College of Liberal Arts and Sciences
Department of Mathematics, Physics, and Statistics

MATH 250

Data Analysis (3 units)
Fall 2022
Section 01 Class #: 11324
Segerstrom 174 TR 4:20PM – 5:45 PM

Instructor
Katie Fitzgerald, PhD Statistics
Assistant Professor of Statistics
Email: kfitzgerald@apu.edu
Office: Segerstrom 112
Office Hours: Tue 10am – 12pm, Th 1-2:20pm
Office Phone: (626) 815-6000 x6529

APU Mission Statement

Azusa Pacific University is an evangelical Christian community of disciples and scholars who seek to advance the work of God in the world through academic excellence in liberal arts and professional programs of higher education that encourage students to develop a Christian perspective of truth and life.

Course Description

This course features hands-on experience using statistical tools to answer real-world questions. Emphasis is on analysis of actual data using statistical software. Statistical topics include numerical/graphical summaries, measures of association, and statistical techniques including chi-square tests, t-tests, ANOVA, and regression. Focus is on interpretation, not calculation.

Prerequisite: MATH 130 or MATH 361

APU Credit Hour Policy

Following the APU Credit Hour policy, to meet the identified student learning outcomes of this course, the expectations are that this 3 unit course, delivered over a 15 week term will approximate 3 hours/week classroom or direct faculty instruction. In addition, out-of-class student work will approximate a minimum of 6 hours each week.

Other Important Policies and Information

https://goo.gl/2uDWh7

Course Objectives and Desired Student Learning Outcomes for Probability and Statistics

Student Learning Outcome "By the end of this course, students should be able to…"	IDEA Objective	Assignments Used to Assess
Prepare data for statistical analysis	Developing specific skills needed by professionals in the field	Labs, Homework, Project
Use methods for effectively organizing, interpreting, and presenting quantitative data based on type of data	Learning to Apply Course Materials	Labs, Homework, Project, Exam
Identify appropriate measures of association applied to different data types	Gaining a basic understanding of the subject	Labs, Homework, Project, Exam
Select and use hypothesis tests for different data types and scientific questions	Learning appropriate methods for collecting, analyzing, and interpreting numerical information	Labs, Homework, Project
Apply concepts of power, errors, assumptions, and biases in statistical analyses	Learning to Apply Course Materials	Labs, Homework, Project

Required Course Materials

Books: This course has two textbooks, both of which are freely available online:

- R for Data Science by Hadley Wickham & Garrett Grolemund
- Introduction to Modern Statistics by Mine Çetinkaya-Rundel and Johanna Hardin

Software: This course will utilize the statistical software R via the web version RStudio Cloud. Students will receive instructions in the first week of class for how to set up an RStudio Cloud account. A $5/month subscription fee may apply (max $20 for the semester).

Hardware: Students are expected to bring a laptop to all class sessions. If access to a laptop is an issue, then please contact the course instructor and an accommodation will be made. This requirement will not prevent students from taking this course.

Campuswire account: This term we will be using Campuswire as our preferred platform for questions about homework, labs, and general course questions. The system is highly catered to getting you help quickly and efficiently from classmates and the instructor. Rather than emailing questions to the instructor, you should post your questions on Campuswire. ***See Canvas for Enrollment Code.***

Course design[1]

Modern data analysis is inherently tied to coding and computational tools, and coding is learned by doing. To facilitate this type of learning, this course utilizes a flipped learning design. Each week, you will watch a series of lecture videos and complete a preparation quiz ***before*** class on Tuesdays. Then class time will be dedicated to hands-on application exercises and group lab assignments. The activities and assessments follow a *prepare, practice, perform* format and are designed to help you develop the foundational skills of a modern data scientist.

- Prepare: Includes lecture videos and optional readings to introduce new concepts and a short quiz ensure a basic comprehension of the material. The goal is to help you prepare for the in-class activities.

- Practice: Includes in-class application exercises where you will begin to apply the concepts and methods introduced in the lecture videos. The activities will be graded for completion, as they are designed for you to gain experience with the statistical and computing techniques before working on graded assignments.

- Perform: Includes labs, homework, one exam, and the final project. These assignments build upon the prepare and practice assignments and are the opportunity for you to demonstrate your ability to apply course material to analyze real-world data.

Preparation Quizzes (Prepare)

At the beginning of each week, you will watch a series of lecture videos and take a low-stakes preparation quiz in Canvas to check your understanding. Each quiz can be attempted up to 3 times before the deadline, and your score will be the average of your attempts. The quiz is due by class-time on Tuesdays (4:20pm).

