Symphony Hopkins

DSCI 502: R Programming

Professor Xiaobo Li

November 9, 2022

Project 4

1. Read the dataset in loan.csv. Call the loaded data, loan. Make sure that you have the directory set to the correct location for the data.
   1. See R script.
2. Which variables (columns) are continuous/numerical variables? Which columns are factors (categorical variables)?
   1. Continuous/Numerical Variables: loan\_amnt, int\_rate, installment, annual\_inc,
   2. Factors/Categorical Variables: id, term, grade, emp\_length, home\_ownership, verification\_status, loan\_status
3. Calculate the minimum, maximum, mean, median, standard deviation and three quartiles (25th, 50th and 75th percentiles) of loan\_amnt.
   1. Graphical user interface, text, application

      Description automatically generated
4. Calculate the minimum, maximum, mean, median, standard deviation and three quartiles (25th, 50th and 75th percentiles) of int\_rate.
   1. Graphical user interface, text, application

      Description automatically generated
5. Calculate the correlation coefficient of the two variables: int\_rate and installment. Do they have a strong relationship?
   1. Graphical user interface, application

      Description automatically generated
   2. The correlation coefficient for the two variables is approximately 0.13, which suggests that they do not have a strong relationship.
6. Calculate the frequency table of term? What's the mode of term variable?
   1. Graphical user interface, text, application

      Description automatically generated
   2. The mode for term is 36 months.
7. Calculate the proportion table of loan\_status? What's the mode of loan\_status variable?
   1. Text

      Description automatically generated
   2. The mode for loan\_status is “Current.”
8. Calculate the cross table of term and loan\_status. Then produce proportions by row and column respectively.
   1. Cross Table for term and loan\_statusText

      Description automatically generated with medium confidence
   2. Proportions by RowText, letter

      Description automatically generated
   3. Proportions by ColumnText, letter

      Description automatically generated
9. The data is stored in the data frame, loan. Please summarize all the variables using one command.
   1. Table

      Description automatically generated