Study Questions For Exam 1

1. The following is some computer output from a regression of U.S. residents' 1982 income (in thousands of dollars) on their occupational prestige and on their gender:

<table>
<thead>
<tr>
<th>MULTIPLE R</th>
<th>AAAAAA</th>
<th>ANALYSIS OF VARIANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>R SQUARE</td>
<td>BBBBBB</td>
<td>DF SUM OF SQUARES MEAN SQUARE</td>
</tr>
<tr>
<td>ADJUSTED R SQUARE</td>
<td>.41764</td>
<td>2 6352.17709 CCCCCCCCC</td>
</tr>
<tr>
<td>STANDARD ERROR</td>
<td>8.70875</td>
<td>RESIDUAL 8494.33377 75.84227</td>
</tr>
<tr>
<td>F = EEEEEEEE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---------------------- VARIABLES IN THE EQUATION ----------------------

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>B</th>
<th>SE B</th>
<th>BETA</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRESTIGE</td>
<td>.383478</td>
<td>.053334</td>
<td>.513902</td>
<td>FFFFFF</td>
</tr>
<tr>
<td>SEX</td>
<td>-9.467689</td>
<td>GGGGGGGG</td>
<td>-.406666</td>
<td>-5.690</td>
</tr>
<tr>
<td>(CONSTANT)</td>
<td>HHHHHHHHH</td>
<td>3.540808</td>
<td>3.774</td>
<td></td>
</tr>
</tbody>
</table>

a. Find values for each of the eight numbers that has been replaced by an "AAAAAA," "BBBBBB," etc. in the computer output.

b. Perform a single test of the null hypothesis that this regression model explains none of the variance in income. (Use the .01 significance level.)

c. Again using the .01 significance level, test the null hypothesis that the partial slope between income and occupational prestige equals zero.

2. The United States government recently considered a law that would ensure health care for all American citizens. Whether or not there will be such a law depends upon the attitudes that U.S. citizens have about universal health care. On the one hand, people are motivated by a "need for security" (i.e., a need for assurances that future health catastrophes can be afforded). On the other hand, people are motivated by a "need for solvency" (i.e., a need to pay for present debts). The stronger each of these needs, the more likely one is to be in favor of a law that guarantees universal health care.

You have data on the following three variables:

\[ Y = \] respondents' agreement/disagreement with the statement, "The United States needs a law that guarantees universal health care to all its citizens." Values on this variable range from 1 = strongly agree to 5 = strongly disagree.

\[ W = \] respondents' agreement/disagreement with the statement, "I'm the kind of person, who never gets sick." Values on this variable range from 1 = strongly agree to 5 = strongly disagree.

\[ X = \] the amount of the respondent's debt (in thousands of U.S. dollars)

Based on a sample of 150 adult U.S. citizens, data on these variables are as follows:
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Correlation Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>2.3</td>
<td>2.7</td>
<td>Y</td>
</tr>
<tr>
<td>W</td>
<td>4.1</td>
<td>3.1</td>
<td>W</td>
</tr>
<tr>
<td>X</td>
<td>2.44</td>
<td>1.87</td>
<td>X</td>
</tr>
</tbody>
</table>

a. What is the increment in the proportion of variance explained between a reduced model in which Y is regressed on W, and a complete model in which Y is regressed on W and X?

b. Is this increment a statistically significant amount of variance at the .05 level of significance?

c. Find the unstandardized regression coefficient associated with the variable, X, from the regression of Y on W and X. Express this coefficient in words that a layperson could understand.

d. Do the data provide evidence that the stronger one's "need for solvency" (as measured by X), the more likely one is to be in favor of a law that guarantees universal health care? (Be sure to JUSTIFY YOUR ANSWER with reference to your data.)

3. You theorize that an individual's research productivity is positively associated with both his/her intelligence and his/her energy. To test this you collect data from a random sample of 5 ISU social scientists on (a) the number of their publications in refereed journals in the past three years, (b) their IQ score (high scores mean high intelligence), and (c) the amount of time (in hours) they spend exercising each day. (NOTE: The more exercise someone gets, the more energy you believe they will have.) Your data are as follows:

<table>
<thead>
<tr>
<th>Data</th>
<th>Sums</th>
</tr>
</thead>
<tbody>
<tr>
<td>publications (P): 1 11 9 2 6</td>
<td>29 243</td>
</tr>
<tr>
<td>IQ (I): 104 115 110 106 109</td>
<td>544 59,258</td>
</tr>
<tr>
<td>exercise (E): 5 0 2 6 3</td>
<td>16 74</td>
</tr>
</tbody>
</table>

Cross products:
\[ \sum P_i \cdot I_i = 3225 \]
\[ \sum P_i \cdot E_i = 53 \]
\[ \sum I_i \cdot E_i = 1703 \]

a. Calculate the unstandardized regression equation that tests your theory.

b. Referring to the regression equation in part a, express each regression coefficient (and the constant) in the equation in words.

c. Referring to the regression equation in part a, explain how the data do or do not support your theory. (Hint: You will NOT need to test the significance of any estimates in getting your answer.)
d. Referring to the regression equation in part a, find a 95% confidence interval for the unstandardized partial slope between IQ and publications.

e. What is the proportion of the variance in publications that is explained by the marginal linear effects of IQ and exercise (i.e., by the regression equation in part a)?

f. Test whether the amount of variance found in part e is significant at the .05 level of significance.

g. What is the increment to the regression sum of squares due to exercise?

h. Is the increment found in part g significant at the .05 level of significance?

4. The "free rider effect" refers to the tendency for people to benefit from others' efforts. As people's behaviors become increasingly popular, theorists sometimes speak of a "free rider threshold" that must be met before the behaviors begin attracting the broad masses of society. For example, it was less than six months before Black Monday (i.e., the U.S. stock market crash in October 1987) that "average Americans" began buying stocks in an effort to benefit from an economic boom that professional investors had been benefiting from for more than a year. You believe that the value of stocks decreases to the extent that nonprofessional investors (i.e., free riding "average Americans") become involved in (usually risky) stock ventures. In contrast, a more classical theory contends that stock values decrease as inflation increases. Your variables are . . .

\[ Y = \text{the monthly increase (or decrease, if negative) in points of the Dow-Jones Industrial Average} \]
\[ W = \text{the monthly increase (or decrease, if negative) in the inflation index} \]
\[ X = \text{the monthly increase (or decrease, if negative) in the proportion of nonprofessional investors in the U.S. stock market} \]

You have data on these variables from each of the 48 months prior to the stock market crash. The variables are related as follows:
a. Find the standardized and unstandardized regression equations appropriate for evaluating the classical and free rider theories.

b. Express each standardized and unstandardized regression coefficient in words that a layperson could understand. (You may assume that this lay person knows what a standard deviation is.)

c. Consider a month during which the inflation index remained unchanged. What change in the Dow-Jones Industrial Average would you predict if during this month there was a 10% increase in the proportion of nonprofessional investors in the U.S. stock market?

d. How much of the variance in monthly increases in points of the Dow-Jones Industrial Average is explained both by increases in inflation rate and by increases in the proportion of nonprofessional investors in the U.S. stock market? Is this a statistically significant amount of variance at the .05 level?

e. What is the increment in the proportion of variance explained that is due to changes in the proportions of nonprofessional investors in the U.S. stock market? Is this a statistically significant amount of variance at the .05 level?

f. Do the data support your theory about the role of free-riding nonprofessional investors in determining stock prices? (Be sure to JUSTIFY YOUR ANSWER with reference to your data.)

5. Consider the following correlation matrix based on a random sample of 45 adult residents of San Francisco:

<table>
<thead>
<tr>
<th></th>
<th>INCOME</th>
<th>EDUCATION</th>
<th>MOTIVATION</th>
<th>ANXIETY</th>
<th>AGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCOME</td>
<td>1.00</td>
<td>.45</td>
<td>.36</td>
<td>-.28</td>
<td>.37</td>
</tr>
<tr>
<td>EDUCATION</td>
<td>1.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>MOTIVATION</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>ANXIETY</td>
<td>- .28</td>
<td>.00</td>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>AGE</td>
<td>.37</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(Be sure that you correctly read this matrix. For example, the correlation between EDUCATION and INCOME is \( r = .45 \).)

a. Give the standardized regression equation for the regression of INCOME on EDUCATION, MOTIVATION, ANXIETY, and AGE. (Assume that each of these
variables has a mean of zero and a variance of one.)

b. What proportion of the variance in INCOME is explained by EDUCATION, MOTIVATION, ANXIETY, and AGE?

c. Is the proportion found in part b statistically significant at the .05 level of significance?

6. Consider the following four variables:

Y = Gross national product (an annual measure of economic well-being) in billions of dollars.

W = Average family size in a country.

X = Annual foreign investment in millions of dollars.

Z = Number of ethnic or tribal minorities in a country.

You execute the following SPSS commands:

REGRESSION VARIABLES=Y,W,X,Z/DEP=Y/ENTER
REGRESSION VARIABLES=Y,W,X/DEP=Y/ENTER
REGRESSION VARIABLES=Y,W,Z/DEP=Y/ENTER
REGRESSION VARIABLES=Y,X,Z/DEP=Y/ENTER
REGRESSION VARIABLES=Y,W/DEP=Y/ENTER
REGRESSION VARIABLES=Y,X/DEP=Y/ENTER
REGRESSION VARIABLES=Y,Z/DEP=Y/ENTER

Parts of your output are as follows:

On the first regression:

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>B</th>
<th>STD ERROR B</th>
<th>BETA</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>-.00451</td>
<td>.03834</td>
<td>-.00317</td>
</tr>
<tr>
<td>X</td>
<td>.29631</td>
<td>.03968</td>
<td>.19576</td>
</tr>
<tr>
<td>Z</td>
<td>.58730</td>
<td>.04565</td>
<td>.34645</td>
</tr>
</tbody>
</table>
On the second regression:  On the fifth regression:

R-SQUARE    .03833  R-SQUARE    .00001

On the third regression:  On the sixth regression:

R-SQUARE    .12004  R-SQUARE    .03832

On the fourth regression:  On the seventh regression:

R-SQUARE    .15835  R-SQUARE    .12003

a. You decide to change the metric on X from millions of dollars to thousands of dollars. After making this change, what would be the value of the unstandardized regression coefficient associated with X that would be generated using the first regression statement?

b. Which of the seven regression models best fits the data? (Use the .05 level of significance and SHOW YOUR WORK.)

7. Individuals only pursue goals if they are willing to employ the means to these goals. The goal of "community health," for example, is less likely to be pursued by individuals with an aversion to associated behaviors (e.g., donating blood) than by individuals without such an aversion. Your theory is that exposing individuals to negative means-related imagery (e.g., regarding painful needle punctures) will lessen these individuals' willingness to pursue goals related to these means (e.g., blood donation). You design an experiment to test this theory. The experiment is performed on a random sample of 15 ISU undergraduates, each of whom was involved in the research during two consecutive days. On the first day each student completed a "willingness to donate blood" (WDB) survey. On the second day each student watched a movie with highly negative needle-puncture imagery. After the movie they each completed the same WDB survey. A data matrix with two variables and 30 cases was then generated. The following are data on your two variables:

W = the student’s WDB score (on a scale from 0 = no willingness to 100 = total willingness)

T = the time when the WDB survey was completed (0 = on the first day, 1 = on the second day)

The correlation between W and T is $r_{WT} = -0.2$. Means and standard deviations for the variables are as follows:
a. Find the unstandardized regression equation, and express the unstandardized slope (\( \hat{b} \)) in words that a lay person could understand.

b. What is the students' average WDB score on day 1? How much larger (or smaller) than this is their average score on day 2? (Hint: You will need to use the regression equation found in part a.)

c. Are any of the assumption(s) that underlie OLS regression analysis violated in this research? If so, which? Be sure to justify your answer based on the above description of the research.

8. You are doing research with an ISU professor who teaches meteorology. She is interested to learn whether her students' knowledge of weather processes improves if they use weather simulation software that she has developed. To make it easier for her students to gain access to this software, the professor has placed it on the university computer system. She does not require her students to use the simulation software, however.

Prior to accessing the simulation software, students are required to enter an ID and password. This allows you to monitor exactly how much time each student has used the software. Moreover each student’s knowledge of weather processes is measured at the end of the semester using his or her score on the course’s final exam. Finally, because you believe that smarter students will obtain high scores on the final exam (even if they did not use the simulation program), your analysis uses data on the following three variables:

\[ T = \text{Time (in minutes) using the weather simulation software} \]

\[ S = \text{Score on the final exam (ranging from zero to 100 points)} \]

\[ G = \text{Grade point average (ranging from zero to 4.0) as a measure of how smart the student is} \]

At the end of the semester you assemble the following data on the 30 students in the professor’s meteorology class:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>48.5</td>
<td>7.65</td>
</tr>
<tr>
<td>T</td>
<td>0.5</td>
<td>0.09</td>
</tr>
</tbody>
</table>

W = Standard Variable
T = Variable

\[ b \] = Unstandardized slope
\[ \hat{b} \] = Unstandardized slope
a. Find the unstandardized regression equation appropriate for testing whether students' knowledge of weather processes improves to the extent that they use the simulation software. In obtaining this equation be sure to take into account your belief that smarter students will obtain high scores on the final exam even if they did not use the simulation program.

b. A colleague looks over your work in part a and argues that your analysis is much more complicated than necessary. He advises you to ignore “how smart students are” (as measured by G) in the analysis and to focus entirely on the other two variables (i.e., T and S). Your response to this argument is that omitting G from the analysis would violate one of the assumptions underlying OLS regression analysis. State which assumption this is (writing it in matrix notation), and explain why you believe omitting G would result in the assumption’s violation. (Hint: Your explanation should do more than merely rephrase the assumption or repeat your belief that smart students get high exam scores. It should argue that these variables’ interrelations--possibly bolstered with references to your data--are such that the assumption’s violation would result if G were omitted.)

c. Referring to the regression equation in part a, express the partial regression coefficient between S and T in words that a layperson could understand.

d. Of the variance in final exam scores (S) not explained by grade point average (G), what proportion of this unexplained-by-G variance is explained by time using the simulation software (T)? Is this proportion statistically significant at the .05 level?

e. The professor with whom you are doing this research considers your sample of 30 students to have been drawn from the population of “all her (past and future) students.” Referring back to the description on page 1 of how your data were assembled, explain what assumption of OLS regression analysis was violated as data representative of this population were assembled? State the assumption using matrix notation, and explain what consequences this assumption’s violation has for your analysis.

9. Last year the National Institute for Mental Health (NIMH) awarded you a multi-million dollar grant to study whether or not Parenthood Training Intervention (PTI) is effective in preventing drug abuse among Iowa teenagers in grades 9 through 12. With this money you have begun gathering data in Mason City, Iowa, on high school students and their parents. The administrators at Mason City’s high school have cooperated with you in providing you access to their records on all the school’s students and their families.
You began the study last year by having all Mason City 8th graders (i.e., all Mason City students who were to enter high school during the next year) take notices home to their parents. The notices were invitations for the parents to participate in free parenthood training sessions that might help them in “preparing for the challenges of raising a teenager.” Training was only given to both of the student’s parents; no single parents were trained. In your study’s first statistical analysis, you examine data on the 30 couples who volunteered for and completed this training during the previous year (i.e., during the first year of your study). Each couple’s child graduated from 8th grade, and thus is now a 9th grader. To obtain a comparison group, you randomly sample 30 other couples who are parents of 9th graders within the Mason City school system. Your sample is thus of 60 Mason City 9th graders and their parents. Your variables are as follows:

\[ T = \text{A dummy variable for which } 1=\text{the student’s parents completed parenthood training and } 0=\text{the student/parents are in the control group} \]

\[ C = \text{The parents’ combined score on a 100-point “concern scale” on which } 0=\text{no concern that their 9th grade child might use drugs, and } 100=\text{total paranoia that their 9th grade child might use drugs} \]

\[ D = \text{The number of drug-related “incidents” regarding the student that were reported to high school authorities during year of the student’s 9th grade in high school} \]

a. In your statistical analysis you will be drawing inferences about some population. What population is this?

b. If you were to regress \( D \) on \( T \) and \( C \), could you assume that the randomness assumption had been met? Explain your answer.

c. If you were to regress \( D \) on \( T \) and \( C \), could you assume that your design matrix (i.e., \( X \)) is fixed? Explain your answer.

d. If you were to regress \( D \) on \( T \) only, could you assume that \( T \) (i.e., your measure of parents’ training) is uncorrelated with the errors associated with \( D \)? Explain your answer.

Now let’s assume that data on your 3 variables are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Deviation</th>
<th>Standard Correlation Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>.500</td>
<td>.504</td>
<td>( T ) 1.00, ( C ) 0.35, ( D ) 0.36</td>
</tr>
<tr>
<td>C</td>
<td>33.8</td>
<td>14.4</td>
<td>( C ) 1.00, ( D ) 0.99</td>
</tr>
<tr>
<td>D</td>
<td>3.00</td>
<td>1.39</td>
<td>( D ) 1.00</td>
</tr>
</tbody>
</table>

d. What proportion of the variance in the number of drug-related incidents (\( D \)) is explained by parents’ training (\( T \)) and concern (\( C \))? (Note: A single proportion is called for here.) Is this proportion statistically significant at the .05 level?
e. Find the unstandardized slope associated with T from the regression of D on T and C, and express this slope in words that a layperson could understand.

f. Based on the regression model estimated in part f, do your data provide statistically significant evidence that parenthood training reduces drug-related incidents among Mason City 9th graders? Use the .05 significance level.

You are doing research for the physical therapy center, “Back in Motion.” The center’s specialty is in back care, and the therapists there have been encouraging their patients to spend time walking on a treadmill at home. (Poor patients are provided treadmills at government expense, so you are sure that every patient in your study has a treadmill at home.) Your objective is to determine whether or not listening to music while walking on a treadmill increases physical therapy patients’ enjoyment of this exercise. In your analysis you also wish to take into account the intensity of the patient’s exercising experience. For this reason, you have equipped each treadmill with a measuring device that records not only how long the treadmill is used, but also how fast patients must have walked when they were using the treadmill. The device also measures the weight of the person on the treadmill, so that you can determine the times when it was being used by the patient (i.e., when the weight measure was close to the patient’s weight). You assume that the device’s time, speed, and weight measures are accurate, and that nobody else of the patient’s weight had access to each treadmill while your measurements were made.

Each treadmill was also equipped with headphones, such that the treadmill would only work if the person using the treadmill was simultaneously wearing the headphones. The headphones for half of the treadmills played classical music while the treadmill was in use. For the other half, low-level static (or “white noise”) was played through the headphones. All patients were told that they were to wear headphones during their exercise, “to keep from being distracted.”

After installing this equipment and after randomly determining whether each patient’s earphones would play music versus static, you began collecting data from all 87 physical therapy patients in the clinic with a treadmill at home. One month later, you assemble data from each treadmill on the following variables:

S = Sound (1=music; 0=static)
E = Enjoyment (the patient’s response to the question, "While walking on your treadmill during the past month, what percent of the time did you enjoy this exercise?" on a scale from 0=never to 100=always)

I = Intensity (the patient’s average miles per hour [mph] while on the treadmill during the month on a scale from 1=one mph to 20=twenty mph)

Your data on these variables are as follows:

<table>
<thead>
<tr>
<th>Correlation Coefficients</th>
<th>Sound (S)</th>
<th>Enjoyment (E)</th>
<th>Intensity (I)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound (S)</td>
<td>1.0</td>
<td>0.6</td>
<td>0.3</td>
<td>0.500</td>
<td>0.503</td>
</tr>
<tr>
<td>Enjoyment (E)</td>
<td>0.6</td>
<td>1.0</td>
<td>-0.4</td>
<td>40.0</td>
<td>15.5</td>
</tr>
<tr>
<td>Intensity (I)</td>
<td>0.3</td>
<td>-0.4</td>
<td>1.0</td>
<td>6.10</td>
<td>2.75</td>
</tr>
</tbody>
</table>

a. Find the unstandardized regression equation for the regression of E on S and I.

b. Express in words the meaning of the partial slope associated with S found in part a. (Hint: In your answer you may wish to deviate from the template in your notes.)

c. Explain (i.e., do more than merely state) whether or not the randomness assumption is met for the regression model estimated in part a.

d. Explain (i.e., do more than merely state) whether or not the assumption, E(X)=X, is met for the regression model estimated in part a. (Hint: This assumption has to do with X being fixed.)

e. Explain why the assumption, \(E(X^T e) = E(X^T) * E(e)\), may not be met for the regression model estimated in part a. (Hint: This assumption has to do with X not being correlated with the errors of Y.)

f. Despite any reservations you might have about the assumptions of regression analysis not being met in this case, does your analysis provide significant evidence (at the .05 significance level) that listening to music while walking on a treadmill increases physical therapy patients’ enjoyment of this exercise?

11. United States citizens are supposed to be flattered when Republican politicians speak of their “faith in the American people” as wise investors of their own money. Americans should be trusted, they argue, to invest that portion of their paychecks that now goes for their “Social Security.” By making wise investment decisions, every American would then take responsibility for financing her or his own retirement. (And, of course, this responsibility would thereby no longer belong to the U.S. government.) Your concern is that among Americans, women often lack the investment skills required for such a scheme to work.
Women are often less involved in family finances than their husbands, leaving them with relatively little investment experience. Moreover, women’s typically small incomes (relative to men) leave them hesitant to invest their scarce resources in the first place. Given that the U.S. government may soon require women to take charge of financing their own retirements, you decide to undertake an investigation into U.S. women’s investment behaviors. To ensure that your subjects have sufficient resources to make investments after household bills have been paid, you only sample women with incomes above $35,000 per year. Based on a random sample of U.S. e-mail addresses obtained from InfoUSA, you invite 400 women to respond to your web survey. In a preliminary analysis, you only consider data on the 114 single (i.e., unmarried and without live-in partner) respondents in your sample who provided data on the following four variables:

I = INVEST (amount of money that the respondent deposited in retirement accounts or pension plans during the past year, in thousands of dollars)

R = RINCOM (respondent’s income in thousands of dollars earned during the past year)

A = AGE (respondent’s age in years)

M = MAGE (mother’s age when the respondent was born, in years)

Your data on these variables are as follows:

<table>
<thead>
<tr>
<th>Correlation Coefficients</th>
<th>INVEST (I)</th>
<th>RINCOM (R)</th>
<th>AGE (A)</th>
<th>MAGE (M)</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>INVEST (I)</td>
<td>1.0</td>
<td>0.5</td>
<td>0.6</td>
<td>0.2</td>
<td>2.1</td>
<td>0.72</td>
</tr>
<tr>
<td>RINCOM (R)</td>
<td>0.5</td>
<td>1.0</td>
<td>0.3</td>
<td>0.1</td>
<td>62.0</td>
<td>12.0</td>
</tr>
<tr>
<td>AGE (A)</td>
<td>0.6</td>
<td>0.3</td>
<td>1.0</td>
<td>–0.5</td>
<td>42.5</td>
<td>10.0</td>
</tr>
<tr>
<td>MAGE (M)</td>
<td>0.2</td>
<td>0.1</td>
<td>–0.5</td>
<td>1.0</td>
<td>24.3</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Note that the negative correlation between AGE and MAGE (i.e., rAM=–0.5) indicates that older respondents were more likely born to younger mothers than were younger respondents.

a. The initial step of your analysis is to construct a measure of investment behavior (call it N, for ‘income-Neutral-investment’) that is unrelated to your respondents’ incomes (R). Explain how you would obtain this adjusted measure. (Hint: Provide an equation, possibly in the form of an SPSS compute statement—with numbers, please—for calculating N from data on the variables, I and R.) [weight 2]

b. Building on your answer in part a, answer the following question: “If during the past year a respondent’s income was $50,000 and she had deposited $2,000 in retirement accounts and pension plans, what would be her value on this new measure?” (Be sure to show your work.)
c. Correlations of N (i.e., the adjusted variable constructed in part a) with “respondent’s age” (A) and “mother’s age when the respondent was born” (M) are $r_{NA}=0.50$ and $r_{NM}=0.26$ respectively. Given that the mean value of N is 2.0 and its standard deviation is 1.0, find the unstandardized regression equation for the regression of N on A and M.

d. Referring to the regression model obtained in part c, what proportion of the variance in N is explained by A and M? (Hints: Find only a single proportion, and show how it was calculated.)

e. In interpreting your findings, you consider A to be a measure of “the respondent’s financial concerns as they approach retirement age” and M to be a measure of “the respondent’s exposure during early adulthood to the mother’s financial concerns as the mother approached retirement age.” (Your presumption with the latter measure is that women raised by elderly parents were exposed relatively early to financial concerns regarding retirement—exposure that, you believe, increased their financial concerns [and their investment motivations] for their own retirements.) Keeping in mind these interpretations of your independent variables as well as the meaning of your adjusted dependent variable, express in words the meaning of the partial slope associated with respondent’s age (A) found in part c.

f. Considering the partial slope associated with respondent’s age (A) found in part c (which is, incidentally, the same partial slope that was interpreted in part e), is this slope significantly larger than zero? (Hints: Use the .05 significance level and show your work.)

g. After looking at your above results, a colleague tells you that you have violated the assumption, $E(X^T e) = E(X^T)E(e)$. His argument is as follows:

“Because your respondents’ incomes (RINCOM) are correlated with both their investment behavior (INVEST, at $r_{IR}=0.5$) and their ages (AGE, at $r_{AR}=0.3$), you are erroneously attributing some of the variation in your respondents’ investment behaviors to their ages. That is, by not including RINCOM as an independent variable, you have left AGE associated with variations in respondents’ investment behavior (I) that are due to their larger incomes (R) rather than to their financial concerns as they approach retirement age (A).”

Your colleague is wrong, of course. What has he overlooked in your analysis? How does this oversight undermine his argument?