

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" "island" "culmen_length" "culmen_depth" "flipper_length" "body_mass" "sex"
 $ species      : chr  "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" "Adelie" ...
 $ island       : chr  "Torgersen" "Torgersen" "Torgersen" "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 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" "MALE" "MALE" "FEMALE" ...
> 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)
```

```
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
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
```

Name - Mithil Kadam
Roll No - S083

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

The screenshot displays the RStudio interface with the following components:

- Source:** Contains R code for three t-tests. The first two tests are for `culmen_depth` (comparing group High and group Low) with `alternative = "greater"` and `alternative = "less"`. The third test is for `flipper_length` with `alternative = "greater"`. The fourth test is for `body_mass` with `alternative = "greater"`.
- Console:** Displays the output of the t-tests. The first test shows a p-value of $2.2e-16$ and a 95% confidence interval of 3.134589 to Inf . The second test shows a p-value of 1 and a 95% confidence interval of $-Inf$ to 3.516575 . The third test shows a p-value of $2.2e-16$ and a 95% confidence interval of 22.24178 to Inf . The fourth test shows a p-value of $2.2e-16$ and a 95% confidence interval of 1236.204 to 1425.834 .
- Environment:** Shows the data environment with 344 observations and 11 variables.
- Files:** Shows a list of files in the project directory, including `RData`, `Rhistory`, `Adobe`, `Cancer.dataset.csv`, `Custom Office Templates`, `demo`, `Desktop`, `desktop.ini`, `Hakathon.csv`, `hotel_booking.csv`, `html`, `Indian_Climate_Dataset_2024_2025.csv`, `netflix_titles.csv`, `Practical_No_14A_RR`, `r programming`, and `Rockstar Games`.

Name - Mithil Kadam
Roll No - S083