

# STATISTICA MEDICA

**Dott.ssa Marta Di Nicola**

N.P.D. 3° Blocco 2° piano

0871-3554007  
m.dinicola@unich.it


<http://www.biostatistica.unich.it>

Dott.ssa Marta Di Nicola

## Alcuni obiettivi specifici

- Descrivere ed interpretare una tabella o un grafico
- Costruire una tabella o un grafico per rappresentare un fenomeno biomedico
- Descrivere un fenomeno biomedico per mezzo di misure di sintesi
- Analizzare con strumenti semplici la relazione fra due variabili
- Spiegare i concetti fondamentali dell'inferenza: valori attesi, stima, intervallo di confidenza, precisione, significatività statistica, errore di primo e secondo tipo
- Utilizzare il calcolo statistico per il confronto di due o più gruppi

Dott.ssa Marta Di Nicola



M Pagano, K Gauvreau. BIOSTATISTICA. IDELSON-GNOCCHI

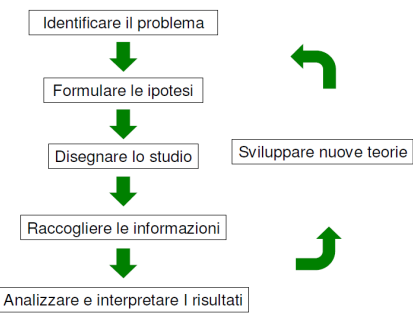
Dott.ssa Marta Di Nicola

## Perché la Statistica Medica è necessaria?

- per leggere, comprendere ed analizzare criticamente relazioni e documenti scientifici;
- per avere la possibilità di effettuare in proprio ricerche che comportino l'acquisizione, l'elaborazione e l'analisi dei dati.

Dott.ssa Marta Di Nicola


## Informazioni, nuove conoscenze, decisioni



Dott.ssa Marta Di Nicola

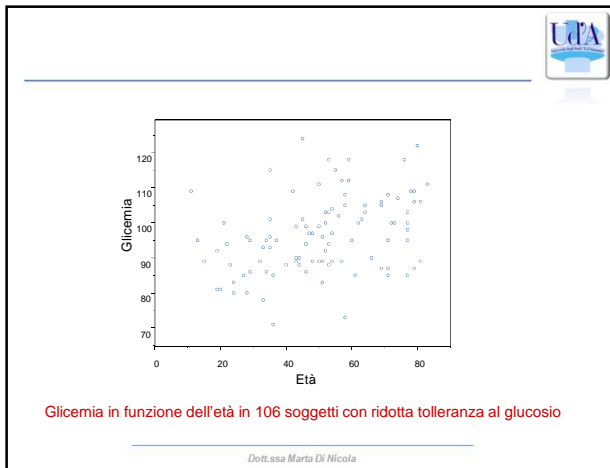
## LA STATISTICA

Sai ched'è la statistica? E' 'na cosa che serve pé fa' un conto in generale de la gente che nasce, che sta male, che more, che va in carcere e che sposa. Ma pe' me la statistica curiosa è dove c'entra la percentuale, pe' via che, li, la media è sempre eguale puro co' la persona bisognosa. Me spiego: da li conti che se fanno seconno le statistiche d'adesso risurta che te tocca un pollo all'anno: e, se nun entra ne le spese tue, t'entra ne la statistica lo stesso perché c'è un altro che ne magna due.



TRILUSSA (1871 - 1950)

