Università di Cassino

Esercitazioni di Statistica 1 del 12 Febbraio 2010

Dott. Mirko Bevilacqua

ESERCIZIO N° 1

Si consideri un campione di 30 sportelli bancari e sia X il numero di operazione effettuate presso uno sportello bancario nell'ultima settimana.

X_i	n_i
[60 – 62]	3
[63 – 65]	13
[66 – 72]	4
[73 – 81]	10

Misurare la mutua variabilità attraverso la differenza semplice media e il grado di concentrazione.

In presenza di caratteri trasferibili (reddito, risorse energetiche, consumo di beni) è di maggiore interesse lo studio della variabilità tra le singole unità statistiche piuttosto che la variabilità rispetto a un centro.
Un carattere quantitativo trasferibile *X* con *n* valori osservati $x_1, x_2, x_3, \ldots, x_n$ si dice equidistribuito se ognuna delle *n* unità statistiche possiede *1/n* dell'ammontare complessivo del carattere

$$A = \sum_{i=1}^{n} x_i{}'$$

ossia per ogni *i* si ha che $X_i = A/n = \mu$.
La situazione di massima concentrazione si ha quando l'intero ammontare del carattere, A, è posseduto da una sola unità del collettivo e cioè:
$X_1 = X_2 = \ldots = X_{n-1} = 0$ e $X_n = A$

DIFFERENZA MEDIA SEMPLICE

Il calcolo si basa sulle differenze tra tutte le coppie di unità statistiche.
In caso di successioni di valori basta applicare la seguente formula:

$$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{n(n-1)} = \frac{2\sum_{i>j}|x_i - x_j|}{n(n-1)} \quad .$$

Per le distribuzioni di frequenza, è di immediata applicazione la formula:

$$\Delta = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}|x_i - x_j| \cdot (n_i \cdot n_j)}{n(n-1)} = \frac{2\sum_{i>j}^{k}|x_i - x_j| \cdot (n_i \cdot n_j)}{n(n-1)}$$

DIFFERENZA MEDIA SEMPLICE:
1° Metodo

X_i	C_i	n_i	$\lvert C_i-C_j \rvert$	$n_i \cdot n_j$	$\lvert C_i-C_j \rvert \cdot n_i \cdot n_j$
[60 – 62]	61	3	3	39	117
[63 – 65]	64	13	8	12	96
[66 – 72]	69	4	16	30	480
[73 – 81]	77	10	3	39	117
		n=30	5	52	260
			13	130	1690
			8	12	96
			5	52	260
			8	40	320
			16	30	480
			3	130	1690
			8	40	320
					$\sum_{i=1}^{k}\sum_{j=1}^{k}\lvert C_i - C_j\rvert \cdot (n_i \cdot n_j) = 5926$

$$\Delta = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}\lvert x_i - x_j\rvert \cdot (n_i \cdot n_j)}{n(n-1)} = \frac{5926}{30(30-1)} = 6{,}81$$

DIFFERENZA MEDIA SEMPLICE
2° Metodo

$$\begin{array}{ccc|cccc|}
 & & n_J & 3 & 13 & 4 & 10 \\
n_i & X_i & X_j & 61 & 64 & 69 & 77 \\
\hline
3 & 61 & & 0 & 117 & 96 & 480 \\
13 & 64 & & 117 & 0 & 260 & 1690 \\
4 & 69 & & 96 & 269 & 0 & 320 \\
10 & 77 & & 480 & 1690 & 320 & 0 \\
\hline
\end{array}$$

$$\Delta = \frac{2\sum_{i>j}\lvert x_i - x_j\rvert (n_i \cdot n_j)}{n(n-1)} = \frac{2(117+260+96+480+1690+320)}{30 \cdot 29} = \frac{5926}{870} = 6{,}81$$

$\mu = 68{,}7$ $\qquad\qquad$ $\Delta_{MAX} = 2\mu = 2 \times 68{,}7 = 137$

Differenza media semplice normalizzata $0 \le R \le 1$:

$$R = \frac{\Delta}{\Delta_{max}} = \frac{6{,}81}{137} = 0{,}05$$

ESERCIZIO N° 2

Nella tabella osserviamo il peso di 10 persone in relazione all'attività sportiva praticata.

Sport	*Atletica*	*Calcio*	*Calcio*	*Calcio*	*Calcio*	*Nuoto*	*Nuoto*	*Nuoto*	*Nuoto*	*Nuoto*
Peso	62	77	71	69	65	67	88	78	72	75

Verificare la proprietà di scomponibilità della varianza.

Le modalità del carattere Sport definiscono 3 gruppi, di seguito riportati con i rispettivi valori del carattere peso:

Sport	*Atletica*	*Calcio*	*Calcio*	*Calcio*	*Calcio*	*Nuoto*	*Nuoto*	*Nuoto*	*Nuoto*	*Nuoto*
Peso	62	77	71	69	65	67	88	78	72	75

MEDIA E VARIANZA TOTALE:

$$\mu = \frac{\sum_{i=1}^{10} x_i}{10} = 72,4 \qquad \sigma_X^2 = \frac{\sum_{i=1}^{10} (x_i - \mu)^2}{10} = \frac{508,4}{10} = 50,84$$

MEDIE DEI GRUPPI:

$$\mu_{ATL} = 62 \qquad \mu_{CAL} = \frac{\sum_{i=1}^{n_{CAL}} x_{CAL_i}}{n_{CAL}} = \frac{282}{4} = 70,5 \qquad \mu_{NUO} = \frac{\sum_{i=1}^{n_{NUO}} x_{NUO_i}}{n_{NUO}} = \frac{380}{5} = 86$$

VARIANZA <u>TRA I GRUPPI</u> O <u>ESTERNA</u> (σ_{EXT}^2):

$$\sigma_{EXT}^2 = \frac{\sum_{i=1}^{G} (\mu_i - \mu)^2 n_i}{n} =$$

$$\frac{1}{10}\left[(62-72,4)^2 \times 1 + (70,5-72,4)^2 \times 4 + (76-72,4)^2 \times 5\right] = \frac{187,4}{10} = 18,74$$

VARIANZA <u>ENTRO I GRUPPI</u> O <u>INTERNA</u>, (σ^2_{INT}):

$$\sigma^2_{INT} = \frac{\sum_{i=1}^{G}\sum_{j=1}^{n_i}\left(x_{ij}-\mu_i\right)^2}{20} =$$

$$= \frac{1}{10}\left\{\left[(62-62)^2\right]+\left[(77-70,5)^2+....+(65-70,5)^2\right]+\left[(67-76)^2+.....+(75-76)^2\right]\right\} =$$

$$= \frac{321}{10} = 32,1$$

Verifica della proprietà della scomponibilità:

$$\sigma^2_{EXT} + \sigma^2_{INT} = \sigma^2_X$$

$$\sigma^2_X = 18,74 + 32,1$$

ESERCIZIO N° 3

I salari mensili per i 5 dipendenti dell'azienda informatica "APIWEB" sono i seguenti:

{1150; 1250; 1500; 1750; 2500}

a) determinare i rapporti di concentrazione.
b) costruire la curva di Lorenz;

X_i	n_i	f_i	$F_i = P_i$	A	q_i	$p_i - q_i$
1150	1	0,2	0,2	1150	(1150/8150)= 0,14	0,06
1250	1	0,2	0,4	2400	(2400/8150)= 0,29	0,11
1500	1	0,2	0,6	3900	(3900/8150)= 0,48	0,12
1750	1	0,2	0,8	5650	(5650/8150)= 0,69	0,11
2500	1	0,2	1	8150	(8150/8150)= 1	$\sum_{i=1}^{n-1}(p_i - q_i) = 0,39$

$$\mu = \frac{\sum_{i=1}^{5} x_i}{5} = 1630$$

DIFFERENZA MEDIA SEMPLICE NORMALIZZATA:

		n_J	1	1	1	1	1
n_i	X_i	X_j	1150	1250	1500	1750	2500
1	1150		0				
1	1250		100	0			
1	1500		350	250	0		
1	1750		600	500	250	0	
1	2500		1350	1250	1000	750	0

$$\Delta = \frac{2\sum_{i>j}|x_i - x_j|(n_i \cdot n_j)}{n(n-1)} = \frac{2(100+350+250+600+500+250+1350+1250+1000+750)}{5 \cdot 4} =$$

