**Week 8: Non-parametric Hypothesis Testing**

1. If you sample with replacement, are you likely to have more, fewer, or less of the observations in the original sample?

2. For the following statements, describe whether it’s true or false and provide a one sentence explanation why:

1. Using the CLT and null distribution, you construct a 90% CI between [1, 5]. The middle point of this confidence interval is your point estimate for the sample mean
2. Using the bootstrap, you construct a 90% CI between [1, 5]. The middle point of this confidence interval is your point estimate for the sample mean
3. The bootstrap requires the mean of your random sample to not be too extreme
4. NHST requires the mean of your random sample to not be too extreme
5. You have to assume a null distribution to conduct a permutation test
6. A parametric bootstrap does not require a full sample to use
7. Both the bootstrap and NHST can estimate the median
8. A permutation test will always have the same or fewer false positives as two-sample NHST
9. Samples for a bootstrap must be drawn from a normal distribution
10. It is an example of systematic error if your sample is biased.

3. You are a political consultant, trying to figure how long (measured in days) it takes for a news story about your candidate to stop receiving coverage on television. You believe this follows an exponential distribution, since you are measuring the amount of time (continuous) it takes for something to be out of the news. The only parameter in an exponential distribution is the rate of change, approximated by 1/mean(X). In your sample of 20 news stories, there was a rate of change of 0.5, in other words 1 / (2 days) on average that it took a news story to stop airing. Construct a 90% confidence interval for the rate of change using a parametric bootstrap. You do not have to show your excel sheet, but provide a step by step accounting for what you are doing to receive partial credit.

4. You are measuring the approval rating of your candidate after an ad airs. Before the ad, your candidate’s approval rating was 52%. You conduct a new focus group of 30 voters. Using this focus group, you resample 20 times, resulting in the following table of means for each resample:

54.3 53.7 51.9 51.6 59.2 60.2 53.4 55.5 54.9 54.5

55.7 73.6 57.5 65.0 50.2 53.1 45.1 56.0 47.4 52.1

Calculate a 90% confidence interval for the shift in the population before and after you ran your ad.

5. What is the main assumption you had to make about the focus group in problem 4?

6. Compare the pros and cons of the bootstrap vs NHST. Your answer should not exceed a short paragraph (you can use bullet points or a list too).

7. When would you use a parametric bootstrap instead of a regular bootstrap?

8. You are comparing the differences in two means. Using a bootstrap you get a 95% confidence interval of [1, 6]. A NHST provides a confidence interval of [2, 5]. Interpret both of these results. Use one sentence for each.

9. Why would you use an empirical null distribution instead of the CLT based null distribution?

10. If a permutation test involves only 100 observations, what is the smallest p-value you could find?