

Appunti di Misure Elettriche

Capitolo 1 - Misura e incertezza (parte II)

Valutazione dell'incertezza	1
Deviazione standard della media	1
Definizione dell'incertezza	5
<i>Valutazione Tipo A dell'incertezza (valutazione di tipo statistico)</i>	5
<i>Valutazione Tipo B dell'incertezza (valutazione di tipo aprioristico)</i>	7
<i>Ulteriori considerazioni</i>	10
Raccomandazioni sull'incertezza.....	10
Incertezza standard combinata	11
<i>Grandezze di ingresso scorrelate</i>	13
<i>Grandezze di ingresso correlate</i>	15
Incertezza estesa	17
<i>Livelli ed intervalli di confidenza</i>	18
Uso della distribuzione t.....	19
Presentazione dei risultati.....	21
<i>Rappresentazione mediante tabelle</i>	24
<i>Rappresentazione mediante grafici</i>	24
Cifre significative e scrittura dell'incertezza	25
Esempio: misura (indiretta) di resistenza.....	27
Eliminazione di dati	30
Metodo dei minimi quadrati	31
Rette di regressione e coefficiente di correlazione	34

Valutazione dell'incertezza

DEVIAZIONE STANDARD DELLA MEDIA

Supponiamo di dover misurare una data quantità e di compiere N **misure ripetute** di tale quantità, ottenendo il seguente **campione di misura**: (X_1, X_2, \dots, X_N) . Sappiamo di poter calcolare la media e la deviazione standard di questo campione di misura:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \qquad \sigma \cong \sqrt{\frac{1}{N} \sum_{i=1}^N d_i^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$$

dove ricordiamo che $d_i = X_i - \bar{X}$ è la deviazione della misura i -sima rispetto alla media \bar{X} .

E' importante distinguere la media (o la deviazione standard) del campione di misura dalla media (o deviazione standard) della

variabile aleatoria X con cui indichiamo il valore del misurando in questione. Quando calcoliamo la media e la deviazione standard di un campione di misura, usiamo tali valori come **stima** della media e della deviazione standard di X . Dobbiamo allora chiederci se possiamo ottenere valori più attendibili della media e della deviazione standard di X se consideriamo altri campioni della **popolazione**, ossia se andiamo a compiere ulteriori misure in modo da avere ulteriori campioni di misura.

Per spiegarci meglio, supponiamo di compiere un numero totale $M \cdot N$ di misure della stessa quantità, in modo da disporre dei seguenti M campioni di misura:

$$(X_{11}, X_{12}, \dots, X_{1N}), (X_{21}, X_{22}, \dots, X_{2N}), \dots, (X_{j1}, X_{j2}, \dots, X_{jN}), \dots, (X_{M1}, X_{M2}, \dots, X_{MN})$$

Se calcoliamo la media e la varianza di due distinti campioni di misura, in generale i valori ottenuti non coincideranno. Quindi, possiamo pensare alla media del generico campione di misura come ad una variabile aleatoria, che assume valori in generale diversi per ciascun campione¹. Un indicatore della variabilità di questa variabile aleatoria è la sua deviazione standard: se la deviazione standard è piccola, significa che la media delle N misure del generico campione è attendibile.

In altre parole, la deviazione standard della media del generico campione è un indicatore della attendibilità della media stessa come stima del misurando. Andiamo allora a calcolare questa deviazione standard.

Cominciamo a chiarire la simbologia di cui faremo uso:

- X_{ji} è la misura i -sima appartenente al j -simo campione di misura;
- $\bar{X}_j = \frac{1}{N} \sum_{i=1}^N X_{ji}$ è la media delle misure del j -simo campione di misura;
- $\mu = \frac{1}{N \cdot M} \sum_{j=1}^M \sum_{i=1}^N X_{ji} = \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N X_{ji} = \frac{1}{M} \sum_{j=1}^M \bar{X}_j$ è la media di tutte le $M \cdot N$ misure, che poi coincide, come si vede, con la media delle medie degli M campioni;
- $d_{ji} = X_{ji} - \mu$ è la deviazione della misura X_{ji} rispetto alla media μ di tutte le misure;
- $D_{j,m} = \bar{X}_j - \mu$ è la deviazione della media \bar{X}_j rispetto a μ .

Per prima cosa, immaginando che tutte le misure costituiscano un unico campione di misura (costituito perciò da $M \cdot N$ misure), possiamo calcolare la varianza di tale campione unico: si ha ovviamente che

$$\sigma^2 = \frac{1}{N \cdot M} \sum_{j=1}^M \sum_{i=1}^N d_{ji}^2 = \frac{1}{N \cdot M} \sum_{j=1}^M \sum_{i=1}^N (X_{ji} - \mu)^2$$

Se invece consideriamo un campione $(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_M)$ costituito dalle medie dei singoli campioni, la varianza di tale campione (cioè la **varianza della media**) sarà

$$\sigma_m^2 = \frac{1}{M} \sum_{j=1}^M D_{j,m}^2 = \frac{1}{M} \sum_{j=1}^M (\bar{X}_j - \mu)^2$$

¹ Vale la pena ricordare un semplice ma importante concetto a proposito delle variabili aleatorie. Consideriamo il lancio di un dado (a 6 facce); l'esito del lancio è una variabile aleatoria X che, ad ogni lancio, assume un preciso valore. Ecco perché parliamo di variabili aleatorie e di valori da esse assunti: il risultato della misura di un dato misurando è una variabile aleatoria che, dopo ogni misura, assume un preciso valore.

Ci conviene fare qualche passaggio per esplicitare meglio l'espressione della deviazione $D_{j,m} = \bar{X}_j - \mu$: abbiamo infatti che

$$D_{j,m} = \bar{X}_j - \mu = \frac{1}{N} \sum_{i=1}^N X_{ji} - \mu = \frac{1}{N} \sum_{i=1}^N (X_{ji} - \mu) = \frac{1}{N} \sum_{i=1}^N d_{ji}$$

Sostituendo nell'espressione della varianza della media, abbiamo dunque che

$$\sigma_m^2 = \frac{1}{M} \sum_{j=1}^M D_{j,m}^2 = \frac{1}{M} \sum_{j=1}^M \left(\frac{1}{N} \sum_{i=1}^N d_{ji} \right)^2 = \frac{1}{M} \frac{1}{N^2} \sum_{j=1}^M \left(\sum_{i=1}^N d_{ji} \right)^2$$

Possiamo adesso ipotizzare che le misure effettuate siano tra loro **mutuamente incorrelate**², per cui possiamo approssimare il quadrato della sommatoria con la sommatoria dei quadrati, in modo da scrivere che

$$\sigma_m^2 = \frac{1}{M} \frac{1}{N^2} \sum_{j=1}^M \left(\sum_{i=1}^N d_{ji} \right)^2 = \frac{1}{M} \frac{1}{N^2} \sum_{j=1}^M \sum_{i=1}^N d_{ji}^2 = \frac{1}{N} \frac{1}{M \cdot N} \sum_{j=1}^M \sum_{i=1}^N d_{ji}^2$$

Quella ottenuta, a meno del termine $1/N$, è evidentemente la varianza di tutto l'insieme delle misure, per cui concludiamo che

$$\sigma_m^2 = \frac{\sigma^2}{N}$$

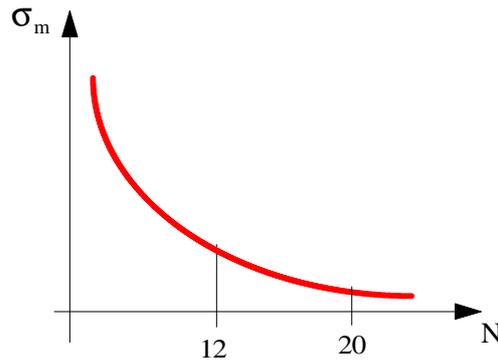
Abbiamo cioè trovato che *la varianza della media di M campioni (di N misure ciascuno) è pari alla varianza di tutte le misure, ridotta del numero N di misure per campione. Se passiamo alla deviazione standard, otteniamo evidentemente*

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

Dato questo legame, la misura più appropriata dell'incertezza del risultato di misura è legata alla deviazione standard s_m delle medie piuttosto che alla varianza s dei singoli campioni di misura.

Non solo, ma quella relazione ci aiuta anche a capire che è più opportuno aumentare il numero N di misure per campione invece che il numero M di campioni di misura. Generalmente, un campione di 12÷20 misure è buono per darci una stima attendibile della media: si è infatti visto che, aumentando il numero di misure oltre 20, il valore di σ_m rimane praticamente costante e molto prossimo a zero:

² il che equivale a dire, come si ricorderà, che è nullo il coefficiente di correlazione mutua tra ciascuna coppia di misure.



D'altra parte, se usiamo un solo campione di misura, sia pure composto da un numero N elevato di misure, avremo il problema che, da esso, potremo solo ottenere una stima sia della media sia della deviazione standard (e quindi dell'incertezza) di X. Dovremo infatti procedere nel modo seguente:

- sia (X_1, X_2, \dots, X_N) l'unico campione di misura a nostra disposizione;
- per prima cosa, possiamo prima stimare la varianza di X tramite la nota relazione

$$s_X^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

dove ovviamente $\bar{X}_j = \frac{1}{N} \sum_{i=1}^N X_{ji}$ è una stima della media di X (detta anche *media sperimentale di X*), in quanto è calcolata facendo la media aritmetica delle misure a nostra disposizione; s_X^2 è in pratica la varianza sperimentale delle N osservazioni;

- in secondo luogo, avendo trovato che $\sigma_m^2 = \frac{\sigma^2}{N}$, possiamo stimare la varianza della media (ottenendo la cosiddetta *varianza sperimentale della media*) ponendo semplicemente

$$s_{\bar{X}}^2 = \frac{s_X^2}{N} = \frac{1}{N(N-1)} \sum_{i=1}^N (X_i - \bar{X})^2$$

- infine, facendo la radice quadrata di questa quantità, avremo una **stima della deviazione standard della media** (detta anche *deviazione standard sperimentale della media*):

$$s_{\bar{X}} = \sqrt{s_{\bar{X}}^2} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (X_i - \bar{X})^2}$$

Come vedremo tra poco, questa relazione è usata per il calcolo dell'incertezza con valutazione di tipo A.

DEFINIZIONE DELL'INCERTEZZA

Abbiamo in precedenza posto in risalto il fatto che i termini *accuratezza* e *precisione* sono solo parametri qualitativi con cui possiamo definire la bontà di una misura:

- per **accuratezza** (*accuracy*) di una misura intendiamo il grado di approssimazione di un insieme ripetuto di misure della stessa quantità al valore del misurando;
- per **precisione** di una misura intendiamo il grado di approssimazione di un insieme ripetuto di misure della stessa quantità al valore medio dell'insieme di misure.

In entrambi i casi, quindi, il concetto è puramente qualitativo. Se invece vogliamo fornire una valutazione sia qualitativa sia soprattutto quantitativa, dobbiamo passare all'**incertezza**. A livello qualitativo, possiamo intenderla come il grado di dubbio sulla validità del risultato di una misura. A livello quantitativo, è necessario invece fare qualche discorso in più, anche se possiamo anticipare sin d'ora che *il valore numerico dell'incertezza di una misura ci dice sostanzialmente quale è l'ultima cifra significativa da considerare nel risultato della misura stessa.*

In altre parole ancora, l'incertezza di una misura è una valutazione, eseguita dall'operatore, sulle cause che incidono significativamente sulla misura fornita e quantifica l'imperfetta conoscenza che si è ottenuta del misurando.

Appare evidente, dunque, che il risultato di misura debba necessariamente essere accompagnato dalla propria incertezza, la quale caratterizza la *dispersione* dei risultati ragionevolmente attribuibili al misurando³.

Si parla di **incertezza standard** quando la si esprime in termini di deviazione standard. Questo concetto sarà chiaro tra poco.

Esistono due modi diversi di valutare l'incertezza⁴: **valutazione Tipo A** e **valutazione Tipo B**. Il concetto è che anche l'incertezza fa riferimento a dei modelli matematici, per cui l'uso di modelli differenti porta a valutazioni di tipo differente.

Valutazione Tipo A dell'incertezza (valutazione di tipo statistico)

La **valutazione Tipo A** dell'incertezza è ottenuta utilizzando i metodi dell'analisi statistica di una serie di osservazioni.

Come si è detto, la migliore stima disponibile del valore atteso di una grandezza X che *varia casualmente* e della quale sono state ottenute N osservazioni indipendenti (nelle stesse condizioni sperimentali) è la **media aritmetica** delle N osservazioni:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

Quindi, la valutazione Tipo A dell'incertezza parte dal fatto di fornire, come stima del misurando, la media aritmetica delle misure effettuate.

