Esercitazione 4 del corso di Statistica (parte 1)

Dott.ssa Paola Costantini

25 Ottobre 2011

- Gli indici di variabilità analizzati sono tutti non robusti, ovvero risentono della presenza di valori anomali
- Essi sono infatti basati sul calcolo di una media aritmetica di una qualche funzione degli scostamenti dalla media o dalla mediana
- Un indice di variabilità robusto è il MAD (Median Absolute Deviation), ovvero lo scostamento mediano dalla mediana:

$$MAD = Me|x_i - Me_x|$$

Esercizio n 1

Di seguito abbiamo la distribuzione dei redditi di 12 individui e gli scostamenti semplice medio e mediano.

	x_i	$x_i - Me$	$\lvert x_i - Me \rvert$
	2500	-500	500
	4000	1000	1000
	3500	500	500
	3000	0	0
	3100	100	100
	3000	0	0
	4000	1000	1000
	2500	-500	500
	3500	500	500
	3000	0	0
	2800	-200	200
	3000	0	0
somma	**37900**	1900	4300

Dati ordinati

2500 2500 2800 3000 3000 3000 3000 3100 3500 3500 4000 4000

Mediana = 3000

Scostamenti [redacted] ordinati

0	0	0	0	100	200	500	500	500	500	1000	1000
1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°

$$MAD = Me\left|x_i - Me_X\right| = \frac{\left|x_i - Me_X\right|_{(6)} + \left|x_i - Me_X\right|_{(7)}}{2} = \frac{200+500}{2} = 350$$

Ricordiamo che per questi dati

$$S_\mu = \frac{\sum_{i=1}^{k}\left|x_i - \bar{x}\right| \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{4733,33}{12} = 394,44 \qquad S_{Me} = \frac{\sum_{i=1}^{k}\left|x_i - Me\right| \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{4300}{12} = 358,33$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = \frac{2749166.67}{12} = 229097.2 \qquad \rightarrow \sigma = 478,641$$

Esercizio n 2

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.

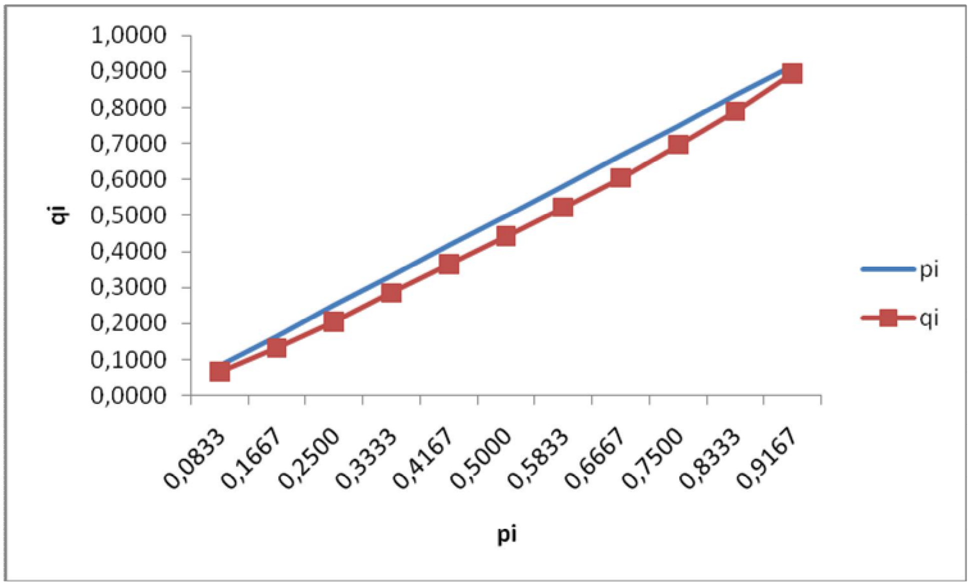
$$R = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

$0 \le R \le 1$

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

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**,

Esercizio n 3

Data la seguente tabella

Settore Reddito	Agricoltura	Industria	Altre attività	Totale
Fino a 15	50	116	160	326
15 —⊣ 30	90	140	241	471
30 —⊣ 45	20	200	260	480
45 —⊣ 65	1	280	200	481
Totale	161	736	861	1758

Determinare la varianza per ciascuna distribuzione parziale e successivamente verificare la scomposizione della varianza.

