6. МЕТОДИ АНАЛІЗУ ВЗАЄМОЗВ'ЯЗКІВ

6.1. Дисперсний аналіз

Дисперсійний аналіз дає змогу знайти вплив систематичної та випадкової варіації у загальній варіації та відповідно з'ясувати вплив факторної ознаки на зміну результативної ознаки. Використовується правило складання дисперсій, згідно з яким загальна дисперсія сукупності дорівнює підсумку двох дисперсій: середній з внутрішньогрупових дисперсій та міжгрупової дисперсії:

$$\sigma^2 = \bar{\sigma}^2 + \delta^2. \qquad (6.1)$$

З метою з'ясування кількості зв'язку між факторною та результативною ознаками використовується кореляційне відношення:

$$\eta^2 = \frac{\delta^2}{\sigma^2}, \qquad (6.2)$$

яке змінюється у межах від «0» до «1».

Як $\eta^2 = 0, (\delta^2 = 0)$ – факторна ознака не впливає на результативну.

Як $\eta^2 = 1, (\delta^2 = \sigma^2)$ – так між факторною та результативною ознаками існує функціональний зв'язок.

Для одержання висновків про практичну значимість кореляційного відношення використовують шкалу Чедока.

Таблиця 6.1

Шкала Чедока

Рівень щільності зв'язку, η^2	**0,1-0,3**	**0,3-0,5**	**0,5-0,7**	**0,7-0,9**	**0,9-1,0**
Характеристика щільності зв'язку, мовна, якісна	слабка	помірна	помірна	висока	значно висока

Окрім якісної оцінки, за шкалою Чедока, дисперсійний аналіз дає можливість оцінити щільність зв'язку з певною імовірністю. Критичне значення кореляційного відношення наведено у таблицях в залежності від відсотка забезпеченості та кількості ступенів вільності факторної дисперсії $k_1 = m - 1,$ та відповідно для випадкової дисперсії

$$k_2 = n - m, \qquad (6.3)$$

де n – кількість елементів сукупності; m – кількість груп, на яку поділена генеральна сукупність. Для α % дані наведені в таблиці.

Таблиця 6.2

Критичне значення $\eta^2_{кр}$, $\alpha = 95\%$

k_2 \ k_1	**1**	**2**	**3**	**4**	**5**	**6**	**8**
3	0,771	0,865	0,903	0,924	938	947	959
4	658	776	832	865	887	902	924
5	569	699	764	806	835	854	885
6	500	632	704	751	785	811	847
7	444	575	651	702	739	768	810
8	399	527	604	657	697	729	775
9	362	488	563	618	656	692	742
10	332	451	527	582	624	659	711

$\alpha = 99\%$

k_2 \ k_1	**1**	**2**	**3**	**4**	**5**	**6**	**8**
3	0,919	954	967	975	979	982	987
4	841	900	926	941	951	958	967
5	765	842	879	901	916	928	943
6	696	785	830	859	879	894	915
7	636	732	784	818	842	860	887
8	585	684	740	778	806	827	858
9	540	641	700	741	771	795	829
10	501	602	663	706	738	764	802

Якщо розраховане значення $\eta^2 \geq \eta^2_{кр}$, так з імовірністю α % можна стверджувати, що зв'язок між факторною та результативною ознаками істотний, а з імовірністю (100 – α) % – ні.

6.2. Перевірка істотності зв'язку за допомогою коефіцієнтів: лінійної кореляції, детермінації та рангової кореляції

Лінійний коефіцієнт кореляції Пірсона визначається формулою:

$$r = \frac{\sum_{1}^{n} xy - n\bar{x} \cdot \bar{y}}{n\sqrt{\sigma_x^2 \cdot \sigma_y^2}}.$$

Коефіцієнт детермінації визначається формулою:

$$R^2 = \frac{\delta_y^2}{\sigma_y^2}.$$

де δ_y^2 – факторна дисперсія результативної ознаки, σ_y^2 – загальна дисперсія результативної ознаки, які обчислюються формулами:

$$\delta_y^2 = \frac{1}{n}(a\sum y + b\sum xy) - (\bar{y})^2;$$

$$\sigma_y^2 = (\overline{y^2}) - (\bar{y})^2 = \frac{\sum y^2}{n} - (\bar{y})^2.$$

Критичне значення коефіцієнта детермінації залежить від відсотка забезпеченості α та кількості степенів вільності дисперсії факторної $k_1 = m-1,$ та результативної ознак – $k_2 = n-1,$ де n – кількість елементів сукупності; m – кількість груп, на яку поділена сукупність. Критичне значення коефіцієнта детермінації знаходиться за таблицею 6.2.

Якщо розраховане значення $R^2 \geq R_{êð}^2$, так з імовірністю α % можна стверджувати, що зв'язок між факторною та результативною ознаками істотний, а з імовірністю (100 – α) % – ні.

Коефіцієнт рангової кореляції Спірмена визначається за формулою:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_j^2}{n(n-1)},$$

де n – кількість рангів, d_j – відхилення рангів $d_j = (R_x - R_y)$.

Критичне значення коефіцієнта рангової кореляції, що залежить від відсотка забезпеченості α та кількості рангів n , наведено у таблиці 6.3.

Таблиця 6.3

Число рангів	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
$\rho_{кр}$	0,900	0,828	0,714	0,642	0,600	0,563	0,527	0,497

Якщо розраховане значення $\rho \geq \rho_{êð}$, так з імовірністю α % можна стверджувати, що зв'язок між ознаками істотний, а з імовірністю (100 – α) % – ні.

6.3. Перевірка істотності зв'язку за допомогою критерію Фішера

Критерій Фішера розраховується за формулою:

$$F = \frac{\delta^2}{\overline{\sigma}_2} \cdot \frac{n-m}{m-1} = \frac{\eta^2}{1-\eta^2} \cdot \frac{n-m}{m-1}. \qquad (6.4)$$

Критичні значення критерію Фішера в залежності від $\alpha\%, k_1, k_2$ ($k_1 \neq k_2$) надані у таблиці 6.4.

