Università di Cassino

Esercitazioni di Statistica 1 del 25 Febbraio 2011

Dott. Mirko Bevilacqua

ESERCIZIO n° 1

La tabella che segue raccoglie i caratteri "Tipo Scuola" e "Punteggio Test" del dataset Test d'ingresso.

Tipo Scuola	GE	LC	LS	LS	LS	LS	LS	LS	MP	MP	TC	TC	TC	TC	TC	TC
Punteggio Test	15	14,75	15	15,75	17,25	17,5	18,5	28,8	7,25	8,75	5	8,5	9,25	17,25	22,3	26,15

Verificare la proprietà di scomponibilità della varianza.

MEDIA E VARIANZA TOTALE:

(Punteggio Test)^2	225	217,6	225	248,1	297,6	306,3	342,3	829,4	52,6	76,6	25	72,3	85,6	297,6	497,3	683,8

$$\mu_x = \frac{\sum_{i=1}^{16} x_i}{16} = 15,44 \qquad \mu_x^2 = \frac{\sum_{i=1}^{16} x_i}{16} = 280,11$$

$$\sigma_X^2 = \mu_x^2 - (\mu_x)^2 = 280,11 - (15,44)^2 = 41,79$$

MEDIE DEI GRUPPI:

$$\mu_{GE} = 15 \qquad \mu_{LC} = 14,75 \qquad \mu_{LS} = 18,8 \qquad \mu_{MP} = 8 \qquad \mu_{TC} = 14,74$$

$$\mu_{GE^2} = 225 \qquad \mu_{LC^2} = 217,6 \qquad \mu_{LS^2} = 374,8 \qquad \mu_{MP^2} = 64,6 \qquad \mu_{TC^2} = 276,9$$

VARIANZA <u>TRA I GRUPPI</u> O <u>ESTERNA</u> (σ_{EXT}^2):

$$\sigma_{EXT}^2 = \frac{\sum_{i=1}^{G} (\mu_i - \mu)^2 n_i}{n} =$$

$$\frac{1}{16}\left[(15-15,44)^2 + (14,75-15,44)^2 + (18,80-15,44)^2 \times 6 + (8-15,44)^2 \times 2 + (14,74-15,44)^2 \times 6\right] = 11,38$$

VARIANZA ENTRO I GRUPPI O INTERNA, (σ^2_{INT}):

$$\sigma^2_{INT} = \frac{\sum_{i=1}^{G}\sum_{j=1}^{n_i}\left(x_{ij} - \mu_i\right)^2}{16} =$$

$$= \frac{1}{16}\begin{bmatrix} (15 - 15)^2 + (14{,}75 - 14{,}75)^2 + (15 - 18{,}8)^2 + \\ + \ldots + (28{,}8 - 18{,}8)^2 + (7{,}25 - 8)^2 + \\ (8{,}75 - 8)^2 + (5 - 14{,}74)^2 + \ldots + (26{,}15 - 14{,}7)^2 \end{bmatrix} = 30{,}41$$

Verifica della proprietà della scomponibilità:

$$\sigma^2_{EXT} + \sigma^2_{INT} = \sigma^2_X$$

$$\sigma^2_X = 11{,}38 + 30{,}41 = 41{,}79$$

ESERCIZIO n° 2

Considerando le seguenti classi per la variabile "punteggio Test", si può affermare che punteggio test dipende in media dal Tipo scuola?

Punteggio Test (Y) / Tipo scuola (X)	5 |-| 8,75	8,75 -| 15	15 -| 17,5	17,5 -| 28,8	**Totale**
TC	2	1	1	2	**6**
LC	0	1	0	0	**1**
LS	0	1	3	2	**6**
GE	0	1	0	0	**1**
MP	2	0	0	0	**2**
Totale	**4**	**4**	**4**	**4**	**16**

Y (carattere quantitativo) è indipendente in media da X (carattere qualitativo) se tutte le medie condizionate della variabile Y sono fra loro uguali e uguali quindi anche alla media marginale:
$\mu_{Y/X=xi}=\mu_Y$ per ogni i = 1,2,…,k

L' indice relativo di dipendenza in media è

$$\eta_{Y|X} = \frac{\sigma^2_{ext_Y}}{\sigma^2_Y} = \frac{\sum_{i=1}^{r}\left(\mu_{Y|X=x_i} - \mu_Y\right)^2 \cdot n_{i.}}{\sum_{j=1}^{c}\left(y_j - \mu_Y\right)^2 \cdot n_{.j}}$$

$\eta_{Y|X} = 0$ perfetta indipendenza in media di Y da X

$\eta_{Y|X} = 1$ perfetta dipendenza in media di Y da X

1. Media generale di Y:

$$\mu_Y = \frac{\sum_{j=1}^{5} y_j n_{.j}}{N} = \frac{(6,87 \cdot 4)+(11,87 \cdot 4)+(16,25 \cdot 4)+(23,15 \cdot 4)}{16} = 14,54$$

2. Medie di Y condizionate alle modalità di X

$$\mu_{PT/TC} = \frac{\sum_{j=1}^{6} y_{TC} n_{1j}}{n_{1.}} = \frac{(6,87 \cdot 2)+(11,87)+(16,25)+(23,15 \cdot 2)}{6} = 15$$

$$\mu_{PT/LC} = \frac{\sum_{j=1}^{1} y_{LC} n_{2j}}{n_{2.}} = 11,87$$

$$\mu_{PT/LS} = \frac{\sum_{j=1}^{6} y_{LS} n_{3j}}{n_{3.}} = \frac{(11,87)+(16,25 \cdot 3)+(23,25 \cdot 2)}{6} = 17,82$$

$$\mu_{PT/GE} = \frac{\sum_{j=1}^{1} y_{GE} n_{4j}}{n_{4.}} = 11,87$$

$$\mu_{PT/MP} = \frac{\sum_{j=1}^{2} y_{LS} n_{5j}}{n_{5.}} = \frac{(6,87 \cdot 2)}{2} = 6,87$$

3. Calcolo del numeratore dell'indice:

$$\sum_{i=1}^{r} \left(\mu_{Y|X=x_i} - \mu_Y\right)^2 \cdot n_{i.} =$$

$$(14,7-14,54)^2 \cdot 6 + (11,87-14,54)^2 \cdot 1 +$$

$$+(17,82-14,54)^2 \cdot 6 + (11,87-14,54)^2 + (6,87-14,54)^2 = 196,44$$

4. Calcolo del denominatore dell'indice:

$$\sum_{j=1}^{c} \left(y_j - \mu_Y\right)^2 \cdot n_{.j} =$$

$$(6,87-14,54)^2 \cdot 4 + (11,87-14,54)^2 \cdot 4 +$$

$$+(16,25-14,54)^2 \cdot 4 + (23,15-14,54)^2 \cdot 4 = 571,64$$

5. Calcolo dell'indice:

$$\eta_{Y|X} = \frac{\sigma^2_{ext_Y}}{\sigma^2_Y} = \frac{\sum_{i=1}^{r} \left(\mu_{Y|X=x_i} - \mu_Y\right)^2 \cdot n_{i.}}{\sum_{j=1}^{c} \left(y_j - \mu_Y\right)^2 \cdot n_{.j}} = \frac{196,44}{571,64} = 0,34$$

Il valore dell'indice indica un basso grado di dipendenza in media.

ESERCIZIO n°3:

Da un collettivo di 15 studenti della facoltà di economia sono stati rilevati altezza e peso. Si calcoli la concordanza o la discordanza tra i due caratteri mediante il calcolo della covarianza.

Altezza X	Peso Y	$X_i-\mu_X$	$Y_i-\mu_Y$	$(X_i-\mu_X)\cdot(Y_i-\mu_Y)$	$X_i \cdot Y_i$
179	65	7,0	2,2	15,4	11635
180	62	8,0	-0,8	-6,4	11160
165	50	-7,0	-12,8	89,6	8250
160	49	-12,0	-13,8	165,6	7840
160	47	-12,0	-15,8	189,6	7520
160	48	-12,0	-14,8	177,6	7680
164	56	-8,0	-6,8	54,4	9184
170	59	-2,0	-3,8	7,6	10030
180	73	8,0	10,2	81,6	13140
186	86	14,0	23,2	324,8	15996
170	66	-2,0	3,2	-6,4	11220
180	68	8,0	5,2	41,6	12240
180	85	8,0	22,2	177,6	15300
176	56	4,0	-6,8	-27,2	9856
170	72	-2,0	9,2	-18,4	12240
				$\sum_{i=1}^{n}(x_i-\mu_x)\cdot(y_i-\mu_y)=1267$	$\sum_{i=1}^{n}(y_i \cdot x_i)=163291$

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

$$cov(x,y)=\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu_y)\cdot(x_i-\mu_x)$$

Quest'indice misura la concordanza o la discordanza tra due caratteri quantitativi.

$$cov(x,y)=\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu_y)\cdot(x_i-\mu_x)=\frac{1267}{15}=84,47$$

Formula alternativa per la covarianza:

$$cov(x,y)=\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu_y)\cdot(x_i-\mu_x)=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\mu_x\cdot\mu_y$$

$$\mu_X=\frac{\sum_{i=1}^{15}x_i}{15}=172 \qquad \mu_y=\frac{\sum_{i=1}^{15}y_i}{15}=62,8$$

Media dei prodotti:

$$\frac{1}{n}\sum_{i=1}^{n}x_iy_i=\frac{163291}{15}=10886,1$$

$$cov(x,y)=10886,1-(172\cdot 62,8)=84,47$$