

## 4. РЯДИ РОЗПОДІЛУ. АНАЛІЗ ВАРІАЦІЇ ТА ФОРМИ РОЗПОДІЛУ

---

Групування за однією ознакою, що характеризує склад (структуру) явища або процесу в даний період часу, зветься *рядом розподілу*. В залежності від того, яка ознака (якісна або кількісна) береться за основу групування, ряди розподілу бувають якісними (атрибутивними) чи кількісними (варіаційними).

Ряд розподілу складається з варіант та частот:  $x_j$  – *варіанта* – окреме значення ознаки, що змінюється;  $f_j$  – *частота* – вказує скільки разів повторюється вказана варіанта;  $d_j$  – *відносна частка* – вказує, яку долю сукупності складає варіанта  $x_j$  у всій сукупності. Підсумок усіх часток налічує повну сукупність. Підсумок відносних часток дорівнює одиниці або 100%:

$$\sum_1^m f_j = n, \quad \sum_1^m d_j = 1, \quad (4.1)$$

$$d_j = \frac{f_j}{\sum_1^m f_j} = \frac{f_j}{n}, \quad (4.2)$$

де  $m$  – кількість груп,  $n$  – обсяг сукупності.

Аналіз рядів розподілу зручніше виконувати за допомогою графіків.

### 4.1. Графічний аналіз рядів розподілу

На першій стадії використовується діаграма «казусів» – на діаграму наносять значення ознаки, що змінюється, у тій послідовності у якій вони отримані. Тобто наносять первинні дані.

На другій стадії будується *ранжований ряд* – ряд, у якому окремі значення ознаки розташовані у порядку зростання, тобто за ранжиром.

Крапки поєднуються відрізками і утворюють ламану лінію, яка називається *огівом*. За нахилом огіву можливо отримати уяву про ступень різномірності сукупності.

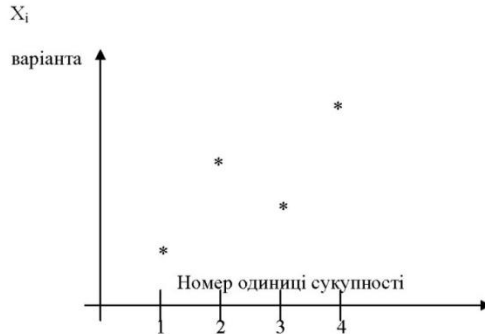


Рис. 4.1 Діаграма «казусів»

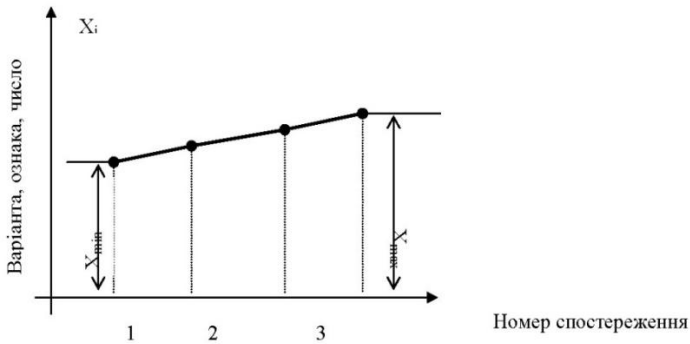
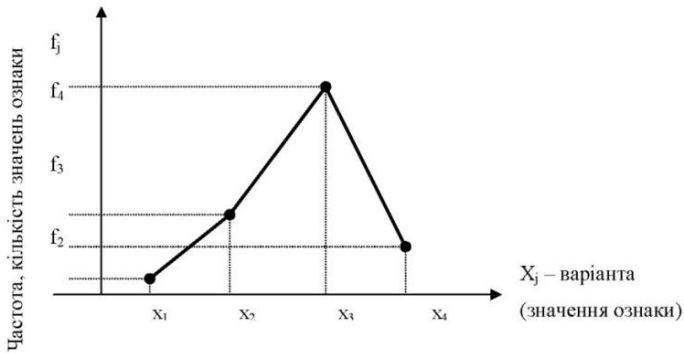


Рис. 4.2. Огіва

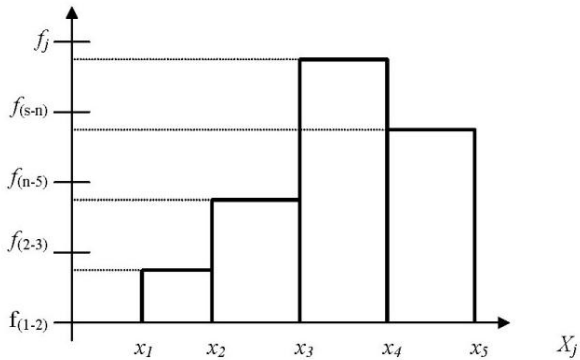
На третій стадії вивчення ряду розподілу будується **полігон** – графік, на якому ряд зображений у вигляді лінійної діаграми. На осі абсцис відкладають значення ознаки  $X_j$  у порядку зростання, а по осі ординат частоту відповідної варіанти (ознаки)  $f_j$ . Крпки поєднуються відрізками та утворюють ламану лінію.

Для інтервальних рядів розподілу будується **гістограма** – графік, на якому ряд розподілу зображується у вигляді розташованих один біля одного стовпчиків. Висота стовпчика пропорційна частоті, тобто кількості ознак, що потрапили до інтервалу ознаки. Ширина стовпчика дорівнює ширині інтервалу розподілу.

Щоб уникнути впливу нерівномірного інтервалу розподілу на висоту стовпчиків, висоту стовпчика зображають пропорційно не частотам, а **густині розподілу**.



**Рис. 4.3. Полігон**



**Рис. 4.4. Гістограма**

**Густина розподілу** – кількість випадків, що припадає на одиницю ширини інтервалу ознаки, яка змінюється. Частота не має розміру – вона вимірюється у кількості випадків. На відміну від частоти, густина розподілу має розмір:

$$\frac{\text{кількість\_випадків}}{\text{розмір інтервалу}}$$

На четвертій стадії вивчення ряду розподілу будується **кумулята** – графік, на якому на осі абсцис відкладають значення ознаки у порядку зростання, а по осі ординат підсумок накопичених частот. Накопичені частоти підраховують шляхом послідовного складання. Ознаці, яка має максимальне значення відповідає відносна частка 1 або 100 %, а підсумок частот дорівнює  $n$  (кількості одиниць усієї сукупності).

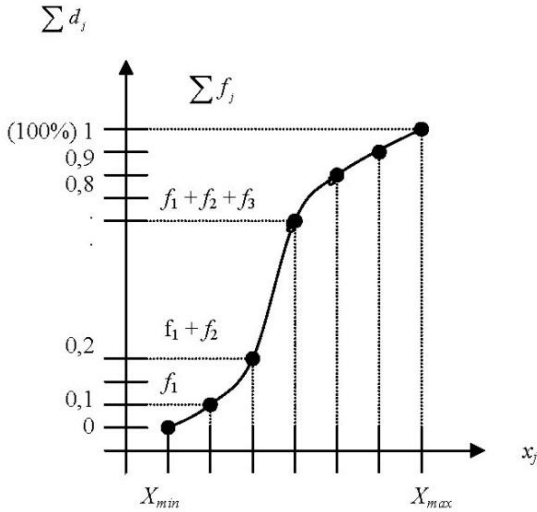


Рис. 4.5. Кумулята

На *n*'ятій стадії вивчення ряду розподілу будується інтегральна функція розподілу.

$$P_j = \lim_{\Delta x_j \rightarrow 0} \frac{\Delta f_j}{\Delta x_j} = \frac{\partial(f_j)}{\partial x_j}$$

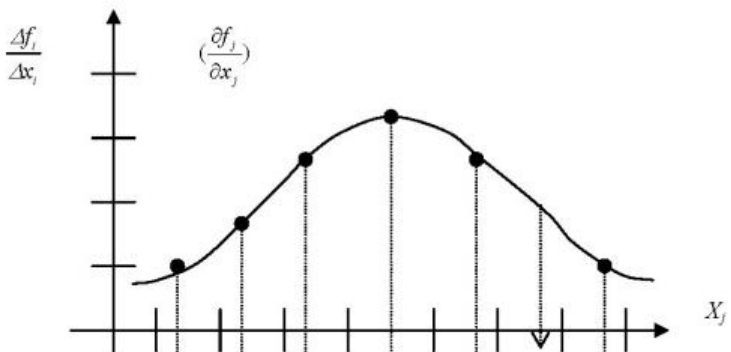


Рис. 4.6. Густина розподілу (щільність ймовірностей, диференціальна функція розподілу)

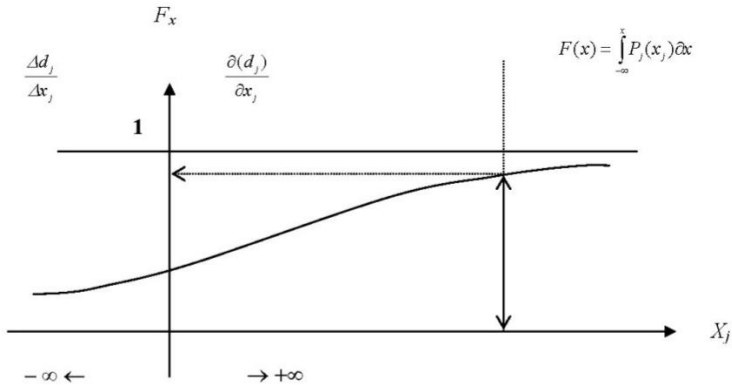


Рис. 4.7. Інтегральна функція розподілу  
(емпірична функція розподілу)

## 4.2. Мода, медіана

**Модою** називають найбільш поширене значення ознаки ( $M_0$ ). У дискретних рядах модальне значення визначають безпосередньо за найбільшою частотою (часткою). В інтервальному ряду визначають модальний інтервал, а конкретне значення модальної ознаки розраховують за формулою:

$$M_0 = x_0 + h \frac{f_{m0} - f_{m0-1}}{(f_{m0} - f_{m0-1}) + (f_{m0} - f_{m0+1})}, \quad (4.3)$$

де  $x_0$  – нижня межа модального інтервалу;  $h$  – ширина модального інтервалу;  $f_{m0}$ ;  $f_{m0-1}$ ;  $f_{m0+1}$  – частоти модального, передмодального, післямодального інтервалів.

**Медіана** – значення варіюючої ознаки, яка припадає на середину ранжованого ряду, тобто ділить його на дві рівні частини: одна частина ряду має значення варіаційної ознаки менші, ніж медіана, а друга – більші. Медіана вказує на значення варіаційної ознаки, якого досягла половина одиниць сукупності. В інтервальному ряду значення медіани розраховують за формулою:

$$Me = x_0 + h \frac{0,5 \sum_1^m f_j - S f_{me-1}}{f_{me}}, \quad (4.4)$$

де  $x_0$  – нижня межа медіанного інтервалу;  $h$  – ширина медіанного інтервалу;  $f_{me}$  – частота медіанного інтервалу;  $Sf_{me-1}$  – кумулятивна частота передмедіанного інтервалу.

Як розподіл ознаки має два максимуми або більше – так помилково виконано групування. Приклад – розподіл розміру взуття студентів без урахування статі.

### 4.3. Показники варіації

**Варіація ознаки** – під цим терміном у статистиці мають на увазі такі кількісні зміни ознаки у межах однорідною сукупності, які обумовлені дією різних факторів і мають випадковий характер. Статистичний аналіз варіації дає можливість вивчити і кількісно розрахувати ступень залежності змін ознаки у статистичній сукупності в залежності від визначаючих її (залежність) факторів.

До статистичних характеристик варіації належать:

**Розмах варіації** – діапазон варіації, це різниця між максимальним та мінімальним значенням ознаки:

$$R = x_{max} - x_{min} \quad (4.5)$$

До уваги приймаються тільки два елементи сукупності і, якщо крайні значення ознаки не типові для сукупності, так відкидають крайні значення, або використовують:

**Квартильний розмах** –  $R_q = Q_3 - Q_1$  тотожно 50% сукупності.

Квартілі розподілу визначають за формулами:

перший квартиль:

$$Q_1 = x_0 + h \frac{0,25 \sum_1^m f_j - S_{Q_{1-1}}}{f_{Q_1}},$$

другий квартиль:

$$Q_2 = x_0 + h \frac{0,5 \sum_1^m f_j - S_{Q_{2-1}}}{f_{Q_2}},$$

третій квартиль:

$$Q_3 = x_0 + h \frac{0,75 \sum_1^m f_j - S_{Q_{3-1}}}{f_{Q_3}},$$

де  $x_0$  та  $h$  – відповідно нижня межа та ширина кватильного інтервалу;  $f_{Q1}, f_{Q2}, f_{Q3}$  – частота кватильного інтервалу;  $S_{Q1-1}, S_{Q1-2}, S_{Q1-3}$  – кумулятивна частота передкватильного інтервалу.

**Децильні розмахи варіації** –  $R_D = D_9 - D_1$  (80 % сукупності);  $R_D = D_8 - D_2$  (60 % сукупності); ...

Перший дециль обчислюється за формулою:

$$D_1 = x_{D1} + h \frac{0,1 \sum_1^m f_{Dj} - S_{D1-1}}{f_{D1}},$$

другий:

$$D_2 = x_{D2} + h \frac{0,2 \sum_1^m f_{Dj} - S_{D2-1}}{f_{D2}},$$

третій:

$$D_3 = x_{D3} + h \frac{0,3 \sum_1^m f_{Dj} - S_{D3-1}}{f_{D3}}, \dots,$$

$$D_j = x_{Dj} + h \frac{0,1 \cdot j \sum_1^m f_{Dj} - S_{Dj-1}}{f_{Dj}}, \dots,$$

де  $x_{Dj}$  та  $h$  – відповідно нижня межа та ширина децильного інтервалу;  $f_{Dj}$  – частота децильного інтервалу;  $S_{Dj-1}$  – кумулятивна частота переддецильного інтервалу.

**Середнє лінійне відхилення** – узагальнююча міра варіації, бо враховує усі елементи сукупності:

$$\bar{l} = \frac{\sum_1^n |x_i - \bar{x}|}{n} \equiv \frac{\sum_1^m |x_j - \bar{x}| f_j}{\sum_1^m f_j}. \quad (4.6)$$

**Середнє квадратичне відхилення** – також узагальнююча міра варіації:

$$\sigma = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{n}}. \quad (4.7)$$

**Середній квадрат відхилення** – дисперсія:

$$\sigma^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n} = \frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j} = \overline{(x^2)} - (\bar{x})^2. \quad (4.8)$$

#### 4.4. Дисперсія альтернативної ознаки

Хай альтернативна ознака може приймати два значення – одинця та нуль:

$x_1 = \langle 1 \rangle \rightarrow d_1$  – частка ознак, що приймають значення  $\langle 1 \rangle$ ,

$x_2 = \langle 0 \rangle \rightarrow d_0$  – частка  $\langle 0 \rangle$  ознак, що приймають значення.

