Università degli Studi di Cassino
Anno accademico 2003-2004
Corsi di Statistica 1, II (Prof. G. Prozio) e Statistica 1, IV (Dott. D. Vistocco)

Esercitazione del 16/2/2004
Dott. Claudio Conversano

Esercizio 1

Si considerino le seguenti due distribuzioni, relative al voto riportato all'esame di Statistica da un campione di 200 studenti che hanno sostenuto l'esame nell'a.a. 2001/2002 e da un campione di studenti che hanno sostenuto l'esame nell'a.a. 2002/2003.

Voto esame di Statistica a.a. 2001/02	Frequenze assolute
21	6
23	26
24	39
25	57
26	44
27	19
28	9
Totale	200

Voto esame di Statistica a.a. 2002/03	Frequenze assolute
21	4
23	31
24	41
25	47
26	47
27	24
28	4
Totale	200

Misurare la forma delle due distribuzioni.

Svolgimento

Si tratta delle due distribuzioni già considerate nell'esercitazione precedente, per le quali sono stati calcolati gli indici di variabilità.

I risultati ottenuti sono stati i seguenti:

Voto esame di Statistica a.a. 2001/02

$Range = x_{\max} - x_{\min} = 28 - 21 = 7$; $\quad Q_1 = 24; \quad Q_2 = Me = 25; \quad Q_3 = 26;$[1]**:**

$$IQR = Q_3 - Q_1 = 26 - 24 = 2; \qquad \sigma_X^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right)^2 = 2,20$$

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{2,20} = 1,48 \qquad CV = \frac{\sigma_X}{\mu_X} = \frac{1,48}{24,97} = 0,06$$

[1] Per le modalità di calcolo dei quartili consultare l'esercitazione sugli indici di posizione.

Voto esame di Statistica a.a. 2002/03

$Range = x_{\max} - x_{\min} = 29 - 21 = 8$ **;** $\quad Q_1 = 24; \quad Q_2 = Me = 25; \quad Q_3 = 26;$[2]**:**

$IQR = Q_3 - Q_1 = 26 - 24 = 2; \qquad \sigma_X^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right)^2 = 2,20$

$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{2,20} = 1,48 \qquad CV = \frac{\sigma_X}{\mu_X} = \frac{1,48}{24,98} = 0,06$

La forma delle due distribuzioni sarà analizzata in 3 modi:

a) graficamente, attraverso gli istogrammi normalizzati di frequenza;
b) graficamente, attraverso i diagrammi a scatola (boxplot)
c) analiticamente, attraverso l'indice di Yule e Bowley
d) analiticamente, attraverso l'indice di Fisher

Istogrammi normalizzati di frequenza: si costruisce per le due distribuzioni la distribuzione di frequenza considerando 5 classi equiampie e su tali distribuzioni si disegnano gli istogrammi normalizzati di frequenza.

Distribuzione del **Voto esame di Statistica a.a. 2001/02**.

Ampiezza delle classi: $d_i = \frac{x_{\max} - x_{\min}}{k} = \frac{28 - 21}{5} = 1,40$

Distribuzioni in classi

Classi	Frequenze assolute	Frequenze relative	Frequenze relative cumulate	Densità di frequenza
21-|22,40	6	0,03	0,03	0,021
22,40-|23,80	26	0,13	0,16	0,093
23,80-|25,20	96	0,48	0,64	0,343
25,20-|26,60	44	0,22	0,86	0,157
26,60-|28,00	28	0,14	1,00	0,100
Totale	200	1,00		

Distribuzione del **Voto esame di Statistica a.a. 2001/02**.

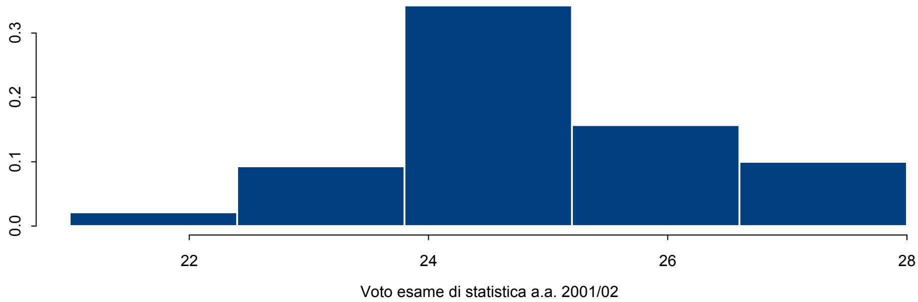
Ampiezza delle classi: $d_i = \frac{x_{\max} - x_{\min}}{k} = \frac{29 - 21}{5} = 1,60$

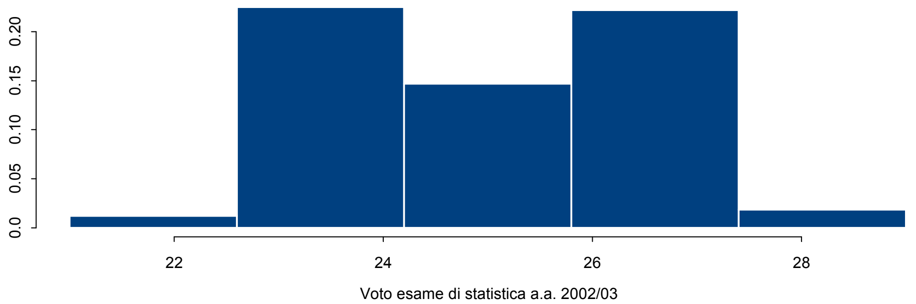
[2] Per le modalità di calcolo dei quartili consultare l'esercitazione sugli indici di posizione.

Distribuzioni in classi

Classi	Frequenze assolute	Frequenze relative	Frequenze relative cumulate	Densità di frequenza
21-|22,60	4	0,02	0,02	0,013
22,60-|24,20	72	0,36	0,38	0,225
24,20-|25,80	47	0,24	0,62	0,147
25,80-|27,40	71	0,36	0,97	0,222
27,40-|29,00	6	0,03	1,00	0,019
Totale	200	1,00		

Istogrammi normalizzati di frequenza





Dagli istogrammi normalizzati di frequenza si evince che le due distribuzioni hanno una forma diversa.

Diagrammi a scatola:

Distribuzione del **Voto esame di Statistica a.a. 2001/02**.

$Q_1 = 24; \quad Q_2 = Me = 25; \quad Q_3 = 26;$

$\alpha = Q_1 - 1,5\left(Q_3 - Q_1\right) = 24 - 1,5(26 - 24) = 21$

$\beta = Q_3 + 1,5(Q_3 - Q_1) = 26 + 1,5(26 - 24) = 29$

a = 21 (minore dei valori maggiori o uguali ad α)

b = 28 (maggiore dei valori minori o uguali ad β)

Distribuzione del **Voto esame di Statistica a.a. 2002/03**.

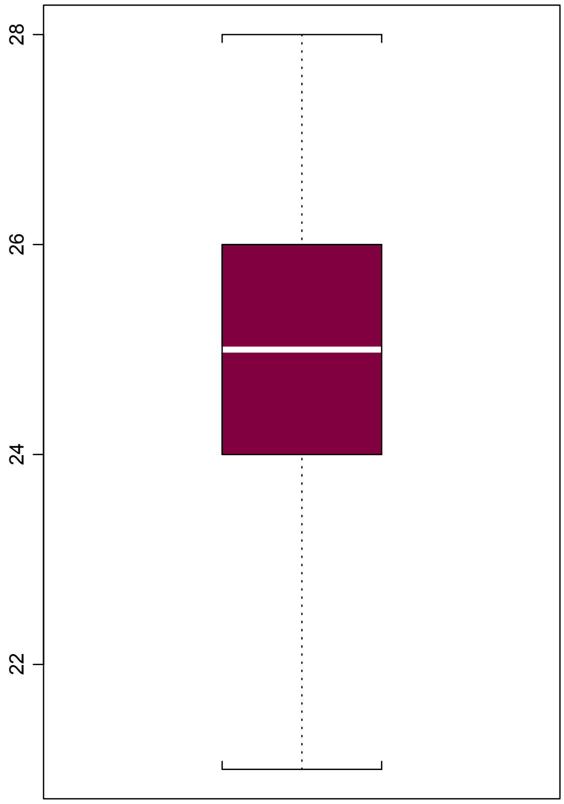
$Q_1 = 24; \quad Q_2 = Me = 25; \quad Q_3 = 26;$

