

Zadanie 1

Badania demograficzne przeprowadzone w 1988 roku w USA wykazały, że wśród kobiet (mających 18 i więcej lat) było: 17364 tyś. panien, 56128 tyś. mężatek, 11239 tyś. wdów i 8170 tyś. rozwódek.

- a) Utworzyć wykres kołowy dla stanu cywilnego danej grupy kobiet. Porównać różne formy opisu wykresu.

```
panie = c(17364, 56128, 11239, 8170) /// cos = c() - tworzenie wektora
stan = c("panny", "mężatki", "wdowy", "rozwodki")
```

```
pie(panie, labels = stan) /// wykres kołowy dla wektora panie
```

```
pie(panie, /// wektor z danymi
labels = stan, /// wektor z etykietami
col = kol, /// wektor z kolorami
main = "stan cywilny", /// tytuł wykresu
col.main = "violet", /// kolor tytułu
radius = 1) // promień koła(wykresu)
```

- b) Utworzyć wykres słupkowy dla stanu cywilnego danej grupy kobiet. Porównać różne rodzaje wykresów i formy ich opisu.

```
barplot(panie,
names.arg = panie, /// podpisy 'słupkow'
legend.text = stan, /// wektor z etykietami
col = kol)
```

Zadanie 2

Koncern paliwowy planuje otworzyć nową stację benzynową w pewnym mieście. Rozważane są cztery możliwe

lokalizacje stacji – w południowej, północnej, zachodniej i wschodniej dzielnicy miasta (oznaczenia dzielnic: S – południowa, N – północna, W – zachodnia i E – wschodnia). W ramach badania opinii społecznej odnośnie preferowanej lokalizacji stacji zapytano o to tysiąc kierowców. Ich odpowiedzi znajdują się w pliku **stacje.csv**. Utworzyć wykres słupkowy i wykres kołowy dla badanych preferencji.

```
stacja = read.csv('P:/smwd/stacje.csv')
dim(stacja) /// rozmiar
ts = table(stacja) /// tabelaryzacja
```

```
barplot(ts, names.arg=ts)
pie(ts)
```

Zadanie 3

Poniższe dane odpowiadają notowaniom pewnej spółki (w PLN) w kolejnych 20 dniach:

23,30 24,50 25,30 25,30 24,30 24,80 25,20 24,50 24,60 24,10, 24,30 26,10 23,10 25,50 22,60 24,60 24,30 25,40 25,20 26,80

Utworzyć wykres cen akcji jako funkcję czasu (szereg czasowy).

```
ceny=c(23.30,24.50,25.30,25.30,24.30,24.80,25.20,24.50,24.60,24.10,24.30,26.10,23.10,25.50,22.60,24.60,24.30,25.40,25.20,26.80)
```

```
plot(ceny,
type="b", /// typ – bez kropki, l - kreski, b – kropki i kreski
col = 3, pch = 16, /// kolor lini i styl kropki
xlab = "dzień",
ylab="ceny akcji" )
```

Zadanie 4

Wytrzymałość na ciśnienie wewnętrzne szkła butelek jest ich ważną charakterystyką jakościową. W celu zbadania wytrzymałości butelek umieszcza się je w maszynie hydrostatycznej, po czym zwiększa się ciśnienie

aż do zniszczenia butelki. Plik **butelki.csv** zawiera dane opisujące graniczną wytrzymałość na ciśnienie wewnętrzne szkła badanej partii butelek (mierzone w psi).

```
butelki = read.csv('P:/smwd/butelki.csv')
```

a) Utworzyć zmienną o nazwie *cisnienie*, opisującą wytrzymałość na ciśnienie wewnętrzne szkła butelek mierzone w MPa

Wskazówka:

1psi = 0,0068947 MPa

```
cis = 0.0068947*butelki$strength
```

b) Utworzyć histogram dla danych opisujących wytrzymałość butelek. Prześledzić wpływ liczby klas na kształt histogramu. Porównać różne rodzaje histogramów.

```
hist(cis, labels =T, prob = T)
```

breaks – liczba klas ale nie do końca bangła

```
h=hist(cis,labels=T)
```

h /// wyświetla info o histogramie

c) Utworzyć wykres łamanej licznosci i nałożyć go na wykres histogramu.

```
xll=c(1.1, h$mids, 2.7)
```

```
ylc=c(0,h$intensities,0)
```

```
lines (xll,yll, col=4)
```

```
====Licznosc skumulowana====
```

```
lk =numeric(length(7))
```

```
lk[1]= h$counts[1]
```

```
for (i in 2:7 ) lk [i]=h$counts[i]+lk[i-1]
```

```
lk2=c(0,lk)
```

```
barplot(lk2,space=0,names.arg=lk2)
```

