Statistica Finita

La media μ e la deviazione standard σ sono parametri pienamente descrittivi delle caratteristiche statistiche di una popolazione a patto di elaborare TUTTI i dati disponibili

Non sempre è possibile, o conveniente, accedere a tutta la popolazione

Non sempre è ragionevole acquisire la funzione di densità di probabilità: è possibile adottare opportune ipotesi

Se ciò è possibile sarà sufficiente identificare/stimare i parametri descrittivi della funzione adottata

Per motivi pratici avremo un campione costituito da numero di eventi, n, al più uguale, ma normalmente inferiore, alla dimensione della popolazione N

Nel caso di una misura la popolazione è, teoricamente, infinita e il limite al numero di misure è solo pratico e/o economico

Avremo a disposizione solo un'approssimazione costituita dalla media, o media campionaria, e dalla deviazione standard del campione

$$\mu \approx \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma \approx S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Cosa succede se non possiamo operare su tutta la popolazione e ci dobbiamo accontentare di un campione?

In alcuni casi è possibile far riferimento alla sola aleatorietà della variabile casuale: è quanto avviene nel rilevamento degli eventi di un campione di una variabile casuale discreta (es. l'esito del lancio di un dado): il singolo dato di lettura del dado è esatto e la statistica fornirà indicazioni relative alla sola variabile casuale

Nel caso di misura di una variabile casuale continua, la sua misurazione comporta un'incertezza (variabilità intrinseca, effetti del processo di misura, effetti dello strumento): il risultato dell'indagine statistica fornisce un risultato complessivo. Non avremo problemi se la dispersione sui singoli valori dovuti al processo di misurazione è piccola rispetto alla dispersione intrinseca del misurando o viceversa.

Nel caso di misure il problema è complicato dalla presenza di effetti casuali su più fronti (misurando, processo di misura e strumento) che concorrono a influenzare la media e, soprattutto, la deviazione standard.

Media e deviazione standard di un campione caratterizzano i dati acquisiti, risentono di tutti gli effetti ricordati, e possono essere impiegati, mediante un opportuno fattore di confidenza/significatività, per prevedere la probabilità di rilevare ulteriori misure in un certo intervallo (es 2σ, P>95%).

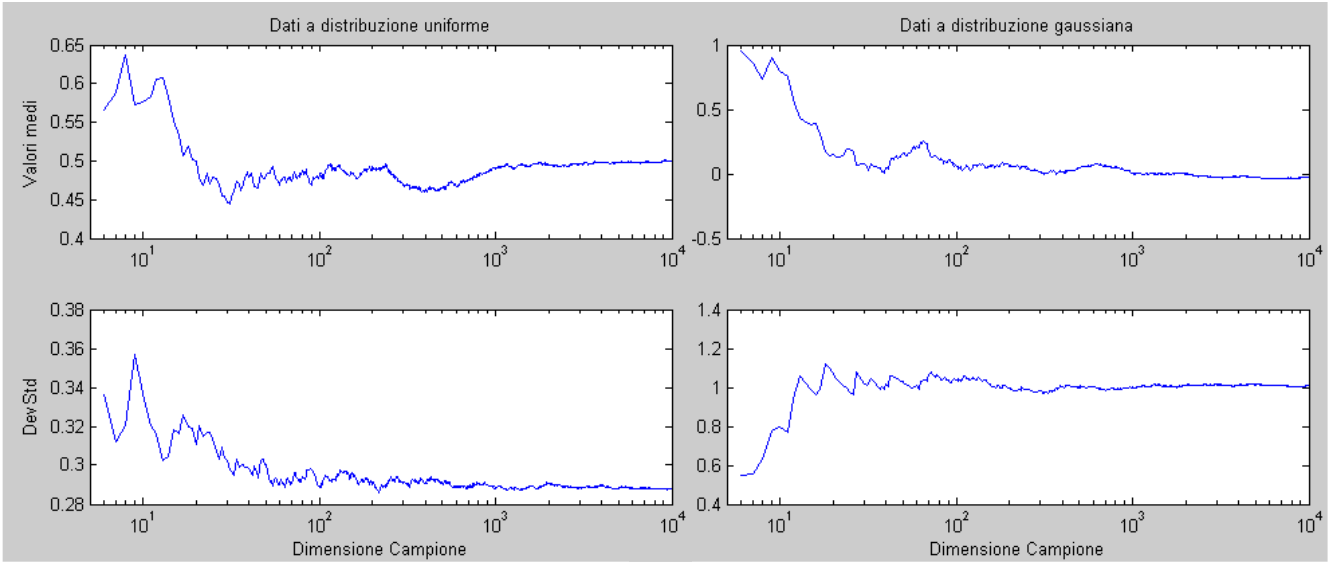
Ma non è questo l'obiettivo di un'operazione di misura: il problema è stimare i parametri relativi al misurando depurandoli degli effetti del processo di misura e fornirne il valore più probabile e un indicatore di dispersione.

Problema: Come possiamo utilizzare la media campionaria e la deviazione standard di un campione per fornire la migliore stima possibile della variabile e della sua distribuzione?

