



**INFOMAAT**  
CENTRO RIPETIZIONI  
SINCE 2011



STATISTICA · UNIVERSITÀ · ESERCIZI SVOLTI

# Correlazione e regressione lineare

Tre esercizi svolti passo per passo: retta di regressione, covarianza, coefficiente di correlazione e determinazione, le due rette di regressione e le previsioni a partire dai dati.

Retta di regressione

Covarianza e correlazione

Previsioni

---

**INFOMAAT — Centro di ripetizioni**

Matematica e fisica · liceo e università · dal 2011

infomaat.it

# Correlazione e regressione in breve

Date due variabili  $X$  e  $Y$  osservate su  $n$  soggetti, misuriamo **quanto** e **come** variano insieme e costruiamo la retta che meglio approssima la nuvola di punti.

## MEDIE E CENTRO DI GRAVITÀ

La retta passa sempre per il punto  $(\bar{x}, \bar{y})$ .

$$\bar{x} = \frac{\sum x_i}{n} \quad \bar{y} = \frac{\sum y_i}{n}$$

## COVARIANZA E VARIANZE

Misurano la variabilità congiunta e quella di ciascuna variabile.

$$s_{xy} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y}$$

$$s_x^2 = \frac{\sum x_i^2}{n} - \bar{x}^2 \quad s_y^2 = \frac{\sum y_i^2}{n} - \bar{y}^2$$

## CORRELAZIONE E DETERMINAZIONE

$$r = \frac{s_{xy}}{s_x s_y} \quad -1 \leq r \leq 1$$

$r^2$  (determinazione) è la quota di variabilità di  $Y$  spiegata da  $X$ .

## LE DUE RETTE DI REGRESSIONE

Sono **diverse**: una stima  $Y$  da  $X$ , l'altra  $X$  da  $Y$ .

$$y = \frac{s_{xy}}{s_x^2} (x - \bar{x}) + \bar{y}$$

$$x = \frac{s_{xy}}{s_y^2} (y - \bar{y}) + \bar{x}$$

## COME SI LEGGE $r$

Il **segno** dà il verso del legame (diretto se  $r > 0$ , inverso se  $r < 0$ ); il **modulo** la forza:  $|r|$  vicino a 1 → legame lineare forte, vicino a 0 → legame lineare debole o assente. La covarianza  $s_{xy}$  ha lo stesso segno di  $r$  ma dipende dalle unità di misura, quindi non se ne confronta direttamente l'intensità.



## PROBLEMA 1

# Vendite e reddito nazionale

### CONSEGNA

Un'azienda vuole prevedere le vendite annue  $Y$  (migliaia di €) di un prodotto a partire dal reddito nazionale  $X$  (milioni di €),  $n = 11$  anni (1990–2000). Calcolare: **a)** la retta di regressione di  $Y$  su  $X$ ; **b)** il coefficiente di correlazione e interpretarlo; **c)** la previsione delle vendite se nel 2001  $X = 325$ .

|     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $X$ | 189 | 190 | 208 | 227 | 239 | 252 | 257 | 274 | 293 | 308 | 316 |
| $Y$ | 402 | 404 | 412 | 425 | 429 | 436 | 440 | 447 | 458 | 469 | 469 |

### A RETTA DI REGRESSIONE DI $Y$ SU $X$

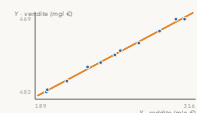
Statistiche di base:  $\bar{x} = 250,27$ ,  $\bar{y} = 435,55$ ,  $s_{xy} = 969,40$ ,  $s_x^2 = 1812,02$ .

$$b = \frac{s_{xy}}{s_x^2} = \frac{969,40}{1812,02} = 0,535 \quad a = \bar{y} - b\bar{x} = 301,65$$

### RETTA

$$y = 0,535x + 301,65$$

### LA NUVOLA DI PUNTI E LA RETTA



Punti quasi allineati: legame lineare diretto fortissimo.

### B COEFFICIENTE DI CORRELAZIONE

$s_x = 42,57$ ,  $s_y = 22,81$ , quindi  $r = \frac{s_{xy}}{s_x s_y} = \frac{969,40}{42,57 \cdot 22,81} = 0,998$ . Essendo  $r$  positivo e quasi pari a 1, esiste un legame lineare **diretto e fortissimo**: al crescere del reddito le vendite crescono in modo molto regolare.

### C PREVISIONE PER $X = 325$

$$y = 0,535 \cdot 325 + 301,65 = 475,52$$

Vendite previste per il 2001: circa **476 mila €**.



