ANALISI STATISICA DEGLI ERRORI

1. TIPI DI GRANDEZZE MISURATE

In generale possiamo trovare due tipi di grandezze: quelle discrete e quelle continue.

1.1 Grandezze discrete

Si dicono grandezze discrete quelle grandezze che si basano su eventi elementari che assumono valori distinti.
Per esempio e' possibile avere 2 o 3 bambini in una famiglia ma non certo 2,5. Analogamente quando tiriamo un dado ci si aspetta un valore da 1 a 6, cioe' non e' possibile avere un valore, ad esempio, di 4,5.
Il modo piu' semplice di presentare tali dati e' di ordinarli in ordine crescente e disegnarli con un diagramma di frequenza assoluta, con la grandezza misurata come ascissa e la frequenza con cui essa si presenta come ordinata, Alternativamente, l'ordinata del diagramma delle frequenze puo' essere resa adimensionale dividendo la frequenza dell'evento per il numero totale degli eventi. Questo diagramma e' detto diagramma delle frequenze relative (vedi FIG. 1).

1.2 Grandezze distribuite con continuita'

Le grandezze distribuite con continuita' sono quelle in cui gli eventi possono assumere qualsiasi valore tra limiti dati. Ossia, si puo' dire che e' sempre possibile avere un valore intermedio tra altri due valori, non importa come essi siano stati presi.

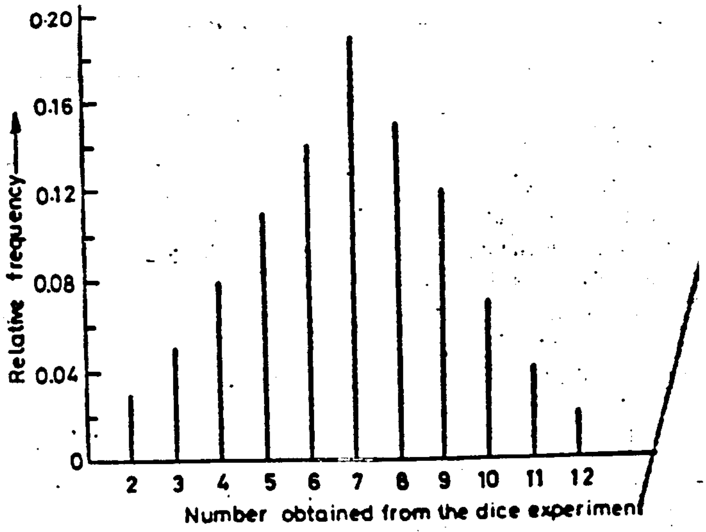
1.2.1 Istogramma

Un istogramma e' una rappresentazione grafica di dati misurati (che sono generalmente relativi a grandezze continue) in cui l'ascissa indica i valori misurati e l'ordinata la frequenza d'occorenza in un range specifico di valori misurati. Percio' e' importante nel costruire un istogramma suddividere le osservazioni in gruppi adeguati (intervalli, classi o celle) scegliendo i limiti adatti. Per prima cosa, quindi, determiniamo il "range" di valori R(x) che e' dato da:

$$R(x) = (X_{max} - X_{min}) \qquad (1)$$

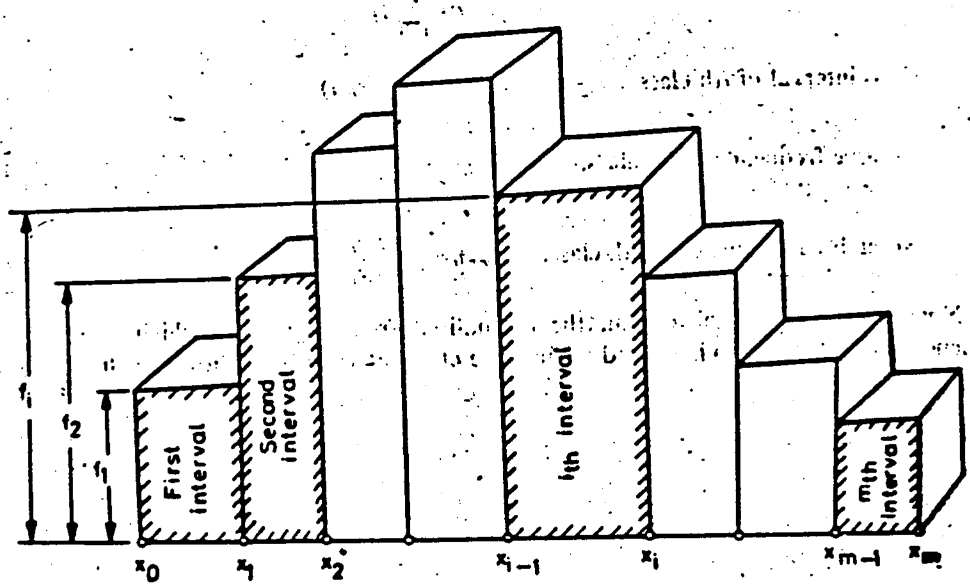
dove X_{max} e' il valore piu' elevato e X_{min} e' il piu' basso.
Inoltre

