Basic data analysis: drilling machine replacement

Practical

Week 5

Learning objectives

By the end of this lab session you will be able to:

- load csv data files into R as `data.frames`;
- use your knowledge from engineering statistics and R to perform basic exploratory analysis on actual data sets; and
- use the analysis to draw appropriate conclusions and answer given industrial engineering questions.

Prerequisites

It is assumed that students are familiar with the following from BES220:

Engineering statistics:

- Sections 1.6 to 1.7 of Chapter 1 of either of the prescribed textbooks (Chapter 1 is the same for both textbooks).
- Data Analysis and Statistical Inference Chapter 2: Introduction to data (DataCamp)

Case study

We are responsible for a relatively simple manufacturing line that produces engines for remote control vehicles. Over the last three months a number of the engines have been returned by clients who claim that the engines do not work. An investigation found that the fault can be traced back to a hole in one of the components that did not meet specifications. In the manufacturing process, the hole is drilled by a single drilling machine that has been in operation for over a year and cannot be repaired. The machine can last another 365 days, but based on the product returns, management has instructed us to investigate the issue and answer the following:

- How well is the drilling machine performing?
- Should it be replaced, and if so, what new machine should we get?

Specifications for the hole are that it should have a diameter of 10cm, and that it has a tolerance of 0.5cm. If the hole is too large or too small the engine will be defective. The new machines all have a life-time of 365 days, after which they are scrapped. In our manufacturing line a machine has to drill holes for 9 hours a day, and it takes 3 minutes to drill a hole. If a part does not meet specifications, the client is reimbursed, which costs R100 per defective product. Purchase prices for the machines can be found in `MachineHolesPrices.csv`.

Study design

To answer management's questions we will need data on the current drilling machine's performance. And if it has to be replaced we need data on the performance of new machines.

To test the current drilling machine we performed a simple test:

- we took 30 pieces of raw material and then used the machine to drill a hole in each; and
- we measured the diameter for each hole and recorded it.

Available data

We saved the sample number and diameter for each drilled hole drilled in the `currentMachineHoles.csv` file. The file is available on clickUP and has to be downloaded and saved.

We further identified three other machines that are popular for this type of job, contacted the distributors of the machines and asked them to do the same experiment. We obtained price information for each machine, which we will use later. Results for the hole sizes from the three machines can be found in `MachineHoles1.csv`, `MachineHoles2.csv`, `MachineHoles3.csv`, and price information in `MachineHolesPrices.csv`, also available from clickUP.

Instructions

For the first part, we will ignore machine costs and do the following:

- use the `currentMachineHoles.csv` data to analyse the performance of the existing machine;
- find a suitable replacement using the `MachineHoles1.csv`, `MachineHoles2.csv`, and `MachineHoles3.csv` data; and
- use our analysis to make a recommendation to management.

Since we are reporting our results to management, we have to be concise. Therefore we should keep our analysis as concise as possible. However, the included analysis should still be accurate, fact based, and complete. We can't overlook important factors.

For the second part we will use our results from the first part and take costs into consideration and compare the current and three replacement machines.

A good place to start is first to view the data, to get a feel for it, and then to decide on which analysis in the form of figures and tables will be useful in our analysis. Here you will have to think back to all the engineering statistics data analysis techniques that we discussed. Don't be scared to try different techniques. The idea is to try different stuff and then keep the best ones to show to management.

Useful commands

The data is saved in `.csv` format, which base R can easily deal with, or you can use the `readr` library.

Functions, as described in the DataCamp chapters that may be useful include:

- `names(mydata)`
- `head(mydata)`
- `tail(mydata)`
- `summary(mydata$someField)`
- `subset(mydata, mydata$someField > someValue & mydata$someField < someValue)`
- `hist(mydata$someFieldX)`
- `barplot(mydata$someFieldX, mydata$someFieldY)`
- `boxplot(mydata$someField)`
- `boxplot(mydata$someField~mydata$someCategory)`
- `table(mydata$someField, mydata$someCategory)`

Other plotting functions:

- `abline(h = y_intercept)`
- `abline(v = x_intercept)`

The above two commands *adds* horizontal (`h`) or vertical (`v`) lines to an *existing* plot and should be called after calling a main plot, such as `hist()`. The functions are useful for showing tolerance limits in an existing plot.

Another useful function is:

- `rbind(dataFile1, dataFile2, dataFile3,...)`

This function binds different data.frames into a single frame. You can think of it as taking the a frame and then adding all the records of another frame to the bottom of it. This can be done for multiple data.frames, as well as each one has the same amount of columns and the data.frames have identical column names. Hint: in certain instances you may need to add some sort of identifier field to the data.frames before binding them, otherwise you won't be able to distinguish what the records refer to during your analysis:

- `dataFile1$ID <- 'Name for data file 1';`
- `dataFile2$ID <- 'Name for data file 2';`
- `dataFile3$ID <- 'Name for data file 3';`
- `...`

The last function that may be useful is:

- `prop.table(mytable,1)`

The function calculates the probabilities per row of a table. If you want to calculate it per column use:

- `prop.table(mytable,2)`