

Spring 2021 Course Syllabus PSY 525/625 Categorical Data Analysis

Instructor

Jason T. Newsom, Ph.D., Office: 317F Cramer Hall Center (3rd floor), Phone: 503-725-5136, Email: newsomj@pdx.edu. Office hours: by appointment. Website: <http://web.pdx.edu/~newsomj>

Meeting Times and Location

Class: Tu-Th 10:00–11:50 AM, online via Zoom.

Text

Azen, R., & Walker, C. M. (2011). *Categorical data analysis for the behavioral and social sciences*. New York: Routledge. ISBN: 1848728360

Recommended, Optional Texts

Agresti, A. (2013). *Categorical data analysis*. New York: Wiley. ISBN: 0470463635

Bilder, C.R. (2015). *Analysis of categorical data with R*. Boca Raton, FL: CRC Press. ISBN: 1439855676

Long, J.S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Menard, Scott. (2010). *Logistic Regression: From Introductory to Advanced Concepts and Applications*. Thousand Oaks, CA: Sage. ISBN: 1412974836

Wickens, T.D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum. ISBN: 0805803777

Overview

This course is intended to introduce students to categorical data analyses. The general goal is to provide a thorough background in the conceptual aspects, statistical underpinnings, and application of this method rather than a tutorial on a specific software package. By the end of the course, students should be able analyze use categorical data analysis methods to analyze real data using current statistical software, write about, critique applications of, and read methodological articles about categorical data analysis. Prerequisites:

Students should have at least one graduate statistics course covering chi-square, ANOVA, regression, and logistic regression analysis, such as PSY 521/621 and PSY 522/622

Homeworks

There will be three homework assignments consisting primarily of data analysis and reporting of categorical analyses using R, SPSS, or SAS. I will supply homework questions and data files (usually 10-12 questions) and you will be asked to analyze your own data for at least one question on each assignment. I can assist you in finding a data set if needed. You will be able to choose among the software programs for many questions, but some analyses may not be available or convenient in all software programs. Illustrations are available in the text or will be provided in class for all types of analyses on the homework assignment. **Homework due dates are: Thurs 4/22, Thurs 5/13, Tues 6/6.** Late assignments are not accepted without substantial penalty except for cases of illness or family emergencies. Please contact me ahead of time if you are going to miss the deadline for any reason. You are welcome to work with others when running the analyses or consult them on interpretation, but **your assignment must be written in your own words.**

Grades

Grades are based on an average of the three homework assignments with total percentages assigned the following grades: ≥ 90 = A, 85-89.9 = B+, 80-84.9 = B, 75-79.9 = C+, 70-74.9 = C.

Other Resources

There are several useful electronic links on the class website.

Disabilities

I am happy to make any necessary arrangements with students who have a disability and are in need of academic accommodations. If you have not done so already, please contact the Disability Resource Center, 116 Smith Memorial Student Union, <http://www.pdx.edu/drc/>, Email: drc@pdx.edu, for assistance and any testing arrangements. I would appreciate it if you would check with me as soon as possible to discuss any needed accommodations and to make sure that I have received a faculty notification letter. If any aspects of instruction or course design result in barriers to your inclusion or learning, please let me know.

Sexual Harassment, Sexual Violence, and Discrimination

As an instructor, one of my responsibilities is to help create a safe learning environment for my students and for the campus as a whole. Please be aware that as a faculty member, I have the responsibility to report any instances of sexual harassment, sexual violence and/or other forms of prohibited discrimination. If you would rather share information about sexual harassment, sexual violence or discrimination to a confidential employee who does not have this reporting responsibility, you can find a list of those individuals or contact a confidential advocate at 503-725-5672. For more information about Title IX please complete the required student module Creating a Safe Campus in your D2L.

Spring 2021 Course Readings

Azen & Walker is a required text. All other readings available online (password protected) at the class website: <http://web.pdx.edu/~newsomj/cdaclclass>

4/1,4/6 Descriptive, Univariate Statistics, and Two Categorical Variables

Levels of scale/measurement, review of probability, descriptive statistics, distributions for binary and categorical variables, test of single proportion, univariate chi-square, estimation basics, contingency of 2 x 2 using Pearson and likelihood ratio chi-square comparing two proportions, I x J tables, measures of association, tetrachoric correlations, interrater agreement statistics

Azen & Walker, Chapters 1-3. "Introduction and Overview," "Probability Distributions," and "Proportions, Estimation, and Goodness-of-Fit"

Chapter 6 "Categorical Outcomes" In Kline, Rex B. (2013). In *Beyond significance testing: Statistics reform in the behavioral sciences*. (2nd ed.). Washington, DC: American Psychological Association.

Azen & Walker, Chapter 4 "Association between Two Categorical Variables"

Optional readings: MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, 7(1), 19-40.; Groebner, D. F., Shannon, P. W., Fry, P. C., & Smith, K. D. (2014). Chapter 2 "Graphs, Charts, and Tables—Describing your data," Business statistics. Pearson Education UK.; Banerjee, M., Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*, 27(1), 3-23.; Young, F. W., Valero-Mora, P. M., & Friendly, M. (2006). Chapter 5 "Seeing Frequency Data" in *Visual statistics: seeing data with dynamic interactive graphics*. New York: John Wiley & Sons.

