Дослідження закономірностей статистичного розподілу

Обчислення асиметрії розподілу:

$$A = \frac{m_3}{\sigma^3}, \; m_3 = \frac{\sum \left(\overline{x}_i - \overline{x}\right)^2 \cdot f}{\sum f}. \tag{1}$$

Обчислення ексцесу статистичного розподілу:

$$E = \frac{m_4}{\sigma^4} - 3, \quad m_4 = \frac{\sum \left(\overline{x}_i - \overline{x}\right)^4 \cdot f}{\sum f}. \tag{2}$$

Основні теоретичні розподіли

Основні теоретичні розподіли:
1)нормальний розподіл;
2)біноміальний розподіл;
3)розподіл Пуассона.

Нормальний розподіл описує залежність між змінною ознакою сукупності і щільністю розподілу у вигляді диференціальної функції Лапласа:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \tag{3}$$

де $t = \dfrac{x - \overline{x}}{\sigma}$ – нормоване відхилення.

Інтегральна функція Лапласа:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt. \qquad (4)$$

Ймовірність P потрапляння випадкової величини (значення ознаки) у проміжок $[x_1; x_2]$:

$$P = F\left(\frac{x_2 - \overline{x}}{\sigma}\right) - F\left(\frac{x_1 - \overline{x}}{\sigma}\right). \qquad (5)$$

Ймовірність настання події A у серії з n незалежних випробувань m разів обчислюється за формулою Бернуллі:

$$P_{n,m} = \frac{n!}{m!(n-m)!} p^m q^{n-m}, \quad m = 0,1,...,n. \qquad (6)$$

$$n! = 1 \cdot 2 \cdot ... \cdot n.$$

Обчислення ймовірності для розподілу Пуассона:

$$P_{n,m}=\frac{\lambda^m\cdot e^{-\lambda}}{m!}. \qquad (7)$$

Приклад 2.13. Підприємство виробляє продукцію з часткою неякісних виробів $p=0,02$. Визначити ймовірність того, що з 1000 відібраних випадковим чином одиниць продукції кількість неякісних виробів становитиме 25 одиниць.

$$n=1000;\ m=25;\ p=0,02;\ q=1-p=0,98.$$

Формула Бернуллі:

$$P_{1000,25}=\frac{1000!}{25!\cdot 975!}\cdot\left(0,02\right)^{25}\cdot\left(0,98\right)^{975}=0,045.$$

Формула Пуассона:

$$\lambda=n\cdot p=1000\cdot 0,02=20.$$

$$P_{1000,25}=\frac{\left(20\right)^{25}e^{-20}}{25!}=0,04459\approx 0,045.$$

Розрахунок теоретичних частот:

$$f_{теор.}=\varphi\left(t\right)\cdot\frac{n\cdot h}{\sigma}, \qquad (8)$$

де $\varphi(t)$ – значення диференціальної функції Лапласа у точці t, n – загальне число спостережень, h – довжина інтервалу зміни ознаки у інтервальному варіаційному ряді, σ – середнє квадратичне відхилення.

$$f_{теор.} = p \cdot n = n \cdot \left[F(t_2) - F(t_1) \right], \tag{9}$$

де n – обсяг вибірки (сума емпіричних частот). Тут

$$t_1 = \frac{x_1 - \overline{x}}{\sigma}, t_2 = \frac{x_2 - \overline{x}}{\sigma},$$

значення x_1 та x_2 – відповідно нижня та верхня межі інтервалу $[x_1; x_2]$ зміни ознаки.

Критерій Пірсона:

$$\chi^2 = \sum \frac{\left(f_{теор.} - f_{емп.} \right)^2}{f_{теор.}}, \tag{10}$$

де $f_{емп.}$ – емпірична частота, $f_{теор.}$ теоретична частота для відповідного інтервалу.

Для нормального розподілу кількість ступенів вільності визначають за формулою:

$$V = m - 3, \tag{11}$$

де V – кількість ступенів вільності, m – кількість інтервалів у варіаційному ряді.

Критерій Романовського:

$$R = \frac{\chi^2 - V}{\sqrt{2V}}. \tag{12}$$

При $R \leq 3$ розбіжність між теоретичними та емпіричними частотами вважається випадковою, а розподіл – близьким до теоретичного. При $R > 3$ розподіл вважається відмінним від теоретичного.

Критерій Колмогорова:

$$\lambda = \frac{D}{\sqrt{n}}, \quad D = \max\left|S_f^{теор.} - S_f^{емп.}\right|,$$

де λ – показник Колмогорова, n – кількість спостережень (сума емпіричних частот, $n \geq 100$), $S_f^{теор.}$ – сума накопичених теоретичних частот, $S_f^{емп.}$ – сума накопичених емпіричних частот, D – максимальна за абсолютним значенням різниця між відповідними накопиченими частотами.

Приклад 1. У таблиці 1 наведено розподіл працівників підприємства за розміром заробітної плати. Оцінити відповідність емпіричного розподілу працівників за рівнем заробітної платні теоретичному нормальному розподілу.

Таблиця 1. Розподіл працівників підприємства за розміром заробітної платні

Заробітна плата x, у.г.о.	Серед-ня в інтер-валі заро-бітна плата $\overline{x}_i$, у.г.о.	Кількість працівників f, чол.	Частка Працівників, %	Нако-пичена частота, чол.	Нако-пичена частка, %	$\left(\overline{x}_i - \overline{x}\right)^2 \cdot f$
До 120	104,5	8	12,3	8	12,3	40943,77
120-150	134,5	10	15,4	18	27,7	17255,72
150-180	164,5	15	23,1	33	50,8	1997,57
180-210	194,5	18	27,7	51	78,5	6133,89
210-240	224,5	9	13,8	60	92,3	21135,34
240 і вище	254,5	5	7,7	65	100	30779,86
Разом	–	65	100	–		118246,15

$$\overline{x} = \frac{\sum\left(\overline{x}_i \cdot f\right)}{\sum f} = 176,5 \text{ у.г.о.},$$

$$\sigma = \sqrt{\frac{\sum\left(\overline{x}_i - \overline{x}\right)^2 \cdot f}{\sum f}} = 42,65 \text{у.г.о.}$$

$$n = 65 > 50, \quad f_{\max} \geq 5.$$

Результати розрахунків заносимо у таблицю 2. У цьому прикладі отримуємо:

$$n = 65; h = 29; \sigma = 42,65; \overline{x} = 176,04; V = 6 - 3 = 3.$$

Таблиця 2.11. Розрахунок критерію Пірсона

x	$f_{емп.}$	$t = \frac{x - \overline{x}}{\sigma}$	$\varphi(t)$	$f_{теор.} = \varphi(t) \cdot \frac{n \cdot h}{\sigma}$	$\frac{\left(f_{теор.} - f_{емп.}\right)^2}{f_{теор.}}$
104,5	8	-1,68	0,09728	4,3	3,184
134,5	10	-0,97	0,24923	11	0.091
164,5	15	-0,27	0,38466	17	0,235
194,5	18	0,43	0,36371	16,1	0,224
224,5	9	1,14	0,20831	9,2	0,004
254,5	5	1,84	0,07341	3,2	1,013
Разом	65	–	–	60,8	$\chi^2 = 4,751$

$$\chi^2_{теор.} = 7,81.$$