

Biometrika 2017.04.19

Rimantas Eidukevicius

April 18, 2017

Porinė imtis, t-testas

Duomenų lentelėje du kintamieji: $V1$ ir $V2$, pvz., pinigai kairiojoje ir dešiniojoje to pačio švarko kišenėse.

H_0 : vidutiniškai pinigų abiejose kišenėse vienodai, H_1 : ... skirtingai.

```
load("~/MIF/Destymas/2017Pavasaris/Biometrika/PriklImtys.RData")
print(PriklImtys)
```

```
##   V1 V2
## 1 11 12
## 2 15 15
## 3 16 15
## 4 18 19
## 5 10 11
## 6 13 12
## 7 19 22
```

```
with(PriklImtys, (t.test(V1, V2, alternative='two.sided',
conf.level=.9, paired=TRUE)))
```

```
##
## Paired t-test
##
## data:  V1 and V2
## t = -1.082, df = 6, p-value = 0.3208
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  -1.5976625  0.4548053
## sample estimates:
## mean of the differences
##                -0.5714286
```

Sukeiskime kintamuosius vietomis:

```
with(PriklImtys, (t.test(V2, V1, alternative='two.sided',
conf.level=.9, paired=TRUE)))
```

```
##
## Paired t-test
##
## data:  V2 and V1
## t = 1.082, df = 6, p-value = 0.3208
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  -0.4548053  1.5976625
## sample estimates:
## mean of the differences
##                0.5714286
```

Palyginti pasiklaidinuosius intervalus, pokyčio vidurkius.

Dvi nepriklausomos imtys, dispersijų ir vidurkių palyginimas

Duomenys:

```
bw <- read.table("C:/Users/rimas/OneDrive/Documents/MIF/Destymas/2017Pavasaris/Biometrika/birthwt.txt",
  header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
head(bw)
```

```
##   low age lwt race smoke ptl ht ui ftv  bwt
## 1   0  19 182   2    0  0  0  1  0 2523
## 2   0  33 155   3    0  0  0  0  3 2551
## 3   0  20 105   1    1  0  0  0  1 2557
## 4   0  21 108   1    1  0  0  1  2 2594
## 5   0  18 107   1    1  0  0  1  0 2600
## 6   0  21 124   3    0  0  0  0  0 2622
```

```
bw <- within(bw, {smokef <- factor(smoke, labels=c('NoSmoke', 'Smoke'))})
bw <- within(bw, {racef <- as.factor(race)})
head(bw)
```

```
##   low age lwt race smoke ptl ht ui ftv  bwt  smokef racef
## 1   0  19 182   2    0  0  0  1  0 2523 NoSmoke    2
## 2   0  33 155   3    0  0  0  0  3 2551 NoSmoke    3
## 3   0  20 105   1    1  0  0  0  1 2557  Smoke    1
## 4   0  21 108   1    1  0  0  1  2 2594  Smoke    1
## 5   0  18 107   1    1  0  0  1  0 2600  Smoke    1
## 6   0  21 124   3    0  0  0  0  0 2622 NoSmoke    3
```

```
summary(bw)
```

```
##           low           age           lwt           race
##  Min.   :0.0000   Min.   :14.00   Min.   : 80.0   Min.   :1.000
## 1st Qu.:0.0000   1st Qu.:19.00   1st Qu.:110.0 1st Qu.:1.000
##  Median :0.0000   Median :23.00   Median :121.0 Median :1.000
##  Mean   :0.3122   Mean   :23.24   Mean   :129.8 Mean   :1.847
## 3rd Qu.:1.0000   3rd Qu.:26.00   3rd Qu.:140.0 3rd Qu.:3.000
##  Max.   :1.0000   Max.   :45.00   Max.   :250.0 Max.   :3.000
##           smoke           ptl           ht           ui
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000 Min.   :0.0000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000 1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.00000 Median :0.0000
##  Mean   :0.3915   Mean   :0.1958   Mean   :0.06349 Mean   :0.1481
## 3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.00000 3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :3.0000   Max.   :1.00000 Max.   :1.0000
##           ftv           bwt           smokef           racef
##  Min.   :0.0000   Min.   : 709   NoSmoke:115   1:96
## 1st Qu.:0.0000   1st Qu.:2414   Smoke  : 74   2:26
##  Median :0.0000   Median :2977                   3:67
##  Mean   :0.7937   Mean   :2945
## 3rd Qu.:1.0000   3rd Qu.:3475
##  Max.   :6.0000   Max.   :4990
```

Palyginkime rūkančių ir nerūkančių mamų amžiaus dispersijas (alternatyva dvipusė):

```
with(bw, tapply(age, smokef, var, na.rm=TRUE))
```

```
## NoSmoke    Smoke  
## 29.89580 25.47649
```

```
var.test(age ~ smokef, alternative='two.sided', conf.level=.95, data=bw)
```

```
##  
## F test to compare two variances  
##  
## data: age by smokef  
## F = 1.1735, num df = 114, denom df = 73, p-value = 0.464  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.7649058 1.7656682  
## sample estimates:  
## ratio of variances  
## 1.173466
```

