Problems: MODEL BUILDING & FITTING

1. Use the 1984 NORC data to test whether prestigious jobs pay off more for men than for women. (NOTICE that this test is not of whether women earn more than men.) To do this you must create a new variable as follows:

\[ \text{compute interact} = (\text{sex} - S) \times (\text{prestige} - P) \]

A few comments on this "interact" variable are in order:

(i) You will want to substitute 'S' with the mean score on the "sex" variable and 'P' with the mean score on the "prestige" variable.

(ii) This variable measures the interaction between sex and prestige. You will notice that for women, "interact" takes larger scores as "prestige" increases. For men, "interact" takes smaller scores as "prestige" increases.

(iii) An interaction measure is only included in a regression equation AFTER the variables from which it was constructed (here, "sex" and "prestige") are already in the equation. Thus to test whether prestigious jobs pay off more for men than for women, you must use the following regression command:

\[ \text{regression variables=rincome,sex,prestige,interact/dep=rincome/enter.} \]

a. If prestigious jobs pay off more for men than for women, would you expect "interact" to be positively or negatively correlated with "rincome"? Explain your answer.

b. Evaluate whether prestigious jobs pay off more for men than for women. What are your findings?

2. You have a sample of 26 wealthy ISU alumni who donate money to the university once each year. You want to predict the best time to ask them for money. First you note how much money each alumnus gave in 1985 and the increase or decrease in the Dow-Jones Industrial Average (an indicator of economic health or decline) during the month prior to their donation. You also note how many other charities have made appeals for money during the same month. Your three variables and related statistics are as follows:

\[ Y = \text{amount of money donated by alumni in thousands of dollars.} \]

\[ X = \text{number of "points" increase in the Dow-Jones Industrial Average (DJIA) in the month prior to the donation.} \]
\( W \) = number of other charities making appeals for money in the month prior to the donation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>3.2</td>
<td>2.3</td>
</tr>
<tr>
<td>X</td>
<td>6.4</td>
<td>5.1</td>
</tr>
<tr>
<td>W</td>
<td>4.1</td>
<td>3.2</td>
</tr>
</tbody>
</table>

\[ r_{YX} = .698 \quad r_{YW} = .451 \quad r_{XW} = .000 \]

a. You wish to evaluate whether one's appeals for money are more effective if many charities simultaneously make their appeals during a period when the economy is improving (i.e., when there is an increase in the DJIA). What variables would you include in a regression equation that could be used to estimate this combined effect? (Do NOT calculate any parameter estimates!) Be sure to show how all variables would be derived.

b. Obtain estimates for each of the three parameters in the model,

\[ Y = \hat{a} + \hat{b}_1 X + \hat{b}_2 W + \hat{e}, \]

where \( \hat{e} \) is the error term.

c. Referring to the regression equation in part b, identify what each parameter represents in terms of the dependent variable, \( Y = \) money donated in thousands of dollars. (Be sure to specify the units of each variable you identify.)

d. Convert the regression model in part b to a standardized regression model (i.e., obtain estimates for the two corresponding standardized regression slopes).

3. You wish to follow-up on a recent study that found the death of one spouse not to be immediately grieved by the surviving spouse when there were children living with the couple prior to the death. In their conclusions, the study's authors theorized that this delay-in-grieving was due to the surviving spouse's distraction from her/his own loss, while she/he helped the children deal with their grief over a lost parent.

You conduct face-to-face interviews with a random sample of 216 adult Iowans, whose spouse died during the past month. Half of these adults had children living with them prior to their spouse's death; the other half did not. Each adult was asked, "How difficult are you finding it today to deal with your spouse's recent death?" Interviewers then ranked responses to this question on a 10-point
scale from 1 = no grief to 10 = consumed by grief. In your data set this variable is named, GRIEF, and may be assumed to be an interval-level measure. Your data set also includes TIME (the number of days since the spouse's death) and KIDS (a dummy variable for which 1 = children living with the couple prior to the spouse's death and 0 = no children living with the couple prior to the spouse's death). Correlations, means, and standard deviations among these variables are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Correlation Coefficients</th>
<th>Means</th>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GRIEF</td>
<td>TIME</td>
<td>KIDS</td>
</tr>
<tr>
<td>GRIEF</td>
<td>1.00</td>
<td>-.32</td>
<td>-.11</td>
</tr>
<tr>
<td>TIME</td>
<td>-.32</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>KIDS</td>
<td>-.11</td>
<td>.00</td>
<td>1.00</td>
</tr>
<tr>
<td>MYVAR</td>
<td>.22</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

(Make sure you read these data correctly. For example, the zero-order correlation between GRIEF and TIME is –.32; the mean GRIEF score is 5.5; the standard deviation for GENDER is 2.5; etc.)

a. What complete and reduced models would you use to evaluate whether surviving spouses with live-in children are more likely than other surviving spouses to delay their own grieving? (Hints: To answer this question, you must explain how a fourth variable, MYVAR, would be computed. Assume that MYVAR has correlations, mean, and standard deviation as given above. Finally, neither this part nor part b require that you calculate any numbers!)

b. If your data were consistent with those of the previous study (i.e., of the study that yours "follows-up"), what would be the sign of the partial slope associated with the variable, MYVAR? Explain your answer.

c. Using the .05 significance level, test whether your data provide significant evidence that surviving spouses with live-in children are more likely than other surviving spouses to delay their own grieving. (Hint: The last seventeen words in this sentence are identical to those underlined in part a of this question.)

4. With the fall of the Soviet Union in 1991, Uzbekistan (a small country located a few hundred miles east of the Caspian Sea) became an independent state. Political independence brought the country fundamental economic changes as a controlled communist economy was replaced by a free market economy.
Independence also allowed the emergence of cultural norms that had been illegal under Soviet rule. For example, in rural towns such as Namangan "religious patrols" now roam the streets enforcing the veiling of women, the ban of alcohol, and other Islamic practices. In Uzbekistan's cities Islamic norms are still prevalent, but are less militantly enforced. An Islamic norm of particular interest to you in your research is the secondary role of women to men in economic matters. It is because of this secondary role, you believe, that Uzbek women are less likely to be employed in rural areas than in urban areas. In contrast to this belief, other researchers have argued that this tendency for urban women to be employed results simply because there are more jobs available in cities than in rural areas. You decide to investigate whether cultural biases or a lack of jobs undermines Uzbek women's labor market participation. You obtain data on the following three variables for each of 21 Uzbek cities and towns:

**RELIG**
number of persons per 1,000 who visit a mosque at least weekly
(This is your measure of the prevalence of Islamic norms in a region.)

**NOEMPLOY**
number of unemployed persons per 1,000 population (This is your measure of the availability of jobs. Your position is that if a region's unemployment rate is high, the region will have fewer jobs available.)

**WOMEN**
number of women employed per 1,000 workers (This is your measure of women's participation in the labor force.)

Correlations, means, and standard deviations among these variables are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Correlation Coefficients</th>
<th>Means</th>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>RELIG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOEMPLOY</td>
<td>.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMEN</td>
<td>-.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RELIG</td>
<td>1.00</td>
<td>400</td>
<td>100</td>
</tr>
<tr>
<td>NOEMPLOY</td>
<td>.80</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>WOMEN</td>
<td>-.40</td>
<td>200</td>
<td>72</td>
</tr>
</tbody>
</table>

(Be sure that you read these data correctly. For example, the zero-order correlation between RELIG and NOEMPLOY is .80; the mean RELIG score is 400 persons; the standard deviation for RELIG is 100 persons; etc.)

a. Find the unstandardized regression equation (i.e., compute numbers, please) for the regression of WOMEN on RELIG and NOEMPLOY. Express in words the meaning of the partial slope associated with NOEMPLOY).
b. Find the total, regression, and error sums of squares for the regression described in part a. Do the marginal effects of both RELIG and NOEMPLOY explain (at $\alpha = .05$) a significantly large proportion of the variance in WOMEN? Be sure to state your conclusion in words. (Hints: The phrase "marginal effects of both" refers to the total effects of two variables, not part of the effect of one, or of some single measure constructed out to the two. The total sum of squares can be obtained based on your knowledge of the sample size and the standard deviation of the variable, WOMEN.)

c. At the .05 significance level and based on coefficients computed in part a, determine if either the partial effects of cultural biases or those of a lack of jobs undermines Uzbek women's labor force participation. (Hints: In your answer you will need to perform one or more hypothesis tests related to coefficient[s] found in part a. Be sure and remember to state your conclusion in words.)

d. Using the data provided at the beginning of this problem, obtain an equation (with numbers, please) that would convert NOEMPLOY into a new variable, NOEMPLOYADJUST, such that this new variable has the linear effects of RELIG removed. What would be the value of the slope (again, a number is called for here) in the bivariate regression of WOMEN on NOEMPLOYADJUST? Explain your answer, please.

e. You decide to expand your analysis to test your belief that Islamic cultural norms only work to Uzbek women's disadvantage in cities/towns with high unemployment, but that they do not have this effect in cities/towns with low unemployment. What complete and reduced models would you estimate in evaluating this hypothesis? (Hints: To answer this question, you must explain how a fourth variable, MYVAR, would be computed. Note that no numbers need be calculated to answer either this part nor part f of this question.)

f. In the complete model described in part e, what would you expect to be the sign of the slope associated with the variable, MYVAR, if it were true that Islamic cultural norms only work to Uzbek women’s disadvantage in cities/towns with high (but not low) unemployment? Be sure to show how you arrived at your answer.

5. Polynomial terms in regression equations allow you to estimate nonlinear relations between variables. The following exercise illustrates how the shapes of regressions lines change as do the regression coefficients associated with the linear and polynomial terms in regression equations.
a. Draw rough sketches of the following four mathematical functions on the same set of axes, for integer values of X between 0 and 10. Be sure to notice how the function changes shape as the quadratic coefficient changes.

\[
\begin{align*}
Y &= 5 + .5 \times X \\
Y &= 5 + .5 \times X + .1 \times (X - 5)^2 \\
Y &= 5 + .5 \times X + .2 \times (X - 5)^2 \\
Y &= 5 + .5 \times X + .5 \times (X - 5)^2
\end{align*}
\]

b. Now do the same on a second set of axes with the following four functions:

\[
\begin{align*}
Y &= 5 + .5 \times X \\
Y &= 5 + 1.0 \times X - .2 \times (X - 5)^2 \\
Y &= 5 + .5 \times X - .2 \times (X - 5)^2 \\
Y &= 5 - .2 \times (X - 5)^2
\end{align*}
\]

c. On a third set of axes, plot the following functions:

\[
\begin{align*}
Y &= 5 + 0.5 \times X \\
Y &= 5 + 0.5 \times X + 0.1 \times (X - 5)^3
\end{align*}
\]

6. In the jungles of southern Mexico, members of the Yucanotec tribe spend countless hours grooming each other. This grooming behavior has many functions. It provides occasions for discourse on tribal folkways, daily activities, etc. And it also serves to rid tribal members of vermin.

A recent beetle infestation of the Yucanotechs' tribal grounds has led to an alarming increase in instances of Lyme disease there. Before rushing to aid the Yucanotechs, the Mexican government has consulted both medical and anthropological experts. The medical experts have developed some computer software that would allow the Yucanotechs to "teach themselves" how to identify and remove the stag beetles (or ticks) that are the carriers of Lyme disease. However, the anthropologists argue that computers should not be used to teach Yucanotechs about ticks, because, according to Professor Inobetta of Mexico University, "Computers are addictive. Once Yucanotechs start using the computers, they will use them more and more, with the unfortunate result that little time will be left for grooming behavior. As grooming behavior declines, the process of information and folkway transmission will be undermined."

