

Statistica

A.A. 2011/12



Qualche informazione di carattere organizzativo

Docenti: Alessandra Bianchi e Paolo Dai Pra

Orario di ricevimento (Dai Pra): Martedì ore 15:00 e per appuntamento (saranno possibili cambiamenti (vedi pag. web del corso))

email: daipra@math.unipd.it

Pagina web: www.math.unipd.it/~daipra/didattica/bio12

Ufficio: Torre Archimede, V piano, corridoio CD

Testo: Sheldon M.Ross, Introduzione alla Statistica, Apogeo.

Modalità d'esame: Prova scritta.



Cos'è la Statistica?

Nel linguaggio comune tale termine viene associato a liste di numeri, dati, tabelle, grafici....

In altre parole, ai metodi per [organizzare](#), [sintetizzare](#) e [presentare](#) i dati provenienti da esperimenti, indagini....

Questi metodi costituiscono la [Statistica Descrittiva](#).

Lo scopo di questo corso è fornire un'introduzione ad una statistica più moderna e [attiva](#).

FARE STATISTICA



Consideriamo la seguente osservazione, quasi ovvia: la maggior parte dei dati sperimentali su cui si basano teorie scientifiche, terapie mediche, politiche economiche o ambientali...., sono soggetti ad [incertezza](#) ineliminabile. Le cause di tale incertezza sono le più varie, e spesso di difficile identificazione. Alcune di queste cause sono:

- ▶ imprecisione degli strumenti e dei metodi di misura;
- ▶ impossibilità di ottenere misure per tutti gli individui di una [popolazione](#) numerosa, e pertanto la necessità di “scegliere” un piccolo gruppo rappresentativo ([campionamento](#));
- ▶ il fatto che alcuni fenomeni naturali, come la trasmissione ereditaria dei caratteri o lo sviluppo di una malattia in un determinato individuo, sono [intrinsecamente soggetti ad una componente casuale](#).

Lo scopo della Statistica è la gestione [ottimale](#) dell'incertezza, cioè il ricavare da dati soggetti ad incertezza la maggiore [informazione possibile](#).



Lo scopo ultimo del corso è di fornire le nozioni essenziali del procedimento statistico di maggior utilizzo nelle Scienze, in particolare nelle Scienze della Vita: [la verifica delle ipotesi](#).

Per avere un'idea di cosa si tratti, usiamo l'analogia con i [procedimenti giudiziari](#). Gli ingredienti essenziali di un processo sono:

- ▶ un'[ipotesi investigativa](#), cioè un [crimine](#) e un [imputato](#);
- ▶ i dati di un'indagine;
- ▶ il verdetto.

Benché l'assoluta certezza viene difficilmente raggiunta anche al termine di accurate indagini, il compito della giuria è di “minimizzare” l'impatto di eventuali errori: l'imputato viene dichiarato colpevole se i dati forniti dalle indagini ne mostrano la colpevolezza [oltre ogni ragionevole dubbio](#).

Una [verifica di ipotesi](#) ha una struttura analoga:

- ▶ un'[ipotesi statistica](#); ad esempio: i pesticidi (imputato) favoriscono l'insorgere della leucemia (il crimine);
- ▶ i dati di una serie di esperimenti;
- ▶ la conclusione, cioè l'accoglimento o il rifiuto dell'ipotesi.

La formulazione precisa di un'ipotesi statistica, così come la procedura di interpretazione dei dati che garantisce di “minimizzare l'impatto di eventuali errori”, richiedono il linguaggio del [Calcolo della Probabilità](#), la teoria matematica che tratta l'incertezza.

Statistica in Biologia?

La disponibilità di un enorme numero di dati di natura biologica e biomedica ha reso indispensabile l'uso della statistica nelle Scienze Biologiche.

Le principali applicazioni sono:

- ▶ in ambito medico: [epidemiologia](#), [test clinici su farmaci](#)...
- ▶ in Genetica: ed esempio studio delle relazioni tra [genotipo](#) e [fenotipo](#), con applicazioni in medicina, agricoltura...
- ▶ nelle Scienze Ambientali ed Ecologia: studio dei fattori che determinano cambiamenti nell'ambiente, ad esempio clima, inquinamento....
- ▶ nello studio delle sequenze di geni nel DNA, in collaborazione con la [Bioinformatica](#).

Statistica descrittiva

Obiettivo: presentare i dati di una, ricerca, indagine..... in modo sintetico attraverso grafici, tabelle e altre forme di sintesi.

