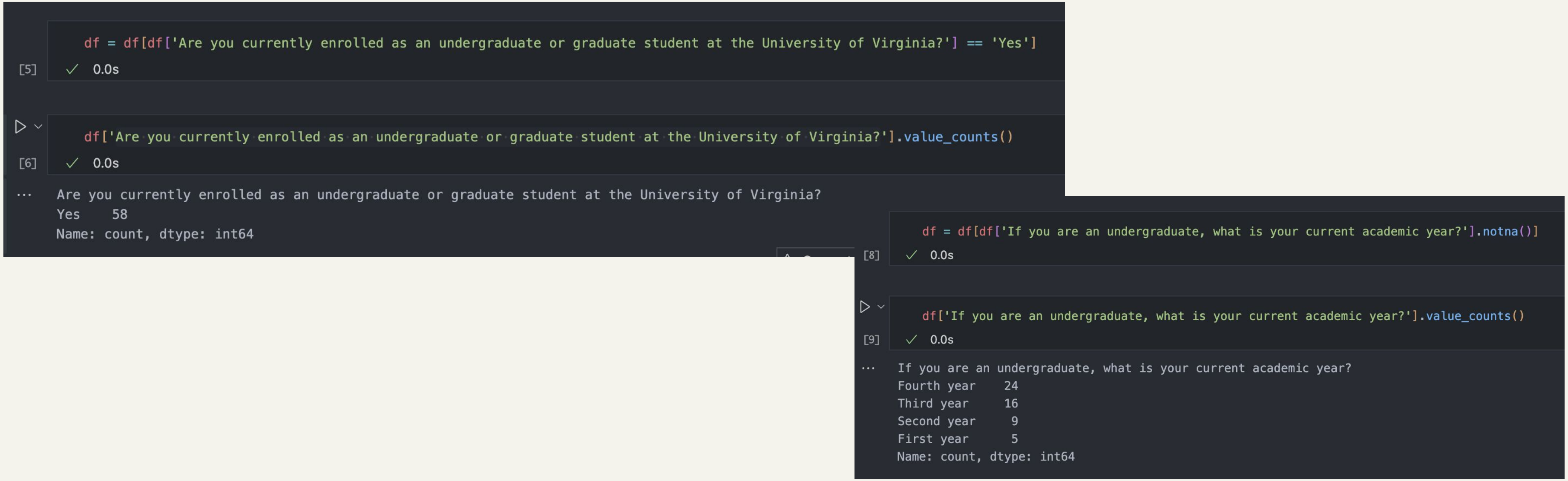
IS A HOTDOG A SANDWICH??

Pre-analysis

- Since our hypothesis is “1/2 of UVA full-time undergraduate students believe a hot dog is a sandwich”, we need to clean our data to only include UVA undergraduate students
- This involves only including rows where the answer to “Are you currently enrolled as an undergraduate or graduate student at the University of Virginia?” is “Yes” (to ensure they are enrolled in UVA) and where the answer to “If you are an undergraduate, what is your current academic year?” is not null (indicating they are indeed an undergraduate student)



```
[5]  df = df[df['Are you currently enrolled as an undergraduate or graduate student at the University of Virginia?'] == 'Yes']
     ✓ 0.0s
```

```
[6]  df['Are you currently enrolled as an undergraduate or graduate student at the University of Virginia?'].value_counts()
     ✓ 0.0s
```

```
Are you currently enrolled as an undergraduate or graduate student at the University of Virginia?
Yes    58
Name: count, dtype: int64
```

