BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Ms. Yu, Fall 2003

Problem Set 5 Solutions

Problem sets should be your own work. You may work together with classmates, but if you're not figuring this out on your own, you will eventually regret it.

1. (**6.4**)

(i) Holding all other factors fixed we have

$$\Delta \log(wage) = \beta_1 \Delta educ + \beta_2 \Delta educ \cdot pareduc = (\beta_1 + \beta_2 pareduc)\Delta educ$$

Dividing both sides by $\Delta educ$ gives the result. The sign of β_2 is not obvious, although $\beta_2 > 0$ if we think a child gets more out of another year of education the more highly educated are the child's parents.

(ii) We use the values $pareduc = 32$ and $pareduc = 24$ to interpret the coefficent on $educ \cdot pareduc$. The difference in the estimated return to education is $.00078(32 - 24) = .0062$, or about .62 percentage points.

(iii) When we add *pareduc* by itself, the coefficient on the interaction term is negative. The t-statistic on $educ \cdot pareduc$ is about -1.33, which is not significant at the 10% level against a two-sided alternative. Note that the coefficient on *pareduc* is significant at the 5% level against a two-sided alternative. this provides a good example of how omitting a level effect (*pareduc* in this case) can lead to biased estimation of the interaction effect.

2. (**6.9**)

(i)
```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/WAGE1

. regress lwage educ exper expersq

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  3,   522) =   74.67
       Model |  44.5393702     3  14.8464567           Prob > F      =  0.0000
    Residual |  103.790392   522  .198832168           R-squared     =  0.3003
-------------+------------------------------           Adj R-squared =  0.2963
```

```
       Total |  148.329762   525   .28253288           Root MSE      =  .44591

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0903658    .007468    12.10   0.000     .0756948    .1050368
       exper |   .0410089   .0051965     7.89   0.000     .0308002    .0512175
     expersq |  -.0007136   .0001158    -6.16   0.000     -.000941   -.0004861
       _cons |   .1279975   .1059323     1.21   0.227    -.0801085    .3361034
------------------------------------------------------------------------------
```

The estimated equation is

$$\begin{aligned}\log(\widehat{wage}) &= \underset{(.106)}{.128} + \underset{(.0075)}{.0904}\, educ + \underset{(.0052)}{.0410}\, exper - \underset{(.000116)}{.000714}\, exper^2 \\ n &= 526, R^2 = .300, \overline{R}^2 = .296.\end{aligned}$$

(ii) The t-statistic on $exper^2$ is about -6.16 which has a p-value of essentially zero. So $exper$ is significant at the 1% level (and much smaller significance levels).

(iii) To estimate the return to the fifth year of experience, we start at $exper = 4$ and increase $exper$ by one, so $\Delta exper = 1$:

$$\%\Delta\widehat{wage} \approx 100[.0410 - 2(.000714)4] \approx 3.53\%$$

Similarly, for the 20^{th} year of experience,

$$\%\Delta\widehat{wage} \approx 100[.0410 - 2(.000714)19] \approx 1.39\%$$

(iv) The turnaround point is about $.041/[2(.000714)] \approx 28.7$ years of experience. In the sample, there are 121 people with at least 29 years of experience. This is a fairly sizeable fraction of the sample.

3. (**6.10**)

(i) Holding $exper$ (and the elements in u) fixed, we have

$$\Delta\log(wage) = \beta_1\Delta educ + \beta_3(\Delta educ)exper = (\beta_1 + \beta_3 exper)\Delta educ,$$

or

$$\frac{\Delta \log(wage)}{\Delta educ} = (\beta_1 + \beta_3 exper).$$

This is the approximate proportionate change in *wage* given one more year of education.

(ii) $H_0 : \beta_3 = 0$. If we think that education and experience interact positively – so tat people with more experience are more productive when given another year of education – then $\beta_3 > 0$ is the appropriate alternative.

(iii)
```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/WAGE2

. gen eduexper= educ* exper

. regress lwage educ exper eduexper

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  3,   931) =   48.41
       Model |  22.3529774     3  7.45099246           Prob > F      =  0.0000
    Residual |  143.303317   931  .153924078           R-squared     =  0.1349
-------------+------------------------------           Adj R-squared =  0.1321
       Total |  165.656294   934  .177362199           Root MSE      =  .39233

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0440498   .0173911     2.53   0.011     .0099195    .0781801
       exper |  -.0214959   .0199783    -1.08   0.282    -.0607036    .0177118
    eduexper |    .003203   .0015292     2.09   0.036      .000202     .006204
       _cons |   5.949455   .2408264    24.70   0.000     5.476829     6.42208
------------------------------------------------------------------------------
```

The estimated equation is

$$\begin{aligned}\widehat{\log(wage)} &= \underset{(0.24)}{5.95} + \underset{(.0174)}{.0440}\, educ + \underset{(.0200)}{.0215}\, exper - \underset{(.00153)}{.00320}\, educ \cdot exper \\ n &= 935, R^2 = .135, \overline{R}^2 = .132.\end{aligned}$$

The t-statistic on the interaction term is about 2.09, which gives a p-value below .036 against $H_1 : \beta_3 > 0$. Therefore, we reject $H_0 : \beta_3 = 0$ against $H_1 : \beta_3 > 0$ at the 3.6% level.

