MULTIPLE REGRESSION OF CPS DATA

A further inspection of the relationship between hourly wages and education level can show whether other factors, such as gender and work experience, influence wages. Linear regression showed that hourly wages increase substantially with education, but there was still considerable variation in wages among people with the same level of education. This variation may be due to many factors, such as work experience, occupation and type of industry. Using multiple regression, we can evaluate which factors account for the variation in hourly wages for people with similar education.

Table 6.1 reports mean hourly wages classified simultaneously by years of education and work experience. The margins of the table replicate the mean hourly wages in Table 3.2 for years of education and work experience. The body of the table shows that for those with identical years of education, hourly wages increase with work experience, indicating that some of the variation within education levels is explained by time in the labor force. Table 6.2 shows that for those with similar years of education, hourly wages are always higher for males than females, so that gender also explains part of the variation in hourly wages at each education level.

Separate linear regressions of log hourly wages against years of education and gender (Table 6.3) show that education alone explains about 16.2% of the variation in wages while gender alone explains about 7% (Table 6.3A and B). Both education and gender are significant univariate predictors of hourly wages. Overall, female wages are exp(– 0.319)=0.73 that of male wages, with a 95% confidence interval of (exp(– 0.392, – 0.247))= (0.68, 0.78) (Table 6.3B). Alternatively, one can say that males wages are higher than females by a factor of exp(0.319)=1.38, with a 95% confidence interval of (exp(0.392, 0.247))= (1.28, 1.48). Note that these factors are not symmetric about 1.0 but are inverses of one another so that 1/0.73=1.38.

Multiple regression indicates that about 22.9% of the variation in hourly wages is explained by education and gender combined (Table 6.3C). Gender thus explains an additional 6.7% of the variation in hourly wages beyond that explained by education alone (16.2%). The variation explained by the combined variables is almost equal to the sum of the variation explained by the variables separately. This is because gender is uncorrelated with years of education (r= – 0.01) since males and females attain similar levels of education. This lack of correlation leads to an important result in multiple regression: the regression coefficient of one variable does not change in the presence of the other variable. In fact, controlling for education, females still earn 73% (68%, 78%) that of males (Table 6.3C).

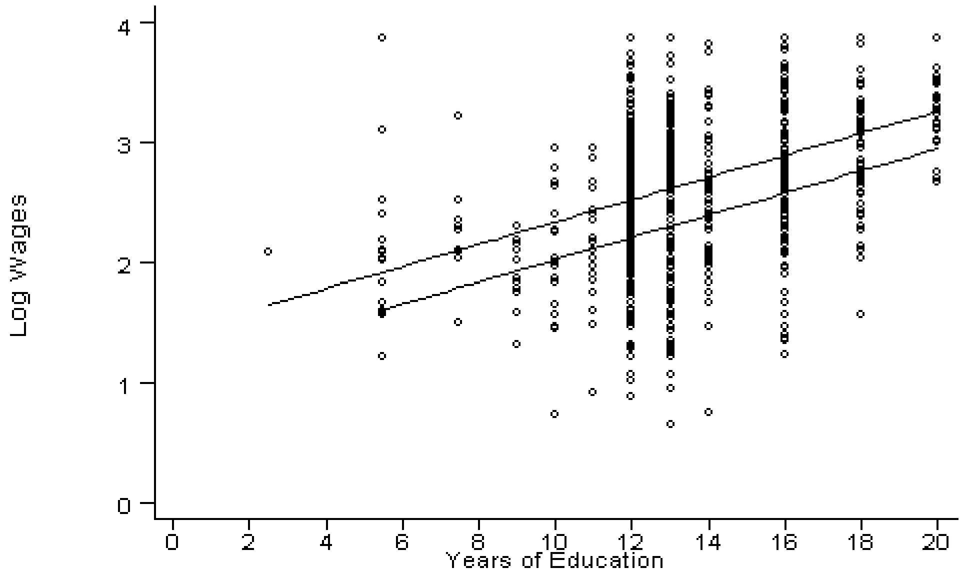

Figure 6.1 Plot of log wages against years of education, superimposing the fitted regression lines of log wages against education for males and females.

