1 Parametrické úlohy o dvou nezávislých náhodných výběrech z normálních rozdělení a jednom náhodném výběru z alternativního rozdělení

Parametrické úlohy o dvou nezávislých náhodných výběrech z normálních rozdělení

Příklad 1.1. Interval spolehlivosti pro parametrickou funkci $\mu_1 - \mu_2$: Bylo vylosováno 11 stejně starých selat téhož plemene. Šesti z nich byla předepsána výkrmná dieta č. 1 a zbylým pěti výkrmná dieta č. 2. Průměrné denní přírůstky v Dg za dobu půl roku jsou následující:

dieta č. 1	62	54	55	60	53	58
dieta č. 2	52	56	49	50	51	

Zjištěné hodnoty považujeme za realizace dvou nezávislých náhodných výběrů pocházejících z rozdělení $N(\mu_1, \sigma^2)$ a $N(\mu_2, \sigma^2)$. Sestrojte 95 % empirický interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$.

```
## [1] "dh = 0.992"
## [1] "hh = 9.808"
```

$$IS = (0.9920\,;\,9.8080)$$

S pravděpodobností alespoň 0.95 platí, že $0.99\,\mathrm{Dg} < \mu_1 - \mu_2 < 9.81\,\mathrm{Dg}$.

Příklad 1.2. Testování hypotéz o parametrických funkcích $\mu_1 - \mu_2$, σ_1^2/σ_2^2:

1. Pro datový soubor z příkladu 1.1 testujte na hladině významnosti $\alpha = 0.05$ hypotézu, že
   a) rozptyly hmotnostních přírůstků selat při obou výkrmných dietách jsou shodné;
   b) obě výkrmné diety mají stejný vliv na hmotnostní přírůstky selat.
2. Výsledek testování podpořte krabicovým diagramem.

Shapirův test normality

```
## [1] "Dieta 1: 0.6195"
## [1] "Dieta 2: 0.4272"
```

ad a) Testování hypotézy o shodě rozptylů.

i. Testování pomocí kritického oboru

```
## [1] "t0 = 1.7534"
## [1] "w1 = 0.1354"
## [1] "w2 = 9.3645"
```

ii. Testování pomocí intervalu spolehlivosti

```
## [1] "dh = 0.1872"
## [1] "hh = 12.9541"
```

iii. Testování pomocí p-hodnoty

```
## [1] "p.val = 0.6063"
```

H_0 o shodě rozptylů σ_1^2 a σ_2^2 nezamítáme na hladině významnosti $\alpha = 0.05$.

Upozornění: V případě zamítnutí hypotézy o shodě rozptylů by bylo zapotřebí použít test se samostatnými odhady rozptylu.

ad b) Testování hypotézy o shodě středních hodnot

i. Testování pomocí kritického oboru

```
## [1] "t0 = 2.7712"
## [1] "w1 = -2.2622"
## [1] "w2 = 2.2622"
```

ii. Testování pomocí intervalu spolehlivosti
V příkladu 1.1 jsme zjistili, že 95 % oboustranný interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$ má tvar

$$IS = (0.9920\,;\,9.8080).$$

iii. Testování pomocí p-hodnoty

```
## [1] "p.val = 0.0217"
```

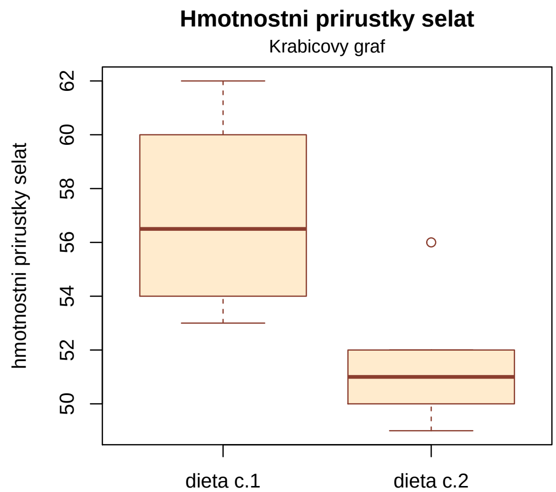
H_0 o shodě středních hodnot μ_1 a μ_2 zamítáme na hladině významnosti $\alpha = 0.1$.

Poznámka: K otestování nulové hypotézy o rozdílu středních hodnot $\mu_1 - \mu_2$ dvou nezávislých náhodných výběrů z normálních rozdělení můžeme použít funkci t.test(x,y) s argumentem alternative='two.sided' (oboustranná alternativa) a argumentem var.equal=T (rozptyly obou náhodných výběrů si jsou rovné).

```
x <- c(62, 54, 55, 60, 53, 58)
y <- c(52, 56, 49, 50, 51)
t.test(x, y, alternative='two.sided', var.equal=T)
```

Upozornění: Pokud bychom na hladině významnosti $\alpha = 0.05$ zamítli nulovou hypotézu o shodě rozptylů σ_1^2 a σ_2^2, mohli bychom k otestování nulové hypotézy o shodě středních hodnot μ_1 a μ_2 použít opět funkci t.test s argumentem alternative='two.sided' (oboustranná alternativa) a argumentem var.equal=F. Tento argument modifikuje klasický t-test na t-test s Welschovou aproximací stupňů volnosti, která se používá v případě, že rozptyly obou náhodných výběrů nejsou shodné.

