BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Ms. Yu, Fall 2003

Problem Set 2 Solutions

Problem sets should be your own work. You may work together with classmates, but if you're not figuring this out on your own, you will eventually regret it.

1.

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/ceosal2,clear

. type http://fmwww.bc.edu/ec-p/data/wooldridge/ceosal2.des
CEOSAL2.DES

salary    age       college   grad      comten    ceoten    sales     profits
mktval    lsalary   lsales    lmktval   comtensq  ceotensq  profmarg

  Obs:   177

  1. salary                    1990 compensation, $1000s
  2. age                       in years
  3. college                   =1 if attended college
  4. grad                      =1 if attended graduate school
  5. comten                    years with company
  6. ceoten                    years as ceo with company
  7. sales                     1990 firm sales, millions
  8. profits                   1990 profits, millions
  9. mktval                    market value, end 1990, mills.
 10. lsalary                   log(salary)
 11. lsales                    log(sales)
 12. lmktval                   log(mktval)
 13. comtensq                  comten^2
 14. ceotensq                  ceoten^2
 15. profmarg                  profits as % of sales
```

a.

```
. summarize  salary ceoten

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
      salary |     177    865.8644   587.5893        100       5299
      ceoten |     177    7.954802   7.150826          0         37
```

The average salary of CEOs' is 865.8644. The average tenure of them is 7.954802.

b.

```
. ttest  salary=1000
```

```
One-sample t test

------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  salary |     177    865.8644    44.16591    587.5893    778.7015    953.0274
------------------------------------------------------------------------------
Degrees of freedom: 176

                           Ho: mean(salary) = 1000

    Ha: mean < 1000           Ha: mean ~= 1000           Ha: mean > 1000
       t =  -3.0371              t =  -3.0371              t =  -3.0371
   P < t =   0.0014         P > |t| =   0.0028          P > t =   0.9986
```

We reject the null hypothesis that the average CEO salary is a million dollars, at a 95% confidence level.

c.

```
. ttest  salary, by( college)

Two-sample t test with equal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |       5      1096.2    283.2906    633.4569    309.6593    1882.741
       1 |     172    859.1686    44.74565    586.8337    770.8437    947.4936
---------+--------------------------------------------------------------------
combined |     177    865.8644    44.16591    587.5893    778.7015    953.0274
---------+--------------------------------------------------------------------
    diff |            237.0314    266.7294                -289.389    763.4518
------------------------------------------------------------------------------
Degrees of freedom: 175

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0               Ha: diff ~= 0              Ha: diff > 0
       t =   0.8887              t =   0.8887              t =   0.8887
   P < t =   0.8123         P > |t| =   0.3754          P > t =   0.1877
```

We cannot reject the null hypothesis that CEOs that went to college make as much money as those who did not, at a 95% confidence level.

d.

```
. tabulate grad

       grad |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         83       46.89       46.89
```

```
          1 |         94       53.11      100.00
------------+-----------------------------------
      Total |        177      100.00

. ttest  salary, by( grad)

Two-sample t test with equal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      83    867.7349    74.11516    675.2212    720.2963    1015.174
       1 |      94    864.2128    51.71468    501.3924    761.5177    966.9079
---------+--------------------------------------------------------------------
combined |     177    865.8644    44.16591    587.5893    778.7015    953.0274
---------+--------------------------------------------------------------------
    diff |            3.522174    88.75501               -171.6458    178.6902
------------------------------------------------------------------------------
Degrees of freedom: 175

                      Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0              Ha: diff ~= 0              Ha: diff > 0
       t =   0.0397              t =   0.0397              t =   0.0397
   P < t =   0.5158          P > |t| =   0.9684          P > t =   0.4842
```

We cannot reject the null hypothesis that CEOs attended grad school make as much money as those who did not, at a 95% confidence level.

e.

```
. correlate  profmarg salary
(obs=177)

             | profmarg   salary
-------------+------------------
    profmarg |   1.0000
      salary |  -0.0289   1.0000
```

2. (**2.1**)

(i) Income, age, and family background (such as number of siblings) are just a few possibilities. It seems that each of these could be correlated with years of education. (Income and eduation are probably positively correlated; age and education may be negatively correlated because women in more recent cohorts have, on average, more education; and number of siblings and education are probably negatively correlated.)

(ii) Not if the factors we listed in part (i) are correlated with *educ*. Because we would like to hold these factors fixed, they are part of the error term. But if u is correlated with *educ* then $E(u|educ) \neq 0$, and so SLR.3 fails.

3. (**2.2**) In the equation $y = \beta_0 + \beta_1 x + u$, add and subtract α_0 from the right hand side to get $y = (\alpha_0 + \beta_0) + \beta_1 x + (u - \alpha_0)$. Call the new error $e = u - \alpha_0$, so that $E(e) = 0$. The new intercept is $\alpha_0 + \beta_0$, but the slope is still β_1.

4. (**2.3**)

Note that it would be easiest to do this problem in Stata, typing the data into the Data Editor, and using the `regress` and `predict` commands to generate the desired solutions.

(i) Let $y_i = GPA_i$, $x_i = ACT_i$, and $n = 8$. Then $\overline{x} = 25.875$, $\overline{y} = 3.2125$, $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 5.8125$, and $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 56.875$. From equation (2.9), we obtain the slope as $\hat{\beta}_1 = 5.8125/56.875 \approx .1022$, rounded to four places after the decimal. From (2.17), $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} \approx 3.2125 - (.1022)25.875 \approx .5681$. So we can write

$$\begin{aligned}\widehat{GPA} &= .5681 + .1022ACT \\ n &= 8.\end{aligned}$$

The intercept does not have a useful interpretation because ACT is not close to zero for the population of interest. If ACT is 5 points higher, $\widehat{GPA}$ increases by $.1022(5) = .511$. The effect of five units' increase of the x variable can be calculated after the regression as `display 5*_b[ACT]`.

(ii) The fitted values and residuals — rounded to four decimal places — are given along with the observation number i and GPA in the following table:

i	GPA	$\widehat{GPA}$	$\hat{u}$
1	2.8	2.7143	.0857
2	3.4	3.0209	.3791
3	3.0	3.2253	-.2253
4	3.5	3.3275	.1725
5	3.6	3.5319	.0681
6	3.0	3.1231	-.1231
7	2.7	3.1231	-.4231
8	3.7	3.6341	.0659

You can verify that the residuals, as reported in the table, sum to $-.0002$, which is pretty close to zero given the inherent rounding error. These could be calculated with the two commands `predict GPAhat` and `predict GPAres, resid`.

(iii) When $ACT = 20$, $\widehat{GPA} = .5681 + .1022(20) \approx 2.61$. This can be calculated by adding a 9th observation on ACT in the Data Editor and then doing `predict GPA20 in 9/9`.

(iv) The sum of squared residuals, $\sum_{i=1}^{n}\hat{u}_i^2$ is about .4347 (rounded to four decimal places), and is given in the ANOVA table regression output as the Residual SS. The total sum of squares, $\sum_{i=1}^{n}(y_i - \overline{y})^2$, is about 1.0288, and is given as the Total SS. So the R-squared from the regression is

$$R^2 = 1 - SSR/SST \approx 1 - (.4347/1.0288) \approx .577.$$

Therefore, about 57.7% of the variation in GPA is explained by ACT in this small sample of students.

5. (**2.5**)

(i) The intercept implies that when $inc = 0$, $cons$ is predicted to be negative \$124.84. This, of course, cannot be true, and reflects that fact that this consumption function might be a poor predictor of consumption at very low-income levels. On the other hand, on an annual basis, \$124.84 is not so far from zero.

(ii) Just plug 30,000 into the equation: $\widehat{cons} = -124.84 + .853(30,000) = 25,465.16$ dollars.

(iii) The MPC and the APC are shown in the following graph. Even though the intercept is negative, the smallest APC in the sample is positive. The graph starts at an annual income level of \$1,000 (in 1970 dollars).

6. (**2.8**)

(i) From equation (2.66),

$$\tilde{\beta}_1 = \Big(\sum_{i=1}^{n} x_i y_i\Big) / \Big(\sum_{i=1}^{n} x_i^2\Big).$$

Plugging in $y_i = \beta_0 + \beta_1 x_i + u_i$ gives

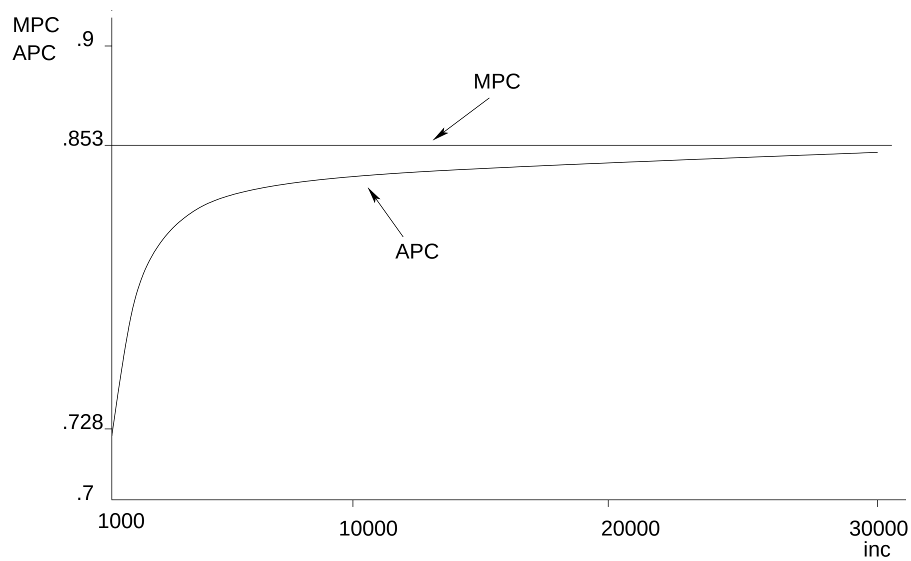
$$\tilde{\beta}_1 = \Big(\sum_{i=1}^{n} x_i(\beta_0 + \beta_1 x_1 + u_i)\Big) / \Big(\sum_{i=1}^{n} x_i^2\Big).$$

After standard algebra, the numerator can be written as

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} x_i u_i.$$

Putting this over the denominator shows we can write $\tilde{\beta}_1$ as

$$\tilde{\beta}_1 = \beta_0 \Big(\sum_{i=1}^{n} x_i\Big) / \Big(\sum_{i=1}^{n} x_i^2\Big) + \beta_1 + \Big(\sum_{i=1}^{n} x_i u_i\Big) / \Big(\sum_{i=1}^{n} x_i^2\Big).$$



Conditional on the x_i, we have

$$E(\tilde{\beta}_1) = \beta_0 \Big(\sum_{i=1}^{n} x_i \Big) / \Big(\sum_{i=1}^{n} x_i^2 \Big) + \beta_1$$

Because $E(u_i) = 0$ for all i. Therefore, the bias in $\tilde{\beta}_1$ is given by the first term in this equation. The bias is obviously zero when $\beta_0 = 0$. It is also zero when $\sum_{i=1}^{n} x_i = 0$, which is the same as $\overline{x} = 0$. In the later case, regression through the origin is identical to regression with an intercept.

7. (**2.11**)

(i)

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/ceosal2,clear

. summarize salary ceoten

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      salary |       177    865.8644    587.5893        100       5299
      ceoten |       177    7.954802    7.150826          0         37
```

