8.1 This question looks at log differences and how well they approximate percentage changes.

a. The first question indicates the Sales$_{2001}$ = 196 million and Sales$_{2002}$ = 198 million. You are asked to compare the percentage change versus the $100 \times$ the log difference. In this case we have

$$100 \times \frac{Sales_{2002} - Sales_{2001}}{Sales_{2001}} = 1.020$$

(1)

versus

$$100 \times [\ln(Sales_{2002}) - \ln(Sales_{2001})] = 1.015$$

(2)

In this case, the approximation is quite close.

b. The second part of question asks a similar question, only they ask you to consider three alternative levels for Sales$_{2002}$: 205 million, 250 million and 500 million. For Sales$_{2002}$ = 205 we get:

$$100 \times \frac{Sales_{2002} - Sales_{2001}}{Sales_{2001}} = 4.592$$

(3)

versus

$$100 \times [\ln(Sales_{2002}) - \ln(Sales_{2001})] = 4.490$$

(4)

For Sales$_{2002}$ = 250 we get:

$$100 \times \frac{Sales_{2002} - Sales_{2001}}{Sales_{2001}} = 27.551$$

(5)

versus

$$100 \times [\ln(Sales_{2002}) - \ln(Sales_{2001})] = 24.335$$

(6)

For Sales$_{2002}$ = 500 we get:

$$100 \times \frac{Sales_{2002} - Sales_{2001}}{Sales_{2001}} = 155.102$$

(7)

versus

$$100 \times [\ln(Sales_{2002}) - \ln(Sales_{2001})] = 93.649$$

(8)

c. Clearly, the approximation is quite good when the change is small, but deteriorates as the change increases in size.

8.5 The next question asks you to use the material from the “Demand for Economics Journals” in Section 8.3.

a. In this first question, you are asked to justify the three conclusions reached in the story. In these cases we have:

- The demand for older journals is less elastic than for younger journals because the interaction term between the log of journal age and price per citation is positive.
- There is a linear relationship between log price and log of quantity follows because the estimated coefficients on log price squared and log price cubed are both insignificant.
- The demand is greater for journals with more characters follows from the positive and statistically significant coefficient estimate on the log of characters.

b. This question focuses on the estimated elasticity for 80-year old journals (-0.28).

i. You are first asked how this value is determined. The effect of $ln(Price \text{ per citation})$ is given by $[-0.899 + 0.141 \times ln(Age)] \times ln(Price \text{ per citation})$. Using Age=80, the elasticity is $[-0.899 + 0.141 \times \ln(80)] = 0.28$.

ii. Next, you are asked how the standard error for the estimated elasticity (0.06) would be calculated. As described in equation (8.8) and the footnote on page 263, the standard error can be found by dividing 0.28, the absolute value of the estimate, by the square root of the F-statistic testing $\beta_{ln(Price per citation)} + ln(80) \times \beta_{ln(Age)ln(Price per citation)} = 0$. 

c. Finally, you are asked how the parameters would change if the variable Characters was divided by 1,000 rather than by 1,000,000. Since $\ln(\frac{\text{Characters}}{a}) = \ln(\text{Characters}) - \ln(a)$ for any constant $a$. Thus, estimated parameter on Characters will not change and the constant (intercept) will change.

8.7 This question focuses in on a study of the “gender gap.”

a. In part a, you are given the results from a simple regression of log-earnings on a gender dummy variable.
   i. You are asked to interpret the coefficient on the gender dummy variable, Female. The coefficient in this case indicates that $\ln(\text{Earnings})$ for females are, on average, 0.44 lower for men than for women.
   ii. The error term has a standard deviation of 2.65 (measured in log-points).
   iii. You are asked if the regression suggests that female top executives earn less than top male executives. The answer is “yes,” but the regression does not control for many factors (size of firm, industry, profitability, experience and so forth).
   iv. Finally, you are asked whether the results indicate gender discrimination. In this case, the answer would be “no.” In isolation, these results do imply gender discrimination. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. Thus, it is also important to control for characteristics of the workers that may affect their productivity (education, years of experience, etc.) If these characteristics are systematically different between men and women, then they may be responsible for the difference in mean wages. (If this were true, it would raise an interesting and important question of why women tend to have less education or less experience than men, but that is a question about something other than gender discrimination.) These are potentially important omitted variables in the regression that will lead to bias in the OLS coefficient estimator for Female. Since these characteristics were not controlled for in the statistical analysis, it is premature to reach a conclusion about gender discrimination.