$$R(x) = \text{numero di classi} * \text{ampiezza della classe.}$$

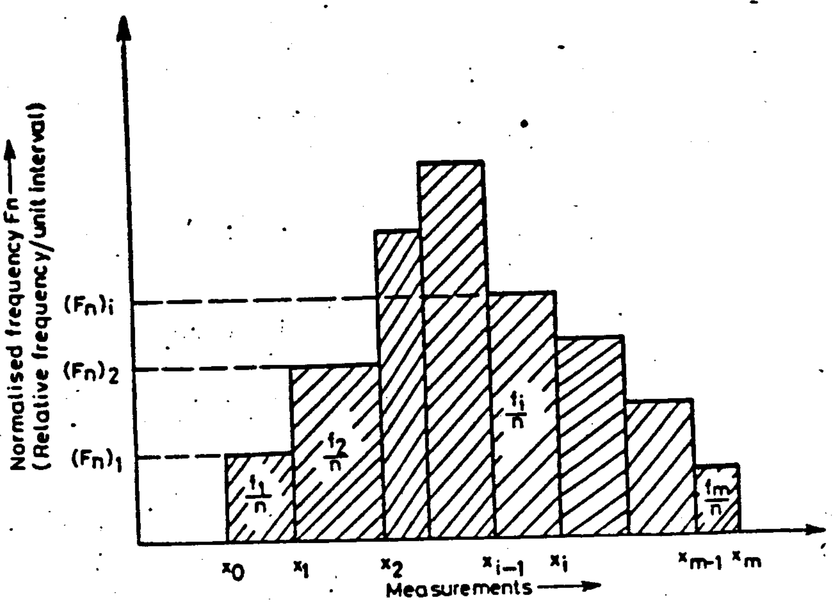


N	F	f
2	6	0,03
3	10	0,05
4	16	0,08
5	22	0,11
6	28	0,14
7	38	0,19
8	30	0,15
9	24	0,12
10	14	0,07
11	8	0,04
12	4	0,02
TOT.	200	1,00

— FIG. 1 —



— FIG 2 —



— FIG 3 —

Prima di decidere l'ampiezza, il numero e gli estremi delle classi, bisogna tener conto delle seguenti osservazioni:

1) Le classi devono contenere tutti i dati
2) Esse hanno usualmente almeno di 6 intervalli e non più di 16.
3) Ogni dato deve essere contenuto in uno ed in uno solo degli intervalli.

Per soddisfare i criteri su menzionati di solito si segue la seguente procedura. Il valore di R(x) ottenuto dalla (1) e' diviso da un valore opportuno di ampiezza così che il numero delle classi stia tra 6 e 16. Se questo numero ottenuto non e' intero allora si prende il primo intero maggiore del numero per rappresentare il numero di classi. E' bene notare che non c'e' alcuna regola che obblighi a mantenere uniforme l'ampiezza delle classi. Percio' se la situazione richiede un numero fissato di classi, le loro ampiezze potranno essere disuguali per poterle adattare al range dato dei valori. Comunque, in pratica esse sono mantenute uguali poiche' questo facilita il confronto della frequenza tra classi diverse, eliminando la necessita' di tener conto delle differenze nella misura dell'intervallo di classe.
Dopo aver deciso l'intervallo di classe e il numero di classi, i punti medi della classe e gli estremi risultano fissati. Il punto medio o marker della classe è il valore della variabile che si trova a meta' tra l'estremo superiore e quello inferiore della classe stessa.
Nel determinare gli estremi della classe occorre fare in modo che nessuna misura cada sugli stessi così' che senza difficolta' possiamo determinare a quale classe appartiene una particolare osservazione. Percio' normalmente fissiamo il valore di estremita' della classe con un numero di cifre significative maggiore di quello con cui effettuiamo le misure. Questo assicura che nessuno dei dati coincida con gli estremi.
Dopo di cio', i dati sperimentali vengono raggruppati nelle varie classi scelte. Il numero delle osservazioni che cade all'interno degli estremi di una particolare classe e' noto come la frequenza della classe. A questo punto possono essere disegnati gli istogrammi delle frequenze assolute, frequenze relative o frequenze percentuali.(vedi FIG. 2)
Nella pratica, tali distribuzioni di frequenze sono normalizzate rispetto all'ampiezza della classe e sono dunque rappresentate sotto forma di istogrammi normalizzati.

