Criminal Justice 405
Graduate Introduction to Criminal Justice Statistical Analysis: Simple and Multiple Regression and Beyond
Fall 2001 Final Exam

This exam has three parts. NOTE the percentage each contributes to your total grade. PLAN YOUR TIME ACCORDINGLY. You have two hours.
PART 1. (80%) You answer various results about different sets of OLS multiple regression results, and one logistic regression.
PART 2 (10%) You answer some questions about simple and multiple regression terms.
PART 3 (10%) You answer some questions about the Best book.

PART 1

You will read below the methods section from a research article. This is the text as it appears in the journal article. The citation for the article is:

Krivo, Lauren J; Peterson, Ruth D (1996) Extremely disadvantaged neighborhoods and urban crime. Social Forces 75: 619-648.

At the end of the article I have a few summary comments. Then there are some questions I want you to answer. Those are followed by a table of results. Bear in mind that this is an ecological analysis; the unit of analysis is the census tract.

Following the Krivo and Petersen results is a table of results from the ecological data file and some questions.

Data and Methods

SAMPLE AND DATA The concern in this analysis is with neighborhood disadvantage and crime. The actual units examined are census tracts in the city of Columbus, Ohio for 1990. Census tracts do not necessarily correspond to neighborhoods in a socially meaningful sense. However, they are the best local areas for which the required data are available, and they have been used in prior analyses of urban crime (Crutchfield 1989; Kohfeld & Sprague 1988, 1990; McClain 1989). We examine areas within the central city of the metropolitan area because Wilson's (1987) discussion of the consequences of social isolation focuses on the urban core.

There are a total of 215 census tracts in Columbus although many are split across municipal boundaries and thus are only partially within the city limits. Our analysis includes the 177 tracts (or portions of tracts) with at least 700 persons within the city. This minimum size allows us to construct reliable crime rates and other aggregate characteristics. Applying a widely used categorization of neighborhood poverty levels into low (less than 20%), high (2039%), and extreme (40% or more) Uargowsky & Bane 1990, 1991; Kasarda 1993; Ricketts & Sawhill 1988; Wilson 1987), 54 of Columbus' census tracts have high (N 32) or extreme (N = 22) poverty rates. Twenty-six of Columbus' tracts are at least 70% black and 122 are at least 70% white (the

remainder are more racially mixed). As expected a much higher proportion of the black (38.5%) than white (7.4%) tracts have extremely high levels of poverty, but the number of black and white tracts with extreme rates of poverty are nearly identical.

Data for the independent variables are from the 1990 U.S. Censuses of Population and Housing Summary Tape File 3A (U.S. Bureau of the Census 1991). The Columbus Police Department (1994) provided counts for a variety of types of reported crime for tracts within the city. These data are the same as those reported in the Federal Bureau of Investigation's (FBI) Uniform Crime Reports (UCR) except that they are broken down by census tract.

CRIME RATES

Rates for the FBI's Index Crimes (homicide, forcible rape, robbery, aggravated assault, burglary, larceny, and vehicle theft) provide the dependent variables. We construct separate rates for property (burglary, larceny, and vehicle theft) and violent (homicide, rape, robbery, and aggravated assault) index crimes. Following common practice, three-year (1989--91) average crimes per 1,000 population are calculated to minimize the impact of annual fluctuations and increase the likelihood of having sufficient incidents to construct reliable rates for small areas (e.g., Messner & Golden 1992; Sampson 1985, 1987). Wilson's perspective indicates that poverty and disadvantage contribute to crime by creating structural conditions that enhance both criminal vulnerability and criminal offending. However, our data are for reported victimizations only. While it is important to study offending, the links hypothesized should be evident in analyses of these rates.

NEIGHBORHOOD POVERTY AND DISADVANTAGE

To examine the hypothesis that crime is most pronounced in areas with very high poverty rates, we use dummy variables contrasting high (20%-39%) and extreme (more than 40%) to low (less than 20%) poverty neighborhoods. This categorization is used widely in research on urban poverty and the underclass (Jargowsky & Bane 1990, 1991; Kasarda 1993; Ricketts&Sawhill 1988; Wilson 1987). In addition to extreme levels of poverty, the literature on urban social dislocation emphasizes the pernicious consequences of living in areas with widespread family disruption and male joblessness, and a dearth of middleclass role models such as persons in professional and managerial occupations (e.g., Sampson & Wilson 1995; Wilson 1995; Wilson 1987). Therefore, our analysis includes tract-level measures of: (1) family disruption - the percent of families headed by females; (2) male joblessness - the percent of civilian noninstitutionalized males age 16 and older who are either unemployed or not in the labor force; and (3) occupational composition - the percent of persons age 16 and older who are employed in professional or managerial occupations. As with poverty, we operationalize each of these with a three-group categorization (i.e., two dummy variables distinguishing high and extreme from low disadvantage).

Unfortunately, past research does not indicate appropriate cut-off points for contrasting neighborhoods with distinct levels of these three aspects of disadvantage. In the absence of such standards, we take an empirical approach. Neighborhoods with family disruption or male joblessness at least one standard deviation above the mean, and where percent professionals is at least one standard deviation below the mean are regarded as extremely disadvantaged along the respective dimensions. High levels of disadvantage are defined as between the mean and one standard deviation above the mean for family disruption and male joblessness, and between the mean and one standard deviation below the mean for the percent professionals. The exact cut-off points for each variable for high and extremely disadvantaged tracts, respectively, are as follows: 25% and 42% for female-headed families, 29% and 42% for male joblessness, and 17% and 6% for professionals and managers

ADDITIONAL INDEPENDENT VARIABLES

We include two indicators of community instability that have been examined widely in prior crime studies (Crutchfield 1989; Messner & Tardiff 1986; Patterson 1991; Roncek & Lobosco 1983; Roncek & Maier 1991; Taylor & Covington 1988): (1) rental occupancy - the percent of dwelling units that are renter occupied; and (2) the vacancy rate - the percent of all dwelling units that are vacant. Finally, two control variables are included: the percent of the tract population that is male and in the crime prone ages (15-24), and the percent of the tract population that is black.

STATISTICAL ANALYSES

Our basic model examines crime as a function of social disadvantage, community instability, and control variables. Using ordinary least squares (OLS) regression, we estimate separate models of property and violent index crimes for the total sample of 177 tracts. The property but not the violent crime rate has a skewed distribution with a relatively small number of tracts having particularly high rates. Therefore, the property rate variable is transformed logarithmically.

In the estimated OLS models, we simultaneously include disadvantage along with the community instability and population control variables. A more complex model is possible. In light of arguments that the effects of disadvantage on crime are mediated by processes of social disorganization, it would be appropriate to include direct measures of this construct as endogenous variables. Along these lines, Sampson (1987) and Shihadeh and Steffensmeier (1994) have argued that family disruption serves this crucial mediating role because of its potential to affect formal and informal social control in the community. We agree that processes of social disorganization may provide the link between disadvantage and crime. However, in our view, empirically modeling family disruption in this manner places too high a demand on the data, reifying family disruption as lack of social control. In fact, the prevalence of female-headed families is an indirect measure of community control just as poverty and other disadvantages are indirect indicators of aggregate criminal motivation (the conditions that encourage crime) and social control (the conditions that discourage crime).

A few notes: all of the census tracts within the city limits meeting the researchers' criteria are used. The property crime rate but not the violent crime rate has been logged. In the table below coefficients marked with an asterisk are ones where t observed exceeded t critical at $p < .05$, two tailed. In each panel of the table the first columns show the impact of a specific aspect of disadvantage; the last column shows the impact of a disadvantage index which combines the different, separate components of disadvantage. High and extreme each refer to a dummy variable for disadvantage; the reference string or the base cell are the tracts that are low on that aspect of disadvantage.

1. In panel A, the researchers report an R squared of .38 in the last column. What is this number telling you?

2. In panel A, looking at the regression in the right-most column, the coefficient for percent rental households is significant, i.e., the unstandardized coefficient is more than two times its standard error and the t observed exceeds the t critical. Completely state the null hypothesis that this t observed is testing.

3. Based on this t test for percent rental households, will you reject the above null hypothesis, or fail to reject it?

4. Interpret the unstandardized coefficient for percent rental households in panel A in the right hand (last) column.

5. In **panel B**, looking at the right-most column reporting the results for the disadvantage index, provide an interpretation for the b weight for the percent of all dwellings that are vacant. Make this interpretation as specific as you can.

6. In **panel B**, looking at the right-most column reporting the results for the disadvantage index, the t-test for the vacancy rate variable is significant (t observed greater than t critical). What null hypothesis is being tested here?

7. Will you reject or fail to reject the null hypothesis you have just described above, given the t-test results?

8. Looking at panel B, **the left-most regression**, which looks at just the impacts of the poverty component of disadvantage, and remembering that each of the first two variables shown are dummy variables, as described above in the methods section: would you say that residents living in high poverty areas or living in extremely high poverty areas have higher violent crime rates?; why (refer to specific results)?

