SINTESI DEL CAPITOLO α

**La statistica si occupa di raccogliere ed elaborare i dati** relativi ad un singolo individuo o ad un gruppo di individui o oggetti per studiarne le abitudini, trarne conclusioni e fare previsioni.

Un’**indagine statistica**, perciò, si articola in due fasi distinte:

* FASE 1 raccolta di informazioni (questionari, ricerche su internet….)
* FASE 2 elaborazione delle informazioni

L’insieme di individui o oggetti sui quali desideriamo fare l’indagine si dice **popolazione statistica.**

Se le informazioni raccolte riguardano l’intera popolazione allora si parla di **censimento,** se si raccolgono informazioni solo su un sottoinsieme significativo della popolazione, allora si parla di **indagine a campione**. La scelta del campione è molto importante al fine dell’affidabilità del risultato.

La **statistica induttiva** si occupa di studiare dei metodi perché il passaggio dal campione all’intera popolazione sia affidabile, la **statistica descrittiva** si occupa delle tecniche di rappresentazione dei risultati ottenuti tramite tabelle o grafici (noi studieremo solo queste tecniche).

Ogni elemento della popolazione si chiama **unità statistica**, ogni caratteristica che noi vogliamo studiare si chiama **carattere**, il carattere è espresso da una **modalità** che può essere descritta da un numero (**carattere quantitativo**) o da un attributo non numerico (**carattere qualitativo**).

**Esempio 1**

Voglio studiare l’età degli abitanti di un paese.

Popolazione= insieme degli abitanti del paese

Unità statistica= il singolo abitante

Carattere=età

Modalità=1,2,…..,15,…..70,…

Si tratta di un carattere quantitativo

**Esempio 2**

Voglio studiare il tipo di vegetazione nel monte del Monte Amiata

Popolazione= insieme degli alberi e delle piante della zona

Unità statistica= la singola pianta

Carattere=stato di conservazione

Modalità=buono, ottimo, malato,……

Si tratta di un carattere qualitativo

La **frequenza** è il numero di volte che una modalità si presenta

Nell’esempio 2 la frequenza delle piante amiatine con stato di conservazione buono è il numero di piante che hanno quello stato

Per descrivere un dato carattere della popolazione si costruiscono delle **tabelle di frequenza** in cui si specifica la frequenza per ogni modalità.

**Esempio 2, una possibile tabella di frequenza:**

|  |  |
| --- | --- |
| **Stato di conservazione** | **Frequenza** |
| buono | 150 |
| ottimo | 72 |
| malato | 80 |
| Su cui intervenire | 103 |

Si sono presi in considerazione 405 alberi, che rappresenteranno il campione dell’intera popolazione delle piante Amiatine, sono stati analizzati e sono stati contate le unità statistiche per ogni modalità.

La frequenza si indica con la lettera **F,** il numero totale di unità statistiche prese in considerazione si indica con la lettera **T,** si definisce **frequenza relativa il rapporto fra F e T**

**f=F/T**

**Esempio 2, una possibile tabella di frequenza, CON FREQUENZE RELATIVE**

|  |  |  |
| --- | --- | --- |
| **Stato di conservazione** | **Frequenza** | **Frequenza RELATIVA** |
| buono | 150 | 150/405=0,37=37% |
| ottimo | 72 | 72/105=0,18=18% |
| malato | 80 | 80/405=0,20=20% |
| Su cui intervenire | 103 | 103/405=0,25=25% |

Dividendo le frequenze delle singole modalità per 405 si ottengono le frequenze relative

Le frequenze relative sono numeri <1 e sono espressi in percentuale.

A volte è utile dividere le modalità in **classi,** per esempio se si vuole studiare l’altezza degli individui di una determinata popolazione si considerano degli intervalli di altezza, ad esempio: da 1,50 mt a 1,60, da 1,60 a 1,70 ecc.

L’estremo inferiore non è compreso nella classe e quello superiore si

Il **valore centrale** di ogni classe è dato dalla somma degli estremi divisa per 2.

Se le tabelle di frequenza sono relative ad un carattere quantitativo si dicono **serie**, può essere utile vedere la variazione del fenomeno negli anni e quindi si parla di **serie storiche**. Se i valori che un carattere può assumere sono ben definiti, numeri interi, la modalità si dice **discreta**, se invece i valori assunti sono numeri reali all’interno di un intervallo allora si definisce **continua**.

