

Department of Sociology
American University
Comprehensive Exam in Social Research Methods
August 18, 2000

This exam has four parts. You must answer one question in each part for a total of four questions. Your answers should: (1) demonstrate the full range and depth of your understanding of the relevant research techniques, (2) refer to appropriate academic studies and or research traditions, and (3) minimize the degree of repetition in your responses. Be sure to budget your time; please *double-space* the document or write on every *other* line of the paper.

Part I

1. Imagine you are advising the government minister in charge of social development in a poor country of about 30 million people. The minister needs to know: To what extent does the education of girls (through primary school) "pay off" in increases in *societal members' well-being*? You must provide quantitative evidence, and there are resources to do either (1) a survey or (2) compile and analyze existing statistics but not both.

In order to decide which to use, address the following issues:

- a. Discuss the relative merits of each strategy for answering the question.
- b. If you opt for a survey, who would you survey and what information would you ask for on the survey?
- c. If you opt to compile and analyze existing statistics, what sources of data would you tap, and what kinds of variables would be appropriate indicators of well-being?
- d. Discuss the limitations associated with the evidence produced by each strategy.

2. A national youth organization runs a large summer camp where groups of boys and girls come to camp and hike. The camp philosophy is dedicated to developing leadership skills of the boys and girls. The participants are age 14-16, and they come in mixed-gender groups of 8-12 boys and girls. In this setting, without going into details about specific substantive hypotheses or theories, describe at least 2 different ways that you could collect data on leadership skills. Compare the strengths and weaknesses of each method of data collection. From the following list discuss the 4-5 issues that best illuminate the differences between the techniques you have chosen: resources needed for collection and analysis, validity, reliability, generalizability to other populations, precision in control and measurement of variables, realism of description, prior research, and the goals of the study. Your main discussion should focus on tradeoffs: what do you gain and lose with different techniques.

Part II

1. What problems of data quality might arise when investigating the topics listed below? Choose **three (3)** and discuss the problems involved. (You may assume a U.S. context for the research.) What can researchers do to minimize the data quality problems in each of the topics you discuss?

- a. incidence of child abuse
- b. attitudes toward gun control
- c. practicing "safe sex"
- d. holding racist stereotypes
- e. quality of primary education

2. A goal of all research is accurate data collection. Researchers have developed many methods to improve data accuracy. Discuss at least 5 of the issues involved when researcher try to avoid bias and be systematic: forced-choice questions, choice of interviewer(s), choice of informants, open-ended questions, question wording, statements to participants about the research, question order, the interview guide, monetary incentives, obtrusive vs. unobtrusive research, or other issues you think are important.

Part III.

1. Compare and contrast the statements below with regard to their implications for the enterprise of social research.

- a. "We cannot discover what is meaningful to us by means of a 'presuppositionless' investigation of empirical data."
- b. "Statements [about the social world] that are true are those that correspond to objectively verified events and conditions 'out there'."

2. Durkheim argued that, "far from being the product of the individual's ideas of will, opinion or caprice, social phenomena have a constraining influence upon the individual and even on the aggregate of these individuals." Explain what this means to you as a sociologist and develop a research question to examine how social constraints effect individuals' attitudes and behavior.

Part IV

1. Table 2 (reproduced below) is from an article entitled "Bad Jobs in America" (Kalleberg et al. *ASR* 65 (2): 260). Interpret the Table and address the following questions:

- Which categories of workers are more likely to have jobs characterized as "bad", that is, with low wages and/or lack of benefits? Refer to figures in the table.
- What does it mean to say that the differences between the genders and between racial/ethnic groups are significant?
- Which has the greater effect on bad jobs -- gender or race/ethnicity? Justify your answer based on information in the table.
- What other information would you want in order to draw firm conclusions about these findings?

260 AMERICAN SOCIOLOGICAL REVIEW

Table 2. Characteristics of Bad Jobs, by Sex and Race/Ethnicity: Current Population Survey, February 1995

Racial/ Ethnic Group ^a	Percent Low Wages			Percent No Health Insurance			Percent No Pension			Mean Number of Bad Job Characteristics		
	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
White	13.5 †††	23.4	18.1	36.3 †††	49.3	42.3	46.3 †††	52.0	48.9	.95 †††	1.22	1.07
Black	23.3 †††	34.1	29.0	43.5	45.6	44.6	50.0	50.4	50.2	1.25	1.30	1.27
Hispanic	29.3 ††	36.1	32.1	53.8	54.7	54.2	70.4 ††	66.5	66.8	1.47	1.56	1.51
Other	21.5	24.2	22.8	40.1 †††	51.5	45.4	54.1 ††	60.8	57.2	1.21	1.29	1.25
Total	16.1 †††	25.7	20.6	38.7 †††	49.3	43.6	49.1 †††	53.1	51.0	1.03 †††	1.26	1.14
Number of cases	11,607	9,792		28,973	26,411		28,522	26,060		10,982	9,317	

