

Dopo che ciascuna misura è stata registrata nella classe appropriata per mezzo di una sbarretta, come mostrato nella tabella 1, i risultati della classificazione sono registrati sotto forma di una tabella di frequenza come mostrato nella seconda metà della tabella 1.

### **Rappresentazione grafica di distribuzioni empiriche (o osservate)**

Un'idea approssimativa di come sono distribuiti i valori di una variabile casuale si può ottenere esaminando il corrispondente istogramma. L'istogramma per i dati di tabella 1 è mostrato nella figura 2, si può osservare che i marchi di classe sono nei punti di mezzo delle basi dei rettangoli che compongono l'istogramma.

Fortunatamente molte importanti distribuzioni che si presentano in natura e nella industria hanno forme relativamente semplici, queste forme variano di solito da quella a campana, come in figura 2, a quella che rassomiglia alla metà di una campana.

Una distribuzione dell'ultimo tipo si dice che è *asimmetrica*, il che significa che manca di simmetria rispetto ad un asse verticale.

Risulta che variabili con distribuzione aventi gradi di asimmetria crescenti sono ad esempio la statura, il peso, l'età di matrimonio, l'età di mortalità per certe malattie e la ricchezza, le figure 2, 3 e 4 rappresentano 3 tipiche distribuzioni con gradi crescenti di asimmetria.

Si può osservare che la distribuzione di figura 2 ha la forma di una curva normale e che perciò la distribuzione normale potrebbe essere un modello soddisfacente per la distribuzione delle misure dei diametri delle barrette. Molte misure lineari di tipo industriale possiedono distribuzioni di questo tipo e vengono trattate con successo usando come modello una curva normale.

La distribuzione di figura 4 possiede una forma che suggerisce la possibilità di usare come modello una distribuzione esponenziale, mentre la figura 3 richiede un modello più sofisticato, poiché una distribuzione  $\gamma$  con i suoi due parametri possiede moltissima flessibilità, essa potrebbe servire da modello per la distribuzione di figura 3.

### **Momenti empirici**

Sebbene un istogramma, come quelli mostrati nelle figure 2, 3 e 4, fornisca una notevole quantità di informazioni generali che riguardano la distribuzione di un insieme di misure campionarie, informazioni più precise e utili per studiare una distribuzione si possono ottenere da una descrizione aritmetica della distribuzione. Per esempio se fosse disponibile l'istogramma dei pesi di un campione di 200 individui di un istituto per confrontarlo con l'istogramma di un campione simile di un altro istituto, potrebbe essere difficile stabilire, eccetto in termini molto generali, fino a che punto le due distribuzioni differiscono, piuttosto che confrontare le due distribuzioni dei pesi in questo modo potrebbe bastare confrontare i pesi medi e la variazione dei pesi dei due gruppi.

La natura di un problema statistico determina largamente se alcune semplici proprietà aritmetiche della distribuzione saranno sufficienti a descriverla in modo soddisfacente.

Spesso la maggior parte dei problemi richiedono per la loro soluzione soltanto alcune proprietà fondamentali della distribuzione.

Per semplici distribuzioni di frequenza, come quelle indicate nelle figure 2, 3 e 4, questa descrizione si compie in modo soddisfacente per mezzo dei momenti di ordine basso della distribuzione.

In molti problemi lo statistico è interessato soltanto ai momenti del primo e del secondo ordine, in alti problemi esso usa i primi 4 momenti, ma raramente ne usa più di 4,

Una ragione di ciò è che i momenti di ordine più elevato sono così instabili in ripetuti esperimenti di campionamento che da essi si possono ricavare poche ulteriori informazioni attendibili.

Indichiamo con  $x_1, x_2, \dots, x_n$  i valori osservati di un campione di dimensione  $n$  della variabile casuale  $X$ , allora per analogia con i momenti teorici i momenti empirici dall'origine sono definiti come segue: **Definizione**

***Il momento d'ordine  $k$  dall'origine di una distribuzione empirica è dato da***



$$(1) m'_k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

I momenti empirici sono chiamati anche momenti del campione o momenti campionari perché essi si basano su valori campionari.

Il momento del primo ordine dall'origine  $m'_1$  si indica tradizionalmente col simbolo  $\bar{x}$  ( $m'_1 \rightarrow \bar{x}$ ), esso dà il centro di gravità di una distribuzione empirica proprio come fa la media  $\mu$  per una distribuzione teorica e serve per misurare dove è centrata la distribuzione empirica, esso è chiamato **media del campione**.

Per analogia con la definizione data per le distribuzioni di probabilità, i momenti empirici dalla media sono definiti come segue: **Definizione**

**Il momento d'ordine  $k$  dalla media di una distribuzione empirica è dato da**

$$(2) m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

Poiché il momento del secondo ordine dalla media si usa molto spesso, adesso viene assegnato il simbolo particolare  $s^2$  ( $m_2 \rightarrow s^2$ ) e viene chiamato **varianza del campione**, corrispondentemente  $s$  prende il nome di **scarto quadratico medio o deviazione standard del campione**.

Per calcolare la varianza  $s^2$  è spesso conveniente usare la formula seguente che è l'analogia di una formula già vista per le distribuzioni di probabilità ( $\sigma^2 = \mu'_2 - \mu^2$ ) e precisamente:  $s^2 = m'_2 - \bar{x}^2$ . Se i valori di osservazione  $x_1, x_2, \dots, x_n$  sono stati classificati in una tabella di frequenza con  $x_i$  che rappresenta l' $i$ -esimo marchio di classe,  $f_i$  che rappresenta il numero di osservazioni nell' $i$ -esimo intervallo, cioè la frequenza assoluta di  $x_i$  e  $N_c$  che rappresenta il numero degli intervalli di classe, allora le definizioni dei momenti assumeranno le forme seguenti:

$$(3) m'_k = \frac{1}{n} \sum_{i=1}^{N_c} x_i^k f_i$$

e

$$(4) m_k = \frac{1}{n} \sum_{i=1}^{N_c} (x_i - \bar{x})^k f_i$$

Il valore di  $\bar{x}$  nella (4) è quello ricavato dalla (3) e non dalla (1).

A rigor di termini le formule (3) e (4) definiscono i momenti soltanto per le distribuzioni classificate e sono soltanto approssimazioni dei valori dati dalle (1) e (2), ma le approssimazioni sono di solito così buone che non si fa in genere nessuna differenza fra i due tipi di formule, per esempio  $\bar{x}$  e  $s^2$  sono chiamati **la media e la varianza campionaria** sia che essi siano stati ottenuti dalle formule (1) e (2), che dalle formule (3) e (4).

Non c'è nessuna ragione di classificare i dati se si desiderano soltanto i momenti di una distribuzione empirica, la classificazione ha lo scopo di osservare geometricamente la natura della distribuzione empirica. Se i dati sono già stati classificati per questo scopo e se si desidera ad



esempio il valore di  $\bar{x}$  e  $s^2$  allora può essere più facile calcolarli con le formule (3) e (4) piuttosto che con le (1) e (2) come si è già detto le differenze sono di solito trascurabili.

Si fa osservare infine che in molti testi si usa di solito invece di  $f_i$   $\varphi(x_i)$  che rappresenta il valore della funzione di frequenza assoluta in  $x_i$ , questa funzione viene indicata con  $\varphi(x)$  anziché con  $f(x)$  perché questa ultima si usa in genere per rappresentare la funzione di densità di probabilità.

