In this fifth computer project, you will run a multivariate OLS regression models in SPSS for Windows. Choose only the data set you are working with for your paper. Answer all questions and hand in your output.

TO ACCESS THE DATA FROM SPSS

1) Click the "Start" button visible in the bottom lefthand corner of your computer screen.
2) Click on Computer
3) Double Click on drive Academic Data(I):, then on Public Shares, then on Class Data, then on GSSW.
4) Double click on PSID1992to2009_1.sav, Adolescent_1.sav, or CDS1997to2007_1.sav (depending on which data you are working with). You can also use the alternative versions of these data sets, such as CDS1997to2007_2.sav or PSID929903_7.sav).

MULTIPLE OLS REGRESSION ANALYSIS:

To run a multiple regression analysis, go to Analyze, then to Regression, then to Linear. Right click on the variables to sort them alphabetically and to display the variable name.

In the PSID1992to2009 data: Choose income09 as your dependent variable (as you did in assignment 4). For your independent variables, choose Education09, femalehead09, and K09 (# of kids in 2009). Click OK.

In the Adolescent data: Choose Endinc (avg. income as an adult) as your dependent variable (as you did in assignment 4). For your independent variables choose NevMarr (has the head of household never married), and UnempAD (the unemployment rate for the state of residence during adulthood). Click OK.

For the Child Development data: Choose Sumprob (the total number of problems for the child, as identified by the primary caregiver for the child) as your dependent variable and Hdavg97 (Average hours of work per week for the head of household), and pcgdepre (does the primary caregiver feel depressed) as your independent variables. Click OK.

Notice that Poverty in the femalehead09, NevMarr, and PCGDEPRE are not interval or ratio scale variables but are nominal scale variables. They go into the regression analysis as dummy variables and measures the difference in the mean value of the dependent variable between groups. A dummy variable must have a value of either 0 or 1. Thus, in an OLS regression model, only interval/ratio scale variables and dummy variables are allowed as independent variables. Only interval/ratio scale variables are allowed as dependent variables.
In the PSID1992to2009 data, the parameter estimate for femalehead09 will indicate the difference in income for families who are headed by females versus those families who are headed by males. A positive b coefficient for FEMALEHEAD09 indicates that female headed families have higher income levels than male headed families. A negative b coefficient for FEMALEHEAD09 indicates that female headed families have lower levels of income than male headed families. For example, if the value of the b coefficient for FEMALEHEAD09 is 1000, this means that families in female headed families have incomes that are $1000 higher than male headed families. If the value of the b coefficient for FEMALEHEAD09 is -1000, this would mean that female headed families have income levels that are 1000 less than male headed families. The coefficient estimate for this dummy variable is determined while controlling for the effects of all the other independent variables in the model. This interpretation of a dummy variable is thus different than for an interval scale independent variable.

The interpretation for the b coefficient for (years of education of the head of household), an interval scale variable, is that for each year increase in education of the head of household, income changes by whatever the b coefficient is. If the b coefficient is 5,145.388, then for each year of education of the head of household, income will increase by $5145.388. Notice that the coefficient estimate for EDUCATION09 has changed in the multiple regression model from the bivariate model you ran in assignment 4. This is because you are now controlling for other variables within this second model.

In the Adolescent data, the coefficient estimate for the variable NevMarr measures the difference in adult income between adolescents who grow up in households that have a head who has never been married and those who grow up in households have a head who had been married at some point. A positive coefficient indicates that the level of income as an adult is higher for those growing up in households where the head has never been married. A negative coefficient indicates that level of income as an adult is lower for those who grow up in households with a head who has never been married relative to those who grow up in households where the head has been married. Hence, a coefficient estimate of 1,000 for NevMarr indicates that the mean level of adult income for those who grew up in families where the head had never been married is $1,000 more than it is for those who grew up in households where the head had been married at some point. A b coefficient estimate of -5,000 indicates that the mean level of income for those who grew up in households where the head had never been married is $5,000 less than it is for those who grew up in households where the head had been married at some point. The coefficient estimates are determined while controlling for all other variables in the model. This interpretation of a dummy variable is thus different than for an interval scale independent variable.

