Testing psicologico

Modelli e metodi statistici per la misurazione in psicologia

Antonio Calcagnì

Dipartimento di Psicologia dello Sviluppo e della Socializzazione (DPSS)
Università di Padova

A.A. 2019/2020

Introduzione

«When you can measure what you are speaking about and express it in a number, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of science»

(Lord Kelvin, 3 May 1883)

Introduzione

La misurazione è in relazione al processo di scoperta scientifica.

La **metrologia** è la disciplina che si occupa di studiare la misurazione da una prospettiva scientifica, includendo sia la parte teorica (modelli matematici della misurazione), quella sperimentale (metodi per la misurazione) e quella tecnica (strumenti per la misurazione).

Introduzione

La misurazione è in relazione al processo di scoperta scientifica.

La **metrologia** è la disciplina che si occupa di studiare la misurazione da una prospettiva scientifica, includendo sia la parte teorica (modelli matematici della misurazione), quella sperimentale (metodi per la misurazione) e quella tecnica (strumenti per la misurazione).

Enti di riferimento per la misurazione:

- Bureau International des Poids et Mesures (BIPM)
- International Union of Pure and Applied Physics (IUPAP)
- International Union of Pure and Applied Chemistry (IUPAC)
- International Organization for Standardization (ISO)

Introduzione

La misurazione è in relazione al processo di scoperta scientifica.

La **metrologia** è la disciplina che si occupa di studiare la misurazione da una prospettiva scientifica, includendo sia la parte teorica (modelli matematici della misurazione), quella sperimentale (metodi per la misurazione) e quella tecnica (strumenti per la misurazione).

Enti di riferimento per la misurazione:

- Bureau International des Poids et Mesures (BIPM)
- International Union of Pure and Applied Physics (IUPAP)
- International Union of Pure and Applied Chemistry (IUPAC)
- International Organization for Standardization (ISO)

Riviste scientifiche di riferimento:

- Measurement (Elsevier)
- Metrologia (IOP)

Introduzione

Bureau International des Poids et Mesures (BIPM): ente intergovernativo che si occupa di misurazione e comprende diverse unità di lavoro, ad esempio CCAUV (acustica), CCEM (elettromagnetismo), CCL (lunghezza), CCM (massa), CCQM (biologia e chimica), CCT (termometria), ecc.

BIPM online: `https://www.bipm.org/en/`

Ogni stato partecipa al BIPM attraverso enti metrologici statali. L'Italia, ad esempio, partecipa al BIPM attraverso l'Istituto Nazionale di Ricerca Metrologica (INRM, `https://www.inrim.it/`)

Introduzione

Il BIPM è responsabile per gli standard di misurazione attraverso la *Guide to the Expression of Uncertainty in Measurement* (GUM):

«This Guide establishes general rules for evaluating and expressing uncertainty in measurement that are intended to be applicable to a broad spectrum of measurements»

GUM online: `https://www.bipm.org/en/publications/guides/gum.html`

Risorsa utile messa a disposizione da BIPM: *dizionario di metrologia* (VIM, `https://jcgm.bipm.org/vim/en/`)

Introduzione

Gli standard metrologici GUM-BIPM sono attualmente utilizzati dalla fisica e dall'ingegneria elettronica (*hard measurement*).

La teoria della misurazione non è unitaria: diversi modelli matematici sono utilizzati per caratterizzare diversi problemi.

Perché interessarsi della GUM:

- utilizza un approccio statistico alla misurazione e alla gestione dell'incertezza
- contiene molti elementi usati, spesso inconsapevolmente e in maniera errata, anche dalla metodologia psicologica (*soft measurement*)

Introduzione

Alcune **letture utili** per il problema della *soft measurement*:

- Mari, L. et al. (2009). Measurement in soft systems: epistemological framework and a case study. *Measurement*, 42(2), 241-253*
- Ferrero, A. et al. [Eds] (2015). *Modern measurements: fundamentals and applications*. John Wiley & Sons (cap. 7)*
- Berglund, B. et al. [Eds.] (2012). *Measurement with persons: theory, methods, and implementation areas*. Psychology Press (capp. 1,2,6,9)*

*Il materiale è disponibile nell'area APPROFONDIMENTI su Moodle del corso.

Introduzione

La misurazione in psicologia ha una lunga storia che risale ancor prima della GUM.

Diversi approcci proposti al problema della misurazione:

- approccio rappresentazionale (Suppes, Luce)
- approccio basato sullo scaling (Stevens, Guttman)
- approccio basato sulle quantità latenti (Spearman, Rasch)
- approccio geometrico (Townsend)
- ...

Ritorneremo sul problema della misurazione in psicologia nei moduli [C] e [D] del corso. Per ora ci concentreremo su alcuni elementi di base della misurazione secondo GUM.

Lessico metrologico

FONTI: G(1.2,1.3,1.4,1.6), K(1,2)

La **misurazione** è un insieme di regole attuate per ottenere informazioni e produrre conoscenza su una **quantità** oggetto di interesse. Una **quantità** è una proprietà misurabile di un fenomeno che può essere rappresentata in termini di numeri e operazioni matematiche. Le operazioni matematiche devono rappresentare azioni ammissibili sulle proprietà del fenomeno oggetto di studio.

La misurazione è fatta rispetto ad un **riferimento metrologico**: la *massa di un corpo* è una quantità misurabile facendo riferimento ad una unità di misura (kg); la *durezza* è una proprietà di un materiale che può essere misurata con riferimento ad una procedura (es.: scala Rockwell). In questo caso, la procedura è un modello del comportamento della proprietà oggetto di interesse.

Lessico metrologico

FONTI: G(1.2,1.3,1.4,1.6), K(1,2)

La misurazione è fatta rispetto ad un **riferimento metrologico**: l'**unità di misura** è una quantità convenzionale impiegata come riferimento per le misurazioni. Esempi: metro per la lunghezza, chilogrammo per la massa, kelvin per la temperatura, ecc.

Dire che la matita è lunga 10 cm significa che la sua lunghezza è 10^{-2} volte ($= 0.01$) l'unità di riferimento convenzionale che è il metro.

La quantità da misurare è chiamata **misurando** η e il suo valore vero η^* è generalmente ignoto e non definibile se non tramite una procedura con un certo riferimento teorico: occorre conoscere la teoria di riferimento di η, la scala su cui η si esprime, le eventuali relazioni di η con altre proprietà misurabili $\xi_1, \ldots, \xi_J$.

Lessico metrologico

FONTI: G(1.2,1.3,1.4,1.6), K(1,2)

Obiettivo della misurazione: fornire una stima $\hat{\eta}$ alquanto **precisa** ed **accurata** del misurando η utilizzando un **metodo sperimentale**.

Trueness (VIM 2.14): differenza tra il risultato di una misurazione e il suo valore vero ($\hat{\eta} - \eta^*$). Nota: *«Measurement accuracy should not be used for measurement trueness»*

Per calcolare $\hat{\eta}$ occorre utilizzare un **campione** di misurazioni $\mathbf{x} = (x_1, \ldots, x_n)$ estratto casualmente dalla popolazione che rappresenta il misurando Ω_η. Il campione è finito ed ha cardinalità n.

Una buona stima di η^* è ottenuta utilizzando la media campionaria:

$$\hat{\eta} = \text{mean}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Lessico metrologico

FONTI: G(1.2,1.3,1.4,1.6), K(1,2)

L'**errore** di misura è definito come $\epsilon_\eta = (\hat{\eta} - \eta^*)$ ed è dovuto a scarsa accuratezza, non ripetibilità delle misurazioni, scarsa risoluzione degli strumenti utilizzati per la misurazione, condizioni ambientali, malfunzionamento degli strumenti di misurazione.

Possiamo individuare tre tipologie generali di errore:

- casuale (random error)
- sistematico (systematic errors)
- spurio (mistakes)

Lessico metrologico

FONTI: G(1.2,1.3,1.4,1.6), K(1,2)

L'**errore casuale** in una serie di misurazioni $(x_1, \ldots, x_n)$ effettuate nelle medesime condizioni è una quantità reale $\delta_\eta \in \mathbb{R}$ che varia in maniera non predicibile in segno e magnitudine. Può dipendere da diverse fonti tra cui la mutabilità delle condizioni ambientali entro cui la misurazione è effettuata e dalle caratteristiche stesse del misurando (esempio: moto browniano).

In generale $\lim_{n\to\infty} \delta_\eta = 0$, ossia l'errore casuale tende a diminuire all'aumentare del numero n di misurazioni effettuate.

Lessico metrologico

FONTI: G(1.2,1.3,1.4,1.6), K(1,2)

L'**errore sistematico** in una serie di misurazioni $(x_1, \ldots, x_n)$ effettuate nelle medesime condizioni è una quantità reale $\kappa_\eta \in \mathbb{R}$ che rimane costante in segno e magnitudine (errore sistematico costante) oppure che varia in maniera predicibile (errore sistematico variabile). Può dipendere da diverse fonti tra cui l'uso di uno strumento difettoso, l'utilizzo di uno strumento in condizioni diverse da quelle di taratura, errore dell'operatore che effettua la misurazione.

In generale $\lim_{n\to\infty} \kappa_\eta = \kappa^*$, ossia l'errore sistematico rimane costante all'aumentare del numero n di misurazioni effettuate.

Lessico metrologico

FONTI: G(1.2,1.3,1.4,1.6), K(1,2)

La definizione dell'errore casuale (o errore statistico) si basa sul concetto primitivo di **ripetibilità** della misurazione:

> *«condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time» (VIM 2.20)*

La ripetibilità si valuta analizzando quanto le misurazioni effettuate $(x_1, \ldots, x_n)$ siano vicine tra loro:

> *«closeness between the results of successive measurements of the same measurand» G(1.3.15)*

Lessico metrologico

FONTI: G(1.2,1.3,1.4,1.6), K(1,2)

Rispetto ad uno strumento di misura, la **ripetibilità** è definita come la capacità di fornire medesimi risultati di misurazione a parità di condizioni di misurazione mentre l'**accuratezza** è definita come la capacità di fornire risultati corretti. Le due proprietà differiscono: «Accuracy is a measure of an instrument's ability to tell the truth, while repeatability is a measure of its ability to indicate the same value of the measured quantity» G(1.6.3).

Una buona ripetibilità è condizione necessaria ma non sufficiente di buona accuratezza: buona ripetibilità non indica buona accuratezza (esempio: κ_η rimane costante all'aumentare di *n*).

Analisi degli errori

FONTI: HH(1,2)

Ripetere un esperimento di misurazione nelle medesime condizioni non assicura l'eliminazione dell'errore casuale che rimane una sorta di *rumore bianco*. È dunque sempre presente e sembra poter essere in relazione alle caratteristiche naturali dei misurandi e della realtà empirica che li contiene.

Esempio HH(1.2):

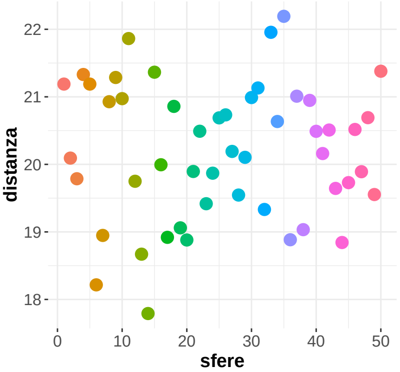
Proviamo a sparare *n* sfere da un lanciatore a molla nelle stesse ed esatte condizioni sperimentali: le leggi di Newton del moto che governano il comportamento delle sfere rimangono invariate ad ogni lancio/misurazione.

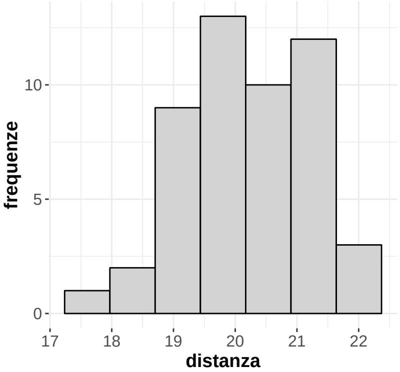
Le distanze delle *n* sfere dalla loro base di lancio al punto di arresto sono tuttavia diverse *a parità di condizioni sperimentali*.

Analisi degli errori

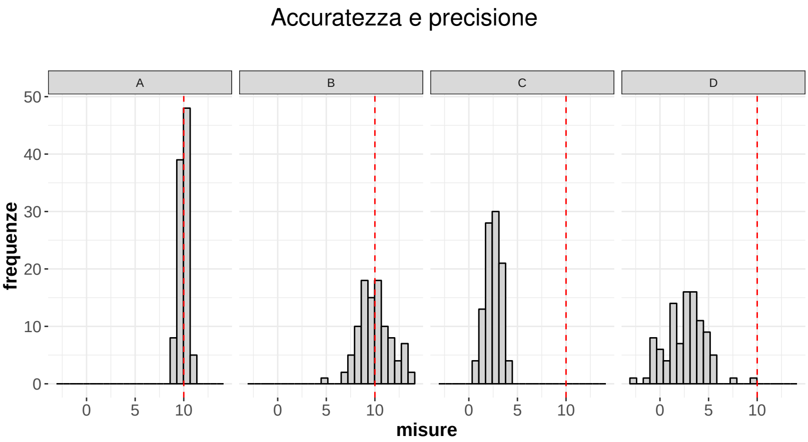
FONTI: HH(1,2)

Rappresentazione grafica delle distanze misurate dal punto di lancio





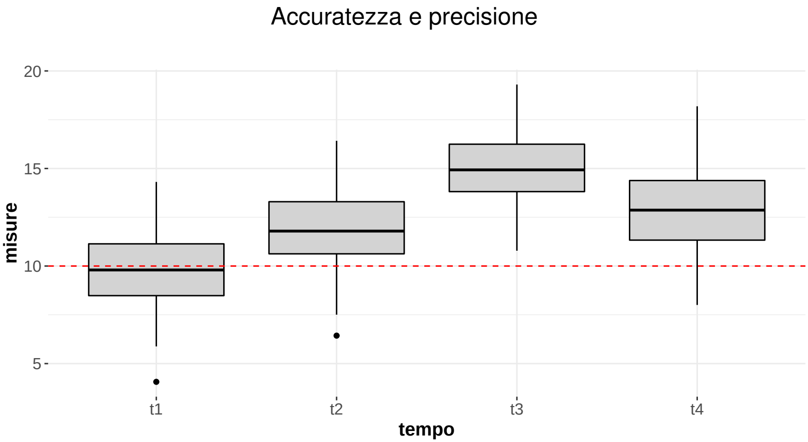
Analisi degli errori
FONTI: HH(1,2)



* *la linea rossa tratteggiata indica il vero valore di misura* η^*
A: accurato e preciso
B: accurato e impreciso
C: inaccurato e preciso
D: inaccurato e impreciso

Analisi degli errori

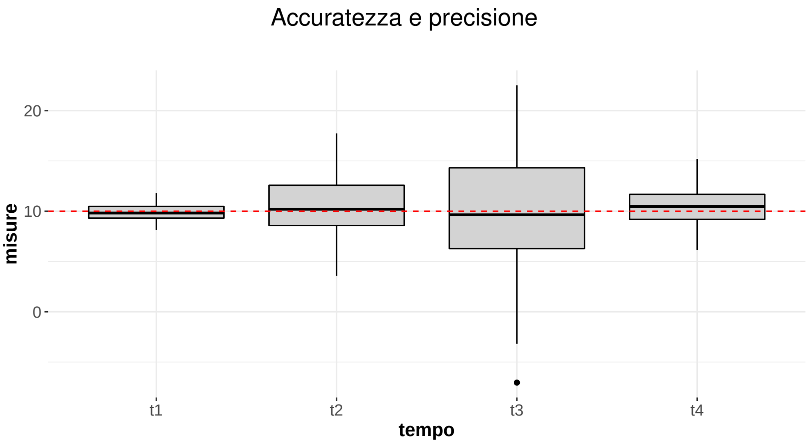
FONTI: HH(1,2)



* *la linea rossa tratteggiata indica il vero valore di misura* η^*
Andamento di *n* misure effettuate in tempi diversi: lo scostamento da η^* (accuratezza) indica una possibile presenza di errore sistematico κ_η. Da notare come la precisione resta invariata al variare del tempo.

