Esercizio 1

La seguente tabella riporta la distribuzione di un collettivo di 70 studenti per numero di anni impiegati per conseguire la laurea (Y) e titolo di studio (X)

X $\backslash Y$	3	4	5
Liceo Classico	15	10	0
Liceo Scientifico	5	3	12
Altro	5	15	5

Calcolare:

1. le distribuzioni marginali, $n_{i.}$ e $n_{j.}$;
2. le frequenze relative congiunte f_{ij}
3. le distribuzioni di frequenza marginali relative, $f_{i.}$ e $f_{.j}$;
4. le distribuzioni relative condizionate.

Soluzione

1. le distribuzioni marginali di X e Y si ottengono calcolando i totali delle frequenze assolute per riga e per colonna, rispettivamente, , $n_{i.} = \sum_{j=1}^{K} n_{ij}$ e $n_{.j} = \sum_{i=1}^{H} n_{ij}$

X $\backslash Y$	3	4	5	Totale
Liceo Classico	15	10	0	**25**
Liceo Scientifico	5	3	12	**20**
Altro	5	15	5	**25**
Totale	**25**	**28**	**17**	70

2. Le frequenze relative congiunte si ottengono come $f_{ij} = \dfrac{n_{ij}}{n}$
3. mentre quelle marginali come $f_{i.} = \dfrac{n_{i.}}{n}$ e $f_{.j} = \dfrac{n_{.j}}{n}$

X $\backslash Y$	3	4	5	Totale
Liceo Classico	0.2143	0.1429	0	**0.3572**
Liceo Scientifico	0.0714	0.0429	0.1714	**0.2857**
Altro	0.0714	0.2143	0.0714	**0.3571**
Totale	**0.3571**	**0.4001**	**0.2428**	1

Calcoliamo la distribuzione titolo di studio condizionata alle diverse modalità di Y.

$X \mid Y = 3$	$n_{i1}/n_{\cdot 1}$
Liceo Classico	0.6
Liceo Scientifico	0.2
Altro	0.2
Totale	1

$X \mid Y = 4$	$n_{i2}/n_{\cdot 2}$
Liceo Classico	0.357
Liceo Scientifico	0.107
Altro	0.536
Totale	1

$X \mid Y = 5$	$n_{i3}/n_{\cdot 3}$
Liceo Classico	0
Liceo Scientifico	0.706
Altro	0.294
Totale	1

Calcoliamo la distribuzione degli anni impiegati per conseguire la laurea condizionata alle diverse modalità di X.

$Y \mid X = Lic.Classico$	$n_{1j}/n_{1\cdot}$
3	0.6
4	0.4
5	0
Totale	1

$Y \mid X = Lic.Scientifico$	$n_{2j}/n_{2\cdot}$
3	0.25
4	0.15
5	0.60
Totale	1

$Y \mid X = Altro$	$n_{3j}/n_{3\cdot}$
3	0.20
4	0.60
5	0.20
Totale	1

Esercizio 2

Con riferimento alla seguente distribuzione di un collettivo di individui secondo il sesso (X) ed il salario (Y).

X $\backslash Y$	1	2	3
M	0	10	10
F	10	10	0

Calcolare:

1. le distribuzioni marginali di X e Y;
2. le distribuzioni di Y condizionate alle diverse modalità di X;
3. le medie e le varianze di Y condizionate alle diverse modalità di X;
4. la media e la varianza di Y.

Soluzione

Otteniamo le distribuzioni marginali di X e Y sommando rispetto alle colonne e alle righe, rispettivamente.

	1	2	3	Totale
M	0	10	10	**20**
F	10	10	0	**20**
Totale	**10**	**20**	**10**	**40**

Otteniamo le distribuzioni condizionate di Y dato X

Y| X=M	n_{1j}
1	0
2	10
3	10
Totale	**20**

Y| X=F	n_{2j}
1	0
2	10
3	10
Totale	**20**

Calcoliamo le medie e le varianze condizionate di $Y \mid X = x_i$ con $i = 1,\dots,H$. Dato che X è un carattere qualitativo sconnesso non possiamo calcolare le medie e le varianze condizionate di $X \mid Y = y_j$ con $j = 1,\dots,K$.

Y|X=M	n_{1j}	$y_j \cdot n_{1j}$	$y_j^2 \cdot n_{1j}$
1	0	0	0
2	10	20	40
3	10	30	90
Totale	**20**	**50**	**130**

$$\bar{y}_{X=x_i} = \frac{1}{n_{i.}} \sum_{j=1}^{K} y_j n_{ij}$$
$$\bar{y}_{X=M} = \frac{50}{20} = 2.5$$
$$\sigma^2_{Y|X=x_i} = \frac{1}{n_{i.}} \sum_{j=1}^{K} y_j^2 n_{ij} - \bar{y}^2_{X=x_i}$$
$$\sigma^2_{Y|X=M} = \frac{130}{20} - 2.5^2 = 0.25$$

