

COURSE OUTLINE: MH3511

Course Title	Data Analysis with Computer		
Course Code	MH3511		
Offered	Study Year 3, Semester 2		
Course Coordinator	Yeo Kwee Poo (Asst Prof)	kweepoo@ntu.edu.sg	6513-7456
Pre-requisites	MH2500 OR BS1008		
AU	3		
Contact hours	Lectures: 26, Laboratories: 24		
Approved for delivery from	AY 2021/22 semester 2		
Last revised	8 Dec 2021, 09:16		

Course Aims

In today's business, data analysis plays an important role in making decisions more scientific and helping the business achieve effective operation. By closely examining data we can find patterns to perceive information, and the information can be used to enhance knowledge. This course provides basic concepts for data analysis with the usage of the R programming language. You will learn the skills of plotting, summarising, making inferences, and presenting various types of data.

Intended Learning Outcomes

Upon successfully completing this course, you should be able to:

1. Evaluate the value of Mathematical functions using R. Writing R program to perform a given algorithm.
2. Distinguish between different types of measurement scales.
3. Explain the meaning of statistical quantities, such as mean, median, variance, etc., and compute the sample values of a given dataset.
4. Use R to construct histogram, boxplot, scatterplot, qq-plot, etc.
5. Construct point and confidence interval estimates for the population parameters using R.
6. Explain the meaning of Type I and Type II errors, and perform statistical hypothesis testing for various types of dataset.
7. Perform statistical inference on categorical dataset.
8. Use parametric methods as an alternative approach to data analysis.
9. Perform linear regression and check for model assumptions.

Course Content

Basic of R programming • Basic R syntax • Write mathematical expression in R language • Variable, vector, matrix and dataframe, and their operations in R • Importing dataset into R, subsetting dataset • Basic loops in R

Describing Data • Basic parameters such as mean, median, standard deviation, variance, inter-quartile range • Boxplot, histogram, stem-leaf plot • Normality checks, qq-plot, outlier, transformation

Statistical Inference • Sample, sampling distribution • Central Limit Theorem • Confidence Interval • Statistical hypothesis testing, Type I and Type II errors, p-value

Categorical Data • Proportion estimate, testing of proportion parameter • Goodness-of-fit test • Two-way contingency table • Paired 2-way contingency table

Multiple Samples • Two independent samples, inference on mean difference • Two dependent samples • Multiple (>2) independent samples, ANOVA test • Multiple (>2) dependent samples

Nonparametric Tests • Quantile test • Wilcoxon rank-sum test • Kruskal-Wallis Test • Sign test, Wilcoxon signed-rank test • Friedman test

Correlation and Regression • Correlation coefficient, its confidence interval and statistical test • Simple linear regression model • Inference on the parameters of linear model • Prediction inference • Model checking

Assessment

Component	Course ILOs tested	SPMS-MAS Graduate Attributes tested	Weighting	Team / Individual	Assessment Rubrics
Continuous Assessment					
Laboratories					
Written Report	1, 2, 3, 4, 5, 6, 7, 8, 9	1. a, b, c, d 2. a, b, d 3. a, b 4. a 5. a	20	individual	See Appendix for rubric
Mid-semester Quiz					
Short Answer Questions 1	1, 2, 3, 4	1. a, b, c, d 2. a, b, d 3. a	15	individual	See Appendix for rubric
Short Answer Questions 2	5, 6, 7	1. a, b, c, d 2. a, b, d 3. a	15	individual	See Appendix for rubric
Examination (2 hours)					
Short Answer Questions	1, 2, 3, 4, 5, 6, 7, 8, 9	1. a, b, c 2. b, d 3. a	50	individual	See Appendix for rubric
Total			100%		

These are the relevant SPMS-MAS Graduate Attributes.

1. Competence

- a. Independently process and interpret mathematical theories and methodologies, and apply them to solve problems
- b. Formulate mathematical statements precisely using rigorous mathematical language
- c. Discover patterns by abstraction from examples
- d. Use computer technology to solve problems, and to communicate mathematical ideas

2. Creativity

- a. Critically assess the applicability of mathematical tools in the workplace
- b. Build on the connection between subfields of mathematics to tackle new problems
- d. Critically analyse data from a multitude of sources

3. Communication

- a. Present mathematics ideas logically and coherently at the appropriate level for the intended audience
- b. Work in teams on complicated projects that require applications of mathematics, and communicate the results verbally and in written form

4. Civic-mindedness

- a. Develop and communicate mathematical ideas and concepts relevant in everyday life for the benefits of society

5. Character

- a. Act in socially responsible and ethical ways in line with the societal expectations of a mathematics professional, particularly in relation to analysis of data, computer security, numerical computations and algorithms

Formative Feedback

For midterm Quizzes: Feedback on common mistakes and the level of difficulty of the problems will be given.

For Project Work: You will receive individual written and/or verbal feedback about their project.

For Final Exam: the Examiner's Report will be provided.

Learning and Teaching Approach

Lectures (26 hours)	This is intended to help you understand the motivation and definitions of the concepts and notions, approaches to solving problems in pursuant to learning outcomes.
Laboratories (24 hours)	This will help to develop your problem solving and computing skills, reinforce the understanding of the concepts and notions.

Reading and References

1 Hothorn & Everitt: A Handbook of Statistical Analysis Using R, 3rd Edition, CRC Press 2014. ISBN-10: 1482204584, ISBN-13: 978-1482204582

2 Michael J. Crawley: Statistics, An Introduction using R, Wiley 2005. ISBN-10: 0470022981, ISBN-13: 978-0470022986

Course Policies and Student Responsibilities

(1) General

You are expected to complete all lab assignments, take the quizzes, and complete the project. You are expected to take responsibility to follow up with course notes, assignments and course related announcements if you are absent.

(2) Absenteeism

Absence from quizzes and examination without a valid reason will affect your overall course grade. Valid reasons include falling sick supported by a medical certificate and participation in NTU's approved activities supported by an excuse letter from the relevant bodies.

(3) Absence Due to Medical or Other Reasons

If you are sick and not able to attend the midterm, you have to submit the original Medical Certificate (or another relevant document) to the administration to obtain official leave. In this case, the missed assessment component will not be counted towards the final grade. There are no make-up midterm.

Academic Integrity

Good academic work depends on honesty and ethical behaviour. The quality of your work as a student relies on adhering to the principles of academic integrity and to the NTU Honour Code, a set of values shared by the whole university community. Truth, Trust and Justice are at the core of NTU's shared values.

As a student, it is important that you recognize your responsibilities in understanding and applying the principles of academic integrity in all the work you do at NTU. Not knowing what is involved in maintaining academic integrity does not excuse academic dishonesty. You need to actively equip yourself with strategies to avoid all forms of academic dishonesty, including plagiarism, academic fraud, collusion and cheating. If you are uncertain of the definitions of any of these terms, you should go to the [Academic Integrity website](#) for more information. Consult

your instructor(s) if you need any clarification about the requirements of academic integrity in the course.

Course Instructors

Instructor	Office Location	Phone	Email
Yeo Kwee Poo (Asst Prof)	SPMS-MAS-04-16	6513-7456	kweepoo@ntu.edu.sg

Planned Weekly Schedule

Week	Topic	Course ILO	Readings/ Activities
1	Basic of R programming <ul style="list-style-type: none"> • Basic R syntax • Write mathematical expression in R language • Variable, vector, matrix and dataframe, and their operations in R 	1	Lecture notes
2	Basic of R programming <ul style="list-style-type: none"> • Importing dataset into R, subsetting dataset • Basic loops in R 	1	Lecture notes / Lab Assignment
3	Describing Data <ul style="list-style-type: none"> • Basic parameters such as mean, median, standard deviation, variance, inter-quartile range • Boxplot, histogram, stem-leaf plot • Normality checks, qq-plot, outlier, transformation 	2, 3, 4	Lecture notes / Lab Assignment
4	Statistical Inference <ul style="list-style-type: none"> • Sample, sampling distribution • Central Limit Theorem • Confidence Interval • 	5, 6, 7	Lecture notes / Lab Assignment
5	Statistical Inference <ul style="list-style-type: none"> • Statistical hypothesis testing, Type I and Type II errors, p-value 	5, 6, 7	Lecture notes / Lab Assignment
6	Categorical Data <ul style="list-style-type: none"> • Proportion estimate, testing of proportion parameter • Goodness-of-fit test 	5, 6, 7	Lecture notes / Lab Assignment
7	Categorical Data <ul style="list-style-type: none"> • Two-way contingency table • Paired 2-way contingency table 	5, 6, 7	Lecture notes / Lab Assignment
8	Multiple Samples <ul style="list-style-type: none"> • Two independent samples, inference on mean difference • Two dependent samples 	5, 6, 7	Lecture notes / Lab Assignment
9	Multiple Samples <ul style="list-style-type: none"> • Multiple (>2) independent samples, ANOVA test • Multiple (>2) dependent samples 	5, 6, 7	Lecture notes / Lab Assignment
10	Nonparametric Tests <ul style="list-style-type: none"> • Quantile test • Wilcoxon rank-sum test • Kruskal-Wallis Test 	5, 6, 8	Lecture notes / Lab Assignment
11	Nonparametric Tests <ul style="list-style-type: none"> • Sign test, Wilcoxon signed-rank test • Friedman test 	5, 6, 8	Lecture notes / Lab Assignment
12	Correlation and Regression <ul style="list-style-type: none"> • Correlation coefficient, its confidence interval and statistical test • Simple linear regression model 	5, 6, 9	Lecture notes / Lab Assignment
13	Correlation and Regression <ul style="list-style-type: none"> • Inference on the parameters of linear model • Prediction inference • Model checking 	5, 6, 9	Lecture notes / Lab Assignment

