

Capitolo 6 | Gli indici di variabilità

Sommario | 1. Introduzione. - 2. Il campo di variazione. - 3. La differenza interquartile. - 4. Gli scostamenti medi. - 5. La varianza, lo scarto quadratico medio e la devianza. - 6. Le differenze medie. - 7. Il coefficiente di variazione. - 8. La concentrazione. - 9. I momenti. - 10. Cenni sugli indici di mutabilità. - **Questionario**.

1. Introduzione

La sintesi di una distribuzione operata attraverso gli indici di posizione fa perdere informazioni rilevanti sulla stessa; gli **indici di variabilità** tendono a colmare questa perdita.

La **variabilità** è l'attitudine di un fenomeno quantitativo ad assumere diverse modalità. Nella metodologia statistica si distinguono due aspetti della variabilità:

- ✓ la **dispersione** che caratterizza il maggiore o minore addensamento delle osservazioni intorno ad una media prestabilita;
- ✓ la **disuguaglianza** (o variabilità reciproca) che evidenzia la diversità delle varie osservazioni tra loro.

Gli indici di variabilità misurano, quindi, la variabilità di una distribuzione di frequenza:

- ✓ rispetto ad un centro rappresentativo (**dispersione**) e sono detti **scostamenti medi** e si ottengono determinando gli scarti $x_i - x_M$ tra le modalità del carattere e una sua media x_M e sintetizzandoli con la stessa o con altra media;
- ✓ tra le unità statistiche a due a due (**disuguaglianza**) e sono detti **differenze medie** e si ottengono determinando le differenze in valore assoluto $|x_i - x_j|$ delle modalità del carattere prese a due a due, e sintetizzandole con opportuna media.

Si distinguono in:

- ✓ **indici assoluti di variabilità**, che sono espressi in una unità di misura che dipende dall'unità di misura del fenomeno in esame, come la *varianza*, lo *scarto quadratico medio* etc.;
- ✓ **indici relativi di variabilità**, che prescindono dall'unità di misura, per cui sono particolarmente adatti per effettuare confronti tra fenomeni diversi. Essi si possono ottenere rapportando un indice assoluto ad una media o al suo massimo.

Infine, si considerano gli **indici normalizzati** come il *rapporto di concentrazione*.

2. Il campo di variazione

Il **campo di variazione** (o **range**) è l'indice assoluto di variabilità più semplice da calcolarsi, è dato dalla differenza tra il valore massimo e il valore minimo delle modalità di un carattere; in simboli:

$$\omega = x_{\max} - x_{\min} \quad (2.1)$$

L'indice in questione è poco utilizzato in quanto prende in considerazione solo la dispersione esistente tra i valori estremi della distribuzione, non tenendo conto della sua struttura. Esso risente di eventuali valori anomali o di errori di registrazione presenti nei dati.

ESEMPIO

Determinare il campo di variazione del seguente insieme di numeri:

35, 4, 28, 76, 12, 5, 7

Per ottenere il campo di variazione bisogna comporre in serie l'insieme; ossia:

4, 5, 7, 12, 28, 35, 76

quindi, fare la differenza tra il numero più grande e il numero più piccolo:

$$\omega = 76 - 4 = 72$$

3. La differenza interquartile

Un indice di variabilità meno sensibile del campo di variazione a valori estremi eccezionali è la **differenza interquartile** data dalla differenza tra il terzo e il primo quartile; in simboli:

$$\delta_Q = Q_3 - Q_1 \quad (3.1)$$

Si tratta di una misura di variabilità della parte centrale di una distribuzione, in quanto, nell'intervallo espresso dall'indice, è compreso il 50% delle osservazioni.

ESEMPIO

Determinare la differenza interquartile della distribuzione riportata nella tabella 6 del quinto capitolo.

Dall'esempio 2 del paragrafo 10 del quinto capitolo si evince che il primo e il terzo quartile sono, rispettivamente:

$$Q_1 = 180,292 \quad \text{e} \quad Q_3 = 190,271$$

per cui la differenza interquartile è:

$$\delta_o = Q_3 - Q_1 = 190,271 - 180,292 = 10,429$$

4. Gli scostamenti medi

Il campo di variazione e la differenza interquartile presentano il comune inconveniente di non tenere conto di tutte le informazioni su una variabile statistica. Gli indici di variabilità di cui ci occuperemo in questo paragrafo, anche se poco usati nella pratica, coinvolgono nel loro calcolo, tutte le determinazioni della variabile considerata e sono: lo *scostamento medio dalla media aritmetica* e lo *scostamento medio dalla mediana*. Si preferisce l'uno all'altro a seconda che sia stata scelta la media aritmetica o la mediana per misurare la dispersione dei dati.