Application Exercises (Practice)

The majority of class-time on Tuesdays will be dedicated to working on Application Exercises (AEs) in RStudio Cloud, designed to help you practice the new skills, code, and concepts introduced in that week's lecture videos. AEs are due by class-time on Thursdays the same week they are assigned. AEs are graded on completion; demonstrating that a good faith effort has been made on all parts of the assignment will earn full credit.

[1] Much of the course design, activities, and assessments are adapted from Mine Çetinkaya-Rundel's *Data Science in a Box* curriculum and Duke University's open source Introduction to Data Science course under the Creative Commons Attribution Share Alike 4.0 International.

Labs (Perform)

The majority of class-time on Thursdays will be dedicated to team lab assignments. The labs are a more in-depth application of the week's material that will have you complete scaffolded analyses of a real dataset using RStudio Cloud. Teams of ~3 will be assigned by the instructor and will rotate approximately every two weeks. Labs are due by midnight the following Tuesday. *The lowest lab grade will be dropped at the end of the semester.*

Homework (Perform)

Approximately every other week, an individual homework assignment will be due. Homework will be similar in nature to the labs but are to be completed on your own outside of class. You are still encouraged to work together, but homework will be submitted and graded individually. Homework will be assigned on Thursdays and due by midnight the following Thursday.

Exam (Perform)

There will be one take-home mid-term exam for you to demonstrate your individual ability to apply what you have learned. It will consist of an analysis and computational tasks related to the material in the prepare, practice, perform assignments. The exam is take-home and open-notes but must be completed individually, without discussing or seeking assistance from anyone inside or outside the course.

You will be given the opportunity to submit annotated test corrections to earn up to 1/3 of the points back on your exam. Specific instructions and expectations will be provided when the exams are graded and returned.

Statistics Experiences (Practice)

The world of statistics and data science is vast and dynamic! The goal of the Statistics Experience assignments is to help you engage with the statistics and data science communities outside of the classroom. There are a variety of ways you can participate in "Statistics Experiences," which include but are not limited to listening to a podcast, reading a book, getting coffee with a statistician, joining the American Statistical Association, or participating in #TidyTuesday. Each experience is worth a varying number of points (see course website for details), and you are required to collect at least 5 "Statistics Experience points" throughout the semester. Up to 10 additional points can be collected and added to your lowest homework grade(s).

Project (Perform)

The purpose of the final project is to apply what you've learned throughout the semester to investigate an interesting data-driven research question by analyzing a real-world dataset of your choosing. The project will be completed in self-assigned teams of 2-3. You will be asked to present your findings in a written report and an oral presentation. More details will be provided during the semester.

Grading

Preparation Quizzes	5%
Application Exercises	3%
Engagement & Participation	2%
Statistics Experiences	5%
Homework	20%
Labs	25%
Exam	15%
Project	25%

Grading criteria and scale

- A Superior knowledge regarding details, principles, terms, and notation; superior skill in computation and application of the material.
- B More than adequate knowledge regarding the major themes; ability to compute correct answers and apply the material.
- C Basic knowledge and skill needed to solve problems relating to probability and statistics.
- D Serious gaps in knowledge, confusion of concepts, inability to recall basic information, inadequate skill in computation or application.
- F Absence of knowledge, incapable of correct computation, misunderstands most concepts.

Final letter grades will be assigned approximately as indicated in the table below.

A 93-100%	B+ 87-89%	C+ 77-79%	D+ 67-69%	F 0-59%
A- 90-92%	B 83-86%	C 73-76%	D 63-66%	
	B- 80-82%	C- 70-72%	D- 60-62%	

Late work & extension policies

The coding skills gained in this course build cumulatively week-by-week, so it is important to complete and submit all assignments on time. Due dates are there to help you keep up with course material and ensure you receive feedback in a timely manner.

- Late work will not be accepted for preparation quizzes or application exercises, as these are designed to help you prepare for the in-class labs each week.
- Labs and homework may be turned in up to 3 days late, and there will be a 5% reduction for each 24 hour period the assignment is late. Note that the lowest lab and homework grades will be dropped at the end of the semester.
- Labs are intended to be completed in teams, and a majority of each lab will be completed in-class on Thursdays. If you have an unexcused absence for a Thursday lab session, you must submit the assignment individually, and a 15% deduction will apply. The deduction will be waived for excused absences.
- Late exams will not be accepted unless specifically coordinated with the instructor in advance.
- Incompletes are rare and are available only in “special or unusual circumstances” as negotiated with the instructor prior to the end of the term. See the Catalog for policies regarding Withdrawals and grade record permanence.

