**Suggested Answers, Problem Set 5**
**ECON 30331**

**Bill Evans**
**Spring 2018**

1. a) The $R^2$ measures the fraction of the variation in Y explained by the model. In this case, $R^2$=SSM/SST. You are given that SSM = 3.059 but not SST. However, note that SST=SSM+SSE so SST=3.059+3.650 = 6.709. Therefore, $R^2$= 3.059/6.709 = 0.460

b) $\hat{\sigma}_{\varepsilon}^2$ =SSE/(n-k-1). In this case, SSE = 3.650, n=30, k=5, so n-k-1=24, so

$$\hat{\sigma}_{\varepsilon}^2 = SSE/(n-k-1) = 3.650/24 = 0.152$$

c) 99% confidence interval is $\hat{\beta}_1 \pm t_{\alpha/2}(n-k-1)[se(\hat{\beta}_1)]$. $\hat{\beta}_1$=0.0928, $se(\hat{\beta}_1) = 0.0336$ and with 24 degrees of freedom and α=0.01, the appropriate critical value of the t-distribution is 2.797. So $\hat{\beta}_1 \pm t_{\alpha/2}(n-k-1)[se(\hat{\beta}_1)]$=0.0928 ± 2.797(0.0336) = (-0.002, 0.187). Since the 99% confidence interval contains 0, we CANNOT REJECT the null hypothesis.

d) No calculation is necessary. Since a p-value of 0.074 is given, p-value < 0.10 and we can reject the null.

e) $\hat{t} = \hat{\beta}_5 / se(\hat{\beta}_5) = 0.363/0.164 = 2.21$ With 24 degrees of freedom and α=0.05, the appropriate critical value of the t-distribution is 2.064 so, since $|\hat{t}| > t_{\alpha/2}(n-k-1)$ at the 95% confidence level, one CAN REJECT the null that $\beta_2$=0.e)

f) $\hat{F} = \dfrac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)}$. $SSE_r$=4.850, $SSE_u$=3.650, q=3, n-k-1=24, so

$$\hat{F} = \frac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)} = \frac{(4.850-3.650)/3}{3.650/24} = 2.63$$

If the null is correct, the F-test statistic is distributed as an F distribution with 3 and 24 degrees of freedom. The 95% critical value would then be 301 and since $\hat{F} < F_{\alpha}$, we CANNOT REJECT the null.

g) No calculations are necessary. STATA reports the f-test on this null hypothesis as .008 so one can easily reject the null that all coefficients are zero.

2. The F-test is defined as $\hat{F} = \dfrac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)}$. The $R^2$ for the unrestricted model is by definition $R_u^2 = 1-(SSE_u/SST)$ so therefore, $SSE_u = SST(1-R_u^2)$ and likewise $SSE_r = SST(1-R_r^2)$. Note that SST is the same in both the restricted and unrestricted models. Substituting these values into the definition of the F-test

$$\hat{F} = \frac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)} = \frac{[SST(1-R_r^2) - SST(1-R_u^2)]/q}{(SST(1-R_u^2))/(n-k-1)}$$

$$= \frac{[(1-R_r^2)-(1-R_u^2)]/q}{(1-R_u^2)/(n-k-1)} = \frac{(R_u^2 - R_r^2)/q}{(1-R_u^2)/(n-k-1)}$$

3. a. The confidence interval is by definition $\hat{\beta}_1 \pm t_{\alpha/2}(t-k-1)se(\hat{\beta}_1)$. Looking at the printout, $\hat{\beta}_1 = 34.781 \, and \, se(\hat{\beta}_1) = 13.244.$ The regressions has n=24 k=3 and n-k-1=20. The appropriate critical value of the t-distribution is therefore 2.086. Therefore, the 95% confidence interval is $34.781 \pm 2.086(13.244) = (7.15, 62.41).$ Since the interval does not contain zero, we can reject the null.