b. In the second part of this question, you are asked to look at a more complex model of executive pay.
   i. The first question here asks you to interpret the coefficient on $\ln(\text{Market Value})$. The coefficient indicates that if MarketValue increases by 1%, then earnings increase by 0.37%
   ii. You are asked here to explain the change in the coefficient on Female from part (a). Female is correlated with the two new included variables and at least one of the variables is important for explaining $\ln(\text{Earnings})$. Thus the regression in part (a) suffered from omitted variable bias.
The two empirical exercises in this homework use the same dataset: TeachingRatings. The data can be downloaded from the Web site listed in the assignment (which you can also reach from the class website). A program that carries all of the tasks for problems in E8.2 is appended to this answer sheet. The parameter estimates for the various models are indicated in the following table.

<table>
<thead>
<tr>
<th>Regressor</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beauty</td>
<td>0.166**</td>
<td>0.160**</td>
<td>0.231**</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.030)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Intro</td>
<td>0.011</td>
<td>0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>OneCredit</td>
<td>0.635**</td>
<td>0.620**</td>
<td>0.657**</td>
</tr>
<tr>
<td></td>
<td>(0.108)</td>
<td>(0.109)</td>
<td>(0.109)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.173**</td>
<td>-0.188**</td>
<td>-0.173**</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.052)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Minority</td>
<td>-0.167*</td>
<td>-0.180**</td>
<td>-0.135</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.069)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>NNEnglish</td>
<td>-0.244**</td>
<td>-0.243*</td>
<td>-0.268**</td>
</tr>
<tr>
<td></td>
<td>(0.094)</td>
<td>(0.096)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>Age</td>
<td>0.020</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Age²</td>
<td>-0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Female × Beauty</td>
<td></td>
<td>-0.141*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.063)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.068**</td>
<td>3.677**</td>
<td>4.075**</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.550)</td>
<td>(0.037)</td>
</tr>
</tbody>
</table>

E8.2

a. The first task you have is to estimate a model of CourseEval as a function of a series of specified variables. This is the base model reported in column (1) of the above table.

b. In the second model, you are asked to add Age and Age². The results are reported in column (2). The coefficient on Age² is not statistically significant, so there is no evidence of a nonlinear effect. The coefficient on Age is not statistically significant and the F-statistic testing whether the coefficients on Age and Age² are zero does not reject the null hypothesis that the coefficients are zero. Thus, Age does not seem to be an important determinant of course evaluations.

c. In the third part, you are asked to modify the model in part (a) to allow for an interaction effect between Beauty and Female. The The coefficient on the interaction term is statistically significant at the 5% level. The magnitude of the coefficient in investigated in parts (d) and (e).

d. Recall that the standard deviation of Beauty is 0.79. Thus Professor Smith’s course rating is expected to increase by 0.231 × (2 × 0.79) = 0.37. The 95% confidence interval for the increase is (0.231 ± 1.96 × 0.048) × (2 × 0.79) or 0.22 to 0.51.

e. Professor Smith’s course rating is expected to increase by (0.231 − 0.173) × (2 × 0.79) = 0.09. To construct the 95% confidence interval, we need the standard error for the sum of coefficients β_{Beauty} + β_{Beauty×Female}. This can be done using the lincom command in Stata. The 95% confidence interval is (0.090 ± 1.96 × 0.040) × (2 × 0.79) or 0.02 to 0.27.
Problem Set #5

