

# **Test 30.05.2022**

Jindřich Kvita

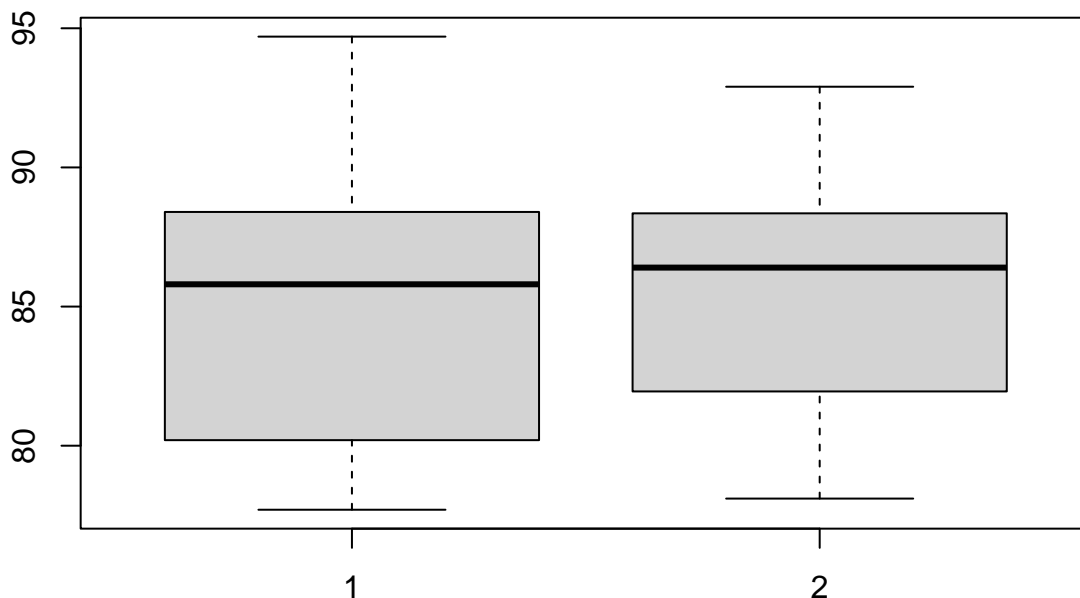
# Zadání B

## 4. Úloha

Analýzujte počáteční hmotnost probandů zařazených do skupin KETO+HIIT a KONTROL. Obsahují-li data odlehlá pozorování, do analýzy je nezařazujte. Nezapomeňte na ověření předpokladů pro použití metod statistické indukce.

- a) Pro probandy ze skupiny KETO+HIIT a skupiny KONTROL určete bodové a 95% oboustranné intervalové odhady střední hodnoty, popř. mediánu, počáteční hmotnosti. Výsledek pro probandy ze skupiny KETO+HIIT interpretujte. (3b)

```
data_zacatek_op <-  
  data %>%  
    group_by(skupina) %>%  
    identify_outliers(m0)  
  
data_bez_op <- data %>%  
  filter(!(ID %in% data_zacatek_op$ID))  
  
boxplot(data_bez_op$m0[data_bez_op$skupina=="KETO+HIIT"],  
        data_bez_op$m0[data_bez_op$skupina=="KONTROL"] )
```



```
zacatek_shapiro <- tapply(data_bez_op$m0, data_bez_op$skupina, shapiro.test)

bodovKh<-mean(data_bez_op$m0[data_bez_op$skupina == "KETO+HIIT"], na.rm=T)
bodovKo<-mean(data_bez_op$m0[data_bez_op$skupina == "KONTROL"], na.rm=T)
khttest<-t.test(data_bez_op$m0[data_bez_op$skupina == "KETO+HIIT"],
  conf.level = 0.95,
  alternative = "two.sided")
kottest<-t.test(data_bez_op$m0[data_bez_op$skupina == "KONTROL"],
  conf.level = 0.95,
  alternative = "two.sided")
```

Skupina	Shapiro-Wilkův test (p-hodnota)
KETO+HIIT	0.200
KONTROL	0.458

Na základě Shapiro-Wilkova testu na hladině významnosti 5 % nezamítáme nulovou hypotézu o normálním rozdělení datových souborů.

Bodovým odhadem počáteční hmotnosti pro KETO+HIIT je 85.1533333Kg a pro KONTROL 85.415 Kg. #todo: chybí jednotky!!, přidat do tabulky průměr / bodový odhad střední hodnoty (podle toho, co bude zrovna lepší)

95% oboustranným intervalovým odhadem je:

Skupina	95% Intervalový odhad středních hodnot
KETO+HIIT	(83.439; 86.868)
KONTROL	(83.313; 87.517)

Průměrná hmotnost ze skupiny keto+hiit je 85.2 Kg, 95% IO je (83.439; 86.868)Kg.

