

# Esercitazione 1 del corso di Statistica (parte 1)

*Dott.ssa Paola Costantini*

24 Gennaio 2012

## Le componenti fondamentali dell'analisi statistica

- Unità statistica

Oggetto dell'osservazione di ogni fenomeno individuale che costituisce il fenomeno collettivo

- Carattere

Caratteristica di ciascuna unità statistica

- Popolazione o collettivo statistico

Insieme di unità statistiche omogenee rispetto ad una o più caratteristiche

- Modalità

Il modo in cui si presenta un carattere. Ciascun carattere è presente in ogni unità con una determinata modalità

- Tipi di caratteri
  - Qualitativo: sconnesso (scala nominale), ordinato (scala ordinale)
  - Quantitativo: discreto (scala ad intervalli), continuo (scala ad intervalli). In questo tipo di caratteri le modalità sono espresse numericamente.

### Esempio

Sconnessi	Ordinati	Discreti	Continui
Sesso	Grado di soddisfazione	n. figli	Peso
Settore Attività	Posizione in graduatoria	n. pezzi prodotti	Altezza
Luogo di nascita	Titolo di studio	n. risposte esatte ad un test	Reddito
Stato civile	Livello di qualità	n. apparecchi telefonici	Distanza
Religione	Gerarchia militare		
Colore occhi			

## Esercizio 1

La seguente tabella raccoglie i dati dei primi 5 classificati nella gara individuale di lancio del disco.

NOME	GRADUATORIA	ALTEZZA (CM)	RESIDENZA	PUNTEGGIO
ALESSIA	1	170	ROMA	180
SARA	2	172	MILANO	167
MARCO	3	180	TORINO	150
ANDREA	4	183	NAPOLI	134
STEFANO	5	175	ROMA	110

a) Indicare qual'è l'unità statistica e quante unità statistiche sono presenti.

**Soluzione.** Unità statistica: singolo atleta; numero unità statistiche: 5.

b) Indicare quali sono i caratteri e per ogni carattere indicare se è quantitativo, discreto o continuo, oppure qualitativo, sconnesso o ordinato.

**Soluzione.** I caratteri sono: Graduatoria (qualitativo ordinato), Altezza (quantitativo continuo), Residenza (qualitativo sconnesso), Punteggio (quantitativo discreto poiché assume numeri interi).

## Esercizio 2

Data la seguente tabella:

Studente	Voto	Sesso	Occhi	Soddisfazione	Altezza	Peso
1	18	M	Marrone	Niente	173	65
2	24	F	azzurro	Abbastanza	176	70
3	30	F	Marrone	Molto	165	56
4	28	M	Marrone	Molto	181	78
5	24	M	Marrone	Poco	178	76
6	18	F	Azzurro	Niente	167	65
7	24	F	Marrone	abbastanza	173	70
8	30	M	verde	Molto	181	80

a) Indicare i tipi di caratteri considerati (qualitativo, ordinabile o sconnesso, oppure quantitativo, continuo o discreto).

**Soluzione.** Rispettivamente: discreto, sconnesso, sconnesso, ordinabile, continuo, continuo.

b) Costruire le distribuzioni di frequenze (assoluta, relativa e percentuale) per ogni carattere considerato (se il carattere considerato è continuo si consideri il raggruppamento in classi).

Voto	$n_i$	$f_i$	%
18	2	0,25	25%
24	3	0,375	37,5%
28	1	0,125	12,5%
30	2	0,25	25%
<b>Totale</b>	<b>20</b>	<b>1</b>	<b>100%</b>

Sesso	$n_i$	$f_i$	%
M	4	0,5	50%
F	4	0,5	50%
<b>Totale</b>	<b>20</b>	<b>1</b>	<b>100%</b>

Occhi	$n_i$	$f_i$	%
Marrone	5	0,625	62,5%
Azzurro	2	0,25	25%
verde	1	0,125	12,5%
<b>Totale</b>	<b>20</b>	<b>1</b>	<b>100%</b>

Soddisfazione	$n_i$	$f_i$	%
niente	2	0,25	25%
poco	1	0,125	12,5%
abbastanza	2	0,25	25%
molto	2	0,375	37,5%
<b>Totale</b>	<b>20</b>	<b>1</b>	<b>100%</b>

Peso	$n_i$	$f_i$	%
56	1	0,125	12,5%
65	2	0,25	25%
70	2	0,25	25%
76	1	0,125	12,5%
78	1	0,125	12,5%
80	1	0,125	12,5%
<b>Totale</b>	<b>8</b>	<b>1</b>	<b>100%</b>

Per il carattere **Peso** suddividiamo la distribuzione in 3 classi equiampie riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.

L'ampiezza costante delle tre classi si ottiene come:

$$\text{range (PESO)} = 80 - 56 = 24$$

L'ampiezza delle classi della prima distribuzione di frequenza è pari a:

$$\text{Ampiezza classi} = \text{range}/3 = 24/3 = 8$$

Le 3 classi sono, dunque, delimitate dai seguenti estremi:

$$C_1 = [56; 64]$$

$$C_2 = ] 64; 72 ]$$

$$C_3 = ] 72; 80 ]$$

$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$	$\hat{C}_i$
$C_1 = [56; 64]$	1	0,125	1	0,125	8	0,0156	60
$C_2 = ] 64; 72 ]$	4	0,5	5	0,625	8	0,0625	68
$C_3 = ] 72; 80 ]$	3	0,375	8	1	8	0,0468	76
<b>Totali</b>	<b>20</b>	<b>1</b>					

Altezza	$n_i$	$f_i$	%
165	1	0,125	12,5%
167	1	0,125	12,5%
173	2	0,25	25%
176	1	0,125	12,5%
178	1	0,125	12,5%
181	2	0,25	25%
<b>Totale</b>	<b>8</b>	<b>1</b>	<b>100%</b>

Per il carattere Altezza suddividiamo la distribuzione in 3 classi equifrequenti riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.

$N$  (numerosità totale) = 8

$8/3 = 2,67$

Avremo pertanto due classi con frequenza pari a 3 e una classe con frequenza pari a 2. La classe con frequenza 2 è la prima.

$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$	$\hat{c}_i$
$C_1 = [165; 167]$	2	0,25	2	0,25	2	0,125	166
$C_2 = ] 167; 176]$	3	0,375	5	0,625	9	0,04167	171,5
$C_3 = ] 176; 181 ]$	3	0,375	8	1	5	0,075	178,5
<b>Totali</b>	<b>8</b>	<b>1</b>					

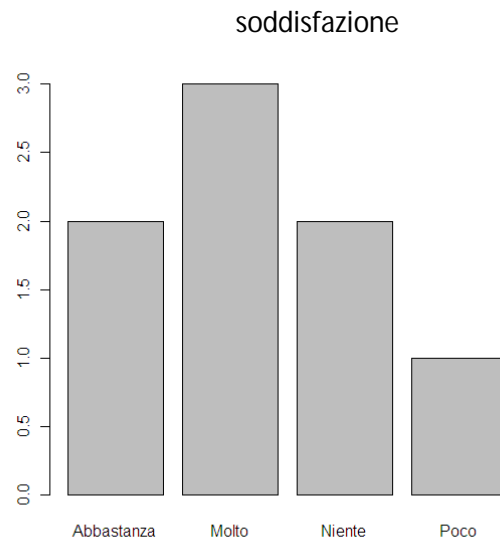
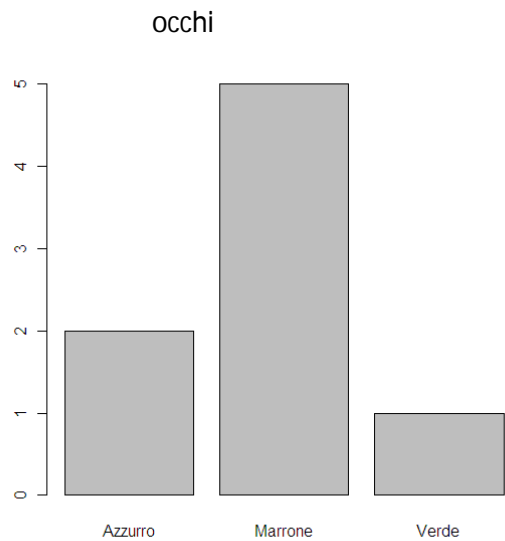
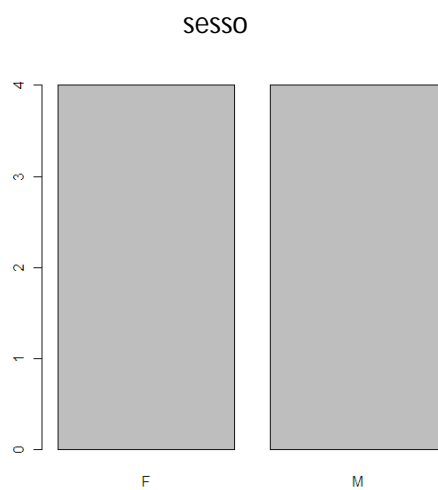
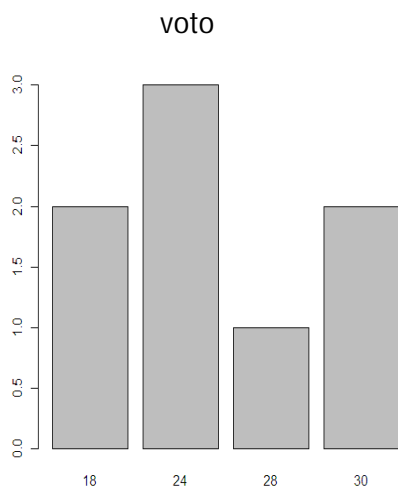
### Esercizio 3

Utilizzando le distribuzioni di frequenze dell'esercizio 2, rappresentare i caratteri Voto, Sesso, Occhi, Soddifazione, Altezza e Peso mediante un'opportuna rappresentazione grafica.

**Soluzione.** Voto: grafico a barre, Sesso: grafico a barre; Occhi: grafico a barre; Soddifazione: grafico a barre; Altezza: istogramma; Peso: istogramma.

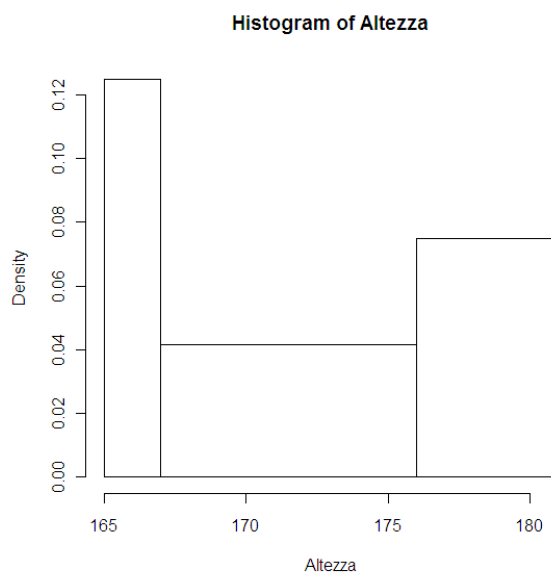
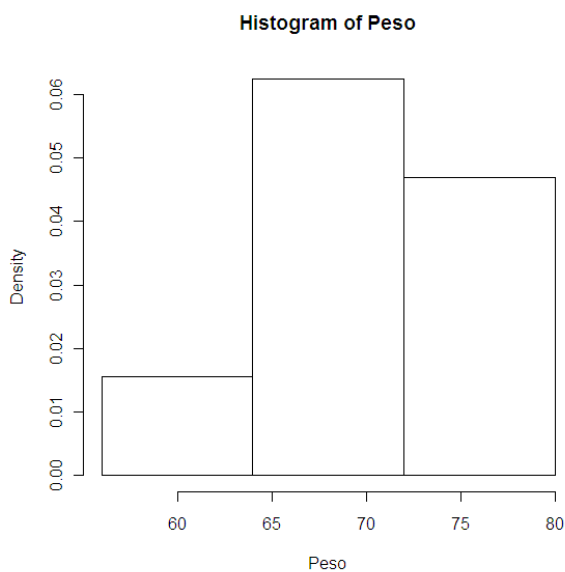
Le rappresentazioni grafiche consentono di cogliere con evidenza visiva la struttura e l'andamento di uno o più fenomeni, il confronto tra più distribuzioni. Attraverso un grafico si riescono ad evidenziare aspetti fondamentali di una distribuzione di frequenza.

