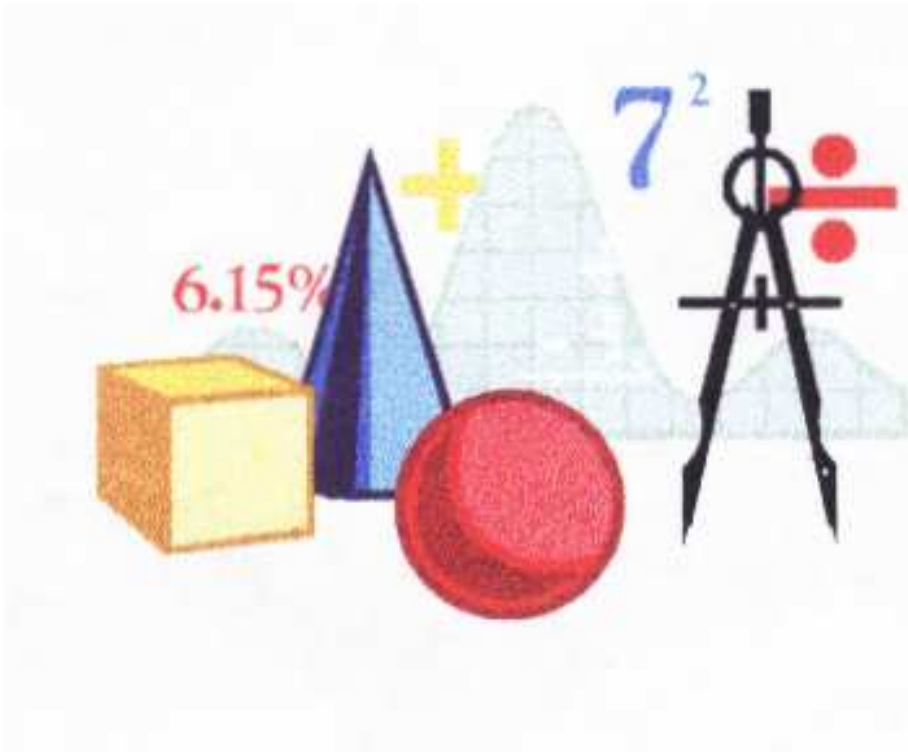
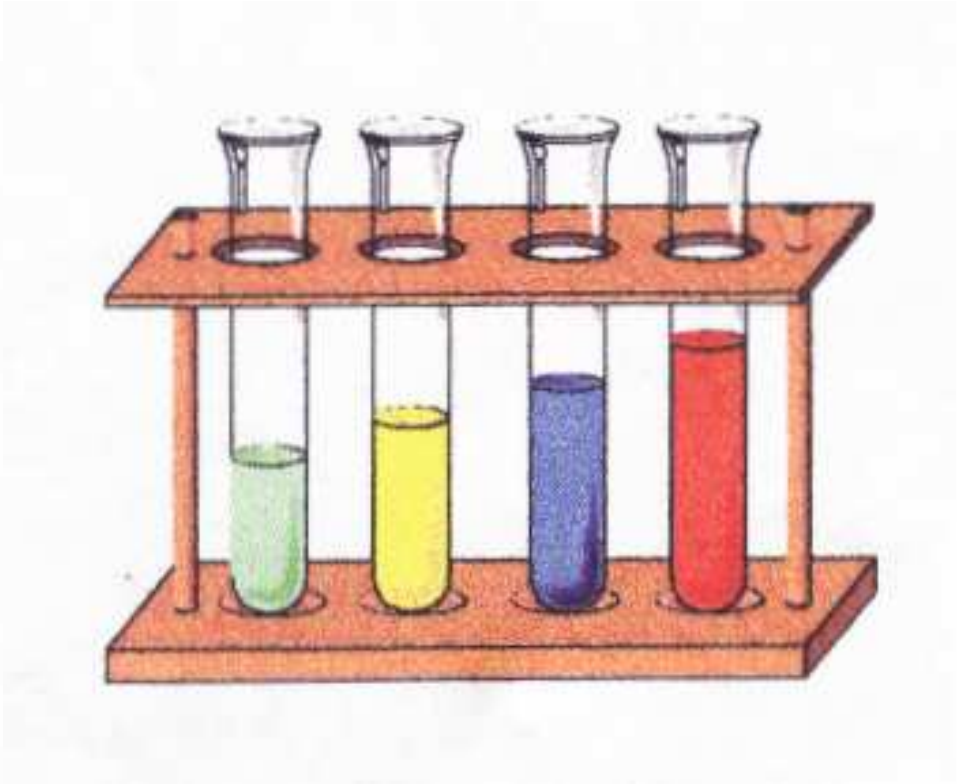
ELEMENTI DI STATISTICA
IN
CHIMICA ANALITICA





ERRORI NELLE DETERMINAZIONI ANALITICHE SPERIMENTALI

Qualsiasi determinazione analitica-sperimentale comporta un certo errore, anche se condotta con la massima cura

L'interesse del chimico analitico è rivolto alla ricerca di un metodo che fornisca risultati attendibili

Cioè risultati sperimentali il più possibile rispondenti al valore "vero" (supposto tale, stimato)

L'attendibilità di un risultato è condizionata da molti fattori, alcuni dei quali dipendono dal metodo, altri dall'esecuzione e dalla efficienza della strumentazione

I fattori più importanti che incidono sull'attendibilità di un'analisi sono:

- •La sensibilità che esprime la più piccola quantità di sostanza che si riesce a determinare con un certo metodo.
- •La specificità che è la possibilità di dosare, con un dato metodo, una specie in presenza di altre senza interferenze.
- •L'accuratezza che è la concordanza tra la media dei risultati ottenuti ed il valore "vero“ (supposto tale, stimato)
- Essa dipende sia dal metodo usato sia dall'esecuzione
- •La precisione indica l'accordo tra i vari risultati sperimentali ottenuti
- Essa non dipende dal metodo, ma dall'esecuzione

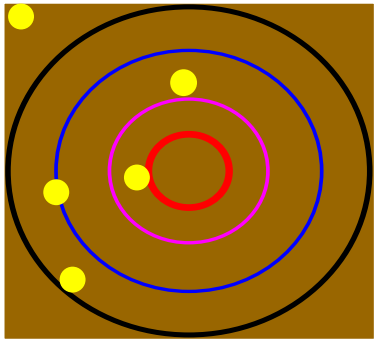
Precisione: bontà dell'accordo tra i risultati di misurazioni successive.

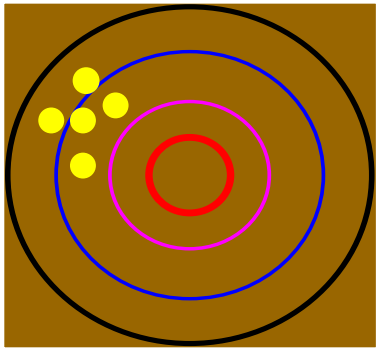
Esattezza*: bontà dell'accordo tra il risultato, x_i, o il valore medio dei risultati di un'analisi, ed il valore "vero" (supposto tale).

- Gli errori possono essere **errori casuali** o **errori sistematici**.

- Gli errori **casuali** influenzano la **precisione**, quelli **sistematici** l'**esattezza**.

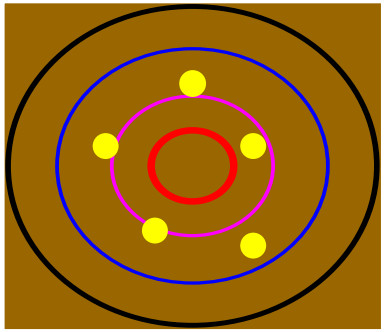
Né esatto né preciso

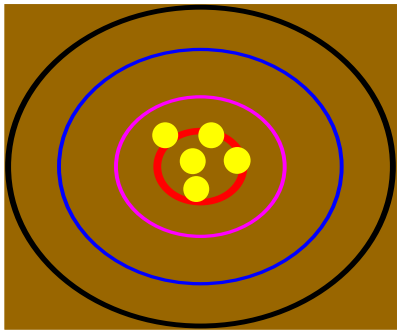




Non esatto ma preciso

Esatto ma non preciso





Esatto e Preciso *affidabile

* In base alla definizione più moderna, l'esattezza corrisponde alla vecchia accuratezza mentre l'accuratezza è la somma di esattezza e precisione.

Qualunque sia il metodo utilizzato e
Comunque avvenga l'esecuzione l'errore indica

LA DIFFERENZA TRA IL VALORE "VERO" ED IL RISULTATO SPERIMENTALE

- **L'errore è una risultante, dunque, che si ottiene dall'accumularsi di errori di natura diversa che si possono classificare in due categorie:**
- **ERRORI DETERMINATI O SISTEMATICI**
- **ERRORI INDETERMINATI O CASUALI**
- **Errori determinati o sistematici (eliminabili)**
- **Sono dovuti sempre ad una causa nota o individuabile**
- **Si ripetono ogni volta che si effettua la stessa determinazione**
- **Con lo stesso metodo e nelle stesse condizioni**

Gli errori sistematici possono essere attribuibili:

1) Al metodo, non all'abilità operativa e quindi non si possono evitare a priori , tuttavia si possono prevedere e correggere se si conosce la legge secondo la quale si verificano (parziale solubilizzazione; non completezza di una reazione; decomposizione di un precipitato; reazione secondaria; ecc.).

2) Ai reattivi o agli strumenti impiegati, e possono essere di natura chimica o fisica (reagente impuro, inesatta concentrazione analitica, variazione di volume, bilancia, burette, pipette ecc. strumenti di misura ponderale o volumetrica starati)

3) All'esecuzione, (incapacità dell'operatore ad apprezzare correttamente la variazione del colore di un indicatore, vetreria non adeguatamente pulita)

Errori indeterminati o casuali ? ≠ Sempre diversi

"Pur eseguendo la stessa analisi ed utilizzando lo stesso metodo"

- **Dovuti all'effetto di variabili incontrollate**
- **Legati a fluttuazioni indefinite di una miriade di parametri sperimentali**
- **Non possono essere individuati né corretti**

PERCHÉ

- **Non si può conoscere la causa che li ha originati**
- **Né la legge secondo cui si verificano**

TUTTAVIA POSSONO ESSERE NOTEVOLMENTE RIDOTTI

1) Operando con la massima attenzione e cura

2) Ripetendo più volte la stessa analisi con lo stesso metodo

3) Calcolando la media dei singoli risultati ottenuti etc.

- ♣INOLTRE♣
- **L'INFLUENZA degli errori indeterminati**
- **sui risultati può essere**
- **“STIMATA TEORICAMENTE”**
- **APPLICANDO l'analisi statistica alla serie**
- **di valori sperimentali ottenuti**
- **IN CONCLUSIONE**
- **un risultato analitico sperimentale è sicuramente affetto da errori sistematici che, in linea di principio, possono essere corretti conoscendone le cause che li hanno prodotti, e da errori accidentali che sono valutabili e quantificabili solo**
- **mediante l'analisi statistica.**

VALUTAZIONE STATISTICA DEGLI ERRORI INDETERMINATI

RIBADIAMO:

- ↳ **Ripetendo una stessa analisi molte volte**
  - ↳ **e nelle stesse condizioni**
- **(anche se si eliminano gli errori sistematici)**
- ↳ **non si ottengono quasi mai risultati coincidenti**

- Gli errori sperimentali si combinano tra loro in modo da rendere ogni nuova misura più o meno diversa dalla precedente.

- **Supponiamo di aver eseguito 37 volte uno stesso dosaggio, di raccogliere in tabella i risultati ottenuti in ordine crescente e di raggrupparli in un certo numero di classi (di solito da 5 a 10)**
- **CLASSE: insieme di risultati compresi in un intervallo prefissato**
- **INTERVALLO = ampiezza di classe (limite max-min)**
- **FREQUENZA: numero di risultati raggruppati in ciascuna classe**

Risultato	Frequenza	Risultato	Frequenza	Risultato	Frequenza
74,6	1	78,0	9	79,0	7
		78,0		79,1	
75,3	2	78,1		79,1	
75,4		78,1		79,1	
		78,1		79,2	
76,0	4	78,2		79,3	
76,1		78,2		79,3	
76,2		78,3			
76,3		78,4		80,0	4
				80,1	
77,0	7			80,2	
77,0				80,3	
77,1					
77,1				80,9	2
77,2				81,2	
77,3					
77,3				82,4	1

VALUTAZIONE STATISTICA DEI RISULTATI SPERIMENTALI (raggruppamento di 37 determinazioni)

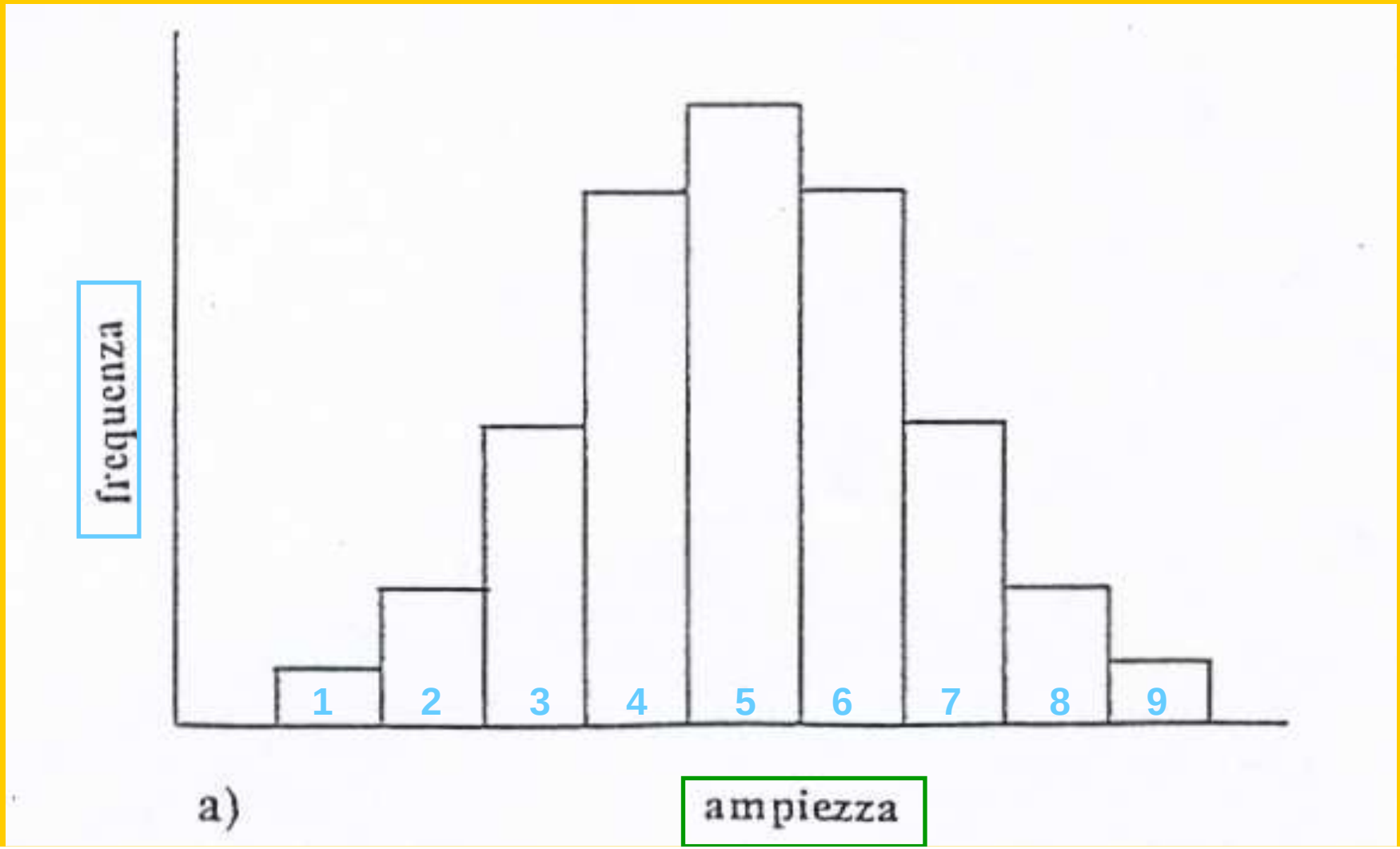
Nella tabella i risultati ottenuti sono stati raggruppati in 9 classi di ampiezza 0,5 e di frequenza diversa

