Elementi di statistica

Federico Bellisardi

Università di Bologna - FABIT

federico.bellisardi2@unibo.it

A.A. 2023 - 2024

Popolazione e Variabili

Popolazione Statistica

Un insieme su cui viene effettuato lo studio di un fenomeno.

- ES. Insieme degli studenti di una classe.

Popolazione e Variabili

Popolazione Statistica

Un insieme su cui viene effettuato lo studio di un fenomeno.

- ES. Insieme degli studenti di una classe.

Variabile Statistica

Una caratteristica assegnata a ogni elemento di una popolazione statistica data. Definiamo **variabile statistica** una rilevazione di una variabile sulla popolazione oppure su un campione, cioè un sottoinsieme della popolazione in cui è possibile stimare la caratteristica che si vuole rilevare.

Popolazione e Variabili

Popolazione Statistica

Un insieme su cui viene effettuato lo studio di un fenomeno.

- ES. Insieme degli studenti di una classe.

Variabile Statistica

Una caratteristica assegnata a ogni elemento di una popolazione statistica data. Definiamo **variabile statistica** una rilevazione di una variabile sulla popolazione oppure su un campione, cioè un sottoinsieme della popolazione in cui è possibile stimare la caratteristica che si vuole rilevare.

Sia N il numero di elementi nel campione, una variabile statistica è identificata da

$$Y = (y_1, y_2, \cdots, y_N) \tag{1}$$

- ES. I valori rilevati per l'altezza, il peso, i voti, il colore degli occhi.

Frequenze Assolute e in Percentuale

Modalità e Frequenza assoluta

Sia $Y = (y_1, y_2, \cdots, y_N)$ una variabile statistica discreta. Definiamo

- **modalità** i valori distinti tra $(y_1, y_2, \cdots, y_N)$;
- **frequenza assoluta** di una modalità il numero di volte che viene osservata nell'espressione della variabile statistica.

Frequenze Assolute e in Percentuale

Modalità e Frequenza assoluta

Sia $Y = (y_1, y_2, \cdots, y_N)$ una variabile statistica discreta. Definiamo

- **modalità** i valori distinti tra $(y_1, y_2, \cdots, y_N)$;
- **frequenza assoluta** di una modalità il numero di volte che viene osservata nell'espressione della variabile statistica.

Frequenza Relativa e Frequenza in Percentuale

Sia $Y = (y_1, y_2, \cdots, y_N)$ una variabile statistica discreta e f la frequenza assoluta della modalità z. Definiamo

- **frequenza relativa** della modalità z il rapporto f/N, dove N è il numero di elementi del campione;
- **frequenza percentuale** $f_{relativa} * 100$.

Esempio

Considerare la variabile statistica rappresentante il voto ad un esame di N studenti iscritti al primo anno di università:

$$Y = (18, 20, 26, 30, 30, 24, 25, 21, 24, 19, 24, 22, 22, 27, 30, 18, 24, 24, 28, 27)$$

- Qual è il numero di studenti del campione?

Esempio

Considerare la variabile statistica rappresentante il voto ad un esame di N studenti iscritti al primo anno di università:

$$Y = (18, 20, 26, 30, 30, 24, 25, 21, 24, 19, 24, 22, 22, 27, 30, 18, 24, 24, 28, 27)$$

- Qual è il numero di studenti del campione? 20
- Esprimere in termini di modalità, frequenza assoluta, frequenza relativa e frequenza percentuale il campione.

Esempio

Considerare la variabile statistica rappresentante il voto ad un esame di N studenti iscritti al primo anno di università:

$$Y = (18, 20, 26, 30, 30, 24, 25, 21, 24, 19, 24, 22, 22, 27, 30, 18, 24, 24, 28, 27)$$

- Qual è il numero di studenti del campione? 20
- Esprimere in termini di modalità, frequenza assoluta, frequenza relativa e frequenza percentuale il campione.

