DISTRIBUZIONI MULTIVARIATE

CdS Fisica
Laboratorio Meccanica e Termodinamica

Distribuzioni Multivariate

- Alcuni fenomeni richiedono, per la loro descrizione, l'utilizzo di più variabili casuali.

- Le distribuzioni di probabilità di due (o più) variabili casuali si chiamano **distribuzioni multivariate**.

- Tratteremo solo il caso di due variabili, le cui distribuzioni sono dette **bivariate**.

- In particolare, noi siamo interessati **a grandezze misurate indirettamente a partire da grandezze dirette, tutte soggette a errori casuali.**

Distribuzioni Multivariate (II)

Ad esempio,
per ottenere la velocità iniziale v_0 di un proiettile misuriamo ripetutamente la gittata R e il tempo di volo T. L'angolo di emissione è mantenuto costante in tutte le misure, ma non è noto.
La relazione che fornisce v_0 in funzione di R e T è

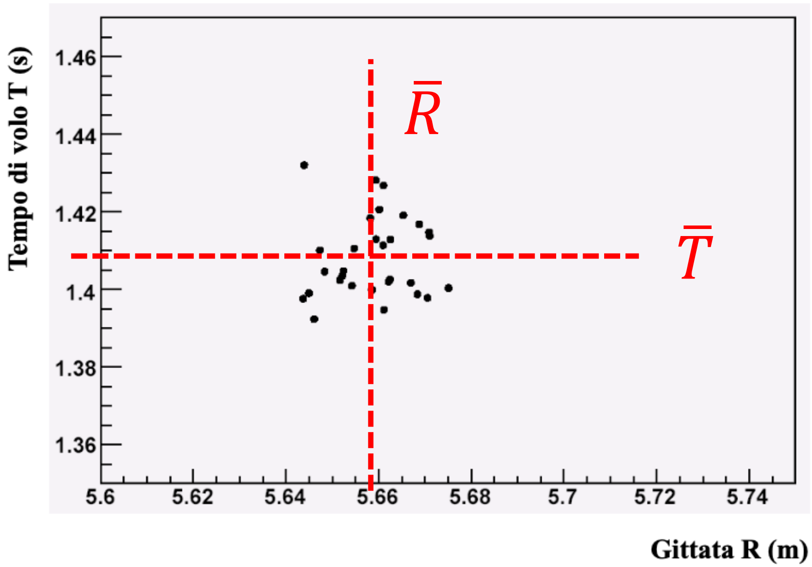
$$v_0(R,T) = \sqrt{\left(\frac{R}{T}\right)^2 + \left(\frac{Tg}{2}\right)^2}$$

Facciamo 30 lanci e rappresentiamo il campione di coppie (R_i,T_i) tramite tabelle o istogrammi.

Distribuzioni Multivariate (III)

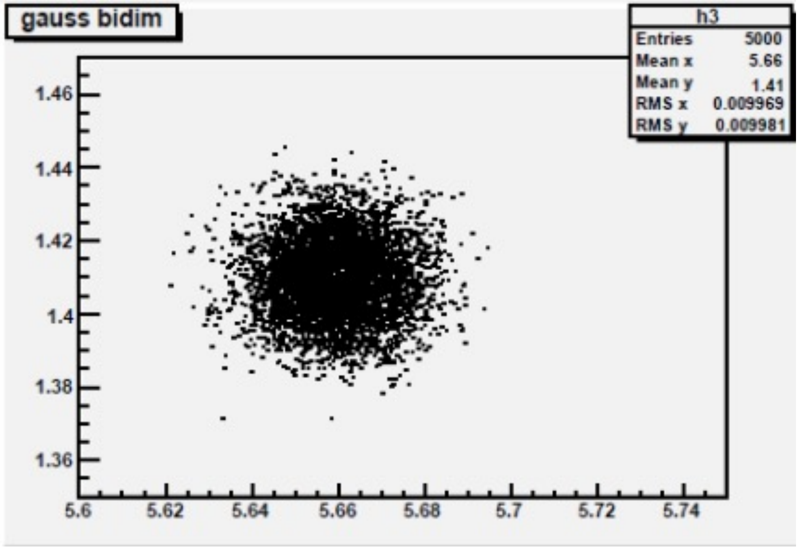
R (m)	T (s)	R (m)	T (s)
5,676	1,400	5,668	1,417
5,662	1,412	5,661	1,395
5,657	1,419	5,659	1,413
5,660	1,412	5,645	1,393
5,649	1,405	5,652	1,402
5,654	1,401	5,644	1,398
5,662	1,402	5,644	1,399
5,665	1,419	5,671	1,398
5,666	1,403	5,651	1,405
5,660	1,427	5,661	1,421
5,671	1,414	5,659	1,428
5,644	1,433	5,652	1,403
5,663	1,403	5,659	1,400
5,647	1,411	5,669	1,398
5,672	1,415	5,655	1,411

Grafico di dispersione
(**Scatter plot**)



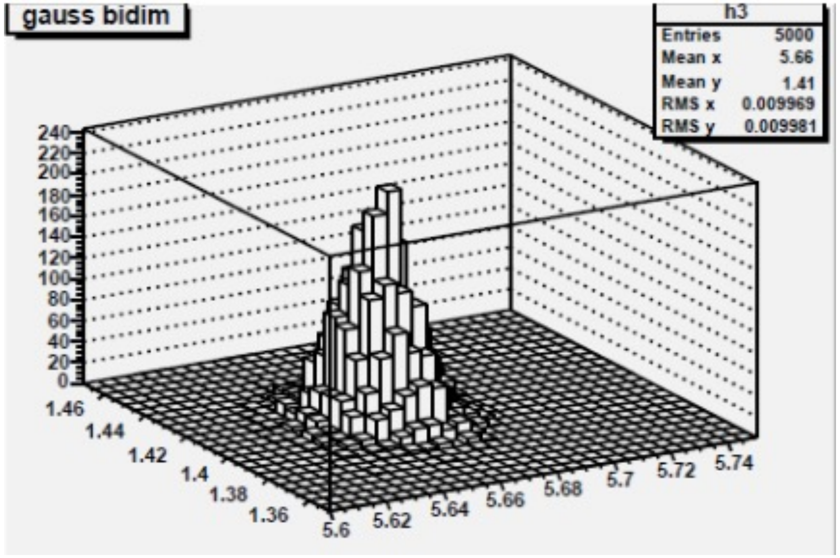
Distribuzioni Multivariate (IV)

Aumentando le misure a 5000 il grafico diventa:



Lego plot

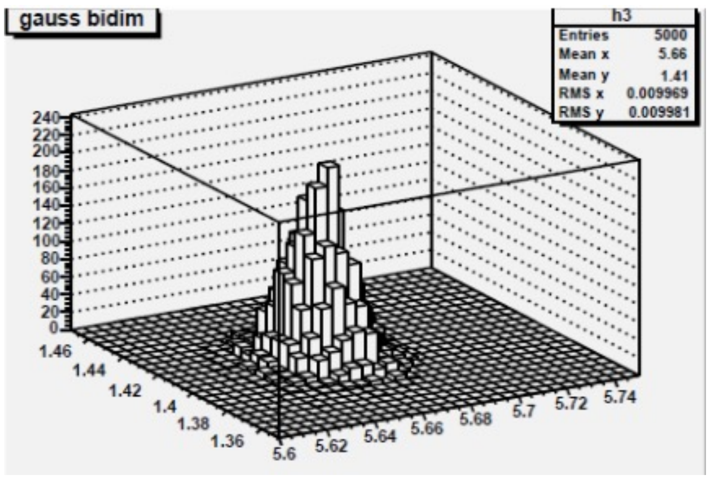
(istogramma bidimensionale, necessita di binnaggio)

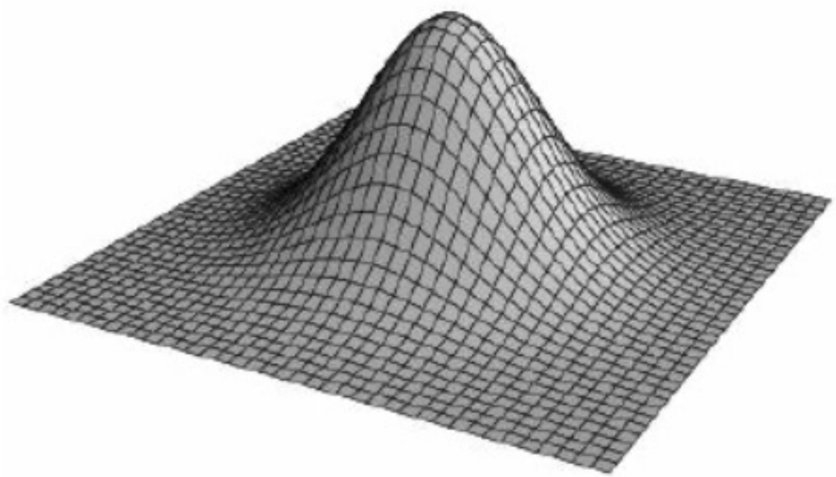


Distribuzioni Multivariate (IV)

In analogia con quanto visto nel caso unidimensionale, nel limite di infinite misure e ampiezza dei bin tendente a zero, il lego-plot in densità di frequenza tende ad una **superficie limite** (**PDF bidimensionale**)

$$\Phi_{j,k} = \frac{n_{jk}}{N\Delta x \Delta y} \xrightarrow[\substack{\Delta x \to 0 \\ \Delta y \to 0}]{N \to \infty} \Phi(x, y)$$





Distribuzioni Multivariate (V)

- A partire dal campione di misure di gittata e tempo di volo, come otteniamo il valore ottimale della velocità di lancio e la sua incertezza ?
- Potremmo calcolare, in base alla formula, v_{0i} da ogni coppia (R_i,T_i) per poi calcolare media e deviazione standard della media del campione di v_{0i}
- Ma si può ottenere il risultato anche direttamente dal campione di coppie (R_i,T_i) e dalla funzione che definisce v_0
- Le incertezze associate ai valori medi di R e di T sono casuali, non sono incertezze massime. Come si propagano queste incertezze sulla misura di v_0 ?

Distribuzioni Multivariate (VI)

Da un campione di N coppie x_i, y_i di misure soggette ad errori casuali, se q = q(x,y) risulta che:

$$\bar{q} = q(\bar{x}, \bar{y})$$

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)\sigma_{xy}$$

Dove il termine

$$\boxed{\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}$$

È detto **covarianza**

Si dimostra che

$$\boxed{\sigma_{xy} = \overline{xy} - \bar{x}\bar{y}}$$

La covarianza è uguale alla media del prodotto meno il prodotto delle medie

Distribuzioni Multivariate (VII)

$\boldsymbol{\bar{q} = q(\bar{x}, \bar{y})}$ Dimostrazione

Sia dato un campione $\{(x_1, y_1), (x_2, y_2), \ \dots . (x_N, y_N)\}$

$$q_i = q(x_i, y_i) = q(\bar{x}, \bar{y}) + \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x} \\ y=\bar{y}}} (x_i - \bar{x}) + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x} \\ y=\bar{y}}} (y_i - \bar{y})$$

Sviluppo di Taylor di una funzione in due variabili

$$\bar{q} = \frac{\sum_{i=1}^{N} q_i}{N} = \frac{\sum_{i=1}^{N} \left[q(\bar{x}, \bar{y}) + \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x} \\ y=\bar{y}}} (x_i - \bar{x}) + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x} \\ y=\bar{y}}} (y_i - \bar{y}) \right]}{N} =$$

$$= \frac{\sum_{i=1}^{N} \boxed{q(\bar{x}, \bar{y})}}{N} + \frac{\sum_{i=1}^{N} \boxed{\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x} \\ y=\bar{y}}}} (x_i - \bar{x})}{N} + \frac{\sum_{i=1}^{N} \boxed{\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x} \\ y=\bar{y}}}} (y_i - \bar{y})}{N} =$$

Sommatoria di una somma = somma di sommatorie

Distribuzioni Multivariate (VIII)

$$= q(\bar{x},\bar{y})\frac{\overbrace{\boxed{\sum_{i=1}^{N} 1}}^{=\mathrm{N}}}{N} + \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}} \frac{\sum_{i=1}^{N}(x_i-\bar{x})}{N} + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}} \frac{\sum_{i=1}^{N}(y_i-\bar{y})}{N} =$$

$$= q(\bar{x},\bar{y}) + \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}} \left(\underbrace{\boxed{\frac{\sum_{i=1}^{N} x_i}{N}}}_{=\bar{x}} - \frac{\sum_{i=1}^{N} \boxed{\bar{x}}}{N}\right) + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}} \left(\underbrace{\boxed{\frac{\sum_{i=1}^{N} y_i}{N}}}_{=\bar{y}} - \frac{\sum_{i=1}^{N} \boxed{\bar{y}}}{N}\right) =$$

$$= q(\bar{x},\bar{y}) + \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}} \left(\bar{x} - \bar{x}\frac{\overbrace{\boxed{\sum_{i=1}^{N} 1}}^{=\mathrm{N}}}{N}\right) + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}} \left(\bar{y} - \bar{y}\frac{\overbrace{\boxed{\sum_{i=1}^{N} 1}}^{=\mathrm{N}}}{N}\right) =$$

$$= q(\bar{x},\bar{y}) + \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}} \cancel{\left(\bar{x} - \bar{x}\frac{N}{N}\right)} + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}} \cancel{\left(\bar{y} - \bar{y}\frac{N}{N}\right)} = q(\bar{x},\bar{y}) \qquad \text{(Cvd)}$$

Distribuzioni Multivariate (IX)

$$\boldsymbol{\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)\sigma_{xy}}$$

Dimostrazione

$$s_q^2 = \frac{\sum_{i=1}^{N}(q_i - \bar{q})^2}{N-1}$$

$$q_i = q(x_i, y_i) = \underbrace{\boxed{q(\bar{x}, \bar{y})}}_{=\bar{q}} + \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}}(x_i - \bar{x}) + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}}(y_i - \bar{y}) \Longrightarrow$$

Sviluppo di Taylor di una funzione in due variabili

$$q_i - \bar{q} = \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}}(x_i - \bar{x}) + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}}(y_i - \bar{y})$$

$$s_q^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}}(x_i - \bar{x}) + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\ y=\bar{y}}}(y_i - \bar{y})\right]^2$$

Distribuzioni Multivariate (X)

$$s_q^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}(x_i-\bar{x}) + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}(y_i-\bar{y})\right]^2 =$$

$$\frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\right)^2(x_i-\bar{x})^2 + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}^2(y_i-\bar{y})^2 + 2\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}(x_i-\bar{x})\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}(y_i-\bar{y})\right]$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N}\boxed{\left(\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\right)^2}(x_i-\bar{x})^2 + \sum_{i=1}^{N}\boxed{\left(\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\right)^2}(y_i-\bar{y})^2\right.$$

$$\left. + \sum_{i=1}^{N}\boxed{2\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}}(x_i-\bar{x})(y_i-\bar{y})\right]$$

Distribuzioni Multivariate (XI)

$$= \frac{1}{N-1}\left[\left(\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\right)^2 \sum_{i=1}^{N}(x_i-\bar{x})^2 + \left(\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\right)^2 \sum_{i=1}^{N}(y_i-\bar{y})^2\right.$$

$$\left. + 2\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})\right]$$

$$= \left(\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\right)^2 \overbrace{\boxed{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N-1}}}^{=s_x^2} + \left(\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\right)^2 \overbrace{\boxed{\frac{\sum_{i=1}^{N}(y_i-\bar{y})^2}{N-1}}}^{=s_y^2}$$

$$+ 2\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\boxed{\frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{N-1}} = s_{xy}$$

Covarianza di x,y

$$= \left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}^2 s_x^2 + \left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}^2 s_y^2 + 2\left[\frac{\partial q}{\partial x}\right]_{\substack{x=\bar{x}\\y=\bar{y}}}\left[\frac{\partial q}{\partial y}\right]_{\substack{x=\bar{x}\\y=\bar{y}}} s_{xy} \qquad \text{(Cvd)}$$

Distribuzioni Multivariate (XII)

$$\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\boxed{N-1}}$$ Da dimostrazione

$$\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\boxed{N}}$$ Da definizione

- Per arrivare alle due varianze campionarie σ_x^2 e σ_y^2 c'è bisogno di (N-1), per non andare in contraddizione con la correttezza della definizione. Di conseguenza, c'è (N-1) anche al denominatore della covarianza.

- Una valutazione "ragionevole" della covarianza può essere fatta solo con campioni numerosi, cioè per N grande, quindi con $N-1 \approx N$.
- La valutazione della varianza anche con campioni poco numerosi può essere frequente, per la valutazione della incertezza da associare ad un valore medio, mentre la valutazione della covarianza da campioni ridotti è molto meno frequente

Distribuzioni Multivariate (XIII)

$\boldsymbol{\sigma_{xy} = \overline{xy} - \bar{x}\bar{y}}$ Dimostrazione

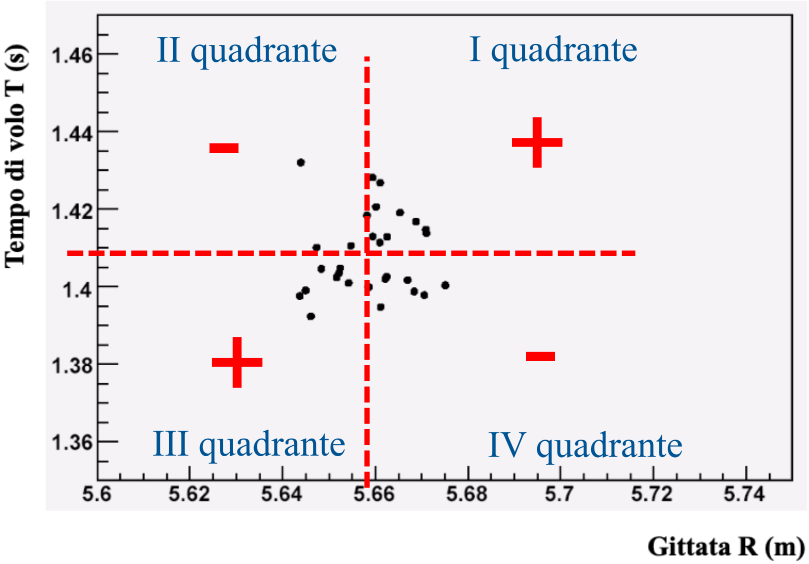
$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) = \frac{\sum_{i=1}^{N}(x_iy_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})}{N} =$$

$$= \underbrace{\boxed{\frac{\sum_{i=1}^{N}(x_iy_i)}{N}}}_{=\overline{xy}} - \frac{\sum_{i=1}^{N}(x_i\boxed{\bar{y}})}{N} - \frac{\sum_{i=1}^{N}(\boxed{\bar{x}}y_i)}{N} + \frac{\sum_{i=1}^{N}\boxed{(\bar{x}\bar{y})}}{N} =$$

$$= \overline{xy} - \bar{y}\underbrace{\boxed{\frac{\sum_{i=1}^{N}(x_i)}{N}}}_{=\bar{x}} - \bar{x}\underbrace{\boxed{\frac{\sum_{i=1}^{N}(y_i)}{N}}}_{=\bar{y}} + \bar{x}\bar{y}\frac{\underbrace{\boxed{\sum_{i=1}^{N}1}}_{=N}}{N} =$$

$$= \overline{xy} - \bar{y}\bar{x} - \bar{x}\bar{y} + \bar{x}\bar{y}\frac{\cancel{N}}{\cancel{N}} = \overline{xy} - \cancel{\bar{y}\bar{x}} - \bar{x}\bar{y} + \cancel{\bar{x}\bar{y}} = \overline{xy} - \bar{y}\bar{x}$$ Cvd

Covarianza $\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N}$



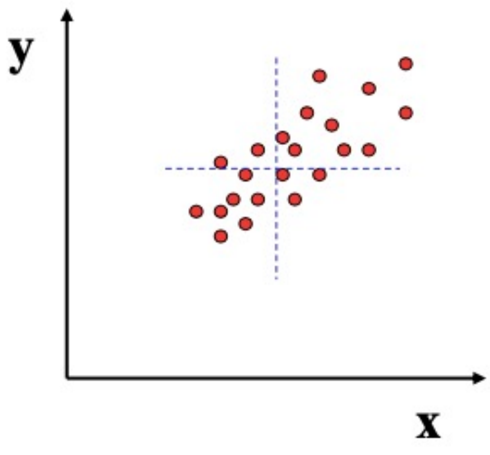
- Primo quadrante: $x_i - \bar{x} > 0$ e $y_i - \bar{y} > 0$ $\Rightarrow \sigma_{xy} > 0$
- Secondo quadrante: $x_i - \bar{x} < 0$ e $y_i - \bar{y} > 0$ $\Rightarrow \sigma_{xy} < 0$
- Terzo quadrante: $x_i - \bar{x} < 0$ e $y_i - \bar{y} < 0$ $\Rightarrow \sigma_{xy} > 0$
- Quarto quadrante: $x_i - \bar{x} > 0$ e $y_i - \bar{y} < 0$ $\Rightarrow \sigma_{xy} < 0$

Covarianza totale NULLA

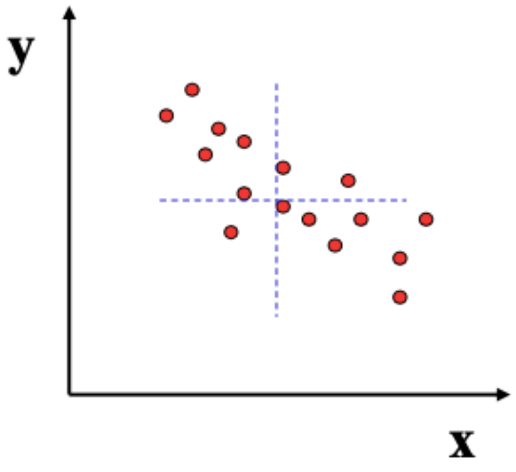
Nel caso in cui gli scarti $(x_i - \bar{x})$ siano **indipendenti** dagli scarti $(y_i - \bar{y})$, la **covarianza è nulla** perché fissato un generico x e quindi un dato $(x - \bar{x})$, ad ogni punto di ordinata y che giace sopra $\bar{y}$ di una quantità $|\bar{y} - y|$ corrisponde un punto la cui ordinata giace sotto $\bar{y}$ della stessa quantità $|\bar{y} - y|$ in modo che i termini $(x - \bar{x})(y - \bar{y})$ si cancellino a vicenda.

Covarianza (II)

Se, invece, la distribuzione delle misure delle variabili x e y si presentasse in uno dei modi seguenti, se cioè gli scarti **non** fossero **indipendenti**, la covarianza σ_{xy} non sarebbe nulla.



correlazione tra le misure di x e y (covarianza positiva)



anti-correlazione tra le misure di x e y (covarianza negativa)

Covarianza (III)

- La covarianza, quindi, misura il grado di correlazione tra le misure di x e y.

- Se le misure delle due grandezze sono **indipendenti** (cioè non correlate), nel senso che lo scarto, rispetto alla media, sulla misura di x non influenza lo scarto sulla misura di y (e viceversa) allora la **covarianza è "nulla".**

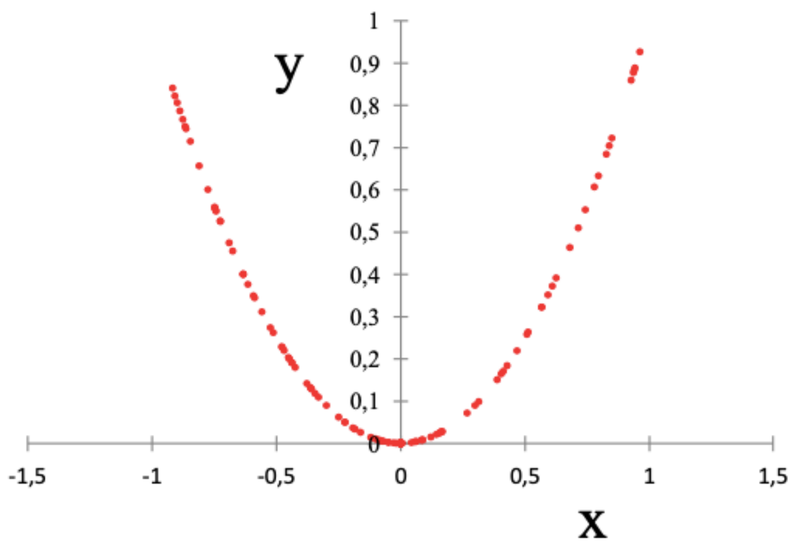
- Se, al contrario, le misure sono **correlate (non indipendenti)** allora **covarianza è diversa da zero**;

- $\boldsymbol{\sigma_{xy} > 0} \rightarrow$ **correlazione positiva**
  $\boldsymbol{\sigma_{xy} < 0} \rightarrow$ **correlazione negativa** (anticorrelazione).

Covarianza (IV)

Abbiamo visto che **se due variabili casuali x e y sono indipendenti** (cioè non correlate) allora la **loro covarianza è nulla**: $\sigma_{xy} = 0$.
Non è vero il contrario (non è un se e solo se).

Esempio: x uniforme fra -1 e +1 e $y = x^2$



Sono evidentemente correlate, ma la loro covarianza è nulla.

Incertezze in Misure Indipendenti

Nel caso (frequente) in cui le misure delle variabili x e y siano indipendenti, la deviazione standard della q si riduce quindi a:

$$\boxed{\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2}$$

La formula definisce la somma in quadratura (o somma quadratica) delle incertezze, nel caso di incertezze casuali e indipendenti delle misure dirette.

La trattazione fatta per due variabili può essere **generalizzata** ad un numero generico di variabili → **matrice di covarianza**.

Se tutte le misure sono casuali e indipendenti, la formula della somma in quadratura **si estende a tutte le variabili**.

Esercizio 1

Supponiamo che la massa di una persona adulta segua una legge gaussiana con media 75 kg e deviazione standard 5 kg.
In un certo tipo di ascensore, il numero massimo di persone ammesse in cabina è nove. Il meccanismo, però, non si avvia se la massa totale delle persone presenti è superiore a 700 kg.
Qual è la probabilità che la massa totale di nove persone superi i 700 kg, bloccando così l'avvio dell'ascensore ?

Deviazione Standard della Media

Abbiamo semplicemente definito che: $\boxed{\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}}}$

Deviazione standard campionaria

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

Deviazione standard della media:

$x_{best} = \bar{x} \pm \sigma_{\bar{x}}$

Dimostrazione:

Sia dato un campione di misure $\{x_1, x_2, \ldots . x_N\}$

Media $\quad \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + \cdots x_N}{N} = \frac{x_1}{N} + \frac{x_2}{N} + \cdots \frac{x_N}{N}$

Deviazione standard campionaria $\quad \sigma_x^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$ Stessa per tutti i valori

Ovvero $\quad x_1 \pm \sigma_x \;,\; x_2 \pm \sigma_x \;\ldots .\; x_N \pm \sigma_x$

Deviazione Standard della Media (II)

$$\sigma_{\bar{x}}^2 = \left(\frac{\partial \bar{x}}{\partial x_1}\right)^2 \sigma_x^2 + \left(\frac{\partial \bar{x}}{\partial x_2}\right)^2 \sigma_x^2 + \cdots . \left(\frac{\partial \bar{x}}{\partial x_N}\right)^2 \sigma_x^2 =$$

$$= \left(\frac{1}{N}\right)^2 \sigma_x^2 + \left(\frac{1}{N}\right)^2 \sigma_x^2 + \cdots . \left(\frac{1}{N}\right)^2 \sigma_x^2 = \frac{\sigma_x^2}{N^2} + \frac{\sigma_x^2}{N^2} + \cdots \frac{\sigma_x^2}{N^2}$$

$$= N \frac{\sigma_x^2}{N^2} = \frac{\sigma_x^2}{N}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} \qquad \text{(Cvd)}$$

Incertezze in Misure Dipendenti

Cosa succede se non si ha la certezza che le misure dirette siano indipendenti e/o non si è in grado di valutare la covarianza ?

Disuguaglianza di Schwarz: $|\sigma_{xy}| \leq \sigma_x \sigma_y$

Si può dimostrare che, in ogni caso, cioè anche per misure non indipendenti, risulta che:

$$\boxed{\sigma_q \leq \left|\frac{\partial q}{\partial x}\right| \sigma_x + \left|\frac{\partial q}{\partial y}\right| \sigma_y}$$

La **somma lineare** delle incertezze può essere utilizzata, nel caso di **misure correlate, come limite superiore dell'errore** al posto della somma in quadratura, che potrebbe sottostimare l'incertezza su q.

Incertezze in Misure Dipendenti (II)

$$\boldsymbol{\sigma_q \leq \left|\frac{\partial q}{\partial x}\right| \sigma_x + \left|\frac{\partial q}{\partial y}\right| \sigma_y}$$ Dimostrazione:

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)\sigma_{xy}$$

$$\leq \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)|\sigma_{xy}|$$

Covarianza può essere positiva o negativa $\Longrightarrow$ valore assoluto e segno di $\leq$

$$\leq \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2\left(\frac{\partial q}{\partial x}\right)\left(\frac{\partial q}{\partial y}\right)\sigma_x\sigma_y =$$

Disuguaglianza di Schwarz

$$= \left(\frac{\partial q}{\partial x}\sigma_x + \frac{\partial q}{\partial y}\sigma_y\right)^2$$

$$\sigma_q^2 \leq \left(\frac{\partial q}{\partial x}\sigma_x + \frac{\partial q}{\partial y}\sigma_y\right)^2 \Longrightarrow \sigma_q \leq \left|\frac{\partial q}{\partial x}\right|\sigma_x + \left|\frac{\partial q}{\partial y}\right|\sigma_y$$ (Cvd)

Combinazione Lineare

Un caso particolare importante della somma in quadratura delle incertezze è il seguente:

la combinazione lineare $q = c_1 x + c_2 y$ di due variabili gaussiane indipendenti x e y è, a sua volta, una variabile gaussiana di media $\mu = c_1 \mu_x + c_2 \mu_y$ e varianza $\sigma^2 = c_1^2 \sigma_x^2 + c_2^2 \sigma_y^2$ (dove c sono costanti).

Il risultato si può **generalizzare** ad una combinazione lineare di un numero qualsiasi di variabili gaussiane indipendenti.

Combinazione Lineare (II)

Dimostrazione

<u>Media</u>:

$$q = c_1 x + c_2 y \Longrightarrow q_i = c_1 x_i + c_2 y_i$$

$$\bar{q} = \frac{\sum_{i=1}^{N} q_i}{N} = \frac{\sum_{i=1}^{N}(c_1 x_i + c_2 y_i)}{N} = \frac{\sum_{i=1}^{N} \boxed{c_1} x_i + \sum_{i=1}^{N} \boxed{c_2} y_i}{N} =$$

$$= \frac{c_1 \sum_{i=1}^{N} x_i + c_2 \sum_{i=1}^{N} y_i}{N} = c_1 \underbrace{\boxed{\frac{\sum_{i=1}^{N} x_i}{N}}}_{= \bar{x}} + c_2 \underbrace{\boxed{\frac{\sum_{i=1}^{N} y_i}{N}}}_{= \bar{y}} = c_1 \bar{x} + c_2 \bar{y} \quad \text{(Cvd)}$$

Combinazione Lineare (II)

Varianza:

$$\sigma_q^2 = \frac{\sum_{i=1}^{N}(q_i - \bar{q})^2}{N-1} = \frac{\sum_{i=1}^{N}(q_i^2 + \bar{q}^2 - 2q_i\bar{q})}{N-1} =$$

$$= \frac{\sum_{i=1}^{N}[(c_1x_i + c_2y_i)^2 + (c_1\bar{x} + c_2\bar{y})^2 - 2(c_1x_i + c_2y_i)(c_1\bar{x} + c_2\bar{y})]}{N-1} =$$

$$= \frac{\sum_{i=1}^{N}\left[\begin{array}{l} c_1^2x_i^2 + c_2^2y_i^2 + 2c_1c_2x_iy_i + c_1^2\bar{x}^2 + c_2^2\bar{y}^2 + 2c_1c_2\bar{x}\bar{y} \\ \quad -2c_1^2x_i\bar{x} - 2c_1c_2x_i\bar{y} - 2c_1c_2y_i\bar{x} - 2c_2^2y_i\bar{y} \end{array}\right]}{N-1} =$$

$$= \frac{\sum_{i=1}^{N}[c_1^2(x_i^2 + \bar{x}^2 - 2x_i\bar{x})]}{N-1} + \frac{\sum_{i=1}^{N}[c_2^2(y_i^2 + \bar{y}^2 - 2y_i\bar{y})]}{N-1} +$$

$$+ \frac{\sum_{i=1}^{N}[2c_1c_2(x_iy_i + \bar{x}\bar{y} - x_i\bar{y} - y_i\bar{x}]}{N-1}$$

Combinazione Lineare (III)

$$\sigma_q^2 = \frac{\sum_{i=1}^{N}\left[\boxed{c_1^2}\left(x_i^2 + \bar{x}^2 - 2x_i\bar{x}\right)\right]}{N-1} + \frac{\sum_{i=1}^{N}\left[\boxed{c_2^2}\left(y_i^2 + \bar{y}^2 - 2y_i\bar{y}\right)\right]}{N-1} + \frac{\sum_{i=1}^{N}\left[\boxed{2c_1c_2}\left(x_iy_i + \bar{x}\bar{y} - x_i\bar{y} - y_i\bar{x}\right)\right]}{N-1}$$

$$= c_1^2 \underbrace{\boxed{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}}_{=\sigma_x^2} + c_2^2 \underbrace{\boxed{\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}}}_{=\sigma_y^2} + 2c_1c_2\frac{\sum_{i=1}^{N}\left[x_i(y_i - \bar{y}) - \bar{x}(y_i - \bar{y})\right]}{N-1}$$

$$= c_1^2\sigma_x^2 + c_2^2\sigma_y^2 + 2c_1c_2 \underbrace{\boxed{\frac{\sum_{i=1}^{N}\left[(x_i - \bar{x})(y_i - \bar{y})\right]}{N-1}}}_{=\sigma_{xy}} = c_1^2\sigma_x^2 + c_2^2\sigma_y^2 + 2c_1c_2\cancel{\sigma_{xy}}$$

Variabili indipendenti $\Longrightarrow$ covarianza nulla

$$\sigma_q^2 = c_1^2\sigma_x^2 + c_2^2\sigma_y^2 \qquad \text{(Cvd)}$$

Combinazione Lineare (II)

Varianza:
Allo stesso risultato si arrivava direttamente utilizzando la somma in quadratura nella propagazione delle incertezze (le variabili *x* e *y* sono per definizione indipendenti e caratterizzate solo da incertezze casuali).

$$q = c_1 x + c_2 y$$

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 = \left(\frac{\partial (c_1 x + c_2 y)}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q(c_1 x + c_2 y)}{\partial y}\right)^2 \sigma_y^2 =$$

$$= c_1^2 \sigma_x^2 + c_2^2 \sigma_y^2 \qquad \text{(Cvd)}$$

Esercizio 2

In blocchi metallici viene realizzato un foro cilindrico e un pistone cilindrico viene poi inserito nel foro. Il “gioco” tra foro e pistone è definito come la metà della differenza tra il diametro del foro e quello del pistone.
Il diametro del foro è una variabile distribuita normalmente con media 15.000 cm e deviazione standard 0.025 cm, mentre il diametro del pistone è una variabile distribuita normalmente con media 14.880 cm e deviazione standard 0.015 cm.
Determinare la probabilità che, prendendo a caso un blocco forato e un pistone, il gioco sia compreso tra 0.05 cm e 0.09 cm (condizione richiesta per il corretto funzionamento del sistema).

Table C.4. Values of the integral $\Phi^*(z) = \int_0^z \phi(z')\,dz'$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.9015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4776	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987									
3.5	0.4998									
4.0	0.4999									

Tavola della probabilità della curva normale standardizzata

Tavola delle probabilità cumulate per valori NEGATIVI di z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Coefficiente di Correlazione Lineare

- Situazione analoga (ma diversa) si ha quando una correlazione tra due grandezze fisiche è dovuta ad una **relazione funzionale**, in particolare di tipo lineare.

- In questi casi è utile introdurre il **coefficiente di correlazione lineare:**

$$\boxed{r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}}$$

- È un numero compreso tra -1 e $+1$; Dim: $|r| = \frac{|\sigma_{xy}|}{\sigma_x \sigma_y} \le \frac{\sigma_x \sigma_y}{\sigma_x \sigma_y} = 1$

- Può essere utilizzato per **controllare la validità della ipotesi di dipendenza lineare** tra le variabili x e y.

Coefficiente di Correlazione Lineare (II)

Il valore numerico di r fornisce, in funzione del numero N di coppie (x_i, y_i), la risposta alla domanda:

"se le due variabili x e y non sono legate da una relazione lineare, qual è la probabilità che le N misure forniscano un r uguale o maggiore di quello effettivamente ottenuto ?"

se **P < 5%** si dice che la correlazione è **significativa**

se **P < 1%** si dice che la correlazione è **altamente significativa**

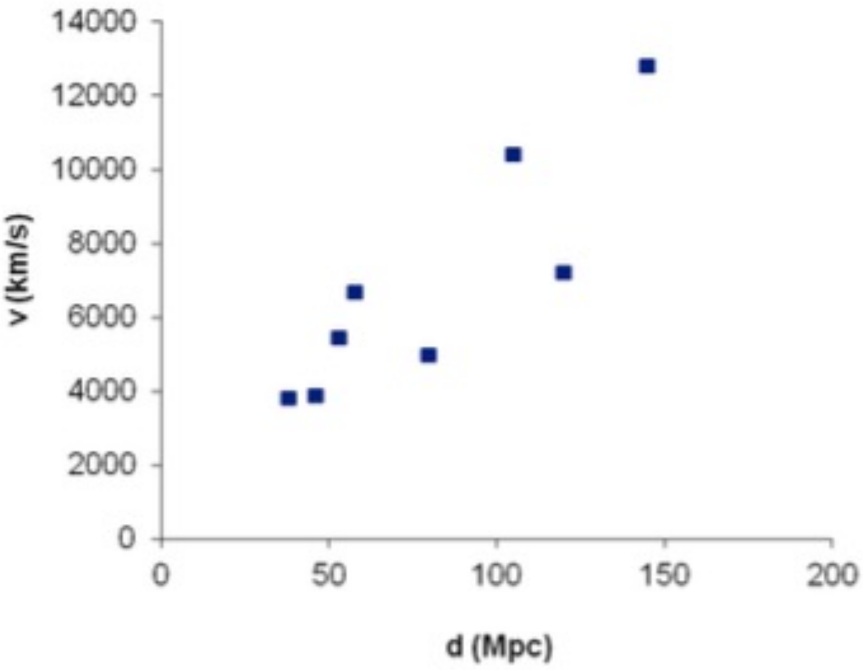
Table C.8. The table gives the percent probability that, for N observations, the value r for two completely uncorrelated variables is larger in absolute value than the observed value r_o: $P_N(|r| \geq |r_o|)$.

$\lvert r_o \rvert \rightarrow$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
N=3	100	94	87	81	74	67	59	51	41	29	0
N=4	100	90	80	70	60	50	40	30	20	10	0
N=5	100	87	75	62	50	39	28	19	10	3.7	0
N=6	100	85	70	56	43	31	21	12	5.6	1.4	0
N=7	100	83	67	51	37	25	15	8.0	3.1	0.6	0
N=8	100	81	63	47	33	21	12	5.3	1.7	0.2	0
N=9	100	80	61	43	29	17	8.8	3.6	1.0	0.1	0
N=10	100	78	58	40	25	14	6.7	2.4	0.5		0
N=11	100	77	56	37	22	12	5.1	1.6	0.3		0
N=12	100	76	53	34	20	9.8	3.9	1.1	0.2		0
N=13	100	75	51	32	18	8.2	3.0	0.8	0.1		0
N=14	100	73	49	30	16	6.9	2.3	0.5	0.1		0
N=15	100	72	47	28	14	5.8	1.8	0.4			0
N=16	100	71	46	26	12	4.9	1.4	0.3			0
N=17	100	70	44	24	11	4.1	1.1	0.2			0
N=18	100	69	43	23	10	3.5	0.8	0.1			0
N=19	100	68	41	21	9.0	2.9	0.7	0.1			0
N=20	100	67	40	20	8.1	2.5	0.5	0.1			0
N=25	100	63	34	15	4.8	1.1	0.2				0
N=30	100	60	29	11	2.9	0.5					0
N=35	100	57	25	8.0	1.7	0.2					0
N=40	100	54	22	6.0	1.1	0.1					0
N=45	100	51	19	4.5	0.6						0

Esercizio 3

Utilizzando i dati della tabella seguente, calcolare il coefficiente r delle variabili d e v, dove d rappresenta la distanza di una galassia e v la velocità di allontanamento.

d (Mpc)	v (km/s)
38	3810
46	3860
53	5430
80	4960
58	6660
105	10400
120	7200
145	12800



Esercizio 3

d	v	d-d_medio	v-v_medio	(d-d_medio)(v-v_+medio)	(d-d_medio)^2	(v-v_medio)^2		
38	3810	-42,625	-3080	131285	1816,890625	9486400		
46	3860	-34,625	-3030	104913,75	1198,890625	9180900		
53	5430	-27,625	-1460	40332,5	763,140625	2131600		
80	4960	-0,625	-1930	1206,25	0,390625	3724900		
58	6660	-22,625	-230	5203,75	511,890625	52900		
105	10400	24,375	3510	85556,25	594,140625	12320100		
120	7200	39,375	310	12206,25	1550,390625	96100		
145	12800	64,375	5910	380456,25	4144,140625	34928100		
media	media			somma	somma	somma	prodotto	radice
80,625	6890			761160	10579,875	71921000	7,60915E+11	872304,528
r	$\boxed{0,87258518}$							

$\lvert r_o \rvert \rightarrow$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$N=3$	100	94	87	81	74	67	59	51	41	29	0
$N=4$	100	90	80	70	60	50	40	30	20	10	0
$N=5$	100	87	75	62	50	39	28	19	10	3.7	0
$N=6$	100	85	70	56	43	31	21	12	5.6	1.4	0
$N=7$	100	83	67	51	37	25	15	8.0	3.1	0.6	0
$N=8$	100	81	63	47	33	21	12	5.3	1.7	$\boxed{0.2}$	0
$N=9$	100	80	61	43	29	17	8.8	3.6	1.0	0.1	0

Con 8 coppie di dati, la probabilità che variabili d e v indipendenti tra loro (cioè non legate da relazione lineare) forniscano un valore di r maggiore o uguale di 0,87 è inferiore a 0.2 %. Quindi la correlazione è **altamente significativa**.

Esercizio 4

Si studia la correlazione tra le temperature medie nei mesi di febbraio e di maggio in sei diverse città. I dati sono riassunti in tabella. Il coefficiente di correlazione lineare r_0 vale 0.94. Come cambierebbe il valore di r_0 se le temperature medie fossero misurate in gradi Fahrenheit (°F) ? (Per trasformare la temperatura da °C a °F si moltiplica per 1.8 e si somma 32). La correlazione tra le due temperature è significativa ?

	città 1	città 2	città 3	città 4	città 5	città 6
T febbraio (°C)	2	8	11	15	12	18
T maggio (°C)	8	10	14	17	18	23

	Febbraio °C	Maggio °C	febbraio -media °C	maggio - media °C	(febbraio-media)(maggio-media) °C	(febbraio -media)^2 °C	(maggio -media)^2 °C		
città 1	2	8	-9	-7	63	81	49		
città 2	8	10	-3	-5	15	9	25		
città 3	11	14	0	-1	0	0	1		
città 4	15	17	4	2	8	16	4		
città 5	12	18	1	3	3	1	9		
città 6	18	23	7	8	56	49	64		
					somma	somma	somma	prodotto	Radice
Media	11	15			145	156	152	23712	153,987012
r	0,94163785								

	Febbraio °F	Maggio °F	febbraio -media °F	maggio - media °F	(febbraio-media)(maggio-media) °F	(febbraio -media)^2 °C	(maggio -media)^2 °C		
città 1	35,6	46,4	-16,2	-12,6	204,12	262,44	158,76		
città 2	46,4	50	-5,4	-9	48,6	29,16	81		
città 3	51,8	57,2	0	-1,8	0	0	3,24		
città 4	59	62,6	7,2	3,6	25,92	51,84	12,96		
città 5	53,6	64,4	1,8	5,4	9,72	3,24	29,16		
città 6	64,4	73,4	12,6	14,4	181,44	158,76	207,36		
media	51,8	59			somma	somma	somma	prodotto	radice
					469,8	505,44	492,48	248919,0912	498,91792
r	0,94163785								

$\lvert r_o \rvert \rightarrow$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$N=3$	100	94	87	81	74	67	59	51	41	29	0
$N=4$	100	90	80	70	60	50	40	30	20	10	0
$N=5$	100	87	75	62	50	39	28	19	10	3.7	0
$N=6$	100	85	70	56	43	31	21	12	5.6	$\boxed{1.4}$	0
$N=7$	100	83	67	51	37	25	15	8.0	3.1	0.6	0
$N=8$	100	81	63	47	33	21	12	5.3	1.7	0.2	0
$N=9$	100	80	61	43	29	17	8.8	3.6	1.0	0.1	0

Con 6 coppie di dati, la probabilità che variabili indipendenti tra loro (cioè non legate da relazione lineare) forniscano un valore di r maggiore o uguale di 0,94 è superiore a 1.4 %. Quindi la correlazione è **significativa**.

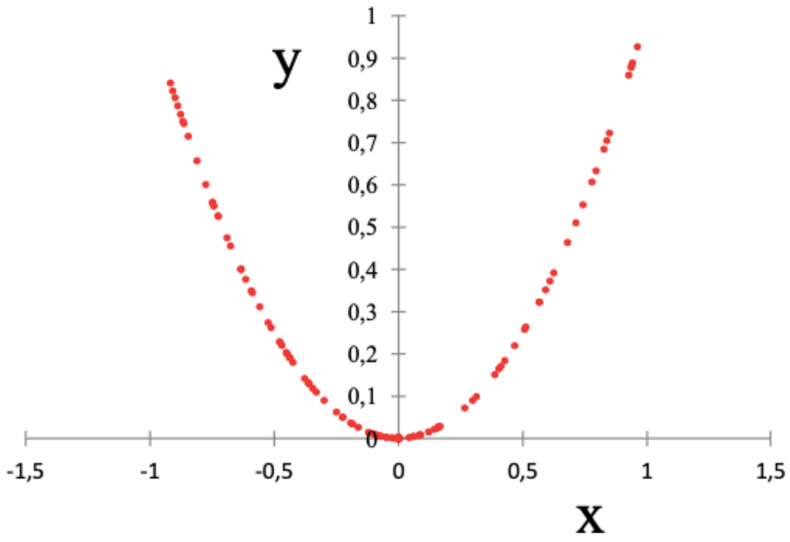
Non cambia nulla se si usano i °C o i °F: **il coefficiente di correlazione con cambia se si effettua una trasformazione lineare nella variabile x**

Esercizio 5

Una lampada è composta da due lampadine, una di tipo A e una di tipo B. Le lampadine A hanno una durata distribuita normalmente con media 800 ore e deviazione standard di 100 ore mentre le lampadine B hanno una durata distribuita normalmente con media 900 ore e deviazione standard di 150 ore. Le durate delle lampadine sono indipendenti. Qual è la probabilità che la lampadina B duri più della lampadina A, in una determinata lampada ?

Esercizio 6

Dimostrare che la covarianza di x (uniforme tra –1 e +1) e $y = x^2$ è nulla usando la identità $\sigma_{xy} = \overline{xy} - \overline{x}\,\overline{y}$



Esercizio 7

In un esperimento di diffrazione si misura la lunghezza d'onda di una luce monocromatica tramite la relazione $\lambda = d \sin(\theta)$ dove:

θ = angolo del massimo di diffrazione = $16.5° \pm 0.5°$ (l'incertezza è pari ad una deviazione standard della media)

d = passo del reticolo = (2.0 ± 0.1) μm (ottenuto da un campione di 25 misure).

Le misure di d e di θ sono indipendenti.

Quante misure di d sarebbero necessarie per rendere trascurabile il suo contributo alla incertezza di λ ?

Nota: imporre che il termine, nella propagazione, corrispondente a d sia almeno un ordine di grandezza inferiore a quello di θ.

Media Pesata di Misure Indipendenti

- Consideriamo due misure indipendenti della stessa grandezza X: $(x_1 \pm \sigma_1)$ e $(x_2 \pm \sigma_2)$.

- Sappiamo, da un test svolto in precedenza, che le misure sono compatibili tra loro, ma hanno incertezze diverse.

- La migliore stima di X si ottiene combinando le due misure con una **media pesata**

$$\boxed{\hat{X} = x_p = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2}}$$

Dove i **pesi** valgono $\boldsymbol{w_1 = \frac{1}{\sigma_1^2}}$ e $\boldsymbol{w_2 = \frac{1}{\sigma_2^2}}$

L'incertezza da associare a x_p è $\boxed{\sigma_{x_p} = \frac{1}{\sqrt{w_1 + w_2}}}$

Media Pesata di Misure Indipendenti (II)

Generalizzando a N misure indipendenti

$$\boxed{\hat{X} = x_p = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}}$$

Dove i **pesi** valgono $\boldsymbol{w_i = \frac{1}{\sigma_i^2}}$

L'incertezza da associare a x_p è

$$\boxed{\sigma_{x_p} = \frac{1}{\sqrt{\sum_{i=1}^{N} w_i}}}$$

Media Pesata di Misure Indipendenti (III)

Dimostrazione: <u>media</u>

Supponiamo di avere due misure: $x_1 \pm \sigma_1$ e $x_2 \pm \sigma_2$. Ogni misura quindi è governata da una distribuzione normale.
La probabilità di ottenere contemporaneamente le due misure risulta quindi:

$$P(x_{1,}x_2) = P(x_1)P(x_2) \propto \frac{1}{\sigma_1} e^{-\frac{(x_1-x)^2}{2\sigma_1^2}} \frac{1}{\sigma_2} e^{-\frac{(x_2-x)^2}{2\sigma_2^2}} \propto e^{-\frac{(x_1-x)^2}{2\sigma_1^2}-\frac{(x_2-x)^2}{2\sigma_2^2}}$$

Uso il criterio della massima verosimiglianza: massimizzo la probabilità minimizzando l'esponente.

$$\frac{d}{dx}\left(-\frac{(x_1-x)^2}{2\sigma_1^2}-\frac{(x_2-x)^2}{2\sigma_2^2}\right) = 0 \Rightarrow -\cancel{2}\frac{(x_1-x)(-1)}{\cancel{2}\sigma_1^2} - \cancel{2}\frac{(x_2-x)(-1)}{\cancel{2}\sigma_2^2} = 0$$

Media Pesata di Misure Indipendenti (IV)

$$\frac{(x_1 - x)}{\sigma_1^2} + \frac{(x_2 - x)}{\sigma_2^2} = 0 \Rightarrow \frac{x_1}{\sigma_1^2} - \frac{x}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} - \frac{x}{\sigma_2^2} = 0$$

$$x\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) = \frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} \Rightarrow \hat{x} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

Definisco pesi $w_1 = \frac{1}{\sigma_1^2}$ e $w_2 = \frac{1}{\sigma_2^2}$ allora

$$\hat{x} = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2}$$ Cvd

Media Pesata di Misure Indipendenti (V)

Dimostrazione: incertezza

$$\hat{x} = \frac{w_1 x_1 + w_2 x_2}{w_1 + w_2}$$

Utilizzo la somma in quadratura per ricavare l'incertezza

$$\sigma_{\hat{x}}^2 = \left(\frac{\partial \hat{x}}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial \hat{x}}{\partial x_2}\right)^2 \sigma_{x_2}^2 = \left(\frac{w_1}{w_1 + w_2}\right)^2 \sigma_{x_1}^2 + \left(\frac{w_2}{w_1 + w_2}\right)^2 \sigma_{x_2}^2 =$$

$$= \frac{w_1^2}{(w_1 + w_2)^2} \frac{1}{w_1} + \frac{w_2^2}{(w_1 + w_2)^2} \frac{1}{w_2} = \frac{w_1}{(w_1 + w_2)^2} + \frac{w_2}{(w_1 + w_2)^2} =$$

$$= \frac{w_1 + w_2}{(w_1 + w_2)^2} = \frac{1}{w_1 + w_2} \Longrightarrow \sigma_{\hat{x}} = \frac{1}{\sqrt{w_1 + w_2}} \qquad \text{Cvd}$$

Esercizio 8

Misuriamo la portata $P = \frac{dV}{dt}$ di un corso d'acqua in tre punti differenti, ottenendo i seguenti valori:

1. $(101 \pm 2)\ m^3/s$
2. $(97 \pm 1)\ m^3/s$
3. $(108 \pm 3)\ m^3/s$

Ipotizzando che la portata sia la stessa, cioè che il flusso sia stazionario, quale valore otteniamo, con una media pesata ?

Regressione Lineare

Dato un insieme di misure: x_i e $(y_i \pm \sigma_y)$ dove i = 1,...N, nella ipotesi che:

- gli errori sulle misure delle variabili x_i siano trascurabili;
- le grandezze y_i siano estratte da popolazioni gaussiane;
- gli errori sulle varie misure delle y siano tutti uguali;
- **le grandezze x e y siano legate da una relazione lineare:**
  y =A+ Bx,

quali sono le **migliori stime dei parametri A e B**?

Regressione Lineare (II)

Le migliori stime dei parametri si trovano col metodo della massima verosimiglianza:

$$\boxed{A = \frac{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i}{\Delta}}$$

Ordinata all'origine

$$\boxed{B = \frac{N \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\Delta}}$$

Coefficiente angolare

$$\Delta = N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2$$

Regressione Lineare (III)

Le incertezze sui parametri sono

$$\sigma_A = \sigma_y \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{\Delta}}$$

$$\sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}$$

Nel caso in cui le incertezze sulle y non siano conosciute, queste possono essere stimate tramite i parametri come:

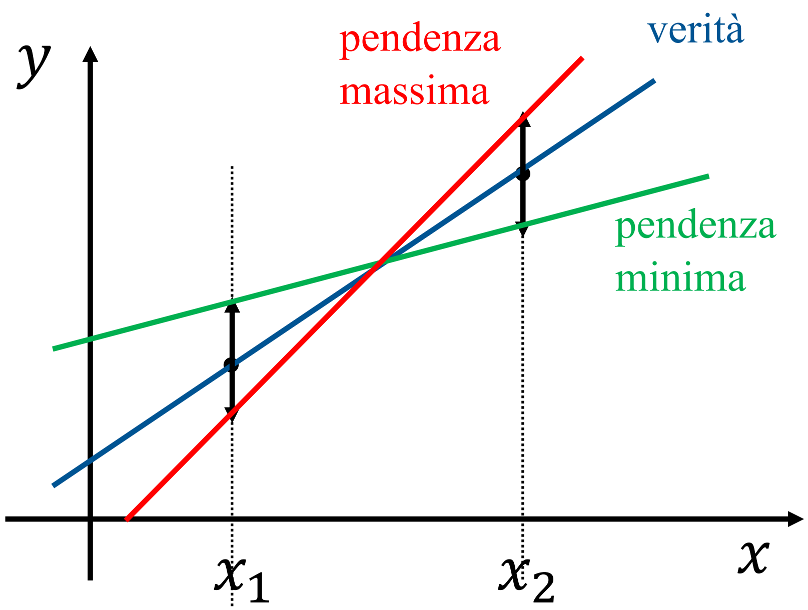
$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{N}(y_i - A - Bx_i)^2}{N-2}}$$

Regressione Lineare (IV)

Dimostrazione

Se conoscessimo le costanti A e B allora per ogni valore di x_i che abbiamo assunto non ha incertezza potremmo calcolare il valore vero di y_i come

$(valore\ vero\ y_i) = A + Bx_i$



Non conosciamo la verità. Abbiamo solo i punti sperimentali con le loro incertezze.

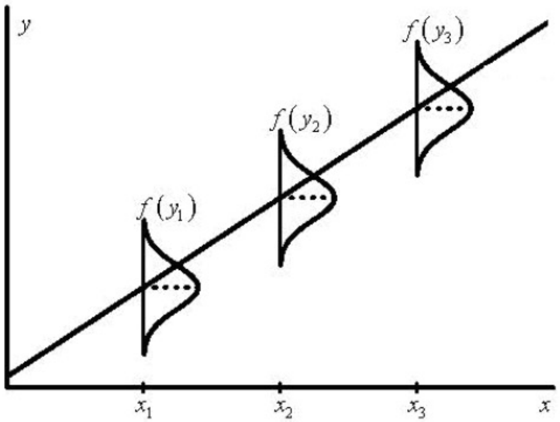
$$\hat{A} = \frac{A_{max} + A_{min}}{2} \pm \frac{A_{max} - A_{min}}{2}$$

$$\hat{B} = \frac{B_{max} + B_{min}}{2} \pm \frac{B_{max} - B_{min}}{2}$$

Regressione Lineare (V)

I punti sperimentali sono spostati rispetto al valore «vero» a causa di effetti casuali.

Ipotesi: ogni valore di y è governato da una distribuzione normale centrata sul valor vero di y_i e parametro di larghezza (errore sul valore vero) è l'incertezza σ_y ovvero la deviazione standard.



La probabilità di ottenere il valore osservato di y_i è

$$P_{A,B}(y_i) \propto \frac{1}{\sigma_y} e^{-\frac{(y_i - y_{i,vero})^2}{2\sigma_y^2}} \propto \frac{1}{\sigma_y} e^{-\frac{(y_i - A - Bx_i)^2}{2\sigma_y^2}}$$

La probabilità dipende dai due parametri A e B.

Regressione Lineare (VI)

La probabilità di ottenere l'intero campione di misure y_1, $y_2 \ldots$. y_N è quindi

$$P_{A,B}\left(y_{1,}y_{2,}\ldots .y_N\right) = P_{A,B}(y_1)P_{A,B}(y_2)\ldots .P_{A,B}(y_N)$$

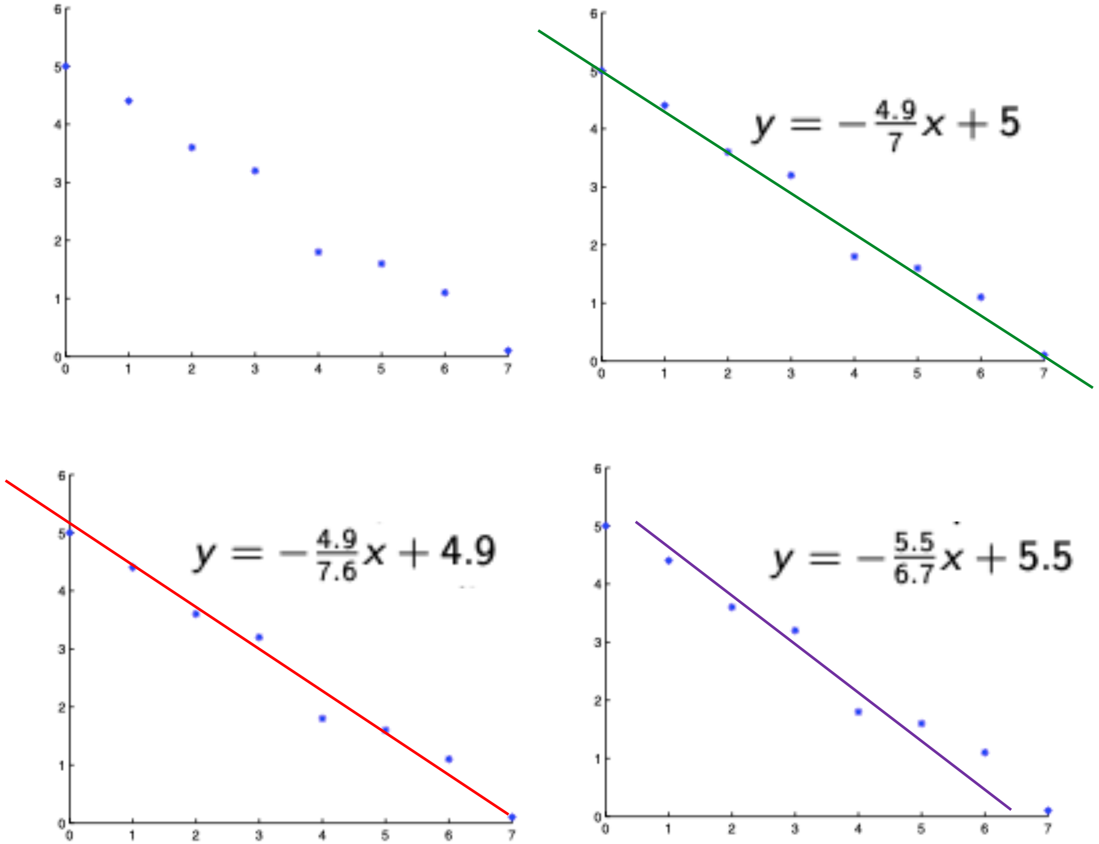
$$\propto \frac{1}{\sigma_y} e^{-\frac{\left(y_1 - y_{1,vero}\right)^2}{2\sigma_y^2}} \frac{1}{\sigma_y} e^{-\frac{(y_2 - A - Bx_2)^2}{2\sigma_y^2}} \ldots \frac{1}{\sigma_y} e^{-\frac{(y_N - A - Bx_N)^2}{2\sigma_y^2}} = \frac{1}{\sigma_y^N} e^{-\frac{\sum_{i=1}^N (y_i - A - Bx_i)^2}{2\sigma_y^2}}$$

Usiamo il criterio della massima verosimiglianza: la probabilità è massima quando l'esponenziale è minimo nei parametri A e B, ovvero

$$\frac{\partial\left(-\frac{\sum_{i=1}^N (y_i - A - Bx_i)^2}{2\sigma_y^2}\right)}{\partial A} = 0 \qquad \frac{\partial\left(-\frac{\sum_{i=1}^N (y_i - A - Bx_i)^2}{2\sigma_y^2}\right)}{\partial B} = 0$$

Adattamento dei **minimi quadrati**

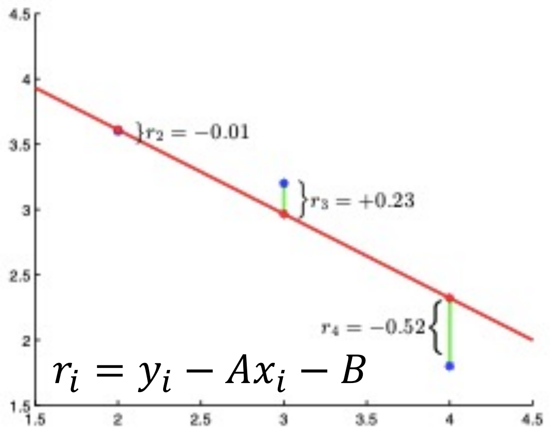
Minimi Quadrati

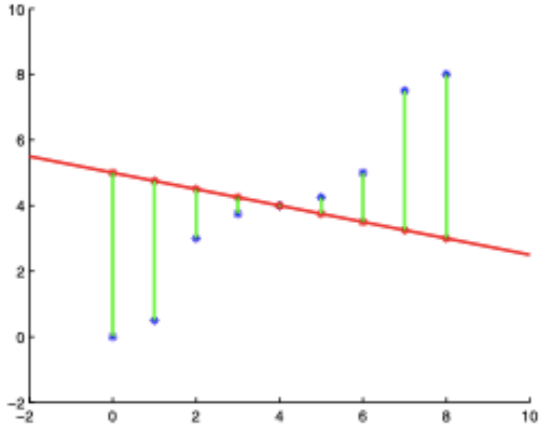


Quando si ha a che fare con dati sperimentali, dall'analisi dei grafici si ha una forte impressione che una relazione lineare debba esistere, ma è difficile determinare ad occhio la posizione della retta.

1) retta passante per il primo e l'ultimo dei punti,
2) retta che lascia sopra e sotto di sé un numero uguale di punti
3) Retta che dà più importanza ai punti al centro

Minimi Quadrati (II)





- Il **residuo r_i** rappresenta la **distanza (con segno) tra il dato (x_i,y_i) e il punto sulla retta (x_i , $Ax_i + B$)** corrispondente al valore x_i della variabile.

- Il residuo r_i ha valore positivo o negativo dipendentemente dal fatto che il punto plottato giace sopra o sotto la retta.

- In alcuni casi il residuo totale $\sum_{i=1}^{N} r_i = 0$

- Utilizzare i quadrati evita l'azzerarsi dei termini nella somma.

- **Se la somma dei quadrati dei residui è minima, allora ogni residuo r_i è vicino a zero** per la proprietà dei numeri reali positivi $\Longrightarrow$ **minimi quadrati.**

Regressione Lineare (VII)

$$
\begin{cases}
\dfrac{\partial\left(-\dfrac{\sum_{i=1}^{N}(y_i - A - Bx_i)^2}{2\sigma_y^2}\right)}{\partial A} = 0 \Rightarrow +\cancel{2}\dfrac{\sum_{i=1}^{N}(y_i - A - Bx_i)(+1)}{\cancel{2}\cancel{\sigma_y^2}} = 0 \\[2ex]
\dfrac{\partial\left(-\dfrac{\sum_{i=1}^{N}(y_i - A - Bx_i)^2}{2\sigma_y^2}\right)}{\partial B} = 0 \Rightarrow +\cancel{2}\dfrac{\sum_{i=1}^{N}(y_i - A - Bx_i)(+x_i)}{\cancel{2}\cancel{\sigma_y^2}} = 0
\end{cases}
$$

$$
\begin{cases}
\sum_{i=1}^{N}(y_i - A - Bx_i) = 0 \\
\sum_{i=1}^{N}x_i(y_i - A - Bx_i) = 0
\end{cases}
\qquad
\begin{cases}
\sum_{i=1}^{N}y_i - \sum_{i=1}^{N}A - \sum_{i=1}^{N}Bx_i = 0 \\
\sum_{i=1}^{N}x_iy_i - \sum_{i=1}^{N}Ax_i - \sum_{i=1}^{N}Bx_i^2 = 0
\end{cases}
$$

Regressione Lineare (VIII)

$$\begin{cases} \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} \boxed{A} - \sum_{i=1}^{N} \boxed{B} x_i = 0 \\ \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} \boxed{A} x_i - \sum_{i=1}^{N} \boxed{B} x_i^2 = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{N} y_i - A \sum_{i=1}^{N} 1 - B \sum_{i=1}^{N} x_i = 0 \\ \sum_{i=1}^{N} x_i y_i - A \sum_{i=1}^{N} x_i - B \sum_{i=1}^{N} x_i^2 = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{N} y_i - AN - B \sum_{i=1}^{N} x_i = 0 \\ \sum_{i=1}^{N} x_i y_i - A \sum_{i=1}^{N} x_i - B \sum_{i=1}^{N} x_i^2 = 0 \end{cases}$$

$$\begin{cases} AN + B \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i \\ A \sum_{i=1}^{N} x_i + B \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i \end{cases}$$

Regressione Lineare (IX)

Stima del parametro A

$$\sum_{i=1}^{N} x_i^2 \quad \begin{cases} AN + B\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i \\ A\sum_{i=1}^{N} x_i + B\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i \end{cases} \quad \sum_{i=1}^{N} x_i$$

$$\begin{cases} AN\sum_{i=1}^{N} x_i^2 + \cancel{B\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} x_i} = \sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i \\ A\left(\sum_{i=1}^{N} x_i\right)^2 + \cancel{B\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} x_i} = \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i \end{cases} \quad -$$

$$AN\sum_{i=1}^{N} x_i^2 - A\left(\sum_{i=1}^{N} x_i\right)^2 = \sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i$$

Regressione Lineare (X)

$$AN\sum_{i=1}^{N}x_i^2 - A\left(\sum_{i=1}^{N}x_i\right)^2 = \sum_{i=1}^{N}x_i^2\sum_{i=1}^{N}y_i - \sum_{i=1}^{N}x_i\sum_{i=1}^{N}x_iy_i$$

$$A\left(N\sum_{i=1}^{N}x_i^2 - \left(\sum_{i=1}^{N}x_i\right)^2\right) = \sum_{i=1}^{N}x_i^2\sum_{i=1}^{N}y_i - \sum_{i=1}^{N}x_i\sum_{i=1}^{N}x_iy_i$$

$$\hat{A} = \frac{\sum_{i=1}^{N}x_i^2\sum_{i=1}^{N}y_i - \sum_{i=1}^{N}x_i\sum_{i=1}^{N}x_iy_i}{N\sum_{i=1}^{N}x_i^2 - \left(\sum_{i=1}^{N}x_i\right)^2}$$ Cvd

Regressione Lineare (XI)

Stima del parametro B

$$\sum_{i=1}^{N} x_i \quad \begin{cases} AN + B\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i \\ A\sum_{i=1}^{N} x_i + B\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i \end{cases}$$

$$N$$

$$\begin{cases} \cancel{AN\sum_{i=1}^{N} x_i} + B\left(\sum_{i=1}^{N} x_i\right)^2 = \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i \\ \cancel{AN\sum_{i=1}^{N} x_i} + NB\sum_{i=1}^{N} x_i^2 = N\sum_{i=1}^{N} x_i y_i \end{cases} \quad -$$

$$B\left(\sum_{i=1}^{N} x_i\right)^2 - NB\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i - N\sum_{i=1}^{N} x_i y_i$$

Regressione Lineare (XII)

$$B\left(\sum_{i=1}^{N} x_i\right)^2 - NB\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i - N\sum_{i=1}^{N} x_i y_i$$

$$B\left(\left(\sum_{i=1}^{N} x_i\right)^2 - N\sum_{i=1}^{N} x_i^2\right) = \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i - N\sum_{i=1}^{N} x_i y_i$$

$$B\left(-\left(\sum_{i=1}^{N} x_i\right)^2 + N\sum_{i=1}^{N} x_i^2\right) = -\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i + N\sum_{i=1}^{N} x_i y_i$$

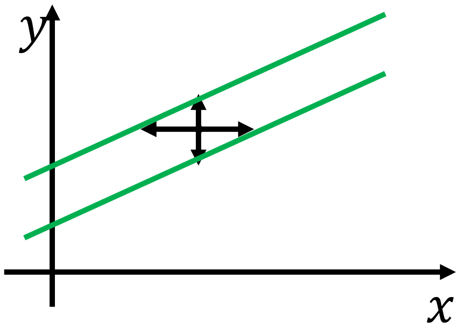
$$B = \frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$ Cvd

