# Applied Methods and Statistics in Secondary Data Analysis

CRIM 795-001, 3 Credits, Spring 2021

Instructor: Evan Lowder, Ph.D.
Email: elowder@gmu.edu
Office: Enterprise 308 and via Zoom
Office Hours: By appointment

**Course Overview:**
This course is designed to provide a graduate-level introduction to applied research methods and statistics for secondary data. Topics will include methodological considerations for developing and testing research questions with secondary data, including threats to internal validity, data quality and pre-processing, and model development. Analytic strategies for secondary data analysis will include hierarchical and multivariable regression modeling, propensity score analysis, multilevel modeling, and other forms of longitudinal data analysis. This course emphasizes student-driven learning. Active, high-quality student participation is expected.

**Student Learning Outcomes:**
1. Develop theoretically, empirically, and/or practically relevant research questions that can be answered with secondary data.
2. Interpret methodological and statistical recommendations from peer-reviewed research.
3. Identify, design, and implement a modeling approach based on data assumptions and research questions.
4. Communicate research findings via oral presentation and written work.

**Course Description (from University Catalog):**
Recent developments in field, or topics not covered by regularly listed courses. Notes: Topics vary. May be repeated when topic is different. May be repeated within the term for a maximum 15 credits.

**Recommended Prerequisites:**
CRIM 780 (Research Methods), CRIM 782 (Statistics I), CRIM 783 (Statistics II)

**Required Text:**
Readings as assigned.

**Required Technology:**
Students will need access to a laptop with Microsoft Office installed (recommended). Students have access to Microsoft Office 365 subscription free of charge with their Mason ID. This class will rely heavily on Stata and R/R Studio. R and R Studio can be downloaded free of charge at: https://www.r-project.org/. Students have the option to purchase a Stata student license at: https://www.stata.com/order/new/edu/profplus/student-pricing/. However, purchasing a license is not required. Stata can be accessed via Mason's Citrix Virtual Lab. Instructions for accessing the Citrix Virtual Lab and connecting to Mason's Virtual Private Network (VPN) are available here: https://its.gmu.edu/service/citrix-virtual-lab/.

**Course Management:**
Blackboard will be used for hosting course materials (i.e., syllabus, readings, assignment descriptions, supplementary materials) and assignment submissions. To access Blackboard, navigate to https://mymasonportal.gmu.edu/ and login with your Mason NetID and password, then navigate to 'Courses' → 'Fall 2021 → '202170.78673 CRIM-795-001 (Fall 2021).' If necessary, remote course sessions will take place via Zoom.

**Course Format:**
This course is designed as a synchronous, in-person course. Course sessions will follow a similar format:

1. Lecture: Dr. Lowder will provide a brief lecture on the week's topic.
2. Article Review and Discussion: Two students each week will lead a review and discussion on methods readings.
3. Secondary Data Application: Dr. Lowder will lead tutorials on methodological strategies and data analysis. Students will replicate approaches and work collaboratively to apply strategies/code to their own datasets.

The course plan is subject to change with advance notice.

**A Note About COVID-19:**
I acknowledge that the return to campus comes with unique challenges and uncertainties. If necessary, I will revise the course format and schedule to accommodate changing COVID-19 circumstances. You are encouraged to reach out and communicate with me regarding your changing individual circumstances. To the extent possible and working within the constraints of University policy, I will try to accommodate your needs to help you be successful in this course. Note that per University policy, you are required to complete the daily Mason COVID Health Check and wear a mask during class sessions. Students who do not comply with these policies will be asked to leave the classroom.

**Grading:**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A+ | 97-100% | B+ | 87-89% | C+ | 77-79% | D | 60-69% |
| A | 93-96% | B | 83-86% | C | 73-76% | F | <60% |
| A- | 90-92% | B- | 80-82% | C- | 70-72% | | |

Note: Grades will not be automatically rounded (e.g., a 92.99 will be graded as an A-, not an A).

**Course Requirements:**

***Article Review and Discussion – (10% of final grade; 30 points)***
One of the goals of this class is to increase your comfort with reading and extracting key conclusions from statistical methods articles. During select class periods, two students will be assigned to provide a review of assigned readings and lead a short class discussion on key findings. Each student will present twice over the course of the semester.

***Secondary Data Analysis Critique – (10% of final grade; 30 points)***
During the week of the American Society of Criminology (ASC) conference, you will attend a presentation involving secondary data and prepare a brief write-up evaluating the authors' methodological and analytic decisions. If you do not attend the ASC conference, you may substitute the presentation with a paper reporting on secondary data findings.

***Statistics Practice Assignments – (30% of final grade; 90 points)***
Over the course of the semester, students will complete six statistics practice assignments on analytic strategies reviewed in class. These assignments are designed to be completed during class time and to provide students with hands-on experience working with Stata and R code to conduct analyses, interpreting output, and writing results. Each assignment will be worth 15 points for 90 points total.

***Research Paper – (50% of final grade; 150 points)***
The major assignment will be a 25-page, double-spaced research paper applying relevant methodological and analytic strategies discussed in class to a discrete research question using a student-identified secondary dataset. Students will have the opportunity to make gradual progress on the project throughout the semester through the following assignments:

- Dataset Description (10 points): Students will identify a secondary dataset that they have permission to use, explain the dataset structure, key variables, sample size, sampling strategy, and population from which the sample was drawn.

- Research Questions and Variables (10 points): Students will identify 1-3 research questions that will be asked using the secondary data and variables used to answer the question, including operationalization.

- Concept Proposal (30 points): Students will prepare a 2-3 page proposal to describe their research project. Sections will include an Introduction (Literature Review, Study Context [if relevant], Problem Statement, Present Study); Method (Data Source and Sample, Variables, Analytic strategy); and Significance for Research, Practice, and Policy

- Research Paper (100 points): Students will prepare a 25-page, double-spaced research paper (exclusive of tables, figures, and references) containing an Introduction, Method, Results, and Discussion section. Final presentation points will be included in the research paper grade. Additional guidelines and rubric are available on Blackboard.

**Contacting the Instructor:**
The best way to contact me is via email or by appointment. Please email if you would like to set up a time to meet. I will check email regularly during the workweek. You can expect a response within 24 hours of sending me an email during the workweek, though often sooner. There is no guarantee of a response over weekends and holidays. Be proactive! Do not save your questions or concerns until the last minute.

**Late Assignments:**
Late assignments will only be accepted in the case of a documented emergency or at my discretion.

**Academic Misconduct Policy:**
Academic misconduct is a serious offense, and I will take seriously any instances of plagiarism in this course. Every assignment you submit via Blackboard Collaborate for this course is reviewed for possible plagiarism. If you are found responsible for academic misconduct, the default sanction will be a failing grade (i.e., "0" on that assignment). Exceptions to the policy will be made on a case by case basis. Acting with academic integrity means that you (1) submit your work and your work alone; 2) submit new work for this class (i.e., do not "recycle" previous assignments or self-plagiarize); and 2) use a documented citation style (e.g., APA style) to properly attribute the works and ideas of others. If you are ever in doubt, err on the side of over-citing. Please seek clarification from me when you are unclear about the rules or guidelines for a given assignment. You are expected to abide by the Mason Honor Code. For more information on the Honor Code, please see: https://oai.gmu.edu/mason-honor-code/

**Course Schedule**

Subject to change with advance notice.

| Date | Topic | Assignment |
|---|---|---|
| **Part 1: Planning and Design** | | |
| Week 1: 8/26 | Introduction to Secondary Data Analysis | |
| Week 2: 9/2 | Dataset Identification and Sampling | |
| Week 3: 9/9 | Developing Research Questions and Hypotheses | Dataset Due |
| Week 4: 9/16 | Assessing Secondary Data | Research Questions and Variables Due |
| Week 5: 9/23 – REMOTE CLASS | Designing a Model | |
| **Part 2: Secondary Data Analysis** | | |
| Week 6: 9/30 | Multivariable Modeling Review Part 1 | Concept Proposal Due |
| Week 7: 10/7 | Multivariable Modeling Review Part 2 | SPA #1 |
| Week 8: 10/14 | Multivariable Modeling Extensions | SPA #2 |
| Week 9: 10/21 | Propensity Score Analysis | SPA #3 |
| Week 10: 10/28 | Econometrics Approaches | SPA #4 |
| Week 11: 11/4 | Multilevel Models | |
| Week 12: 11/11 | Longitudinal/Panel Data | SPA #5 |
| Week 13: 11/18 – (ASC – Class TBD) | Methods and Results Reporting | SPA #6<br><br>Secondary Data Analysis Critique Due 11/22 |
| NO CLASS Thanksgiving Break: 11/25 | | |
| Week 14: 12/2 | Research Presentations | Research Presentation Due |
| Week 15: 12/9 – NO CLASS | Final Exam Period | Final Research Paper Due |

