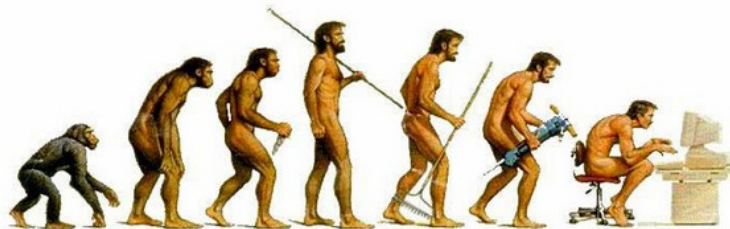


Masarykova univerzita v Brně  
Přírodovědecká fakulta

# APLIKOVANÁ STATISTIKA PRO ANTROPOLOGY

Zadání domácího úkolu – Skupina B



Brno, 2017

## Pokyny k řešení domácího úkolu

Domácí úkol sestává z šesti příkladů. Za vyřešení příkladů lze získat  $10+20+25+75+40+10=180$  bodů + 20 bodů za celkovou úpravu úkolu, kódu, komentáře k postupům, apod. Celkem lze tedy získat 200 bodů.

Aby byl úkol uznán za splněný, je potřeba získat alespoň **150 bodů (75 %)**. Pokud nebude dodatečně stanoveno jinak, má student na vyřešení domácího úkolu 14 dní počínaje dnem zadání domácího úkolu. Pokud student potřebných 120 bodů nezíská, bude mu úkol navrácen a student dostane jeden týden na dořešení příkladů. Pokud ani potom student kýženého počtu nedosáhne, bude psát na konci semestru prověřovací písemku na látku obsaženou v celém semestru.

Řešení domácího úkolu vložte, prosím, do odevzdávací skřínky k předmětu MAS10c (cvičení z AS). (Pozor! nesplést s odevzdávací skřínkou předmětu MAS01 - přednášky!).

Kompletním řešením je míněno:

- dodání **zcela funkčního** R-Skriptu s názvem AS-2017-skupina-X-prijmeni-jmeno.R. Namísto X vložte verzi vypracovávaného domácího úkolu (A nebo B). Před odesláním skriptu vyučující si vyčistíte workspace (V RStudio: Session → Clear Workspace) a všechny příkazy finálně projděte ještě jednou, abyste měli jistotu, že vše funguje, jak má. **Příklady, jejichž RSkript bude vyhazovat chybové hlášky, nebudou kontrolovány a automaticky budou vráceny k přepracování.**
- dodání textového souboru (Word) s názvem AS-2017-skupina-X-prijmeni-jmeno.docx, obsahujícího všechny potřebné komentáře, **interpretace výsledků**, popisy postupů, apod. Jako alternativu je možné odeslat také pouze funkční R-Skript obsahující popisy postupů formou R-kových komentářů (za symbolem #). V tom případě není nutné odevzdávat Wordovský dokument.

Na co si dát závěrem ještě pozor.

- Při programování domácího úkolu nepoužívejte diakritiku. Použití diakritiky vede v důsledku nekonzistentního kódování různých systémů (Windows, Mac, Linux) ke špatnému načítání R Skriptu. Domácí **úkoly s diakritikou budou vráceny k přepracování.**
- Při programování dodržujte jistou přehlednost kódu. Před a za symbolem <- uveďte vždy mezeru, taktéž jednotlivé argumenty funkcí odděľujte mezerami. Příklad správně a přehledně naprogramovaného kódu je k nahlédnutí níže. Správné naprogramování kódu vede ke zvýšení jeho přehlednosti a větší radosti programátora z výsledné práce. :)
- interpretace výsledků jsou nedílnou součástí příkladu a jsou hodnoceny celkem vysokým počtem bodů. **Absence interpretací výsledků tedy může výrazně snížit celkový počet bodů** z jinak správně vypracovaného příkladu.

```
x <- 1:15
px <- dbinom(x, size = 15, p = 0.5)

plot(x, px, type = 'h', lty = 2, lwd = 1,
     main = 'Pravdepodobnostni funkce binomickeho rozdeleni',
     cex.main = 0.9)
points(x, px, pch = 21, col = 'red', bg = 'salmon')

legend('topright', fill = c('salmon'), legend = c('binom'), bty = 'n')
```

Domácí úkoly budou opravovány hromadně po obdržení všech (nebo alespoň významné většiny) řešení úkolů, přičemž na opravení úkolů si vyučující vyhraňuje 14denní lhůtu :).

