

## 7 Jednovýběrové parametrické testy

### 7.1 Test o rozptylu $\sigma^2$

Nechť  $X_1, \dots, X_n$  je náhodný výběr z  $N(\mu, \sigma^2)$ , kde  $\mu$  neznáme a  $\sigma_0^2$  je konstanta. Na hladině významnosti  $\alpha = 0.05$  testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

$$\begin{array}{lll} H_{01} : \sigma^2 = \sigma_0^2 & \text{oproti} & H_{11} : \sigma^2 \neq \sigma_0^2 \quad (\text{oboustranná alt.}) \\ H_{02} : \sigma^2 \leq \sigma_0^2 & \text{oproti} & H_{12} : \sigma^2 > \sigma_0^2 \quad (\text{pravostranná alt.}) \\ H_{03} : \sigma^2 \geq \sigma_0^2 & \text{oproti} & H_{13} : \sigma^2 < \sigma_0^2 \quad (\text{levostranná alt.}) \end{array}$$

Test nazýváme jednovýběrový  $F$ -test. Testovací statistika má tvar

$$F_W = \frac{(n-1)S^2}{\sigma_0^2}, \quad (7.1)$$

kde  $S$  je výběrová směrodatná odchylka,  $n$  je rozsah náhodného výběru a  $\sigma_0$  je konstanta z nulové hypotézy. Za platnosti nulové hypotézy pochází statistika  $F_W$  z  $\chi^2$  rozdělení o  $n-1$  stupních volnosti, tj.

$$F_W = \frac{(n-1)S^2}{\sigma_0^2} \stackrel{H_0}{\sim} \chi_{n-1}^2.$$

Kritický obor podle zvolené alternativní hypotézy má tvar

$$\begin{array}{ll} H_{11} : \sigma^2 \neq \sigma_0^2 & W = (0; \chi_{n-1}^2(\alpha/2)) \cup \langle \chi_{n-1}^2(1-\alpha/2); \infty \rangle \\ H_{12} : \sigma^2 > \sigma_0^2 & W = \langle \chi_{n-1}^2(1-\alpha); \infty \rangle \\ H_{13} : \sigma^2 < \sigma_0^2 & W = (0; \chi_{n-1}^2(\alpha)) \end{array}$$

kde  $\chi_{n-1}^2(\alpha/2)$ ,  $\chi_{n-1}^2(1-\alpha/2)$ ,  $\chi_{n-1}^2(\alpha)$ ,  $\chi_{n-1}^2(1-\alpha)$  jsou kvantily  $\chi^2$  rozdělení o  $n-1$  stupních volnosti, jejichž hodnoty získáme pomocí softwaru  $\mathbb{R}$  a implementované funkce `qchisq()`.

Interval spolehlivosti má podle zvolené alternativní hypotézy jeden z následujících tvarů

$$\begin{array}{ll} H_{11} : \sigma^2 \neq \sigma_0^2 & (d, h) = \left( \frac{(n-1)s^2}{\chi_{n-1}^2(1-\alpha/2)}, \frac{(n-1)s^2}{\chi_{n-1}^2(\alpha/2)} \right) \\ H_{12} : \sigma^2 > \sigma_0^2 & (d, \infty) = \left( \frac{(n-1)s^2}{\chi_{n-1}^2(1-\alpha)}, \infty \right) \\ H_{13} : \sigma^2 < \sigma_0^2 & (0, h) = \left( 0, \frac{(n-1)s^2}{\chi_{n-1}^2(\alpha)} \right) \end{array}$$

*Poznámka:* Protože parametr  $\sigma^2$  je vždy větší než 0, můžeme dolní hranici pravostranného intervalu spolehlivosti zvolit 0 namísto mínus nekonečna.

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

$$\begin{array}{ll} H_{11} : \sigma^2 \neq \sigma_0^2 & p\text{-hodnota} = 2 \min\{\Pr(F_W \leq f_W), \Pr(F_W > f_W)\} \\ H_{12} : \sigma^2 > \sigma_0^2 & p\text{-hodnota} = \Pr(F_W > f_W) = 1 - \Pr(F_W \leq f_W) \\ H_{13} : \sigma^2 < \sigma_0^2 & p\text{-hodnota} = \Pr(F_W \leq f_W) \end{array}$$

kde  $F_W$  je náhodná veličina,  $f_W$  je realizace testovací statistiky  $F_W$  (viz vzorec 7.1), tedy konkrétní číslo, a  $\Pr(F_W \leq f_W)$  je distribuční funkce  $\chi^2$  rozdělení o  $n-1$  stupních volnosti, jejíž hodnotu získáme pomocí  $\mathbb{R}$  a implementované funkce `pchisq()`.

*Poznámka:* Test o rozptylu můžeme využít také jako test o směrodatné odchylce.

### Příklad 7.1. Test o rozptylu $\sigma^2$ (oboustranná alternativa)

Mějme datový soubor 01-one-sample-mean-skull-mf.txt a proměnnou skull.L popisující největší délku mozkovny v mm jedinců starověké egyptské populace (viz sekce ??). Dále máme k dispozici údaje o největší délce mozkovny mužů novověké egyptské populace ( $m_m = 177.468$  mm,  $s_m = 7.526$  mm,  $n_m = 88$ ). Na hladině významnosti  $\alpha = 0.01$  zjistěte, se rozptyly největší délky mozkovny mužů starověké a novověké egyptské populace liší.

#### Řešení příkladu 7.1

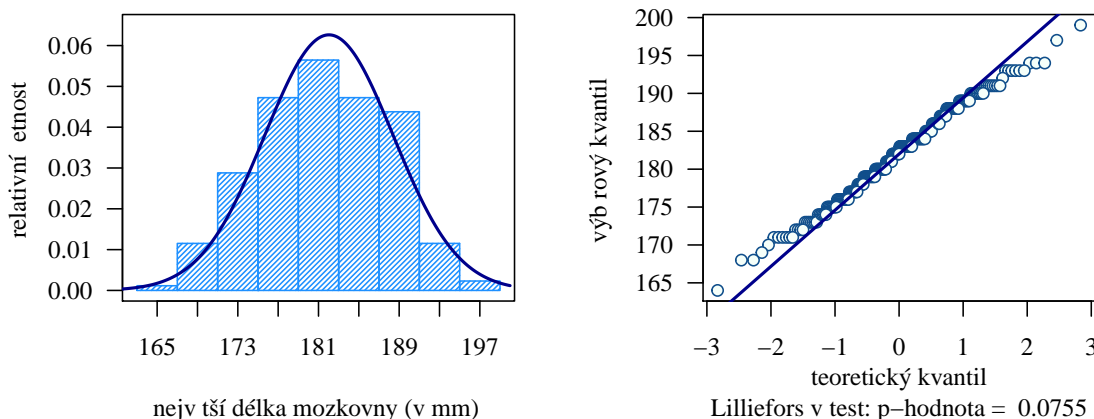
Pomocí příkazu read.delim() načteme datový soubor a pomocí operátoru [] vybereme z datové tabulky naměřené hodnoty největších délek mozkovny (skull.L) mužů (sex == 'm'). Nakonec příkazem na.omit() odstraníme z vektoru naměřených hodnot chybějící údaje a příkazem length() zjistíme rozsah náhodného výběru.

```
1 data <- read.delim('00-Data//01-one-sample-mean-skull-mf.txt')
2 skull.LM <- data[data$sex == 'm', 'skull.L']
3 skull.LM <- as.numeric(na.omit(skull.LM))
4 n <- length(skull.LM) # 217
```

Datový soubor obsahuje údaje o největší délce mozkovny 217 mužů starověké egyptské populace.

Naším úkolem ze zadání je porovnat rozptyly dvou egyptských populací, přičemž u mužů ze starověké egyptské populace máme k dispozici naměřené hodnoty. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení, tj. zda náhodná veličina  $X$  popisující největší délku mozkovny mužů starověké egyptské populace pochází z normálního rozdělení, tj.  $X \sim N(\mu, \sigma^2)$ . Druhá, novověká egyptská populace je reprezentována pouze hodnotou aritmetického průměru ( $m_m = 177.568$  mm) a směrodatnou odchylkou ( $s_m = 7.526$  mm). O jejím rozdělení přesnější informace nemáme. Řešení příkladu tedy vede na situaci, kdy rozptyl jednoho náhodného výběru porovnáváme s konkrétním číslem, tedy na jednovýběrový test o rozptylu  $\sigma^2$ . Jediným předpokladem k použití tohoto testu je normální rozdělení náhodného výběru naměřených délek mozkovny mužů starověké egyptské populace. Před použitím testu tedy tento předpoklad ověříme.

Hladinu významnosti  $\alpha$  pro test normality stanovíme standartně, tj.  $\alpha = 0.05$ . Jelikož rozsah náhodného výběru je větší než 30, ověříme předpoklad normality Lillieforsovým testem. Grafické ověření provedeme na základě QQ-diagramu a histogramu superponovaného křivkou normálního rozdělení, jejíž parametry odhadneme pomocí výběrového průměru a výběrového rozptylu (viz obrázek 1). Datový soubor rozdělíme do devíti ekvidistatních intervalů s šířkou 4 mm prostřednictvím stanovených hranic 163, 167, ..., 199.



Obrázek 1: Histogram a QQ-diagram největší délky mozkovny mužů starověké egyptské populace

Protože  $p$ -hodnota = 0.0755 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z histogramu na obrázku 1 vidíme, že naměřené hodnoty celkem věrně kopírují tvar teoretické křivky hustoty normálního rozdělení. Na QQ-diagramu je zřejmé odchýlení bodů od referenční křivky na obou chvostech. Podle Lillieforsova testu nemá však toto odchýlení statisticky významný vliv na normální rozdělení náhodného

výběru. Náhodný výběr největších délek mozkovny mužů ze starověké egyptské populace pochází z normálního rozdělení.

Protože náhodný výběr pochází z normálního rozdělení, můžeme hypotézu ze zadání otestovat pomocí parametrického testu o rozptylu  $\sigma^2$ . Testování provedeme v posloupnosti sedmi kroků. Naším úkolem je zjistit, zda se rozptyly největší délkou mozkovny mužů u starověké a novověké egyptské populace liší. Tato věta bude součástí alternativní hypotézy, neboť odlišnost implikuje nerovnost a nerovnost je vždy součástí alternativní hypotézy. Nulovou hypotézu potom stanovíme jako doplněk alternativní hypotézy.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Rozptyl největší délkou mozkovny mužů starověké egyptské populace je shodný s rozptylem největší délkou mozkovny mužů novověké egyptské populace.

$H_1$  : Rozptyl největší délkou mozkovny mužů starověké egyptské populace není shodný s rozptylem největší délkou mozkovny mužů novověké egyptské populace.

- **matematická formulace** nulové a alternativní hypotézy

$H_0 : \sigma^2 = \sigma_0^2$ , kde  $\sigma_0^2 = 7.526^2$

$H_1 : \sigma^2 = \sigma_0^2$ , kde  $\sigma_0^2 = 7.526^2$  (oboustranná alternativa)

### 2. Volba hladiny významnosti

- Hladinu významnosti volíme v souladu se zadáním jako  $\alpha = 0.01$ .

### 3. Testování kritickým oborem

- **Testovací statistika**

$$F_W = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(217-1)6.370398^2}{7.526^2} = \frac{216 \times 40.58197}{56.64068} = 154.7599.$$

```
5 alpha <- 0.01
6 sigma_0 <- 7.526
7 s <- sd(skull.LM)
8 fw <- ((n - 1) * s ^ 2) / (sigma_0 ^ 2) # 154.7599
```

- **Kritický obor**

$$\begin{aligned} W &= (0; \chi_{n-1}^2(\alpha/2)) \cup \langle \chi_{n-1}^2(1 - \alpha/2); \infty \rangle \\ &= (0; \chi_{217-1}^2(0.01/2)) \cup \langle \chi_{217-1}^2(1 - 0.01/2); \infty \rangle \\ &= (0; \chi_{216}^2(0.005)) \cup \langle \chi_{216}^2(0.995); \infty \rangle \\ &= (0; 81.1329) \cup \langle 138.6506; \infty \rangle \end{aligned}$$

```
9 qchisq(alpha / 2, n - 1) # 166.2201
10 qchisq(1 - alpha / 2, n - 1) # 273.2856
```

- **Závěr testování**

Protože realizace testovací statistiky  $f_W = 154.7599$  náleží do kritického oboru, tj.  $f_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (d, h) &= \left( \frac{(n-1)s^2}{\chi_{n-1}^2(1-\alpha/2)}, \frac{(n-1)s^2}{\chi_{n-1}^2(\alpha/2)} \right) \\
 &= \left( \frac{(217-1)6.370398^2}{\chi_{217-1}^2(1-0.01/2)}, \frac{(217-1)6.370398^2}{\chi_{217-1}^2(0.01/2)} \right) \\
 &= \left( \frac{216 \times 40.58197}{\chi_{216}^2(0.995)}, \frac{216 \times 40.58197}{\chi_{216}^2(0.005)} \right) \\
 &= \left( \frac{8765.706}{273.2856}, \frac{8765.706}{166.2201} \right) \\
 &= (32.07524, 52.7355)
 \end{aligned}$$

```

11 dh <- (n - 1) * s ^ 2 / qchisq(1 - alpha / 2, n - 1) # 32.07525
12 hh <- (n - 1) * s ^ 2 / qchisq(alpha / 2, n - 1) # 52.73553

```

- Závěr testování

Protože  $\sigma_0^2 = 7.526^2 = 56.64068$  nenáleží do Waldova 99% empirického oboustranného intervalu spolehlivosti, tj.  $\sigma_0^2 = 56.64068 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned}
 p\text{-hodnota} &= 2 \min\{\Pr(F_W \leq f_w), \Pr(F_W > f_w)\} \\
 &= 2 \min\{\Pr(F_W \leq f_w), 1 - \Pr(F_W \leq f_w)\} \\
 &= 2 \min\{\Pr(F_W \leq 154.7599), 1 - \Pr(F_W \leq 154.7599)\} \\
 &= 2 \min\{0.00057779, 0.9994222\} \\
 &= 2 \times 0.00057779 = 0.00115558 \doteq 0.001156
 \end{aligned}$$

```

13 p.hodnota <- 2 * min (pchisq(fw, n - 1), 1 - pchisq(fw, n - 1)) # 0.00115558

```

- Závěr testování

Protože  $p$ -hodnota = 0.001156 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 nulovou hypotézu na hladině významnosti  $\alpha = 0.01$ . Mezi rozptily největší délky mozkovny mužů starověké a novověké egyptské populace existuje statisticky významný rozdíl.

*Poznámka:* Test rozptylu  $\sigma^2$  můžeme provést pomocí funkce `varTest()` z knihovny `EnvStats`. Vstupními parametry budou vektor reprezentující náhodný výběr (`skull.LM`), hodnota parametru  $\sigma_0^2$  z nulové hypotézy zadaná argumentem `sigma.squared = 7.526^2`, hodnota hladiny významnosti  $\alpha$  zadaná prostřednictvím koeficientu spolehlivosti  $1 - \alpha$  nastavením hodnoty argumentu `conf.level = 0.99` a typ zvolené alternativní hypotézy (oboustranná) zadaný pomocí argumentu `alternative = 'two.sided'`.

```

14 EnvStats::varTest(skull.LM, sigma.squared = 7.526 ^ 2, conf.level = 0.99, alt = 'two.sided')

```

Součástí výstupu je hodnota výběrového rozptylu `variance = 40.58197`, hodnota testovací statistiky `Chi-squared = 154.76`, počet stupňů volnosti Studentova rozdělení `df = 216`, hranice 95% Waldova empirického oboustranného

```
Chi-Squared Test on Variance
data: skull.LM
Chi-Squared = 154.76, df = 216, p-value = 0.001156
alternative hypothesis: true variance is not equal to 56.64068
99 percent confidence interval:
 32.07525 52.73553
sample estimates:
variance
40.58197
```

15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

intervalu spolehlivosti 32.07525 a 52.73553 a  $p$ -hodnota  $p\text{-value} = 0.001156$ . Jediné, co musíme stanovit zvlášť, jsou dolní a horní hranice kritického oboru. ★

### Příklad 7.2. Test o rozptylu $\sigma^2$ (pravostranná alternativa)

Mějme datový soubor 19-more-samples-correlations-skull.txt a proměnnou nose.B popisující šířku nosu v mm mužů z německé, malajské, čínské, peruánské a bantuské populace (viz sekce ??). Dále máme k dispozici údaje o šířce nosu mužů a žen z pozdně středověké a novověké populace Polska z lokality Lekno ( $m_m = 25.60$  mm,  $s_m = 2.32$  mm,  $n_m = 59$ ,  $m_f = 24.70$  mm,  $s_f = 2.52$  mm,  $n_f = 30$ , Piontek et al. (2008)). Na hladině významnosti  $\alpha = 0.05$  testujte hypotézu, že rozptyl šířky nosu mužů německé populace je menší než rozptyl šířky nosu mužů polské populace z lokality Lekno.

### Řešení příkladu 7.2

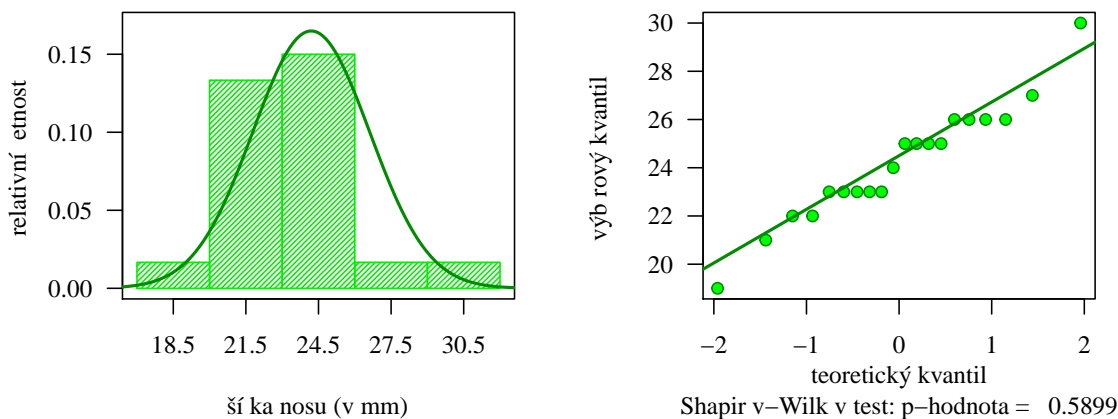
Pomocí příkazu read.delim() načteme datový soubor a pomocí operátoru [] vybereme z datové tabulky naměřené hodnoty největších délek mozkovny (skull.L) mužů (sex == 'm'). Nakonec příkazem na.omit() odstraníme z vektoru naměřených hodnot chybějící údaje a příkazem length() zjistíme rozsah náhodného výběru.

```
26 data <- read.delim('00-Data//19-more-samples-correlations-skull.txt')
27 nose.BN <- data[data$pop == 'nem', 'nose.B']
28 nose.BN <- na.omit(nose.BN)
29 n <- length(nose.BN) # 20
```

Datový soubor obsahuje údaje o šířce nosu 20 mužů německé populace.

Naším úkolem ze zadání je porovnat rozptyly německé a polské populace, přičemž u mužů z německé populace máme k dispozici naměřené hodnoty. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení, tj. zda náhodná veličina  $X$  popisující šířku nosu mužů německé populace pochází z normálního rozdělení, tj.  $X \sim N(\mu, \sigma^2)$ . Druhá, polská populace z lokality Lekno je reprezentována pouze hodnotou aritmetického průměru ( $m_m = 25.60$  mm), směrodatnou odchylkou ( $s_m = 2.32$  mm) a počtem pozorování ( $n_m = 59$ ). O jejím rozdělení přesnější informace nemáme. Řešení příkladu tedy vede na situaci, kdy rozptyl jednoho náhodného výběru porovnáváme s konkrétním číslem, tedy na jednovýběrový test o rozptylu  $\sigma^2$ . Před použitím testu musíme nejprve ověřit předpoklad normálního rozdělení náhodného výběru šířek nosu německé populace.

Hladinu významnosti  $\alpha$  pro test normality zvolíme  $\alpha = 0.05$ . Jelikož rozsah náhodného výběru je menší než 30, ověříme předpoklad normality Shapirovým-Wilkovým testem. Grafické ověření provedeme na základě QQ-diagramu a histogramu superponovaného křivkou normálního rozdělení, jejíž parametry odhadneme pomocí výběrového průměru a výběrového rozptylu (viz obrázek 2). Datový soubor rozdělíme do pěti ekvidistatních intervalů s šířkou 3 mm prostřednictvím stanovených hranic 17, 19, ..., 32.



Obrázek 2: Histogram a QQ-diagram šířky nosu mužů německé populace

Protože  $p$ -hodnota = 0.5899 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z histogramu na obrázku 2 vidíme, že naměřené hodnoty celkem věrně kopírují tvar teoretické křivky hustoty normálního rozdělení. Ani na QQ-diagramu není zřejmé žádné odchýlení bodů od referenční křivky. Náhodný

výběr šířek nosu mužů z německé populace pochází z normálního rozdělení.

Protože náhodný výběr pochází z normálního rozdělení, můžeme hypotézu ze zadání ověřit pomocí parametrického testu o rozptylu  $\sigma^2$ . Naším úkolem je otestovat hypotézu, že rozptyl šířky nosu mužů německé populace je menší než rozptyl u mužů pozdně středověké a novověků polské populace z lokality Lekno. Protože hypotézou myslíme vždy nulovou hypotézu, bude předchozí věta zněním nulové hypotézy. V rámci nulové hypotézy potom testujeme, zda je rozptyl šířky nosu mužů německé populace menší nebo roven rozptylu u mužů pozdně středověké a novověků polské populace z lokality Lekno, neboť rovnost je vždy součástí nulové hypotézy. Alternativní hypotézu potom sestavíme tak, aby byla doplňkem nulové hypotézy.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Rozptyl šířky nosu mužů německé populace je menší nebo rovný rozptylu šířky nosu mužů pozdně středověké a novověků polské populace z lokality Lekno.

$H_1$  : Rozptyl šířky nosu mužů německé populace je větší než rozptyl šířky nosu mužů pozdně středověké a novověků polské populace z lokality Lekno.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $\sigma^2 \leq \sigma_0^2$ , kde  $\sigma_0^2 = 2.32^2$

$H_1$  :  $\sigma^2 > \sigma_0^2$ , kde  $\sigma_0^2 = 2.32^2$  (pravostranná alternativa)

### 2. Volba hladiny významnosti

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

### 3. Testování kritickým oborem

- **Testovací statistika**

$$F_W = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(20-1)2.419221^2}{2.32^2} = \frac{19 \times 5.85263}{5.3824} = 20.65992.$$

```
30 alpha <- 0.05
31 sigma_0 <- 2.32
32 s <- sd(nose.BN)
33 fw <- ((n - 1) * s ^ 2) / (sigma_0 ^ 2) # 20.65993
```

- **Kritický obor**

$$\begin{aligned} W &= (0; \chi_{n-1}^2(\alpha/2)) \cup (\chi_{n-1}^2(1-\alpha/2); \infty) \\ &= (0; \chi_{20-1}^2(0.05/2)) \cup (\chi_{20-1}^2(1-0.05/2); \infty) \\ &= (0; \chi_{19}^2(0.025)) \cup (\chi_{19}^2(0.975); \infty) \\ &= (0; 8.906516) \cup (32.85233; \infty) \end{aligned}$$

```
34 qchisq(alpha / 2, n - 1) # 8.906516
35 qchisq(1 - alpha / 2, n - 1) # 32.85233
```

- **Závěr testování**

Protože realizace testovací statistiky  $f_W = 20.6599$  nenáleží do kritického oboru, tj.  $f_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (d, h) &= \left( \frac{(n-1)s^2}{\chi_{n-1}^2(1-\alpha/2)}, \frac{(n-1)s^2}{\chi_{n-1}^2(\alpha/2)} \right) \\
 &= \left( \frac{(20-1)2.419221^2}{\chi_{20-1}^2(1-0.05/2)}, \frac{(20-1)2.419221^2}{\chi_{20-1}^2(0.05/2)} \right) \\
 &= \left( \frac{19 \times 5.852632}{\chi_{19}^2(0.975)}, \frac{19 \times 5.852632}{\chi_{19}^2(0.025)} \right) \\
 &= \left( \frac{111.2}{32.85233}, \frac{111.2}{8.906516} \right) \\
 &= (3.384844, 12.48524)
 \end{aligned}$$

```

36 dh <- (n - 1) * s ^ 2 / qchisq(1 - alpha / 2, n - 1) # 3.384844
37 hh <- (n - 1) * s ^ 2 / qchisq(alpha / 2, n - 1) # 12.48524

```

- Závěr testování

Protože  $\sigma_0^2 = 2.32^2 = 5.3824$  náleží do Waldova 95% empirického levostranného intervalu spolehlivosti, tj.  $\sigma_0^2 = 5.3824 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned}
 p\text{-hodnota} &= \Pr(F_W > f_W) \\
 &= 1 - \Pr(F_W \leq f_W) \\
 &= 1 - \Pr(F_W \leq 20.65993) \\
 &= 0.355892 \doteq 0.3559
 \end{aligned}$$

```

38 p.hodnota <- 1 - pchisq(fw, n - 1) # 0.355892

```

- Závěr testování

Protože  $p$ -hodnota = 0.3559 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  $\alpha = 0.05$ . Zde je důležité si uvědomit, že pokud nulovou hypotézu nezamítáme, neznamená to, že nulová hypotéza platí. Pouze nemáme dostatek indicií k zamítnutí této hypotézy ve prospěch hypotézy alternativní. Proto jediný relevantní závěr, který z testování vyplývá je, že naměřená data nepodpořila znění alternativní hypotézy, tedy že ozpytl šířky nosu mužů německé populace není statisticky významně větší než rozpytl šířky nosu mužů pozdně středověké a novověké polské populace z lokality Lekno.

*Poznámka:* Test rozptylu  $\sigma^2$  můžeme provést pomocí funkce `varTest()` z knihovny `EnvStats`. Vstupními parametry budou vektor reprezentující náhodný výběr (`nose.BN`), hodnota parametru  $\sigma_0^2$  z nulové hypotézy zadaná argumentem `sigma.squared = 2.32^2`, hodnota hladiny významnosti  $\alpha$  zadaná prostřednictvím koeficientu spolehlivosti  $1 - \alpha$  (`argument conf.level = 0.95`) a typ zvolené alternativní hypotézy (pravostranná; `argument alternative = 'greater'`).

```

39 EnvStats::varTest(nose.BN, sigma.squared = 2.32 ^ 2, conf.level = 0.95, alt = 'greater')

```

Součástí výstupu je hodnota výběrového rozptylu `variance = 5.852632`, hodnota testovací statistiky `Chi-squared = 20.66`, počet stupňů volnosti Studentova rozdělení `df = 19`, hranice 95% Waldova empirického levostranného intervalu spolehlivosti `3.689018` a `lnf` a  $p$ -hodnota `p-value = 0.3559`. Jediné, co musíme stanovit zvlášť, jsou dolní a horní hranice kritického oboru. ★



Chi-Squared Test on Variance	40
	41
data: nose.BN	42
Chi-Squared = 20.66, df = 19, p-value = 0.3559	43
alternative hypothesis: true variance is greater than 5.3824	44
95 percent confidence interval:	45
3.689018 Inf	46
sample estimates:	47
variance	48
5.852632	49
	50

### Příklad 7.3. Test o rozptylu $\sigma^2$ (levostranná alternativa)

Mějme datový soubor 18-more-samples-variances-clavicle.txt a proměnnou cla.L popisující největší délku klíční kosti z pravé strany v mm (viz sekce ??). Dále máme k dispozici údaje o největší délce klíční kosti z pravé strany mužů z populace severozápadní Indie ( $m_{nwI} = 146.89$  mm,  $s_{nwI} = 9.23$  mm,  $n_{nwI} = 100$ , Kaur et al. (1997)). Na hladině významnosti  $\alpha = 0.05$  zjistěte, zda je rozptyl největší délky klíční kosti z pravé strany u mužů populace z Amritsaru menší než rozptyl největší délky klíční kosti z pravé strany mužů populace severozápadní Indie.

#### Řešení příkladu 7.3

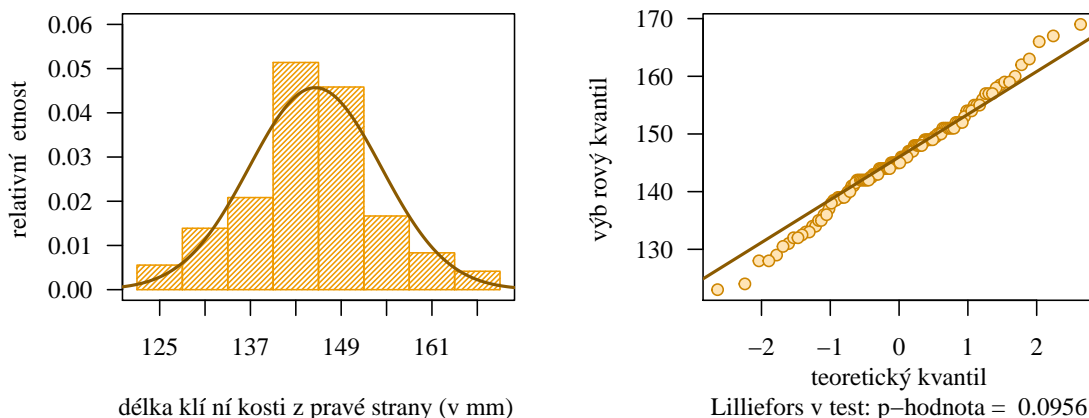
Příkazem read.delim() načteme datový soubor a pomocí operátoru [] vybereme z datové tabulky naměřené délky klíční kosti (cla.L) mužů populace z Amritsaru (pop == 'ind1'). Nakonec z vektoru naměřených hodnot odstraníme chybějící údaje (na.omit()) a zjistíme jeho délku (length()).

```
51 data <- read.delim('00-Data//18-more-samples-variances-clavicle.txt')
52 data <- na.omit(data)
53 cla.LI <- data[data$pop == 'ind1', 'cla.L']
54 n <- length(cla.LI) # 120
```

Datový soubor obsahuje údaje o délce klíční kosti z pravé strany u 120 mužů populace z Amritsaru.

Naším úkolem ze zadání je porovnat rozptyly dvou indických populací, přičemž u mužů z populace z Amritsaru známe naměřené hodnoty. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení. Druhá populace, a sice ze severní Indie je reprezentována pouze hodnotou aritmetického průměru ( $m_{nwI} = 146.89$  mm), směrodatnou odchylkou ( $s_{nwI} = 9.23$  mm) a počtem pozorování ( $n_{nwI} = 100$ ). O jejím rozdělení přesnější informace nemáme. Řešení příkladu vede na případ, kdy rozptyl jednoho náhodného výběru porovnáváme s konkrétním číslem, tedy na jednovýběrový test o rozptylu  $\sigma^2$ . Předpokladem k použití tohoto testu je normalita náhodného výběru délek klíčních kostí z pravé strany u mužů z populace z Amritsaru. Před použitím testu je třeba tento předpoklad ověřit.

Jelikož rozsah náhodného výběru je větší než 30, ověříme předpoklad normality Lillieforsovým testem ( $\alpha = 0.05$ ). Grafické ověření provedeme QQ-diagramem a histogramem (viz obrázek 3). Datový soubor rozdělíme do osmi ekvidistatních intervalů s šířkou 6 mm prostřednictvím stanovených hranic 122, 130, ..., 170.



Obrázek 3: Histogram a QQ-diagram největší délky klíční kosti z pravé strany u mužů z populace v Amritsaru

Protože  $p$ -hodnota = 0.0956 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z histogramu na obrázku 3 vidíme, že naměřené hodnoty kopírují vhodně tvar křivky hustoty normálního rozdělení. Na QQ-diagramu je zřejmý trend odlehlosti bodů od referenční křivky na pravé i levé straně. Tato odlehlost však není fatální pro předpoklad normality náhodného výběru. Náhodný výběr délek klíčních kostí z pravé strany u mužů z populace z Amritsaru pochází z normálního rozdělení.

Jelikož je předpoklad normality náhodného výběru splněn, můžeme hypotézu ze zadání otestovat pomocí parametrického testu o rozptylu  $\sigma^2$ . Naším úkolem je zjistit, zda je rozptyl největší délky klíční kosti z pravé strany u mužů populace z Amritsaru menší než rozptyl největší délky klíční kosti z pravé strany mužů populace severozápadní Indie. Tato věta je zněním alternativní hypotézy. Nulovou hypotézu stanovíme jako doplněk k tomuto tvrzení.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Rozptyl největší délky klíční kosti z pravé strany u mužů populace z Amritsaru je větší nebo roven rozptylu největší délky klíční kosti z pravé strany mužů z populace severozápadní Indie.

$H_1$  : Rozptyl největší délky klíční kosti z pravé strany u mužů populace z Amritsaru je menší než rozptyl největší délky klíční kosti z pravé strany mužů z populace severozápadní Indie.

- **matematická formulace** nulové a alternativní hypotézy

$H_0 : \sigma^2 \geq \sigma_0^2$ , kde  $\sigma_0^2 = 9.23^2$

$H_1 : \sigma^2 < \sigma_0^2$ , kde  $\sigma_0^2 = 9.23^2$  (levostranná alternativa)

### 2. Volba hladiny významnosti

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

### 3. Testování kritickým oborem

- **Testovací statistika**

$$F_W = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(120-1)8.733432^2}{9.23^2} = \frac{119 \times 76.27283}{85.1929} = 106.5402.$$

```
55 alpha <- 0.05
56 sigma_0 <- 9.23
57 s <- sd(c1a.LI)
58 fw <- ((n - 1) * s ^ 2) / (sigma_0 ^ 2) # 106.5402
```

- **Kritický obor**

$$\begin{aligned} W &= (0; \chi_{n-1}^2(\alpha)) \\ &= (0; \chi_{120-1}^2(0.05)) \\ &= (0; 94.8112) \end{aligned}$$

```
59 qchisq(alpha, n - 1) # 94.81124
```

- **Závěr testování**

Protože realizace testovací statistiky  $f_W = 106.5402$  nenáleží do kritického oboru, tj.  $f_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

### 4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

$$\begin{aligned} (0, h) &= \left(0, \frac{(n-1)s^2}{\chi_{n-1}^2(\alpha/2)}\right) \\ &= \left(0, \frac{(120-1)8.733432^2}{\chi_{120-1}^2(0.05)}\right) \\ &= \left(0, \frac{119 \times 76.27283}{94.81124}\right) \\ &= (0, 95.73197) \end{aligned}$$

```
60 hh <- (n - 1) * s ^ 2 / qchisq(alpha, n - 1) # 295.73197
```

- **Závěr testování**

Protože  $\sigma_0^2 = 9.23^2 = 85.1929$  náleží do Waldova 95% empirického pravostranného intervalu spolehlivosti, tj.  $\sigma_0^2 = 85.1929 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$p\text{-hodnota} = \Pr(T_W \leq 106.5402) = 0.2135755 \doteq 0.2136$$

```
61 p.hodnota <- pchisq(fw, n - 1) # 0.2135755
```

- **Závěr testování**

Protože  $p$ -hodnota = 0.2136 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  $\alpha = 0.05$ . Rozptyl největší délky klíční kosti z pravé strany u mužů z populace z Amritsaru není statisticky významně menší než rozptyl délky klíční kosti z pravé strany u mužů z populace severozápadní Indie.

*Poznámka:* Test rozptylu  $\sigma^2$  můžeme provést pomocí funkce `varTest()` z knihovny `EnvStats`. Vstupními parametry budou vektor reprezentující náhodný výběr (`cla.LI`), hodnota parametru  $\sigma_0^2$  (argument `sigma.squared = 9.23^2`), hodnota hladiny významnosti  $\alpha$  zadaná prostřednictvím koeficientu spolehlivosti  $1 - \alpha$  (argument `conf.level = 0.95`) a typ zvolené alternativní hypotézy (levostranná; argument `alternative = 'less'`).

```
62 EnvStats::varTest(cla.LI, sigma.squared = 9.23^2, alt = 'less')
```

```
Chi-Squared Test on Variance
data: cla.LI
Chi-Squared = 106.54, df = 119, p-value = 0.2136
alternative hypothesis: true variance is less than 85.1929
95 percent confidence interval:
 0.00000 95.73197
sample estimates:
variance
76.27283
```

63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73

Součástí výstupu je hodnota výběrového rozptylu `variance = 76.27283`, hodnota testovací statistiky `Chi-squared = 106.54`, počet stupňů volnosti Studentova rozdělení `df = 119`, hranice 95% Waldova empirického pravostranného intervalu spolehlivosti 0 a `95.73197` a  $p$ -hodnota `p-value = 0.2136`. Jediné, co musíme stanovit zvláště, je horní hranice kritického oboru. ★

#### Příklad 7.4. Test o rozptylu $\sigma$ (nesplnění předpokladů)

Mějme datový soubor 18-more-samples-variances-clavicle.txt a proměnnou cla.L popisující největší délku klíční kosti z pravé strany v mm (viz sekce ??). Dále máme k dispozici údaje o největší délce klíční kosti z pravé strany mužů z populace severní Indie ( $m_{nI} = 148.0$  mm,  $s_{nI} = 8.60$  mm,  $n_{nI} = 260$ , Jit, Sahni (1983)). Na hladině významnosti  $\alpha = 0.05$  testujte hypotézu, že rozptyl největší délky klíční kosti z pravé strany u mužů populace z Varanasi je větší nebo rovný rozptylu největší délky klíční kosti z pravé strany mužů populace severní Indie.

#### Řešení příkladu 7.4

Načteme datový soubor a vybereme z datové tabulky naměřené délky klíční kosti (cla.L) mužů populace z Varanasi (pop == 'ind2'). Nakonec z vektoru naměřených hodnot odstraníme chybějící údaje a zjistíme rozsah náhodného výběru.

```
74 data <- read.delim('00-Data//18-more-samples-variances-clavicle.txt')
75 cla.LV <- data[data$pop == 'ind2', 'cla.L']
76 cla.LV <- as.numeric(na.omit(cla.LV))
77 n <- length(cla.LV) # 81
```

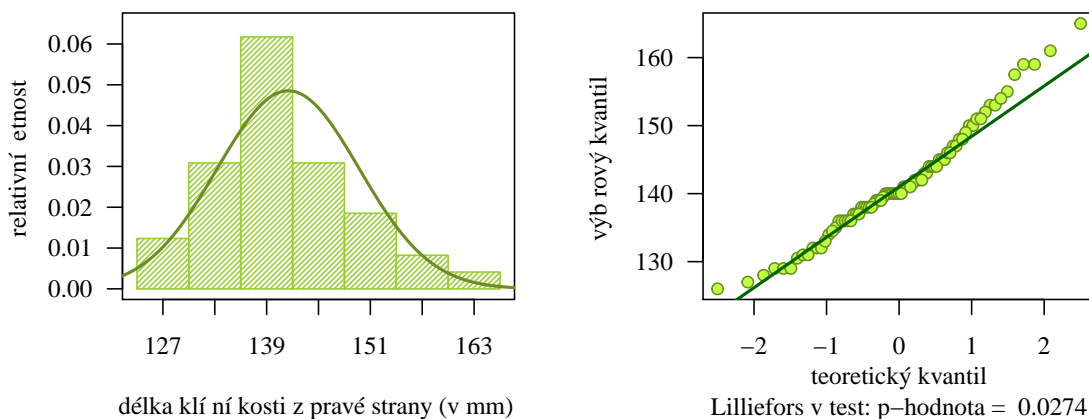
Datový soubor obsahuje údaje o délce klíční kosti z pravé strany u 81 mužů populace z Varanasi.