³ Ricordiamo, infatti, che il risultato di una misura dovrebbe essere la migliore stima del valore V del misurando, che è impossibile da conoscere per i problemi di cui si è parlato all'inizio del capitolo.

⁴ E' opportuno però sottolineare che l'incertezza è una ed una sola, ma è possibile esprimerla in modi diversi, come si vedrà nei prossimi due paragrafi.

Il motivo per cui tali singole misure *differiscono* tra di loro è nelle variazioni casuali delle cosiddette **grandezze di influenza**, cioè quelle grandezze che non sono il misurando ma alterano comunque il risultato della misura.

Per stimare la varianza di X, si usa la cosiddetta **varianza sperimentale** delle osservazioni, precedentemente definita:

$$s_X^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

Facendo la radice quadrata, otteniamo la **deviazione standard sperimentale** delle osservazioni, anch'essa già citata:

$$s_X = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2}$$

Le quantità s_X^2 e s_X caratterizzano quindi la variabilità dei valori osservati X_i o, più specificatamente, la loro dispersione intorno alla media \bar{X} .

Queste quantità servono a loro volta a stimare l'attendibilità di \bar{X} come stima del misurando. Infatti, come si è visto nel precedente paragrafo, la migliore stima della varianza di \bar{X} è data dalla varianza sperimentale della media, ossia

$$s_{\bar{X}}^2 = \frac{s_X^2}{N} = \frac{1}{N(N-1)} \sum_{i=1}^N (X_i - \bar{X})^2$$

mentre la radice quadrata di questa quantità (cioè la deviazione standard sperimentale della media) è la migliore stima della deviazione standard di \bar{X} :

$$s_{\bar{X}} = \sqrt{s_{\bar{X}}^2} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (X_i - \bar{X})^2}$$

*Queste ultime due quantità quantificano bene quanto \bar{X} stimi il valore dell'aspettazione di X, per cui entrambi possono essere adottati come incertezza della misura di \bar{X} . Per questi motivi, $s_{\bar{X}}$ è spesso chiamata **deviazione standard di tipo A** e, analogamente, $s_{\bar{X}}^2$ è spesso chiamata **varianza di tipo A**.*

Notiamo che la grandezza primitiva fondamentale è la varianza; tuttavia, la deviazione standard $s_{\bar{X}}$ è più conveniente dal punto di vista pratico, in quanto ha la stessa dimensione di X ed inoltre il suo valore è più facilmente interpretabile rispetto a quello della varianza. Il risultato della misura potrà essere quindi fornito nel modo seguente⁵:

$$X = \bar{X} \pm s_{\bar{X}}$$

Talvolta, alla quantità $s_{\bar{X}}$ si dà il nome di **scarto tipo sperimentale della media**.

Segnaliamo inoltre che, quando si forniscono valutazioni di tipo A dell'incertezza, è sempre necessario fornire il **numero di gradi di libertà**. A tal proposito, ricordiamo che, per una variabile

⁵ Sul modo con cui scrivere il risultato di una misura si tornerà più avanti:

aleatoria discreta, il numero dei gradi di libertà è quello degli addendi di una somma, diminuito del numero dei vincoli esistenti tra gli addendi stessi. Allora, nel nostro caso, i gradi di libertà della media aritmetica di N osservazioni indipendenti è $v = N - 1$.

Quindi, in definitiva, per una **valutazione tipo A dell'incertezza**, bisogna fornire

- la **media aritmetica** come stima del misurando;
- la **deviazione standard tipo A** come incertezza (in questo caso come stima della deviazione standard di X);
- il **numero di gradi di libertà**.

Valutazione Tipo B dell'incertezza (valutazione di tipo aprioristico)

La valutazione Tipo B dell'incertezza si basa ancora su metodi statistici (come previsto dalle **norme**), ma diversi da quelli dell'analisi statistica di una serie di osservazioni. E' una valutazione di tipo aprioristico, che non fa uso di misure ripetute. Vediamo di chiarire il senso di questa affermazione.

Sia sempre X la grandezza da stimare. In questo contesto, non abbiamo più a disposizione, come nel caso precedente, delle osservazioni ripetute, ma un'unica osservazione. Di conseguenza, dobbiamo seguire altre strade per valutare l'incertezza di misura, che comunque esprimeremo ancora in termini di deviazione standard (detta in questo caso **deviazione standard di tipo B**) e di varianza (detta in questo caso **varianza di tipo B**).

Ciò che dobbiamo valutare è la distribuzione della grandezza incognita X : nota questa distribuzione, potremo fornire una stima sia di X sia della sua deviazione standard e/o varianza; la stima della deviazione standard (oppure anche della varianza) rappresenterà la nostra incertezza di misura.

La distribuzione di X sarà ottenuta, a priori, tramite un giudizio scientifico basato su tutte le informazioni disponibili sulla possibile variabilità di X .

Queste informazioni possono essere di vario tipo e citiamo perciò le principali:

- dati provenienti da misure precedenti quella in corso;
- esperienza o conoscenza generale del comportamento e/o delle proprietà dei materiali e degli strumenti di interesse;
- specifiche tecniche degli strumenti fornite direttamente dal costruttore;
- dati forniti in certificati di taratura e simili;
- incertezze assegnate a valori di riferimento presi da manuali o comunque da apposite banche dati.

Appare evidente che l'uso delle informazioni disponibili per una valutazione tipo B dell'incertezza richiede notevole esperienza da parte dell'operatore. Inoltre, *l'attendibilità di una valutazione tipo B può essere analoga a quella di una valutazione tipo A, specialmente quando quest'ultima si basa su un numero N relativamente ridotto di osservazioni*.

Vediamo adesso con maggiore dettaglio quali casi si possono presentare.

Una prima possibilità è quello di compiere la propria stima sulla base di una specifica del costruttore dello strumento impiegato oppure sulla base di un certificato di taratura, di un manuale o

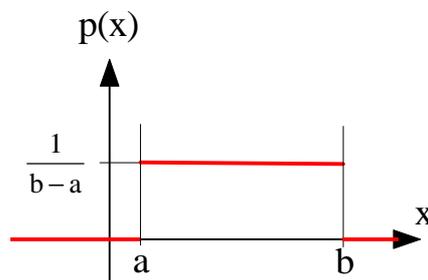
di una fonte simile; se, inoltre, l'incertezza della stima è espressa in termini di un particolare multiplo di una deviazione standard, allora l'incertezza standard si calcola semplicemente dal valore dichiarato dividendolo per il coefficiente moltiplicativo della deviazione standard. Facendo poi il quadrato, si ottiene la stima della varianza.

A volte, l'incertezza indicata nelle specifiche che accompagnano uno strumento è data da un **intervallo** con il suo **livello di confidenza**. In altre parole, si dice che la grandezza X cade in un dato intervallo [a,b] con una certa probabilità. Per esempio, la specifica potrebbe dire che la grandezza X può cadere nell'intervallo [a,b] con una probabilità del 50% (e quindi con la stessa probabilità può cadere al di fuori di tale intervallo). Per risalire, da questa informazione, all'incertezza standard, è necessario conoscere o comunque ipotizzare la distribuzione di probabilità di X; spesso, si può ipotizzare che X abbia una **distribuzione normale**: allora, per esempio nel caso del livello di confidenza del 50%, si può prendere, come migliore stima di X, il punto medio $\frac{a+b}{2}$ dell'intervallo e come incertezza standard la quantità

$$s_x = 1,48 \frac{b-a}{2}$$

Questo numero si spiega molto facilmente: per una distribuzione normale, l'unico intervallo ad avere un livello di confidenza del 50% è l'intervallo $[\mu - 1.48\sigma, \mu + 1.48\sigma]$.

Un altro caso più o meno frequente è quello in cui le specifiche indicano solo l'intervallo [a,b] in cui la grandezza X cade con probabilità unitaria. Se non vengono fornite altre informazioni, l'unica scelta plausibile è quella di ipotizzare una **distribuzione uniforme** (detta anche *rettangolare*) di X nel suddetto intervallo [a,b]:



Sotto questa ipotesi, la stima di X coincide ancora con il valor medio $\frac{a+b}{2}$ dell'intervallo, mentre l'incertezza associata a tale stima è la varianza, che per una distribuzione uniforme vale notoriamente

$$s_x^2 = \frac{(b-a)^2}{12}$$

Possiamo fare un esempio pratico di quest'ultimo caso. Consideriamo infatti uno strumento che fornisca i risultati delle proprie misure su un **display digitale**, con un certo numero di cifre. Supponiamo di compiere delle misure ripetute e che la lettura dello strumento sia sempre la stessa, ad esempio

3	2	5	3	4
---	---	---	---	---

Questo strumento avrà una certa **risoluzione**, che indichiamo con Δx : ciò significa che una stessa indicazione Y sullo strumento si ottiene se il valore della sollecitazione in ingresso cade

nell'intervallo $\left[Y - \frac{\Delta x}{2}, Y + \frac{\Delta x}{2} \right]$. Detto in altre parole, il valore della sollecitazione in ingresso, che produce una indicazione Y sul display, può cadere con uguale probabilità in qualunque punto dell'intervallo $\left[Y - \frac{\Delta x}{2}, Y + \frac{\Delta x}{2} \right]$.

Ad esempio, se consideriamo l'indicazione

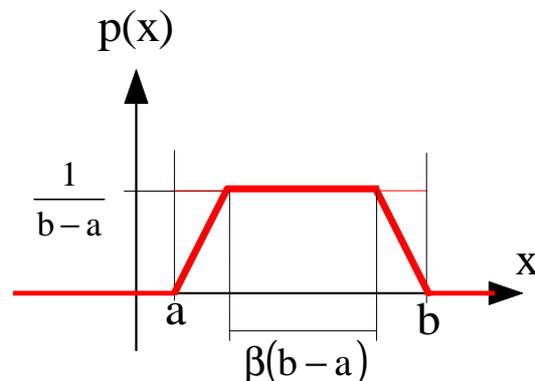
3	2	5	3	4
---	---	---	---	---

 e supponiamo che lo strumento abbia una risoluzione $\Delta x=4$, possiamo star certi che la stessa indicazione sul display si otterrebbe per una sollecitazione in ingresso che varia da 32532 ($=Y-\Delta x/2$) a 32536 ($=Y+\Delta x/2$).

In una situazione come questa, possiamo descrivere la sollecitazione in ingresso come una distribuzione uniforme, su un intervallo di ampiezza Δx , con varianza $\Delta x^2/12$, come visto prima. L'incertezza standard potrà allora essere valutata come la radice quadrata di $\Delta x^2/12$, ossia **0.29 Δx** , per qualsiasi indicazione letta sullo strumento.

Una valutazione empirica dell'incertezza si può ottenere anche incrementando la grandezza di ingresso di piccoli passi fino ad osservare una variazione Δy della lettura dello strumento; essendo Δy la più piccola variazione della grandezza di uscita ottenibile in questo modo, la si può prendere come misura dell'incertezza: quindi, la varianza sarà $\Delta y^2/12$ e l'incertezza standard sarà $0.29\Delta y$.

C'è però adesso da fare una osservazione: nella pratica, le discontinuità a scalino (come quelle di un rettangolo) in una distribuzione di probabilità hanno poco significato fisico; al contrario, è più realistico aspettarsi che i valori prossimi al valore centrale dell'intervallo siano più probabili di quelli vicini invece agli estremi. Questo si traduce nel sostituire, alla distribuzione rettangolare, una **distribuzione trapezoidale isoscele**, del tipo seguente⁶:



I lati obliqui sono uguali e l'estensione della base minore è β volte quella della base maggiore⁷, dove ovviamente $0 < \beta < 1$.

Per una simile distribuzione, si può verificare che il valore medio (cioè il valore della stima) è ancora il punto medio $\frac{a+b}{2}$ dell'intervallo, mentre la varianza standard (da usare come misura dell'incertezza) risulta essere

$$s_x^2 = \frac{(b-a)^2(1+\beta)^2}{24}$$

La scelta di una distribuzione rettangolare o trapezoidale sull'intervallo $[a,b]$ specificato presuppone che i limiti di tale intervallo siano simmetrici rispetto alla stima di X ; in caso contrario,

⁶ Notiamo che il trapezio assomiglia più o meno ad una gaussiana, con la differenza che ha un valore centrale più probabile.

