**Assignment 3**

***Advanced Statistics***

***Roll number:21L-5654***

***Name: Muhammad Hamza Khan***

Consumer Research, Inc. is an independent agency that researches consumer attitudes and behaviors for various firms. In one study, a client asked for an investigation of consumer characteristics that can be used to predict the amount charged by credit card users. Data were collected on annual income, household size, and annual credit card charges for a sample of 50 customers.

Question1 :

1. *Use methods of descriptive statistics to summarize the data. Comment on the findings.*

I will first analyze the three variables separately, to see how the data is distributed.

Answer:

summary(Dataset)

Income\n($1000s) Household\nSize Amount\nCharged ($)

Min. :21.00 Min. :1.00 Min. :1864

1st Qu.:30.25 1st Qu.:2.00 1st Qu.:3130

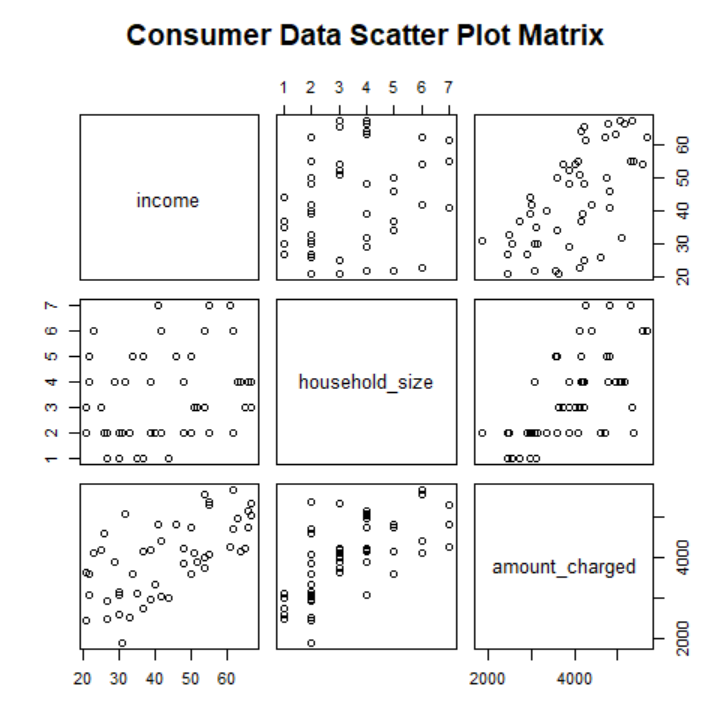
Median :42.00 Median :3.00 Median :4090

Mean :43.48 Mean :3.42 Mean :3964

3rd Qu.:54.75 3rd Qu.:4.75 3rd Qu.:4733

Max. :67.00 Max. :7.00 Max. :5678

Scatter plots:

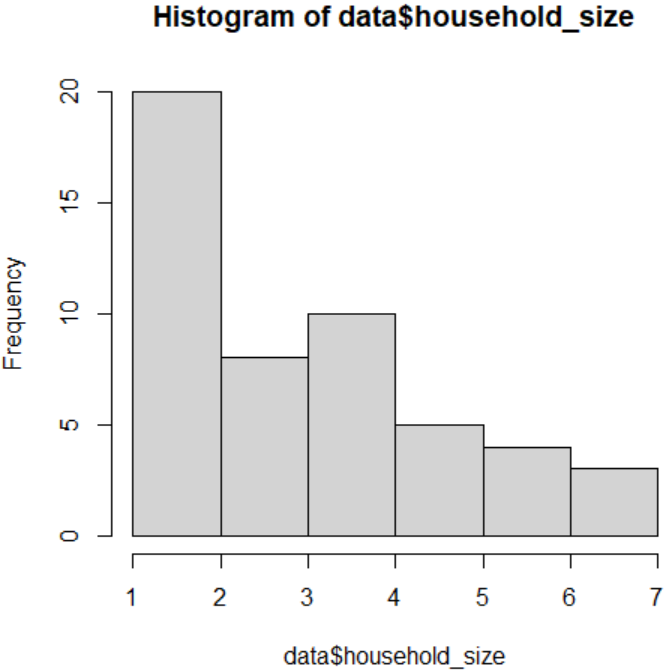
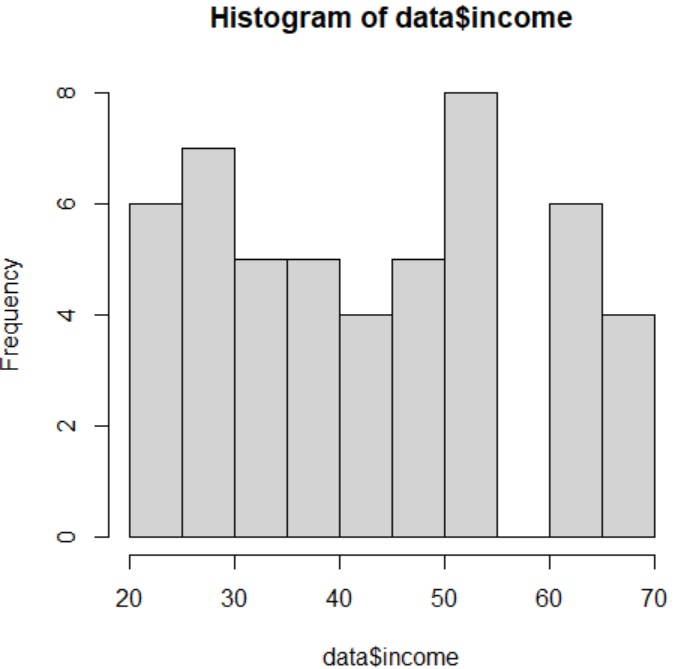


