MA256 Course Guide

Update: 2020-07-22

Task

Develop critical statistical thinkers who can generate precise research questions; identify, collect, and analyze relevant data; and translate this analysis to a complete and correct response that answers the original question.

Purpose

Our daily lives are full of question whose answers should be informed by the growing amount of data surrounding us. Politicians, scientists, sports teams, and military leaders are all looking toward quantitative analysis to provide them with unbiased information with which to make the correct decision. In this course you will gain valuable experience using data in an investigation process to collect, analyze, and report your results to help answer relevant research questions.

You will gain experience in not only completing statistical analysis but also asking the important follow-up questions about why you are getting these results and how you can expect them to change if the experiment was repeated. You will also gain experience in reading, comprehending, and reproducing the results of published scientific papers. All of these tools will help to equip you to be successful in the course project where you are provided the opportunity to do research on a question that interests you. Perhaps you will discover, like some of your fellow cadets that, indeed, more cadets do report to sick call on Mondays.

Grading

Below is the point break-down for this course:

Category	Points
Homework	**350**
Block 1	50
Block 2	125
Block 3	125
Block 4	50
Midterm	**150**
Course Project	**250**
Proposal	25
IPR 1	25
IPR 2	25
Presentation	75
Report	100
TEE	**250**
Total	**1000**

How to use this guide:

The purpose of this course guide is to give you a single document that lists what you need to complete for this course. Given the nature of the learning environment we are in this semester, you can expect that there will be changes to this guide. I will attempt to notify you whenever there are large updates but the version

available on my GitHub (and at the top of the Teams) will always be the latest version.

You will see that there are "Before Class Activities" and "After Class Activities." Hopefully these are pretty self-explanatory. I include "After Class Activities" because I feel that a lot of the "homework" you have will be easier if you complete/think about right after a given class period as opposed to right before the next. I won't be checking though so you do what works best for you.

I have adopted a "Read - Watch - Do" model for laying out your activities. These tasks are the minimum you are expected to complete. There will be some videos assigned that are located under the "Modules" section under our course in WileyPlus. There are a lot more videos available on WileyPlus and please feel free to watch extra videos if it will aid in your understanding. Your textbook also has a lot of exercises if you want some extra practice.

You will see references to example code in many of the lessons. I have seperated out this code to shorten the course guide a little bit. These files are available for you in the main folder of the MA256 GitHub repository. You will see there are **.rmd** files and **.pdf** files. Hopefully you are familiar with PDF files but these RMD (or RMarkdown) files are what I use to produce the PDF files. You will be expected to learn how to use RMarkdown in this course so these files will provide several examples for you (in addition to the countless examples on the internet). You will also notice that I haven't hidden any code from you, even when it's ugly and complicated. I want you to see the types of things you can do with *R*.

I want to instill a sense of responsibility and ownership of your learning in this course. I have a high expectation of your level of mastery of the concepts, you do what is necessary to get yourself to that level.

External Resources:

There are a lot of resources available to you on the internet on the topics we will cover, but I've included a few here that I really like. **Please** take a few minutes to look at these in the beginning of the semester. You never know when a few minutes of seeing what's out there now will save you hours finding an answer later.

- Take a look at Hadley Wickham's book *R for Data Science* which is available at the link below. This is a great resource for your coding questions as it was written by the guy who wrote a large part of the *Tidyverse*. https://r4ds.had.co.nz/

- *Statistical Inference via Data Science*: I recently found this book (and the associated *ModernDive* and *Infer* libraries) but it's got much of the same simulation-based ethos that we have in this course. There are a lot of code examples and descriptions. https://moderndive.com/index.html

- *Introductory Statistics with Randomization and Simulation*: This is a free textbook that covers a lot of the material from the course in a slightly different manner. This website also offer videos, slides, and labs that may help you understand material better. https://www.openintro.org/book/isrs/

Block 1: Question Formulation and Data Exploration

Lesson 1: Course Introduction and RStudio Familiarization

Objectives:

- Introduction to MA256 and course expectations
- Installation and familiarization with RStudio and R
- Tidyverse Tutorial

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Preliminary 1 (P.1)
- Watch: WileyPlus - Videos P.1.1 - P.1.4 (Optional but recommended)
- Do: Consult *Introduction to RStudio and Tidyverse* (AKA *Tidyverse Tutorial*) and follow the "Download R and RStudio" instructions.

