Misure Meccaniche e Collaudi

(A. A. 2020/21)

Errori di misura ed elaborazione dei dati

M. De Lucia / D. Vangi

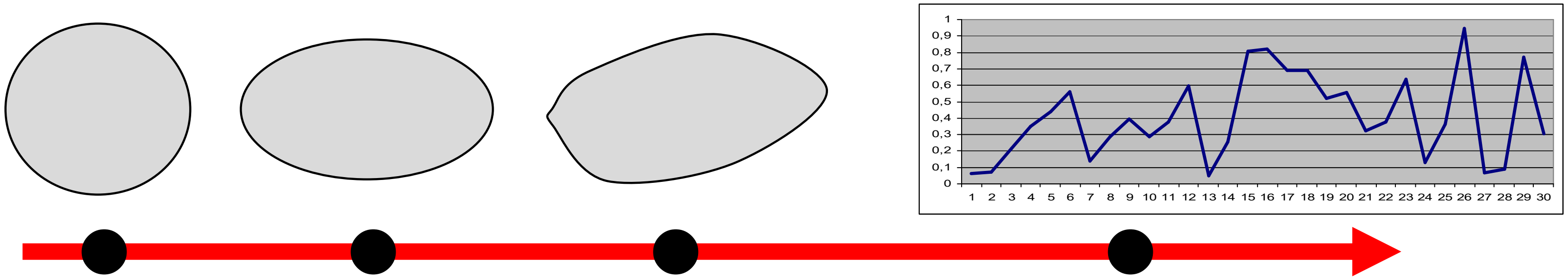
Variabilità delle misure

Eseguendo più *misure* della stessa grandezza in condizioni "nominalmente uguali" si ottiene un insieme di *valori* distribuiti su un intervallo: ciò è dovuto a:

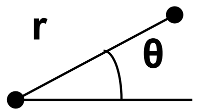
- **non "invarianza nel tempo" dello stato del sistema oggetto della misura**
- **incompletezza del modello utilizzato per descrivere il sistema oggetto della misura**
- **perturbazione del fenomeno per effetto della operazione di "misura" o per fenomeni esterni (rumore)**
- **….**

“Niente è più misurabile della sua definizione”

Es. sezione di una barra



- **Sezione circolare** — **Φ** — **diametro**
- **Sezione ellittica** — **a $\perp$ b** — **semiassi**
- **Sezione** — **r(θ)**



- **Rugosità** — **??** — **?**

Approccio “classico”

- Valore “vero” = misura “ideale” della grandezza, in condizioni “ideali” del fenomeno “ideale”.

- Valore “vero” x_v: metrologicamente inaccessibile per la “imperfezione” propria delle realizzazioni umane

- Dallo strumento x_m: valore misurato
- Errore ε; $x_v = x_m \pm \varepsilon$

Insieme di misure

$$x_1 \; x_2 \; \ldots\ldots \; x_i \; \ldots\ldots \; x_n$$

- **Ottenuto un insieme di misure x_i della "stessa grandezza in condizioni nominalmente uguali"**
- **Non esiste il "valore vero" = un solo valore avrebbe bisogno di un "infinito" contenuto informativo (quindi impossibile)**
- **Passare ad un valore x_o il più possibile (compatibilmente con il problema in esame) vicino al "vero" con la valutazione della incertezza ad esso associabile:**

$$x_0 \pm \Delta$$

Dispersione delle misure

$$x_0 \pm \Delta$$

- **Intervallo Δ**
  - **"insieme" che contiene il non noto valore "vero" del misurando - *errore***
  - **"indice" della "quantità di informazione" acquisita sul misurando - *incertezza***

Dispersione delle misure

$$x_0 \pm \Delta$$

- **Intervallo Δ**
  - **"range" di "ignoranza" = imperfezione dei risultati**
  - **"range" di "indifferenza" / "equivalenza" = possibilità di considerare equivalenti due misure - non c'è nessun motivo per scegliere all'interno dell'intervallo un valore piuttosto che un altro**

Errori: classificazione per “tipo”

- **Sistematici** = presenti in ogni misura – costanti in valore e segno

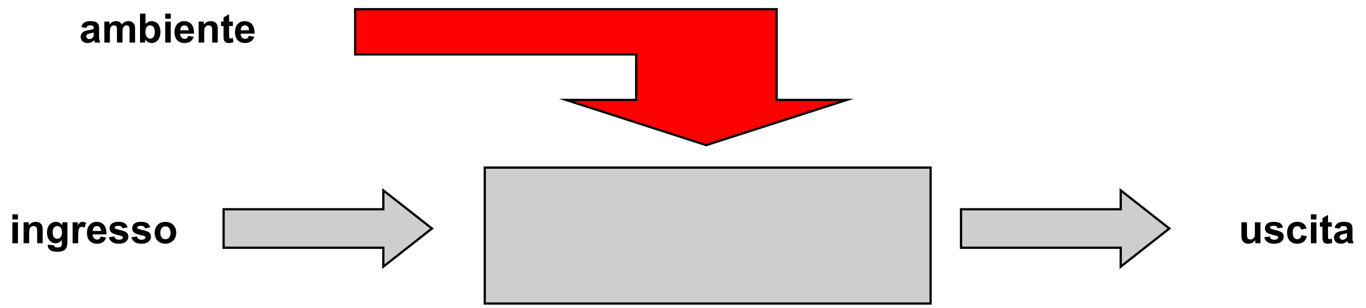
  (es. errore sullo zero, errore di sensibilità, errore di linearità, isteresi?, ecc.)
- **Parassiti** = presenti in una sola misura (valore anomalo)

  (es. errore di lettura, uso errato dello strumento, ecc.)
- **Casuali** (accidentali) = presenti in ogni misura – variabili – di valori sia positivi che negativi - di modesta entità – valore medio nullo; si trattano mediante la teoria degli errori

Errori:
classificazione per “fonte”

- **Connessi con lo strumento** di misura = non idoneità dello strumento (es. isteresi, deriva, errore di inversione, etc.)
- **Connessi con l’operazione di misura** = inserzione nel fenomeno della catena di misura (errori di disturbo)
- **Connessi con l’ambiente** = effetti perturbativi dell’ambiente sulla catena di misura e sul fenomeno

Errori: connessi con l'ambiente



Strategie:

- **Protezione passiva (in ingresso)**
  - Isolare lo strumento / l’ambiente
  - Insensibilizzare lo strumento
- **Protezione attiva (in uscita)**
  - Compensando l’uscita
  - Correggendo l’uscita
- **Taratura nelle condizioni ambientali**

Errori:
classificazione per "caratteristica"

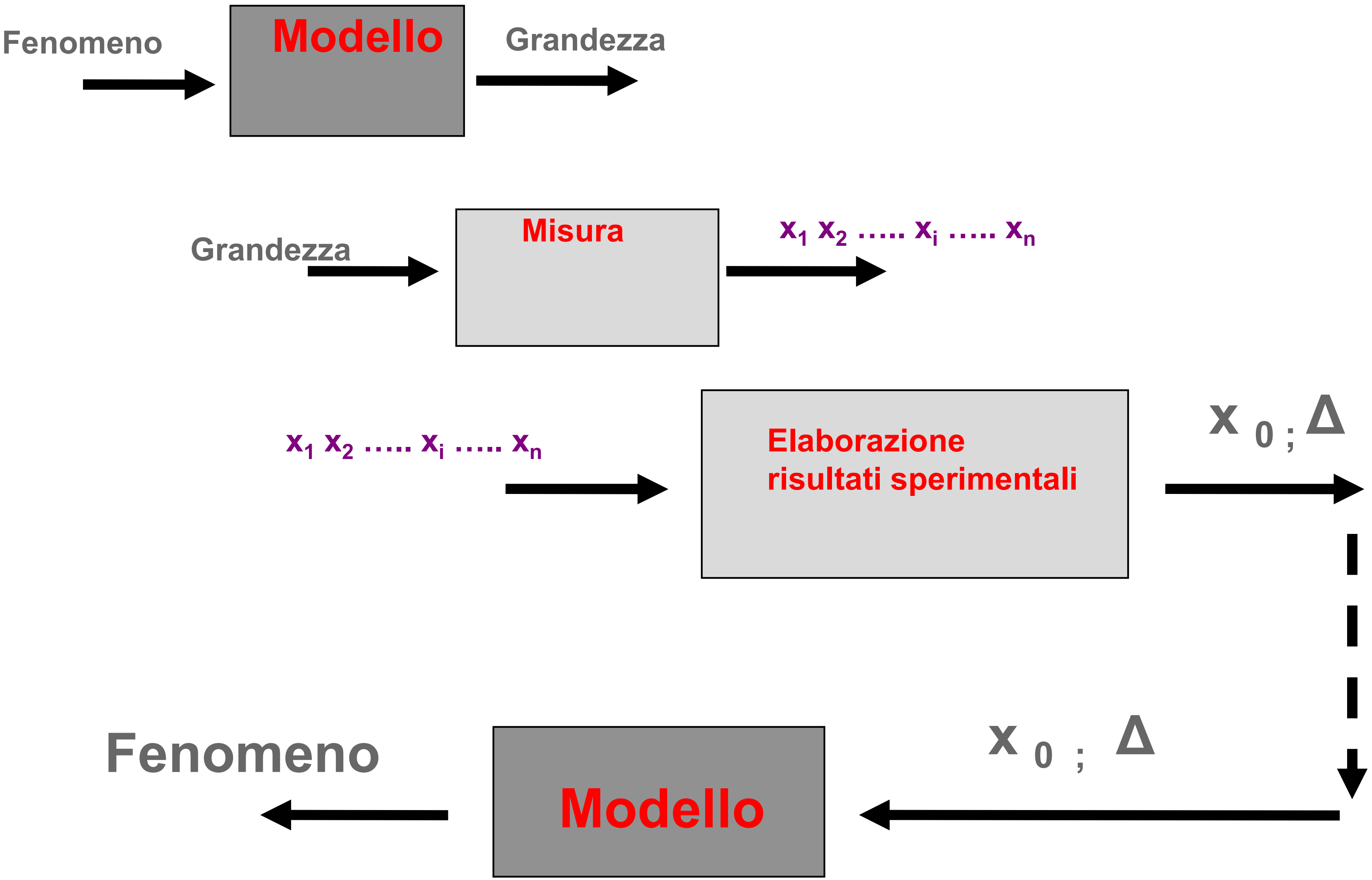
- **Identificabili ➔ evitabili**
- **Identificabili ➔ compensabili o correggibili**
- **Identificabili & incorreggibili ➔ valutabili nell'ambito della "incertezza" associata alla misura**
- **Non identificabili ➔ non valutabili & (in ipotesi) insignificanti**

Errori

- **Errori riducibili ripetendo le misure - legate ad un effettivo contenuto informativo che cresce all’aumentare del numero delle misure - (errori casuali)**

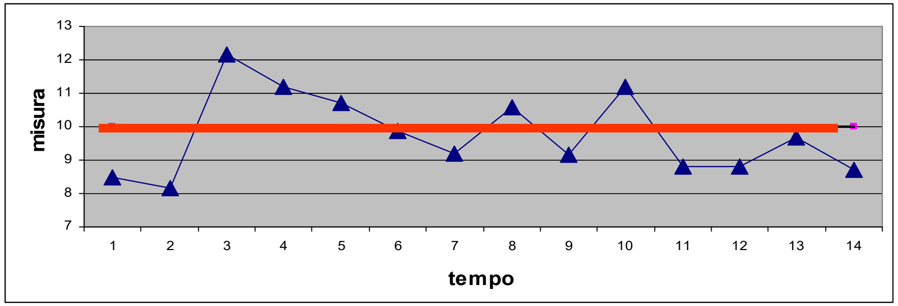
- **Errori non riducibili ripetendo le misure - insensibili al ripetersi delle misure - (errori sistematici)**

Procedura


Fenomeno
Modello
Grandezza
Grandezza
Misura
$x_1\ x_2\ .....\ x_i\ .....\ x_n$
$x_1\ x_2\ .....\ x_i\ .....\ x_n$
Elaborazione
risultati sperimentali
$x_0\ ;\ \Delta$
$x_0\ ;\ \Delta$
Fenomeno
Modello

