

7 Úvod do testování hypotéz

Dataset: 11-two-samples-means-skull.txt

Datový soubor 11-two-samples-means-skull.txt obsahuje původní kranio-metrické údaje o basion-bregmatické výšce lebky u 215 dospělých mužů a 107 dospělých žen ze starověké egyptské populace. Data pochází z archivních materiálů (Schmidt, 1888).

Popis proměnných v datasetu:

- id ... pořadové číslo;
- pop ... populace (egant - egyptská starověká);
- sex ... pohlaví jedince (m - muž, f - žena);
- skull.H ... basion-bregmatická výška lebky (v mm).

Příklad 7.1. Test o normalitě dat

Načtete datový soubor 11-two-samples-means-skull.txt. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že náhodný výběr basion-bregmatických výšek lebky (skull.H) mužů starověké egyptské populace pochází z normálního rozdělení.

Řešení příkladu 7.1

- H_0 : Data pochází / nepochází z normálního rozdělení.
- H_1 : Data pochází / nepochází z normálního rozdělení.

Hladina významnosti $\alpha = \dots\dots\dots$

```
      n min max
1 215 119 146
```

1
2

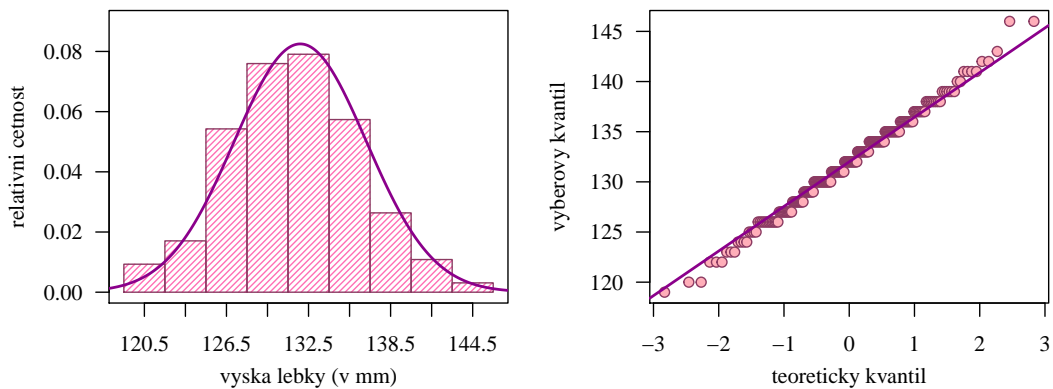
Náhodný výběr obsahuje údaje o basion-bregmatické výšce lebky mužů starověké egyptské populace. Naměřené hodnoty se pohybují v rozmezí-..... mm. Protože rozsah náhodného výběru výšek lebky $n \leq 30 / 30 < n \leq 100 / n > 100$ použijeme na testování hypotézy o normalitě dat Shapirův-Wilkův / Andersonův-Darlingův / Lillieforsův test.

```
[1] 0.1262537
```

3

Protože p -hodnota $p = \dots\dots\dots$ je větší / menší než $\alpha = 0.05$, H_0 zamítáme / nezamítáme na hladině významnosti $\alpha = 0.05$.

Grafická vizualizace rozdělení náhodného výběru



Interpretace výsledků: Náhodný výběr basion-bregmatických výšek lebky mužů starověké egyptské populace pochází / nepochází z normálního rozdělení. ★

Dataset: 15-anova-means-skull.txt

Datový soubor 15-anova-means-skull.txt obsahuje původní kranioметриcké údaje o výšce horní části tváře mužů z německé, malajské, čínské, peruánské a bantuské populace. Data pochází z archivních materiálů (Schmidt, 1888).

Popis proměnných v datasetu:

- id ... pořadové číslo;
- pop ... populace (nem - německá, mal - malajská, cin - čínská, per - peruánská, ban - bantuská);
- sex ... pohlaví jedince (m - muž);
- upface.H ... výška horní části tváře, přímá vzdálenost mezi body *nasion* a *prosthion* (v mm).

Příklad 7.2. Test o normalitě dat

Načtěte datový soubor 15-anova-means-skull.txt. Na hladině významnosti $\alpha = 0.10$ testujte hypotézu, že náhodný výběr výšek horní části tváře (upface.H) mužů německé populace pochází z normálního rozdělení.

Řešení příkladu 7.2

- H_0 : Data z normálního rozdělení.
- H_1 : Data z normálního rozdělení.

Hladina významnosti $\alpha =$

```
n min max
1 19 62 76
```

4
5

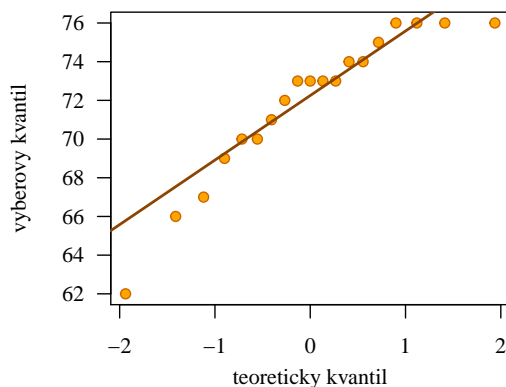
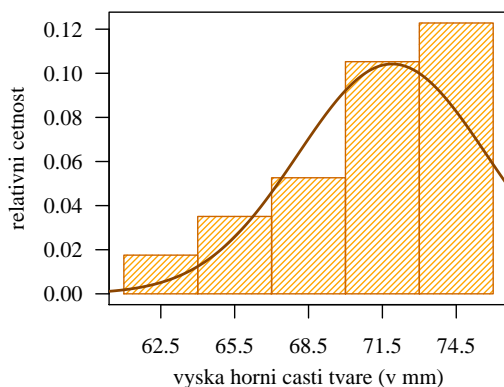
Náhodný výběr obsahuje údaje o výšce horní části tváře mužů německé populace. Naměřené hodnoty se pohybují v rozmezí-..... mm. Protože rozsah náhodného výběru výšek horní části tváře $n \leq 30$ / $30 < n \leq 100$ / $n > 100$ použijeme na testování hypotézy o normalitě dat Shapirův-Wilkův / Andersonův-Darlingův / Lillieforsův test.

```
[1] 0.04190113
```

6

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

Grafická vizualizace rozdělení náhodného výběru



Interpretace výsledků: Náhodný výběr výšek horní části tváře mužů německé populace z normálního rozdělení. ★

Dataset: 19-more-samples-correlations-skull.txt

Datový soubor 19-more-samples-correlations-skull.txt obsahuje údaje o šířce nosu a o interorbitální šířce mužů z německé, malajské, čínské, peruánské a bantuské populace. Data pochází z archivních materiálů (Schmidt, 1888).

Popis proměnných v datasetu:

- pop ... populace (nem - německá, mal - malajská, cin - čínská, per - peruánská, ban - bantuská);
- sex ... pohlaví jedince (m - muž);
- nose.B ... šířka nosu (v mm);
- intorb.B ... interorbitální šířka (v mm).

Příklad 7.3. Test o dvourozměrné normalitě dat

Načtete datový soubor 19-more-samples-correlations-skull.txt. Nechť náhodná veličina X popisuje šířku nosu a náhodná veličina Y popisuje interorbitální šířku mužů peruánské populace. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že náhodný vektor $(X, Y)^T$ pochází z dvourozměrného normálního rozdělení.

Řešení příkladu 7.3

- H_0 : Data pochází / nepochází z dvourozměrného normálního rozdělení.
- H_1 : Data pochází / nepochází z dvourozměrného normálního rozdělení.

Hladina významnosti $\alpha = \dots\dots\dots$

	n	min_X	max_X	min_Y	max_Y
1	46	19	26	19	28

7
8

Náhodný výběr obsahuje údaje o šířce nosu a interorbitální šířce mužů peruánské populace. Hodnoty šířky nosu se pohybují v rozmezí–..... mm, hodnoty interorbitální šířky se pohybují v rozmezí–..... mm.

Dvourozměrnou normalitu otestujeme Mardiovým testem. Ten sestává z dvou částí: (a) z testu šikmosti

- H_{0a} : Data vykazují / nevykazují kladné nebo záporné zešikmení.
- H_{1a} : Data vykazují / nevykazují kladné ani záporné zešikmení.

(b) z testu špičatosti

- H_{0b} : Data vykazují / nevykazují zešpičatění nebo zploštění.
- H_{1b} : Data vykazují / nevykazují zešpičatění ani zploštění.

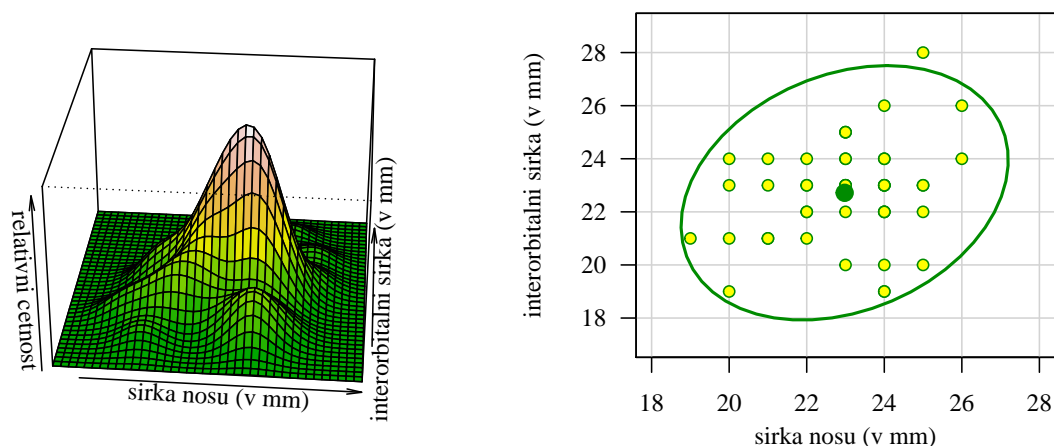
Data pochází z dvourozměrného normálního rozdělení, pokud nevykazují ani zešikmení ani zešpičatění.

	Test	Statistic	p value	Result
1	Mardia Skewness	4.27819772855481	0.369663150730262	YES
2	Mardia Kurtosis	-0.0684871107744411	0.945397880096616	YES
3	MVN	<NA>	<NA>	YES

9
10
11
12

Protože p -hodnota testu šikmosti $p = \dots\dots\dots$ je menší / větší než $\alpha = 0.05$, H_{0a} zamítáme / nezamítáme na hladině významnosti $\alpha = 0.05$. Data vykazují / nevykazují výrazné zešikmení. Protože p -hodnota testu špičatosti $p = \dots\dots\dots$ je menší / větší než $\alpha = 0.05$, H_{0b} zamítáme / nezamítáme na hladině významnosti $\alpha = 0.05$. Data vykazují / nevykazují výrazné zešpičatění nebo zploštění.

Grafická vizualizace rozdělení náhodného výběru



Mimo 95% elipsu spolehlivosti leží z celkového počtu pozorování, tj.%, tedy více / méně než 5% dat.

Interpretace výsledků: Náhodný výběr šířky nosu a interorbitální šířky mužů peruánské populace pochází / nepochází z dvourozměrného normálního rozdělení. ★

Dataset: 16-anova-head.txt

Datový soubor 16-anova-head.txt obsahuje antropometrické údaje mladých dospělých lidí, převážně studentů vysokých škol z Brna a Ostravy (Králík, nepublikovaná data).

Popis proměnných v datasetu:

- sex ... pohlaví jedince (m - muž, f - žena);
- sexor ... sexuální orientace (op - výlučně opačné pohlaví, sa - ostatní)
- obra ... existence staršího bratra (yes - ano, no - ne)
- head.L ... délka hlavy, vzdálenost mezi body *glabella* a *opisthocranion* (v mm);
- bizyg.W ... šířka tváře, vzdálenost obou bodů *zygion dakryon* (v mm).

Příklad 7.4. Test o dvourozměrné normalitě dat

Načtete datový soubor 16-anova-head.txt. Nechť náhodná veličina X popisuje *délku hlavy* a náhodná veličina Y popisuje *šířku tváře* žen. Na hladině významnosti $\alpha = 0.10$ testujte hypotézu, že náhodný vektor $(X, Y)^T$ pochází z dvourozměrného normálního rozdělení.

Řešení příkladu 7.4

- H_0 : Data z dvourozměrného normálního rozdělení.
- H_1 : Data z dvourozměrného normálního rozdělení.

Hladina významnosti $\alpha = \dots\dots\dots$

	n	min_X	max_X	min_Y	max_Y
1	100	170	205	120	151

13
14

Náhodný výběr obsahuje údaje o délce hlavy a šířce tváře žen. Naměřené hodnoty délky hlavy se pohybují v rozmezí-..... mm, naměřené hodnoty šířky tváře se pohybují v rozmezí-..... mm.

Dvourozměrnou normalitu otestujeme Mardiovým testem. Ten sestává z dvou částí: (a) z testu šikmosti

- H_{0a} : Data kladné ani záporné zešikmení.
- H_{1a} : Data kladní ani záporné zešikmení.

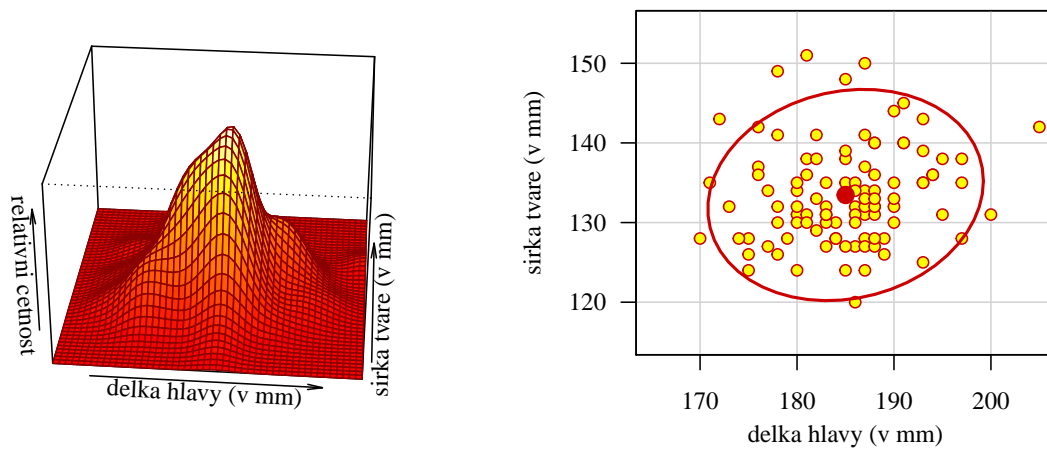
(b) z testu špičatosti

- H_{0b} : Data zešpičatění nebo zploštění.
- H_{1b} : Data zešpičatění nebo zploštění.

	Test	Statistic	p value	Result
1	Mardia Skewness	9.8656660128771	0.042752347786637	NO
2	Mardia Kurtosis	0.604444695534365	0.545548027690992	YES
3	MVN	<NA>	<NA>	NO

15
16
17
18

Protože p -hodnota testu šikmosti $p = \dots\dots\dots$ je větší / menší než $\alpha = 0.10$, H_{0a} zamítáme / nezamítáme na hladině významnosti $\alpha = 0.10$. Data vykazují / nevykazují výrazné zešikmení. Protože p -hodnota testu špičatosti $p = \dots\dots\dots$ je větší / menší než $\alpha = 0.10$, H_{0b} zamítáme / nezamítáme na hladině významnosti $\alpha = 0.10$. Data vykazují / nevykazují výrazné zešpičatění nebo zploštění.



Mimo 90% elipsu spolehlivosti leží z celkového počtu pozorování, tj. %, tedy než 10 % dat.

Interpretace výsledků: Náhodný výběr délek hlavy a šířek tváře žen z dvourozměrného normálního rozdělení. ★