Таблиця 6.4

Критичне значення критерію Фішера $F_{êð}$, α = 95 %

k_2 \ k_1	**1**	**2**	**3**	**4**	**5**	**6**	**8**
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82
6	5,99	5,14	4,76	5,53	4,39	4,28	4,15
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23
10	4,96	4,10	3,71	3,48	3,33	3,	3

Якщо $F \geq F_{êð}$ – так з імовірністю α % можливо стверджувати, що зв'язок між факторною та результативною ознаками істотний, а з імовірністю (100 – α) % – ні.

6.4. Перевірка випадків (артефактів)

Нормоване відхилення допомагає визначити ознаки, які суттєво відрізняються від усіх інших ознак у групі, що може вказувати на помилку спостереження. Перевірка артефактів здійснюється за критерієм Грабса:

$$T = \frac{x_i - \overline{x}}{\sigma} \geq T_{st}; \qquad (6.5)$$

$$як - T \geq T_{st} - артефакт,$$

де T – розраховане значення критерію Грабса, T_{st} – стандартне значення критерію випадків з таблиці; x_i – ознака, яку перевіряють.

Таблиця 6.5

Стандартне значення критерію Грабса T_{st} ($\alpha = 90\%$)

n	2	3-4	5-9	10-15	16-20	21-28	29-34	35-46
T_{st}	2,0	2,1	2,2	2,3	2,4	2,5	2,6	2,7
n	47-66	67-84	85-104	105-124	125-174	175-349	350-599	>599
T_{st}	2,8	2,9	3,0	3,1	3,2	3,3	3,4	3,5

n – кількість елементів сукупності.

6.5. Перевірка закону розподілу на відповідність до нормального закону розподілу за критерієм Пірсона

Критерій Пірсона обчислюється за формулою:

$$\chi^2 = \sum_{1}^{m} \frac{(f_i - f_i')^2}{f_i'}, \qquad (6.6)$$

де f_i – фактична частота «i»-ї групи статистичної сукупності; f'_i – теоретична частота «i»-ї групи нормального закону розподілу; m – кількість груп даних; $r = 3$ – кількість обмежених зв’язків; число ступенів вільності $k = m – r$.

Частоти теоретичної кривої нормального розподілу визначаються формулою:

$$f_i' = (\sum_{1}^{m} f_i) \cdot (\frac{\Delta h}{\sigma}) \cdot f_i(t), \qquad (6.7)$$

де

$$f_i(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}; \qquad (6.8)$$

Δh – розмір інтервалу кожної групи;

$$t = \frac{x' - \bar{x}}{\sigma}.$$

Таблиця 6.3

Критичні значення критерію Пірсона χ^2 (α % = 95 %)

k	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$\chi^2_{êð}$	3,84	5,99	7,82	9,49	11,07	12,59	14,07	15,51	16,92	18,31

Якщо $\chi^2 \le \chi^2_{òàáë}$ – так при прийнятому рівні значущості приймають гіпотезу про нормальний закон розподілу.

Якщо $\chi^2 > \chi^2_{êð}$, так з надійністю α % можна вважати, що нормальний розподіл не збігається з експериментом – статистичним спостереженням.

6.6. Кореляційно-регресійний аналіз

Теоретична лінія регресії – описується певною функцією $y = y(x)$. У статистиці використовують ***рівняння регресії:***

– лінійне $y = a + bx$;
– степеневе $y = ax^b$;
– гіперболічне $y = a + b/x$;
– параболічне $y = a + bx + cx^2$, та інші, де a, b, c – коефіцієнти (параметри) регресії, які знаходять за умов мінімізації квадратів відхилень значень рівняння регресії від значення відповідних ознак:

$$\sum_{1}^{\text{ò}}(y - y_s)^2 \rightarrow \min, \qquad (6.9)$$

де y – значення знайдені із рівняння регресії, y_s – значення ознаки, які отримані унаслідок статистичного спостереження.

Для лінійного рівняння регресії $y = a + bx$ коефіцієнти (параметри регресії) дорівнюють:

$$b = \frac{n(\sum_{1}^{n} x_i y_i) - (\sum_{1}^{n} x_i)(\sum_{1}^{n} y_i)}{n(\sum_{1}^{n} x_i^2) - (\sum_{1}^{n} x_i)^2}, \qquad (6.10)$$

$$a = \bar{y} - b\bar{x}.$$

Важливою характеристикою регресійної моделі є відносний ефект впливу фактора «х» на результат «y» – ***коефіцієнт еластичності:***

$$\gamma = b \times \frac{\bar{x}}{\bar{y}}, \qquad (6.11)$$

який показує на скільки змінюється результат «у» при зміні фактора «x».

Межі випадкових коливань коефіцієнта «b» регресії ***(довірчі межі)*** становлять:

$$b_0 = b \pm \Delta b = b \pm t\mu_b, \qquad (6.12)$$

де $\Delta b = t\mu_b$,

$$\mu_b = \sqrt{\frac{\sigma_e^2}{\sigma_x^2(n-2)}}; \qquad (6.13)$$

$$\sigma_e^2 = \frac{1}{n}\sum_{1}^{n}(y - y_s)^2 \text{ – залишкова дисперсія;} \qquad (6.14)$$

$$\sigma_x^2 = \frac{1}{n}\sum_1^n (x - \bar{x})^2$$ – загальна дисперсія фактора; (6.15)

t – коефіцієнт довіри за Стьюдентом, ступінь вільності $k = n - 2$.

Довірчі межі коефіцієнта «a» становлять:

$$a \pm t\mu_a,$$

де

$$\mu_a = \sqrt{\frac{\sigma_{\hat{a}}^2}{(n-2)}(1 + \frac{\bar{x}^2}{\sigma_x^2})}, \qquad (6.16 – 6.18)$$

$$(\bar{x}^2) = \frac{1}{n}\sum_1^n x_i^2;$$

Довірчі межі результативної ознаки «y_i»:

$$y_{0i} = y_i \pm \Delta y_i;$$

$$y_i = a + bx_i; \qquad (6.19 – 6.21)$$

$$\Delta y_i = \pm t\sigma_e \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_1^n (x_i - \bar{x})^2}}.$$

Завдання для самоконтролю

1. Як розраховується і який зміст має кореляційне відношення?
2. Що таке випадок (артефакт)?
3. Як розраховується критерій Грабса?
4. Як розраховується і який зміст має критерій Пірсона?
5. У чому сутність методу аналітичних групувань?
6. У чому сутність кореляційно-регресійного аналізу?
7. У чому сутність дисперсійного аналізу?
8. Що таке теоретична лінія регресії?
9. Які форми регресійних рівнянь використовують у статистиці?
10. Як знаходяться параметри регресії?
11. Дайте визначення коефіцієнта еластичності.
12. Що таке залишкова дисперсія?
13. Як визначити довірчі межі коефіцієнтів регресії?
14. Як знаходиться і який зміст має кореляційне відношення?
15. У чому полягає правило складання дисперсій?
16. Що оцінюється за допомогою шкали Чедока?
17. Як оцінюється щільність зв’язку за допомогою кореляційного відношення?
18. Як оцінюється істотність зв’язку за допомогою критерію Фішера?

Задачі

6.1. Урожайність цукрових буряків за природнокліматичними зонами становить :

Таблиця 6.1

Зона зволоження	**Розмір збиральної площі, тис. га**	**Урожайність цукрових буряків, ц/га**
Недостатнього	21	200
Нестійкого	74	220
Достатнього	55	246
У цілому	150	226,7

Загальна дисперсія врожайності становить 323. Визначте міжгрупову дисперсію та кореляційне відношення, поясніть його зміст. Перевірте істотність зв’язку з імовірністю 0,95.

Розв’язок:

$$\delta^2 = \frac{\sum_1^m (\bar{y}_i - \bar{y})^2 \, \mathrm{f_i}}{\sum_1^m \mathrm{f_i}} = \frac{(200-226{,}7)^2 \cdot 21 + \ldots + (246-226{,}7)^2 \cdot 55}{150} = 258{,}5.$$

$\eta^2 = \dfrac{\delta^2}{\sigma^2} = \dfrac{258{,}5}{323} = 0{,}8$ – щільність зв’язку є досить суттєвою η^2 [0; 1]; варіація урожайності буряків на 80 % пояснюється варіацією основної ознаки – зони зволоження.

$\eta^2_{0,95}(2,\ 148) \approx \eta^2_{0,95}(2,\ 120) = 0{,}049.$

$\eta^2 > \eta^2_{0,95}(2,\ 148)$ – зв’язок суттєвий.

6.2. У результаті обстеження продуктивності 62 верстатів з різним строком експлуатації встановлено:

Таблиця 6.2

Строк експлуатації верстатів, роки	**Кількість верстатів**	**Виробництво деталей за зміну в розрахунку на один верстат, шт.**
До 7	10	110
7-14	15	96
14-20	25	70
20 і більше	12	66
У цілому	62	82

Загальна дисперсія продуктивності верстатів становить 470. Визначте міжгрупову дисперсію та кореляційне відношення, поясніть його зміст. Перевірте істотність зв'язку з імовірністю 0,95.

Розв'язок:

$$\delta^2 = \frac{\sum_1^m (\bar{y}_i - \bar{y})^2 \mathrm{f_i}}{\sum_1^m \mathrm{f_i}} = \frac{(110-82)^2 \cdot 10 + \ldots + (66-82)^2 \cdot 12}{62} = 281{,}48.$$

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{281{,}48}{470} = 0{,}6;$$

$$\eta^2_{0,95}(3,\ 59) \approx \eta^2_{0,95}(3,\ 60) = 0{,}121.$$

$$\eta^2 > \eta^2_{0,95}(3,\ 59) \text{ – зв'язок суттєвий.}$$

Щільність зв'язку є середньою η^2 [0; 1]; варіація у на 60 % пояснюється варіацією основної ознаки. Загальна дисперсія дорівнює 470, міжгрупова – 281. Перша дисперсія вказує на варіацію результуючої ознаки *y* під впливом усіх можливих факторів, на другу ж впливає виключно варіація *x*. Так як значення η^2 більше табличного, то зв'язок є суттєвим.

6.3. За результатами перевірки якості 20 партій твердих сирів виявлено залежність якості сиру від строку зберігання.

Таблиця 6.3

Строки зберігання, міс.	**Кількість партій**	**Зниження якості, бали**	**Групова дисперсія зниження якості**
До 2	7	1,3	0,08
2-4	8	2,8	0,13
4 і більше	5	4,1	0,20
Разом	20	2,6	Х

Визначте міжгрупову, середню з групових, загальну дисперсії зниження якості сиру. Розрахуйте кореляційне відношення, поясніть його зміст. Перевірте істотність зв'язку з імовірністю 0,95.

Розв'язок:

$$\delta^2 = \frac{\sum_1^m (\bar{y}_i - \bar{y})^2 \mathrm{f_i}}{\sum_1^m \mathrm{f_i}} = \frac{(1{,}3-2{,}6)^2 \cdot 7 + \ldots + (4{,}1-2{,}6)^2 \cdot 5}{20} = 1{,}17.$$

$$\bar{\sigma}^2 = \frac{\sum_1^m \sigma_i^2 f_i}{\sum_1^m f_i} = \frac{0{,}08 \cdot 7 + \ldots + 0{,}20 \cdot 5}{20} = 0{,}13.$$

$$\sigma^2 = \delta^2 + \bar{\sigma}^2 = 1{,}3.$$

$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{1{,}17}{1{,}3} = 0{,}9$ (щільність зв'язку є високою, η^2 [0; 1]; варіація y на 90 % пояснюється варіацією основної ознаки x, й лише на 10 % варіацією сторонніх факторів).

