# Subject Description Form

| | |
|---|---|
| **Subject Code** | AMA4001 |
| **Subject Title** | Statistical Modeling for Discovery |
| **Credit Value** | 3 |
| **Level** | 4 |
| **Pre-requisite** | Statistics for Data Science (AMA3631) or equivalent |
| **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:<br>a) Formulate regression models that describe relationships between variables and understand the models' statistical foundations;<br>b) Perform a complete regression analysis and communicate the results in both statistical and problem-specific terms;<br>c) Evaluate and compare different regression models using formal statistical methods and graphical techniques;<br>d) Present the results using available statistical software;<br>e) Recognize the ethical responsibility of data collection and processing |
| **Subject Synopsis/ Indicative Syllabus** | *Ethical issues in data collection and processing*<br>Professionalism, responsibilities, obligations, roles.<br><br>*Statistical models*<br>Additive models; multiplicative models; sources of variation; hierarchical models and structured dependence<br><br>*Linear regression*<br>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; test for lack of fit; examination of residuals.<br><br>*Multiple linear regression models*<br>An extension of the linear regression model and as a special case of the general linear model; estimation and inference on the parameters; partial F-tests; polynomial regression.<br><br>*Indicator Variables*<br>Concept of indicator variables; use of indicator variables. |

*Multicollinearity*
The problem of multicollinearity; multicollinearity diagnostics; solutions to multicollinearity.

*Autocorrelation*
Sources and effects of autocorrelation; detecting the presence of autocorrelation; parameter estimation procedures with autocorrelated errors.

*Nonparametric inference*
Permutation and randomization tests. Use of ranks and randomization; robustness.

**Teaching/Learning Methodology**

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.

**Assessment Methods in Alignment with Intended Learning Outcomes**

| Specific assessment methods/tasks | % weighting | Intended subject learning outcomes to be assessed (Please tick as appropriate) | | | | |
|---|---|---|---|---|---|---|
| | | a | b | c | d | e |
| 1. Assignments | 20% | ✓ | ✓ | ✓ | ✓ | |
| 2. Mini-project | 20% | ✓ | ✓ | ✓ | ✓ | ✓ |
| 3. Examination | 60% | ✓ | ✓ | ✓ | | |
| Total | 100 % | | | | | |

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:
The subject focuses on knowledge, skills and understanding of **Statistical Modeling for Discovery,** thus, **Exam-based assessment** is the most appropriate assessment method, including 60% examination. Continuous Assessment comprises of individual assignments (20%) and mini-project (20%) are included so as to keep the students in progress. A written examination is held at the end of the semester.

| | |
|---|---|
| Class contact: | |
| • Lecture | 26 Hrs. |
| • 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 |