

STATISTICA

INTRODUZIONE

La **statistica** si occupa di studiare qualsiasi tipo di fenomeno eseguendo delle misurazioni.

Questi fenomeni possono riguardare collettivi o popolazioni che sono formati da singoli elementi distinguibili che prendono il nome di **unità statistiche**.

L'insieme di tutte le unità statistiche formano la **POPOLAZIONE** la quale ha caratteristiche comuni.

Possiamo distinguere la popolazione in: FINITA, INFINITA E INDETERMINATA.

- ✚ La popolazione si definisce *finita* quando può essere numerata come per esempio gli abitanti di Roma
- ✚ La popolazione si definisce *infinita* quando è impossibile numerarla come per esempio quante sono le stelle dell'universo
- ✚ La popolazione si definisce *indeterminata* se le unità si possono contare ma il conteggio è impossibile come per esempio un censimento dei clochard nella regione Lazio.

Lo studio che si può fare su una popolazione può riguardare la popolazione intera o può riguardare solo una parte di essa (prendo un CAMPIONE rappresentativo della popolazione) andando poi ad estendere i risultati che ottengo alla popolazione intera. (Esempio: se voglio studiare le abitudini dei cittadini di Roma non vado a intervistare tutte gli abitanti della capitale ma vado a scegliere un "campione" rappresentativo della popolazione)

La statistica quindi può essere **DESCRITTIVA** se lo studio viene fatto su tutta la popolazione altrimenti si definisce **INFERENZIALE** se invece lo studio viene condotto solo su un campione della popolazione. Quando si studia una popolazione devo definire che cosa voglio studiare e l'oggetto di questo studio prende il nome di **CARATTERE**.

ESEMPIO: Voglio studiare la nazionalità dei calciatori della scuola calcio nati nel 1993 della polisportiva comunale di Ariccia. Il CARATTERE è la nazionalità, UNITÀ STATISTICA sono i calciatori mentre la POPOLAZIONE sono i calciatori della polisportiva comunale di Ariccia.

LE 4 FASI PER LO STUDIO STATISTICO

Come abbiamo detto all'inizio la statistica ha lo scopo di studiare i fenomeni.

Lo studio si articola in 4 fasi:

1. Rilevazione dei dati
2. Elaborazione e interpretazione dei dati
3. Sintesi dei dati
4. Presentazione dei dati

Esaminiamoli uno per uno

1. RILEVAZIONE DEI DATI

Per potere iniziare ad effettuare uno studio ho bisogno di reperire dati. I dati statistici possono essere **raccolti direttamente** attraverso questionari, interviste o **raccolti indirettamente** come per esempio accedendo a database privati.

I dati (le informazioni) possono essere di tipo continuo, periodico o occasionale.

Le informazioni di tipo continuo sono i dati che vengono raccolti in maniera ininterrotta (esempio: le nascite registrate all'ufficio anagrafe)

Quelle di tipo periodiche sono raccolte periodicamente come succede per il censimento in Italia che avviene ogni 10 anni

Mentre quelle occasionali sono quelle che vengono raccolte occasionalmente

2. ELABORAZIONE E INTERPRETAZIONE DEI DATI

Una volta raccolti i dati ho bisogno di elaborarli e interpretarli. Prendiamo per esempio la tabella sottostante

Intervista ai 15 dipendenti del supermercato TOCINS di Ariccia					
	Nome	Sesso	Sport	Età	Peso
1	Angelo	M	Arrampicata	38	82
2	Francesca	F	Pallavolo	37	61
3	Gianni	M	Calcio	47	75
4	Michele	M	Tennis	58	71
...
15	Grazia	F	Pallavolo	62	55

Nella tabella ogni riga rappresenta un'unità statistica mentre ogni colonna rappresenta un carattere della popolazione che si è deciso di analizzare.

Diamo adesso la definizione di **FREQUENZA ASSOLUTA**: la frequenza che indica il numero delle volte in cui una certa modalità del carattere considerato si manifesta prende il nome di frequenza assoluta.

Molto spesso però avremo a che fare con la **FREQUENZA RELATIVA** f_r che indica quanto incide ogni modalità sul totale della popolazione.