I **grafici a barre** (barplot) per tutti i caratteri tranne che caratteri quantitativi continui. Le barre presentano tutte uguale larghezza; in ordinata appaiono le frequenze assolute o quelle relative.



Gli **istogrammi** per caratteri quantitativi continui. Definiamo la densità di frequenza (indicata con  $d_i$ ) come il rapporto tra la frequenza (assoluta o relativa) e l'ampiezza (indicata con  $a_i$ ) di una classe. Ad ogni classe  $\mu_i$  è associato un rettangolo, tale che: la base è pari a  $a_i$  e l'altezza  $\mu_i$  è pari a  $d_i$ .

- L'istogramma è profondamente diverso dal grafico a barre
- Le barre possono avere larghezze diverse tra loro
- E' possibile richiedere l'istogramma delle frequenze assolute solo se le classi hanno la stessa larghezza.



Il seguente dataset descrive alcune caratteristiche di 20 studenti.

<i>N</i>	<i>SESSO</i>	<i>ETA'</i>	<i>PESO</i>	<i>ALTEZZA</i>	<i>DIPLOMA</i>	<i>COMPONENTI</i>	<i>OCCHIALI</i>	<i>FUMO</i>
1	0	20,6	65	180	Ist.Tecnico	6	0	1
2	0	20,2	75	180	Liceo	4	0	0
3	0	20,3	60	173	Ist.Tecnico	4	1	0
4	0	23,9	93	193	Liceo	8	0	1
5	0	21,4	66	164	Ist.Tecnico	5	0	0
6	0	25	84	188	Ist.Tecnico	4	0	0
7	0	20,8	67	175	Altro dipl.	4	0	1
8	0	20,6	89	170	Liceo	3	1	0
9	0	27,1	71	180	Liceo	1	0	1
10	0	23,3	63	170	Liceo	4	0	0
11	1	20,5	51	161	Ist.Tecnico	4	0	1
12	1	19,1	58	167	Ist.Tecnico	5	1	1
13	1	22,1	67	165	Altro dipl.	5	1	1
14	1	21,8	51	153	Ist.Tecnico	4	0	0
15	1	19,2	60	170	Ist.Tecnico	5	1	1
16	1	20,8	55	165	Liceo	4	1	1
17	1	21	55	155	Liceo	5	1	0
18	1	20,9	58	170	Liceo	5	1	1
19	1	22,7	76	170	Liceo	6	1	0
20	1	21	55	165	Liceo	7	0	0

**Esercizio 4** A partire dalla distribuzione di frequenza del carattere Altezza, calcolare la media aritmetica, la media troncata al 20% e la moda.

$X_i$	$n_i$
153	1
155	1
161	1
164	1
165	3
167	1
170	5
173	1
175	1
180	3
188	1
193	1
tot	20

### Media aritmetica

$$\mu = \frac{\sum_{i=1}^n x_i \times n_i}{n} = \frac{153 \times 1 + 155 \times 1 + 161 \times 1 + \dots + 193 \times 1}{20} = 170,7$$

### Moda

$$Mo_1 = 170$$

### Esercizio n. 5

Calcolare media e moda della variabile Altezza ripartita in 3 classi equifrequenti.

$$N/3 = 6,67$$

Per una distribuzione in classi di frequenza, la media si calcola:

$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$\hat{C}_i$	$a_i$	$d_i$
$C_1 = [153; 165]$	7	0,35	7	0,35	159	12	0,029
$C_2 = ] 165; 173 ]$	7	0,35	14	0,70	169	8	0,043
$C_3 = ] 173; 193 ]$	6	0,30	20	1	183	20	0,015
<b>Totali</b>	<b>20</b>	<b>1,00</b>					

$$\mu_x = \frac{\sum_{i=1}^k \hat{x}_i \cdot n_i}{n}$$

## Media

dove

$$\hat{x}_i = \frac{x_{i-1} + x_i}{2}$$

$$\mu_x = \frac{\sum_{i=1}^k (159 \times 7 + 169 \times 7 + 183 \times 6)}{20} = 169,7$$

**Classe Modale** =  $C_2 = ] 165; 173 ]$  = la classe modale è quella cui corrisponde la densità di frequenza più elevata.