Дисперсія згідно з формулою буде складати:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 \cdot d_1 + (x_2 - \bar{x})^2 \cdot d_0}{d_1 + d_0}, \quad (4.9)$$

$$\sigma^2 = \frac{d_1(1-d_1)^2 + (0-d_1)^2 \cdot d_0}{d_1 + d_0} = \frac{d_1 d_0^2 + d_0 d_1^2}{d_1 + d_0} = d_1 d_0.$$

Середнє арифметичне значення відповідно складатиме:

$$\bar{x} = \frac{x_1 d_1 + x_2 d_0}{d_0 + d_1}, \quad (4.10)$$

$$\bar{x} = \frac{1 \cdot d_1 + 0 \cdot d_0}{d_0 + d_1} = \frac{d_1 \cdot 1}{1} = d_1.$$

#### 4.5. Властивості дисперсії

1. Якщо всі значення ознаки у сукупності зменшити чи збільшити на певну постійну, так дисперсія не зміниться.

2. Якщо всі значення ознаки змінити у  $k$ -разів, так дисперсія зміниться у  $k^2$ -разів.

3. У разі заміни частот частками дисперсія не зміниться.

4. Сума квадратів відхилень ознаки від її середньої завжди менша суми квадратів відхилень ознаки від любого іншого числа:

$$\sum_1^n (x_i - \bar{x})^2 < \sum_1^n (x_i - a)^2, \quad (4.11)$$
$$\bar{x} \neq a.$$

#### 4.6. Коефіцієнти варіації

*Лінійний коефіцієнт варіації*

$$v_i = \frac{\bar{l}}{\bar{x}} 100\%. \quad (4.12)$$



### **Квадратичний коефіцієнт варіації**

$$v_{\sigma} = \frac{\sigma}{\bar{x}} 100\%. \quad (4.13)$$

### **Коефіцієнт осциляції**

$$v_r = \frac{R}{\bar{x}} 100\%. \quad (4.14)$$

### **Центрована статистична ознака**

$$x_i^0 = x_i - \bar{x}. \quad (4.15)$$

### **Стандартизоване відхилення (нормоване відхилення)**

$$t = \frac{x - \bar{x}}{\sigma}. \quad (4.16)$$

## **4.7. Закони розподілу Гауса і Стьюдента**

Закон розподілу Гауса у диференціальній та інтегральній формах для нормованого відхилення має відповідно вигляд:

$$\begin{aligned} f(t) &= \frac{1}{\sqrt{2\pi}} \exp(-t^2/2), \\ F(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp(-t^2/2) dt. \end{aligned} \quad (4.17)$$

Закон розподілу Гауса у диференціальній та інтегральній формах для ненормованого відхилення має вигляд:

$$\begin{aligned} f(x) &= \frac{1}{\sigma_x \sqrt{2\pi}} \exp(-(x - \bar{x})^2 / (2\sigma_x^2)), \\ F(x) &= \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^x \exp(-(x - \bar{x})^2 / (2\sigma_x^2)). \end{aligned} \quad (4.18)$$

Графічне зображення закону розподілу Гауса у диференціальній формі надано на рисунку 4.6, а графічне зображення закону розподілу Гауса у інтегральній формі – на рисунку 4.7.

Табличне зображення закону розподілу Гауса у диференціальній та інтегральній формах надано відповідно у таблицях 4.1 та 4.2.

Таблиця 4.1

**Закон розподілу Гауса у диференціальній формі**

t	0,0	0,5	1,0	1,5	2,0	2,5	3,0
f(t)	0,399	0,3521	0,242	0,1295	0,054	0,0175	0,0044

Таблиця 4.2

**Закон розподілу Гауса у інтегральній формі**

t	0,0	0,5	1,0	1,5	2,0	2,5	3,0
F(t)	0,5	0,692	0,841	0,933	0,977	0,994	0,999

$F(-t) = 1 - F(t)$  – висновок з симетричності закону розподілу.

Довірче число (коефіцієнт довіри, квантиль розподілу)  $t$  вказує, як співвідносяться гранична та стандартна похибки. Коефіцієнт довіри залежить від ймовірності, з якою гранична похибка не вийде за межі довірчого інтервалу, та закону розподілу. Якщо закон розподілу наближається до нормального (закону розподілу Гауса), так коефіцієнт довіри та ймовірність граничної похибки пов'язані таким чином, як це наведено у таблиці.

Таблиця 4.3

**Закон розподілу Гауса**

t	Ймовірність	Гранична похибка, $\Delta$
1	0,683 (68,3%)	$\pm 1 \cdot \mu$
2	0,954 (95,4%)	$\pm 2 \cdot \mu$
3	0,997 (99,7%)	$\pm 2 \cdot \mu$

Якщо обсяг вибірки малий (менший за 30 одиниць), так замість закону розподілу Гауса слід використовувати закон розподілу Стюдента, для якого коефіцієнт довіри  $t$  залежить ще і від кількості елементів вибірки (дивись таблицю).

Таблиця 4.4

**Закон розподілу Стюдента**

t	Гранична похибка, $\Delta$	Ймовірність для вибірки з «n» спостережень					
		30 і більше	20	10	5	4	2
1	$\pm 1 \cdot \mu$	0,683	0,670	0,657	0,637	0,609	0,500
2	$\pm 2 \cdot \mu$	0,954	0,940	0,923	0,898	0,861	0,705
3	$\pm 2 \cdot \mu$	0,997	0,993	0,985	0,970	0,942	0,795

## 4.8. Характеристики форми розподілу

Головна передумова використання статистичних методів – однорідність сукупності. **Однорідність сукупності** – це, якщо всі елементи сукупності мають спільні властивості (риси) і належать до одного типу або класу; є певна наявність у них загального в істотному, головному. Показником однорідності сукупності є одновершинність розподілу (одномодальність). Багатовершинність свідчить про неоднорідний склад сукупності, про різнотипність окремих складових. У такому разі необхідно перегрупувати дані, виділити однорідні групи.

Критерієм однорідності сукупності є квадратичний коефіцієнт варіації, який у симетричному розподілі становить  $V_G = 0,33$ .

Пригадаємо:

$$v_\sigma = \frac{\sigma}{\bar{x}} = \frac{\sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{n}}}{\frac{\sum_1^n x_i}{n}}. \quad (4.19)$$

Серед одновершинних розподілів є симетричні та асиметричні. У симетричному розподілі рівновіддалені від центра значення ознаки мають однакові частоти та середня арифметична співпадає з модою та медіаною:  $\bar{x} = Mo = Me$ .

**Асиметрію розподілу** вимірюють за допомогою стандартного відхилення:

$$A_s = \frac{\mu_3}{\sigma_3} = \frac{\frac{\sum_1^m (x_j - \bar{x})^3 f_j}{\sum_1^m f_j}}{\left( \frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j} \right)^{3/2}}, \quad (4.20)$$

або

$$A_s = \frac{\bar{x} - Mo}{\sigma}.$$

Якщо  $A_s > 0$  – правостороння асиметрія (додатна),  $A_s < 0$  – лівостороння асиметрія (від'ємна).

**Скошеність функції (ексцес)** вимірюють за допомогою моменту четвертого порядку:

$$E_k = \mu_4 / \sigma^4 = \frac{\sum_1^m (x_j - \bar{x})^4 f_j}{\frac{\sum_1^m f_j}{\left(\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}\right)^2}} \quad (4.21)$$

Нормальний розподіл має  $E_k$  приблизно 3,0; гостровершинний розподіл має  $E_k$  більше 3; плосковершинний розподіл має  $E_k$  менше 3.

Більш детально форма розподілу характеризується моментом розподілу:

$$\mu_k = \frac{\sum_1^m (x_j - \bar{x})^k f_j}{\sum_1^m f_j} = \frac{\sum_1^n (x_i - \bar{x})^k}{n}. \quad (4.22)$$

Протилежною характеристикою однорідності сукупності є **нерівномірність розподілу** ознак між складовими сукупності, яка оцінюється **коефіцієнтом концентрації**:

$$K = \frac{1}{2} \sum_{j=1}^m |d_j - D_j|, \quad (4.23)$$

де  $d_j$  – частка однієї ознаки сукупності у групі « $j$ »,  $D_j$  – частка другої ознаки сукупності у тій же групі « $j$ »;  $m$  – кількість груп;  $(d_j - D_j)$  відхилення часток двох різних розподілів, ознак однієї групи « $j$ ».

$$\sum_1^m d_j = 1; \quad \sum_1^m D_j = 1. \quad (4.24)$$

$K = 0$  – рівномірний розподіл обох ознак за усіма групами сукупності,  $K = 1$  – повна концентрація першої ознаки у одній групі одного розподілу, а у іншій групі – другої ознаки іншого розподілу.

**Коефіцієнт локалізації:**

$$L_j = \frac{D_j}{d_j} \cdot 100, \quad (4.25)$$

характеризує для однієї сукупності співвідношення часток двох різних розподілів ознак однієї групи « $j$ ».

Порівняння двох структур **різних сукупностей**, що поділені на однакову кількість груп « $m$ » здійснюється за допомогою **коефіцієнту подібності (схожості)** структури двох сукупностей:

$$P = 1 - \frac{1}{2} \sum_1^m |d_j - d_k|, \quad (4.26)$$

де  $d_j$  – частка ознак однієї сукупності у групі,  $d_k$  – частка ознак другої сукупності у такій самій групі. Якщо структури однакові, тоді  $P = 1$ ; якщо протилежні  $P = 0$ .

Зміна окремих складових (часток) однієї сукупності свідчить про структурні зрушення, що відбуваються з сукупністю в часі.

**Інтенсивність структурних зрушень** – оцінюється за допомогою **середнього відхилення часток**:

$$l_d = \frac{\sum_1^m |d_{j1} - d_{j0}|}{m}, \quad (4.27)$$

або **середнього квадратичного відхилення часток**:

$$d_d = \sqrt{\frac{\sum_1^m |d_{j1} - d_{j0}|^2}{m}}, \quad (4.28)$$

де  $m$  – число складових (груп) сукупності;  $d_{j1}$  – частки поточного періоду;  $d_{j0}$  – частки базисного періоду.

## 4.9. Види та взаємозв'язок дисперсій

### Правило складання дисперсій.

Розглянемо структуровану сукупність, яка поділена на групи за факторною ознакою « $\alpha$ », відповідна їй результативна ознака « $y$ » теж буде поділена на групи. У кожній групі розраховано групову середню арифметичну результативної ознаки  $\bar{y}_j$ ; частота  $f_j$  групи « $j$ ». Загальна середня результативних ознак усієї сукупності  $\bar{y}$ .

У кожній групі обчислюється **внутрішньогрупова дисперсія**:

$$\sigma_j^2 = \frac{\sum_1^{f_j} (y_j - \bar{y}_j)^2}{f_j}, \quad (4.29)$$

де  $y_j$  – значення ознаки елементів « $j$ » групи сукупності, їх кількість  $f_j$ .

Для всіх груп у цілому, тобто сукупності, обчислюється **середня з внутрішньогрупових дисперсій**:

$$\bar{\sigma}_j^2 = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j}; \quad (4.30)$$

$m$  – кількість груп на яку поділена сукупність.

**Міжгрупова дисперсія** обчислюється за формулою:

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 \cdot f_j}{\sum_1^m f_j}; \quad (4.31)$$

$$\sum_1^m f_j = n;$$

де  $n$  – кількість елементів сукупності.

**Загальна дисперсія сукупності** за правилом складання дисперсій

$$\sigma^2 = \delta^2 + \bar{\sigma}_j^2; \quad (4.32)$$

чи безпосередньо:

$$\sigma^2 = (\bar{y}^2) - (\bar{y})^2 = \frac{\sum_1^m (y_j - \bar{y})^2 f_j}{\sum_1^m f_j} = \frac{\sum_1^n (y_i - \bar{y})^2}{n}. \quad (4.33)$$

$$\text{Кореляційне відношення: } \eta^2 = \frac{\delta^2}{\sigma^2}; \quad (4.34)$$

Кореляційне відношення вимірює вплив факторної ознаки « $x$ », за якою поділена генеральна сукупність на групи, на результативну « $y$ », або щільність зв'язку між факторною та результативною ознаками.

Внутрішньогрупова дисперсія вимірює розшарування ознак у межах кожної з груп і пояснює дію різнобічних факторів у межах групи.

Міжгрупова дисперсія вимірює розшарування ознак, яке викликано фактором, за яким сукупність поділяли на окремі групи.

## **Завдання для самоконтролю**

1. Що називають рядом розподілу?
2. Надайте визначення варіанти, частоти і частки.
3. Чим варіаційний ряд розподілу відрізняється від атрибутивного?
4. Що то є діаграма «казусів», огіва, полігон, гістограма?
5. Що то є кумулята?
6. Що то є густина розподілу, диференціальна крива розподілу, інтегральна крива розподілу?

7. Які існують способи утворення вторинних групувань статистичних даних?
8. Як розраховується мода ряду розподілу неперервної, дискретної та альтернативної ознак?
9. Як розраховується медіана ряду розподілу неперервної та дискретної ознак?
10. Що називають відкритим та закритим інтервалами ряду розподілу?
11. Як визначити розмір відкритого інтервалу?
12. Що є варіацією ознаки та від чого залежить її розмах?
13. Як обчислюється квартальний розмах варіації?
14. Як обчислюється децильний розмах варіації?
15. Що є середнім лінійним відхиленням?
16. Що є середнім квадратичним відхиленням?
17. Що є середнім квадратом відхилення (дисперсія)?
18. Перелічить коефіцієнти варіації.
19. Як розраховується стандартизоване (нормоване) відхилення?
20. На що вказує довірче число (коефіцієнт довіри, квантиль розподілу)?
21. За якою формулою розраховується диференціальний закон Гауса?
22. За якою формулою розраховується інтегральний закон Гауса?
23. Чим закон розподілу Стьюдента відрізняється від закон розподілу Гауса?
24. У яких випадках використовують закон розподілу Стьюдента, а в яких законом розподілу Гауса?
25. Назвіть властивості дисперсії.
26. Що характеризує і як розраховується асиметрія?
27. Що характеризує і як розраховується скошеність функції (ексцес)?
28. Що характеризує і як розраховується коефіцієнт концентрації?
29. Що характеризує і як розраховується коефіцієнт локалізації?
30. Що характеризує і як розраховується коефіцієнт подібності?
31. Як розраховується дисперсія альтернативної ознаки?
32. У чому полягає правило складання дисперсій?
33. Як розраховується і який зміст має внутрішньогрупова дисперсія?
34. Як розраховується і який зміст має середня з внутрішньогрупових дисперсій?
35. Як розраховується і який зміст має міжгрупова дисперсія?
36. Як розраховується загальна дисперсія?
37. Як розраховується і який зміст має кореляційне відношення?
38. Що характеризує і як розраховується асиметрія?

