

# Corso di Statistica Sociale

---

CORSO DI LAUREA: SCIENZE DELL'EDUCAZIONE

DOCENTE: FRANCESCO SANTELLI

# Il concetto di *variabilità*

---

- ❑ La maggior parte delle variabili presenta nelle unità statistiche di un collettivo valori o attributi in tutto o in parte **diversi**.
- ❑ Si chiama **variabilità** (e si tratta di numeri, variabili quantitative) e **mutabilità** (nel caso qualitativo, cioè categorie), l'attitudine dei caratteri o fenomeni ad assumere modalità differenti.
- ❑ Noi ci concentreremo sulla variabilità e sui modi più diffusi di valutarla
- ❑ Fino ad ora ci siamo concentrati su indici di **posizione**: medie, mediana, quartili, moda ecc.
- ❑ Qui cercheremo di rispondere ad un'altra domanda: quanto la variabile si concentra (distribuisce) intorno ad un dato valore usato come riferimento?
- ❑ Intuitivamente, tanto più gli individui tendono a presentare sempre gli stessi valori, tanto più sarà bassa la variabilità...

# Come interpretare la variabilità?

---

1. Variabilità come misura di **imprecisione**. Esempi (soprattutto valutazioni/rilevazioni fisiche): temperatura rilevata da 10 diversi termometri (ma anche stesso termometro in istanti diversi) sullo stesso bambino, voto espresso da 1 a 30 da un Prof e il suo assistente sullo stesso candidato.
2. Variabilità come misura di **incertezza**. Esempi (fenomeni soggetti a variazioni non prevedibili): tempo impiegato dal Prof. per arrivare da Trieste a Portogruaro, numero di studenti che proveranno l'esame al primo appello ecc.
3. Variabilità come **concentrazione**. Esempi (soprattutto redditi e ricchezze): quanto capitale è concentrato nelle mani di poche persone particolarmente ricche? Quanta quota delle quote di una azienda è concentrata nelle mani del socio di maggioranza?
4. Variabilità come **variabilità** di un fenomeno. Esempi: come varia altezza tra studenti stessa classe, come varia il voto all'esame di Statistica, come varia lo stipendio degli educatori socio-pedagogici in diverse scuole private ecc.

# Come la valutiamo/calcoliamo?

---

Principalmente attraverso **indici di variabilità**.

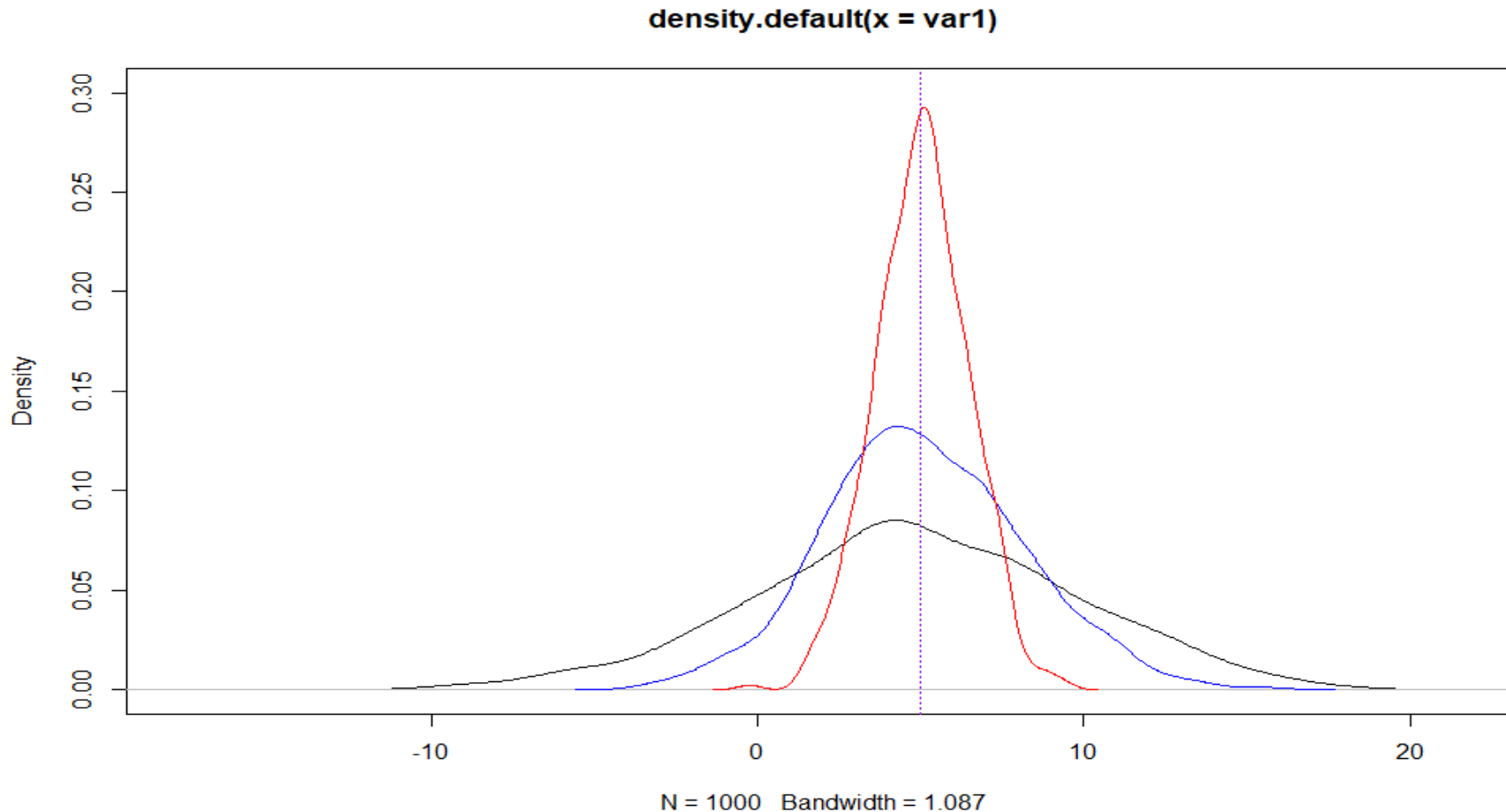
Stesso concetto degli indici che abbiamo visto fino ad ora per la tendenza centrale: cerchiamo di sintetizzare un aspetto dei dati con un solo numero (o pochi numeri)

