Università di Cassino

Esercitazione di Statistica 1 del 27 novembre 2006

Dott.ssa Simona Balzano

Esercizio 1

Determinare in quale misura i caratteri peso e altezza della seguente distribuzione doppia sono correlati:

Peso (y) \ Altezza (x)	160 |-| 164	164 -| 170	170 -| 178	178 -| 186	**Totale**
46 |-| 56	5	1	1	0	**7**
56 -| 66	0	2	0	3	**5**
66 -| 76	0	2	1	2	**5**
76 -| 86	0	0	1	2	**3**
Totale	**5**	**5**	**3**	**7**	**20**

Soluzione

Quando i caratteri sono espressi come distribuzione doppia di frequenza il coefficiente di correlazione $\rho = \frac{\mathrm{cod}(x,y)}{\mathrm{dev}(x)\cdot\mathrm{dev}(y)}$ si può calcolare <u>con la formula abbreviata</u>:

$$\rho = \frac{\mu_{XY} - \mu_X\mu_Y}{\sqrt{\left(\mu_{2X} - \mu_X^2\right)\left(\mu_{2Y} - \mu_Y^2\right)}} = \frac{\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{h}\hat{x}_i\hat{y}_j n_{ij} - \mu_X\mu_Y}{\sqrt{\frac{1}{n}\sum_{i=1}^{k}x_i^2 n_{i.} - \mu_X^2}\sqrt{\frac{1}{n}\sum_{j=1}^{h}y_j^2 n_{.j} - \mu_Y^2}}$$

che consente di ridurre il numero di operazioni necessarie.

Valori centrali per il peso: $x_1 = 51$; $x_2 = 61$; $x_3 = 71$; $x_4 = 81$

Valori centrali per l'altezza: $y_1 = 162$; $y_2 = 167$; $y_3 = 174$; $y_4 = 182$

Le quantità $\hat{x}_i \hat{y}_j n_{ij}$ sono raccolte nella tabella che segue:

$\hat{x}_i \hat{y}_j n_{ij}$	**162**	**167**	**174**	**182**
51	41.310	8.517	8.874	0
61	0	20.374	0	33.306
71	0	23.714	12.354	25.844
81	0	0	14.094	29.484

Totale generale: 217.871

Da cui deriva:

$$\mu_{XY} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{h} x_i y_j n_{ij}}{n} = \frac{217.871}{20} = 10.894$$

Le altre quantità necessarie sono contenute nella seconda tabella:

x_i	$n_{i.}$	y_j	$n_{.j}$	$x_i n_{i.}$	$y_j n_{.j}$	x_i^2	$x_i^2 n_{i.}$	y_j^2	$y_j^2 n_{.j}$
51	7	162	5	357	810	2601	18207	26244	131220
61	5	167	5	305	835	3721	18605	27889	139445
71	5	174	3	355	522	5041	25205	30276	90828
81	3	182	7	243	1274	6561	19683	33124	231868
Totali	**20**		**20**	**1.260**	**3.441**		**81.700**		**593.361**

$$\mu_y = \frac{1}{n}\sum_{j=1}^{c} y_j n_{.j} = \frac{3.441}{20} = 172,05 \qquad \text{altezza media}$$

$$\mu_x = \frac{1}{n}\sum_{i=1}^{r} x_i n_{i.} = \frac{1260}{20} = 63 \qquad \text{peso medio}$$

$$\mu_{2y} = \frac{1}{n}\sum_{j=1}^{c} y_i^2 n_{.j} = \frac{593.361}{20} = 29.668,05$$

$$\mu_{2x} = \frac{1}{n}\sum_{i=1}^{r} y_i^2 n_{i.} = \frac{81.700}{20} = 4.085$$

Sostituendo i valori ottenuti nella formula:

$$\operatorname{cov}(x,y) = \mu_{XY} - \mu_x\mu_y = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{h} \hat{x}_i \hat{y}_j n_{ij} - \mu_X\mu_Y = 10.894 - 172,05 \times 63 = 54,4$$

$$\rho = \frac{\mu_{XY} - \mu_X\mu_Y}{\sqrt{\left(\mu_{2X} - \mu_X^2\right)\left(\mu_{2Y} - \mu_Y^2\right)}} = \frac{54,4}{\sqrt{\left(29.668,05 - 172,05^2\right) \times \left(4.085 - 63^2\right)}} = \frac{54,4}{88,06} = \mathbf{0,62}$$

Tale valore va confrontato con l'intervallo [-1, 1], quindi indica una correlazione lineare positiva abbastanza forte.

