

Podstawy analizy danych

Ćwiczenie 1: Porządkowanie danych i miary położenia

1. Grupy krwi 25 dawców są następujące:

| | | | | |
|----|---|----|----|---|
| AB | B | A | 0 | B |
| 0 | B | 0 | A | 0 |
| B | 0 | B | B | B |
| A | 0 | AB | AB | 0 |
| A | B | AB | 0 | A |

Nie są one szczególnie przejrzyste. Proszę więc je uporządkować wypełniając poniższą tabelę:

| Grupa krwi | Liczebność | Liczebność względna (%) |
|--------------|------------|-------------------------|
| A | | |
| B | | |
| 0 | | |
| AB | | |
| Razem | | |

Otrzymuje się w ten sposób *rozkład liczebności* cechy jakościowej, jaką jest grupa krwi.

Przedstawić rozkłady na wykresach słupkowych. Czym będą się one różnić?

2. Pracownik kontroli jakości w ciągu 25 dni zaobserwował następujące liczby produktów z wadami wytwarzanych w pewnym procesie produkcyjnym:

| Dzień | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|-------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Wady | 10 | 10 | 6 | 12 | 6 | 9 | 16 | 20 | 11 | 10 | 11 | 11 | 9 |
| Dzień | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | |
| Wady | 12 | 11 | 7 | 10 | 11 | 14 | 21 | 12 | 6 | 10 | 11 | 6 | |

Uporządkować te dane wypełniając tabelę o poniższym wzorcu:

| Liczba wad | Liczebność | Liczebność względna (%) |
|--------------|------------|-------------------------|
| ... | ... | ... |
| Razem | | |

W jakim porządku najprzejrzyściej wpisać wartości w pierwszej kolumnie? Mówimy, że cecha jest

w tym przypadku *porządkowa*. Jak będzie wyglądał wykres słupkowy?

3. Zarejestrowano wiek 20 pracowników zgłaszających się na okresowe badania w pewnym zakładzie pracy. Oto wyniki w latach

36, 41, 33, 34, 38, 26, 33, 36, 30, 48, 39, 31, 35, 36, 38, 37, 22, 31, 25, 32.

Ponieważ jest bardzo mało powtarzających się wartości, nie ma sensu tworzenie rozkładu wieku tak jak w poprzednim zadaniu. Zamiast tego, proszę wypełnić poniższą tabelę dotyczącą przedziałów wiekowych:

| Wiek | Liczebność | Liczebność względna (%) |
|-------|------------|-------------------------|
| 20–25 | | |
| 25–30 | | |
| 30–35 | | |
| 35–40 | | |
| 40–45 | | |
| 45–50 | | |
| Razem | | |

Proszę potraktować przedziały jako lewostronnie domknięte i prawostronnie otwarte (taką konwencję przyjmuje się najczęściej). Jak narysować wykres słupkowy liczebności w tym przypadku? Wykres ten nazywa się *histogramem*.

Zasady określania liczby przedziałów i ich granic zostaną omówione na wykładzie.

4. Najczęstszym parametrem położenia jest *średnia arytmetyczna*, oznaczana symbolem \bar{x} jeśli cechę oznacza się symbolem x . Obliczyć wartość średniej dla danych

3 8 6 14 0 -4 0 12 -7 0 -10.

Proszę dla każdej danej wyznaczyć jej odchylenie od średniej i te odchylenia zsumować. Wynik jest łatwą do zapamiętania liczbą, która – jak się okazuje – nie zależy od danych.

5. Wyznaczyć średnią dla danych reprezentowanych w postaci poniższej tabeli:

| Obserwacja | Liczebność |
|------------|------------|
| 20 | 2 |
| 29 | 4 |
| 30 | 4 |
| 39 | 3 |
| 44 | 2 |

6. **Medianą** nazywa się parametr, do którego wyznaczenia należy najpierw posortować dane w kolejności niemalejącej. Gdy liczba danych jest nieparzysta, mediana jest wartością środkową. Gdy liczba danych jest parzysta – jest średnią dwóch wartości środkowych. Będziemy oznaczać ją symbolem *Me*.

Wyznaczyć medianę dla danych z poprzedniego zadania. Czy wynik jest taki sam jak średnia?

Wyznaczyć również medianę wieku ośmiu studentów:

23 19 32 25 26 22 24 20.

6. **Modalną** (oznacza się ją symbolem *Mo*) nazywa się wartość występującą w zbiorze danych najczęściej. Modalna może nie istnieć (jest tak, gdy wszystkie obserwacje mają tę samą liczebność) lub może być wiele modalnych (w sytuacji, gdy więcej niż jedna obserwacja pojawia się najczęściej).

Wyznaczyć modalną dla poniższych danych:

- a) 3 5 1 4 2 9 6 10
 b) 3 5 1 4 2 9 6 10 5 3 4 3 9
 3 6 1
 c) 6 10 5 3 4 3 9 3 6 1 6

7. Zaproponować sposób oceny średniej, mediany i modalnej w sytuacji, gdy dane są dostępne jedynie w postaci szeregu przedziałowego:

| Wyniki egzaminu (pkt) | Liczebność |
|-----------------------|------------|
| 10–15 | 2 |
| 15–20 | 4 |
| 20–25 | 4 |
| 25–30 | 3 |
| 30–35 | 2 |

Literatura

1. W. Kryszicki, J. Bartos, W. Dyczka, K. Królikowska, M. Wasilewski (2011): *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Część 2: Statystyka matematyczna*, Wydawnictwo Naukowe PWN, Warszawa
2. S. Ostasiewicz, Z. Rusnak, U. Siedlecka (1997): *Statystyka: Elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław
3. J. Koronacki, J. Mielniczuk (2001): *Statystyka dla studentów kierunków technicznych i przyrodniczych*, Wydawnictwa Naukowo-Techniczne, Warszawa