

## 12 Dvouvýběrové neparametrické testy

### 12.1 Wilcoxonův dvouvýběrový exaktní test

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  $\alpha$  testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

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

Test nazýváme Wilcoxonův dvouvýběrový exaktní test o rozdílu mediánů  $\tilde{x}_1 - \tilde{x}_2$ . Testovací statistika má tvar

$$S_E = \sum_{j=1}^{n_2} S_j \quad (12.1)$$

kde  $n_2$  je rozsah druhého náhodného výběru a  $S_j$ ,  $j = 1, \dots, n_2$ , 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  $W$  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

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

kde  $w_{n_1, n_2}(\alpha/2)$ ,  $w_{n_1, n_2}(1 - \alpha/2)$ ,  $w_{n_1, n_2}(\alpha)$ ,  $w_{n_1, n_2}(1 - \alpha)$  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  $Y_j - X_i$ ,  $i = 1, \dots, n_1$ ,  $j = 1, \dots, n_2$ , a  $U^{(x)}$  značí  $x$ -tý seřazený rozdíl.

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

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

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 `R` a implementované funkce `pwilcox()`.

1 # ZKONTROLOVAT BARVY

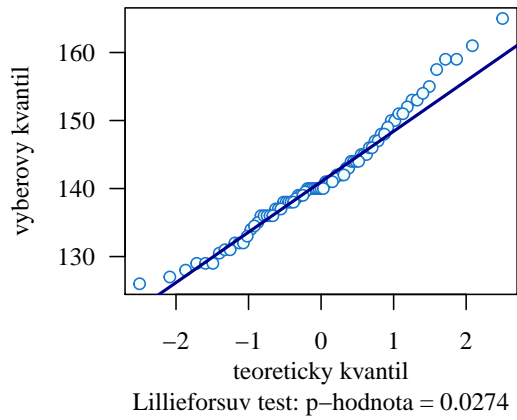
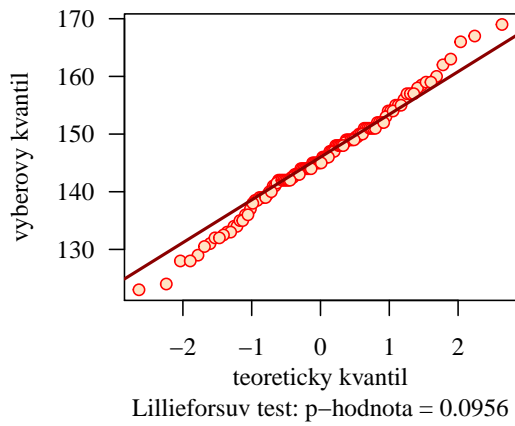
## Příklad 12.1. Wilcoxonův dvouvýběrový exaktní test

### Řešení příkladu 12.1

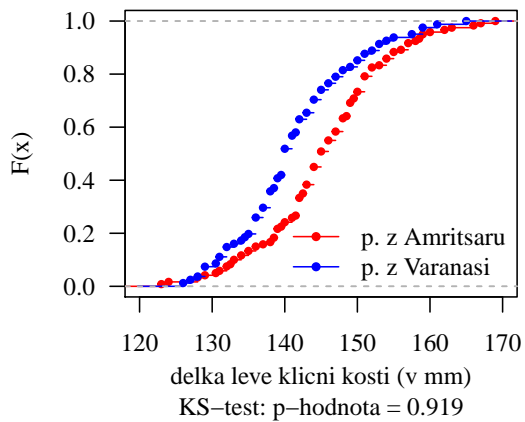
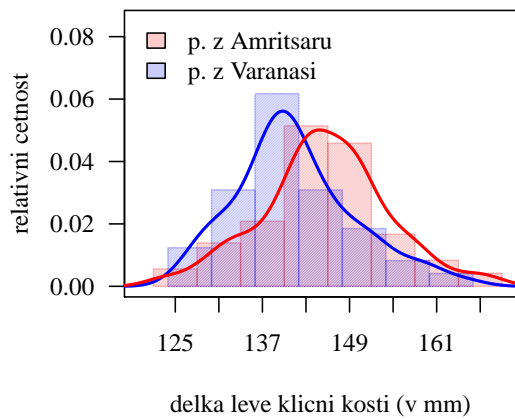
```
2 data <- read.delim('00-Data//18-more-samples-variances-clavicle.txt')
3 head(data)
4 cla.LI1 <- data[data$pop == 'ind1', 'cla.L']
5 cla.LI1 <- na.omit(cla.LI1)
6 cla.LI2 <- data[data$pop == 'ind2', 'cla.L']
7 cla.LI2 <- na.omit(cla.LI2)
8 n1 <- length(cla.LI1)
9 n2 <- length(cla.LI2)

10 par(mar = c(5, 4, 1, 2), family = 'Times')
11 qqnorm(cla.LI1, main = '', xlab = '', pch = 21, col = 'red', bg = 'bisque',
12        ylab = 'vyberovy kvantil', las = 1)
13 mtext('teoreticky kvantil', side = 1, line = 2)
14 mtext(bquote(paste('Lillieforsuv test: p-hodnota = ', .(round(nortest::lillie.test(cla.
15        LI1)$p.val, 4)))),
16        side = 1, line = 3.2)
16 qqline(cla.LI1, lwd = 2, col = 'darkred')
17
18 qqnorm(cla.LI2, main = '', xlab = '', pch = 21, col = 'dodgerblue3', bg = 'mintcream',
19        ylab = 'vyberovy kvantil', las = 1)
20 mtext('teoreticky kvantil', side = 1, line = 2)
21 mtext(bquote(paste('Lillieforsuv test: p-hodnota = ', .(round(nortest::lillie.test(cla.
22        LI2)$p.val, 4)))),
22        side = 1, line = 3.2)
23 qqline(cla.LI2, lwd = 2, col = 'darkblue')
```





Obrázek 1: Kvantilové diagramy délky klíční kosti z levé strany u mužů indické populace z Amritsaru (vlevo) a indické populace z Varanasi (vpravo)



