Brian Tehrani

3-11-19

UID: 604715464

Stats 12 Lab 5 Submission

1. Exercise 1
   a. $H_o: p_o = 0.10\ \&\ H_a: p_a > 0.10$: This is a one-sided test because we are checking for a proportion that is greater.

```
> ## b)
> # proportion of dangerous lead levels in flint
> sample_p_hat_flint <- mean(flint$Pb >= 15)
> sample_p_hat_flint
[1] 0.1238447
>
> # sd of flint lead levels
> sd_p_hat_flint <- sd(flint$Pb >= 15)
> sd_p_hat_flint
[1] 0.3297092
```

   b.

```
> ## c)
> # Standard Error SE = sqrt(p_o(1-p_o)/n)
> SE <- sqrt((p_o*(1-p_o))/n_1)
> SE
[1] 0.01289801
> #z-value for this test
> z_score <- (sample_p_hat_flint - p_o) / SE
> z_score
[1] 1.848714
```

   c.

```
> ## d)
> p_value <- 1 - pnorm(z_score)
> p_value
[1] 0.03224953
```

   d.
   e. Since the p-value is less than the significance level (0.03225 < 0.05), we have enough evidence to reject the null hypothesis.

```
> prop.test(x = sum(flint$Pb >= 15), n = n_1, p = p_o, alt = "greater" )

        1-sample proportions test with continuity correction

data:  sum(flint$Pb >= 15) out of n_1
X-squared = 3.1579, df = 1, p-value = 0.03778
alternative hypothesis: true p is greater than 0.1
95 percent confidence interval:
 0.101559 1.000000
sample estimates:
        p
0.1238447
```

   f.

```
> # No the p-value increases a little but is relatively close to part d).
> # h) for a confidence level of 99%
> prop.test(x = sum(flint$Pb >= 15), n = 541, p = 0.1, alt = "greater",
+           conf.level = 0.99)

        1-sample proportions test with continuity correction

data:  sum(flint$Pb >= 15) out of 541
X-squared = 3.1579, df = 1, p-value = 0.03778
alternative hypothesis: true p is greater than 0.1
99 percent confidence interval:
 0.09376523 1.00000000
sample estimates:
        p
0.1238447
```

   g.

2. Exercise 2

a. $H_o: p_1 = p_2$ & $H_a: p_1 \neq p_2$: This is a two-sided test.

b.

```
> # b)
> Conc_Pb_North <- flint$Pb[flint$Region == "North"]
> n1 <- length(flint$Pb[flint$Region == "North"])
> n1
[1] 261
>
>
> temp <- Conc_Pb_North >= 15
> success_p_hat_1 = 0
> for(i in 1:length(temp)){
+   if (temp[i] == TRUE){
+     success_p_hat_1 = success_p_hat_1 + 1
+   }
+ }
> #Conc_Pb_North
> p_hat_1 <- success_p_hat_1 / n1
> p_hat_1
[1] 0.1762452
>
> temp2 <- Conc_Pb_South >= 15
> success_p_hat_2 = 0
> for(i in 1:length(temp2)){
+   if (temp2[i] == TRUE){
+     success_p_hat_2 = success_p_hat_2 + 1
+   }
+ }
>
> Conc_Pb_South <- flint$Pb[flint$Region == "South"]
> n2 <- length(flint$Pb[flint$Region == "South"])
> n2
[1] 280
>
> #Conc_Pb_South
> p_hat_2 <- success_p_hat_2 / n2
> p_hat_2
[1] 0.075
>
> p_hat <- (success_p_hat_1 + success_p_hat_2)/(n1+n2)
> p_hat
[1] 0.1238447
>
> SE_two_prop <- sqrt(p_hat*(1-p_hat)*((1/n1)+(1/n2)))
> SE_two_prop
[1] 0.02834188
>
> z_score_2_prop <- (p_hat_1 - p_hat_2)/SE_two_prop
> z_score_2_prop
[1] 3.572283
```

c.

```
> # c) P-value
> p_value_2_prop <- 2*(1-pnorm(z_score_2_prop))
> p_value_2_prop
[1] 0.0003538831
```

d. Since the p-value is less than the test statistic (0.000353 < 0.05), we have enough evidence to reject the null hypothesis.

```
> prop.test(x = c(success_p_hat_1, success_p_hat_2), n = c(n1,n2), alt = "two
.sided")

        2-sample test for equality of proportions with continuity
        correction

data:  c(success_p_hat_1, success_p_hat_2) out of c(n1, n2)
X-squared = 11.845, df = 1, p-value = 0.0005781
alternative hypothesis: two.sided
95 percent confidence interval:
 0.04196839 0.16052203
sample estimates:
   prop 1    prop 2
0.1762452 0.0750000
```

e. Although the p-value is slightly higher, the results do not change.

3. Exercise 3
   a. $H_o: \mu = 40$ & $H_a: \mu \neq 40$: This is a two-sided test because we are finding a difference in Cu levels.

   b.
```
> # b) sample mean and sample sd of Cu levels
> samp_mean_Cu <- mean(flint$Cu)
> samp_mean_Cu
[1] 54.58102
> samp_sd_Cu <- sd(flint$Cu)
> samp_sd_Cu
[1] 133.3042
```

   c.
```
> # c) SE for sample mean for Cu
> SE_Cu <- samp_sd_Cu/sqrt(n_2)
> SE_Cu
[1] 5.731197
```

   d.
```
> test_stat <- (samp_mean_Cu - 40)/ SE_Cu
> test_stat
[1] 2.54415
> p_val <- 2*(1-pt(test_stat, df = n_2-1))
> p_val
[1] 0.01123183
```

   e. Since the p-value is greater that the significance level (0.1123 > 0.01), we fail to reject the null hypothesis and believe the average copper levels in Michigan is 40 ppm.

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

        One Sample t-test

data:  flint$Cu
t = 2.5441, df = 540, p-value = 0.01123
alternative hypothesis: true mean is not equal to 40
95 percent confidence interval:
 43.32285 65.83920
sample estimates:
mean of x
 54.58102

> # The results do not change if the significance level is 0.01, however
> # if the significance level is 0.05 we have enough evidence to reject
> # the null hypothesis (0.01123 < 0.05).
```