First Exam: Economics 388, Econometrics Fall 2003 in R. Butler's class

YOUR NAME:____

Section I (30 points) Questions 1-10 (3 points each)
Section II (40 points) Questions 11-14 (10 points each)
Section III (30 points) Question 15

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

1. second quartile as a measure of central tendency -

2. panel data -

3. null hypothesis vs. alternative hypothesis -

4. non-nested models -

5. variance of a predicted (forecasted) value of Y -

6. LaGrange Multiplier test -

7. formulas for population variance and sample variance of a random variable -

8. Gauss-Markov theorem (BLUE theorem) -

9. micronumerosity -

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

II. Some Concepts

11. **The next 16 lines is the shazam code for the HPRICE1 data from class (with variable descriptions given in the program as comment statements; CAREFULLY LOOK AT THE GENR STATEMENTS):**

```
read(hprice1.raw) price assess bdrms lotsize sqrft colonial  lprice    lassess  &
llotsize  lsqrft
*  1. price                  price, in dollars
*  3. bdrms                    number of bedrooms
*  5. sqrft                  size of house, square feet
stat price bdrms sqrft
*next generate deviations from the means
genr bdrm_dev = bdrms - 3.5682
genr sqrf_dev = sqrft - 2013.7
genr inter_dv = bdrm_dev*sqrf_dev
ols price bdrms sqrft inter_dv
test
   test bdrms=0
   test inter_dv=0
end
stop
```

THIS IS THE RELEVANT OUTPUT FROM THE PROGRAM ABOVE:

NAME	N	MEAN	ST. DEV	VARIANCE	MINIMUM	MAXIMUM
PRICE	88	0.29355E+06	0.10271E+06	0.10550E+11	0.11100E+06	0.72500E+06
BDRMS	88	3.5682	0.84139	0.70794	2.0000	7.0000
SQRFT	88	2013.7	577.19	0.33315E+06	1171.0	3880.0

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 84 DF	P-VALUE	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
BDRMS	11262.	9407.	1.197	0.235	0.130	0.0923	0.1369
SQRFT	116.29	14.48	8.033	0.000	0.659	0.6535	0.7977
INTER_DV	23.448	10.16	2.308	0.023	0.244	0.1732	0.0204
CONSTANT	13210.	0.3340E+05	0.3955	0.693	0.043	0.0000	0.0450

```
|_test
|_test bdrms=0
|_test inter_dv=0
|_end
F STATISTIC =   4.0141144     WITH    2 AND   84 D.F.  P-VALUE= 0.02163
WALD CHI-SQUARE STATISTIC =   8.0282288     WITH    2 D.F.  P-VALUE= 0.01806
UPPER BOUND ON P-VALUE BY CHEBYCHEV INEQUALITY = 0.24912
```

a) Interpret the BDRMS and SQRFT coefficients?

b) What does the INTER_DV estimate?

c) What does the F STATISTIC indicate (be specific about the null hypothesis being tested)?

12. Indicate whether the following statement is True, False or Uncertain and explain why. You are graded only on the explanation for your answer.

a. “If all my sample observations have exactly the same level of ability, then omitting ability from the empirical specification where I regress wages on education will lead to no omitted variable bias.”

b. “The most important model assumption is that the population errors are normally distributed. Unless this assumption holds, we can do no testing in our regression models.”

13. The estimators (i.e., the formulas from which we calculate the estimates) for the OLS slope and intercept coefficients in the simple regression model ($y_i = \beta_0 + \beta_1 x_i + \mu_i$) are derived from the following "normal equations":

$$-2\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$-2\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0$$

(note that we can easily divide out the –2s). But how do we get the normal equations? You show that we can get these normal equations either by
a) Calculus (minimizing the sum of squared residuals), or
b) from the orthogonality condition on the residuals.
Show it both ways. Be explicit in your explanations (on each step): again, you don't have to solve the normal equations for the OLS estimators, you have to derive the normal equations in two different ways.

14. For the simple regression, $y_i = \beta_0 + \beta_1 x_i + \mu_i$, we can regress the dependent variable on the independent variable and get $\hat{\beta}_1$, as we did in problem 13. But we can also divide through the dependent variable by its standard deviation (call it σ_y) to get "$y_i / \sigma_y = y_i^*$," and similarly divide through the independent variables by its standard deviation to get "$x_i / \sigma_y = x_i^*$." Then we have a model of normalized regressors: $y_i^* = \alpha_0 + \alpha_1 x_i^* + \varepsilon_i$. Prove that following relationship exists between the slope coefficient of the simple model and the slope coefficient of the normalized model: $\hat{\alpha}_i = \hat{\beta}_i \dfrac{\sigma_x}{\sigma_y}$.

15. Prove that under the usual model assumptions that the least squares estimator of the multiple regression model, $\hat{\beta}$, is (a) an unbiased estimator, and (b) a consistent estimator of the population coefficient vector β.