- b) Pro probandy ze skupiny KETO+HIIT určete, zda se pozorovaná počáteční hmotnost, popř. medián počáteční hmotnosti, statisticky významně liší od 83,5 kg. K ověření využijte intervalový odhad i příslušný čistý test významnosti. (3b)

```
bttest<-t.test(data_bez_op$m0[data_bez_op$skupina=="KETO+HIIT"],
               conf.level = 0.95,
               mu=83.5,
               alternative = "two.sided")
```

Na hladině významnosti 95 % se dle intervalového odhadu ((83.44; 86.87)) váha probandů KETO+HIIT statisticky významně neliší od 83.5kg. Stejně potvrzuje i p-hodnota čistého testu (0.058).

- c) Na hladině významnosti 5 % určete, zda se střední hodnota, popř. medián, počáteční hmotnosti pro probandy ze skupiny KETO+HIIT a pro probandy ze skupiny KONTROL statisticky významně liší. Pro ověření použijte příslušný intervalový odhad. (4b)

```
cvtest<-var.test(data_bez_op$m0[data_bez_op$skupina == "KETO+HIIT"],
                 data_bez_op$m0[data_bez_op$skupina == "KONTROL"],
                 ratiom=1, alternative = "two.sided", conf.level = 0.95)
cttest<- t.test(data_bez_op$m0[data_bez_op$skupina == "KETO+HIIT"],
                 data_bez_op$m0[data_bez_op$skupina == "KONTROL"],
                 conf.level = 0.95, alternative = "two.sided", var.equal = T
                 )
```

Na základě testu shody rozptýlů s p-hodnotou 0.940 na hladině významnosti 95 % nezamítáme nulovou hypotézu. Test intervalového odhadu ukazuje na hladině významnosti 95 %, s p-hodnotou 0.843, že se probandi KETO+HIIT a KONTROL statisticky od sebe neliší.

## 5. úloha

Na hladině významnosti 5 % rozhodněte, zda existuje statisticky významný rozdíl mezi středními hodnotami, popř. mediány, koncové hmotnosti probandů v rámci skupin, do nichž byli probandi zařazeni. Pokud ano, zjistěte, zda lze některé skupiny probandů z hlediska střední hodnoty, popř. mediánu, koncové hmotnosti označit za homogenní a skupiny probandů případně dle sledovaného parametru a výsledků post-hoc analýzy seřadte sestupně. Obsahují-li data odlehlá pozorování, do analýzy je nezařazujte. Nezapomeňte ověřit předpoklady pro použití zvoleného testu. Poznámka: V tomto příkladu do srovnání zařad'te všechny analyzované skupiny probandů (KETO, HIIT, KETO+HIIT, KONTROL). (10b)

```
# nalezení odlehlých pozorování
data_m1_op <-
  data %>%
    group_by(skupina) %>%
    identify_outliers(m1)

# odfiltrování odlehlých hodnot
data_bez_op <- data %>%
  filter(!(ID %in% data_m1_op$ID))

# test normálního rozdělení
shapiro <- tapply(data_bez_op$m1, data_bez_op$skupina, shapiro.test)

# test shody rozptylů
bartlet<- bartlett.test(data_bez_op$m1 ~ data_bez_op$skupina)
# Dle výsledku Bartlettova testu na hladině významnosti 5 % zamítáme shodu rozptylů.
# Kvůli zamítnutí shody rozptylů bude použit Kruskal-Wallisův test pro shodu mediánů

# shoda rozptylů: NE ----> Kruska-Wallis
kruskal.test(data_bez_op$m1 ~ data_bez_op$skupina)

##
## Kruskal-Wallis rank sum test
##
## data: data_bez_op$m1 by data_bez_op$skupina
## Kruskal-Wallis chi-squared = 54.245, df = 3, p-value = 9.947e-12

# VÝSLEDEK: zamítáme hypotézu o shodě MEDIÁNŮ
# Na základě p hodnoty menší, než 0.05 zamítáme na hladině významnosti 5% shodu o rozpty
```

```
# seřazení podle mediánu
data_bez_op %>%
  group_by(skupina) %>%
  summarize(median = median(m1)) %>%
  arrange(desc(median))
```

```
## # A tibble: 4 x 2
##   skupina   median
##   <chr>     <dbl>
## 1 KONTROL    86.6
## 2 HIIT       80.2
## 3 KETO       76.6
## 4 KETO+HIIT  75.6
```

```
# VÝSLEDEK:
# 1. KONTROL
# 2. HIIT
# 3. KETO
# 4. KETO+HIIT
# Na základě
# analýza toho, které se vzájemně liší
dun <- dunnTest(m1 ~ skupina, data = data_bez_op, method = "bonferroni")
# VÝSLEDEK:
# KETO - KETO+HIIT stejné, všechny ostatní nehomogenní

# CELKOVÝ VÝSLEDEK:
# VÝSLEDEK:
# 1. KONTROL
# 2. HIIT
# 3. KETO/KETO+HIIT (homogenní)
```

skupina	p-hodnota
HIIT	0.489
KETO	0.244
KETO+HIIT	0.114
KONTROL	0.160

Jelikož jsou P-hodnoty všech skupin větší než 0.05, nelze zamítnout nulovou hypotézu a lze rozdělení všech skupin modelovat pomocí normálního rozdělení.

-	KONTROL	HIIT	KETO	KETO+HIIT
KONTROL	-	<0.001	<0.001	<0.001
HIIT	-	-	<0.001	<0.001
KETO	-	-	-	1.000
KETO+HIIT	-	-	-	-

Dle Dunnové post-hoc analýzy jee najevo, že pouze skupiny KETO a KETO+HIIT se statisticky významně neliší. Po seřazení dle velikosti mediánů hodnotím sestupně výsledky skupin takto:

1. KONTROL
2. HIIT
3. KETO/KETO+HIIT (homogenní)

# Zadání A

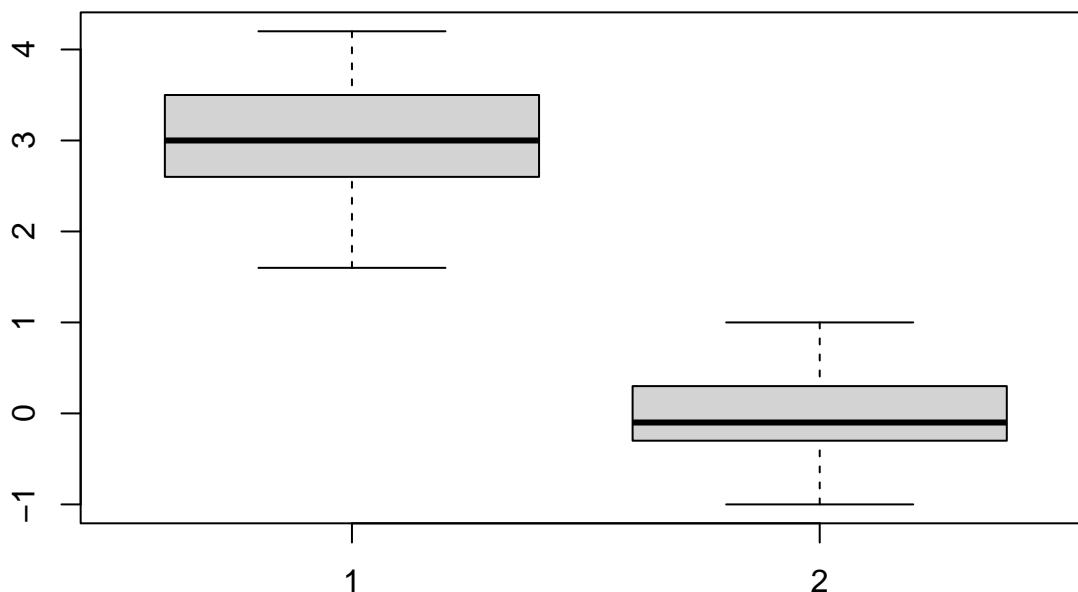
## 4. úloha

Analyzujte pokles BMI (tj. BMI.0-BMI.1) probandů zařazených do skupin KETO a KONTROL. Obsahují-li data odlehlá pozorování, do analýzy je nezařazujte. Nezapomeňte na ověření předpokladů pro použití metod statistické indukce.

