

Errori di misura: Distribuzioni

Ing. Michela Gelfusa

Email: gelfusa@ing.uniroma2.it

Ufficio 3° Piano – Ingegneria Industriale

Tel.: 06 7259 7210

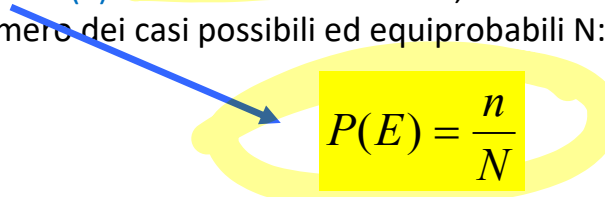
Probabilità, pdf, Gaussiana

Introduciamo il concetto di
probabilità in maniera
empirica ed intuitiva.

Consideriamo un dato sistema, in cui il verificarsi di un **evento E** presenta **n casi favorevoli** su un numero **N di casi possibili** ed equiprobabili:

$$\begin{cases} n & \text{casi favorevoli} \\ N & \text{casi possibili} \end{cases}$$

Possiamo definire la **probabilità $P(E)$** che l'evento si verifichi, come il rapporto fra il numero dei casi favorevoli n , ed il numero dei casi possibili ed equiprobabili N :


$$P(E) = \frac{n}{N}$$

Probabilità, pdf, Gaussiana

Per come è stata definita, la probabilità comporta la conoscenza a priori del sistema “**probabilità a priori**” e ciò non sempre è possibile. Inoltre nella definizione si introduce il concetto di eventi equiprobabili utilizzando di conseguenza il concetto di probabilità che si vorrebbe definire.

In generale, però, la probabilità non è nota a priori.

Supponiamo quindi di eseguire un numero N di prove di un certo sistema, e verifichiamo che un dato evento E si verifica n volte. Si definisce frequenza il rapporto tra il numero di prove in cui l'evento si è verificato ed il numero totale di prove eseguite. La frequenza n/N è anche detta probabilità empirica o “**probabilità a posteriori**”.

$$FREQ = \frac{\# \text{ PROVE IN CUI EVENTO SI VERIFICA}}{\# \text{ PROVE FATTE}}$$

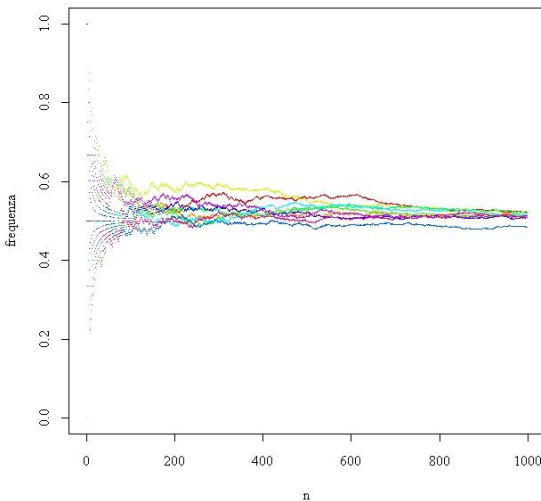
La legge empirica del caso stabilisce che se si conosce a priori la probabilità che un dato evento si verifichi, la frequenza tenderà alla probabilità al crescere delle prove eseguite.

Frequenza $\xrightarrow{N \rightarrow \infty}$ Probabilità

Probabilità, pdf, Gaussiana

Il valore fluttuerà di conseguenza intorno ad un valore ben determinato che è la probabilità a priori e tenderà a stabilizzarsi col crescere delle prove.

Quindi quanto più grande è il numero N di prove, più la frequenza (o probabilità a posteriori) tende alla probabilità a priori.



$$P(E) = \frac{n}{N} \quad \text{per } N \rightarrow \infty$$

Probabilità, pdf, Gaussiana

Definiamo probabilità di un evento E un numero reale $P(E)$ che soddisfi i **3 assiomi** seguenti:

- 1) La probabilità di un dato evento è sempre positivo o nullo: $P(E) \geq 0$
- 2) La somma delle probabilità di tutti gli eventi possibili è uguale ad 1.
- 3) Se indichiamo con E_1 ed E_2 due eventi *mutuamente esclusivi*, la probabilità che si verifichi o l'evento E_1 o l'evento E_2 è pari alla **somma** delle loro probabilità.

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) \quad \text{con } E_1 \cap E_2 = \emptyset.$$

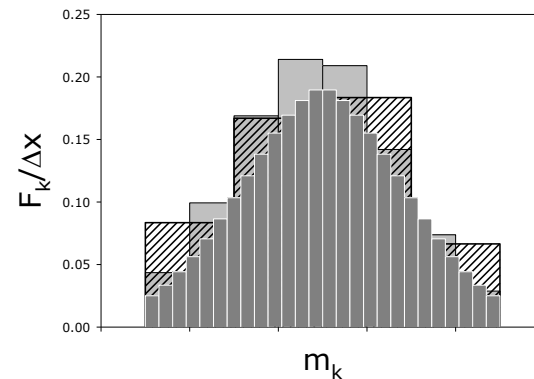
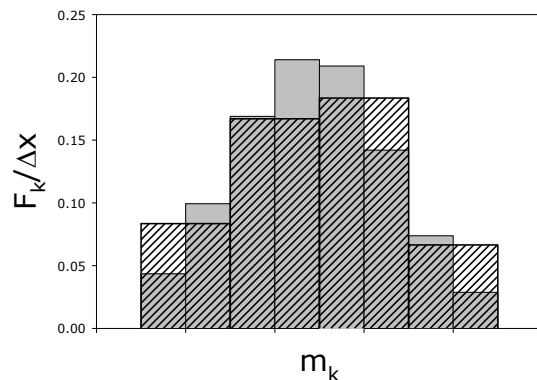
La definizione assiomatica di **probabilità** non fornisce alcuna indicazione di come la probabilità $P(E)$ associata ad un dato evento E debba essere valutata. Tale valutazione deve essere ricercata in altri ambiti; come ad esempio la ricerca della frequenza relativa facendo un numero di prove sufficientemente elevato o ove possibile attraverso la probabilità a priori.

La funzione di densità di probabilità

Ritorniamo ora all'esempio delle variabili distribuite con continuità, e consideriamo l'ipotesi in cui gli errori siano esclusivamente di tipo casuale.

Se si aumentano le misure, si verifica che la distribuzione delle stesse attorno al valor medio assume una forma maggiormente simmetrica.

Aumentare il numero delle misure, permette anche di ridurre la dimensione del singolo intervallo scelto per la costruzione dell'istogramma, quindi di ottenere una informazione maggiormente puntuale sulla forma di questa curva.



La funzione di densità di probabilità

Nell'ipotesi limite di un **numero infinito** di misure, potremmo idealmente far tendere a zero la larghezza dell'intervallo, ottenendo una **informazione puntuale** sulla forma della curva, e sostituire la serie di valori con una funzione vera e propria, chiamata **funzione di densità di probabilità** che indichiamo con **$f(x)$** , ove x rappresenta la variabile misurata.

Se questa funzione deve descrivere una distribuzione di probabilità, allora deve valere:

la **condizione di normalizzazione**, che in questo caso si scrive come:

$$\sum_{k=1}^M F_k = 1 \quad \xrightarrow{M \rightarrow \infty} \quad \int_{-\infty}^{+\infty} f(x) dx = 1$$

Abbiamo in pratica sostituito la **frequenza** con la **probabilità** (area di altezza $f(x)$ e larghezza dx).

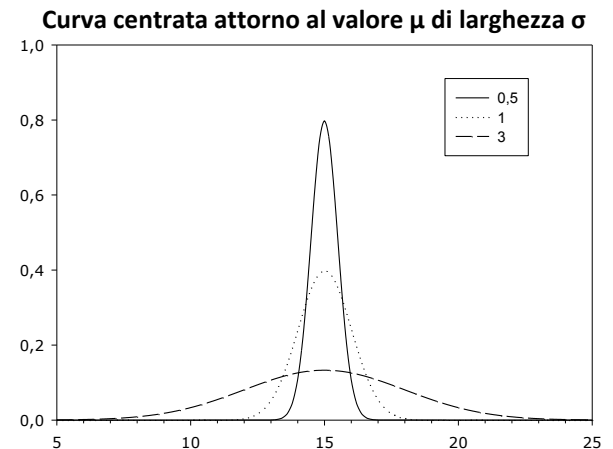


La funzione di densità di probabilità

Nel caso di misure con **errori casuali**, si può dimostrare che la distribuzione di probabilità assume la caratteristica forma a “campana” detta **distribuzione gaussiana o distribuzione normale** espressa mediante la seguente espressione:

$$G(x) = f_{\mu, \sigma}(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Fattore di normalizzazione, che garantisce l'integrale della curva sia uguale a 1



La funzione di densità di probabilità

Qual è il significato di questa curva nel caso della conduzione di misure con errore casuale?

La curva $f(x) \cdot dx$ dà la probabilità di ottenere un certo valore x effettuando una misura

Il valore μ , attorno al quale la curva è centrata, è identificabile col valore vero della grandezza che vogliamo misurare, mentre la larghezza σ è in qualche modo legata alla precisione sulla misura.

$$\bar{x} = \sum_{k=1}^M m_k \cdot F_k \quad \xrightarrow{M \rightarrow \infty} \quad \bar{x} = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

La funzione di densità di probabilità

e inoltre:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 \cdot f(x) dx$$

Andando a risolvere l'integrale presente nell'espressione precedente, si trova che:

- *nel caso ipotetico di un numero **infinito** di misure:*
 - il valor medio \bar{x} risulta essere uguale ad il valore vero μ
- *nel caso reale di un numero **finito** di misure:*
 - il valor medio \bar{x} calcolato con la formula nota risulta essere la miglior stima di μ .
 - la deviazione standard S_x risulta la miglior stima di σ

Compatibilità di due misure

Due misure, supposte affette da errori casuali, si dicono tra loro compatibili quando la loro differenza può essere ricondotta ad una pura **fluttuazione statistica** attorno al valore nullo

(ovvero, se possono essere considerate uguali, nei limiti dei rispettivi errori sperimentali).

Il concetto di compatibilità può essere **quantificato** per mezzo del "livello di **confidenza**" (CL,) che esplicita il valore di probabilità con cui si vuole essere sicuri ("confidenti") che le due misure siano compatibili, ed indica la probabilità che la loro differenza sia una fluttuazione statistica intorno al valore nullo.

Compatibilità di due misure

Si parte dall'ipotesi che, se due misure

$$\begin{cases} x_1 \pm S_1 \\ x_2 \pm S_2 \end{cases}$$

Se si riferiscono allo stesso valore vero, la loro differenza deve essere distribuita normalm. attorno al valore 0.

Si calcola, quindi:

a) la differenza $\Delta = |x_2 - x_1|$

b) l'errore su questa differenza $S_{diff} = \sqrt{S_1^2 + S_2^2}$

c) il rapporto $t = \frac{|x_2 - x_1|}{S_{diff}}$

e si ricava dalla tabella della gaussiana la probabilità di ottenere una differenza grande come quella osservata o più grande di quella osservata , per una distribuzione delle differenze con **valore centrale 0** e $\sigma = S_{diff}$.

In genere due misure si dicono **compatibili se $CL > 5\%$ ($\approx 2 \sigma$)** e **incompatibili se $CL < 0.3\%$ ($\approx 3 \sigma$)**

Esempio

Due gruppi di studenti fanno due misure della stessa grandezza, trovando i seguenti valori:

$$\begin{cases} 35 \pm 3 \\ 29.1 \pm 0.2 \end{cases}$$

Discutere la compatibilità dei due risultati.

Abbiamo: $|x_2 - x_1| = 35 - 29.1 = 5.9$

$$S_{diff} = \sqrt{(S_{x_2})^2 + (S_{x_1})^2} = \sqrt{3^2 + (0.2)^2} = 3.01$$



$$t = \frac{|x_2 - x_1|}{S_{diff}} = \frac{5.9}{3.01} = 1.96$$

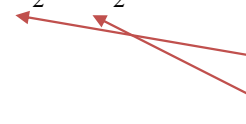
Questo vuol dire che la differenza (5.9) “dista” dal valore atteso 0 di un fattore che è pari a 1.96 volte la deviazione standard sulla differenza:

Ora, dalla tabella della gaussiana ricaviamo che la probabilità di avere un valore che “dista” dal valore atteso di un fattore inferiore a 1.96 deviazioni standard è 0.05 (5%). Quindi possiamo dire che i due valori sono tra loro compatibili con un livello di confidenza del 5%.

Medie Pesate

Consideriamo N studenti ($1, 2, \dots, N$) che misurano la stessa grandezza ottenendo i seguenti risultati:

$$\left\{ \begin{array}{l} \text{Studente 1: } x_1 \pm \delta_1 \\ \text{Studente 2: } x_2 \pm \delta_2 \\ \text{Studente } N: x_N \pm \delta_N \end{array} \right.$$



Miglior media di tutte le misure dello studente 2

Deviazione standard della della media

Supponiamo che le misure effettuate dagli N studenti siano **consistenti**:
cioè la differenza tra x_1, x_2, \dots, x_N non sia significativamente più grandi di $\delta_1, \delta_2, \dots, \delta_N$

Ci poniamo il problema di trovare il modo migliore di **combinare** x_1, x_2, \dots, x_N per ottenere una singola **miglior stima di μ**

Se uno dei due studenti ha eseguito una misura con una **precisione maggiore** degli altri (cioè il suo δ è minore), sarà lecito aspettarsi che alla sua misura debba essere dato un **maggior peso**.

In qualche modo bisognerebbe "**privilegiare**" le informazioni fornite dalle misure più precise.
Fare semplicemente la media dei singoli valori tratterebbe tutti i dati come equivalenti

Medie Pesate

Ognuna delle singole misure è riferita allo stesso **valor vero μ** ; ci aspettiamo quindi che ad ognuna di esse sia associata una distribuzione normale centrata attorno a tale valor vero.

Quindi, possiamo scrivere che la **probabilità** di effettuare la misura x_i con associato l'errore σ_i è proporzionale alla relativa funzione gaussiana centrata su μ :

$$P(x_i) \propto \frac{1}{\sigma_i} e^{-\frac{(x_i - \mu)^2}{2\sigma_i^2}}$$

Quindi, alla serie di misure $x_1 \pm \sigma_1, x_2 \pm \sigma_2, \dots, x_N \pm \sigma_N$ sono associate le probabilità $P(x_1), P(x_2), \dots, P(x_N)$

La **probabilità congiunta** di avere la serie di misure $x_1 \pm \sigma_1, x_2 \pm \sigma_2, \dots, x_N \pm \sigma_N$ è data dal seguente prodotto:

Se indipendenti!

$$P(x_1, x_2, \dots, x_N) = P(x_1) \times P(x_2) \times \dots \times P(x_N) \propto$$

$$\propto \frac{1}{\sigma_1} e^{-\frac{(x_1 - \mu)^2}{2\sigma_1^2}} \cdot \frac{1}{\sigma_2} e^{-\frac{(x_2 - \mu)^2}{2\sigma_2^2}} \cdot \dots \cdot \frac{1}{\sigma_N} e^{-\frac{(x_N - \mu)^2}{2\sigma_N^2}} = \prod_{i=1}^N \left(\frac{1}{\sigma_i} e^{-\frac{(x_i - \mu)^2}{2\sigma_i^2}} \right)$$

N: #studenti
 μ : valore vero

Medie pesate

la precedente espressione può essere riscritta come:

$$P(x_1, x_2, \dots, x_N) \propto \frac{1}{\prod_{i=1}^N \sigma_i} e^{-\sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma_i^2}} = \frac{1}{\prod_{i=1}^N \sigma_i} e^{-\frac{\chi^2}{2}}$$

Dove abbiamo indicato

$$\chi^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{\sigma_i^2} \quad (CHI \quad QUADRATO)$$

Il **principio di Massima Verosimiglianza** asserisce che la miglior stima per il valor vero μ è quella che massimizza la probabilità congiunta di aver effettuato la serie di misure $x_1 \pm \sigma_1, x_2 \pm \sigma_2, \dots, x_N \pm \sigma_N$.

La miglior stima del valore vero μ si trova andando a minimizzare il valore del CHI QUADRATO.

Ricordiamo che per trovare i punti di minimo, si deve porre uguale a zero la derivata del CHI QUADRATO rispetto alla variabile considerata (in questo caso, μ):

$$\frac{d\chi^2}{d\mu} = 0$$

Medie Pesate

Si può dimostrare che tale condizione viene soddisfatta in corrispondenza del seguente valore:

$$x_{best} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \frac{\sum_{i=1}^N x_i \cdot w_i}{\sum_{i=1}^N w_i}$$

Dove abbiamo indicato con x_{best}
la miglior stima del valore vero μ

$$w_i = \frac{1}{\sigma_i^2} \quad \text{"peso"}$$

Si può poi dimostrare che l'incertezza vale:

$$\sigma_{x_{best}} = \frac{1}{\sqrt{\sum_{i=1}^N \frac{1}{\sigma_i^2}}} = \frac{1}{\sqrt{\sum_{i=1}^N w_i}}$$

Medie Pesate

Con pure semplificazioni algebriche, le formule precedenti si riducono alla formula generale della media aritmetica se tutte le incertezze sono uguali tra loro:

$$x_{best} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \xrightarrow{\sigma_1=\sigma_2=\dots=\sigma_N=\sigma} x_{best} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma^2}}{\sum_{i=1}^N \frac{1}{\sigma^2}} = \frac{\frac{1}{\sigma^2} \sum_{i=1}^N x_i}{\frac{N}{\sigma^2}} = \frac{\frac{1}{\sigma^2} \sum_{i=1}^N x_i}{\frac{1}{\sigma^2} N} = \frac{\sum_{i=1}^N x_i}{N}$$

che è appunto la definizione di media aritmetica.

Media Pesata



Media Aritmetica

Relazione funzionale

Consideriamo il caso in cui vogliamo verificare una **relazione funzionale** tra due grandezze **x** e **y**:

$$y = f(x)$$

Possiamo misurare i valori di **y** in corrispondenza di diversi valori di **x**:

$$\left\{ \begin{array}{l} y_1 = f(x_1) \\ y_2 = f(x_2) \\ y_3 = f(x_3) \\ \cdot \\ \cdot \\ y_n = f(x_n) \end{array} \right.$$

In genere, gli x_i sono supposti noti con errore trascurabile, mentre agli y_i viene associato un errore sperimentale δ_i .

Relazione funzionale

Consideriamo ad **esempio** un **grave che cade** dalla cima di un palazzo. Supponiamo di rilevare la posizione del grave ad intervalli (ad esempio regolari) di tempo; registriamo cioè la coppia di valori (tempo, spazio) più volte.

➤ Se riportiamo su un grafico lo **spazio** percorso dal grave in **funzione del tempo** otterremo che i punti si disporranno su una parabola secondo la ben nota relazione:

$$s = \frac{1}{2}gt^2 + v_0t + s_0$$

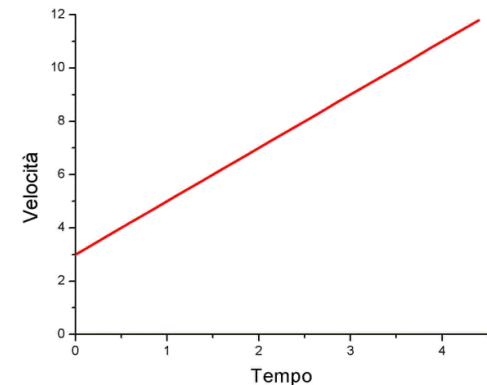
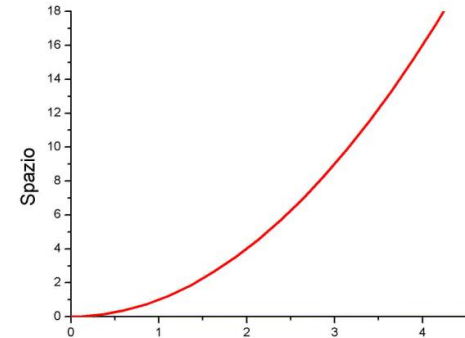
$$s = f(t)$$

➤ Se invece registriamo la coppia di valori (tempo, velocità) cioè rileviamo la **velocità** del grave in **funzione del tempo**, troveremo che i punti si disporranno su di una retta secondo la relazione lineare:

$$v = gt + v_0$$

$$v = h(t)$$

$$y = f(x)$$



Relazione funzionale

La funzione f che lega y a x dipende da una serie di parametri (nel caso delle leggi del moto, posizione iniziale x_0 , velocità iniziale v_0 , accelerazione g).

Nel caso in cui questi parametri siano incogniti, li possiamo ricavare dalle misure effettuate ricordando che la funzione di **distribuzione di probabilità per una misura y_i è data dalla funzione gaussiana centrata sul valore vero y_0 e con larghezza σ** . Il valore vero attorno a cui la distribuzione è centrata corrisponde a quello previsto dalla relazione funzionale $f(x_i)$, mentre la larghezza della distribuzione corrisponde all'errore sperimentale δ_i . Quindi, avendo misurato il valore y_i con errore σ_i , la probabilità di quella misura è esprimibile come:

$$f(y_i) = \frac{1}{\delta_i \cdot \sqrt{2\pi}} e^{-\frac{(y_i - f(x_i))^2}{2\delta_i^2}}$$

Questo discorso vale per ognuno dei valori misurati y_i . La probabilità di tutta la serie di misure y_1, y_2, \dots, y_N , è data dal prodotto delle singole probabilità:

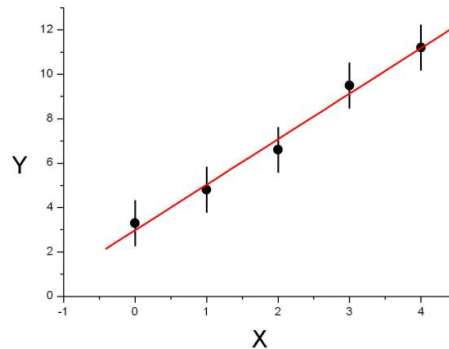
$$\prod_i f(y_i) = \frac{1}{\prod_i \delta_i \cdot \sqrt{(2\pi)^n}} e^{-\sum_i \frac{[y_i - f(x_i)]^2}{2\delta_i^2}}$$

Applicando di nuovo il principio della massima verosimiglianza, discende che i valori incogniti dei parametri che caratterizzano la relazione funzionale studiata si trovano massimizzando tale probabilità.

Relazione funzionale

Limitiamo l'analisi al caso di una **relazione di tipo lineare**:

$$y = A + B \cdot x$$



B coefficiente angolare

A Ordinata all'origine

Lo scopo è trovare la retta che meglio si adatta alle misure.

Ciò significa trovare la miglior stima delle costanti A e B basandoci sui dati $(x_1, y_1)(x_2, y_2).....(x_N, y_N)$

Possiamo riscrivere la probabilità introdotta precedentemente esplicitando la relazione lineare:

$$\prod_i f(y_i) = \frac{1}{\prod_i \delta_i \cdot \sqrt{(2\pi)^n}} e^{-\sum_i \frac{[y_i - (A + B \cdot x_i)]^2}{2\delta_i^2}}$$

$$f(x_i) = A + B \cdot x_i$$

Relazione funzionale

Si dimostra che i valori di **A** e **B** che massimizzano la probabilità sono dati da:

$$A = \frac{\sum_i \frac{x_i^2}{\delta_i^2} \cdot \sum_i \frac{y_i}{\delta_i^2} - \sum_i \frac{x_i y_i}{\delta_i^2} \cdot \sum_i \frac{x_i}{\delta_i^2}}{\sum_i \frac{x_i^2}{\delta_i^2} \cdot \sum_i \frac{1}{\delta_i^2} - \left(\sum_i \frac{x_i}{\delta_i^2} \right)^2} \quad B = \frac{\sum_i \frac{x_i y_i}{\delta_i^2} \cdot \sum_i \frac{1}{\delta_i^2} - \sum_i \frac{x_i}{\delta_i^2} \cdot \sum_i \frac{y_i}{\delta_i^2}}{\sum_i \frac{x_i^2}{\delta_i^2} \cdot \sum_i \frac{1}{\delta_i^2} - \left(\sum_i \frac{x_i}{\delta_i^2} \right)^2}$$

Gli **errori su A e B**, che nel caso più comune in cui gli errori sperimentali sono casuali corrispondono ad una deviazione standard e per questo li indichiamo con σ , sono dati da:

$$\sigma_A^2 = \frac{\sum_i \frac{x_i^2}{\delta_i^2}}{\sum_i \frac{x_i^2}{\delta_i^2} \cdot \sum_i \frac{1}{\delta_i^2} - \left(\sum_i \frac{x_i}{\delta_i^2} \right)^2} \quad \sigma_B^2 = \frac{\sum_i \frac{1}{\delta_i^2}}{\sum_i \frac{x_i^2}{\delta_i^2} \cdot \sum_i \frac{1}{\delta_i^2} - \left(\sum_i \frac{x_i}{\delta_i^2} \right)^2}$$

Relazione funzionale

Siccome i due parametri A e B sono stati trovati in maniera non indipendente, ma entrambi ricavati dallo stesso set di dati, si verifica che il **termine di covarianza** risulta essere diverso da 0:

$$\sigma_{AB} = - \frac{\sum_i \frac{x_i}{\delta_i^2}}{\sum_i \frac{x_i^2}{\delta_i^2} \cdot \sum_i \frac{1}{\delta_i^2} - \left(\sum_i \frac{x_i}{\delta_i^2} \right)^2}$$

Si definisce inoltre il **termine di correlazione**

$$\rho_{AB} = \frac{\sigma_{AB}}{\sigma_A \sigma_B}$$

Diagram illustrating the components of the correlation coefficient ρ_{AB} :

- The numerator σ_{AB} is labeled "COVARIANZA AB" (Covariance AB).
- The denominator σ_A is labeled "COVARIANZA A" (Covariance A).
- The denominator σ_B is labeled "COVARIANZA B" (Covariance B).

che è un indice della correlazione tra i parametri A e B.

Relazione funzionale

Quando σ_y^2 non viene specificato, deve essere ricavato a posteriori mediante la formula:

$$\sigma_y = \sqrt{\frac{1}{N-2} \sum_i [y_i - (A + B \cdot x_i)]^2}$$

A denominatore c'è un termine $N-2$ invece che N . Ricordando la definizione di **gradi di libertà** (pari al numero di osservazioni meno i vincoli), vediamo che in questo caso i gradi di libertà sono $N-2$ (sono infatti 2 i parametri ricavati dai dati, A e B).

Se avessimo considerato due sole coppie di valori (come sappiamo dalla geometria elementare per due punti passa una ed una sola retta), la formula di σ_y , con N al denominatore darebbe come risultato il valore 0 il che è assurdo; invece con $N-2$ otterremmo sotto la radice il termine

$$\sigma_y = \sqrt{\frac{0}{0}}$$

che essendo una forma di indecisione indica che per solo due misure il valore σ_y è giustamente indeterminato.

Il test di chi quadrato

Abbiamo visto il procedimento da seguire per trovare i valori incogniti dei parametri che caratterizzano una relazione funzionale, quando sono a disposizione una serie di misure sperimentali.

Il fatto di avere delle formule o degli algoritmi che permettono di ricavare i valori incogniti dei parametri **non significa** automaticamente che le misure sperimentali sono in accordo con la relazione funzionale ipotizzata.

Un **metodo quantitativo** e statisticamente corretto per verificare l'accordo dei dati con una determinata relazione funzionale è il **test di chi quadrato**

Per trovare i valori incogniti dei parametri abbiamo infatti utilizzato il **principio della massima verosimiglianza**, andando a massimizzare la probabilità congiunta di trovare i valori misurati

$$\prod_i f(y_i) = \frac{1}{\prod_i \delta_i \cdot \sqrt{(2\pi)^n}} e^{-\sum_i \frac{[y_i - f(x_i)]^2}{2\delta_i^2}}$$

Trovare il massimo della probabilità equivale a minimizzare il termine all'esponente:

$$\chi^2 = \sum_i \frac{[y_i - f(x_i)]^2}{\delta_i^2}$$

Il test di chi quadrato

$$\chi^2 = \sum_i \frac{[y_i - f(x_i)]^2}{\delta_i^2}$$

che non è altro che, per ogni valore misurato, il rapporto tra

la differenza tra misura e valore previsto dalla funzione

l'errore di misura

entrambi elevati al quadrato

Idealmente, il termine a numeratore dovrebbe essere 0. In realtà, ci si aspetta che la differenza tra la misura e la previsione sia dello stesso ordine di grandezza dell'errore sperimentale, pertanto ci si aspetta che ogni termine nella somma che definisce il χ^2 sia uguale a 1. Di conseguenza è ragionevole aspettarsi che il termine all'esponente sia uguale al numero di misure effettuate.

