3

variabilità

“ Senza deviazione dalla norma il progresso non è possibile”

(Frank Zappa)

3. variabilità

Definizioni

Variabilità

E' l'attitudine di un fenomeno ad assumere diverse modalità.
Essa è misurata mediante indici di variabilità che si distinguono in indici di dispersione e di disuguaglianza.
Gli indici di dispersione misurano la distanza delle osservazioni rispetto ad un valore medio, mentre gli indici di disuguaglianza misurano la diversità tra le varie osservazioni.
Un'ulteriore distinzione degli indici di variabilità riguarda la variabilità assoluta e quella relativa.
Gli indici assoluti di variabilità sono la varianza, lo scarto quadratico medio, il campo di variazione, la differenza semplice media, ecc.
Gli indici relativi di variabilità sono gli indici di variabilità assoluta rapportati, ad esempio, al proprio valor medio (es. il coefficiente di variazione) oppure al proprio massimo. Il rapporto di concentrazione del Gini, ne è un caso particolare.

Il campo di variazione

E' un indice assoluto di variabilità dato dalla differenza tra il valore massimo e il valore minimo assunti dalle modalità di un carattere; in simboli: $W = x_s - x_1$.

Differenza interquartilica

Rappresenta la differenza tra due diversi quartili della distribuzione. Ad esempio lo scarto tra il terzo ed il primo quartile di una distribuzione, $\delta = Q_3 - Q_1$, è l'intervallo che comprende il 50 per cento delle osservazioni e quindi è una misura di variabilità della parte centrale di una distribuzione statistica.

Scostamento semplice medio dalla media aritmetica

Indice di variabilità dato dalla media aritmetica dei valori assoluti degli scarti dalla media aritmetica: $S_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$. Per distribuzioni di frequenza si ha: $S_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| n_i}{N}$.

Scostamento semplice medio dalla mediana

Indice di variabilità dato dalla media aritmetica dei valori assoluti degli scarti dalla mediana:

$S_{Me} = \frac{\sum_{i=1}^{n} |x_i - Me|}{n}$. Per distribuzioni di frequenza delle modalità, si ha: $S_{Me} = \frac{\sum_{i=1}^{n} |x_i - Me| n_i}{N}$.

Varianza

E' l'indice di variabilità più utilizzato in statistica. E' un indice quadratico e il suo campo di variazione è compreso tra 0 e +∞. Esso è definito come la media aritmetica del quadrato degli scarti dalla media aritmetica: $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$. Nel caso si abbia una distribuzione di frequenza si ha: $\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2 n_i}{N}$.

La varianza può essere calcolata come momento secondo meno il momento primo al quadrato o in altri termini come la media quadratica meno la media aritmetica al quadrato.
E'molto importante sottolineare che la varianza è un indice quadratico,pertanto il suo valore esprime il quadrato dell'unità di misura della variabile oggetto di studio.

Scarto quadratico medio (s.q.m.)

E' la radice quadrata della varianza. A differenza della varianza lo s.q.m. è esprimibile con la stessa unità di misura della variabile osservata.

La sua espressione analitica è: $\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$. Nel caso di distribuzione di frequenza si ha: $\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2 n_i}{N}}$.

Esso misura il grado di dispersione medio di una variabile attorno alla propria media aritmetica e,come dicevamo, il valore dello s.q.m. è espresso nella stessa unità di misura del carattere osservato.

Devianza

E' il numeratore della varianza. La sua espressione analitica è: $Dev(X) = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Per distribuzioni di frequenza si ha: $Dev(X) = \sum_{i=1}^{n} (x_i - \bar{x})^2 n_i$.

Differenza semplice media senza ripetizione

E' l'indice di variabilità che si ottiene facendo la media aritmetica tra tutte le possibili *n (n* - 1) differenze in valore assoluto tra le modalità differenti x_i ed x_j di un carattere quantitativo, cioè:

$$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{n(n-1)}.$$

Tale indice può essere espresso in termini quadratici.

Coefficiente di variazione

E' un indice relativo di variabilità definito dal rapporto tra lo scarto quadratico medio e la media aritmetica. La sua espressione analitica è la seguente:

$$CV = \frac{\sigma}{\mu}\cdot 100$$

Distribuzione massimante

Data una generica distribuzione di un carattere con modalità ordinate e con un totale di N frequenze, si definisce distribuzione massimante della variabilità una nuova distribuzione con due sole modalità, di cui la prima è la modalità più piccola e la seconda quella più grande,con frequenze assolute rispettivamente pari a p e q , la cui somma corrisponde al totale delle frequenze osservate nella distribuzione di partenza. In tal caso,mantenendo inalterata la media della prima distribuzione,si ha che la distribuzione massimante assume il massimo della variabilità.

Concentrazione

Un carattere quantitativo trasferibile e ordinato in senso crescente, si definisce più o meno concentrato se l'ammontare dello stesso è posseduto da un numero più o meno equo di unità. Quando tutte le unità di un collettivo possiedono lo stesso ammontare del carattere si ha una misura della concentrazione nulla (o equidistribuzione del carattere trasferibile) mentre se una sola unità possiede l'intero ammontare del carattere si parla di concentrazione massima. La misura della concentrazione è data dal rapporto di concentrazione.

Curva di Lorenz

Detta anche curva di concentrazione, viene usata per rappresentare graficamente il grado di concentrazione di un carattere quantitativo trasferibile e ordinato in senso crescente.

Momenti

Valori caratteristici della distribuzione di frequenza di una variabile. Si distingue tra momenti rispetto all'origine e momenti rispetto alla media. Il momento primo rispetto all'origine zero coincide con la media aritmetica ed il momento secondo rispetto alla media coincide con la varianza, mentre il momento secondo rispetto all’origine zero coincide con la media quadratica.

3. variabilità

Cenni Metodologici

Le principali caratteristiche degli indici di variabilità riguardano la dispersione e la disuguaglianza.

Dispersione: misura il maggiore o minore addensamento delle osservazioni rispetto ad un valore medio;
Disuguaglianza: evidenzia le diversità delle varie osservazioni tra loro.

Gli **indici di variabilità** si distinguono in:

- **scostamenti medi** da un valore di sintesi;
- **differenze medie**, che si ottengono calcolando le medie degli scarti in valore assoluto delle modalità del carattere prese a due.

Indici assoluti di variabilità. Sono espresse nelle stessa unità di misura del carattere osservato e il loro campo di variazione è compreso tra 0 e ∞.

Indici relativi di variabilità. Prescindono dall'unità di misura del carattere e si possono ottenere rapportando un indice assoluto di variabilità ad una valor medio o al suo massimo. Pertanto gli stessi indici consentono il confronto della variabilità di caratteri espressi in unità di misura diversa

Dispersione	Disuguaglianza
Indici Variabilità Assoluti Scostamenti semplici medi Varianza e Devianza Scarto quadratico Medio **Indici Variabilità Relativi** Indici assoluti rispetto alla media Indici assoluti rispetto al massimo Rapporto di concentrazione	**Differenze Medie** Differenza semplice media senza ripetizione Differenza semplice media con ripetizione Differenza quadratica media senza ripetizione Differenza quadratica media con ripetizione

Il campo di variazione è dato dalla differenza tra il valore massimo e il valore minimo assunti dalle modalità di un carattere; in simboli:

$W = x_n - x_1$

La differenza interquartilica è data dalla differenza tra il terzo e il primo quartile:

$\delta = Q_3 - Q_1$.