c) Utworzyć wykres łodygowo-liściowy.

```
stem(cis)
```

d) Utworzyć i zinterpretować wykres skrzynkowy dla wytrzymałości butelek.

```
boxplot(cis,horizontal=T)
```

```
summary(cis)
```

f) Wyznaczyć i zinterpretować podstawowe statystyki próbkowe dla danych opisujących wytrzymałość butelek.

```
boxplot(cis,horizontal=T)$out /// outsider
```

```
mean(cis) /// srednia
```

```
median(cis) /// mediana
```

```
var(cis) /// wariancja
```

```
sd(cis) /// odchylenie standardowe
```

```
IQR(cis)
```

```
quantile(cis, 0.75) /// kwantyl rzędu 75% - 75% obserwacji ma mniejsza wartość
```

g) Obliczyć i zinterpretować 5, 10, 25, 50, 75, 90 i 95 percentyl dla rozważanych danych.

```
rz=c(.05,.1,.25,.75,.9,.95)
quantile(cis, rz)
```

h) Wyznaczyć 10% średnią uciętą dla danych opisujących wytrzymałość butelek. Porównać średnią uciętą ze średnią arytmetyczną i medianą. Prześledzić, jak zmienia się wartość średniej wraz ze zmianą stopnia ucięcia próbki.

```
mean(cis,trim=0.1)
```

Zadanie 6

W pliku **samochody.csv** zamieszczono dane dotyczące parametrów samochodów kilku wybranych marek.

```
samochody= read.csv2("P:/smwd/samochody.csv")
```

a) Zmienna *mpg* zawiera dane odpowiadające liczbie mil, przejechanych przez dany samochód na galonie paliwa. Utworzyć zmienną *zp* opisującą zużycie paliwa mierzone w litrach na 100 kilometrów.

Wskazówka:

1 mila = 1609 m

1 galon (amerykański) = 3,785 l

```
zp=378.5/(1.609*samochody$mpg)
samochody=cbind(samochody,zp) ///dodaje kolumnę zp
```

b) Utworzyć wykres łądogowo-liściowy dla zużycia paliwa.

```
stem(zp)
```

c) Utworzyć histogram dla danych opisujących zużycie paliwa.

```
hist(zp)
```

d) Utworzyć wykres skrzynkowy dla zużycia paliwa.

```
boxplot(zp,horizontal=T)
```

e) Wyznaczyć i zinterpretować podstawowe statystyki próbkowe dla danych opisujących zużycie paliwa (takie jak: **średnia**, mediana, wariancja, odchylenie standardowe, **rozstęp**, kwartyle, rozstęp międzykwartyłowy, **wartości ekstremalne**, **współczynnik asymetrii**, **kurtoza**, **współczynnik zmienności**).

```
mean(zp,na.rm=T) /// na.rm=T usuwa wartosci NA przed liczeniem
```

```
zp2=na.omit(zp) /// na.omit - pominięcie wartości NA
```

```
max(zp2)-min(zp2) /// rozstęp
```

```
range(zp2) /// wartości ekstremalne
```

```
mean((zp2-mean(zp2))^3)/sd(zp2)^3 /// współczynnik asymetrii
```

```
mean((zp2-mean(zp2))^4)/sd(zp2)^4-3 /// kurtoza
```

```
sd(zp2)/mean(zp2) /// współczynnik zmienności
```

f) Obliczyć i zinterpretować 5, 10, 90 i 95 percentyl dla rozważanych danych.

g) Wyznaczyć 5% średnią uciętą dla danych opisujących zużycie paliwa.

```
mean(cis,trim=0.05)
```

Zadanie 7

Analizowane w poprzednim zadaniu dane dotyczące zużycia paliwa postanowione poddać kategoryzacji tworząc następujące klasy:

Zużycie paliwa [litry na 100 km]	Kod opisujący kategorię zużycia paliwa
nie więcej niż 7	mało
więcej niż 7 ale nie więcej niż 10	średnio
więcej niż 10	dużo

Utworzyć wykres słupkowy dla wyznaczonych w ten sposób kategorii i wskazać, jaki procent badanych samochodów należy do każdej kategorii.

```
m=zp2[zp2<=7]
s=zp2[zp2>7 & zp2<=10 ]
d=zp2[zp2>10 ]

sp=c(length(m),length(s),length(d))

barplot(sp,names.arg=sp)
prop.table(sp) /// procent całości
round(prop.table(sp),2) /// zaokrąglone
```

Zadanie 8

Przeprowadzić wstępną analizę statystyczną danych dotyczących zużycia paliwa oddzielnie dla samochodów produkowanych w Europie, Ameryce i Japonii (wykorzystać zmienne producent i legenda). Zestawić wykresy skrzynkowe zużycia paliwa dla samochodów produkowanych w Europie, Ameryce i Japonii.

```
boxplot(zp~samochody$producent)

attach(samochody) // załącz

amer = subset(samochody,producent == 1) /// subset zwraca wektor spełniający warunek
eu = subset(samochody,producent == 2)
jap = subset(samochody,producent == 3)

mean(jap$zp)
mean(amer$zp)
mean(eu$zp, na.rm=T )

ale też

mean(zp[producent==3], na.rm=T)
```

Zadanie 9

Przeprowadzić wstępną analizę statystyczną danych dotyczących zużycia paliwa przez samochody o jednakowej liczbie cylindrów (dane dotyczące liczby cylindrów znajdują się w zmiennej cylindry).

```
median(zp[cylindry==6],na.rm=T)
summary(zp[cylindry==6],na.rm=T)
```

Zadanie 10

Przeprowadzić wstępną analizę statystyczną danych dotyczących zużycia paliwa przez samochody mające wyłącznie 6 lub 8 cylindrów.

```
median(zp[cylindry==6|cylindry==8],na.rm=T)
summary(zp[cylindry==6|cylindry==8],na.rm=T)
```

Zadanie 11

Porównać przyspieszenie samochodów produkowanych w Ameryce i Japonii (dane dotyczące przyspieszenia znajdują się w zmiennej przysp).

```
przyspa=c(przysp[producent==1])
przyspj=c(przysp[producent==3])
```

```
boxplot(przyspa, przyspj)
```

Zadanie 12

Przeprowadzić wstępną analizę statystyczną danych dotyczących zużycia paliwa wyłącznie dla samochodów ważących mniej niż 2500 funtów (wykorzystać zmienną `waga`).

```
zpw=zp[waga<2500]  
summary(zpw)
```

KOŁOKWIUM

Zadanie 1

Zbiór **Orange** zawiera dane dotyczące wieku (zmienna `age`) mierzonego w dniach i obwodu pnia (zmienna `circumference`) mierzonego w mm, 5 gatunków drzewek pomarańczowych (zmienna `Tree`)

Zbiór ten można otworzyć poleceniem `data(Orange)`

a) Podać wartość obwodu pnia której nie przekracza 70% badanych drzewek pomarańczowych

```
quantile(Orange$circumference, 0.7)
```

b) Podać średni wiek oraz odchylenie standardowe wieku drzewek pomarańczowych 1 gatunku

```
mean(Orange$age[Orange$Tree==1])  
sd(Orange$age[Orange$Tree==1])
```

c) Narysować i opisać wykres skrzynkowy dla obwodu pnia drzewek mających co najmniej 500 ale nie więcej niż 1300 dni.

```
drzewka= Orange$circumference [Orange$age >=500 & Orange$age <=1300 ]  
boxplot(drzewka ,horizontal=T)  
summary(drzewka)
```

Zadanie 2

Analizowaną w poprzednim zadaniu zmienną `circumference` (ze zbioru **Orange**), opisującą obwód pnia drzewek pomarańczowych poddać następującej kategoryzacji:

Obwód	Kategoria
Mniej niż 100mm	A
Co najmniej 100mm ale mniej niż 200mm	B
Co najmniej 200 mm	C

Utworzyć wykres kołowy dla wyznaczonych w ten sposób kategorii i wskazać, jaki procent drzewek pomarańczowych należy do danej kategorii

```
A = subset(Orange,Orange$circumference< 100)  
B = subset(Orange,Orange$circumference>= 100 & Orange$circumference < 200 )  
C = subset(Orange,Orange$circumference>= 200)
```

```
drzewkaZkategoryzowane = c(length(A$age),length(B$age),length(C$age))
```

```
pie(drzewkaZkategoryzowane, labels = drzewkaZkategoryzowane)  
prop.table(drzewkaZkategoryzowane)  
round(prop.table(drzewkaZkategoryzowane),2)
```