Università di Cassino

Esercitazioni di Statistica 1 del 12 Marzo 2010

Dott. Mirko Bevilacqua

ESERCIZIO n° 1

Sono stati intervistati i laureati di una Facoltà di Economia e dalle risposte ai quesiti è stata costruita la seguente tabella doppia di frequenze

	Tempo (in mesi) per trovare lavoro dopo la laurea				
Residenza	**6**	**12**	**18**	**24**	**Totale**
Nord-ovest	3	2	0	0	5
Nord –est	1	2	1	0	4
Centro	45	26	11	23	105
Sud	9	5	2	5	21
Totale	58	35	14	28	135

Calcolare l'indice Chi-quadrato di Pearson e l'indice di Fisher e commentare i risultati ottenuti.

<u>Soluzione</u>:

La misura dell'associazione tra due caratteri qualitativi sconnessi avviene tramite l'indice Chi-quadrato di Pearson:

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

Nella tabella che segue sono rappresentate le frequenze teoriche $\hat{n}_{ij} = \frac{n_{i.} \times n_{.j}}{N}$

Residenza \ **Tempo** *(in mesi)*	**6**	**12**	**18**	**24**	**Totale**
Nord – ovest	2,1	1,3	0,5	1,0	5
Nord – est	1,7	1,0	0,4	0,8	4
Centro	45,1	27,2	10,9	21,8	105
Sud	9,0	5,4	2,2	4,4	21
Totale	58	35	14	28	135

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(n_{ij} - \hat{n}_{ij}\right)^2}{\hat{n}_{ij}} = \frac{(3-2,1)^2}{2,1} + \frac{(2-1,3)^2}{1,3} + \ldots\ldots + \frac{(2-2,2)^2}{2,2} + \frac{(5-4,4)^2}{4,4} = 5,396$$

Indice di Fisher: $$\phi^2 = \frac{\chi^2}{n} = \frac{5,396}{135} = 0,04$$;

ossia, quasi assenza di associazione tra i caratteri: il tempo occorso per trovare lavoro dopo la laurea non dipende dal luogo di residenza.

ESERCIZIO n° 2

Nella seguente tabella è data la distribuzione congiunta secondo il sesso, il peso e la statura (in cm) di un collettivo di individui.

Statura	Peso maschi			Peso femmine		
	40-45	**50- 75**	**75-90**	**40-45**	**50- 75**	**75-90**
160 -165	6	27	10	25	21	1
165 -170	3	34	13	10	33	5
170-175	1	29	37	5	18	5
175-180	0	9	52	1	9	8

Si calcolino covarianza e correlazione del carattere statura sia con il peso dei maschi che con quello delle donne. Commentare i risultati ottenuti.

Soluzione:

a) Calcolo covarianza e correlazione del carattere altezza relativamente al carattere "Peso maschi"

Statura \ Peso maschi	**42,5**	**62,5**	**82,5**	**Tot.**
162,5	6	27	10	43
167,5	3	34	13	50
172,5	1	29	37	67
177,5	0	9	52	61
Tot.	10	99	112	221

- **Medie e Varianze per i caratteri statura e peso in riferimento al collettivo di sesso maschile:**
  (i simboli PM e SM indicano, rispettivamente, i caratteri peso dei maschi e statura maschi)

$$\mu_{PM} = \frac{(42,5 \cdot 10) + ... + (82,5 \cdot 112)}{221} = 71,7$$

$$\mu_{SM} = \frac{(162,5 \cdot 43) + ... + (177,5 \cdot 61)}{221} = 170,8$$

$$\sigma^2_{SM} = \mu_{SM^2} - (\mu_{SM})^2 = 29202,86 - (170,8)^2 = 29,1$$
$$\sigma_{SM} = \sqrt{29,1} = 5,4$$

$$\sigma^2_{PM} = \mu_{PM^2} - (\mu_{PM})^2 = 5280,9 - (71,7)^2 = 135,6$$
$$\sigma_{PM} = \sqrt{135,6} = 11,6$$