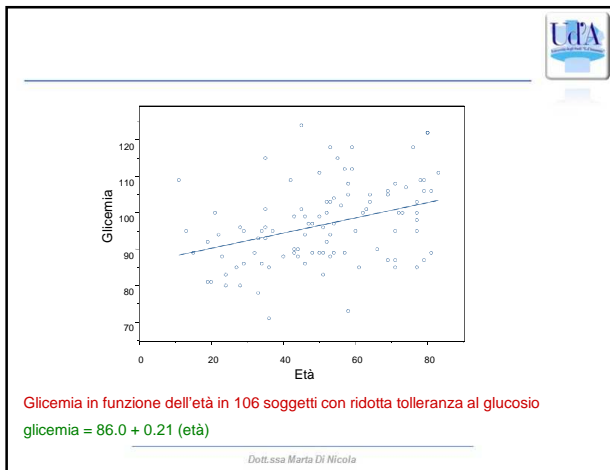
Dott.ssa Marta Di Nicola



### LA STATISTICA

- La Statistica ha come scopo la *conoscenza quantitativa dei fenomeni collettivi*. L'analisi statistica mira ad individuare **modelli** di interpretazione della realtà, attraverso canoni e tecniche che sono astrazioni, semplificazioni di una moltitudine di aspetti e di manifestazioni del reale.
- E' costituita da un insieme dei metodi che consentono di raccogliere, ordinare, riassumere, presentare ed analizzare dati e informazioni, trarne valide conclusioni e prendere decisioni sulla base di tali analisi e risultati.

Dott.ssa Marta Di Nicola



### Statistica e matematica

<b>MATEMATICA</b>	<b>STATISTICA</b>
<ul style="list-style-type: none"> <li>funzione nota <math>y = a + bx</math></li> <li>parametri (a, b) noti</li> <li>x o y ignota</li> </ul> <p>Stime dei parametri</p>	<ul style="list-style-type: none"> <li>modello ignoto <i>lineare?</i> <math>Y = \alpha + \beta x</math></li> <li>parametri (<math>\alpha, \beta</math>) ignoti</li> <li>x o y entrambe note <math>y_i = \alpha + \beta x + \varepsilon_i</math> <math>y_i = \alpha + \beta x</math></li> </ul>

Dott.ssa Marta Di Nicola

### Parametri

Quantità che definiscono un modello teorico. Non si riferiscono ai dati effettivamente osservati, ma possono essere usati per predire il comportamento atteso nel singolo soggetto

### Stime

Quantità, ottenute a partire dai dati effettivamente osservati, per stimare ('indovinare') i valori dei parametri

Dott.ssa Marta Di Nicola

### Statistica

L'obiettivo è riconoscere le cause rilevanti di variabilità

- dimostrare che esiste un'associazione tra fattore studiato (esposizione) e fenomeno in esame (esito)
- dimostrare che l'eventuale associazione osservata non è in realtà un artefatto
- stimare l'entità dell'associazione fra esposizione ed esito

Dott.ssa Marta Di Nicola

## Statistica medica

The diagram shows two blue arrows pointing downwards from the title 'Statistica medica' to two boxes. The left box is titled 'Funzione Descrittiva' and the right box is titled 'Funzione Inferenziale'. Each box contains a brief description of its function.

### Funzione Descrittiva

Offre il metodo per riassumere le informazioni in modo da renderle utilizzabili più facilmente. Riduce i dati in forma maneggevole, sostituendo a molti numeri poche misure.

### Funzione Inferenziale

Permette di generalizzare le informazioni, ricavando proprietà e leggi generali sulla base di dati rilevati solamente su una parte (campione) della popolazione.

Dott.ssa Marta Di Nicola

## UNITA' STATISTICA

**DEFINIZIONE:** Ogni elemento o caso appartenente alla popolazione oggetto diretto della osservazione da cui si raccolgono i dati.

**Esempi:** un singolo paziente o soggetto;  
una ASL;  
un occhio;  
un campione di sangue;  
etc

Dott.ssa Marta Di Nicola

## CAMPIONE E UNIVERSO

Un **campione rappresentativo** è un insieme di unità statistiche tratte da un **universo** (o **popolazione**). Un **universo** consiste della **totalità** delle **unità statistiche** che posseggono determinate caratteristiche. (Non necessariamente una popolazione deve essere composta da un numero elevato di elementi).

**Esempi:**

- **Universo:** tutti i pazienti adulti con una determinata malattia.
- **Campione:** 120 pazienti con quella malattia, inclusi in una sperimentazione clinica.
  
- **Universo:** un lotto di 5000 impianti.
- **Campione:** 10 impianti di quel lotto sottoposti a controllo di qualità.

Il campione è soltanto una parte del tutto.

