The Basics of Simple Linear Regression

1. Terminology

Differences in hourly wages might be determined by differences in education.[1] In particular, the relationship between the hourly wage and education can be thought of as a mathematical function:

$$Wage = f(Education)\,. \tag{1}$$

$Wage$ is the *dependent variable*. The dependent variable measures the behavior or outcome to be explained. The letter Y is often used as mathematical shorthand for a generic dependent variable.

$Education$ is an *explanatory variable*. Explanatory variables measure the factors that explain the behavior or outcome.[2] The letter X is used often as mathematical shorthand for a generic explanatory variable. For now, assume that the dependent variable is explained by just one explanatory variable.

f translates education into hourly wages. To understand why wages differ across people, one needs to know how education differs across people *and* the numerical relationship between wages and education, f. Because f is a function, it can be described by a formula. By far, the most popular is the *linear* function, whose formula is

$$Y = \beta_1 + \beta_2 X\,. \tag{2}$$

β_2 is the slope and β_1 the intercept. Technically, β_1 and β_2 are *population parameters*.[3] Parameters are numbers that characterize the numerical relationship between the dependent variable (the behavior to be explained) and the explanatory variable (the factor that explains behavior) for the entire statistical population.

In the earnings example,

$$Wage = \beta_1 + \beta_2 Education\,. \tag{3}$$

β_2, the slope, shows how the hourly wage changes when education changes:

$$\beta_2 = \frac{\Delta Wage}{\Delta Education}\,. \tag{4}$$

If education were to rise by one year, the hourly wage would rise by β_2. Because $Wage$ is measured in dollars per hour, β_2 is measured in dollars per hour as well. For example, β_2 might be a number like 1.05. If so, then an additional year of education would raise wages by $1.05/hour.

β_1, the intercept, tells how much someone with no education would earn per hour. Or, what someone with no brains and all brawn would get. In some sense, this is the base

[1] The terms "explained by," "accounted for by," and "a function of," may be interchanged with "determined by."
[2] The terms *independent variable* and *regressor* are used interchangeably with *explanatory variable*.
[3] They are also known as *coefficients*.

wage in the economy. For example, β_1 might be a number like 0.50, or 50 cents per hour. This is somewhat difficult to conceptualize, because virtually every worker in the American economy has had some formal education.

2. Random Variation

There are two ways that wages vary. First, *across* education categories, people with higher education have higher wages. Second, *within* education categories, people with the *same* level of education are paid *differently*. The linear relationship between wages and education

$$Wage = \beta_1 + \beta_2 Education \ , \tag{5}$$

accounts for the first type of variation, but not the second. For example, the relationship says all individuals with a high school diploma (i.e., $Education = 12$) earn

$$\beta_1 + \beta_2 ?12$$

dollars per hour. If $\beta_1 = 0.50$ and $\beta_2 = 1.05$, then the expected wage for someone with a high school diploma would be

$$\beta_1 + \beta_2 ?12 = 0.50 + 1.05 ?12 = 13.10 \ ,$$

or $13.10 per hour. In reality, however, some are paid more than others. Many factors could account for this second type of variation, such as differences in work experience, industry, occupation, unionization, gender, race, etc. In addition, wages might differ for purely random reasons.

To reflect randomness, an additional term, u, is added:

$$Wage = \beta_1 + \beta_2 Education + u \tag{6}$$

This new relationship is true for everyone in the population. As a notational convenience, statisticians use the subscript letter i to index each individual in the population.[4] Hence, the relationship often is written formally as

$$Wage_i = \beta_1 + \beta_2 Education_i + u_i \tag{7}$$

This equation is known as a statistical or econometric *model* of the determinants of wages. It is a complete description of the hourly wage for each individual in the population. It implies that the dependent variable is a function of the explanatory variable, the unknown population parameters, and randomness. u is referred to as the *disturbance term* or *error term*. Generically, this relationship can be written as

$$Y_i = \beta_1 + \beta_2 X_i + u_i \ . \tag{8}$$

Technically, u is a *random variable.* As such, it has an average (or expected) value and a variance. Two assumptions are made about u. First, u can be positive or negative, but its average value is zero. For a given education level, some individuals will have randomly higher wages and others randomly lower wages. Simply, think of it as