Per generalizzare il discorso, risulta utile considerare il **chi quadrato ridotto**, cioè il rapporto tra χ^2 e il numero di gradi di libertà:

$$\tilde{\chi}^2 = \frac{\chi^2}{n_g}$$

Ci aspettiamo quindi che il valore di chi quadrato ridotto, pari al rapporto tra il numero di misure fatte e numero di gradi di libertà, sia poco superiore a 1.

Esistono delle tabelle che danno, la probabilità, in funzione del numero di gradi di libertà, di trovare un valore di chi quadrato ridotto maggiore o uguale a valori prefissati. E' quindi possibile, usando questa tabella, trovare la probabilità che in effetti le misure fatte siano regolate dalla relazione che era stata ipotizzata.

Correlazioni lineari

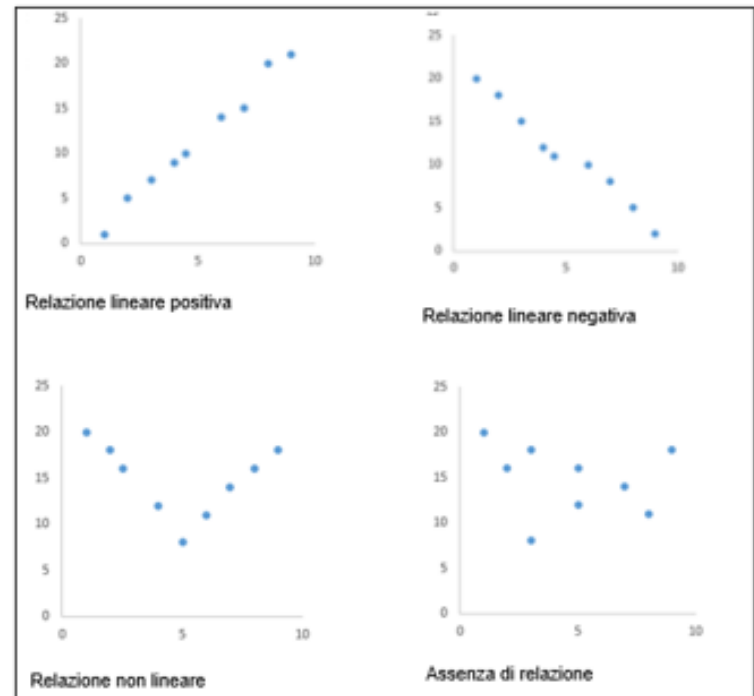
Correlazioni

Nel caso che la relazione che vogliamo verificare sia lineare, possiamo usare un metodo alternativo al chi quadrato per verificarne l'esistenza.

Lo stabilire se tra le variabili x e y esista una correlazione lineare è reso difficile quando i valori sono molto dispersi e non si ha alcuna informazione sulla loro incertezza.

La correlazione indica la tendenza che hanno due variabili (X e Y) a variare insieme, ovvero, a co-variare.

La relazione analizzata può essere di tipo lineare se, rappresentata su assi cartesiani, si avvicina alla forma di una retta e in questo caso, all'aumentare (o al diminuire) di X aumenta (o diminuisce) Y .



Covarianza

Quando si osservano due caratteri diversi, riferiti ad una medesima popolazione, dei quali conosciamo i valori numerici, si può fare riferimento a indici statistici che possono descrivere come i due insiemi di dati variano tra loro.

Un indice simmetrico per misurare la concordanza o la discordanza tra due caratteri quantitativi è la **covarianza**, ovvero la media dei prodotti degli scostamenti delle variabili X e Y dalle rispettive medie:

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y)$$

Covarianza

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y)$$

È un indice simmetrico che misura la concordanza o la discordanza tra due caratteri quantitativi.

È positiva se al numeratore prevalgono i prodotti di scostamenti concordi (tutti e due positivi o tutti e due negativi) mentre è negativa se prevalgono i prodotti di scostamenti discordi.

Una covarianza pressoché uguale a zero indica che i dati non sono in relazione diretta tra loro.

Si può dimostrare che la formula della covarianza si può scrivere come

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \mu_X \cdot \mu_Y = \mu_{XY} - \mu_X \cdot \mu_Y$$

per una maggiore facilità di calcolo.

Il coefficiente di Pearson

La covarianza dipende dall'unità di misura delle osservazioni cosicché non è corretto confrontarne il valore su diverse distribuzioni doppie.

Per ovviare a tale inconveniente è opportuno trasformare la covarianza in un indice relativo

Si può quindi introdurre un indice relativo, il **coefficiente di correlazione lineare** di Bravais e Pearson che esprime l'intensità del legame lineare tra due variabili.

$$\rho_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \mu_X \mu_Y}{\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \mu_X^2} \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - \mu_Y^2}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$-1 \leq \rho_{xy} \leq 1$$

Perché la covarianza assume valori all'interno di: $-\sigma_x \sigma_y \leq \sigma_{xy} \leq \sigma_x \sigma_y$

Il coefficiente di Pearson

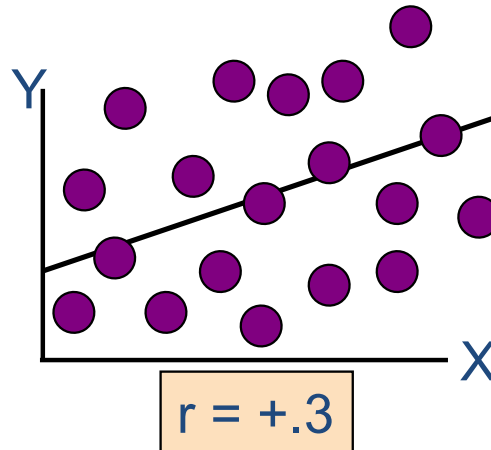
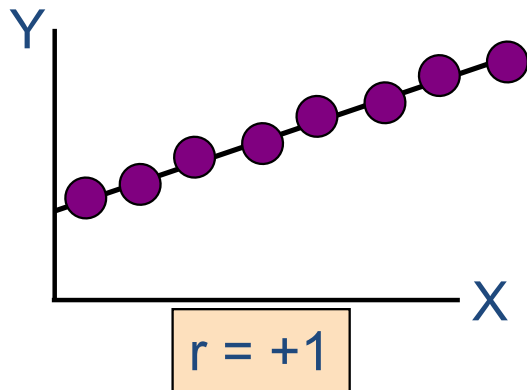
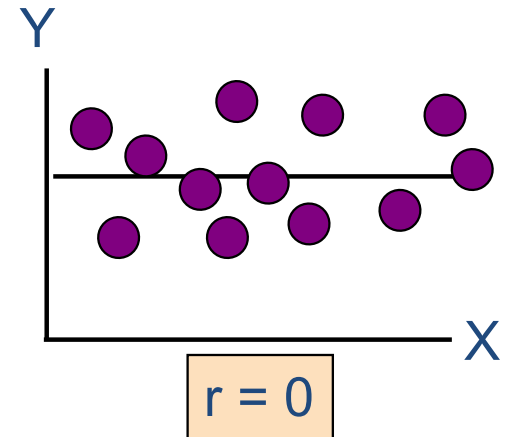
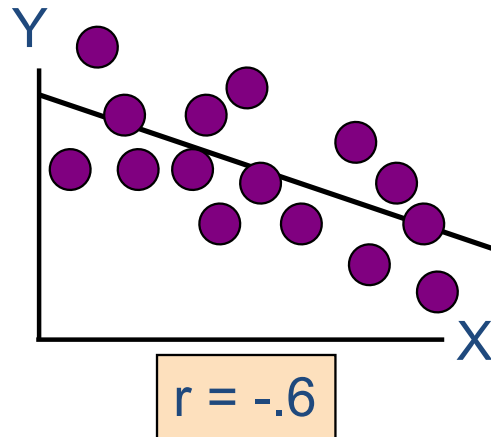
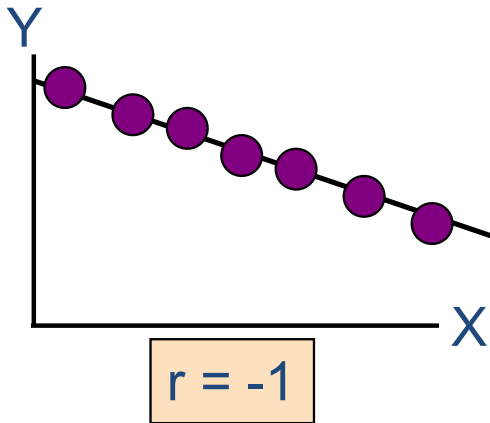
$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Il segno presentato dal coefficiente di correlazione corrisponde al segno della covarianza giacché al suo denominatore vi sono quantità sempre positive.

Proprietà del coefficiente di correlazione:

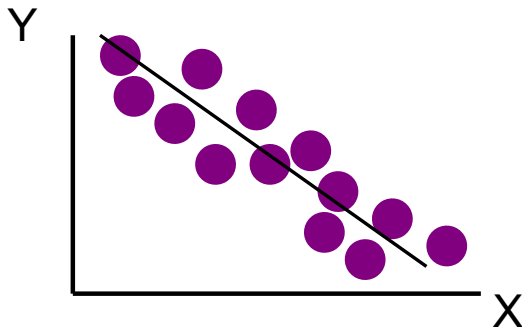
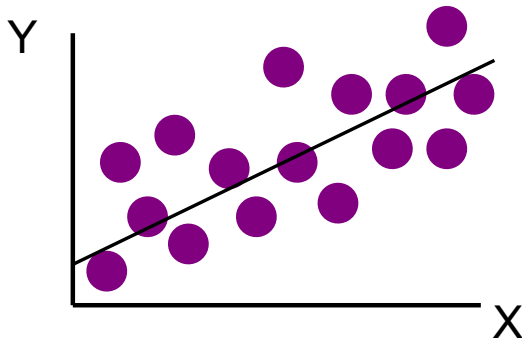
- ✦ $\rho_{XY} = 1$ se tra X e Y sussiste un perfetto **legame lineare** e i due caratteri sono concordi
- ✦ $\rho_{XY} = -1$ se tra X e Y sussiste un perfetto **legame lineare** e i due caratteri sono discordi
- ✦ $\rho_{XY} = 0$ se i due caratteri sono indipendenti oppure se la loro relazione non è lineare

Coefficienti di Correlazione: esempi

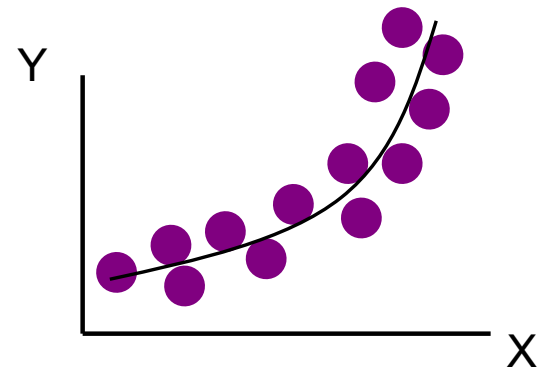
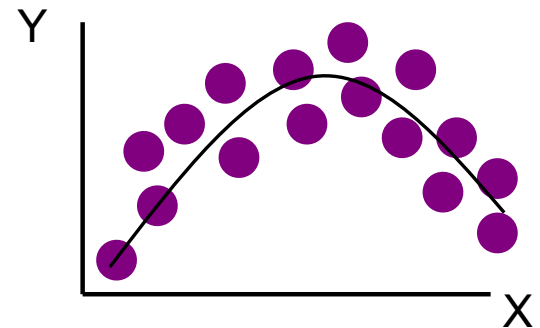


Correlazione lineare: esempi

Linear relationships

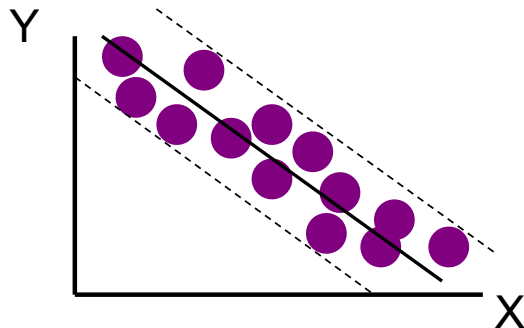
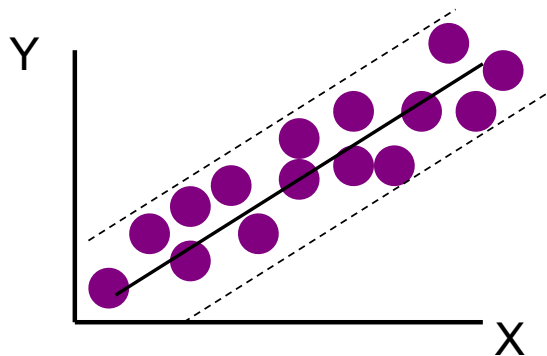


Curvilinear relationships

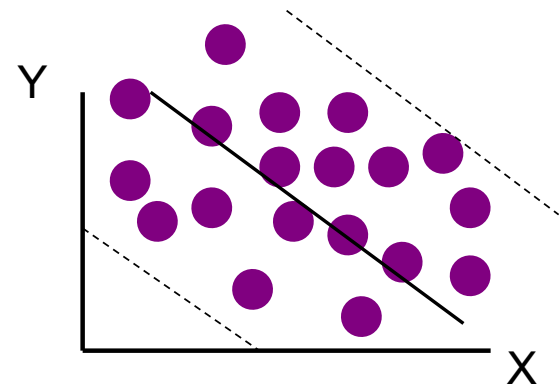
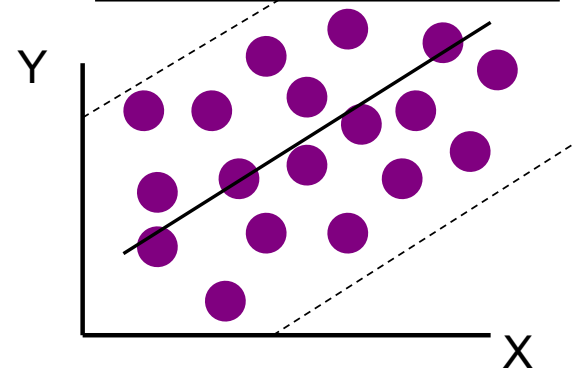


