

Problema n.2

- Trovare la concentrazione di NAD e NADH in una soluzione miscelata basandosi sui seguenti dati ottenuti in una cuvetta da 1 cm,
- I coefficienti di estinzione molare a 260 nm sono 1.5×10^{-4} per il NADH e 1.84×10^{-4} per il NAD. A 340 nm solo il NADH assorbe con ϵ di 6.22×10^{-3} .
- La soluzione miscelata da assorbanze di 1.36 e 0,382 a 260 e 340 nm, rispettivamente.

$$A_{260\text{nm}} = 1.36$$

$$A_{340\text{ nm}} = 0.382$$

	$\epsilon_{260\text{ nm}}$	$\epsilon_{340\text{ nm}}$
NADH	1.5×10^4	6.22×10^3
NAD	1.84×10^4	----

$$A_{340\text{nm}} \quad A = c\epsilon\lambda \rightarrow [\text{NADH}] = 0.382 : 6.22 \times 10^3 = 61.41 \times 10^{-6} \text{ M}$$

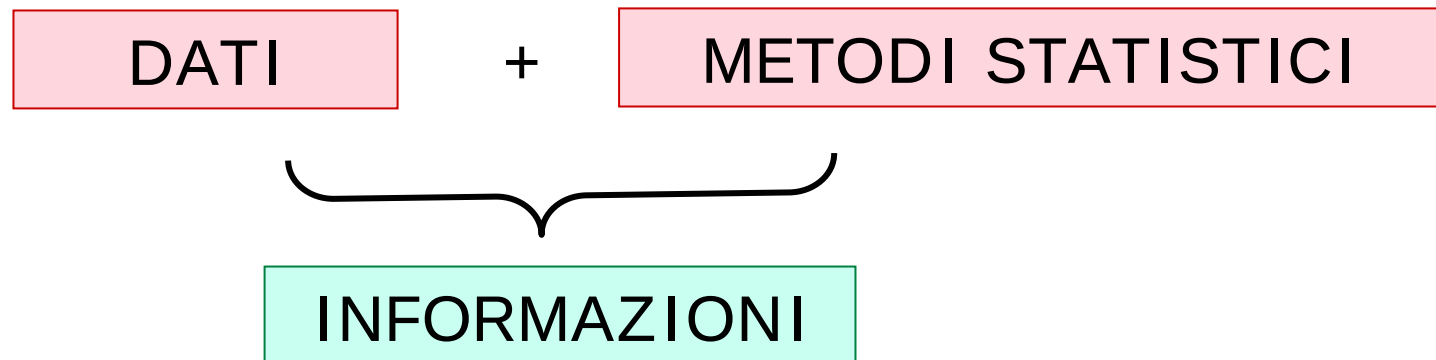
$$A_{260\text{ nm}} \quad A = A_{\text{NAD}} + A_{\text{NADH}}$$

$$\rightarrow 1.36 = [\text{NAD}] \times 1.84 \times 10^4 + [\text{NADH}] \times 1.5 \times 10^4$$

$$\begin{aligned} \rightarrow [\text{NAD}] &= [1.36 - (61.41 \times 10^{-6} \text{ M} \times 1.5 \times 10^4)] : 1.84 \times 10^4 = \\ &= [1.36 - 0.92] : 1.84 \times 10^4 = 2.4 \times 10^{-5} \text{ M} \end{aligned}$$

L'acquisizione di una osservazione come DATO SCIENTIFICO richiede che :

- osserviamo un fenomeno
- lo quantifichiamo correttamente
- lo confrontiamo con altri fenomeni
- lo interpretiamo correttamente con criteri precisi



