8.2 This question focuses on the hedonic regression model results in Table 8.2.

a. The first part of this question asks you to predict the change in the price of a home from building a 500-square addition to the house. According to the regression results in column (1), the house price is expected to increase by 21% (= 100% × 0.00042 × 500), assuming all other factors are held constant. The 95% confidence interval for the percentage change is 100% × 500 × (0.00042 ± 1.96 × 0.000038) or [17.276% to 24.724%].

b. Because the regressions in columns (1) and (2) have the same dependent variable, $R^2$ can be used to compare the fit of these two regressions. The log-log regression in column (2) has the higher $R^2$, so it is better so use ln(Size) to explain house prices.

c. With the addition of a swimming pool, the house price is expected to increase by 7.1% (= 100% × 0.0071 × 1). The 95% confidence interval for this effect is 100% × (0.071 ± 0.034) or [0.436% to 13.764%].

d. The addition of a single bedroom is expected to increase the price of a house by 0.36% (= 100% × 0.0036 × 1). However, the effect is not statistically significant at a 5% significance level (with a corresponding t-statistic of only $t^{act} = \frac{0.0036}{0.034} = 0.097 < 1.96$). Note that this coefficient measures the effect of an additional bedroom holding the size of the house constant.

e. The quadratic term ln(Size)^2 is not statistically significant at a 5% significance level (with a corresponding t-statistic in column (4) of only $t^{act} = \frac{0.0078}{0.144} = 0.056 < 1.96$).

f. The expected change in the price when a pool is added to a house with a view is 7.1% (= 100% × 0.071 × 1) when a swimming pool is added to a house without a view and other factors are held constant. The house price is expected to increase by 7.32% (= 100% × (0.071 × 1 + 0.0022 × 1)) when a swimming pool is added to a house with a view and other factors are held constant. The difference in the expected percentage change in price is 0.22%. The difference is not statistically significant at a 5% significance level (with a corresponding t-statistic of only $t^{act} = \frac{0.0022}{0.10} = 0.022 < 1.96$).

8.4 This question focuses on the returns to education.

a. In this first part of the question, you are asked to predict the impact of an additional year of experience on the logarithm of average hourly earnings (AHE) for a male with 16 years of education and 2 years of experience who is from a western state. Using the ideas from the Key Concept 8.1, we have that:

$$\Delta \text{ln}(AHE) = \begin{bmatrix} 1.215 + 0.0899(0) - 0.521(0) + 0.0207(0)(16) + 0.0232(3) - 0.000368(3^2) \\ -0.058(0) - 0.078(0) - 0.030(1) \end{bmatrix}$$

$$= 2.60 - 2.578$$

$$= 0.022 \ (or \ 2.2\%)$$

b. Repeating this exercise for someone with 10 years of experience, we get that:

$$\Delta \text{ln}(AHE) = \begin{bmatrix} 1.215 + 0.0899(0) - 0.521(0) + 0.0207(0)(16) + 0.0232(11) - 0.000368(11^2) \\ -0.058(0) - 0.078(0) - 0.030(1) \end{bmatrix}$$

$$= 2.744 - 2.729$$

$$= 0.015 \ (or \ 1.5\%)$$

c. The results in (a) and (b) are different because the model is nonlinear, including a quadratic expression in experience.
d. The differences in the answers to (a) and (b) are statistically significant since the coefficient on the quadratic term for potential experience is significant at the 1% level.

e. This question asks if your answers would change if the person was a woman. The answer is that it would not. The gender affect would affect the level of ln(AHE), but not the change associated with another year of experience.

f. Finally, you are asked what you would change in terms of the regression equation if you suspected that the effect of experience on earnings was different for men than for women. In this case, you should include interaction terms Female × PotentialExperience and Female × (PotentialExperience)^2.