[4] If there are n individuals in the population, i is an integer number that runs from 1 to n.

good luck or bad luck. In layman's terms, a zero average says that even though there is some randomness in wages, this washes out in the population as a whole. On average, there is neither good nor bad luck in the population. Second, u has a variance that is constant across subgroups of the population. This says that the random variation in wages is the same regardless of whether you consider high school dropouts or Ph.D.'s.

One implication of our statistical model is that the wage can be broken into two parts. The first is the average wage for someone with that level of education:

$$\beta_1 + \beta_2 Education_i \, .$$

This is sometimes referred to as the *systematic* part of the hourly wage, because, quite obviously, it depends systematically on education! The second is a *random* component or luck:

$$u_i \, .$$

For a generic model, $Y_i = \beta_1 + \beta_2 X_i + u_i$,

$$\beta_1 + \beta_2 X_i$$

is the systematic part and

$$u_i$$

the random part of the dependent variable, respectively.

3. The Population Regression Line

The two parts of the hourly wage can be illustrated graphically as well. Figure 1 is a scatter plot of hourly wages and education for a population of workers. The hourly wage (dependent variable) is measured along the vertical axis and education (explanatory variable) along the horizontal axis. In Figure 2, the upward-sloping line labeled

$$Average\ \ Wage = \beta_1 + \beta_2 Education \tag{9}$$

is the *population regression line*.[5] Equation (9) is known as the *population regression equation*. It shows the average (or expected) value of the hourly wage for each value of education. For all workers in the population with a high school diploma, the average wage is

$$\beta_1 + \beta_2 \, ?12 \; .$$

However, any specific worker with a high school diploma is likely to have a different hourly wage. For example, for worker A, the hourly wage is greater than the average. In fact, the vertical distance between any point and the population regression line, e.g.,

$$Worker\ A\ 's\ Wage\ \text{-}\ Average\ Wage \; ,$$

[5] This is sometimes referred to as the *true regression line*.

is the value of the disturbance term, *u*, for that observation. That is, the disturbance term for worker A is

$$u_{WorkerA} = Worker\ A\text{'}s\ Wage\ -\ Average\ Wage\ . \tag{10}$$

This is shown in Figure 3. The disturbance term is the random part of the hourly wage. For some workers with high school diplomas, it is positive (Worker A) and for others, it is negative (Worker B).

4. The Sample Regression Line

In principle, if the researcher could obtain information on wages and education for every worker in the population, the values of β_1 and β_2 could be determined. Then the population regression line would tell everything needed about the simple relationship between wages and education. However, this is usually not the case because most statistical populations are so large that it is prohibitively costly to gather information for every unit. This means that the values of the parameters are unknown. Consequently, researchers rely on *random samples*. These samples are used to estimate the unknown values of the parameters β_1 and β_2 and make inferences about the population as a whole.

To reduce any confusion in what follows, clear notational distinctions are drawn between the population and the sample. All population parameters will be written as Greek letters, e.g., β_1 and β_2. All estimates based on a sample will be denoted by a carat or "hat," ^ , above the term. For example, the estimate of β_1 is denoted as $\hat{\beta}_1$. The estimate of β_2 is denoted as $\hat{\beta}_2$. In addition, all sample observations will be indexed by the letter j. j is an integer number that runs from 1 to *N*. All population observations will be indexed by the letter i.