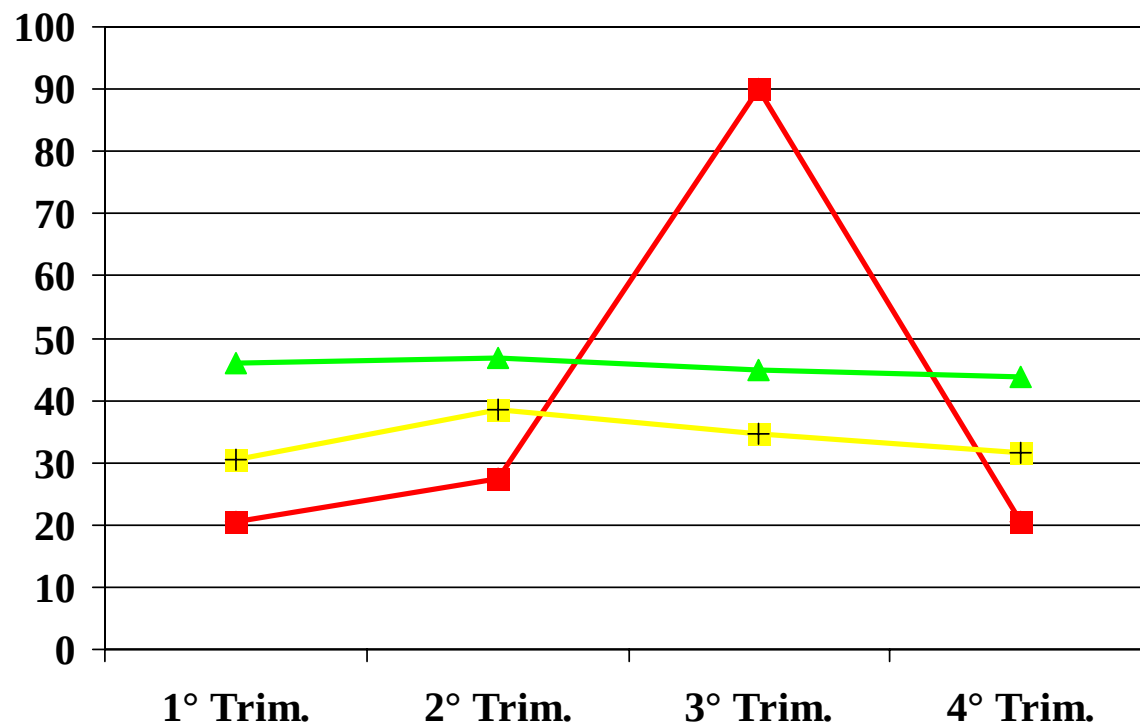
Arrotondamento dei dati

27.435 $\longrightarrow$ 27.43 $\longrightarrow$ 27.4

21.467 $\longrightarrow$ 21.47 $\longrightarrow$ 21.5

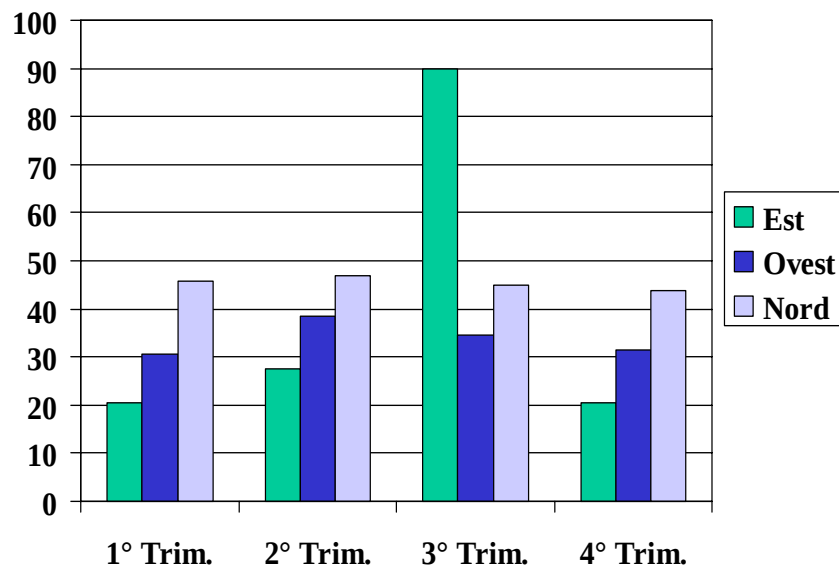
Il numero di cifre significative **non può** essere superiore alla precisione dello strumento, che dipende dalle **caratteristiche tecniche** dello strumento

Rappresentazioni grafiche di frequenza di un evento: il grafico dei dati in funzione del tempo

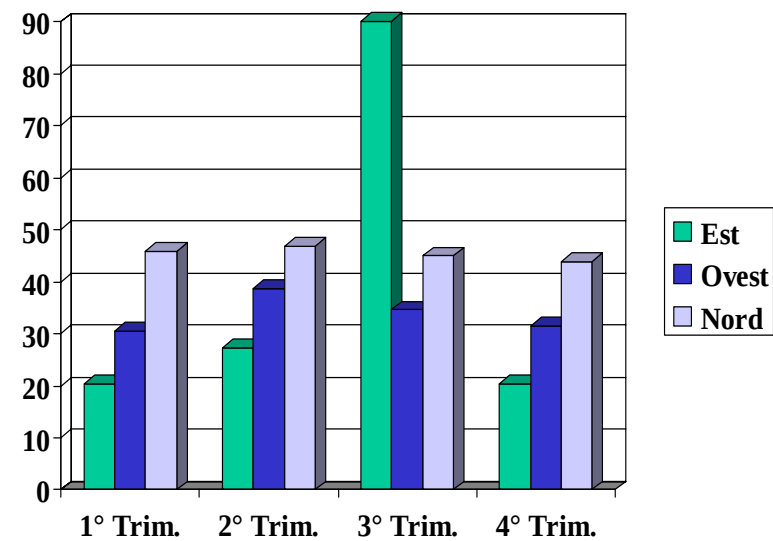


Rappresentazioni grafiche di frequenza di un evento: L'**istogramma** a barre

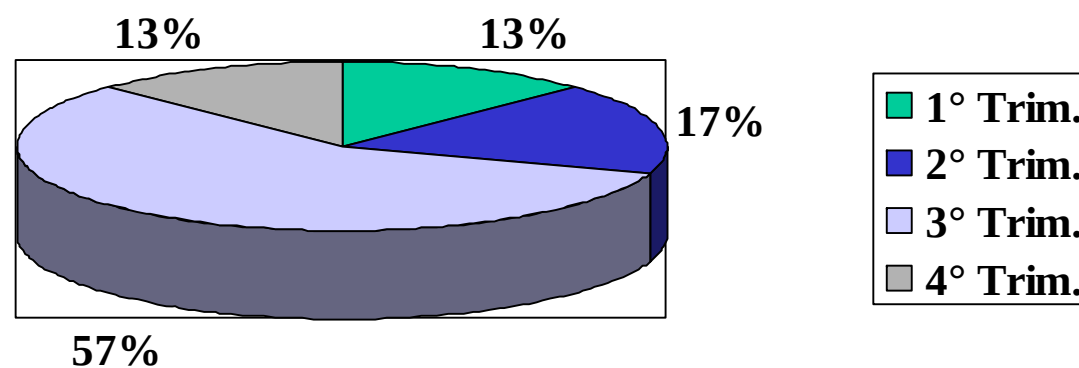
SI



NO



Rappresentazioni grafiche di frequenza di un evento:
il **diagramma a torta**



“Tutta la scienza è fondata sulle misure” (Helmholtz)

La **probabilità** p che un evento si verifichi può essere definita
come
la proporzione di volte in cui l'evento avviene su un numero
totale di prove **sufficientemente grande**.