Krabicový diagram



Příklad k samostatnému řešení

Příklad 1.3. Načtěte datový soubor vyska.txt, který obsahuje údaje o výšce 48 studentek VŠE v Praze (proměnná vyska) a obor jejich studia (1 – národní hospodářství, 2 – informatika).

a) Pomocí S-W testu ověřte na hladině významnosti $\alpha = 0.1$ předpoklad o normalitě výšek v obou skupinách studentek.

b) Na hladině významnosti $\alpha = 0.1$ testujte hypotézu o shodě rozptylů výšek studentek v daných dvou oborech studia.

c) Na hladině významnosti $\alpha = 0.1$ testujte hypotézu o shodě středních hodnot výšek studentek v daných dvou oborech studia.

d) Výpočet doplňte krabicovými diagramy.

Shapirův test normality

```
## [1] "Narodni hospodarstvi: 0.6068"
## [1] "Informatika: 0.1119"
```

Testování hypotézy o shodě rozptylů

i. Testování pomocí kritického oboru

```
## [1] "t0 = 1.9873"
## [1] "w1 = 0.5033"
## [1] "w2 = 2.0905"
```

ii. Testování pomocí intervalu spolehlivosti

```
## [1] "dh = 0.9506"
## [1] "hh = 3.9487"
```

iii. Testování pomocí p-hodnoty

```
## [1] "p.val = 0.1249"
```

H_0 o shodě rozptylů σ_1^2 a σ_2^2 nezamítáme na hladině významnosti $\alpha = 0.1$.

Testování hypotézy o shodě středních hodnot

i. Testování pomocí kritického oboru

```
## [1] "t0 = 1.744"
## [1] "w1 = -1.6787"
## [1] "w2 = 1.6787"
```

ii. Testování pomocí intervalu spolehlivosti

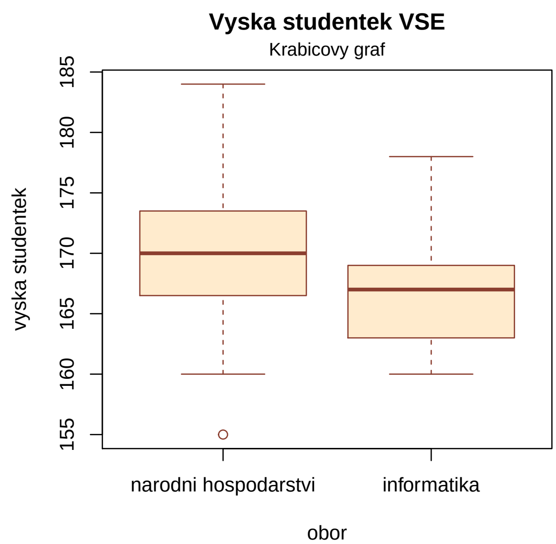
```
## [1] "dh = 0.1095"
## [1] "hh = 5.7334"
```

iii. Testování pomocí p-hodnoty

```
## [1] "p.val = 0.0878"
```

H_0 o shodě středních hodnot μ_1 a μ_2 zamítáme na hladině významnosti $\alpha = 0.1$.

Krabicový diagram



1.1 Parametrické úlohy o jednom náhodném výběru z alternativního rozdělení

Příklad 1.4. Asymptotický interval spolehlivosti pro parametr θ alternativního rozdělení: Může politická strana, pro niž se v předvolebním průzkumu vyslovilo 60 z 1000 dotázaných osob, očekávat se spolehlivostí 0.95, že by v této době ve volbách překročila 5 % hranici pro vstup do parlamentu?

Ověření podmínky $n\theta(1-\theta) > 9$: $1000 * 0.06 * 0.94 = 56.4 > 9$.

```
## [1] "dh = 0.0476"
```

95 % empirický interval spolehlivosti má tvar:

$$(0.0476. \, ; \, \infty)$$

S pravděpodobností přibližně 0.95 je tedy $\theta > 0.0476$. Protože tento interval zahrnuje i hodnoty nižší než 0.05, nelze vyloučit, že strana získá méně než 5 % hlasů.

Příklad k samostatnému řešení

Příklad 1.5. Přírůstky cen akcií na burze (v %) u 10 náhodně vybraných společností dosáhly těchto hodnot: 10, 16, 5, 10, 12, 8, 4, 6, 5, 4. Sestrojte 95 % asymptotický empirický interval spolehlivosti pro pravděpodobnost, že přírůstek ceny akcie překročí 8.5 %.

```
## [1] "dh = 0.0964"
## [1] "hh = 0.7036"
```

$0.096 < \theta < 0.704$ s pravděpodobností aspoň 0.95.

Příklad 1.6. Testování hypotézy o parametru θ alternativního rozdělení: Určitá cestovní kancelář organizuje zahraniční zájezdy podle individuálních přání zákazníků. Z několika minulých let ví, že 30 % všech takto organizovaných zájezdů má za cíl zemi X. Po zhoršení politických podmínek v této zemi se cestovní kancelář obává, že se zájem o tuto zemi mezi zákazníky sníží. Ze 150 náhodně vybraných zákazníků v tomto roce má 38 za cíl právě zemi X. Potvrzují nejnovější data pokles zájmu o tuto zemi? Volte hladinu významnosti $\alpha = 0.05$.

Splnění podmínky $n\theta(1-\theta) > 9$: $150 * 0.3 * 0.7 = 31.5 > 9$.

a) Testování pomocí kritického oboru

```
## [1] "t0 = -1.2472"
## [1] "w1 = -1.6449"
```

b) Testování pomocí intervalu spolehlivosti

```
## [1] "hh = 0.3117"
```

$$IS = (-\infty\,; 0.3117)$$

c) Testování pomocí p-hodnoty

```
## [1] "p.val = 0.1062"
```

H_0 nezamítáme na asymptotické hladině významnosti $\alpha = 0.05$.