Gli indici in questione sono calcolati considerando i valori assoluti degli scarti, in quanto, per la media aritmetica, la media degli scarti, presi con il loro segno, è nulla.

4.1 Scostamento medio dalla media aritmetica

Lo **scostamento medio dalla media aritmetica** è un indice di variabilità dato dalla media aritmetica dei valori assoluti degli scarti dalla media aritmetica:

$$S_\mu = \frac{\sum_{i=1}^n |x_i - \mu|}{n} \quad (4.1)$$

Se si dispone della distribuzione di frequenza delle modalità è:

$$S_\mu = \frac{\sum_{i=1}^k |x_i - \mu| n_i}{\sum_{i=1}^k n_i} \quad (4.2)$$

ESEMPIO 1

Determinare lo scostamento medio dalla media aritmetica del seguente insieme di numeri.

3, 5, 15, 23, 28

Per determinare l'indice di variabilità richiesto occorre calcolare la media aritmetica:

$$\mu = \frac{3 + 5 + 15 + 23 + 28}{5} = 14,8$$

Pertanto, lo scostamento medio dalla media aritmetica è:

$$S_{\mu} = \frac{|3-14,8|+|5-14,8|+|15-14,8|+|23-14,8|+|28-14,8|}{5} = \frac{11,8+9,8+0,2+8,2+13,2}{5} = \frac{43,2}{5} = 8,64$$

4.2 Scostamento medio dalla mediana

Lo **scostamento medio dalla mediana** è un indice di variabilità dato dalla media aritmetica dei valori assoluti degli scarti dalla mediana:

$$S_{Me} = \frac{\sum_{i=1}^n |x_i - Me|}{n} \quad (4.3)$$

Se si dispone della distribuzione di frequenza delle modalità è:

$$S_{Me} = \frac{\sum_{i=1}^k |x_i - Me| n_i}{\sum_{i=1}^k n_i} \quad (4.4)$$

ESEMPIO 2

Determinare lo scostamento medio dalla mediana del seguente insieme di numeri:

3, 5, 15, 23, 28

Per calcolare l'indice richiesto è necessario dapprima calcolare la mediana. Siccome i dati sono disposti in ordine crescente e sono in numero dispari (5), la mediana è quel numero che occupa il 3° posto, ossia:

$$Me = 15$$

Lo scostamento medio della mediana è dunque:

$$S_{Me} = \frac{|3-15|+|5-15|+|15-15|+|23-15|+|28-15|}{5} = \frac{12+10+0+8+13}{5} = 8,6$$

Dalla proprietà della mediana, per cui la somma dei valori assoluti degli scarti dalla mediana è un minimo, si evince che per il medesimo insieme di numeri si ha:

$$S_{Me} < S_{\mu}$$

5. La varianza, lo scarto quadratico medio e la devianza

5.1 Varianza

La **varianza** di un insieme di dati o di una distribuzione di frequenza è una misura di dispersione che si ottiene come **media dei quadrati degli scarti dalla media aritmetica**; in simboli:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} \quad (5.1)$$

Se si dispone, invece, della distribuzione di frequenza:

$$\sigma^2 = \frac{\sum_{i=1}^k (x_i - \mu)^2 n_i}{\sum_{i=1}^k n_i} \quad (5.2)$$

Per evitare calcoli laboriosi della varianza è possibile usare la formula semplificata:



$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n^2} \quad (5.3)$$

Il primo termine a secondo membro è il quadrato della media quadratica, mentre il secondo termine è il quadrato della media aritmetica, per cui:

$$\sigma^2 = Q^2 - \mu^2$$

La varianza presenta, tuttavia, un notevole inconveniente: è espressa nel quadrato dell'unità di misura delle osservazioni, ad esempio, se le osservazioni sono espresse in metri, la varianza è espressa in metri al quadrato. Pertanto, non è mai possibile rappresentare su uno stesso diagramma la varianza e la distribuzione delle osservazioni.

5.2 Scarto quadratico medio

Per ovviare all'inconveniente visto, in Statistica si preferisce usare la radice quadrata della varianza, espressa nella stessa unità di misura delle determinazioni di una variabile e denominata **scarto quadratico medio** o **deviazione standard**.

Per un insieme di dati la sua espressione analitica è la seguente:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \quad (5.4)$$

Se si dispone, invece, della distribuzione di frequenza, lo scarto quadratico medio è:

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

Come vedremo meglio in seguito, nelle distribuzioni di frequenza unimodali e non asimmetriche, circa i due terzi delle osservazioni sono compresi tra: $\mu - \sigma$ e $\mu + \sigma$. Inoltre, quasi tutte le osservazioni sono comprese nell'intervallo: $\mu \pm 3\sigma$.

