Problem Set 4, ECON 30331
(Due on October 29, 2009)

Dan Hungerman
Fall 2009

1. To test a set of q restrictions in a linear regression model, we use the F-statistic which is constructed as

$$\hat{F} = \frac{(SSE_r - SSE_u)/q}{SSE_u/(n-k-1)}$$

Show that the test statistic can be calculated as

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

Where R_u^2 *and* R_r^2 and the R^2's from the restricted and unrestricted models, respectively.

2. In STATA, load the cigarette tax data, state_cig_data.dta, keep only data for 1988 by typing the statement

```
keep if year==1988
```

then run a regression of retail_price (y) on state_tax (x). Using a t-test, and a 95% confidence level, test the hypothesis that the coefficient on state_tax =1, H_0:β_1=1. What is the correct degree of freedom for the t-test? Search the web and find the CORRECT critical value for the t-statistic (the book table only gives values for 40 and 60 degrees of freedom).

3. Suppose you have a regression of the form $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + x_{4i}\beta_4 + \varepsilon_i$

   a) What would the restricted model look like if one were to test the null hypothesis H_o: $\beta_1=(1/2)\beta_2=3\beta_3$

   b) What would the restricted model look like if one were to test the null hypothesis H_o: $\beta_4=1- 4\beta_1 - \beta_2-2\beta_3$

4. (Kinda hard) Consider a regression of y_i on a dummy variable (x_i). The regression is of the form $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$ and we know that OLS estimate for β_1 is

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

Show that because x_i is a dummy variable that the OLS estimate for β_1 equal to $\hat{\beta}_1 = \bar{y}_1 - \bar{y}_0$. All terms were defined in class.

5. Listed below are results from STATA where using 24 observations, y is regressed on x_1 x_2 x_3 and a constant. I have "whited out" some of the results. Using the results in panel a, answer the following questions:

A) Construct a 95% confidence interval for the parameter on x_1? Using this confidence interval, can you reject or not reject the null hypothesis that H_o: $\beta_1=0$?

B) Using a t-test and a 95% confidence interval, test the null hypothesis H_o: $\beta_1=0$? What is the appropriate value of the t-statistic in this case?

C) How do your results in part b) change if you change the confidence level to 99%?

In panel b) of the results, I report the estimates of a model where y is regressed on x_1 and constant.

D) Using the results from panels a) and b), use and F-test and a 95% confidence level to test the null hypothesis that H_o: $\beta_2= \beta_3=0$. What are the degrees of freedom of the critical value of the F in this context and can you reject or not reject the null?

Panel A

```
. reg y x1 x2 x3

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  3,    20) =    2.89
       Model |  407067.668     3  135689.223           Prob > F      =  0.0608
    Residual |  938379.666    20  46918.9833           R-squared     =  0.3026
-------------+------------------------------           Adj R-squared =  0.1979
       Total |  1345447.33    23  58497.7101           Root MSE      =  216.61

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   34.78137   13.24421
          x2 |  -8.757655   30.83438
          x3 |    .161071   .3664861
       _cons |   83.06376   627.1563
------------------------------------------------------------------------------
```

Panel B

```
. reg y x1

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  1,    22) =    6.80
       Model |  317743.343     1  317743.343           Prob > F      =  0.0161
    Residual |  1027703.99    22  46713.8177           R-squared     =  0.2362
-------------+------------------------------           Adj R-squared =  0.2014
       Total |  1345447.33    23  58497.7101           Root MSE      =  216.13

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |    34.4568   13.21172
       _cons |   24.97369    182.131
------------------------------------------------------------------------------
```

6. On the class web page is a data set named meps_2005.dta. The data set contains 3167 observations on total annual medical expenditures for US adults aged 65 and older. The data set has 9 variables and detailed definitions for these variables are listed below.

Variable	Definition
Totalexp	Annual total expenditures on medical care
Income	Annual family income
Age	Age in years

Educ	Years of education
Male	Dummy variable, =1 if male, =0 otherwise
Bmi	Body mass index (weight in kg/height in cm^2
Srhealth	Self reported health, =1 if excellent, 2=very good, 3=good, 4=fair, and 5=poor
Region	Region of the country, 1= northeast, 2=Midwest, 3=south, 4=west
Race	Categorical variable, 1=white, non-Hispanic, 2=black, non-Hispanic, 3=other race, 4=Hispanic

Generate the following 12 variables:
3 dummy variables for white, black and other race, respectively
4 dummy variables for very good, good, fair and poor health, respectively
The natural log of income (ln_income)
The natural log of total medical expenditures, ln_totalexp
3 dummy variables for Midwest, south and west of the country, respectively.

Run a regression with the dependent variable being ln_totalexp and include 15 covariates plus the constant: age, educ, ln_income, bmi, male, 4 self reported health dummies, 3 race dummies, and 3 region dummies. From this regression, answer the following questions

a) What is the SSE and the R^2 for this model?
b) Provide interpretations (a one unit change in x will produce….) for the following coefficients: male, bmi and ln_income?
c) Using a t-statistic and a 95% confidence level, can you reject or not reject the null that $\beta_{ln\_income}=0$? What is the appropriate value of the critical value for the t-test in this case?
d) Using a 95% confidence level, test the null hypothesis that the regional effects are all zero, $H_o:\beta_{region2}= \beta_{region3}= \beta_{region4}=0$. What is the critical value of the F-distribution in this case? Can you reject or not reject the null.
e) How does your answer for part c) change when you use an 80% confidence level?

7. On the next page are STATA results for two OLS models:

Model 1: $y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \varepsilon_i$
Model 2: $y_i = \beta_0 + x_{1i}\beta_1 + \varepsilon_i$

On some places in the printout, I have "whited-out" some of the results. Please use the results on the next page to answer the following questions. Please show all work.

A) Define what is measured by the R^2 and what is the R^2 for model 1 in this case?
B) What is the estimate for $\hat{\sigma}_\varepsilon^2$ from Model 1?
C) Using the results from Model 1 construct a **95% confidence interval** for the coefficient on $\mathbf{x_1}$. What are the appropriate degrees of freedom and the critical value of the t-distribution used in this case? Using this confidence interval, can you reject or not reject the null hypothesis that the true coefficient on $\mathbf{x_1}$ is zero, H_o: $\beta_1=0$?
D) Using the results from Model 1 and a **99%** confidence level, use **a t-test** to test the null hypothesis that the coefficient coefficient on $\mathbf{x_2}$ is zero, H_0: $\beta_2=0$. What are the appropriate degrees of freedom and the critical value of the t-distribution used in this case? Can you reject or not reject the nul?
E) Using the results from models (1) and (2) and a 95% confidence level, test the null hypothesis that H_o: $\beta_2 = \beta_3=0$. What is the estimate of the F test statistic $(\hat{F})$? Specify the degrees of freedom used in the test and the critical value of the F-distribution used in this test. Can you reject or not reject the null?

Results for Question 7

Model 1

```
. reg y x1 x2 x3

      Source |       SS       df       MS              Number of obs =      29
-------------+------------------------------           F(  3,    25) =   29.87
       Model |                 3                       Prob > F      =  0.0000
    Residual |  3134.42453    25        B              R-squared     =      A
-------------+------------------------------           Adj R-squared =
       Total |  14370.9875    28  513.249554           Root MSE      =  11.197

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.4489035   .1635961                                C
          x2 |   .6675618   .3231477        D
          x3 |  -.2015971   .1935028
       _cons |   665.5063   10.50777    63.33   0.000     643.8651    687.1474
------------------------------------------------------------------------------
```

Model 2

```
. reg y x1

      Source |       SS       df       MS              Number of obs =      29
-------------+------------------------------           F(  1,    27) =   78.64
       Model |                 1                       Prob > F      =  0.0000
    Residual |  3673.06239    27                       R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |  14370.9875    28  513.249554           Root MSE      =  11.664

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |  -.7197798   .0811675
       _cons |   685.7047   4.094316   167.48   0.000     677.3039    694.1056
------------------------------------------------------------------------------
```