جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Course Syllabus: Linear Models

STAT 230 - Lecture STAT 230

Course Information	
Comprehensive Course Description	**Objective:** This course is an introduction to the formulation and use of linear models (and generalizations) including parameter estimation and inference for such models in a variety of settings. Emphasis will be split between understanding the theoretical foundations of the models and the ability to apply the models to answer scientific questions. **Tentative outline:** 1. Introduction: General context; Linear algebra; Multivariate Gaussian distribution; 2. Linear Regression Models: Simple/Multiple normal linear regression; Likelihood estimators; Geometrical intuition; (Weighted/Generalized) Least squares estimators; 3. Properties of LSE: Distribution of the LSE; Confidence and prediction intervals; Optimality of LSE; Gauss-Markov Theorem; 4. Diagnostics and Testing: Linearity; Homoskedasticity; Gaussianity; Independence; Coefficient of determination R2; Residuals; Outliers; Leverage points; Hypothesis tests; 5. ANOVA: One-way analysis of variance; Orthogonality; F-tests; 6. Model Selection: Sequential (forward/backward/stepwise) model selection; Information Criteria (AIC/BIC/Cp); Cross-validation; 7. Multicollinearity: Diagnostics for detection; Ridge regression; LASSO; 8. Robust Regression: L1 regression; Trimmed least squares; M-estimators; 9. Non-Linear Regression: Newton-Raphson algorithm; Nonlinear least squares; 10. Non-Parametric Regression: Kernel smoothing; Splines; Projection-Pursuit regres- sion; Additive models; backfitting algorithm. 11. Generalized Linear Models: Exponential Families; GLMs; Logistic regression.
Course Description from Program Guide	This course is an introduction to the formulation and use of the general linear model, including parameter estimation, inference and the use of such models in a variety of settings. Emphasis will be split between understanding the theoretical formulation of the models and the ability to apply the models to answer scientific questions.Multivariate models; Inference about independence.
Goals and Objectives	The overall goal of this course is to master the theory of linear models (and their various generalizations covered during the course), their formulation, their estimation and inference, as well as their application to real datasets. In addition, the students will be able to perform simple data analyses in practice using the statistical software R, and be able to understand and correctly interpret the results and outputs (diagnostics, figures, tables, etc.) from R. All the material covered during the course (lectures and homework) will be mastered by the students.
Required Knowledge	Advanced and multivariate calculus, Linear algebra, Probability and Statistics.

Reference Texts	1. Christensen (2011) Plane Answers to Complex Questions: the Theory of Linear Models,Springer; ebook available 2. Wood (2015) Core Statistics, Cambridge University Press; e-book available 3. Seber and Lee (2003) Linear Regression Analysis, Wiley; e-book available 4. Hocking (1996) Methods and Applications of Linear Models: Regression and the Anal- ysis of Variance, Wiley 5. McCullagh and Nelder (1989) Generalized Linear Models, Chapman & Hall/CRC 6. Kariya and Kurata (2004) Generalized Least Squares, Wiley; e-book available 7. Hastie and Tibshirani (1990) Generalized Additive Models, Chapman & Hall/CRC 8. Davison (2003) Statistical Models, Cambridge University Press; e-book available 9. Faraway (2005) Linear Models with R, Chapman & Hall/CRC 10. Faraway (2006) Extending the Linear Model with R, Chapman & Hall/CRC
Method of evaluation	**25.00**% - Homework /Assignments **25.00**% - Midterm exam **50.00**% - Final exam
Nature of the assignments	Assignments consist of weekly homework (theoretical and practical exercises) to be completed individually.

Note

The instructor reserves the right to make changes to this syllabus as necessary.