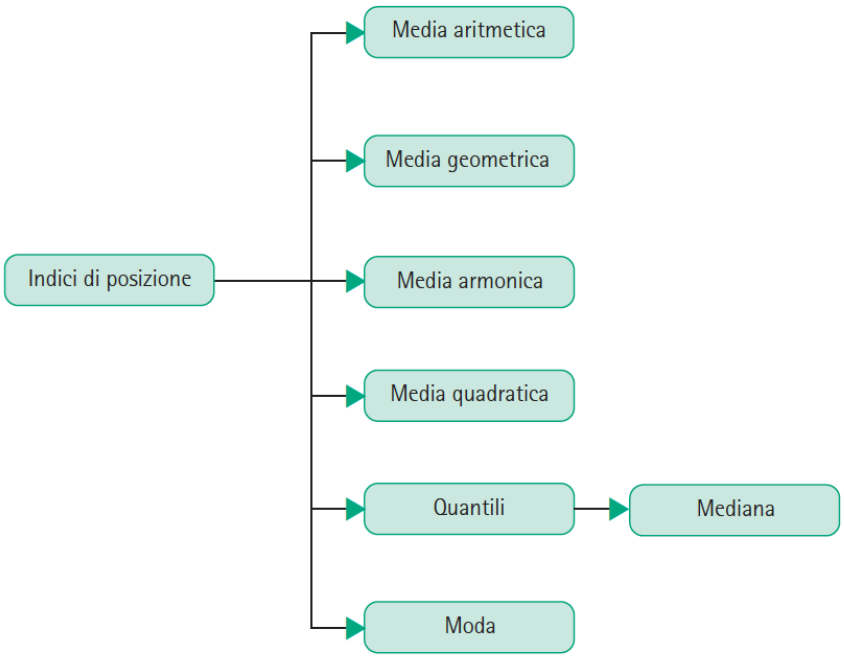
4. GLI INDICI DI POSIZIONE, DI VARIABILITÀ E DI FORMA



1) INDICI DI POSIZIONE

Gli **indici di posizione** possono essere ricavati:

— da tutti i termini della serie stessa, come la *media aritmetica*, la *media geometrica*, la *media armonica*, la *media quadratica*;

— in base alla scelta di uno dei termini della serie, che si differenzia dagli altri per alcune caratteristiche particolari, come la *mediana*, la *moda* e i *quantili*.

Di seguito esponiamo gli indici di posizione suddetti con esempi numerici che ne chiariscono il significato.

MEDIA ARITMETICA

È la *media più usata*. Quando si parla di media, senza ulteriori specificazioni, ci si vuole riferire alla media aritmetica.
La **media aritmetica** è eguale alla *somma dei termini divisa per il loro numero.* Se si indicano gli n termini di una serie con x_1, x_2, …, x_n, la media aritmetica (M) è:

$$M = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Il simbolo Σ, detto *sommatoria*, si ritrova spessissimo nelle formule statistiche, indica la somma dei termini considerati (nel nostro esempio la somma dei vari x_1, x_2 etc.).
Gli indici posti sopra e sotto il simbolo indicano che i assume tutti i valori da 1 a n, e quindi per $\sum_{i=1}^{n} x_i$ deve intendersi la sommatoria di *tutti* i termini della serie (già indicati nel loro numero massimo con n).

Esempio

Dati i valori 1, 2, 3, 7, 10, 13, specificando la formula sopra enunciata, la media aritmetica è:

$$M = \frac{1+2+3+7+10+13}{6} = \frac{36}{6} = 6$$

IL MATERIALE È STATO TRATTO DA "**STATISTICA...IN TASCA**", EDITO DA
che ne ha concesso la pubblicazione su **www.maturansia.it**
EDIZIONI SIMONE
Gruppo Editoriale **Simone**
Copyright © 2015 Simone S.r.l.

Un aspetto che bisogna considerare quando si parla di media aritmetica è che quest'ultima non esprime realmente l'andamento del fenomeno quando nella distribuzione dei dati è presente un valore eccezionale che, di fatto, rende non significativo l'indice considerato.

MEDIA GEOMETRICA

La **media geometrica** si ottiene estraendo la radice ennesima del prodotto degli *n* termini:

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot x_4 \ldots \cdot x_n}$$

Il che si scrive graficamente così:

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Il simbolo Π indica il *prodotto* dei termini da 1 a n, rappresentati dalle x_i.

Esempio

La media geometrica dei valori 3, 5, 7, 9, 13, 17, 21 è:

$$G = \sqrt[7]{3 \cdot 5 \cdot 7 \cdot 9 \cdot 13 \cdot 17 \cdot 21} = \sqrt[7]{4.385.745} \cong 8,889$$

La media geometrica è usata nei casi in cui i termini raccolti sono tra loro in progressione geometrica (esatta o solo approssimata). È particolarmente usata per i fenomeni economici.

