

# STATISTICA DESCRITTIVA

## RAPPRESENTAZIONE DEI DATI

Prof. Rosario Lo Franco – Lezione 1

**Riferimenti:** [1] Sheldon M. Ross, *Introduzione alla statistica*, Apogeo Editore;  
[2] Maria Garetto, *Statistica*, Università di Torino

# Prof. Rosario Lo Franco



Modulo di **Statistica**  
3 CFU

## CONTATTI

- Dipartimento di Fisica e Chimica, via Archirafi 36, 90123 Palermo
- **Tel.** 091/23891709
- **Email:** [rosario.lofranco@unipa.it](mailto:rosario.lofranco@unipa.it)

**URL personale:** <http://www.fisica.unipa.it/~lofranco/index.html>

**URL docente:** <http://portale.unipa.it/persone/docenti/l/rosario.lofranco/>

# Introduzione alla statistica: Problemi

1. La prossima settimana si terranno le elezioni presidenziali americane, ed intervistando un campione di elettori vorremmo stabilire se prevarrà il candidato repubblicano o quello democratico. Quale dei seguenti metodi di selezione produrrà più facilmente un campione rappresentativo?
  - (a) Intervistare tutti gli spettatori di maggiore età ad una partita di basket tra college.
  - (b) Intervistare tutte le persone di maggiore età che escono da un lussuoso ristorante del centro.
  - (c) Ottenere una copia dell'elenco degli elettori, sceglierne 100 a caso ed intervistarli.
  - (d) Usare i risultati di un sondaggio televisivo basato sulle telefonate dei telespettatori.
  - (e) Scegliere dei nomi dall'elenco telefonico e intervistare queste persone.

# Introduzione alla statistica: Problemi

2. L'approccio suggerito nel punto (e) del Problema 1 portò ad una predizione disastrosa in occasione delle elezioni presidenziali americane del 1936, quando Franklin Roosevelt ottenne una vittoria schiacciante su Alfred Landon. La rivista *Literary Digest* predisse infatti la vittoria di Landon, basandosi sulle preferenze di un campione di elettori che era stato scelto sugli elenchi del telefono e tra i proprietari di automobili.
- (a) Per quale motivo la predizione fu tanto scorretta?
  - (b) Dal 1936 ad oggi è intervenuto qualche cambiamento che autorizzi a pensare che un approccio di questo tipo sia ora più affidabile di allora?

# Introduzione alla statistica: Problemi

**4. Per determinare la percentuale di fumatori tra i residenti di un comune, si decide di fare un sondaggio, intervistando i frequentatori di uno dei luoghi seguenti:**

- (a) una piscina;**
- (b) un bowling;**
- (c) un centro commerciale;**
- (d) una biblioteca.**

**Quale di essi ha più chances di fornire una buona approssimazione della percentuale in esame? Perché?**

# Introduzione alla statistica: Problemi

**Tabella 1.2** Tabella delle mortalità di John Graunt

(Le classi di età arrivano fino all'estremo destro escluso. Ad esempio 0-6 significa tutte le età dagli 0 ai 5 anni.)

Tempo di vita	Numero di decessi su 100 nascite
0-6	36
6-16	24
16-26	15
26-36	9
36-46	6
46-56	4
56-66	3
66-76	2
76 o più	1

Basandoti sulla tabella di mortalità di Graunt, rispondi alle seguenti domande.

- (a) Quale frazione della popolazione raggiungeva l'età di 6 anni?
- (b) Quale frazione raggiungeva i 46 anni?
- (c) Quale frazione moriva tra i 6 ed i 45 anni?

