STA 6208 – Spring 2003 – Final Exam

Print Name: **SSN:**

1) Consider the regression through the origin simple regression model:

$$Y_i = \beta_1 X_i + \varepsilon_i \qquad X_i \geq 0 \forall i \qquad \varepsilon \sim N(0, \sigma^2)$$

a) Derive the least squares estimator $\hat{\beta}_1$.

b) Derive the mean and variance of $\hat{\beta}_1$.

c) Consider the estimator $\tilde{\beta}_1 = \sum Y_i / \sum X_i$. Derive its mean and variance.

2) If a random variable Y follows a Poisson distribution with parameter λ, then its mean and variance are both λ. What transformation of Y will make its variance constant, approximately?

3) A large consumer products company wants to compare the effectiveness of radio/television advertising and newspaper advertising. They sample $n = 22$ cities with comparable populations and allocate specific levels of radio/television advertising (X_1, in \$1000s) and newspaper advertising (X_2, in \$1000s). For each city, they observe sales (Y, in \$1000s).

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 156.4 \\ 13.1 \\ 16.8 \end{bmatrix} \qquad (\mathbf{X'X})^{-1} = \begin{bmatrix} 0.63633 & -0.00586 & -0.01126 \\ -0.00586 & 0.00012 & 0.00002 \\ -0.01126 & 0.00002 & 0.00035 \end{bmatrix} \qquad \mathbf{Y'(I-P)Y} = 479760$$

a) For the first city, there was no radio/television advertising and they spent \$40,000 on newspaper ads. Their observed sales were \$973,000. Give the predicted value and residual for this city (in the units the regression model was fit).

b) Give a 95% confidence interval for the change in mean sales when increasing newspaped advertising by one unit, controlling for radio/television advertising.

c) Use the general linear test to test whether the effects of each type of advertising are the same in terms of sales. Clearly state and complete all aspects of the test.

4) A regression model was fit among a sample of naval hospitals, relating monthly man-hours (Y) to the following set of predictors: total number of bed-days, number of xrays, typical length of stay, population in the area, and total number of patients. The analysis of variance, sequential (Type I), and partial (Type II) sums of squares are given below.

```
                              The REG Procedure
                                Model: MODEL1
                        Dependent Variable: man_hour

                             Analysis of Variance

                                    Sum of           Mean
Source                   DF        Squares         Square    F Value    Pr > F

Model                     5      490177488       98035498     237.79    <.0001
Error                    11        4535052         412277
Corrected Total          16      494712540


             Root MSE             642.08838    R-Square     0.9908
             Dependent Mean      4978.48000    Adj R-Sq     0.9867
             Coeff Var             12.89728


                             Parameter Estimates

                      Parameter      Standard
Variable     DF        Estimate         Error    t Value    Pr > |t|      Type I SS

Intercept     1      1962.94816    1071.36170       1.83      0.0941      421349473
bed_days      1         1.58962       3.09208       0.51      0.6174      480950232
xrays         1         0.05593       0.02126       2.63      0.0234        7189926
len_stay      1      -394.31412     209.63954      -1.88      0.0867        1658984
pop           1        -4.21867       7.17656      -0.59      0.5685         367483
patient       1       -15.85167      97.65299      -0.16      0.8740          10863

                             Parameter Estimates

                         Variable     DF    Type II SS

                         Intercept     1       1383997
                         bed_days      1        108962
                         xrays         1       2853834
                         len_stay      1       1458572
                         pop           1        142465
                         patient       1         10863
```

a) Test whether the partial regression coefficients for population (pop) and patients are simultaneously 0 at the $\alpha = 0.05$ significance level, controlling for all other predictors (including intercept) in the model.

b) Give the analysis of variance table for this data, had the regression model been forced to go through the origin.

5) The authors of an article conducting a regression analysis with n observations and p predictors report only Adjusted-R^2. Show how to derive the F-statistic for testing $H_0 : \beta_1 = \cdots = \beta_p = 0$, assuming the model does have an intercept term.

6) A dose response study related steady state cotinine replacement (Y, as a percentage of baseline) to cigarettes smoked per day prior to trial (X) in smokers using a nicotine patch (at doses, 11, 22, and 44 mg). The fitted regression equations for each dose were:

$$\hat{Y}_{44} = 152.7 - 1.34X \qquad \hat{Y}_{22} = 82.7 - 1.05X \qquad \hat{Y}_{11} = 48.1 - 0.55X$$

a) The overall mean cigarettes per day was 25.5, give the adjusted mean response for each dose.

b) Write out a statistical model representing their analysis.

c) Briefly describe how you would test whether there was an interaction between cigarettes smoked per day and dose with respect to cotinine replacement. Write out your procedure assuming you have access to their data set containing Y, X, and dummy variables for the 11 and 22 mg doses, and any standard regression computer package.