1. The three 1st order conditions are:

\[
\begin{align*}
(1) \quad & \quad \frac{\partial \text{SSR}}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} \left( y_i - \hat{\beta}_0 - x_{1i} \hat{\beta}_1 - x_{2i} \hat{\beta}_2 \right) x_{1i} = 0 \\
(2) \quad & \quad \frac{\partial \text{SSR}}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{n} \left( y_i - \hat{\beta}_0 - x_{1i} \hat{\beta}_1 - x_{2i} \hat{\beta}_2 \right) x_{2i} = 0 \\
(3) \quad & \quad \frac{\partial \text{SSR}}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} \left( y_i - \hat{\beta}_0 - x_{1i} \hat{\beta}_1 - x_{2i} \hat{\beta}_2 \right) = 0
\end{align*}
\]

Equation (3) can be reduced to read \[ \sum_{i=1}^{n} \left( y_i - \hat{\beta}_0 - x_{1i} \hat{\beta}_1 - x_{2i} \hat{\beta}_2 \right) = 0 \]. Dividing by \( n \) and solving for \( \hat{\beta}_0 \) we find that \( \hat{\beta}_0 = \overline{y} - \overline{x}_1 \hat{\beta}_1 - \overline{x}_2 \hat{\beta}_2 \) and because we have assumed that \( \overline{y} = \overline{x}_1 = \overline{x}_2 = 0 \) then \( \hat{\beta}_0 = 0 \). Equation (1) can be re-written to read \[ \sum_{i=1}^{n} y_i x_{1i} - \hat{\beta}_0 \sum_{i=1}^{n} x_{1i} - \hat{\beta}_1 \sum_{i=1}^{n} x_{1i}^2 - \hat{\beta}_2 \sum_{i=1}^{n} x_{1i} x_{2i} = 0 \]. Since \( \hat{\beta}_0 = 0 \) and \( \sum_{i=1}^{n} x_{1i} x_{2i} = 0 \) this reduces to \[ \sum_{i=1}^{n} y_i x_{1i} - \hat{\beta}_1 \sum_{i=1}^{n} x_{1i}^2 = 0 \] and therefore \( \hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_{1i}}{\sum_{i=1}^{n} x_{1i}^2} = 120 / 40 = 3 \). Using the same procedure, you can also demonstrate that \( \hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_{2i}^2}{\sum_{i=1}^{n} x_{2i}^2} = 160 / 80 = 2 \).

2. Below are the results for this regression. Given the regression

\[
\ln(\text{weekly earn}) = \beta_0 + \text{age}_i \beta_1 + \text{age}_i^2 \beta_2 + \text{educ}_i \beta_3 + \epsilon_i ,
\]

the derivative with respect to age is

\[
\frac{\partial \ln(\text{weekly earn})}{\partial \text{age}} = \beta_1 + 2 \beta_2 \text{age}
\]

This means that the derivative is a function of age. Given estimates the three derivatives are:

At age 21: \( 0.071 - 2(0.00071)21 = 0.041 \) -- an additional year of age increases ages by 4.1%

At age 35: \( 0.071 - 2(0.00071)35 = 0.021 \) -- and additional year of age increases wages by 2.1%

At age 50: \( 0.071 - 2(0.00071)50 = -0.001 \) -- an additional year of age decreases wages by 0.1%
3. True. Remember, the definition of the $R^2$ is $1 - \text{SSR}/\text{SST}$ – by adding more variables to the system SSR can never go up -- no matter how irrelevant the variables are that are added to the system. The worst that would ever happen by adding more variables is that the computer would set the estimated coefficients for the new variables to zero and obtain the original SSR and hence the original and $R^2$. Therefore, the $R^2$ can only increase when more variables are added to the system.

4. A sample program that generates results for this question is called house_price.do.

Model 1:

Source | SS   df       MS
--------+-------------------
Model   | 942250.712     4 235562.678
Residual| 3086043.86   109 28312.329
Total   | 4028294.57  113 35648.6246

---

price | Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------+--------------------------------------
bedrooms | 26.05118   18.12206     1.44   0.153     -9.866149    61.96852
bathrooms | 109.7691    27.9523     3.93   0.000      54.3685    165.1696
otherrooms | 32.03491   13.73668     2.33   0.022     4.809249    59.26057
age | .3275602   .4960419     0.66   0.510    -1.310699     .6555788
_cons | -14.03946   72.17339    -0.19   0.846   -129.0058    157.0848

a) Remember, house prices are measured in thousands of dollars. Each additional bedroom increase house prices by $26K. Every year increase in age increase house prices by $328.

b) Notice that when sq_feet is added to the model, the coefficients on bedrooms, bathrooms and otherrooms decline so much that the signs are all now negative. This makes sense because sq_feet is positively correlated with these three variables so adding it to the model should decrease the coefficients on the other three variables. To many this was counterintuitive – why would more bedrooms be bad? Remember that the coefficients are assuming all else is held constant. Therefore, how do you get another bedroom “holding square feet” constant? You can only do this by having smaller bedrooms – which home buyers find a negative attribute.

c) Notice that the $R^2$ for model 3 is 0.3903 while the $R^2$ for model 2 is 0.3982, not much of a change. In this sample, once one controls for sq_feet, adding information about the number of rooms does not add much explanatory power to the model

Model 2:

Source | SS   df       MS
--------+-------------------
Model   | 1604241.53     5 320848.306
Residual| 2424053.05   108 22444.9356
Total   | 4028294.57  113 35648.6246

---

price | Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------+--------------------------------------
bedrooms | -21.91485   18.12206    -1.19   0.236    -58.37592    14.54622
### Model 3

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 114</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1572268.9</td>
<td>2</td>
<td>786134.448</td>
<td>F( 2, 111) = 35.53</td>
</tr>
<tr>
<td>Residual</td>
<td>2456025.68</td>
<td>111</td>
<td>22126.3575</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>4028294.57</td>
<td>113</td>
<td>35648.6246</td>
<td>R-squared = 0.3903</td>
</tr>
</tbody>
</table>

| price | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-------|-------|-----------|-------|------|------------------------|
| age   | -0.236 | .418     | -0.56 | 0.573 | -1.064057 .5920842    |
| sq_feet | .1796 | .0214987 | 8.36  | 0.000 | .1370547 .222257     |
| _cons | 40.325 | 46.3245 | 0.87  | 0.386 | -51.46961 132.1204    |

5. A sample program that generates results for this question is on the class web page. The program is called `law_school.do`.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5.34106991</td>
<td>4</td>
<td>1.33526748</td>
<td>F( 4, 90) = 95.30</td>
</tr>
<tr>
<td>Residual</td>
<td>1.2609981</td>
<td>90</td>
<td>0.01401109</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>6.60206802</td>
<td>94</td>
<td>0.07023476</td>
<td>R-squared = 0.8090</td>
</tr>
</tbody>
</table>

| lsalary | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|---------|-------|-----------|-------|------|------------------------|
| lcost   | -0.007 | .0361431  | -0.19 | 0.846 | -0.078843 .0647607    |
| lsat    | 0.018  | .0042339  | 4.23  | 0.000 | 0.0094868 .0263097    |
| rank    | -0.0035 | .0004302 | -8.39 | 0.000 | -0.0044635 .0027543   |
| age     | 0.0005 | .0003653  | 0.73  | 0.466 | -0.0004581 .0009934   |
| _cons   | 8.0383 | .7234791  | 11.11 | 0.000 | 6.601066 9.475701     |

a) The elasticity of salaries with respect to the cost of law school is -0.007 or a 1% increase is cost is estimated to reduce salaries by 0.07 percent.

b) A one unit increase in rank (moving from 5th to 6th for example) is estimated to reduce salaries by .36 percent.

c) Below is the matrix of correlation coefficients. Just like is predicted by the first order conditions, the covariance between the estimated residuals and the x’s is by construction equation to zero

```
<table>
<thead>
<tr>
<th></th>
<th>resl</th>
<th>lsal</th>
<th>lcost</th>
</tr>
</thead>
<tbody>
<tr>
<td>resl</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
d) The correlation coefficient between actual and predicted y is 0.8994 and this number squared is 0.908 which is exactly the $R^2$ in the model.

<table>
<thead>
<tr>
<th></th>
<th>lsalary</th>
<th>pred</th>
</tr>
</thead>
<tbody>
<tr>
<td>lsalary</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>pred</td>
<td>0.8994</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

e) Below are the results when LSAT is removed from the model. Note that the correlation coefficient between lsat and rank is -0.73. We know that ln(salaries) are negatively related to rank and negatively correlated with the lsat so taking rank out of the model would put more weight on the lsat variable in the regression and increase its value, which is exactly what happens. Notice that the coefficient on lsat doubles when school rank is eliminated from the model.

```
* run model deleting lsat from basic model
reg lsalary lcost lsat age
```

```
Source | SS      | df   | MS             | Number of obs = 95
-------|---------|------|-----------------|
Model  | 4.35484336 | 3    | 1.45161445     | F( 3, 91) = 58.78
Residual | 2.24722465 | 91   | .024694776     | Prob > F = 0.0000
Total  | 6.60206802 | 94   | .070234766     | R-squared = 0.6596

|       | Coef.    | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|-------|----------|-----------|-------|-------|----------------------|
| lcost | .0847587 | .0457317  | 1.85  | 0.067 | -.0060817 .1755991 |
| lsat  | .0388551 | .0045385  | 8.56  | 0.000 | .0298399 .0478703  |
| age   | .0015209 | .0004426  | 3.44  | 0.001 | .0006418 .0024001  |
| _cons | 3.469744 | .6323767  | 5.49  | 0.000 | 2.213605 4.725882  |
```

f) Below are the results of part f). Note that when we use the residuals from a regression of lcost on the other covariates from the model in part a) we obtain the exact same coefficient as we do for the beta on lcost in model a. When estimating beta, the regression only uses the portion of x that is NOT predicted by other covariates in the model.

```
predict error_lcost, residual
```

```
. reg lsalary error_lcost
```

```
Source | SS      | df   | MS             | Number of obs = 95
-------|---------|------|-----------------|
Model  | .000532152 | 1    | .000532152     | F( 1, 93) = 0.01
Residual | 6.60153586 | 93   | .070984257     | R-squared = 0.0001
Total  | 6.60206802 | 94   | .070234766     | Root MSE = .26643

|       | Coef.    | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|-------|----------|-----------|-------|-------|----------------------|
| error_lcost | -.0070438 | .0813524  | -0.09 | 0.931 | -.1685935 .1545059  |
| _cons  | 10.55491 | .027335   | 386.13| 0.000 | 10.50063 10.60919   |
```
6. a) Since \( x_{1i} \) is randomly assigned then we expect it to be uncorrelated with all of the possible covariates. As a result, adding these new variables to the model is not expected to change the estimate on \( \hat{\beta}_1 \).

b) In a simple bivariate model, the variance on \( \hat{\beta}_1 \) would be \( \hat{V}(\hat{\beta}_1) = \frac{\hat{\sigma}_{\varepsilon}^2}{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2} \). In the multivariate model where \( \hat{V}(\hat{\beta}_i) = \frac{\hat{\sigma}_{\varepsilon}^2}{(1 - R_i^2) \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2} \), since we expect that \( x_{1i} \) will be uncorrelated with all of the possible covariates, then \( R_i^2 \) should be pretty close to zero and the variance in the multivariate case should look a lot like the variance in the simple bivariate regression model, or \( \hat{V}(\hat{\beta}_1) = \frac{\hat{\sigma}_{\varepsilon}^2}{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2} \). However, recall that \( \hat{\sigma}_{\varepsilon}^2 = \frac{SSE}{(n - k - 1)} \) and adding covariates to the model should reduce the SSE and therefore, if the reduction in SSE is larger than the increase in the change in degrees of freedom, it should reduce the estimated variance on \( \hat{\beta}_1 \). In Random Assignment Clinical Trials, we typically add covariates because they reduce the objective function (SSE) which – hopefully, reduces estimated variances.

7. In a bivariate regression model, we know that \( Var(\hat{\beta}_1) = \frac{\sigma_{\varepsilon}^2}{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2} \) whereas in a multivariate regression model, we know that \( Var(\hat{\beta}_1) = \frac{\sigma_{\varepsilon}^2}{(1 - R_i^2) \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2} \) where \( R_i^2 \) is the \( R^2 \) from a regression of \( x_{1i} \) on \( x_{2i} \).

Note that in results, we see the correlation coefficient between \( x_{1i} \) on \( x_{2i} \) is 0.9994 which means that \( R_i^2 \) should be very close to 1. Therefore, by adding \( x_{2i} \) to the model, a variable highly correlated with \( x_{1i} \), the numerator in \( Var(\hat{\beta}_1) \) in model (2) blows up because \( 1 - R_i^2 \) approaches zero.

8. If Model (2) is the correct model, we know the expected bias generated in model (1) is 
\[ E[\hat{\beta}_1] = \beta_1 + \beta_2 \hat{\delta}_1 \] where \( \hat{\delta}_1 \) is the coefficient from the regression \( x_{2i} = \delta_0 + \delta_1 x_{1i} + \phi \). In this case, we expect that \( \hat{\delta}_1 < 0 \) – people with more medical conditions are less likely to take advantage of the free exercise classes. We are also expect that \( \beta_2 > 0 \) (more poor health conditions tend to increase medical care costs). Therefore, because the product \( \beta_2 \hat{\delta}_1 \) is a negative value, the estimate for \( \hat{\beta}_1 \) would be biased down –by ignoring the fact that healthier people tend to enroll in the exercises classes, we are attributing too much to the exercise class.
If Model (2) is the correct model, we know the expected bias generated in model (1) is therefore
$$E[\hat{\beta}_1] = \beta_1 + \beta_2 \hat{\delta}_1$$
where $\hat{\delta}_1$ is the coefficient from the regression $x_{2i} = \delta_0 + \delta_1 x_{1i} + \epsilon_i$. In this case, we expect that $\hat{\delta}_1 > 0$ -- Higher skilled students will attend better schools. We are also expect that $\beta_2 > 0$ (more skilled students will earn more in the workforce). Therefore, the estimate for $\hat{\beta}_1$ would be biased up --by ignoring the fact that higher test score kids both attend better schools and tend to make higher earnings, we overstate the impact of school quality on earnings.

The correlation coefficients at the end of the printout indicate that $x_2$, $x_3$ and $x_4$ are weakly correlated with $x_1$ at best and therefore, the inclusion of these variables in the model, no matter how well correlated they are with $Y$, will not change the coefficient on $\beta_1$. 