Lancaster University

Quantitative Research Methods

Matthew Ivory

matthew.ivory@lancaster.ac.uk

1. Introduction to the course

- Session 1: Introduction to quantitative research methods using R
- Session 2: Data management and data wrangling
- Session 3: Exploratory data analysis
- Session 4: Data visualization
- Session 5: Live Coding Walkthrough
- Session 6: Probability and distributions
- **Session 7: Tests for discrete variables: Analysing contingency tables**
- Session 8: Correlations and t-tests.
- Session 9: ANOVA and linear regression
- Session 10: Multiple regression, introduction to generalised linear regression

In this session

- Hypothesis Testing
  - Inferential statistics
  - Probabilities
  - Null effects and null hypotheses
- Parametric vs Non-parametric data
- Testing Nominal data
  - Chi-squared test
  - Cramer’s V

The point of a statistical test

- To tell us whether there is a pattern in our data

- A test gives us two things: the size of the effect, and a p-value, which tells us about the significance

- If a test is significant, then it tells us there is a pattern, or some structure in our data
- If a test is not significant, then this could be for two reasons
  - There is no pattern in the data
  - There is a pattern but we can’t find it through the noise
  - ***A statistical test cannot decide between these alternatives***

Inferential Statistics – What is it?

- Inferential statistics lets us make *inferences* or *predictions* about a population based upon our sample
- Unrealistic to collect data from an entire population
  - Too costly in time, money, etc.
- Instead, we sample the population at random and using that data, we can make *inferences* based upon probabilities

Hypothesis Testing

- When we are not interested in exact values, but rather asserting some statement regarding the parameter of interest
- Examples: the average UK income is more than £35,000, male academics get paid more on average than female academics, an increased visible police presence reduces petty crime rates

Hypothesis Testing

- More specifically, we are interested in null hypothesis significance testing
- The main idea of NHST is that you are testing that there is no significant difference between two values
  - Such as between two means
- The null hypothesis states that there is no difference between the two means (or groups) of interest
- Statistical tests seek to determine the probability of finding a difference in your *sample*, if no real difference exists in the *population*
  - *If no real effect exists, what are the chances that I would detect an effect in my sample?*

Hypothesis Testing

- Hypotheses can be directional or non-directional
- Directional means that we are predicting a difference between groups as well as the direction of the effect
  - Exam scores will be higher for students who studied more than four hours a week than those who studied less than four hours a week
- Non-directional means we predict a difference but don't know which way it will go
  - Exam scores will be different for students who studied more than four hours a week than those who studied less than four hours a week
- Hypothesis **must** be worded so that it can be tested and must include both independent and dependent variables
  - **Exam scores (DV)** will be higher for **students who studied more than four hours a week (IV)** than those who did not

Hypothesis Testing

- Independent Variables (IV) – is independent to other variables in the study
- Dependent Variables (DV) – is dependent on changes in independent variables
  - The DV is what we are interested in explaining
- We measure both to determine the effect of the IV on the DV
  - Study time on test scores
  - Room temperature on plant growth
  - Population density on violent crime

Hypothesis Testing

- We collect test scores from two classes on the same test
  - class1 <- c(8,6,4,9,5,8,6,8,9,7,5,7)
  - class2 <- c(3,6,3,7,2,5,2,1,2,7,9,5)
  - Mean of class1 == 6.83
  - Mean of class2 == 4.33
- The difference between class1 and class2 is -2.5
- in terms of the Null Hypothesis, we now want to know: What is the probability of finding a difference between means of -2.5 (or larger) if no real difference between the two groups exists in the population?

Hypothesis Testing and *p*-values

- The ***p*-value** is the probability of finding a difference equal to or greater than the one found if no difference exists in the population. Let's say our *p*-value = .010
- This indicates a very small probability of finding a difference equal or greater in the population if there was no difference. The obtained *p*-value is also smaller than the standard cut-off that we use in Psychology of .05
- As such we would reject our null hypothesis and suggest that there is a significant difference between the two groups.
- *If* the p-value was *greater* than .05 (e.g. .1),we would *fail to reject* the null hypothesis, **not** *accept the null*