Scostamento semplice medio dalla media aritmetica

E' dato dalla media aritmetica dei valori assoluti dagli scarti dalla media aritmetica, in simboli: $S_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$. Per distribuzioni di frequenza: $S_{\bar{x}} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| n_i}{N}$.

Scostamento semplice medio dalla mediana

E' dato dalla media aritmetica dei valori assoluti degli scarti dalla mediana, in simboli: $S_{Me} = \frac{\sum_{i=1}^{n} |x_i - Me|}{n}$. Per distribuzioni di frequenza si ha : $S_{Me} = \frac{\sum_{i=1}^{n} |x_i - Me| n_i}{N}$.

NB. *L'uso di un indice di variabilità basato sulla mediana può essere particolarmente vantaggioso rispetto ad altri scostamenti medi, quando si sospetta la presenza di valori anomali all'interno della distribuzione.*

La varianza

La varianza è un indice assoluto di variabilità dato dalla media dei quadrati degli scarti dalla media aritmetica, in simboli: $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$. Per frequenze si ha: $\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2 n_i}{N}$.

Un espressione alternativa è la seguente: $\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n} \right)^2$,

alla quale si perviene sviluppando il quadrato del numeratore del rapporto $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$:

$\sigma^2 = \frac{\sum_{i=1}^{n} \left(x_i^2 - 2x_i\bar{x} + \bar{x}^2 \right)}{n}$, ossia: $\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2}{n}$

Considerando che $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, il numeratore della frazione può scriversi nel modo seguente:

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^{\ 2}}{n} - 2\frac{\sum_{i=1}^{n} x_i}{n} \cdot \frac{\sum_{i=1}^{n} x_i}{n} + n \cdot \frac{\left(\frac{\sum_{i=1}^{n} x_i}{n} \right)^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n} \right)^2$$

Pertanto l'espressione analitica della varianza è la seguente: $\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2$,

dove il primo termine a secondo membro è la media quadratica al quadrato (momento secondo), mentre il secondo termine è il quadrato della media aritmetica (momento primo al quadrato).

NB. *Il presente indice rispetto agli scostamenti semplici medi ha amplifica le maggiori distanze dalla media aritmetica che risultano espresse al quadrato, mentre l'inconveniente da un punto di vista interpretativo consiste nel fatto che il valore dell'indice è espresso in termini quadratici dell' unità di misura del carattere osservato.*

Lo scarto quadratico medio o deviazione standard

E' un indice assoluto di variabilità dato dalla radice quadrata della varianza:

s.q.m $= \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$. Per distribuzioni di frequenza si ha: s.q.m.= $\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2 n_i}{N}}$.

NB. *Per ovviare all'inconveniente riscontrato per la varianza, il valore dell'indice s.q.m. è esprimibile nella stessa unità di misura del carattere osservato.*

La devianza

E' un indice assoluto di variabilità, analiticamente è il numeratore della varianza; in simboli: Dev(X)= $\sum_{i=1}^{n} (x_i - \bar{x})^2$. Per distribuzioni di frequenza: Dev(X)= $\sum_{i=1}^{n} (x_i - \bar{x})^2 n_i$.

Un espressione alternativa è la seguente $Dev(X) = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$.

A tal fine si consideri la formula della devianza per una distribuzione di frequenza e si proceda allo sviluppo del quadrato di un binomio:

Dev(X)= $\sum_{i=1}^{n}(x_i-\bar{x})^2 n_i = \sum_{i=1}^{n}\left(x_i^2-2\bar{x}x_i+\bar{x}^2\right)n_i = \sum_{i=1}^{n}x_i^2 n_i - 2\bar{x}\sum_{i=1}^{n}x_i n_i + n\bar{x}\sum_{i=1}^{n}x_i$

Dalla definizione di media aritmetica si ha: $\sum_{i=1}^{n}x_i n_i = \bar{x}\sum_{i=1}^{n}n_i$.

Quindi si ha che:

$$Dev(X) = \sum_{i=1}^{n}(x_i-\bar{x})^2 n_i = \sum_{i=1}^{n}x_i^2 n_i - 2\bar{x}^2\sum_{i=1}^{n}n_i + \bar{x}^2\sum_{i=1}^{n}n_i \ ;$$

$$Dev(X) = \sum_{i=1}^{n}x_i^2 n_i - \bar{x}^2\sum_{i=1}^{n}n_i = \sum_{i=1}^{n}x_i^2 - n\bar{x}^2$$

Differenza semplice media senza ripetizione

E' un indice assoluto di variabilità proposto da C. Gini, e si ottiene come media aritmetica degli scarti in valore assoluto tra i termini della distribuzione, in simboli:

$$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{n(n-1)} \ .$$

Il denominatore sta a significare che non vengono presi in considerazione gli n scarti in valore assoluto di ciascun termine con se stesso.

Differenza semplice media con ripetizione

E' un indice assoluto di variabilità ottenuto con la media aritmetica degli scarti ciascuna modalità e se stessa, effettuando in totale n^2confronti, in simboli:

$$\Delta_R = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{n^2}$$

Differenza quadratica media senza ripetizione

$$^2\Delta = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{n(n-1)}}$$

Differenza quadratica media con ripetizione

$$^2\Delta_R = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2}{n^2}}$$

Si ha che: ${}^{2}\Delta_R{}^{2} = 2\sigma^2$.

NB. *I criteri di scelta degli indici assoluti di variabilità, così come quelli di scelta dei valori medi, possono essere espressi nei seguenti termini. Se la distribuzione è costituita da grandezze oggetto di misurazione (come la distanza da un punto, le stime dell'altezza di una montagna all'orizzonte, il consumo di un genere alimentare tra un collettivo di famiglie), in tal caso con un indice di variabilità assoluto si tende a misurare di quanto in media le misure differiscono dalla vera grandezza. Ovverosia si fornisce una valutazione dell'errore medio che può essere commesso nelle varie misurazioni. In questi casi, l'indice di variabilità assoluto più adeguato sono gli scostamenti.*
Se invece i termini della distribuzione sono grandezze osservate e non oggetto di misurazione, come i voti riportati dagli studenti nel superamento dei loro esami, il reddito di un collettivo di individui,il numero degli esami sostenuti, ecc., in questi casi le differenze medie sono da preferirsi agli scostamenti in quanto misurano come queste grandezze differiscono in media tra loro.

Il coefficiente di variazione

E' un indice relativo di variabilità dato dal rapporto tra lo scarto quadratico medio e la media aritmetica:

$C.V. = \frac{\sigma}{\bar{x}} \cdot 100$, espresso in percentuale.

Non essendo espresso in alcuna unità di misura consente di effettuare confronti tra distribuzioni espresse in diverse unità di misura.

NB. *L'indice può essere utilizzato solo quando tutti i valori della distribuzione sono positivi come ad esempio per reddito, statura, ecc. Il coefficiente di variazione ha un minimo uguale a 0 e un massimo non definito. Può essere utilizzato solo per vedere quale delle due distribuzioni è maggiormente variabile ma non fornisce informazioni sul grado di variabilità.*

Distribuzione massimante della variabilità

Data una distribuzione di un carattere con modalità ordinate e con un totale di N frequenze, la distribuzione massimante della variabilità si ha quando il totale delle frequenze si ripartisce nella frequenza p in corrispondenza del valore minimo x_1, e nella frequenza q = N-p in corrispondenza del valore massimo x_n.
Se imponiamo che la distribuzione massimante abbia la stessa media aritmetica $\bar{x}$ della distribuzione di partenza si ricava che:

$$\bar{x} = \frac{x_1 \cdot p + x_n \cdot (N - p)}{N},$$

da cui risolvendo rispetto a p, si ottiene:

$$p = \frac{N(x_n - \bar{x})}{x_n - x_1}$$

$$q = \frac{N(\bar{x} - x_1)}{x_n - x_1}$$

Variabilità relativa rispetto alla media

E’ possibile costruire differenti indici di variabilità rispetto alla media aritmetica a partire dai più noti indici di variabilità assoluta, quali lo scarto semplice medio, lo scarto quadratico medio e la differenza media del Gini.
Come riportato di seguito:

$$\frac{\delta}{\bar{x}} \qquad \frac{\sigma}{\bar{x}} \qquad \frac{\Delta}{\bar{x}}$$

Il più utilizzato di tale categoria di indici è il rapporto tra lo scarto quadratico medio e la media aritmetica, proposto da Karl Pearson denominato coefficiente di variazione.

La concentrazione

La concentrazione è un aspetto della variabilità ed in particolare della variabilità relativa; essa si riferisce esclusivamente a caratteri quantitativi ordinabili e trasferibili**,** quali ad esempio il reddito.
Dato un carattere X, esso si dice concentrato se l’ammontare complessivo è posseduto da un numero limitato di unità statistiche.

Si parla di:

- **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, ecc.) tali che:

$$x_1 \leq x_2 \leq \cdots \leq x_n$$

Si definiscono $\boldsymbol{p_i}$ le frazioni cumulate delle n unità osservate:

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

Il loro significato è immediato: infatti $p_i,\ i = 1,2,\ldots,n$ rappresenta la frazione cumulata delle prime unità statistiche (posizione nella graduatoria).
Si definiscono $\boldsymbol{q_i}$ le frazioni cumulate del carattere posseduto dalle prime unità statistiche:

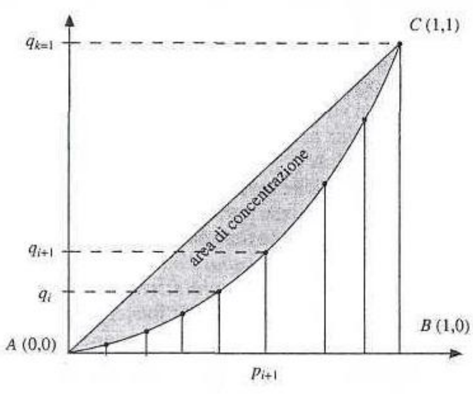
$$q_1 = \frac{A_1}{A_n};\ q_2 = \frac{A_2}{A_n};\ \ldots;\ q_i = \frac{A_i}{A_n};\ \ldots;\ q_n = \frac{A_n}{A_n} = 1,$$

dove:

$A_1 = x_1$
$A_2 = x_1 + x_2$
$\ldots \quad \ldots$
$A_i = x_1 + x_2 + \ldots + x_i$
$\ldots \quad \ldots$
$A_n = x_1 + x_2 + \ldots + x_n$

rappresentano l'ammontare cumulato del carattere X.

Una rappresentazione grafica di tali frazioni è stata proposta dallo statistico statunitense Lorenz nel 1905, attraverso una **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 delle q_i, ottenendo una serie di punti che uniti formano la curva di concentrazione di un dato carattere.
La retta di equidistribuzione è, evidentemente, 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 carattere analizzato è massima, tutto l'ammontare del carattere risulta concentrato in una sola unità; quindi tutti i punti della curva di Lorenz cadono 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.



Indice di concentrazione

Tra gli indici di concentrazione proposti, consideriamo **l'indice R di Gini**, la cui espressione risulta essere la seguente:

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

Tale indice vale:

- 0 in caso di equidistribuzione;
- 1 in caso di massima concentrazione;
- è compreso tra 0 e1 nelle situazioni intermedie.

NB. *L'indice di concentrazione del Gini può essere inteso come un indice di variabilità relativo rispetto al massimo, in quanto l'espressione che compare al denominatore rappresenta il valore massimo che l'indice può assumere.*

I momenti

Alcuni valori medi e alcuni indici di variabilità possono essere espressi attraverso particolari grandezze detti momenti.
Sono definiti come la media aritmetica potenziata di ordine r, semplice o ponderata, degli scarti tra le modalità di una generica variabile X e un particolare valore caratteristico della distribuzione *h*; in simboli si ha: ${}_h m_r = \frac{\sum_{j=1}^{n}(x_i - h)^r}{n}$.

Per una distribuzione di frequenza si ha: ${}_h m_r = \frac{\sum_{j=1}^{n}(x_i - h)^r n_i}{N}$.

Momenti rispetto all'origine ($h = 0$)

$$m_r = \frac{\sum_{j=1}^{n} x^r n_i}{N}$$

- se r = 0 allora $m_0 = 1$
- se r = 1 allora $m_1 = \mu$ (media aritmetica)
- se r = 2 allora $m_2 = Mq^2$ (quadrato della media quadratica).

Momenti centrali o rispetto alla media ($h = \mu$)

$${}_\mu m_r = \frac{\sum_{j=1}^{n}(x_i - \mu)^r n_i}{N}$$

- se r = 0 allora ${}_{\mu}m_0 = 1$
- se r = 1 allora ${}_{\mu}m_1 = 0$ (per la prima proprietà della media aritmetica)
- se r = 2 allora ${}_{\mu}m_2 = \sigma^2$ (varianza).

3. variabilità

Esercizi per l'utilizzo del metodo

Esercizio 1

Si consideri il carattere X =”tempo di connessione ad internet “ rilevato in uno stesso giorno su due differenti laboratori.

Laboratorio A	20	15	60	150	205	0	35	90	80	55
Laboratorio B	73	77	75	65	79	74	72	66	64	65

La media aritmetica di ciascuna delle due distribuzioni è pari a 71 minuti, anche se le due distribuzioni risultano molto differenti tra loro.

La caratteristica che differenzia le due distribuzioni è denominata variabilità che, come definito in precedenza, rappresenta l'attitudine di un carattere ad assumere differenti modalità.

Ricorriamo pertanto al calcolo di uno dei più semplici indici di variabilità assoluta quale lo scarto quadratico medio per evidenziare la differente struttura delle due distribuzioni.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{10}(x_i - \bar{x})^2}{10}}$$

Calcolo scarto quadratico distribuzione “tempo di connessione ad Internet laboratorio A”

X	$(x_i - \mu)$	$(x_i - \mu)^2$
20	-51	2601
15	-56	3136
60	-11	121
150	79	6241
205	134	17956
0	-71	5041
35	-36	1296
90	19	361
80	9	81
55	-16	256
		37090

$$\sigma = \sqrt{\frac{37090}{10}} = 60{,}90$$

Calcolo scarto quadratico distribuzione “tempo di connessione ad Internet laboratorio B”

X	$(x_i - \mu)$	$(x_i - \mu)^2$
73	2	4
77	6	36
75	4	16
65	-6	36
79	8	64
74	3	9
72	1	1
66	-5	25
64	-7	49
65	-6	36
Totale		276

$$\sigma = \sqrt{\frac{276}{10}} = 5{,}23$$

Il calcolo dello scarto quadratico medio evidenzia la maggiore variabilità della distribuzione osservata sul laboratorio A rispetto a quella dello stesso carattere rilevato sul laboratorio B.

Esercizio 2

Sia X= “incasso settimanale (migliaia di euro) osservato nel reparto di profumeria di un centro commerciale

X : 5;   7; 16; 14; 17; 23; 4; 10

L’incasso medio settimanale risulta pari a :

$$\bar{x} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{96}{8} = 12 \qquad \text{migliaia di euro}$$

Lo Scarto quadratico medio come noto è pari a:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{8} (x_i - \bar{x})^2}{8}} \qquad\qquad \sigma = \frac{\sum_{i=1}^{8} (xi - \bar{x})^2}{8}$$

	X	$(x_i - \bar{x})^2$
	5	49
	7	25
	16	16
	14	4
	17	25
	23	121
	4	64
	10	4
Totale	96	308

$$\sigma = \sqrt{\frac{308}{8}} = 6.2025 \qquad \text{migliaia di euro}$$

$$\sigma^2 = \frac{308}{8} = 38.5 \qquad \text{migliaia euro}^2$$

Esercizio 3

Sia X =ore dedicate all'attività sportiva ogni settimana di un collettivo di 32 donne

Ore Attività Sportiva	N° donne
0 - 2	8
2 - 4	12
4 - 6	8
6 - 8	4
	32

Per il calcolo del tempo medio dedicato all'attività sportiva, dello scarto quadratico medio è necessario individuare il valore centrale della classe.

Ore Attività Sportiva	N° donne	Valore Centrale x'	$x_i' n_i$	$(xi - \bar{x})^2 n_i$
0 - 2	8	1	8	50
2 - 4	12	3	36	3
4 - 6	8	5	40	18
6 - 8	4	7	28	49
	32		112	120

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i' n_i}{N} = \frac{112}{32} = 3{,}5 \text{ ore}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 n_i}{N}} = \sqrt{\frac{120}{32}} = 1{,}936 \text{ ore}$$

Esercizio 4

Sia X la variabile "introito pubblicitario" (migliaia di euro) di 5 emittenti radiofoniche.

X= 330; 510; 270; 470 ;420

Si vuole misurare se l'introito è equidistribuito o concentrato tra le emittenti radiofoniche.

i	*X*	*A*	*qi=Ai/An*	*pi=i/n*	*(pi-qi)*
1	*270*	*270*	*0,135*	*0,2*	*0,065*
2	*330*	*600*	*0,3*	*0,4*	*0,1*
3	*420*	*1020*	*0,51*	*0,6*	*0,09*
4	*470*	*1490*	*0,745*	*0,8*	*0,055*
5	*510*	*2000*			
Totali				*2*	*0,31*

Utilizziamo l'indice di concentrazione di Gini:

$$R = \frac{\sum_{i=1}^{n-1}(pi - qi)}{\sum_{i=1}^{n-1} p_i} = \frac{0{,}31}{2} = 0{,}155$$

Come si evince dal valore dell'indice di concentrazione l'introito pubblicitario tra le 5 emittenti radiofoniche risulta prossimo ad una situazione di equidistribuzione.

Esercizio 5

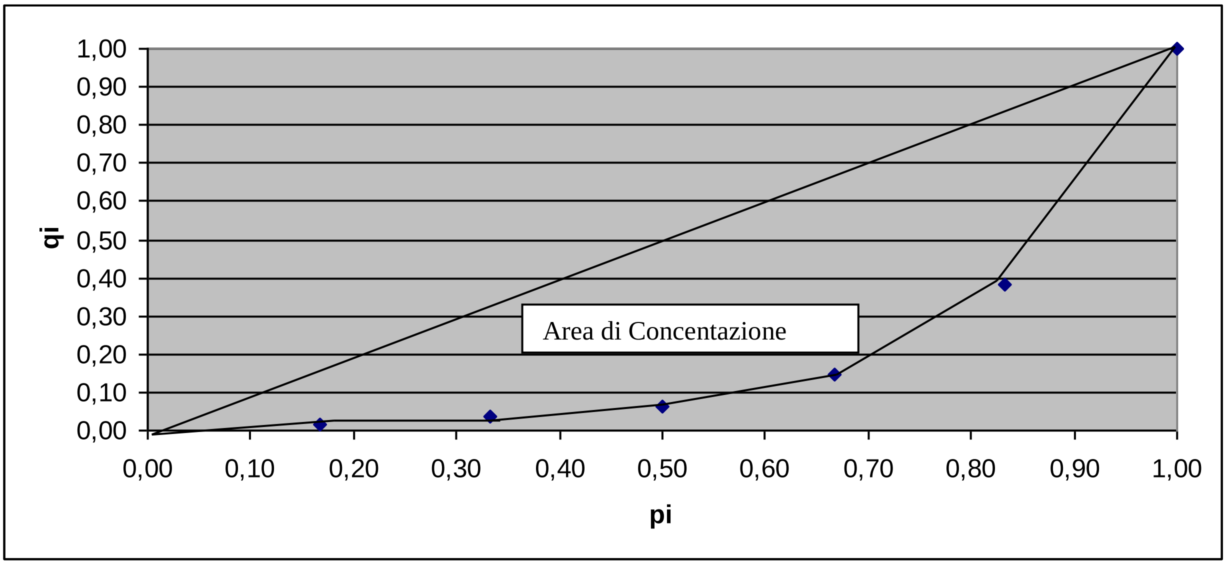
I consumi di energia elettrica nel 2000 per categoria di utilizzatori (in miliardi di Kw) è la seguente

Utilizzatori	Kw
Agricoltura	2,6
Industria	100
Trasporti	5.9
Commercio e servizi	14,5
Illuminazione pubblica	2,5
Usi domestici	38,1

Calcolare il valore del rapporto di concentrazione di Gini e costruire la curva di Lorenz

i	**Utilizzatori**	Kw	pi=i/n	Ai	qi=Ai/An	(pi-qi)
1	Illuminazione pubblica	2,50	0,17	2.50	0,02	0,15
2	Agricoltura	2,60	0,33	5.10	0,04	0,30
3	Trasporti	5,90	0,50	11.00	0,06	0,44
4	Commercio e servizi	14,50	0,67	25.50	0,15	0,52
5	Usi domestici	38,10	0,83	63.60	0,38	0,45
6	Industria	100,00		163.60	1,00	
	Totali		2,50			1,85

$$R = \frac{\sum_{i=1}^{n-1}(pi - qi)}{\sum_{i=1}^{n-1} p_i} = \frac{1{,}85}{2.5} = 0{,}74$$



Esercizio 6

Sia X = “ temperatura media giornaliera” (°C) un carattere quantitativo osservato nell’ultima settimana di gennaio del 2009 in due differenti citta

Tempi	L’Aquila X	Pescara Y
1	3	8
2	2	8
3	2	7
4	-1	6
5	-2	5
6	-4	7
7	0	8

Si vuole individuare quale delle due distribuzioni presenta maggiore variabilità

A tale proposito bisogna osservare che:

a) sebbene tutti i valori siano espressi nella stessa unità di misura, il confronto della variabilità attraverso un indice di variabilità assoluto non è opportuno in quanto le due distribuzioni presentano intensità medie differenti tra loro;
b) non è possibile utilizzare il coefficiente di variazione in quanto una delle due distribuzioni ha un valore medio pari a 0.

Occorre pertanto calcolare una misura di variabilità rispetto al massimo ad esempio

$$V_r = \frac{\sigma}{Max(\sigma)}$$

Procediamo innanzitutto per il calcolo dello scarto quadratico medio per ciascuna distribuzione

Tempi	L’Aquila X	$(x_i - \mu)^2$
1	3	9
2	2	4
3	2	4
4	-1	1
5	-2	4
6	-4	16
7	0	0

$$\sigma_x = \sqrt{\frac{\sum_{1=1}^{n}(x-\bar{x})^2}{n}} = \sqrt{\frac{38}{7}} = 2{,}33$$

Tempi	Pescara Y	$(x_i - \mu)^2$
1	8	1
2	8	1
3	7	0
4	6	1
5	5	4
6	7	0
7	8	1

$$\sigma_y = \sqrt{\frac{\sum_{1=1}^{n}(y-\bar{y})^2}{n}} = \sqrt{\frac{8}{7}} = 1{,}069$$

Procediamo a questo punto alla costruzione della distribuzione massimante per il carattere temperatura giornaliera osservata a L'Aquila

X	ni
-4	3
3	4
	7

$$\sigma_x \max = \sqrt{\frac{(-4-0)^2 3 + (3-0)^2 4}{7}} = \sqrt{\frac{84}{7}} = 3{,}464$$

$V_{x_r} = \frac{2{,}33}{3{,}464} = 0{,}673$ rappresenta l'indice di variabilità relativo ottenuto come rapporto tra lo scarto quadratico medio della distribuzione e quello massimo calcolato sulla distribuzione massi mante della variabilità.

Analogamente si procede per il calcolo della seconda distribuzione massimante:

Y	ni
5	2,33
8	4,67
	7

$$\sigma_y \max = \sqrt{\frac{(5-7)^2 2{,}33 + (8-7)^2 4{,}67}{7}} = \sqrt{\frac{13{,}99}{7}} = 1{,}41$$

$$V_{y_r} = \frac{1{,}069}{1{,}414} = 0{,}756$$

L’indice di variabilità relativo assume un valore più elevato per la distribuzione della temperatura osservata a Pescara per cui è possibile concludere che tale distribuzione presenta maggiore variabilità.

3. variabilità

Esercizi di autovalutazione

RACCOMANDAZIONE: ***Gli esercizi di autovalutazione sono utili per la verifica del livello di conoscenza degli argomenti solo dopo aver assimilato i concetti esposti nella parte metodologica.***

Una misura della variabilità deve:

- ❑ A) Crescere all'aumentare della disuguaglianza tra le modalità della distribuzione.
- ❑ B) Diminuire all'aumentare della disuguaglianza tra le modalità della distribuzione.
- ❑ C) Non variare all'aumentare della disuguaglianza tra le modalità della distribuzione.
- ❑ D) Poter assumere valori negativi in assenza di variabilità.

1. Risposta esatta: A

La variabilità misura l'attitudine del fenomeno ad assumere diverse modalità. Da un punto di vista operativo, occorre definire una misura della variabilità e questo può avvenire mediante la costruzione di indici statistici che devono possedere due proprietà essenziali:

- essere nulli, in assenza di variabilità, quando i termini della distribuzione sono uguali tra loro
- assumere valori infinitamente grandi in relazione al grado di variabilità dei termini della distribuzione

2. La varianza è nulla se:

- ❑ A) La media aritmetica è nulla.
- ❑ B) Le modalità sono uguali tra loro e uguali alla media aritmetica .
- ❑ C) La media aritmetica coincide con lo scarto quadratico medio.
- ❑ D) La media aritmetica è massima .

2. Risposta esatta: B

La varianza è sempre non negativa in quanto è una media di scarti dalla media al quadrato. Essa assume valore minimo 0 quando tutte le modalità sono uguali tra loro e uguali alla media aritmetica per cui tutti gli scarti sono nulli. La varianza aumenta all'aumentare della differenza dei valori osservati. La varianza non ammette un massimo assoluto perché gli scarti dalla media aritmetica possono essere infinitamente grandi.

3. Qual è l'unità di misura della varianza?

- ❑ A) Non ha unità di misura.
- ❑ B) Ha la stessa unità di misura del carattere.
- ❑ C) Ha la stessa unità di misura del carattere rapportata al numero delle osservazioni.
- ❑ D) Ha l'unita di misura del carattere elevata al quadrato.

3. Risposta esatta: D

La varianza $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$ e lo scarto quadratico medio $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$ sono tra gli indici più utilizzati in Statistica per sintetizzare la variabilità di una distribuzione. Va sottolineata una certa difficoltà nella interpretazione della varianza che deriva dal fatto che essa è espressa nella stessa unità di misura della variabile osservata elevata al quadrato. Ad esempio, una varianza di una distribuzione relativa alla variabile statura misurata in cm risulterà espressa in cm^2.
Da qui un maggiore utilizzo dello scarto quadratico medio come misura della variabilità assoluta di una distribuzione.

4. La varianza di un carattere può essere ottenuta come differenza tra:
- ❑ A) Il quadrato della media quadratica e il quadrato della media aritmetica.
- ❑ B) La media quadratica e il quadrato della media aritmetica.
- ❑ C) Il quadrato della media quadratica e la media aritmetica.
- ❑ D) La media quadratica e la media aritmetica.

4. Risposta esatta: A
Dato un insieme di n osservazioni (o modalità di un carattere X) $x_1, x_2, ..., x_n$ aventi media aritmetica μ, la formula della varianza è la seguente:

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

Per la distribuzione di frequenza di un carattere X, con n modalità $x_1, x_2, ..., x_n$ e rispettive frequenze $n_1, n_2, ..., n_n$ tali che $\sum_{i=1}^{n} n_i = n$, la sua formula è, invece, la seguente:

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

Un'espressione alternativa più semplice della varianza che ne facilita il calcolo è la seguente:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i)^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2$$

Ovvero, la varianza è data dal quadrato della media quadratica meno il quadrato della media aritmetica:

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

5. Si può misurare la variabilità di un carattere qualitativo con la varianza?
- ❑ A) Si, sempre.
- ❑ B) Dipende dall'unità di misura del carattere.
- ❑ C) No, la varianza è adatta solo per caratteri quantitativi.
- ❑ D) Si, se il carattere è qualitativo ordinabile.

5. Risposta esatta: C
La varianza è uno degli indici di variabilità più utilizzati. Tale indice rientra nella categoria degli indici di dispersione che misurano la variabilità del carattere tramite una sintesi di misure di diversità tra ogni termine della distribuzione ed una media (scostamenti medi). La varianza, la cui formula è la seguente:

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

è la media dei quadrati degli scarti dalla media aritmetica. Per questa misura della dispersione la diversità tra una modalità e la media aritmetica viene misurata elevando al quadrato la differenza tra due valori. Data la sua espressione formale, che richiede operazioni algebriche tra le modalità, è facile intuire che si tratta di un indice adatto solo per caratteri quantitativi. Quando la diversità concerne gli attributi di una mutabile si parla di eterogeneità.

6. Conoscendo la varianza di una distribuzione è possibile ricavarne la devianza?
- ❑ A) Si, sempre.
- ❑ B) Si, se si conosce il numero delle osservazioni.
- ❑ C) No.
- ❑ Si, se si conosce lo scarto quadratico medio.

6. Risposta esatta: B
La devianza è un indice assoluto della variabilità che rappresenta di fatto il numeratore della varianza. Indicando con X il carattere considerato, con *n il* numero di unità del collettivo osservato, con x_i le singole osservazioni del carattere e con μ la loro media aritmetica, la devianza è data dalla seguente espressione:

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

Per una distribuzione di frequenza l'espressione precedente si pondera con le frequenze

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

La varianza si ottiene rapportando la devianza al numero delle osservazioni

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

Ne consegue che data la varianza di una distribuzione sarà possibile ricavarne la devianza solo se si conosce il numero delle osservazioni, come risulta dalla relazione sotto indicata:

$$Dev(X) = n * \sigma^2$$

7. Var c(X) è pari

- ❑ A) $c^2 Var(X)$
- ❑ B) $Var(X)$
- ❑ C) $c\,Var(X)$
- ❑ D) $\frac{Var(X)}{Var(c)}$

7. Risposta esatta: A
Una nota proprietà della varianza è la seguente:

$$Var(cX) = c^2 Var(X)$$

Dimostrazione
Si indichi con μ la media del carattere X. È noto che la media di (cX)=c μ. Ricordando che la varianza è la media degli scarti dalla media al quadrato si potrà scrivere: $Var(cX) = \frac{1}{n}\sum_{i=1}^{n}(cx_i - c\mu)^2 = c^2 \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = c^2 Var(X)$

8. Lo scostamento semplice medio dalla media aritmetica
- ❑ A) È un indice di dispersione.
- ❑ B) È un indice di disuguaglianza.
- ❑ C) È un intervallo di variazione.
- ❑ D) Non rientra in nessuna delle categorie di indici di variabilità precedentemente indicate.

8. Risposta esatta: A
Tra le misure di variabilità, e più precisamente tra gli indici di dispersione, troviamo gli scostamenti semplici medi che si ottengono come media aritmetica delle differenze, in valore assoluto, tra i valori osservati ed una media. A seconda della media scelta si può ottenere uno specifico scostamento semplice medio. Ad esempio, se come media scegliamo la media aritmetica, si ha lo scostamento semplice medio dalla media aritmetica. Dato un insieme di n modalità di un carattere X ($x_{1,} x_2,...,x_s$), avente media aritmetica μ, lo scostamento semplice medio dalla media aritmetica ha la seguente espressione:

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

Nel caso di una distribuzione di frequenze lo scostamento semplice medio dalla media aritmetica sarà dato da:

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

9. Se si vuole pervenire ad una sintesi di misure di diversità di tutti i termini della distribuzione tra loro è opportuno calcolare
- ❑ A) La differenza semplice media con ripetizione.
- ❑ B) Il coefficiente di variazione.
- ❑ C) L'indice di concentrazione.
- ❑ D) La differenza interquartile.

9. Risposta esatta: A
Le differenze medie sono indici di variabilità in cui non si considera la dispersione dei termini rispetto ad una media, ma la disuguaglianza dei termini fra loro. La differenza semplice media con ripetizione si ottiene confrontando tutti i termini a due a due (non escludendo neppure il confronto di un termine con se stesso). La differenza semplice media con ripetizione è definita dalla seguente formula:

$$\Delta_R = \frac{\sum_i \sum_j |x_i - x_j|}{n^2}$$

Al denominatore di questo indice figurano anche le n differenze nulle generate dal confronto di ciascuna modalità con se stessa. Se, invece, si escludono le n differenze nulle di ogni valore con se stesso, si può calcolare la differenza semplice media senza ripetizione la cui espressione è:

$$\Delta = \frac{\sum_i \sum_j |x_i - x_j|}{n\,(n-1)}$$

10. Siano Δ e Δ_R, rispettivamente, la differenza semplice media e la differenza semplice media con ripetizione delle n modalità di un carattere statistico X, qual è la relazione tra i due indici di variabilità?

- ❑ A) $\Delta_R = \Delta\, n$
- ❑ B) $\Delta_R = \Delta \dfrac{n-1}{n}$
- ❑ C) $\Delta_R = \Delta \dfrac{1}{n}$
- ❑ D) $\Delta_R = \Delta \dfrac{n}{n-1}$

10. Risposta esatta: B

La differenza semplice media ha la seguente espressione:

$$\Delta = \frac{\sum_i \sum_j |x_i - x_j|}{n\,(n-1)}$$

Se, invece, si vogliono conteggiare anche le n differenze nulle di ogni valore con se stesso, allora le differenze da valutare sono n^2. Rapportando la somma delle differenze a tale numero si calcola la differenza semplice media con ripetizione, che per un insieme di dati è espressa dalla seguente formula:

$$\Delta_R = \frac{\sum_i \sum_j |x_i - x_j|}{n^2}$$

Tra la differenza semplice con ripetizione e quella senza ripetizione esiste un'ovvia relazione:

$$\Delta_R = \Delta \frac{n-1}{n}$$

È facile intuire che quando il collettivo è di numerosità sufficientemente elevata, il rapporto $\frac{n-1}{n}$ tende ad 1 sicché la differenza tra l'indice con ripetizione e quello senza ripetizione tende ad annullarsi.

11. Sia dato un carattere X, siano Q_1, Q_2,,e Q_3, rispettivamente, il primo, il secondo e il terzo quartile della distribuzione, qual è l'esatta espressione della differenza interquartile?

- ❑ A) $\dfrac{Q_2 - Q_1}{2}$
- ❑ B) $Q_3 - Q_1$
- ❑ C) $Q_3 - Q_2$
- ❑ D) $\dfrac{Q_3 - Q_1}{2}$

11. Risposta esatta: B
La differenza interquartile è un indice di variabilità assoluta, ottenuto dalla differenza tra il terzo e il primo quartile, cioè tra quella modalità che rappresenta il limite superiore dei primi ¾ della distribuzione e quella che rappresenta il limite superiore del primo quarto della stessa, mettendo in evidenza l'addensamento dei valori osservati nella parte centrale della distribuzione. Nell'intervallo così determinato cade la metà dei casi osservati. Questo indice di variabilità è ricompreso nella categoria degli intervalli di variazione, che si basano sulla differenza tra due particolari valori della distribuzione. Questo indice di variabilità, rispetto al campo di variazione, è meno sensibile all'influenza dei valori cosiddetti anomali. Una variante della differenza interquartile è data dalla semidifferenza interquartile ottenuta dividendo la differenza interquartile per due (di cui alla risposta errata D).

12. Il campo di variazione è uguale:
- ❑ A) Alla differenza tra il terzo ed il primo quartile.
- ❑ B) Al rapporto tra lo scarto quadratico medio e la media aritmetica.
- ❑ C) Alla differenza tra il valore massimo ed il valore minimo assunto dal carattere.
- ❑ D) Al rapporto tra la varianza e la media aritmetica.

12. Risposta esatta: C
Dato un insieme di n valori osservati $x_1, x_2, ..., x_n$, ordinati in senso crescente, il campo di variazione è ottenuto come differenza tra il più grande e il più piccolo di tali valori:

$$W = x_n - x_1$$

Il minimo del campo di variazione è 0 e questo si verifica solo se tutte le unità presentano lo stesso valore. È evidente che tale indice, basandosi solo su 2 degli n valori osservati è una misura della variabilità piuttosto approssimativa. Gli inconvenienti nell'uso di tale indice sono dovuti innanzitutto al fatto che spesso le distribuzioni hanno come modalità estreme delle classi aperte, di conseguenza la scelta del valore rappresentante la modalità estrema sarà inficiata da un certo grado di arbitrarietà; inoltre, poiché esso tiene conto solo dei valori estremi, altro inconveniente è la sua sensibilità alla presenza nella distribuzione di fattori eccezionali (dati anomali), cioè di dati che si discostano in maniera notevole dal resto delle osservazioni. L'unico pregio di questo indice risiede nella semplicità di calcolo.

13. Gli indici di variabilità relativa si ottengono:
- ❑ A) Attraverso la radice quadrata degli indici di variabilità assoluta
- ❑ B) Moltiplicando gli indici di variabilità assoluta per una costante
- ❑ C) Rapportando l'indice assoluto ad una media oppure al suo massimo
- ❑ D) Rapportando l'indice assoluto al numero delle osservazioni

13. Risposta esatta: C
Gli indici variabilità relativa si ottengono rapportando un indice di variabilità assoluta ad una media oppure al massimo che l'indice di variabilità assoluto può assumere (distribuzione massimante). In tal modo l'unità di misura scompare e si ottiene un valore numerico, indipendente da essa, idoneo a consentire confronti tra distribuzioni di variabili espresse con unità di misura diverse. Infatti, è opportuno utilizzare gli indici di variabilità relativa quando si ha la necessità di confrontare le variabilità di due distribuzioni secondo caratteri misurati in unità di misura diverse non trasformabili in una stessa unità di misura oppure confrontare le variabilità di due distribuzioni secondo caratteri misurati nella stessa unità di misura ma con intensità medie differenti.

14. La distribuzione massimante viene utilizzata
- ❑ A) Per calcolare il coefficiente di variazione.
- ❑ B) Per calcolare la differenza interquartile.
- ❑ C) Per calcolare gli indici di variabilità relativi al massimo.
- ❑ D) Per calcolare la relazione tra lo scostamento semplice medio e lo scostamento quadratico medio.

14. Risposta esatta: C
Gli indici di variabilità relativi al massimo variano tra 0 e 1. Tale categoria di indici di variabilità relativa si ottengono rapportando l'indice di variabilità assoluto al valore massimo che lo stesso può assumere. Essi sono strettamente dipendenti dalle ipotesi in base alle quali si determina il massimo dell'indice, cioè dalle ipotesi che sono state poste alla base della costruzione della distribuzione massimante. Uno dei metodi più utilizzati per la costruzione della distribuzione massimante parte dal presupposto che il massimo grado di variabilità si ha quando il carattere assume due sole modalità estreme (x_1 e x_n) con rispettive frequenze pari a P e a (N-P), tali che l'intensità media del carattere, rappresentata dalla media aritmetica, resti immutata. L'imposizione di invarianza della media aritmetica, da un lato consente di determinare univocamente le due frequenze assolute P e (N-P) tramite le quali risalire al massimo dell'indice di variabilità assoluto prescelto e dall'altro consente di concentrare tutta l'attenzione sulla variabilità della distribuzione, visto che l'intensità media viene mantenuta costante.

15. Il coefficiente di variazione è definito
- ❑ A) Per caratteri qualitativi.
- ❑ B) Per caratteri quantitativi tali che $\mu \leq 0$.
- ❑ C) Per tutti i tipi di caratteri.
- ❑ D) Per caratteri quantitativi tali che $\mu > 0$.

15. Risposta esatta: D
II coefficiente di variazione è un indice relativo di variabilità, ottenuto attraverso il rapporto tra lo scarto quadratico medio e la media, moltiplicato per 100. Esso consente il confronto tra distribuzioni anche di tipo molto diverso non dipendendo dall'unità di misura e dall'ordine di grandezza della variabile. Il coefficiente di variazione è normalmente utilizzato solo quando tutti i valori della distribuzione sono positivi, come, per esempio, nel caso delle variabili statura, età, peso, salario. Infatti per caratteri che assumono valori negativi e positivi, la media aritmetica non rappresenta l'ordine di grandezza effettivo (si pensi a quei caratteri che hanno media nulla anche se, in realtà, presentano valori molto grandi e molto piccoli ma che fra loro si compensano). Quindi il coefficiente di variazione possiede alcune limitazioni, in particolare non è definito per caratteri quantitativi tali che $\mu \leq 0$, perché non è possibile interpretare una variabilità negativa, né dividere un numero per zero.

16. Il rapporto di concentrazione di Gini
- ❑ A) È un indice di disuguaglianza.
- ❑ B) È un indice di variabilità assoluta.
- ❑ C) È un indice di variabilità relativo alla media.
- ❑ D) È un indice di variabilità relativo al massimo.

16. Risposta esatta: D
La necessità di misurare la concentrazione di un carattere X deriva dalla trasferibilità dell'ammontare del carattere da una unità statistica ad un'altra.

Uno degli indici più utilizzati per la misura della concentrazione è quello introdotto dal Gini e definito attraverso la seguente espressione:

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

Si tratta di un indice adimensionale che assume valori nell'intervallo [0,1], i cui estremi corrispondono rispettivamente al caso di equidistribuzione ed a quello di massima concentrazione.
La concentrazione varia tra due valori estremi: concentrazione nulla (o equidistribuzione) quando l'ammontare complessivo è distribuito tra le n unità statistiche in modo uguale e concentrazione massima quando l'intero ammontare complessivo del carattere è posseduto da una sola unità statistica mentre le altre ne sono sprovviste. Nei casi intermedi, vi saranno unità statistiche che possiedono il carattere in misura superiore e altre in misura inferiore alla media..
Il rapporto di concentrazione introdotto da Gini è un indice di variabilità relativo al massimo. È facile verificare, infatti, che il denominatore è il massimo del numeratore poiché corrisponde al caso di concentrazione massima in cui $q_i = 0$ per i da 1 a $n-1$ e $q_n = p_n = 1$.

17. In caso di equidistribuzione di un carattere la curva di concentrazione coincide con

- ❑ A) La bisettrice del primo quadrante di un sistema ortogonale cartesiano.
- ❑ B) Una parabola.
- ❑ C) Una retta avente pendenza negativa.
- ❑ D) Un triangolo.

17. Risposta esatta: A

Siano $x_1, x_2, ..., x_n$ le modalità disposte in senso non decrescente di un carattere trasferibile X= reddito, con $x_1 \le x_2 \le ... x_n$. Una rappresentazione grafica della concentrazione del carattere è ottenuta calcolando:

- le frazioni cumulate dei primi *i* redditieri:

$$p_i = \frac{i}{n} \text{ con } i = 1,2,\ldots,n$$

- le frazioni cumulate del reddito posseduto dai primi *i* redditieri:

$$q_i = \frac{\sum_{j=1}^{i} x_j}{\sum_{j=1}^{n} x_j}$$

La rappresentazione grafica delle funzioni suddette è fornita dalla spezzata di concentrazione o curva di Lorenz. Essa si ottiene ponendo, in un sistema di assi cartesiani, sull'asse delle ascisse le frazioni cumulate dei redditieri p_i e sull'asse delle ordinate le frazioni cumulate del reddito q_i.

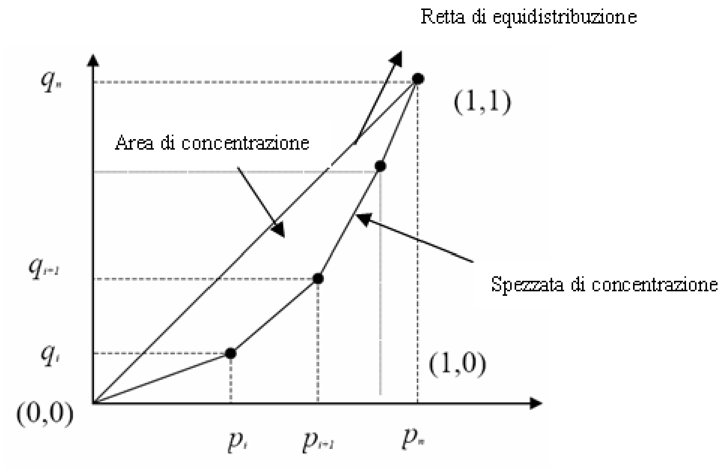


Grafico della curva di Lorenz

Nel caso di equidistribuzione si ha che $p_i=q_i$ e quindi i punti si dispongono sulla bisettrice del I quadrante. Il segmento che unisce i punti di coordinate (0,0) e (1,1) viene chiamato segmento di equidistribuzione. Se non vi è equidistribuzione i punti di coordinate (p_i,q_i) si trovano nel triangolo di vertici (0,0), (1,0) e (1,1). Unendo tali punti si ottiene una linea chiamata spezzata di concentrazione o curva di Lorenz. In generale quanto è maggiore la concentrazione del carattere tanto più la spezzata di concentrazione risulta vicina all'asse dell'ascisse e quindi tanto è più grande
l'area della superficie compresa fra il segmento di equidistribuzione e la spezzata di concentrazione.

18. In caso di equidistribuzione di un carattere quale valore assume il rapporto di concentrazione di Gini:
❑ A) 1
❑ B) 0
❑ C) Non ammette minimo assoluto
❑ D) Non ammette massimo assoluto

18. Risposta esatta: B
Uno degli indici più utilizzati per misurare la concentrazione è il rapporto di concentrazione di Gini la cui espressione è la seguente:

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

Tale indice vale 0 in caso di equidistribuzione; 1 in caso di massima concentrazione; è compreso tra 0 e 1 nelle situazioni intermedie.
Il minimo dell'indice (zero) si ha in caso di equidistribuzione. Infatti se il carattere è equidistribuito, l'ammontare complessivo dello stesso è ripartito in parti uguali tra le n unità statistiche, per cui le modalità ad esse afferenti assumono tutte valore uguale alla media aritmetica, quindi $x_1 = x_2 = \ldots, x_n = \bar{x}$. In questo caso risulta $p_i = q_i$ per ogni $i = 1,2,\ldots,n$. e quindi

$\sum_{i=1}^{n-1}(p_i - q_i)$=0. Si avrà concentrazione massima se l'ammontare complessivo del carattere è attribuito ad una sola unità che ne possiede in misura pari a $n\bar{x}$ mentre le restanti n-1 ne posseggono 0, cioè se $x_1 = x_2 = \ldots, x_{n-1} = 0$ e $x_n = n\bar{x}$. In questo caso risulta

$p_i = \frac{i}{n}$ per ogni $i = 1,2,\ldots,n$

$q_1 = q_2 = \ldots, = q_{n-1} = 0$ e $q_n = 1$

Nei casi intermedi il carattere è tanto più concentrato quando maggiore è la differenza tra $(p_i - q_i) \geq 0$.

VERIFICA

Indicare se le seguenti affermazioni sono vere o false.

		V	F
1	La differenza interquartile è sempre <0.25.		
2	La deviazione standard è sempre minore dello scostamento semplice medio dalla mediana.		
3	La differenza interquartile può essere calcolata solo per dati positivi.		
4	Il coefficiente di variazione dipende dall'unità di misura.		
5	Il campo di variazione è sempre maggiore o uguale alla differenza interquartile.		
6	Un indice di variabilità deve assumere il suo valore minimo se e solo se tutte le modalità osservate sono uguali a zero.		
7	Nel calcolo della varianza, gli scarti più elevati (tra i valori osservati e la media) acquistano molta più importanza rispetto agli scarti più piccoli.		
8	La devianza è il numeratore della deviazione standard.		
9	Il coefficiente di variazione non è adatto ad analizzare una distribuzione che presenti anche valori negativi.		
10	Gli scostamenti semplici medi si ottengono come media aritmetica dei quadrati delle differenze tra i valori osservati e una media.		
11	Quando un carattere è equidistribuito, la variabilità è massima.		
12	La situazione di massima concentrazione si ha quando l'intero ammontare del carattere è posseduto da una sola unità.		
13	L'area di concentrazione evidenzia la differenza tra il grado di concentrazione osservato e la situazione di equidistribuzione.		