Lezione 26 09/01/24

$U_1, U_2, U_3, \dots \quad U_N$ UNITA' STATISTICHE

ESEMPIO UNITA' STATISTICA : "LANCIO DI DUE DADI"

$X = \{x_1, x_2, x_3, \dots \dots, x_N\}$ risultato ottenuto

X è UNA VARIABILE QUANTITATIVA DISCRETA

$N = 250$

$X = \{6, 6, 5, 6, 3, 4, 8, \dots \dots 8, 10, 8\}$

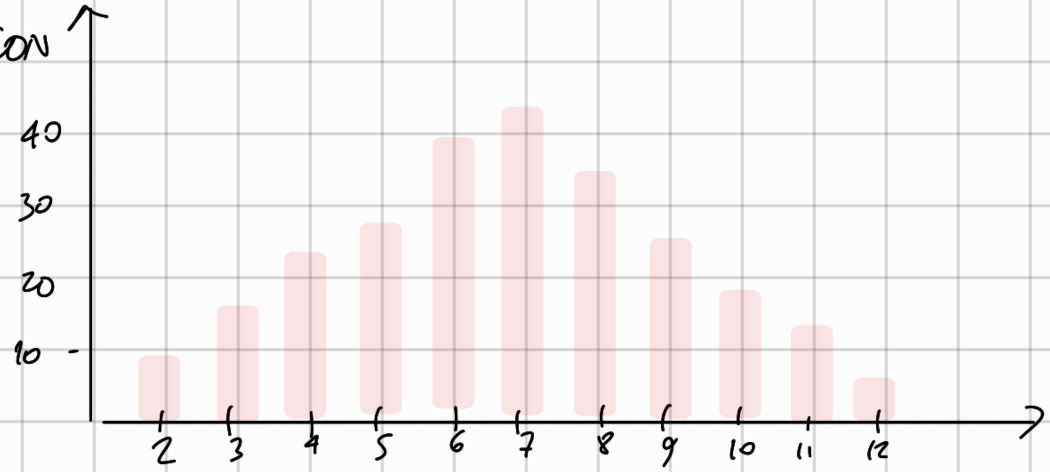
RAPPRESENTAZIONE CON UNA TABELLA

risultato	2	3	4	5	6	7	8	9	10	11	12
frequenza ASSOLUTA	9	15	22	28	39	43	34	25	17	13	5
	n_2	n_3	n_4	.		.		.		.	n_{12}

DEFINIZIONE

L'insieme delle coppie $\{(2,9), (3,15), \dots, (12,5)\}$ si chiama DISTRIBUZIONE DI FREQUENZE

RAPPRESENTAZIONE CON UN ISTOGRAMMA



frequenza RELATIVA $f_i = \frac{n_i}{N}$ $i = 2, \dots 12$

$\frac{9}{250}$	$\frac{15}{250}$	$\frac{22}{250}$	$\frac{28}{250}$	$\frac{39}{250}$	$\frac{43}{250}$	$\frac{34}{250}$	$\frac{25}{250}$	$\frac{17}{250}$	$\frac{13}{250}$	$\frac{5}{250}$
$=$										
0.036	0.006	0.880	0.112	0.156	0.172	0.136	0.100	0.068	0.052	0.02
$=$										

frequenza PERCENTUALE

3.6 %	6 %	8.8 %	11.2 %	15.6 %	17.2 %	13.6 %	10 %	6.8 %	5.2 %	2 %

Se la variabile è di tipo <u>CONTINUO</u>
non si può fare la distribuzione di frequenze

ESEMPIO UNITÀ STATISTICHE : SPIGOLE $N = 300$
VARIABILE PESO IN GRAMMI

$$X = \{217, 250, 297, \dots \dots 256\}$$

COME POSSIAMO RAPPRESENTARE QUESTI DATI?