Reproducibility – variation and reliability

Lancaster University

- Results will not replicate perfectly, because responses contain noise
- observed scores are a combination of **true score** and **error**
- If the error is random then
  - some people will accidentally score more, and some less, than their true score
  - true differences between participants' scores will be unrelated to error differences
- What we call *error* is also *noise*, and represents the deviation from what we consider the true value
- Average UK height in 2021 was 169.15cm. Your own height is 169.15 ± error

Effect sizes and *p*-values

- When we construct an experiment, we construct a research hypothesis:
  - Two variables are related
  - An experimental manipulation will affect another variable

- We also construct a null hypothesis:
  - Two variables are not related
  - The experimental manipulation has no effect

- *p*-value: how confident can we be in rejecting the null hypothesis
  - A p-value of 0.05 tells us that 5% of the time when we find an effect like this, the null hypothesis is true (a **false positive**)

  - If the p-value is less than 0.001, we write by convention: $p < .001$
    - If there's a chance less than one in a thousand that we find this effect when the null hypothesis is true then we're quite confident in our result

Reminder: Types of Data

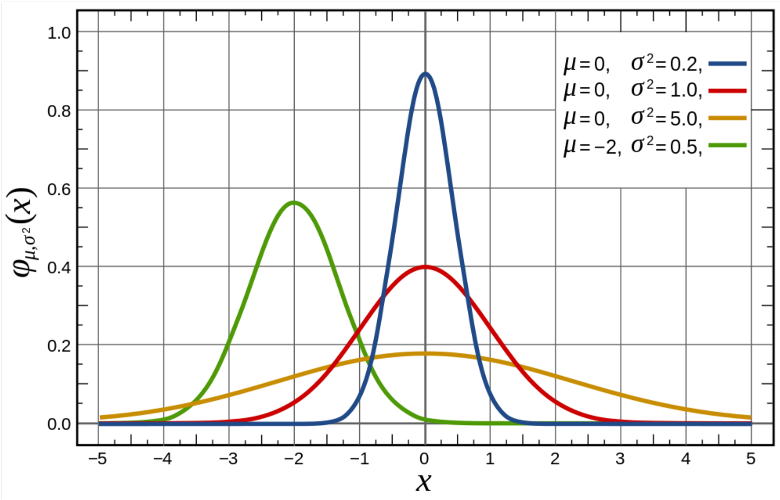
Numbers have three important functions for researchers:

1. Classify or categorise – **Nominal/Categorical data**
   - E.g. male/female, young/old, red/green/blue
   - Each case can only be in one category.

2. Rank or order - **Ordinal data**
   - E.g. first, second, third…
   - Does not indicate magnitude of differences between ranks.

3. Score – **Continuous/Count**
   - E.g. number of correct answers, reaction time
   - Tells us about the order of data but also the distance between them.

Reminder: Normal distribution



- We assume (or test) that the actual data comes from a "normal distribution" of data
- The shape of the normal distribution depends on two factors:
  - **Mean (μ)** – Measure of central tendency
  - **Standard Deviation (σ)** – Measure of variability

Parametric and non-parametric

- If data fit a normal distribution, then we can use “parametric” statistical tests
- If data do not fit a normal distribution, then we need to use “non-parametric” statistical tests – they are *“distribution free”*

- Categorical data are non-parametric
- One non-parametric test is Chi-squared, which is useful for determining if categorisations into two or more groups are different than chance

Example: Titanic

- The ship Titanic sank in 1912 with the loss of most of its passengers
- 809 of the 1,309 passengers and crew died

= 61.8%

- **Research question:** Did class (of travel) affect survival?

