Stat 332 (Regression Analysis)

Assignment 1: (12 Marks)

Problem 1

Suppose that $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ is a data set to which we fit a simple regression model. Let β_1 be the least square estimate of the model slop and r be the simple correlation coefficient between y_1, \dots, y_n and x_1, \dots, x_n .

- (a) Show that $\hat{\beta}_1 = \frac{s_y}{s_x}r$, where s_y and s_x are the sample standard deviations of y_1, \dots, y_n and x_1, \dots, x_n , respectively.
- (b) Show that $\hat{\beta}_1$ is an unbiased estimate of β_1 .
- (c) Show that $\frac{\hat{\beta}_1}{S.E.(\hat{\beta}_1)} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$
- (d) In the estimated model, prove that the sum of the residuals equal zero and estimated line pass through the point $(\overline{X}, \overline{Y})$.

Problem 2

Tamoxifen is a drug often used to treat cancer patients. One effect of the drug is to change the levels of cortisol-binding globulin (CBG). One study attempted to see if the effect of Tamoxifen depends on the patient's age. The response variable (Y) is the change in CBG and the covariate (X) is the age. The following summary statistics were reported.

$$n = 26, \quad \sum x_i = 1613, \quad \sum y_i = 281.9$$

 $S_{XX} = 3756.96, \quad S_{yy} = 465.34, \quad S_{XY} = -757.64$

- (a) Find the least square estimates of the intercept and slope.
- (b) Give the standard errors for your estimates in (a).
- (c) Construct 95% confidence intervals for the true intercept and slope.
- (d) Discuss the efficiency of the estimated model based on T and F test.
- (e) Calculate the coefficient of determination and interpret the results.

Problem 3

Consider the following data

Х	10	85	20	25	30	35
Y	73	85	90	86	75	61
	78	87	92	87	76	63

(a) Estimate the simple linear regression model

(b) Perform the lack of fit test for the model.

Problem 4

Use the data in (cars) in R to estimate the suitable simple regression line. The find the point and 90% confidence intervals of the overall response variable when the independent variable is equal to 26, 28 and 30.

Problem 5

A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where X is the year (coded) and Y is sales in thousands

i	1	2	3	4	5	6	7	8	9	10
X _i	0	1	2	3	4	5	6	7	8	9
Y _i	98	135	162	178	221	232	283	300	374	395

(a) Fit the simple linear model of the given data.

- (b) Fit the simple linear model under the following transformations $Y' = \sqrt{Y}$, $Y' = \log 10(Y)$ and Y' = 1/Y.
- (c) Compare between the results in (a) and (b) using the coefficient of determination.
- (d) What is the expect Y for each model when X=12?

best wishes

(Dr. Majdi Naji)