Přeji vám hodně zdaru při řešení příkladů :).

**Příklad 1 (10 b).** U šestnácti mladých dospělých mužů byla změřena délka pravé dolní končetiny (znak  $X$  - v mm) a délka trupu (znak  $Y$  - v mm). Naměřené hodnoty jsou uvedeny v následující tabulce.

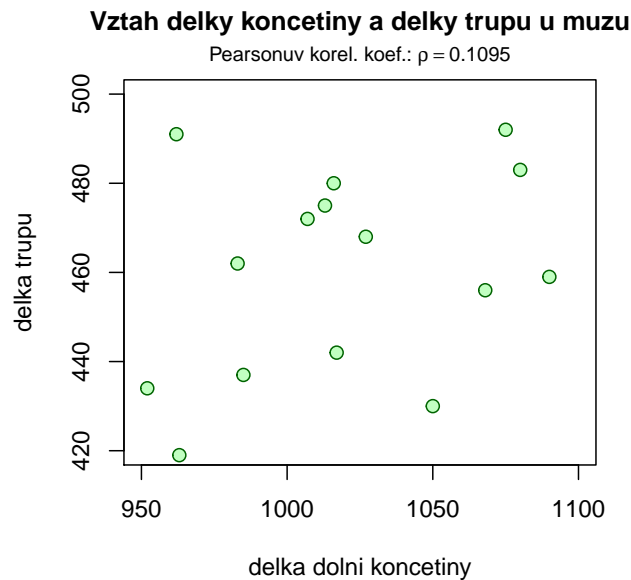
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
d.končetiny	952	1027	1013	1080	1007	1016	983	1068	1114	1017	1090	963	1075	1050	962	985
d.trupu	434	468	475	483	472	480	462	456	428	442	459	419	492	430	491	437

Ze zadaných údajů byly dopočítány následující charakteristiky:  $m_1 = 1025.125$ ,  $m_2 = 458.000$ ,  $s_1 = 49.845$ ,  $s_2 = 23.718$ ,  $s_{12} = 129.400$ .

1. Stanovte hodnotu odhadu korelačního koeficientu  $\rho$  a řádně ji interpretujte. (5 b)

```
## [1] 0.1095
```

2. Nakreslete dvourozměrný tečkový diagram vizualizující vztah mezi délkou dolní končetiny a délkou trupu. (5 b)



**Tip:** Podnadpis do tečkového grafu můžete vložit pomocí příkazu `mtext(bquote(paste('Pear.koef.: ', rho==.(cor.m))), line=0.3, cex=0.8)`, kde `cor.m` je hodnota korelačního koeficientu zaokrouhlená na čtyři desetinná místa.

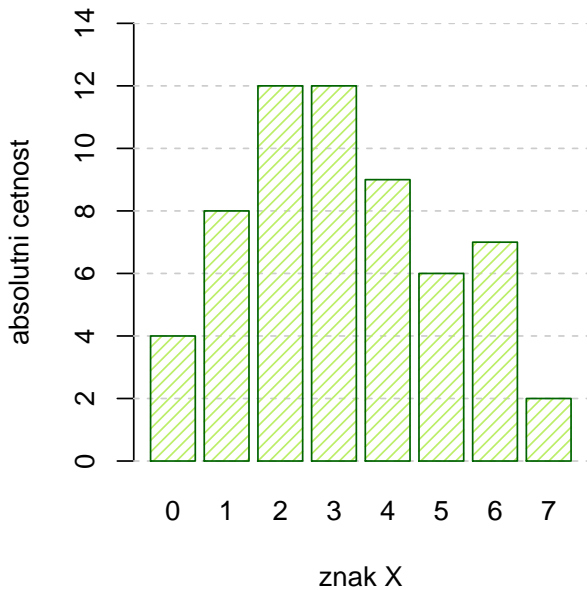
**Příklad 2 (20 b).** V rámci průzkumu ke studii byla sledována četnost znaku  $X$  u celkového počtu 60 subjektů. Hodnoty nasbírané v rámci průzkumu jsou uvedeny v následující tabulce.

