

Università di Cassino (sede di Terracina)

Esercitazione di Statistica 1 del 17 ottobre 2007

Dott.ssa Paola Costantini

La seguente tabella (tab1) riporta le informazioni relative a 15 aziende operanti nel settore dell'*information technology*.

Appartenenza gruppo	E' un'azienda indipendente o appartiene ad un gruppo di aziende? (1=si, 0=no)
Fatturato	Fatturato dell'azienda (in milioni di euro)
% Fatturato hardware	% Fatturato in hardware
% Fatturato Software	% Fatturato in Software
Numero Clienti	Numero Clienti
Numero dipendenti	Numero di dipendenti
Localizzazione clienti	Localizzazione dei clienti (Nazionale, Locale)
Area attività clienti	Area di attività dei clienti (principale): Banche e Assicurazioni, Società di servizi, Industria e Commercio, Pubblica Amministrazione)
Importanza mercato	Indice di importanza sul mercato nazionale (misurata su una scala che va da 1 a 4, dove 1 rappresenta notevole importanza e 4 nessuna importanza)

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

Esercizio n.1

Per ognuna delle variabili che compongono il **dataset 1**, si definisca il tipo di variabile (quantitativo/qualitativo1) elencando l'insieme delle possibili modalità/intensità.

Esercizio n. 2

Costruire le distribuzioni di frequenza per i caratteri qualitativi riportando le frequenze assolute e le frequenze relative. Rappresentare graficamente le distribuzioni ottenute.

Esercizio n. 3

Nel **dataset 2**, sono riportati i dati concernenti 9 caratteri statistici rilevati sulle 20 regioni italiane: la superficie delle aree naturali protette, la superficie qualitativa (0-200.000 ettari = bassa; 200.000-500.000 ettari = media; oltre 500.000 ettari = alta), la superficie dei parchi nazionali per regione, la superficie delle riserve naturali regionali, la superficie delle aree naturali marine protette, la quantità di aziende agrituristiche presenti in ciascuna delle regioni italiane, e la ripartizione delle aziende agrituristiche in base alla destinazione: all'alloggio, alla ristorazione, ad altre attività.

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

3) Per la variabile "superficie qualitativa a terra", si consideri soltanto la modalità "bassa"; costruire la distribuzione di intensità per il carattere NUMERO DI PARCHI NAZIONALI.

3.1) Suddividere la distribuzione in 3 classi equampie riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.

Rappresentare graficamente le due distribuzioni attraverso l'istogramma normalizzato.

3.2) Suddividere la distribuzione in 3 classi equipendenti riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.
Rappresentare graficamente le due distribuzioni attraverso l'istogramma normalizzato.

Esercizio n. 4

Si calcolino la mediana e la moda della variabile % Fatturato hardware e Localizzazione Clienti. (dataset 1)

Esercizio n. 5

Si calcolino la mediana e la moda per le distribuzioni equiampie ed equipendenti, di cui ai punti 3.1 e 3.2. (dataset 2).

SOLUZIONE

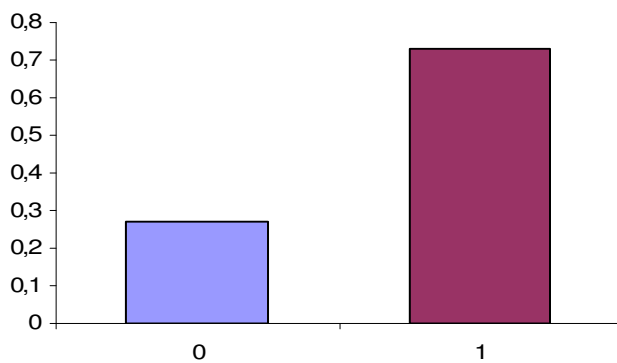
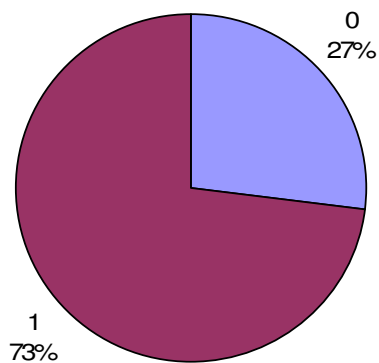
1)

Carattere	Tipo	Modalità
Appartenenza gruppo	Qualitativo	{0;1}
Fatturato	Quantitativo	{1,5; 32,0}
% Fatturato hardware	Quantitativo	{0; 33}
% Fatturato Software	Quantitativo	{15; 90}
Numero Clienti	Quantitativo	{10; 3000}
Numero dipendenti	Quantitativo	{7; 350}
Localizzazione clienti	Qualitativo	{N;L}
Area attività clienti	Qualitativo	{1;2; 3; 4}
Importanza mercato	Qualitativo	{1;4}

