# Soc504: Conclusions

Brandon Stewart

Princeton

May 1, 2017

# Soc500-504 as a Beginning

# Soc500-504 as a Beginning

- We can't possibly teach you everything you would ever need to know about statistics, but we wanted to give you a foundation to learn more

# Soc500-504 as a Beginning

- We can't possibly teach you everything you would ever need to know about statistics, but we wanted to give you a foundation to learn more
- Major content this year:
  - Probability $\rightsquigarrow$ Inference $\rightsquigarrow$ Regression

# Soc500-504 as a Beginning

- We can't possibly teach you everything you would ever need to know about statistics, but we wanted to give you a foundation to learn more
- Major content this year:
  - Probability $\rightsquigarrow$ Inference $\rightsquigarrow$ Regression
  - Regression in Practice $\rightsquigarrow$ Causal Inference

# Soc500-504 as a Beginning

- We can't possibly teach you everything you would ever need to know about statistics, but we wanted to give you a foundation to learn more
- Major content this year:
  - Probability $\rightsquigarrow$ Inference $\rightsquigarrow$ Regression
  - Regression in Practice $\rightsquigarrow$ Causal Inference
  - Maximum Likelihood $\rightsquigarrow$ Generalized Linear Models

# Soc500-504 as a Beginning

- We can't possibly teach you everything you would ever need to know about statistics, but we wanted to give you a foundation to learn more
- Major content this year:
  - Probability $\rightsquigarrow$ Inference $\rightsquigarrow$ Regression
  - Regression in Practice $\rightsquigarrow$ Causal Inference
  - Maximum Likelihood $\rightsquigarrow$ Generalized Linear Models
  - Advanced Topics: Missing Data, Matching, Mediation, Regularization

# Soc500-504 as a Beginning

- We can't possibly teach you everything you would ever need to know about statistics, but we wanted to give you a foundation to learn more
- Major content this year:
  - Probability $\rightsquigarrow$ Inference $\rightsquigarrow$ Regression
  - Regression in Practice $\rightsquigarrow$ Causal Inference
  - Maximum Likelihood $\rightsquigarrow$ Generalized Linear Models
  - Advanced Topics: Missing Data, Matching, Mediation, Regularization
- Useful skills like: simulation, visualization, reproducible research

# Soc500-504 as a Beginning

- We can't possibly teach you everything you would ever need to know about statistics, but we wanted to give you a foundation to learn more
- Major content this year:
  - Probability $\rightsquigarrow$ Inference $\rightsquigarrow$ Regression
  - Regression in Practice $\rightsquigarrow$ Causal Inference
  - Maximum Likelihood $\rightsquigarrow$ Generalized Linear Models
  - Advanced Topics: Missing Data, Matching, Mediation, Regularization
- Useful skills like: simulation, visualization, reproducible research
- Also a basis in working in `R` that should be (somewhat) transferable to other languages.

# Tips for Poster Session

# Tips for Poster Session

- Use the poster session as an opportunity to vary how you pitch the work. Try different variations and see what sticks with people.

# Tips for Poster Session

- Use the poster session as an opportunity to vary how you pitch the work. Try different variations and see what sticks with people.
- Make sure you start pitches with what the paper is about and 'whose mind it will change about what.' It is temping to jump right into methods, but you have to set the stage.

# Tips for Poster Session

- Use the poster session as an opportunity to vary how you pitch the work. Try different variations and see what sticks with people.
- Make sure you start pitches with what the paper is about and 'whose mind it will change about what.' It is temping to jump right into methods, but you have to set the stage.
- You don't necessarily need to even talk about the paper you are replicating. It might be relevant, but remember you are pitching your paper, not someone else's.

# Tips for Poster Session

- Use the poster session as an opportunity to vary how you pitch the work. Try different variations and see what sticks with people.
- Make sure you start pitches with what the paper is about and 'whose mind it will change about what.' It is temping to jump right into methods, but you have to set the stage.
- You don't necessarily need to even talk about the paper you are replicating. It might be relevant, but remember you are pitching your paper, not someone else's.
- A note on setup timing.

# Tips for the Paper

# Tips for the Paper

- Be clear- show the work to other people and make sure its readable.