(iv) We rewrite the equation as

$$\log(wage) = \beta_0 + \theta_1 educ + \beta_2 exper + \beta_3 educ(exper - 10) + u,$$

and run the regression $\log(wage)$ on *educ*, *exper*, and $educ(exper-10)$. We want the coefficient on *educ*.

```
. gen exper_10=exper-10

. gen eduexper_10= educ* exper_10

. regress lwage educ exper eduexper_10

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  3,   931) =   48.41
       Model |  22.3529774     3  7.45099246           Prob > F      =  0.0000
    Residual |  143.303317   931  .153924078           R-squared     =  0.1349
-------------+------------------------------           Adj R-squared =  0.1321
       Total |  165.656294   934  .177362199           Root MSE      =  .39233

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0760795   .0066151    11.50   0.000     .0630974    .0890617
       exper |  -.0214959   .0199783    -1.08   0.282    -.0607036    .0177118
 eduexper_10 |    .003203   .0015292     2.09   0.036      .000202     .006204
       _cons |   5.949455   .2408264    24.70   0.000     5.476829     6.42208
------------------------------------------------------------------------------
```

or using the lincom command after the orignial regression

```
. regress lwage educ exper eduexper

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  3,   931) =   48.41
       Model |  22.3529774     3  7.45099246           Prob > F      =  0.0000
    Residual |  143.303317   931  .153924078           R-squared     =  0.1349
-------------+------------------------------           Adj R-squared =  0.1321
       Total |  165.656294   934  .177362199           Root MSE      =  .39233

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0440498   .0173911     2.53   0.011     .0099195    .0781801
       exper |  -.0214959   .0199783    -1.08   0.282    -.0607036    .0177118
```

```
    eduexper |    .003203   .0015292     2.09   0.036      .000202     .006204
       _cons |   5.949455   .2408264    24.70   0.000     5.476829     6.42208
------------------------------------------------------------------------------

. lincom educ+10* eduexper

 ( 1)  educ + 10.0 eduexper = 0.0

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .0760795   .0066151    11.50   0.000     .0630974    .0890617
------------------------------------------------------------------------------
```

We obtain $\hat{\theta}_1 \approx .0761$ and $se(\hat{\theta}_1) \approx .0066$. The 95% CI for θ_1 is about .063 to .089.

4. (**6.16**)

(i)

```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/NBASAL

. regress points exper expersq age educ

      Source |       SS       df       MS              Number of obs =     269
-------------+------------------------------           F(  4,   264) =   10.85
       Model |  1317.59877     4  329.399693           Prob > F      =  0.0000
    Residual |  8013.59211   264  30.3545156           R-squared     =  0.1412
-------------+------------------------------           Adj R-squared =  0.1282
       Total |  9331.19088   268  34.8178764           Root MSE      =  5.5095

------------------------------------------------------------------------------
      points |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |   2.363631   .4054974     5.83   0.000      1.56521    3.162051
     expersq |  -.0770269   .0234833    -3.28   0.001    -.1232652   -.0307885
         age |  -1.073958   .2950722    -3.64   0.000    -1.654953   -.4929638
        educ |  -1.286255   .4505921    -2.85   0.005    -2.173466    -.399043
       _cons |   35.21831   6.986731     5.04   0.000      21.4615    48.97512
------------------------------------------------------------------------------
```

The estimated equation is

$$\widehat{points} = \underset{(6.99)}{35.22} + \underset{(.405)}{2.364}\, exper - \underset{(.0235)}{.0770}\, exper^2$$

$$- 1.074\, age - 1.286\, edu$$
$$(.295) \qquad (.451)$$
$$n = 269, R^2 = .141, \overline{R}^2 = .128.$$

(ii) The turnaround point is $2.364/[2(.0770)] \approx 15.35$. So, the increase from 15 to 16 years of experience would actually reduce points. This is a very high level of experience, and we can essentially ignore this prediction: only two players in the sample of 269 have more than 15 years of experience.

(iii) Many of the most promising players leave college early, or, in some cases, forego college altogether, to play in the NBA. These top players command the highest salaries. it is not more college than hurts salary, but less college is indicative of super-star potential.

