Due: 07.02.2022 23.59

STAT 363 LINEAR MODELS

2021-2022 ACADEMIC YEAR FALL SEMESTER

HOMEWORK 5

Instructions

* Please, do the questions by yourself. If you need help, ask me.
* Please, write your names and ID numbers on the document you are going to submit.
* Submit your assignments via <https://odtuclass.metu.edu.tr>.
* You will submit your assignments through turnitin.
* Use R and Latex functionalities only by creating an RMarkdown document for your homework. This means no handwriting/photo/image/camscanner output for the mathematical operations in your homework. I sent you a document to show how to write Latex code in R. After you complete your homework in R, just knit and this will give you a Word document or you can get a PDF output as well.
* Solutions provided as an image/photo/camscanner output or in handwriting will not be accepted.
* Submissions will be in PDF or Word format.
* Do not add your R codes in the homework document. Submit them in the other assignment created for that.
* Late submissions will not be accepted.

## Computational Questions (40 pts)

## 1. The concentration of in a tube - flow reactor as a function of several controllable variables is shown in the table below.

data <- MPV::table.b6  
knitr::kable(data)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| y | x1 | x2 | x3 | x4 |
| 0.000450 | 0.0105 | 90.9 | 0.0164 | 0.0177 |
| 0.000450 | 0.0110 | 84.6 | 0.0165 | 0.0172 |
| 0.000473 | 0.0106 | 88.9 | 0.0164 | 0.0157 |
| 0.000507 | 0.0116 | 488.7 | 0.0187 | 0.0082 |
| 0.000457 | 0.0121 | 454.4 | 0.0187 | 0.0070 |
| 0.000452 | 0.0123 | 439.2 | 0.0187 | 0.0065 |
| 0.000453 | 0.0122 | 447.1 | 0.0186 | 0.0071 |
| 0.000426 | 0.0122 | 451.6 | 0.0187 | 0.0062 |
| 0.001215 | 0.0123 | 487.8 | 0.0192 | 0.0153 |
| 0.001256 | 0.0122 | 467.6 | 0.0192 | 0.0129 |
| 0.001145 | 0.0094 | 95.4 | 0.0163 | 0.0354 |
| 0.001085 | 0.0100 | 87.1 | 0.0162 | 0.0342 |
| 0.001066 | 0.0101 | 82.7 | 0.0162 | 0.0323 |
| 0.001111 | 0.0099 | 87.0 | 0.0163 | 0.0337 |
| 0.001364 | 0.0110 | 516.4 | 0.0190 | 0.0161 |
| 0.001254 | 0.0117 | 488.4 | 0.0189 | 0.0149 |
| 0.001396 | 0.0110 | 534.5 | 0.0189 | 0.0163 |
| 0.001575 | 0.0104 | 542.3 | 0.0189 | 0.0164 |
| 0.001615 | 0.0067 | 98.8 | 0.0163 | 0.0379 |
| 0.001733 | 0.0066 | 84.8 | 0.0162 | 0.0360 |
| 0.002753 | 0.0044 | 69.6 | 0.0163 | 0.0327 |
| 0.003186 | 0.0073 | 436.9 | 0.0189 | 0.0263 |
| 0.003227 | 0.0078 | 406.3 | 0.0192 | 0.0200 |
| 0.003469 | 0.0067 | 447.9 | 0.0192 | 0.0197 |
| 0.001911 | 0.0091 | 58.5 | 0.0164 | 0.0331 |
| 0.002588 | 0.0079 | 394.3 | 0.0177 | 0.0674 |
| 0.002635 | 0.0068 | 461.0 | 0.0174 | 0.0770 |
| 0.002725 | 0.0065 | 469.2 | 0.0173 | 0.0780 |

#### a. Fit a multiple regression model relating concentration of (y) to concentration of , (x1) and mole fraction (x4).

#### b. Test for significance of regression.

#### c. Calculate and for this model.

#### d. Using t tests, determine the contribution of x1 and x4 to the model. Are both regressors x1 and x4 necessary?

#### e. Is multicollinearity a potential concern in this model?

#### f. Construct a normal probability plot of the residuals. Does there seem to be any problem with the normality assumption?

#### g. Construct and interpret a plot of the residuals versus the predicted response.

#### h. Construct the partial regression plots for this model. Does it seem that some variables currently in the model are not necessary?

#### i. Perform a thorough influence analysis. Discuss your results.

## 2. Consider the pressure drop data in the following table.

data <- MPV::table.b9

#### a. Perform a thorough residual analysis of these data.

#### b. Identify the most appropriate transformation for these data. Fit this model and repeat the residual analysis.

## 3. Consider the automobile gasoline mileage data.

data <- MPV::table.b3

#### a. Build a linear regression model relating gasoline mileage y to vehicle weight x10 and the type of transmission x11 . Does the type of transmission significantly affect the mileage performance?

#### b. Modify the model developed in part a to include an interaction between vehicle weight and the type of transmission. What conclusions can you draw about the effect of the type of transmission on gasoline mileage? Interpret the parameters in this model.

## 4. The carbonation level of a soft drink beverage is affected by the temperature of the product and the filler operating pressure. Twelve observations were obtained and the resulting data are shown below.

data <- data.frame(carbonation=c(2.6, 2.4, 17.32, 15.6,16.12,5.36,6.19,10.17,2.62,2.98,6.92,7.06), temperature=c(31,31,31.5,31.5,31.5,30.5,31.5,30.5,31,30.5,31,30.5), pressure=c(21,21,24,24,24,22,22,23,21.5,21.5,22.5,22.5))

### a. Fit a second - order polynomial model.

#### b. Test for significance of regression.

#### c. Does the interaction term contribute significantly to the model?

#### d. Do the second-order terms contribute significantly to the model?

#### e. Compute the residuals from the second - order model. Analyze the residuals and comment on the adequacy of the model.

## Proof Questions (20 points)

### a. Let and represent the slopes in the regression of Y on X and X on Y, respectively. Show that , where r is the coefficient of correlation between X and Y.

### b. Show that studentized residuals are generally larger than the corresponding standardized residuals.

### c. Consider the following model. . Show that the coefficient of simple determination between Y and equals the coefficient of multiple determination R2.

### d. Show that SSR(X1, X2, X3, X4 ) = SSR(X1) + SSR(X2,X3|X1) + SSR(X4|X1,X2,X3)

### e. Show that residuals are not independent.

## Reasoning Questions - Give short answers with your reasoning. (20 pts)

### a. What is the difference between the population and sample regression functions? Is this a distinction without difference?

### b. Why do we need to check the signs of estimated coefficients?

### c. How and under which circumstances can you decide an independent variable is the one that affects the response variable the most? What are the possible problems that makes it difficult to come to a conclusion like this?

### d. How can we use regression analysis as ANOVA?

### e. In your own words, state the steps of regression analysis. Where do we start? Where do we finish analysis?

## Research Question (20 pts)

influence.measures function in R gives you a list. In it, there is a matrix that tells you whether an observation is influential or not. Do a research on the cutoff values used by this function and comment on the possible problems that might arise when using built-in functions.