Per quanto riguarda il valore atteso è ragionevole ritenere che la media campionaria sia una sua ragionevole approssimazione: l'operatore è lineare

Al crescere degli eventi la media e la deviazione tendono a stabilizzarsi identificando i parametri del processo casuale

A Sx distribuzione uniforme 0-1 (μ=0.5 , σ=0.5/sqrt(3)); a Dx distribuzione gaussiana (μ=0,σ=1)



Si nota che a partire da campioni di 20/30 individui i valori sono nell'intorno di quelli di convergenza.

Lecito chiedersi qual sia l'effetto della scelta di un particolare campione quindi come variano media e deviazione standard se, a parità di numero di eventi, si considerano campioni diversi

L'obiettivo, essendo l'interesse posto sul valore medio, è quello di arrivare a definire un *intervallo di confidenza* dentro il quale sicuramente cade la migliore stima possibile della variabile, in modo da rendere la stima indipendente dal numero delle misure piuttosto che da uno specifico campione

$$\mu = \bar{x} \pm \delta \qquad \Rightarrow \qquad \bar{x} - \delta \le \mu \le \bar{x} + \delta$$

$$\delta = \text{incertezza} \qquad \bar{x} = \text{stima di } \mu \text{ (media)}$$

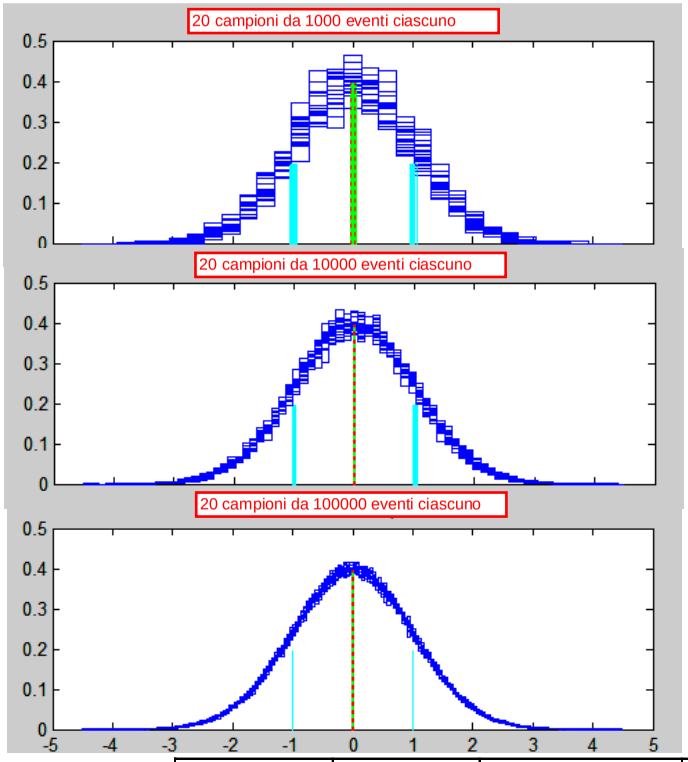
Analizziamo i risultati della statistica per una serie di misure (campioni) di una variabile casuale ($\mu=0,\sigma=1$), campioni tutti di eguale numero di individui.

Campioni diversi producono stime differenti anche con un numero di eventi uguale:

Campioni diversi di una popolazione, per es. di dimensione *n*, forniranno ciascuno una propria media ed una propria deviazione std e non possiamo aspettarci che coincidano: la variabile è *casuale*.

Possiamo solo aspettarci che all'aumentare di numero di eventi di ciascun campione la differenza sia sempre più contenuta

SPECIFICITA' DEL CAMPIONE: la caratterizzazione di cui disponiamo sono il valor medio e la deviazione standard del campione e non della popolazione; questa stima dipende dal numero di eventi disponibili e dal fatto che i campioni sono quelli disponibili e non altri, peraltro da ritenersi equivalenti.



Esaminiamo alcuni istogrammi corrispondenti a 20 campioni di numero crescente di eventi

Si nota che, anche quando gli istogrammi sono significativamente diversi, i valori medi e le deviazioni standard sono poco variabili

