

INTRODUZIONE

L'indagine statistica è lo strumento statistico mediante il quale si acquisiscono informazioni su uno o più fenomeni attinenti ad una popolazione. Può essere effettuata sull'intera popolazione di riferimento (es. censimento della popolazione) oppure su parte di essa, detta campione. L'oggetto dell'osservazione di ogni fenomeno individuale che costituisce il fenomeno collettivo è detto unità statistica. L'insieme delle unità statistiche costituisce il *collettivo o popolazione*.

La rilevazione statistica quindi, permette di determinare le modalità con cui ciascun carattere indagato (*variabile*) si presenta in ogni unità facente parte del collettivo.

I risultati del processo di indagine portano ad avere la determinazione della distribuzione del collettivo secondo i caratteri rilevati, ossia il modo in cui le unità del collettivo si distribuiscono secondo le varie modalità del o dei caratteri rilevati.

Se il carattere secondo cui è fatta la distribuzione è uno soltanto, allora la distribuzione è detta *semplice*, o *univariata*; se è fatta congiuntamente per due caratteri (ad esempio sesso e ore di studio) allora è *doppia* o *bivariata*; se è fatta congiuntamente da n caratteri allora è *multivariata*.

Le variabili sono classificabili in *qualitative* o *quantitative*.

Le variabili *qualitative* sono quelle le cui modalità assumono attributi non numerici di varia tipologia.

A loro volta si possono suddividere in caratteri con *scala nominale* e caratteri con *scala ordinale*:

Con scala nominale: le modalità non assumono alcun ordine preconstituito. L'unico confronto possibile tra due unità statistiche è quello di vedere se appartengono o meno alla stessa modalità (es. città di residenza)

Con scala ordinale: pur non facendo riferimento a valori numerici, assumono modalità che si distribuiscono logicamente in sequenza. Non è possibile effettuare operazioni aritmetiche tra le modalità, ma è possibile stabilire non solo la diversità ma anche una relazione d'ordine.

Le variabili *quantitative* hanno per modalità dei numeri che esprimono una misura o una quantità; per le modalità di questi caratteri dunque, sarà possibile effettuare varie operazioni tra essi: sommare, sottrarre ecc.

A loro volta i caratteri quantitativi si differenziano per il fatto di essere con *scala di intervallo* o con *scala di rapporto*:

Con scala di intervallo: consentono un confronto solo per differenza tra le modalità assunte dal carattere sulle varie unità statistiche. Essi infatti fanno riferimento a un'origine arbitraria, frutto di una convenzione in quanto lo zero delle scale intervallari non è uno zero assoluto ma convenzionale (es. la temperatura, può essere misurata ed espressa in gradi Celsius o gradi Fahrenheit).

Con scala di rapporto: in questo caso è possibile anche un confronto in rapporto

tra le modalità in quanto la definizione dello zero non è frutto di arbitrarie convenzioni; infatti si tratta di uno zero assoluto, qualsiasi unità di misura si utilizzi (es. Età anagrafica).

Tablelle

Tablelle semplici

Una tabella semplice è costituita da due righe (o colonne): la prima riporta le modalità della variabile in esame, la seconda le frequenze. A volte, per maggiore comprensione, si aggiungono alcune righe (o colonne) per inserire diverse tipologie di frequenza. Se la variabile in esame è quantitativa i valori si raggruppano in classi.

Le distribuzioni statistiche nelle quali i dati sono frequenze sono dette distribuzioni di frequenza.

La frequenza può essere:

Frequenza assoluta: il numero delle unità statistiche aventi una data modalità.

Frequenza relativa: il rapporto fra una frequenza assoluta e la somma di tutte le frequenze.

Frequenza percentuale: il rapporto espresso nella frequenza relativa viene convertito in percentuale (x100).

Frequenza cumulata: è di due tipi, relativa o assoluta. Si somma la rispettiva frequenza relativa, o assoluta, con le frequenze dei valori precedenti.

Si propone la popolazione italiana con età 15 anni o più suddivisa per titolo di studio.

	F. Assoluta	F. Relativa	F. Perc.	F. Cum. Ass.	F. Cum. %
<i>Licenza elementare, nessun titolo</i>	11.004	0,211	21,1%	11.004	21,1%
<i>Licenza media</i>	16.498	0,317	31,7%	27.502	52,8%
<i>Diploma 2-3 anni (qualifica professionale)</i>	3.083	0,059	5,9%	30.585	58,7%
<i>Diploma 4-5 anni (maturità)</i>	15.243	0,293	29,3%	45.828	88,0%
<i>Laurea e post-laurea</i>	6.261	0,120	12,0%	52.089	100,0%
Totale	52.089	1,000	100,0%		

N.B. Le frequenze sono in questo ordine da sinistra a destra: assolute, relative, percentuali, cumulate assolute, cumulate percentuali. I dati sono stati riportati in migliaia.

