

Ritorniamo su la propagazione degli errori

- La trattazione sulla propagazione dell'errore massimo, commesso quando si eseguono operazioni con grandezze a, b, c, \dots affette da errori $\Delta a, \Delta b, \Delta c, \dots$, si applica quando non si hanno strumenti sensibili e non si hanno fluttuazioni casuali, per cui $\Delta a, \Delta b, \Delta c, \dots$ sono le risoluzioni degli strumenti stessi.
- Quando si hanno misure più raffinate, si hanno fluttuazioni casuali e in genere si conosce non *gli errori della grandezza misurata*, (altrimenti si conoscerebbe il valore "vero" della grandezza), ma una **stima dell'errore**.
- Si è visto che una buona stima dell'errore è data dalla **deviazione standard**. COME COMBINARE LE DEVIAZIONI STANDARD?

Propagazione degli errori (dimostrazione formule anticipate nelle trasparenze lezioni n 2)

❶ Consideriamo dapprima il caso semplice: $y = f(x)$. Sia X il valor medio di x e σ_x la sua deviazione standard.

Anche per y possiamo ipotizzare l'esistenza di una funzione di distribuzione che dia la probabilità di determinare i vari valori della y e quindi definire:

$$\begin{aligned} Y &= E[y] \\ \sigma_y^2 &= E[(y - Y)^2] \\ \sigma_y &= \sqrt{E[(y - Y)^2]} \end{aligned}$$

Il problema della propagazione degli errori è ricondotto a vedere se Y e σ_y^2 possono essere legate a X e σ_x^2 e sotto quali ipotesi lo sono. Utilizzando lo sviluppo in serie di Taylor, la funzione y , per i valori di x intorno ad X , è:

$$y = f(X) + (x - X) \left(\frac{df}{dx} \right)_X + \frac{1}{2} (x - X)^2 \left(\frac{d^2 f}{dx^2} \right)_X + \dots$$

$$E[y] = E[f(X)] + E \left[(x - X) \left(\frac{df}{dx} \right)_X \right] + E \left[\frac{1}{2} (x - X)^2 \left(\frac{d^2 f}{dx^2} \right)_X \right] + \dots$$

Essendo $f(X)$, $(df/dx)_X$, $(d^2 f/dx^2)_X$ costanti ne segue che:

$$E[y] = f(X) + 0 + \frac{1}{2} \left(\frac{d^2 f}{dx^2} \right)_X E[(x - X)^2] + \dots$$

Se ne deduce che

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

quando la funzione $f(x)$ è lineare o comunque quando si possono trascurare i termini del 2° ordine o successivi. Assunto $Y = f(X)$ si ha:

$$\sigma_y^2 = E \left[[f(x) - f(X)]^2 \right]$$

e trascurando i termini al secondo ordine nello sviluppo in serie di $f(x)$ ne segue:

$$\sigma_y^2 = E \left[(x - X)^2 \left(\frac{df}{dx} \right)_X^2 \right] = \left(\frac{df}{dx} \right)_X^2 E[(x - X)^2] = \left(\frac{df}{dx} \right)_X^2 \sigma_x^2$$

$$\sigma_y = \left| \frac{df}{dx} \right| \sigma_x$$

- Nel caso di più variabili INDIPENDENTI, per es sia $f = f(x, y)$. nelle stesse approssimazioni del caso precedente, si può dimostrare che

$$\sigma_f^2 = f_x^2 \sigma_x^2 + f_y^2 \sigma_y^2$$

con

$$f_x = \left(\frac{\partial f}{\partial x} \right)_{x,y}, \quad f_y = \left(\frac{\partial f}{\partial y} \right)_{x,y}$$

In generale

$$\sigma_f^2 = \sum_{i=1}^n f_{x_i}^2 \sigma_{x_i}^2, \quad \sigma_f = \sqrt{\sum_{i=1}^n f_{x_i}^2 \sigma_{x_i}^2}$$

C'è da osservare che anche nel caso in cui si prende come errore la risoluzione dello strumento, quando si calcola una grandezza che è funzione di molte altre, conviene usare

$$\sigma_f^2 = \sum_{i=1}^n f_{x_i}^2 \sigma_{x_i}^2, \quad \sigma_f = \sqrt{\sum_{i=1}^n f_{x_i}^2 \sigma_{x_i}^2}$$

) perchè è pessimistico pensare che tutti gli errori contribuiscano nello stesso verso; è più significativo l'errore statistico.

