Suggested Answers, Problem Set 2
ECON 30331

Bill Evans
Spring 2018

1. a) $\hat{\sigma}_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = 2500/100 = 25$

b) $$\hat{\rho}(x,y) = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y-\bar{y})}{(\text{n}-1)}}{\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right)^{0.5}\left(\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}\right)^{0.5}} = \frac{\frac{1500}{100}}{(25)^{0.5}\left(\frac{3600}{100}\right)^{0.5}} = 0.5$$

c/d) $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{1500}{2500} = 0.60 \qquad \hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 = 40 - 60(0.6) = 4$$

e) $$R^2 = 1 - \frac{SSE}{SST_y}, SSE = SST(1-R^2)$$
$$SST_y = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 3600 \quad so \quad SSE = 3600*(1-0.2) = 2880$$

2. In class, we demonstrated that the OLS estimate for β_1 can be written as $\hat{\beta}_1 = \frac{\hat{\rho}_x \hat{\sigma}_y}{\hat{\sigma}_x}$ so

$$\hat{\beta}_1 = \frac{\hat{\rho}_x \hat{\sigma}_y}{\hat{\sigma}_x} = 0.4176(13.37)/27.27 = 0.205$$

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) $SST = SSM + SSE \text{ so } SSM = SST - SSE = 0.652 - 0.090 = 0.562$

b) $R^2 = \frac{SSM}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{0.090}{.652} = 0.862$

c) $\hat{\beta}_1 = \frac{\hat{\rho}_x \hat{\sigma}_y}{\hat{\sigma}_x} = .929*0.118/0.505 = 0.217$

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 = $\sqrt{\hat{\sigma}_\varepsilon^2} = \sqrt{\dfrac{SSE}{N-2}} = \sqrt{\dfrac{0.089}{46}} = 0.044$

```
. sum x y

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           x |        48          .5    .5052912          0          1
           y |        48    7.879975    .1177725   7.674153   8.037867

. corr x y
(obs=48)

             |        x        y
-------------+------------------
           x |   1.0000
           y |   0.9285   1.0000


. reg y x

      Source |       SS       df       MS              Number of obs =      48
-------------+------------------------------           F(  1,    46) =  287.76
       Model |  .562059213     1  .562059213           Prob > F      =  0.0000
    Residual |   .08984759    46  .001953208           R-squared     =  0.8622
-------------+------------------------------           Adj R-squared =  0.8592
       Total |  .651906803    47  .013870358           Root MSE      =   .0442

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .2164215    .012758    16.96   0.000     .1907409    .2421021
       _cons |   7.771764   .0090213   861.49   0.000     7.753605    7.789923
------------------------------------------------------------------------------
```

4. a) $\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$ so $\hat{\beta}_0 = \bar{y}$

b) $R^2 = SSM / SST = \sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2$

$\hat{y}_i = \hat{\beta}_0 + x_i\hat{\beta}_1$ but because $\hat{\beta}_1$=0, $\hat{y}_i = \hat{\beta}_0$ and hence $\bar{\hat{y}} = \hat{\beta}_0$ and $\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2 = 0$

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

$\hat{\varepsilon}_i = y_i - \hat{y}_i$ so $y_i = \hat{y}_i + \hat{\varepsilon}_i$

Take the means of both sides

$(1/n)\sum_i y_i = (1/n)\sum_i \hat{y}_i + (1/n)\sum_i \hat{\varepsilon}_i$

Which produces

$\bar{y} = \bar{\hat{y}} + \bar{\hat{\varepsilon}}$ and from the first first-order condition we know that $\bar{\hat{\varepsilon}} = 0$ so $\bar{y} = \bar{\hat{y}}$.

