Università di Cassino

Corso di Statistica

Esercitazione del 19 febbraio 2010

Simona Balzano

ESERCIZIO 1

Nella tabella che segue è riportata la distribuzione doppia di frequenza per i caratteri AREA GEOGRAFICA e LIVELLO DI INQUINAMENTO relativi alle imprese delle 20 regioni:

Area geografica \ Livello di inquinamento	Alto	Basso	Medio	**Totale**
Nord	3	2	2	**7**
Centro	3	4	0	**7**
Sud	1	1	2	**4**
Isole	2	0	0	**2**
Totale	**9**	**7**	**4**	**20**

Misurare il grado di connessione tra i due caratteri.

Soluzione

Il grado di connessione tra due caratteri qualitativi si misura mediante l'indice chi-quadro, che può essere calcolato in due modi.

Metodo 1: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-c_{ij})^2}{c_{ij}}$

Le frequenze teoriche c_{ij} sono ottenute come: $c_{ij} = \frac{n_{i.} \times n_{.j}}{N}$

Le c_{ij} sono nella tabella:

c_{ij}	Alto	Basso	Medio	Totale
Nord	3.15	2.45	1.4	**7**
Centro	3.15	2.45	1.4	**7**
Sud	1.8	1.4	0.8	**4**
Isole	0.9	0.7	0.4	**2**
Totale	**9**	**7**	**4**	**20**

L'indice è dunque pari a:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij}-c_{ij}\right)^2}{c_{ij}} =$$

$$= \frac{(3-3,15)^2}{3,15} + \frac{(2-2,45)^2}{2,45} + \frac{(2-1,4)^2}{1,4} + \cdots + \frac{(0-0,4)^2}{0,4} =$$

$$= 0,01 + 0,08 + 0,26 + \cdots + 0,4 = 7,45$$

A partire da tale valore si possono ottenere gli indici Φ^2 (fi-quadro) e T normalizzato:

$$\Phi^2 = \frac{\chi^2}{N} = \frac{7,45}{20} = 0,37$$

$$T = \frac{\chi^2}{N \times \min(r-1; c-1)} = \frac{\Phi^2}{\min(r-1; c-1)} = \frac{7,45}{20 \times 2} = \frac{0,37}{2} = 0,195$$

<u>Metodo 2</u>: $\chi^2 = n\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}} - 1\right)$

n_{ij}^2	Alto	Basso	Medio
Nord	9	4	4
Centro	9	16	0
Sud	1	1	4
Isole	4	0	0

$n_{i\bullet}n_{\bullet j}$	Alto	Basso	Medio
Nord	63	49	28
Centro	63	49	28
Sud	36	28	16
Isole	18	14	8

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

$$= 20 \times \left[\left(\frac{9}{63} + \frac{4}{49} + \frac{4}{28} + \cdots + \frac{0}{8}\right) - 1\right] = 20 \times (0,14 + 0,08 + 0,14 + \cdots + 0 - 1) = 7,45$$

Seguono le usuali normalizzazioni (Φ^2 e T).

ESERCIZIO 2

La tabella che segue è riportata la distribuzione doppia relative ai caratteri AREA GEOGRAFICA e VALORE AGGIUNTO delle imprese delle 20 regioni:

Valore aggiunto / Area Geografica	0-25	25-50	50-100	100-300	**Totale**
Centro	3	1	1	2	**7**
Isole	0	1	1	0	**2**
Nord	1	3	0	3	**7**
Sud	1	1	2	0	**4**
Totale	**5**	**6**	**4**	**5**	**20**

Determinare in che misura il VALORE AGGIUNTO dipende in media dall'AERA GEOGRAFICA.

Soluzione

Il grado di dipendenza in media è misurate dall'indice eta:

$$\eta = \frac{\sigma^2_{EXT}}{\sigma^2} = \frac{\frac{1}{n}\sum_{i=1}^{k}\left(\mu_i - \mu_Y\right)^2 n_{i\bullet}}{\frac{1}{n}\sum_{j=1}^{h}\left(c_j - \mu\right)^2 n_{\bullet j}} = \frac{\sum_{i=1}^{k}\left(\mu_i - \mu\right)^2 n_{i\bullet}}{\sum_{j=1}^{h}\left(c_j - \mu\right)^2 n_{\bullet j}}$$

È necessario calcolare le quantità: μ (media generale), μ_i (medie dei gruppi), σ^2_{ext} (varianza esterna) e σ^2 (varianza totale).

I valori centrali delle 4 classi sono:

12.5 37.5 75 200

<u>Media generale</u>:

$$\mu = \frac{1}{n}\sum_{j=1}^{c} c_j n_{\bullet j} = \frac{1}{20}\left(12,5\times 5 + 37,5\times 6 + 75\times 4 + 200\times 5\right) = 79,38$$

<u>Medie di gruppo</u>: $\mu_i = \frac{1}{n_{i.}}\sum_{j=1}^{c} c_j n_{ij}$

Centro: $\mu_1 = \frac{1}{n_{1.}}\sum_{j=1}^{c} c_j n_{1j} = \frac{1}{7}\left(12,5\times 3 + 37,5\times 1 + 75\times 1 + 200\times 2\right) = 78,57$

Isole: $\mu_2 = \frac{1}{n_{2.}}\sum_{j=1}^{c} c_j n_{2j} = \frac{1}{2}\left(12,5\times 0 + 37,5\times 1 + 75\times 1 + 200\times 0\right) = 56,25$

Nord: $\mu_3 = \frac{1}{n_{3.}}\sum_{j=1}^{c} c_j n_{3j} = \frac{1}{7}(12,5\times 1 + 37,5\times 3 + 75\times 0 + 200\times 3) = 103,57$

Sud: $\mu_4 = \frac{1}{n_{4.}}\sum_{j=1}^{c} c_j n_{4j} = \frac{1}{4}(12,5\times 1 + 37,5\times 1 + 75\times 2 + 200\times 0) = 50$

Varianza totale:

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{h}(c_j - \mu)^2 n_{\bullet j} =$$

$$= \frac{1}{20}\left[(12,5-79,38)^2\times 5 + (37,5-79,38)^2\times 6 + (75-79,38)^2\times 4 + (200-79,38)^2\times 5\right] =$$

$$= \frac{1}{20}(22361,33 + 10521,09 + 76,56 + 72751,95) = 5285,55$$

Varianza esterna:

$$\sigma^2_{EXT} = \frac{1}{n}\sum_{i=1}^{k}(\mu_i - \mu_Y)^2 n_{i\bullet} =$$

$$= \frac{1}{20}\left[(78,57-79,38)^2\times 7 + (56,25-79,38)^2\times 2(103,57-79,38)^2\times 7 + (50-79,38)^2\times 4\right] =$$

$$= \frac{1}{20}(4,52 + 1069,53 + 4098,27 + 3451,56) = 431,19$$

Eta:

$$\eta = \frac{\sigma^2_{EXT}}{\sigma^2} = \frac{431,19}{5285,55} = 0,08$$