NOTA: ULTERIORI TIPI DI MEDIE

- **MEDIA ARITMETICA**: $$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- **MEDIA PESATA**: x_i UNITA' STATISTICHE, p_i PESI

$$\overline{x_p} = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i} = \frac{p_1 x_1 + \ldots + p_n x_n}{p_1 + p_2 + \ldots + p_n}$$

- **MEDIA QUADRATICA**: $$\overline{x_Q} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

- **MEDIA ARMONICA**: $$\overline{x_e} = \frac{1}{\frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i}}$$

(VARIABILI CONTINUE CON DISTR. $f(x) \rightsquigarrow$

- **MEDIA INTEGRALE**: $$\overline{x_I} = \frac{\int_a^b f(x)\,dx}{b-a}$$)

ESEMPI: 1) LIBRETTO UNIVERSITARIO:

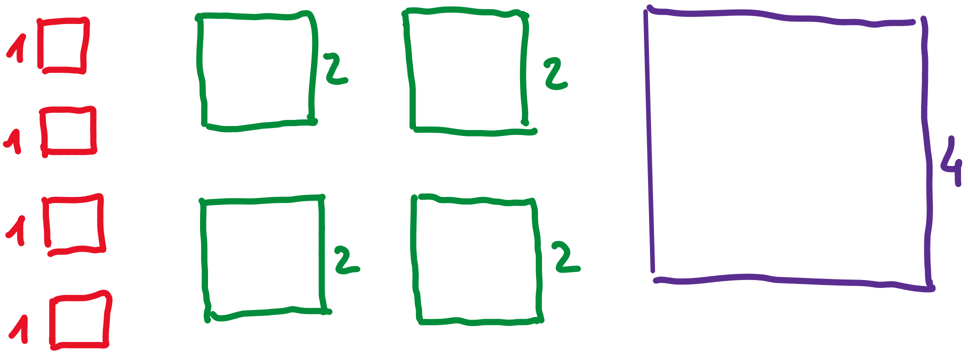
VOTAZIONE	30	28	27	30	18
N° DI CFU	8	6	6	12	4

$\bar{x}_p = ?$ STO PENSANDO x_i = VOTI

p_i = CFU

$$\bar{x}_p = \frac{30\cdot 8 + 28\cdot 6 + 27\cdot 6 + 30\cdot 12 + 18\cdot 4}{8+6+6+12+4} = \dots$$

