Sociology 593
Exam 2 Answer Key
March 28, 2003

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

1. WHITE is coded 1 = white, 0 = nonwhite. X1 is a continuous variable ranging from 0 to 100. WHITEX1 = WHITE * X1. A researcher regresses Y on WHITE, X1, and WHITE*X1. If the coefficient for WHITEX1 is insignificant, this means that X1 does not significantly affect whites.

False. This means that the effect of X1 is the same for both whites and blacks.

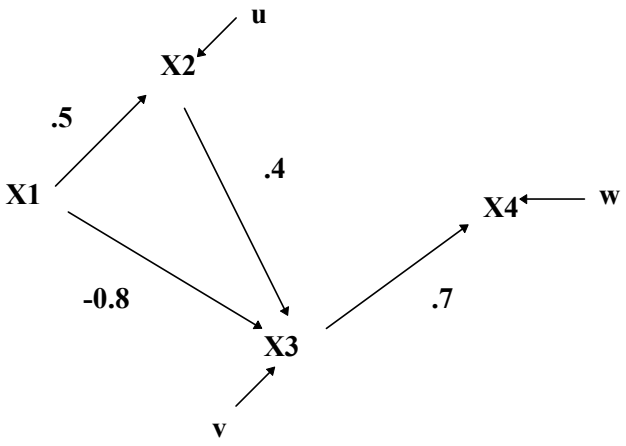
2. A larger sample size will help to reduce the problems caused by adding extraneous variables to a model.

True. Extraneous variables increase standard errors and widen confidence intervals. A larger sample size will cause the standard errors to be smaller.

3. A researcher is unsure whether a quadratic model or an exponential growth model is most appropriate for her data. An incremental F test should be used to decide which model is best.

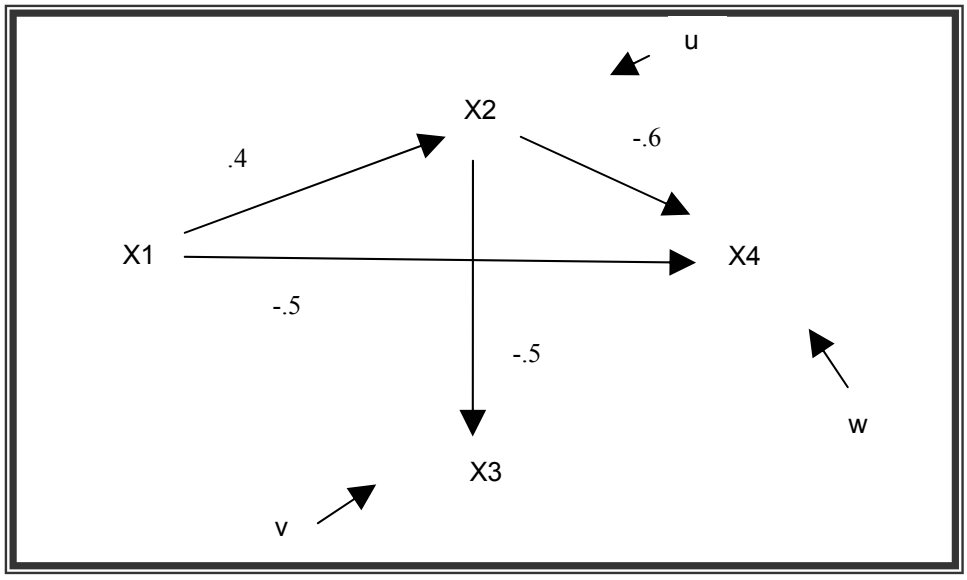
False. These models are not nested, hence it is not possible to use an incremental F test to choose between them. You should rely on theory when deciding which approach is best.

4. All variables are standardized. A researcher has estimated the following model. According to this model, if X1 increases by 1 standard deviation, the expected value of X3 will decrease by .8 standard deviations.



False. While the direct effect of X1 on X3 is -.8, the Total Effect is -.6. Increases in X1 will also produce increases in X2 which in turn will produce increases in X3.

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



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

$X2 = \beta_{21}X1 + u = .4 * X1 + v$

$X3 = \beta_{32}X2 + u = -.5 * X2 + v$

$X4 = \beta_{41}X1 + \beta_{42}X2 + w = -.5 * X1 - .6* X2 + w$

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.)