$x_i$	0	1	2	3	4	5	6	7
$n_i$	4	8	12	12	9	6	7	2

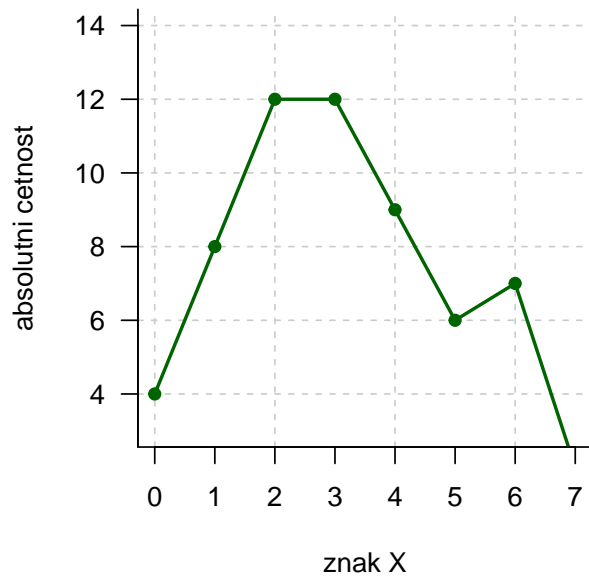
- Vypočítejte dolní kvartil  $x_{0.25}$ , třetí decil  $x_{0.3}$ , medián  $x_{0.5}$ , horní kvartil  $x_{0.75}$  a oktil  $x_{0.8}$  znaku  $X$ . Hodnoty vložte do přehledné tabulky. (5 b)
- Hodnoty  $x_{0.25}$ ,  $x_{0.3}$ ,  $x_{0.5}$ ,  $x_{0.75}$  a  $x_{0.8}$  **řádně interpretujte.** (5 b)
- Vykreslete sloupcový graf a polygon četností znaku  $X$ . (10 b)

