

Errori sistematici e casuali



Errori Casuali

Tempo di reazione nel far partire o fermare l'orologio:

Può essere sia in eccesso che in difetto (ad esempio partenza e arrivo), quindi l'errore può avere segno positivo o negativo

Errori sistematici

Vanno sempre nella stessa direzione. Possono essere legati allo strumento. Ad esempio il cronometro può misurare sempre un tempo con un certo errore percentuale

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Errori sistematici e casuali

Ripetizione delle misure

Nel caso in cui stiamo misurando qualcosa che dovrebbe dare lo stesso valore (beh, certamente non stiamo parlando di una prestazione sportiva, ma ad esempio di una lunghezza di un percorso) allora possiamo avere una idea sulla grandezza dell'errore sperimentale

Ad esempio, da una serie di misure otteniamo questi risultati (in metri)

3.01 3.03 3.00 3.05 3.06 3.04

Possiamo dire ad esempio che la media è una buona stima della misura

E che probabilmente il valore vero sta tra 3.01 e 3.06

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Errori sistematici e casuali

Stima migliore \pm errore

Se effettuiamo diverse misure possiamo dire che la stima migliore sia la media (non è sempre così)

Ad esempio

2.1 2.3

Potremmo dire che la stima migliore sia 2.2

e indicare in qualche modo l'errore

2.2 ± 0.1

Quindi il nostro valore vero (ignoto ... sempre ignoto) sarà

“probabilmente” tra

$2.2 - 0.1 = 2.1$ e $2.2 + 0.1 = 2.3$

Vedremo più avanti qual è il modo più corretto di stimare l'errore, nel caso ad esempio di errori casuali gaussiani

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Errori sistematici e casuali

Stima migliore \pm errore

Vedremo più avanti qual è il modo più corretto di stimare l'errore, nel caso ad esempio di errori casuali gaussiani

Inoltre, l'intervallo nel quale è compresa la misura vera (ignota...sempre ignota) è un'intervallo possibile, ma non possiamo dire con assoluta certezza che la misura ricada al suo interno

Vedremo più avanti che nell'ipotesi di dati gaussiani sarà possibile

- arrivare ad una definizione migliore dell'errore
- stimare la probabilità che il valore vero (ignoto...sempr..) si trovi all'interno di un dato intervallo

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Cifre significative

Gli errori dovrebbero essere arrotondati ad una cifra significativa

Misuriamo una velocità

$$10.44 \pm 0.01285838 \text{ m/s}$$

Non ha senso portarsi dietro tutte le cifre ma si deve arrotondare (ad esempio alla seconda cifra decimale)

$$10.44 \pm 0.01 \text{ m/s}$$

Attenzione, nel caso in cui abbiamo un errore pari a 0.14 allora approssimare a 0.1 può portare ad una approssimazione eccessiva (riduzione del 40%)

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Cifre significative

Altro caso

$$120.01 \pm 2 \text{ km/h}$$

In questo caso non ha senso riportare le cifre significative nella stima ma è meglio scrivere

$$120 \pm 2 \text{ km/h}$$

L'ultima cifra significativa della stima di una misura dovrebbe essere dello stesso ordine di grandezza dell'errore

N.B.

Indicare $120.0 \pm 2.0 \text{ km/h}$ può avere un senso, visto che indica che fornisce una informazione più precisa sulla grandezza dell'incertezza rispetto al caso precedente

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Propagazione dell'errore

Misure Indirette

Misura tramite una relazione matematica tra altre misure in gioco

(ad esempio stima della velocità come rapporto tra spazio percorso e tempo impiegato)

In questo caso l'errore in ciascuna delle misure in gioco (ad esempio quella relativa alla misura dello spazio e quella relativa al tempo) si combinano

- 1) Si devono stimare gli errori che si commettono nelle singole misure
- 2) In funzione della relazione tra le grandezze, stimare come l'errore si propaga

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Propagazione dell'errore

Esempio

Grandezza relativa alla somma di due variabili, x e y. Dalle misure si ottiene che

$$\begin{aligned}x_{migliore} &\pm \delta x \\ y_{migliore} &\pm \delta y\end{aligned}$$

In questo caso il valore più alto è

$$x_{migliore} + y_{migliore} + \delta x + \delta y$$

il valore più basso

$$x_{migliore} + y_{migliore} - \delta x - \delta y$$

Quindi possiamo individuare l'intervallo probabile per la somma

$$x_{migliore} + y_{migliore} \pm (\delta x + \delta y)$$

Possono essere trovate formule per prodotti, quozienti, e funzioni generiche di una variabile

