

12 Dvouvýběrové neparametrické testy

12.1 Wilcoxonův dvouvýběrový test (Mannův-Whitneyův U test) - exaktní varianta

Nechť X_{11}, \dots, X_{1n_1} , $n_1 \geq 2$ je náhodný výběr ze spojitého rozdělení s distribuční funkcí $F_1(x)$ a X_{21}, \dots, X_{2n_2} , $n_2 > 2$ je na něm nezávislý náhodný výběr ze spojitého rozdělení s distribuční funkcí $F_2(x)$. O distribučních funkcích $F_1(x)$ a $F_2(x)$ předpokládáme, že se navzájem liší pouze posunutím. Nechť dále \tilde{x}_1 je medián náhodného výběru X_{11}, \dots, X_{1n_1} a \tilde{x}_2 je medián náhodného výběru X_{21}, \dots, X_{2n_2} a \tilde{x}_0 je konstanta. Na hladině významnosti α testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

$$\begin{array}{lll} H_{01} : \tilde{x}_1 - \tilde{x}_2 = \tilde{x}_0 & \text{oproti} & H_{11} : \tilde{x}_1 - \tilde{x}_2 \neq \tilde{x}_0 \quad (\text{oboustranná alt.}) \\ H_{02} : \tilde{x}_1 - \tilde{x}_2 \leq \tilde{x}_0 & \text{oproti} & H_{12} : \tilde{x}_1 - \tilde{x}_2 > \tilde{x}_0 \quad (\text{pravostranná alt.}) \\ H_{03} : \tilde{x}_1 - \tilde{x}_2 \geq \tilde{x}_0 & \text{oproti} & H_{13} : \tilde{x}_1 - \tilde{x}_2 < \tilde{x}_0 \quad (\text{levostranná alt.}) \end{array}$$

Název testu není v tomto případě jednotný. Některé zdroje uvádí název Wilcoxonův dvouvýběrový test, jiné zdroje používají název Mannův-Whitneyův U test. Oba názvy jsou si rovnocenné a je velmi důležité mít na paměti, že za oběma se skrývá stejný test. V tomto textu budeme preferovat název Wilcoxonův dvouvýběrový test (konkrétně jeho exaktní varianta). Tomu odpovídá i zvolené značení testovací statistiky, která v případě Wilcoxonova dvouvýběrového testu bývá označována jako W nebo S_E , zatímco v případě Mannova-Whitneyova U testu bývá značena jako U .

$$S_E = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2 \quad (12.1)$$

kde n_1 je rozsah prvního náhodného výběru, n_2 je rozsah druhého náhodného výběru a $T_2 = \sum_{j=1}^{n_2} S_j$, kde S_j , $j = 1, \dots, n_2$, je pořadí náhodných veličin X_{21}, \dots, X_{2n_2} druhého náhodného výběru v seřazeném vektoru všech $n_1 + n_2$ hodnot z obou náhodných výběrů. Statistika T_2 je tedy součet pořadí hodnot X_{21}, \dots, X_{2n_2} , v seřazeném vektoru všech $n_1 + n_2$ hodnot. Kritický obor podle zvolené alternativní hypotézy má tvar

$$\begin{array}{ll} H_{11} : \tilde{x}_1 - \tilde{x}_2 \neq \tilde{x}_0 & W = (-\infty; w_{n_1, n_2}(\alpha/2)) \cup \langle w_{n_1, n_2}(1 - \alpha/2); \infty \rangle \\ H_{12} : \tilde{x}_1 - \tilde{x}_2 > \tilde{x}_0 & W = \langle w_{n_1, n_2}(1 - \alpha); \infty \rangle \\ H_{13} : \tilde{x}_1 - \tilde{x}_2 < \tilde{x}_0 & W = (-\infty; w_{n_1, n_2}(\alpha)) \end{array}$$

kde n_1 je rozsah prvního náhodného výběru, n_2 je rozsah druhého náhodného výběru a $T_2 = \sum_{j=1}^{n_2} S_j$, tj. T_2 je součet pořadí S_j , $j = 1, \dots, n_2$, náhodných veličin X_{21}, \dots, X_{2n_2} druhého náhodného výběru v seřazeném vektoru všech $n_1 + n_2$ hodnot z obou náhodných výběrů. Statistika W je tedy součet pořadí hodnot X_{21}, \dots, X_{2n_2} , v seřazeném vektoru všech $n_1 + n_2$ hodnot. Kritický obor podle zvolené alternativní hypotézy má tvar kde $w_{n_1, n_2}(\alpha/2)$, $w_{n_1, n_2}(1 - \alpha/2)$, $w_{n_1, n_2}(\alpha)$, $w_{n_1, n_2}(1 - \alpha)$ jsou tabelované kvantily pro dvouvýběrový Wilcoxonův test, jejichž hodnoty získáme příkazem `qwilcox()`.

Interval spolehlivosti má podle zvolené alternativní hypotézy jeden z následujících tvarů

$$\begin{array}{ll} H_{11} : \tilde{x}_1 - \tilde{x}_2 \neq \tilde{x}_0 & (d, h) = (U(w_{n_1, n_2}(\alpha/2)); U(w_{n_1, n_2}(1 - \alpha/2) + 1)) \\ H_{12} : \tilde{x}_1 - \tilde{x}_2 > \tilde{x}_0 & (d, \infty) = (U(w_{n_1, n_2}(\alpha)); \infty) \\ H_{13} : \tilde{x}_1 - \tilde{x}_2 < \tilde{x}_0 & (-\infty, h) = (-\infty; U(w_{n_1, n_2}(1 - \alpha) + 1)) \end{array}$$

kde $U^{(1)} \leq \dots \leq U^{(n_1 n_2)}$ značí vzestupně seřazené rozdíly $X_{2j} - X_{1i}$, $i = 1, \dots, n_1$, $j = 1, \dots, n_2$, a $U^{(x)}$ značí x -tý seřazený rozdíl.

p -hodnota má v závislosti na zvolené alternativní hypotéze jeden z následujících tvarů

$$\begin{array}{ll} H_{11} : \tilde{x}_1 - \tilde{x}_2 \neq \tilde{x}_0 & p\text{-hodnota} = 2 \min\{\Pr(S_E \leq s_E), \Pr(S_E \geq s_E)\} \\ H_{12} : \tilde{x}_1 - \tilde{x}_2 > \tilde{x}_0 & p\text{-hodnota} = \Pr(S_E \geq s_E) \\ H_{13} : \tilde{x}_1 - \tilde{x}_2 < \tilde{x}_0 & p\text{-hodnota} = \Pr(S_E \leq s_E) \end{array}$$

kde S_E je náhodná veličina, s_E je realizace testovací statistiky S_E (viz vzorec 12.1), tedy konkrétní číslo, a $\Pr(S_E \leq s_E)$ je distribuční funkce tabelovaného rozdělení pro Wilcoxonův dvouvýběrový exaktní test, jejíž hodnotu

získáme pomocí  a implementované funkce `pwilcox()`.

```
1 # ZKONTROLOVAT BARVY
```

Příklad 12.1. Wilcoxonův dvouvýběrový exaktní test (oboustranná alternativa)

Mějme datový soubor 21-goldman-measures.txt obsahující údaje o délce holenní kosti z pravé strany (tibia.LR) a z levé strany (tibia.LL) u mužů a žen z dvou japonských populací Tsugumo Shell Mound a Yoshigo Shell Mound (viz sekce ??). Na hladině významnosti $\alpha = 0.05$ testujte hypotézu o shodě délky holenní kosti z levé strany u mužů z populace Tsugumo Shell Mound a u mužů z populace Yoshigo Shell Mound.

Řešení příkladu 12.1

Datový soubor 21-goldman-measures.txt načteme příkazem `read.delim()`. Následně z načtených dat vybereme naměřené hodnoty délky holenní kosti z levé strany (tibia.LL) u mužů (`pop == 'm'`) z japonské populace Tsugumo Shell Mound (`pop == 'Tsugumo Shell Mound'`) a vložíme je do proměnné `tibia.LLMT`. Dále z datové tabulky vybereme naměřené hodnoty délky holenní kosti z levé strany (tibia.LL) u mužů (`pop == 'm'`) z japonské populace Yoshigo Shell Mound (`pop == 'Yoshigo Shell Mound'`) a vložíme je do proměnné `tibia.LLMY`. Příkazem `na.omit()` odstraníme z vektorů `tibia.LLMT` a `tibia.LLMY` chybějící pozorování, příkazem `length()` zjistíme rozsahy obou náhodných výběrů a příkazem `range()` rozpětí naměřených hodnot v obou náhodných výběrech.

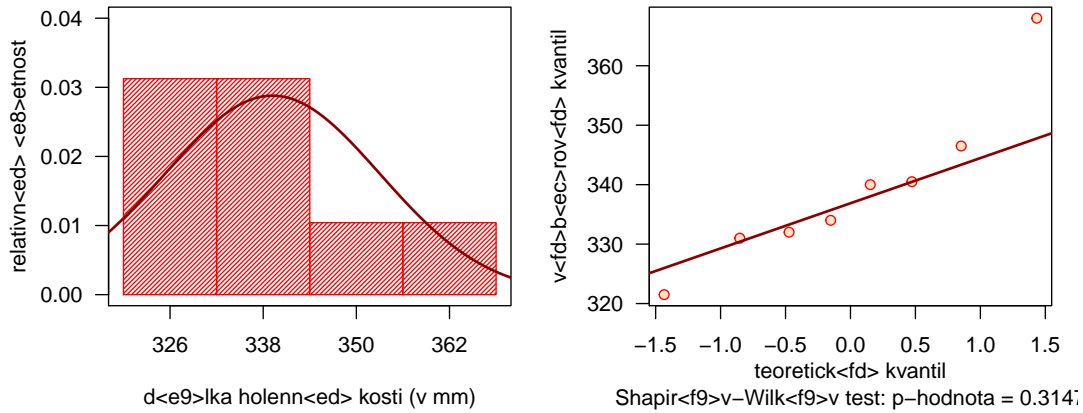
```
2 data <- read.delim('00-Data//21-goldman-measures.csv', sep = ';', dec = '.')
3 tibia.LLMT <- data[data$sex == 'm' & data$pop == 'Tsugumo Shell Mound', 'tibia.LL']
4 tibia.LLMY <- data[data$sex == 'm' & data$pop == 'Yoshigo Shell Mound', 'tibia.LL']
5 tibia.LLMT <- na.omit(tibia.LLMT)
6 tibia.LLMY <- na.omit(tibia.LLMY)
7
8 n1 <- length(tibia.LLMT) # 8
9 n2 <- length(tibia.LLMY) # 7
10 range(tibia.LLMT) # 321.5-368.0
11 range(tibia.LLMY) # 320-356.5
```

Datový soubor obsahuje naměřené hodnoty délky holenní kosti z levé strany u 8 mužů z japonské populace Tsugumo Shell Mound a u 7 mužů z japonské populace Yoshigo Shell Mound. Naměřené délky holenní kosti u mužů z populace Tsugumo Shell Mound nabývají hodnot v rozsahu 321.5–368.0 mm, naměřené délky holenní kosti u mužů z populace Yoshigo Shell Mound nabývají hodnot v rozsahu 320.0–356.5 mm.

Naším úkolem je porovnat střední hodnoty dvou japonských populací, přičemž u obou populací máme k dispozici naměřené hodnoty. Řešení příkladu vede na parametrický test o rozdílu středních hodnot $\mu_1 - \mu_2$ (viz sekce ?? nebo sekce ??). Před provedením tohoto parametrického testu však musíme ověřit, zda oba náhodné výběry pocházejí z normálních rozdělení. Na hladině významnosti $\alpha = 0.05$ testujeme nejprve hypotézu H_{01} : *Náhodný výběr délek holenních kostí z levé strany u mužů z populace Tsugumo Shell Mound pochází z normálního rozdělení*, oproti alternativní hypotéze: H_{11} : *Náhodný výběr délek holenních kostí z levé strany u mužů z populace Tsugumo Shell Mound nepochází z normálního rozdělení*. Protože rozsah náhodného výběru délek holenních kostí je roven 8, což je méně než 30, použijeme na otestování hypotézy o normalitě náhodného výběru Shapirov-Wilkův test. Graficky zhodnotíme normalitu náhodného výběru pomocí histogramu superponovaného křivkou normálního rozdělení a QQ-diagramem (viz obrázek 1). Pro účely histogramu rozdělíme náhodný výběr délek holenních kostí do 4 ekvidistantních třídicích intervalů o šířce 12 mm pomocí stanovených hranic 320, 332, ..., 368 mm.

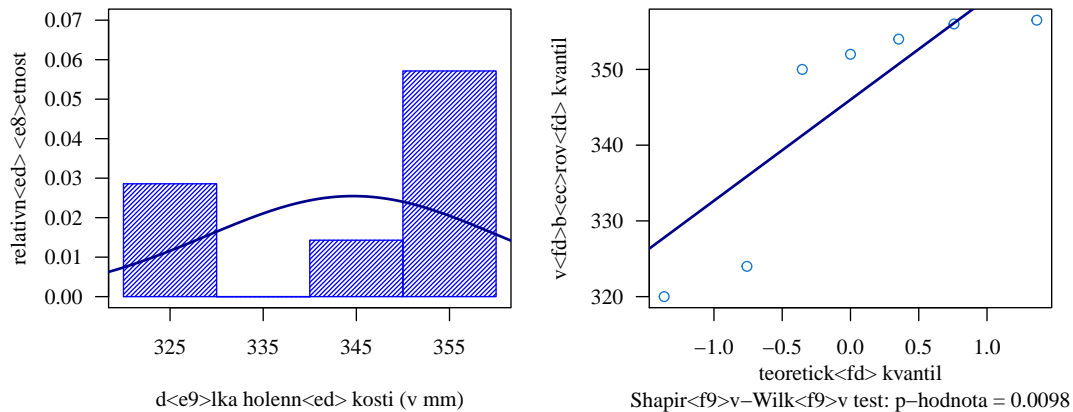
Protože p -hodnota Shapirova-Wilkova testu $p = 0.3147$ je větší než 0.05, hypotézu H_0 nezamítáme na hladině významnosti $\alpha = 0.05$. Při takto malém rozsahu náhodného výběru ($n_1 = 8$) je velmi těžké rozpoznat normálních charakter dat na základě grafické vizualizace. Porušení normality by při takto malém rozsahu náhodného výběru mohlo způsobit maximálně tak nějaké extrémně odlehle pozorování, nebo rozdělení datového souboru do dvou shluků, které by od sebe byly dostatečně vzdálené. Nicméně v histogramu ani v QQ-diagramu žádné z těchto porušení normality nepozorujeme. Náhodný výběr délek holenní kosti z levé strany u mužů z populace Tsugumo Shell Mound pochází z normálního rozdělení.

Analogicky testujeme hypotézu H_{02} : *Náhodný výběr délek holenních kostí z levé strany u mužů z populace Yoshigo Shell Mound pochází z normálního rozdělení*, oproti alternativní hypotéze: H_{12} : *Náhodný výběr délek holenních kostí z levé strany u mužů z populace Yoshigo Shell Mound nepochází z normálního rozdělení*. Protože rozsah náhodného výběru délek holenních kostí ($n_2 = 7$) je menší než 30, použijeme na otestování hypotézy o normalitě náhodného



Obrázek 1: Histogram (vlevo) a kvantilový diagram (vpravo) délky holenní kosti z levé strany u mužů z populace Tsugumo Shell Mound

výběru opět Shapirov-Wilkův test. Graficky zhodnotíme normalitu náhodného výběru pomocí histogramu a QQ-diagramu (viz obrázek 2). Pro účely histogramu rozdělíme náhodný výběr délek holenních kostí do 4 ekvidistantních intervalů o šířce 10 mm pomocí stanovených hranic 320, 330, ..., 360 mm.



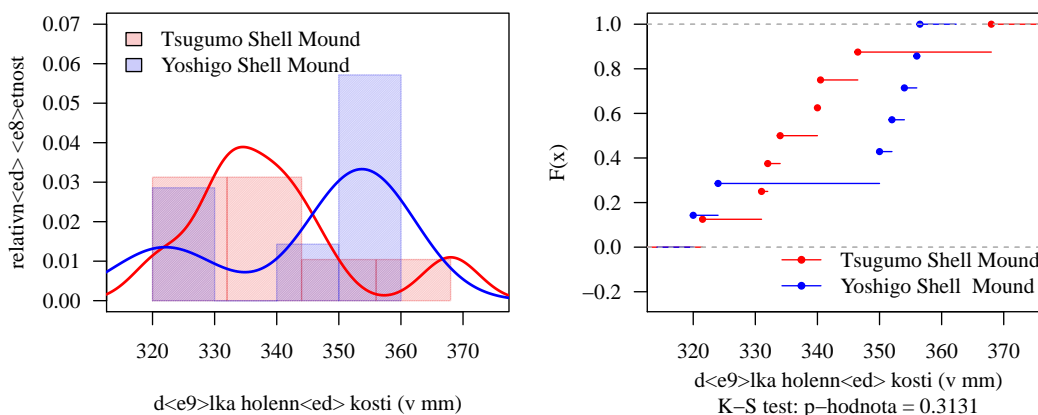
Obrázek 2: Histogram (vlevo) a kvantilový diagram (vpravo) délky holenní kosti z levé strany u mužů z populace Yoshigo Shell Mound

Protože p -hodnota Shapirova-Wilkova testu $p = 0.0098$ je menší než 0.05 , hypotézu H_0 zamítáme na hladině významnosti $\alpha = 0.05$. V grafu histogramu jsou viditelné dva shluky dat, které jsou od sebe poměrně vzdálené. Toto odloučení je pravděpodobně důvodem porušení normality náhodného výběru. Porušení normality je viditelné také v QQ-diagramu, kde se vykreslené body při sebelepší vůli nedrží referenční přímky. Náhodný výběr délek holenní kosti z levé strany u mužů z populace Yoshigo Shell Mound nepochází z normálního rozdělení.

Protože náhodný výběr délek holenních kostí z levé strany u mužů z populace Yoshigo Shell Mound nepochází z normálního rozdělení, nemůžeme hypotézu ze zadání příkladu otestovat pomocí parametrického testu o rozdílu středních hodnot $\mu_1 - \mu_2$. K otestování hypotézy musíme použít neparametrický test. Vzhledem k nízkým rozsahům obou náhodných výběrů otestujeme hypotézu ze zadání pomocí Wilcoxonova dvouvýběrového exaktního testu. Tento test nevyžaduje, aby byl u obou náhodných výběrů splněn předpoklad normálního rozdělení (de facto u tohoto testu nemusí ani jeden náhodný výběr pocházet z normálního rozdělení). Jediným předpokladem, který musí být pro použití Wilcoxonova dvouvýběrového testu splněn, je shoda distribučních funkcí $F_1(x)$ a $F_2(x)$ (s výjimkou posunutí o vzdálenost Δ), kde $F_1(x)$ je distribuční funkce náhodného výběru z populace Tsugumo Shell Mound a $F_2(x)$ je distribuční funkce náhodného výběru z populace Yoshigo Shell Mound. Shodu distribučních funkcí

$F_1(x)$ a $F_2(x)$ ověříme pomocí Kolmogorova-Smirnova testu (viz sekce ??).

Oba náhodné výběry nejprve metodou centrování posuneme do nuly, čímž odfiltrujeme vliv posunutí o vzdálenost Δ mezi oběma distribučními funkcemi. Na hladině významnosti $\alpha = 0.05$ potom testujeme hypotézu H_{03} : *Distribuční funkce centrované délky holenní kosti z levé strany u mužů z populace Tsugumo Shell Mound a z populace Yoshigo Shell Mound jsou shodné.* oproti alternativní hypotéze: H_{13} : *Distribuční funkce centrované délky holenní kosti z levé strany u mužů z populace Tsugumo Shell Mound a z populace Yoshigo Shell Mound nejsou shodné.* Ze sekce ?? již víme, že nulová hypotéza H_{03} odpovídá hypotéze o shodě dvou distribučních funkcí původních náhodných výběrů s výjimkou posunutí o vzdálenost Δ . Kolmogorův-Smirnovův test provedeme pomocí funkce `ks.test()` analogicky jako v sekci ??. Graficky případnou podobnost distribučních funkcí $F_1(x)$ a $F_2(x)$ znázorníme dvěma histogramy superponovanými příslušnými jádrovými odhady hustot a grafem zobrazujícím výběrové distribuční funkce $F_1(x)$ a $F_2(x)$ (viz obrázek 3).



Obrázek 3: Porovnání histogramů (vlevo) a graf distribučních funkcí (vpravo) délky holenní kosti z levé strany u mužů z populací Tsugumo Shell Mound a Yoshigo Shell Mound

Protože p -hodnota K-S testu $p = 0.3131$ je větší než 0.05, nulovou hypotézu o shodě dvou distribučních funkcí centrovaných náhodných výběrů nezamítáme na hladině významnosti $\alpha = 0.05$. To tedy znamená, že distribuční funkce náhodného výběru délek holenních kostí z levé strany u mužů z populace Tsugumo Shell Mound a z populace Yoshigo Shell Mound se liší pouze posunutím.

Předpoklad umožňující otestování hypotézy uvedené v zadání pomocí Wilcoxonova dvouvýběrového testu je splněn. Můžeme tedy přistoupit k samotnému procesu testování hypotézy ze zadání. Naším úkolem je otestovat hypotézu o shodě délky holenní kosti z levé strany u mužů z populace Tsugumo Shell Mound a z populace Yoshigo Shell Mound. Jak již bylo několikrát naznačeno, věta ze zadání je zněním nulové hypotézy, a to jednak proto, že v zadání se přímo vyskytuje slovo hypotéza, které je synonymem pro nulovou hypotézu, a dále proto, že v zadání mluvíme o shodě, která je synonymem rovnosti a rovnost je vždy součástí nulové hypotézy. Zbývá dodefinovat znění alternativní hypotézy tak, aby bylo přesným doplňkem znění nulové hypotézy. Závěrem poznamenejme, že zatímco u parametrických testů (viz kapitola ??) figuruje v hypotézách pojem *střední hodnota*, u neparametrických testů tento pojem nahrazujeme neparametrickým ekvivalentem, kterým je *medián*. Testování provedeme v posloupnosti sedmi kroků.

