Praca domowa nr 2

Mateusz Markowski gr 1.2/4

Zadanie 1

Przygotowanie danych oraz obliczenie dla nich odpowiednio wielkości próbki, średniej wartości próbki, wariancji oraz odchylenia standardowego.

```
men=c(9.88, 9.97, 10.00, 10.02, 10.04, 10.06, 10.07, 10.08, 10.09, 10.09,
  10.11, 10.11, 10.12, 10.13, 10.13, 10.14, 10.16, 10.16, 10.19, 10.20, 10.21,
  10.22, 10.22, 10.22, 10.22, 10.23, 10.24, 10.26, 10.27, 10.28, 10.29, 10.30,
  10.30, 10.31, 10.31, 10.31, 10.34, 10.35, 10.35, 10.38, 10.40, 10.41, 10.48,
  10.53, 10.54, 10.55, 10.59, 10.65, 10.65, 10.69, 10.71, 10.84, 10.90, 10.94)
kob= c(10.83, 10.93, 10.94, 10.96, 10.97, 10.97, 10.99, 11.00, 11.01,
       11.04, 11.06, 11.07, 11.07, 11.07, 11.08, 11.12, 11.13, 11.14, 11.14, 11.18,
       11.19, 11.22, 11.22, 11.27, 11.28, 11.31, 11.32, 11.34, 11.35, 11.36, 11.38,
       11.39, 11.41, 11.41, 11.41, 11.41, 11.42, 11.43, 11.44, 11.46, 11.48, 11.49,
       11.52, 11.56, 11.59, 11.62, 11.63, 11.70, 11.81, 11.82, 11.86, 11.89, 11.90,
       11.92, 11.98, 12.06, 12.35)
length(men)
length(kob)
meanMen=mean(men)
meanKob=mean(kob)
menVar=var(men)
kobVar=var(kob)
menSd=sd(men)
kobSd=sd(kob)
```

Następnym krokiem było sprawdzenie czy dane posiadają rozkład normalny, w tym celu wykorzystano test Shapiro-Wilka.

```
shapiro.test(men) #badanie czy men posiada rozkład normalny
shapiro.test(kob) #badanie czy kob posiada rozkład normalny
```

Wyniki dla przeprowadzonych testów:

```
> shapiro.test(men) #badanie czy men posiada rozkład normalny

        Shapiro-Wilk normality test

data:  men
W = 0.93164, p-value = 0.004242
```

```
> shapiro.test(kob) #badanie czy kob posiada rozkład normalny

        Shapiro-Wilk normality test

data:  kob
W = 0.94886, p-value = 0.01746
```

Z otrzymanych wyników widać, że zarówno próbki mężczyzn jak i kobiet nie posiadają rozkładu normalnego, dlatego w celu sprawdzenia hipotezy o równości wariancji nie możemy posłużyć się testem var. Do potwierdzenia bądź odrzucenia hipotezy posłużymy się testem Lavene'a. Można wykorzystać też np.: test Bartletta. Przed przystąpieniem do testu utworzymy jeszcze tabelę wynikową dla danych zawierających m.in. p-value.

```
dane=rbind(c(length(kob),meanKob,kobVar,kobSd,shapiro.test(kob)$p.value)
            ,c(length(men),meanMen,menVar,menSd,shapiro.test(men)$p.value))
dane
rownames(dane)=c('Kobiety','Mężczyźni')
dane
colnames(dane)=c('n','mean','var','sd','p.value')
dane
```

```
            n     mean        var        sd     p.value
Kobiety    57 11.36667 0.11052976 0.3324602 0.017463297
Mężczyźni  54 10.30074 0.05687114 0.2384767 0.004242284
>
```

W celu przeprowadzenia testu Levene’a należy pogrupować dane oraz długość odpowiednich wektorów zastąpić dowolną zmienną. W naszym przypadku 1 oznaczono mężczyzn, a 2 kobiety.

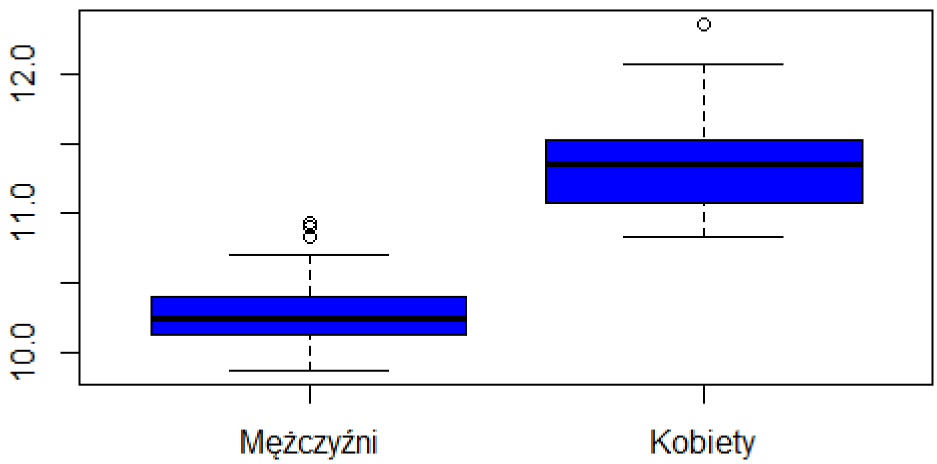
```
group=c(rep(1,length(men)),rep(2,length(kob)))

levene.test(c(men,kob),group)
```

Wyniki testu Levene’a:

```
        Modified robust Brown-Forsythe Levene-type test based on the absolute deviations
        from the median

data:  c(men, kob)
Test Statistic = 5.4319, p-value = 0.02161
```

Z otrzymanych wyników że wartość p-value<0.05 więc należy odrzucić hipotezę o równości wariancji w tych populacjach.

```
boxplot(men,kob,col='blue',names = c('Mężczyźni','Kobiety'))
```



Zadanie 2

Pierwszym krokiem jest zbadanie czy rozkłady w prywatnych i publicznych koledżach posiadają rozkład normalny. W tym celu wykorzystano test Shapiro-Wilka

```
publ=c(4.2, 6.1, 4.9, 8.5, 4.6, 9.1, 7.7, 6.5, 6.2, 10.2, 11.6, 10.4,
       5.0, 10.4, 8.1)
priv=c(13.0, 18.8, 13.2, 14.4, 17.7, 17.7, 17.6, 19.8, 16.8, 16.1)

shapiro.test(publ)
shapiro.test(priv)

var.test(publ,priv)

publSd=sd(publ)
privSd=sd(priv)
meanPubl=mean(publ)
meanPriv=mean(priv)
```

Wyniki testu Shapiro-Wilka dla publicznych danych p-value>0.05 czyli posiada rozkład normalny

```
> shapiro.test(publ)

        Shapiro-Wilk normality test

data:  publ
W = 0.93737, p-value = 0.3505
```

Wyniki testu Shapiro-Wilka dla prywatnych danych p-value>0.05 czyli posiada rozkład normalny.

```
> shapiro.test(priv)

        Shapiro-Wilk normality test

data:  priv
W = 0.92813, p-value = 0.4297
```

Z dwóch powyższych zrzutów ekranu widać, czemu zdecydowano się na wykonanie testu var.test do analizy wariancji. Wyniki var.test przedstawiono poniżej:

```
> var.test(publ,priv)

        F test to compare two variances

data:  publ and priv
F = 1.0936, num df = 14, denom df = 9, p-value = 0.9203
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2879463 3.5097118
sample estimates:
ratio of variances
          1.093607
```

Wartość p-value >0.05 więc możemy potwierdzić hipotezę o równości wariancji. Teraz przechodzimy do badania równości średnich. W tym celu zastosowano tsum.test, do którego potrzebujemy danych dotyczących odchylenia standardowego oraz średnich dla próbek prywatnych i publicznych.

