

Notazioni

$$\mathbb{1}_A(x) := \begin{cases} 0 & \text{se } x \notin A \\ 1 & \text{se } x \in A \end{cases} \quad \text{funzione indicatrice dell'insieme } A$$

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} \quad \text{con } k \in \{0, \dots, n\} \quad \text{coefficiente binomiale di } n \text{ su } k$$

Densità

Densità discrete

Sono indicati solo i valori di k per cui $p_X(k) \neq 0$.

Bernoulliana di parametro $p \in (0, 1)$

Simbolo	Densità	Media	Varianza
$X \sim B(1, p)$	$p_X(k) = \begin{cases} 1-p & \text{se } k=0 \\ p & \text{se } k=1 \end{cases}$	$\mathbb{E}(X) = p$	$\text{Var}(X) = p(1-p)$

Binomiale di parametri $n \in \mathbb{N}$, $p \in (0, 1)$

Simbolo	Densità	Media	Varianza
$X \sim B(n, p)$	$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ con $k \in \{0, \dots, n\}$	$\mathbb{E}(X) = np$	$\text{Var}(X) = np(1-p)$

Poissoniana di parametro $\lambda > 0$

Simbolo	Densità	Media	Varianza
$X \sim \mathcal{P}(\lambda)$	$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ con $k \in \{0, 1, 2, \dots\}$	$\mathbb{E}(X) = \lambda$	$\text{Var}(X) = \lambda$

Densità assolutamente continue

Uniforme continua in (a, b) , con $a < b$

Simbolo	Densità	Media	Varianza
$X \sim \mathcal{U}(a, b)$	$f_X(z) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(z)$	$\mathbb{E}(X) = \frac{a+b}{2}$	$\text{Var}(X) = \frac{(b-a)^2}{12}$

Esponenziale di parametro $\lambda > 0$

Simbolo	Densità	Media	Varianza
$X \sim \mathcal{E}(\lambda)$	$f_X(z) = \lambda e^{-\lambda z} \mathbb{1}_{[0,+\infty)}(z)$	$\mathbb{E}(X) = \frac{1}{\lambda}$	$\text{Var}(X) = \frac{1}{\lambda^2}$

Normale di parametri μ , σ^2

Simbolo	Densità	Media	Varianza
$X \sim N(\mu, \sigma^2)$	$f_X(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(z-\mu)^2}{2\sigma^2} \right]$	$\mathbb{E}(X) = \mu$	$\text{Var}(X) = \sigma^2$

Intervalli di confidenza di livello γ

Media

Con varianza σ^2 nota, campione normale o numeroso

$$\left(\bar{X}_n - z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \right) \quad \left(\bar{X}_n - z_{\gamma} \frac{\sigma}{\sqrt{n}}, +\infty \right) \quad \left(-\infty, \bar{X}_n + z_{\gamma} \frac{\sigma}{\sqrt{n}} \right)$$

Con varianza incognita, campione normale

$$\left(\bar{X}_n - t_{\frac{1+\gamma}{2}}(n-1) \frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{\frac{1+\gamma}{2}}(n-1) \frac{S_n}{\sqrt{n}} \right) \\ \left(\bar{X}_n - t_{\gamma}(n-1) \frac{S_n}{\sqrt{n}}, +\infty \right) \quad \left(-\infty, \bar{X}_n + t_{\gamma}(n-1) \frac{S_n}{\sqrt{n}} \right)$$

Con varianza incognita, campione numeroso

$$\left(\bar{X}_n - z_{\frac{1+\gamma}{2}} \frac{S_n}{\sqrt{n}}, \bar{X}_n + z_{\frac{1+\gamma}{2}} \frac{S_n}{\sqrt{n}} \right) \quad \left(\bar{X}_n - z_{\gamma} \frac{S_n}{\sqrt{n}}, +\infty \right) \quad \left(-\infty, \bar{X}_n + z_{\gamma} \frac{S_n}{\sqrt{n}} \right)$$

Campione di Bernoulli numeroso

$$\left(\bar{X}_n - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \right) \\ \left(\bar{X}_n - z_{\gamma} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, 1 \right) \quad \left(0, \bar{X}_n + z_{\gamma} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \right)$$