Se costruiamo un grafico riportando:

IN ORDINATE LE FREQUENZE

IN ASCISSE LE AMPIEZZE

otteniamo l'istogramma a)

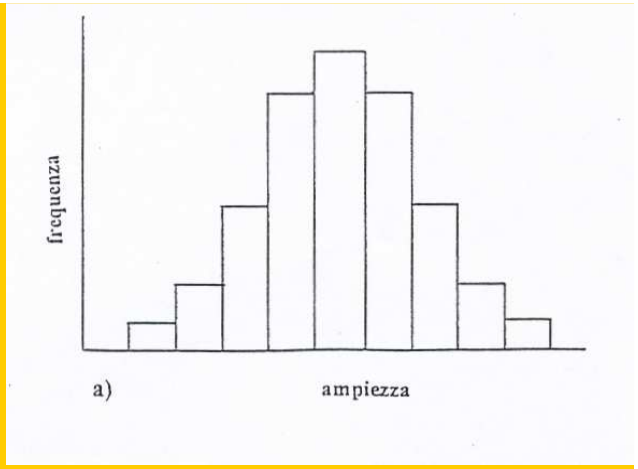


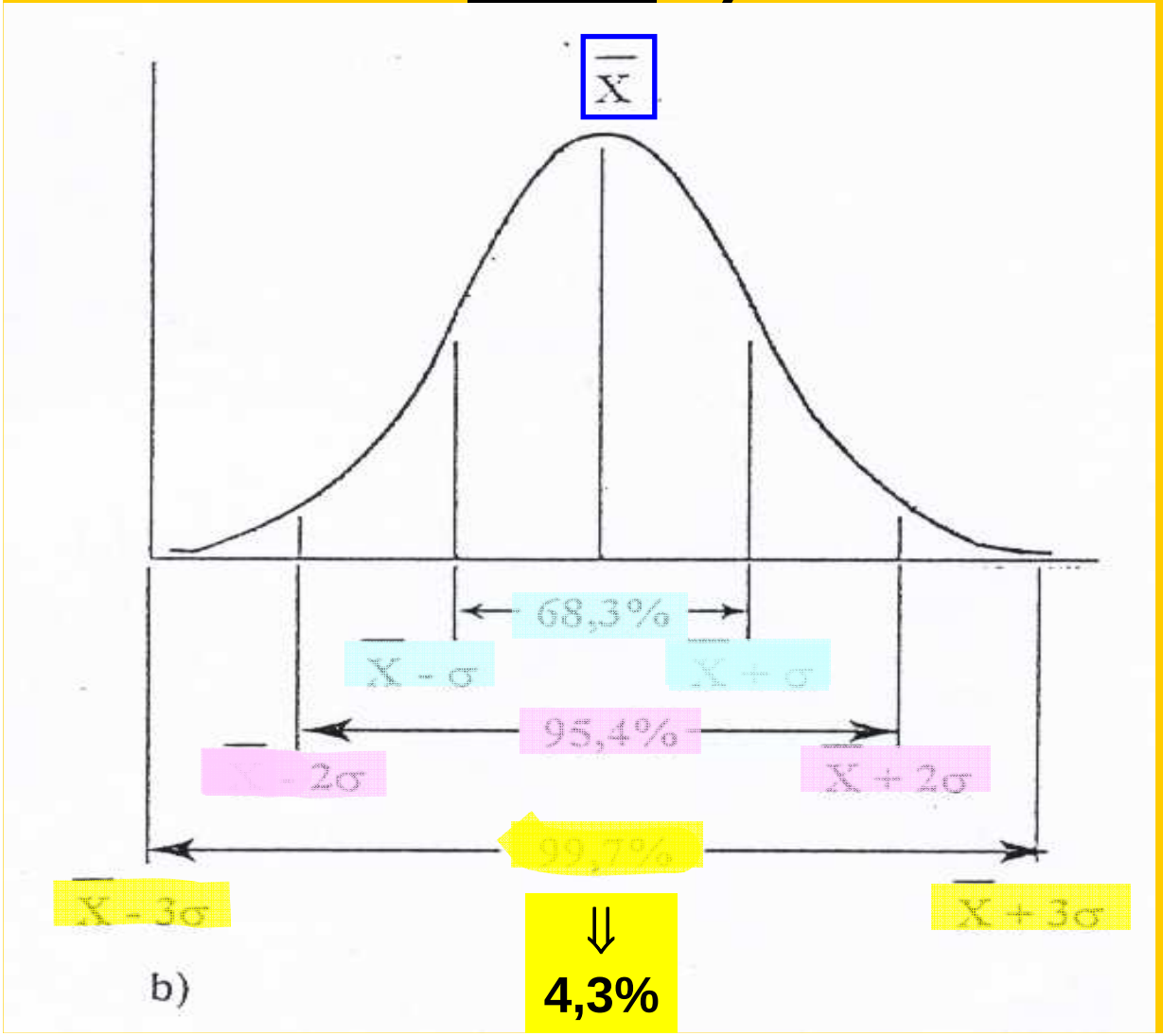
CURVA DI DISTRIBUZIONE

Idealizzando l'istogramma a)

per infinite determinazioni

si ottiene la curva b)





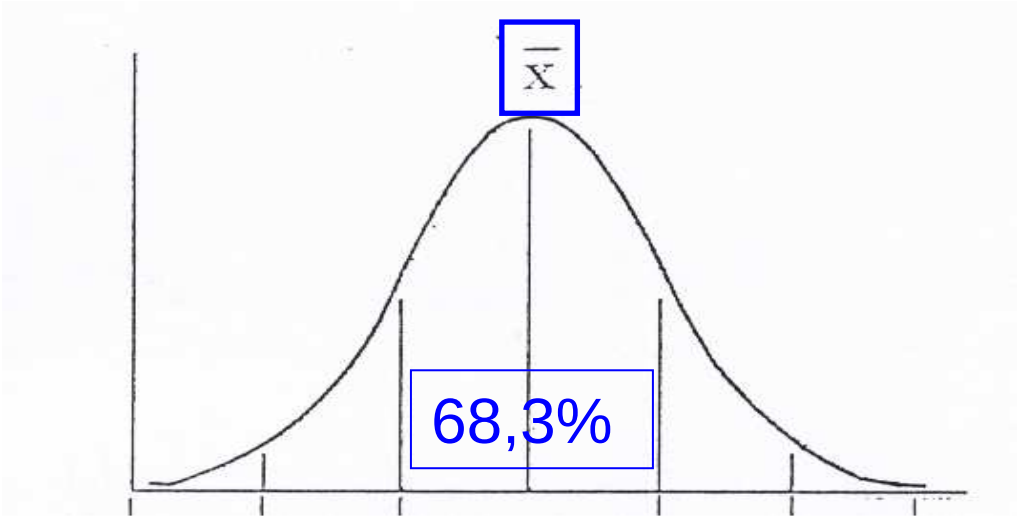
⇓

4,3%

CURVA NORMALE DI DISTRIBUZIONE, curva di Gauss o curva delle probabilità

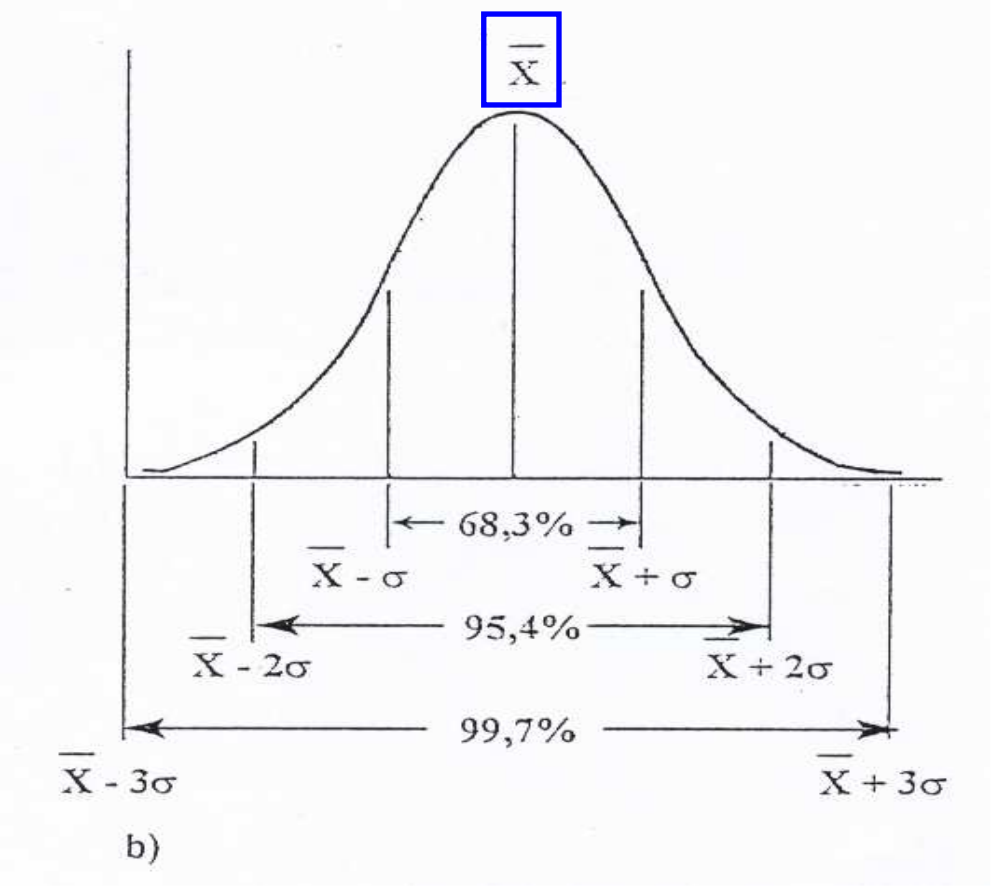
DALL'ANALISI DEL GRAFICO SI PUÒ RILEVARE CHE:

- ❶ **La curva ha forma di campana ed è simmetrica intorno ad un valore centrale $\overline{X}$; in questa curva infatti il massimo della distribuzione delle frequenze è centrale e coincide con la media (curva unimodale).**
- ❷ **La curva presenta due punti di flesso; la distanza fra ciascuno dei due punti di flesso ed il valore medio $\overline{X}$ (X barrato) si chiama deviazione standard vera ed è indicata con σ. Questa grandezza è una misura della "dispersione" dei valori intorno a $\overline{X}$ ed è pertanto un indice di precisione delle misure sperimentali.**



DALL'ANALISI DEL GRAFICO SI PUÒ RILEVARE CHE:

- ❸ **Dalle proprietà della curva di distribuzione risulta che per una distribuzione normale con un numero infinito di misure,**
- **il 68,3% delle misure cade nell'intervallo $\overline{X} \pm \sigma$;**
- **il 95,4% nell'intervallo**
- **$\overline{X} \pm 2\sigma$ (27,1%)**
- **ed infine il 99,7% entro**
- **$\overline{X} \pm 3\sigma$ (4,3%)**

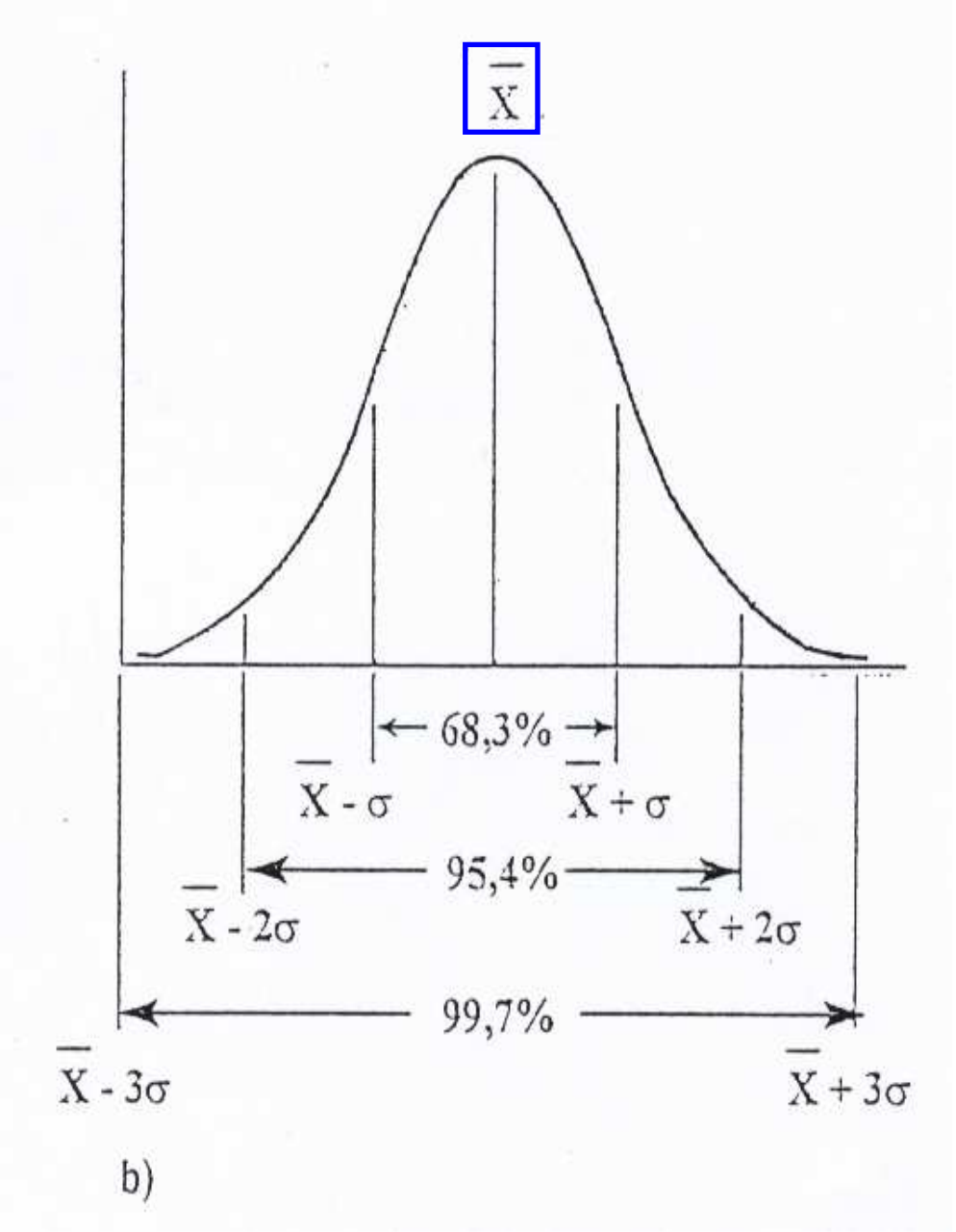


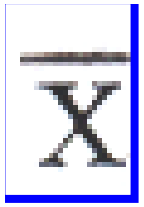
DALL'ANALISI DEL GRAFICO SI PUÒ RILEVARE ANCORA CHE:

- ❶ **Le deviazioni positive e negative (con eguale valore assoluto) rispetto ad $\overline{x}$ sono egualmente probabili.**
- ❷ **Le piccole deviazioni sono più frequenti di quelle grandi.**
- ❸ **L'esame generale della curva mette in evidenza che i singoli risultati sperimentali si addensano intorno al valore medio $\overline{x}$ che quindi è il più probabile. Per un numero infinito di valori, la probabilità di errore è nulla per il valore medio e cresce in valore assoluto allontanandosi da esso: cioè tutti i valori sono possibili, ma non egualmente probabili.**

❹ **Quanto più la curva è stretta tanto più le misure sono precise ed esatte (accurate), viceversa quando è larga.**

Inoltre più piccolo è il valore di σ più esatta e precisa è l'analisi cui esso si riferisce.





MEDIA, MODA O VALORE NORMALE

CORRISPONDE ALLA MASSIMA FREQUENZA

RAPPRESENTA IL VALORE PIÙ RICORRENTE

- ● **Si ha una distribuzione unimodale se vi è un solo valore massimo**
- ● **Bimodale se ve ne sono due non coincidenti**
- ➢ **Tutte le curve plurimodali si possono ricondurre mediante calcoli matematici ad una unimodale**
- ➢ **È chiaro che in questo caso la media non coinciderà con la moda di nessuno dei picchi di massima**

$\tilde{x}$

MEDIANA

VALORE ATTORNO AL QUALE GLI ALTRI SONO EGUALMENTE DISTRIBUITI

- **Metà sono numericamente più grandi (maggiori)**
- **Metà numericamente più piccoli (minori)**
- **Per una serie costituita da valori dispari la scelta del valore mediano è immediata (centrale)**
- **Per una serie di misure pari si prende il valore medio della coppia centrale**

Dati:10,10,12,13,13,13,15,**18**,25,26,26,27,28,28,35

- la media è **19,93** e la mediana è **18**.

CALCOLO

- **Sia dato un insieme di misure $x_1, x_2, \ldots x_N$.**
- **Media:** $$\bar{x} = \frac{\sum_i x_i}{N}$$
- **Mediana: avendo ordinato le misure in ordine crescente**
- **N pari** $$\hat{x} = \frac{x_{\left|\frac{N}{2}\right|} + x_{\left|\frac{N}{2}\right|+1}}{2}$$ **N dispari** $$\hat{x} = x_{\left|\frac{N}{2}\right|+1}$$

- **Date le misure: 1, 3, 4, 5, 7, 8**
- **la media è 4,6 e la mediana è 4,5 cioè (4+5)/2**

DATO CHE:

❶ Non potremo mai disporre di un numero infinito di valori sperimentali

❷ Non potremo mai conoscere il valore vero

❸ Di conseguenza non potremo mai conoscere l'errore vero

❹ Potremo solo avere una "stima" di esso

INOLTRE:

➲ **Affinché i dati di cui disponiamo siano chiari occorre che sia evidenziata tutta l'informazione in essi contenuta e per ottenere ciò essi vanno elaborati.**

➲ **Esiste un numero enorme di parametri statistici, nessuno di essi però contiene tutta l'informazione contenuta nei dati iniziali.**

➲ **I risultati possono essere rappresentati sotto forma di curve, istogrammi, ecc., presentazioni, queste che possono essere molto utili e suggestive e di facile comprensione, tuttavia non permettono né calcoli né confronti ulteriori.**

- **SI DICE, IN LINGUAGGIO STATISTICO, CHE QUESTE PRESENTAZIONI NON SONO "EFFICACI"**

EFFICACIA

l'efficacia di un parametro indica l'entità di informazione in esso contenuta
entità che i matematici sono capaci di valutare anche quantitativamente
(noi ci accontenteremo di alcune nozioni intuitive)

- **UN PARAMETRO STATISTICO**
- **È TANTO PIÙ EFFICACE**

- ✓**Quanto meglio riassume il contenuto informativo dei dati iniziali, con la minor perdita di informazione**
- ✓**Quanto meglio si presta ai calcoli ed ai test ulteriori**

Utilizzando rigorose
dimostrazioni matematiche
si è potuto dimostrare che,
per i tipi di problemi di cui ci occupiamo,

i parametri statistici più efficaci sono:

- ❶ **la media aritmetica** $\overline{X}$,
- ❷ **la varianza** s^2,
- ❸ **la deviazione standard** $\sqrt{s^2}$,
  - **l'insieme di** ❷ e ❸ **è quello che contiene**
  - **la maggior quantità di informazioni utili**

media aritmetica

⌦ Punti fondamentali:

La media (è di origine intuitiva)
è una "stima" del valore centrale
attorno a cui oscillano i valori trovati
Ma racchiude solo una parte dell'informazione
contenuta nei dati e non indica il grado di oscillazione dei vari risultati attorno ad essa

- ✓ **Cioè non fornisce alcuna informazione su quella che si chiama "dispersione" delle misure (informazione essenziale in quanto ci dà un'indicazione del maggiore o minore affollamento dei valori)**
- ✓ **Dal punto di vista pratico è il calcolo più semplice:**
- **è il valore numerico che si ottiene dividendo la somma dei singoli risultati per il numero totale delle determinazioni**

In una serie di analisi tutti i risultati ottenuti hanno uguale peso statistico ma il valore medio è più probabile di ogni singolo risultato

- ➔ **Per dimostrazione matematica la media di n valori, egualmente accurati e precisi (cioè egualmente probabili), è $\sqrt{n}$ volte più probabile di ogni singola misura**
- **Es.: la media di 9 risultati ha "probabilità tripla" di essere il valore esatto, rispetto ad ogni singolo valore)**
- **Se si hanno n risultati la media $\overline{X}$ è data dall'equazione:**
- $$\overline{X} = \frac{X_1 + X_2 + \ldots X_{n-1} + X_n}{n}$$
-

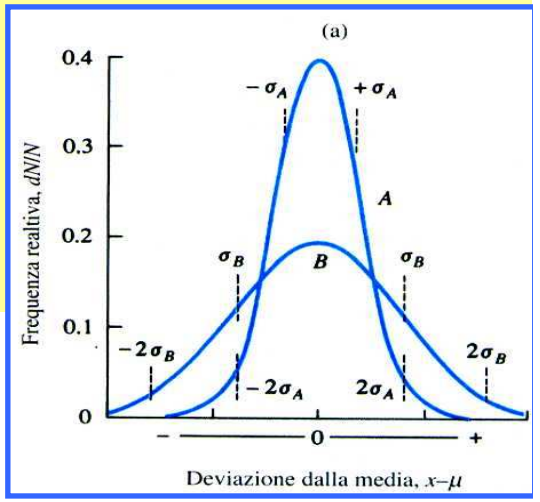
ALTRI PARAMETRI STATISTICI EFFICACI

intervallo di variazione, nozione semplice,

Definito dalle due misure estreme

(massima e minima), è intuitiva,

non necessita di calcolo



INCONVENIENTI:

1) utilizzando solo i due valori estremi, tutta l'informazione contenuta negli altri si perde;

2) i valori estremi, sono i più rari, i più influenzati dalle oscillazioni accidentali, con range tanto più ampio quanto maggiore è il numero delle misure eseguite o quanto minore è l'accuratezza;

3) non sono possibili ulteriori calcoli e test rigorosi di confronto.

"ALTRI PARAMETRI STATISTICI EFFICACI"

LO SCARTO O DEVIAZIONE:

"differenza tra ogni singolo dato iniziale X

e la **media** $\bar{X}$" $(X - \bar{X} = x)$

DEVIAZIONE DALLA MEDIA O SCARTO MEDIO

"media degli scarti presi in valore assoluto"

$(Sx/n = d)$

⌦ VANTAGGI ⌫

➔ Facile da comprendere e si calcola senza fatica

➔ Ci fornisce una certa "stima della dispersione"

➔utile per valutare se un risultato è *aberrante*

INCONVENIENTI: per motivi matematici, è inutilizzabile nei calcoli statistici di significatività

un risultato *aberrante* è un valore da scartare (non per comodità!!) perché inattendibile in quanto derivante sicuramente da un qualche errore grossolano, *aberranti* sono i dati che hanno uno scarto (x) > 4d (il quadruplo della deviazione media) essi vengono scartati e non mediati.

PER IL CALCOLO SI PROCEDE COSÌ:

❶si calcola la media degli scarti in valore assoluto (d) inserendo il dato sospetto, ❷successivamente si fa la media aritmetica dei dati sperimentali senza utilizzare tale valore ❸e si calcola la differenza tra il dato in questione e la media calcolata senza di esso, ❹se quest'ultima è maggiore di 4d il dato deve essere scartato.

- **I PARAMETRI STATISTICI PIU' EFFICACI:**
- **"VARIANZA"**
- **"DEVIAZIONE STANDARD"**
- **➔ Sono i parametri più efficaci in grado di fornire informazioni attendibili sulla dispersione attorno alla media ←**
- **➔ Sebbene derivati da considerazioni matematiche non sono affatto di origine intuitiva ←**
- **Questi due parametri sono strettamente imparentati infatti**
- **➔ la deviazione standard ($\sqrt{s^2} = s$) è la radice quadrata della varianza (s^2) ←**

"Varianza" (s^2)

- ❶ **La varianza è detta anche**
- **scarto quadratico medio**
- ❷ **definizione matematica:**
- **la varianza (s^2) si ottiene dividendo la somma dei quadrati degli scarti (Sx^2)**
- **per il numero dei "gradi di libertà"**
- ◈ **La somma dei quadrati degli scarti (Sx^2 = devianza) è uno dei valori chiave nei più diversi calcoli statistici**

"DEVIAZIONE STANDARD" ($\sqrt{s^2} = s$)

- ◈ **La deviazione standard in realtà si chiama più precisamente "stima della deviazione standard", perché la deviazione standard vera (σ) si può calcolare solo con un numero infinito di determinazioni e quindi con infiniti risultati analitici in una condizione ideale.**

Spiegazione Dettagliata del Calcolo Dei Parametri Statistici Di Una Serie Di Misure

deviazione media d (Sx/n) = 0, 0971

Calcolo Dei Parametri Statistici Di Una Serie Di Misure

Valori (X)	**Scarti (X- $\overline{x}$ = x)**	**Quadrati degli Scarti (x^2)**
3,592	3,592 - 3,4725 = + 0,1195	0,01428025
3,447	3,447 - 3,4725 = - 0,0255	0,00065025
3,570	3,570 - 3,4725 = + 0,0975	0,00950625
3,488	3,488 - 3,4725 = + 0,0155	0,00024025
3,202	3,202 - 3,4725 = - 0,2705	0,07317025
3,536	3,536 - 3,4725 = + 0,0635	0,00403225
20,835 Totale **SX**	**Sx - n $\overline{X}$ = 0**	0,10129450 **Sx^2** **(devianza)**
(SX/n) $\overline{X}$ = 3,4725		

Spiegazione Dettagliata del Calcolo Dei Parametri Statistici Di Una Serie Di Misure

deviazione media d (Sx/n) = 0, 0971

Simbolo	Come si calcola
$\bar{x}$ media	SX/n
Devianza	Sx^2
(s^2) Varianza	$Sx^2 / (n-1)$
(s) Deviazione standard (stimata)	$\sqrt{Sx^2 / (n-1)}$ $\sqrt{s^2}$
(s_m) Errore standard della media (stimato)	$s/\sqrt{n}$ (è tanto minore quanto più numerose sono le misure)
(v) Coefficiente di variazione (RSD o %)	$s / \bar{x} \cdot 100$ ($s:\bar{x} = v:100$)

❖ **Calcolare la deviazione standard e la RSD% (Cv) dei seguenti risultati.**

$X_1 = 23{,}23;$

$X_2 = 21{,}29;$

$X_3 = 20{,}66;$

$X_4 = 29{,}05;$

$X_5 = 23{,}33;$

$i := 1..5$

$\mathbf{X}_1 =$

23.23
21.29
20.66
29.05
23.33

$$\overline{X} := \frac{\sum_i x_i}{5} \qquad x_m = 23.512$$

$$s := \sqrt{\frac{\sum_i (X_1 - \overline{X})^2}{5-1}}$$

$s = 3.311$

$$RSD\% := s \cdot \frac{100}{\overline{X}}$$

$RSD\% = 14.083$

Deviazione std stimata

$\sqrt{Sx^2/(n-1)}$

(s^2) **varianza** $Sx^2/n-1$

❖ **Calcolare errore standard della media dei dati dell'esercizio precedente.**

(s_m) Errore standard della media (stimato) $s/\sqrt{n}$

$$\frac{s}{\sqrt{5}} = 1.481$$

ALCUNI CHIARIMENTI

- con X si indica il risultato sperimentale di ciascuna determinazione;
- con x si indica la deviazione dalla media (scarto), il più semplice simbolo statistico;
- tutti i valori dei parametri statistici sono "stimati" perché riferiti ad un numero limitato di valori;
- la trasformazione in quadrati non è casuale ma giustificata dall'utilità di eliminare il segno del numero ed evitare calcoli complessi con numeri relativi.