1.2.2 Istogrammi normalizzati

In un istogramma normalizzato, si utilizza come ordinata la frequenza normalizzata invece della frequenza assoluta o percentuale.
La frequenza normalizzata e' ottenuta come segue.

Avendo deciso di raggruppare i dati in m classi, siano X0, X1,..., Xm gli estremi delle classi e f1, f2,..., fm le frequenze di occorrenza nelle varie classi.

Il numero totale delle misure e' $n = \sum_{i=1}^{m} f_i$

l'ampiezza della classe i-esima è $(x_i - x_{i-1})$

la frequenza relativa della i-esima classe e' $(f_r)_i = f_i / \sum_{i=1}^{m} f_i$

la frequenza normalizzata della i-esima classe e' $(F_n)_i = \dfrac{(f_r)_i}{(x_i - x_{i-1})}$

Ora, in un istogramma normalizzato, la frequenza normalizzata (Fn)i (che e' la frequenza relativa per un intervallo unitario) e' tracciata come e' mostrato in Fig. 3.
L'area della i-esima banda in un istogramma normalizzato, Ai sarà pari a

Ai = ordinata * ampiezza della classe =

$$A_i = \frac{1}{(x_i - x_{i-1})} \frac{f_i}{\sum_{i=1}^{m} f_i} (x_i - x_{i-1}) = \frac{f_i}{\sum_{i=1}^{m} f_i} = \frac{f_i}{n} = (f_r)_i$$

Percio', l'area assegnata ad ogni banda in un istogramma normalizzato rappresenta la frequenza di occorrenza relativa (fr)i degli eventi che cadono all'interno dell'intervallo della classe data.
Nel caso in cui tutti gli intervalli delle classi sono uguali, l'ordinata di ogni intervallo di classe e' proporzionale alla frequenza relativa. L'area dell'istogramma e' data da:

$$\sum_{i=1}^{m} A_i = \sum_{i=1}^{m} \frac{f_i}{n} = \frac{n}{n} = 1$$

L'area di un istogramma normalizzato e' dunque unitaria.
Si può dimostrare che anche l'area di una distribuzione gaussiana e' unitaria. Percio', se i dati distribuiti con continuita' devono essere comparati con una distribuzione Gaussiana, allora l'istogramma normalizzato diventa piu' appropriato degli istogrammi ordinari.

2. TENDENZA CENTRALE DEI DATI

Uno fra gli importanti parametri che descrivono le informazioni numeriche riguarda la locazione della tendenza centrale dei dati.
Ci sono svariati modi attraverso i quali e' possibile determinare la tendenza centrale dei dati (mediana, moda, media aritmetica). Comunque, puo' essere dimostrato che ciascuno dei modi precedentemente citati puo' essere piu' appropriato per descrivere un tipo di dati e completamente inadatto per altri.

2.1 Moda

Possiamo dire che la moda e' il valore che si presenta con maggiore frequenza. Ad esempio consideriamo un set di numeri contenente i seguenti valori discreti : 15, 18, 17, 14, 17, 13, 16. E' ovvio che in questo set di valori la frequenza del numero 17 e' la massima. Percio' puo' essere definito come valore modale.
Puo' essere notato dall'esempio precedente che, in un set di valori discreti, il valore modale puo' essere determinato con estrema facilita'. Comunque nel caso di dati distribuiti con continuita', il valore modale si trova nel gruppo di dati che hanno la maggiore densita' di frequenza. In altre parole, esso si trova all'interno del piu' alto rettangolo dell'istogramma.
Per determinare il valore modale consideriamo la parte di istogramma contenente il rettangolo di frequenza maggiore insieme con le classi adiacenti (vedi FIG. 4)
Noi definiamo la moda come l'ascissa X' del punto di intersezione P dei segmenti QS e RT. Puo' essere dimostrato che il valore della moda X' e' dato dalla seguente formula:

$$\text{Moda } X' = L_1 + \left[\frac{\Delta_1}{\Delta_2 + \Delta_1}\right] \cdot C$$

laddove L1 e' l'estremo della classe piu' bassa

Δ_1 = eccesso della frequenza modale sulla frequenza della classe adiacente piu' bassa

