

Variabilità

Scuola Second 1° grado; Argomento: Leggiamo - Variabilità; (30.10.14); Pacchetto: S1.A.4

Strumento

LA VARIABILITÀ

Note metodologiche: definizioni

Quando si pensa alla statistica si pensa immediatamente alla *media* e in genere alla *media aritmetica*. Ma non sempre basta!

La media sintetizza la realtà fornendo indicazioni su ciò che può considerarsi "tipico" o centrale in un insieme di dati, cioè fornisce informazioni sull'ordine di grandezza del fenomeno studiato, tuttavia non basta da sola a sintetizzare i dati.

La media può assumere, infatti, lo stesso valore in distribuzioni che presentano valori diversi.

Ad esempio le due distribuzioni:

(2,4,6) e (4,4,4)

hanno la stessa media aritmetica pari a 4 ma differenti modalità.

Naturalmente la distribuzione è ben rappresentata dalla media quando gran parte delle unità presentano modalità vicina alla media: nel secondo caso la media sintetizza meglio che nel primo caso infatti la prima distribuzione presenta una maggiore dispersione nei confronti del valore medio.

Per comprendere la realtà che ci circonda non possiamo utilizzare soltanto la media; è invece necessario sfruttare l'informazione che deriva dalla diversità che sussiste fra i dati, cioè analizzarne la *variabilità*, considerando che un valore molto diverso dagli altri è portatore di un'informazione maggiore e quindi è da tenere nella massima considerazione. **La variabilità è l'attitudine di un fenomeno a presentarsi con modalità diverse.**

Gli indici di variabilità misurano la dispersione dei termini rispetto ad una media, o quanto differiscono fra loro le modalità presenti nelle unità, oppure la diversità fra due quantità, termini specifici osservati nel caso di caratteri quantitativi si utilizza o il valore assoluto o il quadrato della loro differenza. In tal modo l'aumentare di ogni "diversità" così misurata incrementa il valore dell'indice.

In sintesi possiamo individuare tre categorie di indici che misurano la variabilità del carattere mediante:

- una misura di diversità tra due valori particolari della distribuzione.
- una sintesi delle misure di diversità tra ogni valore osservato e una media
- una sintesi delle misure di tutti i termini della distribuzione, ossia tra tutte le possibili coppie di unità.

Un indice di variabilità deve soddisfare alcuni requisiti:

- deve essere nullo se e solo se tutte le unità presentano la stessa modalità del carattere, cioè non c'è variabilità, in questo caso si dice che la variabile non varia ed è una variabile degenere.
- deve essere positivo e aumentare all'aumentare della diversità tra le modalità assunte dalle varie unità.

Un indice di variabilità di rapido calcolo è rappresentato dal **campo di variazione (range)**:

RANGE o CAMPO DI VARIAZIONE = MAX-min

Si ottiene sottraendo il valore più basso del carattere da quello massimo presente nel collettivo. Per esempio, se in una classe della scuola media abbiamo rilevato le altezze in cm di 5 studenti

170, 174, 168, 179, 172,

il range, dato dalla differenza tra il massimo (179) e il minimo (168), è pari a 11 cm.

Questo indicatore ci dice che nel collettivo c'è una variabilità tra il più alto e il più basso di 11 cm.

E' evidente che basandosi solamente su 2 degli n valori osservati è piuttosto indicativo circa la reale variabilità dei dati, tutti gli altri valori osservati non vengono presi in considerazione.

Inoltre può accadere che i valori estremi siano dei valori anomali conducendo così ad una misura della variabilità molto grossolana. L'unico pregio risiede nella semplicità di calcolo.

Per ovviare a questo inconveniente **le differenze medie assolute** sono indici di variabilità che si basano sulla distanza di tutte le osservazioni tra di loro. Sono una sorta di media di scarti di ogni osservazione con tutte le altre senza tener conto del confronto di ogni termine con se stessi il numero totale delle differenze o scarti delle osservazioni sono $n \cdot (n-1)$

La formula della differenza semplice media è la seguente:

$$\Delta = \frac{\sum_{i=1}^n \sum_{h=1}^n |x_i - x_h|}{n \cdot (n-1)}$$

Dove il simbolo Σ (lettera greca sigma) indica l'operazione di somma, x_i e x_h le modalità assunte da ciascuna unità, n la numerosità del collettivo.

Considerando i dati delle altezze dei 5 studenti, per calcolare tutte le differenze in valore assoluto dei valori osservati è utile una tabella di questo tipo:

-	170	174	168	179	172
170	--	4	2	9	2
174	4	--	6	5	2
168	2	6	--	11	4
179	9	5	11	--	7
172	2	2	4	7	--

$$\Delta = \frac{\sum_{i=1}^n \sum_{h=1}^n |x_i - x_h|}{n \cdot (n-1)} = \frac{104}{5 \cdot 4} = 5,2 \text{ cm}$$

Un indice di variabilità meno "laborioso" dal punto di vista dei calcoli può essere ottenuto considerando gli scarti da ogni osservazione rispetto alla media aritmetica elevati al quadrato, questo indice si chiama **varianza**

Variabilità

Scuola Second 1° grado; Argomento: Leggiamo - Variabilità; (30.10.14); Pacchetto: S1.A.4

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

Per i dati relativi all'esempio sopraportato il valore della media è 172,6 cm, la varianza si ottiene sommando gli scarti al quadrato dei valori osservati dalla media e dividendo il tutto per il numero totale delle osservazioni

$$\sigma^2 = \frac{(170 - 172,6)^2 + (174 - 172,6)^2 + \dots + (172 - 172,6)^2}{5} = 14,24 \text{ cm}^2$$

Una varianza maggiore comporta maggiore dispersione dei dati rispetto alla media della popolazione.

La varianza è espressa nel quadrato dell'unità di misura della variabile osservata, quindi il suo valore non è di immediata interpretazione. Per questo motivo si usa più comunemente lo **sarto quadratico medio (s.q.m.)**, espresso dalla radice quadrata della varianza

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - M)^2}{n}} = 3,77 \text{ cm}$$

Questo indice è chiamato anche deviazione standard o scostamento quadratico medio.

Lo sarto quadratico medio viene generalmente indicato con σ ("sigma") e la varianza con σ^2 .

Il risultato è espresso nella stessa unità di misura del carattere (cm) e informa che le altezze dei 5 studenti si discostano mediamente dalla media aritmetica di 3,77 cm.

Nel caso di una *trasformazione lineare dei dati* la varianza risente solamente del coefficiente di scala elevato al quadrato, in altre parole

$$Y = a + bX$$

$$\sigma^2_Y = b^2 \sigma^2_X$$

Nel caso **dati in distribuzione di frequenza** le formule vanno modificate ponderando ogni sarto con la relativa frequenza per la varianza

$$\sigma^2 = \frac{\sum_{j=1}^k (x_j - M)^2 \cdot n_j}{n}$$

Dove l'indice $j=1\dots K$ indica il numero delle modalità, n_j le frequenze assolute.

Gli indici finora esaminati sono **indici assoluti** perché espressi nella stessa unità di misura del fenomeno in esame.

Se dobbiamo confrontare la variabilità di un carattere in collettivi diversi per grandezza media o di più caratteri espressi in diverse unità di misura (peso, altezza, reddito,...) lo sarto quadratico medio non è un buon indicatore perché è

espresso nell'unità di misura del carattere rilevato e non ha senso confrontare una variabilità espressa in cm con una variabilità espressa in kg o in euro.

In questi casi si ricorre agli **indici di variabilità relativa**. Tali indici si calcolano facendo il rapporto fra gli indici di variabilità assoluta e l'intensità media del fenomeno. Il più usato è il coefficiente di variabilità del Pearson dato dall'espressione:

$$CV = \frac{\sigma}{|M|} \cdot 100$$

Tanto σ che M sono espressi nell'unità di misura del carattere osservato, il numero risultante dal loro rapporto è un numero puro, adimensionale. Il coefficiente di variazione esprime la percentuale della variabilità per ogni unità di valore medio, dividendo lo s.q.m. per la media aritmetica si "neutralizza" l'effetto della scala di misura.

Ad esempio, se si vanno a misurare i pesi dei bambini nati in un certo periodo in un grande ospedale, e simultaneamente i pesi delle loro mamme, si osserverà certamente una deviazione standard molto inferiore nell'insieme dei bambini. Per forza! Infatti i bambini appena nati pesano soltanto due-tre o quattro chili ...quindi anche gli scarti dalla media dei loro pesi saranno piccolini!!! Volendo confrontare le due "variabilità" (quella dei pesi dei neonati con quella dei pesi delle mamme) si farà ricorso allora al coeff. di variazione.

Spunti per riflessioni

La variabilità può essere vista anche in modi diversi non solo in termini di dispersione dei valori osservati.

Per **mutabilità** si intende l'attitudine di un carattere qualitativo ad assumere modalità differenti.

Una popolazione è omogenea, rispetto ad un carattere, se la totalità delle frequenze è concentrata in una sola modalità in questo caso c'è assenza di variabilità.