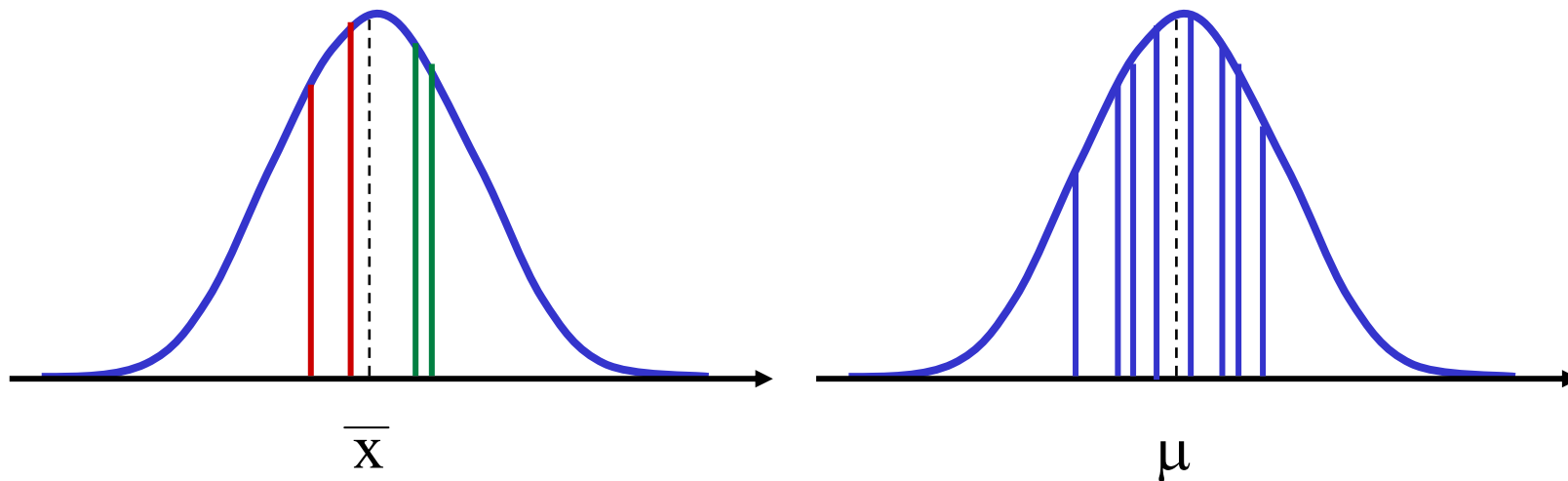
$$0 < p < 1$$

In una determinazione di una grandezza, la misura tanto più si avvicina all'evento certo ($p = 1$) quanto più grande è il numero di volte in cui la misura è ripetuta (n) e cioè tanto più la media osservata $\bar{x}$ delle n misure è vicina alla media vera (μ).

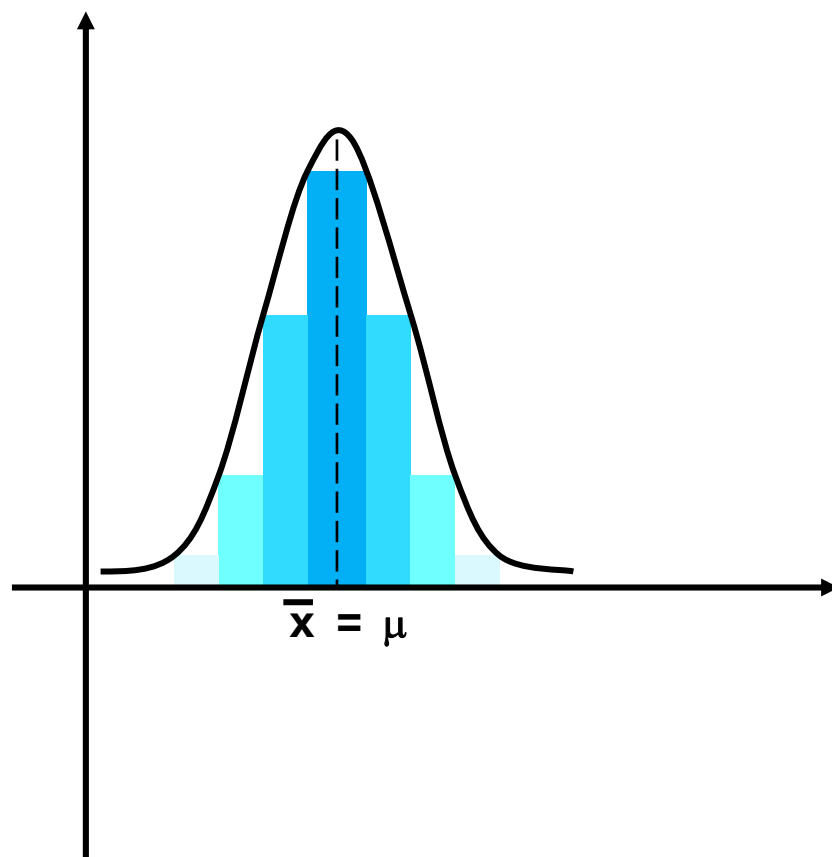
ATTENZIONE : il numero di osservazioni
(misure) deve essere sempre e comunque

