Lezione sugli Indici di Dispersione: Media, Varianza e Deviazione Standard

1. Introduzione alla Media, Varianza e Deviazione Standard

- **Media**: La media rappresenta il valore medio di un insieme di dati, ottenuta sommando tutti i valori e dividendo per il numero totale di elementi. Indica il centro della distribuzione dei dati.

  Formula della media per un campione $(\bar{x})$:

  $$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

  Formula della media per una popolazione (μ):

  $$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- **Varianza**: La varianza misura quanto i dati siano diffusi attorno alla media, considerando la somma dei quadrati delle deviazioni di ciascun valore dalla media.

  Formula della varianza per un campione (s^2):

  $$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

  Formula della varianza per una popolazione (σ^2):

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

- **Deviazione Standard**: La deviazione standard è la radice quadrata della varianza e rappresenta la dispersione dei dati in unità comparabili con i valori originali.

  Formula della deviazione standard per un campione (s):

  $$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

  Formula della deviazione standard per una popolazione (σ):

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

2. Differenza tra Varianza di Popolazione e di Campione

La differenza tra le formule per la varianza (e di conseguenza per la deviazione standard) di popolazione e campione è data dal divisore:

- **Popolazione (N)**: si usa l'intero numero di elementi (N).
- **Campione (n-1)**: si usa $n - 1$ per evitare la sottostima della varianza. Questo correttore $(n - 1)$, detto **correttore di Bessel**, aumenta la varianza del campione per renderla un'approssimazione più accurata della varianza della popolazione.

3. Esempio Pratico

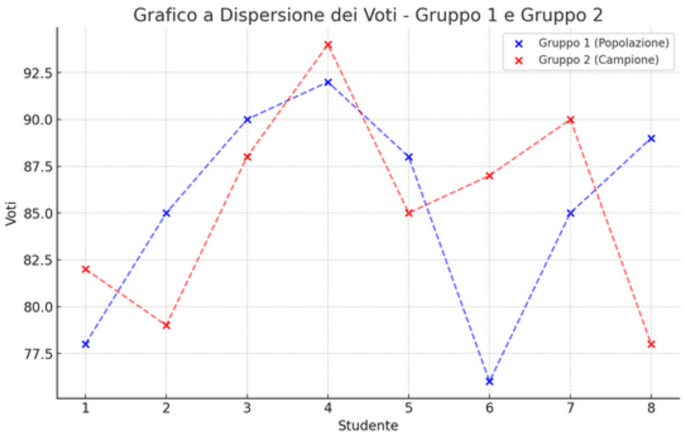
Confrontiamo la dispersione dei voti di due gruppi di studenti. Supponiamo che il primo gruppo sia una popolazione e il secondo un campione.

Studente	Gruppo 1 (Popolazione)	Gruppo 2 (Campione)
1	78	82
2	85	79
3	90	88
4	92	94
5	88	85
6	76	87
7	85	90
8	89	78

Calcolo della Media, Varianza e Deviazione Standard

1. **Media**:
   - Gruppo 1 (Popolazione): $\mu = \frac{78+85+90+92+88+76+85+89}{8} = 85.38$
   - Gruppo 2 (Campione): $\bar{x} = \frac{82+79+88+94+85+87+90+78}{8} = 85.38$
2. **Varianza**:
   - Gruppo 1 (Popolazione): $\sigma^2 = \frac{(78-85.38)^2+(85-85.38)^2+...+(89-85.38)^2}{8} = 30.86$
   - Gruppo 2 (Campione): $s^2 = \frac{(82-85.38)^2+(79-85.38)^2+...+(78-85.38)^2}{8-1} = 34.17$
3. **Deviazione Standard**:
   - Gruppo 1 (Popolazione): $\sigma = \sqrt{30.86} \approx 5.56$
   - Gruppo 2 (Campione): $s = \sqrt{34.17} \approx 5.85$

Grafico A Dispersione Dei Voti - Gruppo 1 E Gruppo 2



Commento del Grafico a Dispersione

Il grafico a dispersione mostra i voti degli studenti di **due gruppi**: il **Gruppo 1** (considerato una popolazione, in blu) e il **Gruppo 2** (considerato un campione, in rosso). Ogni punto rappresenta il voto di uno studente, mentre le linee tratteggiate indicano l'andamento dei voti all'interno di ciascun gruppo.

