Lab 7

1. Create four data sets, each having three variables and a total sample size of ten--five units of analysis in each of two levels of a control variable. When you choose names and values for each variable, make sure they have substantive meanings. (That is, do NOT call the variables X, Y, etc., but something like INCOME, PRESTIGE, etc.) The four data sets should differ in that the correlation between each independent (X) and dependent (Y) variable should change in one of the following four specific ways when you remove the effects of a control (Z) variable in a partial correlation:

a. EXPLANATION: The correlation between X and Y is positive (r > .9). The partial correlation between X and Y controlling for Z is less than .1 in absolute value. PLUS the substantive relation among X, Y, and Z is that Z is causally prior to both X and Y.

b. INTERPRETATION: The correlation between X and Y is negative (r < -.9). The partial correlation between X and Y controlling for Z is less than .1 in absolute value. PLUS the substantive relation among X, Y, and Z is that Z is causally prior to Y, but X is causally prior to Z.

c. DISTORTION: The correlation between X and Y is positive (r > .8). The partial correlation between X and Y controlling for Z is negative (r < -.8).

d. SUPPRESSION: The correlation between X and Y is less than .1 in absolute value. The partial correlation between X and Y controlling for Z has absolute value greater than .9.

In each data set choose a control variable that is binomial (i.e., that takes only two values--for example, 1 = Yes and 2 = No). For each of the four plots, state in words what theoretical relation among the variables is implied by the pattern you have created. An applet is available (http://www.public.iastate.edu/~carlos/404/labs/applet/) to help you obtain the requested correlations and plots. If the applet does not appear on your computer, switch to a different version of the applet via the very last link at the bottom of the web page. The advantage of using the first version is that its "Print" button allows you to print only the part of the web page associated with the applet. (I.e., it saves a tree.) Before plotting data, please enter labels for the 3 variables (plus meaningful ranges of values for X and Y) in the spaces provided. Also, if the printed version is not in color, please circle and label either its blue or red dots.
The president of the Eugenio Widget Corporation believes that widget quality is likely to improve as worker motivation improves. And indeed, you find a large positive correlation between your measures of worker motivation (MOTIVATE = a 100-point scale for which high scores indicate high motivation) and widget quality (QUALITY = a 100-point scale for which high scores indicate high quality). However, when you plot quality by motivation you notice a lot more variation in widget quality when the widgets are produced by highly motivated workers than when they are produced by less highly motivated workers. (As it turns out, quality is uniformly low when the work force is not motivated, but quality ranges from high to low when produced by highly motivated workers. Evidently, workers' different abilities to produce quality widgets only become apparent once they are motivated to do quality work.) You analyze your data with the following SPSS compute and command statements:

```
compute motivat1 = arsin(sqrt(motivate))
compute motivat2 = 1/motivate
compute motivat3 = lg10(motivate)
compute motivat4 = sqrt(motivate)
compute motivat5 = motivate/quality
compute quality1 = arsin(sqrt(quality))
compute quality2 = 1/quality
compute quality3 = lg10(quality)
compute quality4 = sqrt(quality)
compute quality5 = quality/motivate
pearson corr vars = motivat1 motivat2 motivat3 motivat4 motivat5 with quality1 quality2 quality3 quality4 quality5
```

Means and standard deviations for each of the variables are as follows:

<table>
<thead>
<tr>
<th>MOTIVATE1</th>
<th>MOTIVATE2</th>
<th>MOTIVATE3</th>
<th>MOTIVATE4</th>
<th>MOTIVATE5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>.94</td>
<td>.02</td>
<td>-.30</td>
<td>.77</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>.33</td>
<td>.03</td>
<td>.41</td>
<td>.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUALITY1</th>
<th>QUALITY2</th>
<th>QUALITY3</th>
<th>QUALITY4</th>
<th>QUALITY5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>61.47</td>
<td>.02</td>
<td>1.79</td>
<td>7.82</td>
</tr>
<tr>
<td>Std. dev.</td>
<td>7.70</td>
<td>.01</td>
<td>.06</td>
<td>.50</td>
</tr>
</tbody>
</table>

The requested correlation matrix looks as follows:

```
QUALITY1  QUALITY2  QUALITY3  QUALITY4  QUALITY5
MOTIVATE1  .76      -.71     .74      .77      -.63
MOTIVATE2  -.44     .46      -.45     -.44     .75
MOTIVATE3  .67      -.68     .63      .68      -.84
MOTIVATE4  .73      -.71     .73      .72      -.70
MOTIVATE5  .56      -.53     .56      .51      -.62
```
a. Sketch a plot of QUALITY by MOTIVATE that shows data points in the pattern described in first paragraph of this problem.

b. Which variance stabilizing transformation is appropriate given the pattern of heteroscedasticity that appears in part a?

c. Using the correlation matrix on this page and the list of means and standard deviations, find the appropriate unstandardized regression equation for estimating the relation between QUALITY and MOTIVATE.

d. Using on the unstandardized regression equation found in part c, express the relation (based on a constant and a slope) between QUALITY and MOTIVATE in words.