84-701 Regression Analysis in Politics Science (RAPS) I – Fall 2020 Syllabus

Daniel Silverman
dmsilver@andrew.cmu.edu

Course: Tuesdays and Thursdays 1:30–2:50pm, Posner 152 or via Zoom

Office Hours: by appointment via Zoom

Course Description:

This is the first half of the quantitative methods sequence for graduate students in political science. In this course (RAPS I), students will learn the basic techniques of quantitative analysis in political and social science, focusing on exploratory data analysis, hypothesis testing, bivariate associations, ANOVA, and linear regression. In the next course (RAPS II), students will build on this foundation to explore a wider range of statistical tools, including topics such as survival models, count models, time series models, and panel models. Together, these courses will give students a solid grounding in quantitative methods and data analysis that should equip them to engage with statistical research in the field and prepare them to analyze quantitative data (if they so choose) in their thesis projects as well as their future careers.

Learning Objectives:
1. Identify the core challenges of social science research, and understand the different types of research designs (experimental, observational, etc.) traditionally used to address them. Consider where quantitative methods fit in this picture
2. Understand the key ideas behind statistical inference (i.e., sampling, confidence intervals, hypothesis testing), what they do, and why they matter
3. Describe and visualize data, both for individual variables and relationships between them
4. Appropriately use, interpret, and understand simple and multiple linear regression models in applied research
5. Diagnose and address common regression problems, and add common extensions such as interaction terms that appear in real-world applications
6. Gain preliminary exposure to key non-linear classes of models, and build a foundation that will allow you to explore them more fully in the future
7. Conduct basic data wrangling, analysis, and visualization in R

Course Materials:

Good news: all of the materials used in the course are free (you're welcome)! For the readings, we will use a free, online statistics textbook well-suited for our purposes, supplemented by occasional journal articles. The textbook is listed below:

-*Quantitative Research Methods for Political Science, Public Policy, and Public Administration, 4th Ed. with Applications in R,* by Jenkins-Smith et al. (henceforth QRMPS)
        -Link: https://bookdown.org/josiesmith/qrmbook/

For data analysis, we will use the popular open-source computing program R. You can download R for free from www.r-project.org. R is widely used for statistics and data analysis in government, the private sector, and academia, combining the ability to run conventional statistical analyses with strong graphing (as well as scraping and programming) capabilities. If you want to supplement the exposure to R that you get in the class, there are many great resources online. In particular, I would recommend the following:

-R for Data Science, by Grolemund and Hadley-Wickham (a comprehensive, general intro to R)
    -Link: https://r4ds.had.co.nz
-Quantitative Politics with R, by Larsen and Fazekas (a streamlined intro for political scientists)
    -Link: http://qpolr.com

We will be working with R through RStudio, which is the premium "front end" and user interface for the program. RStudio offers a more intuitive and user-friendly experience as well as a number of other nice features for users. It is available here: https://rstudio.com

<div align="center">Course Evaluation:</div>

Problem Sets:
The main assignments in the course will be a series of six problem sets that make students put the techniques and models we learn in class into practice. The problem sets will be assigned every two to three weeks as listed below in the course schedule portion of the syllabus. They will cover the following six topics: (1) data manipulation, (2) data visualization, (3) basic inference, (4) bivariate associations, (5) simple linear regression, and (6) multiple linear regression. Each problem set will be worth 10% of your final grade. You may discuss these assignments with other students, but you must write and submit your own work. Note that the problem sets should be completed with R and submitted in MS Word (after the first one, we will use R Markdown).

Final Project:
There are three options for the final project in the course, all of which force you to extensively use techniques from class on a research problem. Whichever option you choose, the assignment is due **Wednesday, December 18th.** It should be submitted in Word and should be at least 10-pages long (double-spaced, with 12-point Times New Roman font and 1-inch margins).
    (1) The first option is to analyze one of several datasets that I give you. I have not yet fully decided which ones these will be, but they will likely include popular survey databases like the American National Election Studies (ANES) and World Values Survey (WVS). These datasets will have outcomes that can be readily analyzed using linear regression. You will pose a research question about what causes one of these outcomes, formulate hypotheses that can be tested using the data, and use a mixture of techniques from class – including exploratory data analysis, bivariate association, and multiple regression – to investigate them. Note that this is probably the easiest option, since you do not have to find or clean your own data.
    (2) The second option is to replicate and extend an existing research paper in the field that uses multiple linear regression (it will usually just say "OLS" in the paper). To do this, you must first identify a suitable article (e.g., must be published in a scholarly journal, use the appropriate method, and have the data and code posted online). Then, you must

