**EPSY 5261 Discussion Assignment #6: More Hypothesis Tests**

Carefully watch the Week 7 and Week 8 Lectures and Media before you begin this assignment. We will revisit students’ scores on the PISA exam (from Discussion Assignment #4). This time, we’ll use StatKey to evaluate specific hypotheses about each context.

**PISA Scores**

**Background**: In Discussion Assignment #4, we examined data representing the PISA score for 100 randomly selected students from Finland to another 100 randomly selected students from Spain. We calculated a 95% Confidence Interval for the difference in mean PISA scores for all students in Finland compared to Spain as between 6.6pts to 20.4pts higher for Finland. For the Finnish students, the sample mean was 81.35pts with a sample standard deviation of 23.08pts. For the Spanish students, the sample mean was 67.83pts with a sample standard deviation of 26.35pts.

**Assignment**: Professor Ramsey suggests that your analysis from Discussion #4 confirms that the Finnish school system is better than the Spanish school system. [Professor Blackwell](https://mathshistory.st-andrews.ac.uk/Biographies/Blackwell/) suggests that because the study was observational, we should not infer a causal claim such as “The Finnish school system *caused* Finnish students’ scores to be higher”. Professor Blackwell suggests that they might be able to conduct an experiment to test the hypothesis that there really is no difference between the school systems. Professor Blackwell notes that such an experiment would need to use random allocation to either Finland or Spain, which can be done if they partner with the US Foreign Service. 52 Foreign Service Officers (FSOs) and their families volunteer to participate in the study. The total sample size is 80 children. Starting in 2010, 22 FSOs were randomly assigned to the US Embassy in Finland, and their 36 children attended Finnish schools for the next 6 years. The remaining 30 FSOs were randomly assigned to the US Embassy in Spain, and their 44 children attended Spanish schools for the next 6 years. In 2016, all 80 children took the PISA test. Their scores are in the data file is called **Pisa80.csv**. There are 4 variables in the dataset:

* **Parent**: De-identified parent ID
* **Student:** De-identified student ID
* **Country:** Finland, or Spain
* **Score:** Score on the PISA test, out of 219 points.

**Group Discussion Questions**

1. Compare the distribution of PISA scores in the sample from each country. Include a side-by-side dotplot, a side-by-side boxplot, and a two-row summary table.
2. Draw a design model representing each step of Professor Blackwell’s experiment.
   1. Identify the random source of variation in the experiment.
   2. Identify the difference in exposure that is a source of variation in the experiment.
3. Identify the null hypothesis that Professor Blackwell wants to test first.
   1. To what extent does Professor Blackwell’s null hypothesis assume that each source of variation (random and experimental) contributes to the variation in students’ PISA scores?
   2. Draw a hypothetical dotplot for students’ PISA scores in each experimental group based on this experimental null hypothesis. Label the hypothesized mean and the hypothesized range of typical PISA scores for an individual student in each group (i.e., the middle 95%).
4. Explain how you would use re-randomization resampling via the paper pieces and bag method to generate one re-randomization sample statistic for the difference in the means between the two groups (i.e., ).
5. Use StatKey to evaluate the experimental null hypothesis.
   1. Describe what each dot in the randomization dotplot represents.
   2. Verify that the center of the randomization dotplot is equal to experimental null hypothesis’s claim. Highlight this point on the randomization dotplot. Also highlight the sample mean difference observed in experiment on the randomization dotplot.
   3. If experimental null hypothesis is true, is it *possible* to see a sample mean difference like the one we observed in real life from Professor Blackwell’s experiment of 80 students?
   4. Obtain a *p*-value. Use it to evaluate experimental null hypothesis. Is there statistical evidence that suggests the null hypothesis may be incorrect? How strong is the evidence?