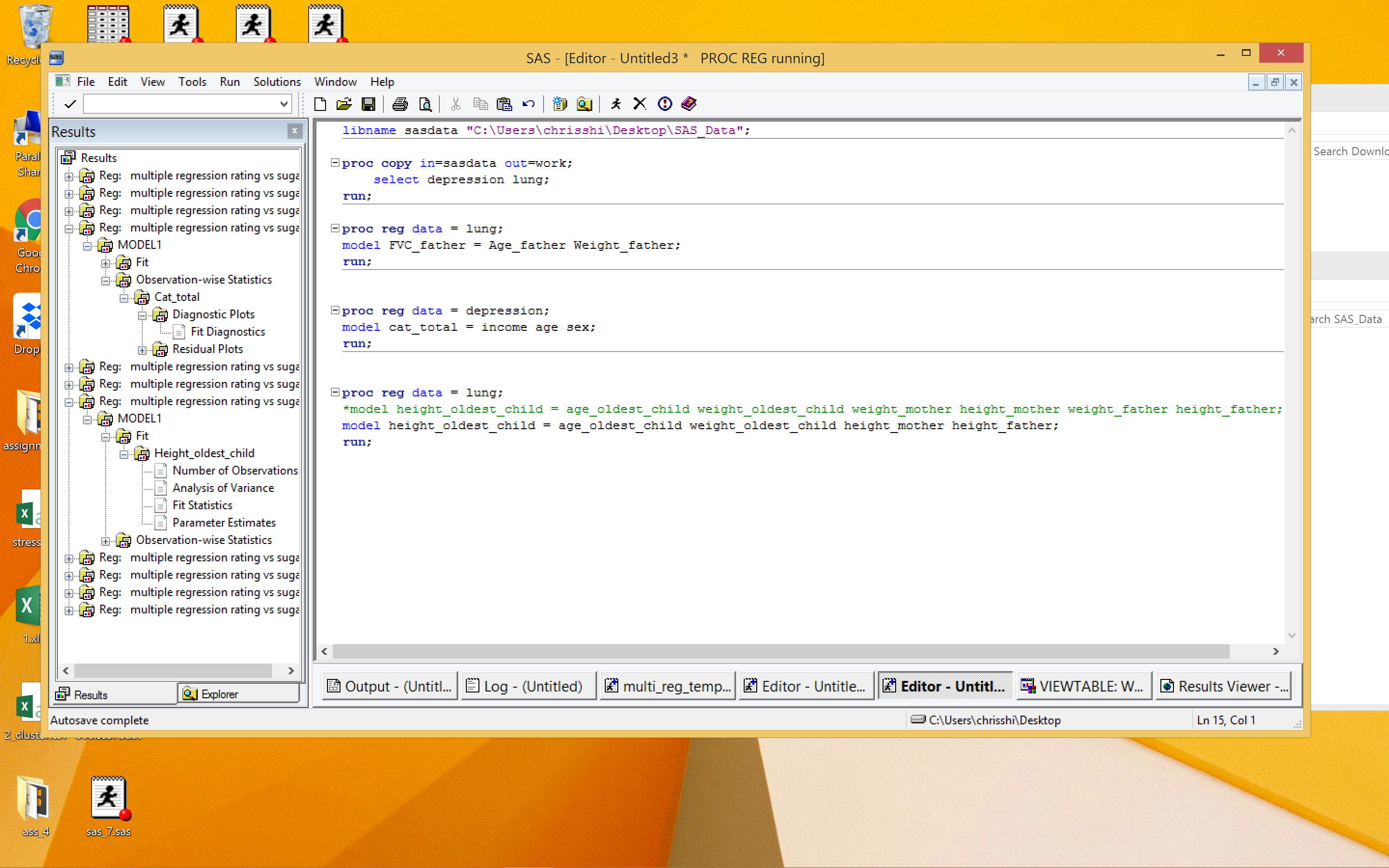
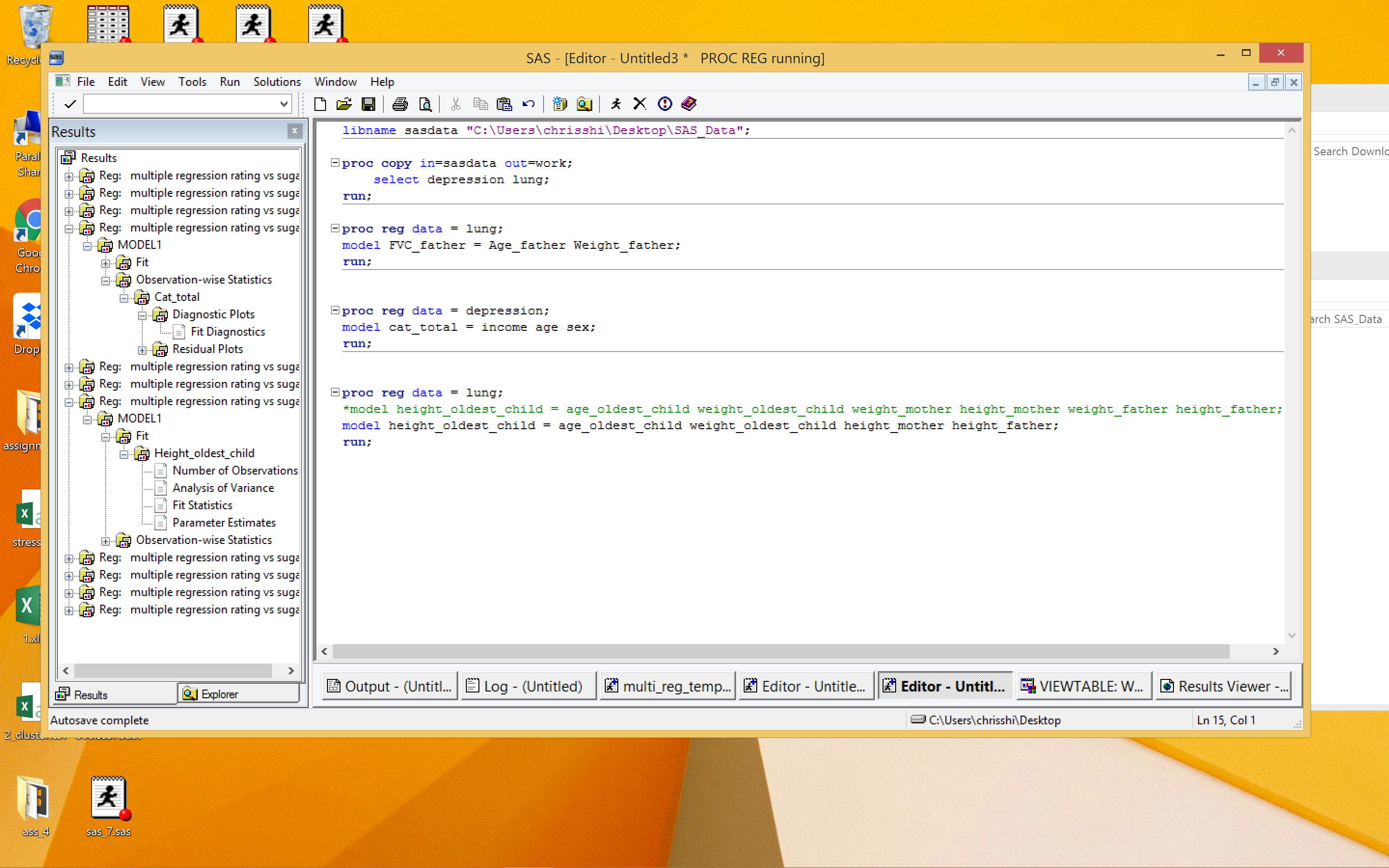
1. Using the “Family Lung” data set fit the regression plane for the father using FVC of father as the