Soc 63993, Homework #6 Answer Key:
Interaction effects and group comparisons

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised February 20, 2015

Problem 1. Download *gender.dta* and/or *gender.sav* from the course web page. This is the hypothetical data on gender, income, education, and job experience that you used in homework 5. You will once again examine group differences in the parameters of this model, this time using dummy variables and interaction effects.

1. You are interested in the effects of education and job experience on income, and whether and if there are any differences in the models for men and women. Estimate the following three models using dummy variables and interaction effects (use Stata's factor variable notation to do so):

a. There are no differences by gender – the models are identical for men and women.

When we estimate the constrained model, we get

```
. reg  income educ jobexp

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,   497) =  239.86
       Model |  22352.7545     2  11176.3773           Prob > F      =  0.0000
    Residual |  23157.8824   497  46.5953368           R-squared     =  0.4912
-------------+------------------------------           Adj R-squared =  0.4891
       Total |  45510.6369   499  91.2036811           Root MSE      =  6.8261

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.309229   .0838474    15.61   0.000      1.14449    1.473968
      jobexp |   .8533107   .0670888    12.72   0.000     .7214982    .9851233
       _cons |  -1.076636   1.205717    -0.89   0.372    -3.445568    1.292295
------------------------------------------------------------------------------
. est store baseline
```

b. The intercepts differ by gender, but the effects of education and job experience are the same for both men and women.

If we regress income on education, job experience, and female, the model is

```
. reg  income educ jobexp i.female

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  3,   496) =  189.85
       Model |  24326.2478     3  8108.74928           Prob > F      =  0.0000
    Residual |   21184.389   496  42.7104618           R-squared     =  0.5345
-------------+------------------------------           Adj R-squared =  0.5317
       Total |  45510.6369   499  91.2036811           Root MSE      =  6.5353

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.281368   .0803805    15.94   0.000      1.12344    1.439296
      jobexp |   .7738483   .0652862    11.85   0.000     .6455767      .90212
    1.female |  -4.071767   .5990074    -6.80   0.000    -5.248671   -2.894862
       _cons |   2.511457   1.269321     1.98   0.048     .0175474    5.005367
------------------------------------------------------------------------------
. est store intonly
```

Note that the t-value for female is significant, suggesting intercepts differ by gender. But, just to be sure, we can also do Wald tests and incremental F tests and LR tests.

```
. testparm i.female

 ( 1)  1.female = 0

       F(  1,   496) =   46.21
            Prob > F =    0.0000

. ftest baseline intonly
Assumption: baseline nested in intonly

F(  1,     496) =     46.21
       prob > F =    0.0000

. lrtest baseline intonly

Likelihood-ratio test                                 LR chi2(1)  =     44.54
(Assumption: baseline nested in intonly)              Prob > chi2 =    0.0000
```

c. The intercepts and slopes differ by gender, i.e. all model parameters are free to differ by gender.

When we also add the the interaction terms, the unconstrained model is

```
. reg  income educ jobexp i.female i.female#c.educ i.female#c.jobexp

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  5,   494) =  179.36
       Model |    29345.68     5    5869.136           Prob > F      =  0.0000
    Residual |  16164.9569   494  32.7225848           R-squared     =  0.6448
-------------+------------------------------           Adj R-squared =  0.6412
       Total |  45510.6369   499  91.2036811           Root MSE      =  5.7204

---------------------------------------------------------------------------------
         income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           educ |   .8195378   .0904314     9.06   0.000     .6418602    .9972154
         jobexp |   1.384972   .0756042    18.32   0.000     1.236426    1.533517
       1.female |   6.399958   2.315577     2.76   0.006     1.850364    10.94955
                |
  female#c.educ |
             1  |   .7060444   .1522692     4.64   0.000     .4068693    1.005219
                |
female#c.jobexp |
             1  |  -1.389892   .1209307   -11.49   0.000    -1.627494    -1.15229
                |
          _cons |  -.9294128   1.264878    -0.73   0.463    -3.414617    1.555792
---------------------------------------------------------------------------------
```