try to *replicate* the authors' results, and *extend* them in some way. This could be adding new variables or interaction effects to their models, checking for influential outliers or other regression problems, or in some other way building on or tweaking what they did. If you wish to pursue this option, you must submit a one-page proposal identifying and briefly summarizing the article, and explaining how you plan to replicate and extend it, by **Friday, November 20th.** This option is provided so that those who wish to analyze a paper (and dataset) in their specific area of interest can do so, though it is likely to be more work than the first option.

(3) The third option is to conduct an analysis of your own data using techniques from class, including multiple regression. This requires that you already have or can find data that are amenable to multiple regression and are otherwise suitable to analyze. As with the first option, you must pose a research question about the causes of an outcome in these data, develop testable hypotheses about what is driving it, and examine your hypotheses using descriptive and explanatory methods, judging whether they are supported or not. If you do wish to pursue this option, you must submit a two-page proposal identifying and briefly summarizing the data, and explaining how you plan to analyze it, by **Friday, November 20th**. This option is provided so that students who are very comfortable with the material and have their own data in hand, or at least realistically in sight, can work on their own research; it will be the most work of the three.

Note that if you submit a proposal for option #2 or option #3, I will do my best to help shape it and set it up for success if possible, but also reserve the right to "reject" it if it is not feasible or suitable for the course (i.e., data can't be collected, linear regression can't be used, etc.). If that is the case, you can of course pursue the idea in the future, but will have to do option #1 in this course.

Attendance and Participation:
While quantitative methods courses often require somewhat of a traditional lecture format, active and thoughtful participation by students is critical to your (and our general) success this semester. In terms of participation, students should carefully read the assigned readings prior to each course meeting and come ready to discuss them, whether in person or via Zoom. In terms of attendance, students are allowed two free absences, but will be penalized for (unexcused) absences after that. The attendance and participation policy is also designed to be flexible and accommodate different types of learners. In particular, students who do not attend a class either in-person or via Zoom can earn attendance and participation credit by watching the lecture and posting a brief, one-paragraph reaction on our Canvas discussion board within the next week. In these posts, you can simply share a thought or reaction to the lecture (something that made sense, something you didn't fully grasp, a question you had, a new idea or thought that struck you, etc.).

Problem sets	60% (10% each)
Final project	30%
Attendance/Participation	10%
Total	100%

Grading Policies

All assignments have due dates indicated on the syllabus. In general, submitting assignments on time helps me provide feedback in a more timely and efficient manner (this is especially crucial

with the problem sets, as I generally won't be able to go over them until they are all submitted). However, I do know that sometimes life happens. If you cannot submit an assignment on time, the default will be that you will be eligible for 90% of the grade the first 48 hours that the assignment is late. If you have to submit beyond 48 hours past a due date, please contact me as soon as possible so we can make arrangements.

Any grievances you may have with a grade must be submitted in writing to the instructor within one week of grading and will lead if appropriate to a regrading of your assignment. Do note that your grade may be raised or lowered in this process. Final grades will be based on the following grade scale, with rounding at instructor discretion:

Undergraduate Students Grading Scale:
A = 90-100   B = 80-89.9   C = 70-79.9
D = 60-69.9   R = <60

Graduate Students Grading Scale:

		B+	87 to 89.9	C+	77 to 79.9	D+	67 to 69.9
A	93 to 100	B	83 to 86.9	C	73 to 76.9	D	60 to 66.9
A-	90 to 92.9	B-	80 to 82.9	C-	70 to 72.9	E/F	0 to 60

Expectations for In-Person Participation:

In order to attend class in person, I expect you will abide by all behaviors indicated in **A Tartan's Responsibility**, including any timely updates based on the current conditions. In terms of specific expectations for in-person students, this includes:
o   entering the classroom via the designated ingress route with appropriate physical distancing,
o   wearing a facial covering throughout class,
o   sitting in the seats with appropriate spacing (and not moving furniture),
o   using the sanitizing wipes available in the classroom to wipe surfaces (e.g., your desk, tablet arm) upon entry and exit,
o   exit the classroom at my direction, proceeding in row-by-row fashion, following the designated egress route and maintaining proper distancing.
o   Facial coverings. If you do not wear a facial covering to class, I will ask you to put one on (and if you don't have one with you, I will direct you to a distribution location on campus). If you do not comply, please remember that you will be subject to student conduct proceedings, up to and including removal from CMU. Accordingly, I will be obliged to take other measures for the safety of the whole class.

Expectations for Zoom Participation:

In our course, we will be using Zoom for synchronous (same time) sessions. The link is available on our course page on Canvas. Please make sure that your Internet connection and equipment are set up to use Zoom and able to share audio and video during class meetings. (See *this page* from Computing Resources for information on the technology you are likely to need.) Let me know if there is a gap in your set-up as soon as possible, and we can see about finding solutions.

Sharing video: Being able to see one another helps to facilitate a better learning environment and promote more engaging discussions. Therefore, students are encouraged to have their cameras on during lectures and discussions. However, I also understand there may be reasons students would not want to have their cameras on. If you have any concerns about sharing your video, please email me as soon as possible and we can discuss possible adjustments. Note: You may use a background image in your video if you wish; just check in advance that this works with your set-up.

Other guidelines: During our class meetings, please keep your mic muted unless you are sharing with the class or your breakout group. If you have a question or want to answer a question, please use the chat or the "raise hand" feature (available when the participant list is pulled up). I will do my best to monitor these channels in order to call on students to contribute.

Transition to Fully Remote After Thanksgiving:

Following university-wide policies this semester, we will transition to a fully remote format after Thanksgiving (that is, for the last two weeks of the semester, 12/1-12/3 and 12/8-12/10). We will meet synchronously on Zoom at the usual class times for these two weeks and otherwise conduct our business as normal. Students are encouraged to attend these meetings live, but are allowed to earn attendance and participation credit by watching the recorded lectures and posting short, one-paragraph responses on Canvas (that is, doing the normal alternative attendance and participation option available during the semester). Course expectations will otherwise be the same.

Other Potential Shifts to Fully Remote Format:

If the class needs to go fully remote at any other point this fall, you will receive an email from me and an announcement will be posted on Canvas. If this occurs, we will plan to use the same Zoom link available on Canvas for remote learners throughout the semester for our new course meetings. At any point during the semester, you can also choose to participate in the class remotely yourself. If you decide to switch to remote for a sustained period of time, please try to let me know by email so that I can prepare the breakout rooms and in-class sessions appropriately. In addition, if you are able, let me know the expected length of your remote engagement.

Inappropriate Technology Use:

This semester involves regular use of technology in class – both for in-person and remote students. Research has shown that divided attention is detrimental to learning, so I encourage you to close any windows not directly related to what we are doing while you are in class. Please turn off your phone notifications and limit other likely sources of technology disruption, so you can fully engage with the material, each other, and me. This will create a better learning environment for everyone.

Note on Recording of Lectures:

All synchronous classes will be recorded on Zoom so that students in this course (and only students in this course) can watch or re-watch past class sessions. Please note that breakout rooms will not be recorded. I will make recordings available on Canvas as soon as possible after each class session (usually within 3 hours of the class meeting). Recordings will live in our Canvas website. Please

note that you are not allowed to share these recordings. This is to protect your FERPA rights and those of your fellow students.

Academic Integrity/Misconduct:

All work you do in this course is expected to be your own. Absolutely no cheating or plagiarism (use of someone else's words or ideas without properly citing them) will be tolerated. Any cases of cheating or plagiarism will be reported to the university according to the University Policy on Academic Integrity. For questions about this policy, see http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html

Accommodations for Disabilities:

If you have a disability and have an accommodations letter from the Disability Resources office, please discuss your accommodations or needs with me as early in the semester as possible. I will work with you to ensure appropriate accommodations are provided. If you suspect you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.

Maintaining Personal Wellness:

General guidance: Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep, and taking some time to relax. This will help you achieve your goals and cope with stress. All of us benefit from support during times of struggle. You are not alone. There are many helpful resources on campus, and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is key. If you or anyone you know experiences any academic stress, difficult life events, or feelings of anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help. Call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty, or a family member you trust for help in getting connected to support services.