Varianza

Con media incognita, campione normale

$$\left(\frac{(n-1)S_n^2}{\chi_{\frac{1+\gamma}{2}}^2(n-1)}, \frac{(n-1)S_n^2}{\chi_{\frac{1-\gamma}{2}}^2(n-1)} \right) \quad \left(\frac{(n-1)S_n^2}{\chi_{\gamma}^2(n-1)}, +\infty \right) \quad \left(0, \frac{(n-1)S_n^2}{\chi_{1-\gamma}^2(n-1)} \right)$$

Test di ipotesi di livello di significatività α

Media

Con varianza σ_0^2 nota, campione normale o numeroso:

H_0	H_1	rifiuto H_0 se
$\mu = \mu_0$ $\mu \leq \mu_0$	$\mu > \mu_0$	$Z_0 > z_{1-\alpha}$
$\mu = \mu_0$ $\mu \geq \mu_0$	$\mu < \mu_0$	$Z_0 < -z_{1-\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$ Z_0 > z_{1-\frac{\alpha}{2}}$

$$Z_0 := \frac{\bar{X}_n - \mu_0}{\sigma_0} \sqrt{n}$$

Con varianza incognita, campione normale:

H_0	H_1	rifiuto H_0 se
$\mu = \mu_0$ $\mu \leq \mu_0$	$\mu > \mu_0$	$T_0 > t_{1-\alpha}(n-1)$
$\mu = \mu_0$ $\mu \geq \mu_0$	$\mu < \mu_0$	$T_0 < -t_{1-\alpha}(n-1)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$ T_0 \geq t_{1-\frac{\alpha}{2}}(n-1)$

$$T_0 := \frac{\bar{X}_n - \mu_0}{S_n} \sqrt{n}$$

Con varianza incognita, campione numeroso:

H_0	H_1	rifiuto H_0 se
$\mu = \mu_0$ $\mu \leq \mu_0$	$\mu > \mu_0$	$Z_0 > z_{1-\alpha}$
$\mu = \mu_0$ $\mu \geq \mu_0$	$\mu < \mu_0$	$Z_0 < -z_{1-\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$ Z_0 > z_{1-\frac{\alpha}{2}}$

$$Z_0 := \frac{\bar{X}_n - \mu_0}{S_n} \sqrt{n}$$

Campione di Bernoulli numeroso:

H_0	H_1	rifiuto H_0 se
$p = p_0$ $p \leq p_0$	$p > p_0$	$Z_0 > z_{1-\alpha}$
$p = p_0$ $p \geq p_0$	$p < p_0$	$Z_0 < -z_{1-\alpha}$
$p = p_0$	$p \neq p_0$	$ Z_0 > z_{1-\frac{\alpha}{2}}$

$$Z_0 := \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$$

Varianza

Con media incognita, campione normale:

H_0	H_1	rifiuto H_0 se
$\sigma^2 = \sigma_0^2$ $\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$X_0^2 > \chi_{1-\alpha}^2(n-1)$
$\sigma^2 = \sigma_0^2$ $\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$X_0^2 < \chi_{\alpha}^2(n-1)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$X_0^2 < \chi_{\frac{\alpha}{2}}^2(n-1)$ oppure $X_0^2 > \chi_{1-\frac{\alpha}{2}}^2(n-1)$

$$X_0^2 := \frac{(n-1)S_n^2}{\sigma_0^2}$$

Differenza tra medie $\mu_X - \mu_Y$

Con varianze σ_X^2 e σ_Y^2 note, campioni indipendenti normali:

H_0	H_1	rifiuto H_0 se
$\mu_X - \mu_Y = \delta_0$ $\mu_X - \mu_Y \leq \delta_0$	$\mu_X - \mu_Y > \delta_0$	$Z_0 > z_{1-\alpha}$
$\mu_X - \mu_Y = \delta_0$ $\mu_X - \mu_Y \geq \delta_0$	$\mu_X - \mu_Y < \delta_0$	$Z_0 < -z_{1-\alpha}$
$\mu_X - \mu_Y = \delta_0$	$\mu_X - \mu_Y \neq \delta_0$	$ Z_0 > z_{1-\frac{\alpha}{2}}$

$$Z_0 := \frac{\bar{X}_m - \bar{Y}_n - \delta_0}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}}$$

