1 Population Regression

When we estimate things, our estimation is going to depend on whatever sample we happen to have obtained. That sample is usually not going to be a perfect representation of the population, so any given sample will differ from the population in random ways.

To illustrate, suppose you have a population of 100 people and you want to estimate their income. You take a sample of $n = 20$ randomly selected observations, someone else takes a sample of a different $n = 20$ randomly selected observations. Chances are you won't sample the exact same 20 people and hence your estimates will be a bit different. We need to account for that sampling variability.

In the context of regressions, we will look at a scatterplot of data for variables x and y. The **regression** line is just the line of best fit through the scatterplot of data. Like any line, it can be described in terms of a y-intercept and a slope, i.e. $y = mx + b$. Our terminology will use β_1 as the intercept coefficient and β_2 as the slope coefficient *for the line that best fits the entire population of data*; but since we only have a random sample, we instead use our sample data to derive respective estimates b_1 and b_2.

Our estimation method is called **ordinary least squares (OLS)**. Since we have a line of *best* fit, but not a line of *perfect* fit, the regression line will not lie exactly on top of all data points. Intuitively the line of best fit will, in some overall sense as described in a bit, be closest to the data points, which is the objective of the OLS estimation procedure. In other words, the line of best fit is the one that minimizes how far off it is from the data. To use OLS estimates, we will rely on a number of assumptions in order to make sure our estimates for each β have nice properties.

2 Unbiased Estimators

It might help to first refer to Figure 1 below to get a visual feel for what's going on.

2.1 Assumption OLS1: Linear True Population Model

Again, a regression is just the line of *best* fit — it is not the line of *perfect* fit. When we talk about a specific data point i, we assume that the true population model has linear form

$$y_i = \beta_1 + \beta_2 x_i + u_i. \tag{1}$$

What this says is we use the line $\beta_1 + \beta_2 x_i$ to best "predict" what y_i should be for a given value of x_i; but since the regression line doesn't perfectly capture all data points, the prediction will be off by u_i. Accordingly, u_i is called the **error term**.

Violations of OLS1. Note that the term *linear* here refers to the fact that the model is *linear in parameters*, that is, linear in β_1 and β_2; there is no restriction placed upon the functional forms of the variables, however. For example, the model is still linear if we have something of the form

$$\log(y_i) = \beta_1 + \beta_2 \log(x_i) + u_i,$$

because we can simply define $v \equiv \log(y_i)$ and $w \equiv \log(x_i)$ and express the model as

$$v_i = \beta_1 + \beta_2 w_i + u_i.$$

On the other hand, this assumption would be violated if the true population model actually has form of, say,

$$y_i = \beta_1 + x_i^{\beta_2} + u_i,$$

because β_2 enters the model exponentially instead of linearly.

2.2 Assumption OLS2: Zero Conditional Mean

The zero conditional mean assumption states that

$$E[u_i|x_i] = 0 \quad \text{for all } i. \tag{2}$$

Remember, u_i represents how wrong the line of best fit is for data point (x_i, y_i). The zero conditional mean assumption means that, given x_i, we expect the line of best fit to not be wrong *on average*.

It's important because we'd like an answer for the question, "what do I expect y to be, given any value of x?" Consider a specific $x = x^*$, where x^* is just some number for

which we want to predict y.[1] This allows us to write

$$\begin{aligned} E[y|x = x^*] &= E[\beta_1|x = x^*] + E[\beta_2 x|x = x^*] + E[u|x = x^*] \\ &= \beta_1 + \beta_2 x^*. \end{aligned}$$

This is true because β_1 and β_2 are just numbers — there is nothing random about them — so we, uh, expect them to be themselves, regardless of what x is. And because of our zero conditional mean assumption, the error term drops out. Thus, the regression line is what we expect y to be for a given value of x.

Violations of OLS2. For more intuition, suppose $E[u_i|x_i] \neq 0$. Then when we plug in some data point x_i, we expect the model's prediction of y to be wrong on average. It makes for a pretty lousy model when we expect it to be wrong, on average.

OLS assumption 2 is equivalent to saying that the error term u is uncorrelated with the regressor x. The idea is that we use x to explain y, and u is all of the other stuff that explains y that we haven't included in our model. If changing x has no effect on how u in turn affects y, then OLS assumption 2 is satisfied.

To illustrate a failure of OLS assumption 2, consider the regression

$$wage = \beta_1 + \beta_2 education + u.$$

I can think of other things besides years of education that might affect someone's wage, e.g. their IQ. Here, IQ is part of u because it is "other stuff" besides education that explains *wage*. But IQ is likely correlated with years education. So when you consider different *education*, you're also implicitly considering different IQ, and hence you do not get the direct relationship between *education* and *wage*. In other words, we are not holding IQ constant, and that gives us a biased estimate of how *education* relates to *wage*.

2.3 Summary of OLS1-2

So to summarize the implications of the first two OLS assumptions:

a. The actual value y_i is given by $y_i = \beta_1 + \beta_2 x_i + u_i$.

b. The regression line is what we expect y_i to be, given x_i: $E[y_i|x_i] = \beta_1 + \beta_2 x_i$.

c. Hence the error term is given by $u_i = y_i - E[y_i|x_i]$.

[1] For example, we might want to predict how many cavities a person has (y) if they eat 200 grams of sugar per day ($x = 200$).

OLS assumptions 1 and 2 imply that OLS estimates (explained soon) b_1 and b_2 of β_1 and β_2 are *unbiased*; in other words, we expect the estimates to be their true values. In maths, $E[b_1] = \beta_1$ and $E[b_2] = \beta_2$.

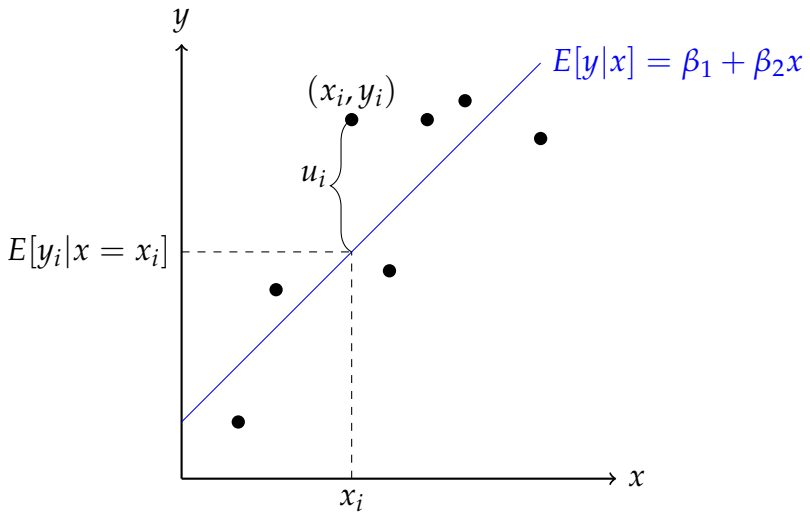

FIGURE 1: Suppose the dots here capture the entire population of data. Pick some arbitrary data point (x_i, y_i). The regression line tells us $E[y_i|x = x_i]$, that is, what value we expect y_i to be for independent variable x_i. This is the **conditional mean** of y_i given x_i. But the regression line is a line of *best* fit, not a line *perfect* fit, so the actual value of y_i will in general be different than what we expect it to be based on the regression line. The difference between what y_i actually is and what we expect y_i to be based on the regression, $y_i - \beta_1 - \beta_2 x_i$, is the error term, u_i.

3 BLUE

We can throw down two more assumptions to make analysis even nicer.

3.1 Assumption OLS3: Homoskedasticity

The variation of u_i given x_i is the same number σ_u^2 for any x_i. In math,

$$\text{Var}(u_i|x_i) = \sigma_u^2 \quad \text{for all } i. \tag{3}$$

This condition is illustrated in Figure 2.

Violations of OLS3. OLS assumption 3 fails often in practice, but can easily be accommodated by using **heteroskedasticity-robust standard errors**. Mathematically it is beyond the scope of this course, but it is very easy to implement in Stata using the command

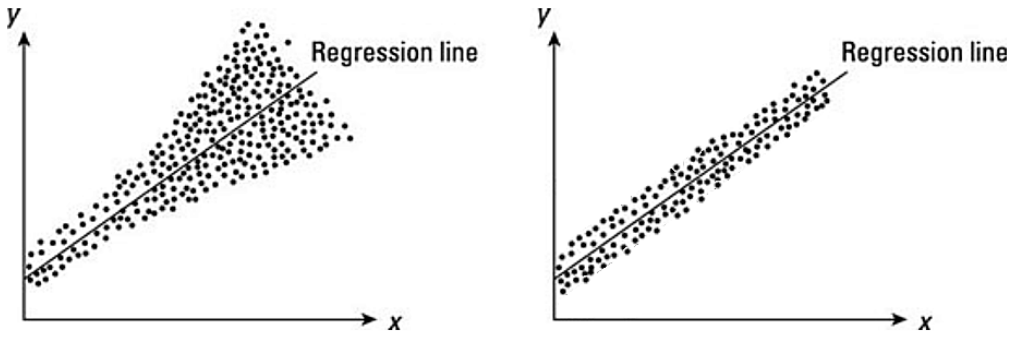

FIGURE 2: The figure on the left is an example of heteroskedasticity; the right an example of homoskedasticity. The left is heteroskedastic because the variation of errors around the regression line gets bigger as x increases.

`regress y x, vce(robust)`. (So you should know that heteroskedasticity is a thing that we should be concerned about.)

3.2 Assumption OLS4: Independent Errors

Errors for different observations are statistically independent, that is,

$$u_i \perp u_j \quad \text{whenever } i \neq j. \tag{4}$$

Violations of OLS4. When dealing with time series data, errors are often correlated over time. For instance, a positive error for GDP data indicates above-average GDP; but if GDP is above-average in one period, there is a good chance it will be above-average next period as well. Which is to say, a positive error term this quarter predicts a positive error term in the next quarter (as well as the previous quarter), so error terms are correlated.

3.3 Summary of OLS1-4

Adding OLS assumptions 3 and 4 allows us to say that the variation of y given x is also constant, and specifically, $\text{Var}(y|x) = \sigma_u^2$. OLS assumptions 1-4 imply also imply that estimates are **consistent**, provided the variances of the estimates go to zero as $n \to \infty$. Put somewhat crudely, this means that our estimates get arbitrarily close (in probability) to their true population values as the sample size increases. Basically the law of large numbers again. In math, we write $b \xrightarrow{p} \beta$.

We can go even further, however. Under OLS assumptions 1-4, the estimates are said to be **BLUE**, which stands for

- **B**est (estimates have the smallest standard errors...)
- **L**inear (among linear models...)
- **U**nbiased (that give unbiased...)
- **E**stimator (um, estimates.)

4 BUE

We can make a fifth assumption for one more nice result, although we're pushing it with this assumption and, frankly, it's not a particularly important assumption for this course.

Assumption OLS5: Normally Distributed Errors. Error terms have normal distribution with some variance σ^2,

$$u_i \sim \mathcal{N}\left(0, \sigma^2\right). \tag{5}$$

This allows us to say that OLS estimates are **BUE**, which means that they have the smallest standard errors among unbiased models, even when compared to nonlinear models. Also note that this is an essential condition if we want to do inference on small sample sizes because it allows us to say that it is exactly true that

$$\frac{b_2 - \beta_2}{\text{se}(b_2)} \sim T(n-2).$$

5 Recapping the OLS Assumptions

We have five OLS assumptions that give the following implications:

- OLS1 (linear model) and OLS2 (zero conditional mean) imply unbiased OLS estimates and $E[y|x] = \beta_1 + \beta_2 x$.
- Adding OLS3 (homoskedasticity) and OLS4 (independent errors) imply that $\text{Var}(y|x) = \sigma_u^2$ is constant, and that the estimate b_i for each β_i is consistent.
- OLS assumptions 1-4 therefore imply that OLS estimates of β are BLUE.
- Adding OLS5 (normality of errors) implies that OLS estimates of β are BUE and

$$\frac{b_2 - \beta_2}{\text{se}(b_2)} \sim T(n-2) \text{ exactly.}$$

We will weaken many of these assumptions as we go further into the course, but it's almost always best to start with the easiest result and then break it down from there.

6 OLS Estimation of a Regression

Again, b_1 is the estimate of β_1 and b_2 the estimate of β_2. Intuitively, we want a model that makes the fewest mistakes possible with the data. We quantify "fewest" by considering the difference between the actual values y_i and the **fitted values** as predicted by the model, given by $\widehat{y}_i = b_1 + b_2 x_i$; this is referred to as the **residual**, denoted e_i, defined as

$$e_i \equiv y_i - \widehat{y}_i = y_i - [b_1 + b_2 x_i]. \tag{6}$$

Think of the residual as capturing how wrong the estimated line of best fit is. Minimizing residuals overall should like a good idea. To summarize,

a. The actual value y_i is given by $y_i = b_1 + b_2 x_i + e_i$.

b. The regression line is what we expect y_i to be, given x_i: $\widehat{y}_i = b_1 + b_2 x_i$.

c. Hence the residual is given by $e_i = y_i - \widehat{y}_i$.

Hopefully you've noticed that residual e_i is like the sample analogue of the error u_i. Indeed, in many places e_i written as $\widehat{u}_i$ instead, as the hat notation often refers to estimates.

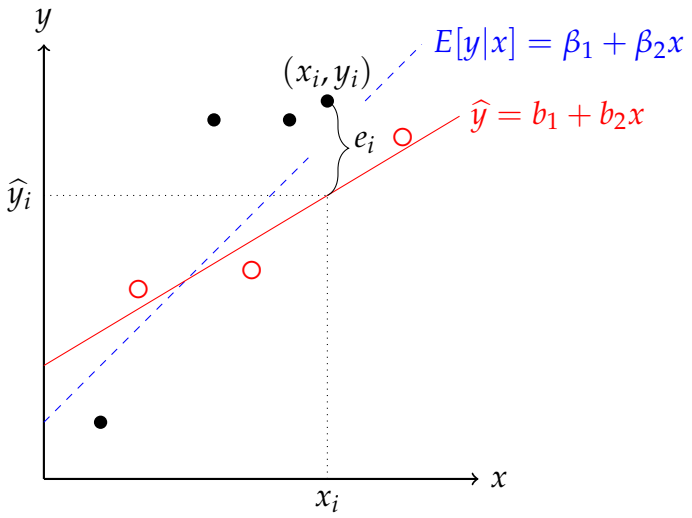

FIGURE 3: Suppose our sample consists of only the hollow red dots. Thus the estimated regression line (the solid red line) is different than the true population regression line (dashed in blue). For data point x_i, it gives us a prediction for y_i, i.e. the fitted value $\widehat{y}_i$. The fitted value will not in general be exactly the true value y_i, and the difference between the true value and the fitted value is the residual, $e_i = y_i - \widehat{y}_i$. This example illustrates a positive residual, $e_i > 0$.

We square each residual to ensure that it's positive, then we add the squared terms all up: this is the **residual sum of squares (RSS)**. We want the estimates that *minimize the residual sum of squares.* In mathspeak, we want to solve

$$(b_1, b_2) = \arg\min_{b_1,b_2} \sum_{i=1}^{n} (y_i - [b_1 + b_2 x_i])^2 = \arg\min_{b_1,b_2} \sum_{i=1}^{n} e_i^2. \tag{7}$$

The solution to this is the **ordinary least squares (OLS)** estimation for a linear regression. I omit the details, but explicitly solving OLS (using calculus to find critical points) gives formulas

$$b_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} = r_{xy} \times \frac{s_y}{s_x}, \tag{8}$$

$$b_1 = \bar{y} - b_2 \bar{x}, \tag{9}$$

where s_{xy} is the **sample covariance** of x and y, defined by

$$s_{xy} \equiv \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}), \tag{10}$$

and r_{xy} is the **sample correlation coefficient** of x and y, defined by

$$r_{xy} \equiv \frac{s_{xy}}{s_x s_y}. \tag{11}$$

I give several different expressions for b_2 because, depending on how a problem is worded, one expression might be more applicable than the others. Note that the first expression is the one given on the exam formula sheet. The last one is useful because it shows the relationship between the slope and the correlation coefficient, specifically, you adjust the correlation coefficient by the ratio of standard deviations.

Again, under OLS assumptions 1 and 2, the estimates will be unbiased: $E[b_1] = \beta_1$ and $E[b_2] = \beta_2$. That said, they will be different in generality than their population analogues because, well, they're estimates. Hence our estimated regression line will be more or less different than the population regression line, depending on how closely our sample reflects the population. This is illustrated in Figure 3.

Furthermore, OLS assumptions 3 and 4 imply that the variance of the slope estimate

b_2 will be

$$\text{Var}(b_2) = \frac{\sigma_u^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \equiv \sigma_{b_2}^2. \tag{12}$$

7 Explained and Unexplained Variation

To reiterate, we define the **residual sum of squares** to be

$$\text{RSS} \equiv \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2. \tag{13}$$

This is the overall measure of how far off the estimated regression line is relative to the data; each residual is squared so that the "mistakes" are positive. You can think of this as being the variation of data around $\bar{y}$ that cannot be explained by x.

Dividing RSS by $n-2$ (because we are estimating two parameters, one for each β) and taking the square root gives the **standard error of the regression**,

$$s_e \equiv \sqrt{\frac{\text{RSS}}{n-2}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}, \tag{14}$$

which is the sample analogue of σ_u as used in OLS3 and OLS4. This is sometimes called the **standard error of the residuals** or **root mean squared error (RMSE)**.

Take my word for it that $\bar{e} = 0$, that is, the mean of residuals is zero.[2] Then if we were to write down the standard deviation of residuals, taking into account that we now have $n-2$ degrees of freedom, we would write

$$\begin{aligned}
\text{SD}(e) &= \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(e_i - \bar{e})^2} \\
&= \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(e_i)^2} \\
&= \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2},
\end{aligned}$$

which is precisely s_e as given above. The takeaway is that the standard error of the re-

[2]This was given in lecture as an optional exercise; I leave it as such. But the key is to take the derivative of equation (7) with respect to b_1 and set it equal to zero, which must be the case from minimization of OLS (again think back to calculus and critical points).

gression is really just the standard deviation of the residual.

On the other hand, the variation of data around $\bar{y}$ that can be explained by x is the **explained sum of squares**,

$$\text{ESS} \equiv \sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2. \tag{15}$$

Finally, the total variation of data around $\bar{y}$ is given by the **total sum of squares**,

$$\text{TSS} \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2. \tag{16}$$

Based on the intuition it should not be surprising (not difficult to show either) that

$$\text{TSS} = \text{ESS} + \text{RSS}. \tag{17}$$

Total variation is explained variation plus unexplained variation. Great.

The proportion of explained variation around $\bar{y}$ is called the **R-squared** or **coefficient of determination**, defined as

$$R^2 \equiv \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}. \tag{18}$$

If R^2 is high, then x explains a lot about what's going on with y; if R^2 is low, then it doesn't. There is no cutoff for what should be considered "high" or "low," however. Note that R^2 also equals the squared correlation between y and x, that is, $R^2 = r_{xy}^2$. Also note that R^2 is only valid if the regression includes the intercept.

8 Estimator Properties and Inference

We are primarily interested in β_2 because it captures the relationship between x and y. Under OLS assumptions 1-4, our slope estimator b_2 has expected value of β_2 because it is unbiased; and it also has variance $\sigma_{b_2}^2$. Thus we can write

$$b_2 \sim (\beta_2, \sigma_{b_2}^2). \tag{19}$$

For sufficiently large sample size (greater than 30), the z-score is approximately standard normal, that is,

$$z \equiv \frac{b_2 - \beta_2}{\sigma_{b_2}},$$

which is drawn from a $\mathcal{N}(0,1)$ distribution, approximately.