Naším úkolem ze zadání je porovnat rozptyly dvou indických populací, přičemž u mužů z populace z Varanasi známe naměřené hodnoty. Populace mužů ze severní Indie je naproti tomu reprezentována pouze hodnotou aritmetického průměru ( $m_{nI} = 148.00$  mm) a směrodatnou odchylkou ( $s_{nI} = 8.60$  mm). Řešení příkladu vede na případ, kdy rozptyl jednoho náhodného výběru porovnáváme s konkrétním číslem, tedy na jednovýběrový test o rozptylu  $\sigma^2$ . Před použitím testu ověříme předpoklad normality náhodného výběru délek klíčních kostí z pravé strany u mužů populace z Varanasi.

Předpoklad normality ověříme Lillieforsovým testem ( $\alpha = 0.05$ ), QQ-diagramem a histogramem (viz obrázek 4). Datový soubor rozdělíme do sedmi ekvidistančních intervalů s šířkou 6 mm prostřednictvím stanovených hranic 124, 130, ..., 166.

```
[1] 0.2361389
```

78



Obrázek 4: Histogram a QQ-diagram největší délky klíční kosti z pravé strany u mužů z populace z Varanasi

Protože  $p$ -hodnota = 0.0274 je menší než 0.05, nulovou hypotézu o normalitě dat zamítáme na hladině významnosti  $\alpha = 0.05$ . Z histogramu na obrázku 4 vidíme posunutí histogramu doleva oproti křivce hustoty normálního rozdělení a výraznou špičatost náhodného výběru ( $b_2 = 0.2361$ ). Z QQ-diagramu je zřejmá příchylnost bodů k referenční přímce pouze ve středové oblasti grafu. Všechny tyto aspekty podporují výsledek Lillieforsova testu. Náhodný výběr délek klíčních kostí z pravé strany u mužů z populace z Varanasi nepochází z normálního rozdělení.

Jelikož je předpoklad normality náhodného výběru není splněn, nemůžeme hypotézu o rozptylu uvedenou v zadání příkladu otestovat pomocí parametrického testu o rozptylu. ★

### Příklad 7.5. Test o směrodatné odchylce $\sigma$ (oboustranná alternativa)

Mějme datový soubor 11-two-samples-means-skull.txt a proměnnou skull.H popisující basion–bregmatickou výšku lebky v mm jedinců starověké egyptské populace (viz sekce ??). Dále máme k dispozici údaje o basion–bregmatické výšce lebky mužů novověké egyptské populace ( $m_m = 133.977$  mm,  $s_m = 5.171$  mm,  $n_m = 87$ ). Na hladině významnosti  $\alpha = 0.05$  zjistíte, zda jsou směrodatné odchylky basion–bregmatické výšky lebky mužů starověké a novověké egyptské populace shodné.

#### Řešení příkladu 7.5

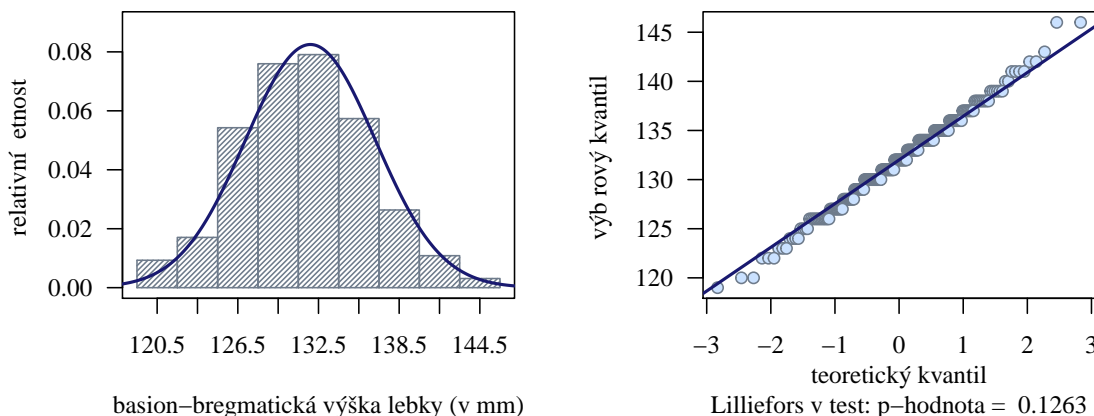
Nejprve načteme datový soubor. Pomocí operátoru `[]` vybereme z datové tabulky naměřené basion–bregmatické výšky lebky (skull.H) mužů (`sex == 'm'`). Nakonec z vektoru naměřených hodnot odstraníme chybějící údaje a zjistíme počet naměřených údajů.

```
79 data <- read.delim('00-Data//11-two-samples-means-skull.txt')
80 skull.HM <- data[data$sex == 'm', 'skull.H']
81 skull.HM <- as.numeric(na.omit(skull.HM))
82 n <- length(skull.HM) # 215
```

Datový soubor obsahuje údaje o basion-bregmatické výšce lebky 215 mužů starověké egyptské populace.

Naším úkolem ze zadání je porovnat směrodatné odchylky dvou egyptských populací, přičemž u mužů ze starověké egyptské populace máme k dispozici naměřené hodnoty. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení. Druhá, novověká egyptská populace je reprezentována pouze hodnotou aritmetického průměru ( $m_m = 133.977$  mm) a směrodatnou odchylkou ( $s_m = 5.171$  mm). O jejím rozdělení přesnější informace nemáme. V rámci příkladu tedy porovnáváme směrodatnou odchylku jednoho náhodného výběru s konkrétním číslem. K tomu, po vhodné modifikaci nulové a alternativní hypotézy, použijeme jednovýběrový test o rozptylu  $\sigma^2$ . Jediným předpokladem k použití tohoto testu je normální rozdělení náhodného výběru basion–bregmatických výšek lebky mužů starověké egyptské populace. Před použitím testu tedy tento předpoklad ověříme.

Vzhledem k rozsahu náhodného výběru ověříme předpoklad normality Lillieforsovým testem ( $\alpha = 0.05$ ). Graficky zhodnotíme náhodný výběr QQ-diagramem a histogramem (viz obrázek 5). Datový soubor rozdělíme do devíti ekvidistatních intervalů s šířkou 3 mm prostřednictvím stanovených hranic 119, 122, ..., 146.



Obrázek 5: Histogram a QQ-diagram basion–bregmatické výšky lebky mužů starověké egyptské populace

Protože  $p$ -hodnota = 0.1263 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z histogramu na obrázku 5 vidíme, že naměřené hodnoty věrně kopírují tvar křivky hustoty normálního rozdělení. Na QQ-diagramu je viditelná příchýlnost bodů k referenční křivce. Náhodný výběr basion–bregmatické výšky lebky mužů ze starověké egyptské populace tedy pochází z normálního rozdělení.

Protože náhodný výběr pochází z normálního rozdělení, můžeme hypotézu ze zadání otestovat pomocí paramet-

rického testu. Naším úkolem je zjistit, zda jsou směrodatné odchylky basion–bregmatické výšky lebky mužů starověké a novověké egyptské populace shodné. Jelikož rozptyl není nic jiného než mocnina směrodatné odchylky, můžeme bez újmy na obecnosti znění této věty modifikovat na otázku, zda jsou rozptyly basion–bregmatické výšky lebky mužů starověké a novověké egyptské populace shodné. Tato věta potom bude součástí  $H_0$ , neboť shoda implikuje rovnost a rovnost je vždy součástí nulové hypotézy. Alternativní hypotézu potom stanovíme jako doplněk nulové hypotézy.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Rozptyl basion–bregmatické výšky lebky mužů starověké egyptské populace a rozptyl basion–bregmatické výšky lebky novověké egyptské populace jsou shodné.

$H_1$  : Rozptyl basion–bregmatické výšky lebky mužů starověké egyptské populace a rozptyl basion–bregmatické výšky lebky novověké egyptské populace nejsou shodné.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $\sigma^2 = \sigma_0^2$ , kde  $\sigma_0^2 = 5.171^2$

$H_1$  :  $\sigma^2 \neq \sigma_0^2$ , kde  $\sigma_0^2 = 5.171^2$  (oboustranná alternativa)

### 2. Volba hladiny významnosti

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

### 3. Testování kritickým oborem

- **Testovací statistika**

$$F_W = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(215-1)4.835494^2}{5.171^2} = \frac{214 \times 23.382}{26.73924} = 187.1313.$$

```
83 alpha <- 0.05
84 sigma_0 <- 5.171
85 s <- sd(skull.HM)
86 fw <- ((n - 1) * s ^ 2) / (sigma_0 ^ 2) # 187.1313
```

- **Kritický obor**

$$\begin{aligned} W &= (0; \chi_{n-1}^2(\alpha/2)) \cup \langle \chi_{n-1}^2(1 - \alpha/2); \infty \rangle \\ &= (0; \chi_{215-1}^2(0.05/2)) \cup \langle \chi_{215-1}^2(1 - 0.05/2); \infty \rangle \\ &= (0; \chi_{214}^2(0.025)) \cup \langle \chi_{214}^2(0.975); \infty \rangle \\ &= (0; 175.3782) \cup \langle 256.4079; \infty \rangle \end{aligned}$$

```
87 qchisq(alpha / 2, n - 1) # 175.3782
88 qchisq(1 - alpha / 2, n - 1) # 256.4079
```

- **Závěr testování**

Protože realizace testovací statistiky  $f_W = 187.1313$  nenáleží do kritického oboru, tj.  $f_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

### 4. Testování intervalem spolehlivosti



- Interval spolehlivosti

$$\begin{aligned}
 (d, h) &= \left( \frac{(n-1)s^2}{\chi_{n-1}^2(1-\alpha/2)}, \frac{(n-1)s^2}{\chi_{n-1}^2(\alpha/2)} \right) \\
 &= \left( \frac{(215-1)4.835494^2}{\chi_{215-1}^2(1-0.05/2)}, \frac{(215-1)4.835494^2}{\chi_{215-1}^2(0.05/2)} \right) \\
 &= \left( \frac{214 \times 23.382}{\chi_{214}^2(0.975)}, \frac{214 \times 23.382}{\chi_{214}^2(0.025)} \right) \\
 &= \left( \frac{5003.748}{256.4079}, \frac{5003.748}{175.3782} \right) \\
 &= (19.5148, 28.53119)
 \end{aligned}$$

```

89 dh <- (n - 1) * s ^ 2 / qchisq(1 - alpha / 2, n - 1) # 19.5148
90 hh <- (n - 1) * s ^ 2 / qchisq(alpha / 2, n - 1) # 28.5312

```

- Závěr testování

Protože  $\sigma_0^2 = 5.171^2 = 26.7392$  náleží do Waldova 95% empirického oboustranného intervalu spolehlivosti, tj.  $\sigma_0^2 = 26.73924 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned}
 p\text{-hodnota} &= 2 \min\{\Pr(F_W \leq f_W), 1 - \Pr(F_W \leq f_W)\} \\
 &= 2 \min\{\Pr(T_W \leq 187.1313), 1 - \Pr(T_W \leq 187.1313)\} \\
 &= 2 \min\{0.09260461, 0.9073954\} \\
 &= 2 \times 0.09260461 = 0.1852092 \doteq 0.1852
 \end{aligned}$$

```

91 p.hodnota <- 2 * min(pchisq(fw, n - 1), 1 - pchisq(fw, n - 1)) # 0.1852092

```

- Závěr testování

Protože  $p$ -hodnota = 0.1852 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 rozptyly basion-bregmatické výšky lebky mužů starověké a novověké egyptské populace neexistuje statisticky významný rozdíl. Výsledný závěr můžeme bez újmy na obecnosti modifikovat na závěr týkající se směrodatné odchylky. Mezi směrodatnými odchylkami basion-bregmatické výšky lebky mužů starověké a novověké egyptské populace neexistuje statisticky významný rozdíl.

*Poznámka:* Test o směrodatné odchylce  $\sigma$  můžeme, analogicky jako test o rozptylu  $\sigma^2$ , provést pomocí funkce `varTest()` s vstupním vektorem `dat.skull.HM` a argumenty `sigma.squared = 5.171^2`, `conf.level = 0.95` a `alternative = 'two.sided'`.

```

92 EnvStats::varTest(skull.HM, sigma.squared = 5.171^2, alt = 'two.sided')

```



Chi-Squared Test on Variance	93
data: skull.HM	94
Chi-Squared = 187.13, df = 214, p-value = 0.1852	95
alternative hypothesis: true variance is not equal to 26.73924	96
95 percent confidence interval:	97
19.5148 28.5312	98
sample estimates:	99
variance	100
23.382	101
	102
	103

### Příklad 7.6. Test o směrodatné odchylce $\sigma$ (pravostranná alternativa)

Mějme datový soubor 03-paired-means-clavicle2.txt sestávající z osteometrických dat klíční kosti anglického souboru dokumentovaných skeletů (Parsons, 1916; viz sekce ??). Dále máme k dispozici údaje o největší délce klíční kosti z pravé strany u mužských a ženských skeletů ze středověké anglické populace ( $m_m = 148.7$  mm,  $s_m = 8.5$  mm,  $n_m = 25$ ,  $m_f = 133.9$  mm,  $s_f = 6.7$  mm,  $n_f = 23$ , Atterton et al. (2016)). Na hladině významnosti  $\alpha = 0.10$  zjistíte, zda je směrodatná odchylka největší největší délky klíční kosti z pravé strany u žen z anglického souboru dokumentovaných skeletů větší než směrodatná odchylka největší délky klíční kosti u žen středověké anglické populace.

#### Řešení příkladu 7.6

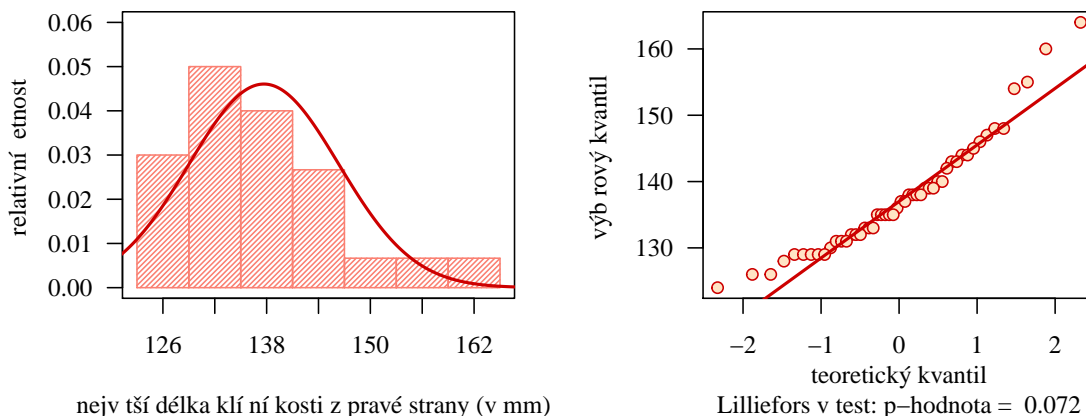
Načteme datový soubor a z datové tabulky vybereme údaje o největší délce klíční kosti z pravé strany (length.R) u žen (sex == 'f'). Z vektoru hodnot následně odstraníme chybějící hodnoty a zjistíme rozsah náhodného výběru.

```
104 data <- read.delim('00-Data//03-paired-means-clavicle2.txt')
105 length.RF <- data[data$sex == 'f', 'length.R']
106 length.RF <- as.numeric(na.omit(length.RF))
107 n <- length(length.RF) # 50
```

Datový soubor obsahuje 50 údajů o největší délce klíční kosti z pravé strany u žen.

Naším úkolem je porovnat směrodatné odchylky dvou anglických populací, přičemž u žen z anglického souboru dokumentovaných skeletů máme k dispozici naměřené hodnoty. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení. Oproti tomu středověká anglická populace je reprezentována pouze hodnotou aritmetického průměru ( $m_f = 133.9$  mm), směrodatnou odchylkou ( $s_f = 6.7$  mm) a počtem pozorování ( $n_f = 23$ ). O jejím rozdělení přesnější informace nemáme. Řešení příkladu tedy vede na situaci, kdy porovnáváme směrodatnou odchylku jednoho náhodného výběru s konkrétním číslem. K tomu po modifikaci nulové a alternativní hypotézy použijeme jednovýběrový test o rozptylu  $\sigma^2$ . Před použitím testu ověříme splnění předpokladu o normalitě náhodného výběru největších délek klíčních kostí z pravé strany u žen ze souboru dokumentovaných skeletů.

Předpoklad normality ověříme Lillieforsovým testem ( $\alpha = 0.05$ ), QQ-diagramem a histogramem (obrázek 6). Datový soubor rozdělíme do sedmi ekvidistantních intervalů s šířkou 4 mm prostřednictvím stanovených hranic 123, 129, ..., 165.



Obrázek 6: Histogram a QQ-diagram největší délky klíční kosti z pravé strany u žen z anglického souboru dokumentovaných skeletů

Protože  $p$ -hodnota = 0.0720 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z histogramu na obrázku 6 vidíme, že naměřené hodnoty mají tvar charakteristický pro normální rozdělení, pouze jsou oproti teoretické křivce normálního rozdělení posunuté doleva. Na QQ-diagramu je zřejmé odchýlení bodů od referenční křivky na levém i pravém chvostu. Podle Lillieforsova testu nemá však toto odchýlení statisticky významný vliv na normální rozdělení náhodného výběru. Náhodný výběr největších délek klíčních kostí z pravé

strany u žen z anglického souboru dokumentovaných skeletů pochází z normálního rozdělení.

Jelikož náhodný výběr pochází z normálního rozdělení, můžeme otázku ze zadání prověřit pomocí parametrického testu. Naším úkolem je zjistit, zda je směrodatná odchylka největší délky klíční kosti z pravé strany u žen z anglického dokumentovaného souboru skeletů větší než směrodatná odchylka největší délky klíční kosti zpravé strany u žen středověké anglické populace. Tato věta bude zněním alternativní hypotézy. Bez újmy na obecnosti můžeme znění alternativní hypotézy modifikovat na tvrzení o rozptylu, a sice: Rozptyl největší délky klíční kosti z pravé strany u žen z anglického dokumentovaného souboru skeletů je větší než rozptyl největší délky klíční kosti zpravé strany u žen středověké anglické populace. Nulovou hypotézu potom stanovíme jako doplněk k alternativní hypotéze.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Rozptyl největší délky klíční kosti z pravé strany u žen z anglického dokumentovaného souboru skeletů je menší nebo rovný rozptylu největší délky klíční kosti zpravé strany u žen středověké anglické populace.

$H_1$  : Rozptyl největší délky klíční kosti z pravé strany u žen z anglického dokumentovaného souboru skeletů je větší než rozptyl největší délky klíční kosti zpravé strany u žen středověké anglické populace.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $\sigma^2 \leq \sigma_0^2$ , kde  $\sigma_0^2 = 6.7^2$

$H_1$  :  $\sigma^2 > \sigma_0^2$ , kde  $\sigma_0^2 = 6.7^2$  (pravostranná alternativa)

### 2. Volba hladiny významnosti

- Hladinu významnosti volíme v souladu se zadáním jako  $\alpha = 0.10$ .

### 3. Testování kritickým oborem

- **Testovací statistika**

$$F_W = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(109-1)4.695991^2}{5.361^2} = \frac{108 \times 22.05233}{28.74032} = 82.86796 \doteq 82.8680.$$

```
108 alpha <- 0.10
109 sigma_0 <- 6.7
110 s <- sd(length.RF)
111 fw <- ((n - 1) * s ^ 2) / (sigma_0 ^ 2) # 81.90867
```

- **Kritický obor**

$$\begin{aligned} W &= \langle \chi_{n-1}^2(1 - \alpha/2); \infty \rangle \\ &= \langle \chi_{50-1}^2(1 - 0.10/2); \infty \rangle \\ &= \langle \chi_{49}^2(0.95); \infty \rangle \\ &= \langle 62.03754; \infty \rangle \end{aligned}$$

```
112 qchisq(1 - alpha, n - 1) # 62.03754
```

- **Závěr testování**

Protože realizace testovací statistiky  $f_W = 81.90867$  náleží do kritického oboru, tj.  $f_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.10$ .

### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (d, \infty) &= \left( \frac{(n-1)s^2}{\chi_{n-1}^2(1-\alpha)}, \infty \right) \\
 &= \left( \frac{(50-1)8.662469^2}{\chi_{50-1}^2(1-0.10)}, \infty \right) \\
 &= \left( \frac{49 \times 75.03837}{\chi_{49}^2(0.90)}, \infty \right) \\
 &= \left( \frac{3676.88}{62.03754}, \infty \right) \\
 &= (59.26863, \infty)
 \end{aligned}$$

```
113 dh <- (n - 1) * s ^ 2 / qchisq(1 - alpha, n - 1) # 59.26863
```

- Závěr testování

Protože  $\sigma_0^2 = 6.7^2 = 44.89$  nenáleží do Waldova 90% empirického levostranného intervalu spolehlivosti, tj.  $\sigma_0^2 = 44.89 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.10$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$p\text{-hodnota} = 1 - \Pr(F_W \leq f_W) = 1 - \Pr(F_W \leq 81.90867) = 0.002219677 \doteq 0.002220$$

```
114 p.hodnota <- 1 - pchisq(fw, n - 1) # 0.002219677
```

- Závěr testování

Protože  $p$ -hodnota = 0.00222 je menší než  $\alpha = 0.10$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.10$ .

## 6. Interpretace výsledků

Za základě všech tří typů testování zamítáme nulovou hypotézu na hladině významnosti  $\alpha = 0.10$ . Rozptyl největší délky klíční kosti z pravé strany u žen z anglického souboru dokumentovaných skeletů je statisticky významně větší než rozptyl největší délky klíční kosti z pravé strany u žen středověké anglické populace. Bez újmy na obecnosti můžeme tento závěr rozšířit na závěr o směrodatné odchylce: Směrodatná odchylka největší délky klíční kosti z pravé strany u žen z anglického souboru dokumentovaných skeletů je statisticky významně větší než směrodatná odchylka největší délky klíční kosti z pravé strany u žen středověké anglické populace.

*Poznámka:* Test o směrodatné odchylce  $\sigma$  můžeme, analogicky jako test o rozptylu  $\sigma^2$ , provést pomocí funkce `varTest()` s vstupním vektorem `dat` `length.RF` a argumenty `sigma.squared = 6.7^2`, `conf.level = 0.90` a `alternative = 'greater'`.

```
115 EnvStats::varTest(length.RF, sigma.squared = 6.7^2, conf.level = 0.90, alt = 'greater')
```



```
Chi-Squared Test on Variance
data: length.RF
Chi-Squared = 81.909, df = 49, p-value = 0.00222
alternative hypothesis: true variance is greater than 44.89
90 percent confidence interval:
 59.26863      Inf
sample estimates:
variance
75.03837
```

116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126

### Příklad 7.7. Test o směrodatné odchylce $\sigma$ (levostranná alternativa)

Mějme datový soubor 21-goldman-measures.csv a proměnnou tibia.LR popisující největší délku holenní kosti z pravé strany v mm u mužských skeletů z Kanárských ostrovů (viz sekce ??). Dále máme k dispozici údaje o největší délce holenní kosti z pravé strany u mužských a ženských skeletů z prehispanického období z ostrova Gran Canaria ( $m_m = 369.12$  mm,  $s_m = 26.02$  mm,  $n_m = 45$ ;  $m_f = 334.71$  mm,  $s_f = 13.76$  mm,  $n_f = 14$ , Gonsález-Reimers et al. (2000)). Na hladině významnosti  $\alpha = 0.01$  testujte hypotézu, že největší délka holenní kosti z pravé strany u mužských skeletů z Kanárských ostrovů je větší nebo rovna největší délce holenní kosti u mužských skeletů z ostrova Gran Canaria.

#### Řešení příkladu 7.7

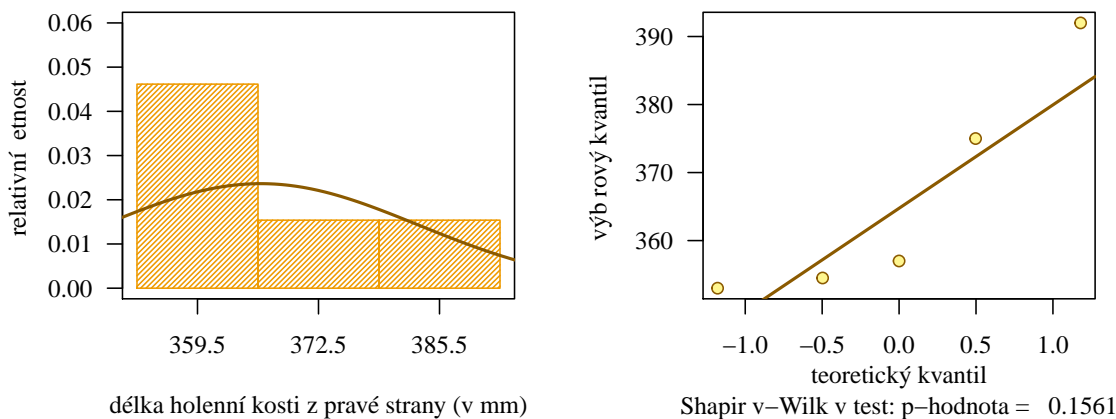
Nejprve načteme datový soubor. Pomocí operátoru `[]` vybereme z datové tabulky naměřené délky holenní kosti z pravé strany (tibia.LR) u mužských skeletů (`sex == 'm'`) z Kanárských ostrovů (`pop == 'Canary Islander'`). Nakonec z vektoru naměřených hodnot odstraníme chybějící údaje a zjistíme počet naměřených údajů.

```
127 data <- read.delim('00-Data//21-goldman-measures.csv', sep = ';')
128 tibia.LRM <- data[data$sex == 'm' & data$pop == 'Canary Islander', 'tibia.LR']
129 tibia.LRM <- as.numeric(na.omit(tibia.LRM))
130 n <- length(tibia.LRM) # 5
```

Datový soubor obsahuje údaje o délce holenní kosti z pravé strany naměřené u pěti mužských skeletů z Kanárských ostrovů.

Naším úkolem ze zadání je porovnat směrodatné odchylky populace z Kanárských ostrovů a z ostrova Gran Canaria, přičemž u mužských skeletů z Kanárských ostrovů máme k dispozici naměřené hodnoty. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení. Naopak populace z ostrova Gran Canaria je reprezentována pouze hodnotou aritmetického průměru ( $m_m = 369.12$  mm), směrodatnou odchylkou ( $s_m = 26.02$  mm) a počtem pozorování ( $n_m = 45$ ). O jejím rozdělení přesnější informace nemáme. V rámci příkladu tedy porovnáváme směrodatnou odchylku jednoho náhodného výběru s konkrétním číslem. K tomu, po vhodné modifikaci nulové a alternativní hypotézy, použijeme jednovýběrový test o rozptylu  $\sigma^2$ . Před testováním nulové hypotézy nejprve ověříme předpoklad normality náhodného výběru délek holenních kostí z pravé strany u mužských skeletů z Kanárských ostrovů.

Vzhledem k rozsahu náhodného výběru ověříme předpoklad normality Shapirovým-Wilkovým testem ( $\alpha = 0.05$ ). Graficky zhodnotíme náhodný výběr QQ-diagramem a histogramem (viz obrázek 7). Datový soubor rozdělíme do tří ekvidistatních intervalů s šířkou 13 mm prostřednictvím stanovených hranic 353, 366, 379 a 392.



Obrázek 7: Histogram a QQ-diagram největší délky holenní kosti u mužských skeletů z Kanárských ostrovů

Protože  $p$ -hodnota = 0.1516 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Jelikož datový soubor sestává pouze z pěti naměřených hodnot, je potřeba mít napaměti, že pro takto malý

rozsah náhodného výběru je test normality (zde konkrétně Shapirův-Wilkův test) velmi benevolentní a hypotézu o normálním rozdělení náhodného výběru zamítá jen v takových případech, kde data normální rozdělení vůbec nepřipomínají. Konkrétním příkladem může být situace, kdy jsou data tvořena dvěma shluky, jeden je tvořený dvěma pozorováními, druhý je tvořený třemi pozorováními a oba shluky jsou od sebe velmi vzdálené. Taktéž je potřeba si uvědomit, že při malém počtu pozorování nám histogram i QQ-diagram dávají pouze velmi hrubý pohled na potenciální rozdělení náhodného výběru. Ačkoli se z obrázku 7 nezdá, že by histogram byl dostatečně dobře superponován křivkou normálního rozdělení, v případě takto malého počtu pozorování je situace uspokojivá. Z QQ-diagramu je potom zřejmé, že body se relativně dobře drží v okolí referenční přímky. Souhlasíme tedy se závěrem Shapirova-Wilkova testu, že datový soubor délek holenních kostí z pravé strany u mužských skeletů z Kanárských ostrovů pochází z normálního rozdělení.

Protože náhodný výběr pochází z normálního rozdělení, můžeme hypotézu ze zadání otestovat pomocí parametrického testu. Naším úkolem je otestovat, zda je směrodatná odchylka největší délky holenní kosti z pravé strany u mužských skeletů z Kanárských ostrovů větší nebo rovna směrodatné odchylce největší délky holenní kosti z pravé strany u mužských skeletů z ostrova Gran Canaria. Tato věta je zněním nulové hypotézy. Navedou nás na to dvě indicie. Jednak slovo *otestovat* je vždy spjato s nulovou hypotézou a navíc zmiňovaná rovnost je vždy součástí nulové hypotézy. Nyní si stačí uvědomit, že rozptyl není nic jiného než mocnina směrodatné odchylky, můžeme proto bez újmy na obecnosti znění nulové hypotézy modifikovat tak, že slovní spojení „*směrodatná odchylka*“ nahradíme slovem „*rozptyl*“. Alternativní hypotézu potom stanovíme jako doplněk nulové hypotézy.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Rozptyl největší délky holenní kosti z pravé strany u mužských skeletů z Kanárských ostrovů je větší nebo roven rozptylu největší délky holenní kosti z pravé strany u mužských skeletů z ostrova Gran Canaria.

$H_1$  : Rozptyl největší délky holenní kosti z pravé strany u mužských skeletů z Kanárských ostrovů je menší než rozptyl největší délky holenní kosti z pravé strany u mužských skeletů z ostrova Gran Canaria.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $\sigma^2 \geq \sigma_0^2$ , kde  $\sigma_0^2 = 26.02^2$

$H_1$  :  $\sigma^2 < \sigma_0^2$ , kde  $\sigma_0^2 = 26.02^2$  (levostranná alternativa)

### 2. Volba hladiny významnosti

- Hladinu významnosti volíme v souladu se zadáním jako  $\alpha = 0.01$ .

### 3. Testování kritickým oborem

- **Testovací statistika**

$$F_W = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(5-1)16.87306^2}{26.02^2} = \frac{4 \times 284.7002}{677.0404} = 1.682028.$$

```
131 alpha <- 0.01
132 sigma_0 <- 26.02
133 s <- sd(tibia.LRM)
134 fw <- ((n - 1) * s ^ 2) / (sigma_0 ^ 2) # 1.682027
```

- **Kritický obor**

$$\begin{aligned} W &= (0; \chi_{n-1}^2(\alpha)) \\ &= (0; \chi_{5-1}^2(0.01)) \\ &= (0; \chi_4^2(0.01)) \\ &= (0; 0.2971095) \end{aligned}$$



```
135 qchisq(alpha, n - 1) # 0.2971095
```

- **Závěr testování**

Protože realizace testovací statistiky  $f_W = 1.68203$  nenáleží do kritického oboru, tj.  $f_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.01$ .

#### 4. Testování intervalem spolehlivosti

- **Interval spolehlivosti** Oproti levostranné alternativě postavíme pravostranný 99% interval spolehlivosti.

$$\begin{aligned}(0, h) &= \left( 0, \frac{(n-1)s^2}{\chi_{n-1}^2(\alpha)} \right) \\ &= \left( 0, \frac{(5-1)16.87306^2}{\chi_{5-1}^2(0.01)} \right) \\ &= \left( 0, \frac{4 \times 284.7002}{\chi_4^2(0.01)} \right) \\ &= \left( 0, \frac{1138.801}{0.2971095} \right) \\ &= (0, 3832.934)\end{aligned}$$

```
136 hh <- (n - 1) * s ^ 2 / qchisq(alpha, n - 1) # 3832.931
```

- **Závěr testování**

Protože  $\sigma_0^2 = 26.02^2 = 677.0404$  náleží do Waldova 99% empirického pravostranného intervalu spolehlivosti, tj.  $\sigma_0^2 = 677.0404 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.01$ .

#### 5. Testování $p$ -hodnotou

- **$p$ -hodnota**

$$\begin{aligned}p\text{-hodnota} &= \Pr(F_W \leq f_W) \\ &= \Pr(F_W \leq 1.682027) \\ &= 0.2060201 \doteq 0.2060\end{aligned}$$

```
137 p.hodnota <- pchisq(fw, n - 1) # 0.2060201
```

- **Závěr testování**

Protože  $p$ -hodnota = 0.2060 je větší než  $\alpha = 0.01$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.01$ .

#### 6. Interpretace výsledků

Na základě všech tří typů testování nezamítáme nulovou hypotézu na hladině významnosti  $\alpha = 0.01$ . Nezamítáme tedy nulovou hypotézu ve prospěch alternativní hypotézy. Jediný závěr, který z testování tedy vyplývá, je, že rozptyl největší délky holenní kosti z pravé strany u mužských skeletů z Kanárských ostrovů není statisticky významně menší než rozptyl největší délky holenní kosti z pravé strany u mužských skeletů z ostrova Gran Canaria. Tento závěr můžeme rozšířit i pro směrodatnou odchylku, která je odmocninou z rozptylu, tj. směrodatná odchylka největší délky holenní kosti z pravé strany u mužských skeletů z Kanárských ostrovů není statisticky významně menší než směrodatná odchylka největší délky holenní kosti z pravé strany u mužských skeletů z ostrova Gran Canaria. Porovnáme-li hodnoty směrodatných odchylek (směrodatná odchylka největší délky holenní kosti u mužských skeletů z Kanárských ostrovů je 16.87, směrodatná odchylka

největší délky holenní kosti u mužských skeletů z ostrova Gran Canaria je 26.02), mohl by nás závěr testování poněkud překvapit. Vzhledem k vysokému rozdílu hodnot obou směrodatných odchylek bychom předpokládali, že rozdíl mezi směrodatnými odchylkami bude statisticky významný. Nicméně zde hraje důležitou roli právě nízký rozsah náhodného výběru. Při takto malém rozsahu náhodného výběru by rozdíl mezi směrodatnými odchylkami musel být ještě markantnější, aby vyšel statisticky významný. Stejně tak, by rozdíl mezi stávajícími směrodatnými odchylkami (s hodnotami 16.87 a 26.02) vyšel jako statisticky významný, kdyby byl podpořen větším množstvím pozorování.

*Poznámka:* Test o směrodatné odchylce  $\sigma$  můžeme, analogicky jako test o rozptylu  $\sigma^2$ , provést pomocí funkce `varTest()` s vstupním vektorem `dat.tibia.LRM` a argumenty `sigma.squared = 26.02^2`, `conf.level = 0.99` a `alternative = 'less'`.

```
138 EnvStats::varTest(tibia.LRM, sigma.squared = 26.02 ^ 2 , alt = 'less')
```

```
Chi-Squared Test on Variance
data:  tibia.LRM
Chi-Squared = 1.682, df = 4, p-value = 0.206
alternative hypothesis: true variance is less than 677.0404
95 percent confidence interval:
 0.000 1602.312
sample estimates:
variance
 284.7
```

139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149



## 7.2 Test o parametru $\mu$ když $\sigma^2$ známe

Nechť  $X_1, \dots, X_n$  je náhodný výběr z  $N(\mu, \sigma^2)$ , kde  $\sigma^2$  známe a  $\mu_0$  je konstanta. Na hladině významnosti  $\alpha = 0.05$  testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

$$\begin{array}{lll} H_{01} : \mu = \mu_0 & \text{oproti} & H_{11} : \mu \neq \mu_0 \quad (\text{oboustranná alt.}) \\ H_{02} : \mu \leq \mu_0 & \text{oproti} & H_{12} : \mu > \mu_0 \quad (\text{pravostranná alt.}) \\ H_{03} : \mu \geq \mu_0 & \text{oproti} & H_{13} : \mu < \mu_0 \quad (\text{levostranná alt.}) \end{array}$$

Test nazýváme jednovýběrový  $Z$ -test. Testovací statistika má tvar

$$Z_W = \frac{M - \mu_0}{\sigma} \sqrt{n}, \quad (7.2)$$

kde  $M$  je výběrový průměr,  $\sigma = \sqrt{\sigma^2}$  je známá směrodatná odchylka,  $n$  je rozsah náhodného výběru a  $\mu_0$  je konstanta z nulové hypotézy. Za platnosti nulové hypotézy pochází statistika  $Z_W$  ze standardizovaného normálního rozdělení, tj.

$$Z_W = \frac{M - \mu_0}{\sigma} \sqrt{n} \stackrel{H_0}{\sim} N(0, 1).$$

Kritický obor podle zvolené alternativní hypotézy má tvar

$$\begin{array}{ll} H_{11} : \mu \neq \mu_0 & W = (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty) \\ H_{12} : \mu > \mu_0 & W = (u_{1-\alpha}; \infty) \\ H_{13} : \mu < \mu_0 & W = (-\infty; u_{\alpha}) \end{array}$$

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

Interval spolehlivosti má podle zvolené alternativní hypotézy jeden z následujících tvarů

$$\begin{array}{ll} H_{11} : \mu \neq \mu_0 & (d, h) = \left( m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, m - \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \right) \\ H_{12} : \mu > \mu_0 & (d, \infty) = \left( m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}, \infty \right) \\ H_{13} : \mu < \mu_0 & (-\infty, h) = \left( -\infty, m - \frac{\sigma}{\sqrt{n}} u_{\alpha} \right) \end{array}$$

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

$$\begin{array}{ll} H_{11} : \mu \neq \mu_0 & p\text{-hodnota} = 2 \min\{\Pr(Z_W \leq z_W), \Pr(Z_W > z_W)\} \\ H_{12} : \mu > \mu_0 & p\text{-hodnota} = \Pr(Z_W > z_W) = 1 - \Pr(Z_W \leq z_W) \\ H_{13} : \mu < \mu_0 & p\text{-hodnota} = \Pr(Z_W \leq z_W) \end{array}$$

kde  $Z_W$  je náhodná veličina,  $z_W$  je realizace testovací statistiky  $Z_W$  (viz vzorec 7.2), tedy konkrétní číslo, a  $\Pr(Z_W \leq z_W)$  je distribuční funkce standardizovaného normálního rozdělení, jejíž hodnotu získáme pomocí  $\mathbb{R}$  a implementované funkce `pnorm()`.

*Poznámka:* Teorie k testu o parametru  $\mu$  když  $\sigma^2$  známe slouží zejména jako základ pro odvozování pokročilých testovacích statistik a intervalů spolehlivosti (zejména jde o skóre a věrohodnostní testy či intervaly spolehlivosti) nebo jako základ pro různé simulační studie. Ty využíváme k hlubšímu studiu testovacích statistik a jejich rozdělení, ke zkoumání aktuální hladiny významnosti nebo aktuální hodnoty koeficientu spolehlivosti a v mnoha dalších případech. Při analýze reálných dat tento test však nevyužíváme, jelikož skutečnou hodnotu parametru  $\sigma^2$  neznáme. Proto dáváme přednost testu uvedenému v sekci 7.3, který nezalost skutečného rozptylu  $\sigma^2$  zohledňuje. Protože však znalost testu o parametru  $\mu$  když  $\sigma^2$  známe patří k základnímu statistickému vzdělání, uvádíme zde alespoň jeden ilustrační příklad pro získání povědomí o tomto testu, jeho zkonstruování a použití.

### Příklad 7.8. Test o $\mu$ když $\sigma^2$ známe (oboustranná alternativa)

Deklarovaná hmotnost jednoho balení hydroxidu sodného je 1 kg s povolenou odchylkou 0.01 kg (1%). Po zakoupení 10 balení hydroxidu sodného byl zvážen obsah každého balení. Navážené hodnoty jsou uvedeny v tabulce 1.

Tabulka 1: Naměřené hodnoty hmotnosti deseti balení hydroxidu sodného

balení	1	2	3	4	5	6	7	8	9	10
hmotnost (kg)	1.0051	0.9880	0.9921	0.9933	0.9766	0.9976	0.9819	0.9992	1.0053	1.0066

Za předpokladu, že náhodná veličina  $X$  popisující hmotnost jednoho balení hydroxidu sodného pochází z normálního rozdělení se střední hodnotou  $\mu = 1$  a rozptylem  $\sigma^2 = 0.01^2$ , tj.  $X \sim N(1, 0.01^2)$  testujte hypotézu, že hmotnost jednoho balení hydroxidu sodného je rovna deklarovanému 1 kg. Hladinu významnosti zvolte  $\alpha = 0.05$ .

### Řešení příkladu 7.8

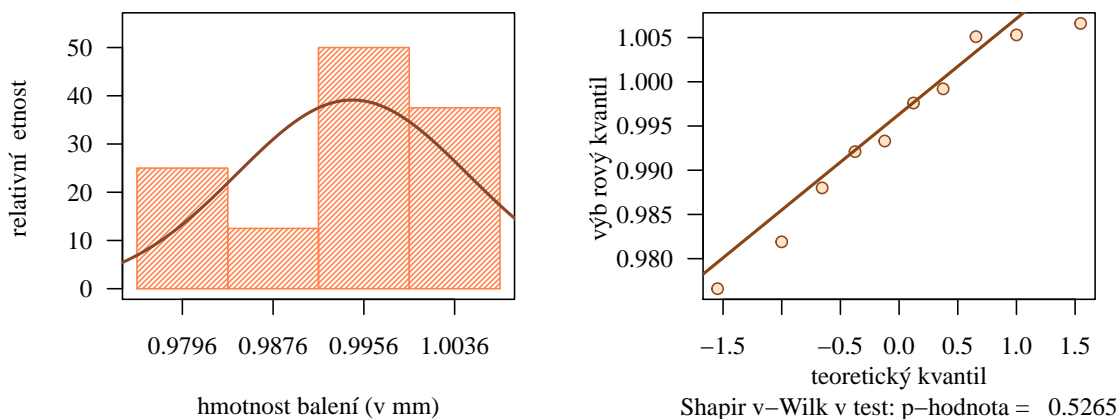
Do proměnné `weight` si vložíme údaje z tabulky 1.

```
150 weight <- c(1.0051, 0.9880, 0.9921, 0.9933, 0.9766,  
151           0.9976, 0.9819, 0.9992, 1.0053, 1.0066)  
152 n <- length(weight) # 10
```

K dispozici máme údaje o hmotnostech deseti balení hydroxidu sodného, jejichž hodnoty se pohybují v rozmezí 0.9766–1.0066 g.

Naším úkolem je porovnat hmotnost balíků hydroxidu sodného s deklarovanou hmotností 1 kg, což vede na test o střední hodnotě  $\mu$ . Abychom mohli tento test použít, musíme nejprve ověřit normalitu náhodného výběru.

Vzhledem k nízkému rozsahu náhodného výběru použijeme na test o normalitě náhodného výběru Shapirův-Wilkův test ( $\alpha = 0.05$ ). Dále vykreslíme histogram a QQ-diagram (obrázek 8). Datový soubor rozdělíme do čtyř ekvidistantních intervalů s šířkou 0.008 g prostřednictvím stanovených hranic 0.9756, 0.9836, ..., 1.0076.



Obrázek 8: Histogram a QQ-diagram hmotnosti balení hydroxidu sodného

Protože  $p$ -hodnota Shapirova-Wilkova testu, tj. 0.5265, je větší než 0.05, hypotézu o normalitě náhodného výběru nezamítáme na hladině významnosti  $\alpha = 0.05$ . Při pohledu na histogram bychom o normalitě náhodného výběru spíše pochybovali. Nezapomínejme však, že při takto malém rozsahu náhodného výběru může být tvar histogramu zavádějící. Při pohledu na QQ-diagram jsme optimističtější. Vzhledem k malému počtu hodnot se body drží dostatečně blízko referenční přímky. Náhodný výběr naměřených hmotností balení hydroxidu sodného pochází z normálního rozdělení.

To, že náhodný výběr pochází z normálního rozdělení, je v souladu s předpokladem normálního rozdělení náhodné veličiny  $X$ . Proto budeme předpokládat, že toho rozdělení má deklarovaný rozptyl  $\sigma^2 = 0.01$ . Hypotézu ze zadání

potom otestujeme pomocí parametrického testu o střední hodnotě  $\mu$ , když rozptyl  $\sigma^2$  známe. Naším úkolem je otestovat, zda je hmotnost jednoho balení hydroxidu sodného rovna deklarovanému 1 kg. Tato věta je zněním nulové hypotézy, neboť shoda implikuje rovnost a rovnost je vždy součástí nulové hypotézy. Alternativní hypotézu potom stanovíme jako doplněk k nulové hypotéze. Testování provedeme v posloupnosti sedmi kroků.

## 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Střední hodnota hmotnosti jednoho balení hydroxidu sodného je shodná s deklarovanou střední hodnotou jednoho balení hydroxidu sodného.

$H_1$  : Střední hodnota hmotnosti jednoho balení hydroxidu sodného není shodná s deklarovanou střední hodnotou jednoho balení hydroxidu sodného.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $\mu = \mu_0$ , kde  $\mu_0 = 1$

$H_1$  :  $\mu \neq \mu_0$ , kde  $\mu_0 = 1$  (oboustranná alternativa)

## 2. Volba hladiny významnosti

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

## 3. Testování kritickým oborem

- **Testovací statistika**

$$T_W = \frac{M - \mu_0}{\sigma} \sqrt{n} = \frac{0.99457 - 1}{0.01} \sqrt{10} = \frac{-0.00543}{0.01} \times 3.162278 = -1.717117 \doteq -1.7171$$

```
153 mu0 <- 1
154 sigma <- 0.01
155 alpha <- 0.05
156 n <- length(weight)
157 m <- mean(weight)
158 zw <- (m - mu0) / sigma * sqrt(n) # -1.717117
```

- **Kritický obor**

$$\begin{aligned} W &= (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty) \\ &= (-\infty; u_{0.05/2}) \cup (u_{1-0.05/2}; \infty) \\ &= (-\infty; -1.959964) \cup (1.959964; \infty) \end{aligned}$$

```
159 qnorm(alpha / 2) # -1.959964
160 qnorm(1 - alpha / 2) # 1.959964
```

- **Závěr testování**

Protože realizace testovací statistiky  $z_W = -1.7171$  nenáleží do kritického oboru, tj.  $z_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (d, h) &= \left( m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, m - \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \right) \\
 &= \left( 0.99457 - \frac{0.01}{\sqrt{10}} u_{1-0.05/2}, 0.99457 - \frac{0.01}{\sqrt{10}} u_{0.05/2} \right) \\
 &= \left( 0.99457 - \frac{0.01}{3.162278} u_{0.975}, 0.99457 - \frac{0.01}{3.162278} u_{0.025} \right) \\
 &= (0.99457 - 0.003162277 \times 1.959964, 0.99457 - 0.003162277 \times (-1.959964)) \\
 &= (0.9883721, 1.000768)
 \end{aligned}$$

```

161 dh <- m - sigma / sqrt(n) * qnorm(1 - alpha / 2) # 0.988372
162 hh <- m - sigma / sqrt(n) * qnorm(alpha / 2) # 1.000768

```

- Závěr testování

Protože  $\mu_0 = 1$  náleží do Waldova 95% empirického oboustranného intervalu spolehlivosti, tj.  $\mu_0 = 1 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned}
 p\text{-hodnota} &= 2 \min\{\Pr(Z_W \leq z_W), \Pr(Z_W > z_W)\} \\
 &= 2 \min\{\Pr(Z_W \leq z_W), 1 - \Pr(Z_W \leq z_W)\} \\
 &= 2 \min\{\Pr(Z_W \leq -1.717117), 1 - \Pr(Z_W \leq -1.717117)\} \\
 &= 2 \min\{0.04297892, 0.9570211\} \\
 &= 2 \times 0.04297892 = 0.08595784 \doteq 0.08596
 \end{aligned}$$

```

163 p.hodnota <- 2 * min(pnorm(zw), 1 - pnorm(zw)) # 0.08595784

```

- Závěr testování

Protože  $p$ -hodnota = 0.08596 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 skutečnou a deklarovanou hmotností balení hydroxidu sodného neexistuje statisticky významný rozdíl.

## 7. Grafická vizualizace výsledku testování

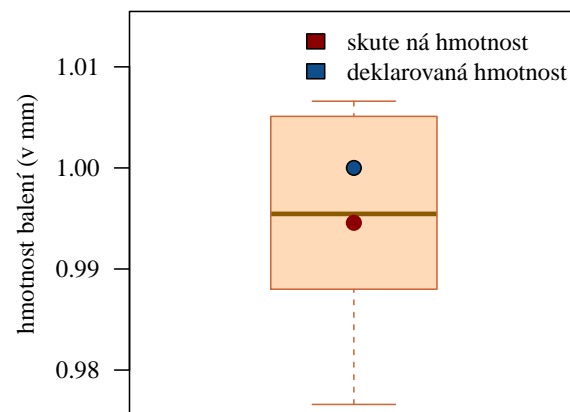
Porovnání náhodného výběru s konstantou  $\mu_0 = 1$  zobrazíme pomocí krabicového diagramu (viz obrázek 9).

```

164 par(mar = c(2, 4, 1, 1), family = 'Times')
165 boxplot(weight, col = 'peachpuff1', ylim = c(0.977, 1.014),
166         xlab = '', ylab = 'hmotnost balení (v mm)', las = 1,
167         medcol = 'orange4', border = 'sienna3')
168 points(mean(weight), bg = 'darkred', pch = 21, cex = 1.3, col = 'darkred')
169 points(1, bg = 'dodgerblue4', pch = 21, cex = 1.3, col = 'black')
170 legend('topright', fill = c('darkred', 'dodgerblue4'), bty = 'n',
171       legend = c('skutečná hmotnost', 'deklarovaná hmotnost'))

```





Obrázek 9: Krabicový diagram hmotnosti balení hydroxidu sodného

### 7.3 Test o parametru $\mu$ když $\sigma^2$ neznáme

Nechť  $X_1, \dots, X_n$  je náhodný výběr z  $N(\mu, \sigma^2)$ , kde  $\sigma^2$  neznáme a  $\mu_0$  je konstanta. Na hladině významnosti  $\alpha = 0.05$  testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

$H_{01} : \mu = \mu_0$	oproti	$H_{11} : \mu \neq \mu_0$	(oboustranná alt.)
$H_{02} : \mu \leq \mu_0$	oproti	$H_{12} : \mu > \mu_0$	(pravostranná alt.)
$H_{03} : \mu \geq \mu_0$	oproti	$H_{13} : \mu < \mu_0$	(levostranná alt.)

Test nazýváme jednovýběrový  $t$ -test. Testovací statistika má tvar

$$T_W = \frac{M - \mu_0}{S} \sqrt{n}, \quad (7.3)$$

kde  $M$  je výběrový průměr,  $S$  je výběrová směrodatná odchylka,  $n$  je rozsah náhodného výběru a  $\mu_0$  je konstanta z nulové hypotézy. Za platnosti nulové hypotézy pochází statistika  $T_W$  ze Studentova rozdělení o  $n - 1$  stupních volnosti, tj.

$$T_W = \frac{M - \mu_0}{S} \sqrt{n} \stackrel{H_0}{\sim} t_{n-1}.$$

Kritický obor podle zvolené alternativní hypotézy má tvar

$H_{11} : \mu \neq \mu_0$	$W = (-\infty; t_{n-1}(\alpha/2)) \cup (t_{n-1}(1 - \alpha/2); \infty)$
$H_{12} : \mu > \mu_0$	$W = (t_{n-1}(1 - \alpha); \infty)$
$H_{13} : \mu < \mu_0$	$W = (-\infty; t_{n-1}(\alpha))$

kde  $t_{n-1}(\alpha/2)$ ,  $t_{n-1}(1 - \alpha/2)$ ,  $t_{n-1}(\alpha)$  a  $t_{n-1}(1 - \alpha)$  jsou kvantily Studentova rozdělení o  $n - 1$  stupních volnosti, jejichž hodnoty získáme pomocí  $\mathbb{R}$  a implementované funkce `qt()`.

Interval spolehlivosti má podle zvolené alternativní hypotézy jeden z následujících tvarů

$H_{11} : \mu \neq \mu_0$	$(d, h) = \left( m - \frac{s}{\sqrt{n}} t_{n-1}(1 - \alpha/2), m - \frac{s}{\sqrt{n}} t_{n-1}(\alpha/2) \right)$
$H_{12} : \mu > \mu_0$	$(d, \infty) = \left( m - \frac{s}{\sqrt{n}} t_{n-1}(1 - \alpha), \infty \right)$
$H_{13} : \mu < \mu_0$	$(-\infty, h) = \left( -\infty, m - \frac{s}{\sqrt{n}} t_{n-1}(\alpha) \right)$

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

$H_{11} : \mu \neq \mu_0$	$p\text{-hodnota} = 2 \min\{\Pr(T_W \leq t_W), \Pr(T_W > t_W)\}$
$H_{12} : \mu > \mu_0$	$p\text{-hodnota} = \Pr(T_W > t_W) = 1 - \Pr(T_W \leq t_W)$
$H_{13} : \mu < \mu_0$	$p\text{-hodnota} = \Pr(T_W \leq t_W)$

kde  $T_W$  je náhodná veličina,  $t_W$  je realizace testovací statistiky  $T_W$  (viz vzorec 7.3), tedy konkrétní číslo, a  $\Pr(T_W \leq t_W)$  je distribuční funkce Studentova rozdělení o  $n - 1$  stupních volnosti, jejíž hodnotu získáme pomocí  $\mathbb{R}$  a implementované funkce `pt()`.

*Poznámka:* Princip testu o střední hodnotě  $\mu$  když rozptyl  $\sigma^2$  neznáme používáme také při analýze párových dat. Blíže se této situaci věnujeme v sekci 7.4.



### Příklad 7.9. Test o střední hodnotě $\mu$ když $\sigma^2$ neznáme (oboustranná alternativa)

Mějme datový soubor 18-more-samples-variances-clavicle.txt a proměnnou cla.L popisující největší délku klíční kosti z pravé strany v mm u mužů indické populace z Amritsar (viz sekce ??) naměřené v roce 1966. Dále máme k dispozici údaje ze studie (Kaur et al.) z roku 1997, v rámci které byly měřeny délky klíční kosti z pravé strany mužů ze severoindické populace ( $m_R = 146.89$  mm,  $s_R = 9.23$  mm,  $n_R = 100$ ). Na hladině významnosti  $\alpha = 0.05$  zjistíte, zda existuje rozdíl mezi největší délkou klíční kosti mužů indické populace z Amritsaru a mužů severoindické populace.

#### Řešení příkladu 7.9

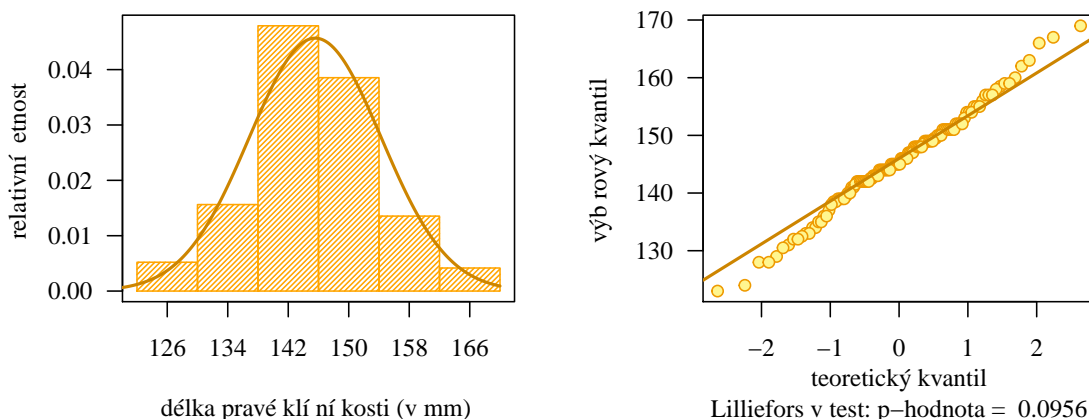
Pomocí příkazu read.delim() načteme datový soubor a pomocí operátoru [] vybereme z datové tabulky údaje o největší délce klíční kosti z pravé strany (cla.L) u jedinců indické populace z Amritsaru (pop == 'ind1'). Z vektoru naměřených hodnot odstraníme chybějící údaje (na.omit()) zjistíme rozsah náhodného výběru (length()).

```
172 data <- read.delim('00-Data//18-more-samples-variances-clavicle.txt')
173 cla.Li <- data[data$pop == 'ind1', 'cla.L']
174 cla.Li <- as.numeric(na.omit(cla.Li))
175 n <- length(cla.Li) # 120
```

Datový soubor obsahuje údaje o největší délce klíční kosti z pravé strany u 120 jedinců indické populace z Amritsaru.

Naším úkolem ze zadání je porovnat střední hodnoty dvou indických populací, přičemž u jedné populace (indická populace z Amritsaru) máme k dispozici naměřené hodnoty. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení, tj. zda náhodná veličina  $X$  popisující největší délku klíční kosti z pravé strany u mužů indické populace v Amritsaru pochází z normálního rozdělení, tj.  $X \sim N(\mu, \sigma^2)$ , kde skutečný rozptyl  $\sigma^2$  neznáme. Druhá populace (severoindická populace) je reprezentována pouze hodnotou aritmetického průměru ( $m_R = 146.89$  mm) a směrodatnou odchylkou ( $s_R = 9.23$  mm). O jejím rozdělení žádné další informace nemáme. Řešení příkladu tedy vede na situaci, kdy střední hodnotu jednoho náhodného výběru (jehož skutečnou hodnotu rozptylu neznáme) porovnáváme s konkrétním číslem, tedy na jednovýběrový test o střední hodnotě  $\mu$  při neznámém rozptylu  $\sigma^2$ . Jediným předpokladem k použití tohoto testu je normalita náhodného výběru naměřených délek klíčních kostí. Před použitím testu tedy tento předpoklad ověříme.

Hladinu významnosti  $\alpha$  pro test normality stanovíme standartně, tj.  $\alpha = 0.05$ . Protože rozsah náhodného výběru je větší než 30, ověříme předpoklad normality Lillieforsovým testem. Grafické ověření provedeme na základě QQ-diagramu a histogramu superponovaného křivkou normálního rozdělení, jejíž parametry odhadneme pomocí výběrového průměru a výběrového rozptylu. Datový soubor rozdělíme do osmi ekvidistatních intervalů s šířkou 8 mm prostřednictvím stanovených hranic 122, 130, ..., 170.



Obrázek 10: Histogram a QQ-diagram délky pravé klíční kosti u mužů indické populace

Protože  $p$ -hodnota = 0.0956 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti

$\alpha = 0.05$ . Náhodný výběr největších délek klíčních kostí z pravé strany u jedinců indické populace z Amritsaru pochází z normálního rozdělení.

Protože náhodný výběr pochází z normálního rozdělení, můžeme hypotézu ze zadání otestovat pomocí parametrického testu o střední hodnotě  $\mu$  když rozptyl  $\sigma^2$  neznáme. Testování provedeme v posloupnosti sedmi kroků. Naším úkolem je zjistit, zda existuje rozdíl mezi největší délkou klíční kosti z pravé strany u mužů indické populace z Amritsaru a u mužů severoindické populace. Tato věta bude součástí alternativní hypotézy, neboť rozdíl implikuje nerovnost a nerovnost je vždy součástí alternativní hypotézy. Nulovou hypotézu potom stanovíme jako doplněk k alternativní hypotéze.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Střední hodnota největší délky klíční kosti na pravé straně mužů indické populace z Amritsaru je shodná se střední hodnotou největší délky klíční kosti na pravé straně mužů severoindické populace.

$H_1$  : Střední hodnota největší délky klíční kosti na pravé straně mužů indické populace z Amritsaru není shodná se střední hodnotou největší délky klíční kosti na pravé straně mužů severoindické populace.

- **matematická formulace** nulové a alternativní hypotézy

$H_0 : \mu = \mu_0$ , kde  $\mu_0 = 146.89$

$H_1 : \mu \neq \mu_0$ , kde  $\mu_0 = 146.89$  (oboustranná alternativa)

### 2. Volba hladiny významnosti

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

### 3. Testování kritickým oborem

- **Testovací statistika**

$$T_W = \frac{M - \mu_0}{S} \sqrt{n} = \frac{145.5667 - 146.89}{8.733432} \sqrt{120} = \frac{-1.3233}{8.733432} \times 10.95445 = -1.659831 \doteq -1.6598$$

```
176 alpha <- 0.05
177 mu_0 <- 146.89
178 m <- mean(c1a.Li)
179 s <- sd(c1a.Li)
180 tw <- (m - mu_0) / s * sqrt(n) # -1.659873
```

- **Kritický obor**

$$\begin{aligned} W &= (-\infty; t_{n-1}(\alpha/2)) \cup (t_{n-1}(1 - \alpha/2); \infty) \\ &= (-\infty; t_{120-1}(0.05/2)) \cup (t_{120-1}(1 - 0.05/2); \infty) \\ &= (-\infty; t_{119}(0.025)) \cup (t_{119}(0.975); \infty) \\ &= (-\infty; -1.9801) \cup (1.9801; \infty) \end{aligned}$$

```
181 qt(alpha/2, n - 1) # -1.9801
182 qt(1 - alpha/2, n - 1) # 1.9801
```

- **Závěr testování**

Protože realizace testovací statistiky  $t_W = -1.6598$  nenáleží do kritického oboru, tj.  $t_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

### 4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

$$\begin{aligned}
 (d, h) &= \left( m - \frac{s}{\sqrt{n}} t_{n-1}(1 - \alpha/2), m - \frac{s}{\sqrt{n}} t_{n-1}(\alpha/2) \right) \\
 &= \left( 145.5667 - \frac{8.733432}{\sqrt{120}} t_{120-1}(1 - 0.05/2), 145.5667 - \frac{8.733432}{\sqrt{120}} t_{120-1}(0.05/2) \right) \\
 &= \left( 145.5667 - \frac{8.733432}{10.95445} t_{119}(0.975), 145.5667 - \frac{8.733432}{10.95445} t_{119}(0.025) \right) \\
 &= (145.5667 - 0.7972497 \times 1.9801, 145.5667 - 0.7972497 \times (-1.9801)) \\
 &= (143.9881, 147.1453)
 \end{aligned}$$

```

183 dh <- m - s / sqrt(n) * qt(1 - alpha / 2, n - 1) # 143.988
184 hh <- m - s / sqrt(n) * qt(alpha / 2, n - 1) # 147.1453

```

- **Závěr testování**

Protože  $\mu_0 = 146.89$  náleží do Waldova 95% empirického oboustranného intervalu spolehlivosti, tj.  $\mu_0 = 146.89 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- **$p$ -hodnota**

$$\begin{aligned}
 p\text{-hodnota} &= 2 \min\{\Pr(T_W \leq t_W), \Pr(T_W > t_W)\} \\
 &= 2 \min\{\Pr(T_W \leq t_W), 1 - \Pr(T_W \leq t_W)\} \\
 &= 2 \min\{\Pr(T_W \leq 146.89), 1 - \Pr(T_W \leq 146.89)\} \\
 &= 2 \min\{0.04978648, 0.9502135\} \\
 &= 2 \times 0.04978648 = 0.09957296 \doteq 0.09957
 \end{aligned}$$

```

185 p.hodnota <- 2 * min(pt(tw, n - 1), 1 - pt(tw, n - 1)) # 0.09957296

```

- **Závěr testování**

Protože  $p$ -hodnota = 0.09957 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 střední hodnotou největší délky klíční kosti na pravé straně u mužů indické populace z Amritsaru a u mužů severoindické populace neexistuje statisticky významný rozdíl.

## 7. Grafická vizualizace výsledku testování

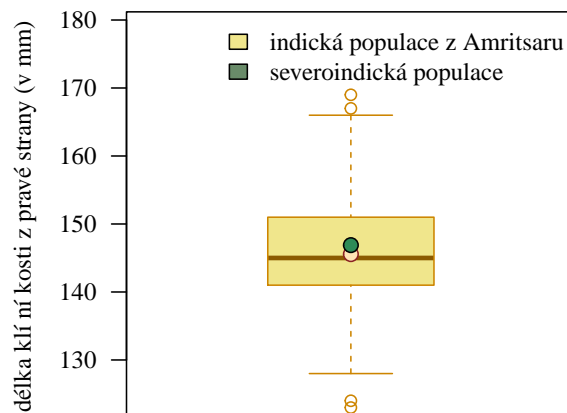
Porovnání náhodného výběru s konstantou  $\mu_0 = 146.89$  zobrazíme nejlépe pomocí krabicového diagramu (viz obrázek 11).

*Poznámka:* Test o střední hodnotě  $\mu$  při neznámém rozptylu  $\sigma^2$  můžeme provést pomocí funkce `t.test()`. Vstupními parametry budou vektor reprezentující náhodný výběr (`cla.Li`), hodnota parametru  $\mu_0$  z nulové hypotézy zadaná argumentem `mu = 146.89`, hodnota hladiny významnosti  $\alpha$  zadaná prostřednictvím koeficientu spolehlivosti  $1 - \alpha$  nastavením hodnoty argumentu `conf.level = 0.95` a typ zvolené alternativní hypotézy (oboustranná) zadaný pomocí argumentu `alternative = 'two.sided'`.

```

186 t.test(cla.Li, mu = 146.89, conf.level = 0.95, alternative = 'two.sided')

```



Obrázek 11: Krabicový diagram délky klíční kosti z pravé strany u mužů indické populace z Amritsaru a u mužů severoindické populace

```

One Sample t-test

data: cla.Li
t = -1.6599, df = 119, p-value = 0.09957
alternative hypothesis: true mean is not equal to 146.89
95 percent confidence interval:
 143.9880 147.1453
sample estimates:
mean of x
 145.5667

```

187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197

Součástí výstupu je hodnota testovací statistiky  $t = -1.6599$ , počet stupňů volnosti Studentova rozdělení  $df = 119$ , hranice 95% Waldova empirického oboustranného intervalu spolehlivosti 143.9880 a 147.1453 a  $p$ -hodnota  $p$ -value = 0.09957. Jediné, co musíme stanovit zvlášť, jsou dolní a horní hranice kritického oboru. ★

### Příklad 7.10. Test o střední hodnotě $\mu$ když $\sigma^2$ neznáme (pravostranná alternativa)

Mějme datový soubor 01-one-sample-mean-skull-mf.txt a proměnnou skull.B popisující největší šířku mozkovny mužů starověké egyptské populace (viz sekce ??). Dále máme k dispozici údaje o největší šířce mozkovny novověké egyptské mužské populace ( $m_m = 136.402$  mm,  $s_m = 6.411$  mm,  $n_m = 87$ ). Na hladině významnosti  $\alpha = 0.01$  zjistěte, zda je šířka mozkovny starověké egyptské mužské populace větší než šířka mozkovny novověké egyptské mužské populace.

#### Řešení příkladu 7.10

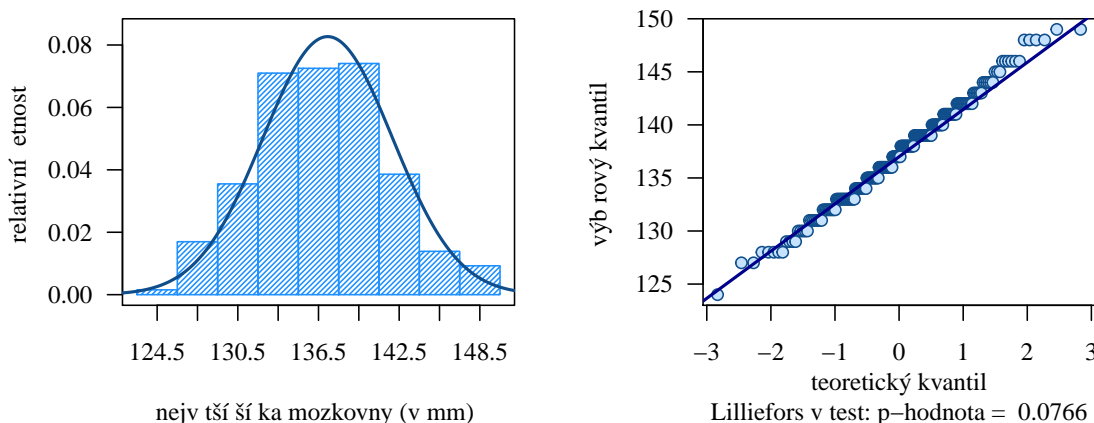
Pomocí příkazu read.delim() načteme datový soubor a operátorem [] vybereme z datové tabulky údaje o největší šířce mozkovny (skull.B) mužů (sex == 'm'). Z vektoru naměřených údajů odstraníme příkazem na.omit() NA hodnoty a příkazem length() zjistíme rozsah náhodného výběru.

```
198 data <- read.delim('00-Data//01-one-sample-mean-skull-mf.txt')
199 skull.BM <- data[data$sex == 'm', 'skull.B']
200 skull.BM <- as.numeric(na.omit(skull.BM))
201 n <- length(skull.BM) # 216
```

Datový soubor obsahuje údaje o největší šířce mozkovny u 216 mužů starověké egyptské populace.

Naším úkolem ze zadání je porovnat šířku mozkovny mužů novověké a starověké egyptské populace. U starověké populace máme k dispozici naměřené hodnoty, pomocí kterých můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení, tj. zda náhodná veličina  $X$  popisující největší šířku mozkovny mužů starověké egyptské populace pochází z normálního rozdělení, tj.  $X \sim N(\mu, \sigma^2)$ , kde skutečný rozptyl  $\sigma^2$  neznáme. Druhá, novověká populace je reprezentována pouze aritmetickým průměrem ( $m_m = 136.402$  mm) a směrodatnou odchylkou ( $s_m = 6.411$  mm). O jejím rozdělení žádné další informace nemáme. Řešení příkladu vede na jednovýběrový test o střední hodnotě  $\mu$  při neznámém rozptylu  $\sigma^2$ . Před použitím tohoto testu musíme nejprve ověřit vyžadovaný předpoklad normálního rozdělení náhodného výběru.

Protože rozsah náhodného výběru je větší než 30, ověříme předpoklad normality Lillieforsovým testem ( $\alpha = 0.05$ ). Graficky zhodnotíme potenciální normalitu náhodného výběru QQ-diagramem a histogramem. Datový soubor rozdělíme do devíti ekvidistatních intervalů s šířkou 3 mm prostřednictvím stanovených hranic 123, 126, ..., 150.



Obrázek 12: Histogram a QQ-diagram délky pravé klíční kosti u mužů indické populace

Protože  $p$ -hodnota = 0.0766 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z obrázku 12 je patrné, že histogram naměřených hodnot dostatečným způsobem kopíruje křivku hustoty normálního rozdělení a že body v QQ-diagramu se pohybují velmi blízko referenční křivky. Závěr tedy je, že náhodný výběr největších šířek mozkovny mužů starověké egyptské populace pochází z normálního rozdělení.

Protože náhodný výběr pochází z normálního rozdělení, můžeme hypotézu ze zadání otestovat pomocí paramet-

rického testu o střední hodnotě  $\mu$  když rozptyl  $\sigma^2$  neznáme. Naším úkolem je zjistit, zda je šířka mozkovny starověké egyptské mužské populace větší než šířka mozkovny novověké egyptské mužské populace. Tato věta bude součástí alternativní hypotézy, nulová hypotéza bude potom doplňkem k alternativní hypotéze.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Střední hodnota největší šířky mozkovny starověké egyptské mužské populace je menší nebo rovna střední hodnotě šířky mozkovny novověké egyptské mužské populace.

$H_1$  : Střední hodnota největší šířky mozkovny starověké egyptské mužské populace je větší než střední hodnota šířky mozkovny novověké egyptské mužské populace.

- **matematická formulace** nulové a alternativní hypotézy

$H_0 : \mu \leq \mu_0$ , kde  $\mu_0 = 171.962$

$H_1 : \mu > \mu_0$ , kde  $\mu_0 = 171.962$  (pravostranná alternativa)

### 2. Volba hladiny významnosti

- Hladinu významnosti volíme podle zadání  $\alpha = 0.01$ .

### 3. Testování kritickým oborem

- **Testovací statistika**

$$T_W = \frac{M - \mu_0}{S} \sqrt{n} = \frac{137.1852 - 136.402}{4.824642} \sqrt{216} = \frac{0.7832}{4.824642} \times 14.69694 = 2.385803 \doteq 2.3858$$

```
202 alpha <- 0.01
203 mu_0 <- 136.402
204 m <- mean(skull.BM)
205 s <- sd(skull.BM)
206 tw <- (m - mu_0) / s * sqrt(n) # 2.385757
```

- **Kritický obor**

$$\begin{aligned} W &= (t_{n-1}(1 - \alpha); \infty) \\ &= (t_{216-1}(1 - 0.01); \infty) \\ &= (t_{215}(0.99); \infty) \\ &= (1.6520; \infty) \end{aligned}$$

```
207 qt(1 - alpha, n - 1) # 2.343817
```

- **Závěr testování**

Protože realizace testovací statistiky  $t_W = 2.385757$  náleží do kritického oboru, tj.  $t_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (d, \text{infity}) &= \left( m - \frac{s}{\sqrt{n}} t_{n-1}(1 - \alpha/2), \infty \right) \\
 &= \left( 137.1852 - \frac{4.824642}{\sqrt{216}} t_{216-1}(1 - 0.01/2), \infty \right) \\
 &= \left( 137.1852 - \frac{4.824642}{14.69694} t_{215}(0.995), \infty \right) \\
 &= (137.1852 - 0.3282753 \times 2.343817, \infty) \\
 &= (136.4158, \infty)
 \end{aligned}$$

```
208 dh <- m - s / sqrt(n) * qt(1 - alpha, n - 1) # 136.4158
```

- Závěr testování

Protože  $\mu_0 = 136.402$  nenáleží do Waldova 99% empirického levostranného intervalu spolehlivosti, tj.  $\mu_0 = 136.402 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.09$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$p\text{-hodnota} = \Pr(T_W > t_w) = 1 - \Pr(T_W \leq t_w) = 1 - \Pr(T_W \leq 136.402) \doteq 0.008956$$

```
209 p.hodnota <- 1 - pt(tw, n - 1) # 0.008955785
```

- Závěr testování

Protože  $p$ -hodnota = 0.008956 je menší než  $\alpha = 0.01$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

## 6. Interpretace výsledků

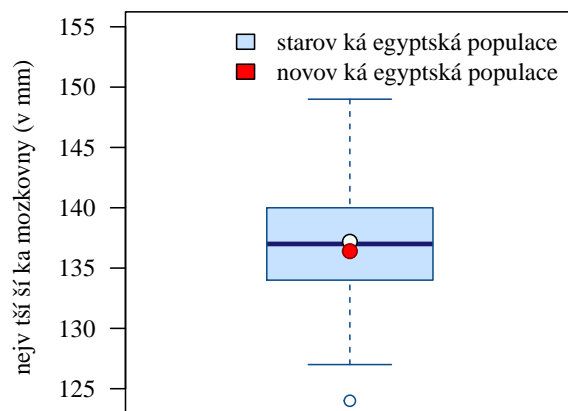
Za základě všech tří typů testování zamítáme nulovou hypotézu na hladině významnosti  $\alpha = 0.01$ . Největší šířka lebky starověké egyptské mužské populace je statisticky významně větší než největší šířka lebky novověké egyptské mužské populace.

## 7. Grafická vizualizace výsledku testování

Rozdíl mezi starověkou a novověkou egyptskou mužskou populací zobrazíme krabicovým diagramem (viz obrázek 13).

```
210 par(mar = c(2, 4, 1, 1), family = 'Times')
211 boxplot(skull.BM, col = 'slategray1', xlab = '',
212         xlab = '', ylab = 'největší šířka mozkovny (v mm)', las = 1,
213         ylim = c(124, 155), medcol = 'midnightblue', border = 'dodgerblue4')
214 points(mean(skull.BM), bg = 'mintcream', pch = 21, cex = 1.3, col = 'black')
215 points(136.402, bg = 'red', pch = 21, cex = 1.3, col = 'red4')
216 legend('topright', fill = c('slategray1', 'red'), bty = 'n',
217        legend = c('starověká egyptská populace', 'novověká egyptská populace'))
```

*Poznámka:* Test o střední hodnotě  $\mu$  při neznámém rozptylu  $\sigma^2$  můžeme provést pomocí funkce `t.test()`. Vstupními parametry budou vektor reprezentující náhodný výběr (`skull.BM`), hodnota parametru  $\mu_0$  z nulové hypotézy (`mu = 136.402`), hodnota hladiny významnosti  $\alpha$  zadaná prostřednictvím koeficientu spolehlivosti  $1 - \alpha$  (`conf.level = 0.99`) a typ zvolené alternativní hypotézy (`alternative = 'greater'`).



