Uporabna statistika

Gregor Dolinar

Fakulteta za elektrotehniko
Univerza v Ljubljani

11. december 2013

Parni t-test (test dvojic)

$$D_j = X_{1j} - X_{2j}$$

$$T_0 = \frac{\overline{D} - \triangle_0}{S_D/\sqrt{n}}$$

$$T_0 = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n} + \frac{1}{n}}}$$

$2n - 2$ prostostnih stopenj

$$T_0 = \frac{\overline{D} - \triangle_0}{S_D / \sqrt{n}}$$

$n - 1$ prostostnih stopenj

Primer

91.50, 89.19

94.18, 90.95

92.18, 90.46

95.39, 93.21

91.79, 97.19

89.07, 97.04

94.72, 91.07

89.21, 92.75

$\overline{x_1} = 92.255$, $\overline{x_2} = 92.733$, $s_1 = 2.39$, $s_2 = 2.98$

Razmerje 2 varianc

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

F- porazdelitev

$$F = \frac{W/u}{Y/v}$$

Razmerje 2 deležev populacije

Standardizirana normalna porazdelitev.

Linearna regresija

Primer

Poskus, pri katerem dobimo dve količini x in y pri vsaki ponovitvi.
Na primer, pri kemični reakciji merimo čistost nastalega kisika y, ki ga dobimo pri dani količini ogljikovodikov x.
Radi bi raziskali, v kakšni odvisnosti sta x in y.
Če poznamo, kako sta količini x in y povezani, lahko pri danem x napovemo vrednost y.

V primeru linearne regresije imamo eno samo neodvisno (pojasnjevalno) spremenljivko x in odvisno spremenljivko Y.
Denimo, da sta x in Y linearno odvisni (graf je premica) in da je Y slučajna spremenljivka pri vsaki vrednosti x.
Torej je pričakovana vrednost slučajne spremenljivke Y pri vsaki vrednosti x enaka

$$E(Y|x) = \beta_0 + \beta_1 x,$$

kjer sta β_0 in β_1 regresijska koeficienta.

Če pišemo

$$Y = \beta_0 + \beta_1 x + \epsilon,$$

potem je ϵ slučajna spremenljivka (napaka) s povprečno vrednostjo 0 in neznano varianco σ^2.

Metoda najmanjših kvadratov (Gauss)

Imamo n parov $(x_1, y_1), \ldots, (x_n, y_n)$. Iščemo β_0 in β_1, da bo premica $\beta_0 + \beta_1 x$ najbolje aproksimirala podatke.
Iščemo taka β_0 in β_1, da bo vsota kvadratov vertikalnih odstopanj od premice najmanjša.

Če je

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

potem iščemo taka β_0 in β_1, da je

$$L(\beta_0, \beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

najmanj.
Iščemo minimum funkcije dveh spremenljivk $L(\beta_0, \beta_1)$.

Rešitvi enačb:

$$\frac{\partial L}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

in

$$\frac{\partial L}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

označimo z $\hat{\beta}_0$ in $\hat{\beta}_1$.

Definiramo

$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

in

$$S_{xy} = \sum_{i=1}^{n} y_i (x_i - \overline{x})^2.$$

Potem je

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

in

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

Torej

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

in

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \quad i = 1, \ldots, n,$$

kjer je $e_i = y_i - \hat{y}_i$ residual (odklon).
Označimo še s

$$SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

vsoto kvadratov napak.

Velja

$$E(SS_E) = (n-2)\sigma^2,$$

zato je nepristranska cenilka za σ^2 enaka

$$\hat{\sigma}^2 = \frac{SS_E}{n-2}.$$

Velja še

$$V(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}.$$