**Introduction to Regression Analysis (STAT 3220)**

**Fall 2024**

*“All models are wrong, but some are useful.” –George Box*

**What is regression and why should you care?**

Will UVA win March Madness this year? How much money you will spend on a vacation? What makes a Spotify song “danceable”? Is there really evidence of global warming? What factors affect your chance of being accepted to graduate school? If you have ever asked yourself any of these questions, then you might have tried to use regression! Regression is one of the most fundamental analysis techniques in statistics that uses a set of explanatory variables to try to predict or estimate a certain outcome. Regression, just as the field of statistics in general, can be used in any discipline to answer many interesting questions. In this course, we will explore three foundational techniques of regression: multiple linear regression, logistic regression, and ANOVA.

**What will you be able to do?**

During this course, you will develop the fundamental practical and interpersonal skills to think and act like a statistician, meaning you value the importance of and are confident in your ability to evaluate evidence.

By the end of this course you will be able to:

* Compare and contrast regression techniques in order to decide which is the most appropriate given a particular research question or data set
* Use a statistical analysis to evaluate a research claim of interest to you personally
* Perform a regression analysis using R Studio
* Clearly and efficiently present your findings through writing a semi-formal statistical report, an oral discussion, and academic poster presentation
* Implement a strategy for learning an unfamiliar regression topic
* Establish and maintain effective working relationships with others
* Develop an understanding of the obligations of a statistician for reporting findingsethically

**What tools do you need to be successful?**

* **Canvas:** All of the resources you should need or want for this course will be posted to Canvas. This site can also be used to communicate with your peers, upload assignments, and download supplemental notes. Additionally, announcements will be posted via Canvas. It is your responsibility to check for course updates through email or announcements.
* **Textbook**: The primary recommended text for this course is *A Second Course in Statistics: Regression Analysis (7th ed) ISBN:* 978-0321831453. There is a more recent edition, but the 7th will suffice. The secondary recommended text for this course is *Practicing Statistics: Guided Investigation for the Second Course ISBN: 978-0321586018***.** 
  + The sections listed in the syllabus correspond with the *Regression Analysis* book*.*
  + While I will provide supplemental notes, I recommend you read the textbook to create your own collection of notes.
* **R Studio:** The software introduced and featured in this course is R. RStudio is a software that runs R with additional user-friendly features. RStudio will be used for all in-class R demonstrations and examples. Both R and RStudio are free for download for all operating systems. R will need to be [downloaded](https://cloud.r-project.org/) first. RStudio Desktop is the recommended version for [download](https://posit.co/download/rstudio-desktop/). Be sure to download the version that match the specifications of your computer. **There was an update to Base R in the summer, you will want to download that new version (4.4.1 “Race for your life”)**
* **Gradescope:** You will use Gradescope to submit most assignments. Please familiarize yourself with the appropriate submission procedure. If an assignment is not submitted correctly, it will not be graded.

**Who/What is here to help you be successful?**

* **Instructor**: Krista Varanyak (that’s me). My research is Statistics Education (how to teach statistics) and I am currently learning more about intermediate stats courses across the country. When I am not teaching (,or prepping, or grading, or meeting with students,…) I enjoy dancing, reading, am a certified fitness instructor, and own a local studio.
  + **Email**: [kristav@virginia.edu](mailto:kristav@virginia.edu) (same as kmv9q@virginia.edu)
  + **Office**: Halsey 211
* **Undergraduate Course Assistants**: Great resources in each class meeting + office hours
  + 8:00am Section - Elizabeth Stevenson (ybp7zy@virginia.edu)
  + 9:30am Section - Divya Ramakrishnan (hzy8ha@virginia.edu)
  + 11:00am Section – Rowan Rosenblum (qtg3zr@virginia.edu)
* **Office Hours:** You may attend any of the hours listed below or email Prof V to schedule an appointment. Please include your availability when sending an email. **Halsey 120 is the first floor conference room in Halsey. Halsey B002 is the basement.**
  + **Monday 1-2pm in Halsey 120 (Prof V)**
  + **Tuesday 6-7pm in Halsey 120 (Divya)**
  + **Wednesday 11-12pm in Halsey 120 (Prof V)**
  + **Wednesday 5-6pm in B002 (Rowan)**
  + **Thursday 12:30-1:30pm in Halsey 211 (Prof V)**
  + **Friday 2-3pm in Halsey B002 (Ellie)**
* **YOU (Discussion Boards):** I think students are sometimes the best source of information, so we will utilize Discussion Boards as our question forum to communicate outside of class. You can post and reply to questions about content, assignments (except where noted), R Studio support and more. Posting and replying in discussion boards will be part of your final grade in the course. Please see the discussion board for more specifics. The course assistants and myself will monitor these about once a day as well.
* **Internet Sources**: Google searching is a skill and will be helpful for troubleshooting and learning new skills. We cover some of the basics, but you may also want to refer to some R Studio handy sheets ([Base R](https://iqss.github.io/dss-workshops/R/Rintro/base-r-cheat-sheet.pdf), [R Markdown](https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf), [Graphics](https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf), and [many others](https://www.rstudio.com/resources/cheatsheets/))

**How will you achieve the goals?**

Ultimately, developing the skills required to think and act like a statistician comes down to **learning** the course material, **recalling** the course material, and **applying** the course material. Therefore, in order to demonstrate you are meeting the course objectives, which lead to the overall course goal, you will be assessed on your demonstration of content engagement, your content knowledge, and content application.

* **Content Engagement:** The purpose of each aspect of content engagement is to demonstrate that you are learning the course material. Activities are designed for you to engage with the material to help you discover underlying themes in regression. The primary purpose of these activities is to inform me about your learning and give an opportunity for you to reflect on your own progress.
  + Pre-Class Readings & Videos: Preparing for each class meeting (reading the relevant textbook sections, watching lecture videos, etc.) will help you get the most out of the class meeting. You are expected to read the appropriate sections listed on the calendar and be prepared to answer the unit questions and watch the pre-recorded lecture video (usually about 30 minutes).
  + Live Class Meeting: Each class meeting will be spent reviewing an example from the corresponding pre-recorded lecture. There will also be time to answer questions from the lecture and work in groups. Class meetings are not recorded.
  + Classwork Assignment: The goal of the classwork activities is to expose you to the nuances of the regression process, develop your own understanding of the techniques, and provide low-stakes situations for you to explore techniques through practice. Classwork activities will vary in structure for each class meeting (individually or in groups; may use R; may be discussion or written; etc). The daily assignments are anticipated to be completed in the class meeting. Therefore, if you attend the meeting, you will complete the assignment during that time. Otherwise, you can complete it on your own time. Your unit classwork will be collected and scored individually at the end of each unit. There will be some classwork assignments due the day of class.
  + Practice Set: There will be 4 unit practice sets. These will not be graded, nor collected. The purpose it to provide you additional exercises to help prepare for quizzes and lab activities. Solutions will be posted.
* **Content Knowledge**: Content knowledge assessments are designed for you to show me what you know through quizzes and labs. The purpose of each content knowledge assessment is to demonstrate that you can recall the course material and perform related tasks correctly.
  + Quizzes: There will be a total of 9 quizzes that correspond with the previous week’s material. The purpose of the quizzes is for you to show you can correctly apply the techniques discussed through in class activities. **Typically, quizzes will open after each Thursday meeting and close at 11:59pm on Tuesday.** Quizzes will be cumulative in nature and there will be 10 multiple choice questions and you will be timed for 30 minutes for each quiz. Additionally, there will be 1 exit survey which will be given at the end of the term that will count as a quiz grade. Therefore, there are a total of 10 assignments in this category.
  + Lab Activities: You will complete the Lab Activities individually (4 Units, 1 Writing Lab, 1 Stat Scholar). The purpose of the labs is for you to demonstrate your ability to apply and extend the concepts we discuss through an exploration and analysis of a real-world problem using R. With the exception of the Writing Lab and Stat Scholar, Unit Labs are due at the end of the class session. If you are not present in class, you will not complete the lab.
* **Content Application**: The purpose of each aspect of content application is to demonstrate that you can apply and synthesize the course material. Projects are designed for you to expose yourself to the real-world intricacies of being a statistician through application. As a group of 3-4 students you will select a research topic (and data set) of interest, perform a statistical analysis, and communicate your findings through a written report, and/or a presentation. **(Specific details to follow).**
  + Project Part 1: You will use methods covered in the exploratory data analysis phase to collect (or compile) your own data and present it in a thoughtful way without inference.
  + Project Part 2: You will apply and extend the multiple linear regression and logistic regression techniques covered and present in an academic poster session and report.

**How will your objectives be measured?**

**NOTE: Please refer to Gradescope for the due dates for each assignment. With the exception of Labs, typically assignments are due at 11:59pm on the due date listed.**

* Classwork (10%): Your classwork score will be comprised of: your scores on the 4 unit assignments, scores on various project work day assignments, or your attendance. If you are unable to submit the activity on time, you cannot make it up for credit, but it should still be completed for your own understanding. To account for short-term illnesses, emergencies, interviews, etc, classwork scores will be weighted to allow for 2 class meetings throughout the term to be missed without penalty.
  + Your classwork is scored out of points, not per assignment. Therefore, your lowest scores are not excluded from your grade, instead your score will be measured out of 200 points less than what was submitted.
* Discussion Board Posts (5%): Discussion Boards are worth 5% of your participation grade in this course. To earn your full 5% you will earn 5 points across the semester. You can see the point break down on the discussion board. You cannot earn more than 2 points in the first two weeks or last two weeks of the semester or on any specific board. Your goal is to space your questions and replies throughout the semester.
* Quizzes (15%): Quiz questions are scored as correct or incorrect. You are permitted to use any notes produced in the course, but are not permitted to work with other students or seek outside help. To account for short-term illnesses, overload weeks, emergencies, interviews, etc, one (out of the 10) quiz grades will be dropped.
* Labs (35%): The labs will be graded with an emphasis on correct logic in addition to a correct final solution. Labs cannot be made up, please mark these dates on your calendar. To account for short-term illnesses, overload weeks, emergencies, interviews, etc, one (out of the 6) quiz grades will be dropped.
* Project 1 (15%): Details and rubrics will be given at the date listed on the calendar.
* Project 2 (20%): Details and rubrics will be given at the date listed on the calendar.
* Grade Thresholds: The default university thresholds are used in this course, for more information see <https://virginia.service-now.com/its?id=itsweb_kb_article&sys_id=1153c16fdba41f444f32fb671d961934#undergraduate>

