Cvičení z matematické statistiky 08

Poslední úprava dokumentu: 8. dubna 2025.

Jednovýběrový t-test: p-hodnota, síla testu, ověřování normality Neparametrické jednovýběrové testy, párový t-test

---

Dnes budeme pokračovat v testování hypotéz u výběrů z normálního rozdělení a zopakujeme si pojmy, které s tím souvisejí. V závěru se pak 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.

Rozcvička

Načtěte si data `Deti.RData`, se kterými jsme pracovali minule. Jejich kompletní popis naleznete v pracovním listu k 6. cvičení.

```
load("data/Deti.RData")
DetiH <- subset(Deti, Hoch == "hoch")
attach(DetiH)
```

Ověřte hypotézu, že střední hmotnost chlapců ve 24. týdnu je 8 100 g. Hladinu uvažujte 5%. Nezapomeňte ověřit předpoklady testu! Hodnotu testové statistiky zkuste spočítat též ručně. 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`.

1 p-hodnota testu

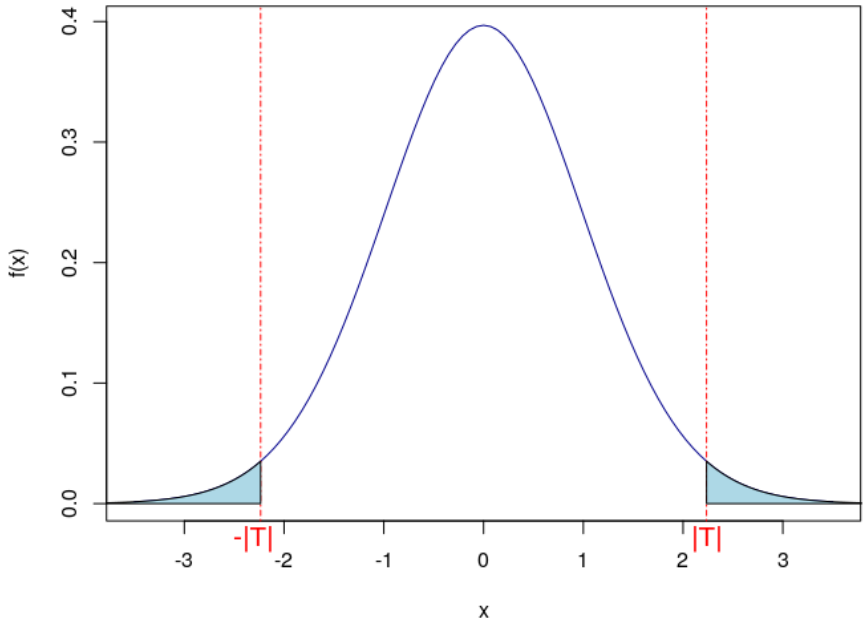
Pro lepší pochopení p-hodnoty se nyní podíváme na její ruční výpočet.

1) 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 odpovídají 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`.

2) 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 Ověřování 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

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.

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) Ujistěte se, že předpoklady testu jsou splněné.
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[1] 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.

4 Síla jednovýběrového t-testu

Posledním pojmem, který je s teorií testování hypotéz neodmyslitelně spjat, je *síla* testu. Připomeňte si jeho význam (např. text *ÚTH*: str. 5-6).

10) Pokusme se udělat si představu o síle testu z Rozcvičky (tj. test hypotézy $H_0 : \mu = 8100$ g proti alternativě $H_1 : \mu \neq 8100$ g). Existuje vztah, který dává do souvislosti sílu testu, hladinu α a počet pozorování n:

$$n \geq \left(\frac{qnorm(1-\alpha/2) + qnorm(\texttt{síla})}{\mu_1 - \mu_0} \right)^2 \sigma^2, \tag{1}$$

kde μ_0 je testovaná hodnota (z nulové hypotézy) a μ_1 je skutečná střední hodnota (tu samozřejmě neznáme), σ^2 je pak skutečný rozptyl daného znaku (ten taky neznáme). *qnorm* značí příslušný kvantil normovaného normálního rozdělení $N(0, 1)$.

Máme-li požadavek na `sílu` a zajímá nás jaký počet pozorování je k dosažení této síly potřebný, ze vzorečku (1) můžeme vypočítat potřebné n. Máme-li naopak dané n, můžeme úpravou vzorečku (1) vypočítat `sílu`. Naštěstí nemusíme výpočet provádět ručně, v R je na to příkaz `power.t.test`.

✧ Odhadněte sílu testu (připomeňme, že počet chlapců byl 49):

```
power.t.test(n = 49, delta = 8100-mean(hmotnost), sd = sd(hmotnost),
sig.level = 0.05, type = "one.sample", alternative = "two.sided")
```

argument `delta` odpovídá rozdílu $\mu_1 - \mu_0$, `sd` je odhad σ a `sig.level` udává hladinu. Vychází nám, že síla je pouhých 28 %.

✧ Odhadněte, kolik pozorování by bylo potřeba, aby měl příslušný t-test sílu 0.9?

```
power.t.test(power = 0.9, delta = 8100-mean(hmotnost), sd = sd(hmotnost),
sig.level = 0.05, type = "one.sample", alternative = "two.sided")
```

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

Bylo by potřeba 260 pozorování. (Zvýšit počet pozorování je v praxi také jediný způsob, jak sílu testu ovlivnit.)

11) Odhadněte sílu testu z bodu 7) ze sekce 2

```
power.t.test(n = 49, delta = 8100-mean(hmotnost), sd = sd(hmotnost),
sig.level = 0.05, type = "one.sample", alternative = "one.sided")
```

Uvědomte si, že vypočtené hodnoty síly a počtu pozorování jsou pouhými odhady těchto hodnot, protože ve vzorečku (1) se vyskytují neznámé parametry skutečného rozdělení (μ_1 a σ^2), které jsme museli při výpočtu odhadnout jejich výběrovými protějšky. To nám do výpočtu síly, popř. potřebného počtu pozorování, vneslo nepřesnost.

5 Neparamterické jednovýběrové testy

Pokusíme se zjistit, zda je střední porodní délka dětí rovna 51 cm. Označme jako X porodní délku náhodně vybraného dítěte. Chceme testovat hypotézu $H_0 : \mathsf{E}(X) = 51$ cm proti $H_1 : \mathsf{E}(X) \neq 51$ cm, kde $\mathsf{E}(X)$ značí střední hodnotu veličiny X. Hladinu testu uvažujme opět 5 %.

1) Spočtěte základní popisné statistiky pro porodní délku dětí.

2) Obrázky i formálním testem zjistěte, zda je smysluplné předpokládat normální rozdělení pro porodní délku dětí. Pro posouzení normality nakreslete mimo jiné též histogram.

✧ Patrně jste zjistili, že rozdělení porodní délky není normální, ale je mírně zešikmené. K otestování naší hypotézy H_0 tedy musíme použít nějaký neparametrický/pořadový test.

Jednovýběrový Wilcoxonův test

Pokud chceme použít neparametrický test, musíme přeformulovat testované hypotézy pomocí populačního mediánu. Budeme nyní testovat nulovou hypotézu

H_0 : populační medián X je roven 51 cm, proti alternativě
H_1 : populační medián X není roven 51 cm.

3) Jednovýběrový Wilcoxonův test zjišťující, zda je typická porodní délka dítěte rovna 51 cm provedeme pomocí příkazu

```
wilcox.test(Deti$por.delka, mu=51)
```

Hodnota V ve výstupu je testová statistika Wilcoxonova testu (před vynormováním) a udává součet pořadí od těch hodnot X_i, které byly větší než 51 (po vyloučení hodnot rovných 51). V textu *NJT* na str. 1 je tato statistika označena jako W.

✧ Jaký je závěr (zamítáme/nezamítáme H_0 na požadované hladině)?

✧ Jak byste formulovali svůj závěr bez použití výrazů „zamítáme/nezamítáme H_0“?

Nicméně i použití Wilcoxonova testu je v této situaci diskutabilní. Jednovýběrový Wilcoxonův test sice nepředpokládá normální rozdělení, avšak stále předpokládá symetrické rozdělení (což nebyl úplně případ porodní délky dětí, jak jsme viděli na histogramu).

Znaménkový test

4) Nejobecnějším (avšak zároveň nejslabším) testem pro tuto situaci je test znaménkový.

✧ K provedení znaménkového testu je nejprve potřeba zjistit, u kolika jedinců nastala situace, kdy je dítě po porodu delší než 51 cm.

```
U <- sum(Deti$por.delka > 51)        # v textu NJT oznaceno jako U
```

✧ Dále je potřeba zjistit, u kolika jedinců je porodní délka odlišná od 51 cm, neboť tato pozorování jsou pro znaménkový test bezcenná (jako by v datech nebyly).

```
n2 <- sum(Deti$por.delka  != 51)   # n* z textu NJT
```

✧ Znaménkový test (asymptotický, s Yatesovými korekcemi) nyní provedeme takto:

```
prop.test(U, n2)
```

✧ X-squared ve výstupu je druhá mocnina statistiky Z_3 z textu *NJT*, vzorec (3).

✧ Význam ostatních hodnot ve výstupu si ozřejmíme později, v rámci cvičení o kategoriálních datech.

✧ Jaký je závěr (zamítáme/nezamítáme H_0 na požadované hladině)?

✧ Jak byste formulovali svůj závěr bez použití výrazů „zamítáme/nezamítáme H_0“?

6 Párový t-test (= převlečený jednovýběrový t-test)

Pokusíme se zjistit, zda jsou otcové v průměru o více jak 14 cm vyšší než matky.