SI FISSANO DEGLI INTERVALLI $I = [\min X, \max X]$

$$\ell = \max X - \min X = 450 - 150 = 300$$

SI SUDDIVIDE I in k parti (non necessariamente uguali)
SOLITAMENTE SÌ

$$X_{MIN} = S_0 < S_1 < S_2 < S_3 < \dots S_k = X_{MAX}$$

NELL'ESEMPIO

$$150 = S_0 < 200 < 250 < 300 < 350 < 400 < 450 = S_k = X_{MAX}$$

$$I_j = [S_{j-1}, S_j) \quad \forall j = 1 \dots k-1$$

$$I_k = [S_{k-1}, S_k]$$

	n_k	f_k
$I_1 = [150, 200)$	4	1.33 %
$I_2 = [200, 250)$	41	13.66 %
$I_3 = [250, 300)$	98	32,66 %
$I_4 = [300, 350)$	108	36 %
$I_5 = [350, 400)$	43	14.33 %
$I_6 = [400, 450]$	6	2 %

PER CASA
DISEGNARE ISTOGRAMMA

ESEMPIO: UNITÀ STATISTICA: PAZIENTI $N = 100$

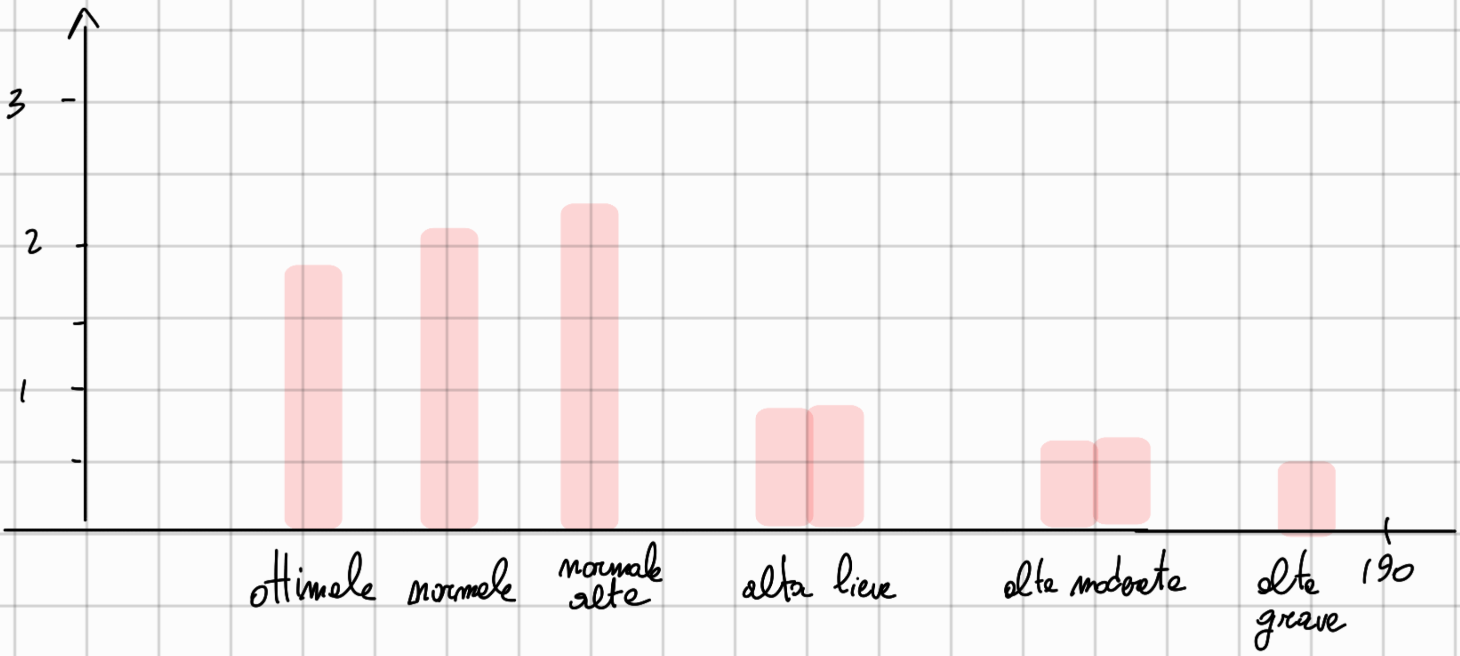
VARIABILE PRESSIONE SISTOLICA

$$X = \{110, 110, 110, 111, 112, \ldots, 186, 186, 187\}$$

↑ LI SCRIVO ORDINATI

ALTEZZA

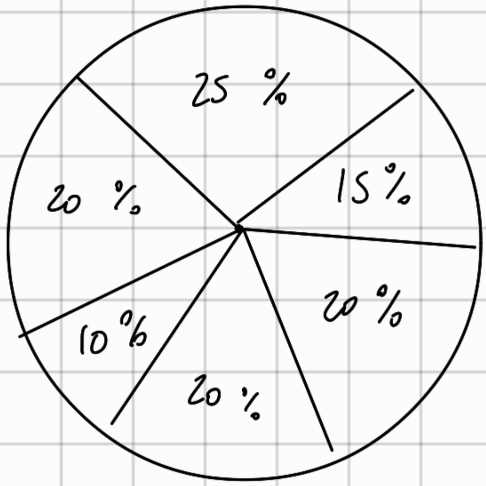
CLASSIFICAZIONE				f_k	h_k
ottimale	< 120	[110, 120)	10	18%	$\frac{18}{10} = 1.8$
normale	120 - 129	[120, 130)	10	23%	$\frac{23}{10} = 2.3$
normale alta	130 - 139	[130, 140)	10	26%	$\frac{26}{10} = 2.6$
alta lieve	140 - 159	[140, 160)	20	16%	$\frac{16}{20} = 0.8$
alta moderata	160 - 179	[160, 180)	20	12%	$\frac{12}{20} = 0.6$
alta grave	> 180	[180, 190]	10	5%	$\frac{5}{10} = 0.5$



LA FREQUENZA OTTENUTA DEVE ESSERE PROPORZIONALE ALL'AREA DEL RETTANGOLO

ESEMPIO: VARIABILI QUALITATIVE NON ORDINATE

DIAGRAMMA A TORTA



