

STATISTICA DESCRITTIVA

DEFINIZIONI

- UN' UNITA' STATISTICA è la minima unità delle quale si raccolgono i dati. es. PERSONA, CORSO DI STUDI, UN RICCIO DI MARE.

- **POPOLAZIONE**

L'insieme di tutte le unità statistiche

NB. NON SOLO DI QUELLE
DI CUI HO I DATI

- **CAMPIONE**

L'insieme delle unità statistiche delle quali raccolgo i dati.

- **CARATTERI o VARIABILI**

Le caratteristiche di ogni unità statistica

PERSONA
↓
ETÀ, PESO

CORSO DI STUDI
↓
NUMERO DI ISCRITTI,
VOTO MEDIO DI LAUREA

RICCI DI MARE
↓
DIAMETRO, PESO, ETÀ

PERCENTUALE DI
OCCUPATI A UN ANNO
DALLA LAUREA

VARABILI

QUALITATIVE

"CATEGORIE"

CARATTERI CHE SI POSSONO AVERE O NON AVERE

ORDINATE

NON ORDINATE

• TITOLO DI STUDIO

0 NIENTE

1 ELEMENTARE

2 MEDIE

3 DIPLOMA

4 LAUREA T

5 LAUREA M

6 DOTTORATO

• COLORE DEGLI OCCHI

• REGIONE DI NASCITA

• GRUPPO SANGUIGNO

QUANTITATIVE

SI MISURANO NUMERICAMENTE

DISCRETE $\in \mathbb{Z}$
SOLO VALORI INTERI

• VOTO DI LAUREA

• NUMERO DI FIGLI

CONTINUE $\in \mathbb{R}$

ALTEZZA,

PESO,
ETA'

SIAMO U_1, U_2, \dots, U_N unite statistiche

$N \in \mathbb{N}$

↑
numero di elementi della popolazione

ESEMPIO (1) "LANCIO DI DUE DADI"

U_i "LANCIO"

VARIABILE "NUMERO USCITO"

QUANTITATIVA DISCRETA

NON È UNA CATEGORIA ORDINATA PERCHÉ HA SENSO

IL DOPPIO DI UN NUMERO (esempio 10 è il doppio di 5)

MENTRE NELL'ESEMPIO DEL TITOLO DI STUDIO 4 NON È IL DOPPIO DI 2, MA SOLO UN LIVELLO SUPERIORE (LAUREA T) (MEDIE)

$$N = 250$$

INSIEME DI VARIABILI

$$X = (6, 6, 5, 6, 3, 4, 8, \dots, 10, 8)$$

PER DESCRIVERE QUESTO TIPO DI DATI CALCOLO LA

DISTRIBUZIONE DI FREQUENZE

L'insieme di coppie ordinate (modalità, frequenze assolute)
di occorrenze

NELL'ESEMPLO (1) LE POSSIBILI MODALITÀ SONO 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