Obrázek 2: Porovnání histogramů a graf distribučních funkcí délky klíční kosti z levé strany u mužů indické populace z Amritsaru a indické populace z Varanasi

```

24 par(mar = c(5, 4, 1, 2), family = 'Times')
25 hist(cla.LI1, prob = T, breaks = seq(122, 170, by = 6), main = '', axes = F,
26     col = rgb(0.9, 0, 0, 0.2), density = 80, xlim = c(120, 170),
27     xlab = 'delka leve klicni kosti (v mm)', ylim = c(0, 0.085),
28     ylab = 'relativni cetnost')
29 hist(cla.LI2, prob = T, breaks = seq(124, 166, by = 6), main = '', axes = F,
30     col = rgb(0, 0, 0.9, 0.2), density = 80, add = T)
31 box(bty = 'o')
32 axis(1, seq(125, 167, by = 6))
33 axis(2, las = 1)
34 lines(density(cla.LI2), lwd = 2, col = 'blue')
35 lines(density(cla.LI1), lwd = 2, col = 'red')
36 legend('topleft', fill = c(rgb(0.9, 0, 0, 0.2), rgb(0, 0, 0.9, 0.2)),
37     legend = c('p. z Amritsaru', 'p. z Varanasi'), bty = 'n')
38
39 plot(ecdf(cla.LI1), col = 'red', main = '', xlim = c(120, 170), ylab = 'F(x)',
40     xlab = '', las = 1, cex = 0.6)
41 plot(ecdf(cla.LI2), add = T, col = 'blue', cex = 0.6)
42 ks.val <- round(ks.test(cla.LI2-mean(cla.LI2), cla.LI1-(mean(cla.LI1)))$p.val, 4)
43 mtext('delka leve klicni kosti (v mm)', side = 1, line = 2.1)
44 mtext(bquote(paste('KS-test: p-hodnota = ', .(ks.val))), side = 1, line = 3.3)
45 legend('bottomright', col = c('red', 'blue'), lty = 1, pch = 20,
46     legend = c('p. z Amritsaru', 'p. z Varanasi'), bty = 'n')

```

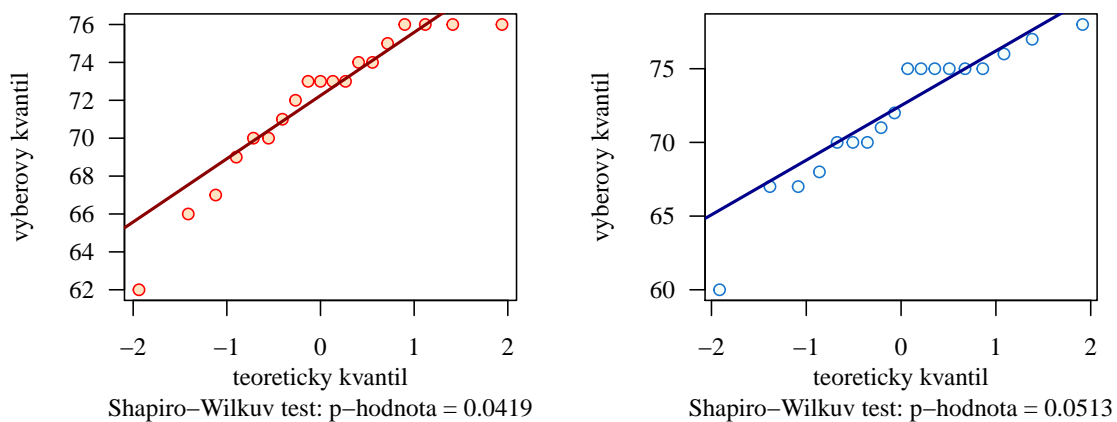
## Příklad 12.2. Wilcoxonův dvouvýběrový exaktní test

### Řešení příkladu 12.2

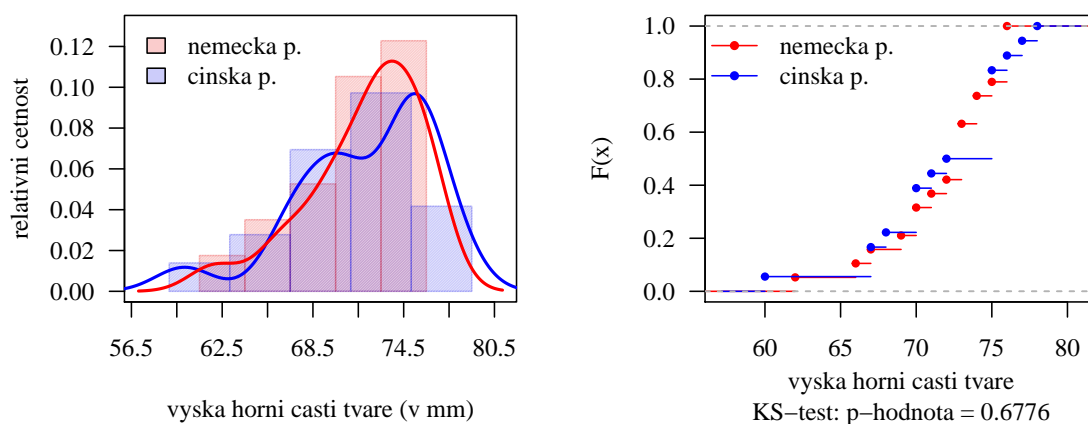
```
47 data <- read.delim('00-Data//15-anova-means-skull.txt')
48 head(data)
49 upface.HN <- data[data$pop == 'nem', 'upface.H']
50 upface.HN <- na.omit(upface.HN)
51 upface.HC <- data[data$pop == 'cin', 'upface.H']
52 upface.HC <- na.omit(upface.HC)
53 n1 <- length(upface.HN)
54 n2 <- length(upface.HC)

55 par(mar = c(5, 4, 1, 2), family = 'Times')
56 qqnorm(upface.HN, main = '', xlab = '', pch = 21, col = 'red', bg = 'bisque',
57        ylab = 'vyberovy kvantil', las = 1)
58 mtext('teoreticky kvantil', side = 1, line = 2)
59 mtext(bquote(paste('Shapiro-Wilkuv test: p-hodnota = ', .(round(shapiro.test(upface.HN)$
60        p.val, 4)))),
61        side = 1, line = 3.2)
61 qqline(upface.HN, lwd = 2, col = 'darkred')
62
63 qqnorm(upface.HC, main = '', xlab = '', pch = 21, col = 'dodgerblue3', bg = 'mintcream',
64        ylab = 'vyberovy kvantil', las = 1)
65 mtext('teoreticky kvantil', side = 1, line = 2)
66 mtext(bquote(paste('Shapiro-Wilkuv test: p-hodnota = ', .(round(shapiro.test(upface.HC)$
67        p.val, 4)))),
68        side = 1, line = 3.2)
68 qqline(upface.HC, lwd = 2, col = 'darkblue')
```





