Karta wzorów: statystyka

Podręcznik Ideis, 1 rok informatyki stosowanej. Karta od 6 października 2026, ostatnia aktualizacja 24 października 2026, 11:20. Najnowsza wersja: plan-ideis.vercel.app/wzory

3 Statystyka opisowa

Przepisane z pliku WZORY w Moodle (Statystyka opisowa), z tymi samymi oznaczeniami.

Oznaczenia

Dane i szeregi

n	liczebność: liczba wszystkich obserwacji
$x_1, \ldots, x_n$	szereg szczegółowy: kolejne obserwacje (do mediany i kwartyli ustawione rosnąco)
$x_i,\ f_i$	szereg punktowy: wariant i jego liczebność, $n = \sum f_i$
k	liczba wariantów albo przedziałów
x'_i	szereg przedziałowy: środek i-tego przedziału
h	rozpiętość przedziału
$x_{\min},\ x_{\max}$	najmniejsza i największa obserwacja

Przedział, w którym leży miara

$x_s,\ f_s$	dolna granica i liczebność przedziału mediany
$x_m,\ f_m$	to samo dla przedziału dominanty (największa liczebność)
$f_{m-1},\ f_{m+1}$	liczebności przedziałów przed i za przedziałem dominanty
$x_r,\ f_r$	przedział kwartyla pierwszego
$x_p,\ f_p$	przedział kwartyla trzeciego
$\sum_{i=1}^{s-1} f_i$	liczebność skumulowana: suma liczebności wszystkich przedziałów przed przedziałem mediany

Miary

$\bar{x}$	średnia arytmetyczna
M_e	mediana (wartość środkowa), to samo co Q_2
M_o	dominanta (moda): wartość najczęstsza
$Q_1,\ Q_3$	kwartyl pierwszy i trzeci
R	rozstęp
$s^2,\ s$	wariancja, odchylenie standardowe
d	odchylenie przeciętne
Q	odchylenie ćwiartkowe
V_s, V_d, V_Q	współczynniki zmienności
A_s	współczynnik asymetrii

Korelacja i regresja

$x,\ y$	dwie badane cechy; $\bar{y}$ średnia y
r_{xy}	współczynnik korelacji Pearsona
$r_s,\ d_i$	korelacja rang Spearmana, różnica rang
$c_{ix},\ c_{iy}$	rangi: miejsca i-tej obserwacji po ustawieniu wg x i wg y
$\hat{y}$	wartość y odczytana z prostej regresji (teoretyczna)
$a_1,\ a_0$	współczynnik kierunkowy i wyraz wolny prostej
$s_x,\ s_y$	odchylenia standardowe x i y
$y_i - \hat{y}_i$	reszta: różnica między wartością prawdziwą a z prostej
$\varphi^2,\ R^2$	współczynnik zbieżności, współczynnik determinacji
s_e	błąd standardowy reszt

Dynamika i trend

y_t	poziom zjawiska w okresie t; $t = 1, \ldots, n$
$\Delta_{t/1},\ \Delta_{t/t-1}$	przyrost absolutny do okresu bazowego, do poprzedniego
$d_{t/1},\ d_{t/t-1}$	przyrosty względne
$i_{t/1},\ i_{t/t-1}$	indeksy o stałej podstawie, łańcuchowe
$\bar{i}$	średnie tempo zmian
$p,\ q,\ w$	cena, ilość, wartość
$p_0,\ p_1$	indeks 0 to okres bazowy, 1 badany
$I^L,\ I^P$	indeks agregatowy Laspeyresa, Paaschego
$\bar{y}_t^{(3)}$	średnia ruchoma (tu 3-okresowa)
$S_t,\ \alpha$	wartość wygładzona, stała wygładzania
$\hat{y}_t$	wartość trendu w okresie t

3.1 Średnia, mediana, dominanta, kwartyle

Średnia arytmetyczna

szczegółowy

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

punktowy

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i$$

przedziałowy

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x'_i$$

> suma wartości przez liczebność; w szeregach rozdzielczych każdy wariant (albo środek przedziału) mnożysz przez jego liczebność f_i

Mediana: szereg szczegółowy

$$M_e = \begin{cases} x_{\frac{n+1}{2}} & n \text{ nieparzyste} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & n \text{ parzyste} \end{cases}$$

> dane ustawione rosnąco

Mediana: szereg przedziałowy

$$M_e = x_s + \frac{h}{f_s}\left(\frac{n}{2} - \sum_{i=1}^{s-1} f_i\right)$$

> połowa obserwacji leży poniżej. Najpierw znajdź przedział, w którym liczebność skumulowana przekracza $\frac{n}{2}$

Dominanta: szereg przedziałowy

$$M_o = x_m + \frac{(f_m - f_{m-1}) \cdot h}{(f_m - f_{m-1}) + (f_m - f_{m+1})}$$

