

Homework 6

1. Download this data: bit.ly/illschooldata.csv (Data set is from here: <https://ww2.amstat.org/censusatschool/RandomSampleExport.cfm>)
 - Describe the data. Who is in this data set? What are some of the interesting characteristics of this data set?
 - Perform the appropriate test to test the null hypothesis that handedness (i.e. the variable named Handed) is independent of favorite season vs the alternative hypothesis that there is some dependence. Perform this test after removing responses that are blank. Do you think it is ok here to remove the blanks? Explain why or why not. Explain your reasoning for the test you chose and state your conclusions.
 - Build a simple linear regression model with height as your response and arm span as your predictor. First, you need to clean the data, then use MICE to impute missing values using a CART model. Estimate the simple linear regression model on each of the completed data sets and use Rubin's combining rules to combined estimates across imputations. State your final estimates for each of the slope and intercept parameters as well as standard errors for each of these combined estimates.
 - Repeat the previous problem, but use a random forest for imputation in MICE instead of a cart model.
 - Finally, put your code and results in a github repository. In the final version of your homework that you submit to Sakai, the answer to this part will simply be a link to that github repository.