

Sociology 593
Exam 2 Answer Key [Draft]
April 1, 2005

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

1. A researcher regresses political liberalism on age and income. If the effect of age is 0, this means that as people get older, their political liberalism does not change.

False. It could mean that. But, it could also be that the effect of age is indirect, i.e. age affects income and income affects liberalism. If so, then people's liberalism will change as they age.

2. In order to use a Wald test, you need to estimate both the constrained and unconstrained models.

False. The Wald test requires only the estimation of the unconstrained model.

3. A Chow test can be used to test whether coefficients differ across groups, e.g. men and women.

True.

4. A researcher believes that years in the labor force has a stronger (more positive) effect on the occupational prestige of women than it does for men. If she is right, and if all other variables have the same effect on men and women, this means that women have higher occupational prestige than do men.

False. If women have fewer years in the labor force than men do, they could still wind up with lower levels of occupational prestige.

5. A researcher hypothesizes that Parental Socio-Economic Status (PSES) has a stronger effect on the educational achievement of blacks than it does for whites (White = 1 if white, 0 if black). She finds that

$$\beta_{\text{PSES}} = 4$$

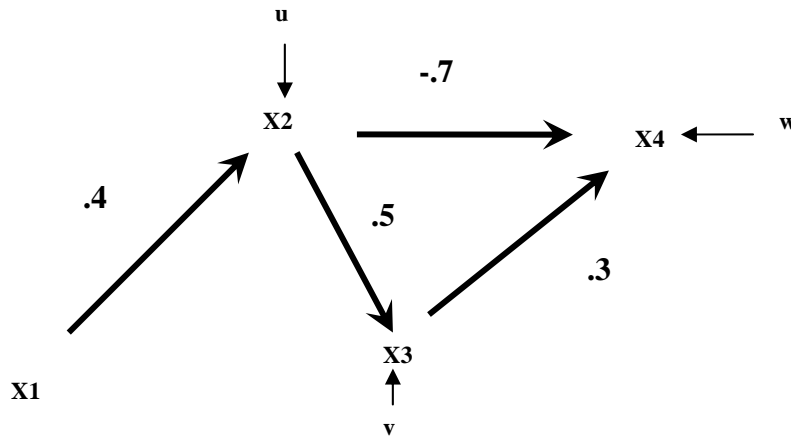
$$\beta_{\text{White}} = 0$$

$$\beta_{\text{White*PSES}} = 1$$

The results are consistent with her hypothesis.

False. The interaction effect shows that Parental SES has a larger effect on whites than it does blacks. It would have to be negative to support the researcher's argument.

II. Path Analysis/Model specification (30 pts). 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. (10 pts) Write out the structural equation for each endogenous variable.

$$X_2 = \beta_{21}X_1 + u = .4X_1 + u$$

$$X_3 = \beta_{32}X_2 + v = .5X_2 + v$$

$$X_4 = \beta_{42}X_2 + \beta_{43}X_3 + w = -.7X_2 + .3X_3 + w$$

b. (10 pts) Part of the correlation matrix is shown below. 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.)

	x1	x2	x3	x4
x1	1.0000			
x2	0.4000	1.0000		
x3	?	?	1.0000	
x4	-0.2200	?	?	1.0000

Here is the complete correlation matrix:

	x1	x2	x3	x4
x1	1.0000			
x2	0.4000	1.0000		
x3	0.2000	0.5000	1.0000	
x4	-0.2200	-0.5500	-0.0500	1.0000

c. (5 pts) Decompose the correlation between X3 and X4 into

- Correlation due to direct effects

.3

- Correlation due to indirect effects

0

- Correlation due to common causes

-.35

d. (5 pts) 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.

In the mis-specified model, the standardized coefficient will be the same as the bivariate correlation, -.05. The researcher will therefore conclude that increases in X3 lead to decreases in X4. However, in the correctly specified model, we see that the effect of X3 on X4 is .3, i.e. increases in X3 lead to increases in X4. By failing to take into account the common cause, X2, the research will not only mis-estimate the magnitude but also the direction of the effect of X3 on X4. If policy issues were involved, policy makers might do exactly the opposite of what they should do.

III. Group comparisons (30 points).

A researcher has constructed a 100 point scale (workmom) that measures the level of support for a mother of young children being in the labor force. She believes that education (ed, measured in years) will affect the level of support. She also believes that there may be differences in the model for men and women (male = 1 if male, 0 if female), but she is not sure what those differences will be. She estimates the following models.

```
. * Model 1
. reg workmom ed
```

Source	SS	df	MS			
Model	80897.1481	1	80897.1481	Number of obs =	2293	
Residual	2069606.67	2291	903.363889	F(1, 2291) =	89.55	
Total	2150503.82	2292	938.265191	Prob > F =	0.0000	
				R-squared =	0.0376	
				Adj R-squared =	0.0372	
				Root MSE =	30.056	

workmom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ed	1.87957	.1986202	9.46	0.000	1.490076	2.269065
_cons	30.08284	2.50661	12.00	0.000	25.16738	34.9983

. * Model 2

. reg workmom ed male

Source	SS	df	MS	Number of obs =	2293
Model	151260.364	2	75630.1821	F(2, 2290) =	86.63
Residual	1999243.45	2290	873.032076	Prob > F =	0.0000
Total	2150503.82	2292	938.265191	R-squared =	0.0703
				Adj R-squared =	0.0695
				Root MSE =	29.547

workmom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ed	1.941475	.1953789	9.94	0.000	1.558337 2.324613
male	-11.11334	1.237904	-8.98	0.000	-13.54087 -8.685811
_cons	34.49301	2.512658	13.73	0.000	29.56568 39.42033

. test male

(1) male = 0

F(1, 2290) = 80.60
Prob > F = 0.0000

. * Model 3

. gen edmale = ed * male

. reg workmom ed male edmale

Source	SS	df	MS	Number of obs =	2293
Model	157638.208	3	52546.0693	F(3, 2289) =	60.35
Residual	1992865.61	2289	870.627178	Prob > F =	0.0000
Total	2150503.82	2292	938.265191	R-squared =	0.0733
				Adj R-squared =	0.0721
				Root MSE =	29.506

workmom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ed	2.561016	.3007722	8.51	0.000	1.971202 3.150831
male	1.945303	4.980628	0.39	0.696	-7.821713 11.71232
edmale	-1.069662	.395208	-2.71	0.007	-1.844665 -.2946587
_cons	26.98782	3.739688	7.22	0.000	19.65429 34.32135

. test edmale

(1) edmale = 0

F(1, 2289) = 7.33
Prob > F = 0.0068

. test male edmale

(1) male = 0

(2) edmale = 0

F(2, 2289) = 44.07
Prob > F = 0.0000

. * Model 4 - alternate model suggested by smart (?) graduate student

. egen edmean = mean(ed)

. gen xed = ed - edmean

. gen xedmale = xed * male

```
. reg workmom xed male xedmale
```

Source	SS	df	MS	Number of obs = 2293		
Model	157638.208	3	52546.0693	F(3, 2289)	=	60.35
Residual	1992865.61	2289	870.627178	Prob > F	=	0.0000
				R-squared	=	0.0733
				Adj R-squared	=	0.0721
Total	2150503.82	2292	938.265191	Root MSE	=	29.506