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Propagazione dell'errore

Un altro approccio è quello, anche in misure indirette, di fare diverse prove e valutare la dispersione dei valori

Quindi anche in questo caso utilizzare un approccio statistico per trattare i valori finali, senza studiare la propagazione dell'errore

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Errore relativo

L'errore relativo fornisce una indicazione della bontà di una misura indipendentemente dalla dimensione della grandezza misurata

Si esprimono in valori percentuali

$$x_{misurato} = x_{migliore} \pm \delta x$$

Errore relativo $\frac{\delta x}{|x_{migliore}|}$

Nel caso di 120 ± 2 km/h

L'errore relativo è pari a $\frac{2}{|120|} \cong 1.7\%$

Errori inferiori allo 1% sono molto difficili da ottenere

Errori del 10% sono indice di misure non di alta qualità

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Media e Deviazione Standard

Dato un insieme di N misure $\{x_1, x_2, \dots, x_N\}$ supponendo di avere minimizzato le sorgenti di errore sistematico

Allora l'errore casuale può essere individuato ripetendo la misura

Si dice media aritmetica semplice di N numeri il numero che si ottiene dividendo la loro somma per N.

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

Questa è la migliore stima della grandezza x

UNIPI-SM-Caratteristiche della strumentazione biomedicale

Media e Deviazione Standard

La differenza tra la media e la i-esima misura si chiama deviazione o residuo

$$d_i = x_i - \bar{x}$$

Per avere una stima dell'attendibilità delle misure non si possono mediare i d_i (fornirebbe un valore nullo)

Ma si usa la seguente grandezza, detta deviazione standard

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i)^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Il quadrato si chiama varianza

UNIPI-SM-Caratteristiche della strumentazione biomedicale

Deviazione Standard e incertezza

Se le misure sono distribuite come una gaussiana allora possiamo usare la deviazione standard come una misura dell'incertezza

In questo caso possiamo determinare l'intervallo

$$\bar{x} \pm \sigma_x$$

E associare a questo intervallo una probabilità: in questo intervallo è presente il valore vero con una probabilità circa del 70%

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Rigetto di Dati

Queste valutazioni possono essere utili nel caso di misure che si discostano dalle altre, come in questo caso

2.2.2.5.2.1 2.6.2.7.2.3.5

È una fase molto critica e lo sperimentatore deve valutare se esistono delle cause che possono spiegare questa differenza

Nel caso non si trovi, la decisione è lasciata allo sperimentatore che può decidere di scartare o meno i dati

Un modo è quello di confrontare il valore sospetto con il valore medio e verificare quanto sia distante, in relazione alla deviazione standard delle misure

$$\bar{x} \pm \sigma_x$$

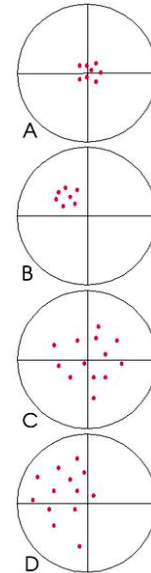
E associare a questo intervallo una probabilità in questo intervallo è

UNIFI-SM-Caratteristiche della strumentazione biomedicale

Accuratezza e precisione

Facciamo un test chiedendo a 4 persone di lanciare un certo numero di freccette verso un bersaglio.

Supponiamo che i risultati siano quelli rappresentati nelle figure A, B, C e D.



UNIFI-SM-Caratteristiche della strumentazione biomedicale

Accuratezza

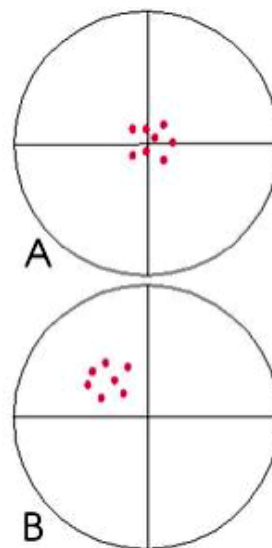
Prendiamo il caso A e B.

Nel caso A vediamo che le freccette sono attorno al centro. Nell'insieme dei nostri lanci le freccette stanno tutte attorno al centro del bersaglio, cioè attorno al **centro vero** che volevo colpire.

Nel caso B invece vediamo che nell'insieme dei nostri lanci le freccette hanno colpito il bersaglio lontano dal centro, cioè mediamente lontano dal **centro vero**.

Potremo dire che l'insieme di lanci del **caso A e' piu' accurato del caso B** intendendo che "mediamente" le freccette A stanno attorno al centro vero mentre le freccette B stanno tutte lontane dal centro.

Da notare che non ho considerato la dispersione delle freccette sul bersaglio.



UNIFI-SM-Caratteristiche della strumentazione biomedicale

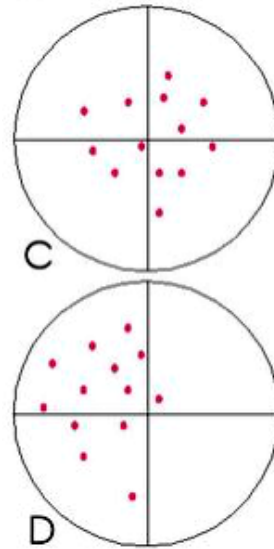
Accuratezza

Consideriamo ora il caso C e D.

Nel caso C, come in A, le freccette cadono “mediamente” attorno al **centro vero**; questa volta però i lanci colpiscono il bersaglio con una maggiore dispersione.

Anche nel caso D, come in B, le freccette hanno colpito il bersaglio “mediamente” lontano dal **centro vero**, e anche qui con maggiore dispersione.

Facendo riferimento a dove cadono le freccette rispetto al **vero centro**, come nel caso precedente possiamo ancora dire che il **caso C è più accurato del caso D**.



UNIFI-SM- Caratteristiche della strumentazione biomedicale

Precisione

Riguardando i 4 casi si vede comunque che sono molto diversi tra di loro, anche se ho potuto dire che il giocatore A è più accurato di B, ed il giocatore C più accurato di D.

In più potrei dire che, considerando la distanza “media” dei lanci di ciascun giocatore dal centro vero, il giocatore A è accurato quanto il giocatore C, ed il giocatore B quanto D.

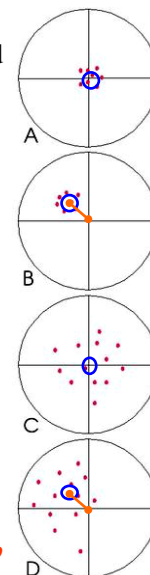
Prendiamo ora in considerazione, per ogni giocatore, la **“dispersione” dei lanci rispetto alla loro “media”**.

Con riferimento alle figure, qualitativamente si può dire che la dispersione è piccola nei casi A e B, ed è più grande nei casi C e D.

Questa **dispersione è strettamente legata al concetto di Precisione**, tanti più piccola è la dispersione tanto maggiore sarà la precisione.

Si può quindi affermare che i giocatori A e B sono più precisi dei giocatori C e D.

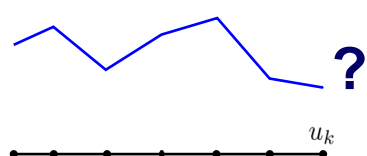
Da notare che nella definizione di Precisione non viene considerato il fatto che i lanci siano più o meno vicini al centro vero.



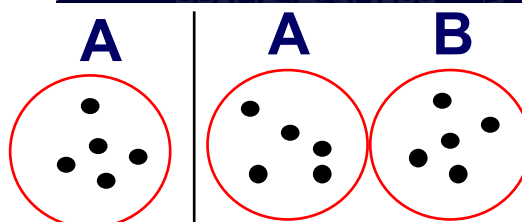
UNIFI-SM- Caratteristiche della strumentazione biomedicale

Statistica Descrittiva ed Inferenziale

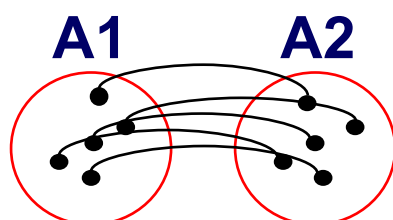
Why Statistics?



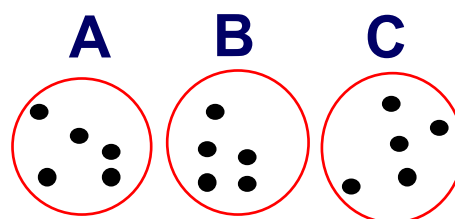
Description and Prediction



Samples Analysis

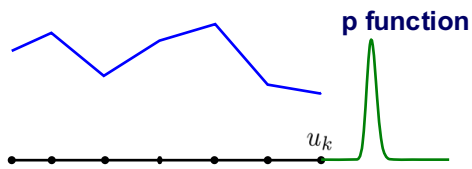


Paired Samples Analysis

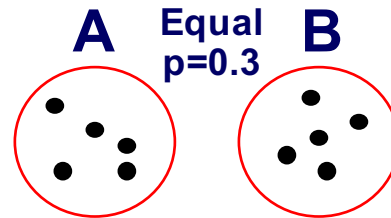


MultiSamples Analysis

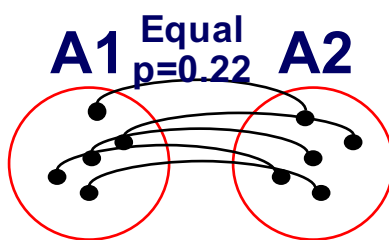
Why Statistics?