Obrázek 13: Krabicový diagram největší šířky mozkovny mužů starověké a novověké egyptské populace

```

One Sample t-test

data:  skull.BM
t = 2.3858, df = 215, p-value = 0.008956
alternative hypothesis: true mean is greater than 136.402
99 percent confidence interval:
 136.4158      Inf
sample estimates:
mean of x
 137.1852

```

219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229

```
218 t.test(skull.BM, mu = 136.402, conf.level = 0.99, alternative = 'greater')
```

Součástí výstupu je hodnota testovací statistiky  $t = 2.3858$ , počet stupňů volnosti Studentova rozdělení  $df = 215$ , hranice 99% Waldova empirického oboustranného intervalu spolehlivosti 136.4158 a  $\lnf$  a  $p$ -hodnota  $p\text{-value} = 0.008956$ . Jediné, co musíme stanovit zvlášť, je dolní hranice kritického oboru. ★



### Příklad 7.11. Test o střední hodnotě $\mu$ když $\sigma^2$ neznáme (levostranná alternativa)

Mějme datový soubor 11-two-samples-means-skull.txt a proměnnou skull.H popisující basion–bregmatickou výšku lebky žen starověké egyptské populace (viz sekce ??). Dále máme k dispozici údaje o basion–bregmatické výšce lebky žen novověké egyptské populace ( $m_f = 126.942$  mm,  $s_f = 4.430$  mm,  $n_f = 52$ ). Na hladině významnosti  $\alpha = 0.05$  zjistíte, zda je basion–bregmatická výška lebky žen starověké egyptské populace je menší než basion–bregmatická výška lebky žen novověké egyptské populace.

#### Řešení příkladu 7.11

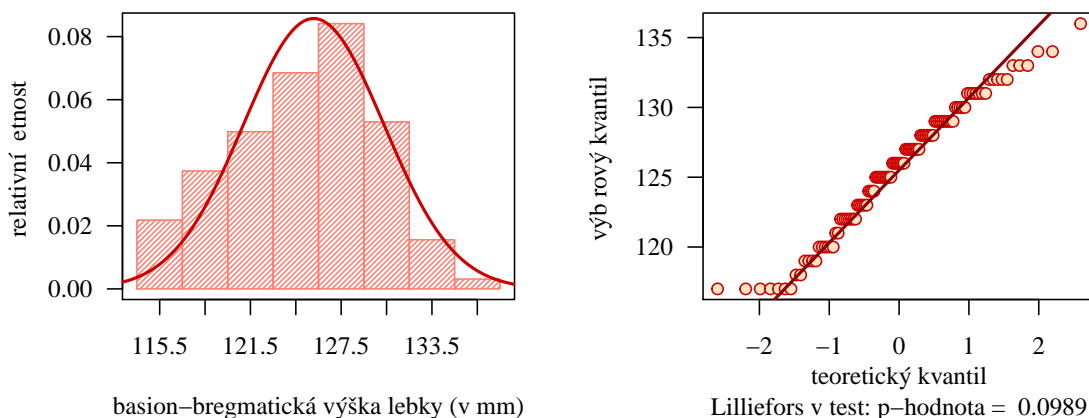
Nejprve načteme datový soubor a vybereme z datové tabulky údaje o basion–bregmatické výšce lebky (skull.H) žen (sex == 'f'). Z vektoru naměřených hodnot odstraníme chybějící údaje a zjistíme počet naměřených hodnot.

```
230 data <- read.delim('00-Data//11-two-samples-means-skull.txt')
231 skull.HF <- data[data$sex == 'f', 'skull.H']
232 skull.HF <- as.numeric(na.omit(skull.HF))
233 n <- length(skull.HF) # 107
```

Datový soubor obsahuje naměřené hodnoty basion–bregmatické výšky lebky 107 žen starověké egyptské populace.

Naším úkolem je porovnat výšku lebky žen dvou egyptských populací. U starověké populace máme k dispozici naměřené hodnoty, pomocí kterých můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení, kde skutečný rozptyl  $\sigma^2$  neznáme. Druhá, novověká populace je reprezentována pouze aritmetickým průměrem ( $m_f = 126.942$  mm) a směrodatnou odchylkou ( $s_f = 4.430$  mm). O jejím rozdělení žádné další údaje nemáme. Řešení příkladu vede na jednovýběrový test o střední hodnotě  $\mu$  při neznámém rozptylu  $\sigma^2$ . Před použitím tohoto testu musíme ověřit předpoklad normality náhodného výběru.

Protože rozsah náhodného výběru je větší než 30, ověříme normální rozdělení náhodného výběru Lillieforsovým testem ( $\alpha = 0.05$ ). Graficky zhodnotíme rozdělení náhodného výběru QQ-diagramem a histogramem (viz obrázek 14). Datový soubor rozdělíme do osmi ekvidistančních intervalů s šířkou 3 mm prostřednictvím stanovených hranic 114, 117, ..., 138.



Obrázek 14: Histogram a diagram basion–bregmatické výšky lebky žen starověké a novověké egyptské populace

Protože  $p$ -hodnota = 0.0989 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z obrázku 14 vidíme, že histogram naměřených hodnot je oproti křivce hustoty normálního rozdělení mírně posunutý doprava. Taktéž v QQ-diagramu vidíme, že body umístěné na koncích se vzdalují od referenční křivky. Naštěstí tyto odchylky nejsou tak závažné, aby fatálně narušily normální rozdělení datového souboru. Náhodný výběr basion–bregmatické výšky lebky žen starověké egyptské populace tedy pochází z normálního rozdělení.

Protože náhodný výběr naměřených hodnot splňuje předpoklad normality, můžeme otázku ze zadání ověřit pomocí parametrického testu o střední hodnotě  $\mu$  když rozptyl  $\sigma^2$  neznáme. V zadání příkladu máme tentokrát přímo uve-

dené znění alternativní hypotézy. Zbývá tedy dodefinovat nulovou hypotézu tak, aby byla doplňkem k alternativní hypotéze.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Střední hodnota basion–bregmatické výšky lebky žen starověké egyptské populace je větší nebo rovna střední hodnotě basion–bregmatické výšky lebky žen novověké egyptské populace.

$H_1$  : Střední hodnota basion–bregmatické výšky lebky žen starověké egyptské populace je menší než střední hodnota basion–bregmatické výšky lebky žen novověké egyptské populace.

- **matematická formulace** nulové a alternativní hypotézy

$H_0 : \mu \geq \mu_0$ , kde  $\mu_0 = 126.942$

$H_1 : \mu < \mu_0$ , kde  $\mu_0 = 126.942$  (levostranná alternativa)

### 2. Volba hladiny významnosti

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

### 3. Testování kritickým oborem

- **Testovací statistika**

$$T_W = \frac{M - \mu_0}{S} \sqrt{n} = \frac{125.6822 - 126.942}{4.653256} \sqrt{107} = \frac{-1.2598}{4.653256} \times 10.34408 = -2.800506 \doteq -2.8005$$

```
234 alpha <- 0.05
235 mu_0 <- 126.942
236 m <- mean(skull.HF)
237 s <- sd(skull.HF)
238 tw <- (m - mu_0) / s * sqrt(n) # -2.80041
```

- **Kritický obor**

$$\begin{aligned} W &= (-\infty; t_{n-1}(\alpha)) \\ &= (-\infty; t_{107-1}(0.05)) \\ &= (-\infty; -1.659356) \end{aligned}$$

```
239 qt(alpha, n - 1) # -1.659356
```

- **Závěr testování**

Protože realizace testovací statistiky  $t_W = -2.8005$  náleží do kritického oboru, tj.  $t_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (-\infty; h) &= \left( -\infty, m - \frac{s}{\sqrt{n}} t_{n-1}(\alpha) \right) \\
 &= \left( -\infty, 125.6822 - \frac{4.653256}{\sqrt{107}} t_{107-1}(0.05) \right) \\
 &= \left( -\infty, 125.6822 - \frac{4.653256}{10.34408} t_{106}(0.05) \right) \\
 &= (-\infty, 125.6822 - 0.4498473 \times (-1.659356)) \\
 &= (-\infty, 126.4287)
 \end{aligned}$$

```
240 hh <- m - s / sqrt(n) * qt(alpha, n - 1) # 126.4287
```

- Závěr testování

Protože  $\mu_0 = 126.942$  nenáleží do Waldova 95% empirického pravostranného intervalu spolehlivosti, tj.  $\mu_0 = 126.942 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$p\text{-hodnota} = \Pr(T_W \leq t_W) = \Pr(T_W \leq -2.80041) = 0.003033381 \doteq 0.003033$$

```
241 p.hodnota <- pt(tw, n - 1) # 0.003033381
```

- Závěr testování

Protože  $p$ -hodnota = 0.003033 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  $\alpha = 0.05$ . Basion-bregmatická výška lebky žen starověké egyptské populace je statisticky významně menší než basion-bregmatická výška lebky žen novověké egyptské populace.

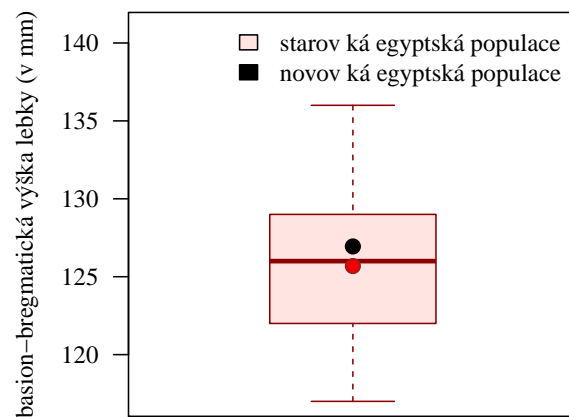
## 7. Grafická vizualizace výsledku testování

Rozdíl středních hodnot basion-bregmatické výšky lebky žen obou populací zobrazíme pomocí krabicového diagramu (viz obrázek 15).

*Poznámka:* Test o střední hodnotě  $\mu$  při neznámém rozptylu  $\sigma^2$  můžeme provést pomocí funkce `t.test()` s argumenty `mu = 126.942`, `conf.level = 0.95` a `alternative = 'less'`.

```
242 t.test(skull.HF, mu = 126.942, conf.level = 0.95, alternative = 'less')
```

Součástí výstupu je hodnota testovací statistiky  $t = -2.8004$ , počet stupňů volnosti Studentova rozdělení  $df = 106$ , hranice 95% Waldova empirického oboustranného intervalu spolehlivosti  $-\lnf$  a 126.4287 a  $p$ -hodnota `p-value = 0.003033`. Jediné, co musíme stanovit zvlášť, je horní hranice kritického oboru. ★



Obrázek 15: Krabicový diagram basion-bregmatické výšky lebky žen starověké a novověké egyptské populace

One Sample t-test	243
	244
data: skull.HF	245
t = -2.8004, df = 106, p-value = 0.003033	246
alternative hypothesis: true mean is less than 126.942	247
95 percent confidence interval:	248
-Inf 126.4287	249
sample estimates:	250
mean of x	251
125.6822	252
	253

### Příklad 7.12. Test o střední hodnotě $\mu$ když $\sigma^2$ neznáme (nesplnění předpokladů)

Mějme datový soubor 15-anova-means-skull.txt a proměnnou upface.H popisující výšku horní části tváře mužů německé populace (viz sekce ??). Dále máme k dispozici údaje o výšce horní části tváře mužů Černjachovské kultury z území dnešní Ukrajiny ( $m_{ck} = 70.00$  mm,  $n_{ck} = 99$ ). Na hladině významnosti  $\alpha = 0.10$  testujte hypotézu, že výška horní části tváře německé mužské populace je menší nebo rovna výšce horní části tváře mužské populace z Černjachovské kultury.

#### Řešení příkladu 7.12

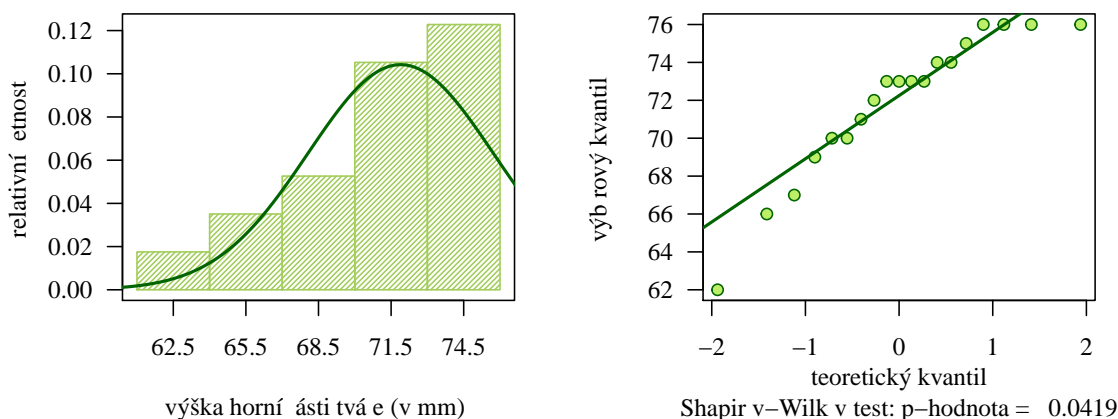
Nejprve načteme datový soubor a z datové tabulky vybereme údaje o výšce horní části tváře (upface.H) mužů německé populace (pop == 'nem'). Dále z vektoru naměřených hodnot odstraníme NA hodnoty a zjistíme rozsah náhodného výběru.

```
254 data <- read.delim('00-Data//15-anova-means-skull.txt')
255 upface.HN <- data[data$pop == 'nem', 'upface.H']
256 upface.HN <- as.numeric(na.omit(upface.HN))
257 n <- length(upface.HN) # 19
```

Datový soubor obsahuje údaje o výškách horní části tváře 19 mužů německé populace.

Naším úkolem je porovnat výšku horní části tváře u mužů ze dvou různých populací. U německé populace známe naměřené hodnoty, pomocí kterých můžeme zjistit, zda náhodný výběr pochází z normálního rozdělení s nám neznámým rozptylem  $\sigma^2$ . U mužské populace z Černjachovské kultury známe pouze aritmetickým průměr ( $m_{ck} = 70.00$  mm). Rozdělení náhodného výběru, na základě kterého byly publikované statistiky získány, nám není známo. Řešení příkladu vede na jednovýběrový test o střední hodnotě  $\mu$  při neznámém rozptylu  $\sigma^2$ . Před použitím tohoto testu ověříme předpoklad normality náhodného výběru.

Protože rozsah náhodného výběru výšek horní části tváře mužů německé populace je menší než 30, ověříme normalitu tohoto náhodného výběru Shapirovým-Wilkovým testem ( $\alpha = 0.05$ ). Graficky zhodnotíme rozdělení náhodného výběru QQ-diagramem a histogramem (viz obrázek 16). Datový soubor rozdělíme do pěti ekvidistatních intervalů s šířkou 3 mm prostřednictvím stanovených hranic 61, 64, ..., 76.



Obrázek 16: Histogram a diagram výšky horní části tváře mužů německé populace

Protože  $p$ -hodnota = 0.0419 je menší než 0.05, nulovou hypotézu o normalitě dat zamítáme na hladině významnosti  $\alpha = 0.05$ . Z obrázku 16 vidíme, že histogram naměřených hodnot nekopíruje tvar křivky hustoty normálního rozdělení. Odchylka nastává zejména na pravém konci, kde oproti očekávání nedochází ke snížení počtu hodnot. QQ-diagram vizualizuje odlehlost bodů od referenční přímky na pravém i na levém chvostu. Náhodný výběr výšky horní části tváře mužů německé populace tedy nepochází z normálního rozdělení.

Protože náhodný výběr naměřených výšek horní části tváře mužů německé populace nespĺňuje předpoklad normality,

nemůžeme k otestování hypotézy ze zadání použít parametrický test o střední hodnotě  $\mu$  když rozptyl  $\sigma^2$  neznáme. Hypotézu bychom otestovali vhodnou metodou z kapitoly ??.



## 7.4 Párový test

Nechť  $(X_1, Y_1)^T \dots (X_n, Y_n)^T$  je náhodný výběr z  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , kde  $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$  a  $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ , přičemž  $\mu_1$  je střední hodnota náhodné veličiny  $X$ ,  $\mu_2$  je střední hodnota náhodné veličiny  $Y$ ,  $\sigma_1^2$  je rozptyl náhodné veličiny  $X$ ,  $\sigma_2^2$  je rozptyl náhodné veličiny  $Y$  a  $\rho$  je korelační koeficient. Potom náhodný výběr  $Z = (Z_1, \dots, Z_n)^T$ , kde  $Z_i = Y_i - X_i$ ,  $i = 1, \dots, n$ , pochází z normálního rozdělení se střední hodnotou  $\mu = \mu_1 - \mu_2$  a rozptylem  $\sigma^2 = \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}$ , tj.  $Z \sim N(\mu, \sigma^2) = N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2})$ . Na hladině významnosti  $\alpha$  testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

$H_{01} : \mu = \mu_0$	oproti	$H_{11} : \mu \neq \mu_0$ (oboustranná alt.)
$H_{02} : \mu \leq \mu_0$	oproti	$H_{12} : \mu > \mu_0$ (pravostranná alt.)
$H_{03} : \mu \geq \mu_0$	oproti	$H_{13} : \mu < \mu_0$ (levostranná alt.)

kde  $\mu$  je střední hodnota rozdílů  $Z_1, \dots, Z_n$  a  $\mu_0$  je konstanta, jejíž hodnotu nejčastěji volíme jako  $\mu_0 = 0$ . Tato volba odpovídá hypotéze, že rozdíl mezi  $X$  a  $Y$  neexistuje (resp. hypotéze, že střední hodnota náhodné veličiny  $X$  je menší, resp. větší, než střední hodnota náhodné veličiny  $Y$ ). Vzhledem k tomu, že jde o test hypotézy o parametru  $\mu$ , přičemž skutečná hodnota rozptylu  $\sigma^2$  rozdílů  $Z_1, \dots, Z_n$  není známá, testujeme hypotézy o střední hodnotě rozdílů  $X - Y$  pomocí jednovýběrového  $t$ -testu, analogicky jako je uvedeno v sekci 7.3.

Výše popsaný test, v rámci kterého převádíme problém porovnávání dvou náhodných veličin  $X$  a  $Y$  na problém srovnávání jejich rozdílů  $Z$  s konstantou  $\mu_0$  a následně jej řešíme pomocí jednovýběrového  $t$ -testu, nazýváme párový Studentův  $t$ -test.

*Poznámka:* Jak již bylo zmíněno v úvodu kapitoly, předpokladem k použití parametrického párového testu je normální rozdělení rozdílů  $X - Y$ , tj. normální rozdělení náhodné veličiny  $Z$ . V praxi bývá často tento předpoklad mylně zaměňován s předpokladem normality náhodné veličiny  $X$  a předpokladem normality náhodné veličiny  $Y$ . Otestování předpokladu normality zvlášť pro každý z obou náhodných výběrů je však hrubá chyba ukazující na nepochopení základního principu párového testu.

*Poznámka:* Jak napovídá název testu, nejčastější realizace párového testu je v situacích, kdy na jednom jedinci porovnáváme sledovaný párový znak ze dvou stran. Typickým příkladem je například porovnání délkových nebo šířkových rozměrů na pravé a levé straně těla jedince (porovnání délky pravé a levé klíční kosti, porovnání míry přilehlosti ušního lalůčku na pravé a levé straně nebo porovnání 2D:4D poměru na pravé a levé ruce).

Trochu netypickým ale i tak dobře fungujícím příkladem využití párového testu může být také porovnání sledovaného znaku u sourozenců (resp. dvojčat). Sledovaný znak u mladšího a staršího sourozence (dvojčete) považujeme též za párový případ, kdy subjektem spojujícím "obě strany" párového znaku je společná matka obou sourozenců (konkrétně si uveďme kupříkladu porovnání porodní hmotnosti staršího a mladšího dvojčete, porovnání výše inteligenčního kvocientu osmiletých dvojčat, nebo porovnání výšky v 15 letech u staršího a mladšího sourozence).

V neposlední řadě používáme párový test na porovnání měření téhož znaku na týž jedincích dvěma různými výzkumníky, tedy na zjištění interindividuální chyby měření, nebo na porovnání dvou opakovaných měření téhož znaku na týchž jedincích jedním výzkumníkem, tedy na zjištění intraindividuální chyby měření. V obou případech je subjektem spojujícím "obě strany" párového testu jedinec, na kterém byla opakovaná měření (ať už jedním výzkumníkem dvakrát, nebo dvěma výzkumníky jednou) provedena.

### Příklad 7.13. Párový test (oboustranná alternativa)

Máme datový soubor 21-goldman-measures.csv obsahující údaje o délce stehenní kosti z pravé strany (femur.LR) a levé strany (femur.LL) u žen aljašské populace z kmene Ipituaq. Více informací viz sekce ???. Na hladině významnosti  $\alpha = 0.05$  testujte hypotézu o shodě střední hodnoty délky stehenní kosti na pravé a levé straně u žen z kmene Ipituaq.

#### Řešení příkladu 7.13

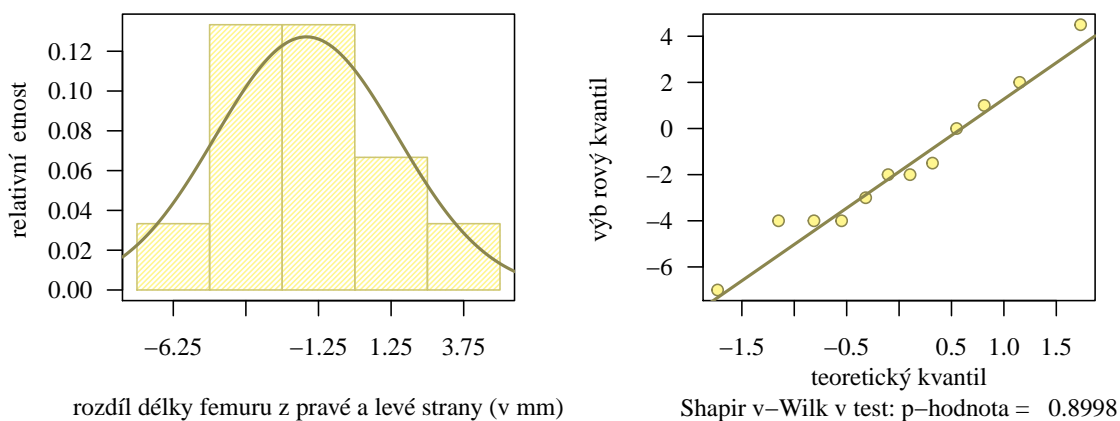
Nejprve příkazem read.delim() načteme datový soubor a pomocí operátoru [] z něj vybereme pouze údaje o délce stehenní kosti na pravé straně a levé straně (proměnné femur.LR a femur.LL) naměřené u žen sex == 'f' z kmene Ipituaq (pop == 'Ipituaq'). Údaje vložíme do proměnné data.LRF. Z datové tabulky následně odstraníme chybějící údaje a zjistíme rozsah náhodného výběru.

```
258 data <- read.delim('00-Data\\21-goldman-measures.csv', sep = ';', dec = '.')
259 data.LRF <- data[data$sex == 'f' & data$pop == 'Ipituaq', c('femur.LR', 'femur.LL')]
260 data.LRF <- na.omit(data.LRF)
261 dim(data.LRF) # 12x2
```

Datový soubor obsahuje kompletní údaje o délce stehenních kostí u 12 žen. Naším úkolem ze zadání je porovnat měření na pravé a levé straně, proto je vhodné použít na tuto situaci párový test. Prvním krokem k provedení tohoto testu je vytvoření rozdílů hodnot z levé a pravé strany. V druhém kroku je potřeba ověřit jediný předpoklad, který musí být splněn, abychom mohli párový test provést, a sice ověřit normální rozdělení těchto rozdílů.

Vzhledem k rozsahu náhodného výběru ( $n = 12 < 30$ ) použijeme k ověření normality rozdílů Shapirův-Wilkův test ( $\alpha = 0.05$ ). Graficky zhodnotíme rozdělení náhodného výběru QQ-diagramem a histogramem (viz obrázek 17). Datový soubor rozdělíme do pěti ekvidistatních intervalů s šířkou 2.5 mm prostřednictvím stanovených hranic  $-7.5, -5, \dots, 5$ .

```
262 femur.RF <- data.LRF$femur.LR
263 femur.LF <- data.LRF$femur.LL
264 femur.RLF <- femur.RF - femur.LF
```



Obrázek 17: Histogram a diagram rozdílů délky stehenní kosti z pravé a levé strany u žen aljašské populace z kmene Ipituaq

Protože  $p$ -hodnota = 0.8998 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z grafů vidíme, že normalita náhodného výběru rozdílů z pravé a levé strany je v pořádku. Histogram ukazuje, že rozdílů jsou symetrické a mají klasický kopcovitý tvar. V QQ-diagramu není patrná odchylka bodů od referenční přímky. V souladu s výsledkem Shapirova-Wilkova testu docházíme k závěru, že náhodný výběr rozdílů pochází z normálního rozdělení.

Předpoklad normality pro použití parametrického testu je tedy splněn. Naším úkolem je otestovat (nulovou) hy-



potézu o shodě střední hodnoty délky stehenní kosti z pravé a levé strany. Tento problém jsme převedli na analogický problém, kdy porovnáváme střední hodnotu rozdílů pravé a levé strany s konstantou  $\mu_0 = 0$ . Původní i analogické tvrzení jsou zněním nulové hypotézy. Zbývá dodefinovat alternativní hypotézu. Proces testování si předvedeme v posloupnosti sedmi kroků.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Střední hodnota rozdílů měření z pravé a levé strany je rovná nule.

$H_1$  : Střední hodnota rozdílů měření z pravé a levé strany není rovná nule.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $\mu = \mu_0$ , kde  $\mu_0 = 0$

$H_1$  :  $\mu \neq \mu_0$ , kde  $\mu_0 = 0$  (oboustranná alternativa)

### 2. Volba hladiny významnosti

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

### 3. Testování kritickým oborem

- **Testovací statistika**

$$T_W = \frac{M - \mu_0}{S} \sqrt{n} = \frac{-1.666667 - 0}{3.135815} \sqrt{12} = \frac{-1.666667}{3.135815} \times 3.464102 = -1.84115 \doteq -1.8412$$

```
265 alpha <- 0.05
266 mu_0 <- 0
267 m <- mean(femur.RLF)
268 s <- sd(femur.RLF)
269 tw <- (m - mu_0) / s * sqrt(n) # -1.841149
```

- **Kritický obor**

$$\begin{aligned} W &= (-\infty; t_{n-1}(\alpha/2)) \cup (t_{n-1}(1 - \alpha/2); \infty) \\ &= (-\infty; t_{12-1}(0.05/2)) \cup (t_{12-1}(1 - 0.05/2); \infty) \\ &= (-\infty; t_{11}(0.025)) \cup (t_{11}(0.975); \infty) \\ &= (-\infty; -2.2010) \cup (2.2010; \infty) \end{aligned}$$

```
270 qt(alpha / 2, n - 1) # -2.200985
271 qt(1 - alpha / 2, n - 1) # 2.200985
```

- **Závěr testování**

Protože realizace testovací statistiky  $t_W = -1.8412$  nenáleží do kritického oboru, tj.  $t_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

### 4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

$$\begin{aligned}
 (d, h) &= \left( m - \frac{s}{\sqrt{n}} t_{n-1}(1 - \alpha/2), m - \frac{s}{\sqrt{n}} t_{n-1}(\alpha/2) \right) \\
 &= \left( -1.666667 - \frac{3.135815}{\sqrt{12}} t_{12-1}(1 - 0.05/2), -1.666667 - \frac{3.135815}{\sqrt{12}} t_{12-1}(0.05/2) \right) \\
 &= \left( -1.666667 - \frac{3.135815}{3.464102} t_{37}(0.995), -1.666667 - \frac{3.135815}{3.464102} t_{37}(0.005) \right) \\
 &= (-1.666667 - 0.9052317 \times 2.200985, -1.666667 - 0.9052317 \times (-2.200985)) \\
 &= (-3.6591, 0.3257)
 \end{aligned}$$

```

272 dh <- m - s / sqrt(n) * qt(1 - alpha / 2, n - 1) # -3.659068
273 hh <- m - s / sqrt(n) * qt(alpha / 2, n - 1) # 0.3257349

```

- **Závěr testování**

Protože  $\mu_0 = 0$  náleží do Waldova 95% empirického oboustranného intervalu spolehlivosti, tj.  $\mu_0 = 0 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned}
 p\text{-hodnota} &= 2 \min\{\Pr(T_W \leq t_w), \Pr(T_W > t_w)\} \\
 &= 2 \min\{\Pr(T_W \leq t_w), 1 - \Pr(T_W \leq t_w)\} \\
 &= 2 \min\{\Pr(T_W \leq -1.841149), 1 - \Pr(T_W \leq -1.841149)\} \\
 &= 2 \min\{0.04635293, 0.9536471\} \\
 &= 2 \times 0.04635293 = 0.09270586 \doteq 0.0927
 \end{aligned}$$

```

274 p.hodnota <- 2 * min (pt(tw, n - 1), 1 - pt(tw, n - 1)) # 0.09270586

```

- **Závěr testování**

Protože  $p$ -hodnota = 0.0927 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 stehenní kosti na pravé a levé straně u žen aljašské populace z kmene Ipituaq neexistuje statisticky významný rozdíl.

## 7. Grafická vizualizace výsledku testování

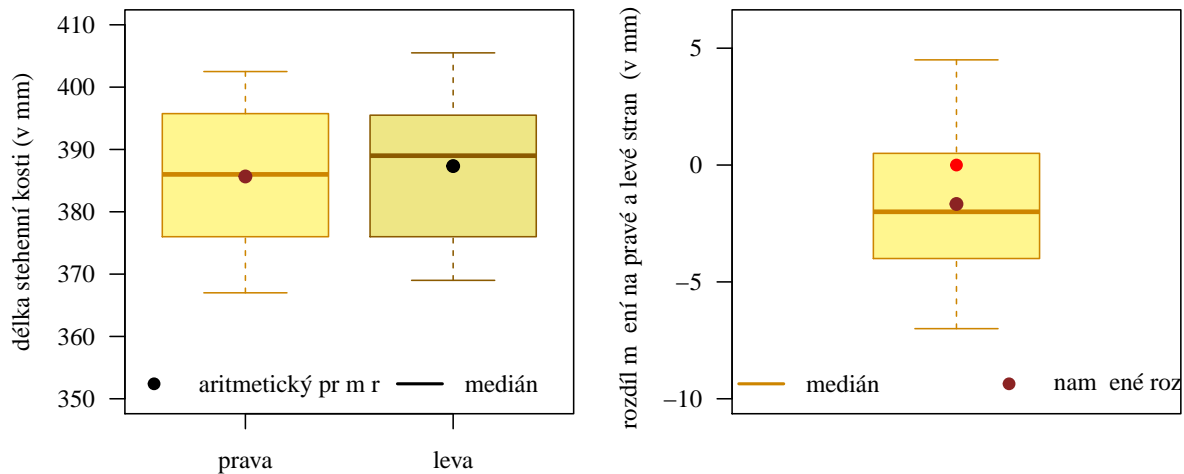
Porovnání délek stehenních kostí na pravé a levé straně nejlépe vizualizujeme pomocí krabicového diagramu. Vybrat si můžeme mezi dvěma variantami diagramů. Prvním, který porovnává vzájemně měření na pravé a levé straně, a druhým který porovnává rozdíly s konstantou  $\mu_0 = 0$ .

*Poznámka:* Párový test můžeme provést pomocí funkce `t.test()`. Vstupními parametry budou vektor reprezentující měření na pravé straně (`femur.RF`), vektor reprezentující měření na levé straně (`femur.LF`), argument `paired = T` určující, že oba vektory považujeme za párová pozorování, hodnota hladiny významnosti  $\alpha$  zadaná prostřednictvím koeficientu spolehlivosti  $1 - \alpha$  nastavením hodnoty argumentu `conf.level = 0.95` a typ zvolené alternativní hypotézy (oboustranná) zadaný pomocí argumentu `alternative = 'two.sided'`.

```

275 t.test(femur.RF, femur.LF, paired = T, conf.level = 0.95, alternative = 'two.sided')

```



Obrázek 18: Krabicový diagram rozdílů délky stehenní kosti z pravé a levé strany u žen aljašské populace z kmene Ipituaq

```

Paired t-test

data: femur.RF and femur.LF
t = -1.8411, df = 11, p-value = 0.09271
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.6590682  0.3257349
sample estimates:
mean of the differences
 -1.666667

```

276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286

Součástí výstupu je hodnota testovací statistiky  $t = -1.8411$ , počet stupňů volnosti Studentova rozdělení  $df = 11$ , hranice 95% Waldova empirického oboustranného intervalu spolehlivosti  $-3.6590682$  a  $0.3257349$  a  $p$ -hodnota  $p\text{-value} = 0.09271$ . Jediné, co musíme stanovit zvlášť, jsou dolní a horní hranice kritického oboru.

Druhou možností provedení párového testu je opět pomocí funkce `t.test()`, kde vstupními parametry budou vektor rozdílů naměřených hodnot na pravé a levé straně (`femur.RLF`), argument `mu = 0` určující, že rozdíl porovnáváme s konstantou  $\mu_0 = 0$ , hodnota hladiny významnosti  $\alpha$  zadaná prostřednictvím koeficientu spolehlivosti  $1 - \alpha$  (`conf.level = 0.95`) a typ zvolené alternativní hypotézy (`alternative = 'two.sided'`).

```
287 t.test(femur.RLF, mu = 0, conf.level = 0.95, alternative = 'two.sided')
```

```

One Sample t-test

data: femur.RLF
t = -1.8411, df = 11, p-value = 0.09271
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -3.6590682  0.3257349
sample estimates:
mean of x
 -1.666667

```

288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298

Výstup tohoto příkazu je totožný s výše uvedeným výstupem. Záleží tedy na nás, jakou syntaxi k zadání párového testu použijeme. ★

### Příklad 7.14. Párový test (pravostranná alternativa)

Máme datový soubor 03-paired-means-clavicle2.txt obsahující údaje o délkách klíční kosti (*clavicula*) z pravé strany (`length.R`) a levé strany (`length.L`) z anglického souboru dokumentovaných skeletů (Parsons, 1916, viz soubor D-03-paired-means-clavicle2). Na hladině významnosti  $\alpha = 0.05$  testujte hypotézu, že střední hodnota délky klíční kosti u mužů z levé strany je menší nebo rovna střední hodnotě délky klíční kosti u mužů z pravé strany.

#### Řešení příkladu 7.14

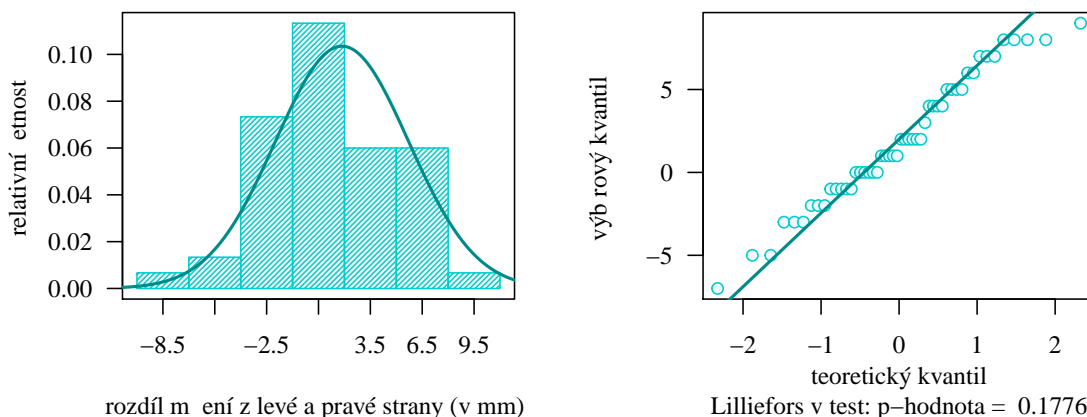
Nejprve příkazem `read.delim()` načteme datový soubor a pomocí operátoru `[]` z něj vybereme pouze údaje o délce klíční kosti z levé strany (sloupce `length.L` resp. z pravé strany (`length.R`) u mužů `sex == 'm'`. Údaje vložíme do proměnné `length.LRM`. Z datové tabulky následně odstraníme chybějící údaje a zjistíme rozsah náhodného výběru.

```
299 data <- read.delim('00-Data\\03-paired-means-clavicle2.txt')
300 data.LRM <- data[data$sex == 'm', c('length.L', 'length.R')]
301 data.LRM <- na.omit(data.LRM)
302 dim(data.LRM) # 50x2
```

Datový soubor obsahuje údaje o délce klíční kosti z levé a pravé strany u 50 mužů. Úkolem ze zadání je porovnat naměřené hodnoty na levé a pravé straně. Jde tedy o měření stejného znaku (délka klíční kosti) sledovaného na stejných subjektech (muži), proto použijeme na tuto situaci párový test. Prvním krokem k tohoto testu je vytvoření rozdílů hodnot naměřených na levé a pravé straně. V druhém kroku je potřeba ověřit předpoklad normality rozdílů.

Vzhledem k rozsahu náhodného výběru použijeme k ověření normality rozdílů Lillieforsův test ( $\alpha = 0.05$ ) v kombinaci s QQ-diagramem a histogramem (viz obrázek 19). Datový soubor rozdělíme do šesti ekvidistatních intervalů s šířkou 0.51 mm prostřednictvím stanovených hranic  $-2.195, -2.405, \dots, 0.865$ .

```
303 length.RM <- data.LRM[, 'length.R']
304 length.LM <- data.LRM[, 'length.L']
305 length.LRM <- length.LM - length.RM
```



Obrázek 19: Histogram a diagram rozdílů měření délky klíční kosti na levé a na pravé straně u mužů

Protože  $p$ -hodnota = 0.1776 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Zatímco histogram křivku hustoty normálního rozdělení přesně nevystihuje, v QQ-diagramu je zřejmá dostatečná příchylnost bodů k referenční přímce a výjimkou tří bodů na pravém chvostu. Náhodný výběr rozdílů délek klíčních kostí z levé a pravé strany tedy pochází z normálního rozdělení.

Naším úkolem je otestovat hypotézu, která tvrdí, že střední hodnota délky klíční kosti u mužů z levé strany je menší nebo rovna střední hodnotě délky klíční kosti u mužů z pravé strany. Tuto hypotézu převedeme na analogickou hypotézu, v rámci které porovnáme střední hodnotu rozdílů měření z levé a pravé strany s konstantou  $\mu_0 = 0$ . Nakonec dodefinujeme znění alternativní hypotézy.

## 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Střední hodnota délky klíční kosti u mužů z levé strany menší nebo rovná střední hodnotě délky klíční kosti u mužů z pravé strany.

$H_1$  : střední hodnota délky klíční kosti u mužů z levé strany větší než střední hodnota délky klíční kosti u mužů z pravé strany.

- **matematická formulace** nulové a alternativní hypotézy

$H_0 : \mu \leq \mu_0$ , kde  $\mu_0 = 0$

$H_1 : \mu > \mu_0$ , kde  $\mu_0 = 0$  (pravostranná alternativa)

## 2. Volba hladiny významnosti

- Hladinu významnosti volíme  $\alpha = 0.05$  (viz zadání příkladu).

## 3. Testování kritickým oborem

- **Testovací statistika**

$$T_W = \frac{M - \mu_0}{S} \sqrt{n} = \frac{1.86 - 0}{3.854549} \sqrt{50} = \frac{1.86}{3.854549} \times 7.071068 = 3.412121 \doteq 3.4121$$

```
306 alpha <- 0.05
307 mu_0 <- 0
308 m <- mean(length.LRM)
309 s <- sd(length.LRM)
310 tw <- (m - mu_0) / s * sqrt(n) # 3.41212
```

- **Kritický obor**

$$\begin{aligned} W &= \langle t_{n-1}(1 - \alpha); \infty \rangle \\ &= \langle t_{50-1}(1 - 0.05); \infty \rangle \\ &= \langle t_{49}(0.95); \infty \rangle \end{aligned}$$

```
311 qt(1 - alpha, n - 1) # 1.676551
```

- **Závěr testování**

Protože realizace testovací statistiky  $t_W = 3.4121$  náleží do kritického oboru, tj.  $t_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

## 4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

$$\begin{aligned} (d, h) &= \left( m - \frac{s}{\sqrt{n}} t_{n-1}(1 - \alpha/2), \infty \right) \\ &= \left( 1.86 - \frac{3.854549}{\sqrt{50}} t_{50-1}(1 - 0.05), \infty \right) \\ &= \left( 1.86 - \frac{3.854549}{7.071068} t_{49}(0.95), \infty \right) \\ &= (1.86 - 0.5451155 \times 1.676551, \infty) \\ &= (0.9460861, \infty) \end{aligned}$$

```
312 dh <- m - s / sqrt(n) * qt(1 - alpha, n - 1) # 0.9460859
```

- **Závěr testování**

Protože  $\mu_0 = 0$  nenáleží do Waldova 95% empirického jednostranného intervalu spolehlivosti, tj.  $\mu_0 = 0 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$p\text{-hodnota} = 1 - \Pr(T_W \leq t_W) = 1 - \Pr(T_W \leq 3.41212) = 0.0006503568 \doteq 0.0006504$$

```
313 p.hodnota <- 1 - pt(tw, n - 1) # 0.0006503568
```

- **Závěr testování**

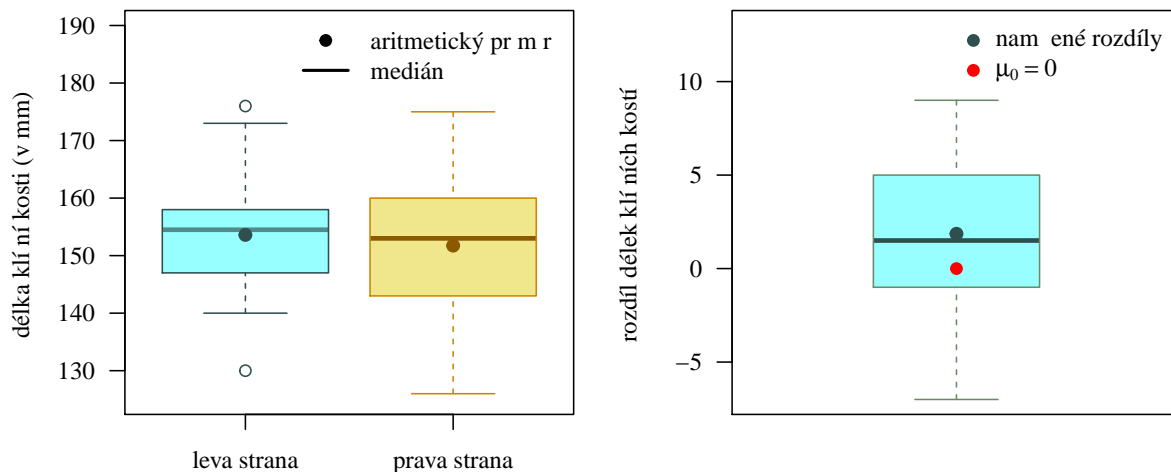
Protože  $p$ -hodnota = 0.0006504 je menší než  $\alpha = 0.05$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

## 6. Interpretace výsledků

Za základě všech tří typů testování zamítáme nulovou hypotézu na hladině významnosti  $\alpha = 0.05$ . Střední hodnota délky klíční kosti mužů na levé straně je statisticky významně větší než střední hodnota délky klíční kosti mužů na pravé straně.

## 7. Grafická vizualizace výsledku testování

Porovnání měření obou výzkumníků nejlépe vizualizujeme pomocí krabicového diagramu.



Obrázek 20: Krabicový diagram rozdílů délky klíční kosti u mužů na levé a na pravé straně

*Poznámka:* Párový test provedeme také pomocí funkce `t.test()`. Vstupními parametry budou vektor reprezentující rozdíly měření délky klíční kosti na levé straně a pravé straně (`length.LRM`), hodnota konstanty  $\mu_0 = 0$  (`mu = 0`), hodnota hladiny významnosti  $\alpha$  (`conf.level = 0.95`) a volba pravostranné alternativní hypotézy (`alternative = 'greater'`).

```
314 t.test(length.LRM, mu = 0, conf.level = 0.95, alternative = 'greater')
```

Součástí výstupu je hodnota testovací statistiky  $t = 3.4121$ , počet stupňů volnosti Studentova rozdělení  $df = 49$ , hranice 95% Waldova empirického oboustranného intervalu spolehlivosti 0.9460859 a lnf a  $p$ -hodnota  $p\text{-value} = 0.0006504$ . Jediné, co musíme stanovit zvlášť, je dolní hranice kritického oboru. ★

One Sample t-test	315
	316
data: length.LRM	317
t = 3.4121, df = 49, p-value = 0.0006504	318
alternative hypothesis: true mean is greater than 0	319
95 percent confidence interval:	320
0.9460859            Inf	321
sample estimates:	322
mean of x	323
1.86	324
	325

### Příklad 7.15. Párový test (levostranná alternativa)

Máme datový soubor 02-paired-means-clavicle.txt obsahující údaje o hodnotách vertikálního průměru středu délky těla klíční kosti z pravé a levé strany (*clavicula*) z pohřebiště u Sv. Jakuba v Brně, převážně z období středověku, naměřené jedním výzkumníkem ve dvou opakovaných měřeních (hodnoty naměřené při prvním opakování jsou uloženy v proměnné `simd.1`, hodnoty naměřené při druhém opakování jsou uloženy v proměnné `simd.2`) a naměřené druhým výzkumníkem v jednom měření (`simd`). Více informací viz sekce ???. Na hladině významnosti  $\alpha = 0.01$  zjistíte, zda jsou měření levé strany prvního výzkumníka statisticky významně nižší než měření levé strany druhého výzkumníka.

#### Řešení příkladu 7.15

Nejprve příkazem `read.delim()` načteme datový soubor a pomocí operátoru `[]` z něj vybereme pouze údaje naměřené prvním výzkumníkem (sloupce `simd.1` a `simd.2`) a druhým výzkumníkem (`simd`) na levé straně `side == 'L'`. Údaje vložíme do proměnné `data.12R`. Z datové tabulky následně odstraníme chybějící údaje a zjistíme rozsah náhodného výběru.

```
326 data <- read.delim('00-Data\\02-paired-means-clavicle.txt')
327 data.12L <- data[data$side == 'L', c('simd.1', 'simd.2', 'simd')]
328 data.12L <- na.omit(data.12L)
329 dim(data.12L) # 38x3
```

Datový soubor obsahuje 38 opakovaných měření vertikálního průměru středu délky těla klíční kosti z levé strany, a to dvakrát prvním výzkumníkem a jedenkrát druhým výzkumníkem. Nyní pomocí funkce `apply()` vytvoříme z obou měření prvního výzkumníka aritmetické průměry (`simd.AL`), které budeme následně porovnávat s naměřenými údaji druhého výzkumníka (`simd.BL`).

```
330 simd.AL <- apply(data.12L[, c('simd.1', 'simd.2')], 1, mean)
331 simd.BL <- data.12L$simd
```

Naším úkolem ze zadání je porovnat měření provedená dvěma různými výzkumníky. Jde ale o měření stejného znaku sledovaného na stejných subjektech, proto je vhodné použít na tuto situaci párový test. Nejprve tedy vytvoříme vektor rozdílů hodnot naměřených prvním a druhým výzkumníkem. V druhém kroku ověříme jediný předpoklad, který musí být splněn, abychom mohli párový test provést, a sice normalita těchto rozdílů.

Vzhledem k rozsahu náhodného výběru ( $n = 38 > 30$ ) použijeme k ověření normality rozdílů Lillieforsův test ( $\alpha = 0.05$ ). Graficky zhodnotíme rozdělení náhodného výběru QQ-diagramem a histogramem (viz obrázek 21). Datový soubor rozdělíme do šesti ekvidistatních intervalů s šířkou 0.51 mm prostřednictvím stanovených hranic  $-2.195, -2.405, \dots, 0.865$ .

```
332 simd.ABL <- simd.AL - simd.BL
333 nortest::lillie.test(simd.ABL)$p.val # 0.05234526
```

Protože  $p$ -hodnota = 0.0523 je větší než 0.05, nulovou hypotézu o normalitě dat nezamítáme na hladině významnosti  $\alpha = 0.05$ . Z grafů vidíme, že normalita náhodného výběru není příliš přesvědčivá, však jsme byli pouze kousek k zamítnutí hypotézy o normálním rozdělení náhodného výběru. Lillieforsův test však přiklonil zrníčko vah směrem k normálnímu rozdělení, proto v souladu s výsledkem testu předpoklad normality výběru rozdílů hodnot měření obou výzkumníků nezamítáme.

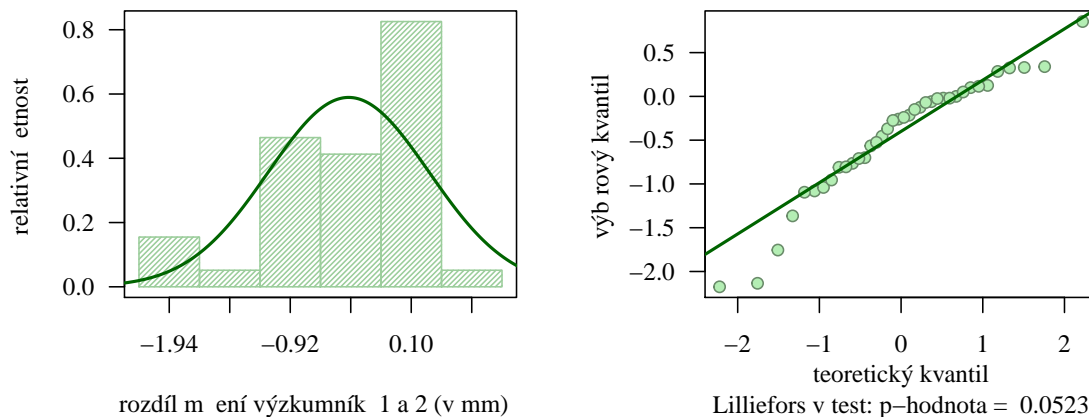
Předpoklad normality pro použití parametrického testu je tedy splněn. Naším úkolem je zjistit, zda jsou měření levé strany prvního výzkumníka statisticky významně nižší než měření levé strany druhého výzkumníka. Tento problém jsme převedli na analogický problém, kdy zjišťujeme, zda je střední hodnota rozdílů měření prvního a druhého výzkumníka menší než konstanta  $\mu_0 = 0$ . Původní i analogické tvrzení jsou zněním alternativní hypotézy. Zbývá dodefinovat nulovou hypotézu.

#### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Střední hodnota rozdílů měření prvního a druhého výzkumníka je větší nebo rovná nule.





Obrázek 21: Histogram a diagram rozdílů měření prvního a druhého výzkumníka hodnot vertikálního průměru ve středu délky těla klíční kosti na levé straně

$H_1$  : Střední hodnota rozdílů měření prvního a druhého výzkumníka je menší než nula.

- **matematická formulace** nulové a alternativní hypotézy

$H_0 : \mu \geq \mu_0$ , kde  $\mu_0 = 0$

$H_1 : \mu < \mu_0$ , kde  $\mu_0 = 0$  (levostranná alternativa)

## 2. Volba hladiny významnosti

- Hladinu významnosti volíme v souladu se zadáním jako  $\alpha = 0.01$ .

## 3. Testování kritickým oborem

- **Testovací statistika**

$$T_W = \frac{M - \mu_0}{S} \sqrt{n} = \frac{-0.4273684 - 0}{0.6768715} \sqrt{38} = \frac{-0.4273684}{0.6768715} \times 6.164414 = -3.892136 \doteq -3.8921$$

```
334 alpha <- 0.01
335 mu_0 <- 0
336 m <- mean(simd.ABL)
337 s <- sd(simd.ABL)
338 tw <- (m - mu_0) / s * sqrt(n) # -3.892136
```

- **Kritický obor**

$$\begin{aligned} W &= (-\infty; t_{n-1}(\alpha)) \\ &= (-\infty; t_{38-1}(0.01)) \\ &= (-\infty; t_{37}(0.001)) \\ &= (-\infty; -2.4314) \end{aligned}$$

```
339 qt(alpha, n - 1) # -2.431447
```

- **Závěr testování**

Protože realizace testovací statistiky  $t_W = -3.8921$  náleží do kritického oboru, tj.  $t_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

#### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}(d, h) &= \left( -\infty, m - \frac{s}{\sqrt{n}} t_{n-1}(\alpha) \right) \\ &= \left( -\infty, -0.4273684 - \frac{0.6768715}{\sqrt{38}} t_{38-1}(0.01) \right) \\ &= \left( -\infty, -0.4273684 - \frac{0.6768715}{6.164414} t_{37}(0.01) \right) \\ &= (-\infty, -0.4273684 - 0.109803 \times (-2.431447)) \\ &= (-\infty, -0.1603882)\end{aligned}$$

```
340 hh <- m - s / sqrt(n) * qt(alpha, n - 1) # -0.1603881
```

- Závěr testování

Protože  $\mu_0 = 0$  nenáleží do Waldova 99% empirického oboustranného intervalu spolehlivosti, tj.  $\mu_0 = 0 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

#### 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned}p\text{-hodnota} &= \Pr(T_W \leq t_W) \\ &= \Pr(T_W \leq -3.892136) \\ &= 0.0001999986 \doteq 0.0002\end{aligned}$$

```
341 p.hodnota <- pt(tw, n - 1) # 0.0001999986
```

- Závěr testování

Protože  $p$ -hodnota = 0.0002 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 nulovou hypotézu na hladině významnosti  $\alpha = 0.01$ . Měření levé strany prvního výzkumníka jsou statisticky významně nižší než měření levé strany druhého výzkumníka.

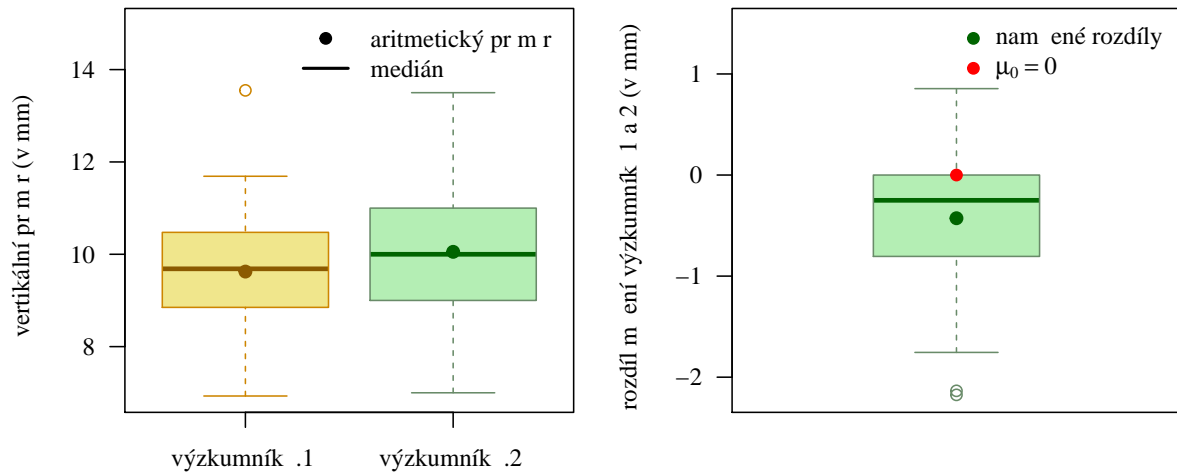
#### 7. Grafická vizualizace výsledku testování

Porovnání měření obou výzkumníků vizualizujeme pomocí krabicového diagramu. Vybrat si můžeme mezi dvěma variantami diagramů. Prvním, který porovnává vzájemně měření prvního a druhého výzkumníka, a druhým který porovnává rozdíly s konstantou  $\mu_0 = 0$ .

*Poznámka:* Párový test provedeme také pomocí funkce `t.test()`. Vstupními parametry budou vektor reprezentující měření prvního výzkumníka na levé straně (`simd.AL`), vektor reprezentující měření druhého výzkumníka na levé straně (`simd.BL`), volba párového testu (`paired = T`), hodnota hladiny významnosti  $\alpha$  (`conf.level = 0.99`) a volba levostranné alternativní hypotézy (`alternative = 'less'`).

```
342 t.test(simd.AL, simd.BL, paired = T, conf.level = 0.99, alternative = 'less')
```

Součástí výstupu je hodnota testovací statistiky  $t = -3.8921$ , počet stupňů volnosti Studentova rozdělení  $df = 37$ , hranice 99% Waldova empirického oboustranného intervalu spolehlivosti  $-\lnf$  a  $-0.1603881$  a  $p$ -hodnota  $p\text{-value} = 2e-04 = 2 \times 10^{-4} = 0.0002$ . Jediné, co musíme stanovit zvláště, je horní hranice kritického oboru. ★



Obrázek 22: Krabicový diagram rozdílů měření prvního a druhého výzkumníka hodnot vertikálního průměru ve středu délky těla klíční kosti na levé straně

```

Paired t-test
data: simd.AL and simd.BL
t = -3.8921, df = 37, p-value = 2e-04
alternative hypothesis: true difference in means is less than 0
99 percent confidence interval:
  -Inf -0.1603881
sample estimates:
mean of the differences
  -0.4273684

```

343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353

**Příklad 7.16. Párový test (nesplnění předpokladů)** Máme datový soubor 02-paired-means-clavicle.txt obsahující údaje o hodnotách vertikálního průměru středu délky těla klíční kosti z pravé a levé strany (*clavicula*) z pohřebiště u Sv. Jakuba v Brně, převážně z období středověku, naměřené jedním výzkumníkem ve dvou opakovaných měřeních (viz sekce ??). Hodnoty naměřené při prvním opakování jsou uloženy v proměnné `simd.1`, hodnoty naměřené při druhém opakování jsou uloženy v proměnné `simd.2`. Můžeme zjistit, že aritmetický průměr hodnot získaných v rámci prvního měření je větší než aritmetický průměr hodnot získaných v rámci druhého měření. Na hladině významnosti  $\alpha = 0.05$  zjistěte, zda je střední hodnota prvního měření větší než střední hodnota druhého měření vertikálního průměru délky těla klíční kosti na levé straně provedené tímto výzkumníkem.

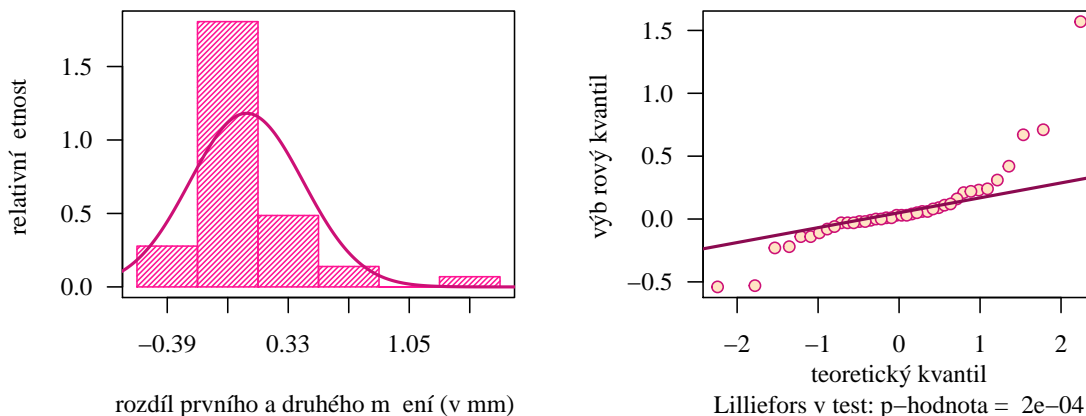
### Řešení příkladu 7.16

Nejprve příkazem `read.delim()` načteme datový soubor a pomocí operátoru `[]` z něj vybereme pouze údaje naměřené sledovaným výzkumníkem (sloupce `simd.1` a `simd.2`) na levé straně `side == 'L'`. Údaje vložíme do proměnné `data.12L`. Z datové tabulky následně odstraníme chybějící údaje a zjistíme rozsah náhodného výběru.

```
354 data <- read.delim('00-Data\\02-paired-means-clavicle.txt')
355 data.12L <- data[data$side == 'L', c('simd.1', 'simd.2')]
356 data.12L <- na.omit(data.12L)
357 dim(data.12L) # 40x2
```

Datový soubor obsahuje 40 opakovaných měření délky těla klíční kosti z levé strany. Protože naším úkolem ze zadání je porovnat opakovaná měření provedená na jednom subjektu, použijeme k tomuto porovnání párový test. Prvním krokem k provedení tohoto testu je vytvoření rozdílů hodnot získaných v prvním a druhém měření. V druhém kroku je potřeba ověřit jediný předpoklad, který musí být splněn, abychom mohli párový test provést, a sice ověřit normalní rozdělení těchto rozdílů.

Vzhledem k rozsahu náhodného výběru ( $n = 40 > 30$ ) použijeme k ověření normality rozdílů Lillieforsův test ( $\alpha = 0.05$ ). Graficky zhodnotíme rozdělení náhodného výběru QQ-diagramem a histogramem (viz obrázek ??). Datový soubor rozdělíme do šesti ekvidistatních intervalů s šířkou 0.36 mm prostřednictvím stanovených hranic  $-0.57, -0.21, \dots, 1.59$ .



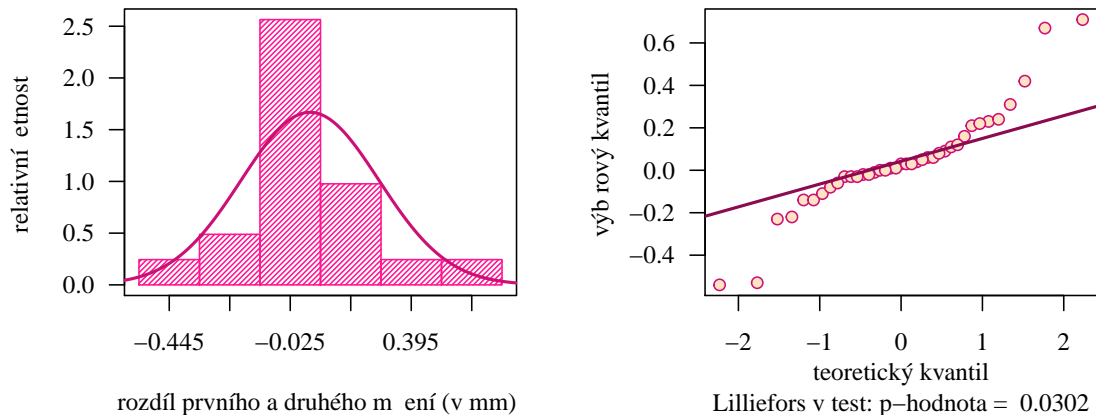
Obrázek 23: Histogram a diagram rozdílů prvního a druhého měření hodnot vertikálního průměru ve středu délky těla klíční kosti na levé straně provedené jedním výzkumníkem

Protože  $p$ -hodnota = 0.00020223 je menší než 0.05, nulovou hypotézu o normalitě rozdílů zamítáme na hladině významnosti  $\alpha = 0.05$ . Histogram naměřených hodnot ukazuje na vyšikmení rozdílů směrem doleva s prodlouženým pravým koncem. QQ-diagram ukazuje na odlehlost bodů od referenční přímky a minimálně jeden extrémně odlehlý bod na pravé straně. Náhodný výběr rozdílů hodnot získaných v prvním a druhém měření nepochází z normálního rozdělení.

Předpoklad normality pro použití parametrického testu není splněn, proto není možné tvrzení ze zadání ověřit po-

mocí parametrického párového testu. Jednou z možností, jak příklad dále řešit je zkusit odstranit rozdíl s nejvyšší hodnotou vyskytující se ve vektoru `simd.12L`. Vypsáním vektoru rozdílů `simd.12L` můžeme zjistit, že odlehlé pozorování nabývá hodnoty 1.57 a je umístěno na 36. pozici ve vektoru `simd.12L`. Pomocí operátoru `[]` toto pozorování odstraníme a provedeme opětovně Lillieforsův test normality a zobrazíme histogram a QQ-diagram (viz obrázek 24).

```
358 simd.12L2 <- simd.12L[-36]
```



Obrázek 24: Histogram a diagram rozdílů prvního a druhého měření hodnot vertikálního průměru ve středu délky těla klíční kosti na levé straně po odstranění nejodlehlejšího pozorování

Protože  $p$ -hodnota = 0.0302 je menší než 0.05, nulovou hypotézu o normalitě rozdílů zamítáme na hladině významnosti  $\alpha = 0.05$ . Histogram zobrazující rozdílů prvního a druhého měření sice vypadá lépe, ale u QQ-diagramu stále vidíme fatální odchylky bodů od referenční přímky. Náhodný výběr rozdílů hodnot získaných v prvním a druhém měření nepochází z normálního rozdělení.

Předpoklad normality pro použití parametrického testu není ani po odstranění nejodlehlejšího rozdílu splněn, proto není možné tvrzení ze zadání ověřit pomocí parametrického párového testu. Otázku ze zadání bychom tedy ověřili pomocí neparametrické alternativy párového testu, tj. metodami uvedenými v kapitole ??.

★

## 7.5 Test o korelačním koeficientu $\rho$

Nechť  $(X_1, Y_1)^T \dots (X_n, Y_n)^T$  je náhodný výběr z  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  a necht'  $\rho_0$  je konstanta. Na hladině významnosti  $\alpha$  testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

$$\begin{array}{lll} H_{01} : \rho = \rho_0 & \text{oproti} & H_{11} : \rho \neq \rho_0 \quad (\text{oboustranná alt.}) \\ H_{02} : \rho \leq \rho_0 & \text{oproti} & H_{12} : \rho > \rho_0 \quad (\text{pravostranná alt.}) \\ H_{03} : \rho \geq \rho_0 & \text{oproti} & H_{13} : \rho < \rho_0 \quad (\text{levostranná alt.}) \end{array}$$

Test nazýváme jednovýběrovým  $Z$ -testem o korelačním koeficientu  $\rho$ . Testovací statistika má tvar

$$Z_W = \sqrt{n-3}(Z_R - \xi_0), \quad (7.4)$$

kde  $Z_R = \frac{1}{2} \ln \frac{1+R}{1-R}$  je Fisherova  $Z$ -transformace výběrového korelačního koeficientu  $R$  a  $\xi_0 = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}$  je Fisherova  $Z$ -transformace konstanty  $\rho_0$  z nulové hypotézy a  $n$  je rozsah náhodného výběru. Testovací statistika  $Z_W$  pochází asymptoticky ( $n \geq 10$ ) ze standardizovaného normálního rozdělení, tj.

$$Z_W = \sqrt{n-3}(Z_R - \xi_0) \overset{A}{\sim} N(0, 1).$$

Kritický obor podle zvolené alternativní hypotézy má tvar

$$\begin{array}{ll} H_{11} : \rho \neq \rho_0 & W = (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty) \\ H_{12} : \rho > \rho_0 & W = (u_{1-\alpha}; \infty) \\ H_{13} : \rho < \rho_0 & W = (-\infty; u_{\alpha}) \end{array}$$

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

$$\begin{array}{ll} H_{11} : \rho \neq \rho_0 & (d, h) = \left( \tanh \left( z_R - \frac{u_{1-\alpha/2}}{\sqrt{n-3}} \right); \tanh \left( z_R - \frac{u_{\alpha/2}}{\sqrt{n-3}} \right) \right) \\ H_{12} : \rho > \rho_0 & (d, 1) = \left( \tanh \left( z_R - \frac{u_{1-\alpha}}{\sqrt{n-3}} \right); 1 \right) \\ H_{13} : \rho < \rho_0 & (-1, h) = \left( -1; \tanh \left( z_R - \frac{u_{\alpha}}{\sqrt{n-3}} \right) \right) \end{array}$$

kde  $\tanh$  je hyperbolický tangens, jehož hodnotu získáme pomocí  $\mathbb{R}$  a implementované funkce `tanh()`.

*Poznámka:* Dá se ukázat, že mezi korelačním koeficientem  $\rho_0$  a  $\xi_0$  platí vztah  $\rho_0 = \tanh(\xi_0)$ , který je inverzí ke vztahu  $\xi_0 = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}$ . Všimněme si, že  $z_R - \frac{u_{1-\alpha/2}}{\sqrt{n-3}}$  a  $z_R - \frac{u_{\alpha/2}}{\sqrt{n-3}}$  jsou dolní a horní hranice intervalu spolehlivosti pro  $Z$ -transformaci  $\xi_0$ , neboť  $z_R$  je také  $Z$ -transformace. Hyperbolický tangens potom funguje jako zpětná transformace, která převede hranice intervalu spolehlivosti pro  $\xi_0$  zpátky na hranice intervalu spolehlivosti pro korelační koeficient  $\rho_0$ .

*Poznámka:* Protože parametr  $\rho$  je korelační koeficient, platí, že  $\rho \in (-1; 1)$ . Proto levostranný interval spolehlivosti omezíme shora hodnotou 1, namísto nekonečnem, a pravostranný interval spolehlivosti omezíme zdola hodnotou -1, namísto mínus nekonečnem.

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

$$\begin{array}{ll} H_{11} : \rho \neq \rho_0 & p\text{-hodnota} = 2 \min\{\Pr(Z_W \leq z_W), \Pr(Z_W > z_W)\} \\ H_{12} : \rho > \rho_0 & p\text{-hodnota} = \Pr(Z_W > z_W) = 1 - \Pr(Z_W \leq z_W) \\ H_{13} : \rho < \rho_0 & p\text{-hodnota} = \Pr(Z_W \leq z_W) \end{array}$$

kde  $Z_W$  je náhodná veličina,  $z_W$  je realizace testovací statistiky  $Z_W$  (viz vzorec 7.4), tedy konkrétní číslo, a  $\Pr(Z_W \leq z_W)$  je distribuční funkce standardizovaného normálního rozdělení, jejíž hodnotu získáme pomocí  $\mathbb{R}$  a implementované funkce `pnorm()`.

### Příklad 7.17. Test o korelačním koeficientu $\rho$ (oboustranná alternativa)

Mějme datový soubor 05-one-sample-correlation-skull-mf.txt, proměnnou skull.pH popisující největší výšku mozkovny a proměnnou face.H popisující morfologickou výšku tváře (viz sekce ??) starověké egyptské populace. Současně máme k dispozici hodnotu korelačního koeficientu mezi oběma znaky a údaje o počtu případů ze vzorku novověké egyptské mužské populace. Na hladině významnosti  $\alpha = 0.05$  testujte hypotézu o shodě korelačního koeficientu největší výšky mozkovny a morfologické výšky tváře u mužů starověké a novověké egyptské populace.

### Řešení příkladu 7.17

Datový soubor načteme příkazem `read.delim()`. Pomocí operátoru `[]` vybereme z tabulky údaje o největší výšce mozkovny (`skull.pH`) a morfologické výšce tváře `face.H` u mužů `sex == 'm'`. Údaje vložíme do proměnné `data.M`. Pomocí příkazu `na.omit()` odstraníme z tabulky `data.M` chybějící údaje. Nakonec příkazem `dim()` zjistíme rozsah náhodného výběru a příkazem `range()` rozsahy naměřených hodnot obou proměnných.

```
359 data <- read.delim('00-Data\\05-one-sample-correlation-skull-mf.txt')
360 data.M <- data[data$sex == 'm', c('skull.pH', 'face.H')]
361 data.M <- na.omit(data.M)
362 skull.pHM <- data.M$skull.pH
363 face.HM <- data.M$face.H
364 range(skull.pHM) # 127-149
365 range(face.HM) # 100-136
```

Datový soubor obsahuje údaje o největší výšce mozkovny a morfologické výšce tváře u 164 mužů, přičemž naměřené největší výšky mozkovny nabývají hodnot v rozmezí 127–149 mm a naměřené morfologické výšky tváře nabývají hodnot v rozmezí 100–136 mm.

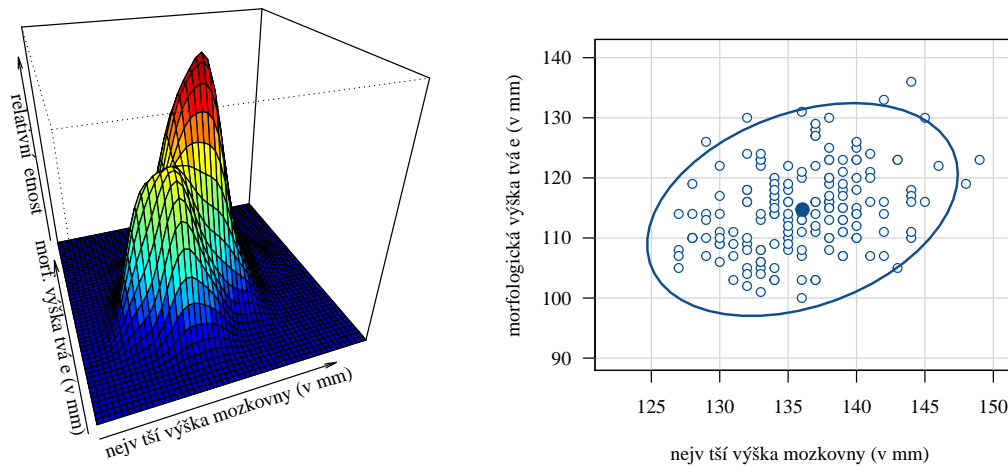
Naším úkolem je porovnat korelační koeficienty dvou egyptských populací, přičemž u mužů ze starověké egyptské populace známe naměřené hodnoty. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z dvourozměrného normálního rozdělení. Druhá, novověká egyptská populace je reprezentována pouze hodnotou korelačního koeficientu ( $r_0 = 0.251$ ). O jejím rozdělení přesnější informace nemáme. Řešení příkladu vede na případ, kdy korelační koeficient jednoho náhodného výběru porovnáváme s konkrétním číslem, tedy na jednovýběrový test o korelačním koeficientu  $\rho$ . Předpokladem k použití tohoto testu je dvourozměrná normalita náhodného výběru největších výšek mozkovny a morfologických výšek tváře mužů starověké egyptské populace. Před použitím testu je třeba tento předpoklad ověřit.

Závěr o dvourozměrné normalitě obou náhodných výběrů stanovíme na základě Mardiova testu ( $\alpha = 0.05$ ) v kombinaci s grafickou vizualizací dat pomocí 3D grafu a tečkového diagramu superponovaného 95 % elipsou spolehlivosti, analogicky, jako je uvedeno v sekci ??.

```
366 MVN::mvn(data.M, mvnTest = 'mardia')$multivariateNormality
367 # šikmost: 0.3852790 # špicatost: 0.1998495
```

Protože  $p$ -hodnota testu o nevýznamnosti koeficientu šikmosti, tj. 0.3853, je větší než 0.05, hypotézu o nevýznamnosti koeficientu šikmosti nezamítáme na hladině významnosti  $\alpha = 0.05$ . Dále protože  $p$ -hodnota testu o nevýznamnosti koeficientu špicatosti, tj. 0.1998, je větší než 0.05, nezamítáme hypotézu o nevýznamnosti koeficientu špicatosti. Protože náhodný výběr nevykazuje statisticky významné známky zešikmení ani zešpicatění, nezamítáme na hladině významnosti  $\alpha = 0.05$  hypotézu o dvourozměrné normalitě náhodného výběru největších výšek mozkovny a morfologických výšek tváře u mužů starověké egyptské populace. Ke stejnému závěru bychom došli také použitím Henzova-Zirklerova testu ( $p$ -hodnota = 0.64065 > 0.05) i Roystonova testu ( $p$ -hodnota = 0.05850 > 0.05).

Nyní se podíváme na grafickou vizualizaci náhodného výběru (viz graf 25) prostřednictvím 3D grafu obarveného pomocí 20 odstínů z palety `matlab.like2` z knihovny `colorRamps`. 3D graf nám ukazuje kopcovitý tvar náhodného výběru složený z vyššího a nižšího kopce, které splývají v jeden objekt. Vzhledem k tomu, že rozsah náhodného výběru je 164, požadujeme, aby elipsa spolehlivosti pokrývala alespoň 156 bodů. Zbylých 8 bodů smí ležet mimo elipsu spolehlivosti. V našem případě leží mimo elipsu spolehlivosti právě osm bodů, což podporuje výsledek Mardiova testu. Náhodný výběr největších výšek mozkovny a morfologických výšek tváře u mužů starověké egyptské populace pochází z dvourozměrného normálního rozdělení.



Obrázek 25: 3D graf a tečkový diagram s 95% elipsou spolehlivosti pro největší výšku mozkovny a morfologickou výšku tváře mužů starověké egyptské populace (v mm)

Jelikož je předpoklad dvourozměrné normality náhodného výběru splněn, můžeme hypotézu ze zadání otestovat pomocí parametrického testu o korelačním koeficientu  $\rho$ . Naším úkolem otestovat (nulovou) hypotézu o shodě korelačního koeficientu největší výšky mozkovny a morfologické výšky tváře u mužů starověké a novověké egyptské populace. Zbývá tedy stanovit znění alternativní hypotéz tak, aby bylo doplňkem k nulové hypotéze. Testování provedeme v posloupnosti sedmi kroků.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Korelační koeficient proměnných největší výška mozkovny a morfologická výška tváře u mužů starověké egyptské populace je rovný hodnotě 0.251.

$H_1$  : Korelační koeficient proměnných největší výška mozkovny a morfologická výška tváře u mužů starověké egyptské populace není rovný hodnotě 0.251.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $\rho = \rho_0$ , kde  $\rho_0 = 0.251$

$H_1$  :  $\rho \neq \rho_0$ , kde  $\rho_0 = 0.251$   
(oboustranná alternativa)

### 2. Volba hladiny významnosti

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

### 3. Testování kritickým oborem

- Fisherova  $Z$ -transformace výběrového korelačního koeficientu

K výpočtu Fisherovy  $Z$ -transformace potřebujeme znát hodnotu výběrového korelačního koeficientu. Tuto hodnotu získáme pomocí příkazu `cor()` s nastaveným argumentem `method == 'pearson'`. Výběrový korelační koeficient největší výšky mozkovny a morfologické výšky tváře mužů starověké egyptské populace  $r = 0.33064$ . Nyní můžeme vypočítat Fisherovu  $Z$ -transformaci korelačního koeficientu a Fisherovu  $Z$ -transformaci konstanty  $\rho_0 = 0.251$  z nulové hypotézy.

$$\begin{aligned} Z_R &= \frac{1}{2} \ln \frac{1+R}{1-R} \\ &= \frac{1}{2} \ln \frac{1+0.3306431}{1-0.3306431} = \frac{1}{2} \ln \frac{1.3306431}{0.6693569} = \frac{1}{2} \ln 1.987943 \\ &= 0.5 \times 0.6871002 = 0.3435501 \doteq 0.3436 \end{aligned}$$



$$\begin{aligned}
\xi_0 &= \frac{1}{2} \ln \frac{1 + \rho_0}{1 - \rho_0} \\
&= \frac{1}{2} \ln \frac{1 + 0.251}{1 - 0.251} = \frac{1}{2} \ln \frac{1.251}{0.749} = \frac{1}{2} \ln 1.670227 \\
&= 0.5 \times 0.5129595 = 0.2564798 \doteq 0.2565
\end{aligned}$$

```

368 alpha <- 0.05
369 n     <- length(skull.pHM)
370 rho0  <- 0.251
371 r     <- cor(skull.pHM, face.HM, method = 'pearson')
372 zR    <- 1 / 2 * log((1 + r) / (1 - r)) # 0.3435501
373 ksi0  <- 1 / 2 * log((1 + rho0) / (1 - rho0)) # 0.2564798

```

- Testovací statistika

$$\begin{aligned}
Z_W &= \frac{Z_R - \xi_0}{\sqrt{n - 3}} \\
&= (0.3435501 - 0.2564798) \sqrt{164 - 3} \\
&= 0.0870703 \times \sqrt{161} \\
&= 0.0870703 \times 12.68858 \\
&= 1.104798
\end{aligned}$$

```

374 Zw <- (zR - ksi0) * sqrt(n - 3) # 1.104799

```

- Kritický obor

$$\begin{aligned}
W &= (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty) \\
&= (-\infty; u_{0.05/2}) \cup (u_{1-0.05/2}; \infty) \\
&= (-\infty; u_{0.025}) \cup (u_{0.975}; \infty) \\
&= (-\infty; -1.959964) \cup (1.959964; \infty)
\end{aligned}$$

```

375 qnorm(alpha / 2) # -1.959964
376 qnorm(1 - alpha / 2) # 1.959964

```

- Závěr testování

Protože realizace testovací statistiky  $z_W = 1.1048$  nenáleží do kritického oboru, tj.  $z_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

#### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
(d, h) &= \left( \tanh \left( z_R - \frac{u_{1-\alpha/2}}{\sqrt{n-3}} \right); \tanh \left( z_R - \frac{u_{\alpha/2}}{\sqrt{n-3}} \right) \right) \\
&= \left( \tanh \left( 0.3435501 - \frac{u_{1-0.05/2}}{\sqrt{164-3}} \right); \tanh \left( 0.3435501 - \frac{u_{0.05/2}}{\sqrt{164-3}} \right) \right) \\
&= \left( \tanh \left( 0.3435501 - \frac{u_{0.975}}{\sqrt{161}} \right); \tanh \left( 0.3435501 - \frac{u_{0.025}}{\sqrt{161}} \right) \right) \\
&= \left( \tanh \left( 0.3435501 - \frac{1.959964}{12.68858} \right); \tanh \left( 0.3435501 - \frac{-1.959964}{12.68858} \right) \right) \\
&= (\tanh(0.3435501 - 0.1544668); \tanh(0.3435501 - (-0.1544668))) \\
&= (\tanh(0.1890833); \tanh(0.4980169)) \\
&= (0.1869; 0.46056)
\end{aligned}$$

```

377 dh <- tanh(zR - qnorm(1 - alpha / 2) / sqrt(n - 3)) # 0.1868617
378 hh <- tanh(zR - qnorm(alpha / 2) / sqrt(n - 3)) # 0.4605561

```

- **Závěr testování**

Protože  $\rho_0 = 0.251$  náleží do Waldova 95% empirického oboustranného intervalu spolehlivosti, tj.  $\rho_0 = 0.251 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- **$p$ -hodnota**

$$\begin{aligned}
p\text{-hodnota} &= 2 \min\{\Pr(Z_w \leq z_w), \Pr(Z_W > z_w)\} \\
&= 2 \min\{\Pr(Z_w \leq z_w), 1 - \Pr(Z_W \leq z_w)\} \\
&= 2 \min\{\Pr(Z_w \leq 1.104799), 1 - \Pr(Z_W \leq 1.104799)\} \\
&= 2 \min\{0.8653766, 0.1346234\} \\
&= 2 \times 0.1346234 \\
&= 0.2692467 \doteq 0.2692
\end{aligned}$$

```

379 2 * min(pnorm(zw), 1 - pnorm(zw)) # 0.2692467

```

- **Závěr testování**

Protože  $p$ -hodnota = 0.2692 je větší než  $\alpha = 0.05$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 6. Grafická vizualizace výsledků testování

Vhodným grafem ukazujícím míru závislosti mezi největší výškou mozkovny a morfologickou výškou tváře je tečkový diagram superponovaný lineární regresní přímkou prokládající zobrazené body. Nejprve příkazem `plot()` vykreslíme body největší výšky mozkovny a morfologické výšky tváře mužů starověké egyptské populace. Koeficienty lineární regresní přímky získáme pomocí funkce `lm()`, jejímž jediným argumentem bude vztah `face.HM skull.pHM`, tj. vztah vyjadřující závislost mezi proměnnou `face.HM` na ose  $y$  a proměnnou `skull.pHM` na ose  $x$ . Koeficienty regresní přímky, které jsou vloženy v položce `coefficients`, získáme z výstupu funkce `lm` pomocí odkazu `$coef`. Dále vytvoříme posloupnost tisíce bodů  $x$  v rozsahu hodnot proměnné `skull.pHM` a vypočítáme hodnoty regresní přímky v bodech posloupnosti  $x$ , které vložíme do proměnné  $y$ . Nakonec vykreslíme lineární regresní přímkou v bodech  $x, y$  příkazem `lines()`. Nakonec do grafu doplníme pod osu  $x$  popisek obsahující hodnotu výběrového korelačního koeficientu  $r = 0.33064$  (příkaz `mtext()`).

```

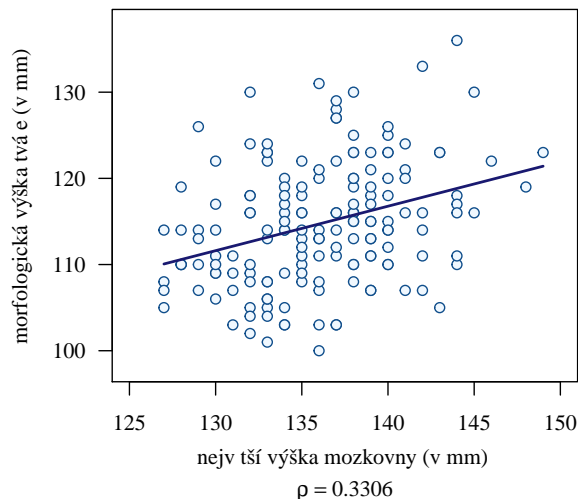
380 par(mar = c(5, 4, 1, 1), family = 'Times')

```

```

381 plot(skull.pHM, face.HM, xlab = '', ylab = 'morfologická výška tváře (v mm)',
382       xlim = c(125, 150), ylim = c(98, 138),
383       las = 1, pch = 21, col = 'dodgerblue4', bg = 'aliceblue')
384 k <- lm(face.HM ~ skull.pHM)$coef
385 x <- seq(min(skull.pHM), max(skull.pHM), length = 1000)
386 y <- k[1] + x * k[2]
387 lines(x, y, col = 'midnightblue', lwd = 2)
388
389 r <- round(r, digit = 4)
390 mtext('největší výška mozkovny (v mm)', side = 1, line = 2.3)
391 mtext(bquote(paste(rho == .(r))), side = 1, line = 3.7)

```



Obrázek 26: Tečkový diagram s lineární regresní přímkou pro největší výšku mozkovny a morfologickou výšku tváře mužů starověké egyptské populace (v mm)

## 7. Interpretace výsledků:

Na základě všech tří způsobů testování nezamítáme hypotézu o shodě korelačních koeficientů pro největší výšku mozkovny a morfologickou výšku tváře mužů starověké a novověké egyptské populace. Mezi korelačním koeficientem největší výšky mozkovny a morfologické výšky tváře mužů starověké a novověké egyptské populace neexistuje statisticky významný rozdíl. Mezi největší výškou mozkovny a morfologickou výškou tváře mužů starověké egyptské populace existuje nízký stupeň přímé lineární závislosti ( $\rho_1 = 0.33064$ ). Taktéž mezi největší výškou mozkovny a morfologickou výškou tváře mužů starověké egyptské populace existuje nízký stupeň přímé lineární závislosti ( $\rho_2 = 0.2510$ ) (viz stupnice míry závislosti pro Pearsonův korelační koeficient, kapitola ??).

★

### Příklad 7.18. Test o korelačním koeficientu $\rho$ (pravostranná alternativa)

Mějme datový soubor 13-two-samples-correlations-trunk.txt, proměnnou lowex.L popisující délku dolní končetiny (v mm) a proměnnou tru.L popisující délku trupu (v mm) (viz sekce ??). Na hladině významnosti  $\alpha = 0.01$  zjistěte, zda mezi délkou dolní končetiny a délkou trupu žen existuje přímá lineární závislost.

#### Řešení příkladu 7.18

Datový soubor načteme příkazem read.delim(). Pomocí operátoru [] vybereme z tabulky údaje o délce dolní končetiny (lowex.L) a délce trupu tru.L žen sex == 'f'. Údaje vložíme do proměnné data.F. Z tabulky data.F odstraníme chybějící údaje (na.omit()). Nakonec zjistíme rozsah náhodného výběru (dim()) a rozsahy naměřených hodnot obou proměnných (range()).

```
392 data <- read.delim('00-Data\\13-two-samples-correlations-trunk.txt')
393 data.F <- data[data$sex == 'f', c('lowex.L', 'tru.L')]
394 data.F <- na.omit(data.F)
395 dim(data.F) # 100x2
396
397 lowex.LF <- data.F$lowex.L
398 tru.LF <- data.F$tru.L
399 range(lowex.LF) # 836-1076
400 range(tru.LF) # 323-492
```

Datový soubor obsahuje údaje o délce dolní končetiny a délce trupu u 100 žen, přičemž naměřené délky dolní končetiny nabývají hodnot v rozmezí 836–1 076 mm a naměřené délky trupu nabývají hodnot v rozmezí 323–492 mm.

Naším úkolem je porovnat korelační koeficient populace žen s konstantou  $\rho_0 = 0$ . Řešení příkladu tedy vede na jednovýběrový test o korelačním koeficientu  $\rho$ . Předpokladem k použití tohoto testu je dvourozměrná normalita náhodného výběru délek dolní končetiny a délek trupu žen. Před použitím testu je třeba tento předpoklad ověřit.

Závěr o dvourozměrné normalitě obou náhodných výběrů stanovíme na základě Henzova-Zirklerova testu ( $\alpha = 0.05$ ) v kombinaci s grafickou vizualizací dat pomocí 3D grafu a tečkového diagramu 95 % elipsou spolehlivosti.

```
401 MVN::mvn(data.F, mvnTest = 'hz')$multivariateNormality # 0.587225
```

Protože  $p$ -hodnota Henzova-Zirklerova testu, tj. 0.587225, je větší než 0.05, hypotézu o dvourozměrné normalitě náhodného výběru délek dolních končetin a délek trupu žen nezamítáme na hladině významnosti  $\alpha = 0.05$ . Ke stejnému závěru bychom došli také použitím Mardiova testu ( $p$ -hodnota pro koeficient šikmosti = 0.1769 > 0.05,  $p$ -hodnota pro koeficient špičatosti = 0.8360 > 0.05) i Roystonova testu ( $p$ -hodnota = 0.6339 > 0.05).

Nyní se podíváme na grafickou vizualizaci náhodného výběru (viz graf 27) prostřednictvím 3D grafu obarveného 20 odstíny z palety matlab.like2. Odstíny tentokrát používáme v obráceném pořadí (viz funkce rev()).

3D graf nám ukazuje kopcovitý tvar s několika málo odlehlými hodnotami. Vzhledem k tomu, že rozsah náhodného výběru je 100, je třeba, aby elipsa spolehlivosti pokrývala alespoň 95 bodů. Zbýlých 5 bodů se smí realizovat mimo 95 % elipsu spolehlivosti. Z tečkového diagramu vidíme, že mimo elipsu spolehlivosti se realizují pouze 4 body. Na základě Henzova-Zirklerova testu a grafické vizualizace docházíme k závěru, že náhodný výběr délek dolních končetin a délek trupu pochází z dvourozměrného normálního rozdělení.

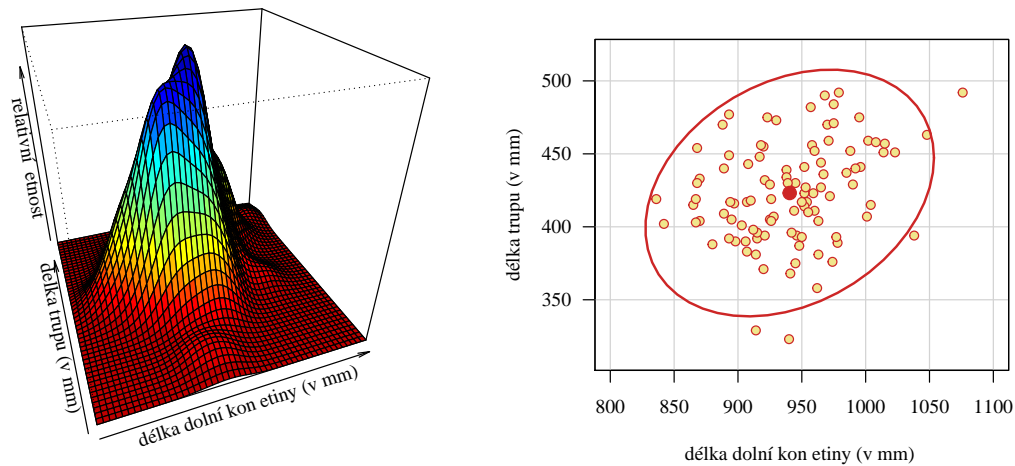
Jelikož je předpoklad dvourozměrné normality náhodného výběru splněn, můžeme použít parametrický test o korelačním koeficientu  $\rho$ . Naším úkolem zjistit, zda mezi délkou dolní končetiny a délkou trupu žen existuje statisticky významná přímá lineární závislost. Protože ukazatelem přímé závislosti je kladná hodnota korelačního koeficientu, je naším úkolem zjistit, zda je korelační koeficient délky dolní končetiny a délky trupu žen větší než 0. Analogicky jako v předchozích příkladech je toto tvrzení je zněním alternativní hypotézy. Nulovou hypotézu dodefinujeme jako doplněk k tomuto tvrzení.

#### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Korelační koeficient proměnných délka dolní končetiny a délka trupu je menší nebo rovný 0.

$H_1$  : Korelační koeficient proměnných délka dolní končetiny a délka trupu je větší než 0.



Obrázek 27: 3D graf a tečkový diagram s 95% elipsou spolehlivosti pro délku dolní končetiny a délku trupu žen (v mm)

- **matematická formulace** nulové a alternativní hypotézy

$$H_0 : \rho \leq \rho_0, \text{ kde } \rho_0 = 0$$

$$H_1 : \rho > \rho_0, \text{ kde } \rho_0 = 0$$

(pravostranná alternativa)

## 2. Volba hladiny významnosti

- Hladinu významnosti volíme podle zadání  $\alpha = 0.01$ .

## 3. Testování kritickým oborem

- Fisherova  $Z$ -transformace výběrového korelačního koeficientu

K výpočtu Fisherovy  $Z$ -transformace zjistíme nejprve hodnotu výběrového korelačního koeficientu pomocí příkazu `cor()`. Výběrový korelační koeficient délky dolní končetiny a délky trupu žen  $r = 0.2853$ . Nyní můžeme vypočítat Fisherovu  $Z$ -transformaci korelačního koeficientu a konstanty  $\rho_0 = 0$  z nulové hypotézy.

$$\begin{aligned} Z_R &= \frac{1}{2} \ln \frac{1+R}{1-R} \\ &= \frac{1}{2} \ln \frac{1+0.285256}{1-0.285256} = \frac{1}{2} \ln \frac{1.285256}{0.714744} = \frac{1}{2} \ln 1.798205 \\ &= 0.5 \times 0.5867888 = 0.2933944 \doteq 0.2934 \end{aligned}$$

$$\begin{aligned} \xi_0 &= \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0} \\ &= \frac{1}{2} \ln \frac{1+0}{1-0} = \frac{1}{2} \ln 1 \\ &= 0.5 \times 0 = 0 \end{aligned}$$

```
402 alpha <- 0.01
403 n <- length(lowex.LF)
404 rho0 <- 0
405 r <- cor(lowex.LF, tru.LF, method = 'pearson')
406 zR <- 1 / 2 * log((1 + r) / (1 - r)) # 0.2933943
407 ksi0 <- 1 / 2 * log((1 + rho0) / (1 - rho0)) # 0
```

- Testovací statistika

$$\begin{aligned}
 Z_W &= \frac{Z_R - \xi_0}{\sqrt{n-3}} \\
 &= (0.2933943 - 0)\sqrt{100-3} \\
 &= 0.2933943 \times \sqrt{97} \\
 &= 0.2933943 \times 9.848858 \\
 &= 2.889599 \doteq 2.8896
 \end{aligned}$$

```
408 Zw <- (zR - ksi0) * sqrt(n - 3) # 2.889599
```

- Kritický obor

$$\begin{aligned}
 W &= \langle u_{1-\alpha}; \infty \rangle \\
 &= \langle u_{1-0.01}; \infty \rangle \\
 &= \langle u_{0.99}; \infty \rangle \\
 &= \langle 2.326348; \infty \rangle
 \end{aligned}$$

```
409 qnorm(1 - alpha) # 2.326348
```

- Závěr testování

Protože realizace testovací statistiky  $z_W = 2.8896$  náleží do kritického oboru, tj.  $z_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

#### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (d, 1) &= \left( \tanh \left( z_R - \frac{u_{1-\alpha}}{\sqrt{n-3}} \right); 1 \right) \\
 &= \left( \tanh \left( 0.2933943 - \frac{u_{1-0.01}}{\sqrt{100-3}} \right); 1 \right) \\
 &= \left( \tanh \left( 0.2933943 - \frac{u_{0.99}}{\sqrt{97}} \right); 1 \right) \\
 &= \left( \tanh \left( 0.2933943 - \frac{2.326348}{9.848858} \right); 1 \right) \\
 &= (\tanh(0.2933943 - 0.2362048); 1) \\
 &= (\tanh(0.0571895); 1) \\
 &= (0.05712723; 1)
 \end{aligned}$$

```
410 dh <- tanh(zR - qnorm(1 - alpha) / sqrt(n - 3)) # 0.05712723
```

- Závěr testování

Protože  $\rho_0 = 0$  nenáleží do Waldova 99% empirického levostranného intervalu spolehlivosti, tj.  $\rho_0 = 0 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned} p\text{-hodnota} &= 1 - \Pr(Z_W \leq z_w) \\ &= 1 - \Pr(Z_W \leq 2.889599) \\ &= 0.001928667 \doteq 0.001929 \end{aligned}$$

```
411 1 - pnorm(Zw) # 0.001928667
```

- Závěr testování

Protože  $p$ -hodnota = 0.001929 je menší než  $\alpha = 0.01$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

## 6. Grafická vizualizace výsledků testování

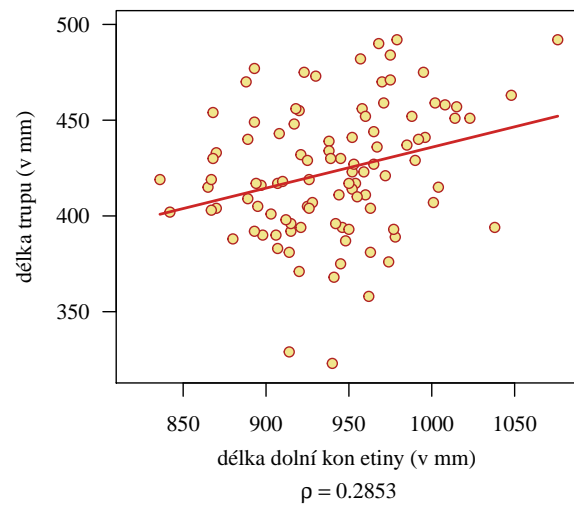
Míru závislosti mezi délkou dolní končetiny a délkou trupu vizualizujeme tečkovým diagramem superponovaným lineární regresní přímkou, která prokládá zobrazené body. Nejprve vykreslíme body délky dolní končetiny a délky trupu (příkaz `plot()`). Koeficienty lineární regresní přímky získáme pomocí funkce `lm()`, jejímž argumentem bude vztah `tru.LF ~ lowex.LF`, tj. vztah vyjadřující závislost mezi proměnnou `tru.LF` na ose  $y$  a proměnnou `lowex.LF` na ose  $x$ . Koeficienty regresní přímky získáme z výstupu funkce `lm` pomocí odkazu `$coef`. Dále vytvoříme posloupnost tisíce bodů  $x$  v rozsahu hodnot proměnné `lowex.LF` a vypočítáme hodnoty regresní přímky v bodech posloupnosti  $x$ , které vložíme do proměnné  $y$ .

```
412 par(mar = c(5, 4, 1, 1), family = 'Times')
413 plot(lowex.LF, tru.LF, xlab = '', ylab = 'délka trupu (v mm)',
414       xlim = c(820, 1080), ylim = c(320, 500),
415       las = 1, pch = 21, col = 'firebrick', bg = 'khaki')
416 k <- lm(tru.LF ~ lowex.LF)$coef
417 x <- seq(min(lowex.LF), max(lowex.LF), length = 1000)
418 y <- k[1] + x * k[2]
419 lines(x, y, col = 'firebrick3', lwd = 2)
420
421 r <- round(r, digit = 4)
422 mtext('délka dolní končetiny (v mm)', side = 1, line = 2.3)
423 mtext(bquote(paste(rho == .(r))), side = 1, line = 3.7)
```

## 7. Interpretace výsledků:

Na základě všech tří způsobů testování zamítáme hypotézu  $H_0$ . Korelační koeficient délky dolní končetiny a délky trupu žen je statisticky významně větší než 0. To znamená, že mezi délkou dolní končetiny a délkou trupu žen existuje statisticky významná přímá závislost. Interpretací výběrového korelačního koeficientu můžeme stanovit, že mezi oběma znaky existuje nízký stupeň přímé lineární závislosti ( $\rho_1 = 0.2853$ ), který je statisticky významný.





Obrázek 28: Tečkový diagram s lineární regresní přímkou pro délku dolní končetiny a délku trupu žen (v mm)



### Příklad 7.19. Test o korelačním koeficientu $\rho$ (levostranná alternativa)

Mějme datový soubor 06-lin-uhl-fm.txt, proměnnou skull.H popisující výšku lebky v mm a proměnnou base.B popisující šířku lebeční báze na spojnici obou bodů porion v mm (viz sekce ??). Na hladině významnosti  $\alpha = 0.10$ , zda mezi výškou lebky a šířkou lebeční báze žen existuje nepřímá závislost.

#### Řešení příkladu 7.19

Nejprve načteme datový soubor a vybereme z něj údaje o výšce lebky (skull.H) a šířce lebeční báze base.B žen sex == 'f'. Údaje vložíme do proměnné data.F. Z tabulky data.F následně odstraníme NA hodnoty a zjistíme rozsah náhodného výběru a rozsahy naměřených hodnot obou proměnných.

```
424 data <- read.delim('00-Data\\06-lin-uhl-fm.txt')
425 data.F <- data[data$sex == 'f', c('skull.H', 'base.B')]
426 data.F <- na.omit(data.F)
427 dim(data.F) # 20x2
428 skull.HF <- data.F$skull.H
429 base.BF <- data.F$base.B
430 range(skull.HF) # 115.4675-142.0617
431 range(base.BF) # 108.5226-123.0860
```

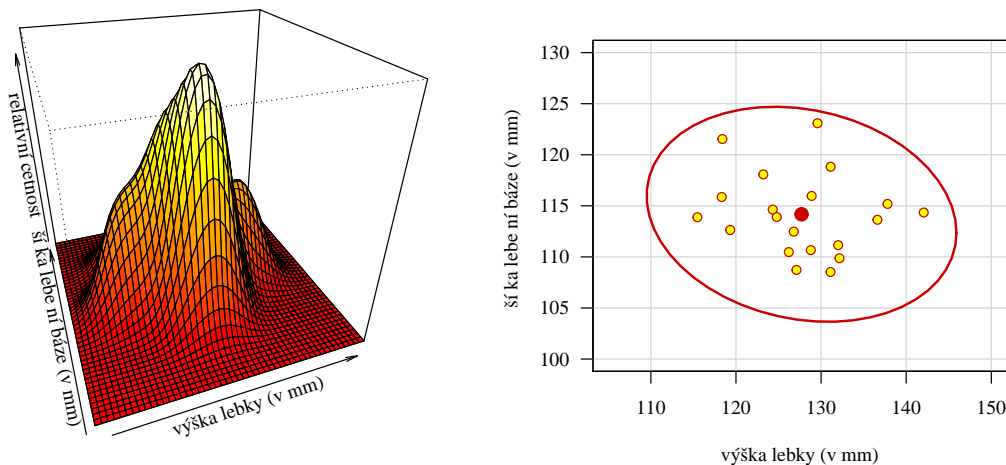
Datový soubor obsahuje údaje o výšce lebky a šířce lebeční báze u 20 žen, přičemž naměřené výšky lebky nabývají hodnot v rozmezí 115.46–142.062 mm a naměřené šířky lebeční báze nabývají hodnot v rozmezí 108.523–123.086 mm.

Naším úkolem je porovnat korelační koeficient populace žen s konstantou  $\rho_0 = 0$ . Řešení příkladu tedy vede na jednovýběrový test o korelačním koeficientu  $\rho$ . Před použitím tohoto testu je třeba ověřit předpoklad normality náhodného výběru výšek lebky a šířek lebeční báze žen. Předpoklad normality ověříme Roystonovým testem ( $\alpha = 0.05$ ) v kombinaci s 3D grafem a tečkovým diagramem s 95% elipsou spolehlivosti.

```
432 MVN::mvn(data.F, mvnTest = 'royston')$multivariateNormality # 0.6845014
```

Protože  $p$ -hodnota Roystonova testu, tj. 0.6845, je větší než 0.05, hypotézu o dvourozměrné normalitě náhodného výběru výšek lebky a šířek lebeční báze žen nezamítáme na hladině významnosti  $\alpha = 0.05$ . Ke stejnému závěru bychom došli také použitím Mardiova testu ( $p$ -hodnota pro koeficient šikmosti = 0.7607 > 0.05,  $p$ -hodnota pro koeficient špičatosti = 0.3897 > 0.05) i Henzova-Zirklerova testu ( $p$ -hodnota = 0.3559 > 0.05).

Nyní se podíváme na grafickou vizualizaci náhodného výběru (viz graf 29).



Obrázek 29: 3D graf a tečkový diagram s 95% elipsou spolehlivosti pro výšku lebky a šířku lebeční báze žen (v mm)

Grafy na obrázku 29 podporují výsledek testování. 3F graf zobrazuje normální rozdělení náhodného výběru tvořené jedním pospolitým kopcem bez výraznějších odlehklých pozorování. Z tečkového diagramu je navíc patrné,

že všechny body leží uvnitř elipsy spolehlivosti. Náhodný výběr výšek lebky a šířek lebeční báze žen pochází z dvourozměrného normálního rozdělení.

Jelikož je předpoklad dvourozměrné normality náhodného výběru splněn, můžeme použít parametrický test o korelačním koeficientu  $\rho$ . Naším úkolem zjistit, zda mezi výškou lebky a šířkou lebeční báze žen existuje statisticky významná nepřímá lineární závislost. Protože ukazatelem nepřímé závislosti je záporná hodnota korelačního koeficientu, je naším úkolem zjistit, zda je korelační koeficient výšky lebky a šířky lebeční báze žen menší než 0. Analogicky jako v předchozích příkladech je toto tvrzení je zněním alternativní hypotézy. Nulovou hypotézu dodefinujeme jako doplněk k tomuto tvrzení.

### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Korelační koeficient výšky lebky a šířky lebeční báze žen je větší nebo rovný 0.

$H_1$  : Korelační koeficient výšky lebky a šířky lebeční báze žen je menší než 0.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $\rho \geq \rho_0$ , kde  $\rho_0 = 0$

$H_1$  :  $\rho < \rho_0$ , kde  $\rho_0 = 0$

(levostranná alternativa)

### 2. Volba hladiny významnosti

- Hladinu významnosti volíme v souladu se zadáním  $\alpha = 0.10$ .

### 3. Testování kritickým oborem

- Fisherova  $Z$ -transformace výběrového korelačního koeficientu

Nejprve vypočítáme hodnotu výběrového korelačního koeficientu  $R$  výšky lebky a šířky lebeční báze ( $r = -0.1713$ ). Nyní můžeme vypočítat Fisherovu  $Z$ -transformaci korelačního koeficientu a Fisherovu  $Z$ -transformaci konstanty  $\rho_0 = 0$ .

$$\begin{aligned} Z_R &= \frac{1}{2} \ln \frac{1+R}{1-R} \\ &= \frac{1}{2} \ln \frac{1+(-0.1712964)}{1-(-0.1712964)} = \frac{1}{2} \ln \frac{0.8287036}{1.171296} = \frac{1}{2} \ln 0.70751 \\ &= 0.5 \times (-0.3460036) = -0.1730018 \doteq -0.1730 \end{aligned}$$

$$\begin{aligned} \xi_0 &= \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0} \\ &= \frac{1}{2} \ln \frac{1+0}{1-0} = \frac{1}{2} \ln 1 = 0.5 \times 0 = 0 \end{aligned}$$

```
433 alpha <- 0.10
434 n     <- length(skull.HF)
435 rho0  <- 0
436 r     <- cor(skull.HF, base.BF, method = 'pearson') # -0.1712964
437 zR    <- 1 / 2 * log((1 + r) / (1 - r))           # -0.173002
438 ksi0  <- 1 / 2 * log((1 + rho0) / (1 - rho0)) # 0
```

- **Testovací statistika**

$$\begin{aligned}
Z_W &= \frac{Z_R - \xi_0}{\sqrt{n-3}} \\
&= (-0.173002 - 0)\sqrt{20-3} \\
&= -0.173002 \times \sqrt{17} \\
&= -0.173002 \times 4.123106 \\
&= -0.7133056 \doteq -0.7133
\end{aligned}$$

```
439 Zw <- (zR - ksi0) * sqrt(n - 3) # -0.7133054
```

- Kritický obor

$$\begin{aligned}
W &= (-\infty; u_\alpha) \\
&= (-\infty; u_{0.10}) \\
&= (-\infty; -1.281552)
\end{aligned}$$

```
440 qnorm(alpha) # -1.281552
```

- Závěr testování

Protože realizace testovací statistiky  $z_W = -0.7133$  nenáleží do kritického oboru, tj.  $z_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.10$ .

#### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
(-1, h) &= \left(-1; \tanh\left(z_R - \frac{u_\alpha}{\sqrt{n-3}}\right)\right) \\
&= \left(-1; \tanh\left(-0.173002 - \frac{u_{0.10}}{\sqrt{20-3}}\right)\right) \\
&= \left(-1; \tanh\left(-0.173002 - \frac{-1.281552}{4.123106}\right)\right) \\
&= (-1; \tanh(-0.173002 - (-0.310822))) \\
&= (-1; \tanh(0.13782)) \\
&= (-1; 0.136954)
\end{aligned}$$

```
441 hh <- tanh(zR - qnorm(alpha) / sqrt(n - 3)) # 0.1369539
```

- Závěr testování

Protože  $\rho_0 = 0$  náleží do Waldova 90% empirického pravostranného intervalu spolehlivosti, tj.  $\rho_0 = 0 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.10$ .

#### 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned}
 p\text{-hodnota} &= \Pr(Z_w \leq z_w) \\
 &= \Pr(Z_w \leq -0.7133054) \\
 &= 0.2378284 \doteq 0.2378
 \end{aligned}$$

```
442 pnorm(Zw) # 0.2378284
```

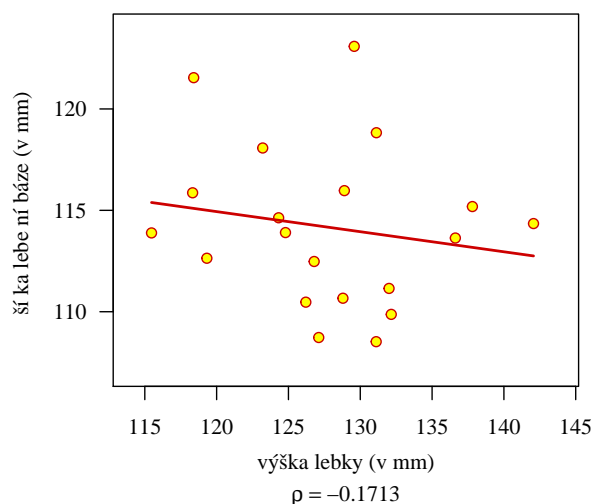
- **Závěr testování**

Protože  $p$ -hodnota = 0.2378 je větší než  $\alpha = 0.10$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.10$ .

## 6. Grafická vizualizace výsledků testování

Míru závislosti mezi výškou lebky a šířkou lebeční báze vizualizujeme tečkovým diagramem superponovaným lineární regresní přímkou, která prokládá zobrazené body. Koeficienty lineární regresní přímky získáme pomocí funkce `lm()`, jejímž argumentem bude vztah `base.BF ~ skull.HF`, tj. vztah vyjadřující závislost mezi proměnnou `base.BF` na ose `y` a proměnnou `skull.HF` na ose `x`. Dále vytvoříme posloupnost bodů `x` v rozsahu hodnot proměnné `skull.HF` a vypočítáme hodnoty regresní přímky v bodech této posloupnosti.

```
443 par(mar = c(5, 4, 1, 1), family = 'Times')
444 plot(skull.HF, base.BF, xlab = '', ylab = 'šířka lebeční báze (v mm)',
445       xlim = c(114, 144), ylim = c(107, 124),
446       las = 1, pch = 21, col = 'red3', bg = 'yellow')
447 k <- lm(base.BF ~ skull.HF)$coef
448 x <- seq(min(skull.HF), max(skull.HF), length = 1000)
449 y <- k[1] + x * k[2]
450 lines(x, y, col = 'red3', lwd = 2)
451
452 r <- round(r, digit = 4)
453 mtext('výška lebky (v mm)', side = 1, line = 2.3)
454 mtext(bquote(paste(rho == .(r))), side = 1, line = 3.7)
```



Obrázek 30: Tečkový diagram s lineární regresní přímkou pro výšku lebky a šířku lebeční báze žen (v mm)

## 7. Interpretace výsledků:

Na základě všech tří způsobů testování nezamítáme hypotézu  $H_0$ . Korelační koeficient výšky lebky a šířky

lebeční báze žen není statisticky významně menší než 0. To znamená, že mezi výškou lebky a šířkou lebeční báze neexistuje statisticky významná nepřímá závislost. Interpretací výběrového korelačního koeficientu můžeme stanovit, že mezi oběma znaky existuje nízký stupeň nepřímé lineární závislosti ( $\rho_1 = -0.1713$ ), který není statisticky významný.



### Příklad 7.20. Test o korelačním koeficientu $\rho$ (nesplnění předpokladů)

Mějme datový soubor 16-anova-head.txt, proměnnou bigo.W popisující šířku dolní čelisti v mm a proměnnou bizyg.W popisující šířku tváře v mm (viz sekce ??). Na hladině významnosti  $\alpha = 0.10$  testujte hypotézu o nepřímé závislosti mezi šířkou dolní čelisti a šířkou tváře u mužů.

#### Řešení příkladu 7.20

Nejprve načteme datový soubor a vybereme z něj údaje o šířce dolní čelisti (bigo.W) a šířce tváře bizyg.W mužů sex == 'm'. Údaje vložíme do proměnné data.M. Z tabulky data.M následně odstraníme chybějící údaje, zjistíme rozsah náhodného výběru a rozsahy naměřených hodnot obou proměnných.

```
455 data <- read.delim('00-Data\\16-anova-head.txt')
456 data.M <- data[data$sex == 'm', c('bigo.W', 'bizyg.W')]
457 data.M <- na.omit(data.M)
458 dim(data.M) # 75x2
459 bigo.WM <- data.M$bigo.W
460 bizyg.WM <- data.M$bizyg.W
461 range(bigo.WM) # 94 126
462 range(bizyg.WM) # 113-155
```

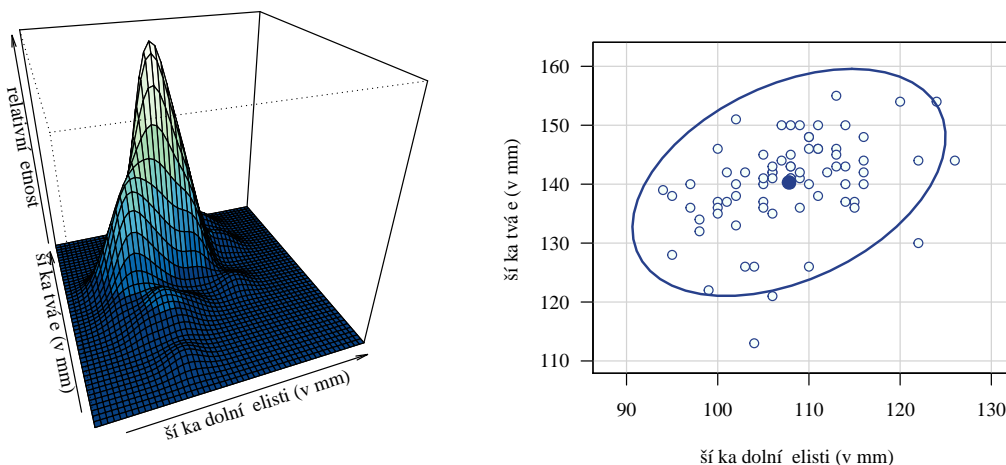
Datový soubor obsahuje údaje o šířce dolní čelisti a šířce tváře 75 mužů, přičemž naměřené šířky dolní čelisti nabývají hodnot v rozmezí 94–126 mm a naměřené šířky tváře nabývají hodnot v rozmezí 113–155 mm.

Naším úkolem je porovnat korelační koeficient populace mužů s konstantou  $\rho_0 = 0$ . Řešení příkladu vede na jednovýběrový test o korelačním koeficientu  $\rho$ . Před použitím tohoto testu je třeba ověřit předpoklad normality náhodného výběru šířek dolní čelisti a šířek tváře mužů. Předpoklad normality ověříme Henzovým-Zirklerovým testem ( $\alpha = 0.05$ ) v kombinaci s 3D grafem a tečkovým diagramem s 95 % elipsou spolehlivosti.

```
463 MVN::mvn(data.M, mvnTest = 'hz')$multivariateNormality # 0.006343478
```

Protože  $p$ -hodnota Henzova-Zirklerova testu, tj. 0.006343, je menší než 0.05, hypotézu o dvourozměrné normalitě náhodného výběru šířek dolní čelisti a šířek tváře mužů zamítáme na hladině významnosti  $\alpha = 0.05$ . Ke stejnému závěru bychom došli použitím Mardiova testu ( $p$ -hodnota pro koeficient šikmosti = 0.002218 < 0.05,  $p$ -hodnota pro koeficient špičatosti = 0.05248 > 0.05) i Roystonova testu ( $p$ -hodnota = 0.006445 < 0.05).

Před stanovením závěru o dvourozměrné normalitě vyhodnotíme také grafickou vizualizace náhodného výběru (viz graf 31).



Obrázek 31: 3D graf a tečkový diagram s 95% elipsou spolehlivosti pro výšku lebky a šířku lebeční báze žen (v mm)

3D graf zobrazuje dvourozměrné normální rozdělení tvořené jedním vrcholem s několika odlehlými hodnotami a jedním výrazně odlehlým pozorováním. Zda je počet odlehlých pozorování unosný pro předpoklad normality

ověříme na základě tečkového diagramu. Aby data pocházela z dvourozměrného normálního rozdělení, je potřeba, aby alespoň 71 bodů (95% bodů) leželo uvnitř elipsy spolehlivosti. Zbylé čtyři body mohou ležet mimo elipsu spolehlivosti. Bohužel mimo elipsu spolehlivosti se tentokrát nachází 5 bodů. Náhodný výběr šířek dolní čelisti a šířek lebeční báze žen nepochází z dvourozměrného normálního rozdělení.

Jelikož předpoklad dvourozměrné normality náhodného výběru není splněn, nemůžeme k otestování hypotézy ze zadání použít jednovýběrový test o korelačním koeficientu  $\rho$ .

*Poznámka:* V současné chvíli se nabízí více postupů, jak dále řešit nastalou situaci. Jedním z nich by bylo odstranění nejodlehlejšího pozorování a následná kontrola, zda se odstraněním tohoto pozorování nespravila normalita náhodného výběru. V případě, že ano, můžeme nový datový soubor s rozsahem 74 hodnot použít k otestování hypotézy ze zadání pomocí parametrického testu o korelačním koeficientu  $\rho$ , analogicky jako v příkladech 7.17, 7.18 a 7.19. V případě, že odstranění nevedlo k zlepšení normality náhodného výběru, mohli bychom vyzkoušet odstranění dalších odlehlých pozorování a to až do výše 10% naměřených hodnot. Pokud předpoklad normality stále není splněn, vrátíme se k původnímu souboru 75 naměřených hodnot a k otestování hypotézy ze zadání použijeme neparametrický test o korelačním koeficientu založený na Spearmanově koeficientu pořadové korelace (viz kapitola ??). ★

## 7.6 Test o nulovém korelačním koeficientu $\rho$ (Test o nezávislosti)

Nechť  $(X_1, Y_1)^T \dots (X_n, Y_n)^T$  je náhodný výběr z  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Na hladině významnosti  $\alpha$  testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

$$\begin{array}{lll} H_{01} : \rho = 0 & \text{oproti} & H_{11} : \rho \neq 0 \quad (\text{oboustranná alt.}) \\ H_{02} : \rho \leq 0 & \text{oproti} & H_{12} : \rho > 0 \quad (\text{pravostranná alt.}) \\ H_{03} : \rho \geq 0 & \text{oproti} & H_{13} : \rho < 0 \quad (\text{levostranná alt.}) \end{array}$$

Test nazýváme jednovýběrovým testem o nezávislosti. Testovací statistika má tvar

$$T_W = \frac{\sqrt{n-2}R}{\sqrt{1-R^2}} \quad (7.5)$$

kde  $R$  je výběrový korelační koeficient  $n$  je rozsah náhodného výběru. Testovací statistika  $T_W$  pochází ze Studentova  $t$ -rozdělení o  $n-2$  stupních volnosti, tj.

$$T_W = \frac{\sqrt{n-2}R}{\sqrt{1-R^2}} \sim t_{n-2}.$$

Kritický obor podle zvolené alternativní hypotézy má tvar

$$\begin{array}{ll} H_{11} : \rho \neq 0 & W = (-\infty; t_{n-2}(\alpha/2)) \cup (t_{n-2}(1-\alpha/2); \infty) \\ H_{12} : \rho > 0 & W = (t_{n-2}(1-\alpha); \infty) \\ H_{13} : \rho < 0 & W = (-\infty; t_{n-2}(\alpha)) \end{array}$$

kde  $t_{n-2}(\alpha/2)$ ,  $t_{n-2}(1-\alpha/2)$ ,  $t_{n-2}(\alpha)$ ,  $t_{n-2}(1-\alpha)$  jsou kvantily Studentova rozdělení o  $n-2$  stupních volnosti, jejichž hodnoty získáme pomocí  $\mathcal{R}$  a implementované funkce  $qt()$ .

Interval spolehlivosti má podle zvolené alternativní hypotézy jeden z následujících tvarů

$$\begin{array}{ll} H_{11} : \rho \neq 0 & (d, h) = \left( \frac{t_{n-2}(\alpha/2)}{\sqrt{t_{n-2}^2(\alpha/2)+n-2}}; \frac{t_{n-2}(1-\alpha/2)}{\sqrt{t_{n-2}^2(1-\alpha/2)+n-2}} \right) \\ H_{12} : \rho > 0 & (-1, h) = \left( -1; \frac{t_{n-2}(1-\alpha)}{\sqrt{t_{n-2}^2(1-\alpha)+n-2}} \right) \\ H_{13} : \rho < 0 & (d, 1) = \left( \frac{t_{n-2}(\alpha)}{\sqrt{t_{n-2}^2(\alpha)+n-2}}; 1 \right) \end{array}$$

*Poznámka:* Při testování hypotézy o nulovém korelačním koeficientu  $\rho$  pomocí intervalu spolehlivosti rozhodujeme o zamítnutí nebo nezamítnutí  $H_0$  v závislosti na tom, zda hodnota výběrového korelačního koeficientu  $R$  náleží nebo nenáleží do intervalu spolehlivosti, nikoli v závislosti na tom, zda  $\rho_0 = 0$  náleží nebo nenáleží do intervalu spolehlivosti. Jedná se o výjimku.

*Poznámka:* Protože parametr  $\rho$  je korelační koeficient, platí, že  $\rho \in (-1; 1)$ . Proto levostranný interval spolehlivosti omezíme shora hodnotou 1, namísto nekonečnem, a pravostranný interval spolehlivosti omezíme zdola hodnotou -1, namísto mínus nekonečnem.

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

$$\begin{array}{ll} H_{11} : \rho \neq 0 & p\text{-hodnota} = 2 \min\{\Pr(T_W \leq t_W), \Pr(T_W > t_W)\} \\ H_{12} : \rho > 0 & p\text{-hodnota} = \Pr(T_W > t_W) = 1 - \Pr(T_W \leq t_W) \\ H_{13} : \rho < 0 & p\text{-hodnota} = \Pr(T_W \leq t_W) \end{array}$$

kde  $T_W$  je náhodná veličina,  $t_W$  je realizace testovací statistiky  $T_W$  (viz vzorec 7.5), tedy konkrétní číslo, a  $\Pr(T_W \leq t_W)$  je distribuční funkce Studentova rozdělení o  $n-2$  stupních volnosti, jejíž hodnotu získáme pomocí  $\mathcal{R}$  a implementované funkce  $pt()$ .



### Příklad 7.21. Test o korelačním koeficientu $\rho$ (oboustranná alternativa)

Mějme datový soubor 19-more-samples-correlations-skull.txt, proměnnou nose.B popisující šířku nosu a proměnnou intorb.B popisující interorbitální šířku (viz sekce ??) malajské populace. Na hladině významnosti  $\alpha = 0.05$  testujte hypotézu o nezávislosti šířky nosu a interorbitální šířky u mužů malajské populace.

#### Řešení příkladu 7.21

Datový soubor načteme příkazem read.delim(). Pomocí operátoru [] vybereme z tabulky údaje o šířce nosu (nose.B) a interorbitální šířce intorb.B u mužů malajské populace pop == 'mal'. Údaje vložíme do proměnné data.M. Pomocí příkazu na.omit() odstraníme z tabulky data.M chybějící údaje. Nakonec příkazem dim() stanovíme rozsah náhodného výběru a příkazem range() zjistíme rozsahy naměřených hodnot obou proměnných.

```
464 data <- read.delim('00-Data\\19-more-samples-correlations-skull.txt')
465 data.M <- data[data$pop == 'mal', c('nose.B', 'intorb.B')]
466 data.M <- na.omit(data.M)
467 dim(data.M) # 73x2
468 nose.BM <- data.M$nose.B
469 intorb.BM <- data.M$intorb.B
470 range(nose.BM) # 21-30
471 range(intorb.BM) # 18-29
```

Datový soubor obsahuje údaje o šířce nosu a interorbitální šířce 73 mužů malajské populace, přičemž naměřené šířky nosu nabývají hodnot v rozmezí 21–30 mm a naměřené interorbitální šířky nabývají hodnot v rozmezí 18–29 mm.

Naším úkolem je porovnat korelační koeficient malajské populace s konstantou  $\rho_0 = 0$ , která reprezentuje nezávislost obou znaků. U mužů malajské populace máme k dispozici náhodný výběr naměřených hodnot. Na základě těchto hodnot můžeme zjistit, zda náhodný výběr pochází z dvourozměrného normálního rozdělení. Řešení příkladu vede na jednovýběrový test o nulovém korelačním koeficientu (neboli na test o nezávislosti). Předpokladem k použití tohoto testu je dvourozměrná normalita náhodného výběru šířek nosu a interorbitálních šířek mužů malajské populace. Před použitím testu je třeba tento předpoklad ověřit.

Závěr o dvourozměrné normalitě obou náhodných výběrů stanovíme na základě Henzova-Zirklerova testu ( $\alpha = 0.05$ ) v kombinaci s 3D grafem a tečkovým diagramem superponovaným 95 % elipsou spolehlivosti, analogicky, jako je uvedeno v sekci ??.

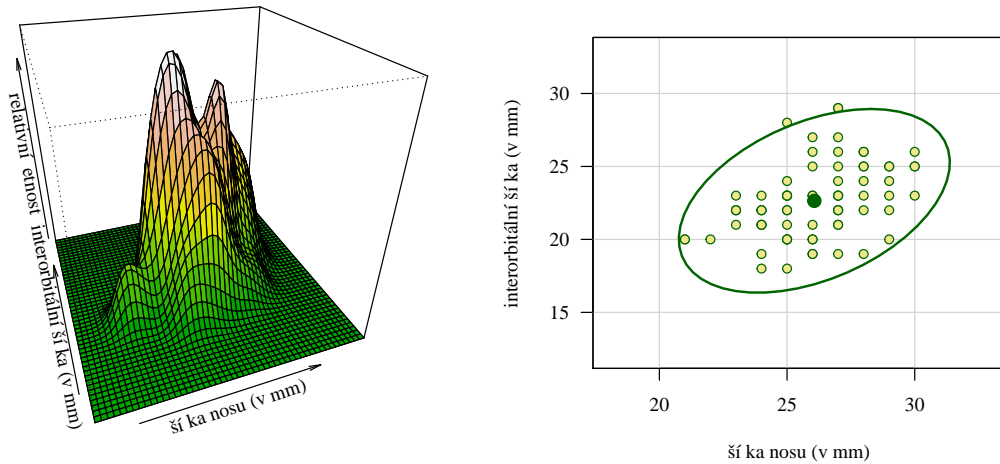
```
472 MVN::mvn(data.M, mvnTest = 'hz')$multivariateNormality # 0.269426
```

Protože  $p$ -hodnota Henzova-Zirklerova testu, tj. 0.2694, je větší než 0.05, hypotézu o dvourozměrné normalitě náhodného výběru šířek nosu a interorbitálních šířek nezamítáme na hladině významnosti  $\alpha = 0.05$ . Ke stejnému závěru bychom došli použitím Mardiova testu ( $p$ -hodnota pro koeficient šikmosti = 0.54502 > 0.05,  $p$ -hodnota pro koeficient špičatosti = 0.2638 > 0.05). Na základě Rystonova testu bychom dvourozměrnou normalitu náhodného výběru zamítli ( $p$ -hodnota = 0.02814 < 0.05). Před vytvořením závěru o dvourozměrné normalitě zhodnotíme ještě grafickou vizualizaci náhodného výběru (viz graf 32).

3D graf zobrazuje pospolitý kopcovitý tvar náhodného výběru bez viditelných odlehlých pozorování. Na tečkovém diagramu vidíme, že přeci jen dvě odlehlá pozorování jsou součástí náhodného výběru. Vzhledem k tomu, že pro podpoření předpokladu normality stačí, aby mimo elipsu spolehlivosti ležely nejvýše tři body (alespoň 69 bodů se má realizovat uvnitř elipsy spolehlivosti), je normalita náhodného výběru z hlediska odlehlých pozorování v pořádku. Náhodný výběr šířek nosu a interorbitálních šířek u mužů malajské populace pochází z dvourozměrného normálního rozdělení.

Jelikož je předpoklad dvourozměrné normality náhodného výběru splněn, můžeme hypotézu ze zadání otestovat pomocí parametrického testu o nezávislosti. Naším úkolem otestovat (nulovou) hypotézu o nezávislosti šířky nosu a interorbitální šířky, tj. hypotézu o shodě korelačního koeficientu šířky nosu a interorbitální šířky s konstantou  $\rho = 0$ . Zbývá tedy stanovit znění alternativní hypotézy tak, aby bylo doplňkem k nulové hypotéze. Testování provedeme v posloupnosti sedmi kroků.

#### 1. Stanovení hypotéz



Obrázek 32: 3D graf a tečkový diagram s 95% elipsou spolehlivosti pro šířku nosu a interorbitální šířku mužů malajské populace (v mm)

- **slovní formulace** nulové a alternativní hypotézy  
 $H_0$  : Korelační koeficient proměnných šířka nosu a interorbitální šířka tváře je rovný 0.  
 $H_1$  : Korelační koeficient proměnných šířka nosu a interorbitální šířka tváře není rovný 0.
- **matematická formulace** nulové a alternativní hypotézy  
 $H_0$  :  $\rho = 0$   
 $H_1$  :  $\rho \neq 0$   
 (oboustranná alternativa)

## 2. Volba hladiny významnosti

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

## 3. Testování kritickým oborem

### • Testovací statistika

K výpočtu testovací statistiky potřebujeme nejprve stanovit hodnotu výběrového korelačního koeficientu. K tomu využijeme funkci `cor()` s argumentem `method = 'pearson'`. Hodnota výběrového korelačního koeficientu mezi šířkou nosu a interorbitální šířkou je  $r = 0.41753$ . Nyní již můžeme vypočítat hodnotu testovací statistiky

$$\begin{aligned}
 T_W &= \frac{\sqrt{n-2}R}{\sqrt{1-R^2}} \\
 &= \frac{\sqrt{73-2} \times 0.4175294}{\sqrt{1-0.4175294^2}} \\
 &= \frac{\sqrt{71} \times 0.4175294}{\sqrt{0.8256692}} \\
 &= \frac{8.42615 \times 0.4175294}{0.9086634} \\
 &= 3.871803 \doteq 3.8718
 \end{aligned}$$

```

473 alpha <- 0.05
474 r <- cor(nose.BM, intorb.BM, method = 'pearson')
475 n <- length(nose.BM)
476 tw <- sqrt(n - 2) * r / sqrt(1 - r ^ 2) # 3.871803

```

- Kritický obor

$$\begin{aligned}
 W &= (-\infty; t_{n-2}(\alpha/2)) \cup (t_{n-2}(1 - \alpha/2); \infty) \\
 &= (-\infty; t_{73-2}(0.05/2)) \cup (t_{73-2}(1 - 0.05/2); \infty) \\
 &= (-\infty; t_{71}(0.025)) \cup (t_{71}(0.975); \infty) \\
 &= (-\infty; -1.9939) \cup (1.9939; \infty)
 \end{aligned}$$

```

477 qt(alpha / 2, n - 2) #-1.993943
478 qt(1 - alpha / 2, n - 2) # 1.993943

```

- Závěr testování

Protože realizace testovací statistiky  $t_W = 3.8718$  náleží do kritického oboru, tj.  $t_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

#### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (d, h) &= \left( \frac{t_{n-2}(\alpha/2)}{\sqrt{t_{n-2}^2(\alpha/2) + n - 2}}; \frac{t_{n-2}(1 - \alpha/2)}{\sqrt{t_{n-2}^2(1 - \alpha/2) + n - 2}} \right) \\
 &= \left( \frac{t_{73-2}(0.05/2)}{\sqrt{t_{73-2}^2(0.05/2) + 73 - 2}}; \frac{t_{73-2}(1 - 0.05/2)}{\sqrt{t_{73-2}^2(1 - 0.05/2) + 73 - 2}} \right) \\
 &= \left( \frac{t_{71}(0.025)}{\sqrt{t_{71}^2(0.025) + 71}}; \frac{t_{71}(1 - 0.025)}{\sqrt{t_{71}^2(1 - 0.025) + 71}} \right) \\
 &= \left( \frac{-1.993943}{\sqrt{(-1.993943)^2 + 71}}; \frac{1.993943}{\sqrt{1.993943^2 + 71}} \right) \\
 &= \left( \frac{-1.993943}{\sqrt{74.97581}}; \frac{1.993943}{\sqrt{74.97581}} \right) \\
 &= \left( \frac{-1.993943}{8.658857}; \frac{1.993943}{8.658857} \right) \\
 &= (-0.2302779; 0.2302779)
 \end{aligned}$$

```

479 dh <- qt(alpha / 2, n - 2) / sqrt(qt(alpha / 2, n - 2) ^ 2 + n - 2) #
      -0.2302779
480 hh <- qt(1 - alpha / 2, n - 2) / sqrt(qt(1 - alpha / 2, n - 2) ^ 2 + n - 2) #
      0.2302779

```

- Závěr testování

Protože  $r = 0.4175$  nenáleží do Waldova 95% empirického oboustranného intervalu spolehlivosti, tj.  $r = 0.4175 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

#### 5. Testování $p$ -hodnotou

- *p*-hodnota

$$\begin{aligned}
 p\text{-hodnota} &= 2 \min\{\Pr(T_w \leq t_w), \Pr(T_W > t_w)\} \\
 &= 2 \min\{\Pr(T_w \leq t_w), 1 - \Pr(T_W \leq t_w)\} \\
 &= 2 \min\{\Pr(T_w \leq 3.871803), 1 - \Pr(T_W \leq 3.871803)\} \\
 &= 2 \min\{0.999946, 5.401663e - 05\} \\
 &= 2 \times 5.401663e - 05 \\
 &= 0.0001080333 \doteq 0.000108033
 \end{aligned}$$

```
481 2 * min(pnorm(tw), 1 - pnorm(tw)) # 0.0001080333
```

- **Závěr testování**

Protože *p*-hodnota = 0.0001080333 je menší než  $\alpha = 0.05$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

## 6. Grafická vizualizace výsledků testování

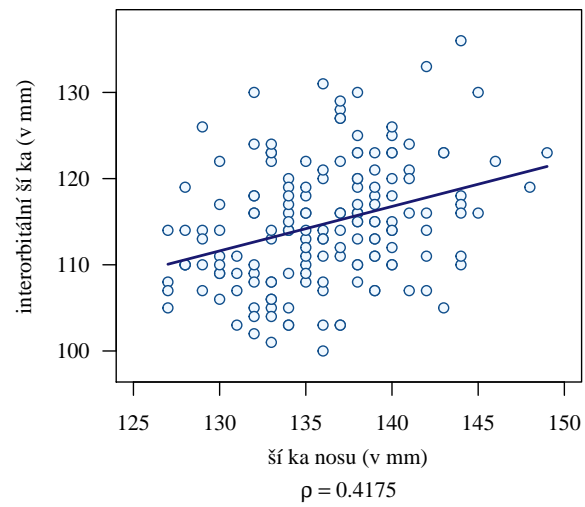
Míru závislosti mezi šířkou nosu a interorbitální šířkou vizualizujeme tečkovým diagramem superponovaným lineární regresní přímkou prokládající zobrazené body. Nejprve příkazem `plot()` vykreslíme body šířky nosu a interorbitální šířky mužů malajské populace. Koeficienty lineární regresní přímky získáme pomocí funkce `lm()`, jejímž jediným argumentem bude vztah `intorb.BM ~ nose.BM`, tj. vztah vyjadřující závislost mezi proměnnou `intorb.BM` na ose *y* a proměnnou `nose.BM` na ose *x*. Koeficienty regresní přímky, které jsou vloženy v poloze `coefficients`, získáme z výstupu funkce `lm` pomocí odkazu `$coef`. Dále vytvoříme posloupnost tisíce bodů *x* v rozsahu hodnot proměnné `nose.BM` a vypočítáme hodnoty regresní přímky v bodech posloupnosti *x*, které vložíme do proměnné *y*. Nakonec vykreslíme lineární regresní přímkou v bodech *x*, *y* příkazem `lines()`. Nakonec do grafu doplníme pod osu *x* popisek obsahující hodnotu výběrového korelačního koeficientu  $r = 0.4175$  (příkaz `mtext()`).

```
482 par(mar = c(5, 4, 1, 1), family = 'Times')
483 plot(skull.pHM, face.HM, xlab = '', ylab = 'interorbitální šířka (v mm)',
484       xlim = c(125, 150), ylim = c(98, 138),
485       las = 1, pch = 21, col = 'dodgerblue4', bg = 'aliceblue')
486 k <- lm(face.HM ~ skull.pHM)$coef
487 x <- seq(min(skull.pHM), max(skull.pHM), length = 1000)
488 y <- k[1] + x * k[2]
489 lines(x, y, col = 'midnightblue', lwd = 2)
490
491 r <- round(r, digit = 4)
492 mtext('šířka nosu (v mm)', side = 1, line = 2.3)
493 mtext(bquote(paste(rho == .(r))), side = 1, line = 3.7)
```

## 7. Interpretace výsledků:

Na základě všech tří způsobů testování zamítáme hypotézu o shodě korelačního koeficientu šířky nosu a interorbitální šířky u mužů malajské populace s nulou. Mezi šířkou nosu a interorbitální šířkou mužů malajské populace existuje statisticky významná závislost. Interpretací výběrového korelačního koeficientu zjistíme, že mezi šířkou nosu a interorbitální šířkou existuje mírný stupeň přímé lineární závislosti ( $\rho_2 = 0.4175$ ) (viz stupnice míry závislosti pro Pearsonův korelační koeficient, kapitola ??).





Obrázek 33: Tečkový diagram s lineární regresní přímkou pro šířku nosu a interorbitální šířku mužů malajské populace (v mm)

### Příklad 7.22. Test o korelačním koeficientu $\rho$ (pravostranná alternativa)

Mějme datový soubor 13-two-samples-correlations-trunk.txt, proměnnou lowex.L popisující délku dolní končetiny (v mm) a proměnnou tru.L popisující délku trupu (v mm) (viz sekce ??). Na hladině významnosti  $\alpha = 0.01$  zjistěte, zda mezi délkou dolní končetiny a délkou trupu žen existuje přímá lineární závislost.

#### Řešení příkladu 7.22

Zadání příkladu je shodné se zadáním příkladu 7.18. Taktéž datový soubor je totožný. Naším úkolem je porovnat korelační koeficient populace žen s konstantou  $\rho_0 = 0$ , tentokrát prostřednictvím jednovýběrového testu o nezávislosti. Předpokladem k použití tohoto testu je dvourozměrná normalita náhodného výběru délek dolní končetiny a délek trupu žen. Splnění tohoto předpokladu jsme ověřili v rámci řešení příkladu 7.18. Ze zadání příkladu víme, že máme zhodnotit, zda mezi délkou dolní končetiny a délkou trupu žen existuje statisticky významná přímá lineární závislost. Protože ukazatelem přímé závislosti je kladná hodnota korelačního koeficientu, je naším úkolem zjistit, zda je korelační koeficient délek dolní končetiny a délek trupu žen větší než 0. Toto tvrzení je opět zněním alternativní hypotézy, zatímco nulovou hypotézu dodefinujeme jako doplněk k tomuto tvrzení.

#### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Korelační koeficient proměnných délka dolní končetiny a délka trupu je menší nebo rovný 0.

$H_1$  : Korelační koeficient proměnných délka dolní končetiny a délka trupu je větší než 0.

- **matematická formulace** nulové a alternativní hypotézy

$H_0 : \rho \leq 0$

$H_1 : \rho > 0$

(pravostranná alternativa)

#### 2. Volba hladiny významnosti

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

#### 3. Testování kritickým oborem

- **Testovací statistika** K výpočtu testovací statistiky potřebujeme nejprve stanovit hodnotu výběrového korelačního koeficientu (viz funkce cor()). Hodnota výběrového korelačního koeficientu mezi šířkou nosu a interorbitální šířkou je  $r = 0.2853$ . Nyní již můžeme vypočítat hodnotu testovací statistiky

$$\begin{aligned} T_W &= \frac{\sqrt{n-2}R}{\sqrt{1-R^2}} \\ &= \frac{\sqrt{100-2} \times 0.285256}{\sqrt{1-0.285256^2}} \\ &= \frac{\sqrt{98} \times 0.285256}{\sqrt{0.918629}} \\ &= \frac{9.899495 \times 0.285256}{0.9584514} \\ &= 2.946305 \doteq 2.9463 \end{aligned}$$

```
494 data <- read.delim('00-Data\\13-two-samples-correlations-trunk.txt')
495 data.F <- data[data$sex == 'f', c('lowex.L', 'tru.L')]
496 data.F <- na.omit(data.F)
497 dim(data.F) # 100x2
498 lowex.LF <- data.F$lowex.L
499 tru.LF <- data.F$tru.L
500 alpha <- 0.01
501 r <- cor(lowex.LF, tru.LF, method = 'pearson') # 0.285256
502 n <- length(lowex.LF) # 100
503 tw <- sqrt(n - 2) * r / sqrt(1 - r ^ 2) # 2.946305
```

- Kritický obor

$$\begin{aligned}
 W &= \langle t_{n-2}(1 - \alpha); \infty \rangle \\
 &= \langle t_{100-2}(1 - 0.01); \infty \rangle \\
 &= \langle t_{98}(0.99); \infty \rangle \\
 &= \langle 2.365002; \infty \rangle
 \end{aligned}$$

```
504 qt(1 - alpha, n - 2) # 2.365002
```

- Závěr testování

Protože realizace testovací statistiky  $t_W = 2.9463$  náleží do kritického oboru, tj.  $t_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

#### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (-1, h) &= \left( -1; \frac{t_{n-2}(1 - \alpha)}{\sqrt{t_{n-2}^2(1 - \alpha) + n - 2}} \right) \\
 &= \left( -1; \frac{t_{100-2}(1 - 0.01)}{\sqrt{t_{100-2}^2(1 - 0.01) + 100 - 2}} \right) \\
 &= \left( -1; \frac{t_{98}(0.99)}{\sqrt{t_{98}^2(0.99) + 98}} \right) \\
 &= \left( -1; \frac{2.365002}{\sqrt{103.5932}} \right) \\
 &= \left( -1; \frac{2.365002}{10.17807} \right) \\
 &= (-1; 0.2323625)
 \end{aligned}$$

```
505 hh <- qt(1 - alpha, n - 2) / sqrt(qt(1 - alpha, n - 2) ^ 2 + n - 2) # 0.2323624
```

- Závěr testování

Protože  $r = 0.2853$  nenáleží do Waldova 99% empirického pravostranného intervalu spolehlivosti, tj.  $r = 0.2853 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

#### 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$p\text{-hodnota} = \Pr(T_W > t_w) = 1 - \Pr(T_W \leq t_w) = 1 - \Pr(T_W \leq 2.946305) = 0.002009162 \doteq 0.00200916$$

```
506 1 - pt(tw, n - 2) # 0.002009162
```

- **Závěr testování**

Protože  $p$ -hodnota = 0.00200916 je menší než  $\alpha = 0.01$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

## 6. Grafická vizualizace výsledků testování

Viz příklad 7.18.

## 7. Interpretace výsledků:

Na základě všech tří způsobů testování zamítáme hypotézu  $H_0$ . Korelační koeficient délky dolní končetiny a délky trupu žen je statisticky významně větší než 0. To znamená, že mezi délkou dolní končetiny a délkou trupu žen existuje statisticky významná přímá závislost. Mezi oběma znaky existuje nízký stupeň přímé lineární závislosti ( $\rho_1 = 0.2853$ ), který je statisticky významný. Ke stejnému závěru jsme dospěli též v příkladu 7.18.





### Příklad 7.23. Test o korelačním koeficientu $\rho$ (levostranná alternativa)

Mějme datový soubor 06-lin-uhl-fm.txt, proměnnou skull.H popisující výšku lebky v mm a proměnnou base.B popisující šířku lebeční báze na spojnici obou bodů porion v mm (viz sekce ??). Na hladině významnosti  $\alpha = 0.10$  zjistěte, zda mezi výškou lebky a šířkou lebeční báze žen existuje nepřímá závislost.

#### Řešení příkladu 7.23

Zadání příkladu je shodné se zadáním příkladu 7.23. Stejně tak datový soubor je totožný. Naším úkolem je porovnat korelační koeficient populace žen s konstantou  $\rho_0 = 0$ , a to prostřednictvím jednovýběrového testu o nezávislosti. Jeho jediným předpokladem je dvourozměrná normalita náhodného výběru výšek lebky a šířek lebeční báze žen, kterou jsme ověřili v rámci řešení příkladu 7.23. Podle zadání máme zjistit, zda mezi výškou lebky a šířkou lebeční báze žen existuje statisticky významná nepřímá lineární závislost. Ukazatelem nepřímé závislosti je záporná hodnota korelačního koeficientu. Naším úkolem je tedy zjistit, zda je korelační koeficient výšky lebky a šířky lebeční báze žen menší než 0, což je znění alternativní hypotézy.

```
507 data <- read.delim('00-Data\\06-lin-uhl-fm.txt')
508 data.F <- data[data$sex == 'f', c('skull.H', 'base.B')]
509 data.F <- na.omit(data.F)
510 dim(data.F) # 20x2
511 skull.HF <- data.F$skull.H
512 base.BF <- data.F$base.B
513 range(skull.HF) # 115.4675-142.0617
514 range(base.BF) # 108.5226-123.0860
```

#### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy  
 $H_0$  : Korelační koeficient výšky lebky a šířky lebeční báze žen je větší nebo rovný 0.  
 $H_1$  : Korelační koeficient výšky lebky a šířky lebeční báze žen je menší než 0.
- **matematická formulace** nulové a alternativní hypotézy  
 $H_0$  :  $\rho \geq \rho_0$ , kde  $\rho_0 = 0$   
 $H_1$  :  $\rho < \rho_0$ , kde  $\rho_0 = 0$   
(levostranná alternativa)

#### 2. Volba hladiny významnosti

- Hladinu významnosti volíme v souladu se zadáním  $\alpha = 0.10$ .

#### 3. Testování kritickým oborem

- **Testovací statistika** Nejprve vypočítáme hodnotu výběrového korelačního koeficientu  $R$  výšky lebky a šířky lebeční báze ( $r = -0.1713$ ). Nyní můžeme vypočítat testovací statistiku.

$$\begin{aligned} T_W &= \frac{\sqrt{n-2}R}{\sqrt{1-R^2}} \\ &= \frac{\sqrt{20-2} \times (-0.1712964)}{\sqrt{1-(-0.1712964)^2}} \\ &= \frac{\sqrt{18} \times (-0.1712964)}{\sqrt{0.9706575}} \\ &= \frac{4.242641 \times (-0.1712964)}{0.9852195} \\ &= -0.737652 \doteq -0.7377 \end{aligned}$$

```
515 alpha <- 0.1
516 r <- cor(skull.HF, base.BF, method = 'pearson') # -0.1712964
```

```
517 n <- length(skull.HF)
518 tw <- sqrt(n - 2) * r / sqrt(1 - r ^ 2) # -0.737652
```

- Kritický obor

$$\begin{aligned} W &= (-\infty; t_{n-2}(\alpha/2)) \\ &= (-\infty; t_{20-2}(0.10/2)) \\ &= (-\infty; t_{18}(0.05)) \\ &= (-\infty; -1.33039) \end{aligned}$$

```
519 qt(alpha, n - 2) # -1.330391
```

- Závěr testování

Protože realizace testovací statistiky  $t_W = -0.737652$  nenáleží do kritického oboru, tj.  $t_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.10$ .

#### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned} (d, 1) &= \left( \frac{t_{n-2}(\alpha)}{\sqrt{t_{n-2}^2(\alpha) + n - 2}}; 1 \right) \\ &= \left( \frac{t_{20-2}(0.10)}{\sqrt{t_{20-2}^2(0.10) + 20 - 2}}; 1 \right) \\ &= \left( \frac{t_{18}(0.10)}{\sqrt{t_{18}^2(0.10) + 18}}; 1 \right) \\ &= \left( \frac{-1.330391}{\sqrt{(-1.330391)^2 + 18}}; 1 \right) \\ &= \left( \frac{-1.330391}{\sqrt{19.76994}}; 1 \right) \\ &= \left( \frac{-1.330391}{4.44634}; 1 \right) \\ &= (-0.2992104; 1) \end{aligned}$$

```
520 dh <- qt(alpha, n - 2) / sqrt(qt(alpha, n - 2) ^ 2 + n - 2) # -0.2992103
```

- Závěr testování

Protože  $r = -0.1713$  náleží do Waldova 90% empirického jednostranného intervalu spolehlivosti, tj.  $r = 0.1713 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.10$ .

#### 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$p\text{-hodnota} = \Pr(T_w \leq t_w) = \Pr(T_w \leq -0.737652) = 0.230363 \doteq 0.23036$$

## 6. Grafická vizualizace výsledků testování

Viz příklad 7.19.

## 7. Interpretace výsledků:

Na základě všech tří způsobů testování nezamítáme hypotézu  $H_0$ . Korelační koeficient výšky lebky a šířky lebeční báze žen není statisticky významně menší než 0. Mezi výškou lebky a šířkou lebeční báze neexistuje statisticky významná nepřímá závislost. Interpretací výběrového korelačního koeficientu můžeme stanovit, že mezi oběma znaky existuje nízký stupeň nepřímé lineární závislosti ( $\rho_1 = -0.1713$ ), který není statisticky významný. Ke stejnému závěru jsme dospěli také v příkladu 29.



## 7.7 Test o parametru $p$ alternativního rozdělení

Nechť  $X_1, \dots, X_N$  je náhodný výběr sledující nastání úspěchu ( $X_i = 1$ ), nebo nastání neúspěchu ( $X_i = 0$ ) v  $i$ -tém náhodném pokusu,  $i = 1, \dots, N$ . Náhodný výběr  $X_1, \dots, X_N$  potom pochází z alternativního rozdělení, tj.  $X \sim \text{Alt}(p)$ , kde  $p$  je pravděpodobnost nastání úspěchu v jednom náhodném pokusu. Na hladině významnosti  $\alpha = 0.05$  testujeme jednu z následujících tří hypotéz oproti příslušné alternativní hypotéze.

$H_{01} : p = p_0$	oproti	$H_{11} : p \neq p_0$ (oboustranná alt.)
$H_{02} : p \leq p_0$	oproti	$H_{12} : p > p_0$ (pravostranná alt.)
$H_{03} : p \geq p_0$	oproti	$H_{13} : p < p_0$ (levostranná alt.)

Test nazýváme testem o pravděpodobnosti  $p$ . Testovací statistika má tvar

$$Z_W = \frac{M - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}}, \quad (7.6)$$

kde  $M = \frac{1}{N} \sum_{i=1}^N X_i$ , kde  $X_i$  nabývá hodnot 0 nebo 1, je výběrový průměr,  $N$  je rozsah náhodného výběru a  $p_0$  je konstanta z nulové hypotézy. Testovací statistika  $Z_W$  pochází asymptoticky ze standardizovaného normálního rozdělení, tj.

$$Z_W = \frac{M - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \stackrel{\mathcal{A}}{\approx} N(0, 1).$$

To znamená, že pro malý rozsah náhodného výběru  $N$  nemusí mít rozdělení statistiky  $Z_W$  charakter standardizovaného normálního rozdělení. To je ale problém, neboť testování kritickým oborem, intervalem spolehlivosti i  $p$ -hodnotou je založeno na předpokladu, že  $Z_W$  má charakter standardizovaného normálního rozdělení. Proto pro malý rozsah náhodného výběru nemůžeme statistiku  $Z_W$  k testování nulové hypotézy použít, neboť závěry takového testování by mohly být mylné. S rostoucím rozsahem náhodného výběru  $N$  se však rozdělení statistiky  $Z_W$  čím dál více blíží ke standardizovanému normálnímu rozdělení, získává tak všechny jeho vlastnosti a závěry testování se stávají spolehlivě správnými.

Zda máme dostatečný počet pozorování k provedení testu o pravděpodobnosti  $p$  prověříme podmínkou dobré aproximace. Ta má tvar

$$Np_0(1 - p_0) > 9 \quad (7.7)$$

a pokud je splněna, můžeme test o pravděpodobnosti  $p$  použít, aniž bychom se vystavili riziku zavádějících výsledků testování. Pokud by však podmínka dobré aproximace nebyla splněna, nemůžeme test použít, dokud nerozšíříme náš datový soubor o další pozorování a nezvýšíme tak rozsah náhodného výběru  $N$ .

*Poznámka:* Problému s nedostatkem pozorování při použití testu o pravděpodobnosti  $p$  se můžeme vyvarovat, pokud si ve fázi plánování experimentu spočítáme minimální potřebný rozsah náhodného výběru  $N$ . Ve fázi sběru dat si potom již snadno ohlídáme, aby náš datový soubor obsahoval potřebný počet pozorování, ideálně s nějakou rezervou. K výpočtu minimálního rozsahu náhodného výběru potřebujeme znát pouze předpokládanou hodnotu pravděpodobnosti  $p_0$  (viz příklady 7.28 a 7.29).

Kritický obor podle zvolené alternativní hypotézy má tvar

$$\begin{array}{ll} H_{11} : p \neq p_0 & W = \langle -\infty; u_{\alpha/2} \rangle \cup \langle u_{1-\alpha/2}; \infty \rangle \\ H_{12} : p > p_0 & W = \langle u_{1-\alpha}; \infty \rangle \\ H_{13} : p < p_0 & W = \langle -\infty; u_{\alpha} \rangle \end{array}$$

kde  $u_{\alpha/2}$ ,  $u_{1-\alpha/2}$ ,  $u_{\alpha}$ ,  $u_{1-\alpha}$  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

$$\begin{aligned}
H_{11} : p \neq p_0 & \quad (d, h) = \left( M - \sqrt{\frac{M(1-M)}{N}} u_{1-\alpha/2}; M - \sqrt{\frac{M(1-M)}{N}} u_{\alpha/2} \right) \\
H_{12} : p < p_0 & \quad (d, h) = \left( M - \sqrt{\frac{M(1-M)}{N}} u_{1-\alpha}; 1 \right) \\
H_{13} : p > p_0 & \quad (0, h) = \left( 0; M - \sqrt{\frac{M(1-M)}{N}} u_{\alpha} \right)
\end{aligned}$$

*Poznámka:* Protože parametr  $p$  značí pravděpodobnost úspěchu v jednom pokusu, platí, že  $p \in (0; 1)$ , a tedy horní hranice levostranného intervalu spolehlivosti je 1. Analogicky dolní hranice pravostranného Waldova empirického intervalu spolehlivosti je 0.

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

$$\begin{aligned}
H_{11} : p \neq p_0 & \quad p\text{-hodnota} = 2 \min\{\Pr(Z_W \leq z_W), \Pr(Z_W > z_W)\} \\
H_{12} : p > p_0 & \quad p\text{-hodnota} = \Pr(Z_W > z_W) = 1 - \Pr(Z_W \leq z_W) \\
H_{13} : p < p_0 & \quad p\text{-hodnota} = \Pr(Z_W \leq z_W)
\end{aligned}$$

kde  $Z_W$  je náhodná veličina,  $z_W$  je realizace testovací statistiky  $Z_W$  (viz vzorec 7.6), tedy konkrétní číslo, a  $\Pr(Z_W \leq z_W)$  je distribuční funkce standardizovaného normálního rozdělení, jejíž hodnotu získáme pomocí  a implementované funkce `pnorm()`.

### Příklad 7.24. Test o parametru $p$ alternativního rozdělení (oboustranná alternativa)

Mějme datový soubor `25-one-sample-probability-dermatoglyphs.txt` obsahující údaje o frekvenci výskytu dermatoglyfického vzoru *smyčka* na prstech 470 jedinců (235 mužů a 235 žen) populace Bagathů z Araku Valley (viz sekce ??). Současně máme k dispozici hodnotu pravděpodobnosti výskytu dermatoglyfického vzoru *smyčka* u jedinců z populace Lambadis ( $p_m = 0.5618$ ,  $p_f = 0.6233$ ). Na hladině významnosti  $\alpha = 0.05$  zjistíte, zda existuje rozdíl mezi frekvencemi výskytu dermatoglyfického vzoru *smyčka* u mužů bagathské populace z Araku Valley a mužů z populace Lambadis.

#### Řešení příkladu 7.24

Příkazem `read.delim()` načteme datový soubor, přičemž nastavením argumentu `row.names = 1` specifikujeme, že první sloupec tabulky má být zvolen jako záhlaví řádků. Pomocí funkce `sum()` zjistíme celkový počet všech pozorování. Dále, aplikováním funkce `FUN = sum` na sloupce (`MARGIN = 2`) tabulky `data` prostřednictvím příkazu `apply()` zjistíme počet údajů pro muže a pro ženy.

```
522 (data <- read.delim('00-Data\\25-one-sample-probability-dermatoglyphs.txt',  
523 row.names = 1))
```

	m	f
whorl	1053	880
loop	1246	1349
arch	51	121

524  
525  
526  
527

```
528 sum(data) # 4700  
529 apply(data, MARGIN = 2, FUN = sum) # m 2350 # f 2350
```

Z tabulky vidíme, že z celkového počtu 4700 otisků patří 2350 otisků mužům, čemuž odpovídá 10 otisků prstů na jednoho jedince mužského pohlaví. Z celkového počtu 2350 otisků byl na 1246 otiscích rozpoznán vzor *smyčka* a na zbylých 1104 otiscích byl rozeznán jiný vzor než *smyčka* (tj. *oblouček* nebo *vír*). Rozsah náhodného výběru  $N = 2350$ .

Naším úkolem je porovnat frekvenci výskytu dermatoglyfického vzoru *smyčka* dvou indických populací. U bagathské populace máme k dispozici počet úspěchů, tj. počet výskytů vzoru *smyčka* na 2350 prstech. Náhodná veličina  $X$  popisující frekvenci výskytu vzorů *smyčka* u mužů bagathské poulace tedy pochází z alternativního rozdělení, tj.  $X \sim \text{Alt}(p)$ , kde  $p$  je pravděpodobnost nastání úspěchu, tedy pravděpodobnost výskytu vzoru *smyčka*. Druhá, populace z Lambadis je reprezentována pouze pravděpodobností ( $p_m = 0.5618$ ). Řešení příkladu vede na situaci, kdy pravděpodobnost  $p$  porovnáváme s konstantou 0.5618, tedy na jednovýběrový test o parametru  $p$  alternativního rozdělení. K použití tohoto testu je nejprve potřeba ověřit podmínku dobré aproximace pro náhodný výběr mužů bagathské populace, což znamená ověřit, zda platí  $Np_0(1 - p_0) > 5$ , kde  $p_0 = 0.5618$ .

$$Np_0(1 - p_0) = 2350 \times 0.5618 \times (1 - 0.5618) = 2350 \times 0.5618 \times 0.4382 = 578.5248.$$

Jelikož  $578.5248 > 5$ , je podmínka dobré aproximace pro muže bagathské populace splněna.

```
530 N <- 2350  
531 x <- 1246  
532 p0 <- 0.5618  
533 N*p0*(1-p0) # 578.5248
```

Protože je podmínka dobré aproximace pro muže bagathské populace splněna, můžeme se zaměřit na otázku ze zadání. Naším úkolem je zjistit, zda existuje rozdíl mezi frekvencemi výskytu dermatoglyfického vzoru *smyčka* u mužů bagathské populace z Araku Valley a mužů z populace Lambadis. Toto tvrzení je zněním alternativní hypotézy, neboť rozdíl implikuje nerovnost a nerovnost je vždy součástí alternativní hypotézy. Zbývá dodefinovat nulovou hypotézu. Proces testování provedeme v posloupnosti šesti kroků.

#### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Frekvence výskytu dermatoglyfického vzoru smyčka u mužů багаžtské populace a populace Lambadis jsou shodné.

$H_1$  : Frekvence výskytu dermatoglyfického vzoru smyčka u mužů багаžtské populace a populace Lambadis nejsou shodné.

- **matematická formulace** nulové a alternativní hypotézy

$H_0$  :  $p = p_0$ , kde  $p_0 = 0.5618$

$H_1$  :  $p \neq p_0$ , kde  $p_0 = 0.5618$   
(oboustranná alternativa)

## 2. Volba hladiny významnosti

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

## 3. Testování kritickým oborem

- Testovací statistika

$$\begin{aligned} Z_W &= \frac{M - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \\ &= \frac{0.5302128 - 0.5618}{\sqrt{\frac{0.5618(1-0.5618)}{2350}}} \\ &= \frac{-0.0315872}{\sqrt{\frac{0.5618 \times 0.4382}{2350}}} \\ &= \frac{-0.0315872}{\sqrt{0.0001047578}} \\ &= \frac{-0.0315872}{0.01023512} \\ &= -3.086158 \doteq -3.08616 \end{aligned}$$

```
534 alpha <- 0.05
535 m <- x / N #0.5302128
536 zw <- (m - p0) / (sqrt(p0 * (1 - p0) / N)) # -3.08616
```

- Kritický obor

$$\begin{aligned} W &= (-\infty; u_{\alpha/2}) \cup (u_{1-\alpha/2}; \infty) \\ &= (-\infty; u_{0.05/2}) \cup (u_{1-0.05/2}; \infty) \\ &= (-\infty; u_{0.025}) \cup (u_{0.975}; \infty) \\ &= (-\infty; -1.9560) \cup (1.9560; \infty) \end{aligned}$$

```
537 qnorm(alpha / 2) # -1.959964
538 qnorm(1 - alpha / 2) # 1.959964
```

- **Závěr testování**

Protože realizace testovací statistiky  $z_W = -3.08616$  náleží do kritického oboru, tj.  $z_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

## 4. Testování intervalem spolehlivosti

- **Interval spolehlivosti**

$$\begin{aligned}
 (d, h) &= \left( M - \sqrt{\frac{M(1-M)}{N}} u_{1-\alpha/2}; M - \sqrt{\frac{M(1-M)}{N}} u_{\alpha/2} \right) \\
 &= \left( 0.5302128 - \sqrt{\frac{0.5302128(1-0.5302128)}{2350}} u_{1-0.05/2}; 0.5302128 - \sqrt{\frac{0.5302128(1-0.5302128)}{2350}} u_{0.05/2} \right) \\
 &= \left( 0.5302128 - \sqrt{\frac{0.5302128 \times 0.4697872}{2350}} u_{0.975}; 0.5302128 - \sqrt{\frac{0.5302128 \times 0.4697872}{2350}} u_{0.025} \right) \\
 &= \left( 0.5302128 - \sqrt{0.0001059945} \times 1.959964; 0.5302128 - \sqrt{0.0001059945} \times (-1.959964) \right) \\
 &= (0.5302128 - 0.01029536 \times 1.959964; 0.5302128 - 0.01029536 \times (-1.959964)) \\
 &= (0.5100343; 0.5503913)
 \end{aligned}$$

```

539 dh <- m - sqrt(m * (1 - m) / N) * qnorm(1 - alpha / 2) # 0.5100342
540 hh <- m - sqrt(m * (1 - m) / N) * qnorm(alpha / 2) # 0.5503913

```

- **Závěr testování**

Protože  $p_0 = 0.5618$  nenáleží do Waldova 95% empirického oboustranného intervalu spolehlivosti, tj.  $p_0 = 0.5618 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- **$p$ -hodnota**

$$\begin{aligned}
 p\text{-hodnota} &= 2 \min\{\Pr(Z_W \leq z_W), \Pr(Z_W > z_W)\} \\
 &= 2 \min\{\Pr(Z_W \leq z_W), 1 - \Pr(Z_W \leq z_W)\} \\
 &= 2 \min\{\Pr(Z_W \leq -3.08616), 1 - \Pr(Z_W \leq -3.08616)\} \\
 &= 2 \min\{0.001013798, 0.9989862\} \\
 &= 2 \times 0.001013798 = 0.002027595 \doteq 0.0020276
 \end{aligned}$$

```

541 p.hodnota <- 2 * min (pnorm(zw), 1 - pnorm(zw)) # 0.002027595

```

- **Závěr testování**

Protože  $p$ -hodnota = 0.0020276 je menší než  $\alpha = 0.05$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.05$ .

## 6. Grafická vizualizace výsledků testování

Výsledek testování vizualizujeme pomocí sloupcového diagramu relativních četností, který vygenerujeme příkazem `rel.barplot()`. Funkce `rel.barplot()` je implementována v RSkriptu `Sbirka-AS-I-2018-funkce.R`. Před použitím samotné funkce je tedy potřeba načíst RSkript příkazem `source()`.

```

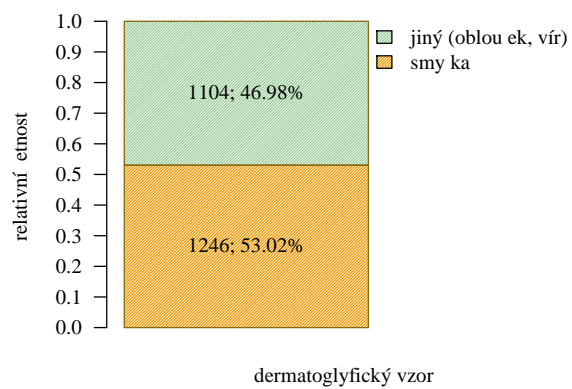
542 source('Sbirka-AS-I-2018-funkce.R')
543 par(mar = c(5, 4, 1, 1), family = 'Times')
544 rel.barplot(c(x, N - x), col = c('orange2', 'darkseagreen3'),
545           border = 'goldenrod4',
546           names = c('smyčka', 'jiný (oblouček, vír)'),
547           density = 60, main = '',
548           xlab = 'dermatoglyfický vzor', ylab = 'relativní četnost')

```

## 7. Interpretace výsledků:

Na základě všech tří způsobů testování zamítáme hypotézu  $H_0$ . Mezi frekvencí výskytu dermatoglyfického znaku *smyčka* u mužů baghatské populace z Araku Valley a frekvencí výskytu vzoru *smyčka* u mužů z populace Lambadis existuje statisticky významný rozdíl.





Obrázek 34: Sloupcový diagram relativních četností zastoupení dermatoglyfického vzoru *smyčka* v populaci bagathských mužů z Araku Valley



### Příklad 7.25. Test o parametru $p$ alternativního rozdělení (pravostranná alternativa)

Pravděpodobnost narození chlapce je o něco málo vyšší, než pravděpodobnost narození děvčete (obecný poměr je 51:49). V datovém souboru 08-one-sample-probability-sexratio.txt jsou uvedeny údaje o pohlaví novorozenců narozených během jednoho roku v jedné krajské nemocnici (viz sekce ??). Potvrzují nasbírané údaje tvrzení, že se v tomto okrese rodí statisticky významně více chlapců než děvčat? Hladinu významnosti zvolte  $\alpha = 0.05$ .

#### Řešení příkladu 7.25

Příkazem `read.delim()` načteme datový soubor. Pomocí funkce `head()` vypíšeme první čtyři řádky z načtené tabulky. Tabulka obsahuje pouze jeden sloupec `sex` obsahující údaje o binární proměnné *pohlaví* s kódováním `m` = muž a `f` = žena. Pomocí příkazu `table()` zjistíme, kolik pozorování přísluší každé variantě znaku `sex`.

```
549 data <- read.delim('00-Data\\08-one-sample-probability-sexratio.txt')
550 head(data, n = 4)
```

```
  sex
1   m
2   m
3   f
4   m
```

551  
552  
553  
554  
555

```
556 table(data)
```

```
data
  f  m
674 729
```

557  
558  
559

V krajské nemocnici se v průběhu jednoho roku narodilo celkem 729 chlapců a 674 děvčat. Rozsah náhodného výběru  $N = 1403$ .

Naším úkolem je porovnat frekvenci výskytu narození chlapců s konstantou 0.5. Ta reprezentuje 50% šanci na narození chlapce a 50% šanci na narození děvčete. K dispozici máme počet úspěchů, tj. počet narozených chlapců z 1403 narozených jedinců. Náhodná veličina  $X$  popisující narození chlapce tedy pochází z alternativního rozdělení, tj.  $X \sim \text{Alt}(p)$ , kde  $p$  je pravděpodobnost nastání úspěchu, tedy pravděpodobnost narození chlapce. Řešení příkladu vede na situaci, kdy pravděpodobnost  $p$  porovnáváme s konstantou 0.5, tedy na jednovýběrový test o parametru  $p$  alternativního rozdělení. K použití tohoto testu je nejprve potřeba ověřit podmínku dobré aproximace, tj. ověřit, zda platí  $Np_0(1 - p_0) > 5$ , kde  $p_0 = 0.5$ .

$$Np_0(1 - p_0) = 1403 \times 0.5 \times (1 - 0.5) = 1403 \times 0.5 \times 0.5 = 350.75.$$

Jelikož  $350.75 > 5$ , je podmínka dobré aproximace splněna.

```
560 N <- 1403
561 x <- 729
562 p0 <- 0.5
563 N * p0 * (1 - p0) # 350.75
```

Protože podmínka dobré aproximace je splněna, můžeme se zaměřit na otázku ze zadání. Naším úkolem je zjistit, zda se ve sledovaném okrese rodí statisticky významně více chlapců než děvčat. Toto tvrzení odpovídá hypotéze, že frekvence narození chlapců je větší 0.5 a je zněním alternativní hypotézy, neboť v zadání není zmínka, že máme testovat hypotézu nebo nulovou hypotézu. Zbývá tedy dodefinovat znění  $H_0$ .

#### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy

$H_0$  : Frekvence výskytu narození chlapce je menší nebo rovna 0.5.

$H_1$  : Frekvence výskytu narození chlapce je větší než 0.5.

- **matematická formulace** nulové a alternativní hypotézy  
 $H_0 : p \leq p_0$ , kde  $p_0 = 0.5$   
 $H_1 : p > p_0$ , kde  $p_0 = 0.5$   
 (pravostranná alternativa)

## 2. Volba hladiny významnosti

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

## 3. Testování kritickým oborem

- **Testovací statistika**

$$\begin{aligned}
 Z_W &= \frac{M - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \\
 &= \frac{0.5196009 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{1403}}} \\
 &= \frac{0.0196009}{\sqrt{\frac{0.25}{1403}}} \\
 &= \frac{0.0196009}{\sqrt{0.0001047578}} \\
 &= \frac{0.0196009}{0.01334877} \\
 &= 1.468367 \doteq 1.4684
 \end{aligned}$$

```

564 alpha <- 0.05
565 m <- x / N # 0.5196009
566 zw <- (m - p0) / (sqrt(p0 * (1 - p0) / N)) # 1.468364

```

- **Kritický obor**

$$\begin{aligned}
 W &= \langle u_{1-\alpha}; \infty \rangle \\
 &= \langle u_{1-0.05}; \infty \rangle \\
 &= \langle u_{0.95}; \infty \rangle \\
 &= \langle 1.6449; \infty \rangle
 \end{aligned}$$

```

567 qnorm(1 - alpha) # 1.644854

```

- **Závěr testování**

Protože realizace testovací statistiky  $z_W = 1.4684$  nenáleží do kritického oboru, tj.  $z_W \notin W$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}
 (d, h) &= \left( M - \sqrt{\frac{M(1-M)}{N}} u_{1-\alpha/2}; 1 \right) \\
 &= \left( 0.5196009 - \sqrt{\frac{0.5196009(1-0.5196009)}{1403}} u_{1-0.05}; 1 \right) \\
 &= \left( 0.5196009 - \sqrt{\frac{0.5196009 \times 0.4803991}{1403}} u_{0.95}; 1 \right) \\
 &= \left( 0.5196009 - \sqrt{0.0001779158} \times 1.644854; 1 \right) \\
 &= (0.5196009 - 0.01333851 \times 1.644854; 1) \\
 &= (0.497661; 1)
 \end{aligned}$$

```
568 dh <- m - sqrt(m * (1 - m) / N) * qnorm(1 - alpha) # 0.497661
```

- Závěr testování

Protože  $p_0 = 0.5$  náleží do Waldova 95% empirického levostranného intervalu spolehlivosti, tj.  $p_0 = 0.5 \in IS$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$p\text{-hodnota} = 1 - \Pr(Z_W \leq z_W) = 1 - \Pr(Z_W \leq 1.468364) = 0.07100263 \doteq 0.0710026$$

```
569 p.hodnota <- 1 - pnorm(zw) # 0.07100263
```

- Závěr testování

Protože  $p$ -hodnota = 0.07100263 je větší než  $\alpha = 0.05$ ,  $H_0$  nezamítáme na hladině významnosti  $\alpha = 0.05$ .

## 6. Grafická vizualizace výsledků testování

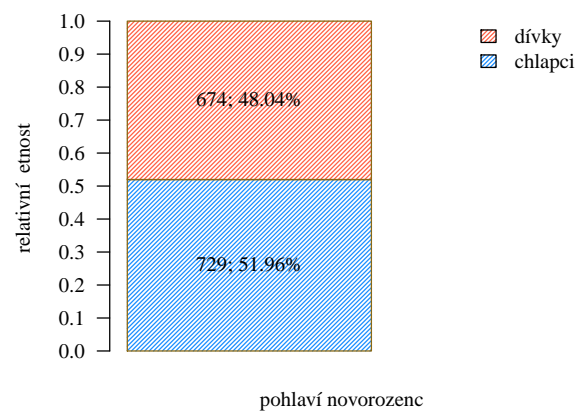
Výsledek testování vizualizujeme pomocí sloupcového diagramu relativních četností.

```
570 source('Sbirka-AS-I-2018-funkce.R')
571 par(mar = c(4, 4, 1, 1), family = 'Times')
572 rel.barplot(c(x, N - x), col = c('dodgerblue', 'tomato'),
573           border = 'goldenrod4',
574           names = c('chlapci', 'dívky'),
575           density = 40, main = '',
576           xlab = 'pohlaví novorozenců', ylab = 'relativní četnost')
```

## 7. Interpretace výsledků:

Na základě všech tří způsobů testování nezamítáme hypotézu  $H_0$ . Nasbírané údaje nepotvrzují, že by se ve sledovaném kraji rodilo statisticky významně více chlapců než děvčat.





Obrázek 35: Sloupcový diagram relativních četností zastoupení narození chlapce a děvčete v krajské nemocnici během jednoho roku

### Příklad 7.26. Test o parametru $p$ alternativního rozdělení (levostranná alternativa)

Mějme k dispozici údaje o frekvenci výskytu epigenetického znaku *sutura metopica* (binomické proměnná) na lebkách Ainů z ostrova Hokkaido (viz datový soubor 09-one-sample-probability-sutmet.txt a jeho popis v sekci ??). Dále mějme k dispozici údaje o výskytu epigenetického znaku *sutura metopica* na lebkách jedinců ze současné japonské populace ( $p_{jap} = 0.091$ ). Na hladině významnosti  $\alpha = 0.01$  zjistíte, zda je frekvence výskytu epigenetického znaku *sutura metopica* u populace Ainů statisticky významně nižší než u současné japonské populace.

#### Řešení příkladu 7.26

Příkazem `read.delim()` načteme datový soubor, přičemž nastavením argumentu `row.names = 1` specifikujeme, že první sloupec tabulky má být zvolen jako záhlaví řádků.

```
577 (data <- read.delim('00-Data/09-one-sample-probability-sutmet.txt',  
578 row.names = 1))
```

```
      n met  
Ain 184  6
```

579  
580

Z tabulky vidíme, že z celkového počtu 184 lebek populace Ainů byl zaznamenán výskyt epigenetického znaku *sutura metopica* na 6 lebkách. Rozsah náhodného výběru lebek Ainské populace  $N = 184$ .

Naším úkolem je porovnat frekvenci výskytu epigenetického znaku *sutura metopica* dvou japonských populací. U populace Ainů máme k dispozici počet úspěchů, tj. počet výskytů epigenetického znaku *sutura metopica* u 184 lebek. Náhodná veličina  $X$  popisující frekvenci výskytu znaku *sutura metopica* u mužů populace Ainů tedy pochází z alternativního rozdělení, tj.  $X \sim \text{Alt}(p)$ , kde  $p$  je pravděpodobnost nastání úspěchu, tedy pravděpodobnost výskytu znaku *sutura metopica*. Současná japonské populace je oproti tomu reprezentována pouze pravděpodobností ( $p_{jap} = 0.091$ ). Řešení příkladu vede na situaci, kdy pravděpodobnost  $p$  porovnáváme s konstantou 0.091, tedy na jednovýběrový test o parametru  $p$  alternativního rozdělení. Před použitím tohoto testu nejprve ověříme podmínku dobré aproximace pro náhodný výběr mužů populace Ainů, tj.  $Np_0(1 - p_0) > 5$ , kde  $p_0 = 0.091$ .

$$Np_0(1 - p_0) = 184 \times 0.091 \times (1 - 0.091) = 184 \times 0.091 \times 0.909 = 15.2203.$$

```
581 N <- 184  
582 x <- 6  
583 p0 <- 0.091  
584 N * p0 * (1 - p0) # 15.2203
```

Jelikož číslo  $15.2203 > 5$ , je podmínka dobré aproximace pro populaci Ainů splněna a my se můžeme zaměřit na řešení otázky ze zadání. Zde poznamenejme, že v zadání příkladu není zmínka o znění nulové hypotézy, pouze o záměru zjistit, zda je frekvence výskytu znaku *sutura metopica* u populace Ainů nižší než u současné japonské populace. Toto tedy bude znění alternativní hypotézy, zatímco nulovou hypotézu musíme vhodně dodefinovat.

#### 1. Stanovení hypotéz

- **slovní formulace** nulové a alternativní hypotézy  
 $H_0$  : Frekvence výskytu epigenetického znaku *sutura metopica* u populace Ainů je vyšší nebo rovna frekvenci znaku *sutura metopica* u současné japonské populace.  
 $H_1$  : Frekvence výskytu epigenetického znaku *sutura metopica* u populace Ainů je nižší než u současné japonské populace.
- **matematická formulace** nulové a alternativní hypotézy  
 $H_0$  :  $p \geq p_0$ , kde  $p_0 = 0.091$   
 $H_1$  :  $p < p_0$ , kde  $p_0 = 0.091$   
(levostranná alternativa)

#### 2. Volba hladiny významnosti

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

### 3. Testování kritickým oborem

- Testovací statistika

$$\begin{aligned}Z_W &= \frac{M - p_0}{\sqrt{\frac{M(1-M)}{N}}} \\&= \frac{0.0326087 - 0.091}{\sqrt{\frac{0.091(1-0.091)}{184}}} \\&= \frac{-0.0583913}{\sqrt{\frac{0.082719}{184}}} \\&= \frac{-0.0583913}{\sqrt{0.0004495598}} \\&= \frac{-0.0583913}{0.02120283} \\&= -2.753939 \doteq -2.75394\end{aligned}$$

```
585 alpha <- 0.01
586 m <- x / N # 0.0326087
587 zw <- (m - p0) / (sqrt(p0 * (1 - p0) / N)) # -2.75394
```

- Kritický obor

$$\begin{aligned}W &= (-\infty; u_\alpha) \\&= (-\infty; u_{0.01}) \\&= (-\infty; -2.3263)\end{aligned}$$

```
588 qnorm(alpha) # -2.326348
```

- Závěr testování

Protože realizace testovací statistiky  $z_W = -2.75394$  náleží do kritického oboru, tj.  $z_W \in W$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

### 4. Testování intervalem spolehlivosti

- Interval spolehlivosti

$$\begin{aligned}(-1, h) &= \left( 0; m - \sqrt{\frac{m(1-m)}{N}} u_\alpha \right) \\&= \left( 0; 0.0326087 - \sqrt{\frac{0.0326087(1-0.0326087)}{184}} u_{0.01} \right) \\&= \left( 0; 0.0326087 - \sqrt{0.0001714422} \times (-2.326348) \right) \\&= \left( 0; 0.0326087 - 0.01309359 \times (-2.326348) \right) \\&= (0; 0.06306895) \doteq (0; 0.063069)\end{aligned}$$

```
589 HH <- m - sqrt(m * (1 - m) / N) * qnorm(alpha) # 0.06306895
```

- **Závěr testování**

Protože  $p_0 = 0.091$  nenáleží do Waldova 99% empirického pravostranného intervalu spolehlivosti, tj.  $p_0 = 0.091 \notin IS$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

## 5. Testování $p$ -hodnotou

- $p$ -hodnota

$$\begin{aligned} p\text{-hodnota} &= \Pr(Z_W \leq z_W) \\ &= \Pr(Z_W \leq -2.326348) \\ &= 0.002944129 \doteq 0.002944 \end{aligned}$$

```
590 p.hodnota <- pnorm(zw) # 0.002944129
```

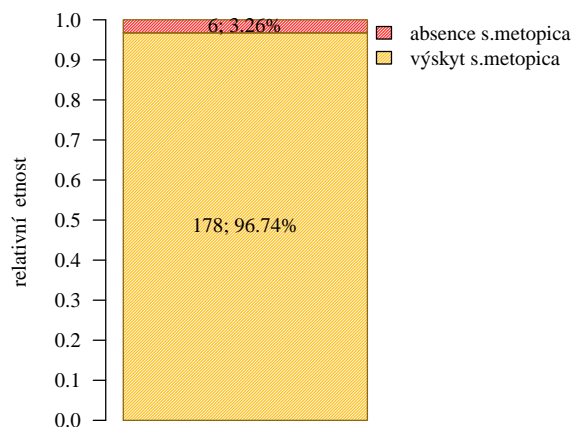
- **Závěr testování**

Protože  $p$ -hodnota = 0.002944 je menší než  $\alpha = 0.01$ ,  $H_0$  zamítáme na hladině významnosti  $\alpha = 0.01$ .

## 6. Grafická vizualizace výsledků testování

Výsledek testování vizualizujeme pomocí sloupcového diagramu relativních četností, který vygenerujeme příkazem `rel.barplot()`.

```
591 source('Sbirka-AS-I-2018-funkce.R')
592 par(mar = c(1, 4, 1, 1), family = 'Times')
593 rel.barplot(c(N - 6, 6), col = c('goldenrod1', 'firebrick1'),
594           border = 'goldenrod4',
595           names = c('výskyt s.metopica', 'absence s.metopica'),
596           density = 60, main = '',
597           xlab = '', ylab = 'relativní četnost')
```



Obrázek 36: Sloupcový diagram relativních četností zastoupení epigenetického znaku *sutura metopica* v populaci Ainů

7. **Interpretace výsledků:** Na základě všech tří způsobů testování zamítáme hypotézu  $H_0$ . Frekvence výskytu epigenetického znaku *sutura metopica* u populace Ainů je statisticky významně menší než u současné japonské populace.





### Příklad 7.27. Test o parametru $p$ alternativního rozdělení (nesplnění předpokladů)

Mějme k dispozici údaje o frekvenci výskytu epigenetického znaku *sutura metopica* (binomické proměnná) na lebkách jedinců z Anatolské populace z oblasti Erzurum. Z celkového počtu 47 jedinců byl zaznamenán výskyt tohoto znaku u 7 jedinců. Dále mějme k dispozici údaje o výskytu epigenetického znaku *sutura metopica* na lebkách jedinců z moderní Anatolské populace ( $p_a = 0.093$ ). Na hladině významnosti  $\alpha = 0.10$  testujte hypotézu, že frekvence výskytu epigenetického znaku *sutura metopica* u Anatolské populace z oblasti Erzurum je menší nebo rovna frekvenci výskytu epigenetického znaku *sutura metopica* u moderní Anatolské populace.

### Řešení příkladu 7.27

Nejprve si do proměnné  $N$  vložíme celkový počet jedinců (rozsah náhodného výběru  $N = 47$ ) a do proměnné  $x$  vložíme počet úspěchů, tj. počet výskytů epigenetického znaku *sutura metopica* na 47 zkoumaných lebkách ( $x = 7$ ). Náhodná veličina  $X$  popisující počet výskytů epigenetického znaku *sutura metopica* pochází z Alternativního rozdělení, tj.  $X \sim \text{Alt}(p)$ , kde  $p$  je pravděpodobnost úspěchu, tj. pravděpodobnost výskytu epigenetického znaku *sutura metopica*.

```
598 N <- 47
599 x <- 7
```

Naším úkolem je porovnat frekvenci výskytu epigenetického znaku *sutura metopica* u dvou Anatolských populací. U populace z oblasti Erzurum máme k dispozici počet úspěchů. Naopak u moderní Anatolské populace máme k dispozici pouze pravděpodobnost výskytu ( $p_a = 0.093$ ). Řešení příkladu vede na situaci, kdy pravděpodobnost  $p$  porovnáváme s konstantou 0.093, tedy na jednovýběrový test o parametru  $p$  alternativního rozdělení. Před použitím testu nejprve ověříme podmínku dobré aproximace, tj. zda  $Np_0(1 - p_0) > 5$ , kde  $p_0 = 0.093$ .

$$Np_0(1 - p_0) = 47 \times 0.093 \times (1 - 0.093) = 47 \times 0.093 \times 0.907 = 3.9645.$$

Jelikož  $3.9645 < 5$ , podmínka dobré aproximace není splněna a k otestování hypotézy ze zadání není tedy možné použít test o parametru  $p$  alternativního rozdělení.

*Poznámka:* Protože rozsah náhodného výběru není dostatečně velký, nemůžeme k testování hypotézy ze zadání použít parametrický test. V opačném případě by získané výsledky nebyly spolehlivé. Jedinou možností, jak získat spolehlivé výsledky je rozšířit datový soubor o další pozorování. Tomuto procesu bychom se vyhnuli, kdybychom v počáteční fázi plánování experimentu spočítali, jak minimálně velký rozsah náhodného výběru je potřeba získat, aby byla podmínka dobré aproximace splněna a výsledky stanovené za základě testu o parametru  $p$  byly spolehlivé (viz příklady 7.28 a 7.29). ★

**Příklad 7.28. Minimální rozsah náhodného výběru pro test o parametru  $p$  alternativního rozdělení (oboustranná alternativa)**

Mějme k dispozici údaje o frekvenci výskytu polydaktylie u jedinců Československé populace z roku 1966 ( $p_{cs} = 0.0005$ ,  $N_{cs} = 20\,074$ ). Dále předpokládejme, že náhodný výběr  $X$  popisující výskyt polydaktylie u jedinců Polské populace pochází z Alternativního rozdělení, tj.  $X \sim \text{Alt}(p)$ , kde  $p$  je pravděpodobnost výskytu polydaktylie u jedinců Polské populace. Vypočítejte, jak velký minimální rozsah náhodného výběru je potřeba k otestování hypotézy o shodě frekvence výskytu polydaktilie u jedinců Československé a Polské populace.

**Řešení příkladu 7.28**

Naším úkolem je porovnat frekvenci výskytu polydaktylie u Československé a Polské populace, přičemž momentálně jsme ve fázi plánování experimentu. Zatímco u Polské populace budeme mít po nasbírání dat k dispozici údaje o počtu úspěchů, tj. o počtu výskytů polydaktilie u Československé populace máme k dispozici údaj o výskytu polydaktylie z roku 1966 ( $p = 0.0005$ ). Celá studie spěje k situaci, kdy budeme porovnávat pravděpodobnost  $p$  s konstantou 0.0005, tedy na jednovýběrový test o parametru  $p$  alternativního rozdělení. K tomu, abychom tento test mohli použít, potřebujeme, aby byla splněna podmínka dobré aproximace, tj. aby  $Np_0(1-p_0) > 5$ , kde  $p_0 = 0.0005$ . Z nerovnice si tedy vyjádříme a následně dopočítáme hodnotu  $N$ .

$$\begin{aligned} Np_0(1-p_0) &> 5 \\ N &> \frac{5}{p_0(1-p_0)} \\ N &> \frac{5}{0.0005(1-0.0005)} \\ N &> \frac{5}{0.0005 \times 0.9995} \\ N &> 10\,005 \end{aligned}$$

```
600 p0 <- 0.0005
601 5 / (p0 * (1 - p0)) # 10005
```

Aby byla splněna podmínka dobré aproximace, je potřeba získat jako základ k otestování nulové hypotézy ze zadání náhodný výběr o rozsahu alespoň 10 006 jedinců. ★

**Příklad 7.29. Minimální rozsah náhodného výběru pro test o parametru  $p$  alternativního rozdělení (jednostranná alternativa)**

Mějme k dispozici údaje o frekvenci výskytu dermatoglyfického vzoru *víř* na palci pravé ruky u mužů České populace ( $p_{cs} = 0.533$ ). Dále předpokládejme, že náhodný výběr  $X$  popisující výskyt dermatoglyfického vzoru *víř* na palci pravé ruky u mužů Slovenské populace pochází z Alternativního rozdělení, tj.  $X \sim \text{Alt}(p)$ , kde  $p$  je pravděpodobnost výskytu dermatoglyfického vzoru *víř* na palci pravé ruky u mužů Slovenské populace. Vypočítejte, jak velký minimální rozsah náhodného výběru je potřeba k otestování hypotézy, že frekvence výskytu vzoru *víř* na palci pravé ruky mužů Slovenské populace je větší nebo rovna frekvenci výskytu vzoru *víř* na palci pravé ruky mužů České populace.

**Řešení příkladu 7.29**

Naším úkolem je porovnat frekvenci výskytu vzoru *víř* na palci pravé ruky mužů České a Slovenské populace, přičemž jsme teprve ve fázi plánování experimentu. Zatímco u Slovenské populace budeme mít po nasbírání dat disponovat údaji o počtu úspěchů, tj. o počtu výskytů dermatoglyfického vzoru *víř* na palci pravé ruky u mužů Slovenské populace, u mužů České populace máme k dispozici pouze pravděpodobnost výskytu vzoru *víř* na palci pravé ruky ( $p = 0.533$ ). Celá studie spěje k případu porovnávání pravděpodobnosti  $p$  s konstantou 0.533, tedy na jednovýběrový test o parametru  $p$  alternativního rozdělení. K použití tohoto testu, potřebujeme splnit podmínku dobré aproximace, tj. zajistit, aby  $Np_0(1 - p_0) > 5$ , kde  $p_0 = 0.533$ .

$$\begin{aligned} Np_0(1 - p_0) &> 5 \\ N &> \frac{5}{p_0(1 - p_0)} \\ N &> \frac{5}{0.533(1 - 0.533)} \\ N &> \frac{5}{0.533 \times 0.467} \\ N &> 20.0875 \end{aligned}$$

```
602 p0 <- 0.533
603 5 / (p0 * (1 - p0)) # 20.0875
```

Aby byla splněna podmínka dobré aproximace, je potřeba získat jako základ k otestování nulové hypotézy ze zadání náhodný výběr o rozsahu alespoň 21 jedinců. ★

*Poznámka:* Z příkladů 7.28 a 7.29 vidíme, že rozsah náhodného výběru závisí pouze na očekávané pravděpodobnosti  $p_0$  z nulové hypotézy. Čím je  $p_0$  blíže k 0 nebo 1, tím větší počet pozorování bude potřeba k zajištění podmínky dobré aproximace. Naopak, čím bude hodnota  $p_0$  blíže k číslu 0.5, tím menší rozsah náhodného výběru je potřebný ke splnění podmínky dobré aproximace.