Sociology 593
Exam 3
May 5, 1998

I. True-False. (25 points) Indicate whether the following statements are true or false. If false, briefly explain why.

1. Two stage least squares should be used when a model is under-identified.

2. Pairwise deletion, listwise deletion, and mean substitution are means for dealing with missing data in both logistic and OLS regression.

3. When comparing two populations, the structural effects could actually be smaller in one population, yet the R^2 in that population could be larger.

4. In both recursive and non-recursive models, all pairs of error terms in the model are assumed to be uncorrelated.

5. For a dichotomous dependent variable, if the OLS assumptions of linearity and additivity are NOT met, then WLS is preferable to OLS regression.

II. Short answer. (15 pts each; 45 pts total; up to 10 points extra credit). Answer *three* of the following (up to 10 pts. extra credit for getting all 4 right).

1. A graduate school has collected data on its applicants for admission. The variables are: LOWINC (1 = Applicant is low income, 0 = Not Low Income), BLACK (1 = Applicant is black, 0 = Not black) and PRESTIGE (prestige of the student's undergraduate program, where scores range from a low of 1 to a high of 10). The dependent variable is ADMIT (1 = Applicant was admitted, 0 = Not admitted). The logistic regression gives the following:

$b_{Lowinc} = -3.0$ $b_{Black} = -1.0$ $b_{Prestige} = 0.2$ $a = 2.0$

Complete the following table:

Lowinc	Black	Prestige	Log odds	Odds	P(Getting admitted)
Low income	Black	1			
Low income	Black	10			
Not low income	Black	1			
Not low income	Not Black	10			

2. *Subprime lenders* specialize in making loans to low income/ higher risk applicants. In exchange, they typically charge a higher rate of interest than do other lenders.

The following analyses were done using the 1996 Home Mortgage Disclosure Act Data for the state of Indiana. All home purchase applications that were either approved or denied are included. The dependent variable is *Denial* (1 if the loan application was denied, 0 otherwise). The independent variables are *Sublend* (1 = subprime lender, 0 = regular lender), *Black* (1 = black applicant, 0 = nonblack applicant), *Applinc* (Applicant income, in thousands of dollars), and *Blacksub* (*Black * Sublend*).

Review the following logistic regressions, and answer the following questions. Be sure to cite evidence from the printout that supports your claims.

- What percentage of home mortgage loan applications were denied in Indiana in 1996?
- Was the denial rate higher for subprime lenders or for regular lenders? If the subprime rate was higher, was this due to differences across lenders in the racial and economic characteristics of loan applicants?
- Subprime lenders claim to be more "color blind" than other types of lenders – blacks who apply to subprime lenders are not disadvantaged relative to whites. Do the analyses support their claims?

Logistic Regression - Model 1 – Sublend only

```
Dependent Variable..   DENIAL     Was loan denied?

Beginning Block Number  0.  Initial Log Likelihood Function

-2 Log Likelihood   105027.01

* Constant is included in the model.


Beginning Block Number  1.  Method: Enter

Variable(s) Entered on Step Number
1..       SUBLEND   Is this a subprime lender?

-2 Log Likelihood    83559.633

Classification Table for DENIAL
The Cut Value is .50
                   Predicted
                  No      Yes     Percent Correct
                    N  I    Y
Observed       +-------+-------+
   No      N   I 78406 I  6820 I   92.00%
               +-------+-------+
   Yes     Y   I  9312 I 11503 I   55.26%
               +-------+-------+
                          Overall  84.79%
```

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

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

SUBLEND        2.6533     .0188 19904.29     1    .0000   .4353   14.2014
Constant      -2.1306     .0110 37783.80     1    .0000
```

Logistic Regression - Model 2 – Sublend + Black, Income

```
Beginning Block Number  2.  Method: Enter

Variable(s) Entered on Step Number
1..        APPLINC   Applicant income
           BLACK     Is applicant black?

-2 Log Likelihood    80907.977

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

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

SUBLEND        2.4062     .0194 15309.06     1    .0000   .4280   11.0922
APPLINC        -.0210     .0005 1922.084     1    .0000  -.1516     .9792
BLACK           .2442     .0443  30.4623     1    .0000   .0185    1.2766
Constant      -1.1959     .0224 2851.154     1    .0000
```

Logistic Regression - Model 3 – Model 2 + Sublend * Black

```
Beginning Block Number  3.  Method: Enter

Variable(s) Entered on Step Number
1..        BLACKSUB  Black * Sublend Interaction

-2 Log Likelihood    80866.354

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

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

SUBLEND        2.4289     .0198 15065.45     1    .0000   .4315   11.3468
APPLINC        -.0209     .0005 1904.506     1    .0000  -.1533     .9793
BLACK           .3791     .0472  64.5166     1    .0000   .0278    1.4609
BLACKSUB       -.7034     .1076  42.7337     1    .0000  -.0224     .4949
Constant      -1.2074     .0225 2882.821     1    .0000
```

3. A sociologist and a psychologist both looked at the following results from a 2 population comparison:

Population 1

Descriptive Statistics

	Mean	Std. Deviation	N
X1	.000000	1.000000	500
X2	.000000	7.141428	500
X3	.000000	45.552170	500
X4	.000000	112.42330	500

Correlations

		X1	X2	X3	X4
Pearson Correlation	X1	1.000	.140	.110	.151
	X2	.140	1.000	.476	.582
	X3	.110	.476	1.000	.905
	X4	.151	.582	.905	1.000

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.317		.000	1.000
	X1	1.000	.317	.140	3.156	.002

a. Dependent Variable: X2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	1.792		.000	1.000
	X1	2.000	1.812	.044	1.104	.270
	X2	3.000	.254	.470	11.823	.000

a. Dependent Variable: X3

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	1.955		.000	1.000
	X1	4.000	1.979	.036	2.021	.044
	X2	3.000	.313	.191	9.575	.000
	X3	2.000	.049	.810	40.875	.000

a. Dependent Variable: X4

Population 2

Descriptive Statistics

	Mean	Std. Deviation	N
X1	.000000	2.236068	500
X2	.000000	7.416198	500
X3	.000000	46.636890	500
X4	.000000	117.45210	500

Correlations

		X1	X2	X3	X4
Pearson Correlation	X1	1.000	.302	.240	.324
	X2	.302	1.000	.506	.614
	X3	.240	.506	1.000	.908
	X4	.324	.614	.908	1.000

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.317		.000	1.000
	X1	1.000	.142	.302	7.057	.000

a. Dependent Variable: X2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	1.792		.000	1.000
	X1	2.000	.842	.096	2.376	.018
	X2	3.000	.254	.477	11.823	.000

a. Dependent Variable: X3

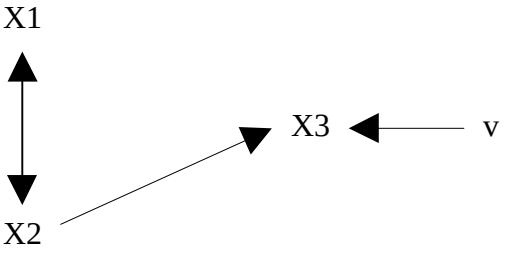
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	1.955		.000	1.000
	X1	4.000	.923	.076	4.333	.000
	X2	3.000	.313	.189	9.575	.000
	X3	2.000	.049	.794	40.875	.000

a. Dependent Variable: X4

The sociologist concluded that the two populations were very similar. The psychologist concluded that the two populations were very different, at least with regards to the effect that X1 has in both populations. Explain what you think each scholar was basing her conclusions on. Then explain which you feel has the better case. HINT: The sociologist is correct of course.

4. A researcher believes in the following model:



A sample of 100 cases is collected. When she regresses X3 on X1 and X2, she gets $b_{31} = .2$, $b_{32} = .3$, $r_{12} = .4$. All variables are in standardized form. Test whether the over-identifying restriction in her preferred model appears reasonable.

III. *Essay.* (30 points) Answer *one* of the following questions.

1. Present a substantive problem, real or hypothetical, where a nonrecursive model might be appropriate. Explain why you think the model should be nonrecursive. What problems might you encounter if you tried to use OLS regression to estimate this model? Even if you are correct in saying the model is nonrecursive, explain why it might be difficult for you to estimate your model.

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