1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

H_0 : *Medián délky holenní kosti z levé strany u mužů z populace Tsugumo Shell Mound je shodný s mediánem délky holenní kosti z levé strany u mužů z populace Yoshigo Shell Mound.*

H_1 : *Medián délky holenní kosti z levé strany u mužů z populace Tsugumo Shell Mound není shodný s mediánem délky holenní kosti z levé strany u mužů z populace Yoshigo Shell Mound.*

- **matematická formulace** nulové a alternativní hypotézy

$$H_0 : \tilde{x}_1 = \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 = \tilde{x}_0, \text{ kde } \tilde{x}_0 = 0$$

$$H_0 : \tilde{x}_1 \neq \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 \neq \tilde{x}_0, \text{ kde } \tilde{x}_0 = 0 \text{ (oboustranná alternativa)}$$

2. Volba hladiny významnosti

- Hladinu významnosti volíme v souladu se zadáním jako $\alpha = 0.05$.

3. Testování kritickým oborem

• Testovací statistika

Nejprve vytvoříme společný vektor délek holenní kosti z levé strany u mužů z obou populací tak, že za sebe poskládáme hodnoty délky holenní kosti naměřené na mužích z populace Tsugumo Shell Mound (X_{1i} , $i = 1, \dots, 8$) a následně na mužích z populace Yoshigo Shell Mound (X_{2j} , $j = 1, \dots, 7$). Délka společného vektoru $n = n_1 + n_2 = 8 + 7 = 15$. Následně každému pozorování přiřadíme hodnotu indikační funkce I , která nabývá hodnoty 1, pokud měření pochází z druhého náhodného výběru, nebo hodnoty 0, pokud měření pochází z prvního náhodného výběru. Nakonec stanovíme pořadí S_k , $k = 1, \dots, 15$ všech 15 hodnot od nejmenší po největší (viz tabulka 1).


Tabulka 1: Naměřené hodnoty X_{1i} , $i = 1, \dots, 8$, a X_{2j} , $j = 1, \dots, 7$, obou náhodných výběrů, indikační funkce I příslušnosti k druhému náhodnému výběru a pořadí S_k , $k = 1, \dots, 15$, naměřených hodnot seřazených vzestupně

měření	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_{1i}, X_{2j}	340	368	321.5	334	332	340.5	346.5	331	320	356.5	354	350	356	352	324
I	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
S_k	7	15	2	6	5	8	9	4	1	14	12	10	13	11	3

Z tabulky 1 vidíme, že sedmi hodnotám příslušícím druhému náhodnému výběru patří ve vektoru patnácti seřazených hodnot pořadí 1, 14, 12, 10, 13, 11 a 3. Statistiku T_2 získáme sečtením těchto pořadí. Hodnotu testovací statistiky S_E potom získáme dosazením hodnot n_1 , n_2 a statistiky T_2 do vzorce 12.1.

$$T_2 = \sum_{j=1}^{n_2} S_j = 1 + 14 + 12 + 10 + 13 + 11 + 3 = 64$$

$$\begin{aligned} S_E &= n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2 \\ &= 8 \times 7 + \frac{7(7 + 1)}{2} - 64 \\ &= 56 + \frac{56}{2} - 64 \\ &= 56 + 28 - 64 \\ &= 20 \end{aligned}$$

Výpočet testovací statistiky si provedeme také pomocí softwaru . Vzestupné pořadí 15 naměřených hodnot obou náhodných výběrů stanovíme pomocí funkce `rank()`. Indikační funkci I vytvoříme spojením osmi nul a sedmi jedniček. Opakující se posloupnost nul a jedniček vytvoříme pomocí funkce `rep()`. Vektor měření, indikačních hodnot a pořadí potom spojíme příkazem `rbind()` do matice, kterou příkazem `data.frame()` převedeme na tabulku. Výpočet statistiky T_2 je potom součtem pořadí `S` na těch pozicích, na nichž jsou ve vektoru indikačních hodnot 1 jedničky. Stanovení hodnoty testovací statistiky `SE` je potom pouhým přepisem vzorce 12.1.

- **Kritický obor**

```

12 S <- rank(c(tibia.LLMT, tibia.LLMY))
13 I <- c(rep(0, n1), rep(1, n2))
14 tab <- rbind("Xi, Yj" = c(tibia.LLMT, tibia.LLMY), "I" = I, "Si, Sj" = S)
15 tab <- data.frame(tab)
16 names(tab) <- 1 : 15
17 #
18 # Xi, Yj 340 368 321.5 334 332 340.5 346.5 331 320 356.5 354 350 356 352 324
19 # I      0  0  0.0  0  0  0.0  0.0  0  1  1.0  1  1  1  1  1
20 # Sk     7 15  2.0  6  5  8.0  9.0  4  1 14.0 12 10 13 11  3
21 T2 <- sum(S[I == 1]) # 64
22 SE <- n1 * n2 + n2 * (n2 + 1) / 2 - T2 # 20

```

$$\begin{aligned}
W &= (-\infty; w_{n_1, n_2}(\alpha/2)) \cup (w_{n_1, n_2}(1 - \alpha/2); \infty) \\
&= (-\infty; w_{8,7}(0.05/2)) \cup (w_{8,7}(1 - 0.05/2); \infty) \\
&= (-\infty; w_{8,7}(0.025)) \cup (w_{8,7}(0.975); \infty) \\
&= (-\infty; 11) \cup (45; \infty)
\end{aligned}$$

K výpočtu kvantilů $w_{n_1, n_2}(\alpha/2)$ a $w_{n_1, n_2}(1 - \alpha/2)$ použijeme funkci `qwilcox()`. Vstupními argumenty funkce jsou hodnota kvantilu ($\alpha/2$, resp. $1 - \alpha/2$) a rozsahy náhodných výběrů n_1 a n_2 .

```

23 alpha <- 0.05
24 qwilcox(alpha / 2, n1, n2) # 11
25 qwilcox(1 - alpha / 2, n1, n2) # 45

```

- **Závěr testování**

Protože realizace testovací statistiky $s_E = 20$ nenáleží do kritického oboru, tj. $s_E \notin W$, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

Ke stanovení dolní a horní hranice 95% intervalu spolehlivosti je třeba nejprve vypočítat $n_1 \times n_2 = 8 \times 7 = 56$ rozdílů $X_{1i} - X_{2j}$, $i = 1, \dots, 8$, $j = 1, \dots, 7$, a ty následně vzestupně seřadit. Rozdíly $X_{1i} - X_{2j}$ vypočítáme hromadně pomocí funkce `outer()`. Prvními dvěma vstupními argumenty této funkce budou vektory délek holenních kostí `tibia.LLMT` a `tibia.LLMY`. Nakonec specifikujeme hodnotu argumentu `FUN`, kde defaultní nastavení `FUN = '*'` změním na nastavení `FUN = '-'`. Tím zajistíme výpočet rozdílů $X_{1i} - X_{2j}$ namísto defaultně nastaveného výpočtu součinů $X_{1i} X_{2j}$, $i = 1, \dots, 8$, $j = 1, \dots, 7$. Získané rozdíly $X_{1i} - X_{2j}$, $i = 1, \dots, 8$, $j = 1, \dots, 7$ seřadíme příkazem `sort()` a vložíme do proměnné `U`.

```

26 rozdily <- outer(tibia.LLMT, tibia.LLMY, FUN = '-')
27 U <- sort(rozdily)
28 # [1] -35.0 -34.5 -32.5 -30.5 -28.5 -25.5 -25.0 -24.5 -24.0 -23.0 -22.5 -22.0 -22.0
29 # [14] -21.0 -20.0 -20.0 -19.0 -18.0 -18.0 -16.5 -16.0 -16.0 -16.0 -15.5 -14.0 -13.5
30 # [27] -12.0 -11.5 -10.0 -10.0 -9.5 -9.5 -7.5 -5.5 -3.5 -2.5 1.5 7.0 8.0
31 # [40] 10.0 11.0 11.5 12.0 12.0 14.0 14.0 16.0 16.0 16.5 18.0 20.0 20.5
32 # [53] 22.5 26.5 44.0 48.0

```

Hranice intervalu spolehlivosti potom tvoří ty hodnoty, které se v seřazeném vektoru rozdílů `U` nachází na $(w_{n_1, n_2}(\alpha/2))$ -té pozici a na $(w_{n_1, n_2}(1 - \alpha/2))$ -té pozici. Hodnoty $w_{n_1, n_2}(\alpha/2)$ a $w_{n_1, n_2}(1 - \alpha/2)$ nalezneme opět pomocí funkce `qwilcox()`.

$$\begin{aligned}
(d, h) &= \left(U^{(w_{n_1, n_2}(\alpha/2))}; U^{(w_{n_1, n_2}(1-\alpha/2))} \right) \\
&= \left(U^{(w_{8,7}(0.05/2))}; U^{(w_{8,7}(1-0.05/2))} \right) \\
&= \left(U^{(w_{8,7}(0.025))}; U^{(w_{8,7}(0.975))} \right) \\
&= \left(U^{(11)}; U^{(45)} \right) \\
&= (-22.5; 14)
\end{aligned}$$

```

33 qwilcox(alpha / 2, n1, n2) # 11
34 qwilcox(1 - alpha / 2, n1, n2) # 45
35 U[11] # -22.5
36 U[45] # 14

```

- **Závěr testování**

Protože $\tilde{x}_0 = 0$ náleží do 95% empirického oboustranného intervalu spolehlivosti, tj. $\tilde{x}_0 = 0 \in IS$, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

5. Testování p -hodnotou

Příslušnou p -hodnotu vypočítáme pomocí vzorce $2 \min\{\Pr(S_E \leq s_E), \Pr(S_E \geq s_E)\}$. Zde si uvědomme, že realizace testovací statistiky $s_E = 20$. Zároveň S_E je diskrétní náhodná veličina. Z vlastností pravděpodobnostní funkce diskrétních náhodných veličin víme, že $\Pr(S_E \geq 20) = 1 - \Pr(S_E < 20) = \Pr(S_E \leq 19)$. Viz kapitola ??.

- **p -hodnota**

$$\begin{aligned}
p\text{-hodnota} &= 2 \min\{\Pr(S_E \leq s_E), \Pr(S_E \geq s_E)\} \\
&= 2 \min\{\Pr(S_E \leq 20), 1 - \Pr(S_E \leq 19)\} \\
&= 2 \min\{0.198446, 0.8321678\} \\
&= 2 \times 0.198446 = 0.396892 \doteq 0.3969
\end{aligned}$$

Hodnoty distribuční funkce rozdělení Wilcoxonovy testovací statistiky S_E , tj. $\Pr(S_E \leq 20)$ a $\Pr(S_E \leq 19)$ vypočítáme pomocí funkce `pwilcox()`. Na první pozici zde bude hodnota testovací statistiky S_E , na druhé pozici rozsah prvního náhodného výběru n_1 a na druhé pozici rozsah druhého náhodného výběru n_2 .

```

37 p.hodn <- 2 * min(pwilcox(SE, n1, n2), 1 - pwilcox(SE - 1, n1, n2)) # 0.396892

```

- **Závěr testování**

Protože p -hodnota = 0.3969 je větší než $\alpha = 0.05$, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

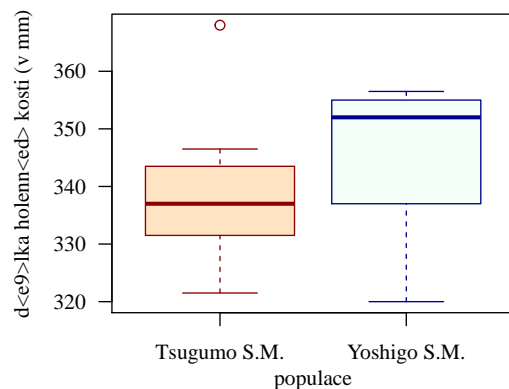
6. Interpretace výsledků

Za základě všech tří typů testování nezamítáme nulovou hypotézu na hladině významnosti $\alpha = 0.05$. Mezi délkou holenní kosti z levé strany u mužů z japonské populace Tsugumo Shell Mound a mužů z japonské populace Yoshigo Shell Mound neexistuje statisticky významný rozdíl.

7. Grafická vizualizace výsledků testování

Vhodným grafem porovnávajícím oba náhodné výběry a jejich mediány je krabicový diagram (viz obrázek 4).

Poznámka: Wilcoxonův dvouvýběrový exaktní test můžeme provést také pomocí funkce `wilcox.test()`. Vstupními parametry budou vektor měření délek holenních kostí z levé strany u mužů z populace Tsugumo Shell Mound (`tibia.LLMT`), vektor měření délek holenních kostí z levé strany u mužů z populace Yoshigo Shell Mound (`tibia.LLMY`), požadavek na výpočet hranic intervalu spolehlivosti zadaný nastavením argumentu `conf.int = T`, hodnota hladiny



Obrázek 4: Krabicový diagram délky holenní kosti z levé strany u mužů z populací Tsugumo Shell Mound a Yoshigo Shell Mound

významnosti α zadaná prostřednictvím koeficientu spolehlivosti $1 - \alpha$ nastavením hodnoty argumentu `conf.level = 0.95`, typ zvolené alternativní hypotézy (oboustranná) zadaný pomocí argumentu `alternative = 'two.sided'` a argument `correct = F`, který zakazuje aplikování spojité korekce na výsledné intervaly spolehlivosti a p -hodnotu.

```
38 wilcox.test(tibia.LLMT, tibia.LLMT, conf.int = T, conf.level = 0.95,
39             alternative = 'two.sided', correct = F)
```

```

      Wilcoxon rank sum exact test
data:  tibia.LLMT and tibia.LLMT
W = 20, p-value = 0.3969
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -22.5  14.0
sample estimates:
difference in location
      -10.75
```

40
41
42
43
44
45
46
47
48
49
50

Součástí výstupu funkce je hodnota rozdílu mediánů obou náhodných výběrů -10.75 , hodnota testovací statistiky S_E značená jako W , tj. $W = 20$, hranice 95% Waldova empirického oboustranného intervalu spolehlivosti -22.5 a 14.0 a p -hodnota $p\text{-value} = 0.3969$. Jediné, co musíme stanovit zvlášť, jsou dolní a horní hranice kritického oboru.

★

Příklad 12.2. Wilcoxonův dvouvýběrový exaktní test (pravostranná alternativa)

Mějme datový soubor 19-more-samples-correlations-skull.txt obsahující údaje o interorbitální šířce (intorb.B) u mužů z bantuské, čínské, malajské, německé a peruánské populace (viz sekce ??). Na hladině významnosti $\alpha = 0.01$ otestujte, zda je interorbitální šířka u mužů bantuské populace menší nebo rovna interorbitální šířce u mužů peruánské populace.

Řešení příkladu 12.2

Z načteného datového souboru 19-more-samples-correlations-skull.txt vybereme naměřené hodnoty interorbitální šířky (intorb.B) u mužů z bantuské populace `pop == 'ban'`, které vložíme je do proměnné `intorb.BB`, a naměřené hodnoty interorbitální šířky (intorb.B) u mužů z peruánské populace `pop == 'per'`, které vložíme je do proměnné `intorb.BP`. Z obou proměnných odstraníme chybějící pozorování (`na.omit()`) a zjistíme rozsahy obou náhodných výběrů (`length()` a `range()`) v obou náhodných výběrech.

```
51 data <- read.delim('00-Data//19-more-samples-correlations-skull.txt')
52 intorb.BB <- data[data$pop == 'ban', 'intorb.B']
53 intorb.BB <- na.omit(intorb.BB)
54 intorb.BP <- data[data$pop == 'per', 'intorb.B']
55 intorb.BP <- na.omit(intorb.BP)
56 n1 <- length(intorb.BB) # 14
57 n2 <- length(intorb.BP) # 46
58 range(intorb.BB) # 21-30
59 range(intorb.BP) # 19-28
```

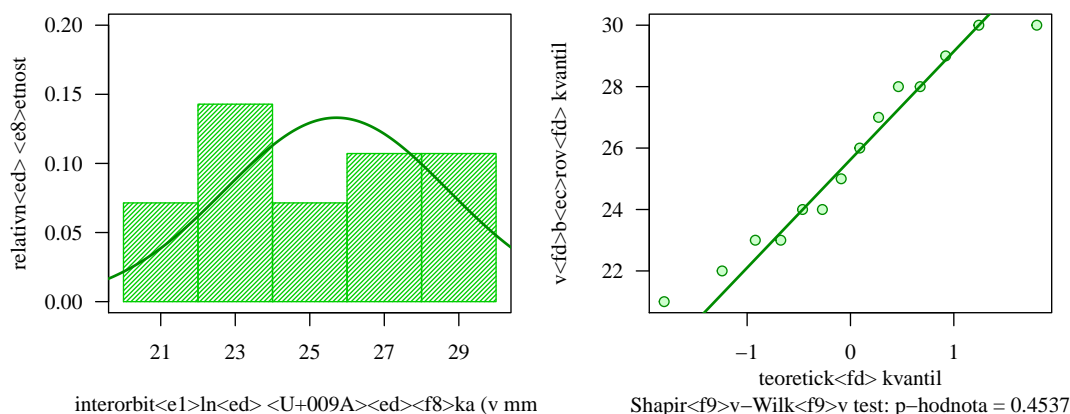
Datový soubor obsahuje naměřené hodnoty interorbitální šířky u 14 mužů bantuské populace a u 46 mužů peruánské populace. Naměřená interorbitální šířka u mužů bantuské populace nabývá hodnot v rozsahu 21–30 mm, naměřená interorbitální šířka u mužů peruánské populace nabývá hodnot v rozsahu 19–28 mm.

V příkladu máme za úkol porovnat střední hodnoty dvou populací, přičemž u obou populací máme k dispozici naměřené hodnoty. Nulovou hypotézu bychom ideálně chtěli otestovat parametrickým testem o rozdílu středních hodnot $\mu_1 - \mu_2$ (viz sekce ?? nebo sekce ??). Nejprve však musíme ověřit, zda oba náhodné výběry pochází z normálních rozdělání. Bez splnění tohoto předpokladu nemůžeme parametrický test použít. Na hladině významnosti $\alpha = 0.05$ testujeme nejprve hypotézu H_{01} : *Náhodný výběr interorbitálních šířek u mužů bantuské populace pochází z normálního rozdělání.* oproti alternativní hypotéze: H_{11} : *Náhodný výběr interorbitálních šířek u mužů bantuské populace nepochází z normálního rozdělání.* K otestování hypotézy použijeme Shapirov-Wilkův test, neboť rozsah náhodného výběru interorbitálních šířek bantuské populace $n_1 = 14 < 30$. Graficky zhodnotíme normalitu náhodného výběru pomocí histogramu superponovaného křivkou normálního rozdělání a QQ-diagramem (viz obrázek 9). Pro účely histogramu rozdělíme náhodný výběr interorbitálních šířek do 5 ekvidistantních třídicích intervalů o šířce 2 mm stanovením hranic 20, 22, ..., 30 mm.

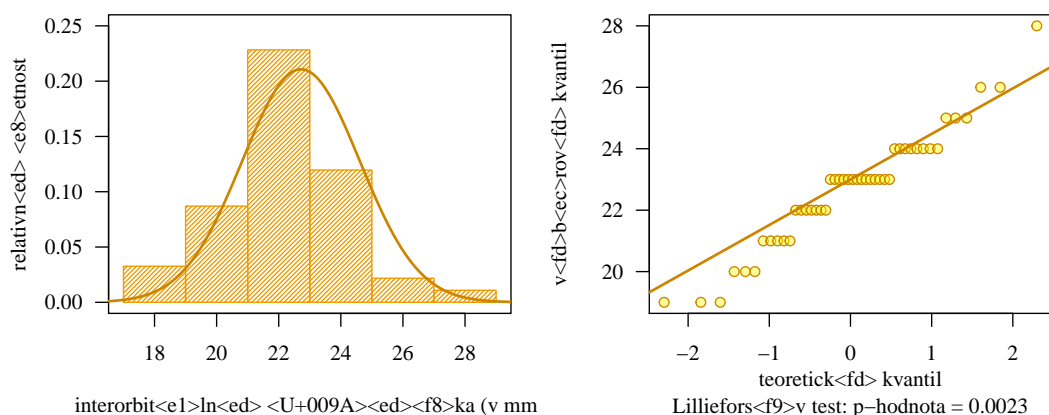
Protože p -hodnota Shapirova-Wilkova testu $p = 0.4537$ je větší než 0.05, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$. Z histogramu ani QQ-diagramu nepozorujeme žádné porušení normálního rozdělání náhodného výběru. Vzhledem k malému počtu pozorování je vzhled histogramu uspokojující. V qq-diagramu zase vidíme, že body se pěkně drží referenční přímky. Náhodný výběr interorbitálních šířek u mužů bantuské populace pochází z normálního rozdělání.

Analogicky testujeme hypotézu H_{02} : *Náhodný výběr interorbitálních šířek u mužů z peruánské populace pochází z normálního rozdělání.* oproti alternativní hypotéze: H_{12} : *Náhodný výběr interorbitálních šířek u mužů z peruánské populace nepochází z normálního rozdělání.* K otestování H_{02} použijeme vzhledem k rozsahu náhodného výběru ($n_2 = 46 > 30$) Lillieforsův test. Pro účely grafické vizualizace prostřednictvím histogramu rozdělíme náhodný výběr interorbitálních šířek do 6 ekvidistantních intervalů o šířce 2 mm stanovením hranic 17, 19, ..., 29 mm. Nakonec vykreslíme QQ-diagram (oba grafy viz obrázek 6).