Esistono molti indici di variabilità, ma presentano tutti due caratteristiche in comune:

- 1) Sono sempre maggiori o uguali a zero (non ha senso la variabilità negativa)
- 2) Valgono zero se e solo se la variabile presenta sempre una sola modalità per tutte le unità statistiche

Ma anche alcune rappresentazioni grafiche dicono qualcosa sulla variabilità

# Da un grafico cosa si evince?



---

Sono curve di densità  
(simili a istogrammi)

3 gruppi di persone  
Centrate sul valore 5 ma con  
Varianze piuttosto diverse

Quale presenta maggiore variabilità?

# Cosa aggiunge la variabilità?

---

Poniamo di avere tre studenti tutti con la stessa media, uguale a 24: Giorgio, Alfredo e Laura.

Tutti e 3 hanno fatto 4 esami. Voi riterreste le loro «carriere universitarie» equivalenti giusto?

Questi sono i voti da loro presi:

<u><b>Giorgio</b></u>	<u><b>Alfredo</b></u>	<u><b>Laura</b></u>
24	22	18
24	23	18
24	25	30
24	26	30

Medie uguali...ma moda?

Quale dei 3 studenti mostra minore variabilità?

Quale invece la maggiore variabilità?

Interpretando solo la media, li avremmo considerati **UGUALI**, MA IN REALTA' SONO STUDENTI CON PERCORSI DIVERSI

# L'indice più importante: la varianza

---

Si legge «sigma quadro» o «varianza»

E' il nostro caro amico sommatoria, da 1 a N  
(cioè sommiamo per tutti i dati)

E' la parte da calcolare: differenze di ogni numero dalla media e poi al quadrato.

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$$

E' come al solito N, la numerosità totale,  
Il numero di osservazioni (dati)

# Esercizio 1

---

<u>Giorgio</u>	<u>Alfredo</u>	<u>Laura</u>
24	22	18
24	23	18
24	25	30
24	26	30

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

- 1) Mettere tutti i voti insieme e calcolare la **varianza generale**
- 2) Calcolare **mediana** e **moda** di tutti i voti insieme
- 3) Calcolare la varianza di **ogni singolo studente**



# Dalla varianza allo scarto quadratico medio

---

Quando facciamo la varianza, abbiamo preso la distanza dalla media di ogni osservazione e l'abbiamo elevata al quadrato. Perché?

Proviamo a non fare i quadrati..

<b><i>Giorgio</i></b>	<b><i>Alfredo</i></b>	<b><i>Laura</i></b>
24	22	18
24	23	18
24	25	30
24	26	30

Prendiamo la media, facciamo ogni valore meno la media, e sommiamo tutto...

Risultato!?

Questo è il motivo per cui nella formula si calcolano gli scarti al quadrato, ed è per questo che si chiama  $\sigma^2$

$$S.Q.M = \sigma = \sqrt{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

# Come si interpretano?

---

Questi indici (varianza e scarto quadratico medio) come detto non possono mai essere negativi; tanto più sono grandi, tanto più il fenomeno è variabile

Tuttavia, la **varianza** utilizzando i quadrati ha una interpretazione ancora più difficile...

Anche se non scritta esplicitamente, la varianza ha anche l'unità di misura del fenomeno al quadrato!

Dire che la varianza delle temperature a Trieste nel mese di Ottobre è stata di 100 gradi centigradi ( $^{\circ}$ )<sup>2</sup> in realtà non ha alcun senso!

**Per questo si preferisce, solitamente, lo scarto quadratico medio sigma!**

Tuttavia, quando si devono fare confronti tra gruppi diversi, anche *S. Q. M.* può fallire..

# Oltre lo scarto quadratico medio

---

<b><i>Stipendi Lire</i></b>	<b><i>Stipendi Euro</i></b>
1000000	516
2000000	1033
3000000	1549

3 manovali edili percepiscono il 31 dicembre 2000  
Il loro stipendio in Lire, mentre il 1 gennaio 2001 in  
Euro.

Calcoliamo varianza e scarto quadratico medio

Che conclusioni traiamo!?

Si sono ridotte le distanze tra i manovali in termini di  
Stipendi?

**COSA PROPONETE DI FARE?**

# Il Coefficiente di Variazione

---

E' la deviazione standard sigma divisa la media. In questo modo sono possibili i confronti anche utilizzando *unità di misura diverse*!!

$$C.V. = \frac{\sigma_x}{\bar{x}}$$

Ovviamente non può essere negativo, ma assume valori abbastanza piccoli rispetto a varianza e deviazione standard

media	2000000	1032,913798
sigma	816496,5809	421,6852923
CV	0,40824829	0,40824829

# Esercizio 2

Ore Studio	Ore Notte
8	6
8	7
10	7
30	7
30	7
22	8

Calcolare:

- 1) media, moda e mediana di ore di studio settimanali e di ore di sonno
- 2) Calcolare varianza e scarti quadratici medi
- 3) Calcolare coefficienti di variazione
- 4) Era necessario o potevamo confrontarli anche senza?
- 5) Rappresentare uno graficamente con il boxplot e uno con ...