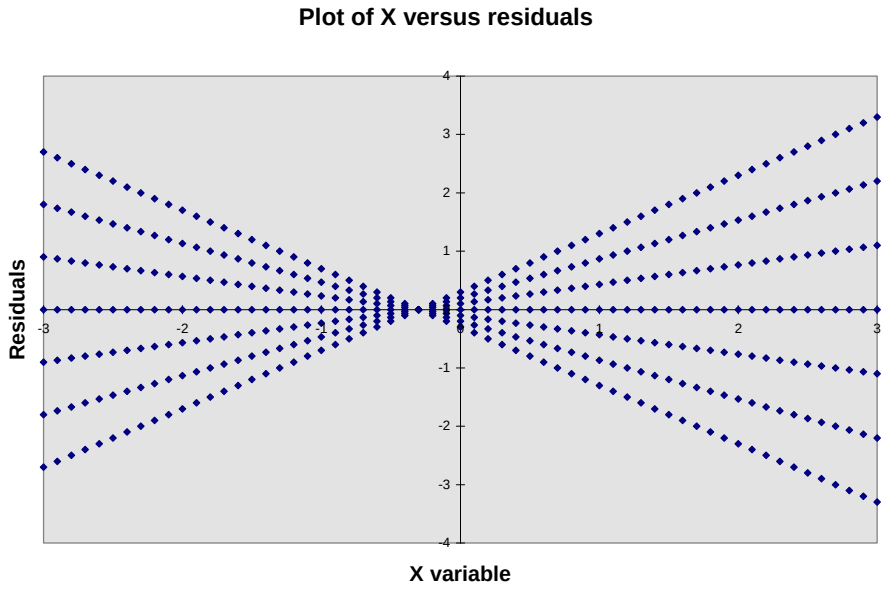
Sociology 593
Exam 1
February 17, 1995

I. True-False. (25 points) Indicate whether the following statements are true or false. If false, briefly explain why.

1. A researcher regressed Y on X. When he plotted the residuals against X, he got the following. He should now use the GQ test to determine whether heteroscedasticity is present.



2. In a regression, if the alternative hypothesis is two-tailed *and* there is only one IV, either an F test or a T test is appropriate.

3. Random measurement error results in biased estimates of means, variances and covariances.

4. If more variables are added to an equation, the F value will always either stay the same or increase.

5. A researcher found that the reliability of a measuring instrument was higher for women than it was for men. This must mean that there is less error variance in women's responses, i.e. women provide more accurate answers than do men.

II. *Short answer*. Answer *three* of the following four questions. (25 points each; up to 10 points extra credit if you do all 4).

1. A researcher collected the following data:

Case #	Y	X1	X2	X3
1	30	2	Missing	12
2	37	2	1	Missing
3	41	3	1	20
4	42	1	Missing	16
5	45	3	2	Missing
6	49	1	2	27
7	51	Missing	1	30
8	55	3	2	33
9	58	Missing	2	19
10	60	2	Missing	24

a. Suppose the researcher believes that data are missing on a *random* basis, i.e. those who did not respond are no different than those who did. What would you recommend for her—pairwise deletion of missing data, or listwise deletion? Why?

b. Suppose the researcher believes that data may be missing on a *non-random* basis. What would you recommend for her—substitution of the mean for MD cases, or substitution of the mean plus including missing data dichotomies. Why?

2. A researcher obtained the following printout:

```
Listwise Deletion of Missing Data

            Mean  Std Dev  Label

Y         79.000    9.400
V1        14.000    2.700
V2        32.000    5.600
V3        42.000    7.100

N of Cases =   200

Correlation:

                  Y         V1         V2         V3

Y             1.000       .240       .250       .270
V1             .240      1.000       .810       .850
V2             .250       .810      1.000       .900
V3             .270       .850       .900      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..   Y

Multiple R           .27102          Analysis of Variance
R Square             .07345                          DF      Sum of Squares      Mean Square
Adjusted R Square    .05927          Regression       3          1291.53396        430.51132
Standard Error      9.11718          Residual       196         16292.10604         83.12299

                                     F =       5.17921       Signif F =  .0018

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

Variable              B        SE B       Beta  Tolerance        VIF         T  Sig T

V1              .115363     .463386    .033136    .266842      3.748      .249  .8037
V2              .048669     .270006    .028994    .182703      5.473      .180  .8571
V3              .285627     .237076    .215740    .147427      6.783     1.205  .2297
(Constant)    63.831186    3.920948                                    16.280  .0000
```