(iv)
```
. regress points exper expersq age agesq educ

      Source |       SS       df       MS              Number of obs =     269
-------------+------------------------------           F(  5,   263) =    8.92
       Model |  1353.54692     5  270.709385           Prob > F      =  0.0000
    Residual |  7977.64396   263   30.333247           R-squared     =  0.1451
-------------+------------------------------           Adj R-squared =  0.1288
       Total |  9331.19088   268  34.8178764           Root MSE      =  5.5076

------------------------------------------------------------------------------
      points |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |   2.863828   .6127241     4.67   0.000     1.657359    4.070297
     expersq |  -.1280723   .0524378    -2.44   0.015    -.2313237   -.0248209
         age |  -3.983695   2.689078    -1.48   0.140    -9.278557    1.311168
       agesq |   .0535514   .0491917     1.09   0.277    -.0433083    .1504112
        educ |  -1.312604   .4510841    -2.91   0.004    -2.200799    -.424408
       _cons |   73.59034   35.93341     2.05   0.042     2.836555    144.3441
------------------------------------------------------------------------------
```

When age^2 is added to the regression from part (i), its coefficient is .0536 (se=.0492). Its t statistic is barely above one, so we are justified in dropping it. The coefficient on age in the same regression is -3.984 (se = 2.689). Together, these estimates imply a negative, increasing, return to age. The turning point is roughly at 74 years old. In any case, the linear function of age seems sufficient.

(v)
```
.regress  lwage  points exper expersq age educ
```

```
      Source |       SS       df       MS              Number of obs =     269
-------------+------------------------------           F(  5,   263) =   50.10
       Model |  101.561351     5  20.3122701           Prob > F      =  0.0000
    Residual |  106.627377   263  .405427287           R-squared     =  0.4878
-------------+------------------------------           Adj R-squared =  0.4781
       Total |  208.188727   268  .776823609           Root MSE      =  .63673

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      points |   .0777297   .0071128    10.93   0.000     .0637243     .091735
       exper |   .2178447   .0497877     4.38   0.000     .1198115     .315878
     expersq |  -.0070821   .0027687    -2.56   0.011    -.0125338   -.0016305
         age |  -.0481375   .0349466    -1.38   0.170    -.1169481    .0206732
        educ |  -.0402709   .0528725    -0.76   0.447    -.1443781    .0638364
       _cons |   6.779038   .8454209     8.02   0.000     5.114384    8.443693
------------------------------------------------------------------------------
```

The OLS results are:

$$\begin{aligned} \log(\widehat{wage}) &= \underset{(.85)}{6.78} + \underset{(.007)}{.078}\, points + \underset{(.050)}{.218}\, exper - \underset{(.0028)}{.0071}\, exper^2 \\ &\quad - \underset{(.035)}{.048}\, age - \underset{(.053)}{040}\, edu \\ n &= 269, R^2 = .488, \overline{R}^2 = .478. \end{aligned}$$

(vi)
```
. test age educ

 ( 1)  age = 0.0
 ( 2)  educ = 0.0

       F(  2,   263) =    1.19
            Prob > F =    0.3061
```

The joint F test produced by Stata is about 1.19. With 2 and $263df$, this gives a p-value of roughly .31. Therefore, once scoring and years played are controlled for, there is no evidence for wage differnetials depending on age or years played in college.

5. (7.3)

(i) The t statistic on $hsize^2$ is over four in absolute value, so there is very strong evidence that it belongs in the equation. We obtain this by finding the turnaround point; this is the value of $hsize$ that maximizes $\widehat{sat}$

(other things fixed): $19.3/(2 \cdot 2.19) \approx 4.41$. Because $hsize$ is measured in hundreds, the optimal size of graduating class is about 441.

(ii) This is given by the coefficient on $female$ (since $black = 0$): nonblack females have SAT scores about 45 points lower than nonblack males. The t statistic is about -10.51, so the difference is very statically significant. (The very large sample size certainly contributes to the statistical significance.)

(iii) Because $female = 0$, the coefficient on $black$ implies that a black male has an estimated SAT score almost 170 points less than a comparable nonblack male. The t statistic is over 13 in absolute value, so we easily reject the hypothesis that there is no ceteris paribus difference.

(iv) We plug in $black = 1$, $female = 1$ for black females and $black = 0$ and $female = 1$ for nonblack females. The difference is therefore $-169.81 + 62.31 = -107.50$. Because the estimate depends on two coefficients, we cannot construct a t statistic from the information given. The easiest approach is to difine dummy variables for three of the four race/gender categories and choose nonblack females as the base group. We can then obtain the t statistic we want as the coefficient on the black females dummy variable.

6. (**7.5**)

(i) Following the hint,

$$\begin{aligned} \widehat{colGPA} &= \hat{\beta}_0 + \hat{\delta}_0(1 - noPC) + \hat{\beta}_1 hsGPA + \beta_2 ACT \\ &= (\hat{\beta}_0 + \hat{\delta}_0) - \hat{\delta}_0 noPC + \hat{\beta}_1 hsGPA + \beta_2 ACT \end{aligned}$$

For the specific estimates in equation (7.6), $\hat{\beta}_0 = 1.26$ and $\hat{\delta}_0 = .157$, so the new intercept is $1.26 + .157 = 1.417$. The coefficient on $noPC$ is $-.157$.

(ii) Nothing happens to the R-squared. Using $noPC$ in place of PC is simply a different way of including the same information on PC ownership.

(iii) It makes no sense to include both dummy variables in the regression: we cannot hold $noPC$ fixed while changing PC, we have only two groups based on PC owership so, in addition to the overall intercept, we need only to include one dummy variable. If we try to include both along with an intercept we have perfect multicollinearity (the dummy variable trap).