Tablelle a doppia entrata

Una tabella a doppia entrata è costituita da due variabili e permette il confronto fra di esse, ponendone una in ascissa e l'altra in ordinata. Si applica nell'analisi delle due variabili per individuare le relazioni esistenti fra di esse.

In una scuola media, per l'anno scolastico 2012-2013, si sono iscritti 200 studenti. Vediamo come sono ripartiti per sesso e classe:

	Maschio	Femmina	Totale (per classe)
<i>Prima</i>	<u>40</u>	<u>25</u>	65
<i>Seconda</i>	<u>26</u>	<u>34</u>	60
<i>Terza</i>	<u>39</u>	<u>36</u>	75
<i>Totale (per sesso)</i>	105	95	200

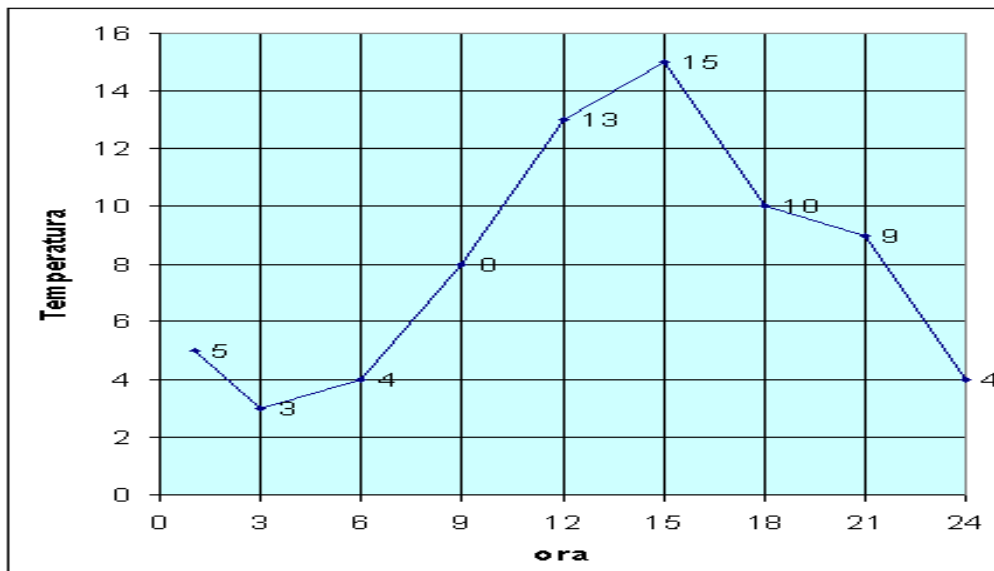
GRAFICI

I grafici vengono utilizzati come tecnica di presentazione dei dati che affianca la forma tabellare di questi, con lo scopo di semplificare il messaggio verso gli utenti, poiché la lettura delle tabelle può essere più complessa.

I più semplici, ma anche i più usati, sono: il diagramma cartesiano, il diagramma a settori circolari, l'istogramma e il diagramma a barre

Diagramma cartesiano

I diagramma cartesiani utilizzano semplici linee che descrivono l'andamento di un fenomeno. Questo tipo di rappresentazione, basato su di un piano cartesiano, utilizza dei punti, posti in corrispondenza di coppie ordinate di valori, uniti tra di loro da una linea spezzata che facilita la comprensione dell'andamento.

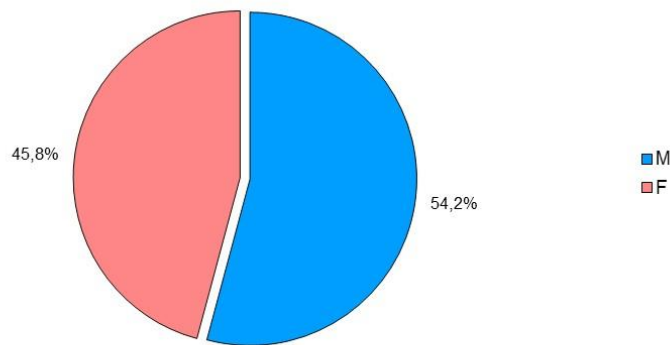


Il grafico illustra le temperature registrate ogni 3 ore in una giornata di febbraio nella città Z. Nell'asse delle ascisse si trova l'orario e nell'asse delle ordinate la temperatura. I punti rappresentano le coordinate (x,y). Per esempio il punto (18,10) indica che alle ore 18 c'erano 10 gradi.

