

Vivienne Hughes
Ethan Ogilvie
Sophie Phillips
Seblework Alemu
Avery Donmoyer

Pre-Analysis Steps

In order to conduct our analysis we first needed to clean the data set. To start data cleaning, we dropped all rows where a student answered 'no' for an undergraduate. Then we dropped all rows where there was no response for whether or not a hotdog is a sandwich. After this, we scanned both columns for the letter 'y' or 'n', replacing them with either 'yes' or 'no' to ensure answers were documented uniformly across the dataset.

- 64 → removed responses that were not UVa undergraduates, or did not answer sandwich prompt → 58
- Clean Data: organize the Yes/No responses to a binary system, remove irrelevant columns
- Conduct exploratory analysis by producing graphs to observe trends in variables of interest. Plot response by count by undergraduate year and whether or not that are a Virginia resident.
- Normalize dependent variable to frequency of responses by undergraduate year

Analysis Methods

To conduct analysis, we compared the proportion of survey respondents who either agreed or disagreed with the statement "A hot dog is a sandwich." To calculate these proportions, we conducted a chi-squared distribution test and compared the results by the respondents' year in school and reported gender. Chi-square result produced p value of 0.3, greater than alpha value 0.05, providing evidence we fail to reject the null hypothesis, that 0.5 students believe a hot dog is considered a sandwich. Compared the result with a one-sample proportion z test and found that there was not a significant difference between proportions ($p=0.4$), further strengthening the conclusion.

We decided to sort the data by year in school in order to control for the respondents' age/education level while assessing their responses.

Evaluation of Success

Chi-Square Test Results:

Chi-Square Statistic: 4.8615221987315

P-value: 0.3018035506086386

Degrees of Freedom: 4

Expected Frequencies:

[[15.01587302 6.98412698]

```
[ 2.73015873  1.26984127]
[ 0.68253968  0.31746032]
[ 0.68253968  0.31746032]
[23.88888889 11.11111111]]
```

There is **no significant association** between gender and the responses.

```
prop.test(24, 58, p = 0.5, alternative = c("two.sided"), conf.level = 0.95, correct = TRUE)
```

```
1-sample proportions test with continuity correction

data:  24 out of 58, null probability 0.5
X-squared = 1.3966, df = 1, p-value = 0.2373
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.2885656 0.5504354
sample estimates:
              p
0.4137931
```

Associated Graphs





