Second Exam: Economics 388, Econometrics Fall 2003 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. log-likelihood ratio test-

2. dummy variable trap-

3. Goldfeld-Quandt test-

4. probit-

5. maximum likelihood estimation--

6. "hetcov" option in Shazam-

7. Chow statistic--

8. Feasible GLS estimator--

9. self selection--

10. "diagnos / het" option in Shazam-

II. Some Fun Stuff: True, False or Uncertain Questions (11-14); you are graded on your explanation and not on whether you guessed T or F correctly.

11. T,F or U: "To do a LM test of a restriction in the standard OLS model, you follow this procedure: 1) regress Y on the restricted (say "q" restrictions, or q-coefficients forced to be zero, for example) set of regressors, then 2) take the residuals from step 1 and regress them on all of the regressors (including those restricted from the first stage), then 3) take n (the sample size) and multiply it by the adjusted R-square from the second regression. This last statistic (n*adj. R-square) will be distributed as chi-square distribution with q-degrees of freedom under the null hypothesis."

12. T,F or U: "To predict a value of the dependent variable, you must have the regression coefficients and the values of the independent variables you want to predict with."

III. Really Fun Stuff

13. Regressing a dummy variable for marriage (=1 if married; 0 otherwise) on body mass index, body mass index squared, earnings last year, and books read last year. (The normal range for body mass index is 19 to 25. Higher than 25 is considered overweight.)

Here's the code:

```
* next calculate the body mass index for each student
genr bmi=703*(wght/(height**2))
genr bmi_sq=bmi**2
ols married bmi bmi_sq earnings books_yr /predict=yhat
if (YHAT.GE.1) YHAT=.999
if (YHAT.LE.0) YHAT=.001
GENR WT = 1 / (YHAT*(1-YHAT))
ols married bmi bmi_sq earnings books_yr / WEIGHT = WT
```

Here's the output (from the second "ols" command):

```
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
   NAME    COEFFICIENT   ERROR      23 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
 BMI      -0.35410     0.6642E-01  -5.331     0.000-0.743     -5.5621   -13.3688
 BMI_SQ    0.69985E-02 0.1310E-02   5.344     0.000 0.744      5.5755     7.2122
 EARNINGS  0.10947E-01 0.3714E-02   2.947     0.007 0.524      0.1544     0.0828
 BOOKS_YR -0.55471E-03 0.2152E-03  -2.577     0.017-0.473     -0.3266    -0.1616
 CONSTANT   4.8113     0.7224       6.660     0.000 0.811      0.0000     7.2354
```

a. What does the Shazam code do? why?

b. What does the output indicate about each of the regressors influence on the probability of being married (including the "optimal" bmi)?

14. Here's some shazam code and output from Butler's Pension consulting firm, where hours of consulting (consult) is regressed on number of employees, whether the firm employs mostly blue collar workers (blue_col=1 if it is blue collar; 0 otherwise), and an interactive term.

Here's the code:

```
sample 1 10
read consult employee blue_col
2.9   2  1
3  6  0
4.8   8  1
1.8   3  0
2.9   2  1
4.9   7  1
4.2   9  0
4.6   8  0
4.4   4  1
4.5   6  1
genr em_blue=employee*blue_col
ols consult employee blue_col em_blue
test
    test blue_col=0
    test em_blue=0
end
```

Here's the output:

```
VARIABLE   ESTIMATED  STANDARD   T-RATIO        PARTIAL STANDARDIZED ELASTICITY
  NAME    COEFFICIENT   ERROR       6 DF   P-VALUE CORR. COEFFICIENT  AT MEANS
EMPLOYEE  0.45714     0.8917E-01   5.127     0.002 0.902      1.1177     0.6617
BLUE_COL   2.0237     0.7241       2.795     0.031 0.752      0.9855     0.3195
EM_BLUE  -0.12313     0.1142      -1.078     0.322-0.403     -0.3650    -0.0940
CONSTANT  0.42857     0.6145       0.6974    0.512 0.274      0.0000     0.1128
 |_test
 |_test blue_col=0
 |_test em_blue=0
 |_end
F STATISTIC =    11.446130     WITH    2 AND    6 D.F.   P-VALUE= 0.00896
```

What do the results – the individual regressor results and the joint test at the end – indicate?

15. Suppose that for the general linear regression model, $y = X\beta + \mu$, the error term is heteroskedastic:

$$[Variance(\mu)]^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_3^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_4^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma_n^2} \end{bmatrix}$$

. That is, $\mu \approx N(0, \Sigma)$ where the inverse of the variance matrix Σ is given above. Derive the elements of a nxn matrix T that transforms the model such that the transformed model is homoskedastic: T such that $Ty = TX\beta + T\mu$ such that "$T\mu$" is homoskedastic. Show your work and be explicit at each step; no credit for showing T without describing how it is derived.