$$n \geq 3$$

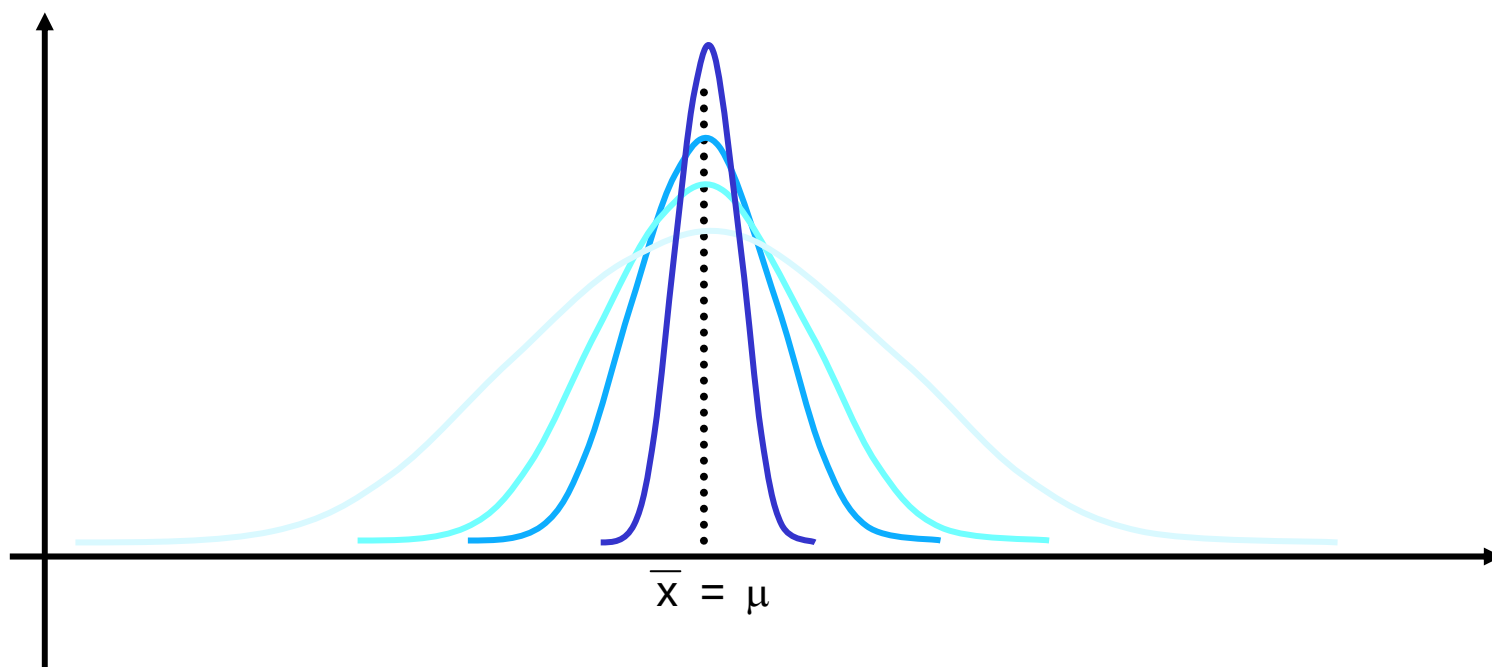
la curva BINOMIALE dell'ERRORE è una
gaussiana



Per un n sufficientemente grande,
la distribuzione degli errori è una **gaussiana**
la cui media è il valore vero



L'**ampiezza** della gaussiana può dipendere dallo sperimentatore, dallo strumento e dalla variabilità del fenomeno misurato.



Per essere sottoposto ad una analisi statistica il fenomeno deve essere :

- osservato in circostanze omogenee
- ripetibile (riproducibile)
- variabile

La **descrizione della variabilità** del fenomeno è lo scopo della statistica

STATISTICA

Che cos'è la statistica? A cosa serve?

Esempio:

Gene **A**: v.a. X $P(X = x_i)$

Domanda: Il gene **A** è espresso oppure no nel topo wt? Quanto è espresso?

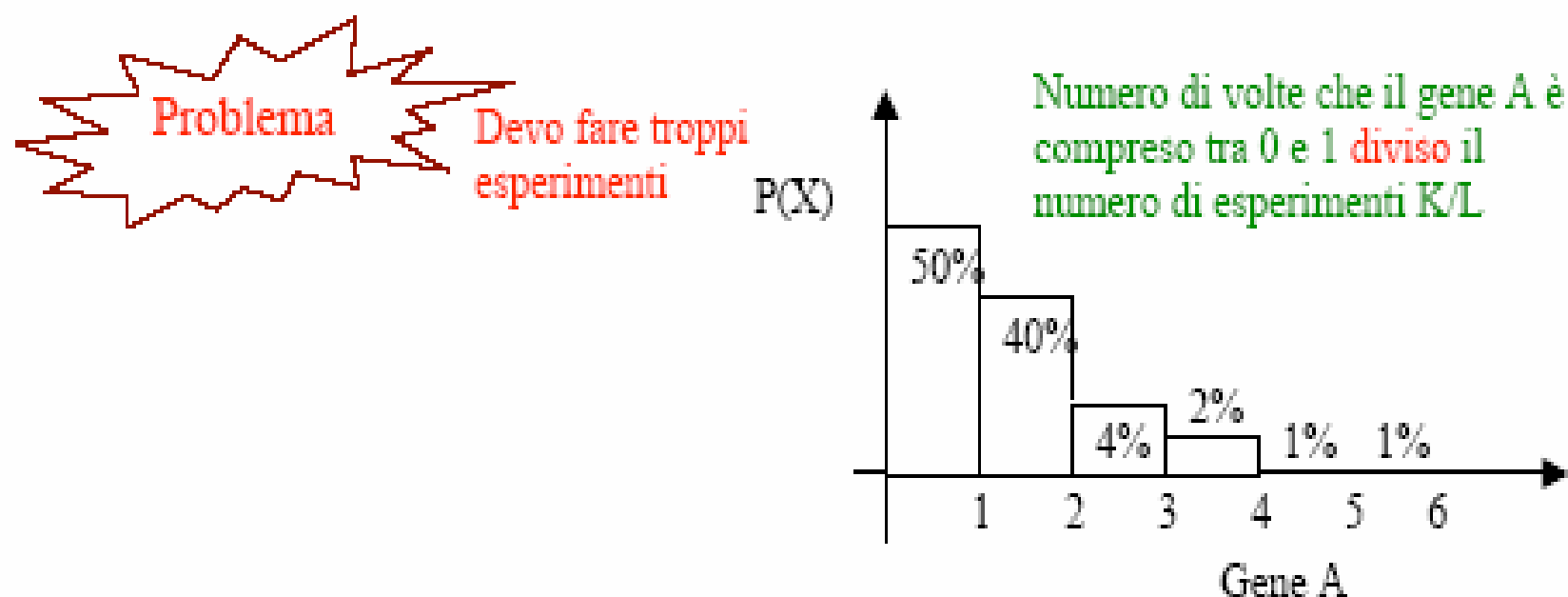
Esperimento:	è espresso?	Quanto?
Risposta classica	SI oppure NO	gene A = 4
Risposta statistica	Si (96%) e NO (4%)	gene A = $4,1 \pm 0,2$

INFERENZA STATISTICA: stima di $P(X)$ dalle misure sperimentali

Per dare la risposta statistica dobbiamo conoscere $P(X)$. MA NON LA CONOSCIAMO!

