

Zajęcia 2

Statystyka opisowa

Zadanie 1. Zmienna WYNIK w pliku Ankieta.txt opisuje wyniki badania ankietowego działalności prezydenta pewnego miasta. Wybrano w sposób losowy 100 mieszkańców miasta i zadano im następujące pytanie: Jak oceniasz działalność prezydenta miasta? Do wyboru były następujące odpowiedzi: zdecydowanie dobrze (a), dobrze (b), źle (c), zdecydowanie źle (d), nie mam zdania (e).

- (a) Zaimportuj dane z pliku Ankieta.txt do zmiennej Ankieta. **Wskazówka:** `setwd()`, `read.table()`

```

      PLEC SZKOLA WYNIK
1      m      p      d
2      m      s      e
3      m      w      a
4      m      s      d
5      m      p      c
6      m      w      c
7      m      s      c
8      k      w      c
9      m      w      e
10     k      s      c
...

```

- (b) Przedstaw rozkłady odpowiedzi dla wszystkich osób oraz wyłącznie dla osób z wykształceniem podstawowym w postaci szeregu rozdzielczego. **Wskazówka:** `table()`, `prop.table()`

```

  a  b  c  d  e              a  b  c  d  e
10 23 29 17 21             0.10 0.23 0.29 0.17 0.21

  a  b  c  d  e              a  b  c  d  e
2  3  4  7  1              0.12 0.18 0.24 0.41 0.06

```

- (c) Zilustruj wyniki badania (wg wykształcenia) przy pomocy wykresów kołowych. **Wskazówka:** `pie()` **Odpowiedź:** patrz Rysunek 1
- (d) Utwórz ramkę danych Kobiety zawierającą dane dotyczące wyłącznie kobiet z wykształceniem co najmniej średnim. **Wskazówka:** `subset()`

```

8      k      w      c
10     k      s      c
11     k      w      b
12     k      s      c
16     k      s      e
17     k      s      a
...
99     k      s      e

```

- (e) Dla danych z ramki Ankieta dokonaj kategoryzacji zmiennej WYNIK w taki sposób, aby kategoria 1 obejmowała odpowiedzi a i b, 2 odpowiedzi c i d oraz 3 odpowiedź e. **Wskazówka:** `recode()` z pakietu `car`

```

[1] 2 3 1 2 2 2 2 3 2 1 2 2 1 2 3 1 3 1 2 1 3 3 2 1 3 1 1 1 2 2 1 2 1 2 1 2
[38] 3 2 1 3 2 1 1 2 2 2 1 2 3 2 2 2 1 1 1 3 2 3 2 1 1 3 2 3 2 2 1 2 1 2 1 3 2
[75] 2 2 2 2 1 3 2 1 1 2 2 3 3 3 2 1 3 2 2 1 1 1 1 2 3 2
Levels: 1 2 3

```

- (f) Zilustruj wyniki ankiety przy pomocy wykresu słupkowego, z podziałem na kobiety i mężczyzn. **Wskazówka:** `barplot()` **Odpowiedź:** patrz Rysunek 2

Zadanie 2. Zbadano 200 losowo wybranych 5-sekundowych odcinków czasowych pracy pewnej centrali telefonicznej. Notowano liczbę zgłoszeń. Wyniki zawarto w pliku `Centrala.RData`.

- (a) Przedstaw rozkład liczby zgłoszeń w tej próbie w postaci szeregu rozdzielczego. **Wskazówka:** `table()`, `prop.table()`, `cbind()`, `data.frame()`

	Liczebność	Procent
0	32	0.160
1	67	0.335
2	49	0.245
3	31	0.155
4	15	0.075
5	6	0.030

- (b) Zilustruj graficznie ten rozkład przy pomocy wykresu słupkowego. **Wskazówka:** `barplot()`
Odpowiedź: patrz Rysunek 3
- (c) Podaj średnią liczbę zgłoszeń, medianę liczby zgłoszeń, odchylenie standardowe liczby zgłoszeń oraz współczynnik zmienności liczby zgłoszeń. **Wskazówka:** `mean()`, `median()`, `sd()`

	Średnia	Mediana	Odch.stand.	Wsp.zm.
1	1.74	2	1.28086	0.7361266

Zadanie 3. Zmienna w pliku `Awarie.txt` opisuje wyniki 50 pomiarów czasu bezawaryjnej pracy pewnego urządzenia (w h).

