

# Università di Cassino (sede di Terracina)

## Esercitazione di Statistica 1 del 24 ottobre 2007

**Dott.ssa Paola Costantini**

### Esercizio n. 1

Utilizzando il dataset n. 2, si suddivida la distribuzione Riserve Naturali (soltanto per le Regioni con superficie qualitativa BASSA) in 3 classi equifrequenti.  
Rappresentare graficamente le due distribuzioni attraverso l'istogramma normalizzato.

### Dataset 2

REGIONI	SUPERFICIE AREE NATURALI PROTETTE (in migliaia di ettari)	SUPERFICIE A TERRA QUALITATIVA	PARCHI NAZIONALI	RISERVE NATURALI STATALI	AREE NATURALI MARINE PROTETTE	AZIENDE AGRITURISTICHE	AZIENDE AGRITURISTICHE DESTINATE ALL'ALLOGGIO	AZIENDE AGRITURISTICHE DESTINATE ALLA RISTORAZIONE	AZIENDE AGRITURISTICHE DESTINATE AD ALTRE ATTIVITA'
Piemonte	167,2	bassa	44,9	3,3	0,0	MOLTE	548,0	480,0	722,0
Valle d'Aosta	42,9	bassa	36,7	0,0	0,0	POCHE	44,0	26,0	28,0
Lombardia	130,2	bassa	59,7	0,2	0,0	MOLTE	447,0	697,0	413,0
Trentino-Alto Adige	283,4	media	73,6	0,0	0,0	MOLTISSIME	2.509,0	494,0	1.444,0
Veneto	93,4	bassa	15,1	19,4	0,0	MOLTE	536,0	611,0	678,0
Friuli-Venezia Giulia	55,1	bassa	0,0	0,4	1,3	ABBASTANZA	186,0	321,0	186,0
Liguria	577,9	alta	3,5	0,0	552,5	ABBASTANZA	257,0	212,0	71,0
Emilia-Romagna	88,1	bassa	30,7	8,2	0,0	MOLTE	456,0	538,0	519,0
Toscana	1.800,0	alta	39,9	11,0	1.640,1	MOLTISSIME	3.505,0	761,0	3.663,0
Umbria	63,3	bassa	17,9	0,0	0,0	MOLTISSIME	890,0	250,0	1.012,0
Marche	89,1	bassa	61,1	6,1	0,0	ABBASTANZA	444,0	287,0	213,0
Lazio	217,6	media	26,6	25,8	4,2	ABBASTANZA	316,0	290,0	264,0
Abruzzo	303,2	media	219,4	17,7	0,0	ABBASTANZA	404,0	281,0	198,0
Molise	6,4	bassa	4,0	1,2	0,0	POCHE	54,0	69,0	66,0
Campania	327,0	media	185,4	2,0	1,7	MOLTE	528,0	531,0	704,0
Puglia	149,1	bassa	120,0	9,9	20,3	ABBASTANZA	206,0	127,0	199,0
Basilicata	125,2	bassa	88,4	0,9	0,0	ABBASTANZA	222,0	133,0	226,0
Calabria	264,1	media	232,0	16,2	14,7	ABBASTANZA	267,0	261,0	154,0
Sicilia	346,7	media	0,0	0,0	75,9	ABBASTANZA	313,0	284,0	313,0
Sardegna	602,2	alta	84,5	0,0	509,7	MOLTE	461,0	548,0	224,0

## Esercizio n. 2

Calcolare mediana, moda, primo e terzo quartile della variabile Riserve Naturali per le sole Regioni con Superficie Qualitativa BASSA. (dataset n.2 )

## Esercizio n. 3

Per la variabile Numero Clienti (inclusa nell'insieme dati "Information Technology" (dataset 1):

