

## The small biostatistician

### STATISTICA DESCRITTIVA **Descrive i dati.**

**Variabile nominale:** categorie con nomi distinti, tra le quali non è possibile stabilire un criterio di ordinamento logico.

**Variabile ordinale:** serie ordinata di categorie, la differenza tra le quali non può essere considerata costante.

**Variabile di intervallo:** i valori possibili si presentano a intervalli uguali.

Grafico a barre: dati categorici. Istogramma: dati continui.

$X$  = valore di una variabile per un soggetto. Numerosità: numero di soggetti.

$n$  = numerosità di un gruppo del campione.  $N$  = numerosità dell'intero campione (somma delle numerosità dei gruppi). Nota: non vi è accordo su queste sigle.

**INDICI DI TENDENZA CENTRALE** - valori "tipici" con cui possiamo riassumere i dati.

**Media:** la media (aritmetica) è l'indice di tendenza centrale per dati di intervallo (somma dei valori di ogni elemento della popolazione / numero di elementi)

**Mediana:** indice di tendenza centrale per dati ordinali (valore rispetto al quale il 50% dei valori risulta superiore e l'altro 50% inferiore)

**Moda:** indice di tendenza per dati nominali (categoria che si presenta con maggior frequenza) (se sono 2: distribuzione bimodale)

**INDICI DI DISPERSIONE** - indicano quanto i dati si raggruppano strettamente intorno alle misure di tendenza centrale.

**Deviazione standard:** radice quadrata della varianza (la **varianza** è la media del quadrato degli scarti di ogni valore rispetto alla media).

**Intervallo minimo-massimo:** differenza tra il valore massimo e quello minimo - è sempre un numero solo.

**Intervallo interquartile:** (midspread): differenza tra quartile superiore e inferiore (comprende il 50% centrale dei dati).

### STATISTICA INFERENZIALE **Stabilisce la probabilità che una deduzione basata sui dati (raccolti per un campione) sia corretta.**

Il **campione** è costituito dai soggetti sotto osservazione. La **popolazione** descrive gli ipotetici (e di solito infiniti) soggetti di interesse per lo studio.

Il punto chiave è la quantificazione del livello di imprecisione della stima basata sui dati ottenuti dal campione.

**Ipotesi nulla:** non c'è differenza tra il campione e la popolazione.

**Errore standard della media:** deviazione standard delle medie di campioni casuali (di una data numerosità) estratti dalla popolazione.

$P$  (oppure  $p$ ) = probabilità di sbagliare affermando che esiste una differenza reale (ossia rifiutando l'ipotesi nulla).

$p=0,05$ : c'è differenza, oppure si è verificata una coincidenza che non si verifica più di una volta ogni 20 esperimenti.

Per convenzione, se  $p<0,05$  la differenza è statisticamente significativa; se  $p<0,01$  la differenza è altamente significativa.

Una differenza statisticamente significativa può non essere clinicamente significativa.

Test a due code: valuta qualsiasi differenza tra i gruppi, indipendentemente dalla direzione della differenza stessa.

Intervallo di confidenza (C.I.) [intorno alla media]: limiti (inferiore e superiore) entro i quali è compresa con il (95)% di probabilità [la media della popolazione].

**Errore di tipo I:** affermare che esiste una differenza significativa quando non è vero. Misurato dalla probabilità alfa ( $\alpha$ ). Spesso si sceglie  $\alpha=0,05$

**Errore di tipo II:** affermare che non esiste una differenza significativa quando invece essa esiste. Misurato dalla probabilità beta ( $\beta$ ). Spesso si sceglie  $\beta=0,20$

**Potenza:** probabilità di concludere che c'è una differenza quando di fatto la differenza c'è davvero ( $=1-\beta$ ).

È in relazione con l'entità della differenza che si vuole rilevare e con la numerosità del campione.

Scelta del test: soggetti? → variabili? ↓	2 gruppi di soggetti diversi	3 o più gruppi di soggetti diversi	Un trattamento sugli stessi soggetti	Più trattamenti sugli stessi soggetti	Associazione tra 2 variabili	
<b>Variabile di intervallo</b> (e distribuzione normale)	Test $t$ per dati indipendenti ( <i>unpaired</i> )	Analisi della varianza (ANOVA)	Test $t$ per dati appaiati ( <i>paired</i> )	ANOVA per misure ripetute	Regressione lineare e correlazione di Pearson	
<b>Variabile nominale</b>	$\chi^2$ (chi quadro) Test esatto di Fisher	$\chi^2$ (chi quadro)	Test di McNemar	Test $Q$ di Cochran	-Rischio relativo (RR) -Odds ratio (OR)	(Prospettico) (Retrospectivo)
<b>Variabile ordinale</b>	Test di Mann-Whitney	Test di Kruskal-Wallis	Test di Wilcoxon	Test di Friedman	Correlazione dei ranghi di Spearman	

Fonti (sito del corso: <http://apollo11.isto.unibo.it/summa/it/metodo/stat.htm>)

Glantz SA. *Primer of Biostatistics*. McGraw-Hill, 2012 (7<sup>th</sup> Ed).

Glantz SA. *Statistica per discipline biomediche*. McGraw-Hill, 2007 (6<sup>a</sup> Ed).

Norman GR, Streiner DL. *Biostatistics: The Bare Essentials*. Pmhp USA, 2014 (4<sup>th</sup> Ed).

Norman GR, Streiner DL. *Biostatistica. Quello che avreste voluto sapere...* CEA, 2015.