Econ 31

An F test example using the 1978 Current Population Survey

```
. use cps78
. summarize

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+--------------------------------------------------------
      ED |     550    12.53636   2.772087          1         18
   SOUTH |     550    .2963636    .457069          0          1
   NONWH |     550    .1036364   .3050657          0          1
    HISP |     550    .0654545   .2475513          0          1
      FE |     550    .3763636    .484914          0          1
    MARR |     550    .6527273   .4765367          0          1
  MARRFE |     550    .1890909   .3919373          0          1
      EX |     550    18.71818   13.34653          1         55
    EXSQ |     550    528.1764   625.4931          1       3025
   UNION |     550    .3054545    .461019          0          1
  LNWAGE |     550    1.681002    .490157       -.47     3.3514
     AGE |     550    36.25455   12.65432         18         64
    NDEP |     550    .9890909      1.286          0          8
   MANUF |     550         .28   .4494076          0          1
  CONSTR |     550    .0727273   .2599247          0          1
   MANAG |     550    .1036364   .3050657          0          1
   SALES |     550    .0527273   .2236919          0          1
    CLER |     550    .2018182   .4017226          0          1
    SERV |     550    .0872727   .2824912          0          1
    PROF |     550         .18   .3845372          0          1
    wage |     550    6.062766   3.257956   .6250023   28.54266
    nwed |     550    1.214545   3.739876          0         18
    edsq |     550    164.8309   68.61239          1        324
```

For our base equation we use the basic human capital equation

```
. regress wage ED EX EXSQ

  Source |       SS       df       MS                  Number of obs =     550
---------+------------------------------                F(  3,   546) =   59.44
   Model |   1434.6765     3  478.225501                Prob > F      =  0.0000
Residual |  4392.56015   546  8.04498196                R-square      =  0.2462
---------+------------------------------                Adj R-square  =  0.2421
   Total |  5827.23665   549  10.6142744                Root MSE      =  2.8364

------------------------------------------------------------------------------
    wage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ED |   .4869247    .047362     10.281   0.000       .3938907    .5799587
      EX |   .1917173   .0342828      5.592   0.000       .1243749    .2590596
    EXSQ |  -.0020404   .0007417     -2.751   0.006      -.0034974   -.0005834
   _cons |  -2.552409   .6890515     -3.704   0.000      -3.905926   -1.198893
------------------------------------------------------------------------------
```

Note that each of the variables is significant at better than the 5% level. Also note the value of R^2 and Adj R^2.

Now we will investigate whether female wage rates are determined differently than males wage rates. To do so we first simply add the dummy variable for female to the base equation.

```
. regress wage ED EX EXSQ FE
```

```
  Source |       SS       df       MS                  Number of obs =     550
---------+------------------------------               F(  4,   545) =   66.18
   Model |  1905.16608     4  476.291519               Prob > F      =  0.0000
Residual |  3922.07058   545  7.19645977               R-square      =  0.3269
---------+------------------------------               Adj R-square  =  0.3220
   Total |  5827.23665   549  10.6142744               Root MSE      =  2.6826

------------------------------------------------------------------------------
    wage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ED |   .4959596   .0448087     11.068   0.000       .4079408    .5839785
      EX |   .1831763   .0324417      5.646   0.000       .1194502    .2469024
    EXSQ |  -.0020183   .0007015     -2.877   0.004      -.0033963   -.0006403
      FE |  -1.922942   .2378212     -8.086   0.000        -2.3901   -1.455784
   _cons |  -1.793759   .6584209     -2.724   0.007      -3.087112   -.5004053
------------------------------------------------------------------------------
```

We see that females earned $1.92 less than males in 1978 and that the difference was statistically significantly different from zero at better than the 5% level.

Now we wish to investigate whether the way characteristics are rewarded are different for females and males or whether there is just a constant difference in wages regardless of characteristics. The previous equation allows only for a constant difference regardless of characteristics.

First we create variables which are interactions of FE with each of the characteristics.

```
. gen EDF=ED*FE

. gen EXF=EX*FE

. gen EXSQF=EXSQ*FE
```

Now we add the new variables to the previous equation.

```
. regress wage ED EX EXSQ FE EDF EXF EXSQF

  Source |       SS       df       MS                  Number of obs =     550
---------+------------------------------               F(  7,   542) =   40.19
   Model |    1991.159     7  284.451286               Prob > F      =  0.0000
Residual |  3836.07765   542  7.07763404               R-square      =  0.3417
---------+------------------------------               Adj R-square  =  0.3332
   Total |  5827.23665   549  10.6142744               Root MSE      =  2.6604

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

```
      ED |     .53327   .0521785     10.220   0.000       .4307732    .6357669
      EX |   .2492387   .0404426      6.163   0.000       .1697953    .3286822
    EXSQ |  -.0030643   .0008552     -3.583   0.000      -.0047443   -.0013844
      FE |    1.86922   1.480418      1.263   0.207      -1.038839    4.777279
     EDF |  -.1589386   .1015433     -1.565   0.118      -.3584052     .04 0528
     EXF |  -.1785797   .0685046     -2.607   0.009      -.3131466   -.0440128
   EXSQF |   .0028976   .0015165      1.911   0.057      -.0000812    .0058765
   _cons |  -2.963195   .7659532     -3.869   0.000      -4.467795   -1.458594
------------------------------------------------------------------------------
```

First we do an F test to see whether the new equation is significantly different from the base equation. The null hypothesesis is that $B_{fe}=B_{edf=}B_{exf}=B_{exsqfe}=0$. The test is whether $F_{sample}= [(R^2_{ur}-R^2_r)/(Q-K)]/[(1-R^2_{ur})/(N-K)]>F_{q-k,n-q}$.
In this case $F_{q-k,n-q}=F_{8-4,550-8}$=about 2.4. For these data F_{sample}={[(0.3417-0.2462)/(8-4)]/[(1-0.3417)/(500-8)]}=17.84 which is clearly larger than the $F_{crtical}$ at the 5% significance level. So we conclude that the structure of rewards for females in wages was different from that for males in 1978.

We can use the last regression to get more detailed insight regarding the differences in structure by looking at the individual coefficient for the added variables and their t values. Recall that the coefficients on each of the added variables is the differences between the effect of that characteristic on the wages of females and its effect on the wages of males. Thus the t test with a null hypothesis of zero difference is simply a test of significance for that coefficient. Looking at the t values for the individual coefficients we see that the coefficient for EXF is significant at better than the 5% level and the coefficient for EXSQF is significant at nearly the 5% level (it is signficiant at the 6% level).

The test for significant difference in structure of wage determination between females and males can be done in a slightly different fashion using what is called a “Chow” test, named after the Princetion econometrician Gregory Chow. First we run separate regessions of the base formulation for females and for males. To do this in STATA we add after the regression specification “if FE==1”. This will run the regrssion using only the observaton for females. Then we run it adding “if FE==0” this will run if using only male observations.

```
. regress wage ED EX EXSQ if FE==1

  Source |       SS       df       MS              Number of obs =     207
---------+------------------------------           F(  3,   203) =   12.22
   Model |  196.503775     3  65.5012582           Prob > F      =  0.0000
Residual |  1088.36597   203  5.36140874           R-square      =  0.1529
---------+------------------------------           Adj R-square  =  0.1404
   Total |  1284.86975   206  6.23723179           Root MSE      =  2.3155

------------------------------------------------------------------------------
    wage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ED |   .3743314   .0758179      4.937   0.000       .2248399     .523823
      EX |    .070659    .048124      1.468   0.144      -.0242281    .1655461
    EXSQ |  -.0001667     .00109     -0.153   0.879      -.0023158    .0019824
   _cons |  -1.093975   1.102622     -0.992   0.322      -3.268035    1.080085
------------------------------------------------------------------------------

. regress wage ED EX EXSQ if FE==0

  Source |       SS       df       MS              Number of obs =     343
---------+------------------------------           F(  3,   339) =   51.74
   Model |  1258.00508     3  419.335025           Prob > F      =  0.0000
```

```
Residual |  2747.71167   339  8.10534417              R-square      =  0.3141
---------+------------------------------              Adj R-square  =  0.3080
   Total |  4005.71675   342  11.7126221              Root MSE      =   2.847

------------------------------------------------------------------------------
    wage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ED |     .53327   .0558384      9.550   0.000       .4234366    .6431034
      EX |   .2492387   .0432794      5.759   0.000       .1641088    .3343687
    EXSQ |  -.0030643   .0009152     -3.348   0.001      -.0048645   -.0012642
   _cons |  -2.963195   .8196792     -3.615   0.000      -4.575493   -1.350897
------------------------------------------------------------------------------
```

Now the F_{sample} is formed as
$F_{sample}=[(SSE_R-\{SSE_{FE=1}+SSE_{FE=0}\})/(Q-K)]/[(SSE_{FE=1}+SSE_{FE=0})/(N-Q)]$

$=[(4392.56-\{1088.36+2747.71\})/8-4]/[(1088.36+2747.71)/(550-8)]= 19.65$
which is clearly greater than the $F_{crtical}$ at the 5% level.

Notice that the values for the coefficients for the females variables -FE, EDF, EXF,EXFSQ - are precisely equal to the differences in the similar variable between the female and male equations i.e. $FE=const_{fe}-const_{male}= -1.093975-(-2.963195)=1.86922$; $EDF=ED_{fe}-ED_{male}=.3743314-.53327=-.1589386$; etc. That is why above we could say we could test for which individual variables had coefficients significantly different for females vs. males by doing t tests on the interaction variables in the regression using EDF, EXF and EXSQF.