Working with Multiple Regression

The purpose of this assignment is to think about multivariate modeling, use a computer package to estimate the regression model, and interpret the regression results. We will be using the Washington State Population Survey data set and SPSS or Excel. You'll need to refer to the information on the data set to identify the outcome (dependent) and explanatory (predictors or independent) variables. [You can find the data set in SPSS and Excel format on the Evans School server using the link from the course web page].

Assignment requirements:
I. Discuss the model: the outcome and predictors that you will use
II. Discuss expected relationships between the outcome and the independent variables
III. Estimate the regression [Directions on running a regression in SPSS and Excel are at the end of this assignment]
IV. Answer the discussion questions

I. The model: the outcome and predictors
Pick one dependent variable (outcome) from the data set and two or three explanatory variables. Your dependent variable should be ordinal or ratio (not categorical). [You'll need to be careful about this -- most of the variables in this data set are categorical.]

Your explanatory variables should be factors that you believe affect the outcome, but are not themselves outcomes (that is, they themselves are not affected by the dependent variable). The explanatory variables can be ordinal, ratio, or dummy variables (that is, coded as a 1 or 0), but not categorical variables with more than two categories.

For example, my model is:

\[ Q3P5 = \beta_0 + \beta_1(PEARN99) + \beta_2(AGE) + \beta_3(Q8P4) + \beta_4(KING) + \varepsilon \]

I believe that the amount of monthly rent (Q3P5) is affected by the individual's income (PEARN99), the person's age (AGE), the minutes required to get to school or work (Q8P4), or whether the person living King County (KING). (From the variable REGION, I created a binary indicator for King County residents, the comparison being all other regions.)

II. The expected relationships
DQ 1: What is your full model? Why and how do you think the explanatory variables affect the dependent variable? What you expect the coefficients and SEs to look like? Do you think this model will explain much of the difference in your outcome?

III. Estimation of the model
Estimate your model with all your explanatory variables. Then drop one explanatory variable and re-estimate the model. What happens to the estimated regression? Why?
IV. More discussion questions

DQ1: Make sure you answer DQ1 above!

DQ 2: What does each coefficient tell you about the relationship of the explanatory factor to your dependent variable (described results)? Are they asked you expected? If not, can you speculate why not?

DQ3: What happened to the regression results when you dropped one explanatory variable? What does this tell you about the relationship between the explanatory variables?

DQ4: What were $R^2$ and the adjusted to $R^2$ for your model (before dropping your variable)? What do they tell you?

DQ5: What are the 95% confidence intervals around each of your coefficients? Can you reject the null hypothesis that each coefficient (individually) has no effect on the outcome?

DQ6: How likely is it that you would get coefficient values this large if there were no linear relationship to the dependent variable (p-values)?

DQ7: If you were to continue with this project, what else would you add to your model? Why?

DQ8: Write a short paragraph describing your findings to a policy maker.

DO NOT HAND IN ALL YOUR PRINT OUT, ONLY A COPY OF THE REGRESSION RESULTS AND THE ANSWERS TO THE QUESTIONS.
RUNNING A REGRESSION IN SPSS FOR WINDOWS:

To run a regression in SPSS, first get into the program and read in your data. You may want to create new variables to use in your regression, as well.

To get your data, click on:

FILE

    OPEN, then fill in or click on file name

To run the regression, click on:

ANALYZE

    REGRESSION

    LINEAR

Pick your outcome variable from the list at the left and click on the arrow for DEPENDENT VARIABLE.

Pick your explanatory variables from the list and click on the arrow to move them to the list of INDEPENDENT VARIABLES.

If you'd like, SPSS will give you descriptives for each variable in the regression. This is helpful because SPSS uses only cases with no missing data on ANY of the variables in the regression. The descriptive statistics it gives you here are only for the cases actually used in the regression (no missing data). Also, SPSS gives you the simple correlations between each pair of explanatory variables. This will be helpful later when we talk about multicollinearity.

To get descriptive statistics, in the regression screen click on STATISTICS and click in the box for descriptives.

You can also get confidence intervals around each coefficient. In the regression screen click on STATISTICS, then click in the box for confidence intervals.

Now you can run your regression by clicking on OK.
RUNNING A REGRESSION IN EXCEL:
Make sure your data are in a spreadsheet with the variable names across the top and each case in a row. Open your dataset and perform any transformations you might need to. Make sure all your independent variables are in adjacent rows.

To run the regression, click on TOOLS > DATA ANALYSIS. In the dialogue box, select “regression.” Make sure your cursor is in the box “Input Y Range” then select the column of data that contains the dependent variable, including the label. Click in the box “Input X Range” and select the column or columns with your independent variables (including the labels). Click the box next to “labels.” In the box next to “new worksheet ply” type the name of the worksheet where your results will appear. Click OK to run the regression.