Statistica Applicata ai Metodi Analitici

Corso di Laurea in Biotecnologie — Università degli Studi di Teramo
Anno Accademico 2026/2027 · Prof. Federico Fanti

CHIMICA ANALITICA · II ANNO

Indice della Lezione

01

Incertezza della misura

Perché ogni misura ha un'incertezza e come si esprime correttamente

02

Dal dato alla conoscenza

Tipologie di dati e segnale analitico

03

Tipi di errore

Grossolani, sistematici e casuali

04

Precisione, esattezza, accuratezza

Definizioni VIM/ISO 5725 e bersagli

05

Statistica descrittiva

Media, mediana, deviazione standard, RSD

06

Distribuzione normale e inferenza

Gaussiana, intervallo di confidenza, test anomalie

07

Incertezza GUM e calibrazione

Contributi Tipo A/B, regressione lineare, analisi residui



Perché Ogni Misura Ha un'Incertezza

È impossibile effettuare un'analisi chimica con risultati privi di incertezza. Ogni misura è affetta da errori di varia natura, e il compito del chimico analitico è minimizzare gli errori e stimarne l'entità con accuratezza accettabile.

Passo 1

Stabilire il **massimo errore tollerabile** nella misura prima di iniziare l'analisi

Passo 2

Eseguire misure **replicate** per stimare la variabilità e calcolare il valore medio

Passo 3

Esprimere il risultato come **valore ± incertezza** con unità di misura e cifre significative corrette

Come Si Esprime Correttamente un Risultato

Un risultato analitico si esprime sempre nella forma:

$$x = \bar{x} \pm U \quad \text{[unità di misura]}$$

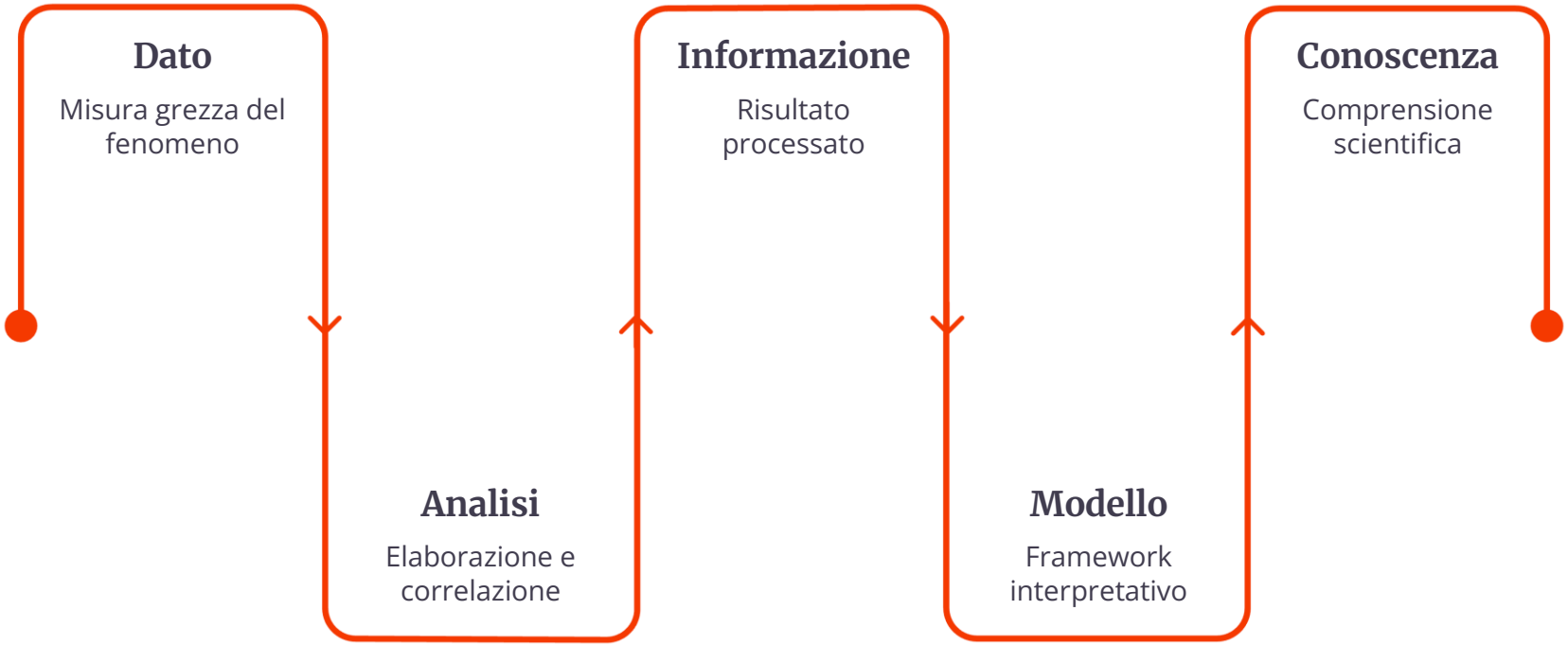
dove **U** è l'incertezza estesa (vedi sezione GUM). Le **cifre significative** dell'incertezza guidano il numero di cifre del valore centrale.

Esempio: concentrazione di glucosio nel sangue = (5,42 ± 0,08) mmol/L

Regole pratiche

- L'incertezza si arrotonda a 1–2 cifre significative
- Il valore centrale si arrotonda alla stessa cifra decimale dell'incertezza
- Indicare sempre l'unità di misura
- Specificare il livello di fiducia (es. k = 2, 95%)

Dal Dato all'Informazione alla Conoscenza



I **dati** sono informazioni elementari che descrivono aspetti particolari di un fenomeno. Di per sé un dato non ha significato: è necessaria un'analisi che lo correli con aspetti significativi del campione. Per esempio, la composizione chimica del sangue acquista senso solo attraverso un modello del corpo umano e delle patologie.

Tipologie di Dati

Dati Quantitativi

Valore numerico + unità di misura
Es.: *la temperatura dell'acqua è 298,2 K*

Base delle **hard sciences**: discipline fondate su dati rigorosi connessi da modelli matematici.