Specify the output file

Read in and summarize the data

Estimate the model for question E8.1a and compute earnings effects

```
delimit ;
clear;
cap log close;

# delimit ;
clear;
cap log close;

* Specify the output file
log using Problemset5.log,replace;
set more off;

* Read in and summarize the data
use     CPS04.dta;
describe;
describe;
summarize;

* Estimate the model for question E8.1a and compute earnings effects
reg     ahe age female bachelor,r;
scalar  drop _all;
scalar  Ahea2526a = _b[age];
scalar  Aheb3334a = _b[age];
scalar  list;
```
**********;
*                Estimate the model for question E8.1b and compute earnings effects
*  
**********;
gen lnahe = ln(ahe);
reg lnahe age female bachelor, r;
scalar drop _all;
scalar Ahea2526b = _b[age];
scalar Aheb3334b = _b[age];
scalar list;
gen fitb = _b[_cons] + _b[age]*age + _b[female]*0 + _b[bachelor]*0;

**********;
*                Estimate the model for question E8.1c and compute earnings effects
*  
**********;
gen lnage = ln(age);
reg lnahe lnage female bachelor, r;
scalar drop _all;
scalar Ahea2526c = _b[lnage]*(ln(26)-ln(25));
scalar Aheb3334c = _b[lnage]*(ln(34)-ln(33));
scalar list;
gen fitc = _b[_cons] + _b[lnage]*lnage + _b[female]*0 + _b[bachelor]*0;

**********;
*                Estimate the model for question E8.1d and compute earnings effects
*  
**********;
gen age2 = age^2;
reg lnahe age age2 female bachelor, r;
scalar drop _all;
scalar Ahea2526d = _b[age]+_b[age2]*(26^2 - 25^2);
scalar Aheb3334d = _b[age]+_b[age2]*(34^2 - 33^2);
scalar list;
gen fitd = _b[_cons] + _b[age]*age + _b[age2]*age2 + _b[female]*0 + _b[bachelor]*0;

********************************************************************************************
******;
*
* Plot the fitted values for question E8.1e
*
********************************************************************************************
**************;
twoway (line fitb age, sort) (line fitc age, sort) (line fitd age, sort),
    ytitle(ln(AHE)) xtitle(Age) legend(on);
log close;
clear;
exit;
set more off;

********************************************************************************;

* Read in and summarize the data
> ********************************************************************************;

use CPS04.dta;

describe;

Contains data from CPS04.dta
obs: 7,986
vars: 4
size: 159,720 (84.8% of memory free)

variable name type format label variable label
ahe float %9.0g
bachelor float %9.0g
female float %9.0g
age float %9.0g

Sorted by:

summarize;

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ahe</td>
<td>7986</td>
<td>16.77115</td>
<td>8.758696</td>
<td>2.097902</td>
<td>61.05769</td>
</tr>
<tr>
<td>bachelor</td>
<td>7986</td>
<td>0.4557976</td>
<td>0.4980735</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>female</td>
<td>7986</td>
<td>0.414851</td>
<td>0.4927272</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>age</td>
<td>7986</td>
<td>29.75445</td>
<td>2.891125</td>
<td>25</td>
<td>34</td>
</tr>
</tbody>
</table>

********************************************************************************;

* Estimate the model for question E8.1a and compute earnings effects
> ********************************************************************************;

reg ahe age female bachelor, r;

 Linear regression
 Number of obs = 7986
 F(  3,  7982) = 545.30
 Prob > F = 0.0000
 R-squared = 0.1900
 Root MSE = 7.8843

| Variable | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|---|-----|---------------------|
| ahe      | 0.4392042 | 0.0301511 | 14.57 | 0.000 | 0.3801001 - 0.4983082 |
| age      | 0.0301511 | 0.0000000 | 0.000 | 1.000 | 0.0000000 - 0.0000000 |

Page 1
Estimate the model for question E8.1b and compute earnings effects

**log**:  
```
. gen lnahe = ln(ahe);
. reg lnahe age female bachelor, r;
```

**Results**:
```
Linear regression                                      Number of obs =    7986
F(  3,  7982) =  634.63
Prob > F      =  0.0000
R-squared     =  0.1924
Root MSE      =   .4571
------------------------------------------------------------------------------
     |               Robust
  lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    age |   .0244429   .0017788    13.74   0.000     .0209561    .0279298
  female |  -.1804636   .0103411   -17.45   0.000    -.2007348   -.1601925
  bachelor |   .4052749   .0103623    39.11   0.000     .3849621    .4255877
       _cons |   1.856457   .0535225    34.69   0.000     1.751539    1.961375
------------------------------------------------------------------------------
```