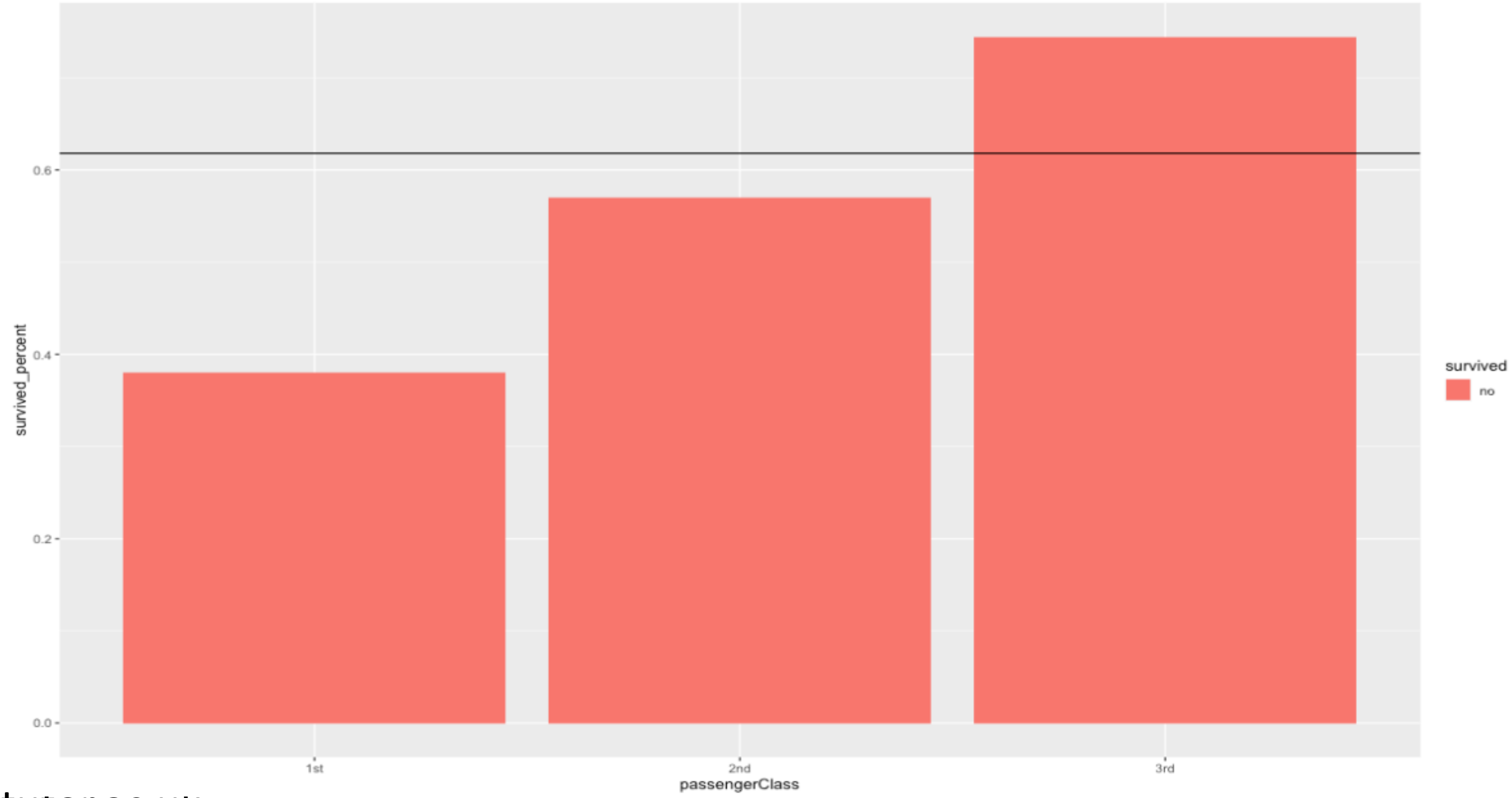
Chi squared Test

- **Null:** There is **NO** association between class and survival
- **Research:** There **IS** an association between class and survival

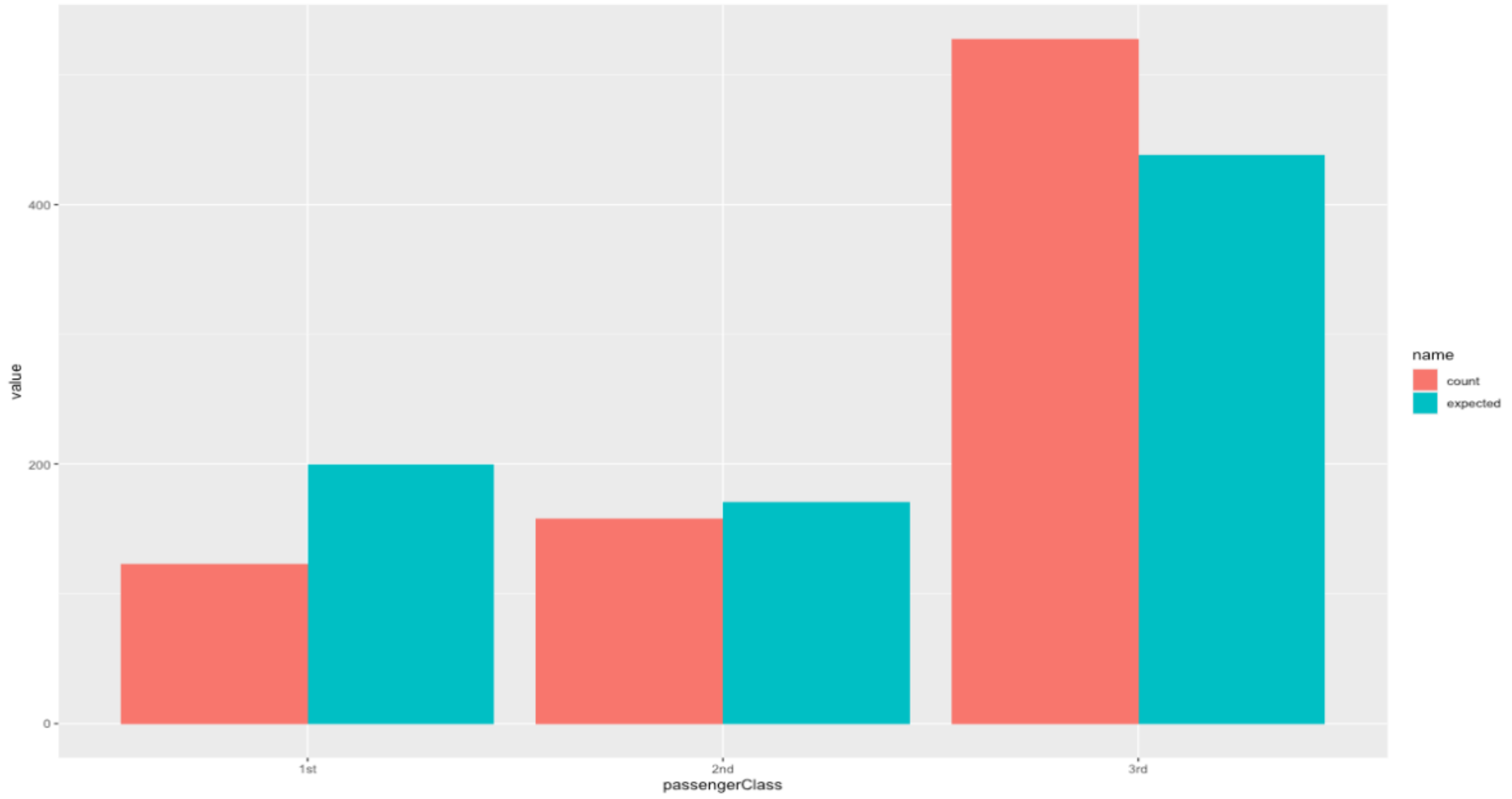
Class	Died	Survived	Total
1st	123	200	323
2nd	158	119	277
3rd	528	181	709
Total	809	500	1309

What would be expected if there is no association? Expected and actual values

- Same proportion of people would have died in each class!
- Overall, 809 people died out of 1309 = 61.8%



Chi-Squared Test Compares Observed and Expected Frequencies



Expected number dying in each class = 0.618 x number in each class

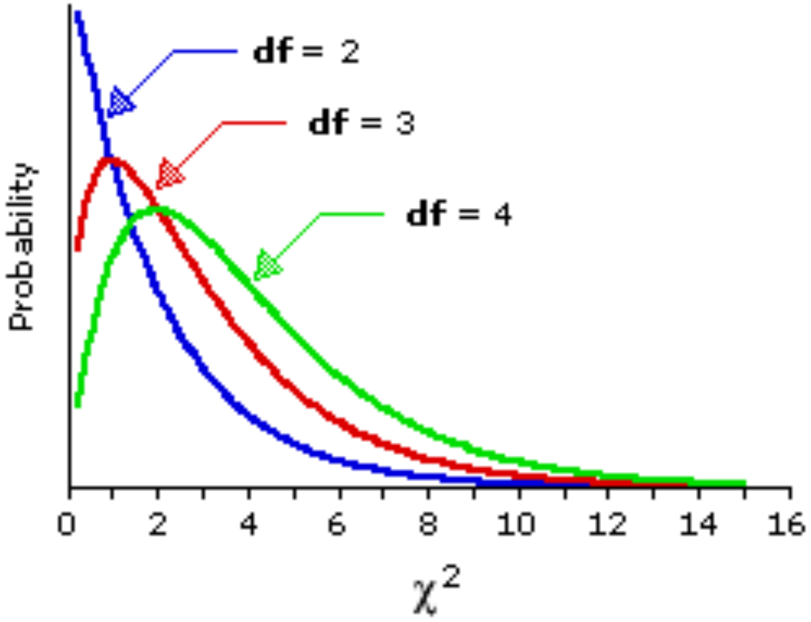
Chi-squared test statistic

- The Chi-squared test is used when we want to see if two categorical variables are related
- The test statistic for the Chi-squared test uses the sum of the squared differences between each pair of observed (O) and expected values (E)

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Chi squared distribution

- The p-value is calculated using the Chi-squared distribution for this test
- Chi-squared is a skewed distribution which varies depending on the degrees of freedom



degrees of freedom = (no. of rows – 1) x (no. of columns – 1)

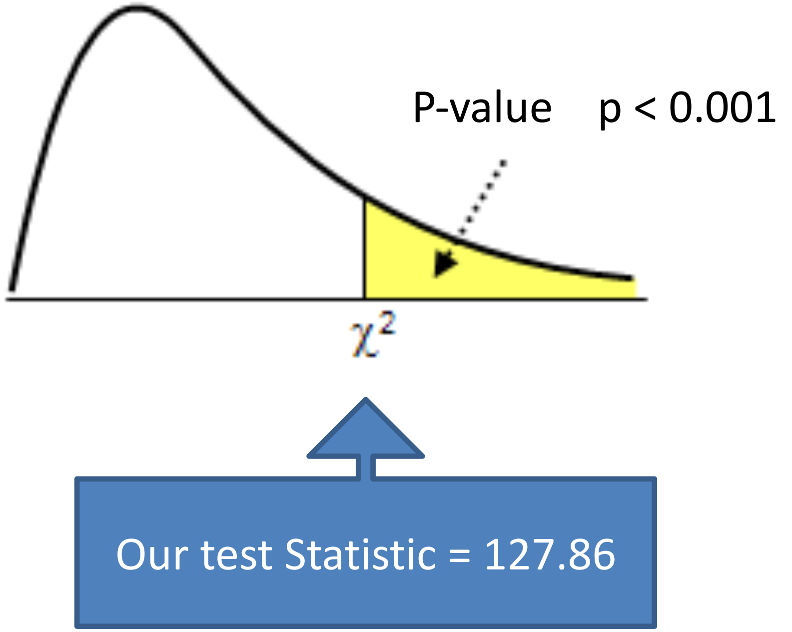
For Titanic example, number of rows = 3 (class), number of columns = 2 (survived or not)

So for Titanic, degrees of freedom = …

What’s a p-value?

In the Titanic example, the probability of getting a test statistic of 127.86 or above (**just by chance**) is < 0.001