6. The regression results are below

a) Every year, population in the US increased by about 2.46 million people
b) The R^2 is 0.997
c) This high R^2 means that population changes are highly predictable.
d) US population in 2017 is about 325.5 million
e) Timetrend in 2017 would be 2017-1949=68. The model predicts population in 2017 will be 152.2 + (68)*(2.46) = 319.5. Pretty close to the actual number – only 1.8% off.

```
. reg population trend

      Source |       SS           df       MS      Number of obs   =        51
-------------+----------------------------------   F(1, 49)        =  14327.28
       Model |  66717.7033         1  66717.7033   Prob > F        =    0.0000
    Residual |  228.177828        49  4.65669036   R-squared       =    0.9966
-------------+----------------------------------   Adj R-squared   =    0.9965
       Total |  66945.8811        50  1338.91762   Root MSE        =    2.1579

------------------------------------------------------------------------------
  population |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trend |   2.457194   .0205285   119.70   0.000      2.41594    2.498447
       _cons |   152.1881   .6133413   248.13   0.000     150.9555    153.4206
------------------------------------------------------------------------------
```

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
-------------+------------------------------           F(  1,  2968) =   19.98
       Model |  3.9532e+09     1  3.9532e+09           Prob > F      =  0.0000
    Residual |  5.8710e+11  2968   197808727           R-squared     =  0.0067
-------------+------------------------------           Adj R-squared =  0.0064
       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) $\hat{\beta}_1 = \dfrac{\partial total\exp}{\partial age}$ =185.25 so each additional year of age increases annual medical expenditures by $185.

c) $\hat{y}_i = \hat{\beta}_0 + x_i\hat{\beta}_1 = \hat{\beta}_0 = -5364.4 + (70)(185.3) = 7603.1$

d) $\hat{y}_i = \hat{\beta}_0 + x_i\hat{\beta}_1 = \hat{\beta}_0 = -5364.4 + (71)(185.3) = 7788.4$

e) Notice that the difference in the answers between parts c and d is 185.3, which is the estimate for $\dfrac{\partial total\exp}{\partial age} = \hat{\beta}_1$. 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 ($age_i = \gamma_0 + totalexp_i\gamma_1 + v_i$) produces an R^2 that also equals 0.0067

c. This takes some effort. $R_y^2 = \dfrac{SSM}{SST} = \dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$. Note that $\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 = \hat{\beta}_1^2\sum_{i=1}^{n}(\mathrm{x}_i - \bar{x})^2$.

Now substitute the definition of $\hat{\beta}_1$ into the equation above/. This produces

$$R_y^2 = \left(\frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})\right)^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}\right)\frac{\sum_{i=1}^{n}(\mathrm{x}_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \left(\frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})\right)^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)}\right)$$

Now consider the R2 from the second regression $R_x^2 = \dfrac{\sum_i(\hat{x}_i - \bar{\hat{x}})^2}{\sum_i(x_i - \bar{x})^2}$. Note that $\hat{x}_i = \hat{\gamma}_0 + y_i\hat{\gamma}_1$ so $\sum_i(\hat{x}_i - \bar{\hat{x}})^2 = \hat{\gamma}_1^2\sum_i(\mathrm{y}_i - \bar{y})^2$. Note that $\hat{\gamma}_1 = \dfrac{\sum_i(x_i - \bar{x})(y_i - \bar{y})}{\sum_i(y_i - \bar{y})^2}$ so substituting this into the into the definition of R_x^2 you get exactly $R_x^2 = \left(\dfrac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})\right)^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)}\right)$.

9. Given the model, $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$ we know that $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = -0.90$

For the second model, $y_i = \beta_0 + x_i^*\beta_1 + \varepsilon_i \quad where \quad x_i^* = x_i / 100$

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

For the third model, $y_i^* = \beta_0 + x_i\beta_1 + \varepsilon_i \quad where \quad y_i^* = y_i / 12$

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

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 $R^2 = SSM / SST$ 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 $SSM = \sum_i (\bar{y} - \bar{y})^2 = 0$ and hence, $R^2=0$.