# DATA MINING PER IL MARKETING (63 ore)

Marco Riani

mriani@unipr.it

Sito web del corso

**http://www.riani.it/DMM**

# Regressione multipla

## Forma generale del modello di regressione

$$y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \ldots + \beta_{k-1} x_{ik-1} + \varepsilon_i$$

- Il modello è ancora lineare nei coefficienti $\beta_j$ ma la sostanziale differenza rispetto al modello lineare semplice risiede nella presenza d'un maggior numero di variabili indipendenti $X_j$

$$y_i = \mathbf{x}_i' \beta + \varepsilon_i$$

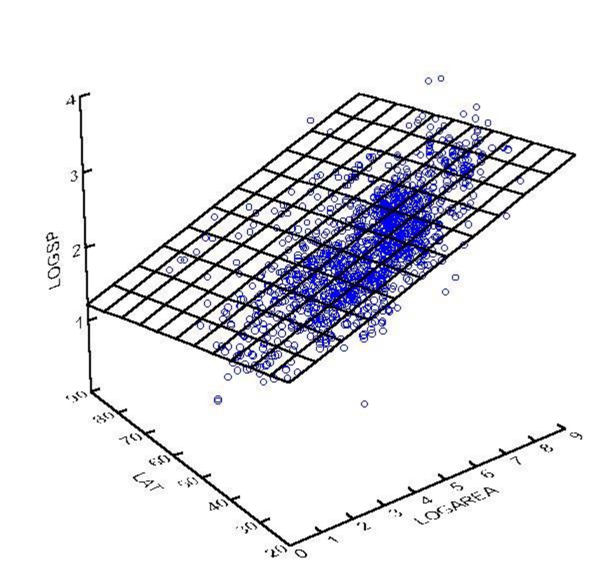
## In forma matriciale

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1,k-1} \\ 1 & x_{21} & \cdots & x_{2,k-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{n,k-1} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

# Obiettivo: adattare un piano “iperpiano” di regressione




# Generalizzazione delle assunzioni per la regr. lineare semplice

$$E[\mathbf{y}] = \mathbf{X}\beta.$$

- che implica

$$E[\varepsilon_i] = 0 \qquad i = 1, 2, ..., n$$

$$E[\varepsilon] = \begin{bmatrix} E[\varepsilon_1] \\ E[\varepsilon_2] \\ \vdots \\ E[\varepsilon_n] \end{bmatrix} = \mathbf{0}$$

# Assunzione sulla varianza dei termini di disturbo

$$E\left[\varepsilon\varepsilon'\right] = \sigma^2 \mathbf{I}$$

$$\begin{aligned} E\left[\varepsilon\varepsilon'\right] &= \begin{bmatrix} E\left[\varepsilon_1\varepsilon_1\right] & E\left[\varepsilon_1\varepsilon_2\right] & \cdots & E\left[\varepsilon_1\varepsilon_n\right] \\ E\left[\varepsilon_2\varepsilon_1\right] & E\left[\varepsilon_2\varepsilon_2\right] & \cdots & E\left[\varepsilon_2\varepsilon_n\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[\varepsilon_n\varepsilon_1\right] & E\left[\varepsilon_n\varepsilon_2\right] & \cdots & E\left[\varepsilon_n\varepsilon_n\right] \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \end{aligned}$$

# Richiami sulla matr. var. cov. di un vettore aleatorio (p. 297)

$$\begin{aligned} var(x) &= \begin{pmatrix} var(X_1) & cov(X_1,X_2) & \dots & cov(X_1,X_n) \\ cov(X_2,X_1) & var(X_2) & \dots & cov(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n,X_1) & cov(X_n,X_2) & \dots & var(X_n) \end{pmatrix} \\ &= E\begin{pmatrix} (X_1-E(X_1))^2 & (X_1-E(X_1))(X_2-E(X_2)) & \dots & (X_1-E(X_1))(X_n-E(X_n)) \\ (X_2-E(X_2))(X_1-E(X_1)) & (X_2-E(X_2))^2 & \dots & (X_2-E(X_2))(X_n-E(X_n)) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n-E(X_n))(X_1-E(X_1)) & (X_n-E(X_n))(X_2-E(X_2)) & \dots & (X_n-E(X_n))^2 \end{pmatrix} \\ &= E\begin{pmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \dots & (X_1-\mu_1)(X_n-\mu_n) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \dots & (X_2-\mu_2)(X_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n-\mu_n)(X_1-\mu_1) & (X_n-\mu_n)(X_2-\mu_2) & \dots & (X_n-\mu_n)^2 \end{pmatrix} \\ &= E\begin{pmatrix} (X_1-\mu_1) \\ (X_2-\mu_2) \\ \vdots \\ (X_n-\mu_n) \end{pmatrix} \begin{pmatrix} (X_1-\mu_1) & (X_2-\mu_2) & \dots & (X_n-\mu_n) \end{pmatrix} \\ &= E(x-\mu)(x-\mu)'. \end{aligned} \qquad \text{(B.1}$$

