

6 Testy o normalitě dat

- Normalita = nepostradatelný předpoklad parametrických testů (jednovýběrových, párových, dvouvýbě., ...)
- Stanovení hypotéz:
 - H_0 : Data pochází z normálního rozdělení.
 - H_1 : Data nepochází z normálního rozdělení.
- Testy normality:
 - Shairo-Wilkův test: `shapiro.test()`
 - Lillie-Forsův test: `lillie.test()` [`nortest`]
 - Anderson-Darlingův test: `ad.test()` [`nortest`]
- výstup testů = p -hodnota: $p > \alpha \rightarrow H_0$ nezamítáme; $p \leq \alpha \rightarrow H_0$ **zamítáme**
- Grafické ověření normality:
 - histogram + křivka hustoty normálního rozdělení
 - Q-Q plot `qqnorm()` a `qqline()`

Příklad 6.1. Test o normalitě dat

Načtěte datový soubor `01-one-sample-mean-skull-mf.txt` a odstraňte z načtených dat NA hodnoty. Mějme náhodnou veličinu X popisující *největší šířku mozkovny* u skeletů mužského pohlaví. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že náhodný výběr naměřených hodnot největší šířky mozkovny u skeletů mužského pohlaví pochází z normálního rozdělení.

Řešení příkladu 6.1

Nejprve stanovíme nulovou a alternativní hypotézu.

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

Hladina významnosti $\alpha = \dots$. Nyní zjistíme rozsah náhodného výběru.

```
[1] 216
```

1

Protože náhodný výběr naměřených hodnot největší šířky mozkovny u skeletů mužského pohlaví má rozsah, což je než 30, použijeme na testování hypotézy o normalitě dat test.

```
[1] 0.07662229
```

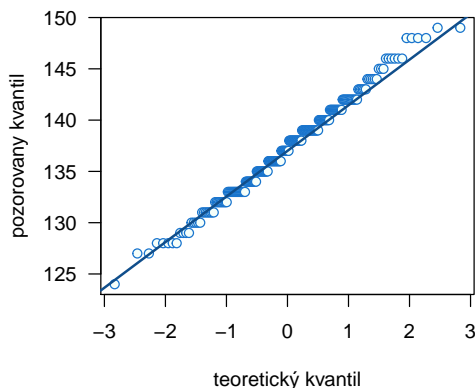
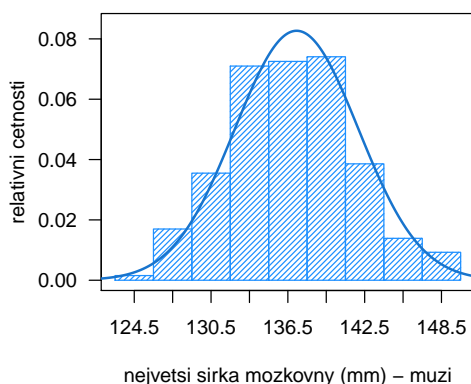
2

P -hodnota vyšla Protože p -hodnota α , H_0 na hladině významnosti $\alpha = 0.05$. Data z normálního rozdělení.

Příklad 6.2. Grafická vizualizace normality dat

V příkladu 6.1 jsme na základě testování stanovili, že naměřené hodnoty největší šířky mozkovny u skeletů mužského pohlaví z normálního rozdělení. Závěr stanovený v příkladu 6.1 podpořte grafickou vizualizací.

Řešení příkladu 6.2



7 Jednovýběrové testy o střední hodnotě, rozptylu a směrodatné odchylce

Prolog: Z archivních materiálů (Schmidt, 1888; 01-one-sample-mean-skull-mf.txt) máme k dispozici původní kranio-metrické údaje o délce a šířce mozkovny ze starověké egyptské populace. Současně máme k dispozici průměrné hodnoty obou rozměrů, hodnoty směrodatné odchylky a počty případů ze vzorku novověké egyptské populace (délka mozkovny: $\bar{x}_m = 177.568$ mm, $\bar{x}_f = 171.962$ mm, $s_m = 7.526$ mm, $s_f = 5.361$ mm, $n_m = 88$, $n_f = 52$; a šířka mozkovny: $\bar{x}_m = 136.402$ mm, $\bar{x}_f = 131.038$ mm, $s_m = 6.411$ mm, $s_f = 5.361$ mm, $n_m = 87$, $n_f = 52$).

Příklad 7.1. Jednovýběrový test o střední hodnotě

Načtěte datový soubor 01-one-sample-mean-skull-mf.txt a odstraňte z načtených dat NA hodnoty. Na hladině významnosti $\alpha = 0.05$ otestujte nulovou hypotézu o shodě střední hodnoty šířky mozkovny starověké egyptské mužské populace se střední hodnotou šířky mozkovny novověké egyptské mužské populace 136.402 mm.

Řešení příkladu 7.1

Nejprve stanovíme nulovou a alternativní hypotézu.

- H_0 :
- H_1 : (..... alternativa).
- Hladina významnosti $\alpha =$
- Test o když známe / neznáme.

