Описова статистика

Формули

Ряд розподілу:

x_i	$\bullet\bullet\bullet$	$\bullet\bullet\bullet$	$\bullet\bullet\bullet$	$\bullet\bullet\bullet$	
n_i	$\bullet\bullet\bullet$	$\bullet\bullet\bullet$	$\bullet\bullet\bullet$	$\bullet\bullet\bullet$	

x_i - значення ознак

n_i - числа значення ознак

$N = \sum n_i$ - число цілого розподілу

Показник інтенсивності при побудові ряду розподільного ряду з нерівними інтервалами:

$$\begin{matrix}\textbf{показник}\\ \textbf{інтенсивності}\end{matrix} = \frac{\textbf{розмір класу} \cdot \textbf{діапазон найвужчого (найширшего) класу}}{\textbf{діапазон класу}}$$

Кількість класів у серії розподілу: $k \le 5\log N$

Діапазон класів у розподільному ряду: $i = \frac{x_{\max} - x_{\min}}{k}$

I. Середні міри

Середня арифметична: $\frac{\sum x_i n_i}{N}$

Середня гармонічна: $\frac{N}{\sum \frac{1}{x_1} n_i}$

Середня геометрична: $\sqrt[N]{\prod x_i^{n_i}}$

Домінанта: $D =$ Значення ознаки x_i, яка має найбільше значення n_i.

У випадку інтервального ряду: $D = x_D + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} i_D$, де:

x_D - нижня межа інтервалу, в якому розташована домінанта

n_D - розмір інтервалу, в якому розташована домінанта

n_{D-1} - розмір інтервалу, що передує інтервалу з домінантою

n_{D+1} - розмір інтервалу, що слідує з домінантою

i_D - проміжок інтервалу, в якому розташована домінанта

Медіана: Медіана ділить інтервал на дві рівні половини.

Коли N непарне: $Me = \frac{x_{N+1}}{2}$ (значення ознаки, яка є „посередині” впорядковує ознаки у порядку зростання)

Коли N парне: $Me = \frac{1}{2}\left(x_{\frac{N}{2}} + x_{\frac{N}{2}+1} \right)$ (середня значення ознаки, які є „посередині”)

Квартилі $Q_1, Q_2 = Me, Q_3$: Вони діляться інтервалом так само, як медіана, але в пропорціях належним чином 25% : 75%; 50% : 50%; 75% : 25%.

У випадку інтервального ряду:

$$Q_1 = x_{Q_1} + \frac{\frac{N}{4} - \sum_{i=1}^{k-1} n_i}{n_{Q_1}} i_{Q_1}$$

$$Q_2 = Me = x_{Me} + \frac{\frac{N}{2} - \sum_{i=1}^{k-1} n_i}{n_{Me}} i_{Me}$$

$$Q_3 = x_{Q_3} + \frac{\frac{3N}{4} - \sum_{i=1}^{k-1} n_i}{n_{Q_3}} i_{Q_3}$$

x_{Q_1}, x_{Me}, x_{Q_3} - нижні межі діапазонів, в яких розташований квартиль

$\sum_{i=1}^{k-1} n_i$ - сума кількості ПОПЕРЕДНІХ інтервалів, у яких розташований квартиль

n_{Q_1}, n_{Me}, n_{Q_3} - кількість інтервалів, у яких розташований квартиль

i_{Q_1}, i_{Me}, i_{Q_3} - поширення діапазонів, в яких розташований квартиль

Квартилі: Вони ділять групу так само, як медіана і квартилі, але будь-яким чином пропорції. Наприклад, квантиль порядку 0,4, тобто $x_{0,4}$ розділяє інтервал в пропорції: 40%: 60%.

I. Міри мінливості (дисперсії)

Емпірична область конвергенції: $R = x_{\max} - x_{\min}$

Середнє відхилення: $d = \frac{\sum |x_i - \bar{X}| n_i}{N}$

Інтегральне відхилення: $Q = \frac{Q_3 - Q_1}{2}$

Типова область варіації з використанням позиційних показників: $Me - Q < x_{typ} < Me + Q$

Варіація: $S^2 = \frac{\sum (x_i - \bar{X})^2 n_i}{N}$

Середнє відхилення: $S = \sqrt{S^2}$

Типова область варіації: $\bar{X} - S < x_{typ} < \bar{X} + S$

Коефіцієнт варіації: $V_S = \frac{S}{\bar{X}} \cdot 100\%, \quad V_Q = \frac{Q}{Me} \cdot 100\%$

I. Міри асиметрії

$D > Me > \overline{X}$ - асиметрія лівобочна

$D < Me < \overline{X}$ - асиметрія правобочна

Індекс асиметрії: $W_S = \overline{X} - D$, або $W_S = (Q_3 - Q_2) - (Q_2 - Q_1)$

$W_S > 0$ - асиметрія правобочна

$W_S = 0$ - асиметрії немає, розклад симетричний

$W_S < 0$ - асиметрія лівобочна

Коефіцієнт асиметрії (кососі): $A_S = \frac{W_S}{S}$, або $A_S = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{2Q}$

$-1 \leq A_S \leq 1$

Крім напрямку асиметрії A_S, це вказує на силу асиметрії. Чим він ближче 0, тим слабше асиметрія. Чим ближче ± 1 асиметрія сильніша.

Центральний момент третього порядку: $m_3 = \frac{\sum \left(x_i - \overline{X}\right)^3 \cdot n_i}{N}$

Тлумачення негативу/позитиву те саме, що W_S.

Стандартизований момент третього порядку: $As = \frac{m_3}{S^3}, \quad -1 \leq As \leq 1$

Тлумачення сили асиметрії те ж саме, що в A_S.

II. Міри концентрації

Концентрація в сенсі поділу фонду характеристик:

Коефіцієнт концентрації Лоренца: $k = \frac{a}{5000}$, де a це поле поміж лінією рівномірного поділ і кривї концентрації Лоренца.

$k = 0$ означає повну відсутність концентрації (кожна одиниця має однакову частину фонду характеристик)

$k = 1$ означає повну концентрацію (одна одиниця має цілий фонд функцій)

Концентрація у значенні концентрації навколо середнього:

Центральний момент 4-го порядку: $m_4 = \frac{\sum\left(x_i - \overline{X}\right)^4 \cdot n_i}{N}$

Стандартизований момент 4-го порядку: $a_4 = \frac{m_4}{S^4}$

Орієнтиром є нормальний розподіл.

$a_4 > 3$ - значення ознак більш зосереджені навколо середнє, ніж у нормальному розподілі (тонкий графік)

$a_4 = 3$ - значення ознак так само зосереджені навколо середнє, ніж у нормальному розподілі (ідентичний графік)

$a_4 < 3$ - значення ознак менш зосереджені навколо середнє, ніж у нормальному розподілі (сплощений графік)