Matematica e Statistica 15/11/17

Statistica elementare

popolazione = insieme delle unità su cui si esegue una certa misura/valutazione

carattere = la quantità / caratteristica che viene rilevata

modalità = i modi in cui il carattere può manifestarsi

Esempi:

1) gli alunni di VC del Liceo... | voto all'ultimo compito di mate | $\underbrace{0, 1, 2, \dots, 10}_{\text{quantitativa discreta}}$

modalità: numeri naturali

2) la popolaz. italiana maggiorenne — altezza — [100, 250] (misure in cm)

quantitativa continua

modalità: numeri reali

3) la popolaz. italiana — religione — nessuna, cattolicesimo, islam . . .

qualitativa

NON CE NE OCCUPIAMO

Supponiamo popolazione con n unità.
Misuriamo un carattere che può assumere

le modalità $v_1, \dots, v_k$. Se la modalità v_j è assunta da n_j unità chiamo

n_j = frequenza assoluta di v_j

n_j/n = frequenza relativa

<u>Es</u>: Popolaz: $\{A, B, C, D, E, F, G, H, I, J\}$

Carattere = numero di lettere con cui si legge la lettera

A	B	C	D	E	F	G	H	I	J
1	2	2	2	1	4	2	4	1	3

modalità	freq. ass.	freq. rel.
1	3	0.3
2	4	0.4
3	1	0.1
4	2	0.2

Def: chiamiamo per una statistica

- moda = la modalità più frequente
- mediana = quella modalità m t.c. la freq. complessiva di tutte le modalità $< m$ e di tutte quelle $> m$ sono entrambe $< 1/2$.
- media (aritmetica)

$$\mu = \frac{1}{n}\sum_{j=1}^{k} n_j v_j = \sum_{j=1}^{k} f_j \cdot v_j$$

frequenze rel   modalità

= somma pesata delle modalità secondo la loro frequenza relativa.

Esempio: Voti nell'ultimo compito di mate di VF:

Voto	3	4	5	6	7	8	9	10
Freq. Ass.	2	1	3	5	4	4	3	2

Tot. Pop. 24

Moda: 6

Mediana: 7 ($11/24 < 1/2$ hanno < 7

$9/24 < 1/2$ hanno > 7)

Media $= \frac{1}{24}\left(2 \cdot 3 + 1 \cdot 4 + 3 \cdot 5 + 5 \cdot 6 + 4 \cdot 7 + 4 \cdot 8 + 3 \cdot 9 + 2 \cdot 10\right)$

---

misura della dispersione di un campione

= "quanto è lontano dalla media"

varianza: $\sigma^2 = \frac{1}{n} \sum_{i=1}^{k} n_i \left(x_i - \mu\right)^2$

Esercizio: $\sum_{i=1}^{k} n_i(v_i - \mu) = 0$.

$\sigma = \sqrt{\sigma^2}$ è detta deviazione standard o scarto quadratico medio.

Legame tra due dati.

Supponiamo di avere due statistiche x, y

sulla stessa popolazione con risultati:

$x_1, \dots, x_n$

$y_1, \dots, y_n$

→ ammetto ripetizioni

→ intendo che le misure sono eseguite sulle unità nel medesimo ordine

Es: popolazione = noi qui ora

x = altezze

y = numero di scarpe

individuo	C.P.	F.G.	S.S.
altezza	188	155	157
scarpe	41	37	37

Def: chiamo covarianza di x, y

$$\sigma_{x,y} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)$$

e correlazione di x e y

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \cdot \sigma_y}.$$

Esempio: se accadesse che $y = m x + q$ allora:

$$\mu_y = m \cdot \mu_x + q \qquad \left(\mu_y = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{i=1}^{n} (m \cdot x_i + q) \right.$$

$$\left. = m \cdot \underbrace{\frac{1}{n} \sum_{i=1}^{n} x_i}_{\mu_x} + q \cdot \underbrace{\frac{1}{n} \sum_{i=1}^{n} 1}_{1} \right)$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu_y)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(m x_i + q - (m \mu_x + q) \right)^2$$

