

# Statistica

negli esperimenti reali si effettuano sempre un numero *finito* di misure,

( spesso molto limitato )

➤ l'insieme delle misure effettuate costituisce il *campione* statistico

## Statistica descrittiva e statistica inferenziale o "predittiva"

- la statistica descrittiva studia i criteri di rilevazione ( campionamento ), di classificazione e di sintesi delle informazioni relative ad una popolazione oggetto di studio

si e' costretti a fare previsoni quando per qualche motivo non si possono conoscere i dati relativi a tutta la popolazione oggetto di studio, ma si hanno a disposizione solo le informazioni fornite da una porzione limitata della popolazione in esame ( campione )

altro esempio: le proiezioni preelettorali

- la statistica predittiva si prefigge di "***stimare***" caratteristiche della "popolazione" incognita basandosi sulle informazioni fornite da un campione limitato di misure ripetute

esempio:

**determinare la statura vera delle matricole iscritte ad Ingegneria a Bologna nel corrente anno accademico**

misurando le stature degli studenti che seguono questa lezione

e facendo la media aritmetica delle misure o la mediana o usando altri stimatori

suggeriti dai metodi della statistica descrittiva, si puo' pensare di poterla ottenere

- ma al massimo si avrebbe una stima della statura media di quegli studenti di Ingegneria Meccanica che sono presenti oggi in aula che di certo non e' quella di tutte le matricole iscritte ad Ingegneria a Bologna

ma esistera' una unica statura **vera** delle matricole iscritte  
ad ingegneria a Bologna ???

in realta' esiste una **distribuzione** di stature caratterizzata

da una densita' di probabilita'  $f(x)$  in generale incognita

a sua volta caratterizzabile tramite valor medio, varianza etc.,

parametri a loro volta incogniti che occorrera' quindi **stimare**

a partire dalle misure effettuate, dai "dati campionari"

si ipotizza che le stature possano essere considerate come una

variabile aleatoria rappresentabile da una densita' di probabilita'  $f_X(x | \theta_1, \theta_2, \dots)$

caratterizzata da una serie di parametri  $\theta_1, \theta_2, \dots, \theta_n$

popolazione "parente" →

insieme delle stature di tutte le matricole di Ingegneria-Bologna

distribuzione "parente" →

distribuzione delle stature di tutte le matricole di ingegneria

→ in generale la forma (ossia l'espressione analitica) della  $f_X(x | \theta_1, \theta_2, \dots, \theta_n)$  e' incognita

parametri della distribuzione "parente",  $\theta_1, \theta_2, \dots, \theta_n$  →

parametri, quali il valor medio  $\mu$ , la varianza  $\sigma^2$ , la *r.m.s.* etc., che caratterizzano univocamente la distribuzione

→ in generale i parametri  $\theta_1, \theta_2, \dots, \theta_n$  della  $f_X(x | \theta_1, \theta_2, \dots, \theta_n)$  sono incogniti

popolazione "campionaria" ( **campione** ) →

stature delle sole matricole del corso di Ingegneria Meccanica presenti oggi in aula

**campione**  $\rightarrow$  insieme delle ***misure*** della statura di  $n$  studenti scelti a caso

$$X_1 = x_1, \quad X_2 = x_2, \quad \dots \dots \dots \quad X_n = x_n$$

attraverso la conoscenza del campione si cerca di **stimare** la realta'

sempre che l'ipotesi che il fenomeno sia aleatorio, sia valida

per stimare il valor medio di solito si fa uso dello “stimatore”

## Media aritmetica

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

la media aritmetica **non** e' l'unico stimatore possibile del valor medio

altri stimatori di centralita' sono la

## Media geometrica

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}}$$

ma con la limitazione che tutti gli  $x_i$  devono essere positivi

significato geometrico: la media geometrica di due numeri è la lunghezza del lato di un quadrato equivalente ( = stessa area ) ad un rettangolo che abbia i lati pari al modulo dei due numeri

# Media armonica

$$\bar{x}_a^{-1} = \frac{1}{n} \left( \frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right)$$

ma con la limitazione che tutti gli  $x_i$  non devono essere nulli

una media armonica darà una media significativa quando si misurano grandezze che si sommano come reciproci in una sequenza

ad esempio, la media armonica dei condensatori in serie rappresenta la capacità che un singolo condensatore avrebbe se venisse usato un solo condensatore invece di un insieme di condensatori in serie

Analogo discorso per le resistenze elettriche collegate in parallelo

altro esempio:

se si desidera la velocità media di un'auto che percorre la stessa distanza (non il tempo!) con varie velocità, l'effetto netto di tutta la guida (il tempo totale impiegato) viene rilevato dividendo la distanza comune per ciascuna velocità per ottenere il tempo per quella parte del viaggio, e poi sommando i tempi.

La velocità costante che richiederebbe lo stesso tempo totale per l'intero viaggio è la media armonica delle velocità

Un'auto percorre all'andata la distanza di 60 km alla velocità di 60 km/h, mentre, a causa del traffico intenso, percorre il ritorno sulla stessa strada a velocità di 30 km/h.

Determinare in  $\text{Km h}^{-1}$  la velocità media dell'auto.

per definizione la velocità media è data da: tutto lo spazio percorso diviso il tempo totale

impiegato a percorrerlo  $\rightarrow$  lo spazio complessivo percorso è  $60 + 60 = 120 \text{ Km}$

mentre il tempo complessivamente impiegato per l'andata e il ritorno è

$$60\left(\frac{1}{60}\right) + 60\left(\frac{1}{30}\right) = 3 \text{ h} \quad \rightarrow \text{ la velocità media è } 120/3 = \mathbf{40 \text{ Km/h}}$$

➤ la velocità media è la media armonica delle velocità

Nota Bene: facendo la media aritmetica delle velocità si avrebbe  $(30 + 60)/2 = 45 \text{ km/h}$  che è un risultato chiaramente scorretto, perché se le ore complessive per andata e ritorno sono 3 alla velocità di 45 km/h l'auto avrebbe percorso  $3 \cdot 45 = 135 \text{ km}$  e non 120 come invece è specificato

## Media quadratica

$$\bar{x}_q = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

in generale la media quadratica e' sempre maggiore della media aritmetica

infine, come indicatori di centralita' di una distribuzione, si possono usare

anche la ***mediana campionaria*** (= 50-esimo percentile ) e la

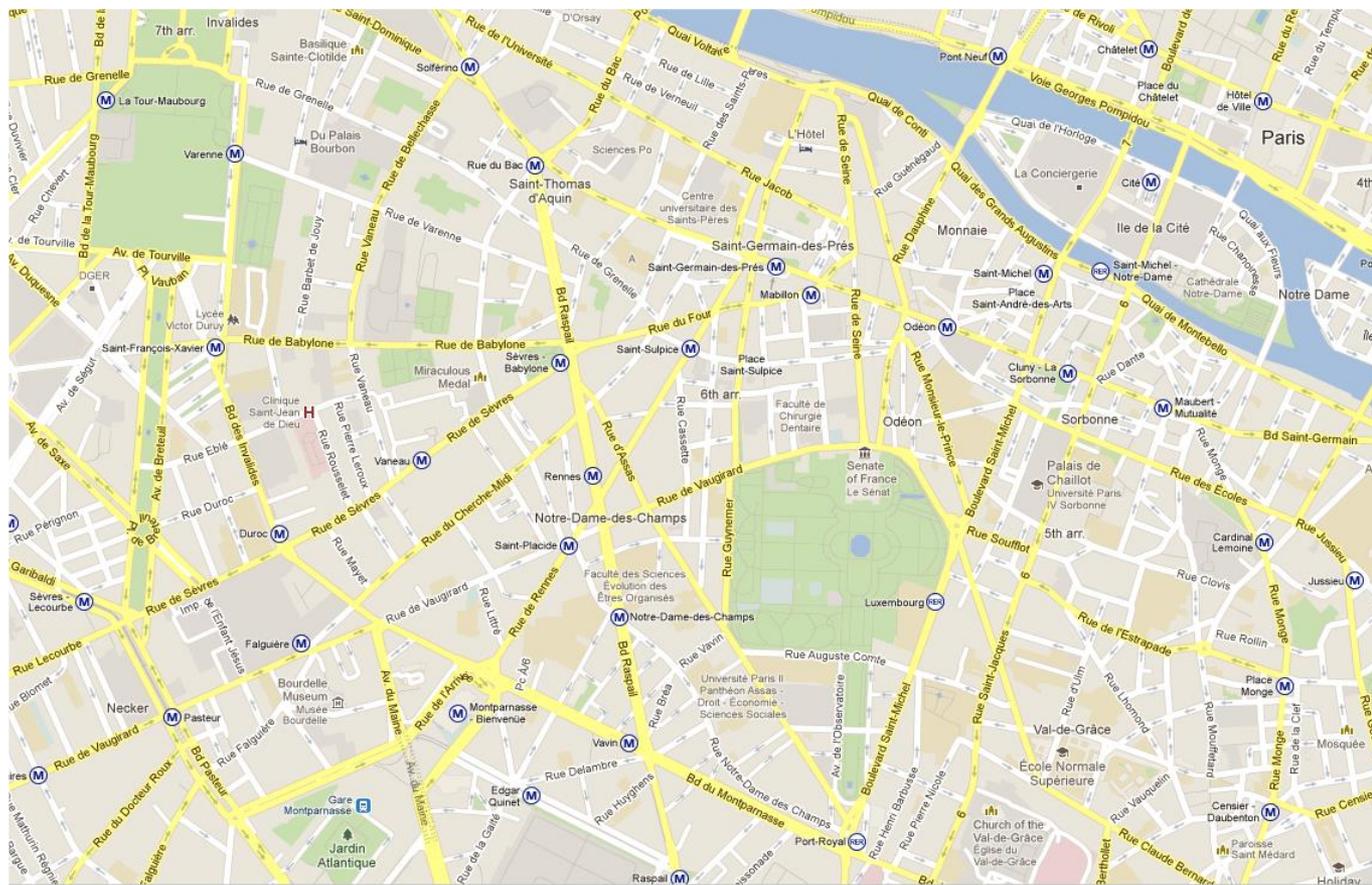
***moda campionaria*** ( = valore piu' probabile)

# Scelta del miglior stimatore di un parametro incognito di una distribuzione aleatoria

es. : media aritmetica o mediana campionaria ?

supponiamo che un'agenzia di viaggi vi possa fornire informazioni sui prezzi degli hotel in alcune delle vie più caratteristiche del quartiere da voi prescelto per trascorrere un week end a Parigi