Y|X=F	n_{2j}	$y_j \cdot n_{2j}$	$y_j^2 \cdot n_{2j}$
1	10	10	10
2	10	20	40
3	0	0	0
Totale	**20**	**30**	**50**

$$\bar{y}_{X=x_i} = \frac{1}{n_{i.}} \sum_{j=1}^{K} y_j n_{ij}$$
$$\bar{y}_{X=F} = \frac{30}{20} = 1.5$$
$$\sigma^2_{Y|X=x_i} = \frac{1}{n_{i.}} \sum_{j=1}^{K} y_j^2 n_{ij} - \bar{y}^2_{X=x_i}$$
$$\sigma^2_{Y|X=F} = \frac{50}{20} - 1.5^2 = 0.25$$

Per calcolare la media e la varianza di Y dobbiamo considerare la distribuzione marginale di Y

y_j	n_j	$y_j \cdot n_j$	$y_j^2 \cdot n_j$
1	10	10	10
2	20	40	80
3	10	30	90
Totale	**40**	**80**	**180**

$$\bar{y} = \frac{80}{40} = 2$$
$$\sigma_y^2 = \frac{180}{40} - 2^2 = 4.5 - 4 = 0.5$$

Esercizio 3

Con riferimento alla distribuzione di un collettivo di famiglie secondo il numero di figli (X) e il reddito (Y), completare la seguente tabella nel caso di indipendenza

X \ Y	0⊣ 2	2⊣ 6	6 ⊣ 10	Totale
1		20		40
2				
3		10		
Totale			30	**100**

Soluzione

$n_{13}: \ \dfrac{30 \cdot 40}{100} = 12$

$n_{11}:$ 40-(12+20)=8

$n_{\cdot 1}: \ \dfrac{40 \cdot x}{100} = 8 \Rightarrow x = 20$

$n_{\cdot 2}:$ 100-(30+20)=50

$n_{22}:$ 50-(20+10)=20

$n_{2 \cdot}: \ \dfrac{50 \cdot x}{100} = 20 \Rightarrow x = 40$

$n_{3 \cdot}:$ 100-(40+40)=20

$n_{33}: \ \dfrac{20 \cdot 30}{100} = 6$

$n_{31}:$ 20-(10+6)=4

$n_{21}:$ 20-(8+4)=8

$n_{23}:$ 40-(20+8)=12

X $\backslash Y$	0⊣ 2	2⊣ 6	6 ⊣ 10	Totale
1	**8**	20	**12**	40
2	**8**	**20**	**12**	**40**
3	**4**	10	**6**	**20**
Totale	**20**	**50**	30	**100**

Esercizio 4

Con riferimento alla seguente distribuzione di un collettivo di individui secondo il sesso (Y) ed il comune di residenza (X)

$X \setminus Y$	M	F
A	0	5
B	5	0
C	10	10

Calcolare le frequenze sotto l'ipotesi di indipendenza tra i due caratteri.

Soluzione

Prima di tutto occorre ricavarsi le distribuzioni marginali di X e Y.

$X \setminus Y$	M	F	Totale
A			5
B			5
C			20
Totale	15	15	30

La frequenza d'indipendenza è pari a $n_{ij}^{\star} = \dfrac{n_{i.}n_{.j}}{n}$. Quindi otteniamo la seguente tabella

$X \setminus Y$	M	F	Totale
A	2.5	2.5	5
B	2.5	2.5	5
C	10	10	20
Totale	15	15	30

Esercizio 5

Nella tabella seguente si riportano l'ammontare degli investimenti (X) e i volumi di vendita (Y)

X	Y
1	11
1	8.6
2	10.5
2.1	12
2.2	12.8
2.9	14.7
3	13.5
3	14
3	12.7
3.1	16.4

Sapendo che $\bar{x} = 2.33$ e $\bar{y} = 12.62$, calcolare una misura della correlazione lineare tra i due caratteri.

Soluzione

X	Y	$x_i \cdot y_i$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	11	11	1.769	2.624
1	8.6	8.6	1.769	16.160
2	10.5	21	0.109	4.494
2.1	12	25.2	0.053	0.384
2.2	12.8	28.16	0.017	0.032
2.9	14.7	42.63	0.325	4.326
3	13.5	40.5	0.449	0.774
3	14	42	0.449	1.904
3	12.7	38.1	0.449	0.006
3.1	16.4	50.84	0.593	14.288
Totale		308.03	5.982	44.992

$$\bar{x} = \frac{23.3}{10} = 2.33$$
$$\bar{y} = \frac{126.2}{10} = 12.62$$
$$\rho_{xy} = \frac{COD(X,Y)}{\sqrt{dev(X)dev(Y)}}$$
$$COD(X,Y) = \sum_i x_i y_i - n\bar{x}\bar{y} = 308.03 - 10 \cdot (2.33 \cdot 12.62) = 13.984$$
$$\rho_{xy} = \frac{13.984}{\sqrt{5.982 \cdot 44.992}} = 0.852$$

Esercizio 6

Data la seguente tabella

X\\Y	-1	1	Totale
-1	30	10	40
1	20	40	60
Totale	50	50	100

Calcolare il coefficiente di correlazione lineare tra i caratteri X e Y.