Using normal equations,

$r12 = \beta_{21}E(X1^2) = .4$

$r13 = \beta_{32}E(X1X2) = -.5*.4 = -.2$

$r23 = \beta_{32}E(X2^2) = -.5$

$r41 = \beta_{41}E(X1^2) + \beta_{42}E(X1X2) = -.5 + (-.6 * .4) = -.74$

$r42 = \beta_{41}E(X1X2) + \beta_{42}E(X2^2) = (-.5 * .4) + -.6 = -.8$

$r43 = \beta_{41}E(X1X3) + \beta_{42}E(X2X3) = (-.5 * -.2) + (-.6 * -.5) = .4$

	X1	X2	X3	X4
X1	1.00			
X2	0.40	1.00		
X3	-.20	-.50	1.00	
X4	-.74	-.80	0.40	1.00

c. Decompose the correlation between X2 and X4 into

- Correlation due to direct effects

-.60

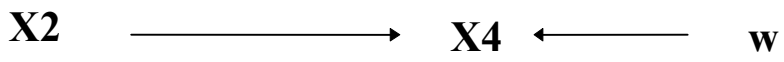
- Correlation due to indirect effects

0

- Correlation due to common causes

-.20

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 X2 on X4? Why would she make these mistakes? Discuss the consequences of this mis-specification.

The estimated effect would be -.8 instead of -.6, i.e. the effect of X2 on X4 would be overestimated. There would be omitted variable bias; X2 would get credit for association that is actually due to X1 being a common cause of X2 and X3. Hence, changes in X2 (unless they were produced by changes in X1) would not produce as much change in X4 as the researcher expected.

We can also see this using our formulas for omitted variable bias. In this case, X1 is being erroneously omitted.

$$\beta_{42}^{*} = \beta_{42} + \beta_{41}\frac{\sigma_{12}}{\sigma_2^2} = -.6 + \left(-.5 * \frac{.4}{1.0}\right) = -.8$$

III. Group comparisons/ Interaction effects (30 points). A recent study done by professors at the University of Chicago Graduate School of Business and the Massachusetts Institute of Technology found that resumes with white-sounding first names (such as Neil, Brett, Greg, Emily, Anne and Jill) received 50% more responses from potential employers than did identical resumes with African American-sounding first names (such as Tamika, Ebony, Aisha, Rasheed, Kareem and Tyrone).

A followup study is extending this research. Fictitious resumes have been constructed such that applicants' qualifications (QUAL) range from extremely low to extremely high (as measured on a scale that ranges from 0 to 100). Otherwise identical resumes are then given white sounding first names or black sounding first names (BLACK, coded 0 if first name is white-sounding, 1 if black-sounding). Two hundred company executives, who know nothing about the purpose of the experiment, are then asked to rate how much they would like to hire the applicant on a scale (HIRE) that again ranges from a low of 0 to a high of 100.

Based on the printout on the next two pages, answer these questions. Be sure to indicate how the printout supports your arguments.

a) The researcher runs two sets of regressions. In the first set of regressions, the coefficient for BLACK changes from being significantly negative (-13.1) in model 2 to being insignificantly positive (1.244) in model 3. How do we interpret this change??? What exactly does the coefficient for BLACK in model 3 tell us about differences between whites and blacks in this first set of regressions?

In Model 2, the coefficient for black is the predicted difference between a white and a black who have comparable levels of qualifications. This includes, but is not limited to, the special case where a white and a black both have a score of 0 on the qualifications scale. In Model 3, once the interaction term is added, the interpretation of BLACK becomes narrower. Now, it is the predicted difference between a white and a black who have 0 qualifications. The fact that the BLACK coefficient is insignificant simply means that there isn't much difference in the likelihood of an employer wanting to hire a white or a black who is totally unqualified.

b) She then computes a new variable, QUALX, computes a new interaction term, BLKQUALX, and runs a second set of regressions using the new variables. Why does she do this??? In model 3 of the 2nd set of regressions, why is the coefficient for BLACK now -13.1? How should we interpret this coefficient, i.e. what does it tell us about differences between whites and blacks? Make clear how and why the interpretation of the BLACK coefficient in Model 3 differs between the two sets of regressions.

She centers the qualifications variable (i.e. subtracts the mean of 53.6758 from every case) to make the interpretation of the effects of race a little easier and more meaningful. The coefficient for BLACK is the predicted difference between a white and a black who score 0 on the QUALX variable. But, because of centering, this now corresponds to the predicted difference between the white and black with average qualifications, not the white and black who have no qualifications. The coefficient tells us that the average black will score 13.1 points lower on the hiring scale, and that this difference is statistically significant.

c) Using the 2nd set of regressions, indicate whether there appear to be statistically significant differences in the determinants of hiring between blacks and whites. If so, tell whether these differences are limited to differences in the intercepts, or whether the effect of qualifications differs between blacks and whites. Briefly discuss the substantive implications of what you think is the best model; in particular, what does the model say about whether or not there is prejudice for or against blacks?

