ELEMENTI DI STATISTICA

Si definisce ***popolazione oggetto*** l'insieme di tutti quegli elementi che hanno in comune almeno una caratteristica (o attributo)
Lo studio di una popolazione è effettuato quindi dal punto di vista di un suo ***attributo***: si valuta come si distribuiscono le varie forme dell'attributo sugli individui della popolazione.

Requisiti dell'attributo x:

1. L'attributo deve poter assumere distinte forme (x_i) tra loro incompatibili. Ogni individuo quindi deve poter possedere solo una forma dell'attributo;

2. Ogni forma x_i deve poter rappresentare un concetto di classe: ovvero deve essere possibile che parecchi individui posseggano la stessa forma dell'attributo;

3. In ogni individuo della popolazione deve essere presente una delle forme x_i;

4. Almeno due individui della popolazione devono essere in possesso di forme diverse dell'attributo;

Siccome è raramente possibile esaminare tutti gli elementi della popolazione lo scopo dell'analisi statistica è fare previsioni sulla popolazione mediante l'esame di solo una parte della popolazione detta campione.

Argomenti discreti

Sia X l'attributo in esame, esso può assumere diverse forme. Si ha quindi:

$$X \quad \begin{cases} x_1, \quad x_2, \quad x_3, \quad \ldots\ldots, x_n \\ \\ f_1, \quad f_2, \quad f_3, \quad \ldots\ldots, f_n \end{cases}$$

x_i = valore argomentale di x
N = numerosità della popolazione
f_i = frequenza assoluta o numerosità degli elementi che assumono la forme i – ma dell'attributo

$$f_i < N$$

$$\sum_{1=i}^{n} f_i = N$$

si definiscono:

$$\frac{f_i}{N} = \text{frequenze relative,} \quad \text{e si ha:} \quad \sum_{i=1}^{n} \frac{f_i}{N} = 1$$

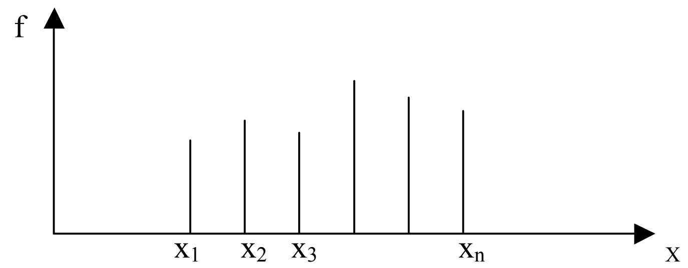
$$X \begin{cases} x_1, \quad x_2, \quad x_3, \quad \ldots\ldots, x_n \\ \\ \dfrac{f_1}{N}, \quad \dfrac{f_2}{N}, \quad \dfrac{f_3}{N}, \quad \ldots\ldots, \dfrac{f_n}{N} \end{cases}$$

X può essere qualitativo o quantitativo e può assumere valori in numero finito (discreto) od infinito

X = variabile statistica ad una dimensione

Rappresentazione grafica

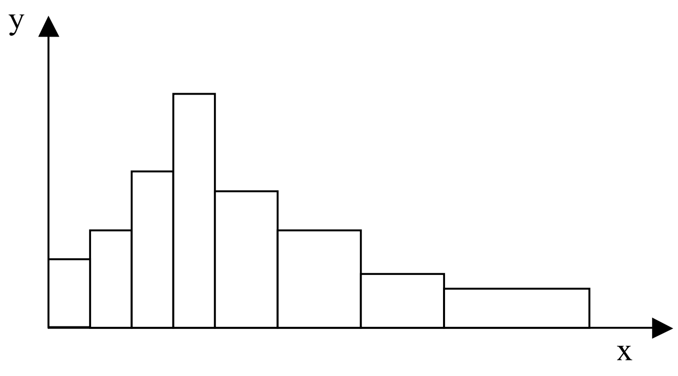
- Diagramma di frequenza



- Istogramma

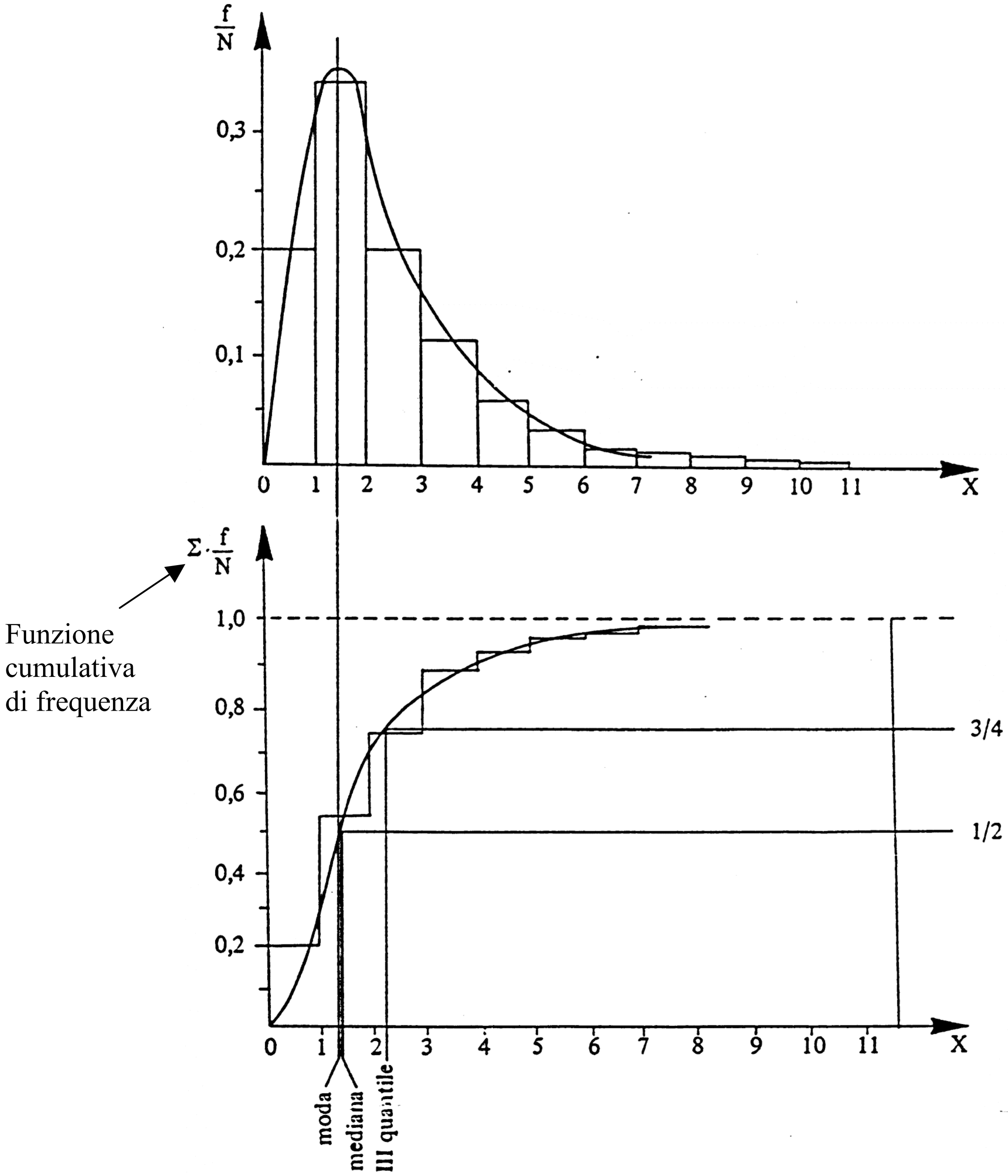
u_o = unità di misura delle x
u_i = ampiezza della classe

$$y_i = \frac{f_i}{u_i/u_0} \qquad \text{oppure} \qquad y_i = \frac{f_i/N}{u_i/u_0}$$



Funzione cumulativa di frequenza o ***funzione di distribuzione***:

$$
X \quad \begin{cases} x_1, \qquad x_2, \qquad \dots \qquad x_n \\[2ex] F_1 = \dfrac{f_1}{N}, \quad F_2 = \dfrac{f_1 + f_2}{N}, \qquad \dots \qquad F_n = \dfrac{f_1 + f_2 + \quad + f_n}{N} \end{cases}
$$



Determinazione di moda, mediana e quantici per una v. s. ordinata sotto forma di istogramma

Rappresentazione sintetica di una variabile statistica a una dimensione

Si definiscono:

- Momento k^{mo} rispetto ad un polo θ

$$m_{k,\theta} = \sum_{i=1}^{n} (x_i - \theta)^k \frac{f_i}{N}$$

- Momento del primo ordine rispetto al polo 0

$$m_{1,0} = \sum_{i=1}^{n} x_i \frac{f_i}{N} = M \quad (MEDIA)$$

- Momento del secondo ordine rispetto al polo 0

$$m_{2,0} = \sum_{i=1}^{n} x_i^2 \frac{f_i}{N} = M_2 \quad (VALORE\ QUADRATICO\ MEDIO)$$

- Momento del secondo ordine rispetto al polo M

$$m_{2,M} = \sum_{i=1}^{n} (x_i - M)^2 \frac{f_i}{N} = \sigma^2 \qquad (VARIANZA)$$

- Scarto: $x_i - M$

- Scarto quadratico medio o deviazione standard: σ

Proprietà

- La media M è il valore del polo per cui è minimo il momento del secondo ordine

$$M_2(\theta) = \sum_{i=1}^{n} (x_i - \theta)^2 \frac{f_i}{N}$$

$$\frac{dM_2(\theta)}{d\theta} = -2\sum_{i=1}^{n} (x_i - \theta) \frac{f_i}{N} = 0$$

$$\theta = M$$

- La variabile statistica formata dagli scarti ha valore medio nullo

$$\sum_{i=1}^{n} (x_i - \theta) \frac{f_i}{N} = 0$$

- Esiste la seguente relazione tra gli indici

$$\sigma^2 = \sum_{i=1}^{n} (x_i - M)^2 \frac{f_i}{N} = \sum_{i=1}^{n} \left(x_i^2 - 2x_iM + M^2\right)\frac{f_i}{N} = M_2 - 2M^2 + M^2 = M_2 - M^2$$

- Se alla variabile statistica sostituisco la stessa aumentata di una costante, la media trasla

$$y_i = x_i + c$$

Se x_i ha media M allora y_i ha media M + c.

$$\sum_i (y_i)\frac{f_i}{N} = \sum_i (x_i + c)\frac{f_i}{N} = \sum_i x_i \frac{f_i}{N} + c\sum_i \frac{f_i}{N} = M + c$$

La varianza non varia.

- Se moltiplico la variabile statistica per una costante, la media è moltiplicata per quella costante , ed anche la varianza viene ad essere moltiplicata:

$$y_i = cx_i$$

Se x_i ha media M allora y_i ha media cM.
Se x_i ha varianza σ^2 allora y_i ha varianza $c^2\sigma^2$.

$$\sum_i (y_i - c \cdot M)^2 \frac{f_i}{N} = \sum_i (c \cdot x_i - c \cdot M)\frac{f_i}{N} = c^2 \sum_i (x_i - M)^2 \frac{f_i}{N}$$

Si definisce ***indice di simmetria***:

$$\gamma = \frac{\sum_{1=I}^{N} (x_i - M)^3 \frac{f_i}{N}}{\sigma^2}$$

$\gamma = 0$	➔	simmetria
$\gamma > 0$	➔	cade nella destra
$\gamma < 0$	➔	cade nella sinistra

ELEMENTI DI TEORIA DELLA PROBABILITÀ

La popolazione è costituita dai risultati di un esperimento ripetitivo di conseguenza:

- si conoscono le leggi che regolano il fenomeno in esame → i risultati saranno prevedibili con certezza

- per l'azione di fattori disturbanti di difficile controllabilità o per eccessiva complessità delle leggi che regolano il fenomeno, non è possibile prevedere con certezza i risultati degli esperimenti.

Nel secondo caso si parla di ***esperimenti casuali o stoacastici***

Il risultato di un esperimento casuale può essere visto come un'estrazione a caso di un individuo dalla popolazione costituita da tutti i risultati che si sarebbero potuti ottenere ripetendo l'esperimento, nelle stesse condizioni, infinite volte.