- **Covarianze tra i caratteri statura e peso per il sesso maschile**

SM \ PM	**42,5**	**62,5**	**82,5**	**Tot.**
162,5	41438	274219	134063	449719
167,5	21356	355938	179644	556938
172,5	7331	312656	526556	846544
177,5	0	99844	761475	861319
				2714519

$$\mu_{SM \cdot PM} = \frac{1}{n}\sum_i \sum_j SM_i \cdot PM_j \cdot n_{ij} = \frac{2714519}{221} = 12282{,}9$$

$$cov(SM;PM) = \mu_{SM \cdot PM} - \mu_{PM} \cdot \mu_{SM} = 12282{,}9 - (71{,}7 \cdot 170{,}8) = 31{,}05$$

- **Correlazione tra i caratteri Statura e altezza in riferimento al sesso maschile:**

$$\rho_{SM,PM} = \frac{cov(PM;SM)}{\sigma_{SM} \cdot \sigma_{PM}} = \frac{31{,}05}{5{,}4 \cdot 11{,}6} = 0{,}49$$

b) Calcolo covarianza e correlazione del carattere altezza relativamente al carattere "Peso maschi"

Statura \ Peso femmine	**42,5**	**62,5**	**42,5**	**Tot.**
162,5	25	21	1	47
167,5	10	33	5	48
172,5	5	18	5	28
177,5	1	9	8	18
Tot.	41	81	19	141

- **Medie e Varianze per i caratteri statura e peso in riferimento al collettivo di sesso femminile** (i simboli PF e SF indicano, rispettivamente, i caratteri peso-femmine e statura-Femmine)

$$\mu_{PF} = \frac{(42{,}5 \cdot 41) + \ldots\ldots + (82{,}5 \cdot 19)}{141} = 59{,}4$$

$$\mu_{SF} = \frac{(162{,}5 \cdot 47) + \ldots\ldots + (177{,}5 \cdot 18)}{141} = 168{,}1$$

$$\sigma^2_{SF} = \mu_{SF^2} - (\mu_{SF})^2 = 28284{,}3 - (168{,}1)^2 = 25{,}7$$

$$\sigma_{SF} = \sqrt{25{,}7} = 5{,}1$$

$$\sigma^2_{PF} = \mu_{PF^2} - (\mu_{PF})^2 = 3686,4 - (59,4)^2 = 160,5$$

$$\sigma_{PF} = \sqrt{160,5} = 12,7$$

- **Covarianze tra i caratteri statura e peso per il sesso femminile**

SF \ PF	**42,5**	**62,5**	**82,5**	**Tot.**
162,5	172656	213281	13406	399344
167,5	71188	345469	69094	485750
172,5	36656	194063	71156	301875
177,5	7544	99844	117150	224538
				1411506

$$\mu_{SF \cdot PF} = \frac{1}{n}\sum_i \sum_j SF_i \cdot PF_j \cdot n_{ij} = \frac{1411506}{141} = 10011$$

$$cov(SF;PF) = \mu_{SF \cdot PF} - \mu_{PF} \cdot \mu_{SF} = 10011 - (59,4 \cdot 168,1) = 28,8$$

- **Correlazione tra i caratteri Statura e altezza in riferimento al sesso femminile:**

$$\rho_{SF,PF} = \frac{cov(PF;SF)}{\sigma_{SF} \cdot \sigma_{PF}} = \frac{28,8}{5,1 \cdot 12,7} = 0,45$$

ESERCIZIO n° 3

Con riferimento alla seguente distribuzione di 5 famiglie secondo il reddito e il consumo medio mensile misurati in migliaia di euro:

Reddito	5	6	8	3	6
Consumo	4	2	5	2	3

a) Calcolare i parametri della retta di regressione del consumo Y sul reddito X
b) Dalla relazione trovata si può affermare che mediamente circa metà del reddito di una famiglia finisce in consumi?
c) La retta spiega più del 50% della variabilità totale?
d) Quale sarebbe il consumo di una famiglia che guadagna 4 mila euro?