Con varianze incognite ma uguali, campioni indipendenti normali:

H_0	H_1	rifiuto H_0 se
$\mu_X - \mu_Y = \delta_0$ $\mu_X - \mu_Y \leq \delta_0$	$\mu_X - \mu_Y > \delta_0$	$T_0 > t_{1-\alpha}(m+n-2)$
$\mu_X - \mu_Y = \delta_0$ $\mu_X - \mu_Y \geq \delta_0$	$\mu_X - \mu_Y < \delta_0$	$T_0 < -t_{1-\alpha}(m+n-2)$
$\mu_X - \mu_Y = \delta_0$	$\mu_X - \mu_Y \neq \delta_0$	$ T_0 > t_{1-\frac{\alpha}{2}}(m+n-2)$

$$T_0 := \frac{\bar{X}_m - \bar{Y}_n - \delta_0}{\sqrt{S_P^2 \left(\frac{1}{m} + \frac{1}{n} \right)}}$$

$$S_P^2 := \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

Con varianze incognite, campioni indipendenti numerosi:

H_0	H_1	rifiuto H_0 se
$\mu_X - \mu_Y = \delta_0$ $\mu_X - \mu_Y \leq \delta_0$	$\mu_X - \mu_Y > \delta_0$	$Z_0 > z_{1-\alpha}$
$\mu_X - \mu_Y = \delta_0$ $\mu_X - \mu_Y \geq \delta_0$	$\mu_X - \mu_Y < \delta_0$	$Z_0 < -z_{1-\alpha}$
$\mu_X - \mu_Y = \delta_0$	$\mu_X - \mu_Y \neq \delta_0$	$ Z_0 > z_{1-\frac{\alpha}{2}}$

$$Z_0 := \frac{\bar{X}_m - \bar{Y}_n - \delta_0}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}$$

Campioni indipendenti di Bernoulli numerosi:

H_0	H_1	rifiuto H_0 se
$p_X - p_Y = \delta_0$ $p_X - p_Y \leq \delta_0$	$p_X - p_Y > \delta_0$	$Z_0 > z_{1-\alpha}$
$p_X - p_Y = \delta_0$ $p_X - p_Y \geq \delta_0$	$p_X - p_Y < \delta_0$	$Z_0 < -z_{1-\alpha}$
$p_X - p_Y = \delta_0$	$p_X - p_Y \neq \delta_0$	$ Z_0 > z_{1-\frac{\alpha}{2}}$

$$Z_0 := \frac{\bar{X}_m - \bar{Y}_n - \delta_0}{\sqrt{\hat{P}(1-\hat{P}) \left(\frac{1}{m} + \frac{1}{n} \right)}}$$

$$\hat{P} := \frac{m\bar{X}_m + n\bar{Y}_n}{m+n}$$

Rapporto tra varianze σ_X^2/σ_Y^2

Con medie incognite, campioni indipendenti normali di numerosità m ed n rispettivamente:

H_0	H_1	rifiuto H_0 se
$\sigma_X^2 = \sigma_Y^2$ $\sigma_X^2 \leq \sigma_Y^2$	$\sigma_X^2 > \sigma_Y^2$	$F_0 > f_{1-\alpha}(m-1, n-1)$
$\sigma_X^2 = \sigma_Y^2$ $\sigma_X^2 \geq \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$	$F_0 < f_{\alpha}(m-1, n-1)$
$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 \neq \sigma_Y^2$	$F_0 < f_{\frac{\alpha}{2}}(m-1, n-1)$ oppure $F_0 > f_{1-\frac{\alpha}{2}}(m-1, n-1)$

$$F_0 := \frac{S_X^2}{S_Y^2}$$

Formule utili

$z_{1-\alpha} = -z_{\alpha}$ $t_{1-\alpha}(n) = -t_{\alpha}(n)$

$t_{\alpha}(n) \simeq z_{\alpha}$ $\chi_{\alpha}^2(n) \simeq \frac{(z_{\alpha} + \sqrt{2n-1})^2}{2}$ se n è grande

$f_{\alpha}(m, n) = \frac{1}{f_{1-\alpha}(n, m)}$

Regressione lineare

Regressione semplice ($k = 1$ predittori)

Modello:

Y_1, Y_2, \dots, Y_n indipendenti con $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ e parametri $\beta_0, \beta_1, \sigma^2$ incogniti.