Figure 4 is a scatter plot of hourly wages and education for a random sample of workers from the population shown in Figure 1. The line through the plot in Figure 5 is a *sample regression line*. It is labeled

$$Predicted\ \ Wage = \hat{\beta}_1 + \hat{\beta}_2 Education\ . \tag{10}$$

It has an intercept, $\hat{\beta}_1$, and a slope, $\hat{\beta}_2$ (note the hats!). The sample regression line shows the hourly wage one would predict using the sample if given a level of education and the choice of estimates, $\hat{\beta}_1$ and $\hat{\beta}_2$. For example, all sample observations with a high school diploma have a predicted wage of

$$Predicted\ \ Wage_{High\ School} = \hat{\beta}_1 + \hat{\beta}_2\ ?12\ .$$

The predicted value of a variable is denoted with a "hat." Thus, the predicted value of the hourly wage, $\hat{Wage}$, for someone with a high school diploma is written typically as

$$\hat{Wage}_{High\ \ School} = \hat{\beta}_1 + \hat{\beta}_2\ ?12\ .$$

Likewise, all sample observations with a Bachelor's degree have an predicted wage of

$$\hat{Wage}_{Bachelor's\ Degree} = \hat{\beta}_1 + \hat{\beta}_2 ?16 .$$

However, any specific sample observation with a high school diploma is likely to have an hourly wage different than the predicted hourly wage for their education level. In fact, the vertical distance between any point and the sample regression line is known as the *residual*, $\hat{u}$, for that observation. For worker A, the residual is

$$\hat{u}_{WorkerA} = Worker\ A's\ Wage - Predicted\ Wage\ . \tag{11}$$

Using the definition of the predicted wage,

$$\hat{u}_{WorkerA} = Worker\ A's\ Wage - \hat{\beta}_1 - \hat{\beta}_2 ?(Worker\ A's\ Education\ Level) \tag{12}$$

It is important to note that once the parameter estimates are chosen and a sample regression line is drawn, *each observation in the sample has a residual*. That is, for an arbitrary sample observation,

$$\hat{u}_j = Wage_j - \hat{\beta}_1 - \hat{\beta}_2 Education_j .$$

For a generic model, $Y_i = \beta_1 + \beta_2 X_i + u_i$, the residual for the jth sample observation is

$$\hat{u}_j = Y_j - \hat{Y}_j = Y_j - \hat{\beta}_1 - \hat{\beta}_2 X_j .$$

The residual can be interpreted in a number of ways, all complementary. First, it is the difference between the actual and average value of the dependent variable. Second, because it measures how the actual differs from the average, the residual is often thought of as an *error*.[6] Third, as the name suggests, the residual is the part of the dependent variable left unexplained by the explanatory variable. It is the leftover after the explanatory variable has done all of the explaining it can. In fact, each sample value of the dependent variable can be decomposed into two parts: one explained and the other unexplained. If

$$Wage_j = \hat{\beta}_1 + \hat{\beta}_2 Education_j + \hat{u}_j ,$$

the explained part is

$$\hat{\beta}_1 + \hat{\beta}_2 Education_j ,$$

and the unexplained part is

$$\hat{u}_j .$$

This is an extremely important interpretation. We will return to it later in detail.

Finally, it should be noted that there is a direct correspondence between the population and sample. Each piece of the population equation has an analog in the sample equation:

[6] In fact, the terms "residual" and "error" often are used interchangeably.

$$Population: \; Wage_i = \beta_1 + \beta_2 Education_i + u_i$$

$$) \qquad\quad) \quad) \quad\;) \qquad) $$

$$Sample: \qquad Wage_j = \hat{\beta}_1 + \hat{\beta}_2 Education_j + \hat{u}_j$$

The sample dependent variable corresponds to the population dependent variable. The sample explanatory variable corresponds to the population explanatory variable $\hat{\beta}_1$ is the sample estimate of the population parameter β_1. $\hat{\beta}_2$ is the sample estimate of the population parameter β_2. Finally, in a loose sense, $\hat{u}$, the residual, can be thought of as the sample analog to the population disturbance term u.

5. The Least Squares Method

An estimator is a mathematical rule or method for calculating estimates of unknown parameters. The idea of estimation is really just to use sample information on the dependent and explanatory variables to guess β_1 and β_2. This is probably best illustrated graphically. In Figure 2, the population regression line roughly went through the middle of the scatter. Therefore, to best estimate β_1 and β_2, it seems you would want to do the same in the sample: have a line run through the middle of the sample scatter. The line in Figure 5 is one such line. However, as shown in Figure 6, it should be apparent that there are many possible lines that could run through the middle of the scatter. Three are shown along with their respective values of $\hat{\beta}_1$ and $\hat{\beta}_2$. Each line has a slightly different slope and intercept, which means it is based on a different choice of numbers for $\hat{\beta}_1$ and $\hat{\beta}_2$, and each line generates a different set of residuals. The question is: Which line should be chosen? This is the same as saying: Which numbers should we choose for $\hat{\beta}_1$ and $\hat{\beta}_2$?

The least squares method of estimation says to choose the values of $\hat{\beta}_1$ and $\hat{\beta}_2$ that minimize the sum of the squared residuals. Recall that residuals can be interpreted as errors, i.e., how far the actual wage is from that predicted by the sample regression line (which is governed by the choice of $\hat{\beta}_1$ and $\hat{\beta}_2$). The square of a big error is much bigger than the square of a small error. So, the least squares method penalizes choices of $\hat{\beta}_1$ and $\hat{\beta}_2$ that generate big errors or residuals.

Using some calculus and a lot of algebra, the mathematical formula for the least squares estimators for β_1 and β_2 are

$$\hat{\beta}_2 = \frac{\left(\sum_{j=1}^{N} Y_j X_j - \frac{\left(\sum_{j=1}^{N} Y_j \right)\left(\sum_{j=1}^{N} X_j \right)}{N} \right)}{\left(\sum_{j=1}^{N} \left(X_j^2 \right) - \frac{\left(\sum_{j=1}^{N} X_j \right)^2}{N} \right)} . \tag{13}$$

and

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} , \tag{14}$$

where Y is a generic dependent variable and X a generic explanatory variable. $\bar{Y}$ is the sample mean of the dependent variable and $\bar{X}$ is the sample mean of the explanatory variable.

Least squares regression has a number of properties. First, (14) implies

$$, \tag{15}$$

which says the least squares sample regression line runs through the point with coordinates equal to the sample means of the dependent and explanatory variables, $(\bar{Y}, \bar{X})$. Second, the residuals,

$$\hat{u}_j = Y_j - \hat{\beta}_1 - \hat{\beta}_2 X_j \ , \tag{16}$$

always sum to zero:

$$\sum_{j=1}^{N} \hat{u}_j = 0 \ . \tag{17}$$

Of course, this implies the sample mean of the residuals is zero:

$$\frac{\sum_{j=1}^{N} \hat{u}_j}{N} = 0 \ . \tag{18}$$

Third, the residual and the explanatory variable are *uncorrelated*. *This is true for every least squares estimation (or regression)*. This must be true because the residuals are defined as that part of the dependent variable, Y, that is left over after the explanatory variable, X, and the intercept have done all they can to explain the dependent variable.

7. Interpreting the Least Squares Estimators

In the least squares formula for $\hat{\beta}_2$, the numerator and denominator contain sums of products of the sample values for the dependent and explanatory variables and the products of sums of these variables, respectively. While mathematically elegant, this depiction lacks intuition because it is not natural to think in terms of sums of products and products of sums. Luckily, with some algebra, (13) can be rewritten as

$$\hat{\beta}_2 = \frac{C\hat{o}v(Y,X)}{V\hat{a}r(X)} . \tag{19}$$

The numerator is the sample covariance between the dependent and explanatory variables. The denominator is the sample variance of the explanatory variable.

It is more intuitive to think in terms of covariances and variances. For example, the covariance is really just a measure of correlation between two variables. In (19), it measures how on average the dependent variable moves relative to its mean when the explanatory variable moves relative to its mean. That is, how the dependent and explanatory variables move together, or "co-vary". Because the value of the explanatory variable itself varies in the sample, least squares says to normalize how the dependent and explanatory variables co-vary relative to how the explanatory variable itself varies. This estimates β_2 !

