ESERCIZIO 4.1

Un istituto di credito intende studiare la relazione tra titolo di studio e tipo di investimento attraverso un campione di 60 suoi clienti che mostrano la seguente distribuzione doppia:

n_{ij}	Investimento			
Titolo di studio	**Azioni**	**Fondi**	**Polizze vita**	**Totale**
ELEMENTARE	0	13	1	**14**
LAUREA	8	10	3	**21**
MATURITÀ	6	3	2	**11**
MEDIA	0	12	2	**14**
Totale	**14**	**38**	**8**	**60**

È possibile sostenere che i due caratteri sono indipendenti?

Soluzione

Trattandosi di due caratteri qualitativi, il loro grado di connessione si misura attraverso l'indice χ^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 $\hat{n}_{ij} = \dfrac{n_{i.} \times n_{.j}}{N}$ sono raccolte nella seguente tabella:

$\hat{n}_{ij}$	Investimento			
Titolo di studio	**Azioni**	**Fondi**	**Polizze vita**	**Totale**
ELEMENTARE	3,27	8,87	1,87	**14**
LAUREA	4,90	13,30	2,80	**21**
MATURITÀ	2,57	6,97	1,47	**11**
MEDIA	3,27	8,87	1,87	**14**
Totale	**14**	**38**	**8**	**60**

Sostituendo nella formula si ha:

$$\chi^2 = \frac{(0-3{,}27)^2}{3{,}27} + \frac{(13-8{,}87)^2}{8{,}87} + \frac{(1-1{,}87)^2}{1{,}87} + \cdots + \frac{(12-8{,}87)^2}{8{,}87} + \frac{(2-1{,}87)^2}{1{,}87}$$

In tabella gli addendi della sommatoria:

$\frac{(n_{ij}-\hat{n}_{ij})^2}{\hat{n}_{ij}}$	**Investimento**			**Totale = χ^2**
Titolo di studio	**Azioni**	**Fondi**	**Polizze vita**	
ELEMENTARE	3,27	1,93	0,40	
LAUREA	1,96	0,82	0,01	
MATURITÀ	4,59	2,26	0,19	
MEDIA	3,27	1,11	0,01	
Totale = χ^2				**19,82**

da cui risulta:

$$\chi^2 = 19{,}82$$

Quindi si ha:

$$\phi^2 = \frac{\chi^2}{N} = \frac{19{,}82}{60} = 0{,}33$$

e l'indice normalizzato:

$$\frac{\phi^2}{min(r-1;c-1)} = \frac{\chi^2}{N \cdot min(r-1;c-1)} = \frac{0{,}33}{2} = 0{,}17$$

ESERCIZIO 4.2

Nella tabella che segue sono riportati i dati relativi al reddito medio (X) ed al consumo medio in generi alimentari (Y) delle famiglie italiane residenti nelle 20 regioni italiane.
a) Si calcoli la covarianza tra X ed Y.
b) Si valuti se c'è indipendenza lineare tra i due caratteri.

Regione	X reddito	Y consumi
Piemonte	31,5	8
Valle d'Aosta	28	6,9
Lombardia	37,9	8,7
Trentino Alto Adige	34,2	8
Veneto	36,9	8,5
Friuli Venezia Giulia	31,1	7,7
Liguria	28,1	8,6
Emilia Romagna	37,3	8,4
Toscana	32,1	8,6
Umbria	32,3	8

Regione	X reddito	Y consumi
Marche	30,9	8,8
Lazio	30,4	8,4
Abruzzo	25,5	7,8
Molise	25,4	8
Campania	20,9	8
Puglia	25,1	8,9
Basilicata	20,9	6,7
Calabria	21,8	7,6
Sicilia	21,7	7,7
Sardegna	23,6	8,4

Soluzione

Le quantità necessarie al calcolo degli indici desiderati sono sintetizzate nella tabella che segue:

Regione	X reddito	Y consumi	$X-\mu_x$	$Y-\mu_y$	$(X-\mu_x)^2$	$(Y-\mu_y)^2$	$(X-\mu_x)(Y-\mu_y)$
Piemonte	31,5	8	2,72	-0,085	7,398	0,007	-0,231
Valle d'Aosta	28	6,9	-0,78	-1,185	0,608	1,404	0,924
Lombardia	37,9	8,7	9,12	0,615	83,174	0,378	5,609
Trentino Alto Adige	34,2	8	5,42	-0,085	29,376	0,007	-0,461
Veneto	36,9	8,5	8,12	0,415	65,934	0,172	3,370
Friuli Venezia Giulia	31,1	7,7	2,32	-0,385	5,382	0,148	-0,893
Liguria	28,1	8,6	-0,68	0,515	0,462	0,265	-0,350
Emilia Romagna	37,3	8,4	8,52	0,315	72,590	0,099	2,684
Toscana	32,1	8,6	3,32	0,515	11,022	0,265	1,710
Umbria	32,3	8	3,52	-0,085	12,390	0,007	-0,299
Marche	30,9	8,8	2,12	0,715	4,494	0,511	1,516
Lazio	30,4	8,4	1,62	0,315	2,624	0,099	0,510
Abruzzo	25,5	7,8	-3,28	-0,285	10,758	0,081	0,935
Molise	25,4	8	-3,38	-0,085	11,424	0,007	0,287
Campania	20,9	8	-7,88	-0,085	62,094	0,007	0,670
Puglia	25,1	8,9	-3,68	0,815	13,542	0,664	-2,999
Basilicata	20,9	6,7	-7,88	-1,385	62,094	1,918	10,914
Calabria	21,8	7,6	-6,98	-0,485	48,720	0,235	3,385
Sicilia	21,7	7,7	-7,08	-0,385	50,126	0,148	2,726
Sardegna	23,6	8,4	-5,18	0,315	26,832	0,099	-1,632
Somme	**575,6**	**161,7**	**0**	**0**	**581,052**	**6,526**	**28,374**
Medie	**28,78**	**8,085**			**29,053**	**0,326**	**1,419**

a)

$$cov(X,Y) = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N} = 1{,}419$$

Nota:
Utilizzando la formula breve (direttamente sulle osservazioni) avremmo:

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

$$= \frac{(31{,}5 \times 8) + (28 \times 6{,}9) + (37{,}9 \times 8{,}7) + \cdots + (23{,}6 \times 8{,}4)}{20} - (28{,}78 \times 8{,}085) =$$

$$= \frac{4.682{,}1}{20} - 232{,}69 = 1{,}419$$

b)
La presenza o assenza di dipendenza lineare si verifica attraverso il coefficiente di correlazione lineare di Bravais-Pearson ρ:

$$\rho = \frac{cod(x,y)}{\sqrt{dev(x) \cdot dev(y)}} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} = \frac{28{,}374}{\sqrt{581{,}052 \cdot 6{,}526}} = \frac{1{,}419}{\sqrt{29{,}053 \cdot 0{,}326}} = 0{,}461$$

ESERCIZIO 4.3

Sulla base dei dati riportati nella tabella precedente, si aggreghino le Regioni per area geografica (Nord, Centro e Sud+Isole) e si costruiscano due tabelle doppie ripartendo le variabili quantitative REDDITO e CONSUMI ALIMENTARI rispettivamente nelle classi:

a) REDDITO: [20 - 30], [30 - 40];
b) CONSUMI ALIMENTARI: [6,7 - 7,7],]7,7 - 8.0],]8,0 - 8,9].

Si valutati, quindi, se i redditi ed i consumi in beni alimentari sono indipendenti in media dall'area geografica.

Soluzione

L'indice di indipendenza in media, o rapporto di correlazione, η^2 si calcola come segue:

$$\eta^2_{Y|X} = \frac{\sigma^2_{est}}{\sigma^2} = \frac{dev_{est}}{dev(Y)} = \frac{\sum_{i=1}^{r}(\mu_i - \mu)^2 n_{i.}}{\sum_{j=1}^{c}(y_j - \mu)^2 n_{.j}}$$

a)

	Classi di REDDITO		
AREA	**20-30**	**30-40**	**Totale**
NORD	2	6	**8**
CENTRO	2	4	**6**
SUD	6	0	**6**
Totale	**10**	**10**	**20**

Considerando che:

$y_1 = 25$; $y_2 = 35$;

$$\mu = \frac{\sum_{j=1}^{2} y_j n_{.j}}{N} = \frac{(25 \times 10) + (35 \times 10)}{20} = 30\ ;$$ Reddito medio generale

$$\mu_1 = \mu_{nord} = \frac{\sum_{j=1}^{2} y_j n_{1j}}{n_{1.}} = \frac{(25 \times 2) + (35 \times 6)}{8} = 32{,}5\ ;$$ Reddito medio nelle regioni del nord

$$\mu_2 = \mu_{centro} = \frac{\sum_{j=1}^{2} y_j n_{2j}}{n_{2.}} = \frac{(25 \times 2) + (35 \times 4)}{6} = 31{,}67\ ;$$ Reddito medio nelle regioni del centro

$$\mu_3 = \mu_{sud} = \frac{\sum_{j=1}^{2} y_j n_{3j}}{n_{3.}} = \frac{(25\times 6)+(35\times 0)}{6} = 25$$; Reddito medio nelle regioni del sud e nelle isole

$$\eta^2_{reddito\,|\,area} = \frac{\left[(32{,}5-30)^2\times 8\right]+\left[(31{,}67-30)^2\times 6\right]+\left[(25-30)^2\times 6\right]}{\left[(25-30)^2\times 10\right]+\left[(35-30)^2\times 10\right]} = \frac{216{,}67}{500} = 0{,}43$$

b)

	Classi di consumo alimentare			
AREA	**[6,7 - 7,7]**	**]7,7 - 8.0]**	**]8,0 - 8,9]**	**Totale**
NORD	2	2	4	**8**
CENTRO	0	3	3	**6**
SUD	3	1	2	**6**
Totale	**5**	**6**	**9**	**20**

Considerando che:

$y_1 = 7{,}2$; $y_2 = 7{,}85$; $y_3 = 8{,}45$;

$$\mu = \frac{\sum_{j=1}^{3} y_j n_{.j}}{N} = \frac{(7{,}2\times 5)+(7{,}85\times 6)+(8{,}45\times 9)}{20} = 7{,}96$$

Consumo medio generale

$$\mu_1 = \mu_{nord} = \frac{\sum_{j=1}^{3} y_j n_{1j}}{n_{1.}} = \frac{(7{,}2\times 2)+(7{,}85\times 2)+(8{,}45\times 4)}{8} = 7{,}99$$

Consumo medio nelle regioni del nord

$$\mu_2 = \mu_{centro} = \frac{\sum_{j=1}^{3} y_j n_{2j}}{n_{2.}} = \frac{(7{,}2\times 0)+(7{,}85\times 3)+(8{,}45\times 3)}{6} = 8{,}15$$

Consumo medio nelle regioni del centro

$$\mu_3 = \mu_{sud} = \frac{\sum_{j=1}^{3} y_j n_{3j}}{n_{3.}} = \frac{(7{,}2\times 3)+(7{,}85\times 1)+(8{,}45\times 2)}{6} = 7{,}73$$

Consumo medio nelle regioni del sud e nelle isole

$$\eta^2_{consumo\,|\,area} = \frac{\left[(7{,}99-7{,}96)^2\times 8\right]+\left[(8{,}15-7{,}96)^2\times 6\right]+\left[(7{,}73-7{,}96)^2\times 6\right]}{\left[(7{,}2-7{,}96)^2\times 5\right]+\left[(7{,}85-7{,}96)^2\times 6\right]+\left[(8{,}45-7{,}96)^2\times 9\right]} = \frac{0{,}09}{0{,}83} = 0{,}11$$

ESERCIZIO 4.4

Si verifichi la presenza di indipendenza assoluta per le distribuzioni doppie (esercizio 4.3):

a) AREA GEOGRAFICA / REDDITO
b) AREA GEOGRAFICA / CONSUMI ALIMENTARI

Soluzione

Si tratta di calcolare per entrambe le distribuzioni l'indice del χ^2:

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

a)

Frequenze osservate:

n_{ij}	Classi di REDDITO		Totale
AREA	**20-30**	**30-40**	
NORD	2	6	**8**
CENTRO	2	4	**6**
SUD	6	0	**6**
Totale	**10**	**10**	**20**

Frequenze teoriche:

$\hat{n}_{ij}$	Classi di REDDITO		Totale
AREA	**20-30**	**30-40**	
NORD	4	4	**8**
CENTRO	3	3	**6**
SUD	3	3	**6**
Totale	**10**	**10**	**20**

Addendi della sommatoria:

$\frac{\left(n_{ij} - \hat{n}_{ij}\right)^2}{\hat{n}_{ij}}$	Classi di REDDITO		Totale $= \chi^2$
AREA	**20-30**	**30-40**	
NORD	1,00	1,00	
CENTRO	0,33	0,33	
SUD	3,00	3,00	
Totale = χ^2			**8,67**

χ^2 = 8,67

$$\Phi^2 = \frac{\chi^2}{N} = \frac{8,67}{20} = 0,43$$

Indice normalizzato: $\frac{\phi^2}{min(r-1; c-1)} = \frac{0,43}{1} = 0,43$

b)

Frequenze osservate:

n_{ij}	Classi di consumo alimentare			Totale
AREA	**[6,7 - 7,7]**	**]7,7 - 8.0]**	**]8,0 - 8,9]**	
NORD	2	2	4	**8**
CENTRO	0	3	3	**6**
SUD	3	1	2	**6**
Totale	**5**	**6**	**9**	**20**

Frequenze teoriche:

$\hat{n}_{ij}$	Classi di consumo alimentare			Totale
AREA	**[6,7 - 7,7]**	**]7,7 - 8.0]**	**]8,0 - 8,9]**	
NORD	2	2,4	3,6	**8**
CENTRO	1,5	1,8	2,7	**6**
SUD	1,5	1,8	2,7	**6**
Totale	**5**	**6**	**9**	**20**

Addendi della somma:

$\frac{(n_{ij}-\hat{n}_{ij})^2}{\hat{n}_{ij}}$	Classi di consumo alimentare			Totale = χ^2
AREA	**[6,7 - 7,7]**	**]7,7 - 8.0]**	**]8,0 - 8,9]**	
NORD	0	0,07	0,04	
CENTRO	1,50	0,80	0,03	
SUD	1,50	0,36	0,18	
Totale = χ^2				**4,48**

$\chi^2 = 4,48$

$$\Phi^2 = \frac{\chi^2}{N} = \frac{4,48}{20} = 0,22$$

Indice normalizzato: $\frac{\phi^2}{min(r-1; c-1)} = \frac{0,22}{2} = 0,11$

Nota: si vedano i risultati dell'η^2.

ESERCIZIO 4.5

Su un campione di 80 studenti che hanno conseguito il diploma di maturità classica sono stati osservati il voto di licenza liceale ed il giudizio ottenuto alla licenza media, i cui valori hanno originato la seguente distribuzione doppia.

	Voto di maturità					Totale
Giudizio licenza media	**36-40**	**41-45**	**46-50**	**51-55**	**56-60**	
Sufficiente	10	6	2	2	0	**20**
Buono	4	5	14	2	0	**25**
Distinto	1	9	9	3	3	**25**
Ottimo	0	0	5	3	2	**10**
Totale	**15**	**20**	**30**	**10**	**5**	**80**

Verificare:

a) se il voto di maturità dipende in senso assoluto dal giudizio conseguito all'esame di licenza media;
b) se il voto di maturità dipende in media dal giudizio conseguito all'esame di licenza media.

Soluzione

a)

Calcolo dell'indice χ^2:

Frequenze teoriche:

$\hat{n}_{ij}$	Voto di maturità					Totale
Giudizio licenza media	**36-40**	**41-45**	**46-50**	**51-55**	**56-60**	
Sufficiente	3,75	5,00	7,50	2,50	1,25	**20**
Buono	4,69	6,25	9,38	3,13	1,56	**25**
Distinto	4,69	6,25	9,38	3,13	1,56	**25**
Ottimo	1,88	2,50	3,75	1,25	0,63	**10**
Totale	**15**	**20**	**30**	**10**	**5**	**80**

Addendi della sommatoria:

$\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$	Voto di maturità					Totale $= \chi^2$
Giudizio licenza media	**36-40**	**41-45**	**46-50**	**51-55**	**56-60**	
Sufficiente	10,42	0,20	4,03	0,10	1,25	
Buono	0,10	0,25	2,28	0,41	1,56	
Distinto	2,90	1,21	0,02	0,01	1,32	
Ottimo	1,88	2,50	0,42	2,45	3,03	
Totale $= \chi^2$						**36,32**

$\chi^2 = 36,32$

$$\Phi^2 = \frac{\chi^2}{N} = \frac{36{,}32}{80} = 0{,}45$$

Indice normalizzato:

$$\frac{\phi^2}{min(r-1; c-1)} = \frac{0{,}45}{3} = 0{,}15$$

b)

Calcolo dell'indice η^2

	Voto di maturità					Totale
Giudizio licenza media	**36-40**	**41-45**	**46-50**	**51-55**	**56-60**	
Sufficiente	10	6	2	2	0	**20**
Buono	4	5	14	2	0	**25**
Distinto	1	9	9	3	3	**25**
Ottimo	0	0	5	3	2	**10**
Totale	**15**	**20**	**30**	**10**	**5**	**80**

$y_1 = 38$; $y_2 = 43$; $y_3 = 48$; $y_4 = 53$; $y_5 = 58$;

$$\mu = \frac{\sum_{j=1}^{5} y_j n_{.j}}{N} = \frac{(38\times 15)+(43\times 20)+(48\times 30)+(53\times 10)+(58\times 5)}{80} = 46{,}125$$

Voto medio generale

$$\mu_1 = \mu_{sufficiente} = \frac{\sum_{j=1}^{5} y_j n_{1j}}{n_{1.}} = \frac{(38\times 10)+(43\times 6)+(48\times 2)+(53\times 2)+(58\times 0)}{20} = 42$$

Voto medio di chi ha avuto sufficiente

$$\mu_2 = \mu_{buono} = \frac{\sum_{j=1}^{5} y_j n_{2j}}{n_{2.}} = \frac{(38\times 4)+(43\times 5)+(48\times 14)+(53\times 2)+(58\times 0)}{25} = 45{,}8$$

Voto medio di chi ha avuto buono

$$\mu_3 = \mu_{distinto} = \frac{\sum_{j=1}^{5} y_j n_{3j}}{n_{3.}} = \frac{(38\times 1)+(43\times 9)+(48\times 9)+(53\times 3)+(58\times 3)}{25} = 47{,}6$$

Voto medio di chi ha avuto distinto

$$\mu_4 = \mu_{ottimo} = \frac{\sum_{j=1}^{5} y_j n_{4j}}{n_{4.}} = \frac{(38\times 0)+(43\times 0)+(48\times 5)+(53\times 3)+(58\times 2)}{10} = 51{,}5$$

Voto medio di chi ha avuto ottimo

$$\eta^2_{voto\ |\ giudizio} = \frac{dev_{est}}{dev(Y)} =$$

$$= \frac{\left[(42-46{,}125)^2 \times 20\right] + \left[(45{,}8-46{,}125)^2 \times 25\right] + \left[(47{,}6-46{,}125)^2 \times 25\right] + \left[(51{,}5-46{,}125)^2 \times 10\right]}{\left[(38-46{,}125)^2 \times 15\right] + \left[(43-46{,}125)^2 \times 20\right] + \left[(48-46{,}125)^2 \times 30\right] + \left[(53-46{,}125)^2 \times 10\right] + \left[(58-46{,}125)^2 \times 5\right]} =$$

$$= \frac{686{,}25}{2.468{,}75} = 0{,}278$$