Osservazioni:

- **Distribuzione e Media**: Entrambi i gruppi hanno una media simile, intorno a **85.38**. Questo valore è rappresentativo del centro dei dati in entrambe le distribuzioni.
- **Dispersione e Varianza**: La dispersione dei dati, indicata dalla variabilità dei punti, è maggiore nel Gruppo 2 (rosso) rispetto al Gruppo 1 (blu). Ciò è confermato dai calcoli della **varianza**:
  - **Varianza del Gruppo 1 (Popolazione)**: circa **30.86**
  - **Varianza del Gruppo 2 (Campione)**: circa **34.17**

  La varianza più alta del Gruppo 2 indica una maggiore diversità nei voti degli studenti, con alcuni studenti che hanno ottenuto punteggi molto più alti o più bassi rispetto alla media.
- **Deviazione Standard**: La **deviazione standard** è leggermente più alta nel Gruppo 2 (circa 5.85 rispetto a 5.56 del Gruppo 1), il che conferma che i punteggi del Gruppo 2 sono più distribuiti rispetto alla media.

Questo grafico e l'analisi dei valori di varianza e deviazione standard aiutano a visualizzare e comprendere meglio la dispersione dei dati in ciascun gruppo, dimostrando come i valori possano differire anche se le medie risultano simili.

Il **correttore di Bessel** è un fattore utilizzato nel calcolo della varianza e della deviazione standard quando si analizza un **campione** anziché una **popolazione**. Questo correttore si applica dividendo la somma dei quadrati delle differenze tra ogni valore e la media campionaria per $n-1$ (dove n è il numero di osservazioni nel campione) invece che per n. In breve, il correttore di Bessel permette di ottenere una stima più accurata della varianza della popolazione a partire da un campione.

Perché è Necessario il Correttore di Bessel?

Quando calcoliamo la varianza di una popolazione, possiamo confrontare ogni valore con la **media della popolazione**. Tuttavia, quando lavoriamo con un campione, non abbiamo la media della popolazione, ma solo la **media campionaria**, che tende a trovarsi più vicino ai valori del campione rispetto alla media reale della popolazione. Questo rende il campione **meno disperso** di quanto non sarebbe l'intera popolazione.

Dividendo per $n-1$ invece che per n, si corregge questa sottostima della dispersione, rendendo la varianza campionaria una **stima imparziale** della varianza della popolazione.

Come Funziona il Correttore di Bessel?

Matematicamente, la varianza campionaria si calcola così:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

dove:

- x_i è ogni singolo valore del campione,
- $\bar{x}$ è la media campionaria,
- n è il numero totale di valori nel campione.

Dividendo per $n-1$ invece che per n, si aumenta leggermente il valore della varianza, compensando la tendenza della media campionaria a trovarsi vicino ai valori osservati, cosa che ridurrebbe artificialmente la varianza.

Esempio Pratico

Supponiamo di avere i seguenti punteggi di tre studenti: 80, 85 e 90. Calcoliamo la varianza considerando sia la popolazione che il campione.

1. **Media**: $\bar{x} = \frac{80+85+90}{3} = 85$
2. **Calcolo della varianza**:
   - Senza il correttore di Bessel (considerando la popolazione):
     $$\sigma^2 = \frac{(80-85)^2 + (85-85)^2 + (90-85)^2}{3} = \frac{25+0+25}{3} = \frac{50}{3} \approx 16.67$$
   - Con il correttore di Bessel (considerando il campione):
     $$s^2 = \frac{(80-85)^2 + (85-85)^2 + (90-85)^2}{3-1} = \frac{25+0+25}{2} = \frac{50}{2} = 25$$

In questo caso, la varianza campionaria risulta maggiore (25) rispetto alla varianza della popolazione (16.67), grazie al correttore di Bessel. Questo approccio permette di compensare l'errore introdotto dal calcolo della media basata solo su un campione e offre una stima più accurata della varianza della popolazione.

Lezione su Media, Varianza e Deviazione Standard: Confronto tra due Gruppi di Dati

Obiettivo della lezione:

Dimostrare come due gruppi di dati possano avere la stessa media ma diverse dispersioni, usando la varianza e la deviazione standard per evidenziare queste differenze.