Esibiscono comunque una distribuzione: sono variabili casuali

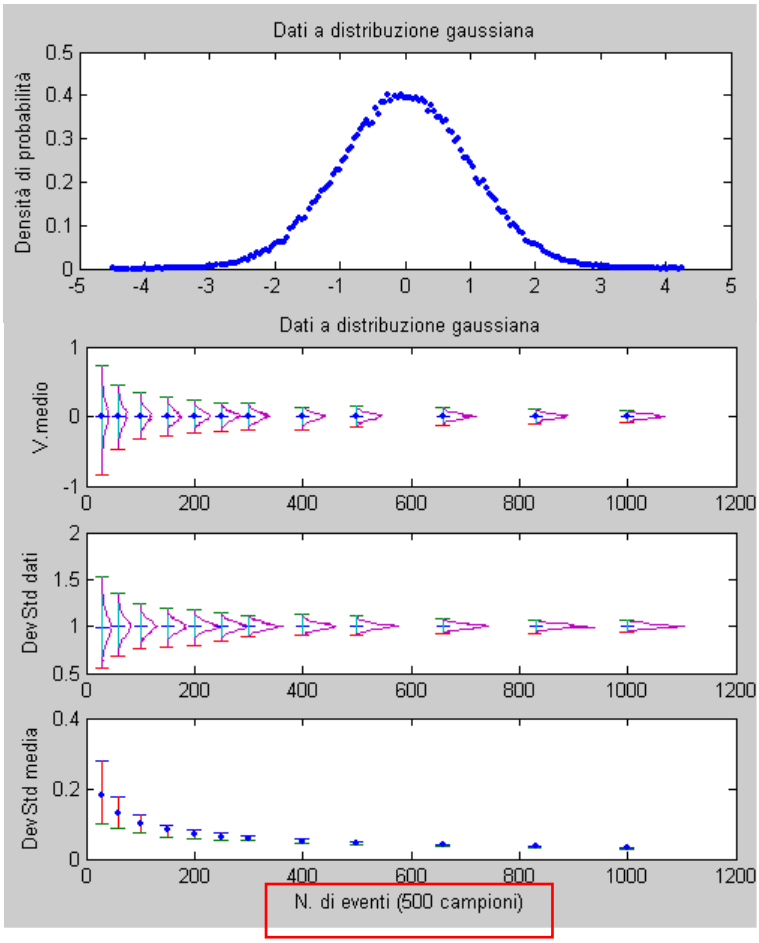
Campioni	Eventi N	$\overline{x}_{Max} - \overline{x}_{Min}$	S_{Media}	$\overline{x}_{Medio}$
20	100	0.077210	0.997277	-0.024008
20	1000	0.023909	0.998705	0.014712
20	10000	0.007946	1.002611	0.005152
20	100000	0.004044	1.000984	-0.000153
	∞	-	σ=1	μ=0

Come dipende la statistica dalla scelta del campione? 500'000 eventi casuali (μ=0,σ=1) elaborati a pacchetti

Distribuzione delle medie dei pacchetti: si riconosce una distribuzione centrata sul valore nullo (Media di tutte le misure disponibili)

Distribuzione delle devstd dei pacchetti: attestate sul valore unitario

DevStd delle medie: minore di quella dei dati e diminuisce all'aumentare del campione



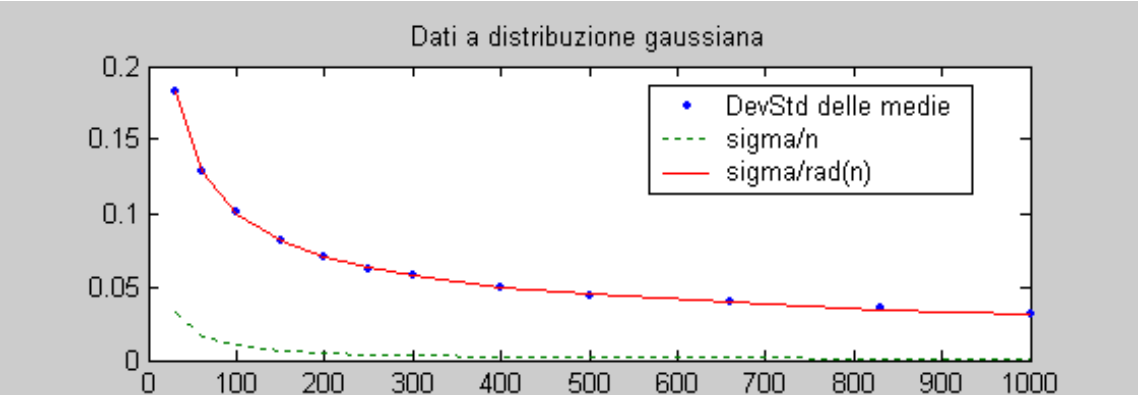
Dimostrazione euristica

All'aumentare del numero di misure la deviazione standard delle medie diminuisce.

L'andamento suggerisce una diminuzione della deviazione all'aumentare della dimensione del campione

Un coefficiente proporzionale all'inverso del numero di misure porta ad una eccessiva diminuzione (linea verde)

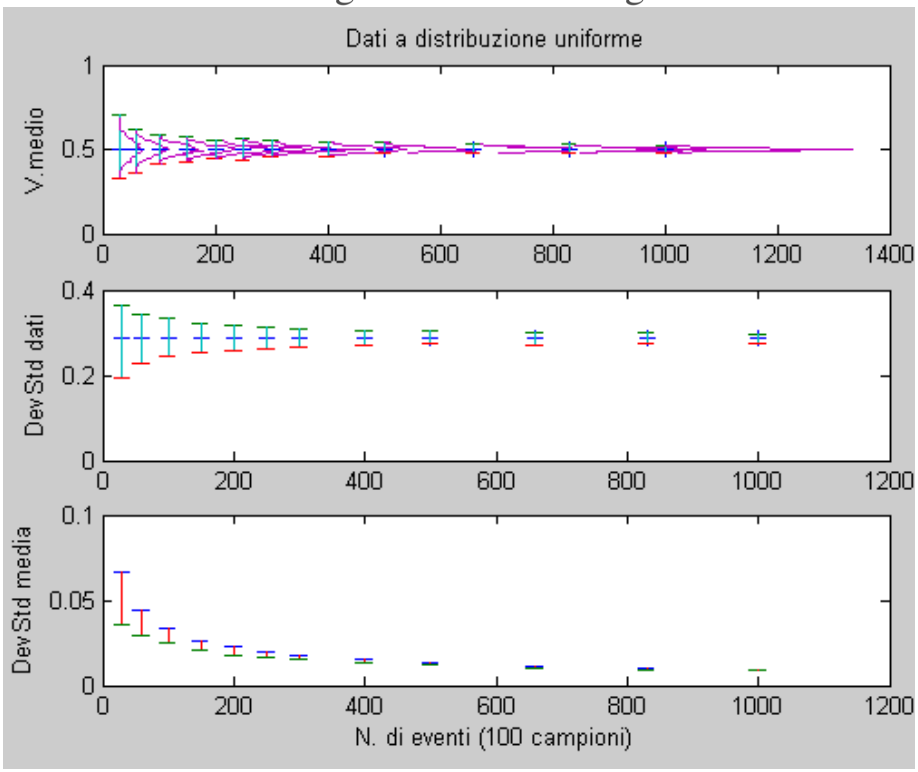
La potenza 0.5 porta ad una corretta interpretazione dell'andamento della deviazione delle medie