After Class Activities:

- Ensure that R/RStudio is installed and functional on your computer.
- Carefully go through the *Tidyverse Tutorial*. Attention now will save you a great deal of time later.
- Consider the dataset that we discussed in class today, what sort of research questions could you ask based on the data and variables present in this dataset. No need to answer the questions, but develop three good research questions before next class.

Lesson 2: Research Questions

Objectives:

- Understanding types of and relationships between variables
- Formulation of specific and focused research questions
- Investigation of the "state of the art"
- Presentation of research in a literature review

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Chapter 4
- Watch: WileyPlus - Videos 4.1.1 - 4.1.5 and Videos 4.2.1 - 4.2.3
- Do: Exploration 4.2 - Problems 1 - 14
- Do: Prepare the three research questions referenced in **Lesson 1 - After Class Activities**.

After Class Activities:

- Pick one of your group's research questions and prepare a literature review for at least two scholarly articles. One submission per group. This will be due at the start of **Lesson 4**.

Lesson 3: Data Exploration - 1

Objectives:

- Understand terminology used in describing distributions
- Understand the information conveyed in a boxplot and five-number summary
- Create appropriate data visualizations in *R*
- Understand reasons for and conclusions drawn from data exploration

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Preliminary 2 (P.2)
- Read: *Introduction to Statistical Investigations* - Section 6.1
- Read: **Project Guidance** document on GitHub under the "Course Project Information" folder
- Watch: WileyPlus - Videos P.2.1 - P.2.5 (Optional but recommended)
- Do: Work on the literature review referenced in **Lesson 2 - After Class Activities**.
- Do: Follow the **Lesson 3 Companion** document on GitHub.

After Class Activities:

- Finalize and submit the literature review referenced in **Lesson 2 - After Class Activities**.

Lesson 4: Data Exploration - 2

Objectives:

- Demonstate proficiency of data exploration skills

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Preliminary 3 (P.3)
- Watch: No videos for this lesson.
- Do: Ensure you have the *NYPD Arrest* data loaded for this lesson.
- Do: Literature review from **Lesson 2** is due before the start of class.

After Class Activities:

- Finalize your work from class. Submission is due before the start of **Lesson 5**. One submission per group.

Block 2: Questions of Single Variables

Lesson 5: Single Categorical Variable - 1

Objectives:

- Understand chance models
- Calculate and utilize various "strength of evidence" measures
- Understand what affects the strength of evidence
- Apply the simulation-based and theory-based approach to solving these problems

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Chapter 1
- Watch: WileyPlus - Videos 1.5.1 - 1.5.3
- Do: *Introduction to Statistical Investigations* - Exploration 1.4
- Do: Submit your work from **Lesson 4** before the start of class.
- Do: Follow the **Lesson 5 Companion** document on GitHub.

After Class Activities:

- Ensure you understand the content discussed in class today... you'll get a chance to prove it next class.

Lesson 6: Single Categorical Variable - 2

Objectives:

- Demonstrate proficiency in answering research questions involving a single categorical variable.

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Chapter 3.1 - 3.2
- Watch: WileyPlus - Videos 3.1.1 - 3.1.3, 3.2.1 - 3.2.2
- Do: *Introduction to Statistical Investigations* - Exploration 3.2, Problems 1 - 11
- Do: Follow the **Lesson 6 Companion** document on GitHub.

After Class Activities:

- Finalize your work from class. Submission is due before the start of **Lesson 7**. One submission per group.

Lesson 7: Single Quantitative Variable - 1

Objectives:

- Understand the concepts of convenience sample, random sample, and sampling variability
- Apply the simulation-based and theory-based approach to solving these problems
- Understand the two types of statistical errors and relate them to the selection of a significance level

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Chapter 2
- Watch: WileyPlus - Videos 2.2.4 - 2.2.5, Videos 2.3.3/2.3.4 - 2.3.5
- Do: *Introduction to Statistical Investigations* - Exploration 2.1B
- Do: Submit your work from **Lesson 6** before the start of class.
- Do: Follow the **Lesson 7 Companion** document on GitHub.

After Class Activities:

- Ensure you understand the content discussed in class today... you'll get a chance to prove it next class.

