Suggested Answers, Problem Set 4
ECON 30331

Dan Hungerman
Fall 2009

1. The F-test is defined as $\hat{F}=\frac{(SSE_r-SSE_u)/q}{SSE_u/(n-k-1)}$. 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

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

2. The results from STATA are below. Given a null hypothesis that H_o:β_1=a, the t-statistic is defined as

$\hat{t}=\frac{\hat{\beta}_1-a}{se(\hat{\beta}_1)}$. We are given a=1 and in the printout below,, $\hat{\beta}_1=1.111685\ and\ se(\hat{\beta}_1)=0.0881146$ so

$$\hat{t}=\frac{\hat{\beta}_1-a}{se(\hat{\beta}_1)}=\frac{1.111685-1}{0.0881146}=1.267$$

There are 49 degrees of freedom in this model and the 95% critical value of a t with 49 degrees of freedom is 2.010. In this case, we cannot reject the null the coefficient is 1.

```
. use state_cig_data

.
. keep if year==1988
(969 observations deleted)

.
. reg retail_price state_tax

      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =  159.17
       Model |  3880.45446     1  3880.45446           Prob > F      =  0.0000
    Residual |  1194.56711    49  24.3789206           R-squared     =  0.7646
-------------+------------------------------           Adj R-squared =  0.7598
       Total |  5075.02157    50  101.500431           Root MSE      =  4.9375

------------------------------------------------------------------------------
retail_price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   state_tax |   1.111685   .0881146    12.62   0.000     .9346125    1.288758
       _cons |   98.35109   1.855326    53.01   0.000     94.62267    102.0795
------------------------------------------------------------------------------

. test state_tax=1

 ( 1)  state_tax = 1

       F(  1,    49) =    1.61
```

```
Prob > F =    0.2110
```

3. 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.

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

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 β_4 and collect like term

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

4. In the simple bivariate regression $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$ we know the estimate for β_1 can be written as

$$\hat\beta_1 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

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

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

Work with the numerator for $\hat\beta_1$ first.

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

Note that $\sum_{i=1}^{n} y_i x_i = n_1\overline{y}_1$ and $\overline{y}$, the sample mean of y, is simply a weighted average of $\overline{y}_1\ and\ \overline{y}_0$ where $\overline{y} = \frac{n_1}{n}\overline{y}_1 + \frac{n_0}{n}y_0$. Therefore, the numerator can be written as

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

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

$$\frac{n_1 n_0}{n}(\bar{y}_1 - \bar{y}_0)$$

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

$\sum_{i=1}^{n}x_i^2 - \bar{x}n_1 = n_1 - \frac{n_1}{n}(n_1) = n_1 - \frac{n_1^2}{n} = \frac{n_1 n - n_1^2}{n} = \frac{n_1(n-n_1)}{n} = \frac{n_1 n_0}{n}$ and therefore

$$\hat{\beta}_1 = \frac{\frac{n_1 n_0}{n}(\bar{y}_1 - \bar{y}_0)}{\frac{n_1 n_0}{n}} = (\bar{y}_1 - \bar{y}_0)$$

5. 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$_o$:β$_1$=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 β$_1$=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.

6. A sample program that generates results and the log from this program are 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 line

   ```
   test midwest south west
   ```

   To perform the f-test. 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 rejected the null at the 95% level, you can also reject the null at the 80% level.

7. a) The R^2 measures the fraction of the variation in Y explained by the model
   In this case, R^2=1-(SSE/SST)=1-(3134.42/14370.98) = 0.782

   b) $\hat{\sigma}_\varepsilon^2$=SSE/(n-k-1). In this case, SSE = 3134.42, n=29, k=3, so n-k-1=25, so
   $\hat{\sigma}_\varepsilon^2 = SSE/(n-k-1) = 3134.42/25 = 125.38$

   c) 95% confidence interval is $\hat{\beta}_1 \pm t_{\alpha/2}(n-k-1)[se(\hat{\beta}_1)]$. $\hat{\beta}_1$=-0.449, $se(\hat{\beta}_1) = 0.164$ and with 25 degrees of freedom and α=0.05, the appropriate critical value of the t-distribution is 2.060. So $\hat{\beta}_1 \pm t_{\alpha/2}(n-k-1)[se(\hat{\beta}_1)] = -0.449 \pm 2.060(.164) = [-0.787, -0.111]$. Since the 95% confidence does not contain 0, we CAN REJECT the null hypothesis.

   d) $\hat{t} = \hat{\beta}_2 / se(\hat{\beta}_2) = 0.668/0.323 = 2.068$. With 25 degrees of freedom and α=0.01, the appropriate critical value of the t-distribution is 2.787 so, since $|\hat{t}| < t_{\alpha/2}(n-k-1)$ at the 99% confidence level, one CANNOT REJECT the null that β_2=0.

   e) $\hat{F} = \frac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)}$. SSE_r=3673.06, SSE_u=3134.42, q=2, n-k-1=25, so

$\hat{F} = \frac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)} = \frac{(3673.06 - 3134.42)/2}{3134.32/25} = 2.15.$ If the null is correct, the F-test statistic is distributed as an F distribution with 2 and 25 degrees of freedom. The 95% critical value would then be 3.39 and since $\hat{F} < F_\alpha$, we CANNOR REJECT the null.