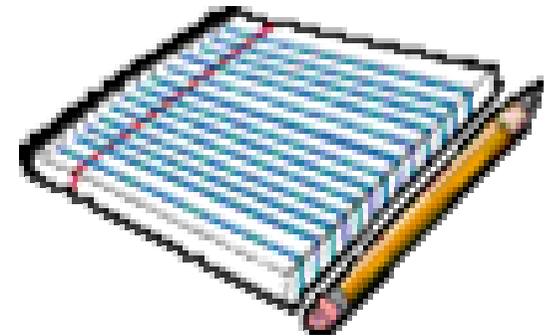


# Lezione 1

## Concetti Fondamentali



# Sonetto di Trilussa

---



Sai ched'è la statistica? E' 'na cosa  
che serve pe' 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, lì la media 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  
perchè c' un antro che se ne magna due.

# Definizione di Statistica

---



Esistono diverse definizioni di statistica  
(oltre 100)

La *Statistica* è lo studio di come si può utilizzare  
l'informazione come supporto alle decisioni, in  
situazioni che implicano incertezza  
(Bartlett 1975).

# La Statistica nella Società dell'Informazione



**Tutti dicono che viviamo nella *società dell'informazione*.**

Ma molti (tutti?) si lamentano che le informazioni sono troppe. E' facile raccoglierle, memorizzarle, distribuirle. E' difficile verificarle ed interpretarle.

La statistica è, in molte situazioni, la *tecnologia* necessaria per risolvere queste difficoltà.

L'informazione, debitamente trattata, offre la possibilità di prendere decisioni in condizioni di incertezza.

# Esempio



L'Organizzazione Mondiale del Turismo (WTO) raccoglie dati relativi al numero dei turisti, entrate ed uscite relative al settore turistico di quasi tutti gli stati del mondo.

Con i suddetti dati si possono analizzare la tendenza dei movimenti dei turisti nelle differenti località del mondo e fare previsioni per il futuro;

Evidentemente non esiste nessuna certezza, anche se esistono nei dati analizzati determinate tendenze, che le stesse si ripetano nel futuro. Possono infatti cambiare all'improvviso, dovute a fenomeni imprevisti quali guerre, disgrazie naturali, cambio nelle preferenze o gusti dei consumatori etc.....

# Classificazioni della Statistica



La Statistica si divide in due parti:

Statistica Descrittiva

Statistica Inferenziale

Il nesso tra la Statistica Descrittiva e la Statistica Inferenziale è rappresentato dalla Teoria del Calcolo delle Probabilità

# Statistica Descrittiva



Tecnica o Metodo utilizzato per raccogliere, riassumere, presentare e analizzare i risultati delle osservazioni dei fenomeni reali.

Nel campo del turismo abbiamo principalmente:

**Analisi Univariata.** Conoscere le caratteristiche individuali di un collettivo, per esempio, la proporzioni di donne, l'età media dei visitatori di un determinato museo;

**Analisi Bivariate.** Relazionare due variabili, per esempio studiare l'influenza che una svalutazione monetaria può avere rispetto al numero di turisti di un determinato paese.

**Analisi Multivariate.** Relazionare più variabili, per esempio, raggruppare i visitatori di un parco naturale in differenti tipologie, prendendo in considerazioni più variabili, affinché alle diverse categorie si possano offrire diversi prodotti.

# Statistica Inferenziale



Scienza che utilizza come strumento il calcolo delle probabilità, con l'obiettivo di studiare le leggi di comportamento di quei fenomeni caratterizzati dall'incertezza.

Nel campo del turismo, la statistica inferenziale è largamente utilizzata per esempio:

Realizzare indagini di mercato, per esempio, per conoscere se un determinato prodotto turistico sarà acquistato.

Realizzare controlli di qualità, per esempio conoscere il grado di soddisfazione di un determinato servizio, dei clienti di un hotel.

Realizzare altri studi, quali per esempio, conoscere le caratteristiche dei visitatori di una certa zona geografica.



# Popolazione e Campione

---

**Popolazione.** Insieme finito o infinito di elementi rispetto ad uno o più caratteri.

**Campione.** Insieme finito di elementi di numerosità ridotta rispetto alla popolazione, estratto dalla popolazione con procedimento casuale.