Parametry ustawiono mu=10, ponieważ interesuje nas różnica równa lub większa10tys. W tym celu parametr alternative ustawiono na ‘greater.

```
publSd=sd(publ)
privSd=sd(priv)
meanPubl=mean(publ)
meanPriv=mean(priv)

tsum.test(meanPubl,publSd,length(publ)
          ,meanPriv,privSd,length(priv),mu = 10,alternative = 'greater')
```

Wyniki testu:

```
> tsum.test(meanPubl,publSd,length(publ)
+           ,meanPriv,privSd,length(priv),mu = 10,alternative = 'greater')

        Welch Two Sample t-test

data:  User input summarized values for x and y
t = -19.712, df = 20.054, p-value = 1
alternative hypothesis: true difference in means is greater than 10
95 percent confidence interval:
 -10.60055       Inf
sample estimates:
mean of x mean of y
 7.566667 16.510000
```

Wartość p-value=1 oznacza, że hipoteza postawiona w zadaniu została potwierdzona.

Zadanie 3

Przygotowanie danych oraz zbadanie czy dane posiadają rozkład normalny. Obliczenie różnicy pomiędzy danymi dotyczącymi zapylenia i nie zapylenia rośliny.

```
zapylona=c(0.78,0.76,0.43,0.92,0.86,0.59,0.68)
niezapylona=c(0.21,0.12,0.32,0.29,0.30,0.20,0.14)

var(zapylona)
var(niezapylona)
sd(zapylona)
sd(niezapylona)

shapiro.test(zapylona)
shapiro.test(niezapylona)

roznica=niezapylona-zapylona
```

Wyniki testu Shapiro-Wilka dla próbki zapylona przedstawiono poniżej. Jak widać wartość p-value>0.05

```
> shapiro.test(zapylona)

	Shapiro-Wilk normality test

data:  zapylona
W = 0.96211, p-value = 0.8366
```

Wyniki testu Shapiro-Wilka dla próbki nie zapylona przedstawiono poniżej. Jak widać wartość p-value>0.05

```
> shapiro.test(niezapylona)

	Shapiro-Wilk normality test

data:  niezapylona
W = 0.90921, p-value = 0.3904
```

Z otrzymanych wyników widać, że obie próbki przeszły próbę rozkładu normalnego więc do sprawdzenia badanej hipotezy posłużono się t.testem.

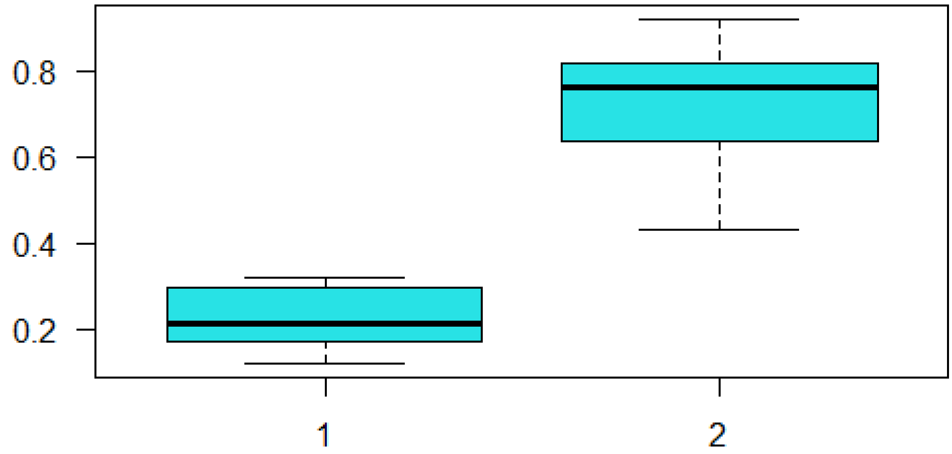
```
t.test(niezapylona,zapylona,alternative = 'greater',paired=TRUE,mu=0)
```

Wyniki przedstawiono poniżej:

```
> t.test(niezapylona,zapylona,alternative = 'greater',paired=TRUE,mu=0)

	Paired t-test

data:  niezapylona and zapylona
t = -6.9451, df = 6, p-value = 0.9998
alternative hypothesis: true mean difference is greater than 0
95 percent confidence interval:
 -0.6289256        Inf
sample estimates:
mean difference
     -0.4914286
```

