Formulario di Probabilità e Statistica

Università degli Studi Milano Bicocca

Regressione Lineare

$$\boxed{A_0 = \overline{Y} - A_1\overline{X} \simeq N(\alpha_0, \sigma^2 \cdot [\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2}])}$$

$$\boxed{A_1 = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2} \simeq N(\alpha_1, \sigma^2 \cdot \frac{1}{\sum_{i=1}^n (x_i - \overline{x})^2})}$$

varianza campionaria dei residui

$$\boxed{S^2_{RES} = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}$$

$$T_0 = \frac{A_0 - \alpha_0}{\sqrt{S^2_{RES} \cdot [\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2}]}} \simeq t_{n-2}$$

$$T_1 = \frac{A_1 - \alpha_1}{\sqrt{S^2_{RES} \cdot [\frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2}]}} \simeq t_{n-2}$$

intervalli di confidenza

$$\boxed{\left[\alpha_0 - t_{1-\frac{\alpha}{2}} \cdot \sqrt{s^2_{RES} \cdot [\frac{1}{n} + \frac{\overline{x}^2}{s^2_x}]}, \alpha_0 + t_{1-\frac{\alpha}{2}} \cdot \sqrt{s^2_{RES} \cdot [\frac{1}{n} + \frac{\overline{x}^2}{s^2_x}]}\right]}$$

$$\boxed{\left[\alpha_1 - t_{1-\frac{\alpha}{2}} \cdot \sqrt{s^2_{RES} \cdot \frac{1}{s^2_x}}, \alpha_1 + t_{1-\frac{\alpha}{2}} \cdot \sqrt{s^2_{RES} \cdot \frac{1}{s^2_x}}\right]}$$

Intervalli di confidenza per i valori dei singoli individui

$$s^2_x = \sum_{i=1}^n (x_i - \overline{x})^2 \qquad \Delta_0 = \frac{(x_0 - \overline{x})^2}{s^2_x}$$

$$\boxed{\left[\widehat{y}_0 - t_{1-\frac{\alpha}{2}} \cdot \sqrt{s^2_{RES} \cdot [\frac{1}{n} + \Delta_0]}, \widehat{y}_0 + t_{1-\frac{\alpha}{2}} \cdot \sqrt{s^2_{RES} \cdot [\frac{1}{n} + \Delta_0]}\right]}$$

$$\boxed{\left[\widehat{y}_0 - t_{1-\frac{\alpha}{2}} \cdot \sqrt{s^2_{RES} \cdot [\frac{1}{n} + 1 + \Delta_0]}, \widehat{y}_0 + t_{1-\frac{\alpha}{2}} \cdot \sqrt{s^2_{RES} \cdot [\frac{1}{n} + 1 + \Delta_0]}\right]}$$

Attendibilità del modello lineare

$$\boxed{\widetilde{F} = (n-2)\frac{D_{SP}}{D_{RES}} = \frac{D_{SP}}{S^2_{RES}} \simeq F_{(1,n-2)}}$$

statistica

$$\widetilde{T} = \frac{A_1}{\sqrt{S^2_{RES} \cdot [\frac{1}{\sum_{i=1}^{n}(X_i - \overline{X})^2}]}} \simeq t_{n-2}$$

$$\widetilde{T}^2 = \widetilde{F}$$

Test parametrico di incorrelazione

coefficiente di correlazione lineare

$$r_{xy} = \frac{c_{xy}}{s_x s_y} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_X)(y_i - \mu_Y)}{\sigma_X \sigma_Y}$$

stimatore

$$R_n = \frac{\sum_{i=1}^{n}(x_i - \overline{X}_n)(y_i - \overline{Y}_n)}{nS_{X,n}S_{Y,n}}$$

$$\widehat{T}_n = R_n\sqrt{\frac{n-2}{1-R_n^2}} \simeq t_{n-2}$$

$$\boxed{C = (-\infty, -t_{1-\frac{\alpha}{2}}) \cup (t_{1-\frac{\alpha}{2}}, +\infty)}$$

Test non parametrico di indipendenza

$$W = n \cdot \sum_{k=1}^{M_X}\sum_{j=1}^{M_Y}\frac{(f_{k,j} - f_k^X \cdot f_j^Y)^2}{f_k^X \cdot f_j^Y} \simeq \chi^2_{(M_X-1)(M_Y-1)}$$

$$\boxed{C = (\chi^2_{1-\alpha}, +\infty)}$$

Test non parametrico di incorrelazione

Coefficiente di correlazione dei ranghi R_S di Spearman

$$\boxed{R_S = 1 - \frac{6\sum_{i=1}^{n}(r_i^X - r_i^Y)^2}{n^3 - n}}$$

$$\widetilde{T}_n = R_S\sqrt{\frac{n-2}{1-R_S^2}} \simeq t_{n-2}$$

$$\boxed{C = (-\infty, -t_{1-\frac{\alpha}{2}}) \cup (t_{1-\frac{\alpha}{2}}, +\infty)}$$

