Austin Strahan

Professor Moore

Probability and Statistics

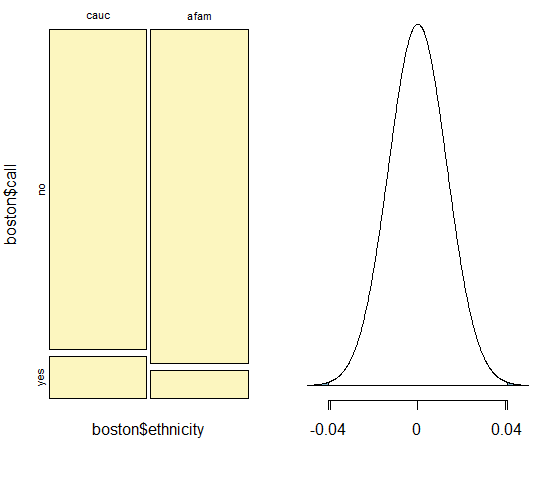
20 November 2020

Project Outline

**Hypothesis Test**

1. The Set Up:
   1. African Americans and Caucasians, but any ethnic group of people could demonstrate interest.
   2. There is no difference between the proportion of people that received a call back based on if their name sounded African American or Caucasian.

There is a difference between the proportion of people that received a call back based on if their name sounded African American or Caucasian.

1. The Hypothesis Test Criteria:
   1. Criteria
      1. Independence: Randomized Controlled Experiment and 4870 < 10% of the total population of African Americans and Caucasians.
      2. 10 Successes/Failures: As shown in the work and highlighted in the report, each population has at least 10 successes and 10 failures.
   2. Most computation done in R. , p-value is calculated using a table.
   3.  The level of confidence is 95% and the significance level is 0.05
2. The Sample Evidence:
   1. Sample information all collected in R for the two populations Boston and Chicago. Information is also in R for whether an applicant got a call back based on if there resume name sounded Caucasian or African American.
   2. Boston: Z = 3.05

Chicago: Z = 2.763

1. The Hypothesis Test:
   1. Boston: p = 0.0022

Chicago: p = 0.0058

* 1. Boston: p-value < 0.05

Chicago: p-value < 0.05

1. The Results:
   1. Decision: Reject the null hypothesis for both Boston and Chicago.
   2. Conclusion: There is a statistically significant difference between the number of callbacks received based on if the applicant’s name sounded Caucasian or African American in both Chicago and Boston.

Resume Ethnicity Report

The research questions this reports aims to cover is whether having a Caucasian or African American sounding first name on a resume affects the chances of getting a call back from an employer. The population of interest would primarily be Caucasians and African Americans, but the results of the study may interest other ethnicities as well. The null hypothesis for the study is:

To clarify, the null hypothesis states that there is no difference between the proportion of people that received a call back based on if their name sounded African American or Caucasian. In more general terms, there would be no differences in the percentages of Caucasians that receive a call back versus the percentage of African Americans that receive a call back. Conversely, the alternative hypothesis for the study is:

To explain in further detail, the alternative hypothesis states that there is a difference between the proportion of people that received a call back based on if their name sounded African American or Caucasian. To elucidate, the percentage of Caucasians that received a call back would be more or less than the percentage of African Americans that received a call back. To perform a hypothesis test, a few conditions must be met.

