

10 Neparametrické testy o mediánech

Neparametrické testy o mediánech se používají v situacích, kdy nejsou splněny předpoklady nutné pro aplikaci parametrických testů (aspoň intervalový charakter zkoumané veličiny, normalita dat). U neparametrických testů o mediánech stačí předpokládat, že daný náhodný výběr má aspoň ordinální charakter a pochází ze spojitého rozložení, eventuálně ještě přistupuje předpoklad o symetrii hustoty kolem mediánu. Tyto testy jsou však slabší než parametrické testy, tj. nepravdivou nulovou hypotézu zamítají s menší pravděpodobností.

10.1 Jednovýběrové a párové testy

X_1, \dots, X_n , $n \geq 2$ je náhodný výběr ze spojitého rozložení s mediánem $x_{0,50}$, c je reálná konstanta, $\alpha \in (0, 1)$ je hladina významnosti. Testujeme $H_0: x_{0,50} = c$ proti $H_1: x_{0,50} \neq c$, resp. proti $H_1: x_{0,50} < c$ resp. proti $H_1: x_{0,50} > c$. Je-li $(X_1, Y_1)^T, \dots, (X_n, Y_n)^T$ náhodný výběr ze spojitého dvourozměrného rozložení, $n \geq 2$, pak zavedeme rozdílový náhodný výběr $Z_1 = X_1 - Y_1, \dots, Z_n = X_n - Y_n$, o němž předpokládáme, že pochází z rozložení s mediánem $z_{0,50}$. Tím jsme od dvourozměrného rozložení přešli k jednorozměrnému rozložení.

Jednovýběrové a párové neparametrické testy jsou obdobou jednovýběrového a párového t-testu.

10.1.1 Znaménkový test

Utvoříme rozdíly $X_i - c$ u jednovýběrového testu resp. $Z_i - c$ u párového testu, $i = 1, \dots, n$. (Jsou-li některé rozdíly nulové, pak za n bereme jen počet nenulových hodnot.) Testovou statistikou je počet kladných rozdílů S_Z^+ . Platí-li H_0 , pak $S_Z^+ \sim \text{Bi}(n, 1/2)$. Stanovíme kritický obor W . Pokud $S_Z^+ \in W$, H_0 zamítáme na hladině významnosti α a přijímáme H_1 .

Pro oboustrannou alternativu $W = \langle 0; k_1 \rangle \cup \langle k_2; n \rangle$, $k_1 = K_{\alpha/2}(S_Z^+) - 1$, $k_2 = K_{1-\alpha/2}(S_Z^+) + 1$, přičemž $K_\alpha(S_Z^+)$ je α -kvantil rozložení $\text{Bi}(n, 1/2)$. Pro levostrannou alternativu: $W = \langle 0; k_1 \rangle$, $k_1 = K_\alpha(S_Z^+) - 1$. Pro pravostrannou alternativu: $W = \langle k_2; n \rangle$, $k_2 = K_{1-\alpha}(S_Z^+) + 1$.

Čísla k_1, k_2 pro oboustranný test i jednostranné testy najdeme ve statistických tabulkách (pro $n \in \{6, 7, \dots, 20\}$) nebo je vypočítáme pomocí software. Pro $n > 20$ lze použít asymptotickou variantu testu: za platnosti H_0 se testová statistika $U_0 = \frac{S_Z^+ - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$ asymptoticky řídí rozložením $N(0, 1)$. Kritické obory pak mají tvar: $W = (-\infty; -u_{1-\alpha/2}) \cup \langle u_{1-\alpha/2}; \infty \rangle$ resp. $W = (-\infty; -u_{1-\alpha})$ resp. $W = \langle u_{1-\alpha}; \infty \rangle$. U asymptotické varianty testu lze využít korekci na nespojnost.

Příklad 10.1. Řešený příklad

Načtete datový soubor 31-goldman-alaska.csv obsahující údaje o délce vřetenní kosti z pravé strany (radius.R) a z levé strany (radius.L) mužů a žen aljašské populace z kmene Ipituaq. Na hladině významnosti α testujte hypotézu, že délka vřetenní kosti z pravé strany a z levé strany u mužů je shodná. Testování proveďte pomocí znaménkového testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,05$.

Řešení příkladu 10.1

Zadání příkladu vede na párový test. Nejprve je třeba ověřit normální rozložení náhodného výběru rozdílů délek vřetenních kostí u mužů z pravé a z levé strany. Na hladině významnosti $\alpha = 0,05$ testujeme H_0 : *Rozdíly délky vřetenní kosti z pravé a z levé strany u mužů aljašské populace z kmene Ipituaq pochází z normálního rozložení.* oproti H_1 : *Rozdíly délky vřetenní kosti z pravé a z levé strany u mužů aljašské populace z kmene Ipituaq nepochází z normálního rozložení.* K otestování normality použijeme vzhledem k rozsahu náhodného výběru ($n = 12$) Shapiro-Wilkův test.

```
1 data <- read.delim('31-goldman-alaska.csv', sep = ';', dec = '.')
2 data.M <- data[data$sex == 'm' & data$pop == 'Ipituaq', c('radius.R', 'radius.L')]
3 data.M <- na.omit(data.M)
4 radius.RIM <- data.M$radius.R
5 radius.LIM <- data.M$radius.L
6 diff <- radius.RIM - radius.LIM
7 shapiro.test(diff)$p.val # 0,040717
```

Náhodný výběr rozdílů délek vřetených kostí z pravé a z levé strany mužů aljašské populace z kmene Ipituaq nepochází z normálního rozložení (p -hodnota = 0,0407).

