

Università di Cassino
Corso di Statistica
Esercitazione del 12 febbraio 2010

Simona Balzano

ESERCIZIO 1

In riferimento al Dataset Emissioni atmosferiche – attività economiche (<http://areadocenti.eco.unicas.it/balzano>), per i 3 caratteri ANIDRIDE CARBONICA, MONOSSIDO DI CARBONIO ed EFFETTO SERRA determinare il coefficiente di variazione.

Soluzione

I dati:

Anidride carbonica	Monossido di carbonio	Effetto serra
19,26	56,77	31,71
0,49	1,76	0,68
56,36	102,93	68,71
32,41	68,18	38,85
3,59	12,92	4,84
10,31	26,03	11,94
16,72	26,15	17,83
29,87	57,80	38,36
23,72	68,67	26,86
5,94	16,05	7,23
6,80	19,38	8,42
28,09	56,95	33,04
5,97	22,51	7,39
3,50	8,40	4,09
11,34	60,74	14,99
55,46	433,04	59,69
3,89	12,78	5,01
8,62	74,05	10,41
38,36	84,62	43,34
22,43	48,68	25,78

$$CV = \frac{\sigma}{|\mu|}$$

Essendo $N = 20$ per i 3 caratteri, ed essendo:

$$0 \leq CV \leq \sqrt{n-1}$$

$$0 \leq CV \leq \sqrt{19}$$

$$0 \leq CV \leq 4,36$$

il risultato del coefficiente di variazione per ciascuno dei 3 caratteri va confrontato con l'intervallo $[0; 4,39]$.

Anidride carbonica (AC)

$$\mu_{AC} = \frac{19,26 + 0,49 + 56,36 + \dots + 22,43}{20} = 19,16$$

$$\begin{aligned}\sigma_{AC} &= \sqrt{\frac{\sum_{i=1}^n x_i^2}{N} - \bar{x}^2} \\ &= \sqrt{\frac{19,26^2 + 0,49^2 + 56,36^2 + \dots + 22,43^2}{20} - (19,16)^2} = \\ &= \sqrt{\frac{370,92 + 0,24 + 3176,04 + \dots + 503,28}{20} - 366,93} = \sqrt{631,87 - 366,93} = 16,28\end{aligned}$$

$$CV_{AC} = \frac{16,28}{19,16} = 0,85$$

Monossido di carbonio (MC)

$$\mu_{MC} = \frac{56,77 + 1,76 + 102,93 + \dots + 48,68}{20} = 62,92$$

$$\begin{aligned}\sigma_{MC} &= \sqrt{\frac{\sum_{i=1}^n x_i^2}{N} - \bar{x}^2} \\ &= \sqrt{\frac{56,77^2 + 1,76^2 + 102,93^2 + \dots + 48,68^2}{20} - (62,92)^2} = \\ &= \sqrt{\frac{3222,85 + 3,1 + 10593,56 + \dots + 2370,16}{20} - 3959,05} = \sqrt{11944,7 - 3959,05} = 89,36\end{aligned}$$

$$CV_{MC} = \frac{89,36}{62,92} = 1,42$$

Effetto serra (ES)

$$\mu_{ES} = \frac{31,71 + 0,68 + 68,71 + \dots + 25,78}{20} = 22,96$$

$$\begin{aligned}\sigma_{ES} &= \sqrt{\frac{\sum_{i=1}^n x_i^2}{N} - \bar{x}^2} \\ &= \sqrt{\frac{31,71^2 + 0,68^2 + 68,71^2 + \dots + 25,78^2}{20} - (22,96)^2} = \\ &= \sqrt{\frac{1005,32 + 0,46 + 4721,71 + \dots + 664,49}{20} - 527,08} = \sqrt{883,21 - 527,08} = 18,87\end{aligned}$$

$$CV_{ES} = \frac{18,87}{22,96} = 0,82$$

ESERCIZIO 2

Si consideri la seguente successione di valori di emissioni di anidride carbonica per le 7 regioni in cui l'indicatore di effetto serra registra i più elevati valori:

7 8 9 11 11 12 24

Si determini la differenza semplice media.

