

36-202 (Fall 2018)

Methods for Statistical Data Science

Location and Times:

Fall 2018

Sunday & Tuesday 4:30 - 5:50 PM

CMB 1185

Instructor Information:

Taeyong Park, Ph.D.

✉taeyongp@andrew.cmu.edu

Office: CMB 2191

Office Hours: Sunday & Tuesday 9:30 - 11:00 AM

TA: Tehseen Niaz (✉tniaz@andrew.cmu.edu)

CA: Faree Mohammed (✉fareedum@andrew.cmu.edu)

TA / CA Office Hours: TBD

1. Course Description

This is the second course in the sequence of statistical data analysis for Information Systems majors. Building on probability theory, core concepts in statistics, and basic computer skills covered in 36-201 (or 70-207), you are expected to learn more advanced statistical methods and computational skills for rigorous statistical analysis. You are also expected to use the methods and skills to conduct your own data analysis. To this end, first, you will begin with reviewing confidence intervals and hypothesis testing and apply these concepts to t tests, z tests, ANOVA and χ^2 tests. And then, you will learn mathematical foundations for linear regression analysis and the ordinary least squares estimator (OLS). In comparison to 36-201 (or 70-207), the assumptions required for linear regression analysis and the OLS will be emphasized, which lead into logistic regression analysis and maximum likelihood estimation (MLE). Finally, you are expected to come up with your own research question and a dataset for your question. You will conduct a data analysis and present your analysis result as your term project. Throughout the semester, you will use statistical environment R (<http://www.r-project.org/>). R is widely used in business, economics, biostatistics, social science, and many other fields. It is a powerful programming language for statistical analysis and data visualization. You will use R to manage and analyze data.

2. Prerequisites

- 36-201 (or 70-207).
- Prior knowledge of R is required.

3. Textbooks

- David R. Anderson et al. 2017. *Essentials of Statistics for Business and Economics*. 8th Edition. CENGAGE Learning.
- John Fox. 2015. *Applied Regression Analysis and Generalized Linear Models*. 3rd Edition. SAGE Publications, Inc.
- Several handouts will be provided by the instructor, including Aoun et al. 2018 “Feedback for thought: examining the influence of feedback constituents on learning experience.” *Studies in Higher Education* 43(1): 72-95 and Khansa et al. 2015. “Understanding Members’ Active Participation in Online Question-and-Answer Communities: A Theory and Empirical Analysis.” *Journal of Management Information Systems* 32(2): 162-203.

4. Requirements and Evaluation (Subject to change)

Reading: No grade on reading. But I recommend you to do the assigned reading as described in the course outline below.

Attendance (5%): Every class, your attendance will be recorded. If you miss one class, you will lose 0.5 % of your final grade, if you miss two classes, you will lose 1 %, and if you miss ten or more classes, you will lose all the attendance credit which is 5 % of your final grade. You are expected to attend class on time and remain in class. If you arrive more than 10 minutes late to class, you will not receive credit. If you leave in the middle of class, you will not receive credit.

Exam 1 (20%): In-class exam. This exam covers the material discussed until the exam day. Further details will be announced when the exam approaches.

Exam 2 (20%): In-class exam. This exam covers the material discussed after Exam 1 and until the exam day. Further details will be announced when the exam approaches.

Exam 3 (20%): In-class exam. This exam covers the material discussed after Exam 2 and until the exam day. Further details will be announced when the exam approaches.

Research Project Proposal and Presentation (15%): You are required to submit your research project proposal (at least 7 pages, excluding references; 12 points, double space, and 1-inch mar-

gins all around). First, your proposal must include a clear research question that can potentially be answered by a data analysis. Thus, you must consider data availability when you come up with your research question. Second, your proposal must explain why your research question is important, what related studies exist, and how your research is different from those existing studies. Third, you need to build your research hypothesis and provide theoretical justification for your hypothesis. Fourth, you must specify your dependent variable, explanatory variable(s), and control variable(s) by which you can test your hypothesis. Finally, your proposal needs to clarify your plan about how to collect the data you will use to measure your variables. You will also present your project proposal using Power Point or any other presentation software for about 15 minutes. Your presentation skill and the quality of presentation material as well as its content will be evaluated.

Paper with Data Analysis and Presentation (20%): As a second step toward completing your project, you are required to submit a draft of your paper (at least 10 pages, excluding references; 12 points, double space, and 1-inch margins all around). In addition to your research question, explanation of why it is an important question, and literature review, your draft must include details of your data, statistical model, analysis results, and R scripts. Your statistical model must be from one of the topics you learned during the first half of this course, depending on your question and data. You must justify your specification of statistical model. Typically, a statistical model is specified by the choice of variables such as the dependent variable, explanatory variable(s), and control variable(s) and the choice of the type of model (i.e., why logistic regression is used rather than linear regression or why ANOVA test is used rather than alternative methods). Besides your draft, you should submit an R file and data so that I can replicate your analysis. You will also present your paper using Power Point or any other presentation software for about 15 minutes. Your presentation skill and the quality of presentation material as well as its content will be evaluated.

