Suggested Answers, Problem Set 2
ECON 30331

Bill Evans
Spring 2016

1. a) \[ \hat{\sigma}_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{250}{100} = 2.5 \]

b) \[ \hat{\rho}(x, y) = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1) \left( \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{0.5} \left( \sum_{i=1}^{n} (y_i - \bar{y})^2 \right)^{0.5}} = \frac{125}{100} \left( \frac{160}{100} \right)^{0.5} = 0.625 \]

\[ (\text{c/d)} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{125}{250} = 0.50 \quad \hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1 = 20 - 30(0.5) = 5 \]

\[ R^2 = 1 - \frac{SSE}{SST_y} \quad , \quad SSE = SST(1 - R^2) \]

e) \[ SSE = \sum_{i=1}^{n} (y_i - \bar{y})^2 = 160 \quad \text{so} \quad SSE = 160 * (1 - 0.2) = 128 \]

2. In class, we demonstrated that the OLS estimate for \( \beta_1 \) can be written as \[ \hat{\beta}_1 = \frac{\hat{\rho}_{xy} \hat{\sigma}_y}{\hat{\sigma}_x} \]

\[ \hat{\beta}_1 = \frac{0.4176(13.37) / 27.27}{0.505} = 0.205 \]

This means that \( dy/dx \) 0.205 and for each additional million dollars in payroll, wins increase by 0.2. Another 15 million will generate \( (15)(0.205) = 3 \) or another 3 wins.

\[ \hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1 = 80.97 - 70.13(0.205) = 66.59 \]

3. a) \[ SST = SSM + SSE \quad \text{so} \quad SSM = SST - SSE = 0.652 - 0.090 = 0.562 \]

b) \[ R^2 = \frac{SSM}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{0.090}{0.652} = 0.862 \]

c) \[ \hat{\beta}_1 = \frac{\hat{\rho}_{xy} \hat{\sigma}_y}{\hat{\sigma}_x} = \frac{0.929 * 0.118}{0.505} = 0.217 \]
d) \( \hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1 \) so \( \bar{y} = \hat{\beta}_0 + \bar{x} \hat{\beta}_1 = 7.77 + 0.5(0.217) = 7.88 \)

e) Root MSE = \( \sqrt{\frac{\hat{\sigma}^2}{N-2}} = \sqrt{\frac{SSE}{N-2}} = \sqrt{\frac{0.089}{46}} = 0.044 \)

```plaintext
.sum x y
```

<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>x</td>
<td>48</td>
<td>.5</td>
<td>.5052912</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>y</td>
<td>48</td>
<td>7.879975</td>
<td>.1177725</td>
<td>7.674153</td>
<td>8.037867</td>
</tr>
</tbody>
</table>

```plaintext
.corr x y
```

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>0.9285</td>
<td>1.000</td>
</tr>
</tbody>
</table>

```plaintext
.reg y x
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>.562059213</td>
<td>1</td>
<td>.562059213</td>
<td>F( 1, 46) = 287.76</td>
</tr>
<tr>
<td>Residual</td>
<td>.08984759</td>
<td>46</td>
<td>.001953208</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>.651906803</td>
<td>47</td>
<td>.013870358</td>
<td>R-squared = 0.8622</td>
</tr>
</tbody>
</table>

|            | y | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|-------------|---|-------|-----------|---|-----|---------------------|
| x           | 0.2164215 | .012758 | 16.96 | 0.000 | 0.1907409 | .2421021 |
| _cons       | 7.771764 | .0090213 | 861.49 | 0.000 | 7.753605 | 7.789923 |

4. There are a number of ways to show this. Here is what I think is the easiest.

\[ \hat{e}_i = y_i - \hat{y}_i \] so \[ y_i = \hat{y}_i + \hat{e}_i \]

Take the means of both sides

\[ \frac{1}{n} \sum_i y_i = \frac{1}{n} \sum_i \hat{y}_i + \frac{1}{n} \sum_i \hat{e}_i \]

Which produces

\[ \bar{y} = \bar{\hat{y}} + \bar{\hat{e}} \] and from the first first-order condition we know that \( \bar{\hat{e}} = 0 \) so \( \bar{y} = \bar{\hat{y}} \).

5. Given the model, \( y_i = \beta_0 + x_i \beta_1 + \epsilon_i \) we know that \( \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = -0.90 \)
For the second model, \( y_i = \beta_0 + x_i^* \beta_1 + \epsilon_i \) where \( x_i^* = x_i / 100 \)

\[
\hat{y}_i^* = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i^* - \bar{x}^*)}{\sum_{i=1}^{n} (x_i^* - \bar{x}^*)^2} = \frac{\sum_{i=1}^{n} (y_i - \bar{y}) \left( \frac{x_i}{100} - \frac{\bar{x}}{100} \right)}{\sum_{i=1}^{n} \left( \frac{x_i}{100} - \frac{\bar{x}}{100} \right)^2} = \frac{1}{100} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - x) = 100 \hat{\beta}_1 = -0.9(100) = -90
\]

For the third model, \( y_i^* = \beta_0 + x_i \beta_1 + \epsilon_i \) where \( y_i^* = y_i / 12 \)

\[
\hat{\alpha}_i^* = \frac{\sum_{i=1}^{n} (y_i^* - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} \left( \frac{y_i^*}{12} - \frac{\bar{y}}{12} \right)(x_i - x)}{\sum_{i=1}^{n} (x_i - x)^2} = \frac{1}{12} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - x) = \frac{\hat{\beta}_1}{12} = -0.9/12 = -0.075
\]

6. Recall that \( \hat{\beta}_0 = \bar{y} - \bar{x} \hat{\beta}_1 \) and if \( \hat{\beta}_1 = 0 \) then \( \hat{\beta}_0 = \bar{y} \). Recall that \( R^2 = \frac{SSM}{SST} \) where 

\[
SSM = \sum_i (\hat{y}_i - \bar{y})^2.
\]

When \( \hat{\beta}_1 = 0 \) then \( \hat{y}_i = \hat{\beta}_0 + x_i \hat{\beta}_1 = \hat{\beta}_0 = \bar{y} \) and then 

\[
SSM = \sum_i (\bar{y} - \bar{y})^2 = 0 \text{ and hence, } R^2=0.
\]
7. The program below will generate the results for this problem

* read in meps_senior
use meps_senior
*describe what is in the data set
desc
* run OLS of totalexp on age
reg totalexp age

Source |       SS       df       MS              Number of obs =    2970
-------------+-----------------------------------------------------------
Model |  3.9532e+09     1  3.9532e+09           Prob > F      =  0.0000
Residual |  5.8710e+11  2968   197808727           R-squared     =  0.0067
-------------+-----------------------------------------------------------
Total |  5.9105e+11  2969   199073579           Root MSE      =   14064

------------------------------------------------------------------------------
totalexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------+---------------------------------------------------------------------
      age |   185.2511   41.43923     4.47   0.000     103.9986    266.5036
      _cons |  -5364.368   3080.472  -1.74   0.082    -11404.44    675.7084

a) \( \hat{\beta}_0 = -5364.4 \) and \( \hat{\beta}_1 = 185.3 \)
b) \( \hat{\beta}_1 = \frac{\partial \text{total exp}}{\partial \text{age}} = 185.25 \) so each additional year of age increases annual medical expenditures by $185.
c) \( \hat{y}_i = \hat{\beta}_0 + x \hat{\beta}_1 = \hat{\beta}_0 - 5364.4 + (70)(185.3) = 7603.1 \)
d) \( \hat{y}_i = \hat{\beta}_0 + x \hat{\beta}_1 = \hat{\beta}_0 - 5364.4 + (71)(185.3) = 7788.4 \)
e) Notice that the difference in the answers between parts c and d is 185.3, which is the estimate for \( \frac{\partial \text{total exp}}{\partial \text{age}} = \hat{\beta}_1 \). This makes sense because as the numbers in parts c) and d) indicate, as a person ages 1 year, expenditures go up by $185.3

8. a. The \( R^2 \) is 0.0067
b. The regression of x on y (\( \text{age}_i = \gamma_0 + \text{totalexp}_i \gamma_1 + v_i \)) produces an \( R^2 \) that also equals 0.0067
c. This takes some effort. \( R^2 = \frac{SSM}{SST} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \). Note that \( \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \).

Now substitute the definition of \( \hat{\beta}_1 \) into the equation above. This produces
\[ R_i^2 = \left( \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)^{\frac{1}{2}}} \right) \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\left(\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})\right)^2}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)\left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)} \]

Now consider the R2 from the second regression \( R_i^2 = \frac{\sum_{i=1}^{n} (\hat{x}_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \). Note that \( \hat{x}_i = \hat{\gamma}_0 + \hat{\gamma}_1 y_i \) so

\[ \sum_{i=1}^{n} (\hat{x}_i - \bar{x})^2 = \hat{\gamma}_1^2 \sum_{i=1}^{n} (y_i - \bar{y})^2. \]

Note that \( \hat{\gamma}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \) so substituting this into the into the definition of \( R_i^2 \) you get exactly \( R_i^2 = \frac{\left(\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})\right)^2}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)\left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)} \).