Sociology 592 - Research Statistics I
Final Exam
December 16, 1994

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. SPSS produced the following computer printout. If $\alpha = .05$ and forward stepwise regression is being used, then X1 should be added to the model next. (Note that, unfortunately, the T values have been accidentally deleted.)

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

Variable             B        SE B       Beta   Correl Part Cor  Partial      Sig T

X2           13.432836    3.516508    .360000  .360000  .360000  .360000      .0002
(Constant)   54.313433    1.572583                                            .0000

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

Variable     Beta In  Partial  Min Toler       Sig T

X1           .165657  .176672    .990000       .0802
X3           .155556  .165899    .990000       .1008
```

b. As we saw, larger samples tend to produce larger F values. This is because, the larger the sample, the larger R^2 tends to be.

c. Y is regressed on X1, X2, and X3. To find b_1, use the formula

$$b_1 = (s_2^2 * s_{y1} - s_{12} * s_{y2}) / (s_1^2 * s_2^2 - s_{12}^2)$$

d. At one point, critics claimed that injuries had ruined Nebraska's chance for the national championship in football. With his team now healthy, Coach Tom Osborne of the #1 ranked Cornhuskers finds himself in the incredible position of having the two best quarterbacks in the country. He has to figure out how to use them in the Orange Bowl game against Miami. He has three options: Play Brook Berenger most of the game; play Tommy Frazier most of the game; or play both Berenger and Frazier equal amounts. The Statistics department at Nebraska has gathered data from Nebraska games played over the last 3 years. Three variables are computed: BROOK (coded 1 if Berenger played most of the game, 0 otherwise), TOMMY (coded 1 if Frazier played most of the game, 0 otherwise) and BOTH (Coded 1 if both played the same amount, 0 otherwise). The dependent variable is SCORE, score of the game. The following results are obtained:

```
Variable         B

BROOK          -8.00
TOMMY          -5.00
(Constant)     46.00
```

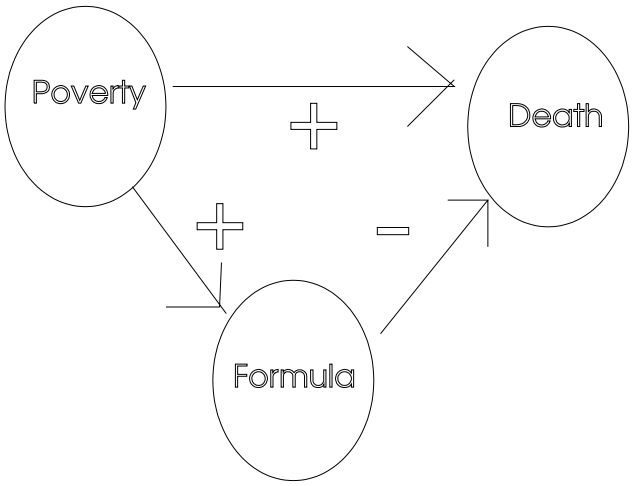
Based on these results, the smartest thing for Osborne to do (assuming he wants to get the highest score possible) is to play Tommy Frazier most of the game.

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. N = 200, K = 11, R^2 = .30, SSE = 77. Construct the ANOVA table.

b. r_{y1} = .24, r_{y2} = .25, r_{12} = .95. Compute the standardized betas. What warning might you want to pass on to others who look at these results?

c. A few years ago, there was a great controversy over the sale of infant formula in third world countries. Critics maintained that the use of formula was less healthy than breastfeeding, and that as a result, millions of children were suffering or dying. Suppose, however, that you believed in the following model:



That is, when all variables are considered simultaneously,
- poverty increases the likelihood of using infant formula
- Poverty also increases infant deaths
- Use of infant formula, however, decreases infant deaths.

Suppose an empirical study found that children fed infant formula were more likely to die than children who were breastfed, i.e. there was a *positive* correlation between infant formula use and infant deaths. Would this result necessarily invalidate the above model? Explain your reasoning.

d. A professor believes that graduate students today aren't as smart as when he was in school. He has a sample of 100 graduate students from the 1960s and 100 students from today. The following variables are measured: COHORT (coded 1 if from the 1960s, 0 if from the 1990s) and GRESCORE (the student's score on the GRE exam). Unfortunately, his bungling graduate assistant has lost the computer printout that has the results. All that is remembered is that, when GRESCORE is regressed on COHORT, a = 1400, b = -100, and the R^2 is .25. Based on this information, should the professor conclude that students today are not as smart as they used to be?

e. A sociologist has collected data from 150 respondents. When she regresses Y on X1 and X2, she gets F = 147. When she regresses Y on X1, X2, X3, and X4, she gets R^2 = .75. Using α = .05, test

H_0: $ß_3 = ß_4 = 0$
H_A: $ß_3$ and/or $ß_4$ do not equal 0.

3. A researcher is investigating home mortgage lending in a medium-sized urban county. The researcher believes that characteristics of the applicant, characteristics of the neighborhood the desired property is located in, and characteristics of the bank at which the application is made all affect whether or not a loan application gets accepted or rejected. Hence, data for 10,376 home mortgage applications over a 3 year period have been collected. For each application, information on the following variables is available:

Denial	Was the loan application denied (i.e. turned down)? 1 = Yes, 0 = No.
Applinc	Applicant's income, in thousands of dollars
Black	Was the applicant black? 1 = Yes, 0 = No
Lowmod	Is the desired property located in an area classified by the Census Bureau as "Low to Moderate Income" (i.e., one of the poorer sections of the county)? 1 = Yes, 0 = No.
NW50	Is the desired property located in an area that is 50% or more nonwhite? 1 = Yes, 0 = No.
SmallBnk	Is the bank the application was made at small in size (i.e. bank assets are less than $100 million)? 1 = Yes, 0 = No.

Following are the results:

```
            Mean  Std Dev  Label

DENIAL      .100     .300
APPLINC   44.996   37.266
BLACK       .033     .179
LOWMOD      .153     .360
NW50        .026     .159
SMALLBNK    .214     .410

Correlation

             DENIAL    APPLINC      BLACK     LOWMOD       NW50   SMALLBNK

DENIAL        1.000
APPLINC       -.115      1.000
BLACK          .073      -.053      1.000
LOWMOD         .115      -.155       .181      1.000
NW50           .053      -.046       .109       .383      1.000
SMALLBNK      -.135      -.034      -.009       .003      -.030      1.000

Equation Number 1    Dependent Variable..   DENIAL

Multiple R           .21077          Analysis of Variance
R Square             .04442                              DF      Sum of Squares      Mean Square
Adjusted R Square    .04396          Regression           5           41.46264          8.29253
Standard Error       .29327          Residual         10370          891.89732           .08601

                                     F =         (1)          Signif F =  .0000

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

Variable             B        SE B       Beta   Correl Part Cor  Partial  Tolerance        VIF         T

APPLINC        -.000828     .000078   -.102872 -.114672 -.101526 -.103303    .973991       (2)    -10.576
BLACK           .084280     .016420    .050169  .073470     (3)   .050341    .964564      1.037     5.133
LOWMOD             (4)      .008852    .088110  .114622  .079613  .081173    .816428      1.225     8.294
NW50            .009354     .019682    .004947  .053107  .004562  .004667    .850372      1.176      .475
SMALLBNK       -.100833     .007028       (5)  -.134729 -.137730 -.139517    .997604      1.002   -14.348
(Constant)      .144532     .005137                                                              28.135
```

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

b. (10 points) Do an F test of the hypothesis
H_0: $ß_{Applinc} = ß_{Black} = ß_{Lowmod} = ß_{NW50} = 0$

(i.e., all coefficients, except the one for Smallbnk, equal 0)

c. (20 points) Interpret the results. Be sure to answer the following questions:

1. What percentage of the loan applications were denied?

2. What proportion of the loan applicants were black? Overall, are blacks more likely or less likely than others to have their applications denied? Can the differences in denial rates between blacks and whites be attributed solely to differences in black and white incomes?

3. What percentage of the applications were for loans in minority areas (i.e. areas that were more than 50% minority)? Overall, are loans from minority areas more likely or less likely to get denied than loans from other areas? Is this true even after controlling for things like the race and income of the applicant, and the income of the tract? If not, offer your explanation as to why.

4. Based on these results, if you were applying for a loan yourself, would you apply at a small bank or a larger bank?