Incertezza in una funzione di parecchie variabili

Supponiamo che x, \dots, z siano misurate con incertezze $\delta x, \dots, \delta z$, e i valori misurati utilizzati per calcolare la funzione $q(x, \dots, z)$. Se le incertezze in x, \dots, z sono indipendenti e casuali, allora l'incertezza in q è

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x} \delta x \right)^2 + \dots + \left(\frac{\partial q}{\partial z} \delta z \right)^2}$$

In ogni caso, essa non è mai più grande della somma ordinaria

$$\delta q \leq \left| \frac{\partial q}{\partial x} \right| \delta x + \dots + \left| \frac{\partial q}{\partial z} \right| \delta z$$

Cenno alla teoria dei campioni

- L'insieme delle misure sperimentali (che è una piccola parte di tutte le misure effettuabili) è un *campione* di tutte le possibili misure
- *Si postula l'esistenza* di una distribuzione dei risultati di tutte le misure effettuabili: questa viene chiamata *distribuzione generatrice (parent distribution)*
- Dal campione dobbiamo approssimare la media μ e la varianza σ^2 della distribuzione generatrice
- Vediamo allora gli estimatori del campione di misure

Stime di tendenza centrale

- Oltre ai grafici è necessario fornire dei numeri riassuntivi al fine di stimare il miglior valore e l'incertezza. Si vuole cioè definire dei *descrittori globali del campione*.

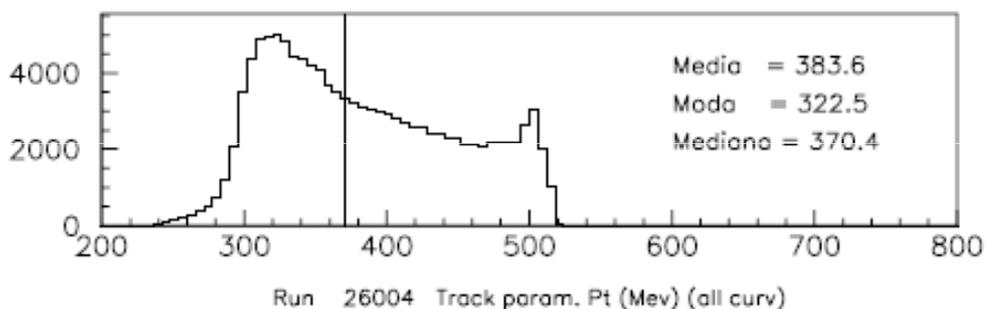
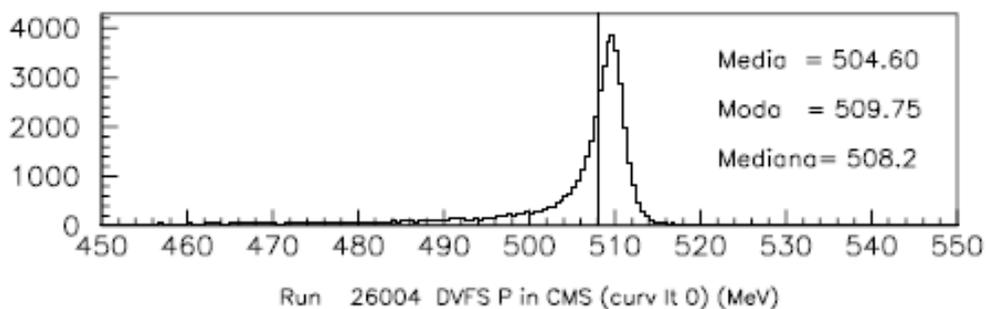
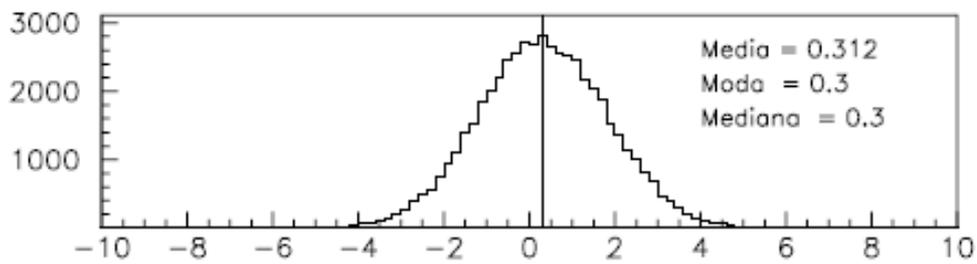
Moda: è il valore corrispondente al massimo della frequenza, cioè il valore che si è presentato il maggior numero di volte: tale stima si chiama *moda del campione*, e si può indicare con il simbolo \hat{x} .

