

Università di Cassino

Esercitazione di Statistica 1 del 16 ottobre 2005

Dott.ssa Simona Balzano

La seguente tabella, chiamata DATASET STUDENTI, riporta i dati concernenti 9 caratteri statistici rilevati su 20 studenti frequentanti un corso di Statistica: sesso, altezza in centimetri, peso in chilogrammi, corso di laurea, numero di scarpa, colore degli occhi, colore dei capelli, attività sportiva praticata, diploma scolastico conseguito.

N	SESSO	ALTEZZA (cm)	PESO (kg)	CORSO LAUREA	NUMERO SCARPA	COLORE OCCHI	COLORE CAPELLI	ATT SPORTIVA	DIPLOMA
1	M	179	65	I	43	1	1	2	3
2	M	180	62	I	43	1	1	2	3
3	F	165	50	I	39	1	2	1	1
4	F	160	49	I	37	2	2	2	3
5	F	160	47	M	37	3	3	2	5
6	F	160	48	M	36	1	3	2	1
7	F	164	56	M	38	2	2	2	2
8	F	170	59	M	38	1	2	2	1
9	M	180	73	M	43	2	2	1	1
10	M	186	86	M	45	3	1	1	3
11	M	170	66	M	42	1	2	2	1
12	M	180	68	I	41	1	1	3	3
13	M	180	85	I	43	1	2	2	3
14	F	176	56	I	37	1	3	2	1
15	M	170	72	I	42	2	2	2	5
16	M	180	65	I	42	1	1	1	3
17	M	170	75	I	41	1	1	2	3
18	M	172	70	I	40	1	2	1	1
19	M	178	80	I	45	1	2	1	3
20	F	162	49	B	37	1	1	2	2

Le modalità dei caratteri CORSO DI LAUREA, COLORE OCCHI, COLORE CAPELLI, ATTIVITÀ SPORTIVA e DIPLOMA sono codificate come risulta dalla seguente tabella:

CORSO LAUREA :	COLORE OCCHI	COLORE CAPELLI	ATT.SPORTIVA	DIPLOMA :
M= matematica B = biologia I = informatica	1 = scuri 2 = verdi 3 = azzurri	1 = scuri 2 = castani 3 = biondi	1 = nulla 2 = media 3 = alta	1 = liceo scientifico 2 = liceo classico 3 = ist. tecnico 4 = ist. magistrale 5 = altro

1) Definire, per ciascuno dei caratteri riportati in tabella, il tipo elencando l'insieme delle possibili modalità/intensità.

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

3) Considerando separatamente gli studenti maschi e femmine, costruire le due distribuzioni di intensità per il carattere PESO (peso dei maschi e peso delle femmine), in modo da favorire il confronto tra i due gruppi.

3.1) Suddividere ciascuna distribuzione in 2 classi equiampie ed effettuare il confronto tra distribuzioni riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.

Rappresentare graficamente le due distribuzioni attraverso l'istogramma normalizzato. Commentare brevemente i risultati ottenuti.

3.2) Suddividere ciascuna distribuzione in 2 classi equifrequenti ed effettuare il confronto tra distribuzioni riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.

Rappresentare graficamente le due distribuzioni attraverso l'istogramma normalizzato. Commentare brevemente i risultati ottenuti.

Soluzione

1)

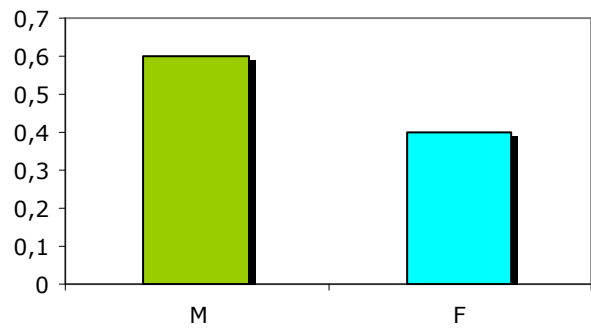
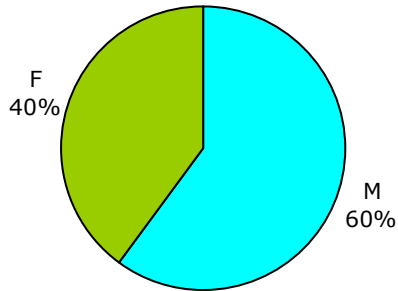
Carattere	Tipo	Modalità
SESSO	qualitativo	{M, F}
ALTEZZA	quantitativo	[160, 186]
PESO	quantitativo	[47, 86]
CORSO LAUREA	qualitativo	{B, I, M}
NUMERO SCARPA	quantitativo	[36, 45]
COLORE OCCHI	qualitativo	{1, 2, 3}
COLORE CAPELLI	qualitativo	{1, 2, 3}
ATTIVITÀ SPORTIVA	qualitativo ordinale	{1, 2, 3}
DIPLOMA	qualitativo ordinale	{1, 2, 3, 4, 5}

In modo equivalente, avremmo potuto definire gli insiemi di modalità delle variabili qualitative riportando i valori originali piuttosto che la codifica numerica.

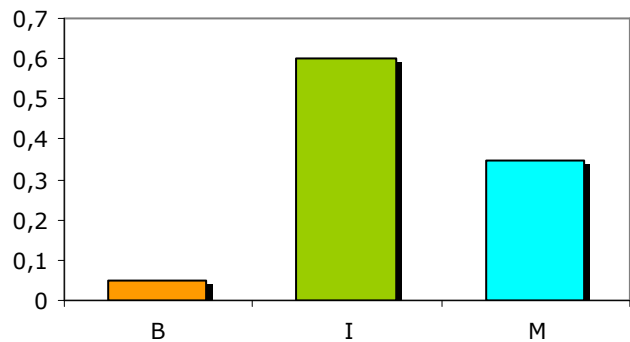
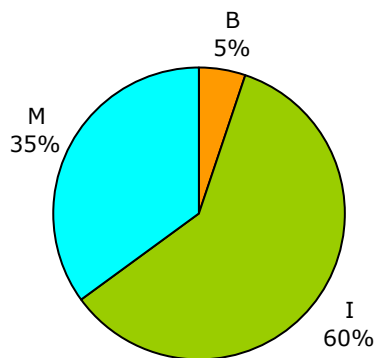
2)

Distribuzioni di frequenza e rappresentazioni grafiche dei caratteri qualitativi:

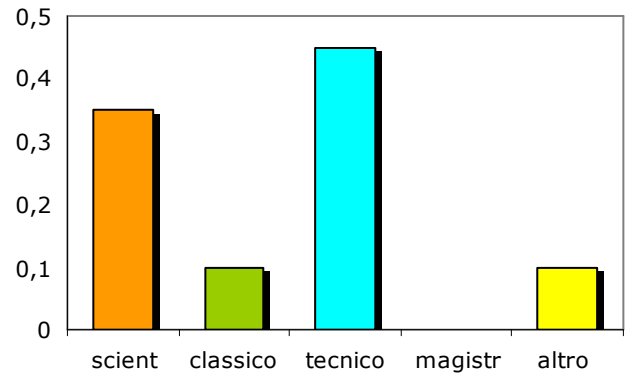
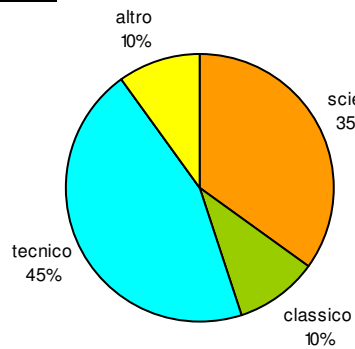
SESSO	n_i	f_i	%
M	12	0,60	60
F	8	0,40	40
Totale	20	1	100



