Statistics For Political Research II: Recitation Period

Alexander (Alex) Dean*

Recitation Period: Likely Friday 2:00pm-3:00pm, Commons Center 349
Office Hours: Likely Friday 3:00pm-5:00pm, Political Science TA Office[1]

Overview of Recitation Period

The recitation period (one hour, weekly) for STAT II serves as an addendum and extension of the course, giving students additional time to review and apply key concepts from the standard instruction time. We will cover the implementation of various causal inference methods from class in `R`, practical skills that will aid in the analysis of causal identification techniques, and student questions from lectures and problem sets.

Learning Objectives

By the end of the semester, students should be able to:

(i) Gain competencies and skillsets for reading a technical methods paper

(ii) Implement causal inference techniques discussed in class/recitation period using `R` for their own research projects

(iii) Use knowledge of assumptions, weaknesses, and strengths of the "canonical" identification methods from class/recitation period to motivate self-exploration of recent methodological advances for causal inference methods for their own projects

Required Software

A majority of recitation periods will involve using `R`, a statistical programming language. As such, it is important that you have a baseline knowledge and familiarity with `R`. By baseline knowledge and familiarity, this involves tasks such as:

(i) Correctly setting a working directory

(ii) Calling packages

*alexander.r.dean@vanderbilt.edu

[1]While these are my "official" office hours (I am also TA'ing for Gosha's undergraduate class and need a time for them to meet with me), feel free to come to my carrel (304E in LAPOP) or shoot me an email/text if you have questions at other points otherwise during the week.

(iii) Having command of basic syntax/data types/etc. (what does df <- dataframe[, -c(5:7)] do?)

(iv) Being comfortable with troubleshooting problems that inevitably arise from the coding process

If you have never worked in `R`, or you feel that you current skills/knowledge of statistical programming in `R` need improvement to reach the baseline familiarity that I have laid out above, please come talk with me and/or Gosha as soon as possible. We can help as well as provide resources that will get you to the baseline knowledge and familiarity outlined above. It is not an issue if this is the case, but it is important that we work so that you can fully participate in the recitation period and have it be as useful as possible to you.[2]

A Tentative Schedule

Disclaimer: This schedule is subject to change, depending on my and Gosha's determinations on student needs and our overall progress through the course together.[3]

Date	Title	Topic	Associated Readings
1/17	Reading Technical Papers & Quarto Overview	Reading a technical paper requires a different skillset than that required to skim a paper for a seminar. We'll go over strategies for maximizing comprehension. We will also discuss the basics of using Quarto, a markdown-based open-source system that is superseding rmarkdown.	• Aronow and Samii (2016) • Quarto Guide
1/31	The Importance of Clarity for Replication	Making your research replicable doesn't stop at making your data public. Clear, concise language of the data cleaning and analysis that you performed is also crucial. We'll go over an exercise that demonstrates its importance.	• Breznau et al. 2022
1/24	Bootstrapping Practicum	Bootstrapping, a type of replacement sampling, has become increasingly important for statistical inference. I will introduce bootstrapping, some notable uses, and demonstrate its utility in practical applications with `R`	• Efron & Tibshirani 1986 • Horowitz 2019
2/7	Designing Pre-Analysis Plans	A crucial step in implementing an experiment, pre-analysis plans are increasingly becoming required for publishing. We'll review some key parts of PAPs, including power calculations and multiple to highlight their utility.	• EGAP Method Guides

[2]Regardless of your existing familiarity and comfortableness with using `R` for your statistical programming needs, I have found the following resources to be immensely helpful as I have learned how to use `R` for my own purposes. `R for Data Science` is a good, practical textbook that is geared towards introductory topics. `Advanced R` is a continuation of `R for Data Science` with an eye for more complex topics. Finally, Nick Klein has a supremely useful Youtube series `Introduction to R for Economists`.

[3]Note that the associated readings are not strictly necessary to be read prior to the recitation period (although I would encourage you to do so), but are the sources for the topics that we encounter on that date.

Date	Title	Topic	Associated Readings
2/14	Matching Practicum	An implementation of Sekhon 2009 with the `Matching` package in `R`	• Sekhon 2009 • Sekhon 2011
2/21	Instrumental Variables Practicum	An implementation of Nunn & Wantchekon 2010 with the `AER` package in `R`	• Nunn and Wantchekon 2010 • Kleiber & Zeileis 2024
2/28	Additional Day	Topic based on student interest	
3/7	Understanding Fixed Effects	Do fixed effects still produced biased ATE estimates? Can we quantify this bias theoretically? This recitation period we will solve a formal proof from Gibbons et al. 2019 and explore alternative methods of estimating FE that are less subject to bias in the point estimate for ATEs	• Gibbons et al. 2019
3/21	DiD Practicum	An implementation of Callaway & Sant'Anna (2021) with the `did` package in `R`	• Callaway & Sant'Anna 2021 • `did` Package setup
3/28	Synthetic Control Practicum	An implementation of Abadie & Gardeazabal 2003 with the `Synth` package in `R`	• Abadie & Gardeazabal 2003 • Abadie et al. 2011
4/4	RDD Practicum	An implementation of Lee 2008 with the `rdrobust` package in `R`	• Lee 2008 • Calonico et al. 2013
4/11	Implementing Sensitivity Analysis	Uses of sensitivity analyses are becoming increasingly common form of testing robustness of models as well as for the causal conclusions we can draw when assumptions, like that of parallel trends, are violated. We'll go over a general overview followed by an implementation of sensitivity analysis, using the `honest` package in `R`	• Ramachan & Roth 2023 • Saltelli 2002

Other Helpful Resources

Note: these resources are ones that I have personally found helpful in my own journey with statistics. You're in no way required to read these, they are placed here for those interested.