Označme jako X výšku náhodně vybraného otce a jako Y výšku jeho partnerky. Data lze tedy považovat za realizaci náhodného výběru z dvourozměrného rozdělení náhodného vektoru $\begin{pmatrix} X \\ Y \end{pmatrix}$. Jak víme, k zodpovězení naší otázky stačí zabývat se náhodnou veličinou $Z = X - Y$, kde Z reprezentuje výškový rozdíl otce a matky náhodně vybraného dítěte. Testujeme pak $H_0 : \mathsf{E}(Z) = 14$ cm proti $H_1 : \mathsf{E}(Z) > 14$ cm, kde $\mathsf{E}(Z)$ značí střední hodnotu veličiny Z, čili její populační průměr. Jestliže lze navíc předpokládat normální rozdělení náhodné veličiny Z, lze k testu těchto hypotéz použít nám dobře známý jednovýběrový t-test (aplikovaný na výškové rozdíly). Hladinu testu uvažujte 5 %.

5) Připravte novou proměnnou vyskaDif, jež bude udávat výškový rozdíl mezi otcem a jeho partnerkou.

```
detach(DetiH)
Deti$vyskaDif <- Deti$vyska.o - Deti$vyska.m
```

6) Spočtěte základní popisné statistiky pro výškové rozdíly mezi otcem a matkou.

```
summary(Deti$vyskaDif)  # charakteristiky polohy
sd(Deti$vyskaDif)       # charakteristika variability
```

7) Pomocí obrázků i formálním testem zjistěte, zda je smysluplné předpokládat normální rozdělení pro výškové rozdíly mezi otcem a matkou.

```
hist(Deti$vyskaDif)          # histogram
qqnorm(Deti$vyskaDif)        # normalni diagram
qqline(Deti$vyskaDif)
shapiro.test(Deti$vyskaDif) # Shapiro-Wilkuv test normality
```

Normální rozdělení zde můžeme předpokládat (p-hodnota Shapiro-Wilkova testu je 0.4836).

8) Pomocí párového t-testu (pro výšky otců a matek), resp. jednovýběrového t-testu pro výškové rozdíly, otestujte $H_0 : \mathsf{E}(Z) = 14$ cm proti $H_1 : \mathsf{E}(Z) > 14$ cm.

```
t.test(Deti$vyskaDif, mu=14, alternative="greater")        # nebo:
t.test(Deti$vyska.o, Deti$vyska.m, mu=14, alternative="greater", paired=TRUE)
```

✧ Jaký je závěr (zamítáme/nezamítáme H_0 na požadované hladině)?
P-hodnota vyšla 0.9802, což je víc než zvolená 5% hladina. Nezamítáme tedy H_0. Na základě našich dat nelze zamítnout hypotézu, že rozdíl populačních průměrů výšek je 14 cm.

✧ Jak byste formulovali svůj závěr bez použití výrazů „zamítáme/nezamítáme H_0“?
Data neprokázala, že by rozdíl populačních průměrů výšek byl různý od 14 cm. (Pokud ten rozdíl není 14 cm, tak ale ne statisticky významně. Na detekci případné odchylky bychom potřebovali více dat.)

✧ Jak souvisí výsledek testu s intervalem spolehlivosti, jež je též uveden ve výstupu?
Je to 95% interval spolehlivosti pro $\mathsf{E}(Z)$, čili je to interval, který s pravděpodobností 0.95 obsahuje skutečnou hodnotu rozdílu populačních průměrů výšek otce a matky. Jelikož hodnota 14 leží v tomto intervalu, má šanci být tím skutečným rozdílem výšek. Docházíme tedy ke stejnému závěru, a to nezamítnutí H_0.

Výsledný interval spolehlivosti má tvar $(10.93, \infty)$, je tedy jednostranný. Je to proto, že jsme zvolili jednostrannou alternativní hypotézu.

7 Samostatná práce

Pokuste se zjistit, zda jsou otcové typicky o více jak 2 roky starší než matky.

Označme jako X věk náhodně vybraného otce a jako Y věk jeho partnerky. Data lze tedy opět považovat za realizaci náhodného výběru z dvourozměrného rozdělení náhodného vektoru $\begin{pmatrix} X \\ Y \end{pmatrix}$. K zodpovězení naší otázky však opět stačí zabývat se náhodnou veličinou $Z = X - Y$, kde Z reprezentuje věkový rozdíl otce a matky náhodně vybraného dítěte. Chceme testovat hypotézu $H_0 : \mathsf{E}(Z) = 2$ roky proti $H_1 : \mathsf{E}(Z) > 2$ roky, kde $\mathsf{E}(Z)$ značí střední hodnotu veličiny Z. Hladinu testu uvažujte opět 5 %.

✧ Připravte novou proměnnou vekDif, jež bude udávat věkový rozdíl mezi otcem a jeho partnerkou.

✧ Spočtěte základní popisné statistiky pro věkové rozdíly mezi otcem a matkou.

✧ Obrázky i formálním testem zjistěte, zda je smysluplné předpokládat normální rozdělení pro věkové rozdíly mezi otcem a matkou.

Patrně jste zjistili, že rozdělení věkových rozdílů mezi otcem a matkou není normální, ale je značně zešikmené. K posouzení věkové odlišnosti mezi otcem a matkou musíme tedy použít nějaký neparametrický/pořadový test.