5.1. The Basics

Education is not the only factor that affects pay. As shown in Figure 4.2, even for workers with the *same* education, there is remarkable variation in wages. Surely, some of this variation is due to work experience, unionization, industry, occupation, region, and demographics, such as gender, race, marital status, etc. These easily can be accounted for using multiple regression.

For example, one could think of wages as a function of education and work experience:

\[
Wage = f(Education, Experience).
\]

The longer one spends on a job, the better one gets. If people are paid for their productivity, then workers with more work experience should be more productive, and therefore, paid more. That is,

\[
\frac{\Delta Wage}{\Delta Experience} > 0,
\]

other things equal.

The complete relationship between wages, education, and experience can be written as

\[
\ln(Wage_i) = \beta_1 + \beta_2 Education_i + \beta_3 Experience_i + u_i,
\]

where wages are measured in natural logs. This is a multiple regression model of wages. Because there is more than one explanatory variable, each parameter is interpreted as a partial derivative, or the change in the dependent variable for a change in the explanatory variable, holding all other variables constant. For example,

\[
\beta_3 = \frac{\partial \ln(Wage)}{\partial Experience} = \frac{\Delta \ln(Wage)}{\Delta Experience} \bigg|_{Education}
\]

is the effect of experience on the log wage, *holding education constant*. Other ways of saying "holding experience constant" are "controlling for experience" or "accounting for the effect of experience." Because pay is measured in natural logs, \( \beta_3 \) also can be interpreted as

\[
\beta_3 = \frac{\% \Delta Wage}{\Delta Experience} \bigg|_{Education},
\]

or the "return to experience" in the labor market.

If we group all workers according to their education level (less than high school, high school, some college, college graduates, and more than college), we can compare wages and work experience *within* education categories. This is really what multiple regression does. By looking *within* categories, you are *holding education constant*. From the univariate analysis in Chapter 4, we know that wages increase with education level. Table 5.1 shows that within any given education category (i.e., reading across rows), hourly wages rise with greater work
experience. This suggests $\beta_1$ is positive, so that wages increase with work experience controlling for education, but also that work experience explains some of the residual variation in wages within education levels.

Table 5.1

Means, Standard Deviations and Frequencies of Hourly Wages

<table>
<thead>
<tr>
<th>Years of Education</th>
<th>exp&lt;=5</th>
<th>5&lt;x&lt;=10</th>
<th>10&lt;x&lt;=20</th>
<th>20&lt;x&lt;=30</th>
<th>exp&gt;30</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>10</td>
<td>22</td>
<td>25</td>
<td>25</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>45</td>
<td>102</td>
<td>93</td>
<td>96</td>
<td>362</td>
</tr>
<tr>
<td></td>
<td>2.2878627</td>
<td>5.0236677</td>
<td>6.807096</td>
<td>8.68351</td>
<td>9.4877654</td>
<td>8.0749169</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>27</td>
<td>67</td>
<td>52</td>
<td>38</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td>5.1686776</td>
<td>4.4477838</td>
<td>10.56787</td>
<td>8.6071216</td>
<td>12.906913</td>
<td>9.5829901</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>38</td>
<td>78</td>
<td>66</td>
<td>31</td>
<td>253</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>15</td>
<td>35</td>
<td>32</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>97</td>
<td>135</td>
<td>304</td>
<td>268</td>
<td>199</td>
<td>1003</td>
</tr>
</tbody>
</table>

Likewise,

$$\beta_2 = \frac{\partial \ln(Wage)}{\partial Education} = \frac{\Delta \ln(Wage)}{\Delta Education}$$

is the effect of education on the log wage, holding experience constant. $\beta_2$ also can be expressed as

$$\beta_2 = \frac{\% \Delta Wage}{\Delta Education}_{Experience}$$

or the return to education in the labor market.
If we group workers according to years of work experience (0-5, 5-10, 11-20, 21-30, >30), we can compare wages and education \textit{within} work experience categories. Again, this is what multiple regression does. By looking \textit{within} experience categories, we are \textit{holding experience constant}. In Table 5.1, within any given experience category (reading down columns), the hourly wage rises with education. This suggests $\beta_2$ is positive, so that wages increase with education even when controlling for work experience.

Importantly, multiple regression recognizes possible \textit{interdependence} among explanatory variables. For example, for any individual, education and work experience are determined \textit{in part} by the underlying decision to allocate time. Individuals can go to school or work. Those with more education will have less work experience, and vice versa, holding other factors such as age constant. Thus, education and experience are interdependent. In fact, they are inversely correlated since the sample correlation coefficient, $r = -0.186$.