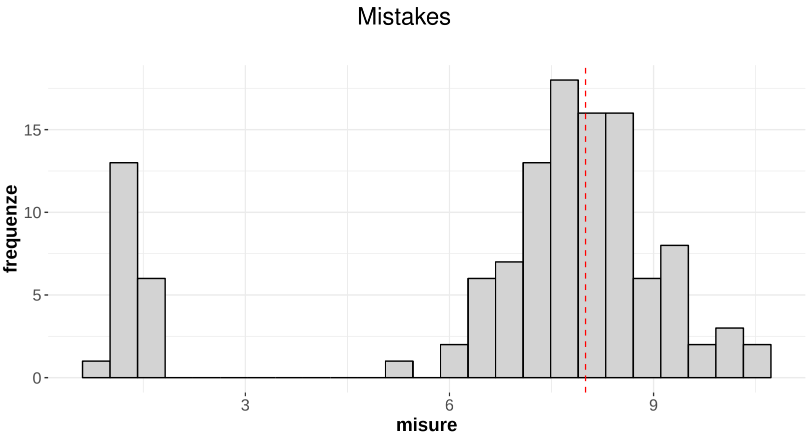
Analisi degli errori

FONTI: HH(1,2)



* *la linea rossa tratteggiata indica il vero valore di misura* η*
Andamento di *n* misure effettuate in tempi diversi: la variazione della precisione indica la presenza di errore casuale δ_η. Da notare come l'accuratezza resta invariata al variare del tempo.

Analisi degli errori
FONTI: HH(1,2)



$*$ *la linea rossa tratteggiata indica il vero valore di misura* η^*
Grafico di n misurazioni di η. La presenza di misurazioni errate è visibile dallo scostamento di queste dalla distribuzione principale delle misurazioni.

Analisi degli errori

FONTI: HH(1,2)

In generale, dato un campione di n misure $(x_1, \ldots, x_n)$:

- l'errore causuale δ_η influenza la **precisione** delle misurazioni e rallenta/velocizza la convergenza al valor vero $\hat{\eta} \to \eta^*$
- l'errore sistematico κ_η influenza l'**accuratezza** delle misurazioni e distorce la convergenza $\hat{\eta} \to \eta^*$ di una quantità costante κ^* tale per cui si ha che $\lim_{n\to\infty} \hat{\eta} = \eta^* + \kappa^*$ e $\frac{1}{n}\sum_i x_i \neq \eta^*$

Stima di η: la media

FONTI: CDM(4.1,4.2), G(1.4.9-1.4.12)

Abbiamo affermato come dato un insieme di misurazioni $\mathbf{x} = (x_1, \ldots, x_n)$ effettuate nelle medesime condizioni, una stima $\hat{\eta}$ della quantità di interesse η è ottenuta tramite la **media aritmetica**:

$$\hat{\eta}_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Proprietà notevoli:

1. internalità: $\min(\mathbf{x}) \leq \hat{\eta}_x \leq \max(\mathbf{x})$
2. invarianza rispetto alla somma: $\sum_{i=1}^{n} x_i = \hat{\eta}_x n$
3. $(\sum_{i=1}^{n} x_i - \hat{\eta}_x) = 0$
4. $\min_{c \in \mathcal{C}} \sum_{i=1}^{n} (x_i - c)^2 = \sum_{i=1}^{n} (x_i - \hat{\eta}_x)^2$ quando c è la media aritmetica
5. associatività: se $\mathbf{x} = \underbrace{\{(x_1, x_2), \ldots, (x_{n-1}, x_n)\}}_{\text{m sottogruppi}}$ allora $\hat{\eta}_x = \frac{1}{\sum_{j=1}^{m} n_j} \sum_{j=1}^{m} \hat{\eta}_j n_j$

Stima di η: la media

FONTI: CDM(4.1,4.2), G(1.4.9-1.4.12)

Abbiamo affermato come dato un insieme di misurazioni $\mathbf{x} = (x_1, \ldots, x_n)$ effettuate nelle medesime condizioni, una stima $\hat{\eta}$ della quantità di interesse η è ottenuta tramite la **media aritmetica**:

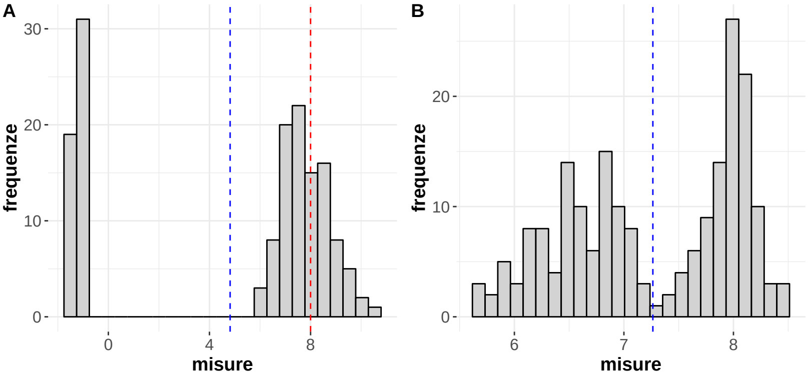
$$\hat{\eta}_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Alcuni limiti:

1. è "attratta" dai valori anomali della distribuzione
2. non è utile quando le distribuzioni delle misurazioni sono non unimodali

Stima di η: la media

FONTI: CDM(4.1,4.2), G(1.4.9-1.4.12)



* *la linea rossa tratteggiata indica il vero valore di misura* η^* *mentre quella blu indica* $\hat{\eta}$

A: caso in cui $\hat{\eta}$ è attratta da misurazioni anomali

B: caso in cui $\hat{\eta}$ è calcolata per misurazioni con distribuzioni bimodali

Stima della precisione

FONTI: CDM(5.2,5.4), HH(2.3)

Dato un insieme di misurazioni $\mathbf{x} = (x_1, \ldots, x_n)$ effettuate nelle medesime condizioni, una stima della **precisione** è fatta sfruttando la distribuzione di $\mathbf{x}$ e $\hat{\eta}_x$.

Un primo modo per calcolare la precisione è sfruttando il **campo di variazione**:

$$\Delta_x = \max(\mathbf{x}) - \min(\mathbf{x})$$

Proprietà notevoli:

1. $\Delta_x \geq 0$
2. $\Delta_x = 0$ se $\mathbf{x} = (x, \ldots, x)$
3. se $\mathbf{y} = a + \mathbf{x}$, allora $\Delta_y = \Delta_x$
4. se $\mathbf{y} = b\mathbf{x}$, allora $\Delta_y = b\Delta_x$

Stima della precisione

FONTI: CDM(5.2,5.4), HH(2.3)

Dato un insieme di misurazioni $\mathbf{x} = (x_1, \ldots, x_n)$ effettuate nelle medesime condizioni, una stima della **precisione** è fatta sfruttando la distribuzione di $\mathbf{x}$ e $\hat{\eta}_x$.

Un primo modo per calcolare la precisione è utilizzando il **campo di variazione**:

$$\Delta_x = \max(\mathbf{x}) - \min(\mathbf{x})$$

Alcuni limiti:

1. dipende unicamente dai valori estremi delle misurazioni
2. assume lo stesso valore per distribuzioni con medesimi estremi

Stima della precisione

FONTI: CDM(5.2,5.4), HH(2.3)

Dato un insieme di misurazioni $\mathbf{x} = (x_1, \ldots, x_n)$ effettuate nelle medesime condizioni, una stima della **precisione** è fatta sfruttando la distribuzione di $\mathbf{x}$ e $\hat{\eta}_x$.

Un secondo modo per calcolare la precisione è utilizzando lo **scarto quadratico medio** (standard deviation):

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{n}(x_i - \hat{\eta}_x)^2}$$

Proprietà notevoli:

1. $\sigma_x \geq 0$
2. $\sigma_x = 0$ se $\mathbf{x} = (x, \ldots, x)$
3. se $\mathbf{y} = a + \mathbf{x}$, allora $\sigma_y = \sigma_x$
4. se $\mathbf{y} = b\mathbf{x}$,s allora $\sigma_y = b\sigma_x$

Stima della precisione

FONTI: CDM(5.2,5.4), HH(2.3)

Dato un insieme di misurazioni $\mathbf{x} = (x_1, \ldots, x_n)$ effettuate nelle medesime condizioni, una stima della **precisione** è fatta sfruttando la distribuzione di $\mathbf{x}$ e $\hat{\eta}_x$.

Un terzo modo per calcolare la precisione è utilizzando il **coefficiente di variazione** (errore relativo):

$$\mathrm{CV}(\mathbf{x}) = \frac{\sigma_x}{|\hat{\eta}_x|}$$

Alcuni pregi:

1. è un indice adimensionale utile per confrontare misurazioni effettuate con diverse unità di misura
2. è utile per confrontare misurazioni con diverse medie
3. è invariante rispetto alla numerosità delle misurazioni n

e difetti:

1. $\mathrm{CV}(\mathbf{x}) \to \infty$ quando $\hat{\eta}_x \to 0$

Stima della precisione

FONTI: https://en.wikipedia.org/wiki/Signal-to-noise_ratio

Dato un insieme di misurazioni $\mathbf{x} = (x_1, \ldots, x_n)$ effettuate nelle medesime condizioni, una stima della **precisione** è fatta sfruttando la distribuzione di $\mathbf{x}$ e $\hat{\eta}_x$.

Un quarto modo per calcolare la precisione è utilizzando il **rapporto segnale-rumore** (signal-to-error ratio):

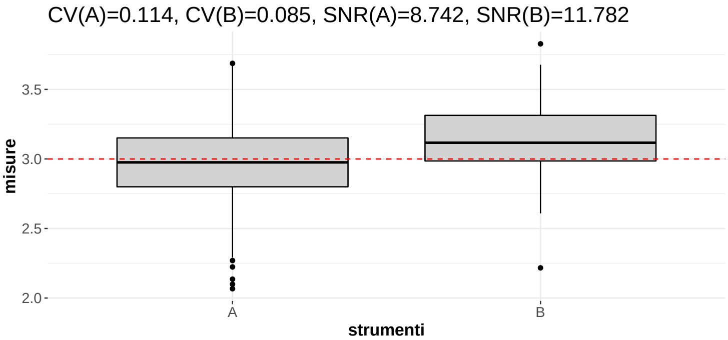
$$\text{SNR}(\mathbf{x}) = \frac{\hat{\eta}_x}{\sigma_x}$$

Valuta il rapporto tra quantità di segnale (stimata da $\hat{\eta}_x$) e rumore (stimata da σ_x). Un valore di $\text{SNR}(\mathbf{x}) \geq 5$ indica un'ottima capacità di distinguere il segnale dal rumore (criterio di Rose).