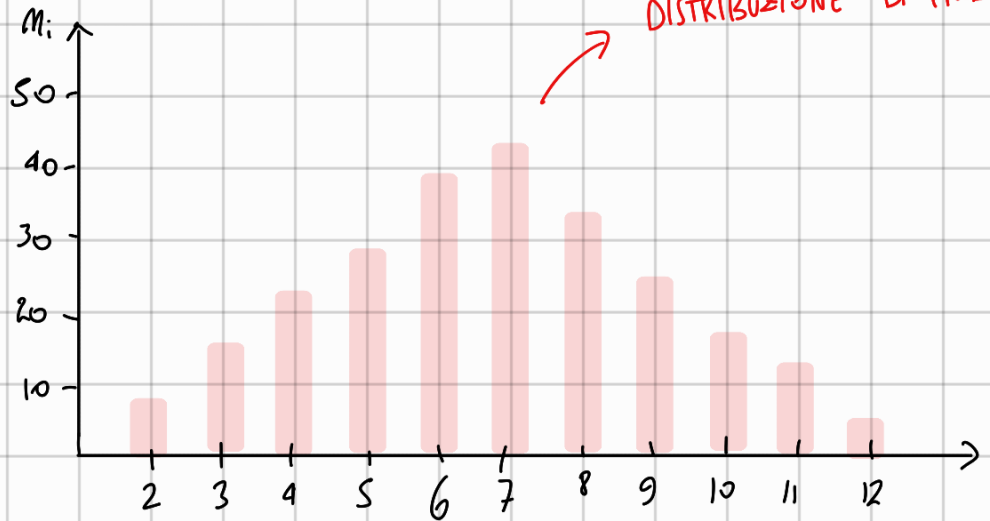
numero uscito
frequenze assolute

numero uscito	2	3	4	5	6	7	8	9	10	11	12
frequenze assolute	9	15	22	28	39	43	34	25	17	13	5
	m_2	m_3	m_4								

TABELLA

È LEGGERMENTE DIVERSA DALLA
DISTRIBUZIONE DI PROBABILITÀ

ISTOGRAMMA



frequenza relativa

$$f_i = \frac{m_i}{N}$$

$$i = 2, 3, \dots, 12$$

$$k = 11$$

NUMERO DI MODALITÀ
DIVERSE CHE PUÒ
ASSUMERE LA VARIABILE

$$f_1 = \frac{9}{250}, \quad f_2 = \frac{15}{250}, \quad \dots$$

$$0.036$$

$$0.060$$

ESEMPIO (3)

UNITA': PAZIENTI

VARIABLE: PRESSIONE SANGUIGNA

$$X = (110, 110, \dots)$$

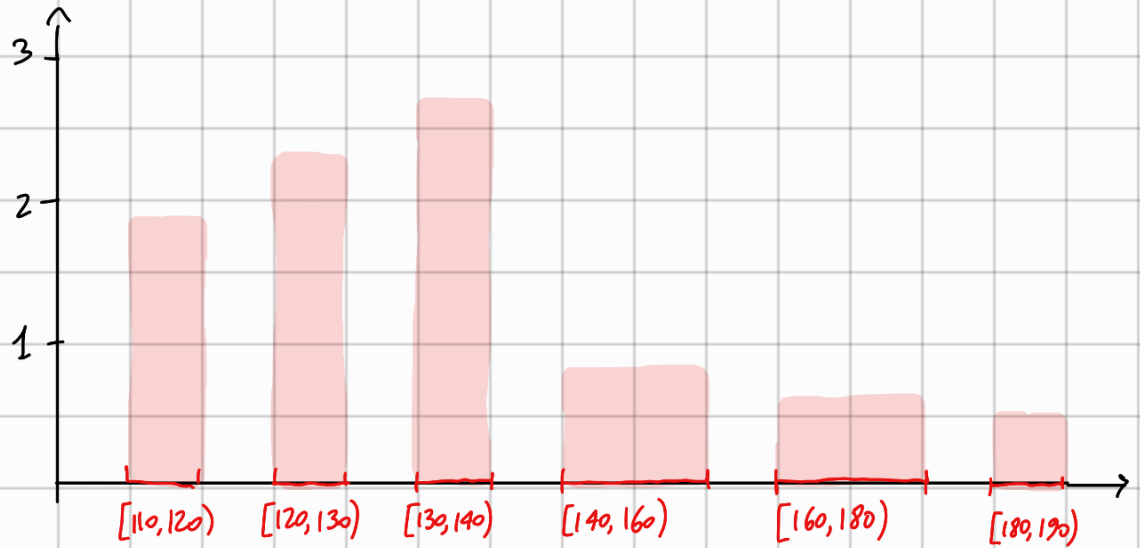
x_{min}
↑

$$(186, \dots, 187)$$

x_{max}
↑

POSSO RISCRIVERE LA SEQUENZA ORDINATA

Descrizione	Intervallo	a_i	m_i	f_i	P_i	P_i ← ALTEZZA COLONNA
ottimale	[110, 120)	10	-	:	18%	$18/10 = 1.8$
normale	[120, 130)	10	:	:	23%	$23/10 = 2.3$
normale alt.	[130, 140)	10	:	:	26%	$26/10 = 2.6$
alte lievi	[140, 160)	20	:	:	16%	$16/20 = 0.8$
alte moderate	[160, 180)	20	:	:	12%	$12/20 = 0.6$
alte gravi	[180, 190)	10	}	:	5%	$5/10 = 0.5$



