

Introduzione alla Statistica

(Metodo delle Scienze Empiriche)

Variabili
Scale di Misura

Sezione di Epidemiologia e Statistica Medica
Università degli Studi di Verona

La Metodologia Statistica

Disciplina, il cui corpus è costituito da metodi, e cioè da un insieme di procedure utili al trattamento delle informazioni relative a fenomeni collettivi, che per loro natura si manifestano con determinazioni tipicamente non costanti



**MODELLO ASTRATTO DI RAGIONAMENTO, CHE
TRAVALICA L'AMBITO SPECIFICO CUI VIENE
APPLICATO**

La Disciplina Statistica

Oggetto della Statistica sono dunque quei fenomeni che presentano caratteri di variabilità all'interno di un collettivo di riferimento (**popolazione statistica**), costituito da **unità statistiche o elementari**

A cosa serve la Statistica Medica nelle professioni sanitarie?

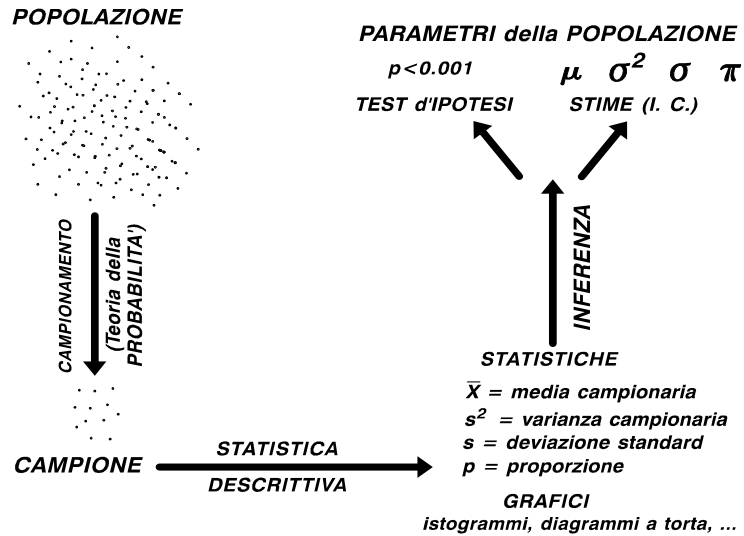
- 1) *Per tenersi aggiornati studiando la letteratura scientifica corrente.*
- 2) *per riflettere sui nostri pazienti, sulle nostre casistiche!*

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 del Sud e dell'Europa del Nord. Rispetto agli Stati Uniti di America i diabetici di tipo II presentano incrementi più contenuti nella glicemia, nell'indice di massa corporea e nella mortalità cardiovascolare (Zoppini G, Verlato G. Mortalità nel diabete. Il Diabete, Dicembre 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 40 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**.

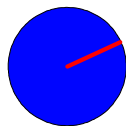
I risultati conseguiti sul campione vengono poi generalizzati alla popolazione di origine attraverso l'**inferenza**. In genere, per esprimere l'incertezza insita nel procedimento, le percentuali di elettori favorevoli ad un determinato partito non vengono espresse con dei valori precisi, ma con degli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

Costanti e Variabili

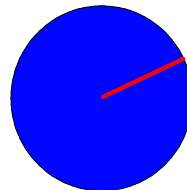
COSTANTE = caratteristica che non cambia nelle diverse unità di osservazione

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

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

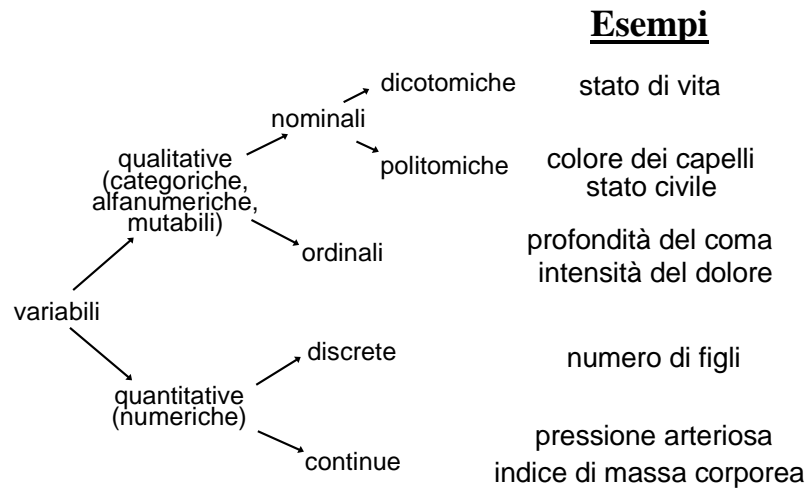
F: $\text{Capacità vitale (l)} = 4.43 * \text{altezza (m)} - 0.026 * \text{età (anni)} - 2.89$

Tipi di Variabili

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	Test non-parametrici
Quantitativa	$\neq < >$	Variabili con distribuzione asimmetrica: Tempo di sopravvivenza, n° linfonodi metastatici	Mediana, range, distanza interquartile	Test non-parametrici trasformazioni per normalizzare i dati
	$+ - * /$	Variabili con distribuzione simmetrica: 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.