The interpretation for UnempAD (the unemployment rate as an adult), an interval/ratio scale variable, is that for each 1 unit increase in the rate of unemployment, income (the dependent variables) change by whatever the b coefficient is. If the b coefficient is -1000, then for increase in unemployment by 1 unit, income will decrease by $1000.
In the Child Development data, the coefficient estimate for the variable PCGDEPRE measures the difference in the number of emotional type problems perceived by primary caregivers who are depressed and those who are not. If PCGDEPRE is positive, this indicates that primary caregivers who are depressed have children with more problems than those primary caregivers who are not depressed. If the coefficient estimate for PCGDEPRE is negative, this indicates that the number of emotional type of problems for children who come from families with depressed primary caregivers is less than the number of problems for children where their primary caregiver is not depressed. If the coefficient estimate for PCGDEPRE is 1.0, this indicates that children with primary caregivers who are depressed have one more problem than children with primary caregivers who are not depressed. If the coefficient estimate for PCGDEPRE is -5, this indicates that children from families with depressed primary caregivers have children with 5 fewer emotional problems than children from families with a primary caregiver who is not depressed.

The interpretation for Hdavg97 (the weekly hours of work for the head of household), an interval/ratio scale variable, is that for each unit increase in the number of hours of work, the number of problems for the child (the dependent variables) changes by whatever the b coefficient is. If the b coefficient is 3, then for an increase in hours of work by 1 unit, the number of problems for the child will increase by 3. If the b coefficient is -2, then for an increase in hours of work by 1 unit, the number of problems for the child will decrease by 2.

Multiple Regression Questions

Questions
1. Is there a significant relationship between your interval/ratio scale independent variable and your dependent variable? At what level is the coefficient estimate significant? What is predicted to happen to your dependent variable when your interval/ratio scale variable increases by 1 unit? Is the model statistically significant?
2. How much of the variation in the dependent variable is explained by the independent variables?
3. Explain the meaning for the b coefficient estimate for the nominal scale or dummy variable.

Note: If your dependent variable in your paper is an interval or ratio scale variable, you will be using OLS regression analysis for your paper. If you would like, go to the appropriate data set and try using your dependent variable with a few of your independent or control variables to run a regression. What do you find?

Either print out the results, or save them to a file (by right clicking when you’re in the output) and send the results with answers typed in.

Multiple Regression in MS Excel
From the Computer Assignment Web page http://www.brynmawr.edu/socialwork/GSSW/Vartanian/Comp131/COMPASS.HTML), go to the Download a small version of the PSID data in MS Excel format and click on this. If you are asked to either save the data set or open it, click on open it and OK. This is a data set of 500 randomly chosen PSID survey respondents. You will determine multiple regression coefficients for variables from this data. At the very top of your screen, go to Data Analysis. Click on Regression and OK. You will first be asked for the Input Y Range, which is the range of values for your dependent variable. Your dependent variable in this case is total family income. Type G1:G501 into this first box. For the input X range, or your independent variables, type in E1:F501. You are using two independent variables in this case -- age of the youngest child in the household and years of education for the head of the household. Click on the box for Labels (if you fail to do this, you won't get any results). Next, click on the circle for Output range, and type in A505 in the box. Click on OK. You will need to go to cell A505 to see your results.

Questions

1. Is there a significant relationship between your interval/ratio scale independent variable and your dependent variable? At what level is the coefficient estimate significant? What is predicted to happen to your dependent variable when your interval/ratio scale variable increases by 1 unit? Is the model statistically significant?

2. How much of the variation in the dependent variable is explained by the independent variables?

Print this MS Excel output or copy the output, put it into a word file, and write up your answers. You can then email the results.