Errori di misura:

di solito vengono effettuate molte misure *indipendenti* della stessa grandezza,

fino a collezionare un numeroso campione di misure

ma a causa degli errori di misura il risultato varia sensibilmente da misura a misura

si definisce **errore = | valore misurato – valore vero |**

cause di errore:

- → Limiti strumentali
- → Metodi di misura errati
- → Cause accidentali

categorizzazione degli errori:

→ **sistematici**

misura di un intervallo di tempo usando un orologio che va troppo lento, o troppo veloce.

misura della lunghezza di un oggetto non in modo perpendicolare all'oggetto (errore di parallasse)

→ **casuali o "statistici"**

incertezze dovute a cause accidentali e alla limitatezza del campione di misure

gli errori statistici sono riducibili aumentando il numero di misure indipendenti della stessa grandezza → aumentando la dimensione del campione

ma , ammesso che esista, quale e' il " **valore vero** "

potrebbe anche non esistere se alla base del fenomeno in esame vi fosse in

gioco una variabile aleatoria in questo caso sarebbe meglio parlare di errore

come di **errore = | valore misurato – valore medio |**

ma quanto vale il valor medio ? per saperlo con certezza si dovrebbero fare

una infinita' di misure ripetute e se si ha un numero finito di misure si puo'

solo tentare di " **stimarlo** " con il minimo margine di errore possibile

si assume come ***stima*** del valor medio (vero ?) μ della grandezza in esame

la media aritmetica dei risultati ottenuti nelle varie misure

- la cosiddetta **media aritmetica campionaria**

es. : si siano effettuate n misurazioni della stessa grandezza fisica $x_1, x_2 \ldots x_n$

la media aritmetica delle n misure e' :

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

la media aritmetica stima il valor "vero" μ , ma con un certo errore ε

$$\mu = (\overline{x} \pm \varepsilon) \quad \textit{unita' di misura}$$

problema : come stimare l'errore statistico ε ?

come indicatore di dispersione di una distribuzione intorno al suo valor medio

si usa la **deviazione standard campionaria** o " errore quadratico medio",

in inglese "root mean square" o *rms* che e' definito come :

$$dev.\ stand.\ camp. \equiv \sigma_{\bar{x}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Nota Bene: $\sum_{i=1}^{n}(x_i - \boxed{\bar{x}})^2$ al posto di $\sum_{i=1}^{n}(x_i - \boxed{\mu})^2$

commento sull'uso di $\bar{x}$ o di μ

Nota Bene: $\boxed{\frac{1}{n-1}}\sum_{i=1}^{n}(x_i - \bar{x})^2$ al posto di $\boxed{\frac{1}{n}}\sum_{i=1}^{n}(x_i - \bar{x})^2$

commento sull'uso di n o di $n-1$

una tra le proprieta’ piu’ importanti della media aritmetica e’ che

l’ **errore statistico** della media aritmetica stessa e’ dato da:

$$\varepsilon(\overline{x}) = \frac{dev.\ stand.\ camp.}{\sqrt{n}} \equiv \frac{\sigma_{\overline{x}}}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Intervallo di confidenza

si sono effettuate n misurazioni indipendenti della stessa grandezza fisica x,

ottenendo : $x_1, x_2 \ldots x_n$ la miglior stima del valor medio incognito

e' la media aritmetica $\overline{x}$ delle n misure dove $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

ma ripetendo una seconda volta n misure della stessa grandezza fisica,

si potrebbe ottenere un diverso insieme di risultati : $x'_1, x'_2 \ldots x'_n$

la miglior stima campionaria del valor medio incognito

continuerebbe ad essere la media aritmetica $\overline{x}'$ dove $\overline{x}' = \frac{1}{n}\sum_{i=1}^{n} x'_i$

ma essendo gli x_i ' diversi dagli x_i la nuova media aritmetica sarebbe diversa

$$\overline{x}' \neq \overline{x}$$

dunque anche la media aritmetica campionaria varia, ***imprevedibilmente***,

da campione a campione ossia e' essa stessa una ***variabile aleatoria***

- come stima della dispersione della v.a. "media aritmetica campionaria"

  si assume la deviazione standard campionaria $\sigma_{\overline{x}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$

- come errore sulla media aritmetica campionaria, si assume la

  deviazione standard campionaria **divisa per la radice di *n***

$$\varepsilon(\overline{x}) = \frac{\sigma_{\overline{x}}}{\sqrt{n}}$$

rilevanza della distribuzione gaussiana

- se la distribuzione di una serie di n variabili aleatorie x_i con $i = 1, n$ segue la forma funzionale gaussiana si puo' dimostrare che la media aritmetica delle n v. a. sara' distribuita in modo gaussiano

- anche se la distribuzione delle n v. a. x_i non fosse gaussiana se $n >> 1$ grazie al teorema del limite centrale, si puo' assumere che la distibuzione della media campionaria sia gaussiana

infine

- istogrammando i risultati di molte misure ripetute ed **indipendenti** tra loro risulta, molto spesso ma non sempre, che le misure si distribuiscono in modo gaussiano

dunque nella maggior parte dei casi, ma non sempre,

- se si stima il valor medio (vero) come

$$\mu = \overline{x} \pm \frac{dev.\ st.(\overline{x})}{\sqrt{n}} \quad \rightarrow \quad \mu = \overline{x} \pm \frac{\sigma_{\overline{x}}}{\sqrt{n}}$$

si ha il 68% di **probabilita'** di fare una stima esatta

- se si stima il valor medio (vero) come

$$\mu = \overline{x} \pm 2\frac{dev.\ st.(\overline{x})}{\sqrt{n}} \quad \rightarrow \quad \mu = \overline{x} \pm \frac{2\sigma_{\overline{x}}}{\sqrt{n}}$$

si ha il 95% di **probabilita'** di fare una stima esatta

- se si stima il valor medio (vero) come

$$\mu = \overline{x} \pm 3\frac{dev.\ st.(\overline{x})}{\sqrt{n}} \quad \rightarrow \quad \mu = \overline{x} \pm \frac{3\sigma_{\overline{x}}}{\sqrt{n}}$$

si ha il 99.7% di **probabilita'** di fare una stima esatta

il valore della percentuale di probabilita’ che si desidera ottenere,

ossia la attendibilita’ (affidabilita’, credibilita’, grado di fiducia …)

della stima del valor “vero” che si desidera raggiungere,

e’ detto “ **livello di confidenza** ”

supponiamo siano state molte fatte misure di precisione ed indipendenti tra loro,

di una stessa grandezza fisica, e' ragionevole attendersi che i risultati

non si riproducano perfettamente

postulando che il fenomeno in esame abbia un carattere aleatorio

ossia che l'esito della misura sia descrivibile in termini di una variabile aleatoria

si potranno applicare i metodi statistici per il trattamento dei dati sperimentali

vale a dire per stimare i parametri della variabile aleatoria incognita

in conclusione:

fare una misura sperimentale e' assimilabile all'effettuare un esperimento aleatorio

il risultato ottenuto in una misura equivale al verificarsi di uno tra i tanti possibili risultati che una opportuna v.a. puo' assumere

Vantaggio : questa impostazione permette di riconciliare la non riproducibilita' degli esperimenti con i principi galileiani

Prezzo da pagare : si deve rinunciare al determinismo della fisica classica

la statistica predittiva, utilizzando i risultati rigorosi della teoria della probabilita',
e' in grado di suggerire:

- quale sia caso per caso il miglior stimatore possibile del valor medio,
  o valor "vero", che di solito, ma non sempre, e' la media aritmetica campionaria

- quale sia il margine di errore con cui si puo' fare la stima in funzione della
  numerosita' del campione, ossia di determinare quale sia l'errore
  sulla media aritmetica

- come valutare quale sia l'attendibilita' di questa misura in termini di probabilita',
  ossia come determinare quale sia il livello di confidenza della stima

Nota bene:

e' chiaro che si tratta di **previsioni** e che per la aleatorieta' stessa del fenomeno

in esame e' **impossibile** prevedere esattamente quale risultato

si presentera' all' atto di ogni singola prova (misura)

