MIE237 Assignment 2

Formal Due Date: 2016-04-08

Instructions

1. You have two datasets. Each name starts with the first two letters of your family name, followed by the first two letters of your given name, followed by the last two digits of your student number, and ends with `A2_1.csv` and `A2_2.csv`. The datasets can be found within `a2_data.zip` in the github repository.
2. The assignment must be done using R Markdown. *No exceptions.* You will submit the R Markdown file and the document it makes. *Don't make any further alterations to the file it makes.* Use the course website to submit your work.
3. You can render the R Markdown file into your choice of HTML, PDF, or Word. Note that PDF format requires a LaTeX installation and Word may require Word to be installed.
4. The beauty of your report will not be judged. Don't feel you need to spend too much time on purely aesthetic matters. You can if you like, but it won't affect your mark.

Questions

For each of the two simulated datasets, select a suitable regression model. The `y` variable is the "output" variable. The others are input variables. Do not consider any higher order terms or any variable transformations. There may be more than one suitable model, which is also an acceptable conclusion.

Do not show all intermediate computations. Only give a description of how you arrived at your final models, perhaps summarized in table form if you like, at most. Pages and pages of regression output is not suitable and will result in a lower mark even if you make a good selection in the end.

There is no need to discuss model assumptions. The datasets have all been simulated to follow the regression model assumptions perfectly.

Note about variable names

I accidentally made the input variable names start with `X` rather than `x`. This might turn out to be wildly annoying. Feel free to change the names to lower cases. Here's a quick way to do it, supposing the dataset has been named `d`:

```
names(d) <- tolower(names(d))
```