dependent variable and age and height of the father as the independent variables (see below)

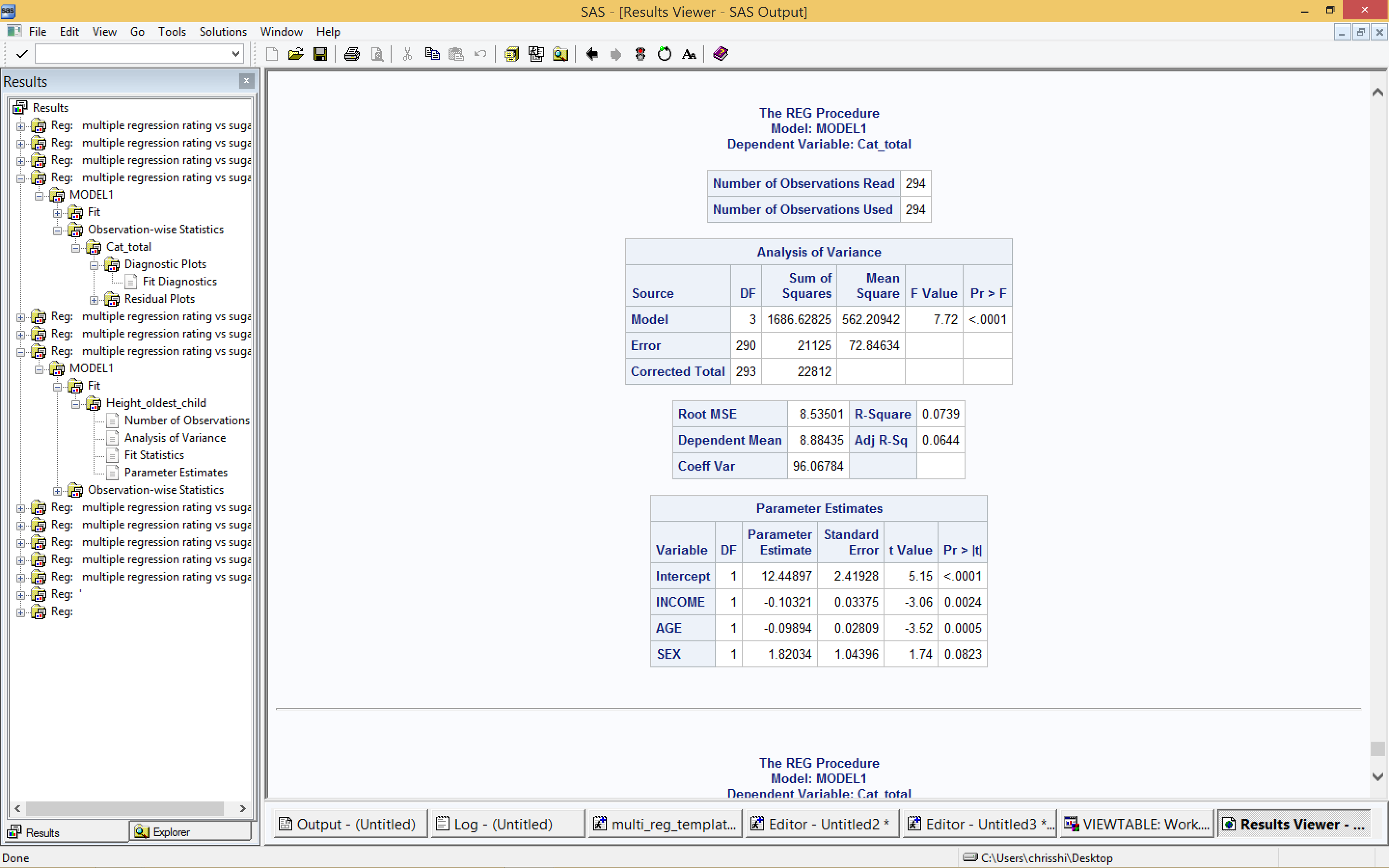
Code:

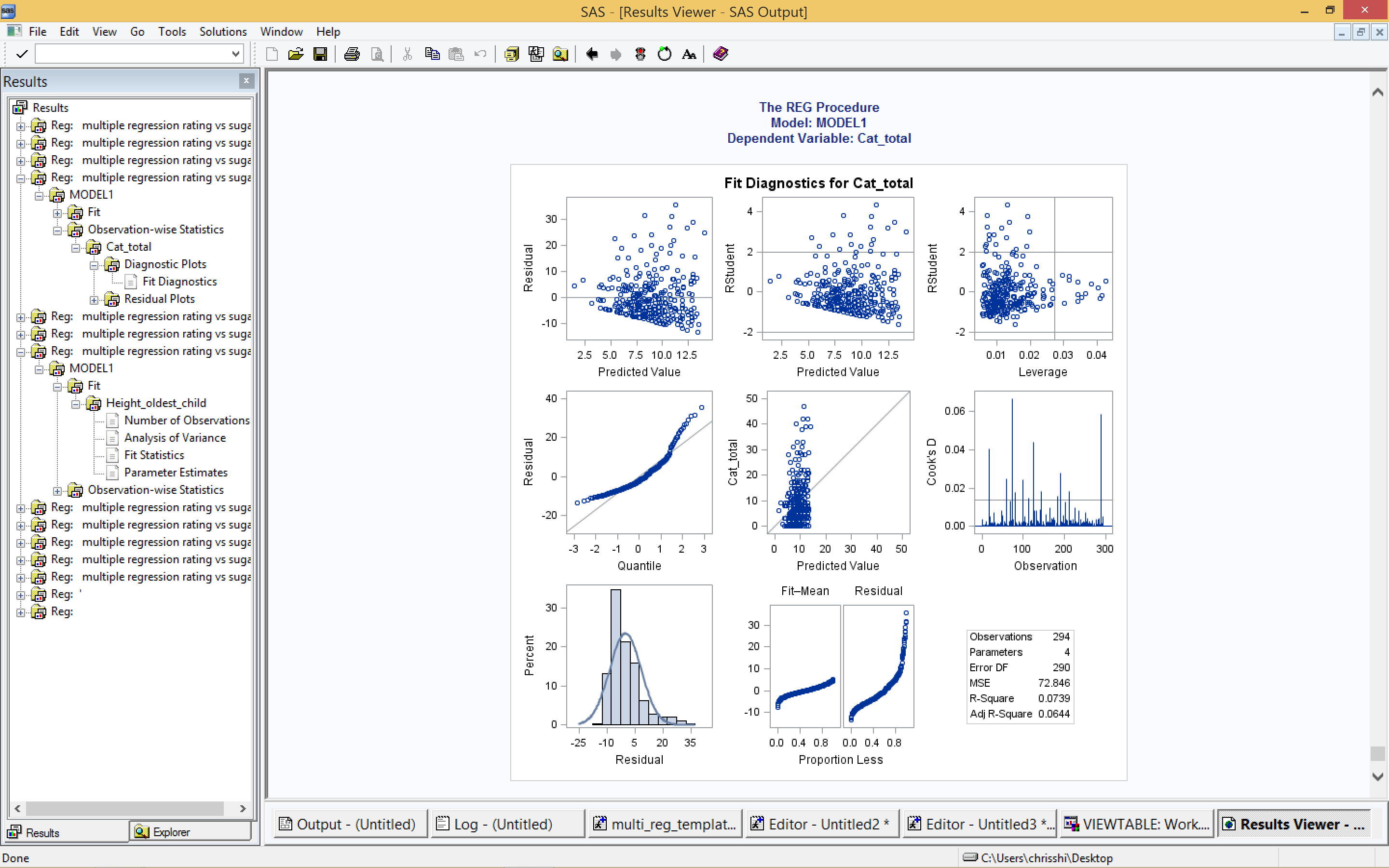