4/8,4/13 Three Categorical Variables, Matched Pairs and Repeated Measures

Mantel-Haenszel, Cochran-Mantel-Haenszel, McNemar's chi-square and related conventional tests, measures analysis for categorical variables

Azen & Walker Chapter 5 "Association Between Three Categorical Variables"

Newsom (2012) Chapter 5 "Basic Longitudinal Analysis Approaches for Continuous and Categorical Variables" (pp. 163-166 only). In J.T. Newsom, R.N. Jones, & S.M. Hofer (Eds.) (2012). *Longitudinal Data Analysis: A Practical Guide for Researchers in Aging, Health, and Social Science*. New York: Routledge.

Agresti (2013). Chapter 11 Models for Matched Pairs (pp. 412-418 only). In *Categorical data analysis, third edition*. John Wiley & Sons.

4/15,4/20 Ordinal Analyses for Contingency Tables and Loglinear Models

Azen & Walker Chapter 7 "Log-Linear Models"

Wickens, T.D. (1989). Chapter 13 "Ordered Categories." *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

Nussbaum. Chapter 6 "Basic Nonparametric tests for Ordinal Data." In Nussbaum, E. M. (2014). *Categorical and nonparametric data analysis: choosing the best statistical technique*. New York: Routledge. (pp. 137-149 only)

Optional: Green, J.A. (1988). Loglinear Analysis of Cross-Classified Ordinal Data: Applications in Developmental Research. *Child Development*, 59, 1-25.

4/22 HW 1 Due

4/22,4/27 Regression Models for Noncontinuous Outcomes I: Logistic Regression Review and Diagnostics

Logistic regression (continuous and binary predictors), interactions with logistic regression, propensity scores, diagnostics for logistic regression

Azen & Walker Chapters 8 & 9 "Logistic Regression with Continuous Predictors" and "Logistic Regression with Categorical Predictors"

Hayes, A.F., & Mathes, J. (2009) Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, 41 (3), 924-936

Menard (2010). Chapter 7 "Logistic Regression Diagnostics and Problems of Inference." In Menard, S. (2002). *Applied logistic regression analysis*, second edition. Sage.

4/29,5/4 Regression Models for Noncontinuous Outcomes II: Mediation and Longitudinal Applications

Mediation, lagged regression, conditional logistic models, GEE

Iacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology*, 22(4), 582-594.

Newsom (2012) Chapter 5 "Basic Longitudinal Analysis Approaches for Continuous and Categorical Variables". In Newsom, J.T., Jones, R.N., & Hofer, S.M. (Eds.) (2012). *Longitudinal Data Analysis: A Practical Guide for Researchers in Aging, Health, and Social Science*. New York: Routledge (pp. 168-170 only)

Hanley, J. A., Negassa, A., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: an orientation. *American journal of epidemiology*, 157(4), 364-375.

5/6,5/11 Regression Models for Noncontinuous Outcomes III: Review of Generalized Linear Models for Ordinal and Multicategory Outcomes, Survival Analysis

Generalized linear models, ordinal logistic and probit models, multinomial logistic, discrete choice, discrete survival analysis

Azen & Walker Chapter 6 "Generalized linear models"

Long 1997 Chapter 5 "Ordinal Outcomes. Ordered Logit and Ordered Probit Analysis" Long, S.J. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Azen & Walker Chapter 10 "Logistic Regression for Multicategory Outcomes"

Graham, S.E., Willett, J.B., & Singer, J.D. (2012) Chapter 11 "Using discrete-time survival analysis to study event occurrence". In Newsom, J.T., Jones, R.N., & Hofer, S.M. (Eds.) (2012). *Longitudinal Data Analysis: A Practical Guide for Researchers in Aging, Health, and Social Science* (pp. 329-371). New York: Routledge.

5/13 HW 2 Due

5/13,5/18 Psychometric Analyses

Some basics of IRT, relationships between IRT and factor analysis

DeMars, C. (2010). Chapter 1 "Introduction" and Chapter 4 "Results," from DeMars, C. (2010). *Item response theory*. Oxford University Press.

Thorpe, G. L. and Favia, A., (2012). Data Analysis Using Item Response Theory Methodology: An Introduction to Selected Programs and Applications. *Psychology Faculty Scholarship*. 20. https://digitalcommons.library.umaine.edu/psy_facpub/20

Immekus, J. C., Snyder, K. E., & Ralston, P. A. (2019, May). Multidimensional item response theory for factor structure assessment in educational psychology research. *Frontiers in Education* (Vol. 4, p. 45). Frontiers.

5/20,5/25 Latent Class and Latent Transition Analysis

Introduction to latent class modeling concepts

Collins, L. M., & Lanza, S. T. (2010). Chapters 2 "The Latent Class Model" and 3 "The Relation Between the Latent Variable and Its Indicators" in *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (Vol. 718). John Wiley & Sons. Clogg, C.C. (1995). Chapter 5 "Latent Class Models" in G. Arminger, C.C.

Lanza, S. T., Patrick, M. E., & Maggs, J. L. (2010). Latent transition analysis: Benefits of a latent variable approach to modeling transitions in substance use. *Journal of Drug Issues*, 40(1), 93-120.

5/27,6/1 Sample Size, Estimation, and Practical Issues

Allison, P.D. (2008). Convergence Failures in Logistic Regression. *SAS Global Forum 2008*. Paper 360-2008

Hagenaars, J. A. (2015). Methodological Issues in Categorical Data Analysis. *Methodology*, 11, 126-141.

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.

6/8 HW 3 due (finals week)