

DS 401 : Statistical modeling for data science : Fall 2021

MWF 8-11:40 am, CST 103

Instructor Anand Pardhanani Email: pardhan@earlham.edu
CST 210 Phone: 765-983-1683

Office hours The following hours are tentative - I'll finalize office hours after the 1st week
M: 1-2 pm. T: 1-2 pm. W: 1-2 pm.

And by appointment or walk-in. The best way to contact me, in order of preference, is: [1] in person, [2] by email, [3] by phone.

Open door policy: I keep my posted office hours to a bare minimum, to avoid being locked into a rigid schedule all semester. However, I am happy to assist students well beyond my office hours. Students are encouraged to just drop by whenever needed. Anytime my office door is open you're welcome to stop by and check whether I am available. Also, please do not hesitate to make an appointment if my posted office hours don't work for you.

Class website <https://cs.earlham.edu/~pardhan/courses/ds401/>

The website is a central component of this class, and you are responsible for regularly checking it for announcements, homework assignments and various supplementary handouts. I prepare for class with the assumption that students have reviewed the website and followed through on posted instructions.

Textbook and reference materials

We will use various open-source, online materials. Our primary textbook and reference resource will be:

[OpenIntro Statistics, 4th Edition, 2020](#), by David Diez, Mine Cetinkaya-Rundel, Christopher Barr, and openintro.org.

Other references that will be useful for certain topics include:

[Principles and techniques of data science, 2021](#), by Sam Lau, Joey Gonzalez, and Deb Nolan.

[Introductory statistics, 2018](#), by Barbara Illowsky, Susan Dean, and openstax.org.

Course credits and work load

This course is worth 4 credits, and will meet for in-person classes for 400 minutes each week for 7 weeks. This is consistent with the standard practice of 4-credit courses meeting for 200 minutes per week during a regular 14-week semester. In addition, students should expect a workload outside class of about 15 hours each week.

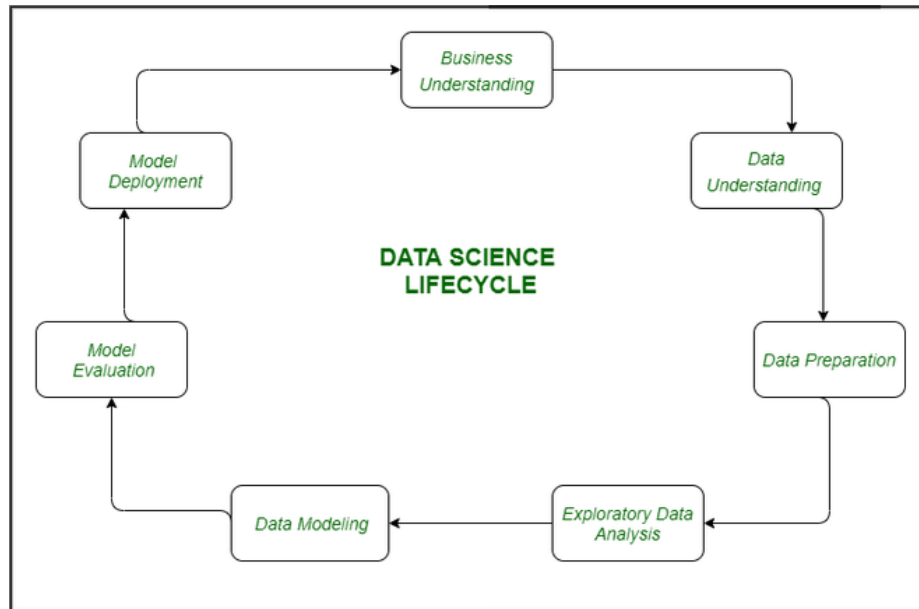
Requirements this class fulfills

This class is required for majoring in data science. In addition, it fulfills the Quantitative Reasoning component of Earlham's General Education requirements.

Description & objectives

The breadth and diversity of real-world data science projects makes it nearly impossible to devise standardized procedures and recipes that would be applicable to all of them. In spite of this diversity, however, it is possible to identify certain common components that frequently arise in

the life cycle of most data science projects. The schematic diagram below shows key components in the life cycle of a typical data science project.