Waiver for Extenuating Circumstances

Life happens, and there are sometimes extenuating circumstances that prevent you from completing an assignment on time. The lowest application exercise, prep quiz, and lab grade will each be dropped at the end of the semester to accommodate such circumstances.

Additionally, you may request a waiver of the late penalty for one assignment. To do so, email the instructor in advance of the deadline, and the waiver will be granted, no questions asked. You may only request the waiver once, so only use it for truly extenuating circumstances.

If there are life circumstances that are having a longer-term impact on your academic performance or well-being, come talk to me, and we can work towards a solution and connect you to the support you need.

Tips for success & how to access support for this class

- Dedicate yourself to being an engaged learner and contributing to a thoughtful learning environment for your peers
- Utilize Campuswire to ask questions, respond to your peers, and upvote others' questions and responses.
- Come to my office hours. Even if you don't know what your specific questions or points of confusion are, we can figure that out together. You are welcome to attend in-person (Segerstrom 112) or on Zoom; Zoom details will be provided on Canvas. You can also email me to set up an appointment at an alternative time.
- Collaborate! Get to know your classmates. This is a very hands-on class, and you will be working together extensively.
- Google is a coder's best friend! Answers & discussions on stackoverflow.com are often particularly helpful for R related questions
- Embrace the struggle & don't shy away from confusion or uncertainty. After all, statistics is the "science of uncertainty," and being "good at math" is being good at being stuck…
- Contact me about any concerns. Best way to reach me is via email (kfitzgerald@apu.edu). I do my best to respond within 24 hours.

Important Dates

August 30 (Tue)	First day our class meets
September 7 (Wed)	Add Deadline
September 9 (Fri)	Drop Deadline
September 29 (Thurs)	Exam 1 assigned
October 4 (Tue)	Exam 1 due
November 21 - 27	Thanksgiving break (NO CLASSES)
December 12 - 16	Final Exams

Course Community & Policies

It is my intent that this course models and fosters justice, equity, diversity, and inclusion. We will engage with these values both in content and in practice. Data and statistics can be tools to tell diverse stories and help us learn about the state of the world from a perspective beyond our own lived experience. When used responsibly and with integrity, they can amplify the experiences of vulnerable and historically excluded populations. For example, they can be used to shed light on disparities in our schools, healthcare system, and criminal justice system.

The research questions we ask, the data we collect, and the way we use that data are infused with (often hidden) values about who and what matters in the world. For example, we should examine if and when marginalized people and their experiences are being excluded from our data, particularly when that data is used in countless ways to drive decision-making and inform society about the state of the world.

You will be asked to continually and critically engage with these ideas with each dataset and analysis you encounter. You are expected to engage your peers and new perspectives with curiosity, empathy, and intellectual humility. It is my intent that all students be well-served by this course, that your learning needs are met inside and outside the classroom, and that the diversity that you bring to this class be valued and utilized as a resource and strength.

I (like many people) am continually learning how to honor diverse perspectives and identities. If something was said in class (by me or a peer) that made you feel uncomfortable, please let me know. You will also have the opportunity to express concerns anonymously via check-in surveys. APU encourages community members to resolve conflicts directly, when possible. If an APU community member perceives that hostile words or behaviors were directed toward an individual or a group based upon that individual's or group's identity, they can submit a Bias Incident Report. Information on the reporting process is available on the website at www.apu.edu/diversity/bias/.

Affirming that diversity is an expression of God's image, love, and boundless creativity, it is APU's aim to collectively nurture an environment that respects each individual's uniqueness while celebrating our collective commonalities. It is in this spirit that we collectively strive to create an inclusive environment in which all students, staff, faculty, and administrators thrive.

Faith Integration Statement

Academic Faith Integration is recognized as an important feature of courses at APU. My identity as a Christ-follower shapes my worldview and therefore shapes my understanding of data and statistics. You can expect to discover how relevant themes from Christianity and data science meaningfully inform one another. I respectfully recognize that students come from a diversity of faith backgrounds and that they have a variety of perspectives. This diversity in perspective will be valued as a strength and resource that enriches the learning community.

Academic Integrity Policy

TL;DR: Don't cheat!

Please abide by the following as you work on assignments in this course:

- You may discuss individual homework and lab assignments with other students; however, you may not directly share (or copy) code or write up with other students. For team assignments, you may

collaborate freely within your team. You may discuss the assignment with other teams; however, you may not directly share (or copy) code or write up with another team. Unauthorized sharing (or copying) of the code or write up will be considered a violation for all students involved.

- You may not discuss or otherwise work with others on the exam. Unauthorized collaboration or using unauthorized materials will be considered a violation for all students involved. More details will be given closer to the exam date.
- **Reusing code**: Unless explicitly stated otherwise, you may make use of online resources (e.g. StackOverflow) for coding examples on assignments. If you directly use code from an outside source (or use it as inspiration), you must explicitly cite where you obtained the code. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism.

Any violations in academic integrity standards as outlined in the APU Academic Integrity Policy and those specific to this course will automatically result in a 0 for the assignment and will be reported to the Office of the Provost for further action.

The mission of Azusa Pacific University includes cultivating in each student not only the academic skills that are required for a university degree, but also the characteristics of academic integrity that are integral to a sound Christian education. It is therefore part of the mission of the university to nurture in each student a sense of moral responsibility consistent with the biblical teachings of honesty and accountability. Furthermore, a breach of academic integrity is viewed not merely as a private matter between the student and an instructor but rather as an act which is fundamentally inconsistent with the purpose and mission of the entire university. A complete copy of the Academic Integrity Policy is available in the Office of Student Life, the Office of the Vice Provost, and online.

Support Services Policy

If there is any portion of this class that is not accessible to you due to course format or challenges with technology, please let me know so we can make appropriate accommodations. If you have a disability that might prevent you from fully demonstrating your abilities, you should meet with an advisor in Accessibility and Disability Resources as soon as possible to initiate disability verification and discuss reasonable accommodations that will allow the opportunity for full participation and for successful completion of course requirements. For more information, please contact Accessibility and Disability Resources by phone at 626-815-3849, or email at disabilityservices@apu.edu.

Bibliography

Çetinkaya-Rundel, M. & Hardin, J. Introduction to Modern Statistics. https://openintro-ims.netlify.app

Kuhn M, Wickham H (2020). *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.* https://www.tidymodels.org.

Tipton, E., Kuyper, A.M., Fitzgerald, K.G. – Adapted from Kim, A.Y. & Ismay, C. Introduction to Statistics and Data Science: A moderndive into R and the tidyverse. https://nustat.github.io/intro-stat-ds/index.html

Wickham, H. & Grolemund, G. (2017). R for Data Science. O'Reilly Media. https://r4ds.had.co.nz

Course Calendar (Tentative)

Week	Dates (T-Th)	Topics	Tuesday	Thursday	Other Assigned
1	Aug 30 – Sept 1	Welcome to MATH 250! Meet the toolkit	AE 01	Lab 01	
2	Sept 6 – 8	Data visualization	AE 02	Lab 02	HW 01
3	Sept 13 – 15	Data wrangling I	AE 03	Lab 03	SE 01
4	Sept 20 – 22	Data wrangling II	AE 04	Lab 04	HW 02
5	Sept 27 – 29	Data importing Data cleaning	AE 05	Lab 05	Exam 1
6	Oct 4 – 6	Scientific Practice Data Science Workflow	Project Work Session	Lab 06	Project Proposal
7	Oct 11 – 13	Data Ethics	Faith Integration Discussion	Project Work Session	HW 03
8	Oct 18 – 20	Simple & Multiple Linear Regression	AE 06	Lab 07	Project cleaning & EDA
9	Oct 25 – 27	Logistic Regression	AE 07	Lab 08	HW 04
10	Nov 1 – 3	Feature Engineering Cross Validation	AE 08	Lab 09	Project Rough Draft
11	Nov 8 – 10	Uncertainty Quantification Bootstrapping	AE 09	Lab 10	HW 05
12	Nov 15 – 17	Inference	AE 10	Lab 11	SE 02
	Nov 21 - 25	THANKSGIVING BREAK			
13	Nov 29 – Dec 1	Special Topics: Shiny apps	AE 11	Project Peer Review	End of Course Reflection
14	Dec 6 - 8	Special Topics: Text analysis	AE 12	Project Work Session	Project Presentation
15	Dec 13 - 15	Exam Week – Final Project Presentations			

Course schedule, topics, exams and assignments may be changed at the instructor's discretion