Львівський державний університет безпеки життєдіяльності

М.Ф. Стасюк
О.О. Карабин
М.І. Кусій

Статистичний аналіз

Львів 2015

Стасюк М.Ф., Карабин О.О., Кусій М.І.

Статистичний аналіз

Теоретичні відомості та приклади розв’язування задач

Для курсантів, студентів, ад’юнктів та слухачів заочного відділення

Рецензенти: Єлейко В.І., доктор економічних наук, професор, завідувач кафедри менеджменту Львівської комерційної академії;
Кухарська Н.П., канд. фіз.-мат. наук, доцент, доцент кафедри управління інформаційною безпекою Львівського державного університету безпеки життєдіяльності.

1. Основні поняття математичної статистики

Математичною статистикою називається наука, яка займається розробкою методів відбору, опису та аналізу статистичних даних з метою вивчення закономірностей масових випадкових явищ.

Вихідними поняттями математичної статистики є поняття **генеральної сукупності та вибірки.**

Нехай потрібно дослідити яку-небудь ознаку, характерну для великої групи однотипних об'єктів.

Вибіркою називається сукупність випадково взятих об'єктів при дослідженні деякої ознаки.

Генеральною сукупністю називається сукупність об'єктів, з яких зроблено вибірку.

Вибірковий метод дослідження випадкових явищ полягає в тому, що з генеральної сукупності береться вибірка відносно невеликого обсягу, обчислюються характеристики цієї вибірки, які приймаються за наближені значення відповідних характеристик генеральної сукупності.

Нехай з генеральної сукупності взята вибірка, яка набула n_1 разів значення x_1, n_2 разів значення $x_2, ..., n_m$ разів значення x_m.

Тоді значення $x_{1,} x_2, ..., x_m$ називаються **варіантами** вибірки, а значення $n_1, n_2, ..., n_m$ **частотами** цих варіант.

Варіанти, що записані в зростаючому порядку, називають **варіаційним рядом.**

Сума всіх частот

$$\sum_{i=1}^{m} n_i = n \tag{1.1}$$

називається **об'ємом вибірки.**

Відношення

$$w_i = \frac{n_i}{n}, \quad i = 1, 2, ..., m \tag{1.2}$$

називається **відносною частотою** варіанти x_k. Очевидно, що

$$\sum_{i=1}^{m} w_i = 1. \tag{1.3}$$

Розмахом вибірки ***r*** називають різницю між найбільшим та найменшим її значеннями:

$$r = x_m - x_1 \tag{1.4}$$

Дискретним статистичним розподілом вибірки називається перелік варіант та відповідних їм частот

x_i	x_1	x_2	$\cdots$	x_m
n_i	n_1	n_2	$\cdots$	n_m

(1.5)

або відносних частот, який подається у вигляді таблиць:

x_i	x_1	x_2	$\cdots$	x_m
w_i	w_1	w_2	$\cdots$	w_m

(1.6)

де $x_1 < x_2 < ... < x_m, \sum_{i=1}^{m} n_i = n,$ а w_k визначені формулами (1.3).

Полігоном частот (відносних частот) статистичного розподілу (1.5) (1.6) називають **ламану,** відрізки якої з'єднують точки $(x_1; n_1)$, $(x_2; n_2)$, …, $(x_m; n_m)$ ($(x_1; w_1)$, $(x_2; w_2)$, …, $(x_m; w_m)$).

Коли ознака X генеральної сукупності має неперервний розподіл, або, коли ознака X має дискретний розподіл і обсяг вибірки – великий, тоді статистичний матеріал, який отримуємо при вивченні випадкової ознаки генеральної сукупності доцільно подати у вигляді

інтервального статистичного розподілу. Для цього потрібно весь інтервал зміни варіант $[x_1; x_m]$ поділити на скінчену кількість інтервалів без спільних точок та підрахувати кількість n_k тих значень варіант x_k, що потрапили в заданий інтервал (або кількість їх відносних частот w_k).

Інтервальний статистичний розподіл подається у вигляді таблиць:

$[z_{i-1}, z_i)$	$[z_0, z_1)$	$[z_1, z_2)$	$\cdots$	$[z_{m-1}, z_m]$
n_i	n_1	n_2	$\cdots$	n_m

(1.7)

або

$[z_{i-1}, z_i)$	$[z_0, z_1)$	$[z_1, z_2)$	$\cdots$	$[z_{m-1}, z_m]$
w_i	w_1	w_2	$\cdots$	w_m

(1.8)

Частинні інтервали можуть мати як різні, так і однакові довжини.

Гістограмою частот (відносних частот) інтервального розподілу (1.7) (1.8) називається східчаста фігура, яка складається з прямокутників, основами яких є частинні інтервали $[z_{i-1}, z_i], i = 1, 2, ..., m,$ а висоти дорівнюють

$$h_i = \frac{n_i}{z_{i-1} - z_i} \qquad \left(\tilde{h}_i = \frac{w_i}{z_{i-1} - z_i} \right).$$

Зауважимо, що площа кожного з прямокутників дорівнює n_k для інтервального розподілу (1.7) і w_k для розподілу (1.8). Площа гістограми частот дорівнює об’єму вибірки n, а площа гістограми відносних частот дорівнює одиниці.

Емпіричною функцією розподілу вибірки (1.5) називається функція

$$F^*(x) = \frac{n_x}{n}, \qquad (1.9)$$

де n_x кількість варіант вибірки, які менші ніж x, n об'єм вибірки.

Для дискретного статистичного розподілу (1.5) чи (1.6) емпірична функція $F^*(x)$ може бути задана конструктивно, а саме:

$$F^*(x) = \begin{cases} 0, x \in (-\infty; x_1]; \\ \frac{n_1}{n}, x \in (x_1; x_2]; \\ \frac{n_1 + n_2}{n}, x \in (x_2; x_3]; \\ \dots\dots\dots\dots\dots\dots\dots \\ \frac{1}{n}\sum_{k=1}^{m-1} n_k, x \in (x_{m-1}; x_m]; \\ 1, x \in (x_m; \infty). \end{cases} \qquad (1.10)$$

Емпірична функція $F^*(x)$, як і теоретична функція розподілу, має такі властивості:

1. $0 \le F^*(x) \le 1$;
2. $F^*(x)$ – неспадна;
3. $F^*(x)$=0, якщо $x \le x_1$ і $F^*(x)$=1, якщо $x > x_m$, де x_1 найменша варіанта, x_m найбільша.

Вибірковим середнім $\bar{x}$ статистичного розподілу (1.5) називається середнє арифметичне варіант x_k з урахуванням їх частот:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{1}{n}\left(x_1 n_1 + x_2 n_2 + \dots + x_m n_m\right). \qquad (1.11)$$

Вибірковою дисперсією $\overline{D}$ статистичного розподілу (1.5) називається середнє значення квадратів відхилень його членів від середнього значення $\bar{x}$:

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{m}\left(x_i - \bar{x}\right)^2 n_i. \qquad (1.12)$$

Для обчислення вибіркової дисперсії зручніше використовувати таку формулу:

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{m} x_i^2 \cdot n_i - \left(\bar{x}\right)^2. \qquad (1.13)$$

Вибіркове середнє квадратичне відхилення $\bar{\sigma}$ визначається рівністю:

$$\bar{\sigma} = \sqrt{\bar{D}}. \qquad (1.14)$$

Вибірковим центральним емпіричним моментом s**-го порядку** $\bar{\mu}_s$ називається величина, яка визначається рівністю:

$$\bar{\mu}_s = \frac{1}{n}\sum_{i=1}^{m}\left(x_i - \bar{x}\right)^s n_i. \qquad (1.15)$$

Зокрема $\bar{\mu}_0 = 1, \quad \bar{\mu}_1 = 0, \quad \bar{\mu}_2 = \bar{D}.$

Для оцінки відхилення статистичного розподілу вибірки від нормального розподілу використовують такі чисельні характеристики, як **асиметрію** і **ексцес.**

Вибірковою асиметрією називають число, яке обчислюється за формулою:

$$\bar{A} = \frac{\bar{\mu}_3}{\bar{\sigma}^3}, \qquad (1.16)$$

де $\bar{\mu}_3 -$ вибірковий центральний момент 3-го порядку, $\bar{\sigma} -$ середнє квадратичне відхилення статистичного розподілу вибірки (1.5) чи (1.6).

Вибірковим ексцесом $\bar{E}$ статистичного розподілу вибірки (1.5) чи (1.6) називається число, яке обчислюється за формулою:

$$\bar{E} = \frac{\bar{\mu}_4}{\bar{\sigma}^3} - 3, \qquad (1.17)$$

де $\bar{\mu}_4$ – вибірковий центральний емпіричний момент 4-го порядку, а $\bar{\sigma}$ – середнє квадратичне відхилення статистичного розподілу вибірки (1.5) чи (1.6).

Якщо випадкова величина *X*, що описує деяку ознаку генеральної сукупності, розподілена за нормальним законом, то її асиметрія та ексцес дорівнюють нулю. Тому, чим більше асиметрія і ексцес статистичного розподілу вибірки віддалені від нуля, тим менше підстав сподіватися, що вибірка, з якої утворено варіаційний ряд, утворена з генеральної сукупності, ознака X якої розподілена за нормальним законом.

Зауважимо, що для знаходження чисельних характеристик інтервального розподілу вибірки (1.7) чи (1.8) його слід замінити на дискретний. Для цього досить частинні проміжки $\left[z_{k-1,} z_k\right)$ замінити числами серединами цих проміжків, тобто числами $x_k = \dfrac{z_{k-1} + z_k}{2}$, а відповідні їм значення частот чи відносних частот залишити без змін.

Приклад 1. Для вивчення прибутків регіону утворено вибірку, яка характеризується такими даними: 8, 7, 6, 9, 10, 9, 11, 8, 9, 10, 8, 9, 6, 9, 8, 10, 12, 7, 10, 7. Виконати такі завдання:

1. Записати дискретний статистичний розподіл вибірки, побудувати полігон частот і відносних частот та емпіричну функцію розподілу.
2. Обчислити чисельні характеристики вибірки: вибіркове середнє, вибіркову дисперсію, середнє квадратичне відхилення, розмах вибірки.
3. Записати інтервальний статистичний розподіл вибірки частот та відносних частот, поділивши проміжок на 4 рівні частини, і побудувати гістограми частот та відносних частот.

Розв'язання. В цьому прикладі генеральною сукупністю є випадкова величина X, яка описує відсоткове відношення прибутку до обсягу виробництва одного підприємства. Обсяг вибірки $n = 20$.

1. Записуємо варіаційний ряд варіант та відповідний ряд частот варіант і записуємо дискретний статистичний розподіл вибірки у формі таблиці:

x_k	6	7	8	9	10	11	12
n_k	2	3	4	5	4	1	1
$w_k = \frac{n_k}{n}$	$^2/_{20}$	$^3/_{20}$	$^4/_{20}$	$^5/_{20}$	$^4/_{20}$	$^1/_{20}$	$^1/_{20}$

Будуємо полігон частот, як ламану, відрізки якої з'єднують точки $(x_k; n_k)$, і полігон відносних частот ламану, відрізки якої з'єднують точки $(x_k; w_k)$ на координатній площині.

Емпірична функція розподілу має такий вигляд:

$$F^*(x) = \begin{cases} 0, & x \le 6; \\ ^2/_{20}, & 6 < x \le 7; \\ ^5/_{20}, & 7 < x \le 8; \\ ^9/_{20}, & 8 < x \le 9; \\ ^{14}/_{20}; & 9 < x \le 10; \\ ^{18}/_{20}, & 10 < x \le 11; \\ ^{19}/_{20}, & 11 < x \le 12; \\ 1, & x > 12. \end{cases}$$

Зображаємо емпіричну функцію графічно:

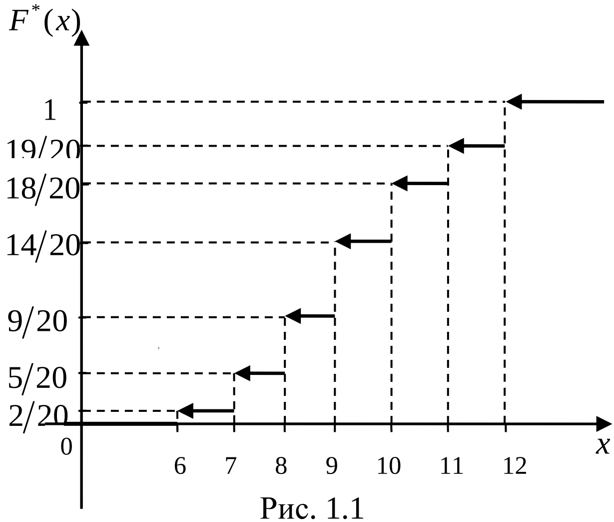


Рис. 1.1

2. Обчислимо чисельні характеристики вибірки:

а) вибіркове середнє обчислюємо за формулою (1.11):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{1}{20}\left(2\cdot 6+3\cdot 7+4\cdot 8+5\cdot 9+4\cdot 10+1\cdot 11+1\cdot 12\right)=$$

$$=\frac{1}{20}\cdot 173 = 8,65;$$

б) вибіркову дисперсію обчислюємо за формулою (1.12):

$$\overline{D} = \frac{1}{n}\sum_{k=1}^{m} n_i\left(x_i - \overline{x}\right)^2 = \frac{1}{20}\Big[\ 2\left(-2,65\right)^2 + 3\left(-1,65\right)^2 + 4\left(-0,65\right)^2 +$$

$$+5\cdot 0,35^2 + 4\cdot 1,35^2 + 1\cdot 2,35^2 + 1\cdot 3,35^2\ \Big] = \frac{1}{20}\cdot 48,55 = 2,4275,$$

або за формулою (1.13):

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \overline{x}^2 =$$

$$= \frac{1}{20}\left(2 \cdot 6^2 + 3 \cdot 7^2 + 4 \cdot 8^2 + 5 \cdot 9^2 + 4 \cdot 10^2 + 1 \cdot 11^2 + 1 \cdot 12^2\right) - 8{,}65^2 =$$

$$= \frac{1}{20} \cdot 1545 - 74{,}8225 = 2{,}4275;$$

в) вибіркове середнє квадратичне відхилення обчислюємо за формулою (1.14):

$$\overline{\sigma} = \sqrt{\overline{D}(x)} = \sqrt{2{,}4275} = 1{,}558;$$

г) розмах вибірки дорівнює:

$$R = 12 - 6 = 6.$$

3. Знаходимо інтервальний статистичний розподіл, поділивши проміжок $[6;\,12]$ на 4 рівні частини довжиною 1,5 і отримуємо інтервальну таблицю частот та відносних частот:

$[z_{i-1}, z_i)$	$[6;7{,}5)$	$[7{,}5;9)$	$[9;10{,}5)$	$[10{,}5;12]$
n_i	5	4	9	2
w_i	$5/20$	$4/20$	$9/20$	$2/20$

Гістограмою відносних частот є східчаста фігура, яка складається з прямокутників з основами $[z_{i-1}; z_i]$ і висотами:

$$h_1 = \frac{5}{1{,}5} = 3{,}3, \quad h_2 = \frac{4}{1{,}5} = 2{,}7, \quad h_3 = \frac{9}{1{,}5} = 6, \quad h_4 = \frac{2}{1{,}5} = 1{,}3.$$

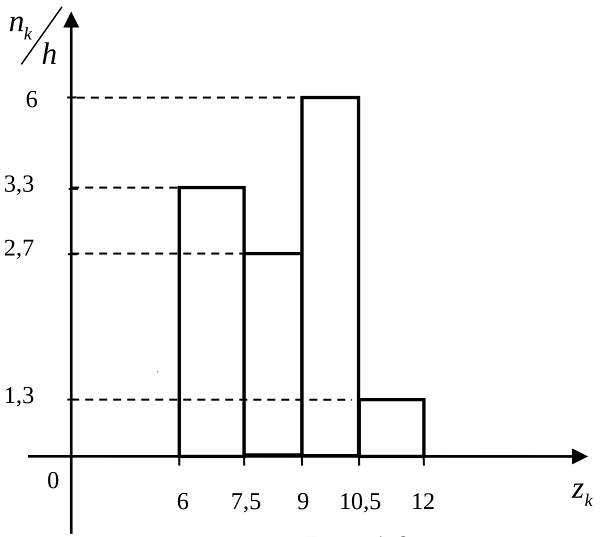
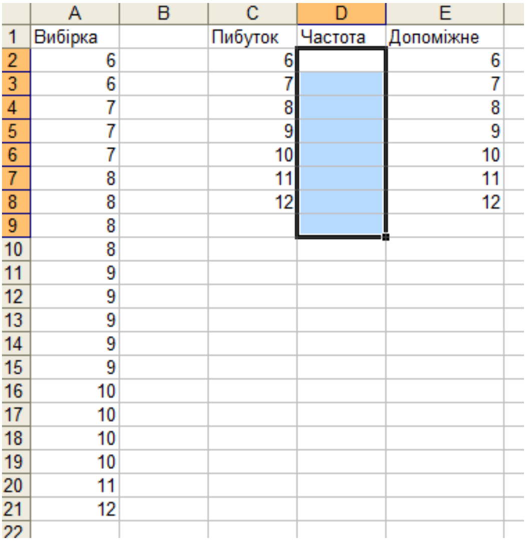


Рис. 1.2.

Якщо обсяг вибірки невеликий, то всі розрахунки і побудови можна виконати вручну. Для обробки вибірок великих обсягів використовують прикладні математичні пакети. Розглянемо розв'язання задачі за допомогою функцій Exel.

Дискретний статистичний розподіл

1. Відкриваємо робочий лист, вводимо дані задачі в стовпець **A2:A21** і сортуємо їх за зростанням. Вводимо в комірку **C1** назву «прибуток», в комірку **D1** «частота», в комірку **E1** "допоміжне".
2. Із відсортованого вмісту стовпця A2:A21 видно, що числові дані змінюються від 6 до 12 з кроком 1. Використовуючи маркер заповнення вводимо всі натуральні числа від 6 до 12 в стовпець **D2:D8**. В стовпець **E2:E8** вводимо натуральні числа від 6 до 12.
3. Виділяємо стовпець **D2:D9** (на одну комірку більше, ніж є чисел в стовпці «допоміжне»).



	A	B	C	D	E
1	Вибірка		Пибуток	Частота	Допоміжне
2	6		6		6
3	6		7		7
4	7		8		8
5	7		9		9
6	7		10		10
7	8		11		11
8	8		12		12
9	8				
10	8				
11	9				
12	9				
13	9				
14	9				
15	9				
16	10				
17	10				
18	10				
19	10				
20	11				
21	12				
22					

Рис. 1.3

4. Вибираємо меню **Вставка → Функція → Статистичні → ЧАСТОТА**

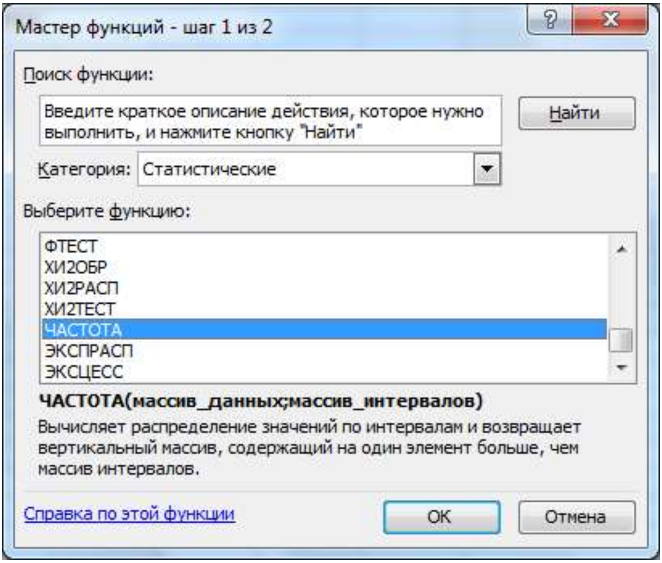


Рис 1.4. Вибір статистичної функції «Частота»

5. У меню функції **ЧАСТОТА**, що з’явилось у вікні, вводимо адреси комірок: **Масив_даних** – комірки **A2:A21**; **Масив_інтервалів** – комірки **E2:E21**. Виставляємо курсор в командну стрічку та натискаємо комбінацію клавіш <**Ctrl + Shift + Enter>** . В результаті в комірках **D2:D9** отримаємо розподіл частот.
6. Стовпці **«Частота»** , **«Допоміжне», «Вибірка»** пов’язані між собою. Це призводить до незручностей під час копіювання вмісту цих стовпців у інший робочий лист. Щоб їх позбутися, потрібно виділити і скопіювати комірки **A2:E21**, а потім зайти в меню **Правка → Спеціальная вставка → Значення.** Тепер можемо скопіювати комірки **C1:D8** вміст яких і є дискретним статистичним розподілом в новий робочий лист.

	A	B
1	Пибуток	Частота
2	6	2
3	7	3
4	8	4
5	9	5
6	10	4
7	11	1
8	12	1

Рис. 1.5

Полігон частот

Використовуємо числові дані розташовані в стовпці **A2:A2**1 та допоміжний стовпчик **E2:E8**.

1. Заходимо в меню **Сервіс → Аналіз даних → Гістограма.** В діалоговому вікні вводимо вхідні дані, як показано на рисунку 1.6

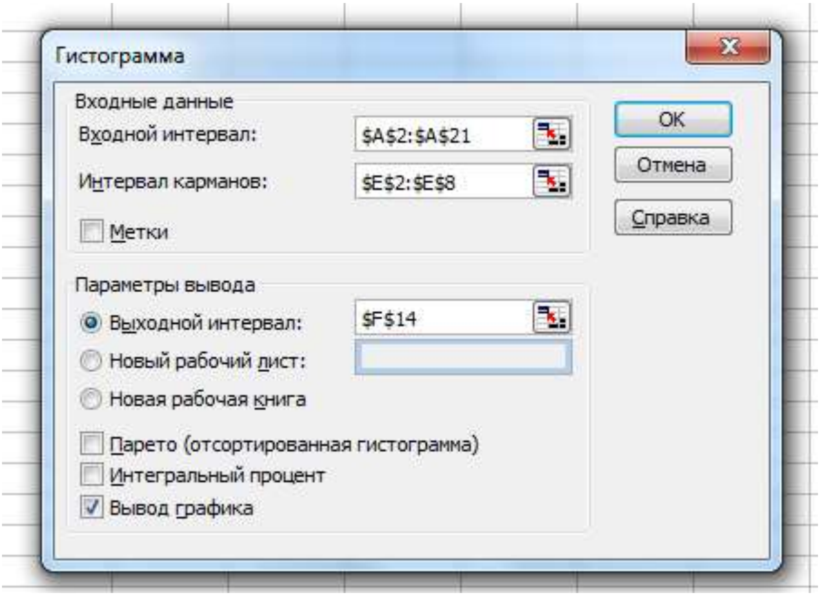


Рис. 1.6

Вхідний інтервал – це комірки **A2:A21**, інтервал кишень – це комірки **E2:E8**. Вихідним інтервалом може бути будь-яка вільна комірка робочого листа, в якій буде розміщено полігон частот. Обов'язково потрібно поставити мітку навпроти **Виведення графіка**. Натискаємо **OK**, одержуємо гістограму частот (рис.1.7), яку потрібно відредагувати.

Карман	*Частота*
6	2
7	3
8	4
9	5
10	4
11	1
12	1
Еще	0

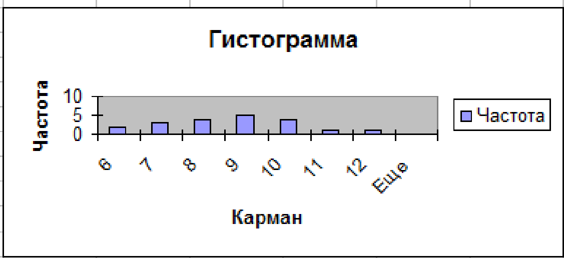


Рис 1.7

2. Вилучаємо надписи **Гистограмма**, **Карман** та **Частота**.

3. Правою клавішею миші клікаємо на одному із стовчиків гістограми У меню, яке з'явилось, вибираємо опцію **«Тип даграммы…»**, а в ньому – опції **«Стандартные»** та «**График**» (рис. 1.8). Натискаємо кнопку **<ОК**>.
4.

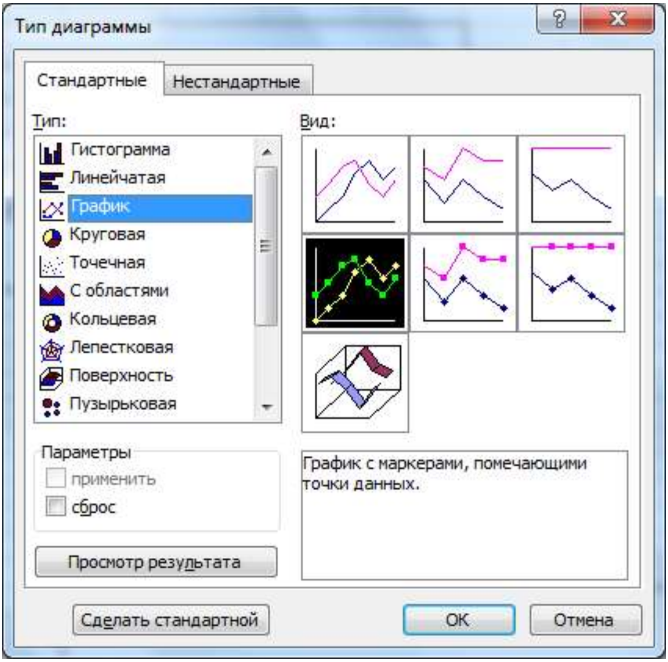


Рис 1.8

Одержуємо зображення полігону частот (рис.1.9)

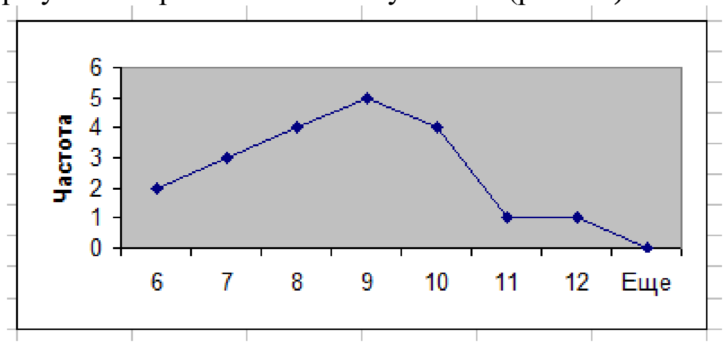


Рис. 1.9.

Емпірична функція розподілу

Для побудови емпіричної функції розподілу потрібно зайти в меню **Сервіс → Аналіз даних → Гістограма,** як на рис. 1.8. і поставити мітку на віконечку **Інтегральний процент.** Одержуємо одночасне зображення гістограми частот та емпіричної функції розподілу. Редагуємо рисунок, вилучаючи із зображення гістограму. Для цього потрібно правою клавішею миші клікнути на будь-якому стовпці гістограми і натиснути клавішу **Del**. Одержуємо графік емпіричної функції розподілу (рис. 1.10).

Карман	Частота	егральный %
6	2	10,00%
7	3	25,00%
8	4	45,00%
9	5	70,00%
10	4	90,00%
11	1	95,00%
12	1	100,00%
Еще	0	100,00%

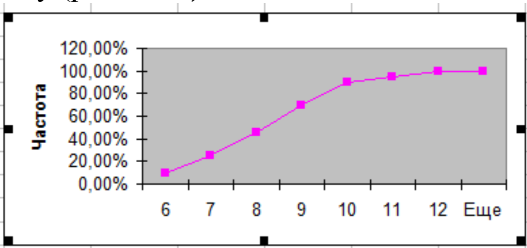


Рис. 1.10

Числові характеристики – вибіркове середнє, дисперсія, середньоквадратичне відхилення обчислюється за допомогою функцій СРЗНАЧ, ДИСП, СТАНДОТКЛОН, відповідно.

Інтервальний статистичний розподіл та гістограма частот

Побудова інтервального статистичного розподілу виконується за допомогою статистичної функції **ЧАСТОТА**.

Вхідні дані для виконання завдання знаходяться в стовпці **A2:A21**. Оскільки в умові задачі вказано розбити проміжок на 4 рівні частини, то в стовпець «Прибуток» вводимо інтервали 6-7,5; 7,5-9; 9-10,5; 10,5-12. Праві кінці інтервалів розподілу є відкритими, отже від кінця кожного з інтервалів робимо невеликий відступ і в стовпець «Допоміжне» вводимо числа 7,4; 8,9; 10,4; 12. В стовпці «Частота» виділяємо комірки, на одну більше, ніж є даних в стовпці «Допоміжне» і викликаємо функцію **ЧАСТОТА** (рис. 1.11). В діалоговому вікні, що з'явилось вводимо адреси комірок **Масив_данних** – комірки **A2:A21**; **Масив_інтервалів** – комірки **E2:E5.** Виставляємо курсор в командну стрічку та натискаємо комбінацію клавіш **<Ctrl + Shift + Enter>.** В результаті у комірках **D2:D5** отримаємо розподіл частот (Рис. 1.12.)

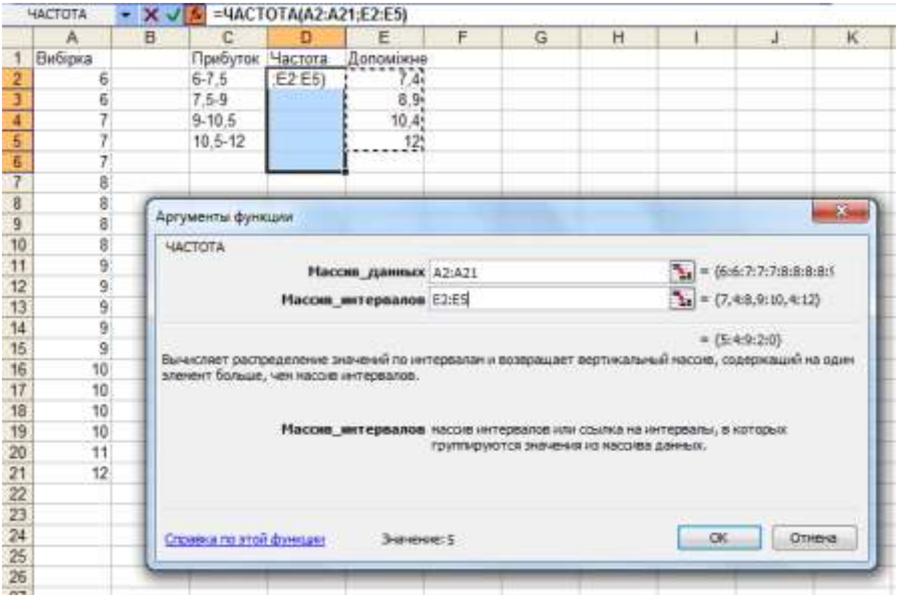


Рис 1.11

C	D
Прибуток	Частота
6-7,5	5
7,5-9	4
9-10,5	9
10,5-12	2
	0

Рис. 1.12

Гістограма частот

Заходимо в меню **Сервіс → Аналіз даних → Гістограма.** В діалоговому вікні вводимо вхідні дані, як показано на рисунку 1.6, лише з однією зміною, що **Інтервал карманів** – це комірки **E2:E5**. Одержану гістограму потрібно відредагувати, видаливши надписи «Гистограмма» та «Карман». Також потрібно зменшити відстань між стовпцями гістограми. Для цього клікаємо правою клавішею миші на одному із стовпців та у випадаючому меню вибираємо опції: **Формат рядів даних → Параметри** і зменшуємо **ширину зазора** до 0. Одержуємо гістограму (рис.1.13)

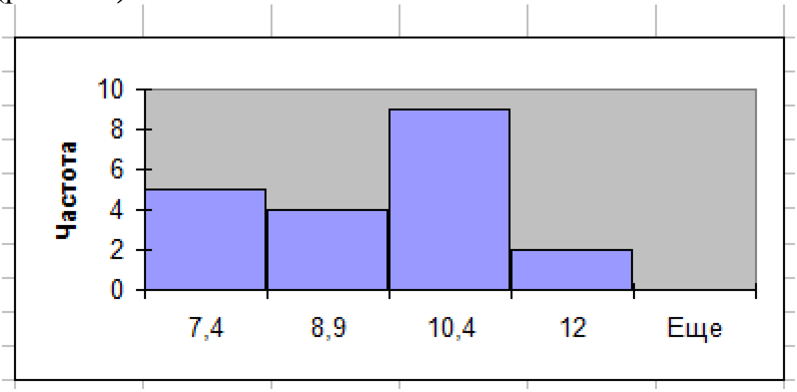


Рис. 1.13

2. Оцінки невідомих параметрів розподілу

2.1. Точкові оцінки

Означення 1. Будь-яку однозначну функцію $\theta^* = \theta^*(x_1, x_2, ..., x_n)$, за допомогою якої знаходять наближене значення невідомого параметра θ відомого розподілу випадкової величини X, називають **точковою оцінкою** цього параметра.

