READING LIST: SPAN 658
Title: Data Science for Linguists

01:940:658:01 - Spring 2025
Meetings: Mondays, 10:20–13:20

Professor: Joseph V. Casillas, PhD
Email: joseph.casillas@rutgers.edu
Office: AB 5174
Office hours: by appointment

Note: Readings must be completed *before* the class that occurs in the assigned week (i.e., readings assigned for week 2 will be discussed *in* the week 2 class, thus they should be read *before*).

Week 2:

Reading

- Wickham, H. (2015). "Teaching Safe-Stats, Not Statistical Abstinence''. In: *The American Statistician, Online Discussion*. https://nhorton.people.amherst.edu/mererenovation/17_Wickham.PDF.
- Wickham, H. and G. Grolemund (2016). "Preface''. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. ix–2. http://r4ds.had.co.nz/introduction.html.
- Wickham, H. and G. Grolemund (2016). "Data Visualization with ggplot2''. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 3–34. http://r4ds.had.co.nz/data-visualisation.html.
- Wickham, H. and G. Grolemund (2016). "Workflow: Basics''. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 37–39. http://r4ds.had.co.nz/workflow-basics.html.
- Johnson, K. (2011). "Fundamentals of Quantitative Analysis''. In: *Quantitative Methods In Linguistics*. Ed. by K. Johnson. Wiley, pp. 1–33.

Additional resources

- Field Guide to the R Ecosystem (highly recommended)
- Data science and R: How do I start?
- GitHub for the useR
- Using Git and GitHub
- Getting help

Week 3:

Reading

- Johnson, K. (2011). "Patterns and Tests''. In: *Quantitative Methods In Linguistics*. Ed. by K. Johnson. Wiley, pp. 34–69.
- Wickham, H. and G. Grolemund (2016). "Data Transformation with dplyr''. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 43–73.

Additional resources

- R for excel users
- RMarkdown cheatsheet
- RMarkdown reference
- Getting started with RMarkdown
- ggplot2 cheatsheet
- Secrets to a happy graphing life
- A Compendium of Clean Graphs in R

Week 4:

Reading

- Johnson, K. (2011). "Patterns and Tests' '. In: *Quantitative Methods In Linguistics*. Ed. by K. Johnson. Wiley, pp. 34–69.
- Wickham, H. and G. Grolemund (2016). "Workflow: Scripts' '. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 77–79.
- Wickham, H. and G. Grolemund (2016). "Exploratory Data Analysis' '. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 81–108.

Additional resources

- The tidyverse style guide
- reprex 1
- reprex 2
- Data structure basics
- Correlation game
- Standardizing and correlations

Week 5:

Reading

- Lewis-Beck, M. (1980). "Bivariate Regression: Fitting a Straight Line' '. In: *Applied Regression: An Introduction*. Ed. by M. Lewis-Beck. Sage University Paper Series on Quantitative Applications in the Social Sciences - 22. Newbury Park, CA: Sage, pp. 9–25. ISBN: 9781483381497.
- Wickham, H. and G. Grolemund (2016). "Workflow: Projects' '. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 111–116.
- Wickham, H. and G. Grolemund (2016). "Tidy Data with tidyr' '. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 147–168.

Additional resources

- tidyr tutorial
- Tidy data
- Non-tidy data
- Creating new variables with mutate and ifelse (advanced)

Week 6:

Reading

- Schroeder, L., D. Sjoquist, and P. Stephan (1986). "Linear Regression' '. In: *Understanding Regression Analysis: An Introductory Guide*. Ed. by L. Schroeder, D. Sjoquist and P. Stephan. Sage University Paper Series on Quantitative Applications in the Social Sciences - 57. Newbury Park, CA: Sage, pp. 11–28. ISBN: 9780803927582.
- Lewis-Beck, M. (1980). "Bivariate Regression: Assumptions and Inferences' '. In: *Applied Regression: An Introduction*. Ed. by M. Lewis-Beck. Sage University Paper Series on Quantitative Applications in the Social Sciences - 22. Newbury Park, CA: Sage, pp. 26–46. ISBN: 9781483381497.
- Wickham, H. and G. Grolemund (2016). "Data Import with readr' '. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 125–145.
- Wickham, H. and G. Grolemund (2016). "R Markdown' '. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 423–438.
- Slidify tutorial

Additional resources

- Fit-a-line
- Simple linear regression game
- Bivariate linear regression (shiny)
- Ordinary least squares regression visualization

Week 7:

Reading

- Lewis-Beck, M. (1980). "Multiple Regression''. In: *Applied Regression: An Introduction*. Ed. by M. Lewis-Beck. Sage University Paper Series on Quantitative Applications in the Social Sciences - 22. Newbury Park, CA: Sage, pp. 47–74. ISBN: 9781483381497.
- Schroeder, L., D. Sjoquist, and P. Stephan (1986). "Multiple Linear Regression''. In: *Understanding Regression Analysis: An Introductory Guide*. Ed. by L. Schroeder, D. Sjoquist and P. Stephan. Sage University Paper Series on Quantitative Applications in the Social Sciences - 57. Newbury Park, CA: Sage, pp. 29–35. ISBN: 9780803927582.
- xaringan tutorial 1
- xaringan tutorial 2

Additional resources

- R Markdown presentations
- Model choices/interactions
- Diagnostics for simple linear regression

Spring break (highly recommended walk through):

Reading

- Wickham, H. and G. Grolemund (2016). "Model Basics with modelr''. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 345–372.
- Wickham, H. and G. Grolemund (2016). "Model Building''. In: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Ed. by H. Wickham and G. Grolemund. O'Reilly Media, pp. 375–396.

