Cvičení z matematické statistiky 07

Poslední úprava dokumentu: 8. dubna 2025.

Úvod do testování hypotéz: jednovýběrový t-test, p-hodnota

Dnes se podíváme na testování hypotéz u výběrů z normálního rozdělení a zopakujeme si pojmy, které s tím souvisejí. V závěru se pak možná dostaneme k tzv. pořadovým (neparametrickým) testům, které použijeme v situaci, kdy předpoklad normálního rozdělení není splněn. K lepšímu pochopení teorie vám může pomoci text *Párové testy* a *Neparametrické jednovýběrové testy* (dále jako *NJT*), které najdete na mých stránkách.

Ještě jednou pro jistotu zdůrazňuji, že populační průměr znaku X = střední hodnota X. Označuje se většinou jako μ (popř. μ_x) nebo $\mathsf{E}(X)$. Všechno to znamená totéž, a to průměrnou hodnotu znaku X v populaci.

Rozcvička

Načtěte si data `Deti.RData`, se kterými jsme pracovali minule. Jejich kompletní popis naleznete v pracovním listu k předchozímu cvičení.

```
load("data/Deti.RData")
```

Spočtěte 99% interval spolehlivosti pro střední hodnotu (tj. populační průměr) porodní hmotnosti dětí (veličina `por.hmotnost`). Nezapomeňte ověřit předpoklad normálního rozdělení!

1 Hmotnost chlapců ve 24. týdnu

1.1 Interval spolehlivosti

V situaci, kdy by pohlaví mělo vliv na hmotnost ve 24. týdnu, vypovídá výše spočtený interval o populačním průměru pouze tehdy, když poměr chlapců a dívek v datech odpovídá poměru chlapců a dívek v populaci. Toto by mělo platit za předpokladu, že děti do studie byly vybírány skutečně náhodně. V každém případě bude zajímavé zjistit, jakých hodnot nabývá populační průměr (střední hodnota) hmotnosti u chlapců, resp. dívek. Odborně řečeno budeme odhadovat podmíněnou (pohlavím) střední hodnotu hmotnosti.

1) Pro počítání intervalů spolehlivosti bude potřeba vytvořit podmnožiny dat obsahující pouze dívky, resp. chlapce. V dalším budeme střídavě pracovat se všemi dětmi, pouze s chlapci, či pouze s dívkami.

```
DetiH <- subset(Deti, Hoch == "hoch")
```

2) Nejprve spočtěme 95% interval spolehlivosti pro střední hmotnost hochů.

```
t.test(DetiH$hmotnost)
```

3) Ještě bychom se měli podívat, zda pro rozdělení hmotnosti u chlapců (popř. dívek) lze předpokládat normální rozdělení. Nakreslete histogram a normální diagram (QQ graf) pro hmotnost chlapců.

```
hist(DetiH$hmotnost)
qqnorm(DetiH$hmotnost)
qqline(DetiH$hmotnost)
```

✧Body na normálním diagramu se od přímky příliš neodchylují. Lze tedy předpokládat, že hmotnosti chlapců by mohly pocházet z normálního rozdělení.

4) Užitečné srovnání průměrů dvou populací poskytuje graf průměrů (Plot of means). K tomu můžete využít knihovnu `RcmdrMisc`.

```
install.packages("RcmdrMisc")
library(RcmdrMisc)                              # otevre knihovnu
plotMeans(Deti$hmotnost, Deti$Hoch, error.bars = "conf.int", level=0.95)
```

Body v grafu představují bodové odhady populační hmotnosti (= odhady střední hmotnosti) v obou skupinách, to jest výběrové průměry. „Fousy“ představují příslušné 95% intervaly spolehlivosti. Každý z těchto intervalů tedy s pravděpodobností 0.95 pokrývá skutečnou hodnotu populační hmotnosti pro dané pohlaví.

1.2 Jednovýběrový t-test (oboustranná alternativa)

5) Zajistěte si přímý přístup k chlapeckým hodnotám.

```
attach(DetiH)
```

6) Naším úkolem bude rozhodnout, zda data podporují hypotézu, že střední (to jest populační) hmotnost chlapců ve 24. týdnu je 8 200 g. Přitom požadujeme, aby k neoprávněnému zamítnutí této hypotézy došlo s pravděpodobností nejvýše 0.05, tj. zvolili jsme si *hladinu testu* $\alpha = 0.05$, neboli 5 %.

✧ Abychom mohli tuto hypotézu (statisticky) otestovat, je potřeba si toto slovní zadání přepsat do řeči matematiky...

7) Za náhodnout veličinu X si označíme hmotnost náhodně vybraného chlapce. Z QQ diagramu, který jsme si vykreslili dříve, víme, že lze předpokládat, že $X \sim \mathcal{N}(\mu, \sigma^2)$, tj. že rozdělení hmotnosti chlapců je normální s nějakou neznámou střední hodnotou μ a neznámým rozptylem σ^2. Ze zadání úlohy je zřejmé, že chceme testovat hypotézu $\mathrm{H}_0 : \mu = 8\,200$ g. Jako alternativní hypotézu si zvolíme $\mathrm{H}_1 : \mu \neq 8\,200$ g.

✧ Za předpokladu normálního rozdělení (který je ale vždy potřeba ověřit) lze k testu této hypotézy využít *jednovýběrový t-test*, jehož princip si prosím připomeňte z přednášky, nebo z textu *Úvod do testování hypotéz* (dále jako *ÚTH*), který najdete na mých stránkách.

8) Jednovýběrový t-test (angl. one-sample t-test) se v R provede jednoduchým příkazem

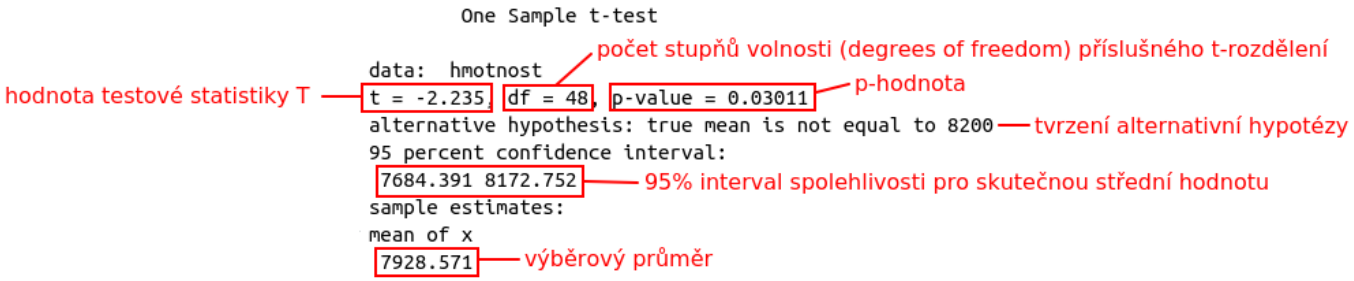
```
t.test(hmotnost, mu=8200)
```

kde do argumentu „`mu=`“ se udává testovaná hodnota z nulové hypotézy, v našem případě tedy 8 200. Řecké μ se totiž v angličtině nazývá „mu“ [mju:]. Nyní se podívejme na jednotlivé údaje ve výstupu:

✧ Jaký je závěr (zamítáme/nezamítáme H_0 na požadované hladině)?
Na základě výstupu z R můžeme o platnosti nulové hypotézy H_0 rozhodnout třemi způsoby (viz text *ÚTH*, str. 6, úplně dole):

(a) Podíváme se na hodnotu testové statistiky $T = -2.235$ a porovnáme ji s příslušným kvantilem[1] t-rozdělení $qt_{n-1}(1-\alpha/2) = qt_{48}(0.975) = 2.010635$ (kde n značí počet

[1]Standarně se kvantil t-rozdělení značí pouze $t_{n-1}(1-\alpha/2)$, prefix q jsem přidala pouze pro názornost, aby se zdůraznilo, že jde o kvantil. Snad to není naopak matoucí :-(



pozorování). Vidíme, že absolutní hodnota testové statistiky $|T| = 2.235$ je větší než tento kvantil (testová statistika překročila kritickou hodnotu), a tudíž zamítáme nulovou hypotézu H_0, že by střední (tj. populační) hmotnost chlapců byla 8 200 g.

(b) Pomocí p-hodnoty: Připomeňme, že p-hodnota je nejmenší hladina, na které bychom H_0 zamítli. Zde tedy bychom zamítli pro hladinu 0.03011 a vyšší. Naše zvolená hladina ze zadání příkladu je $\alpha = 0.05$ a jelikož $0.05 > 0.03011$, tak naše rozhodnutí opět je zamítnout H_0. Jednoduché pravidlo zní: „Je-li $\alpha >$ p-hodnota $\Rightarrow$ zamítáme H_0."

(c) Pomocí intervalu spolehlivosti: Víme, že 95% interval spolehlivosti obsahuje s pravděpodobností 0.95 skutečnou hodnotu střední hmotnosti. Je-li tedy 8200 g skutečná střední hmotnost, měla by v našem intervalu spolehlivosti (7684.391; 8172.752) ležet. Jelikož tam neleží, tak to asi nebude ta skutečná střední hmotnost. Opět tedy závěr zní - zamítnout H_0.