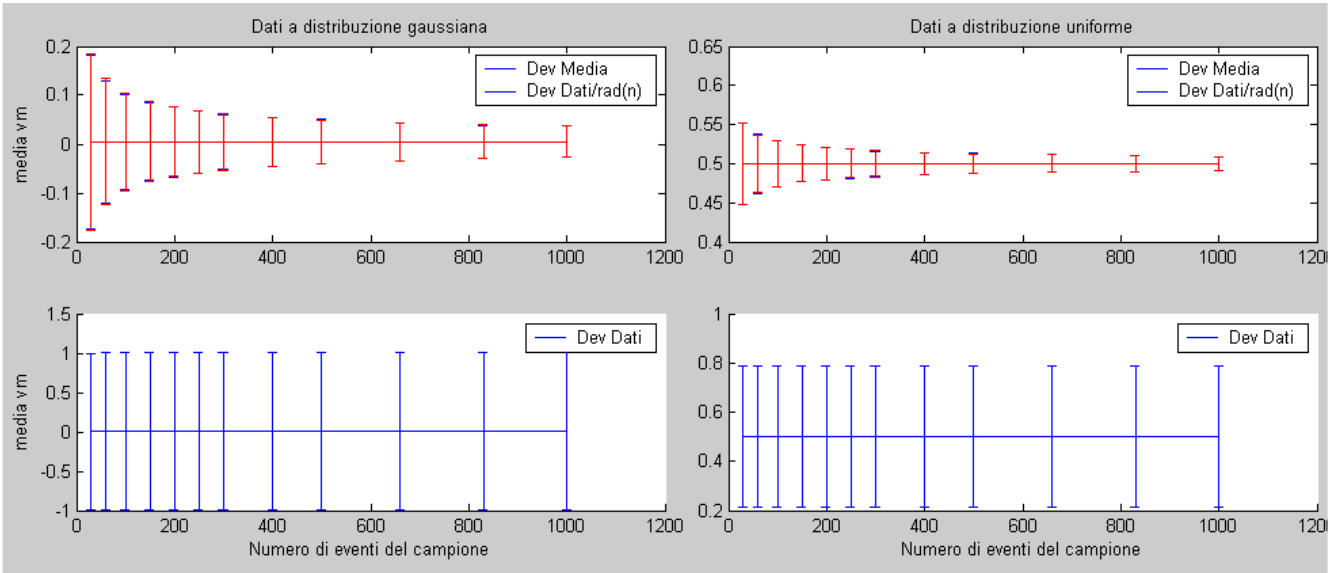
Per *n* sufficientemente grande le stime dei valori medi hanno distribuzione gaussiana con deviazione standard data da (*std deviation of means*):

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{S}{\sqrt{n}}$$

NON è necessario che la popolazione sia a distribuzione normale: per una distribuzione uniforme i grafici sono analoghi



Confronto risultati per distribuzione gaussiana e uniforme



Dimostrazione (cfr. teoremi del limite centrale)

Esperimento di N valutazioni $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$

Ripetizione dell'esperimento $\bar{x}_k = \frac{1}{N}\sum_{i=1}^{N} x_{ki} \quad S_k = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_{ki} - \bar{x}_k)^2}$

k=1,M

(valori medi e dev.std si stabilizzano con M)

Valutazione globale N x M misure

$$m_{\bar{x}} = \frac{1}{M}\frac{1}{N}\sum_{k=1}^{M}\sum_{i=1}^{N} x_{ki} = \frac{1}{M}\sum_{k=1}^{M}\bar{x}_k$$

$$S^2_{MN} = \frac{1}{M}\frac{1}{N}\sum_{k=1}^{M}\sum_{i=1}^{N}(x_{ki} - m_{\bar{x}})^2$$

I valori medi costituiscono un insieme di variabili casuali indipendente che possiamo ipotizzare gaussiano, quindi caratterizzato da un valore medio e da una deviazione standard $m_{\bar{x}}, S_{\bar{x}}$

Occorre identificare questi due parametri

Caratterizzazione del processo di media dei valori medi

L'operatore è lineare: la media dei VM coincide con la media totale

Se N è sufficientemente grande le deviazioni casuali sono già compensate in una serie di dati, quindi:

$$m_{\bar{x}} = \frac{1}{M}\sum_{k=1}^{M}\bar{x}_k \qquad m_{\bar{x}} \approx \bar{x}_k \quad \forall k$$

In particolare: $\boxed{m_{\bar{x}} \approx \bar{x}_1}$ Non serve ripetere l'esperimento!!

