

# Esercitazione 1 del corso di Statistica (parte 1)

*Dott.ssa Paola Costantini*

21 Gennaio 2011

## Dataset Studenti

<i>N</i>	<i>SESSO</i>	<i>ETA'</i>	<i>PESO</i>	<i>ALTEZZA</i>	<i>DIPLOMAI</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	187	Liceo	8	0	1
5	0	21,4	66	164	Ist.Tecnico	5	0	0
6	0	25	84	186	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	156	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	158	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

<i>SESSO</i>	sex degli studenti (= 1 Femmine, = 0 Maschi)
<i>ETA'</i>	età degli studenti (in anni);
<i>PESO</i>	peso degli studenti (in kg);
<i>ALTEZZA</i>	altezza degli studenti (in cm);
<i>DIPLOMA</i>	Liceo; Istituti Tecnici; Altro diploma);
<i>COMPONENTI</i>	numero dei componenti della famiglia (compreso lo studente);
<i>OCCHIALI</i>	studenti con occhiali da vista o lenti a contatto (= 1 Sì, = 0 No);
<i>FUMO</i>	studenti fumatori (= 1 Sì, = 0 No);

### Esercizio n.1

Per ognuna delle variabili che compongono il **dataset Studenti**, si definisca il tipo di variabile (quantitativo/qualitativo) elencando l'insieme delle possibili modalità/intensità.

### SOLUZIONE

Carattere	Tipo	Modalità
SESSO	Qualitativa Nominale (dicotomica)	{0; 1}
ETA'	Quantitativa Continua	{19,1; 27,1}
PESO	Quantitativa Continua	{51; 93}
ALTEZZA	Quantitativa Continua	{156; 187}
TIPO DI DIPLOMA	Qualitativa Nominale	{Liceo; Istituto tecnico; Altro diploma}
NUMERO COMPONENTI	Quantitativa Discreta	{1;8}
OCCHIALI	Qualitativa Nominale (dicotomica)	{0; 1}
FUMO	Qualitativa Nominale (dicotomica)	{0; 1}

### Esercizio n. 2

Costruire le distribuzioni di frequenza per le variabili *Tipo di diploma* e *Numero di componenti*, riportando le frequenze: assolute, assolute cumulate, relative, relative cumulate e i valori percentuali.

Distribuzioni di frequenza dei caratteri qualitativi *Tipo di diploma* e *Numero di componenti*.

Tipo di diploma	$n_i$	$f_i$	%
Liceo	10	0,5	50%
Istituti tecnici	8	0,4	40%
Altro diploma	2	0,1	10%
<b>Totale</b>	<b>20</b>	<b>1</b>	<b>100%</b>

Distribuzione di frequenza del carattere quantitativo *Numero di componenti della famiglia*

Numero di componenti	$n_i$	$f_i$	$N_i$	$F_i$	%
1	1	0,05	1	0,05	5%
3	1	0,05	2	0,1	5%
4	8	0,4	10	0,5	40%
5	6	0,3	16	0,8	30%
6	2	0,1	18	0,9	10%
7	1	0,05	19	0,95	5%
8	1	0,05	20	1	5%
<b>Totale</b>	<b>20</b>	<b>1</b>			<b>100%</b>