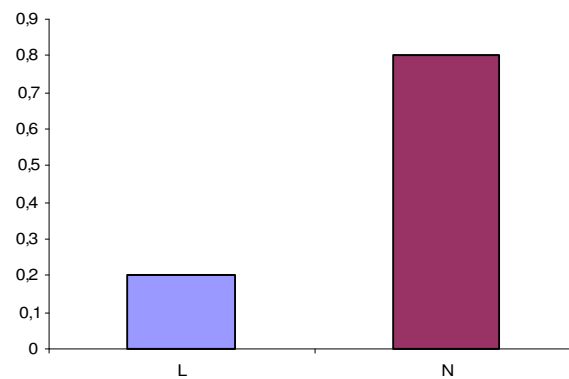
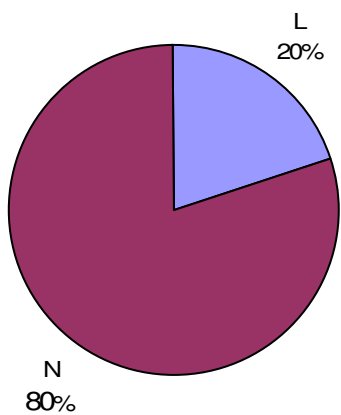
Appartenenza gruppo	Localizzazione Clienti	Area Attività Clienti	Importanza mercato
0= no 1=si	L = Locale N = Nazionale	1 = Pubblica Ammin. 2 = Società Servizi 3 = Industria e Comm. 4 = banche e Ass.	1 = poco 2 = abbastanza 3 = molto 4 = moltissimo

2) Distribuzioni di frequenza e rappresentazioni grafiche dei caratteri qualitativi:

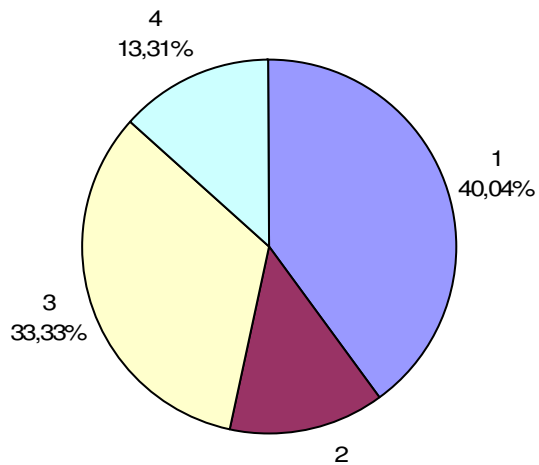
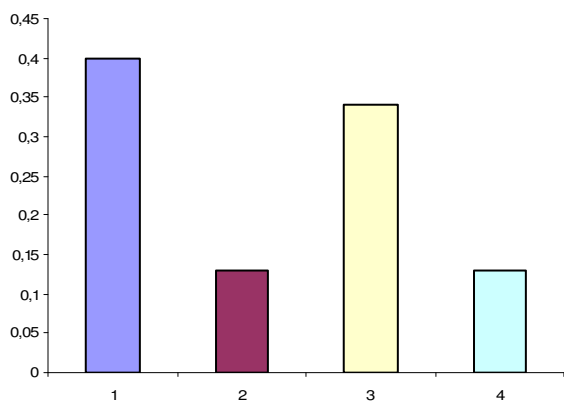
Appartenenza gruppo	n_i	f_i	%
0	4	0,27	27%
1	11	0,73	73%
Totale	15	1	100%



