

# Esercitazione 5 del corso di Statistica (parte 1)

*Dott.ssa Paola Costantini*

18 Febbraio 2011

## Dataset Studenti

<i>N</i>	<i>SESSO</i>	<i>ETA'</i>	<i>PESO</i>	<i>ALTEZZA</i>	<i>DIPLOMAI</i>	<i>COMPONENTI</i>	<i>OCCHIALI</i>	<i>FUMO</i>	<i>REDDITO</i>
1	0	20,6	65	180	Ist.Tecnico	6	0	1	0,7
2	0	20,2	75	180	Liceo	4	0	0	0,2
3	0	20,3	60	173	Ist.Tecnico	4	1	0	1,6
4	0	23,9	93	187	Liceo	8	0	1	2,5
5	0	21,4	66	164	Ist.Tecnico	5	0	0	3,2
6	0	25	84	186	Ist.Tecnico	4	0	0	0,1
7	0	20,8	67	175	Altro dipl.	4	0	1	3,8
8	0	20,6	89	170	Liceo	3	1	0	1,3
9	0	27,1	71	180	Liceo	1	0	1	1,2
10	0	23,3	63	170	Liceo	4	0	0	1,7
11	1	20,5	51	161	Ist.Tecnico	4	0	1	1,9
12	1	19,1	58	167	Ist.Tecnico	5	1	1	0,8
13	1	22,1	67	165	Altro dipl.	5	1	1	0,4
14	1	21,8	51	156	Ist.Tecnico	4	0	0	1,8
15	1	19,2	60	170	Ist.Tecnico	5	1	1	1,9
16	1	20,8	55	165	Liceo	4	1	1	3,2
17	1	21	55	158	Liceo	5	1	0	2,1
18	1	20,9	58	170	Liceo	5	1	1	0,1
19	1	22,7	76	170	Liceo	6	1	0	1,6
20	1	21	55	165	Liceo	7	0	0	2,2

## Esercizio n 1

Data la distribuzione del carattere Reddito 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.

## DATI ORDINATI

<i>N</i>	<i>REDDITO</i>
1	0,1
2	0,1
3	0,2
4	0,4
5	0,7
6	0,8
7	1,2
8	1,3
9	1,6
10	1,6
11	1,7

12	1,8
13	1,9
14	1,9
15	1,9
16	3,2
17	2,1
18	0,1
19	1,6
20	2,2

La concentrazione di un carattere si misura rispetto ad una condizione detta diequidistribuzione.

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.

REDDITO	REDDITO CUMULATO	Pi	Qi	Pi-Qi
0,1	0,1	0,05	0,003	0,047
0,1	0,2	0,1	0,006	0,094
0,2	0,4	0,15	0,012	0,138
0,4	0,8	0,2	0,025	0,175
0,7	1,5	0,25	0,046	0,204
0,8	2,3	0,3	0,071	0,229
1,2	3,5	0,35	0,108	0,242
1,3	4,8	0,4	0,149	0,251
1,6	6,4	0,45	0,198	0,252
1,6	8	0,5	0,248	0,252
1,7	9,7	0,55	0,3	0,25
1,8	11,5	0,6	0,356	0,244
1,9	13,4	0,65	0,415	0,235
1,9	15,3	0,7	0,474	0,226
2,1	17,4	0,75	0,539	0,211
2,2	19,6	0,8	0,607	0,193
2,5	22,1	0,85	0,684	0,166
3,2	25,3	0,9	0,783	0,117
3,2	28,5	0,95	0,882	0,068
3,8	32,5			<b>3,594</b>

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

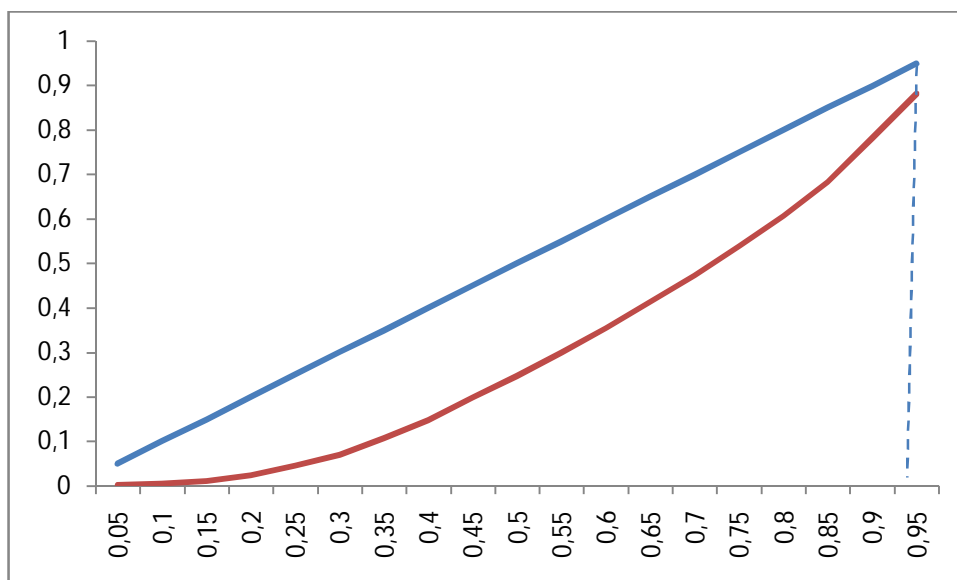
$$R = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

$$= R = \frac{3,594}{9,5} = 0,3783$$

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 (pi,qi) rappresentati sul piano cartesiano.



## Esercizio n 2

Data la distribuzione del carattere Reddito delle sole donne, si analizzi la mutua variabilità e si calcoli il Rapporto di Concentrazione di Gini.

REDDITO donne	Dati ordinati (sole donne)		TABELLA REDDITI DONNE										
0,1													
0,4													
0,8													
1,6	0,1	0,1	0,4	0,8	1,6	1,8	1,9	1,9	2,1	2,2	3,2		
1,8	0,4	0,3	0	0,4	1,2	1,4	1,5	1,5	1,7	1,8	2,8		
1,9	0,8	0,7	0,4	0	0,8	1	1,1	1,1	1,3	1,4	2,4		
1,9	1,6	1,5	1,2	0,8	0	0,2	0,3	0,3	0,5	0,6	1,6		
2,1	1,8	1,7	1,4	1	0,2	0	0,1	0,1	0,3	0,4	1,4		
2,2	1,9	1,8	1,5	1,1	0,3	0,1	0	0	0,2	0,3	1,3		
3,2	2,1	1,8	1,5	1,1	0,3	0,1	1,8	0	0,2	0,3	1,3		
	2,2	2	1,7	1,3	0,5	0,3	0,2	0,2	0	0,1	1,1		
	2,2	2,1	1,8	1,4	0,6	0,4	0,3	0,3	0,1	0	1		
	3,2	3,1	2,8	2,4	1,6	1,4	1,3	1,3	1,1	1	0		

EFFETTUANDO LA SOMMA DEI MARGINALI DI COLONNA

15	12,6	10,2	7	6,6	8,4	6,6	7,4	8	16	<b>SOMMA</b>
										<b>97,9</b>

$$\Delta = \frac{\sum_{i=1}^r \sum_{j=1}^n |x_i - x_j|}{n(n-1)} = \frac{97,9}{10 * 9} = 1,09$$

Un indice normalizzato in [0, 1] di mutua variabilità si ottiene dividendo  $\Delta$  per il suo massimo teorico  $2\mu$ . Tale indice si indica con R.

**$R = \Delta / 2\mu$ , noto come rapporto di concentrazione.**

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.

La media è pari a  $\bar{x} = 1,6$  per cui avremo  $R = \frac{1,09}{1,09 * 1,6} = 0,625$

### Esercizio 3

Considerando il dataset Laureati, si discuta se esiste connessione tra i caratteri Tipo di contratto ed Efficacia della laurea.

### Soluzione

La distribuzione doppia dei caratteri Genere e tipo di diploma è la seguente:

Occhiali \ Tipo di diploma	Tipo di diploma			Tot
	Liceo	Ist. tecnico	Altro diploma	
Si	5	3	1	9
No	5	5	1	11
totale	<b>10</b>	<b>8</b>	<b>2</b>	<b>20</b>

Trattandosi di due caratteri qualitativi, il loro grado di connessione si misura attraverso l'indice  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$$

Le frequenze teoriche  $\hat{n}_{ij} = \frac{n_{i.} \times n_{.j}}{N}$  sono raccolte nella seguente tabella:

		Tipo di diploma		
		Liceo	Ist. tecnico	Altro diploma
Occhiali	Si	4,5	3,6	0,9
	No	5,5	4,4	1,1

$$\chi^2 = \frac{(5-4,5)^2}{4,5} + \frac{(5-5,5)^2}{5,5} + \frac{(3-3,6)^2}{3,6} + \frac{(5-4,4)^2}{4,4} + \frac{(1-0,9)^2}{0,9} + \frac{(1-1,1)^2}{1,1} =$$

Tabella di calcolo

0,056	0,1	0,01	<b>SOMMA</b> <b>0,3</b>
0,045	0,08	0,009	

L'indice  $\chi^2$  dipende dalla numerosità del collettivo, cosicché, a parità di associazione, il suo valore aumenta all'aumentare di N. Generalmente si preferisce utilizzare degli indici "normalizzati" che diano misure

non dipendenti dalla numerosità.  $\phi^2 = \frac{\chi^2}{N} = \frac{0,3}{20} = \mathbf{0,015}$

Possiamo affermare che non c'è connessione tra i due caratteri

Tale valore va confrontato con l'intervallo [0, 1], in quanto

$$0 \leq \phi^2 \leq \min(r - 1; c - 1)$$