[Image courtesy of [geeksforgeeks.org](https://www.geeksforgeeks.org)]

The main objective of this course is to introduce certain essential methods, tools and theoretical concepts relevant to the data analysis, data modeling, and model evaluation phases of the life cycle. In particular, we will focus on statistical methods designed for studying patterns and relationships within data sets, and for modeling such relationships. In addition, we will learn how to use such models for forecasting and decision-making, as well as critically examine questions that relate to quality, reliability, and effectiveness of data models. In terms of software and implementation aspects, an important goal of this course is to train students in the use of R.

Key topics this course will cover include: a review of probability models, distributions, and the central limit theorem; linear regression with two variables; linear regression with many variables; logistic regression; diagnostics and inferences for regression. If time permits, we will also cover topics in principal component analysis and time series regression.

Student learning goals and outcomes

Upon successful completion of this course, students will be able to

1. Understand the key components in the life cycle of typical data science projects.
2. Carry out basic statistical analysis of structured data sets, including exploring and summarizing key trends and patterns.
3. Develop models for studying relationships between variables, including linear regression, multilinear regression, logistic regression.
4. Evaluate the quality, reliability, and effectiveness of regression models.
5. Use the R suite of software tools to carry out statistical analysis and modeling.

These aspirations broadly support all 7 learning goals of an Earlham education (see the Appendix attached to this Syllabus).

Course prerequisites: Elementary Statistics (MATH 120) or Mathematical Statistics (MATH 300), and Calculus A (MATH 180).

Assessment & grading policy

Your final grade will be based on combined performance on: quizzes and classwork, lab projects, homework problems, one exam during the semester, and a final exam. Each will contribute the following proportions:

Quizzes & Classwork	30%
Lab projects	15%
Homework	15%
Mid-term exam	20%
Final exam	20%

Letter grade boundaries for this course are not set in advance. They will be determined at the end of the term, based on factors such as overall class performance, level of difficulty of tests, quizzes, and assigned work, etc. At a minimum, the following standard scale for letter grades will be honored:

A+: 97.0-100; A: 93.0-96.9; A-: 90.0-92.9;
B+: 87.0-89.9; B: 83.0-86.9; B-: 80.0-82.9;
C+: 77.0-79.9; C: 73.0-76.9; C-: 70.0-72.9;
D+: 67.0-69.9; D: 63.0-66.9; D-: 60.0-62.9; F: below 60.

NOTE that all students must also satisfy the following minimum requirements to receive a grade of C- or better:

- * Take both the exams (mid-term and final).
- * Turn in at least 75% of the homework problems.
- * Turn in at least 75% of the quizzes and classwork.
- * Complete all the lab projects.

More details about assessment categories

Quizzes and classwork: In-class quizzes and/or classwork will be frequently assigned throughout the semester. A key purpose of the quizzes is to help us accomplish learning goals 2 through 4 (listed above). Quizzes will typically be short (e.g., 10 minutes), and will focus on conceptual understanding of key ideas. Classwork, on the other hand, will be much like homework problems, and will serve the purpose of hands on learning and practice in class. This will help us accomplish learning goals 1 through 4. Students will sometimes do classwork in teams and, in such cases, turn in a common "team solution" for grading.

Lab projects: The goal of labs is to learn through focused, hands on exploration into application areas, and to enhance conceptualization. In addition, labs will play a key role in helping students learn how to use R. Collectively the labs will broadly support all 5 learning goals.

Homework: The purpose of homework is to help you learn course content and to give you practice applying concepts and solution techniques. Exercises will be assigned from the textbook and other sources at various points throughout the semester. These must be turned in at the beginning of class on the indicated due dates. Homework exercises will help us accomplish learning goals 2 through 4.

Exams: There will be one mid-term exam during the term, plus a final exam at the end of the term. The exams will help fulfill and assess goals 1 through 4. The tentative date of the mid-term exam is October 27.

The final exam date and time is set by the registrar's office. According to their calendar, the final exam will be held **Monday, Nov. 22, at 9 am.**

Important dates

- * Last day to add this course: Oct. 6.
- * Last day to drop: Nov. 5.
- * Date of final exam: Nov. 22.