→ Come posso fare?

Soluzione 1: Ripeto lo stesso esperimento molte volte, (L)



Soluzione 2: Cerco di stimare solo alcune proprietà di X , come la media $E(X)$ e la varianza $E[(X-E(X))^2]$

MISURE DI DISPERSIONE O DI VARIABILITA'

Servono a esprimere il grado di variabilità o dispersione attorno al valore medio.

Naturalmente le fonti di variabilità possono essere tante e si suole raggrupparle in tre grandi gruppi:

- *variazione biologica*
- *temporale*
- *errori di misura.*

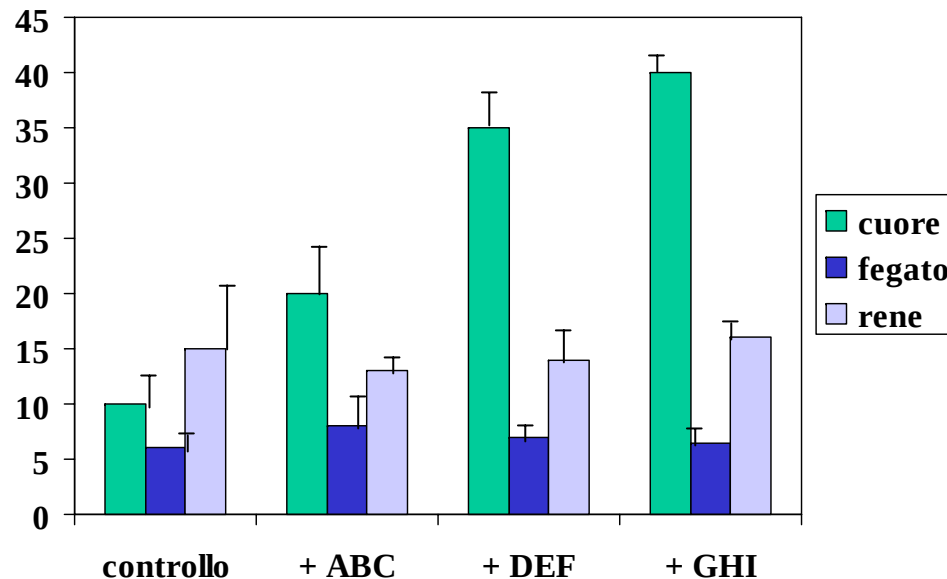
SORGENTI DI VARIAZIONE

In medicina e biologia le misure cliniche quantitative sono esposte a tre tipi fondamentali di variazione:

- 1) **biologica**: tutti i fattori che rendono diversi i vari individui: età, sesso, razza, dieta, ecc; la variazione dovuta alla variabilità biologica è detta anche "variabilità tra soggetti";
- 2) **temporale** : tutti i fattori che determinano variazioni nello stesso individuo da un momento all'altro;
- 3) **errore di misura**: tutti i fattori che tendono a produrre differenze in misure diverse dello stesso fenomeno (strumenti di misura, errori tecnici, diversità dell'osservatore, ecc); tale variabilità è detta anche "variabilità entro soggetti".

L'analisi della varianza è la tecnica statistica che permette di disaggregare la varianza totale in due componenti: quella inter-soggetti e quella intra-soggetti.

Dati scientifici richiedono l'indicazione di
parametri statistici (ad.es. media $\pm$ DS)



Media aritmetica : $\bar{x} = \frac{\sum x_i}{n}$

Deviazione standard : σ

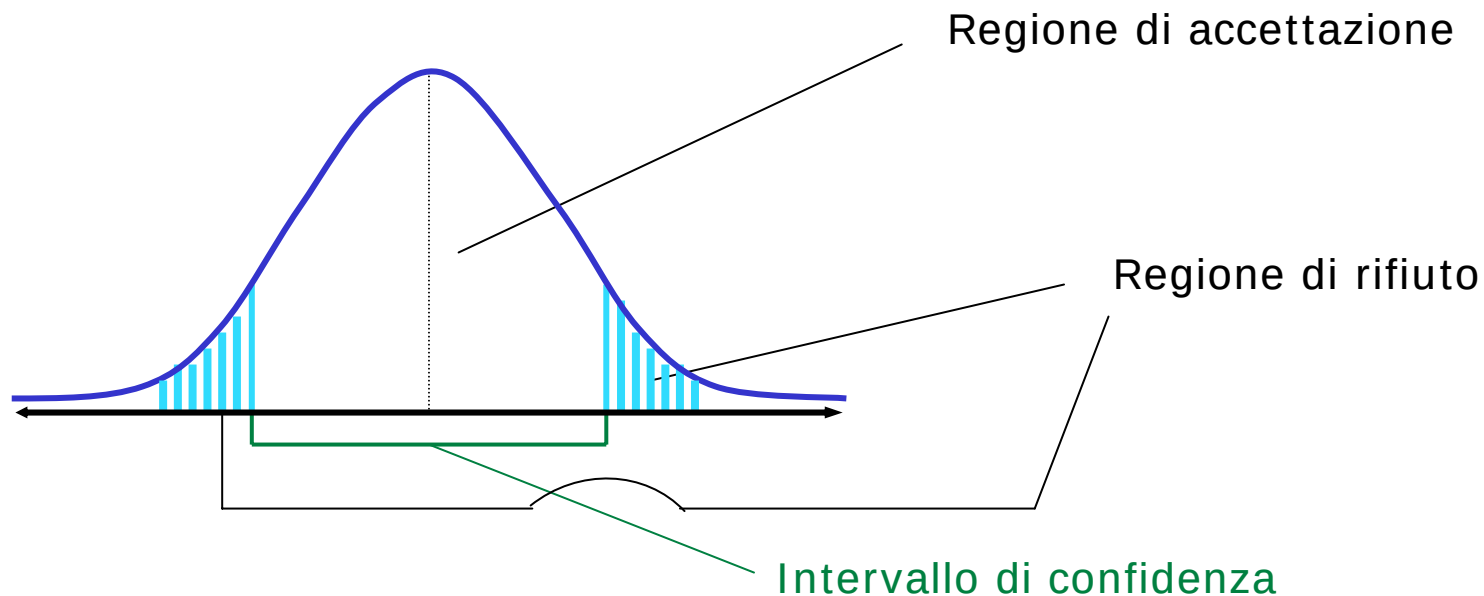
$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