# Tips for the Paper

- Be clear- show the work to other people and make sure its readable.
- Establish why the problem is important, set the stakes for someone who doesn't know the literature. It is hard to do this and might take a few tries.

# Tips for the Paper

- Be clear- show the work to other people and make sure its readable.
- Establish why the problem is important, set the stakes for someone who doesn't know the literature. It is hard to do this and might take a few tries.
- The fact that the paper is short shouldn't make it faster to write. It is difficult to write short pieces, leave yourself some time.

# Tips for the Paper

- Be clear- show the work to other people and make sure its readable.
- Establish why the problem is important, set the stakes for someone who doesn't know the literature. It is hard to do this and might take a few tries.
- The fact that the paper is short shouldn't make it faster to write. It is difficult to write short pieces, leave yourself some time.
- Take your time with figures. It can make sense to create quick dummy versions for placeholders, but be sure to go back and really carefully put them together.

# Tips for the Paper

- Be clear- show the work to other people and make sure its readable.
- Establish why the problem is important, set the stakes for someone who doesn't know the literature. It is hard to do this and might take a few tries.
- The fact that the paper is short shouldn't make it faster to write. It is difficult to write short pieces, leave yourself some time.
- Take your time with figures. It can make sense to create quick dummy versions for placeholders, but be sure to go back and really carefully put them together.
- Allocate space and effort with proportion to what your reader needs to know not how long it took you to do (this is hard!)

# Tips for the Paper

- Be clear- show the work to other people and make sure its readable.
- Establish why the problem is important, set the stakes for someone who doesn't know the literature. It is hard to do this and might take a few tries.
- The fact that the paper is short shouldn't make it faster to write. It is difficult to write short pieces, leave yourself some time.
- Take your time with figures. It can make sense to create quick dummy versions for placeholders, but be sure to go back and really carefully put them together.
- Allocate space and effort with proportion to what your reader needs to know not how long it took you to do (this is hard!)
- Everything doesn't have to be perfect- but you should be making an improvement beyond the original paper in at least one dimension.

# Tips for the Paper

- Be clear- show the work to other people and make sure its readable.
- Establish why the problem is important, set the stakes for someone who doesn't know the literature. It is hard to do this and might take a few tries.
- The fact that the paper is short shouldn't make it faster to write. It is difficult to write short pieces, leave yourself some time.
- Take your time with figures. It can make sense to create quick dummy versions for placeholders, but be sure to go back and really carefully put them together.
- Allocate space and effort with proportion to what your reader needs to know not how long it took you to do (this is hard!)
- Everything doesn't have to be perfect- but you should be making an improvement beyond the original paper in at least one dimension.
- Reread "Publication, Publication" and the paper handout.

# Tips for Talking About Methods

# Tips for Talking About Methods

Not everyone will know what you now know about methods. Be clear, precise and give intuition for what you are doing and why.

# Tips for Talking About Methods

Not everyone will know what you now know about methods. Be clear, precise and give intuition for what you are doing and why.

Start simple and justify added complexity.

# Tips for Talking About Methods

Not everyone will know what you now know about methods. Be clear, precise and give intuition for what you are doing and why.

Start simple and justify added complexity.

Otherwise this will happen...

# Tips for Talking About Methods

Not everyone will know what you now know about methods. Be clear, precise and give intuition for what you are doing and why.

Start simple and justify added complexity.

Otherwise this will happen. . .

# Feedback Assignment (and feedback in general)

# Feedback Assignment (and feedback in general)

- Be charitable

# Feedback Assignment (and feedback in general)

- Be charitable
- Remember the compliment sandwich

# Feedback Assignment (and feedback in general)

- Be charitable
- Remember the compliment sandwich
- Start with a summary of how you see the major contributions

# Feedback Assignment (and feedback in general)

- Be charitable
- Remember the compliment sandwich
- Start with a summary of how you see the major contributions
- Provide suggestions where you can, but often it is helpful just to document how you reacted to the work. Regardless, be concrete

# Questions

Can you give an intuition on the differences between linear regression, fixed-effect model and generalized linear model (of which logistic, multinomial logit, mixed logit, ordered logit and poisson are of greatest interest).