Protože p -hodnota Lillieforsova testu $p = 0.0023$ je menší než 0.05, hypotézu H_0 zamítáme na hladině významnosti $\alpha = 0.05$. Vykreslení grafu histogramu by nás na první pohled mohlo překvapit, protože histogram vypadá docela rozumně. Nenechme se však zmást. Při bližším pohledu vidíme, že histogram je oproti křivce hustoty posunutý



Obrázek 5: Histogram (vlevo) a kvantilový diagram (vpravo) interorbitální šířky u mužů bantuské populace



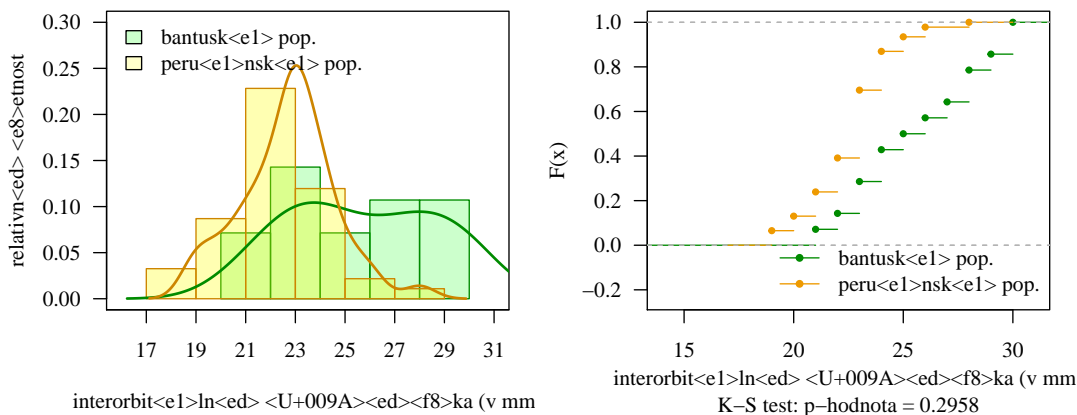
Obrázek 6: Histogram (vlevo) a kvantilový diagram (vpravo) interorbitální šířky u mužů peruánské populace

směrem doleva. Interval se středem 28 potom obsahuje jediné pozorování, které je odlehlé. Porušení normality je potom lépe viditelné v QQ-diagramu, kde pozorujeme, že body umístěné vlevo se výrazněji odchylojí od referenční příčky. Stejně tak se od referenční příčky vzdaluje bod umístěný v pravém horním rohu. Při rozsahu náhodného výběru $n_2 = 46$ máme na normalitu náhodného výběru již větší nároky. Náhodný výběr interorbitálních šířek u mužů z peruánské populace tedy nepochází z normálního rozdělení.

Protože náhodný výběr interorbitálních šířek u mužů peruánské populace nepochází z normálního rozdělení, nemůžeme hypotézu ze zadání testovat pomocí parametrického testu o rozdílu středních hodnot $\mu_1 - \mu_2$. Nmaísto toho tedy použijeme neparametrický test. Vzhledem k nízkému rozsahu náhodného výběru bantuské populace otestujeme hypotézu ze zadání Wilcoxonovým dvouvýběrovým exaktním testem. Jediným předpokladem tohoto testu, je shoda distribučních funkcí $F_1(x)$ a $F_2(x)$ (s výjimkou posunutí o vzdálenost Δ), kde $F_1(x)$ je distribuční funkce náhodného výběru z bantuské populace a $F_2(x)$ je distribuční funkce náhodného výběru z peruánské populace. Shodu distribučních funkcí $F_1(x)$ a $F_2(x)$ ověříme pomocí Kolmogorova-Smirnovova testu (viz sekce ??).

Oba náhodné výběry nejprve metodou centrování posuneme do nuly, čímž odfiltrujeme vliv posunutí o vzdálenost Δ mezi oběma distribučními funkcemi. Na hladině významnosti $\alpha = 0.05$ potom testujeme hypotézu H_{03} : *Distribuční funkce centrované interorbitální šířky u mužů z bantuské a peruánské populace jsou shodné.* oproti alternativní hypotéze: H_{13} : *Distribuční funkce centrované interorbitální šířky u mužů z bantuské a peruánské populace nejsou shodné.* Ze sekce ?? již víme, že tato hypotéza odpovídá hypotéze o shodě dvou distribučních funkcí původních náhodných výběrů s výjimkou posunutí o vzdálenost Δ . Kolmogorův-Smirnovův test provedeme pomocí funkce

ks.test(). Případnou podobnost distribučních funkcí $F_1(x)$ a $F_2(x)$ vizualizujeme dvěma histogramy superponovanými příslušnými jádrovými odhady hustot a grafem zobrazujícím výběrové distribuční funkce $F_1(x)$ a $F_2(x)$ (viz obrázek 7).



Obrázek 7: Porovnání histogramů (vlevo) a graf distribučních funkcí (vpravo) interorbitální šířky u mužů bantuské a peruánské populace

Protože p -hodnota K-S testu $p = 0.2958$ je větší než 0.05 , nulovou hypotézu o shodě dvou distribučních funkcí centrovaných náhodných výběrů nezamítáme na hladině významnosti $\alpha = 0.05$. To tedy znamená, že distribuční funkce náhodného výběru interorbitálních šířek u mužů z bantuské a peruánské populace se liší pouze posunutím.

Předpoklad umožňující testování hypotézy ze zadání pomocí Wilcoxonova dvouvýběrového testu je splněn. Můžeme tedy přistoupit k testování této hypotézy. Naším úkolem je otestovat, zda je interorbitální šířka u mužů bantuské populace menší nebo rovna interorbitální šířce u mužů peruánské populace. Věta ze zadání je zněním nulové hypotézy. Navedou nás na to dvě indicie. Jednak slovo *otestovat* zmíněné v zadání vždy implikuje testování nulové hypotézy a jednak zmíněná rovnost, která je vždy součástí nulové hypotézy. Zbývá dodefinovat alternativní hypotézu tak, aby byla doplňkem k nulové hypotéze. Závěrem připomeňme, že zatímco u parametrických testů (viz kapitola ??) figuruje v hypotézách pojem *střední hodnota*, u neparametrických testů tento pojem nahrazujeme neparametrických ekvivalentem, kterým je *medián*. Testování provedeme opět v posloupnosti sedmi kroků.

1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy
 H_0 : Medián interorbitální šířky u mužů bantuské populace je menší nebo roven mediánu interorbitální šířky u mužů peruánské populace.
 H_1 : Medián interorbitální šířky u mužů bantuské populace je větší než medián interorbitální šířky u mužů peruánské populace.
- **matematická formulace** nulové a alternativní hypotézy
 H_0 : $\tilde{x}_1 \leq \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 \leq \tilde{x}_0$, kde $\tilde{x}_0 = 0$
 H_0 : $\tilde{x}_1 > \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 > \tilde{x}_0$, kde $\tilde{x}_0 = 0$ (pravostranná alternativa)

2. Volba hladiny významnosti

- Hladina významnosti $\alpha = 0.01$ (viz zadání příkladu).

3. Testování kritickým oborem

- **Testovací statistika**
 Nejprve vytvoříme společný vektor interorbitálních šířek u mužů z obou populací tak, že za sebe poskládáme hodnoty délky interorbitální šířky naměřené na mužích z bantuské populace ($X_{1i}, i = 1, \dots, 14$) a následně na mužích z peruánské populace ($X_{2j}, j = 1, \dots, 46$). Délka společného vektoru $n = n_1 + n_2 =$

$14 + 46 = 60$. Následně každému pozorování přiřadíme hodnotu indikační funkce I , která nabývá hodnoty 1, pokud měření pochází z druhého náhodného výběru, nebo hodnoty 0, pokud měření pochází z prvního náhodného výběru. Nakonec stanovíme pořadí S_k , $k = 1, \dots, 60$ všech 60 hodnot od nejmenší po největší (viz tabulka 2).

Tabulka 2: Naměřené hodnoty X_{1i} , $i = 1, \dots, 14$, a X_{2j} , $j = 1, \dots, 46$, obou náhodných výběrů, indikační funkce I příslušnosti k druhému náhodnému výběru a pořadí S_k , $k = 1, \dots, 60$, naměřených hodnot seřazených vzestupně

měření	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_{1i}, X_{2j}	23	28	30	24	28	22	21	30	27	29	26	24	25	23	25
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
S_i, S_j	28.5	56	59.5	41.5	56	16.5	9.5	59.5	54	58	52	41.5	48.5	28.5	48.5
měření	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
X_{1i}, X_{2j}	24	24	22	21	23	22	23	19	23	23	23	25	22	24	21
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	41.5	41.5	16.5	9.5	28.5	16.5	28.5	2	28.5	28.5	28.5	48.5	16.5	41.5	9.5
měření	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
X_{1i}, X_{2j}	23	28	24	20	24	23	22	19	25	23	26	24	21	23	22
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	28.5	56	41.5	5	41.5	28.5	16.5	2	48.5	28.5	52	41.5	9.5	28.5	16.5
měření	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
X_{1i}, X_{2j}	26	20	22	21	21	19	20	23	23	24	23	22	23	24	23
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	52	5	16.5	9.5	9.5	2	5	28.5	28.5	41.5	28.5	16.5	28.5	41.5	28.5

Z tabulky 2 vidíme, že 46 hodnotám příslušícím druhému náhodnému výběru patří ve vektoru patnácti seřazených hodnot pořadí 48.5, 41.5, 41.5, 16.5, 9.5, 28.5, 16.5, 28.5, 2, 28.5, 28.5, 28.5, 48.5, 16.5, 41.5, 9.5, 28.5, 56, 41.5, 5, 41.5, 28.5, 16.5, 2, 48.5, 28.5, 52, 41.5, 9.5, 28.5, 16.5, 52, 5, 16.5, 9.5, 9.5, 2, 5, 28.5, 28.5, 41.5, 28.5, 16.5, 28.5, 41.5, 28.5. Statistiku T_2 získáme sečtením těchto pořadí. Hodnotu testovací statistiky S_E potom získáme dosazením hodnot n_1 , n_2 a statistiky T_2 do vzorce 12.1.

$$T_2 = \sum_{j=1}^{n_2} S_j = 48.5 + 41.5 + 41.5 + 16.5 + \dots + 16.5 + 28.5 + 41.5 + 28.5 = 1220.5$$

$$\begin{aligned}
 S_E &= n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2 \\
 &= 14 \times 46 + \frac{46(46 + 1)}{2} - 1220.5 \\
 &= 644 + \frac{2162}{2} - 64 \\
 &= 644 + 1081 - 1220.5 \\
 &= 504.5
 \end{aligned}$$

Výpočet testovací statistiky si provedeme také pomocí softwaru \mathbb{R} . Nejprve stanovíme vzestupné pořadí 60 naměřených hodnot obou náhodných výběrů (`rank()`). Indikační funkci I vytvoříme spojením 14 nul a 46 jedniček (kombinace funkcí `c()` a `rep()`). Vektor měření, indikačních hodnot a pořadí potom spojíme do matice (`rbind()`), kterou převedeme na tabulku (`data.frame()`). Výpočet statistiky T_2 je potom součtem pořadí `S` na těch pozicích, na nichž jsou ve vektoru indikačních hodnot 1 jedničky. Nakonec stanovíme hodnotu testovací statistiky `SE` přepisem vzorce 12.1.

```

60 S <- rank(c(intorb.BB, intorb.BP))
61 I <- c(rep(0, n1), rep(1, n2))
62 tab <- rbind("Xi, Yj" = c(intorb.BB, intorb.BP), "I" = I, "Si, Sj" = S)
63 tab <- data.frame(tab)
64 names(tab) <- 1 : 60
65 #           1  2    3    4  5    6    7 ...    53  54  55  56  57  58  59  60
66 # Xi, Yj 23.0 28 30.0 24.0 28 22.0 21.0 ... 23.0 23.0 24.0 23.0 22.0 23.0 24.0 23.0
67 # I       0.0 0  0.0  0.0 0  0.0  0.0 ...  1.0  1.0  1.0  1.0  1.0  1.0  1.0  1.0
68 # Si, Sj 28.5 56 59.5 41.5 56 16.5  9.5 ... 28.5 28.5 41.5 28.5 16.5 28.5 41.5 28.5
69 T2 <- sum(S[I == 1]) # 1220.5
70 SE <- n1 * n2 + n2 * (n2 + 1) / 2 - T2 # 504.5

```

- **Kritický obor**

$$\begin{aligned}
W &= \langle w_{n_1, n_2}(1 - \alpha); \infty \rangle \\
&= \langle w_{14, 46}(1 - 0.01); \infty \rangle \\
&= \langle w_{14, 46}(0.99); \infty \rangle \\
&= \langle 454; \infty \rangle
\end{aligned}$$

K výpočtu kvantilu $w_{n_1, n_2}(1 - \alpha)$ použijeme funkci `qwilcox()`. Vstupními argumenty funkce jsou hodnota kvantilu $1 - \alpha$ a rozsahy náhodných výběrů n_1 a n_2 .

```

71 alpha <- 0.01
72 qwilcox(1 - alpha, n1, n2) # 454

```

- **Závěr testování**

Protože realizace testovací statistiky $s_E = 504.5$ náleží do kritického oboru, tj. $s_E \in W$, H_0 zamítáme na hladině významnosti $\alpha = 0.01$.

4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

Ke stanovení dolní hranice 95 % levostranného intervalu spolehlivosti je třeba nejprve vypočítat $n_1 \times n_2 = 14 \times 46 = 644$ rozdílů $X_{1i} - X_{2j}$, $i = 1, \dots, 14$, $j = 1, \dots, 46$, a ty následně vzestupně seřadit. Rozdíly $X_{1i} - X_{2j}$ vypočítáme hromadně pomocí funkce `outer()`. Prvními dvěma vstupními argumenty funkce budou vektory interorbitálních šířek `intorb.BB` a `intorb.BP`. Nakonec specifikujeme hodnotu argumentu `FUN = '-'`. Tím zajistíme výpočet rozdílů $X_{1i} - X_{2j}$ namísto defaultně nastaveného výpočtu součinů $X_{1i}X_{2j}$, $i = 1, \dots, 14$, $j = 1, \dots, 46$. Získané rozdíly $X_{1i} - X_{2j}$, $i = 1, \dots, 14$, $j = 1, \dots, 46$ seřadíme vzestupně příkazem `sort()` a vložíme do proměnné `U`.

```

73 rozdily <- outer(intorb.BB, intorb.BP, FUN = '-')
74 U <- sort(rozdily)

```

Hranice intervalu spolehlivosti potom tvoří hodnota, které se v seřazeném vektoru rozdílů `U` nachází na $(w_{n_1, n_2}(\alpha))$ -té pozici a nekonečno. Hodnotu $w_{n_1, n_2}(\alpha)$ nalezneme opět pomocí funkce `qwilcox()`.

$$\begin{aligned}
(d, h) &= \left(U^{(w_{n_1, n_2}(\alpha))}; \infty \right) \\
&= \left(U^{(w_{14, 46}(0.01))}; \infty \right) \\
&= \left(U^{(190)}; \infty \right) \\
&= (1; \infty)
\end{aligned}$$

```

75 qwilcox(alpha, n1, n2) # 190
76 U[190] # 1

```

- **Závěr testování**

Protože $\tilde{x}_0 = 0$ nenáleží do 99% empirického jednostranného intervalu spolehlivosti, tj. $\tilde{x}_0 = 0 \notin IS$, H_0 zamítáme na hladině významnosti $\alpha = 0.01$.

5. Testování p -hodnotou

Příslušnou p -hodnotu vypočítáme pomocí vzorce $\Pr(S_E \geq s_E)$. Zde si uvědomme, že realizace testovací statistiky $s_E = 504.5$. Zároveň S_E je diskrétní náhodná veličina. Z vlastností pravděpodobnostní funkce diskrétních náhodných veličin víme, že $\Pr(S_E \geq 504.5) = \Pr(S_E \geq 505) = 1 - \Pr(S_E < 505) = \Pr(S_E \leq 504)$. Viz kapitola ??.

- **p -hodnota**

$$p\text{-hodnota} = \Pr(S_E \geq s_E) = 1 - \Pr(S_E \leq 504) = 0.0004942813 \doteq 0.0004943$$

Hodnotu distribuční funkce rozdělení Wilcoxonovy testovací statistiky S_E , tj. $\Pr(S_E \leq 504)$, vypočítáme pomocí funkce `pwilcox()`. Na první pozici zde bude hodnota testovací statistiky S_E , na druhé pozici rozsah prvního náhodného výběru n_1 a na druhé pozici rozsah druhého náhodného výběru n_2 .

```

77 p.hodn <- 1 - pwilcox(504, n1, n2) # 0.0004942813

```

- **Závěr testování**

Protože p -hodnota = 0.0004943 je menší než $\alpha = 0.01$, H_0 zamítáme na hladině významnosti $\alpha = 0.01$.

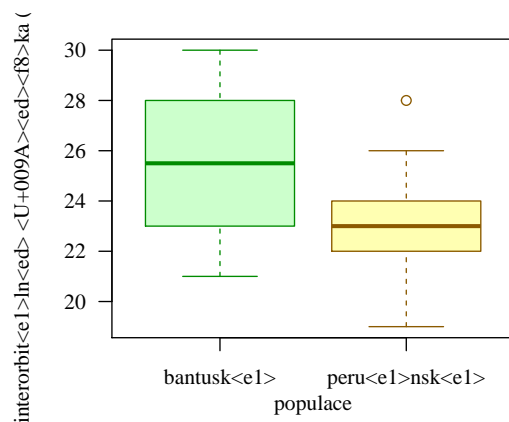
6. Interpretace výsledků

Za základě všech tří typů testování zamítáme na hladině významnosti $\alpha = 0.01$ nulovou hypotézu, že interorbitální šířka u mužů bantuské populace je menší nebo rovna interorbitální šířce u mužů peruánské populace. Interorbitální šířka u mužů bantuské populace je tedy statisticky významně větší než interorbitální šířka u mužů peruánské populace.

7. Grafická vizualizace výsledků testování

Rozdíl ve středních hodnotách (reprezentovaných mediány) interorbitální šířky bantuské a peruánské mužské populace vizualizujeme pomocí krabicových diagramů (viz obrázek 8).

Poznámka: Wilcoxonův dvouvýběrový exaktní test můžeme provést také pomocí funkce `wilcox.test()`. Vstupními parametry funkce budou vektor měření interorbitálních šířek u mužů bantuské populace (`intorb.BB`), vektor měření interorbitálních šířek u mužů peruánské populace (`intorb.BP`), požadavek na výpočet hranic intervalu spolehlivosti (`conf.int = T`, hodnota hladiny významnosti α zadaná prostřednictvím koeficientu spolehlivosti $1 - \alpha$ (`conf.level = 0.99`), typ zvolené alternativní hypotézy (pravostranná; `alternative = 'greater'`) a zákaz aplikování spojité korekce na výsledné intervaly spolehlivosti a p -hodnotu (`correct = F`).



Obrázek 8: Krabicový diagram interorbitální šířky u mužů bantuské a peruánské populace

```
78 wilcox.test(intorb.BB, intorb.BP, conf.int = T, conf.level = 0.99,
79           alternative = 'greater', correct = F)
```

```
Wilcoxon rank sum test
data: intorb.BB and intorb.BP
W = 504.5, p-value = 0.0006093
alternative hypothesis: true location shift is greater than 0
99 percent confidence interval:
 0.9999884      Inf
sample estimates:
difference in location
      2.999984
```

80
81
82
83
84
85
86
87
88
89
90

Součástí výstupu funkce je hodnota rozdílu mediánů obou náhodných výběrů 2.999984, hodnota testovací statistiky S_E značená jako W , tj. $W = 504.5$, interpolované hranice 99% Waldova empirického jednostranného intervalu spolehlivosti 0.9999884 a Inf a interpolovaná p -hodnota $p\text{-value} = 0.0006093$, adaptovaná na situaci, kdy jsou některé hodnoty X_{1i} , $i = 1, \dots, 14$ mezi sebou shodné, X_{2j} , $j = 1, \dots, 46$ mezi sebou shodné nebo X_{1i} , X_{2j} , $i = 1, \dots, 14$, $j = 1, \dots, 46$ navzájem mezi sebou shodné. Jediné, co musíme stanovit zvlášť, jsou dolní a horní hranice kritického oboru.

Poznámka: Výstup funkce `wilcox.test()` provádí tentokrát dvě varovné hlášky. První hláška ve znění: `In wilcox.test.default(intorb.BB, intorb.BP, conf.int = T, conf.level = 0.95,): cannot compute exact p-value with ties.` nás upozorňuje na výskyt duplicitních hodnot v kombinaci s nízkým počtem pozorování obou náhodných výběrů $X_{11} = X_{14} = X_{26} = X_{28} = X_{210} = X_{211} = X_{212} = X_{217} = X_{222} = X_{226} = X_{230} = X_{239} = X_{240} = X_{242} = X_{244} = X_{246} = 23$, $X_{12} = X_{15} = X_{218} = 28$, $X_{13} = X_{18} = 30$, $X_{14} = X_{112} = X_{22} = X_{23} = X_{215} = X_{219} = X_{221} = X_{228} = X_{241} = X_{245} = 24$, $X_{16} = X_{24} = X_{27} = X_{214} = X_{223} = X_{231} = X_{234} = X_{243} = 22$, $X_{17} = X_{25} = X_{216} = X_{229} = X_{235} = X_{236} = 21$, $X_{111} = X_{227} = X_{232} = 26$, $X_{113} = X_{21} = X_{213} = X_{225} = 25$, $X_{29} = X_{224} = X_{237} = 19$, $X_{220} = X_{233} = X_{238} = 20$, a na použití modifikovaného postupu při výpočtu p -hodnoty. Druhá hláška ve znění: `In wilcox.test.default(intorb.BB, intorb.BP, conf.int = T, conf.level = 0.95,): cannot compute exact confidence intervals with ties` nás opět upozorňuje na výskyt duplicitních rozdílů v kombinaci s nízkým počtem pozorování obou náhodných výběrů a na použití modifikovaného postupu při výpočtu hranic intervalu spolehlivosti.

★

Příklad 12.3. Wilcoxonův dvouvýběrový exaktní test (levostranná alternativa)

Mějme datový soubor 16-anova-head.txt obsahující údaje o šířce tváře (bizyg.W) mladých dospělých jedinců, převážně studentů z Brna a Ostravy. Dále máme údaje o sexuální orientaci (sexor; výlučně heretosexuální nebo jiná než výlučně heterosexuální) a o počtu starších sourozenců těchto jedinců (obra; jeden nebo žádný). Více informací o datovém souboru viz sekce ???. Na hladině významnosti $\alpha = 0.05$ zjistíte, zda je šířka tváře žen s výlučně heterosexuální orientací menší než šířka tváře žen s jinou než výlučně heterosexuální orientací.

