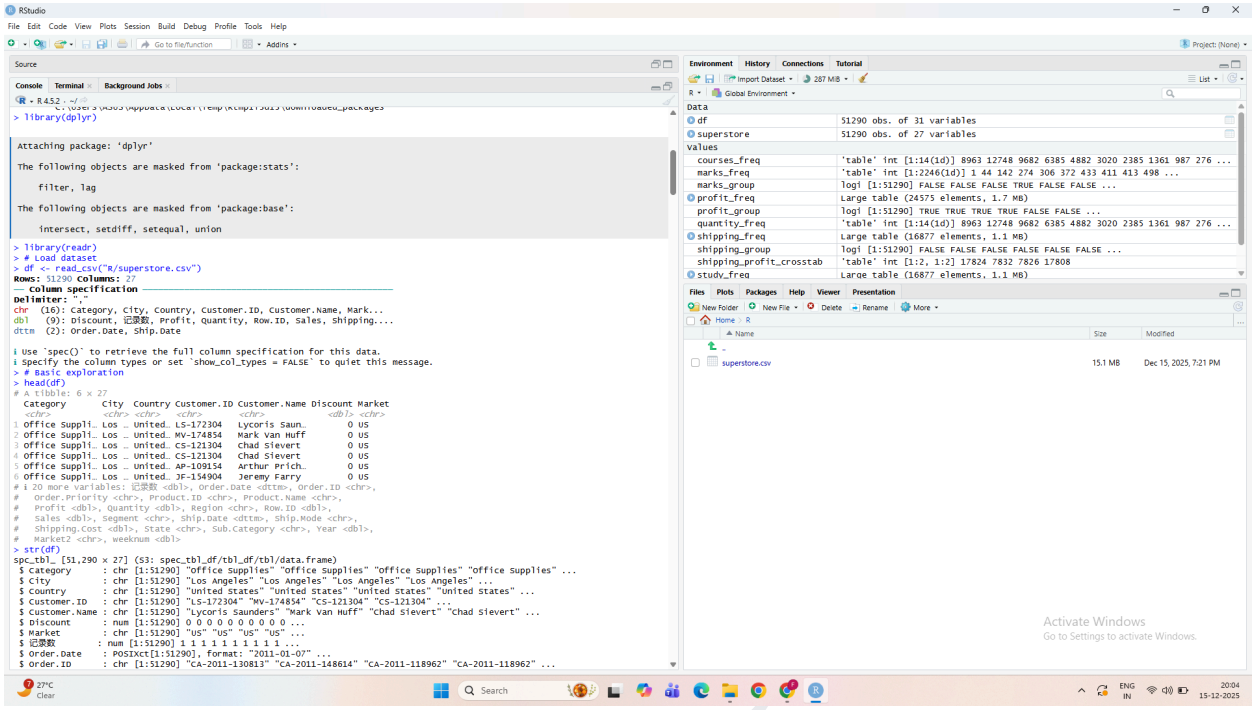
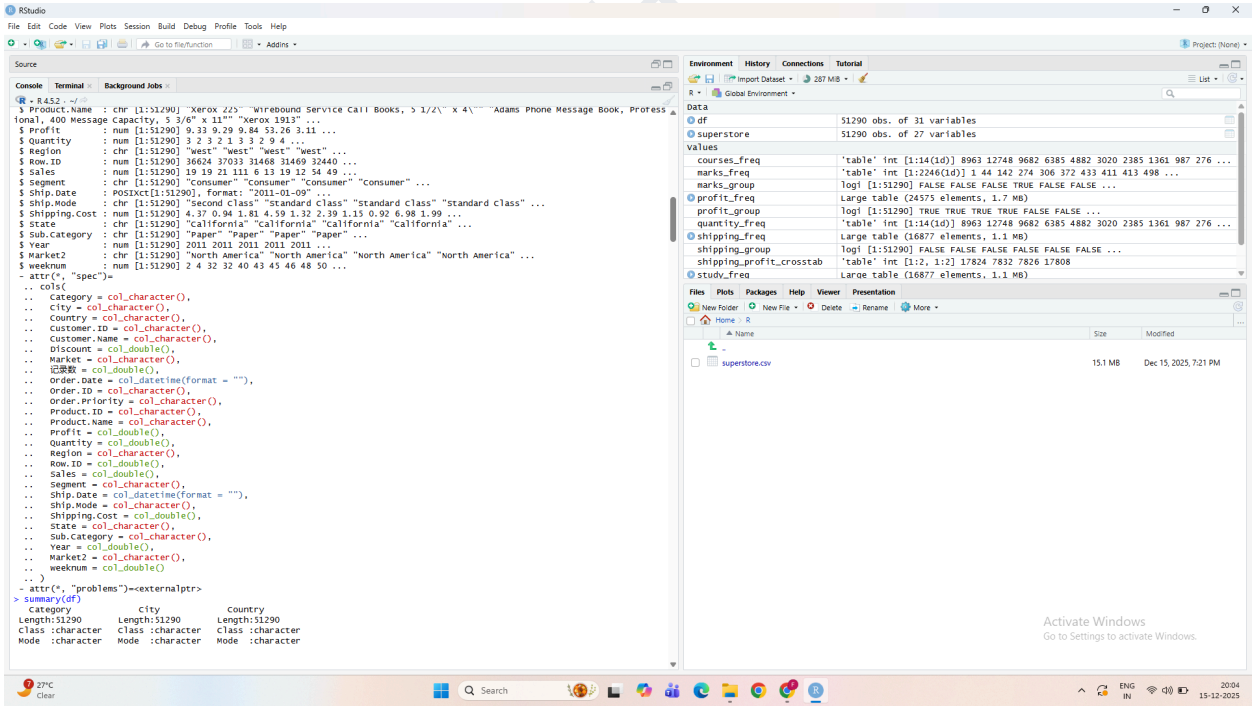