	controllo	+ ABC	+ DEF	+ GHI
cuore	10 $\pm$ 2.8	20 $\pm$ 3.7	35 $\pm$ 3.2	40 $\pm$ 1.8
fegato	7 $\pm$ 1.9	9 $\pm$ 3.0	8 $\pm$ 2.1	7.5 $\pm$ 1.2
rene	15 $\pm$ 4.9	13 $\pm$ 1.2	14 $\pm$ 3.0	16 $\pm$ 1.5

Dati scientifici richiedono una **analisi di significatività**

Il **confronto** statistico tra due serie di dati consente di valutare se quelle due serie sono **significativamente diverse** e con quale **intervallo di confidenza**.

Test del t di Student

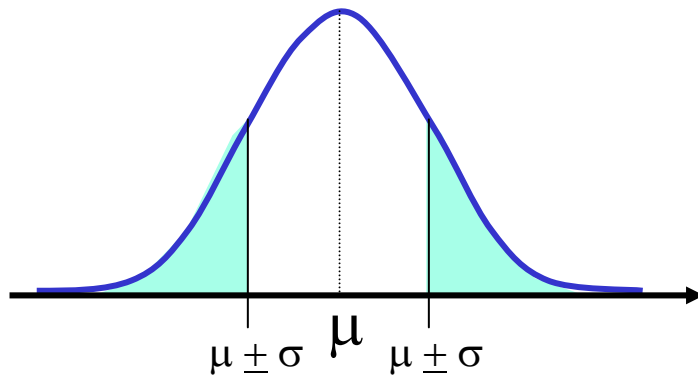


IL TEST DI STUDENT

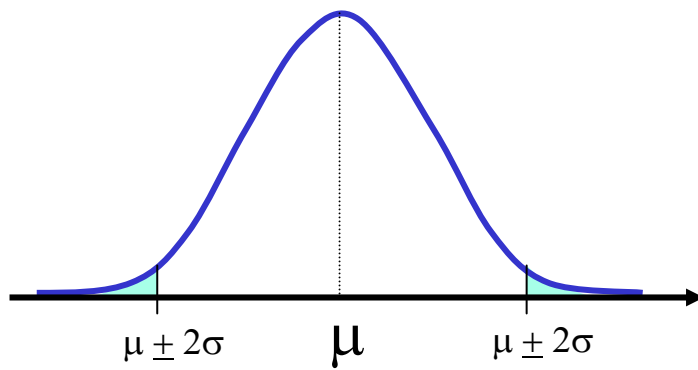
Si parte dal presupposto che il nostro campione discenda dalla popolazione:
e' questa la *ipotesi nulla o zero (H_0)*
l'ipotesi nulla è **vera** quando campione e popolazione coincidono;
invece l'ipotesi nulla è **falsa** o rigettata quando campione e popolazione non coincidono (H_1 ipotesi alternativa).

I test di significatività sono test statistici che quantificano i dati in senso di probabilità: i livelli del 5 % (0.05) e dell' 1 % (0.01) sono livelli accettati come *limiti del tutto convenzionali* per stabilire la significatività di uno scarto dall'ipotesi zero.

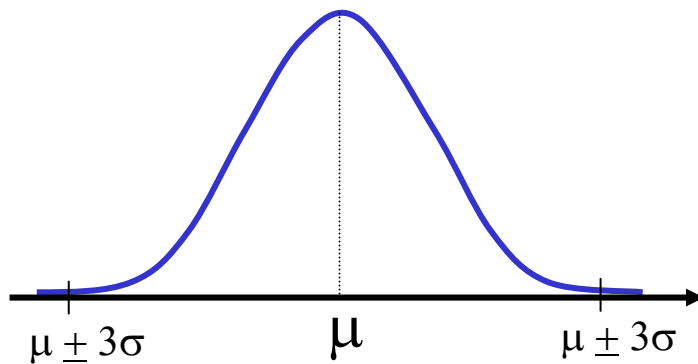
Il livello di 5 % sta ad indicare che ci sono 95 probabilità su 100 che il mio campione non derivi dalla popolazione, cioè che sia diverso. Tuttavia, esistono sempre 5 probabilità su 100 che tali differenze siano casuali e che quindi il campione derivi dalla popolazione.



$$\mu \pm \sigma \quad p = 68.27 \%$$



$$\mu \pm 2\sigma \quad p = 95.45 \%$$



$$\mu \pm 3\sigma \quad p = 99.73 \%$$

POPOLAZIONE (UNIVERSO)

H_0

H_1

CAMPIONE

H_0

Esatto
(vero neg)
(non ci sono
diff)

Errore β
(falso neg)

H_1

Errore α
(falso pos)

Esatto
(vero pos)
(ci sono diff.)

L'errore alfa (detto anche di I° tipo) prende per valide differenze che in realtà (nell'universo) non esistono.

