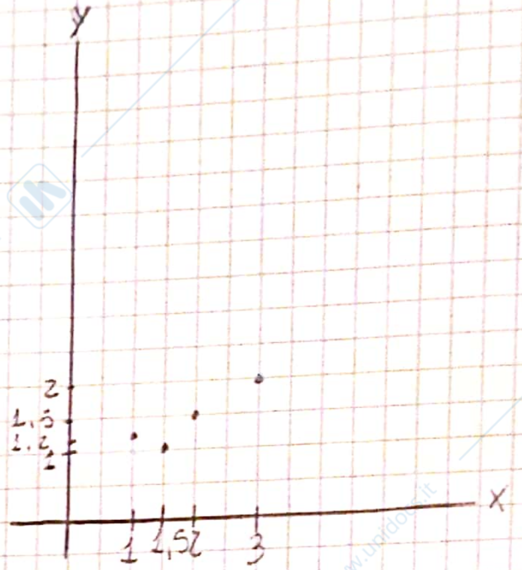
## Il diagramma a dispersione

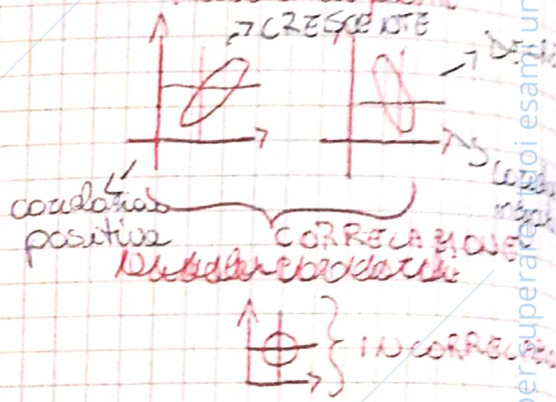
Il diagramma a dispersione, o scatter plot, ci permette di rappresentare sul grafico due variabili

ESEMPIO:

| x | y |
|---|---|
| 1 | 1,2 |
| 1,5 | 1 |
| 2 | 1,5 |
| 3 | 2 |

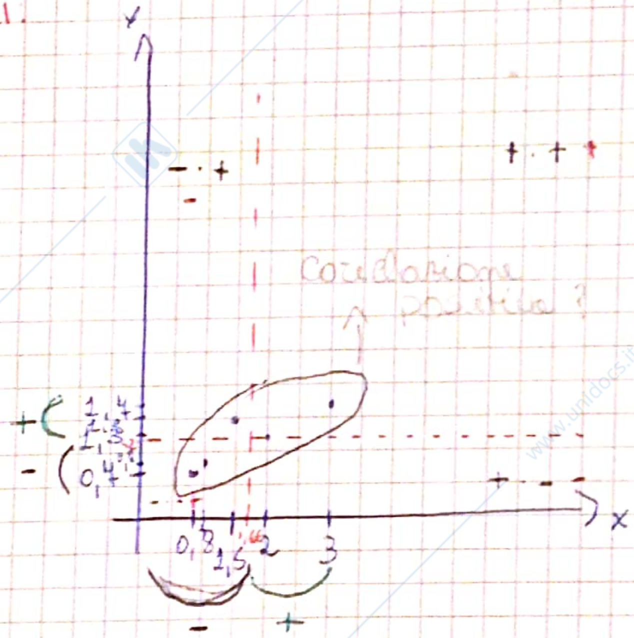



Ogni individuo è visto come un punto su questo diagramma che compongono una nuvola di punti




ESEMPIO 2 VARIABILI:

| | x | y | |
|---|---|---|---|
| - | (0,8 | 0,7) | - |
| + | (2 | 1,3) | + |
| + | (3 | 1,7) | + |
| - | (1 | 0,8) | - |
| - | (1,5 | 1,5) | + |




Quindi

~~(0,8; 1,66)~~

~~(0,8)~~ +

~~(1; 1,66)~~

(0,28)

Calcoli la $M_x$ e la $M_y$

$M_x = 1,66$

$M_y = 1,2$

Sul grafico rappresento le due medie con due assi, con i due assi vado ad individuare gli scarti. Se gli scarti saranno superiori alla media saranno degli scarti concordi, ~~viceversa~~ se saranno inferiori saranno discordi. Il prodotto degli scarti $(x - M_x) \cdot (y - M_y)$ si può dare o ~~...~~ due concordi o due discordi (regola dei segni)

Calcovianza

$\Sigma[(x-M_x)\cdot(y-M_y)]$ -> è un indice che ci misura la correlazione.

Covarianza > 0 -> correlazione positiva

Covarianza < 0 -> correlazione negativa

Covarianza = 0 -> incorrelazione

Per avere un indice ancora più preciso (come nel discorso della devianza), calcolo la covarianza

$$\frac{\Sigma[(x-M_x)\cdot(y-M_y)]}{N} = \sigma_{xy}$$

Ha una forma abbreviata: $\underbrace{\sigma_{xy} = \left(\frac{\Sigma x\cdot y}{N}\right) - (M_x \cdot M_y)}_{\text{Media dei prodotti meno il prodotto tra le medie.}}$

Covarianza nell'esempio:

| x·y |
|---|
| 0,56 |
| 2,6 |
| 5,1 |
| 0,8 |
| 2,25 |
| 11,31 -> $\Sigma(x\cdot y)$ |

$$\left(\frac{11,31}{5}\right) - (1,66\cdot 1,2) = 2,26 - 1,99 = 0,27$$

$\sigma = 0,27$ -> Correlazione positiva

Dobbiamo capire se la correlazione è grande o piccola, per farlo bisogna capire quanto è il massimo che può assumere la correlazione.

Disuguaglianza di Cauchy e Swartz

$|\sigma_{xy}| \leq \sigma_x \cdot \sigma_y$ $\qquad \sigma_x = \sqrt{\frac{\Sigma(x-M_x)^2}{5}}$ $\qquad \sigma_y = \sqrt{\frac{\Sigma(y-M_y)^2}{5}}$

Il massimo che può assumere la covarianza è il prodotto tra le deviazioni standard di x e la deviazione standard di y.

$\sigma_{MAX} = \sigma_x \cdot \sigma_y$

Coefficiente di correlazione lineare di Bravais e Pearson

$\rho = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$
- $\geq -1$ -> correlazione negativa
- $= 0$ -> incorrelazione
- $\leq +1$ -> correlazione positiva.