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

Course Syllabus: Linear Models - STAT 230

Division	Computer, Electrical and Mathematical Sciences & Engineering
Course Number	STAT 230
Course Title	Linear Models
Academic Semester	Fall
Academic Year	2018/2019
Semester Start Date	08/26/2018
Semester End Date	12/11/2018
Class Schedule (Days & Time)	10:30 AM - 12:00 PM | Sun Wed

Instructor(s)

Name	Email	Phone	Office Location	Office Hours
Raphael Georges Huser	raphael.huser@kaust.edu.sa	+966128080682	4125, 1, Al-Khawarizmi (bldg. 1)	

Teaching Assistant(s)

Name	Email

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.
Course Policies	Assignments will be collected at the START of the class on the due date. Late assignments will not be accepted, unless prior arrangements have been made. Staple the pages together, submit the problems in order and write clearly (**illegible handwriting may not be graded**). Students are encouraged to work together on the homeworks, but collaboration with classmates is **strictly limited to discussing problems**, not writing them up, photocopying solutions, or sharing R code.
Additional Information	Class notices and course related information will be posted periodically on the STAT 230 website on Blackboard. Please check regularly for important information. Also, there may be important email communications (like homework hints or a change in the exams date), so it is important to monitor your email account regularly.

Tentative Course Schedule

(Time, topic/emphasis & resources)

Week	Lectures	Topic
1	Sun 08/26/2018 Wed 08/29/2018	Introduction: General context; Linear algebra; Multivariate Gaussian distribution;
2	Sun 09/02/2018 Wed 09/05/2018	Linear Regression Models: Simple/Multiple normal linear regression; Likelihood estimators; Geometrical intuition;
3	Sun 09/09/2018 Wed 09/12/2018	Weighted/Generalized Least squares estimators;
4	Sun 09/16/2018 Wed 09/19/2018	Properties of LSE: Distribution of the LSE; Confidence and prediction intervals; Optimality of LSE; Gauss-Markov Theorem;
5	Sun 09/23/2018 Wed 09/26/2018	Sunday 23rd: Saudi National Day Wednesday 26th: Diagnostics and Testing: Linearity; Homoskedasticity; Gaussianity; Independence;
6	Sun 09/30/2018 Wed 10/03/2018	Coefficient of determination R2; Residuals; Outliers; Leverage points; Hypothesis tests;
7	Sun 10/07/2018 Wed 10/10/2018	ANOVA: One-way analysis of variance; Orthogonality; F-tests;
8	Sun 10/14/2018 Wed 10/17/2018	Sunday 14th: Model Selection: Sequential (forward/backward/stepwise) model selection; Wednesday 17th: Mid-Term Exam
9	Sun 10/21/2018 Wed 10/24/2018	Model Selection: Information Criteria (AIC/BIC/Cp); Cross-validation;
10	Sun 10/28/2018 Wed 10/31/2018	Sunday 28th: Mid-semester break Wednesday 31st: Multicollinearity: Diagnostics for detection; Ridge regression; LASSO;
11	Sun 11/04/2018 Wed 11/07/2018	Robust Regression: L1 regression; Trimmed least squares; M-estimators;
12	Sun 11/11/2018 Wed 11/14/2018	Non-Linear Regression: Newton-Raphson algorithm; Nonlinear least squares;
13	Sun 11/18/2018 Wed 11/21/2018	Non-Parametric Regression: Kernel smoothing; Splines
14	Sun 11/25/2018 Wed 11/28/2018	Non-Parametric Regression: Projection Pursuit regression; Additive models; backfitting algorithm
15	Sun 12/02/2018 Wed 12/05/2018	Generalized Linear Models: Exponential Families; GLMs; Logistic Regression; Quantile Regression
16	Sun 12/09/2018	
17		
18		

Note

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