L'analisi monovariata

- è un'analisi puramente descrittiva dei fenomeni
- informa come ogni variabile è distribuita tra i casi rilevati
- rappresenta un passaggio inevitabile e necessario ad ogni analisi bivariata e multivariata.
- da conto della struttura del campione e la sua rappresentatività, attraverso le distribuzioni di frequenza delle variabili sociografiche (genere, età, etc.).

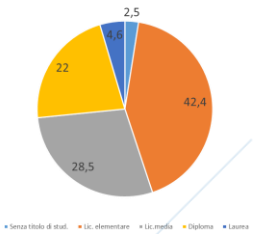
Distribuzione di frequenza

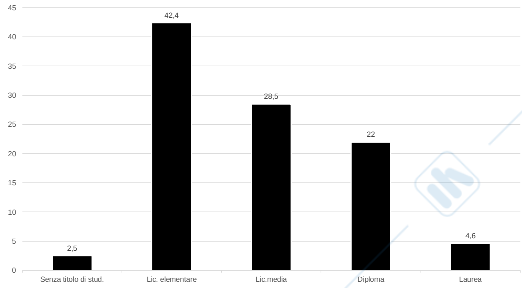
Definizione	Una serie nella quale ad ogni valore della variabile viene associata la frequenza con la quale esso si presenta
Presentazione:	Forma grafica o tabellare.
Frequenze assolute	Il numero di casi che ha lo stesso valore nella modalità della variabile
Frequenze relative	Sono il rapporto (divisione) tra i singoli valori ed il totale dei casi. Moltiplicando il risultato per 100 si ottengono le frequenze percentuali
Frequenza cumulata	la somma delle frequenze corrispondenti ad un valore ed a tutti quelli precedenti

Distribuzione di frequenza del titolo di studio, in forma tabellare

	Freq. Assol.	Freq. Proporz.	Perc. Percent.	Percen. Cumulat.
Senza titolo di stud.	30	0.025	**2.5**	2.5
Lic. elementare	509	0.424	**42.4**	44.9
Lic.media	342	0.285	**28.5**	73.4
Diploma	264	0.220	**22.0**	95.4
Laurea	55	0.046	**4.6**	100.0
Totale	**1.200**	**1**	**100.0**	

…o in forma grafica





Le caratteristiche delle tabelle

distribuzioni di frequenza compatte:
presentare solo le frequenze percentuali, accompagnate dall' indicazione della base del calcolo delle percentuali, cioè del totale del valore assoluto (N);

cifre decimali:
le distribuzioni di frequenza percentuali vengono presentate, per convenzione, con un decimale (caso più frequente) oppure senza decimali (consigliabile se la base delle percentuali è piccola, inferiore a 100)

arrotondamenti
se il decimale da eliminare si colloca tra 0 e 4, si arrotonda per difetto; se si colloca tra 5 e 9 si arrotonda per eccesso;

Le caratteristiche delle tabelle

il decimale zero:
lo 0 va esplicitato. Pertanto il ventidue per cento si scriverà 22.0% e non 22.

quadratura:
se il totale delle frequenze percentuali da come valore 99.9 oppure 100.1, bisogna guardare ai valori percentuali la cui alterazione è meno rilevante.

Indici statistici di sintesi

Per trarre delle indicazioni adeguate quando si considerano dati quantitativi, non è sufficiente rappresentare i dati mediante tabelle ed grafici di frequenza.

Una buona analisi dei dati richiede anche che le caratteristiche principali delle osservazioni siano sintetizzate con opportune misure, dette Indici Statistici, e che tali misure siano adeguatamente analizzate e interpretate.

Tipi di indici:

- Misure di tendenza centrale (Indici di posizione)
- Misure di Variabilità (Indici di dispersione)

8

Misure di Tendenza Centrale

Nella maggior parte degli insiemi di dati, le osservazioni mostrano una tendenza a raggrupparsi attorno a un valore centrale.

Obiettivo di una misura di posizione (*location index*) è quello di sintetizzare in un singolo valore l'intera distribuzione di frequenza per effettuare confronti nel tempo, nello spazio o tra circostanze differenti.

Tale valore descrittivo è una misura di posizione o di tendenza centrale.

Tipi di misure di tendenza centrale:

- Media
- Mediana
- Moda

9

Misure di Tendenza Centrale: la Media

La **media aritmetica** (anche chiamata semplicemente **media**) è la misura di posizione più comune. Si calcola dividendo la somma dei valori osservati per il numero totale di osservazioni.

