Economics 3950
Spring 2008
Dr. Richard Mueller

Assignment #1

Instructions: These questions should be answered using a text editor or a word processor where you can cut and paste output from your statistical program (where necessary). Please mark question numbers clearly. This assignment is **due on Friday, February 1st, 2008 by 12:00 in D-552.**

1. (50 points total) Exercise 3.5, pp. 122-23.

2. (25 points total) Exercise 3.10, p. 123. Assume now that $\alpha = 0$. Would this change your answer? If so, how would it change your answer?

3. (75 points total) Exercise 3.22, pp. 125-26.

4. (30 points total) Exercise 3.33, pp. 131-32.

5. (30 points total) Exercise 3.34, p. 132.

Grand Total: 210 points

Answer Key

1.

a. FALSE. X values closer to their mean implies a larger variance and thus the estimates are less precisely estimated. See Equations 3.19 and 3.20.

b. FALSE because for unbiasedness we need Assumptions 3.3 and 3.4. Violation of Assumption 3.4 implies that unbaisedness is no longer valid.

c. FALSE. We only need Assumption 3.8 for hypothesis testing. An estimator is still BLUE without this restriction.

d. TRUE. This is because t and F distributions for the test statistics were derived from the assumption of normality which is necessary for hypothesis testing.

e. TRUE. The width of a confidence interval directly depends on the standard error of an estimate.

f. TRUE. If Var(X) is large, then from Equation 3.19 and 3.20 we know the variances will be smaller and hence confidence intervals with be narrower.

g. FALSE because a high p-value means rejection of Ho might result in a high probability of a Type I error.

h. TRUE because a higher level of significance means a lower value for t* and hence the actual value |tc| is more likely to be to its right.

i. PARTIALLY TRUE. Violation of Assumptions 3.5 and 3.6 only affects the BLUE property. Thus estimators are still unbiased and consistent but not BLUE.

j. FALSE. The null hypothesis is a statement about whether or not the parameter has a certain value. This is either true or not and therefore it is meaningless to attribute a probability to whether Ho is true or not. However, the rejection of a true hypothesis, a Type I error, is a random event because it can change from trial to trial. The p-value is the probability of making this type of mistake.

2. From the model, $\bar{Y} = \alpha + \beta\bar{X} + \bar{\mu}$. Therefore, $\tilde{\beta} = \bar{Y} / \bar{X} = \beta + (\alpha + \bar{\mu}) / \bar{X}$. Taking the expected value and noting, as before, that X is non-random and that $E(\bar{\mu}) = 0$, we have $E(\tilde{B}) = \beta + \alpha / \bar{X} \neq \beta$. Therefore, $\tilde{B}$ is biased. If $\alpha = 0$ then the estimator would be unbiased.

3. a. (1) The constant term is and estimate of the expected average life insurance a family has when the family income is zero.

a. (2) The coefficient on income is the expected average change in life insurance for each one dollar increase in family income.

a. (3) The value of $\hat{\alpha} + \hat{\beta}x_0$ is an estimate of the expected average life insurance when the family income is x_0.

a. (4) The value of R^2 is a measure of the fraction of the variation in life insurance explained by the model. It is also the square of the correlation coefficient between life insurance and the average value predicted by the estimated equation.

b. (1) The population regression function is $\alpha + \beta income$.

b. (2) As explained Section 3.1, the population error terms arise because of omitted variables, nonlinearities, measurement errors, and unpredictable effects.

c. (1) Unbiasedness of an estimated coefficient means that, although the estimates will differ in repeated trials, the average of those estimates over a large number of trails will be the true population mean. For a regression line, unbiasedness means that the average of the estimated relations from repeated trials will be the population regression function given in 2a.

c. (2) For unbiasedness, we need Assumption 3.3 that $E(\mu_t) = 0$ and Assumption 3.4 that the Xs (which are income values in this example) are given and non-random, of that $Cov(X_t, \mu_t) = 0$, that is, that income and the error term are uncorrelated.

c. (3) The assumption that $E(u_t) = 0$ is not likely to hold here because u_t captures the effects of important omitted variables such as the size of the family and the age distribution of any children. These effects are not likely to be zero.

d. (1) $H_0 : \beta = 5, H_1 : \beta < 5.$

d. (2) The test statistic is given by

$$t_c = \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} = \frac{3.880186 - 5}{0.112125} = -9.987$$

where $\hat{\beta}$ is the estimate of β and $s_{\hat{\beta}}$ is the estimate of the standard error the $\hat{\beta}$. Because the alternative is one-sided, we use a one-tailed t-test. The d.f. are n-2 =18 and the 5% critical value is 1.734. since $t_c < 1.734$, we reject the null hypothesis.

d. (3) The conclusion is that the observed estimate of β is significantly below 5.

d. (4) For a 95% confidence interval we need $t^*_{18}(0.025)$ which is 2.101. We have

$$\hat{\beta} \pm (t^* \times s_{\hat{\beta}}) = 3.880186 \pm (2.101 \times 0.112125) = (3.645, 4.116).$$

e. (1) Because the maximum likelihood method gives the same answers as the OLS procedure, the methodology is sound.

e. (2) Yes, as mentioned earlier, important variables such as the size and age distribution of the children should be included as added variables in the model. Also the wealth of asset position of the family would be important because the higher the wealth of a family the less the need for life insurance.