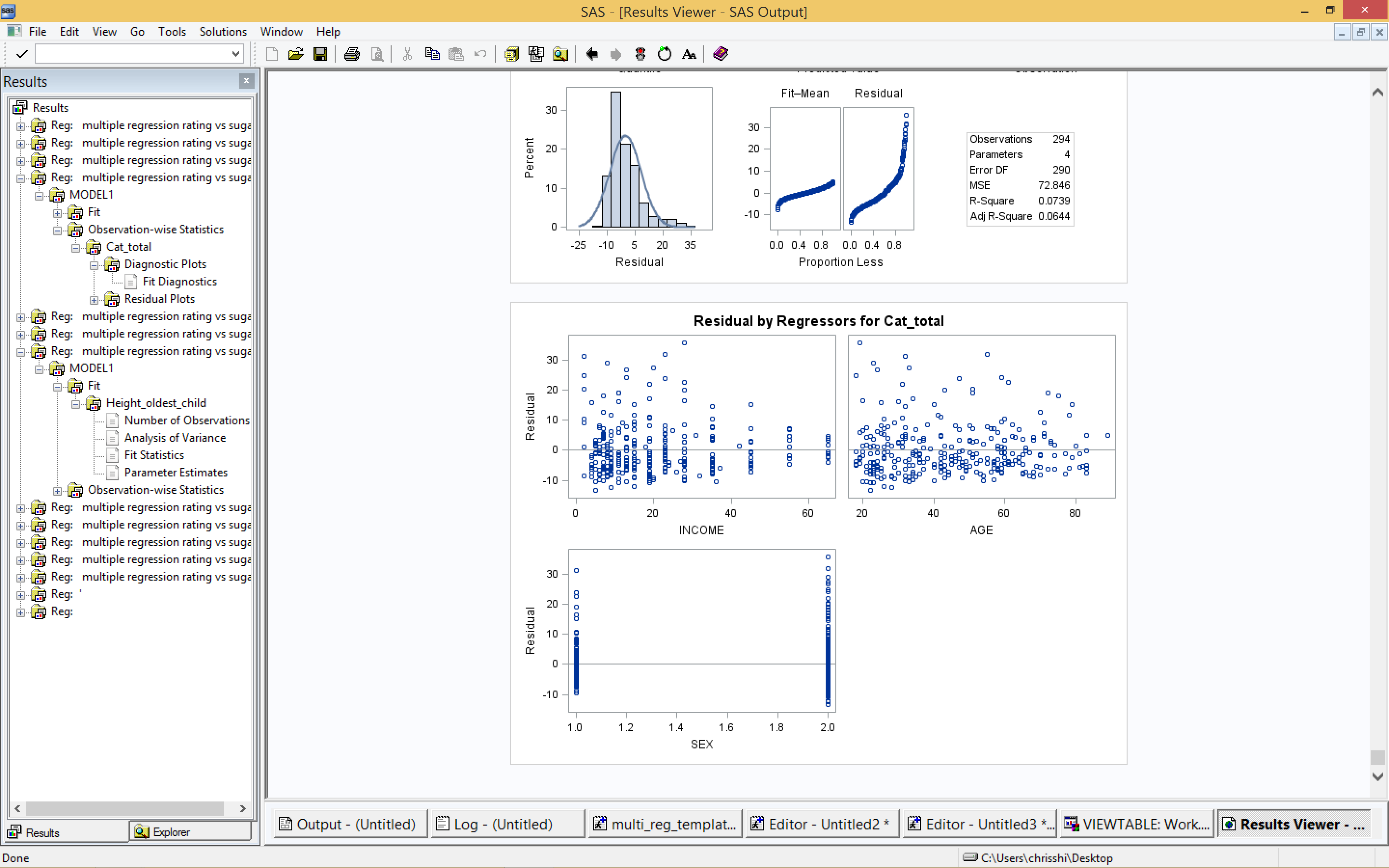
1. For the “Depression” dataset predict the reported level of depression as given by CESD, using income, sex, age as independent variables. Analyze