---

1. Definizione di Media, Varianza e Deviazione Standard

- **Media**: La media rappresenta il valore medio di un insieme di dati. Si ottiene sommando tutti i valori e dividendo per il numero di elementi.

  Formula in Excel per calcolare la media:

  excel Copy code

  ```excel
  =MEDIA(intervallo_dati)
  ```

- **Varianza**: La varianza misura la dispersione dei dati rispetto alla media. Si calcola sommando i quadrati delle differenze tra ciascun valore e la media, diviso per $n - 1$ se è un campione.

  Formula in Excel per la varianza campionaria:

  excel Copy code

  ```excel
  =VAR.S(intervallo_dati)
  ```

  Formula in Excel per la varianza della popolazione:

  excel Copy code

  ```excel
  =VAR.P(intervallo_dati)
  ```

- **Deviazione Standard**: La deviazione standard è la radice quadrata della varianza e fornisce la dispersione dei dati in unità originali.

  Formula in Excel per la deviazione standard campionaria:

  excel Copy code

  ```excel
  =DEV.ST(intervallo_dati)
  ```

  Formula in Excel per la deviazione standard della popolazione:

  excel Copy code

  ```excel
  =DEV.ST.P(intervallo_dati)
  ```

2. Esempio Pratico in Excel

Consideriamo due gruppi di studenti (Gruppo 1 e Gruppo 2) con punteggi simili in media, ma differenze nella dispersione dei valori.

Studente	Gruppo 1	Gruppo 2
1	80	60
2	85	90
3	85	70
4	90	95
5	85	65
6	85	90
7	85	55
8	85	100
9	85	60
10	85	85

Calcoli delle Statistiche

1. **Media**:
   - In Excel, la formula della media per i due gruppi è:
     - **Gruppo 1**: `=MEDIA(B2:B11)` → **85**
     - **Gruppo 2**: `=MEDIA(C2:C11)` → **85**

   Entrambi i gruppi hanno la stessa media, **85**, indicando che il valore centrale dei dati è uguale.

2. **Varianza**:
   - In Excel, calcoliamo la varianza campionaria per i due gruppi:
     - **Gruppo 1**: `=VAR.S(B2:B11)` → **6.67**
     - **Gruppo 2**: `=VAR.S(C2:C11)` → **184.44**

   La varianza del **Gruppo 2** è molto più alta rispetto al **Gruppo 1**, il che indica una maggiore dispersione dei dati.

3. **Deviazione Standard**:
   - In Excel, calcoliamo la deviazione standard campionaria per i due gruppi:
     - **Gruppo 1**: `=DEV.ST(B2:B11)` → **2.58**
     - **Gruppo 2**: `=DEV.ST(C2:C11)` → **13.58**

   Anche la deviazione standard è molto più alta per il **Gruppo 2**, confermando la maggiore dispersione dei voti.

3. Visualizzazione Grafica della Dispersione

Passaggi per creare un grafico a dispersione in Excel:

1. Inserire i dati dei due gruppi in colonne adiacenti (come nella tabella precedente).
2. Selezionare i dati.
3. Andare su **Inserisci** > **Grafico a Dispersione** > **Dispersione con Linee**.

Interpretazione del grafico:

- Nel grafico, i punti del **Gruppo 1** saranno concentrati vicino alla media (85), mentre i punti del **Gruppo 2** saranno più distanti, indicando una maggiore variabilità.

---

4. Commento sui Risultati

- **Media**: Sebbene i due gruppi abbiano la stessa media, la variabilità dei dati è molto diversa. Questo dimostra che la media da sola non è sufficiente per descrivere un dataset, in quanto non fornisce informazioni sulla dispersione.
- **Varianza e Deviazione Standard**: La varianza e la deviazione standard rivelano la **dispersione** all'interno dei dati. Il **Gruppo 1** ha varianza e deviazione standard basse, indicando che i valori sono concentrati attorno alla media. Al contrario, il **Gruppo 2** ha varianza e deviazione standard elevate, suggerendo una forte variabilità.
- **Importanza della Dispersione**: Questo esempio mostra che due gruppi con la stessa media possono differire notevolmente in termini di dispersione. La dispersione è fondamentale per comprendere la consistenza dei dati: valori con bassa dispersione indicano che i dati sono simili tra loro, mentre alta dispersione indica differenze significative.