
SINTETIZZARE I DATI

— Gli indici di posizione e
di variabilità —

PREMESSA

IL PROBLEMA DELLA SINTESI DEI DATI

Al termine di un processo di rilevazione, spesso i dati raccolti sono molto numerosi ed è pertanto necessario procedere a una sintesi.

La statistica offre numerosi **indici di sintesi**; in questa sede ci occuperemo di:

- **Indici di posizione** (media aritmetica, moda e mediana)
- **Indici di variabilità** (scarto quadratico medio, coefficiente di variazione)

PREMESSA

In questa trattazione, a titolo di esempio, costruiremo gli indici di sintesi relativi a una rilevazione che, per un certo Comune italiano, ha individuato il numero di dipendenti assunti dalle cinque principali imprese.

I dati vengono inseriti in Excel, in modo da utilizzarne le funzioni statistiche.

	A	B
1	IMPRESA	NR DIPENDENTI
2	Alfa Spa	200
3	Beta Spa	100
4	Delta Spa	300
5	Gamma Spa	200
6	Lambda Spa	400

La variabile osservata (X) descrive il Nr di dipendenti per ogni impresa.

Per ogni osservazione (Impresa) la variabile assume un valore (x_i) scritto in rosso.

Il numero di osservazioni effettuate (N) coincide con il numero di imprese (5)

LA MEDIA ARITMETICA

- La media aritmetica è calcolato sommando tutti i valori osservati (x) e dividendo la somma così ottenuta per il numero delle osservazioni N.

$$\text{media aritmetica} = \sum_{1}^{N} \frac{x_i}{N}$$

- La media aritmetica è uno strumento utile quando si vuole fornire un valore che si discosti il meno possibile dal complesso dei valori effettivamente osservati.
- Il suo utilizzo è sconsigliato quando vi sono alcuni valori anomali definiti «**outlier**» (molto piccoli o molto grandi) in quanto produrrebbe un indice distorto essendo particolarmente sensibile (influenzato) a tali valori.

LA MEDIA ARTIMETICA

	A	B
1	IMPRESA	NR DIPENDENTI
2	Alfa Spa	200
3	Beta Spa	100
4	Delta Spa	300
5	Gamma Spa	200
6	Lambda Spa	400

Nel nostro esempio la media aritmetica rappresenta il numero medio di dipendenti per ogni impresa.

Si somma il numero di dipendenti osservato per ogni impresa:

$$200 + 100 + 300 + 200 + 400 = 1.200$$

Lo si divide quindi la somma per N (5)

$$1.200 / 5 = 240$$

La media di dipendenti per impresa è 240.

LA MODA

- La moda è la modalità **più frequente** fra quelle osservate ovvero fra un insieme di osservazioni è quella che si presenta con una frequenza maggiore.
- Una distribuzione si dice **unimodale** se ha una sola moda, **bimodale** se ha due mode, **plurimodale** se ne ha più di due.

	A	B
1	IMPRESA	NR DIPENDENTI
2	Alfa Spa	200
3	Beta Spa	100
4	Delta Spa	300
5	Gamma Spa	200
6	Lambda Spa	400

Nel nostro esempio abbiamo osservato:

- 1 impresa con 100 dipendenti
- 2 imprese con 200 dipendenti
- 1 impresa con 300 dipendenti
- 1 impresa con 400 dipendenti

La moda è la modalità più frequente ovvero il numero di dipendenti per impresa osservato più volte: 200.
Si tratta di una distribuzione unimodale

LA MEDIANA

- La mediana è il **valore centrale** in una serie di dati ordinati in ordine crescente.
- Se le modalità osservate (**xi**) sono in numero dispari è immediato calcolare la mediana, viceversa occorre individuare due valori mediani oppure un valore intermedio fra di essi.
- Al contrario della media la mediana non è sensibile ad outlier.

	A	B
	IMPRESA	NR DIPENDENTI
1		
2	Alfa Spa	200
3	Beta Spa	100
4	Delta Spa	300
5	Gamma Spa	200
6	Lambda Spa	400

Ordinando in modo crescente i dati osservati nel nostro esempio, si avrà:

100 - 200 - 200 - 300 - 400

Il valore centrale è pertanto il terzo (200) che lascia alla propria «sinistra» e alla propria «destra» lo stesso numero di osservazioni (2)

GLI INDICI DI POSIZIONE IN EXCEL



MEDIA	Media aritmetica di una distribuzione
MODA	Moda di una distribuzione
MEDIANA	Mediana di una distribuzione



	A	B	C	D	E
	IMPRESA	NR DIPENDENTI			
1					
2	Alfa Spa	200		MEDIA	240
3	Beta Spa	100		MODA	200
4	Delta Spa	300		MEDIANA	200
5	Gamma Spa	200			
6	Lambda Spa	400			

LO SCARTO QUADRATICO MEDIO

- Definito anche deviazione standard
- Esso si calcola come somma delle differenze fra i valori osservati e la media:

$$\text{Deviazione standard} = \sqrt{\sum_{1}^{N} \frac{(xi - media)^2}{N}}$$

- Essa misura la variabilità ovvero in media quanto i valori osservati si discostano dalla media
- Maggiore è questo valore, maggiore è la variabilità

GLI INDICI DI POSIZIONE IN EXCEL



DEV.ST.P

Deviazione standard di una distribuzione (popolazione)



	A	B	C	D	E
1	IMPRESA	NR DIPENDENTI			
2	Alfa Spa	200		MEDIA	240
3	Beta Spa	100		MODA	200
4	Delta Spa	300		MEDIANA	200
5	Gamma Spa	200			
6	Lambda Spa	400		DEVIAZIONE STANDARD	101,98

IL COEFFICIENTE DI VARIAZIONE

- Per esprimere una valutazione più precisa sulla variabilità dei dati si ricorre al coefficiente di variazione calcolato come rapporto percentuale fra scarto quadratico medio e media:

$$\text{Coefficiente di variazione} = (\text{deviazione standard} / \text{media}) \times 100$$

Nel nostro esempio:

$$\text{Coefficiente di variazione} = (101,98 / 240) \times 100 = 42,49\%$$

I dati rilevati mostrano una elevata variabilità pari 42,49% della media.

PROVA TU

Considera i seguenti dati e procedi al calcolo degli indici di posizione e di variabilità.

ALUNNO	VOTO DI DIPLOMA
Marco	100
Matteo	80
Luca	70
Giovanni	60
Mattia	60