the residuals and decide whether or not it is reasonable to assume that they follow a normal distribution.

Code:  


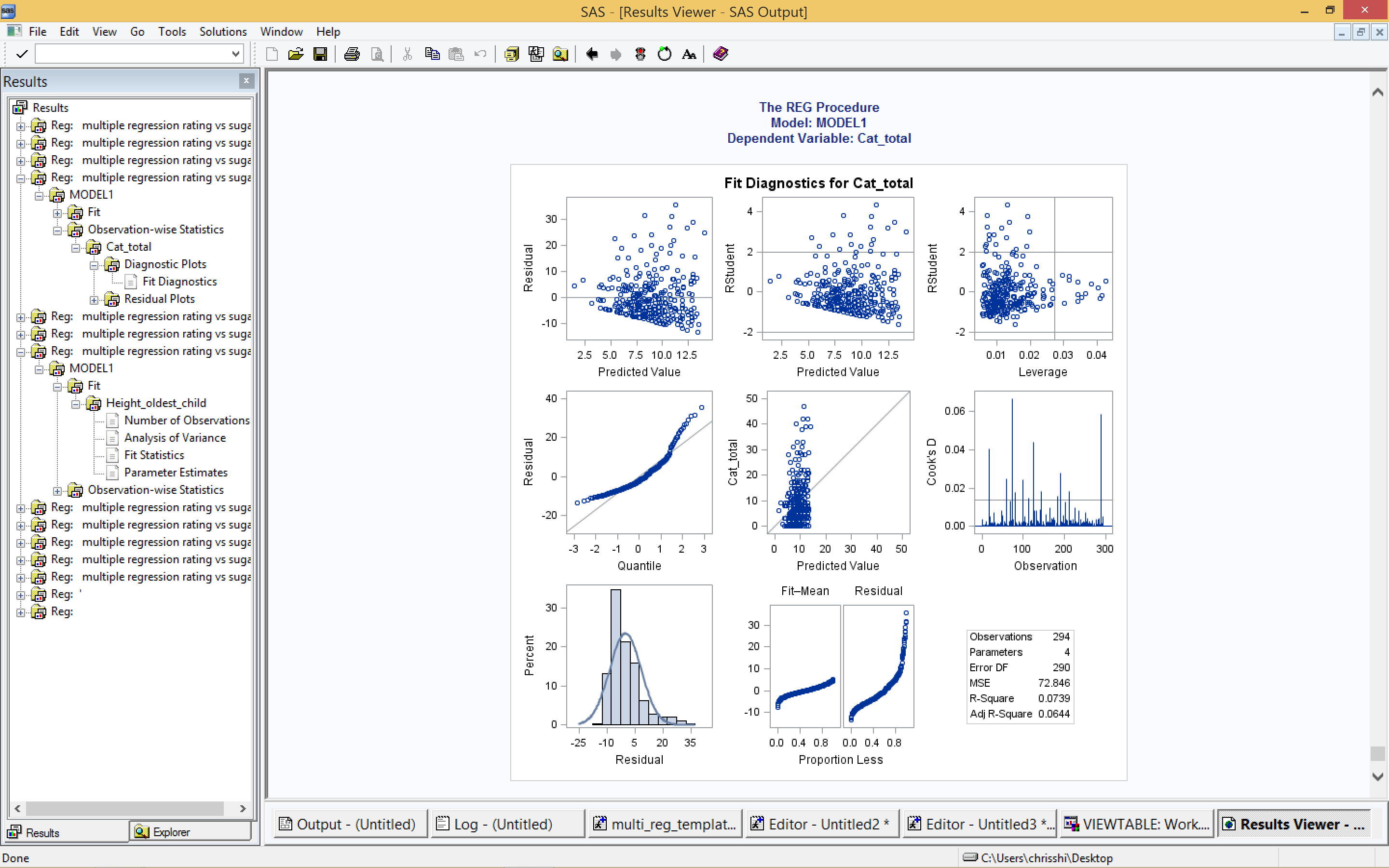
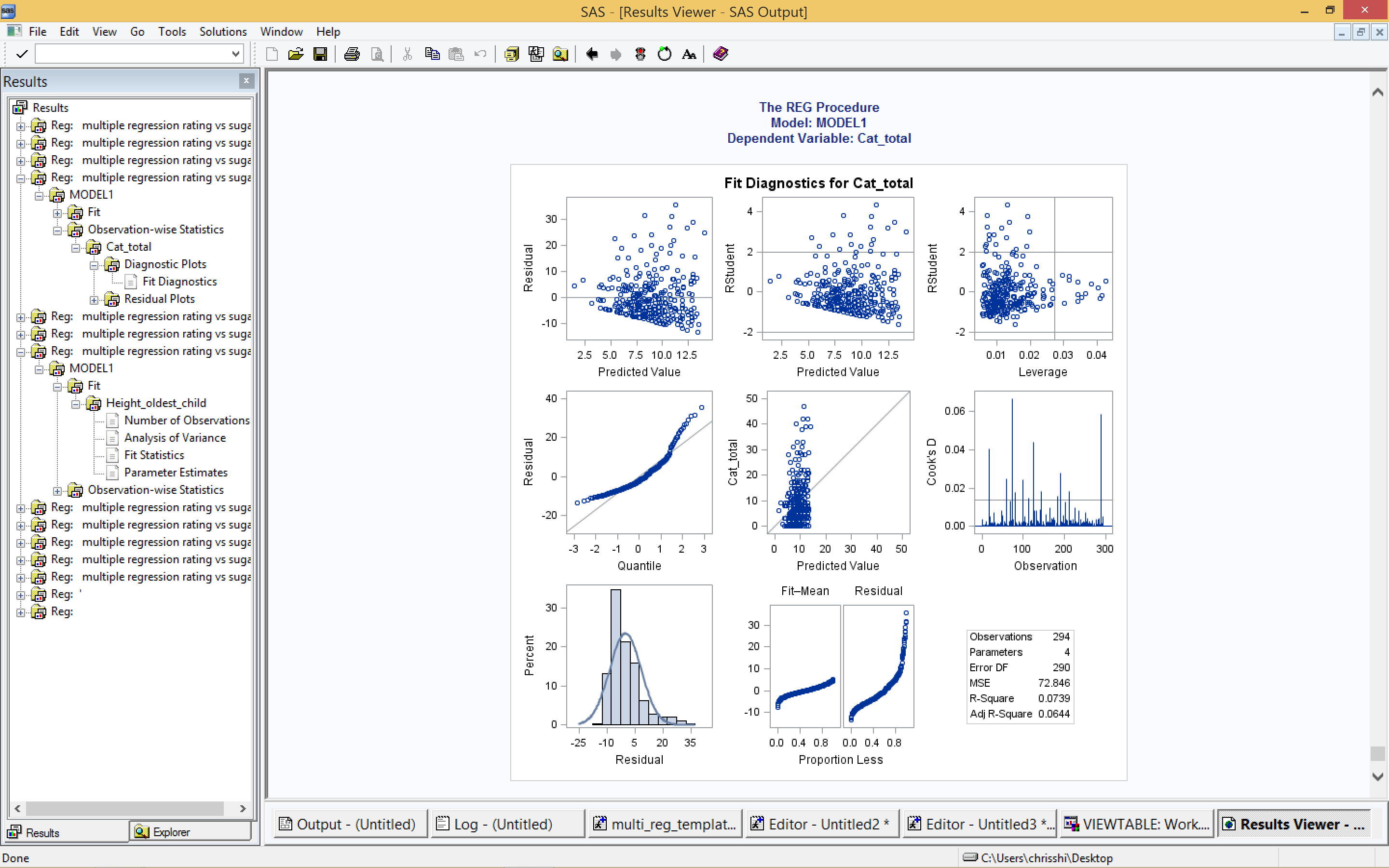
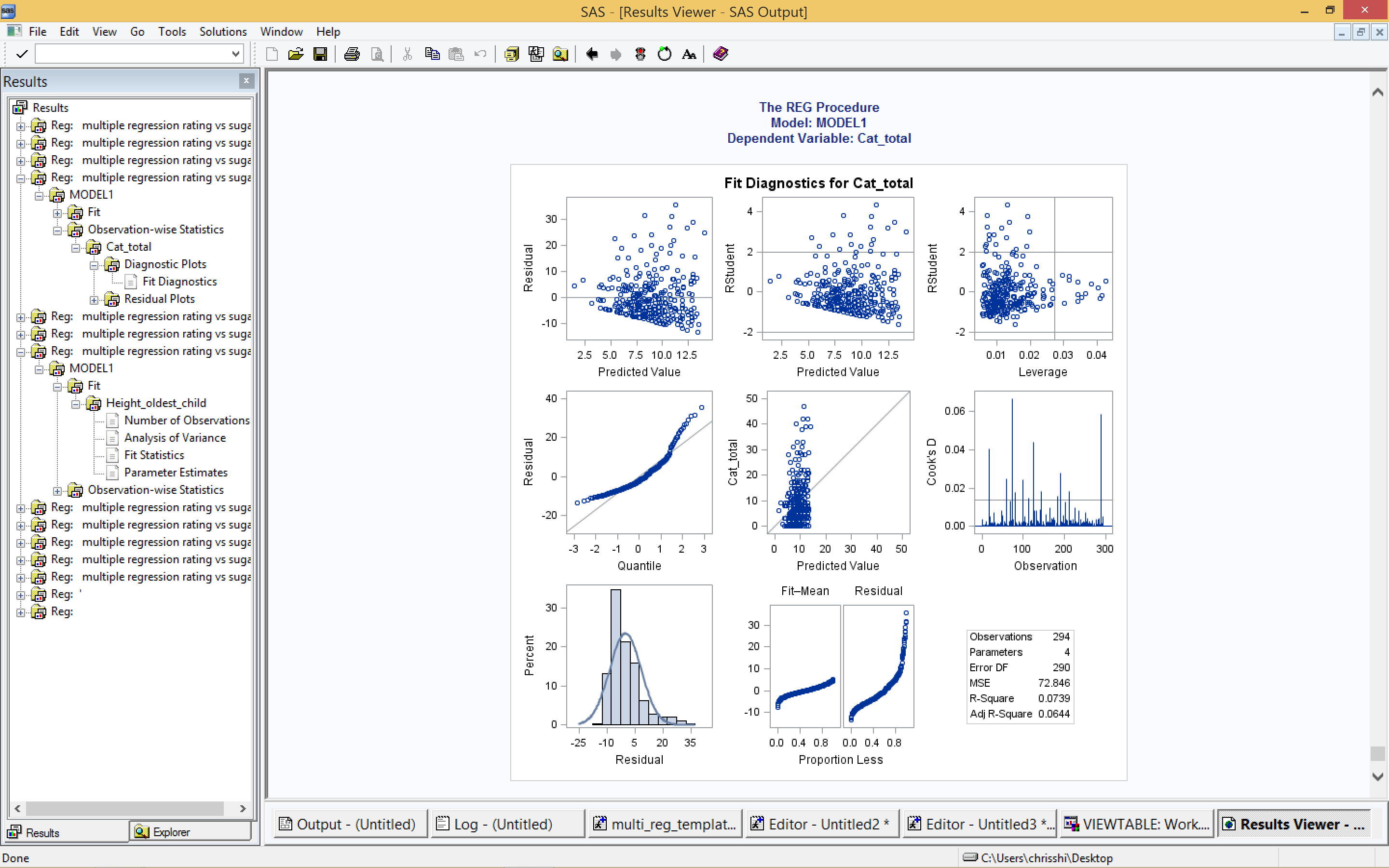
Result:







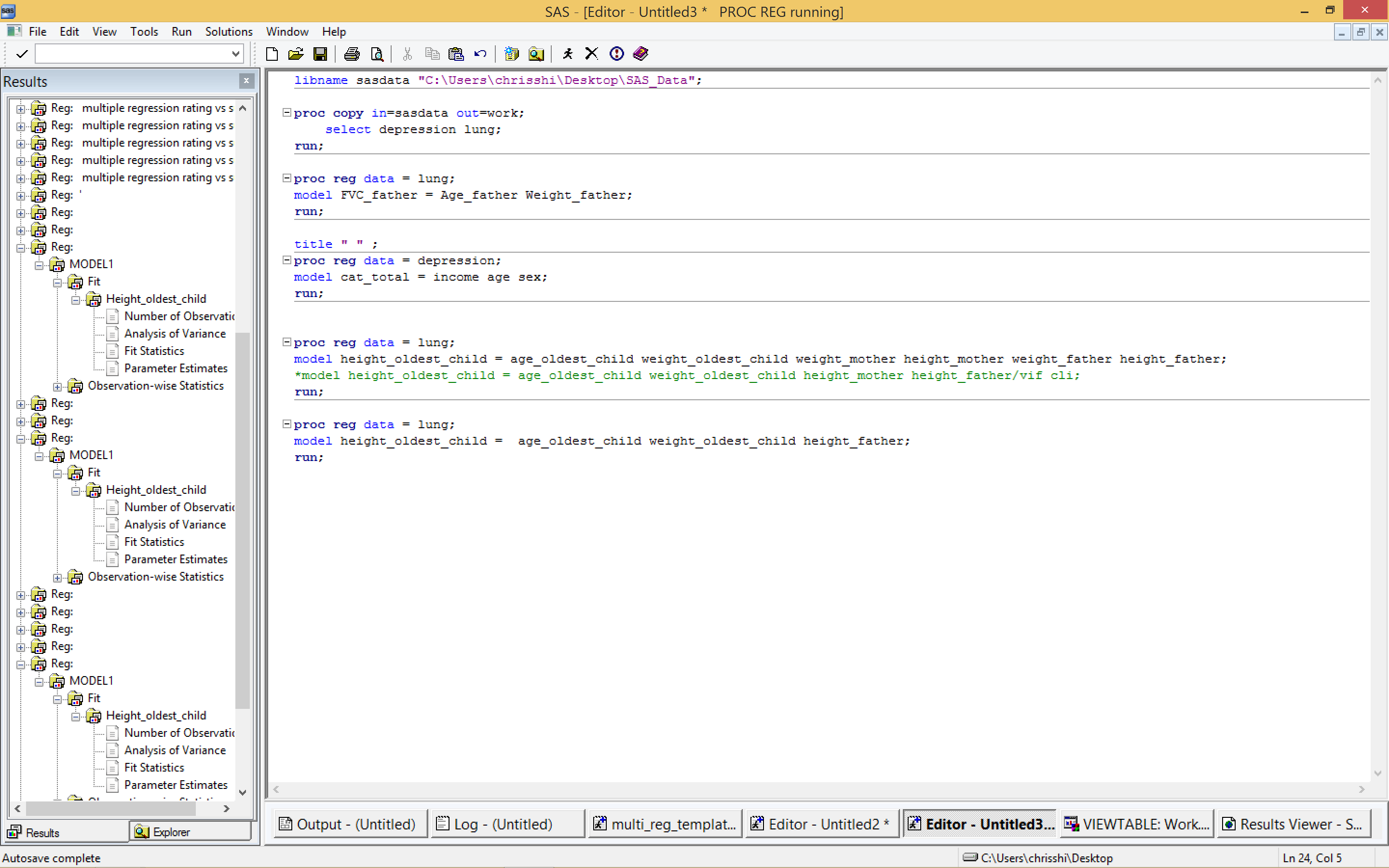
Analysis:



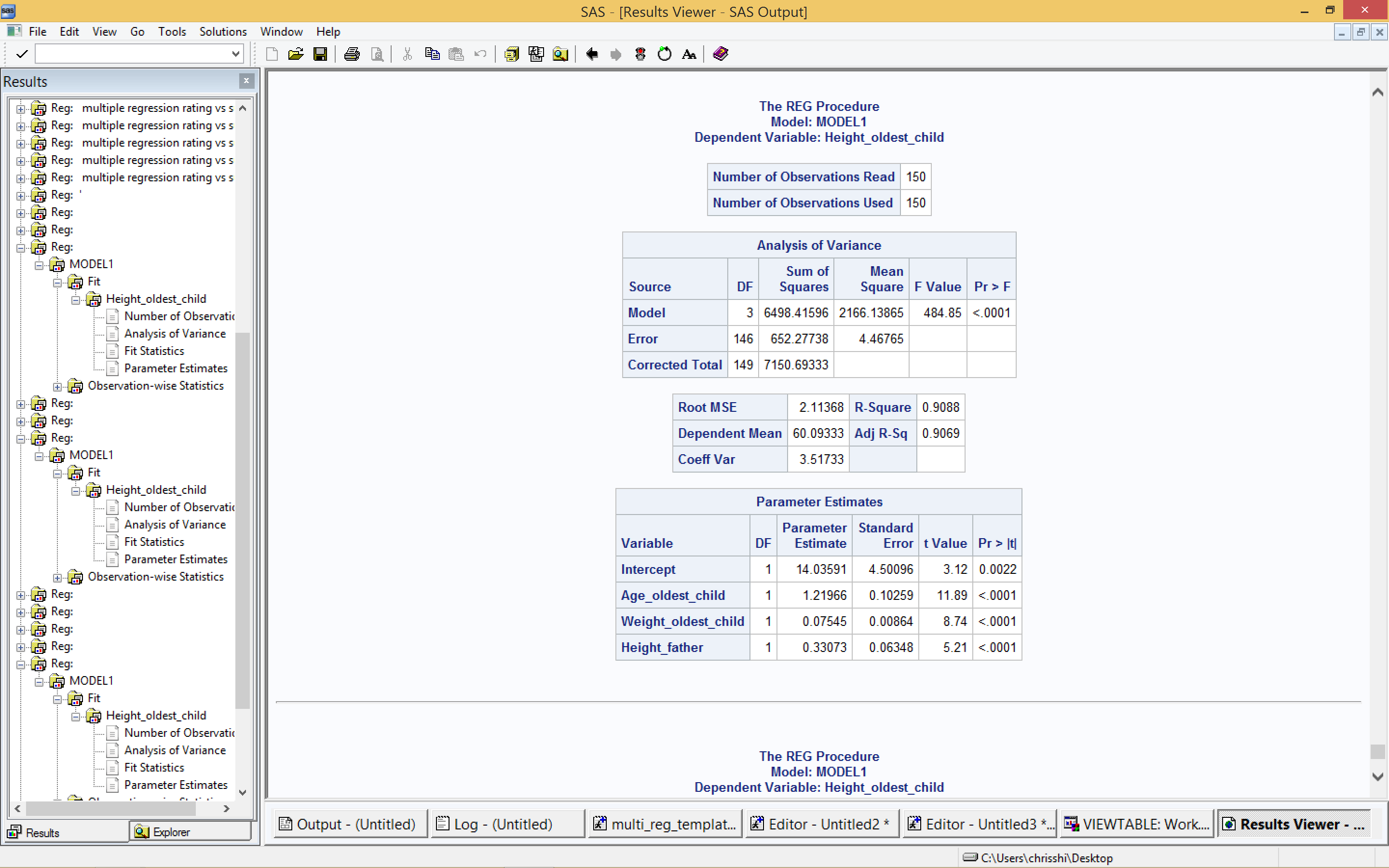