Letter Grade Distribution:

≥ 90.00	A
80.00 - 89.99	B
70.00 - 79.99	C
60.00 - 69.99	D
≤ 59.99	F

5. Grade Appeal Policies

For problem set grade appeals, you should first send an email to CA with a note about why you think your grade should be reconsidered. CA will respond to you with his decision. If you do not agree with his decision regarding your appeal, you can contact me.

6. Electronic Devices and Punctuality

I expect you to be respectful to me and your fellow students to create an environment that is most conducive to learning.

- You will often use your laptop or desktop during class this semester. However, this does not mean that you can feel free to use the computer for whatever you want. It is important to **use it only for class purposes** so that you will not distract yourself and you will not disrupt your classmates. Furthermore, **your cell phone must be turned off** during class. If there is an emergency that might oblige you to be contacted, please talk to me before class. I quote the following passage from the Qatar Business Administration Program Classroom Conduct, which, I believe, must apply to other programs as well:

– Laptops are to be closed. When class is in session, you may use your laptop only as directed by your professor. You should not check email, tweet, text, play games, or surf the Internet, any activity that diminishes your or your classmates' engagement with the classroom content and process. If you are unsure whether a given activity is appropriate, ask your professor. This policy extends to all electronic devices. Be sure that your phones and tablets are silenced and stowed before the class begins. Professors may add specific limits on the sharing or use of personal electronics in exam situations.

- **You must come to class on time and remain in class** once the class has begun. I quote the following passage from the Qatar Business Administration Program Classroom Conduct, which, I believe, must apply to other programs as well:

– In common business culture, punctuality is an important part of showing respect for your colleagues and business partners. Showing up late for a meeting tells the others involved that you do not place much value on their time. QBA students will demonstrate respect for their courses, classmates and professors by arriving for class early enough to get settled and prepared before the scheduled meeting time.

7. Office Hours and Appointments

In general, I encourage you to come to my office whenever you have questions or concerns regarding this course. That said, in order to arrange our meetings efficiently, I have some policies. The best time to meet with me is my office hours: Sunday & Tuesday 9:30 - 11:00 AM (subject to change). During my office hours, I always keep my door open. In addition to my office hours, you are welcome to come in and meet with me if my door is open and I'm not meeting with someone. If my door is closed, it means that I am not available (out of the office, working on research or teaching prep, etc). I encourage you to set up an appointment to make sure I will be available if you want to see me other than during my office hours.

8. Academic Integrity

You should comply with the academic integrity policy. You are required to refer to CMU's general policies on cheating and plagiarism: <http://www.cmu.edu/academic-integrity/valuing/index.html>. Violations of CMU's general policies on cheating and plagiarism carry a range of consequences: <http://www.cmu.edu/academic-integrity/understanding/index.html>.

9. Disability Resources and Health and Well-being

You can find information about disability-related accommodations on <https://scotty.qatar.cmu.edu/health-and-wellness/medical-accommodations/>. You may also consult me or CMUQ staff (Office of Health and Wellness) regarding learning disabilities, health, and wellness.

10. Course Outline (Subject to change)

Date	Topic	Reading
Sep 2, 4	Introduction and Review of Core Concepts	ASWCC 7, 8, 9
Sep 9	Review of Core Concepts	ASWCC 7, 8, 9
Sep 11	Core Concepts and Comparing Two Groups (t / z tests)	ASWCC 7, 8, 9, 10.1-10.2, 10.4
Set 16	Comparing Two Groups (t / z tests)	ASWCC 10.1-10.2, 10.4
Sep 18	Comparing Three or More Groups (ANOVA, χ^2 tests)	ASWCC 12.1, 13.1-13.2
Sep 23	Comparing Three or More Groups (ANOVA, χ^2 tests)	ASWCC 12.1, 13.1-13.2
Sep 25	Exam 1	
Sep 30, Oct 2	Linear Regression and OLS	ASWCC 14, 15; Fox 5, 6
Oct 7, 9	Linear Regression and OLS	ASWCC 14, 15; Fox 5, 6
Oct 14	Linear Regression and OLS	ASWCC 14, 15; Fox 5, 6
Oct 16	Exam 2	
Oct 21, 23	Logistic Regression and MLE	Fox 14.1
Oct 28, 30	Fall Break	
Nov 4, 6	Logistic Regression and MLE	Fox 14.1
Nov 11	Exam 3	
Nov 13	Research Design, Data Collection, Writing a Proposal	Handouts
Nov 18	Research Design, Data Collection, Writing a Proposal	Handouts
Nov 20	Research Proposal Due and Presentation (Group 1)	
Nov 25	Research Proposal Due and Presentation (Group 2)	
Nov 27	Individual Meeting I (Group 1)	
Dec 2	Individual Meeting I (Group 2)	
Dec 4	Paper Due and Presentation (Group 1)	
Dec 9	Paper Due and Presentation (Group 2)	
Dec 11	Individual Meeting II (Groups 1 & 2)	