Jak widać wartość p-value>0.05 stąd nie należy odrzucać hipotezy H0.Wyniki dla próbek zapylona oraz niezapylona przedstawiono za pomocą wykresu boxplot:

```
boxplot(list(przed,po),col=5,las=1)
```



Test Shapiro-Wilka dla danych roznicy=niezapylona-zapylona:

```
shapiro.test(roznica)
```

```
> shapiro.test(roznica)

        Shapiro-Wilk normality test

data:  roznica
W = 0.78853, p-value = 0.03149
```

Wartość p-value>0.05. Po przeprowadzeniu t.test-u dla danych roznicy wyniki przedstawiono na poniższym zrzucie ekranowym:

```
> t.test(roznica,mu=0)

        One Sample t-test

data:  roznica
t = -6.9451, df = 6, p-value = 0.0004419
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.6645690 -0.3182881
sample estimates:
 mean of x
-0.4914286
```

Wyniki dla pairwise.t.test przedstawiono poniżej:

```
pairwise.t.test(niezapylona,zapylona,alternative = 'greater')
```

```
data:  niezapylona and zapylona

     0.43 0.59 0.68 0.76 0.78 0.86
0.59 -    -    -    -    -    -
0.68 -    -    -    -    -    -
0.76 -    -    -    -    -    -
0.78 -    -    -    -    -    -
0.86 -    -    -    -    -    -
0.92 -    -    -    -    -    -

P value adjustment method: holm
```

Zadanie 4

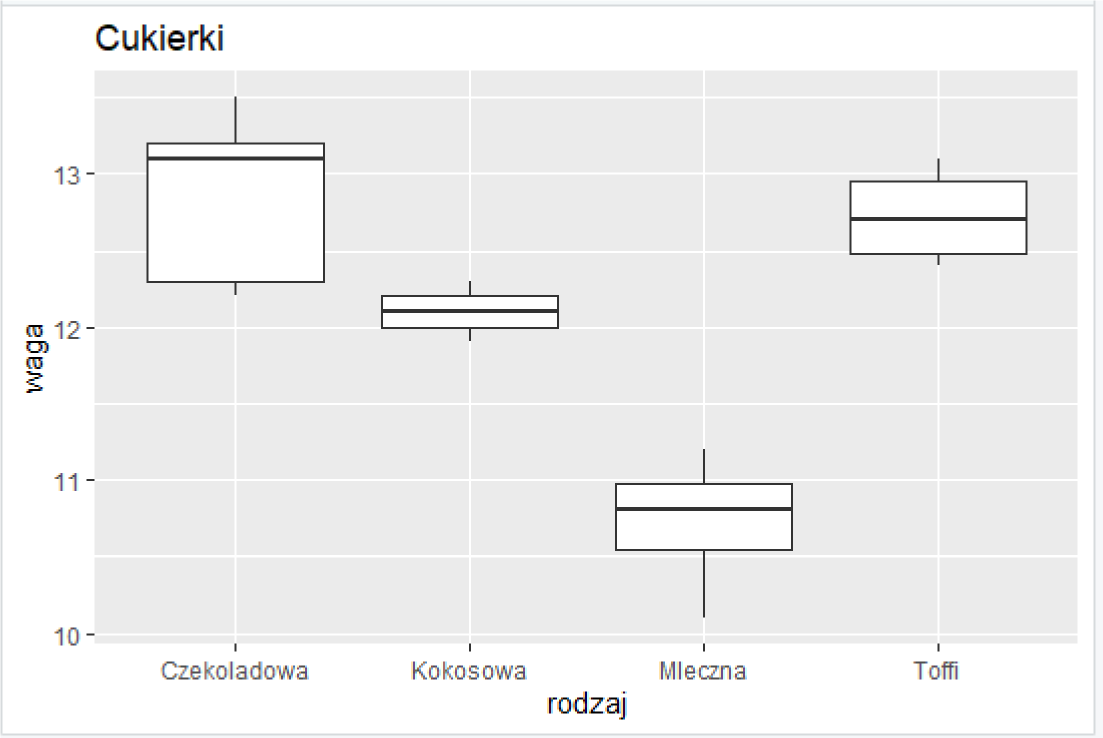
Przygotowanie danych oraz utworzenie wektora z danymi

```
cuk1=c(10.1,10.9,11.2,10.7)
cuk2=c(12.1,11.9,12.3)
cuk3=c(12.3,12.2,13.2,13.5,13.1)
cuk4=c(12.4,12.5,12.9,13.1)

waga=c(cuk1,cuk2,cuk3,cuk4)
```