Errori casuali

Un insieme di misure x_i della stessa grandezza "tempo invariante" in condizioni "nominalmente uguali"



$$x_1 \; x_2 \; \ldots\ldots \; x_i \; \ldots\ldots \; x_n$$

$$x_i = \text{grandezza } x_0 \pm \text{errore casuale } \varepsilon_i$$

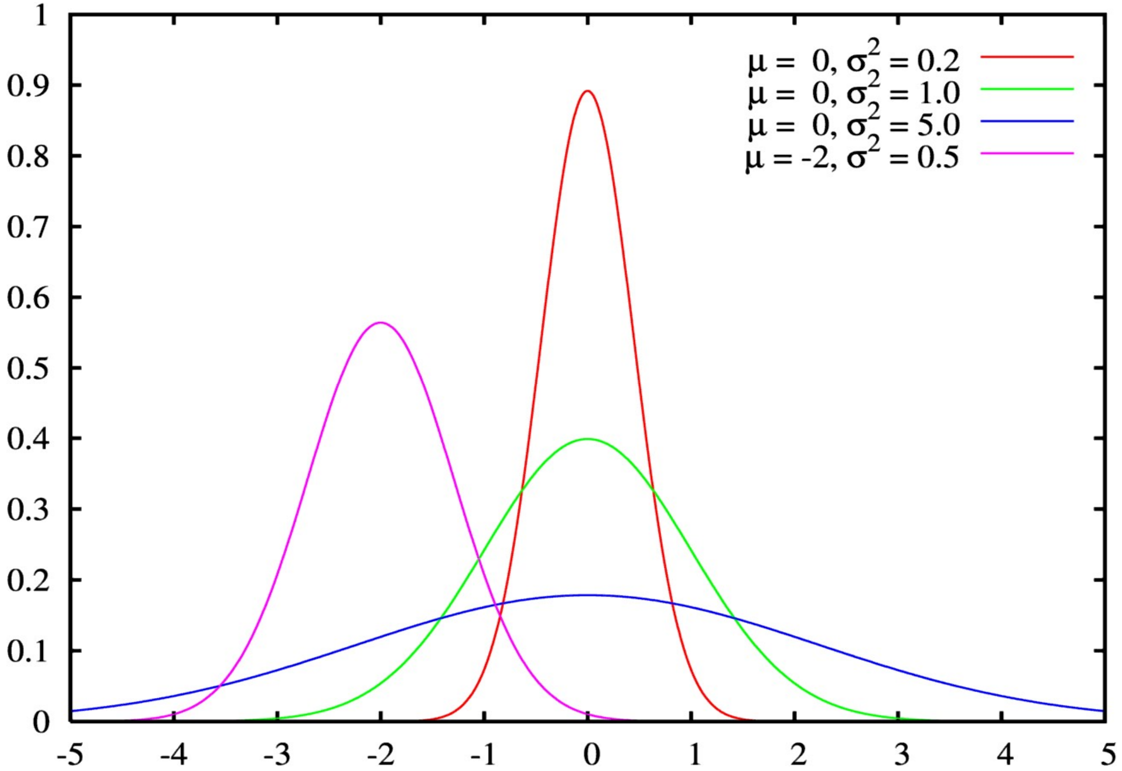
X_i variabile casuale

$$x_i = x_0 \pm \varepsilon_i$$

- variabili “casuali” ⇨ variabili a distribuzione normale / Gauss ⇨ descrivibili in modo sintetico da μ, σ parametri che identificano la distribuzione gaussiana
- x_0 ⇨ valore con maggiore probabilità / frequenza μ – valore tanto più attendibile quanto più frequente ⇨ componente continua del segnale
- ε ⇨ dispersione intorno a μ – varianza σ^2 ⇨ rumore nel segnale

Variabile Casuale Normale
curva di Gauss

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{con} \quad -\infty < x < \infty$$



μ **= media**

σ **= deviazione standard**

(legata all'incertezza)

σ^2 **= varianza**

Variabile Casuale Normale
Curva di Gauss

Per una distribuzione continua:

Media:

$$\mu = \int_{-\infty}^{+\infty} x \cdot p(x) dx$$

Deviazione standard:

$$\sigma = \sqrt{\int_{-\infty}^{+\infty} (x-\mu)^2 \, p(x) dx}$$

Per una distribuzione discreta:

Per la popolazione:

Media:

$$\mu = \frac{\sum_i x_i}{n}$$

Deviazione standard:

$$\sigma = \sqrt{\frac{\sum_i (x_i-\mu)^2}{n}}$$

Per il campione:

Media:

$$m = \frac{\sum_i x_i}{n}$$

Deviazione standard:

$$s = \sqrt{\frac{\sum_i (x_i-\mu)^2}{n-1}}$$

Altre definizioni

Moda:

Per una distribuzione continua: valore corrispondente al massimo della curva di distribuzione

Per una distribuzione discreta: valore più frequente

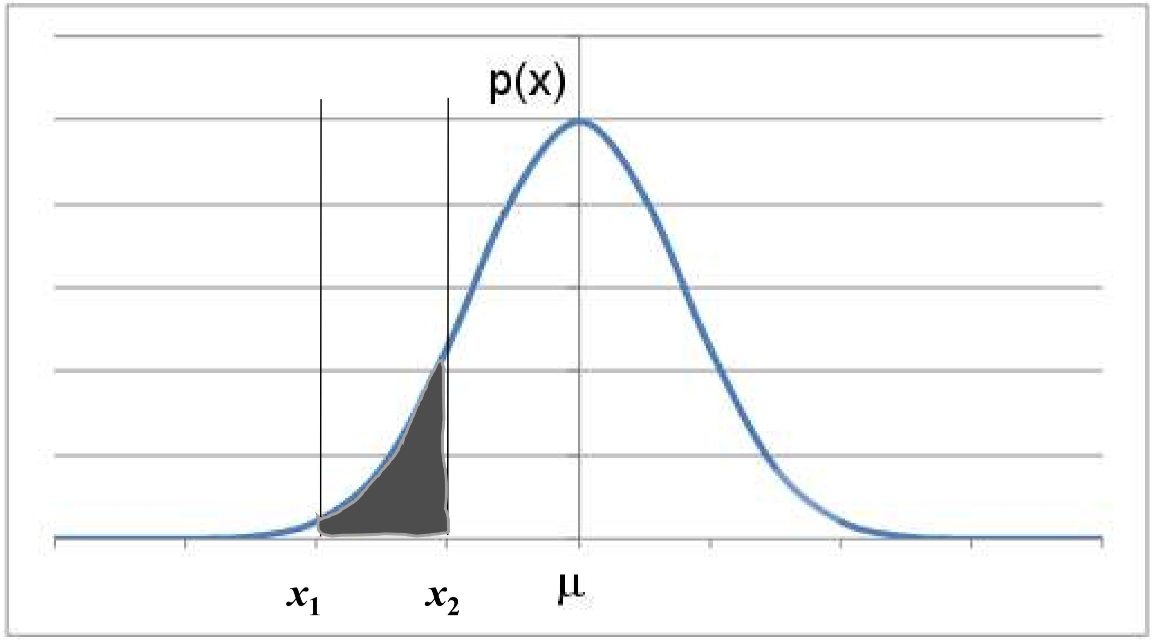
Mediana:

Per una distribuzione continua: valore che divide l'area sottesa dalla curva di distribuzione in due parti uguali

Per una distribuzione discreta: valore centrale dei valori ordinati

Per una distribuzione simmetrica (come quella di Gauss), media, moda e mediana coincidono.

Variabile Casuale Normale
Curva di Gauss



Probabilità:

$$\int_{x_1}^{x_2} p(x)dx = P\{x_1 < X < x_2\}$$

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

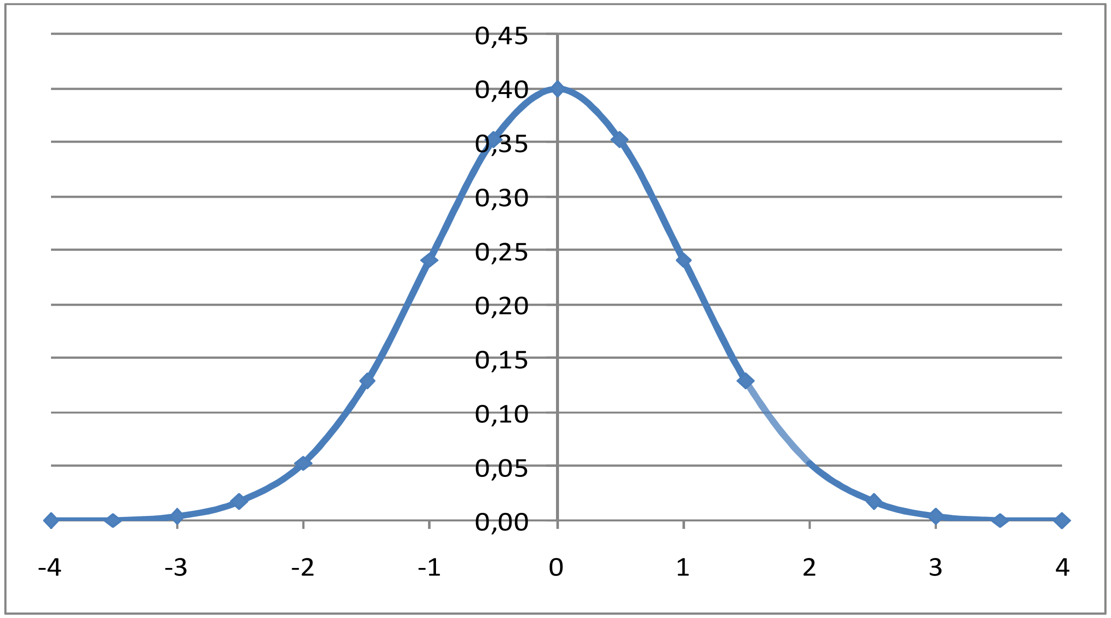
68,3% = P{ μ - 1,00 σ < X < μ + 1,00 σ }
95,0% = P{ μ - 1,96 σ < X < μ + 1,96 σ }
95,5% = P{ μ - 2,00 σ < X < μ + 2,00 σ }
99,0% = P{ μ - 2,58 σ < X < μ + 2,58 σ }
99,7% = P{ μ - 3,00 σ < X < μ + 3,00 σ }

$x_0 \pm \Delta$

$\Downarrow$

$\mu \pm 3\,\sigma$

Variabile Casuale Normale
Curva di Gauss standard



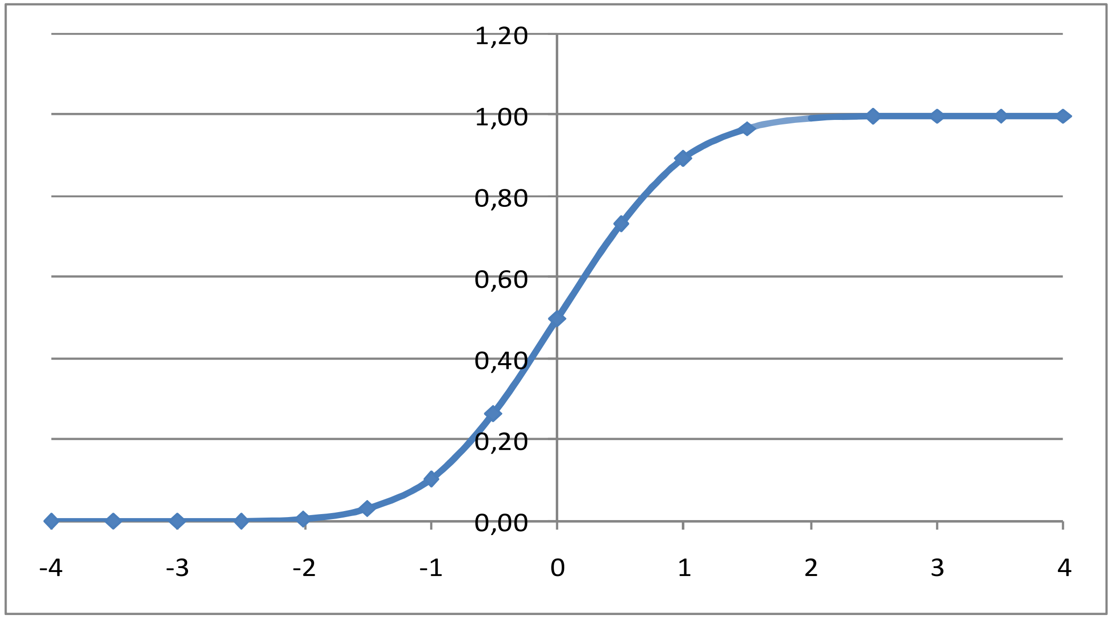
$\sigma = 1, \mu = 0$

$$\sigma \cdot f(z) = \left(\frac{1}{\sqrt{2\pi}} \right) \cdot e^{-z^2/2}$$

$$f(-z) = f(z)$$

68,3% = P{- 1 < z < + 1}
etc.

Variabile Casuale Normale
Curva cumulata di Gauss



Probabilità:

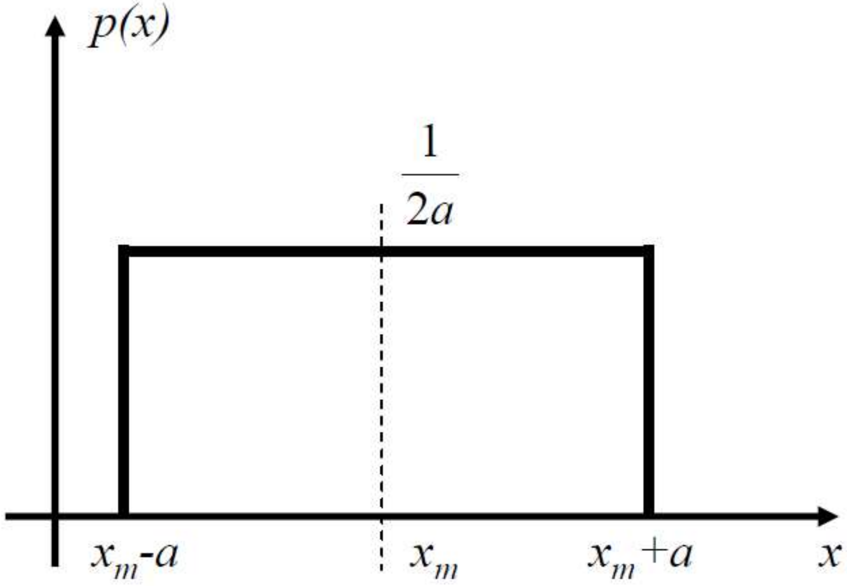
$$F(z) = \int_{-\infty}^{z} f(x)dx$$

$$F(-z) = 1 - F(z)$$

$$\int_{x_1}^{x_2} p(x)dx = F(x_2) - F(x_1)$$

Variabile Casuale Uniforme
Distribuzione rettangolare

Si usa quando tutti i possibili valori di misura hanno la stessa probabilità.
Dalla definizione di varianza:



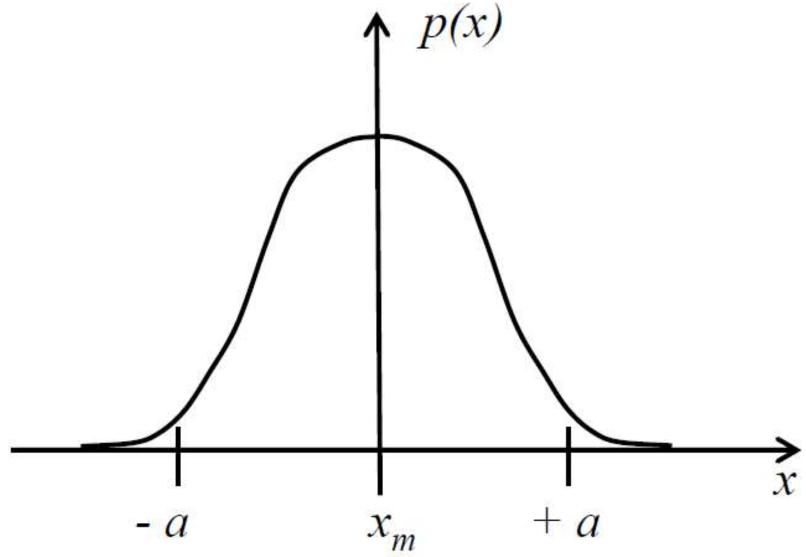
$$\sigma^2 = \int_{-\infty}^{+\infty} (x - x_m)^2 \, p(x) dx = \int_{x_m - a}^{x_m + a} (x - x_m)^2 \frac{1}{2a} dx =$$

$$= \frac{1}{2a} \left[\frac{(x - x_m)^3}{3} \right]_{x_m - a}^{x_m + a} = \frac{a^2}{3}$$

L'incertezza è dunque: $\sigma = \frac{a}{\sqrt{3}} = u$ **(u: uncertainty)**

Variabile Casuale Normale
Distribuzione di Gauss

Si usa quando i valori di misura vicini alla media hanno maggiore probabilità di quelli lontani.



A differenza della distibuzione rettangolare i possibili valori (teorici) non sono limitati; considerando l'intervallo col 95,5% di probabilità (±2σ):

$$\sigma = \frac{a}{2} = u$$

Distribuzione di Gauss per campioni discreti

Esempio: diametro di 500 pistoni prodotti in officina in un certo giorno.

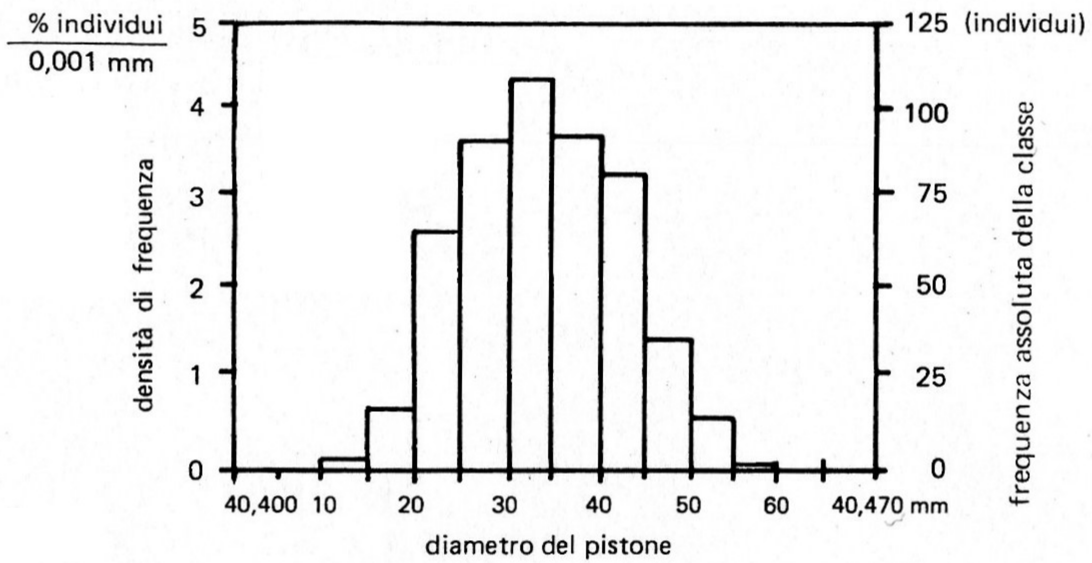
Classe (mm)		Frequenza assoluta	Frequenza %	Frequenza cumulata %	Densità di frequenza $\left(\frac{\text{(\% pistoni)}}{\mu m}\right)$
più di	meno di				
40,410	40,415	3	0,6	0,6	0,12
15	20	17	3,4	4,0	0,68
20	25	66	13,2	17,2	2,64
25	30	89	17,8	35,0	3,56
30	35	109	21,8	56,8	4,36
35	40	92	18,4	75,2	3,68
40	45	73	14,6	89,8	2,92
45	50	34	6,8	96,6	1,36
50	55	15	3,0	99,6	0,60
40,455	40,460	2	0,4	100,0	0,08

densità media (2%)/1 μm

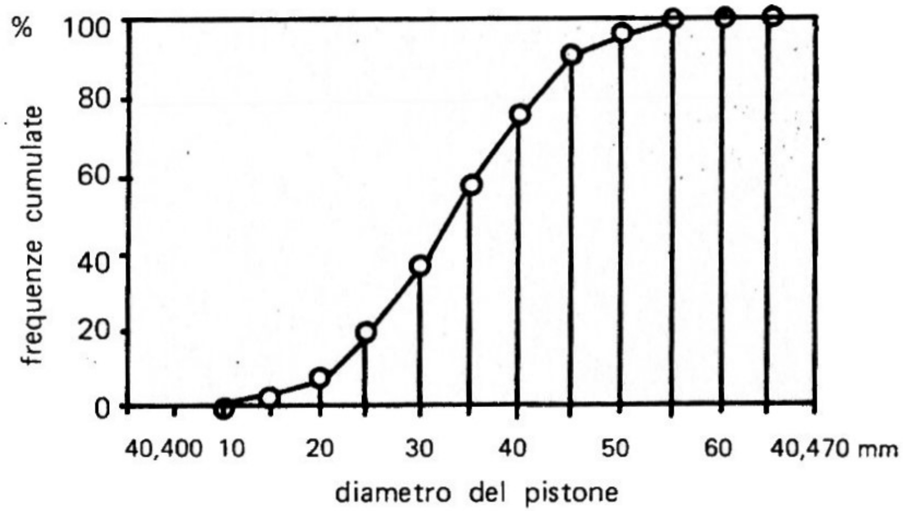
Superficie totale
dens. media $\times$ 50 μm = 100% = 1

Distribuzione di Gauss per campioni discreti

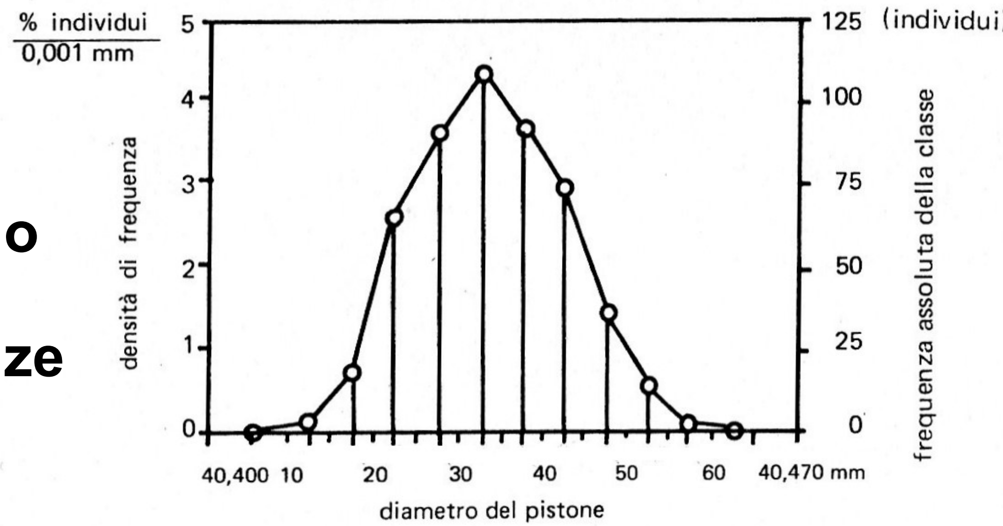
Istogramma



Cumulata



Poligono delle frequenze



Classificazione secondo UNI – CEI – ENV 13005

- **Incertezza di tipo A – riducibili ripetendo le misure – legate ad un effettivo contenuto informativo che cresce all'aumentare del numero delle misure – (errori casuali; seguono la distribuzione di Gauss) – analisi statistica di una serie di misure**

- **Incertezza di tipo B – non riducibili ripetendo le misure – insensibili al ripetersi delle misure (es.: seguono la distribuzione rettangolare) – analisi teorica della distribuzione**

