WZORY NA CHARAKTERYSTYKI OPISOWE W ANALIZIE STRUKTURY ZJAWISK.

1. Miary przeciętnego poziomu.
2. Miary zróżnicowania.
3. Miary asymetrii.
4. Miary spłaszczenia rozkładu - kurtoza.
5. Miary koncentracji.
6. Momenty zwykłe i centralne.
7. Formuły miar z zakresu analizy struktury stosowane w pakiecie Statgraphics.

Materiały pomocnicze do zajęć ze Statystyki

(Instytut Statystyki i Demografii SGH, Warszawa 2014)

Miary przeciętnego poziomu.

- <u>Miary klasyczne.</u>

1.1 Średnia arytmetyczna

- dane indywidualne:

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \qquad \bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j$$

- dane pogrupowane, cecha skokowa:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i \qquad \bar{x} = \sum_{i=1}^{k} x_i w_i \qquad \text{i} = 1,2,...,\text{k}$$

- dane pogrupowane, cecha ciągła:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} \dot{x}_i n_i \qquad \bar{x} = \sum_{i=1}^{k} \dot{x}_i w_i \qquad \text{i} = 1,2,...,\text{k}$$

- średnia ze średnich cząstkowych:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} \bar{x}_i n_i \qquad \bar{x} = \sum_{i=1}^{k} \bar{x}_i w_i \qquad \text{i} = 1,2,...,\text{k}$$

1.2 Średnia harmoniczna

- dane indywidualne:

$$\bar{x}_h = \frac{n}{\sum_{j=1}^{n} \frac{1}{x_j}}$$

- dane pogrupowane, cecha skokowa:

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}} \qquad \bar{x}_h = \frac{1}{\sum_{i=1}^{k} \frac{w_i}{x_i}}$$

- dane pogrupowane, cecha ciągła:

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{\dot{x}_i}} \qquad \bar{x}_h = \frac{1}{\sum_{i=1}^{k} \frac{w_i}{\dot{x}_i}}$$

1.3 Średnia geometryczna

$$\bar{x}_g = \sqrt[n-1]{\frac{y(t=1)}{y(t=0)}\frac{y(t=2)}{y(t=1)}\cdots\frac{y(t=n-1)}{y(t=n-2)}}$$

$$\bar{x}_g = \sqrt[n-1]{\frac{y(t=n-1)}{y(t=0)}}$$

- <u>Miary pozycyjne:</u>

1.4 mediana i kwartyle

mediana - me = kwartyl drugi $k_{0,5}$

- dane indywidualne

$me = x_{(n+1)/2}$ dla n nieparzystych

$me = \frac{x_{n/2} + x_{(n+2)/2}}{2}$ dla n parzystych

- dane pogrupowane

(wzory z wykorzystaniem liczebności i częstości względnych), dla mediany, kwartyla pierwszego i trzeciego.

$$me = x_{0m} + \left(\frac{n}{2} - n(x_{0m})\right)\frac{h_m}{n_m}$$

$$me = x_{0m} + \left(0{,}5 - F_n(x_{0m})\right)\frac{h_m}{w_m}$$

$$Q_1(x) = x_{0Q_1} + \left(\frac{n}{4} - n(x_{0Q_1})\right)\frac{h_{Q_1}}{n_{Q_1}}$$

$$Q_1(x) = x_{0Q_1} + \left(0{,}25 - F_n(x_{0Q_1})\right)\frac{h_{Q_1}}{w_{Q_1}}$$

$$Q_3(x) = x_{0Q_3} + \left(\frac{3n}{4} - n(x_{0Q_3})\right)\frac{h_{Q_3}}{n_{Q_3}}$$

$$Q_3(x) = x_{0Q_3} + \left(0{,}75 - F_n(x_{0Q_3})\right)\frac{h_{Q_3}}{w_{Q_3}}$$

1.5 Dominanta

(wzory z wykorzystaniem liczebności i częstości względnych)

– dane indywidualne

do - wartość cechy występująca najczęściej

– dane pogrupowane, cecha skokowa

do - wartość cechy, której odpowiada najwyższa liczebność (bądź najwyższa wartość wskaźnika struktury)

– dane pogrupowane, cecha ciągła (wzór interpolacyjny)

$$do = x_{0d} + \frac{n_d - n_{d-1}}{(n_d - n_{d-1}) + (n_d - n_{d+1})} h_d$$

$$do = x_{0d} + \frac{w_d - w_{d-1}}{(w_d - w_{d-1}) + (w_d - w_{d+1})} h_d$$

1. Miary zróżnicowania (dyspersji).

	Klasyczne	**Pozycyjne**
Absolutne	wariancja s^2 odchylenie standardowe s odchylenie przeciętne d_x typowy obszar zmienności x_{typ}	rozstęp R(x) odchylenie ćwiartkowe Q typowy obszar zmienności x_{typ}
Względne	Współczynniki zmienności V(x) $V(x) = \frac{s}{\bar{x}} c$ $V(x) = \frac{d_x}{\bar{x}} c$	$V(x) = \frac{Q}{me} c$ $V_{Q_1Q_3} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$ c-stała (np. c=100)

2.1 Wariancja

– dane indywidualne

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2 \qquad s^2 = \frac{1}{N}\sum_{j=1}^{N}(x_j - \overline{x})^2$$

– dane pogrupowane, cecha skokowa

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i \qquad s^2 = \sum_{i=1}^{k}(x_i - \overline{x})^2 w_i$$

– dane pogrupowane, cecha ciągła

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

– formuły uproszczone

$$s^2 = \overline{x}^2 - (\overline{x})^2$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(x_j)^2 - (\overline{x})^2 \qquad \text{i =1,...,k}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i^2 n_i) - (\overline{x})^2 \qquad s^2 = \sum_{i=1}^{k}(x_i^2 w_i) - (\overline{x})^2$$

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

1.2 Odchylenie standardowe

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

1.3 Odchylenie przeciętne

– dane indywidualne

$$d_x = \frac{1}{n}\sum_{j=1}^{n}|x_j - \overline{x}| \qquad d_x = \frac{1}{N}\sum_{j=1}^{N}|x_j - \overline{x}|$$

– dane pogrupowane, cecha skokowa

$$d_x = \frac{1}{n}\sum_{i=1}^{k}|x_i - \overline{x}|n_i \qquad d_x = \sum_{i=1}^{k}|x_i - \overline{x}|w_i$$

– dane pogrupowane, cecha ciągła

$$d_x = \frac{1}{n}\sum_{i=1}^{n}|\dot{x}_i - \bar{x}|n_i \qquad d_x = \sum_{i=1}^{k}|\dot{x}_i - \bar{x}|w_i$$

2.5 Typowy obszar zmienności

(wyznaczony w oparciu o miary klasyczne)

$$\bar{x} - s \le x_{typ} \le \bar{x} + s$$

2.6 Rozstęp

$R(x) = x_{max} - x_{min}$

2.7 Odchylenie ćwiartkowe

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

2.8 Typowy obszar zmienności

(wyznaczony w oparciu o miary pozycyjne)

$me – Q \le x_{typ} \le me + Q$

2. Miary asymetrii

3.1 Współczynnik asymetrii (miara klasyczna)

$$A = \frac{M_3^{/}}{s^3}$$ ($M^{/}_3$ - moment centralny rzędu 3)

– dane indywidualne

$$A(x) = \frac{\sum_{j=1}^{n}(x_j - \bar{x})^3}{(n-1)s^3} \qquad A(x) = \frac{\sum_{j=1}^{N}(x_j - \bar{x})^3}{N\, s^3}$$

– dane pogrupowane, cecha skokowa

$$A(x) = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i}{(n-1)s^3} \qquad A(x) = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 w_i}{s^3}$$

– dane pogrupowane, cecha ciągła

$$A(x)=\frac{\sum_{i=1}^{k}(\dot{x}_i-\bar{x})^3 n_i}{(n-1)s^3} \qquad A(x)=\frac{\sum_{i=1}^{k}(\dot{x}_i-\bar{x})^3 w_i}{s^3}$$

3.2 Wskaźnik asymetrii (skośności) oparty na miarach pozycyjnych

$$A_2=\frac{(Q_3-me)-(me-Q_1)}{2Q}$$

3.3 Wskaźnik asymetrii (skośności) mieszany

$$A_1=\frac{\bar{x}-do}{s}$$

3. Miary spłaszczenia (spiczastości, skupienia) rozkładu - kurtoza.

$$C_x=\frac{M_4^{/}}{s^4}$$ $M^{/}_4$ - moment centralny rzędu czwartego,

– dane indywidualne:

$$C_x=\frac{\sum_{j=1}^{n}(x_j-\bar{x})^4}{(n-1)s^4} \qquad C_x=\frac{\sum_{j=1}^{N}(x_j-\bar{x})^4}{Ns^4}$$

– dane pogrupowane, cecha skokowa:

$$C_x=\frac{\sum_{i=1}^{k}(x_i-\bar{x})^4 n_i}{(n-1)s^4} \qquad C_x=\frac{\sum_{i=1}^{k}(x_i-\bar{x})^4 w_i}{s^4}$$

– dane pogrupowane, cecha ciągła:

$$C_x=\frac{\sum_{i=1}^{k}(\dot{x}_i-\bar{x})^4 n_i}{(n-1)s^4} \qquad C_x=\frac{\sum_{i=1}^{k}(\dot{x}_i-\bar{x})^4 w_i}{s^4}$$

5. Miary koncentracji (nierównomierności rozdziału sum wartości cechy)

5.1 Współczynnik koncentracji

$$K=\frac{T}{0{,}5}$$

$$K = 1 - \sum_{i=1}^{k} \left[z(x < x_{1i}) + z(x < x_{1i-1}) \right] w_i$$

gdzie: $w_i = \frac{n_i}{N}$; $\mathrm{z_i} = \frac{m_i}{M}$; $\mathrm{M} = \sum \mathrm{x_i} n_i$ $\quad M = \sum \dot{x}_i n_i$

$\mathrm{m_i} = x_i n_i \qquad \mathrm{m_i} = \dot{x}_i n_i$

lub: $K = \frac{T}{5000}$

$\frac{\mathrm{z_i} + \mathrm{z_{i-1}}}{2} \mathrm{w_i}$ - pole pojedynczej figury,

$\sum_{i=1}^{k} \left(\frac{z_i + z_{i-1}}{2} \right) w_i$ - suma pól figur,

Pole T oblicza się więc ze wzoru:

$$T = 5000 - \sum_{i=1}^{k} \left(\frac{z_i + z_{i-1}}{2} \right) w_i$$

6. Momenty zwykłe i centralne

6.1 Moment zwykły rzędu k (k =1,...) w rozkładzie empirycznym

– dane indywidualne

$$M_k = \frac{1}{n} \sum_{j=1}^{n} x_j^k$$

– dane pogrupowane

$$M_k = \frac{1}{n} \sum_{i=1}^{r} x_i^k n_i \qquad M_k = \sum_{i=1}^{r} x_i^k w_i$$

Średnia arytmetyczna jest pierwszym momentem zwykłym w rozkładzie empirycznym:($M_1 = \overline{x}$)

6.2 Moment centralny rzędu k (k =1,...) w rozkładzie empirycznym

– dane indywidualne

$$M_k^{/} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - x)^k$$

$$M_k^{/} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - x)^k n_i \qquad M_k^{/} = \frac{1}{n-1}\sum_{j=1}^{n}(\dot{x}_j - x)^k n_i$$

$$M_k^{/} = \sum_{j=1}^{n}(x_j - x)^k w_i \qquad M_k^{/} = \sum_{j=1}^{n}(x_j - x)^k w_i$$

Pomiędzy wariancją s^2 a średnią arytmetyczną $\overline{x}$ zachodzi związek:

$$s^2 = M_2 - (M_1)^2 = \overline{x^2} - (\bar{x})^2$$