NOTE: Last drop date applies to Earlham students only. Students cross-registered through IU-East or other institutions must follow the dates and rules of their own institution.

Academic integrity

After several years of writing standard, boiler-plate stuff in this section, I have decided to replace it with a more authentic message from my heart to yours. Before getting into details, I would like to share 3 key ideas that profoundly shape my thinking, and prompt me to explore more effective ways towards academic integrity:

1. Academic infractions are a much bigger problem at Earlham than many of us would like to believe or admit.
2. The problem is **NOT** our students! Earlham students are as good (or better!) than their peers at other institutions in terms of moral values and ethical standards.
3. Infractions at Earlham can be significantly reduced using a combination of strategies, collectively developed by students and faculty.

These three points summarize my overall perspective, and will frame the rest of my discussion on this subject.

By far the single biggest phenomenon that has radically transformed today's academic integrity / infraction landscape is technology -- particularly the internet and cell phones. In my view, Earlham's traditional approach to academic integrity has been rendered completely obsolete by these technologies. If I were an Earlham student today, I would encounter many situations where the temptation to infract would be extremely high, because these technologies make it so easy, and the risk of getting caught is virtually zero. This is the main reason why I say that you, the student, are not the problem. You are human, just like me and my faculty colleagues. It is a fact of life that many humans succumb to temptation when the rewards are sufficiently high, and the risks sufficiently low.

Yet, the fact remains, a growing rate of academic infractions is a terrible thing to ignore: They sink an institution's reputation, decrease the value of students' education, lower student & faculty morale, and more. Clearly, we need to explore and develop new strategies that are more effective for our times, and also preserve Earlham's distinctive approach to such matters. We will set aside some class time to discuss and formulate specific policies for helping students (joyfully!) meet and exceed the highest standards of integrity in this class. In the meantime, I invite you to reflect on some practical ways that would most help and support you in avoiding the use of inappropriate sources for completing and turning in your graded work.

I would like to conclude with the following excerpt from the Earlham Academic Integrity Policy: "The College trusts students who enroll at Earlham to be honest seekers of truth and knowledge. This trust is extended to all students by other students and by teachers ... Giving or receiving aid

inappropriately on assignments and tests, or plagiarizing by using another person's words or ideas without credit, constitutes a serious breach of our trust in one another and in the integrity of the search for truth. Those who believe they have witnessed violations of academic integrity should feel the obligation to speak about this to the suspected offender. The witness also should feel obligated to report the suspected offender to the instructor if the person fails to offer a satisfactory explanation and refuses to report him or herself. ... Violations of academic integrity, because they undermine our trust in one another and in the credibility of the academic enterprise, are taken very seriously. Penalties for violations range from failing assignments or tests to suspension or expulsion from the College. "

Makeups

In-class items: There will be no makeup for missed in-class items (e.g., quizzes, classwork, class participation, etc.) regardless of reason. I will drop your lowest two scores as an implicit way of making up for missed items.

Homework: Past-due assignments will not be accepted except in rare circumstances, provided the student receives prior consent from the instructor.

Exams: Make-up exams will not be given except in cases of documented illness or emergency.

Academic accommodations

Students with a documented disability (e.g., physical, learning, psychiatric, visual, hearing, etc.) who need to arrange reasonable classroom accommodations must request accommodation memos from the Academic Enrichment Center (main floor of Lilly Library) and contact their instructors each semester. For greater success, students are strongly encouraged to visit the Academic Enrichment Center within the first two weeks of each semester to begin the process. For further details, please visit

<https://earlham.edu/academics/academic-support-and-special-programs/academic-enrichment-center/accessibility-services/>