L'indice può essere calcolato solo quando $\mathbf{x} \in \mathbb{R}^n_+$, ossia per misurazioni sempre positive.

Stima della precisione

FONTI: CDM(5.2,5.4), HH(2.3)



* *la linea rossa tratteggiata indica il vero valore di misura* η^*
Grafico di n misurazioni di η effettuate con due strumenti A e B. Da notare come $\hat{\eta}_A$ è più vicina a η^* rispetto a $\hat{\eta}_B$. Tuttavia lo strumento A presenta maggiore variabilità rispetto allo strumento B, evidenziato dal fatto che CV(A) $>$ CV(B). Allo stesso modo, lo strumento B è migliore nell'estrarre il segnale dal rumore, infatti SNR(A) $<$ SNR(B). Lo strumento B può essere preferito allo strumento A, correggendolo opportunamente $\eta^* = \hat{\eta}_B - \kappa_B$ (κ_B è una distorsione sistematica).

Errore di $\hat{\eta}$

FONTI: HH(2.7)

La ripetibilità delle misurazioni è fondamentale per studiare il comportamento della misura e degli strumenti.

ripetizione	campione	stima di η^*
1	$(x_1, \ldots, x_{n_1})_1$	$\hat{\eta}_{x_1}$
$\vdots$	$\vdots$	$\vdots$
j	$(x_1, \ldots, x_{n_j})_j$	$\hat{\eta}_{x_j}$
$\vdots$	$\vdots$	$\vdots$
J	$(x_1, \ldots, x_{n_J})_J$	$\hat{\eta}_{x_J}$

Le medie delle diverse misurazioni $\hat{\boldsymbol{\eta}} = (\hat{\eta}_{x_1}, \ldots, \hat{\eta}_{x_J})$ possono essere utilizzate per stimare con maggiore accuratezza il valore vero η^*. Sotto determinate condizioni, infatti, la distribuzione di $\hat{\boldsymbol{\eta}}$ è più accurata e precisa (per $J \to \infty$):

$$\hat{\eta}^0 = \left(\frac{1}{J} \sum_{j=1}^{J} \hat{\eta}_{x_j} \right) \to \eta^* \quad \text{e} \quad \hat{\sigma}_{\hat{\eta}^0} = \sqrt{\frac{1}{J} \sum_{j=1}^{J} (\hat{\eta}_{x_j} - \hat{\eta}^0)^2} \to 0$$

Errore di $\hat{\eta}$

FONTI: HH(2.7)

Dato un campione di n misurazioni, l'errore della stima (standard error) $\sigma_{\hat{\eta}}$ è calcolato come

$$\sigma_{\hat{\eta}} = \sigma_x \big/ \sqrt{n}$$

Il fattore $\frac{1}{\sqrt{n}}$ indica che l'errore diventa più piccolo all'aumentare di n.

La scrittura $\hat{\eta} \pm \sigma_{\hat{\eta}}$ (esempio: 13.46$\pm$0.04) indica il range della stima di η^*:

$$\hat{\eta} - \sigma_{\hat{\eta}} \leq \eta^* \leq \hat{\eta} + \sigma_{\hat{\eta}}$$

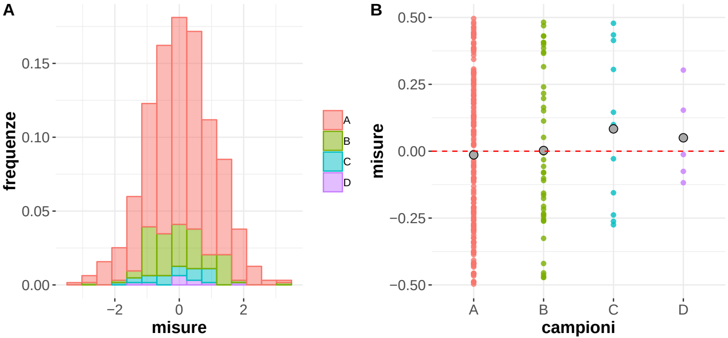
Scrittura analoga di tipo statistico: $\hat{\eta}$ ($\sigma_{\hat{\eta}}$), ad esempio: 13.46(0.04).

Nota bene:

- σ_x indica la dispersione delle misurazioni ($x_1, \ldots, x_n$) intorno a $\hat{\eta}_x$
- $\sigma_{\hat{\eta}}$ indica un'informazione circa la concordanza tra $\hat{\eta}$ e η^*

Errore di $\hat{\eta}$

FONTI: HH(2.7)



* *la linea rossa tratteggiata indica il vero valore di misura* η*
A: istogrammi di quattro campioni di misure con $n_A = 500$, $n_B = 100$, $n_C = 25$, $n_D = 10$. Notiamo come all'aumentare di n la distribuzione delle misurazioni si centra intorno ad un valore medio.
B: stessi campioni visti di profilo: il cerchio grigio indica la media calcolata sul campione. Per n piccolo, la media del campione si allontana da quella vera.

In entrambi i grafici: $\hat{\eta}_A = 0.05 \pm 0.04$, $\hat{\eta}_B = -0.01 \pm 0.09$, $\hat{\eta}_C = -0.03 \pm 0.16$, $\hat{\eta}_D = 0.14 \pm 0.27$

Analisi dell'incertezza

FONTI: G(7.1.5,7.7,7.8,7.14), SS(1.2,1.3)

Il principio della ripetibilità delle misure è alla base della valutazione dell'incertezza: ogni misurazione è portatrice di incertezza e ciò è dovuto a diversi fattori (del misurando, dello strumento, ambientali, ecc).

Prima delle regole GUM, l'incertezza di misura era rappresentata mediante la teoria degli errori classica (errore casuale vs. sistematico) che dava adito a diverse criticità, tra cui:

- scelta del modello matematico per la valutazione dell'errore sistematico
- scelta del metodo per combinare errore casuale e sistematico (ad esempio: somma vs. quadratura)

Le regole GUM cercano di superare queste criticità definendo in generale il concetto di **incertezza** senza una netta distinzione tra parte casuale e sistematica.

Analisi dell'incertezza

FONTI: G(7.1.5,7.7,7.8,7.14), SS(1.2,1.3)

L'incertezza è definita come: «non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used» (VIM 2.26).

«Measurement uncertainty includes components arising from systematic effects, such as components associated with corrections [...]» (NOTE 1)

«The parameter may be, for example, a standard deviation called standard measurement uncertainty (or a specified multiple of it), or the half-width of an interval, having a stated coverage probability» (NOTE 2)

Analisi dell'incertezza

FONTI: G(7.1.5,7.7,7.8,7.14), SS(1.2,1.3)

Definizione di due tipi generali di incertezza:

«Measurement uncertainty comprises, in general, many components. Some of these may be evaluated by **Type A** evaluation of measurement uncertainty from the statistical distribution of the quantity values from series of measurements and can be characterized by standard deviations. The other components, which may be evaluated by **Type B** evaluation of measurement uncertainty, can also be characterized by standard deviations, evaluated from probability density functions based on experience or other information.» (NOTE 3)

Analisi dell'incertezza

FONTI: G(7.1.5,7.7,7.8,7.14), SS(1.2,1.3)

Valutazione dell'**incertezza tipo A**:

Sia $\mathbf{x} = (x_1, \ldots, x_n)$ un campione di n misurazioni utilizzato per portare informazioni sul misurando η. In generale:

$$\underbrace{\eta}_{\text{misurando}} = f \underbrace{(\mathbf{x})}_{\text{input}}$$

Come discusso nelle slide precedenti, la stima migliore di η^* è ancora la media dell'input $\hat{\eta}$ mentre la sua incertezza è espressa ancora dalle quantità $\hat{\sigma}_{\hat{\eta}}$ e $\hat{\sigma}_x$.

Come visto in precedenza, l'incertezza tipo A comprende l'incertezza statistica dovuta al processo di misurazione.

Analisi dell'incertezza

FONTI: G(7.1.5,7.7,7.8,7.14), SS(1.2,1.3)

Valutazione dell'**incertezza tipo B**:

Comprende l'informazione non statistica non compresa nell'incertezza tipo A: informazioni a priori sui bias dello strumento o il processo di misurazione, errore sistematico, ecc.

Dato $\mathbf{x} = (x_1, \ldots, x_n)$ un campione di n misurazioni utilizzato per portare informazioni sul misurando η, la quantità u_x indica tale incertezza.

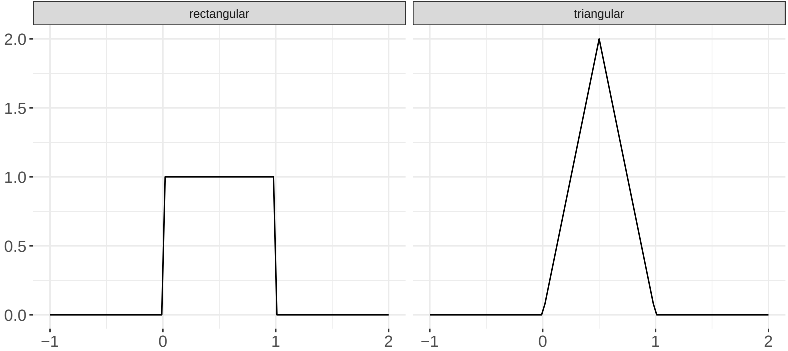
u_x deve essere espresso mediante distribuzioni di probabilità che quantifichino l'incertezza non statistica, ad esempio:

- distribuzione uniforme (o rettangolare): $u_x = \alpha/\sqrt{3}$
- distribuzione triangolare: $u_x = \alpha/\sqrt{6}$

dove α indica il range in cui la distribuzione si esprime.

Analisi dell'incertezza

FONTI: G(7.1.5,7.7,7.8,7.14), SS(1.2,1.3)



Esempio di distribuzione rettangolare e triangolare per la rappresentazione dell'incertezza tipo B.
In entrambi i casi: $\alpha = 1$.

Analisi dell'incertezza

FONTI: G(7.1.5,7.7,7.8,7.14), SS(1.2,1.3)

Calcolo dell'**incertezza totale** (regola della quadratura):

Dato $\mathbf{x} = (x_1, \ldots, x_n)$ un campione di n misurazioni utilizzato per portare informazioni sul misurando η, l'incertezza totale $\hat{\sigma}_{tot}$ è espressa dalla seguente combinazione:

$$\hat{\sigma}_{tot} = \sqrt{\hat{\sigma}_x^2 + u_x^2}$$

dove $\hat{\sigma}_x$ è l'incertezza dovuta alla ripetibilità della misurazione.

Quando $n = 1$, ossia nel caso di una misurazione singola:

$$\hat{\sigma}_{tot} = \sqrt{u_x^2}$$

l'incertezza totale è espressa solo dall'incertezza dovuta allo strumento.

Propagazione dell'incertezza

FONTI: HH(4.1,4.2,4.3)

Quando il misurando è rappresentato in funzione di una serie di quantità misurate $(\mathbf{x}_1, \ldots, \mathbf{x}_J)$ dove il generico $\mathbf{x}_j$ è un vettore di n misurazioni come di consueto:

$$\eta = f(\mathbf{x}_1, \ldots, \mathbf{x}_J)$$

si parla di **misurazione indiretta**. In questo caso il valore stimato del misurando è ottenuto applicando la funzione f sulle J misurazioni mentre l'incertezza $\hat{\sigma}_{\hat{\eta}}$ è ottenuta usando le regole della **propagazione dell'incertezza** come riportato nelle Tabelle 4.1-4.2 di HH(4.1.3,4.2.4).

Per l'obiettivo di questo corso, vedremo solo alcuni semplicissimi esempi durante le attività di laboratorio (utilizzando il pacchetto `propagate` di `R`).

Propagazione dell'incertezza

FONTI: HH(4.1,4.2,4.3)

In generale, la regola di propagazione quando η è *misurato indirettamente* è la seguente:

$$\hat{\sigma}_{\hat{\eta}} = \sum_{j=1}^{J} c_j^2 \hat{\sigma}_{x_j} + 2 \sum_{l=1}^{J-1} \sum_{h=l+1}^{J} c_l c_h \, \hat{\sigma}_{x_l} \hat{\sigma}_{x_h} \hat{\rho}_{x_{lh}}$$

Propagazione dell'incertezza

FONTI: HH(4.1,4.2,4.3)

In generale, la regola di propagazione quando η è *misurato indirettamente* è la seguente:

$$\hat{\sigma}_{\hat{\eta}} = \underbrace{\sum_{j=1}^{J} c_j^2 \hat{\sigma}_{x_j}}_{\text{contributo j-esima misurazione}} + \underbrace{2\sum_{l=1}^{J-1} \sum_{h=l+1}^{J} c_l c_h \, \hat{\sigma}_{x_l} \hat{\sigma}_{x_h} \hat{\rho}_{x_{lh}}}_{\text{correlazione a coppie tra misurazioni l e h}}$$

Note:

- $c_j := \frac{\partial f}{\partial x_j}$ è il coefficiente di sensibilità (sensitivity coefficient) ed è calcolabile usando ad esempio le espressioni indicate nelle Tabelle 4.1-4.2 di HH(4.1.3,4.2.4).
- quando x_l e x_h sono indipendenti il secondo addendo diventa zero e la formula si semplifica: $\hat{\sigma}_{\hat{\eta}} = \sum_{j=1}^{J} c_j^2 \hat{\sigma}_{x_j}$ (solo contributi delle singole misurazioni)

Propagazione dell'incertezza

FONTI: HH(4.1,4.2,4.3)

Esempio:

Si vuole misurare η utilizzando due misure X_1 e X_2. A tal fine si raccolgono due campioni $\mathbf{x}_1$ e $\mathbf{x}_2$ di n misurazioni indipendenti con $\hat{\sigma}_{x_1} = 2.1$ e $\hat{\sigma}_{x_1} = 0.4$. È noto che la funzione che esprime il misurando è la seguente: $\eta = 2.36 + \ln(X_1) + \ln(X_2)$.

L'incertezza delle misure di input è propagata sul misurando usando la formula 14 della tabella presente negli APPROFONDIMENTI:

$$\hat{\sigma}^2_{\hat{\eta}} = \left(\frac{\hat{\sigma}_{x_1}}{\overline{X_1}}\right)^2 + \left(\frac{\hat{\sigma}_{x_2}}{\overline{X_2}}\right)^2$$

Sapendo che $\overline{X_1} = 4.1$ e $\overline{X_2} = 9.3$:

- $\hat{\sigma}_{\hat{\eta}} = \sqrt{\hat{\sigma}^2_{\hat{\eta}}} = 0.514$
- $\hat{\eta} = 2.36 + \log(4.1) + \log(9.3) = 6.001$

Analisi dell'incertezza

Approfondimento sull'analisi dell'incertezza:

Farrance, I., & Frenkel, R. (2012). Uncertainty of measurement: a review of the rules for calculating uncertainty components through functional relationships. *The Clinical Biochemist Reviews*, 33(2), 49.*

*Il materiale è disponibile nell'area APPROFONDIMENTI su Moodle del corso.

Sintesi

- La teoria degli errori classica e l'approccio GUM offrono due approcci per la gestione dell'incertezza di misurazione
- La GUM offre una rappresentazione statistica dell'incertezza di misurazione
- Entrambi sono (inconsapevolmente) condivisi dalla gran parte della metodologia classica per la costruzione dei test comportamentali
- L'incertezza di tipo B può essere agevolmente modellata da approcci probabilistici di tipo Bayesiano
- L'incertezza di tipo B può essere utilizzata per modellare diverse fonti di disturbo nel testing psicologico (es.: fake data)
- Diversi modelli statistici possono innestarsi all'interno della GUM per caratterizzare meglio il problema dell'incertezza - si veda SS(1.4)

Sintesi

- In GUM la teoria della probabilità è impiegata per modellare il processo di misurazione
- La misurazione è rappresentata come esperimento aleatorio
- Le leggi statistiche fondamentali (es.: Teorema del limite centrale) sono alla base della quantificazione del misurando

Nel modulo [B] passeremo in rassegna alcuni dei fondamentali statistici della misurazione, richiamando taluni concetti importanti dell'inferenza statistica (es.: variabili aleatorie, distribuzioni di probabilità).