## In Sintesi.....

---

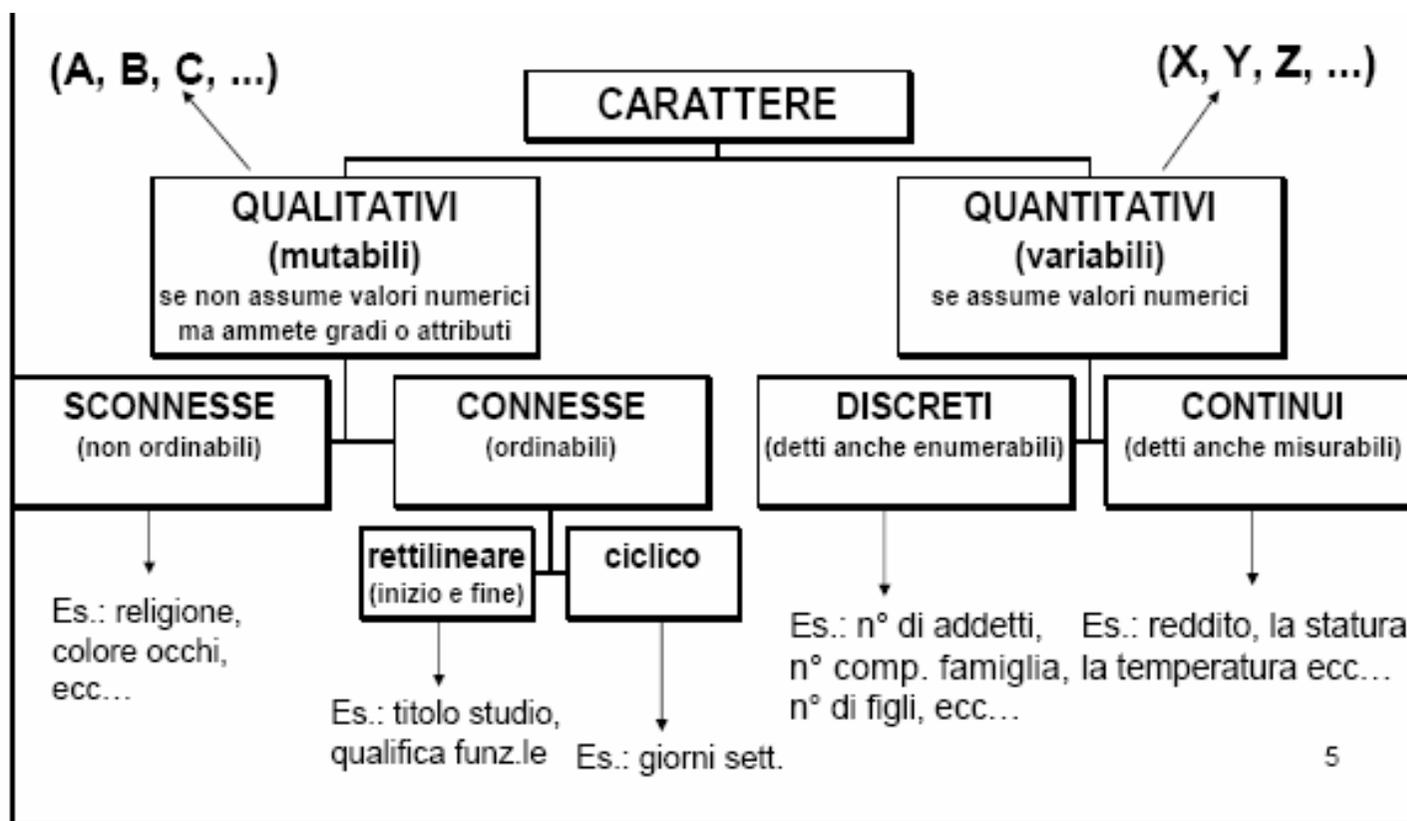
**Descrittiva:** (.quasi. sinonimi: esplorazione statistica dei dati, statistica senza modello probabilistico) Disponiamo di dati riferiti a tutta la popolazione di riferimento.

**Inferenza:** I dati disponibili sono stati rilevati solamente su una parte delle unita statistiche (il *campione da cui indagini campionarie*). Vogliamo utilizzare le informazioni del campione per fare delle affermazioni sulle caratteristiche di tutta la popolazione.



# I Caratteri

I **Caratteri** (o variabili) sono gli aspetti delle unità statistiche oggetto di studio. I diversi modi di presentarsi del carattere si chiamano **modalità**





# Il concetto di Misura

---

Per misura si intende l'operazione di associazione di numeri a concetti e variabili utilizzando una determinata scala.

Esistono 5 differenti scale di misura:

1. **Nominale**
2. **Ordinale**
3. **Intervalli**
4. **Rapporti**
5. **Dicotomica**



# Scala Nominale

Viene utilizzata per le variabili le cui modalità sono nomi (mutabili)

E' possibile tra le modalità stabilire solo uguaglianze o disuguaglianze

Le modalità vengono anche chiamate categorie

## **Esempio:**

- Stato civile
- Genere
- Settore di attività e economica di un'impresa

Canale televisivo

# Scala Ordinale

---



È una variabile qualitativa le cui modalità possono essere messe in sequenza

E' possibile stabilire tra le modalità relazioni d'ordine ma non misurare la grandezza.

- Giudizio (scarso, mediocre, sufficiente, buono, ottimo)
- Titolo di studio (ordinato per anni di studio)



# Scala a intervalli

È una variabile quantitativa le cui modalità sono valori che partono da un'origine arbitraria

I valori sulla scala riflettono l'ordine di grandezza delle osservazioni.