All the coefficients in Model 3 are significant, hence differences go beyond just differences in the intercept. The negative coefficient for BLKQUALX tells us that the effect of qualifications for black applicants is significantly less than the effect of qualifications for white applicants. Since the resumes were otherwise identical, this strongly suggests that there is prejudice against black applicants (or at least black applicants who have black-sounding names.)

```
* Part III.Interaction effects problem.
GET FILE='D:\SOC593\Exams\Exam2-2003.sav'.

Compute Blkqual = Black * qual.

* Regressions - First try.
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT hire
  /METHOD=ENTER qual  /METHOD=ENTER black  /METHOD=ENTER blkqual  .
```

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
HIRE Hireability scale	29.7685	11.06655	200
QUAL Qualifications of Respondent	53.6758	18.27417	200
BLACK Race of Respondent	.5000	.50125	200
BLKQUAL	26.8379	29.84736	200

Model Summary

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.527[a]	.278	.274	9.42921	.278	76.111	1	198	.000
2	.794[b]	.630	.626	6.76790	.352	187.334	1	197	.000
3	.824[c]	.678	.674	6.32335	.049	29.673	1	196	.000

a. Predictors: (Constant), QUAL Qualifications of Respondent

b. Predictors: (Constant), QUAL Qualifications of Respondent, BLACK Race of Respondent

c. Predictors: (Constant), QUAL Qualifications of Respondent, BLACK Race of Respondent, BLKQUAL

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	12.640	2.073		6.096	.000
	QUAL Qualifications of Respondent	.319	.037	.527	8.724	.000
2	(Constant)	19.190	1.563		12.276	.000
	QUAL Qualifications of Respondent	.319	.026	.527	12.155	.000
	BLACK Race of Respondent	-13.100	.957	-.593	-13.687	.000
3	(Constant)	12.018	1.966		6.112	.000
	QUAL Qualifications of Respondent	.453	.035	.748	13.051	.000
	BLACK Race of Respondent	1.244	2.781	.056	.447	.655
	BLKQUAL	-.267	.049	-.721	-5.447	.000

a. Dependent Variable: HIRE Hireability scale

```
* Above results aren't as easy to interpret as we would like -- so we'll try
* something different.
COMPUTE qualx = qual - 53.6758 .
Compute blkqualx = black * qualx.

* Regressions - 2nd try.
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA CHANGE
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT hire
  /METHOD=ENTER qualx  /METHOD=ENTER black  /METHOD=ENTER blkqualx  .
```

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
HIRE Hireability scale	29.7685	11.06655	200
QUALX	.0000	18.27417	200
BLACK Race of Respondent	.5000	.50125	200
BLKQUALX	.0000	12.92179	200

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	.527[a]	.278	.274	9.42921	.278	76.111	1	198	.000
2	.794[b]	.630	.626	6.76790	.352	187.334	1	197	.000
3	.824[c]	.678	.674	6.32335	.049	29.673	1	196	.000

a. Predictors: (Constant), QUALX

b. Predictors: (Constant), QUALX, BLACK Race of Respondent

c. Predictors: (Constant), QUALX, BLACK Race of Respondent, BLKQUALX

Coefficients[a]

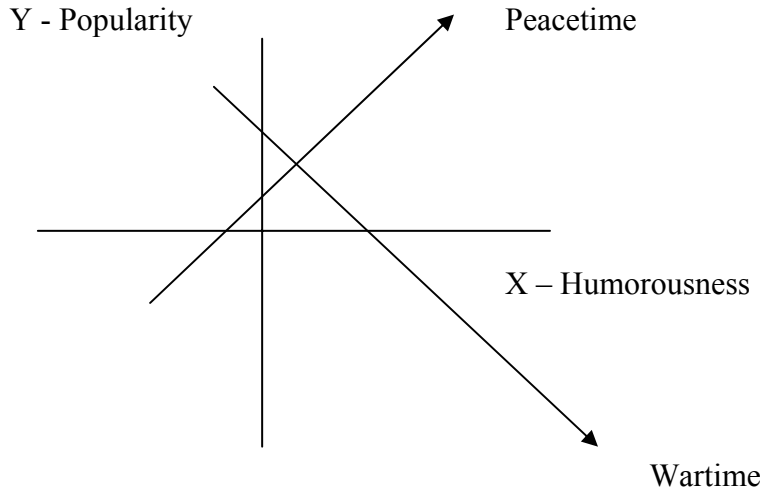
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	29.768	.667		44.647	.000
	QUALX	.319	.037	.527	8.724	.000
2	(Constant)	36.319	.677		53.663	.000
	QUALX	.319	.026	.527	12.155	.000
	BLACK Race of Respondent	-13.100	.957	-.593	-13.687	.000
3	(Constant)	36.319	.632		57.436	.000
	QUALX	.453	.035	.748	13.051	.000
	BLACK Race of Respondent	-13.100	.894	-.593	-14.649	.000
	BLKQUALX	-.267	.049	-.312	-5.447	.000

a. Dependent Variable: HIRE Hireability scale

IV. Short answer. Answer *one* of the following two questions. (20 points; up to 10 points extra credit if you do both).

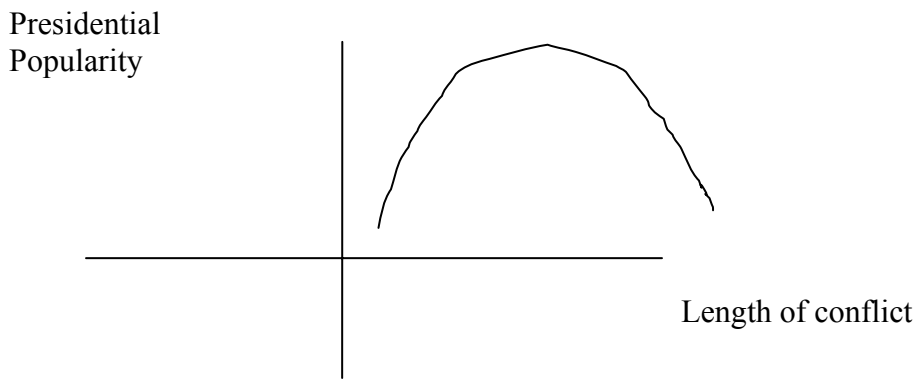
1. Both of the following describe a nonlinear or nonadditive relationship between variables. Draw a scatterplot that illustrates each relationship. Describe the harms that might result if you simply regressed Y on X, e.g. would values be over-estimated, under-estimated, or what? Indicate the model you think should be estimated, e.g. $E(Y) = \alpha + \beta_1 X + \beta_2 X^2$. Explain what variables you would need to compute in order to actually estimate the model, e.g. logs of variables, interaction terms.

a. A company is concerned about its advertising practices during wartime. Six months ago, it found that exposure to humorous ads increased the popularity of its products. However, studies conducted during the first week of the Iraqi war show that exposure to humorous ads actually makes people like its products less.



Interaction effects appear to be present – the effect of humor on a product's popularity depends on whether or not the nation is at war. Perhaps humorous ads are considered inappropriate during wartime. You would have a variable HUMOR that indicates whether the ad is humorous or not (or how humorous it is if humor runs along a continuum.) You would compute a dummy variable, WAR (1 = Wartime, 0 = Peacetime). Then compute WARHUMOR (WAR * HUMOR). You would then regress POPULARITY on HUMOR, WAR, and WARHUMOR. A significant negative coefficient for WARHUMOR would indicate that humorous ads are less effective in wartime (in fact, in this case, they are actually harmful.) Failure to take this interaction into account could cause you to conclude that humor had no effect on ads and/or miss the differing nature of its effects in peacetime and wartime.

b. President Bush is concerned about the effects of the Iraqi conflict on support for him and his policies. Studies of past international conflicts show that support for a nation's leader gradually increases throughout the first month of the conflict. However, after the first month, support starts to decline.



This suggests a curvilinear relationship; conflicts initially produce increases in popularity and support but as they go on longer and longer support begins to erode. A quadratic model sounds appropriate. Y = Leader's popularity, X = Length of conflict so far, X^2 = Length of Conflict squared. Regress Y on X and X^2. A negative coefficient for X^2 will indicate that popularity starts to go down after a certain amount of time. If you didn't take this into account, you might erroneously conclude that length of conflict has no effect on popularity and/or miss how the effect of conflict length changes across time.