Average salary is about 865.864, which means \$865,864 because *salary* is in thousands of dollars. Average *ceoten* is about 7.95.

(ii)

```
. tabulate ceoten

     ceoten |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |          5        2.82        2.82
          1 |         19       10.73       13.56
          2 |         10        5.65       19.21
          3 |         21       11.86       31.07
          4 |         21       11.86       42.94
          5 |         10        5.65       48.59
          6 |         11        6.21       54.80
          7 |          6        3.39       58.19
          8 |         11        6.21       64.41
          9 |          8        4.52       68.93
         10 |          8        4.52       73.45
         11 |          4        2.26       75.71
         12 |          7        3.95       79.66
         13 |          7        3.95       83.62
         14 |          5        2.82       86.44
         15 |          2        1.13       87.57
         16 |          2        1.13       88.70
         17 |          2        1.13       89.83
         18 |          1        0.56       90.40
         19 |          2        1.13       91.53
         20 |          4        2.26       93.79
         21 |          1        0.56       94.35
         22 |          1        0.56       94.92
         24 |          3        1.69       96.61
         26 |          2        1.13       97.74
         28 |          1        0.56       98.31
         34 |          1        0.56       98.87
         37 |          2        1.13      100.00
------------+-----------------------------------
      Total |        177      100.00
```

There are five CEOs with $ceoten = 0$. The longest tenure is 37 years.

