Stat 332 (Regression Analysis)

Assignment 1: (10 Marks)

Due date: 2/ 4/ 2020 (Time 11:30 pm)

**Problem 1**

Suppose that is a data set to which we fit a simple regression model. Let $\hat{β}\_{1}$ be the least square estimate of the model slope and be the simple correlation coefficient between.

1. Show that$\hat{ β}\_{1} $= $\frac{S\_{y}}{S\_{x}} r$, were $S\_{y} $and are the sample standard deviations of respectively.
2. Show that $\hat{β}\_{1} $is an unbiased estimate of.
3. Show that$ \frac{\hat{ β}\_{1}}{S.E(\hat{ β}\_{1})}= \frac{r \sqrt{n-2}}{\sqrt{1- r^{2}}}$.
4. In the estimated model, prove that the sum of the residuals equal zero and estimated line pass through the point.

**Problem 2**

Tamoxifen is a drug often used to treat cancer patients. One effect of the drug isto 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 is the change in CBG and the covariate is the age. The following summary statistics were reported.



1. Find the least square estimates of the intercept and slope.
2. Give the standard errors for your estimates in (a).
3. Construct 95% confidence intervals for the true intercept and slope.
4. Discuss the efficiency of the estimated model based on T and F test.
5. Calculate the coefficient of determination and interpret the results.

**Problem 3**

Consider the following data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | 10 | 85 | 20 | 25 | 30 | 35 |
| Y | 7378 | 8587 | 9092 | 8687 | 7576 | 6163 |

1. Estimate the simple linear regression model
2. 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



1. Fit the simple linear model of the given data.
2. Fit the simple linear model under the following transformations
*Y'* = $\sqrt{Y}$, *Y'* = $log10⁡(Y)$ and *Y'* = $1/Y$.
3. Compare between the results in (a) and (b) using the coefficient of determination.
4. What is the expect Y for each model when X=12?