## Assunzione sulla matrice X

- **X** = fissa non stocastica, non dipende in alcun modo da **ε**

- **X ha rango pieno**

## Ripasso sulle matrici

- Addizione tra matrici
- Moltiplicazione
- Matrice diagonale
- Matrice identità
- Matrice trasposta (trasposta del prodotto)
- Matrice inversa
- Traccia
- Matrice idempotente
- Somma di quadrati (ponderata) in forma matriciale
- Forme quadratiche (positive, negative definite)
- Forme quadratiche idempotenti
- Scomposizione spettrale

## Valore atteso e var di comb. di v.c.

- x vettore casuale
- A = matrice non stocastica
- E(A x) = A E(x)
- var(A x) = A var(x) A'     (v. p. 299)

## Derivate di vettori e matrici (pp. 294-295)

- x vettore casuale

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ \dots \\ x_p \end{pmatrix} \qquad \frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \dots \\ \dots \\ \frac{\partial f}{\partial x_p} \end{pmatrix}$$

## Derivate di vettori e matrici

- Se **a**'=($a_1$, …, $a_p$)

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ \dots \\ x_p \end{pmatrix}$$

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \frac{a_1 x_1 + a_2 x_2 + \dots + a_p x_p}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \dots \\ \dots \\ \frac{\partial f}{\partial x_p} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ \dots \\ a_p \end{pmatrix} = \mathbf{a}$$

## Derivate di vettori e matrici

$$\frac{\partial x' A x}{\partial x} = 2Ax.$$

## Derivate di vettori e matrici

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_1' \\ a_2' \\ a_3' \end{pmatrix}$$

$$x'Ax = x_1^2 a_{11} + 2x_1x_2a_{12} + 2x_1x_3a_{13} + x_2^2a_{22} + 2x_2x_3a_{23} + x_3^2a_{33}.$$

$$\frac{\partial x'Ax}{\partial x} = \begin{pmatrix} \frac{\partial x'Ax}{\partial x_1} \\ \frac{\partial x'Ax}{\partial x_2} \\ \frac{\partial x'Ax}{\partial x_3} \end{pmatrix} = 2\begin{pmatrix} a_1'x \\ a_2'x \\ a_3'x \end{pmatrix} = 2\begin{pmatrix} a_1' \\ a_2' \\ a_3' \end{pmatrix} x = 2Ax.$$

## Stima di β

- Occorre trova il β che minimizza la seguente espressione

$$f\left(\tilde{\beta}\right) = \tilde{\varepsilon}'\tilde{\varepsilon} = \left(\mathbf{y} - \mathbf{X}\tilde{\beta}\right)'\left(\mathbf{y} - \mathbf{X}\tilde{\beta}\right) = \min .$$

$$\begin{aligned} f\left(\tilde{\beta}\right) &= \tilde{\varepsilon}'\tilde{\varepsilon} = \mathbf{y}'\mathbf{y} - \tilde{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\tilde{\beta} + \tilde{\beta}'\mathbf{X}'\mathbf{X}\tilde{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\tilde{\beta}'\mathbf{X}'\mathbf{y} + \tilde{\beta}'\mathbf{X}'\mathbf{X}\tilde{\beta}. \end{aligned}$$

## Stima di β

$$\begin{aligned} f\left(\tilde{\beta}\right) &= \tilde{\varepsilon}'\tilde{\varepsilon} = \mathbf{y}'\mathbf{y} - \tilde{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\tilde{\beta} + \tilde{\beta}'\mathbf{X}'\mathbf{X}\tilde{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\tilde{\beta}'\mathbf{X}'\mathbf{y} + \tilde{\beta}'\mathbf{X}'\mathbf{X}\tilde{\beta}. \end{aligned}$$

- Occorre trova il β che minimizza la seguente espressione

$$\frac{\partial f\left(\tilde{\beta}\right)}{\partial \tilde{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\tilde{\beta} = \mathbf{0}.$$

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

# Stima di β

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$

- Se la matrice X'X quadrata di ordine k ammette inversa, allora la soluzione è

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$