TAMS 65 Assignment 6
Multiple linear regression - Backward elimination

This assignment deals with multiple linear regression as well as Backward elimination.

Instruction

- You are recommended to use software MATLAB to answer questions.
- Attach the project in **pdf file** and name it as **Project.pdf**.
- Make a detailed report on **2 assignments**.
- Give only solutions to the rest of **5 assignments**. Note: I have marked this part of questions in **blue**. That is, you only need to show solutions to the blue part of questions for the rest of 5 assignments.
- **Submit first version** of your report to your **teaching assistants** not later than May 1, 2020.
- **Submit final version** of your report to the lisam: **Lisam - Submissions**.
- **Deadline** for submissions is at 23:00 May 15, 2020. Note: The submission entrance will open at 0:00 May 7, 2020.
- **All codes** that you will need are given either in Lectures or 7 assignments.

Backward elimination

People wants to analyze the chemical yield for an industrial process, but they are not sure which explanatory variables they should use in regression model. In a feasibility study with 30 observations, it is desirable to study which variables should be included in the model. Let

$Y =$ measure of chemical yield,
$x_1 =$ amount of catalyst,
$x_2 = \begin{cases} 1, & \text{preprocessing 2;} \\ 0, & \text{otherwise} \end{cases}$
$x_3 = \begin{cases} 1, & \text{preprocessing 3;} \\ 0, & \text{otherwise} \end{cases}$
$x_4 =$ humidity i %,
$x_5 =$ percentage of oxygen in the surrounding environment,
$x_6 =$ time in seconds for the process,
$x_7 =$ square of time of process,
$x_8 =$ temperature i $^\circ C$.

Download and open the file **Assignment6.m**, then **run** it. Input the codes in the **Command Window** or **Editor** window.

Questions

(a) Scatter plot y against $x_i, i = 1, \ldots, 8$ and calculate their correlations.

(b) Perform a regression analysis with all 8 variables. Calculate the coefficient of determination R^2 and do a residual analysis.

(c) Propose a model by applying backward elimination. Calculate the coefficient of determination R^2 and do a residual analysis.

(d) Compare your proposed model in (c) with the full model using all 8 explanatory variables. Is the full model significantly better?

Steps of Backward Elimination

Step 1: Choose a significance level, for example $\alpha = 5\%$.

Step 2: Analyze the model with all possible explanatory variables.

Step 3: Choose the explanatory variable with biggest p-value, such that

- If p-value $\geq \alpha$, go to Step 4.
- If not, **stop!** We get the model.

Step 4: Remove the explanatory variable.

Step 5: Rebuild and analyze the model with the remaining explanatory variables.

Step 6: Repeat Step 3, Step 4, Step 5 until all p-value $< \alpha$,