- (a) Przedstaw rozkład czasu bezawaryjnej pracy w tej próbie w postaci szeregu rozdzielczego. **Wskazówka:** `cut()`, `table()`, `prop.table()`, `data.frame()`, `cbind()`

	Liczność	Procent
(0,500]	17	0.34
(500,1e+03]	12	0.24
(1e+03,1.5e+03]	7	0.14
(1.5e+03,2e+03]	5	0.10
(2e+03,2.5e+03]	4	0.08
(2.5e+03,3e+03]	3	0.06
(3e+03,3.5e+03]	2	0.04

- (b) Zilustruj graficznie ten rozkład przy pomocy wykresu ramkowego. **Wskazówka:** `boxplot()`
Odpowiedź: patrz Rysunek 4
- (c) Podaj średnią, medianę, odchylenie standardowe, współczynnik zmienności, współczynnik asymetrii oraz współczynnik koncentracji czasu bezawaryjnej pracy. **Wskazówka:** `mean()`, `median()`, `sd()`, `skewness()`, `kurtosis()` z pakietu `e1071`

	Średnia	Mediana	Odch.stand.	Wsp.zm.	Wsp.as.	Wsp.konc.
1	1101.36	734	883.2735	0.8019844	0.9109508	-0.354536

Zadanie 4. Obserwowano plony (w q/ha) bardzo wczesnych odmian ziemniaków. Uzyskano następujące wyniki: Aster: 168, 160, 169, 175, 159, 162; Drop: 130, 136, 140, 137, 124; Frezja: 148, 149, 130, 139, 138, 140; Irys: 126, 128, 131, 130, 127; Ruta: 145, 149, 148, 152, 150, 145.

- (a) W odpowiedni sposób wprowadź zebrane dane do arkusza danych. **Wskazówka:** `rep()`, `data.frame()`

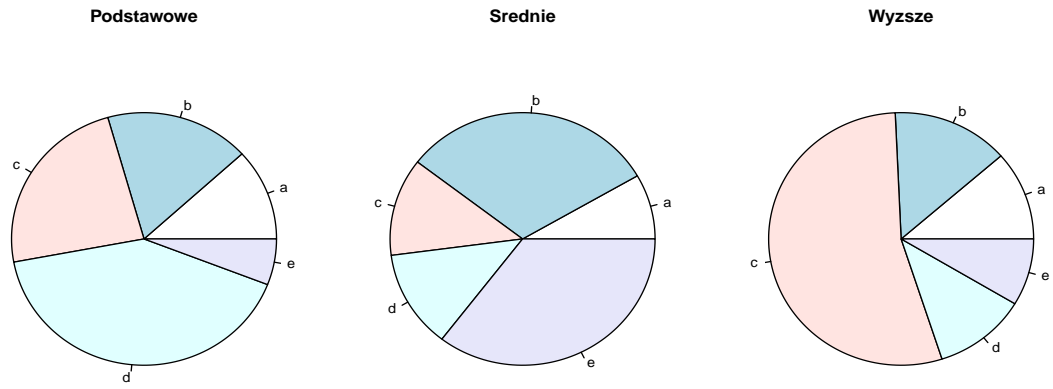
	Plon	Odmiana
1	168	Aster
2	160	Aster
3	169	Aster

4	175	Aster
5	159	Aster
6	162	Aster
7	130	Drop
8	136	Drop
9	140	Drop
10	137	Drop
11	124	Drop
12	148	Frezja
13	149	Frezja
14	130	Frezja
15	139	Frezja
16	138	Frezja
17	140	Frezja
18	126	Irys
19	128	Irys
20	131	Irys
21	130	Irys
22	127	Irys
23	145	Ruta
24	149	Ruta
25	148	Ruta
26	152	Ruta
27	150	Ruta
28	145	Ruta

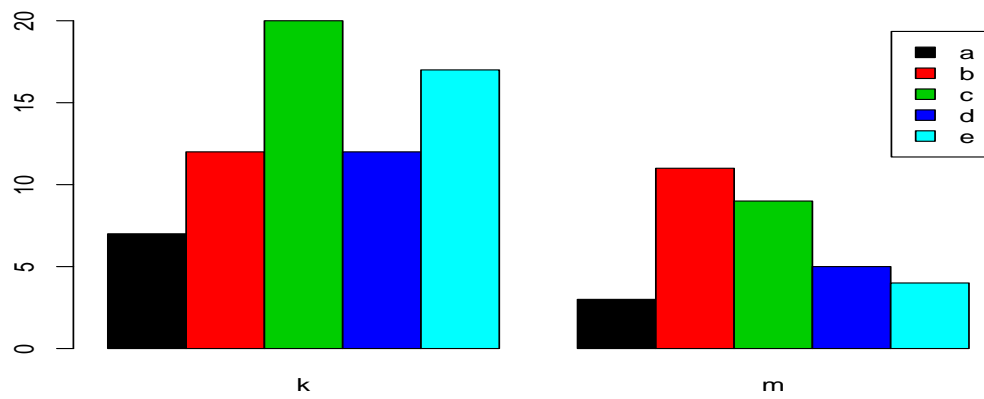
- (b) Zilustruj różnice w plenności tych odmian przy pomocy wykresu ramkowego. **Wskaźówka:** `boxplot()` **Odpowiedź:** patrz Rysunek 5

