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

Course Syllabus: Applied Statistics with R - STAT 215

Division	Computer, Electrical and Mathematical Sciences & Engineering
Course Number	STAT 215
Course Title	Applied Statistics with R
Academic Semester	Fall
Academic Year	2019/2020
Semester Start Date	08/25/2019
Semester End Date	12/10/2019
Class Schedule (Days & Time)	02:30 PM - 04:00 PM | Mon Thu

Instructor(s)

Name	Email	Phone	Office Location	Office Hours
Joaquin Ortega Sanchez	JOAQUIN.ORTEGASANCHEZ@KAUST.EDU.SA		4297, 1, Al-Khawarizmi (bldg. 1)	Sun - Wed 2 - 4 pm

Teaching Assistant(s)

Name	Email
TBA	TBA

Course Information

Comprehensive Course Description	This course is an introduction to practical applied statistics with R, a programming language and software environment for statistical computing, and with RStudio, an integrated development environment for R. Topics include introduction to R, data and programming, summarizing data, probability and statistics in R, simple and multiple linear regression, categorical predictors and interactions, model diagnostics, collinearity, variable selection and model building, selected data analyses.
Course Description from Program Guide	This course is an introduction to practical applied statistics with R, a programming language and software environment for statistical computing, and with RStudio, an integrated development environment for R. Topics include introduction to R, data and programming, summarizing data, probability and statistics in R, simple and multiple linear regression, categorical predictors and interactions, model diagnostics, collinearity, variable selection and model building, selected data analyses.
Goals and Objectives	This course will provide students with an introduction to applied statistics with R and RStudio. Specific objectives: first, to become experienced with the high-level programming language R, an environment for data analysis and graphics, and the use of the RStudio interface and its functionalities to perform reproducible research projects. Second, to become experienced with basic statistical techniques, rigorously implemented in R. Third, to begin the study of more advanced statistical methods on the basis of case studies from challenging real-world problems.
Required Knowledge	The material provided in introductory graduate courses on calculus, basic statistical theory and probability.

Reference Texts	Michael J. Crawley (2013). The R Book, Second Edition, John Wiley & Sons. Peter Daalgard (2008). Introductory Statistics with R, Springer. Ugarte, M.D., Militino, A.F., Arnholt, A.T. (2008). Probability and Statistics with R. CRC Press. **Additional Bibliography** John Maindonald, W. John Braun (2010). Data Analysis and Graphics Using R – an Example-Based Approach, Third Edition, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press. Frank E. Harrell, Jr. (2015). Regression Modeling Strategies, Second Edition, Springer Series in Statistics, Springer. Christopher Gandrud (2015). Reproducible Research with R and RStudio, Second Edition, The R Series, Chapman & Hall/CRC Press. Thomas Rahlf (2017). Data Visualization with R: 100 Examples, Springer. Yangchang Zhao (2013). R and Data Mining: Examples and Case Studies, Academic Press.
Method of evaluation	**35.00**% - Final exam **25.00**% - Midterm exam **40.00**% - Homework /Assignments
Nature of the assignments	Problem lists (5-6 during the course) as homework will be posted through the blackboard page (40%). Solutions should be sent by email or using the blackboard application on the day they are due. Raw computer output is not acceptable. Make it clear what parts of the output are relevant and show how they answer the questions posed. You are encouraged to work together on the homework, but collaboration with classmates is strictly limited to discussing problems, not writing them up or sharing R code. There will be a midterm exam [25%] around week 7 and a final exam [35%] at the end of the term. Precise dates will be announced later; the midterm will be outside of class hours to allow for more time and it could have a take-home component as well.
Course Policies	Attendance is required. Assignments will be posted on the blackboard application on the announced dates throughout the course. They should be sent by email or using the blackboard application on the day that they are due. Raw computer output is not acceptable. Make it clear what parts of the output are relevant and show how they answer the questions posed. Late assignment submissions will not be accepted unless prior arrangements have been made (except in university established cases of illness or emergency). All homework assignments must be neatly typed (LaTeX or RMarkdown are recommended). All projects and homework assignments are required. If you dispute your grade on any homework or project, you may request a regrade (from the TA/instructor for the assignments or the instructor for the exams) only within 48 hours of receiving the graded exam. Incomplete (I) grade for the course will only be given under extraordinary circumstances such as sickness, and these exceptional circumstances must be verifiable. The assignment of an (I) requires first an approval of the Dean and then a written agreement between the instructor and student specifying the time and manner in which the student will complete the course requirements. Collaboration and checking answers on assignments are allowed and encouraged. Of course, copying assignments is not tolerated. In brief, you are allowed to collaborate on all homework problems according to the following rules: you must first attempt to solve each problem on your own. If you get stuck, you can then talk to any student currently enrolled in the class about the issue, as well as the instructor or the TA. However, solutions and R code should not be exchanged (i.e., you still must work through the details of the problem after you have gotten help, write the final answers alone, and understand them thoroughly). Use of cellular phones in the classroom is not allowed.
Additional Information	

Tentative Course Schedule *(Time, topic/emphasis & resources)*		
Week	**Lectures**	**Topic**
1	Mon 08/26/2019 Thu 08/29/2019	Getting started with R. Essentials of the R language. R objects. RStudio
2	Mon 09/02/2019 Thu 09/05/2019	Programming in R. R functions
3	Mon 09/09/2019 Thu 09/12/2019	Graphs with R. Tools for reproducible research
4	Mon 09/16/2019 Thu 09/19/2019	Graphs with ggplot2
5	Mon 09/23/2019 Thu 09/26/2019	Saudi National Day Statistical models. Basic inference concepts
6	Mon 09/30/2019 Thu 10/03/2019	Classical statistical tests. Bootstrap Regression analysis
7	Mon 10/07/2019 Thu 10/10/2019	Regression analysis
8	Mon 10/14/2019 Thu 10/17/2019	Regression analysis Analysis of variance (Anova)
9	Mon 10/21/2019 Thu 10/24/2019	Analysis of variance (Anova) Simple and multiple analysis of covariance
10	Mon 10/28/2019 Thu 10/31/2019	Mid-semester break Simple and multiple analysis of covariance
11	Mon 11/04/2019 Thu 11/07/2019	Generalized linear models Logistic regression
12	Mon 11/11/2019 Thu 11/14/2019	Count data. Proportions Ordinal logistic regression
13	Mon 11/18/2019 Thu 11/21/2019	Probability distributions and random number generation Matrix algebra. Principal component analysis
14	Mon 11/25/2019 Thu 11/28/2019	Simulation Introduction to the Montecarlo method
15	Mon 12/02/2019 Thu 12/05/2019	Parametric survival modeling
16	Mon 12/09/2019	Exams

Note

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