**3.1** Si rappresenti un istogramma con 3 classi equifrequenti;

**3.2** Si rappresenti un istogramma con 2 classi di diversa ampiezza e frequenza;

**3.3** Si calcoli la media e la varianza a partire dai dati grezzi (serie di dati);

**3.4** Si calcoli la media e la varianza a partire dalla distribuzione organizzata in 3 classi equifrequenti;

**3.5** Si calcoli la media e la varianza a partire dalla distribuzione organizzata in due classi (di cui la punto 3.2).

**3.6** Si commentino brevemente eventuali differenze nei risultati ottenuti ai punti 3.3, 3.4, e 3,5.

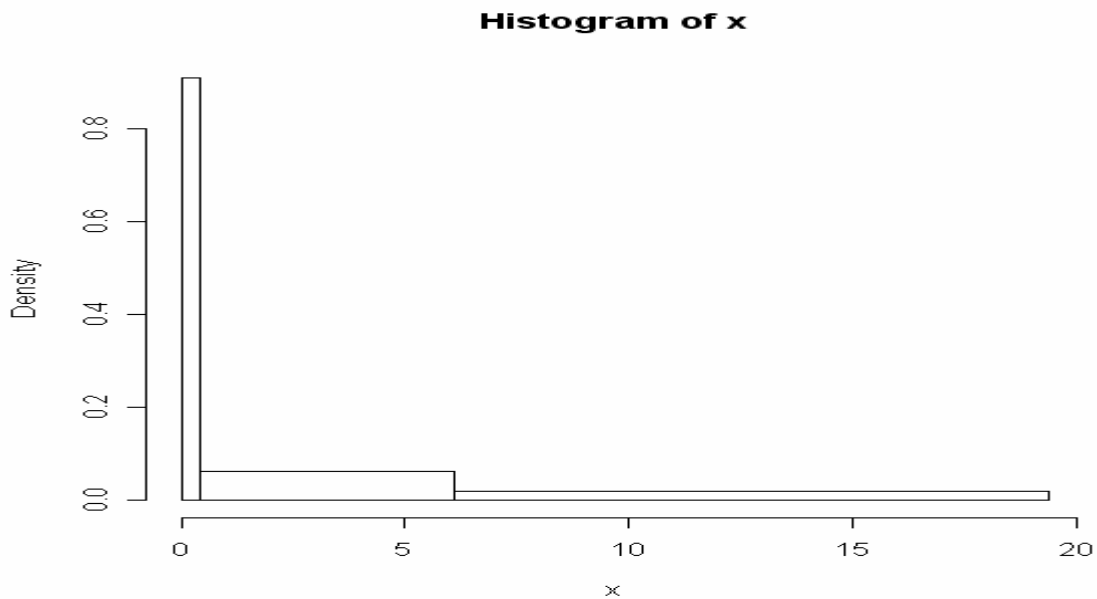
## SOLUZIONE

### Esercizio 1

La numerosità delle Regioni con superficie qualitativa BASSA è 11, quindi la frequenza costante nelle 3 classi dovrebbe essere pari a  $11/3 = 3,6667$ .

Considerando anche l'ampiezza delle classi decidiamo:

$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$
[0; 0,4 ]	4	0,365	4	0,365	0,4	0,9125
]0,4; 6,1 ]	4	0,365	8	0,73	5,7	0,0640
] 6,1; 19,4 ]	3	0,270	11	1	13,3	0,0203
<b>Totali</b>	<b>11</b>	<b>1,00</b>				



## Esercizio n. 2

$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$
$[0; 0,4 ]$	4	0,365	4	0,365	0,4	0,9125
$]0,4; 6,1 ]$	4	0,365	8	0,73	5,7	0,0640
$]6,1; 19,4 ]$	3	0,270	11	1	13,3	0,0203
<b>Totali</b>	<b>11</b>	<b>1,00</b>				

Il calcolo della mediana per la ripartizione di una distribuzione in classi avviene attraverso al formula:

**Mediana**

$$M e \cong x_{M e-1} + (x_{M e} - x_{M e-1}) \frac{0,5 - F_{M e-1}}{F_{M e} - F_{M e-1}}$$

$$M e = 0,4 + (6,1-0,4) \frac{0,5 - 0,4}{0,73- 0,4} = 2,13$$

La **classe modale** è quella con la densità di frequenza più elevata, ossia la prima:  $[0; 0,4 ]$

Il calcolo del primo quartile per la ripartizione di una distribuzione in classi avviene attraverso al formula:

**Primo quartile**

$$Q_1 \cong x_{Q_1-1} + (x_{Q_1} - x_{Q_1-1}) \frac{0,25 - F_{Q_1-1}}{F_{Q_1} - F_{Q_1-1}}$$

$$Q_1 = 0 + (0,4 - 0) \frac{0,25 - 0}{0,365 - 0} = \mathbf{0,27}$$

Il calcolo del terzo quartile per la ripartizione di una distribuzione in classi avviene attraverso al formula:

**Terzo quartile**

$$Q_3 \cong x_{Q_3-1} + (x_{Q_3} - x_{Q_3-1}) \frac{0,75 - F_{Q_3-1}}{F_{Q_3} - F_{Q_3-1}}$$

$$Q_3 = 6,1 + (19,4 - 6,1) \frac{0,75 - 0,73}{1 - 0,73} = \mathbf{7,08}$$

**3.1 Si rappresenti un istogramma con 3 classi equifrequenti della variabile Numero Clienti, inclusa nel dataset 1.**

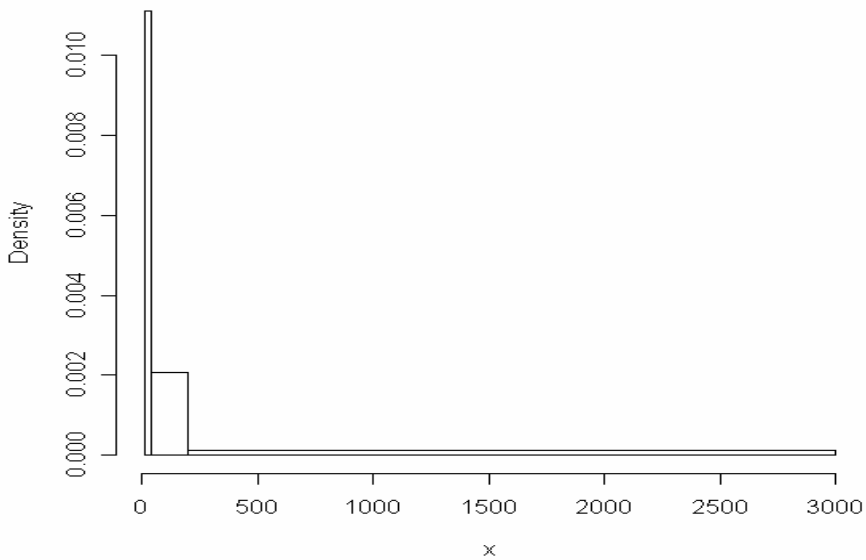
**Dataset 1**

Id Az.	App Gruppo	Fatturato (milioni di euro)	% Fatturato hardware	% Fatturato software	Numero Clienti	Numero dipendenti	Loc Clienti	Area att. Clienti	Importanza mercato
1	1	32,0	15	15	1500	350	N	1	1
2	0	4,0	30	35	40	55	N	2	4
3	1	1,0	33	34	30	7	L	3	2
4	1	17,0	0	40	350	210	N	4	3
5	1	6,0	0	40	30	95	N	4	2
6	1	1,5	30	40	300	12	N	3	1
7	1	2,0	20	55	1000	30	N	1	2
8	0	12,5	5	30	100	100	L	1	2
9	1	25,5	0	40	10	260	L	1	2
10	1	5,0	20	20	200	55	N	1	2
11	1	2,5	0	50	150	7	N	3	3
12	1	23,0	10	85	3000	200	N	3	3
13	0	5,0	5	40	150	100	N	2	3
14	1	2,0	20	40	200	14	N	3	3
15	0	2,5	0	90	40	70	N	1	1

Numero Clienti
10
30
30
40
40
100
150
150
200
200
300
350
1000
1500
3000

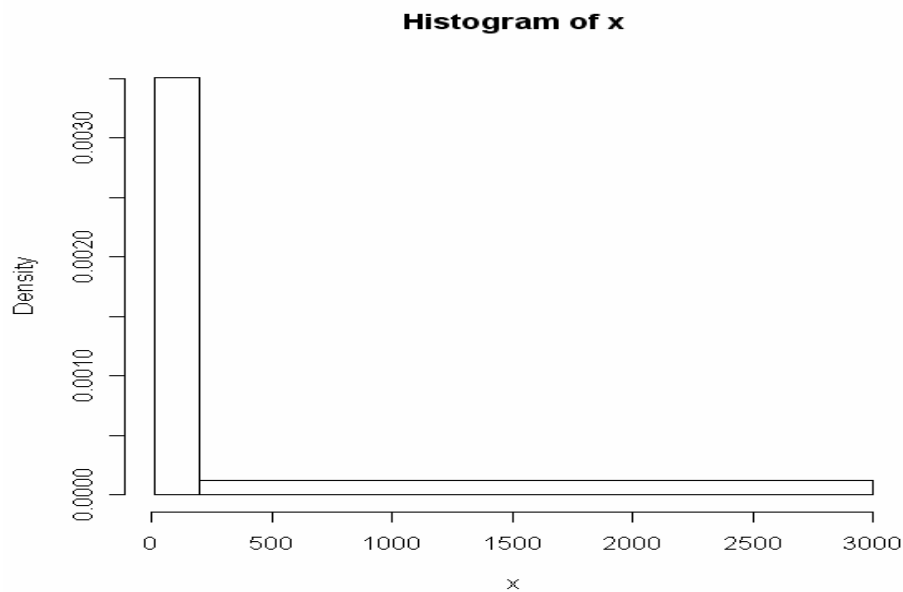
$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$
[10; 40 ]	5	0,33333	5	0,33333	30	0,0111
] 40; 200 ]	5	0,33333	10	0,33333	160	0,0021
] 200; 3000 ]	5	0,33334	15	1	2800	0,0001
<b>Totali</b>	<b>15</b>	<b>1,00</b>				

