

FORMULARIO DEL CORSO DI STATISTICA

Campionamento da popolazioni finite

$$\text{CCS: } \bar{y}_{ccs} = \frac{\sum y_i}{n} \quad V(\bar{y}_{ccs}) = \frac{S^2}{n}(1-f); \quad \text{con } f = \frac{n}{N}; \quad v(\bar{y}_{ccs}) = \frac{s^2}{n}(1-f)$$

$$\text{STR: } \bar{y}_{str} = \sum_h W_h \bar{y}_h; \quad V(\bar{y}_{str}) = \sum W_h^2 \frac{S_h^2}{n_h}(1-f_h); \quad V(\bar{y}_{strp}) = \frac{(1-f)}{n} \sum W_h S_h^2$$

Inferenza statistica: stima per intervallo

$$\text{Intervallo di confidenza per la media: } \bar{y} \pm t(se); \quad \text{con } se = \frac{s}{\sqrt{n}}$$

$$\text{Intervallo di confidenza per la proporzione } \hat{\pi} \pm z(se) \quad \text{con } se = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Analisi dell'associazione tra variabili categoriche

$$\text{Test chi-quadro per } H_0: \text{indipendenza} \quad \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}; \quad gdl = (r-1)(c-1)$$

$$\text{Misure ordinali } \hat{\gamma} = \frac{C-D}{C+D}, \quad -1 \leq \hat{\gamma} \leq 1$$

Regressione lineare e correlazione

Modello di regressione lineare $E(y) = \alpha + \beta x$, equazione di previsione $\hat{y} = a + bx$

$$\text{Stime dei minimi quadrati, } b = \frac{\sum (y - \bar{y})(x - \bar{x})}{\sum (x - \bar{x})^2}; \quad a = \bar{y} - b\bar{x}$$

$$\text{Coefficiente di correlazione lineare, } r = b \left(\frac{s_x}{s_y} \right); \quad -1 \leq r \leq 1$$

$$R^2 = \frac{TSS - SSE}{TSS}, \quad TSS = \sum (y - \bar{y})^2, \quad SSE = \sum (y - \hat{y})^2, \quad 0 \leq R^2 \leq 1$$

$$\text{Test di indipendenza (lineare), } H_0: \beta = 0, \quad t = \frac{b}{se}, \quad gdl = n - 2$$

Regressione multipla e correlazione

Modello di regressione multipla $E(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$

Test complessivo $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$

$$\text{Statistica test } F = \frac{(TSS - SSE)/k}{SSE/[n - (k + 1)]} = \frac{R^2/k}{(1 - R^2)/[n - (k + 1)]};$$

$$gdl_1 = k; \quad gdl_2 = n - (k + 1)$$

$$\text{Test parziale } H_0: \beta_i = 0, \quad \text{statistica test } t = \frac{b_i}{se}, \quad gdl = n - (k + 1)$$

Confronto tra modelli. Test H_0 : parametri dei termini aggiuntivi tutti uguali a 0

$$\text{Statistica test } F = \frac{(SSE_r - SSE_c)/gdl_1}{SSE_c/gdl_2} = \frac{(R_c^2 - R_r^2)/gdl_1}{(1 - R_c^2)/gdl_2}$$

Con $gdl_1 = n$. variabili aggiuntive; $gdl_2 = n - (k + 1)$; con $k = n$. variabili modello completo