Applied Regression **Home Work 4** **Spring**

Name: _____ **Instructor: D. Kushary**

1. A university placement director is interested in the effect that grade point average (GPA) and the number of university activities listed on the resume might have on the starting salaries of this year's graduating class. He has collected these data for a sample of 10 graduates: (Data is in file xr16054).

   a. Determine the multiple regression equation and interpret the partial regression coefficients.

   b. Dave has a 3.6 grade point average and 3 university activities listed on his resume. What would be his estimated starting salary?

   c. Estimate the standard error for the model.

   d. Determine the 95% prediction interval for the starting salary of the student described in part (b).

   e. Determine the 95% confidence interval for the mean starting salary for all students like the one described in part (b).

   f. Determine the 95% confidence interval for the population partial regression coefficients, β_1 and β_2.

   g. Interpret the significance tests in the computer printout.

   h. Analyze the residuals. Does the analysis support the applicability of the multiple regression model to these data?

2 Consider the part of the data (only first 4 data points) of the above problem: (data below)

Graduate	(Thousands)	Average	Activities
1	40	3.2	2
2	46	3.6	5
3	38	2.8	3
4	39	2.4	4

(a) Write down the multiple linear model for this problem and input the data and write down the equation.

(b) Write down the X-Matrix for this data (4 graduates).

(c) Write down the X' matrix (transpose) for this data.

(d) Write down the $X'X$ matrix for this data.

(e) Write down the Y-vector for this data.

(f) Write down the $X'Y$ vector for this data.

(g) Write down the equation $X'X\beta = X'Y$ vector for this data (β vector is unknown) and solve for β vector.

3. (Problem 4.6 - From Book - Page 184). Earnings of Mexican street vendors. Detailed interviews were conducted with over 1,000 street vendors in the city of Puebla, Mexico, in order to study the factors influencing vendors' incomes (World Development, February 1998). Vendors were defined as individuals working in the street, and included vendors with carts and stands on wheels and excluded beggars, drug dealers, and prostitutes. The researchers collected data on gender, age, hours worked per day, annual earnings, and education level. A subset of these data appears in the accompanying table.

(a) Write a first-order model for mean annual earnings, E(y), as a function of age (x_1) and hours worked (x_2).

(b) Find the least squares prediction equation on the printout shown below.

(c) Interpret the estimated β coefficients in your model.

(d) Conduct a test of the global utility of the model (at $\alpha = 0.01$). Interpret the result.

(e) Find and interpret the value of $adj - R^2$.

(f) Find and interpret s, the estimated standard deviation of the error term.

(g) Is age (x_1) a statistically useful predictor of annual earnings? Test using $\alpha = 0.01$.

(h) Find a 95% confidence interval for β_2. Interpret the interval in the words of the problem.