Sociology 593
Exam 1
February 14, 1997

I. True-False. (15 points) Indicate whether the following statements are true or false. If false, briefly explain why.

1. There are 5 IVs in a multiple regression model. If the global F statistic is significant, this means that all IVs affect the DV.

2. If the Goldfield-Quant statistic is not significant, this means that heteroscedasticity is not present in the data.

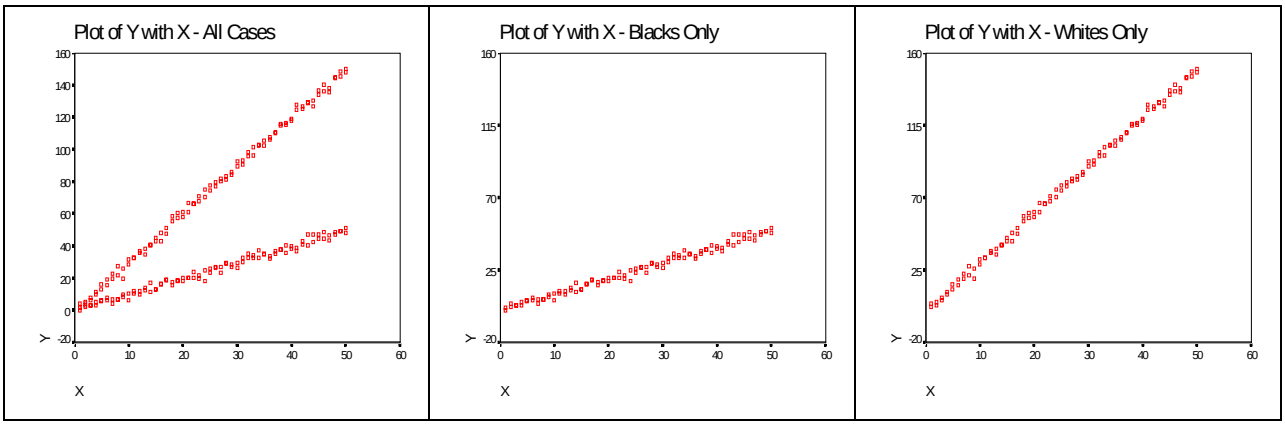
3. If you want to make comparisons across groups, standardized coefficients should be used.

4. Skip patterns are a possible reason for missing data in a survey.

5. A good measuring instrument will always have the same reliability, regardless of what population it is tested on.

II. Short answer. (60 points; up to 10 points extra credit). For <u>three</u> of the following, indicate (i) what problem appears to be present (and how you can tell that from the printout) (ii) why you should be concerned about the problem, i.e. what harmful effects might it have when estimating regression models, and (iii) possible solutions. When discussing solutions, be sure to look carefully at the information presented; if, in this particular case, some solutions appear to be better than others, explain why. (20 points each; up to 10 points extra credit if you get all 4 right.)

a)



b)

Descriptive Statistics

	Mean	Std. Deviation	N
INCOME	42.0000	10.00000	200
EDUC	11.0000	2.000000	200
IQ	102.000	15.00000	200
OCCUP	75.0000	11.00000	200

Correlations

		INCOME	EDUC	IQ	OCCUP
Pearson Correlation	INCOME	1.000	.330	.340	.360
	EDUC	.330	1.000	.810	.850
	IQ	.340	.810	1.000	.900
	OCCUP	.360	.850	.900	1.000

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2636.31	3	878.771	9.977	.000[b]
	Residual	17263.7	196	88.080		
	Total	19900.0	199			

a. Dependent Variable: INCOME

b. Independent Variables: (Constant), OCCUP, EDUC, IQ

Coefficients[a]

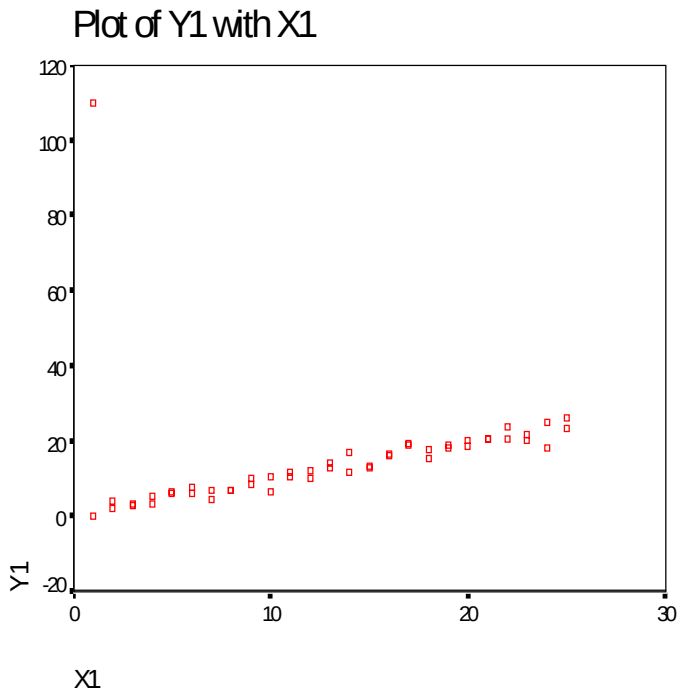
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	17.239	4.731		3.644	.000		
	EDUC	.379	.644	.076	.588	.557	.267	3.748
	IQ	4.4E-02	.104	.066	.426	.671	.183	5.473
	OCCUP	.215	.158	.236	1.362	.175	.147	6.783

a. Dependent Variable: INCOME

c)

