Summary of Results[1]

For this project, we analyzed a dataset of results of the COMPAS risk-assessment algorithm in Broward County, Florida. The algorithm is used during pre-trial and sentencing to help judges predict a defendant's risk of recidivism, violence, and failure to appear. Our dataset was collected by ProPublica[2] by merging the results of a public records request from the Broward County Sheriff's Office with criminal records to create a profile for each defendant including their COMPAS score. The columns that we focused on were the Ethnic_Code_Text column (race), the Raw_Score column (numerical compas scores with high scores indicating higher risk), and the Display_Text column (the display text shown to the judges: Risk of Recidivism, Risk of Violence, and Risk of Failure to Appear). The first piece of analysis we did was to create a histogram of the Raw Scores. We immediately noticed two clusters, which we isolated using R's built-in kmeans function. We found that the cluster of lower scores fit a beta distribution and the cluster of higher scores fit a gamma distribution. We then divided our dataset by the display text and analyzed each subset separately. Using a permutation test, a t test, and a z test, we found that within the display texts of Risk of Recidivism and Risk of Violence, there was a statistically significant difference across race. Namely, non-white people were more likely to get these display texts. Interestingly, there was no significant difference between races within the Risk of Failure to Appear display text. For these tests, classical methods (z and t tests) actually worked better than simulation methods (permutation test) because the p-value was so small that simulation methods gave us a p-value of exactly 0, whereas the classical methods

---

[1] Dataset downloaded here: https://www.kaggle.com/danofer/compass
[2] Larson, Jeff, et al. "How We Analyzed the COMPAS Recidivism Algorithm." *ProPublica*, 13 May 2016.

offered more precision. The last piece of analysis we did was to determine if display text was

race were independent. We found, using a Chi-Square test, that they were perfectly independent,

with a p value of exactly 1. This result was incredibly surprising, both because the results of our

permutation test above seems to indicate that the two are not independent, and because perfect

independence is incredibly rare. Ethical implications of this statistic are discussed in the

Ethics-Results paper.