^a All racial/ethnic group comparisons are significantly different at the $p < .001$ level (***) except for the following: *Percent low wages*: black vs. Hispanic (**), black vs. other, and Hispanic vs. other (**) for males; white vs. other and black vs. Hispanic for females; white vs. other (**), black vs. Hispanic (*), and black vs. other (**) for total. *Percent no health insurance*: white vs. other (*) and black vs. other for males; white vs. other, black vs. other (**), and Hispanic vs. other for females; white vs. black (**), white vs. other (*), and black vs. other for total. *Percent no pension*: black vs. other (*) for males; white vs. black and Hispanic vs. other (**) for females; whites vs. blacks for total. *Mean number of bad job characteristics*: black vs. other for males; white vs. black (*), white vs. other, black vs. other, and Hispanic vs. other (*) for females; black vs. other for total.

† $p < .05$ †† $p < .01$ ††† $p < .001$ (two-tailed tests for sex differences)

* $p < .05$ ** $p < .01$ *** $p < .001$ (two-tailed tests for racial/ethnic group differences)

2. Teachers' pay is influenced by many factors. Below is output from analysis of the 50 states in the U.S. plus DC; assume from prior work with these data that you know case 50 is Alaska, a potential outlier. The variables used in the analysis are:

Spend: spending per pupil, \$s

Inc: median personal income, 100s of \$s

AK: a dummy variable with values of 1 for Alaska and 0 otherwise

4 dummy variables for region of country: Neast, Ncent, South, and West

Based on the output in the figures below, respond to the following:

- a. What do the preliminary plots in Figure 1 tell you? How does this information influence the model?
- b. Interpret the regression output, including the coefficients and measures of fit of the model.
- c. Discuss the collinearity diagnostics.
- d. What does the residuals plot tell you?
- e. What additional test involving dummy variables do you need in order to draw firm conclusions from these results? Describe the test and the information it would give you.

Figure 1

Preliminary plots

(Note that smoothers are LOWESS smoothers)

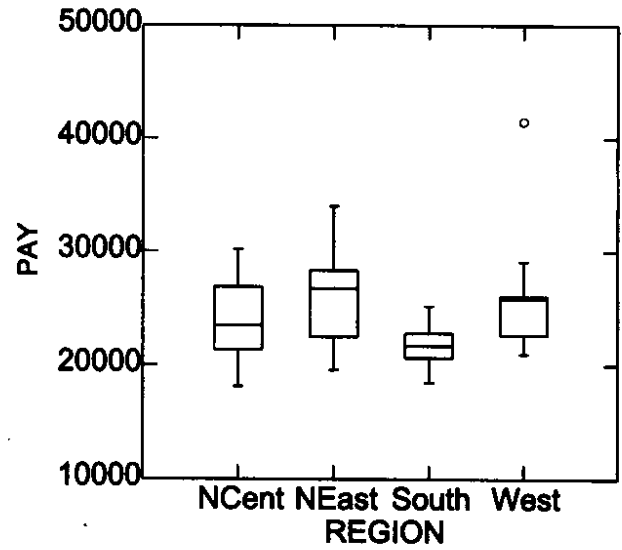
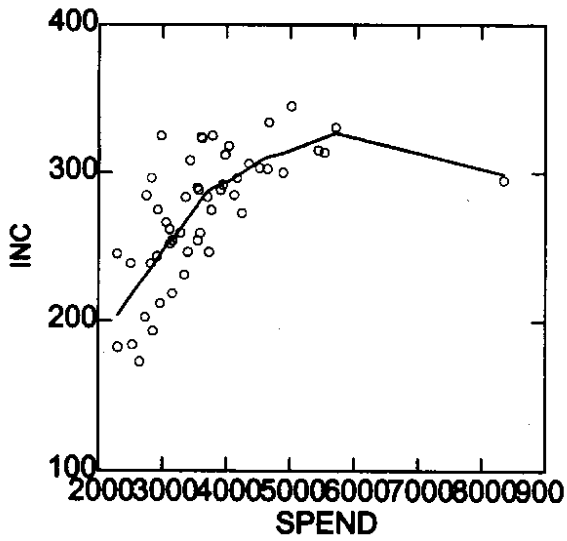
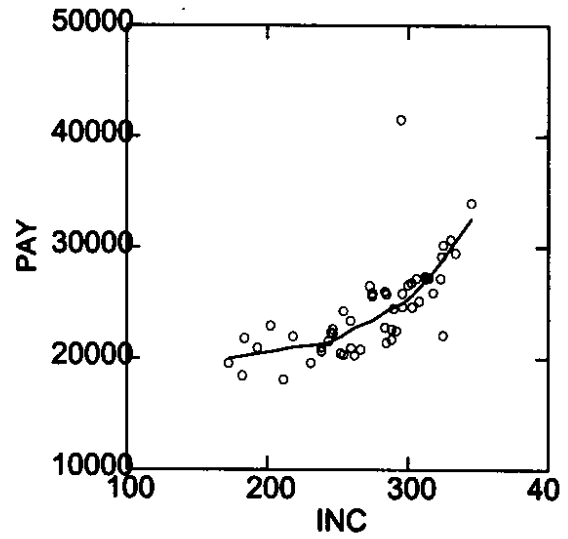
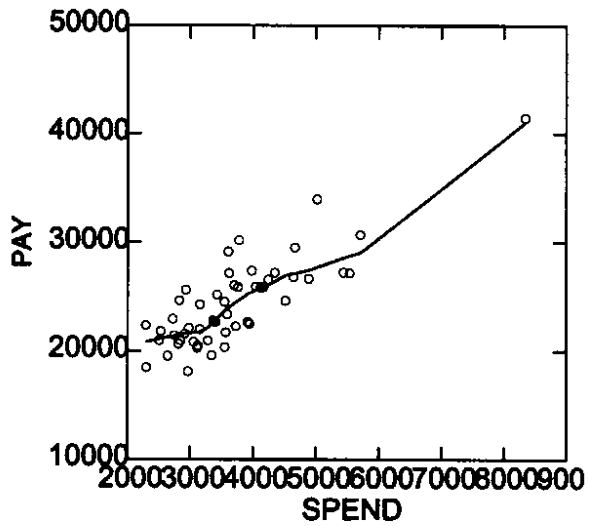