⁷ Ovviamente, per $\beta=0$ il trapezio isoscele diventa un triangolo isoscele. mentre invece per $\beta=1$ riotteniamo ancora un rettangolo

se cioè non è scontata questa simmetria, allora non si può più ipotizzare una distribuzione uniforme. Si può d'altra parte riciclare la formula $s_x^2 = \frac{(b-a)^2}{12}$ per il calcolo della varianza standard: possiamo infatti porre $a = x - d_1$ e $b = x - d_2$, dove ovviamente $d_1 \neq d_2$ perché non ci sia simmetria, da cui ricaviamo che la varianza standard vale

$$s_x^2 = \frac{(b-a)^2}{12} = \frac{(d_1 + d_2)^2}{12}$$

Ulteriori considerazioni

Nei paragrafi precedenti, abbiamo esaminato due metodi di valutare l'incertezza, entrambi basati su un **modello probabilistico**. In effetti, è anche possibile usare un modello di **tipo deterministico**:

- l'uso di un modello probabilistico è spesso quello più opportuno quando sono richiesti piccoli valori di incertezza, che vengono ottenuti tramite misure ripetute (si pone infatti il problema di quante misure effettuare);
- al contrario, l'uso di un modello deterministico, che quindi usa strumenti diversi da quelli dell'analisi statistica, è adatto quando è richiesta una incertezza non troppo piccola (tipicamente, lo si usa per ottenere una stima pessimistica dell'incertezza).

La valutazione dell'incertezza tramite un modello deterministico presuppone un modello semplice, non più usato in misure di una certa complessità oppure quando si deve operare in un contesto legale o commerciale di una certa importanza.

In pratica, quando si hanno pochi contributi di incertezza di cui si tiene conto (e quindi si opera con misure poco accurate) i valori numerici non sono molto differenti da quelli ottenuti con il metodo più raffinato.

In generale, quello deterministico è un modello troppo pessimistico se si usano i risultati per eseguire altre misurazioni successive: in particolare, si adotta una stima pessimistica del contributo delle varie cause di incertezza (ogni contributo di incertezza è stimato nelle condizioni peggiori e sono sommati i valori assoluti dei singoli contributi di incertezza); questo comporta sostanzialmente che l'ampiezza della fascia di valori sia tale da garantire che ("ragionevolmente") il valore del misurando sia compreso all'interno della fascia.

RACCOMANDAZIONI SULL'INCERTEZZA

E' importante sottolineare che *le incertezze, al contrario degli errori, non possono essere classificate come sistematiche o accidentali né esiste alcuna corrispondenza tra gli errori sistematici o accidentali e le valutazioni tipo A o tipo B delle incertezze*. In altre parole, l'incertezza associata ad un errore sistematico può essere calcolata sia con una valutazione tipo A sia con una valutazione tipo B così come anche una incertezza associata ad un errore accidentale. Proprio per evitare queste possibili fonti di confusione, le componenti dell'incertezza non sono state classificate, mentre si è preferito classificare i metodi per valutare tali componenti.

Talvolta, per il calcolo dell'incertezza della stima di una quantità X, ci si avvale di valutazione sia di tipo A sia di tipo B. In questi casi, è necessario combinare le due valutazioni per fornire l'unico

valore di incertezza standard da associare alla stima effettuata. Indichiamo allora con $s_{X,A}$ e $s_{X,B}$ le incertezze standard provenienti rispettivamente da una valutazione tipo A e tipo B. Ovviamente, i quadrati di queste quantità saranno le rispettive varianze, che andranno utilizzate per ottenere il valore finale dell'incertezza standard, che è

$$s_X = \sqrt{s_{X,A}^2 + s_{X,B}^2}$$

Facciamo notare che, laddove fossero stati calcolati altri elementi di incertezza, si potrebbe introdurre anche questi all'interno della somma sotto radice.

In generale, l'incertezza può essere espressa in valore assoluto (s_X), in valore relativo (s_X/x , dove x è la stima) oppure in valore percentuale ($100s_X/x$).

INCERTEZZA STANDARD COMBINATA

Consideriamo una **misura di tipo indiretto**, ossia una misura di una grandezza Y ottenuta come funzione delle misure di altre grandezze misurabili X_1, X_2, \dots, X_N :

$$Y = f(X_1, X_2, \dots, X_N)$$

Si definisce **incertezza standard combinata** della misura di Y la radice quadrata positiva di una somma pesata di addendi, rappresentati dalle varianze o covarianze delle grandezze X_1, X_2, \dots, X_N ; i coefficienti di peso sono determinati in base alla variazione del risultato della misura al variare delle singole grandezze: quanto maggiore è l'influenza della generica grandezza X_i tanto maggiore è il coefficiente di peso della corrispondente varianza o covarianza.

L'*incertezza standard combinata* sarà indicata nel seguito con il simbolo $u_C(y)$. Il quadrato di $u_C(y)$ è detto invece **varianza standard combinata**.

Le grandezze X_1, X_2, \dots, X_N prendono il nome di **grandezze di ingresso**. La misura di ciascuna di esse è sicuramente affetta da una propria incertezza; tali incertezze si propagano perciò anche sulla misura di Y (che è la **grandezza di uscita**) e questa propagazione, come è noto, può essere studiata mediante opportune tecniche matematiche.

Cominciamo intanto a capire come possiamo stimare il misurando Y . Sia perciò y una stima del misurando Y . Tale stima può essere ottenuta da una stima di ciascuna delle grandezze di ingresso X_1, X_2, \dots, X_N : indicando infatti con x_1, x_2, \dots, x_N le stime di tali grandezze, scriveremo che

$$y = f(x_1, x_2, \dots, x_N)$$

In certi casi, si dispone di n osservazioni ripetute delle grandezze di ingresso. Si possono allora sfruttare i risultati di tali osservazioni al fine di migliorare la stima di Y . Esistono in proposito due possibilità:

- la prima è quella di calcolare n stime $y_i = f(x_{i1}, x_{i2}, \dots, x_{iN})$ del misurando Y e poi di farne la media aritmetica: la stima finale sarà

$$y = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n f(x_{i1}, x_{i2}, \dots, x_{iN})$$

- la seconda è quella di calcolare n stime $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ delle grandezze di ingresso e poi di usare tali stime per calcolare la stima di Y:

$$y = f(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$$

E' ovvio che i due modi di procedere sono assolutamente identici se la funzione f è lineare; se, invece, essa non è lineare, allora *generalmente si preferisce adottare il primo metodo, ossia quello del calcolo di y come media aritmetica delle Y_i .*

Una volta capito come stimare Y, dobbiamo capire quale sarà l'incertezza di misura associata a tale stima. In particolare l'**incertezza standard combinata $u_c(y)$** dipenderà dalle deviazioni standard stimate associate alle stime x_i delle grandezze di ingresso. A tali deviazioni standard daremo il nome di **incertezze standard** e le indicheremo con **$u(x_i)$** .

Per trovare i legami tra incertezza standard combinata e incertezze standard sulle stime delle grandezze di ingresso, dobbiamo fare un discorso simile a quello fatto per la propagazione degli errori sulle misure indirette. Le ipotesi sono dunque quelle per cui conosciamo la funzione $Y = f(X_1, X_2, \dots, X_N)$, abbiamo ottenuto delle stime x_1, x_2, \dots, x_N delle grandezze di ingresso e abbiamo valutato le corrispondenti incertezze standard $u(x_1), u(x_2), \dots, u(x_N)$.

Consideriamo la funzione $Y = f(X_1, X_2, \dots, X_N)$. Ne possiamo fare uno sviluppo in serie di Taylor intorno ai valori $\mu_i = E[x_i]$ delle aspettative delle stime delle grandezze di ingresso; se tronchiamo lo sviluppo al primo termine, possiamo confondere la differenza con il differenziale. Questo comporta che, in presenza di piccoli scostamenti delle x_i intorno al loro valor medio, si ottengono piccoli scostamenti di y intorno al valor medio μ_Y e tali scostamenti possono essere valutati nel modo seguente:

$$(y - \mu_Y)^2 = \left[\sum_{i=1}^N \frac{\partial f}{\partial x_i} (x_i - \mu_i) \right]^2$$

Sviluppando il quadrato a secondo membro, abbiamo la somma dei quadrati dei singoli termini, cui vanno aggiunti i doppi prodotti incrociati:

$$(y - \mu_Y)^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 (x_i - \mu_i)^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} (x_i - \mu_i)(x_j - \mu_j)$$

Calcoliamo il valor medio di $(y - \mu_Y)^2$, ossia la varianza di y : applicando la linearità della media, possiamo intanto scrivere che

$$\sigma_Y^2 = E[(y - \mu_Y)^2] = E \left[\sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 (x_i - \mu_i)^2 \right] + 2E \left[\sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} (x_i - \mu_i)(x_j - \mu_j) \right]$$

Continuando ad applicare la linearità della media, possiamo portare l'operatore stesso di *media* all'interno delle varie sommatorie; qui, gli unici termini aleatori (non deterministici) sono quelli del tipo $(x_i - \mu_i)$, per cui abbiamo che

$$\sigma_Y^2 = E[(y - \mu_Y)^2] = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 E[(x_i - \mu_i)^2] + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} E[(x_i - \mu_i)(x_j - \mu_j)]$$

Adesso, il termine $E[(x_i - \mu_i)^2]$ all'interno della prima sommatoria non è altro che la varianza di x_i :

$$\sigma_Y^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} E[(x_i - \mu_i)(x_j - \mu_j)]$$

Analogamente, il termine $E[(x_i - \mu_i)(x_j - \mu_j)]$ presente nell'altra sommatoria è la covarianza di x_i ed x_j ed essa può essere espressa, in funzione del coefficiente di correlazione di x_i ed x_j tramite la relazione $\text{cov}_{ij} = \rho_{ij} \sigma_i \sigma_j$.

Concludiamo perciò che la varianza della nostra stima di Y vale

$$\sigma_Y^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \rho_{ij} \sigma_i \sigma_j$$

Grandezze di ingresso scorrelate

Il caso generale che si può presentare è quello in cui le stime delle grandezze di ingresso sono tra loro correlate, per cui l'unica espressione cui fare riferimento è quella appena ricavata. Al contrario, ci si può trovare anche nel caso semplice in cui non c'è correlazione tra le suddette stime: ciò significa che $\rho_{ij} = 0$ per $\forall i$ e $\forall j$, nel quale caso la relazione di prima si semplifica banalmente nella forma

$$\sigma_Y^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2$$

Avendo detto che l'**incertezza standard combinata** $u_C(y)$ della stima y del misurando è la radice quadrata positiva della varianza standard combinata, da questa relazione scaturisce dunque che

$$u_C^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)$$

Questa relazione ci fa comprendere la definizione data inizialmente circa l'incertezza standard combinata: abbiamo infatti qui che *l'incertezza standard combinata della stima di Y è pari alla somma delle varianze delle stime delle x_i , ciascuna pesata per un coefficiente che porta in conto la variazione ∇f di y dovuta ad una variazione ∇x_i di x_i .*

Come detto, però, questo vale solo se le stime delle grandezze di ingresso sono tra loro incorrelate.

Un esempio molto semplice è quello in cui Y è la media aritmetica delle X_1, X_2, \dots, X_N : in questo caso, risulta evidentemente

$$y = f(x_1, x_2, \dots, x_N) = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\frac{\partial f}{\partial x_i} = \frac{1}{N}$$

da cui quindi consegue, nel caso di x_i mutuamente incorrelate, che

$$\sigma_Y^2 = \frac{1}{N^2} \sum_{i=1}^N \sigma_i^2$$

Se poi le X_1, X_2, \dots, X_N fossero campioni (in numero N) di una stessa popolazione ed avessero la stessa varianza (quindi $\sigma_i^2 = \sigma^2$ per $\forall i$), allora risulterebbe

$$\sigma_Y^2 = \frac{\sigma^2}{N}$$

Questa equazione non è altro che una conferma di una relazione già ottenuta in precedenza.

Si possono fare ulteriori importanti osservazioni sulla relazione

$$u_C^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i)$$

In primo luogo, ricordiamo che ciascuna incertezza standard $u(x_i)$ è ottenuta con valutazioni sia di tipo A sia di tipo B.

In secondo luogo, quella espressione vale rigorosamente quando la funzione f è lineare oppure quando la sua non linearità non è *forte*. Se invece la non linearità di f è significativa, allora è necessario considerare più termini dello sviluppo in serie di Taylor, il che porta ad includere, nell'espressione di $u_C^2(y)$, anche termini di ordine superiore.

In terzo luogo, è bene chiarire cosa indichiamo con il simbolo $\frac{\partial f}{\partial x_i}$: si tratta delle derivate parziali di f rispetto alle grandezze di ingresso, valutate in corrispondenza delle stime delle grandezze di ingresso. In formule, scriviamo cioè che

$$\frac{\partial f}{\partial x_i} = \left. \frac{\partial f}{\partial X_i} \right|_{x_1, x_2, \dots, x_N}$$

Queste derivate sono chiamate **coefficienti di sensibilità** e indicati spesso con c_i : come si vede, essi descrivono come la stima d'uscita varia al variare dei valori delle stime di ingresso.

Supponiamo ad esempio che sia $N=2$ e

$$f(X_1, X_2) = X_1 \cdot X_2^2$$

In questo caso, i coefficienti di sensibilità sono

$$c_1 = \frac{\partial f}{\partial X_1} = \frac{\partial f}{\partial X_1} \Big|_{x_1, x_2} = X_2^2 \Big|_{x_1, x_2} = x_2^2$$

$$c_2 = \frac{\partial f}{\partial X_2} = \frac{\partial f}{\partial X_2} \Big|_{x_1, x_2} = 2X_1 X_2 \Big|_{x_1, x_2} = 2x_1 x_2$$

Andando nell'espressione dell'incertezza standard combinata, scriviamo che

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) = c_1^2 u^2(x_1) + c_2^2 u^2(x_2) = x_2^2 \cdot u^2(x_1) + 2x_1 x_2 \cdot u^2(x_2)$$

Tornando invece all'espressione generale dell'incertezza standard combinata, scriviamo che

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) = \sum_{i=1}^N c_i^2 u^2(x_i) = \sum_{i=1}^N [c_i u(x_i)]^2 = \sum_{i=1}^N u_i^2(y)$$

dove abbiamo indicato con $u_i(y) = |c_i| \cdot u(x_i)$ l'incertezza standard della stima y generata dall'incertezza standard della stima x_i . Questa posizione è valida in quando, nell'ipotesi di avere una piccola variazione di x_i , cui corrisponda una variazione $(\Delta y)_i$ di y , si può scrivere con buona approssimazione che $(\Delta y)_i = c_i \Delta x_i$.

Segnaliamo inoltre che i coefficienti di sensibilità possono talvolta essere calcolati sperimentalmente: il metodo da usare è semplicemente quello di misurare la variazione prodotta su Y da una variazione di una specifica X_i , mentre le altre grandezze di ingresso vengono mantenute costanti.

Grandezze di ingresso correlate

Torniamo alla relazione generale che ci fornisce la varianza della nostra stima di Y in funzione delle varianze e delle correlazioni tra le stime x_i delle grandezze di ingresso X_i :

$$\sigma_Y^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \rho_{ij} \sigma_i \sigma_j$$

Il caso più generale possibile, come già detto in precedenza, è quello in cui esiste una correlazione tra le stime x_i : in questo caso, la doppia sommatoria che compare nella relazione può dare un contributo significativo, non più trascurabile come nel caso di stime incorrelate.

Sostituendo, al posto delle varianze e delle deviazioni standard, le corrispondenti incertezze, riscriviamo quella relazione nella forma

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \rho_{ij} u(x_i) u(x_j)$$

Abbiamo lasciato, in questa relazione, il termine ρ_{ij} (coefficiente di correlazione tra x_i ed x_j generiche), ma è ovvio che di questo coefficiente potremo avere solo una stima: la indichiamo con r_{ij} , per cui scriviamo che

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} r_{ij} u(x_i) u(x_j)$$

Naturalmente, il prodotto $r_{ij} u(x_i) u(x_j)$ non è altro che una stima della covarianza tra X_i ed X_j : la si indica generalmente con $\mathbf{u}(x_i, x_j)$ e si tratta chiaramente di una quantità commutativa. Ne parleremo tra poco.

Un caso molto particolare è quello in cui tutti i coefficienti di correlazione r_{ij} risultano pari ad 1: infatti, se $\mathbf{r}_{ij} = \mathbf{1}$, otteniamo

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i) u(x_j) = \left[\sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right) u(x_i) \right]^2$$

Ricordando che $c_i = \frac{\partial f}{\partial x_i}$, concludiamo dunque che, in questo caso particolare, risulta

$$u_c^2(y) = \left[\sum_{i=1}^N c_i u(x_i) \right]^2$$

Torniamo adesso alla stima della covarianza tra la generica X_i e la generica X_j : abbiamo detto poco fa che si tratta della quantità

$$u(x_i, x_j) = r_{ij} u(x_i) u(x_j)$$

E' possibile ottenere questa stima nel modo seguente: supponiamo di aver effettuato n osservazioni ripetute, in cui abbiamo misurato simultaneamente sia X_i sia X_j ; così facendo, abbiamo ottenuto n coppie di misure, del tipo $(\mathbf{x}_{i,k}, \mathbf{x}_{j,k})$; con tali coppie possiamo costruire la seguente stima:

$$u(x_i, x_j) = \frac{1}{n(n-1)} \sum_{k=1}^n (x_{i,k} - \bar{X}_i)(x_{j,k} - \bar{X}_j)$$

Anche per la covarianza è possibile adottare una *valutazione di tipo A* (sperimentale) oppure una *valutazione di tipo B* (teorica): in quest'ultimo caso, si tratta di sfruttare tutte le informazioni disponibili sulla variabilità correlata delle due grandezze X_i ed X_j in questione.

INCERTEZZA ESTESA

Sia $u_c(y)$ l'incertezza standard combinata associata alla stima y della grandezza incognita Y . Moltiplicando $u_c(y)$ per un fattore k (detto **fattore di copertura**), si ottiene la cosiddetta **incertezza estesa**:

$$U = k \cdot u_c(y)$$

Il significato di U è semplice: quando si fornisce il risultato di una misura (sia diretta sia indiretta), accompagnandolo con la corrispondente incertezza, si scrive

$$Y = y \pm U$$

Così facendo, si intende dire sia che y è la migliore stima del valore attribuibile al misurando Y sia che ci si aspetta che l'intervallo $[y-U, y+U]$ comprenda la maggior parte della distribuzione dei valori ragionevolmente attribuibili ad Y . Quindi, *l'incertezza estesa è quella grandezza che definisce un intervallo, intorno al risultato di una misura, che ci si aspetta contenga una frazione rilevante della distribuzione di valori ragionevolmente attribuibili ad Y .*

La scelta del fattore k , di solito compreso tra 2 e 3, è basata sul **livello di confidenza** (detto anche *probabilità di copertura*) richiesto all'intervallo. Ad esempio, se si vuole un livello di confidenza del 99%, si deve scegliere k in modo che l'intervallo $[y-U, y+U]$ abbia quel livello di confidenza.

Per fare un esempio ancora più concreto, supponiamo che la distribuzione di Y si possa approssimativamente ritenere di tipo normale e supponiamo inoltre che si voglia un livello di confidenza del 99,73%; dallo studio sulla distribuzione normale, sappiamo che l'intervallo con questo livello di confidenza è $[\mu-3\sigma, \mu+3\sigma]$, da cui deduciamo che $k=3$. Se, invece, il livello di confidenza fosse 95,45%, allora sappiamo che l'intervallo da considerare è $[\mu-2\sigma, \mu+2\sigma]$, da cui deduciamo che $k=2$. Quindi, *sommariamente, per una distribuzione normale possiamo dire che un livello di confidenza del 95% corrisponde a $k=2$, mentre un livello di confidenza del 99% corrisponde a $k=3$.*

E' ovvio che questi discorsi non valgono solo per la distribuzione normale: è cioè possibile definire un fattore di copertura per una qualsiasi distribuzione.

Prima di proseguire, osserviamo che l'espressione *livello di confidenza* sarebbe, a rigore, appropriato solo se le incertezze fossero ottenute con valutazioni di tipo A, dato che solo queste ultime si basano sui risultati di osservazioni ripetute. Tuttavia, dato che la definizione di incertezza estesa non fa riferimento ad uno specifico criterio di valutazione dell'incertezza, sarebbe più corretto parlare di **grado di confidenza**.

Quindi, riepilogando, *l'incertezza estesa definisce, intorno al risultato della misura, un intervallo che comprende una gran parte p della distribuzione di probabilità caratterizzata dal risultato stesso e dalla sua incertezza standard combinata, dove p è appunto il grado di confidenza.*

Notiamo inoltre che il valore del fattore di copertura va sempre specificato, in modo che da U si possa sempre risalire ad $u_c(y)$.

Livelli ed intervalli di confidenza

Continuando il discorso sull'incertezza estesa, vogliamo qui fare alcune considerazioni circa alcune situazioni che si possono presentare.

Ricordiamo che stiamo considerando sempre un misurando Y sia che funzione di N distinte (in generale correlate) grandezze di ingresso:

$$Y = f(X_1, X_2, \dots, X_N)$$

Supponiamo che le distribuzioni di tali grandezze di ingresso siano note: in particolare, servono le loro aspettative, le varianze ed i momenti di ordine superiore. Supponiamo inoltre che la funzione f sia lineare. Sotto queste condizioni, si può dimostrare che la distribuzione di probabilità di Y è ottenibile come convoluzione delle distribuzioni di probabilità delle X_i . Quindi, note le distribuzioni delle X_i , si può ottenere la distribuzione di Y e quindi valutare il fattore di copertura k necessario ad ottenere il livello di confidenza desiderato. La cosa è invece più complicata nel caso che f sia una funzione non lineare.

Molto spesso, non è richiesto un elevato grado di accuratezza per il calcolo di k ; di conseguenza, si può talvolta provare a semplificare la determinazione della distribuzione di Y , ricorrendo al cosiddetto **teorema del limite centrale**: esso afferma che è possibile approssimare la distribuzione di Y ad una distribuzione normale, anche quando le distribuzioni delle singole X_i non siano a loro volta perfettamente normali⁸. Si tratta evidentemente di valutare quando l'approssimazione risulti buona. A tal proposito, il teorema dice che risulta determinante il valore delle varianze σ_i^2 delle grandezze di ingresso: tali varianze, pesate per i rispettivi coefficienti di sensibilità, devono essere tutte molto più piccole della varianza effettiva di Y e nessuna di esse deve prevalere sulle altre. Il teorema dice inoltre che, in generale, l'approssimazione è tanto migliore quanto maggiore è il numero N di grandezze di ingresso di cui Y è funzione.

Per avere una idea un po' più quantitativa di quanto appena detto, supponiamo che risulti $N=3$:

$$Y = f(X_1, X_2, X_3)$$

Supponiamo inoltre che le tre grandezze di ingresso abbiano tutte una distribuzione rettangolare⁹ e che tale distribuzione sia uguale per tutte e tre. Indichiamo con C la semiampiezza ($(b-a)/2$) dell'intervallo su cui tale distribuzione è definita, il che comporta, come è noto, che la varianza sia $\sigma_i^2 = C^2/3$.

Effettuando la convoluzione delle tre distribuzioni, si ottiene la distribuzione di Y , che risulta avere una varianza $\sigma_Y^2 = C^2$. Andiamo allora a confrontare i livelli di confidenza della reale distribuzione di X con quelli di una distribuzione normale:

- si trova, ad esempio, che un livello di confidenza del 95% si ottiene in corrispondenza dell'intervallo di confidenza $\mu_Y \pm 1,937\sigma_Y$, quando invece per una distribuzione normale sarebbe $\mu \pm 1,960\sigma$;

⁸ E' noto che la somma di variabili casuali gaussiane è a sua volta una variabile gaussiana. Il teorema dice allora che la risultante della somma tende ad essere gaussiana anche se i singoli addendi non sono perfettamente gaussiani

⁹ La distribuzione rettangolare (o uniforme) è quanto di più lontano ci possa essere da una distribuzione gaussiana

- si trova, inoltre, che un livello di confidenza del 99% si ottiene in corrispondenza dell'intervallo di confidenza $\mu_Y \pm 2,379\sigma_Y$, quando invece per una distribuzione normale sarebbe $\mu \pm 2,576\sigma$.

Come si nota subito, i valori effettivi (cioè basati sulla reale distribuzione di Y) ottenuti per k sono molto simili a quelli ottenibili ipotizzando una distribuzione normale, il che testimonia la veridicità del teorema del limite centrale: *pur considerando distribuzioni profondamente diverse da quella normale e pur considerando poche (appena 3) di queste distribuzioni, abbiamo ottenuto valori del k molto vicini a quelli che si otterrebbero considerando per Y una distribuzione normale.*

Una conseguenza pratica del teorema del limite centrale è la seguente: supponiamo di aver calcolato l'incertezza standard combinata associata alla nostra misura di Y e supponiamo di aver riscontrato che nessuna delle incertezze delle singole grandezze di ingresso abbia dato un contributo significativo¹⁰. Sotto queste condizioni, possiamo applicare il teorema del limite centrale e possiamo andare a calcolare il fattore di copertura ed il livello di confidenza assumendo direttamente una distribuzione normale per Y .

Uso della distribuzione t

Allo scopo di ottenere una migliore approssimazione nel calcolo del fattore di copertura k , si può ricorrere alla *distribuzione t* invece che ad una distribuzione normale.

Il principio di base da applicare è il seguente: si può dimostrare che, se la variabile Y ha distribuzione normale, la nuova variabile

$$\frac{y - Y}{u_c(y)}$$

ha **distribuzione t** , dove ricordiamo ancora che y è la nostra stima di Y e dove $u_c(y)$ è l'incertezza standard combinata ad essa associata, che, nell'ipotesi di grandezze di ingresso scorrelate, è data

dall'equazione $u_c^2(y) = \sum_{i=1}^N u_i^2(y)$.

Si tratta perciò di sfruttare la conoscenza che noi abbiamo della *distribuzione t* . Per comprendere il concetto, consideriamo dapprima un caso particolare e cioè quello in cui Y è una singola grandezza con distribuzione normale. Supponiamo di aver i risultati Y_i di n osservazioni successive di Y , per cui sappiamo che la migliore stima che possiamo dare di questa variabile è la media aritmetica di tali risultati:

$$y = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

Sappiamo anche che la *deviazione standard sperimentale* associata a questa stima è data da

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

¹⁰ Ad esempio, questo accade quando, compiendo valutazioni di tipo A, si hanno a disposizione poche prove ripetute, oppure, compiendo valutazioni di tipo B, si ipotizza una distribuzione di tipo rettangolare

per cui questa quantità coincide, nel nostro caso, con $u_c(y)$:

$$u_c(y) = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Andiamo adesso a costruire la nuova variabile $t = \frac{y - Y}{u_c(y)}$. Avendo supposto che Y abbia distribuzione normale, si può dimostrare che questa nuova variabile ha distribuzione t (o distribuzione di Student), il che significa che si tratta di una variabile continua la cui densità di probabilità è

$$p(t) = \frac{1}{\sqrt{\pi v}} \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$

dove Γ è la **funzione gamma** mentre $v > 0$ rappresenta i **gradi di libertà**.

Nel nostro caso, essendo y data da una media aritmetica di n risultati, il numero di gradi di libertà della distribuzione è $v = n - 1$, ma, come vedremo tra poco, il numero di gradi di libertà va determinato caso per caso.

La variabile t ha una probabilità di occorrenza

$$p = P(-t_p(v) \leq t \leq t_p(v))$$

ossia p è la probabilità che t cada nell'intervallo $[-t_p(v), +t_p(v)]$ ⁽¹¹⁾.

Sostituendo, al posto di t , la sua espressione completa, abbiamo che

$$p = P\left(-t_p(v) \leq \frac{y - Y}{u_c(y)} \leq t_p(v)\right)$$

e questa espressione equivale anche a

$$p = P(-y - t_p(v)u_c(y) \leq -Y \leq t_p(v)u_c(y) - y)$$

ossia anche a

$$p = P(y - t_p(v)u_c(y) \leq Y \leq y + t_p(v)u_c(y))$$

Quindi, **p** è la probabilità che Y cada nell'intervallo $[y - t_p(v)u_c(y), y + t_p(v)u_c(y)]$, il che significa, essendo y la stima che noi abbiamo attribuito ad Y , che l'incertezza estesa associata alla nostra stima è

$$U = ku_c(y) = t_p(v)u_c(y)$$

¹¹ Ricordiamo che la distribuzione di Student è a media nulla.

ossia infine che $t_p(v)$ è il nostro fattore di copertura (che risulta funzione del numero v di gradi di libertà della distribuzione).

Questo discorso vale però nel caso particolare in cui Y non è funzione di alcuna grandezza di ingresso. In caso contrario, subentra il problema per cui la variabile $t = \frac{y - \bar{Y}}{u_c(y)}$ non è distribuita

secondo Student, anche se le grandezze di ingresso sono poche e distribuite normalmente. Per aggirare l'ostacolo, però, è sufficiente considerare un numero di gradi di libertà dato dalla seguente relazione:

$$v_{\text{eff}} = \frac{u_c^4(y)}{\sum_{i=1}^N \frac{u_i^4(y)}{v_i}}$$

dove abbiamo indicato con v_i il numero di gradi di libertà delle N grandezze di ingresso di cui Y è funzione.

Usando quella equazione (ottenuta da **Welch e Satterthwaite**), la variabile $t = \frac{y - \bar{Y}}{u_c(y)}$ è effettivamente distribuita secondo Student e quindi si può continuare ad applicare il discorso fatto prima. Si tratta cioè di calcolare il valore del fattore di copertura $t_p(v)$ in funzione del numero v_{eff} di gradi di libertà e del livello di confidenza p che si desidera. A tal fine, si utilizzano tabelle del tipo seguente:

v_{eff}	$p = 68,27\%$	$p = 9\%$	$p = 95\%$	$p = 99,45\%$	$p = 99\%$	$p = 99,73\%$
1	1,84	6,31	12,71	13,97	63,66	235,8
2	1,32	2,92	4,3	4,53	9,92	19,21
3	1,2	2,35	3,18	3,31	5,84	9,22
4	1,14	2,13	2,78	2,87	4,6	6,62
5	1,11	2,02	2,57	2,65	4,03	5,51
6	1,09	1,94	2,45	2,52	3,71	4,9
7	1,08	1,89	2,36	2,43	3,5	4,53
8	1,07	1,89	2,36	2,43	3,5	4,53
9	1,06	1,83	2,26	2,32	3,25	4,09

Fissati v_{eff} e p , la tabella fornisce il corrispondente valore del fattore di copertura $t_p(v)$ e quindi la possibilità di calcolare l'incertezza estesa.

PRESENTAZIONE DEI RISULTATI

Il **risultato** di una misura va inteso sempre come una *approssimazione* (o **stima**) del valore del misurando ed è quindi completo solo quando è accompagnato dall'indicazione dell'incertezza di quella stima.

Nella presentazione del risultato di una misura, le cifre riportate devono contenere tutte e sole le informazioni che possono essere correttamente utilizzate. E' compito dell'operatore escludere invece le cifre che non contengono utili indicazioni e che quindi sarebbero solo da appesantimento per una eventuale successiva elaborazione. Ad esempio, non ha molto senso scrivere che una certa tensione è

$$V = 15.12215768234 \pm 0.54 \text{ V}$$

dato che la *precisione* con cui è fornita la misura (cioè sostanzialmente il numero di cifre significative) è evidentemente sproporzionata con l'entità dell'incertezza.

L'operatore è chiaramente la persona più qualificata per stabilire l'**accuratezza** del risultato ottenuto, ovvero quanto esso si avvicini al valore del misurando.

Il risultato Y della misura è generalmente scritto come somma e differenza di due grandezze:

$$Y = y \pm U$$

In questa relazione, y è il valore della **stima** attribuita al misurando e corrisponde perciò al valore centrale della **fascia di incertezza** del misurando; U è invece ciò che abbiamo definito **incertezza estesa** e corrispondente alla semiampiezza della suddetta *fascia*.

L'incertezza estesa U è riportata con una o preferibilmente due cifre significative. Ad esempio, consideriamo l'esito di una misura di tensione e supponiamo di aver calcolato una incertezza di 0.5367mV associata ad una stima di 15.1221mV; dovremmo allora scrivere

$$V = 15.1221 \pm 0.5367 \text{ mV}$$

ma questa è una scrittura poco elegante da riportare. Al contrario, più elegante e altrettanto corretta è la seguente scrittura:

$$V = 15.1221 \pm 0.54 \text{ mV}$$

nella quale abbiamo compiuto un arrotondamento per eccesso (se ne parlerà più avanti).

In alternativa, potremmo anche dire, con una leggera sopravvalutazione, che l'incertezza è di 0.6mV oppure, al contrario, con una leggera sottovalutazione, che l'incertezza è di 0.5mV. Nella pratica, si approssima per eccesso (per mettersi nel caso peggiore) e, in presenza di cifre decimali, si preferisce usarne una sola, specialmente quando questa è alta (ad esempio 0.8 oppure 0.9)¹².

 Oltre a questo, la scrittura $V = 15.1221 \pm 0.54 \text{ mV}$ va corretta tenendo conto delle cifre su cui cade l'incertezza. In altre parole, l'ultima cifra significativa da portare in conto nell'indicazione del risultato della misura è quella su cui cade l'incertezza di misura: nel nostro caso, tale incertezza cade sui centesimi di mV, per cui la scrittura più corretta è

$$V = 15.12 \pm 0.54 \text{ mV}$$

Naturalmente, valgono le classiche regole dell'arrotondamento: se la prima cifra da scartare è <5, allora la cifra precedente si lascia inalterata; se la prima cifra da scartare è >5, la cifra precedente si incrementa di 1; se la prima cifra da scartare è =5 ed è seguita da cifre diverse da zero, la cifra precedente si incrementa di 1; se la prima cifra da scartare è =5 ed è seguita da zeri oppure non è seguita da altre cifre, la cifra precedente va arrotondata al numero pari più vicini.

Accanto a queste regole, si può aggiungere quella per cui, nel caso specifico di misura diretta, è lecito troncare il risultato (sempre alla cifra su cui cade l'incertezza).

¹² In alcune situazioni, i risultati di una misura sono espressi usando sia numeri sia simboli. Ad esempio, se si intende dire che due sinusoidi sono in opposizione di fase a meno di una piccola tolleranza e si è in un contesto in cui si usano i radianti, si può scrivere, per esempio, che lo sfasamento è $\varphi = \pi \pm (\pi/100)$, dove quindi l'incertezza è stata valutata come $\pi/100$, che è una quantità che, se esplicitata, possiede infinite cifre significative. Se invece si volesse esprimere il tutto in gradi, allora si dovrebbe scrivere $\varphi = 180^\circ \pm 1.8^\circ$.

Eventualmente, al posto dell'incertezza assoluta U_{ass} , è possibile anche esprimere quella relativa percentuale, che vale $U_{\text{rel}\%} = \frac{U_{\text{ass}}}{X_m} 100$, dove X_m è il valore misurato. Nell'esempio di poco fa, in cui $U_{\text{ass}}=0.54$ e $X_m=15.1221$, abbiamo dunque che $U_{\text{rel}\%} = \frac{0.54}{15.1221} 100 = 3.57\%$, per cui possiamo esprimere il risultato della misura nella forma

$$V = 15.1 \pm 3.57\% \text{ mV}$$

Ad ogni modo, è talvolta necessario conservare anche più di due cifre, in modo da evitare errori di arrotondamento nei calcoli: infatti, i coefficienti di correlazione dovrebbero essere scritti con 3 cifre significative quando i loro valori assoluti sono prossimi ad 1 e quindi è necessario disporre del valore di U con un numero opportuno di cifre.

Sempre riguardo il modo con cui fornire il risultato di una misura, la **Guida** suggerisce di indicare anche le seguenti informazioni:

- una descrizione completa di come è definito il misurando Y ;
- il risultato della misura nella forma $Y = y \pm U$ insieme con le unità di misura di y ed Y ;
- l'**incertezza estesa relativa**, definita come $U/|y|$, ovviamente quando $|y| \neq 0$;
- il grado di confidenza approssimato associato all'intervallo $y \pm U$, specificando il modo in cui esso è stato ottenuto;
- qualora il metodo di misura sia di tipo indiretto, bisogna anche fornire le seguenti informazioni:
 - * i valori di ciascuna stima x_i delle grandezze di ingresso e della corrispondente incertezza standard $u(x_i)$, insieme con la descrizione del modo con cui queste quantità sono state ottenute;
 - * le *covarianze stimate* (oppure i *coefficienti di correlazione stimati*) associati a tutte le stime di ingresso¹³ ed i metodi seguiti per ottenerle;
 - * i gradi di libertà per l'incertezza standard di ciascuna stima di ingresso ed il metodo con il quale sono stati calcolati;
 - * la relazione funzionale $y=f(x_1, x_2, \dots, x_N)$ tra grandezza di uscita e grandezze di ingresso e, quando opportuno, i coefficienti di sensibilità (in particolare, è obbligatorio riportare tali coefficienti quando essi sono stati valutati sperimentalmente). Se la relazione funzionale f è stata ottenuta mediante un programma di calcolo, essa va descritta in termini generali oppure citando il programma usato, chiarendo anche come è stata calcolata la stima y e la sua incertezza.

Per chiarire bene questi concetti, la Guida stessa fornisce un esempio pratico. Si considera la misura di una massa, il cui risultato è fornito nella seguente forma:

$$m = 100,02147 \pm (0,00079) \text{ g}$$

¹³ Ricordiamo che le stime delle grandezze di ingresso possono essere correlate o meno, il che è mostrato da un valore nullo o non nullo dei rispettivi coefficienti di correlazione.

Il numero tra parentesi è l'incertezza estesa $U=ku_c(y)$, calcolata evidentemente tramite il prodotto di un fattore di copertura k ($=2,26$) con una incertezza standard combinata $u_c(y)$ ($=0,35\text{mg}$). Viene anche specificato che il fattore di copertura è stato ricavato usando una distribuzione di Student con $v=9$ gradi di libertà e che U definisce un intervallo che si stima avere un grado di confidenza del 95%.

