

# Introduzione alla Statistica

Variabili  
Scale di Misura

Prof. Giuseppe Verlato

Sezione di Epidemiologia e Statistica Medica  
Università degli Studi di Verona

## La Metodologia Statistica

Disciplina, il cui corpus è costituito da metodi per trattare informazioni relative a fenomeni collettivi, che presentano caratteri di variabilità.

Il gruppo di riferimento viene detto **popolazione statistica**, a sua volta costituita da **unità statistiche o elementari**.

### A cosa serve la Statistica Medica nelle professioni sanitarie?

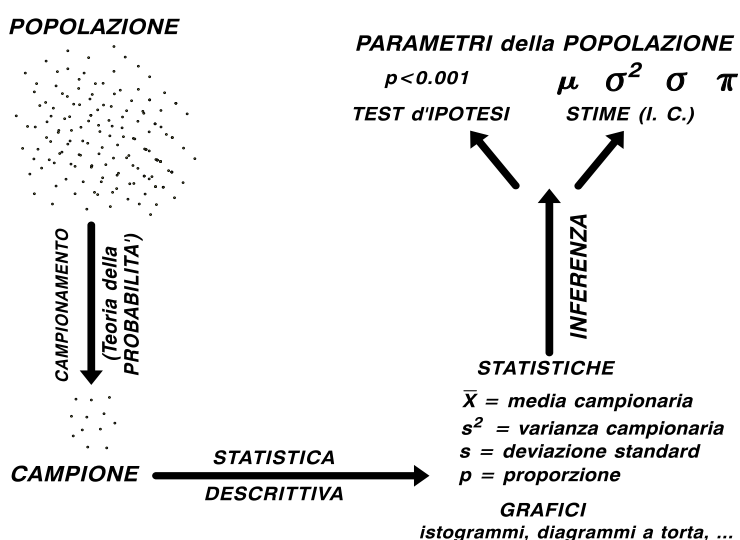
- 1) Per tenersi aggiornati studiando la letteratura scientifica corrente.
- 2) per riflettere sui nostri pazienti, sulle nostre casistiche!

**Perché dobbiamo riflettere sulle nostre casistiche? Per tenersi aggiornati, basta leggere quello che scrivono gli Americani, gli Inglesi, i Tedeschi, ...**

*Dobbiamo riflettere sulle nostre casistiche perchè:*

- 1) Perché i nostri pazienti sono spesso diversi dai pazienti stranieri: ad esempio, le carte del rischio cardiovascolare sono diverse per gli abitanti dell'Europa Occidentale e dell'Europa dell'Est. Rispetto agli Stati Uniti di America i diabetici di tipo II presentavano incrementi più contenuti nella glicemia, nell'indice di massa corporea e nella mortalità cardiovascolare (Zoppini G, Verlatto G. Mortalità nel diabete. *Il Diabete*, Dic 2004, 300-316).
- 2) Perché riflettere sul proprio operato in genere migliora la qualità delle prestazioni.
- 3) Per non accumulare ulteriore ritardo scientifico e tecnologico rispetto agli altri Paesi industrializzati.

## Programma del Corso di Statistica



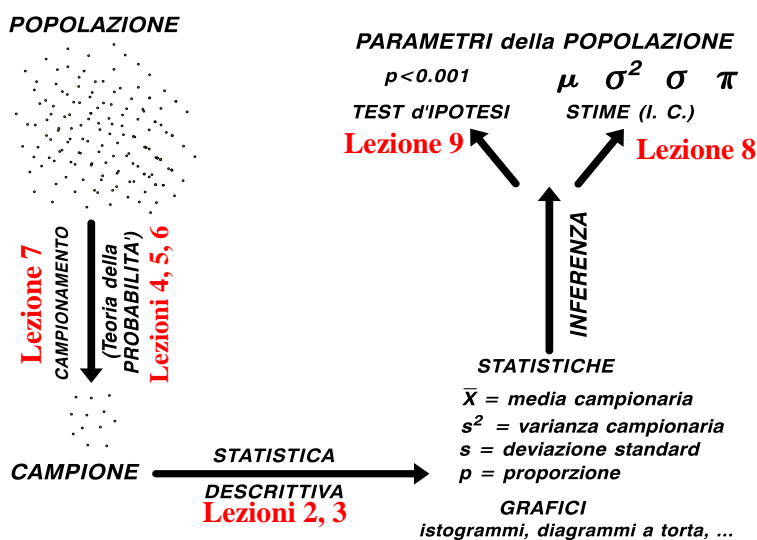
## ESEMPIO

Conoscere l'orientamento politico degli elettori italiani, che sono più di 51 milioni, è un procedimento estremamente dispendioso. Le indagini sull'intera popolazione vengono effettuate ogni 5 anni in corrispondenza delle elezioni politiche.

Nel periodo tra le elezioni, i partiti politici effettuano dei sondaggi. Delle ditte specializzate nei sondaggi di opinione estraggono dei **campioni rappresentativi** della popolazione italiana, i soggetti selezionati vengono intervistati e i risultati vengono sintetizzati attraverso la **statistica descrittiva**.