Per evitare di sceglire un ordine tra le modalità si disegna un diagramma a torta in cui l'area di ogni spicchio è proporzionale alle frequenze ottenute

---

UN ALTRO MODO DI RAPPRESENTARE DEI DATI È UTILIZZANDO UN UNICO NUMERO DETTO INDICATORE DI CENTRALITÀ (VARIABILI QUANTITATIVE)

$$X = \{x_1, x_2, x_3, x_4, \dots \quad x_N\}$$

LA MEDIA ARITMETICA di X è $\bar{x} = \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)$

RIPASSO

$$\sum_{i=1}^{3} x^i = x^1 + x^2 + x^3$$

$$X = \{5, 6, 8, 8, 8, 12, 12, 14\}$$

$$\bar{X} = \frac{1}{8}\left(5+6+8+8+8+12+12+14\right) = \frac{73}{8} = 9.125$$

ORDINATI IN MODO CRESCENTE

LA MEDIANA È IL VALORE CENTRALE DELLA SEQUENZA $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_N$

SE N è DISPARI $\qquad M = x_{\frac{N+1}{2}}$

SE N è PARI $\qquad M = \dfrac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$

UNA PROPRIETÀ IMPORTANTE DELLA MEDIANA È LA ROBUSTEZZA

Cioè se aggiungiamo al set di dati un valore molto diverso da tutti gli altri il suo valore cambia di poco

$$\text{MEDIANA}(X) = \frac{x_{\frac{8}{2}} + x_{\frac{8}{2}+1}}{2} = \frac{x_4 + x_5}{2} = \frac{8+8}{2} = 8$$

ESEMPIO

SE AGGIUNGO $x_9 = 21$

$X = \{5, 6, 8, 8, 8, 12, 12, 14, 21\}$ $\qquad N = 9$ DISPARI

$\text{MEDIANA}(X) = x_{\frac{9+1}{2}} = x_5 = 8$ $\qquad$ NON È CAMBIATA

$\bar{X} = 10.44$ $\qquad$ È CAMBIATA $(+\ 1.32)$

UN INDICATORE DI CENTRALITÀ CHE SI PUÒ UTILIZZARE PER VARIABILI DISCRETE (E CATEGORIALI)

La **MODA** È IL NUMERO CHE SI RIPETE PIÙ VOLTE
(LA MODALITÀ)