Vediamo se possibile ottenere un risultato analogo per la deviazione standard dei valori medi $S_{\bar{x}}$

Dev.Std di tutti i dati $$S_{MN}^2 = \frac{1}{M}\frac{1}{N}\sum_{k=1}^{M}\sum_{i=1}^{N}(x_{ki} - m_{\bar{x}})^2$$

Dev.Std dei valori medi $$S_{\bar{x}}^2 = \frac{1}{M}\sum_{k=1}^{M}(\bar{x}_k - m_{\bar{x}})^2$$

Il termine quadratico può essere sviluppato in termini di *media delle deviazioni delle singole misure dalla media globale*:

$$\bar{x}_k - m_{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}x_{ki} - \frac{N}{N}m_{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}x_{ki} - \frac{1}{N}\sum_{i=1}^{N}m_{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}\left(x_{ki} - m_{\bar{x}}\right)$$

Utilizzando questa relazione della dev.std delle medie:

$$\bar{x}_k - m_{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}\left(x_{ki} - m_{\bar{x}}\right) \qquad S_{\bar{x}}^2 = \frac{1}{M}\sum_{k=1}^{M}(\bar{x}_k - m_{\bar{x}})^2$$

otteniamo

$$S_{\bar{x}}^2 = \frac{1}{M}\sum_{k=1}^{M}(\bar{x}_k - m_{\bar{x}})^2 = \frac{1}{M}\sum_{k=1}^{M}\left(\frac{1}{N}\sum_{i=1}^{N}\left(x_{ki} - m_{\bar{x}}\right)\right)^2 =$$

$$= \frac{1}{M}\frac{1}{N^2}\sum_{k=1}^{M}\left(\sum_{i=1}^{N}d_{ki}\right)^2 = \frac{1}{M}\frac{1}{N^2}\sum_{k=1}^{M}\left(\sum_{i=1}^{N}d_{ki}^2 + \sum_{i=1}^{N}\sum_{j=i}^{N}2d_{kj}d_{ki}\right)$$

Il secondo termine, per N elevato, tende a zero trattandosi di scostamenti casuali

$$S_{\bar{x}}^2 \approx \frac{1}{M}\frac{1}{N^2}\sum_{k=1}^{M}\sum_{i=1}^{N}d_{ki}^2 = \frac{1}{M}\frac{1}{N^2}\sum_{k=1}^{M}\sum_{i=1}^{N}\left(x_{ki} - m_{\bar{x}}\right)^2$$

Caratterizzazione del processo di misura: deviazione standard dei valori medi

$$S_{\bar{x}}^2 \approx \frac{1}{M}\frac{1}{N^2}\sum_{k=1}^{M}\sum_{i=1}^{N}\left(x_{ki} - m_{\bar{x}}\right)^2$$

Ma poichè $$S_{MN}^2 = \frac{1}{M}\frac{1}{N}\sum_{k=1}^{M}\sum_{i=1}^{N}(x_{ki} - m_{\bar{x}})^2$$

Otteniamo $$S_{\bar{x}}^2 \approx \frac{1}{M}\frac{1}{N^2}\sum_{k=1}^{M}\sum_{i=1}^{N}(x_{ki} - \bar{x}_k)^2 = \frac{1}{N^2}S_{MN}^2$$

Assumendo $S_{MN} \approx \frac{1}{M}\sum_{k=1}^{M} S_k \approx S_k \ \forall k$ abbiamo infine $\boxed{S_{\bar{x}} \approx \frac{1}{\sqrt{N}}S}$

Per essere considerato "grande" *n* dovrebbe essere superiore a 30.

Per *n* molto grande la deviazione standard della media tende a zero

Riassumendo:

1. Se la popolazione è a distribuzione normale, la distribuzione delle medie dei campioni è normale.
2. Se la popolazione non è a distribuzione normale ma *n* è grande ($n > 30$), la distribuzione delle medie è normale.
3. Se la popolazione non è normale e $n < 30$, le medie seguono la distribuzione normale solo approssimativamente.

Per la media, potendosi adottare l'ipotesi di distribuzione normale, l'incertezza potrà quindi essere espressa come:

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

si potranno inoltre utilizzare le tabelle di integrazione della funzione probabilità con le medesime procedure operative per stimare l'intervallo di confidenza definendo un opportuno valore per z

Esempio: Si voglia calcolare l'intervallo di confidenza della media di un certo numero di resistenze. Se ne misurano 36; la resistenza media misurata, $\bar{x}$, è pari a 25Ω e la deviazione standard (stimata), S, è pari a 0.5Ω.

Determinare l'intervallo di confidenza della media per una probabilità pari al 90%.

Soluzione:

Dati: Probabilità = 1-α = 0.9 e quindi α = 0.1

Occorre trovare il valore di $z_{\alpha/2}$: cioè l'intervallo che racchiude un'area del 90% (esclude un'area del 10%).

