



Bansilal Ramnath Agarwal Charitable Trust's
Vishwakarma Institute of Information
Technology

**Department of
Artificial Intelligence and Data
Science**

Name: Siddhesh Dilip Khairnar

Class: SY

Division: B

Roll No: 272028

Semester: IV

Academic Year: 2022-2023

Subject Name & Code: ES22201AD: Probability and Statistics

Title of Assignment: Hypothesis testing in R

Date of Performance: 27-04-2023

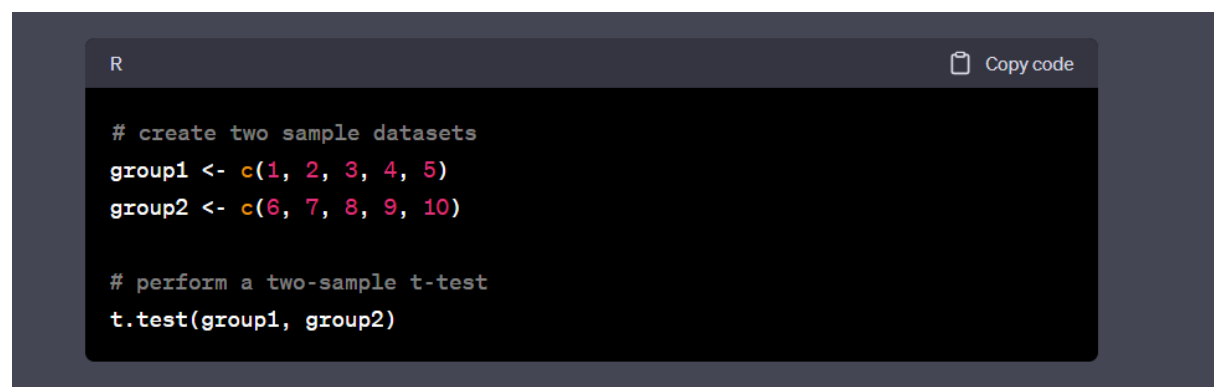
Date of Submission: 27-04-2023

ASSIGNMENT NO. 7

Background information:

Hypothesis testing is a statistical method that allows us to test whether there is a significant difference between two groups, or whether a certain relationship exists between two variables. In hypothesis testing, we formulate a null hypothesis (H_0) that assumes there is no difference or relationship, and an alternative hypothesis (H_a) that assumes there is a difference or relationship. We then use statistical methods to determine the likelihood of the observed data given each hypothesis.

In R, there are several packages that can be used for hypothesis testing, including ``stats``, ``car``, and ``psych``. Here is an example of how to perform a hypothesis test using the `t.test()` function from the ``stats`` package:

A screenshot of an R console window with a dark background. The window has a title bar with 'R' on the left and a 'Copy code' button on the right. The console contains the following R code:

```
# create two sample datasets
group1 <- c(1, 2, 3, 4, 5)
group2 <- c(6, 7, 8, 9, 10)

# perform a two-sample t-test
t.test(group1, group2)
```

This will perform a two-sample t-test between ``group1`` and ``group2``, and output the results including the p-value. The p-value represents the probability of observing the data given the null hypothesis, and a small p-value (typically less than 0.05) indicates that the null hypothesis should be rejected in favor of the alternative hypothesis.

You can also perform other types of hypothesis tests using R, such as chi-square tests or ANOVA tests, depending on your research question and data.

Program and Output:

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs
R 4.2.2 ~/
> sample(100, size=100, replace=TRUE)
mean of x
303.6713

> t.test(turtles, var.equal = FALSE)$p.value
[1] 3.578637e-47
> turtles
[1] 323.0258 328.6194 321.4139 283.0273 315.0791 290.9349 297.1569 310.5372 269.6901 282.3625 296.6962 270.2296 349.7296 324.1544
[15] 305.4443 300.3842 322.2767 322.5484 273.4874 299.5718 302.5579 302.5956 281.8637 307.1443 292.2426 281.4497 307.4346 276.5295
[29] 265.8561 335.0668 305.8242 353.5639 298.0346 286.4092 303.5420 322.6174 310.6682 317.6221 317.1238 292.3365
> #double test
> turtles_1<-c(rnorm(38,mean=305,sd=16.7))
> turtles_1
[1] 320.8620 333.1927 318.7885 314.8289 295.3377 301.7108 328.8536 288.7556 297.6098 298.2688 295.3041 302.2558 299.7324 317.4452
[15] 342.8354 301.7863 299.4526 332.4803 321.2977 301.9436 315.1620 297.7769 293.6954 267.0650 279.1133 280.1733 292.9081 316.8001
[29] 321.5712 295.5713 282.7300 329.4573 292.2955 289.4808 308.8799 321.2473 300.9191 293.6710
> t.test(turtles, turtles_1, mu=310, conf.level = 0.90)

Welch Two Sample t-test

data: turtles and turtles_1
t = -72.205, df = 74.272, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 310
90 percent confidence interval:
-8.544375  5.820739
sample estimates:
mean of x mean of y
303.6713  305.0331

> wilcox.test(turtles, turtles_1, var.equal=FALSE)$p.value
[1] 0.9090028
> |
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs
R 4.2.2 ~/
> turtles<-c(rnorm(40,mean=300,sd=18.5))
> t.test(turtles, mu=310)

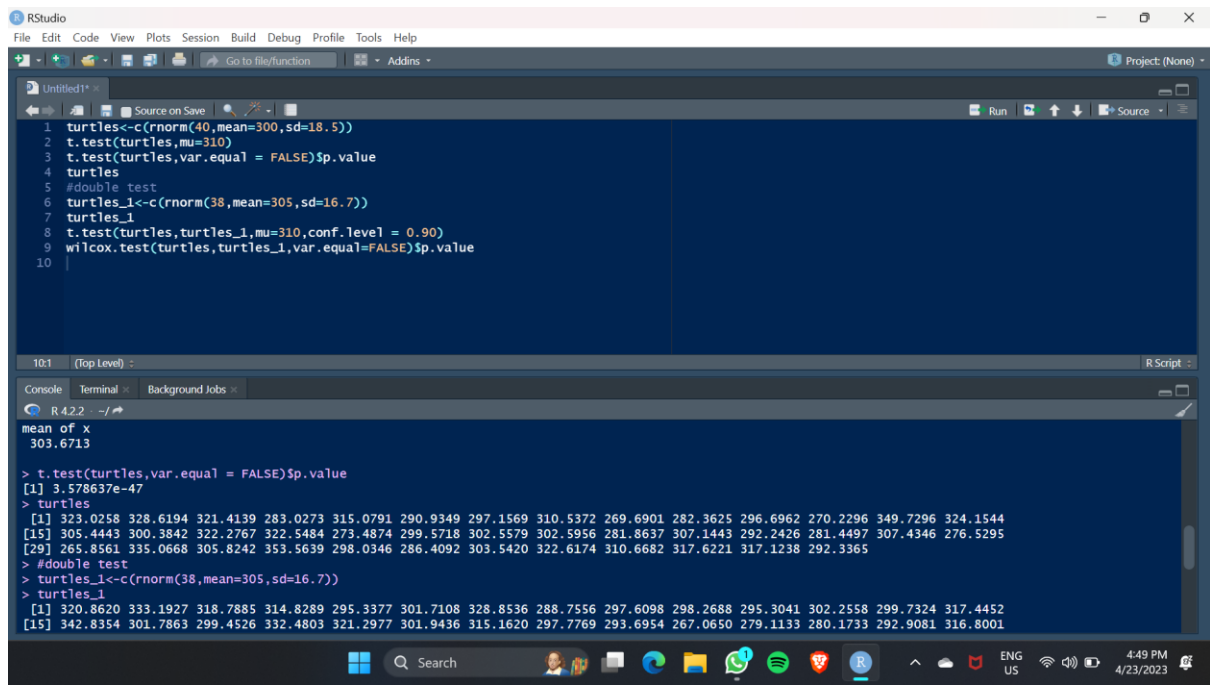
One Sample t-test

data: turtles
t = -1.9117, df = 39, p-value = 0.06329
alternative hypothesis: true mean is not equal to 310
95 percent confidence interval:
296.9750 310.3676
sample estimates:
mean of x
303.6713

> t.test(turtles, var.equal = FALSE)$p.value
[1] 3.578637e-47
> turtles
[1] 323.0258 328.6194 321.4139 283.0273 315.0791 290.9349 297.1569 310.5372 269.6901 282.3625 296.6962 270.2296 349.7296 324.1544
[15] 305.4443 300.3842 322.2767 322.5484 273.4874 299.5718 302.5579 302.5956 281.8637 307.1443 292.2426 281.4497 307.4346 276.5295
[29] 265.8561 335.0668 305.8242 353.5639 298.0346 286.4092 303.5420 322.6174 310.6682 317.6221 317.1238 292.3365
> #double test
> turtles_1<-c(rnorm(38,mean=305,sd=16.7))
> turtles_1
[1] 320.8620 333.1927 318.7885 314.8289 295.3377 301.7108 328.8536 288.7556 297.6098 298.2688 295.3041 302.2558 299.7324 317.4452
[15] 342.8354 301.7863 299.4526 332.4803 321.2977 301.9436 315.1620 297.7769 293.6954 267.0650 279.1133 280.1733 292.9081 316.8001
[29] 321.5712 295.5713 282.7300 329.4573 292.2955 289.4808 308.8799 321.2473 300.9191 293.6710
> t.test(turtles, turtles_1, mu=310, conf.level = 0.90)

Welch Two Sample t-test

data: turtles and turtles_1
```



```
1 turtles<-c(rnorm(40,mean=300,sd=18.5))
2 t.test(turtles,mu=310)
3 t.test(turtles,var.equal = FALSE)$p.value
4 turtles
5 #double test
6 turtles_1<-c(rnorm(38,mean=305,sd=16.7))
7 turtles_1
8 t.test(turtles,turtles_1,mu=310,conf.level = 0.90)
9 wilcox.test(turtles,turtles_1,var.equal=FALSE)$p.value
10
```

```
mean of x
303.6713

> t.test(turtles,var.equal = FALSE)$p.value
[1] 3.578637e-47
> turtles
[1] 323.0258 328.6194 321.4139 283.0273 315.0791 290.9349 297.1569 310.5372 269.6901 282.3625 296.6962 270.2296 349.7296 324.1544
[15] 305.4443 300.3842 322.2767 322.5484 273.4874 299.5718 302.5579 302.5956 281.8637 307.1443 292.2426 281.4497 307.4346 276.5295
[29] 265.8561 335.0668 305.8242 353.5639 298.0346 286.4092 303.5420 322.6174 310.6682 317.6221 317.1238 292.3365
> #double test
> turtles_1<-c(rnorm(38,mean=305,sd=16.7))
> turtles_1
[1] 320.8620 333.1927 318.7885 314.8289 295.3377 301.7108 328.8536 288.7556 297.6098 298.2688 295.3041 302.2558 299.7324 317.4452
[15] 342.8354 301.7863 299.4526 332.4803 321.2977 301.9436 315.1620 297.7769 293.6954 267.0650 279.1133 280.1733 292.9081 316.8001
```

Conclusion: We have successfully implemented hypothesis testing (z-test and t-test) in RStudio.