**When will we do all this? – need to update chapters to correspond with other book.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Day** | **Date** | **Material Covered** | **Chapter/Topic (Class Activity)** | **Assignments & Project** |
| **Unit 1: WHAT IS A MODEL? (Ch 2-4.11)** | | | | |
| Tuesday | 14-Jan | Welcome, Syllabus, Policies, | **Course Introduction** |  |
| Thursday | 16-Jan | Unit 1.1: Ch 2-3; Introduction to R Studio | **What is model?**  Simple linear regression model |  |
| Tuesday | 21-Jan | Unit 1.2: Ch 4.1-4.5 +4.11 | **What is a model not?** Estimating the parameters (betas and sigma) of regression. | Quiz 1 Due (Syllabus) |
|  |
| Thursday | 23-Jan |  | **Writing Lab**  Storyboarding with a partner. |  |  |
|  |
|  |
|  |
| Tuesday | 28-Jan |  | **Writing Continued.** |  |  |
| Thursday | 30-Jan | Unit 1.3 CH 4.6-4.8 | **What makes a good model?**  Assessing the model; parameter inference, R^2 vs adjusted R^2, Global F | Writing Lab Due  Friday 1/31 |  |
| Tuesday | 4-Feb | Unit 1.4: Ch 4.9 | **How do you use a model?**  Prediction vs Estimation | Quiz 2 Due (1.1-1.3)  Unit 1 Classwork Due Wednesday 2/5 |  |
|  |
|
|  |
| Thursday | 6-Feb |  | **Project Work Day**  Visit from Data Librarian | Project 1 Assigned |  |
| Tuesday | 11-Feb |  | **Unit 1 Lab** |  |  |
|  |
| **Unit 2: HOW TO… BUILD A MODEL? (Ch 4.10-5.10)** | | | | |  |
| Thursday | 13-Feb | Unit 2.1: Ch 4.12, 5.7-5.9 | **How to… turn colors into numbers?**  Qualitative variables |  |  |
|  |
|  |
| Tuesday | 18-Feb | Unit 2.2: CH 4.10-4.13 | **How to… change the slopes?**  Interactions | Quiz 4 Due (1.4-2.1) |  |
|  |
| Thursday | 20-Feb | Unit 2.3: CH 5.11 | **How to… Build a model?** Nested Test, Model Validation | Unit 2 Classwork Due Friday 2/21 |  |
| Tuesday | 25-Feb |  | **Project Work Day**  Peer Review Task | Quiz 5 Due (2.2-2.3) |  |
| Thursday | 27-Feb |  | **Unit 2 Lab** |  |  |
|  |
|  |
| **Unit 3: WHY ARE ERRORS “NORMAL”? (Ch 8 + 7.6)** | | | | |  |
| Tuesday | 4-Mar | Unit 3.1: Ch 7.4 + Ch 6.1-6.2 | **What happens when “eXes” aren’t independent?** Multicollinearity + Stepwise Regression |  |  |
| Thursday | 6-Mar | Unit 3.2: Ch 8.1-8.5, 8.7 | **Why are errors “Normal”?** Regression Residuals & assumptions |  |  |
|  |
| Tuesday | 11-Mar | Spring Break | **NO CLASS** |  |  |
| Thursday | 13-Mar | Spring Break | **NO CLASS** |  |  |
| Tuesday | 18-Mar | Unit 3.3: Ch 8.6 | **Can one observation change a model?**  Outliers and Influential Points | Quiz 6 Due (3.1-3.2) |  |
|  |
|  |
| Thursday | 20-Mar | Unit 3.4: Ch 7 + P-values | **Friends don’t let friends extrapolate.**  Regression Pitfalls  **Project Work Day** | Unit 3 Classwork Due Friday 3/21 |  |
|  |
|  |
| Tuesday | 25-Mar |  | **Unit 3 Lab** | Quiz 7 Due (3.3-3.4) |  |
|  |
|  |
| **Unit 4: WHAT ARE THE ODDS? (CH 9.5-9.6, CH 12, Supplemental)** | | | | |  |
| Thursday | 27-Mar | Unit 4.1: Ch 9.5-9.6, Supplemental | **Why not use linear regression?**  Logistic Regression | Project Part 1 Due Friday 3/28 |  |
| Tuesday | 1-Apr | Unit 4.2: Supplemental | **What are the odds?** Logistic regression cont |  |  |
|  |
| Thursday | 3-Apr | Unit 4.3: Ch 12.1-12.3, 12.5 | **One factor. Many means. One test.**  One factor ANOVA | Quiz 8 Due (4.1-4.2) |  |
| Tuesday | 8-Apr | Unit 4.4: Ch 12.7 & 12.9 | **ANOVA Follow Up**  ANOVA assumptions and post hoc analysis | Unit 4 Classwork Due Wednesday 4/9 |  |
| Thursday | 10-Apr |  | **Project Work Day** |  |  |
|  |  |
| Quiz 9 Due (4.1-4.2) |  |
|  |  |
| Tuesday | 15-Apr |  | **Unit 4 Lab** |  |  |
|  |
|  |
| Final Project | | | | |  |
| Thursday | 17-Apr |  | **Project Work Day** |  |  |
| Tuesday | 22-Apr |  | **Project Work Day** |  |  |
| Tuesday | 24-Apr |  | **Project Presentations** | Final Project Due for all |  |
| Thursday | 29-Apr |  | **Project Presentations** | Stat Scholar + Exit Survey Due |  |
|  |