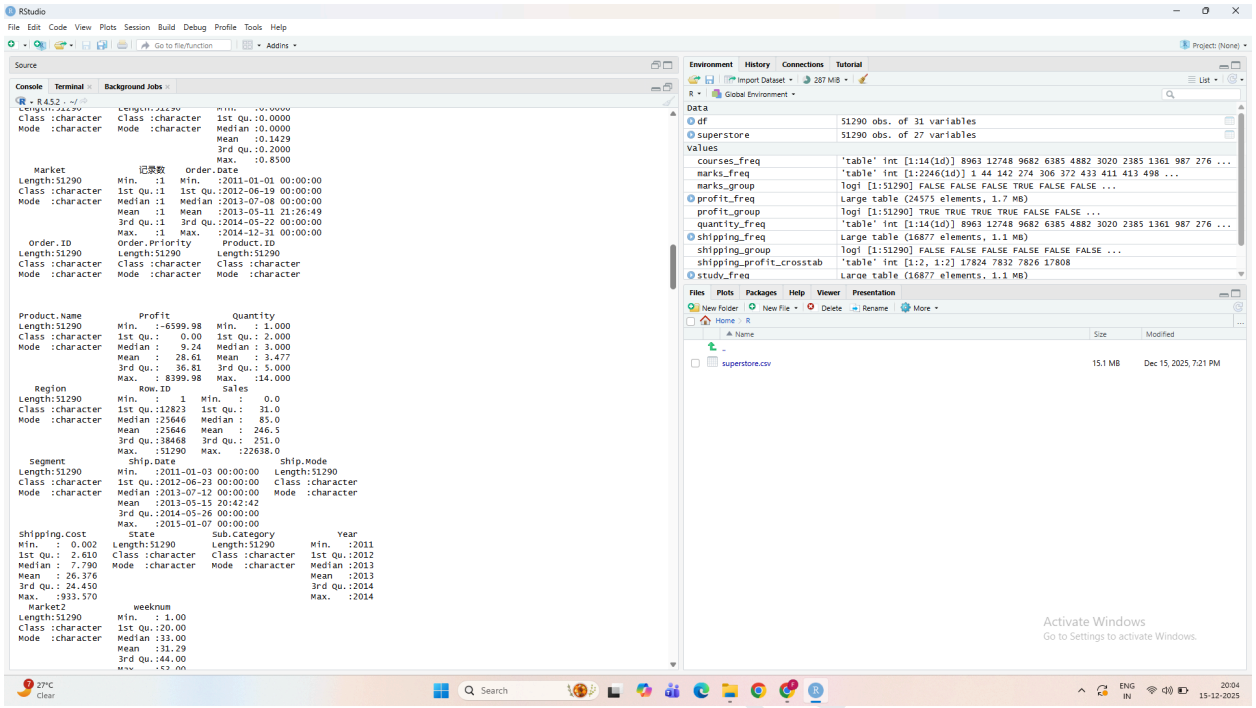
SHETH L.U.J. AND SIR M.V. COLLEGE

PRACTICAL NO 6

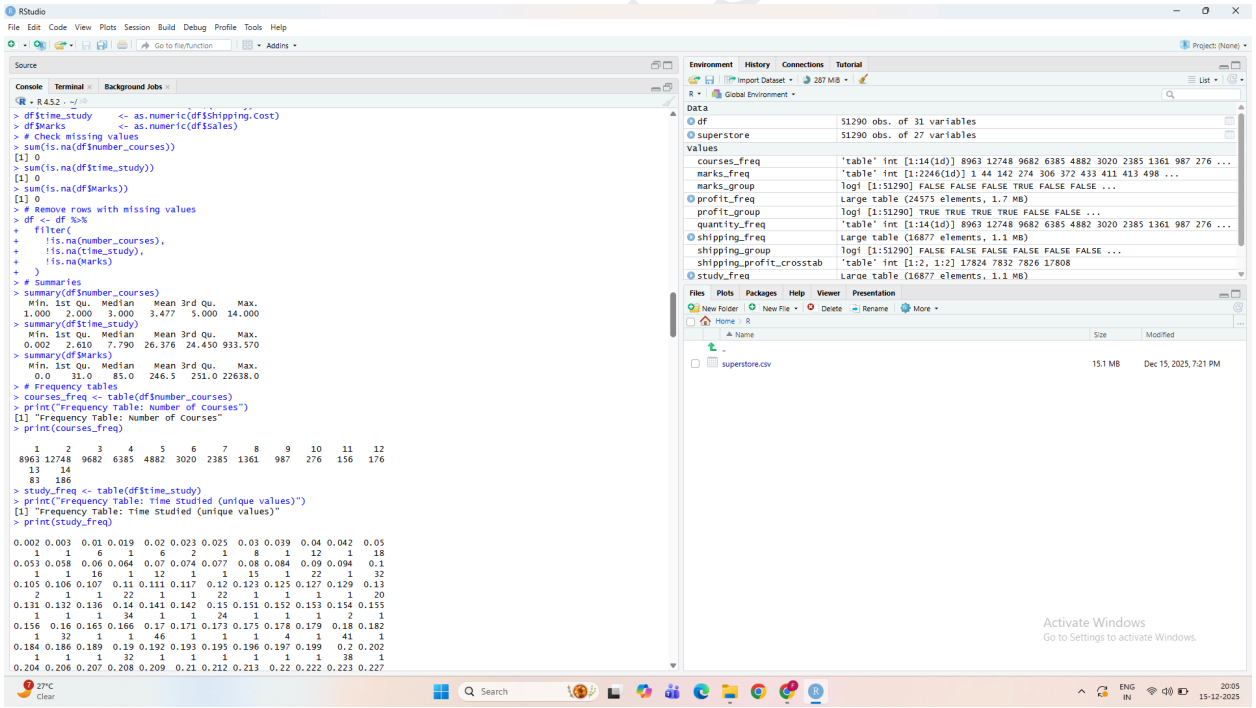
AIM: Performing paired t-tests using t.test(paired=TRUE) (R).