Figure 2

```
>model pay = constant + spend + inc + ncent + south + west + ak
>save resid / data residuals
>estimate
```

Eigenvalues of unit scaled x'x

	1	2	3	4	5
	3.774	1.206	1.002	0.762	0.239
	6	7			
	0.012	0.005			

Condition indices

	1	2	3	4	5
	1.000	1.769	1.941	2.226	3.971
	6	7			
	13.289	33.142			

Variance proportions

	1	2	3	4	5
CONSTANT	0.001	0.000	0.000	0.000	0.001
SPEND	0.001	0.000	0.000	0.000	0.011
INC	0.001	0.000	0.000	0.000	0.003
NCENT	0.007	0.037	0.193	0.034	0.391
SOUTH	0.005	0.028	0.132	0.024	0.245
WEST	0.008	0.087	0.000	0.154	0.412
AK	0.002	0.126	0.000	0.235	0.003

	6	7
CONSTANT	0.611	0.387
SPEND	0.696	0.291
INC	0.011	0.985
NCENT	0.337	0.001
SOUTH	0.547	0.019
WEST	0.330	0.009
AK	0.468	0.166

Dep Var: PAY N: 51 Multiple R: 0.895 Squared multiple R: 0.801

Adjusted squared multiple R: 0.773 standard error of estimate: 1989.308

Effect	Coefficient	Std Error	Std Coef Tolerance	t	P(2 Tail)
CONSTANT	6334.806	2483.073	0.000	2.551	0.014
SPEND	1.773	0.575	0.448	3.084	0.004
INC	39.591	9.741	0.403	4.064	0.000
NCENT	345.354	923.102	0.035	0.374	0.710
SOUTH	498.574	1029.690	0.054	0.484	0.631
WEST	1057.294	926.852	0.111	1.979	0.048
AK	7609.526	3393.940	0.255	2.242	0.030

Effect	Coefficient	Lower	< 95%>	Upper
CONSTANT	6334.806	1330.502		11339.110
SPEND	1.773	0.614		2.933
INC	39.591	19.960		59.222
NCENT	345.354	-1515.035		2205.743
SOUTH	498.574	-1576.630		2573.778
WEST	1057.294	-810.653		2925.241
AK	7609.526	769.489		14449.563

Correlation matrix of regression coefficients

	CONSTANT	SPEND	INC	NCENT	SOUTH
CONSTANT	1.000				
SPEND	-0.313	1.000			
INC	-0.533	-0.616	1.000		
NCENT	-0.452	0.437	-0.123	1.000	
SOUTH	-0.674	0.492	0.032	0.641	1.000
WEST	-0.406	0.465	-0.186	0.612	0.628
AK	0.273	-0.792	0.466	-0.351	-0.401

Correlation matrix of regression coefficients (continued)

	WEST	AK
WEST	1.000	
AK	-0.476	1.000

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	6.99257E+08	6	1.16543E+08	29.450	0.000
Residual	1.74123E+08	44	3957348.005		

*** WARNING ***

Case 50 has large leverage (Leverage = 1.000)

Durbin-Watson D Statistic 1.280

First Order Autocorrelation 0.348

Figure 3

Plot of Residuals against Predicted Values