Distribution of test statistics



Interpretation of the Chi-squared test

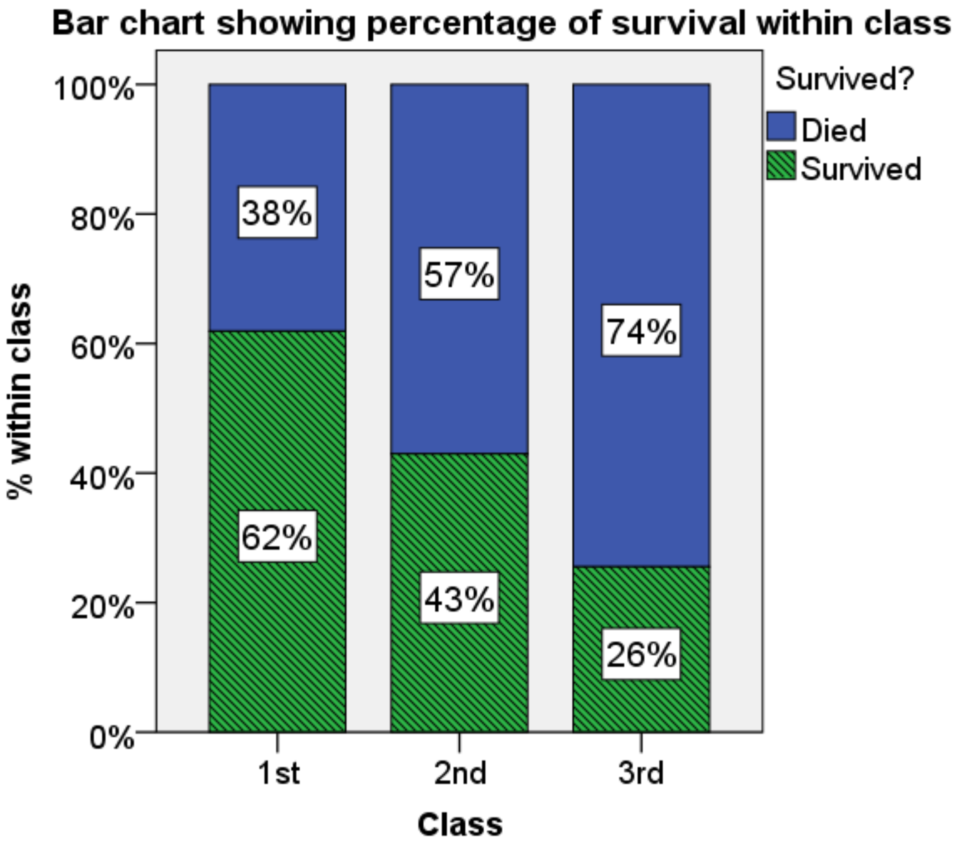


Figure 1. Bar chart showing % of passengers surviving within each class

Data collected on 1309 passengers aboard the Titanic was used to investigate whether class had an effect on chances of survival. There was evidence ($\chi^2(2, N = 1309) = 127.86$, $p < 0.001$) to suggest that there is an association between class and survival.

How would you interpret this finding?

Figure 1 shows that class and chances of survival were related. As class decreases, the percentage of those surviving also decreases from 62% in 1st Class to 26% in 3rd Class.

Assumptions of the Chi-squared test

- Cells should have expected values >= 5
- If any expected cell counts are <1 then cannot use the chi-squared distribution
- In these cases, use **Fishers' Exact test**

That’s the p-value. What about the effect?

- Effect sizes tell us about the magnitude of the difference found.
- How meaningful is the relationship that we have identified
- In essence, how much should we care about our finding in terms of real-world use?

- If a teaching approach improves class scores by 1 mark (and significantly so) compared to another, how useful is this really? Is it a meaningful improvement? What about when you consider the additional resources required?
- If sitting in first class improves your survival likelihood by 2%, is this a useful statistic to consider when buying transatlantic tickets?

That's the p-value. What about the effect? Cramer's V

- Cramer's V adjusts for the number of observations

- V = √ (Chi-squared / (N * w))
  - Where N is the total
  - w is the smaller of the (rows – 1) or (columns -1)

Class	Died	Survived	Total
1st	123	200	323
2nd	158	119	277
3rd	528	181	709
Total	809	500	1309

Class	Died	Survived	Total
1st	61	100	161
2nd	79	60	139
3rd	264	90	354
Total	404	250	654

- Titanic: Chi^2 = 127.86
- V = √ (127.68/(1309*1)) = .31

Teenitinic: Chi^2 = 64.92

V = √ (64.92/(654*1)) = .32

Interpreting the effect: Cramer’s V

- Cramer’s V adjusts for the number of observations

- Cramer’s V is between 0 and 1
- A measure of the association between the row and column categories
  - 0 = no association
  - 1 = perfect association
  - In between … relative strength (but not absolute)

Summary

- Chi-squared test is a measure of association between categories/nominal groups
- Interpretation of Chi-squared and p-values
  - Expected and observed frequencies
- Cramer’s V effect size measure of association

Questions?

- I will be walking around while you work through the worksheet