Regressione Lineare (XIII)

Usando la somma in quadratura:

$$\sigma_y^2 = \left(\frac{\partial y}{\partial x}\right)^2 \sigma_x^2 \Rightarrow \sigma_y = \frac{dy}{dx}\sigma_x = B\sigma_x$$

Essendo B fissato, è fissata anche la pendenza della retta. Si ha quindi un fascio di rette parallele.



$$y = A + Bx$$

$$\bar{y} = A + B\bar{x}$$

$$y - \bar{y} = B(x - \bar{x})$$

Coefficiente di correlazione lineare: $r = \dfrac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2\sum_{i=1}^{N}(y_i-\bar{y})^2}} =$

$$r = \frac{\sum_{i=1}^{N}(x_i-\bar{x})\boxed{B}(x_i-\bar{x})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2\sum_{i=1}^{N}\boxed{B^2}(x_i-\bar{x})^2}} = \frac{B\,\cancel{\sum_{i=1}^{N}(x_i-\bar{x})^2}}{|B|\cancel{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^4}}} = \frac{B}{|B|}$$

$r = 1$
$B > 0$

$r = -1$
$B < 0$

Regressione Lineare (XIV)

Nel caso in cui **gli errori sulle varie misure delle y non siano tutti uguali** si effettua una regressione lineare analoga a quella già vista, ma **pesando** i vari punti in base al valore di σ_{yi}

Si definisce quindi il **peso w_i** come $\boldsymbol{w_i = \frac{1}{\sigma_{yi}^2}}$

$$A = \frac{\sum_{i=1}^{N} w_i x_i^2 \sum_{i=1}^{N} w_i y_i - \sum_{i=1}^{N} w_i x_i \sum_{i=1}^{N} w_i x_i y_i}{\Delta} \qquad \sigma_A = \sigma_y \sqrt{\frac{\sum_{i=1}^{N} w_i x_i^2}{\Delta}}$$

$$B = \frac{\sum_{i=1}^{N} w_i \sum_{i=1}^{N} w_i x_i y_i - \sum_{i=1}^{N} w_i x_i \sum_{i=1}^{N} w_i y_i}{\Delta} \qquad \sigma_B = \sigma_y \sqrt{\frac{\sum_{i=1}^{N} w_i}{\Delta}}$$

$$\Delta = \sum_{i=1}^{N} w_i \sum_{i=1}^{N} w_i \boldsymbol{x_i^2} - \left(\sum_{i=1}^{N} w_i \boldsymbol{x_i} \right)^2$$

Regressione Lineare (XV)

Nel caso in cui **gli errori sulle varie misure delle x $\boldsymbol{\sigma_x}$ non siano trascurabili** ci si riconduce al caso precedente

1) Si introduce un errore equivalente sulla y $\boldsymbol{\sigma_{y,equivalente}}$

$$\sigma_{y,equivalente} = \frac{dy}{dx}\sigma_x = B\sigma_x$$

2) si sommano in quadratura $\sigma_{y,}$ e $\sigma_{y,equivalente}$

$$\boldsymbol{\sigma_{y,tot} = \sqrt{(B\sigma_x)^2 + \sigma_y^2}}$$

Regressione Lineare (XVI)

Si possono linearizzare alcune funzioni non lineari per poi procedere con regressioni lineari per ottenere i parametri della funzione.

Esempio importante è il caso di una dipendenza esponenziale di y da x.

$y(x) = y_0 e^{-\mu x}$ può essere linearizzata facendone il logaritmo:

$$\underbrace{ln\big(y(x)\big)}_{y} = ln(y_0 e^{-\mu x}) = lny_0 + ln(e^{-\mu x}) = \underbrace{lny_0}_{A} \underbrace{- \mu x}_{B}$$

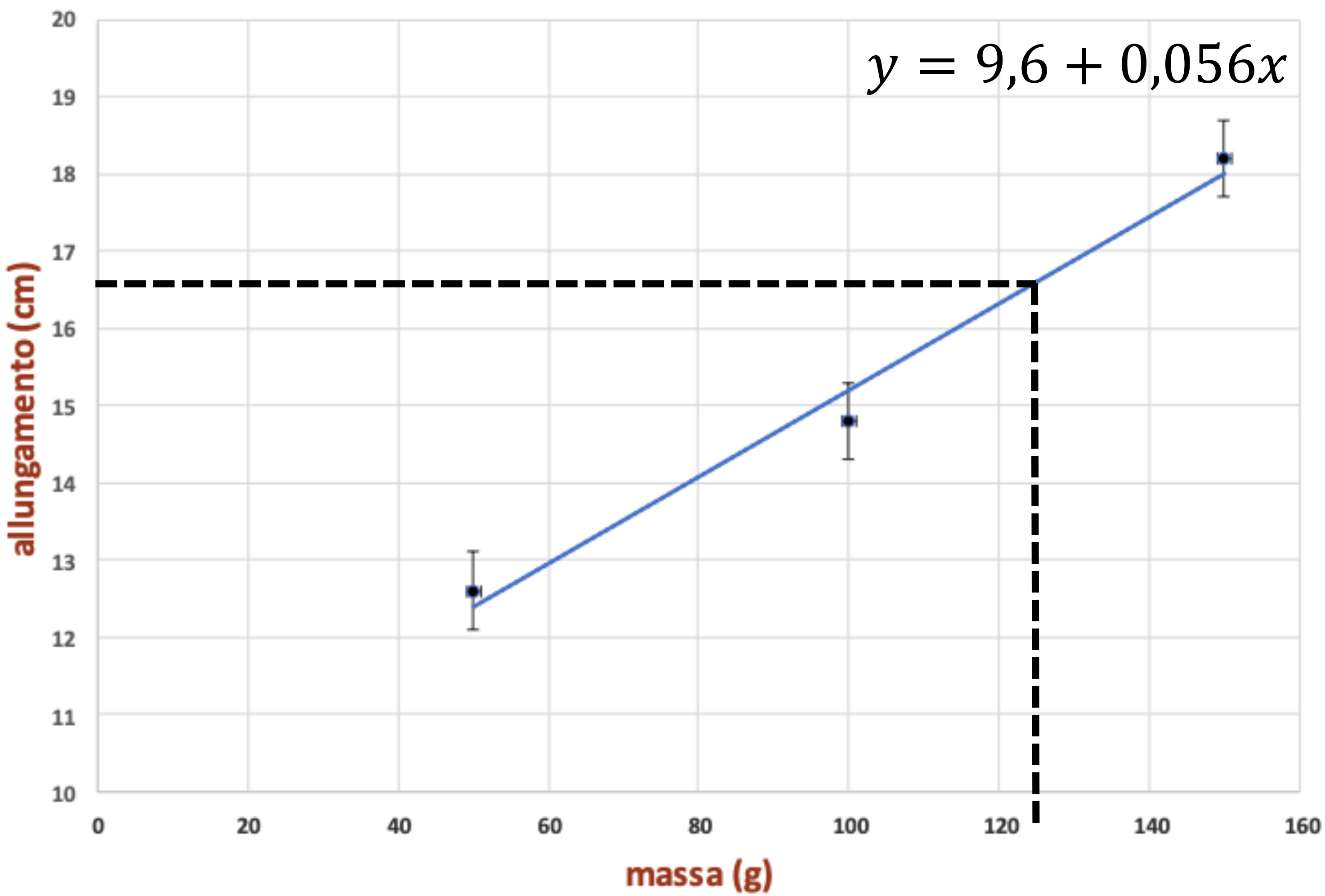
Esercizio 9

Stiamo calibrando un dinamometro utilizzando masse campione la cui incertezza è trascurabile. Otteniamo le misure seguenti:

1) $m = 50$ g $\quad l = (12.6 \pm 0.5)$ cm
2) $m = 100$ g $\quad l = (14.8 \pm 0.5)$ cm
3) $m = 150$ g $\quad l = (18.2 \pm 0.5)$ cm

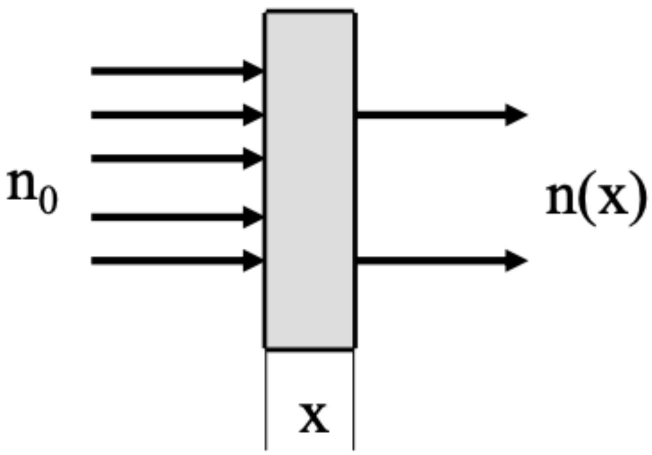
A calibrazione terminata, a quale massa incognita corrisponderà una posizione dell'indice del dinamometro pari a $l = 16.6$ cm ?

Calibrazione Dinamometro



Esercizio 10

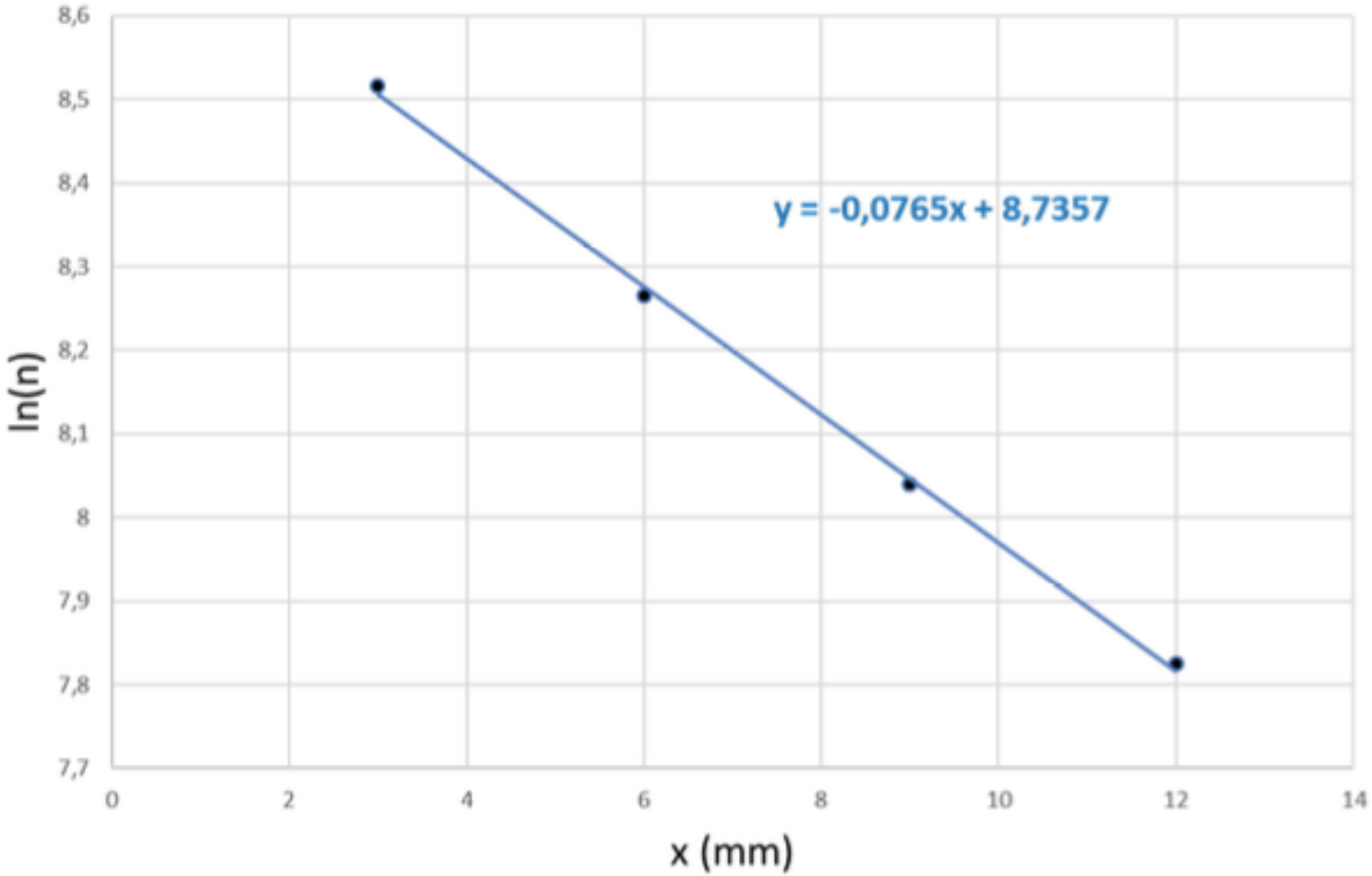
Una sorgente radioattiva emette raggi γ. Tra la sorgente e un rivelatore di radiazione (contatore Geiger) vengono poste lamine metalliche di vario spessore x. Effettuando conteggi su tempi uguali, si ottengono i dati riassunti nella tabella seguente, dove n rappresenta il numero di conteggi.



x (mm)	n(x)
3	4992
6	3887
9	3103
12	2503

Sapendo che la legge che descrive l'attenuazione della radiazione da parte delle lamine assorbitrici è: $n(x) = n_0 e^{-\mu x}$, qual è la migliore stima del parametro μ, che caratterizza il materiale dell'assorbitore ?