# Statistica Descrittiva

## *Esempio 1*

Rilevando con uno strumento di misurazione il numero di particelle cosmiche in 40 periodi consecutivi di un minuto si ottengono i seguenti dati

0	2	1	4	3	1	2	3	8	2	5	2	1	3	3	1	3	2	2	5
4	4	4	2	3	5	5	1	1	2	4	4	2	3	3	3	3	3	3	2

Tabella 1

- **Variabile discreta**
- **Costruiamo tabella di frequenza, diagramma a barre e poligonale**

# Statistica Descrittiva

## *Esempio 2*

I seguenti dati sono il risultato di 80 determinazioni, in una data unità di misura, dell'emissione giornaliera di un gas inquinante da un impianto industriale

15.8	26.4	17.3	11.2	23.9	24.8	18.7	13.9	9.0	13.2
22.7	9.8	6.2	14.7	17.5	26.1	12.8	28.6	17.6	23.7
26.8	22.7	18.0	20.5	11.0	20.9	15.5	19.4	16.7	10.7
19.1	15.2	22.9	26.6	20.4	21.4	19.2	21.6	16.9	19.0
18.5	23.0	24.6	20.1	16.2	18.0	7.7	13.5	23.5	14.5
14.4	29.6	19.4	17.0	20.8	24.3	22.5	24.6	18.4	18.1
8.3	21.9	12.3	22.3	13.3	11.8	19.3	20.0	25.7	31.8
25.9	10.5	15.9	27.5	18.1	17.9	9.4	24.1	20.1	28.5

Tabella 2

- **Variabile continua**
- **Costruiamo tabella di frequenza e istogramma**



# Statistica Descrittiva

## *Esempio 3*

In uno stabilimento vengono registrati i casi di malfunzionamento di una macchina utensile controllata dal computer, e le loro cause. I dati relativi a un certo mese sono i seguenti

fluttuazioni di tensione	6
instabilità del sistema di controllo	22
errore dell'operatore	13
strumento usurato e non sostituito	2
altre cause	5
<i>Totale</i>	48

Tabella 3

- **Variabile nominale**
- **Costruiamo grafico a torta.**

# Statistica Descrittiva

In ciascuno degli esempi si osserva una variabile, che è rispettivamente

- 1 – il numero di particelle rilevate in un intervallo di un minuto;
- 2 – la quantità di gas inquinante emesso in un giorno;
- 3 – la causa di un guasto verificato.

Della variabile in questione abbiamo un insieme di  $n$  osservazioni registrate (negli esempi  $n$  vale, rispettivamente, 40, 80, 48), che costituiscono i dati da analizzare.

variabili	$\left\{ \begin{array}{l} \text{numeriche (quantitative)} \\ \text{non numeriche (qualitative)} \end{array} \right.$	$\left\{ \begin{array}{l} \text{discrete} \\ \text{continue} \end{array} \right.$
-----------	--	---

# Statistica Descrittiva

**Tabella di frequenza (Esempio 1)**

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
0	1	0.025	2.5%
1	6	0.15	15%
2	10	0.25	25%
3	12	0.3	30%
4	6	0.15	15%
5	4	0.1	10%
6	0	0	0%
7	0	0	0%
8	1	0.025	2.5%
<i>Totale</i>	40	1	100%

Tabella 4

# Statistica Descrittiva

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
0	1	0.025	2.5%
1	6	0.15	15%
2	10	0.25	25%
3	12	0.3	30%
4	6	0.15	15%
5	4	0.1	10%
6	0	0	0%
7	0	0	0%
8	1	0.025	2.5%
<i>Totale</i>	40	1	100%

## **Osservazione**

Si osservino le seguenti proprietà dei numeri riportati nella tabella di distribuzione di frequenza (tabella 4): la frequenza assoluta è un numero intero compreso tra 0 e il numero totale di osservazioni; la frequenza relativa è un numero reale compreso tra 0 e 1; la frequenza percentuale è un numero reale compreso tra 0 e 100.

La somma delle frequenze assolute è sempre uguale al numero totale di osservazioni; la somma delle frequenze relative è sempre uguale a 1; la somma delle frequenze percentuali è uguale a 100; i valori ottenuti come quozienti devono essere spesso arrotondati e questo fatto comporta che la somma di tutte le percentuali può non essere esattamente uguale a 100.