$\text{MODA}\{X\} = 8$

LA MEDIA $\boxed{\text{PESATA}}$

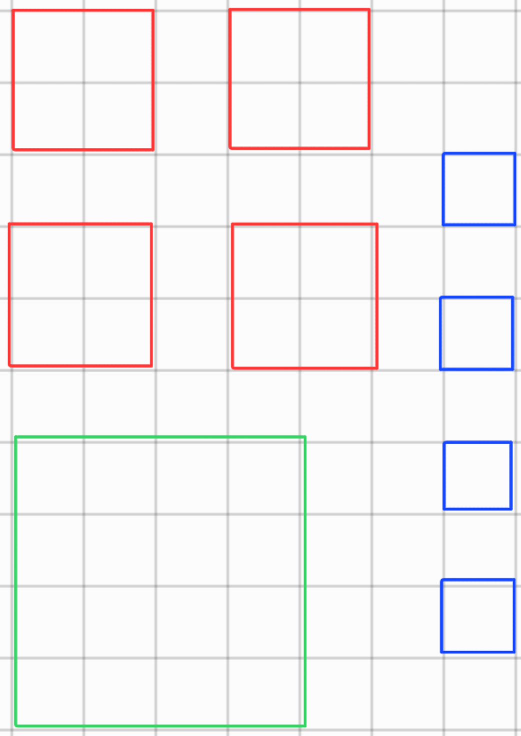
DATI UN INSIEME DI VARIABILI $X_1, X_2, X_3 \cdots X_N$ e i loro relativi PESI $p_1, p_2, p_3 \cdots p_N$

LA MEDIA PESATA

$$\overline{X}_p = \frac{\sum_{i=1}^{N} p_i x_i}{\sum_{i=1}^{N} p_i} \qquad \frac{p_1 X_1 + p_2 X_2 + \cdots + p_N X_N}{p_1 + p_2 + p_3 + \cdots + p_N}$$

LA MEDIA QUADRATICA

$$\overline{X}_q = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}}$$

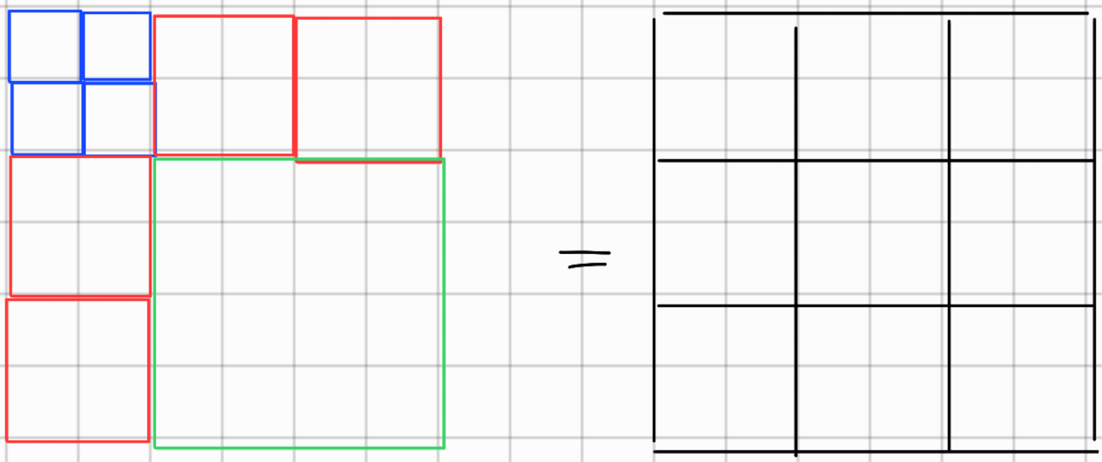


$$X = \{1, 1, 1, 1, 2, 2, 2, 2, 4\}$$

$$\overline{x}_q = \sqrt{\frac{1^2+1^2+1^2+1^2+2^2+2^2+2^2+2^2+4^2}{9}}$$

$$= \sqrt{\frac{4(1)+4(4)+16}{9}} = \sqrt{\frac{36}{9}} = \sqrt{4} = 2$$

Cioè la somma delle aree di questi 9 quadrato corrisponde alle somma delle aree di 9 quadrati uguali di lato $\overline{x}_q = 2$



LA MEDIA ARMONICA $X = \{x_1 \dots x_n\}$

$$\overline{X}_a = \frac{1}{\left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{x_i}\right)}$$

ESEMPIO: Guido per 100 km in una strada,

percorro 21 km a $v_1 = 30 \frac{km}{h}$

21 km a $v_2 = 70 \frac{km}{h}$

A quale velocità costante sarei potuto andare per arrivare nello stesso tempo?

$$\overline{X}_a = \frac{1}{\frac{1}{2}\left(\frac{1}{70} + \frac{1}{30}\right)} = \frac{2}{\frac{1}{30} + \frac{1}{70}} = \frac{2}{\frac{7+3}{210}} = \frac{2 \cdot 21\cancel{0}}{\cancel{10}} = 42 \frac{km}{h}$$