$\eta^2_{0,95}(2,\ 18) = 0{,}283.$

$\eta^2 > \eta^2_{0,95}(2,\ 18)$ – зв'язок суттєвий.

6.4. Заготівля овочевої сировини консервним комбінатом проводиться в радіусі до 200 км. Відстань перевезень впливає на якість заготовленої сировини таким чином:

Таблиця 6.4

№ перевезення	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Радіус перевезень, км	110	42	157	132	126	65	102	148	174	86
Частка нестандартної сировини, %	23	14	26	22	21	17	20	25	28	18

Виходячи з цих даних:

а) опишіть залежність якості овочевої сировини від дальності перевезення лінійною функцією регресії, визначте параметри регресії, поясніть їх зміст;

б) за допомогою лінійного коефіцієнта кореляції оцініть щільність зв'язку між факторною та результуючою ознаками;

в) перевірте істотність зв'язку між факторною та результуючою ознаками з імовірністю 0,95.

Розв'язок:

Так як якість заготівельної сировини залежить від відстані перевезень, то позначимо:

x_j – це факторна ознака, тобто радіус перевезень;

y_j – це результуюча ознака, тобто частка нестандартної сировини.

Для обчислення параметрів лінійної регресії заповнимо допоміжну таблицю.

Таблиця

№ перевезення *n*	**Радіус перевезень, км, *x***	**Частка нестандарт. сировини, %, *y***	***xy***	***x*²**	***y*²**
1	110	23	2530	1210	529
2	42	14	588	1764	196
3	157	26	4082	24649	676
4	132	22	2904	17424	484
5	126	21	2646	15876	441
6	65	17	1105	4225	289
7	102	20	2040	10404	400
8	148	25	3700	21904	625
9	174	28	4872	30276	784
10	86	18	1548	7396	324
Разом	1142	214	26015	146018	4748

Лінійне рівняння: $Y = a + bx$. Визначимо параметри регресії a та b на основі системи нормальних рівнянь:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x};$$

$$a = \bar{y} - b\bar{x}.$$

Обчислимо їх:

$$b = \frac{10 \cdot 26015 - 1142 \cdot 214}{10 \cdot 146018 - 1142 \cdot 1142} = \frac{15762}{156016} = 0{,}101.$$

Розрахуємо середні показники:

$$\bar{y} = \frac{\sum y}{n} = \frac{214}{10} = 21{,}4; \quad \bar{x} = \frac{\sum x}{n} = \frac{1142}{10} = 114{,}2.$$

Отже, $a = \bar{y} - b\bar{x}$ = 21,4 – 0,101 · 114,2 = 21,4 – 11,53 = 9,87.

Лінійне рівняння регресії набуває вигляду: $Y = 9{,}87 + 0{,}101x$.

Проаналізуємо зміст цих параметрів.

Параметр a особливого значення не має, тільки якщо $x = 0$ цей параметр співпадатиме з Y і кількісно виражає цю ознаку.

Параметр b показує, що при збільшенні радіусу на 1 км збільшується відсоток нестандартної сировини на 0,101 %.

Оцінимо щільність зв'язку між ознаками:

Лінійний коефіцієнт кореляції (Пірсона) $r = \dfrac{\sum_{1}^{n} xy - n\bar{x} \cdot \bar{y}}{n\sqrt{\sigma_x^2 \cdot \sigma_y^2}}$.

Дисперсія результуючої ознаки:

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 = \frac{\sum y^2}{n} - \bar{y}^2 = \frac{4748}{10} - (21{,}4)^2 = 16{,}84.$$

Дисперсія факторної ознаки:

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{146018}{10} - (114{,}2)^2 = 1560{,}16.$$

$$r = \frac{26015 - 10 \cdot 114{,}2 \cdot 21{,}4}{10\sqrt{16{,}84 \cdot 1560{,}16}} = \frac{1576{,}2}{1620{,}89} = 0{,}972.$$

Отже, зв'язок між радіусом перевезень та якістю сировини дуже щільний (залежність – 97,2 % і тільки 2,8 % впливу мають решта факторів) та прямий (пряма залежність між ознаками x та y).

Перевіримо цей зв'язок на істотність.

Для цього потрібно перейти до коефіцієнту детермінації (R^2):

$|r| = R;$

$r = R;$

$r^2 = R^2 = (0{,}972)^2 = 0{,}944.$

Порівняємо R^2 та $R^2_{1-\alpha(K_1,K_2)}$ $K_1 = m - 1.$

$K_2 = n - m,$ де m – кількість параметрів.

$K_1 = 2 - 1 = 1.$

$K_2 = 10 - 2 = 8.$

$R^2_{0{,}95(1{,}8)} = 0{,}399.$

Отже, критичне значення (0,399) значно менше фактичного (0,944), тому можна стверджувати, що зв'язок між радіусом перевезень та якістю сировини – істотний з імовірністю 0,95.

6.5. За даними аудиторського звіту про діяльність 12 комерційних банків, встановлено залежність між розміром кредитної ставки та доходністю кредитних операції:

Таблиця 6.5

№ банка ***n***	**Кредитна ставка, %** ***x***	**Дохідність операцій, %,** ***y***
1	59	18
2	61	24
3	64	35
4	66	31
5	68	29
6	61	25

Закінчення табл. 6.5

7	64	36
8	64	32
9	66	30
10	67	31
11	66	30
12	62	28
Разом	768	349

Виходячи з цих даних:

а) визначте функцію, яка описує залежність між розміром кредитної ставки та доходністю від кредитних операцій, обчисліть параметри рівняння, оцініть х зміст;

б) за допомогою коефіцієнта детермінації оцініть щільність зв'язку між ознаками;

в) перевірте істотність зв'язку з імовірністю 0,95.

