

**SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE**  
**SUBJECT - Data Analysis with R/SAS/SPSS**

Aim - Performing independent two-sample t-tests using t.test() with grouping (R).

Output :

```
> data <- read.csv("penguins.csv")
> head(data)
  species  island culmen_length culmen_depth flipper_length body_mass   sex
1  Adelie     Torgersen      39.1        18.7         181     3750  MALE
2  Adelie     Torgersen      39.5        17.4         186     3800 FEMALE
3  Adelie     Torgersen      40.3        18.0         195     3250 FEMALE
4  Adelie     Torgersen       NA         NA          NA       NA <NA>
5  Adelie     Torgersen      36.7        19.3         193     3450 FEMALE
6  Adelie     Torgersen      39.3        20.6         190     3650  MALE
> str(data)
'data.frame': 344 obs. of  7 variables:
 $ species : chr "Adelie" "Adelie" "Adelie" ...
 $ island   : chr "Torgersen" "Torgersen" "Torgersen" ...
 $ culmen_length: num 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
 $ culmen_depth : num 18.7 17.4 18.0 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
 $ flipper_length: int 181 186 195 NA 193 190 181 195 193 190 ...
 $ body_mass   : int 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
 $ sex       : chr "MALE" "FEMALE" "FEMALE" NA ...
> data$group_culmen_length <- ifelse(data$culmen_length >= median(data$culmen_length, na.rm = TRUE), "High", "Low")
> data$group_culmen_depth <- ifelse(data$culmen_depth >= median(data$culmen_depth, na.rm = TRUE), "High", "Low")
> data$group_flipper_length <- ifelse(data$flipper_length >= median(data$flipper_length, na.rm = TRUE), "High", "Low")
> data$group_body_mass <- ifelse(data$body_mass >= median(data$body_mass, na.rm = TRUE), "High", "Low")
> table(data$group_culmen_length)

High    Low
171    171
> table(data$group_culmen_depth)

High    Low
176    166
> table(data$group_flipper_length)

High    Low
176    166
> table(data$group_body_mass)
```

RStudio  
File Edit Code View Plots Session Build Debug Profile Tools Help  
Source  
Console Terminal Background Jobs  
R > 4.5.2 - />

```
Welch Two Sample t-test

data: culmen_length by group_culmen_length
t = 31.413, df = 339.97, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High and group Low is not equal to 0
95 percent confidence interval:
 8.814141 9.991707
sample estimates:
mean in group High mean in group Low
 48.62339          39.22047

> t.test(culmen_length ~ group_culmen_length, data = data, var.equal = TRUE)

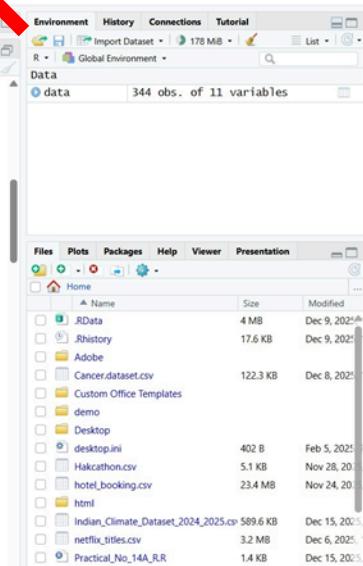
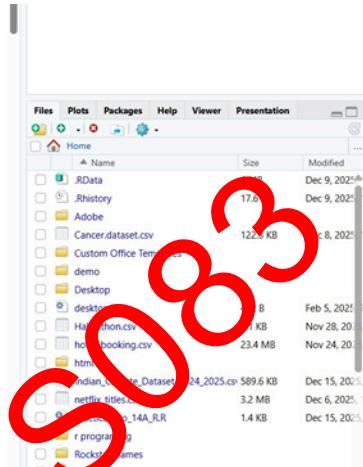
Two Sample t-test

data: culmen_length by group_culmen_length
t = 31.413, df = 340, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High and group Low is not equal to 0
95 percent confidence interval:
 8.814141 9.991707
sample estimates:
mean in group High mean in group Low
 48.62339          39.22047

> t.test(culmen_length ~ group_culmen_length, data = data, alternative = "greater")

Welch Two Sample t-test

data: culmen_length by group_culmen_length
t = 31.413, df = 339.97, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High and group Low is greater than 0
95 percent confidence interval:
 8.909215 Inf
sample estimates:
mean in group High mean in group Low
 48.62339          39.22047
```



MITHIL KADAM S083

**SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE**  
**SUBJECT - Data Analysis with R/SAS/SPSS**



The image displays two side-by-side screenshots of the RStudio IDE. Both screenshots show the R console, environment, and file browser panes. The R console output shows the execution of several Welch Two Sample t-tests on the 'data' dataset. The tests compare 'culmen\_depth' and 'flipper\_length' between 'group\_culmen\_depth' and 'group\_flipper\_length'. The results indicate significant differences for both variables.

```
> t.test(culmen_depth ~ group_culmen_depth, data = data, alternative = "greater")
Welch Two Sample t-test

data: culmen_depth by group_culmen_depth
t = 28.723, df = 320.51, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High and group Low is greater than 0
95 percent confidence interval:
 3.134589   Inf
sample estimates:
mean in group High mean in group Low
18.76534      15.43976

> t.test(culmen_depth ~ group_culmen_depth, data = data, alternative = "less")
Welch Two Sample t-test

data: culmen_depth by group_culmen_depth
t = 28.723, df = 320.51, p-value = 1
alternative hypothesis: true difference in means between group High and group Low is less than 0
95 percent confidence interval:
-Inf 3.516575
sample estimates:
mean in group High mean in group Low
18.76534      15.43976

> t.test(flipper_length ~ group_flipper_length, data = data)
Welch Two Sample t-test

data: flipper_length by group_flipper_length
t = 28.048, df = 274.54, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High and group Low is not equal to 0
95 percent confidence interval:
 21.98250 25.19056
sample estimates:
mean in group High mean in group Low
212.3636      188.7771

> t.test(flipper_length ~ group_flipper_length, data = data, var.equal = TRUE)

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help

Source
Console Terminal Background Jobs
<R + R4.5.2 - />
> t.test(flipper_length ~ group_flipper_length, data = data, var.equal = TRUE)
Welch Two Sample t-test

data: flipper_length by group_flipper_length
t = 28.048, df = 274.54, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High and group Low is greater than 0
95 percent confidence interval:
 22.24178   Inf
sample estimates:
mean in group High mean in group Low
212.3636      188.7771

> t.test(flipper_length ~ group_flipper_length, data = data, alternative = "less")
Welch Two Sample t-test

data: flipper_length by group_flipper_length
t = 28.048, df = 274.54, p-value = 1
alternative hypothesis: true difference in means between group High and group Low is less than 0
95 percent confidence interval:
-Inf 24.93128
sample estimates:
mean in group High mean in group Low
212.3636      188.7771

> t.test(body_mass ~ group_body_mass, data = data)
Welch Two Sample t-test

data: body_mass by group_body_mass
t = 27.641, df = 263.75, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High and group Low is not equal to 0
95 percent confidence interval:
 1236.204 1425.834
sample estimates:
mean in group High mean in group Low
4863.372      3532.359
```

The right screenshot shows the file browser pane with various CSV files listed, including 'Cancer.dataset.csv', 'Custom Office Templates', 'Hakathon.csv', 'Hotel\_Booking.csv', 'HTML', 'Indian\_Climate\_Dataset\_2024\_2025.csv', 'Netflix\_titles.csv', 'Practical\_No\_14A.R.R', and 'Rockstar Games'.