For fall 2020: This semester is unlike any other. We are all under a lot of stress and uncertainty at this time. Attending Zoom classes all day can take its toll on our mental health. Make sure to move regularly, eat well, and reach out to your support system or me if you need to. We can all benefit from support in times of stress, and this semester is no exception.

Note on Diversity and Inclusion:

Respect for Diversity: It is my intent that students from all diverse backgrounds and perspectives be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the efficacy of the course for you personally

or for other students or student groups. In addition, if any of our class meetings conflict with your religious events, please let me know so that we can make arrangements for you.

Course Schedule:

Tuesday, 9/1 & Thursday, 9/3. Introduction and "big picture."
- Reading: Ch. 1 of QRMPS
- R module: intro

Tuesday, 9/8 & Thursday, 9/10. Introduction and "big picture."
- Reading: Ch. 2 of QRMPS
- R module: dplyr

9/15: Problem Set #1 (data manipulation) distributed, due 9/22

Tuesday, 9/15 & Thursday, 9/17. Description/exploratory data analysis.
- Reading: Ch. 3 of QRMPS
- R module: ggplot (part 1)

Tuesday, 9/22 & Thursday, 9/24. Description/exploratory data analysis.
- Reading (optional): Ch. 3 of R for Data Science. https://r4ds.had.co.nz
- R module: ggplot (part 2)

Tuesday, 9/29 & Thursday, 10/1. Probability useful for statistics.
- Reading: Ch. 4 of QRMPS
- R module: getting/importing data

10/5: Problem Set #2 (data visualization) distributed, due 10/12

Tuesday, 10/6 & Thursday, 10/8. Foundations of statistical inference.
- Reading: Ch. 5 of QRMPS (only 5.1-5.3)
- R module: basic inference

Tuesday, 10/13 & Thursday, 10/15. Bivariate associations for continuous DVs.
- Reading: Ch. 5 (only 5.4) and Ch. 6 (only 6.2-6.4) of QRMPS
- R module: correlations and t-tests

10/16: Problem Set #3 (basic inference) distributed, due 10/23

Tuesday, 10/20 & Thursday, 10/22. Bivariate associations for categorical DVs.
- Reading: Ch. 6 (only 6.1) of QRMPS
- Reading: TBD, additional reading on ANOVA
- R module: Chi-squared and ANOVA

Tuesday, 10/27 & Thursday 10/29. <u>Simple linear regression.</u>
- Reading: Ch. 7-9 of QRMPS
- R module: linear regression

11/1: Problem Set #4 (bivariate associations) distributed, due 11/8

Tuesday, 11/3 – no class. Go vote!

Thursday, 11/5. <u>Data merging day.</u>
- Reading (optional): Ch. 13 of R for Data Science. https://r4ds.had.co.nz
- R module: merging data

Tuesday, 11/10 & Thursday 11/12. <u>Regression diagnostics.</u>
- Reading: Ch. 10 of QRMPS
- R module: regression diagnostics

11/16: Problem Set #5 (linear regression) distributed, due 11/23

Tuesday, 11/17 & Thursday, 11/19. <u>Multiple linear regression.</u>
- Reading: Ch 12-13 of QRMPS (Ch. 11 optional)
- R module: multiple regression

11/20: deadline for proposals if doing final paper option #2 or #3

<u>Note: all classes will be online/remote only after this point</u>

Tuesday, 11/24. <u>Multiple regression extensions, part 1.</u>
- Reading: Ch. 14 (only 14.1, 14.3) of QRMPS
- R module: dummies and standardized coefficients

Thursday, 11/26 – no class, Thanksgiving!

Tuesday, 12/1 & Thursday, 12/3. <u>Multiple regression extensions, part 2.</u>
- Reading: Ch. 14 (only 14.2) and Ch. 15 (only 15.2.7) of QRMPS
- Reading: Brambor, Thomas, William Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses" *Political Analysis* 14(1): 63-82.
- R module: multicollinearity and interaction effects

12/6: Problem Set #6 (multiple regression) distributed, due 12/13

Tuesday, 12/8 & Thursday, 12/10. <u>Logistic regression and GLM.</u>
- Reading: Ch. 16 of QRMPS
- R module: logistic regression

Final paper/project due by Wednesday, 12/18