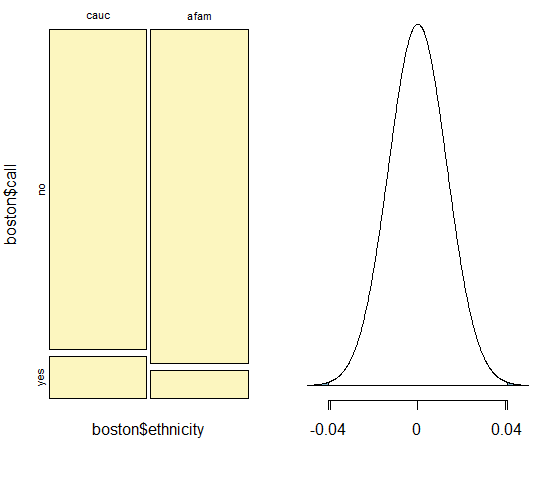
The first criteria for a hypothesis test are to have data that is independent. The data for this study was produced using a randomized controlled experiment. Furthermore, the sample size was 4,870 imaginary resumes, which is less than 10% of the total population of Caucasians and African Americans. This evidence proves that the data was conducted to be independent. Another equally important criteria are having at least 10 successes and 10 failures in the data. For the city of Chicago, the data yielded 109 successes for Caucasians, 73 successes for African Americans, 1243 failures for Caucasians, and 1279 failures for African Americans. Nevertheless, for the city of Boston, the data yielded 126 successes for Caucasians, 84 successes for African Americans, 957 failures for Caucasians, and 999 failures for African Americans. All numbers for successes and failures are greater than 10, thus the condition is met. Therefore, the conditions required for a hypothesis test are fulfilled.

There were a variety of formulas used to calculate the statistics for the hypothesis test. The first equation produces standard error:

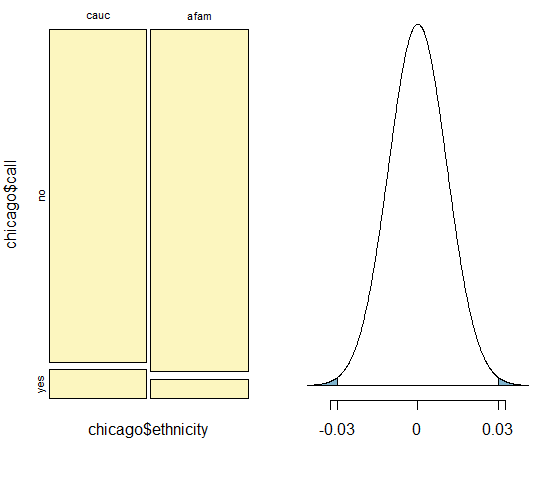
where P1 is the proportion for Caucasians, P2 is the proportion for African Americans, n1 is the sample size for Caucasians, and n2 is the sample size for African Americans. Following, the next equation calculates the z-score:

Where is the proportion of Caucasians, is the proportion of African Americans, and SE is the standard error calculated in the previous equation. The p-value is calculated using a z-score table and the z-score received from the data. Finally, the level of confidence is 95% and the significance level is:

The population distribution for Boston looks like:



In addition to the population distribution for Boston, the distribution for Chicago looks like:



All the data for this study was found in an AER package in the statistical software R. The two populations in this experiment are Boston and Chicago. All variables in this report are categorical, ordinal with the explanatory variable being the ethnicity and the response variable will be whether the applicant gets a call back. Most of the calculations for this study were conducted using R to help prevent human error. As for the results of the study, the z-score obtained from the Boston data is 3.05 and the z-score obtained from the Chicago data is 2.763. The next step will be to convert these z-scores into p-values to determine if they produce statistically significant results.

