



1. Create a working folder and change working directory to the created folder. Import “Data2.xlsx” manually and using syntax. Change data type: case_id to string, and cluster to numeric
2. Rename the following variables as defined:
 - a. V001 to gender
 - b. v002 to age
 - c. V003 to income
 - d. v004 to edu
 - e. V005 to occup
 - f. v025 to residence
 - g. v225 to mar_st
3. Label the following variables as defined:
 - a. Gender to “Sex of the respondent”
 - b. Age to “Age of the respondent”
 - c. Income to “Monthly income”
 - d. Edu to “Highest education achieved”
 - e. Occup to “Current occupation”
 - f. Residence to “Place of residence”
 - g. Mar_st to “Ever married status”
4. Change label values of the following variables as defined:
 - a. Gender: 1 Female, 2 male
 - b. Occup: 1 Agriculture, 2 Business, 3 Private job, 4 Government job, 5 Day laborer
 - c. Residence: 0 Urban, 2 Rural
 - d. Mar_st: 0 No, 1 Yes
5. Find the summary statistics of age, monthly income, and education. Find the 25% quartile and 75% quartile of the variables.
6. Categorize the following variables by generating new variable:
 - a. Income: 1 “Below 15k”, 2 “15k to below 20k”, 3 “20k to below 25k”, 4 “25k to 30k”, “5 above 30k”
 - b. Edu: Uneducated (0), Below primary (class 1 to class 4), Primary pass (class 5 to class 7), Secondary pass or above (class 8 to maximum)
 - c. Age: ‘30 years or below’, and ‘30 years above’
7. Find the frequency table: Residence, Gender, Occupation, generated Income category, and generated Education category.
8. Compute the cross table, row percentage, and do association test:
 - a. Education category and Gender;
 - b. Income category and Residence;
 - c. Income category and Education category.
9. Visualize the followings:
 - a. Bar plot of occupation;
 - b. Pie chart of gender, and gender by residence;
 - c. Histogram of income
10. Is there any association between income with variables age, education, residence, and occupation? Find the regression model when the dependent variable is income and independent variables are age, education, residence, and occupation. Interpret the model.

Please make sure that you comment your name and affiliation in the “syntax.do” file so that we can distinguish you. Syntax submission is mandatory.