2) $X = \{1, 1, 1, 1, 2, 2, 2, 2, 4\} \rightsquigarrow \bar{x}_Q = ?$

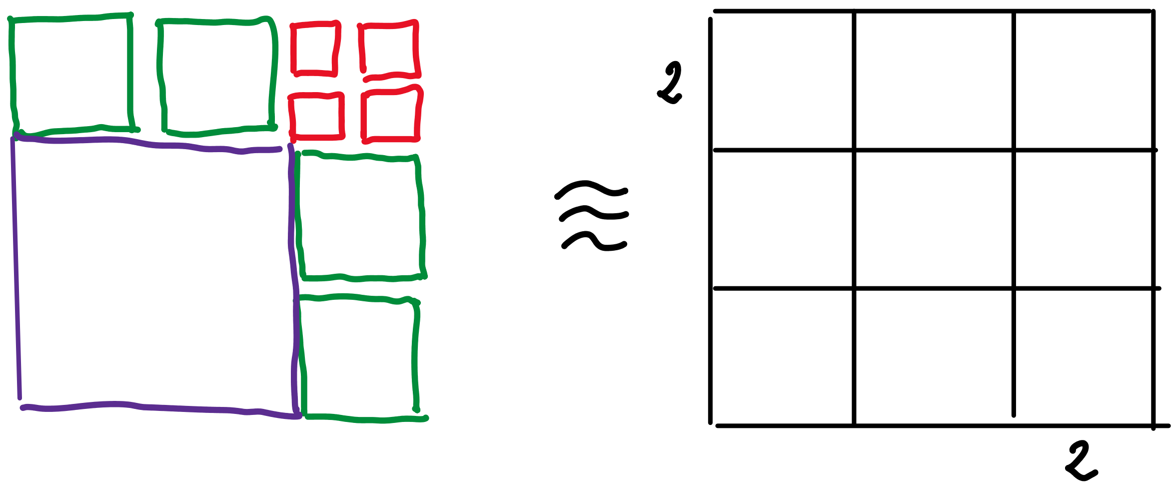


$\Downarrow$

$$\bar{X}_Q = \sqrt{\frac{1^2+1^2+1^2+1^2+2^2+2^2+2^2+2^2+4^2}{9}}$$

STO DICENDO CHE I QUADRATI SOPRA SONO EQUIVALENTI A 9 QUADRATI TUTTI DI LATO 2

$$= \sqrt{\frac{36}{9}} = \sqrt{4} = 2$$

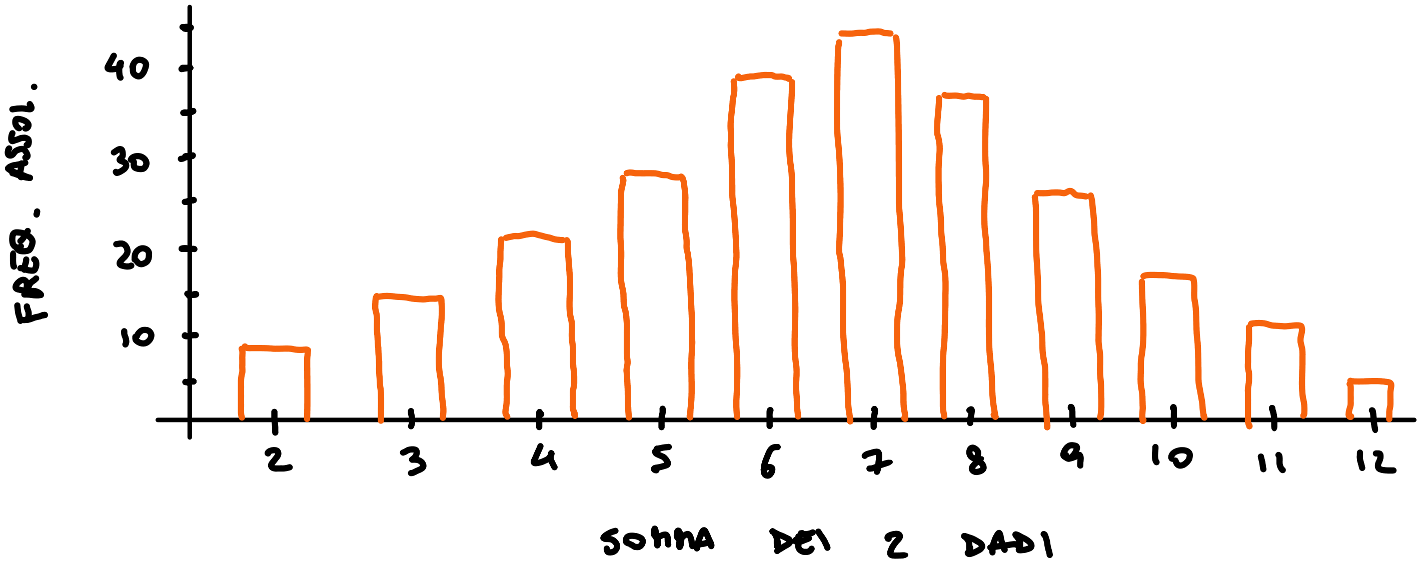


NOTA: RAPPRESENTAZIONE DEI DATI.

FINORA ABBIAMO STUDIATO DATI RAGGRUPPATI IN TABELLE O TESTI. UN MODO EFFICACE PER VISUALIZZARE LE RIPARTIZIONI DI UN CARATTERE NELLE DIVERSE MODALITÀ È L'ISTOGRAMMA; SI RAPPRESENTA CIASCUNA MODALITÀ COME UN RETTANGOLO AVENTE AREA PARI ALLA FREQUENZA CON CUI SI RIPETE.

ES: 1)

SOMMA 2 DADI	2	3	4	5	6	7	8	9	10	11	12
FREQUENZA ASSOLUTA	9	15	22	28	39	43	34	25	17	13	5



2) (VARIABILE QUANTITATIVA CONTINUA) VIENE MISURATA LA PRESSIONE SISTOLICA a $n = 100$ PAZIENTI:

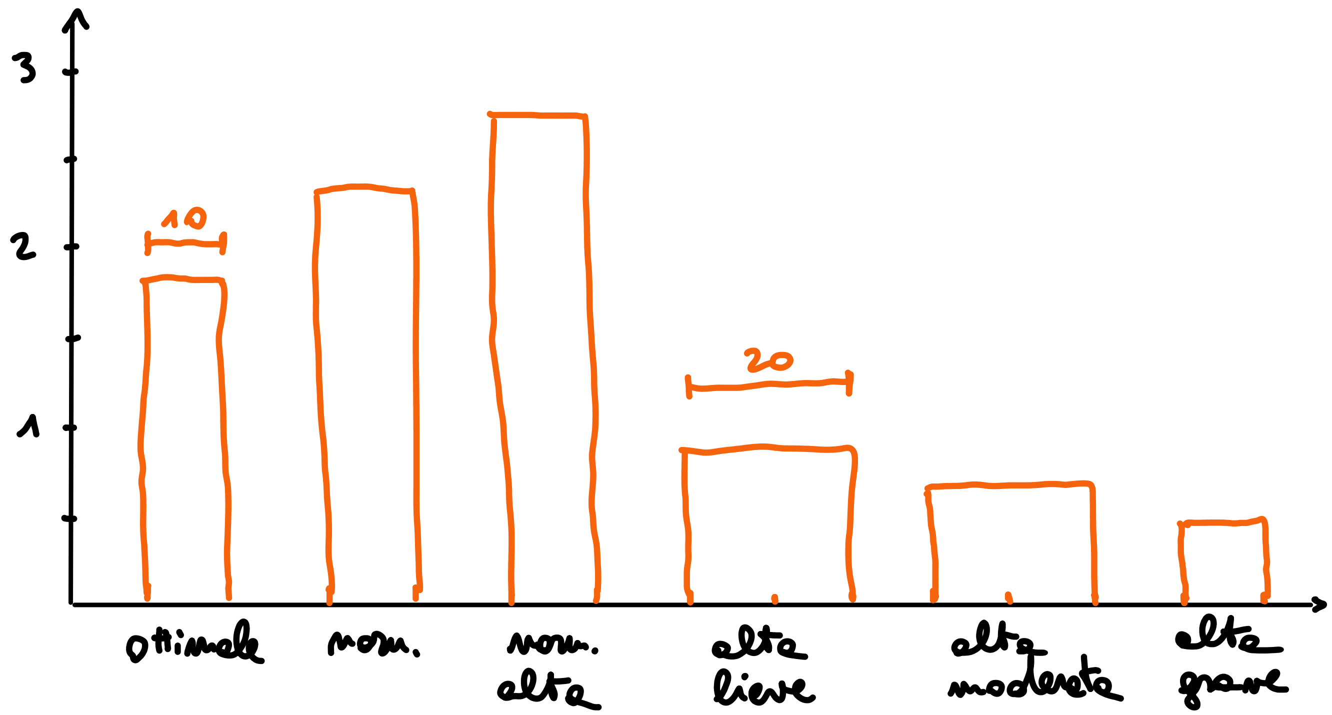
$$X = \{110, 110, 110, 111, 112, \ldots\ldots, 186, 186, 187\}$$

(ORDINATI)

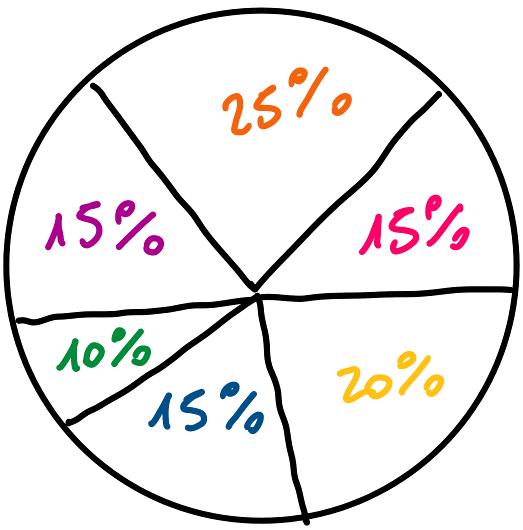
IN QUESTO TIPO DI SITUAZIONE POSSO DIVIDERE IL RANGE DEI VALORI (IN QUESTO CASO 110-190) IN INTERVALLI E CONSIDERARE CIASCUNO COME UNA MODALITÀ.

CLASSIFICAZIONE	INTERVALLO		AMPIEZZA	P_k	h_k (ALTEZZA)
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%	$\frac{12}{20} = 0.6$
[illegible]	>180	[180,190]	10	5%	$\frac{5}{10} = 0.5$

alta moderata	[illegible] [illegible]	[illegible]	[illegible]	[illegible] = 0.6
alta grave	≥180 [180,190]	10	5 %	$\frac{5}{10} = 0.5$



3) <u>DIAGRAMMA A TORTA</u>: PER EVITARE DI SCEGLIERE UN ORDINE TRA LE MODALITÀ SI DISEGNA UN DISCO DIVISO A FETTE, DOVE CIASCUNA DI ESSE RAPPRESENTA UNA MODALITÀ, AVENTI AREA PROPORZIONALE ALLA FREQUENZA.



