

Statistics 12, Lab 4

Michael Wu
UID: 404751542

March 5th, 2019

Exercise 1

a) This is a one sided test, since we only care if the proportion of dangerous lead levels is too high.

$$H_0 : p = 0.1 \quad H_a : p > 0.1$$

b) I ran the following code.

```
flint <- read.csv ("flint_2015.csv", header=TRUE)
p <- mean(flint$Pb >= 15)
```

This gave me the sample proportion $\hat{p} = 0.1238447$. I calculated the standard deviation of the sample proportion of dangerous lead levels by using the null hypothesis value for the true proportion of dangerous lead levels. This yielded the following value.

$$\sqrt{\frac{0.1 \times 0.9}{541}} = 0.012898$$

c) The standard error will be the same as the standard deviation of the sample proportion. So the z-value for this test will be the following value.

$$\frac{0.1238447 - 0.1}{0.012898} = 1.8487$$

d) I ran the following code.

```
pnorm(1.8487, lower.tail = FALSE)
```

This gave me a p-value of 0.03225057.

e) Using a significance level of 0.05, we would reject the null hypothesis.

f) Based on my results, I would recommend that remediation action should be taken.

g) I ran the following code.

```
prop.test(x = sum(flint$Pb >= 15), n = nrow(flint), p = 0.1, alt = "greater")
```

This output a p-value of 0.03778. Using the same significance level, my results stay the same. The difference in p-value is mostly due to the test using continuity correction.

h) I ran the following code.

```
prop.test(x = sum(flint$Pb >= 15), n = nrow(flint), p = 0.1, alt = "greater",  
          conf.level = 0.99)
```

This output the confidence interval (0.09376523, 1). This makes sense since it includes 0.1, the null hypothesis proportion, and we would fail to reject the null hypothesis at the corresponding significance value of 0.01.

Exercise 2

a) This is a two sided test, since we care about differences that are both positive and negative.

$$H_0 : p_1 = p_2 \quad H_a : p_1 \neq p_2$$

b) I ran the following code.

```
p1 <- mean(flint$Pb[flint$Region == "North"] >= 15)
p2 <- mean(flint$Pb[flint$Region == "South"] >= 15)
n1 <- sum(flint$Region == "North")
n2 <- sum(flint$Region == "South")
```

This gave me the following values.

$$\hat{p}_1 = 0.1762452$$

$$\hat{p}_2 = 0.075$$

$$n_1 = 281$$

$$n_2 = 260$$

From previously we have $\hat{p} = 0.1238447$, so we can calculate our test statistic using the following formula.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 3.57179$$

c) I ran the following code.

```
2 * pnorm(3.57179, lower.tail = FALSE)
```

I multiplied by two since this is a two sided test. This gave me a p-value of 0.0003545497.

d) Using a significance level of 0.05, I would reject the null hypothesis. This means that we would conclude that the proportion of houses with dangerous lead levels in the north is not the same as the proportion of houses with dangerous lead levels in the south.

e) I ran the following code.

```
prop.test(x = c(p1 * n1, p2 * n2), n = c(n1, n2), alt = "two.sided")
```

This output a p-value of 0.0005781. Using the same significance level, my results stay the same.

Exercise 3

a) This is a two sided test, since we care about differences that are both positive and negative.

$$H_0 : \mu = 40 \quad H_a : \mu \neq 40$$

b) I ran the following code.

```
mu <- mean(flint$Cu)
s <- sd(flint$Cu)
```

This output the sample mean $\bar{x} = 54.58102$ and the sample standard deviation $s = 133.3042$.

c)

$$\sigma_{\bar{x}} = \frac{133.3042}{\sqrt{541}} = 5.731195$$

d) I calculate the test statistic using the following formula.

$$t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = 2.54415$$

Then I ran the following code.

```
2 * pt(2.54415, df = nrow(flint) - 1, lower.tail = FALSE)
```

I multiplied by two since this is a two sided test. This output a p-value of 0.01123182.

e) Using a significance level of 0.01, I do not reject the null hypothesis. This means that the average copper levels in Flint could be the same as the average copper levels in Michigan.

f) I ran the following code.

```
t.test(flint$Cu, mu = 40, alt = "two.sided")
```

This output a p-value of 0.01123. My results would not change and I would still fail to reject the null hypothesis.