Stats Exploration HW 2

Shalini Mishra

10/8/2019

R Markdown

Question 1

- $H_0 : \mu \geq 22$
- $H_a : \mu < 22$

```
#sample size
n <- 220
q1_data <- read.csv('Question1.csv')
head(q1_data)
```

```
##   Payment
## 1      27
## 2      24
## 3      14
## 4      39
## 5      13
## 6      31
```

```
t.test(q1_data$Payment,alternative="less",mu=22)
```

```
##
##  One Sample t-test
##
## data:  q1_data$Payment
## t = -0.93585, df = 219, p-value = 0.1752
## alternative hypothesis: true mean is less than 22
## 95 percent confidence interval:
##      -Inf 22.28168
## sample estimates:
## mean of x
##  21.63182
```

```
#type 1 more serious
```

As p-value=0.1752>0.05, we failed to reject the null hypothesis. Hence, the CFO can't conclude that the system is profitable.

At 5% significance level, there is not sufficient evidence to support the claim that the proposal of CFO to use SSA (self addressed stamped) envelopes is profitable for the company.

Question 2

- $H_0 : p_r \leq 0.5$
- $H_a : p_r > 0.5$

```
q2_data <- read.csv("Question2.csv")
#(q2_data)
n <- nrow(q2_data)
num_dem <- length(which(q2_data == 1)) #code 1 for democrat votes
num_rep <- length(which(q2_data == 2)) #code 2 for republican votes
prop.test(num_rep, n, alternative = "greater", p = 0.5, conf.level=0.95)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  num_rep out of n, null probability 0.5
## X-squared = 3.0118, df = 1, p-value = 0.04133
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.5016378 1.0000000
## sample estimates:
##         p
## 0.5320261
```

```
#reject null hypothesis
```

At 5% significance level, there is sufficient evidence to support the claim that Republicans will win the elections. Based on the estimation, the network should announce at 8:01 p.m. that the Republican candidate will win.

Question 3
Let μ be the weekly spent average for the town.

- $H_0 : \mu = 150$
- $H_a : \mu \neq 150$

a

```
q3_data <- read.csv("Question3.csv")
#head(q3_data)
#At a significance level of 5%
t.test(q3_data$Weekly_food_expense,alternative="two.sided",mu=150)
```

```
##
##  One Sample t-test
##
## data:  q3_data$Weekly_food_expense
## t = 3.8989, df = 99, p-value = 0.000176
## alternative hypothesis: true mean is not equal to 150
## 95 percent confidence interval:
##  154.5829 164.0819
## sample estimates:
## mean of x
##  159.3324
```

At 5% significance level, p value (say p) =0.000176 - p<0.10 (10% significance) => Reject the null Hypothesis, H_0 - p<0.05 (5% significance) => Reject the null Hypothesis, H_0 - p<0.01 (1% significance) => Reject the null Hypothesis, H_0

Yes, the sample evidence is statistically significant. We can reject the null hypotheses From 1% to 10% significance level, there is sufficient evidence to support the claim that the weekly spent average for the town is different than the national average.

b

```
#1% significance level
t.test(q3_data$Weekly_food_expense,alternative="two.sided",conf.level=0.99,mu=150)
```

```
##
##  One Sample t-test
##
## data:  q3_data$Weekly_food_expense
## t = 3.8989, df = 99, p-value = 0.000176
## alternative hypothesis: true mean is not equal to 150
## 99 percent confidence interval:
##  153.0458 165.6190
## sample estimates:
## mean of x
##  159.3324
```

```
#Critical values for 1% significance
a <- 150-(qt(0.995,nrow(q3_data)-1)*sd(q3_data$Weekly_food_expense)/(sqrt(nrow(q3_data))))
b <- 150+(qt(0.995,nrow(q3_data)-1)*sd(q3_data$Weekly_food_expense)/(sqrt(nrow(q3_data))))
cat("Any value of sample mean below",a,"and any value above",b,"will reject the null hypothesis at 1% si
```

```
## Any value of sample mean below 143.7134 and any value above 156.2866 will reject the null hypothesis
```

```
#10% significance level
t.test(q3_data$Weekly_food_expense,alternative="two.sided",conf.level=0.90,mu=150)
```

```
##
##  One Sample t-test
##
## data:  q3_data$Weekly_food_expense
## t = 3.8989, df = 99, p-value = 0.000176
## alternative hypothesis: true mean is not equal to 150
## 90 percent confidence interval:
##  155.3581 163.3067
## sample estimates:
## mean of x
##  159.3324
```

```
#critical values for 10% significance
a <- 150-(qt(0.95,nrow(q3_data)-1)*sd(q3_data$Weekly_food_expense)/(sqrt(nrow(q3_data))))
b <- 150+(qt(0.95,nrow(q3_data)-1)*sd(q3_data$Weekly_food_expense)/(sqrt(nrow(q3_data))))
cat("Any value of sample mean below",a,"and any value above",b,"will reject the null hypothesis at 10% s
```

```
## Any value of sample mean below 146.0257 and any value above 153.9743 will reject the null hypothesis
```

Question 4

Let σ^2 be the variance of the daily demand of the product.

