Uporabna statistika

Gregor Dolinar

Fakulteta za elektrotehniko
Univerza v Ljubljani

10. december 2013

Statistična analiza dveh vzorcev

Primer

- Pri proizvajanju izdelka bi radi nek sestavni del nadomestili s cenejšim, hkrati pa nočemo, da bi se življenjska doba izdelka spremenila.
- Pri proizvajanju izdelka bi radi nek sestavni del nadomestili z boljšim, a dražjim, zato želimo, da bi se življenjska doba izdelka podaljšala vsaj za pol leta.

Izberemo vzorec n_1 izdelkov s starim sestavnim delom in n_2 izdelkov z novim sestavnim delom. Radi bi videli, če je razlika med življenjskima dobama vzorcev statistično značilna.

Imamo

- $X_{11}, X_{12}, \ldots, X_{1n_1}$ slučajni vzorec 1. populacije
- $X_{21}, X_{22}, \ldots, X_{2n_1}$ slučajni vzorec 2. populacije
- Populaciji X_1 in X_2 sta normalno porazdeljeni in neodvisni

Cenilka razlike $\mu_1 - \mu_2$ je $\overline{X_1} - \overline{X_2}$.

$$E(\overline{X_1} - \overline{X_2}) = E(\overline{X_1}) - E(\overline{X_2}) = \mu_1 - \mu_2$$

$$V(\overline{X_1} - \overline{X_2}) = V(\overline{X_1}) + V(\overline{X_2}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$Z = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Z standardizirana normalna.

Preverjanje domneve

H_0: $\mu_1 - \mu_2 = \triangle_0$
Testna statistika

$$Z = \frac{\overline{X_1} - \overline{X_2} - \triangle_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Alternativne hipoteze

- H_1: $\mu_1 - \mu_2 \neq \triangle_0$, zavrnemo če $z_0 > z_{\alpha/2}$ ali $z_0 < -z_{\alpha/2}$
- H_1: $\mu_1 - \mu_2 > \triangle_0$, zavrnemo če $z_0 > z_\alpha$
- H_1: $\mu_1 - \mu_2 < \triangle_0$, zavrnemo če $z_0 < -z_\alpha$

Interval zaupanja

$$P\left[-z_{\alpha/2} \leq \frac{\overline{X_1} - \overline{X_2} - \triangle_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{\alpha/2}\right] = 1 - \alpha$$

$$\overline{x_1} - \overline{x_2} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2$$

$$\mu_1 - \mu_2 \leq \overline{x_1} - \overline{x_2} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Ločimo dva primera:

- $\sigma_1 = \sigma_2 = \sigma$
- $\sigma_1 \neq \sigma_2$

$\sigma_1 = \sigma_2 = \sigma$

Prvi vzorec ima n_1 elementov, drugi vzorec n_2 elementov.

$$V(\overline{X_1} - \overline{X_2}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$$

Cenilka za σ^2

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$T = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

T je t porazdeljena z $n_1 + n_2$ - 2 prostostnimi stopnjami.

Preverjanje domneve

H_0: $\mu_1 - \mu_2 = \triangle_0$
Testna statistika

$$T = \frac{\overline{X_1} - \overline{X_2} - \triangle_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Alternativne hipoteze

- H_1: $\mu_1 - \mu_2 \neq \triangle_0$, zavrnemo če $t_0 > t_{\alpha/2, n_1+n_2-2}$ ali $t_0 < -t_{\alpha/2, n_1+n_2-2}$
- H_1: $\mu_1 - \mu_2 > \triangle_0$, zavrnemo če $t_0 > t_{\alpha, n_1+n_2-2}$

Interval zaupanja

$$\overline{x_1} - \overline{x_2} - t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2$$

$$\mu_1 - \mu_2 \leq \overline{x_1} - \overline{x_2} + t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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

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

$\sigma_1 \neq \sigma_2$

Prvi vzorec ima n_1 elementov, drugi vzorec n_2 elementov.

$$T = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

T je t porazdeljena z ν prostostnimi stopnjami, pri čemer je

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}.$$