CERTEZZA ⟷ CONCLUSIONE

- ABBASTANZA SICURO
- MOLTO SICURO
- PROBABILE
- IMPROBABILE

(qualitativo)

Nelle analisi chimiche

↳ CERTEZZA DI UN EVENTO IN MODO QUANTITATIVO

come?

<u>TEST STATISTICI!</u>

Domande:

1) di quanto si è sicuri che il valore ottenuto sia vicino al valore vero?
2) e che il valore ottenuto sia uguale o diverso rispetto a quello ottenuto per il medesimo campione in tempi diversi e da altra persona?

▷ (1) DOMANDA FONDAMENTALE x LA SCIENZA SPERIMENTALE

▷ (2) • CONCORDANZA FRA LABORATORI
• VALIDAZIONE DEI METODI
• REGOLE INTERNAZIONALI

NS

ESEMPIO (ORIGINE DEGLI ERRORI)

- Detn del fenolo (m/v) in uno spray x la gola con HPLC.

2 PIPETTE AUTOMATICHE → 1,00 ml ; → 10,0 μl

PROTOCOLLO: 1.00 ml H_2O DEIONIZZ/DIST

↓

FIALA DI VETRO

+

100 μl SPRAY

↓

AUTOCAMPIONATORE

↓

CROMATOGRAFO (UV 210 nm)

ERRORI!

1) Volumi prelevati (2 pipette)
2) Volume iniettato autocamp.
3) Reagenti contaminati
4) Relazione Abs - conc sbagliata
5) contaminazione incrociata (si mescolano 2 campioni ≠ nel cromatografo).

COME VALUTO TALI ERRORI?

REPLICHE!!! + PROVE INDIPEND.

⇓

+ CERTEZZA { PIÙ FIDUCIA CHE IL RISULTATO CADA IN UN INTERVALLO.

~~UNA MISURAZIONE SOLA~~ NO!

\+ MISURAZIONI → x_1, x_2, x_3, x_4, … x_n

DATI DIVERSI !!!

QUALE RIPORTO?

- MEDIA ARITMETICA: (O MEDIA)

$$\boxed{\bar{x} = \frac{\sum_i x_i}{n}}$$

- DEVIAZIONE DALLA MEDIA:

$$\boxed{d_i = x_i - \bar{x}}$$

- INTERVALLO

$$w = x_{\text{MAGGIORE}} - x_{\text{MINORE}}$$

- VARIANZA $= \dfrac{d_1^2 + d_2^2 + d_3^2 + \ldots + d_N^2}{N-1}$

- DEV. STANDARD $= s = \sqrt{\dfrac{d_1^2 + d_2^2 + d_3^2 + \ldots d_N^2}{N-1}}$

- DEV. STD RELATIVA $= \dfrac{s}{\bar{x}}$

- DEV. STD RELAT. % $= \dfrac{s}{\bar{x}} \cdot 100$ (coefficiente di variazione)

AFLATOSSINA NEL CIOCCOLATO

(N.B. CANCEROGENA)

LIMITE PRATICO DI DETN.:

(POCHI μg/Kg

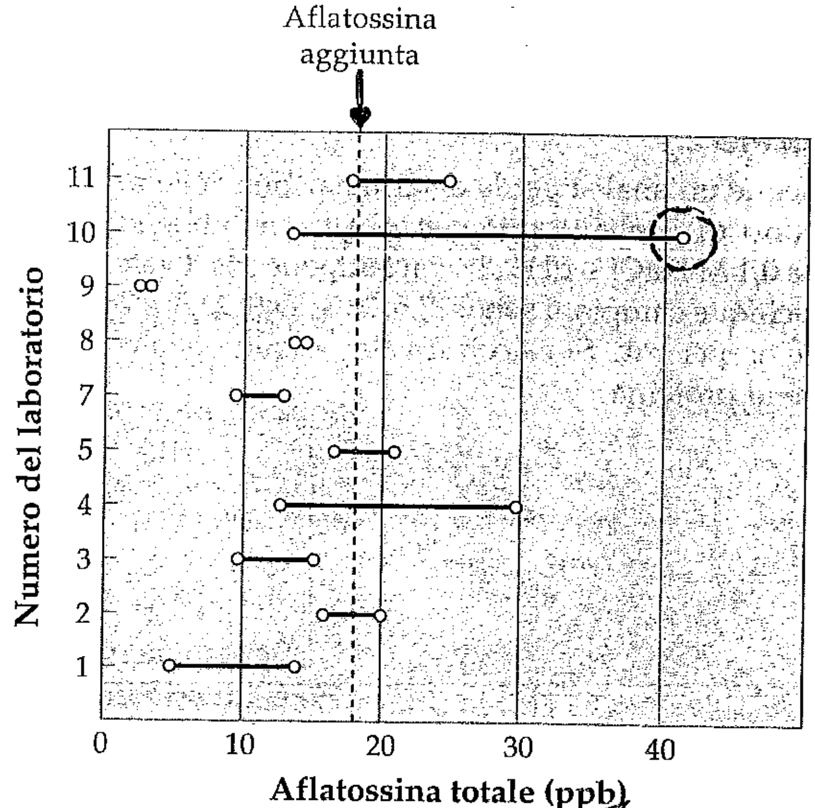


Grafico dei risultati dello studio interlaboratorio sulla determinazione dell'aflatossina nei semi di cacao. La linea tratteggiata verticale indica la quantità di aflatossina aggiunta ai semi non contaminati dai supervisori dell'intero studio. I cerchietti vuoti indicano i valori dei due replicati ottenuti nelle determinazioni fatte ad opera di ciascuno dei dieci laboratori partecipanti al circuito, mentre il valore cerchiato è stato giudicato un dato sospetto. Si noti inoltre che i due risultati ottenuti dal laboratorio 9 sono quasi identici tra loro ma piuttosto lontani dal valore vero, il che indica dati precisi ma non accurati. [Da: HORWITZ, W. 1982. *Anal. Chem.* 54:67A-70A. Con permesso di *Analytical Chemistry*. Copyright © 1982 American Chemical Society.]

[redacted]

[redacted]

ERRORI CASUALI e DISTRIBUZIONE GAUSSIANA

- gli errori casuali sono INDIPENDENTI FRA LORO

ES. MONETA → TESTA
→ CROCE

Il lancio di una moneta può portare a testa o croce, ma il risultato del lancio NON ha influenza su quello successivo!!!

SE GLI ERRORI SONO CASUALI:

→ I VALORI SI DISTRIBUISCONO IN MODO CARATTERISTICO SU ENTRAMBI I LATI DEL VALORE MEDIO

↓

CURVA GAUSSIANA (vedi fig. 63)

INTERVALLO DI FIDUCIA

1 MISURAZ.: INCERTEZZA
2 MISURAZIONI : MEGLIO
3 "
+ "

↓

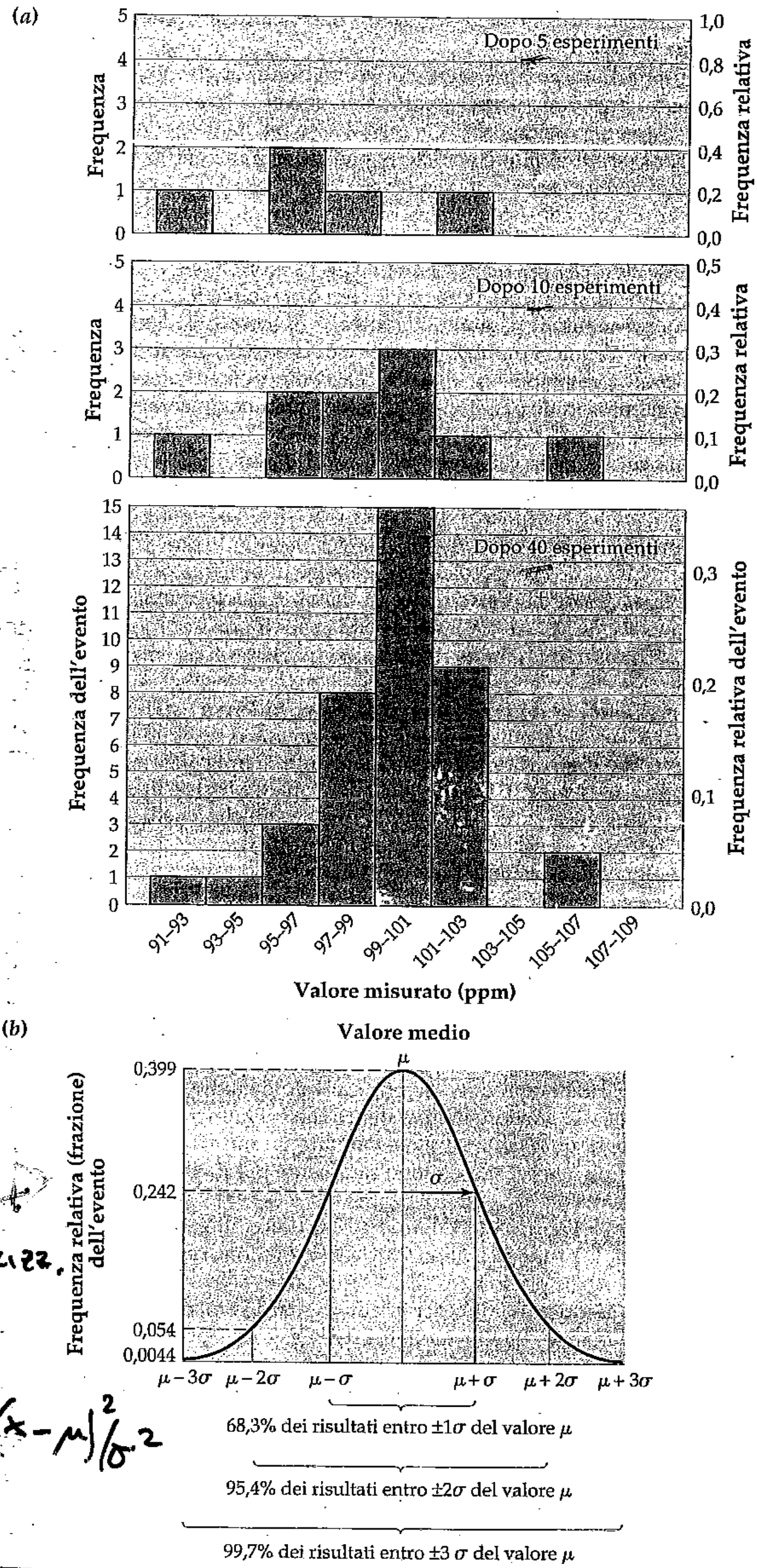
SEMPRE + CERTEZZA

..MA.. OLTRE UN CERTO NUMERO DI MISURAZIONI È INUTILE ANDARE LA CERTEZZA NON AUMENTA PIÙ DI TANTO.

FIGURA 2.2
Grafico dei risultati di una serie di 40 esperimenti classificati per gruppi di valori adiacenti.
(a) La base di ogni barra dell'istogramma rappresenta un intervallo di 2 ppm, mentre l'altezza è il numero di risultati che si presentano in ciascun intervallo. Sull'asse y sono riportate due scale: quella della frequenza o numero di volte con cui si presenta un evento, e quella della frequenza relativa. Il valore medio è 100 ppm e i tre grafici, come riportato su di essi, descrivono la situazione dopo 5, 10, e 40 esperimenti.
(b) Per un numero «molto grande» di misurazioni, la distribuzione dei risultati tende ad avvicinarsi a quella di una curva gaussiana, caratterizzata da una deviazione standard σ e da un massimo in corrispondenza del valore medio. Quando si **normalizza** la gaussiana in modo che l'area sotto la curva sia unitaria, l'altezza del massimo diventa 0,399. Il valore dell'ordinata diventa 0,242 in corrispondenza di 1σ, 0,054 in corrispondenza di 2σ, e 0,004 4 per 3σ. Dopo aver costruito la gaussiana con un numero alto di risultati ottenuti da misurazioni ripetute, per quelle successive, relative allo stesso misurando nello stesso sistema, ci si aspetta questo comportamento:

- il 68,3% dei risultati cadrà entro $\pm 1\sigma$ dal valore medio e
- il 95,4% dei risultati cadrà entro $\pm 2\sigma$

Questi valori di percentuale corrispondono alle aree sotto la curva gaussiana normalizzata, delimitate da valori positivi e negativi di σ.



CURVA GAUSSIANA NORMALIZZ. (CARATTERIZZATA DA σ)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

N.B. n molto grande

CURVA CAMPANULARE (SIMMETRICA)

