Sociology 592 - Research Statistics I
Final Exam
December 16, 2003

Where appropriate, show your work - partial credit may be given. (On the other hand, don't waste a lot of time on excess verbiage.) Do not spend too much time on any one problem. You are free to refer to anything that was demonstrated in the homework or handouts.

1. (5 points each, 20 points total). For each of the following, indicate whether the statement is true or false. If you think the statement is false, indicate how the statement could be corrected.

NOTE: These are all pretty easy, but you could waste a great deal of time on some of them or make stupid mistakes if you don't happen to see what the easiest way to approach each problem is.

a. The correlation between Y and X1 is positive. Therefore, if Y is regressed on X1, X2, and X3, the unstandardized coefficient for X1 will also always be positive (or else zero).

b. A researcher has estimated the following two models:

$$\text{Model 1: } \hat{y} = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$
$$\text{Model 2: } \hat{y} = a + b_1X_1 + b_2X_2 + b_5X_5$$

Assuming that the same cases are analyzed in both models, the R^2 for Model 1 will be at least as large as the R^2 for Model 2.

c. A researcher obtains the following results:

```
------------------------------ Variables in the Equation -------------------------------

Variable             B        SE B       Beta   Correl Part Cor  Partial         T  Sig T

STATUS        8.351967    1.447318    .278956  .400000  .222883  .250827     5.771  .0000
MALE         10.698092    1.450267    .286709  .330000  .284911  .314422     7.377  .0000
ANOMIE        1.282964     .402837    .153956  .350000  .123009  .141563     3.185  .0015
(Constant)   37.213275    3.878287                                           9.595  .0000
```

If she is using backwards stepwise regression with $\alpha = .05$, on the next step R^2 will decline by $.123009^2$.

d. In a bivariate regression with N = 500, the null and alternative hypotheses are

H_0: $\beta = 0$
H_A: $\beta > 0$

It is found that b = 37 and F = 36. The null hypothesis should be rejected.

2. Short answer problems. (10 points each, 30 points total, up to 5 points extra credit). Answer <u>three</u> of the following. You will get up to five points extra credit if you can solve all four problems.

a. In a multivariate regression, n = 150, k = 12, R^2 = .25, SSE = 144. Construct the ANOVA table.

b. Y = income (in thousands of dollars). X1 = 1 if black, 0 otherwise. X2 = 1 if female, 0 otherwise. X3 = X1 * X2. (NOTE: X3 is referred to as an <u>interaction</u> term.) Suppose a = 18, b1 = -4, b2 = -3, b3 = -2. What are the average incomes for black males, black females, nonblack males, and nonblack females?

c. When Y is regressed on X1 and X2, r_{y1} = .5, r_{y2} = .5, TOL_{X1} = .75. X1 and X2 are positively correlated. Compute the semipartial correlations.

d. A researcher obtains the following results. Are there any logical inconsistencies in the results of the significance tests? If so, discuss what might be responsible. Cite evidence from the printouts to support your points.

Regression

Correlations

		Y	X1	X2
Pearson Correlation	Y	1.000	.240	.250
	X1	.240	1.000	.950
	X2	.250	.950	1.000

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.194	2	3.097	3.237	.044[a]
	Residual	92.806	97	.957		
	Total	99.000	99			

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.0000	.0978		.0000	1.0000		
	X1	.0256	.3148	.0256	.0814	.9353	.0975	10.2564
	X2	.2256	.3148	.2256	.7167	.4753	.0975	10.2564

a. Dependent Variable: Y

3. (50 points total; up to 5 points extra credit.) Credit Scores are increasingly being used in the lending industry. These scores supposedly measure how credit-worthy a person is, by taking into account such things as their income, net worth, how well they have handled credit in the past, the stability of their employment record, etc. Proponents of these scores argue that they eliminate racial discrimination in lending: loan applicants are judged only on their credit-worthiness and not on their race. Others, however, express concern because the formulas used to compute credit scores are not public knowledge; therefore it could be that the scores themselves contain a hidden racial bias.

A researcher has therefore collected data from a random sample of 1,605 recent loan applicants at a major bank. Her variables are as follows:

Variable	Description
Black	Coded 1 if the applicant is black, 0 otherwise
Applinc	Applicant Income, in thousands of dollars
Networth	Applicants net worth (assets – debts), measured in thousands of dollars. When debts exceed assets, the value is negative.
CrScore	The applicants credit score, which can range from a low of 0 to a high of 1000. Higher scores indicate the applicant is more credit-worthy.