TABLE 2: Regression of Property and Violent Crime Rates on Discrete Measures of Disadvantage: Census Tracts in Columbus, 1990[a]

Panel A: Property Crime Rate (Ln)

Independent variables	Poverty	Male Jobless	Female Headed	Profes-sional	Dis-advantage
High	.2101*	.3988*	.2956*	.2019*	.2237*
	(.1036)	(.0811)	(.0971)	(.0793)	(.0977)
Extreme	.2486*	.3823*	.3214*	.4203*	.2932*
	(.1349)	(.1074)	(.1517)	(.1146)	(.1420)
Vacancy rate	.0145	.0151*	.0190*	.0160*	.0142
	(.0096)	(.0090)	(.0096)	(.0093)	(.0098)
Percent renters	.0082*	.0115*	.0067*	.0097*	.0083*
	(.0019)	(.0019)	(.0020)	(.0019)	(.0019)
Percent black	.0013	-.0007	-.0004	.0004	.0004
	(.0014)	(.0014)	(.0017)	(.0014)	(.0016)
Percent young males	.0004	-.0036	.0075	.0024	-.0008
	(.0067)	(.0061)	(.0062)	(.0063)	(.0068)
Constant	3.7618	3.5579	3.7340	3.5917	3.7650
R^2	.3800	.4471	.3947	.4117	.3838

Panel B: Violent Crime Rates

Independent variables	Poverty	Male Jobless	Female Headed	Profes-sional	Dis-advantage
High	5.6558*	7.4700*	3.5473*	3.9074*	3.5587*
	(1.4844)	(1.3041)	(1.5552)	(1.2939)	(1.4625)
Extreme	17.3926*	11.4395*	14.2100*	10.8941*	16.2633*
	(1.9328)	(1.7277)	(2.4301)	(1.8702)	(2.1248)
Vacancy rate	.7463*	.8233*	.7662*	.8713*	.6191*
	(.1376)	(.1447)	(.1533)	(.1512)	(.1470)
Percent renters	.0358	.1219*	.0201	.0805*	.0569*
	(.0272)	(.0300)	(.0316)	(.0304)	(.0283)
Percent black	.0938*	.0815*	.0719*	.1193*	.0697*
	(.0200)	(.0223)	(.0269)	(.0224)	(.0233)
Percent young males	-.3265*	-.2059*	.0495	-.0553	-.3148*
	(.0957)	(.0985)	(.1000)	(.1022)	(.1011)
Constant	1.6182	-5.1125	-.2570	-4.6498	1.6916
R^2	.7199	.6852	.6584	.6555	.6965

[a] Entries are unstandardized coefficients with standard errors in parentheses.

* $p < .05$

Here is some background information about a multiple regression from the ecological data file we have been using all semester. The regression contains just regional dummy variables. The dependent variable is the imprisonment rate in 1985 per 100,000 residents.

Note that each region except for the Midwest scores 1 on one and only one of the dummies. The Midwest scores 0 on all three dummies. It is the reference category. This is appropriate, we could argue conceptually, since the midwest is one of the safest regions of the country, and it also represents the "heartland." The terms omitted category, base cell, and reference category or reference string are equivalent.

REGION	REGION$	NEAST	SOUTH	WEST
1	Northeast	1	0	0
2	Midwest	0	0	0
3	South	0	1	0
4	West	0	0	1

Note that each region except for the Midwest scores 1 on one and only one of the dummies. The Midwest scores 0 on all three dummies. It is the reference category. This is appropriate, we could argue conceptually, since the midwest is one of the safest regions of the country, and it also represents the "heartland." The terms omitted category, base cell, and reference category or reference string are equivalent.

DEP VAR:PRISRA85 N: 50 MULTIPLE R: .581 SQUARED MULTIPLE R: .338

VARIABLE	COEFFICIENT	STD ERROR	STD COEF	TOLERANCE	T	P(2 TAIL)
CONSTANT	1.271	0.192	0.000	.	6.625	0.000
SOUTH	1.022	0.254	0.608	0.6302521	4.025	0.000
NEDUM	-0.100	0.293	-0.049	0.6968641	-0.340	0.735
WESTDUM	0.478	0.266	0.268	0.6486486	1.797	0.079

9 Interpet the coefficient for the constant.

10 Interpret the b weight for south.

PART 2

11 What is the definition of a residual?

12 How do you interpret the adjusted R squared term in a multiple regression?

13 When we say that OLS regression assumes that the error terms in are i.i.d., as Hamilton describes it, what are we assuming?

PART 3

14 Best in his book Damned lies and statistics talks about how statistics can become mangled. Can you give an example and describe the process whereby a social problem statistic becomes mangled.