7. (**7.10**)

(i)
```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/WAGE2

. regress lwage educ exper tenure married black south urban

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  7,   927) =   44.75
       Model |  41.8377677     7  5.97682396           Prob > F      =  0.0000
    Residual |  123.818527   927  .133569069           R-squared     =  0.2526
-------------+------------------------------           Adj R-squared =  0.2469
       Total |  165.656294   934  .177362199           Root MSE      =  .36547

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0654307   .0062504    10.47   0.000     .0531642    .0776973
       exper |    .014043   .0031852     4.41   0.000      .007792     .020294
      tenure |   .0117473    .002453     4.79   0.000     .0069333    .0165613
     married |   .1994171   .0390502     5.11   0.000     .1227802    .2760541
       black |  -.1883499   .0376666    -5.00   0.000    -.2622717   -.1144282
       south |  -.0909036   .0262485    -3.46   0.001     -.142417   -.0393903
       urban |   .1839121   .0269583     6.82   0.000     .1310056    .2368185
       _cons |   5.395497    .113225    47.65   0.000      5.17329    5.617704
------------------------------------------------------------------------------
```

The estimated equation is

$$
\begin{aligned}
\log(\widehat{wage}) &= \underset{(0.11)}{5.40} + \underset{(.0063)}{.0654}\, educ + \underset{(.0032)}{.0140}\, exper + \underset{(.0025)}{.0117}\, tenure \\
&\quad + \underset{(0.039)}{.199}\, married - \underset{(.038)}{.188}\, black - \underset{(.026)}{.091}\, south + \underset{(.027)}{.184}\, urban \\
n &= 935, R^2 = .253.
\end{aligned}
$$

The coefficient on *black* implies that, at given levels of the other explanatory variables, black men earn about 18.8% less than nonblack men. The t statistic is about -4.95, and so it is very statistically significant.

(ii)
```
. gen expersq=exper* exper

. gen tenuresq=tenure* tenure

. regress lwage educ exper tenure married black south urban  expersq tenuresq

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  9,   925) =   35.17
       Model |  42.235332     9  4.69281467            Prob > F      =  0.0000
    Residual |  123.420962   925  .133428067           R-squared     =  0.2550
-------------+------------------------------           Adj R-squared =  0.2477
       Total |  165.656294   934  .177362199           Root MSE      =  .36528

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0642761   .0063115    10.18   0.000     .0518896    .0766625
       exper |   .0172146   .0126138     1.36   0.173    -.0075403    .0419695
      tenure |   .0249291   .0081297     3.07   0.002     .0089744    .0408838
     married |    .198547   .0391103     5.08   0.000     .1217917    .2753023
       black |  -.1906636   .0377011    -5.06   0.000    -.2646533    -.116674
       south |  -.0912153   .0262356    -3.48   0.001    -.1427035   -.0397271
       urban |   .1854241   .0269585     6.88   0.000     .1325171    .2383311
     expersq |  -.0001138   .0005319    -0.21   0.831    -.0011576      .00093
    tenuresq |  -.0007964    .000471    -1.69   0.091    -.0017208    .0001279
       _cons |   5.358676   .1259143    42.56   0.000     5.111565    5.605786
------------------------------------------------------------------------------

. test expersq tenuresq

 ( 1)  expersq = 0.0
 ( 2)  tenuresq = 0.0

       F(  2,   925) =    1.49
            Prob > F =    0.2260
```

The F statistic for joint significance of $exper^2$ and $tenure^2$, with 2 and $925 df$, is about 1.49 with p-value $\approx .226$. Because the p-value is above .20, these quadratics are jointly insignificant at the 20% level.

(iii)
```
. gen blackedu= black*educ

. regress lwage educ exper tenure married black south urban  blackedu

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  8,   926) =   39.32
```

```
       Model |  42.0055536     8   5.2506942           Prob > F      =  0.0000
    Residual |  123.650741   926  .133532117           R-squared     =  0.2536
-------------+------------------------------           Adj R-squared =  0.2471
       Total |  165.656294   934  .177362199           Root MSE      =  .36542

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0671153   .0064277    10.44   0.000     .0545008    .0797299
       exper |   .0138259   .0031906     4.33   0.000     .0075642    .0200876
      tenure |    .011787   .0024529     4.81   0.000     .0069732    .0166009
     married |   .1989077   .0390474     5.09   0.000     .1222761    .2755394
       black |   .0948094   .2553995     0.37   0.711    -.4064194    .5960383
       south |  -.0894495   .0262769    -3.40   0.001    -.1410187   -.0378803
       urban |   .1838523   .0269547     6.82   0.000      .130953    .2367516
    blackedu |  -.0226237   .0201827    -1.12   0.263    -.0622327    .0169854
       _cons |   5.374817   .1147027    46.86   0.000     5.149709    5.599924
------------------------------------------------------------------------------
```

We add the interaction $black \cdot educ$ to the equation in part (i). The coefficient on the interaction is about $-.0226$ (se $\approx .0202$). Therefore, the point estimate is that the return to another year of education is about 2.3 percentage points lower for black men than nonblack men. (The estimated return for nonblack men is about 6.7%.) This is nontrivial if it really reflects difference in the population. But the t statistic is only about 1.12 in absolute value, which is not enough to reject the null hypothesis that the return to education does not depend on race.