La frequenza relativa si ottiene dividendo la frequenza assoluta f_a corrispondente a ogni modalità per il totale della popolazione.

$$f_r = \frac{\text{numero di volte che si osserva la modalità}}{\text{numero componenti della popolazione}}$$

Quindi guardando la tabella sopra se voglio sapere la frequenza relativa allo sport pallavolo avrò:

$$f_r = \frac{2 \text{ (persone che giocano a pallavolo)}}{15 \text{ (numero di componenti della popolazione, in questo caso le persone analizzate sono 15)}} = 0,13$$

Spesso si preferisce trasformare la frequenza relativa in **FREQUENZA PERCENTUALE** f_p che si ottiene moltiplicando la frequenza relativa per 100

$$f_p = \frac{\text{numero di volte che si osserva la modalità}}{\text{numero componenti della popolazione}} * 100$$

Quindi: $f_p = \frac{2 \text{ (pallavolo)}}{15} * 100 = 13\%$

N.B. Le frequenze relative sono sempre minori di uno e la somma di tutte le frequenze relative è uguale a 1.

Le frequenze percentuali sono sempre minori di 100 e la somma di tutte le frequenze percentuali è uguale a 100.

3. LA SINTESI DEI DATI

Se nell'analisi statistica lo studioso avesse sempre a disposizione una tabella come sopra riportata che descriva solo 15 persone appartenenti ad un supermercato allora le cose sarebbero sempre molto semplici. Nella realtà quando si esegue uno studio statistico abbiamo a disposizione tantissimi dati e quindi abbiamo bisogno di una sintesi che ci permetta di riassumerli e di confrontarli.

Per poter fare tutto ciò dobbiamo introdurre degli **INDICI** che sono: Media, Moda, Mediana, Campo di variazione e Scarto

- **MEDIA ARITMETICA SEMPLICE**

Dati n valori x_1, x_2, \dots, x_n la media aritmetica si ottiene dividendo la somma dei dati per il loro numero.

Ragazzi la media aritmetica non è nient'altro che la media che fate dei vostri voti.

Se ipotizziamo che Francesco a matematica prende i seguenti voti: 6, 8, 5, 5 la sua media aritmetica semplice sarà:

$$M = \frac{6 + 8 + 5 + 5}{4} = \frac{24}{4} = 6$$

- **MEDIA ARITMETICA PONDERATA**

Dati n valori x_1, x_2, \dots, x_n e le corrispondenti frequenze f_1, f_2, \dots, f_n la media aritmetica ponderata è il valore che si ottiene sommando i prodotti di ciascun dato x_i per il numero di volte f_i in cui si rappresenta e dividendo per il totale ottenuto per la somma delle frequenze.

Facciamo un esempio. Nella tabella sottostante vengono riportate le 30 verifiche di matematica fatte dai ragazzi dei quarti anni del CFP Padri Somaschi di Ariccia.

Voto	Numero studenti
6	18
7	6
8	3
9	2
10	1
	30

La media aritmetica ponderata sarà:

$$M = \frac{6 * 18 + 7 * 6 + 8 * 3 + 9 * 2 + 10 * 1}{30} = \frac{108 + 42 + 24 + 18 + 10}{30} = 6,73$$

- **MEDIANA O VALORE CENTRALE**

Consideriamo n valori x_1, x_2, \dots, x_n disposti in ordine crescente o decrescente.

- Se i dati sono un numero pari, la mediana è pari alla semisomma dei due valori centrali che occupano i posti $(n/2)$ e $[(n/2)+1]$.
- Se i dati sono un numero dispari la mediana è il valore che occupa il posto centrale cioè il posto $(n/2)$.

