This question focuses on what’s called a hedonic regression model; i.e., where the sales price of the home is regressed on the various attributes of the home. The idea of this model is to try to explain what attributes increase (or decrease) the value of the home. These types of models are often used for property tax assessments. In this case, the estimated model is:

\[
\hat{Price} = 119.2 + (0.485)BDR + (23.4)Bath + (0.156)Hsize + (0.002)LSize + (0.090)Age - (48.8)Poor,
\]

with \( \hat{R}^2 = 0.72 \) and \( SER = 41.5 \).

a. The first part of this question asks you to predict the increase in home value associated with adding a bathroom. The question notes that this done by taking up part of the existing bathroom (so that the overall square footage for the house is not changed). From the regression model, the expected change is given by the coefficient on Bath, since if we increase Bath by 1, then the expected Price increases by 23.4 (or $23,400, since price is measured in thousands of dollars).

b. The second part of question asks a similar question, only in this case the additional bedroom is not taken out of the family room, but instead represents a net increase in the size of the house (i.e., \( \Delta Hsize = 100 \) and \( \Delta Bath = 1 \)). In this case, the increase in the price of the house would be:

\[
\Delta\hat{Price} = (23.4)\Delta Bath + (0.156)\Delta Hsize = 23.4 + 15.6 = 39
\]

or $39,000.

c. The third part of the question asks you to imagine that the owner of a home allows the home to deteriorate to a Poor rating (i.e., so that the change is \( \Delta Poor = 1 \)). In this case, the increase in the price of the house would be:

\[
\Delta\hat{Price} = (48.8)\Delta Poor = 48.8
\]

or $48,800.

d. Finally, you are asked to compute \( R^2 \) for the regression. This is relatively easy to do, as you are told that the adjusted \( R^2 \), \( \hat{R}^2 \), is 0.72. What we want to know is:

\[
R^2 = 1 - \frac{SSR}{TSS}.
\]

What we know is

\[
\hat{R}^2 = 1 - \frac{n - 1}{n - k - 1} SSR.
\]

Rearranging the above equation, we have

\[
\frac{n - 1}{n - k - 1} SSR = 1 - \hat{R}^2,
\]

or equivalently

\[
\frac{SSR}{TSS} = \frac{n - k - 1}{n - 1} \left[1 - \hat{R}^2\right].
\]

Substituting this expression into (4) yields:

\[
R^2 = 1 - \frac{n - k - 1}{n - 1} \left[1 - \hat{R}^2\right].
\]

Plugging in our values for \( n \), \( k \) and \( \hat{R}^2 \), we get

\[
R^2 = 1 - \left(\frac{220 - 6 - 1}{220 - 1}\right) \left[1 - 0.72\right] = 1 - \frac{213}{219} \left[1 - 0.72\right] = 0.728.
\]
6.6 This question asks you to consider the causal effect of police on crime using data from a random sample of U.S. counties. You are told that a researcher plans to regress county crime rate on the size of the county’s police force.

a. The first question asks you to explain why this is likely to suffer from omitted variables bias and to suggest what variables you might want to add to control for such biases. There is no “right” answer here, but there are likely to be many factors influencing the county’s crime rate, such as
- population density
- income level in the area
- the fraction of young males in the area, etc.

Each of these factors may also be correlated with the regressor (the size of the county’s police force).

b. In the second part of the question, you are asked to determine the likely direction of the bias. I’ll use just one of the possible omitted variables above. Suppose that the crime rate is positively affected by the fraction of young males in the population, and that counties with high crime rates tend to hire more police. In this case, the size of the police force is likely to be positively correlated with the fraction of young males in the population leading to a positive value for the omitted variable bias so that $\hat{\beta}_1 > \beta_1$.

7.7 This question continues question 6.5 above, providing standard errors for the estimated hedonic regression model. Specifically, we now have

$$\hat{Price} = 119.2 + (0.485)BDR + (23.4)Bath + (0.156)Hsize + (0.002)LSize + (0.090)Age - (48.8)Poor$$

$$(23.90) (2.61) (8.94) (0.11) (0.0048) (0.311) (10.5)$$

a. The first part of the question asks if the coefficient on $BDR$ is statistically significantly different from zero. The corresponding p-value for this test is given by:

$$p-value = 2\Phi(-|t_{act}|) = 2\Phi\left(-\frac{0.485}{2.61}\right) = 2\Phi(-0.186) < 2(0.4247) = 0.8494.$$ 

We would not reject the null hypothesis that the bedroom coefficient is zero.

b. The next part of the question notes that a typical five-bedroom house sells for much more that a two-bedroom house. It is then asked whether the result in part (a) is consistent with this result. In this case, the results are consistent, because the coefficient on $BDR$ measures the partial effect of the number of bedrooms holding house size ($Hsize$) constant. Yet, the typical five-bedroom house is much larger than the typical two-bedroom house. Thus, the results in (a) says little about the conventional wisdom.

c. In this part of the question, you are asked to imagine that the homeowner has purchased 2000 square feet from an adjacent lot and to calculate the change in the value of her house. In this case, the only thing that has changed is the lot size. The 99% confidence interval for effect of lot size on price is $2000 \times [0.002 \pm 2.58(0.00048)]$ or 1.52 to 6.48 (in thousands of dollars).

d. Part (d) of the question asks if changes in how the lot size will affect the results; i.e., would another scale be more appropriate here? The fundamental results would not change with a change of scale. Choosing the scale of the variables should be done to make the regression results easy to read and to interpret. If the lot size were measured in thousands of square feet, the estimate coefficient would be 2 instead of 0.002.

e. Finally, you are told that the $F$-statistic for omitting $BDR$ and $Age$ from the regression is $F = 0.08$. Using this information, you are asked to evaluate whether the coefficients on $BDR$ and $Age$ are jointly statistically different from zero at the 10% level. The 10% critical value from the $F_{2,\infty}$ distribution is 2.30. Because 0.08 < 2.30, the coefficients are not jointly significant at the 10% level.
7.9 This question asks you to consider the regression model:

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \]  

(10)

and to use “Approach 2” to transform the regression model so as to allow a \( t \)-statistic test of the hypothesis of interest.

a. The first hypothesis is in fact the same as the one we went through in class, since \( H_0 : \beta_1 - \beta_2 = 0 \) is the same restriction as \( H_0 : \beta_1 = \beta_2 \) One can rewrite the model as:

\[
Y_i = \beta_0 + \beta_1 X_{1i} + [-\beta_2 X_{1i} + \beta_2 X_{2i}] + u_i \\
= \beta_0 + (\beta_1 - \beta_2) X_{1i} + \beta_2(X_{1i} + X_{2i}) + u_i \\
= \beta_0 + \gamma_1 X_{1i} + \beta_2 W_i + u_i
\]

where \( \gamma_1 \equiv \beta_1 - \beta_2 \) and \( W_i \equiv X_{1i} + X_{2i} \). Our hypothesis can be rewritten as

\[ H_0 : \gamma_1 = 0 \text{ versus } H_1 : \gamma_1 \neq 0 \]  

(11)

involving a simple t-test.

b. In this second question, you are asked to test the hypothesis \( H_0 : \beta_1 + a\beta_2 = 0 \). The trick is to add and subtract \( a\beta_2 X_{1i} \) to our original equation, so that the coefficient of \( X_{1i} \) is our term of interest. Specifically, we get:

\[
Y_i = \beta_0 + \beta_1 X_{1i} + [a\beta_2 X_{1i} - a\beta_2 X_{1i}] + \beta_2 X_{2i} + u_i \\
= \beta_0 + [\beta_1 X_{1i} + a\beta_2 X_{1i}] + [\beta_2 X_{2i} - a\beta_2 X_{1i}] + u_i \\
= \beta_0 + (\beta_1 + a\beta_2) X_{1i} + \beta_2(X_{2i} - aX_{1i}) + u_i \\
= \beta_0 + \delta_1 X_{1i} + \beta_2 \tilde{W}_i + u_i
\]

where \( \delta_1 \equiv \beta_1 + a\beta_2 \) and \( \tilde{W}_i \equiv X_{2i} - aX_{1i} \). Our hypothesis can be rewritten as

\[ H_0 : \delta_1 = 0 \text{ versus } H_1 : \delta_1 \neq 0 \]  

(12)

involving a simple t-test.

c. Finally, you are asked to test the hypothesis \( H_0 : \beta_1 + \beta_2 = 1 \), or equivalent \( H_0 : \beta_1 + \beta_2 - 1 = 0 \). This question is a bit trickier, but the idea is the same. We want to get one of the coefficients in the model to match the parameter combination of interest (i.e., \( \beta_1 + \beta_2 - 1 \)). Suppose we add and subtract \( (\beta_2 - 1)X_{1i} \) to our original equation. We then get

\[
Y_i = \beta_0 + \beta_1 X_{1i} + [(\beta_2 - 1)X_{1i} - (\beta_2 - 1)X_{1i}] + \beta_2 X_{2i} + u_i \\
= \beta_0 + [\beta_1 X_{1i} + (\beta_2 - 1)X_{1i}] + [\beta_2 X_{2i} - (\beta_2 - 1)X_{1i}] + u_i \\
= \beta_0 + (\beta_1 + \beta_2 - 1) X_{1i} + \beta_2(X_{2i} - X_{1i}) + X_{1i} + u_i \\
= \beta_0 + \tau_1 X_{1i} + \beta_2 Z_i + X_{1i} + u_i
\]

where \( \tau_1 \equiv \beta_1 + \beta_2 - 1 \) and \( Z_i \equiv X_{2i} - X_{1i} \). We are almost done, except that we have the extra \( X_{1i} \) on the right hand side of our equation. If we simply subtract it from both sides, we get:

\[
Y_i - X_{1i} = \beta_0 + \tau_1 X_{1i} + \beta_2 Z_i + u_i \\
\Rightarrow \\
\tilde{Y}_i = \beta_0 + \tau_1 X_{1i} + \beta_2 Z_i + u_i
\]

where \( \tilde{Y}_i \equiv Y_i - X_{1i} \). Now our hypothesis can be rewritten as

\[ H_0 : \tau_1 = 0 \text{ versus } H_1 : \tau_1 \neq 0 \]  

(13)

involving a simple t-test.
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 E6.1 and E7.2 is appended to this answer sheet.

E4.1  

a. The first task you are asked to do is to regress the course evaluations (CourseEval) on age (Beauty) and to report the estimated slope. The results are as follows:

\[
\hat{\text{Course\_Eval}} = 4.00 + 0.133 \text{Age}, R^2 = 0.04
\]

The slope, then, for this regression is 0.133.

b. Next, you are asked to run an additional regression including some of the other variables in the data set. In my case, I added all of the other variables except Age. The resulting parameter estimates are:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Est.</th>
<th>Standard Error</th>
<th>p - value</th>
</tr>
</thead>
<tbody>
<tr>
<td>beauty</td>
<td>0.166</td>
<td>0.032</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>intro</td>
<td>0.011</td>
<td>0.056</td>
<td>0.840</td>
</tr>
<tr>
<td>onecredit</td>
<td>0.635</td>
<td>0.108</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>female</td>
<td>-0.173</td>
<td>0.049</td>
<td>0.001</td>
</tr>
<tr>
<td>minority</td>
<td>-0.167</td>
<td>0.0677</td>
<td>0.014</td>
</tr>
<tr>
<td>menglish</td>
<td>-0.2447</td>
<td>0.094</td>
<td>0.009</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.068</td>
<td>0.037</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

The fact that the coefficient on beauty does not change substantially with the addition of these other variables suggests that the original regression does not suffer from important omitted variables bias.

c. Finally, you are asked to predict the evaluations of Professor Smith, who is a black male with average beauty and is a native English speaker. This would correspond to:

\[
\hat{\text{Course\_Eval}} = (0.166 \times 0) + (0.011 \times 0) + (0.634 \times 0) + (0.173 \times 0) + (0.167 \times 1) + (0.244 \times 0) + 4.068 = 3.901
\]

E7.1  

This question continues the analysis from E6.1, reported above.

a. Your first task is to construct a 95% confidence interval around the slope coefficient. This can be read directly from the Stata output from the first regression. Specifically, the 95% confidence interval is given by: (0.069,0.197).

b. The second question asks which variables should be included in the regression. The results seem to support the inclusion of all of the regressors except Intro, which has a large p-value (0.840). One could rerun the model excluding this variable, but the results (in terms of the confidence interval are unchanged out to three significant digits: (0.104,0.228).
# delimit ;
clear;
cap log close;

# delimit ;
clear;
cap log close;

Specify the output file
log using Problemset4.log,replace;
set more off;

Read in and summarize the data
use TeachingRatings.dta;
describe;
summarize;

course_eval beauty,r;

course_eval beauty,r;

Estimate the model for question E6.1b
reg course_eval beauty intro onecredit female minority nnenglish,r;
scalar Smith = _b[_cons] + 1*_b[minority];
scalar list;
reg course_eval beauty onecredit female minority nnenglish,r;
log close;
*clear;
*exit;
. set more off;
.
> *       Read in and summarize the data
> *
.
> *       Estimate the model for question E6.1a
> *
.
  Linear regression
  Number of obs =  463
  F(  1,   461) =  16.94
Problem 4.1b

**Problem 4.1b**

Estimate the model for question E6.1b

```
> * Estimates a regression model
> *
>妪* Estimates a regression model
> *
```

```
.* Estimates the model for question E6.1b
> *
```

```
.reg course_eval beauty intro onecredit female minority nnenglish,r;
```

```
Linear regression
    Number of obs = 463
    F(  6,  456) =   17.03
    Prob > F      =  0.0000
    R-squared     =  0.1546
    Root MSE      =  .51351
------------------------------------------------------------------------------
          |               Robust
course_eval |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       beauty |   .1660434   .0314939     5.27   0.000     .1041526    .2279342
      _cons  |   4.072006   .0323999    125.93   0.000     4.008453     4.13556
------------------------------------------------------------------------------
```

```
.scalar Smith = _b[_cons] + 1* _b[minority];
```

```
scalar list;
Smith = 3. 9016734
```

```
.reg course_eval beauty onecredit female minority nnenglish,r;
```

```
Linear regression
    Number of obs = 463
    F(  5,  457) =   20.17
    Prob > F      =  0.0000
    R-squared     =  0.1546
    Root MSE      =  .51297
------------------------------------------------------------------------------
          |               Robust
course_eval |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       beauty |   .1660434   .0314939     5.27   0.000     .1041526    .2279342
      onecredit |   .6413254   .0966112     6.64   0.000     .4514682    .8311826
       female |  -.174755   .0495102    -3.52   0.000    -.2714713   -.0768797
    minority |  -.1647853   .0681909    -2.42   0.016    -.2989034   -.0309382
    nnenglish |  -.2480077   .0928702    -2.67   0.008    -.4305133   -.0655021
      _cons  |   4.072006   .0323999    125.93   0.000     4.008453     4.13556
------------------------------------------------------------------------------
```