

ASSIGNMENT-2 STATISTICAL INFERENCE

Sourav Goyal (2020341)

Ans - 1

Date _____
DELTA Pg No. _____

Assignment - 2

Ans - 1 Hypothesis :-

→ Let μ be population mean,
 $H_0: \mu \leq 15 \text{ cm}$ v/s $H_1: \mu > 15 \text{ cm}$

2) Information given :- Baker bakes 10 loaves of bread,

- Baker claims that his average bread height is more than 15 cm so we take this in alternative hypothesis.
- Mean height of 10 loaves = 17 cm
- Baker knows from baking hundreds of loaves of bread that the standard deviation for the height is 0.5 cm.
- Level of significance = 5%.

→ So, popⁿ var is known, distribution is normal and 10 loaves selected and each loaves have equal chances so SRS.

→ Assumptions met for 1-Sample test for mean with σ known.

→ Test statistics :- $Z_{\text{test}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

→ From Information given :- $\bar{X} = 17 \text{ cm}$, $\sigma = 0.5 \text{ cm}$, $n = 10$,
 $\mu = 15 \text{ cm}$.

∴ $Z_{\text{test}} = \frac{17 - 15}{0.5} \times \sqrt{10} = 12.6491 \sim N(0,1)$

→ Now, from the Z-table $2_{0.05} = 1.6448 = Z_{\text{cv}}$

∴ Graph will be like :-

(1-tailed)

Decision :- So by critical value approach our Zstat lies in the rejection region therefore we reject null so average bread height is greater than 15 cm which means baker's claim is true.

Result from R :-

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for a hypothesis test. The code calculates a z-test statistic, compares it to a critical value, and prints the results.
- Environment Pane:** Shows the values of variables created in the script.
- Console:** Displays the output of the R script execution.

R Code (Source Editor):

```
9
10 n <- 10
11 sample_mean <- 17
12 pop_mean <- 15
13 pop_sd <- 0.5
14 alpha <- 0.05
15 z_test_st <- (sample_mean - pop_mean)/(pop_sd/sqrt(n))
16 # This is 1-tailed test and critical value will lie in right.
17 critical_val <- qnorm(p=alpha,0,1,lower.tail = FALSE)
18 p_val <- pnorm(z_test_st,0,1,lower.tail = FALSE)
19 cat("Test Statistic Value",z_test_st,"\n")
20 cat("Critical Value",critical_val,"\n")
21 cat("P value: ",p_val,"\n")
22
23 # Conclusion : So p_val < alpha or z_test_st value is lies in right of
24 # critical value so this shows that we reject null which means pop_mean <= 15 cm so
25 # baker's claim is true i.e., pop_mean > 15 cm.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
```

Environment Pane (Values):

Variable	Value
alpha	0.05
critical_val	1.64485362695147
n	10
p_val	5.65741895121656e-37
pop_mean	15
pop_sd	0.5
sample_mean	17
z_test_st	12.6491106406735

Console Output:

```
R 4.2.1 - D:/SouravGoyal_2020341/ >
> rm(list=ls())
> n <- 10
> sample_mean <- 17
> pop_mean <- 15
> pop_sd <- 0.5
> alpha <- 0.05
> z_test_st <- (sample_mean - pop_mean)/(pop_sd/sqrt(n))
> # This is 1-tailed test and critical value will lie in right.
> critical_val <- qnorm(p=alpha,0,1,lower.tail = FALSE)
> p_val <- pnorm(z_test_st,0,1,lower.tail = FALSE)
> cat("Test Statistic Value",z_test_st,"\n")
Test Statistic Value 12.64911
> cat("Critical Value",critical_val,"\n")
Critical Value 1.644854
> cat("P value: ",p_val,"\n")
P value: 5.657419e-37
> # Conclusion : So p_val < alpha or z_test_st value is lies in right of
> # critical value so this shows that we reject null which means pop_mean <= 15 cm so
> # baker's claim is true i.e., pop_mean > 15 cm.
>
```

$$Z_{\text{stat}} = 12.64911$$

$$Z_{\text{cv}} = 1.644854$$

$$P \text{ value} = 5.67 \times 10^{-37}$$

So, From R also we verify our results.

Ans - 2

Ans - 2 Hypothesis :-

$H_0: \mu = 15 \text{ years}$ v/s $H_1: \mu \neq 15 \text{ years}$

1) (2-tailed test)

2) Information given :- Random survey of 75 death row inmates.
• Mean length of time on death row is 17.4 years.
• Sample standard deviation of 6.3 years.
• Hypothesis test to determine if the population mean time on death row could likely be 15 years.

3) Level of significance = 5 percent $\Rightarrow \alpha = 0.05$.

4) So, again popⁿ var is unknown, sample is random and sample size > 30 .

\therefore Assumptions met for 1-sample test for mean with σ unknown.

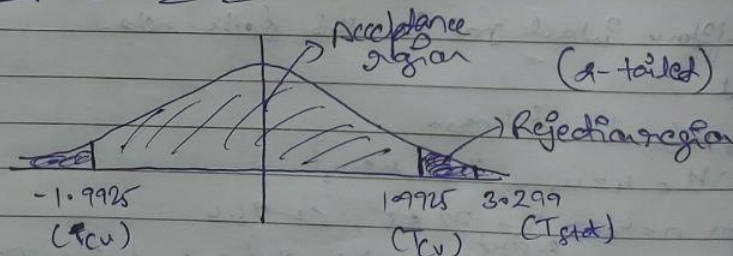
\therefore Test statistics :- $T_{\text{stat}} = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{(n-1)}$

5) From Information given :- $\bar{X} = 17.4 \text{ years}$, $\mu = 15 \text{ years}$, $s = 6.3 \text{ years}$ and $n = 75$.

$$\therefore T_{\text{stat}} = \frac{(17.4 - 15)}{\frac{6.3}{\sqrt{75}}} = 3.299 \sim t_{(74)}$$

\Rightarrow From the T-table $t_{74, 0.05} = 1.9925 = T_{\text{cv}}$

6) Graph will be like :-



Decision :- So by critical value approach our T_{stat} lies in the rejection region therefore we reject null so population mean time on death row will not likely be 15 years.

Result from R :-

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for a t-test. The code calculates the test statistic, critical values, and p-value, and prints the results.
- Environment:** Shows the values of variables defined in the script.
- Console:** Displays the output of the R commands, including the test statistics, critical values, and p-value.

Environment Variables:

Variable	Value
alpha	0.05
left_criticalval	-1.99254349518093
n	75
p_val	0.00149316364683508
pop_mean	15
right_criticalval	1.99254349518093
sample_mean	17.4
sample_sd	6.3
test_statistics	3.29914439536929

Console Output:

```
> pop_mean <- 15
> alpha = 0.05
> test_statistics = (sample_mean - pop_mean)/(sample_sd/sqrt(n))
> right_criticalval = (qt(p=alpha/2,df=n-1,lower.tail = FALSE))
> left_criticalval = (qt(p=alpha/2,df=n-1,lower.tail = TRUE))
> p_val = 2 * pt(q=test_statistics,df = n-1,lower.tail = FALSE)
> cat("Test Statistics : ",test_statistics,"\n")
Test Statistics : 3.299144
> cat("Left Critical Value : ",left_criticalval,"\n")
Left Critical Value : -1.992543
> cat("Right Critical Value : ",right_criticalval,"\n")
Right Critical Value : 1.992543
> cat("P value : ",p_val,"\n")
P value : 0.001493164
>
> # Conclusion : p val < alpha or test_statistics lies after the right critical value.
> # Conclusion : p val < alpha or test_statistics lies after the right critical value.
> # so we reject H0 which means population mean time on death row will not be equal to 15 years.
>
```

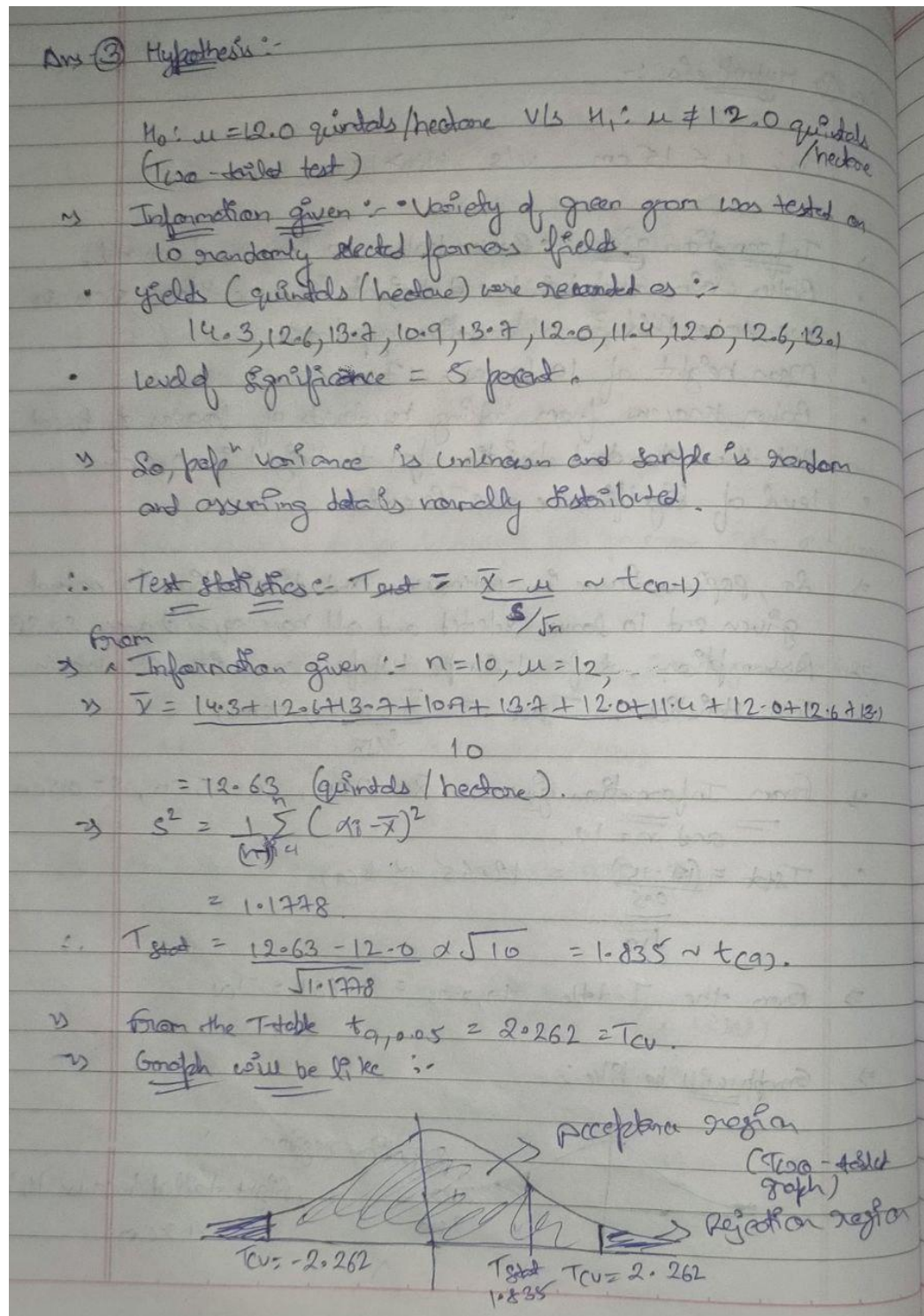
$T_{\text{stat}} = 3.299144$

$T_{\text{cv}} = 1.992543$

P value = 0.001493164

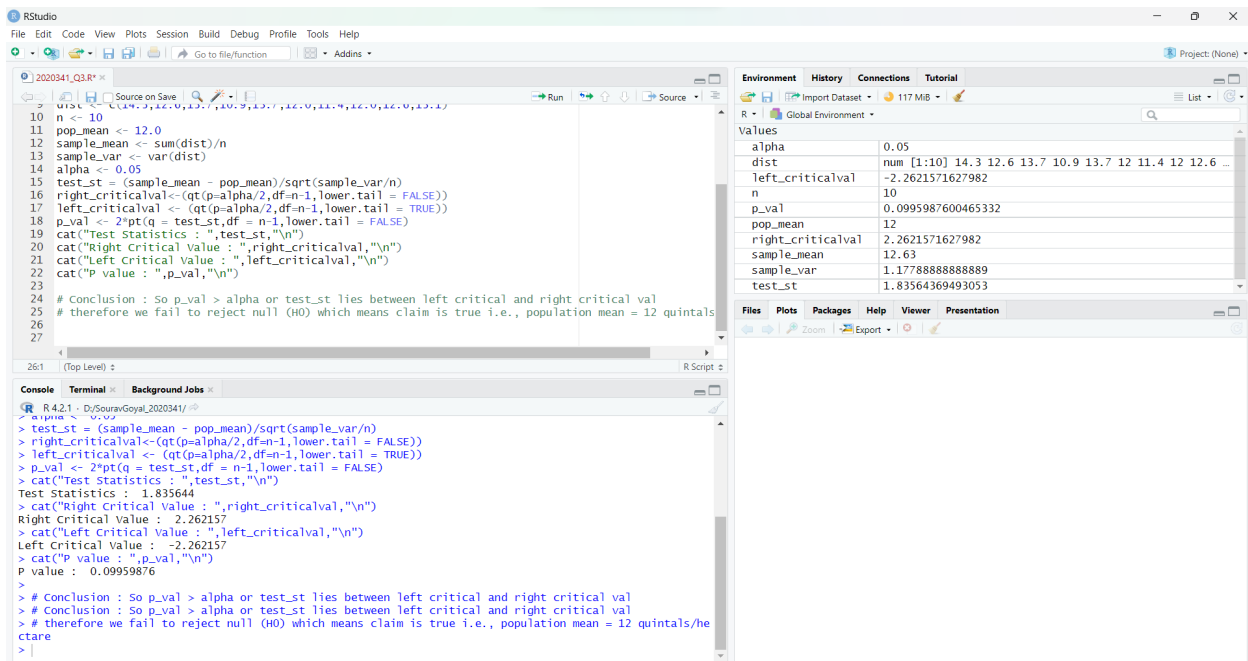
So, From R also we verify our results.

Ans - 3



Decision :- So by critical value approach our T_{stat} lies in the acceptance region therefore we fail to reject null so the claim is true that a new variety of green gram is expected to give a yield of 12.0 quintals per hectare.

Result from R :-



The screenshot displays the RStudio interface. The script editor on the left contains R code for a t-test. The console at the bottom shows the execution output, and the environment pane on the right lists the variables created during the process.

```
2020341_Q3.R  
# Source on Save  
# U.S. ...  
n <- 10  
pop_mean <- 12.0  
sample_mean <- sum(dist)/n  
sample_var <- var(dist)  
alpha <- 0.05  
test_st = (sample_mean - pop_mean)/sqrt(sample_var/n)  
right_criticalval <- (qt(p=alpha/2,df=n-1,lower.tail = FALSE))  
left_criticalval <- (qt(p=alpha/2,df=n-1,lower.tail = TRUE))  
p_val <- 2*pt(q = test_st,df = n-1,lower.tail = FALSE)  
cat("Test Statistics : ",test_st,"\n")  
cat("Right Critical Value : ",right_criticalval,"\n")  
cat("Left Critical Value : ",left_criticalval,"\n")  
cat("P value : ",p_val,"\n")  
# Conclusion : So p_val > alpha or test_st lies between left critical and right critical val  
# therefore we fail to reject null (H0) which means claim is true i.e., population mean = 12 quintals/he  
ctare  
>
```

Environment

Variable	Value
alpha	0.05
dist	num [1:10] 14.3 12.6 13.7 10.9 13.7 12 11.4 12 12.6 ...
left_criticalval	-2.2621571627982
n	10
p_val	0.0995987600465332
pop_mean	12
right_criticalval	2.2621571627982
sample_mean	12.63
sample_var	1.17788888888889
test_st	1.83564369493053

Console

```
R 4.2.1 - D:/SourceGoyal_2020341/  
> alpha <- 0.05  
> test_st = (sample_mean - pop_mean)/sqrt(sample_var/n)  
> right_criticalval <- (qt(p=alpha/2,df=n-1,lower.tail = FALSE))  
> left_criticalval <- (qt(p=alpha/2,df=n-1,lower.tail = TRUE))  
> p_val <- 2*pt(q = test_st,df = n-1,lower.tail = FALSE)  
> cat("Test Statistics : ",test_st,"\n")  
Test Statistics : 1.835644  
> cat("Right Critical Value : ",right_criticalval,"\n")  
Right Critical value : 2.262157  
> cat("Left Critical Value : ",left_criticalval,"\n")  
Left Critical Value : -2.262157  
> cat("P value : ",p_val,"\n")  
P value : 0.09959876  
>  
> # Conclusion : So p_val > alpha or test_st lies between left critical and right critical val  
> # Conclusion : So p_val > alpha or test_st lies between left critical and right critical val  
> # therefore we fail to reject null (H0) which means claim is true i.e., population mean = 12 quintals/he  
ctare  
>
```