## Reading List

***Week 1: Introduction to Secondary Data Analysis***

Woodward, V. H., Webb, M. E., III, O. H. G., & Copes, H. (2016). The current state of criminological research in the United States: An examination of research methodologies in criminology and criminal justice journals. *Journal of Criminal Justice Education*, 27(3), 340–361. https://doi.org/10.1080/10511253.2015.1131312

***Week 2: Dataset Identification and Sampling***

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14(2), 81–100. https://doi.org/10.1037/a0015914

Lynch, J. (2018). Not even our own facts: Criminology in the era of big data. *Criminology*, 56(3), 437–454. https://doi.org/10.1111/1745-9125.12182

Mooney, S. J., & Garber, M. D. (2019). Sampling and sampling frames in big data epidemiology. *Current Epidemiology Reports*, 6(1), 14–22. https://doi.org/10.1007/s40471-019-0179-y

Wallace, D., Walker, J., Nelson, J., Towers, S., Eason, J., & Grubesic, T. H. (2021). The 2020 coronavirus pandemic and its corresponding data boon: Issues with pandemic-related data from criminal justice organizations. *Journal of Contemporary Criminal Justice*, 10439862211027992. https://doi.org/10.1177/10439862211027993

***Week 3: Developing Research Questions and Hypotheses***

Jackson, D., Walter, G., Daly, J., & Cleary, M. (2014). Editorial: Multiple outputs from single studies: acceptable division of findings vs. 'salami' slicing. *Journal of Clinical Nursing*, 23(1–2), 1–2. https://doi.org/10.1111/jocn.12439

Kim, B., Spohn, C., & Hedberg, E. C. (2015). Federal sentencing as a complex collaborative process: Judges, prosecutors, judge–prosecutor dyads, and disparity in sentencing. *Criminology*, 53(4), 597–623. https://doi.org/10.1111/1745-9125.12090

Pratt, T. (2015). Theory testing in criminology. In A. R. Piquero & M. L. Rorie (Eds.), *The Handbook of Criminological Theory*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/gmu/detail.action?docID=4035968

Wooldredge, J., Frank, J., Goulette, N., & Travis, L. (2015). Is the impact of cumulative disadvantage on sentencing greater for black defendants? *Criminology & Public Policy*, 14(2), 187–223. https://doi.org/10.1111/1745-9133.12124

***Week 4: Assessing Secondary Data***

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4), 372–411. https://doi.org/10.1177/1094428114548590

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: A mini-review. *Frontiers in Psychology*, 0. https://doi.org/10.3389/fpsyg.2012.00322

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

*Optional:*

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1), 222. https://doi.org/10.1186/2193-1801-2-222

***Week 5: Designing a Model***

Sauerbrei, W., Perperoglou, A., Schmid, M., Abrahamowicz, M., Becher, H., Binder, H., Dunkler, D., Harrell, F. E., Royston, P., Heinze, G., Abrahamowicz, M., Becher, H., Binder, H., Dunkler, D., Harrell, F., Heinze, G., Perperoglou, A., Rauch, G., Royston, P., … for TG2 of the STRATOS initiative. (2020). State of the art in selection of variables and functional forms in multivariable analysis—Outstanding issues. *Diagnostic and Prognostic Research*, 4(1), 3. https://doi.org/10.1186/s41512-020-00074-3

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, 15(3), 250–267. https://doi.org/10.1037/a0018719

Witte, J., & Didelez, V. (2019). Covariate selection strategies for causal inference: Classification and comparison. *Biometrical Journal*, 61(5), 1270–1289. https://doi.org/10.1002/bimj.201700294

***Week 6: Multivariable Modeling Review (Part 1)***

Kleinbaum, D. G., Kupper, L. L., Nizam, A., & Muller, K. E. (2007). Multiple regression analysis: General considerations. *In Applied Regression Analysis and Other Multivariable Methods* (4th edition, pp. 114–138). Duxbury Press.