Dott.ssa Marta Di Nicola

## VARIABILE STATISTICA

**DEFINIZIONE:** Ciascuna informazione, caratteristica che verrà rilevata sulle unità statistiche sarà chiamata **VARIABILE**.

Le "categorie" della variabile statistica in osservazione costituiscono le MODALITA' della variabile.

I dati sperimentali (variabili) si presentano sotto differenti forme, essi possono essere sia di tipo quantitativo sia di tipo qualitativo, ed essere espressi o con scale continue o con scale discrete.

Dott.ssa Marta Di Nicola

## GLOSSARIO

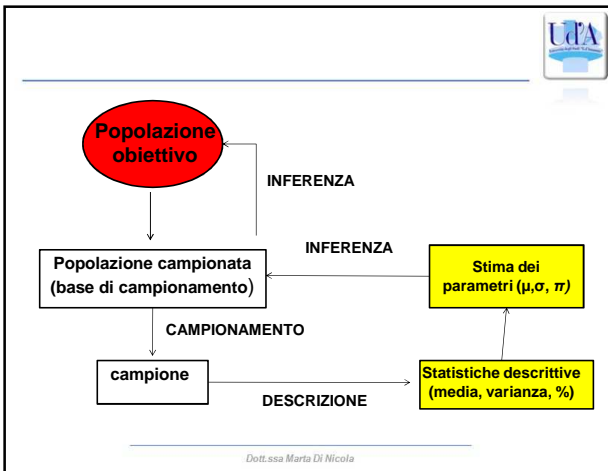
- **POPOLAZIONE:** l'insieme di tutte le unità statistiche oggetto dell'osservazione (es.: medici, paramedici, studenti, diabetici, obesi, addetti all'agricoltura...).
- **CAMPIONE:** la parte delle unità statistiche sottoposte all'osservazione, all'esperimento, etc.
- **UNITA' STATISTICA:** per ogni elemento o caso appartenente alla popolazione oggetto diretto della osservazione da cui si raccolgono i dati.

Dott.ssa Marta Di Nicola

## GLOSSARIO

- **CARATTERE (O VARIABILE):** la caratteristica (attributo o misura) osservata sulla unità statistica.
- **MODALITA':** ogni diversa presentazione del carattere o variabile osservata su ciascuna unità statistica.
- **FREQUENZA:** numero di volte che si presenta una data modalità.

Dott.ssa Marta Di Nicola

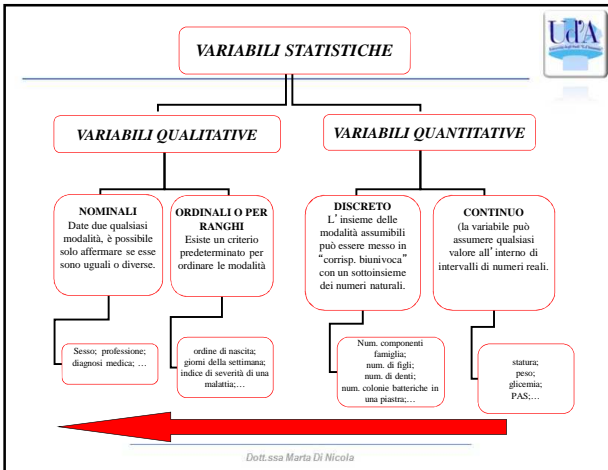


### Esempi

Le **variabili** sesso, età, peso, pressione arteriosa, etc (di pazienti inclusi in uno studio) hanno come **modalità**:

- **maschio** o **femmina** per la variabile "sesso";
- **anni**, per la variabile "età";
- **Kg**, per il "peso corporeo",
- **mmHg**, per la "pressione arteriosa"
- **A, AB, B, 0** per il "gruppo sanguigno",
- **elementare, media inferiore, media superiore, università**, per la variabile "titolo di studio"

Dott.ssa Marta Di Nicola



## LA SINTESI DEI DATI

Dott.ssa Marta Di Nicola


**Esempio 1.** Su un campione di pazienti si rilevino le caratteristiche: sesso, età, altezza, peso, pressione arteriosa sistolica (PAS), tasso glicemico.