$$T_{\text{stat}} = 1.835644$$

$$T_{\text{cv}} = 2.262157$$

$$P \text{ value} = 0.09959876$$

So, From R also we verify our results.

Ans - 4

Date / /
DELTA Pg No.

Ans (i) Hypothesis :-

$H_0: \mu_1 = \mu_2$ V/S $H_1: \mu_1 \neq \mu_2$, where
 μ_1 = average amount of boys spend in playing sports
 μ_2 = average amount of time girls spend in playing sports
 (Two-tailed test).

=> Information given :- So we have data of girls and boys in which we have sample size, average number of hours playing sports per day and sample standard deviation.

we denote 1 in subscript for boys and 2 for girls.

So, we have independent data and normal distribution is given. So Assumptions are met for 2-sample independent and variances are unknown and unequal.

level of significance = 5%.

Test Statistic :- $T_{stat} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{(df)}$

df = $\min(n_1 - 1, n_2 - 1)$,

From Information given :- $\bar{x}_1 = 3.2$ hours, $s_1^2 = 1$, $n_1 = 16$
 and $\bar{x}_2 = 2$ hours, $s_2^2 = 0.75$, $n_2 = 9$.

$\therefore df = \min(15, 8) = 8$.

$T_{stat} = \frac{3.2 - 2}{\sqrt{\frac{1}{16} + \frac{0.75}{9}}} = 3.1423 \sim t_{(8)}$.

From T-table we have $t_{8, 0.05} = 2.306 = T_{CV}$

Graph will be like :-

Decision :- So by critical value approach our T_{stat} lies in the rejection region therefore we reject null so the claim is false that the population mean of playing sports time of boys and girls is same.

Result from R :-

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for a two-sample t-test. The code calculates the t-statistic, degrees of freedom, critical values, p-value, and prints the results.
- Environment Pane:** Shows the values of variables created in the script.

Variable	Value
alpha	0.05
df	8
left_criticalval	-2.30600413520417
n_boys	16
n_girls	9
p_val	0.0137566065495193
right_criticalval	2.30600413520417
sample_mean_boys	3.2
sample_mean_girls	2
sample_var_boys	1
sample_var_girls	0.75
t_test	3.14233761939829
- Console:** Shows the execution output of the R script, including the test statistics, critical values, p-value, and the final conclusion.

$T_{\text{stat}} = 3.142338$

$T_{\text{cv}} = 2.306004$

P value = 0.01375661

So, From R also we verify our results.

Ans - 5

Ans ⑤ Given :- Food A : 49, 53, 51, 52, 47, 50, 52, 53
= Food B : 52, 55, 52, 53, 50, 54, 54, 53
• Differences of pairs is normally distributed.

2) So,

i	x_i	y_i	$d_i = x_i - y_i$
0	49	52	-3
1	53	55	-2
2	51	52	-1
3	52	53	-1
4	47	50	-3
5	50	54	-4
6	52	54	-2
7	53	53	0

4) Hypothesis :-
 $H_0 : \mu_d = 0$ v/s $H_1 : \mu_d \neq 0$
 μ_d = mean value of change in weight of children due to Food B.

5) Assumptions are met as sample data are dependent and difference of pairs is normally distributed.

∴ Test Statistics :- $T_{stat} = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} \sim t_{(n-1)}$

∴ Now, $n = 8$,

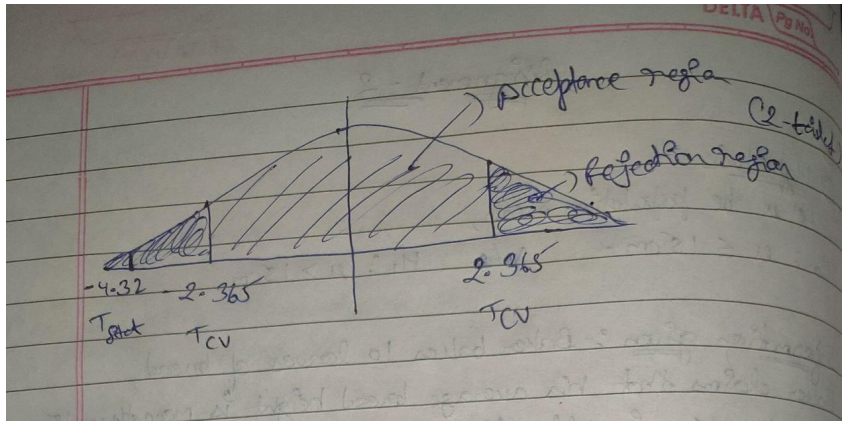
∴ Now, $\bar{d} = \frac{(-3-2-1-1-3-4-2+0)}{8} = -2$

∴ And, $s_d = \sqrt{\frac{1}{n} \sum (d_i - \bar{d})^2}$
 $= 1.3093$

∴ $T_{stat} = \frac{-2 \times \sqrt{8}}{1.3093} = -4.32 \sim t_{(7)}$

∴ From T value table $t_{7, 0.05} = 2.365$

Graph will be like :-



Decision :- So by critical value approach our Tstat lies in the rejection region therefore we reject null so the claim is false which means population mean of difference of pairs is non zero.

Result from R :-

The screenshot shows the RStudio interface. The script editor on the left contains the following R code:

```

9
10 n <- 8
11 FoodA <- c(49,53,51,52,47,50,52,53)
12 FoodB <- c(52,55,52,53,50,54,54,53)
13 alpha <- 0.05
14 test.ans <- t.test(x=FoodA, y=FoodB, paired = TRUE, alternative = "two.sided")
15 t.stat <- test.ans$statistic
16 left_critical_value <- qt(p = alpha/2, df=n-1, lower.tail = TRUE)
17 right_critical_value <- qt(p = alpha/2, df=n-1, lower.tail = FALSE)
18 p_val <- test.ans$p.value
19 cat("Test Statistics : ", t.stat, "\n")
20 cat("Left Critical Value : ", left_critical_value, "\n")
21 cat("Right Critical Value : ", right_critical_value, "\n")
22 cat("P Value : ", p_val, "\n")
23
24 # Conclusion : p_val < alpha or t.stat lies before left critical value which is rejection region
25 # So we reject H0 which means claim is false i.e., average difference of pairs are not zero
26
27

```

The Environment pane on the right shows the following data:

Object	Class	Value
test.ans	htest	htest object
alpha	num	0.05
FoodA	num	[1:8] 49 53 51 52 47 50 52 53
FoodB	num	[1:8] 52 55 52 53 50 54 54 53
left_critical_value	num	-2.36462425159278
n	num	8
p_val	num	0.00347808426500278
right_critical_value	num	2.36462425159278
t.stat	num	-4.32

The Console pane at the bottom shows the output of the R code:

```

R 4.2.1 - DrSouvikGoyal_2020341/
> FoodA <- c(49,53,51,52,47,50,52,53)
> FoodB <- c(52,55,52,53,50,54,54,53)
> alpha <- 0.05
> test.ans <- t.test(x=FoodA, y=FoodB, paired = TRUE, alternative = "two.sided")
> t.stat <- test.ans$statistic
> left_critical_value <- qt(p = alpha/2, df=n-1, lower.tail = TRUE)
> right_critical_value <- qt(p = alpha/2, df=n-1, lower.tail = FALSE)
> p_val <- test.ans$p.value
> cat("Test Statistics : ", t.stat, "\n")
Test Statistics : -4.320494
> cat("Left Critical Value : ", left_critical_value, "\n")
Left Critical Value : -2.364624
> cat("Right Critical Value : ", right_critical_value, "\n")
Right Critical Value : 2.364624
> cat("P Value : ", p_val, "\n")
P Value : 0.003478084
> # Conclusion : p_val < alpha or t.stat lies before left critical value which is rejection region
> # So we reject H0 which means claim is false i.e., average difference of pairs are not zero
>

```

$T_{stat} = -4.320494$

$T_{cv} = -2.262157$

P value = 0.003478084

So, From R also we verify our results.