Nulinės hipotezės apie dispersijų lygybę neatmetame (kodėl neminimas reikšmingumo lygmuo?), todėl lyginsime amžiaus vidurkius (alternatyva dvipusė) su prielaida apie dispersijų lygybę. Kartais patariama šia prielaida apamai nesinaudoti.

```
t.test(age~smokef, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=bw)
```

```
##  
## Two Sample t-test  
##  
## data: age by smokef  
## t = 0.60702, df = 187, p-value = 0.5446  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.080245 2.040528  
## sample estimates:  
## mean in group NoSmoke    mean in group Smoke  
## 23.42609 22.94595
```

Dispersinė analizė (vieno faktoriaus): mamų amžius grupėse pagal rasę

Hipotezės apie dispersijų lygybę tikrinimas:

```
library(car)  
with(bw, tapply(age, racef, var, na.rm=TRUE))
```

```
##      1      2      3  
## 31.97719 26.09846 20.57440
```

```
leveneTest(age ~ racef, data=bw, center="mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")  
##      Df F value Pr(>F)  
## group  2  1.7769 0.172  
##      186
```

Prielaida apie dispersijų lygybę neatmetama.

```
AnovaModel.1 <- aov(age ~ smokef, data=bw)
summary(AnovaModel.1)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## smokef        1    10    10.38   0.368  0.545
## Residuals    187   5268    28.17
```

```
library(RcmdrMisc)
```

```
## Loading required package: sandwich
```

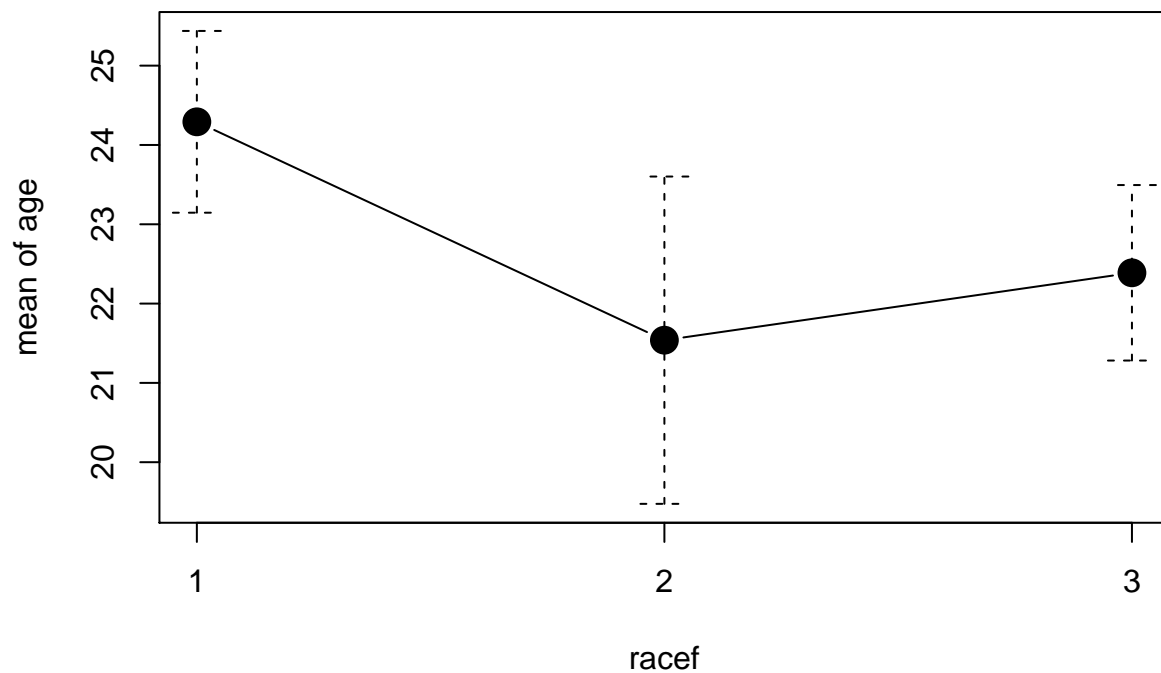
```
with(bw, numSummary(age, groups=smokef, statistics=c("mean", "sd")))
```

```
##              mean      sd data:n
## NoSmoke 23.42609 5.467706   115
## Smoke   22.94595 5.047424    74
```

Paiškinantis grafikas:

```
with(bw, plotMeans(age, racef, error.bars="conf.int", level=0.95))
```

Plot of Means



Dvifaktorisė dispersinė analizė

```
AnovaMode2 <- lm(age ~ racef*smokef, data=bw)
Anova(AnovaMode2)
```

```
## Anova Table (Type II tests)
```

```
##
## Response: age
##           Sum Sq Df F value    Pr(>F)
## racef      280.5  2  5.4628 0.004962 **
## smokef      60.8  1  2.3676 0.125602
## racef:smokef 289.4  2  5.6372 0.004210 **
## Residuals   4698.0 183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

with(bw, (tapply(age, list(racef, smokef), sd, na.rm=TRUE)))

##      NoSmoke      Smoke
## 1 6.017373 4.925807
## 2 3.889623 5.952590
## 3 4.452896 5.107926

xtabs(~ racef + smokef, data=bw)

##      smokef
## racef NoSmoke Smoke
##      1      44     52
##      2      16     10
##      3      55     12
```

Pakartoti skaičiavimus su kūdikių svoriu, rašyti trumpas ataskaitas.