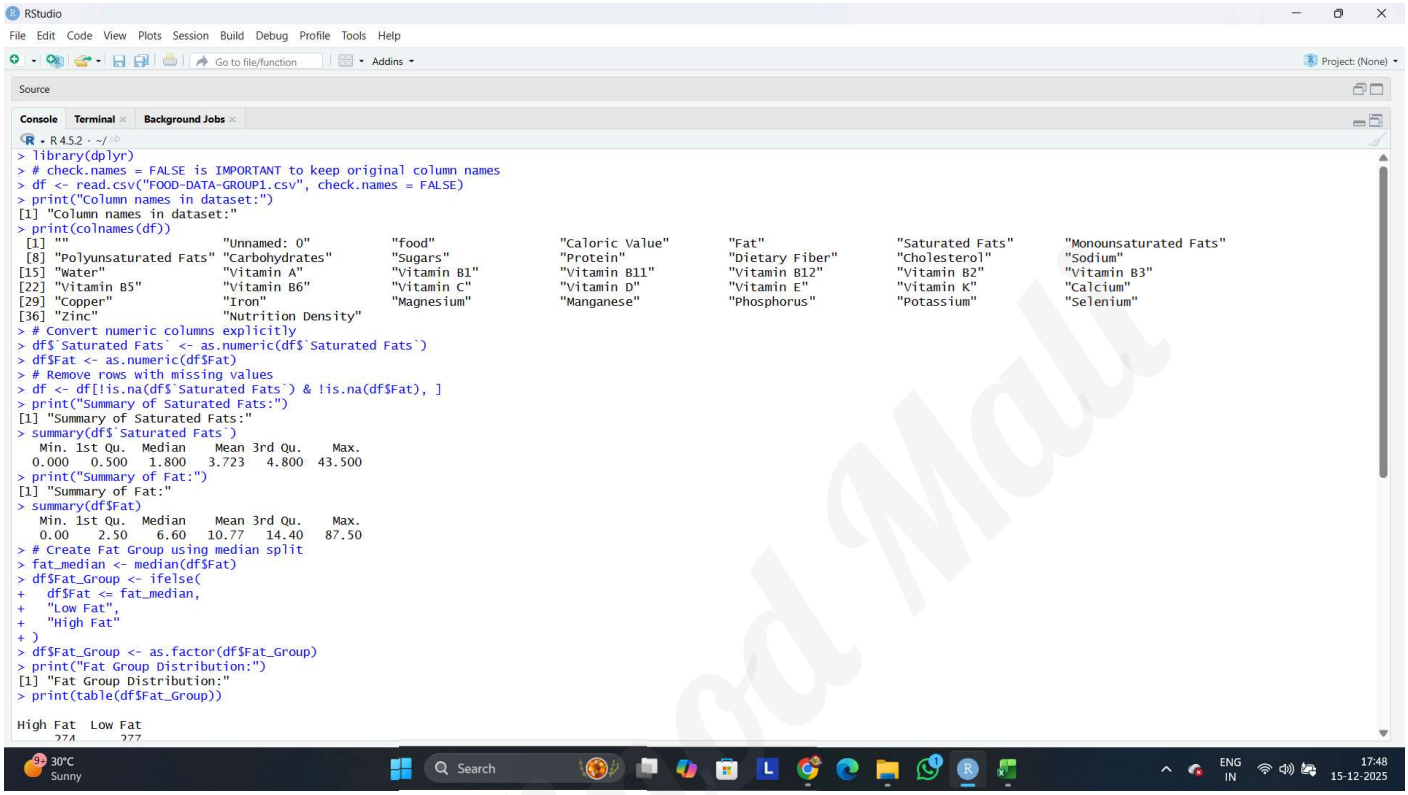
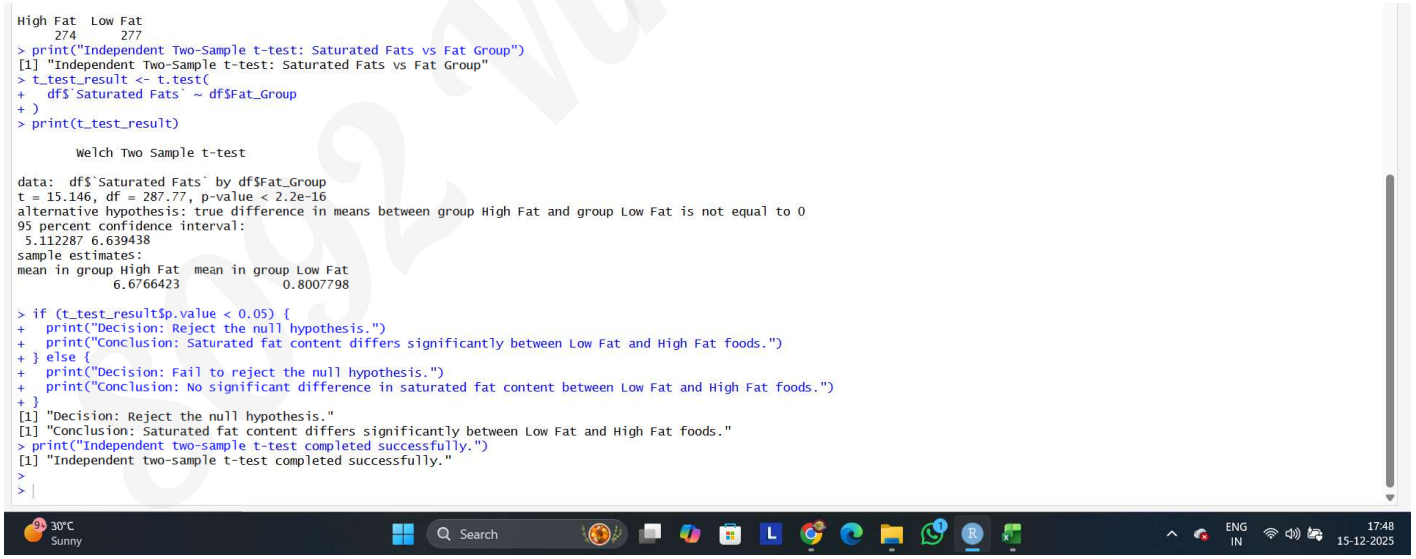
Sheth L.U.J & Sir M.V College Of Science
Subject :- Data Analysis with SAS/SPSS/R
Practical no 5

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



```
> library(dplyr)
> # check.names = FALSE is IMPORTANT to keep original column names
> df <- read.csv("FOOD-DATA-GROUP1.csv", check.names = FALSE)
> print("Column names in dataset:")
[1] "Column names in dataset:"
> print(colnames(df))
 [1] ""                      "Unnamed: 0"            "food"                  "Caloric Value"         "Fat"                   "Saturated Fats"        "Monounsaturated Fats"
 [8] "Polyunsaturated Fats" "Carbohydrates"         "Sugars"                "Protein"               "Dietary Fiber"         "Cholesterol"           "Sodium"
[15] "Water"                "Vitamin A"             "Vitamin B1"            "Vitamin B11"           "Vitamin B12"           "Vitamin B2"            "Vitamin B3"
[22] "Vitamin B5"           "Vitamin B6"            "Vitamin C"             "Vitamin D"             "Vitamin E"             "Vitamin K"             "Calcium"
[29] "Copper"               "Iron"                  "Magnesium"             "Manganese"             "Phosphorus"            "Potassium"             "Selenium"
[36] "Zinc"                 "Nutrition Density"
> # Convert numeric columns explicitly
> df$`Saturated Fats` <- as.numeric(df$`Saturated Fats`)
> df$Fat <- as.numeric(df$Fat)
> # Remove rows with missing values
> df <- df[!is.na(df$`Saturated Fats`) & !is.na(df$Fat), ]
> print("Summary of Saturated Fats:")
[1] "Summary of Saturated Fats:"
> summary(df$`Saturated Fats`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   0.500   1.800   3.723   4.800  43.500
> print("Summary of Fat:")
[1] "Summary of Fat:"
> summary(df$Fat)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    2.50    6.60   10.77   14.40   87.50
> # Create Fat Group using median split
> fat_median <- median(df$Fat)
> df$Fat_Group <- ifelse(
+   df$Fat <= fat_median,
+   "Low Fat",
+   "High Fat"
+ )
> df$Fat_Group <- as.factor(df$Fat_Group)
> print("Fat Group Distribution:")
[1] "Fat Group Distribution:"
> print(table(df$Fat_Group))

High Fat  Low Fat
     274      277
```



```
High Fat  Low Fat
     274      277
> print("Independent Two-Sample t-test: Saturated Fats vs Fat Group")
[1] "Independent Two-Sample t-test: Saturated Fats vs Fat Group"
> t_test_result <- t.test(
+   df$`Saturated Fats` ~ df$Fat_Group
+ )
> print(t_test_result)

        Welch Two Sample t-test

data:  df$`Saturated Fats` by df$Fat_Group
t = 15.146, df = 287.77, p-value < 2.2e-16
alternative hypothesis: true difference in means between group High Fat and group Low Fat is not equal to 0
95 percent confidence interval:
 5.112287 6.639438
sample estimates:
mean in group High Fat  mean in group Low Fat
             6.6766423              0.8007798

> if (t_test_result$p.value < 0.05) {
+   print("Decision: Reject the null hypothesis.")
+   print("Conclusion: Saturated fat content differs significantly between Low Fat and High Fat foods.")
+ } else {
+   print("Decision: Fail to reject the null hypothesis.")
+   print("Conclusion: No significant difference in saturated fat content between Low Fat and High Fat foods.")
+ }
[1] "Decision: Reject the null hypothesis."
[1] "Conclusion: Saturated fat content differs significantly between Low Fat and High Fat foods."
> print("Independent two-sample t-test completed successfully.")
[1] "Independent two-sample t-test completed successfully."
>
>
```


Name :- Vinod Mali
Roll No :- S092