

LAB-9

TWO SAMPLE Z-TEST

SIDDHI SINGH
17BIT0028

Aim:-To test the hypothesis for Large Samples by using Two-sample Z-test

R-Code:-

Tests about a proportion using x and n using the prop.test function:

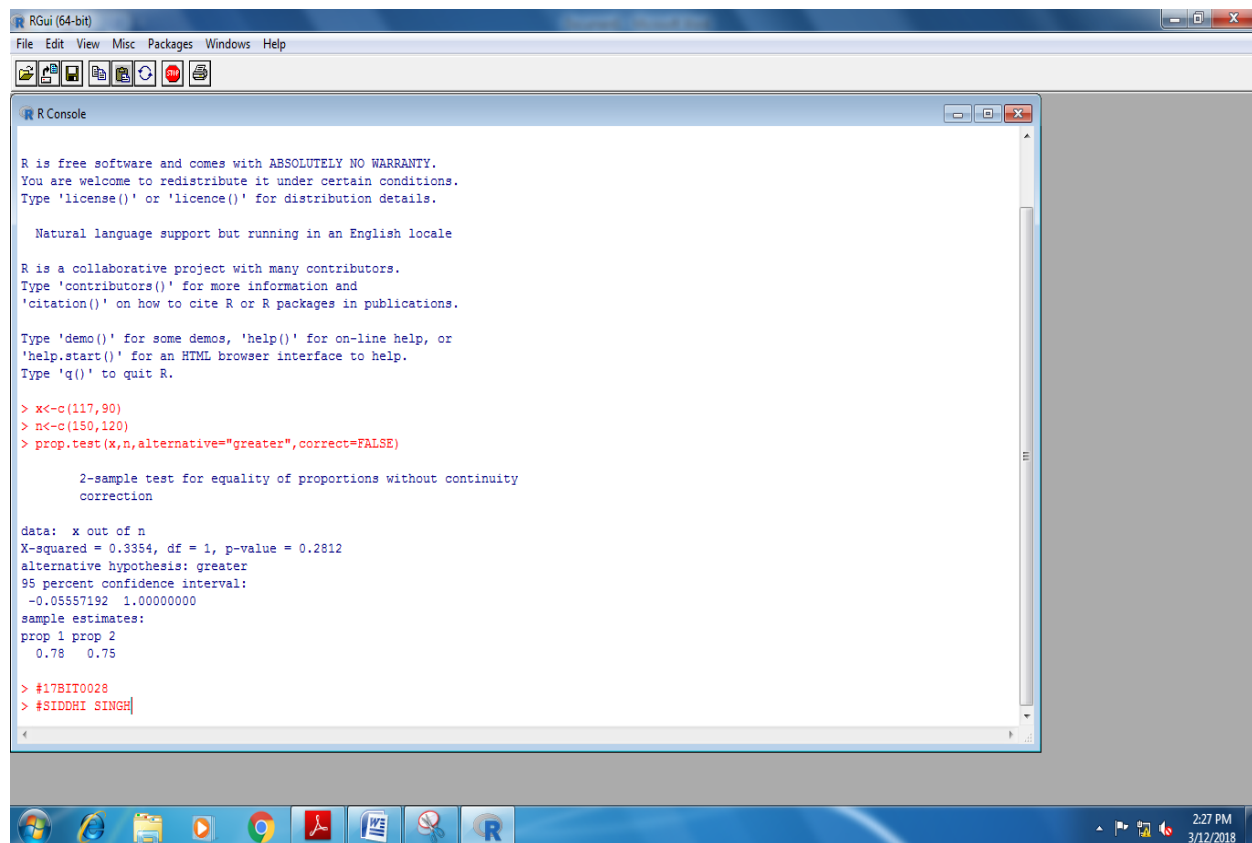
Usage: `prop.test(c(x1,x2), c(n1,n2), correct=, alternate =)`.

1. `x1` and `x2` are the number of successes in sample 1 and 2 respectively.
2. `n1` and `n2` are the sample sizes or number of trials.
3. `correct = TRUE` (use a continuity correction factor) or `FALSE` (do not).
4. `alternate = "two.sided"` (default), `"less"`, or `"greater"`.

Problem : A popular cold-remedy was tested for it's efficacy. In a sample of 150 people who took the remedy upon getting a cold, 117 (78%) had no symptoms one week later. In a sample of 125 people who took the placebo upon getting a cold, 90 (75%) had no symptoms one week later. The table summarizes this information.

<i>Group</i>	<i>#who are symptom Free after one week(x)</i>	<i>Total # in group (n)</i>	<i>Proportion $\hat{p} = x / n$</i>
<i>Remedy</i>	<i>117</i>	<i>150</i>	<i>0.78</i>
<i>Placebo</i>	<i>90</i>	<i>120</i>	<i>0.75</i>

The Test: Test the claim that the proportion of all remedy users who are symptom-free after one week is greater than the proportion for placebo users. Test this claim at the 0.05 significance level.



```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> x<-c(117,90)
> n<-c(150,120)
> prop.test(x,n,alternative="greater",correct=FALSE)

      2-sample test for equality of proportions without continuity
      correction

data:  x out of n
X-squared = 0.3354, df = 1, p-value = 0.2812
alternative hypothesis: greater
95 percent confidence interval:
 -0.05557192  1.00000000
sample estimates:
prop 1 prop 2 
 0.78  0.75 

> #17BIT0028
> #SIDDIHI SINGH
```

We fail reject the null hypothesis because the P-value (.2812) is greater than the significance level. Therefore, we can't support the claim.

Problem 2:-

The Trial Urban District Assessment (TUDA) is a study sponsored by the government of student achievement in large urban school district. In 2009, 1311 of a random sample of 1900 eighth-graders from Houston performed at or above the basic level in mathematics . In 2011, 1440 of a random sample of 2000 eighth-graders from Houston performed at or above the basic level . (The study reports the proportions).

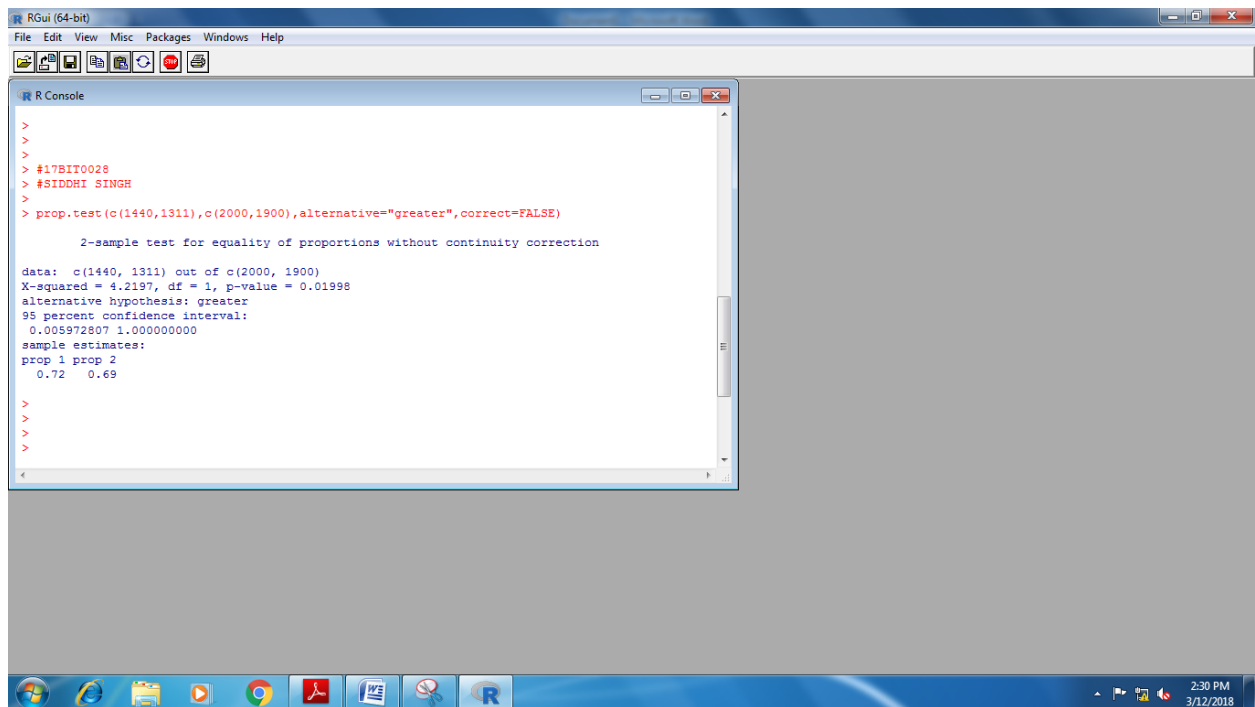
(A)Is there an increase in the proportion of eighth-graders who performed at or above the basic level in mathematics from 2009 to 2011 at the 5% significance level?

(B) Compute the 95% confidence interval for the difference in proportion of eighth-graders who performed at or above the basic level in mathematics from 2009 to 2011.

Solution:-

Let p_1 and p_2 be the proportions of eighth-graders that performed at or above the basic level in mathematics in 2011 and 2009, respectively .We wish the test

$$H_0=p_1=p_2 \text{ against } H_1=p_1> p_2$$



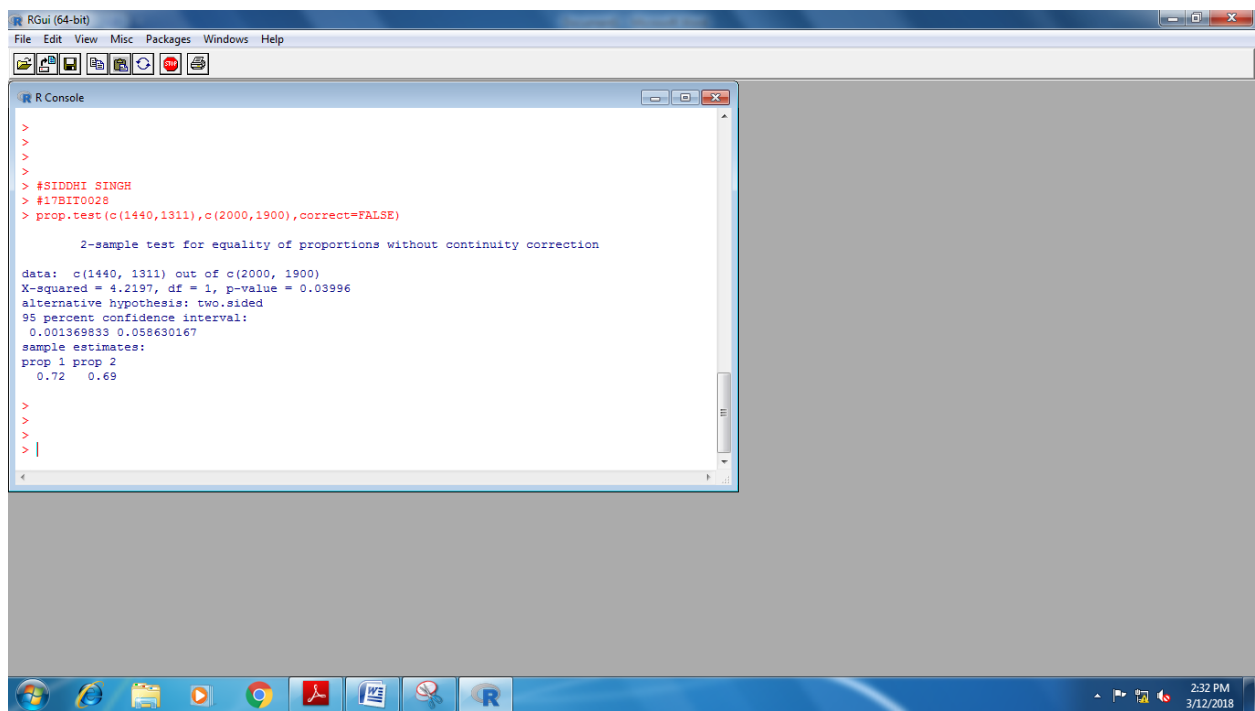
```
>
>
> #17BIT0028
> #SIDDIHI SINGH
> prop.test(c(1440,1311),c(2000,1900),alternative="greater",correct=FALSE)

      2-sample test for equality of proportions without continuity correction

data:  c(1440, 1311) out of c(2000, 1900)
X-squared = 4.2197, df = 1, p-value = 0.01998
alternative hypothesis: greater
95 percent confidence interval:
 0.005972897 1.000000000
sample estimates:
prop 1 prop 2
 0.72  0.69
>
>
>
```

The $p\text{-value}=0.02 < 0.05$ so we reject H_0 . Thus, there is evidence that there is an increase from 2009 to 2011 in the proportion of eighth-graders who performed at or above the basic level at the 5% significance level.

Solution to part (b). We obtain



```
>
>
> #SIDDIHI SINGH
> #17BIT0028
> prop.test(c(1440,1311),c(2000,1900),correct=FALSE)

      2-sample test for equality of proportions without continuity correction

data:  c(1440, 1311) out of c(2000, 1900)
X-squared = 4.2197, df = 1, p-value = 0.03996
alternative hypothesis: two.sided
95 percent confidence interval:
 0.001369833 0.058630167
sample estimates:
prop 1 prop 2
 0.72  0.69
>
>
> |
```

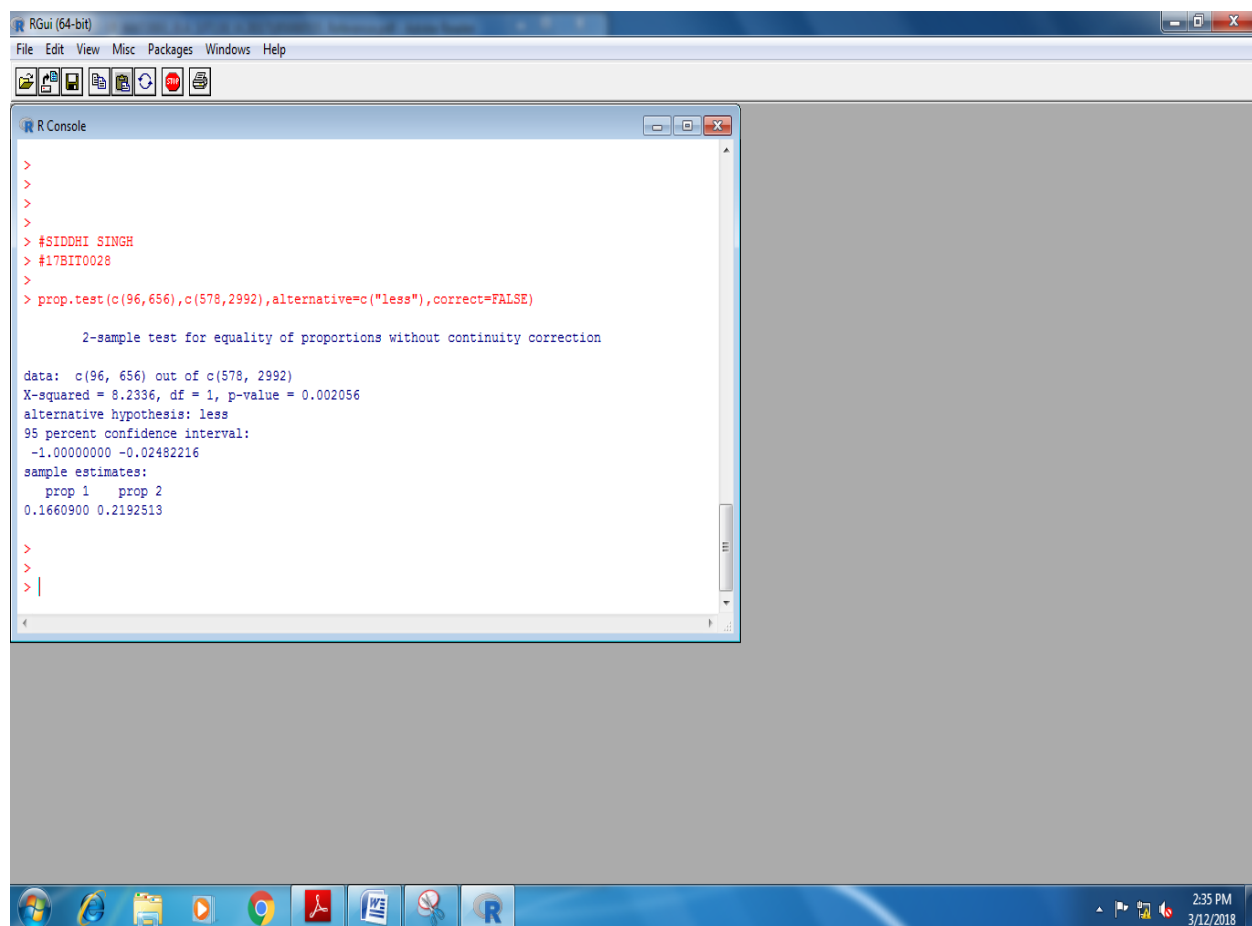
Thus, we are 95% confident that the percent of eighth-graders who performed at or above the basic level in mathematics in 2011 is between 0:14% and 5:86% higher than in 2009.

Problem. The use of helmet among recreational alpine skiers and snowboarders are generally low. A study from Norway wanted to examine if helmet use reduces the risk of head injury. In the study, they compared the helmet use among skiers and snowboarders that was injured with a control group. The control group consisted of skiers and snowboarders that was uninjured. 96 of 578 people with head injuries used a helmet and 656 of 2992 people in the uninjured group used a helmet. Is helmet use lower among skiers and snowboarders who had head injuries?

Solution : Let p_1 be the proportion of helmet use among injured skiers and snowboarders. Let p_2 be the proportion of helmet use among uninjured skiers and snowboarders.

We wish to test

$H_0 : p_1 = p_2$ against $H_a : p_1 < p_2$



The screenshot shows the RGui (64-bit) interface. The R Console window displays the following output:

```
>
>
>
>
> #SIDDHI SINGH
> #17BIT0028
>
> prop.test(c(96,656),c(578,2992),alternative="less",correct=FALSE)

      2-sample test for equality of proportions without continuity correction

data:  c(96, 656) out of c(578, 2992)
X-squared = 8.2336, df = 1, p-value = 0.002056
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000 -0.02482216
sample estimates:
   prop 1   prop 2 
0.1660900 0.2192513 
>
>
> |
```

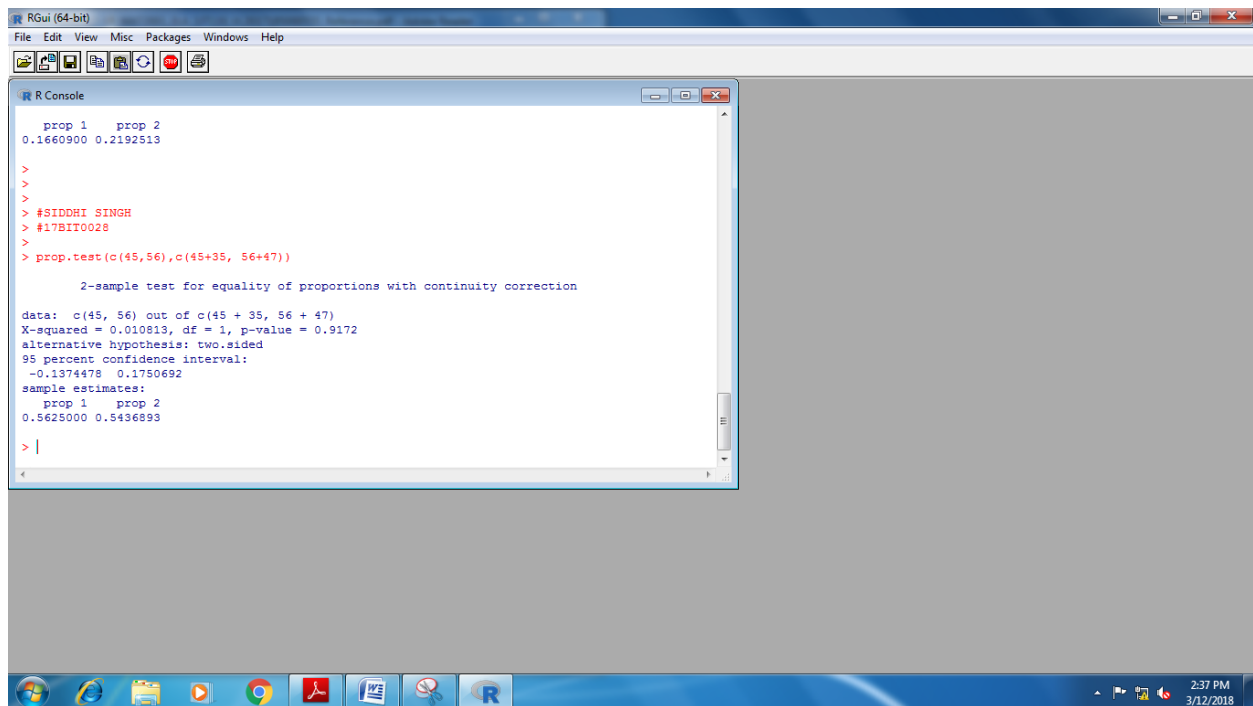
The Windows taskbar at the bottom shows the time as 2:35 PM on 3/12/2018.

The p-value= 0.0021 < 0.01 so we have strong evidence that helmet use is lower among skiers and snowboarders who had head injuries compared to uninjured skiers and snowboarders.

Problem : A survey is taken two times over the course of two weeks. The pollsters wish to see if there is a difference in the results as there has been a new advertising campaign run. Here is the data

	<i>Week1</i>	<i>Week2</i>
<i>Favorable</i>	45	56
<i>Unfavorable</i>	35	47

The standard hypothesis test is $H_0: P_1 = P_2$ against the alternative (two-sided) $H_1: P_1 \neq P_2$. The function `prop.test` is used to being called as `prop.test(x,n)` where x is the number favorable and n is the total. Here it is no different, but since there are two x 's it looks slightly different. Here is how



```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console
> prop.test(c(45,56),c(45+35, 56+47))
2-sample test for equality of proportions with continuity correction
data:  c(45, 56) out of c(45 + 35, 56 + 47)
X-squared = 0.010813, df = 1, p-value = 0.9172
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1374478  0.1750692
sample estimates:
 prop 1    prop 2 
0.5625000 0.5436893 
> |
```

we observe that the p-value is 0.9172 so we accept the null hypothesis that $P_1 = P_2$.

Problem

In the built-in data set named *quine*, children from an Australian town is classified by ethnic background, gender, age, learning status and the number of days absent from school.

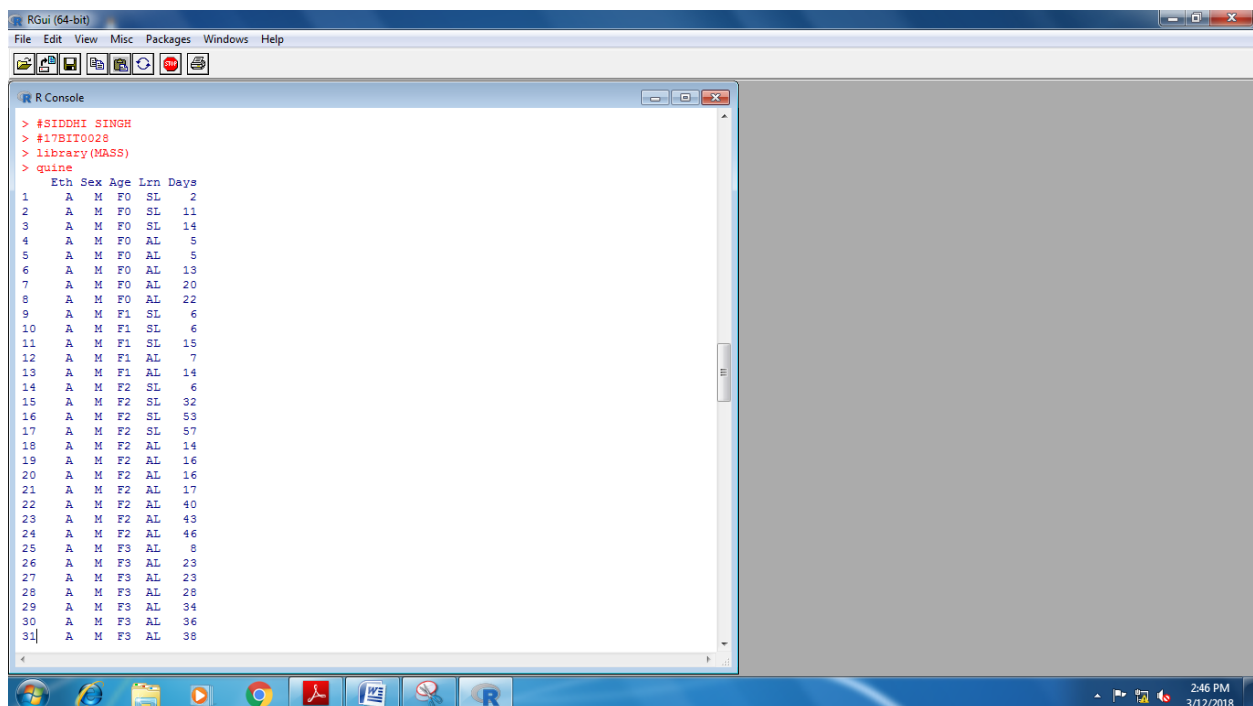
```
> library(MASS)      # load the MASS package
```

```
> quine
```

```
> table(quine$Eth, quine$Sex)
```