Appendix 1: Assessment Rubrics

Rubric for Laboratories: Written Report (20%)

Criteria	Standards		
	Fail standard	Pass standard	High standard
Methods of approach	<ul style="list-style-type: none"> Using methods that are irrelevant or do not apply to the given problem. Invoking theorems whose conditions are not satisfied. 	<ul style="list-style-type: none"> Using relevant methods that help solve the problem. Invoking theorems whose conditions are satisfied. 	Finding methods and utilizing theorems that are both relevant and effective
Validity of reasoning	Reasoning is logically invalid.	Reasoning is logically valid.	Reasoning is logically valid and effective.
Clarity of argument	Reasoning is poorly explained or not explained at all.	Reasoning is clear but may contain some gaps.	Reasoning is clear, precise with no or insignificant gaps.

Rubric for Mid-semester Quiz: Short Answer Questions 1 (15%)

Criteria	Standards		
	Fail standard	Pass standard	High standard
Methods of approach	<ul style="list-style-type: none"> Using methods that are irrelevant or do not apply to the given problem. Invoking theorems whose conditions are not satisfied. 	<ul style="list-style-type: none"> Using relevant methods that help solve the problem. Invoking theorems whose conditions are satisfied. 	Finding methods and utilizing theorems that are both relevant and effective
Validity of reasoning	Reasoning is logically invalid.	Reasoning is logically valid.	Reasoning is logically valid and effective.
Clarity of argument	Reasoning is poorly explained or not explained at all.	Reasoning is clear but may contain some gaps.	Reasoning is clear, precise with no or insignificant gaps.

Rubric for Mid-semester Quiz: Short Answer Questions 2 (15%)

Criteria	Standards		
	Fail standard	Pass standard	High standard
Methods of approach	<ul style="list-style-type: none"> Using methods that are irrelevant or do not apply to the given problem. Invoking theorems whose conditions are not satisfied. 	<ul style="list-style-type: none"> Using relevant methods that help solve the problem. Invoking theorems whose conditions are satisfied. 	Finding methods and utilizing theorems that are both relevant and effective
Validity of reasoning	Reasoning is logically invalid.	Reasoning is logically valid.	Reasoning is logically valid and effective.
Clarity of argument	Reasoning is poorly explained or not explained at all.	Reasoning is clear but may contain some gaps.	Reasoning is clear, precise with no or insignificant gaps.

Rubric for Examination: Short Answer Questions (50%)

Criteria	Standards		
	Fail standard	Pass standard	High standard
Methods of approach	<ul style="list-style-type: none">· Using methods that are irrelevant or do not apply to the given problem.· Invoking theorems whose conditions are not satisfied.	<ul style="list-style-type: none">· Using relevant methods that help solve the problem.· Invoking theorems whose conditions are satisfied.	Finding methods and utilizing theorems that are both relevant and effective
Validity of reasoning	Reasoning is logically invalid.	Reasoning is logically valid.	Reasoning is logically valid and effective.
Clarity of argument	Reasoning is poorly explained or not explained at all.	Reasoning is clear but may contain some gaps.	Reasoning is clear, precise with no or insignificant gaps.