

## Biostat II: Lab 6, Dichotomous variables and Chi-square tests in R

Date: 29 April 2008

1. We have data from a cohort study related to the association between oral contraceptive (OC) use in women and circulatory disease. In the study, 23,000 users of OC were identified and 23,000 nonusers of OC were identified. None of these individuals had circulatory disease at baseline. The women were followed for a fixed period of time and then the presence or absence of circulatory disease was determined for each woman. The study results follow:

		Exposure		Total
		OC use	nonuse	
Disease	Yes	24	5	29
	No	22976	22995	45971
Total		23000	23000	46000

- (a) Calculate and interpret in one sentence the relative risk of developing circulatory disease, comparing OC users to non-users.
  - (b) Calculate and interpret in one sentence the OR of developing circulatory disease, comparing OC users to non-users.
  - (c) Why are the above RR and OR similar?
  - (d) If this had been a case control study, would we have been able to calculate the RR?
  - (e) You calculate the 95% Confidence Interval for the RR and find that it is (1.8, 12.5). What do you conclude about the true relative risk? (*Hint: think about the associated hypothesis test*)
2. One method of potentially increasing response rates in surveys is to contact participants before they are sent the survey and encourage participation. The data below show the response rates for a survey in which some of the participants were contacted before-hand by letter, some were contacted before-hand by phone, and some were not contacted at all. (Adapted from Stafford, 1966, *Journal of Marketing Research*).

Response	Intervention		
	Letter	Phone	None
Yes	171	146	118
No	220	68	455
Total	391	214	573

- (a) What is the response rate for each of the three contact categories? Based on these initial statistics, does it appear that response rate increases with pre-survey contact?
- (b) Reconstruct the table above, but with expected cell counts, assuming that response rate and pre-survey contact are unrelated.
- (c) Compute the chi-square statistic,  $\chi^2$ , for this data.
- (d) If response rate and preliminary contact are unrelated, how likely are you to observe a value as large of larger than what you computed in (c)?
- (e) Based on this probability, is there evidence that pre-survey contact increases response rates?

3. Blood typing is critical for transfusions and organ transplants. Blood type varies worldwide, but is fairly stable within regions. Blood types such as AB or O are defined by the presence of two alleles. The distribution of these alleles in native populations varies widely worldwide.

The following is the Blood Type Distribution in the United States  
(Data from the American Association of Blood Banks):

		B	
		+	-
A	+	AB: 4%	A: 40%
	-	B: 11%	O: 45%

A group of Native Americans were typed during an effort to increase organ donorship in this group. The results from the 300 people typed are below.

- 142 were type A
- 20 were type B
- 9 were type AB
- 129 were type O

Perform a test to see whether or not Native Americans have the same blood type frequencies as seen in the overall US population.