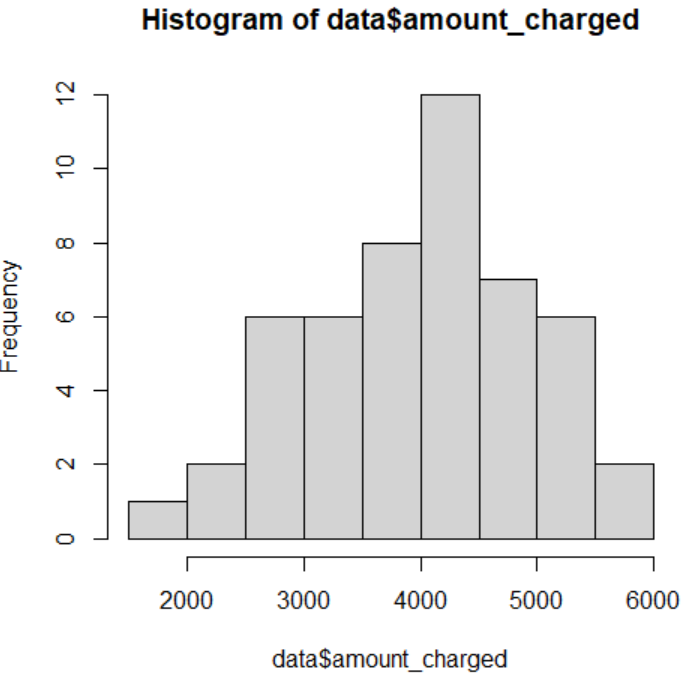
Histograms:

hist(data$household\_size)

hist(data$amount\_charged)

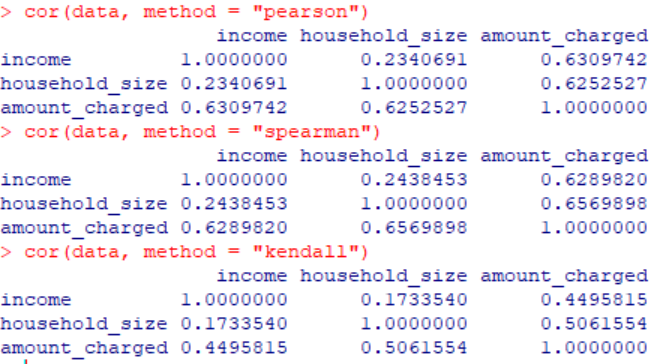
hist(data$income)





Since all the data follows are not having a normal distribution therefore, we apply or use non-parametric tests for correlation.

Correlation:

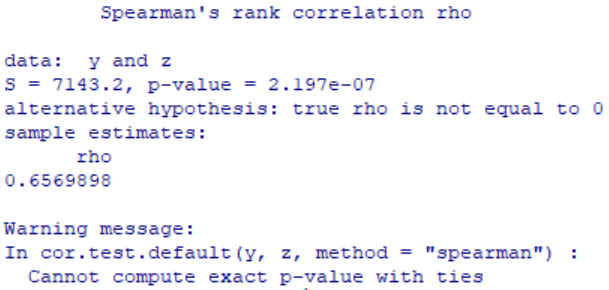
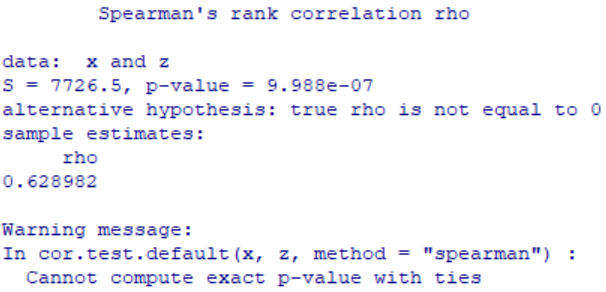
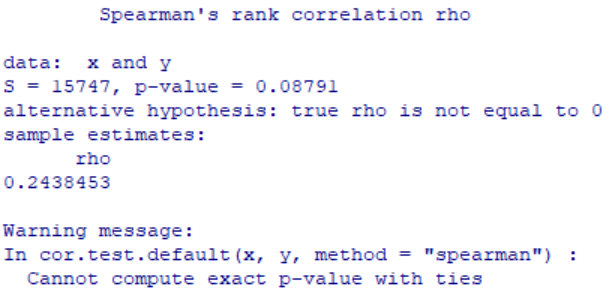