Corresponds with chapters 22, 23, 24, and 25 online.

Week 8:

Reading

- Schroeder, L., D. Sjoquist, and P. Stephan (1986). "Problems and Issues of Linear Regression''. In: *Understanding Regression Analysis: An Introductory Guide*. Ed. by L. Schroeder, D. Sjoquist and P. Stephan. Sage University Paper Series on Quantitative Applications in the Social Sciences - 57. Newbury Park, CA: Sage, pp. 65–80. ISBN: 9780803927582.
- Berry, W. and S. Feldman (1985). "Specification Error''. In: *Multiple Regression in Practice*. Ed. by W. Berry and S. Feldman. Sage University Paper Series on Quantitative Applications in the Social Sciences - 50. Newbury Park, CA: Sage, pp. 18–25. ISBN: 9780803920545.
- Berry, W. and S. Feldman (1985). "Multicollinearity''. In: *Multiple Regression in Practice*. Ed. by W. Berry and S. Feldman. Sage University Paper Series on Quantitative Applications in the Social Sciences - 50. Newbury Park, CA: Sage, pp. 37–50. ISBN: 9780803920545.
- papaja documentation: Read documentation, install package and play with demo example.

Additional resources

- Multicollinearity
- More papaja
- R Markdown with papaja (video pt. 1)
- R Markdown with papaja (video pt. 2)

Week 9:

Reading

- Hardy, M. (1993). "Creating Dummy Variables' '. In: *Regression with Dummy Variables*. Ed. by M. Hardy. Sage University Paper Series on Quantitative Applications in the Social Sciences - 93. Newbury Park, CA: Sage, pp. 7–17. ISBN: 9780803951280.
- Hardy, M. (1993). "Using Dummy Variables as Regressors' '. In: *Regression with Dummy Variables*. Ed. by M. Hardy. Sage University Paper Series on Quantitative Applications in the Social Sciences - 93. Newbury Park, CA: Sage, pp. 18–28. ISBN: 9780803951280.
- Hardy, M. (1993). "Assessing Group Differences in Effects' '. In: *Regression with Dummy Variables*. Ed. by M. Hardy. Sage University Paper Series on Quantitative Applications in the Social Sciences - 93. Newbury Park, CA: Sage, pp. 29–63. ISBN: 9780803951280.

Additional resources

- Model choices/interactions
- StackOverflow question w/ good answer
- Dummy variables (video)

Week 10:

Reading

- Hardy, M. (1993). "Alternative Coding Schemes for Dummy Variables' '. In: *Regression with Dummy Variables*. Ed. by M. Hardy. Sage University Paper Series on Quantitative Applications in the Social Sciences - 93. Newbury Park, CA: Sage, pp. 64–74. ISBN: 9780803951280.

Additional resources

- Coding systems
- Coding schemes for categorical variables in regression

Week 11:

Reading

- McElreath, R. (2015). "Big Entropy and the Generalized Linear Model' '. In: *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Ed. by R. McElreath. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, pp. 267–290. ISBN: 9781482253481.
- Gelman, A. and J. Hill (2007). "Logistic regression' '. In: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Ed. by A. Gelman and J. Hill. Analytical Methods for Social Research. Cambridge University Press, pp. 79–108. ISBN: 9780521686891.
- Gelman, A. and J. Hill (2007). "Generalized linear models' '. In: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Ed. by A. Gelman and J. Hill. Analytical Methods for Social Research. Cambridge University Press, pp. 109–134. ISBN: 9780521686891.

Additional resources

- Generalised Linear Models in R
- Shiny GLM
- When to use logistic regression
- Why regular regression does NOT work
- What is (multivariate) logistic regression

Week 12:

Reading

- Brown, V. A. (2021). "An Introduction to Linear Mixed-Effects Modeling in R''. In: *Advances in Methods and Practices in Psychological Science* 4.1, pp. 1–19. DOI: 10.1177/2515245920960351.

Additional resources

- DeBruine & Barr (2021)
- https://shiny.psy.gla.ac.uk/lmem_sim/
- http://mfviz.com/hierarchical-models/

Week 13:

Reading

- Bayesian thinking for toddlers
- Seeing theory: Frequentist inference
- Seeing theory: Bayesian inference
- Basic Bayesian analysis

Additional reading

- Interactive visualization
- https://www.tjmahr.com/bayesplot-satrdays-2019/#/

Readings to keep you occupied after S2025

Reading

- Common statistical tests are linear models
- Bayesian statistics with R
- An introduction to Bayesian Data Analysis for Cognitive Science