Protože rozdíl hodnot naměřených z pravé a z levé strany nepochází z normálního rozložení, použijeme na ověření zadané hypotézy v souladu se zadáním neparametrický znaménkový párový test. Na hladině významnosti $\alpha = 0,05$ testujeme $H_0: z_{0,50} = 0$ oproti $H_1: z_{0,50} \neq 0$ (oboustranný test), kde $z_{0,50}$ je medián rozdílů. Vzhledem k rozsahu náhodného výběru ($n < 20$) zvolíme exaktní variantu znaménkového párového testu implementovanou ve funkci `signTest()` z knihovny `EnvStats`. Výstupem funkce je realizace testové statistiky S_Z^+ a p -hodnota. Hranice kritického oboru dopočítáme pomocí funkce `qbinom()`.

```
8 c <- 0
9 EnvStats::signTest(diff, mu = c, alternative = 'two.sided', conf.level = 0.95)
10 n <- sum(diff != c) # 12
11 alpha <- 0.05
12 qbinom(alpha / 2, n, 1 / 2) # 3
13 qbinom(1 - alpha / 2, n, 1 / 2) # 9
```

```

      Sign test
data:  diff
# Obs > median = 7, p-value = 0,7744
alternative hypothesis: true median is not equal to 0
sample estimates:
median
 2,25
```

14
15
16
17
18
19
20
21
22

Realizace testové statistiky $s_Z^+ = 7$, kritický obor $W = (-\infty; 3) \cup \langle 9; \infty$). Protože $s_Z^+ \notin W$, H_0 nezamítáme na hladině významnosti $\alpha = 0,05$. Protože p -hodnota = 0,7744 je větší než $\alpha = 0,05$, H_0 nezamítáme na hladině významnosti $\alpha = 0,05$. Mezi délkou vřetení kosti z pravé strany a z levé strany mužů aljašské populace z kmene Ipituaq neexistuje statisticky významný rozdíl.

★

Příklad 10.2. Neřešený příklad

Načtěte datový soubor `10-two-samples-means-birth.txt` obsahující údaje o porodní hmotnosti novorozenců (`birth.W`) narozených v jedné krajské nemocnici v průběhu jednoho roku. Dále mějme k dispozici údaje o průměrné porodní hmotnosti novorozenců (`www.czso.cz`) v České republice ($\bar{x} = 3283$ g). Na hladině významnosti α testujte hypotézu, že porodní hmotnost novorozenců narozených ve sledované krajské nemocnici se neliší od průměrné porodní hmotnosti novorozenců v České republice. Testování proveďte pomocí znaménkového testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,01$.

Výsledky: Lillieforsův test: p -hodnota $< 0,001$, $\alpha = 0,01$; data nepochází z normálního rozložení; asymptotická varianta znaménkového jednovýběrového testu ($n > 20$): $u_0 = -0,4595$, $W = (-\infty; -2,5758) \cup \langle 2,5758; \infty$); p -hodnota = 0,6459, $\alpha = 0,01$; H_0 nezamítáme na hladině významnosti $\alpha = 0,01$.

★

10.1.2 Wilcoxonův test

K předpokladům uvedeným u znaménkového testu přistupuje ještě předpoklad symetrie hustoty podle mediánu. První krok postupu je stejný jako u znaménkového testu. Dále stanovíme pořadí R_i vzestupně uspořádaných absolutních hodnot rozdílů $X_i - c$ resp. $Z_i - c$ a vypočteme statistiky S_W^+ , což je součet pořadí R_i přes kladné hodnoty rozdílů a S_W^- , což je součet pořadí R_i přes záporné hodnoty rozdílů. Testová statistika je rovna $\min(S_W^+, S_W^-)$ pro oboustrannou alternativu, S_W^+ pro levostrannou alternativu a S_W^- pro pravostrannou alternativu. H_0 zamítáme na hladině významnosti α , když testová statistika je menší nebo rovna tabelované kritické hodnotě. Tu najdeme ve statistických tabulkách (pro $n \in \{6, 7, \dots, 30\}$) nebo ji vypočítáme pomocí software. Pro $n > 30$ lze použít asymptotickou variantu testu: za platnosti H_0 se testová statistika $U_0 = \frac{S_W^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ asymptoticky řídí rozložením

$N(0, 1)$. Kritické obory jsou stejné jako u asymptotické varianty znaménkového testu. U asymptotické varianty testu lze využít korekci na nespojitost.

Poznámka: V softwaru \mathbb{R} funguje Wilcoxonův test jinak. Testová statistika je pro všechny tři alternativy S_W^+ , kritické obory mají tvar $W = (-\infty; s_{\alpha/2}(n)) \cup (s_{1-\alpha/2}(n); \infty)$ pro oboustranný test; $W = (-\infty; s_{\alpha}(n))$ pro levostranný test a $W = (s_{1-\alpha}(n); \infty)$ pro pravostranný test. Hodnota $s_{\alpha}(n)$ je α -kvantil rozložení, jímž se za platnosti nulové hypotézy řídí testová statistika Wilcoxonova pořadového testu pro náhodný výběr o rozsahu n (po odstranění shod). Vypočítáme jej pomocí funkce `qsignrank()`.

Příklad 10.3. Řešený příklad

Načtete datový soubor `15-anova-means-skull.txt` obsahující údaje o výšce horní části tváře (`upface.H`) mužů bantuské, čínské, malajské, německé a peruánské populace. Dále mějme k dispozici údaje o výšce horní části tváře mužů ze středověké a novověké populace ze západopolské Cedyňe ($\bar{x}_m = 67,30$ mm, $s_m = 4,44$ mm, $n_m = 189$). Na hladině významnosti α testujte hypotézu, že výška horní části tváře mužů německé populace je menší nebo rovná výšce horní části tváře mužů ze západopolské Cedyňe. Testování proveďte pomocí Wilcoxonova testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,10$.