Obrázek 3: Kvantilové diagramy výšky horní části tváře mužů německé populace (vlevo) a čínské populace (vpravo)



Obrázek 4: Porovnání histogramů a graf distribučních funkcí výšky horní části tváře mužů německé a čínské populace

```

69 par(mar = c(5, 4, 1, 2), family = 'Times')
70 hist(upface.HN, prob = T, breaks = seq(61, 76, by = 3), main = '', axes = F,
71      col = rgb(0.9, 0, 0, 0.2), density = 80, ylim = c(0, 0.13), xlim = c(57, 81),
72      xlab = 'vyska horni casti tvare (v mm)',
73      ylab = 'relativni cetnost')
74
75 hist(upface.HC, prob = T, breaks = seq(59, 79, by = 4), main = '', axes = F,
76      col = rgb(0, 0, 0.9, 0.2), density = 80, add = T)
77 box(bty = 'o')
78 axis(1, seq(53.5, 80.5, by = 3))
79 axis(2, las = 1)
80 lines(density(upface.HC), lwd = 2, col = 'blue')
81 lines(density(upface.HN), lwd = 2, col = 'red')
82 legend('topleft', fill = c(rgb(0.9, 0, 0, 0.2), rgb(0, 0, 0.9, 0.2)),
83      legend = c('nemecka p.', 'cinska p.'), bty = 'n')
84
85 plot(ecdf(upface.HN), col = 'red', main = '', xlim = c(57, 81), ylab = 'F(x)',
86      xlab = '', las = 1, cex = 0.6)
87 plot(ecdf(upface.HC), add = T, col = 'blue', cex = 0.6)
88 ks.val <- round(ks.test(upface.HC - mean(upface.HC), upface.HN - (mean(upface.HN))))$p.
      val, 4)
89 mtext('vyska horni casti tvare', side = 1, line = 2.1)
90 mtext(bquote(paste('KS-test: p-hodnota = ', .(ks.val))), side = 1, line = 3.3)
91 legend('topleft', col = c('red', 'blue'), lty = 1, pch = 20,
92      legend = c('nemecka p.', 'cinska p.'), bty = 'n')

```



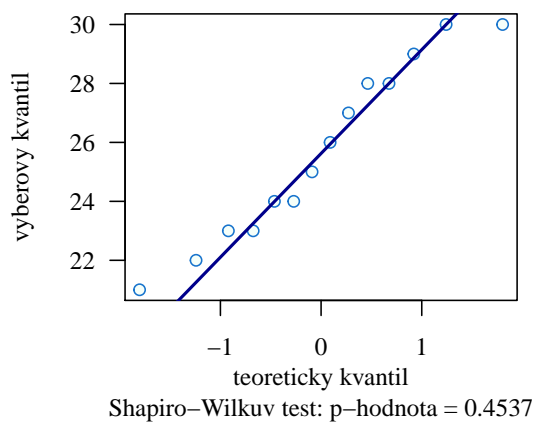
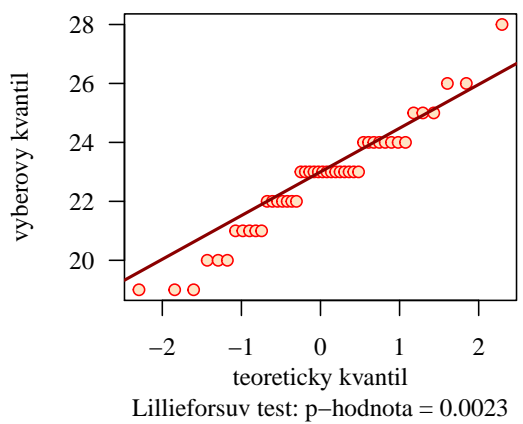
## Příklad 12.3. Wilcoxonův dvouvýběrový exaktní test

