Cvičení z matematické statistiky **07**

Poslední úprava dokumentu: 10. dubna 2024.

Dvouvýběrový t-test a Wilcoxonův test, ANOVA

Ze složky `V:/turcicova` si stáhněte soubor `znamky.txt`. Tento datový soubor obsahuje údaje (hodnoty oddělené mezerou) o prospěchu v 7. a 8. třídě u 181 dětí (93 dívek a 88 chlapců) a hodnotě jejich IQ. Najdeme zde následující proměnné:

`ZAK`	identifikační číslo žáka
`SKOLA`	identifikační číslo školy
`POHLAVI`	pohlaví žáka (0 - dívka, 1 - chlapec)
`IQ`	hodnota IQ;
`ZN7`	průměrná známka na vysvědčení v pololetí 7. třídy;
`ZN7K`	průměrná známka na vysvědčení na konci 7. třídy;
`ZN8`	průměrná známka na vysvědčení v pololetí 8. třídy;

1) Spusťte si RStudio, nastavte pracovní adresář, otevřete si nový skript a uložte si jej. Dále si načtěte datovou tabulku `znamky.txt`.

```
zaci <- read.table("znamky.txt", header = TRUE, sep = " ", dec = ".")
```

2) Podívejte se, které veličiny jsou kategoriální, a případně změňte jejich formát na `factor`:

```
zaci$POHLAVI <- as.factor(zaci$POHLAVI)
zaci$SKOLA <- as.factor(zaci$SKOLA)
```

3) Zajistěte si přímý přístup k datům a data uložte ve formátu `RData`.

```
save(zaci, file = "zaci.RData")
attach(zaci)
```

4) Pomocí vhodného obrázku prozkoumejte, jak spolu souvisí známka z konce 7. třídy a z pololetí 8. třídy. Příslušný vztah popiště také číselně pomocí korelačního koeficientu.

1 Dvouvýběrový t-test

Nyní se zaměříme na testy sloužící k porovnání středních hodnot dvou populací, ze kterých máme k dispozici výběry. Prostudujte si prosím příslušnou teorii v textu *Dvouvýběrové testy* (dále jako *DVT*), který je k dispozici na mých stránkách.

Zajímáme se o porovnání IQ chlapců a dívek. Označme si IQ náhodně vybrané dívky jako X a předpokládejme, že $X \sim N(\mu_x, \sigma_y^2)$. Analogicky označme jako Y IQ náhodně vybraného chlapce a předpokládejme $Y \sim N(\mu_y, \sigma_y^2)$.

1) Pomocí popisných statistik si udělejte představu o tm, zda se typické IQ chlapců a dívek liší.

2) Vykreslete si vhodný ilustrativní obrázek.

```
boxplot(IQ~POHLAVI)
```

Nakreslete graf průměrů (plot of means) s „anténami" odvozenými ze směrodatných chyb průměrů a dále z intervalů spolehlivosti pro střední hodnotu.

```
install.packages("RcmdrMisc")  # nainstaluje knihovnu
library(RcmdrMisc)        # otevre knihovnu
plotMeans(IQ, POHLAVI, error.bars = "conf.int", level=0.95)
plotMeans(IQ, POHLAVI, error.bars = "se")
```

3) Ověřte, zda lze předpokládat, že IQ chlapců a dívek jsou nezávislé výběry z normálního rozdělení se shodných rozptylem.

```
qqnorm(IQ[POHLAVI == 0])
qqline(IQ[POHLAVI == 0])
qqnorm(IQ[POHLAVI == 1])
qqline(IQ[POHLAVI == 1])
shapiro.test(IQ[POHLAVI == 0])   # test normality 1. vyberu
shapiro.test(IQ[POHLAVI == 1])   # test normality 2. vyberu
```

Otestujme nyní pomocí F-testu hypotézu $\mathrm{H}_0 : \sigma_x^2 = \sigma_y^2$ proti $\mathrm{H}_1 : \sigma_x^2 \neq \sigma_y^2$.

```
var.test(IQ~POHLAVI)    # nebo
var.test(IQ[POHLAVI == 0], IQ[POHLAVI == 1])
```

✧ Podporují tato data hypotézu o shodných rozptylech?
P-hodnota vychází 0.2663, takže shodu rozptylů nezamítáme.

✧ Co udává `95 percent confidence interval` ve výstupu?
Je to 95% interval spolehlivosti pro podíl $\frac{\sigma_x^2}{\sigma_y^2}$. Jelikož tento interval obsahuje jedničku (což je teoretická hodnota tohoto podílu za platnosti H_0), tak opět docházíme k závěru, že shodu rozptylů nelze zamítnout.

4) Otestujte, zda je typické IQ chlapců stejné jako typické IQ dívek.

- ✧ Uveďte uvažovaný model. Jsou splněny předpoklady tohoto testu?
- ✧ Formulujte nulovou a alternativní hypotézu.
  ✧Chceme testovat $H_0 : \mu_x = \mu_y$ proti $H_1 : \mu_x \neq \mu_y$.
- ✧ Proveďte dvouvýběrový t-test.

```
t.test(IQ~POHLAVI, var.equal = TRUE)     # prvni moznost
t.test(IQ[POHLAVI == 0], IQ[POHLAVI == 1], var.equal = TRUE)    # druha moznost
```

  Argumentem `var.equal=TRUE` říkáme, že jsme otestovali shodu rozptylů a došli jsme k závěru, že shodné rozptyly lze předpokládat.
- ✧ Jaký je závěr (zamítáme/nezamítáme H_0 na požadované hladině)?
  Nulovou hypotézu zamítáme, protože p-hodnota je menší než zvolená hladina $\alpha = 0.05$.
- ✧ Jak byste formulovali svůj závěr bez použití výrazů „zamítáme/nezamítáme H_0"?
  „Na 5% hladině jsme prokázali, že střední hmotnost dívek a chlapců v populaci není stejná."
- ✧ Jak souvisí výsledek testu s intervalem spolehlivosti, jež je též uveden ve výstupu?
  95% interval spolehlivosti (-8.151, 0.736) obsahuje s pravděpodobností 0.95 skutečnou hodnotu rozdílu $\mu_x - \mu_y$. Má-li platit nulová hypotéza (podle které je $\mu_x - \mu_y = 0$), měla by nula ležet v tomto intervalu. Ta tam opravdu leží, tedy a docházíme tedy ke stejnému závěru jako pomocí p-hodnoty, a to k nezamítnutí H_0.

5) O platnosti předpokladů (a volbě statistického postupu) by se správně mělo rozhodovat ještě předtím, než jsou známa data. Předpoklady bychom tedy měli formulovat např. na základě historických dat apod. V případě t-testu je zejména předpoklad shody rozptylů poměrně zbytečný, neboť t-test, který shodu rozptylů nepředpokládá, je v situaci, kdy shoda rozptylů platí, zpravidla jenom nepatrně slabší (tzn. síla testu je menší) než t-test se shodou rozptylů. Obvykle tedy provádíme rovnou *t-test pro nestejné rozptyly*, který se nazývá Welchův test.

```
t.test(IQ~POHLAVI)
t.test(IQ[POHLAVI == 0], IQ[POHLAVI == 1])
```

✧ O kolik se změnila p-hodnota? Liší se náš závěr?

Jednostranná alternativa

6) Zjistěte, zda mají chlapci nižší IQ než dívky.

- ✧ Odpovězte na zkoumanou otázku pomocí vhodného testu.
- ✧ Poznámka: Kdybyste si nebyli jistí pořadím kategorií nějakého `factor`u, můžete si ho ověřit příkazem

```
levels(as.factor(POHLAVI))
```

2 Dvouvýběrový Wilcoxonův test

Nyní se pokusme zjistit, zda se významně liší průměrná známka na pololetním vysvědčení v 7. třídě u chlapců a dívek. Bude nás tedy zajímat, zda se střední hodnota průměru na vysvědčení v pololetí 7. třídy liší v populaci dívek a chlapců.