Income	*Race*	*Gender*
24	Black	Missing
26	Black	Missing
26	White	Missing
27	White	Missing
32	Black	Male
34	White	Male
35	Black	Female
36	White	Female
37	Black	Female
40	White	Male

d)



[CONTINUED]

Extreme Values

			Case Number	Value
Y1	Highest	1	1	110.00
		2	50	26.28
		3	47	25.12
		4	44	24.19
		5	49	23.48
	Lowest	1	2	.40
		2	4	2.42
		3	5	3.33
		4	7	3.58
		5	6	3.64
X1	Highest	1	50	25.00
		2	49	25.00
		3	47	24.00
		4	48	24.00
		5	46	.[a]
	Lowest	1	1	1.00
		2	2	1.00
		3	4	2.00
		4	3	2.00
		5	6	.[b]

a. Only a partial list of cases with the value 23 are shown in the table of upper extremes.

b. Only a partial list of cases with the value 3 are shown in the table of lower extremes.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.449	4.407		2.144	.037
	X1	.439	.296	.209	1.481	.145

a. Dependent Variable: Y1

Casewise Diagnostics[a]

Case Number	Std. Residual	Y1
1	6.623	110.00

a. Dependent Variable: Y1

III. Computations. (25 points)

a. (5 points) Compute the GQ statistic. Based on the GQ statistic, would you recommend using WLS with this data set?

Model Summary[a,b,c]

Model	Variables Entered	Variables Removed	R XVAR3 <= 26.48 (Selected)	R Square	Adjusted R Square	Std. Error of the Estimate
1	XVAR3, XVAR1, XVAR2[d,e]	.	.755	.570	.553	17.0597

a. Unless noted otherwise, statistics are based only on cases for which XVAR3 <= 26.48.

b. Dependent Variable: YVAR

c. Method: Enter

d. Independent Variables: (Constant), XVAR3, XVAR1, XVAR2

e. All requested variables entered.

ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29300.9	3	9766.96	33.559	.000[c]
	Residual	22118.6	76	291.034		
	Total	51419.5	79			

a. Dependent Variable: YVAR

b. Selecting only cases for which XVAR3 <= 26.48

c. Independent Variables: (Constant), XVAR3, XVAR1, XVAR2

Model Summary[a,b,c]

Model	Variables Entered	Variables Removed	R XVAR3 >= 37.50 (Selected)	R Square	Adjusted R Square	Std. Error of the Estimate
1	XVAR3, XVAR2, XVAR1[d,e]	.	.540	.292	.264	19.5715

a. Unless noted otherwise, statistics are based only on cases for which XVAR3 >= 37.50.

b. Dependent Variable: YVAR

c. Method: Enter

d. Independent Variables: (Constant), XVAR3, XVAR2, XVAR1

e. All requested variables entered.

ANOVA[a,b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11996.7	3	3998.89	10.440	.000[c]
	Residual	29111.4	76	383.045		
	Total	41108.1	79			

a. Dependent Variable: YVAR

b. Selecting only cases for which XVAR3 >= 37.50

c. Independent Variables: (Constant), XVAR3, XVAR2, XVAR1

b. (10 pts.) Fill in the missing entries [1] - [5]

Descriptive Statistics

	Mean	Std. Deviation	N
YVAR	53.8161	33.1173	200
XVAR1	10.1000	5.8016	200
XVAR2	20.7608	11.4587	200
XVAR3	31.4899	18.1460	200

Model Summary[a,b]

Model	Variables Entered	Variables Removed	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	XVAR3, XVAR1, XVAR2[c,d]	.	.818	.668	.663	19.2130

a. Dependent Variable: YVAR

b. Method: Enter

c. Independent Variables: (Constant), XVAR3, XVAR1, XVAR2

d. All requested variables entered.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	**[1]**	3	48634.3	131.750	.000[b]
	Residual	72351.6	196	369.141		
	Total	218255	199			

a. Dependent Variable: YVAR

b. Independent Variables: (Constant), XVAR3, XVAR1, XVAR2

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	2.894	**[2]**		.936	.350					
	XVAR1	.299	.310	.052	.965	.336	.561	.069	.040	**[3]**	1.739
	XVAR2	.631	.180	**[4]**	3.501	.001	.700	.243	.144	.435	2.299
	XVAR3	1.105	.126	.606	**[5]**	.000	.803	.530	.360	.353	2.834

a. Dependent Variable: YVAR

c) (5 pts.) If you were using backwards stepwise regression, what would you do next? Why?

d) (5 pts.) Do an F test of the hypothesis

H_0: $\beta_1 = \beta_2 = 0$

H_A: β_1 and/or $\beta_2 \neq 0$