

Joseph Lally

ADEC7310.02 – Data Analysis [Fulton]

Week 1 Homework

8/29/2017

### 1.1 Migraine and acupuncture

- a) What percent of patients in the treatment group were pain free 24 hours after receiving acupuncture?  
What percent in the control group?

Treatment group =  $(10/43) = 23.25\%$

Control group =  $(2/46) = 4.35\%$

- b) At first glance, does acupuncture appear to be an effective treatment for migraines? Explain your reasoning.

Yes, given the ~19% increase in pain-free patients who received the acupuncture treatment over those who were exposed to the placebo, it appears that the treatment group received effective treatment.

- c) Do the data provide convincing evidence that there is a real pain reduction for those patients in the treatment group? Or do you think that the observed difference might just be due to chance?

Given that the amount of pain free patients in the treatment group resulted in a ~19% increase from the control group, this provides reasonable basis for the claim that the acupuncture treatment does produce real pain reduction for patients who receive the treatment.

### 1.2 Sinusitis and antibiotics

- a) What percent of patients in the treatment group experienced a significant improvement in symptoms?  
What percent in the control group?

Treatment group =  $(66/85) = 77.65\%$

Control group =  $(65/81) = 80.25\%$

- b) Based on your findings in part (a), which treatment appears to be more effective for sinusitis?

Given the ~2.6% difference in favor of the control group's treatment plan, it appears as though the placebo treatment is more effective than the amoxicillin treatment.

- c) Do the data provide convincing evidence that there is a difference in the improvement rates of sinusitis symptoms? Or do you think that the observed difference might just be due to chance?

The ~2.6% variance in the two treatment plans is not convincing, especially given the small sample size from which the researchers studied. As such, it is more likely than not that the observed difference in results is due to chance.

### **1.3 Air pollution and birth outcomes, study components**

- a) Cases: 143,196 births between 1989-1993 in southern California.
- b) The variables and their types:
  - a. Measurements on the levels of carbon monoxide (ppm), nitrogen dioxide (pphm), oxone (pphm), coarse particulate matter (PM10) in  $\mu\text{g}/\text{m}^3$  – continuous numerical.
  - b. Length of gestation – continuous numerical.
- c) The main research question: Is there a relationship between exposure to air pollution and preterm birth rates?

### **1.4 Buteyko method, study components**

- a) Cases: 600 asthma patients aged 18-69 years old who rely on medication for asthma treatment.
- b) The variables and their types:
  - a. Quality of life, activity, asthma symptoms, and medication reduction – each on a scale from 0-10. These are discrete numerical variables, assuming that each case would not be able to score each variable at the decimal point level (eg. 5.5/10). Otherwise, they would be considered to be continuous numerical variables.
- c) The main research question: Does the Buteyko method improve conditions for asthma patients' conditions?

### **1.5 Cheaters, study components**

- a) Cases: 160 children between the ages of 5-15.
- b) The variables and their types:
  - a. Age – discrete numerical
  - b. Sex – categorical
  - c. Flipped coin (white/black) - categorical
  - d. Whether or not the participant was an only child or not – categorical
- c) The main research question: Is there a relationship between honesty, age, and self-control for children ages 5-15?

### **1.6 Stealers, study components**

- a) Cases: 129 University of California undergraduates at Berkley.
- b) The variables and their types:
  - a. Most (least) money – continuous numerical
  - b. Most (least) education – ordinal
  - c. Most (least) respected jobs - ordinal
  - d. Number of candies taken – discrete numerical

- c) Is there a relationship between socio-economic class and unethical behavior?

### 1.7 Fisher's irises

- a) Summary function shows that there are 150 individual records made up of three different species of flowers at 50 records each. This tells us that there are 150 cases in the iris dataset.
- b) Summary function also shows us that there are four determinable continuous numerical variables (eg. Sepal.length, sepal.width, petal.length, & petal.width). They are continuous because the decimal values can be plotted along any point inside of a standard integer/whole number.
- c) Summary function finally shows us that there are one categorical variable: species of flowers, with three levels of categories are setosa, versicolor, and virginica.

>summary(iris)

summary(iris)	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	Species Count
Min.	4.300	2.000	1.000	0.100	setosa	50
1st Qu.	5.100	2.800	1.600	0.300	versicolor	50
Median	5.800	3.000	4.350	1.300	virginica	50
Mean	5.843	3.057	3.758	1.199		
3rd Qu.	6.400	3.300	5.100	1.800		
Max.	7.900	4.400	6.900	2.500		