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

Course Syllabus: Applied Statistics and Data Analysis - STAT 210

Division	Computer, Electrical and Mathematical Sciences & Engineering
Course Number	STAT 210
Course Title	Applied Statistics and Data Analysis
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 | Mon Thu

Instructor(s)

Name	Email	Phone	Office Location	Office Hours
Hernando Catequista Ombao	hernando.ombao@kaust.edu.sa	+966128080213	4126, 1, Al-Khawarizmi (bldg. 1)	TBD.

Teaching Assistant(s)

Name	Email

Course Information

Comprehensive Course Description	For students outside the Statistics Program wishing to obtain an introduction to statistical methods for data analysis. This course is Statistics 110 for AMCS and Statistics students. This course provides fundamentals of probability and statistics for data analysis in research. Topics include data collection, exploratory data analysis, random variables, common discrete and continuous distributions, sampling distributions, estimation, confidence intervals, hypothesis tests, elementary simulation and bootstrapping, distribution-free techniques, linear regression, analysis of variance, two-way tables, and data analysis using statistical software. Formal modeling of data and formal inference will be covered.
Course Description from Program Guide	Provides fundamentals of probability and statistics for data analysis in research. Topics include data collection, exploratory data analysis, random variables, common discrete and continuous distributions, sampling distributions, estimation, confidence intervals, hypothesis tests, linear regression, analysis of variance, two-way tables and data analysis using statistical software.
Goals and Objectives	By the end of this course, the students are expected to have mastered the following: (1.) Statistical visualization methods (2.) Framework for statistical modeling of continuous and discrete-valued types of data (3.) Formal inferential procedures (4.) Various data analytic techniques
Required Knowledge	Undergraduate mathematics including calculus and basic matrix algebra.
Reference Texts	Shahbaba, B. (2012). Biostatistics with R: An Introduction to Statistics Through Biological Data. Springer. ISBN-13: 978-1461413011. ISBN-10: 146141301X. (Required). Weisberg, S. (2013). Applied Linear Regression. Wiley, 4th edition. ISBN: 978-1-118-38608-8. (Required).

Method of evaluation	**10.00**% - Oral presentation **40.00**% - Research Project **30.00**% - Midterm exam **20.00**% - Homework /Assignments
Nature of the assignments	Problem sets [20%] (2-3 for the entire semester) will complement the work in class. They should be handed in class at the start of the lectures on the day that they are due. Arrange the pages in the proper order and staple them (we are not responsible for lost pages). 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 one in-class midterm exam [30%]. Date TBD; time will be outside of class hours to allow for more time. The exam could have a take-home component as well. Each student is required to complete an individual project and submit a final report. More information on the final project will be given in a separate document and will be distributed in class. Grading will include the proposal, draft and final paper. [40%] Each student is required to present his/her paper during the finals week [10%].
Course Policies	Students who miss any homework or exam should expect a grade of zero on that assignment. Your grade on any homework or exam may be submitted for reconsideration only within two business days of receiving the graded exam. If you are unable to take an exam or complete an assignment on time due to circumstances beyond your control, please e-mail me within 24 hours for appropriate arrangements. If you know ahead of time that you will have a university excused absence, homework assignments are due before you leave, and exams will be made up after you return.
Additional Information	The materials used in this course are copyrighted. By materials, I mean all materials generated for this class including syllabi, exams, course notes, computer code, and examples. You do not have the right to copy the handouts or distribute them, unless I expressly grant permission. This syllabus should be taken as a fairly reliable guide for the course content. However, you cannot claim any rights from it and in particular I reserve the right to change due dates or the methods of assessment. Official announcements will ALWAYS be those made in class.

Tentative Course Schedule

(Time, topic/emphasis & resources)

Week	Lectures	Topic
1	Mon 08/27/2018 Thu 08/30/2018	Introduction to data analysis Introduction to R computing
2	Mon 09/03/2018 Thu 09/06/2018	Visualization: scatter plots, histograms, boxplots Data summaries: mean, median, quartiles, variance, cross-correlations
3	Mon 09/10/2018 Thu 09/13/2018	Comparing two independent populations (2 sample t-test) Paired t-test Comparing population proportions
4	Mon 09/17/2018 Thu 09/20/2018	Comparing many populations (ANOVA) F-test Multiple testing
5	Mon 09/24/2018 Thu 09/27/2018	Design of experiments Sample size calculations Power analysis
6	Mon 10/01/2018 Thu 10/04/2018	Simple linear regression Model formulation Formal inference Model diagnostics
7	Mon 10/08/2018 Thu 10/11/2018	Overview of linear models (matrix notation) Midterm exam to be given on this week (outside of class hours)
8	Mon 10/15/2018 Thu 10/18/2018	Multiple linear regression Variable selection Model diagnostics Data analysis and interpretation
9	Mon 10/22/2018 Thu 10/25/2018	Nonparameteric methods for comparing populations Permutation and randomization tests
10	Mon 10/29/2018 Thu 11/01/2018	Nonparametric function estimation Kernel smoothing
11	Mon 11/05/2018 Thu 11/08/2018	Binary regression models Logistic regression Probit regression
12	Mon 11/12/2018 Thu 11/15/2018	Regression models for count data Tests for independence for categorical data
13	Mon 11/19/2018 Thu 11/22/2018	Generalized linear models (GLM) Power analysis and sample size calculations for GLM
14	Mon 11/26/2018 Thu 11/29/2018	Longitudinal data analysis Modeling samples of curves
15	Mon 12/03/2018 Thu 12/06/2018	Modern inferential methods including the bootstrap for regression
16	Mon 12/10/2018	Preparing the statistical sections of grant applications
17		
18		

Note

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