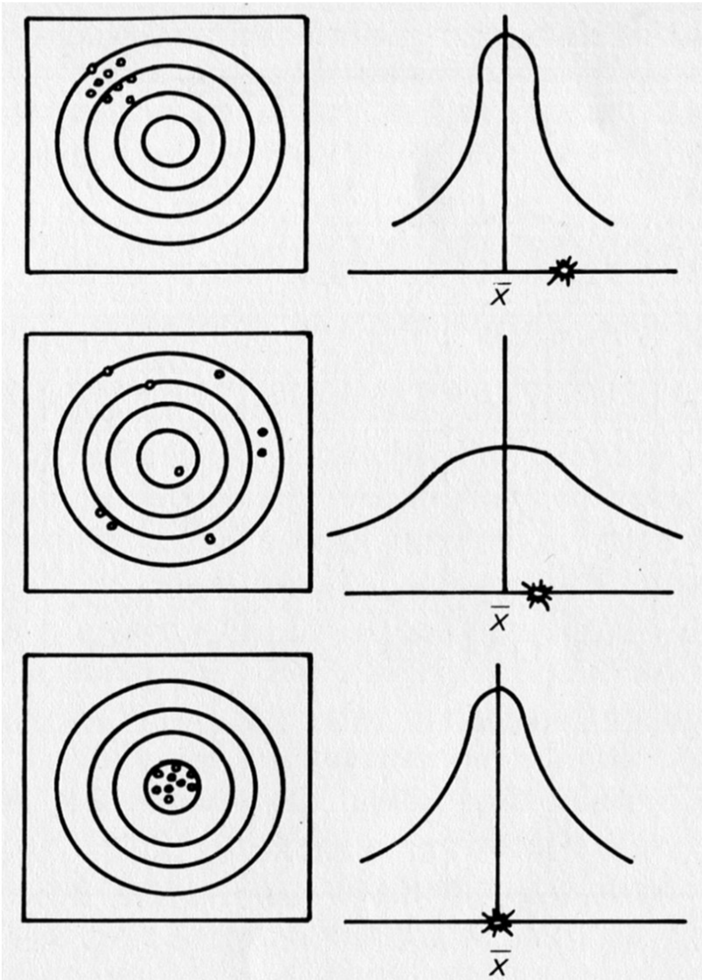
Teorema del Limite Centrale

- **La combinazione di un grande numero di variabili aleatorie X_i indipendenti, indipendentemente dalle forme distributive, è distribuita approssimativamente come una variabile casuale normale**

$$\lim_{n \to \infty} \sum_{i=1}^{n} X_i = N(\mu, \sigma^2),$$

Variabile Casuale Normale
Curva di Gauss

Esempio dei 3 fucili



Ripetibilità

Accuratezza

Precisione

Probabilità - Frequenza

- **Probabilità *p* : un evento *E* può presentarsi in *h* modi su di un totale di *N* ugualmente probabili modi possibili**

$$p_E = \frac{h}{N}$$

- **Frequenza *f* : un evento *E* che si è presentato *a* volte su di un totale di *n* eventi**

$$f_E = \frac{a}{n}$$

Legge empirica del caso

- **In una serie di prove, ripetute un grande numero di volte nelle stesse condizioni, ciascuno dei risultati possibili si manifesta con una frequenza relativa che approssima il valore della probabilità, l'approssimazione *ordinariamente* cresce con il numero delle prove**

$$f_E \xrightarrow{n=\infty} p_E$$

Legge empirica del caso

- **Frequenza : risultanze sperimentali ⇨ *a posteriori***
- **Probabilità : teoria assiomatica ⇨ *a priori***

- **Legge Empirica del Caso:** $$f_E \xrightarrow{n=\infty} p_E$$

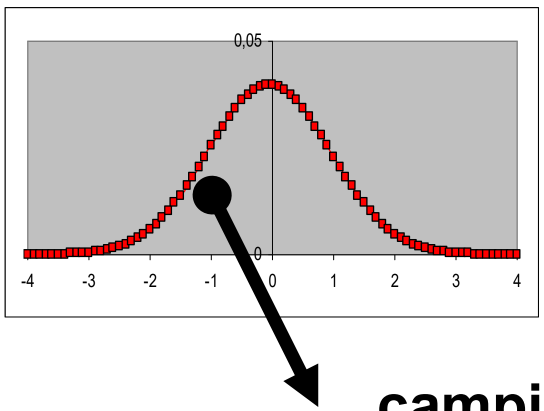
- **Frequenza come stima della Probabilità**
- **Probabilità come previsione della Frequenza**

Proprietà della probabilità (eventi A e B indipendenti):

$$p(A \cup B) = p(A) + p(B) \qquad p(A \cap B) = p(A) \cdot p(B)$$

Misure ripetute - Campione

Universo / popolazione



⇨ **μ , σ , ∞**

campione

x_1			x_i			x_n

⇨ **m , s_{n-1} , n**

$$m = \frac{\sum x_i}{n} \qquad s_n^2 = \frac{\sum (x_i - m)^2}{n}$$

$$s_{n-1}^2 = \frac{\sum (x_i - m)^2}{n-1}$$

$$m \rightarrow \mu$$

$$s_{n-1} \rightarrow \sigma$$

Campione

$$x_{rif} \pm \Delta\ incertezza \Rightarrow m \pm 3\, s_{n-1}$$

- **Insieme di misure $x_1\ x_2\ \ldots\ldots\ x_i\ \ldots\ldots\ x_n$**
- **Valore di "riferimento" x_0 = media dell'universo $\mu \cong$ media delle misure m**
- **Dispersione / incertezza Δ delle misure = distribuzione errori; se considero un intervallo di confidenza al 99,7%**

$$\Longrightarrow \quad 3\,\sigma \cong 3\, s_{n-1}$$

- **N.B. : σ rappresenta l'incertezza con una confidenza del 68,4%**

Incertezza

- **L'incertezza decresce all'aumentare del numero *n* delle misure fino al suo limite "intrinseco"**
  - **limite "intrinseco" = procedura usata e ambiente & strumento (precisione dello strumento)**

- **Incertezza "ottimale" - compatibile con l'uso - contiene l'"ottimale" livello di contenuto di informazioni**

Misure indirette - Propagazione degli errori

- **Se la grandezza da valutare non è misurata direttamente ma misurando altre grandezze, ad essa legate da una relazione analitica:**

$$y = f(x_1, x_2, ..., x_n)$$

la sua incertezza si può derivare dall'incertezza delle singole variabili indipendenti x_i.

Se $\overline{x}_i$ è il risultato della misura i-esima e $u_i(x)$ l'incertezza corrispondente, allora

$$\overline{y} = f(\overline{x}_1, \overline{x}_2, ..., \overline{x}_n)$$

rappresenta il risultato della misura indiretta con incertezza $u_c(y)$, ricavabile dalle singole incertezza e dalla funzione che lega le grandezze.

Misure indirette - Propagazione degli errori

$$x_i = \bar{x}_i \pm \Delta x_i$$

$$\bar{y} = f(\bar{x}_1, \bar{x}_2, ..., \bar{x}_n) \quad \Rightarrow$$

$$\bar{y} + \Delta y = f(\bar{x}_1 + \Delta x_1, \bar{x}_2 + \Delta x_2, ..., \bar{x}_n + \Delta x_n) =$$

(sviluppando in serie di Taylor)

$$= f(\bar{x}_1, \bar{x}_2, ..., \bar{x}_n) + \frac{\partial f}{\partial x_1} \cdot \Delta x_1 + \frac{\partial f}{\partial x_2} \cdot \Delta x_2 + ... + \frac{\partial f}{\partial x_n} \cdot \Delta x_n \quad \Rightarrow$$

$$\Delta y = \frac{\partial f}{\partial x_1} \cdot \Delta x_1 + \frac{\partial f}{\partial x_2} \cdot \Delta x_2 + ... + \frac{\partial f}{\partial x_n} \cdot \Delta x_n$$

Incertezza combinata (errore assoluto)

Propagazione degli errori- Errore massimo

- **Differenziale logaritmico**

$$G = a^{\alpha} b^{\beta} c^{\gamma}$$

$$\ln G = \alpha \ln a + \beta \ln b + \gamma \ln c$$

$$\frac{\Delta G}{G} = \alpha \frac{\Delta a}{a} + \beta \frac{\Delta b}{b} + \gamma \frac{\Delta c}{c}$$

$$\Delta G = \left(\alpha \frac{\Delta a}{a} + \beta \frac{\Delta b}{b} + \gamma \frac{\Delta c}{c} \right) G$$

Legge di Propagazione degli errori casuali di Gauss

$$z = f(x, y)$$

$$z_i = f(\bar{x} + \Delta x_i, \bar{y} + \Delta y_i)$$

$$\Delta z_i = z_i - \bar{z} = (\frac{\partial z}{\partial x}\Delta x_i + \frac{\partial z}{\partial y}\Delta y_i)$$

$$\sigma_z^2 = \frac{\sum_1^n \Delta_z^2}{n-1} \qquad \sigma_z^2 = \left(\frac{\partial z}{\partial x}\right)^2_{x=\bar{x}} \frac{\sum_1^n \Delta_x^2}{n-1} + \left(\frac{\partial z}{\partial y}\right)^2_{y=\bar{y}} \frac{\sum_1^n \Delta_y^2}{n-1}$$

$$\sigma_z^2 = \left(\frac{\partial z}{\partial x}\right)^2_{x=\bar{x}} \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2_{y=\bar{y}} \sigma_y^2$$

Misure indirette - Propagazione degli errori

In termini di medie e di scarti (deviazione standard):

$$\bar{y} = f(\bar{x}_1, \bar{x}_2, ..., \bar{x}_n)$$ **(stima della media)**

$$\sigma_y^2 = \Sigma \left(\frac{\partial y}{\partial x_i} \right)_0^2 \sigma_{x_i}^2$$

Esempi:

- **y = a ± b** ⇨ $$\sigma_y^2 = \sigma_a^2 + \sigma_b^2$$
- **y = a · b** ⇨ $$\sigma_y^2 = b^2 \cdot \sigma_a^2 + a^2 \cdot \sigma_b^2$$

Misure indirette - Propagazione degli errori

Esempi:

- **y = a · b ·c** $\Longrightarrow$ $$\sigma_y^2 = (bc)\cdot\sigma_a^2 + (ca)\cdot\sigma_b^2 + (ab)\cdot\sigma_c^2$$

- **y = a**$^{\alpha}$ $\Longrightarrow$ $$\sigma_y^2 = (\alpha\cdot a^{\alpha-1})^2\cdot\sigma_a^2$$

Incertezza estesa

Se sono coinvolte più grandezze la distribuzione tende alla gaussiana e σ ha il significato di incertezza "standard" $u(y)$, con una confidenza del 68,4%.

Se vogliamo maggiore certezza si deve moltiplicare $u(y)$ per un fattore di copertura k.

$$U(y) = k \cdot u(y)$$

Il fattore di copertura k dipende dalla distribuzione di probabilità ipotizzata.

Esempio

Verifica di un dinamometro usando un altro dinamometro

La differenza fra i valori letti e quelli "veri" è dovuta a due termini:

- **errore del dinamometro in prova (es. classe di precisione ±1%);**
- **errore del dinamometro di riferimento (es. classe di precisione ±0,1%).**

±1% significa che nel 99,7 % dei casi l'errore è inferiore a 1% :

$$\pm 3\sigma_1 = \pm 1\ \% \rightarrow \sigma_1 = 0{,}33\ \%$$

±0,1% significa che nel 99,7 % dei casi l'errore è inferiore a 0,1% :

$$\pm 3\sigma_2 = \pm 0{,}1\ \% \rightarrow \sigma_2 = 0{,}033\ \%$$

Scarto combinato:

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2} = 0{,}332\,\%$$

Usando uno strumento 10 volte più preciso dell'altro, l'errore è coperto da quello meno preciso.