- a) Pro probandy ze skupiny KETO a skupiny KONTROL určete bodové a 95% levostranné intervalové odhady střední hodnoty, popř. mediánu, poklesu BMI. Výsledek pro probandy ze skupiny KETO interpretujte. (3b)

```
a_zacatek_op <-  
  dataA %>%  
    group_by(skupina) %>%  
    identify_outliers(zlepseni)  
  
a_data_bez_op <- dataA %>%  
  filter(!(ID %in% a_zacatek_op$ID))  
  
boxplot(a_data_bez_op$zlepseni[a_data_bez_op$skupina=="KETO"],  
        a_data_bez_op$zlepseni[a_data_bez_op$skupina=="KONTROL"] )
```





```
a_shapiro <- tapply(a_data_bez_op$zlepseni, a_data_bez_op$skupina, shapiro.test)

a_bodovKe<-mean(a_data_bez_op$zlepseni[a_data_bez_op$skupina == "KETO"], na.rm=T)
a_bodovKo<-mean(a_data_bez_op$zlepseni[a_data_bez_op$skupina == "KONTROL"], na.rm=T)
a_kettest<-t.test(a_data_bez_op$zlepseni[a_data_bez_op$skupina == "KETO"],
                  conf.level = 0.95,
                  alternative = "greater")
a_kottest<-t.test(a_data_bez_op$zlepseni[a_data_bez_op$skupina == "KONTROL"],
                  conf.level = 0.95,
                  alternative = "greater")
```

skupina	p-hodnota
HIIT	0.072
KETO	0.123
KETO+HIIT	0.086
KONTROL	0.295

Na základě Shapiro Wilkova testu na hladině významnosti 5% nezamítáme normalitu Bodovým odhadem pro KETO je 3.0510204 a pro KONTROL -0.0066667.

95% oboustranným intervalovým odhadem je:

Skupina	95% Intervalový odhad středních hodnot
KETO	(2.911; Inf)
KONTROL	(-0.106; Inf)

- b) Pro probandy ze skupiny KETO určete, zda je pozorovaný průměrný pokles BMI, popř. medián poklesu BMI, statisticky významný. K ověření využijte intervalový odhad i příslušný čistý test významnosti. (3b)

```
a_bttest<-t.test(a_data_bez_op$zlepseni[a_data_bez_op$skupina=="KETO"],
  conf.level = 0.95,
  mu=0,
  alternative = "two.sided")
```

Na hladině významnosti 95 % se dle intervalového odhadu ((2.88; 3.22)) pokles váhy probandů KETO není statisticky významný. Stejně potvrzuje i p-hodnota čistého testu (0.000).

(PS. zde je na zvážení, jestli nepoužít jen jednostranný, když se řeší pokles, pak by to bylo greather místo two.sided

- c) Na hladině významnosti 5 % určete, zda se střední hodnota, popř. medián, poklesu BMI pro probandy ze skupin KETO a KONTROL statisticky významně liší. Pro ověření použijte příslušný čistý test významnosti. (4b)

```
a_cvtest<-var.test(a_data_bez_op$zlepseni[a_data_bez_op$skupina == "KETO"],
  a_data_bez_op$zlepseni[a_data_bez_op$skupina == "KONTROL"],
  ration=1, alternative = "two.sided", conf.level = 0.95)
a_cttest<- t.test(a_data_bez_op$zlepseni[a_data_bez_op$skupina == "KETO"],
  a_data_bez_op$zlepseni[a_data_bez_op$skupina == "KONTROL"],
  conf.level = 0.95, alternative = "two.sided", var.equal = T
)
```

Na základě testu shody rozptýlů s p-hodnotou 0.074 na hladině významnosti 95 % nezamítáme nulovou hypotézu. na základě čistého testu významnosti a z něj vyplývající p hodnoty 0.000 zamítáme nulovou hypotézu a skupiny se tak významně liší.

