Chapter 1: background

- Nominal, ordinal, interval data.
- Distributions: Poisson, binomial, multinomial.
- Basic ideas on MLE theory, CI's and testing.

  You should know basic MLE/CLT theory as applied to the binomial distribution, and be able to construct simple Wald, LRT, and score tests, e.g. homework problems.

  You will need to be able to perform a simple hypothesis test and form a CI from an estimate $\hat{\beta}_j$ and associated standard error $se(\hat{\beta}_j)$. You should memorize $z_{0.025} = 1.96$.

Chapter 2: $I \times J$ tables

- 2×2 tables: OR θ. $X \perp Y \Leftrightarrow \theta = 1$. Differences in proportions, relative risk. Estimability of odds ratio from case/control data – interpretation 'flips.'
- Types of sampling: multinomial, product multinomial, Poisson.
- $2 \times 2 \times K$ tables. Simpson's paradox: marginal association has different direction than conditional association. Homogeneous association $\theta_{XY(1)} = \cdots = \theta_{XY(K)}$, conditional independence $X \perp Y|Z$. Conditional independence does not imply marginal independence. Death penalty example, homework problems.
- On to $I \times J$ tables. Local odds ratios.
- Ordinal trends: γ measure of concordance.

Chapter 3: Estimation, testing in $I \times J$ tables

- Estimation: OR, RR, difference in proportions for 2×2 tables.
- Testing independence: Pearson and LRT. Large sample $\chi^2_{(I-1)(J-1)}$ approximation.
- If reject $H_0 : X \perp Y$ find out why: residuals and partitioning χ^2.
- $I \times J$ tables with ordinal outcomes. Focusing on a measure of association: $\hat{\gamma}$, polychoric correlation $\hat{\rho}$, Pearson correlation based on replacing outcomes with scores $\hat{\rho}$; all in PROC FREQ.
- Exact tests of $H_0 : X \perp Y$ by conditioning on sufficient statistics (marginal totals).

Chapter 4: How does categorical response or counts change with predictors? GLMs

- GLMs: basic notation. Binomial and Poisson regression; identity and canonical links.
- Crab satellite data!
- Deviance G^2, saturated model.
- Negative binomial regression (just mentioned).
- Bit of GLM theory: moments, fitting procedures, residuals.
- Quasi-likelihood adds dispersion ϕ. MOM estimation from large sample theory, overdispersion $\Rightarrow$ inflate MLE SE's via $\hat{\phi}$. SCALE=PEARSON or SCALE=DEVIANCE.

Chapter 5: logistic regression I

- Logistic regression with one predictor.
- Parameter estimates give odds ratios.
- Case/control (retrospective) studies don't change parameter estimates (only intercept estimate).
- More crab analyses.
- GOF for logistic regression. Grouped data versus ungrouped. Hosmer and Lemeshow.
- Categorical predictors.
- Multiple predictors.
- A bit on fitting.

Studying for Exam I

Focus on homeworks 1-6, examples done in the notes, and Chapters 2 (2.1, 2.2, 2.3, 2.4), 3 (3.1, 3.2, 3.3, 3.5), 4 (4.1, 4.2, 4.3, 4.7), and 5 (5.1, 5.2, 5.3, 5.4).

In particular, you should be able to:

- Briefly describe in words what the polychoric correlation $\hat{\rho}$, gamma statistic $\hat{\gamma}$, and Pearson statistic $\hat{\rho}$ (based on scores) measure. For what type of data are these measures valid?
- Show how odds ratio interpretation 'flips' in 2×2 table.

- Be able to interpret a logistic regression model with numerous categorical predictors involving interactions.
- Be able to coherently interpret the residual and partitioning approaches to following up tests of independence in $I \times J$ tables (Section 3.3). Patterns of residual signs?
- Be able to describe in words what the quasi-likelihood approach to modeling overdispersion does for Poisson and binomial data. That is, what is $\text{var}(Y_i)$ modeled as under the actual (real) probability models versus the quasi-likelihood approaches?

- Be able to interpret output for logistic regression models with logistic and identity links. Be able to interpret output for Poisson regression models with log and identity links.
- Have good working knowledge of what the deviance and Pearson GOF tests measure and when you can trust the p-values. Have an idea of what the Hosmer and Lemeshow GOF test measures. What are the null and alternative hypotheses in all of these tests?
- Have an idea of when asymptotic χ^2 tests for $H_0 : X \perp Y$ are valid.