

ST520, Fall 2008

Homework 1, due: Thursday, 9/4/2008

1. (10 pts) Find an example of an epidemiological study with data collected (from whatever source you can get such as other textbooks or the Internet). Describe the study and state why it is an epidemiological study. Is it a cross-sectional study, prospective study or case-control study, or other study? Are there any exposure variables? What is the disease variable? What are the findings of the study?
2. (10 pts) For the data from a cross-sectional study given in the  $2 \times 2$  table on slide 8, you are given the following facts:

$$n_{1+} = n_{11} + n_{12} \sim \text{binomial}(n_{++}, P[E]), \quad n_{11}|(n_{1+}, n_{1+} > 0) \sim \text{binomial}(n_{1+}, P[D|E]).$$

We are interested in estimating  $P[D|E]$ . Define an estimate as follows:

$$\hat{p} = \begin{cases} \frac{n_{11}}{n_{1+}} & \text{if } n_{1+} > 0 \\ 0 & \text{if } n_{1+} = 0. \end{cases}$$

Show that

$$E(\hat{p}) \longrightarrow P[D|E], \quad \text{as } n_{++} \longrightarrow \infty.$$

**Hint:** Use  $E(\hat{p}) = E[E(\hat{p}|n_{1+})]$  and calculate  $E(\hat{p}|n_{1+})$  for  $n_{1+} = 0$  and  $n_{1+} > 0$ .

3. (10 pts) Suppose in a population there are 30% alcohol drinkers and 70% non-alcohol drinkers. Assume that the chance of developing stomach cancer for alcohol drinkers is 2% and the chance of developing stomach cancer for non-alcohol drinkers is 1% (these numbers are hypothetical and cannot be taken literally).
  - (a) Find the true relative risk and odds-ratio of developing stomach cancer between alcohol drinkers and non-alcohol drinkers. Are they close to each other? Why?
  - (b) Suppose an investigator identified 1,000 alcohol drinkers and 1,000 non-alcohol drinkers who are free of stomach cancer and will follow them until they develop stomach cancer. Find the approximate variance of the log odds-ratio estimate using the data from this study (using expected cell counts).
  - (c) Suppose the investigator used a case-control design with equal number of cases (people with stomach cancer) and controls ( people without stomach cancer). How many of them should

the investigator sample in order to yield the same efficiency as the prospective study (b).  
Again, use expected cell counts.

4. (10 pts) In a study to examine the association between myocardial infarction (MI) and smoking, each of 262 Italian women with MI was matched with about 2 controls (MI-free women). Then their smoking history was identified. The data was summarized in the following table:

Ever Smoker	MI	Controls
Yes	172	173
No	90	346

Answer the following questions:

- (a) Can you estimate the probability that a random Italian woman is a ever smoker?
- (b) Can you estimate the probability of developing MI for an Italian woman?
- (c) Can you estimate the probability of developing MI among smoking Italian women?
- (d) Can you estimate the probability of developing MI among non-smoking Italian women?
- (e) Can you estimate the odds ratio of developing MI between ever smoking Italian women and never smoking Italian women? Construct a 95% CI for the true odds ratio.
- (f) Can you estimate or approximately estimate the relative risk of developing MI between ever smoking Italian women and never smoking Italian women? Under what assumption? Interpret the estimate.