- Literature on Recent Advances in Applied Micro Methods— Christine Cai

  Christine Cai has a continuously updated list of recent methodological advancements in applied microeconomics methods, which largely overlap with the methods that we use in political science. If you want to see the current state of a method, this resource is one of the best places to start. On her website under the tab `Public Goods`, she has compiled an extensive list of helpful resources relating to applied econometrics; coding in `R` and Python; and much more.

- Causal Inference: The Mixtape— Scott Cunningham

  This online textbook is clear and lucid, providing a solid introduction to many concepts that are discussed in class and the recitation period.

- Practical Data Analysis for Political Scientists— Brenton Kinkel

  The textbook for STAT II prior to the restructuring of our statistics sequence, Brenton's book provides a good refresher for topics surrounding regression analysis, including the basics of bivariate and multivariate regression; non-spherical errors; clustered data/standard errors; and much more.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2011. "Synth: An R Package for Synthetic Control Methods in Comparative Case Studies". Pre-published, 2011. SSRN Scholarly Paper. Visited on 01/06/2025. `https://papers.ssrn.com/abstract=1958891`.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country". *American Economic Review* 93 (1): 113–132. `https://doi.org/10.1257/000282803321455188`.

Aronow, Peter M., and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60 (1): 250–267. `https://doi.org/10.1111/ajps.12185`.

Breznau, Nate, et al. 2022. "Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty". *Proceedings of the National Academy of Sciences* 119 (44): e2203150119.

Cai, Christine. 2024. "Literature on Recent Advances in Applied Micro Methods".

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods". *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225, no. 2 (2021): 200–230. `https://doi.org/10.1016/j.jeconom.2020.12.001`.

Calonico, Sebastian, Matias Cattaneo D., and Rocío Titiunik. 2015. "Rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs". *The R Journal* 7 (1): 38. `https://doi.org/10.32614/RJ-2015-004`.

Coppock, Alexander. "10 Things to Know About Pre-Analysis Plans – EGAP". egap. Visited on 01/06/2025. https://egap.org/resource/10-things-to-know-about-pre-analysis-plans/.

Coppock, Alexander. "10 Things to Know About Statistical Power – EGAP". egap. Visited on 01/06/2025. https://egap.org/resource/10-things-to-know-about-statistical-power/.

Cunningham, Scott. 2021. *Causal Inference The Mixtape*.

Efron, B., and R. Tibshirani. 1986. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy". *Statistical Science* 1 (1): 54–75. JSTOR: 2245500.

Gibbons, Charles E, Juan Carlos Suárez Serrato, and Michael B Urbancic. 2019. "Broken or Fixed Effects?" *Journal of Econometric Methods* 8 (1): 20170002.

Horowitz, Joel L. 2019. "Bootstrap Methods in Econometrics". *Annual Review of Economics* 11 (Volume 11, 2019 2019): 193–224. https://doi.org/10.1146/annurev-economics-080218-025651.

Kenkel, Brenton. *Practical Data Analysis for Political Scientists*.

Kleiber, Christian, and Achim Zeileis. 2008a. (AER: Applied Econometrics with R; visited on 01/06/2025). 2008. https://doi.org/10.32614/CRAN.package.AER.

Kleiber, Christian, and Achim Zeileis. 2008b. (AER: Applied Econometrics with R; visited on 01/06/2025). 2008. https://doi.org/10.32614/CRAN.package.AER.

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections". *Journal of Econometrics*, The Regression Discontinuity Design: Theory and Applications, 142, no. 2 (2008): 675–697. https://doi.org/10.1016/j.jeconom.2007.05.004.

Nunn, Nathan, and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa". *American Economic Review* 101 (7): 3221–3252. https://doi.org/10.1257/aer.101.7.3221.

"Quarto". Quarto. Visited on 01/06/2025. https://quarto.org/.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A More Credible Approach to Parallel Trends". *The Review of Economic Studies* 90, no. 5 (2023): 2555–2591. https://doi.org/10.1093/restud/rdad018.

Saltelli, Andrea. 2002. "Sensitivity Analysis for Importance Assessment". *Risk analysis* 22 (3): 579–590.

Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference". *Annual Review of Political Science* 12 (Volume 12, 2009 2009): 487–508. https://doi.org/10.1146/annurev.polisci.11.060606.135444.

Sekhon, Jasjeet S. 2011. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R". *Journal of Statistical Software* 42 (2011): 1–52. https://doi.org/10.18637/jss.v042.i07.

"Treatment Effects with Multiple Periods and Groups". Visited on 01/06/2025. https://bcallaway11.github.io/did/.