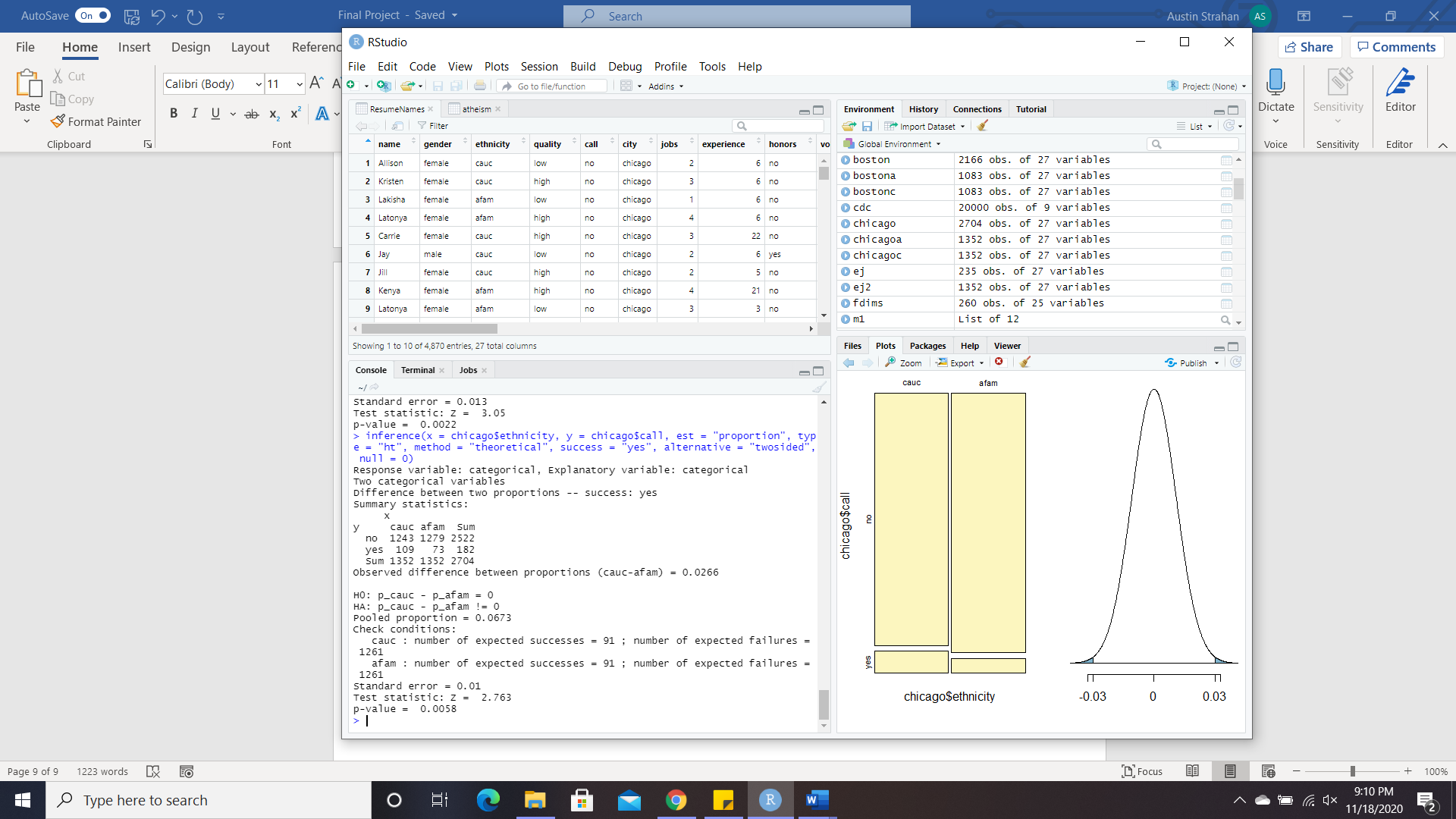
As a result of consulting the z-score table, the p-value found for Boston was 0.0022, while the p-value for Chicago was 0.0058. For a significance level of 0.05, to reject the null hypothesis, the p-value must be less than the significance level. Consequently, both the p-value for Boston and Chicago are less than 0.05. This result means the data for both cities is statistically significant.