This interdependence implies that some of the population variation in education and experience is common. The Venn diagram in Figure 5.1 illustrates this. The two circles represent the variation in education and experience, respectively. Area B is the intersection and represents the variation shared by the variables. This is the co-variation between education and experience. Area A is the remaining variation in education and is due to influences other than experience, and hence, is \textit{independent} of experience. Similarly, area C is the remaining variation in experience, \textit{independent} of education.

When estimating parameters, least squares uses only the \textit{independent} variation in each explanatory variable to estimate that variable's parameter. To estimate $\beta_2$, only the independent part of education is used. The formula for the least squares estimator of $\beta_2$ is

$$\hat{\beta}_2 = \frac{Cov(\ln(Wage), \text{Independent Part of Education})}{Var(\text{Independent Part of Education})}. \quad (6)$$

and for $\beta_3$,

$$\hat{\beta}_3 = \frac{Cov(\ln(Wage), \text{Independent Part of Experience})}{Var(\text{Independent Part of Experience})}. \quad (7)$$

Table 5.2 shows parameter estimates, standard errors and 95% confidence intervals for simple and multiple regression models of the log wage.

-------------------------------

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>0.0933</td>
<td>---</td>
<td>0.1035</td>
</tr>
<tr>
<td></td>
<td>(0.0067)</td>
<td>----</td>
<td>(0.0066)</td>
</tr>
<tr>
<td></td>
<td>(0.0801, 0.1064)</td>
<td>---</td>
<td>(0.0066)</td>
</tr>
</tbody>
</table>

Table 5.2 Regression of Log Wages against Education and Experience
Experience 0.0084 0.0129
----       (0.0017)      (0.0015)
(0.0051, 0.0117)   (0.0098, 0.0159)

Constant 1.2597 2.3456 0.8629
(0.0919)     (0.0389)      (0.1008)
(1.0793, 1.4399)  (2.2692, 2.4220)   (0.6651, 1.0607)

R² 0.162 0.024 0.217

For the simple regression model 1,
\[ \ln(Wage_i) = \beta_1 + \beta_2 Education_i + u_i, \]
an additional year of education is estimated to raise log wages by 0.0933 or in terms of relative change in wages, by a factor of \(\exp(0.0933) = 1.098\) with a 95% confidence interval of \((\exp(0.0801), \exp(0.1064)) = (1.08, 1.11)\). Economists often say that the increase in percent wages is 9.3\%, an approximation. This is a moderately good return to a one-year investment!

Alternatively, for model 2,
\[ \ln(Wage_i) = \beta_1 + \beta_2 Experience_i + u_i, \]
an additional year of experience is estimated to raise the log wage by 0.0084 or to raise the wage by a factor of \(\exp(0.0084)=1.0084\) or 0.84\%. To put this finding in a more meaningful context, an additional 10 years of experience raises the log wages by 0.084, or raises wages by a factor of \(\exp(0.084)=1.088\) or 8.8\%. \(R^2\) is 0.024, which means that variation in experience alone explains just 2.4\% of the sample variation in log wages.

The estimates for the multiple regression model 3,
\[ \ln(Wage_i) = \beta_1 + \beta_2 Education_i + \beta_3 Experience_i + u_i, \]
show that together, education and experience explain 21.7\% of the variation in log wages. This is much more than both explain individually (16.2\% and 2.4\%). So, the whole is greater than the sum of its parts!

Accounting for the effect of experience on wages, an additional year of education is estimated to raise the log wage by 0.1035 or the actual wage by a factor of \(\exp(0.1035)=1.109\) or 10.9\% (Economists 10.4\%). In addition, accounting for the effect of education on wages, an additional year of experience is estimated to raise the log wage by 0.0129 or wages by a factor of 1.013 or 1.3\%. Surprisingly, the return to an additional year of experience is significantly less than the return to an additional year of education. In fact, based on these estimates, it would require an additional 8.4 years of work experience to raise wages by the same percent as an additional year of education (10.9/1.3=8.4). Education seems like a good deal!

Interestingly, the estimated effects of education and experience on wages change substantially from simple to multiple regression. An additional year of education is estimated to raise the wages by 9.8\% in model (1) but by 10.9\% in model (3). That is, the estimated return rises by more than a percentage point once differences in work experience are taken into account.
Because this is a big difference in the return on an investment---you would much prefer a 10.4% to a 9.3% return---it is natural to ask: "Why did this happen?"

The answer is at the heart of multiple regression. There are many less-educated (but more-experienced) workers that earn as much as more-educated (but less-experienced) workers. Without accounting for differences in experience, the better educated appear to get a lower return to education. Is this "low return" really because of education? No, it is because of experience.

Simple linear regression does not account for experience; however, multiple regression does. Once differences in experience across workers are taken into account, an additional year of education has a much bigger payoff and the estimated return to education rises. Because education and experience are correlated (or interdependent), simple regression confuses or "confounds" the effect of education on wages with the effect of experience on wages. By acknowledging potential correlation between the explanatory variables, multiple regression neatly sorts out each variable's independent effect. Section 6 will discuss "confounding effects" in more detail.

5.2. Gender and Wages

A question of great public interest is whether there is gender inequality in earnings, and, if so, what accounts for it. The basic comparison of average wages for men and women in section 3 showed that women earn $4.90 per hour less than men. Because men's average earnings were $17.05, this implies that, on average, women earn about 28.7% less than men ($4.90/17.05=0.287).

If pay is based solely on productivity, then this differential could be economically rational only if there were some innate underlying difference in productivity between the sexes. In this case, men would have to be more productive than women to justify their higher wages. If one believes the sexes are equal, then gender difference in wages must be caused by something else. One view is that there is labor market discrimination against women. Another is that there are other, confounding factors that affect wages but happen to be correlated with gender. Multiple regression can account for these additional factors. If gender-based wage differentials exist even after controlling for many possible confounding influences, then more credence might be given to the discrimination explanation.

The effect of gender on wages can be modeled simply as

$$\ln(Wage_i) = \beta_1 + \beta_2 Female_i + u_i,$$

where Female is an indicator variable that is 1 if the worker is female and 0 otherwise (which, of course, means male). A 0-1 indicator variable such as this is frequently referred to as a categorical or dummy variable. $\beta_2$ is the relative change in the wage from going from the 0 category to the 1 category. The way to think about $\beta_2$ in this context is to observe the wage of worker first as a male, and then "transform" the worker into a female and observe the wage. If the wage rises, $\beta_2$ is positive and women earn more than men. That is, there is a labor market "premium" to being a woman. If the wage does not change, $\beta_2$ is zero and gender has no effect on wages. Finally, if the wage falls, $\beta_2$ is negative and men earn more than women. That is, there is a labor market discount to being a woman.

Based on the comparison of mean wages above, one expects $\beta_2$ to be negative. The least squares estimate is shown in column (1) of Table 5.3.
Table 5.3 Regression of Log Wages against Female, Education and Experience

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.0933</td>
<td>0.1035</td>
<td>(0.0067)</td>
</tr>
<tr>
<td></td>
<td>(0.0801, 0.1064)</td>
<td>----</td>
<td>(0.0066)</td>
</tr>
<tr>
<td></td>
<td>(0.0801, 0.1064)</td>
<td></td>
<td>(0.0801, 0.1064)</td>
</tr>
<tr>
<td>Education</td>
<td>0.0933</td>
<td>0.1035</td>
<td>(0.0067)</td>
</tr>
<tr>
<td></td>
<td>(0.0801, 0.1064)</td>
<td>----</td>
<td>(0.0066)</td>
</tr>
<tr>
<td></td>
<td>(0.0801, 0.1064)</td>
<td></td>
<td>(0.0801, 0.1064)</td>
</tr>
<tr>
<td>Experience</td>
<td>0.0084</td>
<td>0.0129</td>
<td>(0.0017)</td>
</tr>
<tr>
<td></td>
<td>(0.0051, 0.0117)</td>
<td>(0.0015)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0051, 0.0117)</td>
<td></td>
<td>(0.0098, 0.0159)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.2597</td>
<td>2.3456</td>
<td>0.8629</td>
</tr>
<tr>
<td></td>
<td>(0.0919)</td>
<td>(0.0389)</td>
<td>(0.1008)</td>
</tr>
<tr>
<td></td>
<td>(1.0793, 1.4399)</td>
<td>(2.2692, 2.4220)</td>
<td>(0.6651, 1.0607)</td>
</tr>
<tr>
<td>R²</td>
<td>0.162</td>
<td>0.024</td>
<td>0.217</td>
</tr>
</tbody>
</table>

\( \hat{\beta}_2 = -0.3193 \), which says that females earn 31.9% less than males. The 95% confidence interval is (-0.3918, -0.2468) and does not include zero. Therefore, this discount to being female is statistically significantly different from zero.

In theory, one explanation for this wage differential could be differences in education between men and women. If men had more education than women on average, then that could explain the gender differential. Unfortunately, this is not the case. From section 3 above, the sample mean education is 13.48 years for men, but 13.42 years for women. That is, men and women have almost identical education levels! In fact, the sample correlation between Education and Female is -0.01. Effectively, the variables are uncorrelated. Therefore, differences in education will not explain the gender difference in wages.

This can be seen in two ways. First, Table 5.4 compares mean hourly wages for males and females within education categories. Remember, this is akin to what least squares does to estimate the effect of Female on wages holding Education constant in a multiple regression.

Table 5.4

<table>
<thead>
<tr>
<th>Years of Education</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
</tbody>
</table>

---

\( \hat{\beta}_2 = -0.3193 \), which says that females earn 31.9% less than males. The 95% confidence interval is (-0.3918, -0.2468) and does not include zero. Therefore, this discount to being female is statistically significantly different from zero.