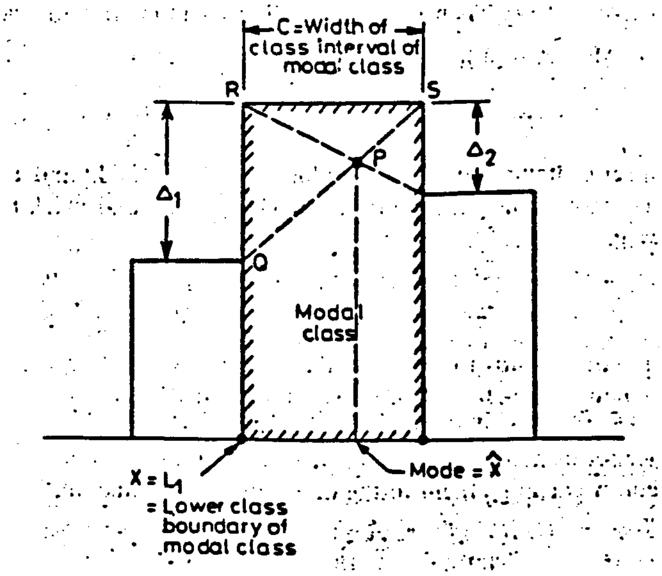
Δ_2 = eccesso della frequenza modale sulla frequenza della classe adiacente piu' alta

C = ampiezza della classe modale.

Occasionalmente una classe puo' essere bimodale, cioe' essa puo' contenere due distinti valori che possono occorrere molto piu' frequentemente degli altri valori della variabile.
Tutto cio' sta usualmente ad indicare che il campione proviene da due popolazioni differenti.

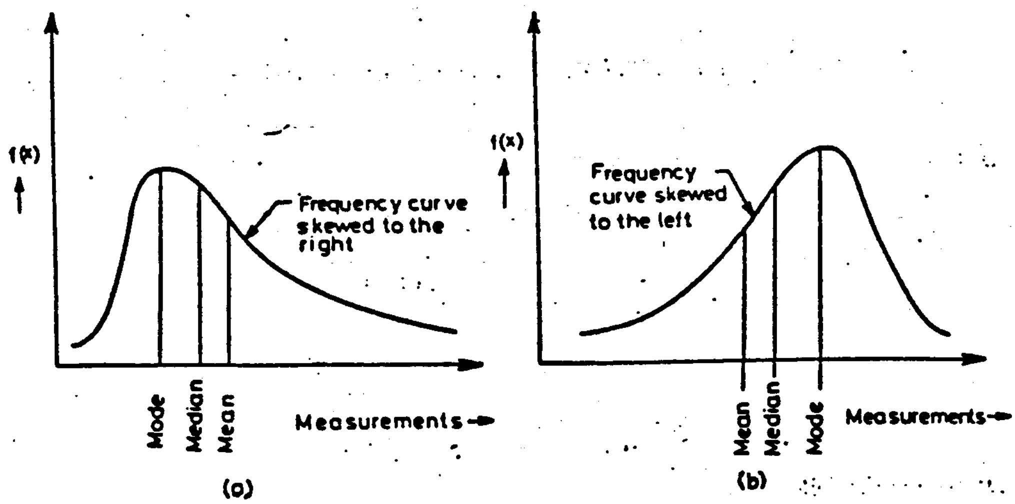
2.2 Mediana

La mediana e' definita come il valore centrale di un gruppo di dati raggruppati in ordine numerico che puo' essere crescente



— FIG 4 —

moda



— FIG 5 —

Posizione relativa di media, moda e mediana

(dal piu' basso valore di osservazione al piu' alto) o viceversa. Nel caso in cui la dimensione del campione e' pari, noi determiniamo la mediana operando la media dei due valori centrali.
La mediana e' anche detta 50-simo percentile e si denota con P_{50}. 50-esimo percentile significa che approsimativamente il 50% delle misure nella distribuzione sono minori di P_{50} ed il 50% delle misure nella distribuzione sono maggiori di P_{50}.
Supponiamo che N osservazioni dei dati siano ordinate in senso crescente o decrescente da 1 a N. Allora possiamo esprimere matematicamente il valore della mediana come:

$$\text{Mediana} = \{X\}_{(n+1)/2} \qquad \text{quando N è dispari}$$

$$\text{Mediana} = \frac{1}{2}\left[\{X\}_{n/2} + \{X\}_{(n/2)+1}\right] \qquad \text{quando N e' pari}$$

La mediana e' una quantita' posizionale, nel senso che il suo valore e' determinato dal valore di osservazione che occupa una particolare posizione nella distribuzione. Puo' essere notato che la mediana non e' influenzata dai valori estremi dei dati.
In taluni casi, la distribuzione dei dati puo' non essere simmetrica. In tali casi la mediana e' considerata come la locazione centrale dei dati.
Geometricamente, il valore della mediana e' ottenuto dalla intersezione dell'ascissa dell'istogramma con una retta verticale che divide l'istogramma stesso in due parti aventi aree uguali.
Comunque, si puo' precisare che il valore ottenuto per la mediana puo' non essere rappresentativo se i dati non si raggruppano al centro della distribuzione.

