Esercitazione 4 del corso di Statistica (parte 1)

Dott.ssa Paola Costantini

21 Febbraio 2012

Esercizio n 1

Data la distribuzione del carattere Reddito di cui all'esercizio precedente se ne misuri il grado di concentrazione.

La concentrazione di un carattere quantitativo è possibile solo se il carattere è trasferibile, cioè quando il carattere può passare da un'unità all'altra del collettivo. Il carattere Reddito è un carattere trasferibile e si parte dalla successione dei valori ordinati in senso non decrescente.

xi	Reddito
1	2500
2	2500
3	2800
4	3000
5	3000
6	3000
7	3000
8	3100
9	3500
1	3500
11	4000
12	4000

La concentrazione di un carattere si misura rispetto ad una condizione detta di equidistribuzione.
Si ha **concentrazione nulla** quando l'ammontare totale del carattere è ripartito in parti uguali tra le unità.
Si ha **concentrazione massima** quando tutto il carattere è posseduto da una sola unità, mentre (n-1) unità non possiedono nulla.

$p_i =$ la frazione cumulata dei primi i redditieri, $i = 1,2,3,\ldots n$ $\qquad p_i = \frac{i}{n}$

$q_i =$ la frazione cumulata del reddito posseduto dai primi i redditieri, $i = 1,2,3,\ldots n$ $\qquad p_i = \frac{i}{n}$

$$q_i = \frac{\sum_{j=1}^{i} x_j}{\sum_{j=1}^{n} x_j}$$

x_i	Reddito	Reddito Cumulato	p_i	q_i	$p_i - q_i$
1	2500	2500	0,083	0,066	0,017
2	2500	5000	0,167	0,132	0,035
3	2800	7800	0,25	0,206	0,044
4	3000	10800	0,33	0,285	0,045
5	3000	13800	0,42	0,364	0,056
6	3000	16800	0,5	0,443	0,057
7	3000	19800	0,58	0,522	0,058
8	3100	22900	0,67	0,604	0,066
9	3500	26400	0,75	0,696	0,054
10	3500	29900	0,83	0,789	0,041
11	4000	33900	0,92	0,89	0,03
12	4000	37900			
					0,497

Possiamo dire che c'è equidistribuzione quando $q_i = p_i$ per ogni $i = 1,2,3,\ldots n$

Un indice che misura la concentrazione è il Rapporto di concentrazione (R) di Gini. Si tratta di un indice relativo che varia tra 0 e 1.

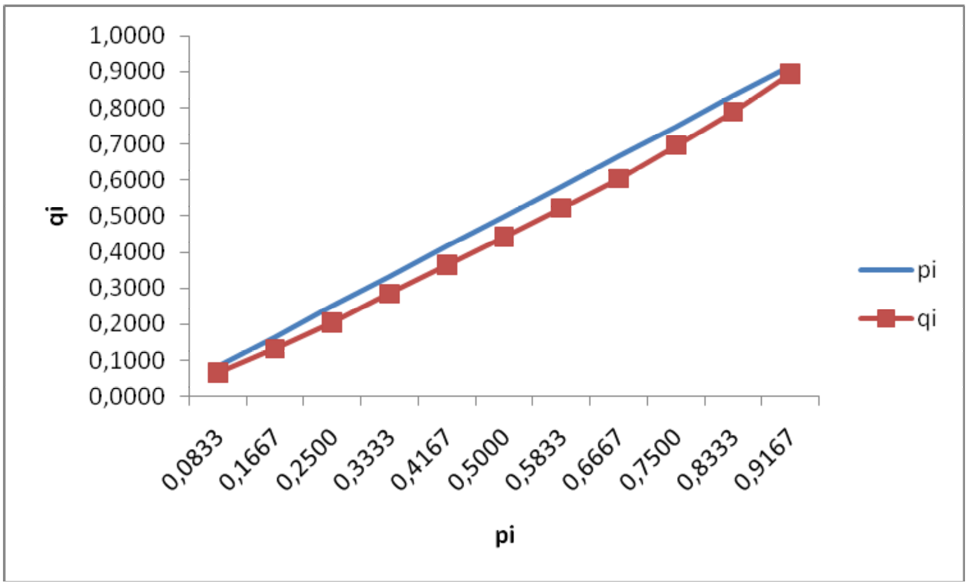
$$R = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

$$0 \leq R \leq 1$$

$$R = \frac{0,497}{5,5} = 0,09$$

R=0 si ha concentrazione minima.
R=1 si ha concentrazione massima.