Correlazione lineare: esempi

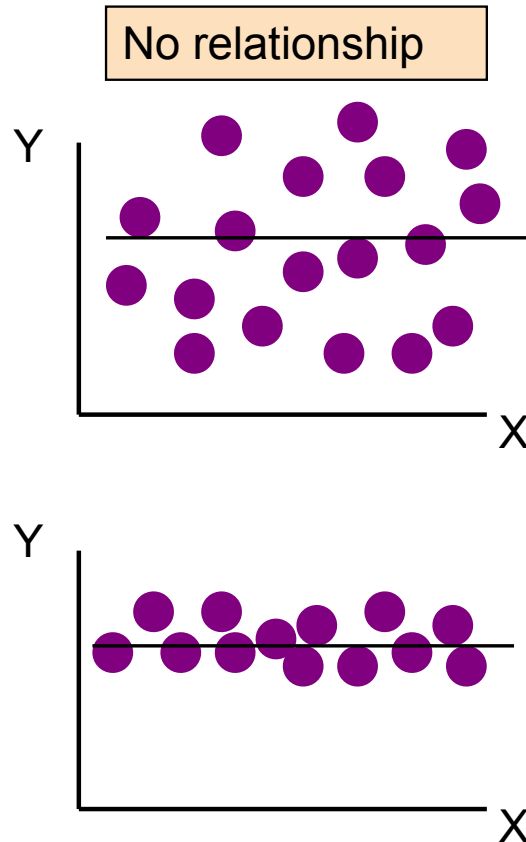
Strong relationships



Weak relationships



Correlazione lineare: esempi



Interpolazione dei dati

L'interpolazione dei dati

Date due variabili, X e Y , si è interessati a comprendere come la variabile Y (**dipendente** o **risposta**) sia influenzata dalla X (**esplicativa** o **indipendente**).

Quando l'analisi delle relazioni tra variabili evidenzia l'esistenza di un legame di Y da X , sorge naturale l'esigenza di formalizzare la natura della dipendenza mediante un opportuno modello di tipo matematico. A tal fine si può ricorrere ad una funzione interpolante del tipo:

$$\hat{Y} = f(X)$$

Dove f denota una qualche *funzione* di X e \hat{Y} indica che si sta approssimando la realtà osservata (Y) con una curva (f) più *semplice e regolare*.

L'interpolazione dei dati

Il più semplice modello di regressione è il modello di **regressione lineare semplice**. In esso si assume che la funzione di regressione $f(X)$ sia lineare, e si considera una sola variabile esplicativa. Questa retta di regressione descrive come cambia una variabile dipendente Y quando cambia la variabile esplicativa X .

$$\hat{Y} = \beta_0 + \beta_1 X$$

β_0 e β_1 sono delle quantità costanti detti **parametri** della retta;

β_0 è chiamata **intercetta** perché è l'altezza alla quale la retta incontra l'asse delle ordinate, ossia è il valore di Y per $X=0$,

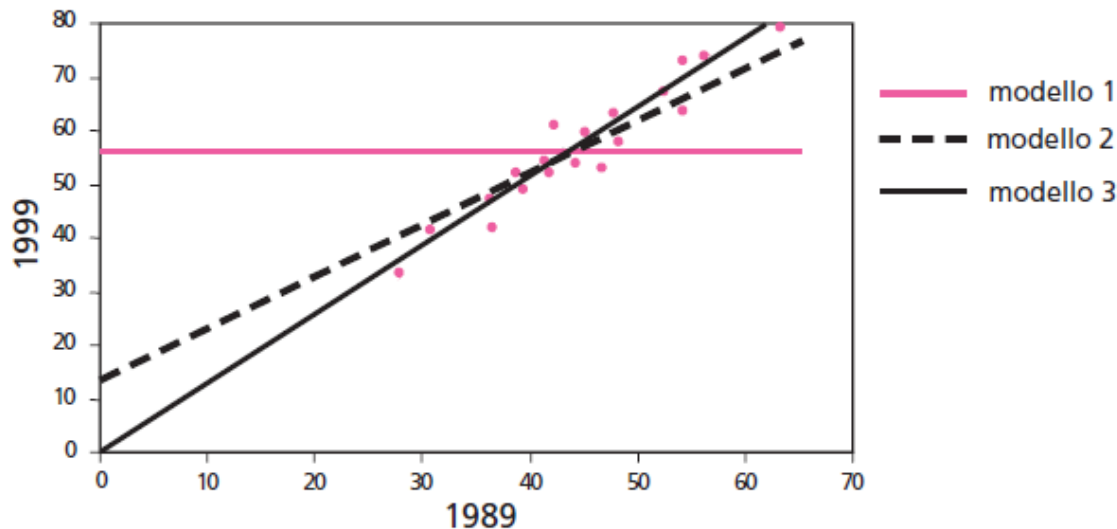
β_1 è chiamato **coefficiente angolare** perché esprime l'incremento di Y per un incremento unitario di X e determina l'inclinazione della retta e la sua pendenza.

Predisporre un modello di regressione lineare semplice significa, pertanto, utilizzare i dati per assegnare un valore ai parametri β_0 e β_1 della retta.

Interpolazione dei dati

Predisporre un modello di regressione lineare semplice significa, pertanto, utilizzare i dati per assegnare un valore ai parametri β_0 e β_1 della retta

$$\hat{Y} = \beta_0 + \beta_1 X$$

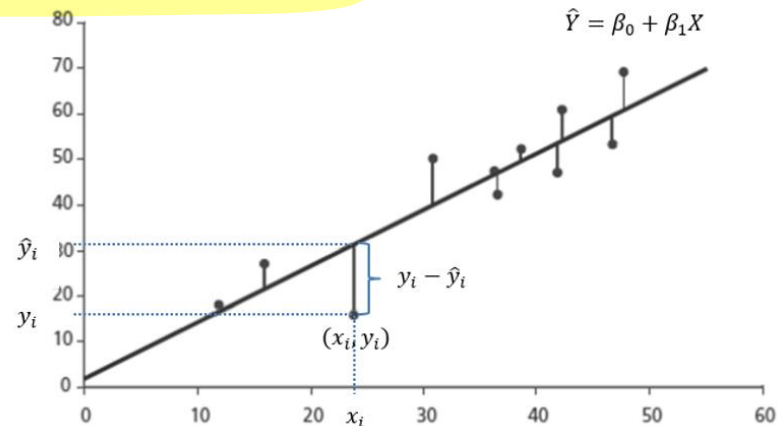


In linea teorica esistono infinite rette con cui interpolare i dati perché ai parametri β_0 e β_1 possiamo assegnare un qualunque valore reale.