For the wages and education example, $\hat{\beta}_2$ is

$$\hat{\beta}_2 = \frac{C\hat{o}v(Wage, Education)}{V\hat{a}r(Education)} . \tag{20}$$

This formula is so simple that you easily could calculate an estimate of β_2 by hand. The sample covariance between the hourly wage and education is 9.982. The sample variance of education is 6.807. Therefore, for our CPS sample,

$$\hat{\beta}_2 = \frac{9.982}{6.807} = 1.466 .$$

That is, the least squares estimate of β_2 is 1.466. Because

$$\beta_2 = \frac{\Delta Wage}{\Delta Education} ,$$

this estimate says that an additional year of education is associated with an additional \$1.47 in the hourly wage. Or, someone with a high school diploma that went to one year of college would see their hourly wage rise by \$1.47 by getting that additional year of schooling. Relative to the sample mean wage of \$14.77 per hour, this marginal effect represents a 10% increase in the hourly wage for an additional year of schooling.

Likewise, $\hat{\beta}_1$ can be calculated as

$$\hat{\beta}_1 = \overline{Wage} - \hat{\beta}_2 \overline{Education} .$$

$\overline{Wage}$ is the sample mean of the hourly wage, which is \$14.77 per hour. $\overline{Education}$ is the sample mean of the education level, which is 13.5 years of education. Therefore, for our CPS sample,

$$\hat{\beta}_1 = 14.77 - 1.47 \cdot 13.5 = -4.96 .$$

Because β_1 tells how much someone with no education would earn per hour, we would expect β_1 to be positive or, at a minimum, zero. But this estimate says that someone with no education would earn *negative* \$4.96 per hour. So, someone with no brains and all brawn would have to *pay an employer* 4.96 per hour to work! This is counterintuitive.

One should be suspicious of this estimate for a number of reasons. First, because there are no sample observations with no education, the estimate is an extrapolation. Second, the assumed linear relationship might not hold at lower education levels. For those with less than primary school education, the regression line may have a shallower slope than the one that fits the rest of the data. Finally, there may also be reporting (or measurement) error associated with the low values of education. Some of these individuals report only six years of education. They would be grade school drop-outs! Because it is difficult not to have attended primary and middle school these days, these individuals may have reported their education incorrectly during the CPS interview. For now, we will look the other way and just remind ourselves that -\$4.96 is just an estimate, and some estimates are better than others. Later, we will return to our discussion of $\hat{\beta}_1$ and show that the non-linearity between wages and education may be very important.

The complete regression results are shown in Table 4.1. There are 1003 persons in the sample. The coefficient of variation, R^2 , is 0.1708: about 17% of the variation in hourly wages is due to variation in years of education across workers. This is somewhat lower than might be expected, given the presumed importance of education. It means that about 83% of wage variation is due to other factors, possibly some that were measured by the CPS.[7]

A test of association between wages and education is performed using the *t*-test for the slope parameter, β_2 . Here, $t = 14.36$ with an observed level of significance, or *p*-value, of $p<.0001$. This means if there were no relationship between wages and education, the chances of obtaining a slope estimate this large (or larger) are less than 1 in 10,000. Because this would be a rare event, we conclude there is strong evidence of a positive relationship between education and wages. In fact, we say the relationship between wages and education is *statistically significant*.

The best single estimate for the true increase in average hourly wages for each additional year of education is \$1.47. But because this is based on a random sample, and different random samples would give different estimates, there is uncertainty associated with the point estimate. This uncertainty is reflected in the *standard error*, and, thus, in the *confidence interval*. An interval estimate, or a range of values where the true value could lie, is computed based on the underlying variability of wages within each education level and the size of the sample, and is called a *confidence interval*. Given the standard error of 0.10, the 95% confidence interval for the effect of an additional year of education on the wage is (\$1.27, \$1.67). The lower end of the interval is far enough from zero to

[7] The regression output also gives other information. For example, the root mean square error (RMSE) and the Analysis of Variance (ANOVA) table. These are often less useful for the practical interpretation of the analysis.

provide assurance that not only more education is associated with higher wages but that wages rise by about \$1.30 and \$1.70 per hour for every additional year.

