**Question 1**

**Syntax:**

# Check the file exist in working directory

if(!file.exists("ME\_DA\_Test.txt")){

# Read the the file if not available in the working directory

row\_data <- read.table(file.choose(), sep = "\t", header = TRUE)

head(row\_data)

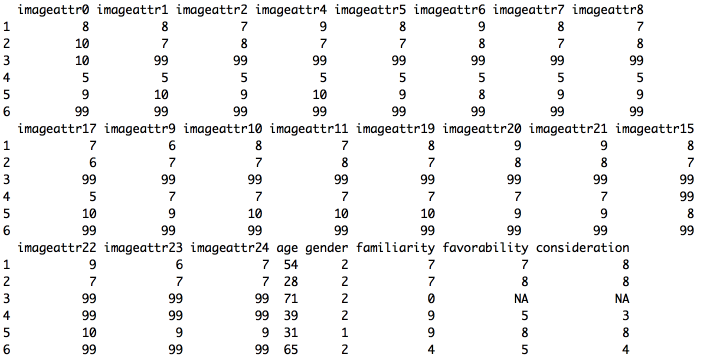
}else{

row\_data <- read.table("ME\_DA\_Test.txt", sep = "\t", header = TRUE)

head(row\_data)

}

**Output:**



**Question 2**

**Syntax:**

# function to identify age category

age\_indicator\_function <- function(data,col){

     no\_row = nrow(data);

     new\_col = c()

     for(i in 1:no\_row)

     {

         if((data[i,col] >= 18) && (data[i,col] <= 24))

         {

             new\_col[i] <- "Age 18-24"

         }

         else if((data[i,col] >= 25) && (data[i,col] <= 44))

         {

             new\_col[i] <- "Age 18-24"

         }

         else if((data[i,col] >= 45) && (data[i,col] <= 64))

         {

             new\_col[i] <- "Age 18-24"

         }

         else

         {

             new\_col[i] <- "Age 18-24"

         }

     }

     return (new\_col)

 }

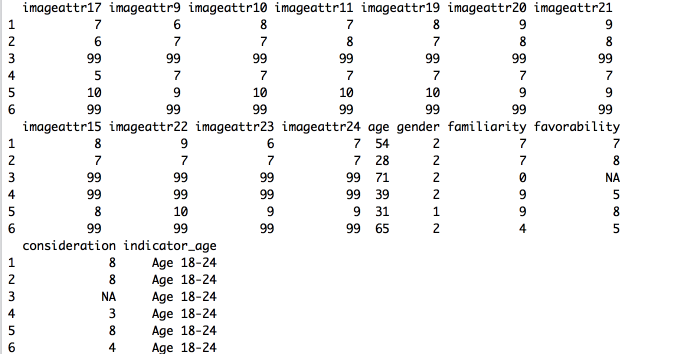
# function call

indicator\_age <- age\_indicator\_function(row\_data,"age")

row\_data <- cbind(row\_data,indicator\_age) #Attach new column to data using cbind function

head(row\_data)

**Output:**



**Question 3**

**Syntax (Function 1 to 3)**

**Function 1: max3\_values function**

#Returns top 3 box

max3\_values <- function(dataset)

{

    # Return top 4 values in specified database  
 max\_3 <- head(unique(sort(dataset, decreasing = TRUE, index.return = FALSE)),4)

   # print(max\_3)

# check correct value.

    if (any(max\_3==99))

    {

        #print("99 is avalable")

        max\_3 <- max\_3[max\_3 != 99]

        #print(max\_3)

    }else{

        #print("99 is not avalable")

        max\_3 <- head(max\_3,3)

        #print(max\_3)

    }

}

**Function 3: check\_top\_3 function**

# Returns the respond value is top 3 box

check\_top\_3 <- function(respondent\_value,data\_set)

{

    top\_3 <- max3\_values(data\_set)

    if(any(top\_3 == respondent\_value))

    {

        return(1)

    }

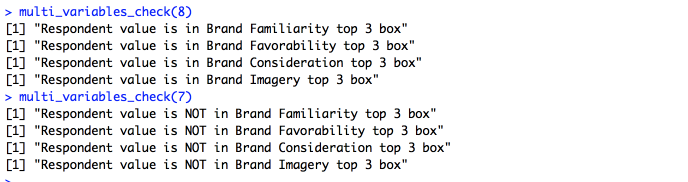
    else{

        return(0)

    }

}

**Output:**



**Function 3**: **multi\_variable\_ckeck function**

# Multi variable check function

multi\_variables\_check <- function(respondent\_value)

{

    #Brand Familiarity

    check\_brand\_familiarity <- check\_top\_3(respondent\_value, row\_data$familiarity)

    if(check\_brand\_familiarity)

    {print("Respondent value is in Brand Familiarity top 3 box")}

else

{print("Respondent value is NOT in Brand Familiarity top 3 box") }

    #Brand Favorability

    check\_brand\_favorability <- check\_top\_3(respondent\_value, row\_data$favorability)

    if(check\_brand\_favorability)

    { print("Respondent value is in Brand Favorability top 3 box")}

else  
 {print("Respondent value is NOT in Brand Favorability top 3 box")}

    #Brand Consideration

    check\_brand\_consideration <- check\_top\_3(respondent\_value, row\_data$consideration)

    if(check\_brand\_consideration)

    { print("Respondent value is in Brand Consideration top 3 box")}

else

{ print("Respondent value is NOT in Brand Consideration top 3 box") }

   #Brand Imagery

   imageatt0 <- check\_top\_3(respondent\_value, row\_data$imageattr0)

   imageatt1 <- check\_top\_3(respondent\_value, row\_data$imageattr1)

   imageatt2 <- check\_top\_3(respondent\_value, row\_data$imageattr2)

   imageatt4 <- check\_top\_3(respondent\_value, row\_data$imageattr4)

   imageatt5 <- check\_top\_3(respondent\_value, row\_data$imageattr5)

   imageatt6 <- check\_top\_3(respondent\_value, row\_data$imageattr6)

   imageatt7 <- check\_top\_3(respondent\_value, row\_data$imageattr7)

   imageatt8 <- check\_top\_3(respondent\_value, row\_data$imageattr8)

   imageatt9 <- check\_top\_3(respondent\_value, row\_data$imageattr9)

   imageatt10 <- check\_top\_3(respondent\_value, row\_data$imageattr10)

   imageatt11 <- check\_top\_3(respondent\_value, row\_data$imageattr11)

   imageatt15 <- check\_top\_3(respondent\_value, row\_data$imageattr15)

   imageatt17 <- check\_top\_3(respondent\_value, row\_data$imageattr17)

   imageatt19 <- check\_top\_3(respondent\_value, row\_data$imageattr19)

   imageatt20 <- check\_top\_3(respondent\_value, row\_data$imageattr20)

   imageatt21 <- check\_top\_3(respondent\_value, row\_data$imageattr21)

  imageatt22 <- check\_top\_3(respondent\_value, row\_data$imageattr22)

  imageatt23 <- check\_top\_3(respondent\_value, row\_data$imageattr23)

  imageatt24 <- check\_top\_3(respondent\_value, row\_data$imageattr24)

  if (imageatt0 || imageatt1 || imageatt2 || imageatt4 || imageatt5 || imageatt6 || imageatt7 || imageatt8 || imageatt9 || imageatt10 || imageatt11 || imageatt15 || imageatt17 || imageatt19 || imageatt20 || imageatt21 || imageatt22 || imageatt23 || imageatt24 )

 { print("Respondent value is in Brand Imagery top 3 box") }

else

{ print("Respondent value is NOT in Brand Services top 3 box")}

}

multi\_variables\_check(8)

multi\_variables\_check(7)

**Question 4**

# get top 3 values  
top\_3\_values <- function(data\_set){

return(head(unique(sort(data\_set, decreasing = TRUE, index.return = FALSE)),3))

}  
# Assumption 1: samples are independent

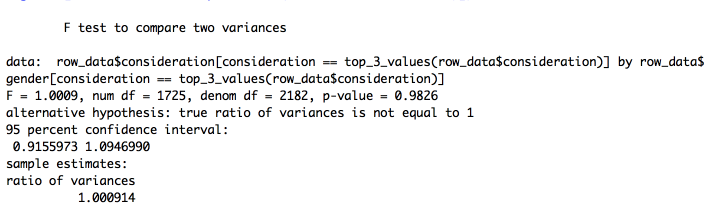
# Assumption 2: Given sample data equal variance

# H0 : variance for both male and female have selected top 3 box consideration box value is same.

# H1 : variance for both male and female have selected top 3 box consideration box value is NOT same.

var.test(row\_data$consideration[consideration == top\_3\_values(row\_data$ consideration)]~row\_data$gender[consideration == top\_3\_values(row\_data$ consideration)])

For this problem, we need to compare difference between male and female So we can either use ANOVA test or T test. I have used T test to compare significance difference.   
Case 1: Significant difference between means in both male and female have selected values top 3 box consideration box.



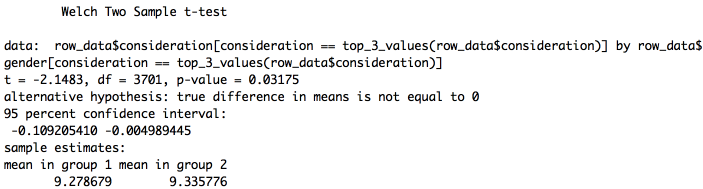
Explanation, As, P value is greater than 0.05 so we fail to reject null hypothesis, in other word variance for both male and female have selected top 3 box consideration box value are equally accepted.