(iv)
```
. gen marrnonblck= married*(1- black)

. gen singblck=(1- married)* black

. gen marrblck= married* black

. regress lwage educ exper tenure south urban marrnonblck singblck marrblck

      Source |       SS       df       MS              Number of obs =     935
-------------+------------------------------           F(  8,   926) =   39.17
       Model |  41.8849419     8  5.23561773           Prob > F      =  0.0000
    Residual |  123.771352   926  .133662368           R-squared     =  0.2528
-------------+------------------------------           Adj R-squared =  0.2464
       Total |  165.656294   934  .177362199           Root MSE      =   .3656

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

```
-------------+----------------------------------------------------------------
        educ |   .0654751    .006253    10.47   0.000      .0532034    .0777469
       exper |   .0141462    .003191     4.43   0.000      .0078837    .0204087
      tenure |   .0116628   .0024579     4.74   0.000       .006839    .0164866
       south |  -.0919894   .0263212    -3.49   0.000     -.1436455   -.0403333
       urban |   .1843501   .0269778     6.83   0.000      .1314053    .2372948
  marrnonblck |   .1889147   .0428777     4.41   0.000      .1047659    .2730635
    singblck |  -.2408201   .0960229    -2.51   0.012     -.4292678   -.0523724
     marrblck |   .0094485   .0560131     0.17   0.866     -.1004788    .1193757
       _cons |   5.403793   .1141222    47.35   0.000      5.179825    5.627761
------------------------------------------------------------------------------
```

We choose the base group to be single, nonblack. Then we add dummy variables *marrnonblck*, *singblck*, and *marrblck* for the other three groups. The result is

$$\begin{aligned}
\log(\widehat{wage}) &= \underset{(0.11)}{5.40} + \underset{(.0063)}{.0655}\, educ + \underset{(.0032)}{.0141}\, exper + \underset{(.0025)}{.0117}\, tenure \\
&\quad - \underset{(0.026)}{.092}\, south + \underset{(.027)}{.184}\, urban + \underset{(.043)}{.189}\, marrnonblck \\
&\quad - \underset{(0.096)}{.241}\, singblck + \underset{(.0560)}{.0094}\, marrblck \\
n &= 935, R^2 = .253.
\end{aligned}$$

We obtain the ceteris paribus differential between married blacks and married nonblacks by taking the difference of their coefficients: $.0094 - .189 = -.1796$, or about $-.18$. That is, a married black man earns about 18% less than a comparable, married nonblack man.

8. (**7.12** using dataset GPA2-20)

(i) The two signs that are pretty clear are $\beta_3 < 0$ (because *hsperc* is defined so that the smaller the number the btter the student) and $\beta_4 > 0$. The effect of size of graduating class is not clear. It is also unclear whether males and females have systematically different GPAs. We may think that $beta_0 < 0$, that is, athletes do worse than other students with comparable characteristics. But remember, we are controlling for ability to some degree with *hsperc* and *sat*.

(ii) `. use http://fmwww.bc.edu/ec-p/data/wooldridge/GPA2-20`

```
. regress  colgpa hsize hsizesq hsperc sat female athlete

      Source |       SS       df       MS              Number of obs =     827
-------------+------------------------------           F(  6,   820) =   49.59
       Model |  90.9288519     6  15.1548087           Prob > F      =  0.0000
    Residual |  250.571787   820   .30557535           R-squared     =  0.2663
-------------+------------------------------           Adj R-squared =  0.2609
       Total |  341.500639   826   .41343903           Root MSE      =  .55279

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0422584   .0378294    -1.12   0.264    -.1165123    .0319955
     hsizesq |   .0023961   .0054545     0.44   0.661    -.0083104    .0131025
      hsperc |  -.0127884   .0012877    -9.93   0.000     -.015316   -.0102608
         sat |   .0013982   .0001478     9.46   0.000     .0011081    .0016882
      female |   .1334382   .0394927     3.38   0.001     .0559196    .2109569
     athlete |   .0035205    .101566     0.03   0.972    -.1958395    .2028805
       _cons |   1.498411   .1753761     8.54   0.000     1.154172    1.842649
------------------------------------------------------------------------------
```

The estimated equation is

$$
\begin{aligned}
\widehat{colgpa} &= \underset{(0.175)}{1.498} - \underset{(.0378)}{.0423}\, hsize + \underset{(.00545)}{.00240}\, hsize^2 - \underset{(.00129)}{.0128}\, hsperc \\
&\quad - \underset{(0.000148)}{.00140}\, sat + \underset{(.0395)}{.133}\, female + \underset{(.102)}{.00352}\, athlete \\
n &= 827, R^2 = .2663.
\end{aligned}
$$

Holding other factors fixed, an athlete is predicted to have a GPA about .00352 points higher than a nonathlete. The t statistic $.0352/.102 \approx .03$, which is very insignificant.

(iii)
```
. regress  colgpa hsize hsizesq hsperc female athlete

      Source |       SS       df       MS              Number of obs =     827
-------------+------------------------------           F(  5,   821) =   37.56
       Model |  63.5774308     5  12.7154862           Prob > F      =  0.0000
    Residual |  277.923208   821  .338517915           R-squared     =  0.1862
-------------+------------------------------           Adj R-squared =  0.1812
       Total |  341.500639   826   .41343903           Root MSE      =  .58182
```

```
------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0400371    .0398156    -1.01   0.315    -.1181895    .0381152
     hsizesq |   .0034527    .0057398     0.60   0.548    -.0078137    .0147191
      hsperc |  -.0160537    .0013058   -12.29   0.000    -.0186167   -.0134907
      female |   .0740543    .0410386     1.80   0.072    -.0064986    .1546072
     athlete |  -.1316444    .1058377    -1.24   0.214    -.3393888    .0760999
       _cons |    3.02014    .0735695    41.05   0.000     2.875733    3.164546
------------------------------------------------------------------------------
```

With *sat* dropped from the model, the coefficient on *athlete* becomes about $-.132$ (se $\approx .106$), the t statistic is -1.24, which is very insignificant.

(iv)
```
. gen femath= female* athlete

. gen maleath=(1- female)* athlete

. gen malenonath=(1- female)*(1- athlete)

. regress  colgpa hsize hsizesq hsperc sat femath maleath malenonath

      Source |       SS       df       MS              Number of obs =     827
-------------+------------------------------           F(  7,   819) =   42.46
       Model |  90.9320164     7  12.9902881           Prob > F      =  0.0000
    Residual |  250.568622   819  .305944594           R-squared     =  0.2663
-------------+------------------------------           Adj R-squared =  0.2600
       Total |  341.500639   826   .41343903           Root MSE      =  .55312

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0424362    .0378927    -1.12   0.263    -.1168144    .0319419
     hsizesq |   .0024077     .005459     0.44   0.659    -.0083075     .013123
      hsperc |  -.0127982     .001292    -9.91   0.000    -.0153343   -.0102621
         sat |   .0013982    .0001479     9.46   0.000     .0011079    .0016884
      femath |  -.0113654    .1781901    -0.06   0.949    -.3611284    .3383977
     maleath |  -.1236811    .1229176    -1.01   0.315    -.3649517    .1175895
  malenonath |  -.1341265    .0400919    -3.35   0.001    -.2128215   -.0554316
       _cons |   1.632741    .1685775     9.69   0.000     1.301846    1.963636
------------------------------------------------------------------------------
```

To facilitate testing the hypothesis that there is no difference between women athletes and women nonathletes, we should choose one of these as the base group. We choose female nonathletes. The estimation equation is

$$
\begin{aligned}
\widehat{colgpa} \;=\; & \underset{(.169)}{1.633} - \underset{(.0379)}{.0424}\, hsize + \underset{(.00546)}{.0024}\, hsize^2 - \underset{(.00129)}{.0128}\, hsperc \\
& + \underset{(0.00015)}{.0014}\, sat - \underset{(.178)}{.0114}\, female - \underset{(.123)}{.124}\, maleath - \underset{(.040)}{.134}\, malenonath \\
n \;=\; & 827, R^2 = .266.
\end{aligned}
$$

The coefficient on $femath = female \cdot athlete$ shows that $colgpa$ is predicted to be about .0114 points lower for a female athlete than a female nonathlete, other variables in the equation fixed.

(v)
```
. gen femsat=female*sat

. regress  colgpa hsize hsizesq hsperc sat female athlete  femsat

      Source |       SS       df       MS              Number of obs =     827
-------------+------------------------------           F(  7,   819) =   42.47
       Model |  90.9524481     7  12.9932069           Prob > F      =  0.0000
    Residual |  250.548191   819  .305919647           R-squared     =  0.2663
-------------+------------------------------           Adj R-squared =  0.2601
       Total |  341.500639   826   .41343903           Root MSE      =   .5531

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0419658   .0378654    -1.11   0.268    -.1162905    .0323589
     hsizesq |   .0023623   .0054589     0.43   0.665    -.0083528    .0130774
      hsperc |  -.0127783    .001289    -9.91   0.000    -.0153084   -.0102483
         sat |   .0014327   .0001932     7.42   0.000     .0010535    .0018119
      female |   .2139498   .2925759     0.73   0.465     -.360337    .7882366
     athlete |   .0050122   .1017651     0.05   0.961    -.1947388    .2047633
      femsat |  -.0000781   .0002812    -0.28   0.781      -.00063    .0004738
       _cons |   1.461552   .2200118     6.64   0.000     1.029698    1.893405
------------------------------------------------------------------------------