GRADI DI LIBERTÀ
La nozione di "gradi di libertà" è una nozione chiave

- **Consideriamo i 6 dati iniziali,**
- **essi sono tutti "indipendenti fra loro":**
- ● **Perché nessun valore della serie è deducibile dalla conoscenza degli altri; cioè conoscendo il primo valore nulla ci può permettere di prevedere o calcolare il secondo, e così di seguito fino all'ultimo**
- ●● **Pertanto i 6 valori sono indipendenti e si dice che il numero dei gradi di libertà nella serie di misure considerata coincide con il numero n delle misure**

- **“GRADI DI LIBERTÀ”**
- ➤ **Consideriamo ora i 6 scarti: la conoscenza del primo non determina il valore del secondo, né quello del terzo e così di seguito fino al penultimo**
- ➤➤ **A questo punto la situazione cambia essendo noti 5 scarti il rimanente può essere calcolato a partire dai 5 valori noti perché la somma algebrica degli scarti deve essere necessariamente zero**
- ➤➤➤ **Il valore numerico dell'ultimo scarto è quindi determinato e, può essere uno ed uno solo.**

Quanto detto è intuitivo, ma anche dimostrabile matematicamente

- $x = X - \bar{X}$ per definizione,
- perciò: $Sx = SX - n\bar{X}$,
- da cui $\bar{X} = Sx + SX/n$

- Dato che $\bar{X} = SX/n$ se ne deduce che Sx deve essere necessariamente uguale a zero
- ➡ Perciò solo 5 dei 6 valori (n-1) di scarto sono indipendenti e non deducibili in nessun modo:
- Essi rappresentano i gradi di libertà

Da quanto detto si evince che i GRADI DI LIBERTÀ
⬇⬇ rappresentano ⬇⬇

- ✷ **Un dato importante in tutte le applicazioni statistiche perchè esprime, il numero di dati effettivamente disponibili per valutare e includere tutte le informazioni nel parametro considerato**
- ✷ **Il semplice buon senso indica che la quantità di informazione contenuta nei dati dipende dal loro numero; ma si capisce che conta il numero dei dati indipendenti, non il numero totale dei dati**
- ✷ **perché quando un dato non è indipendente l'informazione che esso fornisce è già contenuta implicitamente negli altri dati**
- ✷ **perciò esso non apporta nulla di nuovo dal punto di vista dell'informazione e, pertanto, non deve entrare nel conto**

Intervallo fiduciale o Livello di probabilità

- ① È quello entro il quale si vuole trovare, con una data probabilità, il valore “vero”
- ② Non viene utilizzato nei test statistici rigorosi, tuttavia è utile perché fornisce immediatamente al ricercatore le indicazioni più opportune per quanto concerne la precisione delle misure
- ③ I limiti di tale intervallo variano a seconda della probabilità con cui si vuole che il valore "vero" cada nell'intervallo stesso

  ④ L'intervallo deve essere il meno possibile esteso, perché se è troppo ampio è difficile formulare conclusioni di validità generale sulla base dei risultati sperimentali ottenuti

Gli estremi di questo intervallo sono calcolati rispettivamente mediante le formule:

$\bar{X} - t \; s/\sqrt{n}$ = limite inferiore

$\bar{X} + t \cdot s/\sqrt{n}$ = limite superiore

- $t = m-\mu / s_m$
- è la variabile di Student

- i suoi valori dipendono:
- dal numero di gradi di libertà e
- dalla probabilità con cui si vuole trovare il risultato "vero" in quell'intervallo

$t = m-\mu / s_m$

m è la media stimata

(calcolata per ciascuna serie di gruppi limitati di risultati estrapolati a caso dalla grande serie)

μ è la media vera

(calcolata a partire da numero **strabiliante** di valori sperimentali)

$s_m = s/\sqrt{n}$ è l'errore stimato (**s** dev. std stimata)

(rappresenta la **stima dell'errore standard** "vero" perché gli scarti non vengono calcolati rispetto alla media vera, ma a quella di varie serie di gruppi limitati)

- ① Pertanto egli, utilizzando un numero elevatissimo di valori potè calcolare i parametri statistici (μ media e σ
- deviazione std) e considerarli veri
- ② In seguito raggruppò in maniera del tutto casuale i numerosi valori che aveva a disposizione, componendo diverse e numerose serie di quegli stessi valori, come se essi fossero stati ottenuti effettivamente da indagini distinte con un numero di campioni diverso per le diverse indagini caratterizzate così anche da un numero di gradi di libertà variabili

A questo punto denominò t la seguente quantità:

$$t = m-\mu / s_m$$

- ③ **Effettuando questi calcoli per un numero incredibilmente elevato di volte ottenne delle serie di valori di t che da un'esperienza all'altra oscillano da una parte all'altra dello zero, distribuendosi simmetricamente in valori positivi e negativi ed anche l'errore standard stimato oscillerà attorno al valore zero**

➔ **Student effettuò un numero esorbitante di calcoli e pubblicò delle tavole indicanti non una sola distribuzione di probabilità di t ma una famiglia di distribuzioni di t dal momento che la distribuzione si modifica secondo il numero dei gradi di libertà utilizzati nel calcolo**

Tab. 3 Valori di t secondo il numero di determinazioni e il livello di probabilità

N° Determ.	Gradi Libertà	50 %	60 %	70 %	80 %	90 %	95 %	99 %
3	2	0,816	1,061	1,386	1,886	2,920	**4,303**	9,925
6	5	0,727	0,920	1,156	1,476	2,015	**2,571**	4,032
15	14	0,692	0,868	1,076	1,345	1,761	2,145	2,977
20	19	0,688	0,861	1,066	1,328	I,729	**2,093**	2,861
30	29	0,683	0,854	1,055	1,311	1,699	**2,045**	2,756
60	59	0,679	0,848	1,046	1,296	1,671	**2,000**	2,660
121	120	0,677	0,845	1,041	1,289	1,658	**1,980**	2,617
∞	∞	0,674	0,842	1,036	1,282	1,645	**1,960**	2,576

$\bar{X}$ - 2,145 • 1,481 (s/√n) = limite inferiore

$\bar{X}$ + 2,145•• 1,481 (s/√n) = limite superiore

- L'uso di t ci permetterà di ottenere informazioni partendo da un campione i cui parametri veri non ci sono noti in quanto ci permetterà di valutare quanto la media "stimata" ottenuta a partire da un campione si avvicini alla media "vera".
- ➢ Partendo da un numero infinito di misure, si possono conoscere i parametri statistici veri, ma, ovviamente, nessuno nella realtà si potrà mai trovare in questa situazione ideale, piuttosto potrà partire da un numero limitato, di valori sperimentali ed elaborare i parametri statistici stimati

Una notazione frequentemente utilizzata consiste nell'indicare la media $\bar{X}$ aritmetica ed un determinato parametro statistico separati dal segno ±

- **il parametro che segue il segno ± può essere l'errore standard stimato ($s/\sqrt{n}$ o s_m), oppure la deviazione standard stimata, oppure lo scarto medio, perciò è assolutamente indispensabile, quando si usi questa notazione, indicare quale sia quello utilizzato**

➡Il segno ± che accompagna una media
➡non è mai seguito dal cosiddetto
➡intervallo fiduciale o di confidenza della media

RANDOMIZZAZIONE

- **Non vi sono dubbi invece sul parametro che segue**  **±**
- **quando si vogliono confrontare fra loro numerose serie di misure,**
- **ciascuna delle quali prevede un numero uguale di campioni per ogni esperimento,**
- **quando cioè il piano di lavoro necessita di replicati con tecnica di "randomizzazione"**
- **ossia quando il materiale sia stato ripartito rigorosamente a "caso"**
- **(quale che sia il piano sperimentale adottato)**

RANDOMIZZAZIONE

Il materiale (es. principio attivo per lo studio dei residui di pesticidi) viene ripartito fra i gruppi sperimentali in maniera che ciascun trattamento non venga sistematicamente influenzato, in una serie di replicazioni, da fattori di variazione estranei noti o ignoti, nel caso contrario i risultati sono "viziati" e l'analisi statistica non ha più alcun senso

RANDOMIZZAZIONE

- **Nel caso suddetto la media aritmetica seguita dal $\bar{x}$ segno ± è accompagnata da un valore che rappresenta sempre un parametro detto coefficiente di variazione**
  - (RSD) **Cv o semplicemente v (s/ $\bar{x}$ ·100)**
    - **!! Questo parametro statistico è un numero puro, indipendente**
      - **dalla variabile studiata e**
      - **dall'unità di misura utilizzata**
      - **!! Si può quindi usare nelle determinazioni di campioni non solo quantitativamente ma, anche qualitativamente diversi**

Deviazione standard relativa, RSD

$$RSD = \frac{s}{\bar{x}}$$

Deviazione standard relativa o percentuale, DSR o %, coefficiente di variazione, o Cv o v

$$RSD\% = CV = \frac{s}{\bar{x}} \cdot 100$$

- **Il coefficiente di variazione v (s/ $\bar{x}$ ·100)**
  Indica il grado di approssimazione permesso dal metodo adottato ed
- **è un valore relativo che permette il confronto anche tra serie di trattamenti ottenuti con materiali diversi (vari pesticidi) su diversi substrati**

CONCLUSIONI

- **NELL'ESPRIMERE I RISULTATI DI UNA SERIE DI MISURE**
  - **☹ ecco ciò che non si deve fare ☹**
- **☹ Esprimerli indicando solo la media**
- **☹ Esprimerli fornendo la media e i valori estremi,**
- **☹ Oppure la media e la media degli scarti.**

☺ Ciò che si deve fare è ☺

☺ esprimere i risultati indicando ☺

1° il numero delle misure eseguite;

2° la media;

3° uno dei seguenti parametri: varianza, deviazione standard, errore standard.

E' opportuno a questo punto ricordare che

- **Tutti i parametri statistici si riferiscono all'influenza degli errori casuali sulla determinazione sperimentale, ma non tengono in considerazione gli errori determinati o sistematici.**
- **Questi ultimi vengono messi in evidenza eseguendo una stessa determinazione con due o più metodi differenti di analisi e confrontandone i risultati, come più volte detto ed eliminati con la precisione e l'esattezza.**

CIFRE SIGNIFICATIVE

sono quelle necessarie ad esprimere il risultato di una misura con il grado di accuratezza con cui è stata eseguita

- ❶ **Ad esempio, se si esegue una pesata in una bilancia analitica precisa al decimilligrammo, il valore numerico in grammi va espresso con 6 cifre significative e non con un numero di cifre superiore o inferiore perché nel primo caso si attribuirebbe alla misura un grado di precisione troppo elevato, nel secondo si esprimerebbe la misura in modo meno preciso di quanto è stato fatto.**

CIFRE SIGNIFICATIVE

- ❷ **Analogamente, nel riportare qualunque risultato sperimentale è bene indicare tutte le cifre significative che si conoscono in modo che soltanto l'ultima possa essere dubbia.**
- ❸ **Se essa è seguita da una cifra superiore a 5 ne viene fatto l'arrotondamento per eccesso, se è inferiore per difetto. Inoltre se l'ultima cifra da scartare è 5 seguita da zero si arrotonda in eccesso se la rimanente è dispari, in difetto se è pari; se il 5 è seguito da numeri diversi da zero si arrotonda sempre in eccesso.**

CIFRE SIGNIFICATIVE

- ❹ **Nei calcoli che si eseguono per conoscere il risultato finale bisogna tenere presente che esso dovrà essere espresso con il numero di cifre significative relative al dato meno accurato.**
- ❺ **Come regola generale i pesi vanno espressi con 6 cifre significative; i volumi con 4, le percentuali con 4.**

INOLTRE

- **L'espressione di una misura con il corretto numero di cifre significative (esattamente conosciute più quella approssimata) non solo è più corretta, ma semplifica notevolmente i calcoli; è comodo, infatti poter arrotondare i risultati al giusto numero di cifre senza fare prodotti o rapporti con molte cifre che non hanno una corrispondenza con l'accuratezza delle misure.**

INOLTRE

- **Non bisogna dimenticare che la significatività è diversa per gli zeri:**
- **essi sono significativi quando fanno parte di un numero, o ne indicano la grandezza, non lo sono quando precedono le cifre significative di un numero con la virgola della quale individuano la posizione. Quando un valore deve essere espresso con un certo numero di cifre, ma se ne tralascia qualcuna, si commette un errore, perché automaticamente si attribuisce ad esse il valore zero.**

Deviazione assoluta e relativa.

- Detta impropriamente errore assoluto e relativo,
- il primo rappresenta
- ↳ la differenza fra un risultato sperimentale incognito e il valore "vero" stimato solamente col calcolo della media $X_i - \overline{X}$ = errore assoluto
- Il secondo rappresenta
- ↳ Il rapporto fra l'errore assoluto e la grandezza effettiva moltiplicato per 100 rappresenta l'errore relativo espresso in percentuale.
- errore assoluto/valore stimato •100=errore relativo
- In laboratorio è tollerato fino a ± 2%

IMPORTANZA DELLA RETTA DI REGRESSIONE

- **Fino ad ora abbiamo analizzato esempi che riguardavano ogni volta una sola serie di valori**
- **Nell' eseguire una determinazione analitica sperimentale, può essere necessario effettuare le analisi con un criterio diverso, cioè operare su campioni differenti per entità, ma uguali per qualità**
- ◉ **Questo capita tutte le volte che si vuole indagare sulla relazione tra:**
- **la solubilità di un precipitato e la temperatura,**
- **o il pH, o il contenuto di elettroliti, o la concentrazione, o più in generale tra un qualunque parametro e la risposta strumentale.**

◉ **Questo studio viene solitamente compiuto tracciando un grafico che porta**
in ascissa la variabile indipendente
(concentrazione)
in ordinata quella dipendente
(risposta strumentale).

◉ ◉ ◉ **Si può quindi utilizzare il diagramma così ottenuto per ricavare il valore della concentrazione del campione incognito da quello del segnale ad esso relativo.**

◉ ◉ ◉ ◉ **Scegliendo opportunamente le variabili, o calcolandone i logaritmi, o in genere con artifici matematici, è quasi sempre possibile fare sì che i risultati sperimentali vengano rappresentati da punti più o meno allineati sul grafico.**

Anche in questi casi, ovviamente, si commettono sempre errori sistematici ed errori casuali

- **I primi possono essere svelati ed eliminati**
- **proprio mediante la costruzione di curve o grafici di taratura,**
- **utilizzando concentrazioni note di sostanza pura di riferimento (standard analitici primari o secondari),**
- **cosicché la quantità incognita verrà determinata utilizzando il grafico dello standard nei modi in cui viene insegnato dalla elaborazione statistica.**

Una sostanza per essere impiegata come standard primario deve soddisfare i seguenti requisiti:

a) deve essere più pura possibile (impurezza massima 0,01%);

b) non deve essere alterabile all'aria (soprattutto nei confronti di O_2 e CO_2)

c) non deve reagire con il solvente in cui è solubilizzata;

d) non deve essere né igroscopica, né deliquescente, né efflorescente;

e) deve essere facilmente reperibile in commercio;

f) deve essere possibilmente poco costosa;

g) deve reagire in maniera stechiometricamente univoca ;

h) deve avere un peso equivalente sufficientemente alto da minimizzare gli errori di pesata;

i) deve essere stabile: non deve volatilizzare (tensione di vapore troppo alta), né sublimare.