Tale popolazione da cui si effettuano queste estrazioni a caso prende il nome di ***universo.***
Si definiscono inoltre:

Evento: ogni singolo elemento o sottoinsieme, definito da una certa relazione, dell'universo;

Unione di due eventi: è un evento che si verifica qualora si verifichi uno dei due eventi che lo formano, od entrambi contemporaneamente;

Intersezione di due eventi: è un evento che si verifica quando il risultato dell'esperimento è parte di entrambi gli eventi che formano l'evento intersezione;

Evento nullo: è l'evento che non contiene alcun risultato dell'esperimento;

Eventi che si escludono vicendevolmente: sono quelli la cui intersezione è l'evento nullo;

Evento complementare E^C: si verifica solo se non si verifica alcun risultato dell'esperimento contenuto in E;

E è contenuto in F ($E \subset F$ od $F \supset E$): quando i risultati che costituiscono E fanno parte, tutti, anche di F;

Se ***$E \subset F$ ed $F \subset E$*** si dice che E ed F sono ***identici (E=F).***

PROPRIETÀ e LEGGI

- ***Commutativa***

$$E \cup F = F \cup E$$

- ***Associativa***

$$(E \cup F) \cup G = E \cup (F \cup G)$$
$$(EF)G = E(FG)$$

- ***Distributiva***

$$(E \cup F)G = EG \cup FG$$
$$(EF) \cup G = (E \cup G)\,(F \cup G)$$

- ***Leggi di De Morgan***

$$(E \cup F)^C = E^C\, F^C$$
$$(EF)^C = E^C \cup F^C$$

- ***Legge empirica del caso***
  Ripetendo N volte un esperimento nelle medesime condizioni, all'aumentare di N la frequenza relativa con cui un evento E si verifica tende a divenire costante.

- ***Teoria assiomatica della probabilità***
  Ad ogni evento E che appartiene ad un universo S viene associato un numero P(E) che è in accordo con i seguenti assiomi:

  ASSIOMA 1: $0 \le P(E) \le 1$

  ASSIOMA 2: $P(S) = 1$

  ASSIOMA 3: $\forall$ sequenza di eventi che si escludano mutuamente ($E_i \cap E_j = \varnothing\ \forall\ i \neq j$)

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

P(E) = probabilità dell'evento E

- ***Teorema di Bernulli o Legge dei grandi numeri***
  In una serie di prove un evento, che abbia probabilità costante P in ciascuna di esse, si presenti f volte:
  la probabilità della disuguaglianza

$$\left| P - \frac{f}{n} \right| < \varepsilon$$

  con $\varepsilon > 0$ comunque assegnato, tende ad uno con il crescere del numero n di prove.

Proposizioni derivanti dagli assiomi

$$P(E^C) = 1 - P(E)$$

$$E \cup E^C = S$$

essendo E ed E^C insiemi disgiunti

$$P(E \cup E^C) = P(E) + P(E^C) = P(S) = 1$$

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

infatti

$$E \cup F = EF^C \cup E^C F \cup EF$$

essendo EF^C, $E^C F$, EF insiemi disgiunti

$$P(E \cup F) = P(EF^C) + P(E^C F) + P(EF)$$

si ha inoltre:

$$E = EF \cup EF^C$$

$$F = EF \cup E^C F$$

Essendo gli insiemi dei secondi membri di ogni relazione disgiunti:

$$P(E) = P(EF) + P(EF^C)$$

$$P(F) = P(EF) + P(E^C F)$$

per cui:

$$P(E \cup F) = P(E) – P(EF) + P(F) – P(EF) + P(EF) = P(E) + P(F) – P(EF)$$

- ***Spazi equiprobabili***

$S = \{1, 2, \ldots N\}$
$P(\{1\}) = \ldots = P(\{N\}) = p$

Per il 2° ed il 3° assioma si ha:

$$1 = P(s) = P(\{1\}) + \ldots + P(\{N\} = N \cdot p$$

ovvero:

$$P(\{i\}) = p = \frac{1}{N}$$

Quindi in base all'assioma 3 per ogni evento si verifica:

$$P(E) = \frac{\text{numero di punti in E}}{N}$$

- ***Probabilità condizionata***

Si calcola la probabilità che un evento E si verifichi essendosi verificato l'evento F: P(E/F).
In uno spazio equiprobabile risulta:

$$P(E/F) = \frac{n\,(EF)}{n(F)}$$

Sia $n(E_i)$ = numero di punti dell'evento E_i;

Se N è il numero finito di punti dell'evento S, dividendo numeratore e denominatore per N si ottiene:

$$P(E/F) = \frac{P\,(EF)}{P(F)}$$

quindi

$$P(EF) = P(E/F)\,P(F)$$

- ***Eventi indipendenti***

Si suppone che due eventi siano tali da avere

$$P(E/F) = P(E)$$

con P(E) e P(F) maggiori di zero.

