Sociology 593
Exam 2
March 27, 1998

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

1. X1 and X2 are positively correlated. So, if something increases the value of X1, the value of X2 will increase too.

2. A researcher hypothesizes that education positively affects Political Liberalism. Further, she thinks the effect of education is greater for men than it is women. She gets

$\beta_{Education} = 6$

$\beta_{Male} = 0$

$\beta_{Education*Male} = -4$

Male = 1 if male, 0 if female. The standard error for the interaction term is .25. N = 1,200. The evidence supports the researcher's hypothesis.

3. When extraneous variables are added to a model, the expected values of their regression coefficients are zero. Therefore, there is no harm to including extraneous variables in a model.

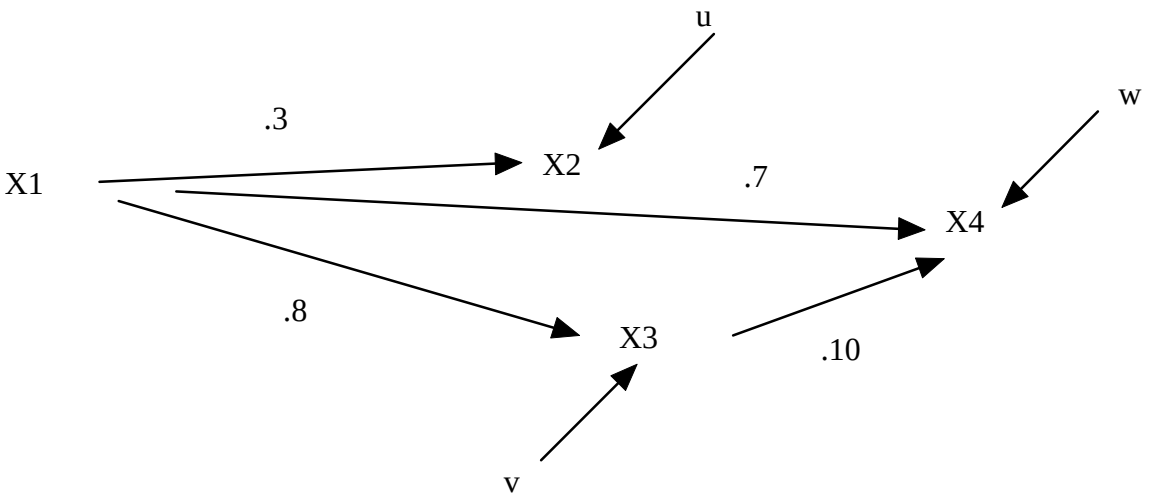
4. A researcher estimates the following two models:

$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

$Y = \beta_1 X_1 + \beta_4 X_4 + \varepsilon$

She can use an incremental F test to determine which of these two models is better.

II. Path Analysis/Model specification. (30 points). A sociologist believes that the following model describes the relationships between X1, X2, X3 and X4. All variables are in standardized form. The hypothesized value of each path is included in the diagram.



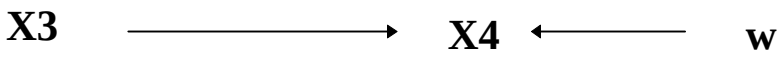
a. Write out the structural equation for each endogenous variable.

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 X3 and X4 into

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

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

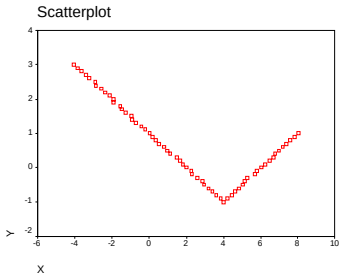


What conclusions would the researcher likely draw? In particular, what would the researcher conclude about the effect of changes in X3 on X4? Why would he make these mistakes? Discuss the consequences of this mis-specification.

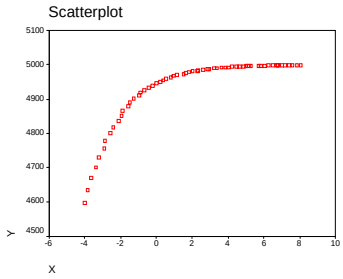
III. *Short answer*. Answer *two* of the following three questions. (25 points each; up to 10 points extra credit if you do all 3).

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

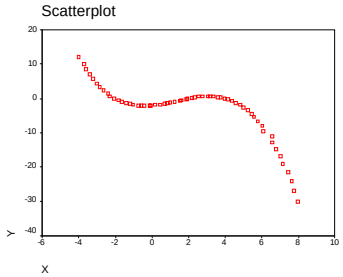
a.



