

STAT8010 Assignment 2 – Due Wednesday December 11th at 9am

The file “Assignment2.csv” contains data for your second assignment. A description of the column headings is as follows:

No: row number
year: year of data in this row
month: month of data in this row
day: day of data in this row
hour: hour of data in this row
pm2.5: PM2.5 concentration ($\mu\text{g}/\text{m}^3$)
DEWP: Dew Point ($^{\circ}\text{f}$)
TEMP: Temperature ($^{\circ}\text{f}$)
PRES: Pressure (hPa)
cbwd: Combined wind direction
lws: Cumulated wind speed (m/s)
ls: Cumulated hours of snow
lr: Cumulated hours of rain

You must read this data into R and complete a number of tasks.

- 1) You must apply a linear regression model with TEMP as your y variable and an appropriate combination of other x variables from the dataset. You should justify your choice in terms of the number of variables.
- 2) You should build a Shiny app or dashboard allowing a scatterplot for any combination of variables to be displayed. Additionally, you should be able to generate histograms, boxplots etc. of your data in this app.
- 3) You should include the ability to fit a linear regression model to the scatterplots generated in (2). The chart should include the fitted line and a table with the slope and intercept should be present within the Shiny App or dashboard.
- 4) Using Monte Carlo simulations, you should attempt to predict the temperature in subsequent years. This should be done using at least two different models (i.e. different collections of variables in part 1). You should clearly state which performs best.
- 5) We now consider two linear models: one where we consider y as temperature and x as pressure and a second with y as temperature and both pressure and wind speed as x variables. You should apply both models and calculate the ESS statistics on the data (some notes on this statistic can be found here: https://www.graphpad.com/guides/prism/7/curve-fitting/reg_howthetestworks.htm?toc=0&printWindow). You are required to generate a distribution for this test statistic by simulation; you may assume the errors and residuals are normal. You should clearly state whether this model is an appropriate fit to the data based on your simulations.

You should generate a short report detailing your work. You should clearly state your conclusion. Your codes should also be included in your submission (R script file or R markdown file or as appendix in report). Submission is by email to justin.mcguinness@cit.ie and is due **9am Wednesday 11th December**.

This is an **individual assignment**. Any collaboration amongst students is forbidden. Plagiarism is strictly prohibited and will be dealt with by the harshest punishments available.