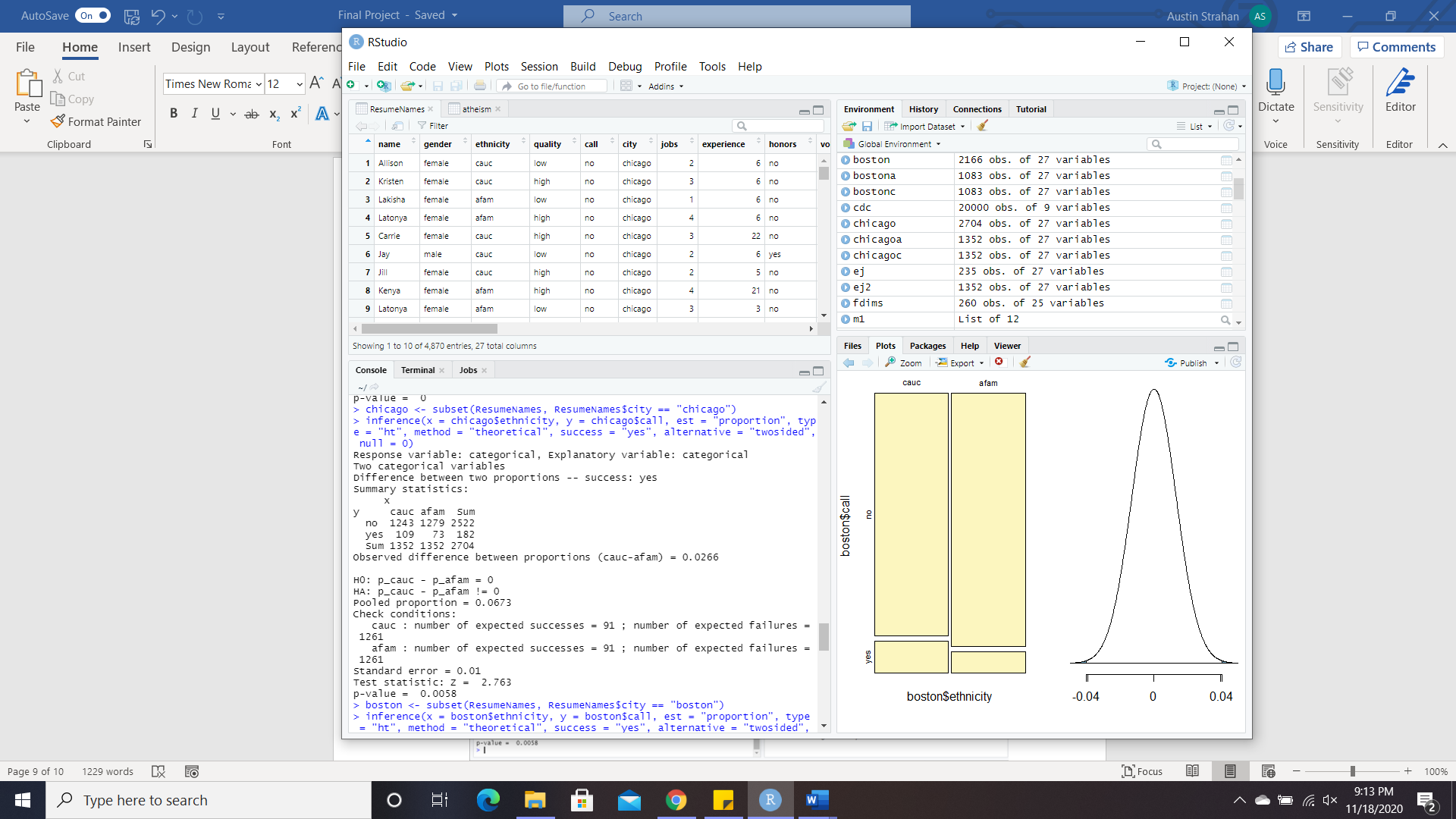
In conclusion, the null hypothesis that there is no difference between the probability of getting a call back depending on if the resume name sounds Caucasian or African American is rejected for the alternative hypothesis. Furthermore, the data provides sufficient evidence to prove that there is a significant difference between the probability of receiving a call back from an employer based off different ethnically sounding names. Considering this finding, new questions regarding the subject of ethnicity and the impact it has on employment abound.

In light of this conclusion, more data should be compiled on different ethical names on job resumes in different cities and countries. While the data provided evidence for two cities in America, the result can only loosely be concluded for the overall population. A possible way to improve this experiment would be to include more locations and more ethnicities. Furthermore, future research should be done regarding college applications to examine if ethnicity plays a role in deciding acceptance. Another way to improve this study would be to consider more variables than just ethnicity to investigate getting a call back from an employer. To this end, the conclusion that there is a cause and effect relationship between ethnically sounding names and job callbacks should just be the push that gets the ball rolling in terms of examining the role racial prejudice can play in American lives. While not proving that racial bias exists, the data found in this study provides convincing evidence that racial bias is a potential issue that needs to be addressed in the employment world.

Calculations for the Study

Calculations for Chicago:





Calculations for Boston:

