Sociology 592 - Research Statistics I
Final Exam
December 10, 1993

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. One of the problems with using listwise deletion of missing data is that it can produce a correlation matrix that would be impossible with any complete set of data.

b. SPSS produced the following computer printout. If $\alpha = .05$ and backwards stepwise regression is being used, then EDUC should be removed from the equation next.

```
Multiple R           .50998          Analysis of Variance
R Square             .26008                          DF      Sum of Squares      Mean Square
Adjusted R Square    .25561          Regression       3         29200.56017       9733.52006
Standard Error     12.94175          Residual       496         83074.43983        167.48879


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

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

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

c. A sociologist believes that women are more religious than are men (where religiosity is measured on a scale that runs from a low of zero to a high of one hundred). In her sample of 300 subjects, when she regresses Religiosity (Y) on Gender (X, where 1 = Female, 0 = Male), she finds that b = 10, F = 25, the mean of Y is 50 and the mean of X is 0.5. Therefore, she should reject the null hypothesis.

d. A researcher is interested in the relationship between Income (measured in dollars) and race (where 1 = White, 0 = NonWhite). Regression, Anova, and a T-Test are all possible means for addressing this issue.

2. Short answer problems. (10 points each, 30 points total, up to 10 points extra credit) Answer three of the following. You will get up to five points extra credit for each additional problem you answer correctly.

a. When Y is regressed on X1 and X2, $r_{y1} = .6$, $r_{y2} = .7$, $TOL_{X1} = .75$. Compute the semipartial and partial correlations.

b. In a multiple regression, N = 160, K = 9, F = 5, SSE = 600. Construct the ANOVA table.

c. A sociologist has collected data from 100 respondents. When she regresses Y on X1 and X2, she gets F = 97. When she adds X3 to the model, $R^2 = .75$. Test

H_0: $\beta_3 = 0$
H_A: $\beta_3 <> 0$.

d. The dean of the college wants to see what relationship, if any, there is between the scholarly output of a faculty member (measured on a scale that runs from 0 to 100) and the religion and gender of the faculty member. Hence, information is collected from 25 Catholic males, 25 Catholic females, 25 NonCatholic males, and 25 NonCatholic females on the following variables:

Y = scholarly output
X1 = 1 if Catholic, -1 otherwise
X2 = 1 if male, -1 otherwise
X3 = X1 * X2 (i.e. X3 is an interaction term)

A regression analysis yields a = 60, b1 = -10, b2 = -10, b3 = 10. Compute the average scores for Catholic males, Catholic females, NonCatholic males, and NonCatholic females. Who is the most "scholarly" of these four groups?

e. A political scientist has collected data from voters concerning their opinion of Ross Perot. Her variables are Y = Level of support for Perot (measured on a scale that runs from a low of zero to a high of 100), X1 = Previously supported Perot (where 1 = voted for Perot in 1992, 0 = did not vote for Perot), and X2 = Gender (1 = Female, 0 = Male). She finds the following:

```
            Mean  Label

Y         35.000
X1          .200
X2          .500

N of Cases =   500

Covariance:

                  Y         X1         X2

Y           100.000      3.000     -1.000
X1            3.000       .160      0.000
X2           -1.000      0.000       .250
```

(a) Compute the standardized coefficients b1' and b2'.
(b) The researcher believes that women are more supportive of Perot than are men. Is the evidence strong enough to support her case?

3. After his exciting come from behind win on the NAFTA vote, Bill Clinton wants to know where he stands with the American public. He has told his pollsters to go out and collect information on four key variables:

Clinton	Level of support for Bill Clinton, where 0 = strongly opposes and 100 = strongly supports.
Female	Coded 1 if the respondent is female, 0 if male
Nafta	Coded 1 if the respondent favored NAFTA, 0 otherwise
Union	Coded 1 if the respondent is a union member, 0 otherwise.

Using *forward* stepwise regression, the *first* and *only* model obtained was as follows:

MODEL I: FORWARD STEPWISE REGRESSION

```
            Mean  Std Dev  Label

CLINTON   47.367   15.426
FEMALE      .500     .509
NAFTA       .500     .509
UNION       .367     .490

N of Cases =    30
```

Correlation:

	CLINTON	FEMALE	NAFTA	UNION
CLINTON	1.000	.705	.305	.159
FEMALE	.705	1.000	.200	-.069
NAFTA	.305	.200	1.000	-.623
UNION	.159	-.069	-.623	1.000

* * * * M U L T I P L E R E G R E S S I O N * * * *

Equation Number 1 Dependent Variable.. CLINTON

Block Number 1. Method: Forward Criterion PIN .0500 FEMALE NAFTA UNION

Variable(s) Entered on Step Number 1.. FEMALE

Multiple R	.70549	Analysis of Variance			
R Square	.49771		DF	Sum of Squares	Mean Square
Adjusted R Square	.47977	Regression	1	3434.70008	3434.70008
Standard Error	11.12633	Residual	28	3466.26656	123.79523

F = 27.74501 Signif F = .0000

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

Variable	B	SE B	Beta	T	Sig T
FEMALE	21.400001	4.062762	.705488	5.267	.0000
(Constant)	36.666666	2.872806		12.763	.0000

---------- Variables not in the Equation ----------

Variable	Beta In	Partial	Min Toler	T	Sig T
NAFTA	.171244	.236742	.960000	1.266	.2163
UNION	.209274	.294576	.995215	1.602	.1209

End Block Number 1 PIN = .050 Limits reached.

a. (15 points) Interpret the results from model I. How many of the sample members supported NAFTA? How many belong to a union? Who supports Clinton more, men or women, and by how much? Did union members tend to support or oppose NAFTA? Why was FEMALE the first and only variable to enter into the forward stepwise regression?

Using *backwards* stepwise regression, the *first* and *only* model obtained was as follows:

MODEL II: BACKWARDS STEPWISE REGRESSION

Multiple R	.82835	Analysis of Variance			
R Square	.68616		DF	Sum of Squares	Mean Square
Adjusted R Square	.64995	Regression	3	4735.19331	1578.39777
Standard Error	9.12683	Residual	26	2165.77333	83.29897

F = 18.94859 Signif F = .0000

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

Variable	B	SE B	Beta	Correl	Part Cor	Partial	Tolerance	VIF	T
FEMALE	(1)	3.410260	.641671	.705488	(2)	.745743	.955000	1.047	5.708
NAFTA	15.060209	4.347246	.496486	.305492	.380611	.561975	(3)	1.702	3.464
UNION	16.143976	4.430058	(4)	.159473	.400374	.581453	.609250	1.641	3.644
(Constant)	24.184991	3.943977	(5)						6.132

------ in -------

Variable	Sig T
FEMALE	.0000
NAFTA	.0019
UNION	.0012
(Constant)	.0000

End Block Number 2 POUT = .100 Limits reached.
No variables removed for this block.

b. (20 points) Fill in the missing items (1) - (5).

c. (10 points) Do an F test of the hypothesis H_0: $ß_{Nafta} = ß_{Union} = 0$.

d. (5 points) In the forward stepwise regression, only Female made it into the final model. Yet, in the backwards stepwise regression, all three variables made it. Why did the final models differ so much?