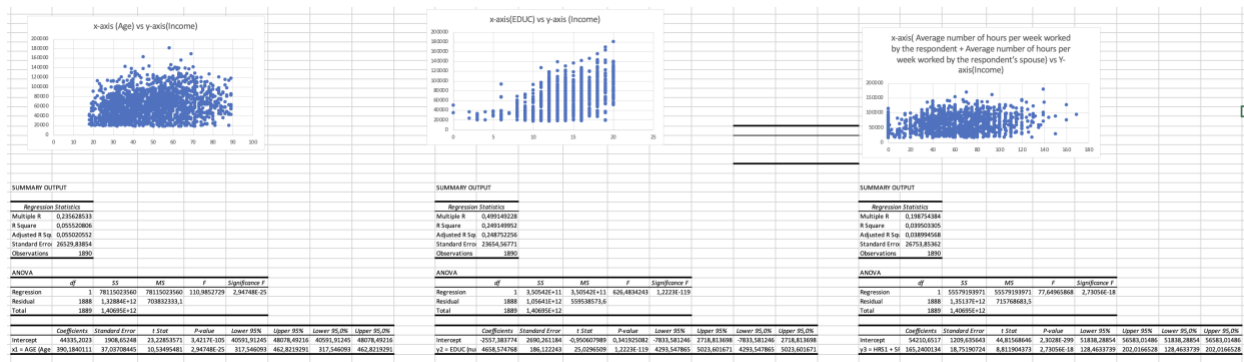


4. The 3 scatter plots of the 3 different regression models are as shown below



R square coefficient says the best fit of the model. Out of the above 3 models, **X (EDUC) vs Y (Income) model can explain the variation better (24.9 %)** compared to the other 2 models. Hence, using X(educ) as a dependent variable results in a better model.

Regression equation is $y = b_0 + b_1(x)$

Variable coefficient interpretation: If x is changed by 1 unit, y would change by b1 unit

1st scatter plot & model: Age vs Income

$Y(\text{Income}) = b_0 + b_1(x = \text{Age}) = 44335.2 + (390.18) \text{ Age}$

Variable coefficient interpretation: If age changes by 1 year, income will change by 390.18 \$

2nd scatter plot & model: Educ (number of years of school completed by respondent) vs Income

$Y(\text{Income}) = b_0 + b_1(x = \text{Educ}) = -2557.8 + (4658.5) \text{ Educ}$

Variable coefficient interpretation: If educ (number of years of school completed by respondent) changes by 1 year, income will change by 390.18 \$

3rd scatter plot & model: HRS1 + SPHRS1 (Avg hrs worked by respondent and spouse together) vs Income

$Y(\text{Income}) = b_0 + b_1(x = \text{Educ}) = 54210.6 + (165.2) \text{ Avg hrs worked by respondent and spouse together}$

Variable coefficient interpretation: If HRS1 + SPHRS1 (Avg hrs worked by respondent and spouse together) changes by 1 year, income will change by 165.2 \$

5. Mean value of **Educ ((number of years of school completed by respondent)** column is 14,15years. Predicted value is 63361,45 \$ \pm 46404,05916 for the annual income of a household with the mean value of the Educ independent variable of 14,15 years.

Predicting the average value for the annual income of a series of households is 63361,45 \pm 1066,832