Точкова оцінка θ^* сама є випадковою величиною, яка для різних вибірок набуває різних значень.

Для того, щоб оцінка θ^* була в певному сенсі «найкращою», тобто мала практичну цінність, вона повинна справджувати умови **незміщеності, змістовності** та **ефективності.**

Означення 2. Точкова оцінка $\theta^* = \theta^*\left(x_1, x_2, ..., x_n\right)$ для параметра θ випадкової величини X називається **незміщеною,** якщо її математичне сподівання дорівнює точному значенню цього параметра, тобто

$$M\left(\theta^*\right) = \theta$$

В інших випадках точкову оцінку θ^* називають **зміщеною.**

Означення 3. Точкова оцінка $\theta^* = \theta^*\left(x_1, x_2, ..., x_n\right)$ параметра розподілу θ називається **змістовною (консистентною)**, якщо θ^* збігається за ймовірністю до параметра θ випадкової величини X за необмеженого зростання числа спостережень, тобто виконується така рівність:

$$\lim_{n\to\infty} P\left\{\left|\theta^* - \theta\right| < \varepsilon\right\} = 1,$$

де $\varepsilon > 0$ - як завгодно мале число.

Для виконання даної граничної рівності достатньо, щоб оцінка θ^* була незміщеною і щоб дисперсія оцінки необмежено наближалась до нуля при $n \to \infty$, тобто :

$$\lim_{n\to\infty} D\left(\theta^*\right) = 0$$

Оскільки θ^* — випадкова величина, значення якої змінюються від вибірки до вибірки, то міра її розсіювання навколо математичного сподівання θ характеризуються дисперсією $D\left(\theta^*\right)$.

Означення 4. Незміщена оцінка $\theta^* = \theta^*\left(x_1, x_2, ..., x_n\right)$ називається **ефективною**, якщо вона має найменшу дисперсію серед усіх незміщених оцінок параметра θ, обчислених за вибірками одного і того ж обсягу.

Через те, що параметри розподілу випадкової величини часто, у простий спосіб, виражаються через її чисельні характеристики, перш за все виникає задача про оцінки основних чисельних характеристик випадкової величини X .

2.2.Точкові оцінки математичного сподівання, дисперсії та середнього квадратичного відхилення

Нехай $x_1,\ldots,x_n$ - вибірка, яка отримана в результаті n незалежних спостережень над випадковою величиною X - деякою ознакою генеральної сукупності, яка має математичне сподівання $M\left(X\right)=a$.

За точкову оцінку M^* математичного сподівання $a=M\left(X\right)$ беруть вибіркове середнє*:*

$$M^* = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.1}$$

для незгрупованої вибірки і

$$M^* = \overline{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i \tag{2.2}$$

для статистичного розподілу вибірки (1.5).

Можна показати, що оцінка $M^*=\overline{x}$ є незміщеною, змістовною за умови, що випадкова величина X має скінчену дисперсію. Якщо випадкова величина X розподілена за нормальним законом з параметрами $M\left(X\right)=a,\quad D\left(X\right)=\sigma^2$, то оцінка $\overline{x}$ є й ефективною.

За точкову оцінку D^* дисперсії $D\left(X\right)=\sigma^2$ беруть вибіркову дисперсію*:*

$$D^* = \overline{D} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 \tag{2.3}$$

у випадку незгрупованої вибірки або

$$D^* = \bar{D} = \frac{1}{n}\sum_{i=1}^{n} n_i (x_i - \bar{x})^2 \qquad (2.4)$$

для статистичного розподілу вибірки (1.5), які є зміщеними оцінками для параметра σ^2.

За точкову оцінку дисперсії $D(X) = \sigma^2$ **беруть** також **виправлені вибіркові дисперсії:**

$$D^* = \tilde{D} = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \text{і} \qquad (2.5)$$

$$D^* = \tilde{D} = \frac{1}{n-1}\sum_{i=1}^{n} n_i (x_i - \bar{x})^2 \qquad (2.6)$$

відповідно, які є незміщеними оцінками параметра $D(X) = \sigma^2$.

Можна показати, що оцінки $\bar{D}$ і $\tilde{D}$ є **змістовними**, проте не є **ефективними.**

Зауважимо, що $\bar{D}$ і $\tilde{D}$ пов’язані співвідношенням:

$$\tilde{D} = \frac{n}{n-1}\bar{D}. \qquad (2.7)$$

Якщо випадкова величина розподілена за нормальним законом і математичне сподівання a - відоме, то **незміщеною, грунтовною та ефективною оцінкою дисперсії** $D(X) = \sigma^2$ є одна з таких оцінок:

$$D^* = \bar{D} = \frac{1}{n}\sum_{i=1}^{n} (x_i - a)^2$$

$$D^* = \bar{D} = \frac{1}{n}\sum_{i=1}^{n} n_i (x_i - a)^2 .$$

Оскільки середнє квадратичне відхилення $\sigma(X)$ дорівнює $\sqrt{D(X)}$, то **за оцінку** σ^* **параметра** σ **можна**

вибрати один із варіантів вибіркового середнього квадратичного відхилення:

$$\sigma^* = \sqrt{D^*}\,, \tag{2.8}$$

де D^* одна з оцінок (2.3)-(2.6).

Слід зауважити, що всі оцінки для σ, які обчислюються за формулою (2.7), не є **незміщеними, ефективними**, а є лише **змістовними**.

Приклад 2.1. Статистичні дослідження величини годинного доходу працюючого офіцера-пожежника дали такі результати:

x_i -дохід у гр.	6	7	8	9	10	11	12	13	14
n_i - кількість офіцер.	1	2	3	20	25	24	15	7	3

Обчислити точкові оцінки чисельних характеристик $M(X)$, $D(X)$, $\sigma(X)$, де X рівень доходу одного працюючого офіцера.

Розв'язання. За точкову оцінку математичного сподівання $M(X)$ беремо вибіркове середнє і обчислюємо її за формулою (2.1):

$$M^* = \overline{x} = \frac{1}{100}(1\cdot 6 + 2\cdot 7 + 3\cdot 8 + 20\cdot 9 + 25\cdot 10 + 24\cdot 11 + 15\cdot 12 +$$

$$+7\cdot 13 + 3\cdot 14) = \frac{1}{100}\cdot 1051 = 10,51.$$

Отже середній дохід працівника протягом дня становить $10,51$ гривні.

Точкову оцінку дисперсії $D(X)$ обчислимо у двох варіантах:

зміщену точкову оцінку обчислимо за формулою (2.4):

$$D^* = \bar{D} = \frac{1}{100}\Big[1\cdot(-4,5)^2 + 2\cdot(-3,51)^2 + 3\cdot(-2,51)^2 + 20\cdot(-1,51)^2 +$$

$$25\cdot(-0,51)^2 + 24\cdot(0,49)^2 + 15\cdot(1,49)^2 + 7\cdot(2,49)^2 + 3\cdot(3,49)^2\Big] =$$

$$= \frac{1}{100}\cdot 234,99 = 2,3499;$$

незміщена оцінка (виправлена дисперсія) обчислюється за формулою (2.7):

$$D^* = \tilde{D} = \frac{100}{99}\cdot 2,3499 = 2,3736.$$

Як видно з попередніх обчислень, відхилення зміщеної оцінки $\bar{D}$ від незміщеної $\tilde{D}$ становить $\tilde{D} - \bar{D} = 0,0237$ і є порівняно мале, бо обсяг вибірки $n = 100$ є досить великим.

Для середнього квадратичного відхилення отримуємо такі оцінки:

$$\sigma^* = \bar{\sigma} = \sqrt{2,3499} = 1,5329,$$

$$\sigma^* = \tilde{\sigma} = \sqrt{2,3736} = 1,5406.$$

Приклад 2.2. Статистичні дослідження зростання продуктивності праці підприємств регіону в поточному році у відсотках до відповідного періоду попереднього року виражаються інтервальним розподілом вибірки:

$[z_{i-1}, z_i]$	$80-90$	$90-100$	$100-110$	$110-120$	120-130
n_i	2	14	60	20	4

Обчислити оцінки для $M(X)$, $D(X)$, $\sigma(X)$, де X зростання продуктивності праці одного підприємства регіону у відсотках до відповідного періоду попереднього року.

Розв'язання. Точкову оцінку математичного сподівання $M\left(X\right)$ обчислимо за формулою (2.1) з врахуванням того, що $x_i = \frac{z_{i-1} + z_i}{2}$:

$$M^* = \overline{x} = \frac{1}{100}\left(2\cdot 85 + 14\cdot 95 + 60\cdot 105 + 20\cdot 115 + 4\cdot 125\right) =$$

$$= \frac{1}{100}\cdot 10600 = 106.$$

Отже середнє зростання продуктивності праці одного підприємства в цьому році у відсотках до відповідного періоду попереднього року становить 106 %.

Зміщену точкову оцінку дисперсії $D\left(X\right)$ обчислимо за формулою :

$$D^* = \overline{D} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \overline{x}^2$$

з таким же вибором x_i, як при знаходженні точкової оцінки $M\left(X\right)$. Маємо:

$$D^* = \overline{D} = \frac{1}{100}\left(2\cdot 85^2 + 14\cdot 95^2 + 60\cdot 105^2 + 20\cdot 115^2 + 4\cdot 125^2\right) - 106^2 =$$

$$= \frac{1}{100}\cdot 1129300 - 11236 = 57.$$

За формулою (2.7) обчислимо незміщену точкову оцінку дисперсії:

$$D^* = \tilde{D} = \frac{100}{99}\cdot \overline{D} = \frac{100}{99}\cdot 57 = 57,(57).$$

Для середнього квадратичного відхилення маємо такі оцінки:

$$\sigma^* = \overline{\sigma} = \sqrt{57} = 7,55, \quad \sigma^* = \tilde{\sigma} = \sqrt{57,(57)} = 7,59.$$

Обчислимо ці числові характеристики за допомогою калькулятора Exel. Для цього відкриємо робочий лист та введемо дані задачі, перейшовши від інтервального до дискретного статистичного розподілу. Замінимо інтервали їх серединами та скористаємось маркером заповнення: 85 – 2 рази, 95 -- 14 разів, 105 – 60 разів, 115 – 20 разів, 125 – 4 рази. Всі дані вводимо в стовпчик. Дані задачі знаходяться в совпчику **А2:А101**. Математичне сподівання та його точкова очінка обчислюються за допомогою функціїї **СРЗНАЧ** (рис. 2.1)

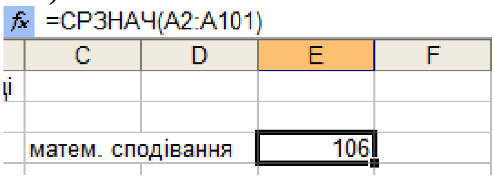


Рис. 2.1

Зміщену оцінку дисперсії обчислюємо за допомогою функції **=ДИСПР** (рис. 2.2)

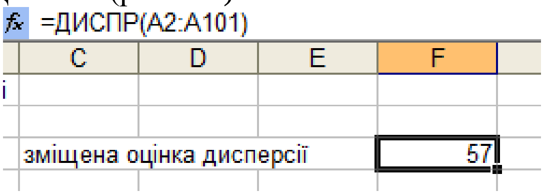


Рис. 2.2

Незміщену оцінку дисперсії обчислюємо за допомогою функції **=ДИСП** (рис. 2.3)

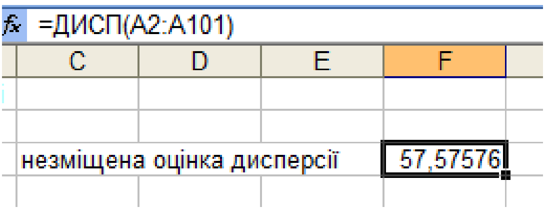


Рис. 2.3

Зміщену оцінку середньоквадратичного відхилення обчислюємо за допомогою функції =СТАНДОТКЛОНП, незміщену оцінку середньоквадратичного відхилення обчислюємо за допомогою функції =СТАНДОТКЛОН.

2.3. Інтервальні оцінки параметрів розподілу

Точкова оцінка θ^* параметра розподілу θ є досить близькою до його точного значення, якщо обсяг вибірки досить великий. Якщо ж обсяг вибірки невеликий, то між точковою оцінкою θ^* і точним значенням параметра θ можуть бути значні розбіжності. У зв'язку з цим виникає питання про надійність точкової оцінки.

Зрозуміло, що точкова оцінка θ^* параметра θ є тим точнішою, чим менша величина різниці $\left|\theta-\theta^*\right|$. Якщо б вдалось встановити, що $\left|\theta-\theta^*\right|<\delta,$ то число $\delta>0$ характеризувало б точність точкової оцінки θ^*. Однак статистичні методи не дозволяють категорично стверджувати, що $\left|\theta-\theta^*\right|<\delta,$ бо θ^* випадкова величина. Можна лише говорити про ймовірність γ, з якою ця нерівність виконується.

Означення 5. Надійністю (довірчою ймовірністю) точкової оцінки θ^* параметра розподілу θ називають ймовірність γ, з якою виконується нерівність $\left|\theta-\theta^*\right|<\delta,$ тобто

$$P\left(\left|\theta-\theta^*\right|<\delta\right)=\gamma. \tag{2.9}$$

На практиці надійність оцінки задається наперед, причому число γ вибирають близьким до одиниці: $\gamma=0,95;\quad \gamma=0,99,\quad \gamma=0,999.$

Співвідношення (2.9) можна записати так:

$$P\left(\theta^*-\delta<\theta<\theta^*+\delta\right)=\gamma. \tag{2.10}$$

Означення 6. Інтервал $\left(\theta^*-\delta,\theta^*+\delta\right),$ для якого виконується рівність (2.10), називається **довірчим (надійним) інтервалом**, а його межі $\theta^*-\delta,$ $\theta^*+\delta$ **довірчими (надійними) межами** для параметра розподілу $\theta.$

Загальний спосіб, за допомогою якого знаходять довірчий інтервал, полягає в розв’язанні рівняння (2.10) і визначення з нього числа δ. Для цього потрібно обчислити ймовірність $P\left(\theta^{*}-\delta<\theta<\theta^{*}+\delta\right)$. Це обчислення можна виконати, якщо відомий закон розподілу точкової оцінки $\theta^{*}\left(x_1, x_2, x_3, \ldots, x_n\right)$ або пов’язаної з нею іншої випадкової величини, бо при цьому можна використати відомі формули теорії ймовірностей:

$$P\left(\alpha \le \theta^{*}<\beta\right)=F\left(\beta\right)-F\left(\alpha\right), \quad P\left(\alpha \le \theta^{*}<\beta\right)=\int_{\alpha}^{\beta} f\left(x\right)dx,$$

де $F\left(x\right)$ функція розподілу, а $f\left(x\right)$ густина розподілу випадкової величини θ^{*}.

При знаходженні довірчих інтервалів точкових оцінок параметрів розподілу поряд з відомими законами розподілу випадкових величин (закон розподілу Пуассона, нормальний закон розподілу, показниковий закон) в статистиці часто застосовують розподіли: **«хі-квадрат», Стьюдента і Фішера-Снедекора.**

2.4. Розподіл χ^2 - «хі-квадрат»

Нехай $X_1, X_2, \ldots, X_n$ незалежні і розподілені за нормальним законом випадкові величини з математичними сподіваннями $M\left(X_i\right)=0, \quad \forall i=1,2,\ldots,n$ та середніми квадратичними відхиленнями $\sigma\left(X_i\right)=1, \quad i=1,2,\ldots,n$.

Випадкова величина $\left[1,4,5\right]$

$$\chi^2=\sum_{i=1}^{n} X_i^2$$

має розподіл χ^2 з n ступенями вільності, який характеризується густиною:

$$R(x,n)=\begin{cases} 0, & x<0; \\ A_n x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x\ge 0, \end{cases}$$

де A_n стала, яка визначається з умови нормування:

$$A_n \cdot \int_0^{\infty} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx = 1.$$

Розподіл «хі-квадрат» залежить від одного параметра n і при $n \to \infty$ він наближається до **нормального** закону розподілу.

2.5. Розподіл Стьюдента

Нехай Z випадкова величина, яка розподілена за нормальним законом, причому $M(Z)=0,\quad \sigma(Z)=1,$ а V незалежна від Z випадкова величина, яка розподілена за законом «хі-квадрат» з n ступенями вільності. Випадкова величина

$$T=\frac{Z}{\sqrt{\frac{V}{n}}}$$

має розподіл Стьюдента з $k=n$ ступенями вільності, який характеризується густиною:

$$S(x,n)=B_n\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}, \quad x\in(-\infty,\infty),$$

де B_n нормуюча константа.

Розподіл Стьюдента також залежить від одного параметра n і при $n \to \infty$ наближається до стандартного **нормального розподілу.**

2.6. Розподіл Фішера-Снедекора

Нехай U і V незалежні випадкові величини, які мають розподіли χ^2 зі ступенями вільності k_1 і k_2 відповідно. Випадкова величина

$$F = \frac{\dfrac{U}{k_1}}{\dfrac{V}{k_2}}$$

залежить від двох параметрів – ступенів вільності k_1 і k_2 і задається густиною ймовірностей:

$$f_F\left(x, k_1, k_2\right) = C_{k_1 k_2} \cdot x^{\frac{k_1}{2}-1}\left(1 + \frac{k_1}{k_2}x\right)^{-\frac{k_1+k_2}{2}}, \quad x \geq 0,$$

де коефіцієнт $C_{k_1 k_2}$ визначається з умови нормування. Розподілу Фішера - Снедекора підпорядковується зокрема відношення дисперсій двох незалежних вибірок обсягів n і m із двох нормально розподілених генеральних сукупностей з однаковими дисперсіями. В цьому випадку $k_1 = n-1$ і $k_2 = m-1$.

2. 7. Інтервальні оцінки математичного сподівання і середнього квадратичного відхилення випадкової величини, яка розподілена за нормальним законом

2.7.1. Інтервальні оцінки математичного сподівання

Нехай випадкова величина розподілена за нормальним законом, тобто характеризується густиною розподілу:

$$f\left(x, a, \sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \left(-\infty, \infty\right).$$

Розглянемо два випадки оцінювання невідомого параметра a розподілу на основі вибірки:

- нехай параметр σ (середнє квадратичне відхилення) — **відомий.** Тоді рівність (2.10) виглядає так:

$$P\left(\bar{x} - \frac{\sigma}{\sqrt{n}}t_\gamma < a < \bar{x} + \frac{\sigma}{\sqrt{n}}t_\gamma\right) = \gamma, \tag{2.11}$$

де $\bar{x}$ – вибіркове середнє (1.11) , n обсяг вибірки, а $t = t_\gamma$ розв'язок рівняння $\Phi(t) = \frac{\gamma}{2}$, де $\Phi(x)$ функція Лапласа, тобто інтервал $\left(\bar{x} - \frac{\sigma}{\sqrt{n}} t_\gamma, \bar{x} + \frac{\sigma}{\sqrt{n}} t_\gamma \right)$ є довірчим інтервалом, що «накриває» невідомий параметр a з надійністю γ;

- нехай параметр σ (середнє квадратичне відхилення) **невідомий**. Тоді рівність (2.10) має вигляд:

$$P\left(\bar{x} - \frac{\tilde{\sigma}}{\sqrt{n}} t_{\gamma,n} < a < \bar{x} + \frac{\tilde{\sigma}}{\sqrt{n}} t_{\gamma,n} \right) = \gamma, \qquad (2.12)$$

де $\bar{x}$ вибіркове середнє, $\tilde{\sigma}$ виправлене вибіркове середнє квадратичне відхилення (2.8), n обсяг вибірки, $t = t_{\gamma,n}$ розв'язок рівняння:

$$\int_0^t S(x,n)\,dx = \frac{\gamma}{2}, \qquad (2.13)$$

де $S(x,n)$ густина розподілу Стьюдента. Розв'язок рівняння (2.13) знаходимо з таблиці додатка 4 за даними значеннями обсягу вибірки n та надійності γ.

Приклад 2.3. Відомо, що випадкова величина X, яка в результаті спостереження набула значень: 98,2; 100,2; 98,1; 96,2; 99,8; 101,2; 99,2; 104,1; 102,6; 103,8; 101,2; 99,4; 106,1; 102,6; 100,6; 98,8; 98,2; 101,1; 100,6; 99,8 розподілена за нормальним законом з $\sigma(X) = 1{,}9$. Знайти довірчий інтервал, який з надійністю $\gamma = 0{,}95$ «накриває» невідоме математичне сподівання випадкової величини X.

Розв'язання. Оскільки середнє квадратичне відхилення $\sigma(X) = 1{,}9$ відоме, то довірчий інтервал шукатимемо за формулою (2.11). Формула (2.11) передбачає, що відомим є $\bar{x}$ і t_γ. Знайдемо $\bar{x}$ за статистичним матеріалом задачі при обсязі вибірки $n = 20$. Маємо:

$$\overline{x} = \frac{1}{20}(98,2+100,2+98,1+96,2+99,8+101,2++99,2+104,1+$$

$$+102,6+103,8+101,2+99,4+106,1+102,6+100,6+98,8+$$

$$+98,2+101,1++100,6+99,8) = \frac{1}{20}\cdot 2011,8 = 100,6.$$

Величину $t = t_\gamma$ знайдемо з рівняння $\Phi(t) = \frac{0,95}{2}$, використавши таблицю додатка 2, з якої знаходимо що $t_\gamma = 1,96$.

За формулою (2.11) маємо:

$$100,6 - \frac{1,9}{\sqrt{20}}\cdot 1,96 < a < 100,6 + \frac{1,9}{\sqrt{20}}\cdot 1,96,$$

Звідки остаточно маємо:

$$99,7 < a < 101,43.$$

Отже інтервал $(99,7;101,43)$ є довірчим інтервалом, який «накриває» невідомий параметр a з надійністю $\gamma = 0,95$.

Приклад 2.4. За спостереженнями випадкова величина X тижневий прибуток фермерів (в тис. грн) характеризується таким статистичним розподілом вибірки:

Прибуток x_i	5	6	7	8	9	10	11	12
К-сть фермерів, n_i	1	2	4	6	7	5	3	2

Припускаючи, що випадкова величина X має нормальний закон розподілу ймовірностей, знайти інтервальну оцінку невідомого математичного сподівання a з надійністю $\gamma = 0,999$.

Розв'язання. В цьому прикладі середнє квадратичне відхилення σ невідоме, тому для отримання інтервальної оцінки для невідомого параметра a використаємо формулу (2.12). Формула (2.12) передбачає, що є відомими вибіркове середнє $\bar{x}$, виправлене середнє квадратичне відхилення $\tilde{\sigma}$ та $t_{\gamma,n}$. Знайдемо перші дві оцінки із заданого статистичного розподілу вибірки:

$$\bar{x} = \frac{1}{30}(1\cdot 5 + 2\cdot 6 + 4\cdot 7 + 6\cdot 8 + 7\cdot 9 + 5\cdot 10 + 3\cdot 11 + 2\cdot 12) = 8,77;$$

$$\tilde{\sigma}^2 = \frac{1}{29}(1\cdot 5^2 + 2\cdot 6^2 + 4\cdot 7^2 + 6\cdot 8^2 + 7\cdot 9^2 + 5\cdot 10^2 + 3\cdot 11^2 + 2\cdot 12^2) -$$

$$-\frac{30}{29}\cdot 8,77^2.$$

Отже $\tilde{\sigma} = \sqrt{\frac{1}{29}\cdot 2395 - 79,57} = 1,74.$

За обсягом вибірки $n = 30$ і надійністю $\gamma = 0,999$ з таблиці додатка 4 знаходимо $t_{\gamma,n} = 3,659.$

Отже за формулою (2.12) маємо:

$$8,77 - \frac{1,74}{\sqrt{30}}\cdot 3,659 < a < 8,77 + \frac{1,74}{\sqrt{30}}\cdot 3,659,$$

Звідки остаточно отримуємо:

$$7,61 < a < 9,93.$$

Таким чином інтервал $(7,61; 9,93)$ це надійний інтервал, який «накриває» невідоме математичне сподівання a (при невідомому σ) з надійністю $\gamma = 0,999$.

2.7.2. Інтервальні оцінки середнього квадратичного відхилення

Оскільки однією з точкових оцінок для середнього квадратичного відхилення σ є виправлене середнє квадратичне відхилення σ^*, яке визначене формулою (2.8), то для знаходження інтервальної оцінки для нього потрібно розв'язати рівняння

$$P\left(\left|\sigma^* - \sigma\right| < \delta\right) = \gamma,$$

або еквівалентне рівняння

$$P\left(\sigma^* - \delta < \sigma < \sigma^* + \delta\right) = \gamma, \qquad (2.14)$$

де γ – надійність.

Для розв'язання рівняння (2.14) використовуємо розподіл «хі-квадрат» і приходимо до такого **висновку:**

- *якщо розв'язок рівняння* $q = q_{\gamma,n}$

$$\int\limits_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} R(t,n)\,dt = \gamma \qquad (2.15)$$

менший від одиниці $\left(q_{\gamma,n} < 1\right)$,*то з надійністю* γ *середнє квадратичне відхилення* σ *випадкової величини* X *справджує нерівність*

$$\sigma^*\left(1 - q_{\gamma,n}\right) < \sigma < \sigma^*\left(1 + q_{\gamma,n}\right), \qquad (2.16)$$

де $R(t,n)$ – *густина розподілу «хі-квадрат»* (див. 2.4).

Розв'язок $q = q_{\gamma,n}$ *знаходимо за даними* n, γ *з таблиці додатка 5.*

- *якщо* $q_{\gamma,n} > 1$, то нерівність (2.16) набуває вигляду

$$0 < \sigma < \sigma^*\left(1 + q_{\gamma,n}\right) \quad , \qquad (2.17)$$

а рівняння (2.15) *зводиться до такого*

$$\int\limits_{\frac{\sqrt{n-1}}{1+q}}^{0} R(t,n)\,dt = \gamma\,. \qquad (2.18)$$

Для знаходження розв'язку рівняння (2.18) *для даних* n, γ *також використовують таблицю додатка 5.*

Приклад 2.5. За даними вибірки прикладу 2.3. знайти інтервальну оцінку середнього квадратичного відхилення σ для заданої надійності $\gamma = 0,95$.

Розв'язання. Вибіркове середнє $\overline{x} = 100,6$ було обчислене в прикладі 2.3. Обчислимо виправлене вибіркове середнє квадратичне відхилення:

$$\sigma^* = \frac{1}{\sqrt{19}}\Big[1\cdot(-4,4)^2 + 1\cdot(-2,5)^2 + 2\cdot(-2,4)^2 + 1\cdot(-1,8)^2 +$$

$$+1\cdot(-1,4)^2 + 1\cdot(-1,2)^2 + 2\cdot(-0,8)^2 + 1\cdot(-0,4)^2 + 1\cdot(-1,4)^2 +$$

$$+2\cdot 0^2 + 1\cdot 0,5^2 + 2\cdot 0,6^2 + 2\cdot 2^2 + 1\cdot 3,2^2 + 1\cdot 3,5^2 + 1\cdot 5,5^2\Big]^{\frac{1}{2}} =$$

$$\left(\frac{1}{19}\cdot 106,92\right)^{\frac{1}{2}} \approx 2,37.$$

За таблицею додатка 5 для значень $n = 20$, $\gamma = 0,95$ знаходимо $q = q_{\gamma,n} = 0,37 < 1.$ Тому довірчий інтервал для σ визначиться нерівністю

$$2,37(1-0,37) < \sigma < 2,37(1+0,37).$$

Отже з ймовірністю $\gamma = 0,95$ середнє квадратичне відхилення потрапляє в інтервал $(1,49;3,25)$.

Приклад 2.6. Відомо,що вибірка обсягу $n = 10$ репрезентує випадкову величину X, яка розподілена за нормальним законом з вибірковим середнім квадратичним відхиленням $\sigma^* = 0,16.$ Знайти довірчий інтервал для σ з надійністю $\gamma = 0,99.$

Розв'язання. За даними значень $n = 10$ і $\gamma = 0,99$ з таблиці додатка 5 знаходимо $q_{\gamma,n} = 1,08.$ Оскільки $q_{\gamma,n} > 1$, то для знаходження довірчого інтервалу використаємо нерівність (2.17). Отже маємо

$$0 < \sigma < 0,16(1+1,08).$$

і остаточно

$$0 < \sigma < 0,33.$$

Тому інтервал $(0; 0,33)$ «накриває» середнє квадратичне відхилення σ з ймовірністю $\gamma = 0,99$.

3. Статистична перевірка гіпотез

3.1. Статистичні гіпотези та їх різновиди

При дослідженні багатьох випадкових явищ необхідно знати закон розподілу генеральної сукупності. Якщо закон розподілу — невідомий, але є міркування для припущення його певного вигляду **А**, наприклад, розподіл — рівномірний, показниковий або нормальний. Тоді висувають гіпотезу**:** ***генеральна сукупність розподілена за законом А*****.**

В цій гіпотезі мова йде про функцію розподілу невідомого розподілу. Проте іноді закон розподілу генеральної сукупності — відомий, але його параметри — невідомі. Якщо є міркування припустити, що невідомий параметр θ дорівнює певному значенню θ_0, то висувають гіпотезу: $\theta = \theta_0$. Ця гіпотеза вказує на припущену величину параметра відомого розподілу.

Статистичними називаються гіпотези про вигляд розподілу генеральної сукупності або про параметри відомих розподілів.

Наприклад, статистичними гіпотезами є такі:

- генеральна сукупність розподілена за нормальним законом;
- дисперсії двох випадкових величин, які розподілені за законом Пуассона, рівні між собою.

Разом з припущеною гіпотезою завжди можна розглядати протилежну до неї гіпотезу. Якщо припущена гіпотеза була відхилена, тоді справедлива протилежна гіпотеза.

Основною (нульовою) називається гіпотеза H_0, яка припускається.

