First Exam: Economics 388, Econometrics   Spring 2006 in R. Butler's class

YOUR NAME:______________________________________

Section I (30 points) Questions 1-10 (3 points each)
Section II (40 points) Questions 11-15 (10 points each)
Section III (30 points) Question 16 (20 points each)

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

1. centered vs. uncentered R^2 -

2. Frisch theorem -

3. population regression function vs. sample regression function-

4. conditional probability density function of y given x-

5. method of moments (approach to parameter estimation)-

6. omitted variable bias-

7. micronumerosity-

8. random sampling -

9. probability significance values (i.e., 'p-values')-

10. SST=SSM + SSR (for regressions)-

II. Some Concepts

11. Indicate whether the following statement is True, False or Uncertain and explain why. You are graded only on the explanation for your answer. "Let there be *n* observations and *k* regressors in the standard linear regression model. The uncentered regression R-square for this regression model $Y = X\beta + \mu$ will be unchanged if there is a nonsingular transformation, *A* (with dimension *kxk*), of the data matrix, i.e., *XA*, and all the values in the *Y* vector are rescaled by the real number "*d*", or in other words, *Y* is postmultiplied by a diagonal matrix *D=dI*, where I is the *nxn* identity matrix and *d* is a real number."

(something of a hint: the uncentered R-square can be written $\frac{\| P_X Y \|^2}{\| Y \|^2}$ before the linear transformations, and as $\frac{\| P_{XA} YD \|^2}{\| YD \|^2}$ after the transformations.)

`

12. Indicate whether the following statement is True, False or Uncertain and explain why. You are graded only on the explanation for your answer. "Suppose the multiple regression model is partitioned as $Y = X_1\beta_1 + X_2\beta_2 + \mu$, where we are interested in testing β_2 =0. Let *X* be the unrestricted regression space *X=[X_1 X_2]*, and X_1 the restricted regression space (where β_2 =0 is imposed). Then the sum of squared residuals from unrestricted regression will be greater then the sum of squared residuals from the restricted regression, that is $\| M_X Y \|^2 > \| M_{X_1} Y \|^2$."

13. My daughter has a pyramid dice, with four sides numbered from 1 to 4. Let W be the random variable corresponding to number that's on the bottom side when the dice is rolled. If the dice is not fair, but the probability that the ides with numbers 1, 2 or 3 will occur is one sixth (for each of these events taken separately) then

a. What is the expected value of the random variable W and what is the variance of W?

b. If we did not know whether the dice were fair or not (i.e., we only suspected that my daughter may have weighted then as described above—and that the alternative was that the dice were fair in the sense that each outcome was equally probable), how could we test for that?

14. a. Explain what a positive definite matrix is (or positive semi-definitive if you prefer).

b. Where might positive definite (or positive semi-definite) matrices be useful in comparing the covariance matrices of alternative estimators of the same model?

c. Prove the Gauss-Markov, or BLUE theorem for the multiple regression model.

15. Interpret the STATA (or SAS) code statements circled below, and the statistical output that follows, in the space below:

```
/*stata code*/ # delimit ;
infile  lsalary  years  games  gamesyr  bavg hrunsyr  rbisyr using "e:\classrm_data\wooldridge\mlb1.raw", clear;
replace bavg=. if bavg<=0;
replace hrunsyr=. if hrunsyr<=0;
replace rbisyr=. if rbisyr<=0;
regress lsalary years gamesyr bavg hrunsyr rbisyr;
test (bavg=0) (hrunsyr=0) (rbisyr=0);
```

```
/* SAS code*/ data one;
infile "e:\classrm_data\wooldridge\mlb1.raw" lrecl=800;
input lsalary years  games  gamesyr  bavg hrunsyr  rbisyr ;
if bavg<=0 then bavg=. ;
if hrunsyr<=0 then hrunsyr=. ;
if rbisyr<=0 then rbisyr=. ;
run;
proc reg;
model lsalary=years gamesyr bavg hrunsyr rbisyr;
test bavg,hrunsyr,rbisyr;
run;
```

```
      Source |       SS       df       MS              Number of obs =     334
-------------+------------------------------           F(  5,   328) =   95.02
       Model |  257.696039     5  51.5392079           Prob > F      =  0.0000
    Residual |  177.906509   328  .542397893           R-squared     =  0.5916
-------------+------------------------------           Adj R-squared =  0.5854
       Total |  435.602548   333  1.30811576           Root MSE      =  .73648

------------------------------------------------------------------------------
     lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       years |   .0676365   .0123464     5.48   0.000     .0433484    .0919247
     gamesyr |   .0127922   .0028692     4.46   0.000     .0071478    .0184367
        bavg |   .0018745   .0014176     1.32   0.187    -.0009143    .0046633
     hrunsyr |   .0170263   .0166954     1.02   0.309    -.0158174    .0498699
      rbisyr |   .0092741   .0075602     1.23   0.221    -.0055985    .0241467
       _cons |   10.98279   .3718074    29.54   0.000     10.25137    11.71422
------------------------------------------------------------------------------

. test (bavg=0) (hrunsyr=0) (rbisyr=0);

       F(  3,   328) =    9.25
            Prob > F =    0.0000
```

16. The last four assumptions for the linear regression model are:

II. X'X is invertible

III. E(u|X)=0

IV. V(u|X)= $\sigma^2 I$

V. u is normally distributed

a. Using the simple regression model (with only one independent), and the slope-intercept type diagram used extensively in the book, draw (three) pictures where assumptions III, IV, and V fail to hold.

b. What do each of these assumptions (II, III, IV, and V) allow us to say about the estimated $\hat{\beta}$ vector? (You don't necessarily have to do proofs, but do have to provide as much intuition as you can)

c. Which assumptions are necessary to make the OLS estimator of the linear model the same as the maximum likelihood estimator of the linear model? Why?