Lesson 8: Single Quantitative Variable - 2

Objectives:

- Demonstrate proficiency in answering research questions involving a single quantitative variable

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Chapter 3.3 - 3.5
- Watch: WileyPlus - Videos 3.3.1 - 3.3.2, 3.4.1.1 - 3.4.1.4
- Do: *Introduction to Statistical Investigations* - Exploration 3.5, Problems 1 - 6
- Do: Follow the **Lesson 8 Companion** document on GitHub.

After Class Activities:

- Finalize your work from class. Submission is due before the start of **Lesson 9**. One submission per group.

Block 3: Questions of Multiple Variables

Lesson 9: Two Quantitative Variables - 1

Objectives:

- Describe the relationship between two quantitative variables using direction, form, strength, and ununsual obervations
- Under the computation of residuals and the formulation of a least squares regression line
- Apply the simulation- and theory-based approach to solving these problems

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Chapter 10
- Watch: WileyPlus - Videos 10.5.1 - 10.5.4
- Do: Submit your work from **Lesson 8** before the start of class.
- Do: Follow the **Lesson 9 Companion** document on GitHub.

After Class Activities:

- Review the material from Chapter 10. This was a lot for one day but we will be practicing more next class.

Lesson 10: Two Quantitative Variables - 2

Objectives:

- Demonstrate proficiency in building simple linear regression models

Before Class Activities:

- Review: *Introduction to Statistical Investigations* - Chapter 10
- Do: Follow the **Lesson 10 Companion** document on GitHub.

After Class Activities:

- Finalize your work from class. Submission is due before the start of **Lesson 11**. One submission per group.

Lesson 11: Multiple Regression

Objectives:

- Understand the motivation for including additional variables in causal models
- Understand and apply the definition of confounding variable
- Build multiple linear regression models in R

Before Class Activities:

- Read: *Advanced Statistical Investigations With Probability* - Chapter 4.1 - 4.3
- Do: Submit your work from **Lesson 10** before the start of class.
- Do: Follow the **Lesson 11 Companion** document on GitHub.

After Class Activities:

- Finalize your work from class. Submission is due before the start of **Lesson 12**. One submission per group.

Lesson 12: Multiple Regression with Interaction

Objectives:

- Understand the motivation for including interaction terms in causal models
- Understand and apply the interpretation of coefficients for interaction terms
- Build multiple liunear regression model in R

Before Class Activities:

- Read: *Advanced Statistical Investigations With Probability* - Chapter 4.4
- Do: Submit your work from **Lesson 11** before the start of class.
- Do: Follow the **Lesson 12 Companion** document on GitHub.

After Class Activities:

- Study for your midterm!

Lesson 13: Midterm

Exam Instructions: This exam will present you with a data set and a set of questions to answer about that data set. The expectation is that

you will use the knowledge and experience gained as a student in MA256 to carefully and completely answer each question. You are authorized to use outside resources on this exam; however, over-reliance on a particular outside resource may result in reduced academic credit. The expectation is that you will utilize the methods taught in this course to answer the questions presented. Utilizing methods gathered from external sources may augment but will not replace the demonstration of your mastery of the primary course material.

Lesson 14: Comparing Two Proportions - 1

Objectives:

- Properly identify the relevant statistic for these problems
- Compose hypothesis for the appropriate test
- Properly interpret p-values and confidence intervals

Before Class Activities:

- Read: *Introduction to Statistical Investigations* - Chapter 5
- Watch: WileyPlus - Videos 5.3.1 - 5.3.3
- Do: Follow the **Lesson 14 Companion** document on GitHub.

After Class Activities:

- Ensure you understand the content discussed in class today. . . you'll get a chance to prove it next class.

Lesson 15: Comparing Two Proportions - 2

Objectives:

- Demonstrate proficiency in answering research questions comparing two proportions.

Before Class Activities:

- Review: *Introduction to Statistical Investigations* - Chapter 5

After Class Activities:

- Finalize your work from class. Submission is due before the start of **Lesson 16**. One submission per group.

Lesson 16: Project IPR 1

Objectives:

Present research question, literature review, dataset, and data exploration Gain feedback and learn about other group's work *Practice providing a professional briefing* Learn to build a markdown presentation

Before Class Activities:

Review project guidance on GitHub Prepare in-progress review presentation

After Class Activities:

*Make any necessary modifications to your project plan based on feedback and/or information gleaned from other group's presentations.

Lesson 17: Comparing Two Means - 1

Objectives:

Before Class Activities:

After Class Activities:

Lesson 18: Comparing Two Means - 2

Objectives:

Before Class Activities:

After Class Activities: