Second Exam: Economics 388, Econometrics Fall 2004 in R. Butler's section

YOUR NAME:____

Section I (30 points) Questions 1-10 (3 points each)
Section II (20 points) Questions 11-12 (10 points each)
Section III (50 points) Questions 13-14 (15 points each)
Question 15 (20 points)

Section I. Define or explain the following terms (3 points each)

1. LaGrange-Multiplier test-

2. beta coefficient-

3. adjusted R-square-

4. "boundedness" problem in the probability model-

5. maximum likelihood estimation--

6. method of moment estimators-

7. dummy variable trap --

8. central limit theorem--

9. asymptotic efficiency-

10. probit model of being married-

II. Some Fun Stuff:

11. True, False or Uncertain question: you are graded on your explanation and not on whether you guessed T or F correctly. "Solomon has estimated β_{age} from a regression where all the assumptions held and got the estimate $\hat{\beta}_{age}$. Since his theory really has to do with the square root of β_{age}, Solomon knew it was OK to use $\sqrt{\hat{\beta}_{age}}$ to test his theory because least squares estimators are unbiased and consistent."

12. Define y=1 if the individual is married; 0 otherwise
age=age of the individual in years
edu=educational attainment in years

Write the Shazam program that estimates the linear probability model (LPM) for being married, correcting for heteroskedasticity (recall the LPM is intrinsically heteroskedastic with variance of error μ_i equal to $P_i * (1 - P_i)$)

III. Really Fun Stuff

13. Suppose that for the standard regression model, $y = X\beta + \mu$, we rescale both the independent variables and the dependent variable by non-zero variables c_0, c_1, c_2, …, c_k by regressing $c_0 y_i$ on a constant c_1, $c_2 x_{2i}$, $c_3 x_{3i}$,…, $c_k x_{ki}$ (so there are k regressors, including an intercept which is measured as c_1 instead of 1, so that the intercept coefficient is β_1). In other words, instead of the OLS estimator $\hat{\beta} = (X'X)^{-1}X'Y$, you do OLS on the transformed data $\hat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$, where the transformed data can be described by:

$$X^* = XC \text{ where } C = \begin{bmatrix} c_1 & 0 & 0 & 0 & 0 \\ 0 & c_2 & 0 & 0 & 0 \\ 0 & 0 & c_3 & 0 & 0 \\ 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & c_k \end{bmatrix} \text{ and } Y^* = C_0 Y \ , C_0 = \begin{bmatrix} c_0 & 0 & 0 & 0 & 0 \\ 0 & c_0 & 0 & 0 & 0 \\ 0 & 0 & c_0 & 0 & 0 \\ 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & 0 & c_0 \end{bmatrix} = c_0 I .$$

Prove that the i[th] beta between the transformed data, and the untransformed data, have the following relationship: $\hat{\beta}_i^* = \frac{c_0}{c_i}\hat{\beta}_i$.

14. Data from GPA2.RAW generated the regression below, where

colgpa=college gpa

hsize=size of the individual's high school graduating class (in hundreds)

hsizesq=square of hsize

female=1 if female, 0 if male

```
|_genr hsiz_fm=hsize*female
|_genr hsiz2_fem=hsizesq*female
|_ols colgpa hsize hsizesq female hsiz_fm hsiz2_fm

 R-SQUARE =   0.0172     R-SQUARE ADJUSTED =   0.0161
 VARIANCE OF THE ESTIMATE-SIGMA**2 =  0.42683
 STANDARD ERROR OF THE ESTIMATE-SIGMA =  0.65332
 SUM OF SQUARED ERRORS-SSE=   1763.2
 MEAN OF DEPENDENT VARIABLE =   2.6527
 LOG OF THE LIKELIHOOD FUNCTION = -4106.10

 VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR    4131 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 HSIZE     0.91537E-01 0.2517E-01   3.637     0.000 0.056     0.2413     0.0966
 HSIZESQ  -0.14978E-01 0.3453E-02  -4.337     0.000-0.067    -0.2871    -0.0613
 FEMALE    0.21530     0.5867E-01   3.670     0.000 0.057     0.1626     0.0365
 HSIZ_FM  -0.63405E-01 0.3778E-01  -1.678     0.093-0.026    -0.1745    -0.0300
 HSIZ2_FM  0.96083E-02 0.5201E-02   1.847     0.065 0.029     0.1459     0.0176
 CONSTANT   2.4951     0.3931E-01   63.47     0.000 0.703     0.0000     0.9406
 |_test
 |_test female=0
 |_test hsiz_fm=0
 |_test hsiz2_fm=0
 |_end
 F STATISTIC =   17.297779     WITH    3 AND 4131 D.F.  P-VALUE= 0.00000
```

a). Should hsizesq be in the regression? Why or why not?

b) Is there an optimal high school size for males? If so, what is it? If not, why not?

c) Is there an optimal high school size for females? If so, what is it? If not, why not?

d) What does the F-statistic at the bottom of the page indicate about the regression results? (Be as specific as you can in answering)

15. Suppose that for the general linear regression model, $y = X\beta + \mu$, the modeling assumptions are satisfied, in particular, μ is normally distribution with mean 0 and variance $\sigma^2 I$ where I is the n by n identity matrix. Prove that $s^2 = \frac{\sum_{i=1}^{n} \hat{\mu}_i^2}{n-k} = \frac{\hat{\mu}'\hat{\mu}}{n-k}$ is a consistent estimator of σ^2. (You can assume that $\text{plim}\left(\frac{\text{X'X}}{\text{n}}\right) = \Omega$, a positive definite, symmetric matrix for any size n. Also recall the $\hat{\mu}$ is the least squares residual, $\hat{\mu} = y - X\beta = (I - X(X'X)^{-1}X')y = (I - (X(X'X)^{-1}X')\mu$.)