Esercizio 2

Sullo spazio campionario:

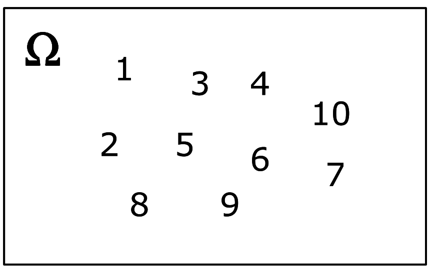
$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

considerando l'esperimento casuale "estrazione di un numero", determinare, rappresentare mediante diagrammi di Venn e calcolare la probabilità dei seguenti eventi:

a) A = estrazione di numero pari;
b) B = estrazione di numero dispari;
c) C = estrazione di numero > 7
d) Negazione di C
e) $A \cap C$
f) $A \cup C$
g) $A \cap B$

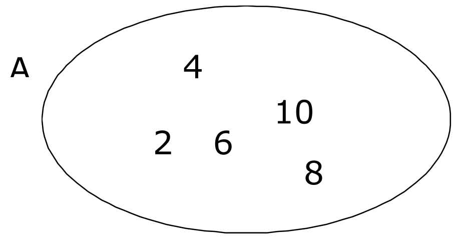
Soluzione

Lo spazio campione è rappresentato graficamente come segue:



a)

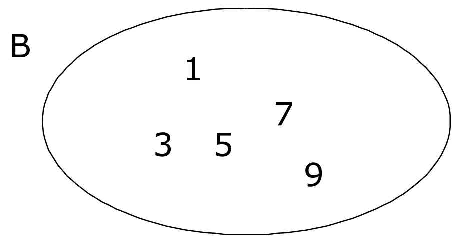
L’evento A = “estrazione di un numero pari” è costituito dagli eventi elementari: {2}, {4}, {6}, {8}, {10} ed è rappresentato graficamente come segue:



$$P(A) = \frac{\text{numero di elementi pari}}{\text{numero totale di elementi di } \Omega} = \frac{5}{10} = 0,5$$

b)

L’evento B = “estrazione di un numero dispari” è costituito dagli eventi elementari: {1}, {3}, {5}, {7}, {9} ed è rappresentato graficamente come segue:



$$P(B) = \frac{\text{numero di elementi dispari}}{\text{numero totale di elementi di } \Omega} = \frac{5}{10} = 0,5$$

Nota:
A e B rappresentano una partizione di Ω, pertanto essi sono due eventi necessari ed incompatibili. Quindi la loro unione è Ω:

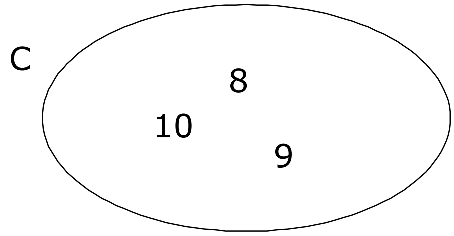
$$A \cup B = \Omega$$

mentre la loro intersezione è nulla:

$$A \cap B = \varnothing$$

c)

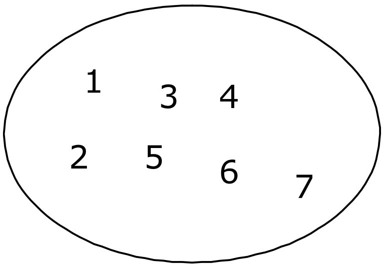
L'evento C = "estrazione di un numero maggiore di 7" è costituito dagli eventi elementari: {8}, {9}, {10} ed è rappresentato graficamente come segue:



$$P(B) = \frac{\text{numero di elementi} > 7}{\text{numero totale di elementi di } \Omega} = \frac{3}{10} = 0,3$$

d)

L'evento $\overline{C}$ è dato degli elementi di Ω che non fanno parte di C, ossia {1}, {2}, {3}, {4}, {5}, {6}, {7}.