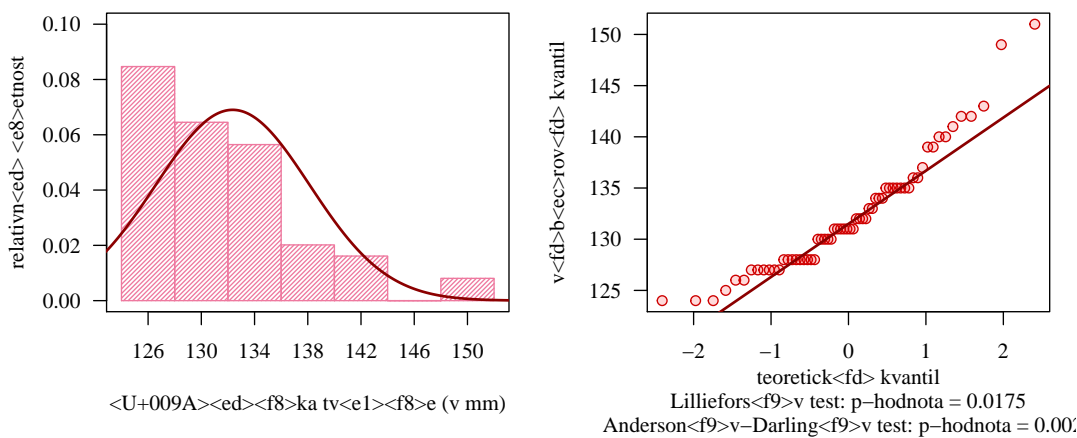
Řešení příkladu 12.3

Z načteného datového souboru 16-anova-head.txt vybereme naměřené hodnoty šířky tváře (bizyg.W) žen sex == 'f' s výlučně heterosexuální orientací (sexor == 'op') a vložíme je do proměnné bizyg.WFO. Dále vybereme naměřené hodnoty šířky tváře (bizyg.W) žen sex == 'f' s jinou než heterosexuální orientací (sexor == 'se') a vložíme je do proměnné bizyg.WFS. Z obou proměnných odstraníme chybějící pozorování a zjistíme počet pozorování a rozpětí naměřených hodnot v obou náhodných výběrech.

```
91 data <- read.delim('00-Data//16-anova-head.txt')
92 bizyg.WFO <- data[data$sex == 'f' & data$sexor == 'op', 'bizyg.W']
93 bizyg.WFS <- data[data$sex == 'f' & data$sexor == 'sa', 'bizyg.W']
94 bizyg.WFO <- na.omit(bizyg.WFO)
95 bizyg.WFS <- na.omit(bizyg.WFS)
96 n1 <- length(bizyg.WFO) # 62
97 n2 <- length(bizyg.WFS) # 23
98 range(bizyg.WFO) # 124-151
99 range(bizyg.WFS) # 120-148
```

Datový soubor obsahuje naměřené hodnoty šířky tváře 62 žen s heterosexuální orientací a 23 žen s jinou sexuální orientací. Naměřená šířka tváře žen s heterosexuální orientací nabývá hodnot v rozsahu 124–151 mm, naměřená šířka tváře žen s jinou sexuální orientací nabývá hodnot v rozsahu 120–148 mm.

V příkladu máme za úkol porovnat střední hodnoty dvou populací, přičemž u obou populací máme k dispozici naměřené hodnoty. Nulovou hypotézu bychom chtěli primárně testovat parametrickým testem o rozdílu středních hodnot $\mu_1 - \mu_2$. Opět tedy nejprve ověříme předpoklad normálního rozdělení obou náhodných výběrů. Na hladině významnosti $\alpha = 0.05$ testujeme nejprve hypotézu H_{01} : *Náhodný výběr šířek tváře žen s heterosexuální orientací pochází z normálního rozdělení*, oproti alternativní hypotéze: H_{11} : *Náhodný výběr šířek tváře žen s heterosexuální orientací nepochází z normálního rozdělení*. K otestování hypotézy použijeme vzhledem k rozsahu náhodného výběru $n_1 = 62 > 30$ Lillieforsův test. Pro účely vizualizace náhodného výběru pomocí histogramu rozdělíme naměřené hodnoty do 7 ekvidistantních třídicích intervalů o šířce 4 mm stanovením hranic 124, 131, ..., 152 mm (viz obrázek 9).

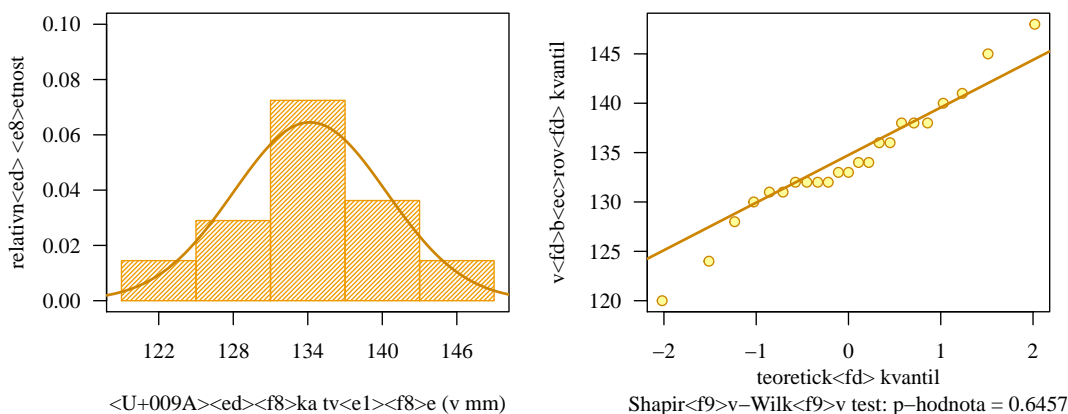


Obrázek 9: Histogram (vlevo) a kvantilový diagram (vpravo) interorbitální šířky u mužů bantuské populace

Protože p -hodnota Lillieforsova testu $p = 0.0175$ je menší než 0.05 , H_0 zamítáme na hladině významnosti $\alpha = 0.05$. Zamítnutí bychom nicméně mohli považovat za hraniční, protože p -hodnota je větší než hladina významnosti $\alpha = 0.01$. Kontrolně tedy otestujeme normalitu náhodného výběru také pomocí Andersonova-Darlingova testu příkazem `ad.test()` z knihovny `nortest` (viz sekce ??). Protože p -hodnota Andersonova-Darlingova testu $p = 0.0175$ je menší než 0.05 , H_0 zamítáme na hladině významnosti $\alpha = 0.05$. Andersonův-Darlingův test tedy možnost hraničně normálního rozdělení náhodného výběru nepotvrdil. Grafická vizualizace taktéž ukazuje na porušení normality. Z histogramu na obrázku 9 pozorujeme nesymetrické rozdělení naměřených hodnot, které jsou kladně vyšikmené s prodlouženým pravým koncem. Taktéž QQ-diagram ukazuje odlehle hodnoty na pravém chvostu. Náhodný výběr šířek tváře u žen s heterosexuální orientací nepochází z normálního rozdělení.

Již v tomto okamžiku víme, že k ověření otázky ze zadání budeme muset použít neparametrický test o rozdílu mediánů. Pro pořádek (a také proto, že před samotným testováním bychom vždy měli mít alespoň základní přehled o rozdělení naměřených hodnot) si ověříme také normalitu druhého náhodného výběru.

Otestujeme tedy hypotézu H_{02} : *Náhodný výběr šířek tváře u žen jiné než heterosexuální orientace pochází z normálního rozdělení.* oproti alternativní hypotéze: H_{12} : *Náhodný výběr šířek tváře u žen jiné než heterosexuální orientace nepochází z normálního rozdělení.* K otestování H_{02} použijeme vzhledem k rozsahu náhodného výběru ($n_2 = 23 < 30$) Shapirov-Wilkův test. Pro účely grafické vizualizace prostřednictvím histogramu rozdělíme náhodný výběr šířek tváře do 5 ekvidistantních intervalů o šířce 6 mm stanovením hranic 119, 125, ..., 149 mm. Nakonec vykreslíme QQ-diagram (oba grafy viz obrázek 10).



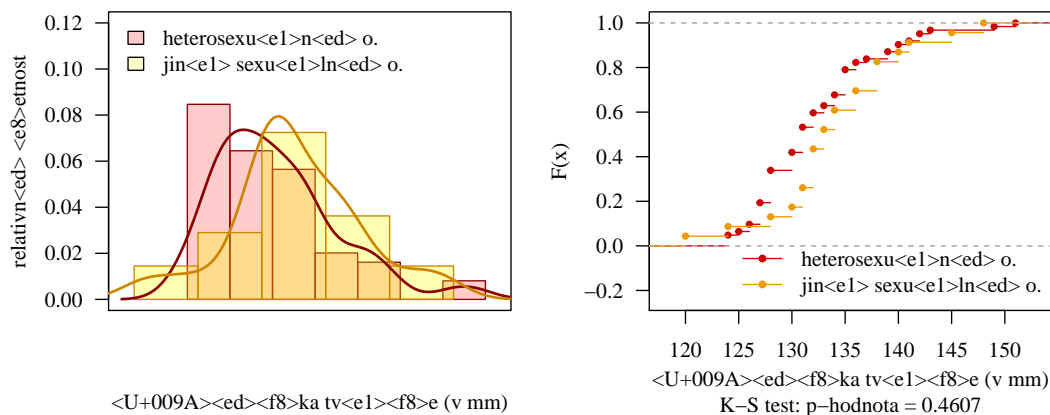
Obrázek 10: Histogram (vlevo) a kvantilový diagram (vpravo) šířky tváře žen s heterosexuální orientací a žen s jinou sexuální orientací

Protože p -hodnota Shapirova-Wilkova testu $p = 0.6457$ je menší než 0.05 , hypotézu H_0 nezamítáme na hladině významnosti $\alpha = 0.05$. Pohledem na histogram vidíme symetrické rozdělení naměřených hodnot, které vykazuje jasné prvky normality. V QQ-diagramu vidíme mírné odchýlení krajních bodů od referenční křivky, které však nemá vzhledem k nižšímu rozsahu náhodného výběru fatální vliv na normální rozdělení naměřených hodnot. Náhodný výběr šířek tváře žen s jinou než heterosexuální orientací pochází z normálního rozdělení.

Jak jsme zmínili výše, kvůli porušení normality prvního náhodného výběru použijeme na ověření otázky ze zadání neparametrický test. Vzhledem k nízkému počtu pozorování v druhém náhodném výběru použijeme Wilcoxonův dvouvýběrový exaktní test. Jediným předpokladem tohoto testu, je shoda distribučních funkcí $F_1(x)$ a $F_2(x)$ (s výjimkou posunutí o vzdálenost Δ), kde $F_1(x)$ je distribuční funkce náhodného výběru žen s heterosexuální orientací a $F_2(x)$ je distribuční funkce náhodného výběru žen s jinou sexuální orientací. Shodu distribučních funkcí $F_1(x)$ a $F_2(x)$ ověříme pomocí Kolmogorova-Smirnovova testu.

Oba náhodné výběry nejprve metodou centrování posuneme do nuly, čímž odfiltrujeme vliv posunutí o vzdálenost Δ mezi oběma distribučními funkcemi. Na hladině významnosti $\alpha = 0.05$ testujeme hypotézu H_{03} : *Distribuční funkce*

centrované šířky tváře žen s heterosexuální orientací a žen s jinou sexuální orientací jsou shodné. oproti alternativní hypotéze: H_{13} : Distribuční funkce centrované šířky tváře žen s heterosexuální orientací a žen s jinou sexuální orientací nejsou shodné. Tato hypotéza odpovídá hypotéze o shodě dvou distribučních funkcí původních náhodných výběrů s výjimkou posunutí o vzdálenost Δ . Kolmogorův-Smirnovův test provedeme pomocí funkce `ks.test()`. Náhled na distribuční funkce $F_1(x)$ a $F_2(x)$ vizualizujeme dvěma histogramy superponovanými příslušnými jádrovými odhady hustot a grafem zobrazujícím výběrové distribuční funkce $F_1(x)$ a $F_2(x)$ (viz obrázek 11).



Obrázek 11: Porovnání histogramů (vlevo) a graf distribučních funkcí (vpravo) šířky tváře žen s heterosexuální orientací a žen s jinou sexuální orientací

Protože p -hodnota K-S testu $p = 0.4607$ je větší než 0.05 , nulovou hypotézu o shodě dvou distribučních funkcí centrovaných náhodných výběrů nezamítáme na hladině významnosti $\alpha = 0.05$. To tedy znamená, že distribuční funkce náhodného výběru šířek tváře žen s heterosexuální orientací a žen s jinou sexuální orientací se liší pouze posunutím.

Předpoklad umožňující ověření otázky ze zadání pomocí Wilcoxonova dvouvýběrového testu je splněn. Naším úkolem ze zadání je zjistit, zda je šířka tváře žen s výlučně heterosexuální orientací menší než šířka tváře žen s jinou než výlučně heterosexuální orientací. Věta ze zadání je zněním alternativní hypotézy, a to proto, že zde není zmínka ani o nulové hypotéze, ani o testování, ani se ve větě nevyskytuje zmínka o rovnosti, která je vždy součástí nulové hypotézy. Nulovou hypotézu tedy dodefinujeme tak, aby byla doplňkem k alternativní hypotéze. Ve finálním znění hypotéz bude opět figurovat pojem *medián* namísto pojmu *střední hodnota*.

1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

H_0 : Medián šířky tváře žen s výlučně heterosexuální orientací je větší nebo roven mediánu šířky tváře žen s jinou než výlučně heterosexuální orientací.

H_1 : Medián šířky tváře žen s výlučně heterosexuální orientací je menší než medián šířky tváře žen s jinou než výlučně heterosexuální orientací.

- **matematická formulace** nulové a alternativní hypotézy

H_0 : $\tilde{x}_1 \geq \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 \geq \tilde{x}_0$, kde $\tilde{x}_0 = 0$

H_0 : $\tilde{x}_1 < \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 < \tilde{x}_0$, kde $\tilde{x}_0 = 0$ (pravostranná alternativa)

2. Volba hladiny významnosti

- Hladina významnosti $\alpha = 0.05$.

3. Testování kritickým oborem

- **Testovací statistika**

V prvním kroku vytvoříme společný vektor šířek tváře žen tak, že za sebe poskládáme hodnoty šířek

tváře žen s heterosexuální orientací ($X_{1i}, i = 1, \dots, 62$) a následně žen s jinou sexuální orientací ($X_{2j}, j = 1, \dots, 23$). Délka společného vektoru $n = n_1 + n_2 = 62 + 23 = 85$. Následně každému pozorování přiřadíme hodnotu indikační funkce I , která nabývá hodnoty 1, pokud měření pochází z druhého náhodného výběru, nebo hodnoty 0, pokud měření pochází z prvního náhodného výběru. Nakonec stanovíme pořadí $S_k, k = 1, \dots, 85$ všech 85 hodnot od nejmenší po největší (viz tabulka 3).


Tabulka 3: Naměřené hodnoty $X_{1i}, i = 1, \dots, 62$, a $X_{2j}, j = 1, \dots, 23$, obou náhodných výběrů, indikační funkce I příslušnosti k druhému náhodnému výběru a pořadí $S_k, k = 1, \dots, 85$, naměřených hodnot seřazených vzestupně

měření	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
X_{1i}, X_{2j}	127	128	127	137	131	130	127	140	133	135	128	135	128	130	128	135	136
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S_i, S_j	11.5	19.5	11.5	68	35	27.5	11.5	75	49.5	60	19.5	60	19.5	27.5	19.5	60	65.5
měření	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
X_{1i}, X_{2j}	132	142	127	143	132	136	140	130	131	131	126	127	135	127	135	135	134
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S_i, S_j	43.5	79.5	11.5	81	43.5	65.5	75	27.5	35	35	35	7.5	11.5	60	11.5	60	54
měření	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
X_{1i}, X_{2j}	134	134	124	124	132	131	128	141	130	132	128	133	151	142	125	139	131
I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S_i, S_j	54	54	3.5	3.5	43.5	35	19.5	77.5	27.5	43.5	19.5	49.5	85	79.5	6	72.5	35
měření	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
X_{1i}, X_{2j}	135	128	131	135	128	128	149	130	139	126	124	134	138	141	138	120	145
I	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
S_i, S_j	60	19.5	35	60	19.5	19.5	84	27.5	72.5	7.5	3.5	54	70	77.5	70	1	82
měření	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85
X_{1i}, X_{2j}	128	140	124	131	130	132	136	136	138	132	131	132	134	133	148	132	133
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	19.5	75	3.5	35	27.5	43.5	65.5	65.5	70	43.5	35	43.5	54	49.5	83	43.5	49.5

Z tabulky 3 vidíme, že 23 hodnotám příslušícím druhému náhodnému výběru patří ve vektoru 85 seřazených hodnot pořadí 54, 70, 77.5, 70, 1, 82, 19.5, 75, 3.5, 35, 27.5, 43.5, 65.5, 65.5, 70, 43.5, 35, 43.5, 54, 49.5, 83, 43.5 a 49.5. Statistiku T_2 získáme sečtením těchto pořadí. Hodnotu testovací statistiky S_E získáme dosazením hodnot n_1, n_2 a statistiky T_2 do vzorce 12.1.

$$T_2 = \sum_{j=1}^{n_2} S_j = 54 + 72 + 77.5 + 70 + \dots + 49.5 + 83 + 43.5 + 49.5 = 1161$$

$$\begin{aligned} S_E &= n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2 \\ &= 62 \times 23 + \frac{23(23 + 1)}{2} - 1161 \\ &= 1426 + \frac{552}{2} - 1161 \\ &= 1426 + 276 - 1161 \\ &= 541 \end{aligned}$$

Výpočet testovací statistiky si provedeme také pomocí softwaru . Nejprve stanovíme vzestupné pořadí 60 naměřených hodnot obou náhodných výběrů (`rank()`). Indikační funkci I vytvoříme spojením 62 nul a 23 jedniček (kombinace funkcí `c()` a `rep()`). Vektor měření, indikačních hodnot a pořadí potom spojíme do

matice (`rbind()`), kterou převedeme na tabulku (`data.frame()`). Výpočet statistiky T_2 je potom součtem pořadí S na těch pozicích, na nichž jsou ve vektoru indikačních hodnot I jedničky. Nakonec stanovíme hodnotu testovací statistiky SE .

```

100 S <- rank(c(bizyg.WFO, bizyg.WFS))
101 I <- c(rep(0, n1), rep(1, n2))
102 tab <- rbind("Xi, Yj" = c(bizyg.WFO, bizyg.WFS), "I" = I, "Si, Sj" = S)
103 tab <- data.frame(tab)
104 names(tab) <- 1 : 85
105 #           1      2      3      4      5      6      7 ... 79      80      81      82      83      84      85
106 # Xi, Yj 127.0 128.0 127.0 137 131 130.0 127.0 ... 131 132.0 134 133.0 148 132.0 133.0
107 # I       0.0  0.0  0.0  0  0  0.0  0.0 ...  1  1.0  1  1.0  1  1.0  1.0
108 # Si, Sj 11.5 19.5 11.5 68 35 27.5 11.5 ... 35 43.5 54 49.5 83 43.5 49.5
109 T2 <- sum(S[I == 1]) # 1161
110 SE <- n1 * n2 + n2 * (n2 + 1) / 2 - T2 # 541

```

• Kritický obor

$$\begin{aligned}
 W &= (-\infty; w_{n_1, n_2}(\alpha)) \\
 &= (-\infty; w_{62, 23}(0.05)) \\
 &= (-\infty; 547)
 \end{aligned}$$

K výpočtu kvantilu $w_{n_1, n_2}(\alpha)$ použijeme funkci `qwilcox()`. Vstupními argumenty funkce jsou hodnota kvantilu α a rozsahy náhodných výběrů n_1 a n_2 .

```

111 alpha <- 0.05
112 qwilcox(alpha, n1, n2) # 547

```

• Závěr testování

Protože realizace testovací statistiky $s_E = 541$ náleží do kritického oboru, tj. $s_E \in W$, H_0 zamítáme na hladině významnosti $\alpha = 0.05$.

4. Testování intervalem spolehlivosti

• Interval spolehlivosti

Ke stanovení dolní hranice 95% pravostranného intervalu spolehlivosti je třeba nejprve vypočítat $n_1 \times n_2 = 62 \times 23 = 1426$ rozdílů $X_{1i} - X_{2j}$, $i = 1, \dots, 62$, $j = 1, \dots, 23$, a ty následně vzestupně seřadit. Rozdíly $X_{1i} - X_{2j}$ vypočítáme hromadně pomocí funkce `outer()`. Prvními dvěma vstupními argumenty funkce budou vektory šířek tváře `bizyg.WFO` a `bizyg.WFS`, třetí argumentem bude argument `FUN = '-'`, kterým zajistíme výpočet rozdílů $X_{1i} - X_{2j}$ $i = 1, \dots, 62$, $j = 1, \dots, 23$. Získané rozdíly seřadíme vzestupně příkazem `sort()` a vložíme do proměnné `U`.

```

113 rozdily <- outer(bizyg.WFO, bizyg.WFS, FUN = '-')
114 U <- sort(rozdily)

```

Hranice intervalu spolehlivosti potom tvoří mínus nekonečno a hodnota, která se v seřazeném vektoru rozdílů `U` nachází na $(w_{n_1, n_2}(1 - \alpha))$ -té pozici. Hodnotu kvantilu $w_{n_1, n_2}(1 - \alpha)$ nalezneme opět pomocí funkce `qwilcox()`.

$$\begin{aligned}
(d, h) &= \left(-\infty; U^{(w_{n_1, n_2}(1-\alpha))}\right) \\
&= \left(-\infty; U^{(w_{62, 23}(1-0.05))}\right) \\
&= \left(-\infty; U^{(w_{62, 23}(0.95))}\right) \\
&= \left(-\infty; U^{(45)}\right) \\
&= (-\infty; 0)
\end{aligned}$$

```

115 qwilcox(1 - alpha, n1, n2) # 879
116 U[879] # 0

```

- **Závěr testování**

Protože $\tilde{x}_0 = 0$ nenáleží do 95% empirického jednostranného intervalu spolehlivosti, tj. $\tilde{x}_0 = 0 \notin IS$, H_0 zamítáme na hladině významnosti $\alpha = 0.05$.


Poznámka: Připomeňme, že kulaté závorky okolo intervalu spolehlivosti $(d; h)$ značí, že do intervalu spolehlivosti spadají veškeré hodnoty v rozmezí hranic $d-h$ s výjimkou samotných hranic d a h . Z toho důvodu hodnota $\tilde{x}_0 = 0 \notin (-\infty; 0)$.

5. Testování p -hodnotou

Příslušnou p -hodnotu vypočítáme pomocí vzorce $\Pr(S_E \leq s_E)$, kde s_E je realizace testovací statistiky $s_E = 541$ diskrétní náhodné veličiny S_E .