Dati Qualitativi

Etichette, descrittori, categorie
Es.: *"l'acqua è calda"*

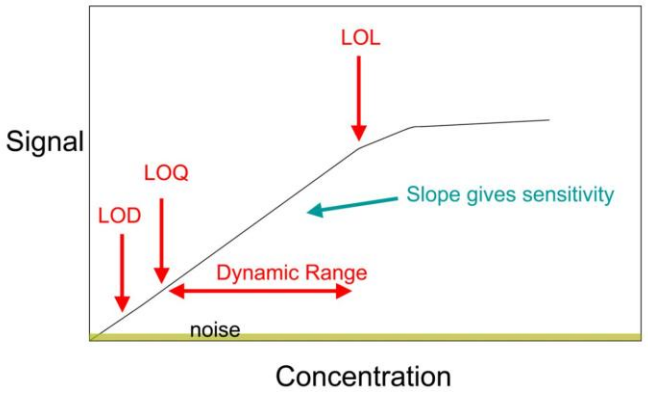
Difficilmente standardizzabili e riproducibili (es. analisi sensoriale).

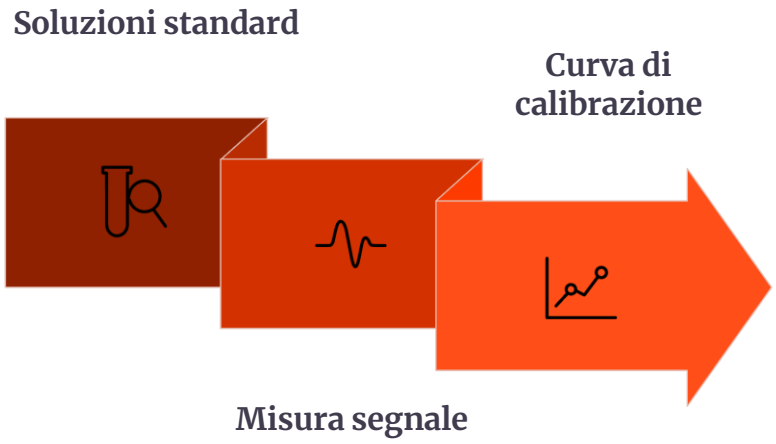
I dati di un individuo includono: altezza, peso, colore della pelle, concentrazione di composti chimici nel sangue, composizione del DNA, taglia di abiti e calzature. La scienza galileiana si fonda sui dati quantitativi.

Il Segnale Analitico e la Curva di Calibrazione

Il **segnale analitico** è una grandezza proporzionale alla quantità di analita presente nel campione. La condizione necessaria è l'esistenza di una relazione stechiometrica (matematica) tra analita e sorgente del segnale.

Aumentando la concentrazione dell'analita, il segnale aumenta proporzionalmente, consentendo di costruire la **curva di calibrazione**.





CAPITOLO 1

Tipi di Errore nelle Analisi Chimiche

Errori Grossolani

Errori macroscopici facilmente individuabili (es. campione sbagliato, contaminazione evidente). Devono essere eliminati.

Errori Sistematici

Agiscono secondo leggi definite, causano sovrastima o sottostima del valore vero. Identificabili e correggibili.

Errori Casuali

Fluttuazioni imprevedibili di parametri sperimentali. Non eliminabili, ma minimizzabili e quantificabili statisticamente.

Il Concetto di Errore: Due Significati

Significato 1 — Errore come differenza

L'errore rappresenta la differenza tra il valore misurato e il valore "vero" o "noto":

$$E = x_i - x_t$$

Significato 2 — Errore come incertezza

L'errore indica l'incertezza stimata in una misura o in un esperimento, espressa come intervallo attorno al valore centrale.

Il termine **bias** indica l'errore sistematico della popolazione:
bias = μ − xt
dove μ è la media della popolazione e xt è il valore vero.

Ciò che si può fare è: **minimizzare gli errori e stimarne l'entità con un'accuratezza accettabile.**

Errori Sistematici: Classificazione



Strumentali

Inesatta calibrazione, uso improprio di vetreria e strumenti, strumenti non idonei.



Di Metodo

Comportamento non ideale di reattivi e reazioni, condizioni sperimentali non idonee (es. solubilità inattesa, tempi di calcinazione inadeguati).



Personali

Distrazione, ignoranza della procedura, difetti fisici (astigmatismo, daltonismo), pregiudizi inconsci (es. terminare la titolazione vicino al volume atteso).

Gli errori sistematici dipendono da cause che agiscono secondo **leggi definite** e possono essere individuati e corretti.

Errori Sistematici: Costanti e Proporzionali

Errori Costanti

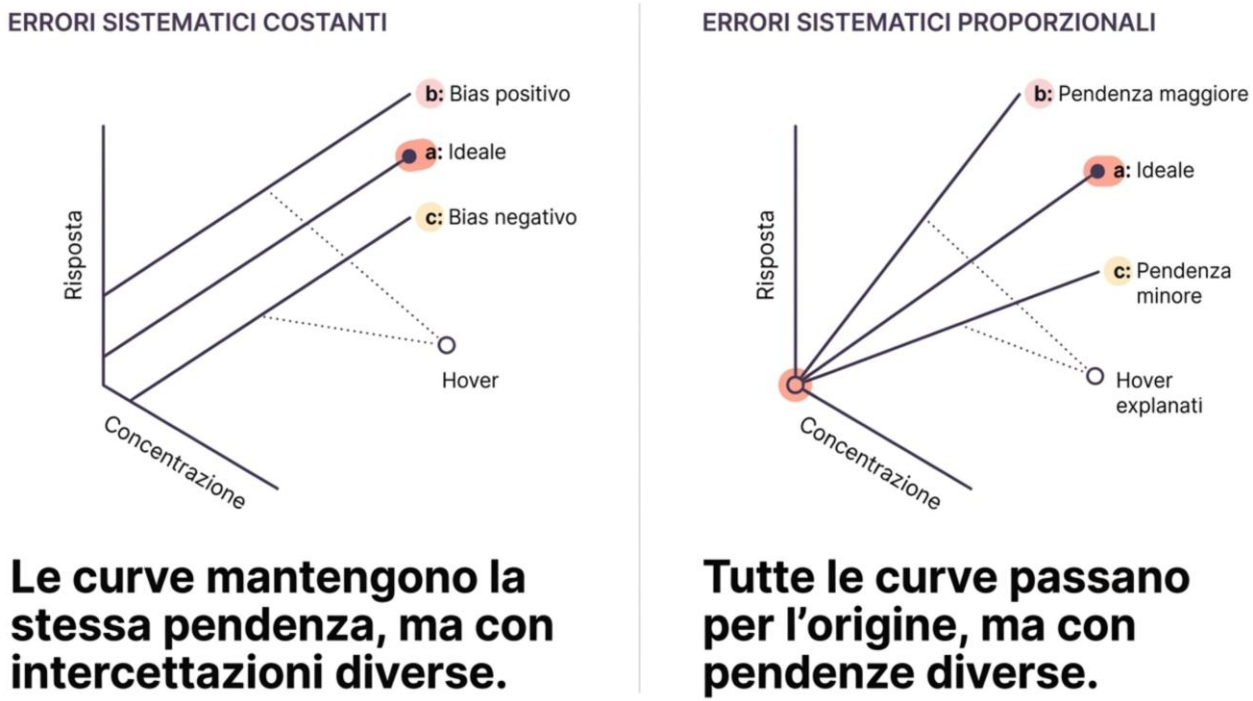
Spostano la retta di calibrazione parallelamente alla retta ideale. L'entità dell'errore è **indipendente dalla concentrazione**.