Una rappresentazione grafica della concentrazione può essere fatta attraverso la curva di Lorenz (curva di concentrazione), ovvero la spezzata che si ottiene unendo i punti di coordinate (p_i, q_i) rappresentati sul piano cartesiano.



La bisettrice rappresentata sul grafico rappresenta la situazione di equidistribuzione. L'area compresa tra la curva di concentrazione e la retta di equidistribuzione viene detta **area di concentrazione**,

In alcune circostanze si pone un maggior interesse sullo studio della variabilità tra le singole unità statistiche, piuttosto che lo studio della variabilità rispetto ad un centro. Lo studio della mutua variabilità è possibile attraverso il calcolo dell'indice Delta (Δ) che rappresenta la media delle differenze, in valore assoluto, di tutte le possibili coppie senza ripetizioni.

Esercizio n 2
Data la distribuzione del carattere Reddito dei soli maschi, espresso in migliaia di euro, se ne misuri il grado di concentrazione.

xi	Sesso	Reddito
1	M	2,5
2	M	2,8
3	M	3,0
4	M	3,0
5	M	3,1
6	M	3,5
7	M	4,0

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

Si dispongono i valori ordinati in una tabella doppia sia in riga che in colonna:

	2,5	2,8	3	3	3,1	3,5	4
2,5							
2,8							
3							
3							
3,1							
3,5							
4							

In modo da calcolare facilmente le differenze di tutte le possibili coppie

	2,5	2,8	3	3	3,1	3,5	4
2,5	0	0,3	0,5	0,5	0,6	1	1,5
2,8	0,3	0	0,2	0,2	0,3	0,7	1,2
3	0,5	0,2	0	0	0,1	0,5	1
3	0,5	0,2	0	0	0,1	0,5	1
3,1	0,6	0,3	0,1	0,1	0	0,4	0,9
3,5	1	0,7	0,5	0,5	0,4	0	0,5
4	1,5	1,2	1	1	0,9	0,5	0

Effettuando la somma degli elementi di ogni colonna (marginale di colonna) si ha:

4,4	2,9	2,3	2,3	2,4	3,6	6,1

Effettuando la somma dei marginali di colonna si ottiene il valore 24 che corrisponde al numeratore dell'indice Delta Δ. Adesso disponiamo di tutti i dati per poter calcolare l'indice.

$$\Delta = \frac{\sum_{i \neq j = 1}^{n} \left| x_i - x_j \right|}{n(n-1)} = \frac{24}{42} = 0{,}57$$

Quando tutte le modalità coincidono, l'indice Delta assume valore zero (Δ = 0).In presenza di caratteri trasferibili è possibile calcolare la variabilità massima e quindi il valore massimo di Delta .
L'indice Delta assume il valore massimo quando tutte le modalità tranne una sono nulle (Δ = 2μ).
Il Rapporto di Concentrazione (R) di Gini si ottiene dividendo Delta per il suo valore massimo.

$$R = \frac{\Delta}{2\mu}$$

Si tratta di un indice relativo che varia tra 0 ed 1. $0 \le R \le 1$

R=0 si ha concentrazione minima.
R=1 si ha concentrazione massima.