	F	M
A	38	31
N	42	35

Assuming that the data in *quine* follows the normal distribution, find the 95% confidence interval estimate of the difference between the female proportion of Aboriginal students and the female proportion of Non-Aboriginal students, each within their own ethnic group.



```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console
> #SIDDIHI SINGH
> #17BIT0028
> library(MASS)
> quine
  Eth Sex Age Lrn Days
1   A  M  F0  SL    2
2   A  M  F0  SL   11
3   A  M  F0  SL   14
4   A  M  F0  AL    5
5   A  M  F0  AL    5
6   A  M  F0  AL   13
7   A  M  F0  AL   20
8   A  M  F0  AL   22
9   A  M  F1  SL    6
10  A  M  F1  SL    6
11  A  M  F1  SL   15
12  A  M  F1  AL    7
13  A  M  F1  AL   14
14  A  M  F2  SL    6
15  A  M  F2  SL   32
16  A  M  F2  SL   53
17  A  M  F2  SL   57
18  A  M  F2  AL   14
19  A  M  F2  AL   16
20  A  M  F2  AL   16
21  A  M  F2  AL   17
22  A  M  F2  AL   40
23  A  M  F2  AL   43
24  A  M  F2  AL   46
25  A  M  F3  AL    8
26  A  M  F3  AL   23
27  A  M  F3  AL   23
28  A  M  F3  AL   28
29  A  M  F3  AL   34
30  A  M  F3  AL   36
31  A  M  F3  AL   38
```

```

RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console
136 N F F2 AL 1
137 N F F3 AL 1
138 N F F3 AL 9
139 N F F3 AL 22
140 N F F3 AL 3
141 N F F3 AL 3
142 N F F3 AL 5
143 N F F3 AL 15
144 N F F3 AL 18
145 N F F3 AL 22
146 N F F3 AL 37
> table(quine$Eth, quine$Sex)
Error: unexpected ')' in "table(quine$Eth, quine$Sex)"
> table(quine$Eth, quine$Sex)

  F M
A 38 31
N 42 35
>
> prop.test(table(quine$Eth, quine$Sex), correct=FALSE)
Error in table(quine$Eth, quine$Sex) :
  all arguments must have the same length
> prop.test(table(quine$Eth, quine$Sex), correct=FALSE)

2-sample test for equality of proportions without continuity correction

data: table(quine$Eth, quine$Sex)
X-squared = 0.0040803, df = 1, p-value = 0.9491
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1564218  0.1669620
sample estimates:
 prop 1    prop 2 
0.5507246 0.5454545 
>

```

Conclusion:

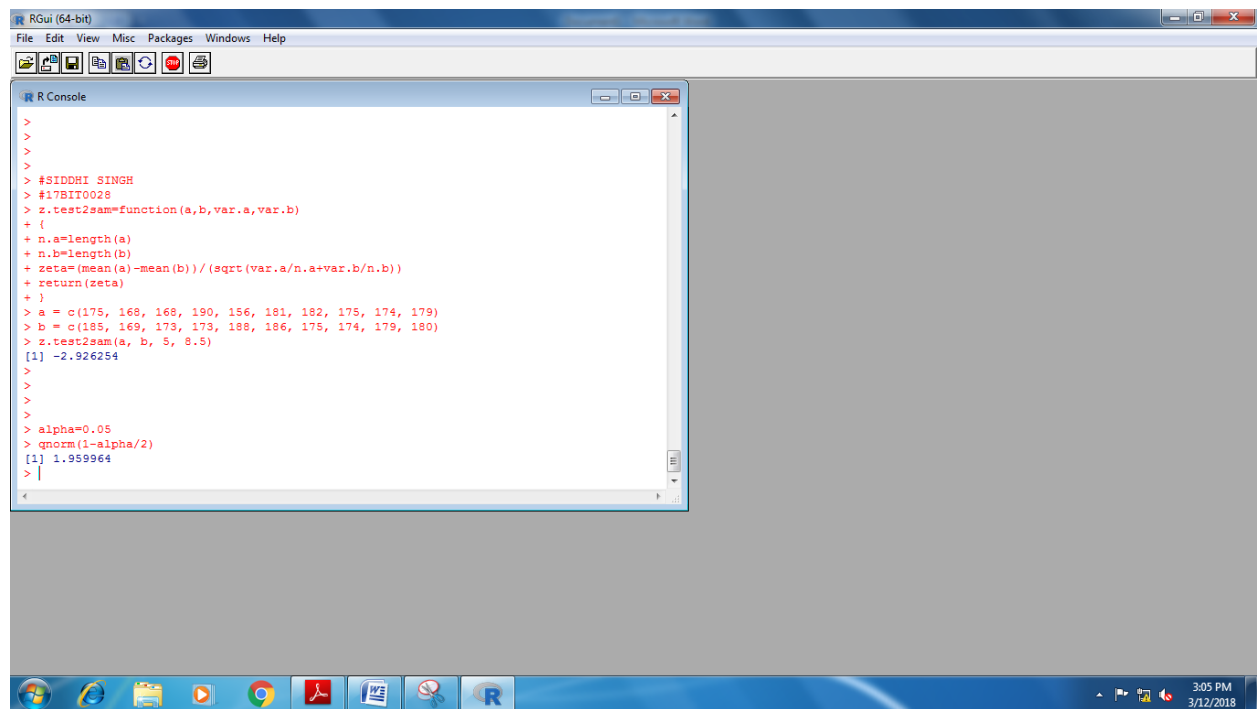
The 95% confidence interval estimate of the difference between the female proportion of Aboriginal students and the female proportion of Non-Aboriginal students is between -15.6% and 16.7%.

Comparison of two independent sample means, taken from two populations with known variance

Problem: The following data shows the heights of individuals of two different countries with the population variance of 5 and 8.5 respectively. Is there any significant difference between the average heights of two groups.

A: 175	168	168	190	156	181	182	175	174	179
B: 185	169	173	173	188	186	175	174	179	180

Since we have the variance of the population, we must proceed with a two sample Z-test.



```
>
>
>
> #SIDDIHI SINGH
> #17BIT0028
> z.test2sam=function(a,b,var.a,var.b)
+ {
+   n.a=length(a)
+   n.b=length(b)
+   zeta=(mean(a)-mean(b))/(sqrt(var.a/n.a+var.b/n.b))
+   return(zeta)
+ }
> a = c(175, 168, 168, 180, 156, 181, 182, 175, 174, 179)
> b = c(185, 169, 173, 173, 188, 186, 175, 174, 179, 180)
> z.test2sam(a, b, 5, 8.5)
[1] -2.926254
>
>
>
> alpha=0.05
> qnorm(1-alpha/2)
[1] 1.959964
>
```

Inference :-

The value of z is greater than the critical value of z tabulated for alpha equal to 0.05 (z-tabulated = 1.96 for a two-tailed test): we reject the null hypothesis and conclude that the two means are significantly different.

CAB-9

Lab-9 — 12/13/18 — A

17BIT0028

```
> x <- c(117, 190)
> u <- c(150, 120)
> prop.test(x, u, alternative = "greater", correct = FALSE)

> prop.test(c(1440, 1311), c(2000, 1900), alternative =
  "greater", correct = FALSE)

> prop.test(c(1440, 1311), c(2000, 1900), correct = FALSE)

> prop.test(c(96, 656), c(578, 2992), alternative = c("less"),
  correct = FALSE)

> prop.test(c(45, 56), c(45+35, 56+47))

> library(MASS)
> quine
> table(quine$eth, quine$sex)
> prop.test(table(quine$eth, quine$sex), correct =
  FALSE)

> alpha = 0.5
> qnorm(1 - alpha/2)

[1] 1.959964
```