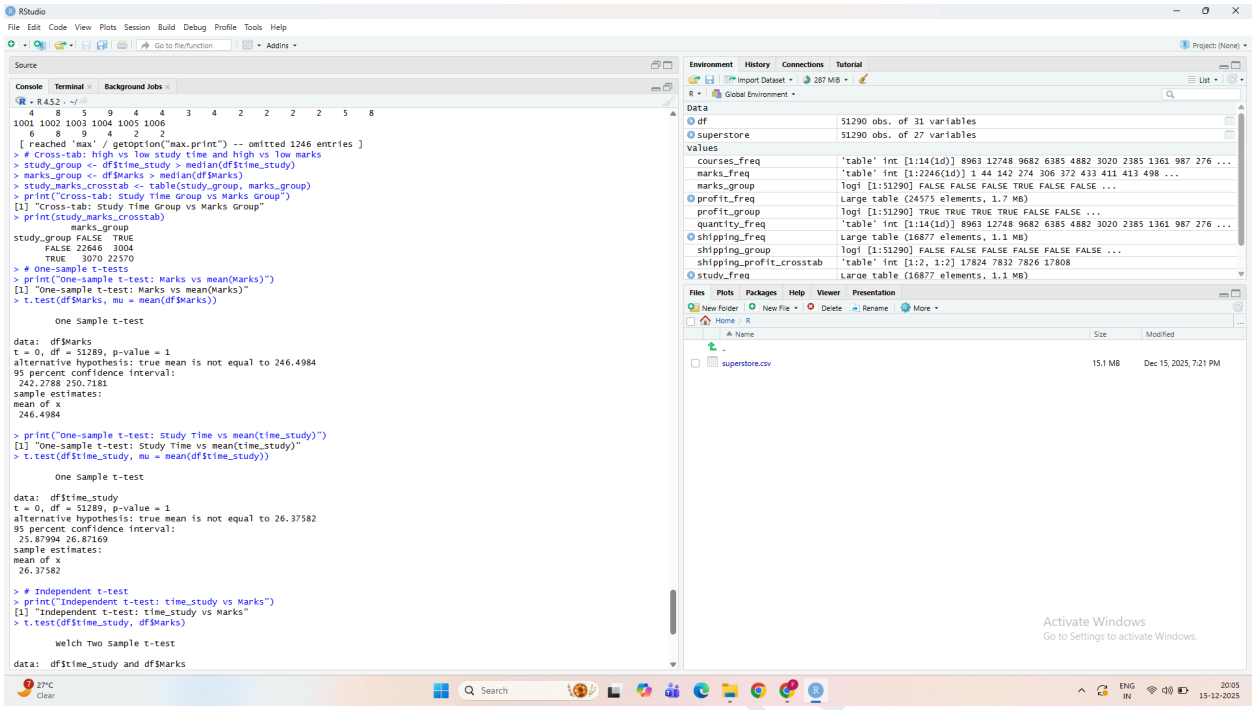


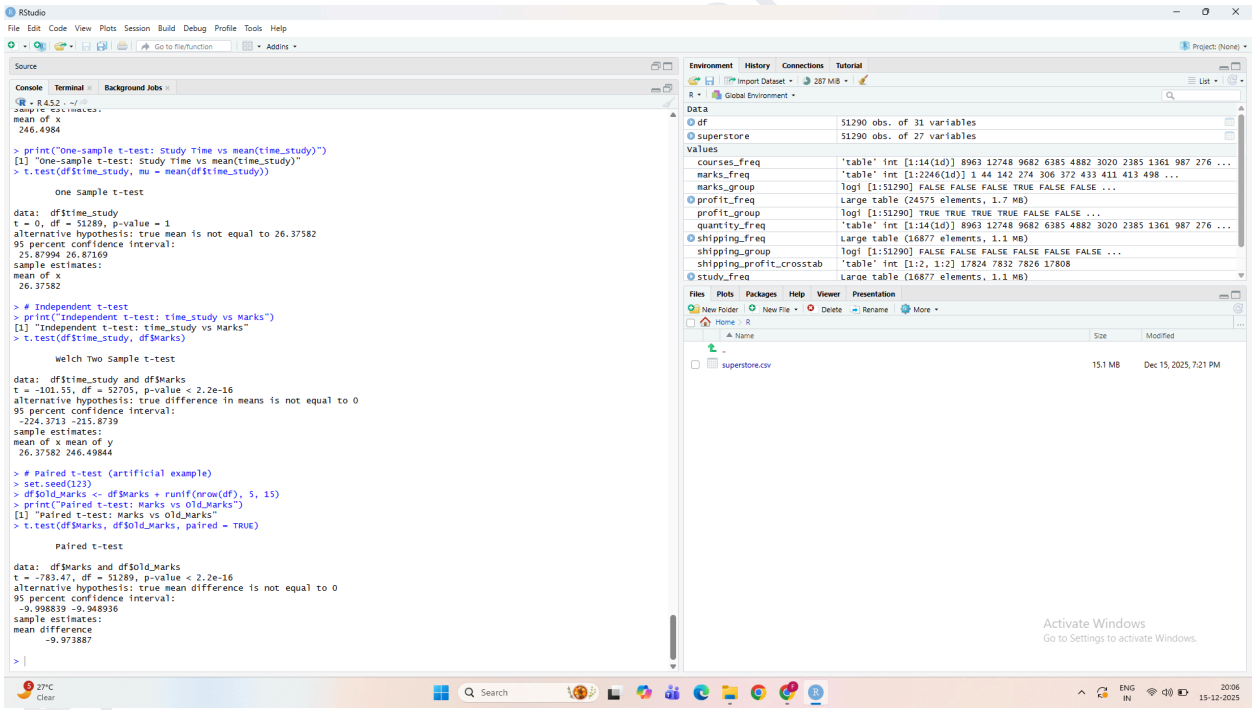


```
  Class :character   Class :character   1st Qu.:0.0000
  Mode  :character   Mode  :character   Median :0.0000
                                        Mean   :0.1429
                                        3rd Qu.:0.2000
                                        Max.   :0.8500
    Market              记录数      Order.Date
 Length:51290       Min.   :1   Min.   :2011-01-01 00:00:00
 Class :character   1st Qu.:1   1st Qu.:2012-06-19 00:00:00
 Mode  :character   Median :1   Median :2013-07-08 00:00:00
                    Mean   :1   Mean   :2013-05-11 21:26:49
                    3rd Qu.:1   3rd Qu.:2014-05-22 00:00:00
                    Max.   :1   Max.   :2014-12-31 00:00:00
  Order.ID          Order.Priority      Product.ID
 Length:51290       Length:51290       Length:51290
 Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character


  Product.Name          Profit             Quantity
 Length:51290       Min.   :-6599.98   Min.   : 1.000
 Class :character   1st Qu.:    0.00   1st Qu.: 2.000
 Mode  :character   Median :    9.24   Median : 3.000
                    Mean   :   28.61   Mean   : 3.477
                    3rd Qu.:   36.81   3rd Qu.: 5.000
                    Max.   : 8399.98   Max.   :14.000
    Region            Row.ID           Sales
 Length:51290       Min.   :    1   Min.   :    0.0
 Class :character   1st Qu.:12823   1st Qu.:   31.0
 Mode  :character   Median :25646   Median :   85.0
                    Mean   :25646   Mean   :  246.5
                    3rd Qu.:38468   3rd Qu.:  251.0
                    Max.   :51290   Max.   :22638.0
   Segment            Ship.Date                  Ship.Mode
 Length:51290       Min.   :2011-01-03 00:00:00   Length:51290
 Class :character   1st Qu.:2012-06-23 00:00:00   Class :character
 Mode  :character   Median :2013-07-12 00:00:00   Mode  :character
                    Mean   :2013-05-15 20:42:42
                    3rd Qu.:2014-05-26 00:00:00
                    Max.   :2015-01-07 00:00:00
 Shipping.Cost         State            Sub.Category            Year
 Min.   :  0.002   Length:51290       Length:51290       Min.   :2011
 1st Qu.:  2.610   Class :character   Class :character   1st Qu.:2012
