ECN627

Problem Set III

Saumitra Mazumder
500720916
Ryerson University
November 20, 2017

I hereby declare that I am the sole author of this work.

5.2

$$\hat{wage} = 12.52 + 2.12 \times \text{male}, R^2 = 0.06, SER = 4.2$$

a)

$$\hat{wage} = 12.52 + 2.12 \times 0 = 12.52$$
$$\hat{wage} = 12.52 + 2.12 \times 1 = 13.64$$
Hence, the wage gap is $13.64 - 12.52 = 2.12$

b)

$$\begin{aligned} \text{H}_0 &: \text{wage}_1 = \text{wage}_2 \\ \text{H}_1 &: \text{wage}_1 \neq \text{wage}_2 \\ t &= \frac{2.12}{0.36} = 5\frac{8}{9} \\ \text{p-value} &= 2\Phi\left(-\left|5\frac{8}{9}\right|\right) < 0.0001 \end{aligned}$$

Hence at 95% confidence, H_0 is rejected

c)

At 95% confidence:
$$2.12 \pm 1.96 \times 0.36 = (1.41, 2.83)$$

d)

The estimated mean wage of women is 12.52
The estimated mean wage of men is 13.64

e)

$$\hat{wage} = 13.64 - 2.12 \times \text{female}, R^2 = 0.06, SER = 4.2$$

5.5

$$\hat{wage} = 918.0 + 13.9 \times \text{Small Class Size}, R^2 = 0.01, SER = 74.6$$

a)

$test\hat{s}core = 918.0 + 13.9 \times 0 = 918.0$

$test\hat{s}core = 918.0 + 13.9 \times 1 = 931.9$

Hence, the score gap is $931.9 - 918.0 = 13.9$

The regression estimates that a smaller class size increases the test scores by 13.9 points

b)

$$\begin{aligned}
&\mathrm{H_0} : \text{test score}_1 = \text{test score}_2 \\
&\mathrm{H_1} : \text{test score}_1 \neq \text{test score}_2 \\
&t = \frac{13.9}{2.5} = 5.56 > 1.96 \\
&\text{Hence, at 95\% confidence, } \mathrm{H_0} \text{ is rejected.}
\end{aligned}$$

c)

$$\begin{aligned}
&\text{At 99\% confidence:} \\
&13.9 \pm 2.575 \times 2.5 = (7.46, 20.34)
\end{aligned}$$

5.8

$$\hat{Y} = 43.2 + 61.5 \times \mathrm{X}, R^2 = 0.54, SER = 1.52$$

a)

$$\begin{aligned}
&t = \frac{43.2}{10.2} \approx 4.23 \\
&\text{At 95\% confidence:} \\
&43.2 \pm 1.96 \times 10.2 = (23.21, 63.19)
\end{aligned}$$

b)

$$
\begin{aligned}
&\mathrm{H}_0 : \beta_1 = 55 \\
&\mathrm{H}_1 : \beta_1 \neq 55 \\
&t = \frac{55}{7.4} \approx 7.43 > 1.96 \\
&\text{Hence, reject } \mathrm{H}_0
\end{aligned}
$$

c)

$$
\begin{aligned}
&\mathrm{H}_0 : \beta_1 = 55 \\
&\mathrm{H}_1 : \beta_1 > 55 \\
&t = \frac{55}{7.4} \approx 7.43 > -1.645 \\
&\text{Hence, there is not sufficient evidence to reject } \mathrm{H}_0
\end{aligned}
$$

5.9

a)

$$
\begin{aligned}
\bar{\beta} &= \frac{\bar{Y}}{\bar{X}} \\
&= \frac{1}{n \times \bar{X}} \times [Y_1 + ,\ ...,\ + Y_n] \\
&\text{Hence, } \bar{\beta}\text{is a linear function of } Y_1 + ,\ ...,\ + Y_n]
\end{aligned}
$$

b)

$$
\begin{aligned}
E[\bar{\beta}|X_1, ,\ ...,\ + Y_n] &= E\left[\frac{1}{n \times \bar{X}} \times [Y_1 + ,\ ...,\ + Y_n]\,|[\,X_1 + ,\ ...,\ + X_n]\right] \\
&= \frac{\beta \times [X_1 + ,\ ...,\ + X_n]}{n \times \bar{X}} \\
&= \beta
\end{aligned}
$$

E5.1

a i)

```
reg earnings height

     Source |       SS           df       MS              Number of obs   =    17870
```

```
-------------+------------------------------           F(  1, 17868) =  196.46
       Model |  1.4086e+11     1  1.4086e+11           Prob > F      =  0.0000
    Residual |  1.2812e+13 17868   717020563           R-squared     =  0.0109
-------------+------------------------------           Adj R-squared =  0.0108
       Total |  1.2953e+13 17869   724863544           Root MSE      =   26777

------------------------------------------------------------------------------
    earnings |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   707.6716   50.48922    14.02   0.000     608.7078    806.6353
       _cons |  -512.7336   3386.856    -0.15   0.880    -7151.299    6125.832
------------------------------------------------------------------------------
```

p-value $= 0.0000 < 0.01$

Hence, the regression is of significance.

aii)

$$608.7078 \leq \widehat{height} \leq 806.6353$$

b i)

```
reg earnings height if sex == 0

      Source |       SS       df       MS              Number of obs =    9974
-------------+------------------------------           F(  1,  9972) =   26.72
       Model |  1.9194e+10     1  1.9194e+10           Prob > F      =  0.0000
    Residual |  7.1628e+12  9972   718288013           R-squared     =  0.0027
-------------+------------------------------           Adj R-squared =  0.0026
       Total |  7.1820e+12  9973   720140552           Root MSE      =   26801

------------------------------------------------------------------------------
    earnings |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |   511.2222   98.89631     5.17   0.000     317.3654    705.0789
       _cons |   12650.86   6383.741     1.98   0.048     137.4364    25164.28
------------------------------------------------------------------------------
```

p-value $= 0.0000 < 0.01$

Hence, the regression is of significance.

bii)

$$317.3654 \leq \widehat{height} \leq 705.0789$$

c i)

```
reg earnings height if sex ==1

      Source |       SS       df       MS              Number of obs =    7896
-------------+------------------------------           F(  1,  7894) =  168.20
       Model |  1.1965e+11     1  1.1965e+11           Prob > F      =  0.0000
    Residual |  5.6155e+12  7894   711357894           R-squared     =  0.0209
-------------+------------------------------           Adj R-squared =  0.0207
       Total |  5.7351e+12  7895   726423097           Root MSE      =   26671

------------------------------------------------------------------------------
    earnings |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      height |    1306.86   100.7662    12.97   0.000     1109.332    1504.388
       _cons |  -43130.34   7068.481    -6.10   0.000    -56986.43   -29274.25
------------------------------------------------------------------------------
```

p-value $= 0.0000 < 0.01$
Hence, the regression is of significance.

cii)

$$1109.332 \leq \widehat{height} \leq 1504.388$$

d)

$$\mathrm{H}_0 : \beta_{1,m} - \beta_{1,w} = 0$$
$$\mathrm{H}_1 : \beta_{1,m} - \beta_{1,w} \neq 0$$
$$t = \frac{1306.86 - 511.2222}{\sqrt{100.7662^2 + 98.89631^2}} \approx 5.635 > 1.96$$

Hence, there is not suficient evidence to reject H_0

d)

If 4.3 holds, then $E[X_i|u_i] = 0$. If it does not hold, then linear regression is inappropriate.

E5.2

a)

```
 drop in 65
(1 observation deleted)

reg growth tradeshare

      Source |       SS       df       MS              Number of obs =      64
-------------+------------------------------           F(  1,    62) =    2.90
       Model |  9.28031557     1  9.28031557           Prob > F      =  0.0937
    Residual |  198.527844    62  3.20206201           R-squared     =  0.0447
-------------+------------------------------           Adj R-squared =  0.0292
       Total |   207.80816    63  3.29854222           Root MSE      =  1.7894

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  tradeshare |   1.680905   .9873624     1.70   0.094    -.2928046    3.654614
       _cons |   .9574107   .5803727     1.65   0.104    -.2027378    2.117559
------------------------------------------------------------------------------
```

p-value $= 0.0937 < 0.1$

Hence, cannot reject H_0 at 99% or 95% confidence. However can reject H_0 at 90%.

b)

$$\text{p-value} = 0.094 < 0.1$$

c)

$$-.2928046 \leq \widehat{tradeshare} \leq 3.654614$$

6.1

$$\begin{aligned}
1 : \bar{R}^2 &= 1 - \frac{7440 - 1}{7440 - 2 - 1} \times (1 - 0.162) \approx 0.162 \\
2 : \bar{R}^2 &= 1 - \frac{7440 - 1}{7440 - 3 - 1} \times (1 - 0.180) \approx 0.180 \\
3 : \bar{R}^2 &= 1 - \frac{7440 - 1}{7440 - 6 - 1} \times (1 - 0.182) \approx 0.181
\end{aligned}$$

6.2

b)

$$\begin{aligned}
\Delta \widehat{TestScores} &= -5.82 \times 4 \\
&= -23.28
\end{aligned}$$

$$\begin{aligned}
\widehat{TestScores}_{av} &= 520.4 - 5.82 \times 21.4 \\
&= 395.852
\end{aligned}$$

6.3

a)

On average, a worker earns \$ 0.51 more for every year.

b)

$1.87 + 8.32 \times 1 - 3.81 \times 1 + 0.51 \times 29 = 21.17$
Sally predicted to earn \$ 21.17

$1.87 + 8.32 \times 1 - 3.81 \times 1 + 0.51 \times 34 = 23.72$
Betsy predicted to earn \$ 23.72

Hence the difference is \$ 2.55

6.9

Omitted variable bias occurs where X_1 is correlated with the omitted variable and said variable is a determinant of the dependent variable. β_1 does not have a omitted variable bias, since X_1 and X_2 are not correlated.

E6.1

a)

```
. reg birthweight smoker,r

Linear regression                                      Number of obs =    3000
                                                       F(  1,  2998) =   89.21
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0286
                                                       Root MSE      =  583.73

------------------------------------------------------------------------------
             |               Robust
 birthweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |  -253.2284   26.81039    -9.45   0.000     -305.797   -200.6597
       _cons |    3432.06   11.89053   288.64   0.000     3408.746    3455.374
------------------------------------------------------------------------------
```

The estimated effect on smoking is a lose of 253.22 grams

bi)

```
. reg birthweight smoker alcohol nprevist, r

Linear regression                                      Number of obs =    3000
                                                       F(  3,  2996) =   59.48
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0729
                                                       Root MSE      =  570.47

------------------------------------------------------------------------------
             |               Robust
 birthweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
      smoker |  -217.5801   26.10764    -8.33   0.000    -268.7708   -166.3894
     alcohol |  -30.49129   72.59671    -0.42   0.675    -172.8357    111.8531
    nprevist |   34.06991   3.608326     9.44   0.000     26.99487    41.14496
       _cons |   3051.249   43.71445    69.80   0.000     2965.535    3136.962
------------------------------------------------------------------------------
```

The variable smoker may be correlated with alcohol or nprevist. Furthermore, alcohol or nprevist may be determinants of birthweight. If both conditions are satisfied, then omitted bias occurs.

bii) The omitted variable has not substantially changed the effect of smoker on birthweight. Hence, there may not be omitted variable bias.

biii)

```
. reg birthweight smoker alcohol nprevist

      Source |       SS       df       MS              Number of obs =    3000
-------------+------------------------------           F(  3,  2996) =   78.47
       Model |  76610831.2     3  25536943.7           Prob > F      =  0.0000
    Residual |   975009173  2996  325436.974           R-squared     =  0.0729
-------------+------------------------------           Adj R-squared =  0.0719
       Total |  1.0516e+09  2999  350656.887           Root MSE      =  570.47

------------------------------------------------------------------------------
 birthweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |  -217.5801    26.6796    -8.16   0.000    -269.8923   -165.2679
     alcohol |  -30.49129   76.23405    -0.40   0.689    -179.9677    118.9851
    nprevist |   34.06991   2.854994    11.93   0.000     28.47197    39.66786
       _cons |   3051.249   34.01596    89.70   0.000     2984.552    3117.946
------------------------------------------------------------------------------

. predict bhat
(option xb assumed; fitted values)

. sort nprevist alcohol smoker
```

$$\widehat{birthweight} = 3106.228$$

iv)

R-squared = 0.0729

Adj R-squared = 0.0719

The values are similar because there are few extraneous variables in the model.

c)

```
. reg smoker alcohol nprevist

      Source |       SS       df       MS              Number of obs =    3000
-------------+------------------------------           F(  2,  2997) =   38.97
       Model |  11.8897961     2  5.94489803           Prob > F      =  0.0000
    Residual |  457.202204  2997  .152553288           R-squared     =  0.0253
-------------+------------------------------           Adj R-squared =  0.0247
       Total |     469.092  2999  .156416139           Root MSE      =  .39058

------------------------------------------------------------------------------
      smoker |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     alcohol |    .334529   .0518358     6.45   0.000     .2328917    .4361662
    nprevist |  -.0111667    .001944    -5.74   0.000    -.0149785   -.0073549
       _cons |   .3102729   .0225893    13.74   0.000     .2659807    .3545651
------------------------------------------------------------------------------

. predict xhat, resid


. reg birthweight smoker alcohol nprevist

      Source |       SS       df       MS              Number of obs =    3000
-------------+------------------------------           F(  3,  2996) =   78.47
       Model |  76610831.2     3  25536943.7           Prob > F      =  0.0000
    Residual |   975009173  2996  325436.974           R-squared     =  0.0729
-------------+------------------------------           Adj R-squared =  0.0719
       Total |  1.0516e+09  2999  350656.887           Root MSE      =  570.47

------------------------------------------------------------------------------
 birthweight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      smoker |  -217.5801    26.6796    -8.16   0.000    -269.8923   -165.2679
     alcohol |  -30.49129   76.23405    -0.40   0.689    -179.9677    118.9851
```

```
    nprevist |   34.06991   2.854994    11.93   0.000     28.47197    39.66786
       _cons |   3051.249   34.01596    89.70   0.000     2984.552    3117.946
------------------------------------------------------------------------------

. predict yhat, resid


. reg yhat xhat

      Source |       SS       df       MS              Number of obs =    3000
-------------+------------------------------           F(  1,  2998) =    0.00
       Model |           0     1           0           Prob > F      =  1.0000
    Residual |   975009173  2998  325219.871           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared = -0.0003
       Total |   975009173  2999  325111.428           Root MSE      =  570.28

------------------------------------------------------------------------------
        yhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        xhat |   7.07e-07    26.6707     0.00   1.000    -52.29472    52.29472
       _cons |   1.27e-07   10.41185     0.00   1.000    -20.41509    20.41509
------------------------------------------------------------------------------

.
```