# Statistica Descrittiva

## Tabella di frequenza (Esempio 2)

Nell'esempio 2 la variabile osservata è continua. I valori dei dati sono compresi tra 6.2 e 31.8; il **campo di variazione  $R$  o range** dei dati, cioè la differenza tra il più grande e il più piccolo, vale

$$R = 31.8 - 6.2 = 25.6$$

Scegliamo come classi i 7 intervalli

$$5.0 \leq x \leq 8.9$$

$$9.0 \leq x \leq 12.9$$

$$13.0 \leq x \leq 16.9$$

$$17.0 \leq x \leq 20.9$$

$$21.0 \leq x \leq 24.9$$

$$25.0 \leq x \leq 28.9$$

$$29.0 \leq x \leq 32.9$$

**N.B. Il modo di scegliere il numero di classi  $k$  non è unico.** Suggerimento:

$$k \cong 1 + 3.322 \cdot \log_{10} n$$

**Ampiezza delle classi**  
(uguale per tutte):

$$a \cong \frac{R}{k}$$

**Nel nostro esempio:**

$$k = 1 + 3.322 \cdot \log_{10} 80 \cong 7 \quad a = \frac{25.6}{7} \cong 3.7$$

# Statistica Descrittiva

**Tabella di frequenza (Esempio 2)**

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
$5.0 \leq x \leq 8.9$	3	0.0375	3.75%
$9.0 \leq x \leq 12.9$	10	0.1250	12.5%
$13.0 \leq x \leq 16.9$	14	0.1750	17.5%
$17.0 \leq x \leq 20.9$	25	0.3125	31.25%
$21.0 \leq x \leq 24.9$	17	0.2125	21.25%
$25.0 \leq x \leq 28.9$	9	0.1125	11.25%
$29.0 \leq x \leq 32.9$	2	0.0250	2.5%
<i>Totale</i>	80	1	100%

Tabella 5

Le classi hanno uno “stacco” per evitare ambiguità.

E' più consigliabile scegliere **classi chiuse a sinistra (aperte a destra)**, ad esempio

$$5 \leq x < 9$$

$$9 \leq x < 13$$

.....

$$29 \leq x < 33$$

# Statistica Descrittiva

## Tabella di frequenza (Esempio 3)

Nell'esempio 3 la variabile “tipo di guasto verificato” è non numerica; i dati sono già raggruppati in classi e si ottiene la seguente tabella di distribuzione di frequenza

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
<i>fluttuazioni tensione</i>	6	0.125	12.5%
<i>instabilità</i>	22	0.458	45.8%
<i>errore operatore</i>	13	0.271	27.1%
<i>strumento</i>	2	0.042	4.2%
<i>altro</i>	5	0.104	10.4%
<i>Totale</i>	48	1	100%

Tabella 7

# Statistica Descrittiva

Ci sono altri modi di raggruppare i dati: ad esempio dati “minori di”, “maggiore di”; si ottengono in questo modo le **distribuzioni cumulative**.

La frequenza totale di tutti i valori minori del limite superiore di una data classe è detta **frequenza cumulativa**. Una tabella che presenti frequenze cumulative è detta **tabella di distribuzione cumulativa di frequenza**.

## Esempio 1

<i>Classe</i>	<i>Freq. cumul. assoluta</i>
$x \leq 0$	1
$x \leq 1$	7
$x \leq 2$	17
$x \leq 3$	29
$x \leq 4$	35
$x \leq 5$	39
$x \leq 6$	39
$x \leq 7$	39
$x \leq 8$	40

## Esempio 2

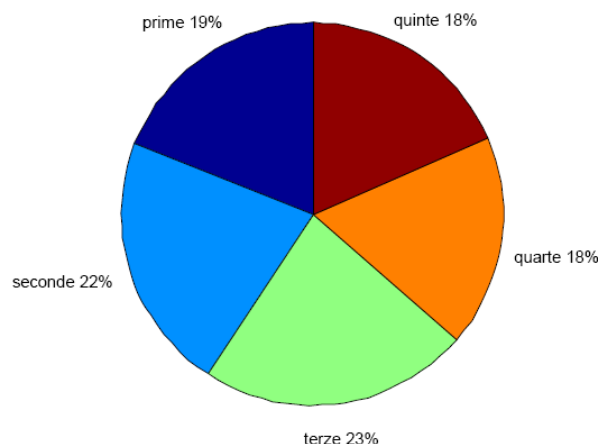
<i>Classe</i>	<i>Freq. cumul. assoluta</i>
$x \leq 4.9$	0
$x \leq 8.9$	3
$x \leq 12.9$	13
$x \leq 16.9$	27
$x \leq 20.9$	52
$x \leq 24.9$	69
$x \leq 28.9$	78
$x \leq 32.9$	80



# Statistica Descrittiva

## Rappresentazione grafica dei dati

### Grafico a torta



Per variabili nominali (per l'esempio 3 quale sarebbe?)

$$f : 100 = g : 360^\circ$$

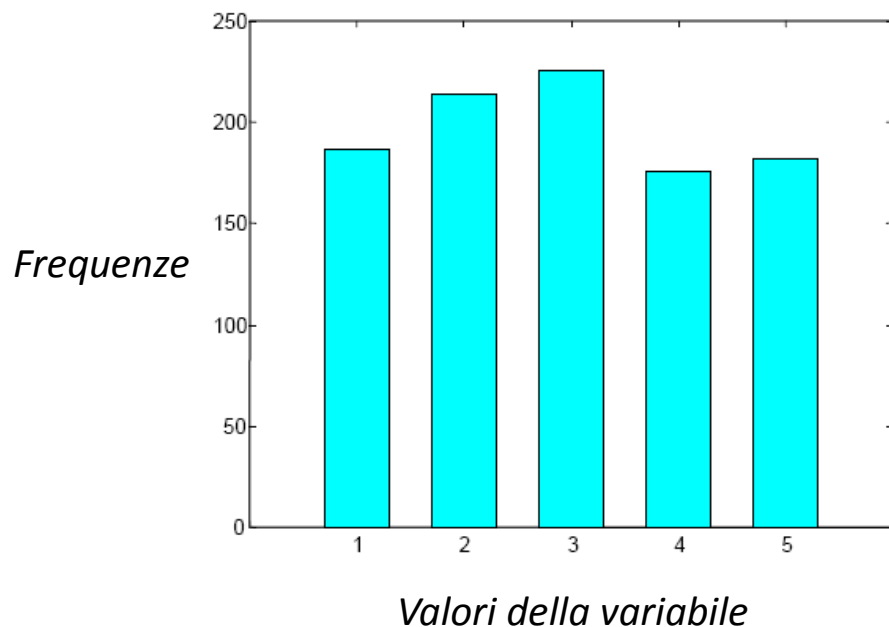
f = frequenza percentuale della variabile