- $H_0 : \sigma^2 = 250$
- $H_a : \sigma^2 \neq 250$

```
q4_data <- read.csv("Question4.csv")
#head(q4_data)
n <- nrow(q4_data)
EnvStats::varTest(q4_data$Demand,alternative = "two.sided",sigma.squared=250)
```

```
##
##  Chi-Squared Test on Variance
##
## data:  q4_data$Demand
## Chi-Squared = 25.976, df = 24, p-value = 0.7088
## alternative hypothesis: true variance is not equal to 250
## 95 percent confidence interval:
##  164.9727 523.6611
## sample estimates:
## variance
## 270.5833
```

p-value=0.708 > 0.05 (5% significance) => Failed to reject H0 The data doesn't provide enough evidence at the 5% significance level to infer that the operations research analyst's assumption about the variance is wrong. At 5% significance level, there is not sufficient evidence to support the claim that variance is different than 250.

Question 5

- $H_0 : p_w \geq 0.14$
- $H_a : p_w < 0.14$

a.
From the alternate hypothesis, we can see it's a left tailed test. To determine the sample proportion values that will lead to the show's cancellation, assuming a 5% significance level, We will have to find the critical value

```
z_score <- qnorm(0.95)
p_hat <- 0.14
p_critical <- p_hat - z_score*(sqrt(p_hat*(1-p_hat)/1500))
p_critical
```

```
## [1] 0.1252635
```

the sample proportion values that will lead to the show's cancellation, assuming a 5% significance level are: Any sample proportion below 12.5%

Hence ,There is sufficient evidence at 5% significance level to support the claim that the show have viewership less than 14% to get it cancelled only when sample proportion < 12.5%

b.
My goal is to find the significance level which will give me the same rejection critical proportion as above

Using the same p-critical value ie. 12.5%, we will try to obtain the corresponding z-score A sample will reject the null hypothesis if the proportion of people watching it will be below 12.5% With a new population proportion of 13.4%, we will find the z-score

```r
z_score <- (0.125-0.134)/(sqrt((0.134*(1-0.134))/1500))

z_score
```

```
## [1] -1.023238
```

```r
pnorm(z_score,mean=0,sd=1,lower.tail=TRUE)
```

```
## [1] 0.1530976
```

The probability that a sample will lead to rejection of the null hypothesis is 15.3%

Question 6

- H_0 : Declare Emergency, fuel=0
- H_a : enough fuel to stay afloat for 15 mins, fuel>0

Error Analysis table is in the word

Question 7

- H_0 : the accused is innocent, not enough votes
- H_a : the accused is guilty, jury votes unanimously

a. According to the question, Evidence in the trial of an innocent suspect is enough to convince 42% of all jurors in the population that the suspect is guilty. Each of these 12 jurors are picked from the population. Probability that a juror is voting for an innocent suspect as guilty =p=0.42 As they are voting unanimously, I am considering each activity of voting as an event

Probability that a jury convicts an innocent suspect = p x p x p x. . . . . . .12 times = p12 = (0.42)12 ~ 0.00003

```r
prob_convict_when_innocent <- dbinom(12,12,0.42)
prob_convict_when_innocent
```

```
## [1] 3.012947e-05
```

b.
Refer to the error analysis table drawn for this question in the word file Type 1 error was committed by the Jury in part (a)

c.
Evidence in the trial of a guilty suspect is enough to convince 99% of all jurors in the population that the suspect is guilty. Probability of a juror voting that the suspect is guilty =0.99 Probability of the jury voting that the suspect is guilty(conviction) = 0.99 X 0.99 X . . . 12 times = 0.9912 = 0.88638487 ~ 0.886 Probability of the jury voting that the suspect is not guilty (acquitted) = 1 − Prob(conviction by jury) = 1-0.886 = 0.114

```r
prob_conv_when_guilty <- dbinom(12,12,0.99)
prob_acquit_when_guilty <- 1- dbinom(12,12,0.99)
prob_acquit_when_guilty
```

```
## [1] 0.1136151
```

d.
From the error analysis table drawn in part(b) We can conclude that the case in part (c) is a type 2 error.