

**ESERCIZIO 1**

Le due distribuzioni seguenti sono relative ad un gruppo di 1570 studenti iscritti al primo anno fuori corso del Corso di Laurea in Economia Aziendale, divisi in matricole pari e dispari e distribuiti secondo il numero di esami sostenuti:

Esami sostenuti	Matricole pari	Matricole dispari
0 - 4	38	12
5 - 9	116	171
10 - 14	168	234
15 - 19	203	300
20 - 24	215	113
<b>Totale</b>	<b>740</b>	<b>830</b>

- 1) precisare il collettivo, il fenomeno e la natura del carattere oggetto di studio
- 2) calcolare il n° medio di esami sostenuti dagli studenti con matricola pari e con matricola dispari
- 3) stabilire quale delle due distribuzioni presenta maggiore variabilità
- 4) studiare la forma delle distribuzioni attraverso il box-plot

**SOLUZIONE**

1) Il collettivo è costituito da tutti gli studenti iscritti al primo anno fuori corso d.el C di L. in Economia Aziendale, divisi però in due gruppi a seconda che la loro matricola sia pari o dispari; il fenomeno studiato è il numero di esami sostenuti alla data di rilevazione, carattere quantitativo di tipo discreto.

2) Per calcolare il numero medio di esami sostenuti dagli studenti dei due diversi collettivi è necessario innanzi tutto calcolare il valore centrale delle diverse classi di voto.

Esami sostenuti	v.c.	Matricole pari	Matricole dispari
0 - 4	2	38	12
5 - 9	7	116	171
10 - 14	12	168	234
15 - 19	17	203	300
20 - 24	22	215	113
<b>Totale</b>		<b>740</b>	<b>830</b>

Calcoliamo la media aritmetica considerando i valori centrali di ciascuna classe e il numero di studenti corrispondente, prima per il collettivo degli studenti con matricola pari e quindi per il collettivo degli studenti con matricola dispari. Il numero medio di esami sostenuti dagli studenti con matricola pari è pari a circa 15, mentre il numero medio di esami sostenuto dagli studenti con matri-

cola dispari è 14. Possiamo anche calcolare il numero medio di esami sostenuto da tutti gli studenti ricordando la proprietà della associatività della media:

$$\bar{x}_a = \frac{\bar{x}_{\text{pari}} \cdot N_{\text{pari}} + \bar{x}_{\text{dispari}} \cdot N_{\text{dispari}}}{N_{\text{pari}} + N_{\text{dispari}}} = \frac{15 \cdot 740 + 14 \cdot 830}{1570} = 14,7 \text{ esami}$$

3) Per poter stabilire quale distribuzione è più variabile (cioè se il carattere numero di esami ha una maggiore variabilità tra gli studenti con matricola pari o tra quelli con matricola dispari), è necessario prima calcolare lo scarto quadratico medio e quindi il coefficiente di variazione.

$$\sigma_{\text{pari}}^2 = \frac{(2-15)^2 38 + (7-15)^2 116 + (12-15)^2 168 + (17-15)^2 203 + (22-15)^2 215}{740} \cong 36 \rightarrow \sigma_{\text{pari}} = 6 \text{ esami}$$

$$\sigma_{\text{dispari}}^2 = \frac{(2-14)^2 12 + (7-14)^2 171 + (12-14)^2 234 + (17-14)^2 300 + (22-14)^2 113}{830} \cong 25 \rightarrow \sigma_{\text{dispari}} = 5 \text{ esami}$$

Abbiamo quindi nel caso delle matricole pari una variabilità di  $\pm 6$  esami (in termini di s.q.m), mentre nel caso delle matricole dispari una variabilità di  $\pm 5$  esami. Possiamo ora calcolare i coefficienti di variazione e confrontare le distribuzioni:

$$CV_{\text{pari}} = \frac{\sigma_{\text{pari}}}{|\bar{x}_{\text{pari}}|} = \frac{6}{15} \cong 0,40 \rightarrow 40\% \quad CV_{\text{dispari}} = \frac{\sigma_{\text{dispari}}}{|\bar{x}_{\text{dispari}}|} = \frac{5}{14} \cong 0,36 \rightarrow 36\%$$

Dal confronto dei due coefficienti di variazione concludiamo che il numero di esami sostenuti dagli studenti con matricola pari ha una maggiore variabilità rispetto al numero di esami sostenuti dagli studenti con matricola dispari.

