MIE237 Assignment 1

Due: 2016-03-11 17:00:00

Instructions

1. Your dataset's 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.
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. **Everyone's dataset will be slightly altered the afternoon before the assignment is due.** It will change all the calculations, but not any conclusions you might make.
4. 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.
5. 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.
6. The `.Rmd` files for the lectures and labs contain examples of everything you'll need for this assignment. Much of the lectures and some of the labs also contain lovely mathematical typesetting, usually enclosed by pairs of `$` or `$$` symbols. You definitely don't need to worry about that. You won't be doing any mathematical typesetting in this assignment. Just focus on the `R` code.

Questions

The assignment is essentially a repeat of the test, plus a regression question. But rather than using pen, paper, and calculator, you'll use the computer.

The dataset has 5 variables: `ID, Manufacturer, Size, Age, Rust`. The variable `ID` contains the serial number of the transformer. The variable `Manufacturer` contains the manufacturer name, one of: `Camtran, Nema, Moloney`. The variable `Size` contains a description of the transformer's power rating. The variable `Age` contains the age in years of the transformer at the time of its failure. The variable `Rust` contains a number from 0 to 100 indicating the amount of corrosion that the transformer has.

1. Produce a 95% confidence interval for the difference in mean age at failure between `Camtran` and `Moloney` transformers, commenting on any relevant assumptions you might have needed to make.
2. The company wants to look at the `Manufacturer` and `Size` variables.
    a. Produce a 95% confidence interval for the proportion of transformers that are manufactured by Nema, commenting on any relevant assumptions you might have needed to make.
    b. Perform the test of independence with null hypothesis (informally) expressed as: H_0 : `Manufacturer` and `Size` are independent, commenting on any relevant assumptions you might have needed to make.
3. The company wants to look at the `Age` and `Rust` variables. Perform the regression analysis to see if `Rust` (output) and `Age` (input) have a linear relationship. See if the slope is different from 0. Comment on any relevant assumptions you might have needed to make.