

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

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R

MODULE 2 - PRACTICAL – 5

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

OUTPUT:-

```
raji - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins +
Source
Console Terminal Background Jobs ×
R - R.4.1.2 - /raji/...
+ }
[1] "Reject the null hypothesis"
> library(dplyr)
>
> # 1. Load dataset
> df <- read.csv("C:\\Users\\IT\\Downloads\\pizza_sales.csv")
> print("Dataset Loaded Successfully")
[1] "Dataset Loaded Successfully"
>
> # 2. Dataset overview
> head(df)
#> #> #> pizza_id order_id pizza_name_id quantity order_date order_time unit_price total_price pizza_size pizza_category
#> #> 1 1 hawaiian_m 1 1/1/2015 11:38:36 13.25 13.25 M classic
#> #> 2 2 classic_dlx_m 1 1/1/2015 11:57:40 16.00 16.00 M classic
#> #> 3 3 5ive_cheese_l 1 1/1/2015 11:57:40 18.50 18.50 L veggie
#> #> 4 4 2 ital_suprl 1 1/1/2015 11:57:40 20.75 20.75 L Supreme
#> #> 5 5 2 mexicana_m 1 1/1/2015 11:57:40 16.00 16.00 M veggie
#> #> 6 6 2 thai_chkn_l 1 1/1/2015 11:57:40 20.75 20.75 L chicken
#> #> pizza_ingredients pizza_name
#> #> 1 Sliced Ham, Pineapple, Mozzarella Cheese The Hawaiian Pizza
#> #> 2 Pepperoni, Mushrooms, Red Onions, Red Peppers, Bacon The Classic Deluxe Pizza
#> #> 3 Mozzarella Cheese, Provolone Cheese, Smoked Gouda Cheese, Romano Cheese, Blue Cheese, Garlic The Five Cheese Pizza
#> #> 4 Calabrese Salami, Capicotto, Tomatoes, Red Onions, Green Olives, Garlic The Italian Supreme Pizza
#> #> 5 Tomatoes, Red Peppers, Jalapeno Peppers, Red Onions, Cilantro, Corn, chipotle Sauce, Garlic The Mexicana Pizza
#> #> 6 Chicken, Pineapple, Tomatoes, Red Peppers, Thai Sweet Chilli Sauce The Thai Chicken Pizza
> str(df)
'data.frame': 48620 obs. of 12 variables:
 $ pizza_id   : num 1 2 3 4 5 6 7 8 9 10 ...
 $ order_id    : num 1 2 2 2 2 2 3 3 4 5 ...
 $ pizza_name_id: chr "hawaiian_m" "classic_dlx_m" "five_cheese_l" "ital_suprl" ...
 $ quantity    : num 1 1 1 1 1 1 1 1 1 1 ...
 $ order_date   : chr "1/1/2015" "1/1/2015" "1/1/2015" "1/1/2015" ...
 $ order_time   : chr "11:38:36" "11:57:40" "11:57:40" "11:57:40" ...
 $ unit_price   : num 13.2 16 18.5 20.8 16 ...
 $ total_price  : num 13.2 16 18.5 20.8 16 ...
 $ pizza_size   : chr "M" "L" "L" "L" ...
 $ pizza_category: chr "classic" "classic" "veggie" "Supreme"
 $ pizza_ingredients: chr "Sliced Ham, Pineapple, Mozzarella Cheese", "Pepperoni, Mushrooms, Red Onions, Red Peppers, Bacon", "Mozzarella cheese, Provolone Cheese, Smoked Gouda Cheese, Romano cheese, Blue Cheese, Garlic", "Calabrese Salami, Capicotto, Tomatoes, Red Onions, Green Olives, Garlic" ...
 $ pizza_name   : chr "The Hawaiian Pizza" "The Classic Deluxe Pizza" "The Five Cheese Pizza" "The Italian Supreme Pizza" ...
> dim(df)
[1] 48620 12
```

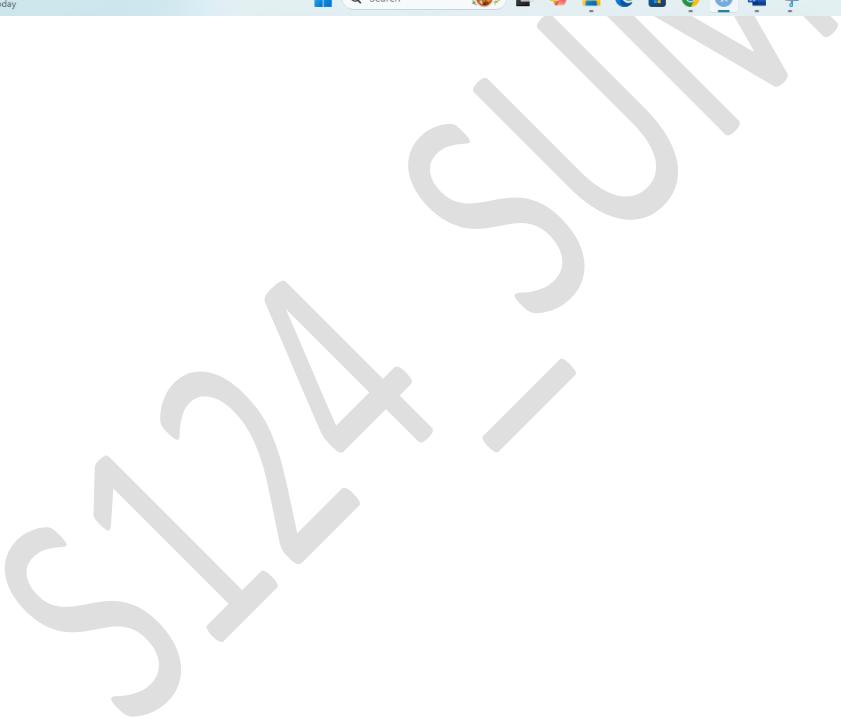
The screenshot shows an RStudio interface with a script named `raji - RStudio`. The code performs several steps:

- Imports the `pizza` dataset.
- Creates a variable `df` containing the first 10 rows of the dataset.
- Prints the first 10 rows of the `df` dataset.
- Creates a variable `dfsize_group` by grouping the `df` dataset by size.
- Prints the distribution of sizes.
- Creates a summary table of unit price by size group.
- Performs an independent two-sample t-test comparing the mean unit price of Large/XL and Small/Medium pizzas.

```
R4.12 - [file]@raji ~
> pizza
#> # A tibble: 14,013 x 10
#>   ingredients     crust toppings    cheese
#>   <chr>           <chr>  <chr>       <chr>
#> 1 Sliced Ham, Pineapple, Mzzarella Cheese Pepperoni, Mushrooms, Red Onions, Red Peppers, Bacon, Mzzarella Cheese, Provolone Cheese, Smoked Gouda Cheese, Roasted Red Peppers, Blue Cheese, Garlic "Calabrese Salami, Capicolla, Tomatoes, Red Onions, Green Olives, Garlic" ...
#> 2 Pizza_name      : chr "The Hawaiian Pizza" "The Classic Deluxe Pizza" "The Five Cheese Pizza" "The Italian Supreme Pizza" ...
#> 3 diameter       : num 12
#> 4
#> # 3. Create Grouping Variable (IMPORTANT)
#> dfsize_group <- ifelse(
#> +   df$pizza_size %in% c("S", "M"),
#> +   "Small/Medium",
#> +   "Large/XL"
#> + )
#> print("Group Distribution:")
#> [1] "Group Distribution:"
#> table(dfsize_group)
#>
#> Large/XL Small/Medium
#> 19098      29522
#>
#> # 4. Summary Statistics by Group
#> # Using unit_price as numeric variable
#> df %>%
#>   group_by(size_group) %>%
#>   summarise(
#>     Mean_Price = mean(unit_price, na.rm = TRUE),
#>     SD_Price = sd(unit_price, na.rm = TRUE),
#>     count = n()
#>   )
#> # A tibble: 2 x 4
#>   size_group  Mean_Price SD_Price count
#>   <chr>        <dbl>     <dbl>  <int>
#> 1 Large/XL      20.0      1.88   19098
#> 2 Small/Medium   14.2      2.51   29522
#>
#> # 5. Independent Two-Sample t-test
#> # H0: Mean unit price of both size groups is equal
#> # H1: Mean unit price of both size groups is not equal
#>
#> print("Independent Two-Sample t-test Result:")
#> [1] "Independent Two-Sample t-test Result:"
#> t_test_result <- t.test(unit_price ~ size_group, data = df)
```

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

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R



```
raji - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
File ~ Go to file/function Addins
Source
Console Terminal Background Jobs
R 4.1.2 - ~/r/j...
+ group_by(size_group) %>
+ summarise(
+   Mean_Price = mean(unit_price, na.rm = TRUE),
+   SD_Price = sd(unit_price, na.rm = TRUE),
+   Count = n()
+ )
# A tibble: 2 × 4
  size_group  Mean_Price SD_Price Count
<chr>          <dbl>     <dbl>  <int>
1 Large/XL      20.0     1.88 19998
2 Small/Medium  14.2     2.51 2922
> 
> # 5. Independent Two-sample t-test
> # H0: Mean unit price of both size groups is equal
> # H1: Mean unit price of both size groups is not equal
> print("Independent Two-Sample t-test Result:")
[1] "Independent Two-Sample t-test Result:"
> 
> t_test_result <- t.test(unit_price ~ size_group, data = df)
> print(t_test_result)

Welch Two Sample t-test

data: unit_price by size_group
t = 248.4, df = 747, p-value < 2.2e-16
alternative hypothesis: true difference in means between group Large/XL and group Small/Medium is not equal to 0
95 percent confidence interval:
 5.715817 5.794040
sample estimates:
mean in group Large/XL mean in group Small/Medium
         19.98852              14.23359

> 
> # 6. Decision Based on p-value
> if (t_test_result$p.value < 0.05) {
+   print("Reject the null hypothesis")
+ } else {
+   print("Fail to reject the null hypothesis")
+ }
[1] "Reject the null hypothesis"
> print(head(Retail_Product))
[1] "Air Poor Today"
```