### Řešení příkladu 12.3

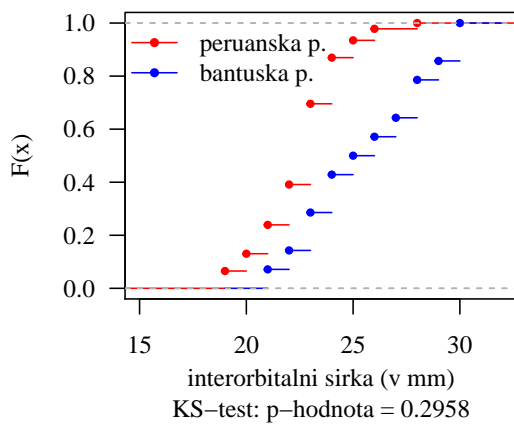
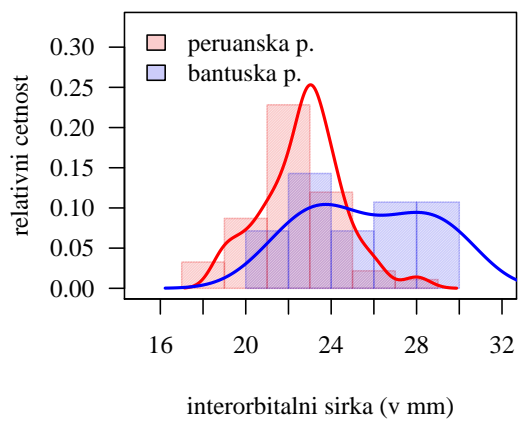
```
93 data <- read.delim('00-Data//19-more-samples-correlations-skull.txt')
94 head(data)
95 intorb.BP <- data[data$pop == 'per', 'intorb.B']
96 intorb.BP <- na.omit(intorb.BP)
97 intorb.BB <- data[data$pop == 'ban', 'intorb.B']
98 intorb.BB <- na.omit(intorb.BB)
99 n1 <- length(intorb.BP) # 46
100 n2 <- length(intorb.BB) # 14

101 par(mar = c(5, 4, 1, 2), family = 'Times')
102 qqnorm(intorb.BP, main = '', xlab = '', pch = 21, col = 'red', bg = 'bisque',
103        ylab = 'vyberovy kvantil', las = 1)
104 mtext('teoreticky kvantil', side = 1, line = 2)
105 mtext(bquote(paste('Lillieforsuv test: p-hodnota = ', .(round(nortest::lillie.test(
106        intorb.BP)$p.val, 4)))),
107        side = 1, line = 3.2)
108 qqline(intorb.BP, lwd = 2, col = 'darkred')
109 qqnorm(intorb.BB, main = '', xlab = '', pch = 21, col = 'dodgerblue3', bg = 'mintcream',
110        ylab = 'vyberovy kvantil', las = 1)
111 mtext('teoreticky kvantil', side = 1, line = 2)
112 mtext(bquote(paste('Shapiro-Wilkuv test: p-hodnota = ', .(round(shapiro.test(intorb.BB)$
113        p.val, 4)))),
114        side = 1, line = 3.2)
114 qqline(intorb.BB, lwd = 2, col = 'darkblue')
```





Obrázek 5: Porovnání histogramů a graf distribučních funkcí interorbitální šířky mužů peruánské a bantuské populace



Obrázek 6: Porovnání histogramů a graf distribučních funkcí interorbitální šířky mužů peruánské a bantuské populace

```

115 par(mar = c(5, 4, 1, 2), family = 'Times')
116 hist(intorb.BP, prob = T, breaks = seq(17, 29, by = 2), main = '', axes = F,
117       col = rgb(0.9, 0, 0, 0.2), density = 80, xlim = c(15, 32), ylim = c(0, 0.33),
118       xlab = 'interorbitalni sirka (v mm)',
119       ylab = 'relativni cetnost')
120 hist(intorb.BB, prob = T, breaks = seq(20, 30, by = 2), main = '', axes = F,
121       col = rgb(0, 0, 0.9, 0.2), density = 80, add = T)
122 box(bty = 'o')
123 axis(1, seq(12, 36, by = 2))
124 axis(2, las = 1)
125 lines(density(intorb.BP), lwd = 2, col = 'red')
126 lines(density(intorb.BB), lwd = 2, col = 'blue')
127 legend('topleft', fill = c(rgb(0.9, 0, 0, 0.2), rgb(0, 0, 0.9, 0.2)),
128       legend = c('peruanska p.', 'bantuska p.'), bty = 'n')
129
130 plot(ecdf(intorb.BP), col = 'red', main = '', xlim = c(15, 32), ylab = 'F(x)',
131       xlab = '', las = 1, cex = 0.6)
132 plot(ecdf(intorb.BB), add = T, col = 'blue', cex = 0.6)
133 ks.val <- round(ks.test(intorb.BP - mean(intorb.BP), intorb.BB - (mean(intorb.BB)))$p.
134       val, 4)
135 mtext('interorbitalni sirka (v mm)', side = 1, line = 2.1)
136 mtext(bquote(paste('KS-test: p-hodnota = ', .(ks.val))), side = 1, line = 3.3)
137 legend('topleft', col = c('red', 'blue'), lty = 1, pch = 20,
138       legend = c('peruanska p.', 'bantuska p.'), bty = 'n')

```

## 12.2 Wilcoxonův dvouvýběrový asymptotický test

Pro náhodné výběry o rozsazích  $n_1 > 30$  a  $n_2 > 30$  máme možnost použít k otestování nulové hypotézy asymptotickou variantu testu. Tuto variantu nazýváme Wilcoxonův dvouvýběrový asymptotický test o rozdílu mediánů  $\tilde{x}_1 - \tilde{x}_2$ . Testovací statistika asymptotického testu má tvar

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

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.

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

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

Interval spolehlivosti má podle zvolené alternativní hypotézy jeden z následujících tvarů

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

kde  $C_{\alpha/2} = \frac{n_1 n_2}{2} - u_{\alpha/2} \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$ ,  $C_{\alpha} = \frac{n_1 n_2}{2} - u_{\alpha} \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$ ,  $U^{(1)} \leq \dots \leq U^{(n_1 n_2)}$  značí vzestupně seřazené rozdíly  $Y_j - X_i$ ,  $i = 1, \dots, n_1$ ,  $j = 1, \dots, n_2$ , a  $U^{(x)}$  značí  $x$ -tý seřazený rozdíl.

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

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

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  $\mathbb{R}$  a implementované funkce `pnorm()`.

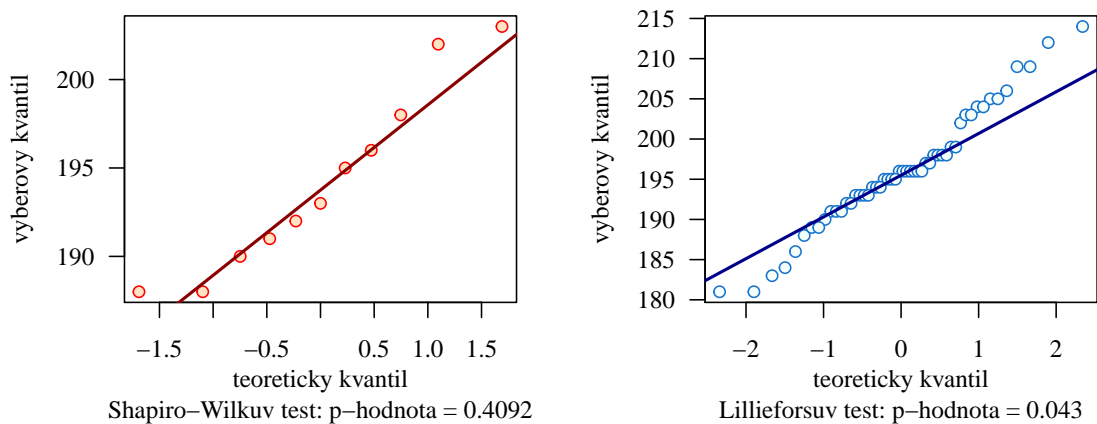
## Příklad 12.4. Wilcoxonův dvouvýběrový asymptotický test

### Řešení příkladu 12.4

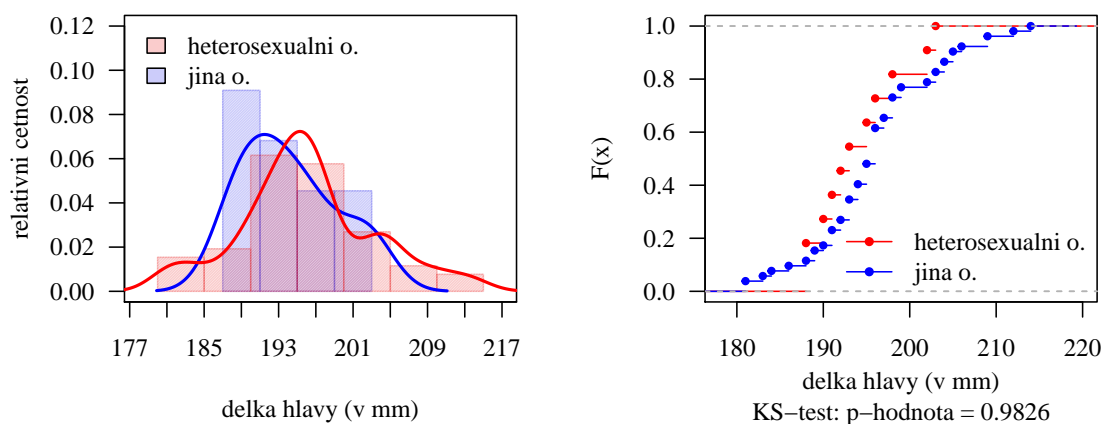
```
139 data <- read.delim('00-Data//16-anova-head.txt')
140 head(data)
141 head.LMSa <- data[data$sex == 'm' & data$sexor == 'sa', 'head.L']
142 head.LMSa <- na.omit(head.LMSa)
143 head.LMOp <- data[data$sex == 'm' & data$sexor == 'op', 'head.L']
144 head.LMOp <- na.omit(head.LMOp)
145 n1 <- length(head.LMSa)
146 n2 <- length(head.LMOp)

147 par(mar = c(5, 4, 1, 2), family = 'Times')
148 qqnorm(head.LMSa, main = '', xlab = '', pch = 21, col = 'red', bg = 'bisque',
149       ylab = 'vyberovy kvantil', las = 1)
150 mtext('teoreticky kvantil', side = 1, line = 2)
151 mtext(bquote(paste('Shapiro-Wilkuv test: p-hodnota = ', .(round(shapiro.test(head.LMSa)$
152   p.val, 4)))),
153       side = 1, line = 3.2)
154 qqline(head.LMSa, lwd = 2, col = 'darkred')
155 qqnorm(head.LMOp, main = '', xlab = '', pch = 21, col = 'dodgerblue3', bg = 'mintcream',
156       ylab = 'vyberovy kvantil', las = 1)
157 mtext('teoreticky kvantil', side = 1, line = 2)
158 mtext(bquote(paste('Lillieforsuv test: p-hodnota = ', .(round(nortest::lillie.test(head.
159   LMOp)$p.val, 4)))),
160       side = 1, line = 3.2)
161 qqline(head.LMOp, lwd = 2, col = 'darkblue')
```





Obrázek 7: Kvantilové diagramy délky hlavy mužů orientovaných výhradně heterosexuálně (vlevo) a mužů orientovaných jinak než výhradně heterosexuálně (vpravo)



Obrázek 8: Porovnání histogramů a graf distribučních funkcí délky hlavy mužů orientovaných výhradně heterosexuálně a mužů orientovaných jinak než výhradně heterosexuálně

```

161 par(mar = c(5, 4, 1, 2), family = 'Times')
162 hist(head.LMSa, prob = T, breaks = seq(187, 203, by = 4), main = '', axes = F,
163       col = rgb(0, 0, 0.9, 0.2), density = 80, xlim = c(178, 217),
164       xlab = 'delka hlavy (v mm)', ylim = c(0, 0.12),
165       ylab = 'relativni cetnost')
166
167 hist(head.LMOp, prob = T, breaks = seq(180, 215, by = 5), main = '', axes = F,
168       col = rgb(0.9, 0, 0, 0.2), density = 80, add = T)
169 box(bty = 'o')
170 axis(1, seq(169, 221, by = 4))
171 axis(2, las = 1)
172 lines(density(head.LMSa), lwd = 2, col = 'blue')
173 lines(density(head.LMOp), lwd = 2, col = 'red')
174 legend('topleft', fill = c(rgb(0.9, 0, 0, 0.2), rgb(0, 0, 0.9, 0.2)),
175       legend = c('heterosexualni o.', 'jina o.'), bty = 'n')
176
177 plot(ecdf(head.LMSa), col = 'red', main = '', xlim = c(178, 220),
178      ylab = 'F(x)', xlab = '', las = 1, cex = 0.6)
179 plot(ecdf(head.LMOp), add = T, col = 'blue', cex = 0.6)
180 ks.val <- round(ks.test(head.LMOp - mean(head.LMOp), head.LMSa - (mean(head.LMSa)))$p.
181                val, 4)
181 mtext('delka hlavy (v mm)', side = 1, line = 2.1)
182 mtext(bquote(paste('KS-test: p-hodnota = ', .(ks.val))), side = 1, line = 3.3)
183 legend('bottomright', col = c('red', 'blue'), lty = 1, pch = 20,
184       legend = c('heterosexualni o.', 'jina o.'), bty = 'n')

