Università di Cassino
Corso di Statistica 1
Esercitazione del 28/11/2006
Dott. Alfonso Piscitelli

Esercizio 1

La tabella che segue riporta la durata in anni di 125 fotografie stampate su tre diversi supporti fotografici.

	Durata			
Supporti	10 anni	12 anni	13 anni	Totale
Supporto A	15	16	14	**45**
Supporto B	12	14	13	**39**
Supporto C	14	14	13	**41**
Totale	**41**	**44**	**40**	**125**

Si studi la relazione statistica tra il supporto e la durata della stampa.

L'indice più opportuno per lo studio della relazione tra una variabile qualitativa ed una quantitativa, dove la variabile quantitativa dipende da quella qualitativa, è l'indice η^2.

$$\eta^2 = \frac{Dev.(B)}{Dev.(T)}$$

Il rapporto è dato dalla *devianza tra i gruppi (Between)*, indicata con Dev.(B) sulla *devianza totale* indicata con Dev.(T).
Se la variabile quantitativa dipendente è posta in colonna valgono le seguenti formule:

$$Dev.(B) = \sum_{i=1}^{R} (M(Y|x_i) - M(y))^2 n_{i.} \qquad Dev.(W) = \sum_{i=1}^{R} [\sum_{j=1}^{C} (y_j - M(Y|x_i))^2 n_{ij}]$$

$$Dev.(T) = \sum_{j=1}^{C} (y_j - M(y))^2 n_{.j}$$

Calcolo della media Generale:

$$M(y) = \frac{1}{N} \sum_{j=1}^{C} y_j n_{.j}$$

y_j	$n_{.j}$	$(y_j * n_{.j})$
10	41	410
12	44	528
13	40	520
	125	1458

$$M(y)=\frac{1458}{125}=11,66$$

Calcolo delle medie parziali:

$$M(Y|x_i)=\sum_{j=1}^{C}\frac{y_j n_{ij}}{n_{i.}}$$

nel nostro esercizio sono tre: la media del supporto A, la media del supporto B, la media del supporto C

y_j	n_{ij}	$(y_j * n_{ij})$
10	15	150
12	16	192
13	14	182
	45	524

y_j	$n_{.j}$	$(y_j * n_{.j})$
10	12	120
12	14	168
13	13	169
	39	457

y_j	$n_{.j}$	$(y_j * n_{.j})$
10	14	140
12	14	168
13	13	169
	41	477

$$M(Y|x_A)=\frac{524}{45}=11,64 \qquad M(Y|x_B)=\frac{457}{39}=11,72 \qquad M(Y|x_C)=\frac{477}{41}=11,63$$

Calcolo della Devianza Totale:

	y_j	$n_{.j}$	$(y_j-M(y))^2 n_{.j}$
	10	41	112,98
	12	44	5,09
	13	40	71,82
Totale (*n*)		125	
Devianza Totale			189,89

Calcolo della Devianza Between:

	$M(Y\|x_i)$	$n_{i.}$	$(M(Y\|x_i)-M(y))^2 n_{i.}$
	11,64	45	0,018
	11,72	39	0,140
	11,63	41	0,037
Totale (*n*)		125	
Devianza Between			0,195

$$\eta^2 = \frac{0{,}195}{189{,}89} = 0{,}001$$

Si può affermare che vi è una indipendenza, in media, della durata dal tipo di supporto.

Anche se non è necessario ai fini dell'esercizio, possiamo effettuare un **controllo sui calcoli** delle devianze utilizzando il "Teorema di decomposizione della devianza"

Dev.(T) = Dev.(B) + Dev.(W)

Calcolo della Devianza Within:

	y_j	n_{ij}	$(y_j - M(Y \vert x_i))^2 n_{ij}$
	10	15	40,34
	12	16	2,07
	13	14	25,89
$n_{i.}$		45	
Devianza Gruppo A			68,3

	y_j	n_{ij}	$(y_j - M(Y \vert x_i))^2 n_{ij}$
	10	12	35,5
	12	14	1,1
	13	13	21,3
$n_{i.}$		39	
Devianza Gruppo B			57,9

	y_j	n_{ij}	$(y_j - M(Y \vert x_i))^2 n_{ij}$
	10	14	37,2
	12	14	1,9
	13	13	24,4
$n_{i.}$		41	
Devianza Gruppo C			63,5

Dev.(W) = 68,3 + 57,9 + 63,5 = 189,7

Dev.(T) = Dev.(B) + Dev.(W) = 0,195 + 189,7 = 189,895

Esercizio 2

Nella tabella seguente sono riportati sia i costi della pubblicità che le vendite effettuate da un'azienda in otto mesi di attività.

Pubblicità	Vendite
1	30
3	40
5	40
4	50
2	35
5	50
3	35
2	25

Verificare l'esistenza di una relazione statistica tra le due variabili.

Considerando che entrambe le variabili sono di natura quantitativa, l'indice più opportuno per verificare l'esistenza di una relazione tra le variabili è il coefficiente di correlazione lineare.

Il coefficiente di correlazione è una misura dell'INTERDIPENDENZA lineare tra due fenomeni e varia tra -1 ed +1 (perfetta correlazione inversa, perfetta correlazione diretta).

Il coefficiente di correlazione di Bravais-Pearson e dato da:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\frac{1}{S}\sum_{i=1}^{S}(x_i - \mu_X)(y_i - \mu_Y)}{\sigma_X \sigma_Y}$$

Il numeratore si chiama **Covarianza** perché è una misura della contemporanea variazione di X e Y in rapporto alle rispettive medie.

Mesi (**s**)	X	Y	XY	X^2	Y^2
1	1	30	30	1	900
2	3	40	120	9	1600
3	5	40	200	25	1600
4	4	50	200	16	2500
5	2	35	70	4	1225
6	5	50	250	25	2500
7	3	35	105	9	1225
8	2	25	50	4	625
Totale	**25**	**305**	**1025**	**93**	**12175**

Si inizia con il calcolare la media per le due variabili

μ x=25/8=3,125

μ y=305/8=38,125

Cov(X,Y)= $\frac{1}{S}\sum_{i=1}^{S} x_i y_i - S\mu_X \mu_Y = \frac{1025\text{ - }8*3,125*38,125}{8} = \frac{1025-953,125}{8} = \frac{71,875}{8}$ =

=8,9843

$$\sigma_X = \sqrt{\mu_{X^2} - \mu_X^2} = \sqrt{11,625-9,765} =$$ 1,3635

$$\sigma_Y = \sqrt{\mu_{Y^2} - \mu_Y^2} = \sqrt{1521,875-1453,516} =$$ 8,2679

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{8,9843}{1,3635*8,2679} = 0,7969$$

Siamo in presenza di un’alta correlazione positiva tra i costi della pubblicità e le vendite effettuate.