**Esercizio 3)** Costruire la distribuzione di intensità per la variabile *Peso*

**3.1)** Suddividere 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:

$$\text{range (PESO)} = 93 - 51 = 42$$

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

$$\text{Ampiezza classi} = \text{range}/3 = 42/3 = 14$$

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

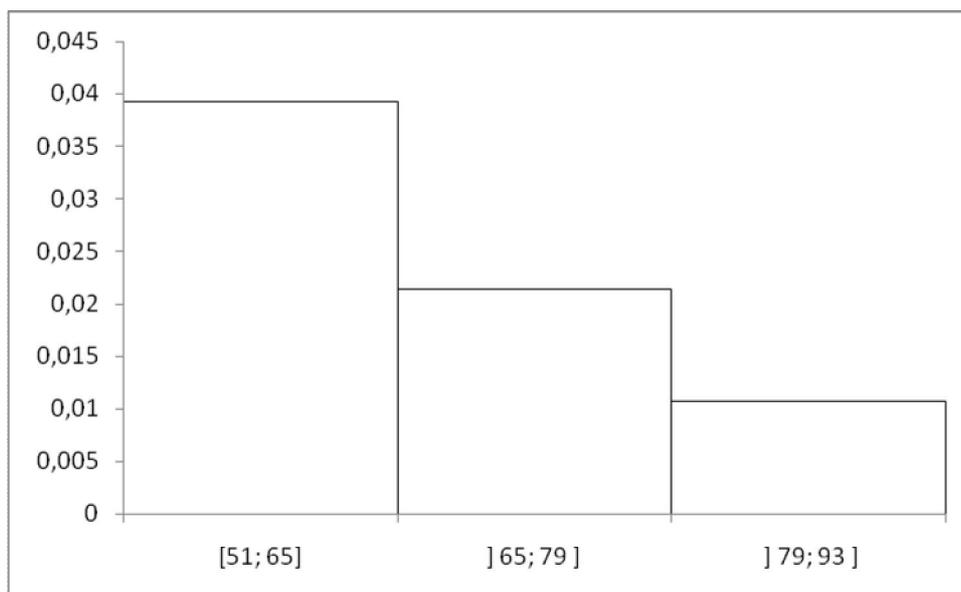
$$C_1 = [51; 65]$$

$$C_2 = ] 65; 79 ]$$

$$C_3 = ] 79; 93 ]$$

$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$	$\hat{c}_i$
$C_1 = [51; 65]$	11	0,55	11	0,55	14	0,03928571	58
$C_2 = ] 65; 79 ]$	6	0,3	17	0,83	14	0,02142857	72
$C_3 = ] 79; 93 ]$	3	0,15	20	1	14	0,01071429	86
<b>Totali</b>	<b>20</b>	<b>1</b>					

Istogramma della variabile PESO (classi equiampie)



**3.2)** Suddividere la distribuzione in 3 classi equifrequenti riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.

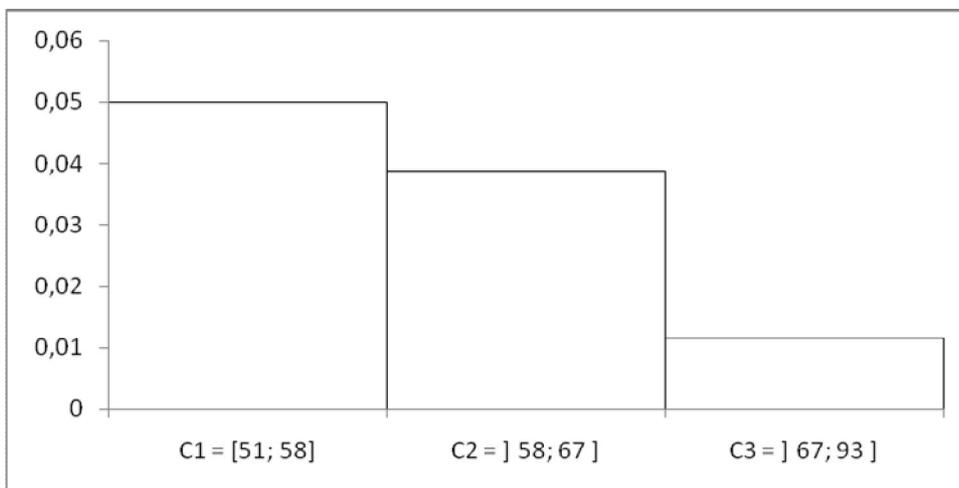
$N$  (numerosità totale) = 20

$20/3 = 6,67$

Avremo pertanto due classi con frequenza pari a 7 e una classe con frequenza pari a 6. La classe con frequenza 6 è la seconda.

$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$	$\hat{C}_i$
$C_1 = [51; 58]$	7	0,35	7	0,35	7	0,05	54,5
$C_2 = ] 58; 67 ]$	7	0,35	14	0,7	9	0,03888889	62,5
$C_3 = ] 67; 93 ]$	6	0,3	20	1	26	0,01153846	80
<b>Totali</b>	<b>20</b>	<b>1</b>					

Istogramma della variabile PESO (classi equifrequenti)



**Esercizio 4** A partire dalla distribuzione di frequenza del carattere *Peso*, calcolare la media e la moda. Elencare alcune delle proprietà della media.

Distribuzione di frequenza del carattere quantitativo *Peso*

<b>Peso</b>	$n_i$
51	2
55	3
58	2
60	2
63	1
65	1
66	1
67	2
71	1
75	1
76	1
84	1
89	1
93	1

## Media aritmetica

$$\mu = \frac{\sum_{i=1}^n x_i \times n_i}{n} = \frac{51 \times 2 + 55 \times 3 + 58 \times 2 + \dots + 93 \times 1}{20} = 69,5$$

## Moda

$$Mo_1 = 55$$

I Proprietà) la somma algebrica degli scarti dei termini della distribuzione è uguale a zero.

$$\sum_{i=1}^N (x_i - \mu_1) = 0$$

Poiché, qualunque sia la distribuzione, solo  $\mu$  gode di tale proprietà, la media aritmetica può anche definirsi come quel valore che rende nulla la somma algebrica degli scarti.

II Proprietà) la somma dei quadrati degli scarti dei termini della distribuzione della media aritmetica è un minimo rispetto alla somma dei quadrati degli scarti da un qualsiasi altro valore.

$$\sum_{i=1}^N (x_i - \mu)^2 = \min$$

Per qualsiasi  $K$  reale si ha che  $\sum_{i=1}^N (x_i - \mu)^2 \leq \sum_{i=1}^N (x_i - K)^2$