Альтернативною (конкурентною) називається гіпотеза H_1**,** що суперечить основній.

Наприклад, якщо $H_0 : M(X) = 5,$ то $H_1 : M(X) \neq 5.$

3.2. Похибки перевірки гіпотез

Статистична гіпотеза, яка висунута, може бути правильною, або неправильною. Тому виникає необхідність її перевірки. Перевірка гіпотези здійснюється за даними вибірки, тобто статистичними методами. Тому перевірку гіпотези за даними вибірки називають *статистичною.*

При перевірці статистичної гіпотези за даними випадкової вибірки можна зробити хибний висновок. При цьому можуть виникнути похибки **першого та другого роду.**

Якщо буде відкинута правильна гіпотеза, то кажуть, що це **похибка першого роду.**

Якщо ж буде прийнята неправильна гіпотеза, то кажуть, що це **похибка другого роду.**

Ймовірність здійснити похибку першого роду позначається грецькою буквою α і називається **рівнем значущості.**

Найчастіше рівень значущості приймають рівним 0,05 або 0,01. Якщо рівень значущості за прийнятої гіпотези дорівнює 0,05, то це означає, що в п’яти випадках із 100 ми ризикуємо одержати похибку першого роду, тобто відкинути правильну гіпотезу.

3.3. Статистичний критерій перевірки основної гіпотези

Перевірку статистичної гіпотези можна здійснити лише з використанням даних вибірки. Для цього слід вибрати деяку випадкову статистичну характеристику (вибіркову функцію), точний або наближений розподіл якої відомий і з допомогою цієї характеристики перевірити основну гіпотезу.

Статистичним критерієм перевірки гіпотези називається випадкова величина *K*, розподіл якої (точний або наближений) відомий, і яка застосовується для перевірки основної гіпотези.

Якщо статистична характеристика розподілена за нормальним законом, то критерій позначають літерами **U** або **Z**. Якщо статистична характеристика розподілена за законом Фішера-Снедекора, то її позначають літерою **F**. У випадку розподілу статистичної характеристики за законом Стьюдента її позначають літерою **T**, а у випадку розподілу «хі-квадрат» χ^2.

Наприклад, для перевірки гіпотез про рівність дисперсії двох нормальних генеральних сукупностей за статистичну характеристику **K** вибирають відношення виправлених вибіркових дисперсій

$$F = \frac{S_1^2}{S_2^2}.$$

У різних дослідах дисперсія буде набувати різних, невідомих наперед, значень, тому ця величина випадкова. Вона розподілена за законом Фішера-Снедекора.

Емпіричним (спостереженим) значенням критерію називається значення відповідного критерію, обчислене за даними вибірки.

Наприклад, якщо за даними вибірок із двох нормальних генеральних сукупностей знайдено виправлені вибіркові дисперсії $S_1^2 = 18$, $S_2^2 = 6$, то спостережним значенням критерію є

$$F_{емп} = \frac{18}{6} = 3.$$

3.4. Критична область

Після вибору певного критерію множину всіх його значень поділяють на дві підмножини: одна з них містить значення критерію, при яких основна гіпотеза відхиляється, а друга при яких вона приймається.

Критичною областю називається сукупність значень критерію, при яких основна гіпотеза відхиляється.

Областю прийняття гіпотези називається сукупність значень критерію, при яких основна гіпотеза приймається.

В курсі статистичного аналізу критерій K це одновимірна випадкова величина, усі значення якої належать даному інтервалу, тому критична область та область прийняття рішень в цьому випадку також будуть інтервалами, тобто існують точки, які відокремлюють ці інтервали.

Критичними точками (межами) критерію *K* називаються точки k_{kp}, які відокремлюють критичну область від області прийняття гіпотези.

Розрізняють **однобічну (правобічну, лівобічну) та двобічну критичні області.**

Правобічною називається критична область, що визначається нерівністю

$$K > k_{kp}.$$

Лівобічною називається критична область, що визначається нерівністю:

$$K < k_{kp}.$$

Двобічною називається критична область, що визначається двома нерівностями: $K < k_1, \quad K > k_2, \quad k_1 < k_2$.

3.5. Знаходження критичних областей

Щоб знайти однобічну критичну область, потрібно знайти критичну точку k_{kp}. Для цього задають рівень значущості α і шукають k_{kp} із співвідношень:

- $P(K > k_{kp}) = \alpha$ для правобічної критичної області;
- $P(K < k_{kp}) = \alpha$ для лівобічної критичної області;
- у випадку двобічної критичної області повинна виконуватись тотожність:

$$P(K < k_1) + P(K > k_2) = \alpha .$$

Для кожного критерію є відповідні таблиці, які дозволяють знайти таку точку k_{kp}, яка виконує потрібну умову.

3.6. Порядок дій при перевірці статистичних гіпотез

Для перевірки правильності основної статистичної гіпотези H_0 необхідно:

- визначити гіпотезу H_1 альтернативну до H_0;
- вибрати статистичний критерій K;
- задати рівень значущості α;
- знайти за відповідною таблицею критичну область (критичну точку) для вибраного критерію.

До критичної області належать такі значення критерію K, при яких гіпотеза H_0 відхиляється на користь альтернативної гіпотези H_1.

3.7. Критерій узгодження Пірсона (χ^2)

Критерій узгодження Пірсона ефективно використовується для перевірки гіпотези про розподіл генеральної сукупності, тобто, про те, що розподіл випадкової величини X має певний функціональний вираз.

Нехай вибірка, яка репрезентує генеральну сукупність X, має дискретний статистичний розподіл

x_i	x_1	x_2	$\ldots$	x_m
n_i	n_1	n_2	...	n_m

або інтервальний статистичний розподіл

$[x_{i-1};x_i]$	$[x_1;x_2)$	$[x_2;x_3)$	$\ldots$	$[x_{m-1};x_m]$
n_i	n_1	n_2	...	n_m

об'єму n.

На підставі цих розподілів і за наперед заданим рівнем значущості α треба перевірити гіпотезу :

- H_0 : закон розподілу генеральної сукупності описується функцією розподілу $F(x)$ або густиною розподілу $f(x)$, якщо альтернативною є гіпотеза:
- H_1 : закон розподілу генеральної сукупності не описується функцією розподілу $F(x)$ або густиною розподілу $f(x)$.

За критерієм Пірсона для перевірки гіпотези H_0 вводиться статистика K :

$$K=\chi^2=\sum_{i=1}^{m}\frac{(n_i-n_i')^2}{n_i'}=n\sum_{i=1}^{m}\frac{\left(w_i-p_i\right)^2}{p_i}, \qquad (3.1)$$

де:

- m кількість груп в статистичному розподілі вибірки;
- n_i частота k -ї групи ;
- $n_i'=np_i$ теоретична частота;
- p_i ймовірність того, що значення ознаки генеральної сукупності X, яка досліджується, набуває значення x_i для дискретного статистичного розподілу

вибірки (1.5) і належить до i-го інтервалу для інтервального розподілу (1.7). p_i обчислюється з допомогою гіпотетичної функції розподілу $F(x)$ або густини розподілу $f(x)$.

В повних курсах математичної статистики доводиться, що статистика K при $n \to \infty$ прямує за розподілом до випадкової величини, яка розподілена за законом χ^2 з $k = m - s - 1$ ступенями вільності, де m - кількість інтервалів статистичного розподілу вибірки, а s - кількість параметрів, що входить до гіпотетичного розподілу F або f і які оцінюються на підставі даних спостереження. Так, наприклад, коли перевіряється, чи ознака X має розподіл Пуассона, то $k = m - 2$, бо розподіл Пуассона має єдиний невідомий параметр λ, який оцінюється за вибірковими даними (тобто шукається його точкова оцінка). Якщо ж перевіряється гіпотеза про нормальний закон розподілу ознаки генеральної сукупності, то $k = m - 3$, бо нормальний закон розподілу має два невідомі параметри α і σ, точкові оцінки для яких знаходяться за вибірковими даними.

Наступний крок – це знайти критичне значення χ^2_{kp} за заданим рівнем значущості α та числом ступенів вільності k. Критичні значення розподілу χ^2 задані таблицею додатка 5 при різних значеннях α і k.

Обчисливши значення

$$K_{емп} = \chi^2_{емп} = \sum_{i=1}^{m} \frac{(n_i - n_i')^2}{n_i'} \qquad (3.2)$$

і знайшовши k_{kp}, приходимо до таких висновків:

- якщо $K_{емп} < k_{kp}$, то гіпотезу H_0 приймаємо;
- якщо $K \geq k_{kp}$, то гіпотезу H_0 відхиляємо.

Критерій χ^2 застосовується при дотриманні таких вимог:

- статистичні дані мають бути незалежними;

- обсяг вибірки повинен бути достатньо великим (не меншим ніж 50 одиниць), а частота кожного інтервалу не меншою за 5. Якщо остання умова не виконується, то проводиться попереднє об’єднання нечисленних груп.

Отже, підсумовуючи все, що було сказане, опишемо схему перевірки статистичної гіпотези за критерієм Пірсона:

- записати вихідні статистичні дані;
- обчислити теоретичні частоти $n_i' = np_i$ для варіант вибірки. Для обчислення ймовірностей p_i використати формули: $p_i = P(x_{i-1} \le X < x_i) = F(x_i) - F(x_{i-1})$ або $p_i = \int\limits_{x_{i-1}}^{x_i} f(x)dx$, $i = 1, 2, ..., m$;
- обчислити $\chi^2_{емп}$ за формулою (3.2);
- знайти ступінь вільності за формулою $k = m - s - 1$;
- знайти з таблиці критичну точку $\chi^2_{кр} = k_{кр}$, яка відповідає заданому рівню значущості α та ступені вільності k;
- порівняти $\chi^2_{кр}$ та $\chi^2_{емп}$ і зробити висновок:

якщо $\chi^2_{емп} < \chi^2_{кр}$, то гіпотеза H_0 приймається;

якщо $\chi^2_{емп} \ge \chi^2_{кр}$, то гіпотеза H_0 відхиляється.

Приклад 3.1. Нехай X - середній час (у хв.) доїзду до місця пожежі. За даними 100 спостережень отримали такий інтервальний розподіл частот:

Таблиця 3.1

$[x_{i-1}; x_i)$	[13,5; 14,5)	[14,5; 15,5)	[15,5; 16,5)	[16,5; 17,5)	[17,5; 18,5)	[18,5; 19,5)	[19,5; 20,5]
n_i	6	10	18	28	20	12	6

За допомогою критерію Пірсона і для заданого рівня значущості $\alpha = 0,05$ перевірити гіпотезу H_0: випадкова величина, яка досліджується в задачі, має **нормальний закон розподілу** ймовірностей.

Розв'язання. Нормальний закон розподілу характеризується, як відомо, густиною розподілу

$$f(x,a,\sigma) = \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad x \in (-\infty; +\infty),$$

де $a = M(X)$ і $\sigma(X) = \sqrt{D(X)}$ параметри розподілу. Оскільки ці параметри невідомі, то знайдемо їх ***точкові оцінки*** на підставі вибіркових даних. Отримаємо:

$$a^* = \overline{x} = \frac{1}{100}\left(14\cdot 6 + 15\cdot 10 + 16\cdot 18 + 17\cdot 28 + 18\cdot 20 + 19\cdot 12 + 20\cdot 6\right) =$$

$$= \frac{1}{100}\cdot 1706 = 17,06.$$

$$D^* = \overline{D} = \frac{1}{99}\left(3,06^2\cdot 6 + 2,06^2\cdot 10 + 1,06^2\cdot 18 + 0,06^2\cdot 28 + 0,94^2\cdot 20 +\right.$$

$$\left. + 1,94^2\cdot 123 + 2,94^2\cdot 6\right) = \frac{233,64}{99} = 2,36.$$

$$\sigma^* = \overline{\sigma} = \sqrt{\overline{D}} = \sqrt{2,36} = 1,54.$$

Заокругливши, отримаємо $a = 17, \sigma = 1,5.$ Перевіримо гіпотезу про те, що випадкова величина X має нормальний закон розподілу ймовірностей із густиною:

$$f(x) = \frac{1}{\sqrt{2\pi}\cdot 1,5} e^{-\frac{(x-17)^2}{2\cdot 1,5^2}}.$$

Для цього обчислимо спочатку статистичні ймовірності w_k та теоретичні ймовірності p_k, з якою випадкова величина X потрапляє в інтервали $[x_{i-1}; x_i)$.

Статистичні ймовірності обчислюємо за формулою:

$$w_i = \frac{n_i}{n}. \qquad (3.3)$$

Отримуємо:

$$w_1 = 0,06, \quad w_2 = 0,10, \quad w_3 = 0,18, \quad w_4 = 0,28,$$
$$w_5 = 0,20, \quad w_6 = 0,12, \quad w_7 = 0,06.$$

Теоретичні ймовірності обчислюємо, використовуючи інтегральну теорему Лапласа :

$$p_k = P\left(x_{k-1} < X \le x_k\right) = \Phi\left(\frac{x_k - \overline{x}}{\overline{\sigma}}\right) - \Phi\left(\frac{x_{k-1} - \overline{x}}{\overline{\sigma}}\right). \qquad (3.4)$$

За формулою (3.4) отримуємо:

$$p_1 = \Phi\left(\frac{14,5-17}{1,5}\right) - \Phi\left(-\infty\right) = \Phi\left(-\frac{2,5}{1,5}\right) - \Phi\left(-\infty\right) = \Phi\left(\infty\right) - \Phi\left(\frac{2,5}{1,5}\right)$$
$$= \Phi\left(\infty\right) - \Phi\left(1,67\right) = 0,5 - 0,4525 = 0,0475 \approx 0,05;$$

$$p_2 = \Phi\left(\frac{15,5-17}{1,5}\right) - \Phi\left(\frac{14,5-17}{1,5}\right) = \Phi\left(-\frac{1,5}{1,5}\right) - \Phi\left(-\frac{2,5}{1,5}\right) =$$
$$= \Phi\left(\frac{2,5}{1,5}\right) - \Phi\left(\frac{1,5}{1,5}\right) = \Phi\left(1,67\right) - \Phi\left(1\right) = 0,4525 - 0,3413 \approx 0,11;$$
$$p_3 = \Phi\left(\frac{16,5-17}{1,5}\right) - \Phi\left(\frac{15,5-17}{1,5}\right) = \Phi\left(-\frac{0,5}{1,5}\right) - \Phi\left(-\frac{1,5}{1,5}\right) =$$
$$= \Phi\left(\frac{1,5}{1,5}\right) - \Phi\left(\frac{0,5}{1,5}\right) = \Phi\left(1\right) - \Phi\left(0,36\right) = 0,3413 - 0,1293 \approx 0,21;$$
$$p_4 = \Phi\left(\frac{17,5-17}{1,5}\right) - \Phi\left(\frac{16,5-17}{1,5}\right) = \Phi\left(\frac{0,5}{1,5}\right) - \Phi\left(-\frac{0,5}{1,5}\right) =$$
$$= 2\Phi\left(\frac{0,5}{1,5}\right) = 2\Phi\left(0,33\right) = 2 \cdot 0,1293 = 0,2586 \approx 0,26;$$

$$p_5 = \Phi\left(\frac{18,5-17}{1,5}\right) - \Phi\left(\frac{17,5-17}{1,5}\right) = \Phi\left(\frac{1,5}{1,5}\right) - \Phi\left(\frac{0,5}{1,5}\right) =$$

$$\Phi(1) - \Phi(0,33) = 0,3413 - 0,1293 \approx 0,21;$$

$$p_6 = \Phi\left(\frac{19,5-17}{1,5}\right) - \Phi\left(\frac{18,5-17}{1,5}\right) = \Phi\left(\frac{2,5}{1,5}\right) - \Phi\left(\frac{1,5}{1,5}\right) =$$

$$= \Phi(1,67) - \Phi(1) = 0,4525 - 0,3413 = 0,1112 \approx 0,11;$$

$$p_7 = \Phi(+\infty) - \Phi\left(\frac{19,5-17}{1,5}\right) = \Phi(+\infty) - \Phi\left(\frac{2,5}{1,5}\right) =$$

$$= \Phi(+\infty) - \Phi(1,67) = 0,5 - 0,45625 \approx 0,05.$$

Зауважимо, що для обчислення першої та останньої ймовірностей p_1 і p_m у формулах (3.4) покладають відповідно $x_0 = -\infty$ і $x_m = +\infty$. Тоді $\sum_{i=1}^{m} p_i = 1$.

Обчислюємо $K_{емп} = \chi^2_{емп}$ за формулою:

$$K_{емп} = n\sum_{i=1}^{m} \frac{(w_i - p_i)^2}{p_i} \tag{3.5}$$

Маємо

$$K_{емп} = \chi^2_{емп} = 100 \times$$

$$\times\left(\frac{0,01^2}{0,05} + \frac{0,01^2}{0,11} + \frac{0,03^2}{0,21} + \frac{0,02^2}{0,26} + \frac{0,01^2}{0,21} + \frac{0,01^2}{0,11} + \frac{0,01^2}{0,05}\right) =$$

$$= 100(0,002 + 0,0009 + 0,004 + 0,0015 + 0,0005 + 0,0009 + 0,002) =$$

$$= 100 \cdot 0,0118 = 2,59.$$

Результати обчислень доцільно подати у вигляді таблиці:

Табл. 3.2.

x_{i-1}	x_i	n_i	$z_{i-1} = \dfrac{x_{i-1} - \overline{x}}{\overline{\sigma}}$	$z_i = \dfrac{x_i - \overline{x}}{\overline{\sigma}}$	$\Phi(z_{i-1})$	$\Phi(z_i)$	$n_i' = n \cdot p_i$
$13,5$	$14,5$	6	$-\infty$	$-1,67$	$-0,49$	$-0,45$	4
$14,5$	$15,5$	10	$-1,67$	$-1,02$	$-0,45$	$-0,35$	11
$15,5$	$16,5$	18	$-1,02$	$-0,36$	$-0,35$	$-0,14$	20
$16,5$	$17,5$	28	$-0,36$	$0,29$	$-0,14$	$0,11$	25
$17,5$	$18,5$	20	$0,29$	$0,94$	$0,11$	$0,33$	21
$18,5$	$19,5$	12	$0,94$	$1,59$	$0,33$	$0,44$	12
$19,5$	$20,5$	6	$1,59$	$+\infty$	$0,44$	$0,5$	4

Для заданого рівня значущості $\alpha = 0,05$ і числа ступенів вільності $k = 7 - 2 - 1 = 4$ із таблиці 5 додатка визначимо: $k_{kp} = \chi^2_{kp} = 9,5$.

Оскільки $K_{емп} = 2,59 < k_{kp} = 9,5$, то гіпотеза H_0: **випадкова величина має нормальний розподіл – приймається.**

Виконаємо це завдання за допомогою Exel

1. Вводимо дані про час доїзду у дискретному вигляді, замінивши інтервали їх серединами: **А1:А6** – 14; **А7:А16** – 15; **А17:А34** – 16; **А35:А62** – 17; **А63:А82** – 18; **А83:А94** – 19; **А95:А100** – 20. Введення даних здійснюємо за допомогою маркера заповнення. Дамо ім’я робочому листу «Дані».

2. У новому робочому листі, який назвемо **«Результати»,** вводимо заголовки стовпців таблиці, в якій будуть проводитись обчислення: **A5** – x_{i-1} , **B5** – x_i , **C5** – n_i , **D5** – z_{i-1}, **E5** – z_i , **F5** – $\Phi\left(z_{i-1}\right)$, **G5** – $\Phi\left(z_i\right)$, **H5** – n_i' , **I5** – середнє, **J5** – σ. Заповнюємо комірки **A6:A12** числами, що відповідають початкам інтервалів, комірки **B6:B12** – числами, що є кінцями інтервалів. У комірки **C6:C12** вводимо відповідні частоти.
3. Переходимо в робочий лист **«Дані»**, в якому обчислюємо середнє значення (**СРЗНАЧ)** і середнє квадратичне відхилення (**СТАНДОТКЛОН**). Результати обчислень копіюємо в робочий лист **«Результати»** в комірки **I6** та **J6** відповідно.
4. У робочому листі **«Чернетка»** будуємо гістограму і копіюємо її в робочий лист **«Результати»**, розмістивши справа від таблиць.
5. Вводимо в комірку **A13** текст **«Всього виїздів»** і в комірці **C13** за допомогою формули **= СУММ(C6:C12)** обчислюємо обсяг вибірки.
6. Редагуємо вміст комірок **A13:C13**, виконавши такі операції: виділяємо комірки, копіюємо їх вміст, викликаємо команду **Правка → Спеціальна вставка** і ставимо позначку “**•**” на **«Значення»**.
7. Вводимо у комірку **D6** формулу **= (A6 – I$6)/J$6**, яку за допомогою маркера заповнення копіюємо в комірки **D7:D12**.
8. Вводимо у комірку **E6** формулу **= (B6 – I$6)/J$6**. Копіюємо за допомогою маркера заповнення цю формулу в комірки **E7:E12**.
9. Вводимо у комірку **F6** формулу **= НОРМСТРАСП(D6) – 0,5** і копіюємо її в комірки **F7:F12**.

10. Вводимо у комірку **G6** формулу **=НОРМСТРАСП(E6) – 0,5** і копіюємо в комірки **G7:G12**.
11. Вводимо в комірку **H6** формулу **= C$13*(G6 – F6)** і копіюємо її в комірки **H7:H12**. Округлюємо отримані теоретичні частоти до цілих чисел.
12. Виділимо вміст стовпців **C5:C12**, **H5:H12** скопіюємо його та за допомогою команд **Правка → Спеціальна вставка → Значення** внесемо у комірки **A22:A29** та **B22:B29**. Введемо в комірку **C22** текст **«хі-квадрат»**, а в комірку **A30** − **«Сума»**.
13. Внесемо в **C23** формулу **= ((A23 – B23)^2)/B23** і копіюємо її у комірки **C24:C29**.
14. Обчислимо в комірці **C30** суму **=СУММ(C23:C29).**
15. Вводимо у комірки **A32** – **Точність**, **A33** – **ступені вільності**, **A34** – **критична точка**, **A35** – **Розподіл**, **C32** – **0,01**, **D32** – **0,05**, **C33** – **4**, **D33** – **4.**
16. Обчислюємо критичну точку критерію хі-квадрат в комірці **C34** за формулою **= ХИ2ОБР(C32; C33)** і копіюємо її вміст у комірку **D34**.
17. У комірці **C35** вводимо формулу **=ЕСЛИ(C$30>C34; "Відмінний від нормального норм";"Норм")** і аналогічні дії виконуємо в комірці **D35**.

Результати обчислень видно на рис. 3.1 та рис. 3.2

	A	B	C	D	E	F	G	H	I	J
4										
5	X_{i-1}	X_i	n_i	Z_{i-1}	Z_i	$Ф(Z_{i-1})$	$Ф(Z_i)$	n'_i	середнє	сігма
6	13,5	14,5	6	-2,32	-1,67	-0,49	-0,45	4	17,06	1,54
7	14,5	15,5	10	-1,67	-1,02	-0,45	-0,35	11		
8	15,5	16,5	18	-1,02	-0,36	-0,35	-0,14	20		
9	16,5	17,5	28	-0,36	0,29	-0,14	0,11	25		
10	17,5	18,5	20	0,29	0,94	0,11	0,33	21		
11	18,5	19,5	12	0,94	1,59	0,33	0,44	12		
12	19,5	20,5	6	1,59	2,24	0,44	0,49	4		
13	Всього виїздів		100							
14										

Рис. 3.1

21				
22	ni	n'i	хі-квадрат	
23	6	4	1,34	
24	10	11	0,05	
25	18	20	0,26	
26	28	25	0,25	
27	20	21	0,08	
28	12	12	0,00	
29	6	4	0,62	
30	Сума		2,59	
31				
32	Точність		0,01	0,05
33	Ступені свободи		4	4
34	критична точка		13,28	9,49
35	Розподіл		нормальний	

Рис. 3.2

Порівнюючи рис. 3.1 та табл 3.2, бачимо, що в Exel немає можливості і потреби вводити $x_0 = -\infty$ і $x_m = +\infty$.

Приклад 3.2. Аналіз декларацій 748 фізичних осіб про сукупний річний дохід подано в таблиці:

x_i	0	1	2	3	4	5
n_i	424	233	68	20	1	1

де x_k відсоток незаявленого сукупного доходу, n_k кількість осіб, які не заявили $x_k\%$ сукупного доходу. Перевірити гіпотезу про те, що відсоток незаявленого сукупного доходу однієї особи під час оформлення нею декларації має **розподіл Пуассона**, якщо рівень значущості $\alpha = 0,01$.

Розв'язання. Нехай випадкова величина X середній відсоток незаявленого сукупного річного доходу однієї фізичної особи під час оформлення декларації. За умовою задачі, потрібно перевірити гіпотезу:

$$H_0 = P(X = i) = \frac{\lambda^i}{i!}e^{-\lambda}, \quad i = 0,1,2,3,...,$$

де $\lambda = M\left(X\right)$ невідомий параметр,

а саме, що розподіл відсотків незаявленого сукупного доходу підлягає розподілу Пуассона.

Точковою оцінкою невідомого параметра λ, як відомо, наприклад з $\left[4,5\right]$, є вибіркове середнє $\overline{x}$. Обчислимо його.

$$\lambda^* = \overline{x} = \frac{1}{747}(424\cdot 0 + 233\cdot 1 + 68\cdot 2 + 20\cdot 3 + 1\cdot 4 + 1\cdot 5) =$$

$$= \frac{1}{747}\cdot 438 = 0,584 \approx 0,6.$$

Обчислимо теоретичні ймовірності p_i за формулою Пуассона, покладаючи в ній $\lambda = 0,6$:

$$p_i = \frac{0,6^i}{i!}e^{-0,6}, \quad i = 0,1,2,3,...,$$

Дві останні варіанти випадкової величини X в статистичному розподілі вибірки мають частоти менші ніж п’ять. Через те об’єднаємо їх з варіантою $X = 3$. Зауважимо також, що остання теоретична ймовірність p_3 визначатиметься як доповнення до одиниці. Отже маємо:

$$p_0 = \frac{0,6^0}{0!}e^{-0,6} = \frac{1}{e^{0,6}} = 0,56;$$

$$p_1 = \frac{0,6^1}{1!}e^{-0,6} = \frac{0,6}{e^{0,6}} = 0,33;$$

$$p_2 = \frac{0,6^2}{2!}e^{-0,6} = \frac{0,36}{e^{0,6}} = 0,09;$$

$$p_3 = 1 - 0,549 - 0,329 - 0,099 = 0,02.$$

Знайдемо теоретичні частоти $n'_k = np_k$:

$$n'_0 = 747\cdot 0,56 = 417,27; \quad n'_1 = 747\cdot 0,33 = 242,99;$$

$$n'_2 = 747\cdot 0,09 = 70,75; \quad n'_3 = 747\cdot 0,02 = 13,73.$$

Обчислимо

$$K_{емп} = \sum_{i=0}^{3} \frac{\left(n_i - n_i'\right)^2}{n_i'} = \frac{\left(424 - 417,27\right)^2}{417,27} + \frac{\left(233 - 242,99\right)^2}{242,99} +$$

$$+ \frac{\left(68 - 70,75\right)^2}{70,75} + \frac{\left(22 - 13,73\right)^2}{17,25} = 5,6.$$

З таблиці додатка 5 критичних точок розподілу χ^2 для $\alpha = 0,01$ і числа ступенів вільності $k = 4 - 1 - 1 = 2$ знаходимо критичну точку $k_{кр} = 9,2$.

Оскільки $K_{емп} = 5,6 < k_{кр} = 9,2$, то сформульована гіпотеза H_0 про те, що описана в умові задачі випадкова величина X розподілена за законом Пуассона за рівня значущості $\alpha = 0,01$ **приймається.**

Як і в попередньому прикладі, всі необхідні для перевірки основної гіпотези викладки можна подати у таблиці:

x_i	n_i	p_i	$n_i' = np_i$	$n_i - n_i'$	$\frac{\left(n_i - n_i'\right)^2}{n_i'}$
0	424	0,56	417,27	6,73	0,11
1	233	0,33	242,99	-9,99	0,41
2	68	0,09	70,75	-2,75	0,11
3 і більше	22	0,02	13,73	8,27	4,98
Сума	750	1	750	-	$\chi^2_{emn} = 5,6$

Реалізуємо цю задачу засобами пакету Exel.

1. Формуємо таблицю для проведення обчислень. Вводимо назви стовпців: **A1** – “xi”, **B1** – “ni”, **C1** – “pi”, **D1** – “ni’”, **E1** – “ni-ni’”, **F1** – “ хі-квадрат”.
2. Вводимо числові дані в комірки **A2:B5** (рис. 3.3)

	A	B
1	xi	ni
2	0	424
3	1	233
4	2	68
5	3	22

Рис. 3.3

3. Обчислюємо обсяг вибірки. В комірку **A6** вводимо назву «всього», а в комірку **B6** вводимо формулу **=СУММ(B2:B5)**.
4. В комірці **B7** обчислюємо вибіркове середнє за формулою **=(A2*B2+A3*B3+A4*B4+A5*22)/B6**.
5. В комірках **C2:C5** обчислюємо імовірності pi. Для цього в комірку **C2** вводимо формулу

   = (B$7^A2)*EXP(-B$7)/ФАКТР(A2)

   і розповсюджуємо її вміст на решту комірок стовпця.
6. В комірках **D2:D5** обчислюємо емпіричні частоти. Для цього в комірку **D2** вводимо формулу **=B$6*C2** і поширюємо її вміст на решту комірок стовпця.
7. В комірку **E2** вводимо формулу **=B2-D2** і поширюємо її в комірки **E3:E5**.
8. В комірку F2 вводимо формулу **=(E2^2)/D2** і розповсюджуємо її в комірки **F3:F5**.
9. В комірці **F6** обчислюємо емпіричне значення критерію за формулою **=СУММ(F2:F5)**.

Результати роботи видно на рис. 3.4

	A	B	C	D	E	F	G
1	xi	ni	pi	ni'	ni-ni'	xi-квадрат	
2	0	424	0,56	417,27	6,73	0,11	
3	1	233	0,33	242,99	-9,99	0,41	
4	2	68	0,09	70,75	-2,75	0,11	
5	3	22	0,02	13,73	8,27	4,98	
6	всього	747				5,60	
7	середнє	0,58					
8							

Рис. 3.4.

10. Знаходимо критичну точку розподілу хі-квадрат для рівня значущості 0,01 за допомогою функції **=ХИ2ОБР(0,01;2)**.
11. Порівнюємо емпіричне значення критерію з критичною точкою і робимо висновки за допомогою логічної функції
=ЕСЛИ (F6<C9;"розподіл Пуассона";"розподіл відмінний від Пуассона")

3.8. Перевірка гіпотез про параметри нормального розподілу

Гіпотези такого типу або **параметричні гіпотези** виникають, наприклад, після застосування критеріїв згоди, тобто, коли встановлено вигляд функції розподілу випадкової величини, але самі параметри, від яких залежить ця функція, невідомі й потребують уточнення.

Перевіряють **параметричні статистичні гіпотези** за загальними правилами перевірки гіпотез.

Припустимо, що генеральна сукупність описується випадковою величиною, що має **нормальний закон розподілу.** Оскільки цей закон розподілу визначається двома параметрами a (математичне сподівання) і σ (середнє квадратичне відхилення), то гіпотези формулюються відносно цих параметрів.

