

Metody analizy danych – ćwiczenia

Mierniki statystyczne

Proszę również samodzielnie wykonać ćwiczenia w oparciu o program Excel. Pomocne będą następujące książki:

- Michael R. Middleton (2004): *Microsoft Excel w analizie danych*. – Wydawnictwo RM: Warszawa,
- Andrzej Obecny (2002): *Statystyka opisowa w Excelu dla szkół. Ćwiczenia praktyczne*. – Helion: Warszawa.

Program ćwiczeń obejmuje następująca zadania:

1. Czasy dojazdu w minutach dwóch grup pracowników zatrudnionych w dwóch zakładach pracy (I i II) mieszczących się w centrum miasta są następujące:

zakład I					zakład II				
22	26	27	27	9	38	37	22	31	64
11	29	33	27	28	39	30	26	42	44
30	31	14	33	30	46	41	27	34	39
16	20	18	34	33	16	46	60	37	36
34	30	29	24	29	30	26	36	51	27
29	27	42	54	60	32	26	20	41	27
52	36	37	43	57	40	44	26	19	18
39	40	36	50	44	54	49	28	42	43
33	31	42	49	32	39	38	41	50	24
36	41	32	48	31	33	38	36	18	53
26	36	38	37	38	50	59	40	36	16
					36	39	48	46	58
					33	43	37	44	40

Utworzyć szeregi rozdzielcze przedziałowe (przyjąć 6 przedziałów w zakresie od 5 do 65 minut), narysować histogramy oraz wyznaczyć podstawowe parametry statystyczne. Porównać obydwie zakłady pod względem czasów dojazdu. W ćwiczeniu wykorzystać program *Excel*.

2. Pokazać, że dla średniej arytmetycznej zachodzi wzór

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

3. W pewnym zakładzie zbadano pracowników bezpośrednio produkcyjnych pod względem stażu pracy. Okazało się, że 25% tych pracowników pracowało poniżej 6 lat, połowa od 6 do 12 lat, natomiast wśród pozostałych najdłuższy staż wyniósł 18 lat. Średni staż pracy pracowników inżynieryjno-technicznych w tym zakładzie wynosił 12 lat. Jaki był średni staż pracy ogółu pracowników tego zakładu, jeśli wiadomo, że grupa pracowników bezpośrednio produkcyjnych była 2.5-krotnie liczniejsza od inżynieryjno-technicznych?
4. Strukturę rodzin wg liczby członków w pewnej miejscowości charakteryzuje rozkład

liczba członków	2	3	4	5	6	7	8
odsetek rodzin	15	30	20	15	10	5	5

Za pomocą miar przeciętnych scharakteryzować ten rozkład. Wyniki zinterpretować.

5. W szeregach rozdzielczych przedziałowych bezpośrednio można określić tylko przedział, w którym znajduje się modalna (jest to przedział o największej liczebności). Konkretną wartość liczbową należącą do tego przedziału i będącą dobrym przybliżeniem modalnej wyznacza się często wg wzoru

$$Mo \approx x_m + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} h_m,$$

gdzie: x_m – dolna granica przedziału, w którym znajduje się modalna, n_m – liczebność przedziału modalnej, n_{m-1} – liczebność przedziału poprzedzającego przedział modalnej, n_{m+1} – liczebność przedziału następującego po przedziale modalnej, h_m – długość przedziału modalnej. Uzasadnić poprawność tego wzoru.

6. W spółdzielni mieszkaniowej dominują mieszkania o powierzchni 45 – 50 m². Na podstawie badań statystycznych ustalono, że najwięcej mieszkań (30%) miało powierzchnię 48 m², a 20% mieszkań miało powierzchnię od 40 do 45 m². Jaki odsetek mieszkań ma powierzchnię od 50 do 55 m²?
7. Udowodnić, że wariancję

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

można równoważnie wyznaczyć ze wzoru

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

Wyznaczyć liczby operacji potrzebnych do wykorzystania obydwu wzorów.

8. Lekkoatleta A uzyskał w skoku w dal następujące wyniki na zawodach w całym sezonie (w m): 6.82, 6.96, 7.23, 7.05, 7.80, 7.75. Lekkoatleta B , startujący w tych samych zawodach, uzyskał takie wyniki, że ich średnia arytmetyczna wyniosła 7.5 m, a suma ich kwadratów 450.2592 m^2 . Który z tych lekkoatletów osiągał regularniejsze wyniki?
9. Dienne zużycie energii elektrycznej (w kWh) w pewnym bloku mieszkalnym kształtowało się następująco:

zużycie	2-4	4-6	6-8	8-10	10-12	12-14
liczba rodzin	6	10	30	40	10	4

Obliczyć średnią arytmetyczną, medianę, modalną, współczynnik zmienności oraz wyznaczyć obszar zmienności.