2.3 Media Aritmetica

La Media Aritmetica e' il piu' comune ed il piu' usato tra i parametri di tendenza centrale dei dati. Per definizione, la media aritmetica di un gruppo di dati e' uguale alla somma dei valori del gruppo diviso per il numero di valori contenuti nello stesso. E' possibile determinare sia la media di un insieme di valori misurati che la media di una popolazione. Il simbolo $\bar{X}$ tradizionalmente denota il primo (la media di un insieme di dati o campione) mentre $\bar{X}$ denota il secondo. Si intende per popolazione il totale dei valori tra cui il campione e' preso. Laddove noi facciamo riferimento ad un campione, noi non intendiamo solo una porzione di popolazione, ma una porzione selezionata in modo tale che i valori osservati non siano affetti da bias. Tale processo e' generalmente definito come campionatura casuale o random.
Supponiamo che un particolare campione abbia un totale di n osservazioni in cui i veri valori osservati sono:
X1 occorso f1 volte, X2 occorso f2 volte, Xm occorso fm volte.
Ora la somma delle frequenze di occorrenza dei valori veri ci da' ovviamente il numero totale di osservazioni del campione.

La media X e' data da: $$\overline{X} = \frac{\sum_{i=1}^{m} f_i \cdot x_i}{\sum_{i=1}^{m} f_i}$$

La stessa espressione puo' anche essere applicata per dati distribuiti con continuita' laddove Xi rappresenta il valore del punto medio della i-esima classe avente la frequenza fi. Similarmenteè possibile calcolare la media di una popolazione. Consideriamo la popolazione costituita da N osservazioni e siano X1, X2,...,Xk le medie relative a k campioni.-.., .. siano le medie dei valori di K campioni. La media della popolazione, X, e' data da:

$$\overline{\overline{X}} = \frac{\sum_{j=1}^{k} X_j}{N}$$

2.4 <u>Media Aritmetica Pesata</u>

A volte, noi associamo ai numeri X1, X2,.., Xm certi coefficienti pesati o pesi W1, W2,.., Wm rispettivamente, a seconda del significato o dell'importanza attribuita al singolo numero. In tal caso, la media aritmetica pesata e' definita come:

$$\overline{X} = \frac{W_1 x_1 + W_2 x_2 + \cdots + W_n x_n}{W_1 + W_2 + \cdots + W_n} = \frac{\sum_{i=1}^{n} W_i \cdot x_i}{\sum_{i=1}^{n} W_i}$$

2.5 <u>Migliore stima del valore vero di una misura</u>

Noi abbiamo fino ad ora descritto tre parametri chiamati moda, mediana e media aritmetica, che possono essere usati per descrivere la tendenza centrale dei dati. Per curve che sono moderatamente asimmetriche, vi e' una relazione empirica che lega questi parametri:

Media - moda = 3 (media-mediana)

La FIG. 5 mostra le relative posizioni di moda, mediana e media per curve di frequenza spostate a destra o a sinistra. Comunque, per curve simmetriche, moda, mediana e media coincidono.
Nell'analisi di dati sperimentali, ci si trova usualmente di fronte al problema della determinazione della migliore stima del valore vero della misura. Supponiamo che X1, X2,..., Xm siano misure di una quantita' il cui valore vero e' X. Allora la misura potrebbe avere i seguenti errori:

Errore in X1, e1 = X1 - X

....

Errore in Xn, en = Xn - X

Alcuni di questi errori potrebbero essere positivi mentre altri potrebbero essere negativi. Ora, se noi diamo eguale importanza agli errori sia negativi che positivi, allora la strada migliore e' quella di elevare al quadrato tutti gli errori.
Ora, il criterio piu' plausibile per la migliore scelta di X sarebbe quello grazie al quale gli errori commessi nelle varie misure risultano minimi. Alternativamente la somma dei quadrati degli errori possa essere minima.
La somma dei quadrati degli errori puo' essere scritta come:

$$S_e = e_1^2 + e_2^2 + \dots + e_i^2 + \dots + e_n^2 = (x_i - X)^2 + (x_2 - X)^2 + \dots + (x_i - X)^2 + \dots + (x_n - X)^2$$

La somma dei quadrati degli errori e' una funzione di X che consente la migliore stima del valore vero. Ora, al fine di determinare il valore di X che renda minima Se si procede come segue.
Detto X = XO il valore di X che minimizza Se si ha:

$$\frac{dS_e}{dX} = -2(x_1 - X) - 2(x_2 - X) \cdots -2(x_i - X) \cdots -2(x_n - X) = 0$$

$$\Rightarrow X_0 = (x_1 + x_2 + \dots + x_i + \dots x_n)/n = \overline{X}_n$$

In generale, quindi, la media aritmetica e' considerata la migliore stima del valore vero di un insieme di valori misurati.

