Sociology 593
Exam 2
March 31, 1995

I. True-False. (20 points) Indicate whether the following statements are true or false. If false, briefly explain why.

1. A variable is called BLACK. This probably means that only blacks have a score on this variable.

2. The correct model is $y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$. The researcher mistakenly believes in and estimates $y = \alpha^* + \beta 1^* X1 + \varepsilon^*$. If X_1 and X_2 are uncorrelated, $E(b_1^*) = \beta_1$.

3. The null and alternative hypotheses are

H_0: $\beta_1 = \beta_2$
H_A: $\beta_1 \neq \beta_2$

The researcher computes SUMX1X2 = X1 + X2. She regresses Y on X1 and SUMX1X2. The T value for SUMX1X2 will tell her whether or not the null hypothesis should be rejected.

4. A researcher is interested in whether the effect of education on political liberalism is different for men than it is for women. Her variables are LIBERALISM (0 = very conservative, 100 = very liberal), FEMALE (1 = female, 0 = male), EDUC (years of education) and FEMEDUC (FEMALE * EDUC). When she estimates the model, she gets the following:

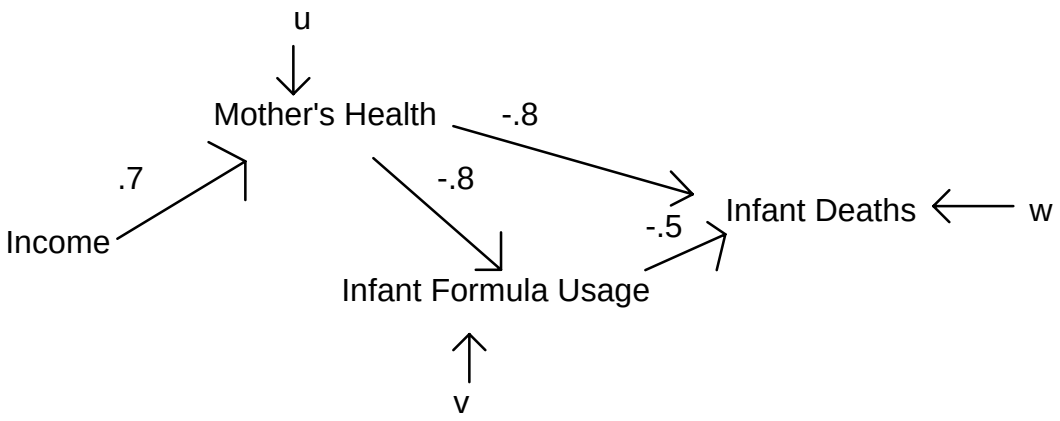
$b_{Educ} = +0.8$

$b_{Female} = 0$

$b_{Femeduc} = -0.4$

This means that education positively affects the liberalism of men but negatively affects the liberalism of women.

II. Path Analysis/Model specification. (30 points). A demographer believes that the following model describes the relationship between Income, Health of the Mother, Use of Infant formula, and Infant deaths. All variables are in standardized form. The hypothesized value of each path is included in the diagram.



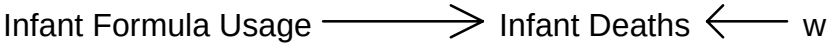
a. Write out the structural equation for each endogenous variable.

b. Determine the complete correlation matrix. (Remember, variables are standardized. You can use either normal equations or Sewell Wright, but you might want to use both as a double-check.)

c. Decompose the correlation between Infant deaths and Usage of Infant formula into

- Correlation due to direct effects
- Correlation due to common causes

d. Suppose the above model is correct, but instead the researcher believed in and estimated the following model:

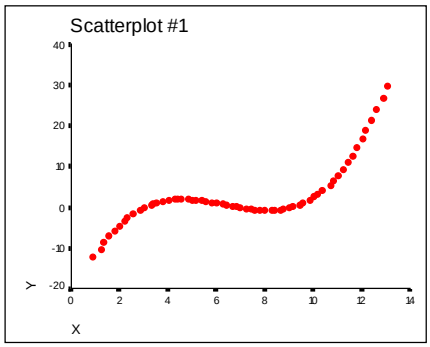


What conclusions would the researcher likely draw? Why would he make these mistakes? Discuss the consequences of this mis-specification.

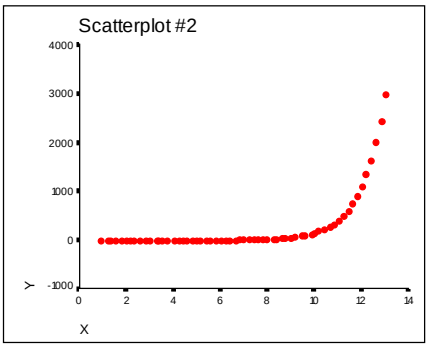
III. *Short answer.* Answer *two* of the following three questions. (25 points each; up to 10 points extra credit if you do all 3).

1. Examine the following plots. What OLS assumptions, if any (e.g., additivity, linearity), would be violated if you simply regressed Y on X? Indicate the model you think should be estimated, e.g. $y = \alpha + \beta_1 X + \beta_2 X^2$.

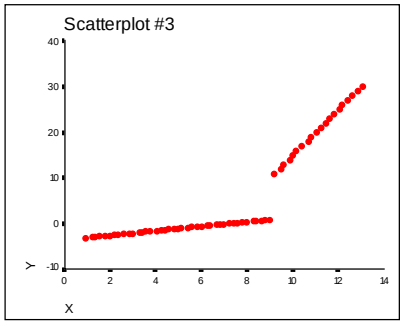
a.



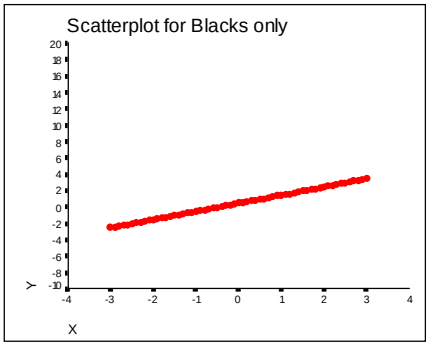
b.

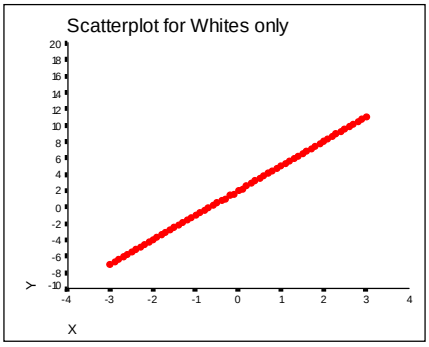


c.



d.



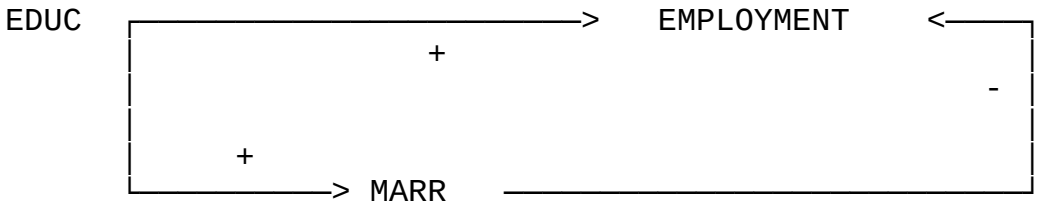


2. A researcher is interested in the relationship between a woman's education, her marital status, and her employment. For each of the following models, indicate which woman would be more likely to be employed. Your options are: Sue is more likely to be employed; Jane is more likely to be employed; Sue and Jane are equally likely to be employed; or DK, don't know, not enough information. (Note: EDUC is coded 1 = educated, 0 = uneducated; MARR is coded 1 = married, 0 = not married; EMPLOYMENT is coded 1 = employed outside the home, 0 = not employed outside the home. All three models agree that educated women are more likely to be married.)

	Sue	Jane	Fig 1	Fig 2	Fig 3
1.	Married	Not Married	a.	b.	c.
2.	Educated Married	Not educated Married	d.	e.	f.
3.	Educated Not Married	Not educated Married	g.	h.	i.

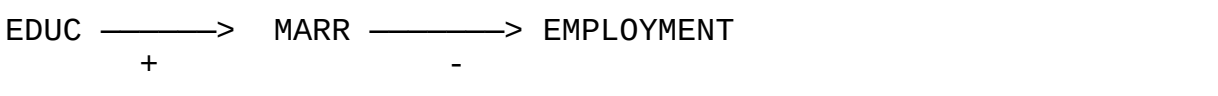
==========

Figure 1.



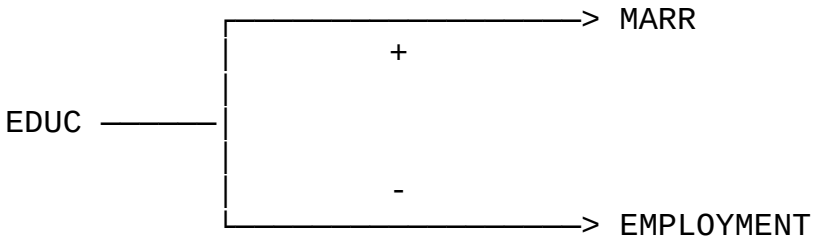
==========

Figure 2.



==========

Figure 3.



3. A researcher has constructed a scale which measures political attitudes (LIBERAL), where high values indicate political liberalism and low values indicate political conservatism. She has also constructed a scale called DEMOCRAT, where the higher the score, the more the person supports the Democratic party. She wants to see whether different models apply to men and women. She therefore runs the following regressions. Indicate whether there appear to be statistically significant differences between men and women. If so, tell whether these differences are limited to differences in the intercepts, or whether the effect of liberalism differs between men and women. Briefly discuss the substantive implications of what you think is the best model.

```
-> compute femlib = female * liberal.
->
-> REGRESSION
->   /STATISTICS COEFF R ANOVA CHA
->   /DEPENDENT democrat
->   /METHOD=ENTER liberal  /METHOD=ENTER female  /METHOD=ENTER femlib  .


* * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   DEMOCRAT

Block Number  1.  Method:  Enter      LIBERAL

Variable(s) Entered on Step Number  1..    LIBERAL

Multiple R           .49135
R Square             .24142          R Square Change    .24142
Adjusted R Square    .23990          F Change        158.49045
Standard Error      8.33429          Signif F Change    .0000

Analysis of Variance
                   DF      Sum of Squares      Mean Square
Regression          1         11008.82192      11008.82192
Residual          498         34591.31558         69.46047

F =     158.49045       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

LIBERAL        1.292002     .102627    .491346    12.589  .0000
(Constant)    31.365000     .372721               84.151  .0000


End Block Number   1   All requested variables entered.
```

[Continued]

```
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Block Number  2.  Method:  Enter      FEMALE

Variable(s) Entered on Step Number  2..    FEMALE

Multiple R           .66916
R Square             .44778          R Square Change    .20636
Adjusted R Square    .44556          F Change        185.72270
Standard Error      7.11806          Signif F Change    .0000

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           2         20418.78074      10209.39037
Residual           497         25181.35676         50.66671

F =     201.50094       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

LIBERAL        1.424367     .088187    .541684    16.152  .0000
FEMALE         8.729493     .640555    .457047    13.628  .0000
(Constant)    27.000253     .451565               59.793  .0000

End Block Number   2   All requested variables entered.


* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Block Number  3.  Method:  Enter      FEMLIB

Variable(s) Entered on Step Number  3..    FEMLIB

Multiple R           .66972
R Square             .44853          R Square Change    .00075
Adjusted R Square    .44519          F Change           .67471
Standard Error      7.12039          Signif F Change    .4118

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           3         20452.98849       6817.66283
Residual           496         25147.14901         50.69990

F =     134.47094       Signif F =  .0000

------------------ Variables in the Equation ------------------

Variable              B        SE B       Beta         T  Sig T

LIBERAL        1.373291     .107928    .522260    12.724  .0000
FEMALE         8.750184     .641260    .458130    13.645  .0000
FEMLIB          .153876     .187333    .033664      .821  .4118
(Constant)    27.020684     .452397               59.728  .0000

End Block Number   3   All requested variables entered.
```