##	dolní kvartil	třetí decil	median	horní kvartil	oktil
## 1	2	2	3	4.5	5

**Sloupcový diagram**



**Polygon četností**



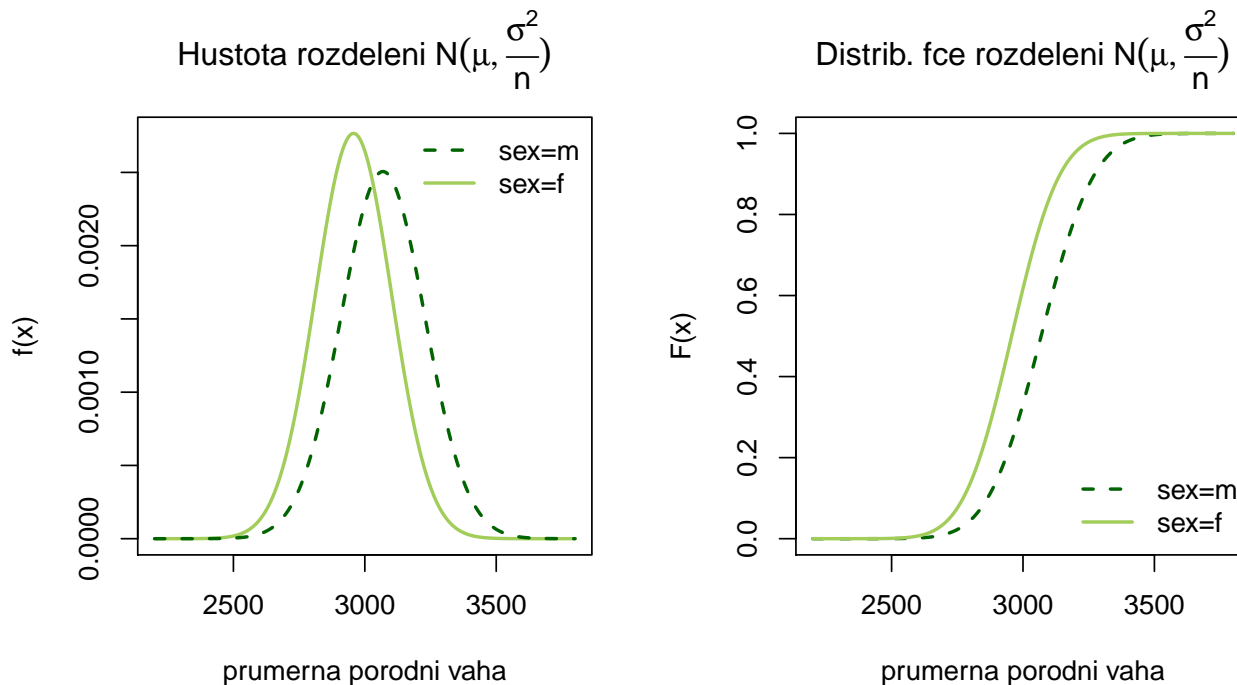
**Příklad 3 (25 b).** Předpokládejme, že porodní hmotnost prvorozených novorozenců – chlapečků je normálně rozdělena okolo střední hodnoty 3069 g se směrodatnou odchylkou 712 g.

1. Vypočítejte pravděpodobnost, že **průměrná hmotnost** dvaceti náhodně vybraných narozených chlapečků bude nejvýše 3000 g. **(7 b)**

**Tip:** S řešením příkladu vám pomůže znění následující věty: Pokud náhodná veličina  $X$  pochází z normálního rozdělení se střední hodnotou  $\mu$  a rozptylem  $\sigma^2$ , tj.  $X \sim N(\mu, \sigma^2)$ , potom průměr  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  pochází z normálního rozdělení se střední hodnotou  $\mu$  a rozptylem  $\frac{\sigma^2}{n}$ , tj.  $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ .

2. Vykreslete graf hustoty normálního rozdělení pro průměrnou hmotnost novorozených chlapečků. Na osu  $x$  naneste posloupnost 1000 hodnot od 2200 g do 3800 g a na osu  $y$  hodnoty hustoty normálního rozdělení průměrné hmotnosti dvaceti novorozených chlapečků ( $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ ). Do grafu dokrešte také křivku hustoty normálního rozdělení pro průměrnou hmotnost dvaceti novorozených holčiček, víte-li, že porodní hmotnost prvorozených novorozenců – holčiček je též normálně rozdělena okolo střední hodnoty 2957 g se směrodatnou odchylkou 645 g. **(9 b)**
3. Vykreslete graf distribuční funkce normálního rozdělení pro průměrnou hmotnost novorozených chlapečků. Na osu  $x$  naneste posloupnost 1000 hodnot od 2200 g do 3800 g a na osu  $y$  hodnoty distribuční funkce normálního rozdělení průměrné hmotnosti dvaceti novorozených chlapečků. Do grafu dokrešte také křivku distribuční funkce normálního rozdělení pro průměrnou hmotnost dvaceti novorozených holčiček, víte-li, že porodní hmotnost prvorozených novorozenců – holčiček je též normálně rozdělena okolo střední hodnoty 2957 g se směrodatnou odchylkou 645 g. **(9 b)**

## [1] 0.3323638



**Příklad 4 (příklad z praxe; 75 b).** Máme k dispozici soubor hodnot délky pravé dolní končetiny a délky trupu u 175 mladých dospělých jedinců (75 mužů ( $m$ ) a 100 žen ( $f$ )). Údaje jsou vloženy v souboru AS-2017-DU-koncetiny-trup.txt. Ve sloupci sex jsou uvedeny údaje o pohlaví, ve sloupci lowex jsou vloženy hodnoty délky pravé dolní končetiny a ve sloupci tru.L jsou vloženy hodnoty délky trupu. Na hladině významnosti  $\alpha = 0.05$  testujte nulovou hypotézu, že délka pravé dolní končetiny u mužů je větší než délka pravé dolní končetiny u žen.

**Tip:** Datový soubor obsahuje neznámé (tzv. NA) hodnoty. Před řešením příkladu je vhodné tyto hodnoty odstranit příkazem `na.omit()`.

V rámci tohoto příkladu proveďte následující úkoly:

1. Pro každé pohlaví sestrojte tabulku základních charakteristik pro délku pravé dolní končetiny. Výsledky v tabulce základních charakteristik zaokrouhlete na čtyři desetinná místa. **(10 b)**

##	n	m	s	min	dolni.kv.	median	horni.kv.	max	sikmost	spicatost
## muzi	75	1008.76	47.0534	915	967	1014	1044	1114	-0.0182	-0.7387
## zeny	100	940.50	45.4712	836	909	943	969	1076	0.2018	0.1084

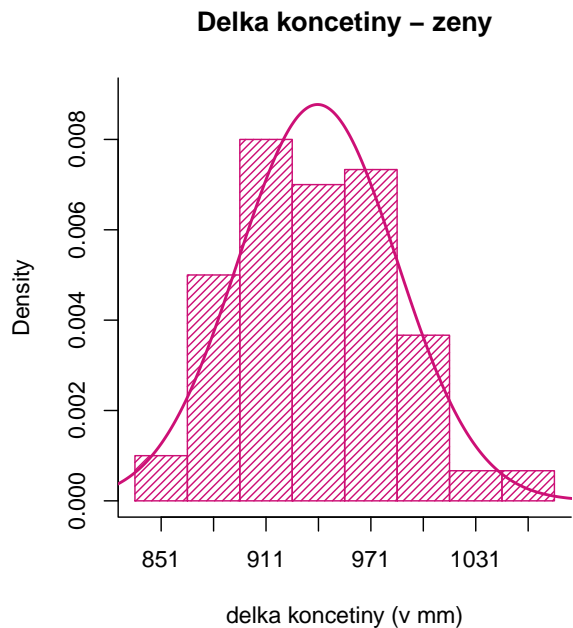
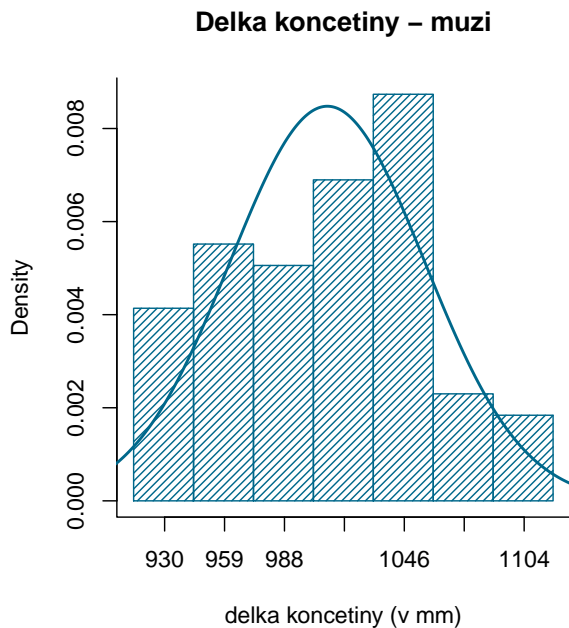
2. Vykreslete histogram pro délku pravé dolní končetiny u mužů. Ohlídejte správný počet třídících intervalů (viz Sturgesovo pravidlo). Histogram superponujte křivkou normálního rozdělení  $N(\mu_m, \sigma_m^2)$ , kde odhad parametřů  $\mu_m$  a  $\sigma_m^2$  získáte z dat. **(5 b)**

**Tip:** Aby se vám křivka vykreslila správně, musíte v histogramu zadat argument `prob=T`. Tento argument převede měřítko  $y$ -ové osy z absolutní škály (na ose  $y$  jsou defaultně nastaveny absolutní četnosti) na relativní škálu (na ose  $y$  budou relativní četnosti).