b. Given a null hypothesis that $H_o$:$\beta_1$=a, the t-statistic is defined as $\hat{t} = \frac{\hat{\beta}_1 - a}{se(\hat{\beta}_1)}$. In the problem, we are given that a=0, $\hat{\beta}_1 = 34.781 \, and \, se(\hat{\beta}_1) = 13.244$ so $\hat{t} = \frac{\hat{\beta}_1 - a}{se(\hat{\beta}_1)} = \frac{34.781}{13.244} = 2.626$. Since $|\hat{t}| > t_{\alpha/2}(n-k-1)$ we can reject the null that $\beta_1$=0.

c. With a 99% confidence level, the critical value of the t-distribution with 20 degrees of freedom is 2.845. In this case, $|\hat{t}| < t_{\alpha/2}(n-k-1)$ so we cannot reject the null.

d. Panel A contains the unrestricted model and Panel B is the restricted model. The F-test is by

$\hat{F} = \frac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)}$ and note that the denominator in the f-test is simple $\hat{\sigma}_\varepsilon^2$ in the unrestricted model, which is label as the MSE or mean squared residual on the printout (46918.9833). In this case, $SSE_u$=938379.666, $SSE_r$=1027703.99, q=2, n-k-1=20.

$$\hat{F} = \frac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)} = \frac{(1027703.99 - 938379.67)/2}{46918.9833} = 0.95$$

The 95% critical value of the F-distribution with 2 and 20 degrees of freedom is 3.49, so since $\hat{F} < F_\alpha$, we cannot reject the null hypothesis.

4. a) We are given the model $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + x_{4i}\beta_4 + \varepsilon_i$ and the null $H_o$: $\beta_1$=(1/2)$\beta_2$=3$\beta_3$. Note that 2$\beta_1$=$\beta_2$ and (1/3)$\beta_1$=$\beta_3$ so substitute these values in above and collect like terms.

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}2\beta_1 + x_{3i}(1/3)\beta_1 + x_{4i}\beta_4 + \varepsilon_i$$
$$y_i = \beta_0 + (x_{1i} + 2x_{2i} + (1/3)x_{3i})\beta_1 + x_{4i}\beta_4 + \varepsilon_i$$
$$y_i = \beta_0 + (x_{5i})\beta_1 + x_{4i}\beta_4 + \varepsilon_i$$
$$where \; x_{5i} = x_{1i} + 2x_{2i} + (1/3)x_{3i}$$

b) The null in this case is $H_o$: $\beta_4$=1- 4$\beta_1$ - $\beta_2$-2$\beta_3$ so substitute 1- 4$\beta_1$ - $\beta_2$-2$\beta_3$ in for $\beta_4$ and collect like term

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + x_{4i}\beta_4 + \varepsilon_i$$
$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + x_{4i}(1-4\beta_1-\beta_2-2\beta_3) + \varepsilon_i$$
$$y_i = \beta_0 + (x_{1i}-4x_{4i})\beta_1 + (x_{2i}-x_{4i})\beta_2 + (x_{3i}-2x_{4i})\beta_3 + x_{4i} + \varepsilon_i$$
$$y_i - x_{4i} = \beta_0 + (x_{1i}-4x_{4i})\beta_1 + (x_{2i}-x_{4i})\beta_2 + (x_{3i}-2x_{4i})\beta_3 + \varepsilon_i$$
$$y_i^* = \beta_0 + x_{1i}^*\beta_1 + x_{2i}^*\beta_2 + x_{3i}^*\beta_3 + \varepsilon_i$$
$$where\ y_i^* = y_i - x_{4i}\ , x_{1i}^* = (x_{1i}-4x_{4i}),\ x_{2i}^* = (x_{2i}-x_{4i}),\ x_{3i}^* = (x_{3i}-2x_{4i})$$

5. A sample program named meps_2005.do that generates results and the log from this program is included on the web page.

a. SSE=10,978.99, $R^2$=0.1193

b. Males have 27.7 percent lower spending than female
a one unit increase in the BMI will increase spending by 2.6%
a 10% increase in income will reduce spending by (0.1)(-0.168)=-0.017 or by 1.7 percent

c. $\hat{t}$ on income is -1.57 and the 95% critical value of the t-distribution with over 3000 degrees of freedom is 1.96 so since $|\hat{t}| < t_{\alpha/2}(n-k-1)$ we cannot reject the null the true parameter is zero.

d. After running the unrestricted model, add the following line to perform the f-test.

test midwest south west

You will see the F-statistic is 3.41. If the null is correct, the test statistic is distributed as an F-distribution with 3 and infinite degrees of freedom and the 95% critical value is 2.60 so we can reject the null.

e. I must admit this is a stupid question on my part. Since you cannot rejected the null at the 95% level, you can also not reject the null at the 99% level.