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

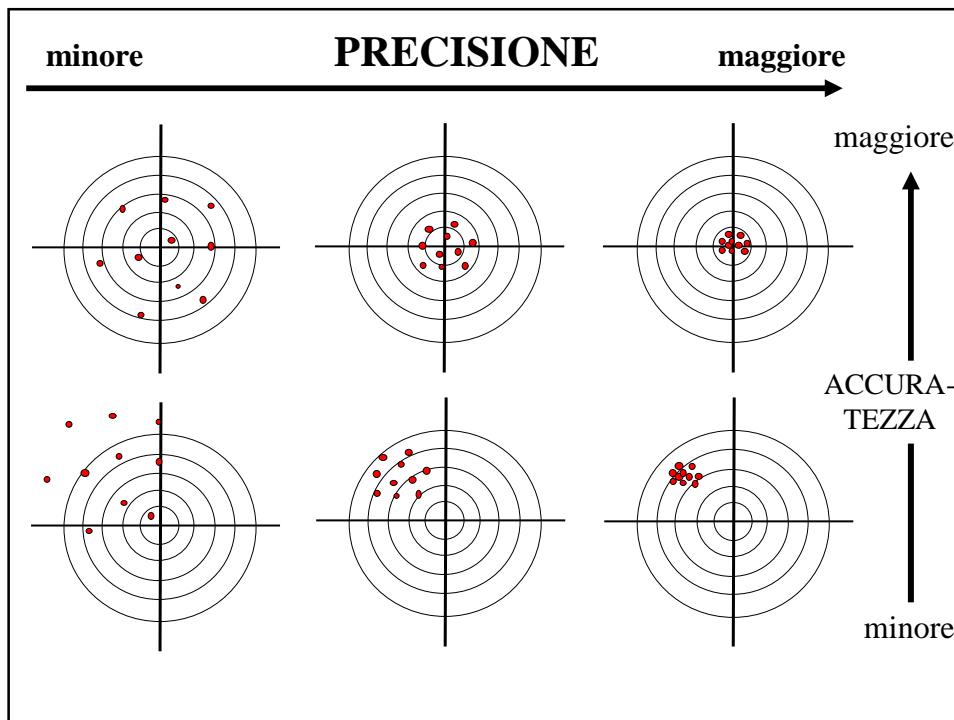
Ci sono 20 variabili dicotomiche, che possono assumere solo 2 valori:
 0 = risposta errata
 1 = risposta corretta

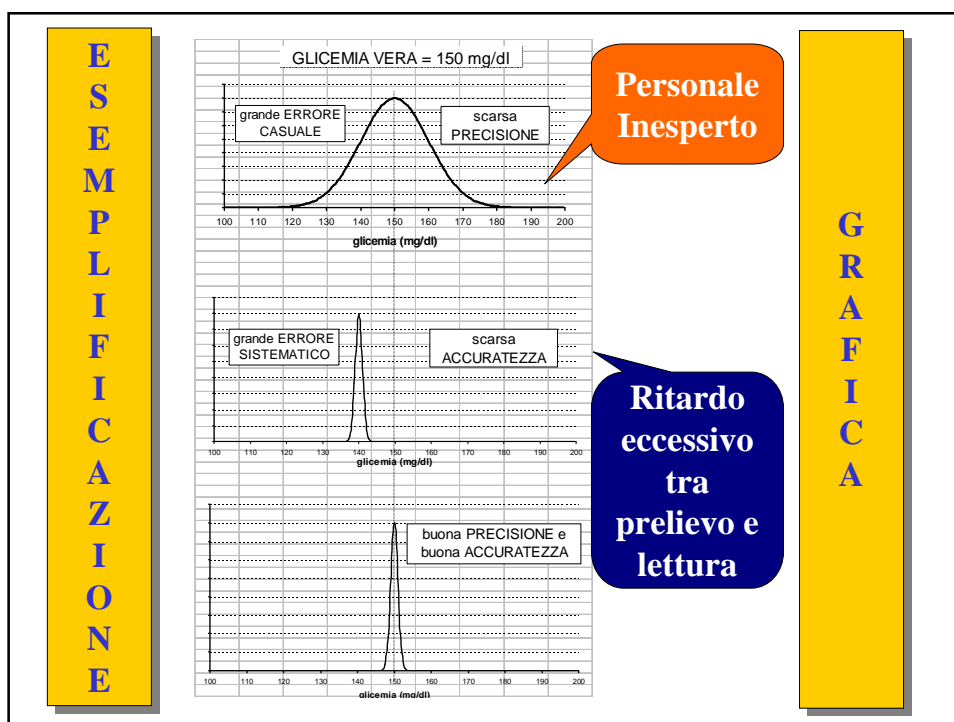
Risposte alle domande di Statistica

Studente	Stat. descrittiva								Probabilità								Inferenza			
(nome fittizio)	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
E' legata all'errore sistematico o distorsione (*bias*)
- PRECISIONE
E' legata alla variabilità accidentale o errore casuale (*errore standard*)
- RIPETIBILITA' / AFFIDABILITA'
La misurazione fornisce risultati sovrapponibili, quando ripetuta in circostanza analoghe
- 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