Sociology 593
Exam 1
February 15, 2002

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

1. The Dubin Watson statistic is used to determine what variable, if any, should be entered next in a forward stepwise regression.

2. In a bivariate regression, the F value is 81. The T statistic for the beta coefficient is therefore 9.

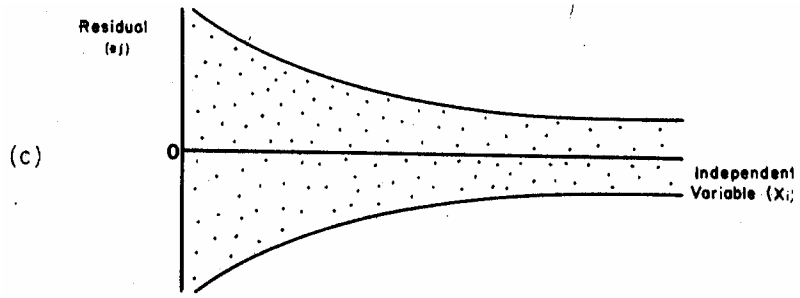
3. High multicollinearity results in biased parameter estimates and can make it more difficult to detect significant relationships.

4. A key problem with listwise deletion is that the pieces put together for the regression analysis refer to systematically different subsets of the population, e.g. the cases used in computing r_{12} may be very different than the cases used in computing r_{34}.

5. A researcher encounters an extreme outlier in her data. She should immediately delete the case and then rerun her analyses.

II. Short answer. Answer both of the following. (20 points each 40 pts total.)

1. Consider the following plot:



a. Explain how a mis-specified model could produce a pattern such as that shown above. How would you try to solve the problem in this case?

b. Explain how a model could be correctly specified and yet you still get a pattern such as the above. Under such conditions, will OLS estimates of the betas be biased? Are there other problems that OLS will have in this situation? Should OLS be used in this case, or is there a superior alternative?

2. A researcher wants to test whether the effect of attitudes on behavior is different for blacks than it is for whites; that is, she wants to test

$$H_0 : \beta^{White} = \beta^{Black}$$
$$H_A : \beta^{White} \neq \beta^{Black}$$

In reality, the null hypothesis is true: the effect of attitudes on behavior is the same for both blacks and whites. Explain how and why the following conditions might lead the researcher to reach an erroneous conclusion.

a) There is a great deal of random measurement error in the white responses, and very little error in the black responses

b) Low-income blacks are disproportionately likely to NOT answer questions on attitudes.

III. Computation and interpretation. (40 points total, 10 pts extra credit)

A researcher is interested in what determines how housework gets divided between couples. She has extracted the following variables from the 1996 General Social Survey (GSS).

Variable	Question
SPWork	How much (of the housework) does your spouse or partner do? The values are 1 "Little or None" 2 "Some" 3 "About half" 4 "Most" 5 "All".
Educ	Years of Education (ranges from 0, no formal education, to 20 years)
White	1 = white, 0 = black or other
Male	1 = Male, 0 = Female
Party	Coded on a 7 point scale where 1 = Strong Republican, 4 = Independent, 7 = Strong Democrat

She begins by running frequencies on the entire sample. She then runs a regression with SPWork as the dependent variable.

```
FREQUENCIES
  VARIABLES=spwork male educ white party
  /Format = Notable / Statistics = Default .
```

Frequencies

Statistics

		How much of the housework does your spouse or partner do?	MALE	Highest year of school completed	WHITE	PARTY
N	Valid	763	2904	2895	2904	2855
	Missing	2141	0	9	0	49
Mean		2.6920	.4425	13.36	.8089	4.1860
Std. Deviation		.95339	.49677	2.929	.39325	1.97977
Minimum		1.00	.00	0	.00	1.00
Maximum		5.00	1.00	20	1.00	7.00

```
REGRESSION
  /MISSING LISTWISE/ Descriptives DEF
  /STATISTICS COEFF OUTS R ANOVA ZPP TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT SpWork
  /METHOD=ENTER male white educ party.
```

Descriptive Statistics

	Mean	Std. Deviation	N
How much of the housework does your spouse or partner do?	2.6909	.94882	744
MALE	.4866	.50016	744
WHITE	.8629	.34418	744
Highest year of school completed	13.53	2.820	744
PARTY	3.9758	2.01928	744

Correlations

		How much of the housework does your spouse or partner do?	MALE	WHITE	Highest year of school completed	PARTY
Pearson Correlation	How much of the housework does your spouse or partner do?	1.000	.479	.047	.019	-.057
	MALE	.479	1.000	-.003	.000	-.079
	WHITE	.047	-.003	1.000	.037	-.255
	Highest year of school completed	.019	.000	.037	1.000	-.076
	PARTY	-.057	-.079	-.255	-.076	1.000

Model Summary

Model		R Square	Std. Error of the Estimate
1	a	[1]	.83366

a. Predictors: (Constant), PARTY, Highest year of school completed, MALE, WHITE

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	155.298	4	38.825	55.863	.000[a]
	Residual	513.600	739	.695		
	Total	668.898	743			

a. Predictors: (Constant), PARTY, Highest year of school completed, MALE, WHITE

b. Dependent Variable: How much of the housework does your spouse or partner do?

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	2.0758	.194		10.713	.000					
	MALE	.9081	.061	.479	14.799	.000	[2]	.478	.477	.993	1.007
	WHITE	.1279	[3]	.046	1.392	.164	.047	.051	.045	.934	1.070
	Highest year of school completed	.0055	.011	.016	.509	.611	.019	.019	.016	.994	[4]
	PARTY	-.0030	.016	[5]	-.192	.848	-.057	-.007	-.006	.924	1.082

a. Dependent Variable: How much of the housework does your spouse or partner do?

Casewise Diagnostics[a]

Case Number	Std. Residual	How much of the housework does your spouse or partner do?	Predicted Value	Residual	DFBETA (Constant)	DFBETA PARTY	DFBETA Highest year of school completed	DFBETA MALE	DFBETA WHITE
62	3.417	5.00	2.1518	2.8482	.0031	.0008	.0019	-.0074	-.0277
176	3.268	5.00	2.2752	2.7248	.0121	-.0019	.0001	-.0078	.0014
1800	3.447	5.00	2.1266	2.8734	.0242	.0016	-.0001	-.0071	-.0261
2518	3.289	5.00	2.2580	2.7420	.0132	.0000	-.0007	-.0072	.0045

a. Dependent Variable: How much of the housework does your spouse or partner do?

a) (10 pts) The researcher is very puzzled by her sample size. As her frequencies show, there are almost 3,000 cases in the 1996 GSS. But, there are only 744 cases in her regression analysis. Explain to the researcher why so many cases are missing. Cite evidence from the printout to support your answer. Other than the reduced sample size, would you yourself be very worried about the missing data, e.g. do you think your results will be seriously biased because of

the MD? Are there any additional simple computer runs you would recommend for checking whether MD may be a problem?

b) (10 pts) Fill in the missing quantities [1] – [5].

c) (10 pts) If you were the researcher, would you be worried about multicollinearity? Why or why not? Is there reason to be greatly concerned about outliers in the data? Assuming the outliers are not coding errors, what do you think would be the better strategy – toss the outliers out, or try to add other explanatory variables to the model?

d) (10 pts) Interpret the results. About how even is the (perceived) housework split – do respondents think they are doing most of the work, do they think their spuse/partner is doing most of the work, or do they think the split is about equal? The researcher chose these variables because she thought a person's gender, race, education and party identification would all affect how much housework their spouse did. To what extent is she right, and to what extent is she wrong? Cite evidence from the printout to support your answer.

e) (10 pts extra credit) A colleague is very critical of the researcher's model. She agrees with the choice of gender as an IV, but argues that, at least as the model is set up, it makes little theoretical sense to expect education, race, and party id to have much of an effect. What do you think she is basing her argument on? [HINT: For each of the IVs in the model, do you think spouses will tend to have similar values, or different values? Do you think respondents' characteristics unilaterally determine how much housework their spouse/partner does?]