3. Vykreslete histogram pro délku pravé dolní končetiny u žen. Ohlídejte správný počet třídících intervalů (viz Sturgesovo pravidlo). Histogram superponujte křivkou normálního rozdělení  $N(\mu_f, \sigma_f^2)$ , kde odhad parametřů  $\mu_f$  a  $\sigma_f^2$  získáte z dat. **(5 b)**

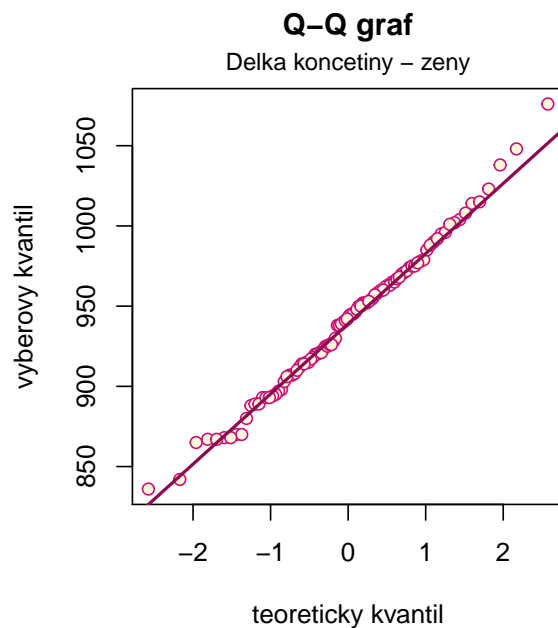
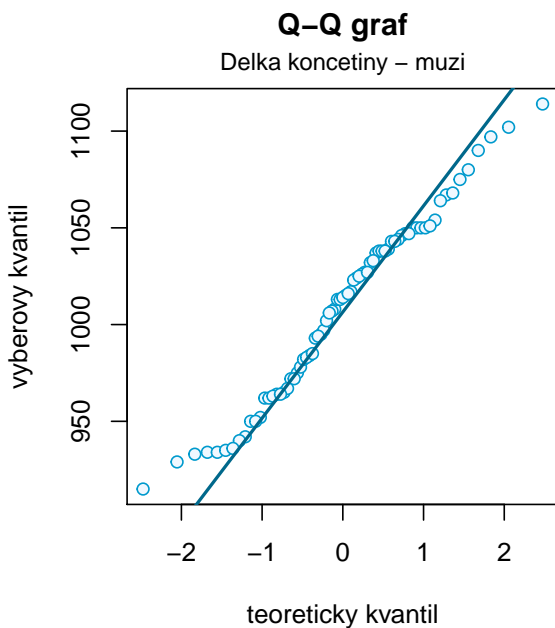
**Tip:** Aby se vám křivka vykreslila správně, musíte v histogramu zadat argument `prob=T`. Tento argument převede měřítko  $y$ -ové osy z absolutní škály (na ose  $y$  jsou defaultně nastaveny absolutní četnosti) na relativní škálu (na ose  $y$  budou relativní četnosti).

```
## [1] 7
## [1] 915 1114
```



4. Otestujte normalitu dat (k otestování zvolte vhodný test a jeho volbu zdůvodněte). K testování patří stanovení hypotéz  $H_0$ ,  $H_1$ , zdůvodněné rozhodnutí o zamítnutí/nezamítnutí  $H_0$  a interpretace výsledku testování. V rámci testu normality vykreslete také pro obě pohlaví Q-Q graf. (10 b)

```
## [1] 0.4271049
## [1] 0.8859276
```