Nel nostro esercizio la media del reddito dei maschi è pari a

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{24}{7} = 3,42$$

Quindi il massimo valore che può assumere Delta sarà:

$$2\mu = 3,42 \cdot 2 = 6,86$$

Per cui il rapporto Concentrazione di Gini (R) sarà pari a

$$R = \frac{\Delta}{2\mu} = \frac{0,57}{6,86} = 0,083$$

Esercizio 3

A partire dalla distribuzione della variabile Età, costruire il boxplot

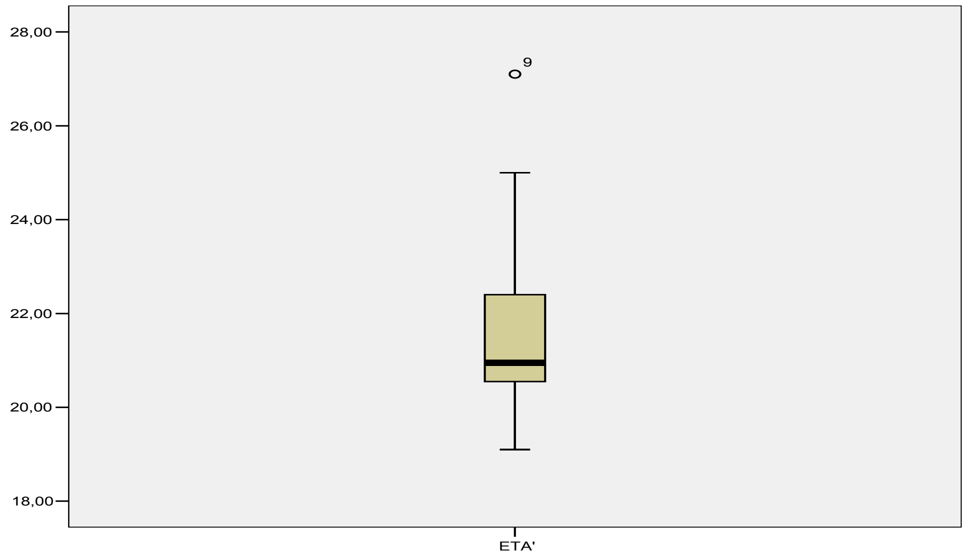
ETA'	ni	fi	Ni	Fi
19,1	1	0,05	1	0,05
19,2	1	0,05	2	0,1
20,2	1	0,05	3	0,15
20,3	1	0,05	4	0,2
20,5	1	0,05	5	0,25
20,6	2	0,1	7	0,35
20,8	2	0,1	9	0,45
20,9	1	0,05	10	0,5
21	2	0,1	12	0,6
21,4	1	0,05	13	0,65
21,8	1	0,05	14	0,7
22,1	1	0,05	15	0,75
22,7	1	0,05	16	0,8
23,3	1	0,05	17	0,85
23,9	1	0,05	18	0,9
25	1	0,05	19	0,95
27,1	1	0,05	20	1
totale	20	1		

Q1 = 20,5 Me= 20,9 Q3= 22,1

a = Q1 – 1,5 (Q3 –Q1) = 20,5 -1,5 (22,1-20,5) = 18
b = Q3 + 1,5 (Q3 –Q1)= 22,1 +1,5 (22,1-20,5) = 24,5

α = min = 19,1

β = max = 27,1



Esercizio n 4

Calcolare gli indici di forma per il carattere Età suddiviso in 3 classi equifrequenti.

a) L'indice di Fisher, è un indice di forma basato sui momenti terzi standardizzati:

$$\gamma = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^3 n_i \qquad \begin{array}{l} \gamma > 0 \rightarrow AsimmetricaPositiva; \\ \gamma = 0 \rightarrow Simmetrica; \\ \gamma < 0 \rightarrow AsimmetricaNegativa; \end{array}$$

Partendo dalla distribuzione in classi del carattere Età,

Ci	ni	fi	Ni	Fi	$\hat{c}_i$
C1 = [19,1; 20,6]	7	0,35	7	0,35	19,85
C2 =] 20,6; 21,8]	7	0,35	14	0,70	21,2
C3 =] 21,8; 27,1]	6	0,30	20	1	24,45
Totali	20	1,00			