Dle

## 5. úloha

Definujte novou dichotomickou proměnnou (ZMENA.kat, varianty proměnné: ANO, NE), která bude definovat, zda u probanda došlo během tří měsíců k poklesu BMI o více než 2,5 kg/m<sup>2</sup>. Následně ověřte, zda existuje statisticky významný vliv typu změny chování probandů na pokles BMI o více než 2,5 kg/m<sup>2</sup>, tj. srovnajte přítomnost poklesu BMI o více než 2,5 kg/m<sup>2</sup> u probandů ze skupiny KETO a skupiny HIIT. Nezapomeňte na ověření předpokladů pro použití metod statistické indukce.

```
dataA <- dataA %>%
  mutate(ZMENA.kat = ifelse(zlepseni > 2.5, "ANO", "NE")) %>%
  select(ID, skupina, zlepсени, "ZMENA.kat")

#da se preskocit, je o ulohu vys
a_zacatek_op <-
  dataA %>%
    group_by(skupina) %>%
    identify_outliers(zlepseni)

a_data_bez_op <- dataA %>%
  filter(!(ID %in% a_zacatek_op$ID))
#koniec preskoku

dataVyhozene <- dataA %>%
  filter(!(ID %in% a_data_bez_op$ID)) %>%
  filter(skupina %in% c("KETO", "HIIT"))
dataVyhozene

## # A tibble: 2 x 4
##       ID skupina zlepсени ZMENA.kat
##   <dbl> <chr>      <dbl> <chr>
## 1   101 KETO        5.2  ANO
## 2   123 HIIT        3.70 ANO

asoc_tabulka <- a_data_bez_op %>%
  group_by(skupina) %>%
  summarise(celkem = length(ID),
            pokles_ano = length(ZMENA.kat[ZMENA.kat == "ANO"]),
            pokles_ne = length(ZMENA.kat[ZMENA.kat == "NE"]),
            pokles_ano_procent = pokles_ano / celkem * 100,
            pokles_ne_procent = pokles_ne / celkem * 100,
            pokles_ano_fmt = sprintf("%d (%0.1f %%)", pokles_ano, pokles_ano_procent),
```

```

        pokles_ne_fmt = sprintf("%d (%0.1f %%)", pokles_ne, pokles_ne_procent)
    )
asoc_tabulka_celkem <- asoc_tabulka %>%
  summarise(pokles_ano_celkem = sum(pokles_ano[1], pokles_ano[2]),
            pokles_ne_celkem = sum(pokles_ne[1], pokles_ne[2]),
            pokles_ano_procent_celkem = pokles_ano_celkem / sum(celkem) * 100,
            pokles_ne_procent_celkem = pokles_ne_celkem / sum(celkem) * 100,
            pokles_ano_celkem_fmt = sprintf("%d (%0.1f %%)", pokles_ano_celkem, pokles_ano_procent_celkem),
            pokles_ne_celkem_fmt = sprintf("%d (%0.1f %%)", pokles_ne_celkem, pokles_ne_procent_celkem)
  )

```

- a) Uveďte asociační tabulku vhodnou pro analýzu závislosti neexistence poklesu BMI o více než 2,5 kg/m<sup>2</sup> na skupině (KETO vs. HIIT), do níž byli probandi zařazeni. Tabulku rozšířte o řádkové relativní četnosti. (2b)

skupina/četnosti	neexistence poklesu (NE)	neexistence poklesu (ANO)	celkem
HIIT	50 (84.7 %)	9 (15.3 %)	59
KETO	12 (24.5 %)	37 (75.5 %)	49
celkem	87 (39.9 %)	46 (21.1 %)	218

- b) Načrtněte graf pro vizualizaci dané závislosti (korespondující s tabulkou uvedenou v bodě a)) a na základě asociační tabulky, daného grafu a vhodné míry kontingence interpretujte závěry, k nimž jste ohledně sledované závislosti došli. (2b)

```

# lab <- factor(c(asoc_tabulka$skupina[1], asoc_tabulka$skupina[2]))
# levels(lab)
#
# a_data_bez_op$ZMENA.kat = factor(a_data_bez_op$ZMENA.kat)

# kont.tab<-matrix(c(asoc_tabulka$pokles_ne_fmt[1], asoc_tabulka$pokles_ano_fmt[1],
#                    asoc_tabulka$pokles_ne_fmt[2], asoc_tabulka$pokles_ano_fmt[2]),
#                  byrow = T, nrow = 2)
# kont.tab <- as.table(kont.tab)
# rownames(kont.tab)<- c(asoc_tabulka$skupina[1], asoc_tabulka$skupina[2])
# colnames(kont.tab)<- c("neexistence poklesu (ANO)", "neexistence poklesu (NE)")
# kont.tab <- as.table(kont.tab)
#
# kont.tab

```

```
# kont.tab.tmp=as.data.frame(kont.tab)
# mosaicplot(kont.tab,
#           las = 1,
#           color = gray.colors(5))
```

- c) Určete bodový a 95% oboustranný intervalový odhad rizika, že po 3 měsících užívání KETO diety nedojde k poklesu BMI o více než 2,5 kg/m<sup>2</sup>. Výsledek interpretujte. (3b)
- d) Na hladině významnosti 5 % rozhodněte, zda je pozorovaná závislost statisticky významná. Pro ověření výzkumné hypotézy použijte odhad příslušného relativního rizika. (Doplňte interpretaci bodového a 95% intervalového odhadu relativního rizika.) (3b)