Legge di Attenuazione



Esercizio 11

Consideriamo un campione di misure di coppie (x_i,y_i).
Si ipotizza che la dipendenza sia del tipo $\boxed{y = Bx}$ (retta passante per l'origine). Quale stima del parametro B si ottiene in base al principio di massima verosimiglianza, nella ipotesi di incertezze trascurabili sulle x_i e tutte uguali sulle y_i ?

Risposta: $$\hat{B} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

La cui incertezza risulta $$\sigma_B = \frac{\sigma_y}{\sqrt{\sum_{i=1}^{N} x_i^2}}$$

Esercizio 12

Si misurano quattro coppie di valori x e y; il campione è riassunto nella tabella ($\Delta x = 0$):

x (cm/s)	y (g/s²)	Δy (g/s²)
0,93	8,1	0,5
1,16	11,5	0,5
1,62	14,4	0,5
3,48	32,2	0,5

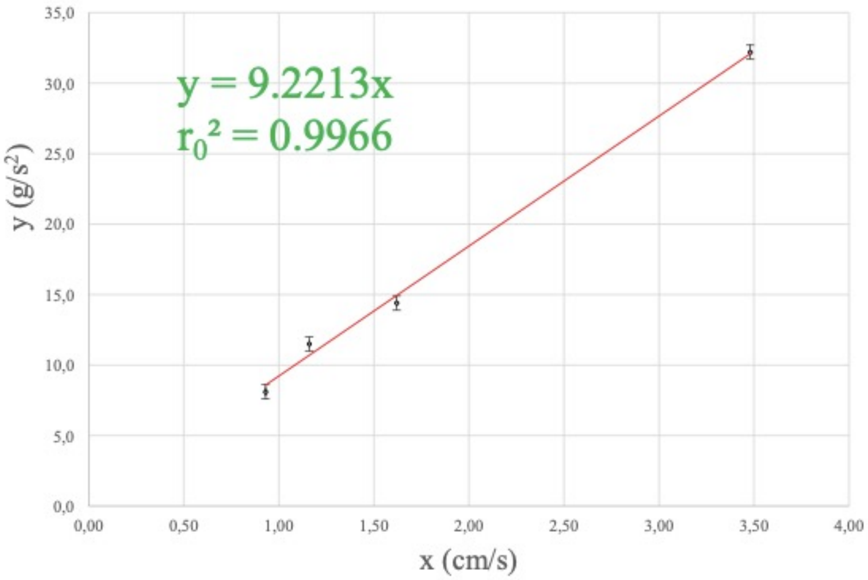
Calcolare il coefficiente di correlazione lineare tra x e y e verificare che l'ipotesi di dipendenza lineare è corretta.
Calcolare poi il coefficiente angolare B e la sua incertezza, nella ipotesi che la dipendenza lineare sia del tipo $y = Bx$.
Il coefficiente B rappresenta la viscosità η di un liquido: esprimere il risultato della misura di η nella unità del SI.

Esercizio 12

$\lvert r_o \rvert \rightarrow$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
N=3	100	94	87	81	74	67	59	51	41	29	0
~~N=4~~	~~100~~	~~90~~	~~80~~	~~70~~	~~60~~	~~50~~	~~40~~	~~30~~	~~20~~	10	0
N=5	100	87	75	62	50	39	28	19	10	3.7	0
N=6	100	85	70	56	43	31	21	12	5.6	1.4	0
N=7	100	83	67	51	37	25	15	8.0	3.1	0.6	0
N=8	100	81	63	47	33	21	12	5.3	1.7	0.2	0
N=9	100	80	61	43	29	17	8.8	3.6	1.0	0.1	0

$r = 0{,}9983$

$N = 4$



Esercizio 13

Effettuando una regressione lineare sui seguenti 4 punti

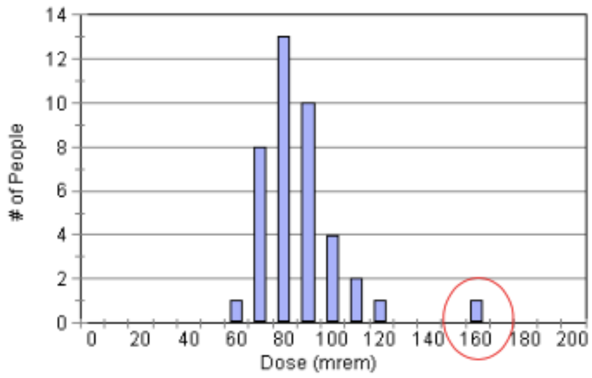
x	y
1.9	37.2
4.1	52.8
6.0	71.0
8.1	85.8

Si ottengono i parametri A = 21.5597 e B = 7.9881 (Δ = 84.11).

Quali incertezze σ_A e σ_B possono essere associate ai parametri ?

Rigetto di Dati in un Campione

Può accadere che una misura appartenente ad un campione di dati gaussiano si discosti in modo evidente dalle altre.



Il criterio di Chauvenet può essere utilizzato per rigettare misure "sospette" di questo tipo:

"Se il numero atteso di misure improbabili (*) almeno quanto la misura sospetta è minore di 0.5, allora la misura sospetta può essere rigettata"

(*) misure che distano dal valore medio più di z deviazioni standard, con $z = \frac{(x_{sospetto} - \bar{x})}{\sigma_x}$

Esercizio 14

In tabella sono presentate 40 misure relative a dosi annuali di radiazioni γ assorbite da abitanti di una città degli USA.
Il dato evidenziato può essere rigettato in base al criterio di Chauvenet ?

Dose annuale (mrem)			
78	76	116	85
84	94	86	73
88	101	91	94
86	83	85	166
90	88	79	70
100	94	85	83
62	102	79	95
88	91	90	116
122	84	103	73
89	90	72	91

Tabella B. La probabilità percentuale, $Q(t) = \int_X^{X+t\sigma} G_{X,\sigma}(x)dx$, come funzione di t.

t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00	0.40	0.80	1.20	1.60	1.99	2.39	2.79	3.19	3.59
0.1	3.98	4.38	4.78	5.17	5.57	5.96	6.36	6.75	7.14	7.53
0.2	7.93	8.32	8.71	9.10	9.48	9.87	10.26	10.64	11.03	11.41
0.3	11.79	12.17	12.55	12.93	13.31	13.68	14.06	14.43	14.80	15.17
0.4	15.54	15.91	16.28	16.64	17.00	17.36	17.72	18.08	18.44	18.79
0.5	19.15	19.50	19.85	20.19	20.54	20.88	21.23	21.57	21.90	22.24
0.6	22.57	22.91	23.24	23.57	23.89	24.22	24.54	24.86	25.17	25.49
0.7	25.80	26.11	26.42	26.73	27.04	27.34	27.64	27.94	28.23	28.52
0.8	28.81	29.10	29.39	29.67	29.95	30.23	30.51	30.78	31.06	31.33
0.9	31.59	31.86	32.12	32.38	32.64	32.89	33.15	33.40	33.65	33.89
1.0	34.13	34.38	34.61	34.85	35.08	35.31	35.54	35.77	35.99	36.21
1.1	36.43	36.65	36.86	37.08	37.29	37.49	37.70	37.90	38.10	38.30
1.2	38.49	38.69	38.88	39.07	39.25	39.44	39.62	39.80	39.97	40.15
1.3	40.32	40.49	40.66	40.82	40.99	41.15	41.31	41.47	41.62	41.77
1.4	41.92	42.07	42.22	42.36	42.51	42.65	42.79	42.92	43.06	43.19
1.5	43.32	43.45	43.57	43.70	43.82	43.94	44.06	44.18	44.29	44.41
1.6	44.52	44.63	44.74	44.84	44.95	45.05	45.15	45.25	45.35	45.45
1.7	45.54	45.64	45.73	45.82	45.91	45.99	46.08	46.16	46.25	46.33
1.8	46.41	46.49	46.56	46.64	46.71	46.78	46.86	46.93	46.99	47.06
1.9	47.13	47.19	47.26	47.32	47.38	47.44	47.50	47.56	47.61	47.67
2.0	47.72	47.78	47.83	47.88	47.93	47.98	48.03	48.08	48.12	48.17
2.1	48.21	48.26	48.30	48.34	48.38	48.42	48.46	48.50	48.54	48.57
2.2	48.61	48.64	48.68	48.71	48.75	48.78	48.81	48.84	48.87	48.90
2.3	48.93	48.96	49.98	49.01	49.04	49.06	49.09	49.11	49.13	49.16
2.4	49.18	49.20	49.22	49.25	49.27	49.29	49.31	49.32	49.34	49.36
2.5	49.38	49.40	49.41	49.43	49.45	49.46	49.48	49.49	49.51	49.52
2.6	49.53	49.55	49.56	49.57	49.59	49.60	49.61	49.62	49.63	49.64
2.7	49.65	49.66	49.67	49.68	49.69	49.70	49.71	49.72	49.73	49.74
2.8	49.74	49.75	49.76	49.77	49.77	49.78	49.79	49.79	49.80	49.81
2.9	49.81	49.82	49.82	49.83	49.84	49.84	49.85	49.85	49.86	49.86
3.0	49.87									
3.5	49.98									
4.0	49.997									
4.5	49.9997									
5.0	49.99997									

Esercizio 14

Esercizio 15

Si misura la pressione sistolica di 200 persone. Il valore medio è $p = (120 \pm 4)mmHg$. Una misura di 200 mmHg sarebbe da rigettare secondo il criterio di Chauvenet?

Esercizio 16

Consideriamo un campione di N = 60 misure di intervalli di tempo, estratto da una popolazione gaussiana.
La media è $\bar{t} = 7.20$ s e la dev. standard è $\sigma_t = 1.3$ s.
Nel campione è presente una misura sospetta $t_S = 10.8$ s.
Verificare che t_S può essere rigettata, in base al criterio di Chauvenet, e ricalcolare la media con le restanti 59 misure.

Esercizio 17

Per misurare la distanza focale f di una lente convergente si misura un campione di distanze oggetto-lente, d, e di corrispondenti distanze lente-immagine, r. I valori medi di d e r sono rispettivamente 2 cm e 10 cm. Le deviazioni standard delle medie sono entrambe uguali a 0.12 cm. La relazione che lega le tre grandezze è:

$$\frac{1}{f} = \frac{1}{d} + \frac{1}{r}$$

Determinare la distanza focale e la sua incertezza.