ROBERT GORDON UNIVERSITY ABERDEEN

Assessment Brief Template

Academic Year	**2023**
Semester	**Semester 2**
Module Number	**CM2605**
Module Title	**Simulation and Modelling Technique**
Assessment Method	
Deadline (time and date)	**23rd April 2023, 12:00 midnight (IST)**
Submission	**Assessment Dropbox in the Module Study Area in CampusMoodle.**
Word Limit (see Assessment Word Limit Statement)	
Module Co-ordinator	**Prashan Rathnayaka**

What knowledge and/or skills will I develop by undertaking the assessment?

To introduce the process of designing models of existing or proposed real-world systems and how to use these models to perform simulations that allow for predictions about the future behavior of the system.

On successful completion of the assessment students will be able to achieve the following Learning Outcomes:

1. Describe and analyse various modeling techniques, types of stochastic processes and apply probabilistic techniques in data modelling.
2. Critically evaluate appropriate techniques for random number generation and apply them to simulate the real-life processes.
3.
4.
5.

Please also refer to the Module Descriptor, available from the module Moodle study area.

What is expected of me in this assessment?

Task(s) - content

To introduce the process of designing models of existing or proposed real-world systems and how to use these models to perform simulations that allow for predictions about the future behavior of the system. Students will also gain a solid understanding of probabilistic data modeling, interpretation, and analysis for solving practical problems more generally in computer science, business and finance, economics and engineering, and daily life.

Task(s) - format

How will I be graded?

A grade will be provided for each criterion on the feedback grid which is specific to the assessment.

The overall grade for the assessment will be calculated using the algorithm below.

A	At least 50% of the feedback grid to be at Grade A, at least 75% of the feedback grid to be at Grade B or better, and normally 100% of the feedback grid to be at Grade C or better.
B	At least 50% of the feedback grid to be at Grade B or better, at least 75% of the feedback grid to be at Grade C or better, and normally 100% of the feedback grid to be at Grade D or better.
C	At least 50% of the feedback grid to be at Grade C or better, and at least 75% of the feedback grid to be at Grade D or better.
D	At least 50% of the feedback grid to be at Grade D or better, and at least 75% of the feedback grid to be at Grade E or better.
E	At least 50% of the feedback grid to be at Grade E or better.
F	Failing to achieve at least 50% of the feedback grid to be at Grade E or better.
NS	Non-submission.

ROBERT GORDON UNIVERSITY ABERDEEN

Feedback grid *Add more rows/criteria if necessary, up to a maximum of 8.*

GRADE	A	B	C	D	E	F
DEFINITION / CRITERIA (WEIGHTING)	**EXCELLENT** Outstanding Performance	**COMMENDABLE/VERY GOOD** Meritorious Performance	**GOOD** Highly Competent Performance	**SATISFACTORY** Competent Performance	**BORDERLINE FAIL**	**UNSATISFACTORY** Fail
CRITERION 1 **(x %)** **Grade:** ☐						
CRITERION 2 **(x %)** **Grade:** ☐						
CRITERION 3 **(x %)** **Grade:** ☐						
CRITERION 4 **(x %)** **Grade:** ☐						

Coursework received late, without valid reason, will be regarded as a non-submission (NS) and one of your assessment opportunities will be lost.

What else is important to my assessment?

What is plagiarism?

"Plagiarism is the practice of presenting the thoughts, writings or other output of another or others as original, without acknowledgement of their source(s) at the point of their use in the student's work. All materials including text, data, diagrams or other illustrations used to support a piece of work, whether from a printed publication or from electronic media, should be appropriately identified and referenced and should not normally be copied directly unless as an acknowledged quotation. Text, opinions or ideas translated into the words of the individual student should in all cases acknowledge the original source" (RGU 2022).

What is collusion?

"Collusion is defined as two or more people working together with the intention of deceiving another. Within the academic environment this can occur when students work with others on an assignment, or part of an assignment, that is intended to be completed separately" (RGU 2022).

For further information please see Academic Integrity.

What is the Assessment Word Limit Statement?

It is important that you adhere to the Word Limit specified above. The Assessment Word Limit Statement lists what is included and excluded from the word count, along with the penalty for exceeding the upper limit.

What if I'm unable to submit?

- The University operates a Fit to Sit Policy which means that if you undertake an assessment then you are declaring yourself well enough to do so.
- If you require an extension, you should complete and submit a Coursework Extension Form. This form is available on the RGU Student and Applicant Forms page.
- Further support is available from your Course Leader.

What else is important to my assessment?
What additional support is available? • RGU Study Skills provide advice and guidance on academic writing, study skills, maths and statistics and basic IT. • RGU Library guidance on referencing and citing. • The Inclusion Centre: Disability & Dyslexia. • Your Module Coordinator, Course Leader and designated Personal Tutor can also provide support.
What are the University rules on assessment? The University Regulation 'A4: Assessment and Recommendations of Assessment Boards' sets out important information about assessment and how it is conducted across the University.

Instructions to Candidates:

- There are **5 questions** in the assignment. **Answer all questions**.
- Total **40 marks**.
- Be sure to use **R** code for all your calculations and the statistical analysis.
- Write **R** Codes for the necessary tasks, obtain **R**-outputs and graphs, and copy & paste on the appropriate places of your solution:
- You can complete the entire assignment in a single **R Markdown** using RStudio and submitted as a **PDF** document. (i.e. write up your works as a .Rmd file, "knit" the results to a PDF file, and submit the PDF file to LMS. Between **R** chunks, you can use **Latex** to write up text or any mathematical equations (Using **R markdown** with **Latex** is optional but can be a useful skill moving forward)

QUESTION 1. (Total 8 marks)

A data scientist is interested in studying the distribution of Body Mass Index (BMI) in a certain population. The BMI is defined as

$BMI = \frac{W}{H^2}$ where W is the weight of a person measured in Kg and H is the height of a person measured in m.

In the population it is assumed that $W \sim N(60, 3^2)$ and $H \sim N(1.6, 0.1^2)$

a) Generate 10000 BMI values. (2 marks)
b) Draw the histogram of the empirical distribution of BMI and comment on it. (2 marks)
c) Find the approximate mean and variance of BMI (2 marks)
d) Estimate $p(BMI \geq 25)$ (2 marks)

QUESTION 2. (Total 8 marks)

A sport analyst wants to study the effect of service rules on winning a game. Example for service sports are Badminton, Table tennis, Volleyball etc. The analyst wishes to study two service rules:

a) **Rule A** Server is the winner of the previous point.
b) **Rule B** Server is the loser of the previous point.

The analyst chooses two players, namely Player 1 and Player 2. It is known that

a) *P*(Player 1 wins a point | player 1 serves)=0.55
b) *P*(Player 1 wins a point | player 2 serves)=0.40

If a player serves and wins the point only the score of the server is increased by 1. The game is played for Best-of-3, in which the first player to reach 2 points wins the game. Assume that a game is always started by the player 1. Write a R program to simulate the game 1000 times and answer the following questions.

a) Estimate the winning probability of a game of Player 1 under each service rule.
b) Estimate the expected length of a game under each service rule.
c) Interpret the results and determine which rule is more competitive.
d) What assumption(s) would make your results valid.

QUESTION 3. (Total 8 marks)

Let $\{X_n, n \geq 1\}$ be a sequence of independent and identically distributed (iid) random variables with common probability distribution

$$X_n = \begin{cases} 1 \quad with\ probability\ \frac{1}{2} \\ -1 \quad with\ probability\ \frac{1}{2}, \end{cases}$$

$for\ n \geq 1. Define$

$$S_0 = 0, S_n = X_1 + X_2 + \cdots + X_n, n \geq 1.$$

The stochastic process $\{S_n, n \geq 1\}$ is called a random walk.

a) Simulate five realizations of the random walk $\{S_n, n \geq 1\}$. (4 marks)
b) Discuss the aspects of the random variable $S_n\ as\ n$ tends to infinity. (4 marks)

QUESTION 4. (Total 8 marks)

Consider a customer who buys beer at regular intervals, say once a week, and every time he has to decide among three brands: A, B and C. From his buying records so far it has been determined that his brand choice in week $n + 1$ depends only on the brand choice in week n, regardless of previous purchases. Let Y_n be the brand he purchases in week n. Then it is obvious that $\{Y_n, n \geq 1\}$ is a discrete time Markov chain with state space $S = \{A, B, C\}$.

The transition probability matrix **P** will look as follows:

$$\mathbf{P} = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{vmatrix} 0.1 & 0.2 & 0.7 \\ 0.2 & 0.4 & 0.4 \\ 0.1 & 0.3 & 0.6 \end{vmatrix} \end{array}$$

Thus, if he purchases beer of brand A in week n, he will purchase beer of brand C in week n+1 with probability 0.7, etc. This type of brand-switching models are used quite often in practice by industries to predict market shares, etc.

a) Simulate 5 possible beer purchases for 10 weeks by assuming that he bought beer brand A for first week. (4 marks)
b) Find the probability that he purchases beer brand A in first week and again buy the same brand in fifth week. (4 marks)

QUESTION 5. (Total 8 marks)

Suppose that the weather can take on only two states: "hot" and "cold". Three out of four hot days are followed by another cold day whereas only one of three cold days is followed by another cold day. On hot days, John is equally likely to wear sandals or flip flops and he never wears boots. On cold days, John will wear boots **80**% of the time and if he does not wear boots, he wears sandals.

Suppose that an observer can just see John's footwear but not the weather.

a) This situation can be modelled with a hidden Markov model (HMM). Define the state space of this HMM.
(4 marks)
b) Formulate the probability transition probability matrix $\boldsymbol{P}$.
(2 marks)
c) If today is a cold day, find the probability that John will wear sandals both today and tomorrow.
(2 marks)