

ASSESSMENT DETAILS

SIT718 Real World Analytics

Assessment Task 3: Problem Solving



Using aggregation functions for data analysis

This document supplies detailed information on assessment tasks for this unit.

Key information

Due: Friday 3rd May 2019 11.30 pm AEST

Weighting: 30%

Reference style: Harvard

Learning Outcomes

This assessment assesses the following Unit Learning Outcomes (ULO) and related Graduate Learning Outcomes (GLO):

Unit Learning Outcome (ULO)	Graduate Learning Outcome (GLO)
<p>ULO1 – assessed through student ability to apply knowledge of multivariate functions, data transformations and data distributions to summarise data sets.</p> <p>ULO2 – assessed through the student ability to analyse datasets by interpreting summary statistics, model and function parameters.</p> <p>ULO4 - assessed through student ability to develop software codes to solve computational problems for real world analytics.</p>	<p>GLO1 - Discipline knowledge and capabilities</p> <p>GLO4 – Critical thinking</p> <p>GLO5 – Problem solving</p>

Purpose

This assignment will test your knowledge and understanding of the aggregation functions and their applications for data summarization and prediction. This assignment will also test your ability in R programming, in using specific R commands as well as R packages.

Instructions

The work is individual. Solutions and answers to the assignment must be explained carefully in a concise manner and presented carefully. Use of books, articles and/or online resources on share price related to SIT718 Real World Analytics is allowed. Students are expected to refer to the suitable literature where appropriate.

The assessment consists of **FOUR tasks**. Students must attempt all tasks and provide an individual written report in appropriate word processor.

The detailed problem description and data set will be released to students on **Friday 5th April**.

ASSESSMENT DETAILS

Submission details

- **No more than 7 A4 sides**, including Figures, Tables, Appendices and References. The report should be typed. Use minimal font 11pt and 2.5cm side margins. If the page limit is exceeded only the first 7 pages will be marked.
- Assignment (a report in pdf format, software code and/or data) must be submitted via the assignment folder in the unit site (accessed via the unit [Program page](#))
- No e-mail or hardcopy submissions are accepted.

Extension requests

Requests for extensions should be made to Unit/Campus Chairs well in advance of the assessment due date. If you wish to seek an extension for an assignment, you will need to apply by email directly to **Prof. Maia Angelova** (maia.a@deakin.edu.au), as soon as you become aware that you will have difficulty in meeting the scheduled deadline, but at least 3 days before the due date. When you make your request, you must include appropriate documentation (medical certificate, death notice) and a copy of your draft assignment.

Conditions under which an extension will normally be approved include:

Medical To cover medical conditions of a serious nature, e.g. hospitalisation, serious injury or chronic illness. Note: Temporary minor ailments such as headaches, colds and minor gastric upsets are not serious medical conditions and are unlikely to be accepted. However, serious cases of these may be considered.

Compassionate e.g. death of close family member, significant family and relationship problems.

Hardship/Trauma e.g. sudden loss or gain of employment, severe disruption to domestic arrangements, victim of crime. Note: Misreading the timetable, exam anxiety or returning home will not be accepted as grounds for consideration.

Special consideration

You may be eligible for special consideration if circumstances beyond your control prevent you from undertaking or completing an assessment task at the scheduled time.

See the following link for advice on the application process:

<http://www.deakin.edu.au/students/studying/assessment-and-results/special-consideration>

Assessment feedback

Students will receive written feedback to aid reflection and analysis of problem strategies and solutions for consideration in the upcoming problem-solving task.

Referencing

You must correctly use the Harvard method in this assessment. See the Deakin [referencing guide](#).

Academic integrity, plagiarism and collusion

Plagiarism and collusion constitute extremely serious breaches of academic integrity. They are forms of cheating, and severe penalties are associated with them, including cancellation of marks for a specific assignment, for a specific unit or even exclusion from the course. If you are ever in doubt about how to properly use and cite a source of information refer to the referencing site above.

ASSESSMENT DETAILS

Plagiarism occurs when a student passes off as the student's own work, or copies without acknowledgement as to its authorship, the work of any other person or resubmits their own work from a previous assessment task.

Collusion occurs when a student obtains the agreement of another person for a fraudulent purpose, with the intent of obtaining an advantage in submitting an assignment or other work.

Work submitted may be reproduced and/or communicated by the university for the purpose of assuring academic integrity of submissions: <https://www.deakin.edu.au/students/study-support/referencing/academic-integrity>

Using aggregation functions for data analysis

Download [SIT718 Assessment-Task 3-T1 2019-data and script.zip](#) it contains the data file [Energy19.txt] and the R code [AggWaFit718.R] to use with the following tasks, include these in your R working directory.