4) ESERCIZIO: SI MISURA IL PESO DI $n = 300$ SPIGOLE OTTENENDO VALORI TRA 150 g E 450 g. IN PARTICOLARE:

	N_J	f_J

	N_J	f_J
$I_1 = [150, 200)$	4	$\frac{4}{300} = 0.0133$
$I_2 = [200, 250)$	41	0.1366
$I_3 = [250, 300)$	98	0.3266
$I_4 = [300, 350)$	108	?
$I_5 = [350, 400)$	43	?
$I_6 = [400, 450]$	6	?

DISEGNARE UN ISTOGRAMMA RELATIVO A TALI DATI.

ABBIAMO GIA' DEFINITO LA VARIANZA

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

TALVOLTA, PER ANALIZZARE CAMPIONI USEREMO LA <u>VARIANZA CAMPIONARIA</u>

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

DOVE $n-1$ SOSTITUISCE n.

<u>STATISTICA INFERENZIALE</u>: STATISTICA BIVARIATA.

CONFRONTO TRA VARIABILI. SUPPONIAMO DI VOLER CONFRONTARE I CARATTERI "ETÁ" E "DIAMETRO" DI UNA POPOLAZIONE DI ALBERI.

SU UN CAMPIONE DI m ESEMPLARI AVREMO m COPPIE $(x_1, y_1), \dots, (x_m, y_m)$, DOVE x_i É L'ETÁ E y_i IL DIAMETRO DELL' i-ESIMO ALBERO.

VOGLIAMO CAPIRE SE LE DUE GRANDEZZE X = "ETÁ" E Y = "DIAMETRO" SONO TRA LORO "LEGATE", OSSIA SE ESISTE UNA CORRELAZIO-NE TRA TALI PROPRIETÁ.

DEF: SI DEFINISCE <u>COVARIANZA TRA X E Y</u> IL NUMERO

$$\sigma_{XY} = \frac{\sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y})}{m}$$

DOVE $\bar{x}$ É LA MEDIA (ARITMETICA) DI X E $\bar{Y}$ QUELLA DI Y. OSSERVIAMO CHE SE $\sigma_{XY} > 0$ ALLORA AVREMO MAGGIORMENTE CHE $x_i > \bar{x}$ CORRISPONDONO A $y_i > \bar{y}$ (E VICEVERSA A $x_i < \bar{x}$ CORRISPONDONO $y_i < \bar{y}$)

VICEVERSA A $x_i < \bar{x}$ CORRISPONDONO $y_i < \bar{y}$)

MENTRE SE $\sigma_{xy} < 0$ ALLORA MAGGIORMENTE CHE A $x_i > \bar{x}$ CORRISPONDONO $y_i < \bar{y}$ (E VICEVERSA A $x_i < \bar{x}$ CORRISPONDONO $y_i > \bar{y}$).

- SE $\sigma_{xy} > 0$, X, Y SI DICONO CONCORDI
- SE $\sigma_{xy} < 0$, X, Y SI DICONO DISCORDI
- SE $\sigma_{xy} = 0$, X, Y SI DICONO INCORRELATI.

RICORDANDO CHE $\bar{x} = \frac{\sum_i x_i}{n}$, $\bar{y} = \frac{\sum_i y_i}{n}$

OTTENIAMO

$$\sigma_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum_i x_i y_i}{n} - \bar{x} \cdot \bar{y}$$

$$= \overline{xy} - \bar{x} \cdot \bar{y}$$

OSSIA σ_{xy} É PARI A "LA MEDIA DEL PRODOTTO MENO IL PRODOTTO DELLE MEDIE".

NOTAZIONE: σ_x^2 = VARIANZA DI X

σ_y^2 = VARIANZA DI Y

$$\sigma_x = \sqrt{\sigma_x^2} \quad , \quad \sigma_y = \sqrt{\sigma_y^2}$$

DEF: SI CHIAMA <u>COEFFICIENTE DI BRAVAIS-PEARSON</u> (O COEFFICIENTE DI CORRELAZIONE) IL VALORE

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2} \cdot \sqrt{\sigma_y^2}} = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2} \cdot \sqrt{\sum_i (y_i-\bar{y})^2}}$$

SI PONE $\rho_{xy} = 0$ SE $\sigma_x^2 = 0$ O $\sigma_y^2 = 0$.

VALE CHE $-1 \leq \rho_{xy} \leq 1$, INOLTRE ESSO QUANTIFICA QUANTO X, Y SONO LEGATI DA UNA RELAZIONE LINEARE, OSSIA UNA RETTA $Y = aX + b$.

- SE $\rho_{xy} = \pm 1 \Rightarrow X, Y$ SONO IN PERFETTA RELAZIONE LINEARE, QUINDI I PUNTI (x_i, y_i) SONO ALLINEATI A FORMARE UNA RETTA (DI COEFF. ANGOLARE $a > 0$ SE $\rho_{xy} = +1$, DI COEFF. ANGOLARE $a < 0$ SE $\rho_{xy} = -1$)
- SE $\rho_{xy} = 0 \Rightarrow \nexists$ RELAZIONE LINEARE TRA X, Y.

ESEMPIO: $\{(1,1),(5,9),(4,7),(10,19),(2,3)\}$

$\rho_{xy} = ?$ $\quad \bar{x} = ?$ $\quad \bar{y} = ?$

$$\bar{x} = \frac{\sum_i x_i}{5} = \frac{1+5+4+10+2}{5} = \frac{22}{5} = 4.4$$

$$\bar{y} = \frac{\sum_i y_i}{5} = \frac{1+9+7+19+3}{5} = \frac{39}{5} = 7.8$$

$$\sigma_x^2 = \frac{\sum_i (x_i-\bar{x})^2}{5} = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+(x_4-\bar{x})^2+(x_5-\bar{x})^2}{5}$$

$$= \frac{(-3.4)^2+(0.6)^2+(-0.4)^2+(5.6)^2+(-2.4)^2}{5} = \frac{49.2}{5} = 9.84$$

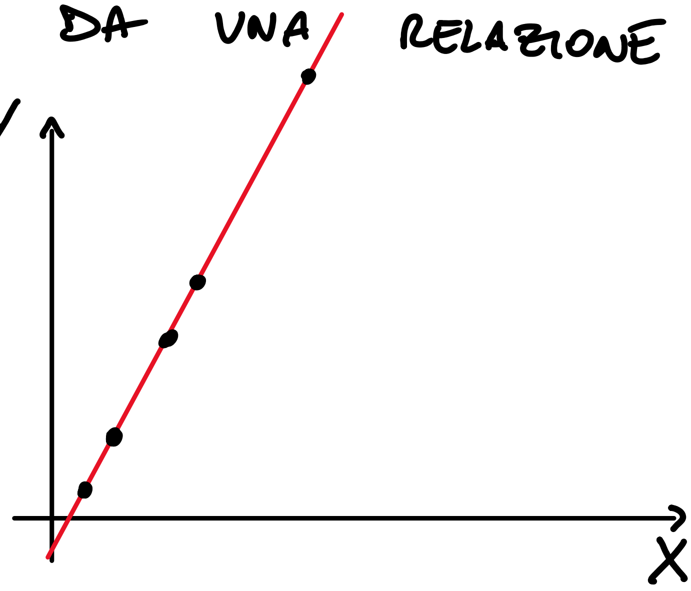
$\sigma_y^2 =$ ESERCIZIO.

$$\rho_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2}\cdot\sqrt{\sigma_y^2}} = \frac{\sum_i \overbrace{(x_i-\bar{x})}(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2}\cdot\sqrt{\sum_i (y_i-\bar{y})^2}} =$$

$$= \frac{\overbrace{(-3.4)}(-6.8)+\overbrace{(0.6)}(1.2)+\overbrace{(-0.4)}(-0.8)+\overbrace{(5.6)}(11.2)+\overbrace{(-2.4)}(-4.8)}{\sqrt{49.2}\cdot\sqrt{196.8}}$$

$$= \frac{98.4}{\sqrt{49.2\cdot 196.8}} = \boxed{1}$$

$\Rightarrow$ X, Y SONO LEGATI DA UNA RELAZIONE LINEARE. IN EFFETTI I PUNTI ANALIZZATI SONO ALLINEATI SULLA RETTA $y = 2x - 1$.



DOMANDA: POSSIAMO APPROSSIMARE CON UNA RETTA LA DISPOSIZIONE DI COPPIE (= PUNTI) ANCHE QUANDO ρ_{XY} NON É ESATTAMENTE ± 1?

RISPOSTA: <u>SÍ</u>, CIÓ PUÓ ESSERE FATTO OGNI VOLTA CHE $\rho_{XY} \neq 0$. LA RETTA APPROSSIMANTE PRENDE IN TAL CASO IL NOME DI <u>RETTA DI REGRESSIONE LINEARE</u>, E I SUOI COEFFICIENTI a, b (NELLA FORMA $Y = aX + b$) SONO TROVATI COL METODO SEGUENTE: CERCHIAMO $a, b \in \mathbb{R}$ CHE MINIMIZZANO

$$\min_{a,b} \sum_{i=1}^{n} \left[y_i - (a x_i - b) \right]^2$$

[METODO DEI MINIMI QUADRATI, GAUß (1794)]