МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД

ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ АГРАРНИЙ УНІВЕРСИТЕТ

Кафедра прикладної
математики та економічної
кібернетики

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

до виконання лабораторно-практичних робіт
з дисципліни: «Біометрія»
для здобувачів вищої освіти першого (бакалаврського) рівня
третього року
спеціальностей «лісове господарство», «садово–паркове господарство»

Змістова частина №2
Аналіз даних та моделювання зв’язку між випадковими величинами

Херсон – 2019

УДК 681.3.04

Розглянуто і схвалено на засіданні кафедри прикладної математики та економічної кібернетики ХДАУ (протокол №___ від «____»_________________2019р.)

Лобода О.М. Методичні рекомендації до виконання лабораторно-практичних робіт
з дисципліни: «Біометрія» для здобувачів вищої освіти факультету ФРГП спеціальностей «лісове господарство», «садово–паркове господарство». Змістова частина 2
Аналіз даних та моделювання зв'язку між випадковими величинами -53с.

Лабораторно-практичне заняття №1

Тема: Оцінювання параметрів розподілу випадкової величини.

Мета: здобути навики оцінювання статистичних параметрів на основі біологічних даних.

Хід виконання роботи:

1. Перевірка гіпотези про вид закону розподілу досліджуваної величини

ПРИКЛАД 1.1. За наданим інтервальний статистичним рядом знайти закон розподілу випадкової величини *X*.

$[a_i;a_{i+1})$	[–2;–1,2)	[–1,2;–0,4)	[–0,4;0,4)	[0,4;1,2)	[1,2;2)
n_i	6	11	21	7	5

Розв'язок. Для визначення виду закону розподілу побудуємо гістограму за даними таблиці. За видом гістограми висуваємо гіпотезу про нормальний закон розподілу даної випадкової величини:

H_0 – випадкова величина X розподілена за нормальним законом;

H_1 – випадкова величина X не розподілена за нормальним законом.

Щільність розподілу випадкової величини, розподіленої за нормальним законом має вид $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$, де a і σ – параметри розподілу.

Знайдемо означені параметри, враховуючи, що $\bar{x}=a;\quad S^2=\sigma^2$. Розрахунки оформимо у вигляді таблиці

Знайдемо вибіркове середнє, вибіркову дисперсію і вибіркове середнє квадратичне відхилення. Отже, параметрами теоретичного закону розподілу є: $\bar{x}=a=-0{,}096;\quad S=\sigma=0{,}886$.

ПРИКЛАД 1.2. Для варіаційних рядів значень досліджуваних біологічних ознак перевірити гіпотези про вид закону розподілу досліджуваної величини

x_i	0	1	2	3	4	5	6	7

n_i	6	27	26	20	10	5	5	1

Розв'язок. Сформулюємо гіпотези:

H_0 – випадкова величина *X* підкоряється закону розподілу Пуассона;

H_1 – випадкова величина *X* не підкоряється закону розподілу Пуассона.

Закон Пуассона має вигляд: $p_k = \frac{e^{-\lambda}\lambda^k}{k!}$, $k = 0,1,\ldots;\lambda > 0$ де λ - параметр розподілу. Крім того, відомо, що $\overline{x} = \lambda$; $S^2 = \lambda$. Отже, для встановлення параметра λ потрібно знайти $\overline{x}$ або S^2.

Випадкова величина *X* – кількість значень в годину; тоді кількість спостережень – це відповідні значенням *X* частоти n_i, а таблиця 2.5 є статистичним рядом і емпіричним законом розподілу величини *X*. Знайдемо $\overline{x}$ і S^2. Отже, $\overline{x} = \frac{1}{n}\sum_{i=1} x_i n_i = \frac{1}{100}\cdot 241 = 2,41$; $S^2 = \frac{1}{n}\sum_{i=1}(x_i - \overline{x})^2 n_i = \frac{1}{100}\cdot 244,19 \approx 2,44$.

Оскільки повинна виконуватися рівність $\overline{x} = \lambda; S^2 = \lambda$ то як параметр можна вибрати або $\overline{x}$, або S^2, або їх середнє арифметичне. Виберемо $\lambda = \frac{\overline{x} + S^2}{2} \approx 2,426$. Таким чином, гіпотеза H_0 – це припущення, що величина *X* розподілена згідно із законом Пуассона: $p_k = \frac{e^{-2,426}2,426^k}{k!}$, $k = 0,1,\ldots$

Перевіримо правильність гіпотези за допомогою критерію Пірсона. Знайдемо теоретичні частоти, використовуючи формулу: $p_k = \frac{e^{-2,426}2,426^k}{k!}$, $k = 0,1,\ldots$ Як k візьмемо значення *X*, тобто x_i.

Відмітимо, що p_i – ймовірність того, що *X* прийме значення x_i, тобто статистично вони є відносними частотами, теоретичні частоти знаходитимемо за формулою: $n_i' = np_i$.

Для зручності при обчисленні теоретичних частот продовжимо таблицю, складену на основі статистичного ряду. Отже, за результатами розрахунків $\chi^2 = \sum_{i=0}^{7} \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}} = 5,739$.

Для даного завдання $l = k - r - 1$=8-1-1=6. Виберемо рівень значущості α=0,01 і знайдемо за допомогою таблиць або функції ХИ2ОБР табличного процесора Excel значення $\chi^2{}_{\alpha,l}$: $\chi^2{}_{0,01;6}$=16,812. Оскільки для такого рівня довіри (0,99) $\chi^2 < \chi^2{}_{\alpha,l}$, то гіпотезу H_0 про розподіл Пуассона можна прийняти.

2. Перевірка гіпотези про рівність генеральних середніх. Критерій Стьюдента

ПРИКЛАД 1.4. Для насадження 10 дерев за першою технологією було витрачено, у середньому, 30хв. Дисперсія часу складала 1хв^2. Для насадження кожного з 16 дерев за другою технологією було витрачено, у середньому, 28хв із дисперсією часу 2хв^2. Чи можна вважати, що у середньому для насадження за першою технологією потрібно більше часу?

Розв'язок. За умов задачі було зроблено дві вибірки: перша – вибірка об'єму n_1=10 значень величини X_1 – часу, потрібному для насадження за першою технологією; друга – вибірка об'єму n_2=16 значень величини X_2 – часу, потрібному для насадження за другою технологією. Відомі вибіркові середні $\overline{x_1}$ =30хв та $\overline{x_2}$ =28хв – середній час, необхідний для насадження за першою і другою технологіями відповідно. Відомі дисперсії часу для вибірок: S_1^2 =1хв^2 та S_1^2 =2хв^2. За питанням задачі потрібно перевірити гіпотезу про рівність генеральних середніх.

Сформулюємо гіпотези:

H_0 – середні двох генеральних сукупностей рівні, тобто $\overline{x_1} = \overline{x_2}$;

H_1 - середні двох генеральних сукупностей не рівні, тобто $\overline{x_1} \neq \overline{x_2}$.

Перед вибором критерію для перевірки потрібно встановити, чи рівні генеральні дисперсії. Скористуємось критерієм Фішера. За формулою обчислимо значення F-критерію: оскільки $S_2^2 > S_1^2$, то $F = \dfrac{S_2^2}{S_1^2} = \dfrac{2}{1} = 2$.

Знайдемо критичне значення розподілу Фішера $F_{крит}$: оберемо рівень значущості α=0,05; врахуємо, що ступені волі l_1=n_1–1=9 та l_2=n_2–1=15. Тоді FРАСПОБР (α; l_2; l_1)=FРАСПОБР (0,05; 15; 9)=3,006. Отже, F< $F_{крит}$, тому генеральні дисперсії можна вважати рівними.

Оскільки генеральні дисперсії рівні (випадок 1), то t-критерій Стьюдента розраховуємо за формулою:

Знайдемо критичне значення розподілу Стьюдента $t_{\text{крит}}$, враховуючи, що $l = n_1 + n_2 - 2 = 10 + 16 - 2 = 24$. Оберемо значення α=0,05. Тоді $t_{\text{крит}}$=СТЬЮДРАСПОБР($\dfrac{\alpha}{2}$;l)=СТЬЮДРАСПОБР(0,025;24)=2,39.

Отже, $|t| > t_{\text{крит}}$, тому гіпотеза H_0 про рівність генеральних середніх відкидається на рівні значущості 0,05 і приймається гіпотеза H_1.

Висновок: для насадження дерев за першою технологією потрібно, у середньому, більше часу.

3. Перевірка статистичних гіпотез із використанням Microsoft Excel

Двохвибірковий F-тест для дисперсій. Перевірку гіпотези про рівність генеральних дисперсій за F-критерієм (Фішера) можна виконати за допомогою пакету аналізу Microsoft Excel. Для використання пакету необхідно:

1) Вибрати в меню послідовно пункти ***Сервис – Анализ данных***, після чого з'явиться вікно для вибору інструменту аналізу.

2) Вибрати у діалоговому вікні інструмент ***Двухвыборочный F-тест для дисперсии***, після чого з'явиться вікно для надання параметрів.

3) Задати всі необхідні параметри і натиснути ОК.

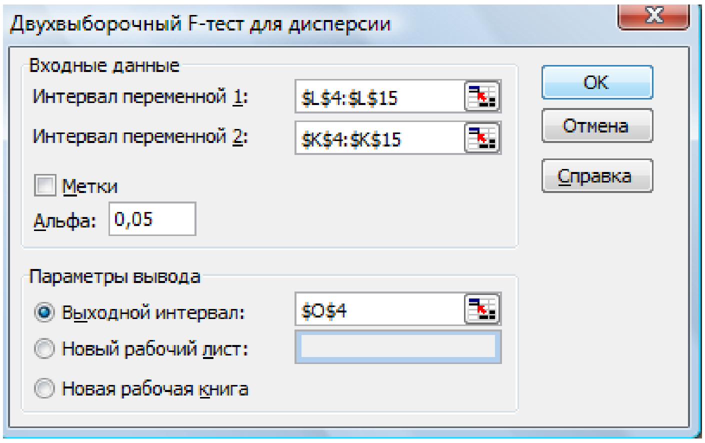


Рис. 1. Вікно надання параметрів

Приклад і результат роботи тесту наведено на рис. 2.

Двохвибірковий t-тест для середніх. Перевірку гіпотез про рівність генеральних середніх за критерієм Стьюдента можна виконати за допомогою пакету аналізу даних Microsoft Excel. Для здійснення перевірки необхідно у вікні для вибору інструменту аналізу вибрати:

1) У випадку рівних генеральних дисперсій – інструмент ***Двухвыборочный t-тест с одинаковыми дисперсиями***;

2) У випадку різних генеральних дисперсій – інструмент ***Двухвыборочный t-тест с разными дисперсиями***;

3) У випадку залежних вибірок – ***Парный двухвыборочный t-тест для средних***.

4) Після вибору інструменту аналізу з’явиться вікно для надання параметрів аналізу, аналогічне зображеному на рис. 1, в якому необхідно задати масиви чарунок із вхідними вибірковими даними і рівень значущості (стандартний рівень – 0,05). Приклад і результат роботи тесту наведено на рис. 3.

Двохвибірковий F-тест для дисперсії

Вхідні дані

Номер з/р	Вибіркові дані: I вибірка	Вибіркові дані: II вибірка
1	0,027	0,075
2	0,036	0,24
3	0,1	0,08
4	0,12	0,105
5	0,32	0,075
6	0,45	0,032
7	0,049	0,06
8	0,105	0,075

Двухвыборочный F-тест для дисперсии

	Переменная 1	*Переменная 2*
Среднее	0,150875	0,09275
Дисперсия	0,023234982	0,003957643
Наблюдения	8	8
df	7	7
F	5,870914325	
P(F<=f) одностороннее	0,016248714	
F критическое одностороннее	3,78704354	

Рис. 2. Перевірка рівності генеральних дисперсій

Двохвибірковий t-тест для середніх

Вхідні дані

Номер з/р	Вибіркові дані: I вибірка	Вибіркові дані: II вибірка
1	0,027	0,075
2	0,036	0,24
3	0,1	0,08
4	0,12	0,105
5	0,32	0,075
6	0,45	0,032
7	0,049	0,06
8	0,105	0,075

Двухвыборочный t-тест с различными дисперсиями

	Переменная 1	*Переменная 2*
Среднее	0,150875	0,09275
Дисперсия	0,023234982	0,003957643
Наблюдения	8	8
Гипотетическая разность средних	0	
df	9	
t-статистика	0,996970694	
P(T<=t) одностороннее	0,172413357	
t критическое одностороннее	1,833112923	
P(T<=t) двухстороннее	0,344826713	
t критическое двухстороннее	2,262157158	

Рис. 3. Перевірка рівності генеральних середніх

Індивідуальні завдання:

1. При дослідженні ефективності фунгіцидів двох типів було сформовано дві групи з 15 рослин. У випадку використання фунгіцидів першого типу строк одужання склав, у середньому, 11 днів з дисперсією 3 дні; другого – 8 днів з дисперсією 4 дні. Перевірити на рівні значущості 0,01 гіпотезу про перевагу другого типу фунгіцидів

2. При дослідженні густоти травостою було перевірено 50 газонів першої фірми і 63 одиниці другої. Середня оцінка першої склала 36 балів, другої – 32 бали. Дисперсія оцінок виявилася рівною 10. Визначити фірму, що виробляє більш якісну продукцію на рівні значущості 0,01.