Metodo di stima dei minimi quadrati

Abbiamo allora bisogno di stabilire un **criterio** con cui scegliere la retta che *meglio approssima* la spezzata di regressione. Il criterio (normalmente) usato per la regressione lineare semplice è il **metodo dei minimi quadrati**. Il metodo dei minimi quadrati consiste nell'esprimere in una formula la *distanza* fra i dati osservati (la realtà) e la retta di regressione (il modello teorico) e nell'assegnare ai parametri del modello il valore che rende **minima** (il più piccolo possibile) tale distanza. Vediamolo in formule:

- valori (reali) osservati $\rightarrow y_j$;
- modello \rightarrow retta di regressione $\hat{Y} = \beta_0 + \beta_1 X$
- valori (teorici) approssimati mediante il modello $\rightarrow \hat{y}_i = \beta_0 + \beta_1 x_i$;
- **distanza fra dati reali e valori teorici** \rightarrow è la differenza $y_i - \hat{y}_i$, va elevata al quadrato per eliminare l'influenza del segno



Metodo di stima dei minimi quadrati

Nel caso della retta di regressione si dimostra che per i parametri β_0 e β_1 esiste un'unica soluzione che coinvolge quantità conosciute.

$$\beta_1 = \frac{\sigma_{xy}}{\sigma_x^2} \qquad \beta_0 = \mu_y - \beta_1 \mu_x$$

Poiché al denominatore di β_1 vi è la varianza di X , che è una quantità sempre positiva, allora β_1 prenderà il segno della covarianza

- correlazione positiva $\Rightarrow \sigma_{XY} > 0 \Rightarrow \beta_1 > 0 \Rightarrow$ retta dei minimi quadrati crescente;
- correlazione negativa $\Rightarrow \sigma_{XY} < 0 \Rightarrow \beta_1 < 0 \Rightarrow$ retta dei minimi quadrati decrescente.

Sostituendo le soluzioni dei minimi quadrati nella retta di regressione si ottiene la **retta dei minimi quadrati** cioè la *sola* retta che, fra le infinite che possono interpolare la spezzata di regressione, rende *minima* la **distanza totale** fra i dati osservati e il modello.

Regressione

Al fine di verificare in che misura la retta di regressione sia una buona interpolante, è necessario individuare quanto il modello si adatti alle coppie di valori, in altre parole di quanto i valori teorici \hat{y}_i siano vicini ai valori reali di y_j .

Come visto in precedenza il metodo dei minimi quadrati garantisce che la distanza totale fra i dati reali osservati e i valori teorici del modello di regressione sia minima, cioè la parte più piccola possibile, ma questo non significa necessariamente che tale distanza sia piccola o magari nulla.

Regressione

- a) La media dei quadrati degli scarti dei valori teorici \hat{y}_i dalla media μ_y , detta **varianza spiegata dalla retta di regressione**.

Rappresenta quella frazione di varianza totale σ_y^2 *spiegata* (cioè posseduta) dalla retta di regressione.

$$\bar{\sigma}_y^2 = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \mu_y)^2$$

- b) La media dei quadrati degli scarti dei valori effettivi y_i dai valori teorici \hat{y}_i , detta **varianza residua dalla retta di regressione**.

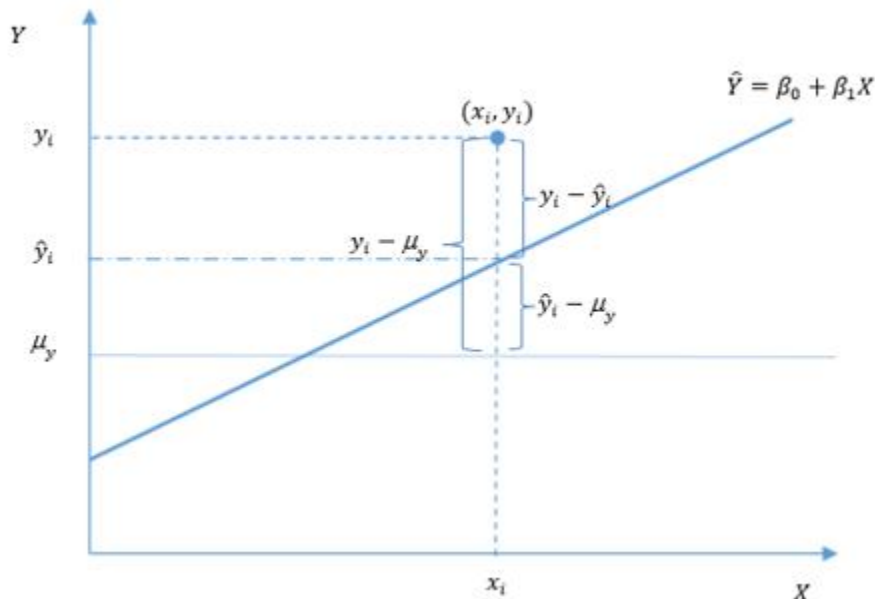
Rappresenta quella frazione di varianza totale σ_y^2 *non spiegata* (e quindi residua) dalla retta di regressione.

$$\tilde{\sigma}_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Regressione

Quindi

$$\frac{1}{n} \sum_{i=1}^n (y_i - \mu_y)^2 = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \mu_y)^2 + \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



$$\sigma_y^2 = \bar{\sigma}_y^2 + \tilde{\sigma}_y^2$$

Varianza
spiegata
dalla retta di
regressione

Varianza
residua
dalla retta di
regressione

Scomposizione della
varianza

Coeff. Di Determinazione

Come misura della «bontà della retta di regressione sembra, quindi, ragionevole utilizzare la frazione normalizzata di varianza totale spiegata dalla retta di regressione.

$$R_{XY}^2 = \frac{\bar{\sigma}_y^2}{\sigma_y^2} = 1 - \frac{\tilde{\sigma}_y^2}{\sigma_y^2}$$

è detto **Coefficiente di Determinazione** ed indica la proporzione di variabilità di Y spiegata dalla variabile esplicativa X, attraverso il modello di regressione.

Coeff. Di Determinazione

Si osservi come ai fini del calcolo, per il modello lineare di primo grado, rivesta particolare importanza l'eguaglianza

$$R_{XY}^2 = \frac{\overline{\sigma}_y^2}{\sigma_Y^2} = \frac{\frac{\sigma_{XY}^2}{\sigma_X^2}}{\sigma_Y^2} = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = \rho^2$$

Infatti con ρ_{XY} si misura la correlazione ed elevandolo al quadrato si ottiene la bontà di adattamento. Questo è vero solo per la retta dei minimi quadrati in quanto essa si adatta tanto meglio alla realtà osservata quanto più elevata è la correlazione tra X e Y , positiva o negativa che sia.

Coeff. Di Determinazione

È noto che se ρ_{XY} assume valori tra -1 e +1, allora la bontà di adattamento ρ^2 assumerà valori fra 0 e 1

- $\rho^2 = 0$ Se $\sigma_y^2 = 0$ cioè se $\sigma_y^2 = \sigma_y^2$ la retta lascia tutto residuo e non spiega nulla della variabilità di Y (X e Y sono incorrelati);
- $\rho^2 = 1$ se $\sigma_y^2 = 0$ cioè se $\sigma_y^2 = \sigma_y^2$ la retta non lascia alcun residuo e spiega perfettamente la variabilità di Y (X e Y sono perfettamente correlati);
- I valori di ρ^2 intermedi tra 0 e 1 sono interpretabili come percentuali di variabilità di Y spiegata dalla retta dei minimi quadrati.

La pendenza della retta di regressione corrisponde al segno del coefficiente di correlazione. Quando questo assume i valori estremi del suo intervallo di variazione, $\rho^2 = \pm 1$, i punti osservati sono allineati perfettamente lungo la retta; se invece $\rho^2 = 0$, la retta di regressione è parallela all'asse delle ascisse e ciò indica che il valore medio di Y non dipende linearmente dalla X.