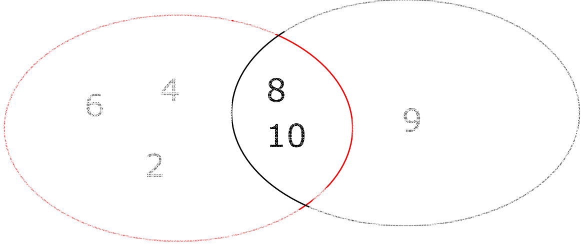


$$P(\overline{C}) = \frac{\text{numero di elementi} \leq 7}{\text{numero totale di elementi di } \Omega} = \frac{7}{10} = 0,7$$

e)

L'evento $E_1 = A \cap C$ è costituito dagli elementi che fanno parte sia di A sia di C, ossia {8}, {10}:

$A \cap C$



$$P(E_1) = P(A \cap C) = \frac{2}{10} = 0,2$$

Avremmo potuto calcolare $P(E_1)$ in base al Teorema delle probabilità composte:

$$P(A \cap C) = P(A \mid C) \times P(C) = P(C \mid A) \times P(A)$$

utilizzando i valori:

$$P(A \mid C) = \frac{2}{3} \qquad \text{ovvero} \qquad P(C \mid A) = \frac{2}{5}$$

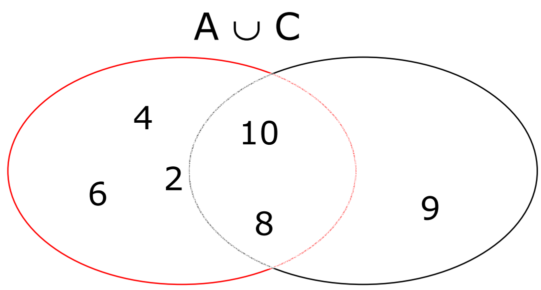
$$P(C) = \frac{3}{10} \qquad \qquad P(A) = \frac{5}{10}$$

da cui:

$$P(A \cap C) = \frac{2}{3} \times \frac{3}{10} = \frac{2}{5} \times \frac{5}{10} = 0,2$$

f)

L'evento $E_2 = A \cup C$ è costituito dagli elementi che fanno parte di A o di C, ossia {2}, {4}, {6}, {8}, {9}, {10}



$$P(E_2) = P(A \cup C) = \frac{6}{10} = 0,6$$

Avremmo potuto calcolare $P(E_2)$ in base al Teorema delle probabilità totali:

$$P(A \cup C) = P(A) + P(B) - P(A \cap B) = 0,5 + 0,3 - 0,2 = 0,6$$

Esercizio 3

Considerando gli eventi elemenatri associati alla seguente distribuzione doppia:

Attività sportiva / Corso laurea	Nulla (N)	Media (M)	Alta (A)	**Totale**
Biologia (B)	0	1	0	**1**
Informatica (I)	4	7	1	**12**
Matematica (Mat)	2	5	0	**7**
Totale	**6**	**13**	**1**	**20**

determinare:

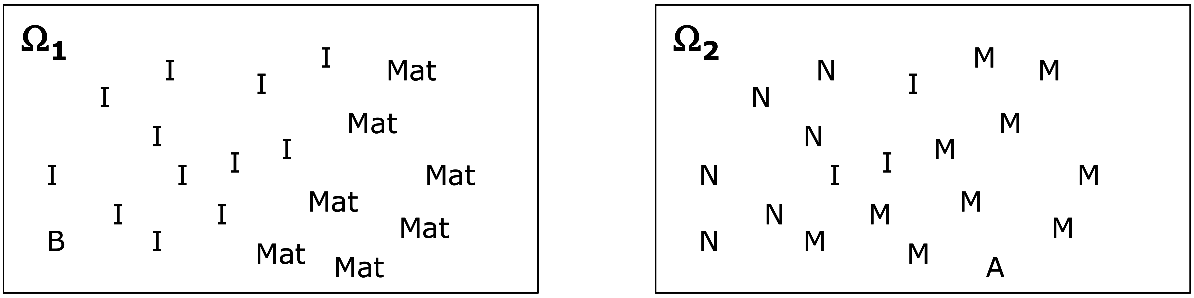
a) le probabilità elementari

Inoltre si determinino le probabilità che, scegliendo uno studente a caso:

b) sia iscritto in matematica e pratichi attività sportiva media;
c) sia iscritto in biologia e pratichi attività sportiva alta;
d) sia iscritto in informatica o pratichi attività sportiva nulla;
e) essendo iscritto in informatica, pratichi attività sportiva nulla;
f) praticando attività sportiva media, sia iscritto in biologia.

Soluzione

Una distribuzione doppia equivale a considerare le associazioni tra 2 spazi campione Ω_1 e Ω_2, che in questo caso sono:



ciascuno, quindi, costituito da 20 elementi.

a)
gli eventi elementari sono riferiti in entrambi gli spazi alla selezione casuale di uno dei 20 studenti.

Gli eventi elementari dello spazio Ω_1 sono:

B = "studente iscritto al CDL in biologia"
I = "studente iscritto al CDLin informatica"
Mat = "studente iscritto al CDL in matematica"

Gli eventi elementari dello spazio Ω_2 sono:

N = "attività sportiva nulla"
M = "attività sportiva media"
A = "attività sportiva alta"

Le probabilità elementari si ottengono come frequenze relative marginali della tabella.

$$P(B) = \frac{1}{20} = 0,05 \qquad P(N) = \frac{6}{20} = 0,3$$
$$P(I) = \frac{12}{20} = 0,6 \qquad P(M) = \frac{13}{20} = 0,65$$
$$P(Mat) = \frac{7}{20} = 0,35 \qquad P(Mat) = \frac{1}{20} = 0,05$$

Naturalmente la somma delle probabilità in ciascuno dei due spazi campionari è pari ad 1:

$$\begin{aligned}\sum_i P(E_i) &= 0,05 + 0,6 + 0,35 = \\ &= 0,3 + 0,65 + 0,05 = \\ &= 1\end{aligned}$$

b) sia "iscritto in matematica" *e* "pratichi attività sportiva media"

Si tratta della probabilità dell'intersezione dei due eventi:
$$P(Mat \cap M) = \frac{5}{20} = 0,25$$

c) sia "iscritto in biologia" *e* "pratichi attività sportiva alta"

Si tratta della probabilità dell'intersezione dei due eventi:
$$P(B \cap A) = \frac{0}{20} = 0$$

d) sia "iscritto in informatica" *o* "pratichi attività sportiva nulla"

Si tratta della probabilità dell'unione dei due eventi:

$$P(I \cup N) = P(I) + P(N) - P(I \cap N) = \frac{12}{20} + \frac{6}{20} - \frac{4}{20} = \frac{14}{20} = 0,7$$

e) *essendo iscritto in informatica*, pratichi attività sportiva nulla:

Si tratta di una probabilità condizionata, dove l'iscrizione in informatica è l'evento condizionante, quindi lo spazio campione di riferimento si riduce a quello costituito dai soli studenti iscritti in informatica.

$$P(N \mid I) = \frac{P(N \cap I)}{P(I)} = \frac{4/20}{12/20} = \frac{4}{12} = 0,\overline{3}$$

f) *praticando attività sportiva media*, sia iscritto in biologia:

Si tratta di una probabilità condizionata, dove il praticare attività sportiva media è l'evento condizionante, quindi lo spazio campione di riferimento si riduce a quello costituito dai soli che praticano attività sportiva media.

$$P(B \mid M) = \frac{P(B \cap M)}{P(M)} = \frac{1/20}{13/20} = \frac{1}{13} = 0,769$$