Due by 4pm on Friday, April 11. Please leave your homework on the table before class begins on Friday or leave it in the dropbox outside my office. Do not forget to attach the honor code.

- (1) (2 points each) True/False?
  - (a) In 2 × 2 tables, statistical independence is equivalent to a population odds ratio value of  $\theta = 1.0$ .
  - (b) We found that a 95% confidence interval for the odds ratio relating having a heart attack (yes, no) to drug (placebo, aspirin) is (1.44, 2.33). If we had formed the table with aspirin in the first row (instead of placebo), then the 95% confidence interval would have been (1/2.33, 1/1.44) = (0.43, 0.69).
  - (c) Using a survey of college students, we study the association between opinion about whether it should be legal to (1) use marijuana, (2) drink alcohol if you are 18 years old. We may get a different value for the odds ratio if we treat opinion about marijuana use as the response variable than if we treat alcohol use as the response variable.
  - (d) Interchanging two rows or interchanging two columns in a contingency table has no effect on the value of the  $\chi^2$  or  $G^2$  chi-squared statistics. Thus, these tests treat both the rows and the columns of the contingency table as nominal scale, and if either or both variables are ordinal, the test ignores that information.
  - (e) Suppose that income (high, low) and gender are conditionally independent, given type of job (secretarial, construction, service, professional, etc.). Then, income and gender are also independent in the  $2 \times 2$  marginal table
- (2) (25 points) A baker is trying to choose between two types of cookies to determine which should be featured in an upcoming marketing ad. The plan is to sample public opinion on the two types of cookies by setting up sampling areas for each flavor in a busy area and asking people at each location if they enjoy the cookie. Suppose that the baker also collected information related to the gender, age, etc. of the person sampling the cookie. Data are given below.

Age	Cookie	Liked	Not Liked
<30	Raspberry Rally	760	140
	Toast-Yay!	600	100
30-49	Raspberry Rally	40	60
	Toast-Yay!	150	150

- (a) Calculate the conditional odds ratios for the two flavors, conditioned on age.
- (b) Calculate the marginal odds ratio for the two flavors.
- (c) Are the public opinion and cookie flavor conditionally independent given age?
- (d) Are the public opinion and cookie flavor marginally dependent? Use a  $\chi^2$  test with  $\alpha = 0.05$ .
- (e) Do we have evidence of Simpson's paradox? Why or why not?
- (3) (15 points) The table below summarizes three studies in China about smoking and lung cancer.

		Lung Cancer		
City	Smoking	Yes	No	Total
Beijing	Yes	126	100	226
	No	35	61	96
	Total	161	161	322
Shanghai	Yes	908	688	1596
	No	497	807	1304
	Total	1405	1495	2900
Shenyang	Yes	913	747	1660
	No	336	598	934
	Total	1249	1345	2594

- (a) Test the null hypothesis of homogeneity of the odds ratio across the different cities. Be sure to state your hypotheses and interpret your result in the context of the problem. Use  $\alpha = 0.01$ .
- (b) Assuming homogeneous association is a valid assumption, calculate an estimate of the common odds ratio.

(4) (20 points) The following table refers to ratings of agricultural extension agents in North Carolina. In each of five districts, agents were classified by their race and by whether they qualified for a merit pay increase. Conduct the Cochran-Mantel-Haenszel test of the hypothesis that the merit pay decision is independent of race, conditional on the district. Interpret.

District	Blacks, Merit Pay		Whites, Merit Pay	
	Yes	No	Yes	No
NC	24	9	47	12
NE	10	3	45	8
NW	5	4	57	9
SE	16	7	54	10
SW	7	4	59	12

Source: J. Gastwirth, Statistical Reasoning in Law and Public Policy, Vol. 1, 1988, p. 268.

(5) (20 points) A recent General Social Survey asked subjects whether they believed in heaven and whether they believed in hell. The following table shows the results.

Believe in	Believe	in Hell
Heaven	Yes	No
Yes	833	125
No	2	160

- (a) Test the hypothesis that the population proportions answering "yes" were identical for heaven and hell. Use a two-sided alternative.
- (b) Find a 90% confidence interval for the difference between the population proportions. Interpret.
- (6) (10 points) Hypertension is defined as having: Systolic BP  $\geq$  160 mm Hg, Diastolic BP  $\geq$  95 mm Hg

The hypertensive status of 20 patients was evaluated by an automated device and a trained observer. Note that this data is paired, since each person is having their blood pressure measured twice, once by the machine and once by a person.

observer					
Person	Hypertensive status			Hypertensive status	
	Computer device	Trained observer	Person	Computer device	Trained
1	-	_	11	+	-
2	-	-	12	+	_
3	+	-	13	_	_
4	+	+	14	+	_
5	-	-	15	-	+
6	+	-	16	+	T
7	-	-	17	+	
8	+	+	18	-	_
9	+	+	19	_	-
10	-	-	20	_	_

Hypertensive status of 20 patients as judged by a computer device and a trained

Test the claim that the human and automated device differ in their ability to detect hypertension.