# Independent sample test

#H0 : There is no significant difference between means in both male and female have selected values top 3 box consideration box.

#H1 : There is significant difference between means in both male and female have selected values top 3 box consideration box.

t.test(row\_data$consideration[consideration == top\_3\_values(row\_data$ consideration)]~row\_data$gender[consideration == top\_3\_values(row\_data$ consideration)])



p-value is less than 0.05, we reject the null hypothesis that there's no difference between the means in both male and female have selected values top 3 box consideration box and conclude that There is significant difference between means in both male and female have selected values top 3 box consideration box.

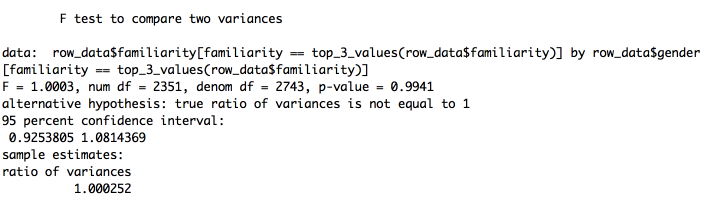
Case 2: significant difference between means in both male and female have selected values top 3 box familiarity box.

# Assumption 1: Given sample data equal variance

# H0 : variance for both male and female have selected top 3 box familiarity box value is same.

# H1 : variance for both male and female have selected top 3 box familiarity box value is NOT same.

var.test(row\_data$familiarity[familiarity == top\_3\_values(row\_data$familiarity)]~row\_data$gender[familiarity == top\_3\_values(row\_data$familiarity)])



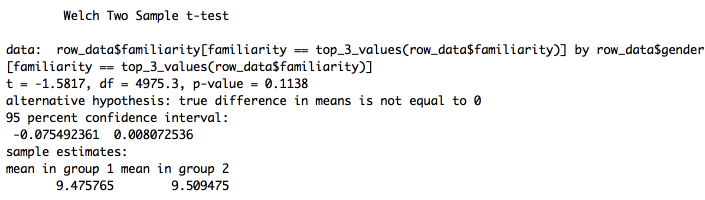
As, P value is greater than 0.05 so we fail to reject null hypothesis, in other word variance for both male and female have selected top 3 box familiarity box values are equally accepted.

# Independent sample test

# H0 : There is no significant difference between means in both male and female have selected values top 3 box familiarity box.

# H1: There is significant difference between means in both male and female have selected values top 3 box familiarity box.

t.test(row\_data$familiarity[familiarity == top\_3\_values(row\_data$familiarity)]~row\_data$gender[familiarity == top\_3\_values(row\_data$familiarity)])



p-value is greater than 0.05, we keep the null hypothesis that there's no significant difference between the means in both male and female have selected values top 3 box familiarity box.

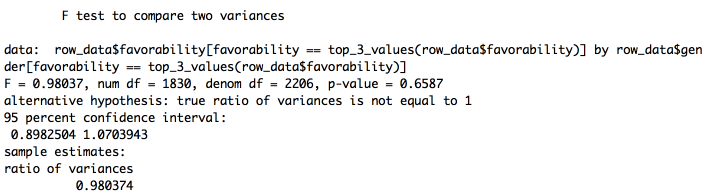
**Case 3**: significant difference between means in both male and female have selected values top 3 box favorability box.

# Assumption 1: Given sample data equal variance

#H0 : There is no significant difference between means in both male and female have selected values top 3 box favorability box.

#H1 : There is significant difference between means in both male and female have selected values top 3 box favorability box.

var.test(row\_data$favorability[favorability == top\_3\_values(row\_data$favorability)]~row\_data$gender[favorability == top\_3\_values(row\_data$favorability)])



As, P value is greater than 0.05 so we fail to reject null hypothesis, in other word variance for both male and female have selected top 3 box familiarity box values are equally accepted.

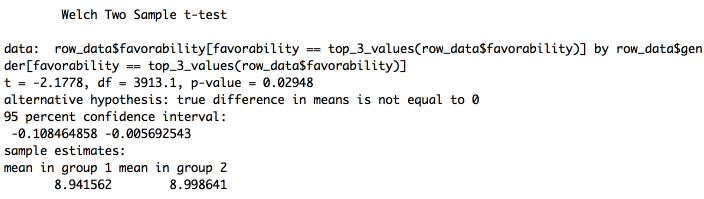
# Independent sample test

#H0 : Both male and female have selected values top 3 box favorability box are same.

#H1 : Both male and female have selected values top 3 box favorability box value are NOT same.

#Welch Two Sample t-test

t.test(row\_data$favorability[favorability == top\_3\_values(row\_data$favorability)]~row\_data$gender[favorability == top\_3\_values(row\_data$favorability)])



p-value is less than 0.05, we reject the null hypothesis that there's no difference between the means in both male and female have selected values top 3 box favorability box and conclude that There is significant difference between means in both male and female have selected values top 3 box favorability box.

**Question 5**

The aim of this study is to find significant relationship between improve overall Brand Consideration in a young audience with Improving brand attitudes by proving or disproving the marketing team’s hypothesis that improving brand image and attitudes among the young audience will have better impact on Consideration.

Assumption 1: The young audience consider age group between 18 to 24 ages.

Assumption 2: The top three boxes are calculated by arranging records in descending order and get top three values.

Assumption 3: Consideration is dependent variable and familiarity and favorability are independent variable in above given problem.

Find correlation between the dependent variable independent variable.

cor(data\_filter$consideration,data\_filter$favorability+data\_filter$familiarity)

Output:

[1] 1

Correlation can take values between -1 to +1. The value is 1 suggest a strong relationship consideration with the favorability and familiarity in young abundance.

Now, find significant relationship between the familiarity and favorability with consideration in young audience. So the following hypothesis need to prove using liner regression analysis.

H0: There is no relationship between overall Brand Consideration in a young audience with Improving brand attitudes

H1: There is no relationship between overall Brand Consideration in a young audience with Improving brand attitudes

Step 1: Subset data as per required column familiarity, favorability, consideration, indicator\_age

Step 2: Filter data with top three box and age group between Age 18-24

Step 3: Perform liner regression model on filtered data.

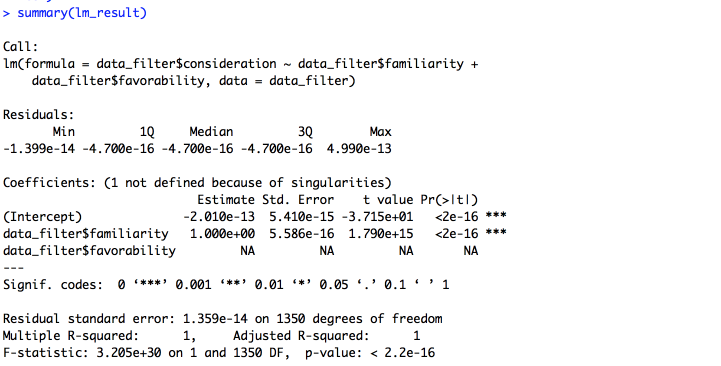
Step 4: Analyze result

data\_filter <- subset(row\_data, select = c(familiarity,favorability,consideration,indicator\_age))  
data\_filter <- subset(data\_filter,consideration == top\_3\_values(data\_filter$consideration) & familiarity == top\_3\_values(data\_filter$familiarity) & indicator\_age == "Age 18-24" & favorability == top\_3\_values(data\_filter$favorability))

lm\_result <- lm(data\_filter$consideration~data\_filter$familiarity+data\_filter$favorability,data=data\_filter)

summary(lm\_result)

Output:



As NA as coefficient indicates that the favorability linearly related to the familiarity. As the p-value is much less than 0.05 for both familiarity and favorability. Secondly t value for the both variable is far away from the zero and closed to Std. Error. So, we reject the null hypothesis that β = 0. Hence there is a significant relationship between the familiarity and favorability with consideration in young audience in the linear regression model of the data set faithful. In other word, **we can say that marketing team’s hypothesis that improving brand image and attitudes among the young audience will have better impact on consideration is right and proven**. The R2 we get is 1. Or roughly 100% of the variance found in the response variable (consideration) can be explained by the predictor variable (familiarity and favorability) in young age group.

**Question 6**

1. Baseline probability of likelihood to purchase is 0%.

**Answer:**

As no value coefficient is zero, baseline probability of likelihood to purchase is not 0%.

2. Younger age groups are more likely to purchase than older age groups.

**Answer:**

As t value coefficient is negative, there is decrease in likelihood to purchase with increase

in age. Hence conclusion is true.

3. Probability of Older age group purchasing is exp(-0.33) = 72%.

**Answer:**

Probability of Older age group in logistic regression is 1/1+e^-e  
False, As we are using above formula to calculate probability in logistic regression.

1/(1+exp(-0.33)) = 0.5817594

So actually probability of Older age group is 58.18%

4. Lower income groups are less likely to purchase across the board.

**Answer:**

Negative t value interprets the decrease trend in likelihood to purchase corresponding to

lower income groups.

5. It is better to improve Attribute B than Attribute A.

**Answer:**

True. As T value depends on slope of regression line. More the t value, more is the linear

trend of increase. Attribute B has more t value, so it’s better to improve attribute B.

6. We should move marketing dollars away from males and low-income groups.

**Answer:**

7. Sports TV is effective for females only.

**Answer:**