Total Marks 100, Weighting 30%

Energy Prediction of Domestic Appliances Dataset

The given dataset, "Energy19.txt", can be used to create models of energy use of appliances in a energy-efficient house. The dataset provides the Energy use of appliances (denoted as Y) using 671 samples. It is a modified version of data used in the study [1]. The dataset includes 5 variables, denoted as X1, X2, X3, X4, X5, and Y, described as follows:

X1: Temperature in kitchen area, in Celsius

X2: Humidity in kitchen area, given as a percentage

X3: Temperature outside (from weather station), in Celsius

X4: Humidity outside (from weather station), given as a percentage

X5: Visibility (from weather station), in km

Y: Energy use of appliances, in Wh

Assignment Tasks

1. Understand the data [20 marks]

- (i) Download the txt file (Energy19.txt) from Future Learn and save it to your R working directory.
- (ii) Assign the data to a matrix, e.g. using
`the.data <- as.matrix(read.table("Energy19.txt"))`
- (iii) The **variable of interest** is Energy use of appliances (Y). To investigate Y, generate a subset of 300 data, e.g. using:

`my.data <- the.data[sample(1:671,300),c(1:6)]`

- (iv) Using scatter plots and histograms, report on the general relationship between each of the variables X1, X2, X3, X4, X5 and the *variable of interest* Y. Include 5 scatter plots, 6 histograms, and 1 or 2 sentences for each of the variables, including the variable of interest Y.

2. Transform the data [20 marks]

- (i) Choose any **four** from the five variables (X1, X2,...,X5). Make appropriate transformations to the chosen four variables and the variable of interest Y so that the values can be aggregated in order to predict the *variable of interest*. Assign your *transformed* data along with your *transformed* variable of interest to an array (it should be 300 rows and 5 columns). Save it to a txt file titled "name-transformed.txt" using
- ```
write.table(your.data,"name-transformed.txt")
```
- where "name" is replaced with your name - you can use your surname or first name.
- (ii) Briefly explain the transformations applied for the selected four variables and the *variable of interest*. (1- 2 sentences each)

## 3. Build models and investigate the importance of each variable [40 marks]

- (i) Download the AggWaFit718.R file (from Future Learn) to your working directory and load into the R workspace using,
- ```
source("AggWaFit718.R")
```
- (ii) Use the fitting functions to learn the parameters for
- A weighted arithmetic mean (WAM)
 - Weighted power means (WPM) with $p = 0.5$, and $p = 2$,
 - An ordered weighted averaging function (OWA), and
 - A Choquet integral.
- (iii) Include two tables in your report - one with the error measures and correlation coefficients, and one summarising the weights/parameters and any other useful information learned for your data.
- (iv) Compare and interpret the data in your tables. Comment on
- a. How good the model is,
 - b. The importance of each of the variables (the four variables that you have selected),
 - c. Any interaction between any of those variables (are they complementary or redundant?) and
 - d. Better models favour higher or lower inputs.
- (1-3 paragraphs for part 3(iv))

4. Use your model for prediction [20 marks]

- (i) Choose your best fitting model.
Using your best fitting model, predict the Energy use of appliances for the following input $X_1=18$; $X_2=44$; $X_3=4$; $X_4=74.8$; $X_5=31.4$.
- (ii) Give your result and comment on whether you think it is reasonable. (1-2 sentences).
- (iii) Comment on the best conditions (in terms of your chosen four variables) under which a **low** Energy use of appliances will occur. (1-2 sentences).

Submit to the **SIT718 Clouddeakin Dropbox**. Your final submission should include the following three files:

1. A report, "**name-report.pdf**", in pdf format (created in any word processor), covering all of the items in above (where "name" is replaced with your name -you can use your surname or first name). With plots and tables it should be **up to 7 pages**.
2. A data file named "**name-transformed.txt**" - just to help us distinguish them!).
3. The R code file (that you have written to produce your results) named "**name-code.R**" (where "name" is replaced with your name - you can use your surname or first name).

References:

1. Luis M. Candanedo, Veronique Feldheim, Dominique Deramaix. Data driven prediction models of energy use of appliances in a low-energy house, Energy and Buildings, Volume 140, 1 April 2017, Pages 81-97, ISSN 0378-7788.
<http://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction>

REMINDER!!

Download [SIT718 Assessment-Task 3-T1 2019-data and script.zip](#) it contains the data file [Energy19.txt] and the necessary R code [aggWaFit718.R] to use with the above tasks, include these in your R working directory.