Tiesine regresija

RE

16 May 2017

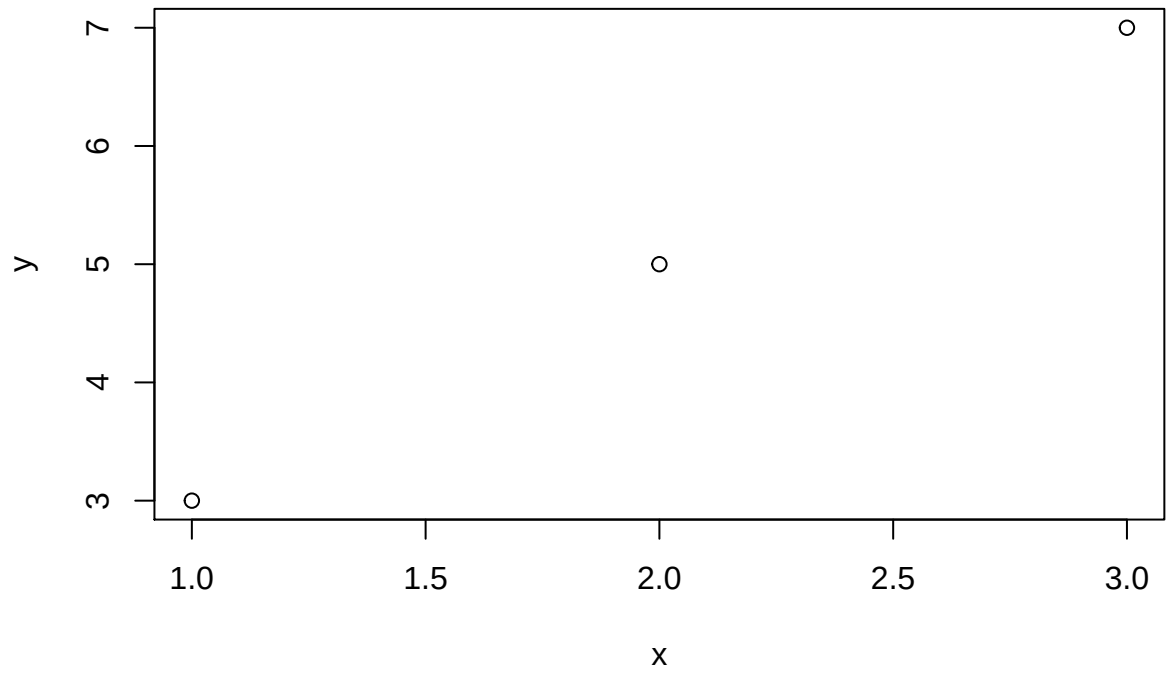
Tiesinė regresija: pavyzdys

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, \ldots, n.$$

"Betos" - nežinomi parametrai:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2.$$

```
y <- c(3, 5, 7)
x <- c(1, 2, 3)
plot(x,y)
```



Nuostoliai:

```
least.squares <- function(p, x, y) {
  sum((y - (p[1] + p[2] * x))^2)}
```

"Blogas" pasirinkimas

```
start.searching.here <- c(intercept = 0, slope = 0)
least.squares(start.searching.here, x, y)
```

```
## [1] 83
```

Optimizavimas:

```
optm <- optim(par = start.searching.here, fn = least.squares,
      x = x, y = y)
optm$par
```

```
## intercept     slope
## 0.9991065 2.0003141
```

```
summary(optm)
```

```
##             Length Class  Mode
## par         2      -none- numeric
## value       1      -none- numeric
## counts      2      -none- numeric
## convergence 1      -none- numeric
## message     0      -none- NULL
```

Kitaip rasime parametrus (tikimybinė interpretacija):

```
mean(y) - cov(x,y) / var(x) * mean(x)   # Beta 0
```

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

```
cov(x,y) / var(x)                       # Beta 1
```

```
## [1] 2
```

Svarbūs požimių x ir y matavimo vienetai (metrai, coliai, tonos ir t.t.). Tarkime, kad $x_{i+1} = x_i + 1$. Tegu triukšmas ϵ nedalyvauja, t.y., skaičiuojame po vidurkinimo ($E(\epsilon_i) = 0$). Rasime skirtumą :

$$y_{i+1} - yi = (\beta_0 + \beta_1(x_i + 1) - (\beta_0 + \beta_1 x_i) = \beta_1.$$

Turime parametro β_1 interpretaciją.

Rasti jo interpretaciją po transformacijos: $z_i = x_i - \bar{x},\ y_i = \beta_0 + \beta_1 z_i + \epsilon_i,\ i = 1, \ldots, n$

Pavyzdys su duomenimis apie kūdikių svorį:

```
load("C:/Users/rimas/OneDrive/Documents/MIF/Destymas/2017Pavasaris/Biometrika/bw.RData")
RegModel.1 <- lm(bwt~lwt, data=bw)
summary(RegModel.1)
```

```
##
## Call:
## lm(formula = bwt ~ lwt, data = bw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2192.18  -503.63    -3.91   508.25  2075.53
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2369.672    228.431  10.374   <2e-16 ***
```

```
## lwt            4.429      1.713   2.586   0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 718.2 on 187 degrees of freedom
## Multiple R-squared:  0.03452,    Adjusted R-squared:  0.02935
## F-statistic: 6.686 on 1 and 187 DF,  p-value: 0.01048
```