Mediana \hat{x} è definita come quel valore che divide l'istogramma dei dati in due parti di uguale area. Nel diagramma della frequenza cumulativa relativa è definita dall'ascissa corrispondente all'ordinata del 50%.

Media aritmetica: la miglior stima della tendenza centrale di un campione è la media aritmetica \bar{x} dei valori osservati, definita attraverso la

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t$$

Tale scelta è legata alle proprietà statistiche della media, come abbiamo visto (teorema di Tschebyscheff)



Tre *esempi* di istogrammi. Per ciascuno sono indicati i valori dei 3 descrittori globali del campione di misure: media, mediana e moda. Le 3 linee sono disegnate in corrispondenza delle mediane

La media aritmetica espressa tramite le frequenze

- Siano x_i , con $i = 1, \dots, N$, gli N valori del campione di cui vogliamo calcolare la media aritmetica; supponiamo che qualcuno dei valori ottenuti sia *ripetuto*, ad esempio che il valore x_1 si sia presentato n_1 volte, x_2 si sia presentato n_2 volte e così via:
la media aritmetica si può calcolare come

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots}{N} \quad (N = n_1 + n_2 + \dots)$$

Siano x_j ($j = 1, 2, \dots, M$) gli M valori distinti di x presenti nel campione:

- n_j è la *frequenza assoluta* con cui abbiamo ottenuto il valore x_j nel corso delle nostre misure,

-ed il rapporto n_j/N è la *frequenza relativa* f_j dello stesso evento casuale

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{N} \sum_{j=1}^M n_j x_j = \sum_{j=1}^M \frac{n_j}{N} x_j = \sum_{j=1}^M f_j x_j$$

Formula di media pesata

Stime di dispersione

•Oltre alla posizione è da stimare la *dispersione* (che è legata alla incertezza sulla misura) , cioè la *larghezza* dell'istogramma o della *banda* di fluttuazione, ovvero la valutazione della larghezza dell'intervallo in *x* in cui le misure stesse sono distribuite attorno al valore centrale.

•*Semidispersione massima e range*

-La più grossolana delle stime statistiche di dispersione si effettua trovando il massimo ed il minimo valore osservato: la *semidispersione massima* è definita come la semidifferenza tra questi due valori,

$$\frac{x_{\max} - x_{\min}}{2}$$

-Difetti:

ignora la maggior parte dei dati e soprattutto quelli prossimi al centro della distribuzione;

inoltre normalmente *aumenta* all'aumentare del numero di misure, invece di tendere ad un valore determinato.

-Il doppio della semidispersione massima

$$R = x_{\max} - x_{\min}$$

viene chiamato *range*.