Estimate the model for question E8.1c and compute earnings effects

```
. gen fitb = _b[_cons] + _b[age]*age + _b[female]*0 + _b[bachelor]*0;
```

**log**:  
```
. gen lage = ln(age);
. reg lnahe lage female bachelor, r;
```

**Results**:
```
Linear regression                                      Number of obs =    7986
F(  3,  7982) =  635.95
Prob > F      =  0.0000
R-squared     =  0.1927
Root MSE      =   .4571
------------------------------------------------------------------------------
     |               Robust
  lnahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   lage |   .0517887   .0017788    29.44   0.000     .0482921    .0552853
  female |  -.1804636   .0103411   -17.45   0.000    -.2007348   -.1601925
  bachelor |   .4052749   .0103623    39.11   0.000     .3849621    .4255877
       _cons |   1.856457   .0535225    34.69   0.000     1.751539    1.961375
------------------------------------------------------------------------------
```
Problemset 5. log

Prob > F      =  0.0000
R-squared     =  0.1927
Root MSE      =  .45701

|               Robust     |
|-------------------|-----------------------|
| lnage |       Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval] |
|-------|-----------------|-----------|--------|-----------------|---------------------|
| lnahe |                  |           |        |                 |

| female | -0.1802958   .0103391   -17.44   0.000    -.2005632   -.1600285 |
| bachelor | 0.4052329  .0103599  39.12  0.000    .3849247    .4255413 |

| _cons | 1.282838  .1774524  0.72  0.470   -.2195692    .4761368 |

---

scalar drop _all;
scalar Ahea2526c = _b[lnage]*(ln(26)-ln(25));
scalar Aheb3334c = _b[lnage]*(ln(34)-ln(33));
scalar list;
Aheb3334c = .02163436
Ahea2526c = .02842314

gen fitc = _b[_cons] + _b[lnage]*lnage + _b[female]*0 + _b[bachelor]*0;

---

> * Estimate the model for question E8.1d and compute earnings effects
> *---------------------------------------------------------------;
gen age2 = age^2;
reg lnahe age age2 female bachelor, r;

--

|               Robust     |
|-------------------|-----------------------|
| lnage |       Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval] |
|-------|-----------------|-----------|--------|-----------------|---------------------|
| lnahe |                  |           |        |                 |

age | 0.1470452  .0416466  3.53  0.000    .0654069    .2286835 |
age2 | -0.0020706  .0007025 -2.95  0.003   -.0034475   -.0006936 |
female | -0.1797868  .0103352 -17.40  0.000   -.2000465    -.1595273 |
bachelor | 0.4050769  .0103574  39.11  0.000    .3847737    .4253801 |
_cons | 0.0587333  .6125979  0.10  0.924   -1.142119    1.259585 |

---

scalar drop _all;
scalar Ahea2526d = _b[age]+_b[age2]*(26^2 - 25^2);
scalar Aheb3334d = _b[age]+_b[age2]*(34^2 - 33^2);
scalar list;
Aheb3334d = .00831759
Ahea2526d = .04144657

gen fitd = _b[_cons] + _b[age]*age + _b[age2]*age2 + _b[female]*0 +
Problemset5.log

_b[ bachelor]* 0;

. ******************************************************************************************;
. *
> * Plot the fitted values for question E8.1e
> *
>**************************************************************************************;
. twoway (line fitb age, sort) (line fitc age, sort) (line fitd age, sort),
> ytitle(ln(AHE)) xtitle(Age) legend(on);

.log close;
log: C:\Documents and Settings\jaherrig\My Documents\Classes\Economics 371\Stata\Problemset5.log
log type: text
closed on: 11 Nov 2008, 19:08:01
------------------------------------------------------------------------------------
Page 4