nome: Rossi Amerigo	nome: Bianchi Paolo
sesso: maschio	sesso: maschio
età: 32	età: 47
altezza: 172 cm.	altezza: 170 cm.
peso: 64 Kg.	peso: 80 Kg.
PAS: 140 mm Hg.	PAS: 148 mm Hg.
glicemia: 190 mg/dl	glicemia: 180 mg/dl

Dott.ssa Marta Di Nicola

nome: Valenzi Alberica	nome: Alinori Alfonso
sesso: femmina	sesso: maschio
età: 45	età: 27
altezza: 168 cm.	altezza: 183 cm.
peso: 51 Kg.	peso: 85 Kg.
PAS: 125 mm Hg.	PAS: 138 mm Hg.
glicemia: 150 mg/dl	glicemia: 170 mg/dl


Dott.ssa Marta Di Nicola



Le informazioni raccolte per essere "trattate" da un computer devono essere organizzate in strutture chiamate comunemente  
**Data Base o File Dati.**


Le informazioni vengono, comunemente, organizzate per riga, cioè su ogni riga, consecutivamente, vengono elencati i dati relativi ad un soggetto.

Dott.ssa Marta Di Nicola



A	B	C	D	E	F	G	H	I	J
ID	Sesso	Età (anni)	Numero figli	Fumatore (0=No; 1=Si)	BMI (kg/m <sup>2</sup> )	Massa grassa (%)	Massa magra (%)	PAS (mm/Hg)	PAD (mm/Hg)
1	M	58	0	0	22	3.83	29.94	130	100
2	M	57	0	0	25	6.44	21.68	131	90
3	M	51	1	0	40	4	27.62	128	90
4	F	36	3	0	38	14.03	32.91	110	80
5	M	50	2	0	35	9.43	28.6	136	85
6	M	64	2	1	29	9.42	26.43	149	100
7	F	39	1	1	24	8.41	26.99	120	80
8	F	38	1	0	18	20.23	33.85	115	75
9	F	70	0	1	19	3.99	24.49	95	50
10	M	49	4	1	31	7.29	30.37	100	40
11	M	53	2	1	30	11.18	26.83	105	40
12	F	56	3	1	22	5.84	26.53	100	60
13	F	54	5	1	39	4.67	28.91	95	45
14	F	45	0	0	24	13.24	27.08	120	90
15	M	51	0	0	25	3.79	29.52	120	80
16	F	54	1	0	18	3.28	30.82	130	70
17	F	39	1	0	20	7.22	27.33	135	70
18	M	68	2	1	29	25.72	31.84	90	60
19	M	61	3	0	21	5.12	24.82	95	60
20	M	59	2	1	22	5.07	25.7	100	80
21	M	60	2	0	20	5.03	30.02	115	75

Dott.ssa Marta Di Nicola




**DISTRIBUZIONI SEMPLICI DI FREQUENZE**

I dati (cioè le informazioni raccolte) spesso sono di non immediata lettura.

Per questo si procede ad una sistematizzazione e sintesi delle informazioni raccolte, cioè alla loro **tabulazione**. Per ogni variabile si calcolano le **frequenze assolute (f.a.)** che rappresentano il numero di u.s. che presentano una stessa modalità del carattere.

Dott.ssa Marta Di Nicola



**Alcune distribuzioni semplici di frequenze**

**TABLE 2.4**  
Cases of Kaposi's sarcoma for the first 2560 AIDS patients reported to the Centers for Disease Control in Atlanta, Georgia

Kaposi's Sarcoma	Number of Individuals
Yes	246
No	2314

**TABLE 2.5**  
Cigarette consumption per person aged 18 or older, United States, 1900-1990


Year	Number of Cigarettes
1900	54
1910	151
1920	665
1930	1485
1940	1976
1950	3522
1960	4171
1970	3985
1980	3851
1990	2828

**TABLE 2.6**  
Absolute frequencies of serum cholesterol levels for 1067 U.S. males, aged 25 to 34 years, 1976-1980

Cholesterol Level (mg/100 ml)	Number of Men
80-119	13
120-159	150
160-199	442
200-239	299
240-279	115
280-319	34
320-359	9
360-399	5
Total	1067