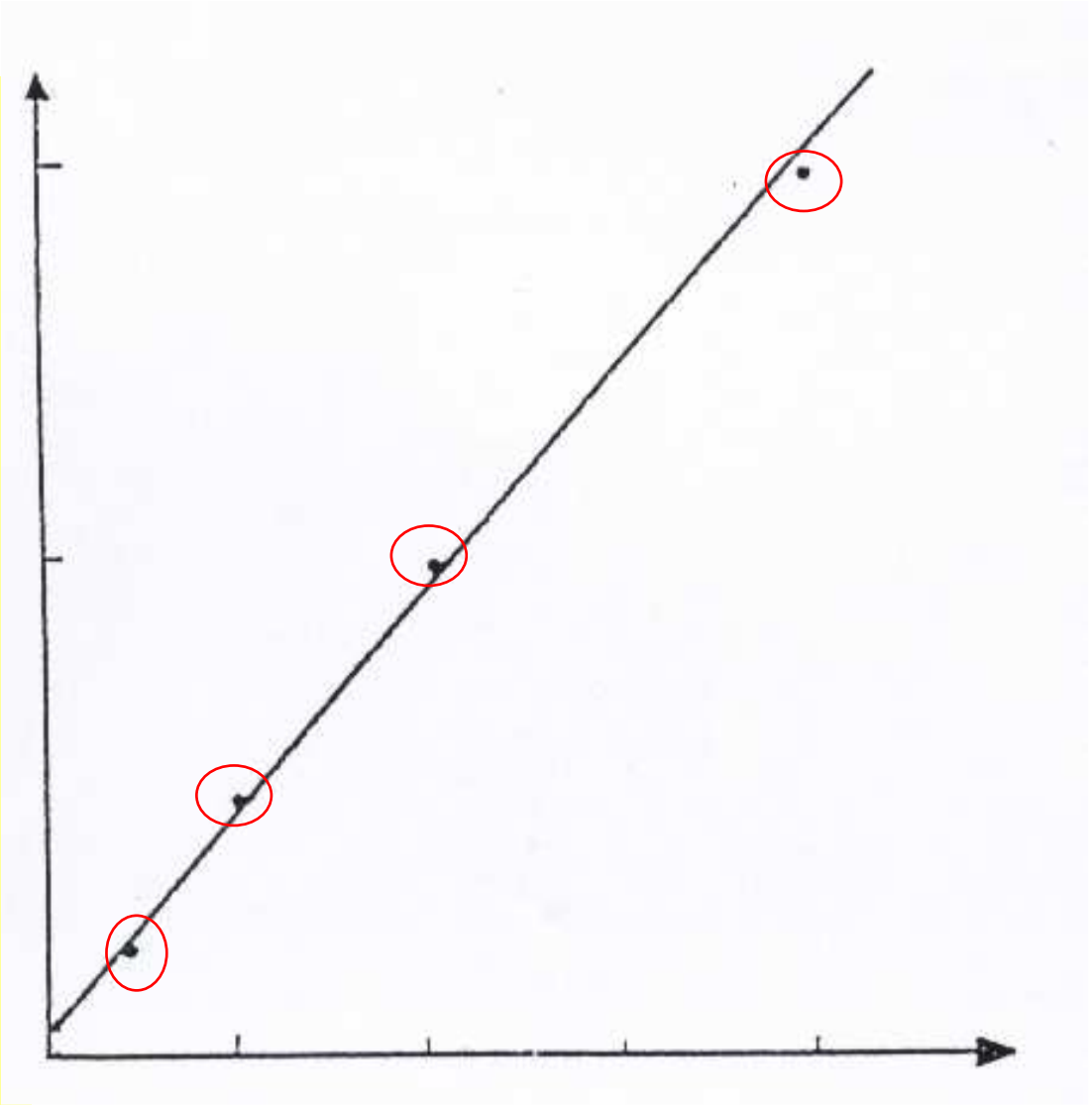
Una sostanza per essere impiegata come standard primario deve soddisfare i seguenti requisiti:

d) non deve essere né igroscopica, né deliquescente, né efflorescente:

- **Igroscopica: assorbe acqua;**
- **Deliquescente: si scioglie perché l'umidità relativa dell'aria raggiunge la tensione di vapore della sua soluzione satura;**
- **Efflorescente perde acqua di cristallizzazione ed i cristalli si sfaldano, quando il valore dell'umidità relativa dell'aria diminuisce andando sotto la tensione di vapore della sua soluzione satura.**

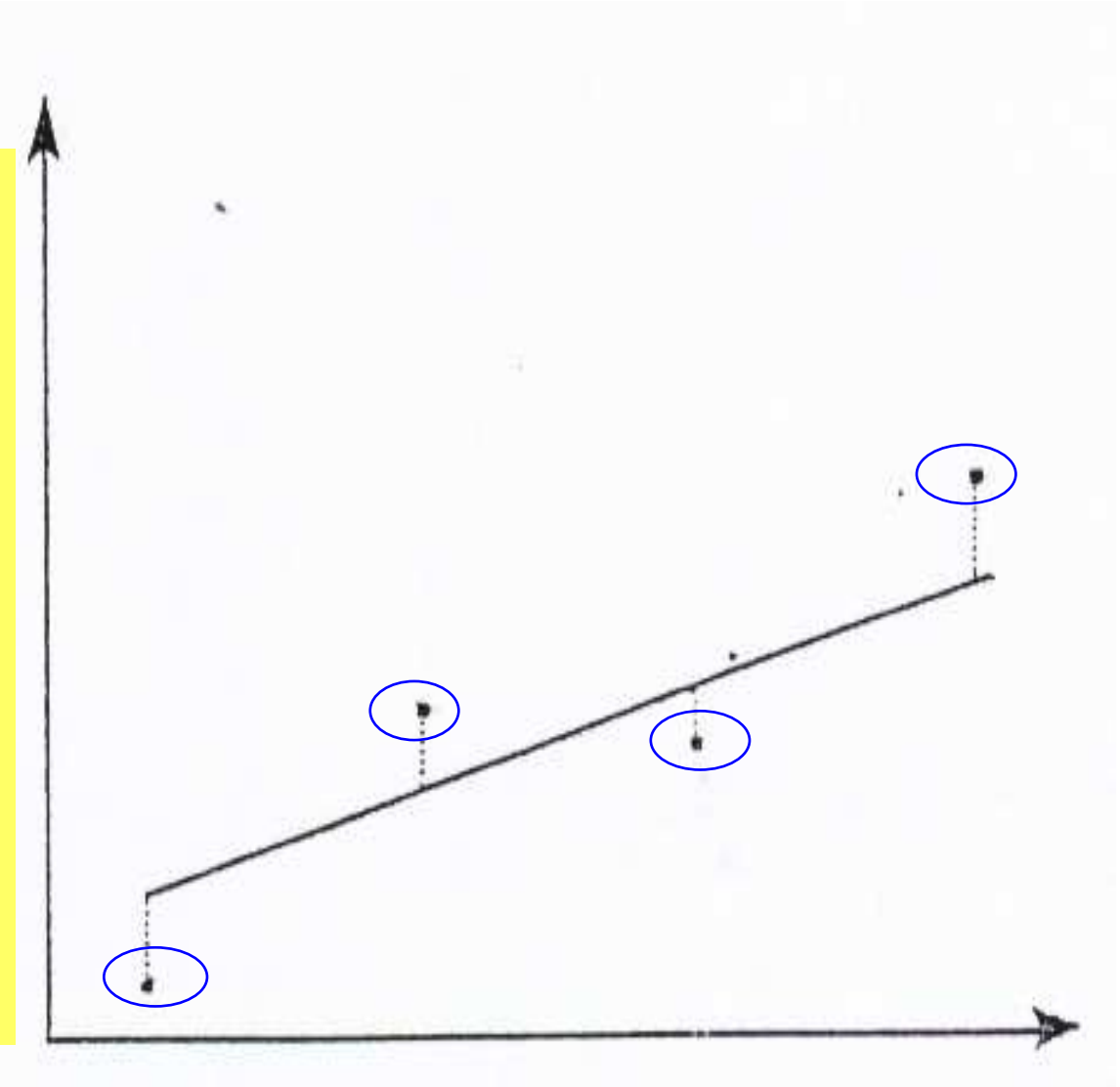
Anche gli errori casuali sono sempre presenti, come dimostra il fatto che quasi sempre i grafici sono costruiti da punti sperimentali non allineati

anche in assenza di errori sistematici, diversamente da come richiederebbe la dipendenza teorica tra la quantità indagata (conc.) e la risposta strumentale, perciò



bisogna minimizzarli cercando di tracciare la retta migliore possibile che comprenda i punti sperimentali.

La retta che unisce questi punti poteva una volta essere tracciata ad occhio e rappresentava una relazione soggettiva, oggi si usano i metodi indicati dalle tecniche dell'analisi statistica dei dati.



Fra questi IL METODO PIÙ SEMPLICE
è quello della
REGRESSIONE LINEARE detta anche
DEI MINIMI QUADRATI

- ❖ **Giacché, dunque, non sempre tutti i punti di un grafico giacciono esattamente su una retta, per consentire di stimare al meglio i valori di Y_i per determinati valori di X (secondo una generica legge $Y = f \cdot X$), si ricorre all'interpolazione lineare della retta che viene effettuata col metodo suddetto che permette di calcolare l'equazione di quella retta che**

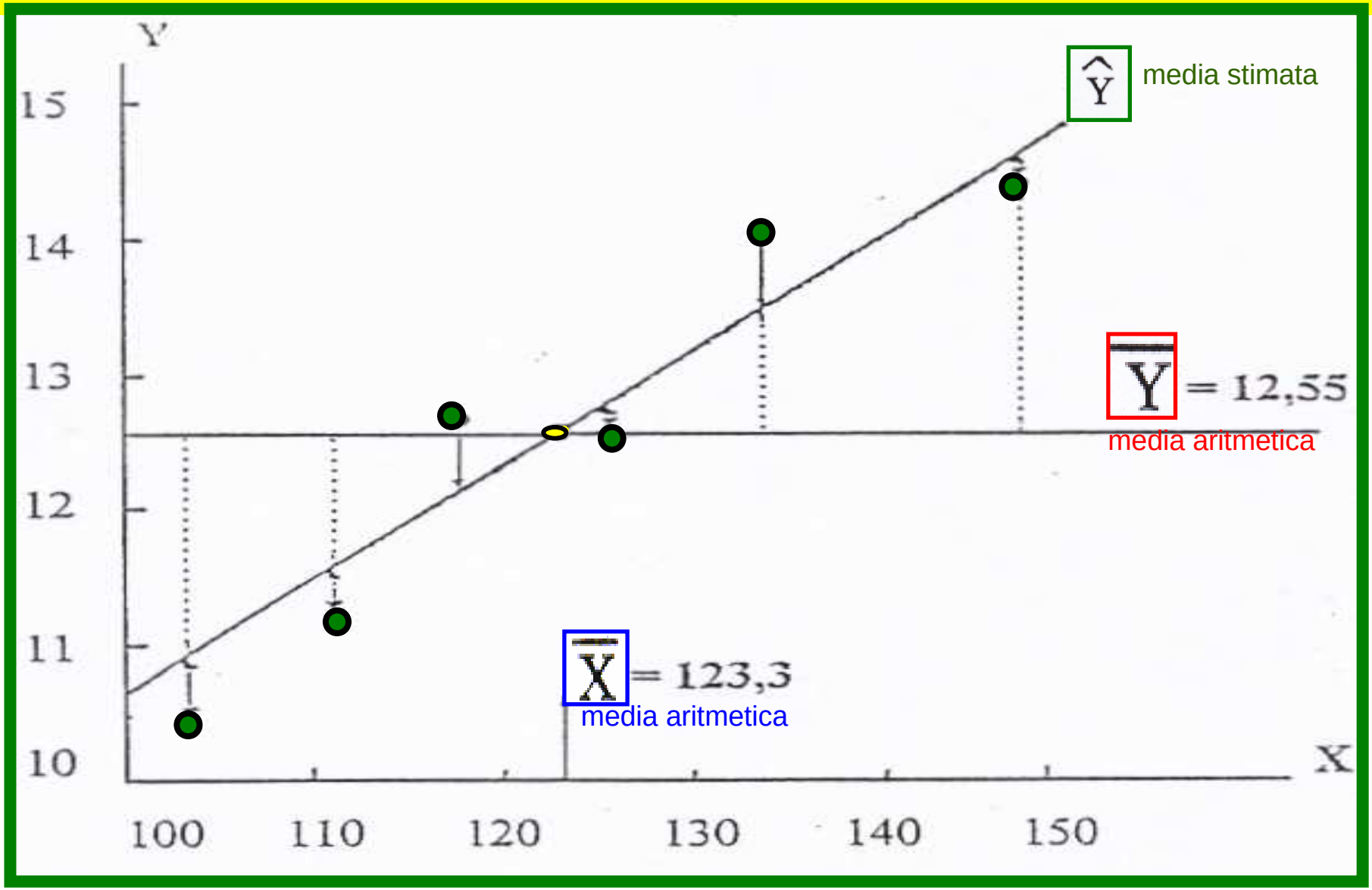
  - **minimizza la somma dei quadrati delle distanze verticali tra i punti e la retta stessa**

❖ In questa eventualità, facendo l'ipotesi che gli errori siano distribuiti secondo la solita Gaussiana, applicando i metodi statistici si possono stabilire i limiti di attendibilità e compiere altre interessanti operazioni. Infatti, quando noi misuriamo due variabili entrambi gli insiemi di misure tendono ad essere distribuiti normalmente e i dati possono essere descritti come normali-bivariati, tuttavia non studieremo tutti gli aspetti dell'analisi delle rette di regressione, che sono numerosi ma solamente quelli che sono utili alla comprensione del nostro problema.

❖ **Anticipiamo a questo punto che i valori della variabile indipendente X (riportati in ascissa e riguardanti lo standard di riferimento) sono considerati "esatti" perciò tutti gli errori casuali ricadranno sui valori della variabile dipendente Y (riportati in ordinata).**

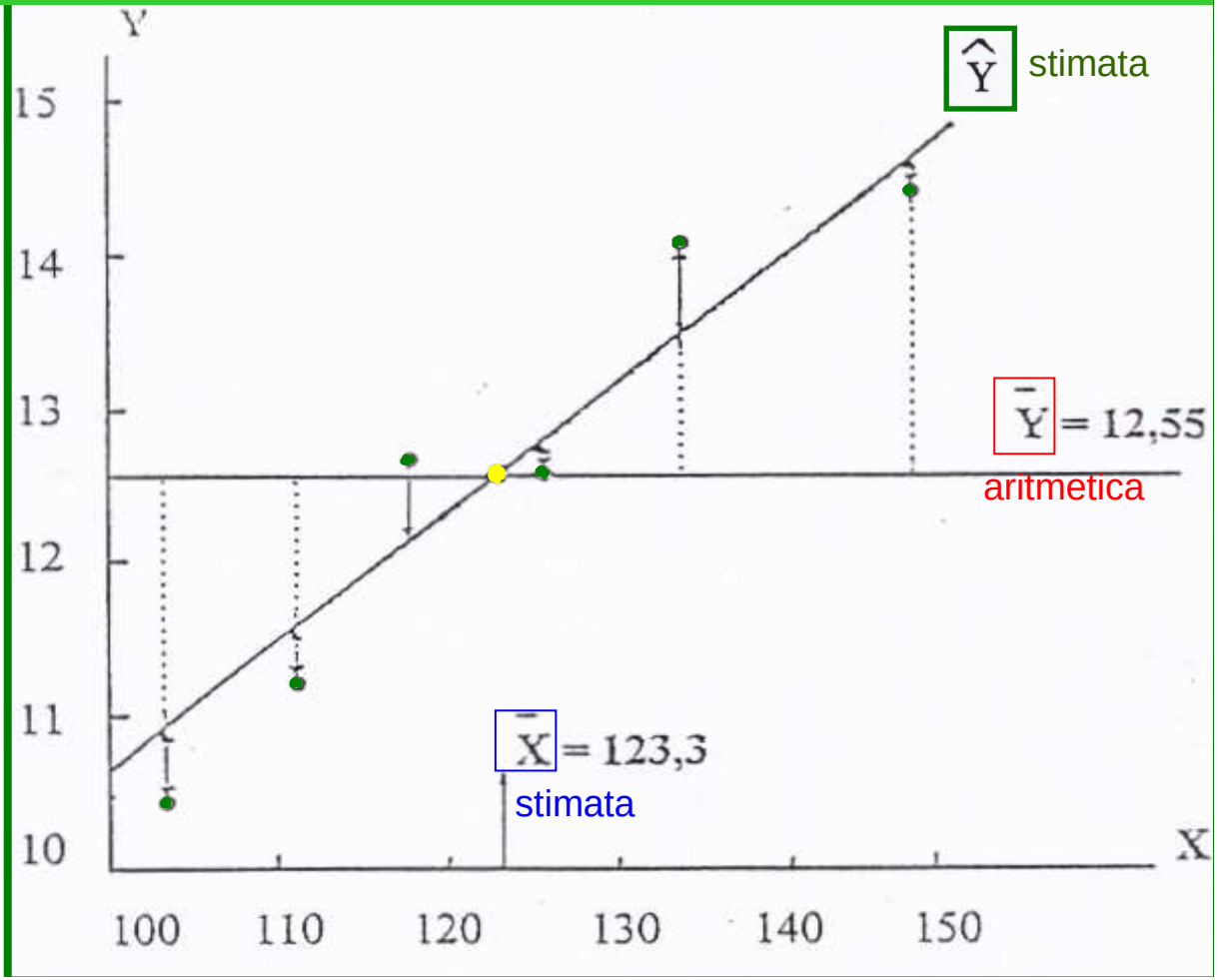
- ❖ **Da essi possiamo calcolare i parametri statistici: media $\bar{Y}$ e la somma dei quadrati (Sy^2 abbreviazione di somma dei quadrati degli scarti della media) a partire dalla quale si calcola la varianza congiunta co-varianza e la deviazione standard stimata.**

Riportiamo sul seguente grafico i valori relativi a Y e a X.



La media delle $\overline{Y}$ è stata rappresentata con una linea orizzontale, **gli scarti** dei valori individuali di Y sulla media $\overline{Y}$ sono rappresentati dalla **distanza di ciascun punto sperimentale** da questa linea orizzontale e indicati con **linee verticali**

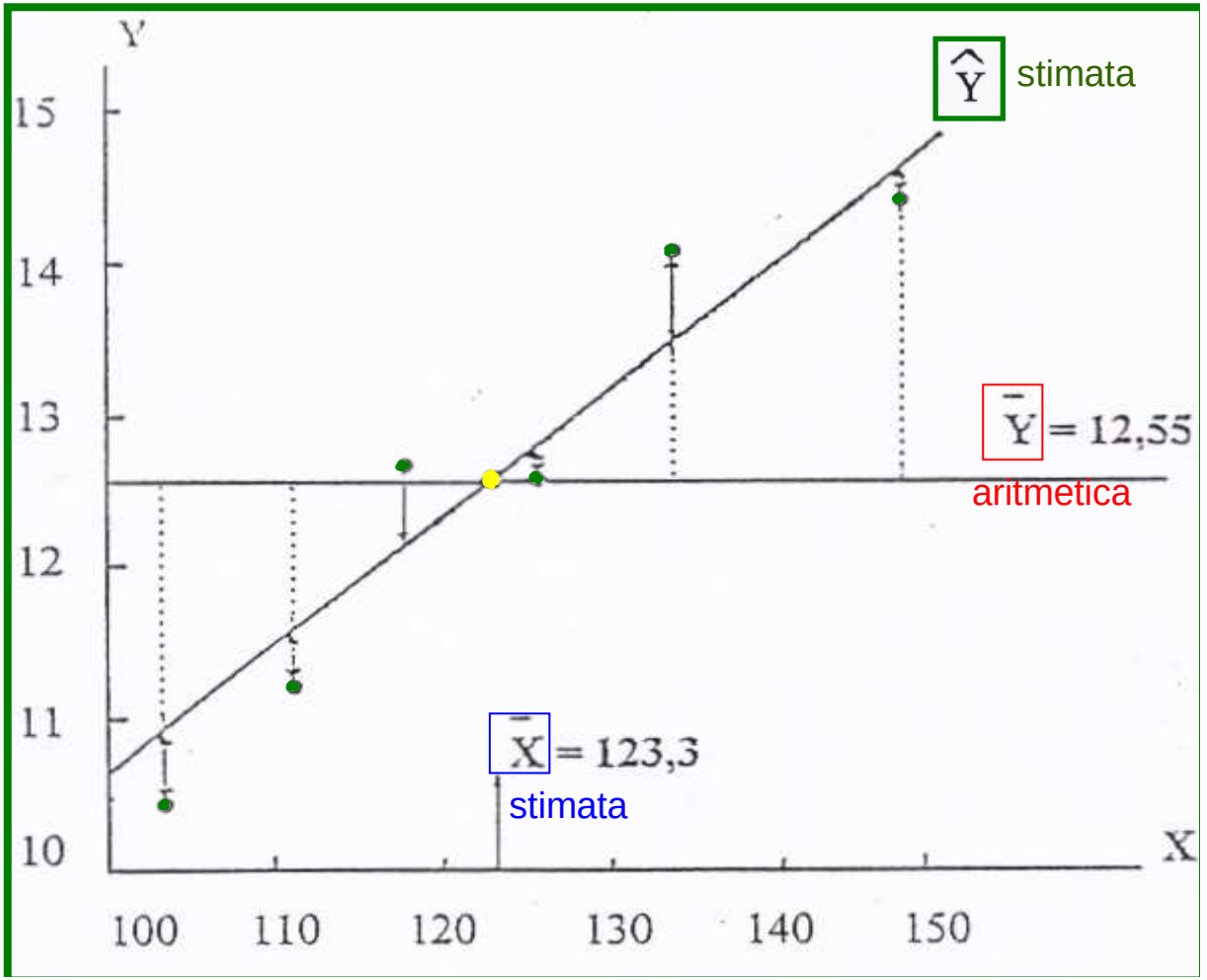
La somma algebrica di queste distanze (scarti), come già sappiamo, **è nulla**, la **"somma dei loro quadrati"** (devianza) è Sy^2



La linea orizzontale $\overline{Y}$ non è però la retta per la quale la somma dei quadrati delle distanze sia minima. La linea che gode di queste proprietà deve essere calcolata, e si chiama linea della regressione lineare delle Y sulle X (nella figura è rappresentata da $\hat{Y}$ con l'accento circonflesso che mostra:

🗣 🗣

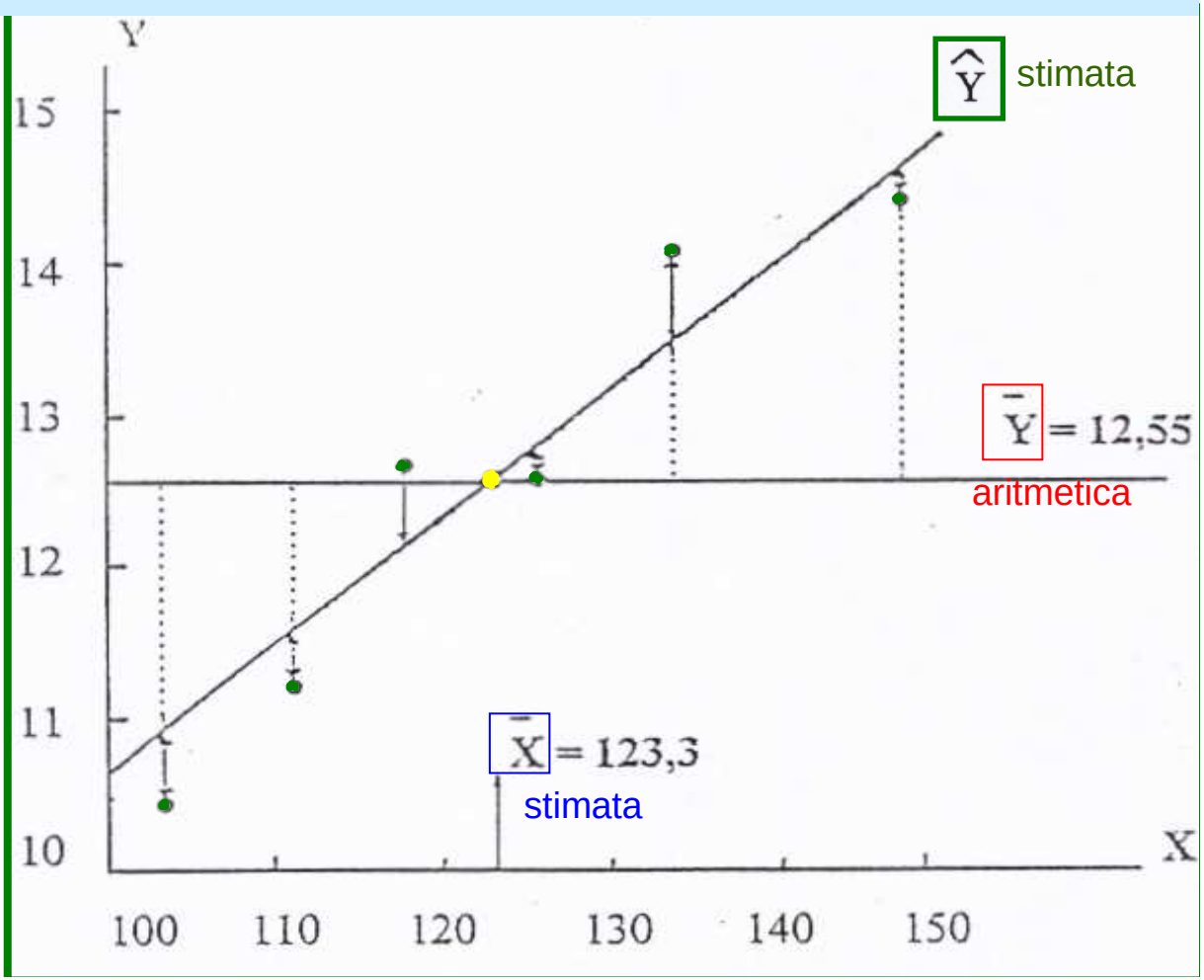
il valore di $\hat{Y}$ è "stimato" come valore medio e non corrisponde al valore medio sperimentale $\overline{Y}$ rispetto a $\overline{X}$, ma è ricavato dal calcolo statistico e stimato, cioè previsto corrispondente al valore "stimato" $\overline{X}$



La **somma degli scarti**, rappresentati dalle distanze verticali continue, è inferiore calcolata su $\hat{Y}$ rispetto a quella calcolata su $\bar{Y}$ perciò è minore anche **la somma dei quadrati degli scarti** Sy^2, che è il tramite per calcolare **varianza** e **deviazione standard "vera"**

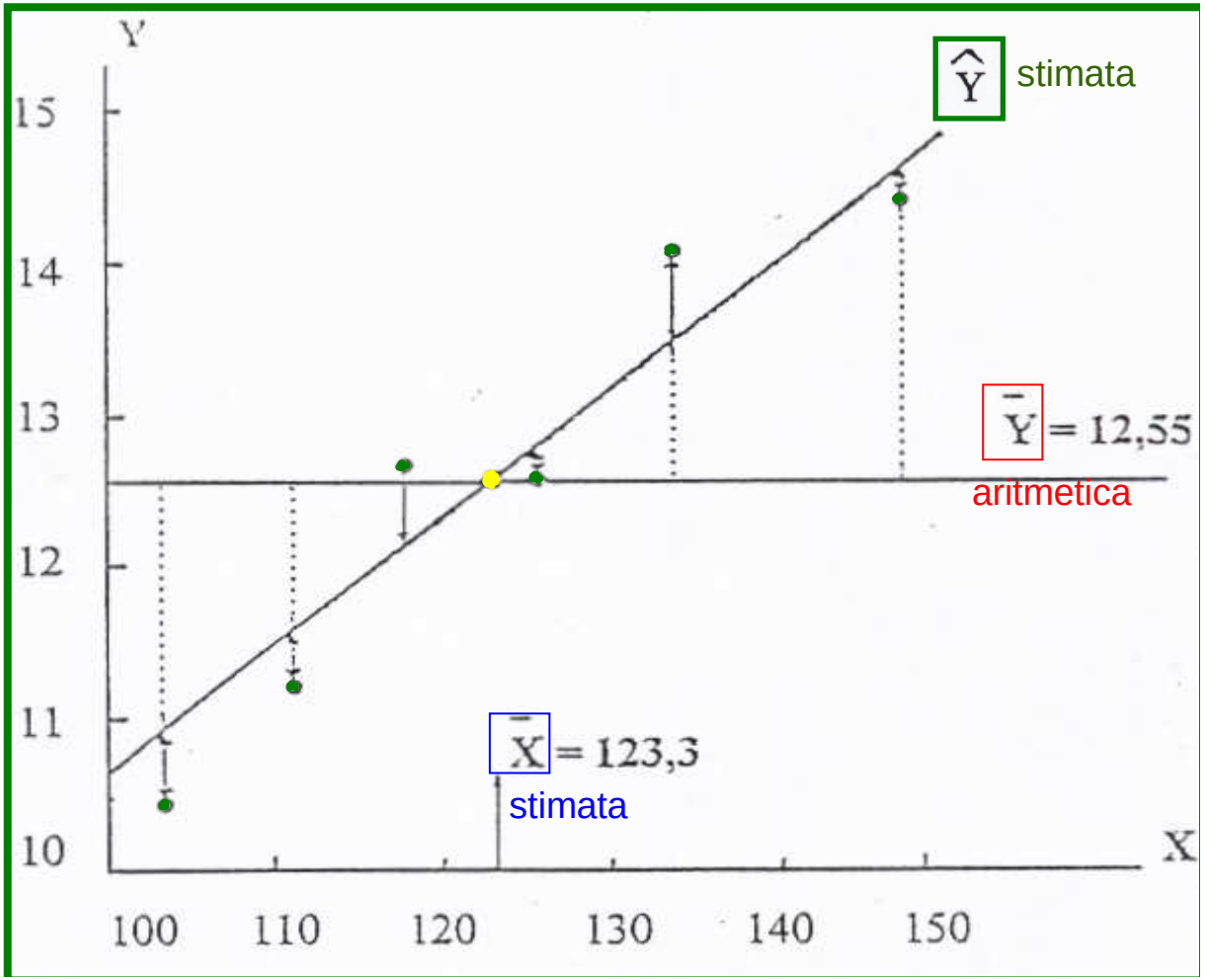
🗣🗣

per questo la retta $\hat{Y}$ si chiama dei "minimi quadrati"



Se i valori relativi alla **X** e quelli relativi alle **Y** fossero stati **esattamente proporzionali**, i punti sperimentali sarebbero stati situati **esattamente sulla retta di regressione**, **non ci sarebbero stati scarti** da essa e la **"somma degli scarti dalla retta di regressione"** sarebbe stata nulla. Ma questa non è una novità (anche la retta orizzontale $\overline{Y}$ gode di questa proprietà e molte altre rette sono possibili), **ma la retta di regressione è l'unica fra tutte a possedere il requisito di essere quella per la quale la somma dei quadrati degli scarti**

Sy^2 è minima !



- **La “somma dei quadrati degli scarti (Sy^2) dalla linea di regressione delle Y sulle X” (purtroppo non esiste un'espressione più abbreviata) rappresenta in rapporto alla linea di regressione $\hat{Y}$ quello che la "somma dei quadrati" rappresenta in rapporto alla media aritmetica.**
- **Cioè:**
- **❖ Come la media è una "stima" della tendenza centrale e la "somma dei quadrati" serve a valutare la "dispersione" in rapporto a questo valore centrale**
- **(come abbiamo visto in precedenza),**
- **❖ Così la "somma dei quadrati degli scarti dalla linea di regressione" serve a stimare la dispersione attorno al valore centrale di cui la linea di regressione fornisce una "stima"**

BISOGNA PERÒ ANCORA PRECISARE CHE:

LINEA DI REGRESSIONE +

SOMMA DEI QUADRATI DEGLI SCARTI DA ESSA

(insieme)

- **forniscono delle informazioni più estese e**
- **più precise (quindi più efficaci) di quelle fornite dall'insieme di media + somma dei quadrati.**

Infatti sono più estese, cioè più ampie, perché ci informano anche sull'influenza di una sorgente di variazione definita, rappresentata dalle X, su quella indefinita Y, mettendo in relazione i valori di X (indipendenti) con quelli di Y (dipendenti).

BISOGNA PERÒ ANCORA PRECISARE CHE:

LINEA DI REGRESSIONE +

SOMMA DEI QUADRATI DEGLI SCARTI DA ESSA

- **sono più precise perché gli scarti in rapporto alla linea di regressione sono più piccoli che non gli scarti in rapporto alla media aritmetica, da ciò deriva anche una minore deviazione standard e di conseguenza una maggiore attendibilità dei risultati.**

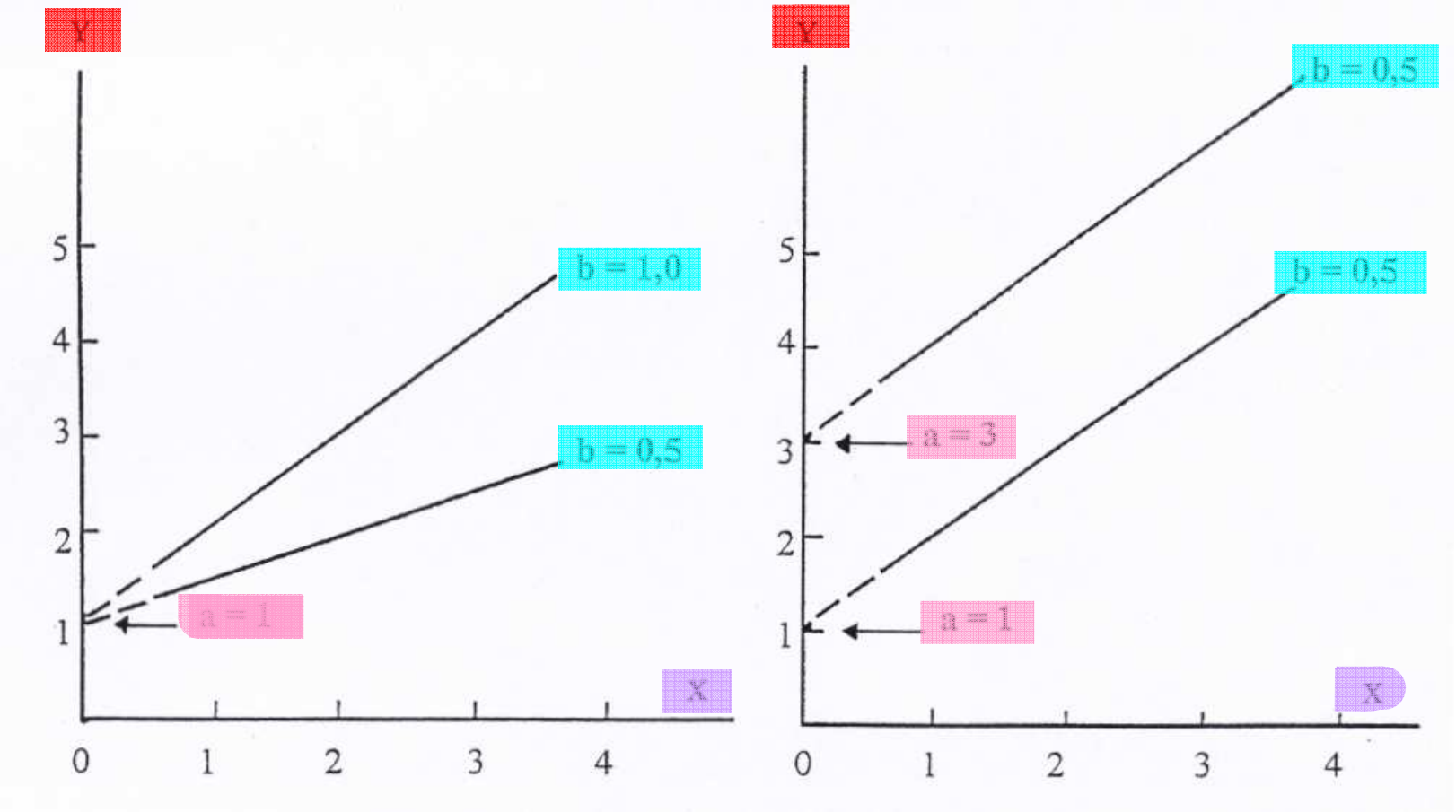
•la media aritmetica è definita da un solo valore, la retta di regressione è necessariamente definita dai due valori di un'equazione di primo grado

•(informazioni più estese e più precise)

CALCOLO DELL'EQUAZIONE DELLA RETTA DI REGRESSIONE

- → **Una retta è rappresentata dall'equazione di primo grado: $Y = a + b X$, dove a e b sono termini costanti e devono essere noti affinché la retta di regressione in questione sia definita**
- → **a) viene chiamata intercetta (intersezione della retta $\hat{Y}$ stimata con l'asse delle ordinate) o termine costante dell'equazione di regressione e definisce la posizione in altezza della retta di regressione in rapporto all'asse delle ascisse.**
- → **b) è chiamata coefficiente angolare o coefficiente di regressione e rappresenta la pendenza della linea di regressione (il suo significato si comprende meglio se si ricorda che b è la quantità di cui varia Y quando X varia di una unità)**

Due rette di regressione che differiscono per il valore di b hanno diversa inclinazione ma stessa altezza. Due rette di regressione che differiscono solamente per il valore di a hanno la stessa inclinazione, ma sono situate a due altezze differenti, sono dunque parallele fra loro.
(ovviamente due rette con valori uguali di a e b coincidono).



•**Per poter procedere all'applicazione della regressione lineare dobbiamo fare delle ipotesi che definiscono il modello matematico e che**

•**devono essere soddisfatte prima che essa venga applicata:**

①**- che la variabile indipendente X (valori in ascisse) sia misurata senza errore;**

②**- che per ciascun valore di X vi sia un solo corrispondente valore "vero" di Y, tale da essere approssimato mediante una relazione lineare tra X e Y;**

③**- che le misure di Y mostrino una variazione "random" e siano distribuite normalmente (dispersione) attorno alla media "vera";**

④**- che tutto l'errore sia nella direzione dell'asse delle ordinate Y;**

⑤**- che la variazione dei valori di Y attorno alla loro media "vera" $\hat{Y}$ sia la stessa per tutti i valori di X.**

I valori di a e di b devono essere "stimati" dai dati del campione, partendo dagli scarti e mediante l'equazione di regressione

$$\bar{Y} = a + b\bar{X}$$

- **Il coefficiente di regressione (b)**
- **è "stimato" dall'espressione:**

$$b = \Sigma xy / \Sigma x^2$$

Sommatoria di $(X - \bar{X})(Y - \bar{Y})$ **/ Sommatoria di** $(X - \bar{X})^2$

Di poi l'intercetta (a) è "stimata" sottraendo il valore calcolato di b nell'equazione:

$$a = \bar{Y} - b\bar{X}$$

Per tutti i calcoli concernenti la regressione è necessario e sufficiente calcolare preliminarmente tre valori di base↴↴↴

↳la somma dei quadrati per le X (Sx^2 o Σx^2)

↳somma dei quadrati per le Y (Sy^2 o Σy^2)

↳la somma dei prodotti (Sxy o Σxy).

- **Alcuni chiarimenti sulla simbologia doperata: l'espressione "somma dei prodotti"**
  - **Sxy = co-devianza (congiunta)**
    - **è un' abbreviazione per**
      - **"somma dei prodotti degli scarti delle X dalla media**
      - **delle $\overline{Y}$ per gli scarti delle Y dalla media delle $\overline{X}$ ".**
- **Le lettere maiuscole X e Y indicano sempre i dati iniziali, mentre le lettere minuscole x e y indicano sempre gli scarti dei dati iniziali dalla media cioè X- $\overline{X}$**

- **Non si devono confondere i simboli**

- SX^2 (somma dei quadrati dei dati)
- $(SX)^2$ (quadrato della somma dei dati)

- Sx^2 (somma dei quadrati degli scarti dei dati sulla media)
- perciò $Sx^2 = S(X-\bar{X})^2$, $Sy^2 = S(Y-\bar{Y})^2$

- $(Sx)^2$ indica il quadrato della somma degli scarti

- Sxy (somma dei prodotti degli scarti congiunti)

Correlazione

- Quando abbiamo un insieme di dati normali-bivariati in cui X e Y appaiono, da un diagramma per punti, essere in relazione tra loro (correlazione) e quando la correlazione mostra di essere rettilinea, possiamo valutare quanto stretta sia questa apparente correlazione e verificare la sua significatività.
- ESISTE una misura del grado in cui X e Y variano congiuntamente. Una tale misura è data dalla somma dei prodotti degli scarti congiunti di X e Y dalle loro rispettive medie divisi per il numero di gradi di libertà, ossia la (co)varianza (C= Sxy/n-2) fra X e Y

Dalla covarianza (C= Sxy/n-2)
si può calcolare la deviazione standard ($\sqrt{C}$)
↳↳ Il numero di gradi di libertà è n-2
perché sono necessariamente noti i valori di a e di b relativi alla retta considerata.

- **La covarianza è positiva quando X e Y tendono a variare nella stessa direzione (proporzionalità diretta) e negativa quando al crescere dell'una l'altra decresce (proporzionalità inversa).**
- **La covarianza non è una misura conveniente della vicinanza alla correlazione ipotizzata perchè (come la varianza) la sua grandezza dipende dalle unità di misura con cui vengono misurate X e Y.**

Questo svantaggio è rimosso esprimendo lo scarto in unità di deviazione standard, che è una specifica unità di misura, la quantità risultante è chiamata coefficiente di correlazione o momento prodotto r

- **e la nostra "stima" del suo valore è data da:**

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

dove con Σx^2 si indica la sommatoria dei quadrati degli scarti di X e con Σy^2 la sommatoria dei quadrati degli scarti di Y.

Essendo r una specifica unità di misura della deviazione std, è il parametro più significativo ed immediato come indice di accuratezza (anche da solo)

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

- Il coefficiente r può avere valori che vanno da
  - +1 a -1 (-1↔0↔+1)
- r = +1 corrisponde ad una correlazione lineare per cui le due variabili sono correlate positivamente;
- r = -1 corrisponde ad una correlazione lineare in cui le due variabili sono correlate negativamente;
- I valori di r molto vicini a +1 e a -1 indicano una forte approssimazione alla correlazione lineare;
- I valori di r intermedi (verso lo zero) possono essere dovuti all'assenza di correlazione o alla esistenza di una correlazione che non è essenzialmente rettilinea.

- **Un particolare parametro denominato coefficiente di determinazione ed indicato con R^2 permette di avere solo correlazioni positive perché assume valori compresi tra 0 e 1.**
  Se $R^2 = 1$ esiste una perfetta correlazione lineare fra X e Y per cui ad un determinato valore di X corrisponde uno ed uno solo valore di Y.
  Se $R^2 = 0$ non esiste alcuna correlazione lineare fra le due variabili. Questo significa che valori di R^2 così come quelli di r forniscono una indicazione della "bontà" dell'equazione di regressione calcolata.
- **Per la retta $R^2 = 0{,}998$, o $r = \pm 0{,}998$ significa che il 99,8% della variabilità dei valori di Y è attribuibile alla sua relazione lineare con la variabile indipendente X**
  - **La retta tracciata è affidabile**
  - **perché calcolata su valori accurati.**

- **Nella pratica le rette di taratura che si tracciano con le varie tecniche strumentali, ben difficilmente presentano valori inferiori e 0,98 e in ogni caso mai al di sotto di 0,95, perché altrimenti vanno comunque scartate dato che la qualità del lavoro svolto risulterebbe scadente. Il valore di questi coefficienti, oltre che dare una misura della qualità del lavoro svolto, sono un'indicazione chiara dell'affidabilità del metodo utilizzato e dell' efficienza della strumentazione.**
  - **In chimica analitica solitamente r o**
  - **R^2 non ha un valore inferiore a 0,999!**

Il valore di R^2, ma altrettanto di r (più semplice da calcolare: oltretutto in pratica le correlazioni sono quasi sempre positive!), oltre che dare una misura della qualità del lavoro svolto, sono una chiara indicazione dell'affidabilità del metodo utilizzato e dell' efficienza della strumentazione.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$$

$$R^2 = \frac{\left[\sum xy - \frac{(\sum x)(\sum y)}{n}\right]^2}{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \bullet \left[\sum y^2 - \frac{(\sum \ )^2}{n}\right]}$$

TARATURA mediante RETTA DI REGRESSIONE E INTERPOLAZIONE LINEARE

Allo scopo di calcolare la retta di regressione (taratura) vanno predisposti opportuni valori di X, almeno tre (meglio di più) per i quali si misurano i corrispondenti valori di Y. Se la relazione tra X e Y stimata statisticamente è di tipo lineare, da essa è possibile calcolare un valore X_i (incognito) corrispondente ad un valore Y_i misurato e compreso fra due valori Y_{n-1} e Y_n (a cui corrispondono X_{n-1} e X_n rispettivamente), in base alla solita equazione:

$$Y_i = a + bX_i \quad \text{da cui} \quad X_i = (Y_i - a)/b$$

Nel nostro caso X_i è la concentrazione incognita di un campione noto qualitativamente, si misura la risposta strumentale (es. assorbanza) e si procede al calcolo utilizzando l'equazione della retta trovata

Dati ottenuti da una determinazione spettrofotometrica e loro elaborazioni matematiche per l'applicazione del metodo dei minimi quadrati

$r = \Sigma xy/\sqrt{\Sigma x^2 y^2}$

$r = 0{,}9874$

A (Y)	C (X)	$x\ (X_i - \overline{X})$	$x^2\ (X_i - \overline{X})^2$	$y\ (Y_i - \overline{Y})$	$x\,y\ (X_i - \overline{X})(Y - \overline{Y})$	$y^2 (Y_i - \overline{Y})^2$
0,10	29,8	-9,3	86,49	-0,30	2,79	0,09
0,20	32,6	-6,5	42,25	-0,20	1,30	0,04
0,30	38,1	-1,0	1,00	-0,10	0,10	0,01
0,40	39,2	0,1	0,01	0,00	0,00	0,00
0,50	41,3	2,2	4,84	0,10	0,22	0,01
0,60	44,1	5,0	25,00	0,20	1,00	0,04
0,70	48,7	9,6	96,16	0,30	2,88	0,09
$\overline{Y} = 0{,}40$ Assorb.	$\overline{X} = 39{,}1$ C = mg/l		$\Sigma = 251{,}75$		$\Sigma = 8{,}29$	$\Sigma\, y^2\ 0{,}28$

$b = 8{,}29/251{,}75 = 3{,}29 \cdot 10^{-2}$

$b = \Sigma xy/\Sigma x^2$

$a = 0{,}40 - 3{,}29 \cdot 10^{-2} \cdot 39{,}1 = -0{,}89$

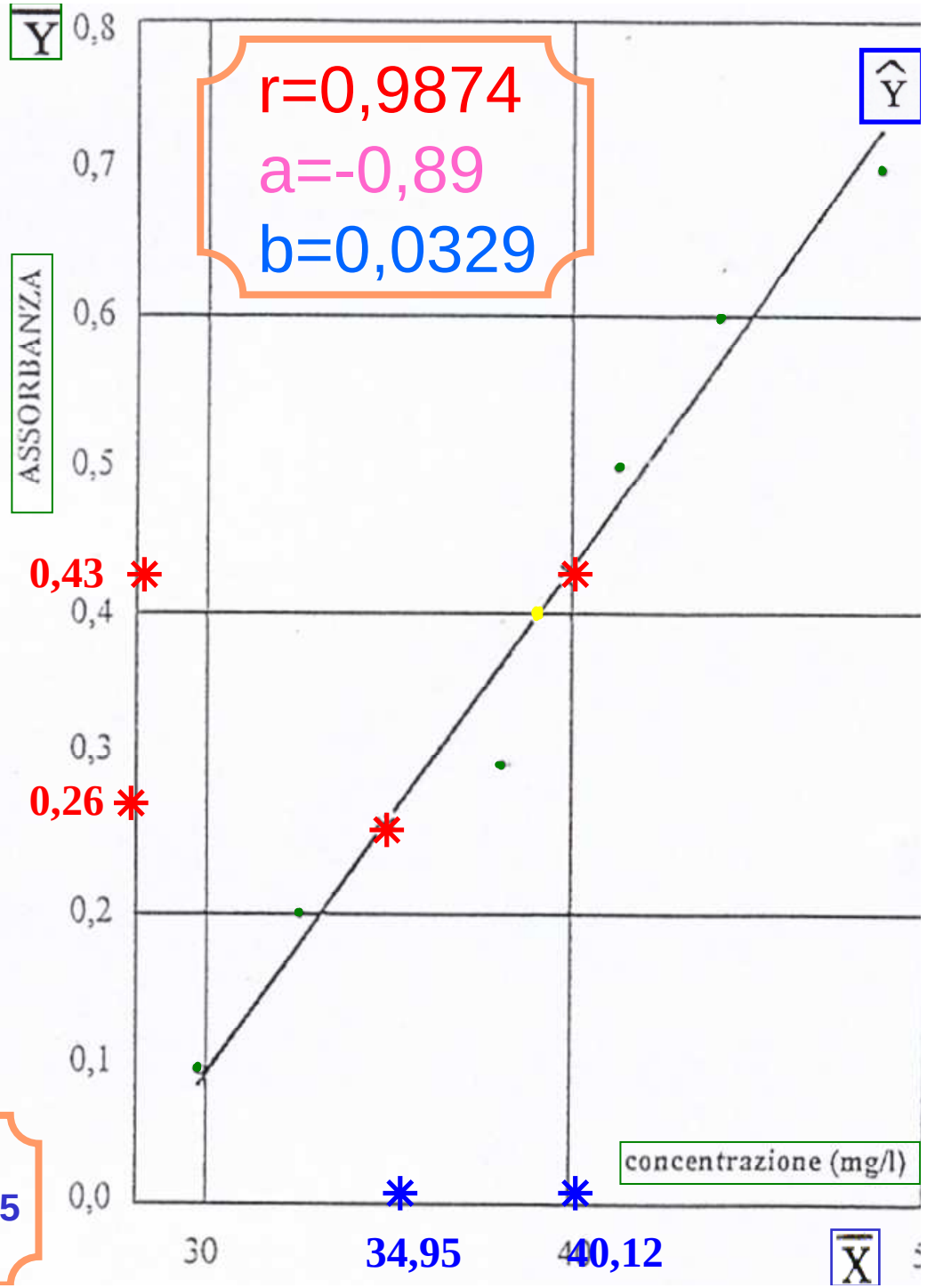
$a = \overline{Y} - b\,\overline{X}$

L'equazione della retta è : $Y = 0{,}0329X - 0{,}89 \Leftarrow\Leftarrow \overline{Y} = a + b\overline{X}$

La retta di taratura è costruita in base ai valori calcolati col metodo della regressione lineare, che sono statisticamente i più probabili. Per tracciarla basta fissare due punti, cioè calcolare i valori di Y (assorbanza) per due valori di concentrazione X (es. 40,00 e 35,00 teorici) compresi nell'intervallo sperimentale. Con lo stesso principio partendo da due valori di assorbanza Y_i si calcolano i corrispondeti valori di X_i

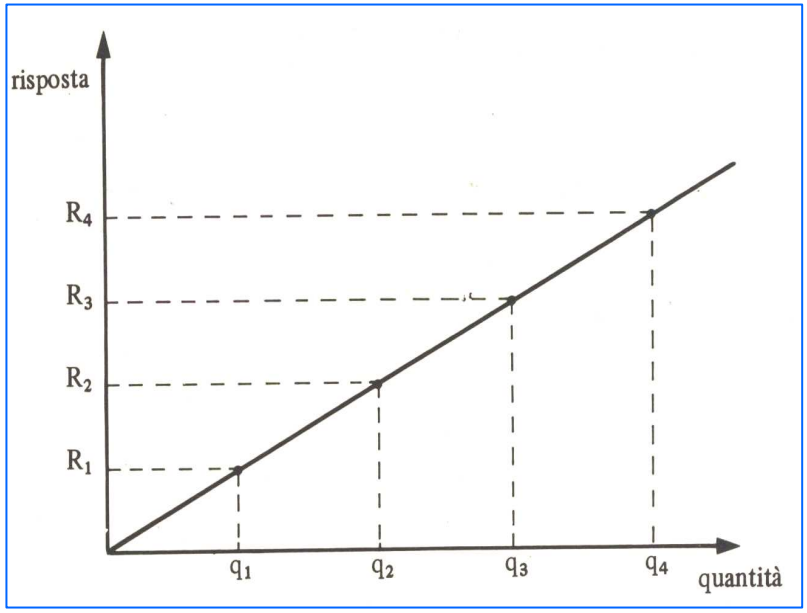
$$X_i = \frac{(Y_i - a)}{b}$$

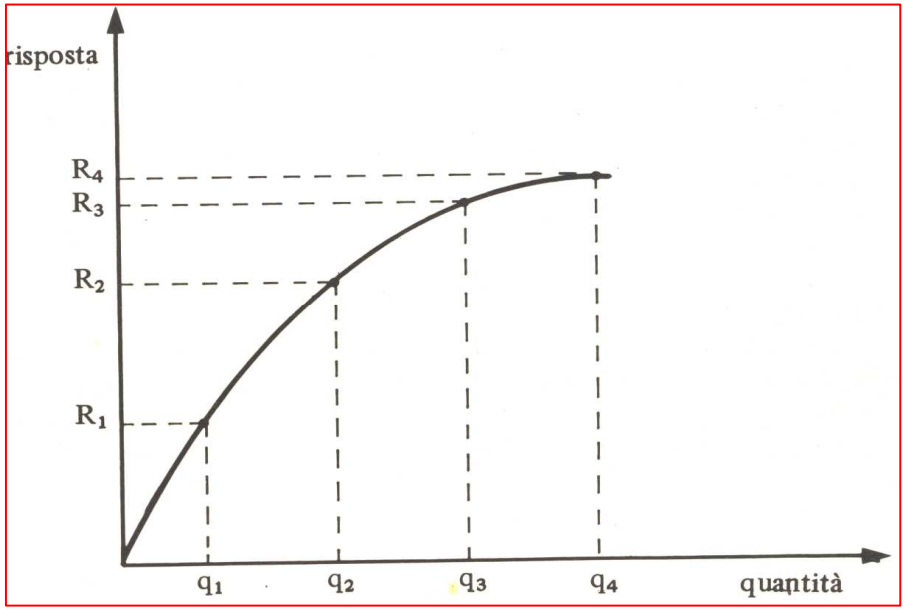
Teorici: 40,00; 35,00
Calcolati: 40,12; 34,95
(con la regressione)



(Premettendo che, calcolata la retta,
si può anche fare a meno della rappresentazione grafica)

La retta di regressione deve essere tracciata nell'intervallo dei punti dai quali è stata calcolata (interpolazione), non oltre poiché questa sarebbe una ingiustificata estrapolazione. Infatti, dall'equazione della retta si possono calcolare solamente i valori ubicati su di essa e correlati linearmente, e non di certo quelli relativi a due grandezze che non fossero correlate o la cui correlazione non fosse rettilinea.





Il metodo dell'interpolazione lineare risulta tanto più preciso quanto più piccolo è l'intervallo Y_n e Y_{n-1}.

Nella tabella sono riportati i dati relativi alla retta di taratura della determinazione spettrofotometrica dell'ammoniaca per calcolare una concentrazione X_i dalla risposta strumentale Y_i in base al calcolo: $X_i = (Y_i - a)/b$ $\{r = \Sigma xy/\sqrt{\Sigma x^2 y^2};\quad b = \Sigma xy/\Sigma x^2;\quad a = \overline{Y} - b\,\overline{X}\}$

	X (ppm NH_4^+)		Y (Assorbanza a 420 nm)
	0,2		0,035
	0,7		0,102
	1,0		0,127
X_{n-1}	1,5	Y_{n-1}.	0,208
X_i	1,74	Y_i	0,24
X_n	2,0	Y_n	0,272
	2,5		0,348

Nella rappresentazione grafica della funzione, occorre tenere presenti alcune avvertenze:

❶- Le variabili dipendente e indipendente vanno riportate rispettivamente in ordinata e in ascissa, in modo chiaro e facilmente individuabile (carta millimetrata).

❷- Le scale vanno scelte in modo tale che i valori possano essere letti velocemente e con facilità.

❸- L'intervallo dei valori deve coprire, se possibile, tutto lo spazio a disposizione sulla carta.

❹- Le scale vanno scelte in modo che la pendenza della retta risulti quanto più possibile vicina all'unità (inclinata di 45º).

❺ - A meno di altre condizioni, la scala prescelta deve permettere di ottenere un grafico il più rettilineo possibile.

❻- Le unità di misura sui due assi non devono consentire una precisione di lettura superiore alla precisione con cui si è effettuata la misura (scala semilogaritmica).

Nel grafico si può applicare l'interpolazione lineare per X_i

ESEMPIO di retta di taratura per la determinazione quantitativa del Cloramfenicolo in coluzione acquosa mediante analisi Spettrofotometrica

λ = 278 nm

retta di taratura del cloroamfenicolo

n° camp.	camp. h	ppm teor	ppm calc	X = ppm	
Std 1	630	19,20	19,11	Y = H camp	
Std 1/2	320	9,60	9,84	Intercept	-9,2609
std1/4	147	4,80	4,67	Slope	33,446
Std 1/8	70	2,40	2,37	Correl	0,9997

Supponiamo di avere 3 campioni a concentrazione incognita X_ia, X_ib, X_ic a cui corrispondono rispettivamente le assorbanze misurate Ya 500,00; Yb 325,00; Yc 124,50. Le concentrazioni calcolate ($X_i=Y_i$-a/b) cadono esattamente sulla retta di regressione e saranno: Xa 15,23; Xb 9,99; Xc 4,00.