Homework #3, ST518

1. (Taken from example 12.7 of Rao's textbook) In this edition of a make-believe problem involving test preparation and outcome, $n = 20$ students are randomized to two groups, one of which receives a smart pill supplement, with the other receiving a placebo. The code below was used to obtain the output that follows.

```
proc reg;
   model y=supp z/ss1 ss2;
run;
```

```
The SAS System
The REG Procedure
                          Analysis of Variance

                                    Sum of           Mean
Source                     DF      Squares         Square    F Value    Pr > F
Model                       2   2039.50067     1019.75034     136.83    <.0001
Error                      17    126.69933        7.45290
Corrected Total            19   2166.20000

                     Parameter      Standard
Variable     DF       Estimate         Error   t Value   Pr > |t|     Type I SS    Type II SS
Intercept     1        0.87294       2.28148      0.38     0.7067        26938       1.09109
supp          1        0.84259       1.22368      0.69     0.5004      24.20000       3.53364
z             1        1.13117       0.06879     16.44     <.0001    2015.30067    2015.30067
```

   (a) Consider a t-test comparing posttest scores (y) with and without the supplement without including pretest scores (z) in the analysis. Report the absolute value of the t-statistic and associated degrees of freedom from such a test.

   (b) Consider a simple linear regression of posttest scores on pretest scores (z). Construct the ANOVA table from such a model. Report the coefficient of determination.

   (c) Consider the general linear model with additive effects for supplement and pretest score.

      i. After controlling for pretest score, is the estimated increase in posttest scores due to the supplement positive or negative? Report the estimate, along with a standard error.

      ii. In the population of students with pretest score $z = 50$, report the estimated mean posttest score for students with and without the supplement.

2. Consider a multiple linear regression with $n = 4$ observations and $p = 2$ explanatory variables:

i	x_{i1}	x_{i2}	y_i
1	1	-2	68
2	0	-1	57
3	0	2	90
4	2	1	112

   (a) Calculate these matrix products by hand (you can use software to check):

$$X, X'X, X'Y$$

   (b) Obtain the inverse of $X'X$. You can use software, but it is doable by hand too. To obtain it by hand, consider augmenting the matrix with the (3×3) identity matrix on the right and then carrying out the following sequence of row operations which should leave the identity matrix on the left, and the inverse of $X'X$ on the right ($R1, R2, R3$ refer to rows 1,2 and 3):

$$R3' = R3/10, R2' = 3R1 - 4R2, R2' = R2/11, R1' = R1 - 3R2, R1' = R1/4.$$

   (c) Obtain the product $(X'X)^{-1}X'Y$. What is this vector product called, (in words, not symbols).

3. Consider the `trees` dataset in BASE R (also available on moodle as "`trees.txt`"). Consider a multiple linear regression of `Volume` on `Girth` and `Height` in an additive model. (To solve this problem using R, you may want to download the `pcor.R` package. In SAS PROC REG, you may want to use the `pcorr1` and `pcorr2` options in the MODEL statement.)

   (a) Obtain the partial correlation coefficient between `Volume` and `Girth` after adjusting for linear dependence on `Height`.

   (b) Obtain the partial correlation coefficient between `Volume` and `Height` after adjusting for linear dependence on `Girth`.

   (c) For part (a), the partial correlation coefficient is calculated as the correlation between which two quantities (residuals from which models)?