Spearman:

cor.test(x,y, method = "spearman")

cor.test(x,z, method = "spearman")

cor.test(y,z, method = "spearman")



Question 2:

data <- data.frame(

income = c(54, 30, 32, 50, 31, 55, 37, 40, 66, 51, 25, 48, 27, 33, 65, 63, 42, 21, 44, 37, 62, 21, 55, 42, 41, 54, 30, 48, 34, 67, 50, 67, 55, 52, 62, 64, 22, 29, 39, 35, 39, 54, 23, 27, 26, 61, 30, 22, 46, 66),

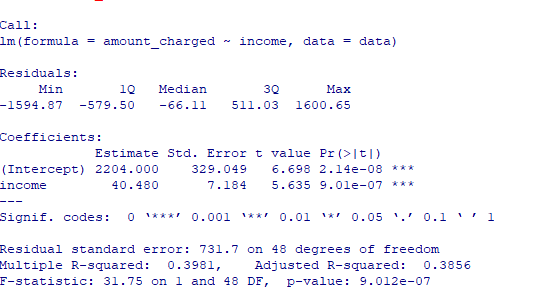
household\_size = c(3, 2, 4, 5, 2, 2, 1, 2, 4, 3, 3, 4, 1, 2, 3, 4, 6, 2, 1, 5, 6, 3, 7, 2, 7, 6, 1, 2, 5, 4, 2, 3, 2, 3, 2, 4, 5, 4, 2, 1, 4, 3, 6, 2, 2, 7, 2, 4, 5, 4),

amount\_charged = c(4016, 3159, 5100, 4742, 1864, 4070, 2731, 3348, 4764, 4110, 4208, 4219, 2477, 2514, 4214, 4965, 4412, 2448, 2995, 4171, 5678, 3623, 5301, 3020, 4828, 5573, 2583, 3866, 3586, 5037, 3605, 5345, 5370, 3890, 4705, 4157, 3579, 3890, 2972, 3121, 4183, 3730, 4127, 2921, 4603, 4273, 3067, 3074, 4820, 5149)

)

lm\_income <- lm(amount\_charged ~ income, data = data)

summary (lm\_income)

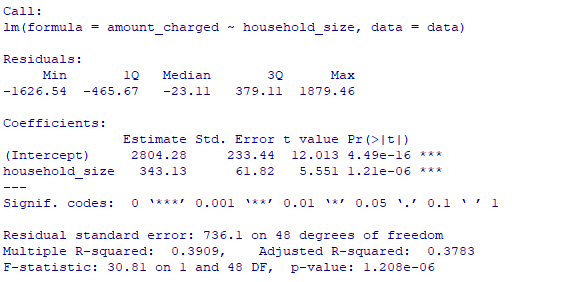


EQUATION:

Amount Charged = 2204 + 40.480 \* Income

lm\_household\_size <- lm(amount\_charged ~ household\_size, data = data)

summary (lm\_household\_size)



EQUATION:

Amount Charged = 2804.28 + 343.13 \* Household Size

By examining the coefficient estimates and p-values, we can determine which variable is a better predictor of annual card charges.

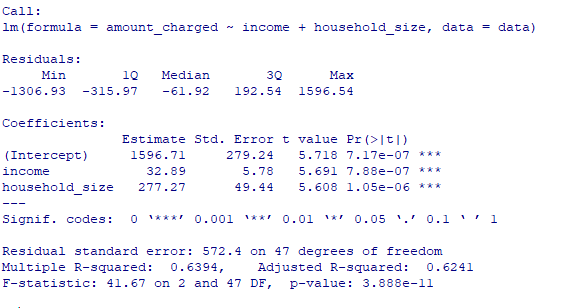
1. Examining the coefficient estimates: Household size has larger coefficient estimates.
2. Examining the p-values: Household size has smaller p value.

Hence, household size is better predictor of annual card charges.