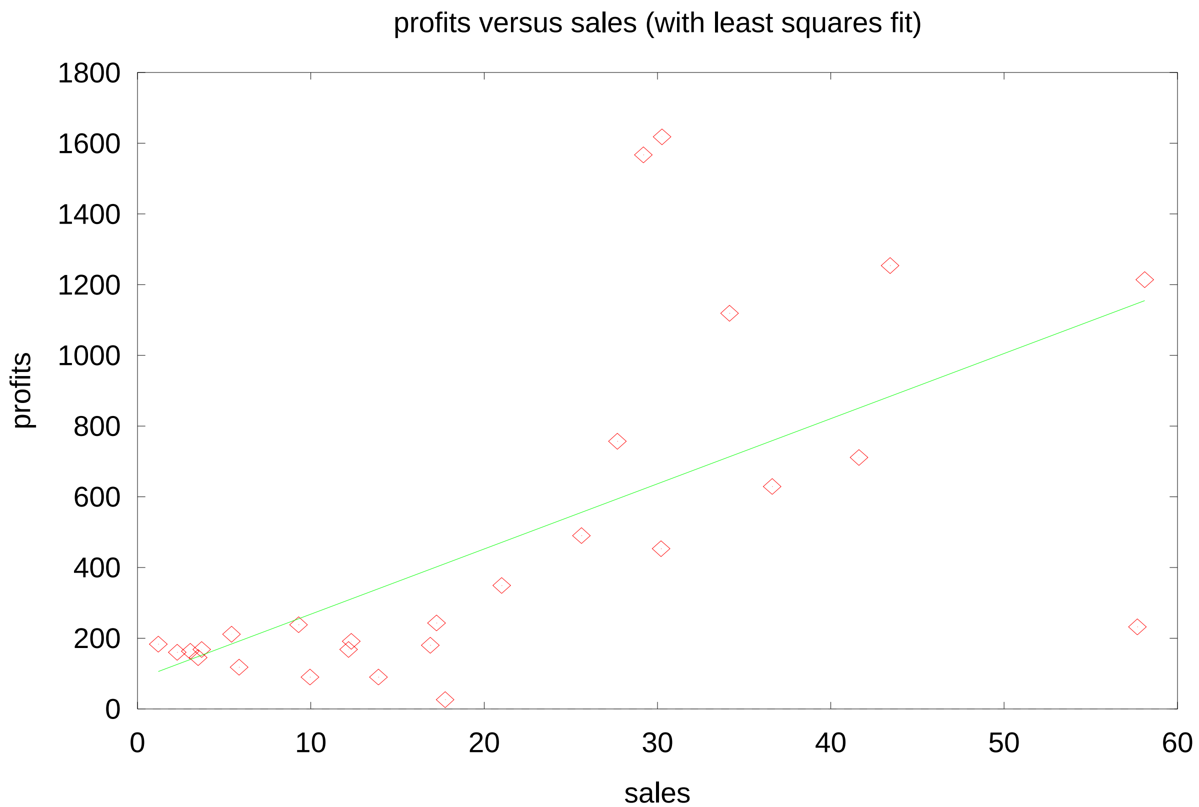
4. a. The GRETL output is as follows:

```
MODEL 1: OLS estimates using the 27 observations 1-27
Dependent variable: profits

      VARIABLE      COEFFICIENT      STDERROR       T STAT    2Prob(t > |T|)

   0)    const          83.5753      118.1309        0.707       0.485812
   2)    sales          18.4338        4.4463        4.146       0.000340 ***

Mean of dep. var.           472.852  S.D. of dep. variable           474.470
Error Sum of Sq (ESS)    3.4685e+006 Std Err of Resid. (sgmahat)    372.4780
Unadjusted R-squared          0.407  Adjusted R-squared                0.384
F-statistic (1, 25)          17.188  p-value for F()                0.000340
Durbin-Watson stat.           1.550  First-order autocorr. coeff       0.210
```

b. The scatterplot is below. The fit does not look particularly good, especially at high levels of sales. These outliers tend to skew the results. The R-squared value of .407 reflects this and is common in cross-sectional data such as these.

c. These are shown above.

d. These are easy to do, since the computer has done most of the work for us. In each case the null hypothesis is that the parameters are equal to zero. The alternative hypothesis is that they are nonzero. In the case of the constant, we cannot reject the null. In the case of the slope, we can reject the null and can do so at a very high level of significance (less than 1%).