Розв'язок:

Для обчислення параметрів лінійної регресії заповнимо допоміжну таблицю.

Таблиця

№ банка ***n***	**Кредитна ставка, %** ***x***	**Дохідність операцій, %,** ***y***	***xy***	***x²***	***y²***
1	59	18	1062	3481	324
2	61	24	1464	3721	576
3	64	35	2240	4096	1225
4	66	31	2046	4356	961
5	68	29	1972	4624	841
6	61	25	1525	3721	625
7	64	36	2304	4096	1296
8	64	32	2048	4096	1024
9	66	30	1980	4356	900
10	67	31	2077	4489	961
11	66	30	1980	4356	900
12	62	28	1737	3844	784
Разом	768	349	22434	49236	10417

Визначимо залежність між ознаками x та y за допомогою лінійної функції.

Лінійне рівняння: $Y = a + bx$. Визначимо параметри a та b на основі системи нормальних рівнянь:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x};$$

$$a = \bar{y} - b\bar{x}.$$

Обчислимо їх:

$$b = \frac{12 \cdot 22434 - 768 \cdot 349}{12 \cdot 49236 - 768 \cdot 768} = 1{,}16.$$

Розрахуємо середні показники:

$$\bar{y} = \frac{\sum y}{n} = \frac{349}{12} = 29{,}08; \quad \bar{x} = \frac{\sum x}{n} = \frac{768}{12} = 64.$$

Отже, $a = \bar{y} - b\bar{x}$ = 29,08 – 1,16·64 = – 45,16.

Лінійне рівняння набуває вигляду: Y= – 45,16 + 1,16x.

Зміст параметра a – тільки кількісний.

Зміст параметра b: зі збільшенням кредитної ставки на 1 % збільшує дохідність від кредитних операцій на 1,16 %.

Оцінимо щільність зв’язку за допомогою коефіцієнта детермінації (R^2):

$$R^2 = \frac{\delta_y^2}{\sigma_y^2},$$

де δ_y^2 – факторна дисперсія; σ_y^2 – загальна дисперсія.

Розрахуємо їх:

$$\delta_y^2 = \frac{1}{n}\left(a\sum y + b\sum xy\right) - \bar{y}^2 = \frac{1}{12}\left(-45{,}16 \cdot 349 + 1{,}16 \cdot 22434\right) - \left(29{,}08\right)^2 = 9{,}57.$$

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 = \frac{\sum y^2}{n} - \bar{y}^2 = \frac{10417}{12} - \left(29{,}08\right)^2 = 22{,}44.$$

Отже, коефіцієнт детермінації:

$$R^2 = \frac{9{,}57}{22{,}44} = 0{,}426.$$

Дохідність від операцій на 42,6 % залежить від розміру кредитної ставки і на 57,4 % – від решти факторів.

Перевіримо цей зв’язок на істотність:

Порівняємо фактичне значення (R^2) та критичне $\left(R^2_{1-\alpha(K_1,K_2)}\right)$:

$$K_1 = m - 1 = 2 - 1 = 1;$$

$$K_2 = n - m = 12 - 2 = 10;$$

$$R^2_{0,95(1,8)} = 0{,}332.$$

Фактичне значення (0,426) більше критичного (0,332), а тому зв’язок є істотним з імовірністю 0,95.

6.6. За даними спостереження окупність витрат на радіоприлади залежить від строку освоєння їх виробництва.

Таблиця 6.6

№ продукції, ***n***	**Строк освоєння, роки** ***x***	**Окупність витрат, тис гр. од.,** ***y***
1	5	10,2
2	4	7,5
3	7	13,9
4	10	12,8
5	1	0,6
6	2	2,8
7	8	13,2
8	12	10,1
9	3	5,4
10	6	12,7
Разом	58	89,2

Виходячи з цих даних:

а) визначте функцію, яка описує залежність між окупністю витрат та строком освоєння виробництва, обчисліть параметри рівняння, оцініть їх зміст;

б) за допомогою коефіцієнта детермінації оцініть щільність зв'язку між ознаками;

в) перевірте істотність зв'язку з імовірністю 0,95.

Розв'язок:

Побудуємо допоміжну таблицю.

Таблиця

№ продукції, ***n***	**Строк освоєння, роки** ***x***	**Окупність витрат, тис гр. од.,** ***y***	xy	x^2	y^2
1	5	10,2	51	25	104,04
2	4	7,5	30	16	56,25
3	7	13,9	97,3	49	193,21
4	10	12,8	128	100	163,84
5	1	0,6	0,6	1	0,36
6	2	2,8	5,6	4	7,84
7	8	13,2	105,6	64	174,24
8	12	10,1	121,2	144	102,01
9	3	5,4	16,2	9	29,16
10	6	12,7	76,2	36	161,29
Разом	58	89,2	631,7	448	992,24

Визначимо залежність між ознаками *x* та *y* за допомогою лінійної функції.

Лінійне рівняння: $Y = a + bx$. Визначимо параметри *a* та *b* на основі системи нормальних рівнянь:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x};$$

$$a = \bar{y} - b\bar{x}$$

Обчислимо їх:

$$b = \frac{10 \cdot 631{,}7 - 58 \cdot 89{,}2}{10 \cdot 448 - (58)^2} = 1{,}024.$$

Розрахуємо середні показники:

$$\bar{y} = \frac{\sum y}{n} = 8{,}92; \qquad \bar{x} = \frac{\sum x}{n} = 5{,}8.$$

Отже, $a = \bar{y} - b\bar{x}$ = 8,92 – 1,024·5,8 = 2,98.