g = ampiezza in gradi

# Statistica Descrittiva

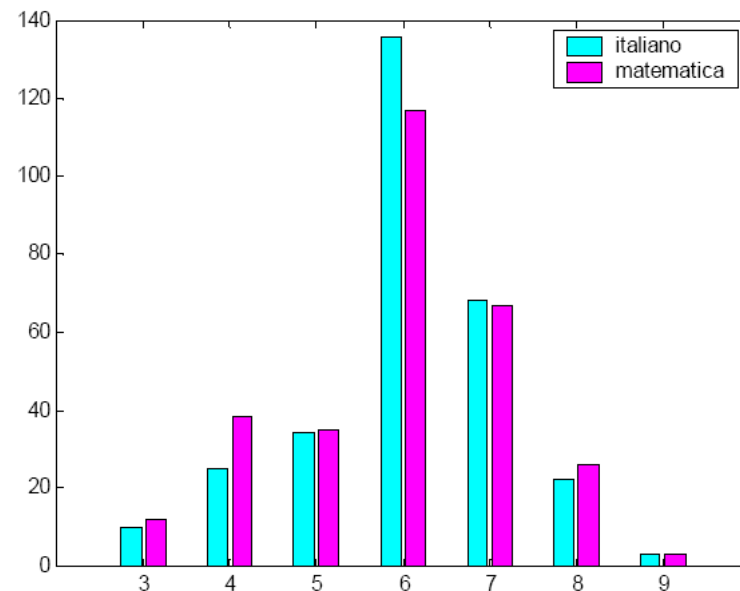
## Rappresentazione grafica dei dati

### Grafico a barre



Per variabili numeriche discrete  
(per l'esempio 1 quale sarebbe?)

Si possono rappresentare  
contemporaneamente due  
diverse distribuzioni di dati per  
un confronto

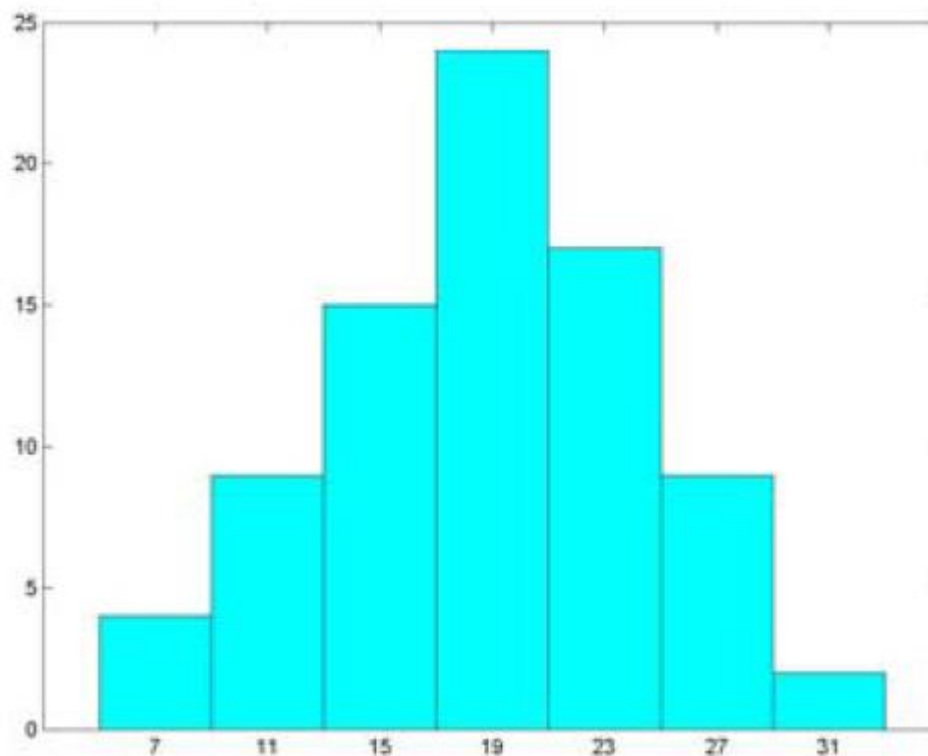


# Statistica Descrittiva

## Rappresentazione grafica dei dati

### Istogramma

*Frequenze o  
frequenze  
relative  
(lungo asse y)*



*Classi della variabile  
(lungo l'asse x)*

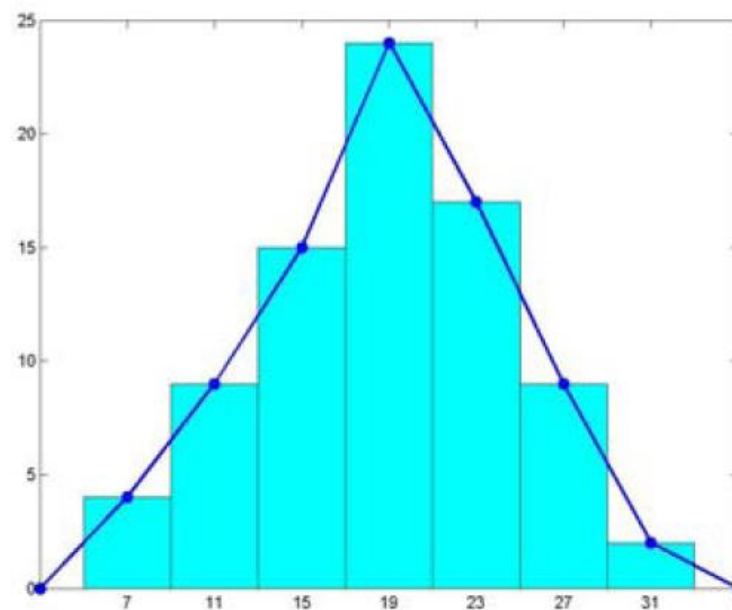
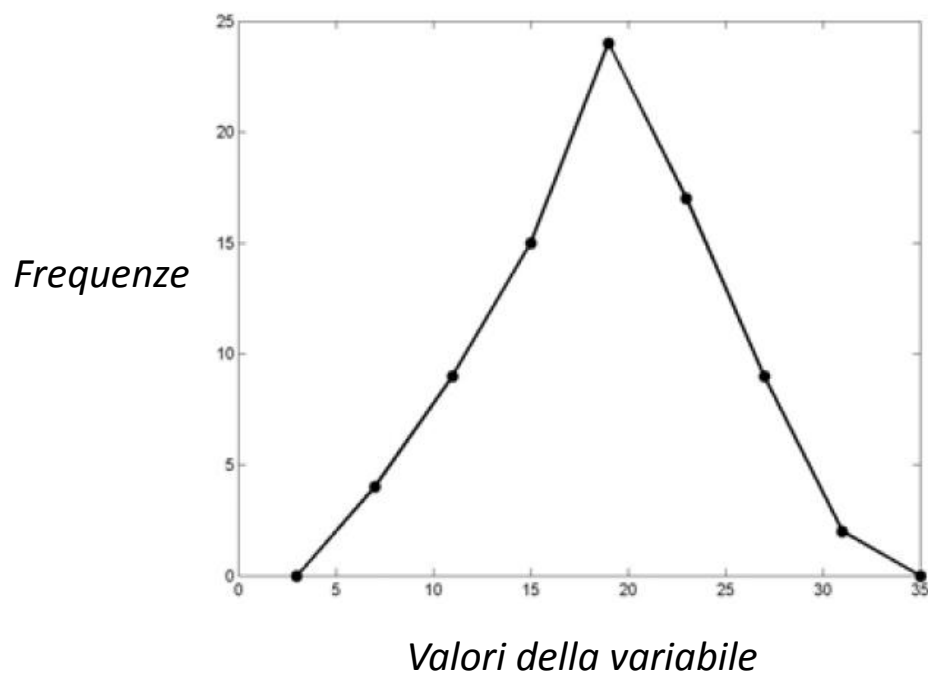
Tipicamente per variabili numeriche continue (per l'esempio 2 quale sarebbe?)

# Statistica Descrittiva

## Rappresentazione grafica dei dati

### Poligono di frequenza

Una distribuzione di frequenza può essere rappresentata graficamente anche con un altro tipo di grafico: il **poligono di frequenza**. Tale poligono si ottiene unendo fra loro i punti aventi come ascissa il valore centrale di ogni classe e come ordinata il corrispondente valore della frequenza.



Quale sarebbe per l'esempio 1?