1. WSTĘPNA ANALIZA DANYCH

ZADANIE 1.1 Wczytać zbiór *pima* znajdujący się w bibliotece *faraway*.

```
> install.packages("faraway")     # instalujemy bibliotekę faraway
> library(faraway)                # uaktywniamy bibliotekę faraway
> data(pima)                      # uaktywniamy zbiór pima
```

(a) Obejrzeć opis zbioru *pima* by sprawdzić jakie informacje zawierają zmienne *test*, *pregnant* i *diastolic*.

(b) Wyznaczyć podstawowe miary liczbowe dla zmiennej *diastolic* i przyjrzeć się czy nie zawiera ona błędów i rzeczy nietypowych. Jeśli je zawiera, wprowadzić stosowne poprawki.

(c) Jakiego typu jest zmenna *test*? Czy została ona dobrze zapisana? Jeśli nie, to wprowadzić stosowne poprawki. Opisać po polsku poziomy tej zmiennej.

(d) Wyznaczyć i zinterpretować średnią, medianę, dolny i górny kwartyl, 1-szy decyl, rozstęp, rozstęp międzywartylowy oraz odchylenie standardowe dla zmiennej *diastolic*.

(e) Wyznaczyć średnie rozkurczowe ciśnienie krwi oraz jego odchylenie standardowe dla kobiet, u których zaobserwowano objawy cukrzycy.

(f) Dla zmiennej *pregnant* sporządzić i opisać wykres skrzynkowy.

(g) Dla zmiennej *test* sporządzić wykres słupkowy i wykres kołowy. Dodać do nich legendy. Wykresy umieścić w jednym oknie.

WSKAZÓWKA: Aby uzyskać liczności poszczególnych grup użyć funkcji `table(test)` lub `summary(test)`.

(h) Dla zmiennej *diastolic* sporządzić histogram częstości oraz narysować jądrowy estymator gęstości.

ZADANIE 1.2 Dane zawarte w pliku *gala_data.txt* zawierają informacje o kilkudziesięciu wyspach.

(a) Wczytać te dane. Uzyskać bezpośredni dostęp do zmiennych w tym zbiorze.

(b) Narysować histogram o trzech klasach dla danych opisujących powierzchnię (zmienna *Area*) 75% najmniejszych wysp. Podpisać osie i umieścić nagłówek.

(c) Narysować wykres skrzynkowy dla danych przedstawiających liczbę gatunków żółwi na wyspach (zmienna *Species*), których powierzchnia jest większa od 1 (km^2) i mniejsza od 25 (km^2).

ZADANIE 1.3 Poniższe dane zawierają wagę (w kg) oraz wzrost (w metrach) 6 mężczyzn:

waga: 60, 72, 57, 90, 95, 72

wzrost: $1,75$; $1,80$; $1,65$; $1,90$; $1,74$; $1,91$

(a) Wpisać powyższe dane. Obliczyć $bmi = waga/(wzrost^2)$ dla każdej z osób.

(b) Dla zmiennej *bmi* wyznaczyć średnią, medianę, dolny i górny kwartyl, wariancję, odchylenie standardowe i rozstęp międzykwartylowy.

(c) Sporządzić wykres skrzynkowy dla zmiennej *bmi*.

(d) Wypisać wzrosty mężczyzn, którzy mają wagę prawidłową. Przyjąć, że mężczyzna ma wagę prawidłową, gdy $20,7 \leqslant bmi < 26,5$.

ZADANIE 1.4 Zbiór *Cars93*, znajdujący się w bibliotece MASS, zawiera dane dotyczące różnych modeli samochodów osobowych.

(a) Wyjaśnić jakie informacje zostały podane w następujących kolumnach zbioru *Cars93*: *Min.Price, MPG.city, MPG.highway, Weight, Origin, Type*.

(b) Utworzyć nowe zmienne opisujące: zużycie paliwa (mierzone w litrach na 100 km) podczas jazdy samochodu w mieście, zużycie paliwa podczas jazdy samochodu na autostradzie, wagę samochodu

w kg oraz cenę wersji podstawowej modelu samochodu w tys. PLN. Przyjąć, że 1 mila to 1,6 km; 1 US gallon to 3,8 litra; 1 funt to 0,4536 kg. Sprawdzić bieżący kurs $ do PLN i użyć go do obliczeń.

(c) Wyznaczyć podstawowe statystyki próbkowe dla danych opisujących cenę w tys. PLN wersji podstawowej samochodu. Obliczyć kwantyl rzędu 0,95 dla tych danych i podać jego interpretację.

(d) Wypisać ceny wersji podstawowej samochodów, które były wyższe od kwantyla wyznaczonego w punkcie (c). Jakich modeli te ceny dotyczą?

(e) Sporządzić wykresy skrzynkowe dla zużycia benzyny podczas jazdy w mieście osobno dla samochodów amerykńskich i nieamerykańskich.

(f) Narysować histogram częstości dla danych dotyczących wagi samochodu. Nanieść na ten histogram jądrowy estymator gęstości.

(g) Narysować wykres słupkowy i kołowy dla zmiennej *Type*. Ile, spośród badanych samochodów, zaliczono do kategorii *sportowe*?

ZADANIE 1.5 Dla następujących danych:

(a) 5, 1, 0, −2, 3, 0, −1, 1, 2, 4;

(b) −1, 5, 1, 0, −2, 3, 0, −1, 1, 2, 4;

(c) 0, −1, 5, 1, 0, −2, 3, 0, −1, 1, 2, 4;

(d) 5, 0, −1, 5, 1, 0, −2, 3, 0, −1, 1, 2, 4
wyznaczyć ręcznie (tzn. bez użycia komputera) medianę oraz dolny i górny kwartyl. Uzyskane wyniki porównać z wartościami tych parametrów podawanymi przez R.