

ESERCIZIO

La tabella indagine *StudentiFlorida*, disponibile sul sito (in formato pdf e xls), riporta i dati di un'indagine condotta su 60 studenti laureati presso la University of Florida, relativamente alle seguenti variabili:

COLONNA	NOME VARIABILE	DESCRIZIONE
1	id	Codice identificativo dello studente
2	genere	Genere dello studente
3	età	Età (in anni compiuti)
4	punteggioHighSchool	Punteggio ottenuto alla scuola superiore (scala da 1 a 4)
5	punteggioCollege	Punteggio ottenuto al College (scala da 1 a 4)
6	distanzaCampus	Distanza (in miglia) del campus dalla città di provenienza dello studente
7	distanzaAula	Distanza (in miglia) della classe dall'attuale luogo di residenza
8	oreTVsettimana	Numero medio di ore settimanali passate a guardare la TV
9	oreSportSettimana	Numero medio di ore settimanali dedicate allo sport
10	quotidianiPerSettimana	Numero di volte per settimana che si legge un quotidiano
11	AH	Numero di persone conosciute morte per AIDS o sieropositive
12	vegetariano	Regime di alimentazione vegetariano (yes, no)
13	affiliazionePolitica	Affiliazione politica (D = Democratico, R = Repubblicano, I = Indipendente)
14	ideologiaPolitica	Ideologia politica (1 = molto liberale, 2 = liberale, 3 = leggermente liberale, 4 = moderato, 5 = abbastanza conservatore, 6 = conservatore, 7 = molto conservatore)
15	religiosità	Partecipazione ad eventi religiosi (0 = mai, 1 = occasionalmente, 2 = la maggior parte delle settimane, 3 = ogni settimana)
16	opinioneAborto	Opinione circa la legalità dell'aborto nei primi tre mesi di gravidanza (yes, no)
17	azioniDiscriminazione	Impegno in azioni contro le discriminazioni (si, no)
18	vitaMorte	Crede nella vita dopo la morte (si -y, incerto - u, no - n)

Fonte: A. Agresti and B. Finlay, *Statistical Methods for the Social Science*, Prentice Hall, 2008

1) A partire dalle seguenti informazioni sulle variabili *affiliazione politica* e *nr di quotidiani per settimana*, calcolare l'indice di dipendenza in media η^2 usando la variabile *nr. di quotidiani per settimana* come variabile di risposta e la variabile *affiliazione politica* come variabile esplicativa:

Affiliazione politica	Frequenza	nr. medio di quotidiani per settimana	Devianza
d	21	4.429	283.143
i	24	3.708	90.958
r	15	4.200	152.400
tutti gli intervistati	60	4.083	532.58

2) A partire dalle seguenti statistiche calcolate per le variabili **età (X)** e **quotidiani per settimana (Y)**, calcolare la covarianza e il coefficiente di correlazione lineare tra le due variabili:

STATISTICA	VALORE
$\sum_i x_i$	1750
$\sum_i y_i$	245
$\sum_i x_i^2$	55280
$\sum_i y_i^2$	1533
$\sum_i (x_i - \mu_X)^2$	4238.33
$\sum_i (y_i - \mu_Y)^2$	532.58
$\sum_i (x_i - \mu_X)(y_i - \mu_Y)$	-15.83
$\sum_i x_i y_i$	7130
Covarianza	
Coefficiente di correlazione lineare	

3) Dimostrare che:

$$\sum_i \frac{(x_i - \mu_X)(y_i - \mu_Y)}{N} = \frac{\sum_i x_i y_i}{N} - \mu_X \mu_Y$$

--	--

4) A partire dalle seguenti statistiche calcolate per le variabili **età (X)** e **quotidiani per settimana (Y)**, calcolare i coefficienti della retta di regressione in cui l'età viene usata come variabile esplicativa e il numero di quotidiani letti per settimana come variabile di risposta:

STATISTICA	VALORE
$\sum_i x_i$	1750
$\sum_i y_i$	245
$\sum_i x_i^2$	55280
$\sum_i y_i^2$	1533
$\sum_i (x_i - \mu_X)^2$	4238.33
$\sum_i (y_i - \mu_Y)^2$	532.58
$\sum_i (x_i - \mu_X)(y_i - \mu_Y)$	-15.83
$\sum_i x_i y_i$	7130

β_0	
β_1	

Calcolare l'indicatore R² usando le differenti formule possibili:

$R^2 = \frac{\sum_i (\hat{y}_i - \mu_Y)^2}{\sum_i (y_i - \mu_Y)^2}$	
$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \mu_Y)^2}$	
$R^2 = \rho_{XY}^2$	
$R^2 = \beta_1^2 \frac{\sum_i (x_i - \mu_X)^2}{\sum_i (y_i - \mu_Y)^2}$	

5) La seguente tabella riporta la distribuzione congiunta delle variabili **punteggioHighSchool** (in 3 classi) e **punteggioCollege** (in 3 classi):

		PUNTEGGIO COLLEGE			TOT.
		0 - 3	3 - 3.5	3.5 - 4	
PUNTEGGIO HIGH SCHOOL	0 - 3	8	6	6	20
	3 - 3.5	3	11	8	22
	3.5 - 4	2	6	10	18
TOT.		13	23	24	60

A partire dalla tabella doppia, calcolare i due coefficienti di regressione e l'indice di bontà di adattamento del modello:

β_0	
β_1	
R^2	