



STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

MASTER OF SCIENCE (DATA SCIENCE AND ANALYTICS)

COURSE OUTLINE

DSA 8305: LINEAR MODELS

LECTURER INFORMATION

Name: Dr. Evans Otieno Omondi.

Office: Sangale Campus, Jasiri Staffroom.

Email: eomondi@strathmore.edu.

Aim: To introduce linear models as a broad class of statistical models applying the general principles of likelihood-based inference and further, to enable the student to acquire deep knowledge in the theory and practice of modern regression modelling within the unifying context of generalized linear models. A second aim is to get students acquainted with research questions in this domain.

Course Learning Outcomes:

At the end of this course the student will be able to:

1. Formulate linear models including model building, parameter estimation, hypothesis testing, model validation, and model diagnostics.
2. Apply various linear models to real life data and interpret the results.
3. Critique research papers in the area of linear models and applications in scientific way.
4. Use R and other software to perform data analysis and interpret the results.

Contact Hours: Lecture hours: 35 Lecture hours, 30 Lab hour

Prerequisite: None.

Course content

| Week | Topic/ Sub-topics | Intended Learning Outcome(s) | Class Activity |
|------|--|---|--|
| 1-4 | Theory of general linear models. <ul style="list-style-type: none">• Model construction.• Full and less than full rank models.• Estimation and properties of estimates.• Hypothesis testing. | By the end of this topic the student will be able to <ul style="list-style-type: none">- Describe what model construction entails.- Formulate linear models.- Learn on methods of estimating parameters and their properties- Formulate and test hypothesis.- Perform variable and model selection- Perform model prediction | <ul style="list-style-type: none">• Lecture illustrations• Group discussions.• Questions and answers sessions• Use numerical methods to solve problems.• Practical assessment using R |

| Week | Topic/ Sub-topics | Intended Learning Outcome(s) | Class Activity |
|------------------------|--|---|--|
| | <ul style="list-style-type: none"> Variable and model selection Prediction | - Solve the formulated models using computer software. | |
| 5-8 | Non-linear models <ul style="list-style-type: none"> Nonlinear models via nonlinear least squares. Basic splines and Generalized Additive Models | By the end of this topic the student will be able to <ul style="list-style-type: none"> Describe nonlinear models and nonlinear least squares. Describe splines and GAM models and be able to analyze them. | <ul style="list-style-type: none"> Lecture illustrations Group discussions. Questions and answers sessions Use numerical methods to solve problems. Practical assessment using R |
| CAT 1 (WEEK 8) | | | |
| 9-11 | Regression <ul style="list-style-type: none"> Ridge Regression and Principal Components Regression for collinear data Models for non-independent errors. | By the end of this topic the student will be able to <ul style="list-style-type: none"> Formulate models relating to ridge regression and solve PCA problems. Solve models with non-independent errors. | <ul style="list-style-type: none"> Lecture illustrations Group discussions. Questions and answers sessions Practical assessment using R |
| 12-14 | Semi parametric regression models <ul style="list-style-type: none"> Scatter plot smoothing. Mixed models. Additive models. Interaction models and generalized regression | By the end of this topic the student will be able to <ul style="list-style-type: none"> Perform smoothing relating to linear models. To understand and be able to analyze mixed models, additive models and generalized models. | <ul style="list-style-type: none"> Lecture illustrations Group discussions. Questions and answers sessions Class presentations Practical assessment using R |
| CAT 2 (WEEK 15) | | | |

Course Delivery Methodology:

- Lectures
- Case Studies
- Group Work and Presentations

Course Materials

| Name | Author | Publisher |
|--|-------------------------------|--|
| Analysis of Categorical Data; 2nd Edition | Agresti A. | Springer Berlin Heidelberg, (2011) ISBN: 978-3-642-04897-5, ISBN: 978-3-642-04898-2 |
| An introduction to generalized linear models | Dobson A. J., & Barnett A. G. | Chapman and Hall/CRC, (2018). |
| An R and S-Plus Companion to Applied Regression; 2nd Edition | Fox John | Sage Publication, (2010) ISBN: 141297514X, ISBN: 978-1412975148 |

Academic Assessment

- Homework assignments and quizzes
- Two sit-in CATS, and the final exam.
- The breakdown for each component is as follows:

| Type | Weighting (%) |
|--|---------------|
| Coursework (CATs and Assignments) | 40 |
| End of semester Examination (Final Exam) | 60 |
| Total | 100 |

Communication

- The primary means of communication outside of class will be by email.
- Students must use the university email accounts to receive important University information, including messages related to this class.

Collaborative work

- **Collaboration** in course work is certainly encouraged as this promotes team spirit and group synergy as long provided originality is preserved.

Computational Software:

- The main computational software that I use is R.
- R is open source and is free.

Document Development Software:

- The main document development software that I use is TeX.
- This is also a free software. One has to install the basic MikTeX before installing any TeX interface.