

N.B. Il formulario contiene alcune formule che non sono state trattate a lezione. Si prega di ignorarle.

FORMULARIO DI STATISTICA DESCRITTIVA

$$d = \frac{x_{\max} - x_{\min}}{k} \quad h_i = \frac{f_i}{d_i} \quad \text{oppure} \quad h_i = \frac{n_i}{d_i} \quad d_i = x_i - x_{i-1}$$

$$Me = \begin{cases} \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{se } n \text{ è pari} \\ x_{\frac{n}{2}+1} & \text{se } n \text{ è dispari} \end{cases} \quad Me \cong x_{Me-1} + (x_{Me} - x_{Me-1}) \frac{0,5 - F_{Me-1}}{F_{Me} - F_{Me-1}}$$

$$Q_1 = \begin{cases} \frac{x_{\frac{n}{4}} + x_{\frac{n}{4}+1}}{2} & \text{se } n \text{ è pari} \\ x_{\frac{n}{4}+1} & \text{se } n \text{ è dispari} \end{cases} \quad Q_1 \cong x_{Q_1-1} + (x_{Q_1} - x_{Q_1-1}) \frac{0,25 - F_{Q_1-1}}{F_{Q_1} - F_{Q_1-1}}$$

$$Q_3 = \begin{cases} \frac{x_{\frac{3n}{4}} + x_{\frac{3n}{4}+1}}{2} & \text{se } n \text{ è pari} \\ x_{\frac{3n}{4}+1} & \text{se } n \text{ è dispari} \end{cases} \quad Q_3 \cong x_{Q_3-1} + (x_{Q_3} - x_{Q_3-1}) \frac{0,75 - F_{Q_3-1}}{F_{Q_3} - F_{Q_3-1}}$$

$$\mu_x = \bar{x} = M(x) = \frac{1}{n} \sum_{i=1}^n x_i n_i \quad \mu_2 = M(x^2) = \frac{1}{n} \sum_{i=1}^n x_i^2 n_i$$

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 n_i = M(x^2) - (\bar{x})^2 \quad \sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 n_i}$$

$$Range(x) = x_{(n)} - x_{(1)} \quad IQR = Q_3 - Q_1 \quad CV = \frac{\sigma_x}{\mu_x}$$

$$G = 1 - \sum_{i=1}^k f_i^2 \quad G^* = G \frac{k}{k-1} \quad H = \sum_{i=1}^k f_i \log(f_i) \quad H^* = \frac{H}{\log(k)}$$

$$A_F = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 n_i}{\sigma_x^3} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right)^3 n_i \quad K = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right)^4 n_i - 3$$

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^h \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} \quad \phi^2 = \frac{\chi^2}{n}$$

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) n_i = M(xy) - M(x)M(y)$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \mu_x \mu_y}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 n_i} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 n_i}}$$

$$R = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i} \quad \approx R = 1 - \sum_{i=0}^{n-1} (p_{i+1} - p_i)(q_{i+1} - q_i)$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \hat{\beta}_1 = \frac{Cov(x,y)}{\sigma_x^2} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

FORMULARIO DI STATISTICA INFERENZIALE

$$P(\emptyset) = 0$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) > 0$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad P(A) > 0$$

$$P(A \cap B) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A)$$

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(H_i|E) = \frac{P(H_i \cap E)}{P(E)} = \frac{P(H_i) \cdot P(E|H_i)}{\sum_{j=1}^m P(H_j) \cdot P(E|H_j)}$$

$$\mu^r = E(X^r) = \sum x_i^r p_i$$

$$\mu^r = E(X^r) = \int_{-\infty}^{+\infty} x^r f(x) dx$$

$$\mu = E(X) = \sum x_i p_i$$

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$VAR(X) = E(X - \mu)^2 = \sum (x_i - \mu)^2 p_i = E(X^2) - [E(X)]^2$$

$$VAR(X) = \int_{-\infty}^{+\infty} (x_i - \mu)^2 f(x) dx$$

$$X \sim U(n) \quad P(X = x) = \frac{1}{n} > 0$$

$$E(X) = \frac{n+1}{2}$$

$$VAR(X) = \frac{n^2-1}{12}$$

$$X \sim Ber(p) \quad P(X = x) = p^x (1-p)^{1-x}$$

$$E(X) = p$$

$$VAR(X) = p(1-p)$$

$$X \sim Bin(n, p) \quad P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E(X) = np$$

$$VAR(X) = np(1-p)$$

$$X \sim Hg(N, M, n) \quad P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

$$E(X) = np$$

$$VAR(X) = np(1-p) \left(\frac{N-n}{N-1} \right)$$

$$X \sim Po(\lambda) \quad P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$E(X) = \lambda$$

$$VAR(X) = \lambda$$

$$X \sim Geo(p) \quad P(X = x) = p(1-p)^{x-1}$$

$$E(X) = \frac{1}{p}$$

$$VAR(X) = \frac{1-p}{p^2}$$

$$X \sim N(\mu, \sigma^2) \quad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$$

$$E(X) = \mu$$

$$VAR(X) = \sigma^2$$

$$Z = \frac{x-\mu}{\sigma}$$

$$X \sim \chi_{(r)}^2$$

$$E(X) = r$$

$$VAR(X) = 2r$$

$$X \sim t_{(r)}$$

$$E(X) = 0$$

$$VAR(X) = \frac{r}{r-2}$$

$$X \sim F_{(r_1, r_2)} \quad \frac{V_1/r_1}{V_2/r_2} = \frac{\chi_{(r_1)}^2/r_1}{\chi_{(r_2)}^2/r_2} \sim F_{(r_1, r_2)}$$

$$\frac{S_n}{n} \sim Bin(n, p) \quad \frac{S_n}{n} \sim N\left(p, \frac{p(1-p)}{n}\right) \text{ per } n \text{ grande}$$

$$\begin{aligned} \bar{X} &\sim N(\mu, \sigma^2) & E(\bar{X}) &= \mu & VAR(\bar{X}) &= \frac{\sigma^2}{n} \\ S^2 &= \frac{1}{n-1} \sum (x_i - \bar{X})^2 & \frac{(n-1)S^2}{\sigma^2} &\sim \chi_{(n-1)}^2 & T &= \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)} \end{aligned}$$

$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$\sigma_1^2, \sigma_2^2 \text{ note} \quad Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

$$\sigma_1^2 = \sigma_2^2 \text{ incognite} \quad T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \sim t_{(n_1+n_2-2)}$$

INTERVALLI DI CONFIDENZA

$$\bar{X} \sim N(\mu = ?, \sigma^2) \quad P\left[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right] = (1 - \alpha)$$

$$\bar{X} \sim N(\mu = ?, \sigma^2 = ?) \quad P\left[\bar{x} - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right] = (1 - \alpha)$$

$$\bar{X} \sim N(\mu, \sigma^2 = ?) \quad P\left[\frac{(n-1)S^2}{\chi_{(n-1), \frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{(n-1), (1-\frac{\alpha}{2})}^2}\right] = (1 - \alpha)$$

$$\frac{S_n}{n} \sim N\left(p = ?, \frac{p(1-p)}{n} = ?\right) \quad P\left[\frac{S_n}{n} - z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{S_n}{n}(1-\frac{S_n}{n})}{n}} \leq p \leq \frac{S_n}{n} + z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{S_n}{n}(1-\frac{S_n}{n})}{n}}\right] = (1 - \alpha)$$

$$X \sim N(\mu_1 = ?, \sigma_1^2) \quad X \sim N(\mu_2 = ?, \sigma_2^2)$$

$$P\left[(\bar{X}_1 - \bar{X}_2) - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = (1 - \alpha)$$

$$X \sim N(\mu_1, \sigma_1^2 = ?) \quad X \sim N(\mu_2, \sigma_2^2 = ?) \quad \sigma_1^2 = \sigma_2^2 = ?$$

$$P\left[(\bar{X}_1 - \bar{X}_2) - t_{t_{(n_1+n_2-2, \frac{\alpha}{2})}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) - t_{t_{(n_1+n_2-2, \frac{\alpha}{2})}} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = (1 - \alpha)$$

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

TEST DELLE IPOTESI

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z = \frac{\frac{S_n}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \cong N(0, 1) \quad Z = \frac{\left(\frac{S_{n1}}{n} - \frac{S_{n2}}{n}\right) - (p_1 - p_2)}{\sqrt{\frac{\frac{S_{n1}}{n}\left(1-\frac{S_{n1}}{n}\right)}{n_1} + \frac{\frac{S_{n2}}{n}\left(1-\frac{S_{n2}}{n}\right)}{n_2}}}$$

$$\frac{(n-1)S^2}{\sigma^2} \cong \chi_{(n-1)}^2$$

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - p_i)^2}{p_i} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad \chi^2 = \sum_{i=1}^k \sum_{j=1}^k \frac{(n_{ij} - c_{ij})^2}{c_{ij}} = \sum_{i=1}^k \sum_{j=1}^k \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$$

Disuguaglianza di Čebyšëv: $P\{|X - \mu| < \varepsilon\sigma\} \geq 1 - \frac{1}{\varepsilon^2}$ $P\{|X - \mu| < k\} \geq 1 - \frac{\sigma^2}{k^2}$

REGRESSIONE

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$Cov(X, Y) = \sum x_i y_i - (\bar{x} \bar{y}) \quad s_x^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \quad s_y^2 = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1}$$

$$\hat{y} = \beta_0^* + \beta_1^* x \quad \beta_1^* = \frac{Cov(x, y)}{\sigma_x^2} \quad \beta_0^* = \bar{y} - \beta_1^* \bar{x}$$

$$s_\varepsilon = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{(n-1)s_y^2 - \frac{Cov(x, y)^2}{s_x^2}}{n-2}}$$

Statistica test: $T = \frac{\beta_1^* - \beta_1}{s_{\beta_1^*}} \quad s_{\beta_1^*} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}}$

Intervallo di confidenza: $\beta_1^* \pm t_{\frac{\alpha}{2}, (n-2)} \cdot s_{\beta_1^*} \quad R^2 = 1 - \frac{SSE}{\sum (y_i - \bar{y})^2}$