Statistinės duomenų analizės praktinės užduotys

2017

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

6. **Parametrinių hipotezių tikrinimas.**

(a) *Hipotezė apie normalaus skirstinio vidurkio reikšmę: Stujudento kriterijus vienai normaliajai imčiai* (ČM III.3.3.4). Užkandžiais prekiaujanti firma nusprenė mėsainius su žuvimi pakeisti mėsainiais 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.

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

mu_0 <- 520
```

Patikrinkime *normalumo prielaidą*. *Kolmogorovo - Smirnovo kriterijus*:

```r
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:

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

```r
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
```

$pv = 0.280576 > \alpha = 0.05$, 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 vartojusiu 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") {
  # Hipothesis 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 paiimtos iš `Pratybos1.pdf`. $pv = 3.8763114 \times 10^{-8} < \alpha = 0.05$. 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
```

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

(d) *Hipotezė apie koreliacijos koeciento lygybę nuliui* (ČM III.3.13). Duomenys apie pardavėjo stažą (metais) ir jo pradinų 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

```r
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

$$H : \rho = 0$$

```r
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
```

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

(e) *Dviejų priklausomų normaliųjų imčių vidurkių palyginimas: Stjudento kriterijus.* Tiriamas fizinių pratimų poveikis svoriui. Parenkamos 5 moterys ir matuojamas ju 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 pasikliautinų intervalą svorių skirtumui ir patikrinti hipotezę: vidutinis svoris nepakinta.

```r
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.

(f) *Dviejų priklausomų normaliųjų imčių vidurkių palyginimas: Stjudento kriterijus.* Krakmolo kiekis bulvėse nustatomas dviem būdais. Norint palyginti tuos būdus, buvo paimta 16 bulvių ir kiekvienoje iš jų krakmolo kiekis nustatytas dviem būdais. Gauti rezultatai (krakmolingumas procentais) surašyti lentelėje.

Eil. Nr.	I būdas	II būdas	Eil. Nr.	I būdas	II būdas
1	21.7	21.5	9	14.0	13.9
2	18.7	18.7	10	17.2	17.0
3	18.3	18.3	11	21.7	21.4
4	17.5	17.4	12	18.6	18.6
5	18.5	18.3	13	17.9	18.0
6	15.6	15.4	14	17.7	17.6
7	17.0	16.7	15	18.3	18.5
8	16.6	16.9	16	15.6	15.5

Laikydami, kad nustatomas krakmolingumo procentas turi normalųjį skirstinį patikrinkite prielaidą, kad abu metodai yra ekvivalentūs.

(g) *Dviejų nepriklausomų normaliųjų imčių vidurkių palyginimas. Stjudento kriterijus.* Matuojamas 16 detalių, pagamintų vieną dieną, atsparumas:

$$13.1, 12.8, 11.9, 12.4, 13.5, 13.5, 12.0, 13.8,$$

$$10.6, 12.4, 13.5, 11.7, 13.9, 11.5, 12.5, 11.9$$

Kitų 9 detalių, pagamintų kitą dieną, atsparumas:

$$13.7, 13.5, 14.2, 15.6, 14.8, 14.3, 15.4, 14.0, 15.1$$

Patikrinkite hipotezę, kad abi dienas buvo gaminamos vidutiniškai vienodo atsparumo detalės.

(h) *Dviejų nepriklausomų normaliųjų imčių vidurkių palyginimas: Stjudento kriterijus.* Lentelėje pateikti dviejų nepriklausomų eksperimentų su musėmis rezultatai. Pirmajame bandyme tam tikrais nuodais musės buvo veikiamos 30 sekundžių, antrajame - 60 sekundžių. Paralyžuojantį nuodų poveikį nusako vadinamasis reakcijos laikas, praėjęs nuo musės sąlyčio su nuodais iki to momento, kai musė jau nebegali stovėti. Reikia patikrinti hipotezę, kad vidutinis reakcijos laikas antrajame bandyme yra trumpesnis.

I bandymas

i	X_i	i	X_i
1	4.9	9	17.1

i	X_i	i	X_i
2	16.2	10	17.9
3	25.4	11	26.6
4	8.6	12	33.7
5	10.9	13	33.9
6	12.5	14	28.1
7	12.9	15	15.9
8	9.8	16	66.2

II bandymas

i	Y_i	i	Y_i
1	10.8	9	30.6
2	10.9	10	36.3
3	13.3	11	26.9
4	13.4	12	22.4
5	17.1	13	51.9
6	19.2	14	23.8
7	25.0	15	26.9
8	26.0		

Nurodymas. Iš histogramų pavidalo matome, kad X ir Y skirstiniai labai asimetriški. Todėl reikėtų atlikti stebimų dydžių transformacijas, kad naujo a.d. skirstiniai būtų patenkinamai aprašomi normaliuoju skirstiniu, po to remtis Stjudento kriterijumi. Nesunki įsitikinti, kad nagrinėjame pavyzdyje $\ln X$ ir $\ln Y$ tiksliau aprašomi normaliuoju skirstiniu, negu a.d. X ir Y. Kitaip sakant, stebimieji a. d. tiksliau aprašomi lognormaliuoju skirstiniu.

(i) *Dviejų nepriklausomų normaliųjų imčių dispersijos palyginimas: Fišerio kriterijus.* Naudojant 6.g pratimo duomenis patikrinti hipotezę, kad abi dienas pagamintų detalių atsparumai turi vienodas dispersijas.

(j) *Proporcijų palyginimas.* Dviejose šalyse buvo pateiktas klausimas "ar bijote vaikščioti gatvėje naktį?". Pirmoje šalyje iš $n_1 = 300$ apklaustųjų buvo gauta $S_1 = 120$ teigiamų atsakymų. Antroje šalyje gauta $S_2 = 148$ teigiamų atsakymų tarp $n_2 = 200$ apklaustųjų. Patikrinti hipotezę, kad abiejose šalyse žmonės vienodai baiminasi vaikščioti gatvėje naktį.

(k) *Proporcijų palyginimas.* (ČM III.4.9) Nepriklausomas ekspertas tiria, kiek kartų garantinio TV taisymo prireikė televizoriams, surinktiems Pietryčių Azijoje, ir kiek - Rytų Europoje. Iš 150 azijinių televizorių garantinio remonto prireikė 4, iš 100 europinių - 2. Ar galima teigti, kad europinams televizoriams garantinio remonto reikia rečiau?

Padaryta su R version 3.4.2 (2017-09-28), x86_64-pc-linux-gnu.