5. Proveďte test o shodě rozptylů  $\sigma_1^2$  a  $\sigma_2^2$ . Testování proveďte kritickým oborem, intervalem spolehlivosti i  $p$ -hodnotou. K testování patří stanovení hypotéz  $H_0$ ,  $H_1$ , zdůvodněné rozhodnutí o zamítnutí/nezamítnutí  $H_0$  (u všech tří typů testování) a interpretace výsledku testování. (20 b)

```
## [1] "Testovací statistika:"
## [1] 1.070804
## [1] "Kritický obor:"
## [1] 0.6468828
## [1] 1.525232
## [1] "Interval spolehlivosti:"
## [1] 0.7020597
## [1] 1.655329
## [1] "p-hodnota:"
## [1] 0.7453007
```

6. Otestujte nulovou hypotézu uvedenou v zadání. Testování proveďte kritickým oborem, intervalem spolehlivosti i  $p$ -hodnotou. K testování patří stanovení hypotéz  $H_0$ ,  $H_1$ , zdůvodněné rozhodnutí o zamítnutí/nezamítnutí  $H_0$  (u všech tří typů testování) a interpretace výsledku testování. (20 b)

```
## [1] "Testovací statistika:"
## [1] 9.681952
## [1] "Kritický obor:"
## [1] -1.653709
## [1] "Interval spolehlivosti:"
## [1] 79.91903
## [1] "p-hodnota:"
## [1] 1
```

7. Závěr testování podložte vykreslením boxplotů pro muže a ženy do jednoho grafu. (5 b)





**Příklad 5 (40 b).** Máme datový soubor obsahující údaje o délce pravé a levé klíční kosti u 33 mužů a 25 žen. Údaje jsou uloženy v souboru AS-2017-DU-klicni-kosti.csv, kde sloupec sex obsahuje údaje o pohlaví, sloupec length.R obsahuje údaje o délce pravé klíční kosti a sloupec length.L obsahuje údaje o délce levé klíční kosti. Na hladině významnosti  $\alpha = 0.05$  testujte nulovou hypotézu, že délka pravé a levé klíční kosti u **žen** jsou stejné.

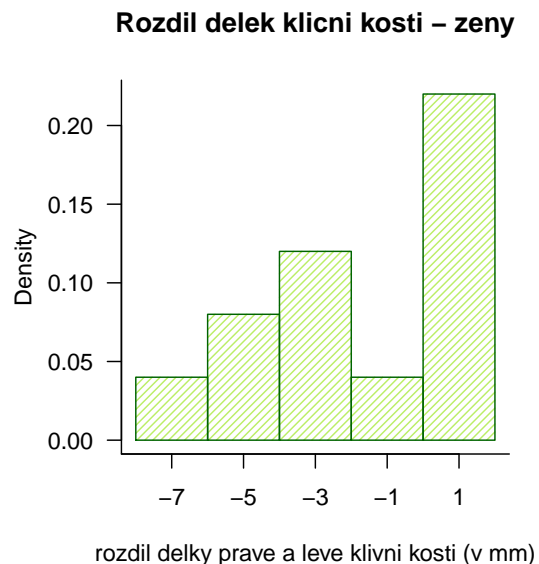
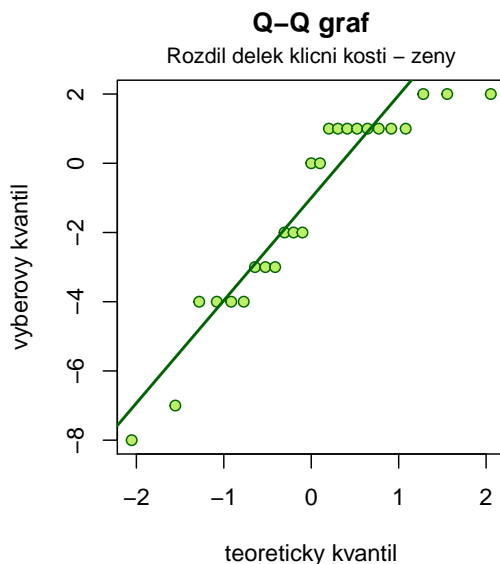
V rámci tohoto příkladu proveďte následující úkoly:

1. Zamyslele se nejprve nad tím, jaký typ dat máte před sebou a jakým typem testu (jednovýběrový, párový, dvouvýběrový) bude nejspíše potřeba nulovou hypotézu otestovat. Typ testu uveďte. **(2.5 b)**
2. Proveďte test normality. **Nápověda:** Nebudete testovat normalitu ani délky pravých klíčních kostí u žen, ani délky levých klíčních kostí u žen, ani normalitu všech údajů dohromady. Čeho tedy máte normalitu testovat? Odpověď souvisí s bodem 1.

K otestování normality zvolte vhodný test a jeho volbu zdůvodněte. K testování patří stanovení hypotéz  $H_0$ ,  $H_1$ , zdůvodněné rozhodnutí o zamítnutí/nezamítnutí  $H_0$  a interpretace výsledku testování. **(5 b)**

```
## [1] 0.007293198
```

3. Testování normality proveďte také graficky, pomocí Q-Q grafu a histogramu. Při vytváření histogramu ohlíďte správný počet třídících intervalů (viz Sturgesovo pravidlo). **Nápověda:** Správným řešením je vykreslení pouze jednoho Q-Q grafu a jednoho histogramu. Oba grafy sestrojte pro tutéž 'veličinu', u které jste testovali normalitu dat. **(10 b)**

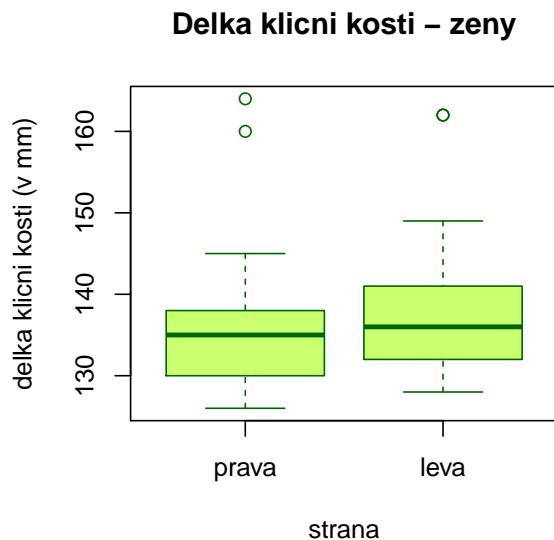


4. Na základě výsledku testování normality vyberte vhodný test k otestování zadané hypotézy. Uveďte jeho celý název. (Začíná na W :)). Testování proveďte kritickým oborem, intervalem spolehlivosti i  $p$ -hodnotou. K testování patří stanovení hypotéz  $H_0$ ,  $H_1$ , zdůvodněné rozhodnutí o zamítnutí/nezamítnutí  $H_0$  (u všech tří typů testování) a interpretace výsledku testování. **(22.5 b)**

```
## [1] "Testovací statistika:"
##      V
## 70.5
## [1] "Kritický obor:"
## [1] 73.4
```

```
## [1] "Interval spolehlivosti:"  
## [1] -2.999972  
## [1] -4.155388e-05  
## [1] "p-hodnota:"  
## [1] 0.0385479
```

5. Výsledek testování podpořte vykreslením krabicového grafu pro délku pravé klíční kosti a pro délku levé klíční kosti u žen. (5 b)



**Příklad 6 (10 b).** Gertruda dostala k svátku balíček sušeného ovoce (rozinky a pomerančová kůra) v mléčné a hořké čokoládě, přičemž 40 % ovoce je zalito v mléčné čokoládě a 60 % ovoce je zalito v hořké čokoládě. V mléčné čokoládě jsou zality rozinky s pravděpodobností 62 %, v hořké čokoládě jsou rozinky zality s pravděpodobností 18 %. S jakou pravděpodobností si Gertruda vytáhne ze sáčku bonbon s pomerančovou kůrou?

```
## [1] 0.644
```

