Biostat 513: **Exercise Set #2** DUE: April 22, 2002

Reading: Rosner 10.8, 10.9, 10.10 (*optional*)

NOTE: Unless explicitly stated, direct computer output is **not** desired. Typically only part of the computer output is asked for (such as a confidence interval) and then proper interpretation of the statistics is requested.

DATA: The data for these exercises can be found on the class web page:
`http://courses.washington.edu/b513/` in the *Homeworks* directory (ie. click on Homeworks from the main Biostat 513 page).

Contingency Tables

1. The data `tuyns.dat` was collected by Tuyns et al. (1977) in the French department of Ille-et-Vilaine (Brittany). Cases in this study were 200 males diagnosed with oesophageal cancer in one of the regional hospitals between January 1972 and April 1974. Controls were a sample of 778 adult males drawn from electoral lists in each commune, of whom 775 provided sufficient data for analysis. Both types of subject were administered a detailed dietary interview which contained questions about their consumption of tobacco and of various alcoholic beverages. The goal of the study is to characterize the cancer risk associated with both exposures.

(a) Summarize the distribution of AGE, TOB (tobacco), and ALC (alcohol).

(b) Compute a test of homogeneity of the distribution of TOB comparing cases and controls. Compute a test of homogeneity of the distribution of ALC comparing cases and controls. State the null and alternative hypotheses, quote and interpret the test statistic for each factor.

(c) Is there evidence for a TOB "dose" effect? Present point estimates and confidence intervals for odds ratios comparing each category of TOB to the lowest category. Interpret these odds ratios.

(d) Is there evidence for a ALC "dose" effect? Present point estimates and confidence intervals for odds ratios comparing each category of ALC to the lowest category. Interpret these odds ratios.

(e) Calculate a trend test for case/control status and TOB. State the null and alternative hypotheses, interpret the test statistic and p-value.

(f) Calculate a trend test for case/control status and ALC. State the null and alternative

hypotheses, interpret the test statistic and p-value.

(g) Based on these summaries, what do you conclude? What are some possible concerns (ie. any confounders to worry about?)

Stratified Tables

2. Beitler and Landis (1985) present data from a randomized, controlled clinical trial conducted at eight clinics. The purpose of the study was to evaluate the effect of a topical cream in curing nonspecific infections. The binary response variable was classified as favorable or unfavorable response to treatment. The data are in the file `landis.dat` on the class web page.

(a) Summarize the response to treatment in a 2×2 table that ignores clinic. What is the success rate for the drug group and the control group? Calculate and interpret a summary odds ratio. Interpret the χ^2 statistic.

(b) Summarize the response to treatment after stratifying on clinic. Present a table that shows: $\widehat{p}_{0i}$ the success rate for the control group at clinic i; $\widehat{p}_{1i}$ the success rate for the drug group at clinic i; the estimated relative risk for each clinic; and the estimated odds ratio for each clinic. Summarize what this table suggests.

(c) Compute the Mantel-Haenszel test. State the null and alternative hypotheses for this example. Interpret the χ^2 statistic and corresponding p-value.

(d) Compute the test of homogeneity of odds ratios. State the null and alternative hypotheses for this example. Interpret the χ^2 statistic and corresponding p-value.

(e) Compute the Mantel-Haenszel odds ratio estimate and corresponding confidence interval. Interpret this CI and compare these conclusions from those obtained in part (a).

(f) Is clinic a confounder for the association between the response to treatment and the intervention arm (ie. exposure groups are drug and control)? Justify your answer.

(g) Summarize the results of your analysis. Specifically, do you conclude that the drug is effective? Can you state a measure of the effectiveness of the drug? Can you state an estimate of the probability that a favorable response is obtained for a patient that you are advising?

3. The following data were taken from a recent study: "Dietary Iron and Coronary Heart Disease: A Study from Greece" by Tzonou et al. (1998) AJE. Because biochemical data suggest that iron is involved in lipid peroxidation and animal experiments have indicated that iron can promote ischemic myocardial injury, these findings have prompted the study of iron in relation to coronary heart disease (CHD).

The following variables are reported in Tzonou et al.:

```
FEMALE:  0 = male
         1 = female

AGE:     1 = <= 49
         2 = 50-59
         3 = 60-69
         4 = >= 70

IRON:  estimated monthly iron intake (mg)
         1 = <= 250
         2 = 251-300
         3 = 301-350
         4 = 351-400
         5 = > 400

CASE:    0 = control
         1 = case

COUNT:  number of subjects
```

The goal of the study was to assess the relationship between elevated iron consumption and the risk of CHD. A total of 338 cases were obtained and 570 community controls were also recruited.

A dichotomous exposure variable was created:

NEWIRON=1 if IRON $>$ 350 mg/month

NEWIRON=0 if IRON $\leq$ 350 mg/month

Question: Is there evidence that iron consumption (defined using NEWIRON) is associated with CHD? Support your conclusions with appropriate analyses and justify the methods that you use in support of your conclusion.