2.5 Esempio

La tabella mostra la distribuzione di frequenza della resistenza di un resistore fabbricato da un pro .

Resistenza [Ω]	Frequenza
93 - 95	4
96 - 98	15
99 - 101	33
102 - 104	21
105 - 107	7

Determinare: a) la media aritmetica, b) la mediana, c) la m d , d) usare la formula empirica per calcolare il valore modale e comparare tale valore con quello precedentemente ottenuto.

a) Per trovare la media aritmetica, noi possiamo supporre che 4 resistori hanno una resistenza di 94Ω, 15 di 97Ω ecc.
La media aritmetica diviene:

$$\frac{4 \times 94 + 15 \times 97 + 3 \times 100 + 21 \times 103 + 7 \times 106}{4 + 15 + 33 + 21 + 7} = \frac{8036}{80} = 100{,}45\ \Omega$$

b) il valore mediano puo' essere determinato attraverso il metodo di interpolazione o attraverso la definizione. Per definizione la mediana e' quel valore tale che la meta' della frequenza totale (80/2 = 40) giace prima di esso e l'altra meta' dopo.
Ora la somma delle frequenze delle due prime classi e' 4+15=19. Cosi' per avere i desiderati 40, abbiamo bisogno di 21 dei 33 campioni della terza classe. Il valore mediano dovrebbe stare ai 21/33 dell'ampiezza dell'intervallo tra 99 e 101.*

$$\text{Mediana} = L_1 + \left\{ \frac{\frac{N}{2} - (\sum f_i)}{f_{mediana}} \right\} \cdot C = 99 + \frac{40-(4+15)}{33}(101-99) = 100{,}273\,\Omega$$

c) Il valore modale per dati distribuito con continuita' e':

$L_1 = 99$

$\Delta_1 = 33-15 = 18$

$\Delta_2 = 33-21 = 12$

$$\hat{X} = L_1 + \left\{ \frac{\Delta_1}{\Delta_1+\Delta_2} \right\} C = 99 + \left(\frac{18}{12+18} \right)(101-99) = 100{,}2\,\Omega$$

d) Usando la formula empirica

Moda = media − 3 (media-mediana)

si ottiene:

Moda = 99,92 Ω

Si puo' notare che c'e' un buon accordo tra il valore modale calcolato per definizione e quello ottenuto usando la relazione empirica per distribuzioni moderatamente asimmetriche.

3. MISURE DI DISPERSIONE

Le sole misure di tendenza centrale usualmente non danno un'adeguata descrizione dei dati sperimentali. Infatti si dovrebbe prendere in considerazione anche la variabilita' dei dati. Per esempio se una persona mette la sua testa in un frigorifero ed i suoi piedi in un forno, allora mediamente si puo' dire che si senta bene. Ma in pratica la persona vorrebbe star meglio, a causa dell'estrema differenza tra le due temperature. Quindi un altro importante parametro dei dati sperimentali e' l'entita' della dispersione o il grado di compattezza dei dati. Nel seguito si esaminano alcune misure di dispersione dei dati.

3.1 Range

Il modo piu' semplice di rappresentare la dispersione e' il range, che non e' altro che la differenza tra i valore massimo

* cioè: $99 + \frac{21}{33}(101-99) = 100{,}273\,\Omega$ _ Utilizzando la definizione:

e minimo assunti dai dati. Il maggiore svantaggio associato al range come misura di dispersione e' il fatto che esso e' basato esclusivamente sulla dispersione dei valori estremi, mentre non fornisce informazioni sulla compattezza, o meno, dei valori compresi tra gli estremi stessi.
Quindi e' azzardato considerare il range come misura di dispersione. Comunque si puo' utilizzare il range per avere una idea approssimativa circa l'estensione dei dati disponibili e della compattezza dei dati.

