TAMS 65 Assignment 4
Multiple linear regression - Dummy variables

This assignment deals with multiple linear regression as well as Dummy variables.

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.

Dummy variables

In a study, one wanted to study whether an active security program has significance for the number of working hours which lost due to accidents at work. 50 companies were randomly selected. Results:

$$
\begin{aligned}
Y &= \text{ number of lost working hours during 1 year,} \\
x_1 &= \text{number of employees,} \\
x_2 &= \begin{cases} 1, & \text{if you have security program;} \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

Now you analyze the data using Dummy variables.

Download and open the file **Assignment4.m**, then **run** it. Input the codes in the **Command Window** or **Editor** window.

Questions

(a) Analyze the data according to the model

$$Y = \gamma_0 + \gamma_1 z_1 + \gamma_2 x_2 + \varepsilon,$$

where $z_1 = x_1/1000$ and $x_2 = 1$ if the person has security program and otherwise $x_2 = 0$. Scatter plot y against z_1. Calculate the coefficient of determination R^2.

(b) Plot the estimated regression lines for $x_2 = 1$ and $x_2 = 0$ together with the observations.

(c) Now analyze the data according to the model

$$Y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \varepsilon,$$

where $z_1 = x_1/1000$ and $z_2 = x_2 z_1$. Calculate the coefficient of determination R^2.

(d) Plot the estimated regression lines for $x_2 = 1$ and $x_2 = 0$ together with the observations.

(e) Does this analysis give any indication that security programs are leading to fewer working hours lost? Construct an appropriate confidence interval to answer the question.