Problems: MULTIPLE REGRESSION

1. An inspection of the box plot on page 147 of the Lecture Notes reveals that the more prestigious a job, the more it pays. Show that this relation no longer holds after "rincome" has been adjusted for linear "agewed" effects.
   a. Begin by using the "plot" command to generate a plot between "rincome" and "agewed" like the one in the Lecture Notes.
   b. Now use the "regression" command to get a slope estimate for the regression of "rincome" on "agewed".
   c. Using the slope estimate from part b, use a "compute" statement to create "incadj" (a measure of "rincome" adjusted for "agewed") and generate a plot between "incadj" and "agewed".
   d. Draw a median trace across the plot generated in part c. Is there still a linear relation between income and prestige after linear prestige effects have been removed from the income measure? Is there a nonlinear relation? Explain your answers.
   e. Describe how the plots generated in parts a and c differ. Explain why these differences occurred.

2. What are the differences among a zero-order coefficient of determination, a multiple coefficient of determination, and a partial coefficient of determination? How is each interpreted?

3. You have assembled a simple random sample of 26 employees at an ice cream company. You are interested in what enhances employees' job satisfaction at the firm. Each respondent's job satisfaction (SATISFY) is measured on a scale from 1 (least satisfied) to 100 (most satisfied). Charles Perrow's Neo-Weberian theory of organizations states that people will be satisfied with their jobs if they are paid well. Argyrus and other proponents of the Human Relations School of organizational theory argue that workers must believe that their work is of value. To evaluate Perrow's ideas you obtain information on how much the company paid each employee (in dollars) during the past year (MONEY); to evaluate Argyrus's ideas you measure how often employees' supervisors express appreciation (in compliments per week) to each employee (STROKES). You generate the following correlation matrix (a matrix of correlation coefficients):

   SATISFY   MONEY   STROKES
   SATISFY   1.000  .600   -.400
   MONEY     .600   1.000  .300
   STROKES   -.400  .300   1.000

(Make sure you read this matrix correctly. For example, the correlation between employees' job satisfaction and their annual pay is .600.)

   a. What proportion of the variance in "employees' job satisfaction that is not explained by supervisor appreciation" is accounted for by their annual
salaries?

b. Is this proportion greater than zero at the .01 level of significance?

c. Find a 95\% confidence interval for the first-order partial correlation between SATISFY and MONEY after controlling for STROKES.

4. You have asked each respondent from a sample of 63 Americans for his/her household income (HINCOME), his/her parents' household income (PINCOME), and his/her education (REDUC). You generate the following correlations, means, and standard deviations:

<table>
<thead>
<tr>
<th></th>
<th>HINCOME</th>
<th>PINCOME</th>
<th>REDUC</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>HINCOME</td>
<td>1.00</td>
<td>.64</td>
<td>.50</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>PINCOME</td>
<td>.64</td>
<td>1.00</td>
<td>.60</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>REDUC</td>
<td>.50</td>
<td>.60</td>
<td>1.00</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>

(Make sure you read the correlation matrix correctly. For example, the correlation between respondents' household incomes and their educations is .50. Both income measures are in thousands of dollars, and the education measure is in years of education.)

a. For the OLS regression of HINCOME on PINCOME and REDUC, find the standardized regression equation, the unstandardized regression equation, and state the meanings of all unstandardized regression coefficients in words.

b. What proportion of the variance in HINCOME is explained by PINCOME and REDUC? Is this proportion significantly large at the .05 significance level?

c. Considering only that variance in respondents' household incomes that is not explained by their parents' incomes, what proportion of this variance is explained by respondents' educations?

d. Find a 95\% confidence interval for the first-order partial correlation between HINCOME and REDUC controlling for PINCOME.

e. At the .05 significance level test the null hypothesis that people with higher educations are more likely to earn more than people with lower educations even after taking into account the tendency for people from affluent families to be more likely than other people to get higher educations.

f. Find the 95\% confidence interval for the partial slope between HINCOME and REDUC while controlling for PINCOME.

5. According to social capital theory, a community's quality of life can be improved, and its poverty reduced, if its citizens become actively involved in community improvement projects. There are two necessary conditions for such community improvements to occur. First, citizens must have skills that enable them to mobilize changes in their community. Second, these skills must be
implemented in concrete, community-oriented actions. In brief, citizens must be both able and motivated to implement change. You wish to estimate the extent to which the ability and motivation of community leaders account for the amount of wealth in their communities.

You obtain data from Vern Ryan (a professor at Iowa State University) on a random sample of 123 Iowa communities. Your analysis makes use of the following three variables from these data:

- **ABILITY**: the number of "leadership workshops" that have been held in the community during the past 5 years
- **MOTIVATION**: the number of the community's local politicians who attended at least one "leadership workshop" during the past 5 years
- **WEALTH**: the community's median income (in thousands of dollars)

Some numbers:

<table>
<thead>
<tr>
<th></th>
<th>ABILITY</th>
<th>MOTIVATION</th>
<th>WEALTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABILITY</td>
<td>1.00</td>
<td>.61</td>
<td>.38</td>
</tr>
<tr>
<td>MOTIVATION</td>
<td>.61</td>
<td>1.00</td>
<td>.27</td>
</tr>
<tr>
<td>WEALTH</td>
<td>.38</td>
<td>.27</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3.2</td>
</tr>
<tr>
<td>4</td>
<td>2.1</td>
</tr>
<tr>
<td>22</td>
<td>9.2</td>
</tr>
</tbody>
</table>

a. What proportion of "the variance in WEALTH that is not explained by MOTIVATION" is explained by ABILITY?

b. The square root of the proportion referred to in part a is itself a statistic. Name this statistic and find its 95% confidence interval.

c. Estimate the unstandardized partial slope between WEALTH and ABILITY from the regression of WEALTH on ABILITY and MOTIVATION.

d. State in words the meaning the regression coefficient (i.e., the slope estimate) obtained in part c.