39. Що характеризує і як розраховується скошеність функції (ексцес)?
40. Що характеризує і як розраховується коефіцієнт концентрації?

## Задачі

**4.1.** За даними опитування 300 осіб індивідуальні оцінки умов життя в цілому розподілились так: нестерпно – 9; погано – 66; посе-редньо – 135; добре – 72; дуже добре – 18. Замініть групові частоти частками, для кожної групи визначте кумулятивні частки; проведіть частотний аналіз розподілу, зробіть висновки.

**Розв'язок:**

Таблиця 4.1

Умови життя	Кількість осіб ( $f_i$ )	Частка ( $d_i$ )	Кумулятивна частка ( $g_i$ )
Нестерпні	9	0,03	0,03
Погані	66	0,22	0,25
Посередні	135	0,45	0,70
Добрі	72	0,24	0,94
Дуже добрі	18	0,06	1
Всього	300	1	-

$$d_i = f_i / 300$$

$$S_1 = d_1, S_2 = d_1 + d_2, S_3 = d_1 + d_2 + d_3 \text{ і т. д.}$$

Найбільше осіб оцінили свої умови життя як посередні – 135, що відповідає частці 0,45, найменше осіб вважають умови життя нестерпними – 9, що відповідає 0,03.

**4.2.** Склад засуджених у регіоні за віком характеризується даними, % до підсумку:

Таблиця 4.2

Вік, років	14-18	18-25	25-30	30-50	50 і старші	Разом
Кількість засуджених, %	12	28	20	34	6	100

Проведіть частотний аналіз розподілу, використовуючи кумулятивні частки та щільності часток.



**Розв'язок:**

Таблиця

**Частотний аналіз розподілу засуджених**

<b>Вік, років</b>	<b>14-18</b>	<b>18-25</b>	<b>25-30</b>	<b>30-50</b>	<b>50 і старші</b>	<b>Разом</b>
Кількість засуджених,%	12	28	20	34	6	100
Частки ( $d_i$ )	0,12	0,28	0,2	0,34	0,06	1
Кумулятивні частки ( $S_d$ )	0,12	0,4	0,6	0,94	1	
Щільність часток ( $g_i$ )	0,03	0,04	0,01	0,017	0,003	

$S_1 = d_1, S_2 = d_1 + d_2, S_3 = d_1 + d_2 + d_3$  і т. д.

$g_i = d_i / h_i$ , де  $h_i$  – ширина  $i$ -го інтервалу.

Найбільша частка засуджених мають 30-50 років – 0,34, найменше засуджених у віці 50 і більше років – 0,06. Половина засуджених припадає на вік 25-30 років, найбільшу щільність на одиницю інтервалу мають засуджені у віці 18-25 років, найменшу – 50 і більше років.

**4.3.** За результатами аналізу плавки легованої сталі вміст нікелю становить, %:

Таблиця 4.3

<b>Вміст нікелю, %</b>	<b>До 4,0</b>	<b>4,0-4,2</b>	<b>4,2-4,4</b>	<b>4,4 і більше</b>	<b>Разом</b>
Кількість плавок	11	47	33	9	100

Визначте середній процент вмісту нікелю, модальний та медіанний інтервали.

**Розв'язок:**

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{3,9 \cdot 11 + 4,1 \cdot 47 + 4,3 \cdot 33 + 4,5 \cdot 9}{100} = \frac{418}{100} = 4,18\%.$$

Мода визначається візуально за максимальною частотою, або часткою.

Модальний інтервал: (4,0 – 4,2), бо  $f_2 = 47$  – найбільша.

Медіаною буде значення ознаки, для якої кумулятивна частота  $Sf_j > 0,5$ .

Медіанний інтервал: (4,0 – 4,2), бо  $S_2 = 58 > 100/2$ .

**4.4.** Закуплені м'ясокомбінатом 200 голів великої рогатої худоби за вагою розподілились так:

Таблиця 4.4

Вага однієї голови, ц	До 3,2	3,2-3,6	3,6-4,0	4,0-4,4	4,4 і більше	Разом
Кількість голів	10	40	85	50	15	200

Визначте середню вагу однієї голови закупленої рогатої худоби.

**Розв'язок:**

$$\bar{x} = \frac{\sum_{j=1}^m x_j f_j}{\sum_{j=1}^m f_j} = \frac{3,0 \cdot 10 + 3,4 \cdot 40 + 3,8 \cdot 85 + 4,2 \cdot 50 + 4,6 \cdot 15}{200} = \frac{768}{200} = 3,84 \text{ ц.}$$

**4.5.** За наведеними даними про попит на взуття для молодших школярів визначте окремо для дівчаток та хлопчиків модальні розміри взуття:

Таблиця 4.5

Розмір	Продано пар взуття для ( $f_i$ )	
	дівчаток	хлопчиків
18	22	–
18,5	33	12
19	20	18
19,5	18	20
20	6	34
20,5	1	16
Разом	100	100

**Розв'язок:**

Мода – значення ознаки, яка найчастіше зустрічається, визначається візуально за максимальною частотою.

Дівчата:  $M_o = 18,5$ , бо  $f_i = 33$  – максимальне,

хлопчики:  $M_o = 20$ , бо  $f_i = 34$  – максимальне.

**4.6.** Опитування жінок різних країн щодо бажаної кількості дітей дало такі результати:

Таблиця 4.6

Бажана кількість дітей	% до загальної кількості опитаних у країнах		
	А	В	С
0	–	1,0	0,5
1	8,8	35,3	28,4
2	22,1	53,5	38,2
3	48,2	9,4	22,6
4	13,7	0,8	8,2
5 і більше	7,2	–	2,1
Разом	100	100	100

Визначте модальну кількість бажаних дітей у кожній країні. У якій країні варіація цього показника найбільша?

**Розв'язок:**

Мода визначається візуально за максимальною частотою. Для країни А:  $M_o = 3$ , бо  $f_i = 48,2$  – максимальне; для країни В:  $M_o = 2$ , бо  $f_i = 53,5$  – максимальне; для країни С:  $M_o = 2$ , бо  $f_i = 38,2$  – максимальне.

Квадратичний коефіцієнт варіації:  $V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%$ .

Для країни А:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{1 \cdot 8,8 + 2 \cdot 22,1 + 3 \cdot 48,2 + 4 \cdot 13,7 + 5 \cdot 7,2}{100} = 2,88.$$

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}}.$$

$$\sigma = \sqrt{\frac{(1-2,88)^2 \cdot 8,8 + (2-2,88)^2 \cdot 22,1 + (3-2,88)^2 \cdot 48,2 + (4-2,88)^2 \cdot 13,7 + (5-2,88)^2 \cdot 7,2}{100}} = 0,99.$$

$$V_\sigma = \frac{0,99}{2,88} 100\% = 34,36\%.$$

Для країни В:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{0 \cdot 1,0 + 1 \cdot 35,3 + 2 \cdot 53,5 + 3 \cdot 9,4 + 4 \cdot 0,8}{100} = 1,737.$$

$$\sigma = \sqrt{\frac{(0-1,737)^2 \cdot 1,0 + (1-1,737)^2 \cdot 35,3 + (2-1,737)^2 \cdot 53,5 + (3-1,737)^2 \cdot 9,4 + (4-1,737)^2 \cdot 0,8}{100}}$$

$$= 0,67.$$

$$V_\sigma = \frac{0,67}{1,737} 100\% = 38,57\%.$$

Для країни С:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{0 \cdot 0,5 + 1 \cdot 28,4 + 2 \cdot 38,2 + 3 \cdot 22,6 + 4 \cdot 8,2 + 5 \cdot 2,1}{100} = 2,159.$$

$$\sigma = \sqrt{\frac{(0-2,159)^2 \cdot 0,5 + (1-2,159)^2 \cdot 28,4 + (2-2,159)^2 \cdot 38,2 + (3-2,159)^2 \cdot 22,6 + (4-2,159)^2 \cdot 8,2 + (5-2,159)^2 \cdot 2,1}{100}} = 1,01.$$

$$V_\sigma = \frac{1,01}{2,159} 100\% = 46,78\%.$$

Отже, для жінок ступінь варіації бажання мати дітей у країні С вищий, ніж у країні В і вищий, ніж у країні А.

**4.7.** Результати тестування професійної психодіагностики такі:

*Таблиця 4.7*

Первинні тестові оцінки, бали	Кількість обстежених
15	7
16	12
17	21
18	39
19	26
20	15
Разом	120

Визначте медіану та кватильні тестові оцінки, поясніть їх зміст.

**Розв'язок:**

Таблиця

**Кумулятивні частоти тестових оцінок**

Первинні тестові оцінки, бали	Кількість обстежених	Кумулятивна частота ( $S_j$ )
15	7	7
16	12	19
17	21	40
18	39	79
19	26	105
20	15	120
Разом	120	

$S_1 = f_1, S_2 = f_1 + f_2, S_3 = f_1 + f_2 + f_3$  і т. д.

Медіаною буде значення ознаки, для якої кумулятивна частота  $Sf_j > 0,5, M_e = 18$  бо  $S_4=79 > 120/2$ .

Медіана свідчить про те, що половина тестових оцінок буде менше 18, а половина – більше.

$Q_1 = 17$ , бо  $S_3 = 40 > \frac{1}{4} \cdot 120$ .

$Q_3 = 19$ , бо  $S_5 = 105 > \frac{3}{4} \cdot 120$ .

Квартильні оцінки свідчать, що 25% отриманих оцінок не перевищують 17, а 75% оцінок мають найбільшою 19, або з 25% найвищих оцінок, найнижчою є 19.

**4.8.** Терміни корисного використання нематеріальних активів фірми, захищених патентами, ліцензіями тощо, на кінець року становили:

Таблиця 4.8

Термін використання, років	1	2	3	4	5	6	Разом
% до загальної суми нематеріальних активів	2,6	10,2	39,2	40,0	6,8	1,2	100

Визначте ряд кумулятивних часток та медіанне значення терміну корисного використання нематеріальних активів фірми.

**Розв'язок:**

Таблиця

**Кумулятивні частки терміну використання активів, %**

Термін використання, років	1	2	3	4	5	6	Разом
% до загальної суми нематеріальних активів ( $d_i$ )	2,6	10,2	39,2	40,0	6,8	1,2	100
Кумулятивні частки ( $S_d$ )	2,6	12,8	52,0	92,0	98,8	100	

$$S_1 = d_1, S_2 = d_1 + d_2, S_3 = d_1 + d_2 + d_3 \text{ і т.д.}$$

Медіаною буде значення ознаки, для якої кумулятивна частка  $S_{d_i} > 0,5$ ,  $M_e = 3$ , бо  $S_3 = 52,0 > 100/2$ .

**4.9.** Розподіл емігрантів за віком характеризується даними, %:

Таблиця 4.9

Вік, років	Вибули у країні регіону	
	А	В
До 10	4,0	2,2
10-20	16,3	11,4
20-30	23,5	38,2
30-40	40,2	36,8
40-50	9,4	9,0
50-60	4,2	2,1
60 і старші	2,4	0,3
Разом	100	100

Визначте модальний вік емігрантів у країні кожного регіону, проведіть їх порівняльний аналіз.

**Розв'язок:**

В інтервальному ряді за найбільшою частотою визначається модальний інтервал. Конкретне значення моди в інтервалі обчислюється за формулою:

$$M_0 = x_0 + h \frac{f_{m0} - f_{m0-1}}{(f_{m0} - f_{m0-1}) + (f_{m0} - f_{m0+1})},$$

де  $x_0$  та  $h$  – відповідно нижня межа та ширина модального інтервалу;  $f_{m0}$ ,  $f_{m0-1}$ ,  $f_{m0+1}$  – це частоти (частки) модального, передмодального та післямодального інтервалу.

Для країни А:

Модальний інтервал дорівнює 30-40, оскільки  $f_i = 40,2$  – максимальне.

$$M_0 = 30 + 10 \frac{40,2 - 23,5}{(40,2 - 23,5) + (40,2 - 9,4)} = 33,5 \text{ років.}$$

Для країни В:

модальний інтервал дорівнює 20-30, оскільки  $f_i = 38,2$  – максимальне.

$$M_0 = 20 + 10 \frac{38,2 - 11,4}{(38,2 - 11,4) + (38,2 - 36,8)} = 29,5 \text{ років.}$$

У країну В в'їжджають переважно молодші люди, ніж у країну А. В країну А найбільше в'їжджає людей у віці 30-40 років, а в країну В – 20-30 років.

**4.10.** Розподіл безробітних за тривалістю перерви в роботі (місяців) характеризується даними:

Таблиця 4.10

Тривалість перерви в роботі, міс.	% до підсумку	
	чоловіки	жінки
До 1	5,3	18,8
1-3	59,6	27,2
3-6	33,4	40,0
6-12	1,7	14,0
Разом	100	100

За кожним рядом визначте кумулятивні частоти та медіанні значення тривалості перерви в роботі. Проведіть порівняльний аналіз.

**Розв'язок:**

Кумулятивні частоти обчислимо за формулами:  $S_1 = f_1$ ,  $S_2 = f_1 + f_2$ ,  $S_3 = f_1 + f_2 + f_3$ , і т.д.

Таблиця

Кумулятивні частоти тривалості перерви в роботі, місяців

Межі інтервалу	Кумулятивні частоти	
	Чоловіки	Жінки
До 1	5,3	18,8
1-3	64,9	46,0
3-6	98,3	86,0
6-12	100	100

Медіаною буде значення ознаки, для якої кумулятивна частота  $S_{dj} > 0,5$ .

Конкретне значення медіани в інтервалі обчислюється за формулою:

$$Me = x_0 + h \frac{0,5 \sum_{j=1}^m f_j - S_{f_{me-1}}}{f_{me}},$$

де  $x_0$  та  $h$  – відповідно нижня межа та ширина медіанного інтервалу;  $f_{me}$  – частота медіанного інтервалу;  $S_{f_{me-1}}$  – кумулятивна частота передмедіанного інтервалу.

Медіанний інтервал для чоловіків буде 1-3, конкретне значення медіани:

$$Me = 1 + 2 \frac{0,5 \cdot 100 - 5,3}{59,6} = 2,5 \text{ міс.}$$

Медіанний інтервал для жінок 3-6, конкретне значення медіани:

$$Me = 3 + 3 \frac{0,5 \cdot 100 - 46}{40} = 3,3 \text{ міс.}$$

Половина жінок мають тривалість перерви у роботі до 3,3 місяців, половина чоловіків – до 2,5 місяців, отже, чоловіки швидше знаходять роботу, ніж жінки.