. regress  colgpa hsize hsizesq hsperc sat femath maleath malenonath  femsat

      Source |       SS       df       MS              Number of obs =     827
-------------+------------------------------           F(  8,   818) =   37.12
       Model |  90.9591932     8  11.3698992           Prob > F      =  0.0000
    Residual |  250.541445   818  .306285386           R-squared     =  0.2664
-------------+------------------------------           Adj R-squared =  0.2592
```

```
       Total |  341.500639   826   .41343903           Root MSE      =  .55343

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0422033   .0379218    -1.11   0.266    -.1166388    .0322323
     hsizesq |   .0023766    .005463     0.44   0.664    -.0083466    .0130998
      hsperc |  -.0127918   .0012929    -9.89   0.000    -.0153297    -.010254
         sat |   .0014357   .0001944     7.39   0.000     .0010542    .0018172
      femath |  -.0168791   .1792476    -0.09   0.925    -.3687186    .3349604
     maleath |  -.2066222   .3043929    -0.68   0.497    -.8041053    .3908609
  malenonath |  -.2220095   .2977459    -0.75   0.456    -.8064454    .3624265
      femsat |  -.0000849   .0002851    -0.30   0.766    -.0006445    .0004746
       _cons |   1.680639   .2330368     7.21   0.000     1.223219     2.13806
------------------------------------------------------------------------------
```

Whether we add the interaction $female \cdot sat$ to the equation in part (ii) or part (iv), the outcome is practically the same. For example, when $female \cdot sat$ is added to the equation in part (ii), its coefficient is about .000078 and its t statistic is about .28. There is very little evidence that the effect of *sat* differs by gender.

9. (**7.12** with dataset GPA2)

(i) The two signs that are pretty clear are $\beta_3 < 0$ (because *hsperc* is defined so that the smaller the number the btter the student) and $\beta_4 > 0$. The effect of size of graduating class is not clear. It is also unclear whether males and females have systematically different GPAs. We may think that $beta_0 < 0$, that is, athletes do worse than other students with comparable characteristics. But remember, we are controlling for ability to some degree with *hsperc* and *sat*.

(ii)
```
. use http://fmwww.bc.edu/ec-p/data/wooldridge/GPA2

. regress  colgpa hsize hsizesq hsperc sat female athlete

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  6,  4130) =  284.59
       Model |  524.819305     6  87.4698842           Prob > F      =  0.0000
    Residual |  1269.37637  4130  .307355053           R-squared     =  0.2925
-------------+------------------------------           Adj R-squared =  0.2915
       Total |  1794.19567  4136  .433799728           Root MSE      =   .5544
```

```
------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0568543   .0163513    -3.48   0.001    -.0889117   -.0247968
     hsizesq |   .0046754   .0022494     2.08   0.038     .0002654    .0090854
      hsperc |  -.0132126   .0005728   -23.07   0.000    -.0143355   -.0120896
         sat |   .0016464   .0000668    24.64   0.000     .0015154    .0017774
      female |   .1548814   .0180047     8.60   0.000     .1195826    .1901802
     athlete |   .1693064   .0423492     4.00   0.000     .0862791    .2523336
       _cons |   1.241365   .0794923    15.62   0.000     1.085517    1.397212
------------------------------------------------------------------------------
```

The estimated equation is

$$
\begin{aligned}
\widehat{colgpa} &= \underset{(0.079)}{1.241} - \underset{(.0164)}{.0569}\, hsize + \underset{(.00225)}{.00468}\, hsize^2 - \underset{(.0006)}{.0132}\, hsperc \\
&\quad - \underset{(0.00007)}{.00165}\, sat + \underset{(.018)}{.155}\, female + \underset{(.042)}{.169}\, athlete \\
n &= 4,137, R^2 = .293.
\end{aligned}
$$

Holding other factors fixed, an athlete is predicted to have a GPA about .169 points higher than a nonathlete. The t statistic $.169/.042 \approx 4.02$, which is very significant.

(iii)
```
. regress  colgpa hsize hsizesq hsperc female athlete

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  5,  4131) =  191.92
       Model |  338.217123     5  67.6434246           Prob > F      =  0.0000
    Residual |  1455.97855  4131   .35245184           R-squared     =  0.1885
-------------+------------------------------           Adj R-squared =  0.1875
       Total |  1794.19567  4136  .433799728           Root MSE      =  .59368

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0534038   .0175092    -3.05   0.002    -.0877313   -.0190763
     hsizesq |   .0053228   .0024086     2.21   0.027     .0006007     .010045
      hsperc |  -.0171365   .0005892   -29.09   0.000    -.0182916   -.0159814
      female |   .0581231   .0188162     3.09   0.002     .0212333     .095013
     athlete |   .0054487   .0447871     0.12   0.903    -.0823582    .0932556
       _cons |   3.047698   .0329148    92.59   0.000     2.983167    3.112229
------------------------------------------------------------------------------
```

With *sat* dropped from the model, the coefficient on *athlete* becomes about .0054 (se $\approx$ .0448), which is practically and statistically not different from zero. this happens because we do not control for SAT scores, and athletes score lower on average than nonathletes. Part (ii) shows that, once we account for SAT differences, athletes do better than nonathletes. Even if we do not control for SAT score, there is no difference.

(iv)
```
. gen femath= female* athlete

. gen maleath=(1- female)* athlete

. gen malenonath=(1- female)*(1- athlete)

. regress colgpa hsize hsizesq hsperc sat femath maleath malenonath

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  7,  4129) =  243.88
       Model |  524.821272     7  74.9744674           Prob > F      =  0.0000
    Residual |   1269.3744  4129  .307429015           R-squared     =  0.2925
-------------+------------------------------           Adj R-squared =  0.2913
       Total |  1794.19567  4136  .433799728           Root MSE      =  .55446

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0568006   .0163671    -3.47   0.001    -.0888889   -.0247124
     hsizesq |   .0046699   .0022507     2.07   0.038     .0002573    .0090825
      hsperc |  -.0132114    .000573   -23.06   0.000    -.0143349    -.012088
         sat |   .0016462   .0000669    24.62   0.000     .0015151    .0017773
      femath |   .1751106   .0840258     2.08   0.037     .0103748    .3398464
     maleath |   .0128034   .0487395     0.26   0.793    -.0827523    .1083591
  malenonath |  -.1546151   .0183122    -8.44   0.000    -.1905168   -.1187133
       _cons |    1.39619   .0755581    18.48   0.000     1.248055    1.544324
------------------------------------------------------------------------------
```

To facilitate testing the hypothesis that there is no difference between women athletes and women nonathletes, we should choose one of these as the base group. We choose female nonathletes. The estimation equation is

$$\begin{aligned}
\widehat{colgpa} &= \underset{(0.076)}{1.396} - \underset{(.0164)}{.0568}\, hsize + \underset{(.00225)}{.00467}\, hsize^2 - \underset{(.0006)}{.0132}\, hsperc \\
&\quad + \underset{(0.00007)}{.00165}\, sat + \underset{(.084)}{.175}\, female + \underset{(.049)}{.013}\, maleath - \underset{(.018)}{.155}\, malenonath \\
n &= 4,137, R^2 = .293.
\end{aligned}$$

The coefficient on $femath = female \cdot athlete$ shows that $colgpa$ is predicted to be about .175 points higher for a female athlete than a female nonathlete, other variables in the equation fixed.

(v)
```
. gen femsat=female*sat

. regress  colgpa hsize hsizesq hsperc sat female athlete  femsat

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  7,  4129) =  243.91
       Model |  524.867644     7  74.981092            Prob > F      =  0.0000
    Residual |  1269.32803  4129  .307417784           R-squared     =  0.2925
-------------+------------------------------           Adj R-squared =  0.2913
       Total |  1794.19567  4136  .433799728           Root MSE      =  .55445

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsize |  -.0569121   .0163537    -3.48   0.001    -.0889741   -.0248501
     hsizesq |   .0046864   .0022498     2.08   0.037     .0002757    .0090972
      hsperc |   -.013225   .0005737   -23.05   0.000    -.0143497   -.0121003
         sat |   .0016255   .0000852    19.09   0.000     .0014585    .0017924
      female |   .1023066   .1338023     0.76   0.445    -.1600179    .3646311
     athlete |   .1677568   .0425334     3.94   0.000     .0843684    .2511452
      femsat |   .0000512   .0001291     0.40   0.692     -.000202    .0003044
       _cons |   1.263743   .0974952    12.96   0.000       1.0726    1.454887
------------------------------------------------------------------------------

. regress  colgpa hsize hsizesq hsperc sat femath maleath malenonath  femsat

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  8,  4128) =  213.37
       Model |  524.873728     8  65.6092161           Prob > F      =  0.0000
    Residual |  1269.32195  4128  .307490781           R-squared     =  0.2925
-------------+------------------------------           Adj R-squared =  0.2912
       Total |  1794.19567  4136  .433799728           Root MSE      =  .55452
```

colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsize	-.0568198	.0163688	-3.47	0.001	-.0889114	-.0247282
hsizesq	.0046773	.002251	2.08	0.038	.0002641	.0090904
hsperc	-.0132236	.0005738	-23.04	0.000	-.0143487	-.0120986
sat	.001624	.0000858	18.93	0.000	.0014558	.0017922
femath	.1779989	.0843247	2.11	0.035	.0126771	.3433207
maleath	.0652958	.1361172	0.48	0.631	-.2015673	.3321589
malenonath	-.0990198	.1358427	-0.73	0.466	-.3653447	.1673051
femsat	.0000539	.0001306	0.41	0.680	-.0002021	.00031
_cons	1.364334	.1079746	12.64	0.000	1.152646	1.576023

Whether we add the interaction $female \cdot sat$ to the equation in part (ii) or part (iv), the outcome is practically the same. For example, when $female \cdot sat$ is added to the equation in part (ii), its coefficient is about .000051 and its t statistic is about .40. There is very little evidence that the effect of sat differs by gender.