

MA206, Lesson 15 - Two Groups - Two Proportions

**Review:** What types of variables do we compare when testing two proportions?

**Review:** What is the parameter of interest when comparing two proportions?

**Review:** What is the statistic of interest used to infer about the parameter?

What is  $\hat{p}_{Total}$ ?

How do we find the standardized statistic for 2 proportions using theory?

How do we calculate the confidence interval for the difference in two proportions?

For several years in the 1990s, Kristen Gilbert worked as a nurse in the intensive care unit (ICU) of the Veterans Administration Hospital in Northampton, Massachusetts. She became one of the nurses that others looked up to as an example of skill as she was particularly good in a crisis. When a victim went into cardiac arrest, for example, she was often the first to notice that something was wrong and calmly apply adrenaline to restart the patient's heart. Over the course of her time there, other nurses came to suspect that she was killing patients by injecting them with the heart stimulant epinephrine. An investigation was launched of 1,641 eight-hour shifts during her time in the ICU to see if there was indeed an increase in death rate while she was on shift. For each shift, two variables were recorded: Did Gilbert work on the shift, and did at least one patient die during this shift.

a) Identify the observational units in this study.

b) List the applicable variables and classify them as categorical or quantitative.

c) Which variable would you classify as explanatory and which is response? Draw the Causal Diagram.

d) In words and symbols, write the null and alternate hypotheses.

The data from this study is available and can be imported using R by the code below. Copy and run this code to begin analysis and provide evidence for or against Gilbert's case at trial.

```
Nurse <- read_delim("http://www.isi-stats.com/isi/data/chap5/Gilbert.txt", delim="\t")
```

e) Generate a Segmented Bar Graph and a Mosaic Plot for the Gilbert data. Interpret the results.

f) Do we meet the validity conditions required to use theoretical methods?

h) What is our statistic (difference of proportions)?

i) Calculate the standard deviation and standard error.

j) Report your standardized statistic and p-value. Interpret the results.

k) Using the applet, visualize this scenario. Do these results line up with the theoretical analysis?

l) Calculate the 90% confidence interval and interpret your results.

j) Based on these results, do we have reasonable suspicion that the differences in death rates is not due to random chance?

k) Based on these results, have we proven that Kristen Gilbert caused the deaths of her patients?

l) If you were on the prosecution for Kristen Gilbert, what might you say to sway the jury towards a guilty verdict? You may refer to your calculations and Figures.

m) If you were on the defense for Kristen Gilbert, what might you say to sway a jury against this result? You may refer to your calculations and Figures.

2) Use your own project dataset to explore the relationship between two categorical variables. If your categorical variables have more than one category, you may wish to transform them into binary variables such that one category is treated as a “success” and the other is treated as a “failure.” Alternatively, you can manually do the math from your resulting tables.

*If your dataset does not have two different categorical variables, you may convert your response variable to a categorical variable to look for an association. For example code, see below and adjust for your variables. The code below takes a numeric response (Score) and converts them to categories based on value (A, B, C, D, or F).*

```
Data2 <- Data1 %>%  
  mutate(ResponseCat = case_when(  
    Score <= 65 ~ "F",  
    Score < 70 ~ "D",  
    Score < 80 ~ "C",  
    Score < 90 ~ "B",  
    TRUE ~ "A"  
  )  
)
```

a) Provide a brief description of your dataset. What are the observational units?

b) Describe the two categorical variables you are comparing. What is your null and alternate hypothesis about their association?

c) Generate a two-way table of your data. Do you meet the validity conditions to calculate using theory?

d) Generate a segmented bar chart. Describe the shape and its implications.

e) Generate a mosaic plot. Describe the shape and its implications. Does this provide more information than the segmented bar chart previously?

f) Using your data and hypothesis above, calculate your observed statistic, standard deviation, standardized statistic, and p-value.

g) Calculate a 99% confidence interval. Interpret your results.