She obtains the following results:

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
CRSCORE Credit Score	455.18	104.965	1605
BLACK Is Applicant black?	.0829	.27577	1605
APPLINC Applicant income	$71.01	$46.281	1605
NETWORTH Net worth, in thousands of dollars	$33.73	$22.785	1605

Correlations

		CRSCORE Credit Score	BLACK Is Applicant black?	APPLINC Applicant income	NETWORTH Net worth, in thousands of dollars
Pearson Correlation	CRSCORE Credit Score	1.000	-.198	.620	.507
	BLACK Is Applicant black?	-.198	1.000	-.123	-.447
	APPLINC Applicant income	.620	-.123	1.000	.653
	NETWORTH Net worth, in thousands of dollars	.507	-.447	.653	1.000

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.198[a]	.039	.039	102.919
2	[b]	**[1]**		80.893

a. Predictors: (Constant), BLACK Is Applicant black?

b. Predictors: (Constant), BLACK Is Applicant black?, APPLINC Applicant income, NETWORTH Net worth, in thousands of dollars

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	692774.580	1	692774.580	65.404	.000[a]
	Residual	16979434.842	1603	10592.286		
	Total	17672209.422	1604			
2	Regression	7195855.943	3	2398618.648	366.558	.000[b]
	Residual	10476353.479	1601	6543.631		
	Total	17672209.422	1604			

a. Predictors: (Constant), BLACK Is Applicant black?

b. Predictors: (Constant), BLACK Is Applicant black?, APPLINC Applicant income, NETWORTH Net worth, in thousands of dollars

c. Dependent Variable: CRSCORE Credit Score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	461.421	2.683		172.011	.000	456.160	466.683					
	BLACK Is Applicant black?	-75.362	9.319	**[2]**	-8.087	.000	-93.640	-57.084	-.198	-.198	-.198	1.000	1.000
2	(Constant)	352.655	4.331		81.432	.000	344.160	361.149					
	BLACK Is Applicant black?	-28.715	**[3]**	-.075	-3.395	.001	-45.303	**[4]**	-.198	-.085	-.065	.750	1.333
	APPLINC Applicant income	**[5]**	.059	.527	20.089	.000	1.078	1.312	.620	.449	.387	.538	1.858
	NETWORTH Net worth, in thousands of dollars	.594	.134	.129	4.433	.000	.331	.857	**[6]**	.110	.085	.437	**[7]**

a. Dependent Variable: CRSCORE Credit Score

a. (21 points) Fill in the missing items [1] – [7].

b. (4 points) Two models are estimated. Do you think the researcher used hierarchical model selection or stepwise selection? Point to evidence from the printout to support your argument.

c. (5 points) Do an incremental F test of the hypothesis

H_0: $\beta_{Applinc} = \beta_{Networth} = 0$

d. (5 points) When Applinc and Networth are added to the model, the effect of Black becomes much smaller. Offer a substantive explanation as to why this might occur. What evidence from the printout might support your argument?

e. (15 points) Interpret the results. Be sure to answer the following questions. Explain what information from the printout supports your conclusions.

1. What proportion of the sample is black?

2. What is the mean credit score for whites? What is the mean credit score for blacks?

3. What would you say is the most important determinant of credit scores? Cite at least two pieces of evidence from the printout to support your argument.

4. Is there evidence to suggest that there may be racial bias in the credit scores? If so, what is that evidence?

5. If you were an industry representative who had to defend credit scores against charges of racial bias, how might you respond to whatever arguments you raised in pt. 4? That is, how would you argue that the analysis has not been done correctly or else that alternative explanations of the results are possible?

f. (5 points extra credit) Suppose you wanted to test

H_0: $\beta_{Applinc} = \beta_{Networth}$

i.e. income and wealth have equal effects on credit scores. How would you go about doing it? [Hint: You could do it with an incremental F test. The unconstrained model is Model 2 above. What would the constrained model be? In practice, how would you estimate the constrained model, using SPSS?] You don't have to give me a value for the test statistic, just describe what additional analyses you would run and how you would then compute the value of the incremental F statistic.