

Politechnika Gdańska



Mariusz Kaszubowski
Katedra Statystyki
Wydział Zarządzania i Ekonomii
Politechnika Gdańska

Analiza struktury – miary położenia

Ćwiczenia nr 3
Podstawy Statystyki

Podstawowe pojęcia

- miary położenia (klasyczne (średnia arytmetyczna, geometryczna, średnia ważona); pozycyjne (dominanta - moda, mediana, kwantyle (kwantyle, decyle, percentyle))
- miary dyspersji (klasyczne (wariancja, odchylenie standardowe, współczynnik zmienności, odchylenie przeciętne); pozycyjne (rozstęp kwartyłowy, współczynnik zmienności dla kwantyli))
- miary asymetrii (rozkład lewostronny (asymetria ujemna), rozkład prawostronny (asymetria dodatnia), brak asymetrii); miary klasyczne asymetrii (moment trzeci centralny, moment trzeci względny); miary pozycyjne asymetrii (kwantylowy współczynnik skośności, decylowy współczynnik skośności))
- miary spłaszczenia – kurtozy (klasyczne (moment czwarty centralny, moment czwarty względny)

Miary tendencji centralnej

- klasyczne
 - średnia arytmetyczna
 - średnia geometryczna
 - średnia harmoniczna
- pozycyjne
 - dominanta (moda)
 - mediana
 - kwartyle (pierwszy, trzeci)
 - decyle
 - percentyle

Parametr	Szereg szczegółowy	Szereg rozdzielczy punktowy	Szereg rozdzielczy przedziałowy
Średnia arytmetyczna	$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$ gdzie $i = 1, 2, \dots, n$	$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \cdot n_i$ gdzie $n = \sum_{i=1}^k n_i$	$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n \dot{x}_i \cdot n_i$ gdzie \dot{x}_i środek i klasy
Dominanta			$D = x_{i_0} + \frac{(n_i - n_{i-1})}{(n_i - n_{i-1}) + (n_i - n_{i+1})} \cdot h_i$
Mediana	$poz. Me = \frac{n+1}{2}$		
			$Me = x_{i_0} + (poz. Me - n_{sk\ i-1}) \cdot \frac{h_i}{n_i}$
Kwartył pierwszy	$poz. Q_1 = \frac{n+1}{4}$		
			$Q_1 = x_{i_0} + (poz. Q_1 - n_{sk\ i-1}) \cdot \frac{h_i}{n_i}$
Kwartył trzeci	$poz. Q_3 = \frac{3 \cdot (n+1)}{4}$		
			$Q_3 = x_{i_0} + (poz. Q_3 - n_{sk\ i-1}) \cdot \frac{h_i}{n_i}$
Decyl rzędu r	$poz. D_r = \frac{r(n+1)}{10}$		
			$D_r = x_{i_0} + (poz. D_r - n_{sk\ i-1}) \cdot \frac{h_i}{n_i}$

Zadanie 1

Poniższy zestaw danych przedstawia wyniki biegu na 60m w sekundach. Wyznacz miary tendencji centralnej (średnia, mediana, dominanta, kwartył pierwszy i trzeci) dla podanego szeregu szczegółowego.

10, 11, 9, 9, 12, 16, 12, 13, 8, 9, 10, 15

Rozwiązanie

\bar{x}	11,17
D	9
<i>poz Me</i>	6,5
Me	10,5
<i>poz Q_1</i>	3,25
Q_1	9
<i>poz Q_3</i>	9,75
Q_3	12,5

