

PRONTUARIO

Stimatori — Standard error

Distribuzioni campionarie ed errore standard degli estimatori · liceo e università

NOTAZIONE E STIMATORI

Dato un campione casuale $x_1, \dots, x_n$:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2, \quad \hat{p} = \frac{X}{n} \quad (X = \text{n. successi}).$$

Per due campioni, con $d_i = x_i - y_i$:

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y}), \quad s_D^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2.$$

σ, σ^2 : deviazione standard / varianza della popolazione (note) · s, s^2 : stime campionarie (varianza non nota). Lo **standard error** $SE(\hat{\theta})$ è la deviazione standard della distribuzione campionaria dello stimatore $\hat{\theta}$.

Singola popolazione

DISTRIBUZIONI CAMPIONARIE

Media campionaria

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

se n sufficientemente elevato (TLC)

Proporzione campionaria

$$\hat{P} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

se n sufficientemente elevato (TLC)

STANDARD ERROR · SINGOLA POPOLAZIONE

1 Media campionaria — varianza **nota**

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

2 Media campionaria — varianza **NON nota**

$$SE(\bar{X}) = \frac{s}{\sqrt{n}}$$

3 Proporzione campionaria

$$SE(\hat{P}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

RIFERIMENTO DISTRIBUTIVO · QUANDO USARE LA T DI STUDENT

Con varianza **nota** (o **n** grande) la statistica standardizzata è $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$. Con varianza **NON nota** (stimata con **s**) e **n** piccolo, il riferimento corretto è la t di Student:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1} \quad (\text{gradi di libertà } n - 1), \quad t_{n-1} \xrightarrow[n \rightarrow \infty]{} N(0, 1).$$

Due popolazioni

Differenze fra medie campionarie

$$E[\bar{X} - \bar{Y}] = E[\bar{X}] - E[\bar{Y}] = \mu_x - \mu_y$$

$$\text{VAR}[\bar{X} - \bar{Y}] = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y} - 2 \text{COV}(\bar{X}, \bar{Y})$$

Campioni indipendenti $\Rightarrow \text{COV}(\bar{X}, \bar{Y}) = 0$.

Differenze fra proporzioni campionarie

$$E[\hat{P}_x - \hat{P}_y] = p_x - p_y$$

$$\text{VAR}[\hat{P}_x - \hat{P}_y] = \frac{p_x(1 - p_x)}{n_x} + \frac{p_y(1 - p_y)}{n_y}$$

Campioni indipendenti.

STANDARD ERROR · DIFFERENZE FRA MEDIE

1 Indipendenti, varianze note

$$\text{SE}(\bar{X} - \bar{Y}) = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

2 Indipendenti, varianze non note (Welch)

$$\text{SE}(\bar{X} - \bar{Y}) = \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

3 Indipendenti, varianze non note ma **uguali** (pooled)

$$\text{SE}(\bar{X} - \bar{Y}) = s_{\text{pool}} \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

4 Dipendenti (appaiati), varianze note

$$\text{SE}(\bar{X} - \bar{Y}) = \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{n} - 2 \frac{\sigma_{xy}}{n}}$$

5 Dipendenti (appaiati), varianze non note

$$\text{SE}(\bar{X} - \bar{Y}) = \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{n} - 2 \frac{s_{xy}}{n}} = \sqrt{\frac{s_D^2}{n}}$$

VARIANZA COMBINATA (POOLED) E GRADI DI LIBERTÀ

$$s_{\text{pool}}^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}, \quad T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\text{SE}(\bar{X} - \bar{Y})} \sim t_{n_x + n_y - 2}.$$

Nei campioni appaiati $n_x = n_y = n$: conviene lavorare sulle differenze $d_i = x_i - y_i$, con

$$T = \frac{\bar{d} - \delta_0}{s_D / \sqrt{n}} \sim t_{n-1}.$$

STANDARD ERROR · DIFFERENZE FRA PROPORZIONI

6 Indipendenti — per intervalli di confidenza (non pooled)

$$\text{SE}(\hat{P}_x - \hat{P}_y) = \sqrt{\frac{\hat{p}_x(1 - \hat{p}_x)}{n_x} + \frac{\hat{p}_y(1 - \hat{p}_y)}{n_y}}$$

