Assignment 1: (12 Marks)

## Problem 1

Suppose that $\left(y_{1}, x_{1}\right),\left(y_{2}, x_{2}\right), \ldots .,\left(y_{n}, x_{n}\right)$ is a data set to which we fit a simple regression model. Let $\beta_{1}$ be the least square estimate of the model slop and $r$ be the simple correlation coefficient between $y_{1}, \ldots, y_{n}$ and $x_{1}, \ldots, x_{n}$,.
(a) Show that $\hat{\beta}_{1}=\frac{s_{y}}{s_{x}} r$, where $s_{y}$ and $s_{x}$ are the sample standard deviations of $y_{1}, \ldots, y_{n}$ and $x_{1}, \ldots, x_{n}$, respectively.
(b) Show that $\hat{\beta}_{1}$ is an unbiased estimate of $\beta_{1}$.
(c) Show that $\frac{\widehat{\beta}_{1}}{\text { S.E. }\left(\widehat{\beta}_{1}\right)}=\frac{r \sqrt{n-2}}{\sqrt{1-r^{2}}}$
(d) In the estimated model, prove that the sum of the residuals equal zero and estimated line pass through the point $(\bar{X}, \bar{Y})$.

## 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.

$$
\begin{aligned}
& n=26, \quad \sum x_{i}=1613, \quad \sum y_{i}=281.9 \\
& S_{X X}=3756.96, S_{y y}=465.34, S_{X Y}=-757.64
\end{aligned}
$$

(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

| X | 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

| $\mathbf{i}$ | 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 $Y^{\prime}=\sqrt{Y}, Y^{\prime}=\log 10(Y)$ and $Y^{\prime}=1 / Y$.
(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)