$\alpha = Q_1 - 1,5(Q_3 - Q_1) = 24 - 1,5(26 - 24) = 21$

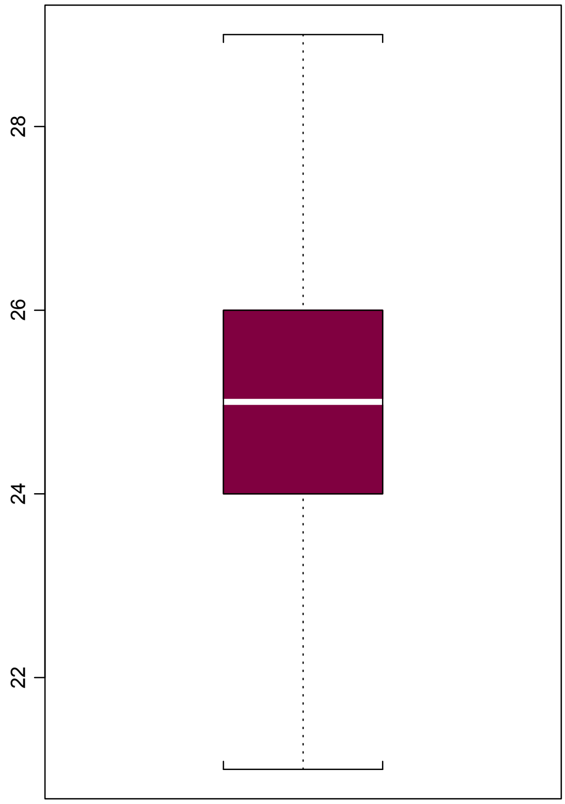
$\beta = Q_3 + 1,5(Q_3 - Q_1) = 26 + 1,5(26 - 24) = 29$

a = 21 (minore dei valori maggiori o uguali ad α)

b = 29 (maggiore dei valori minori o uguali ad β)



Voto esame di statistica a.a. 2001/02



Voto esame di statistica a.a. 2002/03

Le due distribuzioni si differenziano unicamente per il valore del baffo superiore *b* (pari a 28 per la prima distribuzione e 29 per la seconda distribuzione), quindi la forma delle due distribuzioni è diversa.

Indice di Yule e Bowley

Distribuzione del **Voto esame di Statistica a.a. 2001/02**.

$Q_1 = 24; \quad Q_2 = Me = 25; \quad Q_3 = 26;$

$$A_{YB} = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} = \frac{(26-25)-(25-24)}{(26-25)+(25-24)} = 0$$

Distribuzione del **Voto esame di Statistica a.a. 2002/03**.

$Q_1 = 24; \quad Q_2 = Me = 25; \quad Q_3 = 26;$

$$A_{YB} = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)} = \frac{(26-25)-(25-24)}{(26-25)+(25-24)} = 0$$

I valori degli indici di Yule e Bowley indicherebbero che le due distribuzioni sono simmetriche, ma si è già osservato che ciò non corrisponde a realtà (vedi istogrammi normalizzati e boxplot).

Indice di Fisher

Distribuzione del **Voto esame di Statistica a.a. 2001/02**.

Voto esame di Statistica a.a. 2001/02	**Frequenze assolute**	$x_i n_i$	x_i^2	$x_i^2 n_i$	$(x - \mu_X)$	$\left(\frac{x-\mu_X}{\sigma_X}\right)$	$\left(\frac{x-\mu_X}{\sigma_X}\right)^3$	$\left(\frac{x-\mu_X}{\sigma_X}\right)^3 n_i$
21	6	126	441	2646	0,03	0,03	-3,97	-2,68
23	26	598	529	13754	0,13	0,16	-1,97	-1,33
24	39	936	576	22464	0,20	0,36	-0,97	-0,65
25	57	1425	625	35625	0,29	0,64	0,03	0,02
26	44	1144	676	29744	0,22	0,86	1,03	0,69
27	19	513	729	13851	0,10	0,96	2,03	1,37
28	9	252	784	7056	0,05	1,00	3,03	2,04
Totale	200	4994		125140				-46,74

$$\mu_X = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{200}\sum_{i=1}^{7} x_i n_i = \frac{1}{200} 4.994 = 24,97$$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right)^2 = \frac{1}{200}\sum_{i=1}^{7} x_i^2 n_i - \left(\frac{1}{200}\sum_{i=1}^{7} x_i n_i\right)^2 =$$

$$= \frac{1}{200}125.140 - \left(\frac{1}{200}4.994\right)^2 = 2,20$$

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{2,20} = 1,48$$

$$A_F = \frac{1}{n}\sum_{i=1}^{k}\left(\frac{x_i - \mu_X}{\sigma_X}\right)^3 n_i = \frac{1}{200}\sum_{i=1}^{7}\left(\frac{x_i - \mu_X}{\sigma_X}\right)^3 n_i =$$

$$= \frac{1}{200}(-46,74) = -0,233$$

Distribuzione del **Voto esame di Statistica a.a. 2002/03**.

Voto esame di Statistica a.a. 2001/02	**Frequenze assolute**	$x_i n_i$	x_i^2	$x_i^2 n_i$	$(x-\mu_X)$	$\left(\frac{x-\mu_X}{\sigma_X}\right)$	$\left(\frac{x-\mu_X}{\sigma_X}\right)^3$	$\left(\frac{x-\mu_X}{\sigma_X}\right)^3 n_i$
21	4	84	441	1764	-3,98	-2,68	-19,33	-77,30
23	31	713	529	16399	-1,98	-1,34	-2,38	-73,76
24	41	984	576	23616	-0,98	-0,66	-0,29	-11,83
25	47	1175	625	29375	0,02	0,01	0,00	0,00
26	47	1222	676	31772	1,02	0,69	0,33	15,29
27	24	648	729	17496	2,02	1,36	2,53	60,64
28	4	112	784	3136	3,02	2,04	8,44	33,77
29	2	58	841	1682	4,02	2,71	19,91	39,83
Totale	200	4996		125240				-13,37

$$\mu_X = \frac{1}{n}\sum_{i=1}^{k} x_i\, n_i = \frac{1}{200}\sum_{i=1}^{8} x_i\, n_i = \frac{1}{200}4.996 = 24,98$$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right)^2 = \frac{1}{200}\sum_{i=1}^{8} x_i^2 n_i - \left(\frac{1}{200}\sum_{i=1}^{8} x_i n_i\right)^2 =$$

$$= \frac{1}{200}125.240 - \left(\frac{1}{200}4.996\right)^2 = 2,20$$

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{2,20} = 1,48$$

$$A_F = \frac{1}{n}\sum_{i=1}^{k}\left(\frac{x_i - \mu_X}{\sigma_X}\right)^3 n_i = \frac{1}{200}\sum_{i=1}^{8}\left(\frac{x_i - \mu_X}{\sigma_X}\right)^3 n_i =$$

$$= \frac{1}{200}(-13,37) = -0,067$$

Entrambe le distribuzioni sono asimmetriche negative ($A_F<0$). L'asimmetria negativa è più forte nella prima distribuzione (-0,233 < -0,067).