

Metody analizy danych – ćwiczenia

## Mierniki statystyczne (c.d.) i analiza współzależności zjawisk (2 godziny)

Pomocnicze książki do samodzielnego wykonania tego ćwiczenia z zastosowaniem Excela są następujące:

- 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,
- Bernard V. Liengme (2002): *Microsoft Excel w nauce i technice*. – Wydawnictwo Read Me: Warszawa.

Program ćwiczeń obejmuje następująca zadania:

1. W zadaniu chodzi o wstępne zapoznanie się z narzędziami analizy danych z dodatku *Analysis ToolPak* do programu Excel.
  - (a) Z zastosowaniem polecenia *Generowanie liczb pseudolosowych* z narzędzia *Analiza danych*, w komórkach A2 : A101 wygenerować 100 liczb losowych o rozkładzie Poissona, przyjmując wartość  $\lambda = 5$ . Dla tak otrzymanych danych określić rozstęp (zastosować funkcje `min()` i `max()`). Posortować dane, otrzymując w ten sposób szereg szczegółowy. Przy pomocy polecenia `licz.jezeli()` utworzyć odpowiedni szereg rozdzielczy punktowy.
  - (b) Z zastosowaniem polecenia *Generowanie liczb pseudolosowych* z narzędzia *Analiza danych*, w komórkach A2 : A101 Arkusza 1 wygenerować 100 liczb losowych o rozkładzie normalnym, przyjmując wartości  $\text{średnia} = 10$  i  $\text{odchylenie standardowe} = 0.5$ . W komórkach A2 : A101 Arkusza 2 wpisać liczby z odpowiednich komórek Arkusza 1 po uprzednim zaokrągleniu ich do dwóch miejsc po przecinku (funkcja `zaokr()`). Na podstawie tak otrzymanych danych utworzyć szereg rozdzielczy przedziałowy stosując najpierw funkcję `częstość()`, a następnie polecenie *Histogram* z narzędzia *Analiza danych*. Utworzyć szereg skumulowany. Narysować histogram. Zbadać wpływ szerokości przedziałów na otrzymywane rezultaty.

(c) Z zastosowaniem opanowanej wiedzy rozwiązać Zadanie 1 z poprzednich ćwiczeń.

2. Tętna spoczynkowe 16 lekkoatletów są następujące:

67 57 56 57 58 56 54 64 53 54 54 55 57 68 60 58

Określić rozstęp, odchylenie ćwiartkowe, wariancję, odchylenie standardowe i współczynnik zmienności dla tych danych. Znaleźć kwartyle, decyle, 47. percentyl oraz 88. percentyl. Sporządzić (ręcznie) wykres pudełkowy.

3. *Ocenić* rozstęp, odchylenie ćwiartkowe, wariancję, odchylenie standardowe i współczynnik zmienności dla następującego rozkładu:

przedział	liczebność
10–15	2
15–20	4
20–25	4
25–30	3
30–35	2

Ocenić również kwartyle, decyle, 11. percentyl, 22. percentyl, 66. percentyl oraz 88. percentyl. Sporządzić wykres pudełkowy.

4. Która z miar położenia najlepiej charakteryzuje następujący zbiór danych:

1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 100?

Wyznaczyć wartość tej miary.

5. Wędkarz zwykle łowił w ciągu sezonu płocie o następujących długościach: 21 cm, 18 cm, 23 cm, 24 cm, 19 cm. Ocenić, na ile prawdopodobna jest przechwałka, że złowił płoć długości 41 cm?

6. Porównać rozproszenie i asymetrię wieku pracowników trzech przedsiębiorstw dysponując następującymi danymi:

Wiek	A [%]	B [%]	C [%]
poniżej 26 lat	5	5	25
26–28	10	10	10
28–30	10	10	10
30–32	10	10	5
32–34	30	10	5
34–36	10	10	10
36–38	10	20	10
38–40	10	15	10
powyżej 40	5	10	15

7. W grupie 20 studentów i 14 studentek przeprowadzono test. Dla studentów wyniki testu były następujące:  $\bar{x} = 84$ ,  $Me = Mo = 74$ ,  $s = 12$  punktów; dla studentek  $\bar{x} = Me = Mo = 74$ ,  $s = 8$  punktów.
- (a) Prowadzący zajęcia stwierdził, że studenci wypadli lepiej. Czy jego sąd był słuszny?
- (b) Określić średni wynik testu i współczynnik zmienności dla całej grupy.
8. Rozkład jednostkowych kosztów produkcji wyrobu jest lewostronnie asymetryczny o współczynniku skośności równym  $-1$ . Odchylenie standardowe kosztu jednostkowego wynosi 20 gr, a najczęściej spotykany koszt 4 zł 20 gr. Wyprodukowano 1000 sztuk wyrobu. Obliczyć średni koszt jednostkowy wyrobu i łączny koszt 1000 wyrobów.
9. 50 samochodów pewnej marki poddano badaniom pod względem osiągniętej prędkości maksymalnej. Wyniki zawiera tabela:

Prędkość maksymalna (w km/h)	130–140	140–150	150–160	160–170
Liczba samochodów	8	12	20	10

- (a) Przedstawić rozkład i dystrybuantę empiryczną.
- (b) Zmierzyć wielkość i rodzaj asymetrii rozkładu.
- (c) Porównać zmienność osiągniętej prędkości maksymalnej ze zmiennością zużycia paliwa wiedząc, że zmienność zużycia paliwa mierzona współczynnikiem zmienności  $V$  wynosi 48%.
10. Dla danych

$x$	-3	-1	1	3
$y$	8	4	5	-1

Wyznaczyć współczynnik korelacji i równanie najlepiej dopasowanej linii prostej.

11. Zmierzono prędkości pewnego obiektu w kolejnych chwilach czasu. Wyniki przedstawia tabela:

czas	2	4	6	8	10
prędkość	22	42	62	80	100

Wiedząc, że prędkość przy stałym przyspieszeniu  $a$  opisuje równanie

$$v = v_0 + at,$$

należy ocenić wartość  $a$  i prędkość początkową  $v_0$ . W tym celu należy wykorzystać funkcje `nachylenie()` i `odcięta()`. Z zastosowaniem polecenia `reglinw()` wyznaczyć przewidywane prędkości dla tak otrzymanego modelu dla czasów przedstawionych w tabeli. Narysować wykres oryginalnych danych (diagram korelacyjny) i dodać do niego linię trendu. Zinterpretować otrzymaną wartość współczynnika determinacji  $R^2$ . Zapoznać się z funkcją `reglinp` i ponownie wykonać to zadanie z jej wykorzystaniem.

obs.	wzrost	waga	obs.	wzrost	waga	obs.	wzrost	waga
1	160	57.9	25	170	65.3	49	180	71,8
2	160	58.9	26	170	65.2	50	180	72.0
3	160	63.3	27	170	68.3	51	185	74.1
4	160	56.8	28	170	62.3	52	185	74.4
5	160	66.8	29	170	67.8	53	185	72.0
6	160	64.5	30	170	66.5	54	185	66.1
7	160	67.1	31	175	67.4	55	185	69.4
8	160	58.0	32	175	67.7	56	185	71.8
9	160	62.9	33	175	62.6	57	185	73.8
10	160	57.7	34	175	70.6	58	185	69.1
11	165	68.5	35	175	72.0	59	185	72.3
12	165	69.8	36	175	68.3	60	185	72.8
13	165	58.5	37	175	72.9	61	190	72.6
14	165	66.3	38	175	63.4	62	190	81.1
15	165	65.8	39	175	80.7	63	190	78.3
16	165	61.0	40	175	67.3	64	190	72.9
17	165	74.5	41	180	67.4	65	190	79.6
18	165	59.3	42	180	70.6	66	190	77.1
19	165	67.8	43	180	72.4	67	190	84.5
20	165	70.1	44	180	73.2	68	190	74.0
21	170	72.7	45	180	72.8	69	190	77.5
22	170	75.1	46	180	66.4	70	190	75.2
23	170	68.0	47	180	73.0			
24	170	72.2	48	180	78.0			

Tabela 1: Wzrost i waga 70 osób.

12. Tabela 1 przedstawia wzrost i wagę 70 osób. Sprawdzić czy istnieje korelacja liniowa między wagą a wzrostem.