6.
a. A 1 unit increase in horsepower increases prices by $126
b. A 100% increase in MPG (MPG doubles) will increase price by $6,364
c. All wheel drive vehicles cost $469 more than non-AWD vehicles
d. Sedans cost $1,054 less than trucks
e. SUVs cost $674 more than trucks

7. a. The sample program lottery_example.do generates the results for this problem. The results from the unrestricted model are reported below. Note that the coefficient on inc_pupil, K12_earmark_pupil, and not_earmark_pupil are 0.03, 0.78 and 0.39, respectively. This means that if incomes increase by $1 in the state, 3 cents ends up in school spending. In contrast, for each additional dollar (per pupil) in lottery profits hat are generated, 78 cents ends up in school spending. Finally, each additional dollar in general lottery profits that are not earmarked for schools, 39 cents end up in education.

```
. reg exp_pupil inc_pupil k12_earmark_pupil not_earmark_pupil time

      Source |       SS       df       MS              Number of obs =     682
-------------+------------------------------           F(  4,   677) =  721.52
       Model |   966056076     4   241514019           Prob > F      =  0.0000
```

```
    Residual |   226610390   677  334727.312           R-squared     =  0.8100
-------------+------------------------------           Adj R-squared =  0.8089
       Total |  1.1927e+09   681  1751345.77           Root MSE      =  578.56

------------------------------------------------------------------------------
   exp_pupil |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   inc_pupil |   .0306599    .001052    29.15   0.000     .0285944    .0327254
k12_earmar~l |    .777173   .2447934     3.17   0.002     .2965276    1.257818
not_earmar~l |   .3886631      .1586     2.45   0.015      .077256    .7000702
        time |   30.62801   4.077799     7.51   0.000     22.62136    38.63467
       _cons |   804.2216   112.4977     7.15   0.000     583.3352    1025.108
------------------------------------------------------------------------------
```

b. To test the null that $H_o$: $\beta_{K12\_earmark\_pupil}=1$, we can do this three ways. First, we can use a t-statistic. Given a mull hypothesis that $H_0: \beta_j = a$, which can construct the t-test as $\hat{t} = (\hat{\beta}_j - a)/se(\hat{\beta}_j)$ which in this case equals $\hat{t} = (0.777-1)/0.245 = -0.91.$ The critical value for a t with 677 degrees if freedom at the 95% confidence level is roughly 1.96 and since $|\hat{t}| < 1.96$ we cannot reject the null that $\beta_{K12\_earmark\_pupil}=1$. Note as well that the 95% confidence interval for this parameter includes 1 so using the confidence interval, we cannot reject the null. Finally, we can do an f-test after we estimate the unrestricted model

```
. * test for question b using f-test
. test k12_earmark_pupil=1

 ( 1)  k12_earmark_pupil = 1

       F(  1,   677) =    0.83
            Prob > F =    0.3630
```

c. For some reason – I asked b again. Should be a confidence interval. Sorry.

d. Here is the F test

```
. * test for question c
. test k12_earmark_pupil=inc_pupil

 ( 1) - inc_pupil + k12_earmark_pupil = 0

       F(  1,   677) =    9.26
            Prob > F =    0.0024
```

e. To answer this question, we again construct a post-estimation test in STATA, which is illustrated below. In this case, the p-value is 0.0648 which means that we CANNOT reject the null that these two coefficients are the same. Looking at your f-test table, the critical value for an f with 1 and infinite degrees of freedom and an alpha of 0.05 is 3.84.

```
. test k12_earmark_pupil=not_earmark_pupil

 ( 1)  k12_earmark_pupil - not_earmark_pupil = 0

       F(  1,   677) =    3.42
            Prob > F =    0.0648
```

f. Returning to question d), with a 90% confidence level (alpha=0.1), we would still not be able to reject the null since the p-value (0.0024) is less than 0.10.