Označme jako X známku náhodně vybrané dívky a jako Y známku náhodně vybraného chlapce.

7) Je smysluplné předpokládat v této situaci, že X a Y jsou nezávislé náhodné veličiny?

8) Začněte opět tím, že spočítáte základní popisné statistiky pro ZN7 v závislosti na pohlaví žáka.

9) Pokračujte opět tím, že si pomocí krabičkových grafů uděláte představu o souvislostech mezi pohlavím žáka a jeho průměrem na pololetním vysvědčení v 7. třídě.

10) Lze předpokládat, že rozdělení průměrné známky je jak u chlapců tak u dívek normální?

```
shapiro.test(ZN7[POHLAVI == 0])
shapiro.test(ZN7[POHLAVI == 1])
```

11) V předcházejícím kroku jsme zjistili, že u obou pohlaví je rozdělení známek prokazatelně (dokonce na hladinách nižších než 1 %) nenormální. V takové situaci můžeme sáhnout k některému z pořadových testů. Ke zjištění, zda průměrná známka souvisí s pohlavím žáka, bychom mohli použít *dvouvýběrový Wilcoxonův test* (*Mannův-Whitneyův test*, *Wilcoxon rank sum test*). Testovat budeme H_0 : „rozdělení X je shodné s rozdělením Y“ proti H_1 : „rozdělení X se liší od rozdělení Y“. Vzhledem k tomu, že Wilcoxonův test je citlivý zejména na situace, kdy se rozdělení X a Y liší pouze polohou (např. mediánem), specifikují se testované hypotézy často jako $H_0 : \mathsf{med}(X) = \mathsf{med}(Y)$ proti $H_1 : \mathsf{med}(X) \neq \mathsf{med}(Y)$.

Proveďme dvouvýběrový Wilcoxonův test.

```
wilcox.test(ZN7 ~ POHLAVI)   # prvni moznost
wilcox.test(ZN7[POHLAVI==0], ZN7[POHLAVI==1])   # druha moznost
```

Tato funkce počítá testovou statistiku způsobem popsaným Mannem a Whitneyem v roce 1947. Hodnota $W = 5311$ ve výstupu je hodnotou testové statistiky U_y ze vzorce (6) v textu *DVT*.

✧ Jaký je závěr?
P-hodnota vyšla 0.9776. Nezamítáme tedy nulovou hypotézu, že by střední hodnota věku matek byla v populaci dívek i chlapců shodná.

3 Analýza rozptylu

Podrobnější popis teorie analýzy rozptylu jednoduchého třídění naleznete v textu Analýza rozptylu (dále jako ANV), který naleznete na mých stránkách.

Zjistěte, zda se střední hodnota IQ žáků liší na jednotlivých školách.

1) Spočtěte základní popisné statistiky pro IQ žáka v závislosti na tom, jakou školu navštěvuje (SKOLA) a vhodně je graficky znázorněte.

```
tapply(IQ,SKOLA,summary)
boxplot(IQ~SKOLA)
```

2) Znázorněte graficky průměrné IQ v závislosti na navštěvované škole spolu s 95% intervaly spolehlivosti pro střední IQ.

```
plotMeans(IQ, SKOLA, error.bars = "conf.int", level=0.95)
```

Pravděpodobnostní model a formulace hypotéz

Máme k dispozici údaje ze čtyř škol. Označme jako Y_k IQ náhodně vybraného žáka z k-té školy. V případě, že lze předpokládat, že $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $Y_3 \sim \mathcal{N}(\mu_3, \sigma_3^2)$, $Y_4 \sim \mathcal{N}(\mu_4, \sigma_4^2)$, a platí-li navíc $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$, umíme pomocí analýzy rozptylu testovat $\mathrm{H}_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ proti H_1 : *alespoň dvě střední hodnoty se liší.*

Ověřování předpokladů

3) Normalitu rozdělení veličin Y_1, Y_2, Y_3 a Y_4 budeme zkoumat až na konci.

