

1

STATISTICA DESCRITTIVA

Temi considerati

- 1) Aspetti generali
- 2) Distribuzioni statistiche
- 3) Rappresentazioni grafiche
- 4) Misure di tendenza centrale
- 5) Medie ferme o basali
- 6) Medie lasche o di posizione
- 7) Dispersione o variabilità
- 8) Forma della distribuzione

Elementi forniti

- 1) Definizione generale
- 2) Esempi

2

Aspetti generali

Statistica Tecnica speciale per lo studio quantitativo dei fenomeni di massa o collettivi

Fenomeno di massa Fenomeno la cui misura richiede una massa o collezione di osservazioni di altri fenomeni più semplici detti fenomeni singoli o individuali

Unità statistica Risultato di una osservazione sopra uno dei fenomeni individuali da cui risulta il fenomeno collettivo

Dato statistico Risultato di un'operazione compiuta sopra le unità statistiche relative ai fenomeni individuali che rientrano in un fenomeno collettivo

3

Aspetti generali

Modalità Qualunque modo di manifestarsi di un fenomeno, espresso in termini o quantitativi o qualitativi

Variabile statistica Qualunque fenomeno che si manifesta o è capace di manifestarsi in almeno due modi diversi, espressi in termini quantitativi o qualitativi. A volte, quando il fenomeno è un carattere qualitativo si usa indicarlo come Mutabile statistica

Intensità Dato numerico che esprime una modalità quantitativa di un fenomeno

Frequenza Numero di unità statistiche portatrici di una stessa modalità del carattere o fenomeno allo studio

4

Aspetti generali

Popolazione o universo

Insieme delle unità statistiche portatrici di un dato carattere o fenomeno che si manifesta secondo due o più modalità differenti e che si vuole studiare o in relazione allo stesso carattere o in relazione ad un altro fenomeno

Campione

Porzione della popolazione o universo che interessa, estratta seguendo criteri prefissati e che viene effettivamente sottoposta a rilevazione e ad analisi

5

Contenuto della statistica moderna

Raccolta, presentazione ed elaborazione numerica delle informazioni, per agevolare l'analisi dei dati ed i processi decisionali

Statistica descrittiva

Insieme dei metodi che riguardano la raccolta, la presentazione e la sintesi di un insieme di dati per descriverne le caratteristiche essenziali

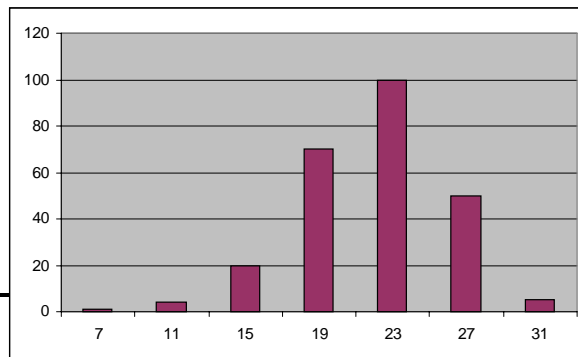
Statistica inferenziale

Insieme dei metodi con cui si possono elaborare i dati dei campioni allo scopo di dedurre omogeneità o differenze nelle caratteristiche analizzate, al fine di estendere le conclusioni alla popolazione

7

Consideriamo un esempio di distribuzione di frequenza. Supponiamo di avere un gruppo di persone che classifichiamo secondo la loro età, con il seguente risultato:

Limiti della classe (in anni)	Modalità o indice della classe : xi	Frequenza o numero dei casi della classe: fi
5 - 9	7	1
9 - 13	11	4
13 - 17	15	20
17 - 21	19	70
21 - 25	23	100
25 - 29	27	50
29 - 33	31	5
		<u>250</u>



Intervallo di classe: 4 anni

Numero delle classi: 7

Valore centrale di classe: semisomma estremi

6

Aspetti generali

Classe o intervallo

Raggruppamento di modalità ordinate di un dato carattere comprese tra un valore inferiore ed uno superiore chiamati *limiti di classe*

Distribuzione di frequenza

Raggruppamento di una serie di dati in classi, contando quanti valori o unità statistiche appartengono ad ogni gruppo o classe di modalità

8

Distribuzioni statistiche

Caratteristiche

1 - Tendenza centrale

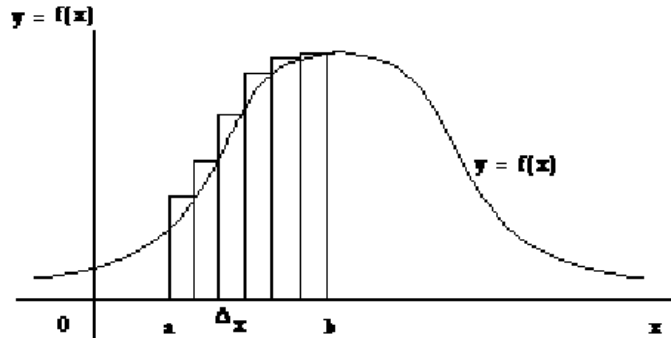
2 - Dispersione o variabilità

3 - Forma della distribuzione

Distribuzioni discontinue o discrete Sono quelle relative a variabili per loro natura discontinue, perché possono assumere solo valori discontinui (isolati), cioè numeri interi e non frazioni di numero. Il suo grafico può essere rappresentato da un istogramma