MEDIA ARMONICA

La **media armonica** è pari al reciproco della media aritmetica dei reciproci dei termini.

Se i termini sono x_1, x_2, x_3, x_4, …, i reciproci sono $\frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_3}, \frac{1}{x_4}$ per cui la media armonica è:

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

Esempio

Se i termini sono 1, 2, 3, 7, 8, 9, si considerano i reciproci:

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{7}, \frac{1}{8}, \frac{1}{9}$$

La media armonica è:

$$H = \frac{6}{1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9}} = 2,712$$

IL MATERIALE È STATO TRATTO DA "**STATISTICA...IN TASCA**", EDITO DA
che ne ha concesso la pubblicazione su **www.maturansia.it**
EDIZIONI SIMONE
Gruppo Editoriale **Simone**
Copyright © 2015 Simone S.r.l.

La media armonica si applica quando il fenomeno che si indaga tende a svilupparsi in progressione armonica: cioè quando i reciproci dei termini sono tra loro in progressione aritmetica (presentano, quindi, una differenza costante fra termini consecutivi). Oppure si usa ancora quando non si vuole ricercare una sintesi del fenomeno esaminato, bensì di un altro fenomeno che costituisce il reciproco del primo.
Ad esempio, i rapporti fra prezzo e potere di acquisto della moneta.

MEDIA QUADRATICA

La **media quadratica** si ottiene estraendo la radice quadrata della media aritmetica dei quadrati dei termini.

Dati i termini x_1, x_2, x_3, ..., x_n la media quadratica è:

$$Q = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2}{n}}$$

Il che, in forma sintetica si scrive:

$$Q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

Esempio

Se i termini sono 1, 1, 2, 2, 3, 3, 4, 4, 5 la media quadratica è:

$$Q = \sqrt{\frac{1+1+4+4+9+9+16+16+25}{10}} = \sqrt{\frac{85}{10}} = 2{,}915$$

La media quadratica è denominata anche *media di precisione*, ed è adoperata tutte le volte che alle differenze tra i termini e il valore medio si dà il significato di *deviazione* o *errore* dal *valore esatto*: essa trova applicazione, in pratica, nel procedimento di ricerca e correzione degli errori.

I QUANTILI

I **quantili** sono le intensità che dividono una distribuzione in due parti, lasciando da una parte una quota dei casi e dall'altra la rimanente.
I quantili si dicono *quartili* se dividono la distribuzione in parti uguali, tali che:

— il primo quartile (Q_1) lascia alla sua sinistra il 25% dei casi e alla sua destra il rimanente 75% dei casi;
— il secondo quartile (Q_2), che coincide con la mediana, lascia alla sua sinistra il 50% dei casi e alla sua destra il rimanente 50%;
— il terzo quartile (Q_3) lascia alla sua sinistra il 75% dei casi e alla sua destra il rimanente 25%.

MEDIANA

La **mediana** è l'indice di posizione che, ordinate le osservazioni di un fenomeno in senso crescente, divide la distribuzione in due gruppi di uguale numerosità. Pertanto, al primo gruppo appartengono le osservazioni inferiori o uguali alla mediana, mentre al secondo gruppo appartengono le osservazioni superiori o uguali alla mediana.

Dato un insieme di n intensità x_1, x_2, …, x_n, la determinazione della mediana è diversa a seconda che sia n dispari o pari; infatti:

— se n è dispari, la mediana è l'intensità individuata dal posto centrale:

$$C = \frac{n+1}{2}$$

Copyright © 2015 Simone S.r.l.

— se *n* è pari, la mediana è data dalla semisomma delle intensità individuate dai due posti centrali:

$$C_1 = \frac{n}{2} \quad \text{e} \quad C_2 = \frac{n}{2} + 1$$

Esempio 1

La mediana delle intensità 3, 15, 9, 2, 6, 12, 5 si ottiene dapprima ordinandole in serie: 2, 3, 5, 6, 9, 12, 15. Sicome le intensità sono 7 (numero dispari), il posto centrale è unico ed è pari a:

$$C = \frac{7+1}{2} = 4$$

La mediana è, quindi, l'intensità individuata dal 4° posto, vale a dire che essa è uguale a:

$$Me = 6$$

Esempio 2

La mediana delle intensità 7, 16, 2, 3, 9, 12, 15, 5, si ottiene dapprima ordinandole in serie: 2, 3, 5, 7, 9, 12, 15, 16. Siccome esse sono 8 (numero pari), i posti centrali sono due e sono:

$$C_1 = \frac{8}{2} = 4 \text{ e } C_2 = \frac{8}{2} + 1 = 5$$

pertanto, la mediana è data dalla semisomma delle intensità individuate dal 4° e dal 5° posto, vale a dire che essa è uguale a:

$$Me = \frac{7+9}{2} = 8$$

MODA

La **moda**, in una successione di dati statistici, è data da quel termine che si presenta il maggior numero di volte.
Può capitare che un insieme di dati abbia una sola moda (*distribuzione unimodale*), o non ne abbia alcuna (*distribuzione zeromodale*).