**There is no final exam for the course, and we will not use our exam period.**

**All work will be completed and submitted by the last day of the semester.**

**How will we all –y’all and I- work together?**

In order to create a positive, equitable learning environment for all students, the following policies are in place and resources are available:

* **Classroom etiquette**-There are no right or wrong answers in Statistics (most times). You are free to ask questions related to our topic and are to treat your peers and instructor with respect.
* **Honor Policy**- By enrolling in this course, you agree to uphold the Honor Policy of the University. Policies are laid out in this syllabus for each type of assignment. Additionally, you are not to share any assignments, or materials including, but not limited to classwork, quizzes, homework, or projects with students from previous or future semesters. You are not permitted to use online question forums as listed in this syllabus. If you have any questions about the policies, it is always better to ask than assume.
  + **Generative AI**  
    Students may use generative AI programs, including ChatGPT, on assignments with some limitations. Generative AI programs are not a replacement for human creativity, originality, critical thinking, and research. Writing, whether text or code, is a craft that must develop over time to refine the author’s individual voice, though, with proper attribution, AI programs may be used as a tool in this process.
  + **Students are NOT permitted to submit any information provided by the instructor, including whole or partial problems, prompts, background information, examples, data, etc., into generative AI programs.**
  + Any submitted work using AI programs must clearly indicate what part is the student’s work and what part is generated by the AI programs with appropriate notation and citations. In such cases, no more than 25% of the submitted work should be generated by AI programs. Students should also indicate how generative AI programs informed their process and their submitted work, including how the generated information was validated.
  + Assignment instructions will provide additional guidance as to how students should provide transparency about the use of generative AI programs in submitted work. Using a generative AI program without proper attribution is a violation of the Honor Code and will be treated as such.
  + Students using generative AI programs should view them as a collaborative tool, understanding that they can be trained on limited datasets that may be out of date. Additionally, generative AI datasets are trained on pre-existing material, that may include copyrighted material and, therefore, relying on generative AI programs may result in plagiarism or copyright violations. Finally, students should keep in mind that the goal of generative AI programs is to produce content that seems to have been produced by a human, not to produce accurate or reliable content and, therefore, relying on generative AI programs may result in submission of inaccurate content. It is the student’s responsibility – not the program’s – to assure the quality, integrity, and accuracy of submitted work.
* **Attendance and Make Ups**- The general policy for all assignments and class meetings is that they are not to be made up and will not be accepted after the due date. Thus if for any reason you will miss class (including short-term illness, overloaded weeks, interviews, or other temporary unforeseen circumstances.) If you do not submit an assignment on time to Gradescope (or whatever platform) please do not send it to me or the grader through email- it will not be graded.
  + In the first few weeks of class, you should find at least two people who you can check in with in the event that you are ever ill. I strongly recommend that at least one of these people is NOT in your classwork group.
  + If you know you will miss an extended period of class due to illness or another reason, please reach out and we will make arrangements to provide a reasonable accommodation for you as you recover. For these situations, you should also be in contact with your Advising Dean, Academic Advisor, and/or and SDAC Advisor. Case-by-case exceptions will not be granted beyond these university channels.
  + Attendance is not a mandatory factor into your course grade. You must only be present on lab days. The item in the gradebook is labelled as such (not factored into your grade). Please do not email or ask about your roll call / attendance score, it is for my tracking only.
* **Class Meetings + On Time Arrival-** Our class meetings begin at 8:00am, 9:30am, 11:00am respectively to your enrolled section and expect to remain in class the entire 75 minute duration. You should be in class, and ready to begin by the start of the class meeting. Arriving regularly well after our start time is disruptive and disrespectful to myself and your peers.
  + There is not a formal policy on late arrivals, but you should be courteous and respectful. I understand occasionally, needing to arrive a few minutes late, but making this a regular occurrence is not advisable nor professional.
  + If you plan to leave class early - It would be appreciated to let me know prior to the start of class. Otherwise, I assume there is a concern.
    - You are of course permitted to leave early as needed, but I do ask you not do so until we are at least working in groups.
    - You should also be aware that if you leave early, you might miss pertinent course information.
  + NOTE: Our class has a total of 29 class meetings- arriving just 3 minutes late to each class is the equivalent to missing 1 full class meeting.
  + I do my best to make attending class a worthwhile use of your time. As such I do not have policies regarding attendance. However, I would appreciate you making the effort to remain actively engaged during the 75 minutes you have committed to being enrolled in this course.
    - Read the university policy on attendance: <https://college.as.virginia.edu/class-attendance>
    - To elaborate on remaining engaged in class: you can do so by answering questions, participating in discussion with your group, and not working on other courses or outside materials while you are in class. Your participation is my primary way to engage your real-time understanding to properly pace the class.
* **Office hours policies:** The following policies will apply to all office hours to be able to answer as many questions for as many students as possible.
  + All office hours will be open-door with a queue. To discuss personal situations, students should schedule an appointment with the instructor.
  + Students should arrive prepared with specific questions.
  + Students should not expect that variations of the questions “Is this correct?”' or “Will this earn full credit?” will be answered.
  + Students should not expect that significant portions of material from a missed class will be explained.
* **Email**- As a student at UVA, I am assuming you are bright and inquisitive and I want you to continue to develop those attributes. Therefore, before sending me or a TA an email, you are expected to consult with at least three other sources to find your answer (Peer, Syllabus, textbook, R help pages, Discussions). Most questions should be posted and answered within Discussion Boards. Assuredly, you are not the only one with that question. If you have a personal question or would like to set up a meeting, please see Canvas for the best way to communicate with me. Emails that are answered in the syllabus, about a specific assignment grading, or submitting a late assignment without proper accommodations may not receive a reply.
* **SDAC**- All students with special needs requiring accommodations should present the appropriate paperwork from the Student Disability Access Center (SDAC). It is the student's responsibility to present this paperwork in a timely fashion and to follow up about the accommodations being offered.

