

STAT8010 Assignment 2 – Due Tuesday December 22nd at 23:59

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

- | | |
|--|---|
| - No: row number | - PRES: Pressure (hPa) |
| - year: year of data in this row | - feed: label of feed used |
| - month: month of data in this row | - undesP: Undesired proteins |
| - day: day of data in this row | - udt: Cumulated hours of unplanned down time |
| - hour: hour of data in this row | - pdt: Cumulated hours of planned down time |
| - tconc: concentration of target product | |
| - Ph: ph reading | |
| - TEMP: Temperature (Celsius) | |

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

- 1) 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.
- 2) 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.
- 3) Using Monte Carlo simulations, you should attempt to predict tconc for the year 2015. This should be done using at least two different models (i.e. different collections of variables or prediction values). You should clearly state which performs best.
- 4) Consider a machine that inserts a needle into test tubes on a conveyer for sampling in a factory process. This machine may become misaligned in the 2 dimensions of the plane of conveyer travel (x and y axes) independently. The machine is realigned to centre at the start of each day and it then samples 200 test tubes throughout the day. The machine fails to sample correctly if it is misaligned in any direction by 2cm or more, as it misses the test tube (possibly colliding with the glass). The x misalignment is 0.1mm on average in the direction of conveyor travel (positive x-direction) for each test, but that this can vary somewhat with a standard deviation of 0.1mm. Similarly, the y misalignment is biased in the negative y-direction, and is much smaller on average; the engineers believe that the average misalignment in the negative y direction is 0.05mm per test, with a standard deviation of 0.05mm.
 - a. Simulate the distribution of misalignments at the end of the day?
 - b. Estimate the likelihood of failure throughout the day?
 - c. Visualise the simulated alignments of the machine at the end of the day on a scatterplot, showing the 2cm limit.
- 5) It costs €50,000 when the machine goes offline due to excessive misalignment and no further batches can be tested for the remainder of the day. Each batch passed through the machine results in gross profit of €400. If a batch is ready for testing but the machine is offline, there is a €500 cost for storage and alternate testing of each untested batch under the target number of tests per day. Given these, use Monte Carlo simulations to find the best strategy - i.e. what is the optimal target number of runs per day before realignment should be done.

You should generate a short report (max 3000 words) detailing your work for parts 3, 4 & 5. You should clearly state your conclusion. Your codes should also be included in your report submission (R markdown file or as appendix in report) along with your R script file (if necessary). Submission is by Canvas and is due **23:59 Tuesday 22nd 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.