N.B. la mediana dipende solo dal dato centrale mentre non tiene conto né dei valori più grande né dei valori più piccoli presenti nella distribuzione dei dati

Esempio:

Vogliamo calcolare la mediana dei turisti presenti a Roma negli ultimi 5 anni

Anno	Turisti
2015	10.000
2016	15.000
2017	12.000
2018	7.000
2019	11.000
2020	14.000

Avendo la seguente distribuzione dei dati per prima cosa li ordino in maniera crescente o decrescente

7.000 – 10.000 – 11.000 – 12.000 – 14.000 – 15.000

Dato che poi ho 6 dati e quindi un numero pari per il calcolo della mediana devo prendere i dati che occupano la terza e la quarta posizione e li divido per due

$$\text{Mediana} = \frac{11.000 + 12.000}{2} = 11.500$$

- **MODA O VALORE CENTRALE**

La moda è il valore che si presenta con la massima frequenza oppure la modalità a cui è associata la frequenza maggiore.

Una stessa distribuzione può avere più mode.

Se riprendiamo la tabella delle 30 verifiche di matematica la moda in questo caso è 6 in quanto la sua frequenza è 18 la più alta di tutte.

Voto	Numero studenti
6	18
7	6
8	3
9	2
10	1
	30

Quindi abbiamo:

➤ MODA = 6 (è il dato con la maggiore frequenza)

➤ MEDIANA = 6 (semisomma del 15° e 16° dato)

6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, **6, 6**, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9, 10

➤ MEDIA PONDERATA = $M = \frac{6 \cdot 18 + 7 \cdot 6 + 8 \cdot 3 + 9 \cdot 2 + 10 \cdot 1}{30} = \frac{108 + 42 + 24 + 18 + 10}{30} = 6,73$

SI può affermare che moda e mediana sono indici poco influenzati dai dati estremi al contrario della media

- **CAMPO DI VARIAZIONE O RANGE**

Questo indice si ottiene dalla differenza tra i valori estremi (il valore più alto e il valore più basso) dei dati considerati.

ESEMPIO

Variazioni del mercato automobilistico in Italia: +1,5 FIAT, -1,2 ALFA ROMEO, +0,7 CITROEN, -0,5 AUDI, +0,1 MERCEDES;

il campo di variazione è: $1,5 - (-1,2) = + 2,7$

- **SCARTO SEMPLICE MEDIO**

A differenza del range che tiene conto solo dei dati estremi della distribuzione lo scarto semplice medio tiene conto di tutti i dati della distribuzione, in particolare dà un'idea di quanto la media rappresenti con affidabile attendibilità tutti i dati, più lo scarto semplice medio è piccolo e più la media è un buon indicatore di sintesi dei valori.

Lo scarto semplice medio si calcola andando a calcolare per prima cosa la media semplice dei dati e successivamente si procede calcolando ogni scarto assoluto (differenza, con segno positivo, tra ogni singolo valore e la media) e si conclude calcolando la media degli scarti assoluti.

Ritornando all'esempio precedente della variazione del mercato automobilistico in Italia abbiamo:

Ambito variazione automobili	Valore
FIAT	+ 1,5
ALFA ROMEO	- 1,2
CITROEN	+ 0,7
AUDI	- 0,5
MERCEDES	+ 0,1

Per prima cosa calcolo la media semplice

$$M = \frac{1,5 + (-1,2) + 0,7 + (-0,5) + 0,1}{5} = 0,12$$

Poi calcolo lo scarto

Ambito variazione automobili	Valore	SCARTO
FIAT	+ 1,5	1,5 - 0,12 = 1,3
ALFA ROMEO	- 1,2	-1,2 - 0,12 = 1,08
CITROEN	+ 0,7	0,7 - 0,12 = 0,58
AUDI	- 0,5	-0,5 - 0,12 = 0,62
MERCEDES	+ 0,1	0,1 - 0,12 = -0,02

Poi vado a prendere solo il valore assoluto ossia i valori negativi li faccio diventare positivi

Ambito variazione automobili	Valore	SCARTO	SCARTO ASSOLUTO
FIAT	+ 1,5	1,5 - 0,12 = 1,3	1,3
ALFA ROMEO	- 1,2	-1,2 - 0,12 = 1,08	1,08
CITROEN	+ 0,7	0,7 - 0,12 = 0,58	0,58
AUDI	- 0,5	-0,5 - 0,12 = 0,62	0,62
MERCEDES	+ 0,1	0,1 - 0,12 = -0,02	0,02