8. Run the following program

use klem_chemicals

```
* take logs of all the key variables
gen ql=ln(q)
gen kl=ln(k)
gen el=ln(e)
gen ml=ln(m)
gen ll=ln(l)

* run unresricted model
reg ql kl ll el ml
test kl ll el
test kl+ll+el+ml=1
```

and you produce the following results

```
. reg ql kl ll el ml

      Source |       SS           df       MS      Number of obs   =        46
-------------+----------------------------------   F(4, 41)        =    758.35
       Model |  5.68192371         4  1.42048093   Prob > F        =    0.0000
    Residual |  .076798444        41  .001873133   R-squared       =    0.9867
-------------+----------------------------------   Adj R-squared   =    0.9854
       Total |  5.75872216        45  .127971604   Root MSE        =    .04328

------------------------------------------------------------------------------
          ql |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          kl |   .1201804   .0956387     1.26   0.216     -.072966    .3133267
          ll |   .0256033   .1166182     0.22   0.827    -.2099119    .2611186
          el |   .1099364   .0910128     1.21   0.234    -.0738678    .2937405
          ml |   .7799035   .1476518     5.28   0.000     .4817146    1.078092
       _cons |   .6130931   .8072122     0.76   0.452    -1.017105    2.243291
------------------------------------------------------------------------------

. test kl ll el

 ( 1)  kl = 0
 ( 2)  ll = 0
 ( 3)  el = 0

       F(  3,    41) =    1.72
            Prob > F =    0.1771

. test kl+ll+el+ml=1

 ( 1)  kl + ll + el + ml = 1

       F(  1,    41) =    0.32
            Prob > F =    0.5718
```

b) The coefficient on ln(m) says that if you double materials, output increases by 78%

c) The null hypothesis is tested in the F test above. The $\hat{F}$ is 1.72 and the p-value is 0.1771 so we CANNOT reject the null? Does this make sense? Note that for all three parameters individually, we cannot reject the null that the coefficient equals zero (e.g., $H_o : \beta_l = 0$ ) so it is no surprise that when we test the joint hypothesis we cannot reject the null.

d) The null hypothesis is tested above. The $\hat{F}$ is 0.32 and the p-value is 0.57 so we CANNOT reject the null. Does this make sense? Note that if we add all four coefficients we get 0.120+0.026+0.110+0.780=1.036 – which is real close to 1 and there are large standard errors – so – we cannot reject.

9. Let case 1 be the situation where we have $n_1$=100 observations. In this case, we get the following result

$$\hat{t}_1 = \frac{\hat{\beta}_1(1)}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2(1)}{\sum_{i=1}^{n_1}(x_i - \overline{x})^2}}} = \frac{\hat{\beta}_1(1)}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2(1)}{(n_1 - 1)\hat{\sigma}_x^2(1)}}} = -1.33$$

Where $\hat{\beta}_1(1)$ $\hat{\sigma}_\varepsilon^2(1)$ and $\hat{\sigma}_x^2(1)$ are the parameters for case 1. Now, we want to increase the sample size in the hopes of increase the t-statistic in absolute value to 2.

$$\hat{t}_2 = \frac{\hat{\beta}_1(2)}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2(2)}{\sum_{i=1}^{n_2}(x_i - \overline{x})^2}}} = \frac{\hat{\beta}_1(2)}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2(2)}{(n_2 - 1)\hat{\sigma}_x^2(2)}}} = -2$$

As the sample size grows from $n_1$ to $n_2$, we expect that with a finite sample we will get different estimates for $\hat{\beta}_1(2)$ $\hat{\sigma}_\varepsilon^2(2)$ and $\hat{\sigma}_x^2(2)$. However, we know that $\hat{\beta}_1(2)$ $\hat{\sigma}_\varepsilon^2(2)$ and $\hat{\sigma}_x^2(2)$ are unbiased estimates of the true underlying population values, just like $\hat{\beta}_1(1)$ $\hat{\sigma}_\varepsilon^2(1)$ and $\hat{\sigma}_x^2(1)$ are as well. Therefore, set $\hat{\beta}_1(2) = \hat{\beta}_1(1)$ $\hat{\sigma}_\varepsilon^2(2) = \hat{\sigma}_\varepsilon^2(1)$ and $\hat{\sigma}_x^2(2) = \hat{\sigma}_x^2(1)$. Therefore

$$\frac{\hat{t}_2}{\hat{t}_1} = 2/1.33 = \frac{\frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{(n_2 - 1)\hat{\sigma}_x^2}}}}{\frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{(n_1 - 1)\hat{\sigma}_x^2}}}} = \frac{\sqrt{(n_2 - 1)}}{\sqrt{(n_1 - 1)}}$$

Noting that $n_1$=100 and solving for $n_2$, we get $n_2$=225. Note – standard errors are roughly proportional to the square root of sample size. If we want the t-statistic to increase by a factor of 1.5, we need the sample size to increase by a factor of $1.5^2 = 2.25$. Since $n_1$=100, $n_2$=225.

10. To answer this question, you must first know what the null hypothesis is. You were walking on the tracks so the null hypothesis must be that you do not expect a train to be coming. The train whistle is data – a new piece of information. What does the data suggest? In this case, a Type I error (false positive) is that you get off the tracks but the train is not coming. A Type II error (false negative) is that you stay on the track and a train is actually coming.

11. In the simple bivariate regression $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$ we know the estimate for $\beta_1$ can be written as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

but in this case $x_i = 1$ or 0. There are n observations in the sample and

$n_1 = \sum_{i=1}^{n} x_i$ observations for which $x_i=1$ and $n_0 = \sum_{i=1}^{n}(1-x_i)$ for which $x_i=0$ and $n_1+n_0=n$. Recall also that

$$\bar{y}_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i} \text{ and } \bar{y}_0 = \frac{\sum_{i=1}^{n} y_i(1-x_i)}{\sum_{i=1}^{n}(1-x_i)}$$

Work with the numerator for $\hat{\beta}_1$ first.

$$\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n}(y_i - \bar{y})x_i = \sum_{i=1}^{n} y_i x_i - \bar{y}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i - \bar{y}n_1$$

Note that $\sum_{i=1}^{n} y_i x_i = n_1\bar{y}_1$ and $\bar{y}$, the sample mean of y, is simply a weighted average of $\bar{y}_1$ *and* $\bar{y}_0$ where

$\bar{y} = \frac{n_1}{n}\bar{y}_1 + \frac{n_0}{n}y_0$. Therefore, the numerator can be written as

$$n_1\bar{y}_1 - n_1\left(\frac{n_1}{n}\bar{y}_1 + \frac{n_0}{n}y_0\right) = n_1\bar{y}_1 - \frac{n_1^2}{n}\bar{y}_1 - \frac{n_1n_0}{n}y_0 = \frac{nn_1\bar{y}_1 - n_1^2\bar{y}_1 - n_1n_0\bar{y}_0}{n} = \frac{n_1(n-n_1)\bar{y}_1 - n_1n_0\bar{y}_0}{n}$$

*and because* $n = n_1 + n_0$ *then* $n_0 = n - n_1$ *and the numerator equals*

$$\frac{n_1n_0}{n}\left(\bar{y}_1 - \bar{y}_0\right)$$

Now work with the denominator. Note that $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})x_i = \sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i$

Remember that $n_1 = \sum_{i=1}^{n} x_i$ and since $x_i = 1$ or zero then $\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i = n_i$ so

$$\sum_{i=1}^{n} x_i^2 - \bar{x}n_1 = n_1 - \frac{n_1}{n}(n_1) = n_1 - \frac{n_1^2}{n} = \frac{n_1n - n_1^2}{n} = \frac{n_1(n-n_1)}{n} = \frac{n_1n_0}{n}$$

and therefore

$$\hat{\beta}_1 = \frac{\frac{n_1n_0}{n}\left(\bar{y}_1 - \bar{y}_0\right)}{\frac{n_1n_0}{n}} = \left(\bar{y}_1 - \bar{y}_0\right)$$