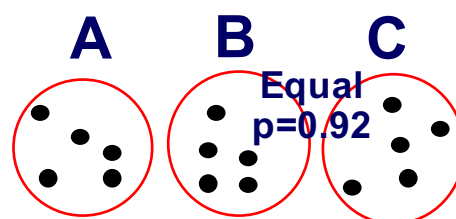
Description and Prediction



Samples Analysis



Samples Analysis



MultiSamples Analysis

Introduzione (1)

Di cosa si occupa la statistica?

Oggetto della statistica sono **fenomeni collettivi** che presentano carattere di **variabilità**

Per fenomeno collettivo si intende un fenomeno che riguarda una grande collezione di elementi = **POPOLAZIONE**

Elementi della popolazione = **UNITÀ STATISTICHE**

Introduzione (2)

La popolazione è troppo vasta per poter essere studiata nella sua globalità → dalla popolazione viene estratto un **CAMPIONE** di n elementi

Sul campione vengono rilevate/misurate alcune caratteristiche. I risultati di questa misura costituiscono i **DATI**

La statistica permette di trarre conclusioni sull'intera popolazione a partire dai dati ottenuti sul campione

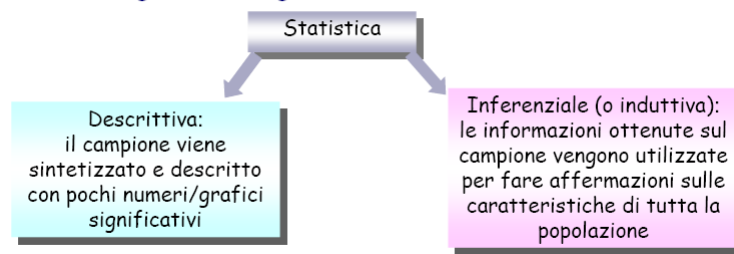
A causa della casistica ridotta non possiamo essere certi delle nostre conclusioni → si specifica il grado di certezza in termini di probabilità

Statistica Descrittiva e Statistica Inferenziale (1)

Scopo principale della statistica consiste nel compiere un'inferenza circa l'intera popolazione a partire dal campione

Per fare questo, per prima cosa, bisogna descrivere e sintetizzare i dati, con pochi numeri o grafici significativi

Si distinguono due grandi rami della statistica:



Statistica Descrittiva e Statistica Inferenziale (2)

Nelle applicazioni, Statistica Descrittiva e Statistica Inferenziale non possono essere completamente separate.

Infatti i problemi di inferenza statistica vengono affrontati secondo il seguente schema:



Statistica Descrittiva di un **CAMPIONE** (**SAMPLE**) di dati

Statistica Descrittiva

Scopo: descrivere il campione (dati) in modo sintetico ed efficace mediante tabelle, grafici, numeri

Premessa: Le caratteristiche che osserviamo sul campione variano da un'unità di osservazione all'altra → variabili

Le variabili possono essere **discrete** o **continue**

Variabili discrete: assumono un numero finito o un'infinità numerabile di valori

Variabili continue: possono assumere qualsiasi valore

Tabelle e Grafici di Frequenza (1)

Un primo utile sistema per riassumere i dati è la costruzione di tabelle e grafici di frequenza

Esempio nel discreto: lancio di un dado

Il risultato del lancio è una variabile discreta (può assumere uno dei seguenti valori 1 2 3 4 5 6)

50 lanci → ottengo una sequenza di 50 (n) numeri

Sintetizzo i dati costruendo una tabella

Tabelle e Grafici di Frequenza (2)

Esempio nel discreto (continua)

risultato del lancio	frequenza (f)	frequenza relativa (f/n)
1	9	0.18
2	12	0.24
3	6	0.12
4	8	0.16
5	10	0.20
6	5	0.10

$$\sum f = n \qquad \sum (f/n) = 1.00$$

prima colonna: possibili risultati del lancio

seconda colonna: numero totale di volte in cui è stato ottenuto quel risultato (**frequenza assoluta**)

terza colonna: frequenza assoluta del risultato divisa per il numero totale (n) di osservazioni (**frequenza relativa**)