STATA Assignment:

Q1. Import the Assignment data from Excel to STATA. This is demographic data on women from different states in India and gives their characteristics such as their age, school years, marital status, wages, etc.

Q2. Label the following variables as follows:

1. *state* – Name of the state
2. *age*- Age in years
3. *education*- Years of schooling
4. *married*- Marital status
5. *children*- Number of children under 12 years old
6. *wage*- Hourly wage in rupees

Q3. What is the total number of observations in this data?

Q4. Obtain a detailed summary of the variable *age*. Answer the following using this output:

1. What is the median age in the sample?
2. Calculate the coefficient of variation of age in the sample.
3. Is *age* variable normally distributed? [Hint: A variable is said to be normally distributed if its skewness is 0 and kurtosis is 3.]

Q4. Create a new value label called **marital** which labels values as follows:

0= unmarried, 1= married

Q5. Attach the above-created value label to the values of the variable *married*.

Q6. Out of the total number of women, how many are married? [Hint: use **tab** command.]

Q7. Out of the total number of married women, what is the percentage of women who don’t have any children?

Q8. The average age of married women is more than the average age of unmarried women. Is this statement TRUE or FALSE?

Q9. Generate a new string variable called *state\_name*, and it should take values in the following way:

|  |  |
| --- | --- |
| county | state\_name |
| 0 | Rajasthan |
| 1 | Maharashtra |
| 2 | Gujarat |
| 3 | Andhra Pradesh |
| 4 | Assam |
| 5 | Bihar |
| 6 | Punjab |
| 7 | Tamil Nadu |
| 8 | Haryana |

 Q10. What are the average schooling years in the sample?

1. Which states have their mean schooling years higher than the sample mean schooling years?
2. Generate a dummy variable which takes the value 1 if states satisfy the condition as in part a) above, and 0 otherwise. Name the variable *higheducation.* Label the variable as “State has higher schooling years”.

Q11. How does the average wage of the high education states (as captured by dummy *higheducation*) compare with that of states that don’t have high education?

Q12. Regress wages on education. Answer the following questions:

1. Why is the total number of observations in this regression 1343 instead of 2000?
2. Using the ANOVA table, calculate the coefficient of determination.
3. What is the coefficient of correlation between wages and education?
4. If education increases by 10 years, then by how much do wages increase per hour?
5. In the regression, the t-statistic for each coefficient is given under the null hypothesis Ho: Bk =0. In the current regression of wages on education, there are 2 parameters, B0 (constant) and B1 (education). What is the t-statistic for coefficient if education under the null hypothesis Ho: B1=2.

Q13. Save all your output in the log file named Assignment.smcl.

Q14. How would you repeat the commands used to answer Q1-Q13, without manually typing them again in the command box?