3.2 La deviazione media

La deviazione media e' un miglioramento sul range in quanto nella sua valutazione si tiene conto di tutti i valori delle misure effettuate.
In essa si considera il valore assoluto della deviazioni di ogni dato dalla media aritmetica, e viene effettuata la media. Quindi:

$$\text{Deviazione Media (DM)} = \frac{\sum_{i=1}^{n} |x_i - \bar{X}|}{n} = \frac{\sum_{i=1}^{n} |d_i|}{n} = \overline{|(x_i - \bar{X})|} \quad (2)$$

Dove d_i e' la deviazione dalla media della i-esima osservazione. Nell'equazione (2) $|x_i - \bar{X}|$ rappresenta il valore assoluto dello scostamento di x_i dalla media.
Se x_1, .., x_n occorrono con frequenze f_1, .., f_n, allora la deviazione media puo' essere scritta come:

$$DM = \frac{\sum_i^n f_i |x_i - \bar{X}|}{\sum_i^n f_i} = \frac{\sum_i^n f_i |d_i|}{N} = \frac{\sum_i^n f_i |(x_i - \bar{X})|}{N} = \overline{f_i (x_i - \bar{X})} \quad (3)$$

dove $\sum_{i=1}^{n} f_i = N$, numero totale di osservazione effettuate.

L'equazione (3) e' anche utile per dati raggruppati, dove x_i rappresenta l'identificativo della classe e f_i la frequenza della classe corrispondente.
Occasionalmente la DM puo' anche essere definita in termini di deviazioni assolute dalla mediana o altre medie. Una proprietà interessante a tale proposito è che la $\sum_i |(x_i - a)|$ assume un valore minimo per a pari alla mediana.
Si puo' notare che sarebbe piu' appropriato usare la terminologia "deviazione media assoluta" piuttosto che "deviazione media." In effetti si puo' dimostrare che la media delle deviazioni reali dai valori medi e' zero:

$$\frac{\sum_i^n d_i}{n} = \sum_{i=1}^{n} \frac{(x_i - \bar{X})}{n} = \frac{\sum x_i}{n} - \frac{n\bar{X}}{n} = \bar{X} - \bar{X} = 0 \quad (4)$$

Nel calcolo della DM si riscontrano alcune deviazioni positive, le rimanenti negative. Possiamo notare nell'equazione (4) che esse si bilanciano l'un l'altra intorno alla media. Pertanto per ogni deviazione deve essere calcolato il valore assoluto che e' piuttosto scomodo ai fini pratici. Percio' la deviazione media e' meno comunemente impiegata come misura di dispersione dei dati, rispetto ai parametri che verranno introdotti nel seguito.

3.3 Varianza

Ora invece di considerare la media delle deviazioni assolute intorno al valore medio, determineremo la deviazione media quadratica intorno al valore medio. L'operazione di elevazione al quadrato evita il problema inerente ai valori assoluti, poiche' tutti i quadrati sono in segno positivi. La deviazione media quadratica anche definita come varianza e' molto usata come misura della variabilita' dei dati. Infatti, i suoi vantaggi sono:

1) impiega tutti i valori dei dati ed e' sensibile al cambiamento di uno qualunque di essi;

2) e' indipendente dalla tendenza centrale, cioe' il valor medio, perche' usa le deviazioni dal valore medio;

3) il suo computo e' ragionevolmente semplice perche' i valori quadrati di ogni deviazione dalla media sono positivi.

La varianza e' generalmente indicata con il simbolo σ^2 ed e' cosi' calcolata:

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

svolgendo il quadrato:

$$\sigma^2 = \frac{1}{n}\sum_i^n(x_i^2-2x_i\bar{X}-\bar{X}^2) = \frac{1}{n}\sum_i^n x_i^2 - 2\bar{X}\sum_i^n x_i + \frac{n\bar{X}^2}{n} = (\bar{X^2})-2(\bar{x}^2)+(\bar{X}^2) = (\overline{X^2})-(\bar{X})^2$$

Quindi potremmo esprimere la varianza come differenza tra la media dei valori al quadrato meno il quadrato del valore medio dell'insieme dei dati.

3.4 Deviazione standard

Molto spesso viene utilizzata la radice quadrata della varianza, denotata con σ e chiamata deviazione standard o radice quadrata dello scarto quadratico medio. Il vantaggio principale della deviazione standard rispetto alla varianza e' che la sua unita' di misura e' la stessa della quantita' misurata. Essa e' definita come:

$$\sigma = \sqrt{\frac{\sum_i^n d_i^2}{n}} = \sqrt{\frac{\sum_i^n (x_i-\bar{X})^2}{n}} \qquad (5)$$

Nell'eventualita' di dati raggruppati si potrebbe scrivere:

$$\sigma = \sqrt{\frac{\sum_i^n f_i(x_i-\bar{X})^2}{n}}$$

Dove $\sum_i^n f_i = n$ e' il numero totale di osservazioni.
Si puo' notare che per distribuzioni di frequenza non eccessivamente asimmetriche, la relazione empirica tra la deviazione media e la deviazione standard e' data da:

$$\text{Deviazione media} = \frac{4}{5}\cdot(\text{deviazione standard})$$

$$i.e. \qquad DM = \frac{4}{5}\sigma$$

3.5 Fattore di dispersione

Il fattore di dispersione e' un numero adimensionale che rappresenta il grado di dispersione dei dati rispetto al valore medio. E' ottenuto dividendo la deviazione standard per il valor medio.

$$C.V. = \frac{\sigma}{\bar{X}}$$

E' generalmente espresso in percentuale ed e' chiamato anche fattore di dispersione percentuale. Si noti che il fattore di dispersione cade in difetto quando X e' prossimo a zero.

