

L'analisi multivariabile

- Prof. Giuseppe Verlato
- Sezione di Epidemiologia e Statistica Medica,
Dipartimento di Medicina e Sanità Pubblica,
Università degli Studi di Verona

Analisi UNIVARIABILE:
sufficiente per una tesi

Una X \longrightarrow Una Y

Analisi MULTIVARIABILE :
necessaria per un articolo scientifico

X1 \longrightarrow
X2 \longrightarrow
X3 \longrightarrow Y

Migrazione di stadio in funzione dell'estensione dell'intervento in pazienti affetti da cancro gastrico

ANALISI BIVARIATA

Linfonodi asportati \longleftrightarrow T di Kendall = 0.192
 $P < 0.001$ linfonodi positivi

N.B. Viene usato una statistica non-parametrica (T , **coefficiente di correlazione per ranghi di Kendall**) anziché una statistica parametrica (r) perché le due variabili in studio sono distribuite in modo fortemente asimmetrico.

Linfonodi asportati \longleftrightarrow $T = 0.192$
 $P < 0.001$ linfonodi positivi
 $T = 0.151$
 $P < 0.001$ Stadio T $T = 0.525$
 $P < 0.001$

Lo stadio T si correla fortemente con il numero di linfonodi positivi e si correla un po' anche con il numero di linfonodi asportati. **Che sia un confondente?**

ANALISI TRIVARIATA

Linfonodi asportati \longleftrightarrow $T = 0.134$
 $P < 0.001$ linfonodi positivi
 $T = 0.060$
 $P = 0.006$ Stadio T $T = 0.511$
 $P < 0.001$

Se si controlla per lo stadio T, attraverso il coefficiente di correlazione parziale per ranghi di Kendall, l'associazione tra linfonodi asportati e linfonodi positivi si indebolisce.

N.B.: Ci sono 921 soggetti per cui anche una correlazione modesta risulta significativa.

G de Manzoni, G Verlato, et al (2003) The new TNM classification of lymph node metastasis minimises stage migration problems in gastric cancer patients. Br J Cancer 87: 171-174

VARIABILI QUANTITATIVE

Analisi univariabile: 1 Y, 1 X

1) Regressione lineare semplice

$$y = \beta_0 + \beta_1 x + \varepsilon, \text{ in cui } X \text{ ed } Y \text{ sono variabili quantitative}$$

Analisi multivariabile: 1 Y, molte X

2) Regressione lineare multipla

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \text{ in cui } X \text{ ed } Y \text{ sono variabili quantitative}$$

3) Analisi della varianza (ANOVA)

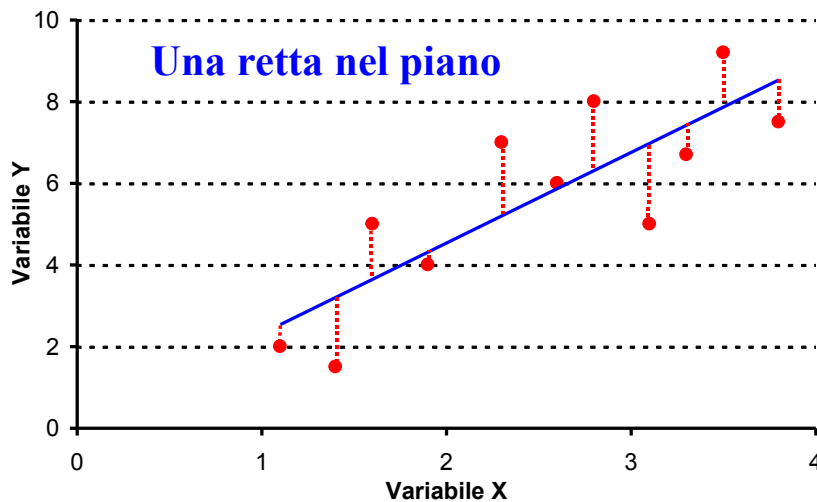
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \text{ in cui } Y \text{ quantitativa, } X \text{ qualitative}$$

4) Analisi della covarianza (ANCOVA)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \text{ in cui } Y \text{ quantitativa, } X \text{ qualitative e quantitative}$$

Regressione lineare semplice

$$y = \beta_0 + \beta_1 x + \varepsilon$$

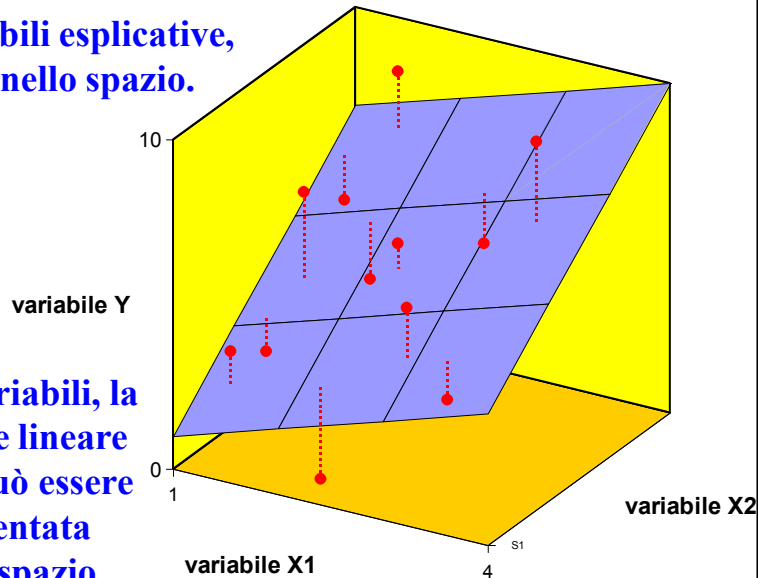


Regressione lineare multipla

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Con 2 variabili esplicative,
un piano nello spazio.

Con più variabili, la
regressione lineare
multipla può essere
rappresentata
nell'iperspazio



E per le variabili qualitative NOMINALI ?

2 VARIABILI (entrambe qualitative):
test del chi-quadrato, test esatto di Fischer

3 VARIABILI qualitative (2 var. + 1 var. di
stratificazione): test di Mantel-Haenszel

MOLTE VARIABILI:

y dicotomica (malato/sano) → modello LOGISTICO

y politomica (fumatore, ex-fumatore, mai-fumatore) →
modello MULTINOMIALE

MODELLO DI REGRESSIONE LOGISTICA

19 / (19+132)	prevalenza
0 / (0+9)	
11 / (11+52)	
6 / (6+97)	

MODELLO LOG-LINEARE

19	132	conteggi
0	9	
11	52	
6	97	

MODELLO di POISSON

19 / 1510 persone-anno	Tassi di incidenza
0 / 90 persone-anno	
11 / 630 persone-anno	
6 / 103 persone-anno	

ANALISI DELLA SOPRAVVIVENZA

1) FASE DESCRITTIVA

Curve di sopravvivenza secondo il metodo del prodotto limite di Kaplan-Meier

2) FASE INFERENZIALE

Stime della sopravvivenza: metodo di Kaplan-Meier

Analisi univariabile: Confronto fra due o più curve di sopravvivenza (log-rank test)

Analisi multivariabile: Valutazione simultanea del significato prognostico di più fattori di rischio (modello di Cox)

Confronto fra due curve di sopravvivenza: il log-rank test