Esempio

Sia data la successione dei termini 3, 5, 9, 3, 5, 7, 3, 2, 9, 3, 4, 3, 6, la moda è il termine *3* in quanto esso compare ben cinque volte.

2) MEDIE PONDERATE

Ciascuna delle medie esaminate può essere *semplice* o *ponderata*. È *semplice* quando (come si è fatto nell'applicare le medie prima esaminate) tutti i termini concorrono alla sua determinazione una sola volta. È *ponderata* quando i termini entrano nel calcolo un numero (intero o decimale) diverso di volte, a seconda del *peso* che deve esser loro attribuito.

MEDIA ARITMETICA PONDERATA

La formula della **media aritmetica ponderata** è la seguente:

$$M = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}$$

dove x_i sono i termini con i=1, 2, …, n, e n_i sono le frequenze con cui essi si presentano.

IL MATERIALE È STATO TRATTO DA "**STATISTICA...IN TASCA**", EDITO DA
che ne ha concesso la pubblicazione su **www.maturansia.it**
EDIZIONI SIMONE
Gruppo Editoriale **Simone**
Copyright © 2015 Simone S.r.l.

Esempio

Data la seguente tabella:

Tabella 1

x_i	1	2	3	4	5
n_i	2	3	2	2	1

la media aritmetica ponderata si ottiene rapportando la somma dei prodotti xn_i alla somma delle frequenze n_i:

$$M=\frac{(1\cdot 2)+(2\cdot 3)+(3\cdot 2)+(4\cdot 2)+(5\cdot 1)}{2+3+2+2+1}=2{,}7$$

MEDIA GEOMETRICA PONDERATA

La formula della **media geometrica ponderata** è la seguente:

$$G=\sqrt[\sum_{i=1}^{k} n_i]{x_1^{n_1}\cdot x_2^{n_2}\cdot \ldots \cdot x_k^{n_k}}$$

O sinteticamente:

$$G=\sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} x_i^{n_i}}$$

Esempio

Considerando la tabella 1, la media geometrica ponderata è:

$$G=\sqrt[10]{1^2\cdot 2^3\cdot 3^2\cdot 4^2\cdot 5^1}=2{,}38$$

MEDIA ARMONICA PONDERATA

La formula della **media armonica ponderata** è la seguente:

$$H=\frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{1}{x_i} n_i}$$

Esempio

Considerando la tabella 1, la media armonica ponderata è:

$$H=\frac{10}{1\cdot 2+\frac{1}{2}\cdot 3+\frac{1}{3}\cdot 2+\frac{1}{4}\cdot 2+\frac{1}{5}\cdot 1}=2{,}05$$

IL MATERIALE È STATO TRATTO DA "**STATISTICA...IN TASCA**", EDITO DA
che ne ha concesso la pubblicazione su **www.maturansia.it**

EDIZIONI SIMONE®
Gruppo Editoriale **Simone**

Copyright © 2015 Simone S.r.l.