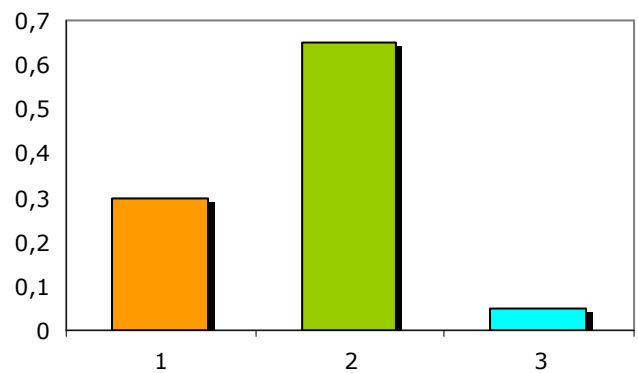
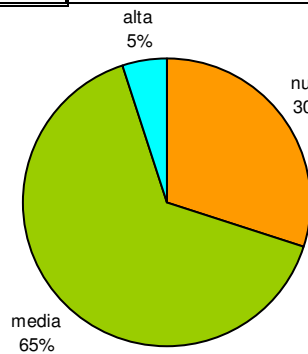
CLAUREA	n_i	f_i	%
B	1	0,05	5
I	12	0,60	60
M	7	0,35	35
Totale	20	1	100



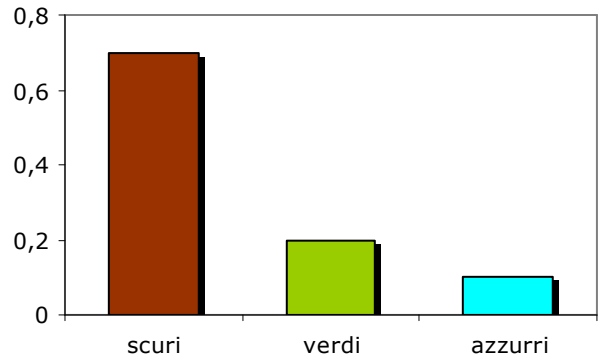
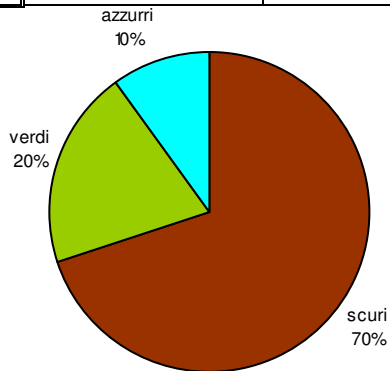
DIPLOMA	n_i	f_i	%
1	7	0,35	35
2	2	0,10	10
3	9	0,45	45
4	0	0,00	0
5	2	0,10	10
Totale	20	1	100



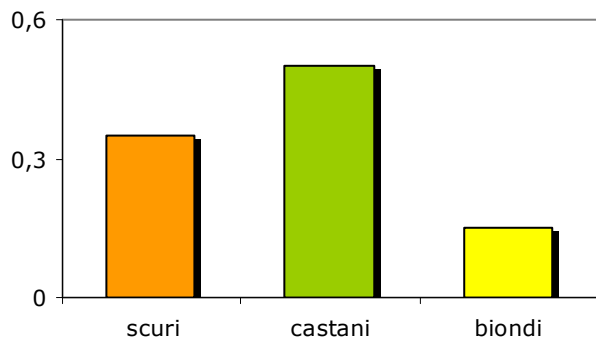
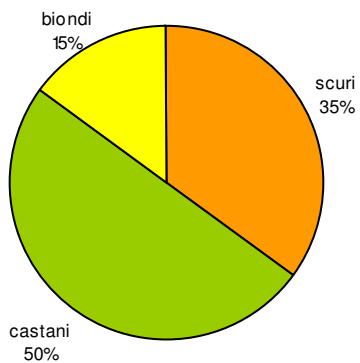
ATT.SPORT.	n_i	f_i	%
1	6	0,30	30
2	13	0,65	65
3	1	0,05	5
Totale	20	1	100