La media aritmetica

La media aritmetica è la somma dei valori divisa per il numero dei valori.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

10

Misure di Tendenza Centrale: la Media

- La media è sempre compresa tra il minimo ed il massimo delle modalità della variabile
- La somma degli *scarti dalla media* ($\sum(x_i - \mu)$) è sempre nulla, per cui la media costituisce il "baricentro" di una distribuzione di frequenza.
- Il calcolo della media si basa su tutte le osservazioni (X_1, X_2, X_3, . . . , X_n) dell'insieme di dati, proprietà questa che non è presentata da nessun'altra misura di posizione comunemente usata.

11

Misure di Tendenza Centrale: la Media

***Commento**: quando usare la Media Aritmetica*

Proprio perché il calcolo della media si basa su tutte le osservazioni, tale misura di posizione risulta influenzata da valori estremi.

In presenza di valori estremi, la media aritmetica fornisce una rappresentazione distorta dei dati ed è pertanto opportuno in questi casi ricorrere ad altre misure di posizione.

12

Misure di Tendenza Centrale: la Mediana

La **mediana** (Galton 1883) è il valore centrale in un insieme di dati ordinati dal valore più piccolo al più grande

La mediana

La mediana è l'osservazione che, nella serie ordinata dei dati, si lascia alla destra il 50% delle osservazioni e a sinistra il 50% delle osservazioni. Quindi, il 50% delle osservazioni risulteranno maggiori della mediana e il 50% risulteranno minori della mediana.

$$\text{Mediana} = \text{osservazione di posto } \frac{n+1}{2} \text{ nella serie ordinata} \qquad (3.2)$$

***Commento**:* La mediana non è influenzata dalle osservazioni estreme di un insieme di dati: nel caso di osservazioni estreme è quindi opportuno descrivere l'insieme di dati con la mediana piuttosto che con la media.

13

Misure di Tendenza Centrale: la Mediana

Per trovare la posizione occupata dal valore mediano nella serie ordinata delle osservazioni si usa l'equazione (3.2) secondo una delle due regole seguenti:

REGOLA 1. Se l'ampiezza del campione è un numero **dispari**, la mediana coincide con il valore centrale, vale a dire con l'osservazione che occupa la posizione (n+1)/2 nella serie ordinata delle osservazioni.

REGOLA 2. Se l'ampiezza del campione è un numero **pari**, la mediana allora coincide con la media dei valori corrispondenti alle due osservazioni centrali.

14

Misure di Tendenza Centrale: la Mediana

Esempio 3.3 ***Il calcolo della mediana in un campione di ampiezza dispari***

Nel nostro esempio del rendimento percentuale a un anno conseguito dai fondi comuni azionari che prelevano le commissioni di commercializzazione direttamente dalle attività del fondo, i dati grezzi sono:

32.2 29.5 29.9 32.4 30.5 30.1 32.1 35.2 10.0 20.6 28.6 30.5 38.0 33.0 29.4 37.1 28.6

Calcolate la mediana.

Soluzione

La serie ordinata è:

	10.0	20.6	28.6	28.6	29.4	29.5	29.9	30.1	30.5	30.5	32.1	32.2	32.4	33.0	35.2	37.1	38.0
									↑ Mediana ↑								
Posizione	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

Mediana = 30.5

Per questi dati il valore centrale coincide con la nona osservazione nella serie ordinata [ossia, $(n + 1)/2 = (17 + 1)/2 = 9$]. Pertanto la mediana è 30.5.

Misure di Tendenza Centrale: la Moda

La **moda** è il valore più frequente in un insieme di dati.

- A differenza della media, la moda non è influenzata dagli outlier.
- Tuttavia tale misura di posizione viene usata solo per scopi descrittivi, poiché è caratterizzata da maggiore variabilità rispetto alle altre misure di posizione (piccole variazioni in un insieme di dati possono far variare in modo consistente la moda).
- La moda può non esistere o non essere unica, se unica la distribuzione è detta unimodale, quando ci sono più mode diverse è detta bimodale o multimodale.

Esempio 3.5 *Il calcolo della moda*

Calcolate la moda dei rendimenti percentuali annui conseguiti dai fondi comuni azionari che prelevano le commissioni direttamente dalle attività del fondo utilizzando la serie ordinata nell'esempio 3.3.

16

Misure di Tendenza Centrale: la Moda

Esempio 3.5 *Il calcolo della moda*

Calcolate la moda dei rendimenti percentuali annui conseguiti dai fondi comuni azionari che prelevano le commissioni direttamente dalle attività del fondo utilizzando la serie ordinata nell'esempio 3.3.