$$= \frac{2 \cdot 6400}{20} = 640$$

$$\Delta_{MAX} = 2\mu = 2 \cdot 1630 = 3260$$

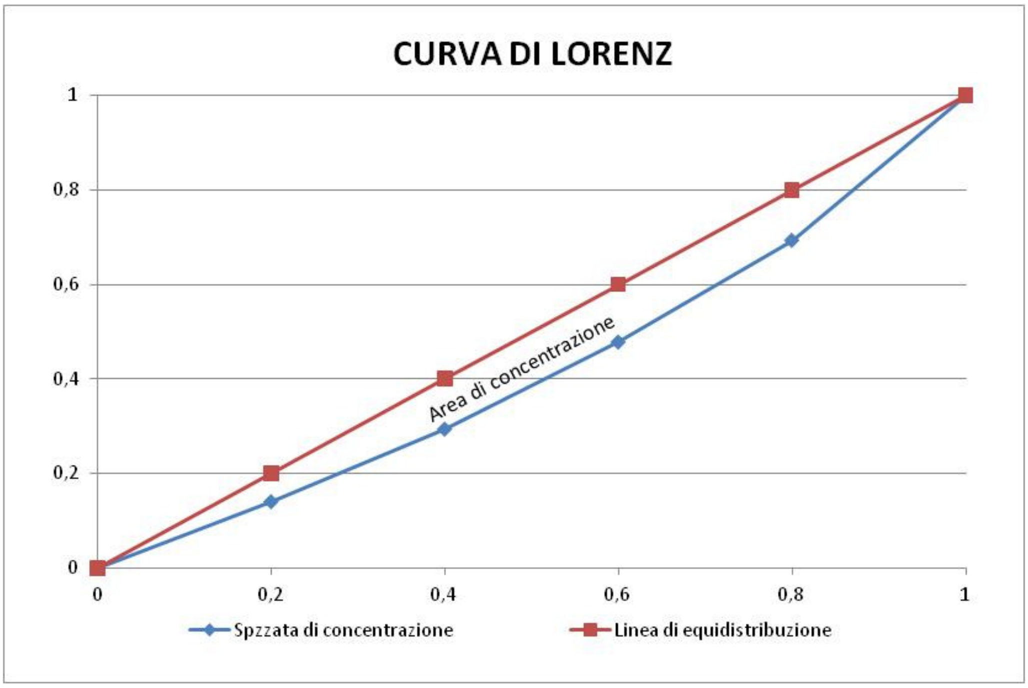
$$R = \frac{\Delta}{\Delta_{max}} = \frac{640}{3260} = 0,19$$

RAPPORTO DI CONCENTRAZIONE DI GINI

Date le distribuzioni delle p_i e delle q_i relative alla distribuzione di un carattere quantitativo trasferibile X osservato su *n* unità con valori ordinati $x_1, x_2, \ldots, x_n$ (con $X_i <= X_{i+1}$) si definisce rapporto di concentrazione di Gini l'indice:

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

Mediante le coppie di valori di p_i e q_i è possibile realizzare un interessante grafico. Consideriamo un piano cartesiano in cui l'asse delle ascisse rappresenti i valori di p_i e l'asse delle ordinate i valori di q_i. In questa maniera ogni coppia di valori è rappresentata da un punto sul piano. I punti limitrofi possono poi essere congiunti da segmenti tali da formare una curva detta **spezzata di concentrazione** o **curva di Lorenz**, dal nome del primo autore che ne propose l'impiego.
Nel grafico oltre alla spezzata di concentrazione viene rappresentata **la linea di equidistribuzione** che è il segmeno che congiunge i punti (0,0) e (1,1). Ogni punto situato su tale segmento ha la proprietà di avere le coordinate uguali cioè $p_i = q_i$. così se l'ammontare del carattere fosse equidistribuito fra tutte le unità del collettivo, i punti corrispondenti giacerebbero sulla linea di equidistribuzione.
Considerando la distribuzione dei valori di p_i e q_i dell'esercizio possiamo otteniamo il seguente grafico:



$$R = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i} = \frac{0{,}39}{2} = 0{,}19$$

Notiamo che non a caso la curva di Lorenz giace sotto la linea di equidistribuzione, poiché come abbiamo visto p_i è sempre maggiore o uguale a q_i.
L'area della superficie compresa tra la curva di Lorenz e la linea di equidistribuzione viene detta **area di concentrazione**.