TEST SU $p_x = p_y$ · PROPORZIONE COMBINATA

Nella **verifica d'ipotesi** $H_0 : p_x = p_y$ si stima l'unica proporzione comune con la *proporzione combinata (pooled)*:

$$\hat{p} = \frac{X_x + X_y}{n_x + n_y} = \frac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y}, \quad \text{SE}_0(\hat{P}_x - \hat{P}_y) = \sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_x} + \frac{1}{n_y} \right)}.$$

La versione *non pooled* (caso 6) resta quella usata per l'intervallo di confidenza.

Statistica standardizzata — uso dello standard error

Test e intervallo di confidenza

$$Z \text{ oppure } T = \frac{\hat{\theta} - \theta_0}{\text{SE}(\hat{\theta})}, \quad \hat{\theta} \pm z_{1-\alpha/2} \text{SE}(\hat{\theta}) \quad (\text{o } t_{\nu, 1-\alpha/2})$$

con $\hat{\theta}$ pari a $\bar{X}$, $\hat{P}$, $\bar{X} - \bar{Y}$ o $\hat{P}_x - \hat{P}_y$ a seconda del caso.

Richiami — covarianza e trasformazioni lineari

Covarianza e correlazione

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$\text{cov}(aX, bY) = ab \text{ cov}(X, Y)$$

$$\text{cov}(X + a, Y + b) = \text{cov}(X, Y)$$

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X) \sigma(Y)}$$

Trasformazione lineare $Z = aX + bY + c$

$$E(Z) = a E(X) + b E(Y) + c$$

$$\text{var}(Z) = a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab \text{cov}(X, Y)$$

Il valore atteso è un operatore lineare; la varianza no (compare il termine di covarianza).

Tabella riassuntiva

Parametro, stimatore e standard error per ogni caso.

Caso	Parametro θ	Stimatore $\hat{\theta}$	Standard error $SE(\hat{\theta})$
Media — varianza nota	μ	$\bar{X}$	$\sigma/\sqrt{n}$
Media — varianza non nota	μ	$\bar{X}$	$s/\sqrt{n}$
Proporzione	p	$\hat{P}$	$\sqrt{\hat{p}(1-\hat{p})/n}$
Diff. medie indep. — var. note	$\mu_x - \mu_y$	$\bar{X} - \bar{Y}$	$\sqrt{\sigma_x^2/n_x + \sigma_y^2/n_y}$
Diff. medie indep. — var. non note	$\mu_x - \mu_y$	$\bar{X} - \bar{Y}$	$\sqrt{s_x^2/n_x + s_y^2/n_y}$
Diff. medie indep. — var. uguali	$\mu_x - \mu_y$	$\bar{X} - \bar{Y}$	$s_{\text{pool}} \sqrt{1/n_x + 1/n_y}$
Diff. medie appaiate	$\mu_x - \mu_y$	$\bar{d}$	$\sqrt{s_D^2/n}$
Diff. proporzioni — IC	$p_x - p_y$	$\hat{P}_x - \hat{P}_y$	$\sqrt{\hat{p}_x(1-\hat{p}_x)/n_x + \hat{p}_y(1-\hat{p}_y)/n_y}$
Diff. proporzioni — test H_0	$p_x - p_y$	$\hat{P}_x - \hat{P}_y$	$\sqrt{\hat{p}(1-\hat{p})(1/n_x + 1/n_y)}$

COME LEGGERE LA TABELLA

Lo standard error è sempre la radice della varianza dello stimatore. Con varianza di popolazione nota si usa σ e la Normale; con varianza stimata si usa s e la t di Student (gradi di libertà indicati nei casi sopra). Per la differenza di proporzioni: versione *non pooled* per l'intervallo di confidenza, *pooled* per il test $H_0 : p_x = p_y$.