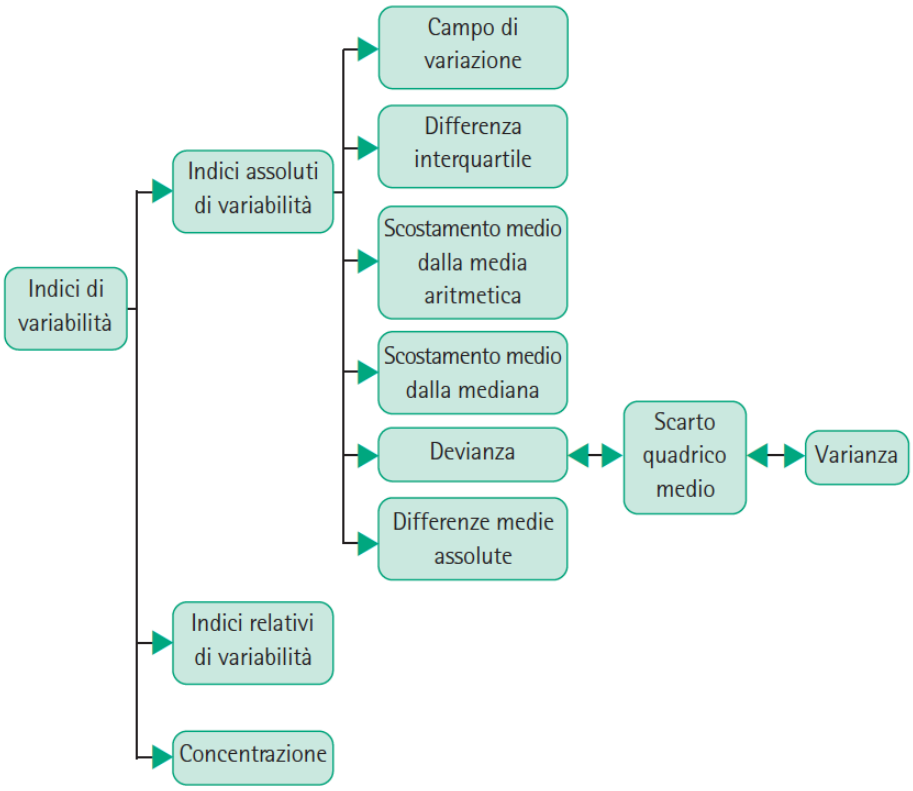
MEDIA QUADRATICA PONDERATA

La formula della **media quadratica ponderata** è la seguente:

$$Q = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i}}$$

Anche per la mediana, per la moda e per i quantili esistono le formule da applicare a distribuzioni di frequenze; tali argomenti esulano, tuttavia, dallo scopo della presente trattazione.

3) INDICI DI VARIABILITÀ



Le medie facilitano la comprensione della *frequenza* dei fenomeni collettivi. Per una più esatta conoscenza di essi occorre però anche studiare la loro *variabilità*, cioè la capacità di assumere differenti valori *quantitativi* in un certo periodo di tempo, o in seguito all'influenza di un altro (o di altri) fenomeno (fenomeni).

DISPERSIONE E DISUGUAGLIANZA

Nella metodologia statistica si distinguono due aspetti della *variabilità*: la *dispersione* e la *disuguaglianza*. La prima pone in rilievo quanto, mediamente, le varie intensità differiscono da una media prestabilita, la seconda evidenzia, invece, la diversità delle varie intensità tra loro.

IL MATERIALE È STATO TRATTO DA "**STATISTICA...IN TASCA**", EDITO DA
che ne ha concesso la pubblicazione su **www.maturansia.it**
EDIZIONI SIMONE®
Gruppo Editoriale **Simone**
Copyright © 2015 Simone S.r.l.

INDICI ASSOLUTI E INDICI RELATIVI

La variabilità si calcola ricorrendo sia agli *indici assoluti di variabilità*, che sono espressi nella stessa unità di misura del fenomeno di cui calcolano la dispersione o la disuguaglianza, sia agli *indici relativi di variabilità* che prescindono dall'unità di misura, per cui sono particolarmente adatti per effettuare confronti tra fenomeni diversi.

3.1 INDICI ASSOLUTI DI VARIABILITÀ

CAMPO DI VARIAZIONE

È l'indice assoluto di variabilità più semplice da calcolarsi. Esso si ottiene dalla differenza tra il valore massimo e quello minimo delle intensità del fenomeno; in simboli:

$$\omega = x_{\max} - x_{\min}$$

Esempio

Se le intensità di un dato fenomeno sono le seguenti:

29, 24, 20, 28, 22, 22, 23

il campo di variazione è:

$$\omega = 29 - 20 = 9$$

L'indice di variabilità in questione è poco utilizzato in quanto è influenzato da valori estremi eccezionali e non tiene conto della struttura della distribuzione dei dati.

DIFFERENZA INTERQUARTILE

È la differenza tra il terzo e il primo quartile di un insieme di osservazioni, in simboli:

$$\delta_Q = Q_3 - Q_1$$

SCOSTAMENTO MEDIO DALLA MEDIA ARITMETICA

Si ottiene calcolando la media aritmetica dei valori assoluti degli n scarti tra le singole intensità x_i, $i = 1, 2, \ldots, n$, e la media aritmetica M; in simboli:

$$S_M = \frac{\sum_{i=1}^{n} |x_i - M|}{n}$$

Esempio

Tenendo conto delle intensità considerate per il campo di variazione, la media aritmetica è 24, per cui lo scostamento medio dalla media aritmetica è:

$$S_M = \frac{|29-24|+|24-24|+|20-24|+|28-24|+|22-24|+|22-24|+|23-24|}{7} = \frac{18}{7} = 2{,}57$$

La formula, *ponderata*, dello scostamento medio dalla media aritmetica è:

$$S_M = \frac{\sum_{i=1}^{k} |x_i - M| n_i}{\sum_{i=1}^{k} n_i}$$

Copyright © 2015 Simone S.r.l.