Esempio di stima dell'incertezza composta – tubo di pitot - 01

Si consideri un tubo di Pitot immerso in una vena di aria; nei limiti dell'ipotesi di fluido non viscoso (R_e elevato) od incomprimibile (Mach basso), si ha

$$C = \sqrt{2\,\frac{\Delta p}{\rho}} = \sqrt{2\,\Delta_p\,\frac{T_a R}{P_a}} = \left(2\,\Delta_p\,R\,\frac{T_a}{P_a}\right)^{1/2}$$

essendo $\Delta_p = P_T - P_a$

Le variabili indipendenti sono Δ_p, T_a, P_a. I disturbi possibili sono:

- disallineamento della sonda con il flusso
- perdite di fluido nelle giunzioni dei tubi
- variazioni nel diametro o nelle condizioni superficiali del manometro
- fluttuazioni di pressione P_a, temperatura T_a, pressione totale P_T

Valutiamo le derivate:

$$\frac{\partial C}{\partial \Delta_p} = \frac{1}{2} \left(2R \frac{T_a}{P_a}\right)^{1/2} \Delta_p^{-1/2} \qquad \frac{\partial C}{\partial P_a} = -\frac{1}{2}(2RT_a\Delta_p)^{1/2} P_a^{-3/2}$$

$$\frac{\partial C}{\partial T_a} = \frac{1}{2} \left(2R \frac{\Delta p}{P_a}\right)^{1/2} T_a^{-1/2}$$

Si ha perciò

$$W_C = \sqrt{\frac{1}{4} \frac{2RT_a}{P_a \Delta_p} W_{\Delta p}^{2} + \frac{1}{4} \frac{2RT_a\Delta_p}{P_a^{3}} W_{Pa}^{2} + \frac{1}{4} \frac{2R\Delta_p}{P_a T_a} W_{Ta}^{2}}$$

Spesso si utilizza l'incertezza relativa, data da:

$$\frac{W_C}{C} = \sqrt{\left(\frac{1}{2} \frac{W_{\Delta p}}{\Delta_p}\right)^2 + \left(\frac{1}{2} \frac{W_{Pa}}{P_a}\right)^2 + \left(\frac{1}{2} \frac{W_{Ta}}{T_a}\right)^2}$$

Si suppone di misurare T_a con un termometro a mercurio, Δ_p con un manometro a ~~V~~ differenziale, P_a con un manometro a tubo di Bourdon. Le incertezze delle misure di base, sempre 20:1, sono:

$\Delta_p = 203 \pm 2.5$ mm H_2O (1.25)   20:1

$T_a = 292.8 \pm 0.1$ °K (0.05)

$P_a = 101000 \pm 2000$ P_a assoluti (1000)

Sostituendo i valori numerici, si ha

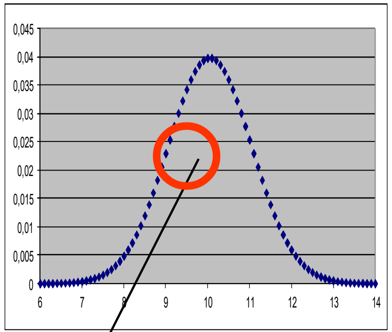
$$\frac{W_C}{C} = \pm .01166 = \sqrt{.00616^2 + .0099^2 + .00001^2}$$

termine da $W_{\Delta p}$ — da W_{Pa} — da W_{Ta}

E' interessante notare come, ad esempio, un termometro più accurato non migliorerebbe la situazione; mentre un barometro più accurato; $W_{Pa} = 500\ P_a$ (±250), darebbe

$$\frac{W_C}{C} = .0066.$$

Distribuzione delle medie



Universo: μ, σ, ∞

Campioni,

X_1			X_i			X_n

m_1, n

X_1			X_i			X_n

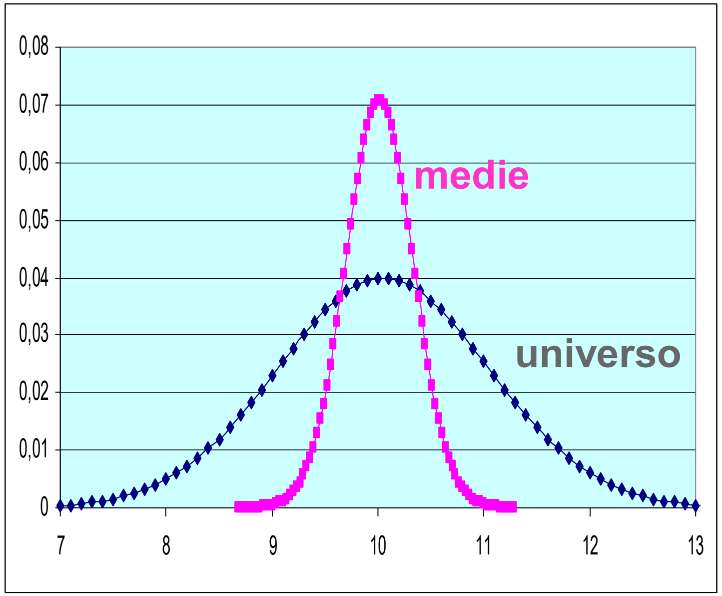
m_2, n

X_1			X_i			X_n

m_i, n

X_1			X_i			X_n

m_∞, n



medie μ, σ/√n , ∞

$$\mu - 3\frac{\sigma}{\sqrt{n}} \leq m_i \leq \mu + 3\frac{\sigma}{\sqrt{n}}$$

Distribuzione di Student

Generalmente sono noti *m* e *s*.

Problema inverso: determinare l'intervallo in cui può cadere la media della popolazione μ noti la media *m* e lo scarto *s* di un campione.

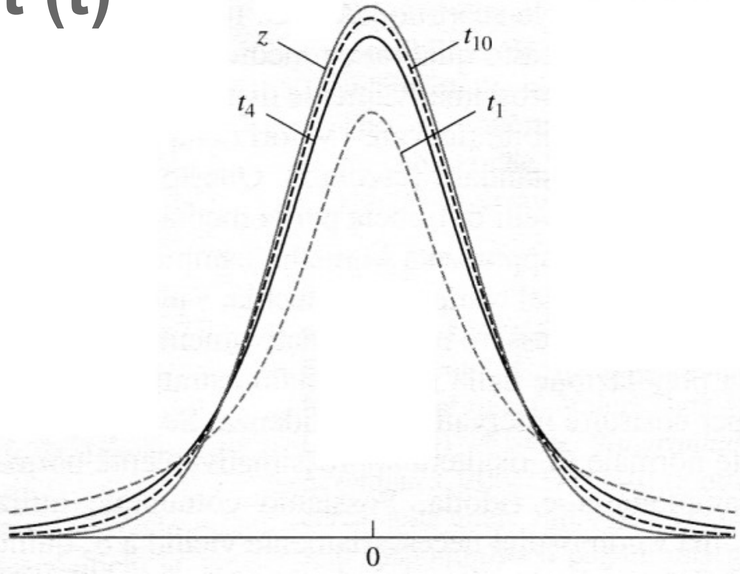
$$m - t\frac{s}{\sqrt{n}} \leq \mu \leq m + t\frac{s}{\sqrt{n}}$$

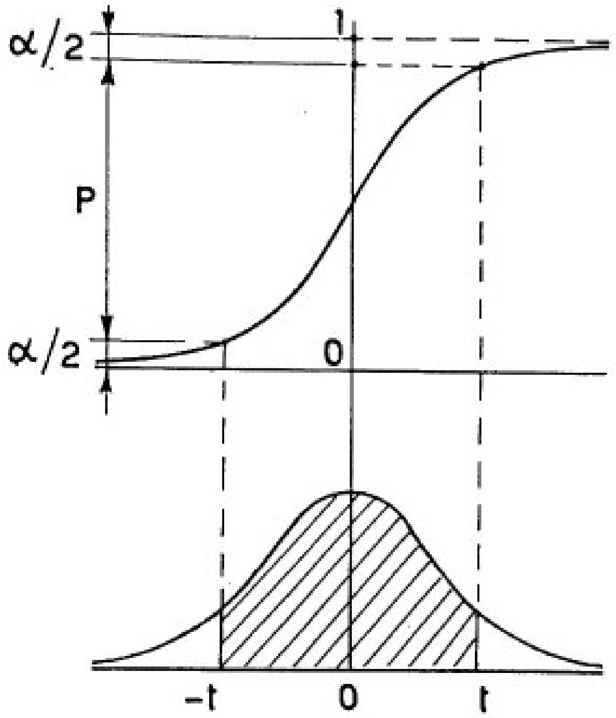
con t (detta t di Student) → z se n → ∞

***m* ed *s* sono una stima di μ e σ**

Distribuzione di Student (t)

E' funzione del numero di gradi di libertà:





ν	P = 10%	20	30	40	50	60	70	80	90	95	98	99
1	·158	·325	·510	·727	1·000	1·376	1·963	3·078	6·314	12·706	31·821	63·657
2	·142	·289	·445	·617	·816	1·061	1·386	1·886	2·920	4·303	6·965	9·925
3	·137	·277	·424	·584	·765	·978	1·250	1·638	2·353	3·182	4·541	5·841
4	·134	·271	·414	·569	·741	·941	1·190	1·533	2·132	2·776	3·747	4·604
5	·132	·267	·408	·559	·727	·920	1·156	1·476	2·015	2·571	3·365	4·032
6	·131	·265	·404	·553	·718	·906	1·134	1·440	1·943	2·447	3·143	3·707
7	·130	·263	·402	·549	·711	·896	1·119	1·415	1·895	2·365	2·998	3·499
8	·130	·262	·399	·546	·706	·889	1·108	1·397	1·860	2·306	2·896	3·355
9	·129	·261	·398	·543	·703	·883	1·100	1·383	1·833	2·262	2·821	3·250
10	·129	·260	·397	·542	·700	·879	1·093	1·372	1·812	2·228	2·764	3·169

(ν = n-1)

Serie di misure

$x_1\ x_2\ .....\ x_i\ .....\ x_n$

- n numerosità misure
- m = valore di riferimento
- s_{n-1} = incertezza sulle misure
- $s_{n-1}/\sqrt{n}$ = incertezza stima della media

$$m = \frac{\sum x_i}{n}$$

$$s_{n-1}^2 = \frac{\sum (m - x_i)^2}{n-1}$$

$$\frac{s_{n-1}^2}{n} = \frac{\sum (m - x_i)^2}{n\,(n-1)}$$

$$m - 3s_{n-1} \le \ x_i \le m + 3s_{n-1}$$

$$m - t\frac{s_{n-1}}{\sqrt{n}} \le \ \mu_i \le m + t\frac{s_{n-1}}{\sqrt{n}}$$

Esempio: sfere - diametro [mm]

53,0
57,3
52,6
54,5
60,9
56,4
58,1
57,5
55,4
58,9

- **Media** $m = 56{,}5$
- **Numerosità** $n = 10$
- **Dev. st.** $s_{n-1} = 2{,}62$ $3 \cdot s_{n-1} = 7{,}9$
- **Dev. st. media** $s_{n-1}/\sqrt{n} = 0{,}83$ $3 \cdot s^* = 2{,}5$

- **Media** 54 – 56,5 – 59
- **Diametri** $48{,}6 \div 64{,}4$

$$\Phi \quad 56{,}5 \pm 7{,}9$$

Rigetto dei valori anomali (errori parassiti)

Criterio di Chauvenet

- **Un valore può essere considerato "anomalo" se la sua deviazione rispetto alla media è tale da avere una probabilità inferiore ad un predeterminato livello**

- **Principio di Chauvenet: se in una serie di N valori la probabilità che uno di essi abbia uno scostamento s_i rispetto alla media maggiore di un certo valore calcolato s_p è ½, allora tutte le misurazioni che presentano uno scostamento maggiore si dovrebbero scartare.**

- **Se P è la probabilità che $s_i < s_p \rightarrow$ 1-P è la probabilità che $s_i > s_p$; se i valori sono N la probabilità complessiva è N (1-P); quando questa è pari a ½ si ha il limite:**

$$N(1 - P) = \frac{1}{2} \rightarrow P = 1 - 1/2N$$

Rigetto dei valori anomali (errori parassiti)

Criterio di Chauvenet

- In pratica N(1 – P) rappresenta il numero di valori attesi fuori dall'intervallo corrispondente a P (area sottesa dalla curva di Gauss in un dato intervallo).

- Ad esempio se N = 10 e considero l'intervallo ± σ interno alla media (P = 0,68):

$$N(1 - P) = 3{,}2$$

Se considero l'intervallo ± 3σ (P=0,997):

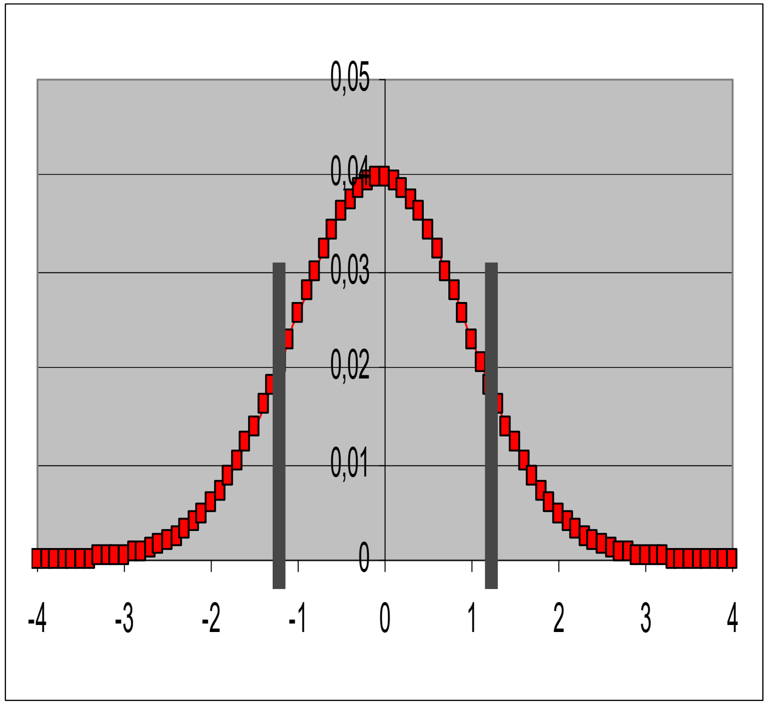
$$N(1 - P) = 0{,}03$$

Il criterio di Chauvenet definisce ½ come numero limite di valori attesi fuori dall'intervallo.