- Bias negativo: perdite per solubilità in gravimetria
- Bias positivo: assorbimenti estranei in spettrofotometria

Errori Proporzionali

Modificano la pendenza della retta di calibrazione. L'entità dell'errore è **proporzionale alla concentrazione**.

- Bias negativo: perdite di analita per estrazioni non efficienti
- Bias positivo: effetto matrice che amplifica il segnale



Come Identificare ed Eliminare gli Errori Sistematici

1 Materiali di Riferimento Certificati (CRM)

Analisi di campioni standard con valore noto certificato da enti accreditati.

2 Metodo Indipendente

Confronto con strumentazione di provata affidabilità o metodo di riferimento.

3 Analisi del Bianco

Soluzione con tutti i componenti eccetto l'analita: corregge il contributo della matrice e del solvente.

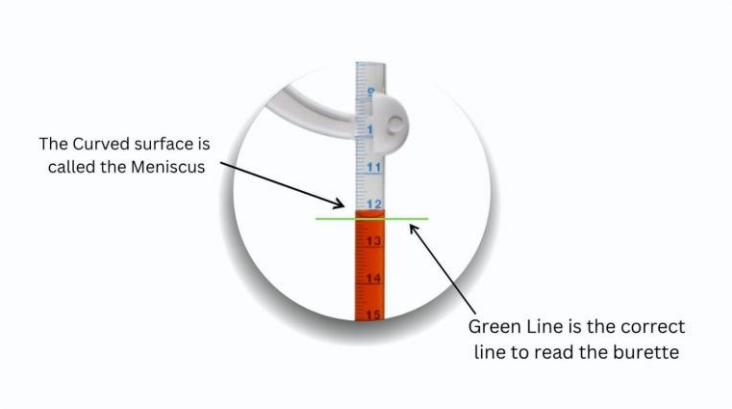
4 Prove di Recupero (Spike)

Aggiunta di quantità note di analita al campione reale per verificare il recupero percentuale.

Errori Casuali: Caratteristiche e Esempio

Gli **errori casuali** (o indeterminati, *random*) causano una dispersione simmetrica dei dati intorno al valore medio. Sono legati a fluttuazioni indefinite di: temperatura, pH, pressione, umidità, forza ionica, tolleranze di bilance e vetreria.

Esempio — lettura della buretta:
Buretta con sensibilità = 0,1 mL. Lettura: 9,68 mL. La stima al livello di 0,01 mL (1/10 della sensibilità) è soggetta ad errore casuale.



Gli errori casuali **non possono essere eliminati**, ma possono essere ridotti operando con cura e quantificati con l'analisi statistica su misure ripetute.

CAPITOLO 2

Precisione, Esattezza e Accuratezza

Le definizioni seguono le norme **VIM** (Vocabulaire International de Métrologie) e **ISO 5725**. L'accuratezza è il concetto più ampio e comprende sia l'esattezza che la precisione.

Precisione

Bontà dell'accordo tra risultati di misurazioni successive nelle stesse condizioni. Quantificata dalla deviazione standard.

Esattezza

Bontà dell'accordo tra il valore medio dei risultati e il valore vero xt. Assenza di errori sistematici.

Accuratezza

Comprende esattezza e precisione. Una misura accurata è sia esatta che precisa.

I Quattro Casi: Illustrazione con i Bersagli



Né esatto né preciso

Errori sistematici e casuali entrambi presenti. Punti dispersi lontano dal centro.



Preciso ma non esatto

Solo errori sistematici. Punti raggruppati ma lontani dal centro.



Esatto ma non preciso

Solo errori casuali. Punti dispersi ma centrati sul valore vero.



Esatto e preciso

Obiettivo ideale. Punti raggruppati e centrati sul valore vero.

Ripetibilità e Riproducibilità

Ripetibilità

Precisione ottenuta nelle **stesse condizioni**: stesso operatore, stesso strumento, stesso laboratorio, breve intervallo di tempo. Variabilità minima attesa.

Riproducibilità

Precisione ottenuta in **condizioni diverse**: operatori, strumenti o laboratori differenti. Variabilità maggiore, ma essenziale per la validazione interlaboratorio.

ISO 5725 distingue chiaramente i due concetti. La riproducibilità è sempre ≥ ripetibilità in termini di varianza. Entrambe si quantificano con la deviazione standard nelle rispettive condizioni.

CAPITOLO 3

Statistica Descrittiva: Campione e Popolazione

Campione

Insieme delle misure in esame (finito). Notazione: $\bar{x}$ per la media, **s** per la deviazione standard.

Popolazione

Insieme di tutte le possibili misure (teoricamente infinito). Notazione: **μ** per la media, **σ** per la deviazione standard.

La distribuzione del campione è rappresentabile in modo **discreto** (istogramma). La distribuzione della popolazione è rappresentabile con una **curva continua** (curva normale o gaussiana). Le statistiche del campione sono stime dei parametri della popolazione.

Media e Mediana

Media Aritmetica

Somma delle misure divisa per il numero di misure:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N}$$

Sensibile ai valori anomali (outlier).

Mediana

Valore centrale dopo ordinamento crescente. Per N pari: media dei due valori centrali. Robusta agli outlier.

Esempio con valore anomalo:

$$X = \{12, 15, 20, 23, 24, 27, \mathbf{200}\}$$

Mediana = **23** · Media = **45,9**

⚠ Il valore anomalo 200 sposta la media a 45,9, ben lontano dal gruppo principale. La mediana (23) è molto più rappresentativa in questo caso.