Much to her dismay, none of the T values for the beta coefficients are statistically significant.

a. What problem might account for this? Point to at least two things in the printout that support your argument.

b. Does this problem cause parameter estimates to be biased? If not, then why should you be concerned about it?

c. Briefly discuss at least three possible ways for dealing with the problem.

3. A researcher fears that heteroscedasticity may be a problem in her data (N = 210). She therefore runs two regressions. Following are part of her results:

Regression 1:

```
Selecting only Cases for which         X LE     1.20

Equation Number 1    Dependent Variable..   Y

Block Number  1.  Method:  Enter      X


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


Multiple R           .45594          Analysis of Variance
R Square             .20788                              DF      Sum of Squares      Mean Square
Adjusted R Square    .19822          Regression           1            10.01000         10.01000
Standard Error       .68202          Residual            82            38.14205           .46515

                                     F =      21.52008       Signif F =  .0000
```

Regression 2:

```
Selecting only Cases for which         X GE     1.90

Equation Number 1    Dependent Variable..   Y

Block Number  1.  Method:  Enter      X


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


Multiple R           .18365          Analysis of Variance
R Square             .03373                              DF      Sum of Squares      Mean Square
Adjusted R Square    .02194          Regression           1            10.01000         10.01000
Standard Error      1.87012          Residual            82           286.78205          3.49734

                                     F =       2.86217       Signif F =  .0945
```

a. Explain, in your own words, the logic behind the researcher's strategy, i.e. why is this a good way for testing for heteroscedasticity?

b. Compute the appropriate test statistic. Based on the test statistic, what should the researcher conclude about heteroscedasticity? [HINT: In case you don't have your tables handy, the critical value for the test statistic is about 1.5]

c. If a problem appears to exist, suggest two or more ways the researcher might try to deal with it.

4. A researcher has collected information on the following variables:

Y	Depression (where high values indicate high levels of depression)
X1	Job dissatisfaction (high values indicate high dissatisfaction)
X2	Physical health (high values indicate good physical health)
X3	Income (measured in thousands of dollars)

She obtains the following results:

```
          Mean  Std Dev

Y         .000    2.128
X1        .000   10.726
X2        .000   79.477
X3        .000     .939

N of Cases =   427

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

Equation Number 1    Dependent Variable..   Y

Block Number  1.  Method:  Enter      X1       X2       X3


Variable(s) Entered on Step Number  1..    X3
                                    2..    X2
                                    3..    X1


                                  Analysis of Variance
R Square            (1)                        DF      Sum of Squares      Mean Square
                                  Regression      3          1323.37441        441.12480
Standard Error      1.19664       Residual      (2)           605.71583          1.43195

                                  F =     308.05831       Signif F =  .0000

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

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

X1                  (3)       .046160    .590600  .759340  .069157  .122489    .013712     72.931     2.538
X2             -.006737       .005013      (4)    .680281 -.036614 -.065203    .021178     47.218    -1.344
X3            1.141410       .165034    .503680  .811429  .188433  .318738    (5)          7.145     6.916
(Constant) -1.18480E-15       .057910                                                              .000
```

a. Fill in the missing quantities (1)-(5)

b. The researcher believes that higher incomes lead to lower levels of depression. Do the results support her belief?

c. If you were doing stepwise regression, what variable would be removed from the equation next?

d. Do an F test of the hypothesis

H_0: $\beta_2 = \beta_3 = 0$

H_A: β_2 and/or $\beta_3 \neq 0$