Calcoliamo dapprima la media aritmetica e la mediana

Media = $\mu_X = \frac{\sum_{i=1}^{k} \hat{x}_i \cdot n_i}{n}$

$$\bar{x} = \frac{\sum_{i=1}^{k}(19{,}85 \times 7 + 21{,}2 \times 7 + 24{,}45 \times 6)}{20} = 21{,}7$$

Mediana $\boxed{Me \cong x_{Me-1} + (x_{Me} - x_{Me-1}) \dfrac{0,5 - F_{Me-1}}{F_{Me} - F_{Me-1}}}$

$$Me = 20,6 + (21,8 - 20,6) \cdot \frac{0,5 - 0,35}{0,70 - 0,35} = 21,11$$

$$Q_1 = 19,1 + (20,6 - 19,1) \cdot \frac{0,25 - 0}{0,35 - 0} = 20,17$$

$$Q_3 = 21,8 + (27,1 - 21,8) \cdot \frac{0,75 - 0,70}{1 - 0,70} = 22,68$$

Poi la **varianza**

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{c}_i - \bar{x}\right)^2 * n_i = \frac{(19,85 - 21,7)^2 \cdot 7 + (21,2 - 21,7)^2 \cdot 7 + (24,45 - 21,7)^2 \cdot 6}{20} = 3,5$$

$$\sigma = \sqrt{\sigma^2} = 1,9$$

A questo punto abbiamo tutti gli elementi utili per calcolare l'indice di asimmetria di Fisher.

xi	ni	$\hat{c}_i$	$\hat{c}_i - \bar{x}$	$Z_i = \frac{(\hat{c}_i - \bar{x})}{\sigma}$	$Z_i = \left(\frac{(\hat{c}_i - \bar{x})}{\sigma}\right)^3$	$(Z_i) \cdot n_i$
C1 = [19,1;20,6]	7	19,85	-1,85	-0,97	-0,91	-6,37
C2 =] 20,6; 21,8]	7	21,2	-0,5	-0,26	-0,0175	-0,1225
C3 =] 21,8; 27,1]	6	24,45	2,75	1,45	3,05	18,3
Totali	20					11,8

$$\gamma = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{c}_i - \bar{x}}{\sigma}\right)^3 \cdot n_i = \frac{11,8}{20} = 0,59$$

Possiamo concludere che la distribuzione è caratterizzata da un'asimmetria positiva (indice maggiore di zero).

Tale risultato è confermato dal **confronto tra la mediana e la media aritmetica**.

$$\bar{x} = 21,7 > Me = 21,11$$

b) **Hotelling-Solomon** =

Hotelling-Solomon = $\frac{21{,}7 - 21{,}11}{1{,}9} = 0{,}31$ = **Asimmetria positiva**

Esercizio 4

Considerando una distribuzione di frequenza relativa alla distribuzione della variabile QUADRATURA ABITAZIONE in 5 classi come di seguito riportata calcolare la differenza media semplice.

Quadratura abitazione	Frequenze assolute
[85, 95[	4
[95, 105[	16
[105, 115[	13
[115, 125[	4
[125, 135[	3
	40

Per una distribuzione di frequenza la differenza media semplice o indice di mutua variabilità si calcola come segue:

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

Si chiama differenza media assoluta la media aritmetica dei valori assoluti delle differenze. Tale differenza si dice con ripetizione se si considerano tutte le n^2 differenze (cioè comprese le nulle); senza ripetizione se si considerano solo le n(n −1) differenze ottenute escludendo i termini nulli della diagonale principale.

	4	16	13	4	3
	90	100	110	120	130
90	0	640	1040	480	480
100	640	0	2080	1280	1440
110	1040	2080	0	520	780
120	480	1280	520	0	120
130	480	1440	780	120	0

$$\Delta = \frac{\sum_{i \neq j=1}^{n} |x_i - x_j| n_i \cdot n_j}{n(n-1)} = \frac{17720}{40(40-1)} = 11{,}3$$