The parameter estimates can be used to construct the sample regression line, which is shown in Figure 7. It gives the *predicted* or *fitted* wage at each education level. Figure 8 superimposes another line. This curvy line runs through the point of mean education for each single year of education.[8] There is a remarkable concordance between the fitted regression line and "mean wage line" at each education level. In fact, tracking mean wages is an important characteristic of regression analysis.

When mean wages show a linear trend, these two curves will be close, as between 9 and 18 years of education. Only at very low and high levels of education do the lines differ greatly. Indeed, wages may be higher than predicted at the extremes. One way to account for this is to modify the initial model to reflect possible non-linearity in the dependent variable and skewness in the distribution of disturbances for each education level. The most common method is to use a natural logarithm transformation of the wage.

8. Log Transformations

The plot of hourly wages against years of education (Figure 1) showed that within each year of education, wages are not symmetrically distributed. In fact, they were right-skewed, so that for the same education level, a few people have very high wages but most are lower than the average. Log transforming hourly wages results in a plot where wages are generally more symmetrically distributed within each education level.

To understand the consequences of the log transformation of wages, first consider the generic model

$$Y_i = \beta_1 + \beta_2 X_i + u_i \ .$$

Taking the derivative of both sides with respect to X yields

$$\beta_2 = \frac{dY}{dX} \ . \tag{21}$$

β_2 is interpreted as the change in the dependent variable for a change in the explanatory variable. This is true for any set of dependent and explanatory variables.

For the model

$$Wage_i = \beta_1 + \beta_2 Education_i + u_i \ , \tag{22}$$

$$\beta_2 = \frac{dWage}{dEducation} \ . \tag{23}$$

That is, a one unit increase in educational attainment (which is one additional year) would raise wages by β_2 dollars per hour. But, for a second model, with a log transformation of the dependent variable,

---

[8] Technically, between years the line is smoothed according to a cubic spline.

$$\ln(Wage_i) = \beta_1 + \beta_2 Education_i + u_i \ , \tag{24}$$

$$\beta_2 = \frac{d\ln(Wage)}{dEducation} \ , \tag{25}$$

because the dependent variable is $\ln(Wage)$. That is, an additional year of educational attainment would raise the natural logarithm of the hourly wage by β_2. This is an awkward interpretation because most people do not think in terms of natural logs. Fortunately, there is a more intuitive interpretation of β_2 here. First, apply the chain rule to (25):

$$\beta_2 = \frac{d\ln(Wage)}{dWage} \cdot \frac{dWage}{dEducation} \ . \tag{26}$$

For any variable z,

$$\frac{d\,ln(z)}{dz} = \frac{1}{z} \ . \tag{27}$$

Therefore, (26) can be rewritten as

$$\beta_2 = \frac{1}{Wage} \cdot \frac{dWage}{dEducation} \ , \tag{28}$$

or,

$$\beta_2 = \frac{\frac{dWage}{Wage}}{dEducation} \tag{29}$$

The numerator can be interpreted as the percentage change in wages, i.e.,

$$\frac{dWage}{Wage} = \%\Delta Wage \tag{30}$$

Therefore,

$$\beta_2 = \frac{\%\Delta Wage}{\Delta Education} \tag{31}$$

That is, an additional year of educational attainment would raise the hourly wage by β_2 percent. Economists refer to this as the *rate of return* to an investment in an additional year of education, or, the return to education.

INSERT THE $e^{\beta_2} \cong 1 + \beta_2$ DISCUSSION HERE.