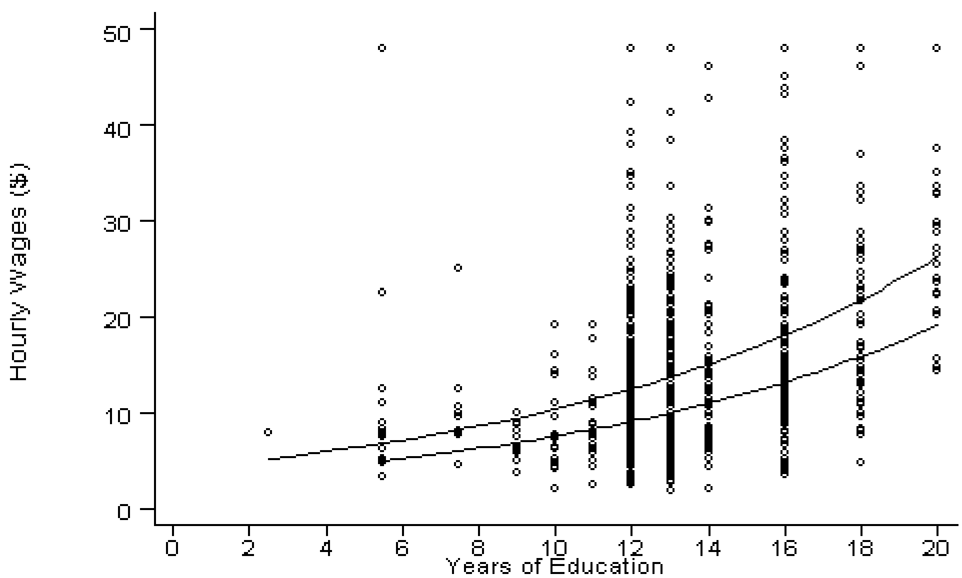

Figure 6.2 Plot of wages against years of education, superimposing the back-transformed fitted regression lines of log wages against education for males and females.

Finally, since the gender effect has not changed in the presence of education, we can also conclude that differences in education do not explain why females generally earn less than males.

Figure 6.1 demonstrates that regressing log wages on education and gender has the effect of fitting separate parallel lines to the relationship between log hourly wages and education for males and females. Parallel lines mean that the increase in log wages for an additional year of education is the same for males and females, and averages about exp(.0925)= 1.10 with a 95% confidence interval of (1.08, 1.11) per year of education (Table 6.3C). The effect has not changed after controlling for gender. The distance between the lines for males and females represents the effect of gender: the line for males is 0.3137 log dollars higher than the line for females (Table 6.3C). This means that for a given education level, wages are higher for males than females by a factor of exp(0.3137)=1.37, with a 95% confidence interval of (1.28 , 1.46). This difference is best seen in the plot of wages against education in Figure 6.2, where the back-transformed fitted regression lines are no longer parallel. Although the relative wage increase of 37% is constant over all years of education, the absolute increase in male salaries over that of females' is much higher for higher salaries because of the effect of compounding described earlier.

We can perform a similar analysis with years of education and experience. From the regression output in Table 6.4, education alone explains about 16.2% of the variation in wages, work experience alone, about 2.4%, but together they explain about 21.7%. The whole is now greater than the sum of its parts! This can happen when two variables are negatively correlated (r= –.186). The negative correlation is due to the fact that those who spend more time in school have less time to spend in the work force, all other things being equal, such as age. Surprisingly, the effect of work experience alone is small: for every 10 years of experience, hourly wages increase by a factor of 1.09 (1.05, 1.12), about the same as for one additional year of education (Table 6.4B). Controlling for education increases the effect since it removes confounding due to education (see Chapter 6). For every 10 years of experience, hourly wages increase by a factor of 1.14 (1.10, 1.17) (Table 6.4C). This effect is far smaller than inflation: an inflation rate of 3% per year would have caused wages to increase 34% over 10 years since 1.03^{10} =1.34.