I risultati conseguiti sul campione vengono poi generalizzati alla popolazione di origine attraverso l'**inferenza**. In genere, per esprimere l'incertezza insita nel procedimento, le percentuali di elettori favorevoli ad un determinato partito non vengono espresse con dei valori precisi, ma con degli intervalli, detti **intervalli di confidenza**.

## Programma del Corso di Statistica



**La Statistica Descrittiva** fornisce una sintesi chiara e concisa della massa dei dati, in forma numerica e/o grafica.

**Il calcolo della probabilità** studia e descrive eventi aleatori (casuali). Un evento è aleatorio quando non si può prevedere con certezza se si avvererà o meno.

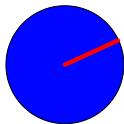
**La Statistica Inferenziale** consente di generalizzare le osservazioni effettuate su un campione, opportunamente sintetizzate attraverso la Statistica Descrittiva, alla popolazione di origine del campione.

## Costanti e Variabili

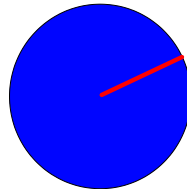
**COSTANTE** = caratteristica che non cambia nelle diverse unità di osservazione

**VARIABILE** = caratteristica che varia da un'unità di osservazione all'altra

$$\text{Area} = \pi * \text{raggio}^2$$



$$\text{Area} = \pi * \text{raggio}^2$$



Capacità vitale attesa

$$\text{M: Capacità vitale (l)} = 5.76 * \text{altezza (m)} - 0.026 * \text{età (anni)} - 4.34$$

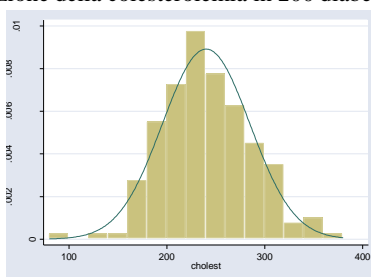
$$\text{F: Capacità vitale (l)} = 4.43 * \text{altezza (m)} - 0.026 * \text{età (anni)} - 2.89$$

## Principali tipi di scale di misura

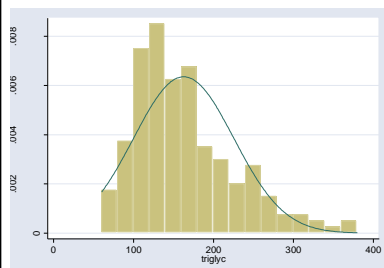
tipo di variabile	operazioni consentite	esempi	statistica descrittiva	statistica inferenziale
Nominale	$= \neq$	Sesso, colore dei capelli, tipo di diabete	Proporzione (prevalenza)	Chi-quadrato, Mantel-Haenszel, modello logistico
Ordinale	$= \neq < >$	Intensità del dolore, profondità del coma, stadio di un tumore	Mediana, range, distanza interquartile	Test non-parametrici
Quantitativa	$= \neq < >$	<b>Variabili con distribuzione asimmetrica:</b> Tempo di sopravvivenza, n° linfonodi metastatici	Mediana, range, distanza interquartile	Test non-parametrici trasformazioni per normalizzare i dati
	$+ - * /$	<b>Variabili con distribuzione simmetrica:</b> glicemia, pressione arteriosa, indice di massa corporea	Media, deviazione standard	Test t, ANOVA, regressione

Riconoscere il tipo di variabile è fondamentale per scegliere correttamente la statistica descrittiva e inferenziale.

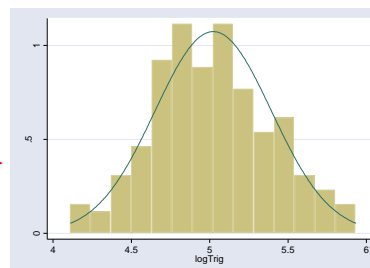
Distribuzione della colesterolemia in 200 diabetici di tipo 2



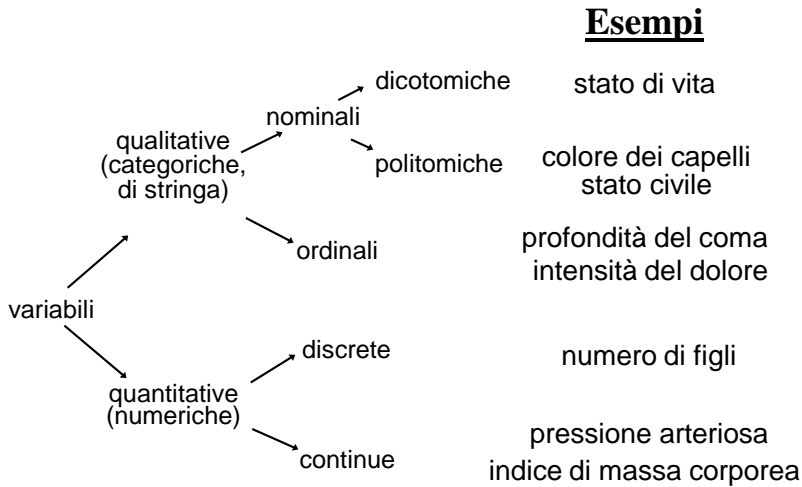
Distribuzione della trigliceridemia in 200 diabetici di tipo 2