Všechny tři způsoby vždy vedou ke stejnému závěru, takže je jedno, který si vyberete. První způsob je poněkud nepraktický, neboť si musíte ručně spočítat kvantil $qt_{48}(0.975)$, a uvedli jsme ho spíš pro úplnost. V praxi je asi nejpoužívanější způsob (b), popř. (c).

✧ Jak byste formulovali svůj závěr bez použití výrazů „zamítáme/nezamítáme H_0"?
„Na 5% hladině jsme prokázali, že střední hmotnost je různá od 8200 g." (Ten rozdíl je statisticky významný).

✧ Jaká jiná hodnota μ při nulové hypotéze by vedla k opačnému závěru?
No nějaká, která leží uvnitř intervalu spolehlivosti. Tak třeba 8 000 g.

✧ Je potřeba znovu počítat, jestliže bychom požadovali přísnější 1% hladinu významnosti, resp. benevolentnější 10% hladinu významnosti?
NE. Na základě výstupu výše lze rozhodnout o platnosti H_0 i na jiné hladině než 5 %. Sice už nelze použít 95% interval spolehlivosti, ale můžeme se řídit p-hodnotou. Je-li p-hodnota menší než zvolená hladina (0.1 nebo 0.01), tak H_0 zamítáme.

Pokud bychom ale přece jen chtěli výstup procedury t-test přepočítat pro hladinu 10 % (tj. $\alpha = 0.1$), použili bychom:

```
t.test(hmotnost, mu=8200, conf.level = 0.9)
```

Argument `conf.level` udává spolehlivost a je rovna vždy $1 - \alpha$.

Ruční výpočet testové statistiky a p-hodnoty

9) Pojďme si osvěžit klasické programování a vypočítat hodnotu testové statistiky a p-hodnoty ručně. Nejprve zkuste příkazy vymyslet sami a až pak se podívat na řešení níže.

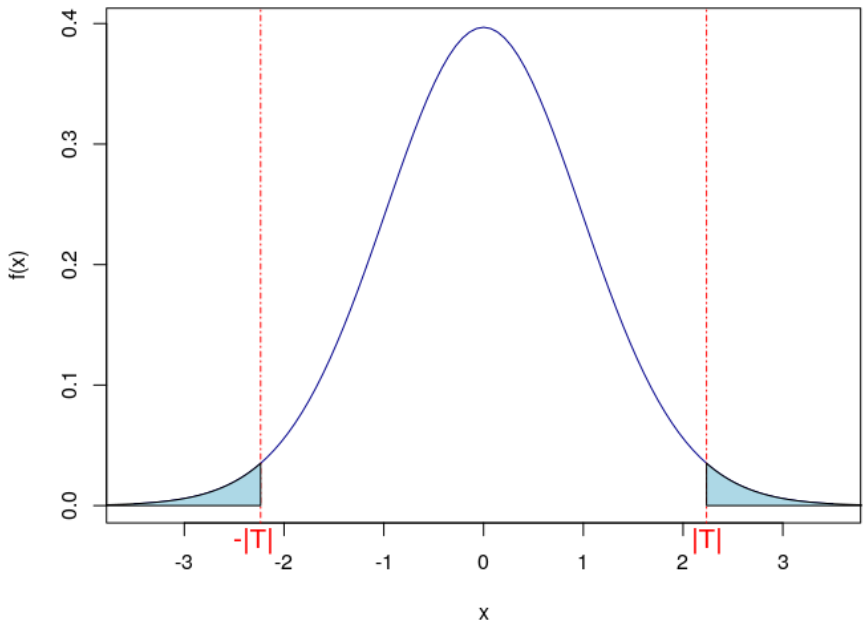
✧ Hodnota testové statistiky. (Příkazy jsou v závorkách, aby se současně s uložením do proměnné vypsala i její hodnota.)

```
(n <- length(hmotnost))          # počet pozorování
(prumer <- mean(hmotnost))       # výběrový průměr
```

```
(se.prumer <- sd(hmotnost)/sqrt(n)) # směrodatná chyba průměru
(odlisnost <- prumer - 8200)        # odchylka půrměru od testované hodnoty
(T <- odlisnost / se.prumer)        # hodnota t-statistiky
```

Ověřte si, že nám vyšla stejná hodnota jako ve výstupu procedury `t.test`.