Лінійне рівняння набуває вигляду: $Y = 2{,}98 + 1{,}042x$.

Зміст параметра *a* – тільки кількісний.

Зміст параметра *b:* зі збільшенням строку освоєння на 1 рік, збільшується окупність витрат на 1,024 тис. гр. од.

Оцінимо щільність зв’язку за допомогою коефіцієнта детермінації (R^2):

$$R^2 = \frac{\delta_y^2}{\sigma_y^2},$$

де δ_y^2 – факторна дисперсія; σ_y^2 – загальна дисперсія.

Так як коефіцієнт показує співвідношення дисперсії факторної і результуючої ознак, то він також визначає щільність зв’язку між ознаками.

Розрахуємо їх:

$$\delta_y^2 = \frac{1}{n}\left(a\sum y + b\sum xy\right) - \bar{y}^2 = \frac{1}{10}\left(2{,}98 \cdot 89{,}2 + 1{,}024 \cdot 631{,}7\right) - (8{,}92)^2 = 11{,}7.$$

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 = \frac{\sum y^2}{n} - \bar{y}^2 = \frac{992{,}24}{10} - (8{,}92)^2 = 19{,}65.$$

Отже, коефіцієнт детермінації:

$$R^2 = \frac{11{,}7}{19{,}65} = 0{,}59.$$

Окупність витрат на 59 % залежить від строку освоєння, і на 41 % від решти факторів.

Перевіримо цей зв’язок на істотність.

Порівняємо фактичне значення (R^2) та критичне $(R^2_{1-\alpha(K_1,K_2)})$:

$K_1 = m-1 = 2-1 = 1;$

$K_2 = n-m = 10-2 = 8;$

$R^2_{0,95(1,8)} = 0,399.$

Фактичне значення (0,59) більше за критичне (0,399), отже, зв'язок між ознаками x та y з ймовірністю 0,95 є істотним.

6.7. За даними лабораторних досліджень вихід хліба з 1 кг борошна залежить від його вологості, про що свідчать наведені дані:

Таблиця 6.7

№ партії хліба, *n*	**Вологість борошна, % *x***	**Вихід хліба, кг, *y***
1	13,1	1,45
2	13,3	1,36
3	13,7	1,32
4	14,1	1,31
5	13,2	1,40
6	13,9	1,32
7	13,1	1,43
8	13,6	1,33
9	14,5	1,31
10	13,2	1,42
11	13,4	1,34
12	14,3	1,31
13	13,2	1,38
Разом	176,6	17,68

Виходячи з цих даних:

а) визначте функцію, яка описує залежність між виходом хліба та вологістю борошна, обчисліть параметри рівняння, оцініть їх зміст;

б) за допомогою коефіцієнта детермінації оцініть щільність зв'язку між ознаками;

в) перевірте істотність зв'язку з імовірністю 0,95.

Розв'язок:

Визначимо залежність між ознаками x та y за допомогою лінійної функції. Для обчислення параметрів лінійної регресії побудуємо допоміжну таблицю.

Таблиця

№ партії хліба, *n*	**Вологість борошна, %, *x***	**Вихід хліба, кг, *y***	***xy***	***x*²**	***y*²**
1	13,1	1,45	18,99	171,6	2,1
2	13,3	1,36	18,08	176,9	1,85
3	13,7	1,32	18,08	187,7	1,74
4	14,1	1,31	18,47	198,8	1,71
5	13,2	1,40	18,48	174,2	1,96
6	13,9	1,32	18,34	193,2	1,74
7	13,1	1,43	18,73	171,6	2,04
8	13,6	1,33	18,08	184,9	1,77
9	14,5	1,31	18,99	210,2	1,71
10	13,2	1,42	18,74	174,2	961
11	13,4	1,34	17,95	179,5	1,79
12	14,3	1,31	18,73	204,5	1,71
13	13,2	1,38	18,21	174,2	1,9
Разом	176,6	17,68	239,8	2401,5	24,03

Лінійне рівняння: $Y = a + bx$. Визначимо параметри a та b на основі системи нормальних рівнянь:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x};$$

$$a = \bar{y} - b\bar{x}.$$

Обчислимо їх:

$$b = \frac{13 \cdot 239{,}8 - 176{,}6 \cdot 17{,}68}{13 \cdot 2401{,}5 - (176{,}6)^2} = -0{,}15.$$

Розрахуємо середні показники:

$$\bar{y} = \frac{\sum y}{n} = \frac{17{,}68}{13} = 1{,}36; \qquad \bar{x} = \frac{\sum x}{n} = \frac{176{,}6}{13} = 13{,}58.$$

Отже, $a = \bar{y} - b\bar{x}$ =1,36 – (– 0,15)·13,58 = 3,39.

Лінійне рівняння набуває вигляду: Y = 3,39 – 0,15x.

Зміст параметра a – тільки кількісний.

Зміст параметра b: зі збільшенням вологості борошна на 1 % зменшується вихід хліба на 0,15 кг.

Оцінимо щільність зв’язку за допомогою коефіцієнта детермінації (R^2):

$$R^2 = \frac{\delta_y^2}{\sigma_y^2},$$

де δ_y^2 – факторна дисперсія; σ_y^2 – загальна дисперсія.

Так як коефіцієнт показує співвідношення дисперсії факторної і результуючої ознак, то він також визначає щільність зв'язку між ознаками.

Розрахуємо їх:

$$\delta_y^2 = \frac{1}{n}\left(a\sum y + b\sum xy\right) - \bar{y}^2 = \frac{1}{13}\left(3{,}39 \cdot 17{,}68 + (-0{,}15)\cdot 239{,}8\right) - (1{,}36)^2 =$$
$$= -0{,}0006.$$

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 = \frac{\sum y^2}{n} - \bar{y}^2 = \frac{24{,}03}{13} - \left(1{,}36\right)^2 = -0{,}001.$$

Отже, коефіцієнт детермінації:

$$R^2 = \frac{0{,}0006}{0{,}001} = 0{,}6.$$

Вихід хліба на 60 % залежить від вологості борошна і на 40 % від решти факторів.