3. Середній річний об’єм виробництва 10 однотипних компаній в регіоні А склав 5900 одиниць продукції; 8 компаній того ж типу в регіоні В – 5690 одиниць. Вибіркова дисперсія об’єму виробництва в регіоні А склала 1000, в регіоні В – 1500. Перевірити гіпотезу про рівність середніх значень на рівні значущості 0,05.

4. Середня урожайність пшениці у фермерському хазяйстві складала 75 ц/га. Після внесення нового добриву середня урожайність підвищилася на 5 ц/га. Вибіркова дисперсія склала 2,5 ц/га. Перевірити на рівні значущості 0,01 гіпотезу про доцільність нового добриву.

5. Середня продуктивність праці на підприємстві складала 30 одиниць продукції за зміну з дисперсією 4. Після внесення змін в організацію праці середня продуктивність склала 34 з дисперсією 7. Перевірити на рівні значущості 0,01 гіпотезу про доцільність внесення змін в організацію праці.

6. Для обробки кожної з n_1=53 рослин за першою технологією було витрачено, у середньому $\overline{x_1}$ секунд часу з дисперсією S_1^2. Для обробки кожної з n_2=43 рослин за другою технологією було витрачено, у середньому $\overline{x_2}$ секунд часу з дисперсією S_2^2. Чи можна зробити висновок, що для обробки рослин за першою технологією потрібно, у середньому, більше часу, ніж за другою. Гіпотезу перевірити на рівні значущості α.

№	$\overline{x_1}$	S_1^2	$\overline{x_2}$	S_2^2	α
1	38	4	31	2	0,05
2	39	5	32	3	0,01
3	33	7	31	8	0,05
4	37	8	34	7	0,01
5	35	4	32	5	0,05
6	37	5	36	4	0,01
7	37	7	35	7	0,05
8	38	8	33	8	0,01

9	42	3	40	5	0,05
10	40	2	34	4	0,01

Контрольні запитання:

1. Довірчі інтервали й довірчі ймовірності.
2. Статистичні гіпотези.
3. Поняття про F-розподіл.
4. Планування обсягу вибірки.

Лабораторно-практичне заняття №2

Тема: Кореляційний аналіз.

Мета: отримати навички розрахунку кореляційної матриці.

Хід виконання роботи:

1. Коефіцієнт кореляції Пірсона.

ПРИКЛАД 2.1. За наявними даними X (%) і продуктивності Y (од. продукції/год.) для 14 однотипних господарств оцінити тісноту зв'язку між X і Y. Визначити можливість розповсюдження результатів розрахунків на всі господарство такого типу.

X	32	30	36	40	41	47	56	54	60	55	61	67	69	76
Y	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Розв'язок. Дані таблиці є вибіркою значень X і відповідних значень Y. Оскільки кількість даних невелика (n=14), то їх можна не групувати. Для оцінки тісноти зв'язку між X і Y розрахуємо коефіцієнт кореляції Пірсона за формулою для незгрупованих даних.

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2}} = \frac{14\cdot 26907 - 724\cdot 492}{\sqrt{14\cdot 40134 - 724^2}\sqrt{14\cdot 18138 - 492^2}}$$

$$= \frac{20490}{\sqrt{37700}\sqrt{11868}} \approx 0{,}969.$$

За значенням коефіцієнта кореляції можна зробити висновок, що між X і Y існує сильний додатній зв'язок.

Перевіримо статистичну значущість знайденого коефіцієнта кореляції Пірсона. Розрахуємо t-статистику за формулою (3.4): $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}969\sqrt{14-2}}{\sqrt{1-0{,}969^2}} \approx 13{,}59$. Знайдемо $t_{крит}$, враховуючи, що $l=n-2=14-2=12$. Оберемо рівень значущості α=0,01. Тоді $t_{крит}$=СТЬЮДРАСПОБР (0,01; 12)=3,055.

Оскільки розраховане значення t-статистики більше критичного $13,59 > 3,055$, то коефіцієнт кореляції можна вважати значимим на обраному рівні α =0,01. Та існує сильний додатній зв'язокю.

ПРИКЛАД 2.2. Визначити кореляційне відношення між довжиною надземних пагонів (x) і довжиною їх коренів (y).

Y \ X	12,5	17,5	22,5	27,5
20,5	1	-	-	-
21,5	-	2	-	-
22,5	-	1	2	-
23,5	-	-	3	3
24,5	-	-	-	8

Розв'язок. Згруповані вибіркові дані запишемо у вигляді кореляційної таблиці

Для розрахунку коефіцієнта кореляції Пірсона скористуємось формулою.

Розрахунки для зручності оформимо у вигляді таблиці

Окремо розрахуємо $\sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j n_{ij}$:

$$\sum_{i=1}^{4}\sum_{j=1}^{5} x_i y_j n_{ij} = 20,5\cdot 12,5 + 21,5\cdot 17,5\cdot 2 + 22,5\cdot 17,5 + 22,5\cdot 22,5 + 23,5\cdot 22,5\cdot 3 +$$

$$+23,5\cdot 27,5\cdot 3 + 24,5\cdot 27,5\cdot 8 = 11330$$

Підставимо знайдені суми в формулу:

$$r = \frac{n\sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j n_{ij} - \left(\sum_{i=1}^{k} x_i n_i\right)\left(\sum_{j=1}^{m} y_j n_j\right)}{\sqrt{n\sum_{i=1}^{k} x_i^2 n_i - \left(\sum_{i=1}^{k} x_i n_i\right)^2}\sqrt{n\sum_{j=1}^{m} y_j^2 n_j - \left(\sum_{j=1}^{m} y_j n_j\right)^2}} = \frac{20\cdot 11330 - 480\cdot 468}{\sqrt{20\cdot 11925 - 480^2}\sqrt{20\cdot 10979 - 468^2}} =$$

$$= \frac{1960}{90\cdot 23,58} \approx 0,924.$$

За значенням коефіцієнта кореляції можна зробити висновок, що між X і

Y існує сильний додатній зв'язок.

Перевіримо статистичну значущість знайденого коефіцієнта кореляції Пірсона. Розрахуємо t-статистику за формулою (2.4): $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}924\sqrt{20-2}}{\sqrt{1-0{,}924^2}} = \frac{3{,}92}{\sqrt{0{,}146}} \approx 10{,}26$. Знайдемо $t_{крит}$, враховуючи, що l=n-2=20-2=18. Оберемо рівень значущості α=0,01. Тоді $t_{крит}$=СТЬЮДРАСПОБР (0,01; 18)=2,88.

Оскільки розраховане значення t-статистики більше критичного $10{,}26 > 2{,}88$, то коефіцієнт кореляції можна вважати значимим на обраному рівні α=0,01.

Висновок. Між X та У сильний додатній зв'язок.

2. Множинний та частинний коефіцієнти кореляції.

ПРИКЛАД 2.2. Для вивчення залежності урожайності зернових культур Z (ц/га) від якості пашні X (бали) і кількості внесеного добриву Y (кг/га) було проведено дослідження 6 фермерських хазяйств, результати якого надано у таблиці. Визначити силу зв'язку між Z та X та Y, використовуючи множинний коефіцієнт кореляції. Порівняти силу зв'язку між Z та X і між Z та Y за частинними коефіцієнтами кореляції.

X	26	35	36	40	41	45
Y	2,1	2,3	2,4	2,6	2,9	3
Z	18	21	22,1	25,3	28	28,5

Розв'язок. За умов задачі необхідно для об'єкту, що характеризується трьома ознаками X, Y та Z (k=3), розрахувати множинний коефіцієнт кореляції R_Z і частинні коефіцієнти кореляції R_{XZ} та R_{YZ} на основі 6 взаємопов'язаних тройок вибіркових даних $(x_i, y_i, z_i), i = \overline{1,n}, n = 6$.

Побудуємо матрицю парних коефіцієнтів кореляції, які обчислимо за формулою. Розрахунки для зручності оформимо у вигляді таблиці

Отже, за формулою маємо:

$$r_{XY} = r_{YX} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} = \frac{6\cdot 579,4 - 223\cdot 15,3}{\sqrt{6\cdot 8503 - 223^2}\sqrt{6\cdot 39,63 - 15,3^2}} \approx$$

$$\approx 0,935;$$

$$r_{XZ} = r_{ZX} = \frac{n\sum_{i=1}^{n} x_i z_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} z_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} z_i^2 - \left(\sum_{i=1}^{n} z_i\right)^2}} = \frac{6\cdot 5441,1 - 223\cdot 142,9}{\sqrt{6\cdot 8503 - 223^2}\sqrt{6\cdot 3489,75 - 142,9^2}} \approx$$

$$\approx 0,954;$$

$$r_{YZ} = r_{ZY} = \frac{n\sum_{i=1}^{n} y_i z_i - \left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} z_i\right)}{\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}\sqrt{n\sum_{i=1}^{n} z_i^2 - \left(\sum_{i=1}^{n} z_i\right)^2}} = \frac{6\cdot 371,62 - 15,3\cdot 142,9}{\sqrt{6\cdot 39,63 - 15,3^2}\sqrt{6\cdot 3489,75 - 142,9^2}} \approx$$

$$\approx 0,991;$$

Таким чином, кореляційна матриця має вигляд:

$$A = \begin{pmatrix} 1 & 0,935 & 0,954 \\ 0,935 & 1 & 0,991 \\ 0,954 & 0,991 & 1 \end{pmatrix}.$$

Знайдемо визначник $|A|$ матриці A та алгебраїчне доповнення $A_{ZZ} = A_{33}$:

$$|A| = \begin{vmatrix} 1 & 0,935 & 0,954 \\ 0,935 & 1 & 0,991 \\ 0,954 & 0,991 & 1 \end{vmatrix} = 1 + 2\cdot 0,935\cdot 0,991\cdot 0,954 - 0,954^2 - 0,991^2 -$$

$$- 0,935^2 \approx 0,0015;$$

$$A_{ZZ} = A_{33} = (-1)^{3+3}\begin{vmatrix} 1 & 0,935 \\ 0,935 & 1 \end{vmatrix} = 1 - 0,935^2 \approx 0,1258;$$

тоді $R_Z = R_3 = \sqrt{1 - \frac{|A|}{A_{33}}} = \sqrt{1 - \frac{0,0015}{0,1258}} \approx 0,994$. Значення множинного коефіцієнта кореляції R_z показує, що величина *Z* сильно пов’язана з *X* та *Y*.

Перевіримо статистичну значущість множинного коефіцієнта кореляції R_Z. Знайдемо t-статистику за формулою:

$$t=\frac{R^2(n-k)}{(1-R^2)(k-1)}=\frac{0{,}994^2(6-3)}{(1-0{,}994^2)(3-1)}\approx 124{,}09\cdot$$

Знайдемо $F_{крит}$, враховуючи, що $l_1=k-1=3-1=2$; $l_2=n-k=6-3=3$. Оберемо рівень значущості α=0,01. Тоді $F_{крит}$=FРАСПОБР (0,01; 2; 3)=30,82. Оскільки t> $F_{крит}$, то множинний коефіцієнт кореляції R_Z є статистично значимим на рівні значущості α=0,01.

Для обчислення частинних коефіцієнтів кореляції $R_{XZ}=R_{13}$ та $R_{YZ}=R_{23}$ знайдемо алгебраїчні доповнення:

$$A_{13}=(-1)^{1+3}\begin{vmatrix}0{,}935 & 1\\ 0{,}954 & 0{,}991\end{vmatrix}=0{,}935\cdot 0{,}991-0{,}954\approx -0{,}027;$$

$$A_{23}=(-1)^{2+3}\begin{vmatrix}1 & 0{,}935\\ 0{,}954 & 0{,}991\end{vmatrix}=(-1)(0{,}991-0{,}935\cdot 0{,}954)\approx -0{,}099;$$

$$A_{11}=(-1)^{1+1}\begin{vmatrix}1 & 0{,}991\\ 0{,}991 & 1\end{vmatrix}=(1-0{,}991^2)\approx 0{,}018;$$

$$A_{22}=(-1)^{2+2}\begin{vmatrix}1 & 0{,}954\\ 0{,}954 & 1\end{vmatrix}=(1-0{,}954^2)\approx 0{,}09.$$

Тоді за формулою маємо:

$$R_{13}=\frac{-A_{13}}{\sqrt{A_{11}A_{33}}}=\frac{-(-0{,}027)}{\sqrt{0{,}018\cdot 0{,}126}}\approx 0{,}577; \qquad R_{23}=\frac{-A_{23}}{\sqrt{A_{22}A_{33}}}=\frac{-(-0{,}099)}{\sqrt{0{,}09\cdot 0{,}126}}\approx 0{,}929\cdot$$

Значення частинних коефіцієнтів кореляції показують, що величина *Z* пов’язана з величиною *Y* сильніше, ніж з величиною *X*.

Перевіримо статистичну значущість частинного коефіцієнта кореляції R_{13}. Знайдемо t-статистику за формулою:

$$t=\frac{R_{ij}\sqrt{n-k+2}}{\sqrt{1-R_{ij}^{\ 2}}}=\frac{0{,}577\sqrt{6-3+2}}{\sqrt{1-0{,}577^2}}\approx 1{,}581\cdot$$

Знайдемо критичне значення $t_{крит}$, враховуючи, що $l=n-k+2=6-3+2=5$. Оберемо рівень значущості α=0,01. Тоді $t_{крит}$=СТЬЮДРАСПОБР(0,01;5)=4,032. Оскільки розраховане значення t-статистики менше критичного $|t|<t_{крит}$, то частинний коефіцієнт кореляції R_{13} не є значимим на рівні значущості α=0,01.