• *Varianza e deviazione standard (campionaria)*

-La più importante (e più usata, non solo in fisica) stima di dispersione è la *deviazione standard (oppure scarto o deviazione quadratica media)* s , che si definisce come la radice quadrata della *varianza*, s^2

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

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

-Il significato della deviazione standard campionaria s è intuitivo, indicando quanto si scarta in media dalla media

-Rispetto alla variazione massima, considera tutte le misure ed è meno soggetta ad eventuali fluttuazioni che si possono presentare al variare del campione considerato

- Altro modo di calcolare la varianza e quindi deviazione standard

$$\begin{aligned}
 N s^2 &= \sum_{t=1}^N (x_t - \bar{x})^2 \\
 &= \sum_{t=1}^N (x_t^2 + \bar{x}^2 - 2 \bar{x} x_t) \\
 &= \sum_{t=1}^N x_t^2 + N \bar{x}^2 - 2 \bar{x} \sum_{t=1}^N x_t \\
 &= \sum_{t=1}^N x_t^2 + N \bar{x}^2 - 2 N \bar{x}^2 \\
 &= \sum_{t=1}^N x_t^2 - N \bar{x}^2
 \end{aligned}$$

da cui

$$s^2 = \frac{1}{N} \sum_{t=1}^N x_t^2 - \bar{x}^2$$

-Questa formula, utilizzando la differenza tra media dei quadrati e il quadrato della media, permette un calcolo più agevole della varianza e della deviazione standard campione

Tuttavia c'è un problema con questa definizione di varianza

- La definizione andrebbe benissimo se le quantità $(x_i - m)$ fossero indipendenti tra loro, ma non lo sono. Infatti:

$$\begin{aligned}\sum(x_i - m) &= \sum x_i - \sum m = \sum x_i - nm = \\ &= \sum x_i - n \sum (x_i/n) = \sum x_i - \sum x_i = 0\end{aligned}$$

- Fisicamente questo dipende dal fatto che il valore di m è stato ricavato proprio dagli x e quindi le quantità indipendenti sono $N-1$
- Allora si prende come varianza campione la quantità

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

•*Elaborazione dati della esercitazione della SOMMA ANGOLI INTERNI DI UN TRIANGOLO :*

Calcolare la media m e la deviazione standard s del campione di somme degli angoli interni

•*Elaborazione dati della esercitazione della MISURA DI g:*

Calcolare la media m e la deviazione standard s del periodo T_i per ciascuna massa

Incertezza sulla media m del campione

- Quasi sempre interessa sapere qual è *la varianza della media m* delle x_i .
- Anche la media campione è infatti una variabile casuale con la sua funzione di distribuzione.
- La stima della varianza s_m^2 della distribuzione della media campione si può effettuare così:

$$m = \frac{1}{n} \sum x_i, \quad s_m^2 = \sum_{i=1}^n \left(\frac{\partial m}{\partial x_i} \right)^2 \sigma_{x_i}^2 = s^2 \sum_{i=1}^n \left(\frac{\partial m}{\partial x_i} \right)^2,$$

dove $\sigma_m^2 = s^2$ è la varianza sulle x_i ed è stata usata una propagazione dell'errore in quadratura dove compare la varianza invece dell'errore massimo

- Poiché $\frac{\partial m}{\partial x_i} = \frac{1}{n}$, si ricavano

$$s_m^2 = s^2 \sum_{i=1}^n \frac{1}{n^2} = \frac{s^2}{n} = \frac{1}{n(n-1)} \sum_{i=1}^n (x_i - m)^2.$$

$$s_m = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - m)^2}.$$

• **Esercizio:** Qual è il significato fisico di s_m e in cosa differisce da s ?

• *Elaborazione dati della esercitazione della SOMMA ANGOLI INTERNI DI UN TRIANGOLO:*

Calcolare l'errore s_m sulla media m

• *Elaborazione dati della esercitazione della MISURA DI g:*

Calcolare l'errore s_m sulla media m del periodo T_i per ciascuna massa

I valori sperimentali in tabella sono ottenuti per il modulo di Young E e quello di shear G dell'acciaio

$E / 10^{10} \text{ N m}^{-2}$	21.1	21.0	20.9	20.6
$G / 10^{10} \text{ N m}^{-2}$	8.12	8.15	8.13	8.08

La formula $(E/2G) - 1 = \nu$ è usata per calcolare il Poisson ratio ν - trova ν e stima l'errore massimo.

$$R \quad \nu = \frac{\langle E \rangle}{2\langle G \rangle} - 1 \quad \text{cm} \langle \rangle = -$$

$$\langle E \rangle = (\quad) / 4 = 20.9$$

$$\langle G \rangle = 8.12$$

$$\Delta E_{\text{MAX}} = \frac{21.1 - 20.6}{2} = 0.25 \quad \frac{\Delta E_{\text{MAX}}}{E} = 1.2\%$$

$$\Delta G_{\text{MAX}} = \frac{8.15 - 8.08}{2} = \frac{0.07}{2} = 0.035 \quad \frac{\Delta G_{\text{MAX}}}{G} = 0.4\%$$

Sì, he allora

$$\frac{\Delta \nu}{\nu} = \left| \frac{\Delta E}{E} \right| + \left| \frac{\Delta G}{G} \right|$$

$$\nu = 0.287 \pm 1.8\% = 0.287 \pm 0.052 \quad (\text{Rin})$$

Ripetere l'esercizio sul Poisson ratio utilizzando la relazione di una media campione, errore standard nella media e propagazione degli errori statistica

$$V = \frac{\langle E \rangle}{2\langle G \rangle} - 1 \quad \langle E \rangle = 20.9 \quad \langle E^2 \rangle = 436.845$$

$$\sqrt{\langle E^2 \rangle - \langle E \rangle^2} = \sigma_E = 0.19 \quad \sigma_m(E) = \frac{0.19}{\sqrt{3}} = 0.11$$

$$\frac{\sigma_m(E)}{\langle E \rangle} = \frac{0.11}{20.9} = 0.5\%$$

$$\langle G \rangle = 8.12 \quad \sigma_m(G)$$

$$\frac{\sigma_m(G)}{\langle G \rangle} = 0.2\%$$

$$\frac{\sigma^2(V)}{V^2} = \frac{\sigma_E^2}{E^2} + \frac{\sigma_G^2}{G^2} \Rightarrow \frac{\sigma(V)}{V} = \sqrt{25 \cdot 10^{-6} + 4 \cdot 10^{-6}} = 10^{-3} \quad 5$$

$$\sigma(V) \approx 0.5\%$$

$$V = 1.287 \approx 0.5\%$$

HA 0.1102
0.2%

- **Nota:** si è trovato dal calcolo precedente che la relazione tra varianza s^2 del campione e la varianza s_m^2 della media campione

$$s_m^2 = s^2 \sum_{i=1}^n \frac{1}{n^2} = \frac{s^2}{n}$$

In altre parole l'operazione di media ha il potere di smorzare le fluttuazioni: facendo la media di tutte le N misure, *questa fluttua N volte meno della singola misura.*

- Quando si hanno misure i cui risultati fluttuano, una volta verificato che tutto è stato fatto con la massima cura, e le variazioni non sono dovute a errori dell'osservatore, ***il risultato delle misure si scrive:***

$$m \pm s_m \quad \text{od anche} \quad m \pm 2s_m$$

- ***Che significato dare a questa scrittura?***
 - Bisogna conoscere la distribuzione della media campione
- 
- ***Teorema del limite centrale***

Premessa: consideriamo una somma di n variabili casuali indipendenti

$$x = x_1 + x_2 + \cdots + x_i + \cdots + x_n$$

dove ogni x_i ha media m_i e deviazione standard s_i . Si può mostrare che la media m e la s.d. s della somma x sono date da:

$$\begin{aligned} m &= m_1 + m_2 + \cdots + m_i + \cdots + m_n, \\ s^2 &= s_1^2 + s_2^2 + \cdots + s_i^2 + \cdots + s_n^2. \end{aligned} \tag{4.5}$$

