Assignment 3

Kyle Ligon

2018-10-16

Problem A

Could we say that the probability of a Texas Tech law school graduate passing the law exam is higher than the state average, which is 70?

1. The hypotheses H_0, H_1, and The level of significance α chosen

   $H_0 : \text{p} \leq 0.70$

   $H_1 : \text{p} > 0.70$

   $\alpha = 0.05$

2. The test statistic T_{Obs}

```
ttech <- c(0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
```

   $T_{Obs} = 18$

3. The critical values (region of rejection)

   Because this is an upper tailed test and n = 20, we'll use the Table A3 value for T.

   Table A3 states a value of n = 20 and p = 0.7 has a Critical Value of between 16 and 17. We will be conservative here, and say if we have a count larger than 17, we can reject he null hypothesis.

4. The p-value

```
library(tidyverse)
library(broom)
r <- binom.test(x = sum(ttech), n = length(ttech),
                p = 0.7, alternative = "greater") %>%
  tidy()
```

   The p-value of the test is 0.0355.

5. The decision rule and the decision.

   With a p-value less than 0.05 and a T_{Obs} greater than 17, we have sufficient evidence to reject the null hypothesis that the Texas Tech graduating class is less than the national average. Instead, we have sufficient evidence to conclude that the Texas Tech's graduating class outperformed the national average.

Problem B

Test the hypothesis that no more than 10% of the cars in the population are unsafe. (Which assumption is more likely to be false in this application?). Consider a level of significance of 5%. Clearly state:

1. The hypotheses H_0, H_1

   $H_0 : \text{p} \geq 0.10$
   $H_1 : \text{p} < 0.10$
   $\alpha = 0.05$

2. The test statistic T_{Obs}

```
cars <- c(0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1)
```

   $T_{Obs} = 6$

3. The critical values (region of rejection)
   Table A3 puts the critical value at 0.

4. The p-value, and

```
t <- binom.test(x = sum(cars), n = length(cars),
                p = 0.10, alternative = "less") %>%
  tidy()
```

   The p-value is 0.9995

5. The decision rule and the decision

   With a p-value larger than 0.05, we cannot reject the null hypothesis that the percent of cars that are unsafe is larger than 0.10. There is evidence conflicting with the idea that the percent of unsafe cars is under 10%.

Problem C

Find a 90% and 95% confidence interval for the probability of a torpedo hitting the target.

1. Use the exact probabilities (Table A4 form Conovers book) to solve this problem.

   According to the table in the back, n = 20, and y = 15, our low bound on the 95% C.I. is 0.509 with a higher boung of 0.913. The 90% C.I. has bounds of 0.544 and 0.896.

2. Use the large sample approximation to solve this problem.

```
prop = 15 / 20
error = sqrt(15*(20-15)/(20^3))
z_score = qnorm(0.975)
lower_bound = round(prop - z_score * error, 4)
upper_bound = round(prop + z_score * error, 4)
```

```
z_score1 = qnorm(0.95)
lower_bound1 = round(prop - z_score1 * error, 4)
upper_bound1 = round(prop + z_score1 * error, 4)
```

The lower and upper bounds baased on the 95% Normal approximation is 0.5602 and 0.9398. The bounds based on the 90% Normal Approximation is 0.5907 and 0.9093.

3. Discuss the assumptions we are making in this problem.

We are making two assumptions in this problem:

- The n trials are mutually independent.
- The probability p of the specified event occurring remains constant from one trial to the next.

With the probability staying consistent from trial to trial, I would certainly see an issue. Unless this is target practice with an object that is not able to be destroyed or maligned by the torpedos, the probability will be fluctuating from attempt to attempt.

Problem D

A random sample of n=20 is defined as follows: 142, 134, 98, 119, 131, 103, 154, 122, 93, 137, 86, 119, 161, 144, 158, 165, 81, 117, 128, 103.

1. Test the hypothesis that the median is 103.

A) Hypotheses:

$H_0 : \_M\_ = 103^1$
$H_1 : M \neq 103^2$
$\alpha = 0.05$

[1] The 50th quantile of X is 103
[2] The 50th quantile of X is not 103

B) Test Statistic

```
sample <- c(142, 134, 98,
            119, 131, 103,
            154, 122, 93,
            137, 86, 119,
            161, 144, 158,
            165, 81, 117,
            128, 103)
T_1 = 0
T_2 = 0
for(i in 1:length(sample)){
```

```
   if(sample[i] < 103) T_1 = T_1 + 1
}


for(i in 1:length(sample)){
   if(sample[i] > 103) T_2 = T_2 + 1
}
```

C) Critical Region/P-Value

According to Table A3, $t_1 = 5$ and $t_2 = 14$.

D) Decision

With $T_1 = 4 < 5$ and $T_2 = 14$, we can reject the the null hypothesis that the M equals 103. We have sufficient evidence to say the $M \neq 103$.

2. Test the hypothesis that the upper quartile is at least 150.

A) Hypotheses:

$H_0 : Q_3 \leq 150^3$

$H_1 : Q_3 > 150^4$

$\alpha = 0.05$

[3] The 75th quantile of X is 150

[4] The 75th quantile of X is not 150

B) Test Statistic

```
sample <- c(142, 134, 98,
            119, 131, 103,
            154, 122, 93,
            137, 86, 119,
            161, 144, 158,
            165, 81, 117,
            128, 103)
T_1 = 0
T_2 = 0
for(i in 1:length(sample)){
   if(sample[i] < 150) T_1 = T_1 + 1
}


for(i in 1:length(sample)){
   if(sample[i] > 150) T_2 = T_2 + 1
}
```

C) Critical Region/P-Value

According to Table A3, $t_2 = 10$.

D) Decision

With $T_2 = 4 < 10 = t_2$, we cannot reject the the null hypothesis that the $Q_3 \leq 150$. We do not have sufficient evidence to say the $Q_3 > 150$.

Problem E

The reaction time before lunch was compared with the reaction time after lunch for a group of 28 office workers. Twenty-two workers found their reaction time before lunch was shorter, and two could detect no difference. Is the reaction time after lunch significantly longer that the reaction time before lunch?

With 28 office workers participating, 22 found their reaction time to be shorter after lunch, 2 couldn't detect a difference. This leaves 4 office workers who saw their time lengthen after lunch. If the probability of longer reaction time is greater than 0.5, then we can reject the null hypothesis and say it's a significant difference.

1. Hypotheses:

   H_0: $p_{Longer} \leq 0.5$
   H_1: $p_{Longer} > 0.5$
   $\alpha = 0.05$

2. Test Statistic:

```
n = 28
p_star = 0.5
q_star = 1 - p_star
z_score = qnorm(0.975)

t_obs = 4
```

   The test statistic is 4.

3. P-Value

```
crit_score = round((t_obs - n * p_star - 0.5)/sqrt(n*p_star*q_star), 4)
```

   The critical value is -3.9686
   The p-value is 1

4. Decision

With a p-value > 0.05, we don't have enough evidence to reject the null hypothesis that the response time is shorter or equal after lunch. There does not seem to be enough evidence to support the claim that response times are longer.

Problem F

What must the sample size be to be 90% sure that at least 95% of the population lies within the sample range?

1. Use the exact table

According to the Table A5, we should plan on a sample of size 29 for this scenario.

2. Use the approximation

```
r = 1
m = 1
q = 0.90
chi_sq = qchisq(p = 0.95, df = 2*(r + m))

n = ceiling((0.25)*chi_sq*((1 + q)/(1-q)) + (0.5)*(r + m - 1))
```

With the approximation, we use the ceiling function to be more conservative in the case of rounding our decimals. That being said, the approximation has 46 as the number for the sample size.

Problem G

A manager writing the acceptance specifications for a load of steel reinforcement rods would like to specify that at least 90% of the rods are between the sixth longest and the sixth shortest rodsin a random sample she selects. In order to have 95% confidence in this statement, what should the sample size be?

```
r = 6
m = 6
q = 0.90
chi_sq = qchisq(p = 0.95, df = 2*(r + m))

n = ceiling((0.25)*chi_sq*((1 + q)/(1-q)) + (0.5)*(r + m - 1))
```

Again, using the ceiling function to be conservative in the case of rounding decimals, the approzimation has us with 179 as the sample size we should be striving for.