Distribuzioni continue Sono relative a variabili continue caratterizzate da valori compresi in certi intervalli di variazione al loro interno continui, cioè costituiti da infiniti valori. Il suo grafico può essere rappresentato da una curva continua

9
Distribuzioni statistiche continue e discrete



$$\int_a^b f(x).dx = P(a < x < b).$$

$$\int_{-\infty}^{+\infty} f(x).dx = 1.$$

$$\lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=n} f_i \cdot \Delta x = \int_a^b f(x).dx$$

10

Rappresentazioni grafiche

Scopo Evidenziare in modo semplice, *a colpo d'occhio*, le caratteristiche fondamentali di una distribuzione di frequenza, cioè: tendenza centrale, variabilità e forma

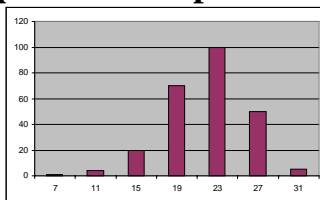
Inconvenienti Mancano di precisione, sono soggettive e permettono letture diverse degli stessi dati

11
Rappresentazioni grafiche

Tipi più frequenti per fenomeni discreti

- Istogramma

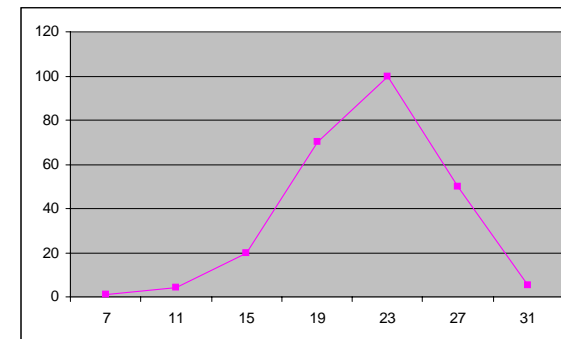
E' un grafico a barre verticali i cui rettangoli vengono costruiti in corrispondenza degli estremi di ciascuna classe. Le misure della variabile o fenomeno sono riportate lungo l'asse orizzontale, la frequenza o numero dei casi in cui si presenta ciascuna classe di misure lungo l'asse verticale. Le superfici dei vari rettangoli sono proporzionali alle frequenze corrispondenti



12

- Poligono di frequenza

Figura simile all'istogramma che è di solito utilizzata per rappresentare valori relativi o percentuali. Può essere ottenuto dall'istogramma corrispondente, unendo con una spezzata le frequenze relative ai punti centrali di ogni classe



Rappresentazioni grafiche

Tipi più usuali per fenomeni qualitativi

- *Diagramma a rettangoli distanziati*

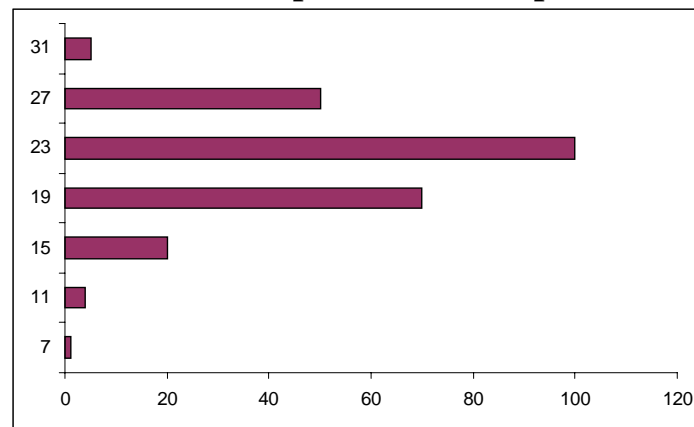
E' un grafico a colonne, formato da rettangoli con basi uguali ed altezze proporzionali alle frequenze dei vari gruppi

- *Diagramma a barre*

Al posto dei rettangoli utilizza linee continue più o meno spesse

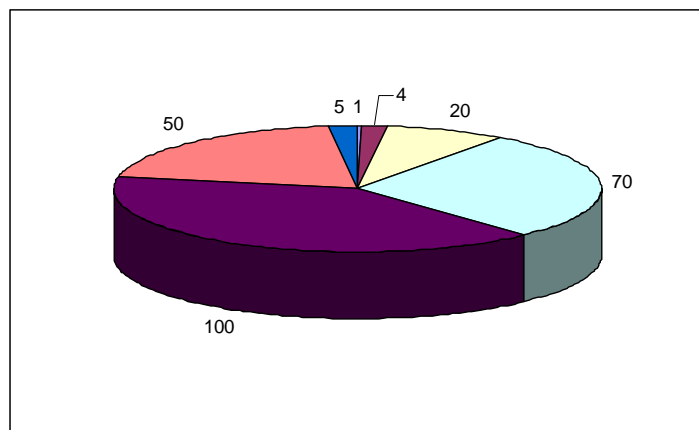
- *Ortogramma o grafico a nastri*

E' uguale ai rettangoli distanziati, ma gli assi sono scambiati per una lettura più facile