3.9.Гіпотези про математичні сподівання

3.9.1. Перевірка гіпотези про значення математичного сподівання за відомої дисперсії

Нехай генеральна сукупність розподілена за нормальним законом з невідомим математичним сподіванням $a = M\left(X\right)$, але відомою дисперсією $\sigma^2 = D\left(X\right)$

Варіант І. На основі вибірки потрібно перевірити нульову гіпотезу $H_0 : a = a_0$ про рівність математичного сподівання a певному числу a_0. При цьому передбачаються відомими такі величини:

- вибірка обсягу n;
- середнє квадратичне відхилення σ;
- гіпотетичне значення математичного сподівання a_0;
- рівень значущості $\alpha(0 < \alpha < 1)$.

Критерій перевірки цієї гіпотези

$$Z = \frac{\left(\overline{x} - a_0\right)\sqrt{n}}{\sigma}$$

має нормальний розподіл з параметрами $(0;1)$.

Перевірку гіпотези $H_0 : a = a_0$ **за альтернативної гіпотези** $H_1 : a \neq a_0$ здійснюємо за таким правилом:

1) обчислюємо емпіричне значення критерію за формулою:

$$Z_{емр} = \frac{\left(\overline{x} - a_0\right)\sqrt{n}}{\sigma}; \tag{3.6}$$

2) знаходимо за таблицею значень функції Лапласа критичне значення z_{kp} з рівняння:

$$\Phi\left(z_{kp}\right) = \frac{1-\alpha}{2}; \tag{3.7}$$

3) робимо висновок про висунуту гіпотезу:

- якщо $\left|Z_{емп}\right| < z_{kp,}$ то гіпотезу H_0 приймаємо;
- якщо $\left|Z_{емр}\right| \geq z_{kp,}$ то гіпотезу відхиляємо на користь альтернативної гіпотези H_1.

Перевірку гіпотези $H_0: a = a_0$ **за альтернативної** $H_1: a > a_0$ **або** $H_1: a < a_0$ проводимо за попередньою схемою з такими змінами:

1) замість рівняння (3.7) для знаходження критичного значення z_{kp} використовуємо рівняння:

$$\Phi\left(z_{kp}\right) = \frac{1-2\alpha}{2};$$

2) робимо висновки відносно висунутої гіпотези H_0:

- якщо $Z_{емп} < z_{kp}$, то гіпотезу H_0 приймаємо; якщо $Z_{емр} \geq z_{kp}$, то гіпотезу H_0 відхиляємо на користь альтернативної гіпотези $H_1: a > a_0$;
- якщо $Z_{емп} > -z_{kp}$, то гіпотезу H_0 приймаємо; якщо $Z_{емп} < -z_{kp}$, то гіпотезу H_0 відхиляємо на користь альтернативної гіпотези $H_1: a < a_0$.

Приклад 3.3. Нехай генеральна сукупність розподілена за нормальним законом з відомим середнім квадратичним відхиленням $\sigma = 5$. З цієї генеральної сукупності одержано вибірку обсягу $n = 50$ і за нею знайдено вибіркове середнє $\overline{x} = 27,7$. У задачі потрібно для рівня значущості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0: a = a_0 = 29$ за альтернативної гіпотези :

- $H_1: a \neq a_0$;
- $H_1: a < a_0$.

Розв'язання. Обчислимо емпіричне значення критерію за формулою (3.6):

$$Z_{емп} = \frac{\left(27,7-29\right)\sqrt{50}}{5} = -1,838.$$

Розглянемо наведені в умові задачі два випадки:

- для альтернативної гіпотези $H_1 : a \neq a_0$ знаходимо z_{kp} за формулою (3.7) з таблиці додатка 2 :

$$\Phi\left(z_{kp}\right) = \frac{1-0,05}{2} = 0,475.$$

Отже z_{kp} =1,96. Оскільки $\left|Z_{емп}\right| = 1,838 < 1,96 = z_{kp}$, то гіпотезу H_0 приймаємо;

- для альтернативної гіпотези $H_1 : a < a_0$ знаходимо z_{kp} за формулою (22) з таблиці додатка 2

$$\Phi\left(z_{kp}\right) = \frac{1-2\cdot 0,05}{2} = 0,45.$$

Отже z_{kp} =1,65. Оскільки $Z_{емп} = -1,838 < -1,65 = -z_{kp}$, то гіпотезу H_0 відхиляємо на користь альтернативної гіпотези H_1.

Варіант ІІ. На основі вибірки потрібно перевірити нульову гіпотезу $H_0 : a = a_0$ за умови, що генеральна сукупність розподілена за нормальним законом з невідомими математичним сподіванням $a = M\left(X\right)$ і дисперсією $D\left(X\right) = \sigma^2$.

При цьому припускаємо, що відомими є лише такі величини:

- дані вибірки обсягу n;
- гіпотетичне значення математичного сподівання a_0;
- рівень значущості $\alpha (0 < \alpha < 1)$.

Критерієм перевірки гіпотези H_0 в цьому випадку використаємо розподіл Стьюдента з числом $k = n - 1$ ступенів вільності.

$$T = \frac{\left(\overline{x} - a_0\right)\sqrt{n}}{\overline{\sigma}}, \qquad (3.8)$$

де $\overline{x}$ вибіркове середнє, а $\overline{\sigma}$ виправлене вибіркове середнє квадратичне відхилення.

Перевірку гіпотези $H_0 : a = a_0$ **за конкуруючої гіпотези** $H_1 : a \neq a_0$ здійснюємо за таким правилом:

1) обчислюємо емпіричне значення критерію за формулою:

$$T_{емп} = \frac{\left(\overline{x} - a_0\right)\sqrt{n}}{\overline{\sigma}}; \qquad (3.9)$$

2) з таблиці критичних точок розподілу Стьюдента (див. додаток) за заданим рівнем значущості α і числом ступенів вільності $k = n - 1$ знаходимо критичну точку $t_{кр} = t_{кр}(\alpha, k)$;

3) робимо висновок про висунуту гіпотезу:

- якщо $\left|T_{емп}\right| < t_{кр,}$ то гіпотезу H_0 приймаємо;
- якщо $\left|T_{емп}\right| \geq t_{кр,}$ то гіпотезу відхиляємо на користь альтернативної гіпотези H_1.

Перевірку гіпотези $H_0 : a = a_0$ **за альтернативної** $H_1 : a > a_0$ **або** $H_1 : a < a_0$ проводимо за попередньою схемою з такими змінами:

1) з таблиці критичних точок розподілу Стьюдента за даним рівнем значущості α і число ступенів вільності $k = n - 1$ знаходимо критичну точку $t_{кр} = t_{кр}(\alpha, k)$;

2) робимо висновки відносно висунутої гіпотези H_0 :

- якщо $T_{емп} < t_{кр}$, то гіпотезу H_0 приймаємо; якщо $T_{емп} \geq t_{кр}$, то гіпотезу H_0 відхиляємо на користь альтернативної гіпотези $H_1 : a > a_0$;
- якщо $T_{емп} > -t_{кр}$, то гіпотезу H_0 приймаємо; якщо $T_{емп} < -t_{кр}$, то гіпотезу H_0 відхиляємо на користь альтернативної гіпотези $H_1 : a < a_0$.

Приклад 3.4. Для вибірки обсягу $n = 16$ значень випадкової величини X, розподіленої за нормальним законом знайдено вибіркове середнє $\overline{x} = 118,2$ та виправлене середнє квадратичне відхилення $\overline{\sigma} = 3,6$. В задачі потрібно для рівня значущості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0 : a = a_0 = 120$ за наявності конкуруючої гіпотези :

- $H_1 : a \neq a_0$;
- $H_1 : a < a_0$.

Розв'язання. Обчислимо емпіричне значення критерію за формулою (3.9):

$$T_{емп} = \frac{(118,2 - 120)\sqrt{50}}{3,6} = -2.$$

Розглянемо наведені в умові задачі два випадки:

- для альтернативної гіпотези $H_1 : a \neq a_0$ за таблицею додатка 2 для числа ступенів вільності $k = n - 1 = 15$ і рівня значущості $\alpha = 0,05$ знаходимо $t_{кр} = t_{кр}(0,05,15) = 2,13$. Оскільки $|T_{емп}| = 2 < 2,13 = t_{кр}$, то гіпотезу H_0 приймаємо;
- для альтернативної гіпотези $H_1 : a < a_0$ за таблицею додатка 2 для числа ступенів вільності $k = 15$ і рівня значущості $\alpha = 0,05$ знаходимо

$t_{kp} = t_{kp}(0,05,15) = 1,75$. Оскільки $|T_{емп}| = 2 > 1,75 = t_{kp}$, то гіпотезу H_0 відхиляємо на користь гіпотези H_1.

3.9.2. Перевірка гіпотези про рівність дисперсій двох незалежних випадкових величин

Задача про перевірку гіпотези про рівність двох дисперсій виникає досить часто. Наприклад, при аналізі стабільності виробничого процесу до і після введення нової технології (коливання у випуску продукції вимірюється за допомогою середнього квадратичного відхилення), при вивченні якості вимірювальних приладів (порівняння дисперсій показників окремих приладів), при аналізі ступеня однорідності двох сукупностей щодо деякої ознаки (кваліфікації робітників, стажу персоналу і т.д.).

Нехай випадкові величини X, Y, що характеризують дві статистичні сукупності, незалежні, нормально розподілені з невідомими дисперсіями $D(X) = \sigma_x^2$, $D(Y) = \sigma_y^2$ відповідно.

Перевірятимемо гіпотезу $H_0 : \sigma_x^2 = \sigma_y^2$ (про рівність дисперсій випадкових величин X і Y).

Вважаємо, що відомими є такі величини:

- вибірки $(x_1, x_2, ..., x_n)$ і $(y_1, xy, ..., y_m)$ обсягів n і m для випадкових величин X і Y відповідно;
- рівень значущості α.

Критерій перевірки гіпотези H_0 базується на співставленні виправлених дисперсій $\tilde{D}_x$ і $\tilde{D}_y$, обчислених за даними вибірок При даних припущеннях критерієм перевірки гіпотези H_0 є випадкова величина:

$$F = \frac{\tilde{D}_x}{\tilde{D}_y}, \text{ де } \tilde{D}_x \geq \tilde{D}_y, \qquad (3.10)$$

яка розподілена за законом Фішера-Снедекора з $k_1 = n-1$ і $k_2 = m-1$ ступенями вільності.

Перевірку нульової гіпотези $H_0 : \sigma_x^2 = \sigma_y^2$ **за конкуруючої** $H_1 : \sigma_x^2 > \sigma_y^2$ **здійснюємо за схемою:**

1. Знаходимо емпіричне значення критерію за формулою (3.10).

2. За таблицею критичних точок розподілу Фішера-Снедекора (додаток 6) для заданого рівня значущості α і ступенів вільності $k_1 = n-1$ і $k_2 = m-1$ знаходимо критичну точку правосторонньої критичної області $f_{kp} = f_{kp}(\alpha, k_1, k_2)$.

3. Робимо висновок щодо прийняття гіпотези H_0:

- якщо $F_{емр} < f_{kp}$, то гіпотезу H_0 приймаємо;
- якщо $F_{емр} \geq f_{kp}$, то гіпотезу H_0 відхиляємо на користь альтернативної гіпотези H_1.

У випадку, коли $\tilde{D}_x < \tilde{D}_y$, критерій згоди задається формулою:

$$F = \frac{\tilde{D}_y}{\tilde{D}_x},$$

де $k_1 = m-1$ і $k_2 = n-1$.

Зауваження. Якщо нульова гіпотеза $H_0 : \sigma_x^2 = \sigma_y^2$, альтернативна гіпотеза $H_1 : \sigma_x^2 \neq \sigma_y^2$, то перевірку гіпотези H_0 здійснюємо за попередньою схемою, в якій змінюється тільки методика знаходження критичного значення f_{kp}, а саме: з таблиці критичних точок розподілу Фішера-Снедекора критичну точку $f_{kp} = f_{kp}(\frac{\alpha}{2}, k_1, k_2)$ визначаємо за рівнем значущості $\frac{\alpha}{2}$ та числом ступенів вільності $k_1 = n-1$ і $k_2 = m-1$.

Приклад 3.5. За даними статистичними розподілами вибірок випадкових величин X і Y,

x_i	1,2	2,2	3,2	4,2	5,2
n_i'	1	2	4	2	3

y_i	0,8	1,6	2,4	3,2	4
n_i''	2	6	1	1	2

які є незалежними та мають нормальні закони розподілу при рівні значущості $\alpha = 0,01$ **перевірити правильність гіпотези** $H_0 : D_x = D_y$ **за альтернативної гіпотези** $H_1 : D_x > D_y$.

Розв'язання. Обчислимо $\tilde{D}_x$ і $\tilde{D}_y$:

$$\overline{x} = \frac{1,2\cdot 1 + 2,2\cdot 2 + 3,2\cdot 4 + 4,2\cdot 2 + 5,2\cdot 3}{12} = \frac{42,4}{12} = 3,53;$$

$$\overline{D}_x = \frac{1,2^2\cdot 1 + 2,2^2\cdot 2 + 3,2^2\cdot 4 + 4,2^2\cdot 2 + 5,2^2\cdot 3}{12} - (3,53)^2 = 1,5791;$$

$$\tilde{D}_x = \frac{12}{11}\cdot \overline{D}_x = \frac{12}{11}\cdot 1,5191 = 1,723;$$

$$\overline{y} = \frac{0,8\cdot 2 + 1,6\cdot 6 + 2,4\cdot 1 + 3,2\cdot 1 + 4\cdot 2}{12} = 2,067;$$

$$\overline{D}_y = \frac{0,8^2\cdot 2 + 1,6^2\cdot 6 + 2,4^2\cdot 1 + 3,2^2\cdot 1 + 4^2\cdot 2}{12} - (2,067)^2 = 1,1175;$$

$$\tilde{D}_y = \frac{12}{11}\cdot \overline{D}_y = \frac{12}{11}\cdot 1,1175 = 1,22.$$

Обчислюємо емпіричне значення критерію Фішера-Снедекора за формулою (3.10):

$$F_{емп} = \frac{\tilde{D}_x}{\tilde{D}_y} = \frac{1,723}{1,22} = 1,41.$$

З таблиці додатка 6 критичних точок розподілу Фішера-Снедекора для заданого рівня значущості $\alpha = 0,01$ знаходимо критичну точку правосторонньої критичної області $f_{kp}(\alpha = 0,01, k_1 = 11, k_2 = 11) = 4,4.$

Оскільки $F_{емп} = 1,41 < f_{kp} = 4,4,$ то нульова гіпотеза H_0 приймається.

4. Однофакторний та двофакторний дисперсійний аналіз

Нехай генеральні сукупності $X_1, X_2, ..., X_m$ розподілені за нормальним законом та мають однакові, хоч і невідомі, дисперсії. Математичні сподівання також невідомі, однак можуть бути різні. Потрібно для заданого рівня значущості за вибірковими середніми перевірити нульову гіпотезу $H_0 : M\left(X_1\right) = M\left(X_2\right) = ... = M\left(X_m\right)$ про рівність математичних сподівань. Здавалося б, що для порівняння кількох вибіркових середніх можна порівнювати їх попарно. Однак із зростанням їх кількості зростає й найбільша різниця між ними: вибіркове середнє нової вибірки може виявитись більшим від найбільшого або меншим від найменшого з вибіркових середніх, отриманих до нового досліду. Тому для порівняння кількох вибіркових середніх використовують інший метод, який базується на порівнянні дисперсій і тому називається **дисперсійним аналізом.**

На практиці дисперсійний аналіз застосовують для того, щоб встановити, чи суттєво впливає деякий фактор F, який має m рівнів $F_1, F_2, ..., F_m$ на випадкову величину X, яку ми досліджуємо. Наприклад, якщо потрібно з'ясувати, який вид піноутворювача найефективніше застосовувати, то фактор це піноутворювач, а його рівні види піноутворювачів.

Основна ідея дисперсійного аналізу полягає в порівнянні «факторної дисперсії», яка породжується дією фактора і «залишковою дисперсією», зумовленою випадковими причинами. Якщо відмінність між цими дисперсіями суттєва, то фактор має суттєвий вплив на випадкову величину X; в цьому випадку групові вибіркові середні також значно відрізняються.

Іноді дисперсійний аналіз застосовується для встановлення однорідності кількох вибірок. Дисперсії цих вибірок рівні за припущенням; якщо дисперсійний аналіз покаже, що й математичні сподівання – однакові, то в цьому сенсі вибірки є однорідними. Однорідні вибірки можна об'єднати в одну і таким чином отримати повнішу інформацію про неї, а, отже й надійніші висновки.

4.1. Однофакторний дисперсійний аналіз

Нехай потрібно дослідити вплив деякого фактора F на випадкову величину X, яка розподілена за нормальним законом і на основі цього дослідження виявити, наскільки цей вплив істотний. Для цього розглядаються різні рівні $F_1, F_2, ..., F_m$ дії фактора і для кожного рівня складають вибірку.

Для прозорості схеми застосування однофакторного дисперсійного аналізу вважатимемо, що вибірки на кожному рівні F_i мають однаковий обсяг n, а загальний обсяг вибірки дорівнює $m \times n$.

Опишемо алгоритм дисперсійного аналізу.

Позначимо середнє значення вибірки, отриманої на рівні F_i, через $\overline{x}_i$ $i = 1, 2, ..., m$. Якщо фактор F не впливає на випадкову величину X, то всі середні значення повинні бути майже однаковими, а у випадку, коли вплив фактора F — істотний, то середні значення значно відрізнятимуться один від одного. Це означає, що потрібно перевірити гіпотезу:

$$H_0 : \overline{x}_1 = \overline{x}_2 = ... = \overline{x}_m$$

За альтернативної гіпотези:

H_1 : умова $\bar{x}_1 = \bar{x}_2 = ... = \bar{x}_m$ не виконується.

Нульову гіпотезу H_0 перевірятимемо за допомогою критерію Фішера-Снедекора, в основі якого є подання загальної дисперсії σ^2 у вигляді суми міжгрупової σ_F та «залишкової» σ_Z^2 дисперсій:

$$\sigma^2 = \sigma_F^2 + \sigma_Z^2. \tag{4.1}$$

Помноживши рівність (4.1) на обсяг вибірки, отримаємо формулу для суми квадратів відхилень (ефектів):

$$Q = Q_F + Q_Z, \tag{4.2}$$

де Q_F розсіяння, спричинене впливом фактора F, а Q_Z розсіяння, спричинене дією інших залишкових факторів.

Величини Q_F і Q_Z можна записати конструктивно, вийшовши з представлення суми квадратів відхилень Q, а саме:

$$Q = \sum_{i=1}^{m}\sum_{j=1`}^{n}\left(x_{ij} - \bar{x}\right)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\left(x_{ij} - \bar{x}_i\right) + \left(\bar{x}_i - \bar{x}\right)\right)^2 =$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2 + n\sum_{i=1}^{m}\left(\bar{x}_i - \bar{x}\right)^2 = Q_F + Q_Z,$$

де

$$Q_F = n\sum_{i=1}^{m}\left(\bar{x}_i - \bar{x}\right)^2, \quad Q_Z = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_{ij} - \bar{x}_i\right)^2. \tag{4.3}$$

Зауважимо, що в останніх формулах використані такі позначення:

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \quad \text{середнє значення } i \text{ - го рівня фактора } F \tag{4.4}$$

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m}\bar{x}_i = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} \quad \text{загальне середнє} \tag{4.5}$$

Зробивши перетворення у формулі (4.3), можна прийти до зручніших формул для обчислення величин Q_F та Q_Z, а саме:

$$Q = C^* - \frac{C^2}{mn}, \quad Q_F = \frac{1}{n}\sum_{i=1}^{m} C_i^2 - \frac{C^2}{mn}, \quad Q_Z = C^* - \frac{1}{n}\sum_{i=1}^{m} C_i^2, \tag{4.6}$$

де

$$C_i = \sum_{j=1}^{n} x_{ij} = x_{i1} + x_{i2} + ... + x_{in}, \quad C_i^* = \sum_{j=1}^{n} x_{ij}^2 = x_{i1}^2 + x_{i2}^2 + ... + x_{in}^2, \tag{4.7}$$

$$C = \sum_{i=1}^{m} C_i = C_1 + C_2 + ... + C_m, \quad C^* = \sum_{i=1}^{m} C_i^* = C_1^* + C_2^* + ... + C_m^*. \tag{4.8}$$

Для зручності дані спостережень записують в таблицю:

Таблиця 4.1

Рівні фактора	x_{ij}	C_i	C_i^2	x_{ij}^2	C_i^*
F_1	$x_{11},..., x_{1j},..., x_{1n}$	C_1	C_1^2	$x_{11}^2,..., x_{1j}^2,..., x_{1n}^2$	C_1^*
F_2	$x_{21},..., x_{2j},..., x_{2n}$	C_2	C_2^2	$x_{21}^2,..., x_{2j}^2,..., x_{2n}^2$	C_2^*
...	....	...	...	......	...
F_i	$x_{i1},..., x_{ij},..., x_{in}$	C_i	C_i^2	$x_{i1}^2,..., x_{ij}^2,..., x_{in}^2$	C_i^*
...	....	...	...	......	...
F_m	$x_{m1},..., x_{mj},..., x_{mn}$	C_m	C_m^2	$x_{m1}^2,..., x_{mj}^2,..., x_{mn}^2$	C_m^*
		C	$\sum C_i^2$		C^*

За відносними значеннями розсіянь Q_F і Q_Z вже можна зробити висновки про вплив фактора F на випадкову величину X, однак значно надійніші висновки отримуємо, обчисливши такі величини:

$$S^2 = \frac{1}{mn-1}Q, \quad S_F^2 = \frac{1}{m-1}Q_{F,} \quad S_Z^2 = \frac{1}{mn-m}Q_Z. \qquad (4.9)$$

Ці величини називаються відповідно виправленою загальною (S^2), виправленою факторною (S_F^2) і виправленою залишковою (S_Z^2) дисперсіями, а числа:

$$\nu = mn-1, \quad \nu_F = m-1, \quad \nu_Z = mn-m \qquad (4.10)$$

- ступенями їх вільності.

Відношення

$$F = F(\nu_F, \nu_Z) = \frac{S_F^2}{S_Z^2}, \qquad (4.11)$$

як відомо, є випадковою величиною, яка розподілена за законом Фішера-Снедекора[1,4] з $\nu_F = m-1$ і $\nu_Z = mn-m$ ступенями вільності. Результати обчислень зручно записати в таблиці:

Таблиця 4.2

Джерела мінливості	Суми квадратів	Ступені вільності	Виправлені дисперсії	$F_{емп}$	$F_{кр}$
F	Q_F	$\nu_F = m-1$	$\frac{Q_F}{\nu_F}$	$F = \frac{S_F^2}{S_Z^2}$	$F_{\text{êð}}$
Z	Q_Z	$\nu_F = mn-m$	$\frac{Q_Z}{\nu_Z}$		
	Q	$\nu = mn-1$	$\frac{Q}{\nu}$		

Для даного рівня значущості α і ступенів вільності ν_F і ν_Z за таблицею додатка 6 критичних значень розподілу Фішера-Снедекора знаходимо $F_{кр}$.

Порівняємо F і $F_{кр}$. Якщо $F > F_{кр}$, то нульова гіпотеза H_0 відхиляється, як малоймовірна, а якщо $F < F_{кр}$, то гіпотеза H_0 для заданого рівня значущості α приймається.

Зауважимо що, якщо нульова гіпотеза H_0 про рівність математичних сподівань справджується, то всі дисперсії (4.6) є незміщеними оцінками генеральної дисперсії.

Поміркуємо над запропонованим алгоритмом дисперсійного аналізу порівняння факторної S_F^2 і залишкової S_Z^2 дисперсій за критерієм Фішера-Снедекора.

1. Нехай нульова гіпотеза H_0 про рівність групових середніх правильна. Тоді факторна і залишкова дисперсії є незміщеними оцінками невідомої дисперсії генеральної сукупності і, отже, відрізняються незначно. Якщо порівняти ці оцінки за критерієм Фішера - Снедекора, то цей критерій вкаже на те, що нульову гіпотезу про рівність факторної та залишкової дисперсій слід прийняти.

2. Нехай гіпотеза H_0 про рівність групових середніх не справджується. В цьому випадку з ростом відмінності між груповими середніми збільшується і факторна дисперсія, а разом з нею і відношення $F_{emp} = \frac{S_F^2}{S_Z^2}$. Отже $F_{емр}$ стає більше ніж F_{kp} і тому гіпотеза про рівність двох дисперсій (факторної та залишкової) не справджується.

Отже, для того, щоб перевірити **нульову гіпотезу** H_0 **про рівність групових середніх вибірок,** які розподілені за нормальним законом з однаковими невідомими дисперсіями, досить перевірити **за критерієм Фішера-Снедекора нульову гіпотезу про рівність факторної** S_F^2 **і залишкової** S_Z^2 **дисперсій.** В цьому й полягає **метод дисперсійного аналізу.**

Зауваження. Якщо факторна дисперсія виявиться меншою ніж залишкова, то це означає, що нульова гіпотеза про рівність групових середніх справджується, і нема потреби застосовувати в цьому випадку критерій Фішера-Снедекора.

Приклад 4.1. Двома приладами по 2 рази виміряно фізичну величину, справжній розмір якої дорівнює x. Розглядаючи як фактор систематичну похибку C, а в якості його рівнів систематичні похибки C_1 і C_2 відповідно першого та другого приладів, показати, що Q_F визначається систематичними, а Q_Z випадковими похибками вимірів.

Розв'язання. Введемо позначення: α_1, α_2 випадкові похибки першого та другого вимірів 1-го приладу, β_1, β_2 випадкові похибки першого та другого вимірів 2-го приладу. Ці дані можна записати в таблицю:

Таблиця 4.3

F_1	$x_{11} = x + C_1 + \alpha_1$	$x_{12} = x + C_1 + \alpha_2$
F_2	$x_{21} = x + C_2 + \beta_1$	$x_{22} = x + C_2 + \beta_2$

Обчислимо групові середні та загальне середнє за формулами (4.4) ,(4.5):

$$\bar{x}_1 = \frac{1}{2}(2x + 2C_1 + \alpha_1 + \alpha_2) = x + C_1 + \alpha; \tag{4.12}$$

$$\bar{x}_2 = \frac{1}{2}(2x + 2C_2 + \beta_1 + \beta_2) = x + C_1 + \beta; \tag{4.13}$$

$$\bar{x} = \frac{(\bar{x}_1 + \bar{x}_2)}{2} = x + \frac{C_1 + C_2}{2} + \frac{\alpha + \beta}{2}. \tag{4.14}$$

Обчислимо факторний ефект за першою з формул (4.3), використовуючи формули (4.12) - (4.14).

$$Q_F = 2\left[\left(\bar{x}_1 - \bar{x}\right)^2 + \left(\bar{x}_2 - \bar{x}\right)^2\right] =$$

$$= \left(C_1 - C_2\right)^2 + 2\left(C_1 - C_2\right)\left(\alpha - \beta\right) + \left(\alpha - \beta\right)^2.$$

Остання формула показує, що факторний ефект Q_F визначається в основному першим доданком, оскільки випадкові помилки α, β досить малі. Отже Q_F дійсно характеризує вплив фактора C.

Обчислимо залишкову суму Q_Z за другою з формул (4.3):

$$Q_Z = \left(x_{11} - \bar{x}_1\right)^2 + \left(x_{12} - \bar{x}_1\right)^2 + \left(x_{21} - \bar{x}_2\right)^2 + \left(x_{22} - \bar{x}_2\right)^2 =$$
$$= \left(\alpha_1 - \alpha\right)^2 + \left(\alpha_2 - \alpha\right)^2 + \left(\beta_1 - \beta\right)^2 + \left(\beta_2 - \beta\right)^2 .$$

Остання формула показує, що Q_Z визначається випадковими помилками вимірювань α і β.

Приклад 4.2. Проведено 4 випробування на кожному з 3-х рівнів F_1, F_2, F_3 фактора F. Результати випробувань наведені в таблиці:

Таблиця 4.4

Рівні фактора	Вибірки	Групові середні	Загальне середнє
F_1	$51; 52; 56; 57$	$\bar{x}_1 = 54$	$\bar{x} = 52$
F_2	$52; 54; 56; 58$	$\bar{x}_2 = 55$	
F_3	$42; 44; 50; 52$	$\bar{x}_3 = 47$	

Потрібно за допомогою методу дисперсійного аналізу за рівня значущості $\alpha = 0,05$ перевірити нульову гіпотезу H_0 про рівність математичних сподівань генеральних сукупностей, з яких зроблені вибірки таблиці, за умови, що вони (генеральні сукупності) розподілені за нормальними законами з рівними дисперсіями.

Розв'язання. Для спрощення розрахунків інколи доцільно від кожної варіанти вибірок наведеної таблиці відняти одне і те ж число C, яке дорівнює загальному середньому $\bar{x}$. При цьому всі розрахунки ефектів Q_F і Q_Z з допомогою формул (4.7) і (4.8) проводяться з заміною варіант x_{ij} на варіанти $y_{ij} - C = y_{ij} - \bar{x}$; $(i = 1, 2, 3;\ j = 1, 2, 3, 4)$.

Результати обчислень, застосування алгоритму знаходження величин Q_F, Q_Z, а також застосування критерію дисперсійного аналізу можна оформити у вигляді таблиць.

Таблиця 4.5

Рівні факто-ра	y_{ij}	C_i	C_i^2	y_{ij}^2	C_i^*
F_1	$-1; 0; 4; 5$	$C_1 = 8$	$C_1^2 = 64$	$1; 0; 16; 25$	$C_1^* = 42$
F_2	$0; 2; 4; 6$	$C_2 = 12$	$C_2^2 = 144$	$0; 4; 16; 36$	$C_2^2 = 56$
F_3	$-10; -8; -2; 0$	$C_3 = -20$	$C_3^2 = 400$	$100; 64; 4; 0$	$C_3^2 = 168$

Таблиця 4.6

Джерела мінли-вості	Суми квадра-тів	Ступенів вільнос-ті	Виправле-ні диспер-сії	$F_{емп}$	$F_{кр}$
F	$Q_F = 152$	$\nu_F = 2$	$S_F^2 = \frac{152}{2} = = 76$	$\frac{S_F^2}{S_Z^2} \approx 6$	$F_{кр} = F(0,05, 2, 9) = = 4,26$
Z	$Q_Z = 114$	$\nu_F = 9$	$S_Z^2 = \frac{114}{9} = = 12,67$		

Оскільки $F_{емп} > F_{кр}$, то нульова гіпотеза H_0 про рівність математичних сподівань не виконується і рівень фактора — істотний.

Зауважимо, що $F_{êð}$ знайдено з таблиці 4.6 додатка

Виконаємо завдання з допомогою програми Exel.