TEOREMA DEL LIMITE CENTRALE: *Qualunque sia la distribuzione delle variabili casuali indipendenti x_i , sotto condizioni molto deboli, la somma $x = x_1 + x_2 + \cdots + x_i + \cdots + x_n$ è asintoticamente normale (m, s) dove m e s sono date dalla (4.5).*

In particolare se le variabili indipendenti x_i hanno la stessa distribuzione di probabilità ed ogni x_i ha media \bar{m} e s.d. \bar{s} finite la somma $x = \sum x_i$ è asintoticamente normale $(n\bar{m}, \sqrt{n}\bar{s})$.

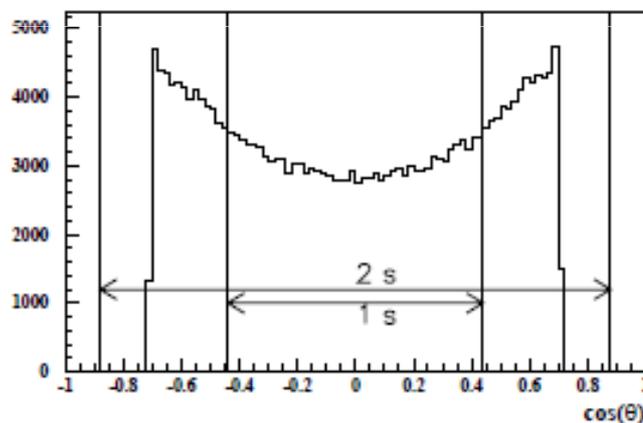
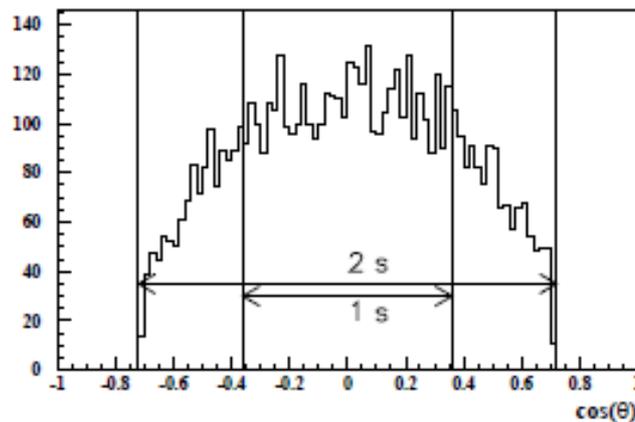
Segue che la media aritmetica $\sum x_i/n$ è asintoticamente normale $(\bar{m}, \bar{s}/\sqrt{n})$ cioè la media campione tende ad avere una distribuzione di Gauss al tendere all'infinito delle dimensioni del campione.

L'importanza del teorema sta inoltre nel fatto che la convergenza è rapida; già per la media di 6 misure la distribuzione è ragionevolmente gaussiana. Quando il campione è di 20 o più misure la distribuzione è praticamente indistinguibile da una gaussiana in quasi tutti i casi sperimentali.

Caratteristiche di s e intervalli di confidenza

- Un intervallo centrato nella media m e di semilarghezza pari alla deviazione standard s , questo intervallo non è un intervallo massimo e rappresenta solo una parte della larghezza della distribuzione del campione.

Esempio:



Due esempi di distribuzioni angolari molto diverse definite però nello stesso intervallo $(-0.7:0.7)$. In entrambi i casi l'intervallo $x \pm 2s$ è sufficiente per includere tutta la distribuzione e costituisce pertanto un *intervallo di certezza*.

L'intervallo $x \pm s$ è invece parziale e racchiude il 61% degli 18 eventi nel grafico in alto ed il 58% in quello in basso.