# Questions

I would like to know the options on future statistics classes, especially if they are somewhat coding oriented, and not math-intensive (there will obviously need to be some math, but your class had a good balance).

# Questions

I understand why we need draws from the multivariate normal to simulate quantities of interest: we draw our coefficients all at once, and the values drawn have to sort of "make sense" with and "fit" with our variance-covariance matrix. But I still really lack intuition for how our variance-covariance matrix (in addition to our mean) produces a multivariate normal and what that multivariate normal "looks like." How does our variance-covariance matrix create or produce the multivariate normal? Could you maybe illustrate with an example?

# Questions

I feel like the missingness assumptions in order of percent of the time they are plausible is probably: (1) NI, (2) MCAR, (3) MAR. Do you agree? I think MAR is probably almost never plausible. But I also think that even if it's implausible, imputation assuming MAR is probably better than nothing (or, more precisely, better than listwise deletion) unless we're certain MCAR holds. And maybe we can incorporate imputation into some sort of sensitivity analysis. We could run our model when we do listwise deletion on our data, run another one where we do multiple imputation, and compare the results. Do you agree?

# Other Questions

# Research Is Iterative

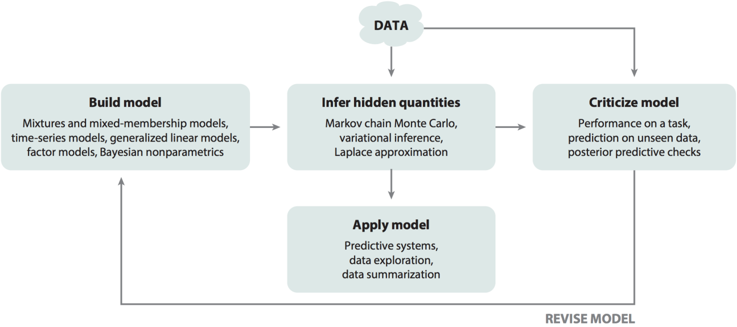


**Figure 1**

Box's loop. Building and computing with models are part of an iterative process for solving data-analysis problems. This is Box's loop, a modern interpretation of the perspective of Box (1976).

From David Blei's Paper "Build, compute, critique, repeat: Data analysis with latent variable models" *Annual Review of Statistics and Its Application*, 2014

# Tips for Balancing More Study of Quantitative Methods with Other Areas

# Tips for Balancing More Study of Quantitative Methods with Other Areas

- Do what you love

# Tips for Balancing More Study of Quantitative Methods with Other Areas

- Do what you love
- Know yourself $\rightsquigarrow$ take classes in what you have a hard time learning alone (corollary: most people don't do any methods training after grad school)

# Tips for Balancing More Study of Quantitative Methods with Other Areas

- Do what you love
- Know yourself $\rightsquigarrow$ take classes in what you have a hard time learning alone (corollary: most people don't do any methods training after grad school)
- Be led by the question

# Tips for Balancing More Study of Quantitative Methods with Other Areas

- Do what you love
- Know yourself $\rightsquigarrow$ take classes in what you have a hard time learning alone (corollary: most people don't do any methods training after grad school)
- Be led by the question
- Understand things before you publish on them

# Tips for Balancing More Study of Quantitative Methods with Other Areas

- Do what you love
- Know yourself $\rightsquigarrow$ take classes in what you have a hard time learning alone (corollary: most people don't do any methods training after grad school)
- Be led by the question
- Understand things before you publish on them
- Identify the kind of work in the field that you want to be doing

# Tips for Balancing More Study of Quantitative Methods with Other Areas

- Do what you love
- Know yourself $\rightsquigarrow$ take classes in what you have a hard time learning alone (corollary: most people don't do any methods training after grad school)
- Be led by the question
- Understand things before you publish on them
- Identify the kind of work in the field that you want to be doing
- Discuss research with colleagues who will help anchor you

# Rules of Thumb

- Analysis

# Rules of Thumb

- Analysis
  - always plot the data

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea
  - if you don't get it, you probably shouldn't do it

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea
  - if you don't get it, you probably shouldn't do it
  - implementations are not models and vice-versa

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea
  - if you don't get it, you probably shouldn't do it
  - implementations are not models and vice-versa
  - think: what's the perfect experiment?, what's the counterfactual?