1. Формуємо таблицю для обчислень. Для цього об'єднуємо комірки **A1:A2**, куди вводимо ім'я «**№**» та комірки **B1:D1**, куди вводимо ім'я «**Рівні фактора**». В комірку **B2** вводимо «**F1**», в комірку **C2** вводимо ім'я «**F2**», в комірку **D2** вводимо ім'я «**F3**». В сформовану таблицю вносимо дані задачі (рис. 4.1)

	A	B	C	D	E
1		Рівні фактора			
2	№	F1	F2	F3	
3	1	51	52	42	
4	2	52	54	44	
5	3	56	56	50	
6	4	57	58	52	
7					

Рис. 4.1

2. Вводимо в комірку **A7** ім'я "***ni=***", а в комірки **B7**, **C7**, **D7** – числа, що відповідають кількостям експериментів на відповідних рівнях фактора, які обчислюємо за формулами:
   B7: **=СЧЁТ(B3:B6)**;
   C7:**=СЧЁТ(C3:C6)**;
   D7:**=СЧЁТ(D3:D16)**.
3. У комірку **A8** вводимо ім'я "**1/ni=**". У відповідних комірках проводимо обчислення:
   B8: **=1/B7**;
   C8: =**1/C7**;
   D8: =**1/D7**.
4. У комірку **A9** вводимо ім'я "**Суми**". Проводимо потрібні обчислення у відповідних комірках:
   - **B9**: **=СУММ(B3:B6)**;
   - **C9**: **=СУММ(C3:C6)**;
   - **D9**: **=СУММ(D3:D6)**.
5. В комірку **A10** вводимо ім'я «**Середні**». Проводимо потрібні обчислення у відповідних комірках:
   - **B10**: **= СРЗНАЧ(B3:B6)**;

- **C10**: **=СРЗНАЧ(C3:C6)**;
- **D10**: **=СРЗНАЧ(D3:D6)**.

6. Об’єднуємо комірки **A11:B11**, вводимо текст “**Загальний обсяг**”. В комірці **C11** вводимо формулу **C11**: **=B7+C7+D7**.
7. Об’єднуємо комірки **A12:B12**, вводимо текст “**Загальна середня**”. В комірці **C12** вводимо формулу **=СРЗНАЧ(B3:B6;C3:C6;D3:D6)**.
8. Об’єднуємо комірки **E1:G1**, вводимо текст “**Квадрати різниць по групах**”. У комірки **E2**, **F2**, **G2** вводимо імена «**F1**», «**F2**» і «**F3**».
9. Об’єднуємо комірки **H1:J1**, вводимо текст “**Квадрати різниць із заг. сер.**”. У комірки **H2, I2, J2** вводимо імена «**F1**», «**F2**» і «**F3**».
10. У комірці **E3** вводимо формулу **=(B3-B$10)^2**. Розповсюджуємо її на комірки **E4: E6**.
11. У комірці **F3** вводимо формулу **=(C3-C$10)^2**. Розповсюджуємо її на комірки **F4: F6**.
12. У комірці **G3** вводимо формулу **=(D3-D$10)^2**. Розповсюджуємо її на комірки **G4: G6**.
13. У комірці **H3** вводимо формулу **=(B3-C$12)^2**. Розповсюджуємо її на комірки **H4: H6**.
14. У комірці **I3** вводимо формулу **=(C3-C$12)^2**. Розповсюджуємо її на комірки **I4: I6**.
15. У комірці **J3** вводимо формулу **=(D3-C$12)^2**. Розповсюджуємо її на комірки **J4: J6**.

Результати обчислень показано на рис. 4.2

	A	B	C	D	E	F	G	H	I	J	K
1		Рівні фактора			Квадрати різниць по групах			Квадрати різниць із заг. середнім			
2	№	F1	F2	F3	F1	F2	F3	F1	F2	F3	
3	1	51	52	42	9	9	25	1	0	100	
4	2	52	54	44	4	1	9	0	4	64	
5	3	56	56	50	4	1	9	16	16	4	
6	4	57	58	52	9	9	25	25	36	0	
7	ni	4	4	4							
8	1/ni	0,25	0,25	0,25							
9	суми	216	220	188							
10	середні	54	55	47							
11	Загальний обсяг		12								
12	Загальна середня		52								
13											

Рис. 4.2.

16. Об’єднуємо комірки **A14:B14; A15:B15; A16:B16; A17:B17**. Вводимо назви: «**Вид дисперсії**», «**Внутрішньогрупова**», «**Міжгрупова**», «**Загальна**».
17. Об’єднуємо комірки **C14:E14**, **F14:G14**, **H14:J14**. Водимо назви: «**Сума квадратів відхилень**», «**Ступені свободи**», «**Статистичні оцінки дисперсій**».
18. Об’єднуємо комірки **D15:E15**, **D16:E16**, **D17:E17**, **F15:G15**, **F16:G16**, **F17:G17**, **I15:J15**, **I16:J16**, **I17:J17**.
19. У комірки **C15**, **C16**, **C17** вводимо назви: «***Dвнутр=***», «***Dміжгруп=***», «***Dзаг=***». В комірки **H15**, **H16**, **H17** вводимо назви «***SSвнутр=***», «***SSміжгруп=***», «***SSзаг=***».
20. Об’єднуємо комірки **A18:B18** і вводимо назву «**Критерій Фішера**». У комірки **A19**, **A20**, **A21** вводимо назви «$F_{емп}$ =», «$F_{0,01}$ =», «$F_{0,05}$ =».
21. У комірку **DE15** вводимо формулу **=СУММ(E3:E6;F3:F6;G3:G6)**.
22. У комірку **DE16** вводимо формулу **=B7*(B10-C12)^2+C7*(C10-C12)^2+D7*(D10-C12)^2.**
23. У комірку **DE17** вводимо формулу **=СУММ(H3:H6;I3:I6;J3:J6)**.
24. У комірки **FG19, FG20, FG21** вводимо формули
    - **FG15: = C11-3**;
    - **FG16: =3-1**;
    - **FG17: =C11-1**.
25. У комірки **IJ19**, **IJ20**, **IJ21** вводимо формули:
    - **IJ15: =D15/F15**;
    - **IJ16: =D16/F16**;
    - **IJ17: =D17/F17**.
26. У комірки **B23**, **B24**, **B25** вводимо формули:
    - **B19: =I16/I15**;

- **B20: = FРАСПОБР(0,01;F16;F15)**;
- **B21: =FРАСПОБР(0,05;F16;F15)**.

27. Робимо висновки на підставі отриманих числових результатів. Якщо емпіричне значення критерію є більшим за критичну точку, то нульову гіпотезу відхиляють (правобічна критична область). Результати обчислень наведено на рис. 4.3

13						
14	Вид дисперсії		Суми квадратів відхилень	Ступені свободи	Статистичні оцінки дисперсій	
15	Внутрішньогрупова	Dвнутр=	114	9	SSвнутр=	12,67
16	Міжгрупова	Dміжгруп=	152	2	SSміжгруп=	76
17	Загальна	Dзаг=	266	11	SSзаг=	24,18
18	Критерій Фішера					
19	Fемп = 6					
20	F0,01 = 8,02					
21	$F_{0,05}$ = 4,26					
22						

Рис. 4.3 Результати дисперсійного аналізу

28. Задачу можна виконати значно швидше, застосувавши пакет «**Аналіз данних**», розділ «**Однофакторний дисперсійний аналіз**». Для цього необхідно внести дані в діалогове вікно так, як показано на рис. 4.4.

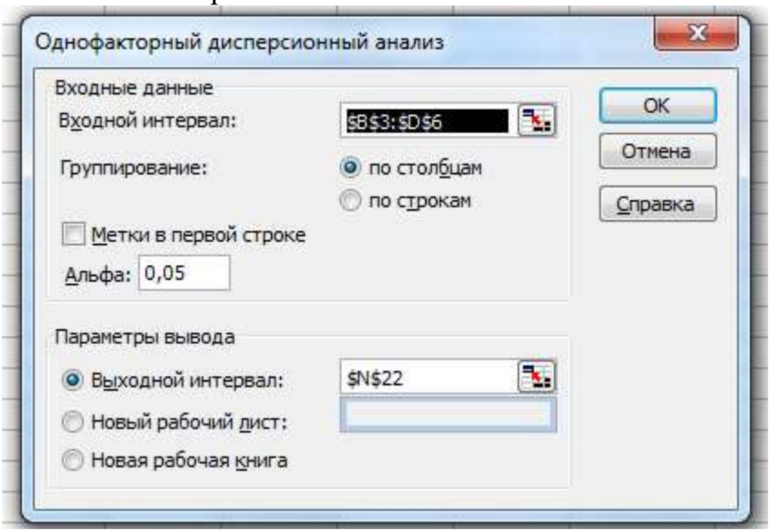


Рис. 4.4. Ввід даних для проведення дисперсійного аналізу

Отримані результати наведені на рис. 4.5

Однофакторный дисперсионный анализ

ИТОГИ

Группы	Счет	Сумма	Среднее	Дисперсия
Столбец 1	4	216	54	8,666666667
Столбец 2	4	220	55	6,666666667
Столбец 3	4	188	47	22,66666667

Дисперсионный анализ

Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	152	2	76	6	0,022085359	4,256494729
Внутри групп	114	9	12,66667			
Итого	266	11				

Рис. 4.5. Результати дисперсійного аналізу

4.2. Двофакторний дисперсійний аналіз

Нехай необхідно визначити вплив двох факторів ***A*** та ***B*** на певну ознаку ***Y***. Для цього необхідно, щоб дослід здійснювався при фіксованих рівнях факторів ***A*** та ***B***, а також їх одночасній дії на ознаку. При цьому дослід здійснюємо однакову кількість разів (*n* разів) для кожного з рівнів факторів ***A*** та ***B***. Загальну кількість експериментів позначимо буквою *N*.

Формулюємо три групи статистичних гіпотез.

Набір А:

- **H_{0A}:** розсіювання ознаки, зумовлене впливом фактора ***A***, не більш виражене, ніж розсіювання, зумовлене впливом випадкових чинників (вплив фактора ***A*** відсутній);
- H_{1A}: розсіювання ознаки, зумовлене впливом фактора ***A***, більш виражене, ніж розсіювання, зумовлене впливом випадкових чинників (фактор ***A*** впливає на результат експерименту).

Набір В:

- **H_{0B}:** розсіювання ознаки, зумовлене впливом фактора ***B***, не більш виражене, ніж розсіювання, зумовлене впливом випадкових чинників (фактор ***B*** не впливає на результат експерименту);

- **H$_{1B}$:** розсіювання ознаки, зумовлене впливом фактора ***B***, більш виражене, ніж розсіювання, зумовлене впливом випадкових чинників (фактор ***B*** впливає на результат експерименту).

Набір АВ:

- **H$_{0AB}$:** вплив фактора ***A*** на результат експерименту однаковий при різних градаціях фактора ***B***, і навпаки, вплив фактора ***B*** на результат експерименту не залежить від градацій фактора ***A***;
- **H$_{1AB}$:** вплив фактора ***A*** різний при різних градаціях фактора ***B***, і навпаки, вплив фактора ***B*** на результат експерименту залежить від градацій фактора ***A***.

Цим гіпотезам відповідає статистичний критерій Фішера з правобічною критичною областю. Обчислюємо емпіричні значення критерію за формулами, наведеними в таблицях 4.7 та 4.8:

Таблиця 4.7

Фактор B	Рівні фактора А				Середня величина за рядками
	A_1	A_2	…	A_p	
B_1	x_{111}, x_{211}, x_{311}, …, x_{n11}, блочна середня: $\bar{x}_{11} = \frac{1}{n}\sum_{i=1}^{n} x_{i11}$	x_{121}, x_{221}, x_{321}, …, x_{n21}, блочнасередня: $\bar{x}_{21} = \frac{1}{n}\sum_{i=1}^{n} x_{i21}$		x_{1p1}, x_{2p1}, x_{3p1}, …, x_{np1}, блочна середня: $\bar{x}_{p1} = \frac{1}{n}\sum_{i=1}^{n} x_{ip1}$	$\bar{y}_1 = \frac{1}{np}\sum_{i=1}^{n}\sum_{k=1}^{p} x_{i1k}$
B_2	x_{112}, x_{212}, x_{312}, …, x_{n12}, блочна середня: $\bar{x}_{12} = \frac{1}{n}\sum_{i=1}^{n} x_{i12}$	x_{122}, x_{222}, x_{322}, …, x_{n22}, блочна середня: $\bar{x}_{22} = \frac{1}{n}\sum_{i=1}^{n} x_{i22}$		x_{1p2}, x_{2p2}, x_{3p2}, …, x_{np2}, блочна середня: $\bar{x}_{p2} = \frac{1}{n}\sum_{i=1}^{n} x_{ip2}$	$\bar{y}_2 = \frac{1}{np}\sum_{i=1}^{n}\sum_{k=1}^{p} x_{i2k}$
…	……………	…………...	.	……………	……………

Bq	x_{1q}, $x_{2 1q}$, x_{31q}, …, x_{n1q}, блочна середня: $\bar{x}_{1q} = \frac{1}{n}\sum_{i=1}^{n} x_{i1q}$	x_{12q}, x_{22q}, x_{32q}, …, x_{n2q}, блочна середня: $\bar{x}_{21} = \frac{1}{n}\sum_{i=1}^{n} x_{ii2q}$		x_{1pq}, x_{2pq}, x_{3pq}, …, x_{npq}, блочна середня: $\bar{x}_{p1} = \frac{1}{n}\sum_{i=1}^{n} x_{i11}$	$\bar{y}_q = \frac{1}{np}\sum_{i=1}^{n}\sum_{k=1}^{p} x_{iqk}$
Середня величина за стовпцями	$\bar{z}_1 = \frac{1}{nq}\sum_{i=1}^{n}\sum_{j=1}^{q} x_{ij1}$	$\bar{z}_2 = \frac{1}{nq}\sum_{i=1}^{n}\sum_{j=1}^{q} x_{ij2}$		$\bar{z}_p = \frac{1}{nq}\sum_{i=1}^{n}\sum_{j=1}^{q} x_{ijp}$	
Загальна середня величина $\bar{x} = \frac{1}{npq}\sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{p} x_{ijk}$					

Таблиця 4. 8

Джерело розсіювання	Сума квадратів відхилень	Ступені свободи	Виправлені дисперсії
Фактор A	$Q_A = nq\sum_{j=1}^{p}(z_j - \bar{x})^2$	$k_A = p-1$	$S_A^2 = \frac{Q_A}{k_A}$
Фактор B	$Q_B = np\sum_{i=1}^{q}(\bar{y}_i - \bar{x})^2$	$k_B = q-1$	$S_B^2 = \frac{Q_B}{k_B}$
Фактор AB (одночасно)	$Q_{AB} = n\sum_{j=1}^{q}\sum_{k=1}^{p}(\bar{x}_{jk} - \bar{z}_k - \bar{y}_j + \bar{x})^2$	$k_{AB} = (p-1)(q-1)$	$S_B^2 = \frac{Q_{AB}}{k_{AB}}$
Випадкове	$Q = \sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{p}(x_{ijk} - \bar{x}_{jk})^2$	$k = N - pq$	$S^2 = \frac{Q}{k}$

$$F_A = \frac{S_A^2}{S^2} \qquad (4.15)$$

характеризує варіативність ознаки, зумовлену дією фактора ***A***,

$$F_B = \frac{S_B^2}{S^2} \qquad (4.16)$$

характеризує варіативність ознаки, зумовлену дією фактора ***B***,

$$F_{AB} = \frac{S_{AB}^2}{S^2} \qquad (4.17)$$

характеризує варіативність ознаки, зумовлену взаємодією факторів ***A*** та ***B***.

З таблиці критичних точок критерію Фішера, або за допомогою статистичної функції Excel **FРАСПОБР**, на заданому рівні значущості α та знайдених ступенях свободи знаходимо критичні точки: $F_{крA}(\alpha;\ k_A;\ k)$, $F_{крB}(\alpha;\ k_B;\ k)$, $F_{крAB}(\alpha;\ k_{AB};\ k)$.

Якщо $F_A > F_{крA}$, то нульову гіпотезу H_{0A} відхиляємо і приймаємо H_{1A}.

Якщо $F_B > F_{крB}$, то нульову гіпотезу H_{0B} відхиляємо і приймаємо H_{1B}.

Якщо $F_{AB} > F_{крAB}$, то нульову гіпотезу H_{0AB} відхиляємо і приймаємо H_{1AB}.

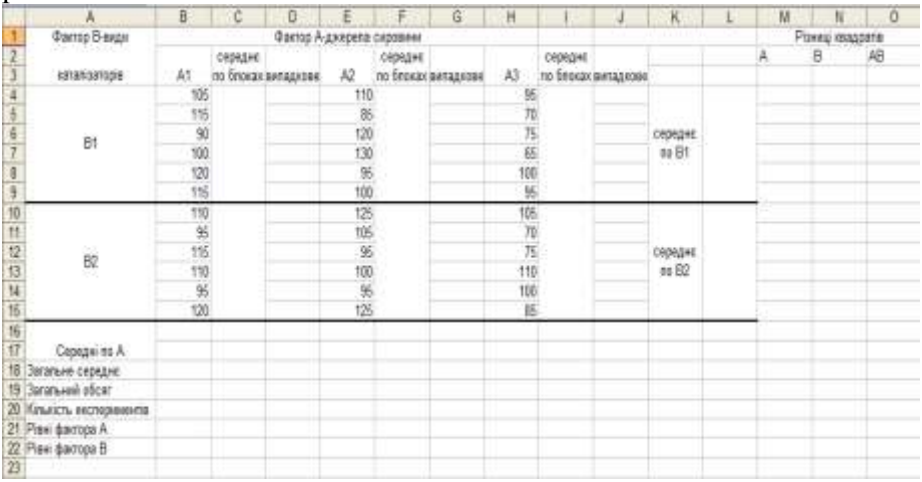
Оскільки реалізація двофакторного дисперсійного аналізу є громіздкою, потребує багато обчислень, розглянемо приклад розв'язування задачі з допомогою Exel.

Приклад 4.3. Перевірити наявність чи відсутність впливу на результат експерименту каталізатора та виду сировини. Результати вимірів наведено в таблиці.

Фактор В	Фактор А – джерела сировини		
	А1	А2	А3
В1	105	110	95
	115	85	70
	90	120	75
	100	130	65
	120	95	100
	115	100	95
В2	110	125	105
	95	105	70
	115	95	75
	110	100	110

	95	95	100
	120	125	85

1. Формуємо таблицю для обчислень, як показано на рис.4.6.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Фактор В-види	Фактор А-джерела сировини											Різниці квадратів		
2			середнє			середнє			середнє				A	B	AB
3	каталізаторів	A1	по блоках	випадкове	A2	по блоках	випадкове	A3	по блоках	випадкове					
4	B1	105			110			95			середнє за B1				
5		115			85			70							
6		90			120			75							
7		100			130			65							
8		120			95			100							
9		115			100			95							
10	B2	110			125			105			середнє за B2				
11		95			105			70							
12		115			95			75							
13		110			100			110							
14		95			95			100							
15		120			125			85							
16															
17	Середні за А														
18	Загальне середнє														
19	Загальний обсяг														
20	Кількість експериментів														
21	Рівні фактора А														
22	Рівні фактора В														
23															

Рис. 4.6. Таблиця для проведення розрахунків двофакторного дисперсійного аналізу

2. В комірку **C4** вводимо формулу **=СРЗНАЧ(B4:B**9**)**;
   у комірку **C10** вводимо формулу **=СРЗНАЧ(B10:B15)**;
   у комірку **F4** вводимо формулу **=СРЗНАЧ(E4:E9)**;
   у комірку **F10** вводимо формулу **=СРЗНАЧ(E10:E15)** ;
   у комірку **I4** вводимо формулу **=СРЗНАЧ(H4:H9)**;
   у комірку **I10** вводимо формулу **=СРЗНАЧ(H10:H15)**.
3. В комірку **L4** вводимо формулу
   =СРЗНАЧ(B4:B9;E4:E9;H4:H9);
   у комірку **L10** вводимо формулу
   =СРЗНАЧ(B10:B15;E10:E15;H10:H15).

Формули цього пункту вводимо, за зразком, наведеним на рисунку 4.7

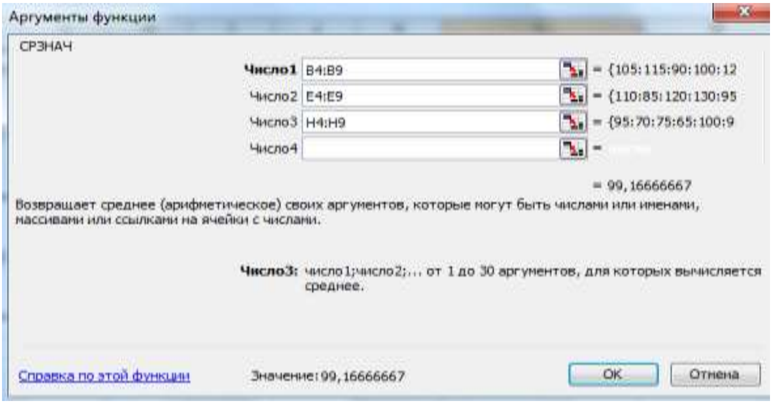


Рис. 4.7. Обчислення середніх значень по рядках

4. У комірку **B17** вводимо формулу **=СРЗНАЧ(B4:B15)**;
   у комірку **E17** вводимо формулу **=СРЗНАЧ(E4:E15)**;
   у комірку **H17** вводимо формулу **=СРЗНАЧ(H4:H35)**.
5. У комірку **C18** вводимо формулу **=СРЗНАЧ(B4:B15;E4:E15;H4:H15)**.
6. У комірку **C19** вводимо формулу **=СЧЁТ(B4:B15;E4:E15;H4:H15)**.
7. У комірку **C20** вводимо формулу **=СЧЁТ(B4:B9)**.
8. У **C21** вводимо число, що відповідає кількості рівнів фактора A. В нашому випадку це число дорівнює 3.
9. У комірку **C22** вводимо число, що відповідає кількості рівнів фактора B. В нашому прикладі це число дорівнює 2.
10. У комірку **M4** вводимо формулу **=(B17-C18)^2**;
   у комірку **M5** вводимо формулу **=(E17-C18)^2**;
   у комірку **M6** вводимо формулу **=(H17-C18)^2**.
13. У комірку **N4** вводимо формулу **=(L4-C18)^2**;
   у комірку **N5** вводимо формулу **=(L10-C18)^2**.
14. У комірку **O4** вводимо формулу **=(C4-B17-L4+C18)^2**;
   у комірку **O5** вводимо **=(C10-B17-L10+C18)^2;**

у комірку **O6** вводимо формулу **=(F4-E17-L4+C18)^2;**

у комірку **O7** вводимо формулу **=(F10-E17-L10+C18)^2;**

у комірку **O8** вводимо формулу **=(I4-H17-L4+C18)^2;**

у комірку **O9** вводимо формулу **=(I10-H17-L10+C18)^2.**

15. У комірку **D4** вводимо формулу **=(B4-C$4)^2** і поширюємо за допомогою маркера заповнення її вміст на весь блок **D4: D9**;

    у комірку **D10** вводимо формулу **=(B10-C$10)^2** і поширюємо за допомогою маркера заповнення її вміст на весь блок **D10: D15**;

    у комірку **G4** вводимо формулу **=(E4-F$4)^2** і поширюємо за допомогою маркера заповнення її вміст на весь блок **G4: G9**;

    у комірку **G10** вводимо формулу **=(E0-F$10)^2** і поширюємо за допомогою маркера заповнення її вміст на весь блок **G10: G15**;

    у комірку **J4** вводимо формулу **=(H4-I$4)^2** і поширюємо за допомогою маркера заповнення її вміст на весь блок **J4: J9**;

    у комірку **J10** вводимо формулу **=(H10-I$10)^2** і поширюємо за допомогою маркера заповнення її вміст на весь блок **J10: J15**.

Результати обчислень показано на рисунку 4.8.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Фактор В-	Фактор А-джерела сировини											Різниці квадратів		
2	види		середнє			середнє			середнє				A	B	AB
3	каналізаторів	A1	по блоках	випадкове	A2	по блоках	випадкове	A3	по блоках	випадкове					
4		105		6,25	110		11,11	95		136,11			46,23	1,93	1,93
5		115		56,25	85		469,44	70		177,78			42,61	1,93	1,93
6	B1	90	107,50	306,25	120	106,67	177,78	75	83,33	69,44	середнє по B1	99,17	101,56		0,96
7		100		56,25	130		544,44	65		336,11					0,96
8		120		156,25	95		136,11	100		277,78					6,57
9		115		56,25	100		44,44	95		136,11					6,57
10		110		6,25	125		306,25	105		200,69					
11		95		156,25	105		6,25	70		434,03					
12	B2	115	107,50	56,25	95	107,50	156,25	75	90,83	250,69	середнє по B2	101,94			
13		110		6,25	100		56,25	110		367,36					
14		95		156,25	95		156,25	100		84,03					
15		120		156,25	125		306,25	85		34,03					
16															
17	Середні по А	107,5			107,08			87,08							
18	Загальне середнє		108,56												
19	Загальний обсяг		36												
20	Кількість експериментів		6												
21	Рівні фактора А		3												
22	Рівні фактора В		2												

Рис. 4.8. Розрахунки двофакторного дисперсійного аналізу

1. Сформуємо таблицю для підсумкових обчислень, як показано на рисунку 4.9.

	A	B	C	D	E	F	G	H	I	J
25	Джерело розсіювання		Сума квадратів відхилень			Ступені свободи		Стат. оцінки дисперсій		
26	Фактор А									
27	Фактор В									
28	Фактор АВ									
29	Випадкове									
30		Критерій Фішера								
31	Емпіричні значення		Критичні точки							
32	FA		F0,05							
33	FB		F0,05							
34	FAB		F0,05							
35										

Рис. 4.9. Таблиця для підсумкових обчислень, пов'язаних з двофакторним дисперсійним аналізом

2. У комірки **C26**, **C27**, **C28**, **C29** вводимо відповідні формули:
=C20*C22*СУММ(M4:M6);
=C20*C21*СУММ(N4:N5);
=C20*СУММ(O4:O9);
=СУММ(D4:D15;G4:G15;J4:J15).
3. У комірки **F26**, **F27**, **F28**, **F29** вводимо відповідні формули:
=C21-1;
=C22-1;
=(C21-1)*(C22-1);

=C19-C21*C22.

4. У комірки **H26**, **H27**, **H28**, **H29** вводимо відповідні формули:
   =C26/F26;
   =C27/F27;
   =C28/F28;
   =C29/F29.
5. У комірку **B32** вводимо формулу **=H26/H29**;
   у комірку **B33** вводимо формулу **=H27/H29**;
   у комірку **B34** вводимо формулу **=H28/H29**;
   у комірку **D32** вводимо формулу
   =FРАСПОБР(0,05;F26;F29);
   у комірку **D33** вводимо формулу
   =FРАСПОБР(0,05;F27;F29)
   у комірку **D34** вводимо формулу
   =FРАСПОБР(0,05;F28;F29).

На рис. 4.10 наведено отримані результати

	A	B	C	D	E	F	G	H	I	J
25	Джерело розсіювання		Сума квадратів відхилень			Ступені свободи		Стат. оцінки дисперсій		
26	Фактор А		3268,06			2		1634,03		
27	Фактор В		69,44			1		69,44		
28	Фактор АВ		101,39			2		50,69		
29	Випадкове		6050,00			30		201,67		
30		Критерій Фішера								
31	Емпіричні значення		Критичні точки							
32	FA	8,10	F0,05	3,31583						
33	FB	0,34	F0,05	4,170877						
34	FAB	0,25	F0,05	3,31583						

Рис. 4.10. Обчислення емпіричних значень критерію Фішера

Формулюємо гіпотези.

- H_{0A}: відмінності в результатах експерименту, зумовлені впливом джерел сировини, не більш виражені, ніж відмінності, викликані впливом випадкових чинників (фактор А не впливає на результат експерименту);
- H_{1A}: відмінності в результатах експерименту, зумовлені впливом джерел сировини, більш виражені, ніж відмінності, викликані впливом

випадкових чинників (фактор А впливає на результат експерименту);

- H_{0B}: відмінності в результатах експерименту, зумовлені впливом каталізатора, не більш виражені, ніж відмінності, викликані впливом випадкових чинників (фактор В не впливає на результат експерименту);
- H_{1B}: відмінності в результатах експерименту, зумовлені впливом каталізатора, більш виражені, ніж відмінності, викликані впливом випадкових чинників (фактор В впливає на результат експерименту);
- H_{0AB}: вплив джерел сировини на результат експерименту однаковий при різних градаціях виду каталізатора, і навпаки, вплив каталізатора на результати експерименту не залежить від градацій фактора джерела сировини (результат експерименту не залежить від сумісної дії факторів А та В);
- H_{1AB}: вплив джерел сировини на результат експерименту різний при різних градаціях виду каталізатора, і навпаки, вплив каталізатора на результати експерименту залежить від градацій фактора джерела сировини (результат експерименту залежить від сумісної дії факторів А та В).

6. **Висновки:** Критерій Фішера має правобічну критичну область. Оскільки емпіричне значення критерію F_A є більшим за відповідну критичну точку, то потрапляємо в критичну область гіпотези H_{1A} і робимо висновок, що відмінності в результатах експерименту, зумовлені впливом джерел сировини, більш виражені, ніж відмінності, викликані впливом випадкових чинників (фактор А впливає на результат експерименту). Оскільки емпіричне значення критерію F_B є меншим за відповідну критичну точку, то потрапляємо в область прийняття гіпотези H_{0B}.і робимо висновок, що відмінності в результатах експерименту, зумовлені впливом каталізатора, не

більш виражені, ніж відмінності, викликані впливом випадкових чинників (фактор В не впливає на результат експерименту). Оскільки F_{AB} є меншим за відповідну критичну точку, то потрапляємо в область прийняття гіпотези H_{0AB} і робимо висновок, що вплив джерел сировини на результат експерименту однаковий при різних градаціях виду каталізатора, і навпаки, вплив каталізатора на результати експерименту не залежить від градацій фактора джерела сировини (результат експерименту не залежить від сумісної дії факторів А та В).

7. Розрахунки двофакторної дисперсійної моделі можна провести за допомогою пакета «**Аналіз даних**», розділ «**Двофакторний дисперсійний аналіз з повтореннями**». Для цього сформуємо нову табличку даних, як показано на рис. 4.11.

	A	B	C	D
37	Фактор В	Фактор А-сировина		
38	каталізатор	А1	А2	А3
39	В1	105	110	95
40		115	85	70
41		90	120	75
42		100	130	65
43		120	95	100
44		115	100	95
45	В2	110	125	105
46		95	105	70
47		115	95	75
48		110	100	110
49		95	95	100
50		120	125	85
51				

Рис. 4.11. Таблиця даних для проведення двофакторного дисперсійного аналізу

Заходимо в меню «**Сервіс**» → «**Аналіз даних**» → «**Двофакторний дисперсійний анализ з повтореннями**», і у діалоговому вікні, що з'явилося вводимо дані, як показано на рис. 4.12

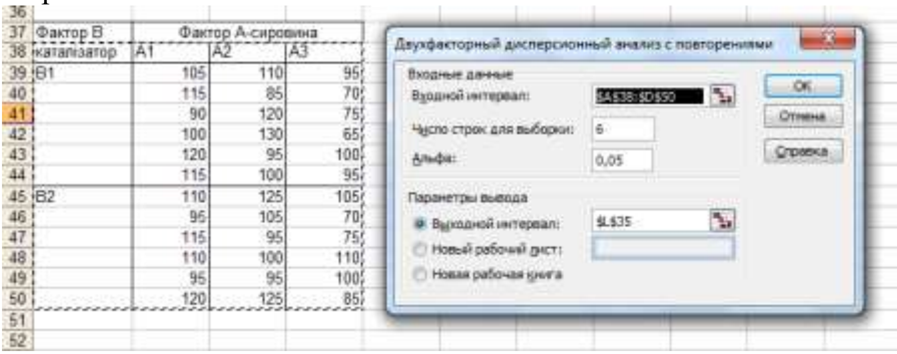


Рис. 4.12. Розрахунок двофакторної моделі за допомогою пакету «Аналіз даних»

Результати дисперсійного аналізу отримуємо у вигляді таблиці, наведеної на рис. 4.13.