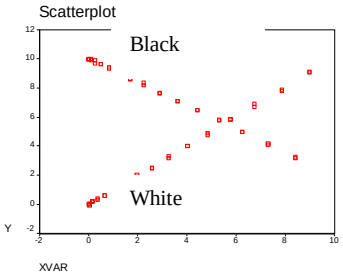
b.



c.



d.



Note: One set of points is for blacks, the other set is for whites

2. The Republican Party wants to take advantage of the alleged sex scandals in the Clinton Presidency. It has therefore come up with an advertising campaign designed to undermine President Clinton's popularity and support. The ads appear in Good Housekeeping, Cosmopolitan, Redbook, and other magazines with large female readerships. Much to its surprise, the party finds that people who read the ads are actually more supportive of President Clinton than people who do not read them. Some critics within the party maintain that the ads are doing Republicans more harm than good and should be stopped, while others continue to insist that the ads are serving their purpose. The party has hired you, a professionally trained social scientist, to give it insight on why these relationships exist. Drawing on your knowledge of <u>the logic of causal order</u>, present

different models that could account for the observed relationships. Indicate what implications the different models have for what should be done about the advertising campaign. To be fair, you will want to present one or more models that suggest that the ads hurt president Clinton, one or more models which imply the ads help him, and one or two models which suggest that the ads are not achieving what the Republicans want but the problems are correctable (i.e. you don't have to completely stop the ads to solve the problem). When presenting your answer, keep in mind that party officials do not know very much about the logic of causal order, so you will have to make things very clear for them.

3. A child psychologist is interested in how grades affect popularity, and how that relationship varies across races. She has collected data on Popularity (scale runs from a low of 0 to a high of 60), Grades (scale ranges from a low of 0 to a high of 25) and Race (Black = 1 if black, 0 if white). She therefore runs the following regressions. Indicate whether there appear to be statistically significant differences in the determinants of popularity between blacks and whites. If so, tell whether these differences are limited to differences in the intercepts, or whether the effect of grades differs between blacks and whites. Briefly discuss the substantive implications of what you think is the best model. Include in your discussion (1) what proportion of the sample is black (2) which race tends to be more popular, (3) which race (if either) tends to get higher grades (4) whether higher grades tend to increase or decrease popularity (5) whether the effect of grades significantly differs across the races, and if so, for which race is the effect stronger? Be sure to indicate how the printout supports your arguments.

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
POPULAR	27.3621	11.1068	250
GRADES	11.1400	3.9779	250
BLACK	.2000	.4008	250
BLACKGRD	2.1000	4.4100	250

Correlations

		POPULAR	GRADES	BLACK	BLACKGRD
Pearson Correlation	POPULAR	1.000	.646	-.223	-.142
	GRADES	.646	1.000	-.081	.022
	BLACK	-.223	-.081	1.000	.954
	BLACKGRD	-.142	.022	.954	1.000

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.646[a]	.417	.415	8.4982	.417	177.327	1	248	.000
2	.668[b]	.446	.442	8.2987	.029	13.064	1	247	.000
3	.669[c]	.447	.440	8.3099	.001	.337	1	246	.562

a. Predictors: (Constant), GRADES

b. Predictors: (Constant), GRADES, BLACK

c. Predictors: (Constant), GRADES, BLACK, BLACKGRD

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12806.403	1	12806.403	177.327	.000[a]
	Residual	17910.347	248	72.219		
	Total	30716.750	249			
2	Regression	13706.136	2	6853.068	99.509	.000[b]
	Residual	17010.613	247	68.869		
	Total	30716.750	249			
3	Regression	13729.418	3	4576.473	66.274	.000[c]
	Residual	16987.331	246	69.054		
	Total	30716.750	249			

a. Predictors: (Constant), GRADES

b. Predictors: (Constant), GRADES, BLACK

c. Predictors: (Constant), GRADES, BLACK, BLACKGRD

d. Dependent Variable: POPULAR

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.278	1.601		4.546	.000
	GRADES	1.803	.135	.646	13.316	.000
2	(Constant)	8.660	1.610		5.380	.000
	GRADES	1.764	.133	.632	13.301	.000
	BLACK	-4.758	1.316	-.172	-3.614	.000
3	(Constant)	8.968	1.696		5.286	.000
	GRADES	1.737	.141	.622	12.335	.000
	BLACK	-7.363	4.676	-.266	-1.575	.117
	BLACKGRD	.246	.424	.098	.581	.562

a. Dependent Variable: POPULAR