

Statistica descrittiva

Misure di tendenza centrale e di dispersione

Per una opportuna descrizione della distribuzione dei valori di una variabile di una popolazione di interesse in uno studio scientifico sono necessarie due misure, una di tendenza centrale, l'altra di dispersione.

La misura di tendenza centrale informa su attorno a quale valore tendono a raggrupparsi le osservazioni, mentre la misura di dispersione ci informa di quanto le singole osservazioni si discostano dal valore centrale.

Per una variabile continua, la misura di tendenza centrale è rappresentata dalla media, mentre la misura di dispersione è rappresentata dalla deviazione standard.

Statistica descrittiva

Misure di tendenza centrale e di dispersione

- Per qualsiasi tipo di variabile è inoltre possibile calcolare la moda, cioè il valore più frequente nella distribuzione.
- Per le variabili ordinali e nominali non è possibile utilizzare misure riassuntive (i valori della variabile sono categorie, ordinate o meno), ma solo la percentuale di casi in ogni categoria.

Variabile	Misura di tendenza centrale	Misura di dispersione
Continua	Media (μ)	Deviazione standard (σ)
Discreta	Mediana	Range

Statistica descrittiva

Misure di tendenza centrale e di dispersione

- La **MEDIA** è la somma dei valori di una data variabile per ogni osservazione divisa per il numero complessivo delle osservazioni
- La **MEDIANA** può essere definita come il valore che divide la popolazione a metà.
- È il valore centrale per un numero dispari di osservazioni, mentre corrisponde alla media dei due valori centrali nel caso di un numero pari di osservazioni. In altre parole, rappresenta quel valore che divide a metà la popolazione, così che il 50% ha un valore pari o più alto della mediana, e il 50% un valore pari o più basso.

Statistica descrittiva

Misure di tendenza centrale e di dispersione

- La **varianza** indica la misura quanto i valori di una distribuzione si allontanano dalla media
- Il **range** invece si riferisce al valore minimo e al valore massimo riscontrati nel campione in studio.
Disponendo in ordine crescente i valori, il range è dato dal primo e dall'ultimo valore nella sequenza.
Nel caso di una variabile ordinale, il range (vale a dire il valore minimo e quello massimo riscontrati) ne misura la dispersione.

Statistica descrittiva

Misure di tendenza centrale e di dispersione

Il **range interquartile** si riferisce ai valori relativi al primo e al terzo quartile (25° e 75° percentile).

Infatti il range potrebbe dare una informazione imprecisa; infatti, il valore più alto e quello più basso potrebbero rappresentare valori estremi atipici, e non riflettere la reale variabilità della misura nella popolazione in oggetto.

Per tale motivo, si utilizza di solito il range interquartile, dove i quartili sono i valori che dividono la popolazione in quattro gruppi, ognuno dei quali contiene il 25% del campione; in tale contesto la mediana è rappresentata dal 50° percentile.

Statistica descrittiva

Misure di tendenza centrale e di dispersione

Infine, la MODA è la misura di tendenza centrale apparentemente meno utilizzata, che non rappresenta altro che il valore che si presenta con maggiore frequenza.

In alcuni casi, dei campioni di dati possono presentare due o più mode; in tali occasioni, utilizzare la media come misura di tendenza centrale potrebbe essere inappropriato dato che, in fondo, esistono due o più punteggi caratteristici

Statistica descrittiva

RIASSUMENDO:

Per ciò che riguarda la scelta delle misure di tendenza centrale:

- La MEDIA dovrebbe essere scelta soltanto quando non sono presenti valori anomali, dato che essa è fortemente influenzata da tali valori
- La MEDIANA è invece un'utile misura di tendenza centrale anche in presenza di valori estremi
- La MODA è invece utile in presenza di dati nominali o categoriali

Statistica descrittiva

RIASSUMENDO:

Per ciò che riguarda la scelta delle misure di dispersione:

- Il RANGE ed il CAMPO DI VARIAZIONE (distanza tra punteggio massimo e minimo di un campione) rappresentano le misure più semplici per indicare la variabilità del campione stesso
- Lo SCARTO INTERQUARTILE invece, che considera un campo di variazione ridotto (trimmed range), risente meno della non uniformità del campione nonché della presenza di punteggi estremi
- Infine la VARIANZA, più della deviazione standard, evita che sia sovrastimata la variabilità presente nell'intera popolazione