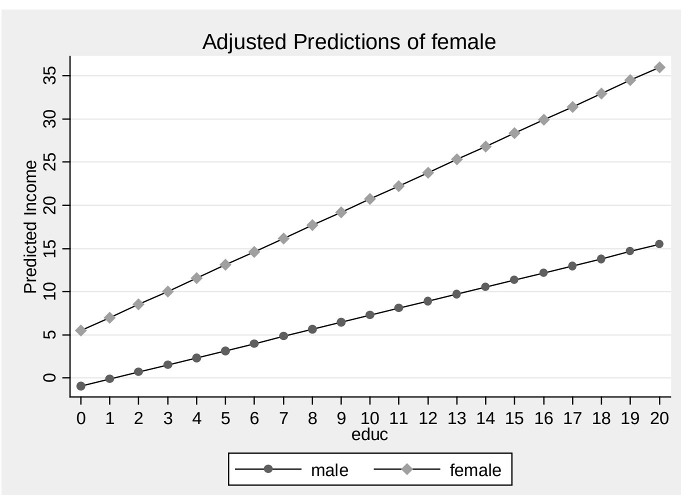
```
. est store slopesdiff
. testparm i.female#c.educ i.female#c.jobexp

 ( 1)  1.female#c.educ = 0
 ( 2)  1.female#c.jobexp = 0

       F(  2,   494) =   76.70
            Prob > F =    0.0000

. ftest intonly slopesdiff
Assumption: intonly nested in slopesdiff

F(  2,     494) =     76.70
       prob > F =    0.0000

. lrtest intonly slopesdiff

Likelihood-ratio test                                 LR chi2(2)  =    135.21
(Assumption: intonly nested in slopesdiff)            Prob > chi2 =    0.0000
```

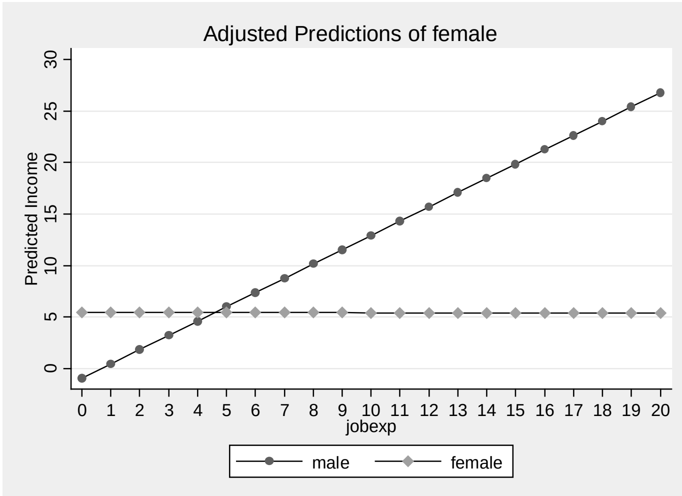
The incremental F is 76.7 with d.f. = 2, 494. This is highly significant. Differences between groups are not just limited to differences in the intercept.

2. Indicate which model you think is best, and why. Briefly discuss the substantive interpretation of what you think is the "best" model. Include in your discussion any insights that the model provides concerning gender differences. To help you with the discussion, run the following commands after your preferred model. Note that, in each case, the variable NOT being graphed is set to zero – which means that the (nonexistent in the data) point where income = 0 and jobexp = 0 is included in each graph.

```
quietly margins female, at(educ=(0(1)20) jobexp=0)
marginsplot, noci ytitle("Predicted Income") ylabel(#10) scheme(sj) name(educ)
quietly margins female, at(jobexp=(0(1)20) educ=0)
marginsplot, noci ytitle("Predicted Income") ylabel(#10) scheme(sj) name(jobexp)
```

The best model is the one that includes all the interaction effects. The graphs look like this:





According to this model, Education has almost twice as large an effect on women as it does men (because the interaction effect FEM*EDUC is almost as large as the main effect of EDUC). On the other hand, job experience has virtually no effect on women (because the B for FEMJOB is almost exactly the opposite of JOBEXP), yet for men job experience actually has a larger effect than does education. Hence, the determinants of income are very different for men than women. Further, if a choice must be made between more education and more job experience, women gain far more from education while men gain somewhat more from job experience. Again, these would be fascinating findings, if only they weren't completely hypothetical.

3. In the models above, the effect of Female changes from negative to positive once interaction terms are added to the model. Explain why this should not concern you. In particular, explain how the interpretation of the coefficient for Female changes once interaction terms are added to the model.

Once interaction effects were added, the effect of female went from being significantly negative to significantly positive. At first, this may seem odd, but it isn't once you understand how to interpret the effects. In the first model, with no interactions, the coefficient for female tells you the expected difference between a man and woman who are otherwise comparable, i.e. have identical values for JOBEXP and EDUC. This includes the special case when JOBEXP and EDUC both equal zero, but is not limited to it. In the second model with interactions, <u>the coefficient for female has a narrower meaning: it is the expected difference between a man and woman who both have 0 years of education and 0 years of job experience</u>. As the following descriptives show, nobody actually has such small values, and zero is far from a typical value for these variables:

```
. sum  educ jobexp income

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        educ |       500        10.9    3.690154          2         17
      jobexp |       500       13.15    4.611945          3         23
      income |       500      24.415    9.550062          5       48.3
```

Hence, we shouldn't pay too much attention to the coefficient for female once interaction effects are added.

4. Center the continuous variables and rerun the three models. How do your results differ from before? Explain how centering makes it easier to interpret the results.

If we want to make the results a little easier to interpret we can center education and jobexp first. In Stata, one approach is

```
. * Center the variables. There is no missing data; if there were you would have
. * to exclude it first
. sum educ, meanonly
. gen educx = educ - r(mean)
. sum jobexp, meanonly
. gen jobexpx = jobexp - r(mean)
. * Redo regressions with centered variables
. reg  income educx jobexpx

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,   497) =  239.86
       Model |  22352.7548     2  11176.3774           Prob > F      =  0.0000
    Residual |   23157.882   497  46.5953361           R-squared     =  0.4912
-------------+------------------------------           Adj R-squared =  0.4891
       Total |  45510.6369   499  91.2036811           Root MSE      =  6.8261

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       educx |   1.309229   .0838474    15.61   0.000      1.14449    1.473968
     jobexpx |   .8533108   .0670888    12.72   0.000     .7214982    .9851233
       _cons |     24.415   .3052715    79.98   0.000     23.81522    25.01478
------------------------------------------------------------------------------

. reg  income educx jobexpx i.female

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  3,   496) =  189.85
       Model |   24326.248     3  8108.74933           Prob > F      =  0.0000
    Residual |  21184.3889   496  42.7104615           R-squared     =  0.5345
-------------+------------------------------           Adj R-squared =  0.5317
       Total |  45510.6369   499  91.2036811           Root MSE      =  6.5353

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       educx |   1.281368   .0803805    15.94   0.000      1.12344    1.439296
     jobexpx |   .7738484   .0652862    11.85   0.000     .6455767    .9021201
    1.female |  -4.071766   .5990074    -6.80   0.000    -5.248671   -2.894862
       _cons |   26.65447   .4404099    60.52   0.000     25.78917    27.51977
------------------------------------------------------------------------------
```

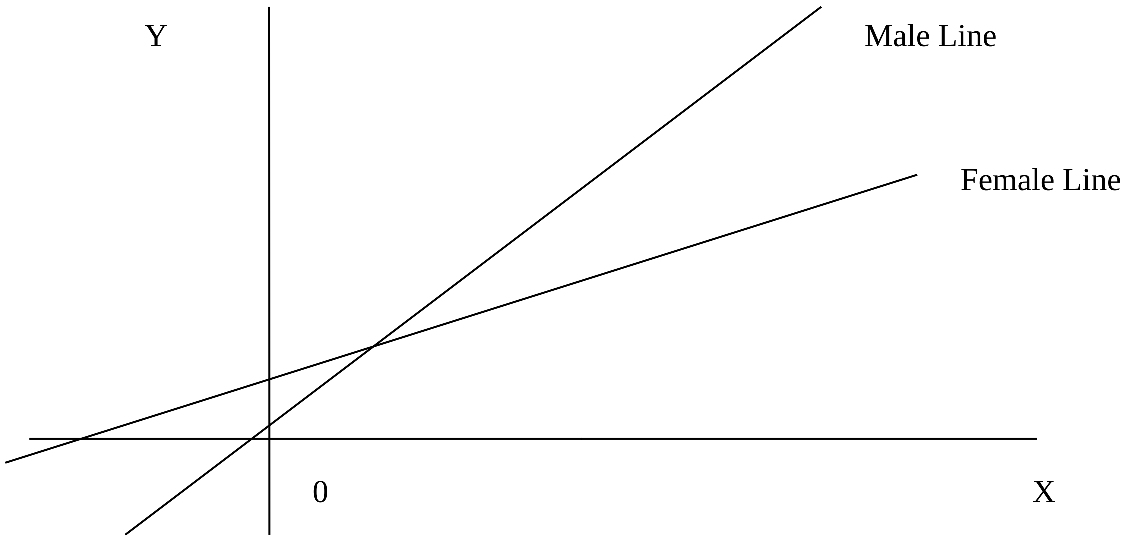
```
. reg  income educx jobexpx i.female i.female#c.educx i.female#c.jobexpx

      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  5,   494) =  179.36
       Model |  29345.6803     5  5869.13606           Prob > F      =  0.0000
    Residual |  16164.9566   494  32.7225842           R-squared     =  0.6448
-------------+------------------------------           Adj R-squared =  0.6412
       Total |  45510.6369   499  91.2036811           Root MSE      =  5.7204

------------------------------------------------------------------------------
          income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
           educx |   .8195378   .0904314     9.06   0.000     .6418602    .9972154
         jobexpx |   1.384972   .0756042    18.32   0.000     1.236426    1.533517
        1.female |  -4.181232   .5259562    -7.95   0.000    -5.214619   -3.147845
                 |
  female#c.educx |
               1 |   .7060443   .1522692     4.64   0.000     .4068693    1.005219
                 |
female#c.jobexpx |
               1 |  -1.389892   .1209307   -11.49   0.000    -1.627494    -1.15229
                 |
           _cons |   26.21593   .3875142    67.65   0.000     25.45455    26.97731
------------------------------------------------------------------------------
```

As we see, the effect of female changes hardly at all between models once variables are centered. Model 3 shows us that, when a man and woman both have average levels of education and job experience (10.9 years of education and 13.15 years of job experience) the woman is predicted to make $4,181 less on average than the man does. However, you can also compute from the above coefficients that if a man and woman both had 0 years of education and job experience, the woman would be predicted to have a $6,400 edge, i.e. regardless of whether you center or not the predictions are the same.

The intercept term also becomes more interpretable. Once we have centered, the intercept tells us the predicted income for a man with average levels of education and jobexp, whereas before centering it gives us the predicted income for a man with 0 years of education and 0 years of job experience. Note that the intercept is slightly lower than the male mean of 27.81 on income. This is because men tend to have above-average levels of education and job experience, i.e. they have higher mean levels of education and job experience than women do. (In other words, the average man is above average.)

```
. tabstat income educ jobexp, by(female) columns(variables)

Summary statistics: mean
  by categories of: female

female |    income      educ    jobexp
-------+------------------------------
  male |  27.81111  11.22222  14.11111
female |  21.63636  10.63636  12.36364
-------+------------------------------
 Total |    24.415      10.9     13.15
--------------------------------------
```

The key thing to realize is, if the male & female lines are not parallel, at some point females have to have a predicted edge over males – although that point may never actually occur within the observed or even any possible data. The following diagram illustrates this in the case where you have one X variable rather than 2:



In the present example, women happen to have a predicted edge over men when job experience and education both equal 0. They'd have an even bigger edge if you extended the lines to include negative values of job experience and education. But, since you don't observe such negative and zero values in reality, the predicted lead for women at these values doesn't mean much.

Problem 2. Get *jgqes2.dta* and *jgqes2.do* from the course web page. Selected variables from The Quality of Employment Survey are contained in jgqes2.dta. Run jgqes2.do and answer the following questions:

1. What is the mean of each group on the dependent variable (jsat = Job Satisfaction)? Is the mean difference between groups statistically significant?

When we regress jsat (job satisfaction) on white, we get

```
. * Regressions, set 1. Mean job satisfaction difference between groups.
. reg jsat white

      Source |       SS       df       MS              Number of obs =    1116
-------------+------------------------------           F(  1,  1114) =   12.97
       Model |  264.913505     1  264.913505           Prob > F      =  0.0003
    Residual |  22749.1511  1114   20.421141           R-squared     =  0.0115
-------------+------------------------------           Adj R-squared =  0.0106
       Total |  23014.0646  1115  20.6404167           Root MSE      =   4.519

------------------------------------------------------------------------------
        jsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   1.489367   .4135134     3.60   0.000     .6780138     2.30072
       _cons |   17.98074    .387499    46.40   0.000     17.22043    18.74105
------------------------------------------------------------------------------
```

This means that non-whites have an average score of 17.98 on the JSAT scale, while whites score an average of 1.49 points higher (i.e. 19.47). The T value shows that this difference is statistically significant.

2. Are there any statistically significant differences in the model parameters between groups?

We'll contrast the model in which there are no differences across groups with the model where all parameters are free to vary.

```
. * Regressions, set 2.  Test for any differences between groups.
. nestreg: reg jsat (goodjob tenure firmsize hrswk) (white goodjobwh tenurewh firmszwh
hrswkwh)

Block  1: goodjob tenure firmsize hrswk

      Source |       SS       df       MS              Number of obs =    1116
-------------+------------------------------           F(  4,  1111) =   12.95
       Model |  1024.96398     4  256.240994           Prob > F      =  0.0000
    Residual |  21989.1006  1111  19.7921698           R-squared     =  0.0445
-------------+------------------------------           Adj R-squared =  0.0411
       Total |  23014.0646  1115  20.6404167           Root MSE      =  4.4488

------------------------------------------------------------------------------
        jsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     goodjob |   1.034038   .2971628     3.48   0.001     .4509745    1.617102
      tenure |   .1036898   .0196212     5.28   0.000     .0651909    .1421887
    firmsize |  -.2064776   .0728452    -2.83   0.005    -.3494073   -.0635479
       hrswk |  -.0294379   .0130543    -2.26   0.024    -.0550518   -.0038239
       _cons |   20.14754   .6331961    31.82   0.000     18.90514    21.38993
------------------------------------------------------------------------------

Block  2: white goodjobwh tenurewh firmszwh hrswkwh

      Source |       SS       df       MS              Number of obs =    1116
-------------+------------------------------           F(  9,  1106) =    7.57
       Model |  1335.39722     9  148.377468           Prob > F      =  0.0000
    Residual |  21678.6674  1106  19.6009651           R-squared     =  0.0580
-------------+------------------------------           Adj R-squared =  0.0504
       Total |  23014.0646  1115  20.6404167           Root MSE      =  4.4273

------------------------------------------------------------------------------
        jsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     goodjob |   1.348528   1.109807     1.22   0.225    -.8290381    3.526093
      tenure |    .150903   .0527798     2.86   0.004     .0473432    .2544629
    firmsize |   .1039604   .2153092     0.48   0.629    -.3185003    .5264211
       hrswk |  -.0754383   .0505458    -1.49   0.136    -.1746147    .0237382
       white |   1.127171   2.403992     0.47   0.639    -3.589728    5.844071
   goodjobwh |   -.448483   1.152123    -0.39   0.697    -2.709077    1.812111
    tenurewh |  -.0533176   .0568131    -0.94   0.348    -.1647912     .058156
    firmszwh |  -.3435501   .2287153    -1.50   0.133    -.7923151    .1052148
     hrswkwh |   .0470711   .0523105     0.90   0.368    -.0555679      .14971
       _cons |   19.33807    2.31209     8.36   0.000     14.80149    23.87464
------------------------------------------------------------------------------
```

```
  +-------------------------------------------------------------+
  |       |          Block  Residual                     Change |
  | Block |       F     df        df   Pr > F       R2    in R2 |
  |-------+-----------------------------------------------------|
  |     1 |   12.95      4      1111   0.0000   0.0445          |
  |     2 |    3.17      5      1106   0.0076   0.0580   0.0135 |
  +-------------------------------------------------------------+
```

The incremental F is 3.17 (see the F change statistic in the printout), with d.f. = 5, 1106. This is highly significant, so we conclude that one or more parameters likely differ across groups. We could have also done

```
. quietly reg jsat goodjob tenure firmsize hrswk white goodjobwh tenurewh firmszwh hrswkwh

. test  white tenurewh firmszwh goodjobwh hrswkwh

 ( 1)  white = 0
 ( 2)  tenurewh = 0
 ( 3)  firmszwh = 0
 ( 4)  goodjobwh = 0
 ( 5)  hrswkwh = 0

       F(  5,  1106) =    3.17
            Prob > F =    0.0076

. reg jsat i.goodjob tenure firmsize hrswk i.white i.white#(i.goodjob c.tenure
c.firmsize c.hrswk)

      Source |       SS       df       MS              Number of obs =    1116
-------------+------------------------------           F(  9,  1106) =    7.57
       Model |  1335.39722     9  148.377468           Prob > F      =  0.0000
    Residual |  21678.6674  1106  19.6009651           R-squared     =  0.0580
-------------+------------------------------           Adj R-squared =  0.0504
       Total |  23014.0646  1115  20.6404167           Root MSE      =  4.4273

------------------------------------------------------------------------------
        jsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.goodjob |   1.348528   1.109807     1.22   0.225    -.8290381    3.526093
      tenure |    .150903   .0527798     2.86   0.004     .0473432    .2544629
    firmsize |   .1039604   .2153092     0.48   0.629    -.3185003    .5264211
       hrswk |  -.0754383   .0505458    -1.49   0.136    -.1746147    .0237382
     1.white |   1.127171   2.403992     0.47   0.639    -3.589728    5.844071
             |
      white#|
     goodjob |
        1 1  |   -.448483   1.152123    -0.39   0.697    -2.709077    1.812111
             |
      white#|
    c.tenure |
          1  |  -.0533176   .0568131    -0.94   0.348    -.1647912     .058156
             |
      white#|
  c.firmsize |
          1  |  -.3435501   .2287153    -1.50   0.133    -.7923151    .1052148
             |
      white#|
     c.hrswk |
          1  |   .0470711   .0523105     0.90   0.368    -.0555679      .14971
             |
       _cons |   19.33807    2.31209     8.36   0.000     14.80149    23.87464
------------------------------------------------------------------------------
```

```
. testparm i.white i.white#(i.goodjob c.tenure c.firmsize c.hrswk)

 ( 1)  1.white = 0
 ( 2)  1.white#1.goodjob = 0
 ( 3)  1.white#c.tenure = 0
 ( 4)  1.white#c.firmsize = 0
 ( 5)  1.white#c.hrswk = 0

       F(  5,  1106) =    3.17
            Prob > F =    0.0076
```

3. If the answer to 2 is yes, are these differences limited to differences in the intercepts? Or are there differences in the effects of the IVs across groups (i.e. are there statistically significant interaction effects? Or is it just the coefficient of the dummy variable for group membership that is statistically significant?)

Even though the incremental F is significant, none of the T values for WHITE or the interaction terms are. It is unlikely that all of the interaction terms belong in the model, and it may be that none of them do. We therefore estimate a more extensive set of models, including one in which only the main effects of the variables (including white) are in the model, and contrast that with the model that also includes interaction terms:

```
. * Regressions, set 3. More detailed tests for differences in effects.
. nestreg: reg jsat (goodjob tenure firmsize hrswk) (white) (goodjobwh tenurewh
firmszwh hrswkwh)

Block  1: goodjob tenure firmsize hrswk

      Source |       SS       df       MS              Number of obs =    1116
-------------+------------------------------           F(  4,  1111) =   12.95
       Model |  1024.96398     4  256.240994           Prob > F      =  0.0000
    Residual |  21989.1006  1111  19.7921698           R-squared     =  0.0445
-------------+------------------------------           Adj R-squared =  0.0411
       Total |  23014.0646  1115  20.6404167           Root MSE      =  4.4488

------------------------------------------------------------------------------
        jsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     goodjob |   1.034038   .2971628     3.48   0.001     .4509745    1.617102
      tenure |   .1036898   .0196212     5.28   0.000     .0651909    .1421887
    firmsize |  -.2064776   .0728452    -2.83   0.005    -.3494073   -.0635479
       hrswk |  -.0294379   .0130543    -2.26   0.024    -.0550518   -.0038239
       _cons |   20.14754   .6331961    31.82   0.000     18.90514    21.38993
------------------------------------------------------------------------------
```

```
Block  2: white

      Source |       SS       df       MS              Number of obs =    1116
-------------+------------------------------           F(  5,  1110) =   12.67
       Model |  1242.38355     5  248.476711           Prob > F      =  0.0000
    Residual |  21771.6811  1110  19.6141271           R-squared     =  0.0540
-------------+------------------------------           Adj R-squared =  0.0497
       Total |  23014.0646  1115  20.6404167           Root MSE      =  4.4288

------------------------------------------------------------------------------
        jsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     goodjob |   .9175087   .2978865     3.08   0.002     .3330246    1.501993
      tenure |   .1041977   .0195334     5.33   0.000     .0658712    .1425243
    firmsize |  -.1999761   .0725431    -2.76   0.006    -.3423132    -.057639
       hrswk |  -.0318354   .0130154    -2.45   0.015     -.057373   -.0062978
       white |   1.363098   .4094135     3.33   0.001      .559786    2.166409
       _cons |   19.06165   .7097216    26.86   0.000     17.66911     20.4542
------------------------------------------------------------------------------

Block  3: goodjobwh tenurewh firmszwh hrswkwh

      Source |       SS       df       MS              Number of obs =    1116
-------------+------------------------------           F(  9,  1106) =    7.57
       Model |  1335.39722     9  148.377468           Prob > F      =  0.0000
    Residual |  21678.6674  1106  19.6009651           R-squared     =  0.0580
-------------+------------------------------           Adj R-squared =  0.0504
       Total |  23014.0646  1115  20.6404167           Root MSE      =  4.4273

------------------------------------------------------------------------------
        jsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     goodjob |   1.348528   1.109807     1.22   0.225    -.8290381    3.526093
      tenure |    .150903   .0527798     2.86   0.004     .0473432    .2544629
    firmsize |   .1039604   .2153092     0.48   0.629    -.3185003    .5264211
       hrswk |  -.0754383   .0505458    -1.49   0.136    -.1746147    .0237382
       white |   1.127171   2.403992     0.47   0.639    -3.589728    5.844071
   goodjobwh |   -.448483   1.152123    -0.39   0.697    -2.709077    1.812111
    tenurewh |  -.0533176   .0568131    -0.94   0.348    -.1647912     .058156
    firmszwh |  -.3435501   .2287153    -1.50   0.133    -.7923151    .1052148
     hrswkwh |   .0470711   .0523105     0.90   0.368    -.0555679      .14971
       _cons |   19.33807    2.31209     8.36   0.000     14.80149    23.87464
------------------------------------------------------------------------------


  +-------------------------------------------------------------+
  |       |          Block  Residual                     Change |
  | Block |       F     df        df   Pr > F       R2    in R2 |
  |-------+-----------------------------------------------------|
  |     1 |   12.95      4      1111   0.0000   0.0445          |
  |     2 |   11.08      1      1110   0.0009   0.0540   0.0094 |
  |     3 |    1.19      4      1106   0.3151   0.0580   0.0040 |
  +-------------------------------------------------------------+
```

Note that in the 2nd model the T value for white is statistically significant (as is the incremental F test for the model). When whites and nonwhites have identical values on other variables, whites still tend to score about 1.36 points higher on the job satisfaction scale, i.e. the intercepts are different across races.

When the interaction effects are added in the unconstrained model, the incremental F is only 1.19 with d.f. = 4, 1106. This is not significant.

4. Briefly discuss the substantive interpretation of what you think is the "best" model for the data set. Include in your discussion any insights that the model provides concerning group differences.

The model with main effects only (including white) is best. Differences between races are limited to differences in the intercepts. Perhaps whites are more satisfied with things in general. Or, perhaps whites tend to receive better treatment on the job simply because they are white, leading to a higher level of satisfaction. All other variables have the same effect on whites that they do on non-whites.

5. Examine the compositional differences (i.e. mean differences) between groups on the independent variables. Discuss how these differences help lead to mean differences on the dependent variable.

```
. * t-tests for compositional differences
. ttest goodjob, by(white)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
NonWhite |     136    .1397059    .0298376    .3479633    .0806963    .1987155
   White |     980    .3173469    .0148756    .4656814    .2881551    .3465388
---------+--------------------------------------------------------------------
combined |    1116    .2956989    .0136668    .4565609    .2688834    .3225145
---------+--------------------------------------------------------------------
    diff |           -.1776411    .0414566               -.2589828   -.0962993
------------------------------------------------------------------------------
    diff = mean(NonWhite) - mean(White)                           t =  -4.2850
Ho: diff = 0                                     degrees of freedom =     1114

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000

. ttest tenure, by(white)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
NonWhite |     136    7.676471    .6264749     7.30589    6.437496    8.915445
   White |     980    7.680612    .2201163     6.89073    7.248658    8.112566
---------+--------------------------------------------------------------------
combined |    1116    7.680108     .207721    6.939249    7.272539    8.087676
---------+--------------------------------------------------------------------
    diff |           -.0041417    .6352679               -1.250598    1.242315
------------------------------------------------------------------------------
    diff = mean(NonWhite) - mean(White)                           t =  -0.0065
Ho: diff = 0                                     degrees of freedom =     1114

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.4974         Pr(|T| > |t|) = 0.9948          Pr(T > t) = 0.5026
```

```
. ttest firmsize, by(white)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
NonWhite |     136    3.558824    .1554264    1.812568    3.251438    3.866209
   White |     980    3.372449    .0595741    1.864965    3.255541    3.489357
---------+--------------------------------------------------------------------
combined |    1116    3.395161    .0556436    1.858862    3.285983    3.504339
---------+--------------------------------------------------------------------
    diff |            .1863745    .1700817                -.147342    .5200911
------------------------------------------------------------------------------
    diff = mean(NonWhite) - mean(White)                           t =   1.0958
Ho: diff = 0                                     degrees of freedom =     1114

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.8633         Pr(|T| > |t|) = 0.2734          Pr(T > t) = 0.1367

. ttest hrswk, by(white)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
NonWhite |     136       40.75    .6536724    7.623064    39.45724    42.04276
   White |     980    43.08765    .3392098    10.61895    42.42199    43.75331
---------+--------------------------------------------------------------------
combined |    1116    42.80278     .309105    10.32614    42.19628    43.40927
---------+--------------------------------------------------------------------
    diff |           -2.337653    .9427301                -4.18738   -.4879263
------------------------------------------------------------------------------
    diff = mean(NonWhite) - mean(White)                           t =  -2.4797
Ho: diff = 0                                     degrees of freedom =     1114

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0066         Pr(|T| > |t|) = 0.0133          Pr(T > t) = 0.9934

. ttest jsat, by(white)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
NonWhite |     136    17.98074     .438593     5.11483    17.11333    18.84814
   White |     980     19.4701     .141528    4.430527    19.19237    19.74784
---------+--------------------------------------------------------------------
combined |    1116     19.2886    .1359963    4.543173    19.02176    19.55544
---------+--------------------------------------------------------------------
    diff |           -1.489367    .4135134                -2.30072   -.6780138
------------------------------------------------------------------------------
    diff = mean(NonWhite) - mean(White)                           t =  -3.6017
Ho: diff = 0                                     degrees of freedom =     1114

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0002         Pr(|T| > |t|) = 0.0003          Pr(T > t) = 0.9998
```

We can note several things from the above:

- More than twice as many whites (about 17.8% more) are in good jobs as are nonwhites. This difference is highly significant.

- However, there are only trivial, and non-significant, differences in job tenure, i.e. whites and nonwhites have been in jobs about equally long.

- Whites tend to work in slightly smaller firms, but the difference is not statistically significant.

- Whites work a statistically significant 2.3 more hours a week.

Hence, compositional (mean) differences in Tenure and Firm Size have virtually no effect on racial differences in Job satisfaction. The longer hours that whites work does tend to reduce their job satisfaction relative to non-whites (because hours worked has a negative effect on Job Satisfaction; an average of 2.3 more hours worked times an effect of -.031835 for hours worked results in a net mean white disadvantage of about .07 on the JSAT scale). However, the much higher proportion of whites in good jobs gives Whites an advantage over non-whites. (An additional 17.8% of whites are in good jobs, the effect of good job is .9175, producing a net white advantage of about .16 on JSAT).

As we saw, overall whites score 1.49 points higher on the JSAT scale. A small part of this advantage is due to the greater likelihood of whites being in good jobs. Most of the difference, however, seems to stem from differences in the intercepts. Even when a white and nonwhite have identical values on all other variables, the white tends to score 1.36 points higher. This may reflect a general attitudinal difference between the races. However, it may also reflect the effects of differential treatment or of other variables that are not considered here.

Following is a copy of jgqes2.do:

```
version 12.1
* Problem 2. Quality of Employment survey.
use https://www3.nd.edu/~rwilliam/statafiles/jqges2.dta, clear
* Tidy up the data for our purposes
keep jsat prof mang tenure firmsize hrswk race
* Compute "Good job" variable (professional or managerial).
gen goodjob=prof+mang
* Compute dummy variable for white/ nonwhite.
recode race (1=1) (else=0), gen(white)
* hrswk (hours work per week) seems to be off by factor of 10,
* so correct.
replace hrswk = hrswk/10.
label define gdjob 0 "Other" 1 "Prof, Manager"
label values goodjob goodjob
label define white 0 "NonWhite" 1 "White"
label values white white
* Limit to cases with complete data
keep if !missing(jsat, goodjob, tenure, firmsize, hrswk, white)

* Compute race interaction terms.
gen tenurewh=tenure*white
gen firmszwh=firmsize*white
gen goodjobwh=goodjob*white
gen hrswkwh=hrswk*white

* Regressions, set 1. Mean job satisfaction difference between groups.
reg jsat white
* Regressions, set 2. Test for any differences between groups.
nestreg: reg jsat (goodjob tenure firmsize hrswk) (white goodjobwh tenurewh firmszwh hrswkwh)
```

```
* Regressions, set 3. More detailed tests for differences in effects.
nestreg: reg jsat (goodjob tenure firmsize hrswk) (white) (goodjobwh tenurewh firmszwh hrswkwh)
* t-tests for compositional differences
ttest goodjob, by(white)
ttest tenure, by(white)
ttest firmsize, by(white)
ttest hrswk, by(white)
ttest jsat, by(white)
```