VALORE MASSIMO DELLO SCARTO QUADRATICO MEDIO

Quando **una sola modalità racchiude l'intero carattere** investigato, mentre le rimanenti $n - 1$ non ne posseggono affatto, lo scarto quadratico medio assume valore massimo:

$$\sigma_{\max} = \mu\sqrt{n-1} \tag{5.5}$$

5.3 Devianza

Prima di illustrare esempi di applicazione delle formule appena date, consideriamo il **numeratore della varianza**: trattasi di un'altra misura di dispersione denominata **devianza**. Per un carattere X la sua espressione analitica è la seguente:

$$D(X) = \sum_{i=1}^n (x_i - \mu)^2 \tag{5.6}$$

In pratica, la devianza è pari alla somma dei quadrati delle deviazioni dei dati dalla relativa media.

Se si dispone della distribuzione di frequenza, la devianza è:

$$D(X) = \sum_{i=1}^k (x_i - \mu)^2 n_i \tag{5.7}$$

La formula della devianza può essere semplificata, onde evitare il calcolo laborioso degli scarti, e la sua espressione è:

$$D(X) = \sum_{i=1}^k x_i^2 n_i - \mu^2 \sum_{i=1}^k n_i \tag{5.8}$$

Per i dati in serie la formula semplificata della devianza è:

$$D(X) = \sum_{i=1}^n x_i^2 - n\mu^2 \tag{5.9}$$

ESEMPIO 1

Sia dato il seguente insieme di numeri:

4,5; 6,2; 7,8; 10,4; 15,9

Determinare:

- a) la devianza, utilizzando anche la formula semplificata;
 - b) la varianza;
 - c) lo scarto quadratico medio.
- a) La media aritmetica della distribuzione è $\mu = 8,96$, per cui la devianza si ottiene considerando il seguente schema:

x_i	$x_i - \mu$	$(x_i - \mu)^2$
4,5	-4,46	19,8916
6,2	-2,76	7,6176
7,8	-1,16	1,3456
10,4	1,44	2,0736
15,9	6,94	48,1636
Totale	—	79,092

Schema 1

La devianza è pari a:

$$D(X) = 79,092$$

Per applicare la formula semplificata della devianza è necessario fare riferimento al seguente schema che contiene i quadrati dei numeri dati:

x_i	x_i^2
4,5	20,25
6,2	38,44
7,8	60,84
10,4	108,16
15,9	252,81
Totale	480,5

Schema 2

Essendo $n\mu^2 = 5 \cdot (8,96)^2 = 401,408$, la devianza è:

$$D(X) = 480,5 - 401,408 = 79,092$$

- b) La varianza è uguale al rapporto tra la devianza e la numerosità n dell'insieme:

$$\sigma^2 = \frac{D(X)}{n} = \frac{79,092}{5} = 15,8184$$

- c) Lo scarto quadratico medio è uguale alla radice quadrata della varianza:

$$\sigma = \sqrt{15,8184} = 3,9772$$

ESEMPIO 2

Sia data la distribuzione riportata nella tabella 6 del quinto capitolo. Determinare:

- a) la devianza, utilizzando anche la formula semplificata;
 - b) la varianza;
 - c) lo scarto quadratico medio.
- a) Per determinare la devianza della distribuzione in esame è necessario, dapprima, calcolare la media aritmetica considerando i valori centrali delle classi (\bar{x}_i). A tal fine si consideri il seguente schema:

	\bar{x}_i	n_i	$\bar{x}_i n_i$
	173	14	2.422
	178	18	3.204
	183	28	5.124
	188	33	6.204
	193	17	3.281
	198	15	2.970
Totale	—	125	23.205

Schema 3

Pertanto, la media aritmetica è:

$$\mu = \frac{23.205}{125} = 185,64$$

Per calcolare la devianza si consideri lo schema a pagina seguente.