Deviazione Standard

La **deviazione standard** (o scarto quadratico medio) misura la dispersione dei dati intorno al valore atteso. Ha la stessa unità di misura dei valori osservati.

Per la Popolazione (σ)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

Si usa quando si conosce l'intera popolazione.

Per il Campione (s)

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

Si usa **n−1** (gradi di libertà) quando si lavora con un campione della popolazione.

Deviazione Standard Relativa (RSD) e Coefficiente di Variazione (CV%)

$$\text{RSD} = \frac{s}{\bar{x}} \qquad \text{CV\%} = \frac{s}{\bar{x}} \times 100$$

Esercizio Numerico — Statistica Descrittiva

Determinazione della concentrazione di Fe(III) in un campione d'acqua (mg/L), 7 replicati:

Replicato	1	2	3	4	5	6	7
Conc. (mg/L)	2,41	2,38	2,45	2,43	2,39	2,44	2,42

Calcoli

$$\bar{x} = \frac{2{,}41 + 2{,}38 + 2{,}45 + 2{,}43 + 2{,}39 + 2{,}44 + 2{,}42}{7} = 2{,}417 \text{ mg/L}$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{6}} = 0{,}024 \text{ mg/L}$$

Risultati

2,417

mg/L — Media

0,024

mg/L — Dev. Std.

1,0%

CV% — Precisione

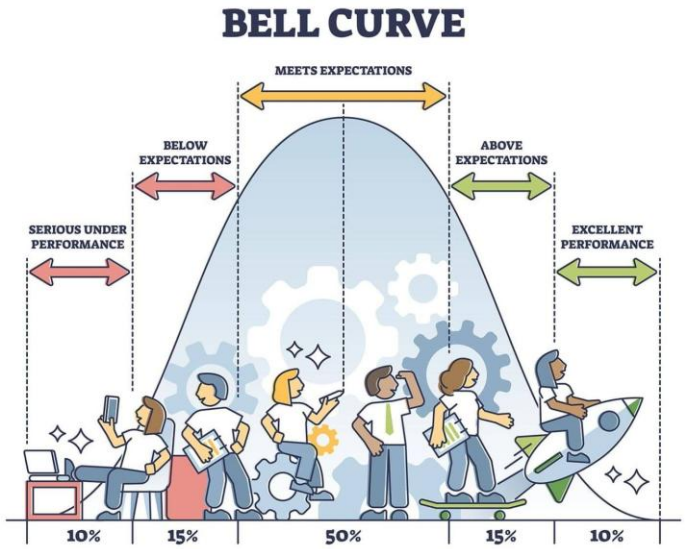
CAPITOLO 4

La Distribuzione Normale (Gaussiana)

La **funzione gaussiana** descrive la distribuzione degli errori casuali:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad -\infty < y < \infty$$

dove **μ** è la media della popolazione e **σ** la deviazione standard. L'area totale sotto la curva è uguale a **1** (rappresenta la probabilità totale).