Calcoliamo le varianze delle distribuzioni parziali utilizzando il metodo indiretto:

$$\sigma_j^2 = \frac{1}{n_{.j}} \sum_{i=1}^{4} y_i^2 \cdot n_{ij} - \bar{y}_j^2 \quad j = 1,2,3.$$

Reddito (valori centrali)	Agricoltura		Industria		Altre attività	
y_i	n_{i1}	$y_i^2 \cdot n_{i1}$	n_{i2}	$y_i^2 \cdot n_{i2}$	n_{i3}	$y_i^2 \cdot n_{i3}$
10	50	5000	116	11600	160	16000
22,5	90	45562,5	140	70875	241	122006,25
37,5	20	28125	200	281250	260	365625
55	1	3025	280	847000	200	605000
Totale	161	81712,5	736	1210725	861	1108631,25

Le medie parziali del reddito annuo risultano essere:

$$\bar{y}_1 = \frac{3330}{161} = 20{,}68 \quad \bar{y}_2 = \frac{27210}{736} = 36{,}97 \quad \bar{y}_3 = \frac{27210}{736} = \frac{27772{,}5}{861} = 32{,}26$$

Le varianze parziali risultano essere:

$$\sigma_1^2 = \frac{81712{,}5}{161} - 20{,}68^2 = 79{,}74$$

$$\sigma_2^2 = \frac{1210725}{736} - 36{,}97^2 = 278{,}22$$

$$\sigma_3^2 = \frac{1108631{,}25}{861} - 32{,}26^2 = 247{,}15$$

La seguente tabella riporta i conti che facilitano il calcolo della varianza dell'intera popolazione:

y_i	$n_{i.}$	$y_i \cdot n_{i.}$	$y_i^2 \cdot n_{i.}$
10	326	3260	32600
22,5	471	10597,5	238443,75
37,5	480	18000	675000
55	481	26455	1455025
totale	1758	58312,5	2401068,75

Il reddito medio aritmetico per l'intera popolazione risulta essere:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{4} y_i \cdot n_i = \frac{58312{,}5}{1758} = 33{,}17$$

La varianza della popolazione totale (calcolata con il metodo indiretto) risulta essere:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{4} y_i^2 \cdot n_{i..} - \overline{y}^2 = \frac{2401068{,}75}{1758} - 33{,}17^2 = 256{,}56$$

Al fine di verificare la scomposizione della varianza, calcoliamo la varianza *interna* ai gruppi:

$$\sigma^2_{INT} = \frac{1}{N}\sum_{j=1}^{3} \sigma_j^2 \cdot n._j = \frac{1}{1758}\left[(79{,}74 \cdot 161) + (278{,}22 \cdot 736) + (247{,}15 \cdot 861)\right] = 244{,}83$$

La varianza *esterna* risulta essere:

$$\sigma^2_{EXT} = \frac{1}{N}\sum_{j=1}^{3} (\overline{y}_j - \overline{y})^2 \cdot n. = \frac{1}{1758} \cdot$$

$$\cdot \left[(20{,}68 - 33{,}17) \cdot 161 + (278{,}22 - 33{,}17) \cdot 736 + (247{,}15 - 33{,}17) \cdot 861\right] =$$

$$= \frac{36456{,}85}{1758} = 20{,}74$$

La varianza totale σ^2_{TOT} risulta essere pari a $\sigma^2_{INT} + \sigma^2_{EXT}$ =244,83+20,74=265,57

che coincide con quanto ricavato in precedenza.