Weekly Tutorial Activity 8

Sean Leggett BDA 201
Winter 2020 - March 28, 2020

Questions

1) We want to test with 95 percent confidence interval whether the volume of a shipment of lumber is as usual (mu=39000 cubic feet). Use rnorm function in R to to generate 75 shipments with mean: 36500 and sd: 2000. Use set.seed(0) before rnorm to regenerate the same data if required. On the simulated data test Ho: mu = 39000

2) The CEO of a large electric utility claims that 79 percent of his 1,000,000 customers are very satisfied with the service they receive. To test this claim, the local newspaper surveyed 100 customers, using simple random sampling. Among the sampled customers, 73 percent say they are very satisfied. Based on these findings, can we reject the CEO's hypothesis that 79% of the customers are very satisfied? Use a 0.04 level of significance.

3) The CEO changed the claim and now says that at least 79% are satisfied, keeping everything else the same, how does this change your analysis?

Answers

Question 1)

Generate test dataset

```
set.seed(0)
shipments <- rnorm(75, 36500, 2000)
shipments
```

```
##  [1] 39025.91 35847.53 39159.60 39044.86 37329.28 33420.10 34642.87
##  [8] 35910.56 36488.47 41309.31 38027.19 34901.98 34204.69 35921.08
## [15] 35901.57 35676.98 37004.45 34716.16 37371.37 34024.92 36051.46
## [22] 37254.79 36766.67 38108.38 36385.79 37507.22 38671.54 35118.09
## [29] 33930.80 36593.45 36028.59 35414.22 35633.38 35201.06 37953.50
## [36] 38803.82 38484.32 35640.97 38976.61 35941.31 40015.81 37621.49
## [43] 35594.43 34835.91 34166.86 34368.82 33372.44 38813.07 38164.09
## [50] 36045.34 37032.27 35746.59 41382.73 34909.32 36390.25 37000.28
## [57] 37736.49 36154.75 32052.20 33972.77 37217.46 36477.91 34618.70
## [64] 36268.35 34870.06 36984.53 33649.80 37231.88 36996.83 36630.58
## [71] 36538.31 37014.68 35201.98 36261.66 37828.27
```

H_0: shipment = 39,000 cu ft H_a: shipment not equal to 39,000 cu ft

Calculate the test statistic

```
n = 75
mu = 39000
s = 2000
smean <- 36500
```

```r
z_score <- (mu - smean) / (s/sqrt(n))
z_score
```

```
## [1] 10.82532
```

A z-score this far off center is hugely unlikely. We can also test for p-value:

```r
pnorm(z_score, lower.tail = FALSE)
```

```
## [1] 1.305861e-27
```

```r
1 - pnorm(z_score)
```

```
## [1] 0
```

Accordingly, we can reject the null hypothesis.

Question 2)

H_0: very satisfied $= 79\%$ H_a: very satisfied $< 79\%$

In this case based on the provided data, we are looking to establish the confidence interval for these proportions.

```r
nn <- 100
pp <- 73/100
qq <- 1-pp
pp - qnorm(.98) * sqrt(pp*qq/nn)
```

```
## [1] 0.6388219
```

```r
pp + qnorm(.98) * sqrt(pp*qq/nn)
```

```
## [1] 0.8211781
```

Which tells us from the television station sample, that their true random mean lies between 64% and 82%, or can be stated as 73% +/- 9%. The executive's claim of 79% very statisfied is on the higher side of this bound but certainly is within tolerance.

Question 3)
We won't perform any math on this change of claim. If the executive would like to change the test and objective of the hypothesis, we can start with a newly commissioned exercise. This is simply gaming the system.

Now this becomes a marketing exercise trying to find the highest percentage that can be attached to a usable marketing term.

However, we can say that reducing the quality factor will likely provide an even more accurate range. We are suspicious of their motives.