



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	2019/2020
Semester Start Date	08/25/2019
Semester End Date	12/10/2019
Class Schedule (Days & Time)	09:00 AM - 10:30 AM Mon Thu

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

Teaching Assistant(s)	
Name	Email
TBA	TBA

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: <ul style="list-style-type: none"> (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	<p>Shahbaba, B. (2012). Biostatistics with R: An Introduction to Statistics Through Biological Data. Springer. ISBN-13: 978-1461413011. ISBN-10: 146141301X. (Required).</p> <p>Weisberg, S. (2013). Applied Linear Regression. Wiley, 4th edition. ISBN: 978-1-118-38608-8. (Required).</p> <p>Additional references</p> <p>Dalgaard, P. (2008) Introductory Statistics with R. Springer.</p> <p>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.</p> <p>Crawley, Michael J. (2015) Statistics, An Introduction Using R, Second Edition. J. Wiley & Sons.</p>
Method of evaluation	<p>30.00% - Course Project(s)</p> <p>30.00% - Homework /Assignments</p> <p>20.00% - Exam 2</p> <p>20.00% - Exam 1</p>
Nature of the assignments	<p>Problem sets [30%] will complement the work in class. 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. 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.</p> <p>There will be two in-class exams [20% each] on weeks 7 and 14. Precise dates will be announced later; time will be outside of class hours to allow for more time. The exam could have a take-home component as well.</p> <p>Each student is required to complete a 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, final paper and oral presentation [30%].</p>
Course Policies	<p>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.</p> <p>Use of cellular phones in the classroom is not allowed.</p>
Additional Information	<p>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.</p> <p>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.</p>

Tentative Course Schedule

(Time, topic/emphasis & resources)

Week	Lectures	Topic
1	Mon 08/26/2019 Thu 08/29/2019	Introduction to data analysis Introduction to R computing
2	Mon 09/02/2019 Thu 09/05/2019	Visualization: scatter plots, histograms, boxplots Data summaries: mean, median, quartiles, variance, cross-correlations
3	Mon 09/09/2019 Thu 09/12/2019	Comparing two independent populations (2 sample t-test) Paired t-test Comparing population proportions
4	Mon 09/16/2019 Thu 09/19/2019	Saudi National Day Comparing many populations (ANOVA), F-test, Multiple testing
5	Mon 09/23/2019 Thu 09/26/2019	Design of experiments Sample size calculations Power analysis
6	Mon 09/30/2019 Thu 10/03/2019	Simple linear regression Model formulation Formal inference Model diagnostics
7	Mon 10/07/2019 Thu 10/10/2019	Overview of linear models (matrix notation) Exam 1 to be given on this week (outside of class hours)
8	Mon 10/14/2019 Thu 10/17/2019	Multiple linear regression Variable selection Model diagnostics Data analysis and interpretation
9	Mon 10/21/2019 Thu 10/24/2019	Nonparametric methods for comparing populations Permutation and randomization tests
10	Mon 10/28/2019 Thu 10/31/2019	Nonparametric function estimation Kernel smoothing
11	Mon 11/04/2019 Thu 11/07/2019	Binary regression models Logistic regression Probit regression
12	Mon 11/11/2019 Thu 11/14/2019	Regression models for count data Tests for independence for categorical data
13	Mon 11/18/2019 Thu 11/21/2019	Generalized linear models (GLM) Power analysis and sample size calculations for GLM
14	Mon 11/25/2019 Thu 11/28/2019	Longitudinal data analysis Modeling samples of curves Exam 2 to be given on this week (outside of class hours)
15	Mon 12/02/2019 Thu 12/05/2019	Modern inferential methods including the bootstrap for regression
16	Mon 12/09/2019	Oral presentations

Note

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