These three graphs show that residual does not follow normal distribution.

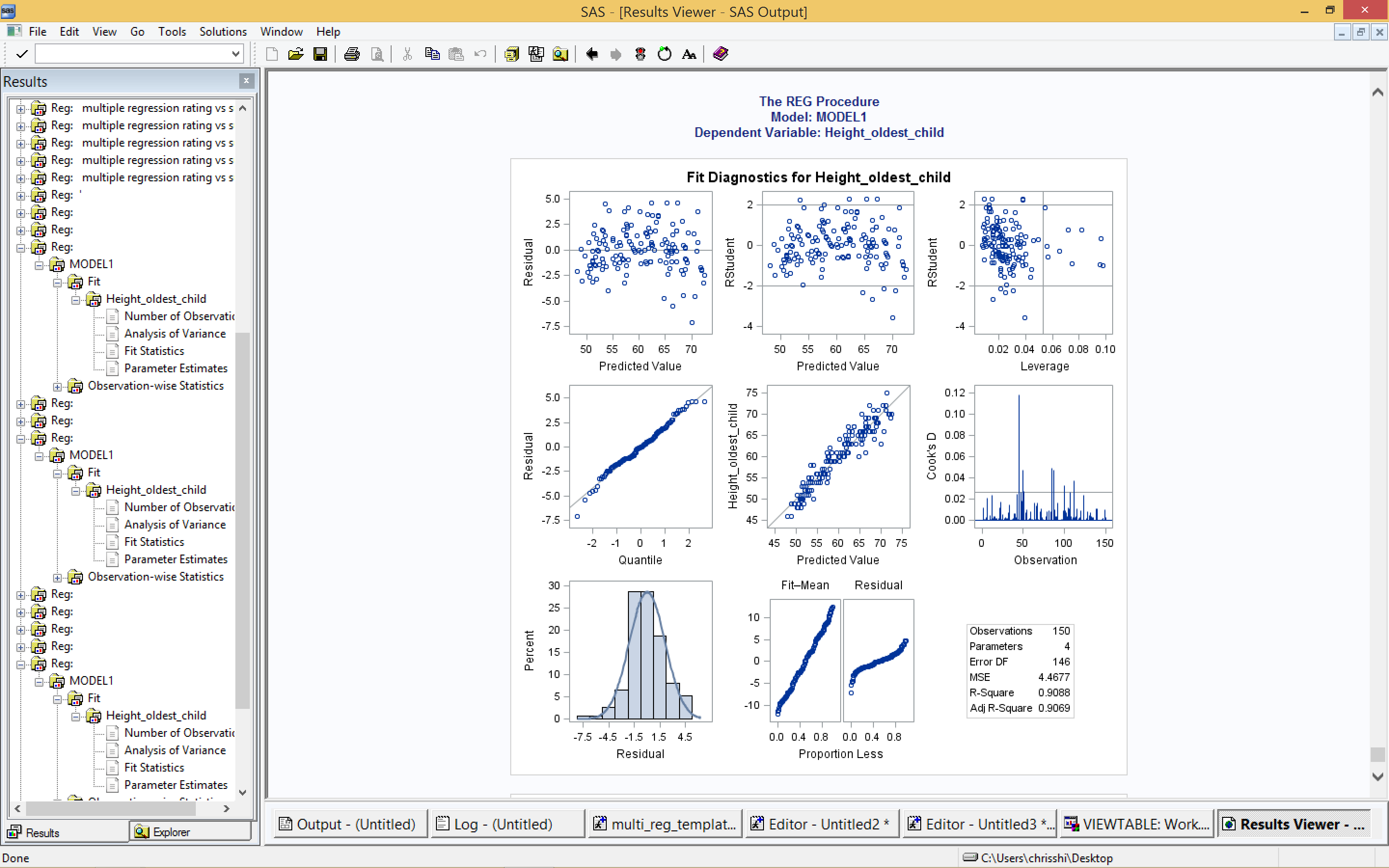
1. Using the “Family Lung” data set fit a regression model to predict height of the oldest child by choosing from among the variables: AGE of oldest child, WEIGHT of oldest child, HEIGHT of the mother, WEIGHT of the mother, HEIGHT of the father, WEIGHT of the father (see below).

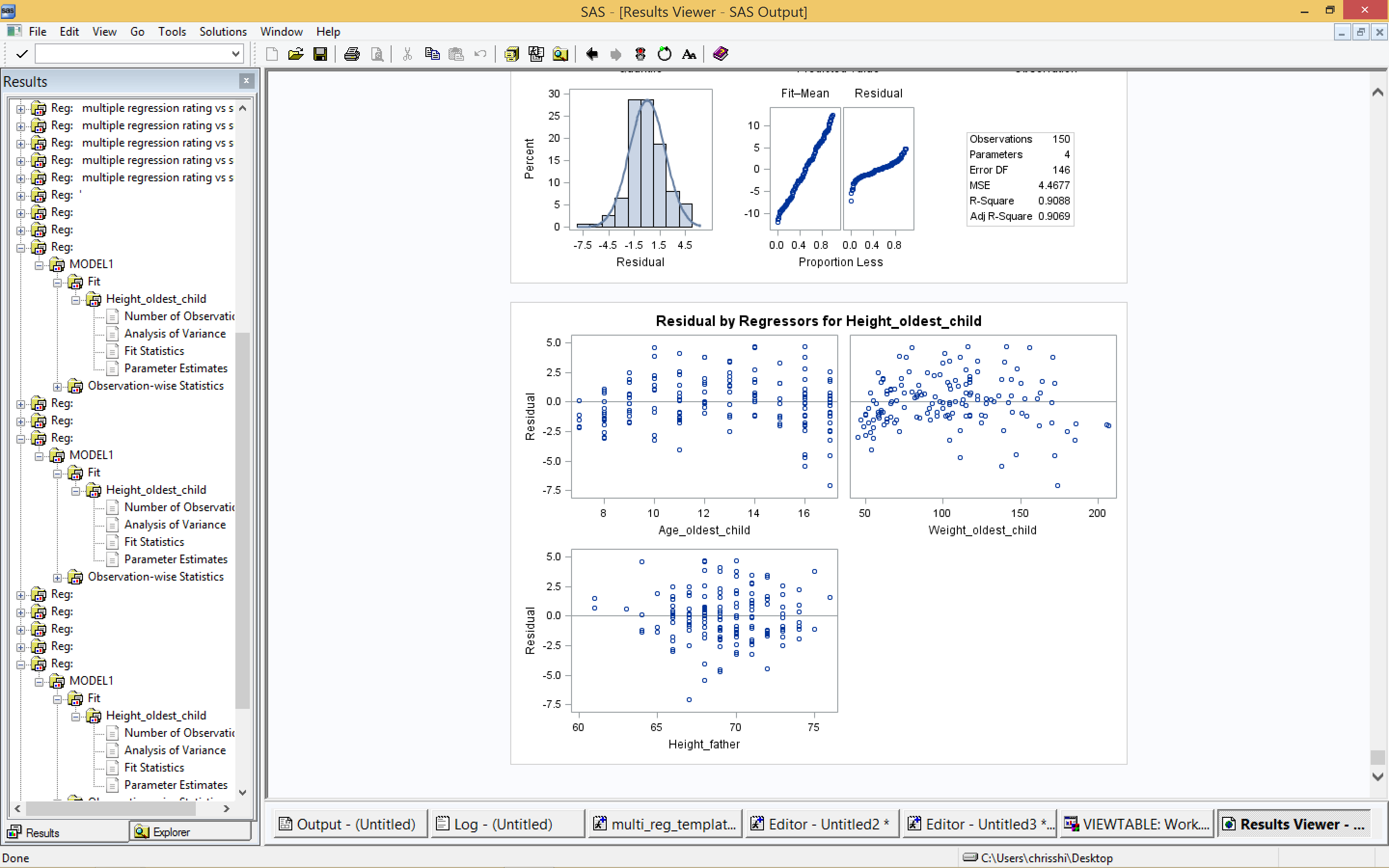
Code:



Result:







Analysis:

I choose age\_oldest\_child, weight\_oldest\_child and height\_father as independent variables. This model is height\_oldest\_child = 14.03591 + 1.21966 \* age\_oldest\_child + 0.07545 \* weight\_oldest\_child + 0.33073 \* height\_father. As shown in the charts above, the “Pr>F” is <.0001 and all of “Pr>|t|” of variables are less than 0.05, which means this is a good model.