SCOSTAMENTO MEDIO DALLA MEDIANA

Si ottiene calcolando la media aritmetica dei valori assoluti degli n scarti tra le singole intensità x_i, i=1, 2, …, n, e la mediana Me; in simboli:

$$S_{Me} = \frac{\sum_{i=1}^{n} |x_i - Me|}{n}$$

Esempio

Considerando sempre le intensità 29, 24, 20, 28, 22, 22, 23, siccome il loro numero $n = 7$ è dispari, la mediana coincide con il valore centrale ottenuto ponendo i termini in ordine non decrescente, vale a dire, 20, 22, 22, 23, 24, 28, 29, essa è pari all'intensità che occupa il 4° posto, per cui è uguale a $Me = 23$, pertanto lo scostamento medio dalla mediana è:

$$S_{Me} = \frac{|29-23|+|24-23|+|20-23|+|28-23|+|22-23|+|22-23|+|23-23|}{7} = \frac{17}{7} = 2{,}43$$

La formula, *ponderata*, dello scostamento medio dalla mediana è:

$$S_{Me} = \frac{\sum_{i=1}^{k} |x_i - Me| n_i}{\sum_{i=1}^{k} n_i}$$

DEVIANZA

Si ottiene calcolando la somma dei quadrati degli n scarti tra le singole intensità x_i, i=1, 2, …, n, e la media aritmetica M; in simboli:

$$D(X) = \sum_{i=1}^{n} (x_i - M)^2$$

Esempio

Considerando le intensità viste nell'esempio sul campo di variazione, gli scarti dalla media aritmetica sono:

5, 0, –4, 4, –2, –2, –1

per cui la devianza è:

$$D(X) = (5)^2 + (-4)^2 + (4)^2 + (-2)^2 + (-2)^2 + (-1)^2 = 66$$

La formula, *ponderata*, della devianza è:

$$D(X) = \sum_{i=1}^{k} (x_i - M)^2 n_i$$

SCARTO QUADRATICO MEDIO

Si ottiene dalla radice quadrata del rapporto tra la devianza e la numerosità n del collettivo; in simboli:

$$\sigma = \sqrt{\frac{D(X)}{n}} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - M)^2}{n}}$$

IL MATERIALE È STATO TRATTO DA "**STATISTICA...IN TASCA**", EDITO DA
che ne ha concesso la pubblicazione su **www.maturansia.it**
EDIZIONI SIMONE
Gruppo Editoriale **Simone**
Copyright © 2015 Simone S.r.l.

Segui **Edizioni Simone** e **Maturansia** su Instagram

Esempio

Considerando le intensità viste nell'esempio sulla devianza, lo scarto quadratico medio è:

$$\sigma = \sqrt{\frac{66}{7}} = 3{,}07$$

La formula, ponderata, dello scarto quadratico medio è:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - M)^2 n_i}{\sum_{i=1}^{k} n_i}}$$

VARIANZA

È il quadrato dello scarto quadratico medio, in simboli:

$$\sigma^2 = \frac{D(X)}{n}$$

Esempio

Considerando le intensità viste nell'esempio sulla devianza, la varianza è pari a:

$$\sigma^2 = (3{,}07)^2 = 9{,}42$$

La formula, *ponderata*, della varianza è:

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - M)^2 n_i}{\sum_{i=1}^{k} n_i}$$

DIFFERENZE MEDIE ASSOLUTE

Evidenziano la disuguaglianza tra le singole intensità; come caso particolare, in questo contesto, ci occuperemo di differenza media con ripetizione e di differenza media senza ripetizione.

— *Differenza media con ripetizione*

Si ottiene calcolando le differenze tra le singole intensità, considerando anche quelle di ciascun termine con se stesso, per cui, se n sono i termini, i confronti da farsi sono n^2, e la differenza media con ripetizione si ottiene rapportando la somma dei valori assoluti di tali differenze a n^2; in simboli:

$$\Delta_R = \frac{\sum_{i,j} |x_i - x_j|}{n^2}$$

in cui le coppie (i, j) comprendono anche i confronti $i=j$.

— *Differenza media senza ripetizione*

Si ottiene in modo analogo alla differenza media con ripetizione, tuttavia si escludono gli n confronti di ciascun termine con se stesso, pertanto i confronti sono $n^2 - n = n \cdot (n - 1)$; in simboli:

$$\Delta = \frac{\sum |x_i - x_j|}{n \cdot (n-1)}$$

Gli indici di posizione, di variabilità e di forma

IL MATERIALE È STATO TRATTO DA "STATISTICA...IN TASCA", EDITO DA
che ne ha concesso la pubblicazione su www.maturansia.it

EDIZIONI SIMONE
Gruppo Editoriale Simone

Copyright © 2015 Simone S.r.l.