**Numerosità campionaria n=2560**

Dott.ssa Marta Di Nicola




**Alcune distribuzioni semplici di frequenze**

**TABLE 2.7**  
Absolute and relative frequencies of serum cholesterol levels for 2294 U.S. males, 1976-1980

Cholesterol Level (mg/100 ml)	Ages 25-34		Ages 55-64	
	Number of Men	Relative Frequency (%)	Number of Men	Relative Frequency (%)
80-119	13	1.2	5	0.4
120-159	150	14.1	48	3.9
160-199	442	41.4	265	21.6
200-239	299	28.0	458	37.3
240-279	115	10.8	281	22.9
280-319	34	3.2	128	10.4
320-359	9	0.8	35	2.9
360-399	5	0.5	7	0.6
Total	1067	100.0	1227	100.0

Dott.ssa Marta Di Nicola



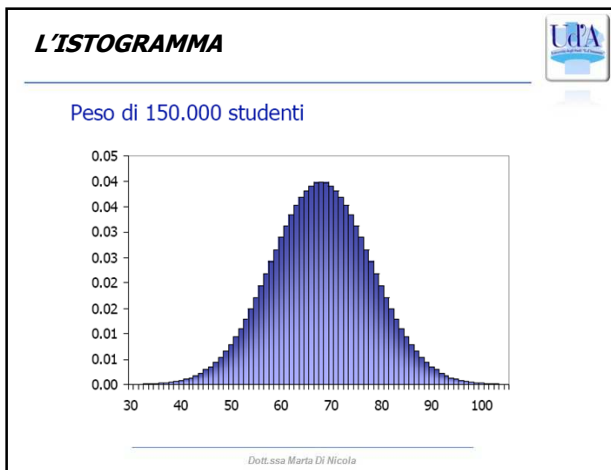
**Alcune distribuzioni semplici di frequenze**

**TABLE 2.8**  
Relative and cumulative relative frequencies of serum cholesterol levels for 2294 U.S. males, 1976-1980

Cholesterol Level (mg/100 ml)	Ages 25-34		Ages 55-64	
	Relative Frequency (%)	Cumulative Relative Frequency (%)	Relative Frequency (%)	Cumulative Relative Frequency (%)
80-119	1.2	1.2	0.4	0.4
120-159	14.1	15.3	3.9	4.3
160-199	41.4	56.7	21.6	25.9
200-239	28.0	84.7	37.3	63.2
240-279	10.8	95.5	22.9	86.1
280-319	3.2	98.7	10.4	96.5
320-359	0.8	99.5	2.9	99.4
360-399	0.5	100.0	0.6	100.0

**15.3=1.2+14.1**

Dott.ssa Marta Di Nicola



### Distribuzione doppia di frequenze assolute

BCO	Fumatori	Non fumatori	Totale
SI	160	100	260
NO	120	70	190
Totale	280	170	450

Dott.ssa Marta Di Nicola

Ci accorgiamo che il confronto **non** può essere effettuato solo con le f.a. in quanto esse si riferiscono a collettivi di numerosità diversa.

Dott.ssa Marta Di Nicola

Se vogliamo confrontare le frequenze le dobbiamo "depurare" dalla numerosità del collettivo; ciò lo si fa dividendo le f.a. per la numerosità (N) della popolazione e moltiplicando per 100 (cioè facendo riferimento ad una ipotetica popolazione di 100 unità).

Le frequenze così calcolate sono le **frequenze percentuali (f.%)**

Dott.ssa Marta Di Nicola

### Distribuzione doppia di frequenze percentuali


BCO	Fumatori		Non fumatori	
	f.a.	f.a. %	f.a.	f.a. %
SI	160	57.1	100	58.8
NO	120	42.8	70	41.2
Totale	280	100	170	100

Dott.ssa Marta Di Nicola

### Distrib. di frequenze assolute, relative e cumulate

Età	f.a.	f%	f.a.cum	f% cum
17	30	13.6	30	13.6
18	60	27.3	90	40.9
19	120	54.6	210	95.5
20	10	4.5	220	100
Totale	220	100		


Dott.ssa Marta Di Nicola



Le frequenze cumulate indicano quante u.s. si presentano **fino a** quella modalità.