L'origine della scala è fissata in modo arbitrario  
esempio: temperature

E' possibile l'operazione di somma



# Scala di rapporti

È una variabile le cui modalità sono valori valutati da un'origine che rappresenta la nullità del fenomeno osservato.

Lo “zero” della scale è univocamente definito e sono consentite operazioni di prodotto e rapporto.

Esempio di scala a rapporti

- Reddito
- Numero di figli per famiglia
- Numero di addetti per imprese
- Età



# Scala Dicotomica

È una variabile con due sole modalità

E' l'espressione del possesso o meno di un determinato attributo oppure l'appartenenza a una categoria di unità.

- Esempio di variabile dicotomica
- Genere
- Avere superato un esame
- Appartenere ad un corso
- Reddito (basso/alto)

# Frequenza

---



Una volta in possesso di dati, la prima forma di sintesi dell'informazione è rappresentata dalle distribuzioni di frequenza.

Per **Intensità** di un carattere si intende il numero che esprime il suo ammontare, la sua misura o la sua grandezza.

Per **Frequenza**, invece, il numero delle unità statistiche che presentano secondo un carattere le stesse modalità.



# Frequenza

## Frequenza Assoluta

Numero delle volte in cui una determinata modalità è stata osservata

## Frequenza Relativa

Rapporto tra la frequenza assoluta ed il totale delle frequenze

$$f_i = n_i/n \quad 0 \leq f_i \leq 1$$

$$\sum_{i=1}^k f_i = \sum_{i=1}^k \frac{n_i}{n} = \frac{1}{n} \sum_{i=1}^k n_i = n \cdot \frac{1}{n} = 1$$



# Esempio

## Variabile Quantitativa

Un ristorante dispone di tre menu, i cui prezzi sono 8, 10, 12 € rispettivamente.

In un determinato giorno vengono serviti 15 menu del primo tipo, 10 del secondo e 5 del terzo.

Variabile di studio prezzo dei menu.

Prezzo	Frequenze assolute
8	15
10	10
12	5
Totale	30

# Esempio



Variabile qualitativa

Nell'Hotel Comodoro, vengono organizzate visite guidate presso tre località dell'isola di Cuba:

<b>Località</b>	<b>Frequenza Assoluta</b>
Varadero	20
Pinar del Rio	16
Camaguey	18
Totale	54

# Frequenze percentuali, assolute cumulate e relative cumulate



Frequenze percentuali

$$f_i = n_i / n \times 100$$

**Frequenze Assolute cumulate**

$$C_1 = n_1$$

$$C_2 = n_1 + n_2$$

.

.

$$C_k = n_1 + n_2 + \dots + n_k = n$$

**Frequenze relative cumulate**

$$F_1 = n_1 / n$$

$$F_2 = n_1 / n + n_2 / n$$

$$F_k = n_1 / n + n_2 / n + \dots + n_k / n = 1$$



# Distribuzioni Statistiche

Classificare le unità statistiche secondo un carattere (distribuzioni semplici) o più caratteri (distribuzioni multiple).

Distribuzione per unità:

Elencazione degli attributi con la quale ciascun carattere si presenta

Se la numerosità del collettivo è elevato si può utilizzare la distribuzione di frequenza.

Se la variabile è quantitativa o presenta numerose modalità, la distribuzione di frequenza necessita di un pretrattamento



# Distribuzioni in classi

Esempio Spesa per la realizzazione di un bar

<b>Classe</b>	<b>Frequenza</b>
0- 200.000	40
200.000- 500.000	30
500000- 1.000.000	10
1.000.000-2.000.000	5
2.000.000-5.000.000	5
totale	90

# Elementi principali della distribuzioni in classi



Per ampiezza della classe si intende la differenza tra l'estremo superiore e l'estremo inferiore della classe.

Il valore centrale della classe è la semisomma dei due estremi.

Per densità di classe si intende il rapporto tra la frequenza e l'ampiezza della classe



# Il raggruppamento in classi

## Formula di Sturges

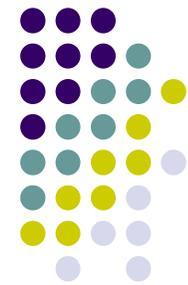
(Numero minimo di classi)

$$S = 1 + \frac{10}{3} \log(n)$$

## Ampiezza della Classe

$$w = \frac{x_{\max} - x_{\min}}{S}$$

# Esercizio



In un determinato libro di 49 pagine sono stati contati il numero di errori di battitura per pagina.

I risultati sono i seguenti

0	2	2	6	5
3	5	7	11	9
1	1	0	20	10
3	2	2	15	7
6	5	2	15	4
4	3	2	16	12
5	7	6	11	16
0	8	0	14	17
2	7	19	16	19
11	12	4	5	

Si costruisca la distribuzione per classi