Zadanie 5. Napisz funkcję `wspolczynnik.zmiennosci()` obliczającą wartość współczynnika zmienności dla zadanego wektora obserwacji. Funkcja powinna posiadać dwa argumenty: `x` - wektor zawierający dane oraz `na.rm` - wartość logiczna (wartość domyślna to `FALSE`), która wskazuje, czy braki danych (obiekty `NA`) mają być ignorowane. Funkcja zwraca wartość współczynnika zmienności wyrażonego w procentach. Ponadto, funkcja ma sprawdzać, czy wektor obserwacji `x` jest wektorem liczbowym. W przeciwnym razie zwracany ma być błąd wraz z następującym komunikatem „argument nie jest wartością liczbową”. **Wskaźówka:** `is.numeric()`, `stop()`, `na.omit()` **Odpowiedź:** Przykładowe wywołania i wyniki funkcji:

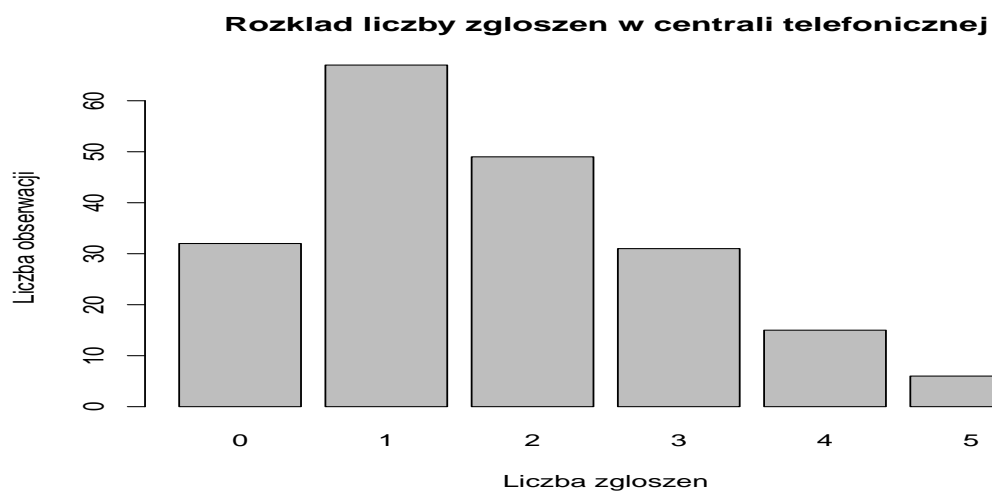
```
x <- c(1, NA, 3)
wspolczynnik.zmiennosci(x)
## [1] NA
wspolczynnik.zmiennosci(x, na.rm = TRUE)
## [1] 70.71068
wspolczynnik.zmiennosci()
## Error in wspolczynnik.zmiennosci() :
##   argument "x" is missing, with no default
wspolczynnik.zmiennosci(c("x", "y"))
## Error in wspolczynnik.zmiennosci(c("x", "y")) :
##   argument nie jest wartością liczbową
```



RYSUNEK 1. Wykres do Zadania 1 (c).



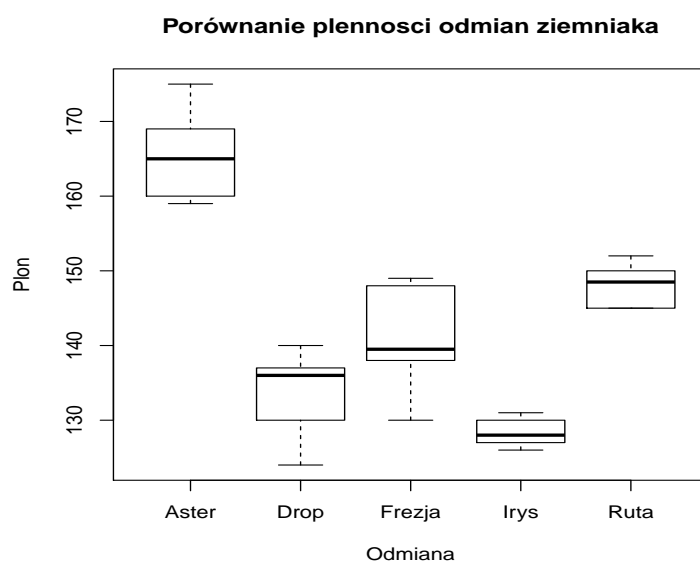
RYSUNEK 2. Wykres do Zadania 1 (f).



RYSUNEK 3. Wykres do Zadania 2 (b).



RYSUNEK 4. Wykres do Zadania 3 (b).



RYSUNEK 5. Wykres do Zadania 4 (b).