**Question 3:**

lm\_model <- lm(amount\_charged ~ income + household\_size, data = data)

summary(lm\_model)



EQUATION:

Amount Charged = 1596.71 + 32.89 \* Income + 277.27 \* Household Size

**Question4:**

Household size=3

Annual income=40,000

Amount Charged = 1596.71 + 32.89 \* 40000 + 277.27 \* 3

Amount Charged=$1318028.52

**Question 5:**

To enhance the predictive power of a model and gain a comprehensive understanding of the factors influencing credit card charges, it is recommended to consider incorporating additional independent variables. Several variables that can potentially improve the prediction of credit card charges include age, occupation, education level, marital status, credit score, geographic location, debt-to-income ratio, and financial literacy.

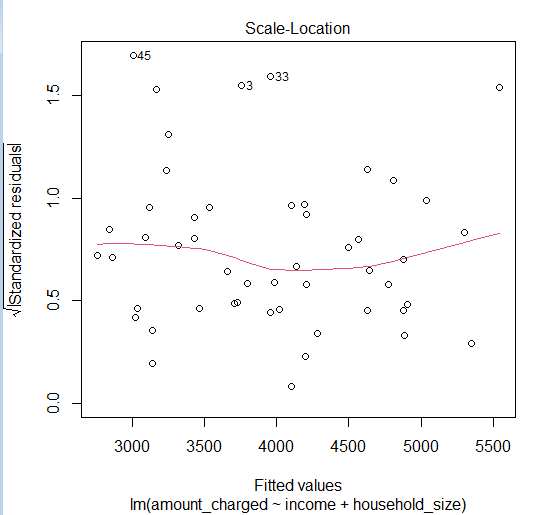
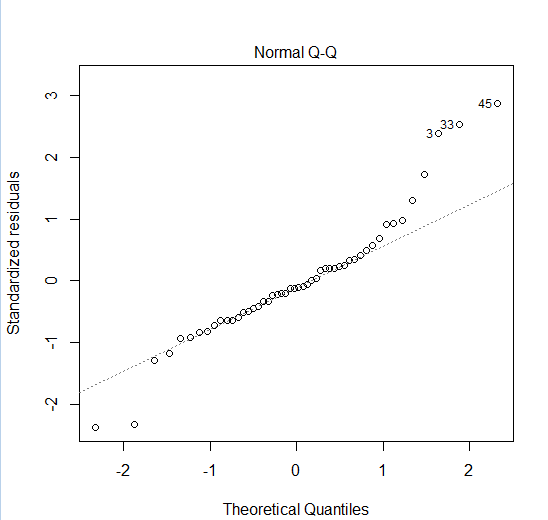
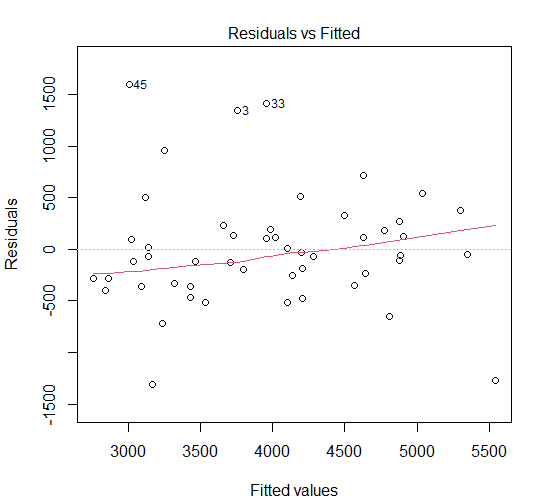
Age is important as spending patterns and financial responsibilities often vary across different age groups. Occupation can impact income levels and spending habits, making it a relevant variable. Education level may correlate with higher incomes and different spending behaviors. Marital status can affect financial commitments and spending patterns.

Credit score reflects creditworthiness and credit history, which can influence credit card charges. Geographic location is significant as consumer behavior and cost of living can vary across regions. Debt-to-income ratio provides insights into an individual's financial obligations and ability to manage credit card charges. Financial literacy and knowledge of personal finance can impact spending behaviors and credit card usage.

It is crucial to select additional variables carefully based on their relevance to the research question and data availability. Additionally, exploring potential interactions between variables and conducting further analysis, such as correlation tests and exploratory data analysis, can help identify the most meaningful independent variables to include in the model.

**Top of Form**

**Question 6:**



The model vs fitted graph shows no clear pattern hence it shows that linear regression is a good fit for the data. It also suggests that there is a positive relationship between household size and annual income and the amount charged by credit card users. The data doesn't follow a normal distribution. The R-squared value of the model is 0.94 which suggests that annual income and household size are important predictors of the amount charged by credit card users.