```

## Příklad 12.5. Wilcoxonův dvouvýběrový asymptotický test

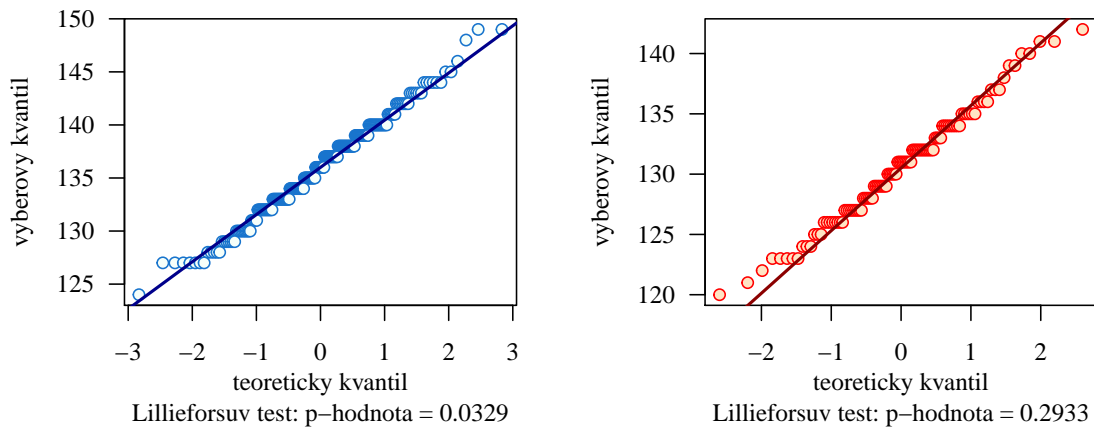
### Řešení příkladu 12.5

```
185 data <- read.delim('00-Data//05-one-sample-correlation-skull-mf.txt')
186 head(data)
187 skull.pHM <- data[data$sex == 'm', 'skull.pH']
188 skull.pHM <- na.omit(skull.pHM)
189 skull.pHF <- data[data$sex == 'f', 'skull.pH']
190 skull.pHF <- na.omit(skull.pHF)
191 n1 <- length(skull.pHM) # 216
192 n2 <- length(skull.pHF) # 107

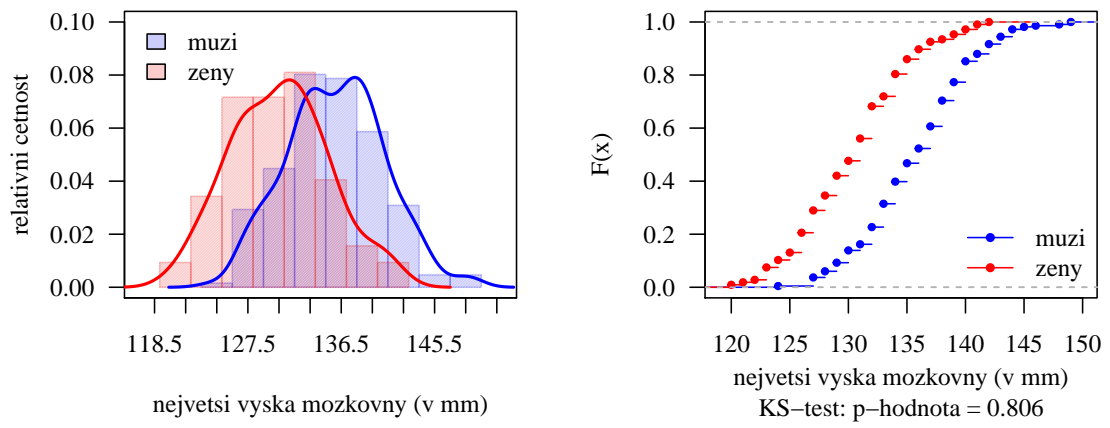
193 par(mar = c(5, 4, 1, 2), family = 'Times')
194 qqnorm(skull.pHM, main = '', xlab = '', pch = 21, col = 'dodgerblue3', bg = 'mintcream',
195        ylab = 'vyberovy kvantil', las = 1)
196 mtext('teoreticky kvantil', side = 1, line = 2)
197 mtext(bquote(paste('Lillieforsuv test: p-hodnota = ', .(round(nortest::lillie.test(skull
198        .pHM)$p.val, 4)))),
199        side = 1, line = 3.2)
200 qqline(skull.pHM, lwd = 2, col = 'darkblue')
201 qqnorm(skull.pHF, main = '', xlab = '', pch = 21, col = 'red', bg = 'bisque',
202        ylab = 'vyberovy kvantil', las = 1)
203 mtext('teoreticky kvantil', side = 1, line = 2)
204 mtext(bquote(paste('Lillieforsuv test: p-hodnota = ', .(round(nortest::lillie.test(skull
205        .pHF)$p.val, 4)))),
206        side = 1, line = 3.2)
207 qqline(skull.pHF, lwd = 2, col = 'darkred')
```







Obrázek 9: Kvantilové diagramy největší výšky mozkovny mužů (vlevo) a žen (vpravo)