Řešení příkladu 10.3

Zadání příkladu vede na jednovýběrový test. Nejprve je třeba ověřit normální rozložení náhodného výběru výšek horní části tváře mužů německé populace. Na hladině významnosti $\alpha = 0,10$ testujeme H_0 : *Data pochází z normálního rozložení.* proti H_1 : *Data nepochází z normálního rozložení.* K otestování normality použijeme vzhledem k rozsahu náhodného výběru ($n = 19$) Shapirov-Wilkův test.

```
23 data <- read.delim('15-anova-means-skull.txt', sep = '\t', dec = '.')
24 upface.HN <- na.omit(data[data$pop == 'nem', 'upface.H'])
25 shapiro.test(upface.HN)$p.val # 0,04190113
```

Náhodný výběr výšek horní části tváře mužů německé populace nepochází z normálního rozložení (p -hodnota = 0,0419).

Protože naměřené hodnoty nepochází z normálního rozložení, použijeme na ověření zadané hypotézy v souladu se zadáním neparametrický Wilcoxonův jednovýběrový test. Na hladině významnosti $\alpha = 0,10$ testujeme H_0 : $x_{0,50} \leq 67,30$ proti H_1 : $x_{0,50} > 67,30$ (pravostranný test). Vzhledem k rozsahu náhodného výběru ($n < 30$) zvolíme exaktní variantu Wilcoxonova testu implementovanou ve funkci `wilcox.test()`. Výstupem funkce je realizace testové statistiky S_W^+ a p -hodnota. Hranici kritického oboru dopočítáme pomocí funkce `qsignrank()`.

```
26 c <- 67.30
27 n <- sum(upface.HN != c) # 19
28 wilcox.test(upface.HN, mu = c, alternative = 'greater', conf.level = 0.90)
29 alpha <- 0.10
30 qsignrank(alpha, n) # 63
```

```

      Wilcoxon signed rank test with continuity correction
data:  upface.HN
V = 179, p-value = 0,0003789
alternative hypothesis: true location is greater than 67,3
31
32
33
34
35
36
```

Realizace testové statistiky $s_W^+ = 179$, kritický obor $W = (63; \infty)$. Protože $s_W^+ \in W$, H_0 zamítáme na hladině významnosti $\alpha = 0,10$. Protože p -hodnota $< 0,001$ je menší než $\alpha = 0,10$, H_0 zamítáme na hladině významnosti $\alpha = 0,10$. Výška horní části tváře u mužů německé populace je statisticky významně větší než u mužů ze západopolské Cedyňe.

Poznámka: Vyskytují-li se v datovém souboru měření se stejnými hodnotami (tzv. shody, anglicky *ties*), vrací funkce `wilcox.test()` p -hodnotu modifikovanou podle algoritmu vhodného pro data se shodami. ★

Příklad 10.4. Neřešený příklad

Načtete datový soubor 02-paired-means-clavicle.txt obsahující údaje o vertikálním průměru ve středu délky těla klíční kosti z pravé a z levé strany měřené dvakrát jedním výzkumníkem (proměnné simd.1 a simd.2). Na hladině významnosti α testujte hypotézu o shodě prvního a druhého měření výzkumníka. Testování proveďte pomocí Wilcoxonova testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,05$.

Výsledky: Andersonův-Darlingův test: p -hodnota $< 0,001$, $\alpha = 0,05$; rozdíly nepochází z normálního rozložení; asymptotická varianta Wilcoxonova párového testu: $u_0 = 0,9038$, $W = (-\infty; -1,96) \cup (1,96; \infty)$; p -hodnota $= 0,3661$, $\alpha = 0,05$; H_0 nezamítáme na hladině významnosti $\alpha = 0,05$. ★

10.2 Dvouvýběrové testy

X_{11}, \dots, X_{1n_1} a X_{21}, \dots, X_{2n_2} jsou dva nezávislé náhodné výběry ze dvou spojitých rozložení s distribučními funkcemi $F(x_1)$ a $G(x_2)$. Na hladině významnosti α testujeme hypotézu, že distribuční funkce těchto rozložení jsou shodné, proti alternativě, že jsou rozdílné. Dvouvýběrové neparametrické testy jsou obdobou dvouvýběrového t -testu.

10.2.1 Dvouvýběrový Wilcoxonův test (Mannův-Whitneyův test)


U tohoto testu předpokládáme, že distribuční funkce daných dvou výběrů se mohou lišit pouze posunutím. Označme $x_{1;0,50}$ medián prvního rozložení a $x_{2;0,50}$ medián druhého rozložení. Vzhledem k tomu, že distribuční funkce se mohou lišit jen posunutím, lze hypotézu o jejich shodě vyjádřit jako hypotézu o shodě mediánů, tedy testujeme $H_0: x_{1;0,50} - x_{2;0,50} = 0$ proti $H_1: x_{1;0,50} - x_{2;0,50} \neq 0$.

Všechny hodnoty X_{11}, \dots, X_{1n_1} a X_{21}, \dots, X_{2n_2} uspořádáme vzestupně podle velikosti. Zjistíme součet pořadí hodnot X_{11}, \dots, X_{1n_1} a označíme ho T_1 . Součet pořadí hodnot X_{21}, \dots, X_{2n_2} označíme T_2 . Vypočteme statistiky $U_1 = n_1 n_2 + n_1(n_1 + 1)/2 - T_1$, $U_2 = n_1 n_2 + n_2(n_2 + 1)/2 - T_2$. Pro tyto statistiky platí vzájemný vztah: $U_2 = n_1 n_2 - U_1$. Testová statistika je rovna $U' = \min(U_1, U_2)$. H_0 zamítáme na hladině významnosti α , když testová statistika je menší nebo rovna tabelované kritické hodnotě (pro dané rozsahy výběrů n_1, n_2 a dané α).

Označení v tabulkách: $n = \min\{n_1, n_2\}$ a $m = \max\{n_1, n_2\}$. Přitom $m \in \{4, 5, \dots, 30\}$, $n \in \{2, 5, \dots, 20\}$.

Pro $n_1, n_2 > 10$ lze použít asymptotickou variantu testu: za platnosti H_0 se testová statistika $U_0 = \frac{U' - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$

asymptoticky řídí rozložením $N(0, 1)$, kde $U' = \min(U_1, U_2)$. Kritický obor pro oboustranný test má tvar: $W = (-\infty; -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}; \infty)$, dále $W = (-\infty; -u_{1-\alpha})$ pro levostranný test a $W = (u_{1-\alpha}; \infty)$ pro pravostranný test. U asymptotické varianty testu lze využít korekci na nespojitost.

Poznámka: V softwaru  funguje Wilcoxonův test jinak. Testová statistika je pro všechny tři alternativy U_2 , kritické obory mají tvar $W = (-\infty; w_{\alpha/2}(n_1, n_2)) \cup (w_{1-\alpha/2}(n_1, n_2); \infty)$ pro oboustranný test; $W = (-\infty; w_{\alpha}(n_1, n_2))$ pro levostranný test a $W = (w_{1-\alpha}(n_1, n_2); \infty)$ pro pravostranný test. Hodnota $w_{\alpha}(n_1, n_2)$ je α -kvantil rozložení, jímž se za platnosti nulové hypotézy řídí testová statistika Wilcoxonova pořadového testu pro dva náhodné výběry o rozsazích n_1 a n_2 (bez odstranění shod). Vypočítáme jej pomocí funkce `qwilcox()`.

Příklad 10.5. Řešený příklad

Načtete datový soubor 31-goldman-alaska.csv obsahující údaje o největší délce hlavice pažní kosti z levé strany (humer.HDL) u žen aljašské populace z kmene Tigara a z kmene Ipituaq. Na hladině významnosti α testujte hypotézu o shodě rozložení největší délky hlavice pažní kosti z levé strany u žen z kmene Ipituaq a u žen z kmene Tigara. Testování proveďte pomocí Wilcoxonova testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,05$.

Řešení příkladu 10.5

Řešení příkladu vede na dvouvýběrový test. Nejprve je třeba ověřit normální rozložení náhodného výběru největších délek hlavice pažní kosti z levé strany žen z kmene Ipituaq (resp. z kmene Tigara). Na hladině významnosti $\alpha = 0,05$ testujeme H_0 : *Data pochází z normálního rozložení.* oproti H_1 : *Data nepochází z normálního rozložení.* K otestování předpokladu normality použijeme vzhledem k rozsahům náhodných výběrů ($n_1 = 10$, resp. $n_2 = 24$) v obou případech Shapirův-Wilkův test.

```

37 data <- read.delim('31-goldman-alaska.csv', sep = ';', dec = '.')
38 humer.HDLI <- na.omit(data[data$pop == 'Ipituaq' & data$sex == 'f', 'humer.HDL'])
39 humer.HDLT <- na.omit(data[data$pop == 'Tigara' & data$sex == 'f', 'humer.HDL'])
40 n1 <- length(humer.HDLI) # 10
41 n2 <- length(humer.HDLT) # 24
42 shapiro.test(humer.HDLI)$p.val # 0,2027266
43 shapiro.test(humer.HDLT)$p.val # 0,04261243

```

Náhodný výběr největších délek hlavice pažní kosti z levé strany žen z kmene Ipituaq pochází z normálního rozložení (p -hodnota = 0,2027). Náhodný výběr největších délek hlavice pažní kosti z levé strany žen z kmene Tigara nepochází z normálního rozložení (p -hodnota = 0,0426).

Protože naměřené hodnoty pro ženy z kmene Tigara nepochází z normálního rozložení, použijeme na ověření zadané hypotézy v souladu se zadáním neparametrický Wilcoxonův dvouvýběrový test. Na hladině významnosti $\alpha = 0,05$ testujeme $H_0: x_{1;0,50} = x_{2;0,50}$ oproti $H_1: x_{1;0,50} \neq x_{2;0,50}$ (oboustranný test), kde $x_{1;0,50}$, resp. $x_{2;0,50}$ je medián naměřených hodnot pro ženy z kmene Ipituaq, resp. z kmene Tigara. Vzhledem k nízkému rozsahu náhodného výběru pro ženy z kmene Ipituaq ($n_1 < 20$) zvolíme exaktní variantu Wilcoxonova dvouvýběrového testu implementovanou ve funkci `wilcox.test()`. Výstupem funkce je realizace testové statistiky U_2 a p -hodnota. Hranice kritického oboru dopočítáme pomocí funkce `qwilcox()`.

```

44 wilcox.test(humer.HDLI, humer.HDLT, alternative = 'two.sided', conf.level = 0.95)
45 alpha <- 0.05
46 qwilcox(alpha / 2, n1, n2) # 68
47 qwilcox(1 - alpha / 2, n1, n2) # 172

```

```

Wilcoxon rank sum test with continuity correction
data: humer.HDLI and humer.HDLT
W = 113,5, p-value = 0,8206
alternative hypothesis: true location shift is not equal to 0

```

48
49
50
51
52
53

Realizace testové statistiky $u_2 = 113,5$, kritický obor $W = (-\infty; 68) \cup (172; \infty)$. Protože $u_2 \notin W$, H_0 nezamítáme na hladině významnosti $\alpha = 0,05$. Protože p -hodnota = 0,8206 je větší než $\alpha = 0,05$, H_0 nezamítáme na hladině významnosti $\alpha = 0,05$. Mezi největší délkou hlavice pažní kosti z levé strany u žen k kmene Ipituaq a u žen z kmene Tigara neexistuje statisticky významný rozdíl.

Poznámka: Vyskytují-li se v datovém souboru měření se stejnými hodnotami (tzv. shody, anglicky *ties*), vrací funkce `wilcox.test()` p -hodnotu modifikovanou podle algoritmu vhodného pro data se shodami. ★

Příklad 10.6. Neřešený příklad

Načtete datový soubor `32-two-samples-whr-mf.csv` obsahující údaje o poměru obvodu pasu a boků (WHR) u chlapců a dívek ve věku od šesti do šestnácti let (*age*). Na hladině významnosti α testujte hypotézu, že poměr obvodu pasu a boků u chlapců starších deseti let (včetně) je větší nebo stejný jako u chlapců mladších deseti let (vyjma). Testování proveďte pomocí Wilcoxonova testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,01$.

Výsledky: Lillieforsův test (starší): p -hodnota < 0,001, $\alpha = 0,01$; data nepochází z normálního rozložení; Andersonův-Darlingův test (mladší): p -hodnota = 0,0039, $\alpha = 0,01$; data nepochází z normálního rozložení; asymptotická varianta Wilcoxonova dvouvýběrového testu: $u_0 = -4,6198$, $W = (-\infty; -2,3263)$; p -hodnota < 0,001, $\alpha = 0,01$; H_0 zamítáme na hladině významnosti $\alpha = 0,01$. ★

10.2.2 Dvouvýběrový Kolmogorovův-Smirnovův test

U tohoto testu předpokládáme, že distribuční funkce daných dvou výběrů se mohou lišit nejen posunutím, ale i tvarem. Označme $F_{n_1}(x_1)$ výběrovou distribuční funkci prvního výběru a $G_{n_2}(x_2)$ výběrovou distribuční funkci druhého výběru. Testovou statistiku $D_{n_1, n_2} = \max_{-\infty < x_1 < \infty} |F_{n_1}(x_1) - G_{n_2}(x_2)|$ porovnáme s tabelovanou kritickou

hodnotou $D_{n_1, n_2}(\alpha)$. Hypotézu o shodě distribučních funkcí $F(x_1)$ a $G(x_2)$ zamítáme na hladině významnosti α , pokud $D_{n_1, n_2} \geq D_{n_1, n_2}(\alpha)$.

Příklad 10.7. Řešený příklad

Načtete datový soubor 31-goldman-alaska.csv obsahující údaje o největší délce hlavice pažní kosti z levé strany (humer.HDL) u žen aljašské populace z kmene Tigara a z kmene Ipituaq. Na hladině významnosti α testujte hypotézu o shodě největší délky hlavice pažní kosti z levé strany u žen z kmene Ipituaq a u žen z kmene Tigara. Testování provedte pomocí Kolmogorova-Smirnovova testu (1) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,05$.

Řešení příkladu 10.7

Zadání příkladu vede na dvouvýběrový test. Nejprve je třeba ověřit normální rozložení náhodného výběru největších délek hlavice pažní kosti z levé strany žen z kmene Ipituaq (resp. z kmene Tigara). Tento předpoklad jsme již ověřili v rámci příkladu 10.5, kde jsme došli k závěru, že náhodný výběr největších délek hlavice pažní kosti z levé strany žen z kmene Ipituaq pochází z normálního rozložení (p -hodnota = 0,2027, $\alpha = 0,05$), zatímco náhodný výběr největších délek hlavice pažní kosti z levé strany žen z kmene Tigara nepochází z normálního rozložení (p -hodnota = 0,0426, $\alpha = 0,05$).

Protože naměřené hodnoty pro ženy z kmene Tigara nepochází z normálního rozložení, použijeme na ověření zadané hypotézy v souladu se zadáním neparametrický Kolmogorův-Smirnovův test. Na hladině významnosti $\alpha = 0,05$ testujeme $H_0: F(x) = G(x)$ oproti $H_1: F(x) \neq G(x)$ (oboustranný test). Kolmogorův-Smirnovův test provedeme pomocí funkce ks.test(). Výstupem funkce je realizace testové statistiky D_{n_1, n_2} a p -hodnota.

```
54 data <- read.delim('31-goldman-alaska.csv', sep = ';', dec = '.')
55 humer.HDLI <- na.omit(data[data$pop == 'Ipituaq' & data$sex == 'f', 'humer.HDL'])
56 humer.HDLT <- na.omit(data[data$pop == 'Tigara' & data$sex == 'f', 'humer.HDL'])
57 ks.test(humer.HDLI, humer.HDLT, alternative = 'two.sided', conf.level = 0.95)
```

```
Two-sample Kolmogorov-Smirnov test

data: humer.HDLI and humer.HDLT
D = 0,21667, p-value = 0,8948
alternative hypothesis: two-sided
```

58
59
60
61
62
63

Protože p -hodnota = 0,8948 je větší než $\alpha = 0,05$, H_0 nezamítáme na hladině významnosti $\alpha = 0,05$. Mezi největší délkou hlavice pažní kosti z levé strany u žen aljašské populace z kmene Ipituaq a u žen z kmene Tigara neexistuje statisticky významný rozdíl.



Příklad 10.8. Neřešený příklad

Načtete datový soubor 10-two-samples-means-birth.txt obsahující údaje o porodní hmotnosti novorozenců narozených v jedné krajské nemocnici během jednoho roku a o počtu starších sourozenců novorozence. Na hladině významnosti α testujte hypotézu o shodě porodní hmotnosti novorozenců s žádným starším sourozencem a s jedním starším sourozencem. Testování provedte pomocí Kolmogorova-Smirnovova testu (1) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,10$.

Výsledky: Lillieforsův test (prvorození): p -hodnota = 0,0014, $\alpha = 0,10$; data nepochází z normálního rozložení; Lillieforsův test (druhorození): p -hodnota < 0,001, $\alpha = 0,10$; data nepochází z normálního rozložení; Kolmogorův-Smirnovův test: p -hodnota = 0,1780, $\alpha = 0,10$; H_0 nezamítáme na hladině významnosti $\alpha = 0,10$.



10.3 Vícevýběrové testy

Je dáno $r \geq 3$ nezávislých náhodných výběrů o rozsazích n_1, \dots, n_r . Předpokládáme, že tyto výběry pocházejí ze spojitých rozložení. Označme $n = n_1 + \dots + n_r$. Zajímá nás, zda všechny výběry pocházejí ze stejného rozložení resp. zda mediány všech r rozložení jsou stejné.

10.3.1 Kruskalův-Wallisův test

Předpokládáme, že distribuční funkce daných r rozložení se mohou lišit jen posunutím. Na asymptotické hladině významnosti α testujeme hypotézu, že všechny tyto výběry pocházejí z téhož rozložení proti alternativě, že aspoň jedna dvojice výběrů pochází z různých rozložení. Kruskalův-Wallisův test je neparametrickou obdobou analýzy rozptylu jednoduchého třídění. Všechny n hodnot seřadíme do rostoucí posloupnosti a určíme pořadí každé hodnoty v tomto sdruženém výběru. Označme T_j součet pořadí těch hodnot, které patří do j -tého výběru, $j = 1, \dots, r$ (kontrola: musí platit $T_1 + \dots + T_r = n(n+1)/2$). Testová statistika: $Q = \frac{12}{n(n+1)} \sum_{j=1}^r \frac{T_j^2}{n_j} - 3(n+1)$. Platí-li H_0 , má statistika Q asymptoticky rozložení $\chi^2(r-1)$. Kritický obor má tvar $W = (\chi_{1-\alpha}^2(r-1); \infty)$, přičemž H_0 zamítáme na asymptotické hladině významnosti α , když $Q \in W$.

Při zamítnutí H_0 je zapotřebí zjistit, které dvojice náhodných výběrů se liší na zvolené hladině významnosti. Testujeme H_0 : k -tý a l -tý náhodný výběr pocházejí z téhož rozložení, $k, l = 1, \dots, r$, $k \neq l$. proti H_1 : Alespoň jedna dvojice výběrů pochází z různých rozložení. K tomu slouží neparametrické metody mnohonásobného porovnávání, kterých je celá řada, např. Neményiho metoda (pro tzv. vyvážené třídění), obecná metoda, metoda párového porovnávání s korekcí na mnohonásobné testování atd.

Příklad 10.9. Řešený příklad

Načtete datový soubor 19-more-samples-correlations-skull.txt obsahující údaje o šířce nosu (nose.B) mužů bantuské, čínské, malajské, německé a peruánské populace. Na hladině významnosti α testujte hypotézu o shodě šířky nosu mužů čínské, malajské a peruánské populace. Testování proveďte pomocí Kruskalova-Wallisova testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,01$. V případě zamítnutí nulové hypotézy zjistíte pomocí Neményiho metody, které dvojice populací se v šířce nosu liší.

Řešení příkladu 10.9

Zadání příkladu vede na vícevýběrový test. Nejprve je třeba ověřit normální rozložení náhodného výběru šířky nosu mužů čínské (resp. malajské, resp. peruánské populace). Na hladině významnosti $\alpha = 0,01$ testujeme H_0 : Data pochází z normálního rozložení. oproti H_1 : Data nepochází z normálního rozložení. K otestování předpokladu normality použijeme vzhledem k rozsahům náhodných výběrů ($n_1 = 19$, resp. $n_2 = 73$, resp. $n_3 = 46$) Shapirov-Wilkův test pro čínskou populaci a Andersonův-Darlingův test pro malajskou (resp. peruánskou) populaci.

```
64 data <- read.delim('19-more-samples-correlations-skull.txt', sep = '\t', dec = '.')
65 nose.BC <- na.omit(data[data$pop == 'cin', 'nose.B'])
66 nose.BM <- na.omit(data[data$pop == 'mal', 'nose.B'])
67 nose.BP <- na.omit(data[data$pop == 'per', 'nose.B'])
68 ni <- c(length(nose.BC), length(nose.BM), length(nose.BP)) # 19; 73; 46
69 shapiro.test(nose.BC)$p.val # 0,1173442
70 nortest::ad.test(nose.BM)$p.val # 0,04117326
71 nortest::ad.test(nose.BP)$p.val # 0,004271102
```

Náhodný výběr šířek nosu mužů čínské populace pochází z normálního rozložení (p -hodnota = 0,1173). Náhodný výběr šířek nosu mužů malajské populace pochází z normálního rozložení (p -hodnota = 0,0412). Náhodný výběr šířek nosu mužů peruánské populace nepochází z normálního rozložení (p -hodnota = 0,0043).

Protože naměřené hodnoty pro muže peruánské populace nepochází z normálního rozložení, použijeme na ověření zadané hypotézy v souladu se zadáním neparametrický Kruskalův-Wallisův test. Na hladině významnosti $\alpha = 0,01$ testujeme H_0 : Všechny tři výběry pochází ze stejného rozložení. oproti H_1 : Alespoň jedna dvojice výběrů pochází z různých rozložení. Kruskalův-Wallisův test provedeme pomocí funkce `kruskal.test()`. Výstupem funkce je realizace testové statistiky Q a p -hodnota. Hranici kritického oboru dopočítáme pomocí funkce `qchisq()`.

```
72 nose.B3 <- c(nose.BC, nose.BM, nose.BP)
73 pop <- rep(1:3, ni)
74 kruskal.test(nose.B3, pop)
75 alpha <- 0.01
76 r <- 3
```



```
77 qchisq(1 - alpha, r - 1) # 9,21034
```

```
Kruskal-Wallis rank sum test

data: nose.B3 and pop
Kruskal-Wallis chi-squared = 50,233, df = 2, p-value = 1,236e-11
```

78
79
80
81
82

Realizace testové statistiky $q = 50,2330$, kritický obor $W = (9,2103; \infty)$. Protože $q \in W$, H_0 zamítáme na asymptotické hladině významnosti $\alpha = 0,01$. Protože p -hodnota $< 0,001$ je menší než $\alpha = 0,01$, H_0 zamítáme na asymptotické hladině významnosti $\alpha = 0,01$. Alespoň jedna dvojice náhodných výběrů pochází z různých rozložení.

Nyní zjistíme, které dvojice náhodných výběrů (populací) se v šířce nosu liší. K tomu použijeme Neményiho metodu mnohonásobného porovnávání implementovanou ve funkci `NemenyiTest()` z knihovny `DescTools`.

```
83 DescTools::NemenyiTest(x = nose.B3, g = factor(pop), dist = 'chisq')
84 # Nemenyi's test of multiple comparisons for independent samples (chisq)
85 #      mean.rank.diff      pval
86 # 2-1      14,23107  0,3911
87 # 3-1     -38,45252  0,0022 **
88 # 3-2     -52,68359  3,5e-11 ***
89 # ---
90 # Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
```

Na hladině významnosti $\alpha = 0,01$ se statisticky významně liší šířka nosu mužů peruánské a čínské populace (p -hodnota = 0,0022) a šířka nosu mužů peruánské a malajské populace (p -hodnota $< 0,001$). Mezi šířkou nosu mužů čínské a malajské populace neexistuje statisticky významný rozdíl (p -hodnota = 0,3911). ★

Příklad 10.10. Neřešený příklad

Načtete datový soubor `31-goldman-alaska.csv` obsahující údaje o délce vřetenní kosti z pravé strany (`radius.R`) a z levé strany (`radius.L`) žen z kmene Tigara, Ipituaq a z oblasti Yukon. Na hladině významnosti α testujte hypotézu o shodě délky vřetenní kosti z pravé strany u žen ze všech tří aljašských populací. Testování proveďte pomocí Kruskalova-Wallisova testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,05$. V případě zamítnutí nulové hypotézy zjistěte pomocí Neményiho metody, které dvojice populací se v délce vřetenní kosti z pravé strany liší.

Výsledky: Shapirův-Wilkův test (Tigara): p -hodnota = 0,0255, $\alpha = 0,05$; data nepochází z normálního rozložení; Shapirův-Wilkův test (Ipituaq): p -hodnota = 0,6489, $\alpha = 0,05$; data pochází z normálního rozložení; Shapirův-Wilkův test (Yukon): p -hodnota = 0,8442, $\alpha = 0,05$; data pochází z normálního rozložení; Kruskalův-Wallisův test: $q = 3,2278$, $W = (5,9915; \infty)$; p -hodnota = 0,1991, $\alpha = 0,05$; H_0 nezamítáme na asymptotické hladině významnosti $\alpha = 0,05$. ★

10.3.2 Mediánový test

Na rozdíl od Kruskalova-Wallisova testu mediánový test nepřepokládá, že distribuční funkce daných r rozložení se mohou lišit jen posunutím, mohou se tedy lišit i tvarem. Na asymptotické hladině významnosti α testujeme hypotézu H_0 : *Mediány všech r rozložení jsou shodné.* oproti alternativě H_1 : *Aspoň jedna dvojice mediánů se liší.*

Všech n hodnot uspořádáme do rostoucí posloupnosti a najdeme medián $x_{0,50}$ těchto n hodnot. Označme P_j počet hodnot v j -tém výběru, které jsou větší nebo rovny mediánu $x_{0,50}$.

Testová statistika: $Q_M = 4 \sum_{j=1}^r \frac{P_j^2}{n_j} - n$. Platí-li H_0 , má statistika Q_M asymptoticky rozložení $\chi^2(r-1)$. Kritický obor je stejný jako u Kruskalova-Wallisova testu. Dojde-li k zamítnutí H_0 , pomocí neparametrické metody mnohonásobného porovnávání identifikujeme dvojice rozložení, které se liší v mediánech.

Příklad 10.11. Řešený příklad

Načtete datový soubor `15-anova-means-skull.txt` obsahující údaje o výšce horní části tváře (`upface.H`) mužů bantuské, čínské, malajské, německé a peruánské populace. Na hladině významnosti α testujte hypotézu o shodě výšky horní části tváře mužů bantuské, čínské a německé populace. Testování proveďte pomocí mediánového testu (1)

kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,10$. V případě zamítnutí nulové hypotézy zjistíte pomocí Neményiho metody, které dvojice populací se ve výšce horní části tváře liší.

Řešení příkladu 10.11

Zadání příkladu vede na vícevýběrový test. Nejprve je třeba ověřit normální rozložení náhodného výběru výšky horní části tváře mužů bantuské (resp. čínské, resp. německé populace). Na hladině významnosti $\alpha = 0,10$ testujeme H_0 : *Data pochází z normálního rozložení*. oproti H_1 : *Data nepochází z normálního rozložení*. K otestování předpokladu normality použijeme vzhledem k rozsahům náhodných výběrů ($n_1 = 13$, resp. $n_2 = 18$, resp. $n_3 = 19$) pro každou populaci Shapirův-Wilkův test.

```
91 data <- read.delim('15-anova-means-skull.txt', sep = '\t', dec = '.')
92 upface.HB <- na.omit(data[data$pop == 'ban', 'upface.H'])
93 upface.HC <- na.omit(data[data$pop == 'cin', 'upface.H'])
94 upface.HN <- na.omit(data[data$pop == 'nem', 'upface.H'])
95 ni <- c(length(upface.HB), length(upface.HC), length(upface.HN))
96 shapiro.test(upface.HB)$p.val # 0,4321368
97 shapiro.test(upface.HC)$p.val # 0,05127186
98 shapiro.test(upface.HN)$p.val # 0,04190113
```

Náhodný výběr výšek horní části tváře mužů bantuské populace pochází z normálního rozložení (p -hodnota = 0,4321). Náhodný výběr výšek horní části tváře mužů čínské populace nepochází z normálního rozložení (p -hodnota = 0,0513). Náhodný výběr výšek horní části tváře mužů německé populace nepochází z normálního rozložení (p -hodnota = 0,0419).

Protože naměřené hodnoty pro muže čínské a německé populace nepochází z normálního rozložení, použijeme na ověření zadané hypotézy v souladu se zadáním neparametrický mediánový test. Na hladině významnosti $\alpha = 0,10$ testujeme H_0 : *Mediány všech tří výběrů jsou shodné*. oproti H_1 : *Alespoň jedna dvojice mediánů se liší*. Mediánový test provedeme pomocí funkce `Median.test()` z knihovny `agricolae`. Výstupem funkce je realizace testové statistiky Q_M a p -hodnota. Hranici kritického oboru dopočítáme pomocí funkce `qchisq()`.

```
99 upface.HBCN <- c(upface.HB, upface.HC, upface.HN)
100 pop <- rep(1:3, ni)
101 agricolae::Median.test(upface.HBCN, trt = pop, console = F)$stat # 1,165992
102 alpha <- 0.10
103 r <- 3
104 qchisq(1 - alpha, r - 1) # 4,60517
```

Chisq	Df	p.chisq	Median
1,165992	2	0,5582234	72,5

105
106

Realizace testové statistiky $q_M = 1,1660$, kritický obor $W = \langle 4,6052; \infty \rangle$. Protože $q_M \notin W$, H_0 nezamítáme na asymptotické hladině významnosti $\alpha = 0,10$. Protože p -hodnota = 0,5582 je větší než $\alpha = 0,10$, H_0 nezamítáme na asymptotické hladině významnosti $\alpha = 0,10$. Mezi mediány výšek horní části tváře mužů bantuské, čínské a malajské populace neexistuje statisticky významný rozdíl.

Poznámka: Vyskytují-li se v datovém souboru měření se stejnými hodnotami (tzv. shody, anglicky *ties*), vrací funkce `Median.test()` testovou statistiku i p -hodnotu modifikovanou podle algoritmu vhodného pro data se shodami. ★

Příklad 10.12. Neřešený příklad

N načtete datový soubor `19-more-samples-correlations-skull.txt` obsahující údaje o interorbitální šířce nosu (`intorb.B`) mužů bantuské, čínské, německé, malajské a peruánské populace. Na hladině významnosti α testujte hypotézu o shodě rozložení interorbitální šířky nosu mužů všech pěti populací. Testování proveďte pomocí mediánového testu (1) kritickým oborem; (2) p -hodnotou. Hladinu významnosti zvolte $\alpha = 0,05$. V případě zamítnutí nulové hypotézy zjistíte pomocí Neményiho metody, které dvojice populací se v interorbitální šířce nosu liší.

Výsledky: Shapirův-Wilkův test (bantuská populace): p -hodnota = 0,4537, $\alpha = 0,05$; data pochází z normálního rozložení; Shapirův-Wilkův test (čínská populace): p -hodnota = 0,3475, $\alpha = 0,05$; data pochází z normálního

rozložení; Shapirův-Wilkův test (německá populace): p -hodnota = 0,3387, $\alpha = 0,05$; data pochází z normálního rozložení; Andersonův-Darlingův test (malajská populace): p -hodnota = 0,0260, $\alpha = 0,05$; data nepochází z normálního rozložení; Andersonův-Darlingův test (peruánská populace): p -hodnota = 0,0221, $\alpha = 0,05$; data nepochází z normálního rozložení; mediánový test: $q_M = 20,1908$, $W = \langle 9,4877; \infty \rangle$; p -hodnota < 0,001, $\alpha = 0,05$; H_0 zamítáme na asymptotické hladině významnosti $\alpha = 0,05$; Neményiho metoda: rozdíl existuje mezi mediány (a) bantuské a malajské populace (p -hodnota = 0,0154); (b) bantuské a peruánské populace (p -hodnota = 0,0425). ★