Miary położenia

Średnia

Dla danych indywidualnych:

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

Dla szeregu rozdzielczego

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i \dot{x}_i,$$

gdzie $\dot{x}_i$ – środek $i-$ tego przedziału, n_i - liczność i-tego przedziału.

Dominanta (moda)

Liczba najczęściej występująca (jeśli taka istnieje) - dla danych idywidualnych.

Dla szeregu rozdzielczego

$$D = x_{ld} + h_d \frac{n_d - n_{d-1}}{2n_d - n_{d-1} - n_{d+1}},$$

gdzie d numer najliczniejszego przedziału, x_{ld} lewy koniec d-tego przedziału, h_d długość d-tego przedziału.

Mediana i kwartyle

Dla danych indywidualnych

$x_{(n+1)/2}$ dla n nieparzystego, $\frac{x_{n/2}+x_{n/2+1}}{2}$ dla n parzystego.

Pierwszy kwartyl Q_1: jeśli $\frac{n}{4}$ jest liczbą całkowitą to $Q_1 = x_{n/4}$, jeśli $k < \frac{n}{4} < k+1$, to $Q_1 = \frac{x_k + x_{k+1}}{2}$.

Trzeci kwartyl Q_3: jeśli $\frac{3n}{4}$ jest liczbą całkowitą to $Q_1 = x_{3n/4}$, jeśli $k < \frac{3n}{4} < k+1$, to $Q_1 = \frac{x_k + x_{k+1}}{2}$.

Są też w użyciu inne wzory na kwartyle dla szeregu indywidualnego.

Kwartyle dla szeregów indywidualnych stosuje się rzadko!

Dla szeregu rozdzielczego

$$Q_i = x_{lm_i} + \left(\frac{ni}{4} - \sum_{j=1}^{m_i-1} n_j\right)\frac{h_{m_i}}{n_{m_i}},$$

gdzie m_i – numer grupy zawierającej daną o numerze $\frac{ni}{4}$; $\sum_{j=1}^{m_i-1}$ – suma liczebności od pierwszego przedziału klasowego do przedziału poprzedzającego przedział kwartyla; n_{m_i} – liczebność przedziału w którym znajduje się kwartyl.

Dla $i = 1$ otrzymujemy kwartyl Q_1, dla $i = 2$ kwartyla Q_2, czyli medianę, dla $i = 3$ kwartyl Q_3.

Miary rozproszenia

Wariancja.

Dla danych indywidualnych:

$$s^2 = \frac{1}{n}\left(\sum_{i=1}^{n}(x_i - \overline{x})^2\right).$$

Dla szeregu rozdzielczego

$$s^2 = \frac{1}{n}\left(\sum_{i=1}^{k} n_i(\dot{x}_i - \overline{x})^2\right),$$

gdzie $\dot{x}_i$- środek i-tego przedziału.

Równie często w literaturze wariancję definiuje się dzieląc przez $n-1$ zamiast przez n.

Odchylenie standardowe: $s = \sqrt{s^2}$.

Odchylenie przeciętne od średniej:

Dla danych indywidualnych:

$$d = \frac{1}{n}\left(\sum_{i=1}^{n}|x_i - \overline{x}|\right).$$

Dla szeregu rozdzielczego

$$d = \frac{1}{n}\left(\sum_{i=1}^{k} n_i|\dot{x}_i - \overline{x}|\right),$$

gdzie $\dot{x}_i$- środek i-tego przedziału.

Typowy klasyczny obszar zmienności

$$\overline{x} - s \leqslant x \leqslant \overline{x} + s.$$

Rozstęp

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

Odchylenie ćwiartkowe

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

Typowy pozycyjny obszar zmienności

$$M - Q \leqslant x \leqslant M + Q.$$

Współczynnik zmienności klasyczny

$$V_x = \frac{s}{\overline{x}}.$$

Współczynnik zmienności pozycyjny

$$V_x = \frac{Q}{M}.$$

Moment centralny rzędu l

Dla danych indywidualnych

$$M_l = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^l$$

Dla szeregu rozdzielczego

$$M_l = \sum_{i=1}^{k}\frac{n_i}{n}(\dot{x}_i - \overline{x})^l$$

Klasyczny współczynnik asymetrii

$$A_x = \frac{M_3}{s^3}.$$

Współczynnik skośności (asymetrii)

$$A_s = \frac{\overline{x} - D}{s}.$$

Pozycyjny współczynnik asymetrii

$$A_p = \frac{Q_3 + Q_1 - 2M}{2Q}.$$

Kurtoza

$$K = \frac{M_4}{s^4}$$

DYNAMIKA ZJAWISK

Indeksy jednopodstawowe o podstawie y_0

$$J_i^P = \frac{y_i}{y_0}.$$

Indeksy łańcuchowe

$$J_i^L = \frac{y_i}{y_{i-1}}.$$

Agregatowy indeks wartości

$$I_w = \frac{\sum_{i=1}^{k} q_{i1}p_{i1}}{\sum_{i=1}^{k} q_{i0}p_{i0}}.$$

Agregatowy indeks ilości Laspeyresa

$$I_q^L = \frac{\sum_{i=1}^{k} q_{i1}p_{i0}}{\sum_{i=1}^{k} q_{i0}p_{i0}}.$$

Agregatowy indeks cen Laspeyresa

$$I_p^L = \frac{\sum_{i=1}^{k} q_{i0}p_{i1}}{\sum_{i=1}^{k} q_{i0}p_{i0}}.$$

Agregatowy indeks ilości Paaschego

$$I_q^P = \frac{\sum_{i=1}^{k} q_{i1}p_{i1}}{\sum_{i=1}^{k} q_{i0}p_{i1}}.$$

Agregatowy indeks cen Paaschego

$$I_p^P = \frac{\sum_{i=1}^{k} q_{i1}p_{i1}}{\sum_{i=1}^{k} q_{i1}p_{i0}}.$$

Agregatowy indeks ilości Fishera

$$I_q^F = \sqrt{I_q^L \cdot I_q^P}.$$

Agregatowy indeks cen Fishera

$$I_p^F = \sqrt{I_p^L \cdot I_p^P}.$$

p_{i0} – cena i-tego produktu w okresie podstawowym
p_{i1} – cena i-tego produktu w okresie badanym
q_{i0} – produkcja i-tego produktu w okresie podstawowym
q_{i1} – produkcja i-tego produktu w okresie badanym

Analiza szeregów czasowych

Metoda mechaniczna

Dany jest ciąg (y_t), $t = 1, 2, \ldots, n$ (t czas dyskretny)

Średnie ruchome - nieparzysta liczba p podokresów

$$\widehat{y}_k = \frac{1}{p}\left(y_{k-\frac{p-1}{2}} + \ldots + y_{k+\frac{p-1}{2}}\right),\ k = \frac{p+1}{2}, \ldots, n - \frac{p-1}{2}$$

- parzysta liczba podokresów

$$\widehat{y}_k = \frac{1}{p}\left(\frac{1}{2}y_{k-\frac{p}{2}} + y_{k-\frac{p}{2}+1} + \ldots + y_{k+\frac{p}{2}-1} + \frac{1}{2}y_{k+\frac{p}{2}}\right),$$

$k = \frac{p}{2} + 1, \ldots, n - \frac{p}{2}$.

Metoda analityczna – liniowa funkcja trendu wyznaczona metodą najmniejszych kwadratów

$$\widehat{y}_t = at + b,$$

gdzie

$$a = \frac{\sum_{t=1}^{n}(t-\overline{t})y_t}{\sum_{t=1}^{n}(t-\overline{t})^2},$$

$$b = \overline{y} - a\overline{t},$$

$$\overline{t} = \frac{1}{n}\sum_{t=1}^{n} t = \frac{n+1}{2},\ \overline{y} = \frac{1}{n}\sum_{t=1}^{n} y_t.$$

Odchylenie standardowe składnika losowego

$$s_u = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \widehat{y}_t)^2}{n-2}}.$$

współczynnik zmienności losowej

$$v_u = \frac{s_u}{\overline{y}} \cdot 100\%.$$

współczynnik zgodności

$$\phi^2 = \frac{\sum_{t=1}^{n}(y_t - \widehat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \overline{y})^2}.$$

współczynnik determinacji

$$\mathbf{R}^2 = 1 - \phi^2.$$

Model uważa się za dopuszczalny, jeśli $\phi^2 < 0{,}2$.

błędy średnie szacunku

$$D(a) = \frac{s_u}{\sqrt{\sum_{t=1}^{n}(t-\overline{t})^2}},$$

$$D(b) = \frac{s_u\sqrt{\sum_{t=1}^{n} t^2}}{\sqrt{n\sum_{t=1}^{n}(t-\overline{t})^2}},$$

Oceny parametrów uznajemy za precyzyjne, jeśli

$$\frac{|a|}{D(a)} > 2, \quad \frac{|b|}{D(b)} > 2.$$

Wahania sezonowe

Dany jest ciąg y_t, $t = 1, 2, \ldots, n = d \cdot m$, gdzie
m - liczba okresów
d - liczba sezonów w każdym okresie.

Krok 1. Wyliczamy $\widehat{y}_t$ metodą średnich ruchomych albo analityczną (przy średnich ruchomych „brakuje” kilku skrajnych danych). Niech N_j ($j = 1, 2, \ldots, d$) oznacza zbiór zawierający wskaźniki dotyczące j-tego sezonu oraz p_j liczność zbioru N_j. Zwykle $N_j = \{m(j-1)+1, m(j-1)+2, \ldots, mj\}$ - wtedy $p_j = d$; przy średnich ruchomych zbiory N_1 oraz N_m są pomniejszone o skrajne liczby i p_1 oraz p_m są mniejsze.

Krok 2. Wyliczamy

$$S_t = \frac{y_t}{\widehat{y}^t}.$$

Krok 3. Liczymy średnie

$$\overline{S}^s_j = \frac{1}{p_j}\sum_{i \in N_j} S_i, \quad j = 1, 2, \ldots, d,$$

zwane **surowymi wskaźnikami sezonowości**

Krok 4. (opcjonalny) Liczymy **oczyszczone wskaźniki sezonowości**:

$$\overline{S}^o_j = \frac{\overline{S}^s_j}{k},$$

gdzie $k = \frac{1}{m}\sum_{j=1}^{d}\overline{S}^s_j$.

Współzależność zjawisk

Tablica korelacyjna

x_i \ y_j	y_1	y_2	$\ldots$	y_l	$n_{i.}$
x_1	n_{11}	n_{12}	$\ldots$	n_{1l}	$n_{1.}$
x_2	n_{21}	n_{22}	$\ldots$	n_{2l}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\ldots$	n_{kl}	$n_{k.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\ldots$	$n_{.l}$	n

x_i - wartości pierwszej cechy
y_j - wartości drugiej cechy
n_{ij} - liczność zbioru z warością cechy pierwszej x_i i drugiej y_j
$n_{i.} = \sum_{p=1}^{l} n_{ip}$
$n_{.j} = \sum_{p=1}^{k} n_{pj}$
$n = \sum_{i=1}^{l} \sum_{j=1}^{k} n_{ij}$.

Cechy mierzalne

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_{i.},\ \overline{y} = \frac{1}{n}\sum_{j=1}^{l} x_j n_{.j}$$

$$s^2(x) = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_{i.},\ \ s^2(y) = \frac{1}{n}\sum_{j=1}^{l}(y_j - \overline{y})^2 n_{.j}$$

Kowariancja

$$C(xy) = C(yx) = \frac{1}{n}\sum_{i=1}^{l}\sum_{j=1}^{k} x_i y_j n_{ij} - \overline{xy}$$

Współczynnik korelacji

$$r_{xy} = r_{yx} = \frac{C(xy)}{S(x)S(y)}$$

$$\overline{x}_j = \frac{1}{n_{.j}}\sum_{i=1}^{k} x_i n_{ij}$$

$$\overline{y}_i = \frac{1}{n_{i.}}\sum_{j=l}^{k} x_j n_{ij}$$

$$s_j^2(x) = \frac{1}{n_{.j}}\sum_{i=1}^{k}(x_i - \overline{x}_j)^2 n_{ij}$$

$$s_i^2(y) = \frac{1}{n_{i.}}\sum_{j=1}^{l}(y_j - \overline{x}_i)^2 n_{ij}$$

Dla cechy ciągłej zastępujemy wartości x_i (y_i) środkami przedziałów klasowych.

Cechy niemierzalne

x_i i y_j są wtedy pewnymi charakterystykami (własnościami).

$$\chi^2 = \sum_{i=1}^{l}\sum_{j=1}^{k} \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}}$$

Współczynnik kontyngencji C_{xy}

$$C_{xy} = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Współczynnik Czuprowa T_{xy}

$$T_{xy} = \sqrt{\frac{\chi^2}{n\sqrt{(k-1)(l-1)}}}$$

Model korelacji i regresji liniowej dwóch zmiennych:

Współczynnik korelacji rang Spearmana

Numerujemy dane x_i oraz y_i wg kolejności rosnącej np. odpowiednio liczbami a_i oraz b_i.

Wtedy

$$R(xy) = R(yx) = 1 - \frac{6\sum_{i=1}^{n}(a_i - b_i)^2}{n(n^2-1)}$$

$$\widehat{y}(x) = ax + b$$

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

$$C(xy) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

$$s^2(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2,\ s^2(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

$$r_{xy} = r_{yx} = \frac{C(xy)}{s(x)s(y)}$$

$$a = \frac{C(xy)}{s^2(x)}$$

$$b = \overline{y} - a\overline{x}$$

Definiujemy:

$$\widehat{y}_i = y(x_i)$$

Odchylenie standardowe składnika losowego

$$s_u = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n-2}}$$

Współczynnik zgodności

$$\phi^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Współczynnik determinacji

$$\mathbf{R}^2 = 1 - \phi^2$$

Błędy średnie szacunku parametrów funkcji regresji

$$D(a) = \frac{s_u}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

$$D(b) = \frac{s_u\sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{n\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

Uznaje się, że parametry są oszacowane precyzyjnie jeśli

$$\frac{a}{D(a)} > 2 \ \text{ oraz } \ \frac{b}{D(b)} > 2$$

Prognoza i błąd prognozy

$$y(x) = ax + b \pm S(y)$$

przy czym

$$S(y) = s_u \cdot \sqrt{\left(1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)}$$