Nutným předpokladem umožňujícím použití parametrického testu na otestování nulové hypotézy je **normalita naměřených hodnot**. Tu jsme ověřili v rámci příkladu 6.1, kde jsme na hladině významnosti $\alpha = 0.05$ nezamítli nulovou hypotézu o normálním rozdělení největší šířky mozkovny u skeletů mužského pohlaví.

a) Testování pomocí kritického oboru

[1] 2.385757 3

[1] -1.971059 4

[1] 1.971059 5

Kritický obor má tvar Protože, H_0 na hladině významnosti $\alpha =$

b) Test pomocí intervalu spolehlivosti

Proti alternativě postavíme IS.

[1] 136.5381 6

[1] 137.8322 7

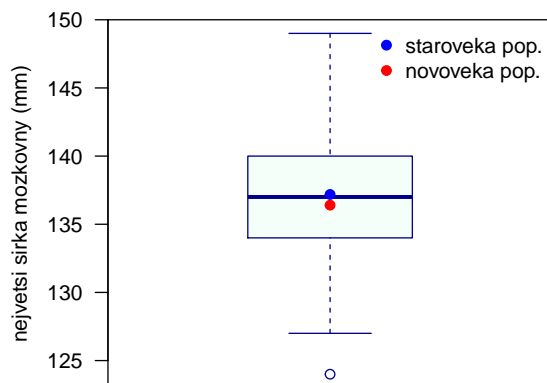
Interval spolehlivosti má tvar Protože, H_0 na hladině významnosti $\alpha =$

c) Test pomocí p -hodnoty

[1] 0.01791157 8

p -hodnota vyšla Protože p -hodnota, H_0 na hladině významnosti $\alpha =$

Interpretace výsledků: Mezi největší šířkou mozkovny u mužů starověké egyptské populace a u mužů novověké egyptské populace existuje / neexistuje statisticky významný rozdíl.



Příklad 7.2. Jednovýběrový test o rozptylu

Načtete datový soubor 01-one-sample-mean-skull-mf.txt a odstraňte z načtených dat NA hodnoty. Na hladině významnosti $\alpha = 0.1$ otestujte nulovou hypotézu o vyšším rozptylu šířky mozkovny starověké egyptské mužské populace vzhledem k rozptylu šířky mozkovny novověké egyptské mužské populace ($s_m = 6.411$ mm).

Řešení příkladu 7.2

Nejprve stanovíme nulovou a alternativní hypotézu.

- H_0 :
- H_1 : (..... alternativa).
- Hladina významnosti $\alpha =$
- Test o když známe / neznáme.

Nutným předpokladem umožňujícím použití parametrického testu na otestování H_0 je naměřených hodnot. Tu jsme ověřili v rámci příkladu 6.1, kde jsme na hladině významnosti $\alpha = 0.05$ nezamítli nulovou hypotézu o normálním rozdělení největší šířky mozkovny u skeletů mužského pohlaví.

a) Testování pomocí kritického oboru

[1] 121.7635 9

[1] 188.8901 10

Kritický obor má tvar Protože, H_0 na hladině významnosti $\alpha =$

b) Test pomocí intervalu spolehlivosti

Proti alternativě postavíme IS.

[1] 26.49473 11

Interval spolehlivosti má tvar Protože, H_0 na hladině významnosti $\alpha =$

c) Test pomocí p -hodnoty

[1] 4.350332e-08 12

p -hodnota vyšla Protože p -hodnota, H_0 na hladině významnosti $\alpha =$

Interpretace výsledků: Rozptyl největší šířky mozkovny u starověké egyptské mužské populace je statisticky významně než rozptyl největší šířky mozkovny u novověké egyptské mužské populace.

Příklad 7.3. Jednovýběrový test o směrodatné odchylce

Načtete datový soubor 01-one-sample-mean-skull-mf.txt a odstraňte z načtených dat NA hodnoty. Na hladině významnosti $\alpha = 0.1$ otestujte, zda je směrodatná odchylka největší šířky mozkovny starověké egyptské mužské populace statisticky významně nižší než směrodatná odchylka šířky mozkovny novověké egyptské mužské populace ($s_m = 6.411$ mm).

Řešení příkladu 7.3

Nejprve stanovíme nulovou a alternativní hypotézu.

- H_0 : \rightarrow
- H_1 : \rightarrow (..... alternativa).
- Hladina významnosti $\alpha =$
- Test o když známe / neznáme.

Nutným předpokladem umožňujícím použití parametického testu na otestování H_0 je naměřených hodnot. Tu jsme ověřili v rámci příkladu 6.1, kde jsme na hladině významnosti $\alpha = 0.05$ nezamítli nulovou hypotézu o normálním rozdělení největší šířky mozkovny u skeletů mužského pohlaví. Protože jsme test o směrodatné odchylce převedli na test o rozptylu, postupovali bychom nyní úplně stejně jako v příkladu 7.2.

Interpretace výsledků: Směrodatná odchylka největší šířky mozkovny u starověké egyptské mužské populace je / není statisticky významně nižší než směrodatná odchylka největší šířky mozkovny u novověké egyptské mužské populace.