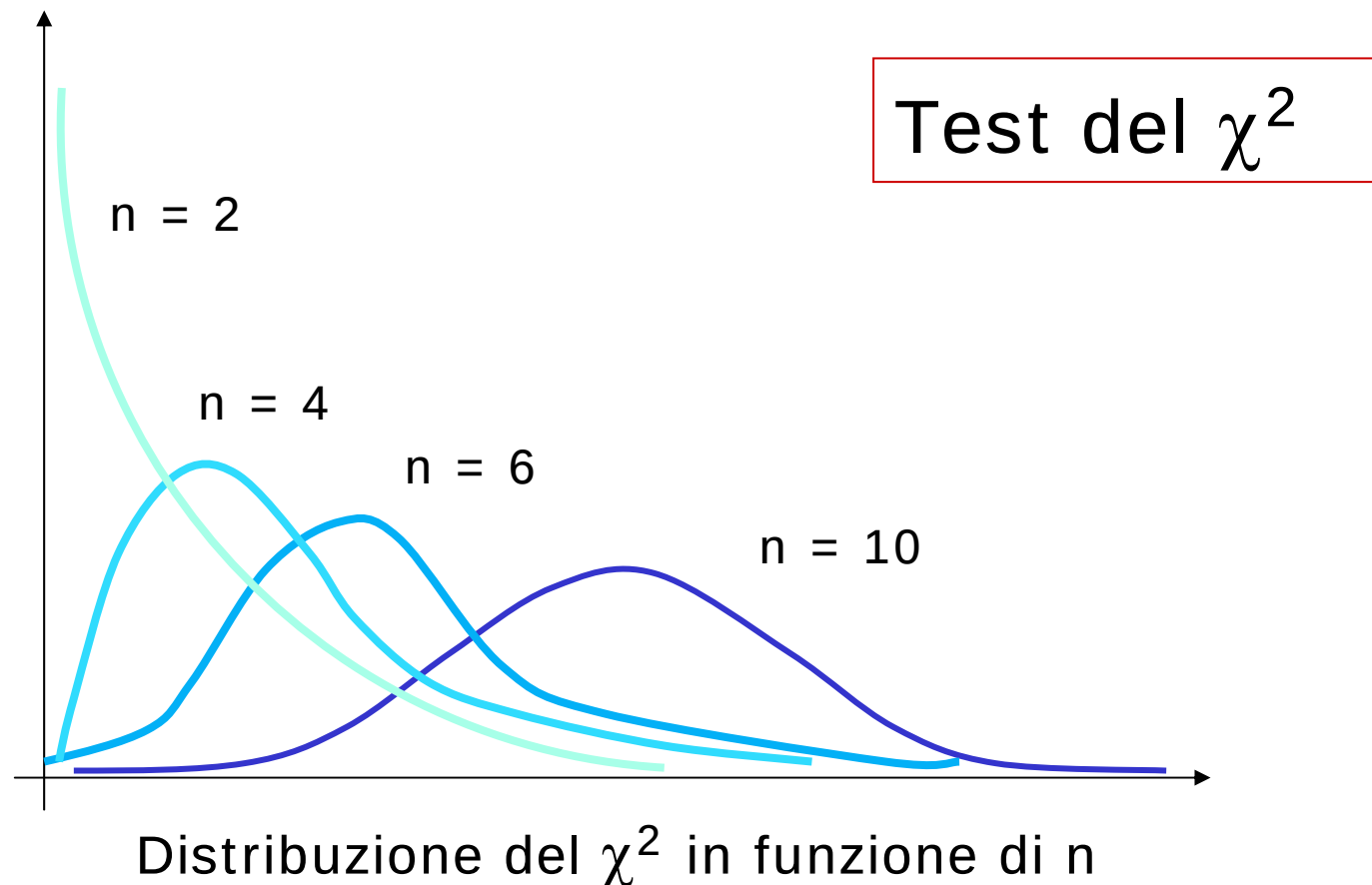
Le tecniche per ridurre l'errore alfa si chiamano "significatività".

L'errore beta (detto anche di II° tipo) non si accorge di differenze che realmente sono presenti nell'universo.

Le tecniche per ridurre l'errore beta si chiamano "potenza".

E' ovvio che più aumenta il numero delle osservazioni e più si riducono gli errori alfa e beta.

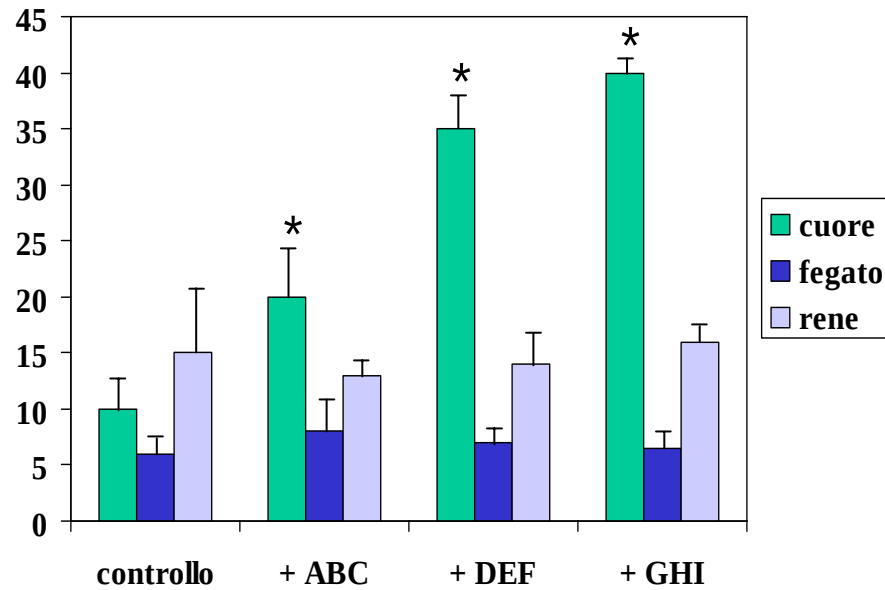
Quando si conosce solo la **media** di un **piccolo campione** rispetto ad una **grande popolazione**, di cui non si conosce la deviazione standard vera σ , si può individuare **l'intervallo di confidenza**, utilizzando il



Il **chi quadro** indica la misura in cui le frequenze osservate in ogni casella della tabella differiscono dalle frequenze che ci aspetteremmo se **non** ci fosse associazione fra i contenuti che definiscono le righe e le colonne.

In altre parole **verifica se esiste una relazione fra la variabile che distingue le righe e la variabile che distingue le colonne.**

Ad esempio :



* significativamente
diverso
dal controllo, $p > 0.01$

	controllo	+ ABC	+ DEF	+ GHI
cuore	10 ± 2.8	20 ± 3.7 *	35 ± 3.2 *	40 ± 1.8 *
fegato	7 ± 1.9	9 ± 3.0	8 ± 2.1	7.5 ± 1.2
rene	15 ± 4.9	13 ± 1.2	14 ± 3.0	16 ± 1.5

La correlazione lineare

E' una procedura statistica parametrica (e quindi usa variabili ad intervalli con distribuzione "normale") che studia le relazioni che intercorrono tra 2 variabili.

Quando la funzione che le lega è di primo grado ($y=a+bx$) la relazione è di tipo lineare.

Se vogliamo descrivere la forza della relazione
fra le due variabili si calcola
il **coefficiente di correlazione** "r" di Pearson,
che è un numero
compreso fra -1 e +1.

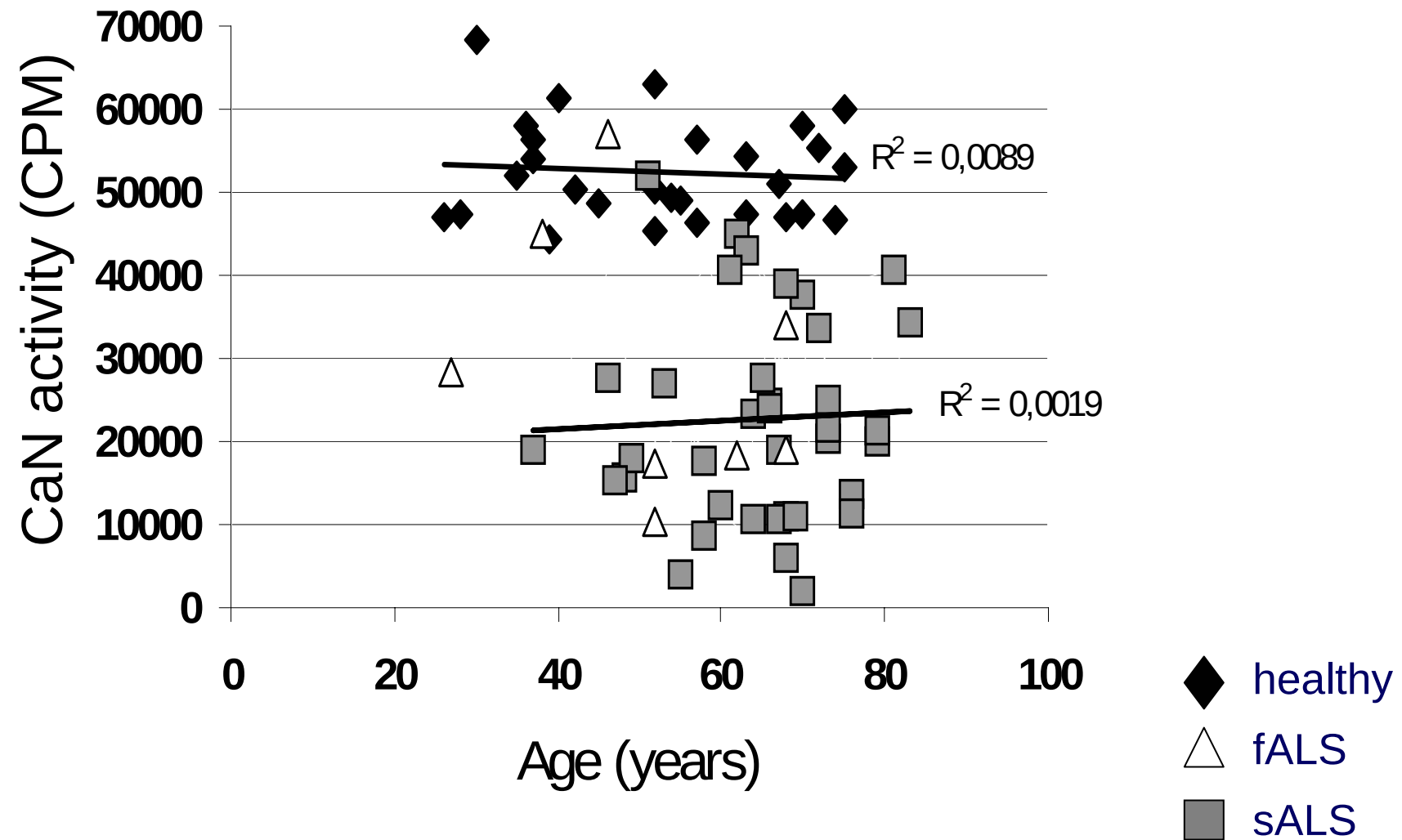
Tanto più è stretta la relazione tanto più "r"
è prossimo a 1 (in valore assoluto);
se è vicino a 0 la relazione è inesistente.

E' importante ricordare che " r " **non misura un rapporto di causa-effetto tra le variabili**, ma solo la forza della dipendenza lineare l'una dall'altra.

Il segno di " r " indica il verso di tale associazione: positivo quando le 2 variabili variano nello stesso senso, negativo quando al crescere dell'una corrisponde il calare dell'altra.

Il coefficiente " r " e' **simmetrico**, cioè il legame che collega X e Y è lo stesso che c'è tra Y e X .

Calcineurin activity is oxidatively impaired in sALS and fALS patients



Problema:

Rappresentare i seguenti dati sperimentali con un istogramma dei valori medi e deviazioni standard delle serie di misure.

	Misura 1	Misura 2	Misura 3	Misura 4
Controllo	1.31	1.47	1.20	0.95
Trattamento B	0.83	1.27	1.13	1.38
Trattamento C	1.52	1.48	1.04	1.35

Come potreste valutare se i tre trattamenti differiscono rispetto ai valori di controllo ?