R Intermediate Project

Remark:

- Save the work using Rmackdown for assessment
- The file name should be your name(s)
- Send the complete project work to codingclub19@gmail.com

You have been asked to study the two data sets provided and answer the questions that follow. You should perform all the statistical analysis required and write up the key findings.

We want to address the following by using the SPSS file **"FinAccess_SPSS.sav"**

1. Import the data to R in csv format (this is SPSS data) and name it financial.csv
2. Conduct exploratory analysis of the data and write at least 5 bullet points on any descriptive statistics (Summary statistics, graphs and tables) you find interesting and why you find them interesting.
3. Create a Cross Table of financial literacy and numeracy
4. Create a new column "average_income" estimated by dividing monthly-income with household_size and perform a one way ANOVA between average_income and credit_usage of the respondent. Does average_income depend on credit_usage?
5. Perform a post hoc test for number 5 and comment on the results
6. Does the data in average_income column pass the normality test? Explain
7. Assuming that the monthly_income of the respondents increased by 12% in the following year after they answered the survey, create a new column with "new_monthly_income". Perform a two sample paired t-test for monthly_income and new_monthly_income. What do you conclude?
8. Write a R program to delete all the entries in the monthly_income column whose value is zero (0) and save the monthly_income as a new file "No_zeros_income"

We want to address the following by using the csv file **"bank_statement.csv"**

1. Import the file bank_statement.csv
2. Rename all the columns of the dataset. For example in the first column, from Total.Assets to total_assets
3. Test for the normality of; Total Assets, Total liabilities, Total Deposits, Total income, Total Expenditure, Total interest income and Profits Before Tax (after exceptional items)
4. Find the regression equation for Profit Before Tax and the other variables. Test of multicollinearity and plot the graphs for the residuals. How do you interpret these results?
5. Compute the correlation between the variables in Q.2 above. Comment on the results