Перевіримо статистичну значущість частинного коефіцієнта кореляції R_{23}. Знайдемо t-статистику:

$$t = \frac{R_{ij}\sqrt{n-k+2}}{\sqrt{1-R_{ij}^{2}}} = \frac{0,929\sqrt{6-3+2}}{\sqrt{1-0,929^2}} \approx 5,614.$$

Оскільки розраховане значення t-статистики більше критичного $|t| > t_{крит}$, то частинний коефіцієнт кореляції R_{23} є значимим на рівні значущості α =0,01.

Висновок: Урожайність зернових культур сильно пов’язана з якістю пашні і кількістю внесеного добриву. При цьому урожайність значно сильніше залежіть від кількості добриву, чим від якості пашні. Сила зв’язку між урожайністю та якістю пашні середня та не є статистично значимою.

3. **Кореляційний аналіз із використанням Microsoft Excel.** Вбудовані сервісні функції Microsoft Excel дозволяють розраховувати парні коефіцієнти кореляції Пірсона. Для отримання матриці парних коефіцієнтів кореляції необхідно:

1) Обрати ***Сервис – Анализ данных.***

2) У діалоговому вікні для вибору інструменту аналізу обрати інструмент ***Корреляция***. З’явиться вікно для надання параметрів.

3) Задати параметри для розрахунку коефіцієнтів кореляції. У графі ***Входной интервал*** вказати масив даних; у графі ***Группирование*** вказати тип групування, наприклад, ***По столбцам***, у графі ***Выходной интервал*** вказати ту клітину, починаючи з якої будуть надаватися вихідні дані – парні коефіцієнти кореляції. Натиснути ***ОК***.

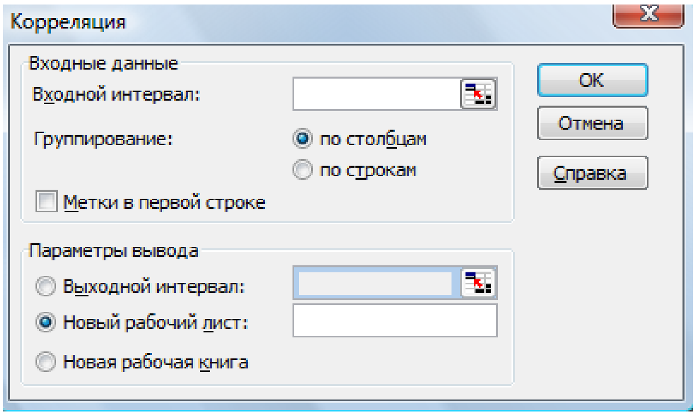


Рис. 1 Вікно надання параметрів кореляційного аналізу

Приклад і результати розрахунків парних коефіцієнтів кореляції надано на рис. 2.

Кореляційний аналіз

Вхідні дані

№	Значення		
i	x1	x2	x3
1	1	328	0,054
2	2	329	0,101
3	3	329	0,099
4	4	345	0,019
5	5	352	0,065
6	6	370	0,053
7	7	377	0,178
8	8	385	0,174
9	9	396	0,289
10	10	399	0,195
11	11	390	0,102
12	12	373	0,138

	Столбец 1	*Столбец 2*	*Столбец 3*
Столбец 1	1		
Столбец 2	0,8913997	1	
Столбец 3	0,5634229	0,692214	1

Рис. 2. Результати розрахунку коефіцієнтів кореляції

Зауваження. 1) В результаті роботи інструменту аналізу даних ***Корреляция*** розраховується матриця парних коефіцієнтів кореляції Пірсона навіть у випадку встановлення зв’язку між двома величинами.

2) Клітини матриці, що розташовані вище головної діагоналі звичайно надаються незаповненими, оскільки матриця симетрична відносно головної діагоналі.

3) Засобами Microsoft Excel неможливо розрахувати парні або множинні коефіцієнти кореляції, однак можна значно спростити розрахунки,

використовуючи вбудовану математичну функцію МОПРЕД, яка дозволяє знайти визначник заданої матриці.

Приклад і результати обчислення визначника матриці надано на рис. 3..

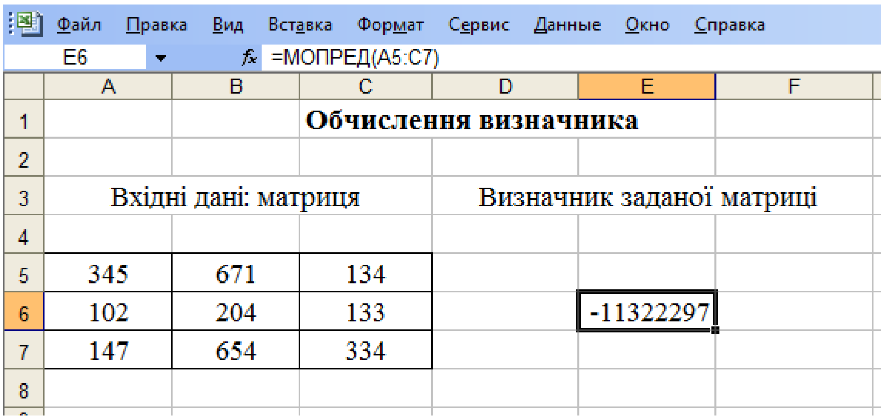


Рис. 3. Обчислення визначника заданої матриці

Індивидуальні завдання:

1. Визначити силу зв'язку між вагою рослини *X* (г) і вагою його насіння *Y* (г) за даними таблиці

X	40	50	60	70	80	90	100
Y	20	25	28	30	35	40	45

2. Обчислити коефіцієнт кореляції. Проаналізувати зв'язок між *Z* та *X* і *Y* за частинними і множинним коефіцієнтами кореляції.

Z	1,2	1,3	2,5	1,4	1,2	0,2	2,4	4,1	1,1
X	1,4	1,4	2,5	1,5	1,3	0,3	2,6	4,2	1,1
Y	1,3	1,3	1,4	1,8	1,5	1,6	1,8	1,9	1,6

3. Обчислити кореляційне відношення між діаметром (хі) і висотою (уі) дерев:

X *Y*	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
1,5 – 2,5	1	-	-	-	-
2,5 – 3,5	2	5	2	-	-

3,5 – 4,5	-	3	3	2	-
4,5 – 5,5	-	-	-	2	-

4. Проаналізувати зв'язок між *Z* та *X* і *Y* за частинними і множинним коефіцієнтами кореляції.

Z	10	12	12	14	16	17	18
X	0,2	0,5	0,3	0,5	0,5	0,6	0,8
Y	0,8	0,2	1	1,2	0,9	1	1,1

5. В таблиці 3.16 наведено дані про рівень витрат *X* (%) та річний доход *Y* (млн. грн.), які було зібрано за 50 крупними магазинами. Визначити силу зв'язку між означеними факторами.

X / *Y*	4 – 6	6 – 8	8 – 10	10 – 12	12 – 14
0,5 – 2,0	-	-	2	3	1
2,0 – 3,5	-	4	5	1	-
3,5 – 5,0	-	8	5	5	-
5,0 – 6,5	3	8	2	-	-
6,5 – 8,0	2	1	-	-	-

6. За даними таблиці проаналізувати тісноту зв'язку між діаметрами і висотами у

10 дерев сосни звичайної та перевірити гіпотезу про наявність лінійного зв'язку.

№	*X*					*Y*					α
1	1	5	3	4	7	1	5	5	2	8	0,05
2	3	6	7	8	7	1	3	5	5	4	0,01
3	4	7	5	4	5	3	1	2	2	1	0,05
4	9	8	3	4	1	0	1	4	3	5	0,01
5	1	0	3	3	0	2	3	5	6	4	0,05
6	0	4	7	8	5	2	6	8	7	5	0,01

7	4	2	3	4	3	8	6	8	7	6	0,05
8	7	5	1	0	3	8	6	4	2	4	0,01
9	3	5	7	2	5	1	3	5	0	1	0,05
10	4	4	8	9	5	6	2	9	9	4	0,01

Контрольні запитання:

1. Сумарний показник зв'язку.
2. Функціональна залежність і кореляція.
3. Коефіцієнт кореляції.
4. Кореляційне відношення та визначення його достовірності.
5. Міра криволінійності зв’язку.

Лабораторно-практичне заняття №3

Тема: Моделі зв’язку. Лінійна регресія

Мета: навчитись визначати коефіцієнти рівняння прямої $y = a + b\,x$ та проводити їх оцінку.

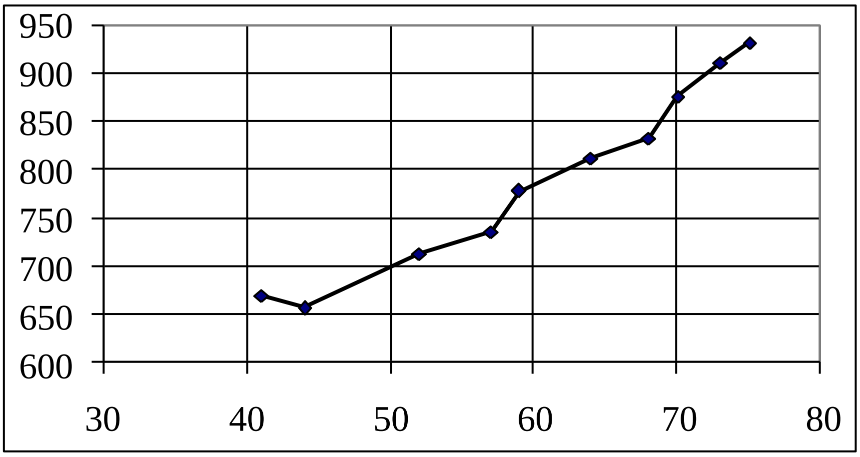
Хід виконання роботи:

ПРИКЛАД 4.1. Побудувати регресійну модель, що описує залежність *Y* від *X*. Відповідні статистичні дані надано у таблиці.

X	41	44	52	57	59	64	68	70	73	75
Y	670	657	713	736	778	812	833	876	911	932

Розв’язок. В таблиці надано вибіркові дані: значення $x_i,\ i=\overline{1,n}$ величини *X* та відповідні значення $y_i,\ i=\overline{1,n}$; кількість пар – $n=10$ невелика, тому для проведення регресійного аналізу їх можна не групувати.

Перший етап аналізу: визначимо вид залежності *Y* від *X*. Побудуємо емпіричну лінію регресії).



Оскільки емпірична лінія регресії наближається до прямої лінії, то висуваємо гіпотезу про лінійну залежність *Y* від *X*, тобто рівняння регресії будемо шукати у вигляді $y = ax + b$.

Другий етап: знайдемо параметри a,b рівняння регресії, для чого складемо систему для даних, що не згруповані. Необхідні розрахунки для зручності оформимо у вигляді таблиці.

Отже, складемо систему для знаходження параметрів рівняння регресії та розв’яжемо її за правилом Крамера:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i \end{cases} \Rightarrow \begin{cases} 37625a + 603b = 487643 \\ 603a + 10b = 7918 \end{cases}.$$

Знайдемо головний визначник системи, що складений із коефіцієнтів перед невідомими: $\Delta = \begin{vmatrix} 37625 & 603 \\ 603 & 10 \end{vmatrix} = 37625 \cdot 10 - 603^2 = 12641$.

Знайдемо допоміжні визначники, що отримуються із головного заміною відповідного стовпця коефіцієнтів на стовпець вільних членів:

$\Delta a = \begin{vmatrix} 487643 & 603 \\ 7918 & 10 \end{vmatrix} = 487643 \cdot 10 - 603 \cdot 7918 = 101876$;

$\Delta b = \begin{vmatrix} 37626 & 487643 \\ 603 & 7918 \end{vmatrix} = 37626 \cdot 7918 - 48743 \cdot 603 = 3866021$.

Знайдемо невідомі за формулами Крамера:

$$a = \frac{\Delta a}{\Delta} = \frac{101876}{12641} \approx 8,06; \quad b = \frac{\Delta b}{\Delta} = \frac{3866021}{12641} \approx 305,83.$$

Отже, шукане рівняння регресії має вигляд $y = 8,06x - 305,83$.

Третій етап: перевіримо правильність побудови моделі та її статистичну значущість за F-статистикою і адекватність вибірковим даним за коефіцієнтом детермінації. Для чого знайдемо загальну варіацію, варіації регресії та залишків; необхідні розрахунки оформимо у вигляді таблиці .

Передусім знайдемо $\bar{y}$: $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \approx 791,8$.

Отже, $Q = 85579,6$; $Q_p = 82103,63$; $Q_o = 3475,974$; тоді основне варіаційне рівняння $Q = Q_p + Q_o$ для побудованої моделі має вигляд: $85579,6 = 82103,63 + 3475,974$ і є тотожністю, тому рівняння регресії побудовано правильно.

Для перевірки статистичної значущості рівняння регресії знайдемо F-статистику, враховуючи, що n=10, l=2 – оскільки шукали рівняння з двома параметрами:

$$F=\frac{Q_p\left(n-l\right)}{Q_o\left(l-1\right)}=\frac{82103(10-2)}{3475,974(2-1)}\approx 188,96\cdot$$

Знайдемо $F_{кр}$: $F_{кр}$= $FРАСПОБР\left(0,001,2-1,10-2\right)\approx 25,41$. Розраховане значення F-статистики більше критичного, тому регресійна модель є статистично значущою на рівні 0,001.