OCCHI	n_i	f_i	%
1	14	0,70	70
2	4	0,20	20
3	2	0,10	10
Totale	20	1	100



CAPELLI	n_i	f_i	%
1	7	0,35	35
2	10	0,50	50
3	3	0,15	15
Totale	20	1	100



3)

Distribuzioni di intensità del carattere PESO per le due modalità di SESSO:

peso maschi	peso femmine
62	47
65	48
65	49
66	49
68	50
70	56
72	56
73	57
75	
80	
85	
86	

3.1)

L'ampiezza costante delle tre classi si ottiene come:

$$\text{range (PESO MASCHI)} = 86 - 62 = 24$$

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

$$a_{\text{maschi}} = \frac{\text{range}}{3} = \frac{24}{3} = 8$$

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

$$C_1 = [62; 70]$$

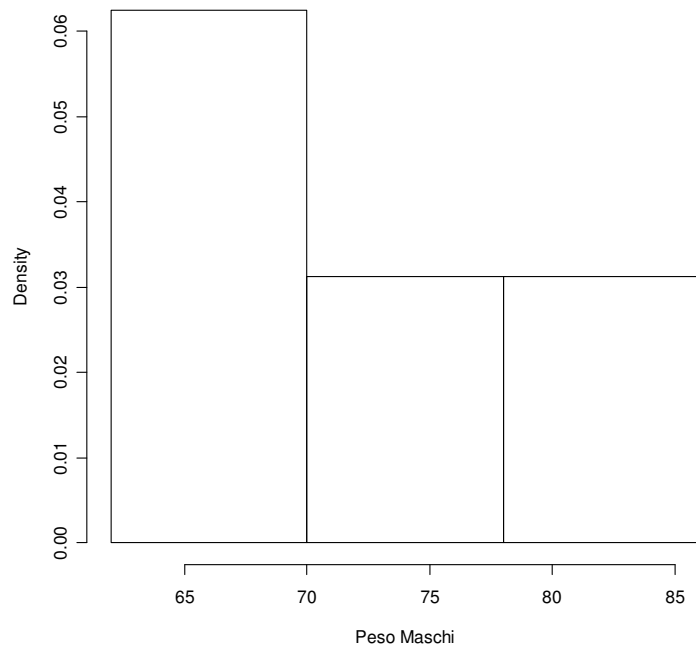
$$C_2 =] 70; 78]$$

$$C_3 =] 78; 86]$$

La relativa distribuzione di frequenza è:

C_i	n_i	f_i	N_i	F_i	a_i	d_i
[62; 70]	6	0,50	6	0,50	8	0,0625
] 70; 78]	3	0,25	9	0,75	8	0,03125
] 78; 86]	3	0,25	12	1	8	0,03125
Totali	12	1,00				

Il relativo istogramma normalizzato è il seguente:



Per quanto riguarda il secondo gruppo:

$$\text{range (FEMMINE)} = 59 - 47 = 12$$

$$a_{\text{femmine}} = \frac{\text{range}}{3} = \frac{12}{3} = 4$$

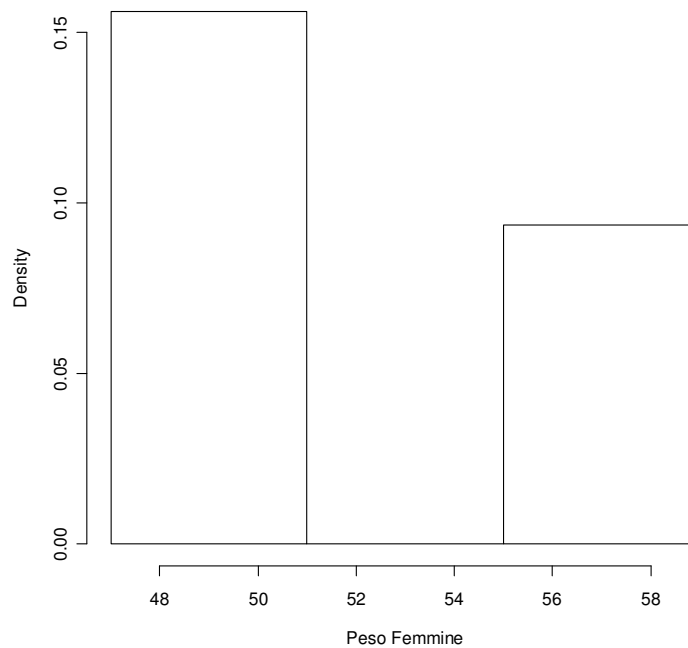
$$C_1 = [47; 51]$$

$$C_2 =] 51; 55]$$

$$C_3 =] 55; 59]$$

C_i	n_i	f_i	N_i	F_i	a_i	dn_i
[47; 51]	5	0,625	5	0,625	4	0,15625
] 51; 55]	0	0,000	5	0,625	4	0
] 55; 59]	3	0,375	8	1	4	0,09375
Totali	8	1,000				

Il relativo istogramma normalizzato è il seguente:



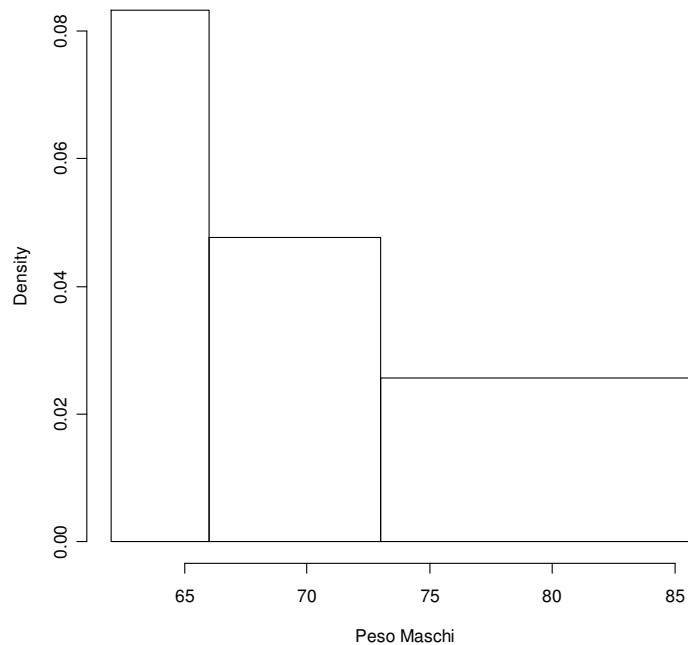
3.2)

La frequenza delle 3 classi equifrequenti è ottenuta dividendo per 3 la numerosità totale di ciascuna distribuzione semplice.

Per i maschi la numerosità è 12, quindi la frequenza costante all'interno delle classi sarà pari a 4.

C_i	n_i	f_i	N_i	F_i	a_i	d_i
[62; 66]	4	0,33	4	0,33	4	0,0825
] 66; 73]	4	0,33	8	0,66	7	0,047143
] 73; 86]	4	0,34	12	1	13	0,026154
Totali	12	1,00				

Il relativo istogramma normalizzato è il seguente:



Per le femmine la numerosità è 8, quindi la frequenza costante nelle 3 classi dovrebbe essere pari a $8/3 = 2,6667$, che ovviamente non ha senso ma suggerisce di scegliere una frequenza pari a 2 o a 3.

Osservando la successione di valori si vede che ci sono valori che si presentano più di una volta e che, quindi, possono vincolare la scelta.

In base a questo la soluzione migliore sembra determinare la prima classe in modo che abbia frequenza pari a 2 e le altre due con frequenza pari a 3, come mostra la tabella:

C_i	n_i	f_i	N_i	F_i	a_i	dn_i
[47; 48]	2	0,250	2	0,250	2	0,125
] 48; 50]	3	0,375	5	0,626	1	0,375
] 50; 59]	3	0,375	8	1	9	0,041
Totali	8	1,00				

Il relativo istogramma normalizzato è il seguente:

