679	1,296	1,671	2,000	2,390	2,660	
80	0,254	0,679	1,292	1,664	1,990	2,374	2,639	
100	0,254	0,678	1,290	1,660	1,984	2,365	2,626	
120	0,254	0,677	1,289	1,658	1,980	2,358	2,467	
200	0,254	0,676	1,286	1,653	1,972	2,345	2,601	

Критичні точки розподілення χ^2

Число степеней вільності k	Рівень значущості α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,0001
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Розподілення Фішера (F-розподіл)

$\alpha = 0,10$		Число степеней вільності ν_1									
		1	2	3	4	5	6	7	8	9	10
Число степеней вільності ν_2	1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19
	2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39
	3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23
	4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92
	5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30
	6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94
	7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70
	8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54
	9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42
	10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32
	11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25
	12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19
	13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14
	14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10
	15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06
	16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03
	17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00
	18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98
	19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96
	20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	

Розподілення Фішера (F-розподіл)

$\alpha = 0,05$		Число степеней вільності ν_1									
		1	2	3	4	5	6	7	8	9	10
Число степеней вільності ν_2	1	161	200	216	225	230	234	237	239	271	242
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	

Распределение Дарбина-Уотсона

$\alpha = 0,01$ (n – объем выборки, m – число объясняющих переменных в уравнении регрессии)

n	$m=1$		$m=2$		$m=3$		$m=4$	
	$d1$	$d2$	$d1$	$d2$	$d1$	$d2$	$d1$	$d2$
6	0,390	1,142						
7	0,433	1,036	0,294	1,676				
8	0,497	1,003	0,343	1,489	0,229	2,102		
9	0,554	0,998	0,408	1,389	0,279	1,873	0,183	2,433
10	0,604	1,001	0,466	1,333	0,340	1,733	0,230	2,193
11	0,633	1,010	0,319	1,297	0,396	1,640	0,286	2,030
12	0,697	1,023	0,369	1,274	0,449	1,373	0,339	1,913
13	0,738	1,038	0,616	1,261	0,499	1,326	0,391	1,826
14	0,776	1,034	0,660	1,234	0,347	1,490	0,441	1,737
15	0,811	1,070	0,700	1,232	0,391	1,464	0,488	1,704
16	0,844	1,086	0,737	1,232	0,633	1,446	0,332	1,663
17	0,874	1,102	0,772	1,233	0,672	1,432	0,374	1,630
18	0,902	1,118	0,803	1,239	0,708	1,422	0,613	1,604
19	0,928	1,132	0,833	1,263	0,742	1,413	0,630	1,384
20	0,932	1,147	0,863	1,271	0,773	1,411	0,683	1,367
21	0,973	1,161	0,890	1,277	0,803	1,408	0,718	1,334
22	0,997	1,174	0,914	1,284	0,831	1,407	0,748	1,343
23	1,018	1,187	0,938	1,291	0,838	1,407	0,777	1,334
24	1,037	1,199	0,960	1,298	0,882	1,407	0,803	1,328
25	1,033	1,211	0,981	1,303	0,906	1,409	0,831	1,323
26	1,072	1,222	1,001	1,312	0,928	1,411	0,833	1,318
27	1,089	1,233	1,019	1,319	0,949	1,413	0,878	1,313
28	1,104	1,244	1,037	1,323	0,969	1,413	0,900	1,313
29	1,119	1,234	1,034	1,332	0,988	1,418	0,921	1,312
30	1,133	1,263	1,070	1,339	1,006	1,421	0,941	1,311
35	1,193	1,307	1,140	1,370	1,083	1,439	1,028	1,312
40	1,246	1,344	1,198	1,398	1,148	1,437	1,098	1,318
50	1,324	1,403	1,283	1,446	1,243	1,491	1,203	1,338
100	1,322	1,362	1,303	1,383	1,482	1,604	1,462	1,623

Навчально-методичне видання

ЕКОНОМЕТРІЯ

Методичні вказівки
до виконання лабораторних і практичних робіт
для студентів спеціальності 6.030601 «Менеджмент»
усіх форм навчання

Укладачі: **Федоренко** Наталія Дмитрівна
Білощицька Світлана Василівна
Білощицький Андрій Олександрович
Доля Олена Вікторівна
Теренчук Світлана Анатоліївна

Комп'ютерна верстка *А.П. Морозюк*

Підписано до друку 8.10. 2012. Формат 60x84
Ум. друк. арк. 1,16. Обл.-вид. арк. 2,5,
Тираж 30 прим. Вид. № 72/Ш-12 Зам. № 136/1-12

КНУБА, Повітрофлотський проспект, 31, Київ, Україна, 03680

E-mai: red_isdat@ua.fm

Видруковано в редакційно-видавничому відділі
Київського національного університету будівництва і архітектури

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи
ДК № 808 від 13.02.2002р.