Pre-Analysis Steps

Before analyzing our dataset, which consisted of 9 columns and 64 entries, we performed several preprocessing steps to ensure clarity and usability. We cleaned the data by removing unnecessary whitespace and non-applicable values. To standardize the "academic year" column, we converted entries such as "First year" to "1" and "Fourth year" to "4." We excluded the timestamp and major columns, as they were not relevant to our analysis. Additionally, we simplified binary yes/no responses in the "Are you an undergraduate?" and "Do you think a hotdog is a sandwich?" columns by encoding them as 1 for "Yes" and 0 for "No." These steps helped streamline the dataset for effective analysis.

Analysis Methods

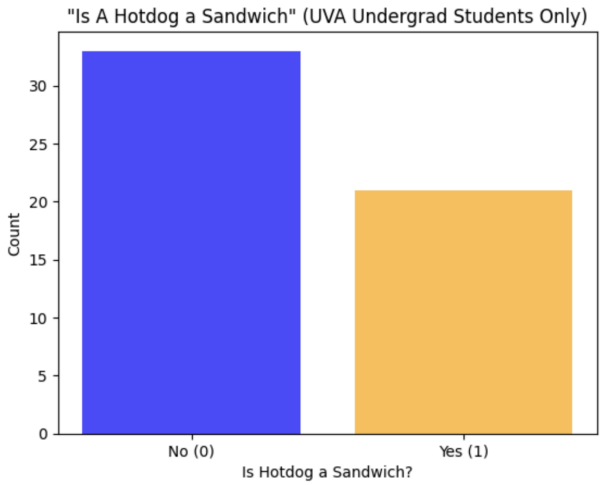
We mainly used Python notebooks along with RStudio to analyze the data. After cleaning the dataset, we first proceeded to find the number of people who agreed that a hot dog was a sandwich vs. did not agree. In order to do this, we had to clean this column by transforming it into a true/false vector, with any answer beginning with “Y” or “y” being TRUE, and any answer otherwise being FALSE. After finding this (and creating a bar chart in the process), we then went on to create bar charts that separated the data by year (which we had to further filter). To test our hypothesis, we will conduct a one-sample z-test for proportions. This test will compare the proportion of students in our sample who believe a hot dog is a sandwich to the hypothesized population proportion of 50%. We will calculate the test statistic using the formula Z = (p-hat - p_0) / sqrt((p_0 * (1 - p_0)) / n), where p-hat is the sample proportion, p_0 = 0.5 is the hypothesized proportion, and n is the sample size. Using RStudio, we will determine the p-value and compare it to

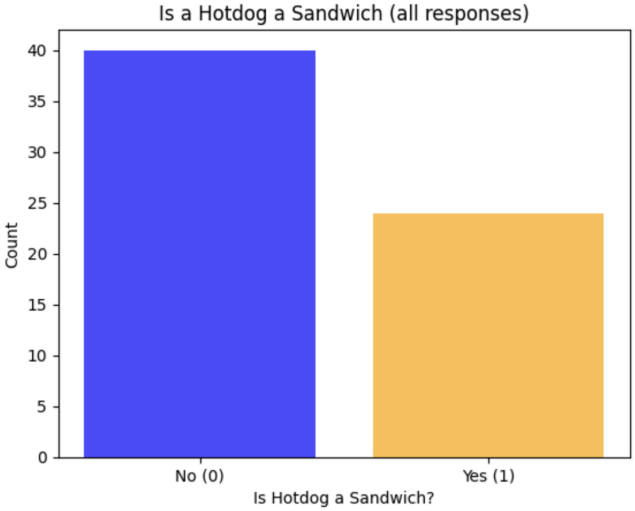
our significance level (alpha = 0.05) to decide whether to reject or fail to reject the null hypothesis. This approach will allow us to statistically evaluate whether the proportion of UVA students who believe a hot dog is a sandwich significantly differs from 50%.

Evaluation of Success

To evaluate the success of our analysis, we will determine whether the results of our one-sample z-test for proportions provide sufficient evidence to reject the null hypothesis. Specifically, success will be defined as obtaining a p-value less than our chosen significance level (alpha = 0.05), which would indicate that the proportion of UVA students who believe a hot dog is a sandwich is significantly different from 50%. Additionally, we will assess the sample proportion (p-hat) to determine whether it suggests a meaningful deviation from the hypothesized 50%, regardless of whether we reject the null hypothesis. While exploring secondary trends, such as correlations between selected variables (e.g., year of study or other demographic factors) and responses to the hot dog question, we will calculate correlation coefficients and interpret their strength and direction. However, these exploratory analyses are supplementary and not directly tied to testing our main hypothesis. The primary measure of success remains the outcome of the z-test for proportions.

Emily McMahon, Max Baskin, Tamer El Aridi, Barbara Uzun





<Figure size 640x480 with 0 Axes>