Biostat 513 Midterm Exam Review

1. R x C contingency table

    - Test for homogeneity (Pearson chi-squared)
    - Test of independence (Pearson chi-squared)
    - Test for trend

2. Single 2 x 2 table

    - Different sampling schemes
        i Cohort (row totals fixed)
        ii Case-control (column totals fixed)
        iii Cross-sectional (grand total fixed)
    - Different measures of association
        i RD (Designs 1 & 3)
        ii RR (Designs 1 & 3)
        iii OR (Designs 1, 2 & 3)
    - Test of association
        i Pearson chi-squared
        ii McNemar's (paired binary outcomes)
        iii Fisher exact (expected cell sizes are small)
    - Rater agreement
        - Kappa to measure agreement greater than chance
        - Test Ho: $\kappa = 0$ equivalent to Pearson χ^2 test of independence
        - Landis and Koch interpretation of κ

3. Series of 2 x 2 tables

    - Confounding, causality
    - Effect modification (interaction)
    - Mantel-Haenszel (combined) OR estimate
    - Mantel-Haenszel (adjusted) test for association (assume OR constant across strata, Ho: OR= 1)
    - Breslow-Day Test for Homogeneity (Interaction, Effect Modification)

4. Logistic Regression

- use when outcome is binary, independent data
- logistic model
  - $\log[\ \pi(X)\ /(1-\pi(X))\ ] = \text{logit}(\pi(X)) = X\beta$
  - $X\beta = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$
  - bounds $\pi(X)$ between 0 and 1
  - log(p/(1-p)) is the "log odds"
  - $\pi(X) = \exp(X\beta)/(1+\exp(X\beta))$ are “probabilities”
  - saturated model has as many parameters (# of β's) equal to number of "cells" in $X_1 \times X_2 \times \ldots \times X_p$ table; such a model reproduces the observed cell probabilities exactly
  - additive vs multiplicative (interaction) models

- odds ratio
  - $\log[\ \pi(X_1)\ /(1-\pi(X_1))\ ] - \log\ [\pi(X_2\ )\ /(1-\pi(X_2))\ ] =$ $\log[\ \pi(X_1)\ (1-\pi(X_2))\ /\pi(X_2)\ (1-\pi(X_1))\ ] = (X_1 - X_2\ )\beta =$ log odds ratio for covariates X_1 vs X_2
  - for X_i coded 0/1, β_i is the (adjusted) log odds ratio, if no interactions
  - confounding
  - effect modification (interaction)

- estimation/testing
  - maximum likelihood used for estimation
  - likelihood ratio and Wald tests used to test hypotheses
  - LR for nested models only
  - Estimation/testing for linear combinations of parameters

- covariates
  - binary (typically coded 0/1)
  - categorical
    - replace with k-1 indicators (unordered categories)
    - replace with ordinal "score" (e.g. 1,2,3 …) (ordered categories)
  - quantitative
    - linear, quadratic …
- other links
  - log link: $\log(\pi(X)) = X\beta$
    - β interpreted as log relative risk
  - identity link: $\pi(X) = X\beta$
    - β interpreted as risk difference
- Prediction
  - $\pi(X)$ is predicted probability
  - Automated procedures (e.g. stepwise, best subsets) for model fitting
  - AIC for model comparison (esp. non-nested models)
  - evaluate using sensitivity, specificity, ROC curve
  - cutoff, "good" values depend on scientific objective

<u>Key Stata Commands (interpret output)</u>

binreg
cc
cs
estat class
estimates store
kap
lfit, *estat gof*
lincom
logit, logistic
lroc
lsense
mcc
mhodds
predict
stepwise
tab
tabodds
test