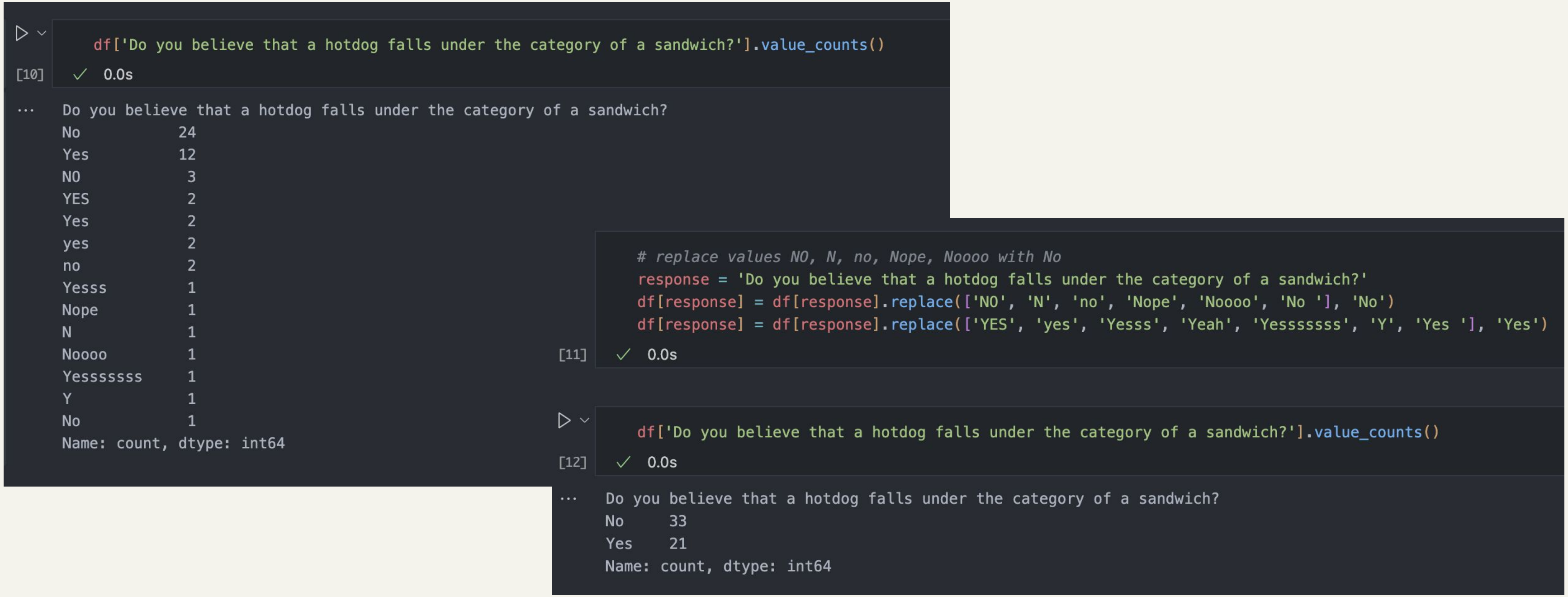
```
[8]  df = df[df['If you are an undergraduate, what is your current academic year?'].notna()]
     ✓ 0.0s
```

```
[9]  df['If you are an undergraduate, what is your current academic year?'].value_counts()
     ✓ 0.0s
```

```
If you are an undergraduate, what is your current academic year?
Fourth year    24
Third year     16
Second year     9
First year      5
Name: count, dtype: int64
```

Pre-analysis

- The previous step left us with 54 valid responses to help test our hypothesis, but we still needed to clean the actual response variable



[10]

```
df['Do you believe that a hotdog falls under the category of a sandwich?'].value_counts()
```

✓ 0.0s

```
Do you believe that a hotdog falls under the category of a sandwich?
No           24
Yes          12
NO            3
YES           2
Yes           2
yes           2
no            2
Yesss         1
Nope          1
N             1
Noooo         1
Yesssssss     1
Y             1
No            1
Name: count, dtype: int64
```

[11]

```
# replace values NO, N, no, Nope, Noooo with No
response = 'Do you believe that a hotdog falls under the category of a sandwich?'
df[response] = df[response].replace(['NO', 'N', 'no', 'Nope', 'Noooo', 'No '], 'No')
df[response] = df[response].replace(['YES', 'yes', 'Yesss', 'Yeah', 'Yesssssss', 'Y', 'Yes '], 'Yes')
```

✓ 0.0s

[12]

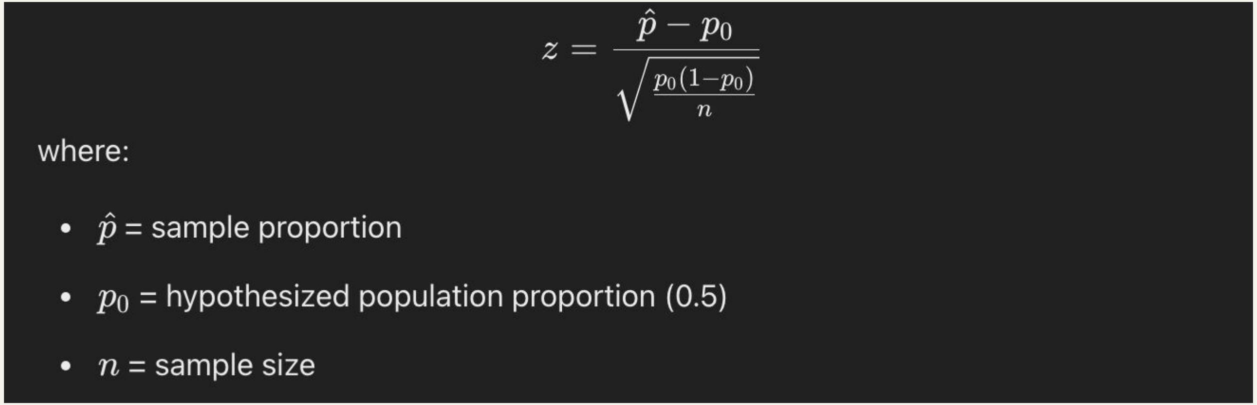
```
df['Do you believe that a hotdog falls under the category of a sandwich?'].value_counts()
```

✓ 0.0s

```
Do you believe that a hotdog falls under the category of a sandwich?
No     33
Yes    21
Name: count, dtype: int64
```

Analysis

- We will use a one-sample z-test for proportions to test our hypothesis.
- Sample proportion: divide the number of students in our sample who believe a hot dog is a sandwich by the total number of students surveyed
- Population proportion: assumed to be 0.5, as that's what the hypothesis proposes.
- We will then calculate a z-score value using the formula:



$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where:

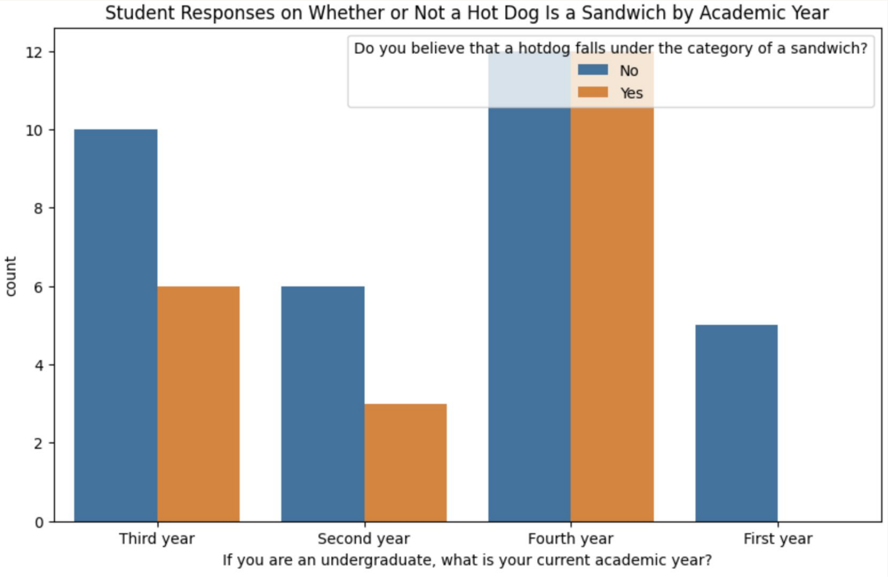
- $\hat{p}$ = sample proportion
- p_0 = hypothesized population proportion (0.5)
- n = sample size

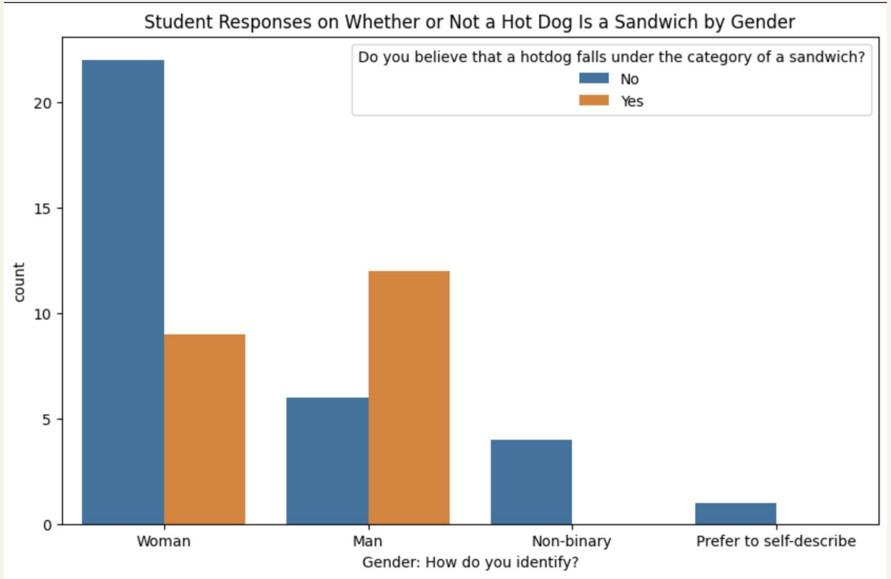
- Using a significance level of 0.05 with a two-tailed test, our decision rule is simply if Z is less than -1.96 or greater than 1.96, we will reject the null hypothesis.

Analysis

- After fully cleaning our data, we found out of 54 students 33 did not believe a hot dog is a sandwich while 21 did believe a hotdog is a sandwich
- This means our sample proportion is 21/54, or 0.388888889
- Follow the formula, we get a z-score of z = (0.388888889 - 0.5) / sqrt((0.5*0.5)/54) = **-1.63299316032**
- Because our z-score is not less than -1.96 or greater than 1.96, we **fail to reject the null hypothesis** that half of UVA full-time undergraduate students believe a hot dog is a sandwich (at a significance level of 0.05).

Visualizations





Link to all code used for this project: https://github.com/amoghghadge/DS-4002/blob/spring-2025_section-18751/data_analysis.ipynb

Thank you for listening