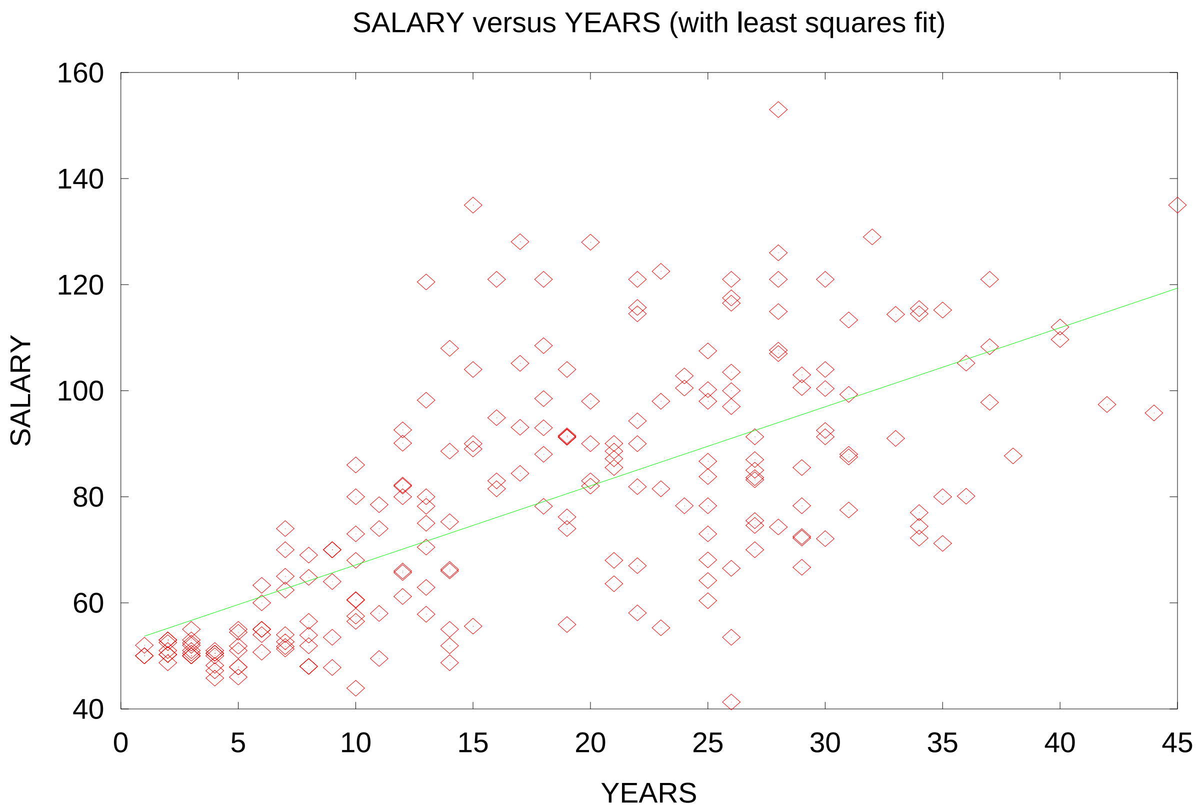
e. All we are doing here is scaling the variables. The result will be changes in the coefficient values, as well as their standard errors, but no change in the t- or F-statistics nor in the R-squared value.

f. Here we might add costs and perhaps some other variables.

5. The results are presented below.

```
MODEL 2: OLS estimates using the 222 observations 1-222
Dependent variable: SALARY

      VARIABLE      COEFFICIENT      STDERROR       T STAT    2Prob(t > |T|)

   0)    const         52.2375         2.3728       22.015        0.000000 ***
   2)    YEARS          1.4911         0.1136       13.131        0.000000 ***

Mean of dep. var.             79.097  S.D. of dep. variable            23.873
Error Sum of Sq (ESS)     70611.3870  Std Err of Resid. (sgmahat)     17.9154
Unadjusted R-squared           0.439  Adjusted R-squared                0.437
F-statistic (1, 220)         172.413  p-value for F()                0.000000
Durbin-Watson stat.            1.346  First-order autocorr. coeff       0.327
```

The scatterplot is shown below and does not show a particularly good fit. The F-statistic has a low p-value, which indicates the model is significant at a high level (less than 1%). In addition, the p-values for both the constant and the slope are very low, also indicating a high level of statistical significance. We could likely improve the fit of the model by including variables for productivity, previous experience, etc. Again, scaling the salary variable would not have any effect on the results.


SALARY versus YEARS (with least squares fit)
SALARY
YEARS
160
140
120
100
80
60
40
0
5
10
15
20
25
30
35
40
45