Now, using the formula for the least squares estimator from (19),

$$\hat{\beta}_2 = \frac{C\hat{o}v(\ln(Wage), Education)}{V\hat{a}r(Education)} . \qquad (32)$$

The sample covariance between the log hourly wage and education is 0.635. The sample variance of education is 6.807. Therefore,

$$\hat{\beta}_2 = \frac{0.635}{6.807} = 0.093 .$$

That is, the least squares estimate of β_2 is 0.093. Because

$$\beta_2 = \frac{\%\Delta Wage}{\Delta Education} ,$$

this estimate says that an additional year of education is associated with a 9.3% increase in the hourly wage. So, the estimated return to a one-year investment in education yields a return of 9.3%.

The complete regression results are shown in Table 4.2. There are still 1003 persons in the sample. Now, the coefficient of variation, R^2, is 0.1620: about 16% of the variation in hourly wages is due to variation in years of education across workers. Much of the variation in the log wage is unexplained. Expressed as an hourly wage, $\hat{\beta}_1$ = 1.2599 implies that a worker with no education earns

$$e^{\hat{\beta}_1} = e^{1.2599} = 3.52 ,$$

or \$3.52 per hour. This is a much more sensible estimate of the base wage in the economy that -\$4.96.

A test of association between wages and education is performed using a *t*-test. Because the natural logarithm is a monotone function, if wages increase according to years of education, the same will be true for log wages. In this case, $t = 13.91$ ($p<.0001$). There is strong evidence of a positive relationship between education and log wages, and, therefore, also between education and wages. Often, the *t*-statistic and observed level of significance are not the same on the two scales, because the data may be more dispersed on one scale; for these data, they happen to be similar. For scales that are monotone transformations of each other, the direction of effect, either positive or negative, is the same, so that inference made on one scale can be applied to the other.

The parameter estimate for education is 0.0933 with a 95% confidence interval of (0.0801, 0.1064). These point and interval estimates can be converted to effects on the original scale by using the exponential function. Average hourly wages increase by a factor of exp(0.0933)=1.10 or 10% for each additional year of education. A 95% confidence interval for this figure is obtained by exponentiating the upper and lower bounds of the 95% confidence interval on the log scale: (exp(0.0801), exp(0.1064)) = (1.08, 1.11). This gives the best interval estimate for the increase in the hourly wages for every additional year of education based on the estimation using the log transformation. The lower bound is far enough from 1.00 (1.00=exp(0), or, no return) that we not only believe there is a positive effect of education, but this effect is on the order of 8 to 11% increase in wages for every additional year of schooling.

The effect on wages of four years of college or four years of high school can also be analyzed. Average wages increase by a factor of about exp(4*0.0933)= 1.45, with a 95% confidence interval of (exp(4*0.0801), exp(4*0.1064))= (1.38, 1.53) for every additional four years of education. This means that high school or college graduates' wages are an average of 45% (38%, 53%) higher than they would be without this education.

The parameter estimates can be used to construct the sample regression line, which is shown in Figure 9. It gives the *predicted* or *fitted* log wage at each education level. Figure 9 superimposes the curvy line that runs through the point of mean education for each single year of education.[9] Again, there is a remarkable concordance between the fitted regression line and "mean wage line" at each education level. At very low and high levels of education, the lines still differ.

Because most people do not think in terms of "log dollars," it is easier to reinterpret the estimates in dollars per hour. The sample regression line implied by the log-linear model, but expressed in terms of the hourly wage and education, is shown in Figure 10. It gives the *predicted* or *fitted* hourly wage at each education level. Its most important feature is that is curved. This is because a linear relationship between the log wage and education is a non-linear relationship between the wage and education. Indeed, the slope of this curve is steeper for higher levels of education. This is because a 9.3% increase in wages is a larger absolute jump when wages are $20 per hour than $10 per hour. For comparison, Figure 10 also superimposes the sample regression line from the linear model discussed above.

[9] Technically, between years the line is smoothed according to a cubic spline.