But we don't actually know what σ_{b_2} because is it is a function of σ_u, which is a unknown population parameter. So instead we must use the sample estimate of σ_u, given earlier as s_e. This then allows us to conclude that the sample standard error of b_2 is

$$\text{se}(b_2) = \frac{s_e}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{20}$$

So under OLS assumptions 1-4, for large enough sample size (which does *not* have a clear cut rule-of-thumb in this case), we appeal to the central limit theorem and conclude that

$$t \equiv \frac{b_2 - \beta_2}{\text{se}(b_2)} \tag{21}$$

is drawn from a $T(n-2)$ distribution, where the distribution is usually approximate. If we add an additional assumption that the error terms are normally distributed (OLS5), then we can say that t is drawn from an exact $T(n-2)$ distribution.

Hence for a hypothesis test with null-hypothesized value β_2^*, you would calculate the t-statistic

$$t = \frac{b_2 - \beta_2^*}{\text{se}(b_2)}$$

and would perform inference about β_2 under the assumption that t was drawn from an approximate $T(n-2)$ distribution. Furthermore, for a two-sided 95 percent confidence intervals, for example, you'd use the formula

$$[b_2 \pm t_{n-2,0.025} \times \text{se}(b_2)], \tag{22}$$

which should look similar (and indeed is analogous) to the confidence intervals of $\bar{x}$.

9 Regression by Hand

Okay, so that's a lot to take in. At this point you should look at the formula sheet on one of the practice exams, because chances are you'll be relying on it when exam time comes

around. To help you become familiar, I provide an example. Consider the following data:

$$(x_1, y_1) = (0, 2),$$
$$(x_2, y_2) = (3, 3),$$
$$(x_3, y_3) = (3, 4).$$

Step 1: Estimate Regression Coefficients. One approach is to use a table to methodically deal with a lot of different parts. For example, after calculating the sample means

$$\bar{x} = \frac{1}{3}[0 + 3 + 3] = 2,$$
$$\bar{y} = \frac{1}{3}[2 + 3 + 4] = 3,$$

we can fill out the following table:

i	x_i	$\bar{x}$	$x_i - \bar{x}$	y_i	$\bar{y}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	0	2	-2	2	3	-1	2	4	1
2	3	2	1	3	3	0	0	1	0
3	3	2	1	4	3	1	1	1	1
$\sum$	6	-	-	9	-	-	3	6	2

Then use the formula given on the formula sheet to get

$$b_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{3}{6} = 0.5,$$
$$b_1 = \bar{y} - b_2\bar{x} = 3 - 0.5(2) = 2.$$

Thus our estimated regression is $y_i = 2 + 0.5x_i + e_i$ with fitted values $\widehat{y}_i = 2 + 0.5x_i$.

Equivalently, you can use sample variances and covariance,

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{6}{2} = 3,$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{2}{2} = 1,$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{3}{2} = 1.5,$$

to get

$$b_2 = \frac{s_{xy}}{s_x^2} = \frac{1.5}{3} = 0.5.$$

Or use correlation coefficient

$$r_{xy} = \frac{1.5}{\sqrt{3}\sqrt{1}} \approx 0.866$$

to calculate

$$b_2 = r_{xy} \times \frac{s_y}{s_x} = 0.866 \times \frac{1}{\sqrt{3}} = 0.5.$$

Depending on what information is given in a question, one approach will generally be faster than the others, and may even be the only tenable one.

Step 2: Calculate Residuals. We can find the standard error of the regression by finding the fitted values, i.e. by plugging each x_i into the regression formula and finding the corresponding $\widehat{y}_i$. Doing so gives

$$\widehat{y}_1 = 2 + 0.5(0) = 2,$$

$$\widehat{y}_2 = 2 + 0.5(3) = 3.5,$$

$$\widehat{y}_3 = 2 + 0.5(3) = 3.5.$$

Note that you can easily calculate fitted values in Stata with command `predict yhat` after having run the regression, where the variable `yhat` will store the fitted values.

The residuals are the difference between the actual y_i and what the regression line

expects y_i to be based on x_i, which in our case are

$$e_1 = y_1 - \widehat{y}_1 = 2 - 2 = 0,$$
$$e_2 = y_2 - \widehat{y}_2 = 3 - 3.5 = -0.5,$$
$$e_3 = y_3 - \widehat{y}_3 = 4 - 3.5 = 0.5.$$

You can easily calculate the residuals in Stata by entering command `predict e, resid` after having run the regression, where the variable `e` will store the residuals.

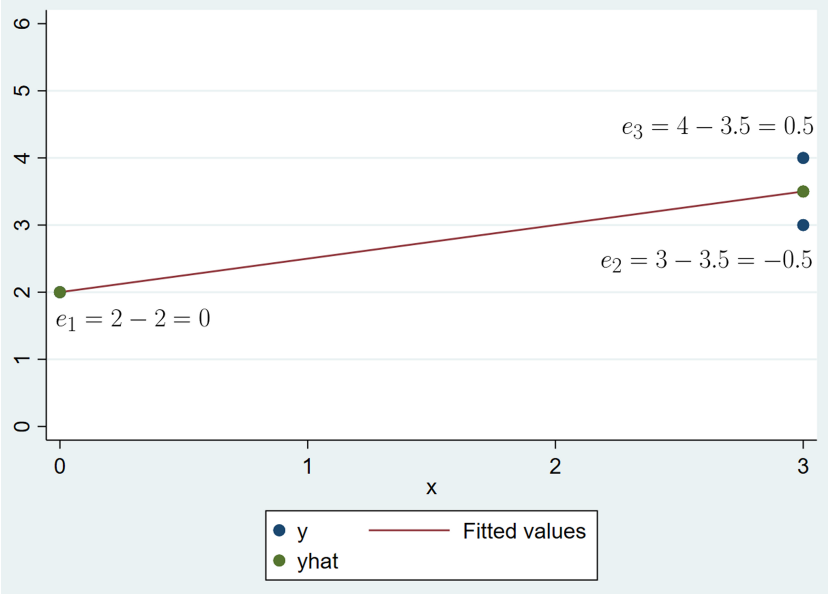

FIGURE 4: The first residual is zero because the regression line falls right on top of the actual point. The second residual is negative because $y_2 < \widehat{y}_2$; and the third residual is positive because $y_3 > \widehat{y}_3$.

Step 3: Calculate Standard Error of Residual. The residual sum of squares (RSS), uh, squares each residual and sums them up, so

$$\text{RSS} = (0)^2 + (-0.5)^2 + (0.5)^2 = 0.5.$$

Now we can find the standard error of the residuals using formula

$$s_e \equiv \sqrt{\frac{\text{RSS}}{n-2}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} = \sqrt{\frac{0.5}{3-2}} = 0.707.$$

Step 4: Calculate Standard Error of Slope Coefficient. The slope coefficient has standard error

$$\text{se}(b_2) = \frac{s_e}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

So we gotta figure out the denominator. Not a big deal, it's pretty much just the calculation for the standard deviation of x but without the division. And in fact we already know that the sum is 6 from the table above. Therefore the standard error of b_2 is

$$\text{se}(b_2) = \frac{0.707}{\sqrt{6}} \approx 0.289.$$

Step 5: Verify in Stata. You can easily verify this all by inputting the data into Stata using the `input` command and regressing y on x with command `reg y x`. Doing so yields the output as seen in Figure 5 below.

```
      Source |       SS           df       MS      Number of obs   =         3
-------------+----------------------------------   F(1, 1)         =      3.00
       Model |         1.5         1         1.5   Prob > F        =    0.3333
    Residual |          .5         1          .5   R-squared       =    0.7500
-------------+----------------------------------   Adj R-squared   =    0.5000
       Total |           2         2           1   Root MSE        =    .70711

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |          .5   .2886751     1.73   0.333    -3.167965    4.167965
       _cons |          2   .7071068     2.83   0.216    -6.984644    10.98464
------------------------------------------------------------------------------
```

FIGURE 5: Stata output for this example. Familiarize yourself by connecting these numbers with the ones just derived. The t-statistic and p-value are for $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$. In other words, Stata by default tests whether x has non-zero explanatory power for y.