Obrázek 10: Porovnání histogramů a graf distribučních funkcí největší výšky mozkovny mužů a žen

```

207 par(mar = c(5, 4, 1, 2), family = 'Times')
208 hist(skull.pHM, prob = T, breaks = seq(123, 150, by = 3), main = '', axes = F,
209       col = rgb(0, 0, 0.9, 0.2), xlim = c(117, 152), density = 80, ylim = c(0, 0.10),
210       xlab = 'nejvetsi vyska mozkovny (v mm)',
211       ylab = 'relativni cetnost')
212
213 hist(skull.pHF, prob = T, breaks = seq(119, 143, by = 3), main = '', axes = F,
214       col = rgb(0.9, 0, 0, 0.2), density = 80,
215       xlab = 'vyska horni casti tvare (v mm)',
216       ylab = 'relativni cetnost', add = T)
217 box(bty = 'o')
218 axis(1, seq(115.5, 157.5, by = 3))
219 axis(2, las = 1)
220 lines(density(skull.pHM), lwd = 2, col = 'blue')
221 lines(density(skull.pHF), lwd = 2, col = 'red')
222 legend('topleft', fill = c(rgb(0, 0, 0.9, 0.2), rgb(0.9, 0, 0, 0.2)),
223       legend = c('muzi', 'zeny'), bty = 'n')
224
225 plot(ecdf(skull.pHM), col = 'blue', main = '', xlim = c(119, 150), ylab = 'F(x)', xlab =
226       '',
227       cex = 0.6, las = 1)
228 plot(ecdf(skull.pHF), add = T, col = 'red', cex = 0.6)
229 ks.val <- round(ks.test(skull.pHM - mean(skull.pHM), skull.pHF - (mean(skull.pHF))))$p.
230       val, 4)
231 mtext('nejvetsi vyska mozkovny (v mm)', side = 1, line = 2.1)
232 mtext(bquote(paste('KS-test: p-hodnota = ', .(ks.val))), side = 1, line = 3.3)
233 legend('bottomright', col = c('blue', 'red'), lty = 1, pch = 20,
234       legend = c('muzi', 'zeny'), bty = 'n')

```

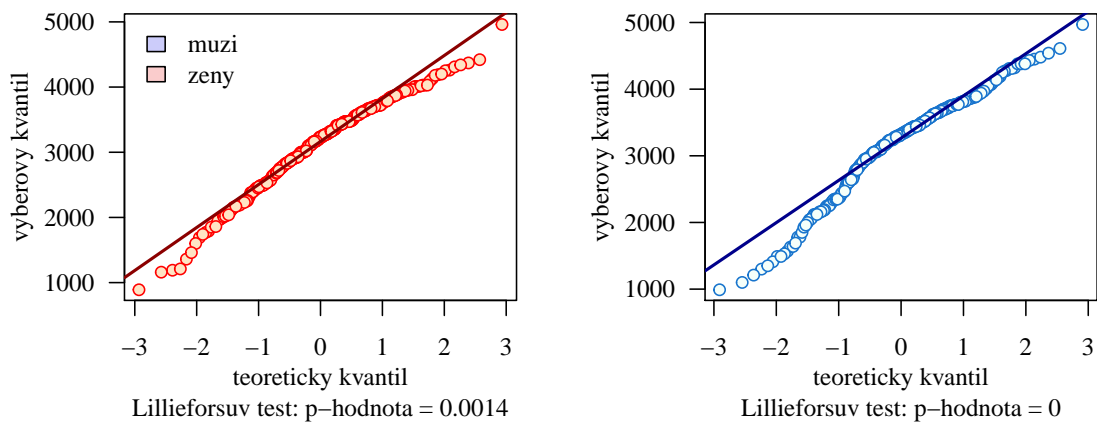
## Příklad 12.6. Wilcoxonův dvouvýběrový asymptotický test

### Řešení příkladu 12.6

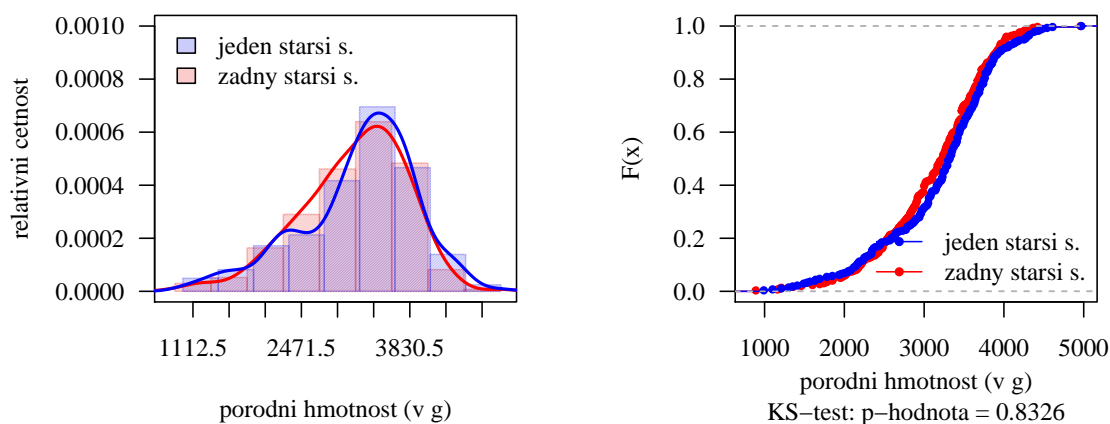
```
233 data <- read.delim('00-Data//10-two-samples-means-birth.txt')
234 head(data)
235 birth.W0 <- data[data$o.sib.N == '0', 'birth.W']
236 birth.W0 <- na.omit(birth.W0)
237 birth.W1 <- data[data$o.sib.N == '1', 'birth.W']
238 birth.W1 <- na.omit(birth.W1)
239 n1 <- length(birth.W0)
240 n2 <- length(birth.W1)

241 par(mar = c(5, 4, 1, 2), family = 'Times')
242 qqnorm(birth.W0, main = '', xlab = '', pch = 21, col = 'red', bg = 'bisque',
243        ylab = 'vyberovy kvantil', las = 1)
244 mtext('teoreticky kvantil', side = 1, line = 2)
245 mtext(bquote(paste('Lillieforsuv test: p-hodnota = ', .(round(nortest::lillie.test(birth
246        .W0)$p.val, 4)))),
247        side = 1, line = 3.2)
248 qqline(birth.W0, lwd = 2, col = 'darkred')
249 legend('topleft', fill = c(rgb(0, 0, 0.9, 0.2), rgb(0.9, 0, 0, 0.2)),
250        legend = c('muzi', 'zeny'), bty = 'n')
251 qqnorm(birth.W1, main = '', xlab = '', pch = 21, col = 'dodgerblue3', bg = 'mintcream',
252        ylab = 'vyberovy kvantil', las = 1)
253 mtext('teoreticky kvantil', side = 1, line = 2)
254 mtext(bquote(paste('Lillieforsuv test: p-hodnota = ', .(round(nortest::lillie.test(birth
255        .W1)$p.val, 4)))),
256        side = 1, line = 3.2)
257 qqline(birth.W1, lwd = 2, col = 'darkblue')
```