3.6 Deviazione standard corretta

La misura di dispersione σ , cade in difetto quando il numero di osservazioni e' piccolo. Per esempio quando si considera una sola osservazione di una grandezza fisica il valore di σ e' zero. Cio' implica che la misura non ha dispersione e conseguentemente non soggetta a nessun altro errore di misura. Al contrario il risultato è indefinibile qualora si consideri una sola osservazione. Comunque l'affidabilita' di σ cresce al crescere del numero di osservazioni. Percio' il valore della deviazione quadratica media σ viene corretto in maniera tale da fornire una stima piu' realistica della precisione delle misure, anche per un numero ridotto di osservazioni. Cio' si ottiene dividendo la somma dei quadrati delle deviazioni dalla media per (n-1) che rappresenta il grado di liberta', anzichè per n che e' il numero totale di osservazioni.
Il numero di gradi di liberta' si riferisce al numero di informazioni indipendenti originate dai dati disponibili. Esso puo' essere determinato usando il ragionamento induttivo che segue.
Un caso semplice costituito da un insieme di 2 misure fornisce solo un input utile rispetto alla dispersione valutata intorno alla popolazione media perche' il set di campioni non fornisce solo informazioni sulla dispersione ma viene anche utilizzato per valutare la variabile rispetto alla quale la dispersione e' misurata, vale a dire X. Cosi' un insieme di 2 misure fornisce solo una osservazione indipendente riguardo la dispersione. Analogamente in un insieme di 10 misure noi possiamo valutare 10 deviazioni rispetto al valor medio. Di conseguenza si può dire che in realtà ci sono 9 deviazioni indipendenti e la decima potrebbe essere determinata dal fatto che la somma delle deviazioni e' uguale a 0. Quindi in generale un insieme di n misure fornisce n-1 osservazioni indipendenti con riferimento alla deviazione standard della popolazione. Infatti l'espressione generale per il grado di

liberta' e' (n-m) dove n e' il numero di osservazioni ed m e' il numero di costanti da valutare dai dati di partenza, allo scopo di generare valori dai quali le deviazioni devono essere calcolate. Nel caso in cui m=1 e cioè in cui solo una costante deve essere calcolata, la media aritmetica, la deviazione standard corretta (denotata con S) detta anche "unbiased" o la migliore stima della precisione di un apparato puo' essere scritta:

$$s = \sqrt{\frac{\sum_i^n (x_i - \bar{X})^2}{(n-1)}} = \sqrt{\frac{\sum_i^n d_i^2}{(n-1)}} \qquad (6)$$

Confrontando le equazioni (5)e (6) la relazione tra S e puo' essere scritta

$$s = \sqrt{\frac{n\sigma}{(n-1)}}$$

La quantita' $\sqrt{\frac{n}{n-1}}$ e' conosciuta come fattore di correzione di Bessel. L'equazione mostra che all'aumentare di n diminuisce la differenza tra S e σ. Quindi l'aumentare del numero di campioni la deviazione standard diventa una stima di precisione piu' affidabile. In pratica per n > 25 il fattore di Bessel diviene tanto vicino all'unità da poter essere trascurato.

4. DEVIAZIONE STANDARD DELLA MEDIA.

Il termine deviazione standard corretta o migliore stima di precisione dell'apparato fornisce un'informazione sugli errori casuali che vengono commessi in ogni misura. In altre parole noi possiamo dire che:

$$x = x_{misurata} \pm s$$

Analogamente, la precisione di un insieme di misure vorrebbe indicare come e' vicina la media delle misure effettuate al valore vero. In altre parole, l'errore standard intrinseco (denotato s_n da alcuni autori) o la migliore stima di incertezza U_n e' data da $\sigma(\bar{X}_n)$.
Quindi,

$$X = (\bar{X}_n)_{misurata} \pm \sigma(\bar{X}_n) = (\bar{X}_n)_{misurata} \pm U_n$$

4.1 Migliore stima dell'incertezza di un insieme di misure

Come prima e' stato spiegato, la migliore stima di incertezza rappresenta la estensione dell'errore casuale nei valori misurati. Essa viene determinata considerando m insiemi di misure di n campioni in una popolazione di (nxm) osservazioni. La media di un insieme di misure varrà dunque:

$$\bar{X}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n}$$