9.3 The key is that the selected sample contains only employed women. Consider two women, Beth and Julie. Beth has no children; Julie has one child. Beth and Julie are otherwise identical. Both can earn $25,000 per year in the labor market. Each must compare the $25,000 benefit to the costs of working. For Beth, the cost of working is forgone leisure. For Julie, it is forgone leisure and the costs (pecuniary and other) of child care. If Beth is just on the margin between working in the labor market or not, then Julie, who has a higher opportunity cost, will decide not to work in the labor market. Instead, Julie will work in “home production,” caring for children, and so forth. Thus, on average, women with children who decide to work are women who earn higher wages in the labor market.

The empirical exercises in this homework uses the dataset: CPS04. 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 problem is appended to this answer sheet. A summary table for the various regressions in provided below:

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.439</td>
<td>0.024</td>
<td>0.147</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.001)</td>
<td>(0.042)</td>
<td></td>
</tr>
<tr>
<td>Age^2</td>
<td>0.725</td>
<td>0.082</td>
<td></td>
<td>-0.0021</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.0007)</td>
<td></td>
<td>(0.0007)</td>
</tr>
<tr>
<td>ln(Age)</td>
<td>-3.158</td>
<td>-0.180</td>
<td>-0.180</td>
<td>-0.180</td>
</tr>
<tr>
<td></td>
<td>(0.176)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Female</td>
<td>6.865</td>
<td>0.405</td>
<td>0.405</td>
<td>0.405</td>
</tr>
<tr>
<td></td>
<td>(0.185)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Bachelor</td>
<td>1.884</td>
<td>1.856</td>
<td>0.128</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>(0.897)</td>
<td>(0.054)</td>
<td>(0.177)</td>
<td>(0.613)</td>
</tr>
</tbody>
</table>

E8.1 a. The regression results for this question are shown in column (1) of the table. If Age increases from 25 to 26, earnings are predicted to increase by $0.439 per hour. If Age increases from 33 to 34, earnings are predicted to increase by $0.439 per hour. These values are the same because the regression is a linear function relating AHE and Age.

b. The regression results for this question are shown in column (2) of the table. If Age increases from 25 to 26, ln(AHE) is predicted to increase by 0.024. This means that earnings are predicted to increase by 2.4%. If Age increases from 34 to 35, ln(AHE) is predicted to increase by 0.024. This means that earnings are predicted to increase by 2.4%. These values, in percentage terms, are the same because the regression is a linear function relating ln(AHE) and Age.

c. The regression results for this question are shown in column (3) of the table. If Age increases from 25 to 26, then ln(Age) has increased by \( \ln(26) - \ln(25) = 0.0392 \) (or 3.92%). The predicted increase in ln(AHE) is 0.725 × (0.0392) = 0.0284. This means that earnings are predicted to increase by 2.84%. If Age increases from 34 to 35, then ln(Age) has increased by \( \ln(35) - \ln(34) = 0.0290 \) (or 2.90%). The predicted increase in ln(AHE) is 0.725 × (0.0290) = 0.0216. This means that earnings are predicted to increase by 2.16%.

d. When Age increases from 25 to 26, the predicted change in ln(AHE) is \((0.147 \times 26 + 0.0021 \times 26^2) - (0.147 \times 25 + 0.0021 \times 25^2) = 0.0414 \). This means that earnings are predicted to increase by 4.14%. When Age increases from 34 to 35, the predicted change in ln(AHE) is \((0.147 \times 34 + 0.0021 \times 34^2) - (0.147 \times 33 + 0.0021 \times 33^2) = 0.0083 \). This means that earnings are predicted to increase by 0.83%.
e. The regressions differ in their choice of one of the regressors. They can be compared on the basis of the $R^2$. The regression in (3) has a (marginally) higher $R^2$ so it is preferred.

f. The regression in (4) adds the variable $Age^2$ to regression (2). The coefficient on $Age^2$ is statistically significant ($t_{act} = 2.95$), and this suggests that the addition of $Age^2$ is important. Thus, (4) is preferred to (2).

g. The regressions differ in their choice of one of the regressors. They can be compared on the basis of the $R^2$. The regression in (4) has a (marginally) higher $R^2$ so it is preferred.

h. The requested figure is provided below. The regression functions using $Age$ (2) and $\ln(Age)$ (3) are similar. The quadratic regression (4) is different. It shows a decreasing effect of $Age$ on $\ln(AHE)$ as workers age. The regression functions for a female with a high school diploma will look just like these, but they will be shifted by the amount of the coefficient on the binary regressor Female. The regression functions for workers with a bachelor’s degree will also look just like these, but they would be shifted by the amount of the coefficient on the binary variable Bachelor.
* Problem Set #5 *

* Specify the output file *

log using Problemset5.log,replace;
set more off;

* Read in and summarize the data *

use CPS04.dta;
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
*  
*****************************************************************************
utoway (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;

 Variable | Obs  Mean    Std. Dev.      Min        Max
-------------+----------------------------------------
ahe | 7986    16.77115    8.758696   2.097902   61.05769
bachelor | 7986    .4557976    .4980735          0          1
female | 7986     .414851    .4927272          0          1
age | 7986    29.75445    2.891125         25         34

* 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 | .4392042  .0301511  14.57   0.000    .3801001    .4983082
age | .4392042  .0301511  14.57   0.000    .3801001    .4983082
Problemset5.log

<table>
<thead>
<tr>
<th>female</th>
<th>-3.157864</th>
<th>.1755882</th>
<th>-17.98</th>
<th>0.000</th>
<th>-3.502063</th>
<th>-2.813665</th>
</tr>
</thead>
<tbody>
<tr>
<td>bachelor</td>
<td>6.86515</td>
<td>.1850291</td>
<td>37.10</td>
<td>0.000</td>
<td>6.502444</td>
<td>7.227855</td>
</tr>
<tr>
<td>_cons</td>
<td>1.883798</td>
<td>.8972419</td>
<td>2.10</td>
<td>0.036</td>
<td>1.249669</td>
<td>3.642626</td>
</tr>
</tbody>
</table>

. scalar drop _all;
. scalar Ahea2526a = _b[age];
. scalar Aheb3334a = _b[age];
. scalar list;
Aheb3334a =  .43920417
Ahea2526a =  .43920417

. ********************************************************************************;
. * Estimate the model for question E8.1b and compute earnings effects
. ********************************************************************************;
. gen     lnahe = ln(ahe);
. reg     lnahe age female bachelor,r;

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
------------------------------------------------------------------------------
. scalar drop _all;
. scalar Ahea2526b = _b[age];
. scalar Aheb3334b = _b[age];
. scalar list;
Aheb3334b =  .02444295
Ahea2526b =  .02444295
. 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;

Linear regression                                      Number of obs =    7986
F(  3,  7982) =  635.95
### Problemset 5.log

**Prob > F** = 0.0000  
**R-squared** = 0.1927  
**Root MSE** = 0.45701

|          | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|------|-----|---------------------|
| **lnahe** |       |           |      |     |                     |
| **lnage** | .7246973 | .0523212 | 13.85 | 0.000 | .6221341 - .8272605 |
| **female** | -.1802958 | .0103391 | -17.44 | 0.000 | -.2005632 - .1600285 |
| **bachelor** | .4052329 | .0103599 | 39.12 | 0.000 | .3849247 - .425541 |
| **_cons** | .1282838 | .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;

**Linear regression**  
Number of obs = 7986  
F( 4, 7981) = 478.66  
Prob > F = 0.0000  
R-squared = 0.1933  
Root MSE = 0.45688

|          | Coef. | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|------|-----|---------------------|
| **lnahe** |       |           |      |     |                     |
| **age**  | .1470452 | .0416466 | 3.53 | 0.000 | .0654069 - .2286835 |
| **age2** | -.0020706 | .0007025 | -2.95 | 0.003 | -.0034475 - .0006936 |
| **female** | -.1797868 | .0103352 | -17.40 | 0.000 | -.2000465 - .159527 |
| **bachelor** | .4050769 | .0103574 | 39.11 | 0.000 | .3847737 - .4253801 |
| **_cons** | .0587333 | .6125979 | 0.10 | 0.924 | -1.1421199 - 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 +
_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