The Mexican government decides to proceed with the medical experts' recommendation to let the Yucanotechs use computers to learn about how to identify and remove ticks. However, they commission you to study the effects that Yucanotechs' introduction to computers has on their grooming behavior. You randomly sample 43 Yucanotechs, who were introduced to computers during the
government’s implementation of the medical experts’ recommendation. You closely monitor these 43 subjects and record data for the following two variables:

**GROOMTIME**  
Hours per day that the subject spent grooming other tribal members

**COMPUTIME**  
Hours per day that the subject spent using the computers provided by the Mexican government

You are surprised to discover that these two variables have a zero correlation with each other. Their means and variances are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROOMTIME</td>
<td>12.5</td>
<td>101.3</td>
</tr>
<tr>
<td>COMPUTIME</td>
<td>3.2</td>
<td>55.7</td>
</tr>
</tbody>
</table>

a. What null and alternative hypotheses correspond to Prof. Inobetta’s views regarding GROOMTIME and COMPUTIME?

b. For this section of this problem only, assume that your sample is in fact unrepresentative of all Yucanotecs. Instead, assume that among all Yucanotecs each additional hour per day using the computer yields a decrease of one hour per day grooming other tribal members. Given this assumption, what is the power of the hypothesis listed in part a? Use the .05 significance level. (Hint: Recall that your data suggest that COMPUTIME explains none of the variance in GROOMTIME. Thus in this case, $SS_{\text{REGRESSION}} = 0$ and $SS_{\text{ERROR}} = SS_{\text{TOTAL}}$.)

c. Upon examining a plot of your data, you discover that grooming behavior was greatest among Yucanotecs who spent moderate amounts of time using the computers. (Presumably, they had learned the tick-grooming message of the computer software.) Yucanotecs who spent large amounts of time using the computers, spent considerably less time grooming others. (A finding consistent with Prof. Inobetta's views.) However, the Yucanotecs who made the least use of the computers were typically upset by the way in which "computer demons" fascinated many of their fellow tribe members and (as a consequence) they rejected the demons' pro-grooming message and displayed less grooming behavior than average. What regression equation would allow you to model this pattern among your data? (Hints: You will need to compute a new variable, NEWVAR, for this model. Be sure and show how NEWVAR would be derived. Do NOT calculate any parameter estimates at this point.)

d. Assume that the mean of NEWVAR is 3.0 and that its variance is 220. Furthermore, assume that NEWVAR has a zero correlation with COMPUTIME and a correlation of −.35 with GROOMTIME. Give (i.e., calculate all coefficients
and display them in equation form) a properly specified unstandardized regression equation that includes NEWVAR.

e. Referring to the equation found in part d, test the null hypothesis that the slope associated with NEWVAR equals zero. (Use $\alpha = .05$.)

7. Consider the following argument: Since high school graduates (i.e., Americans who have completed their twelfth grade of education) are so common, there is less demand for their employment than for the employment of people with either higher or lower educations. As a result, people with around twelve years of schooling are paid less than people who have educations that are much greater or lesser than this. Use the 1984 NORC data to evaluate this argument:

a. You must construct a new variable (call it "newvar") to evaluate the argument. Give the appropriate SPSS "compute" statement needed to construct this variable. (Hint: You may need to use the variable, "educ" [highest grade of education completed], in this problem.)

b. Include this "compute" command in a computer run that evaluates the above argument using the SPSS stepwise regression procedure.

c. Do the data support the above argument? Explain your answer.

In doing this problem you should find that "newvar" does not enter into the regression equation when the .05 significance level (the default level in SPSS) is the criterion of entry. You should be able to get "newvar" to enter the equation by increasing your significance level to the .10 level, however. You can do this by changing your "regression" command to read as follows:

```
regression vars= (variable list) /criteria=pin(.10)/dep=rincome/stepwise.
```

NOTE: You must, of course, change "(variable list)" to a list of the variables required to evaluate the theoretical argument.

8. The older a mother is at childbirth, the greater the chance is that her baby will have a serious birth defect. The proportion of mothers giving birth to babies with serious birth defects increases with the mother's age more slowly at younger than at older ages. For example, the proportion of 35 year-old women who give birth to children with serious birth defect(s) is only 0.4% higher than the proportion of 34 year-old women who give birth to such children. In contrast, the proportion of 40 year-old women who give birth to children with serious birth defect(s) is 2.0% higher than the proportion of 39 year-old women who give birth to such children. That is, the increase in the chance of giving birth to a child with a serious birth defect is five times as large when a woman ages from 39 to 40.
than when she ages from 34 to 35! This increase also extends to the severity of children's birth defects: There is a greater increase in the severity of birth defect(s) with aging among older than among younger women.

You are doing research for Bristol-Myers on a new drug which, when taken orally by pregnant women, is believed to reduce the severity of birth defects among babies that were diagnosed during pregnancy as fetuses having birth defect(s). Your company has received permission from the U.S. Food and Drug Administration to administer the drug to (and collect data from) a sample of 400 pregnant women, who (1) have been diagnosed as having a fetus with a birth defect and (2) volunteer to take part in your research. Half of the women (i.e., 200) take the drug prior to childbirth the other half take a placebo (viz., a sugar-pill). For the purposes of this problem you should assume that this sample of n=400 is a simple random sample.

You enter your data into an SPSS file having three variables:

DEFECT  severity of the child's defects at birth on a scale from 0 = nobirth defect(s) to 100 = 100 or more birth defects. (NOTE: No child had a score of 100 on "defect".)

AGE  age of mother

DRUG  1 = used drug
      2 = did not use drug

NOTE: The average age among the 400 women is 35 and the average severity of the women's children's birth defects is 20.

a. You begin your analysis by estimating the pattern (described in the first paragraph) according to which children born to older women are likely to have more severe birth defects than children born to younger women. What SPSS commands would you use to estimate (or model) this pattern? Be sure to show (in SPSS compute statement(s)) how any variable(s) other than "defect", "age", or "drug" may have been derived.

b. What complete and reduced models would you use to evaluate whether (after taking into account the variation in severity of defect(s) due to mothers' ages) the drug reduces the severity of birth effects? Again, be sure to show (in SPSS compute statement(s)) how any variable(s) other than "defect", "age", or "drug" may have been derived. (Hint: In deriving the reduced model, be careful only to remove variable[s] in the complete model that are relevant to your research question.)
c. Using the following (partial) SPSS output from the complete model, express in words the meaning of the unstandardized partial regression coefficient between "defect" and "drug":

\[
\begin{align*}
R \text{ SQUARE} & \quad .10246 \\
\text{------------------------ VARIABLES IN THE EQUATION ------------------------} \\
\text{VARIABLE} & \quad B & \quad \text{STD ERROR B} \\
\text{DRUG} & \quad 1.154 & \quad 0.414 \\
\text{NEWVAR} & \quad 0.937 & \quad 0.221 \\
\text{AGE} & \quad 0.191 & \quad 0.025 \\
\text{(CONSTANT)} & \quad -0.10436 & \\
\end{align*}
\]

d. If 9% of the variance in "defect" was explained by variables in the reduced model, would you conclude that the drug reduces the severity of birth defects among babies that were diagnosed during pregnancy as fetuses having birth defect(s)? Explain your answer and use the .05 level of significance. Also be sure to specify how you are assuming that NEWVAR was derived.

e. Imagine that you rerun the regression for the complete model described in parts "b" and "c" after inserting the following SPSS statement before the "regression" command:

\[
\text{recode drug ( 2 = -1 ).}
\]

What would be the unstandardized partial regression slope between "defect" and "drug"? Show how you arrived at your answer.

9. In the past when criminals assaulted their victims, these victims' grievances against the criminals were never heard in court. Now in some parts of the United States, victims are allowed to submit "victim impact statements" to judges prior to the sentencing of the criminals that assaulted them. Since some research has shown that these victim impact statements influence judges' sentencing decisions, you undertake research into what factors influence victims' recommendations for the punishments that the criminals should be subjected to.

You draw a random sample of 50 victim impact statements that Chicago residents submitted to Chicago's First District Court during 1989. You assemble data on the following variables from different parts of the statements:

\[
\text{GENDER} = 1 \text{ if female; } 0 \text{ if male.}
\]
WHEREV = 1 if the crime took place in a relatively crime-free part of Chicago; 0 if the crime took place in a part of Chicago with much crime.

CRIMES = the number of criminal acts mentioned in the victim's description of how he/she was victimized. (Note: This is a measure of the severity of the victimization.)

PUNISH = the number of punitive acts mentioned among the victim's recommendations for the criminal's punishment.

Eye-for-an-Eye Theory argues that people, who are severely victimized, will recommend more punishment for the criminal than people, who are less severely victimized. In contrast, I-Asked-for-It Theory argues that people, who are victimized in crime-ridden locations, are less likely to view the criminal as guilty (i.e., the victim views him/herself as partly at fault for the crime, because he/she allowed him/herself to be in a place where such a crime was likely) and thus are less likely than people, who are victimized in crime-free locations, to let the severity of their victimizations influence the amount of punishment they recommend for the criminal.

Your data are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Correlation Coefficients</th>
<th>Means</th>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GENDER WHEREV CRIMES PUNISH NEWVAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>1.00</td>
<td>.00</td>
<td>-.50</td>
</tr>
<tr>
<td>WHEREV</td>
<td>.00</td>
<td>1.00</td>
<td>-.30</td>
</tr>
<tr>
<td>CRIMES</td>
<td>.00</td>
<td>-.30</td>
<td>1.00</td>
</tr>
<tr>
<td>PUNISH</td>
<td>-.50</td>
<td>-.20</td>
<td>.60</td>
</tr>
<tr>
<td>NEWVAR</td>
<td>.00</td>
<td>.10</td>
<td>.15</td>
</tr>
</tbody>
</table>

(Make sure you read these data correctly. For example, the correlation between GENDER and PUNISH is -.50; the mean GENDER score is .76; the standard deviation for GENDER is .43; etc.)

a. Assuming that victims will be more likely to want to punish criminals, the more they view these criminals as guilty, what complete and reduced models would you use to evaluate I-Asked-For-It Theory? (Hints: This problem requires that you explain how you would compute a new variable, NEWVAR. Assume that NEWVAR has the mean, standard deviation, and correlations as given above. Other than in showing how NEWVAR would be computed, neither this part nor part b involves your calculating any numbers.)

b. What signs (if any) do each of the theories indicate will be the signs of the parameters estimated in the complete model described in part a? Explain your answers.
c. Calculate the unstandardized partial regression coefficient for the CRIMES variable from the regression of PUNISH on WHEREV and CRIMES.

d. Say in words the meaning of the coefficient estimated in part c. (Be sure to include the appropriate numerical values in your renderings of these meanings.)

e. A preliminary analysis has shown that 46% of the variance in PUNISH is explained by WHEREV, CRIMES, and NEWVAR. What proportion of the variance in PUNISH that is neither explained by WHEREV nor by CRIMES is explained by NEWVAR? What statistic does this proportion refer to?

f. As is evident from the correlation matrix at the beginning of this problem, your data show women to recommend fewer punishments than men. Noting the −.5 correlation between GENDER and PUNISH, a colleague suggests that your findings on the effects of WHEREV, CRIMES, and NEWVAR on PUNISH might change once the marginal effect of GENDER is estimated in the regression analysis. The colleague is wrong, of course. You point out to him that GENDER is uncorrelated with WHEREV, CRIMES, and NEWVAR. "Sure," the colleague responds, "but GENDER is correlated with PUNISH, right?" How would you go on to explain why adding gender into the regression analysis would not alter your findings on the effects of WHEREV, CRIMES and NEWVAR on PUNISH? (Hint: No calculations are needed to answer this question.)

10. Absenteeism (occurrences of employees’ absence from work) is more common among women than men. There are two theories that explain this. Sincerity Theory states that men are absent less often because they are more serious than women. (Presumably men have a "serious gene" that women lack.) On the other hand, Responsibility Theory states that women with small children miss work more often than other people, because they regularly choose to meet their children’s needs instead of their employer’s needs. (Presumably mothers with children at home are unlike other people in that they feel more responsibility to their children than to their employers.)

You are doing research for a large pharmaceutical company on employee absenteeism. Your data are from a random sample of 150 company employees.

You enter your data into an SPSS file having four variables:

ABSENT number of times the employee was absent from work during the past year
KIDS 1 = employee has children at home
0 = employee does NOT have children at home

GENDER 1 = employee is female
2 = employee is male

AGE age of employee (in years)

a. Using data on both female and male employees, what complete and reduced models would you use to evaluate Responsibility Theory? Show how any variable(s) other than "absent", "age", "kids", or "gender" were derived. (Do NOT attempt to find parameter estimates [i.e., no numbers please].)

b. What (if anything) do Sincerity and Responsibility Theories suggest will be the signs of all slopes in the complete model you have given in part a? Explain your answers.

c. Imagine that part of the SPSS output from the regression of "absent" on "gender" and "kids" is as follows:

<table>
<thead>
<tr>
<th>R SQUARE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.21357</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>B</th>
<th>STD ERROR B</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER</td>
<td>0.755</td>
<td>0.825</td>
</tr>
<tr>
<td>KIDS</td>
<td>1.048</td>
<td>0.332</td>
</tr>
<tr>
<td>(CONSTANT)</td>
<td>5.20547</td>
<td></td>
</tr>
</tbody>
</table>

Express in words the meaning of the unstandardized partial regression coefficient between "gender" and "absent".

d. Imagine that you rerun the regression of "absent" on "gender" and "kids" after inserting the following statement before the "regression" command:

recode gender (2 = –2).

What would be the unstandardized partial regression slope between "absent" and "gender"? Show how you arrived at your answer.

e. Your boss suggests that you begin your analysis by evaluating whether female employees in their child-bearing years (around age 24) have higher absentee rates than both younger and older female employees. What complete and reduced regression models would you use to evaluate this?
In answering this question you may either give the regression equations that you would estimate, or the SPSS commands you would use, in making the evaluation. Be sure to show (e.g., in SPSS compute statement(s)) how any variable(s) other than "absent", "age", "kids", or "gender" may have been derived. (Do NOT attempt to find parameter estimates [i.e., no numbers please].) Also NOTE that this first analysis should be performed on data from WOMEN ONLY. Accordingly, your first SPSS command will be as follows:

select if (gender eq 1).

11. Only four Eastern European countries experienced complete breaks from communist rule during the years immediately following 1989. By October of 1991 Poland, Hungary, Russia (by then an independent nation), and Czechoslovakia (yet to be divided) each had popularly elected governments run by noncommunist majorities. Prior to 1989 the news media in these countries routinely "interpreted" events in ways that justified the ruling communist party's policies. After 1989 the media continued interpreting the news, oftentimes with a lack of regard for professional journalistic standards. Even from within the four countries' journalistic communities, one frequently hears complaints about journalists who "freely mix news with commentary and shortness on objectivity."

You are interested in determining whether the emergence of journalistic professionalism has been hampered by ethnic conflicts. Your reasoning is that when a nation is dealing with ethnic conflicts, its citizens will become less objective (i.e., "we are good; they are bad") in their evaluations not only of the conflicting groups, but also of other individuals and organizations. You theorize that under conditions of ethnic conflict journalists like other citizens will become relatively quick to evaluate others, even in their professional writings. To test your theory you obtain English translations of 273 editorials randomly sampled from among all editorials that appeared between 1990 and 1996 in leading Polish, Hungarian, Russian, and Czech newspapers. From the texts of these editorials you generate data on the following variables:

EVALUATE the number of evaluative statements (i.e., statements that depict an individual or organization as 'good' or 'bad') appearing in each editorial (This number is used to measure how much journalistic professionalism [or objectivity] is evident in the editorial: the more evaluative statements, the less professionalism.)

YEAR the year in which each editorial was published (i.e., 1990, 1991, 1992, 1993, 1994, 1995, or 1996)

CONFLICT a dummy variable that takes the value, 1, if an editorial was published in a country with ethnic conflict (i.e., in Russia or
Czechoslovakia) and the value, 0, if an editorial was published in a country with relatively less ethnic conflict (i.e., in Poland or Hungary)

a. What complete and reduced models would you estimate in order to evaluate the hypothesis that among editorials published in countries with little ethnic conflict there was from 1990 to 1996 a greater decline in evaluative statements than among editorials published in countries with more ethnic conflict? (Hints: You do not need to estimate any parameters [i.e., no numbers, please]. Merely indicate the form of the complete and reduced models in a way that distinguishes which variables are independent and which is dependent. Be sure to indicate how any computed variables were constructed.)

b. Sketch a plot that illustrates how EVALUATE, YEAR, and CONFLICT would be interrelated if the hypothesis in part a (i.e., that among editorials published in countries with little ethnic conflict there was from 1990 to 1996 a greater decline in evaluative statements than among editorials published in countries with more ethnic conflict) were true.

c. How would you test whether your data support the hypothesis stated in part a? (Hints: As in part a, no numbers are required here. However, you will need to specify the strategy you would use to test the hypothesis: What statistics would you need? How would you evaluate them in deciding if you have support for the hypothesis?)

Some numbers for parts d, e, and f:

<table>
<thead>
<tr>
<th></th>
<th>EVALUATE</th>
<th>YEAR</th>
<th>CONFLICT</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVALUATE</td>
<td>1.00</td>
<td>-.32</td>
<td>.29</td>
</tr>
<tr>
<td>YEAR</td>
<td>-.32</td>
<td>1.00</td>
<td>-.12</td>
</tr>
<tr>
<td>CONFLICT</td>
<td>.29</td>
<td>-.12</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9</td>
</tr>
<tr>
<td>2.2</td>
</tr>
<tr>
<td>-.-</td>
</tr>
</tbody>
</table>

d. How much of the variance in EVALUATE is explained (nonexclusively) by YEAR and CONFLICT?

e. Is the amount of variance found in part c significantly large at the .05 level of significance?

f. Consider the regression of EVALUATE on just the two variables, YEAR and CONFLICT. Fixing your significance level at .05 you hypothesize that the slope between EVALUATE and YEAR will be negative, even after adjusting YEAR as if all editorials were published in countries with relatively little ethnic
conflict (i.e., as if CONFLICT=0). If the actual value of the partial slope between EVALUATE and YEAR (from the regression of EVALUATE on YEAR and CONFLICT) equals −0.1, what is the probability that you will make a Type II error in testing the hypothesis? (Note: Like parts d and e, this part of the problem is totally unrelated to the hypothesis considered in parts a, b, and c of this problem.)

12. You have data on a random sample of 22 male college football coaches (hereafter referred to as "the subjects") for the following three variables:

FEAR A measure of the subject's fear of being sued or fined for failing to warn his players to adhere to safe tackling practices: The number of college football coaches that were sued or fined by the subject's local regulatory board during the previous football season for failing to warn their players to adhere to safe tackling practices

CONCERN A measure of the subject's concern for his players: The number of hours per week that the subject spent with his players during the six months after the end of the last football season—and thus before a new football season had begun (Note: These measurements are presumed to be of stable character traits of the coaches. The measurements were taken off-season to assure that time with players was out of genuine concern, rather than as the self-serving means to improving team performance.)

WARNINGS The number of times during the previous football season that the subject warned his players that they must adhere to safe tackling practices

a. You begin your analysis by using SPSS to regress "warnings" on "fear" and "concern". Part of your output looks as follows:

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONCERN</td>
<td>1.500000</td>
</tr>
<tr>
<td>FEAR</td>
<td>2.260925</td>
</tr>
<tr>
<td>(CONSTANT)</td>
<td>−3.999348</td>
</tr>
</tbody>
</table>

In plain English that a nonstatistician would understand, state the meaning of the unstandardized partial slope between "warnings" and "concern". (Hint: Be sure to include the value of the slope [i.e., the number] in your statement. Also, do not assume that "concern" and "fear" are uncorrelated. For example, "the worriers" among the coaches may be worried about both their players and themselves. On the other hand, "concern" and "fear" may be negatively correlated as "the selfless" among the coaches tend to forget about their own fears as they show concern for their players.)
b. In part a the positive partial slope between "fear" and "warnings" suggests that coaches tend to give more warnings as their fear of being fined/sued increases. Yet after looking more closely at a plot of the relation between these two variables, you discover that this tendency levels off for intermediate levels of fear. On the axes below, draw a regression line that would fit this pattern of "increased warnings with increasing fear" at both low and high levels of fear and of "invariant warnings with increasing fear" at intermediate levels of fear.

![Diagram](image)

c. What complete and reduced models would you use to evaluate whether coaches' warnings tend to level off for intermediate levels of fear? (Show how all higher-order variables in the complete model would be derived. Please give no parameter estimates.)

13. Exercise has clear health benefits, particularly for people who are liable to have a sedentary lifestyle. Paraplegics (i.e., people paralyzed from the waist down) are especially liable in this respect. Unable to walk and confined to a wheelchair, paraplegics can easily lose their physical strength unless they regularly engage in some form of exercise.

You convince the administrators of three local hospitals to provide you with names and addresses of all paraplegics listed in their medical records. From this information you compile a list of 60 paraplegics in your county—a list that you assume is representative of the population of all paraplegics nationwide. Every day for a week you begin sending letters and making phone calls to the 60 paraplegics, encouraging them to participate in a special basketball program that you are organizing for their benefit. You provide free transportation to any paraplegic who chooses to attend. The basketball program consists of 30 evenings when the paraplegics can play basketball together. Every evening before play begins, you take a "one repetition maximum" measure of each paraplegic in attendance. Unfortunately 20 of the paraplegics never attend your
basketball program, leaving you with data on the following variables for the 40 paraplegics who attended the program at least once:

ATTENDANCE (A) the number of evenings that the paraplegic attended the program (Note: Values on this variable range from 1 to 30.)

ONE-REP-MAX (O) the paraplegic's "one repetition maximum" (i.e., the maximum number of pounds that can be lifted with a single arm) prior to the last evening that the paraplegic attended the program (Note: This variable measures strength, and is obtained by lifting increasingly heavy weights until a weight is reached that can no longer be lifted. The maximum weight one can lift is one's "one repetition maximum.")

FATIGUE (F) the number of times weights were lifted before a weight was reached that the paraplegic could not lift

Note that the "one repetition maximum" measure is problematic in that one becomes fatigued in the process of lifting weights before a weight is found that one can no longer lift. The idea here is that someone who is more fatigued (due to the fact that s/he has lifted many weights) will have a lower "one repetition maximum" than someone who is less fatigued (given that s/he has lifted few weights). The below correlation, r_{OF} = –.40 , bears this out.

Correlations, means, and standard deviations among these variables are as follows:

<table>
<thead>
<tr>
<th></th>
<th>ATTENDANCE</th>
<th>ONE-REP-MAX</th>
<th>FATIGUE</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTENDANCE</td>
<td>1.00</td>
<td>.60</td>
<td>–.70</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>ONE-REP-MAX</td>
<td>.60</td>
<td>1.00</td>
<td>–.40</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>FATIGUE</td>
<td>–.70</td>
<td>–.40</td>
<td>1.00</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

(Make sure you read the correlation matrix correctly. For example, the correlation between paraplegics' attendance in the program and their strength (as measured by their last "one repetition maximum") is .60. The mean attendance score was 12 evenings, the mean strength score was a "one repetition maximum" of 30 pounds, and the mean fatigue score was 6 lifts.)

a. Explain how to derive a measure of strength from ONE-REP-MAX that is adjusted for paraplegics' tendency to have lower "one repetition maximums" if they are fatigued from lifting many weights. Using this measure, what would be the "adjusted one repetition maximum" of a paraplegic whose score on ONE-REP-MAX was 10 after having lifted weights 5 times (i.e., a paraplegic for whom FATIGUE = 5 )?
b. Of the variance in strength (i.e., in ONE-REP-MAX) not explained by fatigue, what proportion is explained by attendance?

c. Find the unstandardized partial slope associated with ATTENDANCE from the regression of ONE-REP-MAX on ATTENDANCE and FATIGUE. Express the meaning of this partial slope in words that a lay person could understand.

d. Using the partial slope calculated in part c and the .05 significance level, test the hypothesis that the more that paraplegics attended the program the stronger they became.

e. A closer examination of your data suggests that the paraplegics in your study gained most strength during the first 12 evenings that they attended the program. After that their strength increased only slightly. Show how you would construct (i.e., give an equation for) a new independent variable (NEWVAR) to model this pattern in your data. Give the complete and reduced models that you would compare to evaluate whether or not this pattern fits your data. What would be the signs of all partial slopes in the complete model if your data were to fit this pattern? (Be sure that both models take into account the marginal effects of fatigue.)

14. Consider the following six variables from a sample of 26 ISU students:

ANTICIPATE (A) = "joy in knowing STAT 401 is almost at an end" on a scale from 0 = no joy to 100 = extreme joy

NEAREND (N) = "number of STAT 401 classes left to be attended" on a scale from 0 = no more classes to attend to 45 = forty-five more classes left to be attended

REQUIRED (R) = "if STAT 401 was taken as a required course" on a scale from 0 = elective (i.e., not required) to 1 = required

\[ INT (I) = (R - \bar{R}) \times (N - \bar{N}) \]
\[ QUAD (Q) = (N - \bar{N})^2 \]
\[ CUBE (C) = (N - \bar{N})^3 \]

At left below are six plots (lettered a. to f.) of regression lines drawn between ANTICIPATE and NEAREND. At right are fifteen hypothetical regression equations (numbered 1. to 15.).
Note that a pair of regression lines is plotted in two of the plots, one for students who took STAT 401 as an elective (R=0) and one for students who took the course as a requirement (R=1). Also note that in each equation the values (i.e., numbers) of the equation's slopes are given in parentheses. (No values are given for the equations' constants.) Negative slopes are given as negative numbers within parentheses, requiring the plus sign, '+', to always appear outside the parentheses. Only one of the 15 equations could yield the regression line(s) depicted in each plot. In the spaces provided below indicate which equation matches which plot.

a. The number of the equation that matches plot a is _________.

b. The number of the equation that matches plot b is _________.

c. The number of the equation that matches plot c is _________.

d. The number of the equation that matches plot d is _________.
e. The number of the equation that matches plot e is ________.

f. The number of the equation that matches plot f is ________.

15. Talcott Parsons (among other social theorists) distinguishes between people's instrumental versus gratifying motivations. A person's motivations are instrumental insofar as the person thinks of her behavior as useful in attaining something else. For example, a worker's motivations are instrumental to the extent that she views her job as merely the way she gets her paycheck. A person's motivations are for gratification insofar as the person thinks of her behavior as inherently valuable. For example, a worker's motivations are for gratification if she enjoys her work so much that she would keep her job even if she were independently wealthy.

Prior research has provided considerable evidence that variations in instrumental vs. gratifying motivations are strongly related to the agricultural innovations that farmers use. For example, farmers with instrumental motivations are likely to deplete nutrients from the soil, to use chemical fertilizers, and to sell land if its productivity drops. In contrast, farmers whose motivations are for gratification are likely to use sustainable farming methods that utilize the farm's resources in ways that preserve these resources over the long term. For example, dividing farming activities between crops and livestock would allow such sustainable methods as feeding crops to livestock and spreading manure for crops.

Given the long-term environmental advantages of sustainable farming methods, numerous programs have been introduced to institute them in many regions throughout the world. Unfortunately, studies of these programs have repeatedly shown that sustainable farming methods last only as long as the government programs that support them. Once programs end, farmers return to their old methods. Your theory is that when people (including farmers) are left to themselves, they have a natural tendency to shift toward instrumental motivation. You further theorize that motivations for gratification require social support (e.g., discussions about the pleasures of farming like ones experienced while participating in sustainable farming programs). You are interested in investigating the amount of social support that is needed to keep farmers' motivations from shifting from gratifying toward instrumental (due to their natural tendency toward the latter).

Your study is on the effects of the Caracas Rural Development Initiative (CRDI)—a government-sponsored program intended to instill and maintain gratifying motivations in 91 farmers from a rural community south of Caracas, Columbia. These 91 farmers are ones who volunteered to use the sustainable farming methods described to them during the CRDI's initial meeting with community members on July 1, 1996. On the first Saturday of every month after that date, a
CRDI official held a meeting (intended to increase the farmers' gratifying motivations and to answer their questions about sustainable farming methods) at the local community center. During July of 2001, you conducted interviews with the 91 farmers and obtained data on their instrumental versus gratifying motivations for farming. Based on their addresses and attendance records, you also obtained measures of how far each farmer lives from the community center where the monthly CRDI meetings were held, and of how many of the 60 CRDI meetings (between 1996 and 2001) they attended. Your data are as follows:

**MOTIVATE (M)** the farmer's motivation (on a 50-point scale from 1 = instrumental to 50 = gratifying)

**DISTANCE (D)** the distance (in kilometers) between the farmer's home and the community center where the CRDI meetings were held

**ATTEND (A)** the number of CRDI meetings (out of 60) that the farmer attended between July 1996 and July 2001

Correlations, means, and standard deviations among these variables are as follows:

<table>
<thead>
<tr>
<th></th>
<th>MOTIVATE</th>
<th>DISTANCE</th>
<th>ATTEND</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOTIVATE</td>
<td>1.000</td>
<td>-.117</td>
<td>.242</td>
<td>35.70</td>
<td>10.36</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>-.117</td>
<td>1.000</td>
<td>-.394</td>
<td>5.31</td>
<td>2.96</td>
</tr>
<tr>
<td>ATTEND</td>
<td>.242</td>
<td>-.394</td>
<td>1.000</td>
<td>32.68</td>
<td>9.61</td>
</tr>
</tbody>
</table>

(Notes: Please be sure that you read these data correctly. For example, the zero-order correlation between farmers' motivation scores and the distance between their homes and the community center is -.117. The mean motivation score is 35.7 points on the motivation scale; the standard deviation on this scale is 10.36 points.)

a. As just stated, your research objective is to investigate "the amount of social support that is needed to keep farmers' motivations from shifting from gratifying toward instrumental." A ceiling effect (i.e., evidence that gratifying motivation only increases with attendance for a while before it levels off) would provide evidence that monthly CRDI meetings are more frequent than is necessary to increase farmers' gratifying motivation. What complete and reduced models would you compare in a test for such a ceiling effect? If you were to find evidence of a ceiling effect, what would be the signs of all partial slopes in the complete model? (Hint: In this part as well as in parts b and c, you are advised to begin by sketching a plot of the effect described in the problem.)
b. A threshold effect (i.e., evidence that gratifying motivation only increases after attendance reaches a critical amount) would provide evidence that monthly CRDI meetings are less frequent than is necessary to increase farmers’ gratifying motivation. What complete and reduced models would you compare in a test for such a threshold effect? If you were to find evidence of a threshold effect, what would be the signs of all partial slopes in the complete model?

c. According to Cognitive Dissonance Theory, the more effort involved in accomplishing a task, the more likely one is to believe that the task’s objective was valuable. For example, farmers attend CRDI meetings to learn information. Much of this information is about the gratifying nature of farming. The CRDI meetings were carefully prepared to ensure that the more meetings farmers attended, the more they would learn about (and thus gain motivation regarding) the gratifying nature of farming. Yet according to Cognitive Dissonance Theory, the influence that each meeting has on farmers' motivation will be stronger among farmers who expend a lot of effort to attend CRDI meetings (as would happen among farmers who live far from the community center) than among farmers who expend little effort to attend the meetings. What complete and reduced models would you compare in a test of Cognitive Dissonance Theory? If your analysis were to reveal effects consistent with the theory, what would be the signs of all partial slopes in the complete model?

d. Find the unstandardized partial slope associated with the variable, ATTEND, from the regression of MOTIVATE on DISTANCE and ATTEND.

e. Using words that a lay person could understand, express the meaning of the partial slope found in part d.

f. Your theoretical perspective leads you to believe that (after taking into account the distance that farmers live from the community center) attendance at the CRDI meetings strengthens farmers' motivation for gratification. If .5 were the true value of the partial slope estimated in part d, what would be the power of performing a hypothesis test of this perspective? (Hint: Assume that prior to data collection you set .01 as the probability at which you were willing to make a Type I error.)
hand, in most Islamic countries (like Saudi Arabia) people appear ridiculous if they have an unvalued identity (e.g., that of rival for one's true love) and they act in a way inconsistent with that identity (e.g., they step on her/his toes while dancing).

You are interested in evaluating Peculiarity Theory in the light of such cultural differences. Your analysis uses data from a sample of ten (10) political cartoons from the Indian newspaper, Hindustan, and ten (10) political cartoons from the Saudi newspaper, Al Riyadh. Each political cartoon has a "target" (i.e., a person who is depicted as acting in a way that is inconsistent with her/his identity). You ask two panels of experts (3 Hindus for the Indian cartoons and 3 Saudis for the Saudi cartoons) to rank each cartoon according to how humorous it is and how valued its target's identity is. When these rankings are completed, you have data on the following three variables:

FUNNY (F) the humorousness of the cartoon (scored by the appropriate panel on a 10-point scale from 1 = not funny to 10 = hilarious)

VAL-IDENTITY (V) the value of the target's identity (scored by the appropriate panel on a 6-point scale from 0 = strongly unvalued to 5 = strongly valued)

NEWSPAPER (N) 1 = Hindustan; 2 = Al Riyadh

Correlations, means, and standard deviations among these variables are as follows:

<table>
<thead>
<tr>
<th>FUNNY</th>
<th>VAL-IDENTITY</th>
<th>NEWSPAPER</th>
<th>Mean</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUNNY</td>
<td>1.000</td>
<td>.524</td>
<td>-.544</td>
<td>4.300</td>
</tr>
<tr>
<td>VAL-IDENTITY</td>
<td>.524</td>
<td>1.000</td>
<td>-.722</td>
<td>3.250</td>
</tr>
<tr>
<td>NEWSPAPER</td>
<td>-.544</td>
<td>-.722</td>
<td>1.000</td>
<td>1.500</td>
</tr>
</tbody>
</table>

Notes: You may assume that FUNNY and VAL-IDENTITY comprise interval-level measures. Also make sure you read the correlation matrix correctly. For example, the zero-order correlation between the humorousness of a cartoon and the value of its target's identity is .524. The mean humorousness score is 4.3 points on the humorousness scale; the standard deviation on this scale is 2.452 points.)

a. Using the data provided on the previous page, find the unstandardized partial slope associated with the variable, VAL-IDENTITY, from the regression of FUNNY on VAL-IDENTITY and NEWSPAPER.

b. Again using the data provided on the previous page, find the 95% confidence interval for the partial slope calculated in part a.
c. Using words that a lay person could understand, express the meaning of the slope found in part a.

d. What complete and reduced models would you use to evaluate Peculiarity Theory? (Hint: Specifying these models requires that you construct a variable in addition to F [for FUNNY], V [for VAL-IDENTITY], or N [for NEWSPAPER]. Call this variable "X," and show how its values are derived.)

Imagine that you estimate the complete and reduced models described in part d, and that parts of your output (with output for one model on the left and output for the other model on the right) look as follows:

### Regression

#### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.576</td>
<td>.332</td>
</tr>
<tr>
<td></td>
<td>a. Predictors: (Constant), N, V</td>
<td></td>
</tr>
</tbody>
</table>

#### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficients</td>
<td>Beta</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>4.974</td>
</tr>
<tr>
<td>v</td>
<td>.556</td>
<td>.274</td>
</tr>
<tr>
<td>n</td>
<td>-1.654</td>
<td>-.346</td>
</tr>
</tbody>
</table>

a. Dependent Variable: F

### Regression

#### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.791</td>
<td>.626</td>
</tr>
<tr>
<td></td>
<td>a. Predictors: (Constant), X, N, V</td>
<td></td>
</tr>
</tbody>
</table>

#### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficients</td>
<td>Beta</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>3.968</td>
</tr>
<tr>
<td>v</td>
<td>.496</td>
<td>.245</td>
</tr>
<tr>
<td>n</td>
<td>-1.756</td>
<td>-.367</td>
</tr>
<tr>
<td>x</td>
<td>-3.183</td>
<td>-.543</td>
</tr>
</tbody>
</table>

a. Dependent Variable: F

e. Use the .05 significance level to determine whether or not you have statistically significant support for Peculiarity Theory. (Note that the complete model includes X [the variable described in part d] as an independent variable.)

f. Imagine that the variable, NEWSPAPER, were recoded such that 1 = Hindustan and 0 = Al Riyadh. Now imagine that you were to regress FUNNY on VAL-IDENTITY and NEWSPAPER after NEWSPAPER had been recoded in this way. Using the output on the previous page, calculate the value in this regression model of the unstandardized partial slope associated with the recoded NEWSPAPER variable.

g. Calculate the value that the constant from the regression imagined in part f (namely, the regression of FUNNY on VAL-IDENTITY and NEWSPAPER after NEWSPAPER has been recoded such that 1 = Hindustan and 0 = Al Riyadh).

h. Using words that a lay person could understand, express the meaning of the constant found in part g.
17. Years ago while working as a swimming coach you noticed that tall people tend to swim faster than short people. As a consequence, you only allowed tall people on your swimming team until you later started seeing these tall swimmers lose to (i.e., swim slower than) shorter swimmers from other teams. After speaking with the coaches of these other teams, you learned that their short swimmers started swimming faster after they got involved in a weightlifting program.

Last year you intentionally recruited 12 new swimmers who are tall (above 70 inches in height), and 12 who are short (under 66 inches in height). At the time they were recruited, you randomly selected half of the tall swimmers and half of the short swimmers (i.e., 6 each) and involved only these 12 people in a weightlifting program. The other 12 swimmers had no such involvement during the year. Exactly one year after recruiting them, you timed the 24 recruits to see how fast each could swim 100 meters. Means and variances on these swimming speeds are given below for each of four groups of 6 swimmers:

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>tall weightlifters</td>
<td>60</td>
<td>14</td>
</tr>
<tr>
<td>short weightlifters</td>
<td>59</td>
<td>13</td>
</tr>
<tr>
<td>tall nonweightlifters</td>
<td>63</td>
<td>12</td>
</tr>
<tr>
<td>short nonweightlifters</td>
<td>68</td>
<td>15</td>
</tr>
</tbody>
</table>

a. Imagine that you were to analyze the swimming data in a regression program using the two variables, HEIGHT and WEIGHTLIFT, where HEIGHT=1 for all tall swimmers and HEIGHT=−1 for all short swimmers, WEIGHTLIFT=1 for all swimmers whom you involved in a weightlifting program, and WEIGHTLIFT=−1 for all swimmers whom you did not involve in weightlifting. Describe what complete and reduced models you would use to estimate whether involvement in a weightlifting program improves (i.e., decreases) short swimmers' 100 meter swimming time more than it improves tall swimmers' 100 meter swimming time. (Hints: This question requires that you give only the forms of the two regression models [that is, that you clearly indicate what dependent and independent variables would be included in each]. No parameter estimates are needed as part of your answer. In addition to HEIGHT and WEIGHTLIFT, you may wish to consider including a higher-order variable in the models. Be sure to explain how this higher-order variable was constructed.)

b. One of the models described in part a contains a higher-order variable. What would you expect to be the sign of the slope associated with the higher-order variable if involvement in a weightlifting program improves short
swimmers' 100 meter swimming time more than it improves tall swimmers' 100 meter swimming time? (Be sure to explain your answer!)

c. When two variables are uncorrelated (i.e., when they have a correlation between them equal to zero), they are sometimes referred to as being orthogonal. Demonstrate that HEIGHT and WEIGHTLIFT are orthogonal by calculating the correlation between them. (Show your work!)

18. In the mid-19th Century Karl Marx argued that societies were evolving through a capitalist period toward a utopian socialist world. His argument was based on three assumptions about how and why people behave the way they do. First, people's actions are always goal-oriented, where their most basic goal is that of survival. Second, people always pursue their goals creatively, which is to say that people ongoingly strive to attain their goals by ever-more-efficient means. An immediate consequence of these first two assumptions is that societies' technologies improve over time as workers invent ever-more-efficient means of producing goods and services. Marx's third assumption was that capitalist societies are characterized by a parasitic bourgeois class comprised of people who use workers (i.e., who use anyone who does not belong to the bourgeois class) as the means to their own goals. Because these bourgeoisie control the technology (a.k.a. the means of production) on which workers' survival is based, they are able to manipulate workers' interests: If workers are productive, their wages will allow them to survive; if they are not productive, their consequent loss-of-livelihood will leave them dead. Workers develop a false consciousness insofar as they believe that the interests of the ruling class (in having workers produce goods and services for it) are in their own interests (in retaining their livelihoods). Marx believed that at some point each society's workers would come to realize how exploited they were, would take control of the society's technology from the bourgeoisie, and would establish a socialist state in which all shared jointly in control of the society's productive power.

Contemporary social scientists commonly dismiss Marx's ideas, in part because of the recent collapse of the Soviet Union's socialist experiment, but also in part because citizens of 20th Century capitalist societies are more readily seen as members of a vast middle class, rather than as members of either a bourgeois or a working class. Your research is an attempt to see what elements of Marxist theory might still be useful for understanding a capitalist society such as the United States. Your interpretation of Marx (i.e., your theory) is that workers only become aware of their exploitation if their natural creative tendencies are stifled (i.e., if they are not free to invent ever-more-efficient means of producing goods and services). This stifling of creativity results when employers so closely micromanage workers' goals that workers have almost no discretion over how their work is done. Thus, your theoretical argument is for the following 3-part (or 2-step) causal chain: First, the employer's micromanagement over what workers do limits workers' creativity in deciding how they work. Second, the more
workers' creativity has been stifled in this way, the more exploited they will believe themselves to be.

Your data are from a random sample of 32 mid-level employees of U.S. corporations. During face-to-face interviews with the employees you obtain their responses to the following 3 inquiries:

MICROMANAGE (M) "On a scale from 0=no influence to 10=total influence, indicate how much influence you have in determining what projects you spend your time doing while at work."

CREATIVITY (C) "On a scale from 0=no influence to 10=total influence, indicate how much influence you have in determining how you spend your time while doing projects at work."

EXPLOIT (E) "Considering for the moment those people at your workplace who earn more than you do, indicate the extent to which you believe that they are more or less deserving than you on a scale from 0=they deserve to earn more than I do to 10=I should be earning more than they do."

Correlations, means, and standard deviations among these variables are as follows:

<table>
<thead>
<tr>
<th></th>
<th>EXPLOIT</th>
<th>MICROMANAGE</th>
<th>CREATIVITY</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPLOIT</td>
<td>1.000</td>
<td>-.221</td>
<td>-.340</td>
<td>5.70</td>
<td>1.71</td>
</tr>
<tr>
<td>MICROMANAGE</td>
<td>-.221</td>
<td>1.000</td>
<td>.650</td>
<td>4.63</td>
<td>2.78</td>
</tr>
<tr>
<td>CREATIVITY</td>
<td>-.340</td>
<td>.650</td>
<td>1.000</td>
<td>6.28</td>
<td>3.15</td>
</tr>
</tbody>
</table>

(Notes: Please be sure that you understand these data. For example, the zero-order correlation between EXPLOIT and MICROMANAGE is -.221. The mean score on EXPLOIT is 5.7 points; the standard deviation on EXPLOIT is 1.71 points.)

a. Find the \(P\)-value associated with the hypothesis that the more control employers have over what workers do (MICROMANAGE), the less creativity workers have in deciding how they work (CREATIVITY). (Be sure to note that high scores on MICROMANAGE indicate low levels of control by employers.)

b. Compute the unstandardized partial slope associated with CREATIVITY from the regression of EXPLOIT on CREATIVITY and MICROMANAGE.

c. Using words that a lay person could understand, express the meaning of the partial slope found in part b.
d. Is the partial slope found in part b significantly different from zero at the .05 level of significance?

e. Of the variance in workers’ exploitation-beliefs (EXPLOIT) that is not explained by workers' creativity in deciding how they work (CREATIVITY), how much of this residual variance is explained by employers' control over what workers do (MICROMANAGE)?

f. Do the above findings provide statistically significant evidence (at \( \alpha = .05 \)) in support of your theory that workers' exploitation beliefs are determined in a 2-step process from employer micromanagement (M) to workers' lack of creativity (C) and from workers' lack of creativity (C) to workers' exploitation beliefs (E)? Explain your answer by referring to some or all of your answers in parts a through e. (Hint: You may wish to perform a significance test in addition to ones previously done.)

g. After examining a plot of the relation between EXPLOIT and CREATIVITY, you note that exploitation beliefs remain low (i.e., near "they deserve more than I do") until workers' creativity has been severely restricted (i.e., near "no influence"). At that point exploitation beliefs increase dramatically (i.e., toward "I should be earning more"). Give a correctly specified regression equation that would model this pattern. In doing this, specify how any variables other than EXPLOIT, CREATIVITY, or MICROMANAGE would be constructed, and indicate what the sign(s) of the slopes associated with all of the equation’s independent variable(s) would be if your data were consistent with the pattern.

19. Ecotourism refers to tourism that is socially responsible. For example, if fishing is one's mode of tourism, returning oversized fish to the water (so that these mature fish can produce offspring) is a responsible means of helping maintain the ecology of the national parks in which one fishes. Park administrators have found that "education" is required to transform tourists within U.S. national parks into ecotourists.

As part of their plan to educate tourists at Grand Teton National Park, administrators there have put together the "Think Of Others" (TOO) program to motivate fishers to catch no more than 6 nonoversized (i.e., under 12 inches) fish per day. The program involves large "Think of the others: 6 is the limit" billboards (visible upon entering the park) that depict a fisherman returning an oversized fish to the water. In addition, at each entry gate every tourist to the park is handed a flier that discusses fishers' duty to keep sufficient numbers of fish in the park's many lakes and rivers.

You have been hired to evaluate the effectiveness of the TOO program in motivating ecotourism among those who fish in the park. The data for your
evaluation are from records maintained in the park’s administrative offices from
spot-checks made by park rangers of people found fishing in the park. You
consider those who were spot-checked to be representative of all persons fishing
in the park at that time.

Your data are from all 216 spot-checked people who were found during the past 3
years (1½ years before the program began through 1½ years after the program
began) to have had oversized or over-limit fish in their possession. Your
variables are as follows:

DAY (D) The day when the spot-check was made on a scale from 1=first
day of the first year to 1095=last day of the third year (Note:
3x365=1095)

FISH (F) The number of oversized and/or over-limit fish in the spot-
checked person's possession

VISITORS (V) The number of tourists visiting the park on the day when the spot-
check was made

Correlations, means, and standard deviations among these variables are as
follows:

<table>
<thead>
<tr>
<th></th>
<th>DAY</th>
<th>FISH</th>
<th>VISITORS</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAY</td>
<td>1.000</td>
<td>-.318</td>
<td>-.070</td>
<td>540</td>
<td>273</td>
</tr>
<tr>
<td>FISH</td>
<td>-.318</td>
<td>1.000</td>
<td>.850</td>
<td>3.29</td>
<td>1.82</td>
</tr>
<tr>
<td>VISITORS</td>
<td>-.070</td>
<td>.850</td>
<td>1.000</td>
<td>2283</td>
<td>1547</td>
</tr>
</tbody>
</table>

(Notes: Please be sure that you understand these data. For example, the zero-
order correlation between DAY and FISH is –.318. The mean value of DAY is 540
days; the standard deviation of DAY is 273 days.)

a. Your study calls for a measure of "responsible fishing behavior" that does
not vary according to weather conditions. However, the large correlation of
.85 between FISH and VISITORS leads you to suspect that irresponsible
fishers are more likely to be in the park when the park has many visitors,
presumably due to nice weather. (Your suspicion: Responsible fishers are
more likely than less responsible fishers to tolerate personal discomforts such
as bad weather, fishing limits, etc.) How would you adjust FISH such that all
its covariation with VISITORS has been removed? In answering this question,
give an SPSS compute statement that would produce a new variable (call it
"R" for Responsible fishing behavior), that measures this adjusted FISH
measure. (Hint: Your SPSS compute statement should contain a number that
you calculate using the above data.)
PLEASE NOTE: All remaining parts of this problem make use of the adjusted variable, R (for Responsible fishing behavior) obtained in part a, rather than FISH. For use in later calculations, please note that its mean and standard deviation are as follows:

\[
\bar{R} = 3.0 \quad \text{and} \quad \hat{\sigma}_R = 1.7
\]

Parts b, c, d, e, and f all refer to the following regression models in which R (as calculated in part a) is regressed on various combinations of linear, quadratic, cubic, quadruplicate, and quintic measures of D (or DAY):

Model 1: \( \hat{R} = \hat{a} + \hat{b}_1 D \)

Model 2: \( \hat{R} = \hat{a} + \hat{b}_1 D + \hat{b}_2 (D - \bar{D})^2 \)

Model 3: \( \hat{R} = \hat{a} + \hat{b}_1 D + \hat{b}_2 (D - \bar{D})^2 + \hat{b}_3 (D - \bar{D})^3 \)

Model 4: \( \hat{R} = \hat{a} + \hat{b}_1 D + \hat{b}_2 (D - \bar{D})^2 + \hat{b}_3 (D - \bar{D})^3 + \hat{b}_4 (D - \bar{D})^4 \)

Model 5: \( \hat{R} = \hat{a} + \hat{b}_1 D + \hat{b}_2 (D - \bar{D})^2 + \hat{b}_3 (D - \bar{D})^3 + \hat{b}_4 (D - \bar{D})^4 + \hat{b}_5 (D - \bar{D})^5 \)

b. According to Conversion Theory, education is permanent. If you do not tell people their responsibilities, they will not comply; once you do tell them, they will comply. According to this theory, you would expect irresponsible fishing behavior (as evidenced by high scores on the adjusted FISH variable) to be relatively extensive prior to the start of the TOO program, that this behavior would drop immediately once the program had begun, and that responsible behavior would persist inevitably after this. In the space provided below, sketch the relation between R and DAY suggested by Conversion Theory. Which of the above models would be used as the complete model in evaluating this theory?

c. In contrast, Momentum Theory (an alternative to Conversion Theory) suggests that people are slow to change their (e.g., irresponsible) habits. Once a program has begun, it will take many years of continued implementation of the program for irresponsible behaviors to (linearly) subside. In the space provided below, sketch the relation between R and DAY suggested by Momentum Theory. Which of the 5 models would be used as the complete model in evaluating this theory? (Hint: Do not forget that your model should fit data obtained in the 1½ years prior to the program's
implementation as well as in the later years when irresponsible behaviors are theorized to subside.)

d. Finally, according to Rebound Theory (another alternative) programs like TOO only have temporary consequences. At first, it takes a few months before people adopt the responsible behavior espoused by the program. Yet after only a few months of responsible behavior they will have "got used to" the bulletin boards and fliers. After a few more months they will have returned to their previous (for some, irresponsible) habits. In the space provided below, sketch the relation between R and DAY suggested by Rebound Theory. Which of the 5 models would be used as the complete model in evaluating this theory?

e. Based on analyses she ran while you were out to lunch, a colleague reports to you that neither Model 4 nor Model 5 explains significantly more variance than does Model 3. Respective R-squared values for the first 3 models are as follows:

   Model 1: $R^2 = .124$  
   Model 2: $R^2 = .148$  
   Model 3: $R^2 = .152$

Which of these 3 models is most parsimonious? (Use the .05 significance level, and show your work!)

f. Instead of writing $(D - \bar{D})^2$ each time to refer to the quadratic measure in Model 2 through Model 5, let's call this measure, Q. It turns out that Q is slightly collinear with D (i.e., the DAY variable), such that $r_{DQ} = .18$. If the true value of the first partial slope in Model 2 were actually $b_1 = -.001$, find the power in a test of the following hypotheses at $\alpha = .05$:

$$H_0: b_1 = 0$$
$$H_A: b_1 < 0$$

g. You modify the DAY variable by collapsing it into four 9-month time periods using the following recode statement in SPSS:

recode DAY(1 thru 274=1)(275 thru 548=2)(549 thru 822=3)(823 thru 1095=4).

Thus, for example, DAY=1 now means that your data are from the first 9 months of your study period. Since the 549th day of your study period was the first day that the TOO program was implemented, DAY=3 now means that your data are from the first 9 months in which the program was implemented. It turns out that the same number of persons in your sample were spot-checked in each of these 9-month periods (i.e., 54 persons in each), and that the average adjusted FISH scores for the persons within each of these periods were 3.8 for DAY=1, 4.2 for DAY=2, 2.8 for DAY=3, and 1.2 for DAY=4. Using this collapsed DAY variable as your independent variable, construct an
20. With all its benefits, the Internet is open to considerable abuse by crooks and pranksters. Given that any pair of the Internet’s millions of locations (or nodes) can be connected via hundreds of channels, locating the sources of abuse is a difficult challenge. Sophisticated abusers of the Internet can easily alter identifying characters within the "packets" of information that are transferred between nodes. A preliminary step in identifying the abusers is tracing the nodes through which packets pass on their way from one Internet location to another.

Unlike personal computers that usually have only a single connection to the Internet, "routers" are nodes that typically have many connections to other nodes (i.e., to computers and other routers). Each router runs "routing software" that ensures an efficient flow of information-packets between itself and these computers and routers. Although no routing software will send packets to nodes or through channels that are out of action, these programs do differ in how, from among open channels and nodes, they select which channel-node is the one through which a packet is to be transferred.

As depicted in Figure 1, you set up a small network of 16 routers (labeled A to P) into a 4x4 grid such that they are connected by 42 channels. You next simulate a constant state of data-transfers into this network by sending the same long series of packets simultaneously in both directions in all 42 channels. (For example, router F would send the series simultaneously to routers A, B, C, E, G, I, J, and K.) Then during this constant state, you send another series of packets from...
router A to router P and note how many nodes (i.e., routers) the last of these packets passes through before it reaches router P. The smallest this number of nodes can be is 2 (i.e., from A through F through K to P), and the largest it can be is infinite (i.e., the packet could circle among other nodes, never reaching P). Yet routing software makes it unlikely that a packet gets lost in this way.

As developers of routing software point out, their software ensures that the greater the number packets in a series, the fewer nodes through which the series' last packet will pass. When one router sends a packet to another router, the second router acknowledges to the first that the packet was received. With this information, the software on the first router "learns" the shortest path between itself and other routers. For example, after a few acknowledgments router A might learn that packets sent through router F were received at router P sooner than ones sent through router B. In time the network's software would learn that the shortest path from A to P is through F and K, and later packets in each series would be increasingly more likely to be sent along this path. Yet routers only send increasing numbers of a series' packets along the shortest path when the series is not too large (i.e., not larger than around 500 packets in size). To prevent a bottleneck (e.g., at router F) from occurring, routing software is unlikely to send all packets in a large series through the same node. Thus, the more packets in a large series, the more likely routing software will send its packets along next-shortest paths, or even third- or fourth-shortest paths. The point at which this happens differs from one type of routing software to another. For example, MRT (routing software developed at the University of Michigan) is more likely to send smaller series of packets through next-shortest paths than ZebOS (routing software developed by IP Infusion(tm)).

There are 100 series of packets that you send from router A to router P during the constant state described on the previous page. These 100 "sendings" differ according to the following variables:

- ROUTSOFT (R) the type of routing software used in all of the network's 16 routers (1 = MRT; 0 = ZebOS)
- SIZE (S) the number of packets in the series
- NUMBER (N) the number of nodes (i.e., routers) that the last packet passes through on its path from router A to router P

Correlations, means, and standard deviations among these variables are as follows:

<table>
<thead>
<tr>
<th></th>
<th>ROUTSOFT</th>
<th>SIZE</th>
<th>NUMBER</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROUTSOFT</td>
<td>1.000</td>
<td>.000</td>
<td>.460</td>
<td>0.5</td>
<td>0.50</td>
</tr>
<tr>
<td>SIZE</td>
<td>.000</td>
<td>1.000</td>
<td>.330</td>
<td>500</td>
<td>150</td>
</tr>
<tr>
<td>NUMBER</td>
<td>.460</td>
<td>.330</td>
<td>1.000</td>
<td>4.4</td>
<td>0.93</td>
</tr>
</tbody>
</table>

34J
a. What is your interpretation of the fact that 0.5 is the mean of ROUTSOFT?

b. Given the above (rather lengthy) explanation of this study, give a likely reason why the correlation between ROUTSOFT and SIZE equals zero?

c. What proportion of the variance in NUMBER is explained by ROUTSOFT and SIZE? Is this proportion significantly large at the .05 significance level?

d. Of the variance in NUMBER not explained by ROUTSOFT, what proportion of this variance is explained by SIZE?

e. In this study the number of nodes through which the last packet in a series passes (as measured by NUMBER) is influenced by three characteristics of the routing software in use:

   - **First characteristic:** When a series contains relatively few packets (i.e., when SIZE is small), routing software will have less opportunity to "learn" (and will thus send fewer of a series' packets along) the shortest path. More learning is possible when a series has more packets than this.
   - **Second characteristic:** When the series' number of packets (i.e., SIZE) approaches and surpasses 500, routing software will increasingly "avoid" bottlenecks by sending larger series' packets along paths other than the shortest ones.
   - **Third characteristic:** Overall ZebOS has a greater tendency than MRT to send series of packets through shortest paths.

For each of these characteristics, please indicate if the characteristic would result in a positive or a negative linear relation between NUMBER and either SIZE or ROUTSOFT. That is, in each of the following rows circle one sign in the second column and one variable in the third column:

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sign of linear relation between NUMBER</th>
<th>And which variable?</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>+ or –</td>
<td>SIZE or ROUTSOFT</td>
</tr>
<tr>
<td>Second</td>
<td>+ or –</td>
<td>SIZE or ROUTSOFT</td>
</tr>
<tr>
<td>Third</td>
<td>+ or –</td>
<td>SIZE or ROUTSOFT</td>
</tr>
</tbody>
</table>

f. Draw a sketch below that illustrates the relations among NUMBER, SIZE, and ROUTSOFT listed in part e. Be sure to place labels for all three variables on the axes or elsewhere in the sketch.
g. What regression equation would you use to model the pattern depicted in the sketch in part f? Be sure to explain how any variable in this model other than NUMBER, SIZE, and ROUTSOFT were computed.

h. Referring to the regression equation in part g, indicate the sign that each slope would take if you were to find evidence for all of the three routing software characteristics listed in part e. Explain your answers!

21. The phrase, "digital divide," is usually used to distinguish people who live in prosperous Western societies from those living in the developing world. Whereas Westerners generally have access to computers, the Internet, cell telephones, and the like, people in many other societies have few of these things. As a consequence, those in the former group tend to have greater productivity, and thus greater wealth, than those in the latter group. And so "the rich get richer" and the poor remain poor.

Yet there is a second digital divide—a divide between people within Western societies who are physically disabled versus those who are not. Whereas non-physically disabled people can readily make use of widely-marketed digital devices, physically disabled people require special equipment that is tailored to their disabilities. For example, visually disabled people (VDPs) are unable to "read" the contents of a standard computer screen, although they work quite effectively with standard keyboards, mice, disk drives, etc.

You are an employee in the Evaluation Department at Enhance, a company that is currently developing a "screen reader" apparatus that enables people to "read" computer screens with their feet. It is not commonly known that second only to one's fingers, the soles of one's feet have the largest and broadest concentration of nerve endings on the surface of the human body. After being plugged into both a power source and a computer's video port, Enhance's "Podiatric Reader" only requires the user to place her or his bare feet into two slipper-like cuffs.
Beyond this, the user must learn to "recognize" both images and text based on feet-sensations. *Enhance* has begun offering 8-week courses to train VDPs in using the Podiatric Reader.

*Enhance*'s marketing strategy appeals to the Podiatric Reader's capacity to "close the digital divide for the visually disabled." You have been asked to test whether or not there is empirical evidence in support of this appeal. In collaboration with Chicago's Unemployment Office, you obtain a list of all VDPs who are registered there as actively seeking work. From this population of VDPs you randomly sample three groups of 6 VDPs each (i.e., your total sample size is 18). One group is given *Enhance*'s 8-week course on using the Podiatric Reader. The second group is given a 8-week computer-literacy course. Nothing is done with the third group of 6 VDPs. Both courses are run concurrently, and at the end of the 8 weeks all 18 VDPs are asked, "On a scale from 1= 'no confidence' to 10= 'complete confidence', how confident are you that you will be hired within the next year?" Your data are as follows:

Table 1. Visually disabled persons' confidence scores after a Podiatric Reader course, a computer-literacy course, or no course.

<table>
<thead>
<tr>
<th>Podiatric Reader</th>
<th>Computer-Literacy</th>
<th>No Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

NOTE: \[ \sum_{ij} (Y_{ij} - Y_{..})^2 = 134 \]

a. Using the data in Table 1, construct an ANOVA table that lists treatment, error, and total sums of squares, as well as their degrees of freedom, mean squares, and an F-statistic.

b. Is there statistically significant evidence (at the .05 level) that the VDPs in one group are more confident than those in either of the other groups? Explain your answer.

c. Analyze the confidence data by performing a regression analysis with dummy variables. Be sure to indicate for each dummy variable what each of its values means. Also obtain numerical estimates of all coefficients (i.e., slopes and constant) in your regression model.

d. State in words the meaning of each coefficient found in part c.

In your research's second phase, you arrange for the Chicago Unemployment Office to telephone each of your 18 VDPs. During the phone call they are informed of a job opening at a local computer company, and they are encouraged to apply for the job. Only 14 of the VDPs actually follow through and interview for
the job. Thus at this point your sample is 14 (not 18) VDPs. After each interview, you ask the person who interviewed the VDP to evaluate the just-interviewed VDP's "likelihood of being hired, on a scale from 1='no way' to 100='a sure thing'." You proceed with your analysis using the following three variables:

CONFIDENT (C)  "On a scale from 1='no confidence' to 10='complete confidence', how confident are you that you will be hired within the next year?" (Answered by VDP.)

LIKELYHIRE (L)  "likelihood of being hired, on a scale from 1='no way' to 100='a sure thing'" (Evaluated by computer company's interviewer.)

PRCOURSE (P)  1='took Podiatric Reader course' OR 0='otherwise'

You regress L on C and P, and the following is part of your SPSS-PC output:

<table>
<thead>
<tr>
<th>Coefficientsa</th>
<th>Unstandardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>B</td>
</tr>
<tr>
<td>1 (Constant)</td>
<td>-23</td>
</tr>
<tr>
<td>C</td>
<td>9</td>
</tr>
<tr>
<td>P</td>
<td>3</td>
</tr>
</tbody>
</table>

a. Dependent Variable: L

e. Assuming that the correlation between C and P is positive, express the meaning of the partial slope between L and C in words.

The following is a plot of your data with "likelihood of hiring" (i.e., L) on the vertical axis and "confidence" (i.e., C) on the horizontal axis:
f. What regression equation would you use to model the pattern depicted in the above plot? Be sure to explain how any variable in this model other than L and C were computed.

g. Referring to the regression equation in part f, indicate the sign that each slope would take if they were to correspond to the pattern depicted in the above plot.

22. What do cone snails have to do with horse races in Iowa? Cone snails produce a venom that paralyzes their prey. Yet when injected into larger animals this same venom acts as a pain killer. For example, when cone snail venom is injected into a race horse before a race, the horse runs faster during the race because it does not feel the muscle pain that it would normally experience.

“Horse doping” (i.e., the injection of horses with cone snail venom immediately prior to races) is an increasingly popular method of “fixing” horse races in Iowa. Your research group (from the Toxicology Department at Iowa State University) has secured a grant from the Iowa Racing and Gaming Commission to develop a method for detecting cone snail venom in horse blood. This method will be used to evaluate whether winners in Iowa horse races will have won because they were injected with the venom. In developing the method, your immediate research objective involves testing horses’ blood for traces of cone snail venom.

During the past month your research group assembled 55 vials of blood—each from a horse that was injected with between zero and 10 micrograms (mcg) of cone snail venom. (You should consider the 55 horses from which these corresponding vials were obtained to comprise a sample drawn at random from among all horses.) You then analyzed the blood from each vial in two stages:

First, “high performance liquid chromatography” (HPLC) was used to isolate cone snail venom molecules within each vial of blood. In doing this, you obtained data on how many parts per million (ppm) of the venom were detected in each vial of sampled blood.

Second, a “mass spectrometer” was used to estimate the masses of each vial’s detected molecules. (Note: The dalton (Da) is the unit used in reference to molecular masses, and, moreover, the molecular mass of cone snail venom is known to be 2639.13 Da.)

In sum, you have data on the following three variables:

DOSE (D) Micrograms of cone snail venom injected into the horse before obtaining its blood for your study (Values on this variable range from 0="no venom injected" to 10="10 mcg of venom injected.")
AMOUNT (A)  Parts per million of cone snail venom detected in a vial’s blood
(Values on this variable range from 1=“one part per million” to
20=“20 ppm.” Note: “A substance believed to be venom” was
detected in each vial.)

MASS (M)  Molecular mass (in daltons) of the substance detected using
HPLC (Values on this variable range from 2630 Da to 2650 Da.)

Means, variances, and correlations among these variables are as follows:

<table>
<thead>
<tr>
<th>Variables</th>
<th>D</th>
<th>A</th>
<th>M</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>1.00</td>
<td>.95</td>
<td>.00</td>
<td>5.00</td>
<td>10.19</td>
</tr>
<tr>
<td>A</td>
<td>.95</td>
<td>1.00</td>
<td>.10</td>
<td>10.00</td>
<td>25.00</td>
</tr>
<tr>
<td>M</td>
<td>.00</td>
<td>.10</td>
<td>1.00</td>
<td>2639.10</td>
<td>12.25</td>
</tr>
</tbody>
</table>

Note: Be sure that you read this table correctly. For example, the mean
dosage of venom injected into a horse was 5 mcg, the variance among
dosages is 10.19 mcg², and .95 is the correlation between DOSE and AMOUNT
(i.e., \( r_{DA} = .95 \)).

a. Four true/false questions regarding the regression model,
\( \hat{A} = \hat{\alpha} + \hat{b}_1 D + \hat{b}_2 M \):

1. \( R^2_{A,DM} = .95^2 + .10^2 \)  True  False
2. \( \hat{b}_1 \) is a marginal effect.  True  False
3. There is no multicollinearity in this regression model.  True  False
4. The model is misspecified.  True  False

b. Surely it is easier to detect substances with larger rather than smaller
masses. Thus one would expect that when a detected mass’s molecular
weight is large (as measured by MASS), that a larger amount of it (as
measured by AMOUNT) would be detected. Of the variance in “venom
detected” (i.e., in AMOUNT) not explained by “venom injected” (i.e., by DOSE),
what proportion of this unexplained variance is explained by the detected
substance’s molecular weight (i.e., by MASS)?

c. Find a 95% confidence interval for the partial correlation between AMOUNT
and MASS, controlling for DOSE.
d. Clearly the amount of venom detected (or measured) in a horse’s blood will be larger if more venom is injected into it. Yet this association is of less importance to you than understanding how far your measurements deviate from the known amounts of venom that were injected into the 55 horses. To obtain such a deviation measure, you adjust the amount of detected venom (A) for the dosage of venom injected into the horses (D). What formula would you use to adjust AMOUNT in this way? (Hint: The requested formula contains a number. In answering this question, be sure to calculate the number and show how it was obtained.)

e. The adjusted variable obtained in part d (let’s call it $A_{adj}$) can be used in computing a measure of the parts per million, or ppm, that each vial’s “amount of venom detected” (i.e., its AMOUNT) differs from what one would estimate given the linear effect of the known amount of venom (i.e., DOSE) that was injected into the horse from which the vial’s blood was obtained. This new variable (let’s call it NEWVAR) takes values ranging from 0=“AMOUNT equals exactly what one would estimate given DOSE” to 10=“AMOUNT differs 10 ppm from what one would estimate given DOSE.” Thus, NEWVAR measures how well or poorly HPLC has isolated cone snail venom within each vial of blood. (Note that the larger the value of NEWVAR, the poorer the HPLC isolation.) For an extra credit point, explain how to compute NEWVAR.

f. You hypothesize that the closer the detected substance’s mass (i.e., MASS) is to the molecular mass of cone snail venom (namely 2639.13 Da), the closer that AMOUNT is to “what one would estimate given DOSE.” In the space below sketch this relation between NEWVAR and MASS.
g. What complete and reduced models would you estimate in testing the hypothesized relation sketched in part f?

h. The complete model listed in part g contains a higher-order variable. According to the hypothesis in part f, what would be the sign of slope (within the complete model) that is associated with this higher-order variable? Explain your answer.

23. Although you may not think of yourself as a patriot, just travel abroad and you will find yourself energetically arguing against the misconceptions that citizens of other countries have about your country. While abroad, you may also find yourself longing for something unavailable in your new surroundings—be it your favorite foods, a familiar climate, or social recognition. In the US, for example, Indian immigrants of Hindu origins commonly miss the comfortable guidance that others’ expectations provide. Here there is no social hierarchy from which those of greater status can gently suggest what lower-status persons ought to do when difficult life decisions are to be made.

Possibly the most important life decision in every Indian’s life is the selection of a marriage partner. Traditionally this decision is made, or arranged, by the partners’ parents. Yet to most Indian immigrants in the US, parental guidance is only available from occasional phone conversations, and spouses are often sought with the assistance of “matrimonial websites” that cater to typically Indian concerns for finding a “compatible” mate. If one’s spouse is of the wrong caste or with flawed character or in astrological disarray, a lifetime of misery may ensue. No wonder that single Indian immigrants develop “marriage anxiety” when separated from the wisdom of astrologers and parents in their homeland.

As a research project for your master’s thesis, you collect data from face-to-face interviews with two groups of randomly sampled Indians between the ages of 20 and 35. Each group is of 15 unwed (i.e., not married) Indians, yielding a total sample size of 30 single respondents. The first group is of Indian immigrants within a small Midwest university town; the second group is of Indian residents of Mumbai, India, who seek marital guidance from Mukund N. Trivedi—the city’s most famous Vedic astrologer. Although all respondents are single, each indicates a desire to marry at some point. During your interviews you collect data on the following variables:

MARRIAGE ANXIETY (A) respondents’ agreement with the statement, “I am unsure that my future mate and I will be compatible,” on a 5-point scale from 1=strongly disagree to 5=strongly agree.
FAMILY TIES (T) respondents’ agreement with the statement, “I would only consider marrying someone of whom my parents approve,” on a scale from 1=strongly disagree to 5=strongly agree.

RESIDENCE (R) 1=immigrant to the US; 0=resident of Mumbai

You should consider each of these variables to be an interval-level measure.

Your theory is that not only is Indians’ marriage anxiety greater among US immigrants than among Mumbai residents, the tendency for family ties to increase an Indian’s marriage anxiety is stronger for the former than the latter group.

a. In the space provided, sketch a plot of the above-theorized relations among MARRIAGE ANXIETY, FAMILY TIES, and RESIDENCE. (Be sure to place variable names and high and low values on each axis, and to indicate where your theory suggests data points would occur.)

b. What reduced and complete models would you use to evaluate your theory? Note that one of the variables in the complete model will be a higher-order measure that you must calculate. You may call this variable, MYVAR. Be sure to indicate how you would calculate this variable. (No actual calculations—no numbers—are being asked for here. Just indicate what two regression models you would estimate, and how you would go about calculating the higher-order measure.)
In accordance with the models described in part b, output (in part) from estimating the reduced (at left) and complete (at right) models is as follows:

Regression

<table>
<thead>
<tr>
<th>Model Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>a. Predictors: (Constant), T, R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficientsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>1 (Constant)</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>a. Dependent Variable: A</td>
</tr>
</tbody>
</table>

Regression

<table>
<thead>
<tr>
<th>Model Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>a. Predictors: (Constant), T, R, MYVAR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficientsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>1 (Constant)</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>MYVAR</td>
</tr>
<tr>
<td>a. Dependent Variable: A</td>
</tr>
</tbody>
</table>

c. Based on the above output, is the reduced or the complete model more parsimonious? (Use the .05 significance level in giving a statistical justification for your answer.)

d. Does the above output provide empirical support for your theory? (Be sure to show how you arrived at your answer.)

The following means, variances, and correlations among MARRIAGE ANXIETY, FAMILY TIES, and RESIDENCE will be of use in the last parts of this problem:

<table>
<thead>
<tr>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>R</td>
</tr>
</tbody>
</table>

Note: Be sure that you read this table correctly. For example, the mean MARRIAGE ANXIETY score is 2 points (which is a score of “disagree” but not strongly), this variable’s variance is 1.5 squared-points, and .35 is the correlation between MARRIAGE ANXIETY and FAMILY TIES (i.e., $r_{AT} = .35$).

e. Of the variance in MARRIAGE ANXIETY not explained by the linear effects of RESIDENCE, what proportion is explained by FAMILY TIES?
f. Find a 95% confidence interval for the first-order partial correlation between MARRIAGE ANXIETY and FAMILY TIES, controlling for RESIDENCE.

g. When you wrote up your research proposal, your hypothesis was that the partial correlation obtained in part f would be positive (i.e., $H_A: \rho_{AT,R} > 0$). Given this, what is the $P$-value associated with the value of the first-order partial correlation that you obtained (i.e., of $r_{AT,R}$ from part f)? (Hint: The $P$-value will be the same as the one associated with $T(r_{AT,R})$.)

h. If the truth were that $\rho_{AT,R} = 0.60$, what would be the power of the following hypothesis test:

$H_0: \rho_{AT,R} = 0$
$H_A: \rho_{AT,R} > 0$

(Hints: Assume that the test is being conducted at the .05 significance level. Also note that the power will be the same as for $H_0: T(\rho_{AT,R}) = 0$ versus $H_A: T(\rho_{AT,R}) > 0$.)

24. In the late 1990s Iowa State University’s College of Agriculture and Life Sciences (CALS) instituted into its undergraduate program 3 integrated, experiential learning programs (IELPs). Gradually, IELP-associated courses were developed to help students gain skill in utilizing the abstract knowledge they were learning in the classroom toward solving agriculture-related problems. After all, it makes no sense to waste students’ time teaching them about biological and zoological mechanisms that they will have little or no need to know after their graduation. The IELPs were thus intended to increase the relevance of agricultural education within CALS.

This year the dean of CALS asked you to evaluate these programs’ success (or failure). Your initial step was to send out e-mail requests to all 750 students who had both participated in one of the programs and received a BS (i.e., a Bachelor’s degree) anytime between 2001 and 2009. Although each of these students was requested to fill in an online survey, your response rate was only 50%—an amount barely sufficient for making inferences to your imagined population of all potential students in the program. You decide nonetheless to proceed with the data on the 375 students who did provide responses to your online survey.

A preliminary look at these data reveals that after completing one of the IELPs, graduates typically had one of 3 career trajectories: They either (1) got a job in an agriculture-related industry, (2) they continued their education in graduate
school, or (3) they returned to work on the family farm. In addition to their graduation year and their career trajectory, you also have data on (a) the number of credit hours each student earned for having taken newly instituted IELP courses, and (b) how relevant each student perceived her/his undergraduate training to have been to their activities immediately following graduation. More specifically, the four variables in your preliminary analysis are as follows:

YEARGRAD (Y): The student’s year of graduation, ranging from 2001 to 2009

AFTERBS (A): Responses to, “What did you do after getting your Bachelor’s degree?” being one of the following: 1=“I got a job in an agriculture-related industry,” 2=“I entered graduate school,” or 3=“I returned to work on my family’s farm.”

CREDITS (C): The student’s IELP credits (i.e., the number of credits-hours earned for taking newly instituted IELP courses)

RELEVANC (R): Numeric responses to the following: “On a scale from 1 (for total irrelevance) to 10 (for total relevance), how relevant do you perceive your undergraduate training to have been to your activities immediately after getting your Bachelor’s degree?”

a. You note that the students’ earliest and latest graduation years are during years of recession in the US (namely, during either the 2001 Internet bubble or the 2009 housing bubble). You suspect that these two recessions may have influenced how relevant the students may perceive their undergraduate training to have been. In particular, you suspect that in competitive job markets—like those that arise during national recessions—one’s training may be less relevant than other factors (e.g., one’s personal contacts in industries, graduate programs, or at home on the farm). In the space provided below, sketch the relation between YEARGRAD (Y) and RELEVANC (R) that you suspect may have resulted due to differences in the US economic situation at the times the students graduated. (Hint: Be sure to indicate along the below axes which variable is independent, and which is dependent.)

[Diagram showing a scatter plot with YEARGRAD (Y) on the x-axis and RELEVANC (R) on the y-axis, indicating a trend where higher graduation years correlate with lower relevance scores.]

Dependent Variable: __________

Independent Variable: __________
b. What complete and reduced regression models would you estimate to evaluate whether or not there is statistically significant evidence for the relation sketched in part a? (Hint: Do not refer to the values of slope estimates at this point.)

c. Assuming that the 2001 and 2009 recessions influenced students’ “perceived relevance of their training” (i.e., RELEVANC) to an equal degree and assuming that in your sample the same number of students graduated from each year between 2001 and 2009, what would be the signs of the slopes in the complete regression model specified in part b?

Data in the following table will be useful in completing the remaining parts of this exam. In it please find correlations, means, and standard deviations among three of the above-mentioned variables:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Y</th>
<th>C</th>
<th>R</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>1.00</td>
<td>.15</td>
<td>-.20</td>
<td>2005</td>
<td>2.6</td>
</tr>
<tr>
<td>C</td>
<td>.15</td>
<td>1.00</td>
<td>.30</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>R</td>
<td>-.20</td>
<td>.30</td>
<td>1.00</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: Be sure that you read this table correctly. For example, the average year of graduation (YEARGRAD) is 2005, this variable’s standard deviation is 2.6 years, and .15 is the correlation between YEARGRAD and CREDITS (i.e., \( r_{YC} = .15 \)).

d. Calculate the unstandardized partial slope associated with CREDITS from the regression of RELEVANC on both YEARGRAD and CREDITS.

e. State in words the meaning of the unstandardized partial slope associated with CREDITS that you calculated in part d.

f. Is the slope calculated in part d statistically different from zero at the .05 significance level? Be sure to show how you obtain your answer.

g. What complete and reduced models would you use to test whether or not taking additional IELP credits increased the relevance of post-graduation activities only among those students who entered graduate school immediately after getting their Bachelor’s degree? (Hint: Answering this question requires the construction of two new variables. Be sure to explain how each variable was constructed, in addition to describing the regression models in which they are to be used.)

h. As it turns out, the 200 students in your sample who got a job after receiving their BS had an average RELEVANC score of 5, whereas the 100 students who entered graduate school after receiving their BS had an average
RELEVANCE score of 8 and the 75 students who returned to work on their family’s farm after receiving their BS had an average RELEVANCE score of 6. Given this, construct an ANOVA table below (with sums of squares, degrees of freedom, mean squares and F) for testing the null hypothesis that these three means are not significantly different from each other.

Summary table of the group sizes and means described above:

<table>
<thead>
<tr>
<th>Value on AFTERBS (A)</th>
<th>Group size</th>
<th>Group mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>got a job in an agriculture-related industry (A=1)</td>
<td>200</td>
<td>5</td>
</tr>
<tr>
<td>entered graduate school (A=2)</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>returned to work on family’s farm (A=3)</td>
<td>75</td>
<td>6</td>
</tr>
</tbody>
</table>

i. Using the .05 significance level, test the null hypothesis that perceived relevance of undergraduate training is the same among students who had each of the 3 career trajectories measured by AFTERBS (i.e., test the null hypothesis stated at the end of part h).

25. In western Des Moines, Clive’s Department of Community Development (CDCD) has just reviewed a new development project called “Clive Now” in which there would be 22 commercial lots, 32 residential lots, and 8 displaced families. Upon forwarding the project to the city’s Planning and Zoning Commission, CDCD proposed that the project be modified by ensuring both “walkability” and some additional office space as part of the development. Yet final approval of “Clive Now” will depend neither on what the developers want nor on the CDCD’s recommendations, but on how acceptable the project is to the residents of Clive. Often residents’ input is solicited in town meetings requested by Clive’s City Council (CCC). During town meetings residents may provide their arguments for or against any issue on which the CCC must make a decision. Because members of the CCC are concerned that Clive residents will protest the family displacement resulting from “Clive Now,” they called a town meeting that was held last Tuesday. In conjunction with this, the CCC asked you to conduct exit surveys with a random sample of Clive residents as they left that meeting. Your data consist of 44 residents’ agreement or disagreement with the following 3 statements:

OKDISPLACE (O): “A few displaced families’ inconvenience is a small price to pay for the benefits of a carefully planned urban development project.”

CAREFULPLAN (C): “I think that ‘Clive Now’ is a carefully planned development project.”
HOPEAPPROVE (H): “I hope that Clive’s City Council approves the ‘Clive Now’ project.”

Responses to each of these three statements range on a 50-point scale from 1=“strongly disagree” to 50=“strongly agree”—a scale that you may assume comprises an interval-level measure. Correlations, means, and standard deviations among these variables are as follows:

<table>
<thead>
<tr>
<th>Variables</th>
<th>O</th>
<th>C</th>
<th>H</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>1.00</td>
<td>.90</td>
<td>.80</td>
<td>32</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>.90</td>
<td>1.00</td>
<td>.72</td>
<td>27</td>
<td>11</td>
</tr>
<tr>
<td>H</td>
<td>.80</td>
<td>.72</td>
<td>1.00</td>
<td>23</td>
<td>12</td>
</tr>
</tbody>
</table>

Note: Be sure that you read this table correctly. For example, the mean OKDISPLACE score is 32 points (which is a score of “undecided” with a slight leaning toward agree), this variable’s standard deviation is 10 points, and .90 is the correlation between OKDISPLACE and CAREFULPLAN (i.e., $r_{OC} = .90$).

a. Calculate the two unstandardized partial slopes from the regression of HOPEAPPROVE on both OKDISPLACE and CAREFULPLAN.

b. State in words the meaning of the unstandardized partial slope associated with OKDISPLACE that you calculated in part a.

c. In the space provided, sketch a plot of the relations among OKDISPLACE, CAREFULPLAN, and HOPEAPPROVE. (Hints: Be sure that your sketch is consistent with the correlations provided on page 1, and with the partial slopes calculated in part a. Although your sample is much larger than 20, do not place any more than 20 data points between the below axes. Also be sure to indicate whether OKDISPLACE or CAREFULPLAN is measured along the horizontal axis.)

![Sketch of plot](image-url)
d. **Your theory** is that people think rationally. In particular, your argument is that people will only support “Clive now” (by hoping that the CCC approves it) if they agree BOTH that the inconvenience of family displacement is worth “the benefits of a carefully planned urban development project” AND that “‘Clive Now’ is a carefully planned development project.” Their agreement with only one of these positions (i.e., with only O or C) is less important in determining their support than is their agreement with both. In the space below sketch the pattern suggested by this theory. (Hint: The hints provided in part c apply here as well.)

![Diagram showing two axes, one labeled HOPEAPPROVE with values 0 to 50 and the other with values 0 to 50.]

e. To test your theory (as described in part d), you compute a new variable, and name it MYVAR (or M). Give the SPSS compute statement that you would use to calculate this variable.

f. What complete and reduced models would you use to test the theory described in part d?

g. Given that the R-squareds from the complete model and the reduced model (as described in part f) are respectively .73 and .64, determine which model is more parsimonious? (Use the .05 significance level.)

h. It turns out that the unstandardized slope associated with MYVAR (as calculated in part e) equals 3.6 (i.e., $\hat{b}_M = 3.6$ is the value of a partial slope from the complete model in part f). Given this partial slope, do you have empirical support for your theory (as described in part d)?

i. In parts g and h you are performing a one-tailed hypothesis test regarding the partial slope between MYVAR and HOPEAPPROVE (i.e., regarding $\hat{b}_M$ ) from the complete model described in part f. You note from your computer output that the one-tailed $P$-value associated with this partial slope equals .0015 (i.e., $p = .0015$). Explain in words what this $P$-value means. (Hint: Be sure to include the number, .0015, in your explanation.)