Williams, M., Grajales, C., & Kurkiewicz, D. (2019). Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research, and Evaluation*, 18(1). https://doi.org/10.7275/55hn-wk47

***Week 7: Multivariable Modeling Review (Part II)***

MacDonald, J. M., & Lattimore, P. K. (2010). Count models in criminology. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 683–698). Springer. https://doi.org/10.1007/978-0-387-77650-7_32

Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology*, 16(1), 21–43. https://doi.org/10.1023/A:1007521427059

Walters, G. D. (2007). Using Poisson class regression to analyze count data in correctional and forensic psychology: A relatively old solution to a relatively new problem. *Criminal Justice and Behavior*, 34(12), 1659–1674. https://doi.org/10.1177/0093854807307030

*Optional:*

Hoef, J. M. V., & Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology*, 88(11), 2766–2772. https://doi.org/10.1890/07-0043.1

***Week 8: Multivariable Modeling Extensions***

Busenbark, J. R., Graffin, S. D., Campbell, R. J., & Lee, E. Y. (2021). A marginal effects approach to interpreting main effects and moderation. *Organizational Research Methods*, 1094428120976838. https://doi.org/10.1177/1094428120976838

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, 26(1), 67–82. https://doi.org/10.1093/esr/jcp006

***Week 9: Propensity Score Analysis***

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28(25), 3083–3107. https://doi.org/10.1002/sim.3697

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, 163(12), 1149–1156. https://doi.org/10.1093/aje/kwj149

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*: *A Review Journal of the Institute of Mathematical Statistics*, 25(1), 1–21. https://doi.org/10.1214/09-STS313

*Optional:*

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized

boosted models. *Statistics in Medicine*, 32(19), 3388–3414. https://doi.org/10.1002/sim.5753

***Week 10: Econometrics Approaches***

Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: What, why and how. *Journal of Experimental Criminology*, 2(1), 23–44. https://doi.org/10.1007/s11292-005-5126-x

Angrist, J. D., & Pischke, J.-S. (2009). Parallel worlds: Fixed effects, differences-in-differences, and panel data. In *Mostly Harmless Econometrics: An Empiricist's Companion* (1st edition, pp. 221–246). Princeton University Press.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2020). A practical introduction to regression discontinuity designs: Foundations. *Cambridge University Press*. https://doi.org/10.1017/9781108684606

*Optional:*

Baiocchi, M., Cheng, J., & Small, D. S. (2014). Instrumental variable methods for causal inference. *Statistics in Medicine*, 33(13), 2297–2340. https://doi.org/10.1002/sim.6128

Wildeman, C., & Andersen, S. H. (2017). Paternal incarceration and children's risk of being charged by early adulthood: Evidence from a Danish policy shock. *Criminology*, 55(1), 32–58. https://doi.org/10.1111/1745-9125.12124

***Week 11: Multilevel Models***

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., & Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research*, 79(1), 69–102. https://doi.org/10.3102/0034654308325581

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48(1), 85–112. https://doi.org/10.1016/j.jsp.2009.09.002

Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 3, pp. 1570–1573). Wiley.

*Optional:*

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, 26(1), 441–462. https://doi.org/10.1146/annurev.soc.26.1.441

Raudenbush, S. W., & Bryk, A. S. (2001). The logic of hierarchical linear models. In *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd edition, pp. 16–37). SAGE Publications, Inc.

Snijders, T. A. B., & Bosker, R. J. (2011). The random intercept model. In *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Second edition, pp. 41-73). SAGE Publications Ltd.

Snijders, T. A. B., & Bosker, R. J. (2011). The hierarchical linear model. In *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Second edition, pp. 74-93). SAGE Publications Ltd.

***Week 12: Longitudinal/Panel Data***

Dugan, L. (2010). Estimating effects over time for single and multiple units. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 741–763). Springer. https://doi.org/10.1007/978-0-387-77650-7_32

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). Longitudinal data: Basic concepts. In *Applied Longitudinal Analysis* (pp. 19–45). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119513469.ch2

*Optional:*

Raudenbush, S. W., & Bryk, A. S. (2001). Applications in the study of individual change. In *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd edition, pp. 160–202). SAGE Publications, Inc.