workmom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
xed	2.561016	.3007722	8.51	0.000	1.971202	3.150831
male	-11.12389	1.236204	-9.00	0.000	-13.54808	-8.699689
xedmale	-1.069662	.395208	-2.71	0.007	-1.844665	-.2946587
_cons	58.27846	.8429325	69.14	0.000	56.62546	59.93145

```
. * ttest - why did my student think I might want this???
```

```
. ttest ed, by(male)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Women	1227	12.1141	.0799854	2.801774	11.95718	12.27102
Men	1066	12.33771	.1080179	3.526749	12.12576	12.54966
combined	2293	12.21805	.0660083	3.160827	12.08861	12.3475
diff		-.2236116	.1322897		-.4830317	.0358084

Degrees of freedom: 2291

Ho: mean(Women) - mean(Men) = diff = 0

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
t = -1.6903	t = -1.6903	t = -1.6903
P < t = 0.0456	P > t = 0.0911	P > t = 0.9544

Based on the above results, answer the following questions:

a) (15 pts) The researcher begins by estimating three models. Which of these three models do you think is best, and why? Summarize what your preferred model says about the effect of education on feelings toward working mothers, and what it tells you about differences between men and women. Be sure to make clear whether the effect of education differs by gender, and if so what do those differences tell us?

Go with Model 3, which is a statistically significant improvement over models 1 and 2. This model says that the better educated you are, the more supportive you are of working mothers. However, the effect of education is less strong for men than for women.

b) (10 pts) Based on model 3, the researcher had concluded that gender had no effect once other variables were controlled for. However, her graduate student then gave her model 4. Now she does not know what to think. Explain to her why the effect of male seems to differ so much between the two models and how the effects should be interpreted. In particular, what do these results tell us about how the “average” (average on education, that is) man and woman differ in their support for working mothers?

Model 3 says that, for a man and a woman with 0 years of education, the difference in their support for working mothers is small and statistically insignificant. In Model 4,

education has been centered about its mean. Hence, Model 4 tells us that, for a man and a woman who both have average levels of education (12.218 years, as the t-test shows), the average man scores about 11 points lower than the average woman. The two models are identical but Model 4 presents the results in a way that may be easier to interpret.

c) The researcher's assistant also ran a t-test. Why do you think she did that, i.e. how does it add to our understanding of gender differences in support for working mothers? Do you think that differences in the amount of education men and women have plays a major role in determining male/female differences in support for working mothers?

Differences between men and women in support for working mothers could also be due to gender differences in education levels. The t-test shows that such differences are small however, so differences in education are only a small factor with regards to differences in support for working mothers.

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. At first, each additional hour spent studying gradually increases your exam score. However, after 6 hours, each additional hour spent studying actually results in a gradual decrease in your score.

This suggests a curvilinear relationship. Include X^2 in the model. If you don't do this, you will at times underestimate and at other times overestimate the predicted score.

b. Workers are paid an hourly wage, which depends, in part, on their seniority and past job performance. After working 40 hours in a week, the workers receive overtime pay and the hourly wage goes up 50%.

Estimate a piecewise regression model, where the slope for hours is allowed to change after hours = 40.

2. President Bush has been speaking to large audiences across the country, trying to build support for his plan to reform Social Security. He is therefore dismayed by polls which show that support for the plan is eroding. Even worse, the polls show that people who have heard the President speak are less supportive of the plan than people who have not heard him speak. Some of his advisors are urging him to discontinue his talks because they are doing more harm than good, while others maintain that the talks are helping to achieve their intended purpose and should be continued.

The President has therefore called on you, a professionally trained social scientist, to explain why these relationships might 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 talks actually build support for his plan, one or more models which imply that the talks do more harm than good, and one or two models which suggest that the talks are not achieving what the President wants but the problems are correctable (i.e. you don't have to completely scrap the talks to solve the problem). When presenting your answer, keep in mind that the President was only a C student in college, so you will have to make things very clear for him. Don't just draw diagrams; explain substantively what the models mean and why they might be plausible.

Talks are bad. Hearing talks increases informedness. The better informed people are, the less likely they are to support Bush's plan.

Talks are good. The people who are more likely to attend talks are people who are more skeptical of the plan. Attending the talks tends to increase support, but not enough to totally overcome the skepticism people come in with.

The problem is correctable. Attending the talks increases informedness (which hurts support) but also increases liking for Bush (which increases support, but not enough to offset the loss from having more information). Bush should scrap the informative parts of his talk and just concentrate on getting people to like him more.