 Median :  7.790   Mode  :character   Mode  :character   Median :2013
 Mean   : 26.376                                         Mean   :2013
 3rd Qu.: 24.450                                         3rd Qu.:2014
 Max.   :933.570                                         Max.   :2014
   Market2              weeknum
 Length:51290       Min.   : 1.00
 Class :character   1st Qu.:20.00
 Mode  :character   Median :33.00
                    Mean   :31.29
                    3rd Qu.:44.00
                    Max.   :53.00
```



```
> df$time_study      <- as.numeric(df$Shipping.Cost)
> df$Marks           <- as.numeric(df$Sales)
> # Check missing values
> sum(is.na(df$number_courses))
[1] 0
> sum(is.na(df$time_study))
[1] 0
> sum(is.na(df$Marks))
[1] 0
> # Remove rows with missing values
> df <- df %>%
+   filter(
+     !is.na(number_courses),
+     !is.na(time_study),
+     !is.na(Marks)
+   )
> # summaries
> summary(df$number_courses)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   3.000   3.477   5.000  14.000
> summary(df$time_study)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.002   2.610   7.790  26.376  24.450 933.570
> summary(df$Marks)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0    31.0    85.0   246.5   251.0 22638.0
> # Frequency tables
> courses_freq <- table(df$number_courses)
> print("Frequency Table: Number of Courses")
[1] "Frequency Table: Number of Courses"
> print(courses_freq)

    1     2     3     4     5     6     7     8     9    10    11    12
 8963 12748  9682  6385  4882  3020  2385  1361   987   276   156   176
   13    14
   83   186
> study_freq <- table(df$time_study)
> print("Frequency Table: Time Studied (unique values)")
[1] "Frequency Table: Time Studied (unique values)"
> print(study_freq)

0.002 0.003  0.01 0.019  0.02 0.023 0.025  0.03 0.039  0.04 0.042  0.05
    1     1     6     1     6     2     1     8     1    12     1    18
0.053 0.058  0.06 0.064  0.07 0.074 0.077  0.08 0.084  0.09 0.094   0.1
    1     1    16     1    12     1     1    15     1    22     1    32
0.105 0.106 0.107  0.11 0.111 0.117  0.12 0.123 0.125 0.127 0.129  0.13
    2     1     1    22     1     1    22     1     1     1     1    20
0.131 0.132 0.136  0.14 0.141 0.142  0.15 0.151 0.152 0.153 0.154 0.155
    1     1     1    34     1     1    24     1     1     1     2     1
0.156  0.16 0.165 0.166  0.17 0.171 0.173 0.175 0.178 0.179  0.18 0.182
    1    32     1     1    46     1     1     1     4     1    41     1
0.184 0.186 0.189  0.19 0.192 0.193 0.195 0.196 0.197 0.199   0.2 0.202
    1     1     1    32     1     1     1     1     1     1    38     1
0.204 0.206 0.207 0.208 0.209  0.21 0.212 0.213  0.22 0.222 0.223 0.227
```


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
4 8 5 9 4 4 3 4 2 2 2 2 5 8
1001 1002 1003 1004 1005 1006
6 8 9 4 2 2
[reached 'max' / getOption("max.print") -- omitted 1246 entries]
> # Cross-tab: high vs low study time and high vs low marks
> study_group <- df$time_study > median(df$time_study)
> marks_group <- df$Marks > median(df$Marks)
> study_marks_crosstab <- table(study_group, marks_group)
> print("Cross-tab: Study Time Group vs Marks Group")
[1] "Cross-tab: Study Time Group vs Marks Group"
> print(study_marks_crosstab)
marks_group
study_group FALSE TRUE
FALSE 22646 3004
TRUE 3070 22570
> # One-sample t-tests
> print("One-sample t-test: Marks vs mean(Marks)")
[1] "One-sample t-test: Marks vs mean(Marks)"
> t.test(df$Marks, mu = mean(df$Marks))
One Sample t-test
data: df$Marks
t = 0, df = 51289, p-value = 1
alternative hypothesis: true mean is not equal to 246.4984
95 percent confidence interval:
242.2788 250.7181
sample estimates:
mean of x
246.4984
> print("One-sample t-test: Study Time vs mean(time_study)")
[1] "One-sample t-test: Study Time vs mean(time_study)"
> t.test(df$time_study, mu = mean(df$time_study))
One Sample t-test
data: df$time_study
t = 0, df = 51289, p-value = 1
alternative hypothesis: true mean is not equal to 26.37582
95 percent confidence interval:
25.87994 26.87169
sample estimates:
mean of x
26.37582
> # Independent t-test
> print("Independent t-test: time_study vs Marks")
[1] "Independent t-test: time_study vs Marks"
> t.test(df$time_study, df$Marks)
Welch Two Sample t-test
data: df$time_study and df$Marks
Environment History Connections Tutorial
Import Dataset 287 MiB
Global Environment
Data
df 51290 obs. of 31 variables
superstore 51290 obs. of 27 variables
values
courses_freq 'table' int [1:14(1d)] 8963 12748 9682 6385 4882 3020 2385 1361 987 276 ...
marks_freq 'table' int [1:2246(1d)] 1 44 142 274 306 372 433 411 413 498 ...
marks_group logi [1:51290] FALSE FALSE FALSE TRUE FALSE FALSE ...
profit_freq Large table (24575 elements, 1.7 MB)
profit_group logi [1:51290] TRUE TRUE TRUE TRUE FALSE FALSE ...
quantity_freq 'table' int [1:14(1d)] 8963 12748 9682 6385 4882 3020 2385 1361 987 276 ...
shipping_freq Large table (16877 elements, 1.1 MB)
shipping_group logi [1:51290] FALSE FALSE FALSE FALSE FALSE FALSE ...
shipping_profit_crosstab 'table' int [1:2, 1:2] 17824 7832 7826 17808
study_freq Large table (16877 elements, 1.1 MB)
Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home R
Name Size Modified
superstore.csv 15.1 MB Dec 15, 2025, 7:21 PM
Activate Windows
Go to Settings to activate Windows.
27°C Clear
Search
ENG IN
20:05 15-12-2025


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Console Terminal Background Jobs
mean of x
246.4984
> print("One-sample t-test: Study Time vs mean(time_study)")
[1] "One-sample t-test: Study Time vs mean(time_study)"
> t.test(df$time_study, mu = mean(df$time_study))
One Sample t-test
data: df$time_study
t = 0, df = 51289, p-value = 1
alternative hypothesis: true mean is not equal to 26.37582
95 percent confidence interval:
25.87994 26.87169
sample estimates:
mean of x
26.37582
> # Independent t-test
> print("Independent t-test: time_study vs Marks")
[1] "Independent t-test: time_study vs Marks"
> t.test(df$time_study, df$Marks)
Welch Two Sample t-test
data: df$time_study and df$Marks
t = -101.55, df = 52705, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-224.3713 -215.8739
sample estimates:
mean of x mean of y
26.37582 246.49844
> # Paired t-test (artificial example)
> set.seed(123)
> df$Old_Marks <- df$Marks + runif(nrow(df), 5, 15)
> print("Paired t-test: Marks vs Old_Marks")
[1] "Paired t-test: Marks vs Old_Marks"
> t.test(df$Marks, df$Old_Marks, paired = TRUE)
Paired t-test
data: df$Marks and df$Old_Marks
t = -783.47, df = 51289, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
-9.998839 -9.948936
sample estimates:
mean difference
-9.973887
Environment History Connections Tutorial
Import Dataset 287 MiB
Global Environment
Data
df 51290 obs. of 31 variables
superstore 51290 obs. of 27 variables
values
courses_freq 'table' int [1:14(1d)] 8963 12748 9682 6385 4882 3020 2385 1361 987 276 ...
marks_freq 'table' int [1:2246(1d)] 1 44 142 274 306 372 433 411 413 498 ...
marks_group logi [1:51290] FALSE FALSE FALSE TRUE FALSE FALSE ...
profit_freq Large table (24575 elements, 1.7 MB)
profit_group logi [1:51290] TRUE TRUE TRUE TRUE FALSE FALSE ...
quantity_freq 'table' int [1:14(1d)] 8963 12748 9682 6385 4882 3020 2385 1361 987 276 ...
shipping_freq Large table (16877 elements, 1.1 MB)
shipping_group logi [1:51290] FALSE FALSE FALSE FALSE FALSE FALSE ...
shipping_profit_crosstab 'table' int [1:2, 1:2] 17824 7832 7826 17808
study_freq Large table (16877 elements, 1.1 MB)
Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home R
Name Size Modified
superstore.csv 15.1 MB Dec 15, 2025, 7:21 PM
Activate Windows
Go to Settings to activate Windows.
27°C Clear
Search
ENG IN
20:06 15-12-2025