Rappresentazione mediante tabelle

Un modo molto utilizzato per la presentazione delle misure eseguite e dei risultati ottenuti è quello delle **tabelle**, per il quale è opportuno seguire regole precise:

- la tabella deve avere un titolo esplicativo dei dati presentati;
- ogni colonna contenente cifre dovrebbe far riferimento a dei risultati (ottenuti da misure dirette o indirette) associata ad una sola quantità; la colonna dovrà inoltre riportare, in testa, un titolo che consenta di identificare i dati in essa contenuti; è importante inoltre riportare, sotto il titolo della colonna, l'unità di misura dei dati, racchiusa in parentesi tonde o quadre;
- il titolo della tabella ed i titoli delle colonne vanno separati tra loro e dai dati numerici mediante linee orizzontali;
- è buona norma, infine, associare ad ogni colonna di dati le incertezze da cui essi sono affetti, con la loro denominazione; per fare questo, si può utilizzare una colonna aggiuntiva oppure si riporterà, in coda alla singola colonna, solo il valore massimo della fascia di incertezza.

Rappresentazione mediante grafici

Un altro modo possibile per la presentazione delle misure eseguite e dei risultati ottenuti è quello dei **grafici**. Anche se si corre il rischio di perdere informazioni dettagliate circa l'accuratezza dei risultati, i grafici presentano diversi sicuri vantaggi:

- in primo luogo, essi forniscono una rappresentazione visiva dei risultati, che spesso consente una comprensione più immediata del fenomeno fisico;
- in secondo luogo, è possibile verificare, in modo immediato e sintetico, il tipo di relazione esistente tra due variabili.

Anche per l'uso di grafici è opportuno seguire regole precise:

- il grafico deve innanzi tutto contenere indicazioni, attraverso una intestazione o un titolo o una didascalia, sul tipo di misure eseguite, sulla apparecchiatura in prova e sulle condizioni nelle quali sono state eseguite le misure;
- le curve vanno tracciate con cura, scegliendo in modo opportuno le scale;
- su uno steso grafico si possono tracciare più curve, nel qual caso su ciascuna va indicata sinteticamente la grandezza rappresentata;
- sia in ordinate sia in ascisse bisogna indicare, attraverso il simbolo racchiuso in parentesi tonde o quadre, le unità di misura;

- la scelta delle scale, sia in ascisse sia in ordinate, è di cruciale importanza:
 - * essa deve essere fatta in modo da rendere facile la lettura dei valori relativi ad un determinato punto della curva (nel senso che la grandezza deve essere letta direttamente, senza dover fare alcuna operazione di moltiplicazione e/o divisione);
 - * la scala deve inoltre consentire la lettura dei dati con la approssimazione derivante dall'incertezza di misura associata a quella grandezza. Ad esempio, se i risultati di una misura sono noti con una incertezza dell' 1%, è sbagliato scegliere una scala che consenta una lettura con una approssimazione intorno a frazioni dell' 1%;
 - * la scala non va riportata a margine del grafico, ma deve essere comprensibile dalla lettura dei valori numerici riportati sugli assi; questi valori numerici devono essere possibilmente interi ed equispaziati sull'asse;
- importante è anche il criterio con cui rappresentare i punti sperimentali:
 - * i punti sperimentali vanno evidenziati sul grafico, senza l'indicazione dei valori delle coordinate, e, nel caso si riportino più curve, i punti relativi a ciascuna di essi vanno distinti usando simboli diversi;
 - * dato che i fenomeni macroscopici sono notoriamente continui, i punti sperimentali vanno raccordati con curve continue, salvo in quei casi in cui si ha a che fare con caratteristiche intrinsecamente discontinue, nel qual caso si possono usare delle spezzate o degli istogrammi;
 - * quando il raccordo dei diversi punti risulta difficoltoso, è necessaria una operazione di interpolazione, che può essere compiuta ad occhio oppure con particolari algoritmi di cui si parlerà in seguito¹⁴.

CIFRE SIGNIFICATIVE E SCRITTURA DELL'INCERTEZZA

A proposito delle cifre significative da riportare nella presentazione di un risultato, esistono delle precise convenzioni:

- in primo luogo, normalmente, l'ultima cifra del risultato riportato è quella su cui ricade l'incertezza della misura;
- inoltre, nell'arrotondamento di un numero, l'ultima cifra che si conserva è aumentata di 1 se la prima cifra eliminata è maggiore o uguale a 5 (**arrotondamento per eccesso**), mentre invece l'ultima cifra viene lasciata invariata quando la prima cifra eliminata è minore di 5 (**arrotondamento per difetto**); ad ogni modo, ci sono alcuni casi in cui, nella presentazione dell'incertezza, è possibile effettuare un arrotondamento per eccesso anche quando la prima cifra eliminata è minore di 5;
- un'altra convenzione correntemente accettata riguarda gli zeri significativi: si è convenuto di non riportare, alla destra delle cifre significative, zeri che non sia significativi. Ad esempio, il numero 512000 va scritto come $512 \cdot 10^3$

¹⁴ A tal proposito, è importante sottolineare che ogni linea usata per raccordare i punti sperimentali è accettabile, dal punto di vista matematico, come una rappresentazione grafica del fenomeno fisico in esame solo se rispetta una precisa condizione: la massima deviazione dei punti sperimentali dalla curva deve essere inferiore ai valori limite delle possibili incertezze di misura.

Concentriamoci, per un momento, sul problema dell'**arrotondamento**. In precedenza, abbiamo detto che non avrebbe senso riportare, come risultato della misura ad esempio di una tensione, la seguente scrittura:

$$V = 15.12215768234 \pm 0.54 \text{ V}$$

Dato che l'incertezza cade sulla prima cifra a destra della virgola, quel numero con tutte quelle cifre significative non ha senso. Dobbiamo operare un *arrotondamento*, scrivendo ad esempio

$$V = 15.12 \pm 0.54 \text{ V}$$

D'altra parte, il valore $v = 15.12215768234$ è pur sempre il risultato di una misura, ossia la migliore stima del valore vero che stiamo stati in grado di ottenere. Di conseguenza, l'aver troncato tale numero equivale ad aver introdotto una **quantizzazione**, che in questo caso ha un **passo di quantizzazione** $Q = 0.01$. Questo vuol dire comunque peggiore la misura e aumentare l'incertezza.

Di conseguenza, *nel compiere un arrotondamento, dobbiamo sempre fare in modo che l'incertezza aggiunta dall'arrotondamento sia piccola rispetto a quella intrinseca del processo o strumento di misura.*

Per comprendere questo concetto, ricordiamo che l'incertezza va interpretata come un qualsiasi numero maggiorante dell'errore assoluto associato alla misura: se V è il valore vero del misurando ed X è la stima che ne abbiamo fatto, l'errore assoluto è $E = V - X$, mentre l'incertezza U è un qualsiasi numero positivo U tale che $|E| \leq U$. Ovviamente, la valutazione che abbiamo fatto dell'incertezza è tanto migliore quanto più stretta è la maggiorazione espressa da $|E| \leq U$. Indichiamo adesso con X_q il valore troncato (quantizzato) del risultato X della misura (sarebbero i 15.12 V dell'esempio di prima). L'errore assoluto che commettiamo con X_q è dato, in modulo, da

$$|E_q| = |X_q - V| = |X_q - X + X - V| \leq |X_q - X| + |X - V|$$

Il termine $|X_q - X|$ è l'entità dell'arrotondamento effettuato ed è quindi maggiorato dalla metà del passo di quantizzazione Q ; il termine $|X - V|$ è invece, per definizione, il modulo dell'errore assoluto, che è maggiorato (sempre per definizione) dall'incertezza U . Possiamo allora scrivere che

$$|E_q| \leq |X_q - X| + |X - V| \leq \frac{Q}{2} + U$$

In base a questa relazione, la quantità $\frac{Q}{2} + U$ è un maggiorante dell'errore assoluto E_q che commettiamo usando X_q come risultato della misura, per cui si tratta dell'incertezza U_q associata a questa misura:

$$U_q = \frac{Q}{2} + U$$

Riepilogando, quindi, usando X_q al posto di X come risultato della misura, siamo passati dall'incertezza U all'incertezza $U_q = \frac{Q}{2} + U$, ossia abbiamo aumentato l'incertezza di $Q/2$:

$$\text{risultato della misura} = X_q \pm U_q$$

Nell'esempio di prima, dove l'incertezza iniziale era $U=0.54$, il troncamento (con passo $Q=0.01$) porta dunque un aumento dell'incertezza di $0.005 (=Q/2)$. Dovremmo allora scrivere

$$V = 15.12 \pm 0.545 \text{ V}$$

In effetti, però, sappiamo che, quando si parla di incertezza, non ha molto senso distinguere tra 0.54 e 0.545 , per cui è comunque sensato scrivere

$$V = 15.12 \pm 0.54 \text{ V}$$

Questo proprio perché l'arrotondamento è stato fatto in modo che il conseguente incremento di incertezza ($+0.005$) fosse comunque sufficientemente più piccolo dell'incertezza di partenza ($=0.54$).

Come regola pratica (che quindi non è un dogma), diciamo che è opportuno scrivere il risultato della misura con una cifra in più rispetto alla prima cifra significativa dell'incertezza.

Esempio: misura (indiretta) di resistenza

Facciamo un altro esempio concreto per mettere in pratica quanto appena detto ed anche per comprendere alcuni concetti forniti in precedenza.

Supponiamo di voler misurare il valore di una resistenza, tramite una misura indiretta: andiamo perciò a misurare la tensione ai capi della resistenza e la corrente nella resistenza stessa. Supponiamo di aver trovato $V=2\text{V}$ e $I=3\text{mA}$; la resistenza misurata è allora

$$R = \frac{2\text{V}}{3\text{mA}} = \frac{2}{3} \text{k}\Omega = 666.666666\dots\Omega$$

Supponiamo inoltre che ciascuna delle due misure dirette (supposte indipendenti) sia affetta da una *incertezza relativa percentuale* dell' 1%. Avendo a che fare con una misura indiretta ottenuta come rapporto tra misure dirette, sappiamo che l'incertezza relativa $u=U/R$ è data dalla somma delle incertezze relative (regola del caso peggiore), per cui è pari al 2%. La conseguente incertezza è

$$U = u \cdot R = 0.02 \cdot 666.666666\dots = 13.3333333\dots\Omega$$

Siamo dunque in una situazione molto particolare, nella quale sia l'esito della misura sia l'incertezza ad esso associata sono quantità ad infinite cifre decimali. Ci chiediamo come esprimere la misura.

Abbiamo varie possibilità. La prima è quella di arrotondare la misura alla prima cifra significativa dell'incertezza, cioè in questo caso arrotondare sulle decine: dovremmo perciò considerare il valore **670 Ω** . In questo caso, il passo di quantizzazione è $Q=6.6\Omega$, per cui l'arrotondamento aumenta l'incertezza di 3.3Ω . Dovremmo perciò scrivere

$$R = 670 \pm 20\Omega$$

dove ovviamente abbiamo arrotondato (per eccesso) anche l'incertezza, che a rigore sarebbe $16.66\dots\Omega$.

L'altra possibilità è quella di arrotondare semplicemente per eccesso, prendendo **667 Ω** . In questo caso, il passo di quantizzazione è $Q=0.66\Omega$, per cui l'arrotondamento aumenta l'incertezza di 0.33Ω , che è un aumento decisamente minore del precedente. Quest'ultima è dunque la scelta necessaria a scrivere la misura con maggiore precisione:

$$R = 667 \pm 14 \Omega$$

dove ancora una volta abbiamo arrotondato anche l'incertezza (che sarebbe a rigore $13.666\dots\Omega$).

Ci sono invece dei modi sbagliati di riportare l'esito della misura:

$$R = 670 \pm 10 \Omega$$

$$R = 667 \pm 13 \Omega$$

Nel primo caso, abbiamo arbitrariamente abbassato l'incertezza da 20Ω a 10Ω ; nel secondo caso, abbiamo abbassato l'incertezza da 14Ω a 13Ω .

In generale, possiamo dire che *sono completamente sbagliate le forme che forniscono una fascia di incertezza troppo stretta.*

Altrettanto sbagliate, oltre che poco corrette, sarebbero inoltre le seguenti scritte:

$$R = 667 \pm 20 \Omega$$

$$R = 670 \pm 17 \Omega$$

Nel primo caso abbiamo inutilmente sopravvalutato l'incertezza, che abbiamo visto essere di 14Ω ; nel secondo caso abbiamo ancora una volta ridotto l'incertezza.

Ci sono ancora altre possibilità, ridicole prima ancora che sbagliate:

$$R = 670 \pm 16.67 \Omega$$

$$R = 667 \pm 13.67 \Omega$$

$$R = 666.7 \pm 13.37 \Omega$$

Queste scritte darebbero l'impressione di aver confuso l'incertezza con l'errore, quando invece (lo ricordiamo ancora una volta) l'incertezza è solo una maggiorazione dell'errore, per cui non è l'errore.