$$= m^2 \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)^2$$

$$= m^2 \cdot \sigma_x^2$$

$$\Longrightarrow \sigma_y = |m| \cdot \sigma_x$$

$$\sigma_{x,y} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)(m x_i + q - (m \mu_x + q))$$

$$= m \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x)^2$$

$$= m \cdot \sigma_x^2$$

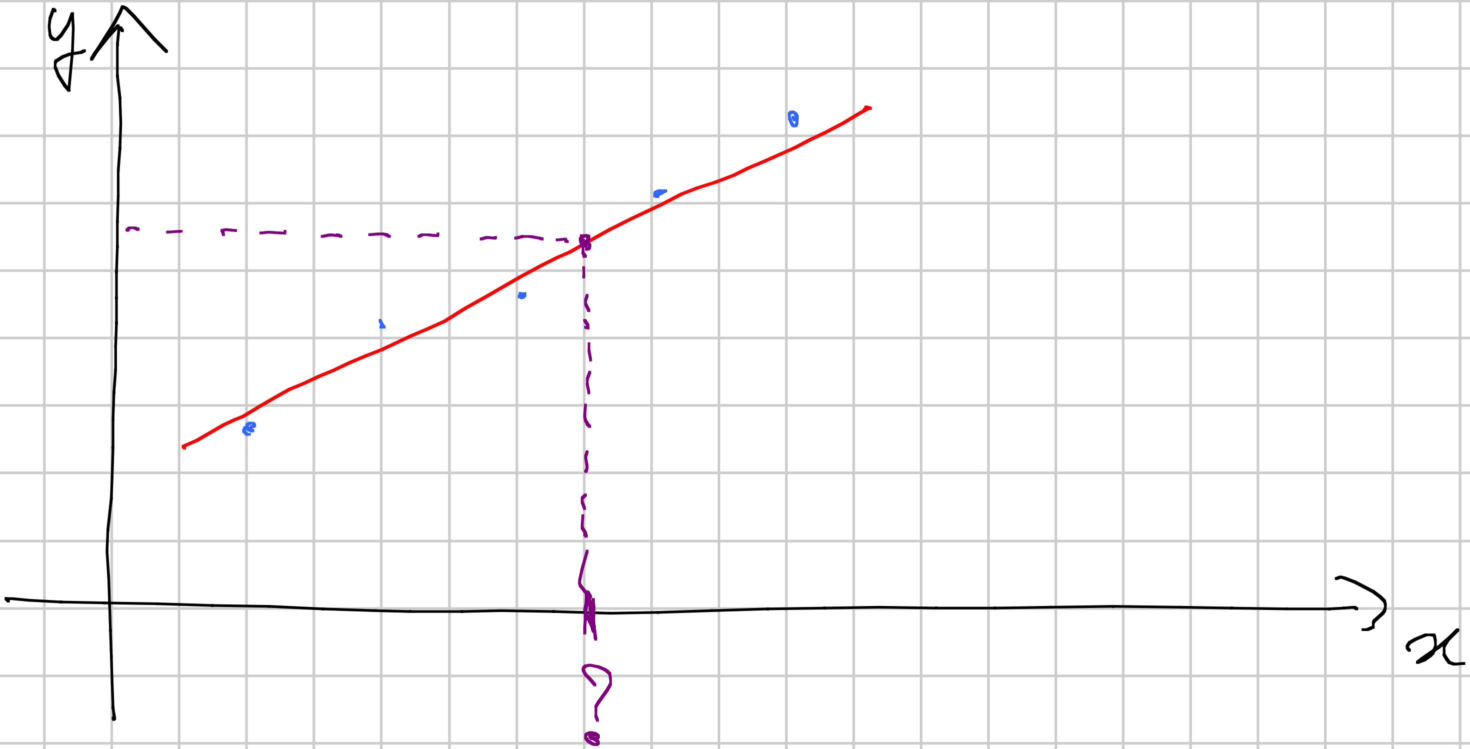
<u>Conclusione</u>: $\rho_{x,y} = \dfrac{\sigma_{x,y}}{\sigma_x \cdot \sigma_y} = \dfrac{m \cdot \sigma_x^2}{\sigma_x \cdot |m| \cdot \sigma_x} = \dfrac{m}{|m|}$

vale $+1$ per m positivo, -1 per m negativo.

Fatto: la correlazione è sempre compresa tra -1 e $+1$.

Interpretazione: se la correlazione è molto vicina a ± 1 possiamo ipotizzare che i dati siano legati da una legge lineare $y = m \cdot x + q$.

Scopo: poter fare previsioni



Dai conti già fatti si ricava che la retta "migliore" è $y = m \cdot x + q$ dove

$$m = \frac{\sigma_{x,y}}{\sigma_x^2} \qquad q = \mu_y - m \cdot \mu_x .$$

<u>Attenzione</u>: se ho un'altra legge di dipendenza $y = f(x)$ (ad $y = \sqrt{x}$ ...)

la correlazione tra x e y può essere bassa.

<u>Es</u>: popolaz: i dadi da gioco a Pisa

x = lato

y = volume

$y = x^3 \implies$ correlazione bassa.

Se non è conosciuta la dipendenza tra x e y ma viene ipotizzata la dipendenza $y = f(x)$, si verifica tale ipotesi se si ottiene una alta correlazione tra y e $f(x)$.

———— o ————

Ese foglio 14/11/17:

$$\lim_{x \to 1} \left(5x^3 - 3\sin\left(\frac{\pi}{2} \cdot x\right)\right) = 5 \cdot 1 - 3 \cdot 1 = 2$$

$$\lim_{x \to \infty} \cos(x)$$ non esiste

$$\lim_{x \to 0} \frac{x+1}{\cos(x)} = \frac{1}{1} = 1$$

$$\lim_{x \to 0} \frac{1}{x^3}$$ non esiste

$$\lim_{x \to 0} \frac{1}{x^2} = +\infty$$

$$\lim_{x \to -\infty} \frac{x^3 - 2x + 4}{x^2 - x + 1} = -\infty$$

$$\lim_{x \to -\infty} \frac{x^3 - 2x + 1}{5x^3 + 2x} = \frac{1}{5}$$

$$\lim_{x \to -\infty} \frac{3x^5 + x^4 - 1}{|x|^5 + 2x - 1} = -3$$

$$\lim_{x \to +\infty} \frac{3x^2 + \cos(x)}{2x^2 - \sin(x)} = \frac{3}{2}$$

$$\lim_{x \to 3^-} \frac{x^2 - 4x + 10}{x^2 - 5x + 6} = \lim_{x \to 3^-} \frac{x^2 - 4x + 10}{(x-2)(x-3)} = -\infty$$

Studiare $\quad f(x) = \dfrac{3x-1}{1+x^2}$

dominio: $\mathbb{R}$

limiti $\quad \lim_{x \to \pm\infty} f(x) = 0^{\pm}$

$$f'(x) = \frac{3\cdot(1+x^2) - 2x(3x-1)}{(1+x^2)^2} = \frac{3+3x^2-6x^2+2x}{(1+x^2)^2}$$