*Note: The syllabus is subject to change. All changes will be submitted in writing.*

*FAQ*

*Consider this to be an extension of the syllabus to clarify all questions.*

* Group Work Submissions
  + **Classwork**- You can work with up to three other students (groups of no more than 4) during class time. However, you will submit the classwork individually. I will randomly assign groups in the first few weeks. Then you will be able to work with any group you wish. You are permitted to change groups from class to class.
  + **Project**- Details about groups are coming. Unless there are any big issues, you will stay in the same group for part 1 and part 2- and you will be able to select your group members.
* Missed class
  + Find a few students in the class to connect with and get notes and information that you miss.
  + The syllabus is designed so that you can “miss” two non-lab class meetings without penalty towards your grade.
  + Late work will not be accepted, and you do not need to notify the professor if you will miss a class meeting.
* Extra Credit
  + Any extra credit opportunities will be disclosed to you if and when they become available.
* Gradescope Submissions
  + To upload your work, you can submit individual jpeg images, or ONE PDF. Please practice doing this before the first assignment.
  + Work that is not properly upload in the proper format or without assigning pages will not be accepted, nor graded. It is your responsibility to submit your work in the appropriate format.
* Grading
  + Labs are primarily graded for correct logic and work- partial credit will be given within reason.
  + Classwork is primarily graded on a complete, thoughtful effort
  + Projects will have specific rubrics for grading
  + Re-scoring of submissions is not accepted.
  + To discuss a grade, you must meet with myself or the TA in person.
* Discussion Board Posting
  + Be sure to search for you question before posting.
  + You can also tag a question that had already been answered.
* Class Meetings
  + As stated, classes meetings will not be recorded.
  + You must attend the section for which you are enrolled. Very rare instance of attending a different section must be approved by the instructor.