

# Cvičení MV011 Statistika I

## 10. Testování statistických hypotéz

Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno

jaro 2019



Funkce pro testování některých statistických hypotéz v *R*:

- `z.test` (v knihovně `BSDA`): Z-test střední hodnoty při známém rozptylu
- `t.test`: T-test střední hodnoty při neznámém rozptylu
- `var.test`: F-test podílu rozptylů dvou výběrů
- `binom.test`: test v binomickém rozdělení

Pro připomenutí – funkce pro kvantily v *R*:

- `qnorm`:  $\mu$ , kvantily standardizovaného normálního rozdělení
- `qt`:  $t$ , kvantily Studentova  $t$ -rozdělení
- `qf`:  $F$ , kvantily Fisherova-Snedecorova  $F$ -rozdělení
- `qchisq`:  $\chi^2$ , kvantily chí-kvadrát rozdělení

Pro řešení využijte vzorce ve slajdech 9. přednášky a *R*-skript z 9. přednášky.

Testujte příslušné statistické hypotézy na hladině významnosti 95 % pomocí:

- pivotové statistiky a kritického oboru,
- $p$ -hodnoty ( $p$ -value),
- intervalu spolehlivosti pro odpovídající parametr.

### Příklad 1 (z 9. přednášky)

*Při 40 hodech mincí byl rub zaznamenán 22krát.*

*Je důvod se domnívat, že rub nepadá stejně často jako líc?*

### Příklad 2 (z 9. přednášky)

*Na základě testu máme na 5% hladině významnosti rozhodnout, zda produkce vajec plemene kornýšek černých je nižší než plemene leghornek bílých. Náhodně jsme vybrali 50 kornýšek a 40 leghornek, u nichž byla zjištěna průměrná roční produkce na slepici: kornýška 275, leghornka 280. Z dřívějších pozorování jsou známy rozptyly  $\sigma_1^2 = 48$  (kornýška),  $\sigma_2^2 = 41$  (leghornka).*

### Příklad 3

*Datový soubor [spotreba.csv](#): spotřeba auta v l / 100 km při 11 nezávislých zkouškách. (a) Může být střední spotřeba paliva rovna 9 l / 100 km?*

*(b) Může být směrodatná odchylka spotřeby paliva rovna 0,5 l / 100 km?*

### Příklad 4

*Testujte nulovou hypotézu, že směrodatná odchylka normálně rozdělené náhodné veličiny je rovna 300, když na vzorku rozsahu 25 byl spočítán výběrový průměr 3118 a výběrová směrodatná odchylka 357.*

### Příklad 5

Datový soubor `pevnost.csv`: pevnost vlákna při zkouškách pevnosti bavlněného vlákna. Rozptyl pevnosti bavlněného vlákna nemá překročit hodnotu 0,36. Je důvod k podezření na vyšší nestejnoměrnost, než je stanovená?

### Příklad 6

Datový soubor `SiO2.csv`: měření obsahu  $\text{SiO}_2$  ve strusce analytickou (A) metodou a fotokolorometrickou (B) metodou. (a) Je mezi rozptyly měření získanými jednotlivými metodami statistický rozdíl? (b) Je mezi středními hodnotami měření získanými jednotlivými metodami statistický rozdíl?

### Příklad 7

Datový soubor `zakaznici.csv`: počty zákazníků odbavených u pokladny ve 20 sledovaných minutových intervalech. Může být střední hodnota doby odbavení jednoho zákazníka rovna 20 sekundám?

### Příklad 8

Datový soubor `kola.csv`: zápůjčky jízdních kol ve Washingtonu D.C. Pro jednotlivé kalendářní měsíce roku (viz sloupec `mnth` = 1, ..., 12) testujte nulovou hypotézu, že střední počet zápůjček kol v daném měsíci je rovný 5 400.

## Výsledky

1. ne
2.  $H_0$  zamítáme, kornýšky mají horší produkci vajec než leghornky
3. (a) ano, (b) ne
4.  $H_0$  nezamítáme
5. ano
6. (a) ne, (b) ano
7. ne
8.  $H_0$  nezamítáme v měsících 5, 7–10, zamítáme v měsících 1–4, 6, 11, 12