**4.11.** Розподіл сільськогосподарських угідь регіону за природною родючістю ґрунтів характеризується даними:

Таблиця 4.11

Природна родючість ґрунтів, бали	% до загальної площі	
	північні райони	південні райони
До 60	13	8
60-70	32	30
70-80	28	50
80-90	19	17
90 і більше	8	5
Разом	100	100

За кожною групою районів визначте модальний рівень родючості ґрунтів. У яких районах ґрунти за родючістю однорідніші?

**Розв'язок:**

В інтервальному ряду за найбільшою частотою визначається модальний інтервал. Конкретне значення моди в інтервалі обчислюється за формулою:

$$M_0 = x_0 + h \frac{f_{m0} - f_{m0-1}}{(f_{m0} - f_{m0-1}) + (f_{m0} - f_{m0+1})},$$

де  $x_0$  та  $h$  – відповідно нижня межа та ширина модального інтервалу;  $f_{m0}$ ,  $f_{m0-1}$ ,  $f_{m0+1}$  – це частоти (частки) модального, передмодального та післямодального інтервалу.

Для північних районів:

Модальний інтервал дорівнює 60-70, оскільки  $f_i = 32$  – максимальне.



Конкретне значення моди дорівнює:

$$M_0 = 60 + 10 \frac{32 - 13}{(32 - 13) + (32 - 28)} = 68,26 \text{ балів.}$$

Для південних районів:

Модальний інтервал дорівнює 70-80, оскільки  $f_i = 50$  – максимальне.

Конкретне значення моди дорівнює:

$$M_0 = 70 + 10 \frac{50 - 30}{(50 - 30) + (50 - 17)} = 73,77 \text{ балів.}$$

Квадратичний коефіцієнт варіації використовують як критерій однорідності сукупності:

$$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%.$$

Для північних районів:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{55 \cdot 13 + 65 \cdot 32 + 75 \cdot 28 + 85 \cdot 19 + 95 \cdot 8}{100} = \frac{7015}{100} = 70,15 \text{ бали.}$$

$$\sigma = \sqrt{\frac{\sum_1^n (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}}.$$

$$\sigma = \sqrt{\frac{(55 - 70,15)^2 \cdot 13 + (65 - 70,15)^2 \cdot 32 + (75 - 70,15)^2 \cdot 28 + (85 - 70,15)^2 \cdot 19 + (95 - 70,15)^2 \cdot 8}{100}} = 11,67 \text{ бали.}$$

$$V_\sigma = \frac{11,67}{70,15} * 100\% = 16,64\%.$$

Для південних районів:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{55 \cdot 8 + 65 \cdot 30 + 75 \cdot 50 + 85 \cdot 17 + 95 \cdot 5}{100} = \frac{8060}{100} = 80,60 \text{ бали.}$$

$$\sigma = \sqrt{\frac{\sum_1^n (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}}.$$

$$\sigma = \sqrt{\frac{(55 - 80,60)^2 \cdot 8 + (65 - 80,60)^2 \cdot 30 + (75 - 80,60)^2 \cdot 50 + (85 - 80,60)^2 \cdot 17 +}{100} + \frac{(95 - 80,60)^2 \cdot 5}{100}} = 12,44 \text{ бали.}$$

$$V_{\sigma} = \frac{12,44}{80,60} * 100\% = 15,43\%.$$

За квадратичним коефіцієнтом варіації в південних регіонах більш однорідні ґрунти.

**4.12.** Розподіл новонароджених за віком матері характеризується даними:

Таблиця 4.12

Вік матері, років	% до підсумку	
	Місто	Село
До 20	14,0	19,5
20-25	40,0	43,1
25-30	27,1	22,1
30-35	13,2	10,4
35-40	4,7	4,0
40 і старші	1,0	0,9
Разом	100	100

Окремо для міст і сіл визначте медіанне значення віку матері, порівняйте їх.

**Розв'язок:**

Таблиця

Кумулятивні частки віку матерів, роки

Вік матері, років	% до підсумку			
	Місто	Кумулятивна частка ( $S_d$ )	Село	Кумулятивна частка ( $S_d$ )
До 20	14,0	14,0	19,5	19,5
20-25	40,0	54,0	43,1	62,6
25-30	27,1	81,1	22,1	84,7
30-35	13,2	94,3	10,4	95,1
35-40	4,7	99,0	4,0	99,1
40 і старші	1,0	100	0,9	100
Разом	100		100	

$S_1 = d_1, S_2 = d_1 + d_2, S_3 = d_1 + d_2 + d_3$  і т. д.

Медіаною буде значення ознаки, для якої кумулятивна частка  $S_{d_j} > 0,5$ . В інтервальному ряду в такий спосіб визначається медіанний інтервал. Конкретне значення медіани в інтервалі обчислюється за формулою:

$$Me = x_0 + h \frac{0,5 \sum_{j=1}^m f_j - S_{f_{me-1}}}{f_{me}},$$

де  $x_0$  та  $h$  – відповідно нижня межа та ширина медіанного інтервалу;  $f_{me}$  – частота медіанного інтервалу;  $S_{f_{me-1}}$  – кумулятивна частота передмедіанного інтервалу.

Для міста:

Медіанний інтервал: (20 – 25), бо  $S_2 = 54 > 100/2$ .

$$Me = 20 + 5 \frac{0,5 \cdot 100 - 14,0}{40,0} = 24,5 \text{ роки.}$$

Для села:

Медіанний інтервал: (20 – 25), бо  $S_2 = 62,6 > 100/2$ .

$$Me = 20 + 5 \frac{0,5 \cdot 100 - 19,5}{43,5} = 23,5 \text{ роки.}$$

У місті половина матерів народжують у віці до 24,5 років, а в селі – до 23,5, отже, в селі народжують дітей раніше.

**4.13.** За наведеними у таблиці даними проведіть порівняльний аналіз варіації показників, зробіть висновки.

Таблиця 4.13

Регіон	Густина населення, чол. на 1 кв. км	Зайнятість працездатного населення, %	Рівень злочинності на 100 000 чол.
А	82,0	80,2	824
В	91,8	68,0	913
С	126,5	72,2	1407
Д	85,4	83,5	585
К	62,9	79,1	450
М	115,2	76,7	1036
Р	70,6	85,8	748
Разом по країні	84,8	81,0	870

**Розв'язок:**

Порівнюючи варіації різних ознак або однієї ознаки у різних сукупностях, використовують відносні характеристики варіації:

1) лінійний коефіцієнт варіації:

$$V_l = \frac{\bar{l}}{\bar{x}} 100\% ,$$

$$\bar{l} = \frac{\sum_{j=1}^n |x_j - \bar{x}|}{n}$$

2) квадратичний коефіцієнт варіації:

$$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\% ,$$

$$\sigma = \sqrt{\frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n}} .$$

3) коефіцієнт осциляції:

$$V_R = \frac{R}{\bar{x}} 100\% ,$$

$$R = x_{max} - x_{min}$$

а) Для густоти населення:

$$\bar{l} = \frac{|82,0 - 84,8| + |91,8 - 84,8| + |126,5 - 84,8| + |85,4 - 84,8| + |62,9 - 84,8| + |115,2 - 84,8| +$$

$$+ |70,6 - 84,8|}{7} = 16,94 \text{ чол. на 1 кв. км.}$$

$$V_l = \frac{16,94}{84,8} 100\% = 20\%$$

$$\sigma = \sqrt{\frac{(82,0 - 84,8)^2 + (91,8 - 84,8)^2 + (126,5 - 84,8)^2 + (85,4 - 84,8)^2 + (62,9 - 84,8)^2 +$$

$$+ (115,2 - 84,8)^2 + (70,6 - 84,8)^2}{7}} = 22,04 \text{ чол. на 1 кв. км.}$$

$$V_\sigma = \frac{22,04}{84,8} 100\% = 25,99\% .$$

$$V_R = \frac{126,5 - 62,9}{84,8} 100\% = 75\% .$$

б) Для зайнятості працездатного населення:

$$\bar{i} = \frac{|80,2 - 81,0| + |68,0 - 81,0| + |72,2 - 81,0| + |83,5 - 81,0| + |79,1 - 81,0| + |76,7 - 81,0| + |85,8 - 81,0|}{7} = 36,1\%.$$

$$V_i = \frac{36,1}{81,0} 100\% = 44,57\%.$$

$$\sigma = \sqrt{\frac{(80,2 - 81,0)^2 + (68,0 - 81,0)^2 + (72,2 - 81,0)^2 + (83,5 - 81,0)^2 + (79,1 - 81,0)^2 + (76,7 - 81,0)^2 + (85,8 - 81,0)^2}{7}} = 6,53\%.$$

$$V_\sigma = \frac{6,53}{81,0} 100\% = 8,06\%.$$

$$V_R = \frac{85,8 - 68,0}{81,0} 100\% = 22\%.$$

в) Для рівня злочинності:

$$\bar{i} = \frac{|824 - 870| + |913 - 870| + |1407 - 870| + |585 - 870| + |450 - 870| + |1036 - 870| + |748 - 870|}{7} = 231,29.$$

$$V_i = \frac{231,29}{870} 100\% = 26,58\%.$$

$$\sigma = \sqrt{\frac{(824 - 870)^2 + (913 - 870)^2 + (1407 - 870)^2 + (585 - 870)^2 + (450 - 870)^2 + (1036 - 870)^2 + (748 - 870)^2}{7}} = 290,91.$$

$$V_\sigma = \frac{290,91}{870} 100\% = 33,44\%.$$

$$V_R = \frac{1407 - 450}{870} 100\% = 110\%.$$

Найбільший лінійний коефіцієнт варіації має зайнятість населення, рівень злочинності найбільше варіює за квадратичним коефіцієнтом варіації та за коефіцієнтом осциляції.

**4.14.** Результати технічного аналізу шахти (7 проб) характеризуються даними, %:

Вологість	8,5;	8,3;	8,8;	8,2;	8,4;	9,1;	8,9
Сірка	1,92;	1,96;	1,88;	1,97;	2,04;	2,02;	2,00

Проведіть порівняльний аналіз варіації показників, зробіть висновки.

**Розв'язок:**

Порівнюючи варіації різних ознак або однієї ознаки у різних сукупностях, використовують відносні характеристики варіації:

1) лінійний коефіцієнт варіації:

$$V_l = \frac{\bar{l}}{\bar{x}} 100\%,$$

$$\bar{l} = \frac{\sum_1^n |x_j - \bar{x}|}{n}.$$

2) квадратичний коефіцієнт варіації:

$$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\%,$$

$$\sigma = \sqrt{\frac{\sum_1^n (x_j - \bar{x})^2}{n}}.$$

3) коефіцієнт осциляції:

$$V_R = \frac{R}{\bar{x}} 100\%,$$

$$R = x_{max} - x_{min}.$$

Для вологості:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{8,5 + 8,3 + 8,8 + 8,2 + 8,4 + 9,1 + 8,9}{7} = \frac{60,2}{7} = 8,6\%.$$

$$\bar{l} = \frac{|8,5 - 8,6| + |8,3 - 8,6| + |8,8 - 8,6| + |8,2 - 8,6| + |8,4 - 8,6| + |9,1 - 8,6| + |8,9 - 8,6|}{7} = 0,29\%.$$

$$V_l = \frac{0,29}{8,6} 100\% = 3,37\%.$$

$$\sigma = \sqrt{\frac{(8,5-8,6)^2 + (8,3-8,6)^2 + (8,8-8,6)^2 + (8,2-8,6)^2 + (8,4-8,6)^2 + (9,1-8,6)^2 + (8,9-8,6)^2}{7}} = 0,3116\%.$$

$$V_\sigma = \frac{0,3116}{8,6} 100\% = 3,62\%.$$

$$V_R = \frac{9,1-8,2}{8,6} 100\% = 10\%.$$

Для рівня сірки:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{1,92 + 1,96 + 1,88 + 1,97 + 2,04 + 2,02 + 2,00}{7} = \frac{13,79}{7} = 1,97\%.$$

$$\bar{j} = \frac{|1,92-1,97| + |1,96-1,97| + |1,88-1,97| + |1,97-1,97| + |2,04-1,97| + |2,02-1,97| + |2,00-1,97|}{7} = 0,04\%.$$

$$V_l = \frac{0,04}{1,97} 100\% = 2,03\%.$$

$$\sigma = \sqrt{\frac{(1,92-1,97)^2 + (1,96-1,97)^2 + (1,88-1,97)^2 + (1,97-1,97)^2 + (2,04-1,97)^2 + (2,02-1,97)^2 + (2,00-1,97)^2}{7}} = 0,05\%.$$

$$V_\sigma = \frac{0,05}{1,97} 100\% = 2,54\%.$$

$$V_R = \frac{2,04-1,88}{1,97} 100\% = 8\%.$$

За всіма показниками найбільше варіює рівень вологості, тож дані про рівень сірки більш точні.

**4.15.** Рівень рентабельності підприємств легкої та харчової промисловості характеризується даними:

Таблиця 4.14

Рівень рентабельності, %	% до підсумку	
	Легка промисловість	Харчова промисловість
До 5	3	8
5-10	8	15
10-15	16	21
15-20	22	26
20-25	24	17
25-30	18	9
30 і вище	9	4
Разом	100	100

Для кожної галузі визначте квартилі рівня рентабельності, поясніть їх зміст. Порівняйте варіацію, зробіть висновки.

**Розв'язок:**

Розрахунок квартилів та децилів ґрунтується на кумулятивних частках:  $S_1 = d_1$ ,  $S_2 = d_1 + d_2$ ,  $S_3 = d_1 + d_2 + d_3$  і т. д.

Таблиця

**Кумулятивні частки рівня рентабельності, %**

Рівень рентабельності, %	% до підсумку			
	Легка промисловість	Кумулятивна частка ( $S_d$ )	Харчова промисловість	Кумулятивна частка ( $S_d$ )
До 5	3	3	8	8
5-10	8	11	15	23
10-15	16	27	21	44
15-20	22	49	26	70
20-25	24	73	17	87
25-30	18	91	9	96
30 і вище	9	100	4	100
Разом	100		100	

Перший та третій квартилі визначаються за формулами:  
перший квартиль:

$$Q_1 = x_o + h \frac{0,25 \sum_1^m f_j - S_{Q_{1-1}}}{f_{Q_1}},$$



третій квартиль:

$$Q_3 = x_o + h \frac{0,75 \sum_1^m f_j - S_{Q_{3-1}}}{f_{Q_3}},$$

де  $x_o$  та  $h$  – відповідно нижня межа та ширина квартильного інтервалу;  $f_{Q_1}, f_{Q_3}$  – частота квартильного інтервалу;  $S_{Q_{1-1}}, S_{Q_{3-1}}$  – кумулятивна частота передквартильного інтервалу.