$$= \frac{3+2x-3x^2}{(1+x^2)^2}$$

$f'(x) = 0$ per $3x^2 - 2x - 3 = 0$

$$x_{1,2} = \frac{1 \pm \sqrt{1+9}}{3} = \frac{1 \pm \sqrt{10}}{3}$$

$1.3\ldots$

$-0.6\ldots$

f decresce su $(-\infty, x_2)$ e $(x_1, +\infty)$

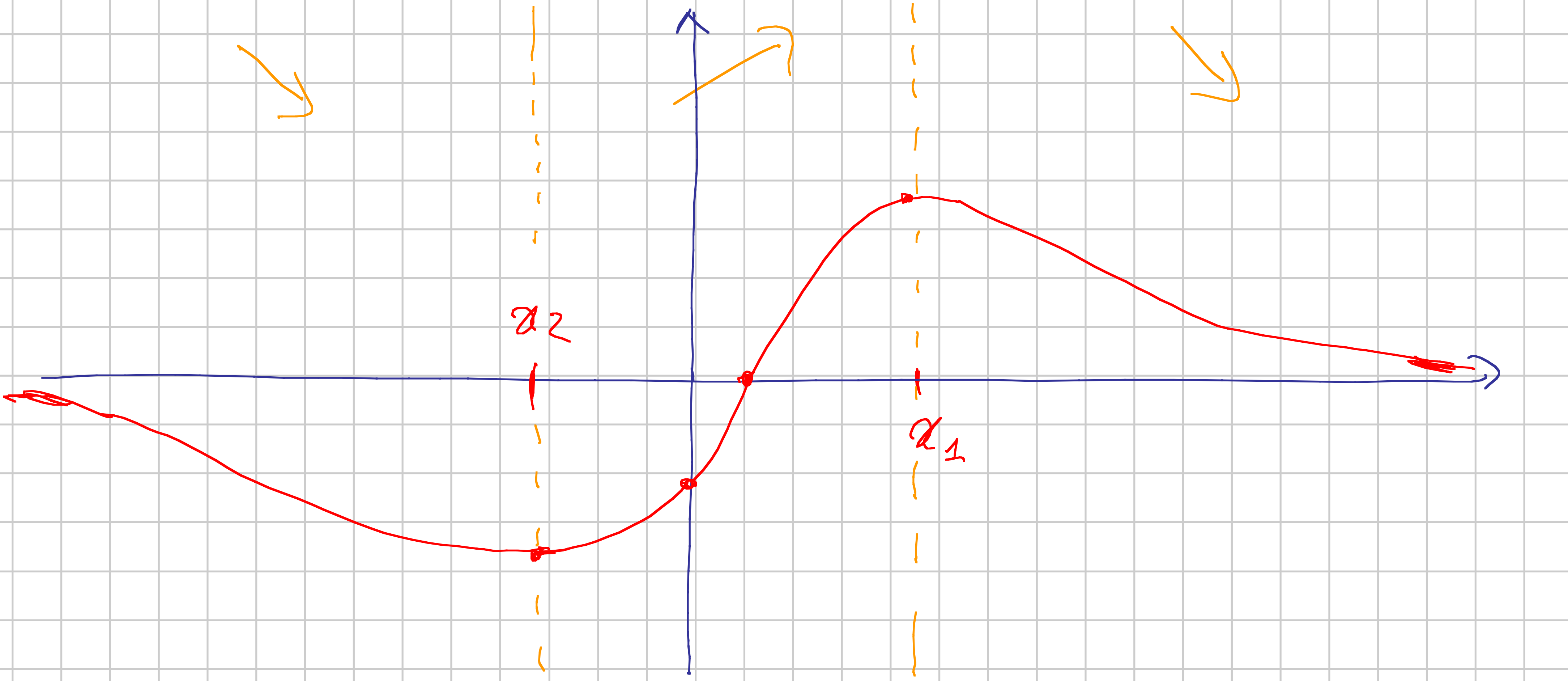
cresce su (x_2, x_1)

Intersez con assi : $f(0) = -1$

$f(x) = 0$ per $x = 1/3$

$f(x_1) = 1,2\ldots$

$f(x_2) = -2\ldots$



x2
x1

$$f(x) = \frac{x^2-4x-5}{x+3}$$

dominio: $x \neq -3$

limiti: $\lim_{x \to \pm\infty} f(x) = \pm\infty$ $\left(\lim_{x \to \pm\infty} \frac{f(x)}{x} = 1 \right)$

$$\lim_{x \to -3^{\pm}} f(x) = \pm\infty$$

$f(0) = -5/3 \qquad f(x) = 0$ per $x = -1$ e $x = 5$

$$f'(x) = \frac{(2x-4)(x+3) - 1\cdot(x^2-4x-5)}{(x+3)^2} =$$

$$= \frac{2x^2+6x-4x-12-x^2+4x+5}{(x+3)^2} = \frac{x^2+6x-7}{(x+3)^2}$$

$$= \frac{(x+7)(x-1)}{(x+3)^2}$$

$\Rightarrow f$ cresce su $(-\infty, -7)$ e $(1, +\infty)$

decresce su $(-7, +1)$