SIT718 REAL WORLD ANALYTICS

Assessment Information
ASSESSMENT TASK 2: Problem solving task

This document supplies detailed information for your second of four assessment task in this unit.

Key information

- **Due:** 11.30 pm AEST on Monday, 16th December 2019.
- Weighting: 25%
- 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)
ULO1 – assessed through student ability to apply knowledge of multivariate functions, data transformations and data distributions to summarise data sets. **ULO2** – assessed through the student ability to analyse datasets by interpreting summary statistics, model and function parameters. **ULO4** - assessed through student ability to develop software codes to solve computational problems for real world analytics. **ULO5** – assessed through student ability to demonstrate professional ethics and responsibility for working with real world data.	**GLO1 -** Discipline knowledge and capabilities **GLO4 –** Critical thinking **GLO5 –** Problem solving **GLO8 –** Global citizenship

Purpose

The aim of this assignment is to analyse the behaviour and performance of one company using historical data. You can use statistical data analysis and features you have been learning in R.

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 several tasks. Students must attempt all tasks and provide an individual written report in an appropriate word processor.

SIT718 REAL WORLD ANALYTICS

ASSESSMENT TASK 2: Problem solving

Submission details

- **No more than 7 A4 pages,** 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 dropbox in the unit site (accessed via DeakinSync).
- 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 **Dr Thanh Thi Nguyen** (thanh.nguyen@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.

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.

SIT718 REAL WORLD ANALYTICS

ASSESSMENT TASK 2: Problem solving

Plagiarism occurs when a student passes off as the student's own work, or copies without acknowledgment 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/

SIT718 REAL WORLD ANALYTICS

ASSESSMENT TASK 2: Problem solving

The aim of this assignment is to analyse the behaviour and pattern of weather time series. You need to download the data from https://www.ncdc.noaa.gov/cag/national/time-series and select **one** of the following periods based on the last digit of your student ID number:

- January 1919 to January 1939, if the last digit of your ID is 0 or 1.
- January 1939 to January 1959, if the last digit of your ID is 2 or 3.
- January 1959 to January 1979, if the last digit of your ID is 4 or 5.
- January 1979 to January 1999, if the last digit of your ID is 6 or 7.
- January 1999 to January 2019, if the last digit of your ID is 8 or 9.

For example, if your student ID number is 218157967, the last digit is 7, thus you need to select the period January 1979 to January 1999.

In the link above, select Parameter = "Average Temperature", Time Scale = "All Months", and the start and end year of your selected period. Note that this is the monthly data, and you only need to use the Date and Value columns, ignore the Anomaly column.

Please use R/R Studio programming language for computational tasks, e.g. Task 2, Task 3, etc… and present results in your report.

By completing this assignment, you are required to write a case study report of no more than 7 pages, which should include the following information:

1. Your student ID, your chosen period and the main characteristics of the weather for this period. Pre-process, prepare and clean the data and convert them to the CSV format.

[5 marks]

2. Visualise weather data for this period. Describe the general trend of the data for this period. Produce your own graph. Print out the mean and variance of the data, and describe the distribution by using the histogram.

[10 marks]

3. One model for modelling weather data is a stochastic model based on the Geometric Brownian motion. You can find more details about this model online.

A simplified version of the model is given below:

$$S_t = S_1 \exp\left[\left(\mu + \frac{\sigma^2}{2}\right) t\right], \qquad \textbf{(Equation 1)}$$

where,

S_1 is the initial value (first data point of your time series).

S_t is the weather data on month t (this would be the t^{th} data point in your analysis)

μ is the mean of the **lognormal distribution**.

σ is the standard deviation of **lognormal distribution** of the data.

This model is under the assumption that data is **lognormally distributed**.

You need to research further and describe the general meaning of μ and σ.

SIT718 REAL WORLD ANALYTICS

ASSESSMENT TASK 2: Problem solving

[10 marks]

4. State the condition that weather data have to satisfy in order to be represented by a Geometric Brownian motion.

- Do your data satisfy this condition? Produce appropriate evidence to support your answer.

Hint: the logarithm of the monthly weather data change, i.e., $\log(S_t/S_{t-1})$, is a normal distribution. Even if your data do not satisfy the condition, continue to use the model and comment in Task 7.

[20 marks]

5. Compute values σ and μ for your selected period. Present the formulas used and the results.

What is the effect of time on σ and μ? Are the parameters constants (with time)? How could this affect the model performance?

[20 marks]

6. Estimate the expected values of the weather data on the next 5 months after your chosen period using the formula given in Equation 1. Use the values of parameters σ and μ computed in Task 5. Consider the last data point of your data as S_1 and $t = 5$. Compare your results with the published real data.

[15 marks]

7. Comment on the model's performance and state the restrictions of the model.

Link your results to the weather behaviour.

Can you relate the findings to the climate change?

Compare your results with other models for weather data, e.g. machine learning models; summarise your findings with no more than 300 words.

[20 marks]

All supporting information should be presented in the pdf report. It will be assessed for style and grammar, professional presentation of figures, tables and references. You should not exceed the 7 A4 page limit (including figures, tables, appendices (if needed) and references).

List and quote in the text the references used, including books, articles and web resources. Use the Harvard style.