8, 9, 9, 9, 10, 10, 11, 12, 12, 13, 15, 16

Zadanie 2

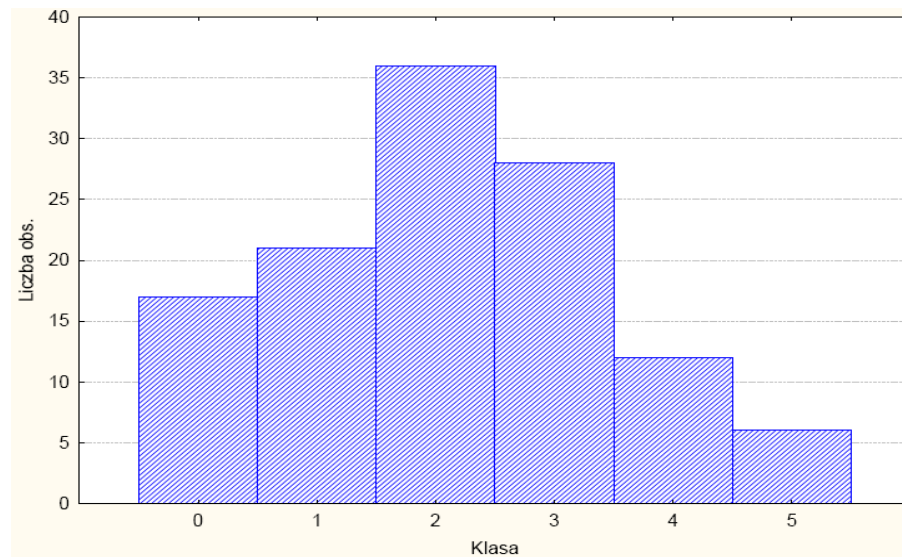
Poniższy szereg przedstawia zestawienie meczy piłki nożnej Ekstraklasy w rundzie jesiennej 2008/2009 pod względem strzelonych bramek w poszczególnych meczach. Wyznacz miary tendencji centralnej (średnia, mediana, dominanta, kwartył pierwszy i trzeci) dla podanego szeregu rozdzielczego punktowego.

liczba bramek w meczu (x_i)	liczba meczy (n_i)	$x_i \cdot n_i$	$n_{sk i}$
0	17		
1	21		
2	36		
3	28		
4	12		
5	6		
Σ			

Rozwiązanie

liczba bramek w meczu (x_i)	liczba meczy (n_i)	$x_i \cdot n_i$	$n_{sk i}$
0	17	0	17
1	21	21	38
2	36	72	74
3	28	84	102
4	12	48	114
5	6	30	120
Σ	120	255	X

\bar{x}	2,125
D	2
poz Me	60,5
Me	2
poz Q_1	30,25
Q_1	1
poz Q_3	90,75
Q_3	3



Zadanie 3

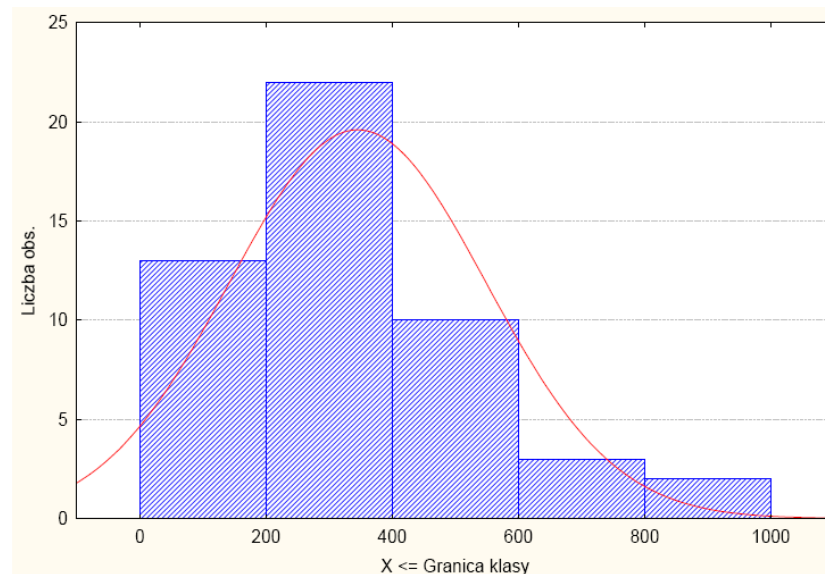
Podany szereg przedstawia zestawienie wyników ankiety przeprowadzonej w grupie 50 kobiet w wieku 20-40 lat odnośnie kwoty miesięcznych wydatków w zł na środki pielęgnacyjne. Wyznacz miary tendencji centralnej (średnia, mediana, dominanta, kwartył pierwszy i trzeci, decyl dziewiąty) dla podanego szeregu rozdzielczego przedziałowego.

wydatki na „waciki” ($x_{0i}; x_{1i}$]	liczba respondentek (n_i)	\hat{x}_i	$\hat{x}_i \cdot n_i$	$n_{zk i}$
0-200	13			
200-400	22			
400-600	10			
600-800	3			
800-1000	2			
Σ	50			

Rozwiązanie

wydatki na „waciki” ($x_{0i}; x_{1i}$]	liczba respondentek (n_i)	\dot{x}_i	$\dot{x}_i \cdot n_i$	$n_{sk i}$
0-200	13	100	1300	13
200-400	22	300	6600	35
400-600	10	500	5000	45
600-800	3	700	2100	48
800-1000	2	900	1800	50
Σ	50	X	16800	X

\bar{x}	336
D	285,71
poz Me	25,5
Me	313,64
poz Q_1	12,75
Q_1	196,15
poz Q_3	38,25
Q_3	465
poz D_9	45,9
D_9	660



Analiza struktury – miary dyspersji, asymetrii i spłaszczenia

Ćwiczenia nr 4
Podstawy Statystyki

Miary dyspersji

- klasyczne
 - wariancja
 - odchylenie standardowe
 - współczynnik zmienności
 - odchylenie przeciętne
- pozycyjne
 - rozstęp
 - rozstęp kwartyłowy
 - odchylenie ćwiartkowe
 - współczynnik zmienności

Parametr	Szereg szczegółowy	Szereg rozdzielczy punktowy	Szereg rozdzielczy przedziałowy
Wariancja – moment drugi centralny	$s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$	$s^2 = \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x})^2 \cdot n_i$	$s^2 = \frac{1}{n} \cdot \sum_{i=1}^k (\dot{x}_i - \bar{x})^2 \cdot n_i$
Odchylenie standardowe	$s = \sqrt{s^2}$		
Typowy obszar zmienności	$\bar{x} - s < x_{op} < \bar{x} + s$		
Rozstęp	$R = x_{\max} - x_{\min}$		
Rozstęp kwartyłowy	$R(Q) = Q_3 - Q_1$		
Odchylenie ćwiartkowe	$Q = \frac{Q_3 - Q_1}{2}$		
Współczynnik zmienności	$V(s) = \frac{s}{\bar{x}} \cdot 100\%$		
	$V(Q) = \frac{Q}{Me} \cdot 100\%$		
Odchylenie przeciętne	$d = \frac{1}{n} \cdot \sum_{i=1}^n x_i - \bar{x} $	$d = \frac{1}{n} \cdot \sum_{i=1}^k x_i - \bar{x} \cdot n_i$	$d = \frac{1}{n} \cdot \sum_{i=1}^k \dot{x}_i - \bar{x} \cdot n_i$

Miary asymetrii

- klasyczne
 - moment trzeci centralny
 - moment trzeci względny
- pozycyjne
 - kwartylowy współczynnik skośności
 - decylowy współczynnik skośności
 - współczynnik asymetrii oparty na średniej i dominancie

Parametr	Szereg szczegółowy	Szereg rozdzielczy punktowy	Szereg rozdzielczy przedziałowy
Moment trzeci centralny	$\mu_3 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^3$	$\mu_3 = \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x})^3 \cdot n_i$	$\mu_3 = \frac{1}{n} \cdot \sum_{i=1}^k (\dot{x}_i - \bar{x})^3 \cdot n_i$
Moment trzeci względny	$\alpha_3 = \frac{\mu_3}{s^3} \quad \text{gdzie} \quad \alpha_3 \in \langle -2; 2 \rangle$		
Kwartylowy współczynnik skośności	$A(Q) = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1} \quad \text{gdzie} \quad A(Q) \in \langle -1; 1 \rangle$		
Decylowy współczynnik skośności	$A(D) = \frac{(D_9 - Me) - (Me - D_1)}{D_9 - D_1} \quad \text{gdzie} \quad A(D) \in \langle -1; 1 \rangle$		
Współczynnik asymetrii	$A(\bar{x}) = \frac{\bar{x} - D}{s} \quad \text{gdzie} \quad A(\bar{x}) \in \langle -1; 1 \rangle$		
Rodzaj asymetrii	<p>współczynnik < 0 $\bar{x} < Me < D$ asymetria ujemna (rozkład lewoskośny)</p> <p>współczynnik > 0 $D < Me < \bar{x}$ asymetria dodatnia (rozkład prawoskośny)</p> <p>współczynnik $= 0$ $\bar{x} = Me = D$ brak asymetrii (rozkład symetryczny)</p>		

Miary spłaszczenia - kurtozy

- klasyczne
 - moment czwarty centralny
 - moment czwarty względny
 - współczynnik ekscesu
- pozycyjne
 - nie stosuję się

Parametr	Szereg szczegółowy	Szereg rozdzielczy punktowy	Szereg rozdzielczy przedziałowy
Moment czwarty centralny	$\mu_4 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^4$	$\mu_4 = \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x})^4 \cdot n_i$	$\mu_4 = \frac{1}{n} \cdot \sum_{i=1}^k (\dot{x}_i - \bar{x})^4 \cdot n_i$
Moment czwarty względny	$\alpha_4 = \frac{\mu_4}{s^4}$		
Współczynnik ekscesu	$K = \alpha_4 - 3$		
Rodzaj koncentracji	współczynnik	$\alpha_4 < 3 \quad (K < 0)$	rozkład spłaszczony
	współczynnik	$\alpha_4 > 3 \quad (K > 0)$	rozkład wysmukły
	współczynnik	$\alpha_4 = 3 \quad (K = 0)$	koncentracja normalna

Moment zwykły i centralny

Momenty	
Moment centralny rzędu r	$\mu_r = \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x})^r \cdot n_i$
Moment zwykły rzędu r	$m_r = \frac{1}{n} \cdot \sum_{i=1}^k x_i^r \cdot n_i$
Zależności	$m_1 = \bar{x}$ $m_2 = (\bar{x})^2$ $\mu_1 = 0$ $\mu_2 = m_2 - m_1^2$ $\mu_3 = m_3 - 3m_2m_1 + 2m_1^3$ $\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$

Zadanie 4

Poniższy szereg rozdzielczy przedstawia dane o zarobkach pracowników pewnej firmy produkcyjnej w skali roku w tysiącach zł. Dokonaj analizy miar położenia, dyspersji, asymetrii i spłaszczenia

$(x_{0i}; x_{1i}]$	n_i	\hat{x}_i	$\hat{x}_i \cdot n_i$	$(\hat{x}_i - \bar{x})$	$(\hat{x}_i - \bar{x})^2 \cdot n_i$	$(\hat{x}_i - \bar{x})^3 \cdot n_i$	$(\hat{x}_i - \bar{x})^4 \cdot n_i$	$ \hat{x}_i - \bar{x} \cdot n_i$
15-20	13							
20-25	48							
25-30	15							
30-35	11							
35-40	8							
40-45	5							
Σ		\bar{x}						

Rozwiązanie

$(x_{0i}; x_{1i}]$	n_i	\hat{x}_i	$\hat{x}_i \cdot n_i$	$(\hat{x}_i - \bar{x})$	$(\hat{x}_i - \bar{x})^2 \cdot n_i$	$(\hat{x}_i - \bar{x})^3 \cdot n_i$	$(\hat{x}_i - \bar{x})^4 \cdot n_i$	$ \hat{x}_i - \bar{x} \cdot n_i$
15-20	13	17,5	227,5	-8,40	917,28	-7705,15	64723,28	109,2
20-25	48	22,5	1080	-3,40	554,88	-1886,59	6414,41	163,2
25-30	15	27,5	412,5	1,60	38,40	61,44	98,30	24
30-35	11	32,5	357,5	6,60	479,16	3162,46	20872,21	72,6
35-40	8	37,5	300	11,60	1076,48	12487,17	144851,15	92,8
40-45	5	42,5	212,5	16,60	1377,80	22871,48	379666,57	83
Σ	100	X	2590	24,60	4444,00	28990,80	616625,92	544,80

\bar{x}	25,9
s^2	44,44
s	6,67
x_{dp}	$19,23 < x_{\text{dp}} < 32,57$
D	22,57
Me	23,91

Q_1	21,28
Q_3	29,92
Q	4,32
$V(s)$	25,75 %
$V(Q)$	18,07 %
d	5,45

μ_3	289,91
α_3	0,98
$A(Q)$	0,39
$A(\bar{x})$	0,5
μ_4	6166,26
α_4	3,12