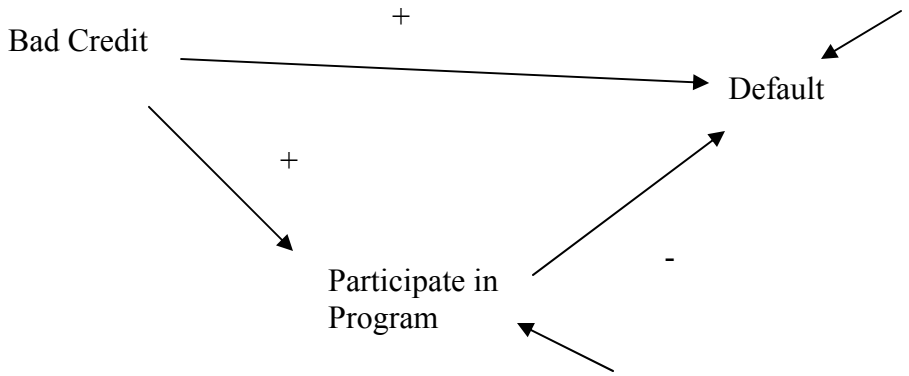
A piecewise regression model might also be plausible.

2. A bank is concerned about rising default rates on its home mortgage loans. It has therefore started a credit counseling program that its borrowers can participate in if they wish. Those who do participate are allowed to make lower down payments than those who do not (which means they do not have to have as much money saved up before getting a loan as they would otherwise).

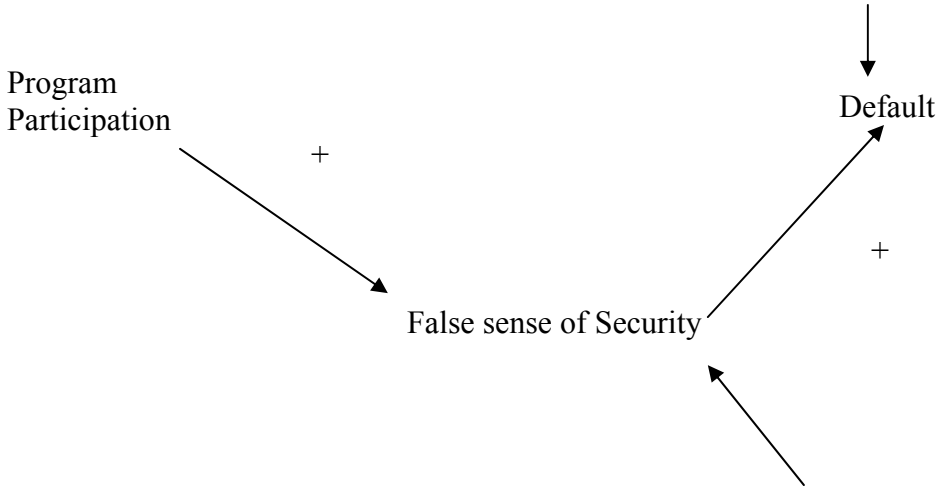
Much to its surprise, borrowers who participate in the program are actually more likely to default on their loans than are its other borrowers. Thus, some critics within the bank maintain that the program is doing more harm than good and should be abandoned, while others continue to insist that the program is serving its purpose.

The bank has hired you, a professionally trained social scientist, to give it insight on why these relationships exist. Drawing on your knowledge of the logic of causal order, present different models that could account for the observed relationships. Indicate what implications the different models have for what should be done about the program. To be fair, you will want to present one or more models that suggest that the program helps borrowers, one or more models which imply that the program does more harm than good, and one or two models which suggest that the program is not achieving what the bank wants but the problems are correctable (i.e. you don't have to completely scrap the program to solve the problem). When presenting your answer, keep in mind that bank staff do not know very much about the logic of causal order, so you will have to make things very clear for them.

* It may be that people who are bad credit risks are more likely to participate in the program. Their default rates would be even higher if the program wasn't there. Hence, the program is good, but not good enough to totally overcome the problems of a bad credit record.



* The program may be harmful. The credit counseling may give inaccurate information. Or, perhaps it creates a false sense of security among participants – they think because they have participated in the program they don't have to worry about default.



* Note that the program actually has two parts. Being in the program lets you make a lower down payment – but a lower down payment could cause you to have less to lose if you go into default, hence you aren't as motivated to avoid default. However, being in the program also gives you more information about handling your finances, and that may be helpful to people in avoiding default. Hence, rather than scrap the whole program, the best strategy may be to keep the counseling part while eliminating the lower down payment component.