Advanced Biostatistics **Quiz 2** **Name** _____
February 23, 2005 **13 Total Points**

Directions: Thoroughly, clearly and neatly answer the following two problems in the space given, showing all relevant calculations. Unless otherwise noted, use $\alpha = 5\%$ throughout.

1. (1 + 1.5 + 5 = 7.5 points) The data analyzed using MLR (multiple linear regression) in Outputs A-C in the Appendix were collected in the summer of 1975 in Florida from an experiment to investigate the use of silver iodide in cloud seeding to increase rainfall, where the response variable Y is the amount of rain (cubic meters x 10^7) that fell in the target area for a 6 hour period on each suitable day. Explanatory variables are:

T	Time: number of days after the first day of the experiment (June 1, 1975);
SNE	a measured suitability criterion;
C	Echo coverage: the percentage cloud cover in the experimental are (measured using radar);
P	Pre-wetness: the total rainfall in the target area 1 hour before seeding (cubic meters x 10^7)
E	Echo motion: an indicator showing whether the radar echo was moving (1) or stationary (2)

   (a) These researchers are concerned that in the MLR involving all five explanatory variables that these explanatory variables are so overly correlated as to cause problems with the analysis. Comment briefly on whether this appears to be a problem here, supporting your claim.

   (b) For the MLR model that includes only T, P and E as explanatory variables, clearly and explicitly interpret the estimated slope parameter for E (echo motion). (You are not being asked to test any hypotheses for this part.)

   (c) Test the compound hypothesis that SNE, C, and E can ***simultaneously*** be dropped from the MLR involving all five explanatory variables, $y = \beta_0 + \beta_1 T + \beta_2 SNE + \beta_3 C + \beta_4 P + \beta_5 E + \varepsilon$, showing all work and clearly giving your detailed conclusion.

   Null Hypothesis _____

   Alternative Hypothesis _____

   Calculated test statistic _____ Degrees of freedom _____

   P-value _____

   Detailed conclusion

2. (1.5 + 1.5 + 2.5 = 5.5 points) Short answer.

(a) Sketch a residual plot that shows a violation of one of the assumptions inherent in linear regression ***and indicate which assumption is being violated***.

Assumption violated _____

Sketch

(b) Sketch a residual plot that shows another (other than the one highlighted in part (a)) violation of one of the assumptions inherent in linear regression ***and indicate which assumption is being violated.***

Assumption violated _____

Sketch

(c) Describe an example (not discussed in class or in the *Course Notes*) illustrating the analysis of covariance design. Please be clear in highlighting the response variable, the factor(s) and the covariate(s).

Response variable _____

Factor(s) _____

Covariate(s) _____

Description

Output A. Regression Analysis: y versus t, sne, c, p, e

```
The regression equation is
y = 1.26 - 0.0363 t + 0.108 sne + 0.072 c + 3.21 p + 1.71 e

Predictor      Coef  SE Coef      T      P  VIF
Constant      1.259    2.036   0.62  0.546
T          -0.03626  0.01650  -2.20  0.044  1.3
SNE          0.1076   0.4482   0.24  0.813  1.2
C            0.0724   0.1387   0.52  0.609  1.1
P             3.209    1.568   2.05  0.059  1.2
E            1.7083   0.9527   1.79  0.093  1.1

S = 1.60991   R-Sq = 51.6%   R-Sq(adj) = 35.5%

Analysis of Variance
Source          DF      SS     MS     F      P
Regression       5  41.452  8.290  3.20  0.037
Residual Error  15  38.877  2.592
Total           20  80.329
```

Output B. Regression Analysis: y versus t, p, e

```
The regression equation is
y = 1.84 - 0.0348 t + 3.49 p + 1.73 e

Predictor      Coef  SE Coef      T      P
Constant      1.844    1.530   1.20  0.245
T          -0.03477  0.01433  -2.43  0.027
P             3.488    1.422   2.45  0.025
E            1.7339   0.9042   1.92  0.072

S = 1.53011   R-Sq = 50.5%   R-Sq(adj) = 41.7%

Analysis of Variance
Source          DF      SS      MS     F      P
Regression       3  40.528  13.509  5.77  0.007
Residual Error  17  39.801   2.341
Total           20  80.329
```

Output C. Regression Analysis: y versus t, p

```
The regression equation is
y = 4.36 - 0.0402 t + 2.65 p

Predictor      Coef  SE Coef      T      P
Constant     4.3629   0.8413   5.19  0.000
T          -0.04019  0.01506  -2.67  0.016
P             2.649    1.451   1.83  0.084

S = 1.63997   R-Sq = 39.7%   R-Sq(adj) = 33.0%

Analysis of Variance
Source          DF      SS      MS     F      P
Regression       2  31.918  15.959  5.93  0.010
Residual Error  18  48.411   2.689
Total           20  80.329
```