Ha senso calcolare le f.cum solamente per le variabili quantitative o qualitative ordinabili.

Dott.ssa Marta Di Nicola




### I GRAFICI STATISTICI

Scopo dei grafici è quello di rendere l'informazione contenuta in una serie di dati:

- ✓ di più facile comprensione;
- ✓ di più diretta lettura.

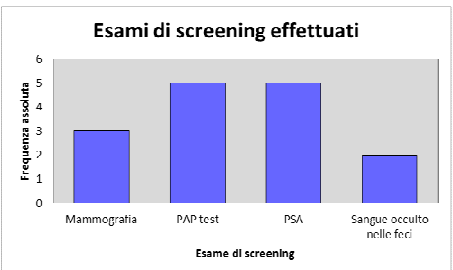
**Pertanto un grafico deve fornire al lettore una informazione sintetica e facile da interpretarsi.**

Dott.ssa Marta Di Nicola



### ORTOGRAMMI-DIAGRAMMA A BARRE


Usati per variabili qualitative, l'altezza delle barre rappresenta la frequenza assoluta o percentuale



**Esami di screening effettuati**

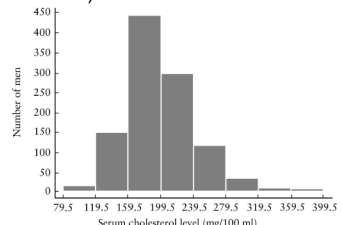
Esame di screening	Frequenza assoluta
Mammografia	3
PAP test	5
PSA	5
Sangue occulto nelle feci	2

Dott.ssa Marta Di Nicola



### ISTOGRAMMI

Indicati per rappresentare distribuzioni in classi (variabili quantitative continue).



**Number of men**

**Serum cholesterol level (mg/100 ml)**

Costituiti da una serie di barre rettangolari contigue ognuna in rappresentanza di una classe e con area proporzionata alla rispettiva frequenza.

Dott.ssa Marta Di Nicola




### GRAFICI PER SPEZZATE-DIAGRAMMI LINEARI

Si ottengono dai grafici per punti congiungendo i vari punti. Indicati per evidenziare una continuità tra valori come ad es. nella rappresentazione delle serie temporali.



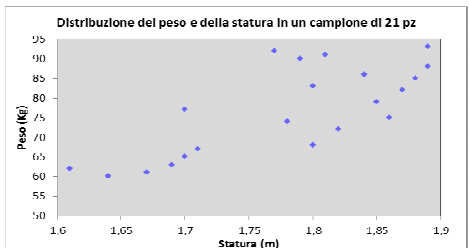
**Glicemia di un ricoverato in 3 giornate consecutive**

Dott.ssa Marta Di Nicola



### DIAGRAMMI DI DISPERSIONE

Costituito dai punti corrispondenti alle diverse coppie di valori rilevati. Indicati per evidenziare le associazioni tra variabili quantitative.



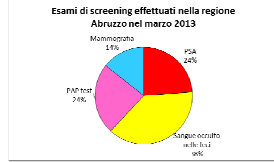
**Distribuzione del peso e della statura in un campione di 21 pz**

Dott.ssa Marta Di Nicola

## DIAGRAMMI A SETTORI CIRCOLARI

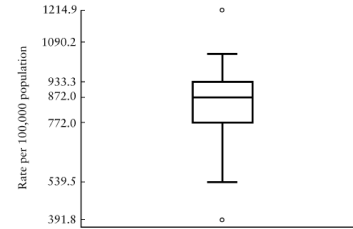


Indicati per variabili qualitative allo scopo di evidenziare le frequenze % delle singole modalità. L'area di un cerchio viene suddivisa in settori proporzionali alle frequenze %



Dott.ssa Marta Di Nicola

## BOX PLOT



**FIGURE 2.8**  
Box plot: Crude death rates for the United States, 1992

Dott.ssa Marta Di Nicola