✧ P-hodnota. Z definice p-hodnoty (text *ÚTH*, str. 4-5) je jasné, že p-hodnota je součtem ploch vymezených vypočtenými hodnotami testové statistiky: Tyto modré plochy odpo-



vídají pravděpodobnostem $P(\text{testová statistika} \leq -|T|)$ a $P(\text{testová statistika} \geq |T|)$. V R tyto pravděpodobnosti spočítáme pomocí distribuční funkce t-rozdělení:

```
(P.vlevo <- pt(-abs(T), df=n-1))      # = P(testová statistika <= -|T|)
(P.vpravo <- 1 - pt(abs(T), df=n-1))  # = P(testová statistika >= |T|)
(P <- P.vlevo + P.vpravo)             # p-hodnota je součtem obou ploch
```

✧ Jak je vidět, obě modré plochy jsou stejné, a tak p-hodnotu lze spočítat najednou:

```
(P <- 2*pt(-abs(T), df=n-1))
```

Ověřte si, že nám vyšla stejná hodnota jako ve výstupu procedury `t.test`.

10) Graficky se lze na P-hodnotu podívat následovně:

```
curve(dt(x,n-1), xlim=c(-3.5, 3.5), xlab="x", ylab="f(x)", col="darkblue")
abline(v=c(-abs(t), abs(t)), col="red", lty=4)
```

2 Testování předpokladu normálního rozdělení: Shapiro-Wilkův test

Ověření předpokadu normálního rozdělení našeho výběru jsme doposud prováděli pomocí QQ diagramu. Tento předpoklad lze ale též formálně otestovat pomocí Shapiro-Wilkova testu.

```
shapiro.test(hmotnost)
```

3 Jednovýběrový t-test - jednostranná alternativa

1) Pokusme se prokázat, že populační hmotnost chlapce ve 24. týdnu je **menší** než 8 100 g. Budeme požadovat, aby k chybnému prokázání tohoto tvrzení došlo s pravděpodobností nejvýše 0.05.

2) Za X označíme hmotnost náhodně vybraného chlapce a budeme předpokládat, že $X \sim \mathcal{N}(\mu, \sigma^2)$, potom můžeme opět pomocí *jednovýběrového t-testu* testovat $\mathrm{H}_0 : \mu = 8\,100$ g proti $\mathrm{H}_1 : \mu < 8\,100$ g. Testujeme na hladině 5 %, tj. $\alpha = 0.05$.

✧ **Důležité připomenutí.** Znaménko (tj. $>$, $<$, $\neq$) v alternativní hypotéze je nutno zvolit podle toho, co nás zajímá a v každém případě ještě před tím než vidíme data! Jinými slovy, volba alternativní hypotézy podle toho, „jak to vyšlo v datech“ je hrubou chybou.

3) Normalitu už jsme ověřovali před chvílí, tak to můžeme přeskočit.

4) Provedeme jednovýběrový t-test, kde do argumentu „`alternative=`“ musíme specifikovat tvar alternativní hypotézy, zde „menší než 8 100 g“, tj. `less`.

```
t.test(DetiH$hmotnost, mu=8100, alternative="less")
```

5) Jaký je závěr (zamítáme/nezamítáme H_0 na požadované hladině)?
P-hodnota vychází 0.08226, tj. je větší než 0.05, a proto nezamítáme H_0. Stejný závěr bychom učinili i na základě příslušného intervalu[2] spolehlivosti $(-\infty; 8132.261)$, který obsahuje hodnotu 8 100 g (ta tedy má šanci být skutečným populačním průměrem hmotnosti).

6) Jak byste formulovali svůj závěr bez použití výrazů „zamítáme/nezamítáme H_0“?
„Data neprokázala, že by střední hmotnost chlapců ve 24. týdnu byla odlišná od 8 100 g.“ (Pokud tam nějaká odlišnost je, tak není statisticky významná. Na její případné prokázání bychom potřebovali více dat.)

7) Jaká jiná hodnota μ při nulové hypotéze by vedla k opačnému závěru?
No každá, která neleží v našem intervalu spolehlivosti. Tak třeba 9 000 g.

8) Je potřeba znovu počítat, jestliže bychom požadovali přísnější 1% hladinu významnosti, resp. benevolentnější 10% hladinu významnosti?
Ne. Stačí příslušnou hladinu porovnat s vypočtenou p-hodnotou 0.08226.

9) Ruční výpočet p-hodnoty by v tomto případě vypadal následovně:

```
pt(-1.4116, df=48)
```

kde -1.4116 je vypočtená hodnota t-statistiky.

[2] `Inf` ve výstupu je zkratka pro `Infinity`, tj. nekonečno.