Soluzione:

a) Calcolo dei coefficienti di regressione

Reddito (R)	Consumo (C)	R · C	R^2	C^2
5	4	20	25	16
6	2	12	36	4
8	5	40	64	25
3	2	6	9	4
6	3	18	36	9
		96	170	58

- **Reddito e consumo medio delle 5 famiglie**

$$\mu_C = \frac{4+2+5+2+3}{5} = 3,2$$

$$\mu_R = \frac{5+6+8+3+6}{5} = 5,6$$

- **Varianze di Reddito e Consumo**

$$\sigma_R^2 = \mu_{R^2} - (\mu_R)^2 = \frac{170}{5} - (5,6)^2 = 2,64$$
$$\sigma_R = \sqrt{2,64} = 1,6$$

$$\sigma_C^2 = \mu_{C^2} - (\mu_C)^2 = \frac{58}{5} - (3,2)^2 = 1,36$$
$$\sigma_C = \sqrt{1,36} = 1,2$$

- **Covarianza Reddito – Consumo**

$$\mu_{C \cdot R} = \frac{96}{5} = 19,2$$
$$cov(C;R) = \mu_{C \cdot R} - \mu_C \cdot \mu_R = 19,2 - (3,2 \cdot 5,6) = 1,28$$

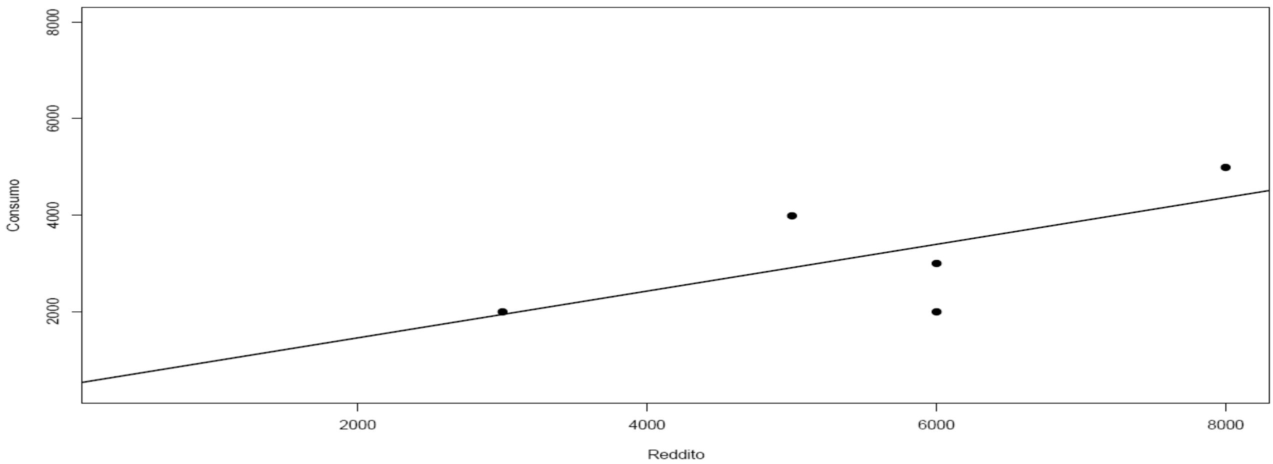
- **Stima dei Coefficiente di Regressione**

$$b = \frac{cov(C,R)}{var(R)} = \frac{1,28}{2,64} = 0,48$$

$$a = \mu_C - b \cdot \mu_R = 3,2 - (0,48 \cdot 5,6) = 0,48$$

- **Retta di Regressione.**

$$\hat{C}_i = 0,48 + 0,48 \cdot R_i$$



b) Si, essendo b=0,48, circa metà del reddito finisce in consumo.

c) Calcolo del coefficiente di determinazione R^2

$$R^2 = \left(\frac{cov(C,R)}{\sigma_C \cdot \sigma_R} \right)^2 = \left(\frac{1,28}{1,6 \cdot 1,2} \right)^2 = 0,456$$

No, il modello spiega soltanto il 45,6% della variabilità totale.

d) Stima del consumo per una famiglia che guadagna 4.000 euro, attraverso il modello di regressione

$$\hat{C}_i = 0,48 + 0,48 \cdot R_i$$

$$\hat{C}_i = 0,48 + 0,48 \cdot 4 = 2,42$$

Il modello stima in 2.420 euro il consumo per una generica famiglia che ha reddito pari a 4.000 euro.