Esercitazione 5 del corso di Statistica (parte 1)

Dott.ssa Paola Costantini

29 Febbraio 2012

Esercizio n 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	Tot
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 le due variabili

Soluzione

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

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

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

Area geografica \ Livello di inquinamento	Alto	Basso	Medio	Tot
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

Sostituendo nella formula si ha:

$$\chi^2 = \frac{(3-3,15)^2}{3,15} + \frac{(2-2,45)^2}{2,45} + \frac{(2-1,4)^2}{1,4} + \ldots\ldots.. + \frac{(2-0,64)^2}{04} = \mathbf{7,45}$$

non dipendenti dalla numerosità $\phi^2 = \frac{\chi^2}{N} = \frac{7,45}{20} = \mathbf{0,37}$

Tale valore va confrontato con l'intervallo [0, 2], in quanto

$$0 \le \varphi^2 \le \min(r-1;\ c-1)$$

V di Cramer= $\frac{\chi^2}{N \times \min(r-1; c-1)} = \frac{\phi^2}{\min(r-1; c-1)} = \frac{7,45}{20 \times 20} = \frac{0,37}{2} = 0,185$

Metodo alternativo per il calcolo del χ^2

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

$ni.n.j$	Alto	Basso	Medio
Nord	63	49	28
Centro	63	49	28
Sud	36	28	16
Isole	18	14	8

$$\chi^2 = n\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 1\right) = 20 \div \left[\left(\frac{9}{63} + \frac{4}{49} + \frac{4}{49} + ... + \frac{0}{8}\right) - 1\right] = 20 \times (0{,}14 + 0{,}08 + 0{,}14 + ... + 0) - 1 = 7{,}45$$

Seguono le normalizzazioni ϕ^2 e V.

Esercizio n 2

La tabella che segue riporta la distribuzione doppia relativa 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'area geografica

Soluzione

Il grado di dipendenza in media è misurato dall'indice eta

$$\eta = \frac{\sigma_{EXT}^2}{\sigma^2} = \frac{\frac{1}{N}\sum_{i=1}^{K}(\mu_i - \mu_Y)^2 n_{i.}}{\frac{1}{N}\sum_{j=1}^{h}(c_j - \mu)^2 n_{.j}} = \frac{\sum_{i=1}^{K}(\mu_i - \mu_Y)n_{i.}}{\sum_{j=1}^{h}(c_j - \mu)n_{.j}}$$

E' necessario calcolare le quantità: μ media generale, μ_i medie dei gruppi, σ_{ext}^2 varianza esterna, e σ^2 varianza totale. I valori centrali delle classi sono 12,5   37,5   75   200

Media generale: $\mu = \frac{1}{n}\sum_{j=1}^{c} c_j n.j = \frac{1}{20}(12{,}5 * 5 + 37{,}5 * 6 + 75 * 4 + 200 * 5 = 79{,}38$

Medie di gruppo $\mu_i = \frac{1}{n_{i\cdot}}\sum_{j=1}^{c} c_j n_{ij}$

Centro: $\mu_1 = \frac{1}{n_{1\cdot}}\sum_{j=1}^{c} c_j n_{1j} = \frac{1}{7}(12{,}5 * 3 + 37{,}5 * 1 + 75 * 1 + 200 * 0) = 78{,}57$

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

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

Nord: $\mu_4 = \frac{1}{n_{3\cdot}} \sum_{j=1}^{c} c_j n_{4j} = \frac{1}{4}(12{,}5*1 + 37{,}5*1 + 75*2 + 200*0) = 50$

Varianza totale

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n}(c_j - \mu)^2 n.j = \frac{1}{20}\left[(12{,}5-79{,}38)^2*5 + (37{,}5-79{,}38)^2*6 + (75-79{,}38)^2*4+\right]$$

$$+(200-79{,}38)^2*5] = 5285{,}55$$

Varianza esterna

$$\frac{\sigma^2{}_{EXT}}{} = \frac{1}{N}\sum_{j=1}^{k}(\mu_j - \mu_Y)^2 \cdot n_{i.} = \frac{1}{20}\cdot$$

$$\cdot\left[(78{,}57-79{,}38)^2\cdot 7 + (56{,}25-79{,}38)^2\cdot 2 + (103{,}57-79{,}38)^2\cdot 7 + (50-79{,}38)^2\cdot 4\right] =$$
$$= 431{,}19$$

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

Esercizio 3

Da un collettivo di 20 studenti della facoltà di economia sono stati rilevati i voi ottenuti nella prova scritta e orale dell'esame di statistica. Si calcoli la concordanza o la discordanza tra i due caratteri mediante il calcolo della covarianza.

studente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
prova scritta	29	29	27	24	24	23	23	23	21	21	19	19	19	18	18	18	18	18	18	17
prova orale	22	28	30	25	27	23	22	25	24	24	25	24	25	23	27	26	22	25	23	18

La covarianza tra due caratteri quantitativi è definita come la media dei prodotti degli scostamenti delle variabili X e Y dalle rispettive medie:

Cov$_{x,y}$= $\mu(x \cdot y) - (\mu_x \cdot \mu_y) =$

Media voto scritto $\mu_x = \frac{426}{20} = 21{,}25$

Media voto orale $\mu_y = \frac{488}{20} = 24,4$

$\mu(x \cdot y)$=(22*29+28*29+30*27+…….18*17)/20=10466/20=523,3

Cov$_{x,y}$ = $\mu(x \cdot y) - (\mu_x \cdot \mu_y) = 522,05,3 - (21,25 * 24,4) = 3,55$