Assessment 2a:  
Questions and Template

Version 1.0  
EGH404: Research in Engineering Practice  
22 July 2019

# About this document

* This document sets out the five (5) questions you are to complete for assessment 2a.
* It also provides the template for your responses.

# Your tasks for Assessment 2a

1. Read the questions in this document
2. Edit your responses into this document
   1. Replace the text in *blue* *italics* with your responses
   2. The associated data and/or original graphs are provided separately in an Excel spreadsheet where necessary.
   3. All necessary calculations and charting should be completed in Matlab.
3. You will submit your responses to this assessment as a PDF or word document which you will upload to TurnItIn. Your document should include:
   1. This template with responses for each question included in the corresponding section.
   2. Your original source code (ie. Matlab scripts) which you generated to complete relevant questions, included either after each question or at the end of the document as an appendix (copy and paste the code into the document, or if using live scripts merge the document PDF with the exported live script PDF). Please note your Matlab code will not be run except under exceptional circumstances but may be used to evaluate the process that you used to answer the question.

# Question 1: Recommending a bolt supplier

Your firm works in construction and buys [ASTM A325M8S](https://en.wikipedia.org/wiki/ASTM_A325) bolts in very large numbers.

You are interested in purchasing bolts from different suppliers and have obtained and tested samples of 200 bolts from three suppliers: Allnutt, Boltzman, Coachers.

You have tested these in your new bolt testing machine, purchased after you had discovered your old bolt testing machine was performing erratically. (For the purposes of this question, you are safe to assume that the new machine works perfectly and that its results are in MPa.)

Your firm wants to ensure that the tensile strengths of the bolts it buys are

* as high as possible
* as consistent as possible (i.e., the variability of bolt strengths is as *low* as possible)

It is critical that the tensile strength of these bolts exceeds 830MPa.

You should draw on the unit content concerning summary statistics to answer this question.

**On the basis of the measurements recorded in Bolts.csv, which supplier (if any) would you recommend, and why?  
(Provide any code or visualisations you use to justify your response.)**

*Edit your response to this question here*

# Question 2: Changes in Air Quality Over Time

The state government’s Department of the Environment and Heritage Protection are concerned about air pollution levels in Brisbane city. You have been provided with data for two sites, the Brisbane CBD and South Brisbane, for 2010 and 2016 in the following four files:

* brisbanecbd-aq-2010.csv
* brisbanecbd-aq-2016.csv
* southbrisbane-aq-2010.csv
* southbrisbane-aq-2016csv

You have been asked to advise the Government:

* If air pollution is related across the two sites, and if so, has this relationship changed between 2010 and 2016
* Has the level of air pollution at either site changed from 2010 to 2016? Have these changes across the two sites been consistent?

You have been advised to focus on the PM10 data series.

You should draw on the unit content concerning summary statistics and correlation to answer this question.

**(Provide any code or visualisations you use to justify your response.)**

*Edit your response to this question here*

# Question 3

## Part A

* Automotive Excellence produces a range of components needed to produce automobile engines.
* Production of different component categories varies from month to month.
* The intent of this graph is to summarise the variation in numbers of component categories from month to month in 2017.

### A.1 What’s wrong with this graphic?

List and briefly explain each of the problems that you detect in the design of this graphic drawing on the principles presented on visualisation and from your wider reading.

*Edit your response to this question here*

### A.2 How would you redesign this graphic?

Develop an alternative graphic that addresses the problems you detected

*Copy and paste your alternative graphic here*

### A.3 What do you now observe in your redesigned graphic?

What are the salient problems or features of the data that you now observe?

*Edit your response to this question here*

# **Part B**

* Automotive Excellence has gaskets which come in a range of sizes for different engine blocks and for different components.
* Through a sophisticated product tracking system, Automotive Excellence monitors the working life of these gaskets and have collected time-to-failure data on hundreds of gaskets.
* The Automotive Excellence product quality team have summarised this data by calculating the mean failure times of a sample of each size. The sizes are reported by their overall length in mm.
* The intent of this graph is to help understand whether there is a relationship between gasket size and failure times.

### B.1 What’s wrong with this graphic?

List and briefly explain each of the problems that you detect in the design of this graphic drawing on the principles presented on visualisation and from your wider reading.

*Edit your response to this question here*

### B.2 How would you redesign this graphic?

Develop an alternative graphic that addresses the problems you detected

*Copy and paste your alternative graphic here*

### B.3 What do you now observe in your redesigned graphic?

What are the salient problems or features of the data that you now observe?

*Edit your response to this question here*

# Question 4: Estimating Concrete Strength

You work for a construction firm who need to be able to accurately predict the compressive strength of concrete given variables including the concrete composition and its age. You are aware that the relationship between these different components and the concrete strength is complex, however you have been asked to investigate how well a simple linear regression model works for prediction. Using the provided data (Concrete\_Data.xls), develop models to predict:

* Concrete strength from the single best indicator variable;
* Concrete strength from all variables.

With the second model, determine if any variables are not contributing significantly to the model, and what impact removing these has on prediction performance. Comment on the final model and its accuracy, and whether it would be appropriate to use this model in practice.

You should draw on the unit content concerning correlation and regression to answer this question. Note that you are not expected to use training/validation/testing data splits, although you are welcome to do so. No marks will be lost/gained for using/not using data splits.

**(Provide any code or visualisations you use to justify your response.)**

*Edit your response to this question here*

# Question 5: Predicting Missing Bikeway Data

Brisbane City Council (BCC) is considering upgrades to the bikeway networks. They are using data they have gathered from sensors placed along the bike paths, which record the number of cyclists, to plan the upgrades, however there have been a number of sensor failures which have resulted in their dataset missing a number of entries. BCC have requested that you investigate if it possible to predict missing data from data gathered from other sensors on the bike path network.

You have been provided with three years’ worth of data (Bike-Ped-Auto-Counts-2014.csv, Bike-Ped-Auto-Counts-2015.csv, and Bike-Ped-Auto-Counts-2016.csv), and the corresponding three years of weather data (the files named IDCJAC00XX\_040913\_201X\_Data.csv). As an initial investigation, you have been asked to consider only these data series from the bikeway data:

* BicentennialBikewayCyclistsInbound
* GoBetweenBridgeCyclistsInbound
* KangarooPointBikewayCyclistsInbound
* NormanParkCyclistsInbound
* RiverwalkCyclistsInbound
* StoryBridgeWestCyclistsInbound

Using the three years data, you are to:

* Determine which counters are best suited to predicting the missing data in others (i.e. which, if any counters, could be used to predict BicentennialBikewayCyclistsInbound).
* Investigate if weather data can be used to help support this prediction, and if so, indicate what weather data is most helpful.
* Predict missing data where appropriate to generate a more “complete” database.
* Comment on the resulting corrected dataset. In particular:
  + What problems, if any, may arise from this approach?
  + How effective has this been in reducing missing data?
  + How trustworthy are the predicted values?

You should draw on the unit content concerning correlation and regression to answer this question. Note that you are not expected to use training/validation/testing data splits, although you are welcome to do so. No marks will be lost/gained for using/not using data splits.

**(Provide any code or visualisations you use to justify your response.)**

*Edit your response to this question here*

# How will my work be graded?

Your responses to questions will be marked using the rubric set out below.

All questions are worth equal marks.

## Passing grades

### Grade 7 work…

* Clear and concise analysis and visualisation (where required) of given data.
* This level of work implies excellence in thinking.
* It is on the whole clear, precise and well-reasoned, and all aspects of the problems are discussed.
* Terms and distinctions are used effectively.
* Work at this level demonstrates a mind beginning to take charge of its own ideas, assumptions, inferences and intellectual processes.

### Grade 6 work…

…demonstrates more strength than weaknesses and is more consistent in high-level performance than grade 5 work. It has some distinctive weaknesses, though no major ones

* Analysis is mostly clear, precise and well-reasoned/ visualisation of given data can be improved
* This level of work implies excellence in thinking.
* Terms and distinctions are used effectively.
* Work at this level demonstrates a mind beginning to take charge of its own ideas, assumptions, inferences and intellectual processes.

### Grade 5 work…

… demonstrates more than a minimal level of skill, but it is also highly inconsistent, with as many weaknesses as strengths.

* Clear analysis with some flaws in methodology/some questions are unexplored/weak visualisation of data
* This level of work shows some signs of critical thinking.
* Clear analysis but occasionally lacks reasoning.
* Terms and distinctions are often used effectively.
* There is some evidence that the student is genuinely engaged in the task of taking charge of his or her thinking.

### Grade 4 work…

… demonstrates only a minimal level of understanding and skill in the subject

* Minimal analysis/weak representation of given data.
* This level of work shows minimal signs of critical thinking.
* Analysis is not clear and often lacks reasoning.
* Some terms and distinctions are used effectively.

## Failing grades

### Grade 3 work…

…demonstrates a pattern of unskilled thinking

* Incorrect analysis/flawed methodology to analyse data/poor representation of data.
* The work suggests that the student is trying to get through the assessment by means of wrote recall, attempting to acquire knowledge by memorisation rather than through comprehension and understanding.
* The work suggests the student is not developing the skills of thought and knowledge requisite to understanding how to read and make sense of data.
* Work at this level, on close examination, typically reveals characteristics including that the student does not understand the basic nature of what it means to think within the context of the assessment, and in any case does not display the thinking skills and abilities which is at the heart of the assessment.
* The work is vague, imprecise, and unreasoned.
* There is little evidence that the student is genuinely engaged in the task of taking charge of his or her thinking.
* The work suggests the student is simply going through the motions without really putting any significant effort into thinking through the questions.
* The work suggests the student is
  + not analysing issues clearly,
  + not formulating information clearly,
  + not accurately distinguishing the relevant from the irrelevant,
  + not identifying key questionable assumptions,
  + not clarifying key concepts,
  + not identifying relevant competing points of view,
  + not reasoning carefully from clearly stated premises,
  + not tracing implications and consequences.
* The work does not display discernible reasoning and problem-solving skills

### Grade 2/1 work…

…demonstrates a pattern of unskilled thinking and fails to do the required work/minimum work done.

### Grade 0 work…

… No relevant response to the questions