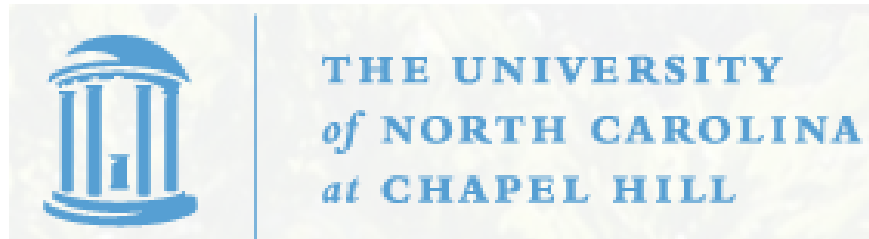


# ***STOR 556: ADV METH DATA ANAL***

***Instructor: Richard L. Smith***

**Class Notes #15:  
February 28, 2019**



## Homework 5

- Chapter 5, Problems 2 and 5 (pages 99–101)
- Due date: Tuesday March 5
- I have posted the TA's solutions to HW1–4. Please treat these as *for personal use only* and do not pass on to anyone outside the class!

## Models for Larger Two-Way Tables

- Graphical summaries: dotchart, mosaic plot
- Classical Pearson  $\chi^2$  test: compute cross-table `ct`, then `summary(ct)`
- Poisson model: fit main effects, no interactions, look at deviance to test fit
- Both our examples so far have indicated strong lack of independence; however, if I simulate data from an independent model, the deviance statistics are not statistically significant.

## Symmetry Models for Matched Pairs

- Symmetry hypothesis:  $p_{ij} = p_{ji}$
- Quasi-symmetry hypothesis:  $p_{ij} = \alpha_i \beta_j \gamma_{ij}$  where  $\gamma_{ij} = \gamma_{ji}$ .
- We can construct an anova test assuming symmetry as the null and quasi-symmetry as the alternative
- Quasi-independence hypothesis: is there independence among people who are asymmetric (not on the diagonal of the contingency table)?

## Simpson's Paradox



Edward H. Simpson (1922-)

## Mantel-Haenszel Test

- Objective: test independence of  $2 \times 2$  tables across  $K$  strata
- Data written  $\{y_{ijk}, i = 1, 2, j = 1, 2, k = 1, \dots, K\}$
- Test is conditional on marginal totals in each table — therefore, it suffices to base the test on the values of  $y_{11k}, k = 1, \dots, K$
- $$T = \frac{(|\sum_k y_{11k} - \sum_k E(y_{11k})| - 1/2)^2}{\sum_k \text{Var}(y_{11k})}$$
- Expectation and variance are computed under null hypothesis of independence in each table, given the marginal totals
- Test statistic  $T$  is approximately  $\chi_1^2$  for large samples; exact p-value calculation is possible for small datasets