Other sources of help

1. **The Academic Enrichment Center:** The Academic Enrichment Center (AEC), located in Lilly Library, provides assistance with study habits and skills as well as a peer tutoring service. The AEC is staffed by trained peer tutors for either pre-arranged group tutoring sessions (provided for many math, science and social science courses) or one-on-one tutoring sessions for other courses. Peer tutoring is a free service offered to all Earlham students. Please visit <https://earlham.edu/academics/academic-support-and-special-programs/academic-enrichment-center/peer-tutoring/> for more information.
 2. **The Earlham Writing Center:** The Writing Center is dedicated to providing students with advice and resources about writing. Students can meet one-on-one with trained consultants who will contribute feedback to writers at any stage of the writing process: brainstorming, drafting, researching, revising, and polishing. This is a free, walk-in service on the main level of Lilly Library. In addition to dropping by, students may also schedule an appointment in advance using the online scheduler found at: <http://www.earlham.edu/writing-center/>. Also, if you want help with specific grammar topics related to your own writing, <https://www.grammarly.com/edu> is available for all Earlham students to proofread their papers and learn more about grammatical errors.
-

Appendix to the Syllabus

Math department student learning goals

Curiosity: an encounter with mathematical meaning, beauty, and joy.

Mathematics is an art as well as a craft. Depending on how one counts, somewhere between five and all of the classical seven liberal arts have significant mathematical content. We'd like to help students encounter seriously both the beauty and the utility of an art without which neither the world around us nor the last two and a half millennia of human culture are at all intelligible.

Craft: using mathematical tools.

Mathematics is a craft as well as an art. Much of our work involves exposing our students to the concrete mathematical tools they need to succeed in other disciplines and in a world shot through by mathematics. Those tools range from fancy machinery for quantum physicists to using simple algebraic or differential equations to model phenomena in biology and economics to techniques to manage randomness in data in the social sciences to the basic vocabulary needed to read newspaper stories with numbers in them.

Confidence: an ability to create and use mathematics themselves.

Mathematics is also the product of human beings, who work individually and collectively to discover or invent mathematical truth. We work to inculcate in our students a spirit of inquiry and to empower them to discover that they are able as individuals and working with others to make mathematical discoveries and to utilize mathematics in creative ways. Math is the work of humans, and as humans, our students have all the prerequisites they need to do and to use mathematics.

Communal Inquiry: mathematical community and communication.

Mathematics is a communal enterprise, and even a glance at the words shows that one can't have community without communication. We'd like our students to practice mathematical communication. This means that they need to learn to verify and to convey to others the results of their mathematical inquiries by writing precise, concise, and completely persuasive arguments in idiomatic mathematical language. In mathematics, the products of this writing are called *proofs*. It also means that students need to practice reading carefully and critically the mathematical works of others so as to be able to share in a community of inquiry and to learn new mathematics on their own.

Continuity: lasting mathematical experience.

Mathematics does not end with what one learns in four years at college. We wish to equip our majors for further study in mathematics at the post-

baccalaureate level, though we accept that with our current program, the path available to most students wishing to do graduate work will require them using as a stepping stone a Masters program at something other than a first-tier university. We would also hope that our non-majors might carry the spirit of joy, skill, confidence, and cooperative inquiry with them after they leave Earlham.

Earlham College learning goals

(See <https://earlham.edu/registrar/curriculum-guide/learning-goals/>)

Students should be able to:

- **Communicate** effectively and work collaboratively across diverse contexts via multiple media. Effective communication involves both social and expressive skills and the ability to communicate in multiple settings and cultures.
- **Investigate** and analyze information, materials, problems and texts using a variety of techniques. Thoughtful and careful analysis requires the ability to collect, understand, interpret and evaluate multiple pieces of evidence, with systematic understanding and overt application of qualitative, quantitative, analytical and abstract reasoning.
- **Integrate** knowledge, experience, and skills across domains and contexts. Integration involves connecting and developing ideas, as well as synthesizing and transferring learning to new and complex situations.
- **Diversify** personal and cultural experiences, ways of knowing, and social relationships. The practice of diversity involves embracing opportunities to explore outside their interests and typical frame of reference.
- **Create** and innovate across a variety of disciplines. Creativity and innovation require a willingness to take risks, be open to new possibilities, and produce new knowledge and artistic and social forms.
- **Reflect** critically on their learning experiences, ethical and vocational choices, lifestyle, and beliefs in light of multiple understandings of the world. Reflection involves the ability to examine past experiences and apply their lessons to future contexts.
- **Apply** knowledge and skills to real world problems and situations as well as to improve their own mental, spiritual and physical well-being. Applying learning effectively is a key skill of a lifelong learner.