Soluzione

Per poter calcolare il coefficiente di correlazione lineare, dobbiamo calcolare le seguenti quantità

$$\bar{x} = \frac{-40+60}{100} = \frac{20}{100} = 0.2$$
$$\bar{y} = \frac{-50+50}{100} = 0$$

x_i	y_j	n_{ij}	$x_i y_j n_{ij}$
-1	-1	30	30
1	-1	20	-20
-1	1	10	-10
1	1	40	40
Totale		100	40

Dalla tabella precedente è immediato verificare che la codevianza sia pari a 40.

$$Dev(x) = \sum_{i=1}^{H}(x_i - \bar{x})^2 n_{i.} = (-1-0.2)^2 40 + (1-0.2)^2 60 = 57.6 + 38.4 = 96$$
$$Dev(y) = \sum_{j=1}^{K}(y_j - \bar{y})^2 n_{.j} = (-1)^2 50 + (1)^2 50 = 50 + 50 = 100$$
$$\rho_{xy} = \frac{40}{\sqrt{100 \cdot 96}} = 0.408$$

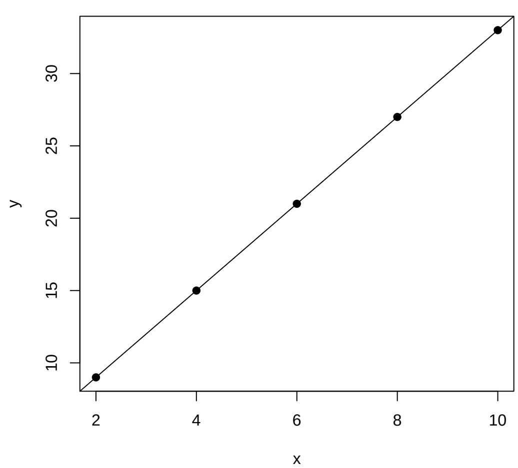
Esercizio 7

La tabella seguente riporta i valori osservati per i caratteri X e Y relativamente a 5 unità statistiche con $Y = 3 + 3X$.

	X	Y=3+3X
A	2	9
B	4	15
C	6	21
D	8	27
E	10	33
Totale	30	105

1. fornire una rappresentazione grafica per i caratteri X e Y
2. calcolare il coefficiente di correlazione lineare

Soluzione



	x_i	y_i	$x_i \cdot y_i$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
A	2	9	18	16	144
B	4	15	60	4	36
C	6	21	126	0	0
D	8	27	216	4	36
E	10	33	330	16	144
Totale	30	105	750	40	360

$$\sigma_{xy} = \frac{750}{5} - (6 \cdot 21) = 150 - 126 = 24$$

$$\sigma_x = \sqrt{\frac{40}{5}} = 2.8284$$

$$\sigma_y = \sqrt{\frac{360}{5}} = 8.4853 \; \rho_{xy} = \frac{24}{2.8284 \cdot 8.4853} = 1$$

$\rho_{xy} = 1$ (come ci aspettavamo) significa che sussiste un perfetto legame lineare tra i due caratteri.
Inoltre verifichiamo che

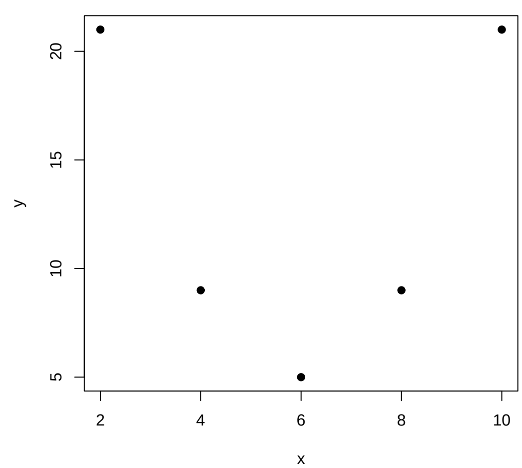
- $\bar{y} = a + b\bar{x} \Rightarrow \bar{y} = 3 + 3 \cdot 6 = 21$
- $\sigma_y^2 = b^2 \sigma_x^2 \Rightarrow \sigma_y^2 = 3^2 \cdot 8 = 72$
- $\sigma_{xy} = b\sigma_x^2 \Rightarrow \sigma_{xy} = 3 \cdot 8 = 24$

Esercizio 8

Ripetere l'esercizio precedente sapendo che $Y = 5 + (X - 6)^2$

	X	Y=5+(X-6)^2
A	2	21
B	4	9
C	6	5
D	8	9
E	10	21
Totale	30	65

Soluzione



$$\bar{x} = \frac{30}{5} = 6$$
$$\bar{y} = \frac{65}{5} = 13$$

	x_i	y_i	$x_i \cdot y_i$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
A	2	21	42	16	64
B	4	9	36	4	16
C	6	5	30	0	64
D	8	9	72	4	16
E	10	21	210	16	64
Totale	30	65	390	40	224

$$\sigma_{xy} = \frac{390}{5} - (6 \cdot 13) = 78 - 78 = 0$$
$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{0}{\sigma_x \sigma_y} = 0.$$

$\rho_{xy} = 0$ (come ci aspettavamo) significa che la relazione tra i due caratteri non è lineare.