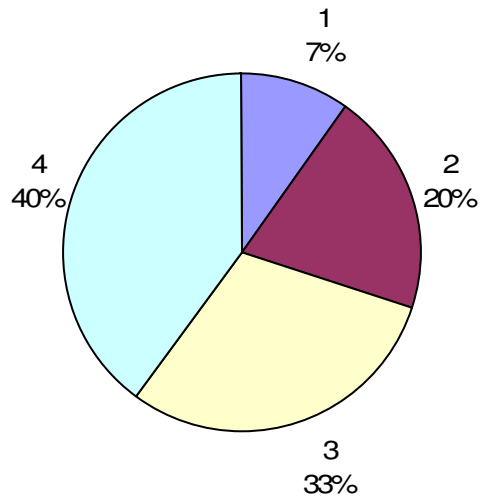
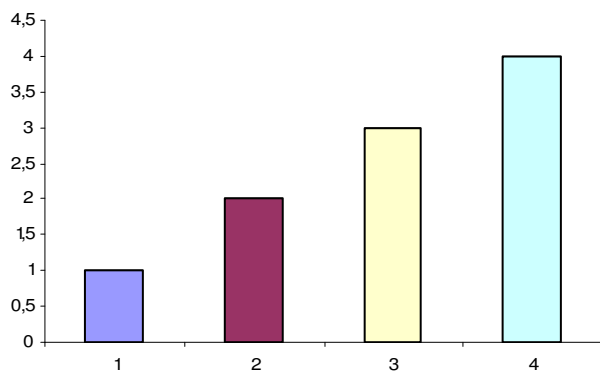
Localizzazione Clienti	n_i	f_i	%
L	3	0,20	20%
N	12	0,80	80%
Totale	15	1	100%



Area attività Clienti	n_i	f_i	%
1	6	0,40	40%
2	2	0,13	13,33%
3	5	0,34	33,33%
4	2	0,13	13,33%
Totale	15	1	100%



Importanza mercato	n _i	f _i	%
1	3	0,2	20%
2	6	0,4	40%
3	5	0,33	33%
4	1	0,07	7%
Totale	15	1	100%



3) Distribuzioni di intensità del carattere PARCHI NAZIONALI per la modalità di SUPERFICIE QUALITATIVA " BASSA "

Migliaia di ettari Di parchi naturali <u>Superficie bassa</u>
0,0
4,0
15,1
17,9
30,7
36,7
44,9
59,7
61,1
88,4
120,0

3.1 L'ampiezza costante delle tre classi si ottiene come:

$$range (SUPERFICIE BASSA) = 120,0 - 0,0 = 120,0$$

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

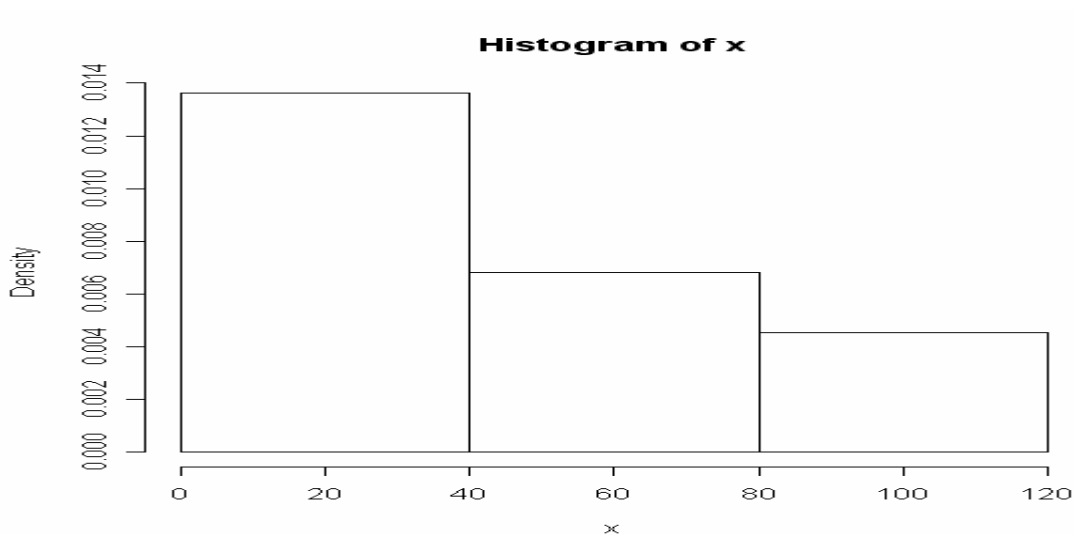
Ampiezza **Sup. Bassa** = range/3 = 120/3 = 40

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

$C_1 = [0; 40]$
 $C_2 =] 40; 80]$
 $C_3 =] 80; 120]$

C_i	n_i	f_i	N_i	F_i	a_i	d_i
[0; 40]	6	0,55	6	0,55	40	0,0137
] 40; 80]	3	0,27	9	0,82	40	0,0067
] 80; 120]	2	0,18	11	1	40	0,0045
Totali	11	1,00				

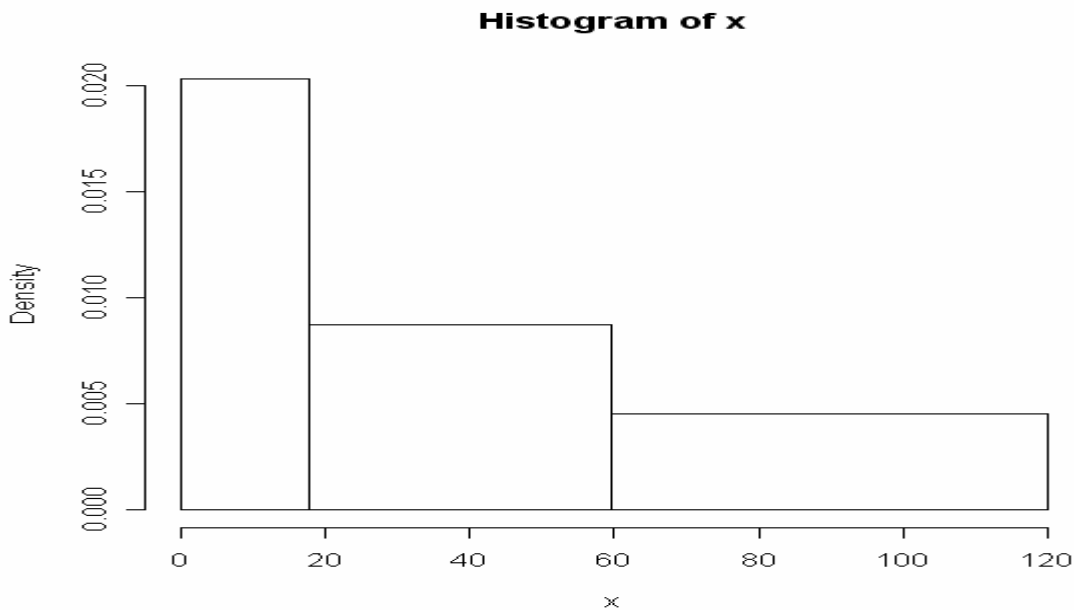
Il relativo istogramma normalizzato è il seguente:



3.2 La numerosità delle Regioni con superficie qualitativa BASSA è 11, quindi la frequenza costante nelle 3 classi dovrebbe essere pari a $11/3 = 3,6667$, che non ha senso ma suggerisce di scegliere una frequenza pari a 3 o a 4. Considerando anche l'ampiezza delle classi decidiamo:

C_i	n_i	f_i	N_i	F_i	a_i	d_i
[0; 17,9]	4	0,365	4	0,365	17,9	0,0203
]17,9; 59,7]	4	0,365	8	0,73	41,8	0,0087
] 59,7; 120]	3	0,270	11	1	60,3	0,0045
Totali	11	1,00				

Il relativo istogramma normalizzato è il seguente:



Esercizio n. 4

Carattere Fatturato hardware: (dati ordinati)

0 0 0 0 0 5 5 **10** 15 20 20 20 30 30 33

Se N è un numero dispari la mediana si calcola con la formula

$$\mathbf{Me} = \frac{X_{N+1}}{2} = \frac{X_{15+1}}{2} = X_{16} = 10$$

Con i dati ordinati, alla 8° posizione c'è 10. La moda è 0 con $n_i=5$

Per la variabile Localizzazione Clienti non ha senso calcolare la mediana perchè è una variabile qualitativa sconnessa. Per la variabile Localizzazione Clienti la moda è **Nazionale** con $n_i = 12$

Esercizio n. 5

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

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

5.1 Per le classi equiampie di cui al punto 3.1:

C_i	n_i	f_i	N_i	F_i	a_i	d_i
[0; 40]	6	0,55	6	0,55	40	0,0137
] 40; 80]	3	0,27	9	0,82	40	0,0067
] 80; 120]	2	0,18	11	1	40	0,0045
Totali	11	1,00				

$$Me = 0 + (40 - 0) * \frac{0,5 - 0}{0,55 - 0} = 36,36$$

La classe modale è quella che presenta la frequenza assoluta più grande: in questo caso la classe modale è la prima: [0; 40];

5.2 Per le classi equifrequenti di cui al punto 3.2, la mediana è:

C_i	n_i	f_i	N_i	F_i	a_i	d_i
[0; 17,9]	4	0,365	4	0,365	17,9	0,0203
]17,9; 59,7]	4	0,365	8	0,73	41,8	0,0087
] 59,7; 120]	3	0,270	11	1	60,3	0,0045
Totali	11	1,00				

$$Me = 17,9 + (59,7 - 17,9) * \frac{(0,5 - 0,365)}{0,73 - 0,365} = 33,3$$

La classe modale è quella con la densità di frequenza più grande, e cioè la prima: [0; 17,9].