A regression of log wages on all three variables explains 28.9% of the variation in wages (Table 6.4D). The effect of gender is similar to before, since gender is uncorrelated with both experience (r=0.04) and education. The effects of education and work experience also do not change when gender is added to the model (compare to Table 6.4C). There is still about 70% unexplained variation in wages!

Chapter 7 explores complex modeling issues such as confounding, multi-colinearity, pooled tests, categorical variables and interactions. With these tools in place, a full analysis of the determinants of wages will be possible.

Table 6.1

Means, Standard Deviations and Frequencies of Hourly Wages

Years of Education	Years of Work Experience					
	exp<=5	5<x<=10	10<x<=20	20<x<=30	exp>30	Total
Educ<12	6.610577	8.3096154	8.506556	8.6632116	11.499039	9.310918
	2.2335515	4.5823646	3.2871646	3.8071621	9.4367496	6.0386959
	4	10	22	25	25	86
Educ=12	8.5617234	10.222842	11.647422	15.137898	13.069812	12.522641
	3.8917755	6.1940197	7.0772693	7.2318973	6.6605462	6.9705726
	26	45	102	93	96	362
Educ=13	6.0346955	11.595442	12.601342	17.252274	16.064233	13.730448
	2.2878627	5.0236677	6.807096	8.68351	9.4877654	8.0749169
	18	27	67	52	38	202
13<Educ<=16	12.018377	11.547343	19.680886	18.701486	18.666967	16.868053
	5.1686776	4.4477838	10.56787	8.6071216	12.906913	9.5829901
	40	38	78	66	31	253
Educ>16	17.574786	22.328942	28.116649	22.697912	26.153953	24.389087
	6.5705128	11.500545	13.368682	10.918406	10.881327	11.893388
	9	15	35	32	9	100
Total	10.274019	12.073586	15.587707	16.724456	14.907941	14.769702
	5.5195622	7.2312446	10.452645	8.8239055	9.4999288	9.257249
	97	135	304	268	199	1003

Table 6.2

Means, Standard Deviations and Frequencies of Hourly Wages

Years of Education	Gender: Male	Gender: Female	Total
Educ<12	10.609443	7.225408	9.310918
	6.9104135	3.46186	6.0386959
	53	33	86
Educ=12	14.280749	10.601934	12.522641
	7.5401576	5.7210515	6.9705726
	189	173	362
Educ=13	16.397839	10.955284	13.730448
	8.9813645	5.8753599	8.0749169
	103	99	202
13<Educ<=16	19.477352	14.110256	16.868053
	10.390313	7.7854763	9.5829901
	130	123	253
Educ>16	26.977737	20.165499	24.389087
	13.00519	8.371838	11.893388
	62	38	100
Total	17.048445	12.14377	14.769702
	10.240742	7.132306	9.257249
	537	466	1003

Table 6.3

reg lnwage educ

```
  Source |       SS       df       MS              Number of obs =    1003
---------+------------------------------           F(  1,  1001) =  193.52
   Model |  59.3347547     1  59.3347547           Prob > F      =  0.0000
Residual |   306.91002  1001  .306603416           R-squared     =  0.1620
---------+------------------------------           Adj R-squared =  0.1612
   Total |  366.244774  1002  .365513747           Root MSE      =  .55372

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .0932696   .0067046     13.911   0.000       .0801129    .1064263
   _cons |   1.259661   .0918705     13.711   0.000       1.079381    1.439942
------------------------------------------------------------------------------
```