Le frequenze possono essere rappresentate con:

* la classica tabella delle frequenze
* Le tabelle a doppia entrata
* Dei grafici
	+ Ortogramma
	+ Istogramma, simile ad ortogramma, ma si usa quando le modalità sono divise in classi e ogni rettangolo ha una dimensione pari all’ampiezza della classe
	+ Areogramma o diagramma a torta
	+ Diagramma cartesiano
	+ Ideogramma o cartogramma

**Esempio 2 ortogramma**

Ottimo

Buono

A volte è utile riassumere una serie di dati attraverso un **valore medio, o indice di posizione centrale,** tale valore si può calcolare in modo diverso:

* media aritmetica: rapporto tra la somma di tutti i valori e il numero dei valori
* media ponderata:rapporto fra la somma di tutti i valori moltiplicati per il loro pesoe la somma di pesi
* mediana: si ordinano i valori in senso crescente, se il numero di valori è dispari si prende il valore centrale, altrimenti la media aritmetica dei valori centrali
* moda: il valore che ha frequenza maggiore. In alcune serie ci possono essere più mode, se le mode sono 2 la serie è detta bimodale

Esempi di esercizi:

Calcola la **media aritmetica** fra i seguenti numeri:

6; 14; 8; 23; 4

Media=(6+14+8+23+4)/5=55/5=11

Un ragazzo partecipa ad un concorso sostenendo 4 prove, ad ogni prova viene assegnato un punteggio e un peso diverso, calcola la **media ponderata**.

PROVA 1 punteggio 27 peso 2

PROVA 2 punteggio 22 peso 4

PROVA 3 punteggio 25 peso 1

PROVA 4 punteggio 16 peso 3

**Media ponderata**=(27x2+22x4+25x1+16x3)/(2+4+1+3)=(54+88+25+48)/10=215/10=21,5

Calcola la **mediana** fra le due serie di valori:

4; 1; 4; 1; 3; 8; 10

22; 28; 30; 40

Mediana prima serie

Ordino i valori in senso crescente: 1; 1; 3, 4; 4; 8; 10

I valori sono 7, quindi prendo il valore centrale ossia il quarto: 4

Mediana seconda serie

I valori sono già ordinati, sono quattro quindi prendo la media fra i due valori centrali: (28+30)/2=29

Determina la **moda** fra la seguente serie di valori: 3; 8; 2; 3; 5; 1; 7; 3; 5; 3; 15; 2; 10; 3; 12; 15;

Per vedere qual è il valore che ha frequenza massima potremmo ordinare la sequenza; ma ci accorgiamo ugualmente che il valore che compare più volte è il 3.

A volte due sequenze pur avendo la stessa media hanno valori singoli molto diversi, per questo è bene tenere in considerazione gli **indici di variabilità**.

Tra gli indici di variabilità abbiamo:

il **campo di variazione** che è la differenza fra il valore più grande e il valore più piccolo della serie

Lo **scarto semplice medio** che la media fra gli scarti di ogni valore rispetto alla media, per scarto rispetto alla media si intende la differenza, in valore assoluto, fra valore e media.

 La **deviazione standard** che è la radice quadrata della media aritmetica dei quadrati degli scarti rispetto alla media

Esercizi sugli indici di variabilità

Determina il **campo di variazione** della seguente serie:

10; 4; 1; 7; 9; 6

Si calcola la differenza fra il valore più grande e quello più piccolo: 10-1=9

Determina lo **scarto semplice** **medio** della seguente serie: 2; 4; 7; 12; 21

Prima si calcola la media Media=(2+4+7+12+21)/5=46/5=9,2

Poi gli scarti rispetto alla media: 9,2-2=7,2 9,2-4=5,2 9,2-7=2,2 12-9,2=2,8 21-9,2=11,8

Quindi si calcola la media degli scarti**: scarto semplice medio**=(7,2+5,2+2,2+2,8+11,8)/5=29,2/5=5,84

Calcola la **deviazione standard** per la seguente sequenza: 5; 7; 9; 11

Calcolo la media: media=(5+7+9+11)/4=32/4=8

Calcolo gli scarti: 8-3=5 8-7=1 9-8=1 11-8=3

Calcolo la media aritmetica dei quadrati degli scarti: (25+1+1+9)/4=36/4=9

La **deviazione standard** è la radice quadrata della media dei quadrati degli scarti, quindi 3

DISTRIBUZIONE GAUSSIANA

Consideriamo l’istogramma relativo alle frequenze e uniamo i punti medi dei lati superiori di tutti i rettangoli, otteniamo così il **poligono delle frequenze**.

Più risultati prendiamo e più il poligono delle frequenze si avvicina ad una curva teorica detta **curva di Gauss. La distribuzione delle frequenze intorno al valore medio è strettamente collegato alla deviazione standard.**

ERRORE STANDARD

Se consideriamo un campione di una determinata popolazione ed effettuiamo la media di un determinato carattere è chiaro che tale valore non sarà proprio uguale alla media effettiva calcolata su tutta la popolazione, quindi in questo calcolo commetteremo un errore che possiamo valutare attraverso un indice di variabilità chiamato **errore standard** che si calcola:

 **deviazione standard/radice quadrata della numerosità del campione-1**