Diagramma a settori circolari

Detto anche "diagramma a torta", ha l'aspetto di un cerchio suddiviso in settori di diversa ampiezza. Il cerchio rappresenta il totale delle unità statistiche facenti riferimento alla variabile studiata. A ogni settore corrisponde una delle diverse modalità con cui si presenta il carattere indagato. L'ampiezza dei singoli settori è proporzionale alla frequenza (assoluta o relativa) della modalità da esso rappresentata. Questo tipo grafico è utile nel caso in cui le modalità siano due o tre.

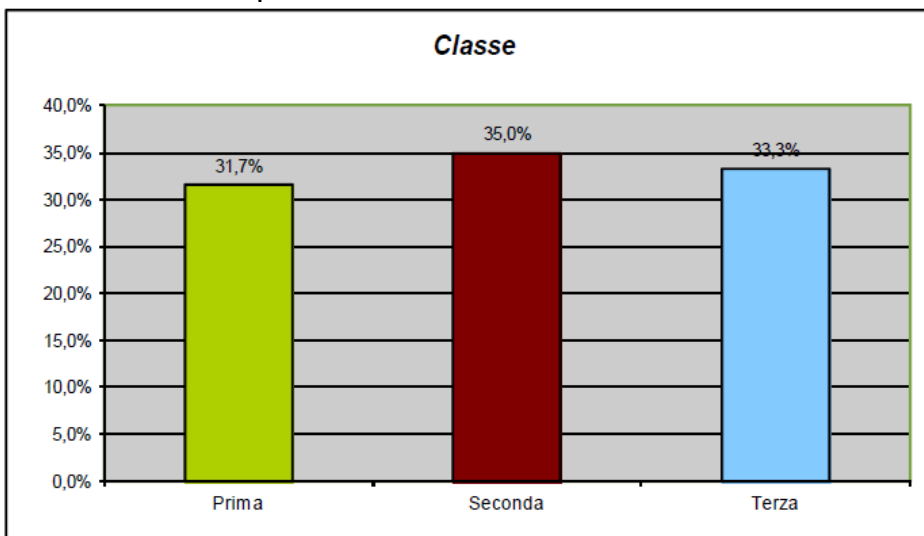
Sesso



In una classe di 24 studenti ci sono 13 maschi (54,2%) e 11 femmine (45,8%).

Istogramma

Il presupposto di un istogramma è un insieme di dati numerici e non, classificati secondo l'appartenenza alle diverse modalità. L'istogramma è una figura geometrica piana formata da rettangoli accostati aventi come altezza la densità di frequenza.

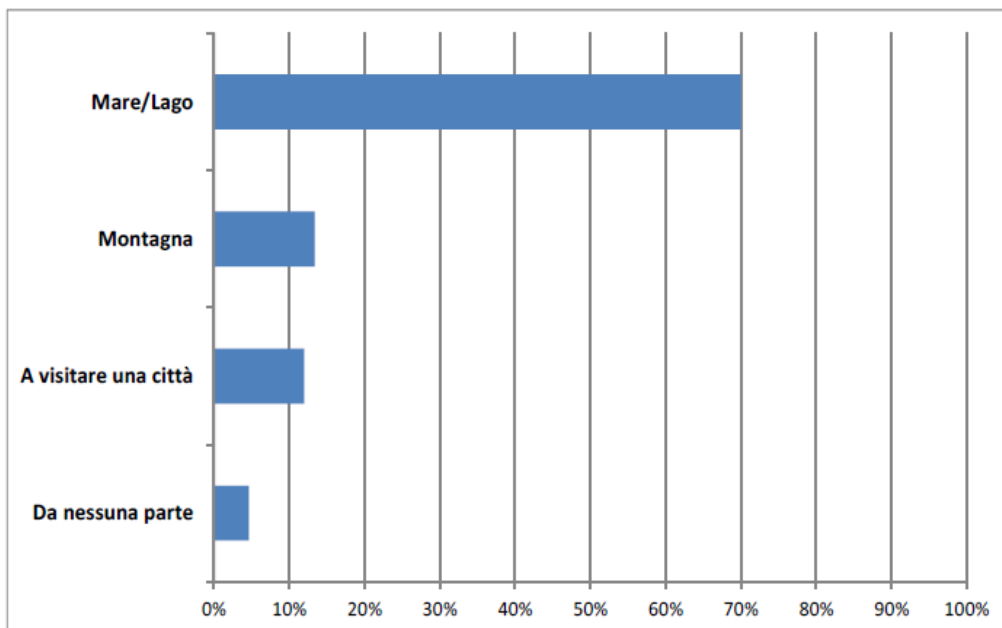


In una scuola secondaria di primo grado nell'anno scolastico 2012/2013 si sono iscritti 120 studenti: 38 (31,7%) in prima, 42 (35,0%) in seconda, 40 (33,3%) in terza.

Diagramma a barre

Molto simile all'istogramma, tranne per il fatto che le barre sono poste in orizzontale; di conseguenza le modalità vengono indicate nell'asse delle ordinate e la scala di misura in quello delle ascisse. Si preferisce questa tipologia di grafico quando le variabili sono qualitative con scala nominale.

Esempio: Dove preferiresti andare in vacanza questa estate?



In un'indagine statistica che ha coinvolto circa 5000 persone di diverse età è stato chiesto loro dove li piacerebbe andare in vacanza quest'estate. Le opzioni di risposta erano quattro. Il 70% di loro ha risposto che il mare o il lago sono le mete preferite. Sia gli amanti della montagna e sia quelli che preferiscono visitare una città sono circa il 12%, mentre c'è un 5% che predilige starsene a casa questa estate.

INDICI DI POSIZIONE

Media aritmetica

La media aritmetica (o semplicemente media) di un insieme di dati numerici è la somma dei dati divisa per il numero di modalità.

Nel caso in cui le modalità siano suddivise in classi, si prende sempre come riferimento il valore medio della classe.

Esempio

In una classe composta da 24 studenti vengono misurate le altezze. Queste vengono suddivise in classi di altezza come segue:

Altezza (in cm)	140-150	150-160	160-170	170-180	180-190
N° Studenti	2	7	9	4	2

Il calcolo della media sarà

$$\text{Media} = (145 \cdot 2 + 155 \cdot 7 + 165 \cdot 9 + 175 \cdot 4 + 185 \cdot 2) / 24 = 163,75$$

Moda

È la modalità che presenta la frequenza, assoluta o relativa, più elevata tra tutte le modalità assunte dalle unità statistiche nel collettivo. Tale indicatore può essere utilizzato per qualsiasi tipologia di scala (comprese quelle qualitative).

In caso di più modalità aventi la frequenza massima, la distribuzione si chiamerà *plurimodale*.

L'uso della moda è rilevante nel caso in cui essa sia unica e con una frequenza ben differenziata rispetto a quelle delle altre modalità.

Esempio

In una classe di 20 studenti viene assegnato dall'insegnante un compito scritto di matematica. I voti sono stati i seguenti:

Voto	3	4	5	6	7	8
N° Studenti	1	2	7	5	4	1

La variabile di riferimento è "Voto" e le modalità sono i vari voti presi. La

moda è in questo caso la modalità "5" con valore 7.

Mediana

È il valore che divide la distribuzione in due parti uguali (il numero centrale di

un insieme di dati numerici posti in ordine crescente). E' intuibile come la mediana sia un indicatore utilizzato per caratteri quantitativi e qualitativi ordinabili.

Se la numerosità del collettivo è un numero *dispari*, la mediana è la modalità relativa all'unità statistica che nell'ordinamento occupa la posizione di mezzo.

Se il numero è *pari*, come valore centrale si prende quello che si ottiene addizionando i valori delle due unità centrali e dividendo la loro somma per due. La mediana è un indice robusto in quanto non è influenzato dai valori

outliers (valori molto grandi o molto piccoli rispetto agli altri). Esempio 1

In un'intervista riservata viene richiesto lo stipendio netto percepito (in euro) a 5 persone (numero dispari): 850 1.200 1.400 1.750 5.300

La **mediana** è rappresentata dallo stipendio che occupa il 3° posto, cioè 1400.

N.B. Da notare come essa non venga influenzata dal valore 5.300, molto più alto degli altri quattro.

Esempio 2

Si riprende l'esempio dei voti del compito di matematica:

Voto	3	4	5	6	7	8
N° Studenti	1	2	7	5	4	1

Il numero di studenti complessivo è 20 (numero pari); dunque la mediana sarà il valore intermedio tra il voto del 10° ed 11° studente.

$$\text{Mediana} = (5 + 6) / 2 = 5,5$$