μ ± 1σ

68,3% delle misure

μ ± 2σ

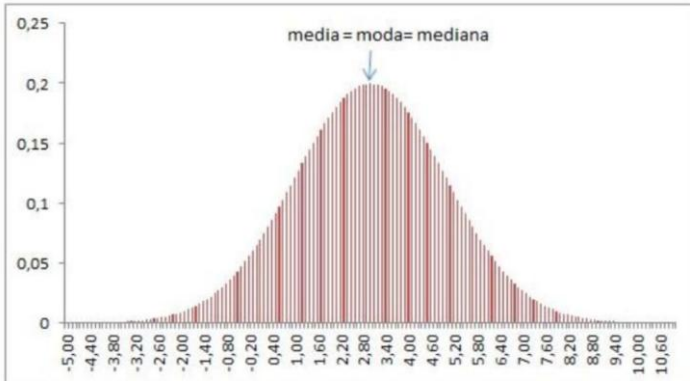
95,4% delle misure

μ ± 3σ

99,7% delle misure

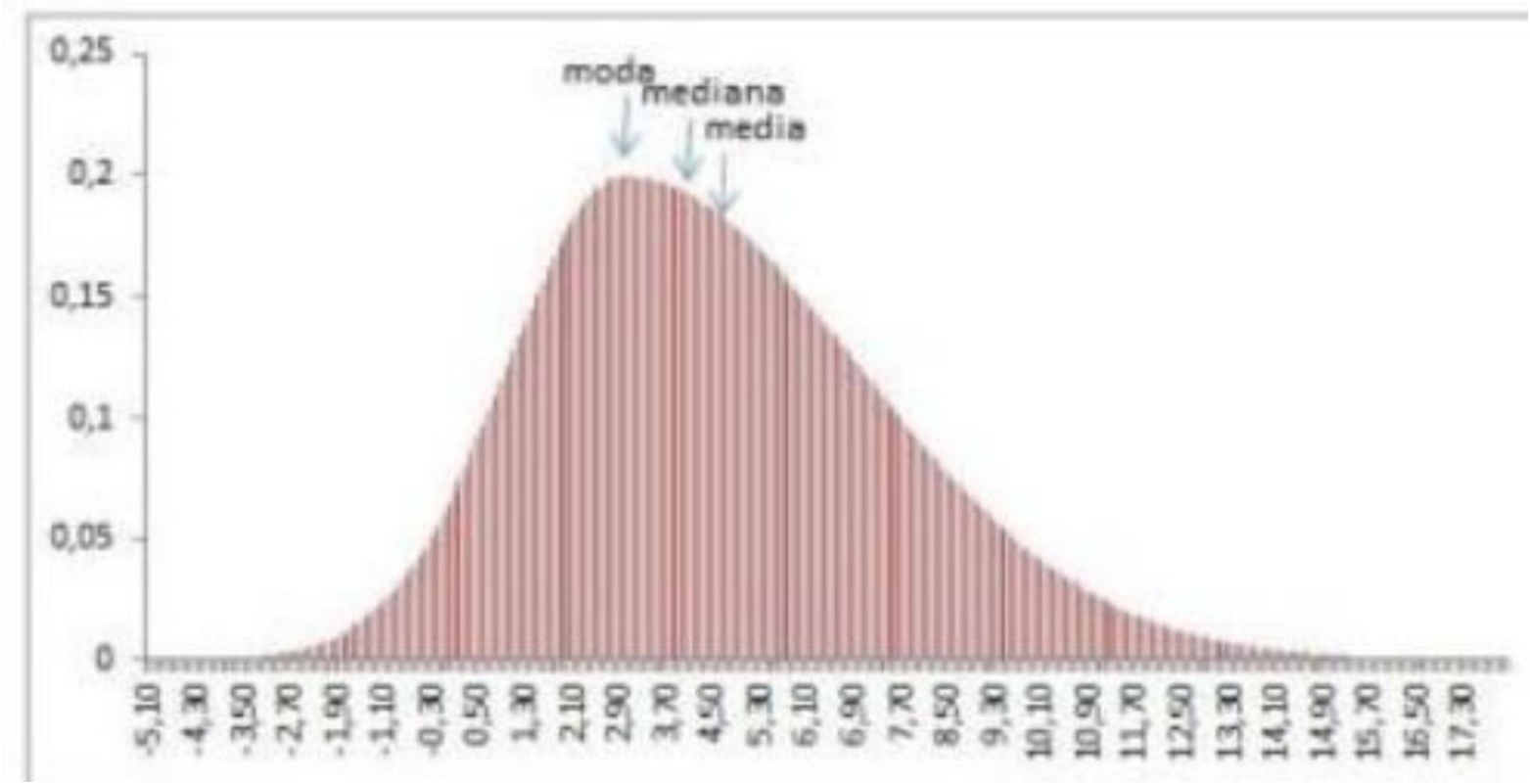
Distribuzioni: Simmetrica

Quando i dati non seguono una distribuzione normale, la posizione relativa di **media, mediana e moda** cambia in modo caratteristico.



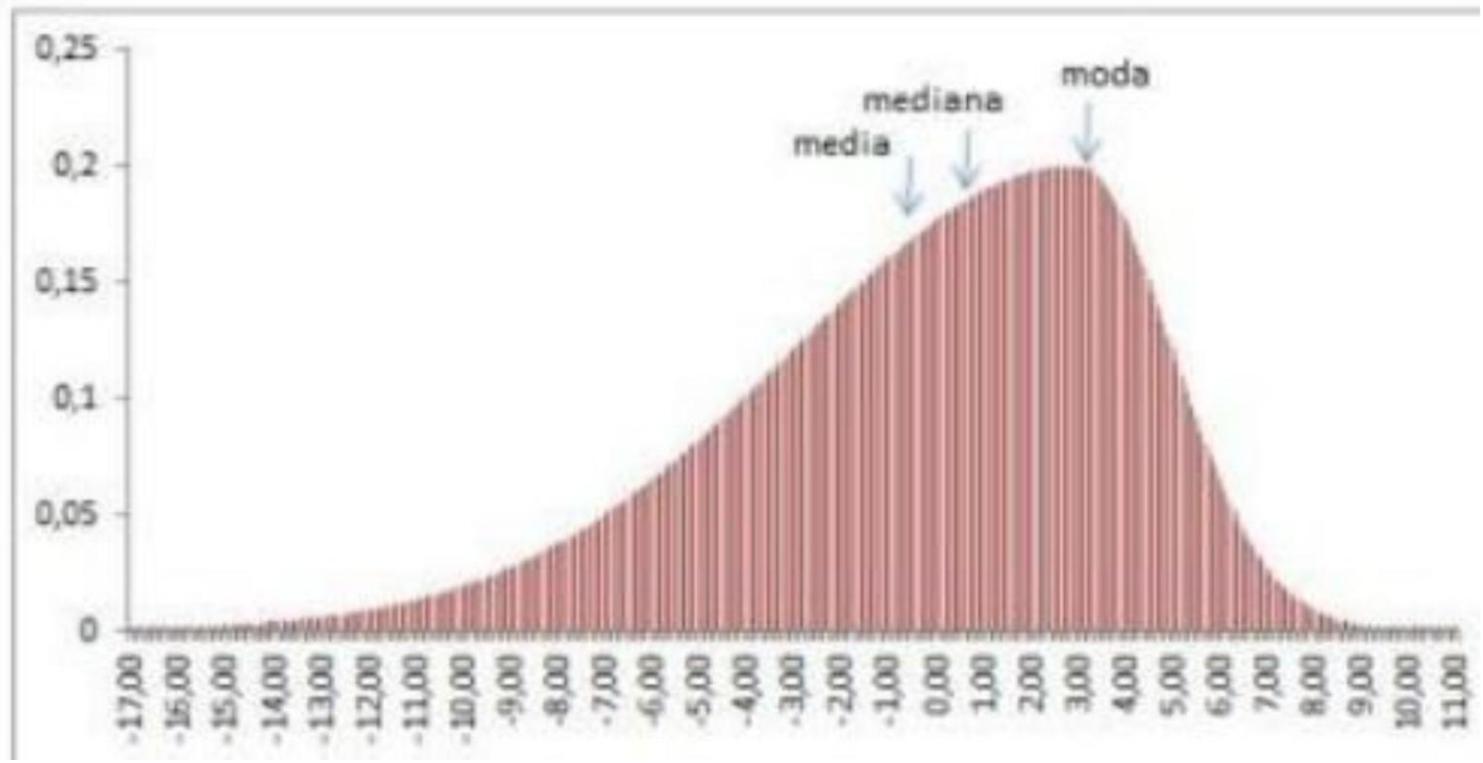
Distribuzione Simmetrica

Media = Mediana = Moda. Tipica degli errori casuali puri.



Asimmetria Positiva (coda destra)

Media > Mediana > Moda. La coda si estende verso destra.



Distribuzioni: Asimmetria Negativa (coda sinistra)

Moda > Mediana > Media. La coda si estende verso sinistra.

Inferenza di Base: Errore Standard e Distribuzione t

Errore Standard della Media

La media di un campione è essa stessa una variabile casuale. La sua deviazione standard è l'**errore standard**:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Aumentando n, l'errore standard diminuisce: più misure → stima più affidabile.

Distribuzione t di Student

Per campioni piccoli (n < 30) con σ ignoto, si usa la distribuzione **t di Student** con ν = n−1 gradi di libertà.

Intervallo di Confidenza

$$\bar{x} \pm t_{\nu,\alpha} \cdot \frac{s}{\sqrt{n}}$$

Il valore t dipende dal livello di fiducia (es. 95%) e dai gradi di libertà.

Test per Valori Anomali: Q di Dixon e Grubbs

Test Q di Dixon

Adatto per campioni piccoli ($3 \leq n \leq 10$). Si calcola il rapporto tra la differenza del valore sospetto dal suo vicino e il range totale:

$$Q = \frac{| x_{sospetto} - x_{vicino} |}{x_{max} - x_{min}}$$

Se Q > Qcritico (tabella), il valore è anomalo al livello di fiducia scelto.

Test di Grubbs

Adatto per campioni più grandi. Si calcola la statistica G:

$$G = \frac{| x_{sospetto} - \bar{x} |}{s}$$

Se G > Gcritico, il valore è statisticamente anomalo. Più potente del test Q per n > 10.

Un valore anomalo va rimosso solo se si identifica una causa sperimentale reale (es. errore di trascrizione, contaminazione). Non si elimina mai solo perché "distante".

CAPITOLO 5

Incertezza di Misura secondo GUM

Il **GUM** (Guide to the Expression of Uncertainty in Measurement, JCGM 100:2008) fornisce il quadro di riferimento internazionale per la valutazione e l'espressione dell'incertezza. Distingue due categorie di contributi.

Contributo di Tipo A

Valutato mediante analisi statistica di serie di osservazioni (es. incertezza di ripetibilità). Si calcola come errore standard della media: uA = s/√n.

Contributo di Tipo B

Valutato con metodi diversi dall'analisi statistica: certificati di taratura, specifiche del costruttore, dati di letteratura, giudizio esperto.

Fonti di Incertezza e Incertezza Estesa

Ripetibilità (Tipo A)

uA = s/√n dalla serie di replicati

Taratura Strumentale (Tipo B)

Dal certificato di taratura dello strumento

Pesata (Tipo B)

Dalla risoluzione e dalla taratura della bilancia

Volume Finale (Tipo B)

Dalla tolleranza della vetreria volumetrica

Standard Interno e Soluzione di Taratura (Tipo B)

Dalla purezza e dalla preparazione degli standard

Incertezza Combinata e Estesa

$$u_c = \sqrt{u_A^2 + u_{B1}^2 + u_{B2}^2 + \dots} \qquad U = k \cdot u_c \quad (k = 2,\ \text{fiducia} \approx 95\%)$$

CAPITOLO 6

Calibrazione e Regressione Lineare

La **regressione lineare** è lo strumento statistico fondamentale per costruire la curva di calibrazione. Permette di prevedere i valori di una variabile dipendente (segnale Y) a partire da una variabile indipendente (concentrazione X).

Regressione Semplice

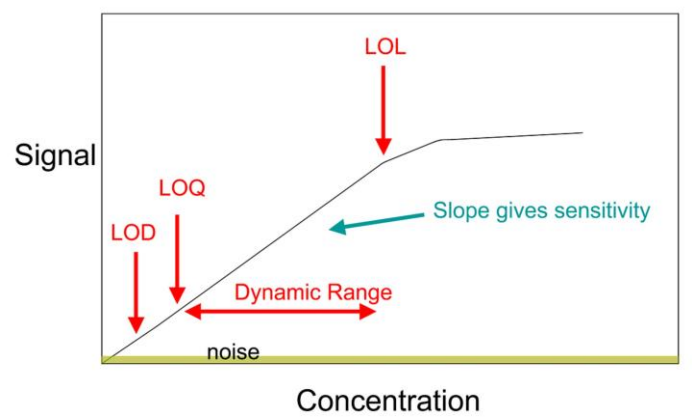
Una sola variabile esplicativa $X \rightarrow Y = \beta_0 + \beta_1 X + \varepsilon$

Regressione Multipla

Diverse variabili esplicative $X_1, X_2, \ldots, X_p \rightarrow$ modelli più complessi

Correlazione

Studio dell'associazione tra variabili quantitative senza previsione causale



Il Modello di Regressione Lineare Semplice

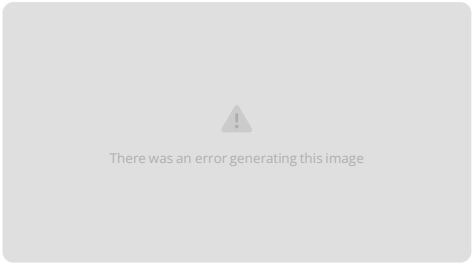
Il modello di regressione lineare semplice è:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

dove:

- $\boldsymbol{\beta_0}$ = intercetta della popolazione (valore di Y quando X = 0)
- $\boldsymbol{\beta_1}$ = pendenza (inclinazione) della popolazione: variazione di Y per unità di X
- $\boldsymbol{\varepsilon_i}$ = errore casuale per l'i-esima osservazione

Il segno di β_1 indica se la relazione è positiva o negativa. In calibrazione, β_1 rappresenta la **sensibilità analitica**.



Metodo dei Minimi Quadrati

Le stime campionarie b_0 e b_1 si ottengono minimizzando la **somma dei quadrati degli errori** (SQE). Le formule sono:

Stima della Pendenza

$$b_1 = \frac{\sum_{i=1}^{n} (X_i - X)(Y_i - Y)}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Stima dell'Intercetta

$$b_0 = \bar{Y} - b_1 \bar{X}$$

Equazione Campionaria della Retta di Regressione

$$\hat{Y}_i = b_0 + b_1 X_i$$

dove $\hat{\mathbf{Y}}_i$ è il valore previsto di Y per l'osservazione i. La retta passa sempre per il punto ($\bar{X}$, $\bar{Y}$

Interpolazione, Estrapolazione e Predizione Inversa

Interpolazione

La previsione di Y corrisponde a un valore di X **interno** all'intervallo usato per stimare il modello. Affidabile e raccomandata.

Estrapolazione

La previsione di Y corrisponde a un valore di X **esterno** all'intervallo. Rischiosa: la linearità potrebbe non valere oltre il range di calibrazione.

Predizione Inversa (Calibrazione)

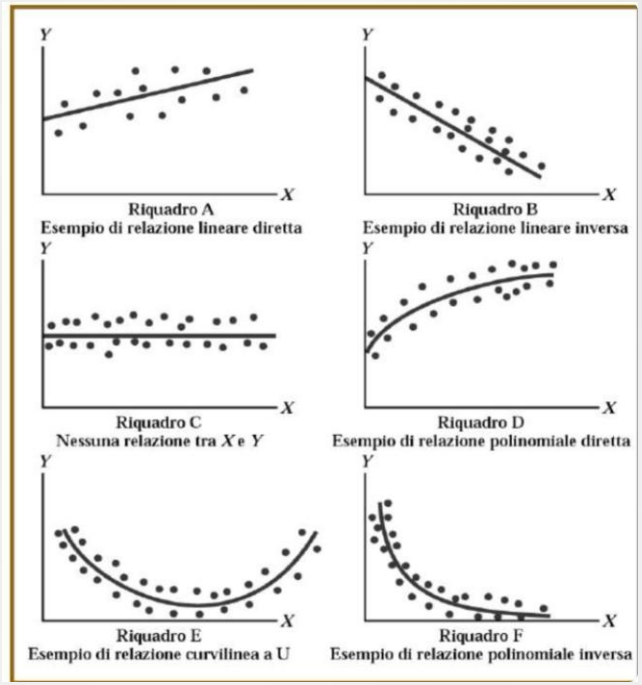
Data una risposta strumentale y_0 misurata su un campione incognito, si ricava la concentrazione:

$$x_0 = \frac{y_0 - b_0}{b_1}$$

Questa è l'operazione fondamentale nell'analisi quantitativa.

Tipi di Relazione nel Diagramma di Dispersione

La scelta del modello matematico appropriato è suggerita dalla distribuzione dei punti nel diagramma di dispersione: relazione lineare diretta (A), inversa (B), assenza di relazione (C), polinomiale diretta (D), curvilinea a U (E), polinomiale inversa (F).



Scomposizione della Variabilità: SQT = SQR + SQE

$$SQT = SQR + SQE$$

SQT — Variabilità Totale

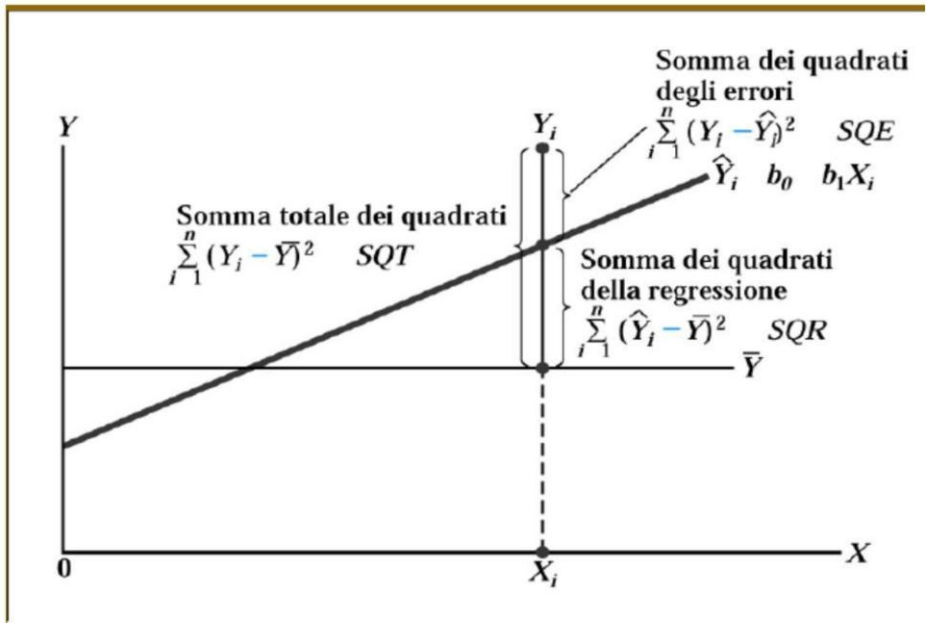
$$SQT = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

SQR — Variabilità Spiegata

$$SQR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

SQE — Variabilità Non Spiegata

$$SQE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$



Coefficiente di Determinazione r^2 e Correlazione di Pearson r

Coefficiente di Determinazione r^2

$$r^2 = \frac{SQR}{SQT}$$

Misura la **proporzione di variabilità di Y spiegata da X**. Varia tra 0 e 1. Un $r^2 = 0{,}999$ indica che il 99,9% della variabilità è spiegata dal modello.

Coefficiente di Correlazione di Pearson r

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \cdot \sum(Y_i - \bar{Y})^2}}$$

Varia tra −1 e +1. Il segno indica la direzione della relazione.

⚠ **Attenzione:** un r^2 elevato non dimostra da solo la linearità del modello. È indispensabile verificare sempre il grafico dei residui per escludere pattern sistematici.

Esercizio Numerico — Curva di Calibrazione

Calibrazione spettrofotometrica per Fe(III) a 5 livelli di concentrazione:

Livello	Conc. X (mg/L)	Assorbanza Y	$\hat{Y}$ (previsto)	Residuo e_i
1	0,50	0,062	0,061	+0,001
2	1,00	0,121	0,120	+0,001
3	2,00	0,238	0,239	−0,001
4	4,00	0,479	0,477	+0,002
5	6,00	0,714	0,715	−0,001

$$b_1 = 0{,}119 \text{ L/mg} \quad b_0 = 0{,}002$$

$$r^2 = 0{,}9999 \quad \Rightarrow \text{ottima linearità}$$

Per un campione incognito con $y_0 = 0{,}350$:

$$x_0 = \frac{0{,}350 - 0{,}002}{0{,}119} = 2{,}92 \text{ mg/L}$$

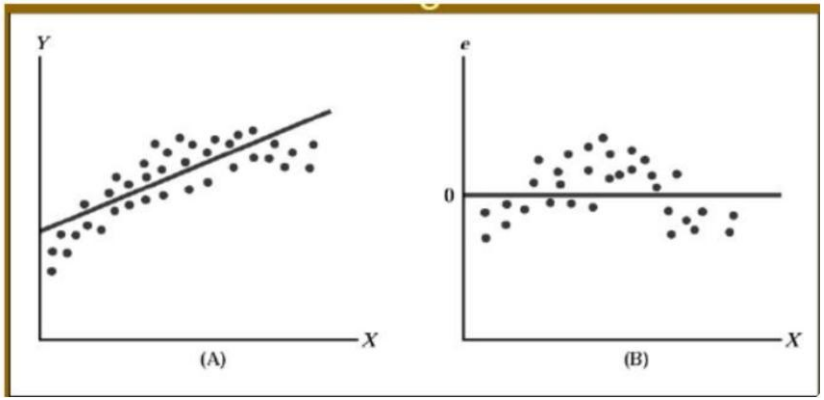
CAPITOLO 7

Analisi dei Residui: Verifica del Modello

Il **residuo** è la stima dell'errore commesso nel prevedere Y_i tramite $\hat{Y}_i$:

$$e_i = Y_i - \hat{Y}_i$$

Il grafico dei residui (e_i) vs X è lo strumento diagnostico fondamentale. Se il modello è adeguato, i residui devono distribuirsi casualmente attorno a zero, senza pattern sistematici.



Se si evidenzia una struttura curvilinea nei residui, il modello lineare non è adeguato: occorre considerare un modello polinomiale o non lineare.

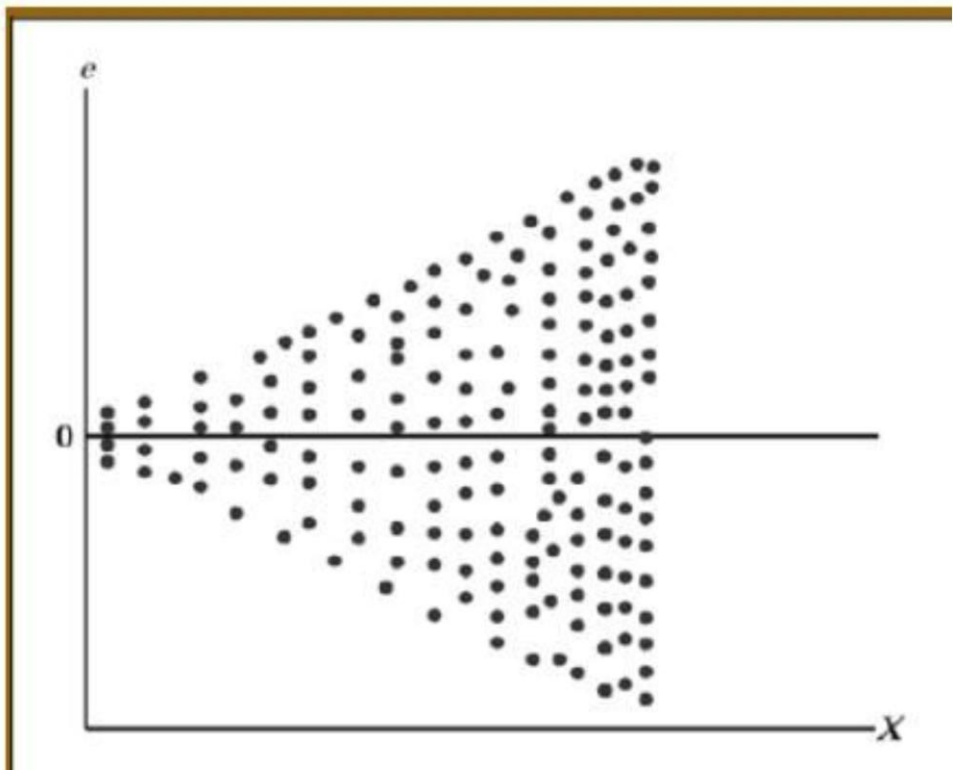
Omoschedasticità ed Eteroschedasticità

Omoschedasticità

La varianza degli errori è **costante** al variare di X. Ipotesi fondamentale della regressione ordinaria ai minimi quadrati (OLS). Il grafico dei residui mostra una banda uniforme attorno a zero.

Eteroschedasticità

La varianza degli errori **aumenta** (o varia) al variare di X. Frequente in chimica analitica ad ampi range di concentrazione. Il grafico dei residui mostra un pattern "a imbuto".



In presenza di eteroschedasticità si ricorre alla **regressione pesata** (weighted least squares), con pesi proporzionali a $1/x$ o $1/x^2$, per dare meno importanza ai punti con maggiore varianza. Questo aspetto è approfondito nella seconda Unità Didattica sulla validazione dei metodi.

Effetto degli Errori Sistematici sulla Curva di Calibrazione

Gli errori sistematici modificano la retta di calibrazione in modo prevedibile, influenzando la stima della concentrazione incognita x_0.

Errore Costante Positivo

Sposta la retta verso l'alto (intercetta maggiore). La concentrazione incognita viene **sottostimata** se non si corregge il bianco.

Errore Costante Negativo

Sposta la retta verso il basso (intercetta minore). La concentrazione incognita viene **sovrastimata**.

Errore Proporzionale Positivo

Aumenta la pendenza (sensibilità apparente maggiore). La concentrazione incognita viene **sottostimata**.

Errore Proporzionale Negativo

Diminuisce la pendenza. La concentrazione incognita viene **sovrastimata**. Tipico di estrazioni non efficienti.



Collegamento alla Validazione dei Metodi

I concetti statistici sviluppati in questa lezione costituiscono la base per la **validazione dei metodi analitici**, trattata nella seconda Unità Didattica. I parametri di validazione principali sono:

Linearità e Range

Verifica di r^2, analisi dei residui, test di mancanza di adattamento (lack of fit)

LOD e LOQ

Limite di rilevazione e di quantificazione, derivati da s del bianco e dalla pendenza

Recupero e Selettività

Prove di spike, interferenti, effetto matrice: tutti valutati con strumenti statistici

Incertezza di Misura

Budget completo secondo GUM, con tutti i contributi Tipo A e Tipo B identificati

Sintesi dei Concetti Chiave

Incertezza

Ogni misura ha un'incertezza. Si esprime come valore ± U con unità e cifre significative corrette.

Errori

Sistematici (bias, correggibili) e casuali (riducibili, quantificabili). Il bias = μ − xt.

Precisione/Esattezza

Precisione = dispersione (s, CV%). Esattezza = assenza di bias. Accuratezza = entrambe.

Gaussiana

Regola 68–95–99,7%. Area totale = 1. Intervallo di confidenza: $\bar{x} \pm t \cdot s/\sqrt{n}$.

GUM

Contributi Tipo A (statistici) e Tipo B (altri). Incertezza estesa U = k·uc con k = 2.

Regressione

$\hat{Y} = b_0 + b_1X$. Minimi quadrati. r^2 e analisi residui. Predizione inversa: $x_0 = (y_0 - b_0)/b_1$.

La statistica non è un optional: è parte integrante del metodo scientifico in chimica analitica. Ogni risultato senza incertezza è incompleto.