Week Thirteen

Last Week

- Multicategory Data
  - Multicategory Regression
  - Ordinal Regression

This Week: Introduction to Correlated Data

- Tuesday: Veteran's Day
- Thursday:
  - Class Recap
  - Lecture on Correlated Data

Next Week: Multicategory Regression

- Models for Matched Pairs (Ch. 8)
- Modeling Correlated, Clustered Responses (Ch. 9)
- Random Effects: Generalized Linear Mixed Models (Ch. 10)

Class Recap

- Week 1. Course Overview

- Week 2. Inference for Proportions

$Y_i \sim \text{Bernoulli}(\pi)$

$\hat{\pi} = \frac{\sum Y_i}{n}$ $\quad\quad$ $\hat{\pi} \pm Z_{\alpha/2}(SE)$

binom.test( )

- Week 3. Bayesian Inference for Proportions

Stat modeling: $Y_i \sim \text{Bernoulli}(\pi)$ ~~$\sim$~~ $\text{Binomial}(n, \pi)$

Prior: $\pi \sim \text{Beta}(a, b)$

Posterior: $\pi|y \sim \text{Beta}(a + \sum Y_i, b + N - \sum Y_i)$

Stat modeling: $Y_i \sim \text{Multinomial}(N, \vec{\pi})$

Prior: $\text{Dirichlet}(a_1, a_2, \ldots a_J)$ $\quad\quad$ $\vec{\pi} = (\pi_1, \ldots \pi_J)$

Posterior $\vec{\pi}|y \sim \text{Dirichlet}(a_1 + \sum_i (Y_i = 1), \ldots a_J + \sum_{(Y_i = J)})$

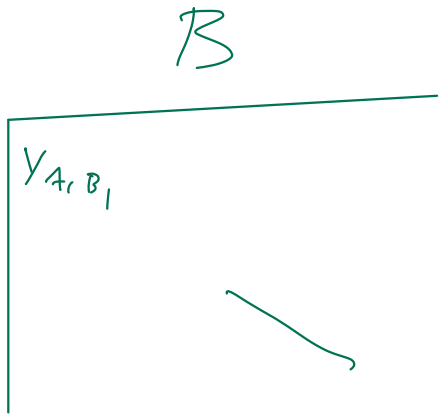
- Week 4 & 5. Contingency Tables

Testing vs. Estimation

joint probabilities

conditional + marginal

χ^2 test



- Week 6 & 7. GLM: Binary Regression

GLMs: 1. random component 2. systematic component 3. link function

$Y_i \sim \text{Bernoulli}(\pi)$

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x + \beta_2 x^2$$

- log odds / odds ratio
- visualize probabilities

- Week 8. Midterm Exam

- Week 9 & 10. GLM: Count Regression

counts in contingency tables

Poisson & Neg. Binomial

$Y \sim \text{Poisson}(\mu)$ or $\text{NegBinom}(\mu, \phi)$

$\mu = \exp(\beta_0 + \beta_1 x)$

- Week 11 & 12. GLM: Multicategory Data

$Y \sim \text{Multinomial}(n, \vec{\pi})$

nominal

$\log\left(\frac{\pi_i}{\pi_J}\right) = \alpha_i + \beta_i x$

ordinal

$\text{logit}(\Pr[Y \le i]) = \alpha_i + \beta x$

$\Pr[Y \le i] = \pi_1 + \cdots + \pi_i$

- Week 13 & 14. Models for Correlated Categorical Data
- Week 15. Fall Break
- Week 16. Machine Learning Tools for Classification
- Week 17. Final Exam

Correlated Data: Matched Pairs

To help the environment

(1) pay higher taxes (2) accept a cut in living standard

cut \ taxes	Y	N
Y		
N		

$H_0: \pi_{ij} = \pi_{ji}$

McNemar's test:

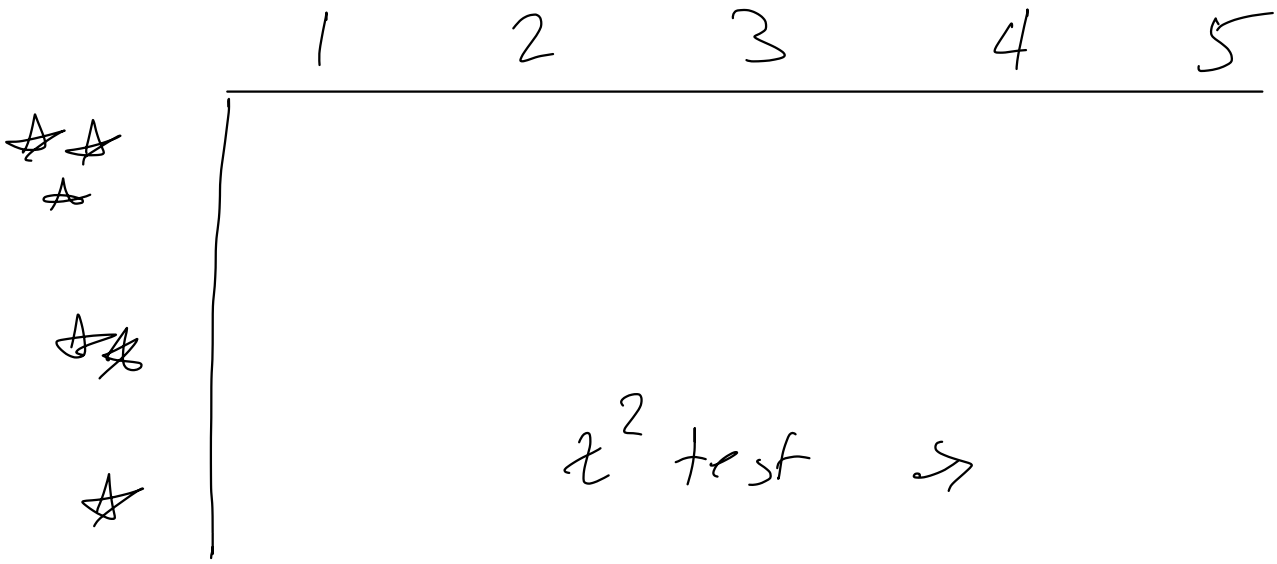
χ^2

$$\left(\pi_{1+} - \pi_{+1}\right) =$$

$$\frac{p_{1+}(1-p_{1+}) + p_{+1}(1-p_{+1}) - 2(p_{11}p_{22} - p_{12}p_{21})}{n}$$

with positive correlation $\rightarrow$ smaller variance

Correlated Data: Rater Agreement



Cohen's Kappa (κ) Measure of agreement
(like a correlation score)

$$\kappa = \frac{\sum \pi_{ii} - \sum \pi_{i+}\pi_{+i}}{1 - \sum \pi_{i+}\pi_{+i}}$$

κ is 0 under independence

κ is 1 perfect agreement

Modeling Correlated Data: Hypothetical Example