

# Statistinės duomenų analizės praktinės užduotys

2017

```
library(knitr)
set.seed(42)
```

## 6. Parametrinių hipotezių tikrinimas.

- (a) *Hipotezė apie normalaus skirstinio vidurkio reikšmę: Stjudento kriterijus vienai normaliajai imčiai* (ČM III.3.3.4). Užkandžiais prekiaujanti firma nusprendė mėšainius su žuvimi pakeisti mėšainiais su bananais. Dvylikoje užkandinių per savaitę buvo parduota atitinkamai 530, 540, 510, 500, 520, 532, 540, 515, 517, 522, 530 ir 510 naujųjų užkandžių. Anksčiau kiekviena užkandinė parduodavo vidutiniškai po 520 senųjų užkandžių per savaitę. Ar naujoji produkcija blogiau perkama? Reikšmingumo lygmuo 0.05.

```
sales <- c(530, 540, 510, 500, 520, 532,
          540, 515, 517, 522, 530, 510)

mu_0 <- 520
```

Patikrinkime normalumo prielaidą. Kolmogorovo - Smirnovo kriterijus:

```
ks.test(sales, pnorm)
```

```
## Warning in ks.test(sales, pnorm): ties should not be present for the
## Kolmogorov-Smirnov test
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: sales
## D = 1, p-value = 7.55e-11
## alternative hypothesis: two-sided
```

Normalumo prielaida galioja. Tikrinkime hipotezę:

$$H : \mu \leq \mu_0, \mu_0 = 520$$

```
t <- t.test(sales, alternative="g", mu=mu_0, conf.level=.95)
pv <- t$p.value
t
```

```
##
## One Sample t-test
##
## data: sales
## t = 0.59924, df = 11, p-value = 0.2806
## alternative hypothesis: true mean is greater than 520
## 95 percent confidence interval:
## 515.6733      Inf
## sample estimates:
## mean of x
## 522.1667
```

$p_v = 0.280576 > \alpha = 0.05$ , duomenys neprieštarauja hipotezei. Statistiškai, naujoji produkcija nėra blogiau perkama (imtis nedidelė).

- (b) *Hipotezė apie proporcijos reikšmę* (ČM III.3.3.7). Naujo medikamento reklamoje teigiama, kad jis sukelia pašalines reakcijas ne daugiau kaip 1 procentui pacientų. Ištyrus 1000 vaistą vartojusių ligonių, nustatyta, kad pašalini poveikį pajuto 32 ligoniai. Ar duomenys neprieštarauja reklaminiam teiginiui? Reikšmingumo lygmuo 0.05.

Tikrinkime hipotezę

$$H : p = p_0, p_0 = 0.01$$

```
proportion.test <- function(p_0, s_n, n, alternative="two.sided") {
  # Hypothesis about proportion value.
  # Returns p-value.

  if (alternative == "two.sided" | alternative == "t"){
    pv <- 2 * min(
      pbeta(p_0, s_n, n - s_n + 1),
      1 - pbeta(p_0, s_n + 1, n - s_n)
    )

  } else if (alternative == "greater" | alternative == "g" ) {
    pv <- pbeta(p_0, s_n, n - s_n + 1)

  } else if (alternative == "less" | alternative == "l") {
    pv <- 1 - pbeta(p_0, s_n + 1, n - s_n)

  } else {
    stop("'alternative' should be one of 'two.sided', 'greater', 'less' or first
      letter of these terms!")
  }

  return(pv)
}

pv <- proportion.test(0.01, 32, 1000)
pv
```

```
## [1] 3.876311e-08
```

Čia formulės paiintos iš Pratybos1.pdf.  $p_v = 3.8763114 \times 10^{-8} < \alpha = 0.05$ . Hipotezę  $H$  atmetame. Statistiškai, galime teigti, kad duomenys prieštarauja reklaminiam teiginiui.

- (c) *Hipotezė apie proporcijos reikšmę* (ČM III.3.10). Ekonomistas nori patikrinti, ar padaugėjo smulkių įmonių (procentais). Prieš dešimt metų jos sudarė 20 procentų visų įmonių. Šiuo metu iš 100 atsitiktinai parinktų įmonių 27 buvo smulkios. Reikšmingumo lygmuo 0.05.

Tikrinkime hipotezę

$$H : p \geq p_0, p_0 = 0.2$$

```
pv <- proportion.test(0.2, 27, 100, alternative="l")
pv
```

```
## [1] 0.9658484
```

$p_v = 0.9658484 > \alpha 0.05$ . Hipotezės nėra pagrindo atmesti. Statistiškai, smulkių įmonių padaugėjo.

- (d) *Hipotezė apie koreliacijos koeficiento lygybę nuliui* (ČM III.3.13). Duomenys apie pardavėjo stažą (metais) ir jo pradinį atlyginimą (sutartiniais vienetais) pateikti lentelėje. Ar atlyginimas priklauso nuo pardavėjo stažo?

Stažas	Atlyginimas	Stažas	Atlyginimas
2	100	8	500
1.5	300	7	400
3	400	5	400
10	600	4	250
12	600	2	200
4	300	1	100
2	100	6	350

```
years <- c(2, 1.5, 3, 10, 12, 4, 2,
           8, 7, 5, 4, 2, 1, 6)
wages <- c(100, 300, 400, 600, 600, 300, 100,
           500, 400, 400, 250, 200, 100, 350)
```

Tikrinkime hipotezę

$$H : \rho = 0$$

```
cor <- cor.test(years, wages)
pv <- cor$p.value
cor

##
## Pearson's product-moment correlation
##
## data: years and wages
## t = 6.9426, df = 12, p-value = 1.555e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.6934363 0.9665141
## sample estimates:
## cor
## 0.8947987
```

$p_v = 1.5546876 \times 10^{-5} < \alpha = 0.05$ . Hipotezę atmetame. Statistiškai, koreliacija nėra lygi nuliui.

- (e) *Dviejų priklausomų normaliųjų imčių vidurkių palyginimas: Studento kriterijus*. Tiriamas fizinių pratimų poveikis svoriui. Parenkamos 5 moterys ir matuojamas jų svoris prieš ir po fizinių pratimų kurso. Gauti rezultatai: 84, 97, 77, 91, 85 (prieš) ir 78, 95, 73, 88, 80 (po). Rasti 0.9 pasikliautinį intervalą svorių skirtumui ir patikrinti hipotezę: vidutinis svoris nepakinta.

```
before <- c(84, 97, 77, 91, 85)
after <- c(78, 95, 73, 88, 80)

t <- t.test(before, after, paired=TRUE, conf.level=0.9)
pv <- t$p.value
t

##
## Paired t-test
##
```

```
## data: before and after
## t = 5.6569, df = 4, p-value = 0.004813
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
## 2.492557 5.507443
## sample estimates:
## mean of the differences
## 4
```

$$H : \mu_1 = \mu_2$$

$pv = 0.0048127 < \alpha = 0.10$ . Hipotezę atmetame. Statistiškai, vidutinis svoris pakito.

---

Padaryta su R version 3.4.2 (2017-09-28), x86\_64-pc-linux-gnu.