а) Для легкої промисловості:

Перший квартильний інтервал: 10-15, бо  $S_3 = 27 > 100/4$ .

$$Q_1 = 10 + 5 \frac{0,25 \cdot 100 - 11}{16} = 14,38 \%$$

Другий квартильний інтервал: 20-25, бо  $S_5 = 73 > 100/2$ .

$$Q_2 = 20 + 5 \frac{0,5 \cdot 100 - 49}{24} = 20,2 \%$$

Третій квартильний інтервал: 25-30, бо  $S_6 = 91 > 3/4 \cdot 100$ .

$$Q_3 = 25 + 5 \frac{0,75 \cdot 100 - 73}{18} = 25,56 \%$$

б) Для харчової промисловості:

Перший квартильний інтервал: 10-15, бо  $S_3 = 44 > 100/4$ .

$$Q_1 = 10 + 5 \frac{0,25 \cdot 100 - 23}{21} = 10,48 \%$$

Другий квартильний інтервал: 15-20, бо  $S_4 = 70 > 100/2$ .

$$Q_2 = 25 + 5 \frac{0,5 \cdot 100 - 44}{26} = 16,5 \%$$

Третій квартильний інтервал: 20-25, бо  $S_5 = 87 > 3/4 \cdot 100$ .

$$Q_3 = 20 + 5 \frac{0,75 \cdot 100 - 70}{17} = 21,47 \%$$

Для легкої промисловості  $1/4$  частина всіх підприємств мають рівень рентабельності до 14,38 %, половина – до 20,2 %, 25 % всіх підприємств із найвищим рівнем рентабельності, найнижчий мають 25,56 %.

Для харчової промисловості  $1/4$  частина всіх підприємств мають рівень рентабельності до 10,48 %, половина – до 16,5 %, 75 % всіх підприємств – до 21,47 %.

Так як центр розподілу представлений медіаною, то використаємо квартильний коефіцієнт варіації:  $V_Q = \frac{Q_3 - Q_1}{2Me}$ .

Для легкої промисловості:

$$V_e = \frac{25,56 - 14,38}{2 \cdot 20,2} = 0,29.$$

Для харчової промисловості:

$$V_e = \frac{21,47 - 10,48}{2 \cdot 16,5} = 0,48.$$

Більшу варіацію мають підприємства харчової промисловості.

**4.16.** Розподіл домогосподарств за рівнем середньодушового доходу характеризується наведеними в таблиці даними.

Для кожного регіону визначте децилі середньодушового доходу, поясніть їх зміст. Оцініть ступінь децильної диференціації домогосподарств за рівнем середньодушового доходу, зробіть висновки.

Таблиця 4.15

Середньодушовий дохід, гр. од.	Кількість домогосподарств, од.	
	регіон А	регіон В
До 20	22	5
20-40	34	12
40-60	62	24
60-80	59	56
80-100	47	109
100-120	20	88
120-140	6	67
140 і вище	–	39
Разом	250	400

**Розв'язок:**

Розрахунок децилів ґрунтується на кумулятивних частотах:  $S_1 = f_1$ ,  $S_2 = f_1 + f_2$ ,  $S_3 = f_1 + f_2 + f_3$  і т. д.

Таблиця

**Кумулятивні частоти середньодушового доходу**

Середньодушовий дохід, гр. од.	Кількість домогосподарств, од.			
	Регіон А	Кумулятивна частота ( $S_f$ )	Регіон В	Кумулятивна частота ( $S_f$ )
До 20	22	22	5	5
20-40	34	56	12	17
40-60	62	118	24	41
60-80	59	177	56	97
80-100	47	224	109	206
100-120	20	244	88	294
120-140	6	250	67	361
140 і вище	–	–	39	400
Разом	250		400	

Перший дециль обчислюється за формулою:

$$D_1 = x_{D_1} + h \frac{0,1 \sum_1^m f_j - S_{D_1-1}}{f_{D_1}},$$

де  $x_{D_1}$  та  $h$  – відповідно нижня межа та ширина децильного інтервалу;  $f_{D_1}$  – частота децильного інтервалу;  $S_{D_1-1}$  – кумулятивна частота децильного інтервалу.

Інші децилі розраховуються аналогічно.

Мірою оцінки розшарування сукупності слугує коефіцієнт децильної диференціації:  $V_D = \frac{D_9}{D_1}$ .

Для регіону А:

$$D_1 = 20 + 20 \frac{0,1 \cdot 250 - 22}{34} = 21,76.$$

$$D_2 = 20 + 20 \frac{0,2 \cdot 250 - 22}{34} = 36,47.$$

$$D_3 = 40 + 20 \frac{0,3 \cdot 250 - 56}{62} = 46,13.$$

$$D_4 = 40 + 20 \frac{0,4 \cdot 250 - 56}{62} = 54,19.$$

$$D_5 = 60 + 20 \frac{0,5 \cdot 250 - 118}{59} = 62,37.$$

$$D_6 = 60 + 20 \frac{0,6 \cdot 250 - 118}{59} = 70,84.$$

$$D_7 = 60 + 20 \frac{0,7 \cdot 250 - 118}{59} = 79,32.$$

$$D_8 = 80 + 20 \frac{0,8 \cdot 250 - 177}{47} = 89,79.$$

$$D_9 = 100 + 20 \frac{0,9 \cdot 250 - 224}{20} = 101.$$

$$V_D = \frac{101}{21,79} = 4,64.$$

Значення децилів вказують на те, що серед 10% домогосподарств з найнижчим середньодушовим доходом, найбільший дохід становить 21,76 грошових одиниць, а серед 10% домогосподарств з найбільшим доходом найменший – 101 грошова одиниця, тобто у 4,6 разів більший.

Для регіону В значення децилів розраховуються аналогічно.

$$D_1 = 59,17;$$

$$D_2 = 73,93;$$

$$D_3 = 84,22;$$

$$D_4 = 91,56;$$

$$D_5 = 98,90;$$

$$D_6 = 107,73;$$

$$D_7 = 116,82;$$

$$D_8 = 127,76;$$

$$D_9 = 139,70;$$

$$V_D = \frac{139,70}{59,17} = 2,36.$$

Значення децилів вказують на те, що серед 10% домогосподарств з найнижчим середньодушовим доходом, найбільший дохід становить 59,17 грошових одиниць, а серед 10% домогосподарств з найбільшим доходом найменший – 139,70 грошових одиниць, тобто у 2,4 рази більший.

Більшу варіацію домогосподарств за середньодушовим доходом має регіон А, більший середньодушовий дохід мають домогосподарства регіону В.

**4.17.** Перевірка інспекцією якості твердих сирів сорту «Р» за вмістом жиру дала такі результати:

Таблиця 4.16

Вміст жиру, %	44	45	46	47	48	Разом
Кількість проб	1	5	9	3	2	20

Визначте середній процент вмісту жиру у твердих сирах та середнє лінійне відхилення.

**Розв'язок:**

1)  $\bar{x}$  – середній процент вмісту жиру у твердих сирах

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{44 \cdot 1 + 45 \cdot 5 + 47 \cdot 3 + 48 \cdot 2}{20} = 25,3 \%$$

2)  $\bar{l}$  – середнє лїнійне вїдхилення

$$\bar{l} = \frac{\sum_1^m |x_j - \bar{x}| f_j}{\sum_1^m f_j} = \frac{|44 - 25,3| \cdot 1 + |45 - 25,3| \cdot 5 + |46 - 25,3| \cdot 9 + |47 - 25,3| \cdot 3 + |48 - 25,3| \cdot 2}{20} = 20,7\%.$$

**4.18.** Розподїл робїтнїкїв двох галузей промисловостї за рївнем квалїфїкацїї характеризується даними, % до пїдсумку:

Таблиця 4.17

Тарифний розряд	Галузь промисловостї	
	А	В
2	9	5
3	20	17
4	35	30
5	24	29
6	12	19
Разом	100	100

Визначте для кожної галузі середній тарифний розряд та середнє лїнійне вїдхилення; порївняйте варїацїю тарифних розрядїв.

**Розв'язок:**

1)  $\bar{x}$  – середній тарифний розряд:

а) галузь А:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{2 \cdot 9 + 3 \cdot 20 + 4 \cdot 35 + 5 \cdot 24 + 6 \cdot 12}{100} = 4,1.$$

б) галузь В:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{2 \cdot 5 + 3 \cdot 17 + 4 \cdot 30 + 5 \cdot 29 + 6 \cdot 19}{100} = 4,4.$$

2)  $\bar{l}$  – середнє лїнійне вїдхилення:

а) галузь А:

$$\bar{l} = \frac{\sum_1^m |x_j - \bar{x}| f_j}{\sum_1^m f_j} = \frac{|2 - 4,1| \cdot 9 + |3 - 4,1| \cdot 20 + |4 - 4,1| \cdot 35 + |5 - 4,1| \cdot 24 + |6 - 4,1| \cdot 12}{100} = 0,89.$$

б) галузь В:

$$\bar{l} = \frac{\sum_1^m |x_j - \bar{x}| f_j}{\sum_1^m f_j} = \frac{|2 - 4,4| \cdot 5 + |3 - 4,4| \cdot 17 + |4 - 4,4| \cdot 30 + |5 - 4,4| \cdot 29 + |6 - 4,4| \cdot 19}{100} = 0,956.$$

3)  $V_l$  – коефіцієнт варіації.

а) галузь А:

$$V_l = \frac{\bar{l}}{\bar{x}} 100\% = \frac{0,89}{4,1} = 0,217 \cdot 100\% = 21,7\%.$$

б) галузь В:

$$V_l = \frac{\bar{l}}{\bar{x}} 100\% = \frac{0,956}{4,4} = 0,2173 \cdot 100\% = 21,73\%.$$

Коефіцієнт варіації є критерієм відхилення індивідуальних значень ознаки від центру розподілу. В галузі А та В вони практично рівні (21,7%), що свідчить про їх однакове відхилення.

**4.19.** Кредитні ставки комерційних банків під короткострокові позики становили:

Таблиця 4.18

Кредитна ставка, %	Суми наданих позик, млн гр. од.	
	I квартал	II квартал
До 30	1	5
30-40	4	11
40-50	9	8
50 і більше	6	6
Разом	20	30

За кожний квартал визначте середню кредитну ставку та середнє лінійне відхилення. Як змінились середній рівень та варіація кредитної ставки?

**Розв'язок:**

1)  $\bar{x}$  – середня кредитна ставка:

а) I квартал:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{25 \cdot 1 + 35 \cdot 4 + 45 \cdot 9 + 55 \cdot 6}{20} = 40,5\%.$$

б) II квартал:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{25 \cdot 5 + 35 \cdot 11 + 45 \cdot 8 + 55 \cdot 6}{30} = 40\%.$$

Середній рівень кредитної ставки зменшився з 40,5 до 40 відсотків, тобто на 1,2%.

2)  $\bar{l}$  – середнє лінійне відхилення:

а) I квартал:

$$\bar{l} = \frac{\sum_1^m |x_j - \bar{x}| f_j}{\sum_1^m f_j} = \frac{|25 - 40,5| \cdot 1 + |35 - 40,5| \cdot 4 + |45 - 40,5| \cdot 9 + |55 - 40,5| \cdot 6}{20} = 8,25\%.$$

б) II квартал:

$$\bar{l} = \frac{\sum_1^m |x_j - \bar{x}| f_j}{\sum_1^m f_j} = \frac{|25 - 40| \cdot 5 + |35 - 40| \cdot 11 + |45 - 40| \cdot 8 + |55 - 40| \cdot 6}{30} = 8,7\%.$$

Варіація кредитної ставки зростає з 8,25 до 8,7 відсотків, тобто на 5%. У II кварталі відхилення значень ознаки від центру розподілу більше.

**4.20.** За даними про ступінь використання виробничого і енергетичного устаткування фірмою, визначте з кожного виду устаткування середній коефіцієнт використання та середнє квадратичне відхилення. Порівняйте варіацію, зробіть висновки.

Таблиця 4.19

Виробниче обладнання		Енергетичне обладнання	
коефіцієнт використання, %	кількість одиниць обладнання	коефіцієнт використання, %	кількість одиниць обладнання
До 80	16	До 60	3
80-85	22	60-70	9
85-90	8	70-80	23
90 і вище	4	80 і вище	15
Разом	50	Разом	50

**Розв'язок:**

1)  $\bar{x}$  – середній коефіцієнт використання:

а) виробниче обладнання:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{77,5 \cdot 16 + 82,5 \cdot 22 + 87,5 \cdot 8 + 92,5 \cdot 4}{50} = 82,5\%.$$

б) енергетичне обладнання:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{55 \cdot 3 + 65 \cdot 9 + 75 \cdot 23 + 85 \cdot 15}{50} = 75\%.$$

2)  $\sigma$  – середньоквадратичне відхилення:

а) виробниче обладнання:

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}} = \sqrt{\frac{(77,5 - 82,5)^2 \cdot 16 + (87,5 - 82,5)^2 \cdot 8 + (92,5 - 82,5)^2 \cdot 4}{50}} = 16,36\%.$$



б) енергетичне обладнання:

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}} = \sqrt{\frac{(55-75)^2 \cdot 3 + (65-75)^2 \cdot 9 + (85-75)^2 \cdot 15}{50}} = 27,2\%.$$

3)  $V_\sigma$  – коефіцієнт варіації:

а) виробниче обладнання:

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\% = \frac{16,36}{82,5} \cdot 100\% = 19,8\%.$$

б) енергетичне обладнання:

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\% = \frac{27,2}{75} \cdot 100\% = 36,3\%.$$

Коефіцієнт варіації використання енергетичного обладнання майже в 2 рази більше, ніж аналогічний коефіцієнт щодо виробничого обладнання. Це свідчить про те, що коефіцієнти використання енергетичного обладнання більше відхиленні від їх центру розподілу.

**4.21.** За процентом високоліквідних активів у сумі поточних активів комерційні банки міста розподіляються так:

Таблиця 4.20

% високоліквідних активів	Кількість банків	
	на початок року	на кінець року
До 10	1	–
10-20	5	1
20-30	9	4
30-40	3	9
40 і вище	2	6
Разом	20	20

Визначте середній процент високоліквідних активів та середнє квадратичне відхилення на початок і кінець року. Проведіть порівняльний аналіз середніх та варіації.

**Розв'язок:**

1)  $\bar{x}$  – середній процент високоліквідних активів.

а) на початок року:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{5 \cdot 1 + 15 \cdot 5 + 25 \cdot 9 + 35 \cdot 3 + 45 \cdot 2}{20} = \frac{500}{20} = 25.$$

б) на кінець року:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{15 \cdot 1 + 25 \cdot 4 + 35 \cdot 9 + 45 \cdot 6}{20} = \frac{700}{20} = 35.$$

Середній процент високоліквідних активів на кінець року на 28,5% більше, ніж на початок року.

2)  $\sigma$  – середньоквадратичне відхилення.

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}}.$$

а) на початок року:

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}} = \sqrt{\frac{(5-25)^2 \cdot 1 + (15-25)^2 \cdot 5 + \dots + (45-25)^2 \cdot 2}{20}} = 10.$$

б) на кінець року:

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}} = \sqrt{\frac{(15-35)^2 \cdot 1 + (25-35)^2 \cdot 4 + \dots + (45-35)^2 \cdot 6}{20}} = 8,37.$$

Середньоквадратичне відхилення на кінець року є меншим, ніж на початок року. Це означає, що сукупність високоліквідних активів, що досліджується, на кінець року є більш однорідною.

**4.22.** За даними обстежень домогосподарств рівень та варіація середньодушового споживання основних продуктів харчування становили:

Таблиця 4.21

Продукт харчування	Середній рівень, кг	Дисперсія
М'ясо та м'ясні продукти	65	169
Молоко та молочні продукти	310	961
Риба та рибні продукти	15	36
Цукор	50	25

Порівняйте варіацію споживання названих продуктів харчування, зробіть висновки.

**Розв'язок:**

Таблиця

Продукт харчування	$\bar{x}$	$\sigma^2$	$\sigma$
М'ясо та м'ясні продукти	65	169	13
Молоко та молочні продукти	310	961	31
Риба та рибні продукти	15	36	6
Цукор	50	25	5

$$\sigma = \sqrt{\sigma^2}$$

$$V_{\sigma} = \frac{\sigma}{\bar{x}} 100\%$$

$$\text{М'ясо та м'ясні продукти: } V_{\sigma} = \frac{13}{65} 100\% = 20\%.$$

$$\text{Молоко та молочні продукти: } V_{\sigma} = \frac{31}{310} 100\% = 10\%.$$

$$\text{Риба та рибні продукти: } V_{\sigma} = \frac{6}{15} 100\% = 40\%.$$

$$\text{Цукор: } V_{\sigma} = \frac{5}{50} 100\% = 10\%.$$

Ступінь варіації споживання риби та рибних продуктів найбільший та становить 40 %, найменший ступінь споживання молока та молочних продуктів і цукру – по 10 %.

**4.23.** Статистичні характеристики розподілу сільськогосподарських підприємств за рівнем ефективності виробництва такі:

Таблиця 4.22

Показник ефективності	Середній рівень	Мода	Середнє квадратичне відхилення
Вихід продукції на 100 га сільськогосподарських угідь, тис. гр. од.	256	240	85
Продуктивність праці, гр. од. на 1 люд-год.	112	100	39
Фондовіддача, гр. од.	0,9	1,0	0,4
Рентабельність виробництва, %	9,6	12,4	7,4

Порівняйте варіацію та асиметрію розподілу сільськогосподарських підприємств за названими показниками ефективності виробництва.

**Розв'язок:**

1)  $V_\sigma$  – варіація розподілу сільськогосподарських підприємств:

а) вихід продукції:

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\% = \frac{85}{256} 100\% = 33,2\%.$$

б) продуктивність праці:

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\% = \frac{39}{112} 100\% = 34,8\%.$$

в) фондовіддача:

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\% = \frac{0,4}{0,9} 100\% = 44,4\%.$$

г) рентабельність виробництва:

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\% = \frac{7,4}{9,6} 100\% = 77,08\%.$$

Найбільша варіація розподілу сільськогосподарських підприємств є за ознакою рентабельність виробництва, тобто тут спостерігається найбільше відхилення підприємств від їх центру розподілу. Найменша варіація – за ознакою вихід продукції.

2)  $A_s$  – асиметрія розподілу сільськогосподарських підприємств:

а) вихід продукції:

$$A_s = \frac{\bar{x} - M_o}{\sigma} = \frac{256 - 240}{85} = 0,188.$$

б) продуктивність праці:

$$A_s = \frac{\bar{x} - M_o}{\sigma} = \frac{112 - 100}{39} = 0,3076.$$

в) фондовіддача:

$$A_s = \frac{\bar{x} - M_o}{\sigma} = \frac{0,9 - 1}{0,4} = -0,25.$$

г) рентабельність виробництва:

$$A_s = \frac{\bar{x} - M_o}{\sigma} = \frac{9,6 - 12,4}{7,4} = -0,378.$$

У розподілі підприємств за виходом продукції та продуктивністю праці спостерігається правостороння асиметрія, а за фондовіддачею та рентабельністю виробництва – лівостороння асиметрія.

**4.24.** На кінець року поточні коефіцієнти ліквідності підприємств-позичальників становили:

Таблиця 4.23

Коефіцієнт ліквідності	До 1,5	1,5-1,7	1,7-1,9	1,9-2,1	2,1 і вище	Разом
Частка підприємств, %	5	10	20	45	20	100

Визначте характеристики розподілу: середню, моду, середнє квадратичне відхилення, коефіцієнти варіації та асиметрії. Зробіть висновок про характер розподілу підприємств-позичальників за рівнем ліквідності.

**Розв'язок:**

1)  $\bar{x}$  – середня зважена:

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{1,4 \cdot 5 + 1,6 \cdot 10 + 1,8 \cdot 20 + 2 \cdot 45 + 2,2 \cdot 20}{100} = \frac{193}{100} = 1,93.$$

2)  $M_0$  – мода:

$$M_0 = x_0 + h \frac{(f_{M_0} - f_{M_0-1})}{(f_{M_0} - f_{M_0-1}) + (f_{M_0} - f_{M_0+1})} = 1,9 + 0,2 \frac{(45 - 20)}{(45 - 20) + (45 - 20)} = 2.$$

Модальний інтервал – 1,9 – 2,1,

де  $x_0 = 1,9$  – нижня границя модального інтервалу;

$h = 0,2$  – розмір модального інтервалу;

$f_{M_0} = 45$  – частота модального інтервалу;

$f_{M_0-1} = 20$  – частота передмодального інтервалу;

$f_{M_0+1} = 20$  – частота післямодального інтервалу.

3)  $\sigma$  – середньоквадратичне відхилення:

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}} = \sqrt{\frac{(1,4 - 1,93)^2 \cdot 5 + (1,6 - 1,93)^2 \cdot 10 + \dots + (2,2 - 1,93)^2 \cdot 20}{100}} = 0,212.$$

3)  $V_\sigma$  – коефіцієнт варіації:

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\% = \frac{0,212}{1,93} 100\% = 11\%.$$

4)  $A_s$  – коефіцієнт асиметрії:

$$A_s = \frac{\bar{x} - M_o}{\sigma} = \frac{1,93 - 2}{0,212} = -0,33.$$

Коефіцієнт асиметрії дорівнює  $-0,33$ . Це свідчить про наявність асиметрії вліво, бо  $A_s < 0$ , тобто в бік, де коефіцієнти ліквідності підприємств-позичальників менші. Спостерігається 11 %-ве відхилення від центру розподілу. Найбільш розповсюдженим є коефіцієнт ліквідності, що дорівнює двом.

**4.25.** За даними обстежень домогосподарств житлова площа у розрахунку на одного члена домогосподарства становила, кв. м:

Таблиця 4.24

Житлова площа, кв.м.	До 5	5- 7	7- 9	9- 11	11- 13	13- 15	15 і більше	Разом
Число домогосподарств	12	34	47	50	26	18	13	200

Визначте характеристики розподілу: середню, медіану, середнє квадратичне відхилення, коефіцієнти варіації та асиметрії. Зробіть висновок про характер розподілу домогосподарств за житловими умовами.

**Розв'язок:**

$$\bar{x} = \frac{\sum_1^m x_j f_j}{\sum_1^m f_j} = \frac{4 \cdot 12 + 34 \cdot 6 + 8 \cdot 47 + 10 \cdot 50 + 12 \cdot 26 + 14 \cdot 18 + 16 \cdot 13}{200} = \frac{1900}{200} = 9,5;$$

$$Me = x_0 + h \frac{0,5 \sum_1^m f_i - S f_{Me-1}}{f_{Me}} = 9 + 2 \frac{0,5 \cdot 200 - (12 + 34 + 47)}{50} = 9,28,$$

де  $x_0$  – нижня границя медіанного інтервалу;

$f_{Me-1}$  – частота передмедіанного інтервалу;  
 $f_{Me}$  – частота медіанного інтервалу;  
 $h$  – величина інтервалу;  
 $\bar{x}$  – середня зважена;  
 $M_e$  – середина 100, вона знаходиться ряду  $12 + 34 + 47 + 50 = 143$ ;  
 $\sigma$  – середньоквадратичне відхилення;  
 $V_\sigma$  – коефіцієнт варіації;  
 $A_s$  – коефіцієнт асиметрії.

$$\sigma = \sqrt{\frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j}} = \sqrt{\frac{(4-9,5)^2 \cdot 12 + (6-9,5)^2 \cdot 34 + \dots + (16-9,5)^2 \cdot 13}{200}} = 3,14.$$

$$V_\sigma = \frac{\sigma}{\bar{x}} 100\% = \frac{3,14}{9,5} 100\% = 33\% = 0,33.$$

$$A_s = \frac{\mu_3}{\sigma^2} = \frac{10,05}{3,14^2} = 1,8.$$

$$\mu_3 = \frac{\sum_1^m (x_j - \bar{x})^3 f_j}{\sum_1^m f_j} = \frac{(4,5-9,5)^3 \cdot 12 + (16-9,5)^3 \cdot 34 + \dots + (16-9,5)^3 \cdot 13}{200} = 10,05.$$

$A_s > 0$  – додатній вправо.

Ряд  $V_\sigma = 0,33$ , тобто розподіл близький до одномодального, а те, що  $A_s = 1,8$ , свідчить про наявність асиметрії вправо, бо  $A_s > 0$ , тобто в бік, де забезпеченість житлом більша.

**4.26.** За результатами перевірки цукристість буряків характеризується такими даними:

Таблиця 4.25

Цукристість, %	12-14	14-16	16-18	18 і вище	Разом
Кількість проб	1	4	3	2	10

Визначте середній рівень та дисперсію цукристості буряків, частку цукрових буряків із цукристістю 18 % і вище та дисперсію частки.

**Розв'язок:**

1)  $\bar{x}$  – середній рівень цукристості буряків:

$$\bar{x} = \frac{\sum_1^m x_i f_i}{\sum_1^m f_i} = \frac{13 \cdot 1 + 15 \cdot 4 + 17 \cdot 3 + 19 \cdot 2}{10} = \frac{162}{10} = 16,2;$$

2)  $\sigma^2$  – дисперсія:

$$\sigma^2 = \frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j} = \frac{(13-16,2)^2 \cdot 1 + (15-16,2)^2 \cdot 4 + (17-16,2)^2 \cdot 3 + (19-16,2)^2 \cdot 2}{10} = 3,36.$$

3) частку цукрових буряків з цукристістю 18 % становить  $2/10 = 20\%$ .

4) дисперсія часток:

$$\sigma^2 = d_i(1-d_i) = 0,2 \cdot (1-0,2) = 0,16.$$

**4.27.** За наведеними даними про частку кредитно-інвестиційного портфеля в активах комерційних банків визначте дисперсії частки КІП для кожного банку. Розмістіть банки за розміром дисперсій. За яких умов дисперсія частки має найбільше значення?

Таблиця 4.26

Банк	А	В	С	Д	К	Ф	М	Н	Р
Частка КІП, %	10	20	30	40	50	60	70	80	90
$d_i$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9

**Розв'язок:**

Дисперсія часток:

$$\sigma^2 = d_i(1-d_i)$$

$$A: \sigma^2 = d_i(1-d_i) = 0,1 \cdot (1-0,1) = 0,09.$$

$$B: \sigma^2 = d_i(1-d_i) = 0,2 \cdot (1-0,2) = 0,16.$$

$$C: \sigma^2 = d_i(1-d_i) = 0,3 \cdot (1-0,3) = 0,21.$$

$$D: \sigma^2 = d_i(1-d_i) = 0,4 \cdot (1-0,4) = 0,24.$$

$$K: \sigma^2 = d_i(1-d_i) = 0,5 \cdot (1-0,5) = 0,25.$$

$$F: \sigma^2 = d_i(1-d_i) = 0,6 \cdot (1-0,6) = 0,24.$$

$$M: \sigma^2 = d_i(1-d_i) = 0,7 \cdot (1-0,7) = 0,21.$$

$$N: \sigma^2 = d_i(1-d_i) = 0,8 \cdot (1-0,8) = 0,16.$$



$$P: \sigma^2 = d_i(1-d_i) = 0,9 \cdot (1-0,9) = 0,09.$$

Розміщення банків за розмірами дисперсій:

1. К.
2. D, F.
3. С, М.
4. В, N.
5. А, Р.

Дисперсія частки має найбільше значення, коли  $d_i$  має значення 50 %, тобто частка кредитно-інвестиційного портфелю становить 50 %. Чим ближче до 50 % знаходиться значення  $d_i$ , тим більше буде дисперсія частки.

**4.28.** Розподіл раціоналізаторів та винахідників за кількістю запатентованих винаходів такий: один винахід – 30 осіб; два – 40; три – 20; чотири – 10. Визначте дисперсію кількості запатентованих винаходів двома методами:

- 1) як середній квадрат відхилень;
- 2) як різницю квадратів.

Зробіть висновки.

**Розв'язок:**

*Таблиця*

**Розподіл раціоналізаторів за кількістю винаходів**

Кількість винаходів, $x_j$	1	2	3	4	Разом
Кількість осіб, $f_j$	30	40	20	10	100
$x_j^2$	1	4	9	16	30

1)  $\sigma^2$  – дисперсія:

$$\sigma^2 = \frac{\sum_1^m (x_j - \bar{x})^2 f_j}{\sum_1^m f_j} = \frac{(1-2,1)^2 \cdot 30 + (2-2,1)^2 \cdot 40 + (3-2,1)^2 \cdot 20 + (4-2,1)^2 \cdot 10}{100} = 0,89.$$

$$\bar{x} = \frac{\sum_1^m x_i f_i}{\sum_1^m f_i} = \frac{30 \cdot 1 + 40 \cdot 2 + 20 \cdot 3 + 10 \cdot 4}{100} = \frac{210}{100} = 2,1.$$

$$2) \sigma^2 = \bar{x}^2 - \left(\bar{x}\right)^2 = 5,3 - 2,1^2 = 0,89.$$

$$\bar{x}^2 = \frac{\sum_1^m x_i^2 f_i}{\sum_1^m f_i} = \frac{1 \cdot 30 + 4 \cdot 40 + 9 \cdot 20 + 16 \cdot 10}{100} = \frac{530}{100} = 5,3.$$

**4.29.** За наведеними даними, обчисліть коефіцієнти галузевої локалізації експорту, зробіть висновки.

Таблиця 4.27

Галузь промисловості	% до підсумку	
	загальний обсяг виробництва, $D_j$	обсяг експорту товарів, $d_j$
Харчова сировина і продукти	17,4	20,4
Промисловість хімічної продукції	7,5	16,6
Метали та металопродукція	23,5	32,7
Промислове устаткування	16,0	16,0
Інше	35,6	14,3
Разом	100	100

**Розв'язок:**

Коефіцієнти галузевої локалізації:

$$L_j = \frac{D_j}{d_j} * 100\%.$$

Харчова сировина і продукти:  $L_j = \frac{D_j}{d_j} 100\% = \frac{17,4}{20,4} 100\% = 85,3\%.$

Промисловість хімічної продукції:

$$L_j = \frac{D_j}{d_j} 100\% = \frac{7,5}{16,6} 100\% = 45,2\%.$$

Метали та металопродукція:  $L_j = \frac{D_j}{d_j} 100\% = \frac{23,5}{32,7} 100\% = 71,8\%.$

Промислове устаткування:  $L_j = \frac{D_j}{d_j} 100\% = \frac{16,0}{16,0} 100\% = 100\%.$

$$\text{Інше: } L_j = \frac{D_j}{d_j} 100\% = \frac{35,6}{14,3} 100\% = 248,9\%.$$

Найбільша концентрація обсягу виробництва та експорту товарів у галузі промислового устаткування, де обсяг виробництва і становить обсяг експорту, найменша концентрація – в галузі промисловості хімічної продукції.

**4.30.** За наведеними даними визначте коефіцієнти регіональної локалізації водо- та теплопостачання міст, зробіть висновки.

Визначте коефіцієнти концентрації водо- та теплопостачання; зробіть висновки.

Таблиця 4.28

Регіони	Частка, % до підсумку		
	міського населення, $D_j$	постачання питної води, $d_{j1}$	постачання теплової енергії, $d_{j2}$
А	46,5	45,2	46,6
В	17,2	19,7	14,8
С	23,6	26,8	26,8
Д	12,7	8,3	11,8
Разом	100	100	100

**Розв'язок:**

1) Коефіцієнти галузевої локалізації:

$$L_j = \frac{D_j}{d_j} 100\%.$$

*Постачання питною водою:*

$$\text{А: } L_j = \frac{D_j}{d_j} 100\% = \frac{45,2}{46,5} 100\% = 97,2\%.$$

$$\text{В: } L_j = \frac{D_j}{d_j} 100\% = \frac{19,7}{17,2} 100\% = 114,5\%.$$

$$\text{С: } L_j = \frac{D_j}{d_j} 100\% = \frac{26,8}{23,6} 100\% = 113,6\%.$$

$$\text{Д: } L_j = \frac{D_j}{d_j} 100\% = \frac{8,3}{12,7} 100\% = 65,3\%.$$

*Постачання тепловою енергією:*

$$A: L_j = \frac{D_j}{d_j} 100\% = \frac{46,6}{46,5} 100\% = 100,2\%.$$

$$B: L_j = \frac{D_j}{d_j} 100\% = \frac{14,8}{17,2} 100\% = 86\%.$$

$$C: L_j = \frac{D_j}{d_j} 100\% = \frac{26,8}{23,6} 100\% = 113,6\%.$$

$$D: L_j = \frac{D_j}{d_j} 100\% = \frac{11,8}{12,7} 100\% = 92,9\%.$$

Розподіл регіонів не є рівномірним, у регіоні С спостерігається найбільша концентрація ознаки постачання тепловою енергією, в регіоні В найбільша концентрація ознаки постачання питною водою.

2) Коефіцієнт концентрації:

$$K = \frac{1}{2} \sum_1^m |D_j - d_j|.$$

*Постачання питною водою:*

$$A: K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |46,5 - 45,2| = 0,65.$$

$$B: K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |17,2 - 19,7| = 1,25.$$

$$C: K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |23,6 - 26,8| = 1,6.$$

$$D: K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |12,7 - 8,3| = 2,2.$$

*Постачання тепловою енергією:*

$$A: K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |46,5 - 46,6| = 0,05.$$

$$B: K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |17,2 - 14,8| = 1,2.$$

$$C: K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |23,6 - 26,8| = 1,6.$$

$$D: K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |12,7 - 11,8| = 0,45.$$

Найбільш високий рівень концентрації постачання питною вагою у регіоні D, найменший – у регіоні A. Найбільший рівень концентрації постачання тепловою енергією в регіоні C, найменший – у регіоні A.

**4.31.** За наведеними даними визначте коефіцієнти концентрації виробництва електроенергії та споживання палива ТЕС. Зробіть висновки.

Таблиця 4.29

Потужність електростанцій, МВт	Число електростанцій, $d_j$	% до підсумку	
		вироблено електроенергії, $D_{j1}$	спожито палива, $D_{j2}$
До 50	20	2	1
50-100	13	3	2
100-200	37	20	13
200-400	11	15	12
400-1000	12	33	30
1000-3000	4	12	15
3000 і вище	3	15	27
Разом	100	10	100

**Розв'язок:**

1) Коефіцієнти галузевої локалізації:

$$L_j = \frac{D_j}{d_j} 100\%.$$

Вироблено електроенергії:

потужність до 50 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{2}{20} 100\% = 10\%$ ;

потужність до 50-100 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{3}{13} 100\% = 23\%$ ;

потужність до 100-200 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{20}{37} 100\% = 54\%$ ;

потужність до 200-400 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{15}{11} 100\% = 136\%$ ;

потужність до 400-1000 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{33}{12} 100\% = 275\%$ ;

потужність до 1000-3000 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{12}{4} 100\% = 300\%$ ;

потужність до 3000 МВт і вище:  $L_j = \frac{D_j}{d_j} 100\% = \frac{15}{3} 100\% = 500\%$ .

*Спожито палива:*

потужність до 50 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{1}{20} 100\% = 5\%$ ;

потужність до 50-100 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{2}{13} 100\% = 15,4\%$ ;

потужність до 100-200 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{13}{37} 100\% = 35\%$ ;

потужність до 200-400 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{12}{11} 100\% = 109\%$ ;

потужність до 400-1000 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{30}{12} 100\% = 250\%$ ;

потужність до 1000-3000 МВт:  $L_j = \frac{D_j}{d_j} 100\% = \frac{15}{4} 100\% = 375\%$ ;

потужність 3000 МВт і вище:  $L_j = \frac{D_j}{d_j} 100\% = \frac{27}{3} 100\% = 900\%$ .

Розподіл електростанцій за виробленою електроенергією та спожитим паливом не є рівномірним, його найбільша концентрація спостерігається в електростанцій з потужністю 3000 МВт і вище.

2) Коефіцієнт концентрації:

$$K = \frac{1}{2} \sum_1^m |D_j - d_j|.$$

*Вироблено електроенергії:*

потужність до 50 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |2 - 20| = 9$ ;

потужність до 50-100 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |3 - 13| = 5$ ;

потужність до 100-200 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |20 - 37| = 8,5$ ;

потужність до 200-400 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |15 - 11| = 2$ ;

потужність до 400-1000 МВт  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |33 - 12| = 10,5$ ;

потужність до 1000-3000 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |12 - 4| = 4$ ;

Потужність до 3000 МВт і вище:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |15 - 3| = 6$ .

*Спожито палива:*

потужність до 50 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |1 - 20| = 9,5$ ;

потужність до 50-100 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |2 - 13| = 5,5$ ;

потужність до 100-200 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |13 - 37| = 12$ ;

потужність до 200-400 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |12 - 11| = 0,5$ ;

потужність до 400-1000 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |30 - 12| = 9$ ;

потужність до 1000-3000 МВт:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |15 - 4| = 5,5$ ;

потужність до 3000 МВт і вище:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |27 - 3| = 12$ .

Найбільш високий рівень концентрації вироблення електроенергії припадає на електростанції, потужність яких від 400 до 1000 МВт, найменший – на електростанції з потужністю 200-400 МВт. Найбільший рівень концентрації споживання палива у електростанції з потужністю 100-200 МВт та 3000 МВт і вище, найменший – на електростанції з потужністю 200-400 МВт.

**4.32.** Концентрація фермерської земельної власності характеризується даними:

*Таблиця 4.30*

Земельна площа, га	% до підсумку			
	2005 р.		2010 р.	
	число ферм, $d_{j1}$	земельна площа, $D_{j1}$	число ферм, $d_{j2}$	земельна площа, $D_{j2}$
До 5	29,5	4,8	21,9	1,9
5-10	29,2	8,8	24,3	4,4
10-25	29,5	26,5	32,0	17,4
25-50	7,5	15,0	12,7	14,8
50-100	2,7	10,6	5,4	12,3
100 і більше	1,6	34,3	3,7	49,2
Разом	100	100	100	100

За кожний рік визначте коефіцієнти концентрації, порівняйте їх; зробіть висновки.

**Розв'язок:**

Коефіцієнт концентрації:

$$K = \frac{1}{2} \sum_1^m |D_j - d_j|.$$

1) 2005 рік:

земельна площа до 5 га:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |4,8 - 29,5| = 12,35;$

земельна площа до 5-10 га:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |8,8 - 29,2| = 10,2;$

земельна площа до 10-25 га:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |26,5 - 29,5| = 1,5;$

земельна площа до 25-50 га:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |15 - 7,5| = 3,75;$

земельна площа до 50-100 га:

$$K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |10,6 - 2,7| = 3,95;$$

Земельна площа 100 га і більше:

$$K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |34,3 - 1,6| = 16,35.$$

2) 2010 рік:

земельна площа до 5 га:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |1,9 - 21,9| = 10;$

земельна площа до 5-10 га:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |4,4 - 24,3| = 9,95;$

земельна площа до 10-25 га:  $K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |17,4 - 32| = 7,3;$

земельна площа до 25-50 га:

$$K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |14,8 - 12,7| = 1,05;$$

земельна площа до 50-100 га:

$$K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |12,3 - 5,4| = 3,45;$$

земельна площа 100 га і більше:

$$K = \frac{1}{2} \sum_1^m |D_j - d_j| = \frac{1}{2} \cdot |49,2 - 3,7| = 22,75.$$



Найбільший рівень концентрації фермерської земельної власності спостерігається в 2005 р. на ділянці площею 100 га і більше, найменший – на ділянці площею 10-25 га. В 2010 р. найбільший рівень концентрації фермерської земельної власності був на ділянці площею 100 га і більше, найменший – на ділянці площею 25-50 га.

**4.33.** Розподіл спеціалістів за галузями економіки у різних регіонах характеризується даними, %:

Таблиця 4.31

Галузь економіки	Регіони				
	А	В	С	Д	К
Матеріальне виробництво	25	28	21	13	33
Освіта	31	26	32	38	21
Охорона здоров'я	18	22	21	23	19
Інші сфери	26	24	26	26	27
Разом	100	100	100	100	100

Оцініть ступінь подібності структур зайнятості спеціалістів. Як базу порівняння використайте регіон А.

**Розв'язок:**

Таблиця

**Розподіл часток спеціалістів за галузями по регіонах**

Галузь економіки	Регіони				
	А, $d_j$	В, $d_{k1}$	С, $d_{k2}$	Д, $d_{k3}$	К, $d_{k4}$
Матеріальне виробництво	0,25	0,28	0,21	0,13	0,33
Освіта	0,31	0,26	0,32	0,38	0,21
Охорона здоров'я	0,18	0,22	0,21	0,23	0,19
Інші сфери	0,26	0,24	0,26	0,26	0,27
Разом	1	1	1	1	1

Коефіцієнт подібності структур:  $P = 1 - \frac{1}{2} \sum_1^m |d_j - d_k|$

1) порівняння А і В:

$$P = 1 - 0,5(|0,28 - 0,25| + |0,26 - 0,31| + |0,32 - 0,18| + |0,24 - 0,26|) = 1 - 0,5 \cdot 0,14 = 0,93.$$

2) порівняння А і С:

$$P = 1 - 0,5(|0,21 - 0,25| + |0,32 - 0,31| + |0,21 - 0,18| + |0,26 - 0,26|) = 1 - 0,5 \cdot 0,08 = 0,96.$$

3) порівняння А і D:

$$P = 1 - 0,5(|0,13 - 0,25| + |0,38 - 0,31| + |0,23 - 0,18| + |0,26 - 0,26|) = 1 - 0,5 \cdot 0,24 = 0,88.$$

4) порівняння А і К:

$$P = 1 - 0,5(|0,33 - 0,25| + |0,21 - 0,31| + |0,19 - 0,18| + |0,27 - 0,26|) = 1 - 0,5 \cdot 0,2 = 0,9.$$

Найбільш подібними є структури зайнятості спеціалістів регіонів А і С, найменш – D і А, але в цілому структури всіх регіонів є значно схожими до регіону А.

**4.34.** За даними про віковий склад парку металообробного устаткування для кожного виду устаткування визначте квадратичний коефіцієнт структурних зрушень, проведіть порівняльний аналіз.

Таблиця 4.32

Вікова група, років	Металорізальне		Ковальсько-пресове	
	1989 р.	1996 р.	1989 р.	1996 р.
До 10	60	45	56	51
10-20	26	34	24	30
20 і старші	14	21	20	19
Разом	100	100	100	100

**Розв'язок:**

Квадратичний коефіцієнт структурних зрушень:

$$\sigma_d = \sqrt{\frac{\sum_1^m (d_{j1} - d_{j0})^2}{m}}.$$

1) для металорізального устаткування:

$$\sigma_d = \sqrt{\frac{(45 - 60)^2 + (34 - 26)^2 + (21 - 14)^2}{3}} = \sqrt{\frac{338}{3}} = \sqrt{112,67} = 10,61;$$

2) для ковальсько-пресового устаткування:

$$\sigma_d = \sqrt{\frac{(51 - 56)^2 + (30 - 24)^2 + (19 - 20)^2}{3}} = \sqrt{\frac{62}{3}} = \sqrt{20,67} = 4,55.$$

Зрушення у структурі металорізального устаткування більш, ніж в 2 рази перевищують зміни структури ковальсько-пресового устаткування протягом зазначеного періоду. Тобто оновлення та перехід устаткування з однієї групи в іншу в першому випадку відбувається вдвічі швидше.

**4.35.** Структура валового споживання енергоресурсів характеризується даними:

Таблиця 4.33

Вид енергоресурсів	% до підсумку			
	країна А		країна В	
	1985 р.	1995 р.	1985 р.	1995 р.
Тверде паливо	32,2	34,9	19,8	18,5
Рідке паливо	49,6	43,2	63,5	51,3
Природний газ	16,5	18,7	12,0	17,4
Атомна та гідроенергетика	1,7	3,2	4,7	12,8

Оцініть інтенсивність структурних зрушень у кожній країні, проведіть порівняльний аналіз.

**Розв'язок:**

1) інтенсивність структурних зрушень валового споживання енергоресурсів у країні А:

$$\bar{i}_d = \frac{\sum_{j=1}^m |d_{j1} - d_{j0}|}{m} = \frac{|34,9 - 32,2| + |43,2 - 49,6| + |18,7 - 16,5| + |3,2 - 1,7|}{4} = 3,2.$$

$$\sigma_d = \sqrt{\frac{\sum_{j=1}^m (d_{j1} - d_{j0})^2}{m}} = \sqrt{\frac{(34,9 - 32,2)^2 + (43,2 - 49,6)^2 + (18,7 - 16,5)^2 + (3,2 - 1,7)^2}{4}} = 3,72.$$

2) інтенсивність структурних зрушень валового споживання енергоресурсів у країні В:

$$\bar{i}_d = \frac{\sum_{j=1}^m |d_{j1} - d_{j0}|}{m} = \frac{|18,5 - 19,8| + |51,3 - 63,5| + |17,4 - 12,0| + |12,8 - 4,7|}{4} = 6,75.$$

$$\sigma_d = \sqrt{\frac{\sum_{j=1}^m (d_{j1} - d_{j0})^2}{m}} = \sqrt{\frac{(18,5 - 19,8)^2 + (51,3 - 63,5)^2 + (17,4 - 12,0)^2 + (12,8 - 4,7)^2}{4}} =$$

$$= 7,83.$$

Інтенсивність структурних зрушень в енергетиці у країні В вдвічі вища, ніж у країні А, що свідчить про дещо більші зміни у використанні енергоресурсів.

**4.36.** У таблиці наведені депозитні процентні ставки 10 відділень банків.

Таблиця 4.34

Відділення комерційних банків	Депозитні процентні ставки в окремих відділеннях, %					
Центральні	29	31	33	27	–	–
Філії	24	27	22	28	23	26

Визначте групові, міжгрупову та загальну дисперсії депозитної процентної ставки, покажіть їх зв'язок.

**Розв'язок:**

$$\bar{y}_j = \frac{\sum_1^{f_j} y_j}{f_j}.$$

$$\bar{y}_1 = \frac{29 + 31 + 33 + 27}{4} = 30.$$

$$\bar{y}_2 = \frac{24 + 27 + 22 + 28 + 23 + 26}{6} = 25.$$

$$\bar{y} = \frac{\sum_1^m y_j f_j}{\sum_1^m f_j} = \frac{120 + 150}{10} = \frac{270}{10} = 27.$$

Внутрішньогрупові дисперсії:

$$\sigma_j^2 = \frac{\sum_1^{f_j} (y - \bar{y}_j)^2}{f_j}.$$

$$\sigma_1^2 = \frac{(29 - 30)^2 + (31 - 30)^2 + (33 - 30)^2 + (27 - 30)^2}{4} = 5.$$

$$\sigma_2^2 = \frac{(24 - 25)^2 + (27 - 25)^2 + (22 - 25)^2 + (28 - 25)^2 + (23 - 25)^2 + (26 - 25)^2}{6} = 4,67.$$

Середня з внутрішньогрупових дисперсій:

$$\bar{\sigma}_j^2 = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j} = \frac{5 \cdot 4 + 4,67 \cdot 6}{10} = \frac{48}{10} = 4,8.$$

Міжгрупова дисперсія:

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^m f_j} = \frac{(30-27)^2 \cdot 4 + (25-27)^2 \cdot 6}{10} = \frac{60}{10} = 6.$$

Загальна дисперсія:  $\sigma^2 = \delta^2 + \bar{\sigma}_j^2 = 6 + 4,8 = 10,8.$

**4.39.** Витрати на демонстрацію рекламних роликів на різних телеканалах характеризуються даними:

Таблиця 4.35

Телевізійні канали	Кількість рекламних роликів у денному ефірі	Витрати на демонстрацію рекламного ролика, тис. ум. гр. од.									
Державні	6	1,18	0,97	1,15	1,20	0,94	1,16	-	-	-	-
Комерційні	10	0,29	0,31	0,26	0,34	0,25	0,30	0,33	0,34	0,28	0,30

Визначте групові, міжгрупову та загальну дисперсії витрат на демонстрацію рекламних роликів, а також питому вагу міжгрупової дисперсії у загальній. Результати проаналізуйте.

**Розв'язок:**

$$\bar{y}_j = \frac{\sum_1^{f_j} y_j}{f_j}.$$

$$\bar{y}_1 = \frac{1,18 + 0,97 + 1,15 + 1,20 + 0,94 + 1,16}{6} = 1,1 \text{ тис. ум. гр. од.}$$

$$\bar{y}_2 = \frac{0,29 + 0,31 + 0,26 + 0,34 + 0,25 + 0,30 + 0,33 + 0,34 + 0,28 + 0,30}{10} = 0,3 \text{ тис. ум. гр. од.}$$

$$\bar{y} = \frac{\sum_1^m y_j f_j}{\sum_1^m f_j} = \frac{6,6 + 3}{16} = \frac{9,6}{16} = 0,6 \text{ тис. ум. гр. од.}$$

Внутрішньогрупові дисперсії:

$$\sigma_j^2 = \frac{\sum_1^{f_j} (y - \bar{y}_j)^2}{f_j}.$$

$$\sigma_1^2 = \frac{(1,18-1,1)^2 + (0,97-1,1)^2 + (1,15-1,1)^2 + (1,2-1,1)^2 + (0,94-1,1)^2 + (1,16-1,1)^2}{6} = 0,0108.$$

$$\sigma_2^2 = \frac{(0,29-0,3)^2 + (0,31-0,3)^2 + (0,26-0,3)^2 + \dots + (0,28-0,3)^2 + (0,3-0,3)^2}{10} = 0,00088.$$

Середня з внутрішньогрупових дисперсій:

$$\bar{\sigma}_j^2 = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j} = \frac{0,0108 \cdot 6 + 0,00088 \cdot 10}{16} = 0,0046.$$

Міжгрупова дисперсія:

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 \cdot f_j}{\sum_1^m f_j} = \frac{(1,1-0,6)^2 \cdot 6 + (0,3-0,6)^2 \cdot 10}{16} = 0,15.$$

Загальна дисперсія:  $\sigma^2 = \delta^2 + \bar{\sigma}_j^2 = 0,15 + 0,0046 = 0,1546$ .

Питома вага міжгрупової дисперсії в загальній:  $\eta^2 = \delta^2 / \sigma^2 = 0,15 / 0,1546 = 0,97$ , тобто результативна ознака (витрати на демонстрацію рекламного ролика) на 97 % визначається факторною (вид телевізійного каналу).

**4.40.** Продуктивність праці на підприємствах вугільної промисловості характеризується даними таблиці.

Визначте міжгрупову, середню з групових та загальну дисперсії видобутку вугілля на одного робітника. Поясніть зміст кожної дисперсії, покажіть їх взаємозв'язок.

Таблиця 4.35

Підприємство	Питома вага у загальній кількості робітників, %	Середньомісячний видобуток вугілля на одного робітника, тис. т	Дисперсія видобутку вугілля
Шахти	75	40	3400
Вугільні розрізи	25	200	2600
За сукупністю в цілому	100	80	-

**Розв'язок:**

Внутрішньогрупові дисперсії характеризують варіацію ознаки  $y$  (видобуток вугілля) за рахунок факторів, що не враховані у групуванні:

$$\sigma_1^2 = 3400; \sigma_2^2 = 2600.$$

Середня з внутрішньогрупових дисперсій:

$$\bar{\sigma}_j^2 = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j} = \frac{3400 \cdot 75 + 2600 \cdot 25}{100} = \frac{320000}{100} = 3200.$$

Міжгрупова дисперсія характеризує варіацію ознаки  $y$  (видобуток вугілля) за рахунок фактора  $x$  (вид підприємства), що покладений в основу групування.

$$\bar{y}_1 = 40; \bar{y}_2 = 200; \bar{y} = 80.$$

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^m f_j} = \frac{(40 - 80)^2 \cdot 75 + (200 - 80)^2 \cdot 25}{100} = 4800.$$

Загальна дисперсія характеризує варіацію ознаки  $y$  (видобуток вугілля) за рахунок впливу всіх факторів:

$$\sigma^2 = \delta^2 + \bar{\sigma}_j^2 = 4800 + 3200 = 8000.$$

**4.41.** За даними обстеження витрати часу жінок на домашню роботу такі:

*Таблиця 4.37*

Тип помешкання	Чисельність жінок, тис. осіб	Середні витрати часу на домашню роботу, год.	Групова дисперсія витрат часу
Індивідуальна квартира	50	6,2	0,01
Приватний будинок	40	7,0	0,04

Визначте міжгрупову, середню з групових та загальну дисперсії витрат часу жінок на домашню роботу. Покажіть взаємозв'язок дисперсій та поясніть зміст кожної з них.

**Розв'язок:**

Внутрішньогрупові дисперсії – характеризують варіацію ознаки  $y$  (середні витрати часу на домашню роботу) за рахунок факторів, що не враховані у групуванні.

$$\sigma_1^2 = 0,01;$$

$$\sigma_2^2 = 0,04.$$

Середня з внутрішньогрупових дисперсій:

$$\bar{\sigma}_j^2 = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j} = \frac{0.01 \cdot 50 + 0.04 \cdot 40}{90} = 0,023.$$

Міжгрупова дисперсія характеризує варіацію ознаки  $y$  (середні витрати часу на домашню роботу) за рахунок фактора  $x$  (тип помешкання), що покладений в основу групування.

$$\bar{y}_1 = 6,2;$$

$$\bar{y}_2 = 7,0.$$

$$\bar{y} = \frac{\sum_1^m \bar{y}_j f_j}{\sum_1^m f_j} = \frac{6,2 \cdot 50 + 7,0 \cdot 40}{90} = \frac{590}{90} = 6,56.$$

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^m f_j} = \frac{(6,2 - 6,56)^2 \cdot 50 + (7,0 - 6,56)^2 \cdot 40}{90} = \frac{14,224}{90} = 0,158.$$

Загальна дисперсія характеризує варіацію ознаки  $y$  (середні витрати часу на домашню роботу) за рахунок впливу всіх факторів.

$$\sigma^2 = \delta^2 + \bar{\sigma}_j^2 = 0,158 + 0,023 = 0,181.$$

**4.42.** Споживання природного газу в житловому секторі характеризується даними таблиці.

Визначте міжгрупову та середню з групових дисперсій споживання газу, якщо відомо, що загальна дисперсія дорівнює 4800.

Таблиця 4.37

Категорія житла	Кількість споживачів, тис.	Середньорічне споживання газу в розрахунку на 1 споживача, м <sup>3</sup>
Квартира з газовою плитою	40	250
Квартира з газовою плитою та колонкою	10	400
У цілому	50	X

**Розв'язок:**

$$\sigma^2 = 4800; \bar{y}_1 = 250; \bar{y}_2 = 400.$$



$$\bar{y} = \frac{\sum_1^m \bar{y}_j f_j}{\sum_1^m f_j} = \frac{250 \cdot 40 + 400 \cdot 10}{50} = 280.$$

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^m f_j} = \frac{(250 - 280)^2 \cdot 40 + (400 - 280)^2 \cdot 10}{50} = 3600.$$

Загальна дисперсія:  $\sigma^2 = \delta^2 + \bar{\sigma}_j^2$ .

$$\bar{\sigma}_j^2 = \sigma^2 - \delta^2 = 4800 - 3600 = 1200.$$

**4.43.** У результаті обстеження затримки літаків в аеропорті через певні метеорологічні умови одержано дані:

Таблиця 4.38

Метеорологічні умови	Кількість літаків	Середній час затримки літаків, год.
Несприятливі	25	8
Нестійкі	35	4
Сприятливі	40	1
У цілому	100	3,8

Визначте міжгрупову та середню з групових дисперсій часу затримки літаків, якщо відомо, що загальна дисперсія дорівнює 10.

**Розв'язок:**

Міжгрупова дисперсія:

$$\bar{y}_1 = 8; \bar{y}_2 = 4; \bar{y}_3 = 1; \bar{y} = 3,8$$

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^m f_j} = \frac{(8 - 3,8)^2 \cdot 25 + (4 - 3,8)^2 \cdot 35 + (1 - 3,8)^2 \cdot 40}{100} = 7,56.$$

Загальна дисперсія:  $\sigma^2 = \delta^2 + \bar{\sigma}_j^2$ ,

середня з групових дисперсій:  $\bar{\sigma}_j^2 = \sigma^2 - \delta^2 = 10 - 7,56 = 2,44.$

**4.44.** За даними таблиці визначте загальну, групові та міжгрупові дисперсії якості роботи ткацьких верстатів, покажіть взаємозв'язок дисперсій.

Таблиця 4.39

№ з/п	Технічний стан верстата	Число обривів нитки на 100 м <sup>2</sup> тканини	№ з/п	Технічний стан верстата	Число обривів нитки на 100 м <sup>2</sup> тканини
1	1	73	11	1	71
2	1	76	12	0	66
3	0	68	13	0	65
4	0	64	14	1	76
5	1	70	15	1	72
6	0	67	16	0	69
7	1	79	17	1	71
8	1	75	18	0	72
9	0	73	19	1	70
10	1	70	20	1	73

Розв'язок:

Таблиця

Вплив планового ремонту на середнє число обривів нитки

Плановий ремонт	Кількість верстатів	Середнє число обривів
Пройшов	12	73
Не пройшов	8	68
У цілому	20	71

$$\bar{y}_j = \frac{\sum_1^{f_j} y_j f_j}{\sum_1^{f_j} f_j}$$

$$\bar{y}_1 = \frac{73 \cdot 2 + 76 \cdot 2 + 70 \cdot 3 + 71 \cdot 2 + 72 + 75 + 79}{12} = 73.$$

$$\bar{y}_2 = \frac{68 + 64 + 67 + 73 + 66 + 65 + 69 + 72}{8} = 68.$$

$$\bar{y} = \frac{\sum_1^m \bar{y}_j f_j}{\sum_1^m f_j} = \frac{73 \cdot 12 + 68 \cdot 8}{20} = \frac{1420}{20} = 71.$$

Внутрішньогрупові дисперсії:

$$\sigma_j^2 = \frac{\sum_1^{f_j} (y - \bar{y}_j)^2}{f_j}$$

$$\sigma_1^2 = \frac{(73-73)^2 + (76-73)^2 + (70-73)^2 + (79-73)^2 + \dots + (73-73)^2}{12} = 7,83.$$

$$\sigma_2^2 = \frac{(68-68)^2 + (64-68)^2 + (67-68)^2 + \dots + (69-68)^2 + (72-68)^2}{8} = 9.$$

Середня з внутрішньогрупових дисперсій:

$$\bar{\sigma}_j^2 = \frac{\sum_1^m \sigma_j^2 f_j}{\sum_1^m f_j} = \frac{7,83 \cdot 12 + 9 \cdot 8}{20} = 8,3.$$

Міжгрупова дисперсія:

$$\delta^2 = \frac{\sum_1^m (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^m f_j} = \frac{(73-71)^2 \cdot 12 + (68-71)^2 \cdot 8}{20} = 6.$$

Загальна дисперсія:  $\sigma^2 = \delta^2 + \bar{\sigma}_j^2 = 6 + 8,3 = 14,3.$