Si dice che l'evento E è indipendente dall'evento F e la condizione di indipendenza è:

$$P(EF) = P(E)\,P(F)$$

Se E è indipendente da F, anche F è indipendente da E, infatti:

$$P(EF) = \frac{P\,(FE)}{P\,(E)} = P(F)$$

Se E ed F sono indipendenti, lo sono anche E ed F^C, infatti:

$$E = EF \cup EF^C$$

I due insiemi a secondo membro sono disgiunti, quindi:

$$P(E) = P(EF) + P(EF^C) = P(E)P(F) + P(EF^C)$$

per cui

$$P(EF^C) = P(E)[1-P(F)] = P(E)\,P(F^C)$$

- ***Formula di Bayes***

Dati due eventi E ed F

$$E = EF \cup EF^C$$

$$P(E) = P(EF) + P(EF^C) =$$

$$= P(E/F)\ P(F) + P(E/F^C)\ P(F^C) =$$

$$= P(E/F)\ P(F) + P(E/F^C)\ [1-P(F)])$$

ovvero la probabilità di un evento E risulta dalla media pesata della probabilità condizionata di E dato che F si sia verificato, e della probabilità condizionata di E dato che F non si sia verificato.

VARIABILI CASUALI

Si consideri una quantità variabile X che possa assumere valori argomentali $x_1, \ldots\ldots, x_n$, ciascuno dei quali può essere estratto a caso dalla popolazione di tutti i valori possibili, con probabilità:

$P_1, \ldots\ldots, P_n$ tali che $\sum_{i=1}^{n} p_i = 1$ $\qquad X \begin{cases} x_1, \ldots\ldots, x_n \\ \\ p_1, \ldots\ldots, p_n \end{cases}$

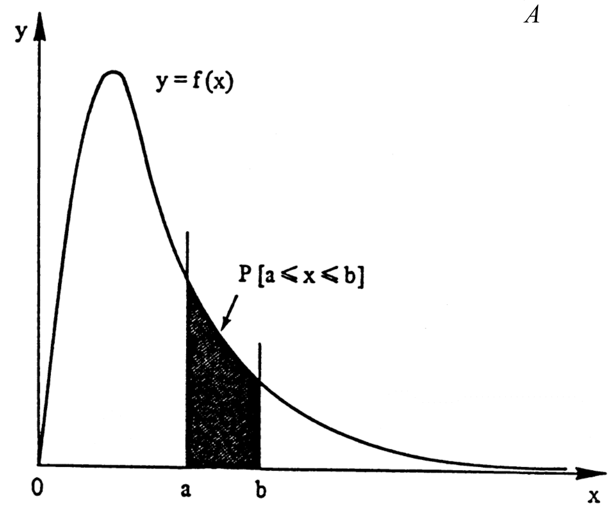
- ***Distribuzioni discrete e continue***

Se X assume valori in numero finito

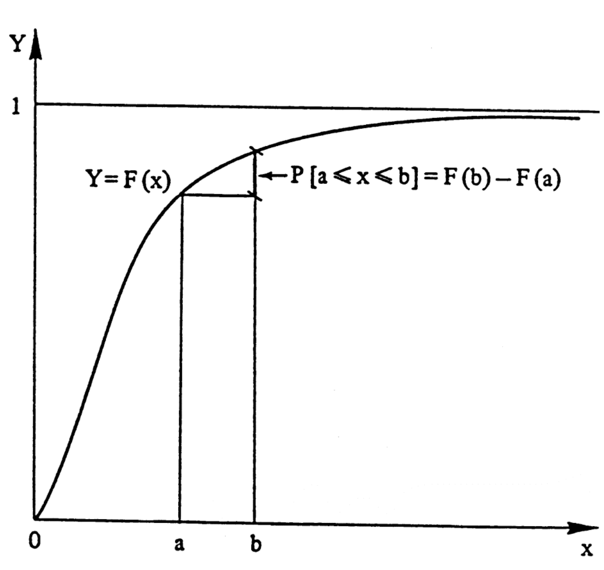
$$F(x_i) = P(x \leq x_i) = \sum_{j=1}^{i} P(x_j)$$