Soluzione

La differenza semplice media senza ripetizione si ottiene come:

$$\Delta = \frac{\sum_{i \neq j=1}^N |x_i - x_j|}{N(N-1)}$$

Gli scarti semplici in valore assoluto sono contenuti nella seguente tabella:

x_i	7	8	9	11	11	12	24
7	0	1	2	4	4	5	17
8	1	0	1	3	3	4	16
9	2	1	0	2	2	3	15
11	4	3	2	0	0	1	13
11	4	3	2	0	0	1	13
12	5	4	3	1	1	0	12
24	17	16	15	13	13	12	0

La loro media è:

$$\Delta = \frac{\sum_{i \neq j=1}^N |x_i - x_j|}{N(N-1)} = \frac{244}{7 \times 6} = 5,81.$$

Media:

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{7 + 8 + 9 + 11 + 11 + 12 + 24}{7} = \frac{82}{7} = 11,71$$

$$\Delta_{\text{MAX}} = 2\mu = 2 \times 11,71 = 23,42$$

Indice normalizzato di mutua variabilità $0 \leq R \leq 1$:

$$R = \frac{\Delta}{\Delta_{\text{max}}} = \frac{5,81}{23,42} = 0,25$$

Conclusione: R è più prossimo a 0 che ad 1 \rightarrow le osservazioni si differenziano poco l'una dall'altra.

ESERCIZIO 3

Si determini la differenza semplice media per la seguente distribuzione di frequenza, relativa al carattere X = Effetto serra relativo alle famiglie nelle 20 regioni:

x_i	n_i
1	3
2	4
3	4
7	3
9	2
11	3
24	1
20	

Soluzione

La differenza semplice media per una distribuzione di frequenze è data da:

$$\Delta = \frac{\sum_{i \neq j=1}^K |x_i - x_j| n_i \cdot n_j}{N(N-1)}$$

dove K indica il numero di modalità che il carattere assume.

La tabella riassume le differenze in valore assoluto tra le modalità, moltiplicate per le rispettive frequenze (riportate nella riga e nella colonna di intestazione, accanto alle modalità stesse):

	n _i	3	4	4	3	2	3	1
n _i	ES	1	2	3	7	9	11	24
3	1	0	12	24	54	48	90	69
4	2	12	0	16	60	56	108	88
4	3	24	16	0	48	48	96	84
3	7	54	60	48	0	12	36	51
2	9	48	56	48	12	0	12	30
3	11	90	108	96	36	12	0	39
1	24	69	88	84	51	30	39	0

$$\Delta = \frac{\sum_{i \neq j=1}^K |x_i - x_j| n_i \cdot n_j}{N(N-1)} = \frac{(|1-2| \cdot 3 \cdot 3) + (|1-3| \cdot 3 \cdot 4) + \dots + (|24-11| \cdot 1 \cdot 3)}{20 \cdot 19} = \frac{2162}{380} = 5,69$$

$$\Delta_{MAX} = 2\mu = 2 \times 5,95 = 11,9$$

Indice normalizzato di mutua variabilità $0 \leq R \leq 1$:

$$R = \frac{\Delta}{\Delta_{max}} = \frac{5,69}{11,9} = 0,49$$

Conclusione: R è centrale nell'intervallo $[0; 1]$ → le osservazioni si differenziano in misura media l'una dall'altra.

ESERCIZIO 4

Sulla successione di valori dell'esercizio 2 calcolare il rapporto di concentrazione utilizzando la regola dei trapezi e disegnare il diagramma di Lorenz.

Soluzione

La regola dei trapezi fornisce la seguente espressione per il rapporto di concentrazione del Gini:

$$R = \frac{\frac{1}{2} - \frac{1}{2} \sum_{i=1}^n (q_i + q_{i-1})(p_i - p_{i-1})}{\frac{1}{2}} = 1 - \sum_{i=1}^n (q_i + q_{i-1})(p_i - p_{i-1})$$

che, trattandosi di una successione di valori, può essere calcolato anche mediante la formula alternativa:

$$R = 1 - \frac{2}{n-1} \sum_{i=1}^{n-1} q_i$$

Nella tabella che segue, accanto ai valori del carattere x_i con le frequenze, vi sono le quantità:

$$S_i = \sum_{s=1}^i x_s$$

$$p_i = \sum_{s=1}^i f_s = F_i$$

$$q_i = \sum_{s=1}^i S_s$$

x_i	n_i	f_i	S_i	$p_i = F_i$	q_i
7	1	0,14*	7	0,14	0,09
8	1	0,14	15	0,29	0,18
9	1	0,14	24	0,43	0,29
11	1	0,14	35	0,57	0,43
11	1	0,14	46	0,71	0,56
12	1	0,14	58	0,86	0,71
24	1	0,14	82	1	1
	7	1			

* Il valore della frequenza relativa costante $f_i = \frac{1}{7} = 0,142857...$ è arrotondato nella tabella a 0,14, per questo il totale della colonna f_i non è esattamente 1; le quantità p_i sono state calcolate con i valori non arrotondati delle f_i .

$$R = 1 - \frac{2}{n-1} \sum_{i=1}^{n-1} q_i = 1 - \frac{2}{7-1} \times 2,26 = 0,25$$

Diagramma di Lorenz