- **p -hodnota**

$$p\text{-hodnota} = \Pr(S_E \leq s_E) = \Pr(S_E \leq 541) = 0.04494765 \doteq 0.04495$$

Hodnotu distribuční funkce rozdělení Wilcoxonovy testovací statistiky S_E , tj. $\Pr(S_E \leq 541)$, vypočítáme v softwaru  pomocí funkce `pwilcox()`.

```

117 p.hodn <- pwilcox(SE, n1, n2) # 0.04494765

```

- **Závěr testování**

Protože p -hodnota = 0.04495 je menší než $\alpha = 0.05$, H_0 zamítáme na hladině významnosti $\alpha = 0.05$.

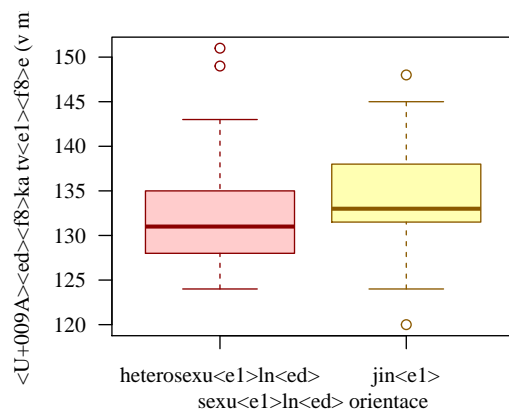
6. Interpretace výsledků

Za základě všech tří typů testování zamítáme nulovou hypotézu na hladině významnosti $\alpha = 0.05$. Šířka tváře žen s heterosexuální orientací je statisticky významně menší než šířka tváře žen s jinou sexuální orientací.

7. Grafická vizualizace výsledků testování

Rozdíl ve středních hodnotách (reprezentovaných mediány) šířky tváře žen s heterosexuální orientací a žen s jinou sexuální orientací vizualizujeme pomocí krabicových diagramů (viz obrázek 12).

Poznámka: Wilcoxonův dvouvýběrový exaktní test provedeme také pomocí funkce `wilcox.test()`. Vstupními parametry funkce budou vektor měření šířek tváře žen s heterosexuální orientací (`bizyg.WFO`) a s jinou sexuální orientací (`bizyg.WFS`), požadavek na výpočet hranic intervalu spolehlivosti (`conf.int = T`, hodnota hladiny významnosti α zadaná prostřednictvím koeficientu spolehlivosti $1 - \alpha$ (`conf.level = 0.95`), typ zvolené alternativní hypotézy (levostranná; `alternative = 'less'`) a zákaz aplikování spojitě korekce na výsledné intervaly spolehlivosti a p -hodnotu (`correct = F`).



Obrázek 12: Krabicový diagram šířky tváře žen s heterosexuální orientací a žen s jinou sexuální orientací

```
118 wilcox.test(bizyg.WF0, bizyg.WFS, conf.int = T, conf.level = 0.95,
119             alternative = 'less', correct = F)
```

```

      Wilcoxon rank sum test
data:  bizyg.WF0 and bizyg.WFS
W = 541, p-value = 0.044
alternative hypothesis: true location shift is less than 0
95 percent confidence interval:
      -Inf -6.897118e-05
sample estimates:
difference in location
      -2.000094
```

120
121
122
123
124
125
126
127
128
129
130

Mezi výstupy funkce můžeme nalézt hodnotu rozdílu mediánů obou náhodných výběrů -0.200094, hodnota testovací statistiky S_E značená jako W , tj. $W = 541$, interpolované hranice 95% Waldova empirického pravostranného intervalu spolehlivosti $-\text{Inf}$ a $-6.897118e-05$ a interpolovanou p -hodnotu $p\text{-value} = -6.897118e-05$. Jediné, co musíme stanovit zvlášť, jsou dolní a horní hranice kritického oboru.

Poznámka: Ačkoli oba náhodné výběry X_{11}, \dots, X_{1n_1} a X_{21}, \dots, X_{2n_2} obsahují shodná měření (ties) uvnitř sebe i navzájem mezi sebou, výstup funkce `wilcox.test()` neprovází tentokrát žádná varovná hláška. Je to proto, že náhodný výběr šířek tváře žen s jinou sexuální orientací obsahuje více než 50 pozorování. V takovém případě funkce `wilcox.test()` nebere shody jako potenciální riziko při výpočtu exaktní p -hodnoty či exaktních hranic intervalu spolehlivosti a chybové hlášky tedy nevypisuje. Aby se chybové hlášky analogické hláškám v příkladu 12.2 nevypsaly, stačí, aby byl počet pozorování v alespoň jednom náhodném výběru větší nebo roven 50.



12.2 Wilcoxonův dvouvýběrový test (Mannův-Whitneyův U test) – asymptotická varianta

Pro náhodné výběry o rozsazích $n_1 > 30$ a $n_2 > 30$ máme možnost použít k otestování nulové hypotézy asymptotickou variantu Wilcoxonova dvouvýběrového testu (resp. Mannova-Whitneyova U testu). Testovací statistika

$$S_A = \frac{S_E - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \quad (12.2)$$

kde S_E je testovací statistika definovaná vztahem 12.1, n_1 je rozsah prvního náhodného výběru, n_2 je rozsah druhého náhodného výběru. Za platnosti nulové hypotézy pochází statistika U_A ze standardizovaného normálního rozdělení, tj.

$$S_A = \frac{S_E - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \stackrel{H_0}{\sim} N(0, 1).$$

Kritický obor podle zvolené alternativní hypotézy má tvar

$$\begin{aligned} H_{11} : \tilde{x}_1 - \tilde{x}_2 &\neq \tilde{x}_0 & W &= (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty) \\ H_{12} : \tilde{x}_1 - \tilde{x}_2 &> \tilde{x}_0 & W &= (u_{1-\alpha}; \infty) \\ H_{13} : \tilde{x}_1 - \tilde{x}_2 &< \tilde{x}_0 & W &= (-\infty; u_{\alpha}) \end{aligned}$$

kde $u_{\alpha/2}$, $u_{1-\alpha/2}$, u_{α} , $u_{1-\alpha}$ jsou kvantily standardizovaného normálního rozdělení, jejichž hodnoty získáme pomocí \mathcal{R} a implementované funkce `qnorm()`.

Interval spolehlivosti má podle zvolené alternativní hypotézy jeden z následujících tvarů

$$\begin{aligned} H_{11} : \tilde{x}_1 - \tilde{x}_2 &\neq \tilde{x}_0 & (d, h) &= (U^{(C_{1-\alpha/2})}; U^{(n_1 n_2 + 1 - C_{1-\alpha/2})}) \\ H_{12} : \tilde{x}_1 - \tilde{x}_2 &> \tilde{x}_0 & (d, \infty) &= (U^{(C_{1-\alpha})}; \infty) \\ H_{13} : \tilde{x}_1 - \tilde{x}_2 &< \tilde{x}_0 & (-\infty, h) &= (-\infty; U^{(n_1 n_2 + 1 - C_{1-\alpha})}) \end{aligned}$$

kde $C_{1-\alpha/2} = \frac{n_1 n_2}{2} - u_{1-\alpha/2} \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$, $C_{1-\alpha} = \frac{n_1 n_2}{2} - u_{1-\alpha} \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$ a $U^{(1)} \leq \dots \leq U^{(n_1 n_2)}$ značí vzestupně seřazené rozdíly $X_{2j} - X_{1i}$, $i = 1, \dots, n_1$, $j = 1, \dots, n_2$, a $U^{(x)}$ značí x -tý seřazený rozdíl. Není-li $C_{1-\alpha/2}$, resp. $C_{1-\alpha}$ celé číslo, zaokrouhujeme jej vždy dolů na nejbližší nižší celé číslo.

p -hodnota má v závislosti na zvolené alternativní hypotéze jeden z následujících tvarů

$$\begin{aligned} H_{11} : \tilde{x}_1 - \tilde{x}_2 &\neq \tilde{x}_0 & p\text{-hodnota} &= 2 \min\{\Pr(S_A \leq s_A), \Pr(S_A > s_A)\} \\ H_{12} : \tilde{x}_1 - \tilde{x}_2 &> \tilde{x}_0 & p\text{-hodnota} &= \Pr(S_A > s_A) = 1 - \Pr(S_A \leq s_A) \\ H_{13} : \tilde{x}_1 - \tilde{x}_2 &< \tilde{x}_0 & p\text{-hodnota} &= \Pr(S_A \leq s_A) \end{aligned}$$

kde S_A je náhodná veličina, s_A je realizace testovací statistiky S_A (viz vzorec 12.2), tedy konkrétní číslo, a $\Pr(S_A \leq s_A)$ je distribuční funkce standardizovaného normálního rozdělení, jejíž hodnotu získáme pomocí \mathcal{R} a implementované funkce `pnorm()`.

Příklad 12.4. Wilcoxonův dvouvýběrový asymptotický test (oboustranná alternativa)

Mějme datový soubor 18-more-samples-variances-clavicle.txt obsahující údaje o délce klíční kosti z pravé strany (cla.L) u mužů z indické populace z Amritsaru, z indické populace z Varanasi a z řecké populace z Atén (viz sekce ??). Na hladině významnosti $\alpha = 0.10$ zjistíte, zda existuje rozdíl mezi délkou klíční kosti z pravé strany u mužů z indické populace z Amritsaru a u mužů z indické populace z Varanasi.

Řešení příkladu 12.4

Datový soubor 18-more-samples-variances-clavicle.txt načteme příkazem read.delim(). Následně z načtených dat vybereme naměřené hodnoty délky klíční kosti z pravé strany (cla.L) u mužů z indické populace z Amritsaru (pop == 'ind1') a vložíme je do proměnné cla.LA. Dále z datové tabulky vybereme naměřené hodnoty délky klíční kosti z pravé strany u mužů z indické populace z Varanasi (pop == 'ind2') a vložíme je do proměnné cla.LV. Z obou proměnných cla.LA a cla.LV odstraníme příkazem na.omit() chybějící pozorování, dále příkazem length() zjistíme počet pozorování v obou náhodných výběrech a příkazem range() rozpětí naměřených hodnot.

```
131 data <- read.delim('00-Data//18-more-samples-variances-clavicle.txt')
132 cla.LA <- data[data$pop == 'ind1', 'cla.L']
133 cla.LA <- na.omit(cla.LA)
134 cla.LV <- data[data$pop == 'ind2', 'cla.L']
135 cla.LV <- na.omit(cla.LV)
136 n1 <- length(cla.LA) # 120
137 n2 <- length(cla.LV) # 81
138 range(cla.LA) # 123-169
139 range(cla.LV) # 126-165
```

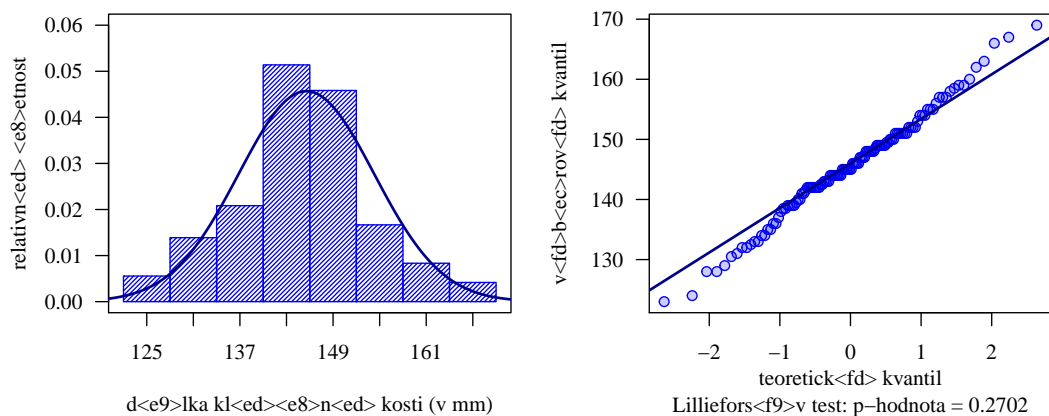
Datový soubor obsahuje naměřené hodnoty délky klíční kosti z pravé strany u 120 mužů z indické populace z Amritsaru u 81 mužů z indické populace z Varanasi. Naměřené délky klíčních kostí z pravé strany u mužů z Amritsaru nabývají hodnot v rozmezí od 123 mm do 169 mm, naměřené délky klíčních kosti z pravé strany u mužů Varanasi nabývají hodnot v rozmezí od 126 mm do 165 mm.

V rámci příkladu budeme chtít porovnat střední hodnoty dvou indických populací, přičemž u obou populací máme k dispozici naměřené hodnoty. Řešení příkladu vede primárně na parametrický test o rozdílu středních hodnot $\mu_1 - \mu_2$ (viz sekce ?? nebo sekce ??). Nejprve však musíme ověřit, zda oba náhodné výběry pochází z normálních rozdělání. Na hladině významnosti $\alpha = 0.05$ testujeme tedy nejprve hypotézu H_{01} : *Náhodný výběr délek klíčních kostí z pravé strany u mužů z indické populace z Amritsaru pochází z normálního rozdělání.* oproti alternativní hypotéze: H_{11} : *Náhodný výběr délek klíčních kostí z pravé strany u mužů z indické populace z Amritsaru nepochází z normálního rozdělání.* Normalitu náhodného výběru otestujeme pomocí Lillieforsova testu a graficky zhodnotíme histogramem superponovaným křivkou normálního rozdělání a QQ-diagramem (viz obrázek 13). Pro účely histogramu rozdělíme náhodný výběr délek klíčních kostí do 8 ekvidistantních třídicích intervalů o šířce 6 mm pomocí stanovených hranic 122, 128, ..., 170 mm.

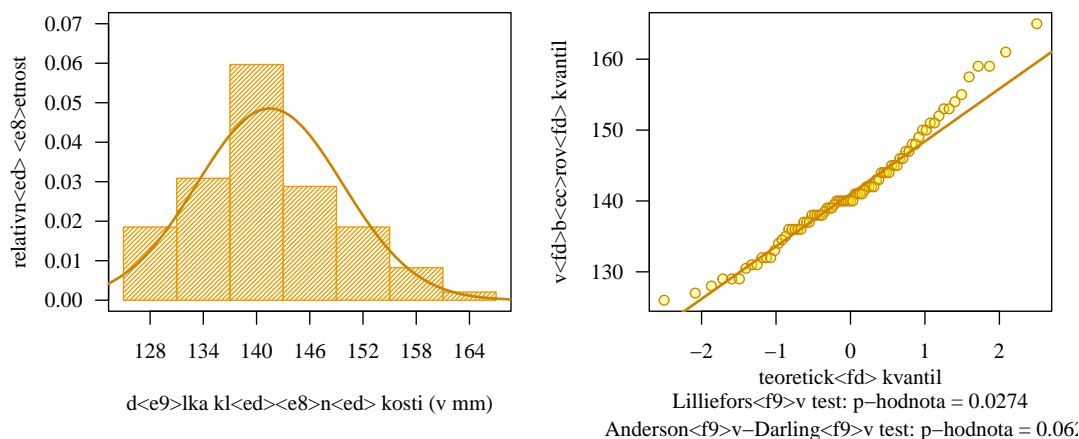
Protože p -hodnota Lillieforsova testu $p = 0.2702$ je větší než 0.05, H_{01} nuzamítáme na hladině významnosti $\alpha = 0.05$. Ani v histogramu ani v QQ-diagramu potom nepozorujeme žádné výrazné porušení předpokladu normality. Náhodný výběr délek klíčních kostí z pravé strany u mužů z populace z Amritsaru pochází z normálního rozdělání.

Analogicky testujeme hypotézu H_{02} : *Náhodný výběr délek klíčních kostí z pravé strany u mužů z indické populace z Varanasi pochází z normálního rozdělání.* oproti alternativní hypotéze: H_{12} : *Náhodný výběr délek klíčních kostí z pravé strany u mužů z indické populace z Varanasi nepochází z normálního rozdělání.* Normalitu náhodného výběru otestujeme pomocí Lillieforsova testu a graficky zhodnotíme histogramem superponovaným křivkou normálního rozdělání a QQ-diagramem (viz obrázek 14). Pro účely histogramu rozdělíme náhodný výběr do 7 ekvidistantních třídicích intervalů o šířce 6 mm prostřednictvím stanovených hranic 125, 131, ..., 167 mm.

Protože p -hodnota Lillieforsova testu $p = 0.0274$ je menší než 0.05, hypotézu H_{02} zamítáme na hladině významnosti $\alpha = 0.05$. Zamítnutí normality je však hraniční, protože p -hodnota neklesla pod hladinu významnosti $\alpha = 0.01$. Proto pro kontrolu otestujeme normalitu také pomocí Andersonova-Darlingova testu. Protože p -hodnota Andersonova-Darlingova testu $p = 0.0628$ je větší než 0.05, hypotézu H_{02} nezamítáme na hladině významnosti $\alpha = 0.05$. An-



Obrázek 13: Histogram (vlevo) a kvantilový diagram (vpravo) délky klíční kosti z pravé strany u mužů z indické populace z Amritsaru

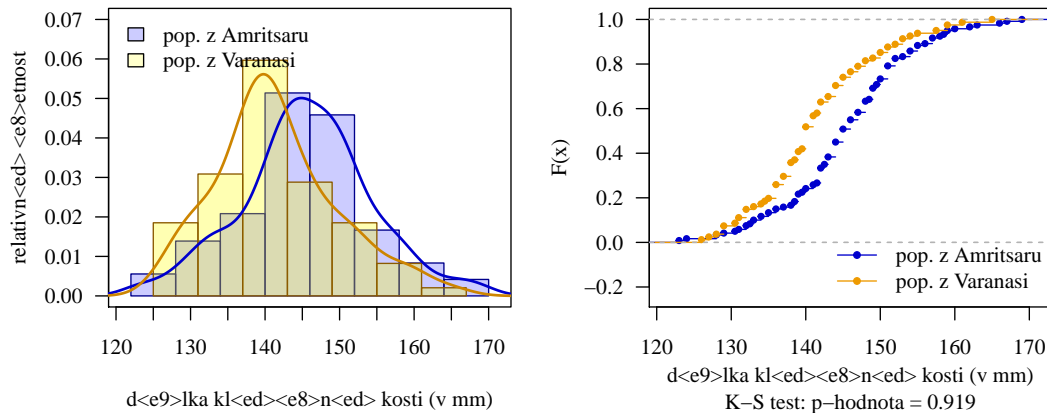


Obrázek 14: Histogram (vlevo) a kvantilový diagram (vpravo) délky klíční kosti z pravé strany u mužů z indické populace z Varanasi

Lilliefors-Darlingův test tedy hraničně normalitu náhodného výběru nezamítá. Bezpochyby se rozdělení náhodného výběru délky klíční kosti z pravé strany u mužů z indické populace z Varanasi pohybuje na pomezí normality. Grafické zhodnocení normality je v tomto případě velmi subjektivní. Z histogramu je patrná normalita náhodného výběru, nicméně je zde viditelné také kladné vyšiknutí hodnot s prodlouženým pravým koncem. QQ-diagram potom zobrazuje stejnou vlastnost prostřednictvím bodů v pravém horním rohu, které se odchylojí od referenční přímky. Situace je tedy opravdu hraniční a záleží pouze na nás, jak se o normalitě náhodného výběru rozhodneme. V praxi bychom pravděpodobně nad normalitou přimhouřili oko a hypotézu bychom vzhledem k sinějším vlastnostem parametrického testu otestovali pomocí klasického dvouvýběrového t -testu o rozdílu středních hodnot $\mu_1 - \mu_2$ tak, jako jsme to provedli v sekci ?? v příkladu ?. Zde si však ukážeme, jak bychom postupovali, kdybychom se rozhodli striktně držet výsledku Lillieforsova testu a přiklonili se k závěru, že náhodný výběr délek klíčních kostí z pravé strany u mužů z indické populace z Varanasi nepochází z normálního rozdělení.

Otázku ze zadání v takovémto případě otestujeme pomocí neparametrického Wilcoxonova dvouvýběrového testu, zde konkrétně vzhledem k vysokým rozsahům obou náhodných výběrů ($n_1 > 30$ i $n_2 > 30$) pomocí jeho asymptotické varianty. Jediným předpokladem, který musí být pro použití tohoto testu splněn, je shoda distribučních funkcí $F_1(x)$ a $F_2(x)$ (s výjimkou posunutí o vzdálenost Δ), kde $F_1(x)$ je distribuční funkce náhodného výběru z populace z Amritsaru a $F_2(x)$ je distribuční funkce náhodného výběru z populace z Varanasi. Shodu distribučních funkcí $F_1(x)$ a $F_2(x)$ ověříme pomocí Kolmogorova-Smirnovova testu (viz sekce ??).

Oba náhodné výběry nejprve metodou centrování posuneme do nuly, čímž odfiltrujeme vliv posunutí o vzdálenost Δ mezi oběma distribučními funkcemi. Na hladině významnosti $\alpha = 0.05$ potom testujeme hypotézu H_{03} : *Distribuční funkce centrované délky klíční kosti z pravé strany u mužů z Amritsar a u mužů z Varanasi jsou shodné.* oproti alternativní hypotéze: H_{13} : *Distribuční funkce centrované délky klíční kosti z pravé strany u mužů z Amritsar a u mužů z Varanasi nejsou shodné.* Ze sekce ?? již víme, že nulová hypotéza H_{03} odpovídá hypotéze o shodě dvou distribučních funkcí původních náhodných výběrů s výjimkou posunutí o vzdálenost Δ . Shodu distribučních funkcí $F_1(x)$ a $F_2(x)$ otestujeme pomocí příkazu `ks.test()` a následně vizualizujeme dvěma histogramy superponovanými příslušnými jádrovými odhady hustot a grafem zobrazujícím obě výběrové distribuční funkce vedle sebe (viz obrázek 15).



Obrázek 15: Porovnání histogramů (vlevo) a graf distribučních funkcí (vpravo) délky klíční kosti z pravé strany u mužů z indické populace z Varanasi

Protože p -hodnota K-S testu $p = 0.919$ je větší než 0.05 , nulovou hypotézu o shodě dvou distribučních funkcí centrovaných náhodných výběrů nezamítáme na hladině významnosti $\alpha = 0.05$. To tedy znamená, že distribuční funkce náhodného výběru délek klíčních kostí z pravé strany u mužů z populace z Amritsar a z populace z Varanasi se liší pouze posunutím.

