

Name: Anne Gidley

UmichID: adirkex

What dataset are you working with:

List 3 questions that you can ask with your dataset.

Q1: Are people with higher education levels more likely to have heard of the Oxford comma?

Q2: Do people who have strong feelings about the oxford comma have higher incomes?

Q3:

List the associated null hypothesis for each question:

Q1: All education levels have the same proportion of familiarity with the Oxford comma

Q2: Average income is the same for all groups of Oxford comma feelings

Q3:

What statistical test(s) will you use to answer each of the questions:

Q1: Chi Square (two categorical variables: familiarity depends on ed level)

Q2: If I knew how to turn the income data into a continuous variable, I would use a One-way ANOVA, but since it is split into income groups, I do a Chi-Square test to compare the two categoricals.

Q3:

Make a visual plot showing the relationship that you will analyze statistically (e.g. boxplot for t-test or ANOVA; scatterplot for regression; table for chi-square).

Q1:

	FALSE	TRUE
Less than high school degree	7	4
High school degree	57	43
Some college or Associate degree	132	163
Bachelor degree	126	218
Graduate degree	89	187

Q2:

	Not at all	Not much	Some	A lot
\$0 - \$24,999	10	31	39	41
\$25,000 - \$49,999	17	33	67	41
\$50,000 - \$99,999	32	82	120	56
\$100,000 - \$149,999	19	37	54	54
\$150,000+	9	26	35	33

Q3:

Do your data meet the assumptions required for the statistical test you want to run? Please state the assumptions you examined and whether or not your data meet those assumptions:

Q1: The data are likely to meet the assumptions, although most would require knowledge about how the data were collected. I expect that the data are from an independent random sample, as required by the Chi-Square test. The sample size is large enough, and there are no cells with values of zero, which would require a correction.

Q2:

Q3:

Run the statistical test! Put your results here:

Q1:

```
Pearson's Chi-squared test  
data: tbl  
X-squared = 25.9, df = 4, p-value = 3.315e-05
```

Q2:

```
Pearson's Chi-squared test  
data: comma_survey$household_income and comma_survey$care_oxford_comma  
X-squared = 19.86, df = 12, p-value = 0.06979
```

Q3:

Interpret your results!

Q1:

Based on the p-value, there is a significant difference in number of “true” and “false” responses to “Have you heard of the Oxford Comma?” depending on the education level of the survey respondent. The table shows that familiarity with the oxford comma increases with education level, although familiarity levels off around bachelor degree.

Based on the number of respondents in each education category, I wonder if higher education was overrepresented. It would be interesting to compare the proportions in each education category of the sample to the proportions in the population.

Q2:

Considering income level as a categorical, there is no significant difference, but a

more accurate analysis would compare the continuous value of income against the categorical variable (ANOVA).

Q3: