



PennState

**STAT 544
CATEGORICAL DATA ANALYSIS I
SPRING 2021**

JOHN HUGHES
JPH264@PSU.EDU

DEPARTMENT OF STATISTICS
EBERLY COLLEGE OF SCIENCE
THE PENNSYLVANIA STATE UNIVERSITY

My website: <http://www.johnhughes.org/stat544.html>
Canvas site: <https://psu.instructure.com/courses/2098146>
Credits: 3
Meeting days: Monday, Wednesday, and Friday
Meeting time: 10:10–11 AM
Meeting place: Zoom
Office hours: 1–2 PM on Wednesday, or by appointment
TA: Nikolas Siapoutis (nzs30@psu.edu)

1. OVERVIEW

1.1. **Description.**

STAT 544 Categorical Data Analysis I, 3 credits.

Prerequisites: STAT 512 and STAT 514, or permission of instructor.

Two-way tables; generalized linear models; logistic and conditional logistic models; loglinear models; fitting strategies; model selection; residual analysis.

1.2. **Course Goals.** Students who successfully complete this course will be able to formulate, fit, select, and assess a wide variety of models for discrete data.

1.3. **Methods of Instruction and Work Expectations.** Class meetings will comprise lecture and discussion. Students are expected to attend class, participate in class discussions, and complete all homework assignments and the capstone project. Students are also expected to complete reading assignments.

1.4. **Text and Materials.** Lecture notes and other materials will be available on the course website. There is no required text. You may find useful one or more of the books listed at the end of this syllabus: Agresti (2013, 2015); Fan (2018); Faraway (2016); Gelman et al. (2014); Hoff (2009); Joe (2014); Robert (2007).

1.5. **Computing.** We will use the R statistical computing environment for all computation in this class.

1.6. **Tentative Outline/Schedule.**

- (1) Introduction
- (2) Foundations of Generalized Linear Models
 - Supplement: Review of Maximum Likelihood Inference
 - Supplement: Maximum Likelihood Inference for Misspecified Models
 - Supplement: The Kolmogorov–Smirnov Test
 - Supplement: L^p Spaces
- (3) Models for Binary Data
 - Supplement: Krippendorff's α
- (4) Models for Count Data
- (5) Quasi-Likelihood Methods
- (6) Models for Dependent Discrete Data
 - Mixed Models
 - Generalized Estimating Equations
 - Copula Models
- (7) Bayesian Generalized Linear Modeling
- (8) Extensions of Generalized Linear Models
 - Basis Expansions and Regularization
 - Local Polynomial Models
 - Generalized Additive Models

2. ASSESSMENT

2.1. **Homework Assignments.** There will be several homework assignments during the semester. Students are encouraged to discuss the homework problems and to work together on the computing. However, each student is expected to write his/her submission independently. Hand in only relevant computer output. Note that submitted code should generally be commented and placed in an appendix.

You will have two weeks to complete each assignment. Late homework will not be accepted without prior approval from the instructor.

2.2. **Capstone Project.** There will be a final project and presentation. Details will be handed out in February.

2.3. **Grading.** The project will be worth 40% of your semester grade, homework 50%, class participation 10%. Each student's percentage grade will be converted to a letter grade according to the following table.

A	[93, 100]	B ⁺	[87, 90]	C ⁺	[77, 80]	D	[60, 70]	F	[0, 60]
A ⁻	[90, 93]	B	[83, 87]	C	[73, 77]				
		B ⁻	[80, 83]	C ⁻	[70, 73]				

3. POLICIES

3.1. Academic Integrity. All University policies regarding academic integrity apply to this course. Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating of information or citations, facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of another person or work previously used without informing the instructor, or tampering with the academic work of other students. All exam answers must be your own, and you must not provide any assistance to other students during exams.

3.2. Disability Services. Penn State welcomes students with disabilities into the University's educational programs. If you have a disability-related need for reasonable academic adjustments in this course, contact the Office for Disability Services (ODS) at 814-863-1807 (V/TTY). For further information regarding ODS, please visit the Office for Disability Services Web site at <http://equity.psu.edu/ods/>. In order to receive consideration for course accommodations, you must contact ODS and provide documentation. If the documentation supports the need for academic adjustments, ODS will provide a letter identifying appropriate academic adjustments.

REFERENCES

- Agresti, A. (2013). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley.
- Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Wiley Series in Probability and Statistics. Wiley.
- Fan, J. (2018). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. CRC Press.
- Faraway, J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Second Edition*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*, volume 2. CRC press Boca Raton, FL.
- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer Science & Business Media.
- Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.
- Robert, C. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Science & Business Media.