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea
  - if you don't get it, you probably shouldn't do it
  - implementations are not models and vice-versa
  - think: what's the perfect experiment?, what's the counterfactual?
- Presentation

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea
  - if you don't get it, you probably shouldn't do it
  - implementations are not models and vice-versa
  - think: what's the perfect experiment?, what's the counterfactual?
- Presentation
  - present your results in graphical form

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea
  - if you don't get it, you probably shouldn't do it
  - implementations are not models and vice-versa
  - think: what's the perfect experiment?, what's the counterfactual?
- Presentation
  - present your results in graphical form
  - always write a caption explaining what someone should take away from your figure or table- if you can't write this it probably needs work

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea
  - if you don't get it, you probably shouldn't do it
  - implementations are not models and vice-versa
  - think: what's the perfect experiment?, what's the counterfactual?
- Presentation
  - present your results in graphical form
  - always write a caption explaining what someone should take away from your figure or table- if you can't write this it probably needs work
  - present the quantity you care about, not the quantity that comes out of the software package

# Rules of Thumb

- Analysis
  - always plot the data
  - simulate to check your estimator
  - there is no panacea
  - if you don't get it, you probably shouldn't do it
  - implementations are not models and vice-versa
  - think: what's the perfect experiment?, what's the counterfactual?
- Presentation
  - present your results in graphical form
  - always write a caption explaining what someone should take away from your figure or table- if you can't write this it probably needs work
  - present the quantity you care about, not the quantity that comes out of the software package
  - explain the intuition of a method (this is hard!), it should be accessible to everyone

# Goal Check

- Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences

# Goal Check

- Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
- Conduct, interpret, and communicate results from analysis using multiple regression (including dummy variables and interactions).

# Goal Check

- Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
- Conduct, interpret, and communicate results from analysis using multiple regression (including dummy variables and interactions).
- Explain the limitations of observational data for making causal claims, and begin to use existing strategies for attempting to make causal claims from observational data.

# Goal Check

- Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
- Conduct, interpret, and communicate results from analysis using multiple regression (including dummy variables and interactions).
- Explain the limitations of observational data for making causal claims, and begin to use existing strategies for attempting to make causal claims from observational data.
- Write clean, reusable, and reliable `R` code.

# Goal Check

- Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
- Conduct, interpret, and communicate results from analysis using multiple regression (including dummy variables and interactions).
- Explain the limitations of observational data for making causal claims, and begin to use existing strategies for attempting to make causal claims from observational data.
- Write clean, reusable, and reliable `R` code.
- Feel empowered working with data

# Goal Check

- Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
- Conduct, interpret, and communicate results from analysis using multiple regression (including dummy variables and interactions).
- Explain the limitations of observational data for making causal claims, and begin to use existing strategies for attempting to make causal claims from observational data.
- Write clean, reusable, and reliable `R` code.
- Feel empowered working with data
- Conduct, interpret, and communicate results from analysis using generalized linear models.

# Goal Check

- Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
- Conduct, interpret, and communicate results from analysis using multiple regression (including dummy variables and interactions).
- Explain the limitations of observational data for making causal claims, and begin to use existing strategies for attempting to make causal claims from observational data.
- Write clean, reusable, and reliable `R` code.
- Feel empowered working with data
- Conduct, interpret, and communicate results from analysis using generalized linear models.
- Conduct additional study of more advanced topics in quantitative methods

# Goal Check

- Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
- Conduct, interpret, and communicate results from analysis using multiple regression (including dummy variables and interactions).
- Explain the limitations of observational data for making causal claims, and begin to use existing strategies for attempting to make causal claims from observational data.
- Write clean, reusable, and reliable `R` code.
- Feel empowered working with data
- Conduct, interpret, and communicate results from analysis using generalized linear models.
- Conduct additional study of more advanced topics in quantitative methods
- Build a solid, reproducible research pipeline to go from raw data to final paper.

# Thank You!

# Thank You!