6/>

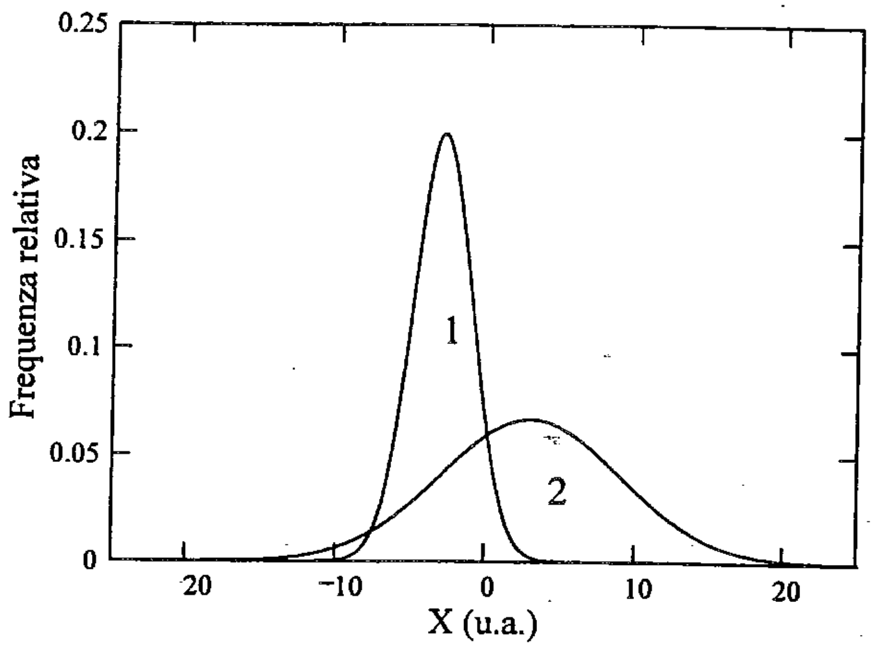


Figura 3.4: Distribuzioni normali di due popolazioni caratterizzate da diverso valore di σ e di μ: $\mu_1 = -3{,}0$; $\sigma_1 = 2{,}0$; $\mu_2 = 3{,}0$ $\sigma_2 = 6{,}0$.

DEVIAZIONE STANDARD DEL VALORE MEDIO (μ)

Viene chiamata σ_m (ERRORE DELLA MEDIA)

Se σ = dev. std di ogni singola misurazione (o replica) $\rightarrow$ $\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

N = numero di misurazioni

N.B.

1) N è sotto radice $\Rightarrow$ i vantaggi in precisione dimin. oltre un certo n° di misure.

2) se voglio raddoppiare la precisione ($\sigma_m \cdot \frac{1}{2}$) devo passare da 4 a 16 REPLICHE.

Perché $\sigma_m < \sigma$?!

ERR. DELLA MEDIA

ERRORE CASUALE MEDIO DI OGNI SINGOLA DETERM.

- Il risultato di ogni singola determ. può essere casualmente alto o basso rispetto al valore vero μ..

.. con + misurazioni gli errori casuali tendono a cancellarsi perché la media degli errori individuali $\rightarrow 0$

X CURVA GAUSSIANA :

- 68.3% DI FIDUCIA che $\bar{X}$ (VALORE MEDIO) sia compreso tra $\pm\sigma_m$
- 95.4% " " $\bar{X}$ " " " $\pm 2\sigma_m$

7/s σ_m dip. dal n° di repliche.

LIMITE DI FIDUCIA: LF 95%

È quello + comune!!
Cioè 95% di probabilità: "limite di fiducia al 95%"

L'intervallo è un po' + ristretto di $\pm 2\sigma_m$ dal valore medio (Ricorda $\underline{\underline{95.4\%}}$)

$\rightarrow \pm 1.96\,\sigma_m$ quindi noto σ:

$$\mu = \bar{x} \pm 1.96 \underbrace{(\sigma/\sqrt{n})}_{\sigma_m}$$

μ: valore vero

$\bar{x}$: valore medio

MA!!!

PROBLEMA PRATICO:
Il valore di σ (errore casuale × ogni misura) NON è NOTO.
e quindi?

× LF 95% t è sempre > di 1.96 visto che N non è ∞.
σ viene sostituita da S

"t di "Student" "

RISULTATO :

$$\mu = \text{media} \pm LF\ 95\%$$
$$= \bar{x} \pm t \cdot \left(s / \sqrt{N} \right)$$
$$= \bar{x} \pm \left(\frac{t}{\sqrt{N}} \right) \cdot s$$

N MISURAZIONI	GRADI DI LIBERTÀ $(n-1)$	t (x LF 95%)	$t/\sqrt{N}$
2	1	12.71	8.99
3	2	4.30	2.48
4	3	3.18	1.59
5	4	2.78	1.24
6	5	2.57	1.05
7	6	2.45	0.936
8	7	2.36	0.831
9	8	2.31	0.77
10	9	2.26	0.71
20	19	2.09	0.467
30	29	2.04	0.372
120	119	1.98	0.180
∞	∞	1.96	–

N.B. ANDAMENTO t.