Se invece le frequenze sono ripartite in parti uguali fra le tutte le modalità del carattere, allora si è in presenza della minima omogeneità e della massima variabilità.

Un particolare aspetto della variabilità di un fenomeno è la **concentrazione**. Lo studio della concentrazione è utile per vedere se il fenomeno è equamente distribuito fra tutte le unità statistiche oppure è concentrato in poche unità. Ad esempio, si può affermare che la distribuzione delle scuole elementari è poco concentrata sul territorio italiano perché quasi tutti i Comuni ne possiedono almeno una, mentre la distribuzione delle Università è molto concentrata perché pochi Comuni sono sedi di Università.

La varianza e lo "sarto quadratico medio" hanno un'importanza colossale in svariati argomenti, come per esempio nella **teoria degli errori di misura**.

Spunti per esercizi o attività

- Effettuare nella classe la rilevazione di uno o più caratteri quantitativi ad esempio peso, altezza, il numero di fratelli. Costruire la relativa tabella di frequenze assolute e a partire da questa calcolare media, scarto quadratico medio e coefficiente di variazione. Individuare la variabile più "variabile".

- Misurare la lunghezza di segmento, o il peso di un corpo attraverso ripetute misurazioni, la media aritmetica di tutte le misure ottenute è la misura attendibile della grandezza osservata, per valutare invece l'errore di rilevazione è utile calcolare lo s.q.m. delle misurazioni.

- Classificare le misurazioni (almeno 100) ripetute della lunghezza di un segmento, o del peso di un corpo in una tabella di frequenza. Può darsi che alcuni di questi numeri coincidano, ma in generale saranno invece un poco diversi fra loro.

Considerare il minimo e il massimo valore rilevato, e suddividere questo intervallo in un certo numero di sottointervalli, circa otto-dieci (in generale, se le misure sono n , si consiglia di far sì che il numero di intervalli non superi \sqrt{n}). Per ciascun sottointervallo, conteggiare il numero di osservazioni che cadono in quel sottointervallo il numero ottenuto è la **frequenza assoluta** di quella classe (o sottointervallo).

Costruire in excel la tabella di frequenza e il relativo istogramma. Osservare che le misure "centrali" della distribuzione saranno in linea di massima più frequenti, e quelle estreme meno. In effetti, nell'atto pratico della misurazione, si commettono sempre errori "casuali" talvolta in difetto talvolta in eccesso, e se il numero di misurazioni effettuate diventa alto, l'istogramma tenderà ad assomigliare a una curva "a campana" detta "gaussiana" (F. Gauss, 1777-1855).

- Considerare il grafico a linee di più indicatori in serie storica individuare dal grafico, calcolando approssimativamente il campo di variazione, l'indicatore più variabile. Grafici a linee quasi paralleli lungo l'asse del tempo evidenziano infatti fenomeni poco variabili, viceversa la presenza di picchi e bruschi crolli caratterizza un andamento più variabile.

- Leggere e soffermare l'attenzione sui grafici dinamici <http://www.istat.it/it/archivio/27587>. Sottolineare come la variabilità sia una caratteristica intrinseca di tutti i fenomeni collettivi studiati attraverso la statistica; cercare di capire le cause di valori eccezionali, anomali, mancanti permette infatti di conoscere meglio la realtà che ci circonda.

Bibliografia essenziale

G. Leti – Statistica descrittiva ed. il Mulino

S. Borra A. Di Caccio – Statistica ed. McGraw-Hill

M. Pelosi (et al.) – Introduzione alla statistica ed. McGraw-Hill

M. Fraire A. Rizzi – Elementi di statistica ed. NIS

Link Utili

<http://www3.istat.it/servizi/studenti/valoredati/>

<http://www.istat.it/it/files/2012/03/statisticaxesempi2.pdf>

http://www3.istat.it/servizi/studenti/valoredati/Cap4/Cap4_5_2.htm

<http://seriestoriche.istat.it/>

<http://noi-italia.istat.it/>

<http://vimeo.com/82504978>

http://www.fga.it/fileadmin/Documenti_Vari/Variabilita_01.pdf

Materiale operativo di riferimento

[*] → Variabilità

Nella cartella Esercizi

[*] _esercizi.xls file da utilizzare su PC o LIM

[*] _esercizi.ods file da utilizzare su PC o LIM

[*] _esxstampa.pdf file per la stampa degli esercizi

Nella cartella Lezioni interattive

[*] _ppt presentazione per PC o LIM

[*] _pdf presentazione per PC o LIM