

Sociology 593
Exam 2
March 28, 2002

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

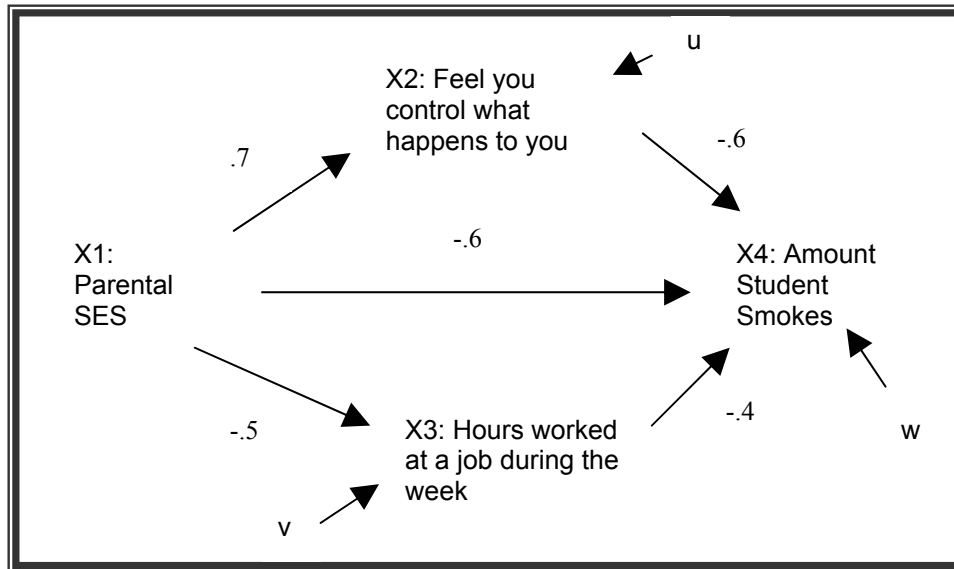
1. A variable is called CATHOLIC. This probably means that only Catholics have a score on this variable.
2. A researcher believes that Education is a cause of earnings: the more educated you are, the higher your earnings tend to be. Further, she knows that blacks make less, on average, than whites do. This means that the effect of education on earnings must be greater for whites than it is for blacks.
3. The correlation between X2 and X4 is .6. Hence, in a path model using standardized variables, the effect of X2 on X4 will be .6.
4. A researcher regresses Y on X1, X2 and X3. In reality, X3 is irrelevant and does not belong in the model. The estimates of the betas will therefore be biased.
5. A researcher computes $SUMX1X2 = X1 + X2$. She regresses Y on X1 and SUMX1X2. She obtains the following. She should conclude that X1 does not affect Y.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.292	2.924		.442	.660
	X1	-.212	.871	-.052	-.243	.808
	SUMX1X2	2.391	.658	.775	3.634	.000

a. Dependent Variable: Y

II. Path Analysis/Model specification. (30 points). A researcher is interested in the determinants of teenage smoking. She has measures of Parental Socio-economic status, hours worked at paying jobs during the school week, feelings of control over your destiny (“Do you feel you control what happens to you in life, or do you think things just happen?”), and the number of cigarettes smoked per week. All her variables are in standardized form. The estimated value of each path in her model is included in the diagram.



- Write out the structural equation for each endogenous variable.
- The correlation between X3 (hours worked) and X4 (# of cigarettes smoked) is

$$\begin{aligned}
 E(X_3 X_4) &= \rho_{43} \\
 &= \beta_{41}\beta_{31} + \beta_{42}\beta_{21}\beta_{31} + \beta_{43} \\
 &= (-.6 * -.5) + (-.6 * .7 * -.5) + (-.4) \\
 &= .30 + .21 - .4 = .11
 \end{aligned}$$

Decompose the correlation between X3 and X4 into

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

c. There is controversy over the harms and benefits of students working at jobs while attending school. Based on these results, do you think that working at a job tends to increase smoking, decrease smoking, or have little or no effect? Explain how, if variables were erroneously omitted from the above model, a researcher might reach a very different, and incorrect, conclusion.

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. Each of the following describes a nonlinear or nonadditive relationship between variables. Draw a scatterplot that illustrates the 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. Studies have found that moderate amounts of alcohol consumption actually tend to increase life expectancy. However, after a certain point, the more alcohol you consume, the more your life expectancy goes down.

b. A political analyst wants to know how TV ads affect the opinions of voters. She finds that, for voters aged 50 and above, the more TV ads they see, the less they like her candidate. For voters younger than 50, just the opposite is true: the more TV ads they see, the more they like the candidate.

c. An employer believes that workers become more skillful and productive the longer they work on the job. Specifically, she believes that each year a worker is on the job, he or she will be 5% more productive than they were the year before.

2. [This problem is modified from Hamilton's Statistics with Stata 5.]. A researcher believes that better students (as measured by Grade Point Average) drink less than do weaker students. However, she suspects that this is less true for men than for women, i.e. the effect of GPA on drinking is smaller for males than it is for females. She tests her ideas with a survey of undergraduate students collected by Ward and Ault (1990). DRINKING is measured on a 33 point scale, where higher values indicate higher levels of drinking. GPA is the student's Grade Point Average, centered so as to have a mean of zero (higher values indicate better grades). MALE is coded 1 if the student is male, 0 if Female. MALEGPA = MALE * GPA. She obtains the following results:

```
REGRESSION
/MISSING LISTWISE/ descriptives
/STATISTICS COEFF OUTS R ANOVA CHA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT drink
/METHOD=ENTER GPA /METHOD=ENTER male /METHOD=ENTER MaleGpa .
```

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
DRINK 33-point drinking scale	18.82	6.739	218
GPA Grade Point Average	.0000	.45917	218
MALE	.4541	.49904	218
MALEGPA	-.0407	.30988	218

Correlations

		DRINK 33-point drinking scale	GPA Grade Point Average	MALE	MALEGPA
Pearson Correlation	DRINK 33-point drinking scale	1.000	-.282	.304	-.170
	GPA Grade Point Average	-.282	1.000	-.178	.687
	MALE	.304	-.178	1.000	-.144
	MALEGPA	-.170	.687	-.144	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	.282 ^a	.080	.075	6.480	.080	18.660	1	216	.000
2	.382 ^b	.146	.138	6.257	.066	16.710	1	215	.000
3	.384 ^c	.148	.136	6.265	.002	.412	1	214	.522

a. Predictors: (Constant), GPA Grade Point Average

b. Predictors: (Constant), GPA Grade Point Average, MALE

c. Predictors: (Constant), GPA Grade Point Average, MALE, MALEGPA

ANOVA^d

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	783.590	1	783.590	18.660	.000 ^a
	Residual	9070.432	216	41.993		
	Total	9854.023	217			
2	Regression	1437.711	2	718.855	18.364	.000 ^b
	Residual	8416.312	215	39.146		
	Total	9854.023	217			
3	Regression	1453.879	3	484.626	12.346	.000 ^c
	Residual	8400.144	214	39.253		
	Total	9854.023	217			

a. Predictors: (Constant), GPA Grade Point Average

b. Predictors: (Constant), GPA Grade Point Average, MALE

c. Predictors: (Constant), GPA Grade Point Average, MALE, MALEGPA

d. Dependent Variable: DRINK 33-point drinking scale

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18.821	.439		42.883	.000
	GPA Grade Point Average	-4.138	.958	-.282	-4.320	.000
2	(Constant)	17.215	.578		29.794	.000
	GPA Grade Point Average	-3.453	.940	-.235	-3.673	.000
	MALE	3.536	.865	.262	4.088	.000
3	(Constant)	17.257	.582		29.640	.000
	GPA Grade Point Average	-4.011	1.282	-.273	-3.129	.002
	MALE	3.553	.867	.263	4.100	.000
	MALEGPA	1.212	1.889	.056	.642	.522

a. Dependent Variable: DRINK 33-point drinking scale

Based on the above results, answer the following questions. Be sure to indicate how the printout supports your arguments.

- Is the researcher correct in hypothesizing that better students tend to drink less?
- Are there significant differences in the determinants of drinking for men and women? If so, are these differences limited to differences in the intercepts, or does the effect of GPA differ by gender? If differences are found, be specific as to what they are, e.g. how much greater (or weaker) is the effect of GPA on men than it is for women.
- Of the three models presented, which one would you say is best, and why? Briefly discuss the substantive implications of your preferred model.

3. A researcher is interested in the following model:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

Using OLS regression, briefly tell her how to go about testing each of the following hypotheses. You don't have to give exact formulas but you should describe the general procedures (e.g. what variables would you need to compute, what models would you need to estimate) and the types of statistics (e.g. T-Test, Global F-test, Incremental F-test) that are appropriate. For example, you might describe the unconstrained and constrained models that need to be estimated, and then say that an incremental F-test should be used to tell whether the two models significantly differ from each other. If necessary, describe any additional variables you would need to compute, such as interaction terms.

- $\beta_1 = 100$
- $\beta_1 = \beta_2$
- $\beta_3 = \beta_4 = 0$
- Suppose X_4 is a dummy variable coded 1 if Black, 0 Otherwise. How would you test the hypothesis that the value of β_1 is different for blacks and whites? Assume that the effects of X_2 and X_3 are the same for both groups.