Sintesi dei dati:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$
$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad s_{xY} = \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) \quad S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Retta di regressione:

$$\hat{Y} = \hat{B}_0 + \hat{B}_1 x \quad (\text{con } \hat{Y}_i = \hat{B}_0 + \hat{B}_1 x_i)$$

dove

$$\hat{B}_0 = \bar{Y} - \frac{S_{xY}}{s_{xx}} \bar{x} \quad \text{stimatore di } \beta_0 \quad \hat{B}_1 = \frac{S_{xY}}{s_{xx}} \quad \text{stimatore di } \beta_1$$

Altre quantità:

$$SS_E = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = S_{YY} - \frac{S_{xY}^2}{s_{xx}} \quad \text{error sum of squares}$$
$$SS_R = \sum_{i=1}^n (\bar{Y} - \hat{Y}_i)^2 = \frac{S_{xY}^2}{s_{xx}} \quad \text{sum of squares due to regression}$$
$$SS_T = S_{YY} \quad \text{total sum of squares}$$
$$\hat{\Sigma}^2 = \frac{SS_E}{n-2} \quad \text{stimatore di } \sigma^2$$
$$\hat{\Sigma} = \sqrt{\hat{\Sigma}^2} \quad \text{residual standard error}$$

Coefficiente di determinazione:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} = \frac{S_{xY}^2}{s_{xx} S_{YY}}$$

Coefficiente di correlazione campionario:

$$R = \frac{S_{xY}}{\sqrt{s_{xx} S_{YY}}}$$

Errori standard:

$$\text{se}(\hat{B}_0) = \hat{\Sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}} \quad \text{se}(\hat{B}_1) = \frac{\hat{\Sigma}}{\sqrt{s_{xx}}}$$

Regressione multipla (k predittori)

Modello:

Y_1, Y_2, \dots, Y_n indipendenti con $Y_i \sim N(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}, \sigma^2) = N((1, \mathbf{x}_i^t)\boldsymbol{\beta}, \sigma^2)$
e parametri $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^t$, σ^2 incogniti.

Sintesi dei dati:

$$\tilde{\mathbf{x}} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} = \begin{pmatrix} (1, \mathbf{x}_1^t) \\ (1, \mathbf{x}_2^t) \\ \dots \\ (1, \mathbf{x}_n^t) \end{pmatrix} \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Retta di regressione:

$$\hat{Y} = \hat{B}_0 + \hat{B}_1 x_1 + \dots + \hat{B}_k x_k = (1, \mathbf{x}^t) \hat{\mathbf{B}} \quad (\text{con } \hat{Y}_i = (1, \mathbf{x}_i^t) \hat{\mathbf{B}})$$

dove

$$\hat{\mathbf{B}} = (\tilde{\mathbf{x}}^t \tilde{\mathbf{x}})^{-1} \tilde{\mathbf{x}}^t \mathbf{Y} \quad \text{e} \quad \mathbf{Y} = (Y_1, \dots, Y_n)^t$$

Altre quantità:

$$SS_E = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \mathbf{Y}^t \mathbf{Y} - \mathbf{Y}^t \tilde{\mathbf{x}} (\tilde{\mathbf{x}}^t \tilde{\mathbf{x}})^{-1} \tilde{\mathbf{x}}^t \mathbf{Y} \quad \text{error sum of squares}$$

$$SS_R = \sum_{i=1}^n (\bar{Y} - \hat{Y}_i)^2 = S_{YY} - SS_E \quad \text{sum of squares due to regression}$$

$$SS_T = S_{YY} \quad \text{total sum of squares}$$

$$\hat{\Sigma}^2 = \frac{SS_E}{n - k - 1} \quad \text{stimatore di } \sigma^2$$

$$\hat{\Sigma} = \sqrt{\hat{\Sigma}^2} \quad \text{residual standard error}$$

Coefficiente di determinazione e di determinazione corretto:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \quad R_A^2 = 1 - \frac{SS_E}{SS_T} \frac{n-1}{n-k-1}$$