X è continua se la sua funzione di distribuzione è ovunque continua, e se ammetterà la derivata continua nell'intervallo di definizione.

$$F(x) = P(x_0 \leq x) = \int_A^x f(t) \qquad\qquad dp = f(x)\,dx$$



Densità di probabilità di una variabile casuale continua



Distribuzione di una variabile casuale continua

MOMENTI

- ***Momenti di ordine k rispetto al polo θ***

$$M_{k,\theta}(x) = \sum_{i=1}^{n} (x_i - \theta)^k p_i; \qquad M_{k,\theta}(x) = \int_{-\infty}^{+\infty} (x-\theta)^k f(x)dx$$

- ***Media***

$$M_k = \sum_{i=1}^{n} x_i p_i ; \qquad M_k = \int_{-\infty}^{+\infty} xf(x)dx$$

- ***Valore quadratico medio***

$$M_2 = \sum_{i=1}^{n} x_i^2 p_i ; \qquad M_2 = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

- ***Varianza***

$$\sigma_x^2 = \sum_{i=1}^{n} (x_i - M)^2 p_i ; \qquad \sigma_x^2 \int_{-\infty}^{+\infty} (x-M)^2 f(x)dx$$

$$\sigma^2 = M_2(x) - M^2$$

- ***Indice di simmetria***

$$\gamma = \frac{\sum_{i=1}^{n} (x_i - M)^3 p_i}{\sigma^2} \qquad \gamma = \frac{\int_{-\infty}^{+\infty} (x_i - M)^3 f(x)dx}{\sigma^2}$$

MEDIA** = **valore del polo che rende minimo il valore di momento del secondo ordine

Detto M il valore medio di una variabile casuale, e σ la deviazione standard, la probabilità che un individuo estratto a caso dalla popolazione possegga un valore argomentale contenuto nell'intervallo [M - λσ; M + λσ], dove λ è un numero qualsiasi maggiore di 1,

è maggiore od uguale a $1 - \frac{1}{\lambda^2}$

VARIABILI CASUALI A DUE DIMENSIONI

Analogamente si hanno:

$$x \begin{cases} x_1, \ldots\ldots, x_m \\ \\ p_1, \ldots\ldots, p_m \end{cases} \qquad \sum_{i=1}^{m} p_i = 1$$

$$y \begin{cases} y_1, \ldots\ldots, y_n \\ \\ q_1, \ldots\ldots, q_n \end{cases} \qquad \sum_{i=1}^{n} q_i = 1$$

- ***Indipendenza***

$$\phi_{ij} = q_i \cdot p_j$$

- ***Variabile casuale somma***

x, y, …, w
z = x + y +…+w

Si consideri la successione ordinata di tutte le somme possibili $x_i + y_j + w_k$ purché diverse fra loro; se le variabili sono indipendenti, la probabilità di una generica somma è:

$$P_i^{(x)} P_j^{(y)} P_k^{(w)}$$

- ***Variabile casuale prodotto***

x, y, …, w
z = x * y *…* w

successione ordinata di tutti i possibili prodotti $x_i * y_j * w_k$ purché diverse fra loro;

- ***Proprietà***

I) $M(kx)=k\,M_x$

II) $M\,(x+y+...+w) = M_x + M_y + \ldots + M_w$

III) $M\,(x\,y\ldots w) = M_x\ \ M_y\ \ \ldots\ M_w$

Nel caso in cui le variabili sono indipendenti

IV) $M_2\,(x+y+...+w) = M_x + M_y + \ldots + M_w$

Nel caso di variabili indipendenti con valore medio nullo

V) Scarti di variabili casuali indipendenti

$$\sigma^2(x+y+....+w)=\sigma_x^2\ \ +\sigma_y^2\ \ +\qquad+\sigma_w^2$$

V_{bis}) $Z = a_1\,x + a_2\,y + \ldots.. + a_n\,w + b$

$$\sigma_Z^2\ \ = a_1^2\sigma_x^2\ \ + a_2^2\sigma_y^2\ \ + \ldots\qquad + a_n^2\sigma_w^2$$

Legge di propagazione degli scarti di n variabili casuali indipendenti