Знайдемо коефіцієнт детермінації R^2: $R^2=\frac{Q_p}{Q}=\frac{82103,63}{85579,6}\approx 0,96$. Значення коефіцієнта детермінації свідчить, що 96% варіації результативної ознаки *Y* пояснюються рівнянням регресії.

Висновок: Сумарні значення *Y* лінійно залежать від *X*. Залежність описується рівнянням $y=8,06x-305,83$, яке є статистично значимим на рівні значущості 0,001 та описує 96% вибіркових даних.

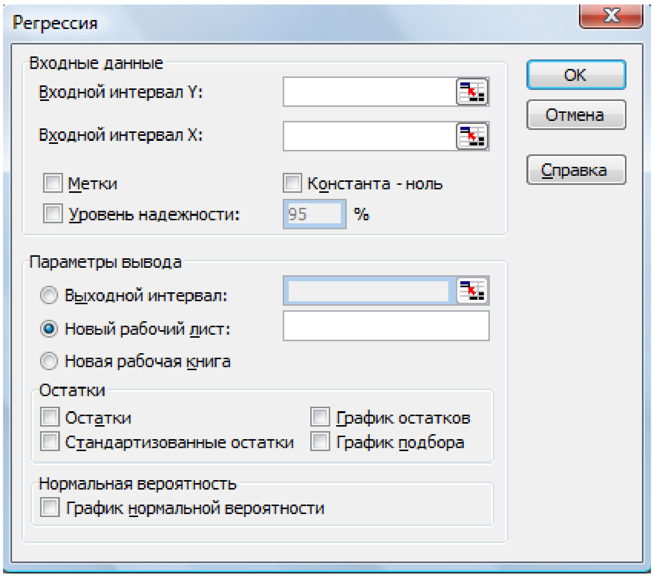
Регресія у Microsoft Excel

Пакет аналізу даних Microsoft Excel надає можливість будувати регресійні моделі, але тільки у випадку лінійної залежності результативної ознаки *Y* від факторної ознаки *X* і тільки для незгрупованих вибіркових даних.

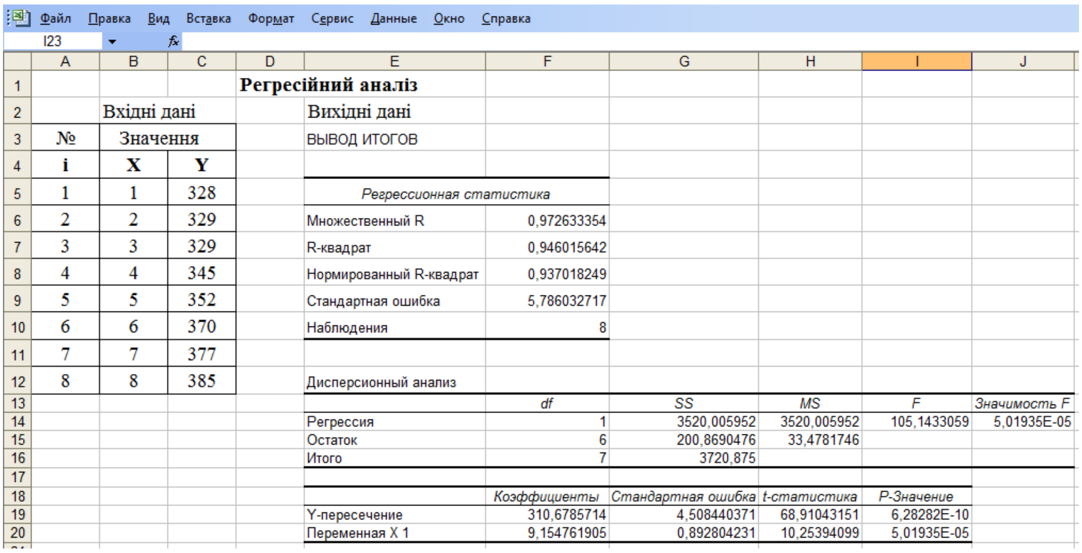
Для побудови лінійної регресійної моделі необхідно:

1) Викликати ***Сервис – Анализ данных – Регрессия – ОК***. З’явиться вікно для надання вхідних даних.

2) У графі ***Входной интервал Y*** та ***Входной интервал X*** вказати відповідні стовпці даних; у графі ***Выходной интервал*** вказати ту клітину, починаючи з якої будуть надаватися вихідні дані – параметри рівняння регресії та результати її статистичного аналізу.



Приклад і результати роботи функції ***Регрессия***



Регресійний аналіз

Вхідні дані		
№	Значення	
i	X	Y
1	1	328
2	2	329
3	3	329
4	4	345
5	5	352
6	6	370
7	7	377
8	8	385

Вихідні дані

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,972633354
R-квадрат	0,946015642
Нормированный R-квадрат	0,937018249
Стандартная ошибка	5,786032717
Наблюдения	8

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	3520,005952	3520,005952	105,1433059	5,01935E-05
Остаток	6	200,8690476	33,4781746		
Итого	7	3720,875			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	310,6785714	4,508440371	68,91043151	6,28282E-10
Переменная X 1	9,154761905	0,892804231	10,25394099	5,01935E-05

В таблиці у графі ***Коэффициенты*** вказані значення параметрів моделі a та b: b - в графі **Y-пересечение**, a - в графі **Переменная X1**. Отже, побудована лінійна регресійна модель має вигляд:

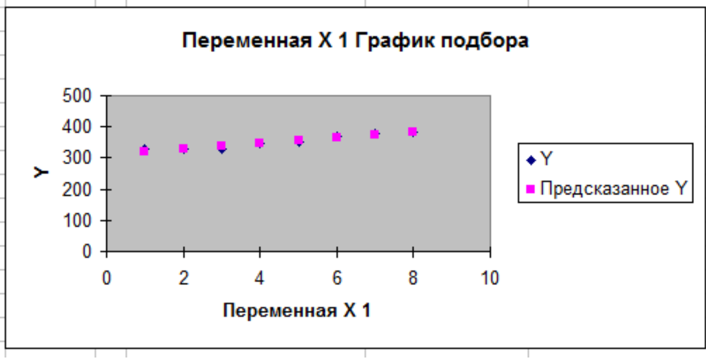
$$y = 69{,}15x + 310{,}68\,.$$

Для перевірки статистичної значущості моделі надається значення F-статистики у графі F : F = 105,14.

Коефіцієнт детермінації моделі R^2 надається у графі **R-квадрат,** R^2=0,97.

Крім того, може бути надано: графік підбору – порівняльна діаграма, що містить емпіричну і теоретичну лінії регресії; таблиця залишків – різниць емпіричних і теоретичних значень *Y*.

	A	B	C	D	E	F	G
1					**Регресійний аналіз**		
2		Вхідні дані			Вихідні дані		
3	№	Значення			ВЫВОД ОСТАТКА		
4	**i**	**X**	**Y**				
5	1	1	328		*Наблюдение*	*Предсказанное Y*	*Остатки*
6	2	2	329		1	319,8333333	8,166666667
7	3	3	329		2	328,9880952	0,011904762
8	4	4	345		3	338,1428571	-9,142857143
9	5	5	352		4	347,297619	-2,297619048
10	6	6	370		5	356,452381	-4,452380952
11	7	7	377		6	365,6071429	4,392857143
12	8	8	385		7	374,7619048	2,238095238
13					8	383,9166667	1,083333333



Індивідуальні завдання:

1. Для парної лінійної регресії на основі статистичних даних показника *y* і фактора *x* знайти оцінки:
   - параметрів лінії регресії b_0, b_1;
   - коефіцієнта кореляції r_{yx};
2. Використовуючи критерій Фішера, з надійністю α=0,05 оцінити адекватність прийнятої моделі статистичним даним.
3. Якщо модель адекватна статистичним даним, то знайти:
   - з надійністю α=0,05 надійні зони базисних даних;
   - прогноз показника та його надійні інтервали;
   - коефіцієнт еластичності для базисних даних і прогнозу.
4. Побудувати графіки:

- статистичних даних;
- лінії регресії і її довірчої зони;
- коефіцієнта еластичності.

Значення показника

N	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
1	7,24	10,89	16,21	12,11	15,21	16,62	10,22	12,50	19,66	14,87	22,68
2	8,02	11,92	17,75	12,3	15,42	17,63	10,58	13,88	20,53	15,78	23,89
3	9,28	12,45	18,39	13,82	16,44	19,22	12,01	15,16	21,31	16,79	24,32
4	10,12	13,27	18,87	14,84	17,93	19,36	12,84	16,06	22,59	18,03	25,97
5	11,12	14,12	19,6	15,86	18,52	20,52	13,28	16,66	23,27	18,29	26,23
6	12,19	15,23	21,21	16,41	19,8	21,95	15,13	17,65	24,44	19,93	27,60
7	13,01	16,07	21,84	17,8	20,76	22,45	15,84	18,46	25,85	20,32	28,13
8	14,12	17,4	23,00	18,61	21,3	23,56	17,08	19,54	26,74	21,18	29,84
9	15,21	18,68	24,44	19,57	22,25	24,9	17,99	20,58	27,36	22,47	30,31
10	16,29	19,46	25,36	21,26	24,14	25,53	18,32	21,77	28,37	23,47	31,52
11	17,01	20,52	25,54	21,08	24,17	26,11	19,49	22,15	29,22	24,07	32,27
12	18,03	21,32	27,14	22,99	25,66	28,02	20,59	23,80	30,50	25,57	33,77
13	19,19	22,58	27,95	23,43	26,5	28,37	21,35	24,79	31,21	27,07	34,66
14	20,21	23,73	28,99	24,63	27,46	29,46	23,20	25,57	32,56	27,62	35,93
15	21,22	25,02	30,8	25,41	29,02	30,42	24,21	27,18	33,66	28,42	36,97

Порядок виконання роботи:

1.1. Запустіть програму EXCEL. Згідно з номером варіанту, виданого викладачем, занесемо початкові значення фактора *x* в блок **A3:A17**, а значення показника *y* в блок **B3:B17**. Надалі, при формуванні розрахункової таблиці рекомендується дотримуватися форми, наведеної в табл.1.3. В першому рядку необхідно написати назву лабораторної роботи, а другий рядок буде використаний для занесення даних, які будуть отримуватися під час розрахунків. В табл.1.3 затемнення комірки означає, що в ній міститься формула, або конкретне числове значення якогось параметра.

1.2. Для обчислення параметрів b_0 та b_1 потрібно підрахувати добутки $x_i y_i$ та x_i^2 і відповідні суми:

- для знаходження добутку $x_i y_i$ використаємо блок **C3:C17**. Для клітинки **C3** формула запишеться: =A3*B3, дану формулу копіюємо в блок **C4:C17**;
- значення x_i^2 занесемо в блок **D3:D17**. Для клітинки **D3** формула запишеться: =A3^2, дану формулу копіюємо в блок **D4:D17**.

1.3. Для підрахунку сум будемо використовувати математичну функцію **СУММ(блок)**:

- в 20 рядку будемо знаходити суми відповідних значень. Значення $\sum x_i$ підрахуємо в **A20** за формулою =СУММ(A3:A17);
- цю формулу копіюємо в комірки **B20** - $\sum y_i$, **C20** - $\sum x_i y_i$, **D20** - $\sum x_i^2$.

1.4. Кількість значень факторів $n=15$ занесемо в комірку **B21**.

1.5. Підставимо адреси знайдених сум в формули (1.2) для розрахунку b_0 та b_1*:*

- формулу для визначення параметра b_1 занесемо в комірку **B22** =(B21*C20-A20*B20)/(B21*D20-A20^2);
- значення параметра b_0 занесемо в **B23** =(B20-B22*A20)/B21.

1.6. Для знаходження розрахункових значень показника $\hat{y}_i$ використовується рівняння регресії (1.1). Занесемо їх в блок **E3:E17**:

- для **E3** маємо формулу =B$22*A3+B$23, копіюємо її в блок **E4:E17;**
- в **E20** підрахуємо суму $\sum \hat{y}_i$ за формулою =СУММ(E3:E17) або просто копіюємо формулу з **D20**;
- при правильному виконанні розрахунків $\sum y_i = \sum \hat{y}_i$, тобто значення комірок **B20** та **E20** повинні співпадати.

1.7. Для оцінки коефіцієнта кореляції будемо використовувати статистичну функцію **КОРРЕЛ(блок знач.X; блок знач.Y)**. Для цього в комірку **F23** заносимо =КОРРЕЛ(A3:A17;B3:B17).

Для оцінки адекватності прийнятої моделі використовують формулу (1.2). Необхідно визначити середнє значення для x та y за допомогою

2.1. статистичної функції **СРЗНАЧ(блок):**

- середнє $\bar{x} = \frac{1}{n}\sum x_i$ визначимо в **D21** =СРЗНАЧ(A3:A17);
- середнє $\bar{y} = \frac{1}{n}\sum y_i$ - в **D22** =СРЗНАЧ(B3:B17).

2.2. Значення $(y_i - \hat{y}_i)^2$ обчислюємо в **F3:F17**. Для комірки **F3** формула =(B3-E3)^2, яку копіюємо в блок **F4:F17**.

2.3. Аналогічно значення $(\hat{y}_i - \bar{y})^2$ обчислюємо в **G3:G17**. Для комірки **G3** формула =(E3-D$22)^2, яку копіюємо в блок **G4:G17**.

2.4. Суми цих величин знайдемо в 20 рядку, $\sum(y_i - \hat{y}_i)^2$ в **F20** і $\sum(\hat{y}_i - \bar{y})^2$ в **G20**.

2.5. Знайдемо розрахункове значення Фішера F в **I22** за формулою =(G20*(B21-2))/(F20).

2.6. Табличне значення $F_{кр}$ знаходимо за допомогою статистичної функції **FРАСПОБР(α;1;n-2)** в **K22** за формулою =FРАСПОБР(0,05;1;13).

3.1 Для визначення надійної зони базисних даних необхідно обчислити:

- значення функції Ст'юдента $t_{(\alpha/2, n-1)}$ за допомогою статистичної функції **СТЬЮДРАСПОБР(α;k)**. Підрахуємо його в **F22** =СТЬЮДРАСПОБР(0,05;14);
- оцінку $\hat{\sigma}_\varepsilon$ в **D23** =КОРЕНЬ(F20/(B21-2));
- значення $(x_i - \bar{x})^2$ в **H3:H17**, занесемо в **H3** формулу =(A3-D$21)^2 і скопіюємо її в блок **H4:H17**;
- підрахуємо суму $\sum(x_i - \bar{x})^2$ в **H20**;
- підставляємо отримані адреси в формулу і знаходимо $\Delta\hat{y}_i$ в **I3:I17**. В **I3** заносимо формулу =F$22*D$23*КОРЕНЬ(1/B$21+H3/H$20).
- значення $(\hat{y}_{min})_i$ знаходимо за формулами (1.7) в **J3:L17**, **J3** =E3-I3. Аналогічно $(\hat{y}_{max})_i$ в **K3:K17**, **K3** =E3+I3.

3.2. Згідно з виразом (1.8) визначаємо коефіцієнт еластичності для 15-ти значень *x* в **L3:L17** за формулою **L3** =B$22*A3/E3.

3.3. Прогнозне значення фактора x_p (табл.1.1) занесемо в 18 рядок, в комірку **А18**.

3.4. Значення прогнозу показника підрахуємо в **Е18** за формулою =В22*А18+В23, або просто копіюємо формулу з **Е17** в **Е18**.

3.5. Для прогнозних значень визначимо: значення $(x_p - \bar{x})^2$ в **Н18**, $\Delta\hat{y}_p$ в **І18**, $\hat{y}_{min}$ в **J18**, $\hat{y}_{max}$ в **K18**, $K_{x_{18}}$ в **L18**. Для цього достатньо просто скопіювати формули з вище розташованих відповідних комірок.

4. Для наочного уявлення результатів будуємо графіки:

- статистичних даних (залежність показника y (В3:В17) від фактора x (А3:А17)) і лінії регресії (залежність значень $\hat{y}$ (Е3:Е17) від фактора x (А3:А17));
- лінії регресії (дублюємо) та її довірчої зони $\hat{y}_{min}$, $\hat{y}_{max}$ (відповідно залежність значень (J3:J17) та (K3:K17) від фактора x (А3:А17)).
- гістограма для коефіцієнта еластичності K_x (L3:L17).

Приклад графіків, що ілюструють результати даної лабораторної роботи, наведено в додатку А.

Орієнтовний порядок побудови графіків може бути наступним

4.1. Наводимо курсор на ***Мастер диаграмм*** і натискаємо ліву клавішу миші.

4.2. Як тільки відпустимо ліву кнопку миші, відкривається перше діалогове вікно ***Тип диаграммы***. На вкладці ***Стандартные*** вибираємо тип діаграми, наприклад, ***Точечная*** і її вид – **5**. Для переходу до наступного діалогового вікна натискаємо кнопку ***Далее>.***

4.3. У вікні ***Источник данных диаграммы*** на вкладці ***Диапазон данных*** вказуємо, що дані знаходяться в стовпцях, а на вкладці ***Ряд*** в однойменному полі, шляхом натискання на кнопку ***Добавить,*** створюємо необхідну кількість рядів, які будуть відображені на графіку. Для роботи з відповідним рядом необхідно його відмітити мишою. Для кожного ряду в полі ***Имя*** вводиться ім'я, а в полях ***Значения X*** та ***Значения Y*** – відповідно діапазони значень фактора та показника. Для цього курсор ставиться у відповідному полі, наприклад, ***Значения X*** після чого натискається кнопка, яка знаходиться в кінці цього рядка. В результаті програма повертається до роботи з таблицею і користувач має змогу відмітити мишею потрібні для побудови графіка блоки, наприклад: *A3-A18.* Або, можна ввести ці ж дані вручну. Для переходу до наступного діалогового вікна натискаємо кнопку ***Далее>.***

4.4. У вікні ***Параметры диаграммы*** на вкладці ***Заголовки*** вводять назву графіка і назву координатних вісей**.** При необхідності, або ж за бажанням користувача, на інших вкладках змінюється решта параметрів діаграми після чого знову таки натискаємо кнопку ***Далее>***.

4.5. У діалоговому вікні ***Размещение диаграмм*** необхідно вказати куди саме ви бажаєте помістити діаграму: на окремий аркуш чи в поле робочого аркуша, наприклад, нижче таблиці. Після останнього вікна натискаємо клавішу ***Готово,*** після чого на робочому аркуші з'являється графік.

Якщо на будь-якому кроці отриманий результат не задовольняє, можна повернутися до попереднього діалогу, натиснувши кнопку ***<Назад.*** Для редагування графіка (або його частин) необхідно навести на нього курсор і натиснути 2 рази на ліву клавішу миші.

5. Перед тим як ***роздрукувати робочий аркуш,*** необхідно в меню ***Файл*** знайти команду ***Параметри страницы,*** після чого відкриється вікно діалогу, яке має такі вкладки: ***Страница, Поля, Колонтитулы, Лист.***

5.1. Вибравши вкладку ***Страница,*** у групі ***Ориентация*** задаємо розташування паперового аркуша при роздрукуванні. За допомогою групи ***Масштаб*** можна зменшити (збільшити) зображення таблиць. При цьому, вибравши опції ***Разместить лист на,*** можна задати, на скількох сторінка паперу розташувати робочий аркуш Excel в довжину та в ширину. Вибір опції ***Установить*** дозволяє задати коефіцієнт масштабування у процентах від нормального розміру таблиці.

5.2. За допомогою вкладки ***Поля*** необхідно встановити відступи від країв паперу та вирівнювання інформації на аркуші. В групі ***Центрирование*** знаходяться перемикачі горизонтальні та вертикальні, які потрібні для визначення положення таблиці на паперовому аркуші по відношенню до його країв.

5.3. За допомогою вкладки ***Лист*** задаємо опції друку для таблиці. У полі ***Выводить на печать диапазон*** вказуємо діапазон комірок, необхідних для друку. Якщо це поле пусте, тоді друкується вся таблиця. При установці перемикача ***Сетка*** в групі ***Печатать*** надрукована таблиця буде мати сітку таку ж, як на екрані, що дуже зручно для зображення таблиць на папері.

5.4. Для друку таблиці вибираємо команду меню ***Файл Печать***. Тут відмічаємо необхідні умови для друку і натискуємо клавішу ***Enter*** або ***Ok***.

Индивидуальні завдання:

1. Обчисліть коефіцієнти регресії прямої для біологічних ознак:

Номер варіанту	X									
a	2	2	4	1	4	2	2	2	3	4
b	3	6	2	5	2	4	4	4	3	2
c	7	4	6	5	6	3	6	6	5	5
d	3	4	4	3	2	7	2	3	4	3
e	2	7	6	9	5	5	3	6	20	5
f	50	35	45	40	40	30	50	45	22	45
g	2	8	2	2	5	10	2	4	8	5
h	1	2	2	4	2	7	1	2	5	2
k	10	10	8	11	8	10	10	8	10	8
m	6	8	6	6	7	8	6	6	6	7
n	4	5	4	4	4	5	4	4	4	4
p	7	6	7	7	7	6	7	7	7	7
q	3	3	4	3	8	3	3	3	3	3

Контрольні запитання

1. Для чого використовується проста лінійна регресія?
2. Що показують нахил і перетин лінії регресії?
3. За яким критерієм і з якою метою проводять перевірку моделі на адекватність?
4. Який метод використовується для обчислення параметрів лінії регресії?
5. Для чого знаходяться інтервали довіри?
6. Що таке прогнозування і для чого воно використовується?
7. Що характеризує коефіцієнт кореляції?
8. Як оцінити вплив будь-якого фактора на показник?

Лабораторно-практичне заняття №4

Тема: Моделі зв'язку. Нелінійна множинна регресія

Мета: вивчити технологію побудови нелінійних регресійних моделей.

Хід виконання роботи:

ПРИКЛАД 4.1. Дано показники X і Y. Знайти регресійну модель, що описує собівартості продукції від об'єму виробництва.

Y / X	10	15	20	25
25	-	-	1	2
50	-	2	2	-
75	-	5	3	1
100	1	3	-	-
125	3	1	1	-

Розв'язок. Для проведення регресійного аналізу за даними побудуємо кореляційну таблицю

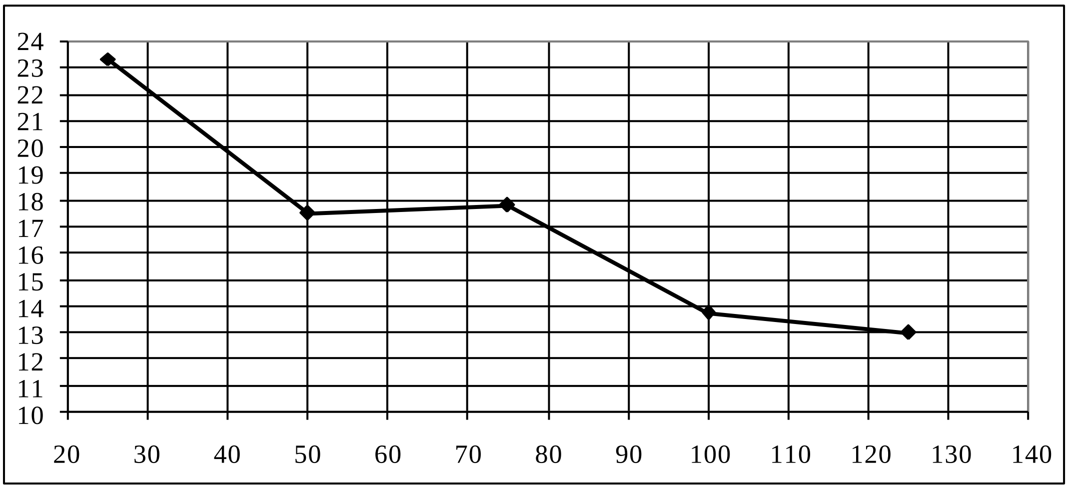
За даними кореляційної таблиці побудуємо ряд, що відображає залежність середнього значення Y від X, для чого знайдемо середні значення $\overline{y}_{x_i}$ для кожного значення x_i, $i=\overline{1,5}$, і заповнимо таблицю:

$$\overline{y}_{x_1}=\frac{y_1n_{11}+y_2n_{12}+y_3n_{13}+y_4n_{14}}{n_1}=\frac{20\cdot1+25\cdot2}{3}\approx23{,}33;\quad \overline{y}_{x_2}=\frac{15\cdot2+20\cdot2}{4}=17{,}5;$$

$$\overline{y}_{x_3}=\frac{15\cdot5+20\cdot3+25\cdot1}{9}=17{,}78;\quad \overline{y}_{x_4}=\frac{10\cdot1+15\cdot3}{4}=13{,}75;$$

$$\overline{y}_{x_5}=\frac{10\cdot3+15\cdot1+20\cdot1}{5}=13.$$

Перший етап аналізу: визначимо вид залежності Y від X. Побудуємо емпіричну лінію регресії.



Оскільки емпірична лінія регресії наближається до гіперболи, то висуваємо гіпотезу про гіперболічну залежність Y від X, тобто рівняння регресії будемо шукати у вигляді $\overline{y_x} = \frac{a}{x} + b$.

Другий етап: знайдемо параметри a, b рівняння регресії, для чого складемо систему для згрупованих даних. Необхідні розрахунки для зручності оформимо у вигляді таблиці, в останньому рядку якої знайдемо відповідні стовпцям суми.

Отже, складемо систему для знаходження параметрів рівняння регресії та розв'яжемо її за правилом Крамера:

$$\begin{cases} a\sum_{i=1}^{k}\frac{1}{x_i^2}n_i + b\sum_{i=1}^{k}\frac{1}{x_i}n_i = \sum_{i=1}^{k}\frac{1}{x_i}\overline{y_{x_i}}n_i \\ a\sum_{i=1}^{k}\frac{1}{x_i}n_i + b\sum_{i=1}^{k}n_i = \sum_{i=1}^{k}\overline{y_{x_i}}n_i \end{cases} \Rightarrow \begin{cases} 0{,}0087a + 0{,}4b = 7{,}4032 \\ 0{,}4a + 25b = 420{,}01 \end{cases}.$$

Головний визначник системи: $\Delta = \begin{vmatrix} 0{,}00872 & 0{,}4 \\ 0{,}4 & 25 \end{vmatrix} = 0{,}00872 \cdot 25 - 0{,}4^2 = 0{,}058$.

Допоміжні визначники: $\Delta a = \begin{vmatrix} 7{,}4032 & 0{,}4 \\ 420 & 25 \end{vmatrix} = 7{,}4032 \cdot 25 - 0{,}4 \cdot 420 = 17{,}076$;

$$\Delta b = \begin{vmatrix} 0{,}00872 & 7{,}4032 \\ 0{,}4 & 420 \end{vmatrix} = 0{,}00872 \cdot 420 - 7{,}4032 \cdot 0{,}4 = 0{,}701207.$$

Формули Крамера: $a = \frac{\Delta a}{\Delta} = \frac{17{,}076}{0{,}058} \approx 294{,}41$; $b = \frac{\Delta b}{\Delta} = \frac{0{,}7012207}{0{,}058} \approx 12{,}09$.

Отже, шукане рівняння регресії має вигляд $\overline{y_x} = \frac{294,41}{x} + 12,09$.

Третій етап: перевіримо правильність побудови моделі за рівнянням, її статистичну значущість за F-статистикою і адекватність вибірковим даним за коефіцієнтом детермінації. Для чого знайдемо загальну варіацію, варіації регресії та залишків; необхідні розрахунки оформимо у вигляді таблиці.

Знайдемо $\bar{y}$: $\bar{y} = \frac{\sum_{i=1}^{k} \overline{y_{x_i}} n_i}{n} \approx 16,8$.

Отже, $Q = 247,936$; $Q_p = 201,096$; $Q_o = 46,840$; тоді основне варіаційне рівняння $Q = Q_p + Q_o$ для побудованої моделі має вигляд: $247,936 = 201,096 + 46,840$ і є тотожністю, тому рівняння регресії побудовано правильно.

Для перевірки статистичної значущості рівняння регресії знайдемо F-статистику, враховуючи, що *n*=25, *l*=2 – оскільки шукали рівняння з двома параметрами:

$$F = \frac{Q_p\left(n-l\right)}{Q_o\left(l-1\right)} = \frac{201,096\left(25-2\right)}{46,840\left(2-1\right)} \approx 98,75.$$

Знайдемо $F_{кр}$: $F_{кр} = FРАСПОБР\left(0,001, 2-1, 25-2\right) \approx 14,20$. Розраховане значення F-статистики більше критичного, тому регресійна модель є статистично значущою на рівні 0,001.

Знайдемо коефіцієнт детермінації R^2: $R^2 = \frac{Q_p}{Q} = \frac{201,096}{247,936} \approx 0,81$. Значення коефіцієнта детермінації свідчить, що 81% варіації результативної ознаки *Y* пояснюється рівнянням регресії.

Висновок: Залежність *Y* від *X* описується рівнянням $\overline{y_x} = \frac{294,41}{x} + 12,09$, яке є статистично значимим на рівні значущості 0,001 та описує 81% вибіркових даних.

Індивидаульні завдання:

1. Для заданої форми стохастичної залежності на основі статистичних даних фактора *x* і показника *y* знайти оцінки:
   - параметрів кривої зростання b_0, b_1;
   - коефіцієнта кореляції r_{yx}.
2. Використовуючи критерій Фішера, з надійністю α=0,05 оцінити адекватність прийнятої моделі статистичним даним.
3. Якщо модель адекватна статистичним даним, то знайти:

- з надійністю α=0,05 надійні зони базисних даних;
- прогноз показника та його надійні інтервали;
- коефіцієнт еластичності для базисних даних і прогнозу.

4. Побудувати графіки:

- статистичних даних;
- кривої зростання і її довірчої зони;
- коефіцієнта еластичності.

Значення фактора

N	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
1	1,01	9,35	9,33	1,15	1,01	1,08	1,10	1,02	0,32	0,11	1,03
2	1,51	15,26	15,72	1,64	1,35	1,53	1,33	1,59	0,93	0,39	1,63
3	2,02	20,67	20,87	2,28	1,71	2,05	1,58	2,12	1,59	0,66	2,16
4	2,51	26,49	26,74	2,77	2,09	2,58	1,81	2,61	2,29	0,89	2,71
5	3,01	33,14	31,95	3,42	2,49	3,02	2,09	3,05	2,92	1,15	3,26
6	3,49	39,39	37,97	3,78	2,85	3,58	2,32	3,56	3,62	1,43	3,77
7	3,98	44,49	44,61	4,49	3,28	4,06	2,59	4,00	4,30	1,67	4,35
8	4,48	50,71	51,32	5,08	3,64	4,56	2,85	4,50	4,99	1,95	4,91
9	4,99	56,01	58,23	5,63	4,04	5,01	3,14	5,03	5,63	2,23	5,50
10	5,49	63,00	64,57	6,11	4,47	5,51	3,43	5,56	6,25	2,45	6,01
11	5,97	68,79	71,32	6,79	4,83	6,06	3,69	6,04	6,90	2,72	6,60
12	6,47	74,97	77,53	7,39	5,20	6,52	3,90	6,50	7,56	2,93	7,20
13	6,98	81,39	84,35	7,84	5,70	7,02	4,20	7,01	8,21	3,13	7,78
14	7,51	87,91	90,87	8,48	6,02	7,53	4,42	7,58	8,91	3,41	8,35
15	7,99	93,94	96,33	8,57	6,52	8,05	4,72	8,04	9,57	3,63	8,90
x_p	**8,45**	**100,15**	**102,68**	**9,46**	**6,90**	**8,48**	**4,99**	**8,52**	**10,20**	**3,85**	**9,42**

Значення показника

N	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}

1	12,03	8,68	9,67	5,44	5,02	4,37	2,72	3,02	5,71	1,78	1,27
2	8,84	9,86	11,82	7,01	5,51	4,01	2,91	2,94	5,97	2,22	1,42
3	6,99	10,65	13,76	10,62	6,52	3,29	3,18	2,71	3,82	4,30	1,93
4	6,03	11,20	13,48	13,40	7,92	3,10	3,50	3,32	3,27	5,49	2,35
5	5,55	11,87	14,99	17,42	9,64	3,22	3,71	2,69	3,06	6,57	2,73
6	5,12	12,32	16,74	21,70	10,06	2,99	3,88	2,99	2,86	7,15	3,93
7	4,63	12,59	16,41	28,60	14,06	2,90	4,06	3,22	2,79	10,48	5,12
8	4,41	12,85	17,88	35,68	19,17	2,37	4,18	3,42	2,63	12,52	6,55
9	4,01	13,18	19,88	41,37	25,58	1,87	4,39	4,00	2,54	17,53	9,05
10	3,96	13,40	20,50	49,45	32,56	1,82	4,44	3,43	2,60	24,23	12,24
11	3,70	13,71	20,01	60,03	37,02	1,89	4,55	4,49	2,41	31,95	17,28
12	3,75	13,91	22,26	68,90	47,97	2,28	4,63	4,44	2,47	41,65	25,25
13	3,72	14,11	22,31	76,80	61,09	1,46	4,61	4,83	2,41	52,00	36,19
14	3,53	14,20	23,34	89,84	70,61	1,56	4,66	4,62	2,50	73,81	52,81
15	3,25	14,38	22,68	92,49	88,60	1,73	4,66	5,01	2,45	95,47	76,20

Порядок виконання роботи

1.1. Запустіть програму EXCEL. Згідно з номером варіанту, виданого викладачем, занесемо початкові значення фактора x в блок **A3:A17**, а значення показника y в блок **B3:B17**. При використанні стохастичних залежностей виду (1)-(5) з табл.2.1 результати дослідження нелінійної регресійної моделі можна подати у формі, представленій в табл.2.4. При використанні інших функцій необхідно самостійно розробити розрахункову таблицю. Таким чином, подальші рекомендації по побудові й дослідженню регресійної моделі будуть стосуватися функцій нелінійних лише за фактором, для яких достатньо зробити тільки одну заміну.

1.2. Прогнозне значення фактора x_p занесемо в 18 рядок, в комірку **A18**.

1.3. Приведемо залежність до лінійного виду, ввівши заміну $z = \varphi(x)$. Для цих даних зарезервуємо блок **C3:C17.** Наприклад, при використанні функції в комірці **C3** записується формула =КОРЕНЬ(A3) і копіюється в решту комірок блоку.

1.4. Знайдемо параметри b_1 і b_0 за допомогою МНК за формулами з урахуванням введеної заміни. Для цього необхідно підрахувати добутки $z_i y_i$ та z_i^2 і відповідні суми.

- Для знаходження добутку $z_i y_i$ використаємо блок **D3:D17**. Для комірки **D3** формула запишеться: =C3*B3, яку копіюємо в блок **D4:D17**.

- Значення z_i^2 занесемо в блок **E3:E17**. Для клітинки **E3** формула запишеться: =C3^2, дану формулу копіюємо в блок **E4:E17**.

1.5. В 20 рядку знаходимо суми відповідних значень, використовуючи математичну функцію **СУММ(блок)**. Так значення $\sum x_i$ підрахуємо в **A20** за формулою =СУММ(A3:A17). Цю формулу копіюємо в комірки **B20** , **C20**, **D20** і **E20**.

1.6. Кількість значень факторів *n=15* занесемо в комірку **B21**. Надалі комірка, яка знаходиться ліворуч відносно комірки, в якій безпосередньо розраховується або ж просто міститься скалярний параметр, буде використовуватися для підпису даних. В даному випадку в комірці **A21** запишемо *n*= (див. табл.2.4).

1.7. Підставимо адреси знайдених сум в формули (1.2) для розрахунку параметрів b_1 та b_0.

- Формулу для визначення параметра b_1 занесемо в комірку **B22** =(B21*D20-C20*B20)/(B21*E20-C20^2).
- Значення параметра b_0 занесемо в **B23** =(B20-B22*C20)/B21.

1.8. Для знаходження розрахункових значень показника $\hat{y}_i$ використовується рівняння регресії (2.2). Підставляючи в формулу (2.2) адреси оцінених параметрів (абсолютне посилання) та плинні значення фактора *z* (відносне посилання) заповнимо блок **F3:F17.** Для комірки **F3** маємо формулу =B$22*C3+B$23;

- в 20-му рядку підрахуємо суму $\sum \hat{y}_i$;
- при правильному виконанні розрахунків $\sum y_i = \sum \hat{y}_i$, тобто значення комірок **B20** та **F20** повинні співпадати.

1.9. Для оцінки коефіцієнта кореляції будемо використовувати статистичну функцію **КОРРЕЛ(блок значень Z; блок значень Y)**. Для цього в комірку **F23** заносимо =КОРРЕЛ(C3:C17;B3:B17).

2.1. Для оцінки адекватності прийнятої моделі використовують критерій Фішера, розрахункове значення якого знаходять за формулою (1.3). Для цього розраховують допоміжні величини:

- середнє значення для z та y за допомогою статистичної функції **СРЗНАЧ(блок)** і заносять їх в комірки **D21** =СРЗНАЧ(С3:С17) та **D22** =СРЗНАЧ(В3:В17) відповідно;
- підставляючи відповідні адреси запишемо в комірці **G3** формулу для розрахунку значень $(y_i - \hat{y}_i)^2$ =(B3-F3)^2, яку скопіюємо у блок **G3:G17**;
- аналогічно значення $(\hat{y}_i - \bar{y})^2$ обчислюємо в блоці **H3:H17** за формулою в комірці **H3** =(B3-D$22)^2;
- суми цих величин знайдемо в 20 рядку.

2.2. Знайдемо розрахункове значення Фішера F в **J21** у відповідності з формулою (1.3), або =H20*(B21-2)/G20.

2.3. Табличне значення $F_{кр}$ знаходимо за допомогою статистичної функції **FРАСПОБР(α;1;n-2)** в **J22.**

3.1. Для визначення надійної зони базисних даних необхідно обчислити:

- значення функції Ст'юдента $t_{(\alpha/2,n-1)}$ за допомогою статистичної функції **СТЬЮДРАСПОБР(α;n-1)**. Підрахуємо його в **F21** =СТЬЮДРАСПОБР(0,05;14);
- оцінку $\hat{\sigma}_\varepsilon$ в **D23** за формулою (1.6), або =КОРЕНЬ(G20/(B21-2));
- блок значень $(z_i - \bar{z})^2$ в **I3:I17** за формулою в **I3** =(C3-D$21)^2 і в 20-му рядку підрахуємо суму $\sum(z_i - \bar{z})^2$;
- підставляємо отримані адреси в другу складову формули (1.5) і знаходимо в блоці **J3:J17** значення $\Delta\hat{y}_i$. Формула для комірки **J3** має вигляд: =F$21*D$23*КОРЕНЬ(1/B21+I3/I$20);
- значення $(\hat{y}_{min})_i$ знаходимо за формулами (1.7) в блоці **К3:К17** (для **К3** формула =F3-J3) та аналогічно значення $(\hat{y}_{max})_i$ в блоці **L3:L17** (для **L3** формула =F3+J3).

3.2. Знаходимо коефіцієнт еластичності для 15-ти значень x в блоці **M3:M17** за відповідною формулою, наведеною в табл.2.1.

3.4. Значення прогнозу показника підрахуємо у 18-му рядку відповідного стовпчика за формулою (2.2), або копіюємо формулу з попередньої комірки.

3.5. У 18-му рядку у відповідних стовпчиках для прогнозного значення визначимо: значення z_p, $(z_p - \bar{z})^2$, $\Delta\hat{y}_p$, $\hat{y}_{min}$, $\hat{y}_{max}$, K_{x_p}.

4. Для наочного уявлення результатів будуємо графіки в електронній таблиці:

- статистичних даних x і y;
- кривої зростання $\hat{y}$;
- довірчої зони $\hat{y}_{min}$, $\hat{y}_{max}$;
- коефіцієнта еластичності K_x.

Контрольні запитання

1. Для чого використовується нелінійна регресія?
2. Відмінність між лінійною та нелінійною регресіями.
3. Назвіть переваги нелінійної регресії.
4. Як розраховуються параметри кривих зростання?
5. Як прогнозувати за кривими зростання?
6. Чи завжди потрібно перевіряти модель на адекватність?

Лабораторно-практичне заняття №5

Тема: Моделі зв’язку. Множинна лінійна

Мета: навчитися будувати регресійні моделі з більш ніж однією незалежною змінною, перевіряти незалежні змінні на наявність мультиколінеарності та оцінювати невідомі параметри моделі.

Завдання до роботи

Показник *y* залежить від трьох факторів (табл.).

1. На основі біологічних даних за 16 періодів побудувати кореляційну матрицю.
2. Використовуючи критерій χ^2 з надійністю *P*=0,95 оцінити наявність загальної мультиколінеарності.
3. Якщо існує загальна мультиколінеарність, то, використовуючи *t*-статистику з надійністю *P*=0,95, виявити пари факторів, між якими існує мультиколінеарність. Якщо такі пари існують, то один із факторів цієї пари виключити із розгляду.
4. Використовуючи МНК в матричній формі знайти параметри множинної регресії.
5. Використовуючи критерій Фішера, з надійністю *P*=0,95 оцінити адекватність прийнятої математичної моделі статистичним даним.
6. Якщо модель адекватна статистичним даним, то, використовуючи *t*-статистику з надійністю *P*=0,95, оцінити значущість параметрів регресії.
7. Знайти значення прогнозу показника для заданих значень факторів та його довірчий інтервал з надійністю *P*=0,95.
8. Обчислити частинні коефіцієнти еластичності для точки прогнозу.

Значення факторів x_1, x_2, x_3 та показника y

Варіант 1				Варіант 2				Варіант 3			
x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3	y
2,31	10,1	8,315	7,627	2,12	9,97	6,28	7,45	2,41	10,25	6,32	7,73
4,67	11,7	7,729	10,7	4,40	11,43	7,64	10,48	4,78	11,95	7,79	10,94

6,17	13,9	8,479	11,53	6,16	13,61	8,25	11,25	6,26	13,87	8,57	11,82
8,7	14,4	8,691	13,4	8,69	14,28	8,61	13,33	8,95	14,52	8,83	13,59
10,7	15,1	10,5	17,02	10,47	14,93	10,21	16,90	10,75	15,18	10,55	17,15
13,5	17,1	10,52	18,75	13,41	17,04	10,43	18,57	13,57	17,41	10,59	18,81
16,2	18,9	11,68	21,14	15,98	18,82	11,54	20,91	18,43	19,01	11,83	21,26
18,3	20,3	13,77	23,37	18,24	20,30	13,73	23,32	18,55	20,52	13,83	23,38
21,2	21,7	13,7	27,45	20,89	21,48	13,64	27,16	21,45	22,02	13,97	27,62
22,7	22,4	14,43	27,13	22,66	22,22	14,41	27,00	22,75	22,53	14,55	27,18
25,1	22,5	14,07	29,61	24,91	22,41	13,98	29,59	25,19	22,64	14,17	29,87
26,1	24,7	16,46	32,52	26,03	24,45	16,45	32,24	26,25	24,89	16,64	32,64
27,5	24,8	15,02	31,8	27,25	24,75	14,83	31,71	27,63	25,02	15,07	32,01
29,9	25	15,27	35,18	29,74	24,89	15,06	35,10	30,16	25,19	15,38	35,25
32,1	26	15,66	37,07	31,80	25,95	15,61	36,77	32,20	26,15	15,72	37,14
33,7	27,4	17,21	38,85	33,66	27,27	17,14	38,70	33,94	27,60	17,27	38,96
35,8	28,9	17,47	?	35,57	28,80	17,39	?	36,00	29,02	17,56	?

Варіант 4				Варіант 5				Варіант 6			
x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3	y
2,37	10,27	6,40	7,73	2,25	9,90	6,09	7,49	2,61	10,35	6,61	7,72
4,77	11,87	7,88	10,85	4,42	11,54	7,49	10,64	4,89	11,78	7,94	10,77
6,24	13,88	8,50	11,54	6,08	13,73	8,46	11,44	6,24	14,09	8,62	11,86
8,70	14,62	8,86	13,52	8,65	14,26	8,59	13,24	9,01	14,64	8,83	13,73
10,79	15,28	10,51	17,13	10,64	14,91	10,43	16,99	10,79	15,17	10,68	17,04
13,60	17,29	10,53	18,75	13,29	17,02	10,52	18,57	13,53	17,42	10,66	18,80
16,31	19,04	11,74	21,15	15,95	18,84	11,65	21,07	16,32	19,24	11,78	21,28
18,40	20,45	13,96	23,49	18,25	20,06	13,55	23,23	18,60	20,60	13,78	23,70
21,25	21,94	13,86	27,50	21,10	21,71	13,67	27,37	21,48	22,04	13,74	27,63
22,87	22,55	14,60	27,16	22,67	22,31	14,33	27,12	23,02	22,69	14,56	27,45
25,15	22,56	14,24	29,73	24,99	22,39	13,95	29,42	25,17	22,65	14,09	29,71
26,27	24,79	16,59	32,71	26,00	24,50	16,34	32,43	26,40	24,83	16,66	32,80
27,70	24,82	15,03	31,83	27,34	24,76	14,81	31,74	27,62	24,82	15,12	31,81
30,00	25,11	15,34	35,18	29,75	24,99	15,13	35,16	30,19	25,17	15,42	35,22
32,25	26,11	15,84	37,12	31,87	25,94	15,46	37,07	32,25	26,22	15,77	37,26
33,85	27,58	17,30	38,97	33,55	27,35	16,96	38,74	33,76	27,72	17,40	39,20
35,96	29,09	17,48	?	35,71	28,83	17,25	?	35,97	29,15	17,77	?

Варіант 7				Варіант 8				Варіант 9			
x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3	y
2,61	10,32	6,71	7,95	2,40	11,27	6,78	8,61	1,43	9,18	6,17	7,29
4,97	11,96	8,14	10,75	5,66	12,48	8,49	11,33	3,92	10,94	7,70	9,79
6,55	13,96	8,95	11,66	6,87	14,11	9,46	12,48	5,49	13,30	8,43	11,26
9,16	14,62	8,83	13,49	9,52	14,64	9,44	14,37	8,17	13,89	8,12	12,42
10,68	15,18	10,95	17,44	11,57	16,17	10,82	17,43	9,68	14,49	10,41	16,04
13,93	17,34	10,92	19,05	13,73	17,47	11,01	19,50	13,42	16,73	10,39	18,34
16,27	19,21	11,75	21,42	16,87	19,85	12,16	21,22	15,92	17,97	11,36	20,94
18,74	20,36	14,05	23,85	18,73	21,43	14,46	23,84	18,04	19,45	13,32	22,74
21,31	22,21	14,09	27,92	22,16	22,12	13,98	27,80	20,69	21,49	12,72	27,09
22,88	22,84	14,60	27,27	23,33	22,98	14,72	27,82	22,68	21,80	14,22	26,43
25,13	22,63	14,38	30,04	25,84	22,61	14,58	30,01	24,33	21,64	13,51	29,39
26,28	24,93	16,57	32,83	26,69	25,03	17,07	33,32	25,64	24,48	15,83	32,37
27,71	24,94	15,51	31,89	27,78	25,16	15,39	31,87	27,14	24,02	14,52	31,52
30,01	25,27	15,44	35,21	30,86	26,05	15,95	35,35	29,22	24,42	15,06	34,89
32,08	26,18	16,00	37,27	32,87	26,44	16,37	37,50	31,09	25,96	15,32	36,33
33,74	27,61	17,59	38,99	34,65	28,09	17,94	39,91	33,34	27,05	16,65	38,35
36,31	29,23	17,97	?	36,25	29,76	18,44	?	35,64	28,12	16,55	?

Порядок виконання роботи

1.1. Запустіть програму EXCEL. Згідно з номером варіанту, виданого викладачем, занесемо початкові значення фактора x_1 в блок **A3:A18**, x_2 в блок **B3:B18**, x_3 в блок **C3:C18**, значення показника y в **D3:D18**. При формуванні розрахункової таблиці рекомендується дотримуватися форми, наведеної в табл.3.2.

1.2. В 20 рядку будемо знаходимо суми відповідних значень. $\sum y_i$ підрахуємо в **D20** за формулою =СУММ(D3:D18).

1.3. Кількість значень факторів або число періодів n=16 занесемо в комірку **B21**, а кількість факторів m=3 – в комірку **D21.**

1.4. Знайдемо кореляційну матрицю, яка визначає зв’язок між трьома факторами.

$$K = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}.$$

Для знаходження коефіцієнтів кореляції використовують статистичну функцію **КОРРЕЛ(блок1;блок2)**.

Наприклад, для знаходження r_{21} необхідно використати вказану функцію і вказати блоки значень x_2 та x_2, формула =КОРРЕЛ(B3:B18;A3:A18).

1.5. Занесемо кореляційну матрицю *К* в блок **A24:C26**.

Формулу заносимо в комірку **A24** =КОРРЕЛ($A3:$A18;A3:A18). Цю формулу копіюємо в комірку **B24** і **C24**.

Для комірки **A25** формула наступна: =КОРРЕЛ($B3:$B18;A3:A18). Цю формулу копіюємо в комірку **B25** і **C25**.

І для комірки **A26** формула: =КОРРЕЛ($C3:$C18;A3:A18). Цю формулу копіюємо в комірку **B26** і **C26**.

Отримана матриця складається зі значень коефіцієнтів кореляції, менших 1, а на головній діагоналі розміщені 1.

2.1. Для дослідження загальної мультиколінеарності між факторами використовується χ^2- критерій, розрахункове значення якого знаходять за формулою (3.3).

2.2. Підрахуємо визначник кореляційної матриці в комірці **E24**, використовуючи математичну функцію =МОПРЕД(A24:C26).

2.3. Значення $\chi^2_{роз}$ занесемо в **E25** =-(B$21-1-(2*D$21+5)/6)*LN(E$24).

2.4. Табличне значення χ^2-критерію знаходять для заданої ймовірності α=0,05 і числа ступенів вільності *k=0,5m(m-1)=3* за допомогою статистичної функції **ХИ2ОБР(α;k)** в комірці **E26**.

Якщо $\chi^2_{роз} > \chi^2_{табл}$, то з надійністю Р=0,95 можна вважати, що між факторами існує загальна мультиколінеарність.

3.1. Для з'ясування питання, між якими факторами існує мультиколінеарність, використовується *t*-критерій.

3.2. Необхідно визначити для трьох пар факторів три розрахункових значення статистик t_{12} , t_{13} , t_{23} за формулами (3.6). Для цього спершу необхідно знайти обернену матрицю $[Z]=[K]^{-1}$ вигляду (3.4), використовуючи математичну функцію **МОБР(блок початк. матриці)**. Нехай матриця Z займає блок **A28:C30**. Введемо формулу =МОБР(A24:C26). Зауважимо, що натискання клавіші Enter не забезпечує введення формули масиву. В цьому випадку необхідно натиснути комбінацію Shift+Ctrl+Enter.

3.3. Значення r_{ij}^* занесемо в **E28:E30**:

$r_{12}^* = -\frac{z_{12}}{\sqrt{z_{11}z_{22}}}$ в **E28** = - B28/КОРЕНЬ(A28*B29); $r_{13}^* = -\frac{z_{13}}{\sqrt{z_{11}z_{33}}}$ в **E29** = - C28/КОРЕНЬ(A28*C30); $r_{23}^* = -\frac{z_{23}}{\sqrt{z_{22}z_{33}}}$ в **E30** = - C29/КОРЕНЬ(B29*C30).

3.4. Тепер є всі дані для знаходження значень статистик t_{12} , t_{13} , t_{23}. Маємо в комірках **G28:G30** відповідно:

$t_{12} = \frac{r_{12}^* \sqrt{n-m-1}}{\sqrt{1-(r_{12}^*)^2}}$ в **G28** =E28*КОРЕНЬ((B$21-D$21-1)/(1-E28^2));

$t_{13} = \frac{r_{13}^* \sqrt{n-m-1}}{\sqrt{1-(r_{13}^*)^2}}$ в **G29** =E29*КОРЕНЬ((B$21-D$21-1)/(1-E29^2));

$t_{23} = \frac{r_{23}^* \sqrt{n-m-1}}{\sqrt{1-(r_{23}^*)^2}}$ в **G30** =E30*КОРЕНЬ((B$21-D$21-1)/(1-E30^2)).

3.5. Далі необхідно визначити табличне значення t-критерію для ймовірності α=0,05 і числа ступенів вільності *k=n-m-1=12*. Значення $t_{табл}$ знайдемо в комірці **I29** за формулою =СТЬЮДРАСПОБР(0,05;12).

3.6. Після цього кожне з розрахункових значень t_{12} , t_{13} , t_{23} порівнюють з табличним $t_{табл}$. Якщо t_{ij} > $t_{табл}$, то між факторами x_i та x_j існує мультиколінеарність. Для усунення мультиколінеарності один з факторів виключають з розгляду.

Наприклад, якщо $t_{13}>t_{табл}$, тоді між факторами x_1 та x_3 існує мультиколінеарність і виключають з розгляду один з пари факторів або x_1, або

x_3. Або, якщо одночасно $t_{12}>t_{табл}$ і $t_{23}>t_{табл}$, то окремо між парами факторів x_1 і x_2 та x_2 і x_3 існує мультиколінеарність і виключають з розгляду той фактор, який повторюється в двох парах факторів. В даному випадку виключають фактор x_2.