Test parametrico sulla differenza delle medie di due popolazioni

$$T_{n,m} = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{(n+m)(nS^2_{X,n}+mS^2_{Y,m})}{nm(n+m-2)}}}$$

$$\boxed{\begin{array}{lr} C' = (-\infty, -t_{1-\frac{\alpha}{2}}) \cup (t_{1-\frac{\alpha}{2}}, +\infty) & \text{(se } H_1' \text{ è l'ipotesi alternativa)} \\ C'' = (-\infty, -t_{1-\alpha}) & \text{(se } H_1'' \text{ è l'ipotesi alternativa)} \\ C''' = (t_{1-\alpha}, +\infty) & \text{(se } H_1''' \text{ è l'ipotesi alternativa)} \end{array}}$$

Test per la differenza delle medie con varianze note

$$Z_{n,m} = \frac{(\overline{X}_n - \overline{Y}_m) - \mu}{\sqrt{\frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{m}}} \simeq N(0,1)$$

Test parametrico sulla differenza delle varianze di due popolazioni

$$V_{n,m} = \frac{\widehat{S}^2_{X,n}}{\widehat{S}^2_{Y,m}}$$

$$\boxed{\begin{array}{lr} H_1' : \sigma^2_Y/\sigma^2_X \neq 1 & C' = [0, f_{\frac{\alpha}{2}}) \cup (f_{1-\frac{\alpha}{2}}, +\infty) \\ H_1'' : \sigma^2_Y/\sigma^2_X > 1 & C'' = [0, f_\alpha) \\ H_1''' : \sigma^2_Y/\sigma^2_X < 1 & C''' = (f_{1-\alpha}, +\infty) \end{array}}$$

Test non parametrico dei ranghi di Wilcoxon

$$U_X = nm + \frac{n(n+1)}{2} - R_X$$

$$U_Y = nm + \frac{m(m+1)}{2} - R_Y$$

$$U = \min\{U_X, U_Y\} \simeq N(\frac{nm}{2}, \frac{nm(n+m+1)}{12})$$

$$\widehat{Z}_{n,m} = \frac{U - \frac{nm}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \simeq N(0,1)$$

$$\boxed{C = (-\infty, -z_{1-\frac{\alpha}{2}}) \cup (z_{1-\frac{\alpha}{2}}, +\infty)}$$

Distribuzioni Notevoli

Distribuzione Triangolare

Densità

$$\boxed{f_X(t) = \begin{cases} \dfrac{4(t-a)}{(b-a)^2} & \text{se } t \in [a, \dfrac{a+b}{2}) \\ \dfrac{4(b-t)}{(b-a)^2} & \text{se } t \in [\dfrac{a+b}{2}, b] \\ 0 & \text{altrimenti} \end{cases}}$$

Funzione di ripartizione

$$\boxed{F_X(t) = \begin{cases} 0 & \text{se}\, t < a \\ 2\dfrac{(t-a)^2}{(b-a)^2} & \text{se}\, t \in [a, \dfrac{a+b}{2}) \\ 2\dfrac{(b-t)^2}{(b-a)^2} & \text{se}\, t \in [\dfrac{a+b}{2}, b] \\ 1 & \text{se}\, t > b \end{cases}}$$

valore atteso e varianza

$$\boxed{\mathbf{E}[X] = \frac{a+b}{2}} \qquad \boxed{\mathbf{V}[X] = \frac{(b-a)^2}{24}}$$

Distribuzione Normale (o di Gauss)

$$\mathbf{P}(X \in [a,b]) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}} dt$$

Indici di tendenza centrale e variabilità per variabili multidimensionali

Covarianza

$$\mathbf{Cov}[X,Y] = \mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])] = \iint_{\mathbb{R}^2} (t-\mathbf{E}[X])(s-\mathbf{E}[Y]) f_{X,Y}(t,s)\, dt\, ds$$

$$\mathbf{Cov}[X,Y] = \mathbf{E}[X \cdot Y] - \mathbf{E}[X] \cdot \mathbf{E}[Y] = \iint_{\mathbb{R}^2} ts\, f_{X,Y}(t,s)\, dt\, ds - \int_{\mathbb{R}} t\, f_X(t)\, dt \cdot \int_{\mathbb{R}} s\, f_Y(s)\, ds$$

Coefficiente di correlazione lineare di Pearson

$$\rho_{XY} = \frac{\mathbf{Cov}[X,Y]}{\sqrt{\mathbf{V}[X] \cdot \mathbf{V}[Y]}}$$

Un ringraziamento a Luca Chiodini per l'attenta lettura della prima versione e la segnalazione degli errori.

Finito di stampare il 2 giugno 2017