

Facoltà di Economia - Corso di Statistica - 2003/2004**Correzione esercitazione n. 4****Es. A)**

- Distribuzione doppia dei caratteri *zona* e *numero di crimini*

| Zona | Numero di Crimini | | | Totale |
|---------------|-------------------|-----------|-----------|-----------|
| | [0;250) | [250;500) | [500;750) | |
| Nord Centro | 3 | 2 | 0 | 5 |
| Nord Est | 4 | 0 | 1 | 5 |
| Ovest | 3 | 2 | 0 | 5 |
| Sud | 5 | 0 | 0 | 5 |
| Totale | 15 | 4 | 1 | 20 |

- Distribuzioni percentuali condizionate del *numero di crimini*

| Zona | Numero di Crimini | | | Totale |
|-------------|-------------------|-----------|-----------|--------|
| | [0;250) | [250;500) | [500;750) | |
| Nord Centro | 60 | 40 | 0 | 100 |
| Nord Est | 80 | 0 | 20 | 100 |
| Ovest | 60 | 40 | 0 | 100 |
| Sud | 100 | 0 | 0 | 100 |

da cui si nota che la regione influenza il numero dei crimini: a differenza delle zone nord e ovest, nel sud ci sono solo aree con bassa criminalità. Quindi c'è connessione tra i due caratteri.

- Distribuzione marginale del *numero di crimini* con frequenze relative e relative cumulate

| Numero di Crimini ($c_{i-1} - c_i$) | n_i | f_i | F_i |
|--|-----------|-------------|----------|
| [0;250) | 15 | 0,75 | 0,75 |
| [250;500) | 4 | 0,20 | 0,95 |
| [500;750) | 1 | 0,05 | 1,00 |
| Totale | 20 | 1,00 | - |

Es. B)

- Distribuzione marginale del *tipo di occupazione* con frequenze assolute cumulate

| Tipo di occupazione | n_i | N_i |
|---------------------|--------------|----------|
| Agricoltura | 288 | 288 |
| Industria | 1.630 | 1.918 |
| Altre attività | 3.286 | 5.204 |
| Totale | 5.204 | - |

- Distribuzioni condizionate percentuali del *tipo di occupazione* data la regione

| Regione | Tipo di occupazione | | | Totale |
|----------------|---------------------|-----------|----------------|--------|
| | Agricoltura | Industria | Altre attività | |
| Piemonte | 5,14 | 39,93 | 54,93 | 100 |
| Emilia Romagna | 7,14 | 34,95 | 57,91 | 100 |
| Lazio | 4,40 | 19,92 | 75,67 | 100 |

da cui si vede che Piemonte e Emilia Romagna hanno una struttura simile, mentre il Lazio ha molta meno industria e più terziario (altre attività).

- Tabella di indipendenza (con frequenze \hat{n}_{ij})

| Regione | Tipo di occupazione | | | Totale |
|----------------|---------------------|------------------|------------------|------------------|
| | Agricoltura | Industria | Altre attività | |
| Piemonte | 93,694 | 530,282 | 1.069,023 | 1.693,000 |
| Emilia Romagna | 93,749 | 530,596 | 1.069,655 | 1.694,000 |
| Lazio | 100,556 | 569,122 | 1.147,322 | 1.817,000 |
| Totale | 288,000 | 1.630,000 | 3.286,000 | 5.204,000 |

Tabella delle contingenze assolute (con elementi c_{ij})

| Regione | Agricoltura | Industria | Altre attività | Totale |
|----------------|--------------|--------------|----------------|--------------|
| Piemonte | -6,694 | 145,718 | -139,023 | 0,000 |
| Emilia Romagna | 27,251 | 61,404 | -88,655 | 0,000 |
| Lazio | -20,556 | -207,122 | 227,678 | 0,000 |
| Totale | 0,000 | 0,000 | 0,000 | 0,000 |

- Per il calcolo degli indici di connessione si imposta una tabella con elementi c_{ij}^2/\hat{n}_{ij}

| Regione | Agricoltura | Industria | Altre attività |
|----------------|-------------|-----------|----------------|
| Piemonte | 0,478 | 40,042 | 18,080 |
| Emilia Romagna | 7,921 | 7,106 | 7,348 |
| Lazio | 4,202 | 75,378 | 45,181 |

da cui $\chi^2 = 205,737$, $\psi = \sqrt{205,737/5.204} = 0,199$, $C = 0,199/\sqrt{2} = 0,141$

Es. C)

- Distribuzioni condizionate percentuali del carattere *anni di reclusione*

| Tipo di reato | Reclusione | | | Totale |
|---------------------|------------|-----------|------------|--------|
| | 0-5 anni | 5-10 anni | 10-20 anni | |
| Omicidio volontario | 18,74 | 18,33 | 62,93 | 100,00 |
| Omicidio colposo | 99,94 | 0,06 | 0,00 | 100,00 |

- Tabella di indipendenza (con frequenze \hat{n}_{ij})

| Tipo di reato | Reclusione | | | Totale |
|---------------------|-----------------|----------------|----------------|-----------------|
| | 0-5 anni | 5-10 anni | 10-20 anni | |
| Omicidio volontario | 630,475 | 22,939 | 77,587 | 731,000 |
| Omicidio colposo | 3107,525 | 113,061 | 382,413 | 3603,000 |
| Totale | 3738,000 | 136,000 | 460,000 | 4334,000 |

tabella delle contingenze assolute (con elementi c_{ij})

| Tipo di reato | Reclusione | | | Totale |
|---------------------|--------------|--------------|--------------|--------------|
| | 0-5 anni | 5-10 anni | 10-20 anni | |
| Omicidio volontario | -493,475 | 111,061 | 382,413 | 0,000 |
| Omicidio colposo | 493,475 | -111,061 | -382,413 | 0,000 |
| Totale | 0,000 | 0,000 | 0,000 | 0,000 |

- Per il calcolo degli indici di connessione si imposta una tabella con elementi c_{ij}^2/\hat{n}_{ij}

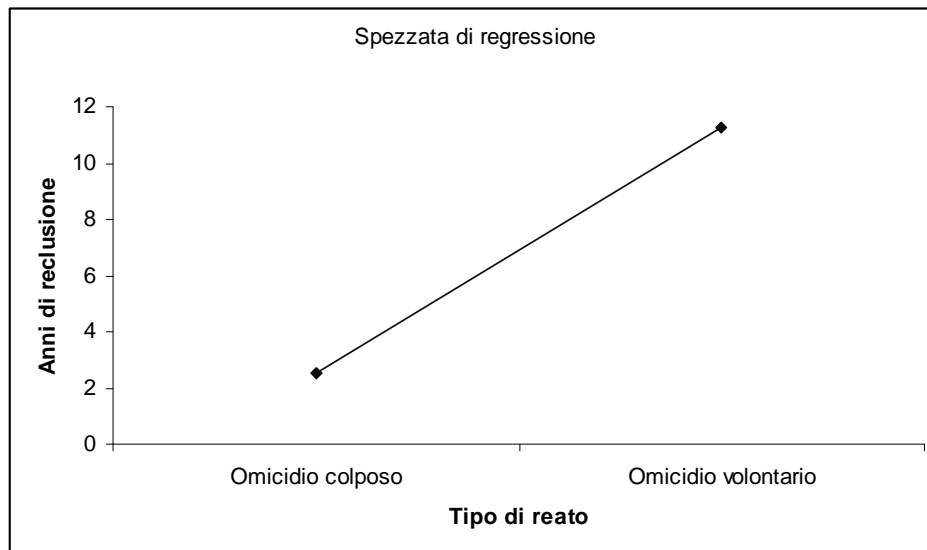
| Tipo di reato | Reclusione | | |
|---------------------|------------|-----------|------------|
| | 0-5 anni | 5-10 anni | 10-20 anni |
| Omicidio volontario | 386,244 | 537,723 | 1.884,864 |
| Omicidio colposo | 78,364 | 109,097 | 382,413 |

da cui $\chi^2 = 3.378,706$, $C = \sqrt{3.378,71/4334} = 0,883$ che indica una connessione molto elevata.

- Le medie condizionate di Y si calcolano secondo lo schema seguente (i valori centrali delle classi sono indicati con y_j)

| Omicidio volontario | | | Omicidio colposo | | |
|---------------------|----------|--------------|------------------|----------|--------------|
| y_j | n_{1j} | $y_j n_{1j}$ | y_j | n_{2j} | $y_j n_{2j}$ |
| 2,5 | 137 | 342,5 | 2,5 | 3.601 | 9.002,5 |
| 7,5 | 134 | 1.005,0 | 7,5 | 2 | 15,0 |
| 15,0 | 460 | 6.900,0 | 15,0 | 0 | 0,0 |
| Totale | 731 | 8.247,5 | Totale | 3.603 | 9.017,5 |

da cui $\bar{y}_1 = 8.247,5/731 = 11,282$ e $\bar{y}_2 = 9.027,5/3603 = 2,503$.



- Per l'indice η^2 si consideri innanzitutto la distribuzione marginale di Y

| y_j | $n_{\circ j}$ | $y_j n_{\circ j}$ | $(y_j - \bar{y})^2 n_{\circ j}$ |
|-------|---------------|-------------------|---------------------------------|
| 2,5 | 3.738 | 9.345 | 8.227,794 |
| 7,5 | 136 | 1.020 | 1.681,632 |
| 15,0 | 460 | 6.900 | 55.825,910 |
| | 4.334 | 17.265 | 65.735,337 |

da cui $\bar{y} = 17.265/4334 = 3,984$, $D_Y = 65.735,337$. Il calcolo della devianza spiegata viene effettuato secondo lo schema

| \bar{y}_i | $n_{i\circ}$ | $(\bar{y}_i - \bar{y})^2 n_{i\circ}$ |
|-------------|--------------|--------------------------------------|
| 11,282 | 731 | 38.942,950 |
| 2,503 | 3.603 | 7.900,998 |
| | 4.334 | 46.843,949 |

da cui $D_S = 46.843,949$ e quindi

$$\eta^2 = \frac{46.843,949}{65.735,337} = 0,713$$

che indica una forte dipendenza in media degli *anni di reclusione* dal *tipo di reato*.

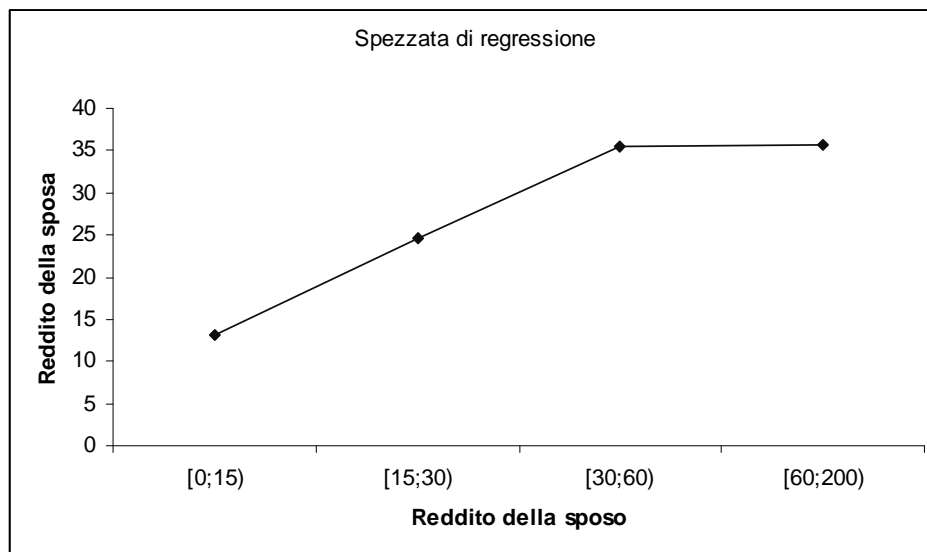
Es. D)

- Si hanno le seguenti medie condizionate del *reddito della sposa* (Y) per ogni modalità del carattere *reddito dello sposo* (X)

| Reddito dello sposo (i) | \bar{y}_i |
|-----------------------------|-------------|
| [0 ; 15) | 13,182 |
| [15 ; 30) | 24,643 |
| [30 ; 60) | 35,440 |
| [60 ; 200) | 35,676 |

Per calcolare la prima media condizionata, ad esempio, si è utilizzata la formula

$$\bar{y}_1 = \frac{7,5 \cdot 25 + 22,5 \cdot 5 + 45 \cdot 3 + 90 \cdot 0}{33} = 13,182.$$



da cui si nota come, a parte i redditi più elevati, al crescere del reddito dello sposo cresce anche quello della sposa.

Per l'indice η^2 si consideri innanzitutto la distribuzione marginale di Y

| y_j | $n_{\circ j}$ | $y_j n_{\circ j}$ | $(y_j - \bar{y})^2 n_{\circ j}$ |
|---------------|---------------|-------------------|---------------------------------|
| 7,5 | 66 | 495,0 | 35.104,008 |
| 22,5 | 86 | 1.935,0 | 5.590,336 |
| 45,0 | 67 | 3.015,0 | 13.965,574 |
| 90,0 | 21 | 1.890,0 | 74.189,145 |
| Totale | 240 | 7.335,0 | 128.849,063 |

da cui $\bar{y} = 7.335 / 240 = 30,563$, $D_Y = 128.849,063$. Il calcolo della devianza spiegata viene effettuato secondo lo schema

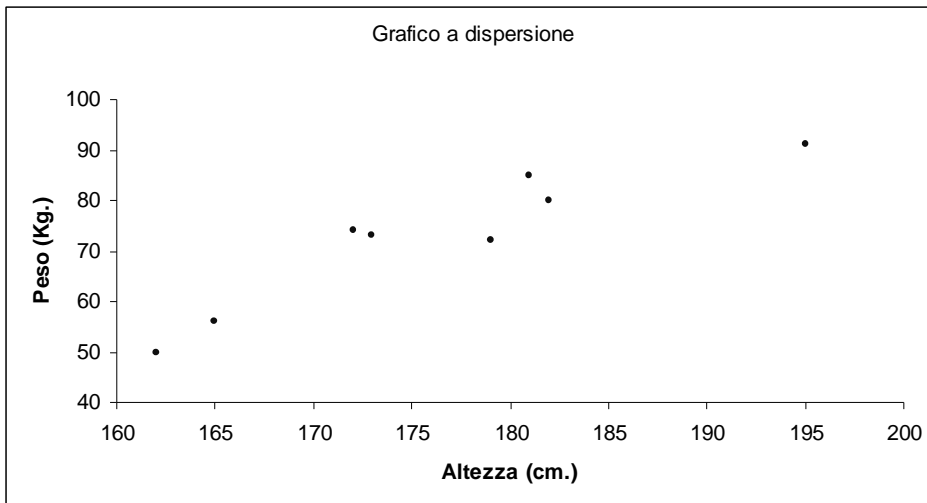
| \bar{y}_i | $n_{i \circ}$ | $(\bar{y}_i - \bar{y})^2 n_{i \circ}$ |
|---------------|---------------|---------------------------------------|
| 13,182 | 33 | 9.968,907 |
| 24,643 | 42 | 1.471,771 |
| 35,440 | 91 | 2.164,500 |
| 35,676 | 74 | 1.934,698 |
| Totale | 240 | 15.539,877 |

da cui $D_S = 15.539,877$ e quindi

$$\eta^2 = \frac{15.539,877}{128.849,063} = 0,121$$

che indica una modesta dipendenza in media del *reddito della sposa* dal *reddito dello sposo*.

Es. E)



- Calcolo dei parametri della retta di regressione del *peso* (Y) dall'*altezza* (X) con il metodo dei minimi quadrati

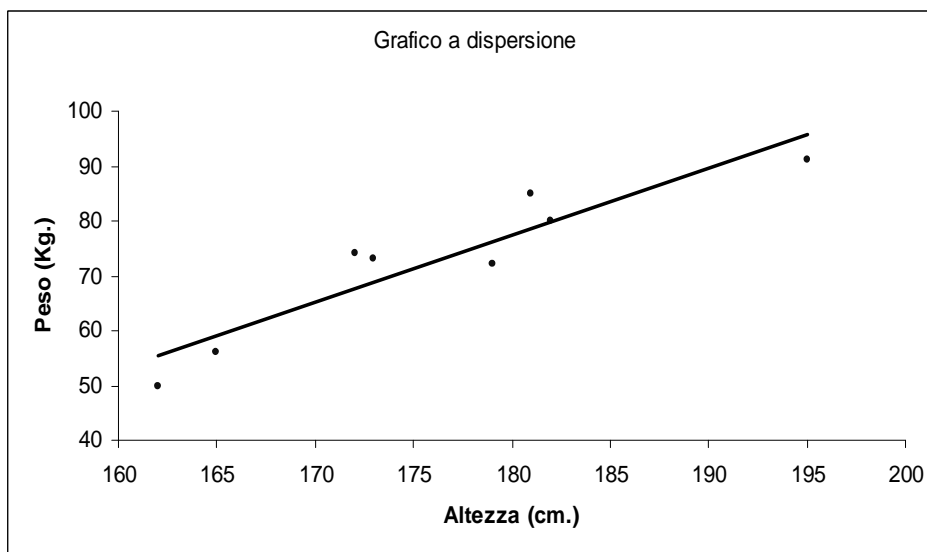
| x_i | y_i | x_i^2 | $x_i y_i$ | y_i^2 |
|-------|-------|---------|-----------|---------|
| 165 | 56 | 27.225 | 9.240 | 3.136 |
| 182 | 80 | 33.124 | 14.560 | 6.400 |
| 173 | 73 | 29.929 | 12.629 | 5.329 |
| 172 | 74 | 29.584 | 12.728 | 5.476 |
| 162 | 50 | 26.244 | 8.100 | 2.500 |
| 179 | 72 | 32.041 | 12.888 | 5.184 |
| 181 | 85 | 32.761 | 15.385 | 7.225 |
| 195 | 91 | 38.025 | 17.745 | 8.281 |
| 1.409 | 581 | 248.933 | 103.275 | 43.531 |

da cui $\bar{x} = 1.409/8 = 176,125$, $\bar{y} = 581/8 = 72,625$. Inoltre si ha

$$C_{XY} = 103.275/8 - 176,125 \cdot 76,625 = 118,297 \text{ e } D_X = 248.933 - 8 \cdot 176,125^2 = 772,875$$

da cui

$$\hat{\beta}_1 = \frac{946,375}{772,875} = 1,224 \quad \text{e} \quad \hat{\beta}_0 = 72,625 - 1,224 \cdot 176,125 = -143,038.$$



- Per calcolare l'indice di correlazione r occorre la devianza di Y che è pari a

$$D_Y = 43.531 - 8 \cdot 72,625^2 = 1.335,875$$

da cui

$$r = \frac{946,375}{\sqrt{772,875 \cdot 1.335,875}} = 0,931$$

che indica una elevata correlazione positiva.

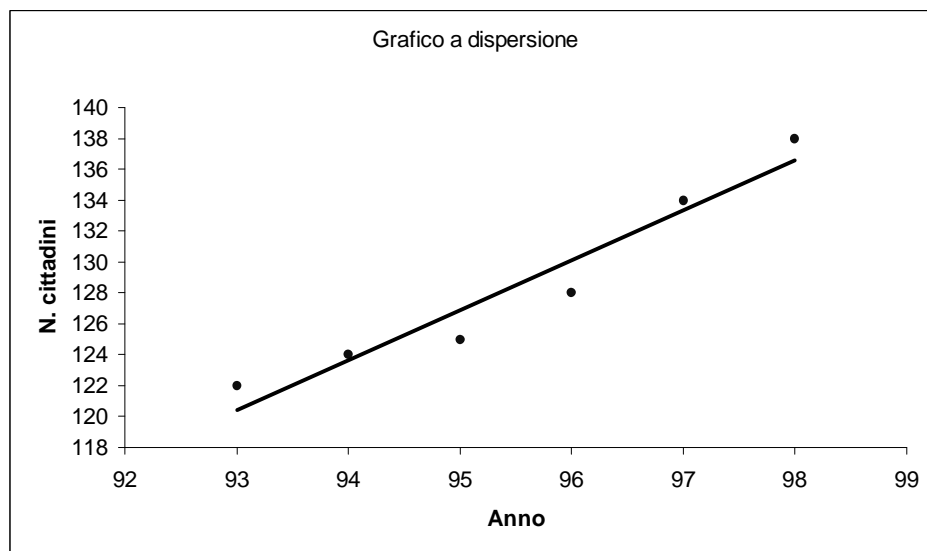
Es. F)

- Calcolo dei parametri della retta di regressione con il metodo dei minimi quadrati

| x_i | y_i | x_i^2 | $x_i y_i$ | y_i^2 |
|-------|-------|---------|-----------|---------|
| 93 | 122 | 8.649 | 11.346 | 14.884 |
| 94 | 124 | 8.836 | 11.656 | 15.376 |
| 95 | 125 | 9.025 | 11.875 | 15.625 |
| 96 | 128 | 9.216 | 12.288 | 16.384 |
| 97 | 134 | 9.409 | 12.998 | 17.956 |
| 98 | 138 | 9.604 | 13.524 | 19.044 |
| 573 | 771 | 54.739 | 73.687 | 99.269 |

da cui $\bar{x} = 95,5$, $\bar{y} = 128,5$. Inoltre si ha $C_{XY} = 56,5$ e $D_X = 17,5$ da cui

$$\hat{\beta}_1 = \frac{56,5}{17,5} = 3,229 \quad \text{e} \quad \hat{\beta}_0 = 128,5 - 3,229 \cdot 95,5 = -179,829.$$



- Per calcolare l'indice di determinazione r^2 occorre la devianza di Y che è pari a $D_Y = 195,5$ da cui

$$r^2 = \frac{56,5^2}{17,5 \cdot 195,5} = 0,933.$$

che indica un ottimo della retta di regressione ai dati.

- Previsione del numero di *cittadini di altri stati della UE* residenti in Italia al 1° gennaio 2004

$$y_7^* = -179,829 + 3,229 \cdot 104 = 155,943.$$