- *Diagramma circolare o a torta*

Si divide un cerchio in parti proporzionali alle frequenze o alla intensità di ciascuna modalità



- *Diagramma a figure*

La frequenza di ogni modalità viene rappresentata da una figura o da simboli che ricordano facilmente l'oggetto, la cui altezza è proporzionale alla frequenza o alla intensità

- *Cartogramma*

Evidenzia distribuzioni territoriali mediante carte geografiche in cui in alcune località sono riportati cerchi proporzionali alle frequenze o alle intensità del fenomeno

Valore centrale o media

E' una misura attraverso la quale si cerca di *sintetizzare* la variabile descritta dall'intera distribuzione di frequenza, attorno alla quale tendono a distribuirsi tutti gli altri valori della variabile compresi nell'intervallo o campo di definizione del fenomeno osservato e ne rappresenta l'ordine di grandezza

Condizioni generali a cui deve rispondere un valore centrale.

Esso deve:

- essere definito in maniera obiettiva
- dipendere da tutte le osservazioni della serie
- avere un significato concreto e facile da concepire
- essere semplice a calcolarsi
- essere poco sensibile alle fluttuazioni del campionamento
- prestarsi con facilità ai calcoli algebrici ulteriori

(Queste condizioni non sono sempre rispettate dalle varie medie)

Misure di tendenza centrale

Il valore medio può essere uno dei valori realmente osservati della serie oppure qualunque altro valore intermedio che non figura effettivamente tra le osservazioni.

I valori medi possono essere raggruppati in due insiemi principali: quello delle medie ferme o basali e quello delle medie lasche o di posizione.

Medie ferme o basali

Sono quelle suscettibili di essere determinate in funzione di tutti i termini della serie a cui si riferiscono

Media lasche o di posizione

Sono quelle il cui valore non dipende da tutti i termini della serie a cui si riferiscono, ma dalla posizione che esse occupano nella serie

Medie ferme o basali

Media aritmetica

La *media aritmetica* di una variabile statistica X è definita come la somma dei valori x_i assunti dalla variabile in tutti gli f_i casi rilevati, divisa per il numero totale degli stessi casi. In formule:

per la semplice:
$$M_1 = \frac{\sum_{i=1}^{i=N} X_i}{N},$$
 con $i = 1, 2, \dots, N.$

per la ponderata:
$$M_1 = \frac{\sum_{i=1}^{i=n} x_i f_i}{\sum_{i=1}^{i=n} f_i},$$
 con $\sum_{i=1}^{i=n} f_i = N.$

Misure di tendenza centrale

Proprietà della media aritmetica

- 1) La *somma algebrica* degli scostamenti dalla media aritmetica di una distribuzione è *uguale a zero*
- 2) La *somma dei quadrati* degli scostamenti dalla media aritmetica di una distribuzione è *un minimo*
- 3) La media aritmetica di una successione a_1, a_2, \dots, a_n , con pesi f_1, f_2, \dots, f_n è l'*ascissa del baricentro* di un sistema di forze di gravità di intensità uguale alle f_i , applicate nei punti di ascissa a_i di un asse orizzontale.
- 4) Si applica di preferenza a *grandezze additive*
- 5) E' particolarmente utile per il fatto che essa effettua la correzione degli errori accidentali di osservazione, per cui essa è la *stima più precisa di misure ripetute*.
- 6) E' la *più semplice* tra le medie algebriche.

Esempio III.2 Si abbia la distribuzione di una data popolazione di individui secondo la statura, descritta nella tavola e si faccia l'ipotesi che le intensità di ogni classe di statura siano concentrate nel valore centrale; per le classi estreme aperte, si ipotizza che il valore centrale sia rispettivamente di 145 e 185 cm. I calcoli per determinare la media aritmetica ponderata della distribuzione sono i seguenti.

Classi di statura (cm)	Valori centrali delle classi (x_j)	Frequenze (n_j)	Prodotti ($n_j x_j$)
meno di 150	145	1.500	217.500
150 --160	155	41.200	6.386.000
160 --170	165	188.300	31.069.500
170 --180	175	125.900	22.032.500
180 e oltre	185	<u>14.500</u>	<u>2.682.500</u>
		<u>371.400</u>	<u>62.388.000</u>

Pertanto la media aritmetica ponderata sarà:

$$\bar{x} = \frac{62.388.000}{371.400} = 167,98.$$