Esempio

sono state fatte 25 misure ripetute ed indipendenti tra loro della stessa grandezza fisica, ad es. il peso di un oggetto misurato con una bilancia di precisione

i risultati, in *grammi*, sono :

1.720, 1.651, 1.811, 1.722, 1.721, 1.670, 1.710,
1.721, 1.740, 1.701, 1.730, 1.700, 1.761, 1.721,
1.751, 1.712, 1.710, 1.721, 1.691, 1.790, 1.740,
1.730, 1.760, 1.731, 1.712

e' evidente che i risultati della misura non si riproducono perfettamente

non avendo altre informazioni a disposizione si dovra' stimare il valor medio,

del peso ossia il peso medio impropriamente detto peso "vero"

della grandezza incognita, usando soli i dati del campione di misure effettuate

→ per stimare il valor medio calcoliamo la media campionaria

→ per stimare l'errore associato a questa misura utilizziamo l' errore sulla media

se x_i e' la i-esima misura $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ⇨ $\bar{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 1.7244$

$$dev.\ st. = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{25-1}\sum_{i=1}^{25}(x_i - 1.7244)^2} = 0.0337$$

l'errore sulla media aritmetica e' $\varepsilon(\bar{x}) = \frac{dev.\ st.}{\sqrt{n}} = \frac{0.0337}{\sqrt{25}} = 0.0074$

arrotondando l'errore ad una sola cifra $\varepsilon = 0.007$

se il livello di confidenza prescelto e' il 68 % il risultato della misura e' :

$$\mu = (1.724 \pm 0.007)\ gm$$

oppure $\mu = (1.72 \pm 0.01)\ gm$ al 95% di livello di confidenza

oppure $\mu = (1.72 \pm 0.02)\ gm$ al 99% di livello di confidenza

da notare la relazione tra la precisione e il grado di fiducia, (livello di confidenza)

a parita' di numerosita' del campione, → a parita' di n, se una cresce l'altra cala

Livello di confidenza	Errore assoluto	Errore relativo
68%	0.007	0.41 %
95%	0.01	0.58 %
99%	0.02	1.16 %

⇒ misura molto precisa ma ci credo poco…

⇒ misura poco precisa ma ci credo molto

Nota sulla presentazione del risultato

se per presentare il risultato ottenuto operando con il 68% di livello di confidenza si utilizzasse la convenzione delle cifre significative

in questo particolare caso il risultato andrebbe presentato come m = 1.72 gm

in effetti $\mu = (1.724 \pm 0.007)\ gm$

significa che, secondo la nostra stima, μ e' compreso tra 1.717 e 1.731

dunque la seconda cifra del risultato, il 2 , non e' una cifra certa

Nota Bene:

se l'errore sulla media fosse risultato minore di 0.004

avremmo dovuto scrivere m = 1.724 gm

infine se avessimo operato al 95 e 99 % di livello di confidenza

avremmo dovuto presentare il risultato come m = 1.72 gm

per farsi un idea della forma della distribuzione dei dati produciamo un istogramma :

1) disponiamo le misure in ordine crescente

2) se necessario raggruppiamo i dati in intervalli (bins) e a definire il bin ci guidano i dati stessi basta valutare quali siano :

→ il binnaggio "naturale" = *larghezza del minimo intervallo tra i dati*

→ l' intervallo di valori = *larghezza del massimo intervallo tra i dati*

→ la numerosita' del campione

dati originali	dati ordinati
1.720	1.651
1.651	1.670
1.811	1.691
1.722	1.700
1.721	1.701
1.670	1.710
1.710	1.710
1.721	1.712
1.740	1.712
1.701	1.720
1.730	1.721
1.700	1.721
1.761	1.721
1.721	1.721
1.751	1.722
1.712	1.730
1.710	1.730
1.721	1.731
1.691	1.740
1.790	1.740
1.740	1.751
1.730	1.760
1.760	1.761
1.731	1.790
1.712	1.811

binnaggio "naturale"
bins di 0.001

intervallo di valori :
da 1.651 a 1.811

⇩

numero di bins
(1.811 - 1.651)/0.001
= 160 bins

ma
la numerosita del camione e' di 25

⇩

160 bins con soli 25 dati a disposizione produrrebbero un istogramma con moltissimi bins vuoti

