Capstone Project 1: Statistical Data Analysis
Predicting clinical genetic variants that will have conflicting classifications by clinicians

Gurdeep Sullan
1/27/19

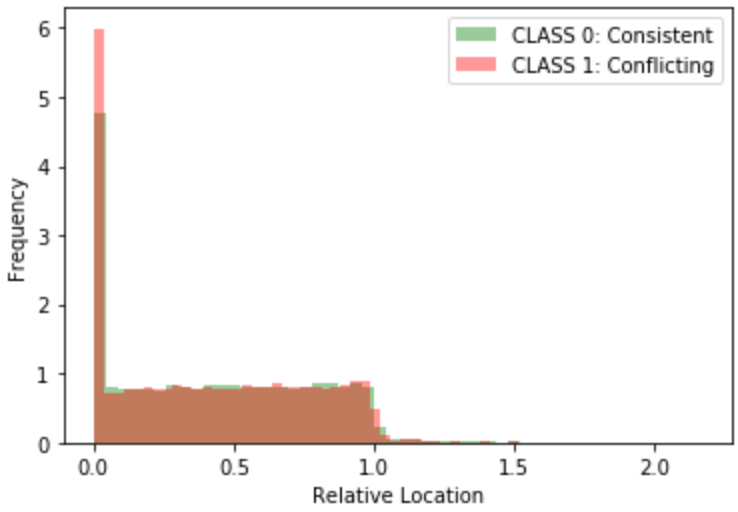
Background:

I will use statistical tests from the scipy.stats package in Python to answer two questions about the classification in relation to different variables. One variable is continuous, so I will use the t-test. The second question deals with a categorical variable, so I will use a chi squared test.

Questions:

1. Is the relative location of the mutation in the length of the entire protein related to the classification?
   Approach: Perform a two tailed t-test using scipy.stats t package.
   - Null Hypothesis: There is no difference in mean relative location in protein between class 0 and class 1
   - Alternate Hypothesis: There is a difference in mean relative location in protein between the two classes
   - Alpha = 0.05

   Results: The p-value is 0.057, greater than the alpha = 0.05. I cannot reject the null hypothesis and conclude that there is no significant difference in mean relative location between the two groups.



2. Are Indels or SNPs (or any other type of variant) more prone to conflicting classification compared to other types of variants? Is there a
   Approach: Perform a chi squared test on a contingency table that contains the types of variants as column names and classification as row labels. I will use the scipy.stats chi2_contingency function and chi2 package
   - Null Hypothesis: There is no difference in classification between different types of variants (i.e. indels/SNPs)
   - Alternate Hypothesis: There is a difference in classification between different groups of variants
   - Alpha = 0.05

   Results: The chi2 test shows that the classification and variant type are not independent of each other. The statistic is greater than the critical value, and the p-value is less than alpha = 0.05. Therefore I reject the null hypothesis that there is no difference in classification across different types of variants.