Essendo la funzione densità simmetrica è sufficiente ricercare nella tabella 0.45.

Il valore di z corrispondente è circa z=1.645, circa a metà tra i valori della tabelle che limitano 0.45.

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
…										
1.4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1.5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1.6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1.7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1.8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
…										

Assumendo una distribuzione di tipo gaussiano e utilizzando S come migliore stima della deviazione standard, tenendo conto dell'intervallo di confidenza e riferendo la deviazione alla media, otteniamo con n=36 :

$$\bar{x} - z_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$25 - 1.645\frac{0.5}{6} \le \mu \le 25 + 1.645\frac{0.5}{6}$$

$$24.86 \le \mu \le 25.14$$

In definitiva la **resistenza media** avrà un valore di:

$$25 \pm 0.14\Omega$$

Con un livello di confidenza pari al 90%, cioè il 90% delle resistenze medie avranno un valore compreso nell'intervallo di 0.14Ω

Student's t distribution

La distribuzione normale rappresenta uno schema corretto quando il numero di misure è elevato; la stima della deviazione costituisce una base valida per la definizione dell'intervallo di confidenza solo in questo caso.

Ma cosa accade per un numero di misure limitato? (si può riconoscere che la trattazione che segue fornisce risultati praticamente sovrapponibili a quelli ottenuti con l'utilizzo della funzione di densità normale, quando si utilizzino almeno 30 campioni)

$$n > 30$$

La deviazione standard della popolazione, σ, non sarà bene approssimata dalla deviazione del campione, S, e, a causa della incertezza nella deviazione standard, ci possiamo aspettare sia necessario avere un intervallo più ampio, per garantire lo stesso livello di confidenza.

Nel caso di poche misure, viene utilizzato un altro indicatore statistico, detto **Student's t** definito come:

$$t = \frac{(\bar{x} - \mu)}{\sigma_{\bar{x}}} = \frac{(\bar{x} - \mu)}{S/\sqrt{n}}$$

t è definito in maniera analoga a *z* : come rapporto tra la deviazione tra la media del campione e quella vera e la Dev. Std. della media stimata

Esiste una famiglia di distribuzioni di *t* le quali, al contrario della distribuzione normale, dipendono dal numero di misure (ν=*N-1*). L'espressione analitica di queste distribuzioni è data da:

$$f(t,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{t^2}{\nu}\right)^{(\nu+1)/2}}$$

La distribuzione è funzione di una funzione matematica (*Gamma function*) e di un parametro ν, che definisce il numero di gradi di libertà (numero di misure meno il numero minimo di misure necessarie a definire un indicatore statistico)

$$f(t,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{t^2}{\nu}\right)^{(\nu+1)/2}}$$

Il numero di gradi di libertà v è dato dal numero di misure ridotto del numero minimo necessario per realizzare la statistica; per il valore medio quindi 1

Per esempio, per definire il diametro di un tubo, il numero minimo di misure necessarie per definire una stima statistica è 1.

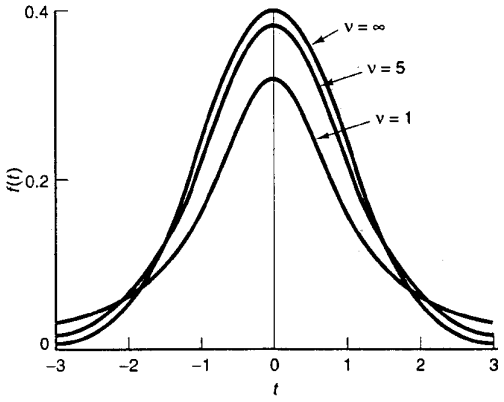
Se eseguiamo 10 misure, il numero di gradi di libertà è 9, cioè ν=10-1.

Espressione della *Gamma function:* $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$

Se n è intero positivo allore: $\Gamma(n) = (n-1)!$

La funzione densità ***t***-Student è simmetrica, inoltre si abbassa e si allarga al diminuire del numero di misure

Graficamente, le distribuzioni ***t*** sono simili alla distribuzione normale, e diventano equivalenti ad essa al crescere del numero delle misure.



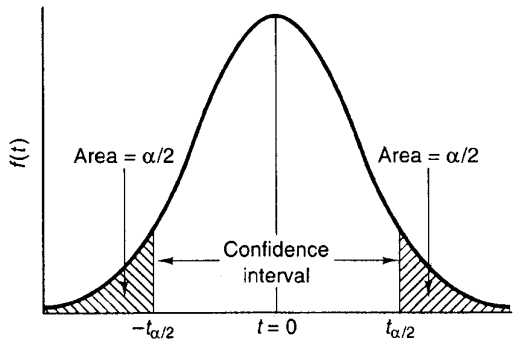
Le distribuzioni ***t***, possono essere utilizzate, analogamente a quelle normali, per stimare l'intervallo di confidenza della media di un certo numero di misure, quando queste sono inferiori a 30.