Criterio di Chauvenet

Procedura:

- x_{min} ……x_i …….. x_{max}
- numerosità n ; media m; scarto s_{n-1}
- P = 1 – 1 / 2n
- P ⇨ *z*
- Intervallo di accettabilità:
  - $m - zs_{n-1} \vdash x_i \dashv m + zs_{n-1}$
- Campione depurato



$\mu \pm z\sigma$

Criterio di Chauvenet

Cosa fare se si trova una misura anomala:

- **si elimina e si ridefinisce il campione;**
- **si esegue ricorsivamente il controllo secondo Chauvenet finché tutti i valori sono compresi nell'intervallo di accettabilità;**
- **considero non significativo tutto il campione (soprattutto se N è piccolo).**

Altre tecniche di rigetto:

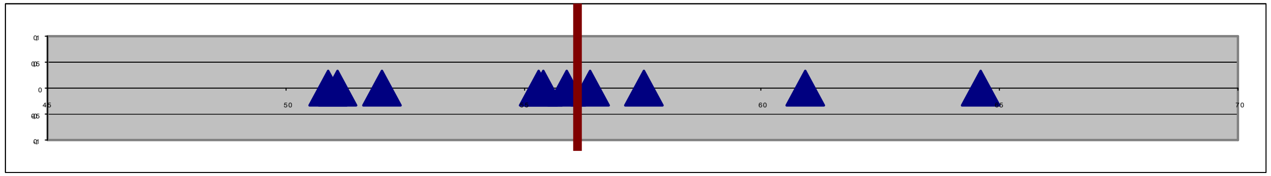
- **scartare il valore più basso e il più alto o il più distante dalla media dei due;**
- **eliminare le "code di Gauss" (cioè i valori fuori da ±3σ, meno di 0,3% di probabilità).**

Criterio di Chauvenet

Il criterio dovrebbe essere applicato solo a distribuzioni gaussiane; in molti casi questa condizione è soddisfatta.

Possibili cause di non normalità:
- **numero di osservazioni insufficiente;**
- **non omogeneità della popolazione;**
- **presenza di errori sistematici;**
- **cause di ordine matematico;**
- **effetto di deriva;**
- **etc.**

Esempio: sfere



m

51,1
55,3
52,0
55,9
60,9
56,4
64,6
57,5
55,4
50,9

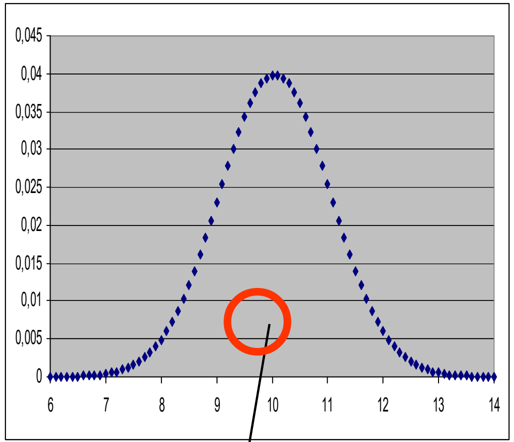
- Media: $m = 56{,}0$
- Numerosità: $n = 10$
- Dev. st.: $s_{n-1} = 4{,}3$
- $P(z) = 1-1/(2\cdot 10)$ P = 95% ⇨

$$z = 1{,}96$$

- Intervallo di accettabilità: 47,6 ├ ─ ┤ 64,4
- Campione ridotto: n=9

  m = 55 dev.st. = 3,3

Medie di campioni estratti dallo stesso universo



Universo $x_i \Rightarrow [\mu, \sigma, \infty]$

Campioni, numerosità = n

x_1			x_i			x_n

$m_i \Rightarrow [\mu, \sigma/\sqrt{n}, \infty]$

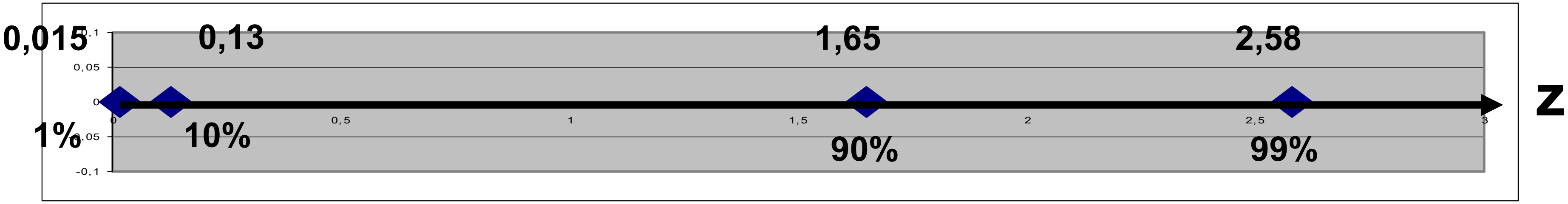
Differenze $z_{ij} = m_i - m_j$

$\Rightarrow \quad [0, \sigma\cdot\sqrt{(2/n)}, \infty]$

Confronto fra le medie

m^*_i m^*_j (ipotesi: stesso universo)

$z^* = |m^*_i - m^*_j| / \sigma^*$



accettabile   prob. non corretto   non corretto

⇨ m^*_i "equivalente" m^*_j

Confronto fra le medie

- ❶

X_1			X_i			X_n

m_1, $s_{m1} = s_{n-1} / \sqrt{n}$

- ❷

X_1			X_i			X_m

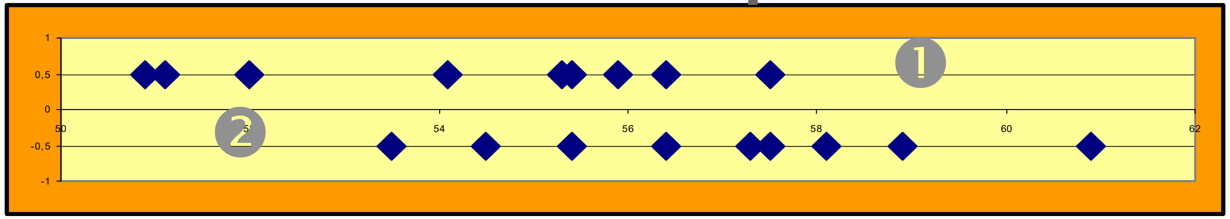
m_2 , $s_{m2} = s_{m-1} / \sqrt{m}$

$z_{12} = m_1 - m_2$

$$\sigma_z = \sqrt{\left(\sigma_1^2 + \sigma_2^2\right)} = \sqrt{\left(\frac{s_{n-1}^2}{n} + \frac{s_{m-1}^2}{m}\right)}$$

Se $\dfrac{z_{12}}{\sigma_z} \leq 1{,}96 \Longrightarrow$ **accettabile 95%**

Confronto fra le medie: esempio



sfere - diametro [mm]

➊	➋
51,1	53,4
55,3	57,3
52,0	54,4
54,1	60,9
56,4	56,4
57,5	58,1
55,4	57,5
50,9	55,4
55,9	59,6

➊	➋
m = 54,3	m = 57,0
n = 9	n = 9
s_{n-1}= 2,41	s_{n-1}= 2,40
$s_{n-1}/\sqrt{n}$ = 0,8	$s_{n-1}/\sqrt{n}$ = 0,8
diff / $\sqrt{(s^{*2}_1 + s^{*2}_2)}$	= 2,7/1,29 = 2,1

2,1 > 1,96 **m ➊ ≠ m ➋**

Analisi di regressione e correlazione

- **Correlazione**:

analizza se esiste una relazione tra due variabili (come e quanto due variabili variano insieme)

- **Regressione**

analizza la forma della relazione tra le variabili

Analisi di regressione

$$y = f(x_i, \beta_i)$$

- serie *j* di misure (x_i^j ; y^j - es. : I / O di uno strumento)
- ipotizzata la forma di f

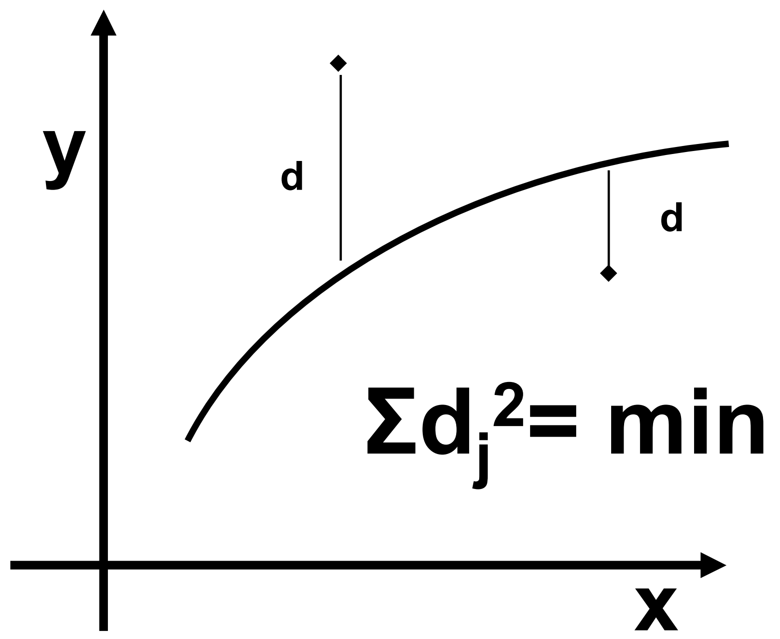
- stimare i valori β_i in f
- avere una stima della “bontà” della scelta di f

Metodo dei Minimi Quadrati

- **j misure** $y^j \; ; x_i^j$

$$\sum_j \left[y^j - f\left(x_i^j, \beta_i\right)\right]^2 = \min$$

$$\frac{\partial \sum_j \left[y^j - f\left(x_i^j, \beta_i\right)\right]^2}{\partial \beta_i} = 0$$



- sistema di equazioni nelle incognite β_i

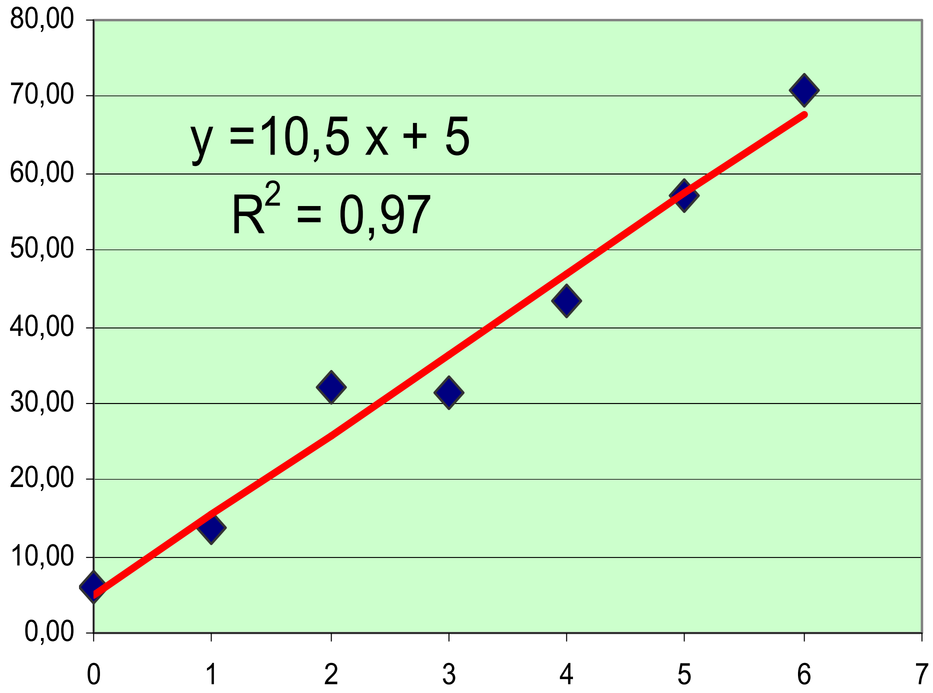
Retta di best fit $y = m x + b$

- N punti (x_i, y_i)

$$\sum\left(\left(mx_i + b\right) - y_i\right)^2 = \min$$

$$\frac{\partial \Sigma}{\partial m} = 0 \qquad \frac{\partial \Sigma}{\partial b} = 0$$

$$\begin{cases} \sum x_i\left(mx_i + b - y_i\right) = 0 \\ \sum \left(mx_i + b - y_i\right) = 0 \end{cases}$$



Coefficiente di correlazione R^2 [0 - 1]

Errore della stima
$$s = \sqrt{\frac{\sum\left(Y - Y_{stim}\right)^2}{N-2}}$$

Errore sulla stima

$$s = \sqrt{\frac{\sum (Y - Y_{stim})^2}{N-2}}$$

- *s* “misura” la dispersione dei dati rispetto alla *f*

  $(Y - Y_{stim})$ = distanza (Punto - *f*) ;

  (N - n) gradi di libertà = N punti sperimentali - n parametri stimati (retta y=mx+b)

- *f* legame fisico = *s* errore sperimentale
- *f* correlazione “empirica” = dati sono intrinsecamente dispersi rispetto alla *f* di regressione

$$y = f(x_i, \beta_i)$$

- funzione di regressione
- definita a monte della sperimentazione

- legge fisica ⇨ “ n minimo ” punti sperimentali ; $R^2 = 0,80 - 0,99$; σ dispersione errori sperimentali
- relazione empirica ⇨ “molti” punti sperimentali ; $R^2 = 0,70 - 0,85$; σ dati intrinsecamente dispersi; funzione “ragionevolmente più semplice”

Regressione lineare

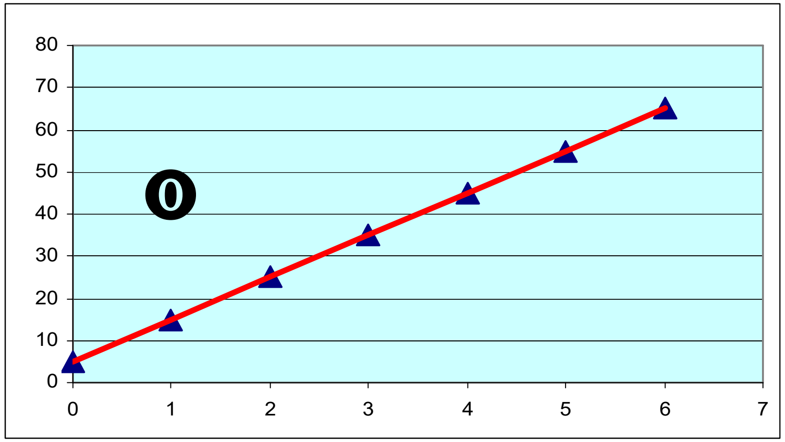
$Y = f(x_i, \beta_i)$	$Y = mx + b$
$Y = a\,x^b$	$\log Y = \log a + b \log x$
$Y = a\,b^x$	$\log Y = \log a + (\log b)\,x$
$Y\,x = a$	$\frac{1}{Y} = \frac{1}{a}\,x$

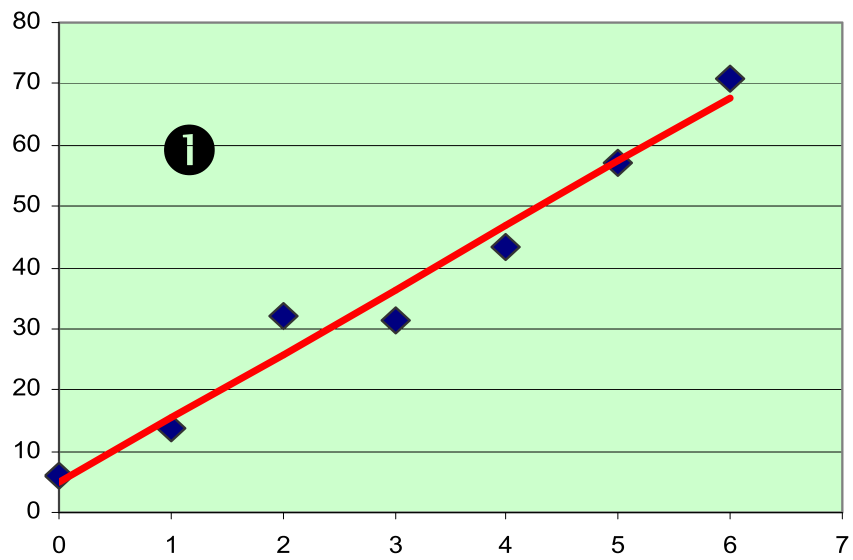
Punti (x_i, y_i)

- **Non conviene usare il polinomio interpolante.**
- **E' opportuno usare almeno un numero di punti 3 – 5 volte superiore al grado della curva.**
- **Si possono usare curve di grado superiore.**
- **Non conviene estrapolare!**
- **In caso di una taratura usare punti distribuiti su tutto il range di funzionamento dello strumento (a salire e a scendere).**
- **$x_i \rightarrow$ ingresso; $y_i \rightarrow$ uscita ; d_i : errore solo sulla y, incertezza sulla x (molto) minore dell'incertezza sulla y ⇨ retta x,y ≠ retta y,x**

- ⓿ retta $y = 10\,x + 5$; $R^2 = 1$
- ❶ retta ⓿ + rumore;

  $y = 10{,}5\,x + 5$; $R^2 = 0{,}97$

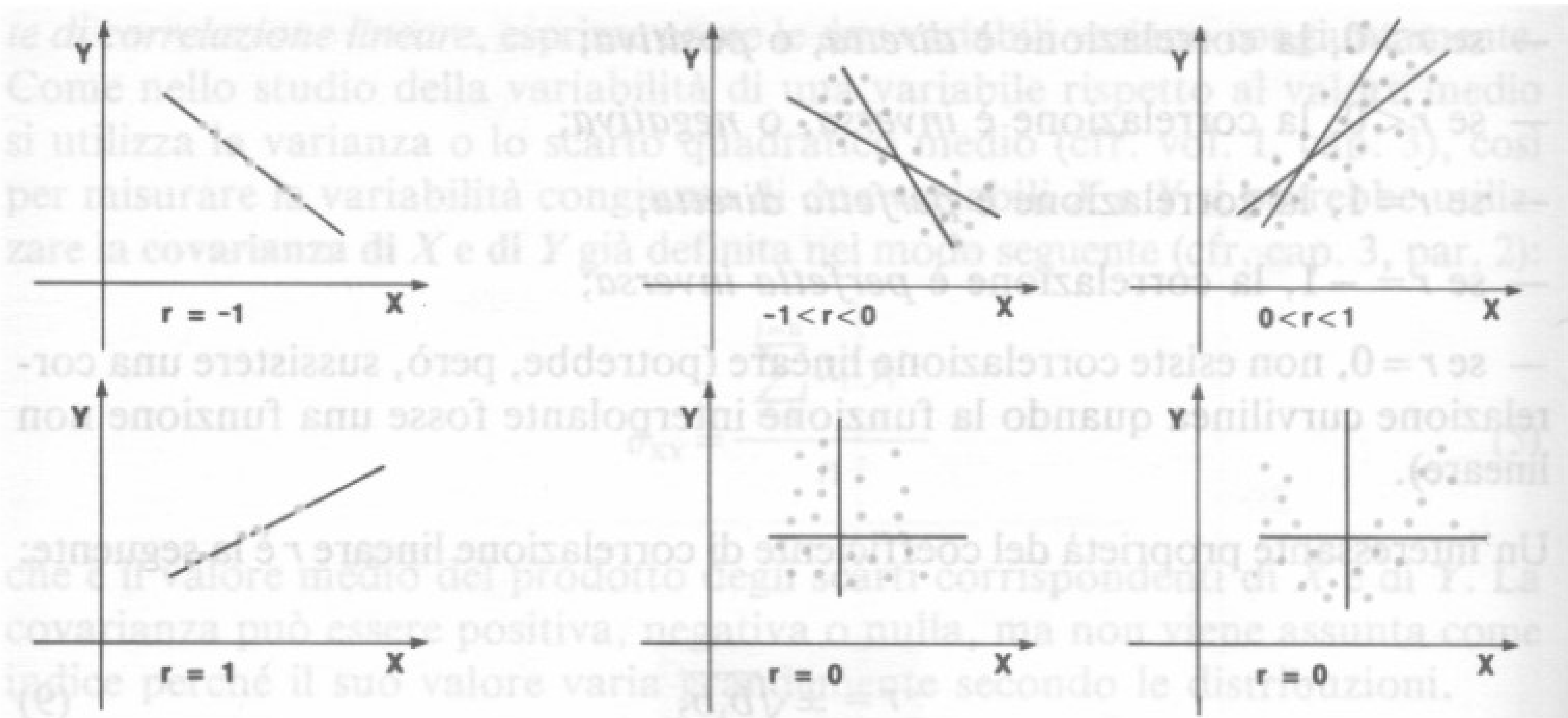




Correlazione

- **Coefficiente di correlazione**

$$-1 \leq \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{n}(y_i - \mu_y)^2}} \leq +1$$



Coefficiente di correlazione

- **Se $R^2 = 1$ le due grandezze (*x* e *y*) sono perfettamente correlate.**
- **Se $R^2 \rightarrow 1$ le due grandezze sono molto correlate.**
- **Se $R^2 \rightarrow 0$ le due grandezze sono poco correlate.**
- **Se $R^2 = 0$ le due grandezze non sono affatto correlate.**

Il coefficiente di correlazione non è una stima dell'errore!

Il coefficiente di correlazione rimane invariato se:

- **moltiplichiamo tutti i valori di una variabile per una costante positiva;**
- **sommiamo una costante a tutti i valori di una variabile;**
- **scambiamo i valori di *x* con quelli di *y*.**

Test di ipotesi

- **serie di eventi & modello di “comportamento” ⇨**
- **“eventi osservati” & “eventi previsti” ⇨**
- **frequenza osservata vs. teorica ⇨**
- **“significatività” della differenza ⇨**
- **accettazione / rigetto della ipotesi**

- **Test del chi – quadro: χ^2**

Test del chi - quadro χ^2

- **eventi organizzati in *k* classi (*k* possibili eventi)**
- **eventi osservati o_i in ciascuna classe**
- **eventi attesi e_i in ciascuna classe a fronte della ipotesi di distribuzione**

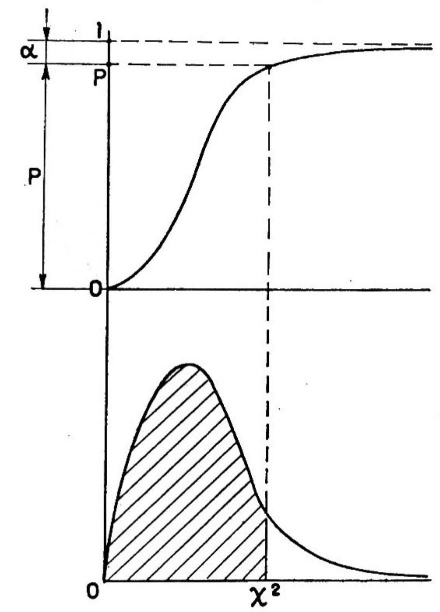
$$\chi_i^2 = \frac{(o_i - e_i)^2}{e_i}$$

$$\chi^2 = \sum_{i=1}^{k} \left[\frac{(o - e)^2}{e} \right]_i$$

$$\chi^2 \to 0 \quad \text{osservato} \equiv \text{previsto}$$

$$\chi^2 \to \infty \quad \text{osservato} \neq \text{previsto}$$

Distribuzione χ^2

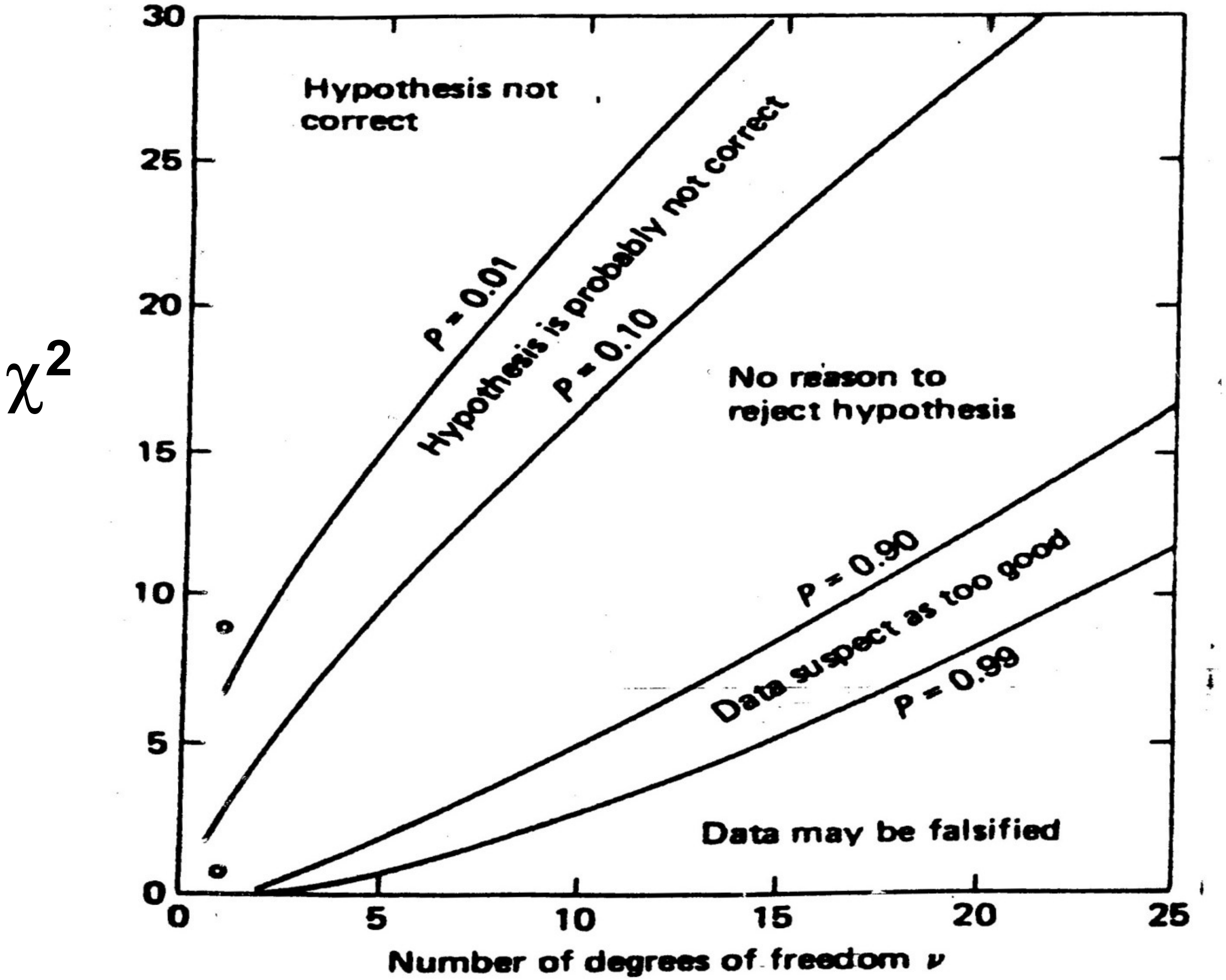


E' funzione dei gradi di libertà

ν	P = 1%	2	5	10	20	30	50	70	80	90	95	97	99
1	·000157	·000628	·00393	·0158	·0642	·148	·455	1·074	1·642	2·706	3·841	5·412	6·635
2	·0201	·0404	·103	·211	·446	·713	1·386	2·408	3·219	4·605	5·991	7·824	9·210
3	·115	·185	·352	·584	1·005	1·424	2·366	3·665	4·642	6·251	7·815	9·837	11·345
4	·297	·429	·711	1·064	1·649	2·195	3·357	4·878	5·989	7·779	9·488	11·668	13·277
5	·554	·752	1·145	1·610	2·343	3·000	4·351	6·064	7·289	9·236	11·070	13·388	15·086
6	·872	1·134	1·635	2·204	3·070	3·828	5·348	7·231	8·558	10·645	12·592	15·033	16·812
7	1·239	1·564	2·167	2·833	3·822	4·671	6·346	8·383	9·803	12·017	14·067	16·622	18·475
8	1·646	2·032	2·733	3·490	4·594	5·527	7·344	9·524	11·030	13·362	15·507	18·168	20·090
9	2·088	2·532	3·325	4·168	5·380	6·393	8·343	10·656	12·242	14·684	16·919	19·679	21·666
10	2·558	3·059	3·940	4·865	6·179	7·267	9·342	11·781	13·442	15·987	18·307	21·161	23·209
11	3·053	3·609	4·575	5·578	6·989	8·148	10·341	12·899	14·631	17·275	19·675	22·618	24·725
12	3·571	4·178	5·226	6·304	7·807	9·034	11·340	14·011	15·812	18·549	21·026	24·054	26·217
13	4·107	4·765	5·892	7·042	8·634	9·926	12·340	15·119	16·985	19·812	22·362	25·472	27·688
14	4·660	5·368	6·571	7·790	9·467	10·821	13·339	16·222	18·151	21·064	23·685	26·873	29·141
15	5·229	5·985	7·261	8·547	10·307	11·721	14·339	17·322	19·311	22·307	24·996	28·259	30·578
16	5·812	6·614	7·962	9·312	11·152	12·624	15·338	18·418	20·465	23·542	26·296	29·633	32·000
17	6·408	7·255	8·672	10·085	12·002	13·531	16·338	19·511	21·615	24·769	27·587	30·995	33·409
18	7·015	7·906	9·390	10·865	12·857	14·440	17·338	20·601	22·760	25·989	28·869	32·346	34·805
19	7·633	8·567	10·117	11·651	13·716	15·352	18·338	21·689	23·900	27·204	30·144	33·687	36·191
20	8·260	9·237	10·851	12·443	14·578	16·266	19·337	22·775	25·038	28·412	31·410	35·020	37·566

χ^2 - ν *chi - quadro / gradi di libertà*

$\nu = \kappa - n$ κ *numero classi - n numero condizioni imposte*



(N.B. : probabilità complementari alla tabella precendente)

$\mathcal{K}$ numero delle classi

- **Coprire l'intero campo degli "eventi"**
- **Distribuzione teorica delle frequenze simmetrica ⇨ $\mathcal{K}$ dispari**
- **Non avere classi "vuote"**
- **$\mathcal{K}$ congruo con la numerosità N degli "eventi" ($\mathcal{K} \approx \sqrt{N}$)**

Studenti promossi / bocciati

Ipotesi : docente a = b = c

Oss	a	b	c	tot
prom	50	47	56	153
bocc	5	14	8	27
tot	55	61	64	180

Prev	a	b	c	tot
prom	47,75	51,85	54,40	153
bocc	8,25	9,15	9,60	27
tot	55	61	64	180

- **Ipotesi ⇨% prom. & % bocc. uguali per i docenti**
- **bocc. 27/180 = 15%**
- **prom. 153/180 = 85 %**
- **$\chi^2 = 4,84 \quad \nu = 2$ (cioè (n-1)(m-1))**

Dalla tabella per ν = 2:

$$\chi^2_{0,90} < \chi^2 < \chi^2_{0,95}$$
$$(4,61) < (4,84) < (5,99)$$

Accettabile con riserva

Gradi di libertà

il rispetto dei totali marginali fa sì che in ogni riga ed in ogni colonna un dato sia vincolato dal valore degli altri. Per cui in una riga di n dati, solo n-1 saranno liberi di variare. Idem, in una colonna di m dati, solo m-1 saranno liberi di variare. In totale i gradi di libertà saranno (n-1)(m-1)

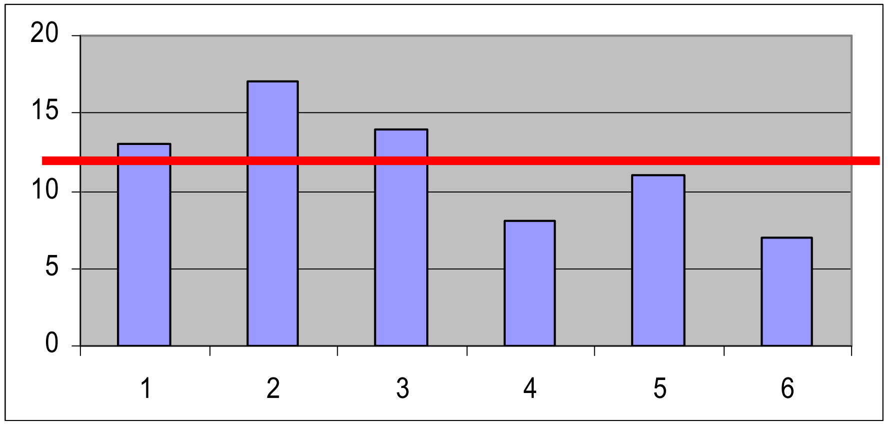
70 estrazioni dado
Ipotesi : dado "onesto"

	Oss	Prev
①	13	11,667
②	17	11,666
③	14	11,667
④	8	11,666
⑤	11	11,667
⑥	7	11,666

- **ipotesi ⇨ p = 1/6**
- **$\chi^2 = 6{,}11 \quad \nu = 5$ (cioè c-1)**

Dalla tabella per $\nu = 5$:

$$\chi^2_{0,10} < \chi^2 < \chi^2_{0,90}$$
$$(1{,}61) < (6{,}11) < (9{,}24)$$



Ipotesi non rigettabile

Confronto fra distribuzioni
Ipotesi : sperimentale = teorica

Dei dati sperimentali sono distribuiti secondo la distribuzione di Gauss?

Centro delle classi	Limiti delle classi	Frequenze assolute	Limiti in unità standard	Area tra z e -inf	Area per classe	Frequenze teoriche	$(fo-ft)^2/ft$
	41,5		-2,614	0,0045			
43		6			0,0498	4,98	0,2083
	44,5		-1,605	0,0543			
46		20			0,2215	22,15	0,2085
	47,5		-0,595	0,2758			
49		40			0,3847	38,47	0,0609
	50,5		0,414	0,6605			
52		27			0,2621	26,21	0,0235
	53,5		1,423	0,9226			
55		7			0,0699	6,99	0,0000
	56,5		2,432	0,9925			

totale	100
media	49,27
dev st	2,97

χ^2 =	0,501
ν =	2
P =	0,778

(3 condizioni: ν = c - 3)

limiti (2,5 % ; 97,5 %)	
0,051	7,378

Dalla tabella per ν = 2:

$$\chi^2_{0,025} < \chi^2 < \chi^2_{0,975}$$
$$(0,051) < (0,501) < (7,378)$$

Ipotesi non rigettabile

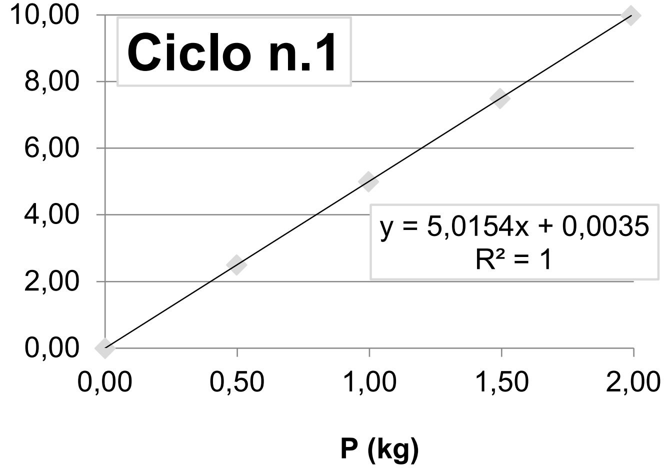
Esempio
Taratura di una cella di carico estensimetrica

- **Fondo scala: 2 kg**
- **Alimentazione: 5 V**
- **Amplificazione (guadagno): 1000**

Pesi campione	
n. 1	0,4976
n. 2	0,4992
n. 3	0,4982
n. 4	0,4948

Ciclo n.1	
P (kg)	Output (V)
0,0000	0,000
0,4976	2,503
0,9968	4,999
1,4950	7,500
1,9898	9,985
1,4950	7,500
0,9968	5,004
0,4976	2,502
0,0000	0,004
n	9
pendenza	5,015
intercetta	0,003

Output a FS (mV/V) 2,006



Errore sulla stima:

$$s = \sqrt{\frac{\sum \left(Y - Y_{stim}\right)^2}{n-2}} = 0{,}0029$$