Obrázek 11: Kvantilové diagramy porodní hmotnosti novorozenců s žádným starším sourozencem (vlevo) a s jedním starším sourozencem (vpravo)



Obrázek 12: Porovnání histogramů a graf distribučních funkcí porodní hmotnosti novorozenců s žádným starším sourozencem a s jedním starším sourozencem

```

256 par(mar = c(5, 5, 1, 2), family = 'Times')
257 hist(birth.W0, prob = T, breaks = seq(886, 4963, by = 453), main = '', axes = F,
258     col = rgb(0.9, 0, 0, 0.2), density = 80, xlim = c(800, 5000),
259     xlab = 'porodni hmotnost (v g)', ylim = c(0, 0.001),
260     ylab = '')
261 hist(birth.W1, prob = T, breaks = seq(986, 4973, by = 443), main = '', axes = F,
262     col = rgb(0, 0, 0.9, 0.2), density = 80, add = T)
263 box(bty = 'o')
264 axis(1, seq(1112.5, 4736.5, by = 453))
265 axis(2, las = 1, at = seq(0, 0.0010, by = 0.0002), labels = c('0.0000', '0.0002', '
0.0004', '0.0006', '0.0008', '0.0010'))
266 lines(density(birth.W0), lwd = 2, col = 'red')
267 lines(density(birth.W1), lwd = 2, col = 'blue')
268 mtext('relativni cetnost', side = 2, line = 4)
269 legend('topleft', fill = c(rgb(0, 0, 0.9, 0.2), rgb(0.9, 0, 0, 0.2)),
270     legend = c('jeden starsi s.', 'zadny starsi s.'), bty = 'n')
271
272 plot(ecdf(birth.W0), col = 'red', main = '', xlim = c(800, 5000),
273     ylab = 'F(x)', xlab = '', cex = 0.6, las = 1)
274 plot(ecdf(birth.W1), add = T, col = 'blue', cex = 0.6)
275 ks.val <- round(ks.test(birth.W1 - mean(birth.W1), birth.W0 - (mean(birth.W0)))$p.val,
276     4)
277 mtext('porodni hmotnost (v g)', side = 1, line = 2.1)
278 mtext(bquote(paste('KS-test: p-hodnota = ', .(ks.val))), side = 1, line = 3.3)
279 legend('bottomright', col = c('blue', 'red'), lty = 1, pch = 20,
280     legend = c('jeden starsi s.', 'zadny starsi s.'), bty = 'n')

```

### 12.3 Mann-Whitneyův dvouvýběrový test

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  $\alpha$  testujeme jednu z následujících tří hypotéz proti příslušné alternativní hypotéze.

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

Test nazýváme Mann-Whitneyův dvouvýběrový test o rozdílu mediánů  $\tilde{x}_1 - \tilde{x}_2$ . Testovací statistika má tvar

$$U = S_E + \frac{n_2(n_2 + 1)}{2}, \quad (12.3)$$

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

$$U = S_E + \frac{n_2(n_2 + 1)}{2} \stackrel{H_0}{\sim} N(0, 1).$$

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

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

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

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

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

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

kde  $U$  je náhodná veličina,  $u$  je realizace testovací statistiky  $U$  (viz vzorec 12.3), tedy konkrétní číslo, a  $\Pr(U \leq u)$  je distribuční funkce tabelovaného rozdělení pro Wilcoxonův dvouvýběrový exaktní test, jejíž hodnotu získáme pomocí `R` a implementované funkce `pwilcox()`.

## Příklad 12.7. Mann-Whitneyův dvouvýběrový test

### Řešení příkladu 12.7



```
280 data <- read.delim('00-Data//18-more-samples-variances-clavicle.txt')
281 head(data)
282 cla.LI1 <- data[data$pop == 'ind1', 'cla.L']
283 cla.LI1 <- na.omit(cla.LI1)
284 cla.LI2 <- data[data$pop == 'ind2', 'cla.L']
285 cla.LI2 <- na.omit(cla.LI2)
286
287 wilcox.test(cla.LI1, cla.LI2, alt = 't')$p.val # Z
```

## Příklad 12.8. Mann-Whitneyův dvouvýběrový test

### Řešení příkladu 12.8



```
288 data <- read.delim('00-Data//15-anova-means-skull.txt')
289 head(data)
290 upface.HN <- data[data$pop == 'nem', 'upface.H']
291 upface.HN <- na.omit(upface.HN)
292 upface.HC <- data[data$pop == 'cin', 'upface.H']
293 upface.HC <- na.omit(upface.HC)
294
295 wilcox.test(upface.HN, upface.HC, alt = 'g')$p.val # N
```



## Příklad 12.9. Mann-Whitneyův dvouvýběrový test

### Řešení příkladu 12.9



```
296 data <- read.delim('00-Data//19-more-samples-correlations-skull.txt')
297 head(data)
298 intorb.BP <- data[data$pop == 'per', 'intorb.B']
299 intorb.BP <- na.omit(intorb.BP)
300 intorb.BB <- data[data$pop == 'ban', 'intorb.B']
301 intorb.BB <- na.omit(intorb.BB)
302
303 wilcox.test(intorb.BP, intorb.BB, alt = 'l')$p.val # Z
```