*YOUR TA’S NAME*:

*Lecture Worksheet*

*Thursday 10/29/2020*

**MAIN POINTS OF LECTURE**

1. Even if there is no association between two discrete variables X and Y in the population, we may observe an association between X and Y in sample data because of random error or sampling variability
2. How can we tell whether the association observed between discrete variables X and Y in sample data is strong enough to rule out the hypothesis that in the population X and Y are statistically independent (or not associated with one another)?
	1. *We begin with the null hypothesis that there is no association between X and Y in the population* … that is, we assume “statistical independence”
	2. We then compute the cell frequencies that we would expect to observe under the null hypothesis and compare them to the actually observed cell frequencies
	3. The 2 test statistic quantifies the degree to which the observed frequencies differ from the frequencies that we would expect to observe under the null hypothesis

$$χ^{2}=Sum of \frac{\left(Expected-Observed\right)^{2}}{Expected}=\sum\_{i=1}^{R}\sum\_{j=1}^{C}\frac{\left(\hat{f}\_{ij}-f\_{ij}\right)^{2}}{\hat{f}\_{ij}}$$

where *R* is the number of rows in the crosstable; *C* is thenumber of columns in the crosstable; “observed” or *f*ij is the number of cases observed in the cell in row *i* and column *j*; and “expected” or $\hat{f}\_{ij}$ is the number of cases we’d expect to see in that cell if the null hypothesis is true.

1. Here, the “expected” cell counts or $\hat{f}\_{ij}$ equal:

$$\hat{f}\_{ij}=\frac{\left(f\_{i•}\right)\left(f\_{•j}\right)}{N}$$

where *fi•* is the number of cases in row *i*, *f•j* is the number of cases in column *j*, and *N* is the sample size

1. For hypothesis tests involving 2, degrees of freedom or *df*=(*R*—1)(*C*—1) where *R*=number of rows in the crosstable and *C*=number of columns in the crosstable
2. When X and Y are ordinal, gamma describes the association between them
	1. Gamma too complex to calculate by hand (in this class)
	2. Gamma = -1 = a perfect negative association (as X goes up, Y goes down)
	3. Gamma = +1 = a perfect positive association (as X goes up, Y goes up)
	4. Gamma = 0 = no association (as X goes up, nothing happens to Y)
3. When X and Y discrete and have only two categories, gamma, relative risk, and odds ratios describe the relationship between them. They each provide different sorts of information
	1. Each requires referring to the 4 cells of the 2-by-2 table as follows:

X=1 X=2

Y=2 a b

Y=1 c d

* 1. Relative Risk
		1. Express the change in the risk of being in a particular category of the “dependent” (Y) variable that result from changing categories of the “independent” (X) variable
		2. 
		3. RR <1 means that the risk that Y=2 is reduced when you move from X=1 to X=2
		4. RR = 1 means that the risk that Y=2 is unchanged when you move from X=1 to X=2
		5. RR >1 means that the risk that Y=2 is increased when you move from X=1 to X=2
	2. Odds Ratio
		1. Express the change in the odds of being in a particular category of the “dependent” (Y) variable that result from changing categories of the “independent” (X) variable
		2. 
		3. OR <1 means that the odds that Y=2 is reduced when you move from X=1 to X=2
		4. OR = 1 means that the odds that Y=2 is unchanged when you move from X=1 to X=2
		5. OR > 1 means that the odds that Y=2 is increased when you move from X=1 to X=2
1. Associations between X and Y that are statistically significant may not necessarily be large or important; always necessary to assess the strength and direction of the association

**QUESTIONS**

1. [From the recorded lecture] What is the critical value of c2 when a=0.05 and…

…the table has 2 rows and 3 columns?

…the table has 5 rows and 3 columns?

1. [From the recorded lecture] What is the critical value of c2 when a=0.01 and…

…the table has 3 rows and 4 columns?

…the table has 4 rows and 4 columns?

1. [From the recorded lecture] Are political views related to whether people view the bible as the literal word of God?



1. [From the recorded lecture] Is there are a relationship between X and Y in the population from which the sample data were collected? Use 2 to test the hypothesis of no association (w/ =0.05). Use RR and OR to describe the association.
2. [From the synchronous session] Is there an association between people’s favorite pet and whether they are right or left handed?