C'è però da osservare, come peraltro già fatto in precedenza, che, dovendo utilizzare il risultato di una misura per ulteriori calcoli, è opportuno utilizzare tutte le cifre ricavate, fino alla precisione del mezzo di calcolo a disposizione. Nel nostro esempio, questo significa sostanzialmente portarsi dietro, come misura di R , direttamente il valore $2/3$.

Continuiamo con l'esempio della resistenza, supponendo questa volta le incertezze relative su tensione e corrente siano dello 0.5% in un caso e dello 0.7% in un altro caso.

Cominciamo da una incertezza relativa dello 0.5% , per cui (sommando) l'incertezza relativa sulla misura di R è dell' 1% , cui corrisponde un valore dell'incertezza (prima della quantizzazione) di $U=6.666666\dots\Omega$. Con passaggi analoghi a prima, scriviamo quanto segue:

$$R = 667 \Omega \rightarrow U_q = 7.000\dots \Omega \rightarrow R = 667 \pm 7 \Omega$$

$$R = 666.7 \Omega \rightarrow U_q = 6.700\dots \Omega \rightarrow R = 666.7 \pm 6.7 \Omega$$

Se invece consideriamo incertezze relative dello 0.7% sulle due misure dirette, otteniamo una incertezza relativa complessiva dell' 1.4% e quindi una incertezza (prima della quantizzazione) $U=9.333333\dots\Omega$. Abbiamo allora le solite due possibilità:

$$R = 667 \Omega \rightarrow U_q = 9.666\dots \Omega \rightarrow R = 667 \pm 10 \Omega$$

$$R = 666.7 \Omega \rightarrow U_q = 9.366\dots \Omega \rightarrow R = 666.7 \pm 9.4 \Omega$$

Si noti che, nei due casi appena considerati, l'incertezza è dell'ordine delle unità, per cui bisogna riportare il risultato almeno fino all'unità. Scrivere $R = 670 \Omega$ sarebbe scorretto.

Possiamo dunque trarre due conclusioni fondamentali:

- bisogna riportare l'incertezza con una o due cifre significative (approssimando per eccesso) e la misura fino all'ultima cifra decimale dell'incertezza (approssimando per arrotondamento);
- la scelta di una o due cifre dipende dalle circostanze (ma usare due cifre non è mai “sbagliato”, al limite è “fuori luogo”).

A questo punto, ci si può porre la seguente questione: in alcuni casi (ad esempio sui cataloghi), l'incertezza non viene riportata perché non interessa; si tratta perciò di scegliere quante cifre usare per riportare il valore numerico di una misura; la “tradizione” vuole che si usi semplicemente l'arrotondamento alla prima cifra significativa dell'incertezza. Quindi, nel caso della resistenza da $2/3 \text{ k}\Omega$, dovremmo scrivere quanto segue:

se $U=2\% \rightarrow R \approx 670\Omega$

se $U=1\%$ o 1.4% ecc. $\rightarrow R \approx 667\Omega$

Tuttavia, si potrebbe obiettare che, nel primo caso, non è chiaro se lo zero finale è significativo oppure no: in base a quanto visto, potrebbe infatti essere che si ha una incertezza dell'ordine delle decine con una misura del tipo $R=666.66\dots\Omega$, ma potrebbe anche essere che si ha una incertezza dell'ordine delle unità con una misura del tipo $R=670.567\dots\Omega$. Per evitare queste incomprensioni, si consiglia, ove possibile, di usare i multipli del Sistema Internazionale quando l'incertezza è di ordine maggiore delle unità (ossia sulle decine, sulle centinaia e così via). Si avrà quindi:

$R \approx 0.67 \text{ k}\Omega$ se $U = 2\%$, cioè dell'ordine delle decine di Ω ;

$R \approx 0.7 \text{ k}\Omega$ se U è dell'ordine delle centinaia di Ω .

I “puristi” consigliano anche l'uso delle potenze di dieci: $R=0.67 \times 10^3 \text{ W}$. Ciò però ha poco senso in un catalogo, una pubblicità, o un qualunque documento di carattere non specialistico, quali sono quelli in cui non si riporta l'incertezza.

E' invece consigliabile non “buttare via” gli zeri decimali nel risultato di una misura se la prima cifra significativa dell'incertezza cade a destra di questi. Ciò serve a sottolineare che quegli zeri non derivano da un arrotondamento ma sono parte del risultato.

E' bene perciò scrivere, per una misura pari a $8140.006\dots \Omega$:

$R \approx 8 \text{ k}\Omega$ (se U è dell'ordine delle migliaia di ohm)

$R \approx 8.1 \text{ k}\Omega$ (se U è dell'ordine delle centinaia di ohm)

$R \approx 8.14 \text{ k}\Omega$ (se U è dell'ordine delle decine di ohm)

$R \approx 8140 \Omega$ (se U è dell'ordine degli ohm)

$R \approx 8140.0 \Omega$ (se U è dell'ordine dei decimi di ohm)

$R \approx 8140.00 \Omega$ (se U è dell'ordine dei centesimi di ohm)

ELIMINAZIONE DI DATI

Supponiamo di compiere una serie di misure ripetute di una stessa grandezza X . Può capitare che un risultato X_k sia molto differente dagli altri, per cui sorge il problema se utilizzarlo oppure scartarlo. La decisione è ovviamente facile se ci si rende conto subito delle cause di questa differenza, mentre, in caso contrario, risulta problematico decidere, proprio perché non esiste alcuna ragione plausibile che induca a ritenere X_k scartabile.

E' chiaro che l'utilizzazione del dato incerto può incidere in modo anche significativo sul valore della media e ancora di più sulla deviazione standard. Viceversa, però, eliminare il dato potrebbe anche portare alla inaccettabile perdita di informazioni utili.

Non è stato raggiunto alcun accordo convenzionale per questo tipo di situazioni ed anzi ci sono posizioni contrapposte:

- la posizione più rigorosa afferma che, nel caso ci sia un dato apparentemente anomalo e non ci sia alcuna ragione per scartarlo, è necessario proseguire l'esperimento, acquisendo dati aggiuntivi in numero tale che il dato anomalo non influenzi in maniera rilevante la media e la deviazione standard;
- una posizione più intermedia è invece quella di calcolare, al posto della media aritmetica, una media pesata, nella quale i coefficienti di pesi "aiutino" a ridurre la rilevanza, nel risultato finale, dei dati anomali o meno significativi;
- tra le posizioni più permissive, invece, ne esiste uno che va sotto il nome di **criterio di Chauvenet**: il criterio si riferisce al caso in cui le misure abbiano delle deviazioni dalla media distribuite secondo la legge normale; sotto questa ipotesi, il criterio fissa un cosiddetto **intervallo di copertura**, all'interno del quale devono essere comprese tutte le misure; in tal modo, *ogni eventuale misura che dovesse cadere al di fuori di questo intervallo può essere eliminata*. Ovviamente, l'intervallo di misura è scelto in modo oculato: in particolare, lo si fissa imponendo che sia $1/2N$ la probabilità che una generica misura cada al di fuori dell'intervallo, dove N è il numero di misure a disposizione. In termini formali, quindi, una volta ottenute le N misure, la condizione da applicare per ottenere l'intervallo di copertura è

$$1 - P(\bar{X} \pm ks_x) \leq \frac{1}{2N}$$

dove \bar{X} e s_x sono, rispettivamente, la media e la deviazione standard del campione di misura a disposizione.

Facciamo un esempio di applicazione di questo criterio. Supponiamo di avere a disposizione $N=20$ misure, per cui deve risultare $1 - P(\bar{X} \pm ks) \leq \frac{1}{40} = 0.025$, ossia $P(\bar{X} \pm ks) \geq 0.975$. Dallo studio della distribuzione normale sappiamo che vale la relazione $P(\mu - k\sigma \leq X \leq \mu + k\sigma) = 2\phi(k) - 1$: allora, ponendo $2\phi(k) - 1 = 0.975$, si trova $\phi(k) = 0.9875$, da cui, usando le apposite tabelle sulla distribuzione normale standard, si trova il valore $k = 2,24$. Quindi, saranno scartate tutte le misure che risultino al di fuori dell'intervallo $(\bar{X} \pm 2,24s)$, ossia tutte le misure la cui deviazione dalla media è maggiore di $2,24s$.

Se riportassimo su una tabella i valori che, al variare di N , si ottengono per k (con il metodo appena visto), troveremmo un interessante risultato: fissati il valor medio \bar{X} e la deviazione standard del campione, si trova che, all'aumentare di N , si allarga l'intervallo in cui possono cadere le misure.

Il criterio di Chauvenet ha un problema fondamentale nel fatto che, dopo aver eliminato eventuali misure anomale, cambia l'insieme dei dati e quindi cambiano anche la media e la deviazione standard; di conseguenza, sul nuovo insieme dei dati, bisogna ripetere il calcolo dell'intervallo di copertura e questo potrebbe risultare tale da dover eliminare ulteriori dati. Così facendo, il procedimento potrebbe continuare ulteriormente, il che evidenzia i limiti del criterio. Allo scopo di evitare situazioni ambigue, l'indicazione comune è quella di applicare non più di una volta il criterio di Chauvenet.

METODO DEI MINIMI QUADRATI

La prima formulazione del **principio dei minimi quadrati** fu la seguente: *il valore più probabile di una qualunque quantità misurata è tale che la somma dei quadrati delle deviazioni delle misure da questo valore è minimo.* Vediamo in cosa si traduce questa frase.

Supponiamo di avere N misure ripetute di una stessa quantità, che formano perciò il campione di misura (X_1, X_2, \dots, X_N) . La differenza (o **scarto**) della generica misura X_i dal valore più probabile X del misurando è $X - X_i$; allora il principio dei minimi quadrati afferma che X è quel valore che minimizza la seguente sommatoria:

$$\sum_{i=1}^N (X - X_i)^2$$

Minimizzare questa sommatoria rispetto ad X significa imporre la seguente condizione:

$$\frac{d}{dX} \sum_{i=1}^N (X - X_i)^2 = 0$$

Calcoliamo allora quella derivata:

$$\frac{d}{dX} \sum_{i=1}^N (X - X_i)^2 = \sum_{i=1}^N \frac{d}{dX} (X - X_i)^2 = \sum_{i=1}^N 2(X - X_i) = 2NX - 2 \sum_{i=1}^N X_i$$

Imponendo che la quantità ricavata sia uguale a zero, si trova evidentemente che X coincide con la media aritmetica delle N misure a nostra disposizione:

$$X = \frac{1}{N} \sum_{i=1}^N X_i$$

Quindi, dall'applicazione del principio ai minimi quadrati deriva, in modo conforme a quanto già visto in precedenza, che il valore più probabile di una quantità misurata più volte è la media aritmetica delle misure ottenute.

Si può allora pensare ad una serie di applicazioni di questo principio. Ci interessiamo, in particolare, all'applicazione del metodo dei minimi quadrati alle equazioni lineari in due variabili.

In forma del tutto generica, una equazione lineare in due variabili può essere scritta nella forma seguente:

$$y = mx + q$$

Su un piano cartesiano, questa è ovviamente l'espressione di una retta con coefficiente angolare m e intercetta sull'asse delle ordinate pari a q .

Supponiamo di aver eseguito diverse misure delle variabili y ed x , al fine di aumentare la precisione con cui identificare i parametri m e q . Dato che le incognite sono due, dobbiamo avere a disposizione un numero N di misure (o meglio coppie di misure) maggiore di 2. Ciascuna coppia di misura ci darà origine ad una equazione del tipo

$$y_i = mx_i + q$$

In realtà, non è proprio così, *in quanto le N equazioni trovate non saranno in genere tutte consistenti, ma esisteranno delle deviazioni rispetto ai valori teorici*. In altre parole, per la generica coppia (x_i, y_i) di misure, risulterà una equazione del tipo

$$y_i + d_i = mx_i + q$$

dove d_i è appunto la deviazione.

A questo punto, in base al principio dei minimi quadrati, i valori più probabili di m e di q , ossia quelli che consentono di individuare la retta che meglio raccordi i punti sperimentali (x_i, y_i) , sono quelli che minimizzano la sommatoria delle deviazioni quadratiche:

$$\sum_{i=1}^N d_i^2$$

Andiamo allora ad esplicitare meglio questa sommatoria e poi a calcolarne le derivate parziali rispetto ad m ed a q , in modo da uguagliarle a zero:

$$\sum_{i=1}^N d_i^2 = \sum_{i=1}^N (mx_i + q - y_i)^2 \longrightarrow$$

$$\longrightarrow \begin{cases} \frac{\partial}{\partial m} \sum_{i=1}^N (mx_i + q - y_i)^2 = \sum_{i=1}^N \frac{\partial}{\partial m} (mx_i + q - y_i)^2 = \sum_{i=1}^N 2x_i (mx_i + q - y_i) \\ \frac{\partial}{\partial q} \sum_{i=1}^N (mx_i + q - y_i)^2 = \sum_{i=1}^N \frac{\partial}{\partial q} (mx_i + q - y_i)^2 = \sum_{i=1}^N 2(mx_i + q - y_i) \end{cases}$$

Uguagliando dunque a zero le due equazioni ottenute, otteniamo il seguente sistema di due equazioni in due incognite:

$$\begin{cases} \sum_{i=1}^N x_i (mx_i + q - y_i) = 0 \\ \sum_{i=1}^N (mx_i + q - y_i) = 0 \end{cases}$$

E' immediato trovare le soluzioni di questo sistema: in primo luogo, possiamo riarrangiare le due equazioni, scrivendo che

$$\begin{cases} m \sum_{i=1}^N x_i^2 + q \sum_{i=1}^N x_i - \sum_{i=1}^N x_i y_i = 0 \\ m \sum_{i=1}^N x_i + Nq - \sum_{i=1}^N y_i = 0 \end{cases}$$

Ponendo $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ e $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$, possiamo ulteriormente scrivere il sistema nella forma

$$\begin{cases} m \sum_{i=1}^N x_i^2 + qN\bar{x} - \sum_{i=1}^N x_i y_i = 0 \\ mN\bar{x} + Nq - N\bar{y} = 0 \end{cases}$$

Ricavando ad esempio m dalla seconda equazione e sostituendo nella prima, otteniamo

$$m = \frac{\bar{y} - q}{\bar{x}} \longrightarrow \frac{\bar{y} - q}{\bar{x}} \cdot \sum_{i=1}^N x_i^2 + qN\bar{x} - \sum_{i=1}^N x_i y_i = 0 \longrightarrow q = \frac{\frac{\bar{y}}{\bar{x}} \cdot \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i y_i}{\frac{1}{\bar{x}} \cdot \sum_{i=1}^N x_i^2 - N\bar{x}} = \frac{\bar{y} \cdot \sum_{i=1}^N x_i^2 - \bar{x} \cdot \sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2 - N\bar{x}^2}$$

Tornando adesso nell'espressione di m , otteniamo

$$m = \frac{\bar{y} - q}{\bar{x}} = \frac{\bar{y} \cdot \sum_{i=1}^N x_i^2 - \bar{x} \cdot \sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2 - N\bar{x}^2} = \frac{1}{\bar{x}} \bar{y} - \frac{\frac{1}{\bar{x}} \bar{y} \cdot \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2 - N\bar{x}^2} = \frac{\sum_{i=1}^N x_i y_i - N\bar{x}\bar{y}}{\sum_{i=1}^N x_i^2 - N\bar{x}^2}$$

Concludiamo dunque che i coefficienti della retta ricercata sono i seguenti:

$$m = \frac{\sum_{i=1}^N x_i y_i - N \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^N x_i^2 - N\bar{x}^2}$$

$$q = \frac{\bar{y} \cdot \sum_{i=1}^N x_i^2 - \bar{x} \cdot \sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2 - N\bar{x}^2}$$

Il metodo dei minimi quadrati può anche essere generalizzato al caso in cui la relazione tra le variabili x ed y non sia lineare. Tipico esempio è quello di una relazione di tipo polinomiale, in cui cioè risulti

$$y = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m$$

In questo caso, le incognite da determinare sono evidentemente gli $m+1$ coefficienti. Servono perciò un numero di coppie di misure (x_i, y_i) in numero N maggiore di $m+1$. Per quanto riguarda il modo di procedere, è lo stesso visto prima, nel senso che bisogna ipotizzare che le varie coppie di misure corrispondano a delle deviazioni tali che

$$y_i - d_i = a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_m x_i^m$$

Esplicitando da qui l'espressione di d_i , bisogna usarla per minimizzare ancora una volta la quantità $\sum_{i=1}^N d_i^2$; bisogna perciò calcolare le $m+1$ derivate parziali di quella quantità, una rispetto a ciascun coefficiente, ed uguagliarle a zero, arrivando così ad un sistema di $m+1$ equazioni in $m+1$ incognite dal quale ricavare i coefficienti desiderati.

Il problema di questa metodologia è che i tempi di calcoli possono anche essere lunghi, per cui sono stati messi a punti degli algoritmi che accelerino la minimizzazione della funzione $\sum_{i=1}^N d_i^2$. queste procedure vanno sotto il nome di **tecniche di ottimizzazione mediante l'LSM** (acronimo inglese di *Least Squares Method*).

RETTE DI REGRESSIONE E COEFFICIENTE DI CORRELAZIONE

Nel paragrafo precedente abbiamo fatto l'ipotesi di conoscere a priori il legame funzionale tra le grandezze x ed y ; tuttavia, in certi casi si può non essere certi su questa relazione funzionale. Allora, allo scopo di trovare la legge che meglio raccorda i punti sperimentali, sono state concepite diverse procedure iterative.

In particolare, si ricorre a **tecniche di regressione**, che consistono nel ricercare una relazione matematica tra le misure di due variabili, in modo tale che il valore di una variabile possa essere predetto dalla misura dell'altra variabile. In questa sede non vogliamo entrare nel merito delle tecniche di regressione, mentre ci concentriamo sulla verifica più semplice, che è quella della **correlazione lineare**.

Il punto di partenza è quello di ipotizzare, tra le due variabili in questione, un legame ancora una volta lineare del tipo

$$y = mx + q$$

Supponendo di aver compiuto misure ripetute di y e di x , si può esprimere quella equazione in altro modo: infatti, è facile verificare che anche la coppia di punti costituiti dalle medie \bar{x} ed \bar{y} dei dati sperimentali cade sulla retta di equazione $y = mx + q$, per cui si può scrivere che

$$\bar{y} = m\bar{x} + q$$

Se ora sottraiamo quest'ultima relazione da $y = mx + q$, otteniamo

$$y - \bar{y} = m(x - \bar{x})$$

Questa è ancora l'equazione di una retta, che viene detta **retta di regressione di y su x** . La possiamo anche scrivere in altra forma e precisamente

$$\frac{1}{m}(y - \bar{y}) = x - \bar{x}$$

e questa prende il nome di **retta di regressione di x su y**.

In generale, dunque, le due rette di regressione sono

$$\begin{aligned} y - \bar{y} &= m(x - \bar{x}) \\ x - \bar{x} &= m'(y - \bar{y}) \end{aligned}$$

Nel caso di perfetta correlazione, come si è visto, risulta $m \cdot m' = 1$. In realtà, la relazione lineare in genere è vera solo approssimativamente. Si definisce perciò **coefficiente di correlazione** la quantità

$$r = \sqrt{m \cdot m'}$$

Nel caso in cui risulti $r=1$, allora vi è perfetta correlazione tra x ed y . Se invece dovesse risultare $r=0$, allora uno tra m ed m' dovrà essere nullo:

- se $m=0$, si ha evidentemente $y = \bar{y}$, il che significa che non c'è correlazione tra i punti e che la retta di regressione è parallela all'asse x ;
- viceversa, se $m'=0$, risulta $x = \bar{x}$ e la retta di regressione è parallela all'asse y .

In genere, r (compreso tra 0 ed 1) non è mai perfettamente uguale ad 1, per cui le due rette di regressione non coincideranno, ma passeranno ambedue per il punto di coordinate (\bar{x}, \bar{y}) . In questa situazione, si assume, come retta che meglio approssima i dati sperimentali, la bisettrice dell'angolo acuto tra le due rette di regressione.

Andiamo allora a ricavare una espressione per il calcolo del coefficiente di correlazione. Avendo detto che $r = \sqrt{m \cdot m'}$, possiamo andare a trovare le espressioni di m e di m' che minimizzano la sommatorie delle deviazioni quadratiche. Per quanto riguarda m , il discorso è stato già fatto nel precedente paragrafo, nel quale abbiamo trovato che

$$m = \frac{\sum_{i=1}^N x_i y_i - N \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^N x_i^2 - N \bar{x}^2}$$

Possiamo fare lo stesso discorso per m' . In particolare, possiamo partire dalla equazione $x - \bar{x} = m'(y - \bar{y})$ e la possiamo utilizzare per calcolare gli scarti d_i in corrispondenza di ciascuna coppia di misure:

$$x_i - \bar{x} - d_i = m'(y_i - \bar{y}) \longrightarrow d_i = (x_i - \bar{x}) - m'(y_i - \bar{y})$$

Dobbiamo come al solito minimizzare la quantità $\sum_{i=1}^N d_i^2$, rispetto ovviamente ad m' :

$$\sum_{i=1}^N d_i^2 = \sum_{i=1}^N [(x_i - \bar{x}) - m'(y_i - \bar{y})]^2$$

$$\frac{\partial}{\partial m'} \sum_{i=1}^N d_i^2 = \frac{\partial}{\partial m'} \sum_{i=1}^N [(x_i - \bar{x}) - m'(y_i - \bar{y})]^2 = \sum_{i=1}^N \frac{\partial}{\partial m'} [(x_i - \bar{x}) - m'(y_i - \bar{y})]^2 =$$

$$= -2 \sum_{i=1}^N [(x_i - \bar{x}) - m'(y_i - \bar{y})](y_i - \bar{y}) = -2 \sum_{i=1}^N [(y_i(x_i - \bar{x}) - \bar{y}(x_i - \bar{x})) - m'(y_i - \bar{y})^2] = \dots$$

Uguagliando a zero quella derivata e facendo ulteriori passaggi, si trova alla fine che

$$m' = \frac{\sum_{i=1}^N x_i y_i - N \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^N (y_i - \bar{y})^2} = \frac{\sum_{i=1}^N x_i y_i - N \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^N y_i^2 - N \bar{y}^2}$$

Note le espressioni di m ed m' (che, come si vede, hanno uguale il numeratore), possiamo trovarci quella di r:

$$r = \sqrt{m \cdot m'} = \frac{\sum_{i=1}^N x_i y_i - N \cdot \bar{x} \cdot \bar{y}}{\sqrt{\left(\sum_{i=1}^N x_i^2 - N \bar{x}^2 \right) \left(\sum_{i=1}^N y_i^2 - N \bar{y}^2 \right)}}$$

Osservando le due quantità che vengono moltiplicate sotto radice, notiamo che si tratta, a meno ciascuna del termine N, delle varianze rispettivamente di x e di y, per cui concludiamo che

$$r = \frac{\sum_{i=1}^N x_i y_i - N \cdot \bar{x} \cdot \bar{y}}{N \cdot \sigma_x \cdot \sigma_y}$$

Il problema fondamentale di questo modo di procedere è che, spesso, l'esame dei punti sperimentali riportati su un grafico mostra che non esiste una relazione lineare tra i suddetti punti. Tuttavia, conoscendo a priori le leggi fisiche che governano il processo in esame, è possibile pensare a delle trasformazioni tali da ottenere una rappresentazione grafica approssimabile con una relazione lineare. Tipico è il caso in cui y ed x sono legate da una legge esponenziale: in questo caso, se usiamo un diagramma cartesiano con le scale in unità logaritmiche, sappiamo di avere una relazione approssimativamente lineare, alla quale quindi applicare i discorsi appena terminati.

Nel caso non si possa in alcun modo ricondursi ad una legge lineare, è necessario ricorrere a relazioni di tipo polinomiale:

$$y = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m$$

Le tecniche di regressione vanno allora applicate stimando il grado ed i parametri di questi polinomi, fino ad ottenere una curva che approssimi i dati sperimentali in modo soddisfacente. Un criterio da poter seguire per arrestare la ricerca è quello per cui la massima deviazione tra i dati e la curva

calcolata deve essere inferiore al valore massimo dell'errore di misura calcolato e che determina la fascia di incertezza.

Tuttavia, non sempre questa condizione è soddisfatta, per cui permane il problema di quale curva scegliere; si tratta anche qui di trovare un opportuno coefficienti di correlazione o qualcosa di simile. Si possono usare le cosiddetta **prove di confidenza**; una di esse dice, per esempio, quanto segue: per prima cosa si calcola la somma S delle deviazioni quadratiche relative ad un dato polinomio con il quale si è tentato di raccordare i dati; si confronta tale somma S con quella S' relativa alla curva di regressione di ordine superiore; se risulta $S < S'$, significa che si può arrestare l'algoritmo, altrimenti si prende la curva di ordine ancora superiore e si fa un ulteriore confronto.

Autore: **SANDRO PETRIZZELLI**
e-mail: sandry@iol.it
sito personale: <http://users.iol.it/sandry>
succursale: <http://digilander.iol.it/sandry1>