I dati si possono riferire a una o più quantità misurate (*variabili*). I dati relativi ad una variabile si presentano come una sequenza x_1, x_2, \dots, x_n di valori della variabile (*campione*), ottenuti in n "misurazioni".

Le variabili vengono divise in due grandi categorie: le variabili *numeriche*, che assumono cioè valori numerici, e le variabili *categoriche* (tutte le altre)

Statistica descrittiva per variabili categoriche

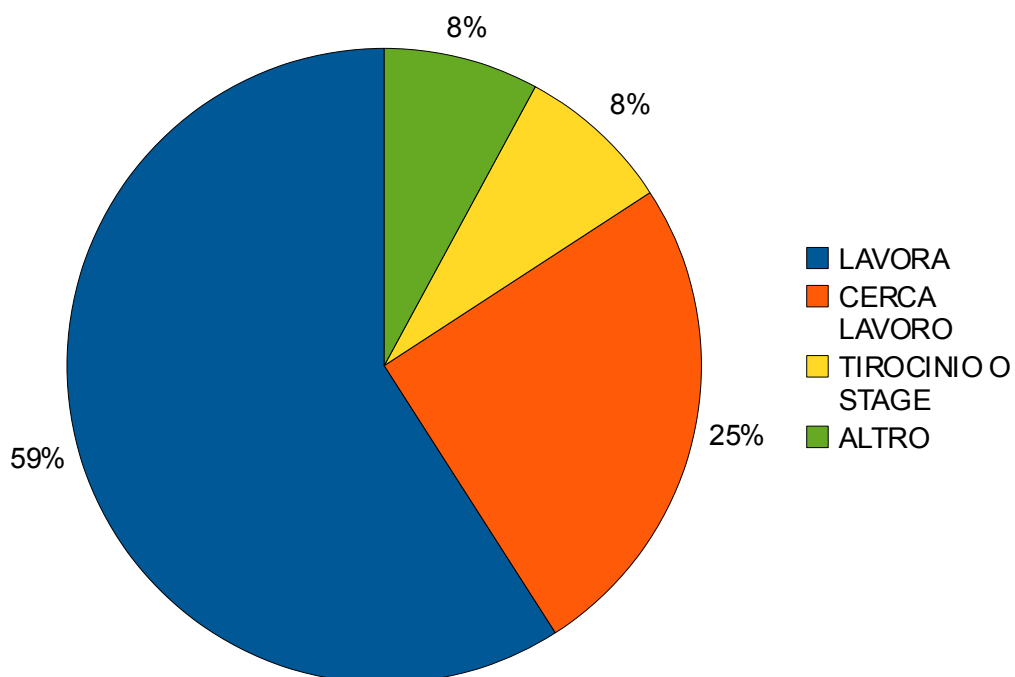
I valori possibili di una variabile categorica vengono chiamati categorie. I dati vengono rappresentati tramite tabelle e grafici. Lo illustro con un esempio, tratto da un'indagine sullo stato occupazionale dei laureati della Facoltà di Scienze MM FF NN a 6 mesi dalla laurea.

	Frequenza assoluta	Frequenza relativa	Frequenza percentuale
LAVORA	127	0.5907	59.07%
CERCA LAVORO	54	0.2512	25.12%
TIROCINIO O STAGE	17	0.0791	7.91%
ALTRO	17	0.0791	7.91%
TOTALI	215	1.0000	100.00%

Frequenza assoluta = numero di occorrenze di un dato valore

Frequenza relativa = $\frac{\text{frequenza assoluta}}{\text{numero di osservazioni}}$

Frequenza percentuale = Frequenza relativa $\times 100\%$



Più variabili categoriche relative ad una medesima popolazione possono venire usate per studi di *correlazione*.

Ad esempio i laureati di cui sopra provenivano dai Corsi di Laurea in Matematica, Fisica e Biologia

	Matematica	Fisica	Biologia	TOTALI
LAVORA	18	17	21	56
CERCA LAVORO	8	7	9	24
TIROCINIO O STAGE	2	4	7	13
ALTRO	1	2	3	6
TOTALI	29	30	40	99

Da tabelle di questo tipo, dette *tabelle di contingenza*, si vogliono rilevare eventuali *correlazioni* tra le due variabili: in questo caso se la condizione occupazionale sia sostanzialmente la stessa nei tre corsi di laurea, o vi siano differenze significative.

Questo tipo di analisi verrà affrontata verso la fine di questo corso.