Homework No. 03

Deadline - Sunday, April 14

Problem 1.

a) Create a random vector X containing 100 observations drawn from a $N(0,1)$ distribution **(score=3).**

b) Create a vector e containing 100 observations drawn from a $N(0, 0.25)$ distribution **(score=3).**

c) Generate a vector Y according to the model **(score=3)**

$$Y = -1 + 0.5X + e \quad (1)$$

d) Create a scatterplot displaying the relationship between X and Y. Comment on what you observe **(score=3).**

e) Fit a least squares linear model to predict Y using X. Comment on the model obtained (p-values, accuracy measures, …). How do $\hat{\beta}_0$ and $\hat{\beta}_1$ compare to β_0 and β_1 **(score=3)?**

f) Display the least squares line on the scatterplot obtained in d). Draw the population regression line on the plot, in a different color. Create appropriate legends **(score=4).**

g) Fit a polynomial regression model that predicts Y using X and X^2. Is there evidence that the quadratic term improves the model fit? Explain your answer **(score=6).**

h) What are the confidence intervals for β_0 and β_1 based on the original data set, the noisier data set, and the less noisy data set? Comment on your results **(score=5).**

Problem 2. Use "*Auto*" data set.

a) Perform a simple linear regression with "*mpg*" as the response and "*horsepower*" as the predictor **(score=2).** Explain the output:

   I. Is there a relationship between the predictor and the response **(score=2)?**

   II. How strong is the relationship between the predictor and the response **(score=2)?**

   III. What is the predicted "*mpg*" associated with a "*horsepower*" of 98 **(score=2)?**

   IV. What are the associated 95% confidence and prediction intervals **(score=2)?**

b) Plot the response and the predictor. Display the least squares regression line **(score=2).**

c) Produce diagnostic plots. Comment on any problems you see with the fit **(score=4).**

d) Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage **(score=4)?**

Problem 3. Use **"*Carseats*"** data set.

a) Fit a multiple regression model to predict "Sales" using "Price", "Urban", and "US" **(score=2).**

b) For which of the predictors can you reject the null hypothesis $H_0: \beta_j = 0$ **(score=2)?**

c) On the basis of your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome **(score=4).**

d) How well do the models in a) and c) fit the data **(score=2)?**

e) Using the model from c), obtain 95% confidence intervals for the coefficient(s) **(score=2).**

f) Produce diagnostic plots of the linear regression fit. Comment on any problems you see with the fit **(score=4).**

g) Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage **(score=4)?**

Problem 4. Use **"*Auto*"** data set

a) Produce a scatterplot matrix which includes all of the variables in the data set **(score=2).**

b) Compute the matrix of correlations between the variables. You will need to exclude the "*name*" variable, which is qualitative **(score=2).**

c) Perform a multiple linear regression with "*mpg"* as the response and all other variables except "*name*" as the predictors **(score=2**). Comment on the output:

   I. Is there a relationship between the predictors and the response **(score=2)?**

   II. Which predictors appear to have a statistically significant relationship to the response **(score=2)?**

   III. What does the coefficient for the "*year*" variable suggest **(score=2)?**

d) Produce diagnostic plots of the linear regression fit. Comment on any problems you see with the fit **(score=4).**

e) Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage **(score=4)?**

f) Add interaction effects. Do any interactions appear to be statistically significant **(score=5)?**

g) Try a few different transformations of the variables, such as $log(X), \sqrt{X}, X^2$. Comment on your findings **(score=5).**