Il modo di procedere è del tutto analogo a quello utilizzato con la distribuzione normale: una volta scelta la curva corrispondente ai gradi di libertà in questione (v), possiamo definire la probabilità che ***t*** cada nell'intervallo:

$$-t_{\alpha/2} \quad \text{e} \quad +t_{\alpha/2}$$

Ovvero:

$$P\left[-t_{\alpha/2} \le t \le +t_{\alpha/2}\right] = 1-\alpha$$



Sostituendo l'espressione di ***t*** otteniamo:

$$P\left[-t_{\alpha/2} \le \frac{\bar{x}-\mu}{S/\sqrt{n}} \le +t_{\alpha/2}\right] = P\left[\bar{x} - t_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right] = 1-\alpha$$

Che si può esprimere come: $\mu = \bar{x} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$ con probabilità 1-α

Dato che le tabelle complete che riportano le distribuzioni ***t*** sono voluminose, vengono solitamente specificati alcuni valori di ***t*** funzioni di ***ν*** ed ***α*** necessari al calcolo degli indici sopra riportati.

	α/2		
ν	0.100	0.050	0.025
1	3.078	6.314	12.706
2	1.886	2.920	4.303
3	1.638	2.353	3.182
4	1.533	2.132	2.776
5	1.476	2.015	2.571
6	1.440	1.943	2.447
7	1.415	1.895	2.365
8	1.397	1.860	2.306
9	1.383	1.833	2.262

...

Dal punto di vista operativo si tratta di introdurre un termine di amplificazione delle incertezze, che permette di compensare una insufficiente disponibilità di dati

Infatti con pochi dati si rischia di avere una non corretta stima delle code, vuoi perché siamo stati fortunati e i dati sono tutti nella zona centrale, vuoi perché siamo stati sfortunati e abbiamo più dati lontano dalla zona centrale

Con il parametro così calcolato e il numero di misure a disposizione (ridotto di 1), da un'apposita tabella si ricava il coefficiente $t_{\alpha/2}$; che garantisce la confidenza desiderata (1-α)

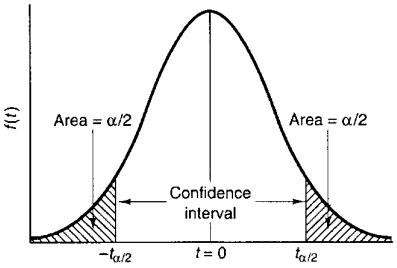
in base a questo coefficiente si esprime l'incertezza:

$$\mu = \bar{x} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$$

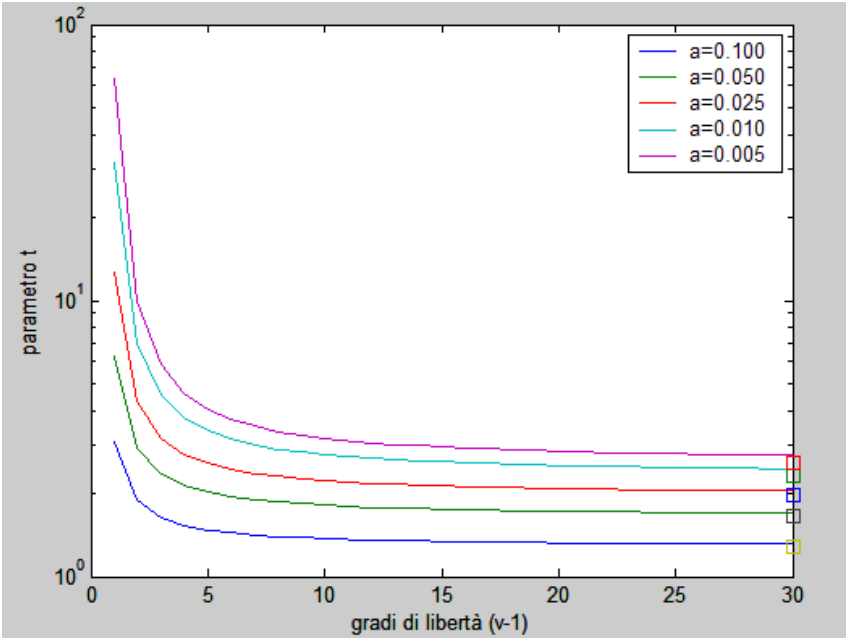
I dati necessari sono: il coefficiente di copertura α (ricavato come 1-livello di confidenza desiderato) e il campioni indipendenti ν (numero di misure-1)

Stralcio della tabella per la determinazione del $t_{\alpha/2}$:

	α/2				
ν	0.100	0.050	0.025	0.010	0.005
1	3.078	6.314	12.706	31.823	63.658
2	1.886	2.920	4.303	6.964	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169



Andamento del coefficiente t-Student al variare del numero di gradi di libertà e del coefficiente di copertura α (P=1- α)



Esempio: Si vuole valutare il tempo medio di guasto di schermi VCR con un intervallo di confidenza del 95%, partendo da 6 misure del tempo di guasto, pari a ore:

1250, 1320, 1542, 1464, 1275 e 1383

Si chiede di stimare la media e l'intervallo di confidenza della media per un livello di confidenza del 95%.

