Sociology 593
**Exam 3**
May 5, 1995

I. True-False. (30 points) Indicate whether the following statements are true or false. If false, briefly explain why.

1. A researcher has a sample of men and a sample of women. For both samples separately, she regresses income on education. The R^2 value is larger for men than it is for women. This means that the structural effect of education is larger for men than it is for women (i.e. $\beta^{Men} > \beta^{Women}$).

2. When using OLS, the "best" model is the model with the largest R^2.

3. One of the advantages of 2 stage least squares is that multicollinearity is not a problem.

4. A physician has developed a new exercise program. She believes that those who participate in the program will be happier, more physically fit, and will work better on the job than those who do not participate. Happiness, physical fitness, and job productivity are all measured on interval-level scales. Participation in the program is coded 0 or 1. Her best strategy is to simply run three different OLS regressions.

5. Ols provides unbiased estimates of recursive models.

6. A researcher wants to estimate the following model. Variables A, B, C, D have interval-level measurement. u and v are disturbance terms. She should use 2 stage least squares.
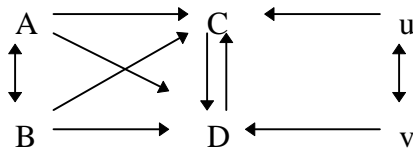
II. *Short answer.* (15 pts each; 45 pts total; up to 10 points extra credit). Answer *three* of the following (up to 5 pts. extra credit for each additional question you answer).

1. A researcher has collected data on the following variables: BLACK (1 = black, 0 = not black), FEMALE (1 = female, 0 = male) and INCOME (measured in thousands of dollars). Her dependent variable is DEMOCRAT (1 = Democrat, 0 = Republican). When she runs her logistic regression, she gets

$b_{Black} = 4.0$
$b_{Female} = 3.2$
$b_{Income} = -0.01$
$a = -3.0$

Complete the following table:

Race	Gender	Income in $1000s	Log odds	Odds	P(Democrat)
Black	Female	20	4.0		
White	Female	20			
White	Male	100			
Black	Male	100	0.0		

2. A researcher obtains the following printout from a logistic regression. Test the hypothesis that X1 and X2 have equal effects.

```
-> Compute SumX1X2 = X1 + X2.
->
-> LOGISTIC REGRESSION y
->    /METHOD=ENTER sumx1x2 x1.

Dependent Variable..    Y

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood    301.41613

* Constant is included in the model.

Beginning Block Number  1.  Method: Enter

Variable(s) Entered on Step Number
1..       SUMX1X2
          X1

-2 Log Likelihood       272.741
```
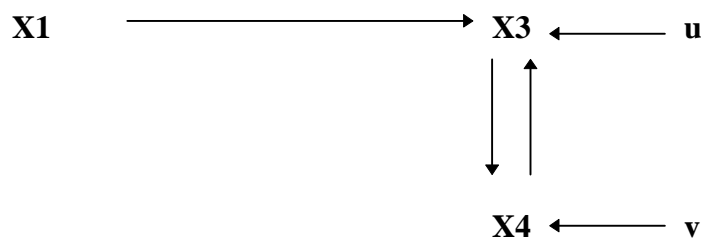
```
                  Chi-Square     df Significance

 Model Chi-Square           28.675     2          .0000
 Improvement                28.675     2          .0000

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

Variable             B       S.E.      Wald    df      Sig        R    Exp(B)

SUMX1X2          1.0062     .3339    9.0816     1     .0026    .1533   2.7352
X1                .4523     .4786     .8931     1     .3446    .0000   1.5719
Constant          .9917     .2157   21.1378     1     .0000
```
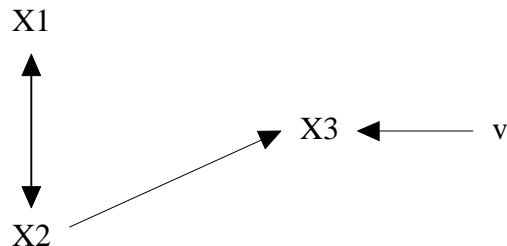
3.   Consider the following model:

Explain why you agree or disagree with the following statement: In order to make this model identified, the researcher should try to find an X2 which has a strong effect on X4 but which has zero correlation with X3.

4.   A researcher believes in the following model:

A sample of 100 cases is collected. It is found that $r_{12} = .50$, $r_{13} = .40$, $r_{23} = .50$. Test whether the overidentifying restriction imposed by this model appears to be valid.

5.   A researcher has collected data on home mortgage loan applications in St. Joseph County, Indiana. Examine the following printout. Based on this printout, are there racial differences in the determinants of having a loan application denied? If yes, do these racial differences occur in the slopes (i.e. the effects of IVs), the intercepts, or both?

```
-> * Extract from the 1990-1992 HMDA data for St. Joseph County.
-> * Applinc = Applicant Income, $1000s of dollars
-> * Localbnk = Is bank locally owned?
-> * Black = Is applicant black?
```

```
-> * Denial = Was loan application denied by bank?
->
-> get file='D:\SOC593\EXAM3.SAV'.

-> compute blkinc = black * applinc.
-> compute blklocal = black * localbnk.

-> LOGISTIC REGRESSION denial
->    /METHOD=ENTER applinc localbnk
->    /METHOD=ENTER black
->    /METHOD=ENTER blkinc blklocal
->    /CRITERIA PIN(.05) POUT(.10) ITERATE(20) .


        Total number of cases:       10980 (Unweighted)
        Number of selected cases:    10980
        Number of unselected cases:  0

        Number of selected cases:                   10980
        Number rejected because of missing data:    576
        Number of cases included in the analysis:   10404

Dependent Variable..   DENIAL

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   6766.9628

* Constant is included in the model.


Beginning Block Number  1.  Method: Enter

Variable(s) Entered on Step Number
1..       APPLINC   Applicant income
          LOCALBNK

-2 Log Likelihood     6474.333

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

Variable          B      S.E.     Wald      df     Sig       R    Exp(B)

APPLINC       -.0242     .0019 162.7114      1    .0000   -.1541    .9761
LOCALBNK      -.4293     .0710  36.5904      1    .0000   -.0715    .6510
Constant     -1.1270     .0727 239.9658      1    .0000


Beginning Block Number  2.  Method: Enter

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

-2 Log Likelihood     6445.640
```

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

Variable            B        S.E.      Wald     df      Sig        R     Exp(B)

APPLINC         -.0235      .0019  154.3192      1     .0000   -.1534    .9768
LOCALBNK        -.4282      .0711   36.2831      1     .0000   -.0728    .6516
BLACK            .7867      .1377   32.6491      1     .0000    .0688   2.1961
Constant       -1.1903      .0738  260.1834      1     .0000


Beginning Block Number   3.   Method: Enter

Variable(s) Entered on Step Number
1..        BLKINC
           BLKLOCAL

-2 Log Likelihood      6443.380

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

Variable            B        S.E.      Wald     df      Sig        R     Exp(B)

APPLINC         -.0237      .0020  146.9443      1     .0000   -.1500    .9766
LOCALBNK        -.4537      .0737   37.8755      1     .0000   -.0746    .6353
BLACK            .5296      .2878    3.3871      1     .0657    .0147   1.6983
BLKINC           .0042      .0079    .2759      1     .5994    .0000   1.0042
BLKLOCAL         .4112      .2849    2.0834      1     .1489    .0036   1.5087
Constant       -1.1770      .0753  244.1027      1     .0000
```

III.   *Essay.* (25 points) Answer *one* of the following questions.

1.   Your psychology professor has told you that you should almost always focus on standardized, rather than unstandardized (metric) coefficients. Explain to your professor (as politely as possible) why he is wrong. Among other things, you may want to discuss the relative strengths and weaknesses of standardized vs. unstandardized coefficients with regard to:
      a.   Variables with arbitrary metrics (e.g. attitudinal scales)
      b.   Structural equation models
      c.   Multiple-group comparisons
      d.   Interpretability of coefficients
      e.   Effect of random measurement error on coefficients

2.   Several assumptions are made when using OLS regression. Discuss TWO of the following. What does the assumption mean? When might the assumption be violated? What effects do violation of the assumption have on OLS estimates? How can violations of the assumption be avoided or dealt with? Be sure to talk about techniques such as 2SLS and logistic regression where appropriate. [NOTE: While the material from the last third of the course is especially relevant here, you should try to tie in earlier material as much as possible too.]
      a.   The effects of the independent variables are linear
      b.   Errors are homoskedastic
      c.   Variables are measured without error
      d.   The X's (independent variables) are uncorrelated with the residuals