trasformazione  
logaritmica



# L'albero delle Variabili



Matrice dei dati

Ogni riga (o più righe vicine) contiene i dati di un soggetto

Ogni colonna (o più colonne vicine) contiene i valori assunti da una variabile

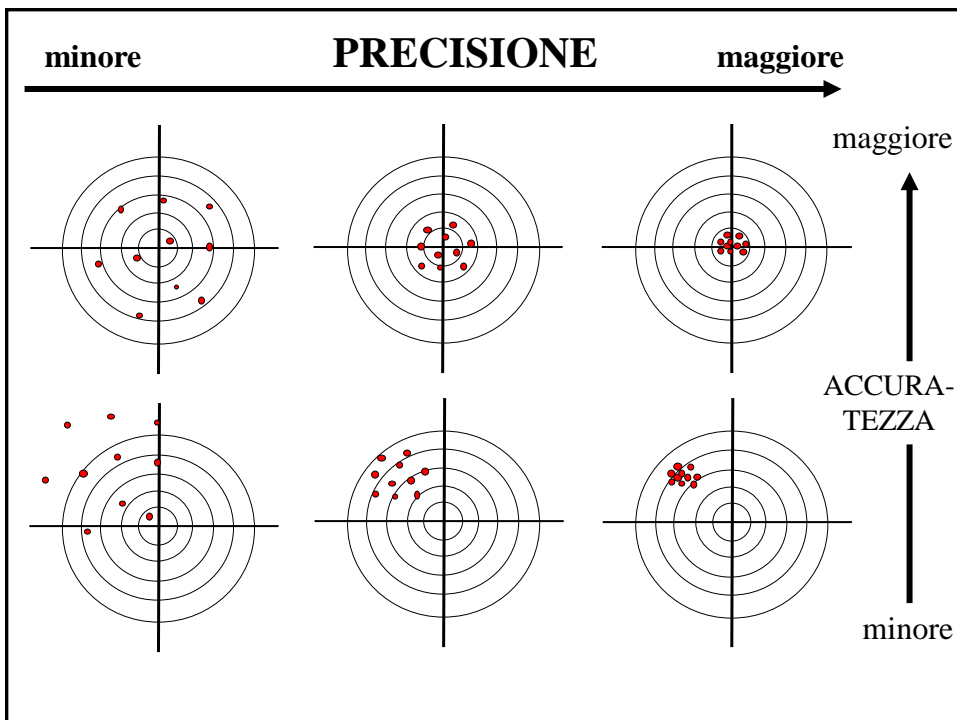
Ci sono 20 variabili dicotomiche, che possono assumere solo 2 valori: 0 = risposta errata  
1 = risposta corretta

Risposte alle domande di Statistica

Studente (nome fittizio)	Stat. descrittiva									Probabilità							Inferenza			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Claudio	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	1	0	0	1	0
Albino	1	1	1	0	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0
Giuseppe	1	1	1	1	1	0	1	1	0	0	0	1	0	1	0	1	1	0	0	0
Antonio	1	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1
Silvia	1	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1
Laura	1	1	1	0	0	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1
Pierina	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
Francesca	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1	0	1
Simone	1	1	1	1	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	1
Elisabetta	0	1	1	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1	1	0
Giovanni	1	1	0	1	0	0	1	1	0	0	1	1	0	0	0	0	0	0	1	0
Andrea	1	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	0	0	0	0

## Principali Proprietà di una Misura

- ACCURATEZZA: Capacità di prevenire/evitare l'errore sistematico o distorsione (*bias*)
- PRECISIONE: Capacità di ridurre/minimizzare la variabilità casuale; è spesso valutata sulla base del coefficiente di variazione. Ripetibilità = Capacità di ottenere lo stesso valore nelle stesse condizioni (stesso operatore, stesso laboratorio, ...). Riproducibilità = Capacità di ottenere lo stesso valore in condizioni differenti.
- VALIDITA': Rappresenta la capacità di una misura di restituire il valore reale del fenomeno per valutare il quale è impiegata





## VALIDAZIONE dello STRUMENTO di MISURA

Lo strumento di misura viene confrontato con un "gold standard".

Ad esempio le risposte ad un questionario sul fumo possono essere confrontate con la concentrazione di CO o di cotinina (metabolita della nicotina) nel sangue (Olivieri et al, 2002).

Analogamente le risposte a un questionario sull'asma possono essere confrontate con le prove allergologiche, la spirometria o la diagnosi clinica di asma (de Marco et al, 1998).

### **Bibliografia**

de Marco R, Cerveri I, Bugiani M, Ferrari M, Verlato G (1998) An undetected burden of asthma in Italy: the relationship between clinical and epidemiological diagnosis of asthma. *Eur Respir J*, 11: 599-605

Olivieri M, Poli A, Zuccaro P, Ferrari M, Lampronti G, de Marco R, Lo Cascio V, Pacifici R (2002) Tobacco smoke exposure and serum cotinine in a random sample of adults living in Verona, Italy. *Arch Environ Health*, 57: 355-359