Esempio

Date le intensità 23, 27, 35, 48 per calcolare l'indice di variabilità in questione si costruisce una matrice quadrata in cui si riportano tutte le differenze tra le varie intensità e nella cui diagonale, chiaramente, figurano tutti zeri, rappresentando essa i confronti di ciascun termine con se stesso:

	23	27	35	48
23	0	– 4	– 12	– 25
27	4	0	– 8	– 21
35	12	8	0	– 13
48	25	21	13	0

Siccome nelle caselle al di sopra della diagonale figurano differenze uguali, in valore assoluto, alle differenze poste nelle caselle simmetriche, al di sotto della diagonale, è sufficiente considerare solo le differenze (in valore assoluto) al di sopra, o al di sotto, della diagonale, moltiplicarle per due e dividere il prodotto ottenuto per $4^2 = 16$ (numero totale dei confronti):

$$\Delta_R = \frac{2\cdot(4+12+25+8+21+13)}{16} = \frac{166}{16} = 10{,}375$$

— Differenza media senza ripetizione

Si ottiene in modo analogo alla differenza media con ripetizione, tuttavia si escludono gli n confronti di ciascun termine con se stesso, pertanto i confronti sono $n^2 - n = n \cdot (n - 1)$; in simboli:

$$\Delta = \frac{\sum_{i,j} |x_i - x_j|}{n\cdot(n-1)}$$

Il numeratore della differenza media senza ripetizione coincide con il numeratore della differenza media con ripetizione.

Esempio

La differenza media con ripetizione vista nell'esempio precedente ha per numeratore 166; è sufficiente dividere 166 per $4 \cdot 3 = 12$, per ottenere l'indice di variabilità in questione:

$$\Delta = \frac{166}{12} = 13{,}83$$

3.2 INDICI RELATIVI DI VARIABILITÀ

Tutti gli indici assoluti di variabilità sono espressi in funzione dell'unità di misura in cui sono espressi i termini della serie.
Per ottenere un indice variabilità non espresso in alcuna unità di misura si pone in rapporto l'indice assoluto di variabilità con la **media** del fenomeno da cui è tratto: si ottiene così l'*indice relativo di variabilità.* Così, se porremo in rapporto con la media aritmetica del fenomeno il *campo di variazione*, lo *scostamento medio* etc., avremo il *campo di variazione relativo* o lo *scostamento medio relativo* etc.
Altre volte, invece, si rapporta un indice assoluto di variabilità al valore che lo stesso assume in caso di **massima variabilità**: il coefficiente che si ottiene varierà tra 0 (assenza di variabilità) e 1 (massima variabilità).

3.3 LA CONCENTRAZIONE

La **concentrazione** è un aspetto della variabilità, essa concerne esclusivamente un carattere **additivo** quindi trasferibile, come il reddito.

Sia dato un carattere *X*, esso si dice *concentrato* se l'ammontare complessivo dello stesso è posseduto da un numero ridotto di unità statistiche. Si parla di:

IL MATERIALE È STATO TRATTO DA "**STATISTICA...IN TASCA**", EDITO DA
che ne ha concesso la pubblicazione su **www.maturansia.it**
EDIZIONI SIMONE®
Gruppo Editoriale **Simone**
Copyright © 2015 Simone S.r.l.

— **concentrazione nulla** (o **equidistribuzione**) quando tutte le unità possiedono il carattere nella stessa misura;
— **concentrazione massima** quando una sola unità possiede l'intero ammontare del carattere.

Si supponga di disporre delle *n* modalità di una variabile statistica *X* (reddito, patrimonio etc.) tali che:

$$x_1 \leq x_2 \leq \ldots \leq x_n$$

Si definiscono **frazioni cumulate delle *n* unità osservate** le seguenti quantità:

$$p_1 = \frac{1}{n}; p_2 = \frac{2}{n}; \ldots; p_i = \frac{i}{n}; \ldots; p_n = \frac{n}{n} = 1$$

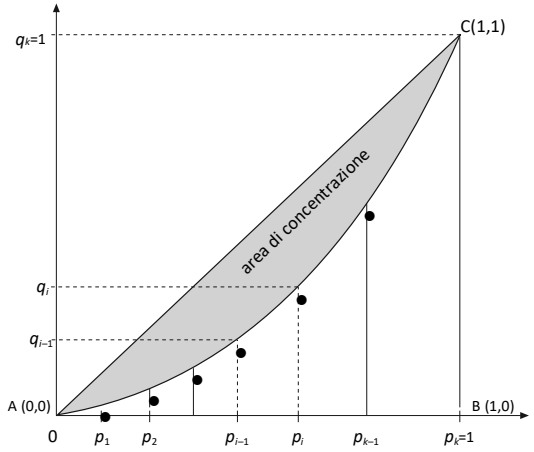
Ciascun termine, p_i, $i = 1, 2, \ldots, n$, rappresenta la frazione cumulata delle prime *i* unità statistiche.

Si definiscono, inoltre, **frazioni cumulate del carattere** posseduto dalle prime unità statistiche le seguenti quantità:

$$q_1 = \frac{x_1}{\sum_{j=1}^{n} x_j}; q_2 = \frac{\sum_{j=1}^{2} x_j}{\sum_{j=1}^{n} x_j}; \ldots; q_i = \frac{\sum_{j=1}^{i} x_j}{\sum_{j=1}^{n} x_j}; \ldots; q_n = \frac{\sum_{j=1}^{n} x_j}{\sum_{j=1}^{n} x_j} = 1$$

CURVA DI LORENZ

Una rappresentazione grafica di tali frazioni è stata proposta dallo statistico statunitense Lorenz nel 1905, attraverso una curva detta **curva di Lorenz**, o **curva di concentrazione**, che si ricava ponendo, in un sistema di assi cartesiani, sulle ascisse i valori p_i e sulle ordinate i valori q_i, ottenendo una serie di punti che, uniti, formano la curva di concentrazione di un dato carattere.
La **retta di equidistribuzione** è la retta che congiunge l'origine (0, 0) con il punto (1, 1).
L'area compresa tra la curva di equidistribuzione e quella di concentrazione è denominata **area di concentrazione**.
Quando la concentrazione del fenomeno analizzato è **massima**, tutto l'ammontare del carattere risulta concentrato in una sola unità, quindi tutti i punti della curva ricadono sull'asse delle ascisse tranne l'ultimo di coordinate (1, 1). In tal caso, la curva di Lorenz assume la forma di un triangolo con vertici ABC. Pertanto, si può affermare che l'area di concentrazione assume valori compresi tra 0 (**equidistribuzione**) e 1/2 (**massima concentrazione**).



(Fig. 1)

Infatti, dall'andamento della curva di Lorenz è possibile desumere la minore o maggiore concentrazione del fenomeno studiato; quanto più la curva si avvicina alla retta di equidistribuzione tanto più si riduce la concentrazione del fenomeno e viceversa quanto più questa si allontana tanto più aumenta la concentrazione.

Copyright © 2015 Simone S.r.l.

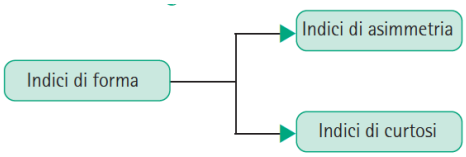
RAPPORTO DI CONCENTRAZIONE DI GINI

Nel 1914 lo statistico Gini si occupò delle differenze $\frac{p_i - q_i}{p_i}$ e constatò che, al crescere di tali differenze, cresceva la disuguaglianza di un particolare carattere trasferibile: il reddito. Operò una media aritmetica ponderata di queste differenze normalizzate, attribuendo a ciascuna un peso p_i, ottenendo il seguente rapporto, detto **rapporto di concentrazione di Gini**:

$$R = \frac{\sum_{i=1}^{n-1}\left(\frac{p_i - q_i}{p_i}\right)p_i}{\sum_{i=1}^{n-1} p_i} = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

Si tratta di un indice variabile tra 0 e 1, che fornisce una misura sintetica della concentrazione di un carattere trasferibile.

4) INDICI DI FORMA



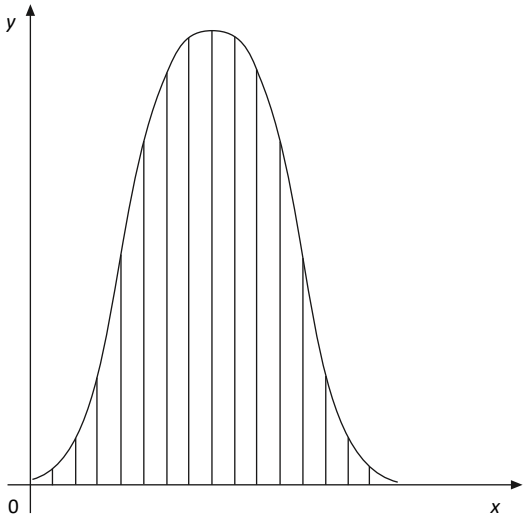
Alcuni aspetti della forma di una distribuzione sono: l'**asimmetria** e la **curtosi**.

4.1 Asimmetria

L'**asimmetria** è un termine statistico che indica l'assenza di specularità di una distribuzione rispetto al suo asse di simmetria, per cui i valori del carattere di una distribuzione asimmetrica sono distribuiti con frequenze differenti attorno al suo valore centrale.