4.1.Тепер показник *y* залежить від двох факторів x_1 і x_3. Припустимо, що між *y* та x_1 і x_3 існує лінійна залежність:

$$\hat{y} = b_0 + b_1 x_1 + b_3 x_3.$$

Знайдемо параметри регресії b_0, b_1, b_3, використовуючи МНК в матричній формі за формулою (3.7).

4.2. Формуємо матрицю $[X]$ в діапазоні **I3:K18**, для цього в блок **I3:I18** занесемо число 1, в блок **J3:J18** копіюємо значення фактора x_1 з блоку **A3:A18** і в блок **K3:K18** копіюємо значення фактора x_3 з блоку **C3:C18**.

4.3. Матриця $[Y]$ включає значення показника *y* і знаходиться в діапазоні **D3:D18**.

4.4. Знайдемо добуток матриць $[X]^T[X]$ в діапазоні **A32:C34** за формулою =МУМНОЖ(ТРАНСП(I3:K18);I3:K18).

4.5. Знайдемо обернену матрицю $\left[[X]^T[X]\right]^{-1}$ в **E32:G34** ввівши формулу =МОБР(A32:C34).

4.6. Знайдемо добуток матриць $[X]^T[Y]$ в діапазоні **I32:I34** за формулою =МУМНОЖ(ТРАНСП(I3:K18);D3:D18).

4.7. Перемножимо матриці $\left[[X]^T[X]\right]^{-1}$ та $[X]^T[Y]$ і знайдемо матрицю $[B]$ в **K32:K34** за формулою =МУМНОЖ(E32:G34;I32:I34). Ця матриця складається з параметрів b_0, b_1, b_3 і вони знаходяться в комірках: **K32, K33** і **K34** відповідно.

4.8. Знайшовши параметри регресії, можна підставити їх в рівняння регресії і знайти розрахункові значення показника *y*. Ці значення знайдемо в діапазоні **E3:E18**, для цього в комірку **E3** занесемо формулу =K$32+K$33*A3+K$34*C3 і потім її копіюємо до **E18**.

4.9. В **E20** знайдемо суму $\sum \hat{y}_i$, копіюючи формулу з **E20**. При правильних розрахунках $\sum y_i = \sum \hat{y}_i$ (тобто D20=E20).

5.1. Оцінимо модель на адекватність, використовуючи критерій Фішера (див. вираз (3.8)).

5.2. Підрахуємо значення $(\hat{y}_i - \bar{y})^2$ в діапазоні **F3:F18**. Для цього запишемо в комірці **F3** формулу =(E3-СРЗНАЧ(D$3:D$18))^2 і копіюємо її до **F18**.

5.3. Підрахуємо значення $(\hat{y}_i - y_i)^2$ в діапазоні **G3:G18**. Для комірки **G3** маємо формулу =(E3-D3)^2, яку копіюємо до **G18**.

5.4. Суми цих величин знайдемо в **F20** і **G20** відповідно.

5.5. Підставимо значення в формулу (3.8) і знайдемо розрахункове значення критерію $F_{розр}$ в **F22** =F20*(B21-2-1)/(G20*2).

5.6. Табличне значення Фішера знаходимо для ймовірності α=0,05 і числа ступенів вільності $k_1=m=2$, $k_2=n-m-1=13$. Підрахуємо $F_{табл}$ в **H22** за формулою =FРАСПОБР(0,05;2;13).

5.7. Якщо $F_{розр}>F_{таб}$, то прийнята модель адекватна і її можна використовувати для економічного аналізу.

6.1. Розглянемо значущість параметрів регресії b_0, b_1, b_3 за допомогою *t*-статистик, які описуються формулами вигляду (3.9).

Для їх розрахунку скористаємося діагональними елементами ζ_{ii} матриці $\left[[X]^T[X]\right]^{-1}$, розміщеної в **E32:G34**.

6.2. За формулою (3.10) знайдемо оцінку середньоквадратичного відхилення помилок. Для цього в комірці **B22** запишемо формулу =КОРЕНЬ(G20/(B21-3)).

6.3. Розрахункові значення *t*-статистики окремо для кожного параметра визначаються наступним чином:

для b_0: $t_0 = \dfrac{|b_0|}{\sigma_\varepsilon \sqrt{\zeta_{11}}}$ в **M32** =ABS(K32)/(B22*КОРЕНЬ(E32));

для b_1: $t_1 = \dfrac{|b_1|}{\sigma_\varepsilon \sqrt{\zeta_{22}}}$ в **M33** =ABS(K33)/(B22*КОРЕНЬ(F33));

для b_3: $t_3 = \dfrac{|b_3|}{\sigma_\varepsilon \sqrt{\zeta_{33}}}$ в **M34** =ABS(K34)/(B22*КОРЕНЬ(G34)).

6.4. Табличне значення значення t-критерію визначаємо для ймовірності α=0,05 і числа ступенів вільності $k=n-m-1=13$. Значення $t_{табл}$ знайдемо в комірці **M36** за формулою =СТЬЮДРАСПОБР(0,05;13).

6.5. Далі кожне з розрахункових значень t_0, t_1, t_2 порівнюють з табличним $t_{табл}$. Якщо $t_i > t_{табл}$, то вплив параметра b_i на показник y значний.

7.1. Для визначення прогнозного значення показника $\hat{y}_p$, необхідно підставити прогнозні значення факторів в рівняння регресії. Прогнозні значення факторів занесемо в 19 рядок. Значення $\hat{y}_p$ знайдемо в **E19**, копіюючи формулу з **E18**.

7.2. Для обчислення довірчого інтервалу точки прогнозу за формулою (3.11) необхідно:

▪занести значення вектора $[X_p]^T$ в блок **A36:C36**: 1 занесемо в комірку **A36**, x_{1p} - в **B36** =A19, x_{3p} - в **C36** =C19;

▪значення функції Ст'юдента, яке визначається за заданими значеннями ймовірності α=0,05 і числа ступенів вільності $k=n-m-1=13$, уже підраховане в комірці **M36**.

▪оцінка $\hat{\sigma}_\varepsilon$ вже знайдена в комірці **B22**.

7.3. Знайдемо добуток матриць $[X_p]^T \left[[X]^T [X]\right]^{-1}$ в діапазоні **A38:C38** за формулою =МУМНОЖ(A36:C36;E32:G34).

7.4. Знайдемо добуток матриць $[X_p]^T \left[[X]^T [X]\right]^{-1} [X_p]$ в комірці **E36**, використовуючи функцію =СУММПРОИЗВ(A38:C38;A36:C36).

7.5. Знайдемо значення $\Delta\hat{y}$ в комірці **E37** за формулою =M36*B22*КОРЕНЬ(E36+1).

7.6. Розраховуємо значення $y_{\min}$ та $y_{\max}$ за формулами

$\hat{y}_{\min} = \hat{y}_p - \Delta\hat{y}$ в комірці **E38** =E19-E37, $\hat{y}_{\max} = \hat{y}_p + \Delta\hat{y}$ в комірці **E39** =E19+E37.

8. Згідно з виразом (3.12) знаходимо частинні коефіцієнти еластичності для точки прогнозу для множинної лінійної регресії в таких комірках: K_{x_1} в **H38**

=К33*А19/Е19 та K_{x_2} в **Н39** =К34*С19/Е19.

9. На підставі отриманих результатів робимо економічні висновки.

Контрольні питання

1. Для чого використовується множинна лінійна регресія? Наведіть приклади.
2. Перелічіть етапи побудови множинної регресійної моделі.
3. Поясніть поняття мультиколінеарність. Опишіть процедуру перевірки факторів на мультиколінеарність.
4. Наведіть математичні вирази для розрахунку параметрів множинної регресії.
5. Як знайти довірчий інтервал точки прогнозу?
6. Що показує частковий коефіцієнт еластичності?

Лабораторно-практичне заняття №7

Тема: Основи дисперсійного аналізу.

Мета роботи: отримати навички вивчення статистичного впливу одного або декількох факторів на результативну ознаку.

Теоретичні відомості:

Дисперсійний аналіз використовує властивість адитивності дисперсії випадкової величини, що обумовлено дією незалежних факторів. В залежності від числа джерел дисперсії розрізняють *однофакторний* та *багатофакторний* дисперсійний аналіз. Дисперсійний аналіз полягає у виділенні і оцінці окремих факторів, що викликають зміну досліджуваної випадкової величини. При цьому проводиться розклад сумарної вибіркової дисперсії на складові, обумовлені незалежними факторами. Кожна з цих складових є оцінкою дисперсії генеральної сукупності. *Факторами* звуться причини зміни характеристик біологічних об'єктів, а ті характеристики біологічних об'єктів, які змінюються під їх впливом, звуться *результативними ознаками.*

Методика дисперсійного аналізу зводиться до деякої загальної схеми – алгоритму.

1. Групування вибіркових матеріалів у комбінаційну таблицю дисперсійного комплексу

2. Визначення значень: середнього арифметичного всього комплексу ($\overline{x}$) і групових середніх за градаціями організованого фактора (x_i)

3. Визначення загальної суми квадратів відхилень (D_y), тобто суми квадратів відхилень варіант від загальної середньої: $D_y = \Sigma x_2 - (\Sigma x)^2/N$

4. Визначення міжгрупової суми квадратів відхилень, яка дорівнює сумі квадратів відхилень групових середніх від загальної середньої з урахуванням статистичної ваги (n_i) групових середніх:

• у випадку рівних чисел варіант в градаціях комплексу – $D_x = n_i \Sigma \overline{x}_i^{\ 2} - (\Sigma x)^2/ N$;

• у випадку різної кількості варіант в градаціях комплексу – $D_x = \Sigma[n_i(x_i - x)^2]$;

5. Визначення внутрішньогрупової суми квадратів, тобто суми квадратів відхилень групових варіант від групових середніх: $D_z = \Sigma x^2 - n\Sigma\overline{x}_i^2$

6. Визначення дисперсій (середніх квадратів відхилень):

• загальна: $\sigma_{заг}{}^2 = D_y / (N\text{-}1)$;

• факторна: $\sigma_{факт}{}^2 = D_x / (a\text{-}1)$, де a – кількість груп;

• остаточна: $\sigma_{ост}{}^2 = D_z / (N\text{-}a)$

7. Визначення фактичного значення критерію $F_{факт.} = \sigma_{факт}{}^2 \text{-} \sigma_{ост}{}^2$;

8. Порівняння фактичного значення критерію F з його табличним (стандартним) значенням для відповідного рівня значимості (p) і даних чисел

Хід виконання роботи:

Завдання 1. Користуючись таблицею алгоритму провести дисперсійний аналіз експериментальних даних наведених в прикладі Досліджувався вплив кількості доглядів (розпушень ґрунту) на ріст 2-х річних сіянців сосни звичайної після висаджування їх в грунт. Дослід організовано на однорідній території з виділенням дослідних площ розміром 0,5 га. Сукупність рослин на кожній дослідній площі є відповідною генеральною сукупністю. Вибіркові сукупності формувались шляхом відбору по 200 сіянців (за принципом імовірності) для кожної дослідної площадки. Дослід закладався в 3-х повторностях з градаціями регульованого фактору (розпушування) – відсутність, два, чотири, шість. Всього закладено 12 пробних площ, на яких у дослідних рослин заміряні восени річні прирости за висотою. Результати замірів згруповані в таблиці.

Заміри приросту сіянців сосни звичайної

Варіанти досліду	Середній приріст рослин за повторностями (варіантами), см				Середній приріст по всіх повторностях
	1	2	3	n_i	
Контроль	6	4	8	3	
Догляди проводилися	8	8	12	3	

два рази					
Чотири рази	12	14	16	3	
Шість разів	12	12	12	3	

Однакова кількість повторностей по всіх варіантах досліду і участь лише одного регульованого фактора підтверджує класифікацію комплексу як однофакторіального і рівномірного.

Завдання 2. Проведіть дисперсійний аналіз вимірів приростів лісових культур у висоту (см/міс) при внесенні різних доз мінеральних добрив.

Варіанти факторів	Повторюваності за варіантами, x_i		
20%	2	4	4
40%	3	5	5
60%	6	7	7

Завдання 3. Провести дисперсійний аналіз впливу на висоту тополевих саджанців заготовки живців.

Фактор (живці різної довжини)	Висота по повторностях					
1	14	22	18	27	6	45
2	26	41	47	32	35	27
3	25	43	28	21	13	26
4	15	16	12	14	20	16

Завдання 4. Вивчався вплив внесення різних доз амонієвої селітри (NH4NO3) на приріст дворічних саджанців дуба звичайного в лісовому розсаднику на середньодернових слабоопідзолених ґрунтах Півдня України.

Дози добрив (д. р.), кг / га	Приріст за						*ni*	Сума (Σ)	Середній приріст, см
	1	2	3	4	5	6			
Контроль (без	4	6	8	4	6	5			

добрив)									
5	6	10	14	8	12	12			
10	10	12	18						
15	12	12							
Σ									

Контрольні запитання:

1. Поняття про дисперсійний аналіз.
2. Передумови та постановка задачі однофакторного дисперсійного аналізу.
3. Загальна, факторна та залишкова суми квадратів відхилень та зв’язок між ними.
4. Загальна, факторна та залишкова дисперсії.
5. Алгоритм однофакторного дисперсійного аналізу за Фішером.