Modalità	18	20	26	30	24	...	25
Freq. Assoluta	2	1	1	3	5	...	1
Freq. Relativa	0.1	0.05	0.05	0.15	0.25	...	0.05
Freq. Percentuale	10	5	5	15	25	...	5

Esempio

Considerare la variabile statistica rappresentante il voto ad un esame di N studenti iscritti al primo anno di università:

$$Y = (18, 20, 26, 30, 30, 24, 25, 21, 24, 19, 24, 22, 22, 27, 30, 18, 24, 24, 28, 27)$$

- Qual è il numero di studenti del campione? 20
- Esprimere in termini di modalità, frequenza assoluta, frequenza relativa e frequenza percentuale il campione.

Modalità	18	20	26	30	24	...	25
Freq. Assoluta	2	1	1	3	5	...	1
Freq. Relativa	0.1	0.05	0.05	0.15	0.25	...	0.05
Freq. Percentuale	10	5	5	15	25	...	5

E se avessimo un campione di 300 studenti?

Classi

Anziché rappresentare la variabile statistica attraverso i suoi 300 valori (cioè i voti conseguiti dagli studenti), utilizziamo una separazione in classi, dove arbitrariamente consideriamo intervalli di 2 punti.

Classi	[18,20)	[20,22)	[22,24)	[24,26)	[26,28)	[28, 30)
Freq. Assoluta	80	55	65	40	35	25
Freq. Relativa	0.27	0.18	0.22	0.13	0.12	0.08

Classi

Anziché rappresentare la variabile statistica attraverso i suoi 300 valori (cioè i voti conseguiti dagli studenti), utilizziamo una separazione in classi, dove arbitrariamente consideriamo intervalli di 2 punti.

Classi	[18,20)	[20,22)	[22,24)	[24,26)	[26,28)	[28, 30)
Freq. Assoluta	80	55	65	40	35	25
Freq. Relativa	0.27	0.18	0.22	0.13	0.12	0.08

Quanti studenti hanno ottenuto un voto inferiore a 24?

Classi

Anziché rappresentare la variabile statistica attraverso i suoi 300 valori (cioè i voti conseguiti dagli studenti), utilizziamo una separazione in classi, dove arbitrariamente consideriamo intervalli di 2 punti.

Classi	[18,20)	[20,22)	[22,24)	[24,26)	[26,28)	[28, 30)
Freq. Assoluta	80	55	65	40	35	25
Freq. Relativa	0.27	0.18	0.22	0.13	0.12	0.08

Quanti studenti hanno ottenuto un voto inferiore a 24?

$$\frac{(27 + 18 + 22)}{100} \times 300 \approx 200$$

Rappresentazione Grafica dei Dati

Consideriamo le diverse altezze degli studenti rilevate in centimetri:

Altezza	157	160	165	173	168	176	184
Freq.Ass	1	2	6	4	3	3	1

- Grafico a Bastoncini;

Rappresentazione Grafica dei Dati

Consideriamo le diverse altezze degli studenti rilevate in centimetri:

Altezza	157	160	165	173	168	176	184
Freq.Ass	1	2	6	4	3	3	1

- Grafico a Bastoncini;
- Grafico Poligonale;

Rappresentazione Grafica dei Dati

Consideriamo le diverse altezze degli studenti rilevate in centimetri:

Altezza	157	160	165	173	168	176	184
Freq.Ass	1	2	6	4	3	3	1

- Grafico a Bastoncini;
- Grafico Poligonale;
- Aerogramma:
  Se q_i sono le frequenze in percentuale, i settori circolari avranno ampiezza:
  $$\frac{q_i}{100} \times 360^\circ$$

Rappresentazione Grafica dei Dati

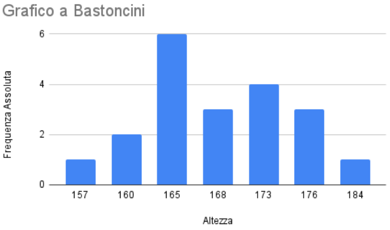
Consideriamo le diverse altezze degli studenti rilevate in centimetri:

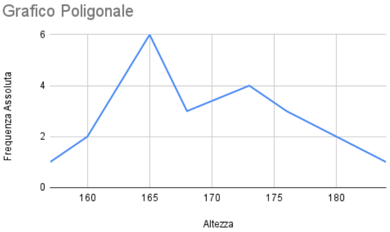
Altezza	157	160	165	173	168	176	184
Freq.Ass	1	2	6	4	3	3	1

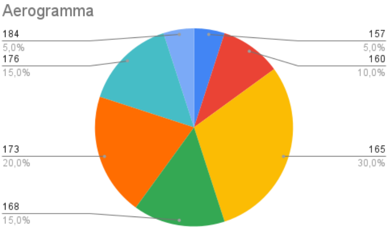
- Grafico a Bastoncini;
- Grafico Poligonale;
- Aerogramma:
  Se q_i sono le frequenze in percentuale, i settori circolari avranno ampiezza:
  $$\frac{q_i}{100} \times 360^\circ$$
- Istogramma:
  Classi modali con le rispettive frequenze.

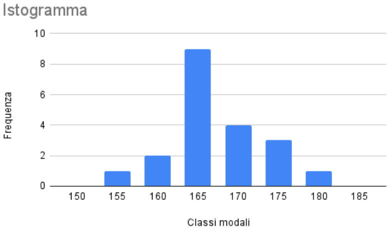
Rappresentazione Grafica dei Dati

▸ Link









Moda, Media e Mediana

Moda

Definiamo la moda di una variabile statistica come la modalità con frequenza più alta. Se le modalità sono raggruppate in classi, la classe modale è la classe con la frequenza più alta.

Moda, Media e Mediana

Moda

Definiamo la moda di una variabile statistica come la modalità con frequenza più alta. Se le modalità sono raggruppate in classi, la classe modale è la classe con la frequenza più alta.

Media Aritmetica

Sia $Y = (y_1, \cdots, y_N)$ una variabile statistica. Definiamo media aritmetica la quantità:

$$\bar{y} = \frac{1}{N}(y_1 + \cdots + y_N)$$

Moda, Media e Mediana

Moda

Definiamo la moda di una variabile statistica come la modalità con frequenza più alta. Se le modalità sono raggruppate in classi, la classe modale è la classe con la frequenza più alta.

Media Aritmetica

Sia $Y = (y_1, \cdots, y_N)$ una variabile statistica. Definiamo media aritmetica la quantità:

$$\bar{y} = \frac{1}{N}(y_1 + \cdots + y_N)$$

Media Geometrica

Sia $Y = (y_1, \cdots, y_N)$ una variabile statistica. Definiamo media geometrica la quantità:

$$\bar{y}_{geom} = \sqrt[N]{y_1 \cdot y_2 \cdot \ldots \cdot y_N}$$

Media Pesata

Data la variabile statistica $Y = (y_1, \cdots, y_N)$, definiamo media pesata (o ponderata) di Y la quantità:

$$\bar{y}_\omega = \frac{\omega_1 y_1 + \omega_2 y_2 + \ldots + \omega_N y_N}{\omega_1 + \omega_2 + \ldots + \omega_N}$$

dove i valori ω_i , $i \in [1, N]$ sono detti *pesi*.

Moda, Media e Mediana

Media Pesata

Data la variabile statistica $Y = (y_1, \cdots, y_N)$, definiamo media pesata (o ponderata) di Y la quantità:

$$\bar{y}_\omega = \frac{\omega_1 y_1 + \omega_2 y_2 + \ldots + \omega_N y_N}{\omega_1 + \omega_2 + \ldots + \omega_N}$$

dove i valori ω_i , $i \in [1, N]$ sono detti *pesi*.

Mediana

La mediana di una variabile statistica $Y = (y_1, \cdots, y_N)$, ove i valori siano ordinati in modo crescente, cioé $y_1 \leq \cdots \leq y_N$, è calcolata nel seguente modo:

- Se N è dispari, la mediana è pari al valore $y_{\frac{N+1}{2}}$;
- Se N è pari, la mediana è pari alla media aritmetica dei due valori $y_{\frac{N}{2}}$, $y_{\frac{N}{2}+1}$.

Esempio

Consideriamo il libretto accademico dello studente X che presenta 7 voti:

Voto	22	28	24	30	18	27	30
CFU	6	3	5	10	12	15	3

Calcolare la media aritmetica, la media geometrica e la media pesata.

Esempio

Consideriamo il libretto accademico dello studente X che presenta 7 voti:

Voto	22	28	24	30	18	27	30
CFU	6	3	5	10	12	15	3

Calcolare la media aritmetica, la media geometrica e la media pesata.
Si ha:

$$\bar{y} = \frac{1}{7}(22 + 28 + 24 + 30 + 18 + 27 + 30) = 25.6$$

$$\bar{y}_{geom} = \sqrt[7]{22 \cdot 28 \cdot 24 \cdot 30 \cdot 18 \cdot 27 \cdot 30} = 25.2$$

$$\bar{y}_{\omega} = \frac{6 \times 22 + 3 \times 28 + 5 \times 24 + 10 \times 30 + 12 \times 18 + 15 \times 27 + 3 \times 30}{6 + 3 + 5 + 10 + 12 + 15 + 3} =$$

$$= 24.9$$

Varianza e Scarto Quadratico Medio

La *varianza* esprime la "variabilità" dei dati che abbiamo a disposizione.

Varianza e Scarto Quadratico Medio

La *varianza* esprime la "variabilità" dei dati che abbiamo a disposizione.

Varianza

Sia $Y = (y_1, \cdots, y_N)$ una variabile statistica con media aritmetica $\bar{y}$. Si defnisce **varianza** di Y la media degli scarti quadratici, cioè delle quantità $(y_i - \bar{y})^2$, con i compreso tra 1 e N.

$$\sigma^2 = \frac{1}{N}\left[(y_1 - \bar{y})^2 + \cdots + (y_N - \bar{y})^2\right]$$

Varianza e Scarto Quadratico Medio

La *varianza* esprime la "variabilità" dei dati che abbiamo a disposizione.

Varianza

Sia $Y = (y_1, \cdots, y_N)$ una variabile statistica con media aritmetica $\bar{y}$. Si defnisce **varianza** di Y la media degli scarti quadratici, cioè delle quantità $(y_i - \bar{y})^2$, con i compreso tra 1 e N.

$$\sigma^2 = \frac{1}{N}\left[(y_1 - \bar{y})^2 + \cdots + (y_N - \bar{y})^2\right]$$

Deviazione Standard

Definiamo **deviazione standard** o **scarto quadratico medio**, la radice quadrata della varianza.

$$\sigma = \frac{1}{\sqrt{N}}\sqrt{(y_1 - \bar{y})^2 + \cdots + (y_N - \bar{y})^2}$$

Proprietà

NOTA: Dalla definizione della varianza si nota che più la varianza è piccola e più i dati sono concentrati "vicino" al valore medio.

Proprietà

NOTA: Dalla definizione della varianza si nota che più la varianza è piccola e più i dati sono concentrati "vicino" al valore medio.

Proposizione

La varianza di una variabile statistica è la differenza tra la media aritmetica dei quadrati delle modalità e il quadrato della media aritmetica delle modalità:

$$\sigma^2 = \overline{y^2} - \bar{y}^2$$

Vedi Esempio

Quartile e Distanza Interquartile

- I quartili sono tre valori che dividono i dati ordinati in quattro parti, ciascuna con un numero uguale di osservazioni. I quartili sono un tipo di quantile.

Quartile e Distanza Interquartile

- I quartili sono tre valori che dividono i dati ordinati in quattro parti, ciascuna con un numero uguale di osservazioni. I quartili sono un tipo di quantile.

Quartile e Distanza Interquartile

- I quartili sono tre valori che dividono i dati ordinati in quattro parti, ciascuna con un numero uguale di osservazioni. I quartili sono un tipo di quantile.

- La distanza interquartile è la differenza tra $Q3$ e $Q1$: $\Delta = Q3 - Q1$. In altri termini la distanza interquartile taglia via il 25% dei valori più bassi e il 25% di quelli più alti.

Quartile e Distanza Interquartile - Esempio

Supponiamo di condurre uno studio sullo sviluppo del linguaggio nei bambini da 1-6 anni. E ottenere la seguente tabella:

Anni	1	2	3	4	5	6
Frequenza	2	3	4	1	2	2

Quartile e Distanza Interquartile - Esempio

Supponiamo di condurre uno studio sullo sviluppo del linguaggio nei bambini da 1-6 anni. E ottenere la seguente tabella:

Anni	1	2	3	4	5	6
Frequenza	2	3	4	1	2	2

- Contiamo il numero di osservazioni:

$$N = 2 + 3 + 4 + 1 + 2 + 2 = 14$$

Quartile e Distanza Interquartile - Esempio

Supponiamo di condurre uno studio sullo sviluppo del linguaggio nei bambini da 1-6 anni. E ottenere la seguente tabella:

Anni	1	2	3	4	5	6
Frequenza	2	3	4	1	2	2

- Contiamo il numero di osservazioni:

$$N = 2 + 3 + 4 + 1 + 2 + 2 = 14$$

- Ordiniamo le osservazioni in ordine crescente:

$$[1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 5, 5, 6, 6]$$

Quartile e Distanza Interquartile - Esempio

Supponiamo di condurre uno studio sullo sviluppo del linguaggio nei bambini da 1-6 anni. E ottenere la seguente tabella:

Anni	1	2	3	4	5	6
Frequenza	2	3	4	1	2	2

- Contiamo il numero di osservazioni:

$$N = 2 + 3 + 4 + 1 + 2 + 2 = 14$$

- Ordiniamo le osservazioni in ordine crescente:

$$[1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 5, 5, 6, 6]$$

- Troviamo il primo quartile

$$N \times (\frac{1}{4}) = 14 \times (\frac{1}{4}) = 3.5$$

3.5 non è intero, pertanto la posizione del primo quartile è la quarta: $Q1 = 2$ anni.

Quartile e Distanza Interquartile - Esempio

Supponiamo di condurre uno studio sullo sviluppo del linguaggio nei bambini da 1-6 anni. E ottenere la seguente tabella

Anni	1	2	3	4	5	6
Frequenza	2	3	4	1	2	2

- Troviamo il secondo quartile (**mediana**)

$$N \times (\frac{2}{4}) = 14 \times (\frac{2}{4}) = 7$$

7 è un intero, pertanto $Q2$ è la media aritmetica tra la settima e l'ottava posizione: $Q2 = (3+3)/2 = 3$ anni.

Quartile e Distanza Interquartile - Esempio

Supponiamo di condurre uno studio sullo sviluppo del linguaggio nei bambini da 1-6 anni. E ottenere la seguente tabella

Anni	1	2	3	4	5	6
Frequenza	2	3	4	1	2	2

- Troviamo il secondo quartile (**mediana**)

$$N \times (\frac{2}{4}) = 14 \times (\frac{2}{4}) = 7$$

  7 è un intero, pertanto $Q2$ è la media aritmetica tra la settima e l'ottava posizione: $Q2 = (3+3)/2 = 3$ anni.

- Troviamo il terzo quartile

$$N \times (\frac{3}{4}) = 14 \times (\frac{3}{4}) = 10.5$$

  10.5 non è un intero, pertanto $Q3$ è il valore alla undicesima posizione: $Q3 = 5$ anni.

Distribuzione Normale o Gaussiana

Distribuzione Normale

Definiamo **distribuzione normale** la seguente funzione:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

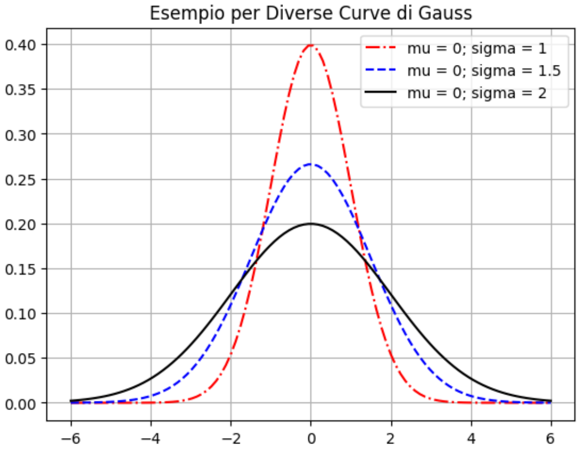
dove

- μ si dice *valore atteso* o *medio*;
- $\sigma^2 \neq 0$ è la *varianza*.

Il grafico della distribuzione normale prende il nome di *curva gaussiana*.

Curva Normale

Rappresentiamo la distribuzione di Gauss centrata nell'origine dell'asse x, ovvero con $\mu = 0$ per diversi valori di σ.



▸ Curve di Gauss

References

Bisi, Fioresi (2022)
Metodi matematici per le scienze applicate

Fine