Podpisanie danych odpowiednim rodzajem cukierków oraz utworzenie ramki danych zawierającej wagi i nazwy cukierków. Przedstawienie danych w formie graficznej w postaci qplot.

```
rodzaj=rep(c("Mleczna","Kokosowa","Czekoladowa","Toffi")
           ,c(length(cuk1),length(cuk2),length(cuk3),length(cuk4)))

ramka=data.frame(waga,rodzaj)
```

```
qplot(rodzaj,waga,data=ramka,geom="boxplot",main="Cukierki")
```



Wykonanie obliczeń średniej oraz wariancji dla każdego rodzaju cukierków:

```
meanMleczna=mean(cuk1)
meanKokos=mean(cuk2)
meanCzekolada=mean(cuk3)
meanToffi=mean(cuk4)

varMleczna=var(cuk1)
varKokos=var(cuk2)
varCzekolada=var(cuk3)
varToffi=var(cuk4)
```

Zbadanie każdego z czterech rodzajów cukierków testem Shapiro-Wilka w celu sprawdzenia czy mamy doczynienia z rozkładem normalnym.

```
shapiroMleczna=shapiro.test(cuk1)
shapiroKokos=shapiro.test(cuk2)
shapiroCzekolada=shapiro.test(cuk3)
shapiroToffi=shapiro.test(cuk4)
```

Wyniki testu Shapiro-Wilka:

```
> shapiroMleczna

        Shapiro-Wilk normality test

data:  cuk1
W = 0.96205, p-value = 0.7918
```

```
> shapiroKokos

        Shapiro-Wilk normality test

data:  cuk2
W = 1, p-value = 1
```

```
> shapiroCzekolada

        Shapiro-Wilk normality test

data:  cuk3
W = 0.87771, p-value = 0.2991
```

```
> shapiroToffi

        Shapiro-Wilk normality test

data:  cuk4
W = 0.91569, p-value = 0.513
```

Dla każdego z przedstawionych powyżej testów Shapiro-Wilka widać, że wartość p-value>0.05 co oznacza, że nie możemy odrzucić hipotezy zerowej, że próbki posiadają rozkład normalny.

Przedstawienie wyników w tabeli.

```
dane=rbind(c(length(cuk1),meanMleczna,varMleczna,shapiroMleczna$p.value),
           c(length(cuk2),meanKokos,varKokos,shapiroKokos$p.value),
           c(length(cuk3),meanCzekolada,varCzekolada,shapiroCzekolada$p.value),
           c(length(cuk4),meanToffi,varToffi,shapiroToffi$p.value))
rownames(dane)=c("Mleczna","Kokosowa","Czekoladowa","Toffi")
dane
colnames(dane)=c("n","mean","var","p.value")
```

```
> dane
            n   mean       var   p.value
Mleczna     4 10.725 0.2158333 0.7918111
Kokosowa    3 12.100 0.0400000 1.0000000
Czekoladowa 5 12.860 0.3330000 0.2990715
Toffi       4 12.725 0.1091667 0.5130401
```

b)

W celu weryfikacji hipotezy o równości wariancji wag wszystkich cukierków należy wykonać test Bartletta. W tym celu grupujemy dane w liste. Następnie na tej liście wykonujemy test Bartletta.

```
lista=list(cuk1,cuk2,cuk3,cuk4)
```

```
bartlett.test(lista)
```

Wyniki testu Bartletta:

```
> bartlett.test(lista)

        Bartlett test of homogeneity of variances

data:  lista
Bartlett's K-squared = 2.3535, df = 3, p-value = 0.5024
```

Z otrzymanych wyników odczytać możemy, że nie możemy odrzucić hipotezy H0 o równości wariancji wszystkich cukierków, ponieważ wartość p-value>0.05.