Перевіримо цей зв'язок на істотність.

Порівняємо фактичне значення (R^2) та критичне $(R^2_{1-\alpha(K_1,K_2)})$:

$K_1 = m - 1 = 2 - 1 = 1;$

$K_2 = n - m = 13 - 2 = 11;$

$R^2_{0{,}95(1{,}8)} = 0{,}307.$

Фактичне значення (0,6) більше за критичне (0,307), отже, зв'язок між ознаками x та y з імовірністю 0,95 є істотним.

6.8. Вибіркове обстеження 15 підприємств з питань впливу агротехнічних робіт на рівень урожайності кормових культур подано в табл. 6.8.

Виходячи з цих даних:

а) визначте функцію, яка описує залежність між урожайністю кормових культур та кількістю внесених мінеральних добрив, обчисліть параметри рівняння, оцініть їх зміст;

б) за допомогою коефіцієнта детермінації оцініть щільність зв'язку між ознаками;

в) перевірте істотність зв'язку з імовірністю 0,95.

Розв'язок:

Добудуємо таблицю первинних даних, здійснемо допоміжні розрахунки.

Таблиця 6.8

№ господарства, *n*	**Внесено мінеральних добрив, ц/га *x***	**Урожайність, ц/га, *y***	***xy***	***x²***	***y²***
1	1,1	25,6	28,16	1,21	655,3
2	1,6	33,3	53,28	2,56	1108,8
3	1,2	28,2	33,84	1,44	795,2
4	1,5	32,0	48	2,25	1024
5	2,1	37,1	77,91	4,41	1376,4
6	1,7	34,2	58,14	2,89	1169,6
7	1,3	29,7	38,61	1,69	882
8	2,2	36,6	80,52	4,84	1339,5
9	2,0	36,9	73,8	4	1361,6
10	1,7	35,1	59,67	2,89	1232
11	1,2	27,0	32,4	1,44	729
12	1,8	35,9	64,62	3,24	1288,8
13	2,2	36,2	79,64	4,84	1310,4
14	1,7	34,7	58,99	2,89	1204
15	1,7	32,5	55,25	2,89	1056,2
Разом	25	495	842,8	43,48	16532,8

Визначимо залежність між ознаками x та y за допомогою лінійної функції. Лінійне рівняння: $Y = a + bx$. Визначимо параметри a та b на основі системи нормальних рівнянь:

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x};$$

$$a = \bar{y} - b\bar{x}.$$

Обчислимо їх:

$$b = \frac{15 \cdot 842{,}8 - 25 \cdot 495}{15 \cdot 43{,}48 - (25)^2} = 9{,}82.$$

Розрахуємо середні показники:

$$\bar{y} = \frac{\sum y}{n} = \frac{495}{15} = 33; \qquad \bar{x} = \frac{\sum x}{n} = \frac{25}{15} = 1{,}67.$$

Отже, $a = \bar{y} - b\bar{x}$ = 33 – 9,82·1,67 = 16,4.

Лінійне рівняння набуває вигляду: Y = 16,4 + 9,82·x.

Зміст параметра a – тільки кількісний.

Зміст параметра b: зі збільшенням мінеральних добрив на 1 ц/га урожайність збільшиться на 9,8 ц/га.

Оцінимо щільність зв'язку за допомогою коефіцієнта детермінації (R^2):

$$R^2 = \frac{\delta_y^2}{\sigma_y^2},$$

де δ_y^2 – факторна дисперсія; σ_y^2 – загальна дисперсія.

Розрахуємо їх:

$$\delta_y^2 = \frac{1}{n}\left(a\sum y + b\sum xy\right) - \bar{y}^2 = \frac{1}{15}\left(16{,}73\cdot 495 + 9{,}8\cdot 842{,}8\right) - \left(33\right)^2 = 3{,}95.$$

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 = \frac{\sum y^2}{n} - \bar{y}^2 = \frac{16532{,}8}{15} - \left(33\right)^2 = 13{,}19.$$

Отже, коефіцієнт детермінації:

$$R^2 = \frac{3{,}95}{13{,}19} = 0{,}299.$$

Урожайність на 30 % залежить від внесених мінеральних добрив і на 70 % залежить від решти факторів. Перевіримо зв'язок на істотність:

Порівняємо фактичне значення (R^2) та критичне $(R^2_{1-\alpha(K_1,K_2)})$:

$K_1 = m - 1 = 2 - 1 = 1;$

$K_2 = n - m = 15 - 2 = 13;$

$R^2_{0,95(1,8)} = 0{,}256.$

Фактичне значення (0,299) більше за критичне (0,265), отже, зв'язок між ознаками *x* та *y* з імовірністю 0,95, – істотний.

6.9. За наведеними результатами тестування країн щодо ефективності економіки та ступеня політичного ризику визначте коефіцієнт рангової кореляції, перевірте його істотність з імовірністю 0,95, зробіть висновки.

Таблиця 6.9

	Ранг країни						
	A	**B**	**C**	**D**	**K**	**M**	**P**
Ефективність економіки	6	7	4	3	5	1	2
Ступінь політичного ризику	1	2	3	5	4	7	6

Розв'язок:

Збудуємо нову таблицю з врахуванням рангів показників (R_x, R_y), відхиленням рангів (d_j) та квадратами відхилення рангів (d_j^2).

Таблиця

Розрахунок рангової кореляції

Країна	R_x	R_y	d_j	d_j^2
A	6	1	5	25
B	7	2	5	25
C	4	3	1	1
D	3	5	– 2	4
K	5	4	1	1
M	1	7	– 6	36
P	2	6	–4	16
Разом			0	108

Відхилення рангів $d_j = R_x - R_y$, таким чином, наприклад, відхилення рангів для країни *A* буде $d_j = 6 - 5 = 1$. В такому порядку розраховується d_j і для інших країн.

Знайдемо коефіцієнт рангової кореляції Спірмена:

$$\rho = 1 - \frac{6\sum_{i=1}^{7} d_j^2}{n(n^2-1)},$$

де n – кількість рангів, у цьому випадку 7.

$$\rho = 1 - \frac{6 \cdot 108}{7(7^2-1)} = -0{,}92.$$

Дане значення коефіцієнта рангової кореляції свідчить про наявність оберненого і досить помітного зв’язку між зазначеними параметрами ефективності економіки та ступеня політичного ризику. За табл. 6.3 значень коефіцієнта рангової кореляції Спірмена критичне значення коефіцієнта при α = 0,05 та n = 7 становить $\rho_{0,95}(7) = 0{,}71$, що значно менше фактичного. Отже, істотність зв’язку доведена з імовірністю 0,95, що позначає обернено пропорційну залежність між економікою і політикою.

6.10. За допомогою коефіцієнта рангової кореляції визначте істотність розбіжностей в оцінках пріоритетів хлопцями та дівчатами старших класів однієї школи:

Таблиця 6.10

Пріоритет	**В % до кількості опитаних**	
	хлопці	**дівчата**
Друзі	97	96
Любов	86	95
Гроші	90	72
Мода	54	67
Навчання	79	65

Розв'язок:

Збудуємо таблицю за наведеними даними та знайдемо ранги та відхилення рангів за формулою, яка зазначена в задачі 6.9:

Таблиця

Розрахунок рангової кореляції

Пріоритет, *n*	R_x	R_y	d_i	d_i^2
1	1	1	0	0
2	3	2	1	1
3	2	3	-1	1
4	5	4	1	1
5	4	5	-1	1
Разом			0	4

Знайдемо коефіцієнт рангової кореляції та оцінимо розбіжність в оцінках пріоритетів.

$$\rho = 1 - \frac{6\sum_{i=1}^{5} d_j^2}{n(n^2-1)};$$

$$\rho = 1 - \frac{6 \cdot 4}{5(5^2-1)} = 0{,}8.$$

Значення коефіцієнта менші за критичне значення $\rho_{0,95}(5) = 0{,}9$. Тому існує неістотний зв'язок між наведеними пріоритетами (імовірність 0,95).

У вузькому розумінні життєві погляди хлопців та дівчат значно різняться.

6.11. За допомогою коефіцієнта рангової кореляції оцініть ступінь узгодженості оцінок двох груп експертів на конкурсі професійної майстерності модельєрів. Висновок зробіть з імовірністю 0,95.

Таблиця 6.11

Модельєри	**Ранг, визначений експертами**	
	Художниками, R_x	**Промисловиками, R_y**
A	5	4
B	1	3
C	6	5
D	3	2
K	2	1
M	4	6
P	7	7

Розв'язок:

Таблиця

Розрахунок відхилення рангів

Модельєри	d_i	d_i^2
A	1	1
B	–2	4
C	1	1
D	1	1
K	1	1
M	–2	4
P	0	0
Разом	0	12

За формулою Спірмена:

$$\rho = 1 - \frac{6\sum_{1}^{n} d_j^2}{n(n^2-1)} = 1 - \frac{6\cdot 12}{7\cdot 48} = 0{,}78.$$

Зв'язок є досить сильним, прямим.

Табличне значення $\rho_{0,95}(7) = 0{,}71$, що менше значення фактичного показника. Тобто суттєвість зв'язку доведена з імовірністю 0,95.

6.12. Розподіл урожайності зернових по домогосподарствах характеризуються даними:

Таблиця 6.12

Урожай, ц/га	**40-42**	**42-44**	**44-46**	**46-48**	**48-50**	**50-52**	**52-54**	**Загалом**
Кількість домогосподарств	4	7	28	35	16	6	4	100

Перевірте розподіл врожайності зернових на відповідність нормальному закону розподілу. Висновок зробіть з надійністю 95%.

Розв'язок:

Допоміжні розрахунки зведемо в таблицю.

Таблиця

x_i	f_i	x'	$x'-\bar{x}$	$t=\frac{x'-\bar{x}}{\sigma}$	$f(t)$	f_i'	$f_i - f_i'$	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
40-42	4	41	5,72	2,22	0,0339	2	2	4	2,0
42-44	7	43	3,72	1,44	0,1415	11	–4	16	1,45
44-46	28	45	1,72	0,67	0,3187	25	3	9	0,36
46-48	35	47	0,28	0,11	0,3965	31	4	16	0,52
48-50	16	49	2,28	0,88	0,2079	21	–5	25	1,19

Закінчення табл.

50-52	6	51	4,28	1,6	0,1006	8	–2	4	0,5
52-54	4	53	6,28	2,43	0,0208	2	2	4	2,0
Разом	100	–	–	–		100	–	–	8,02

$\bar{x} = 46{,}72;$

$\sigma^2 = 6{,}6414;$

$\sigma = 2{,}58;$

$$(\sum_1^m f_i)\frac{\Delta h}{\sigma} = 100 \cdot \frac{(42-40)}{2{,}28} = 77{,}519;$$

$m = 7;$

$r = 3;$

$k = 7 - 3 = 4;$

$$\chi^2 = \sum_1^m \frac{(f_i - f_i')^2}{f_i'}, \ \chi^2 = 8{,}02;$$

$$\chi^2_{табл0,95}(4) = 9{,}49 > (\chi^2 = 8{,}02).$$

Висновок – так, розподіл з ймовірністю 0,95 – нормальний.