DS 100: Principles and Techniques of Data Science **Date: April 20, 2018**

Discussion #11

Name:

Hypothesis Testing

1. In lecture, we saw an example of permutation inference in Boring, Ottoboni, and Stark's (2016) reexamination of Student Evaluation of Teaching (SET) data. In the experiment, 47 students were randomly assigned to one of four sections. In two of the sections, the teaching assistants were introduced using their actual names. In the other two sections, the assistants switched names. Students never met the teaching assistants face-to-face. Instead, they interacted with the students mainly via an online forum. Homework returns were coordinated so that all students received feedback all at the same time. The authors wanted to investigate if gender perception has any effect on SETs.

   (a) What is the model?

   (b) What is the null hypothesis?

   (c) What is the test statistic?

   (d) How did the authors use permutation to compute the sampling distribution under the null hypothesis?

   (e) How is this permutation justified?

   (f) The TAs objectively returned the assignments at the same time. Therefore, assigned TA should also have no effect on promptness. Why, then, do the authors not permute across TA assignments?

2. Suppose we roll a die 10000 times. The first 5000 rolls are done while wearing a fedora, and the latter 5000 rolls are done while wearing a 10-gallon hat. The type of hat does not affect the die. However, we would expect that the null hypothesis—exactly as many even numbers will be rolled while wearing a fedora as while wearing the 10-gallon hat—would not likely be true. Stated differently, random fluctuations in the proportion of even rolls while wearing the different hats imply that the null hypothesis will nearly always be false.

   Comment on the validity of the claim above.

3. The Graduate Division at UC Berkeley compares admission rates for men and women. For one year and one graduate program, this is summarized in the following table:

	Admitted	Rejected	
Men	509	316	825
Women	89	19	108
	598	335	933

   Either argue against the sensibility of the following question or describe a simulation or resampling method that can be used to address it: Is the difference between admission rates for men and women statistically significant?

4. Saccharin is used as an artificial low-calorie sweetener in diet soft drinks. There is some concern that it may cause cancer. Investigators performed an experiment on rats. In the treatment group, the animals got 2% of their daily food intake in the form of saccharin. The treatment group had a higher rate of bladder cancer than the control group, and the difference was highly significant (p=0.01). The investigators concluded that saccharin probably causes cancer in humans. Is this a good way to interpret the p-value? If not, then what, if anything, can the p-value say about this experiment?