Soluzione

La serie ordinata per questi dati è la seguente:

10.0 20.6 28.6 28.6 29.4 29.5 29.9 30.1 30.5 30.5 32.1 32.2 32.4 33.0 35.2 37.1 38.0

Possiamo osservare che ci sono due valori "più tipici" o due mode: 28.6 e 30.5. Questo insieme di dati si dice *bimodale*.

NOTA: un insieme di dati può non avere moda, se nessuno valore è "più tipico".

17

Misure di Tendenza Centrale: i Quartili

Mentre la mediana è un valore che divide a metà la serie ordinata delle osservazioni, i **quartili** sono misure descrittive che dividono i dati ordinati in quattro parti.

Il primo quartile, Q_1

Il **primo quartile**, Q_1, è il valore tale che il 25% delle osservazioni è più piccolo di Q_1 e il 75% è più grande di Q_1.

$$Q_1 = \text{osservazioni di posto } \frac{(n+1)}{4} \text{ nella serie ordinata} \quad (3.4)$$

Il terzo quartile, Q_3

Il **terzo quartile**, Q_3 è il valore tale che il 75% delle osservazioni è più piccolo di Q_3 e il 25% delle osservazioni è più grande di Q_3.

$$Q_3 = \text{osservazioni di posto } \frac{3(n+1)}{4} \text{ nella serie ordinata} \quad (3.5)$$

18

Misure di Tendenza Centrale: i Quartili

Tre sono le regole usate per il calcolo dei quartili.

- *REGOLA 1*. Se il punto di posizionamento è un numero intero, si sceglie come quartile il valore dell'osservazione corrispondente.
- *REGOLA 2*. Se il punto di posizionamento è a metà tra due numeri interi, si sceglie come quartile la media delle osservazioni corrispondenti.
- *REGOLA 3*. Se il punto di posizionamento non è né un intero né a metà tra due numeri interi, una regola semplice consiste nell'approssimarlo per eccesso o per difetto all'intero più vicino e scegliere come quartile il valore numerico dell'osservazione corrispondente.

19

Misure di Tendenza Centrale: i Quartili

Esempio 3.8 *Il calcolo dei quartili*

Calcolate i quartili dei rendimenti percentuali annui conseguiti dai fondi comuni azionari che prelevano le commissioni dalle attività del fondo considerati nell'esempio 3.3.

Soluzione

La serie ordinata è

10.0 20.6 28.6 28.6 29.4 29.5 29.9 30.1 30.5 30.5 32.1 32.2 32.4 33.0 35.2 37.1 38.0

Per questi dati abbiamo

$$Q_1 = \frac{n+1}{4} \text{–esima osservazione ordinata}$$
$$= \frac{17+1}{4} = 4.5\text{–esima osservazione ordinata}$$

Pertanto Q_1, usando la regola 2, può essere approssimato con la media tra la quarta e la quinta osservazione nella serie ordinata.

$$Q_1 = \frac{28.6 + 29.4}{2} = 29.0$$

$$Q_3 = \frac{3(n+1)}{4} \text{–esima osservazione ordinata}$$
$$= \frac{3(17+1)}{4} = 13.5\text{–esima osservazione ordinata.}$$

Pertanto Q_3, usando la regola 2, può essere approssimato con la media tra la tredicesima e la quattordicesima osservazione nella serie ordinata.

$$Q_3 = \frac{32.4 + 33.0}{2} = 32.7$$

Interpretazione della Varianza e dello Scarto Quadratico Medio

- Lo scarto quadratico medio ci aiuta a stabilire se e quanto i dati sono concentrati o dispersi intorno alla loro media.
- Per quasi tutti gli insiemi di dati, la maggior parte dei valori osservati si trova nell'intervallo centrato sulla media e i cui estremi distano dalla media per 1 scarto quadratico medio.

Commento: ***Cosa indica lo scarto quadratico medio***

Nel campione dei 17 fondi comuni azionari che prelevano le commissioni dalle attività del fondo, lo scarto quadratico medio del rendimento percentuale a un anno è 6.42. Ci aspettiamo allora che i rendimenti percentuali annui della maggioranza dei fondi nel campione si raggruppino nel raggio di 6.42 punti dalla media (vale a dire che si raggruppano tra $\bar{X} - 1S = 23.44$ e $\bar{X} + 1S = 36.28$). In effetti, possiamo osservare che i rendimenti del 76.5% dei fondi (13 su 17) cadono in questo intervallo.

21