

ESERCIZIO 1

Nella tabella di seguito sono riportati i dati relativi al tempo necessario a 8 studenti per svolgere un test di valutazione (in ore):

Tempo (in ore)	4,4	3,2	3,6	4,7	5,1	3,2	4,2	2,8
-----------------------	-----	-----	-----	-----	-----	-----	-----	-----

- 1) precisare il collettivo, il fenomeno e la natura del carattere oggetto di studio
- 2) calcolare il tempo medio impiegato dagli studenti per completare il test
- 3) studiare la variabilità del tempo impiegato dagli studenti con un indice opportuno
- 4) data la trasformazione lineare $Z=2X - 3$, calcolare media e varianza del carattere Z

ESERCIZIO 2

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
Totale	740	830

- 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

ESERCIZIO 3

Negli ultimi 20 giorni alle ore 12:00 è stata rilevata la temperatura di una località montana (in °C):

Giorno	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Temperatura (in °C)	0	-3	2	0	-1	-1	2	-2	1	-2	-3	1	-3	2	1	-2	1	2	-3	1

- 1) precisare il collettivo, il fenomeno e la natura del carattere oggetto di studio
- 2) costruire la distribuzione di frequenza
- 3) stabilire in quanti giorni la temperatura si è mantenuta al di sotto dello zero
- 4) determinare la temperatura mediana e studiare la variabilità con un indice opportuno

ESERCIZIO 4

In una indagine sui consumi telefonici è stato chiesto a tutti gli studenti iscritti al primo anno del C. di L. in Economia Aziendale che importo ricaricano sul cellulare in un mese. Nella tabella che segue sono riportati i risultati:

Ricarica (€)	n_i
10	44
25	112
50	176
100	41
150	27
totale	400

- 1) studiare la concentrazione del carattere calcolando l'indice R e utilizzando la Curva di Lorenz
- 2) cosa accade se il prossimo mese ogni studente ricarica il proprio cellulare della metà?

ESERCIZIO 5

È stata analizzata la quantità di ferro (in mg) contenuta in 84 campioni prelevati dal terreno A e in 72 campioni prelevati dal terreno B. I risultati sono riportati nella seguente tabella:

quantità ferro (mg)	TERRENO A		TERRENO B	
	n° campioni	totale ferro (mg)	n° campioni	totale ferro (mg)
0 - 5	15	20	8	40
5 - 15	18	108	8	112
15 - 40	19	304	10	400
40 - 60	12	600	25	1025
60 - 70	6	390	15	975
70 - 100	14	1260	6	426
Totale	84	2682	72	2978

- 1) verificare quale tra le due distribuzioni ha la maggiore variabilità
- 2) verificare in quale terreno il ferro è più concentrato

ESERCIZIO 1

1) Il collettivo oggetto di studio è costituito dagli 8 studenti che hanno svolto il test di valutazione; il fenomeno che stiamo analizzando è il tempo impiegato da ciascun studente per completare il test, carattere quantitativo di tipo continuo.

2) Per calcolare il tempo medio impiegato dagli studenti utilizziamo la media aritmetica:

$$\bar{x}_a = \frac{1}{8}(4,4 + 3,2 + 3,6 + 4,7 + 5,1 + 3,2 + 4,2 + 2,8) = 3,9 \text{ ore}$$

Possiamo quindi concludere che mediamente gli studenti hanno impiegato 3 ore e 54 minuti per completare il test di valutazione ($0,9 \times 60 = 54$).

3) Avendo calcolato il tempo medio, possiamo studiare la variabilità attraverso lo scostamento semplice medio oppure attraverso la varianza:

$$S_{\bar{x}_a} = \frac{[|4,4 - 3,9| + |3,2 - 3,9| + |3,6 - 3,9| + |4,7 - 3,9| + |5,1 - 3,9| + |3,2 - 3,9| + |4,2 - 3,9| + |2,8 - 3,9|]}{8} = 0,7 \text{ ore}$$

$$\sigma^2 = \frac{[(4,4 - 3,9)^2 + (3,2 - 3,9)^2 + (3,6 - 3,9)^2 + (4,7 - 3,9)^2 + (5,1 - 3,9)^2 + (3,2 - 3,9)^2 + (4,2 - 3,9)^2 + (2,8 - 3,9)^2]}{8} = 0,5875$$

Con lo scostamento semplice medio si rileva una variabilità di 42 minuti ($0,7 \times 60$), mentre invece considerando lo scarto quadratico medio (radice quadrata della varianza) si rileva una variabilità di circa 46 minuti ($0,77 \times 60$).

4) Utilizzando le proprietà della media aritmetica e della varianza possiamo facilmente calcolare media e varianza del carattere $Z = 2X - 3$. Infatti avremo che:

$$\bar{z}_a = 2\bar{x}_a - 3 = 2 \cdot 3,9 - 3 = 4,8 \text{ ore}$$

$$\sigma_z^2 = 2^2 \sigma_x^2 = 4 \cdot 0,5875 = 2,35 \rightarrow \sigma_z = \sqrt{2,35} = 1,53 \text{ ore}$$

Dopo la trasformazione avremo quindi un tempo medio pari a 4 ore e 48 minuti, con uno scostamento medio da tale tempo pari a circa ± 1 ora e 32 minuti.

ESERCIZIO 2

1) Il collettivo è costituito da tutti gli studenti iscritti al primo anno fuori corso del CdL in Economia Aziendale, divisi 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
Totale		740	830

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 quello degli studenti con matricola dispari. Il n° medio di esami sostenuti dagli studenti con matricola pari è pari a circa 15, mentre il n° medio di esami sostenuto dagli studenti con matricola dispari è 14. Possiamo anche calcolare il n° medio di esami sostenuto da tutti gli studenti:

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

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 ± 6 esami (in termini di s.q.m), mentre nel caso delle matricole dispari una variabilità di ± 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.

ESERCIZIO 3

- 1) Il collettivo statistico è costituito dai 20 giorni nei quali abbiamo registrato la temperatura; il fenomeno è dato quindi dalla temperatura (in °C), carattere quantitativo di tipo continuo.
- 2) Per costruire la distribuzione di frequenza dobbiamo individuare tutti i diversi modi di manifestarsi del fenomeno, cioè le diverse modalità del carattere temperatura:

Temperatura (in °C)		Giorni
-3	XXXX	4
-2	XXX	3
-1	XX	2
0	XX	2
1	XXXXX	5
2	XXXX	4
		20

- 3) Per stabilire in quanti giorni la temperatura si è mantenuta sotto lo zero è necessario derivare le frequenze cumulate assolute dalle frequenze assolute:

Temperatura (in °C)	Giorni	N_i	F_i
-3	4	4	0,20
-2	3	7	0,35
-1	2	9	0,45
0	2	11	0,55
1	5	16	0,80
2	4	20	1,00
Totale	20	-	

La temperatura si è mantenuta sotto lo zero per 9 giorni (N_3).

- 4) La temperatura mediana è pari a 0 °C, poiché osserviamo dalle frequenze cumulate relative che in corrispondenza di tale valore abbiamo il 55% dei casi, quindi l'unità che bipartisce la distribuzione ha come modalità corrispondente la temperatura 0. Avendo utilizzato la mediana come misura di centralità calcoliamo lo scostamento semplice mediano per valutare la variabilità del fenomeno:

$$S_{Me} = \frac{|0| + |-3| + |2| + |0| + |-1| + |-1| + |2| + |-2| + |1| + |-2| + |-3| + |1| + |-3| + |2| + |1| + |-2| + |1| + |2| + |-3| + |1|}{20} = \frac{33}{20} = 1,65 \text{ °C}$$

ESERCIZIO 4

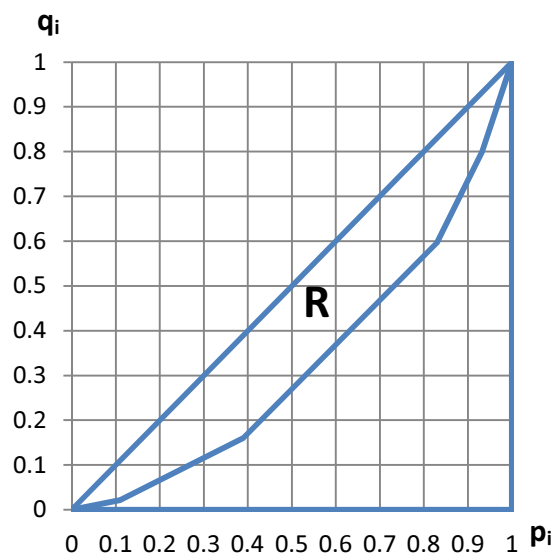
1) Per calcolare la concentrazione abbiamo bisogno di determinare la frazione delle prime i unità statistiche e la frazione di carattere posseduto dalle prime i unità statistiche. La prima quantità (p_i) corrisponde in questo caso alla frequenza cumulata relativa (F_i) poiché stiamo considerando una distribuzione di frequenza; la seconda quantità (q_i) corrisponde invece all'intensità del carattere nelle prime i unità del collettivo, ovvero quanta parte del carattere è posseduta dalle prime i unità del collettivo. Considerando la tabella dei dati avremo quindi:

Ricarica (€)	n_i	$p_i (F_i)$	$A_i (x_i n_i)$	q_i	$p_i - q_i$
10	44	0,11	440	0,02	0,09
25	112	0,39	3240	0,16	0,23
50	176	0,83	12040	0,60	0,23
100	41	0,93	16140	0,80	0,13
150	27	1,00	20190	1,00	0,00
Totale	400	-	-	-	-

Possiamo allora calcolare il rapporto di concentrazione R:

$$R = \frac{\sum_{i=1}^{k-1} (p_i - q_i)}{\sum_{i=1}^{k-1} p_i} = \frac{0,09 + 0,23 + 0,23 + 0,13}{0,11 + 0,39 + 0,83 + 0,93} = 0,3025 \sim 30\%$$

Abbiamo quindi un indice R pari al 30% circa, che indica un basso livello di concentrazione. Possiamo dedurre la stessa informazione disegnando la Curva di Lorenz:



2) Se il prossimo mese ogni studente ricarica sul proprio cellulare la metà di quello che ha ricaricato questo mese allora la concentrazione del carattere non cambia, poiché è invariante per trasformazioni del tipo aX (in questo caso $0,5 \cdot X$).

ESERCIZIO 5

1) E' necessario calcolare il contenuto medio di ferro nei campioni estratti dai terreni A e B:

$$\bar{x}_A = \frac{2,5 \cdot 15 + 10 \cdot 18 + 27,5 \cdot 19 + 50 \cdot 12 + 65 \cdot 6 + 85 \cdot 14}{84} = 34,76 \text{ Mg}$$

$$\bar{x}_B = \frac{2,5 \cdot 8 + 10 \cdot 8 + 27,5 \cdot 10 + 50 \cdot 25 + 65 \cdot 15 + 85 \cdot 6}{72} = 43,19 \text{ Mg}$$

Mediamente i campioni di terreno A presentano un contenuto di 34,76 mg, mentre i campioni di terreno B presentano un contenuto di 43,19 mg. A questo punto è necessario calcolare per ciascun terreno la deviazione standard:

$$\begin{aligned} \sigma_A^2 &= \frac{(2,5^2 \cdot 15) + (10^2 \cdot 18) + (27,5^2 \cdot 19) + (50^2 \cdot 12) + (65^2 \cdot 6) + (85^2 \cdot 14)}{84} - 34,76^2 = \\ &= 848,31 \rightarrow \sigma = \sqrt[2]{848,31} = 29,13 \text{ Mg} \end{aligned}$$

$$\begin{aligned} \sigma_B^2 &= \frac{(2,5^2 \cdot 8) + (10^2 \cdot 8) + (27,5^2 \cdot 10) + (50^2 \cdot 25) + (65^2 \cdot 15) + (85^2 \cdot 6)}{72} - 43,19^2 = \\ &= 601,43 \rightarrow \sigma = \sqrt[2]{601,43} = 24,12 \text{ Mg} \end{aligned}$$

Con la media e la deviazione standard possiamo calcolare i coefficienti di variazione e confrontare la variabilità delle due distribuzioni:

$$CV_A = \frac{29,13}{|34,76|} = 0,84 \sim 84\% \quad CV_B = \frac{24,12}{|43,19|} = 0,57 \sim 57\%$$

Quindi è possibile concludere che il terreno A ha una maggiore variabilità della quantità di ferro rispetto al terreno B.

2) Per calcolare la concentrazione, avendo a disposizione l'ammontare di carattere posseduto da ciascuna classe, assumiamo l'equidistribuzione del carattere in ciascuna classe:

quantità ferro (mg)	n. campioni terreno A	frequenze relative	p_i	totale ferro (mg)	frequenze relative	q_i
0 - 5	15	0,18	0,18	20	0,01	0,01
5 - 15	18	0,21	0,39	108	0,04	0,05
15 - 40	19	0,23	0,62	304	0,11	0,16
40 - 60	12	0,14	0,76	600	0,22	0,38
60 - 70	6	0,07	0,83	390	0,15	0,53
70 - 100	14	0,17	1,00	1260	0,47	1,00
Totale	84	1,00		2682	1,00	

quantità ferro (mg)	n. campioni terreno B	frequenze relative	p_i	totale ferro (mg)	frequenze relative	q_i
0 - 5	8	0,11	0,11	40	0,01	0,01
5 - 15	8	0,11	0,22	112	0,04	0,05
15 - 40	10	0,14	0,36	400	0,13	0,19
40 - 60	25	0,35	0,71	1025	0,34	0,53
60 - 70	15	0,21	0,92	975	0,33	0,86
70 - 100	6	0,08	1,00	426	0,14	1,00
Totale	72	1,00		2978	1,00	

Per calcolare l'indice R applichiamo la formula:

$$R \cong 1 - \sum_{i=0}^{k-1} (q_{i+1} + q_i)(p_{i+1} - p_i)$$

Terreno A
0,001
0,012
0,047
0,078
0,065
0,255
0,459

R
0,541

Terreno B
0,001
0,007
0,033
0,248
0,289
0,155
0,733

R
0,267

Confrontando l'indice R nel primo e nel secondo terreno si deduce che la concentrazione di ferro del terreno A è maggiore (54,1% vs 26,7%).