BISOGNA FARE DIVERSE LE CARCHEZZE DELLE COLONNE IN MANIERA PROPORZIONALE ALLA LARGHEZZA DELL'INTERVALLO.

$$h_i = \frac{f_i}{a_i}$$

← freq. relativa

← ampiezza intervallo

↑
altezza colonna

PER LE VARIABILI QUANTITATIVE
(Specialmente quelle continue)

HA SENSO CALCOLARE GLI INDICATORI DI CENTRALITA'

① MEDIA ARITMETICA

$$X = (x_1, \dots, x_N)$$

$$\bar{x} = \frac{1}{N} \left(\sum_{i=1}^N x_i \right) = \frac{\sum_{i=1}^N x_i}{N}$$

② MEDIANA (Ha senso anche per le variabili qualitative ordinarie)

$$X = (x_1, x_2, \dots, x_N)$$

LO RIORDINO

$$X = (x^{(1)}, x^{(2)}, \dots, x^{(N)})$$

IL PIU' PICCOLO IL PIU' GRANDE

LA MEDIANA e'

SE N DISPARI

$$M = x^{(\frac{N+1}{2})}$$

SE N PARI

$$M = \frac{x^{(\frac{N}{2})} + x^{(\frac{N}{2}+1)}}{2}$$

SE QUANTITATIVA

CASO N DISPARI

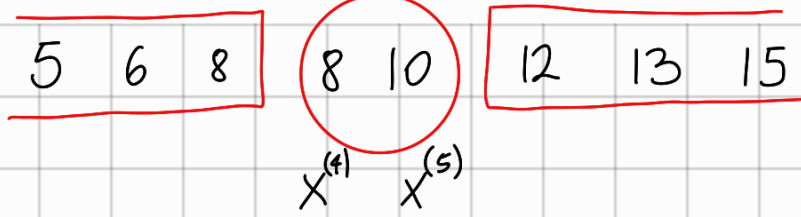
$$x^{(1)} \quad x^{(2)} \quad x^{(3)} \quad x^{(4)} \quad x^{(5)} \quad x^{(6)} \quad x^{(7)}$$

$$\boxed{5 \quad 6 \quad 8} \quad \boxed{8} \quad \boxed{10 \quad 12 \quad 13}$$

$$4 = \frac{7+1}{2}$$

$$M = x^{(4)} = 8$$

N PAM



$$N=8$$

$$M = \frac{X^{(4)} + X^{(5)}}{2} = \frac{8+10}{2} = 9$$

Una PROPRIETA' della mediana e' la **ROBUSTEZZA**
 cioe' se aggiungo un dato MOLTO SBAGLIATO (molto lontano dagli altri)

la mediana non varia di molto mentre la media si

ESEMPIO

(5	6	8	8	10	12	13)	45)	$\bar{X} = 8.85$	$\bar{X} = 13.37$
								$M = 8$	$M = \frac{8+10}{2} = 9$
								(SENZA IL 45)	(CON IL 45)

Per le variabili DISCRETE o QUALITATIVE si puo' definire un altro indice:

③ MODA (e' LA MODALITA' PIU' FREQUENTE)

Nell' ESEMPIO ① DEI DADI LA MODA e' 7 (con 43 occorrenze)

SE $X = (5, 6, 8, 8, 10, 12, 13)$ LA MODA e' 8 (con 2 occorrenze)

Se ci sono due mode si dice che il campione e' **BI MODALE**
 E indichiamo entrambi i valori

ESEMPIO: VARIABILE COLORE OCCHI C → CASTANI | A → AZZURRI
 N → NERI | V → VERDI

N = 18

$X = (C, C, A, N, N, C, V, A, A, N, N, N, C, C, A, V, N, C)$

TABELLA

C	N	A	V
6	6	4	2

MODE = {C, N} (CON 6 OCCORRENZE)