Дисперсионный анализ						
Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Выборка	69,44	1	69,44	0,34	0,56	4,17
Столбцы	3268,06	2	1634,03	8,10	0,00	3,32
Взаимодействие	101,39	2	50,69	0,25	0,78	3,32
Внутри	6050,00	30	201,67			
Итого	9488,89	35				

Рис. 4.13. Результати двофакторного дисперсійного аналізу

Порівнюємо результати, отримані двома способами.

5. Основні поняття кореляційного та регресійного аналізу

5.1. Статистичний опис системи двох випадкових величин

Підставою для статистичного аналізу залежності між випадковими величинами X і Y є дані вибірки, які отримані із спостережень над двовимірною випадковою величиною (X, Y). Елементи вибірки зображуються з допомогою впорядкованих пар чисел (x_i, y_i), $i = 1, 2, ..., n$, де x_i вибіркове значення ознаки X, а y_i вибіркове значення ознаки Y, що відповідають i – му спостереженню, n – обсяг вибірки.

Вихідний статистичний матеріал можна подати у формі таблиці:

Таблиця 5.1

x_i	x_1	x_2	…	x_n
y_i	y_1	y_2	…	y_n

Якщо обсяг вибірки великий, то проводиться групування статистичних даних. Нехай серед вибіркових значень ознаки X можна виділити m різних значень або частинних інтервалів, а серед вибіркових значень ознаки Y можна виділити k різних значень або частинних інтервалів.

Ці згруповані дані двовимірної вибірки можна подати у вигляді таблиці:

Таблиця 5.2

$X = x_i$ / $Y = y_j$	x_1	x_2	…	x_m	n_{y_j}
y_1	n_{11}	n_{21}	…	n_{m1}	n_{y1}
y_2	n_{12}	n_{22}	…	n_{m2}	n_{y_2}
…	…	…	…	…	…
y_k	n_{1k}	n_{2k}	…	n_{mk}	n_{y_k}
n_{x_i}	n_{x_1}	n_{x_2}	…	n_{x_m}	n

Наведену таблицю називають **кореляційною таблицею.** В ній через n_{ij} позначають частоту появи події

$\left(X = x_i, \quad Y = y_j\right), \quad i = 1, 2, ..., m, \quad j = 1, 2, ..., k,$ а числа n_{ij}, n_{x_i}, n_{y_j} пов'язані співвідношеннями:

$$n_{x_i} = \sum_{j=1}^{k} n_{ij}, \quad i = 1, 2, ..., m; \quad n_{y_j} = \sum_{i=1}^{m} n_{ij}, \quad j = 1, 2, ..., k; \qquad (5.1)$$

$$\sum_{i=1}^{m} \sum_{j=1}^{k} n_{ij} = \sum_{i=1}^{m} n_{x_i} = \sum_{j=1}^{k} n_{y_j} = n. \qquad (5.2)$$

Якщо в кореляційній таблиці замість варіант візьмемо частинні інтервали, то вважатимемо, що в таблиці числа x_i та y_j означають середини відповідних інтервалів.

Нагадаємо, що розподіл ймовірностей випадкового вектора (X, Y) характеризується такими чисельними характеристиками, як математичними сподіваннями складових $M(X)$ і $M(Y)$, дисперсіями складових $D(X)$ і $D(Y)$, а також їх коваріацією та коефіцієнтом кореляції r_{xy}. З допомогою двовимірного вибіркового закону розподілу, який задається таблицями, можна обчислювати точкові оцінки для згаданих параметрів. Виявляється, що, як і в одновимірному випадку, точковими оцінками чисельних характеристик $M(X)$, $M(Y)$, $D(X)$, $D(Y)$ випадкового вектора (X, Y) є чисельні характеристики складових X і Y двовимірної вибірки, що задаються формулами:

- у випадку незгрупованої вибірки:

$$\overline{M}(X) = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{M}(Y) = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad (5.3)$$

$$\sigma_x^{*2} = \overline{D}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\overline{x}\right)^2,$$

$$\sigma_y^{*2} = \overline{D}(Y) = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\overline{y}\right)^2. \qquad (5.4)$$

- у випадку згрупованої вибірки:

$$\overline{M}(X)=\overline{x}=\frac{1}{n}\sum_{i=1}^{m}n_{x_i}x_i,\quad \overline{M}(Y)=\overline{y}=\frac{1}{n}\sum_{j=1}^{k}n_{y_j}y_j, \quad (5.5)$$

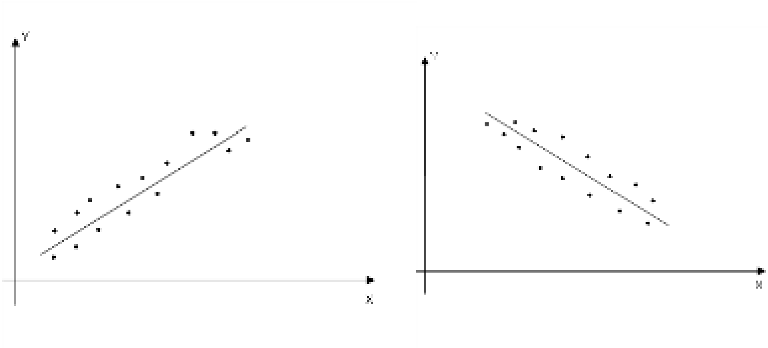
$$\sigma_x^{*2}=\overline{D}(X)=\frac{1}{n}\sum_{i=1}^{m}n_{x_i}x_i^2-\left(\overline{x}\right)^2,$$

$$\sigma_y^{*2}=\overline{D}(Y)=\frac{1}{n}\sum_{j=1}^{k}n_{y_j}y_j^2-\left(\overline{y}\right)^2. \quad (5.6)$$

5.2.Вибірковий коефіцієнт кореляції

Нехай для дослідження залежності між випадковими величинами X та Y (ознаками генеральної сукупності) маємо вибірку (x_i, y_i), $i=\overline{1,n}$ обсягу n. На основі цієї вибірки потрібно визначити напрям і ступінь лінійного кореляційного зв'язку між випадковими величинами X та Y.

У найпростіший спосіб наявність кореляційного зв'язку між випадковими величинами визначається шляхом побудови і візуального аналізу **діаграми розсіювання,** яку отримуємо за допомогою зображення елементів (x_i, y_i) вибірки точками площини *Oxy*. При цьому, якщо точки (x_i, y_i) площини *Oxy* розсіяні вздовж прямої лінії, то можна припустити, що між випадковими величинами X та Y існує тісний кореляційний зв'язок.

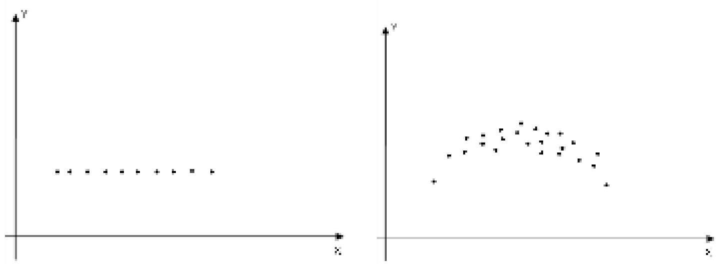


а) б)

Рис.5.1. Лінійна кореляційна залежність між випадковими величинами

Наприклад, розміщення точок $\left(x_i, y_i\right)$, яке зображене на рис.5.1, характеризує лінійну кореляційну залежність між випадковими величинами X та Y, бо ці точки розміщені близько до прямих, які нарисовані суцільними лініями. При цьому, пряма на рис 5.1. а) має додатний кутовий коефіцієнт і лінійну кореляційну залежність в цьому випадку називають ***додатною,*** а пряма на рис 5.1. б) має від'ємний кутовий коефіцієнт і лінійну кореляційну залежність в цьому випадку називають ***від'ємною.***

Буває й так, що між випадковими величинами X та Y немає жодного зв'язку (випадкова величина Y не реагує на випадкову величину X), або є нелінійний кореляційний зв'язок. Такі випадки зображені на рис. 5.2.



а) б)

Рис. 5.2. *Нелінійна кореляційна залежність між випадковими величинами*

З теорії ймовірностей відомо, що ступінь зв'язку між випадковими величинами X та Y визначається такими чисельними характеристиками їх сумісного розподілу, як **кореляційний момент** K_{xy} і **коефіцієнт кореляції** r_{xy}, які обчислюються за формулами:

$$K_{xy} = M\left(X \cdot Y\right) - M\left(X\right) \cdot M\left(Y\right), \quad (5.7)$$

$$r_{xy} = \frac{K_{xy}}{\sigma_x \cdot \sigma_y} = \frac{M\left(X \cdot Y\right) - M\left(X\right) \cdot M\left(Y\right)}{\sigma\left(X\right) \cdot \sigma\left(Y\right)} \quad (5.8)$$

Зауважимо, що коли випадкові величини X і Y незалежні, то $K_{xy} = 0$ і $r_{xy} = 0$. Якщо ж $K_{xy} \neq 0$, то випадкові величини X і Y залежні. Обернені твердження справджуються не завжди, тобто, якщо $K_{xy} = 0$ і $r_{xy} = 0$, то це не означає, що випадкові величини X і Y є незалежними. Якщо залежність між випадковими величинами X та Y лінійна, тобто існують числа A і B такі, що $Y = AX + B$, то $\left|r_{xy}\right| = 1$.

Якщо $\left|r_{xy}\right| = 0$, то випадкові величини X та Y називаються **некорельованими,** а якщо $r_{xy} \neq 0$, то випадкові величини X та Y називаються **корельованими.**

Отже основна задача кореляційного аналізу, яка полягає у виявленні залежності між випадковими величинами X та Y на основі двовимірної вибірки, може бути розв'язана шляхом побудови точкових та інтервальних оцінок коефіцієнта кореляції. Точкова оцінка r^* коефіцієнта кореляції $\overline{r}_{xy}$ обчислюється за формулою:

$$r^* = \overline{r_{xy}} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{\sigma_x} \cdot \overline{\sigma_y}}, \tag{5.9}$$

де $\overline{xy}$ вибіркове середнє добутку випадкової величини $X \cdot Y$, яке обчислюється за формулами:

- $$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i \tag{5.10}$$

у випадку не згрупованої вибірки

- $$\overline{xy} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij} x_i y_i \tag{5.11}$$

у випадку згрупованої вибірки.

Величини $\overline{x}$, $\overline{y}$, $\overline{\sigma_x}$, $\overline{\sigma_y}$ це вибіркові середні та вибіркові середньоквадратичні відхилення випадкових величин X та Y, які обчислюються за формулами (1.11) та (1.14) відповідно.

Означення. Точкова оцінка r^* коефіцієнта кореляції випадкових величин X та Y, яка обчислюється за формулою (5.9), називається **вибірковим коефіцієнтом кореляції.**

Вибірковий коефіцієнт кореляції r^* характеризує зв’язок між ознаками генеральної сукупності X та Y, а саме:

- Якщо $r^* > 0$, то зв’язок між випадковими величинами X та Y додатний;
- Якщо $r^* < 0$, то зв’язок між випадковими величинами X та Y від’ємний;
- Якщо $\overline{r}_{xy} = 0$, то випадкові величини X та Y некорельовані і це означає, між ними відсутній лінійний зв'язок.

Слід зауважити, що вибірковий коефіцієнт кореляції справджує нерівність: $\left|r_{xy}\right| < 1$.

Приклад 5.1. Статистичні дослідження залежності між прибутком підприємства і середньою заробітною платнею працівників на однорідних підприємствах галузі характеризуються наступною вибіркою:

Прибуток y %	94	142	128	70	86	120	76	118	134	112
Зарплата x% в гр..	84	162	136	86	100	150	94	118	138	114

Визначити вибірковий коефіцієнт кореляції та оцінити ступінь лінійного зв'язку між прибутком підприємства і заробітною платою працівників.

Розв'язання. Для зручності обчислень складемо таблицю:

Таблиця 5.3

№	Прибуток (%) y_i	Зарплата (грн.) x_i	$x_i \cdot y_i$	y_i^2	x_i^2
1	94	84	7896	8836	7056
2	142	162	23004	20164	26244
3	128	136	17408	16384	18496
4	70	86	6020	4900	7396
5	86	100	8600	7396	10000
6	120	150	18000	14400	22500
7	76	94	7144	5776	8836
8	118	118	13924	13924	13924
9	134	138	18492	17956	19044
10	112	114	12768	12544	12996
$\sum_{i=1}^{10}$	1080	1182	133256	122280	146492

За допомогою даних таблиці і формул (5.5), (5.6) та (5.11) матимемо:

$$\bar{x} = \frac{1}{10} \cdot 1182 = 118,2; \quad \bar{y} = \frac{1}{10} \cdot 1080 = 108;$$

$$\overline{xy} = \frac{1}{10}\cdot 133256 = 13325,6.$$

$$\overline{\sigma}_x^2 = \frac{1}{10}\cdot 146492 - 118,2^2 \approx 678; \quad \overline{\sigma}_y^2 = \frac{1}{10}\cdot 122280 - 108^2 \approx 564.$$

Підставивши отримані результати в формулу (5.9), обчислимо вибірковий коефіцієнт кореляції:

$$r^* = \overline{r_{xy}} = \frac{13326 - 118\cdot 108}{\sqrt{678}\cdot\sqrt{564}} \approx \frac{13326 - 12744}{26\cdot 24} \approx 0,91.$$

Оскільки $\overline{r}_{xy} \approx 0,91$ це число, яке близьке до одиниці, то залежність між прибутком підприємства і заробітною платою близька до лінійної.

Цю задачу можна швидко вирішити за допомогою калькулятора Exel.

1. Введемо дані задачі в таблицю, як показано на рис. 5.3.

	A	B	C	D	E	F
1	Емпіричні дані			Розрахунки		
2	*i*	*Прибуток*	*Зарплата*	x_i*y_i	y_i^2	x_i^2
3	1	94	84			
4	2	142	162			
5	3	128	136			
6	4	70	86			
7	5	86	100			
8	6	120	150			
9	7	76	94			
10	8	118	118			
11	9	134	138			
12	10	112	114			
13	Суми					
14	Середні					
15	Дисперсії					
16						
17	r_{xy}					
18						

Рис. 5.3

2. В комірку **D3** вводимо формулу **D3:=B3*C3** і поширюємо її вміст на весь стовпець **D3:D12**.
3. В комірку **E3** вводимо формулу **E3:= C3^2** і поширюємо її вміст на весь стовпець **E3:E12**.

4. В комірку **F3** вводимо формулу **F3: =B3^2** і поширюємо її вміст на весь стовпець **F3:F12**.
5. В комірки **B13, C13, D13, E13, F13** вводимо формули:
   B13: =СУММ(B3:B12),
   C13: =СУММ(C3:C12),
   D13: =СУММ(D3:D12),
   E13: =СУММ(E3:E12),
   F13: =СУММ(F3:F12).
6. В комірки **B14**, **C14**, **D14** вводимо формули **B14:=СРЗНАЧ(B3:B12)**,
   C14:=СРЗНАЧ(C3:C12),
   D14:=СРЗНАЧ(D3:D12).
7. В комірки **B15**, **C15** вводимо формули
   B15: =ДИСПР(B3:B12),
   D15: =ДИСПР(C3:C12).
8. В комірку **B17** вводимо формулу
   B17:=(D14-B14*C14)/(КОРЕНЬ(B15)*КОРЕНЬ(C15)).

Результати обчислень наведено на рисунку 5.4.

	A	B	C	D	E	F
1	Емпіричні дані			Розрахунки		
2	*i*	*Прибуток*	*Зарплата*	x_i *yi	y_i^2	x_i^2
3	1	94	84	7896	7056	8836
4	2	142	162	23004	26244	20164
5	3	128	136	17408	18496	16384
6	4	70	86	6020	7396	4900
7	5	86	100	8600	10000	7396
8	6	120	150	18000	22500	14400
9	7	76	94	7144	8836	5776
10	8	118	118	13924	13924	13924
11	9	134	138	18492	19044	17956
12	10	112	114	12768	12996	12544
13	Суми	1080	1182	133256	146492	122280
14	Середні	108	118,2	13325,6		
15	Дисперсії	564	677,96			
16						
17	r_{xy}	0,90562091				

Рис. 5.4

9. Результатом проведених обчислень є коефіцієнт кореляції. Його можна обчислити значно швидше за допомогою функції **=КОРРЕЛ**. В комірці **А21** вводимо формулу **А21:=КОРРЕЛ(В3:В12;С3:С12)**. Діалогове вікно цієї функції показано на рисунку 5.5.

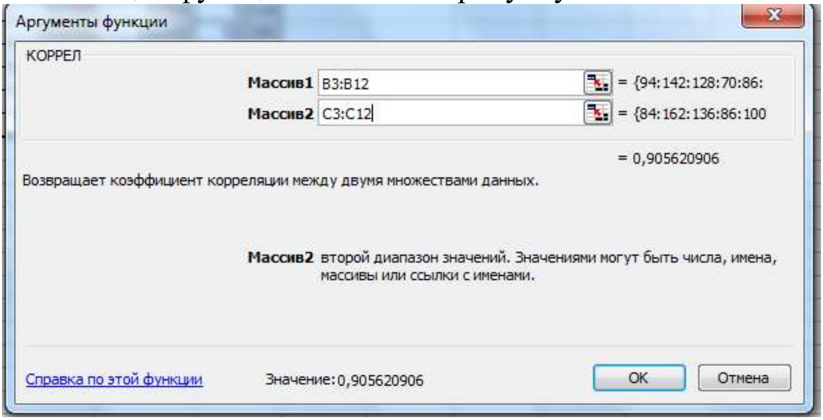


Рис. 5.5

Вибірковий коефіцієнт кореляції $\overline{r}_{xy}$ є точковою оцінкою коефіцієнта кореляції r_{xy} генеральної сукупності. Тому дуже важливо встановити, чи **коефіцієнт кореляції** r_{xy} **є значущим**. Оскільки вибірка є випадковою, то з рівності нулю вибіркового коефіцієнта кореляції $\overline{r}_{xy}$, взагалі кажучи, не випливає, що й коефіцієнт кореляції r_{xy} генеральної сукупності дорівнює нулю. У зв'язку з цим виникає потреба перевірити гіпотезу про значущість вибіркового коефіцієнта кореляції r_{xy}.

Якщо двовимірна генеральна сукупність (X,Y) розподілена за нормальним законом, то за критерій перевірки нульової гіпотези H_0 про рівність нулю коефіцієнта кореляції r_{xy} вибирають випадкову величину T:

$$T = \overline{r_{xy}} \frac{\sqrt{n-2}}{\sqrt{1-(\overline{r}_{xy})^2}},$$

де n – осяг вибірки. Випадкова величина Т має **розподіл Стьюдента** з $k = n - 2$ ступенями вільності. Нульова гіпотеза H_0 про рівність нулю вибіркового коефіцієнта кореляції нормальної двовимірної випадкової величини за конкуруючої гіпотези H_1 про те, що коефіцієнт кореляції r_{xy} не дорівнює нулю, перевіряється за таким правилом:

1. Обчислюємо емпіричне значення критерію $T_{емп}$:

$$T_{емп} = \overline{r_{xy}} \frac{\sqrt{n-2}}{\sqrt{1-\overline{r_{xy}}^2}}. \qquad (5.12)$$

2. Для заданого рівня значущості α і числа $k = n - 2$ ступенів вільності за таблицею критичних точок розподілу Стьюдента (додаток 4) визначаємо критичну точку двосторонньої критичної області.

3. Якщо $\left|T_{емп}\right| < t_{кр}$, то нульову гіпотезу H_0 про рівність нулю коефіцієнта кореляції r_{xy} генеральної сукупності приймають. Якщо ж $\left|T_{емп}\right| \ge t_{кр}$, то нульову гіпотезу H_0 відхиляють.

Якщо обсяг вибірки досить великий, а вибірка є репрезентативною, то висновок про характер лінійної залежності між випадковими величинами X та Y, отриманий на основі вибірки, може бути поширений і на генеральну сукупність. За умови, що обсяг вибірки $n > 100$, для інтервальної оцінки коефіцієнта кореляції можна використати формулу:

$$\overline{r_{xy}} - t_{кр} \frac{1-\overline{r_{xy}}^2}{\sqrt{n}} \le r_{xy} \le \overline{r_{xy}} + t_{кр} \frac{1-\overline{r_{xy}}^2}{\sqrt{n}}, \qquad (5.13)$$

де $t_{кр}$ - розв’язок рівняння $\Phi\left(t_{кр}\right) = \frac{1-\alpha}{2}$.

Перевіримо значущість коефіцієнта кореляції, обчисленого в прикладі 5.1.

В комірку **A21** вводимо назву «емпіричне значення критерію», і в комірку **D21** вводимо формулу **=B17*КОРЕНЬ((A12-2)/(1-B17^2))**.

В комірку **A22** вводимо назву «критична точка», і в комірку **D22** вводимо формулу =СТЬЮДРАСПОБР(0,05;A12-2).

Отримані результати бачимо на рис. 5.6.

21	Емпіричне значення критерію		6,04
22	Критична точка		2,31

Рис. 5.6

Оскільки емпіричне значення критерію є більшим за критичну точку, то приймаємо гіпотезу про відмінність від нуля коефіцієнта кореляції генеральної сукупності.

Приклад 5.2. Із двовимірної нормальної генеральної сукупності зроблено вибірку обсягом $n=122$ і обчислено вибірковий коефіцієнт кореляції $\overline{r_{xy}}=0,4$. Для рівня значущості $\alpha=0,05$ перевірити нульову гіпотезу $H_0: r_{xy}=0$ за конкуруючої гіпотези $H_1: r_{xy}\neq 0$.

Розв'язання. Гіпотезу H_0 перевіримо за правилом, яке сформульоване вище. За формулою (5.12) обчислюємо:

$$T_{емп}=0,4\frac{\sqrt{122-2}}{\sqrt{1-0,4^2}}=0,4\frac{\sqrt{120}}{\sqrt{0,84}}=4,8.$$

Для заданого рівня значущості $\alpha=0,05$ і числа ступенів вільності $k=122-2=120$ за таблицею додатка 4 знаходимо, що $t_{кр}=1,98$. Оскільки $T_{емп}=4,8>t_{кр}=1,98$, то нульову гіпотезу H_0 відхиляємо. Отже вибірковий коефіцієнт кореляції суттєво відмінний від нуля, тому випадкові величини X і Y корельовано.

Знайдемо довірчий інтервал для вибіркового коефіцієнта кореляції. Для його знаходження використаємо формулу (5.13) та таблицю додатка 4, з якої визначимо розв'язок рівняння $\Phi\left(t_{кр}\right)=0,475$ і отримаємо: $z_{кр}=1,96.$ Підставивши ці дані в формулу (5.13), матимемо:

$$0,4-1,96\frac{1-0,16}{\sqrt{122}}\le r_{xy}\le 0,4+1,96\frac{1-0,16}{\sqrt{122}},$$

або остаточно:

$$0,25<r_{xy}<0,55.$$

5.3. Основні поняття і методи регресійного аналізу

Якщо **кореляційний аналіз** досліджує наявність і характер зв'язку між випадковими величинами X і Y, то **регресійний аналіз** встановлює **аналітичну форму цієї залежності**.

Нехай X незалежна змінна (факторна ознака), а Y залежна змінна (результативна ознака) і припустимо, що:

- розподіл результативної ознаки Y генеральної сукупності нормальний;
- дисперсія результативної ознаки Y не залежить від факторної ознаки X;
- між результативною та факторною ознаками існує лінійний зв'язок.

Ці обмеження приводять до дослідження найпростішої **регресійної моделі — лінійної регресії,** коли вибіркове рівняння регресії має вигляд:

$$\bar{y}_x=ax+b. \tag{5.14}$$

В цьому випадку точкові оцінки параметрів a і b справджують основні вимоги до точкових оцінок, які описані в підрозділі 2.2, а тому для них можна побудувати довірчі інтервали та оцінити їх значущість.

Основним методом отримання точкових оцінок для параметрів і рівняння регресії (5.14) є **метод найменших квадратів.**

Нехай задана вибірка (x_1, y_1), (x_2, y_2), ... (x_n, y_n) обсягу n з діаграмою розсіяння, як на рис. 5.1. Ідея методу найменших квадратів полягає в тому, що за точкові оцінки $\overline{a}$ і $\overline{b}$ параметрів a і b вибирають такі числа, для яких пряма $\overline{y}_x = \overline{a}x + \overline{b}$ є «найближчою» до точок (x_1, y_1), (x_2, y_2), ... (x_n, y_n). За міру відхилення шуканої прямої від точок (x_i, y_i) виберемо величину:

$$S(a,b) = \sum_{i=1}^{n} \left[y_i - (ax_i + b) \right]^2, \qquad (5.15)$$

За точкові оцінки $\overline{a}$ і $\overline{b}$ параметрів a і b рівнянні регресії (5.14) виберемо такі числа, для яких функція $S(a,b)$ з (5.15) набуває мінімального значення. Метод знаходження таких оцінок параметрів a і b, які мінімізують функцію $S(a,b)$, називають **методом найменших квадратів.**

Для знаходження точкових оцінок $\overline{a}$ і $\overline{b}$ невідомих параметрів a і b запишемо систему рівнянь:

$$\begin{cases} \dfrac{\partial S(a,b)}{\partial a} = 2\sum_{i=1}^{n} \left[y_i - (ax_i + b) \right](-x_i) = 0 \\ \dfrac{\partial S(a,b)}{\partial b} = (-2)\sum_{i=1}^{n} \left[y_i - (ax_i + b) \right] = 0, \end{cases} \qquad (5.16)$$

яку елементарними перетвореннями зведемо до вигляду:

$$\begin{cases} a\left(\sum_{i=1}^{n} x_i^2\right) + b\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} x_i y_i, \\ a\left(\sum_{i=1}^{n} x_i\right) + b \cdot n = \sum_{i=1}^{n} y_i. \end{cases} \qquad (5.17)$$

Оскільки визначник лінійної відносно невідомих a і b системи рівнянь (5.17) є додатним

$$\begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix} = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 > 0,$$

то система (5.17) має єдиний розв'язок:

$$a = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{\sigma}_x^2}, \quad b = \overline{y} - a \cdot \overline{x}, \qquad (5.18)$$

де

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

$$\overline{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\overline{x}\right)^2, \quad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

Підставивши значення a і b з (5.18) в рівняння (5.14), отримаємо шукане **рівняння лінійної регресії :**

$$\overline{y}_x = ax + b.$$

Коефіцієнт $a = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{\sigma}_x^2}$ називається **коефіцієнтом регресії.**

Зауважимо, що лінійне рівняння регресії можна подати також через точкову оцінку коефіцієнта кореляції, а саме:

$$\overline{y}_x - y = r_{xy}^* \frac{\overline{\sigma}_y}{\overline{\sigma}_x}\left(x - \overline{x}\right). \qquad (5.19)$$

Зауваження. У разі, коли припущення про лінійність зв’язку між ознаками Y та X не справджується, то будують нелінійні регресійні моделі. Ці моделі виражаються, наприклад, рівняннями:

$$\bar{y}_x = ax^2 + bx + c,$$

$$\bar{y}_x = ax^3 + bx^2 + cx + d,$$

$$\bar{y}_x = \frac{a}{x} + b,$$

$$\bar{y}_x = ae^{bx.}$$

Точкові оцінки параметрів у цих нелінійних моделях також можна знайти методом найменших квадратів.

Приклад 5.3. Залежність між обсягом X (тис.грн) товару, який перевозиться через митний пост за кордон, і відсотком Y (%) не задекларованої частини цього обсягу характеризується вибіркою:

$x_i \setminus y_i$	0	5	10	15	20	25
45	4	2	2	1	5	
55	3	5	5	4	7	4
65	2	5	4	5	8	2
75	7	3	3	4	1	3
85	5	2	3	2	2	3
95	1	6	2	1	5	4

Визначити:

а) емпіричний закон розподілу системи випадкових величин X та Y;

б) точкові оцінки чисельних характеристик випадкових величин X та Y;

в) вибірковий коефіцієнт кореляції;

г) при рівні значущості $\alpha = 0,01$ перевірити гіпотезу про статистичну значимість коефіцієнта кореляції.

Розв'язання

а) Знайдемо емпіричний закон розподілу системи випадкових величин у вигляді таблиці:

$x_i \setminus y_i$	0	5	10	15	20	25	n_{x_j}
45	4	2	2	1	5		14
55	3	5	5	4	7	4	28
65	2	5	4	5	8	2	26
75	7	3	3	4	1	3	21
85	5	2	3	2	2	3	17
95	1	6	2	1	5	4	19
n_{y_i}	22	23	19	17	28	16	125

б) за формулами (5.3)-(5.6) знайдемо точкові оцінки чисельних характеристик випадкових величин X та Y. Маємо:

$$\overline{y} = \frac{1}{125}(0 \cdot 22 + 5 \cdot 23 + 10 \cdot 19 + 15 \cdot 17 + 20 \cdot 28 + 25 \cdot 16) = 12{,}16;$$

$$\overline{x} = \frac{1}{125}\left(45 \cdot 14 + 55 \cdot 28 + 65 \cdot 26 + 75 \cdot 21 + 85 \cdot 17 + 95 \cdot 19\right) = 69{,}48;$$

$$\overline{\sigma}_x^2 = \frac{1}{125}\left(0^2 \cdot 22 + 5^2 \cdot 23 + 10^2 \cdot 19 + 15^2 \cdot 17 + 20^2 \cdot 28 + 25^2 \cdot 16\right) -$$

$$-\left(12{,}6\right)^2 = 72{,}13;$$

$$\overline{\sigma}_y^2 = \frac{1}{125}\left(45^2 \cdot 14 + 55^2 \cdot 28 + 65^2 \cdot 26 + 75^2 \cdot 21 + 85^2 \cdot 17 + 95^2 \cdot 19\right) -$$

$$-\left(69.48\right)^2 = 255{,}13;$$

$$\overline{xy} = \frac{1}{125}(45\left(0\cdot 4+5\cdot 2+10\cdot 2+15\cdot 1+20\cdot 5+25\cdot 0\right)+$$

$$+55\left(0\cdot 3+5\cdot 5+10\cdot 5+15\cdot 4+20\cdot 7+25\cdot 4\right)+$$

$$+65\left(0\cdot 2+5\cdot 5+10\cdot 4+15\cdot 5+20\cdot 8+25\cdot 2\right)+$$

$$+75\left(0\cdot 7+5\cdot 3+10\cdot 3+15\cdot 4+20\square 1+25\cdot 3\right)+$$

$$+85\left(0\cdot 5+5\cdot 2+10\cdot 3+15\cdot 2+20\cdot 2+25\cdot 3\right)+$$

$$+95\left(0\cdot 1+5\cdot 6+10\cdot 2+15\cdot 1+20\cdot 5+25\cdot 4\right)) = 846,4.$$

в) вибірковий коефіцієнт кореляції обчислимо за формулою (5.9):

$$\overline{r}_{xy} = \frac{864,4-12,16\cdot 69,48}{8,49\cdot 15,97} = 0,011.$$

г) перевіряємо гіпотезу про статистичну значимість коефіцієнта кореляції. За формулою (5.12) обчислюємо $T_{åìï}$.

$$T_{емп} = 0,011\frac{\sqrt{125-2}}{\sqrt{1-0,011^2}} = 0,1245.$$

Для заданого рівня значущості $\alpha = 0,05$ і числа ступенів вільності $k = 125-2 = 123$ за таблицею додатка 4 знаходимо, що $t_{кр} = 1,97$. Оскільки $T_{емп} = 0,1245 < t_{кр} = 1,97$, то нульова гіпотеза H_0 про некорельованість випадкових величин X, Y приймається.

Розглянемо розв’язування задачі засобами Exel.

Числові дані у вигляді таблиці спряженості вводимо в робочий лист у вигляді парного статистичного розподілу.

1. В комірку **A1** вводимо назву «Обсяг товару, x_i», в комірку **B1** вводимо назву « Незадекларована частина, y_i».
2. В комірку **A2** вводимо число 45, в комірку **B2** вводимо число 0. Оскільки в таблиці спряженості пара (45; 0) має частоту 4, то вміст комірок **A2:B2** за допомогою маркера заповнення поширюємо на діапазон **A3:B5**. Пара (55; 0) має частоту 3, тому

комірки **A6:B8** заповнюємо цією парою чисел, і т.д. Таким чином в діапазоні комірок **A2:B126** утворюється парний статистичной розподіл. Фрагмент бази даних показано на рис. 5.7.

	A	B
1	Обсяг товару, *xi*	Незаде-кларова-на частина, *yi*
2	45	0
3	45	0
4	45	0
5	45	0
6	55	0
7	55	0
8	55	0
9	65	0
10	65	0
11	75	0
12	75	0
13	75	0
14	75	0
15	75	0
16	75	0

Рис. 5.7

3. Щоб візуально оцінити наявність чи відсутність лінійного кореляційного зв'язку, побудуємо діаграму розсіювання. Заходимо в меню **Вставка →Діаграма→Точкова** (рис. 5.8), (рис. 5.9)

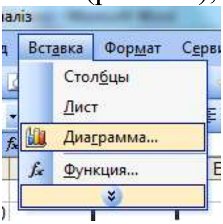


Рис. 5.8

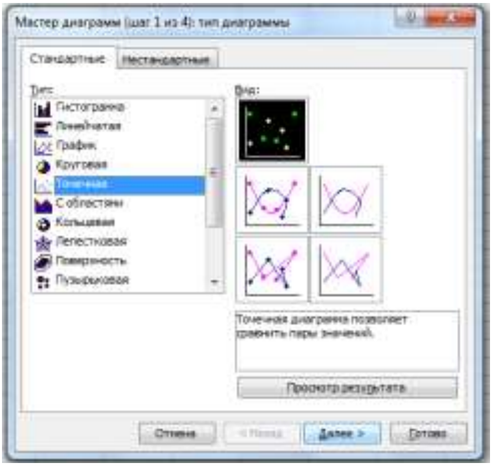


Рис. 5.9

В діалоговому вікні вводимо діапазон даних (рис. 5.10)

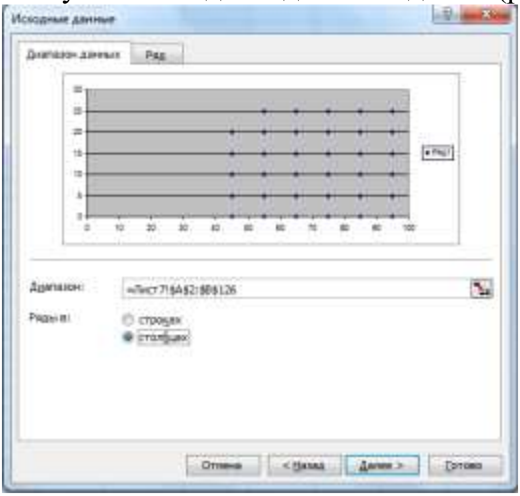


Рис. 5.10

Натиснувши **Далее** відкриємо діалогове вікно, в якому вводимо назву діаграми та назви осей (рис. 5.11)

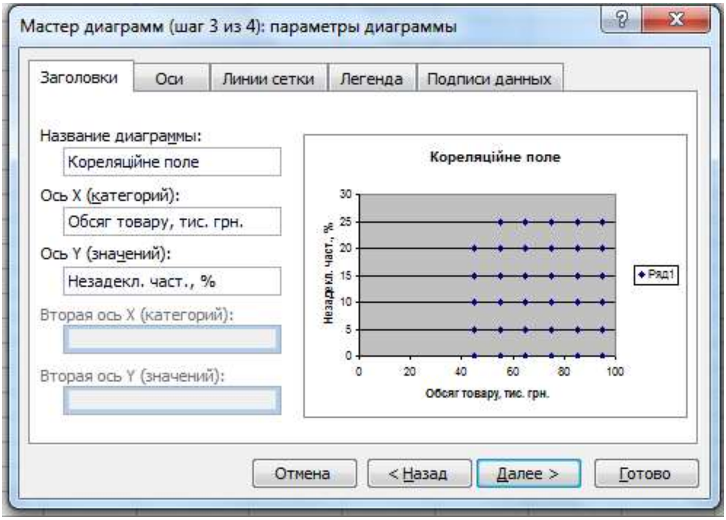


Рис. 5.11

Натискаємо **Далее** і у відкритому діалоговому вікні вибираємо місце розміщення діаграми, натискаємо **Готово** і одержуємо в робочому листі кореляційне поле (рис. 5.12).

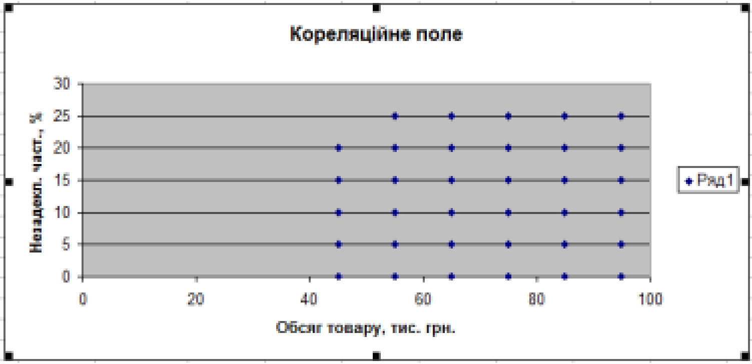


Рис. 5.12

З розташування точок кореляційного поля видно, що лінійний кореляційний зв'язок відсутній. Обґрунтуємо цю гіпотезу числовими розрахунками.

4. Для зручності в роботі перемістимо діаграму з кореляційним полем вниз робочого листа. В комірку **C1** вводимо назву «*xy*». В комірку **C2** вводимо

формулу. **C2:=A2*B2** і поширюємо її вміст в комірки **C2:C126**.

5. Сформуємо табличку для обчислень. В комірку **E2** вводимо назву «*середні*», в комірку **E3** вводимо назву «*дисперсії*», в комірку **E4** вводимо назву «*середньоквадратичні відхилення*», в комірку **E5** вводимо назву «*кореляція*». В комірку **H1** вводимо назву «*x*», в комірку **I1** вводимо назву «*y*», в комірку **J1** вводимо назву «*xy*».
6. Вводимо формули:
   H2: =СРЗНАЧ(A2:A126);
   I2:=СРЗНАЧ(B2:B126);
   J2:=СРЗНАЧ(C2:C126);
   H3:=ДИСПР(A2:A126);
   I3:=ДИСПР(B2:B126);
   H4:=СТАНДОТКЛОНП(A2:A126);
   I4:=СТАНДОТКЛОНП(B2:B126);
   H5:=(J2-H2*I2)/(H4*I4).

В комірці **H5** обчислено коефіцієнт кореляції за формулою (5.9), що потребувало додаткових обчислень. Ці обчислення виконаємо за допомогою функції

I5:=КОРРЕЛ(A2:A126;B2:B126).

Перевіримо гіпотезу про значущість коефіцієнта кореляції.

H_0: коефіцієнт кореляції не значущий (лінійний зв'язок відсутній).

H_1: коефіцієнт кореляції значущий (між ознаками спостерігається лінійний кореляційний зв'язок).

7. В комірку **E8** вводимо назву «емпіричне значення критерію», в комірку **E9** вводимо назву «критична точка». Вводимо формули:

H8:=H5*КОРЕНЬ(СЧЁТ(A2:A126)-2)/КОРЕНЬ(1-H5^2);

H9:=СТЬЮДРАСПОБР(0,05;СЧЁТ(A2:A126)-2).

Результати обчислень видно на рис. 5.13

	E	F	G	H	I	J
				x	y	xy
	середні			69,48	12,16	846,4
	дисперсії			255,1296	72,1344	
	середньоквадратичні відхилення			15,97278	8,493197	
	кореляція			0,011228	0,011228	
	емпіричне значення критерію			0,124533		
	критична точка			1,979439		

Рис. 5.13

Емпіричне значення критерію є меншим за критичну точку, приймаємо нульову гіпотезу і робимо висновок, що між величинами відсутній лінійний кореляційний зв'язок.

Розглянемо задачу 5.3 тільки з іншими числовими даними.

Приклад 5.4. Залежність між обсягом X (тис. грн) товару, який перевозиться через митний пост за кордон, і відсотком $Y\,(\%)$ не задекларованої частини цього обсягу характеризується вибіркою:

$x_i \setminus y_i$	0	5	10	15	20	25
45	5	0	5	0	0	0
55	0	10	5	4	0	0
65	0	2	5	1	4	0
75	0	0	7	10	5	0
85	0	0	0	9	11	0
95	0	0	0	8	5	14

Визначити:

а) вибірковий коефіцієнт кореляції;

б) при рівні значущості $\alpha = 0,05$ перевірити гіпотезу про статистичну значимість коефіцієнта кореляції;

в) написати рівняння парної лінійної регресії.

Розв’яжемо задачу засобами Exel.

1. В новому робочому листі в комірках **A1:B108** формуємо парний статистичний розподіл, так як це описано в п. 1-2 попередньої задачі (рис. 5.14).

	A	B
1	xi	yi
2	45	0
3	45	0
4	45	0
5	45	0
6	45	0
7	55	5
8	55	5
9	55	5
10	55	5
11	55	5
12	55	5
13	55	5
14	55	5
15	55	5
16	55	5
17	65	5
18	65	5
19	45	10
20	45	10
21	45	10
22	45	10
23	45	10
24	55	10
25	55	10

Рис. 5.14

2. Будуємо кореляційне поле користуючись вказівками п. 3 попередньої задачі (рис. 5.15).

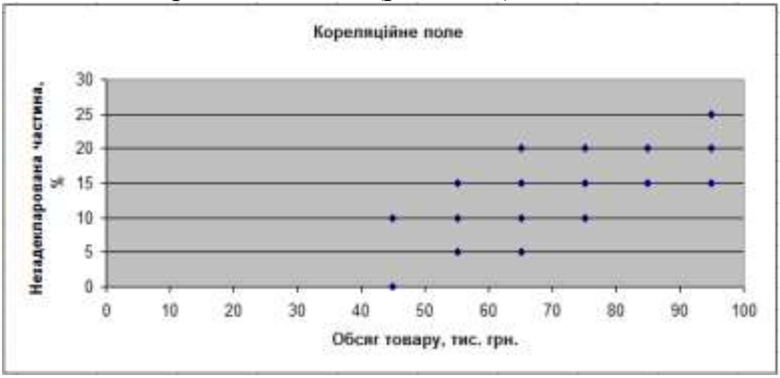


Рис. 5.15

   З вигляду кореляційного поля можна зробити припущення, що між величинами спостерігається лінійна залежність.

3. Обчислюємо коефіцієнт кореляції. В комірку **D2** вводимо назву «**коефіцієнт кореляції**» в комірку **F2** вводимо формулу **F2:=КОРРЕЛ(A2:A108;B2:B108)** (рис. 5.16)

D	E	F
коефіцієнт кореляції		0,783795

Рис. 5.16

4. Перевіримо гіпотезу про значущість коефіцієнта кореляції. Сформулюємо гіпотези

H_0: коефіцієнт кореляції не значущий (лінійний зв'язок відсутній).

H_1: коефіцієнт кореляції значущий (між ознаками спостерігається лінійний кореляційний зв'язок).

Обчислюємо емпіричне значення критерію і знаходимо критичну точку. Вводимо в комірку **D3** назву «**емпіричне значення критерію**», в комірку **D4** вводимо назву «**критична точка**». В комірку **G3** вводимо формулу

G3:=F2*КОРЕНЬ(СЧЁТ(A2:A108)-2)/КОРЕНЬ(1-F2^2),

в комірку G4 вводимо формулу

G4:=СТЬЮДРАСПОБР(0,05;СЧЁТ(A2:A108)-2).(Рис. 5.17)

D	E	F	G
коефіцієнт кореляції		0,783795	
емпіричне значення критерію			13,11621
критична точка			1,982173

Рис. 5.17

Бачимо, що емпіричне значення критерію потрапляє в критичну область і робимо висновок про значущість коефіцієнта лінійної кореляції і наявність зв’язку між величинами.

5. Оскільки коефіцієнт кореляції значущий, доцільно обчислити коефіцієнти парної лінійної регресії. Для цього проведемо додаткові обчислення. В комірку **C1** вводимо назву «***xy***», в комірку **C2** вводимо формулу **C2:=A2*B2** і поширюємо її вміст на весь стовпець **C2:C108**. В комірку **I2** вводимо назву «**середн**і», в комірку **I3** вводимо назву «**дисперсії**», в комірку **J1** вводимо назву «***x***», в комірку **К1** вводимо назву «***y***», в комірку **L1** вводимо назву «***xy***».

Вводимо формули:

J2:=СРЗНАЧ(A2:A108),
К2:=СРЗНАЧ(B2:B108),
L2:=СРЗНАЧ(C2:C108);
J3:=ДИСПР(A2:A108);
К3:=ДИСПР(B2:B108)

6. В комірку **D6** вводимо назву «**коефіцієнти регресії**» в комірку **F6** вводимо назву «***a***», в комірку **G6** вводимо назву «***b***». Вводимо формули

F7:= (L2-J2*K2)/J3;
G7: =K2-F7*J2.

Результати обчислень показано на рис. 5.18.

D	E	F	G	H	I	J	K	L
						x	y	xy
коефіцієнт кореляції		0,783795			середні	74,45455	14,63636	1176,818
емпіричне значення критерію			13,11621		дисперсії	277,8843	44,41322	
критична точка			1,982173					
Коефіцієнти регресії		a	b					
		0,313348	-8,69379					

Рис. 5.18

7. Коефіцієнти регресії можна обчислити швидше за допомогою функції. Для цього виділяємо комірки **F8:G8** і викликаємо функцію **=ЛИНЕЙН()**. Заповнюємо діалогове вікно цієї функції як показано на рис. 5.19.

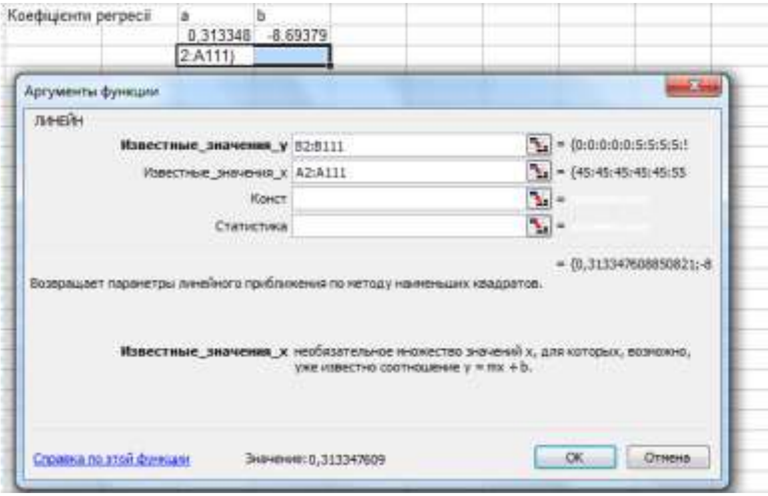


Рис. 5.19

Після цього виставляємо курсор в командну стрічку і натискаємо комбінацію клавіш **CTRL+SHIFT+ENTER**. В комірках **F8:G8** отримуємо коефіцієнти регресії $y = ax + b$. В нашому прикладі рівняння регресії має вигляд $y = 0,31x - 8,69$.

Рівняння регресії можна отримати безпосередньо на діаграмі, побудованій в п. 2. Для цього правою клавішею миші клікаємо на будь-якій точці кореляційного поля після чого з'явиться меню, в якому вибираємо «Додати лінію тренду» (рис. 54), тип лінії «**Лінійна**» і відкриваємо вкладку «**Параметри**», в якій ставимо мітку навпроти команди «**Показувати рівняння на діаграмі**» (рис. 5.20).

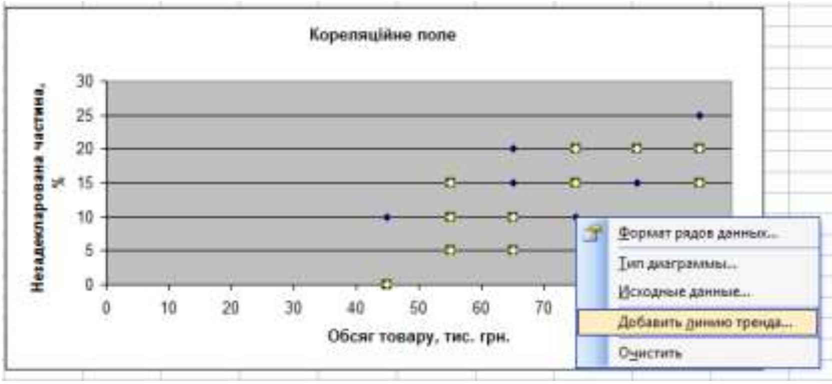


Рис. 5.20

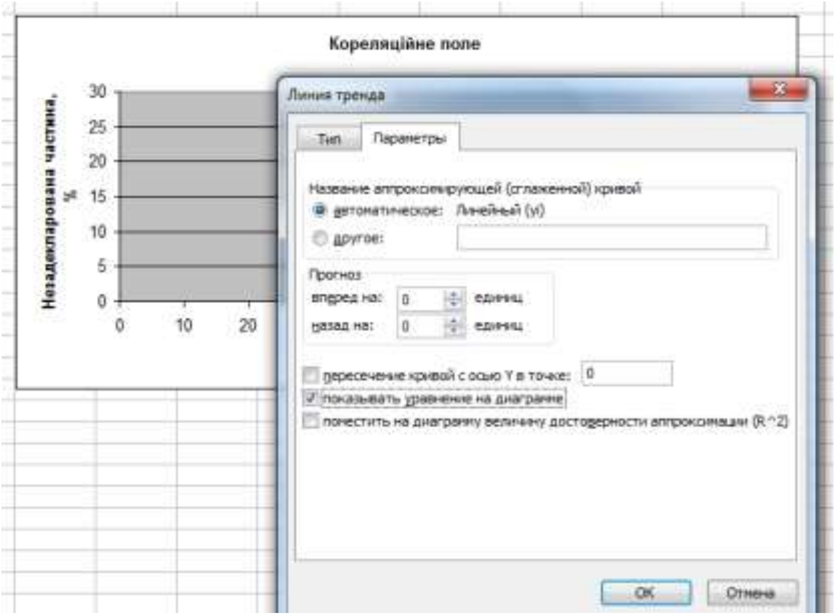


Рис. 5.21

На діаграмі з’явиться лінія тренду і її рівняння (рис. 5.22).

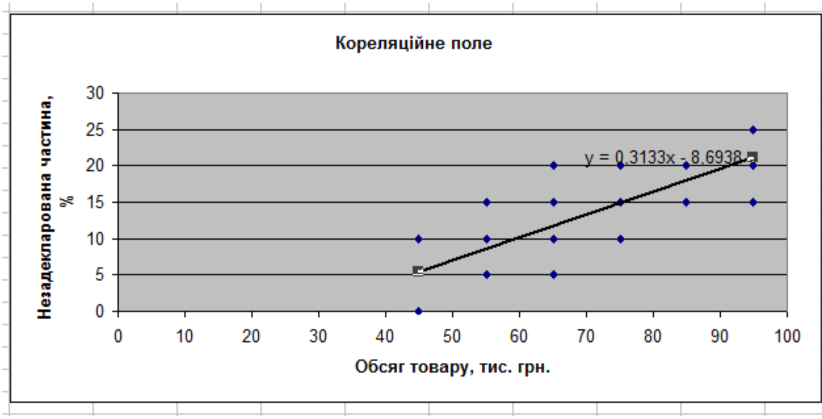


Рис. 5.22

Індивідуальні завдання №1

Для заданих вибірок виконати такі завдання.
а) побудувати інтервальний статистичний розподіл частот;
б) побудувати дискретний статистичний розподіл;
в) побудувати розподіл відносних частот, накопичених частот та накопичених відносних частот.
г) побудувати гістограму частот;
д) побудувати полігон частот;
е) побудувати полігон накопичених частот;
є) обчислити числові характеристики вибіркової сукупності.

1. Проміжок часу (у секундах), витрачений кожним із 124 працівників на вирішення завдання:

52	62	69	129	75	65	11	41	22	27	52	46	106	14	49
73	84	73	47	81	193	119	87	17	24	55	37	56	62	131
69	66	47	60	76	71	91	104	61	59	55	31	52	61	45
53	42	47	53	25	48	87	85	30	40	85	49	54	24	52
33	51	49	42	54	30	53	32,8	58	37	42	38	28	23	24
28	40	41	29	39	28	30	25	30	23	23	35	32	34	21
39	10	23	22	42	27	39	39	46	60	102	22	44	90	53
98	67	49	142	71	30	41	144	50	28	28	27	38	40	35

2**.** Час локалізації пожеж наведено в таблиці:

15	20	14	13	10	9	24	17	11	16
12	15	16	5	15	15	7	9	25	23
20	16	14	17	21	15	10	19	22	25
18	16	10	20	24	11	15	15	17	15
10	11	22	22	22	19	17	20	14	11
14	12	24	16	9	23	13	11	18	17
18	23	22	19	23	26	19	14	22	19
17	21	21	18	16	20	21	15	23	16
15	24	18	14	18	12	20	16	15	20
22	23	16	15	14	5	20	22	20	22

3. У досліді з вивчення амплітудно-частотної характеристики руки людини-оператора значення амплітуд усталених коливань руки одного з досліджуваних становлять:

80	65	62	73	62	71	63	75	66	67	61	64	72	60	67
62	52	68	72	63	66	62	70	64	71	57	55	56	64	61
59	64	65	60	63	59	60	68	66	53	73	67	60	68	60
70	70	63	63	62	62	68	63	67	62	62	58	62	63	55
54	64	65	70	65	58	52	66	58	66	72	69	71	58	60
63	65	60	75	51	64	65	69	65	68	58	55	60	58	57
61	45	46	60	64	67	64	72	58	64	70	73	52	59	69

4. У досліді з вивчення амплітудно-частотної характеристики руки людини-оператора значення амплітуд усталених коливань руки одного з досліджуваних становлять:

56	55	60	54	59	71	63	55	55	58	66	62	82	54	74
55	62	75	62	70	64	63	65	70	65	56	58	56	65	62
62	60	63	68	64	60	71	61	62	55	51	44	73	52	66
62	60	57	60	58	62	58	52	62	61	61	67	60	53	50
42	57	51	66	51	56	55	56	54	53	41	55	54	58	51
58	71	65	62	65	73	66	55	68	62	60	67	57	62	62
49	59	48	52	54	58	62	62	51	62.					

5. Результати опитування групи працівників (80 осіб) за тестом Томаса поведінки в конфліктній ситуації (суперництво), мають такий вигляд:

6	4	4	0	4	3	4	0	8	3	1	2	0	4	0	9	4	0	3	12
2	5	8	0	7	8	3	12	0	0	1	5	5	1	10	2	10	11	3	1
1	1	1	11	10	4	4	7	1	0	3	3	8	1	6	1	6	6	4	9
0	3	5	4	0	1	2	8	5	4	3	2	6	1	8	0	7	1	4	1

6. Результати опитування групи працівників (80 осіб) за тестом Томаса поведінки в конфліктній ситуації (співробітництво), мають такий вигляд:

6	5	5	9	9	6	8	6	8	6	3	9	7	3	7	8	7	5	7	7
10	5	11	10	4	10	6	5	3	4	10	8	6	7	6	7	3	10	5	3
10	7	8	8	7	7	7	5	3	9	8	9	6	7	7	9	10	8	8	7
4	5	10	9	8	6	8	8	10	5	9	8	8	11	6	10	7	6	8	5

7 Результати опитування групи працівників (80 осіб) за тестом Томаса поведінки в конфліктній ситуації (компроміс), мають такий вигляд:

7	8	10	8	9	4	11	9	8	7	10	10	9	7	8	8	8	4	7	6
6	6	5	7	6	8	6	11	10	4	8	8	8	7	6	5	4	5	9	8
9	8	7	6	7	2	6	6	8	5	8	8	7	8	7	6	6	8	4	8
6	8	6	9	5	8	9	5	9	8	9	6	11	8	10	7	10	7	8	9

8. Результати опитування групи працівників (80 осіб) за тестом Томаса поведінки в конфліктній ситуації (уникнення) мають такий вигляд:

5	8	4	6	4	10	3	9	4	6	5	4	9	9	8	2	5	8	4	9
3	4	4	5	8	5	3	2	6	3	4	6	8	4	3	8	6	7	3	4
3	4	5	5	5	4	5	4	7	4	5	6	9	4	7	10	6	7	6	4
8	11	8	5	6	6	8	7	8	7	3	4	2	5	3	5	2	7	5	3

9. Результати опитування групи працівників (80 осіб) за тестом Томаса поведінки в конфліктній ситуації (пристосування), мають такий вигляд:

9	5	3	4	10	4	3	6	11	4	4	4	9	7	8	10	4	8	8	8
2	8	3	4	6	8	4	6	5	8	8	7	11	10	9	7	6	3	4	3
5	8	10	10	6	2	11	7	5	8	7	5	9	1	4	2	6	5	5	4
9	2	2	8	6	7	2	6	3	9	4	3	7	7	6	8	7	4	4	7

10. Результати опитування групи працівників стосовно творчої поведінки (конформізм), мають такий вигляд:

13	19	12	16	13	9	16	10	22	16	6	14	12	17	17	17
15	17	16	5	15	22	14	10	13	18	6	17	17	6	12	14
5	21	19	14	18	23	15	15	8	11	6	15	11	1	8	6
9	10	14	5	16	12	18	16	18	4	18	17	19	16	17	14
17	14	12	16	9	20	7	15	12	8	19	20				

11. Результати опитування групи працівників стосовно творчої поведінки (алгоритмічна поведінка), мають такий вигляд:

16	14	15	17	15	15	17	15	18	14	18	20	15	17	18	12
16	25	20	11	19	18	15	16	15	23	15	12	17	20	20	14
12	14	16	16	22	19	20	17	18	17	19	14	17	17	14	15
17	19	14	17	16	17	21	18	19	16	20	19	23	20	14	20
19	14	18	22	15	16	18	25	18	19	13	12	14	11	13	19

12. Результати опитування групи працівників стосовно творчої поведінки (нонконформізм), мають вигляд:

20	17	17	28	17	14	22	18	19	17	23	20	19	20	15	20
20	25	21	20	23	18	19	33	19	17	18	22	19	26	24	19
23	13	22	18	20	14	19	24	18	26	24	19	17	17	19	22
17	18	24	18	16	21	23	21	26	20	19	20	23	19	16	21
19	28	20	19	25	18	20	19	22	21	21	21	22	21	16	19
22	17	21	25												

13. Результати опитування групи працівників стосовно творчої поведінки (евристична поведінка), мають такий вигляд:

22	22	16	24	20	15	24	18	24	23	21	14	18	19	21	21
16	21	21	18	15	19	24	23	22	24	15	19	16	18	18	17
23	16	17	16	16	14	16	20	20	20	21	13	16	16	17	17
21	22	17	18	19	18	16	17	22	15	23	23	17	23	24	17
19	27	22	14	22	20	21	21	22	23	19	17	13	18	15	15

14. Результати опитування групи працівників стосовно творчої поведінки (спрямованість на відтворення), мають такий вигляд:

25	24	29	29	20	31	29	33	34	32	22	38	32	36	34	24
32	36	30	35	21	32	43	32	29	29	38	25	40	37	20	30
17	35	35	44	36	37	43	32	33	25	30	20	32	28	15	24
29	44	32	26	35	31	24	32	25	45	31	18	31	32	37	40
36	22	37	42	29	33	35	39	30	36	21	19	29	23	19	33

15. Результати опитування групи працівників стосовно творчої поведінки (творча спрямованість), мають такий вигляд:

42	39	33	52	37	29	46	36	43	40	44	34	37	39	36	43
33	39	45	36	31	40	47	44	48	44	34	39	39	37	34	34

46	29	39	34	36	28	35	44	38	46	45	32	33	33	36	38
41	47	38	38	42	36	35	50	41	32	41	45	36	49	48	42
38	55	42	33	47	38	41	40	44	44	40	38	35	39	31	34

16. Результати опитування групи працівників стосовно творчої поведінки (вік), мають такий вигляд:

35	32	33	47	31	35	28	46	35	28	32	34	36	35	35	27
36	23	61	43	57	36	47	59	38	45	40	53	50	53	55	38
37	37	37	41	49	36	42	30	45	44	29	41	36	34	43	38
40	27	39	37	35	33	48	33	29	58	35	36	33	40	37	37
31	34	46	52	32	33	40	36	47	31	35	25	45	27	30	44

17. Результати опитування групи працівників стосовно творчої поведінки (стаж), мають такий вигляд:

12	12	11	21	12	15	7	23	14	4	20	15	17	15	15	3
17	4	40	24	34	16	23	37	14	24	21	31	31	36	36	16
20	17	17	22	26	15	23	10	23	24	9	23	14	14	20	16
19	7	19	15	15	12	26	14	8	40	16	15	13	17	18	17
12	14	17	31	13	12	12	18	29	12	15	6	23	7	10	21

18. Результати опитування групи працівників стосовно творчої поведінки (конформізм), мають такий вигляд:

13	6	18	15	17	12	19	20	11	19	21	14	17	21	16	16
23	14	17	14	13	16	22	14	15	17	9	4	22	13	15	14
2	9	11	14	16	16	8	17	15	8	15	4	17	12	8	17
10	8	10	12	21	17	12	16	11	10	11	9	18	9	17	18
11	19	10	5	15	17	5	11	18	8	14	6	11	22	13	18

19. Результати опитування групи працівників стосовно творчої поведінки (алгоритмічна поведінка), мають такий вигляд:

22	14	22	21	13	20	17	23	20	22	24	18	15	21	16	18
19	19	15	14	17	22	16	20	23	21	17	16	19	16	20	17
16	12	14	15	18	20	18	19	18	9	19	13	21	15	16	21
16	18	20	17	21	21	19	23	15	9	14	12	19	18	21	20
13	15	15	13	21	19	18	15	21	20	11	18	19	25	19	22

20. Результати опитування групи працівників стосовно творчої поведінки (нонконформізм), мають такий вигляд:

20	20	21	17	22	20	22	25	22	22	23	20	22	20	19	27
13	20	19	18	19	15	17	18	23	19	23	19	19	22	16	20
20	23	18	18	21	24	17	18	22	18	19	16	12	24	23	27
21	22	24	21	21	27	19	24	26	19	23	21	26	18	23	23
20	26	25	21	14	23	17	15	10	16	26	24	19	21	15	21

21. Результати опитування групи працівників стосовно творчої поведінки (евристична поведінка), мають такий вигляд:

20	24	25	15	23	22	15	22	20	19	21	19	17	16	10	23
15	17	19	21	19	18	15	19	29	23	19	15	18	21	11	22
23	21	21	16	17	23	19	20	20	15	20	20	12	22	20	28
21	18	19	22	21	22	19	24	28	19	18	19	21	19	26	25
21	23	23	19	15	15	19	11	19	17	22	19	18	20	20	14

22. Результати опитування групи працівників стосовно творчої поведінки (спрямованість на відтворення), мають такий вигляд:

35	20	40	36	30	32	36	43	31	41	45	32	32	42	32	34
42	33	32	28	30	38	38	34	38	38	26	20	41	29	35	31
18	21	25	29	34	36	26	36	33	17	34	17	38	27	24	38
26	26	30	29	42	38	31	39	26	19	25	21	37	27	38	38
24	34	25	18	36	36	23	26	39	28	25	24	30	47	32	40

23. Результати опитування групи працівників стосовно творчої поведінки (творча спрямованість), мають такий вигляд:

40	44	46	32	45	42	37	47	42	41	44	39	39	36	29	50
28	37	38	39	38	33	32	37	52	42	42	34	37	43	27	42
43	44	39	34	38	47	36	38	42	33	39	36	24	46	43	55
42	40	43	43	42	49	38	48	54	38	41	40	47	37	49	48
41	49	48	40	29	38	36	26	29	33	48	43	37	41	35	35

24. Результати опитування групи працівників стосовно творчої поведінки (вік), мають такий вигляд:

33	31	49	40	30	59	47	50	33	60	55	61	53	57	25	24
43	38	46	30	44	24	48	43	34	35	28	30	49	36	28	40
46	38	34	36	41	37	28	47	64	30	35	39	38	34	40	51
44	36	51	59	56	59	52	54	63	51	53	37	53	37	37	30
30	34	44	25	34	41	51	37	48	36	44	57	35	47	49	41

25. Результати опитування групи працівників стосовно творчої поведінки (стаж), мають такий вигляд:

14	16	21	23	9	35	25	23	15	40	35	44	27	44	5	16
24	19	27	8	24	3	28	14	14	11	9	9	24	13	4	20
26	16	12	17	21	18	5	25	43	5	15	19	18	11	17	28
17	13	29	40	30	40	32	33	45	30	12	13	20	12	15	9
8	14	21	5	14	19	20	12	22	19	22	38	12	25	25	14

26. Результати тестування групи молоді (64 чоловік) за шкалою, що характеризує рівень їх суспільно-політичної свідомості, мають такий вигляд:

6	7	7	6	8	6	5	7	5	8	7	7	6	5	6	6
5	7	7	7	6	5	6	6	5	7	6	9	8	6	7	4
4	7	6	7	6	7	7	5	6	7	6	6	9	6	6	5
6	5	7	7	7	6	6	8	5	7	9	7	6	6	6	5

27. Результати тестування групи молоді (64 чоловік) за шкалою, що характеризує рівень їх духовної зрілості, мають такий вигляд:

7	6	5	5	5	5	6	6	6	5	5	6	5	5	5	6
5	6	6	6	7	7	6	7	7	7	7	7	6	6	7	6
5	6	5	6	6	6	6	6	6	5	6	6	7	6	5	5
6	7	6	6	7	6	5	6	6	5	8	7	6	6	5	5

28. Результати тестування групи молоді (64 чоловік) за шкалою, що характеризує рівень їх емоційної зрілості, мають такий вигляд:

6	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5
5	5	5	6	6	6	5	6	7	5	5	6	5	5	7	4
5	5	5	5	5	5	4	5	5	5	5	5	5	5	4	5
5	6	6	6	5	6	7	5	5	6	5	5	7	4	5	5

Завдання 2. Припускаючи, що у наведених задачах вибірка отримана з нормально розподіленої генеральної сукупності, потрібно:

1. знайти точкові оцінки для математичного сподівання і дисперсії випадкової величини, заданої статистичним розподілом вибірки;
2. знайти γ - надійні інтервали для математичного сподівання і середнього квадратичного відхилення випадкової величини, яка задається статистичним розподілом вибірки.