Předpoklad umožňující použití Wilcoxonova dvouvýběrového asymptotického testu je splněn. Můžeme tedy přistoupit k samotnému procesu testování. Naším úkolem ze zadání je zjistit, zda existuje rozdíl mezi délkou klíční kosti z pravé strany u mužů z indické populace z Amritsar a u mužů z indické populace z Varanasi. Tato věta je zněním alternativní hypotézy, a to zejména proto, že slovo *rozdíl* implikuje nerovnost, která se nikdy nevyskytuje v nulové hypotéze. Druhou indicií by potom mohlo být, že nikde v zadání se nevyskytuje slovo *hypotéza*, ani zmínka o testování. Navíc v zadání není nikde zmínka o znění nulové hypotézy. a to jednak proto, že v zadání se přímo vyskytuje slovo *hypotéza*, Zbývá tedy dodefinovat nulovou hypotézu tak, aby byla doplňkem k zadané alternativní hypotéze. Závěrem poznamenejme, že pojem *střední hodnota* figurující v hypotézách, které testujeme parametrickým testem, zde nahrazujeme *mediánem*. Testování provedeme v posloupnosti sedmi kroků.

1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

H_0 : *Medián délky klíční kosti z pravé strany u mužů z populace z Amritsar je shodný s mediánem délky klíční kosti z pravé strany u mužů z populace z Varanasi.*

H_1 : *Medián délky klíční kosti z pravé strany u mužů z populace z Amritsar není shodný s mediánem délky klíční kosti z pravé strany u mužů z populace z Varanasi.*

- **matematická formulace** nulové a alternativní hypotézy

H_0 : $\tilde{x}_1 = \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 = \tilde{x}_0$, kde $\tilde{x}_0 = 0$

H_1 : $\tilde{x}_1 \neq \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 \neq \tilde{x}_0$, kde $\tilde{x}_0 = 0$ (oboustranná alternativa)

2. Volba hladiny významnosti

- Hladina významnosti $\alpha = 0.10$.

3. Testování kritickým oborem

• Testovací statistika

V prvním kroku vytvoříme společný vektor délek klíčních kostí tak, že za sebe poskládáme hodnoty délek klíčních kostí z pravé strany mužů z Amritsru (X_{1i} , $i = 1, \dots, 120$) a následně mužů z Varanasi (X_{2j} , $j = 1, \dots, 81$). Délka společného vektoru $n = n_1 + n_2 = 120 + 81 = 201$. Následně každému pozorování přiřadíme hodnotu indikační funkce I , která nabývá hodnoty 1, pokud měření pochází z druhého náhodného výběru, nebo hodnoty 0, pokud měření pochází z prvního náhodného výběru. Nakonec stanovíme pořadí S_k , $k = 1, \dots, 201$ všech 201 hodnot od nejmenší po největší (viz tabulka 4).

Tabulka 4: Naměřené hodnoty X_{1i} , $i = 1, \dots, 120$, a X_{2j} , $j = 1, \dots, 81$, obou náhodných výběrů, indikační funkce I příslušnosti k druhému náhodnému výběru a pořadí S_k , $k = 1, \dots, 201$, naměřených hodnot seřazených vzestupně


měření	1	2	...	119	120	121	122	123	124	125	126	127	128	129	130
X_i, Y_j	142	146	...	149	148	143	136	147	141	148	141	145	137	143	139
I	0	0	...	0	0	1	1	1	1	1	1	1	1	1	1
S_i, S_j	85.5	125	...	147	138.5	96.5	36	131.5	74.5	138.5	74.5	116.5	41.5	96.5	56
měření	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145
X_i, Y_j	132	142	150	150	138	135	136	127	149	138	145	128	155	140	138
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	19	85.5	155	155	46.5	31	36	4	147	46.5	116.5	6	180.5	66.5	46.5
měření	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
X_i, Y_j	136	140	140	140	142	132	147	129	140	159	134	142	137	140	126
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	36	66.5	66.5	66.5	85.5	19	131.5	9.5	66.5	191.5	27	85.5	41.5	66.5	3
měření	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
X_i, Y_j	144	139	144	138.5	145	153	153	146	138	152	141.5	130.5	131	144	151
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	105.5	56	105.5	51	116.5	173	173	125	46.5	169	78.5	12.5	15	105.5	162
měření	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190
X_i, Y_j	136	137	157.5	165	159	148	129	139.5	132	140	144	139	161	133	151
I	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	36	41.5	187	198	191.5	138.5	9.5	60.5	19	66.5	105.5	56	195	24	162
měření	191	192	193	194	195	196	197	198	199	200	201				
X_i, Y_j	138	141	134.5	136	141	131	129	140	146	142	154				
I	1	1	1	1	1	1	1	1	1	1	1				
S_i, S_j	46.5	74.5	29	36	74.5	15	9.5	66.5	125	85.5	176.5				

Z tabulky 4 vidíme, že 81 hodnotám příslušícím druhému náhodnému výběru patří ve vektoru 201 seřazených hodnot pořadí 96.5, 36, 131.5, 74.5, 138.5, 74.5, 116.5, 41.5, 96.5, 56, 19, 85.5, 155, 155, 46.5, 31, 36, 4, 147, 46.5, 116.5, 6, 180.5, 66.5, 46.5, 36, 66.5, 66.5, 66.5, 85.5, 19, 131.5, 9.5, 66.5, 191.5, 27, 85.5, 41.5, 66.5, 3, 105.5, 56, 105.5, 51, 116.5, 173, 173, 125, 46.5, 169, 78.5, 12.5, 15, 105.5, 162, 36, 41.5, 187, 198, 191.5, 138.5, 9.5, 60.5, 19, 66.5, 105.5, 56, 195, 24, 162, 46.5, 74.5, 29, 36, 74.5, 15, 9.5, 66.5, 125, 85.5 a 176.5. Součtem těchto pořadí získáme hodnotu statistiky T_2 . Následně dosazením hodnot n_1 , n_2 a statistiky T_2 do vzorce 12.1 získáme statistiku S_E . Konečně, dosazením statistiky S_E a rozsahů n_1 a n_2 do vzorce 12.2 získáme hodnotu testovací statistiky Wilcoxonova dvouvýběrového asymptotického testu S_A .

$$T_2 = \sum_{j=1}^{n_2} S_j = 96.5 + 36 + 131.5 + 74.5 + \dots + 66.5 + 125 + 85.5 + 176.5 = 6689.5$$

$$\begin{aligned} S_E &= n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2 \\ &= 120 \times 81 + \frac{81(81 + 1)}{2} - 6689.5 \\ &= 9720 + \frac{6642}{2} - 6689.5 \\ &= 9720 + 3321 - 6689.5 \\ &= 6351.5 \end{aligned}$$

$$\begin{aligned} S_A &= \frac{SE - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \\ &= \frac{6351.5 - \frac{81 \times 120}{2}}{\sqrt{\frac{120 \times 81 (120 + 81 + 1)}{12}}} \\ &= \frac{6351.5 - \frac{9720}{2}}{\sqrt{\frac{120 \times 81 \times 202}{12}}} \\ &= \frac{6351.5 - 4860}{\sqrt{163620}} \\ &= \frac{1491.5}{404.4997} \\ &= 3.687271 \end{aligned}$$

Výpočet testovací statistiky si provedeme také pomocí softwaru . Nejprve stanovíme vzestupné pořadí 201 naměřených hodnot obou náhodných výběrů pomocí funkce `rank()`. Indikační funkci I vytvoříme spojením 120 nul a 81 jedniček prostřednictvím kombinace funkcí `c()` a `rep()`. Vektor měření, indikačních hodnot a pořadí potom spojíme do matice, kterou převedeme na tabulku. Výpočet statistiky T_2 je potom součtem pořadí S na těch pozicích, na nichž jsou ve vektoru indikačních hodnot 1 jedničky. Dále stanovíme hodnotu testovací statistiky SE přímým přepisem vzorce 12.1 a nakonec hodnotu testovací statistiky S_A přímým přepisem vzorce 12.2.

```

140 S <- rank(c(c1a.LA, c1a.LV))
141 I <- c(rep(0, n1), rep(1, n2))
142 tab <- rbind("Xi, Yj" = c(c1a.LA, c1a.LV), "I" = I, "Si, Sj" = S)
143 tab <- data.frame(tab)
144 names(tab) <- 1 : 201
145 #           1     2     3     4     5     6     ...     195 196     197     198 199     200     201
146 # Xi, Yj 142.0 146 136 138.5 129.0 141.0 ... 141.0 131 129.0 140.0 146 142.0 154.0
147 # I       0.0  0   0   0   0   0   0 ...   1   1   1   1   1   1   1
148 # Si, Sj  85.5 125 36  51.0  9.5  74.5 ...  74.5 15  9.5  66.5 125  85.5 176.5
149 T2 <- sum(S[I == 1]) # 6689.5
150 SE <- n1 * n2 + n2 * (n2 + 1) / 2 - T2 # 6351.5
151 SA <- (SE - n1 * n2 / 2) / sqrt(n1 * n2 * (n1 + n2 + 1) / 12) # 3.687271

```

- **Kritický obor**

$$\begin{aligned}
 W &= (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty) \\
 &= (-\infty; u_{0.10/2}) \cup (u_{1-0.10/2}; \infty) \\
 &= (-\infty; u_{0.05}) \cup (u_{0.95}; \infty) \\
 &= (-\infty; -1.644854) \cup (1.644854; \infty)
 \end{aligned}$$

K výpočtu kvantilů $u_{\alpha/2}$ a $u_{1-\alpha/2}$ použijeme funkci `qnorm()` počítající hodnotu $\alpha/2$ -kvantilu (resp. $(1-\alpha/2)$ -kvantilu) standardizovaného normálního rozdělení. Vstupním argumentem funkce `qnorm()` bude pouze hodnota $\alpha/2$, resp. $(1-\alpha/2)$.

```

152 alpha <- 0.10
153 qnorm(alpha / 2) # -1.644854
154 qnorm(1 - alpha / 2) # 1.644854

```

- **Závěr testování**

Protože realizace testovací statistiky $s_A = 3.687271$ náleží do kritického oboru, tj. $s_A \in W$, H_0 zamítáme na hladině významnosti $\alpha = 0.10$.

4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

Ke stanovení dolní a horní hranice 90% intervalu spolehlivosti je třeba nejprve vypočítat $n_1 \times n_2 = 120 \times 81 = 9720$ rozdílů $X_{1i} - X_{2j}$, $i = 1, \dots, 120$, $j = 1, \dots, 81$, a ty následně vzestupně seřadit. Rozdíly $X_{1i} - X_{2j}$ vypočítáme hromadně pomocí funkce `outer()`. Prvními dvěma vstupními argumenty této funkce budou vektory délek klíčích kostí `cla.LA` a `cla.LV`. Dále specifikujeme hodnotu argumentu `FUN`, kde nastavením `FUN = '-'` změním výpočet součinů $X_{1i}X_{2j}$ (defaultní nastavení) na výpočet rozdílů $X_{1i} - X_{2j}$, $i = 1, \dots, 120$, $j = 1, \dots, 81$. Získané rozdíly $X_{1i} - X_{2j}$, $i = 1, \dots, 120$, $j = 1, \dots, 81$ seřadíme příkazem `sort()` a vložíme do proměnné `U`. Následně vypočítáme pozicové souřadnice $C_{1-\alpha/2}$ a $n_1n_2 + 1 - C_{1-\alpha/2}$.

$$\begin{aligned}
 C_{1-\alpha/2} &= \frac{n_1n_2}{2} - u_{1-\alpha/2} \sqrt{\frac{n_1n_2(n_1 + n_2 + 1)}{12}} \\
 &= \frac{120 \times 81}{2} - u_{1-0.10/2} \sqrt{\frac{120 \times 81(120 + 81 + 1)}{12}} \\
 &= 4860 - u_{0.95} \sqrt{\frac{9720 \times 202}{12}} \\
 &= 4860 - 1.644854 \sqrt{\frac{9720 \times 202}{12}} \\
 &= 4860 - 1.644854 \sqrt{163620} \\
 &= 4860 - 1.644854 \times 404.4997 \\
 &= 4194.657 \doteq 4194
 \end{aligned}$$

$$\begin{aligned}
 n_1n_2 + 1 - C_{1-\alpha/2} &= 120 \times 81 + 1 - 4194 \\
 &= 9720 + 1 - 4194 = 5527
 \end{aligned}$$

Hranice intervalu spolehlivosti potom tvoří ty hodnoty, které se v seřazeném vektoru rozdílů `U` nachází na 4194. pozici a na 5527. pozici.

```

155 rozdily <- outer(cla.LA, cla.LV, FUN = '-')
156 U <- sort(rozdily)

```

$$\begin{aligned}
(d, h) &= \left(U^{(C_{1-\alpha/2})}; U^{(n_1 n_2 + 1 - C_{1-\alpha/2})} \right) \\
&= \left(U^{(4194)}; U^{(5527)} \right) \\
&= (3; 6.5)
\end{aligned}$$

```

157 C1 <- floor(n1 * n2 / 2 - qnorm(1 - alpha / 2) * sqrt(n1 * n2 * (n1 + n2 + 1) / 12)) #
    4194
158 C2 <- n1 * n2 + 1 - C1 # 5527
159 U[4194] # 3
160 U[5527] # 6.5

```

- **Závěr testování**

Protože $\tilde{x}_0 = 0$ nenáleží do 90% empirického oboustranného intervalu spolehlivosti, tj. $\tilde{x}_0 = 0 \notin IS$, H_0 zamítáme na hladině významnosti $\alpha = 0.10$.

5. Testování p -hodnotou

Příslušnou p -hodnotu vypočítáme pomocí vzorce $2 \min\{\Pr(S_A \leq s_A), \Pr(S_A \geq s_A)\}$. Zde si uvědomme, že realizace testovací statistiky $s_A = 3.6873$. Zároveň S_A je spojitá náhodná veličina, neboť pochází ze spojitého (standardizovaného normálního) rozdělení $N(0, 1)$. Z vlastností pravděpodobnostní funkce spojitých náhodných veličin víme, že $\Pr(S_A \geq 3.6873) = 1 - \Pr(S_A < 3.6873) = 1 - \Pr(S_A \leq 3.6873)$. Viz kapitola ??.

- **p -hodnota**

$$\begin{aligned}
p\text{-hodnota} &= 2 \min\{\Pr(S_A \leq s_A), \Pr(S_A \geq s_A)\} \\
&= 2 \min\{\Pr(S_A \leq 3.687271), 1 - \Pr(S_A \leq 3.687271)\} \\
&= 2 \min\{0.9998867, 0.0001133359\} \\
&= 2 \times 0.0001133359 = 0.0002266718 \doteq 0.0002267
\end{aligned}$$

Hodnoty distribuční funkce standardizovaného normálního rozdělení v hodnotě $s_A = -4.522871$, tj. $\Pr(S_A \leq -4.522871)$ vypočítáme pomocí funkce `pnorm()` (viz sekce ??).

```

161 p.hodn <- 2 * min(pnorm(SA), 1 - pnorm(SA)) # 0.0002266719

```

- **Závěr testování**

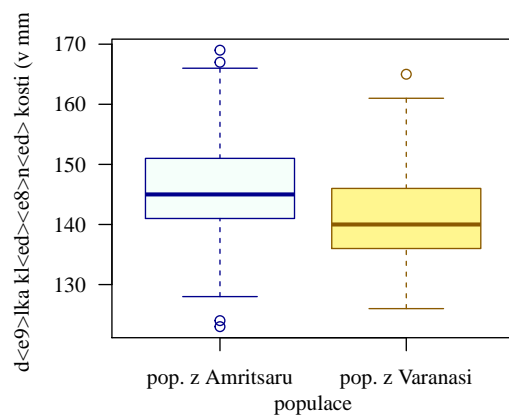
Protože p -hodnota = 0.0002267 je menší než $\alpha = 0.10$, H_0 zamítáme na hladině významnosti $\alpha = 0.10$.

6. Interpretace výsledků

Na základě všech tří typů testování zamítáme nulovou hypotézu na hladině významnosti $\alpha = 0.10$. Mezi délkou klíční kosti z pravé strany u mužů z indické populace z Amritsaru a u mužů z indické populace z Varanasi existuje statisticky významný rozdíl. Ke stejnému závěru jsme došli také v příkladu ??, kde jsme se přiklonili k normalitě náhodného výběru délky klíční kosti z pravé strany u mužů z populace z Varanasi a na otestování nulové hypotézy jsme použili parametrický klasický dvouvýběrový t -test.

7. Grafická vizualizace výsledků testování

Vhodným grafem porovnávajícím oba náhodné výběry právě z hlediska mediánů je krabicový diagram (viz obrázek 16).



Obrázek 16: Krabicový diagram délky klívní kosti z pravé strany u mužů z indických populací z Amritsaru a z Varanasi

Poznámka: Funkce `wilcox.test()` neposkytuje výsledky pro asymptotickou variantu Wilcoxonova dvouvýběrového testu. Funkci můžeme využít maximálně na výpočet testovací statistiky S_E , pomocí které pak snadno vypočítáme přepisem vzorce 12.2 hodnotu testovací statistiky S_A . Hranice kritického oboru, intervalu spolehlivosti a p .hodnoty však musíme vypočítat samostatně. ★

Příklad 12.5. Wilcoxonův dvouvýběrový asymptotický test (pravostranná alternativa)

Mějme datový soubor 22-two-samples-whr-mf.csv obsahující údaje o věku (age) a poměru obvodu pasu a boků (WHR) u dětí (dívek a chlapců) ve věku 6–16 let (viz sekce ??). Na hladině významnosti $\alpha = 0.05$ zjistíte, zda je obvod poměru pasu a boků u chlapců ve věku do deseti let (včetně) větší než u dívek ve věku do deseti let (včetně).

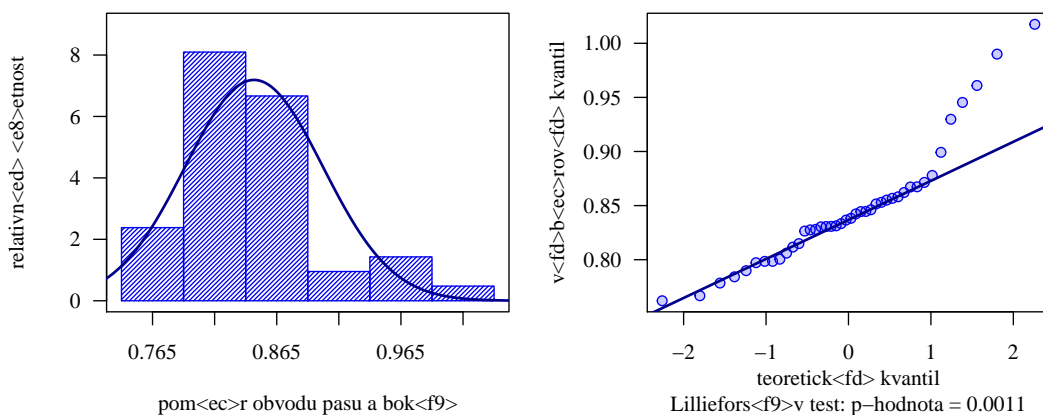
Řešení příkladu 12.5

Nejprve načteme datový soubor 22-two-samples-whr-mf.csv. Následně z načtených dat vybereme hodnoty obvodu pasu a boků (WHR) u chlapců ($\text{sex} == \text{'m'}$) ve věku do deseti let ($\text{age} \leq 10$) a vložíme je do proměnné WHR.M. Dále vybereme z datové tabulky hodnoty obvodu pasu a boků (WHR) u dívek ($\text{sex} == \text{'f'}$) ve věku do deseti let ($\text{age} \leq 10$) a vložíme je do proměnné WHR.F. Z obou proměnných WHR.M a WHR.F odstraníme chybějící pozorování. Nakonec zjistíme rozsahy obou náhodných výběrů a rozpětí naměřených hodnot v obou výběrech.

```
162 data <- read.delim('00-Data//22-two-samples-whr-mf.csv', sep = ';', dec = '.')
163 WHR.M <- data[data$sex == 'm' & data$age <= 10, 'WHR']
164 WHR.F <- data[data$sex == 'f' & data$age <= 10, 'WHR']
165 WHR.M <- na.omit(WHR.M)
166 WHR.F <- na.omit(WHR.F)
167 n1 <- length(WHR.M) # 42
168 n2 <- length(WHR.F) # 47
169 range(WHR.M) # 0.7620438-1.0175439
170 range(WHR.F) # 0.6873315-0.8878981
```

Datový soubor obsahuje hodnoty poměru obvodu pasu a boků u 42 chlapců a 47 dívek ve věku do 10 let. Poměr obvodu pasu a boků u měřených chlapců se pohybuje v rozmezí přibližně 0.76–1.02, poměr obvodu pasu a boků u měřených dívek se pohybuje v rozmezí přibližně 0.69–0.89.

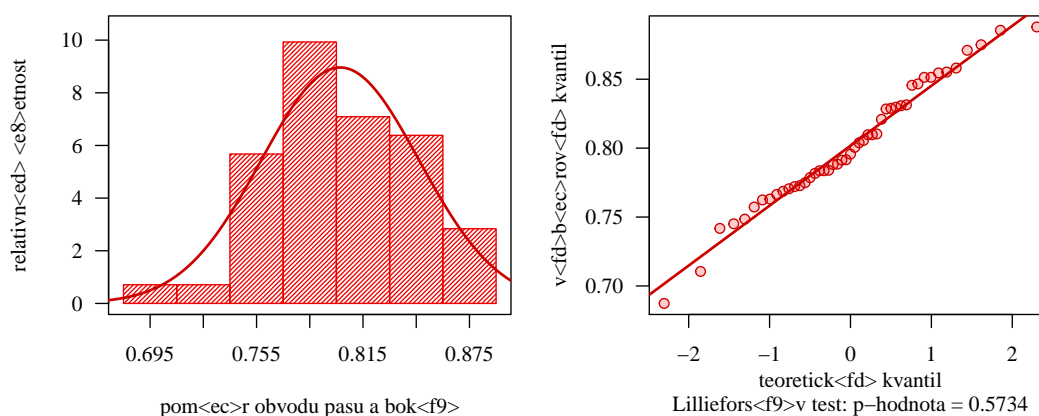
V rámci příkladu budeme chtít porovnat střední hodnotu populace chlapců se střední hodnotou populace dívek, přičemž u obou populací máme k dispozici naměřené hodnoty. Řešení příkladu vede primárně na parametrický test o rozdílu středních hodnot $\mu_1 - \mu_2$. Nejprve však musíme ověřit, zda je splněn nezbytný předpoklad normality obou náhodných výběrů. Na hladině významnosti $\alpha = 0.05$ testujeme tedy nejprve hypotézu H_{01} : *Náhodný výběr poměrů obvodu pasu a boků u chlapců ve věku do 10 let pochází z normálního rozdělení*. oproti alternativní hypotéze: H_{11} : *Náhodný výběr poměrů obvodu pasu a boků u chlapců ve věku do 10 let nepochází z normálního rozdělení*. Normalitu náhodného výběru otestujeme vzhledem k rozsahu náhodného výběru ($n_1 = 42 > 3$) Lillieforsovým testem. Graficky ji potom zhodnotíme histogramem superponovaným křivkou normálního rozdělení a QQ-diagramem (viz obrázek 17). V rámci histogramu rozdělíme náhodný výběr poměrů obvodu pasu a boků do 6 ekvidistantních třídících intervalů o šířce 0.05 pomocí stanovených hranic 0.74, 0.79, ..., 1.04.



Obrázek 17: Histogram (vlevo) a kvantilový diagram (vpravo) poměru obvodu pasu a boků u chlapců ve věku do deseti let

Protože p -hodnota Lillieforsova testu $p = 0.0011$ je větší než 0.05 , H_{01} zamítáme na hladině významnosti $\alpha = 0.05$. Jak v histogramu, tak v QQ-diagramu pozorujeme porušení normality prostřednictvím napravo umístěných odlehlých pozorování. Náhodný výběr poměrů obvodu pasu a boků u chlapců ve věku do deseti let nepochází z normálního rozdělení.

Již v tomto okamžiku můžeme usoudit, že parametrický test nebude možné použít z důvodu výše uvedeného porušení předpokladu normality. Pro pořádek a kompletní náhled na situaci otestujeme a vizualizujeme také normalitu v náhodném výběru dívek. Na hladině významnosti $\alpha = 0.05$ testujeme nyní hypotézu H_{02} : *Náhodný výběr poměrů obvodu pasu a boků u dívek ve věku do 10 let pochází z normálního rozdělení.* oproti alternativní hypotéze: H_{11} : *Náhodný výběr poměrů obvodu pasu a boků u dívek ve věku do 10 let nepochází z normálního rozdělení.* Normalitu náhodného výběru otestujeme vzhledem k rozsahu náhodného výběru ($n_2 = 47 > 30$) opět pomocí Lillieforsova testu. Grafická vizualizace sestávající z histogramu a QQ-diagramu je potom vyobrazena na obrázku 18. Pro účely histogramu rozdělíme náhodný výběr do 7 ekvidistantních třídících intervalů o šířce 0.03 prostřednictvím stanovených hranic $0.68, 0.71, \dots, 0.89$.

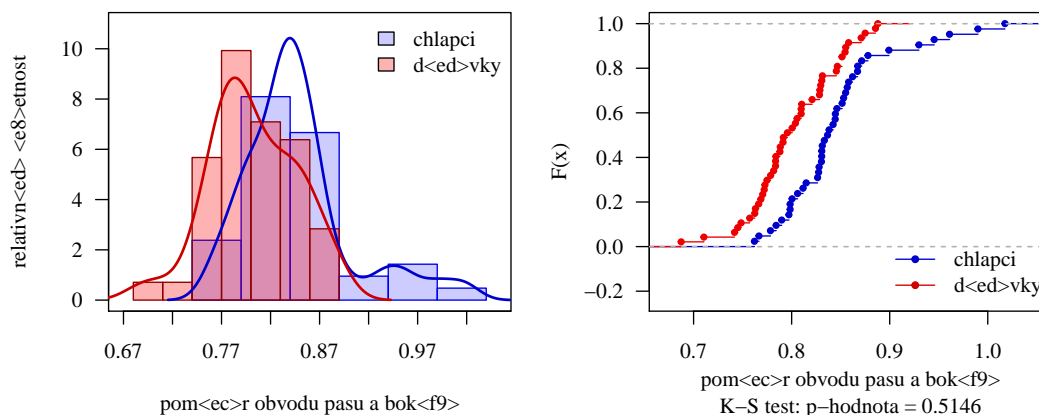


Obrázek 18: Histogram (vlevo) a kvantilový diagram (vpravo) poměru obvodu pasu a boků u dívek ve věku do deseti let

Protože p -hodnota Lillieforsova testu $p = 0.5734$ je větší než 0.05 , hypotézu H_{02} nezamítáme na hladině významnosti $\alpha = 0.05$. Pohled na histogram by nás možná mohl vést k závěru, že normalita náhodného výběru není příliš přesvědčivá a že vykazuje podobné rysy jako vykazoval náhodný výběr poměrů měřených u chlapců. Nicméně, oproti předchozímu náhodnému výběru, obsahuje náhodný výběr poměrů u dívek menší množství odlehlých pozorování a větší koncentraci hodnot okolo střední hodnoty. Z QQ-diagramu je potom patrná téměř ukázková normalita, neboť zobrazené body se drží ve velmi těsné blízkosti referenční příčky. Náhodný výběr poměrů obvodu pasu a boků u dívek ve věku do 10 let tedy pochází z normálního rozdělení.

Vzhledem k zamítnutí hypotézy o normalitě náhodného výběru poměrů obvodu pasu a boků u chlapců budeme otázku ze zadání prověřovat pomocí neparametrického testu. Vzhledem k vyšším rozsahům obou náhodných výběrů ($n_1 > 30$ i $n_2 > 30$) použijeme Wilcoxonův dvouvýběrový asymptotický test. Před jeho použitím ověříme shodu distribučních funkcí $F_1(x)$ a $F_2(x)$ (s výjimkou posunutí o vzdálenost Δ), kde $F_1(x)$ je distribuční funkce náhodného výběru chlapců a $F_2(x)$ je distribuční funkce náhodného výběru dívek.

Oba náhodné výběry nejprve metodou centrování posuneme do nuly, čímž odfiltrujeme vliv posunutí o vzdálenost Δ . Na hladině významnosti $\alpha = 0.05$ potom testujeme hypotézu H_{03} : *Distribuční funkce poměru obvodu pasu a boků u chlapců ve věku do deseti let a u dívek ve věku do 10 let jsou shodné.* oproti alternativní hypotéze: H_{13} : *Distribuční funkce poměru obvodu pasu a boků u chlapců ve věku do deseti let a u dívek ve věku do 10 let nejsou shodné.* Nulová hypotéza H_{03} odpovídá hypotéze o shodě dvou distribučních funkcí původních náhodných výběrů s výjimkou posunutí o vzdálenost Δ . Shodu distribučních funkcí $F_1(x)$ a $F_2(x)$ otestujeme pomocí Kolmogorova-Smirnovova testu a následně vizualizujeme dvěma histogramy superponovanými příslušnými jádrovými odhady hustot a grafem obou výběrových distribučních funkcí (viz obrázek 19).



Obrázek 19: Porovnání histogramů (vlevo) a graf distribučních funkcí (vpravo) poměru obvodu pasu a boků u chlapců ve věku do deseti let a u dívek ve věku do deseti let

Protože p -hodnota K-S testu $p = 0.5146$ je větší než 0.05, nulovou hypotézu o shodě dvou distribučních funkcí centrovaných náhodných výběrů nezamítáme na hladině významnosti $\alpha = 0.05$. To tedy znamená, že distribuční funkce poměrů obvodu pasu a boků u chlapců do deseti let a u dívek do deseti let se liší pouze posunutím.

Předpoklad umožňující použití Wilcoxonova dvouvýběrového asymptotického testu je splněn. Naším úkolem ze zadání je zjistit, zda je obvod poměru pasu a boků u chlapců ve věku do deseti let větší než u dívek ve věku do deseti let. Tato věta je zněním alternativní hypotézy, neboť ve větě není zmínka o znění nulové hypotézy, ani o testování hypotézy, ani o rovnosti. Zbývá pouze dodefinovat znění nulové hypotézy tak, aby byla doplňkem k alternativní hypotéze. Pojem *střední hodnota* figurující v hypotézách, které testujeme parametrickým testem, opět nezapomeneme nahradit *mediánem*.

1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

H_0 : Medián poměru obvodu pasu a boků u chlapců ve věku do deseti let je menší nebo roven mediánu poměru obvodu pasu a boků u dívek ve věku do deseti let.

H_1 : Medián poměru obvodu pasu a boků u chlapců ve věku do deseti let je větší než medián poměru obvodu pasu a boků u dívek ve věku do deseti let.

- **matematická formulace** nulové a alternativní hypotézy

H_0 : $\tilde{x}_1 \leq \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 \leq \tilde{x}_0$, kde $\tilde{x}_0 = 0$

H_1 : $\tilde{x}_1 > \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 > \tilde{x}_0$, kde $\tilde{x}_0 = 0$ (pravostranná alternativa)

2. Volba hladiny významnosti

- Hladina významnosti $\alpha = 0.05$.

3. Testování kritickým oborem

- **Testovací statistika**

V prvním kroku vytvoříme společný vektor poměru obvodu pasu a boků tak, že za sebe poskládáme hodnoty poměrů obvodu pasu a boků u chlapců (X_{1i} , $i = 1, \dots, 42$) a následně u dívek (X_{2j} , $j = 1, \dots, 47$). Délka společného vektoru $n = n_1 + n_2 = 42 + 47 = 89$. Následně každému pozorování přiřadíme hodnotu indikační funkce I , která nabývá hodnoty 1, pokud měření pochází z druhého náhodného výběru, nebo hodnoty 0, pokud měření pochází z prvního náhodného výběru. Nakonec stanovíme pořadí S_k , $k = 1, \dots, 89$ všech 89 hodnot od nejmenší po největší (viz tabulka 5, kde naměřené hodnoty poměru obvodu pasu a boků pro přehlednost zaokrouhluje na čtyři desetinná místa).

Tabulka 5: Naměřené hodnoty X_{1i} , $i = 1, \dots, 42$, a X_{2j} , $j = 1, \dots, 47$, obou náhodných výběrů, indikační funkce I příslušnosti k druhému náhodnému výběru a pořadí S_k , $k = 1, \dots, 89$, naměřených hodnot seřazených vzestupně

měření	1	2	...	41	42	43	44	45	46	47	48
X_i, Y_j	0.7984	0.7667	...	0.8382	0.828	0.7913	0.7836	0.8855	0.8545	0.8308	0.7624
I	0	0	...	0	0	1	1	1	1	1	1
S_i, S_j	31	11	...	58	46	27	21	82	69	52.5	8
měření	49	50	51	52	53	54	55	56	57	58	59
X_i, Y_j	0.8098	0.8297	0.7728	0.8750	0.7485	0.8466	0.7915	0.7956	0.7884	0.8039	0.7836
I	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	39	49	15	80	5	64	28	29	25	35	20
měření	60	61	62	63	64	65	66	67	68	69	70
X_i, Y_j	0.7451	0.8209	0.7418	0.8097	0.7839	0.7817	0.7687	0.763	0.8552	0.8879	0.8289
I	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	4	43	3	38	22	19	12	9	71	83	48
měření	71	72	73	74	75	76	77	78	79	80	81
X_i, Y_j	0.8314	0.7705	0.7573	0.8103	0.8284	0.8514	0.8456	0.7786	0.7664	0.8581	0.6873
I	1	1	1	1	1	1	1	1	1	1	1
S_i, S_j	55	13	6	40	47	65.5	62	18	10	73.5	1
měření	82	83	84	85	86	87	88	89			
X_i, Y_j	0.871	0.7105	0.7721	0.7749	0.7881	0.8006	0.8056	0.8514			
I	1	1	1	1	1	1	1	1			
S_i, S_j	78	2	14	16	24	34	36	65.5			

Z tabulky 5 vidíme, že 47 hodnotám příslušícím druhému náhodnému výběru patří ve vektoru 89 seřazených hodnot pořadí 27, 21, 82, 69, 52.5, 8, 39, 49, 15, 80, 5, 64, 28, 29, 25, 35, 20, 4, 43, 3, 38, 22, 19, 12, 9, 71, 83, 48, 55, 13, 6, 40, 47, 65.5, 62, 18, 10, 73.5, 1, 78, 2, 14, 16, 24, 34, 36 a 65.5. Součtem těchto pořadí získáme hodnotu statistiky T_2 . Následně dosazením hodnot n_1 , n_2 a statistiky T_2 do vzorce 12.1 získáme hodnotu statistiky S_E . Konečně, dosazením statistiky S_E a rozsahů n_1 a n_2 do vzorce 12.2 získáme hodnotu testovací statistiky Wilcoxonova dvouvýběrového asymptotického testu S_A .

$$T_2 = \sum_{j=1}^{n_2} S_j = 31 + 11 + 51 + 77 + \dots + 24 + 34 + 36 + 65.5 = 1661$$

$$\begin{aligned} S_E &= n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2 \\ &= 42 \times 47 + \frac{47(47 + 1)}{2} - 1661 \\ &= 1974 + \frac{2256}{2} - 1661 \\ &= 1974 + 1128 - 1661 \\ &= 1441 \end{aligned}$$

$$\begin{aligned}
S_A &= \frac{SE - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \\
&= \frac{1441 - \frac{42 \times 47}{2}}{\sqrt{\frac{42 \times 47 (42 + 47 + 1)}{12}}} \\
&= \frac{1441 - \frac{1974}{2}}{\sqrt{\frac{42 \times 47 \times 90}{12}}} \\
&= \frac{1441 - 987}{\sqrt{14805}} \\
&= \frac{454}{121.6758} \\
&= 3.731227
\end{aligned}$$

Výpočet testovací statistiky si provedeme také pomocí softwaru \mathbb{R} . Nejprve stanovíme vzestupné pořadí 89 naměřených hodnot obou náhodných výběrů (funkce `rank()`). Indikační funkci I vytvoříme spojením 42 nul a 47 jedniček (kombinace funkcí `c()` a `rep()`). Výpočet statistiky T_2 je potom součtem pořadí S na těch pozicích, na nichž jsou ve vektoru indikačních hodnot I jedničky. Dále přímým přepisem vzorců 12.1 a 12.2 stanovíme hodnoty testovacích statistik SE a SA .

```

171 S <- rank(c(WHR.M, WHR.F))
172 I <- c(rep(0, n1), rep(1, n2))
173 tab <- rbind("Xi, Yj" = c(WHR.M, WHR.F), "I" = I, "Si, Sj" = S)
174 tab <- data.frame(tab)
175 names(tab) <- 1 : 89
176 #           1           2           3           4 ...           86           87           88           89
177 # Xi, Yj  0.7984  0.7667  0.8306  0.8673 ...  0.7881  0.8006  0.8056  0.8514
178 # I       0.0000  0.0000  0.0000  0.0000 ...  1.0000  1.0000  1.0000  1.0000
179 # Si, Sj 31.0000 11.0000 51.0000 77.0000 ... 24.0000 34.0000 36.0000 65.5000
180 T2 <- sum(S[I == 1]) # 1661
181 SE <- n1 * n2 + n2 * (n2 + 1) / 2 - T2 # 1441
182 SA <- (SE - n1 * n2 / 2) / sqrt(n1 * n2 * (n1 + n2 + 1) / 12) # 3.731227

```

• Kritický obor

$$\begin{aligned}
W &= \langle u_{1-\alpha}; \infty \rangle \\
&= \langle u_{1-0.05}; \infty \rangle \\
&= \langle u_{0.95}; \infty \rangle \\
&= \langle 1.644854; \infty \rangle
\end{aligned}$$

K výpočtu kvantilu $u_{1-\alpha}$ použijeme funkci `qnorm()` počítající hodnotu $1 - \alpha$ -kvantilu standardizovaného normálního rozdělení.

```

183 alpha <- 0.05
184 qnorm(1 - alpha) # 1.644854

```

- **Závěr testování**

Protože realizace testovací statistiky $s_A = 3.7312$ náleží do kritického oboru, tj. $s_A \in W$, H_0 zamítáme na hladině významnosti $\alpha = 0.05$.

4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

Oproti pravostranné alternativě postavíme levostranný interval spolehlivosti. Ke stanovení dolní hranice 95 % intervalu spolehlivosti je třeba nejprve vypočítat $n_1 \times n_2 = 42 \times 47 = 1974$ rozdílů $X_{1i} - X_{2j}$, $i = 1, \dots, 42$, $j = 1, \dots, 47$, a ty následně vzestupně seřadit. Rozdíly $X_{1i} - X_{2j}$ vypočítáme hromadně pomocí funkce `outer()`. Prvními dvěma vstupními argumenty této funkce budou vektory hodnot poměrů obvodu pasu a boků WHR.M a WHR.F. Dále specifikujeme výpočet rozdílů $X_{1i} - X_{2j}$, $i = 1, \dots, 42$, $j = 1, \dots, 47$ nastavením argumentu `FUN = '-'`. Získané rozdíly $X_{1i} - X_{2j}$, $i = 1, \dots, 42$, $j = 1, \dots, 47$ seřadíme (`sort()`) a vložíme do proměnné `U`. Následně vypočítáme pozicovou souřadnici $C_{1-\alpha}$.

$$\begin{aligned}
 C_{1-\alpha} &= \frac{n_1 n_2}{2} - u_{1-\alpha} \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \\
 &= \frac{42 \times 47}{2} - u_{1-0.05} \sqrt{\frac{42 \times 47 (42 + 47 + 1)}{12}} \\
 &= 987 - u_{0.95} \sqrt{\frac{1974 \times 90}{12}} \\
 &= 987 - 1.644854 \sqrt{\frac{1974 \times 90}{12}} \\
 &= 987 - 1.644854 \sqrt{14805} \\
 &= 987 - 1.644854 \times 121.6758 \\
 &= 786.8611 \doteq 786
 \end{aligned}$$

```

185 rozdily <- outer(WHR.M, WHR.F, FUN = '-')
186 U <- sort(rozdily)

```

Hranice intervalu spolehlivosti potom tvoří ty hodnoty, které se v seřazeném vektoru rozdílů `U` nachází na 786. pozici a nekonečno.

$$\begin{aligned}
 (d, h) &= (U^{(C_{1-\alpha})}; \infty) \\
 &= (U^{(786)}; \infty) \\
 &= (0.02197848; \infty)
 \end{aligned}$$

```

187 C1 <- floor(n1 * n2 / 2 - qnorm(1 - alpha) * sqrt(n1 * n2 * (n1 + n2 + 1) / 12)) # 786
188 U[786] # 0.02197848

```

- **Závěr testování**


Protože $\tilde{x}_0 = 0$ nenáleží do 95% empirického levostranného intervalu spolehlivosti, tj. $\tilde{x}_0 = 0 \notin IS$, H_0 zamítáme na hladině významnosti $\alpha = 0.05$.

5. Testování p -hodnotou

Příslušnou p -hodnotu vypočítáme pomocí vzorce $\Pr(S_A \geq s_A)$, kde $s_A = -5.53931$ je realizace testovací statistiky S_A . Zároveň S_A je spojitá náhodná veličina, neboť pochází ze spojitého (standardizovaného normálního) rozdělení $N(0, 1)$. Z vlastností pravděpodobnostní funkce spojitých náhodných veličin víme, že $\Pr(S_A \geq -5.53931) = 1 - \Pr(S_A < -5.53931) = 1 - \Pr(S_A \leq -5.53931)$.

- **p -hodnota**

$$p\text{-hodnota} = \Pr(S_A \geq s_A) = 1 - \Pr(S_A \leq -5.53931) = 2 \times 0.198446 = 6.100637 \times 10^{-6}$$

Hodnotu distribuční funkce standardizovaného normálního rozdělení v hodnotě $s_A = 3.731227$, tj. $\Pr(S_A \leq 3.731227)$ vypočítáme pomocí funkce `pnorm()`. Z níže uvedeného výstupu vidíme, že software  zaokrouhlil p -hodnotu na 1. Z kapitoly ?? však již víme, že distribuční funkce nabývá hodnoty 1 pouze v limitním případě, když náhodná veličina $S_A \rightarrow \infty$. Abychom tedy byli korektní, budeme výslednou p -hodnotu prezentovat jako číslo > 0.9999 .

```
189 p.hodn <- 1 - pnorm(SA) # 9.527479e-05
```

- **Závěr testování**

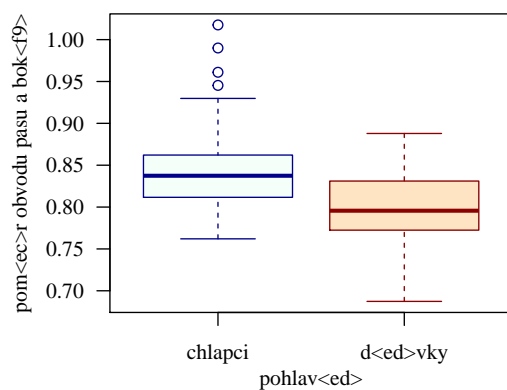
Protože p -hodnota = 9.5275×10^{-5} je menší než $\alpha = 0.05$, H_0 zamítáme na hladině významnosti $\alpha = 0.05$.

6. Interpretace výsledků

Na základě všech tří typů testování zamítáme nulovou hypotézu na hladině významnosti $\alpha = 0.05$. Poměr obvodu pasu a boků u chlapců ve věku do deseti let je statisticky významně větší než poměr obvodu pasu a boků u dívek ve věku do deseti let.

7. Grafická vizualizace výsledků testování

Statisticky významný rozdíl mezi oběma náhodnými výběry vizualizujeme porovnáním krabicových diagramů (viz obrázek 20).



Obrázek 20: Krabicový diagram poměru obvodu pasu a boků u chlapců ve věku do deseti let a u dívek ve věku do deseti let



Příklad 12.6. Wilcoxonův dvouvýběrový asymptotický test (levostranná alternativa)

Mějme datový soubor 10-two-samples-means-birth.txt obsahující údaje o porodní hmotnosti (birth.W) a počtu starších sourozenců (o.sib.N) u novorozenců narozených v krajské nemocnici v průběhu jednoho roku (viz sekce ??). Na hladině významnosti $\alpha = 0.05$ zjistíte, zda je porodní hmotnost prvorozených novorozenců nižší než porodní hmotnost druhorozených novorozenců.

Řešení příkladu 12.6

Z načteného datového souboru 10-two-samples-means-birth.txt vybereme nejprve hodnoty porodní hmotnosti (birth.W) u prvorozených novorozenců (o.sib.N == '0') a vložíme je do proměnné birth.W0. Dále z načtených dat vybereme hodnoty porodní hmotnosti (birth.W) u druhorozených novorozenců (o.sib.N == 1) a vložíme je do proměnné birth.W1. Z obou proměnných birth.W0 a birth.W1 odstraníme chybějící pozorování a zjistíme rozsahy obou náhodných výběrů a rozpětí naměřených hodnot v těchto náhodných výběrech.

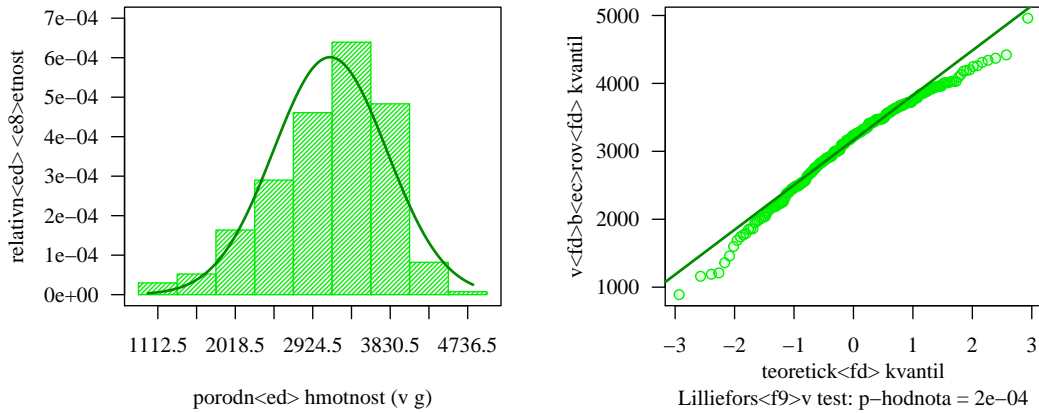
```
190 data <- read.delim('00-Data//10-two-samples-means-birth.txt')
191 birth.W0 <- data[data$o.sib.N == '0', 'birth.W']
192 birth.W1 <- data[data$o.sib.N == '1', 'birth.W']
193 birth.W0 <- na.omit(birth.W0)
194 birth.W1 <- na.omit(birth.W1)
195 n1 <- length(birth.W0) # 297
196 n2 <- length(birth.W1) # 276
197 range(birth.W0) # 890-4960
198 range(birth.W1) # 990-4970
```

Datový soubor obsahuje porodní hmotnosti 297 prvorozených novorozenců a 276 druhorozených novorozenců. Porodní hmotnost prvorozených novorozenců se pohybuje v rozmezí 890–4960 g, porodní hmotnost druhorozených novorozenců se pohybuje v rozmezí přibližně 990–4970 g.

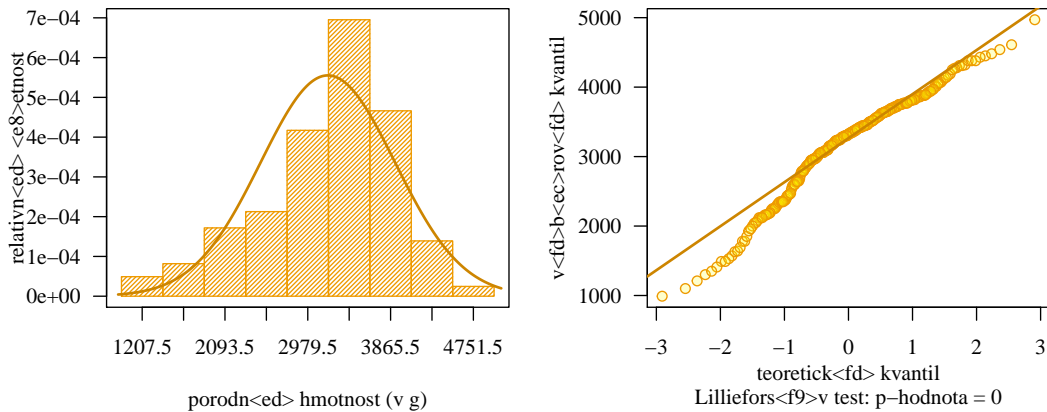
V rámci příkladu budeme chtít porovnat střední hodnotu porodní hmotnosti prvorozených novorozenců se střední hodnotou porodní hmotnosti druhorozených novorozenců. Jelikož u obou populací máme k dispozici naměřené hodnoty, chtěli bychom odpověd' na otázku ze zadání hledat primárně pomocí parametrického testu o rozdílu středních hodnot $\mu_1 - \mu_2$. K tomu je však třeba ověřit, zda oba náhodné výběry pochází z normálních rozdělení. Na hladině významnosti $\alpha = 0.05$ testujeme nejprve hypotézu H_{01} : *Náhodný výběr porodních hmotností prvorozených novorozenců pochází z normálního rozdělení.* oproti alternativní hypotéze: H_{11} : *Náhodný výběr porodních hmotností prvorozených novorozenců nepochází z normálního rozdělení.* Normalitu náhodného výběru otestujeme Lillieforsovým testem ($n_1 = 297 > 30$) a zhodnotíme graficky (viz obrázek 21). Pro účely histogramu rozdělíme náhodný výběr porodních hmotností prvorozených novorozenců do 9 ekvidistantních třídicích intervalů o šířce 453 g pomocí stanovených hranic 886, 1339, ..., 4963 g.

Protože p -hodnota Lillieforsova testu $p = 2 \times 10^{-4}$ je menší než 0.05, H_{01} zamítáme na hladině významnosti $\alpha = 0.05$. Při hodnocení grafické vizualizace je třeba si hned zkrájed' ovědomit, že rozsah náhodného výběru je velmi vysoký ($n_1 = 297$). V takovém případě již musí histogram velmi věrně kopírovat křivku hustoty normálního rozdělení. Z obrázku 21 vidíme, že tomu tak není. Naměřené hodnoty jsou oproti křivce hustoty normálního rozdělení posunuté doprava a vykarují tedy nezanedbatelné množství odlehklých pozorování na levém chvostu. QQ-diagram potom lépe ukazuje, že k porušení normality dochází na obou stranách okrajových hodnot, které se výrazně odchyľují od referenční přímky. Náhodný výběr porodních hmotností prvorozených novorozenců nepochází z normálního rozdělení.

Již v tomto okamžiku je zjevné, že parametrický test nebude možné použít z důvodu porušení normality v prvním náhodném výběru. Pro kompletní náhled na situaci otestujeme a vizualizujeme také normalitu v druhém náhodném výběru. Na hladině významnosti $\alpha = 0.05$ testujeme nyní hypotézu H_{02} : *Náhodný výběr porodních hmotností druhorozených novorozenců pochází z normálního rozdělení.* oproti alternativní hypotéze: H_{11} : *Náhodný výběr porodních hmotností druhorozených novorozenců nepochází z normálního rozdělení.* Normalitu náhodného výběru otestujeme opět pomocí Lillieforsova testu ($n_2 = 276 > 30$) a zhodnotíme graficky (viz obrázek 22). Pro účely histogramu rozdělíme náhodný výběr porodních hmotností do 9 ekvidistantních třídicích intervalů o šířce 443 g prostřednictvím stanovených hranic 986, 1429, ..., 4973 g.



Obrázek 21: Histogram (vlevo) a kvantilový diagram (vpravo) porodní hmotnosti novorozenců s žádným starším sourozencem



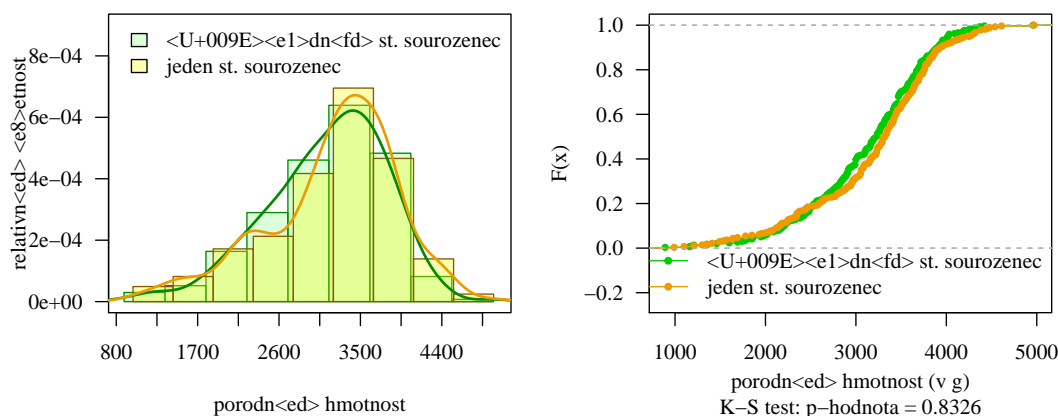
Obrázek 22: Histogram (vlevo) a kvantilový diagram (vpravo) porodní hmotnosti novorozenců s jedním starším sourozencem

Protože p -hodnota Lillieforsova testu $p = 5.7885 \times 10^{-7}$ je menší než 0.05, hypotézu H_{02} zamítáme na hladině významnosti $\alpha = 0.05$. Závažné porušení předpokladu normality je tentokrát patrné hned z pohledu na histogram. Naměřené hodnoty zdaleka nekopírují tvar křivky hustoty normálního rozdělení, vykazují vyšší koncentraci hodnot okolo střední hodnoty, než je únosné, a naopak úbytek hodnot na chvostech. Celkově jsou data vyšikmená doprava s prodlouženým levým koncem. V QQ-diagramu potom vidíme značné odchýlení bodů od referenční příčky minimálně v polovině její délky. Náhodný výběr porodních hmotností druhorozených novorozenců nepochází z normálního rozdělení.

Vzhledem k suverenímu zamítnutí hypotézy o normálním rozdělení obou náhodných výběrů není možné otázku ze zadání prověřovat pomocí parametrického testu. Použijeme tedy neparametrický test, konkrétně potom (vzhledem k velkému rozsahu obou náhodných výběrů) asymptotickou variantu Wilcoxonova neparametrického dvouvýběrového testu. Před jeho použitím pouze ověříme shodu distribučních funkcí $F_1(x)$ a $F_2(x)$ (s výjimkou posunutí o vzdálenost Δ), kde $F_1(x)$ je distribuční funkce náhodného výběru porodních hmotností prvorozených novorozenců a $F_2(x)$ je distribuční funkce porodních hmotností druhorozených novorozenců.

Oba náhodné výběry nejprve metodou centrování posuneme do nuly, čímž odfiltrujeme vliv posunutí o vzdálenost Δ . Na hladině významnosti $\alpha = 0.05$ potom testujeme hypotézu H_{03} : *Distribuční funkce porodních hmotností prvorozených novorozenců a druhorozených novorozenců jsou shodné.* oproti alternativní hypotéze: H_{13} : *Distribuční funkce porodních hmotností prvorozených novorozenců a druhorozených novorozenců nejsou shodné.* Nulová hypotéza H_{03}

odpovídá hypotéze o shodě dvou distribučních funkcí původních náhodných výběrů s výjimkou posunutí o vzdálenost Δ . Shodu distribučních funkcí otestujeme pomocí Kolmogorova-Smirnovova testu a následně vizualizujeme dvojici histogramů superponovaných příslušnými jádrovými odhady hustot a grafem obou výběrových distribučních funkcí (viz obrázek 23).



Obrázek 23: Porovnání histogramů (vlevo) a graf distribučních funkcí (vpravo) porodní hmotnosti novorozenců s žádným starším sourozencem a s jedním starším sourozencem

Protože p -hodnota K-S testu $p = 0.8326$ je větší než 0.05 , nulovou hypotézu o shodě dvou distribučních funkcí centrovaných náhodných výběrů nezamítáme na hladině významnosti $\alpha = 0.05$. Distribuční funkce porodních hmotností u prvorozených novorozenců a u druhorozených novorozenců se liší pouze posunutím.

Předpoklad umožňující použití Wilcoxonova dvouvýběrového asymptotického testu je splněn. Naším úkolem ze zadání je zjistit, zda je porodní hmotnost prvorozených novorozenců nižší než porodní hmotnost druhorozených novorozenců. Tato věta je zněním alternativní hypotézy, neboť ve větě není zmínka ani o znění nulové hypotézy, ani o testování, ani o rovnosti. Dodefinujeme tedy znění nulové hypotézy tak, aby byla doplňkem k alternativní hypotéze. Pojem *střední hodnota* figurující v hypotézách, které testujeme parametrickým testem, nahradíme *mediánem*.

1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

H_0 : Medián porodní hmotnosti prvorozených novorozenců je větší nebo roven mediánu porodní hmotnosti druhorozených novorozenců.

H_1 : Medián porodní hmotnosti prvorozených novorozenců je menší než porodní hmotnosti druhorozených novorozenců.

- **matematická formulace** nulové a alternativní hypotézy

H_0 : $\tilde{x}_1 \geq \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 \geq \tilde{x}_0$, kde $\tilde{x}_0 = 0$

H_1 : $\tilde{x}_1 < \tilde{x}_2 \rightarrow \tilde{x}_1 - \tilde{x}_2 < \tilde{x}_0$, kde $\tilde{x}_0 = 0$ (levostranná alternativa)

2. Volba hladiny významnosti

- Hladina významnosti $\alpha = 0.05$.

3. Testování kritickým oborem

- **Testovací statistika**

V prvním kroku vytvoříme společný vektor porodních hmotností tak, že za sebe poskládáme porodní hmotnosti prvorozených novorozenců (X_{1i} , $i = 1, \dots, 297$) a následně porodní hmotnosti druhorozených novorozenců (X_{2j} , $j = 1, \dots, 276$). Délka společného vektoru $n = n_1 + n_2 = 297 + 276 = 573$. Následně každému pozorování přiřadíme hodnotu indikační funkce I nabývající hodnoty 1, pokud měření pochází

z druhého náhodného výběru, nebo hodnoty 0, pokud měření pochází z prvního náhodného výběru. Nakonec stanovíme pořadí S_k , $k = 1, \dots, 573$ všech 573 hodnot seřazených od nejmenší po největší.

```

199 S <- rank(c(birth.W0, birth.W1))
200 I <- c(rep(0, n1), rep(1, n2))
201 tab <- rbind("Xi, Yj" = c(birth.W0, birth.W1), "I" = I, "Si, Sj" = S)
202 tab <- data.frame(tab)
203 names(tab) <- 1 : 573
204 #      1      2 ... 296      297      298      299      300      301 ...      570      571      572      573
205 # Xi, Yj 3470 3240 ... 3430 3160.0 3650.0 2940 2460 2350.0 ... 3160.0 3440 2950.0 3050.0
206 # I      0      0 ...      0      0      0      1      1      1 ...      1      1      1      1
207 # Si, Sj 369 273 ... 348 244.5 430.5 184 97 79.5 ... 244.5 353 186.5 214.5

```

Jednotlivým 276 hodnotám příslušícím druhému náhodnému výběru patří ve vektoru 573 seřazených hodnot pořadí 430.5, 184, 97, 79.5, ..., 244.5, 353, 186.5, 214.5. Hodnotu statistiky T_2 získáme sečtením těchto pořadí. Následně dosazením hodnot n_1 , n_2 a statistiky T_2 do vzorce 12.1 získáme hodnotu statistiky S_E a následným dosazením statistiky S_E a rozsahů n_1 a n_2 do vzorce 12.2 získáme hodnotu testovací statistiky Wilcoxonova dvouvýběrového asymptotického testu S_A .

$$T_2 = \sum_{j=1}^{n_2} S_j = 430.5 + 184 + 97 + 79.5 + \dots + 244.5 + 353 + 186.5 + 214.5 = 82199.5$$

$$\begin{aligned}
S_E &= n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - T_2 \\
&= 297 \times 276 + \frac{276(276 + 1)}{2} - 82199.5 \\
&= 81972 + \frac{76452}{2} - 82199.5 \\
&= 81972 + 38226 - 82199.5 \\
&= 37998.5
\end{aligned}$$

$$\begin{aligned}
S_A &= \frac{SE - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \\
&= \frac{37998.5 - \frac{297 \times 276}{2}}{\sqrt{\frac{297 \times 276 (297 + 276 + 1)}{12}}} \\
&= \frac{37998.5 - \frac{81972}{2}}{\sqrt{\frac{297 \times 276 \times 574}{12}}} \\
&= \frac{37998.5 - 40986}{\sqrt{3920994}} \\
&= \frac{-2987.5}{1980.15} \\
&= -1.508724
\end{aligned}$$

• Kritický obor

```

208 T2 <- sum(S[I == 1]) # 82199.5
209 SE <- n1 * n2 + n2 * (n2 + 1) / 2 - T2 # 37998.5
210 SA <- (SE - (n1 * n2) / 2) / sqrt(n1 * n2 * (n1 + n2 + 1) / 12) # -1.508724

```

$$\begin{aligned}
W &= (-\infty; u_\alpha) \\
&= (-\infty; u_{0.05}) \\
&= (-\infty; -1.644854)
\end{aligned}$$

K výpočtu kvantilu u_α použijeme funkci `qnorm()` počítající hodnotu α -kvantilu standardizovaného normálního rozdělení.

```

211 alpha <- 0.05
212 qnorm(alpha) # -1.644854

```

- **Závěr testování**

Protože realizace testovací statistiky $s_A = -1.508724$ nenáleží do kritického oboru, tj. $s_A \notin W$, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

Proti levostranné alternativě postavíme pravostranný interval spolehlivosti. Ke stanovení horní hranice 95% intervalu spolehlivosti je třeba nejprve vypočítat $n_1 \times n_2 = 297 \times 276 = 81972$ rozdílů $X_{1i} - X_{2j}$, $i = 1, \dots, 297$, $j = 1, \dots, 276$, a ty následně vzestupně seřadit. Rozdíly $X_{1i} - X_{2j}$ vypočítáme hromadně pomocí funkce `outer()`, následně je seřadíme příkazem `(sort())` a vložíme do proměnné `U`. Nakonec vypočítáme pozicovou souřadnici $C_{1-\alpha}$, pomocí které vypočítáme pozicovou souřadnici C_α . Ta potom figuruje ve výpočtu horní hranice pravostranného intervalu spolehlivosti.

$$\begin{aligned}
C_{1-\alpha} &= \frac{n_1 n_2}{2} - u_{1-\alpha} \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \\
&= \frac{297 \times 276}{2} - u_{1-0.05} \sqrt{\frac{297 \times 276 (297 + 276 + 1)}{12}} \\
&= 40986 - u_{0.95} \sqrt{\frac{297 \times 276 \times 574}{12}} \\
&= 40986 - 1.644854 \sqrt{3920994} \\
&= 40986 - 1.644854 \times 1980.15 \\
&= 37728.94 \doteq 37728
\end{aligned}$$

$$\begin{aligned}
n_1 n_2 + 1 - C_{1-\alpha} &= 297 \times 276 + 1 - 37728 \\
&= 81972 + 1 - 37728 = 44245
\end{aligned}$$

Hranice intervalu spolehlivosti potom tvoří mínus nekonečno a hodnota, která se v seřazeném vektoru rozdílů `U` nachází na 44 245. pozici.

```

213 rozdily <- outer(birth.W0, birth.W1, FUN = '-')
214 U <- sort(rozdily)

```

$$\begin{aligned}
(d, h) &= \left(-\infty; U^{(n_1 n_2 + 1 - C_{1-\alpha})}; \infty\right) \\
&= \left(-\infty; U^{(44245)}; \infty\right) \\
&= (-\infty; 10)
\end{aligned}$$

```

215 C1 <- floor(n1 * n2 / 2 - qnorm(1 - alpha) * sqrt(n1 * n2 * (n1 + n2 + 1) / 12)) # 37728
216 C2 <- n1 * n2 + 1 - C1 # 44245
217 U[44245] # 10

```

- **Závěr testování**

Protože $\tilde{x}_0 = 0$ náleží do 95% empirického levostranného intervalu spolehlivosti, tj. $\tilde{x}_0 = 0 \in IS$, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

5. Testování p -hodnotou

Příslušnou p -hodnotu vypočítáme pomocí vzorce $\Pr(S_A \geq s_A)$, kde $s_A = -1.508724$ je realizace testovací statistiky S_A . Zároveň S_A je spojitá náhodná veličina, neboť pochází ze spojitého (standardizovaného normálního) rozdělení $N(0, 1)$. Z vlastností pravděpodobnostní funkce spojitých náhodných veličin víme, že $\Pr(S_A \geq -1.508724) = 1 - \Pr(S_A < -1.508724) = 1 - \Pr(S_A \leq -1.508724)$.

- **p -hodnota**

$$p\text{-hodnota} = \Pr(S_A \geq s_A) = 1 - \Pr(S_A \leq -1.508724) = 2 \times 0.198446 = 6.100637 \times 10^{-6}$$

Hodnotu distribuční funkce standardizovaného normálního rozdělení v hodnotě $s_A = -1.508724$, tj. $\Pr(S_A \leq -1.508724)$ vypočítáme pomocí funkce `pnorm()`.

```

218 p.hodn <- pnorm(SA) # 0.06568465

```

- **Závěr testování**

Protože p -hodnota = 0.06568 je větší než $\alpha = 0.05$, H_0 nezamítáme na hladině významnosti $\alpha = 0.05$.

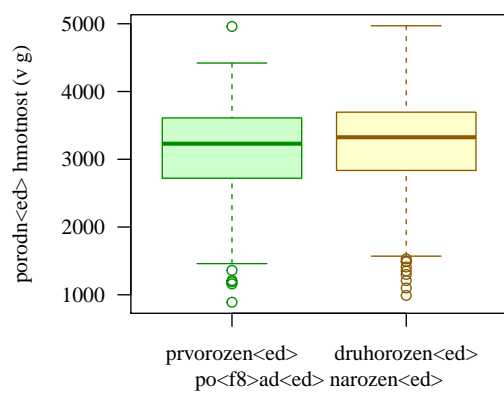
6. Interpretace výsledků

Na základě všech tří typů testování nezamítáme nulovou hypotézu na hladině významnosti $\alpha = 0.05$. Porodní hmotnost prvorozených novorozenců není statisticky významně menší než porodní hmotnost druhorozených novorozenců.

7. Grafická vizualizace výsledků testování

Nakonec vykreslíme krabicový diagram (viz obrázek 24).





Obrázek 24: Krabicový diagram porodní hmotnosti novorozenců s žádným starším sourozencem a s jedním starším sourozencem