**Q2. Does the sex of an individual affect educational attainment? Is there any evidence that the educational attainment of males is different from that of females?**

**Answer: (a)** In this question, we wish to find out whether the educational attainment of an individual is affected by their sex, hence, we wish to find out whether there is any relationship between the sex of an individual and the number of years of schooling. We will be using the following variables to model the relationship.

**Variable Definitions**

|  |  |
| --- | --- |
| **Variable Name** | **Variable Description** |
| S | Number of years of Schooling |
| ASVABC | measure of cognitive ability of a person |
| SM | years of schooling of the respondent’s mother |
| SF | years of schooling of the respondent’s father |
| MALE | dummy variable which is 1 for male respondents and 0 for female respondents |

This question has a variable which is qualitative in nature and hence does not have a continuous numerical value. The variable ‘MALE’ is therefore called a dummy variable because it represents categorical data i.e. sex of the individual. Since the sex of an individual can take on 2 values, we have taken one dummy variable to avoid the dummy variable trap.

So, we can represent our population equation as follows:

|  |  |
| --- | --- |
|  |  |

where β0, β1, β2, β3 and β4 are the coefficients, values of which we will try to estimate though the regression analysis. From the above equation, based on the value of MALE variable, we can get the equations 2 and 3.

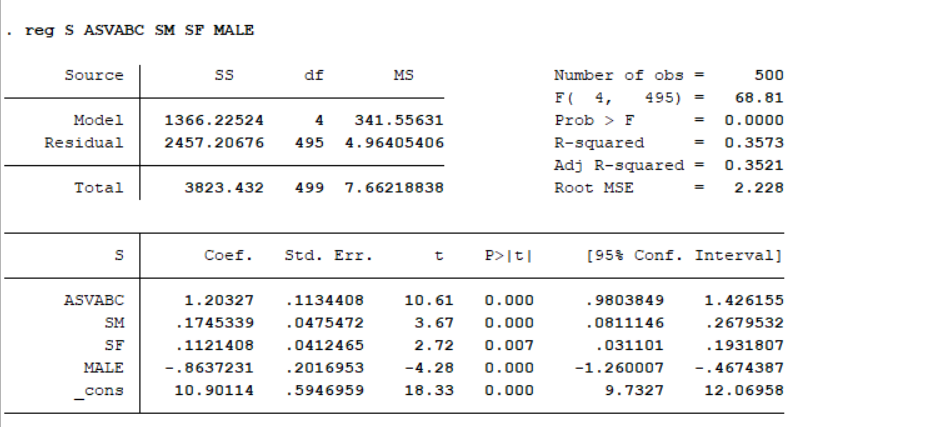
|  |  |
| --- | --- |
|  |  |
| Here, equation 2 is for males and equation 3 is for females. The average value of the error term would be zero so we can eliminate that. On subtracting 3 from 2, we get |  |

|  |  |
| --- | --- |
|  |  |

Here, β4 represents the additional number of years a person might get to attend school for on the basis of their gender. Now, we can write the regression equation as:

|  |  |
| --- | --- |
|  |  |

If the regression coefficients that we estimate are statistically significant then the discrepancy in the number of schooling years is also statistically significant. On running the regression analysis on STATA, we get this output.



Based on the above result, we can write the equation as follows.

|  |  |
| --- | --- |
|  |  |

This means that, holding SM, SF and the cognitive ability of a person constant, Males would attain 0.86 years less of education as opposed to females. However, we need to ascertain whether the coefficient of the MALE dummy variable is statistically significant or not. We can do this by performing a *t*-test on the coefficient of the dummy variable. Our null hypothesis will be that the coefficient of the dummy variable is 0 meaning there is no effect of dummy variable on the years of schooling. Our alternative hypothesis will be that the dummy variable coefficient is not equal to 0 and it does impact the

|  |  |
| --- | --- |
|  |  |

We can calculate the t-stat as follows:

|  |  |
| --- | --- |
| The t-statistic value has been calculated as 2.28 and it can be seen that the STATA output is also the same for this null hypothesis. The degrees of freedom for this sample are 495 and the t-crit value at the 1 percent significance level is 2.586. Since, |t-stat| > |t-crit|, we will reject the null hypothesis at this level. Even at the 0.1 percent, the t-crit value is 3.31 which will lead to a rejection even at the 0.1 percent value with an even smaller risk of Type I error. Thus, we can say that the regression coefficient is statistically significant. The R-squared value is low but not very low to cause a lot concern.  The p-value for this regression is less than 0.001 which means that the probability of obtaining the corresponding t-statistic as a matter of chance is lower than 0.1% so we can reject the null hypothesis at that level.  Before I state my conclusion, I want to test the data for multicollinearity as well to make sure that the explanatory variables aren’t highly correlated. |  |

**Using your EAWE data set, define a slope dummy variable MALEASVC as the product of MALE and ASVABC: MALEASVC = MALE\*ASVABC. Regress S on ASVABC, SM, SF, ETHBLACK, ETHHISP, MALE, and MALEASVC, interpret the equation and perform appropriate statistical tests to comment on the use of the interaction variable in this regression, and whether that improves model specification.**