4) Per poter disegnare il box plot delle due distribuzioni è necessario calcolare mediana, primo e terzo quartile, quindi abbiamo bisogno delle frequenze cumulate relative degli studenti con matricola pari e matricola dispari. Una volta individuate le classi che contengono i valori che ci interessano calcoliamo i diversi indici per approssimazione lineare.

Esami sostenuti	Matricole pari	$f_i$	$F_i$	Matricole dispari	$f_i$	$F_i$
0 - 4	38	0,05	0,05	12	0,01	0,01
5 - 9	116	0,16	0,21	171	0,21	0,22
10 - 14	168	0,23	0,44	234	0,28	0,50
15 - 19	203	0,27	0,71	300	0,36	0,86
20 - 24	215	0,29	1,00	113	0,14	1,00
<b>Totale</b>	<b>740</b>	1,00	-	<b>830</b>	1,00	-

Consideriamo prima il caso degli studenti con matricola pari. Dall'analisi delle frequenze cumulate relative individuiamo nella classe 15 - 19 esami la classe mediana ( $F_4 = 0,71$ ), e quindi possiamo procedere al calcolo della mediana per approssimazione lineare, ottenendo:

$$Me = 15 + \frac{0,5 - 0,44}{0,71 - 0,44} \cdot 4 = 15,89 \text{ esami}$$

Analogamente, per il primo e il terzo quartile otteniamo  $Q_1 = 10,70$  esami e  $Q_3 = 20,55$  esami. In questo caso la Differenza Interquartile è pari a 9,85. Per disegnare il box plot abbiamo bisogno di individuare gli estremi delle code, quindi calcoliamo minimo e massimo teorico della distribuzione:

$$x_{\min}^* = Q_1 - 1,5 \cdot DI = 10,70 - 1,5 \cdot 9,85 = -4,075$$

$$x_{\max}^* = Q_3 + 1,5 \cdot DI = 20,55 + 1,5 \cdot 9,85 = 35,325$$

Poiché il minimo teorico è più piccolo di quello osservato e il massimo teorico è più grande di quello osservato conserviamo il valore minimo e massimo osservati nella distribuzione (cioè 0 e 24).

Dobbiamo ora ripetere gli stessi passaggi per gli studenti con matricola dispari, ottenendo:

$$Me = 14 \quad Q_1 = 10,43 \quad Q_3 = 17,78 \quad DI = 7,35$$

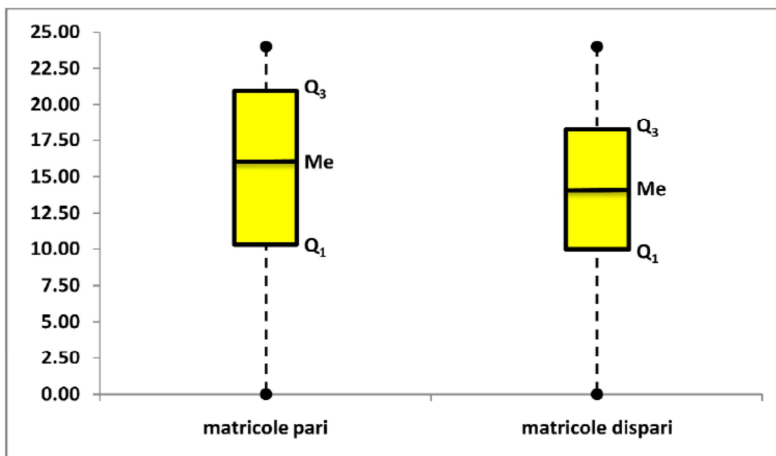
Ancora una volta calcoliamo minimo e massimo teorico della distribuzione:

$$x_{\min}^* = Q_1 - 1,5 \cdot DI = 10,43 - 1,5 \cdot 7,35 = -0,595$$

$$x_{\max}^* = Q_3 + 1,5 \cdot DI = 17,78 + 1,5 \cdot 7,35 = 28,805$$

Come nel caso precedente conserviamo il minimo ed il massimo osservati nella distribuzione.

Possiamo ora disegnare i due box plot, studiare la forma delle distribuzioni e confrontarle:



Si nota come nel caso degli studenti con matricola pari ci sia una più accentuata asimmetria negativa rispetto agli studenti con matricola dispari. Inoltre possiamo osservare come tra gli studenti con matricola pari ci sia una maggior dispersione tra i valori centrali rispetto a quelli rilevati tra gli studenti con matricola dispari.