Histogram of x



**3.2 Si rappresenti un istogramma con 2 classi di diversa ampiezza e frequenza;**

$C_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$
[10; 200 ]	10	0,67	10	0,67	190	0,0035
] 200; 3000 ]	5	0,33	15	1	2800	0,0001
<b>Totali</b>	<b>15</b>	<b>1,00</b>				



**3.3 Si calcoli la media e la varianza a partire dai dati grezzi (serie di dati);**

Calcolo della media e della varianza di una successione di valori:

**Media**

$$\mu_x = \frac{\sum_{i=1}^n x_i}{n} = 473,333$$

**Varianza**

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = 616.128,9$$

**3.4 Si calcoli la media e la varianza a partire dalla distribuzione organizzata in 3 classi equifrequenti;**

<b>C<sub>i</sub></b>	<b>n<sub>i</sub></b>	<b>f<sub>i</sub></b>	<b>N<sub>i</sub></b>	<b>F<sub>i</sub></b>	<b>a<sub>i</sub></b>	<b>d<sub>i</sub></b>
[10; 40 ]	5	0,33333	5	0,33333	30	0,0111
] 40; 200 ]	5	0,33333	10	0,33333	160	0,0021
] 200; 3000 ]	5	0,33334	15	1	2800	0,0001
<b>Totali</b>	<b>15</b>	<b>1,00</b>				

Per una distribuzione in classi di frequenza, la media e la varianza si calcolano seguendo le seguenti formule:

**Media**

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

dove

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

**Varianza**

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^k (c_i - \mu)^2 n_i$$

<b>C<sub>i</sub></b>	<b>n<sub>i</sub></b>	<b>f<sub>i</sub></b>	<b>N<sub>i</sub></b>	<b>F<sub>i</sub></b>	<b>C<sub>i</sub></b>
[10; 40 ]	5	0,33333	5	0,33333	30
] 40; 200 ]	5	0,33333	10	0,33333	120
] 200; 3000 ]	5	0,33334	15	1	1600
<b>Totali</b>	<b>15</b>	<b>1,00</b>			

$$\mu_x = \frac{25 \cdot 5 + 120 \cdot 5 + 1600 \cdot 5}{15} = \mathbf{581,67}$$

$$\sigma_x^2 = \frac{(25 - 581,67)^2 \cdot 5 + (120 - 581,67)^2 \cdot 5 + (1600 - 581,67)^2 \cdot 5}{15} = \mathbf{520.005,55}$$

**3.5 Si calcoli la media e la varianza a partire dalla distribuzione organizzata in due classi (di cui la punto 3.2).**

<b>C<sub>i</sub></b>	<b>n<sub>i</sub></b>	<b>f<sub>i</sub></b>	<b>N<sub>i</sub></b>	<b>F<sub>i</sub></b>	<b>C<sub>i</sub></b>
[10; 200 ]	10	0,67	10	0,67	105
] 200; 3000 ]	5	0,33	15	1	1600
<b>Totali</b>	<b>15</b>	<b>1,00</b>			

$$\mu_x = \frac{105*10+1600*5}{15} = \frac{1050+8000}{15} = \mathbf{603,333}$$

$$\sigma_x^2 = \frac{(105-603,333)^2 * 10 + (1600-603,333)^2 * 5}{15} = \mathbf{496.672,2}$$

**3.6 Si commentino brevemente eventuali differenze nei risultati ottenuti ai punti 3.3, 3.4, e 3,5.**

	<b>Media</b>	<b>Varianza</b>
<b>Dati grezzi</b>	473,333	616.128,9
<b>Due classi</b>	581,67	520.005,55
<b>Tre classi</b>	603,333	496.672,2

Con i dati grezzi i risultati ottenuti sono ESATTI, mentre con il raggruppamento in classi si perde informazione. Con i dati grezzi il valor medio è più basso, mentre la varianza è più ampia rispetto alle distribuzioni suddivise in classi. All'aumentare del numero di classi per distribuzione, migliora l'approssimazione.