Soluzione: Il valor medio e la deviazione del tempo di guasto valgono:

$$\overline{x} = \frac{1250+1320+1542+1464+1275+1383}{6} = 1372h$$

$$S = \sqrt{\frac{\sum_{1}^{n} {}_i d_i^2}{n-1}} = 114\ h$$

I parametri sono intervallo di confidenza e numero di campioni:

$$95\% \Rightarrow \alpha = 0.05 \quad ; \quad v = n-1 = 5$$

Dalla tabelle si ottiene:

$$t_{v,\alpha/2} = t_{5,95\%} = 2.571$$

$$\mu = \overline{x} \pm t_{\alpha/2}\frac{S}{\sqrt{n}} = 1372 \pm 2.571\frac{114}{\sqrt{6}} = 1372 \pm 120h$$

Se non si fosse tenuto conto della correzione t-Student il coefficiente, a parità di intervallo di confidenza, sarebbe stato 1.96 anziché 2.571

Esempio: Se nell'esempio precedente si volesse limitare l'intervallo di confidenza sulla media a ±80 ore, sempre con un livello di confidenza pari al 95%, quante altre misure sono necessarie?

Soluzione: L'intervallo di confidenza è dato da:

$$IC = \pm t_{\nu,\alpha/2} \frac{S}{\sqrt{n}}$$

Avendo eseguito 6 misure, dalla tabella di t-Student per ν=6-1=5 e il 95% otteniamo $t_{5,95\%} = 2.571$; risolvendo rispetto al numero di misure avremo:

$$\pm 80 = \pm 2.571 \frac{114}{\sqrt{n}} \qquad n = \left(\frac{2.571 \times 114}{80} \right)^2 = 13.42 \simeq 14$$

Occorre quindi acquisire altre 8 misure e verificare che la nuova statistica rispetti il requisito IC ≤ 80 con probabilità del 95%

χ^2 distribution

Anche la qualità della stima della varianza può risultare di interesse pratico

In questo caso vengono definiti due coefficienti moltiplicativi della deviazione standard calcolata sul campione, *S*, che permettono di stabilire un valore massimo e minimo all'interno del quale dovrebbe trovarsi la deviazione standard vera, σ, con un livello di confidenza, anche in questo caso, da esprimersi in termini tipo probabilistici, es. 95%

$$\sqrt{\nu} \frac{S}{\chi_{\nu,1-\alpha/2}} \leq \sigma \leq \sqrt{\nu} \frac{S}{\chi_{\nu,\alpha/2}}$$

L'argomento non viene discusso se non per dire che si utilizza una distribuzione di densità di probabilità particolare, denominata χ^2. Si rimanda chi fosse interessato alla bibliografia

Media come miglior stima?

Abbiamo assunto che la media di una serie di misure $\bar{x}$ rappresenti la miglior stima possibile del misurando e che la deviazione standard delle misure σ_x (o delle medie) rappresenti una adeguata stima probabilistica dell'intervallo di confidenza

Possiamo sostanziare queste affermazioni? Hanno dei limiti?

Ammettiamo che le misure abbiano distribuzione Gaussiana, allora la probabilità di ottenere la misura x_1 è:

$$P(x_1 - dx \le x \le x_1 + dx) = p(x_1)\,dx =$$

$$\frac{1}{\sigma_x\sqrt{2\pi}}\exp\left(-\frac{(x_1-\bar{x})^2}{2\sigma_x^2}\right)dx \propto \frac{1}{\sigma_x}\exp\left(-\frac{(x_1-\bar{x})^2}{2\sigma_x^2}\right)$$

La probabilità di osservare tutte le *N* misure è data dal prodotto delle rispettive probabilità

$$P_{\bar{x},\sigma} = \prod_{i=1}^{N} p(x_i)\,dx \propto \frac{1}{\sigma_x^N}\,e^{-\sum_{i=1}^{N}\frac{(x_i-\bar{x})^2}{2\sigma_x^2}}$$

Esiste un Principio, detto di Massima Verosimiglianza, che, applicato al caso in esame, porta a dire che la migliore stima del valore migliore e della deviazione standard si hanno quando è massima la probabilità *P* appena definita

Deve quindi essere minimo l'esponente $-\sum_{i=1}^{N}\frac{(x_i-\bar{x})^2}{2\sigma_x^2}$

Differenziando rispetto a $\bar{x}$, inteso come migliore stima della misura

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

ed eguagliando a zero, si ottiene:

$$\sum_{i=1}^{N}(x_i-\bar{x}) = \sum_{i=1}^{N}x_i - \sum_{i=1}^{N}\bar{x} = \sum_{i=1}^{N}x_i - N\bar{x} = 0 \qquad \boxed{\bar{x} = \frac{1}{N}\sum_{i=1}^{N}x_i}$$

Dev. std come intervallo di confidenza?

Differenziando quindi la funzione di probabilità rispetto a σ_x, inteso come migliore dell'intervallo di confidenza, ed eguagliando a zero

$$\frac{\partial}{\partial \sigma_x}\left(\frac{1}{\sigma_x^N} e^{-\sum_{i=1}^{N} \frac{(x_i-\bar{x})^2}{2\sigma_x^2}} \right) = \ldots$$

si ottiene infine:

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

Osserviamo quindi che il valor medio e la deviazione standard sono la migliore stima della misura e dell'intervallo di confidenza **solo** se gli errori sono a distribuzione Gaussiana

Se la distribuzione non è Gaussiana NON E' DETTO che la media e deviazione standard rappresentino la migliore stima dei parametri statistici.

La media lo è se la distribuzione di probabilità è simmetrica.

Fine presentazione