Week 8 Reading Guide: Hypothesis Tests & Model Conditions

## Section 1 – Baby Birth Weights

**What is the process of obtaining one shuffled sample?**

**With a shuffled resample, once the values are written on the cards what happens? Why is this a necessary step?**

**What process is used to shuffle the cards? Are the cards put back into the hat once they are drawn?**

**How similar was the shuffled slope statistic (in Figure 2) to the original slope statistic found in the original data?**

**Where was the distribution of 10 shuffled slope statistics (in Figure 3) centered? Why was it centered there?**

**What does the word “permute” mean? How is this term related to the shuffling method that was used to obtain the slope statistic?**

## Section 2 – Hypothesis tests

**What general form does a null hypothesis take? i.e. what does the null hypothesis generally assume?**

**What is the difference between a one-sided and a two-sided alternative hypothesis? How do you decide which to use?**

**What is a null distribution?**

**How is a null distribution created?**

**How is a p-value related to a null distribution? i.e. how is a p-value calculated?**

## Section 3 – Conducting a hypothesis test

**When creating a null distribution, what are the sequence of functions you use?**

**Why is the hypothesize() function used to make a null distribution but not for a bootstrap distribution?**

**What does the null = "independence" input in hypothesize() mean? What is it assuming about the variables declared in the specify() step?**

**What method is used to generate() new samples? i.e. what is input as the type?**

**Why is the null distribution centered at 0?**

**How was the p-value calculated using the null distribution?**

**How was the significance threshold (**) used to reach a hypothesis testing decision?

## Section 4 – Interpreting a hypothesis test

**Why do we never accept the null hypothesis?**

**What is a Type I error?**

**What is a Type II error?**

**If you choose a smaller value of** what happens to your Type I error rate?

**If you choose a smaller value of** what happens to your Type II error rate?

## Section 5 – Comparing with confidence intervals

**Why do we look for 0 in our confidence interval?**

**If our confidence interval contains 0 what decision would we have made in our hypothesis test?**

**What is the difference between statistical and practical significance?**

## Section 6 – Theory-based hypothesis tests & confidence intervals

**How is the std\_error column in the regression table calculated?**

**How is the statistic column in the regression table calculated? What type of statistic is it?**

**How is the p\_value column in the regression table calculated?**

**How are the lower\_ci and upper\_ci columns in the regression table calculated?**

**What is the relationship between theory-based methods and simulation-based methods?**

**How is the std\_error column in the regression table calculated?**

## Section 7 – Checking model conditions

**What are the four conditions that need to be evaluated for linear regression?**

**What is a residual?**

**What does it mean for the residuals to be independent?**

**What distribution must the residuals look similar to?**

**What does it mean for there to be “equal variance” in the residuals?**

**What happens if one of these conditions is violated?**