reg lnwage gender

```
  Source |       SS       df       MS              Number of obs =    1003
---------+------------------------------           F(  1,  1001) =   74.73
   Model |  25.4432335     1  25.4432335           Prob > F      =  0.0000
Residual |  340.801541  1001   .34046108           R-squared     =  0.0695
---------+------------------------------           Adj R-squared =  0.0685
   Total |  366.244774  1002  .365513747           Root MSE      =  .58349

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  gender |  -.3193424   .0369406     -8.645   0.000      -.3918323   -.2468524
   _cons |   2.982048   .0571544     52.175   0.000       2.869892    3.094204
------------------------------------------------------------------------------
```

reg lnwage educ gender

```
  Source |       SS       df       MS              Number of obs =    1003
---------+------------------------------           F(  2,  1000) =  148.54
   Model |   83.883979     2  41.9419895           Prob > F      =  0.0000
Residual |  282.360795  1000  .282360795           R-squared     =  0.2290
---------+------------------------------           Adj R-squared =  0.2275
   Total |  366.244774  1002  .365513747           Root MSE      =  .53138

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

educ |	.0925705	.0064345	14.387	0.000	.0799438	.1051973
gender |	-.313703	.0336436	-9.324	0.000	-.3797231	-.247683
_cons |	1.728516	.1014949	17.031	0.000	1.529349	1.927684

Table 6.4

reg lnwage educ

```
  Source |       SS       df       MS              Number of obs =    1003
---------+------------------------------           F(  1,  1001) =  193.52
   Model |  59.3347547     1  59.3347547           Prob > F      =  0.0000
Residual |   306.91002  1001  .306603416           R-squared     =  0.1620
---------+------------------------------           Adj R-squared =  0.1612
   Total |  366.244774  1002  .365513747           Root MSE      =  .55372

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .0932696   .0067046     13.911   0.000       .0801129    .1064263
   _cons |   1.259661   .0918705     13.711   0.000       1.079381    1.439942
------------------------------------------------------------------------------
```

reg lnwage exper

```
  Source |       SS       df       MS              Number of obs =    1003
---------+------------------------------           F(  1,  1001) =   24.56
   Model |   8.7714897     1   8.7714897           Prob > F      =  0.0000
Residual |  357.473285  1001  .357116168           R-squared     =  0.0239
---------+------------------------------           Adj R-squared =  0.0230
   Total |  366.244774  1002  .365513747           Root MSE      =  .59759

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   exper |    .008377   .0016903      4.956   0.000       .0050601    .0116939
   _cons |   2.345581   .0389295     60.252   0.000       2.269188    2.421974
------------------------------------------------------------------------------
```

reg lnwage educ exper

```
  Source |       SS       df       MS              Number of obs =    1003
---------+------------------------------           F(  2,  1000) =  138.21
   Model |  79.3139552     2  39.6569776           Prob > F      =  0.0000
Residual |  286.930819  1000  .286930819           R-squared     =  0.2166
---------+------------------------------           Adj R-squared =  0.2150
   Total |  366.244774  1002  .365513747           Root MSE      =  .53566

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

```
    educ |   .1034977   .0066008     15.680   0.000       .0905448    .1164507
   exper |   .0128666   .0015419      8.345   0.000       .0098408    .0158924
   _cons |   .8628728   .1007955      8.561   0.000       .6650779    1.060668
------------------------------------------------------------------------------
```

reg lnwage educ gender exper

```
  Source |       SS       df       MS              Number of obs =    1003
---------+------------------------------           F(  3,   999) =  135.02
   Model |  105.659586     3  35.2198619           Prob > F      =  0.0000
Residual |  260.585189   999  .260846035           R-squared     =  0.2885
---------+------------------------------           Adj R-squared =  0.2864
   Total |  366.244774  1002  .365513747           Root MSE      =  .51073

------------------------------------------------------------------------------
  lnwage |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .1032311   .0062936     16.402   0.000       .0908808    .1155813
  gender |  -.3252252    .032361    -10.050   0.000      -.3887285   -.2617218
   exper |   .0134428   .0014713      9.137   0.000       .0105556      .01633
   _cons |   1.331179   .1068058     12.464   0.000        1.12159    1.540769
------------------------------------------------------------------------------
```