PRESENTAZIONE DEI RISULTATI ANALITICI

IMPORTANTE!!!

☞ Nessun risultato sperimentale è significativo a meno che ad esso non sia associata la stima dell'errore sperimentale (⇨ esecuzione di repliche).

☞ Il risultato di una singola analisi può cadere in qualunque punto della curva normale dell'errore e quindi, preso come tale, non contiene in sé alcuna informazione circa la sua affidabilità.

☞ Per valutare la precisione di un risultato devono essere specificati: n° repliche, deviazione standard o deviazione standard della media.

☞ In pratica il n° di repliche è sempre basso, e si riporta **l'intervallo di fiducia**, $\mu = x_m \pm ts/n^{1/2}$ (è necessario specificare il livello di fiducia prescelto e il numero di osservazioni).

ESEMPIO:

la standardizzazione di una soluzione di analita è stata effettuata eseguendo "n" titolazioni con una soluzione a titolo noto, il risultato si riporta con il seguente formato:

$$\mathbf{C = (x_m \pm t_{1-\alpha\ v}\ s_m)}\ \boldsymbol{mol/l}\ \mathbf{(n, 1\text{-}\alpha)}$$

$s_m = s/n^{1/2}$
1-α: è il livello di fiducia prescelto
v = n- 1: è il numero di gradi di libertà.

$$\mathbf{C = (0{,}1006 \pm 0{,}0004)\ mol/l\ (n = 11;\ 1\text{-}\alpha = 0.95)}$$

Questo formato permette a chiunque di risalire facilmente alla deviazione standard stimata nel corso delle undici titolazioni:

$$s = 0{,}0004\ (11^{1/2}/2.23) = 0{,}0006\ mol/l.$$

N.B.: l'intervallo ts_m ± 0,0004 mol/l nell'esempio viene riportato ad una sola cifra significativa. Se fosse stato calcolato il valore ±0,00039385, dato che la quarta cifra decimale è la prima cifra incerta non ha alcun senso riportare le cifre successive.

ESEMPIO 2:

si è analizzata una soluzione standard valutando la concentrazione di Pb (certificata ad esempio come 2.25 ppm), e si sono fatte 11 misure ottenendo una media di

y= 2.303 e s= 0.040.

Affinchè non ci siano errori sistematici il valore accertato della concentrazione dello standard deve cadere entro l'intervallo $\bar{x} \pm ts/n^{1/2}$

Se si assume come accettabile il livello di fiducia del 95% dalla tabella si ottiene:

n=11 perciò n-1=10

dalla tabella t=2.23 cioè μ=2.30 ± 0.03

Dato che la concentrazione dello standard (certificata=2.25) cade fuori dell'intervallo di fiducia, esistono meno di 5 probabilità su 100 che la misura non sia affetta da errori sistematici.

DATI DISCORDI : OUTLIER

PROBLEMA: ELIMINARE I DATI ANOMALI!!
(LONTANI DA $\bar{x}$)
???

2 STRADE → BASATA SULLA STATISTICA
→ BASATA SU VALUTAZIONI SOGGETTIVE

TEST STATISTICO: "TEST DI DIXON" o "TEST Q"

1) ORDINARE I DATI + piccolo → + grande
2) d = differenza fra dato sospetto e valore più vicino a esso
3) w = differenza fra + grande e + piccolo
4) $Q = d/w$
5) confrontare Q tabellato in base a N

x es.

N	Q CRITICO
3	0.94
4	0.76
5	0.64
6	0.56
7	0.51

Q critico per rigettare dati con 1 probabilità su 10 di sbagliare

10/5 6) se Q calcolato > Q critico elimino il dato

7) ELIMINATO IL DATO RICALCOLO $\bar{x}$ e S
coi dati rimanenti.

N.B.. Esistono molte tabelle per il test
Q
- Esistono altri Test (Horowitz, Kelly, Grubbs Beck).
  * più sofisticati

... Esempi:

PM_{10} nell'aria. : 50 $\mu g/m^3$

- 10
- 16
- 80
- 21
- 11
- 50

1)

80
50
21
16
11
10

$d = 30$
$w = 70$
$Q = d/w = \frac{30}{70} = 0.43$

$N = 6$

$0,43 < 0.56$ OK.

2)

~~80~~
21
16
11
10

$d = 59$
$w = 70$
$Q = \frac{59}{70} = 0,84$

$N = 5.$

$0.84 > 0.64$ ELIMINARE

$\bar{x} = (21 + 16 + 11 + 10)/4$