Università di Cassino
Corso di Statistica 1
Esercitazione del 19/11/2007
Dott. Alfonso Piscitelli

Esercizio 1

Un'indagine del 2005, sulle iscrizioni universitarie di due città, rileva i seguenti dati:

	Facoltà		
Città	Umanistica	Scientifica	Totale
Milano	250	170	**420**
Napoli	180	220	**400**
Totale	**430**	**390**	**820**

Si stabilisca se la scelta del tipo di facoltà è indipendente dal luogo di residenza e si commenti il risultato.

L'indice più opportuno per lo studio della relazione tra due mutabili è il χ^2, in quanto si tratta di due variabili entrambe di natura qualitativa e quindi non possiamo studiare altro legame che la connessione.

$$\chi^2 = \sum_{i=1}^{R}\sum_{j=1}^{C}\frac{c_{ij}^2}{n_{ij}^{'}}$$

Dove R sono le righe e C le colonne della Tabella; n_{ij} la frequenza doppia; $n_{i.}$ la frequenza marginale di riga; $n_{.j}$ la frequenza marginale di colonna;

$$c_{ij} = n_{ij} - n_{ij}^{'}$$

e

$$n_{ij}^{'} = {(n_{i.}n_{.j})}\big/{N}$$

Calcolo delle frequenze teoriche $n_{ij}^{'}$:

	Facoltà		
Città	Umanistica	Umanistica	Totale
Milano	$(420*430)/820 = 220{,}24$	$(420*390)/820 = 199{,}76$	**420**
Napoli	$(400*430)/820 = 209{,}76$	$(400*390)/820 = 190{,}24$	**400**
Totale	**430**	**390**	**820**

Riportando le frequenze osservate e le frequenze teoriche in un'unica tabella si passa al calcolo del χ^2

	n_{ij}	n'_{ij}	$(n_{ij} - n'_{ij})^2 / n'_{ij}$
	250	220,24	4,02
	170	199,76	4,43
	180	209,76	4,22
	220	190,24	4,65
Totale (*N*)	820	820	

χ^2	17,32

L'indice χ^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.

Perarson:

$$\Phi^2 = \frac{\chi^2}{N}$$

= 0,021

In caso di indipendenza assume il suo valore minimo che è zero. Il valore massimo è pari a 1 solo quando il numero di righe o il numero di colonne è uguale a 2, altrimenti risulta maggiore di 1.

Cramer:

$$V = \frac{\Phi^2}{\min[(R-1),(C-1)]}$$

= 0,021

Proprietà:

$$0 \leq V \leq 1$$

V=0 se vi è indipendenza assoluta o stocastica o in distribuzione tra i caratteri.
V=1 se vi è massima connessione tra i caratteri.

Esercizio 2

La tabella che segue riporta la durata in anni di 125 fotografie stampate su tre diversi supporti fotografici.

	Durata			
Supporti	10 anni	12 anni	13 anni	Totale
Supporto A	15	16	14	**45**
Supporto B	12	14	13	**39**
Supporto C	14	14	13	**41**
Totale	**41**	**44**	**40**	**125**

Si studi la relazione statistica tra il supporto e la durata della stampa.

L'indice più opportuno per lo studio della relazione tra una variabile qualitativa ed una quantitativa, dove la variabile quantitativa dipende da quella qualitativa, è l'indice η^2 .

$$\eta^2 = \frac{Dev.(B)}{Dev.(T)}$$

Il rapporto è dato dalla *devianza tra i gruppi (Between),* indicata con Dev.(B) sulla *devianza totale* indicata con Dev.(T).
Se la variabile quantitativa dipendente è posta in colonna valgono le seguenti formule:

$$Dev.(B) = \sum_{i=1}^{R} (\mu_{y|\mathrm{x_i}} - \mu_y)^2 n_{i.} \qquad Dev.(W) = \sum_{i=1}^{R} [\sum_{J=1}^{C} (y_j - \mu_{y|\mathrm{x_i}})^2 n_{ij}]$$

$$Dev.(T) = \sum_{J=1}^{C} (y_j - \mu_y)^2 n_{.j}$$

Calcolo della media Generale:

$$\mu_y = \frac{1}{N} \sum_{j=1}^{C} y_j n_{.j}$$

y_j	$n_{.j}$	$(y_j * n_{.j})$
10	41	410
12	44	528
13	40	520
	125	1458

$$\mu_y = \frac{1458}{125} = 11{,}66$$

Calcolo delle medie parziali:

$$\mu_{y|x_i} = \sum_{j=1}^{C} \frac{y_j n_{ij}}{n_{i.}}$$

nel nostro esercizio sono tre: la media del supporto A, la media del supporto B, la media del supporto C

y_j	n_{ij}	$(y_j * n_{ij})$
10	15	150
12	16	192
13	14	182
	45	524

y_j	n_{ij}	$(y_j * n_{ij})$
10	12	120
12	14	168
13	13	169
	39	457

y_j	n_{ij}	$(y_j * n_{ij})$
10	14	140
12	14	168
13	13	169
	41	477

$$\mu_{y|x_A} = \frac{524}{45} = 11,64 \qquad \mu_{y|x_B} = \frac{457}{39} = 11,72 \qquad \mu_{y|x_C} = \frac{477}{41} = 11,63$$

Calcolo della Devianza Totale:

	y_j	$n_{.j}$	$(y_j - \mu_y)^2 n_{.j}$
	10	41	112,98
	12	44	5,09
	13	40	71,82
Totale (*n*)		125	
Devianza Totale			189,89

Calcolo della Devianza Between:

	μ_y	$n_{i.}$	$(\mu_{y|x_i} - \mu_y)^2 n_{i.}$
	11,64	45	0,018
	11,72	39	0,140
	11,63	41	0,037
Totale (*n*)		125	
Devianza Between			0,195

$$\eta^2 = \frac{0,195}{189,89} = 0,001$$

Si può affermare che vi è una indipendenza, in media, della durata dal tipo di supporto.

Anche se non è necessario ai fini dell'esercizio, possiamo effettuare un **controllo sui calcoli** delle devianze utilizzando il "Teorema di decomposizione della devianza"

Dev.(T) = Dev.(B) + Dev.(W)

Calcolo della Devianza Within:

	y_j	n_{ij}	$(y_j - \mu_{y\|x_i})^2 n_{ij}$
	10	15	40,34
	12	16	2,07
	13	14	25,89
$n_{i.}$		45	
Devianza Gruppo A			68,3

	y_j	n_{ij}	$(y_j - \mu_{y\|x_i})^2 n_{ij}$
	10	12	35,5
	12	14	1,1
	13	13	21,3
$n_{i.}$		39	
Devianza Gruppo B			57,9

	y_j	n_{ij}	$(y_j - \mu_{y\|x_i})^2 n_{ij}$
	10	14	37,2
	12	14	1,9
	13	13	24,4
$n_{i.}$		41	
Devianza Gruppo C			63,5

Dev.(W) = 68,3 + 57,9 + 63,5 = 189,7

Dev.(T) = Dev.(B) + Dev.(W) = 0,195 + 189,7 = 189,895