## PROBLEMA 2

### Ore di studio e voto d'esame

#### CONSEGNA

Per 8 studenti si registrano le ore dedicate allo studio  $X$  e il voto d'esame  $Y$ . Calcolare: **a)** retta di regressione di  $Y$  su  $X$ , coefficienti di correlazione e determinazione; **b)** la covarianza; **c)** il voto stimato per chi ha studiato 28 ore.

|          |     |    |     |    |    |     |     |    |
|----------|-----|----|-----|----|----|-----|-----|----|
| ORE (X)  | 20  | 16 | 34  | 23 | 27 | 32  | 18  | 22 |
| VOTO (Y) | 6,5 | 6  | 8,5 | 7  | 9  | 9,5 | 7,5 | 8  |

#### A RETTA, CORRELAZIONE, DETERMINAZIONE

$$\bar{x} = 24, \bar{y} = 7,75, s_x^2 = 36,75, s_y^2 = 1,3125.$$

$$b = \frac{s_{xy}}{s_x^2} = \frac{5,75}{36,75} = 0,156 \quad a = \bar{y} - b\bar{x} = 3,99$$

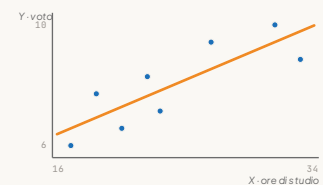
#### RETTA

$$y = 0,156x + 3,99$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{5,75}{6,06 \cdot 1,146} = 0,828 \text{ e } r^2 = 0,685:$$

correlazione positiva **buona**; circa il **69%** della variabilità del voto è spiegato dalle ore di studio.

#### ORE DI STUDIO VS VOTO



Tendenza crescente con dispersione moderata.

#### B COVARIANZA

$$s_{xy} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y} = \frac{1534}{8} - 24 \cdot 7,75 = 5,75$$

Positiva: ore di studio e voto variano nello **stesso verso**.

#### C VOTO STIMATO PER X = 28 ORE

$$y = 0,156 \cdot 28 + 3,99 = 8,38$$

Voto atteso: circa **8,38**.



### PROBLEMA 3

## Voti di Statistica e Chimica

#### CONSEGNA

Voti di 10 studenti in **Statistica** ( $X$ ) e **Chimica** ( $Y$ ). Ottenere: **a)** covarianza e correlazione; **b)** tipo di dipendenza; **c)** centro di gravità; **d)** le due rette di stima; **e)** voto atteso in Statistica con Chimica = 8,3; **f)** voto atteso in Chimica con Statistica = 9,0.

|            |     |     |   |   |     |   |   |    |   |   |
|------------|-----|-----|---|---|-----|---|---|----|---|---|
| STATISTICA | 6   | 4   | 8 | 5 | 3,5 | 7 | 5 | 10 | 5 | 4 |
| CHIMICA    | 6,5 | 4,5 | 7 | 5 | 4   | 8 | 7 | 10 | 6 | 5 |

#### A COVARIANZA E CORRELAZIONE

$\bar{x} = 5,75$ ,  $\bar{y} = 6,30$ ,  $s_x^2 = 3,763$ ,  $s_y^2 = 2,96$ ,  
 $s_x = 1,940$ ,  $s_y = 1,720$ .

$$s_{xy} = 3,075 \quad r = \frac{s_{xy}}{s_x s_y} = 0,921$$

#### B TIPO DI DIPENDENZA

$r = 0,921 > 0$  e vicino a 1  $\rightarrow$  dipendenza lineare **diretta e forte**.  $r^2 = 0,849$ : l'85% della variabilità di una materia è spiegato dall'altra.

#### C CENTRO DI GRAVITÀ

È il punto  $(\bar{x}, \bar{y}) = (5,75; 6,30)$ , per cui passano **entrambe** le rette.

#### D LE DUE RETTE DI STIMA

##### CHIMICA DA STATISTICA

$$y = 0,817x + 1,601$$

##### STATISTICA DA CHIMICA

$$x = 1,039y - 0,795$$

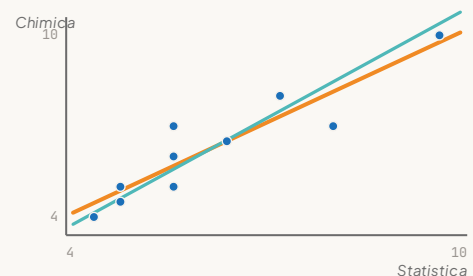
##### E) STATISTICA SE CHIMICA = 8,3

$$x = 1,039 \cdot 8,3 - 0,795 = 7,83$$

##### F) CHIMICA SE STATISTICA = 9,0

$$y = 0,817 \cdot 9,0 + 1,601 = 8,96$$

#### LE DUE RETTE DI REGRESSIONE



— Chimica su Statistica — Statistica su Chimica. Si incrociano in  $(5,75; 6,30)$ .

