**960:553– Categorical Data Analysis– Fall, 2023**

## Projected Schedule and Syllabus

| **Wk** | Day | Date | *Topics* | *Homework* | *Readings* |
|---|---|---|---|---|---|
| **1** | TH | 7 Sep | Response vs. Explanatory, Ordinal vs. Categorical, Introduction, Binomial and Multinomial Distributions, Distributional Assumptions, Inference for a Proportion, Likelihood inference, Poisson Distribution | | A §1.1.1, 1.1.2, 1.1, 1.2.1-1.2.2, 1.2, 1.3, 1.4-1.4.2, 3.5.2, 3.3 |
| **2** | TH | 14 Sep | Proof of $\chi^2$ distribution for Pearson's statistic, Inference on $K > 2$ Counts, One-dimensional ordered exposures | | B&D2 §3.4c–e, 3.4c |
| **3** | TH | 21 Sep | Sampling Models in Two-Way Tables, 2 by 2 table as a difference of proportions, Odds ratio, Confidence Bounds for the Table Effect Sizes, Relative Risk, Equivalence of Odds Ratio for Case–Control Studies, Testing in Multiway Tables, Ordered Categories, Exact Analysis of 2x2 Tables | HW-1 due | A §2.1.5, 2.2.1, 2.3-2.3.2, 2.3.3, 2.3.4, 2.3.5–2.3.6, 2.4, 2.5, 2.6 |
| **4** | TH | 28 Sep | Estimation of the Common Odds Ratio, Controling for additional variables, Marginal and Conditional Independence, Hypotheses of a common odds ratio, The Cochran–Mantel–Haenszel Test, Homogeneity of the Odds Ratio | | A1 §3.2.3, A §2.7–2.7.3, 2.7.4–2.7.5, 2.7.6, 4.3.4, 4.3.5 |
| **5** | TH | 5 Oct | Estimation for the Odds Ratio from Matched Pairs, Modeling Rates, Matched Pairs and McNemar's Test, The log linear model for Poisson regression, Rephrasing current models, Matched groups, perhaps larger than 2 | HW-2 due | A §2.4.3, 4.3, 8-8.2, B&D2 §4.3a, Se §9 pp. 279–280 |
| **6** | TH | 12 Oct | Logistic Regression, Probit Regression, Logistic regression with continuous covariates, Logistic regression for contingency tables, Logistic Regression Approach to Stratified Tables, Rephrasing current models for Three-way Tables, Poisson regression likelihood ratio statistic, Log Linear Models for Two-Way Tables, Proportional Mortality | | A §3.2.3, 3.4.2, 4.1, 4.3–4.3.3, 4.3.4, 7.1.4, 7.2.1, 7-7.1.3, B&D2 §4.7 |
| **7** | TH | 19 Oct | Power Calculations, Paired–preference models, Power in Exponential Family Models | HW-3 due | A §5.5, 8.6 |
| **8** | TH | 26 Oct | Exam | | |
| **9** | TH | 2 Nov | Models and Graphs for Poisson Regression | | A §7-7.1 |
| **10** | TH | 9 Nov | Baseline logit models, Polytomous Regression, Linear-by-Linear Association | HW-4 due | A §6.1, 6.2-6.3, 7.5 |
| **11** | TH | 16 Nov | Marginal Symmetry, Symmetry and Quasi-Symmetry, Quasi-Independence | | A §8.3, 8.4, 8.5.2 |
| **12** | TU | 21 Nov | Exact Methods for Contingency Tables, Exact Methods for Logistic Regression, Agreement and the $\kappa$ statistic, Polychoric Covariance | HW-5 due | A §5.4.2–5.4.3, 5.4.4–5.4.5, 8.5.5, Brown §5.3 |
| **13** | TH | 30 Nov | Latent Class Analysis | | Clogg §6 |
| **14** | TH | 7 Dec | Course Summary | HW-6 due | |
| **15** | TH | 14 Dec | No Class – Reading Day | | |
| **16** | TH | 21 Dec | Exam | | |

**960:553– Categorical Data Analysis– Fall, 2023**

## Course Information

**Instructor:** John E. Kolassa

☎: 445-2692

Office: Hill Center 565

Hours: Wed 9-10pm via zoom, In-person office hours Mon 2-3 pm and Thu 1-2 pm Hill 565.

✉: kolassa@stat.rutgers.edu Please format in plain text or PDF.

**Course Meeting Time:** 6 – 8:50 PM SEC 202 Dates: Thursdays 7 Sep – 7 Dec But not: 23 Nov, and Tuesday 21 Nov.

**Text:**

Required: Available at the university bookstore.

1. Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, New York: John Wiley and Sons.

Recommended: On reserve in the Hill Center Library.

2. Breslow, N.E. and Day, N.E (1980, 1987), *Statistical Methods in Cancer Research*, Oxford: Oxford University Press, **1,2**.
3. Stokes, M.E., Davis, C.S., and Koch, G.G. (1995), *Categorical Data Analysis using the SAS System*, Cary, NC: SAS Institute.
4. Browne, M.W., and Arminger, G. (1995), "Mean- and Covariance-Structure Models," *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (ed. Arminger, G., Clogg, C.C., and Sobel, M.E.), New York: Plenum Press, 185-249.
5. Clogg, C.C. (1995), "Latent Class Models," *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (ed. Arminger, G., Clogg, C.C., and Sobel, M.E.), New York: Plenum Press, 311-360.

**Objectives:** To be able to draw inference from categorical data.

**Requirements:**

Homework: Homework will be assigned approximately biweekly, and will be graded. Solutions will be supplied. Students are encouraged to consult with each other and the instructor when contemplating these problems but are expected to draft their solutions themselves. Students who submit correct homework solutions without fully understanding them can expect substantially lower grades on exams. The lowest homework grade will be dropped. This provision is meant to accommodate personal emergencies, *and constitutes all such indulgence I expect to give in most cases.* Homeworks will constitute 30% of the final grade. Students are responsible for locating computing resources needed for homeworks. I expect (but do not require) that students will perform calculations using *SAS*, which runs on eden.rutgers.edu, a machine routinely acessible by students.

Exams: Exams will be closed–book. Except possibly in the case of a *bona fide* emergency no makeup exams will be given. Absolutely no makeup exam will be given and no alternate source of credit will be given for predictable conflicts with exam times unless such an arrangement is requested within the first two weeks of class.

Midterm: in class, accounting for 30% of the final grade.

FINAL: accounting for 40% of the final grade.
CLASSROOM PARTICIPATION: will not be explicitly graded. Students are responsible for any information and assignments discussed in class.
**Prerequisites:** Introductory statistics 960:484, and either statistical theory (960:583,593) or a regression course (960:563,586).
**Web Page:** `http://www.stat.rutgers.edu/~kolassa/960-553`