Subject Description Form

Subject Code	AMA4601
Subject Title	Statistical Modeling for Discovery
Credit Value	3
Level	4
Pre-requisite	Statistics for Data Science (AMA3631) or equivalent
Exclusion	Statistical Modeling for Discovery (AMA4001)
Objectives	This subject is to acquaint students with Least Square methods and concept of linear regression, correlation, and its applications; to develop the ability to build regression models; to acquaint students with transformations, qualitative variable in the models which broaden the use of linear regression theory; to gain familiarity with use of modern statistical software packages for building a statistical model.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: a) Formulate regression models that describe relationships between variables and understand the models' statistical foundations; b) Perform a complete regression analysis and communicate the results in both statistical and problem-specific terms; c) Evaluate and compare different regression models using formal statistical methods and graphical techniques; d) Present the results using available statistical software; e) Recognize the ethical responsibility of data collection and processing
Subject Synopsis/ Indicative Syllabus	**Uncorrelated data** *Simple Linear regression* Model and assumptions; least squares estimation of parameters; inference on the parameters; coefficient of determination; confidence interval for the mean value of the response variable; prediction interval; sources of variation *Multiple linear regression* An extension of the simple linear regression model; polynomial regression; estimation and inference on the parameters; partial F-tests; visualization for examination of residuals; multicollinearity, diagnostics and solutions; indicator variables; use of indicator variables *Generalized linear model* An extension of the multiple linear regression; exponential family; canonical link; logistic regression; Poisson regression

<table>
<tr><td></td><td>Dependent data

Structured dependence; autocorrelation, sources and effects; detecting the presence of autocorrelation; parameter estimation procedures with autocorrelated errors.

Decomposition models:
Additive models; multiplicative models;

Hierarchical models
Parametric random effects models

Ethical issues in data collection and processing
Professionalism, responsibilities, obligations, roles.</td></tr>
<tr><td>Teaching/Learning Methodology</td><td>The subject will be delivered mainly through lectures and tutorials. The lectures will be conducted to introduce the elements given in the syllabus, which are then reinforced by learning activities involving demonstration, tutorial exercises, assignments, and mini-project.</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Learning Outcomes</td><td>
<table>
<tr><td>Specific assessment methods/tasks</td><td>% weighting</td><td colspan="5">Intended subject learning outcomes to be assessed (Please tick as appropriate)</td></tr>
<tr><td></td><td></td><td>a</td><td>b</td><td>c</td><td>d</td><td>e</td></tr>
<tr><td>1. Assignments</td><td>15%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td></td></tr>
<tr><td>2. Mini-project</td><td>15%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>3. Midterm</td><td>20%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>4. Examination</td><td>50%</td><td>✓</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>Total</td><td>100 %</td><td colspan="5"></td></tr>
</table>

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:
The subject focuses on knowledge, skills and understanding of <u>Statistical Modeling for Discovery</u>, thus, <u>Exam-based assessment</u> is the most appropriate assessment method, including 50% examination. Continuous Assessment comprises of individual assignments (15%), a mini-project (15%), and a midterm (20%) are included so as to keep the students in progress. A written examination is held at the end of the semester.
</td></tr>
<tr><td></td><td>Class contact:</td><td></td></tr>
<tr><td></td><td>• Lecture</td><td>26 Hrs.</td></tr>
</table>

	• Tutorial	13 Hrs.
	Other student study effort:	
	• Assignment	20 Hrs.
	• Mini-project	20 Hrs.
	• Self-study	30 Hrs.
	Total student study effort:	109 Hrs.

Reading List and References

Textbook:

Michael Kutner, Christopher J. Nachtshem, and John Neter	Applied Linear Regression Models, 5th edition	McGraw-Hill, 2005

References:

B.L. Bowerman and R.T. O'Connell	Linear Statistical Models, an applied approach, second edition	Duxbury Press, 2000
D. Montgomery, E. Peck and G. Vining	Introduction to Linear Regression Analysis, fifth edition	Wiley, 2012
	R: A Language and Environment for Statistical Computing, The R Development Core Team Version 2.9.2, 2009	
American Statistical Association	Ethical Guidelines for Statistical Practice	ASA, 1999