Варіанти завдань формуються на підставі поданих нижче задач за значеннями параметра k і надійністі γ за такою таблицею:

Номер варіанта	k	γ
1	40	0,999
2	60	0,95
3	70	0,99
4	30	0,999
5	50	0,95
6	45	0,99
7	1000	0,999
8	1500	0,95
9	2000	0,99
10	2500	0,999
11	1800	0,95
12	3000	0,99
13	1	0,999
14	20	0,99
15	5	0,99
16	15	0,999
17	5	0,95
18	70	0,99
19	2,5	0,999
20	2	0,95
21	1,5	0,99
22	1,2	0,999
23	3	0,95
24	3,4	0,99
25	20	0,999
26	7	0,95
27	14	0,99
28	3	0,999
29	12	0,95
30	35	0,99

Задача 1 (номери варіантів (1-6). Результати перевірки на міцність дроту одного діаметра з різних матеріалів задаються статистичним розподілом вибірки:

Розр. Зусил	[k;k+2)	[k+2;k+4)	[k+4;k+6)	[k+6;k+8)	[k+8;k+10)	[k+10;k+12)	[k+12;k+14)
Кіль. зраз.	20	47	80	89	40	16	8

Задача2 (номери варіантів 7-12). Час роботи електричних ламп різної потужності відображено в таблиці:

Час Роботи у год.	$[k;k+100)$	$[k+100;k+200)$	$[k+200;k+300)$	$[k+300;k+400)$
Кількість ламп	2	4	9	14

Час Роботи Год.	$[k+400;k+500)$	$[k+500;k+600)$	$[k+600;k+700]$
Кількість ламп	8	3	1

Задача 3 (варіанти 13-18). З генеральної нормально розподіленої сукупності з середнім квадратичним відхиленням $\sigma = 14$ зроблено вибірку:

$[x_i;x_{i+1})$	$[k;k+10)$	$[k+10;k+20)$	$[k+20;k+30)$	$[k+30;k+40)$
n_i	20	47	80	89

$[k+40;k+50)$	$[k+50;k+60)$	$[k+60;k+70]$
40	16	8

Задача 4 (варіанти 19-24). Час доїзду до місця праці задається таблицею:

Час доїзду	$[k;k+0,5)$	$[k+0,5;k+1)$	$[k+1;k+1,5)$	$[k+1,5;k+2)$
Число осіб	15	20	30	80

$[k+2;k+2,5)$	$[k+2,5;k+3)$	$[k+3;k+3,5]$
120	50	25

Задача 5 (варіанти 25-30). З генеральної сукупності, яка має нормальний розподіл, з середнім квадратичним відхиленням $\sigma = 7$ отримано вибірку:

$[x_i; x_{i+1})$	$[k+5; k+10)$	$[k+10; k+15)$	$[k+15; k+20)$	$[k+20; k+25)$
n_i	6	8	15	40

$[k+25; k+30)$	$[k+30; k+35)$	$[k+35; k+40]$
16	8	7

Завдання 3. За даними вибірок необхідно:

1. побудувати полігон або гістограму частот;
2. сформулювати гіпотезу про закон розподілу ознаки генеральної сукупності, яка досліджується з допомогою цієї вибірки (в задачах 1-6 рекомендується перевіряти вибірки на нормальний закон, а в задачах 7-12 – на інші закони розподілу,як от, рівномірний,показників,біномний,Пуассона);
3. перевірити сформульовану гіпотезу за критерієм «хі - квадрат», вибравши за рівень значущості $\alpha = 0,5.$

Задачі до завдання 3 визначаються таблицями:

Номер варіанта	Номер задачі	α_k, α_{k+4}, α_{k+8}	α_{k+1}, α_{k+5}, α_{k+9}	α_{k+2}, α_{k+6}, α_{k+10}	α_{k+3}, α_{k+7}, α_{k+11}
1	1, 7	0	0	0	0
2	1, 7	0	-1	0	1
3	1, 7	-1	0	1	0
4	1, 7	1	0	-1	0
5	1, 7	0	1	0	-1
6	2, 8	0	0	0	0
7	2, 8	0	-1	0	1
8	2, 8	-1	0	1	0
9	2, 8	1	0	-1	0
10	2, 8	0	1	0	-1
11	3, 9	0	0	0	0
12	3, 9	0	-1	0	1
13	3, 9	-1	0	1	0
14	3, 9	1	0	-1	0
15	3, 9	0	1	0	-1
16	4, 10	0	0	0	0
17	4, 10	0	-1	0	1
18	4, 10	-1	0	1	0
19	4, 10	1	0	-1	0

20	4, 10	0	1	0	-1
21	5, 11	0	0	0	0
22	5, 11	0	-1	0	1
23	5, 11	-1	0	1	0
24	5, 11	1	0	-1	0
25	5, 11	0	1	0	-1
26	6, 12	0	0	0	0
27	6, 12	0	-1	0	1
28	6, 12	-1	0	1	0
29	6, 12	1	0	-1	0
30	6, 12	0	1	0	-1

Задача 1. Для вивчення технічних властивостей нової марки бетону досліджувались окремі його зразки. Результати перевірки міцності на стиск (кГ/см ^{2}) зразків бетону подано в таблиці:

X , кГ/ см ^{2}	Кількість зразків	X , кГ/ см ^{2}	Кількість зразків
170-180	4	220-230	66+6 α_{k+4}
180-190	8+ α_k	230-240	52+4 α_{k+5}
190-200	28+2 α_{k+1}	240-250	24+2 α_{k+6}
200-210	52+4 α_{k+2}	250-260	10+ α_{k+7}
210-220	70+6 α_{k+3}	260-270	6

Задача 2. Для контролю за готовою продукцією відібрано деталі, що виготовляються на однотипних верстатах –автоматах. Контрольний розмір X деталей подається в таблиці:

X , мм	Кількість деталей	X , мм	Кількість деталей
23,2	1	24,2	$146+8\alpha_{k+3}$
23,4	3	24,4	$75+\alpha_{k+4}$
23,6	$23+2\alpha_k$	24,6	$25+2\alpha_{k+5}$
23,8	$79+4\alpha_{k+1}$	24,8	4
24,0	$141+8\alpha_{k+2}$	25,0	1

Задача 3. Для вдосконалення організації праці на підприємствах торгівлі були відібрані дані про обсяг реалізації X (в умовних одиницях) за місяць товарів у магазинах міста. Ці дані відображені в таблиці:

X , ум.од.	Кількість магазинів	X , ум.од.	Кількість магазинів

28-30	1	38-40	$85+6\alpha_{k+3}$
30-32	2	40-42	$45+3\alpha_{k+4}$
32-34	$10+2\alpha_k$	42-44	$15+2\alpha_{k+5}$
34-36	$51+3\alpha_{k+1}$	44-46	3
36-38	$88+6\alpha_{k+2}$	46-48	1

Задача 4. Під час вивчення рівня продуктивності праці робітників реєструвався час T (хв), затрачений ними на виготовлення однотипної деталі. Ці дані відображені в таблиці:

T , хв.	Кількість робітників	T , хв	Кількість робітників
4,0-4,5	1	6,5-7,0	$116+6\alpha_{k+3}$
4,5-5,0	3	7,0-7,5	$60+4\alpha_{k+4}$
5,0-5,5	$16+2\alpha_k$	7,5-8,0	$23+2\alpha_{k+5}$
5,5-6,0	$66+4\alpha_{k+1}$	8,0-8,5	2
6.0-6,5	$112+6\alpha_{k+2}$	8,5-9,0	1

Задача 5. Для вивчення побутових потреб міста в електроенергії реєструвались її добові витрати приватними споживачами Дані про ці витрати X (кВт./год.) наведені в таблиці:

X , кВт/год	Кількість споживачів	X , кВт/год	Кількість споживачів
1,0-1,5	2	3,5-4,0	$864+8\alpha_{k+4}$
1,5-2,0	$24+\alpha_k$	4,0-4,5	$469+4\alpha_{k+5}$
2,0-2,5	$136+2\alpha_{k+1}$	4,5-5,0	$143+2\alpha_{k+6}$
2,5-3,0	$470+4\alpha_{k+2}$	5,0-5,5	$23+2\alpha_{k+7}$
3,0-3,5	$868+8\alpha_{k+3}$	5,5-6,0	1

Задача 6. За показник якості електроламп вибрано час T (хв.) їх справної роботи. Дані про перевірку їх якості відображено в таблиці:

T , год	Кількість ламп	T , х/год.	Кількість ламп
1000-1100	2	1500-1600	$291+6\alpha_{k+4}$
1100-1200	$7+\alpha_k$	1600-1700	$159+4\alpha_{k+5}$
1200-1300	$50+2\alpha_{k+1}$	1700-1800	$42+2\alpha_{k+6}$
1300-1400	$153+4\alpha_{k+2}$	1800-1900	$9+\alpha_{k+7}$
1400-1500	$286+6\alpha_{k+3}$	1900-2000	1

Задача 7. Для покращення обслуговування сільськогосподарських машин були зібрані дані про вихід з ладу техніки у господарствах регіону за період польових робіт. Ці дані X (в кількостях поломок) наведено в таблиці:

X поломок	Кількість техніки	X поломок	Кількість техніки
0	$2483+32\alpha_k$	3	$142+4\alpha_{k+3}$
1	$1783+16\alpha_{k+1}$	4	$25+2\alpha_{k+4}$
2	$608+8\alpha_{k+2}$	5	$4+2\alpha_{k+5}$

Задача 8. Щоб вивчити схожість насіння кукурудзи на кожному погонному метрі висіяли 8 насінин. Розподіл погонних метрів за кількістю X проросли насінин подано в таблиці:

X , нас.	Кількість пог.метрів	X , нас.	Кількість пог.метрів
0	0	5	$263+4\alpha_{k+3}$
1	3	6	$295+6\alpha_{k+4}$
2	$10+\alpha_k$	7	$198+3\alpha_{k+5}$
3	$47+2\alpha_{k+1}$	8	$51+\alpha_{k+6}$
4	$136+4\alpha_{k+2}$		

Задача 9. На деякому підприємстві досліджувалась кількість поломок деякої техніки протягом року. Дані про щомісячні поломки подано в таблиці:

X поломок	Кількість техніки	X поломок	Кількість техніки
1	$81+\alpha_k$	7	$88+4\alpha_{k+6}$
2	$79+\alpha_{k+1}$	8	$90+\alpha_{k+7}$
3	$80+\alpha_{k+2}$	9	$92+\alpha_{k+8}$
4	$85+\alpha_{k+3}$	10	$75+\alpha_{k+9}$
5	$87+\alpha_{k+4}$	11	$77+\alpha_{k+10}$
6	$88+\alpha_{k+5}$	12	$78+\alpha_{k+11}$

Задача 10. Щоб встановити, яка частка готової продукції заводу потребує додаткового регулювання, утворено вибірки по 7 зразків продукції в кожній. Розподіл цих вибірок за кількістю приладів X , які необхідно регулювати, подано в таблиці:

X приладів	Вибірок,шт.	X приладів	Вибірок,шт.

0	$25+\alpha_k$	4	$29+2\alpha_{k+4}$
1	$74+3\alpha_{k+1}$	5	$8+\alpha_{k+5}$
2	$95+5\alpha_{k+2}$	6	1
3	$68+3\alpha_{k+3}$	7	0

Задача 11. Для покращання обслуговування абонентів на декількох АТС міста були зібрані відомості про випадки несправної роботи апаратури автоматичного телефонного зв'язку. Дані про кількість несправностей X наведені в таблиці:

X несправностей	Одиниць апаратури	X несправностей	Одиниць апаратури
0	$25+\alpha_k$	4	$84+4\alpha_{k+4}$
1	$75+4\alpha_{k+1}$	5	$50+2\alpha_{k+5}$
2	$112+2\alpha_{k+2}$	6	$25+2\alpha_{k+6}$
3	$112+2\alpha_{k+3}$	7 і більше	$5+\alpha_{k+7}$

Задача 12. Для вдосконалення роботи портів реєструвався час T (год.) очікування кораблів на розвантаження в декількох портах країни. Результати річних спостережень подано в таблиці:

T , год.	Кількість кораблів	T , год.	Кількість кораблів
0-6	$518+16\alpha_k$	30-36	$116+4\alpha_{k+5}$
6-12	$384+8\alpha_{k+1}$	36-42	$86+4\alpha_{k+6}$
12-18	$284+8\alpha_{k+2}$	42-48	$63+2\alpha_{k+7}$
18-24	$211+6\alpha_{k+3}$	48-54	$44+2\alpha_{k+8}$
24-30	$156+6\alpha_{k+4}$	54 і більше	$34+\alpha_{k+9}$

Завдання 4

Задача 1. Підприємство отримує електролампи одного типу від постачальників $A_1, A_2, ..., A_m$. Чи справджується гіпотеза про відсутність чинника постачальника на тривалість роботи електролампи?

а)

A_i	Тривалість роботи, год для варіантів $k=1,2,3,4,5$; $n=30k$							
A_1	$n+10$	$n+20$	$n-5$	$n+5$	$n+15$	$n-10$	$n+5$	$n+5$

A_2	$n-10$	$n-15$	$n+5$	$n-10$	$n+10$	$n+5$	$n-10$	$n-10$
A_3	$n+5$	$n+15$	$n+30$	$n-5$	$n+10$	$n-10$	n	$n+5$

б)

A_i	Тривалість роботи, год для варіантів $k=6,7,8,9,10$; $n=20k$					
A_1	$n+20$	$n+25$	$n-30$	$n+10$	$n-15$	$n+5$
A_2	$n+10$	$n-5$	$n+10$	$n+25$	$n-25$	$n-15$
A_3	$n-5$	$n-15$	$n+50$	$n-40$	$n-20$	$n+10$
A_4	$n+30$	$n+15$	$n-5$	$n+10$	$n+25$	$n-15$

Задача 2. Верстати $A_1, A_2, ..., A_m$ виготовляють деталі певного діаметра. Чи справджується гіпотеза про відсутність чинника верстата на діаметр деталі, що виготовляється?

а)

A_i	Діаметр деталі, мм , для варіантів $k=11,12,13,14,15$; $n=20k$						
A_1	$n+10$	$n+15$	$n-5$	$n+5$	$n-15$	$n+35$	$n+10$
A_2	$n-50$	$n-40$	$n+5$	$n-25$	$n-10$	$n+15$	$n-20$
A_3	$n+10$	$n+10$	$n-5$	$n-5$	$n+25$	$n-10$	$n-25$

б)

A_i	Діаметр деталі, мм , для варіантів $k=16,17,18,19,20$; $n=10k$				
A_1	$n+25$	$n+40$	$n-10$	$n+90$	$n+25$
A_2	$n-15$	$n+5$	$n-45$	$n+20$	$n+25$
A_3	$n-10$	$n+5$	$n-10$	$n-5$	$n-20$
A_4	$n+5$	$n+20$	$n-80$	$n-40$	$n+35$

Задача 3. Виробник використовує заготовки одного типу, що надходять від постачальників $A_1, A_2, ..., A_m$. Чи можна підтвердити гіпотезу про те, що отримані заготовки мають однакову масу?

а)

A_i	Маса заготовки, г для варіантів $k = 21, 22, 23, 24, 25$; $n = 10k$								
A_1	$n+5$	$n+10$	$n+25$	$n-15$	$n+10$	$n-25$	$n+10$	$n+5$	$n-15$
A_2	$n+10$	$n-15$	$n-5$	$n+10$	$n-25$	$n+30$	$n+50$	$n-25$	$n+10$
A_3	$n-5$	$n-25$	$n+20$	$n+15$	$n-20$	$n-15$	n	$n+5$	$n+5$
A_4	$n-10$	$n-30$	$n+30$	$n+25$	$n+15$	$n+10$	$n+15$	$n-25$	$n-5$

б)

A_i	Маса заготовки, г $k = 26, 27, 28, 29, 30$; $n = 10k$					
A_1	$n+30$	$n-25$	$n-40$	$n+10$	$n+20$	$n-15$
A_2	$n-10$	$n-40$	$n+5$	$n+30$	$n-25$	$n-40$
A_3	$n+20$	$n+50$	$n-10$	$n-20$	$n-30$	$n-35$

Завдання 5. Припускаючи, що між випадковими X та Y існує лінійний кореляційний зв’язок, який відображений в кореляційній таблиці, необхідно:

1. обчислити коефіцієнт кореляції;
2. скласти рівняння прямих регресії Y на X та X на Y.

Задача 1. (для варіантів $k = 1, 2, ..., 8$) Дано розподіл однотипних підприємств за обсягом продукції X та її собівартістю Y:

Y \X	2	3	4	5	6
100				$11-k$	$k+1$
200			$2k$	k	$10-k$
300			$10-k$	$k+1$	
400		$14-k$	k		
500	$12-k$	k			
600	$k+1$	$12-k$			

Задача 2. (для варіантів $k = 9, 10, ..., 15$) Заданий розподіл підприємств за обсягом їх основних фондів X та готовою продукцією Y:

Y \X	20	30	40	50	60	70
15	$k+1$	$21-k$				

25	$18-k$	$2k$	$k+1$			
35		k	$22-k$	$17-k$		
45				$k+2$	k	$19-k$
55					$24-k$	$k+1$

Задача 3. (для варіантів $k=16,17,...,23$) Заданий розподіл однотипних підприємств за обсягом основних фондів X і собівартістю Y :

Y\X	1,5	2	2,5	3	3,5	4
8					$k-13$	$26-k$
13				$k-15$	$25-k$	$2k-29$
18			$25-k$	$k-13$	$29-k$	
23	$k-10$	$27-k$	$k-14$			
28	$26-k$	$k-7$				

Задача 4. (для варіантів $k=24,25,...,30$) Заданий розподіл накладних витрат X і обсягу виконаних робіт Y :

Y \X	3	5	7	9	11
35	$k-19$	$35-k$			
45	$42-k$	$2k-48$	$k-22$		
55		$k-21$	$46-k$	$39-k$	$33-k$
65			$k-17$	$k-20$	$k-23$
75				$36-k$	$40-k$

\
Додатки

Додаток 1

Значення функції $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

X	0	1	2	3	4	5	6	7	8	9
0,0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3188	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013

3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Додаток 2

Значення функції Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$.

x	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24573	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46582	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47645	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48169
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49609	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807

2,9	49813	49819	49825	49831	49836	49841	49846	49851	49858	49861
3,0	0,49865		3,1	49903	3,2	49931	3,3	49952	3,4	49986
3,5	49977		3,6	49984	3,7	49989	3,8	49993	3,9	49995
4,0	49997									

Додаток 3

Значення розподілу Пуассона $P_{m}(\lambda)=\dfrac{\lambda^{m}}{m!}e^{-\lambda}$

m \ λ	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,904084	0,163746	0,222245	0,268128	0,303265	0,329287
2	0,004524	0,016375	0,033337	0,053626	0,075816	0,098786
3	0,000151	0,001091	0,003334	0,007150	0,012636	0,019757
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964
5		0,000002	0,000015	0,000057	0,000158	0,000356
6			0,000001	0,000004	0,000013	0,000035
7					0,000001	0,000003
m \ λ	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367879	0,135335	0,049787
1	0,347610	0,359463	0,365913	0,367879	0,270671	0,149631
2	0,121663	0,143785	0,164661	0,183940	0,270671	0,224042
3	0,028388	0,038343	0,049398	0,061313	0,180447	0,224042
4	0,004968	0,007669	0,011115	0,015328	0,090224	0,168031
5	0,000695	0,001227	0,002001	0,003066	0,036089	0,100819
6	0,000081	0,000164	0,000300	0,000511	0,012030	0,050409
7	0,000008	0,000019	0,000039	0,000073	0,003437	0,021604
8		0,000002	0,000004	0,000009	0,000859	0,008101
9				0,000001	0,000191	0,002701
10					0,000038	0,000810
11					0,000007	0,000221
12					0,000001	0,000055
13						0,000013
14						0,000002
15						0,000001

Додаток 4

Значення $t = t_{\gamma;n}$ для t-розподілу Стьюдента.

$t_{\gamma;n}$ визначається рівністю $P\left(|\tau_n| < t_{\frac{\alpha}{2},n}\right) = \gamma$, $\gamma = 1 - \alpha$.

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
4	2,78	4,60	8,61	19	2,093	2,861	3,883
5	2,57	4,03	6,86	20	2,086	2,845	3,849
6	2,45	3,71	5,96	25	2,064	2,797	3,745
7	2,37	3,50	5,41	30	2,045	2,756	3,659
8	2,31	3,36	5,04	35	2,032	2,729	3,600
9	2,26	3,25	4,78	40	2,023	2,708	3,558
10	2,23	3,17	4,59	45	2,016	2,692	3,527
11	2,20	3,11	4,44	50	2,009	2,679	3,502
12	2,18	3,06	4,32	60	2,001	2,662	3,464
13	2,16	3,01	4,22	70	1,996	2,649	3,439
14	2,15	2,98	4,14	80	1,991	2,640	3,418
15	2,13	2,95	4,07	90	1,987	2,633	3,403
16	2,12	2,92	4,02	100	1,984	2,627	3,392
17	2,11	2,90	3,97	120	1,980	2,617	3,374
18	2,10	2,88	3,92	∞	1,960	2,576	3,291

Додаток 5

Значення $\chi^2_{\alpha,k}$ для розподілу Пірсона

Число ступенів вільності k \ Ймовірність $(\alpha = 1-\gamma)$	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	0,00016	0,0006	0,0039	0,016	0,064	0,148	0,455	1,07	1,64	2,7	3,8	5,4	6,6	7,9	9,5	10,83
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,6	6,0	7,8	9,2	11,6	12,4	13,8
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,66	4,64	6,3	7,8	9,8	11,3	12,8	14,8	16,3
4	0,30	0,43	0,71	1,06	1,65	2,19	3,36	4,9	6,0	7,8	9,5	11,7	13,3	14, 9	16,9	18,5
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,1	7,3	9,2	11,1	13,4	15,1	16,3	18,9	20,5
6	0,187	1,13	1,63	2,20	3,07	3,83	5,35	7,2	8,6	10,6	12,6	15,0	16,8	18,6	20,7	22,5
7	1,24	1,56	2,17	2,83	3,82	4,67	6,34	8,4	9,8	12,0	14,1	16,6	18,5	20,3	22,6	24,4
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5	11,0	13,4	15,5	18,2	20,1	21,9	24,3	21,6
9	2,09	2,53	3,32	4,17	5,38	6,39	8,35	10,7	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	3,6	4,2	5,2	6,3	7,8	9,0	11,3	14,0	15,8	18,5	21,0	24,1	26,2	28,3	31,0	32,9
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2	18,2	21,1	23,7	26,9	29,1	31,0	34,0	36,1
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6	32,5	35,5	37,7
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0	34,0	37,0	39,2
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4	35,5	38,5	40,8
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8	37,0	40,0	42,3
19	7,6	8,6	10,1	11,7	13,7	18,3	21,7	23,9	27,2	30,1	33,7	36,2	36,2	38,5	41,5	43,8
20	8,3	9,2	10,9	12,4	14,6	19,3	22,8	25,0	28,4	31,4	35,0	37,6	37,6	40,0	43,0	45,3
21	8,9	9,9	11,6	13,2	15,4	20,3	23,9	26,2	29,6	32,7	36,3	38,9	38,9	41,5	44,5	46,8
22	9,5	10,6	12,3	14,0	16,3	21,3	24,9	27,3	30,8	33,9	37,7	40,3	40,3	42,5	46,0	48,3
23	10,2	11,3	13,1	14,8	17,2	22,3	26,0	28,4	32,0	35,2	39,0	41,6	41,6	44,0	47,5	49,7
24	10,9	12,0	13,8	15,7	18,1	23,3	27,1	29,6	33,2	36,4	40,3	43,0	43,0	45,5	48,5	51,2
25	11,5	12,7	14,6	16,5	18,9	24,3	28,1	30,7	34,4	37,7	41,6	44,3	44,3	47,0	50,0	52,6
26	12,2	13,4	15,4	17,3	19,8	25,3	29,3	31,8	35,6	38,9	42,9	45,6	45,6	48,0	51,5	54,1
27	12,9	14,1	16,2	18,1	20,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	47,0	49,5	53,0	55,5
28	13,6	14,8	16,9	18,9	21,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	48,3	51,0	54,5	56,9
29	14,3	15,6	17,7	19,8	22,5	28,3	32,5	35,1	39,1	42,6	46,7	49,6	49,6	52,5	56,0	58,3
30	15,0	16,3	18,5	20,6	23,4	29,3	33,5	36,3	40,3	43,8	48,0	50,9	50,9	54,0	57,5	59,7

Додаток 6

Критичні точки розподілення F Фішера-Снедекора
(k_1 — число степенів свободи великої дисперсії,
k_2 — число степенів свободи малої дисперсії)

Рівень значимості α=0,01

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,63	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10.92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4.71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4.44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Рівень значимості α=0,05

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19.33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8.84	8,81	8,78	8,76,	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4.95	4,88	4,82	4,78	4.74	4,70	4.68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,2)	4,15	4,10	4,06	4,03	4,00
7	5,59	4.74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50.	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2.76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. *Гмурман В.Е.* Теория вероятностей и математическая статистика. – М.: ВШ.2001. – 480 с.
2. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. – М.: ВШ, 2001. – 400 с.
3. *Барковський В.В., Барковська Н.В., Лопатін О.К.* Теорія ймовірностей і математична статистика. – К.; ЦУЛ, 2002. – 448 с.
4. *Жлуктенко В.І., Наконечний С.І.* Теорія ймовірностей і математична статистика: Навч.-метод. Посібник: у 2-х ч. – К.:КНЕУ, 2000. – ч.2. – 334 с.
5. *Бобик О.І., Берегова Г.І., Копитко Б.І.* Теорія ймовірностей і математична статистика. – Київ.: ВД «Професіонал», 2007. – 558 с.

ЗМІСТ		
1.	Основні поняття математичної статистики	
2.	Оцінки невідомих параметрів розподілу	
2.1	Точкові оцінки	
2.2	Точкові оцінки математичного сподівання, дисперсії та середнього квадратичного відхилення	
2.3	Інтервальні оцінки параметрів розподілу	
2.4	Розподіл χ^2 - «хі-квадрат»	
2.5	Розподіл Стьюдента	
2.6	Розподіл Фішера-Снедекора	
2.7	Інтервальні оцінки математичного сподівання випадкової величини, яка розподілена за нормальним законом	
3.	Статистична перевірка гіпотез	
3.1	Статистичні гіпотези та їх різновиди.	
3.2	Похибки перевірки гіпотез	
3.3	Статистичний критерій перевірки основної гіпотези.	
3.4	Критична область	
3.5	Знаходження критичних областей	
3.6	Порядок дій при перевірці статистичних гіпотез	
3.7	Критерій узгодження Пірсона (χ^2).	
3.8	Перевірка гіпотез про параметри нормального розподілу	
3.9	Гіпотези про математичні сподівання	
3.9.1	Перевірка гіпотези про значення математичного сподівання за відомої дисперсії	
3.9.2	Перевірка гіпотези про рівність дисперсій двох незалежних випадкових величин	
4.	Однофакторний дисперсійний аналіз	
5.	Основні поняття кореляційного та регресійного аналізу	
5.1	Статистичний опис системи двох випадкових величин	
5.2	Вибірковий коефіцієнт кореляції	
5.3	Основні поняття і методи регресійного аналізу	
	Додатки	
	Список використаної літератури	