(iii)

```
. regress  lsalary ceoten

      Source |       SS       df       MS              Number of obs =     177
-------------+------------------------------           F(  1,   175) =    2.33
       Model |  .850907685     1  .850907685           Prob > F      =  0.1284
    Residual |  63.7953139   175  .364544651           R-squared     =  0.0132
-------------+------------------------------           Adj R-squared =  0.0075
       Total |  64.6462215   176  .367308077           Root MSE      =  .60378

------------------------------------------------------------------------------
     lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ceoten |   .0097236   .0063645     1.53   0.128    -.0028374    .0222847
       _cons |   6.505498   .0679911    95.68   0.000      6.37131    6.639686
------------------------------------------------------------------------------
```

The estimated equation is

$$\log(\widehat{salary}) = 6.51 + 0.097 ceoten$$

$$n = 177, R^2 = .013$$

We obtain the approximate percentage change in *salary* given $\Delta ceoten = 1$ by multiplying the coefficient on *ceoten* by 100, 100(.0097) = .97%. Therefore one more year as CEO is predicted to increase salary by almost 1%.

8. (**2.13**)

(i)

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/wage2,clear

. summarize wage IQ

    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
        wage |     935    957.9455   404.3608        115       3078
          IQ |     935    101.2824   15.05264         50        145
```

Average salary is about $957.95 and average IQ is about 101.28. The sample standard deviation of IQ is about 15.05, which is pretty close to the population value of 15.

(ii)

```
. regress  wage IQ

      Source |       SS       df       MS              Number of obs =     935
```

```
-------------+------------------------------           F(  1,   933) =   98.55
       Model |  14589782.6     1  14589782.6           Prob > F      =  0.0000
    Residual |   138126386   933  148045.429           R-squared     =  0.0955
-------------+------------------------------           Adj R-squared =  0.0946
       Total |   152716168   934  163507.675           Root MSE      =  384.77

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |   8.303064   .8363951     9.93   0.000     6.661631    9.944498
       _cons |   116.9916   85.64153     1.37   0.172    -51.08078    285.0639
------------------------------------------------------------------------------
```

This calls for a level-level model:

$$\widehat{wage} = 116.99 + 8.30IQ$$

$$n = 935, R^2 = .096.$$

An increase in IQ of 15 increases predicted monthly salary by $8.30(15) =$ \$124.50 (in 1980 dollars). IQ score does not even explain 10% of the variation in *wage*.

(iii)

```
. regress lwage IQ

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  1,   933) =  102.62
       Model |  16.4150981     1  16.4150981           Prob > F      =  0.0000
    Residual |  149.241196   933   .15995841           R-squared     =  0.0991
-------------+------------------------------           Adj R-squared =  0.0981
       Total |  165.656294   934  .177362199           Root MSE      =  .39995

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          IQ |   .0088072   .0008694    10.13   0.000      .007101    .0105134
       _cons |   5.886994   .0890206    66.13   0.000      5.71229    6.061698
------------------------------------------------------------------------------
```

This calls for a log–level (single–log) model:

$$\widehat{\log(wage)} = 5.89 + .0088IQ$$

$$n = 935, R^2 = .099.$$

If $\Delta IQ = 15$ then $\Delta\widehat{\log(wage)} = .0088(15) = .132$, which is the (approximate) proportionate change in predicted wage. The percentage increase is therefore approximaely 13.2.

9. (**3.1**)

(i) *hsperc* is defined so that the smaller it is, the higher the student's standing in high school. Everything else equal, the worse the student's standing in high school, the lower is his/her expected college GPA.

(ii) Just plug these value into the equation:

$$\widehat{colgpa} = 1.392 = .0135(20) + .00148(1050) = 2.676.$$

(iii) The difference between A and B is simply 140 times the coefficient on *sat*, because *hsperc* is the same for both students. So A is predicted to have a xcore $.00148(140) \approx .207$ higher.

(iv) With *hsperc* fixed, $\Delta\widehat{colgpa} = .00148\Delta sat$. Now we want to find Δsat such that $\Delta\widehat{colgpa} = .5$, so $.5 = .00148(\Delta sat)$ or $\Delta sat = .5/(.00148) \approx 338$. Perhaps not surprisingly, a large ceteris paribus difference in SAT score – almost two and one-half standard deviations – is needed to obtain a predicted difference in college GPA or a half a point.