## Quartiere Latino



Bd de Saint-Germain      Bd Saint-Michel      Bd de Montparnasse

**preferireste conoscere il prezzo medio degli hotels in queste strade o il prezzo mediano ?**

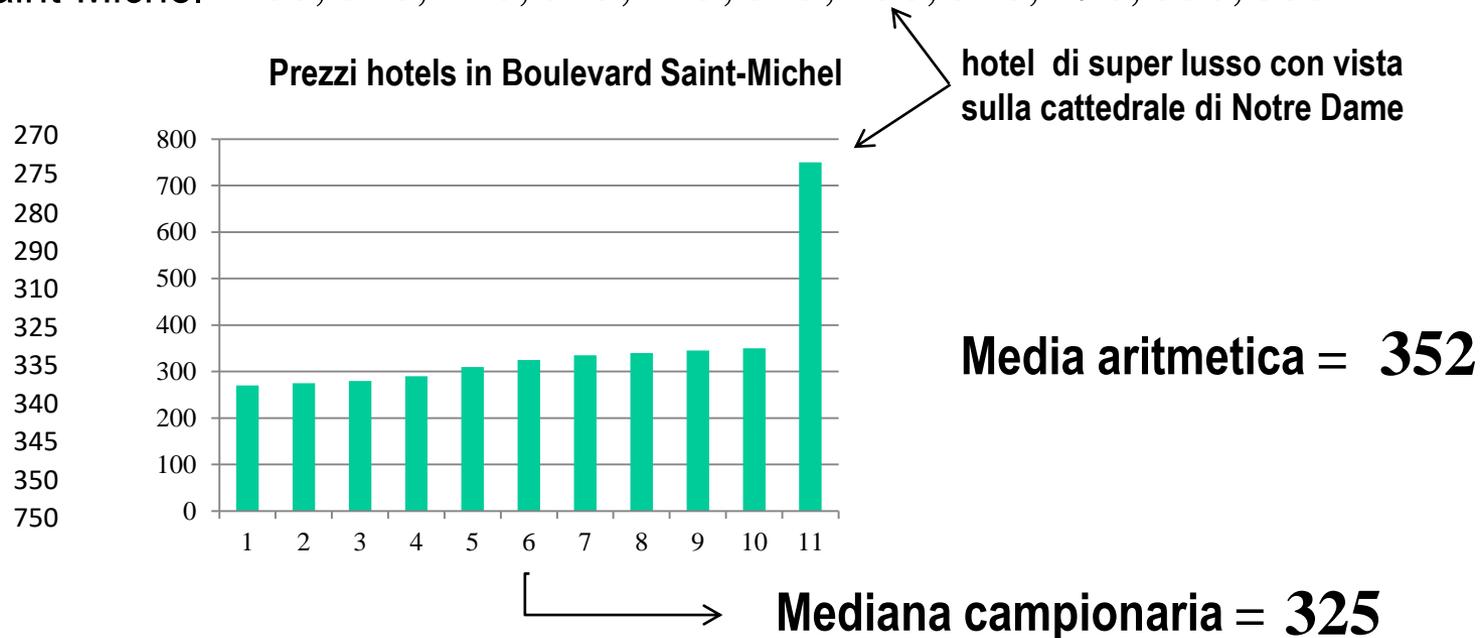
la scelta tra media aritmetica e mediana campionaria dipende dalle “condizioni al contorno”  
 supponiamo si abbia un tetto massimo di spesa per i pernottamenti, un budget, di **330 Euro**  
 e supponiamo non si abbia tempo a disposizione per verificare i costi dei singoli hotels

	Bd de Saint-Germain	Bd Saint-Michel	Bd de Montparnasse	
<b>Media aritmetica</b>	<b>356</b>	<b>352</b>	<b>361</b>	<b>&gt; 350 Euro</b>

Conclusioni : niente week end a Parigi !

ma potrebbe essersi verificata una situazione come questa :

Bd Saint-Michel 280, 310, 270, 345, 275, 325, 750, 340, 290, 350, 335



in questo caso sarebbe convenuto conoscere la *mediana* dei prezzi così da essere certi che esattamente la metà degli hotel di una determinata strada avesse prezzi inferiori alla mediana

la differenza tra la scelta della media o della mediana sarebbe stata irrilevante se si avesse avuto a disposizione il tempo per controllare il prezzo di ciascun hotel, per es. su internet o telefonando direttamente all'hotel ( da cui la: "*time is money*" )

e ovviamente la cosa non avrebbe avuto alcuna rilevanza se si avesse avuto a disposizione un budget illimitato

in conclusione: non esiste il miglior stimatore *in assoluto* esistono pero' **criteri** con i quali " costruire " gli stimatori in modo che possiedano proprieta' " desiderabili "

( breve commento sulle proprieta' degli stimatori )

la ***media aritmetica*** e' lo stimatore piu' usato per la stima del valor medio  $\mu$  della distribuzione "parente" incognita per le sue proprieta' di ***correttezza***, ***consistenza*** ed ***efficienza***

**Backup slides**