→ molto poco significativo

⇨ raggruppiamo i dati in bin di larghezza = 0.01

bin	intervallo
1.65	1.650 ÷ 1.659
1.66	1.660 ÷ 1.669
1.67	1.670 ÷ 1.679
1.68	1.680 ÷ 1.689
1.69	1.690 ÷ 1.699
1.70	1.700 ÷ 1.709
1.71	1.710 ÷ 1.719
1.72	1.720 ÷ 1.729
1.73	1.730 ÷ 1.739
1.74	1.740 ÷ 1.749
1.75	1.750 ÷ 1.759
1.76	1.760 ÷ 1.769
1.77	1.770 ÷ 1.779
1.78	1.780 ÷ 1.789
1.79	1.790 ÷ 1.799
1.80	1.800 ÷ 1.809
1.81	1.810 ÷ 1.819

3) calcoliamo la **frequenza** (F_i) con la quale si presenta un particolare risultato nei vari bins

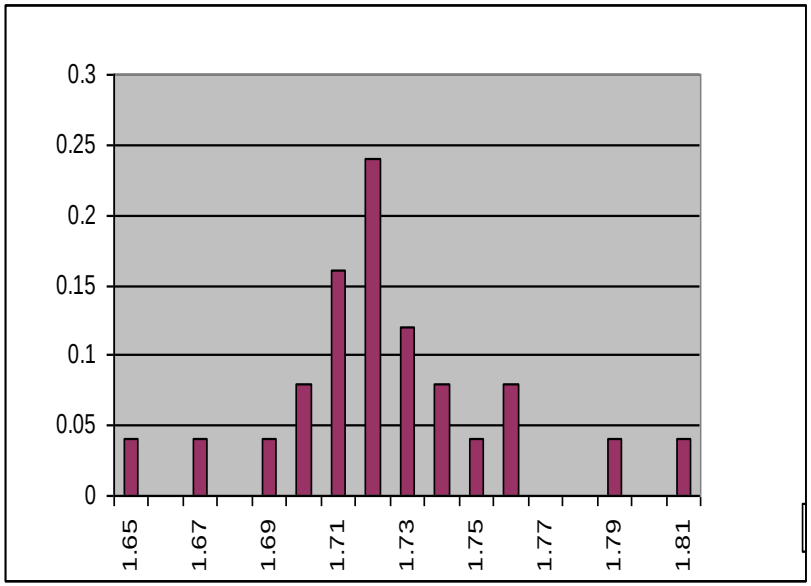
4) Calcoliamo la frequenza relativa (Fr_i) dividendo la frequenza per il numero totale di misure, ossia "**normalizziamo**" i dati binnati alla frequenza relativa

5) grafichiamo la frequenza relativa

Misure "binnate" (x_i)	Frequenza (F_i)	Frequenza relativa $= F_i / N_{tot}$ (Fr_i)
1.65	1	0.04
1.66	0	0
1.67	1	0.04
1.68	0	0
1.69	1	0.04
1.70	2	0.08
1.71	4	0.16
1.72	6	0.24
1.73	3	0.12
1.74	2	0.08
1.75	1	0.04
1.76	2	0.08
1.77	0	0
1.78	0	0
1.79	1	0.04
1.80	0	0
1.81	1	0.04

$N_{tot} = \Sigma F_i = 25$

istogramma delle frequenze relative



la forma dell'istogramma da' una idea di quale sia la distribuzione di probabilita' della variabile misurata

Nota Bene: la frequenza **deve** essere **normalizzata** all'unita affinche' l'istogramma rappresenti una distribuzione di probabilita'

Nota bene : questo costituisce l’approccio frequentistico

ma oggigiorno e’ sempre piu’ adoperato anche l’approccio bayesiano

Ricapitolando:

postulando che effettuare una misurazione sia assimilabile

ad effettuare un esperimento aleatorio

possiamo ”recuperare” la non riproducibilita’ degli esperimenti

Backup Slides