Sociology 593
Exam 3 Answer Key
May 7, 2002

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

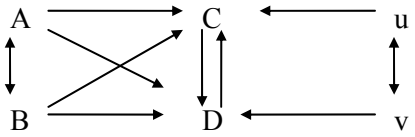
1. The dependent variable is coded 1 = North, 2 = South, 3 = West, 4 = East. Because the DV has more than two categories but is not continuous, an Ordered Logit Model should be used.

False. The categories are not ordered, i.e. the variable is not ordinal. Use multinomial logit instead.

2. In a model with interaction terms, centering the variables will change the predictions that the model makes.

False. The predictions are not changed. But, centering may make the results easier to interpret.

3. The following model is underidentified. One way to solve the identification problem is to add two more variables, e.g. E and F, both of which affect both C and D.



False. If E and F affect both C and D, the model will continue to be underidentified. What would work (at least on paper) is to have E directly affect C and F directly affect D.

4. Censoring and time-dependent explanatory variables can both be addressed via Event History Analysis.

True.

5. Stepwise regression, pairwise deletion of missing data, and analysis of residuals can be done in both OLS and logistic regression.

False. You can't do pairwise deletion of missing data in Logistic Regression.

II. Short answer. (25 pts each; 50 pts total; up to 10 points extra credit). Answer *two* of the following (up to 10 pts. extra credit for getting all 3 right).

II-1. A researcher is interested in the determinants of attitudes toward immigration. Her Variables are: DECRIMM (1 = feels immigration should be decreased, 0 = feels immigration should stay the same or be increased), FORBORN (1 = respondent was born in a foreign country, 0 = respondent was born in the United States), EDUC (years of education), and PARTY (a seven point scale where 1 = Strong Democrat and 7 = Strong Republican). Using the following printout, answer the following questions:

a. What do DEV_M, G_M, and DEV_0 equal?

In the block 1 part of the printout, -2 Log Likelihood corresponds to DEV_M, which is 1667.018. The Model Chi-Square is G_M and equals 67.096. You add these two up to get DEV_0, which is 1734.114. (Note that other statistical routines besides SPSS's Logistic Regression sometimes compute the deviance differently.)

b. What does McFadden's Pseudo R^2 equal?

McFadden's Pseudo $R^2 = G_M/DEV_0 = 67.096/1734.114 = .03869$

c. What is the probability that a Strong Democrat with 16 years of education who was born in a foreign country will want to decrease immigration?

The log odds for such a person are 2.032 – (.118 * 16) – 1.320 + .081 = -1.095.
The odds are exp(-1.095) = .3345.
P(Decrease) = odds/(1 + odds) = .3345/1.3345 = 25.07%

d. What percentage of the respondents feel immigration should be decreased?

As we see in block 0 (the column labeled Exp(B)), the odds of someone wanting to decrease immigration are 1.95. Ergo P(Decrease) = 1.95/(1 + 1.95) = 66.1%.

Logistic Regression

Block 0: Beginning Block

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.668	.057	135.261	1	.000	1.950

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	67.096	3	.000
	Block	67.096	3	.000
	Model	67.096	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1667.018	.048	.067

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	EDUC	-.118	.021	32.024	1	.000	.889
	FORBORN	-1.320	.233	32.113	1	.000	.267
	PARTY	.081	.030	7.503	1	.006	1.085
	Constant	2.032	.299	46.245	1	.000	7.626

a. Variable(s) entered on step 1: EDUC, FORBORN, PARTY.

II-2. For each of the following circumstances describe the statistical technique you would use for revealing the relationship between the dependent and independent variables. Write a few sentences explaining and justifying your answer. In some instances more than one technique may be reasonable.

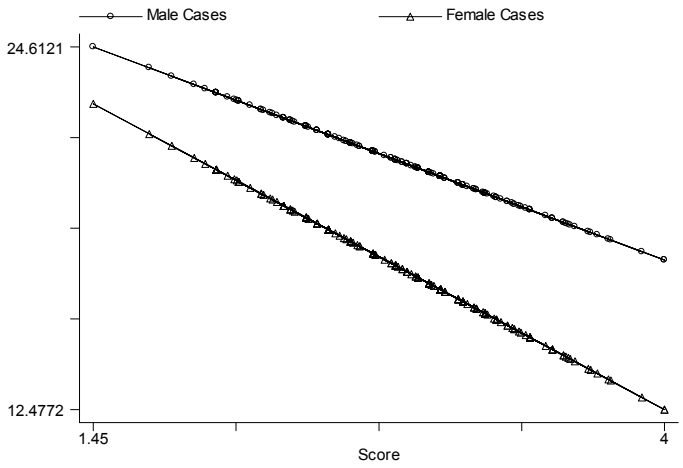
a. The Big Ten Conference is concerned by its dismal performance in College Football over the past three decades. Despite having some of the best talent in the country, the conference has won only one National Championship since 1970 (and even that one was highly controversial). Some claim that the conference's academic standards are too high. Others dispute that, noting that Big 12 member Nebraska, winner of five national titles during this period, has also led the nation in Academic All-Americans. They feel that the problem is a weak non-conference schedule, a consequence, in part, of the Big Ten's long ties to the Rose Bowl and Pac 10. Still others feel that the problem is simply poor coaching. In the hope that it will finally figure out what it needs to do differently, the conference will do a study where the number of national championships won since 1970 is the dependent variable. Data will be collected on all NCAA Division I schools. Independent variables will include academic standards of the school, the number of head coaches during the period studied, and the win-loss records of opponents.

Poisson Regression (i.e. Event Count Model; or, possibly, Negative Binomial). We frequently encounter a Poisson distribution when we are counting the number of times an infrequently occurring event has occurred. Although the Big Ten has done exceptionally poorly, it is still the case that most schools in the country will not experience even one national championship over a 30+ year period. Less but still several will have only 1 or 2 titles, and only the most spectacular programs in the country, such as Nebraska's, will have as many as 5. OLS tends to not work well with event counts until the average number of events gets to be around 8.

b. A prosecutor believes that black women will be harder on alleged criminals in domestic violence cases than will white women. A sample of 30 black women and 30 white women will be drawn. For each respondent, attitudes toward domestic violence will be measured on a scale that ranges from 0 (cares little about domestic violence) to 100 (feels very strongly about domestic violence).

An independent samples T-Test should be used (it probably doesn't much matter whether or not you assume the variances are equal; since the sample sizes are the same, the test statistics will come out the same, although the degrees of freedom may slighty differ). Alternatives that will also work are One-Way Anova and OLS regression using a dummy variable for race.

c. A researcher is interested in testing the relationship between X and Y, both of which are continuous variables. Scatterplots of her data reveal the following:



There appears to be an interaction between Gender and X, i.e. the female slope is different than the male slope. Assuming other assumptions are met, use OLS with a Gender*X interaction term.

d. The airline industry is worried about customer concerns with safety and service. Two experimental advertising campaigns have been developed, one of which will be shown in New York and the other in Los Angeles. In order to assess the effectiveness of these ad campaigns, random samples of residents in both cities will be asked how safe they feel air flight is, how good they think the service provided by airlines is, and how likely they are to fly at least once in the next year. All dependent variables are measured on 100 point scales.

Manova. There are multiple dependent variables. You can also set this up in Lisrel if you prefer, but note that this is NOT a multiple-groups problem. Within each city, only one of the treatments is administered (i.e. people see one ad campaign or the other), hence within each city the value of the IV is a constant.

II-3. A researcher is interested in the relationship between Party Affiliation, Race and Income. She has collected information on the following variables:

- Democrat (1 if Democrat, 0 Otherwise)
- Black (1 = Black, 0 = White)
- Inc (Family Income measured in thousands of dollars. The variable has been centered about its mean)
- BlackInc = Black * Inc

Based on the following printout, answer the following questions. Be sure to cite evidence from the printout to support your answers.

a. Who is more likely to be Democratic, blacks or whites?

The effect of Black in Block 1 is positive and highly significant, so blacks are more likely to be Democrats.

b. Do differences in Income account for racial differences in party affiliation?

No. If they did, the effect of black would become insignificant once income was controlled for. In Block 2, when income is added, the effect of black changes only slightly and remains highly significant.

c. Of the three models presented, which one do you think is best? According to this model, how much does a 1 unit increase in income affect the log odds of a white being a Democrat? How much does a 1 unit increase in income affect the log odds of a black being a Democrat?

All effects are significant in the final model, so we should go with it. This model says that the effect of income varies by race. For whites (who are coded 0 on the interaction term) a 1-unit increase in income produces a .005 drop in the log odds of being a Democrat. For Blacks, to get the effect of income, you add the main effect of income (-.005) to the interaction effect (-.014). Hence, for blacks, the effect of income is even greater: For each 1 unit increase in income, the log odds of being a Democrat drop by .019.

Logistic Regression

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	545.284	1	.000
	Block	545.284	1	.000
	Model	545.284	1	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	7970.261	.085	.113

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	BLACK	2.064	.105	389.234	1	.000	7.880
	Constant	-.302	.028	119.871	1	.000	.739

a. Variable(s) entered on step 1: BLACK.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	34.534	1	.000
	Block	34.534	1	.000
	Model	579.818	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	7935.727	.090	.120

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	BLACK	2.005	.105	363.639	1	.000	7.424
	INC	-.006	.001	33.709	1	.000	.994
	Constant	-.296	.028	114.408	1	.000	.744

a. Variable(s) entered on step 1: INC.

Block 3: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7.783	1	.005
	Block	7.783	1	.005
	Model	587.601	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	7927.944	.091	.122

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	BLACK	1.918	.109	311.933	1	.000	6.810
	INC	-.005	.001	26.272	1	.000	.995
	BLACKINC	-.014	.005	8.080	1	.004	.987
	Constant	-.297	.028	114.792	1	.000	.743

a. Variable(s) entered on step 1: BLACKINC.

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

1. We've talked about several ways that OLS regression can be modified to deal with violations of its assumptions. Some problems, however, require the use of techniques besides OLS. For three of the following, explain why and when the method would be used instead of OLS. Be sure to make clear what assumptions would be violated if OLS was used instead.

a. 2 stage least squares
b. Logistic regression
c. Ordered Logit models
e. Event count models
f. Event History Analysis

2. Path analysis first became popular in Sociology during the 1960s, and has evolved considerably since then.

a. In the early days of path analysis, standardized coefficients were widely used. Give two or three reasons why, in Sociology at least, that practice fell out of favor.

b. In the 1970s, the development of the LISREL program gave new life to path analysis. Discuss some of the key strengths of the LISREL method. Explain how LISREL made it possible to estimate important new sorts of models and how it provided an alternative means for estimating models that could also be approached via other methods.

For the essay questions, see the notes. Certainly, any discussion of LISREL should mention LISREL's ability to deal with measurement error.