

Data Analysis Method

Dustin Pluta

September 18, 2018

Analysis Steps

0. Model selection: choose covariates and models to be used to answer objectives.
 - Consider confounders, precision variables, important adjustment variables, and variables related to the causal pathway or mechanism of action that is of interest.
 - Typically want the simplest model that can answer the question of interest, and which adjusts for all important confounders and precision variables.
 - **Interaction effects should be included rarely and only with strong justification.**
1. Import and Tidy data.
 - Store data in a data frame or tibble.
 - Ideal “tidy” data structure: *Every row is an observation, every column is a variable, every cell is a single value.*
2. Exploratory Analysis
 - Univariate plots: histograms and boxplots
 - Frequency tables
 - Examine structure of missing data
 - Bivariate plots: scatterplots and grouped boxplots
 - Trivariate plots: scatterplots with color
 - Correlation matrix
 - *Are there any significant violations of model assumptions?*
 - *Are there obvious outliers or potentially erroneous observations?*
 - *Should we apply any transformations to the data?*
3. Model fitting.
 - Fit the simplest model that can answer the question of interest first as a benchmark.
 - Fit the “target” model chosen in step (0).
 - (Optional) Fit a “full” model, i.e. the most complex model that is reasonable for the question.
4. Model diagnostics.
 - Plots: residuals vs fitted, qq-plot, Cook’s D
 - Consider the validity of model assumptions.
 - Goodness of fit tests.
 - R^2 , adjusted- R^2 , MSE
5. Test and interpret.
 - Hypothesis tests and confidence intervals for coefficients or contrasts of interest.
 - Plots of best fit line (if relevant).
 - Interpret significant coefficients with respect to the question of interest.
6. *Post-hoc* analysis:
 - Can we find a better model?
 - Did we learn anything that would help us improve the analysis?
 - Do the results change if we remove outliers or influential observations?
 - Did we fully answer the question of interest? Did the analysis motivate any new questions?
 - What type of follow-up study would be recommended from the results of the analysis?
7. Report
 - I. Introduction/Background: Discuss and motivate the problem setting and goals of the current analysis. Cite and comment on relevant prior studies.
 - II. Subjects and Methods: Provide tables and brief discussion of the sample characteristics. How are the measurements distributed. Discuss any findings from the exploratory analysis. Be

explicit about the methods you will use and tests you will perform. Clearly state the covariates and model design for your target model, and justify these choices from both a scientific and statistical perspective.

- III. Results: Present P -values and confidence intervals to answer the questions of interest. Interpret the results in a real-world, practical context. Perform model diagnostics and comment on model fit and possible violations of model assumptions.
- IV. Discussion: discuss the broader context of the present results, including any new questions posed by the analysis. What are the limitations of the present study and analysis? To what population do the results generalize? What type of followup study or other future work is recommended?
- V. Appendix: include additional figures, tables, and other results. Things like diagnostic plots and results from secondary model fits are usually included here.
- Always focus on questions of interest.
- Include 2 - 4 most important plots/figures in text.
- Use tables sparingly in text. Include tables in appendix for reference if needed.