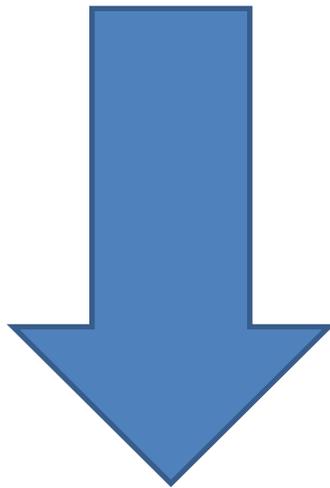
Calcolo poi la media degli scarti assoluti e si ottiene lo scarto semplice medio

$$\text{Scarto semplice medio: } \frac{1,3 + 1,08 + 0,58 + 0,62 + 0,02}{5} = \frac{3,6}{5} = 0,72$$

4. LA RAPPRESENTAZIONE DEI DATI

Una volta quindi ottenuti i dati bisogna rappresentarli e possono essere utilizzati:

- Istogrammi
- Diagrammi a torta
- Grafici a barre
- Diagrammi cartesiani
- Cartogrammi



ESERCIZI

Cari ragazzi, la pagine sopra riportate per alcuni di voi saranno risultate molto ostiche ma con un buon esercizio la varie e tante nozioni che vi ho scritto sembreranno molto più semplici.

Vi riporto adesso una serie di esercizi.

Nel caso in cui ci siano difficoltà o vogliate farmi vedere i vostri passaggi mandatemi tutto tramite mail all'indirizzo e vi risponderò il prima possibile. Tutti voi al fine di una buona preparazione all'esame deve mandarmi l'esercitazione-

a.andreotti@padrisomaschi.it

ES.1

La seguente tabella riporta il numero di libri letti in un anno dagli studenti di una classe

NUMERO DI LIBRI	FREQUENZA ASSOLUTA	FREQUENZA RELATIVA	FREQUENZA PERCENTUALE
1	10		
2	12		
3	6		
4	3		

Quanti sono gli studenti?

Completa inoltre le due colonne relative alle frequenze relative e percentuali

ES. 2

Vero o falso

- La mediana è il valore che in una distribuzione di n dati occupa il valore centrale?
- La moda può coincidere o meno con uno dei dati relativi?
- La media aritmetica semplice si ottiene addizionando tutti i dati e dividendo la somma per il numero dei dati stessi?
- La moda e la mediana, a differenza della media aritmetica, sono indici poco influenzati dai valori più grandi e dai valori più piccoli presenti nella rilevazione?

ES. 3

Completa la seguente tabella che rappresenta le età dei partecipanti a uno spettacolo teatrale. Le diverse età sono raggruppate in classi

Età DEI PARTECIPANTI (IN ANNI)	Età MEDIA DEI PARTECIPANTI PER OGNI INTERVALLO	NUMERO DEI PARTECIPANTI
6-15		20
16-24		60
25-34		85
35-55		72
56-70		42
71-85		18

Calcolare inoltre l'età media dei partecipanti

ES. 4

Calcola media, moda e mediana nei seguenti due casi

A. Numero di televisori per ogni abitazione di un condominio

TELEVISORI PER FAMIGLIA	FREQUENZA
1	5
2	8
3	2

B. Numero di ore settimanali dedicate alle attività sportive di un gruppo di 24 ragazzi

ORE SETTIMANALI	FREQUENZA
1	4
2	6
3	5
4	9

ES. 5

Alfredo ha riportato i seguenti voti a geografia. Calcolare media, moda e mediana

5,5 – 6 – 7 – 5 – 8 – 7 – 6,5

ES. 6

Angelo ha riportato i seguenti voti universitari nella materia scienza delle costruzioni

27 – 28 – 30 – 25 – 23

Che voto deve cercare di prendere al prossimo esame per avere una media del 27?

ES. 7

In una settimana sono state rilevate le seguenti temperature. Calcolare media, moda e mediana

GIORNO	TEMPERATURA (°C)
Lunedì	27
Martedì	26
Mercoledì	25
Giovedì	27
Venerdì	25
Sabato	27
Domenica	29