Michael Rimmey

Professor Jeffrey Ziegler

Quantitative Political Methods

27 March 2019

**Question 1**

**1 a)**

**1 b)**

**1 c)**

**Question 2**

**2 a)**

**2 b)**

**Question 3**

**3 a)**

**3 b)**

**Question 4**

Population mean = sample mean +/- sample error

93.87 +/- 1.812 times 9.50, divided by square root of 11

Sample error = **5.18**

To conduct a 90% confidence interval for number 4, we must use a **T-table** instead of a Z-table. This is because the sample size is incredibly small (11 respondents). The number we get from the T-table is **1.812**, since we are conducting a confidence interval of 90% and the degrees of freedom are equal to **10** (number of observations minus 1). With the T-statistic in place, we multiply it by the sample standard deviation (9.50). We then divide this by the square root of the number of observations (**3.32**). Finally, we add and subtract this sample error from the sample mean and get the following answer:

93.87 - 5.18 = **88.69**

93.87 + 5.18 = **99.05**

Population mean = (**88.69, 99.05**)

**Question 5**

**5 a)**

To complete the “no” portion of the column, we must subtract the known amount (360) from the total number of respondents (781). We then subtract 225 from the total we got for the previous column. We continue the process for the other missing data and get the following results:

Democrats who disapprove of the ACA, expected value: **240.42**

Republicans who approve of the ACA: **139**

Republicans who disapprove of the ACA: **196**

Total number of respondents who disapprove of the ACA: **421**

To calculate the expected frequency, we simply multiply the row total by the column total and divide that sum by the number of observations. For example, for the first expected value in question 5, we multiply 446 by 360 and divide that total by 781.

**5 b)**

To calculate the cell component for Republicans who disapprove of the ACA, we must…

**5 c)**

**Null Hypothesis**: Attitudes toward ACA **not** impacted by party affiliation

**Alternative Hypothesis**: Attitudes toward ACA impacted by party affiliation

X2 = 5.02

To get the degrees of freedom for this test, we simply subtract 1 from the number of rows and the number of columns multiply the two sums.

DF = **2**

We can plug this information into R to get the P-value.

1. pchisq(5.02, 2, lower.tail = FALSE)

By doing this, we get a P-value of **0.081**.

P = 0.081x10-13

We **cannot** reject the null hypothesis.

**5 d)**

The table for question 5 tells us that party affiliation likely plays a role in whether respondents approve of the Affordable Care Act, even though it does not play an overly significant role.

**Question 6**

**6 a)**

3.8 – 3.5 = **0.3**

SD divided by square root of N + SD divided by square root of N = **0.2**

Number of observations – 1 + number of observations – 1 = DF = **528**

DF 528, CI 0.05 = **1.96**

Critical value (1.96) times difference (.2) = **0.392 margin of error**

**0.3 +/- 0.392 = 0.692, 0.092 Confidence Interval**

With repeated random sampling, **95% will fall between 0.092 and 0.692.**

For question 6, we simply conduct a 95% confidence interval for the difference between the two means (3.8 and 3.5). To do this, we first subtract the lower mean and get a difference of **0.3.** The formula then calls on us to divide each sample’s standard deviation (2.2 and 2.4) by the square root of the sample size (242 and 288). We add the two numbers together and get **0.2.** To get the degrees of freedom, we subtract one from each sample and add them together. This leaves us with DF = **528.** For the critical value, we get **1.96** for a confidence interval of 95% with 528 degrees of freedom. Next, we multiply the critical value (**1.96**) by the difference (.**2**) and get a margin of error of **0.392**. Finally, we simply add and subtract the margin of error from the difference between the two numbers to get **0.092** and **0.692**.

**6 b)**

When testing the theory that the Civics class changed party ID, we find that…

**6 c)**

It is … to treat this estimate as casual because…

**Question 7**

**7 a)**

A **sample distribution** is simply the distribution of one sample. For instance, say we have 50 observations in the form of fishermen who went fishing on a particular day in summer. A certain number of the fishermen caught a fish while a certain number of them did not catch a fish. If you were to plot this one sample in something like a histogram, you would get a sample distribution.

A **sampling distribution** would be what we get if we would repeatedly draw samples from a population. For instance, say we have 100 observations in the form of WashU students. A certain number of the students in this sample would be Democrats, and a certain number would be Republicans. The exact number or percentage of how many students identify as Democrats would differ from sample to sample. Therefore, the percentage of people in each sample who identify as Democrats could, and likely will, differ from the population proportion, based on the 100 respondents in a given sample. These are called sample proportions. After many samples, the various sample proportions will form a bell-shaped curve around the population proportion.

The **population distribution** is

**7 b)**

**Question 8**

1. **P-Value**

In a hypothesis test, the **P-value** is the probability that the test statistic equals the observed value or a value even more extreme in the direction predicted by the alternative hypothesis.

1. **Outliers**

Observations are outliers if they fall more than **1.5** IQR **above** the **upper** quartile range or more than **1.5** IQR **below** the **lower** quartile range. The IQR, which helps to define which observations are outliers, is the difference between the upper and lower quartiles. Outliers can be highly influential to the mean of a dataset.

1. **Counterfactual**

A **Counterfactual** outcome is the potential outcome that did not occur. It is the outcome that would have occurred had the treatment been different. Counterfactual outcomes play a large role in causality, as actual outcomes are often compared to counterfactual outcomes.

1. **Autocorrelation**

**Autocorrelation** determines the presence of correlation between the values of variables that are based on associated aspects. This is usually found in datasets that are from the same source instead of random samples. It means that the correlation between the values of the same variables is based on the same source**.**

1. **Standard Error**

The standard deviation of the sampling distribution of y bar is called the **standard error** of y bar. In other words, the standard error of a statistic is simply the standard deviation of the statistic’s sampling distribution.