\bar{x}_i	n_i	$\bar{x}_i - \mu$	$(\bar{x}_i - \mu)^2$	$(\bar{x}_i - \mu)^2 n_i$
173	14	-12,64	159,77	2.236,77
178	18	-7,64	58,3696	1.050,65
183	28	-2,64	6,9696	195,15
188	33	2,36	5,5696	183,80
193	17	7,36	54,1696	920,88
198	15	12,36	152,77	2.291,54
Totale	—	125	—	6.878,80

Schema 4

La devianza è:

$$D(X) = 6.878,80$$

Per applicare la formula semplificata della devianza è necessario calcolare i quadrati dei valori centrali delle classi e fare riferimento al seguente schema:

\bar{x}_i	n_i	\bar{x}_i^2	$\bar{x}_i^2 n_i$
173	14	29.929	417.006
178	18	31.684	570.312
183	28	33.489	937.692
188	33	35.344	1.166.352
193	17	37.249	633.233
198	15	39.204	588.060
Totale	—	125	4.314.655

Schema 5

Pertanto, la devianza è:

$$D(X) = 4.314.655 - (185,64)^2 \cdot 125 = 6.878,80$$

b) La varianza è:

$$\sigma^2 = \frac{D(X)}{n} = \frac{6.878,80}{125} = 55,0304$$

c) Lo scarto quadratico medio è:

$$\sigma = \sqrt{\sigma^2} = \sqrt{55,0304} = 7,418$$

6. Le differenze medie

Le **differenze medie** sono indici di mutua variabilità che esaminano le differenze, in valore assoluto, tra le modalità a due a due $|x_i - x_j|$ con $(i, j) = 1, 2, \dots, n$, e ne operano una sintesi tramite una opportuna media.

In sostanza, il problema è calcolare la somma delle distanze tra ciascun termine e tutti gli altri, ossia:

$$\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \quad (6.1)$$

Le distanze sono calcolate in valore assoluto, in quanto considerandone i valori algebrici la sommatoria si annullerebbe.

Siccome il numero dei confronti tra le n modalità sono $n(n-1)$, non avendo considerato i confronti tra ciascuna modalità e se stessa, la sintesi è effettuata facendo la media aritmetica delle differenze ottenendo la **differenza semplice media**:

$$\Delta = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{n(n-1)} \quad (6.2)$$

Tale indice è una misura della variabilità media interna alla distribuzione, cioè fra i singoli valori tra di loro. Quando il confronto tra tutte le modalità è fatto tenendo conto anche della differenza di una modalità con se stessa allora si parla di **differenza semplice media con ripetizione** definita da:

$$\Delta_R = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{n^2} \quad (6.3)$$

I due indici sono legati dalla relazione:

$$\Delta = \Delta_R \frac{n}{n-1} \quad (6.4)$$

per cui ogni volta che si calcolerà l'uno il risultato potrà estendersi anche all'altro.

La differenza semplice media assume:

✓ **valore minimo** quando tutti i dati sono uguali tra loro per cui:

$$\Delta_{\min} = 0$$

- ✓ **valore massimo** quando $n - 1$ unità statistiche non possiedono alcuna modalità del carattere mentre l' n -esima possiede l'intero ammontare del carattere che è uguale a $n\mu$, per cui è:

$$\Delta_{\max} = 2\mu \quad (6.5)$$

Per **distribuzioni di frequenza** la differenza semplice media con ripetizione può essere calcolata applicando la formula di **De Finetti-Paciello**:

$$\Delta_R = \frac{2}{n^2} \sum_{i=1}^{k-1} n'_i (n - n'_i) (x_{(i+1)} - x_i) \quad (6.6)$$

ESEMPIO 1

Sia dato il seguente insieme di numeri:

4,6; 2,8; 3,5; 7,9; 5,3

Determinare:

- la differenza semplice media;
- la differenza semplice media con ripetizione.

Al fine della determinazione degli indici di variabilità richiesti predisponiamo una matrice delle differenze, dopo aver disposto in ordine crescente i numeri dati:

	2,8	3,5	4,6	5,3	7,9
2,8	0	-0,7	-1,8	-2,5	-5,1
3,5	0,7	0	-1,1	-1,8	-4,4
4,6	1,8	1,1	0	-0,7	-3,3
5,3	2,5	1,8	0,7	0	-2,6
7,9	5,1	4,4	3,3	2,6	0

Dalla matrice si evince che le differenze simmetriche, rispetto alla diagonale principale, sono opposte. Il numeratore delle differenze medie è, quindi, pari al doppio della somma dei valori assoluti delle differenze al di sopra (o al di sotto) della diagonale principale.

- La differenza semplice media è:

$$\Delta = \frac{2 \cdot (0,7 + 1,8 + 2,5 + 5,1 + 1,1 + 1,8 + 4,4 + 0,7 + 3,3 + 2,6)}{5 \cdot 4} = \frac{2 \cdot 24}{20} = 2,4$$

- La differenza semplice media con ripetizione è:

$$\Delta_R = \frac{2 \cdot 24}{5 \cdot 5} = 1,92$$