> najczęstsza wartość: liczysz w przedziale z największą liczebnością f_m, przy równych rozpiętościach przedziałów

Kwartyle

szczegółowy

$$Q_1 = x_{\frac{n+1}{4}} \qquad Q_3 = x_{\frac{3(n+1)}{4}}$$

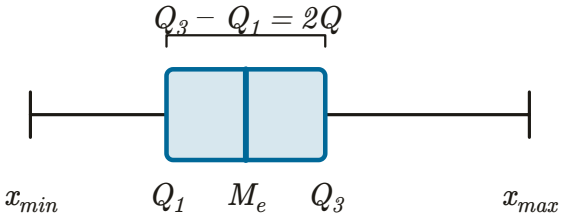
przedziałowy, kwartyl 1

$$Q_1 = x_r + \frac{h}{f_r}\left(\frac{n}{4} - \sum_{i=1}^{r-1} f_i\right)$$

przedziałowy, kwartyl 3

$$Q_3 = x_p + \frac{h}{f_p}\left(\frac{3}{4}n - \sum_{i=1}^{p-1} f_i\right)$$

poniżej Q_1 leży 25% obserwacji, poniżej Q_3 75%; mediana to kwartyl drugi. W przedziałowym liczysz jak medianę, tylko z $\frac{n}{4}$ i $\frac{3}{4}n$



wykres pudełkowy: w pudełku leży środkowe 50% danych, kreska w środku to mediana

3.2 Wariancja, odchylenia, zmienność

Rozstęp

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

największa obserwacja minus najmniejsza; mocno reaguje na wartości odstające

Odchylenie standardowe

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

o ile średnio wartości odchylają się od średniej, w tych samych jednostkach co dane

Wariancja (dzielisz przez n)

szczegółowy

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

punktowy

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i(x_i - \bar{x})^2$$

przedziałowy

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i(x_i' - \bar{x})^2$$

licz po kolei: odchylenia od średniej, ich kwadraty (razy f_i), suma, dzielenie przez n

Wariancja z próby (dzielisz przez n - 1)

szczegółowy

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

punktowy

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i$$

przedziałowy

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i' - \bar{x})^2 f_i$$

wersja dla próby (nieobciążona); przy dużym n wynik prawie taki sam jak przy dzieleniu przez n

Wariancja z poprawką Shepparda

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i' - \bar{x})^2 f_i - \frac{h^2}{12}$$

szereg przedziałowy

Odchylenie przeciętne

szczegółowy

$$d = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

punktowy

$$d = \frac{1}{n}\sum_{i=1}^{k}|x_i - \bar{x}| \cdot f_i$$

przedziałowy

$$d = \frac{1}{n}\sum_{i=1}^{k}|x_i' - \bar{x}| \cdot f_i$$

w ostatnim wierszu pliku z Moodle w punktowym i przedziałowym nad sumą stoi n, poprawnie k (liczba wariantów albo przedziałów)

Odchylenie ćwiartkowe

$$Q = \frac{Q_3 - Q_1}{2}$$

połowa odległości między kwartylami; opisuje środkowe 50% danych, więc wartości skrajne na nie nie wpływają

Współczynniki zmienności

z odchylenia standardowego

$$V_s = \frac{s}{\bar{x}} \cdot 100\%$$

z odchylenia przeciętnego

$$V_d = \frac{d}{\bar{x}} \cdot 100\%$$

pozycyjny

$$V_Q = \frac{Q}{M_e} \cdot 100\%$$

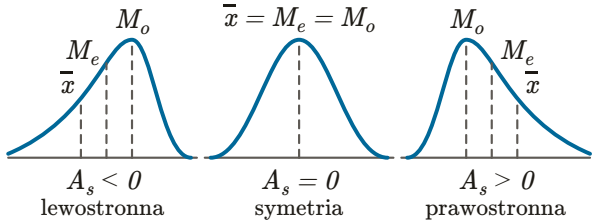
zmienność w procentach średniej (albo mediany); pozwala porównać cechy w różnych jednostkach, im większy, tym bardziej zróżnicowane dane

3.3 Asymetria

Współczynnik asymetrii

$$A_s = \frac{\bar{x} - M_o}{s}$$

$A_s > 0$: asymetria prawostronna, $A_s < 0$: lewostronna



ogon rozkładu ciągnie średnią najdalej: przy prawostronnej $M_o < M_e < \bar{x}$

Rozkład symetryczny

$$\bar{x} = M_e = M_o$$

w rozkładzie symetrycznym z jedną dominantą

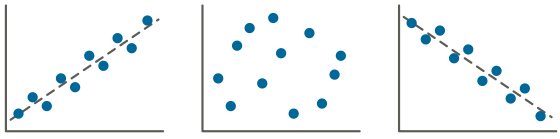
3.4 Korelacja i regresja

Wzory dotyczą prostej regresji $\hat{y} = a_1 x + a_0$. Dla regresji $\hat{x} = b_1 y + b_0$ zamieniasz we wzorach x z y.

Współczynnik korelacji Pearsona

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

od -1 do 1; im bliżej ± 1, tym silniejsza zależność liniowa



r_{xy} mierzy, jak blisko prostej leżą punkty; przerywana to prosta regresji

Prosta regresji

$$a_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad a_0 = \bar{y} - a_1\bar{x}$$

metoda najmniejszych kwadratów; a_1 mówi, o ile średnio zmienia się y, gdy x rośnie o 1

Współczynnik kierunkowy z korelacji

$$a_1 = r_{xy}\frac{s_y}{s_x}$$

s_x, s_y odchylenia standardowe; a_1 ma zawsze ten sam znak co r_{xy}

Korelacja rang Spearmana

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \quad d_i = c_{ix} - c_{iy}$$

c_{ix}, c_{iy} to rangi, czyli numery miejsc obserwacji po ustawieniu

Dopasowanie modelu

$$\varphi^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad R^2 = 1 - \varphi^2$$

R^2 mówi, jaką część zmienności y wyjaśnia prosta

Błąd standardowy reszt

$$s_e^2 = \frac{1}{n-k}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad s_e = \sqrt{s_e^2}$$

k liczba szacowanych parametrów (dla prostej $k = 2$)

3.5 Dynamika zjawisk

Przyrosty absolutne

$$\Delta_{t/1} = y_t - y_1 \qquad \Delta_{t/t-1} = y_t - y_{t-1}$$

pierwszy o stałej podstawie, drugi łańcuchowy (do poprzedniego okresu)

Przyrosty względne

$$d_{t/1} = \frac{\Delta_{t/1}}{y_1} \qquad d_{t/t-1} = \frac{\Delta_{t/t-1}}{y_{t-1}}$$

przyrost absolutny podzielony przez poziom z okresu bazowego; razy 100% daje zmianę w procentach

Indeksy indywidualne

$$i_{t/1} = \frac{y_t}{y_1} \qquad i_{t/t-1} = \frac{y_t}{y_{t-1}}$$

pierwszy dla $t = 1, \ldots, n$, łańcuchowy dla $t = 2, \ldots, n$

Średnie tempo zmian

$$\bar{i} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

średnia geometryczna indeksów łańcuchowych; $(\bar{i} - 1) \cdot 100\%$ to średnia zmiana na okres

Indeksy cen, ilości i wartości

$$I_p = \frac{p_1}{p_0} \qquad I_q = \frac{q_1}{q_0} \qquad I_w = \frac{w_1}{w_0} = \frac{q_1 p_1}{q_0 p_0}$$

p cena, q ilość, w wartość; 0 okres bazowy, 1 badany

Indeksy agregatowe

cen, Laspeyresa

$$I_p^L = \frac{\sum q_0 p_1}{\sum q_0 p_0}$$

cen, Paaschego

$$I_p^P = \frac{\sum q_1 p_1}{\sum q_1 p_0}$$

ilości, Laspeyresa

$$I_q^L = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

ilości, Paaschego

$$I_q^P = \frac{\sum q_1 p_1}{\sum q_0 p_1}$$

sumy po wszystkich wyrobach

3.6 Trend i wygładzanie

Średnia ruchoma 3-okresowa

$$\bar{y}_t^{(3)} = \frac{y_{t-1} + y_t + y_{t+1}}{3}$$

np. $\bar{y}_2^{(3)} = \frac{y_1+y_2+y_3}{3}$; pierwszy i ostatni okres zostają bez średniej

Średnia ruchoma 5-okresowa

$$\bar{y}_t^{(5)} = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}$$

np. $\bar{y}_3^{(5)}$ to średnia z $y_1, \ldots, y_5$

Trend liniowy

$$\hat{y}_t = a_1 t + a_0 \qquad a_1 = \frac{\sum_{t=1}^{n}(t - \bar{t})(y_t - \bar{y})}{\sum_{t=1}^{n}(t - \bar{t})^2}$$

$$a_0 = \bar{y} - a_1\bar{t}$$

t numer okresu $(1, 2, \ldots, n)$, więc $\bar{t} = \frac{n+1}{2}$; to prosta regresji, w której zmienną objaśniającą jest czas

Wygładzanie wykładnicze

$$S_1 = y_1 \qquad S_t = \alpha y_t + (1-\alpha)S_{t-1} \qquad \alpha \in \langle 0{,}1;\ 0{,}4\rangle$$

dla $t = 2, \ldots, n$. W pliku z Moodle przy y_t stoi a, chodzi o α

Własne wzory i notatki