4) Zjistěme si alespoň, zda je smysluplné předpokládat shodnou variabilitu IQ žáků v jednotlivých skupinách dle navštěvované školy. K tomu lze použít např. Leveneův test z knihovny `car`

```
install.packages("car")          # instalace knihovny (trvá dlouho)
library(car)                     # otevření knihovny
leveneTest(IQ ~ SKOLA)   # Leveneův test
```

nebo z knihovny `lawstat`

```
install.packages("lawstat")      # instalace knihovny (rychlejší)
library(lawstat)                 # otevření knihovny
levene.test(IQ, SKOLA)   # Leveneův test
```

✧ Z výstupu nás zajímá pouze p-hodnota, která v tomto případě činí 0.16. Jelikož je p-hodnota větší než zvolená hladina $\alpha = 0.05$, nezamítáme shodu populačních rozptylů.

Obdobně lze použít též Bartlettův test, který je však mnohem citlivější vůči případné nenormalitě.

```
bartlett.test(IQ ~ SKOLA)
```

✧ I zde nás případně zajímá pouze p-hodnota, která vyšla 0.28. Jelikož je to hodnota vyšší než předpokládaná hladina $\alpha = 0.05$, nezamítáme shodu populačních rozptylů IQ v jednotlivých školách a můžeme ji tedy předpokládat.

V praxi samozřejmě stačí použít pouze jeden z těchto testů - buď Leveneův, nebo Bartlettův. Vyšlo nám, že shodu populačních rozptylů hmotností v jednotlivých skupinách lze předpokládat. Označme si hodnotu tohoto společného rozptylu jako σ^2 (tj. předpokládáme $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 =: \sigma^2$).

Test hypotézy

5) Nyní pomocí analýzy rozptylu otestujme

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad \text{proti}$$
$$H_1 : \mu_1 \neq \mu_2 \ \vee \ \mu_1 \neq \mu_3 \ \vee \ \mu_1 \neq \mu_4 \ \vee \ \mu_2 \neq \mu_3 \ \vee \ \mu_2 \neq \mu_4 \ \vee \ \mu_3 \neq \mu_4$$

(jinými slovy: alespoň jedna dvojice středních hodnot se neshoduje).

Hladinu testu budeme uvažovat 5 %, tj. $\alpha = 0.05$.

```
mod01 <- aov(IQ ~ SKOLA)     # vytvoření modelu
summary(mod01)               # tabulka analýzy rozptylu
```

✧ vysvětlení anglických zkratek:

`Df` = degrees of freedom
`Sum Sq` = sum of squares
`Mean Sq` = mean squares
`F value` = value of F-statistics

✧ Porovnejte si tabulku z výstupu s tabulkou z textu *ANV* na str. 4. Vidíme, že:

Součet čtverců $S_A = 3625$
Reziduální součet čtverců $S_e = 38038$
Reziduální rozptyl $s^2 = 214.9$ (v textu *ANV* vzorec (2))
Hodnota testové statistiky $F = 5.622$ (v textu *ANV* vzorec (3))
p-hodnota $= 0.00105$

Celková variabilita S_T v tabulce uvedena není, neboť je součtem S_A a S_e.

✧ Jaký je závěr (zamítáme/nezamítáme H_0 na požadované hladině)?
P-hodnota vyšla 0.00105, což je méně než zvolená hladina 0.05, tudíž nulovou hypotézu zamítáme.

✧ Jak byste formulovali svůj závěr bez použití výrazů „zamítáme/nezamítáme H_0“?
Na hladině 5 % jsme prokázali, že IQ dětí se na jednotlivých školách liší.

✧ Jaký je odhad společného rozptylu hmotností σ^2?
Tím odhadem je reziduální rozptyl $s^2 = \frac{S_e}{f_e} = 214.9$.

6) Jelikož byla nulová hypotéza zamítnuta, potřebujeme zjistit, u kterých skupin se populační průměry liší. K tomu slouží mnohonásobné porovnávání, které se většinou provádí Tukeyho metodou.

```
TukeyHSD(mod01)                    # mnohonásobné porovnání
```

✧ Z dřívějška máme označeno, že μ_1 je populační průměr IQ žáků z 1. školy. Označme si dále $\bar{y}_1$ výběrový průměr IQ žáků z 1. školy v našich datech. Analogicky pak budeme mít μ_2 = populační průměr IQ žáků a $\bar{y}_2$ příslušný výběrový průměr, stejně tak pro $(\mu_2, \bar{y}_3)$ a $(\mu_4, \bar{y}_4)$. Z výstupu výše vidíme, že:

- $\bar{y}_2 - \bar{y}_1 = 6.737183$
- 95% interval spolehlivosti pro $\mu_2 - \mu_1$ je $(-0.282, 13.757)$
- p-hodnota testu hypotézy $H_0 : \mu_2 = \mu_1$ je 0.065.

Analogicky pro ostatní dvojice. Jediný interval spolehlivosti, který neobsahuje nulu, je interval „4-2“, a tedy u porovnání škol č. 2 a 4 nemá nula šanci být skutečným rozdílem příslušných populačních průměrů. Tomu odpovídá fakt, že příšluná p-hodnota je menší než 0.05. U všech ostatních dvojic je p-hodnota větší než 0.05 (a příslušný interval spolehlivosti pro rozdíl středních hodnot obsahuje nulu), a tudíž rozdíl IQ žáků u ostatních dvojic škol není statisticky významný. Závěr tedy je, že IQ žáků se významně odlišuje pouze na školách č. 2 a 4.

✧ Zkratka `p adj` by se dala přečíst jako „p-value **adjusted** for multiple comparisons“ a upozorňuje na fakt, že p-hodnoty berou v potaz, kolik průměrů porovnáváme. Zohledňují to i intervaly spolehlivosti. Spolehlivost 95 % se vztahuje na všechny intervaly současně!

✧ Intervaly spolehlivosti lze přehledně znázornit v obrázku, kde je ihned vidět, které z nich neobsahují nulu, a tudíž které z nich způsobily zamítnutí nulové hypotézy o rovnosti všech středních hodnot:

```
plot(TukeyHSD(mod01))
```

V našem případě nulu obsahují všechny intervaly kromě intervalu pro $\mu_4 - \mu_2$. Dvojice škol č. 2 a 4 tedy způsobila zamítnutí nulová hypotézy $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$.

7) Normální rozdělení lze nyní ověřit „najednou“ pomocí *standardizovaných reziduí* z odhadnutého „modelu“ (viz text *ANV*, str. 4, dole):

```
shapiro.test(rstandard(mod01))
```

P-hodnota vyšla 0.7291, což je více než hladina 0.05, a tudíž normální rozdělení IQ v jednotlivých skupinách nezamítáme.

8) ANOVA je v podstatě dvouvýběrový t-test rozšířený na srovnání více jak dvou skupin. Obdobně jako u t-testu lze vynechat předpoklad shodnosti rozptylů v jednotlivých skupinách a použít příkaz:

```
oneway.test(IQ ~ SKOLA)
```

✧ Výstup této funkce je bohužel trochu chudší než u funkce `aov` a neobsahuje kompletní tabulku analýzy rozptylu.

✧ P-hodnota testu hypotézy, že populační průměry jsou ve všech skupinách stejné, vyšla 0.0003625, což je podobně nízká hodnota jako u ANOVY s předpokladem shodných rozptylů. Náš závěr by byl tedy stejný.