- *Osservazione empirica*: con intervalli di semilarghezza pari a $3s$, si hanno generalmente intervalli all'interno dei quali tutti i valori sono contenuti (*intervalli di quasi certezza*).
-

Costruzione dell'intervallo di confidenza

- Supponiamo di conoscere μ e σ della parent distribution
- Come conseguenza del T. limite centrale la distribuzione delle medie, ottenute da campioni di numerosità N , è gaussiana con media μ (quella della parent distribution) e deviazione standard $\sigma / (N)^{1/2}$
- Allora la variabile z è gaussiana in forma standard

$$z = \frac{X_{\text{med}} - \mu}{\sigma / (N)^{1/2}}$$

- L'integrale da -2 a 2 della $G(z)$ vale 0.9544 .
- Allora con un *livello di confidenza* del 95.44% posso scrivere

$$-2 \leq \frac{X_{\text{med}} - \mu}{\sigma / (N)^{1/2}} \leq +2$$

da cui con semplici calcoli si ricava *l'intervallo di confidenza* del 95.44 per la media

$$X_{\text{med}} - 2 \sigma / (N)^{1/2} \leq \mu \leq X_{\text{med}} + 2 \sigma / (N)^{1/2}$$

- Se non si conosce (come di solito) la deviazione standard, va sostituita con lo scarto quadratico medio s ottenuto dal campione

- Se il campione è di 20 o più misure, la distribuzione della media è praticamente indistinguibile da una gaussiana. Si hanno quindi le probabilità :

$$| x - m | \leq s = 68.26 \%$$

$$| x - m | \leq 2s = 95.44 \%$$

e valgono tutte le considerazioni precedenti

- In presenza di piccoli campioni le cose cambiano drasticamente. Infatti la variabile

$$t = \frac{X_{med} - \mu}{S/(N)^{1/2}}$$

non è più distribuita come una gaussiana in forma standard

- Per qualunque probabilità assegnata , bisogna allora introdurre un fattore correttivo che dipende dal numero di gradi di libertà cioè da $N-1$, numero delle grandezze indipendenti

□ Esempio: Sia $n = 5$. Qual'è l'intervallo intorno a m tale che ci sia il 95% di probabilità di trovarci la μ ? Questo intervallo è:

$$m \pm t_{95, n-1} s_m$$

Tabella del t-di-Student

P $n-1$	50%	80%	90%	95%	98%	99%	99.5%
1	1.000	3.078	6.314	12.71	31.82	63.66	127.32
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372
14	0.692	1.345	1.761	2.145	2.624	2.977	3.326
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286
16	0.690	1.337	1.746	2.120	2.583	2.921	3.252
17	0.689	1.333	1.740	2.110	2.567	2.898	3.223
18	0.688	1.330	1.734	2.101	2.552	2.878	3.197
19	0.688	1.328	1.729	2.093	2.539	2.861	3.174
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153
21	0.686	1.323	1.721	2.080	2.518	2.831	3.135
22	0.686	1.321	1.717	2.074	2.508	2.819	3.119
23	0.685	1.319	1.714	2.069	2.500	2.807	3.104
24	0.685	1.318	1.711	2.064	2.492	2.797	3.090
25	0.684	1.316	1.708	2.060	2.485	2.787	3.078
26	0.684	1.315	1.706	2.056	2.479	2.779	3.067
27	0.684	1.314	1.703	2.052	2.473	2.771	3.057
28	0.683	1.313	1.701	2.048	2.467	2.763	3.047
29	0.683	1.311	1.699	2.045	2.462	2.756	3.038
30	0.683	1.310	1.697	2.042	2.457	2.750	3.030
40	0.681	1.303	1.684	2.021	2.423	2.704	2.971
50	0.679	1.299	1.676	2.009	2.403	2.678	2.937
60	0.679	1.296	1.671	2.000	2.390	2.660	2.915
80	0.678	1.292	1.664	1.990	2.374	2.639	2.887
100	0.677	1.290	1.660	1.984	2.364	2.626	2.871
120	0.677	1.289	1.658	1.980	2.358	2.617	2.860
∞	0.674	1.282	1.645	1.960	2.326	2.576	2.807