In theory, one explanation for this wage differential could be differences in education between men and women. If men had more education than women on average, then that could explain the gender differential. Unfortunately, this is not the case. From section 3 above, the sample mean education is 13.48 years for men, but 13.42 years for women. That is, men and women have almost identical education levels! In fact, the sample correlation between Education and Female is -0.01. Effectively, the variables are uncorrelated. Therefore, differences in education will not explain the gender difference in wages.

This can be seen in two ways. First, Table 5.4 compares mean hourly wages for males and females within education categories. Remember, this is akin to what least squares does to estimate the effect of Female on wages holding Education constant in a multiple regression.

---

\( \hat{\beta}_2 = -0.3193 \), which says that females earn 31.9% less than males. The 95% confidence interval is (-0.3918, -0.2468) and does not include zero. Therefore, this discount to being female is statistically significantly different from zero.

In theory, one explanation for this wage differential could be differences in education between men and women. If men had more education than women on average, then that could explain the gender differential. Unfortunately, this is not the case. From section 3 above, the sample mean education is 13.48 years for men, but 13.42 years for women. That is, men and women have almost identical education levels! In fact, the sample correlation between Education and Female is -0.01. Effectively, the variables are uncorrelated. Therefore, differences in education will not explain the gender difference in wages.

This can be seen in two ways. First, Table 5.4 compares mean hourly wages for males and females within education categories. Remember, this is akin to what least squares does to estimate the effect of Female on wages holding Education constant in a multiple regression.

---

\( \hat{\beta}_2 = -0.3193 \), which says that females earn 31.9% less than males. The 95% confidence interval is (-0.3918, -0.2468) and does not include zero. Therefore, this discount to being female is statistically significantly different from zero.

In theory, one explanation for this wage differential could be differences in education between men and women. If men had more education than women on average, then that could explain the gender differential. Unfortunately, this is not the case. From section 3 above, the sample mean education is 13.48 years for men, but 13.42 years for women. That is, men and women have almost identical education levels! In fact, the sample correlation between Education and Female is -0.01. Effectively, the variables are uncorrelated. Therefore, differences in education will not explain the gender difference in wages.

This can be seen in two ways. First, Table 5.4 compares mean hourly wages for males and females within education categories. Remember, this is akin to what least squares does to estimate the effect of Female on wages holding Education constant in a multiple regression.
It is easy to see that within all categories, males are paid more females. Therefore, education cannot explain the gender wage differential.

Second, column (2) of Table 5.3 gives parameter estimates for the multiple regression model

$$\ln(Wage_i) = \beta_1 + \beta_2 Female_i + \beta_3 Education + u_i.$$  \hspace{1cm} (9)

$\beta_2$ measures the effect of being female on wages (in relative terms), controlling for education. $\beta_2 = -0.3137$, which says that females earn 31.4% less than males, even after accounting for any differences in education between the sexes! This estimate is virtually unchanged from its value in column (1). The 95% confidence interval is (-0.3797, -0.2477) and does not include zero. Therefore, this discount to being female is statistically significantly different from zero.

The estimated relationship between wages, education, and gender also can be illustrated graphically. Figure 5.1 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 (or the return to education) is the same for males and females, and averages about 0.093 (or 9.3%). The distance between the lines for males and females represents the effect of gender: the line for males is 0.3137 log dollars (or 31.37%) higher than the line for females. Figure 5.2 shows the same relationship, but expressed in terms of the wage rather than the log wage.

Another potential explanation for the gender wage differential could be differences in work experience between men and women. If men had more experience...
than women on average, then that could explain the gender differential. Again, this is not
the case. From section 3, the sample mean experience is 19.72 years for men, but even
more, 20.63 years, for women. That is, women have slightly more experience than men.
The sample correlation between *Experience* and *Female* is 0.041. Effectively, the
variables are uncorrelated. Therefore, differences in experience will not explain the
gender difference in wages.

This is illustrated nicely in column (2) of Table 5.2, which gives parameter
estimates for the multiple regression model

\[
\ln(Wage_{ie}) = \beta_1 + \beta_2 Female_i + \beta_3 Education + \beta_4 Experience + u_i. \tag{9}
\]

\(\beta_2\) measures the effect of being female on wages (in relative terms), controlling for
*education and experience*. \(\beta_2^{\prime} = -0.3252\), which says that females earn 32.5% less than
males, even after accounting for any differences in education and experience between the
sexes. The 95% confidence interval is (-0.3887, -0.2617) and does not include zero.
Therefore, this discount to being female is statistically significantly different from zero.
The \(R^2\) is 0.289, which means that variation in education, experience, and gender
explains 28.9% of the sample variation in log wages.