FORMULARIO DI STATISTICA DESCRITTIVA

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

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

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

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

$$\mu_x = \overline{x} = M(x) = \frac{1}{n} \sum_{i=1}^{n} x_i n_i \qquad \mu_2 = M\left(x^2\right) = \frac{1}{n} \sum_{i=1}^{n} x_i^2 n_i$$

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 n_i = M\left(x^2\right) - \left(\overline{x}\right)^2 \qquad \sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 n_i}$$

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

$$G = 1 - \sum_{i=1}^{k} f_i^2 \qquad G^* = G \frac{k}{k-1} \qquad H = \sum_{i=1}^{k} |\; f_i \log\left(f_i\right)| \qquad H^* = \frac{H}{\log(k)}$$

$$R = \frac{\sum_{i=1}^{n-1} \left(p_i - q_i\right)}{\sum_{i=1}^{n-1} p_i} \qquad \widetilde{R} = 1 - \sum_{i=0}^{n-1} \left(p_{i+1} - p_i\right)\left(q_{i+1} + q_i\right)$$

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

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{\left(n_{ij} - \widehat{n}_{ij}\right)^2}{\widehat{n}_{ij}} \qquad \phi^2 = \frac{\chi^2}{n}$$

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

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

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x \qquad \widehat{\beta}_1 = \frac{Cov\,(x, y)}{\sigma_x^2} \qquad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

FORMULARIO DI STATISTICA INFERENZIALE

$P(\varnothing) = 0$

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

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

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

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

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

$\mu^r = E(X^r) = \sum x_i^r p_i$ $\qquad$ $\mu^r = E(X^r) = \int_{-\infty}^{+\infty} x^r f(x)\, dx$

$\mu = E(X) = \sum x_i p_i$ $\qquad$ $\mu = E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx$

$VAR(X) = E(X-\mu)^2 = \sum (x_i - \mu)^2 p_i = E(X^2) - [E(X)]^2$ $\qquad$ $VAR(X) = \int_{-\infty}^{+\infty} (x_i - \mu)^2 f(x)\, dx$

$X \sim U(a, b)$	$P(X = x) = \frac{1}{b-a} \quad (cost.)$	$E(X) = \frac{b+a}{2}$	$VAR(X) = \frac{(b-a)^2}{12}$
$X \sim Ber(p)$	$P(X = x) = p^x (1-p)^{1-x}$	$E(X) = p$	$VAR(X) = p(1-p)$
$X \sim Bin(n\ ,\ p)$	$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$	$E(X) = np$	$VAR(X) = np(1-p)$
$X \sim Hg(N, S, n)$	$P(X = x) = \frac{\binom{S}{x}\binom{N-S}{n-x}}{\binom{N}{n}}$	$E(X) = np \quad \left(p = \frac{S}{N}\right)$	$VAR(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$
$X \sim Po(\lambda)$	$P(X = x) = \lambda^x \cdot \frac{e^{-\lambda}}{x!}$	$E(X) = \lambda$	$VAR(X) = \lambda$
$X \sim Geo(p)$	$P(X = x) = p(1-p)^{x-1}$	$E(X) = \frac{1}{p}$	$VAR(X) = \frac{1-p}{p^2}$

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

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

$X \sim \chi^2_{(r)} \qquad E(X) = r \qquad VAR(X) = 2r$

$X \sim t_{(r)} \qquad E(X) = 0 \qquad VAR(X) = \frac{r}{r-2}$

$X \sim F_{(r_1\ ,\ r_2)} \qquad \frac{V_1/r_1}{V_2/r_2} = \frac{\chi^2_{(r_1)}/r_1}{\chi^2_{(r_2)}/r_2} \sim F_{(r_1,r_2)}$

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

$\overline{X} \sim N\left(\mu\ ,\ \sigma^2\right) \qquad E\left(\overline{X}\right) = \mu \qquad VAR\left(\overline{X}\right) = \frac{\sigma^2}{n}$

$S^2 = \frac{1}{n-1}\sum\left(x_i - \overline{X}\right)^2 \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \qquad T = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$

$\left(\overline{X}_1 - \overline{X}_2\right) \sim N\left(\mu_1 - \mu_2\ ,\ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}\right)$

$\sigma_1^2\ ,\ \sigma_2^2 \quad note \qquad Z = \frac{\left(\overline{X}_1-\overline{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n2}}} \sim N(0,1)$

$\sigma_1^2 = \sigma_2^2 \quad incognite \qquad T = \frac{\left(\overline{X}_1-\overline{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n2}\right)}} \qquad S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \sim t_{(n_1+n_2-2)}$

INTERVALLI DI CONFIDENZA

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

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

$\overline{X} \sim N\left(\mu\ ,\ \sigma^2 =?\right) \qquad P\left[\frac{(n-1)S^2}{\chi^2_{(n-1)',\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{(n-1)'\left(1-\frac{\alpha}{2}\right)}}\right] = (1-\alpha)$

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

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

$$P\left[\left(\overline{X}_1 - \overline{X}_2\right) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}} \leq (\mu_1 - \mu_2) \leq \left(\overline{X}_1 - \overline{X}_2\right) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n2}}\right] = (1-\alpha)$$

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

$$P\left[\left(\overline{X}_1 - \overline{X}_2\right) - t_{t_{\left(n_1+n_2-2,\, \frac{\alpha}{2}\right)}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n2}} \leq (\mu_1-\mu_2) \leq \left(\overline{X}_1 - \overline{X}_2\right) + t_{t_{\left(n_1+n_2-2,\, \frac{\alpha}{2}\right)}} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n2}}\right] = (1-\alpha)$$

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

TEST DELLE IPOTESI

$$Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \qquad T = \frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}$$

$$Z = \frac{\left(\overline{X}_1-\overline{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n2}}} \qquad T = \frac{\left(\overline{X}_1-\overline{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n2}\right)}}$$

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

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

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

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

REGRESSIONE

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

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

$$\widehat{y} = \beta_0^* + \beta_1^* x \qquad \beta_1^* = \frac{Cov(x,y)}{\sigma_x^2} \qquad \beta_0^* = \overline{y} - \beta_1^*\overline{x}$$

$$s_\varepsilon = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{(n-1)s_y^2 - \frac{COV(X,Y)}{s_x^2}}{n-2}}$$

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

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