Graficamente

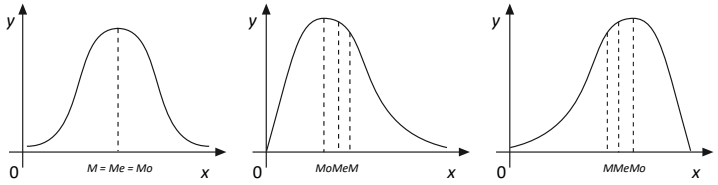
Per valutare l'eventuale asimmetria di una distribuzione si confronta la sua curva rappresentatrice con una curva di frequenza unimodale, simmetrica che assume la caratteristica forma a **campana**, e che è nota, in statistica, con la denominazione di **curva normale** o **gaussiana**. La sua rappresentazione è la seguente:



(Fig. 2)

Copyright © 2015 Simone S.r.l.

Rispetto alla curva normale è possibile evincere se una distribuzione presenta asimmetria, se ha una coda più lunga; precisamente, se la coda più lunga è a sinistra, la distribuzione presenta **asimmetria negativa** (o **a sinistra**), se, invece, la coda più lunga è a destra, allora la distribuzione presenta **asimmetria positiva** (o **a destra**).



Curva normale | Curva asimmetrica a destra | Curva asimmetrica a sinistra

(Fig. 3)

ANALITICAMENTE

Le relazioni esistenti tra media, mediana e moda consentono di verificare se una distribuzione si presenta simmetrica o asimmetrica; infatti, utilizzando la simbologia dei capitoli precedenti, si parla di:

— *simmetria* se $M = Me = Mo$;
— *asimmetria positiva* se $Mo < Me < M$, la distribuzione presenta il ramo destro più allungato di quello sinistro, in altre parole presenta una *coda verso destra*;
— *asimmetria negativa* se $M < Me < Mo$, la distribuzione presenta il ramo sinistro più allungato di quello destro, e si dice che presenta una *coda verso sinistra*.

INDICE DI ASIMMETRIA DI FISHER

Un indice relativo di asimmetria proposto da **Fisher** ha la seguente espressione analitica:

$$\alpha_F = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - M}{\sigma}\right)^3$$

L'espressione entro parentesi è definita **scarto standardizzato**, per cui l'indice di asimmetria è pari alla media dei cubi degli scarti standardizzati.

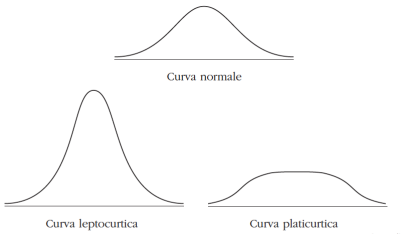
4.2 CURTOSI

La **curtosi** fa riferimento alla maggiore o minore gibbosità di una curva in prossimità del suo massimo e, quindi, alla maggiore o minore lunghezza delle code. La curtosi assume rilievo per una distribuzione di frequenza unimodale, la cui curva è di forma campanulare.

GRAFICAMENTE

Per valutare questo aspetto della forma di una curva, la stessa è paragonata ad una curva normale (detta anche **mesocurtica**) avente la stessa frequenza complessiva, la stessa media e lo stesso scarto quadratico medio; si dice che la curva è:

— **platicurtica** o **iponormale** se è più piatta della curva normale;
— **leptocurtica** o **ipernormale** se è più appuntita della curva normale.



Curva normale | Curva leptocurtica | Curva platicurtica

(Fig. 4)

IL MATERIALE È STATO TRATTO DA "**STATISTICA...IN TASCA**", EDITO DA
che ne ha concesso la pubblicazione su **www.maturansia.it**
EDIZIONI SIMONE®
Gruppo Editoriale **Simone**
Copyright © 2015 Simone S.r.l.

INDICE DI CURTOSI DI PEARSON

Per misurare la curtosi di una curva unimodale di forma campanulare è particolarmente utile l'**indice di curtosi di Pearson** la cui espressione analitica, per una serie, è la seguente:

$$\beta_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - M}{\sigma}\right)^4$$

L'indice:

— vale 3 per una curva normale;

— è maggiore di 3 per una distribuzione leptocurtica;

— è inferiore a 3 per una distribuzione platicurtica.

Per ottenere una misura paragonabile con lo zero, **Fisher** ha proposto un indice che si ottiene sottraendo all'espressione dell'indice di curtosi di Pearson il numero 3, ossia:

$$\gamma_2 = \beta_2 - 3 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - M}{\sigma}\right)^4 - 3$$

Esso vale 0 per una curva normale, è positivo o negativo per una curva, rispettivamente, più appuntita o meno appuntita di una curva normale.

Copyright © 2015 Simone S.r.l.