Errori standard:

$$\text{se}(\hat{B}_r) = \hat{\Sigma} \sqrt{[(\tilde{\mathbf{x}}^t \tilde{\mathbf{x}})^{-1}]_{rr}}$$

Intervalli di confidenza di livello γ

Coefficiente di regressione β_r

$$\begin{aligned} & \left(\hat{B}_r - t_{\frac{1+\gamma}{2}}(n-k-1) \text{se}(\hat{B}_r), \hat{B}_r + t_{\frac{1+\gamma}{2}}(n-k-1) \text{se}(\hat{B}_r) \right) \\ & \left(\hat{B}_r - t_{\gamma}(n-k-1) \text{se}(\hat{B}_r), +\infty \right) \quad \left(-\infty, \hat{B}_r + t_{\gamma}(n-k-1) \text{se}(\hat{B}_r) \right) \end{aligned}$$

Media $\beta_0 + \beta_1 x^*$ di una nuova osservazione (solo caso $k = 1$)

$$\begin{aligned} & \left(\hat{B}_0 + \hat{B}_1 x^* - t_{\frac{1+\gamma}{2}}(n-2) \hat{\Sigma} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}, \right. \\ & \quad \left. \hat{B}_0 + \hat{B}_1 x^* + t_{\frac{1+\gamma}{2}}(n-2) \hat{\Sigma} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}} \right) \\ & \left(\hat{B}_0 + \hat{B}_1 x^* - t_{\gamma}(n-2) \hat{\Sigma} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}, +\infty \right) \\ & \left(-\infty, \hat{B}_0 + \hat{B}_1 x^* + t_{\gamma}(n-2) \hat{\Sigma} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}} \right) \end{aligned}$$

Intervalli di predizione di livello γ per una nuova osservazione (solo caso $k = 1$)

$$\begin{aligned} & \left(\hat{B}_0 + \hat{B}_1 x^* - t_{\frac{1+\gamma}{2}}(n-2) \hat{\Sigma} \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}, \right. \\ & \quad \left. \hat{B}_0 + \hat{B}_1 x^* + t_{\frac{1+\gamma}{2}}(n-2) \hat{\Sigma} \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}} \right) \\ & \left(\hat{B}_0 + \hat{B}_1 x^* - t_{\gamma}(n-2) \hat{\Sigma} \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}, +\infty \right) \\ & \left(-\infty, \hat{B}_0 + \hat{B}_1 x^* + t_{\gamma}(n-2) \hat{\Sigma} \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}} \right) \end{aligned}$$

Test di livello di significatività α

Coefficiente di regressione β_r :

H_0	H_1	rifiuto H_0 se
$\beta_r = \beta_{r0}$ $\beta_r \leq \beta_{r0}$	$\beta_r > \beta_{r0}$	$T_0 > t_{1-\alpha}(n-k-1)$
$\beta_r = \beta_{r0}$ $\beta_r \geq \beta_{r0}$	$\beta_r < \beta_{r0}$	$T_0 < -t_{1-\alpha}(n-k-1)$
$\beta_r = \beta_{r0}$	$\beta_r \neq \beta_{r0}$	$ T_0 \geq t_{1-\frac{\alpha}{2}}(n-k-1)$

$$T_0 := \frac{\hat{B}_r - \beta_{r0}}{\text{se}(\hat{B}_r)}$$

Significatività di un gruppo di predittori:

$$F_0 := \frac{[SS_e(\text{ridotto}) - SS_e(\text{completo})]/r}{SS_e(\text{completo})/(n-k-1)}$$

H_0	H_1	rifiuto H_0 se
$H_0 : \beta_1 = \beta_2 = \dots = \beta_r = 0$	$H_1 : \beta_s \neq 0$ per qualche $s = 1, \dots, r$	$F_0 > f_{1-\alpha}(r, n-k-1)$

Significatività del modello:

$$F_0 := \frac{SS_r/k}{SS_e/(n-k-1)}$$

H_0	H_1	rifiuto H_0 se
$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$	$H_1 : \beta_s \neq 0$ per qualche $s = 1, \dots, k$	$F_0 > f_{1-\alpha}(k, n-k-1)$