Some important properties of summations
ECON 30331/Evans

Definition: The summation sign (Σ) adds up a series of numbers

Suppose there is a sample with n observations and two variables (x_i and y_i). Then

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 \ldots + x_n$$

We can represent sample means with summations:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Throughout the semester when I write at the board, I will shorten the notation some and write $\sum_{i=1}^{n} x_i$ as simply $\sum_i x_i$

Three important properties of summations:

Result (1): $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$. The sum of deviations from means equals zero.

Proof: $\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x} = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - n\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]$

$$= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

Result (2): $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$

Proof: $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i(y_i - \bar{y}) - \sum_{i=1}^{n}\bar{x}(y_i - \bar{y})$

Because $\bar{x}$ is a constant, it can be moved outside the summation sign in the final term above

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i(y_i - \bar{y}) - \bar{x}\sum_{i=1}^{n}(y_i - \bar{y})$$

Given the results from above (summation of deviation from means equal zero), $\bar{x}\sum_{i=1}^{n}(y_i - \bar{y}) = 0$ so

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i(y_i - \bar{y})$$

Following the same logic, we can easily establish that

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$$

Result (3): $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i(x_i - \bar{x})$

Proof: This is the same proof as above. Expand the terms on the right hand side of the equality

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^{n} x_i(x_i - \bar{x}) - \sum_{i=1}^{n}\bar{x}(x_i - \bar{x})$$

In the final term on the right, note that because $\bar{x}$ is a constant, you can take it outside the summation, and

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i(x_i - \bar{x}) - \bar{x}\sum_{i=1}^{n}(x_i - \bar{x})$$

And given result 1 above, $\bar{x}\sum_{i=1}^{n}(x_1 - \bar{x}) = 0$, so

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i(x_i - \bar{x})$$

Deriving the OLS estimates for the Bivariate Regression Model

Model: $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$

The residuals (ε_i) are unobserved, but for candidate values of β_0 and β_1, we can obtain an estimate of the residual.

Estimated residual: $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - x_i\hat{\beta}_1$

Objective is to minimize sum of squared residuals:

$$SSR = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1 \right)^2$$

First order conditions (FOCs):

$$(1)\ \partial SSR / \partial \hat{\beta}_0 = -2\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1 \right) = 0$$

$$(2)\ \partial SSR / \partial \hat{\beta}_1 = -2\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1 \right) x_i = 0$$

Use FOCs to obtain estimate for $\hat{\beta}_0$ and $\hat{\beta}_1$

The estimate for $\hat{\beta}_0$

Working with condition (1), multiply both sides by -1/2

$$(1a)\ \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1 \right) = 0$$

Then divide by n and expand all terms

$$(1b)\ \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_0 - \frac{1}{n}\sum_{i=1}^{n} x_i\hat{\beta}_1 = 0$$

The first term is $\bar{y}$, the second is $\frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_0 = \hat{\beta}_0$ and the third is

$\frac{1}{n}\sum_{i=1}^{n} x_i\hat{\beta}_1 = \frac{\hat{\beta}_1}{n}\sum_{i=1}^{n} x_i = \bar{x}\hat{\beta}_1$ and therefore, we can re-write (1b) as

$$(1c)\ \bar{y} - \hat{\beta}_1 - \bar{x}\hat{\beta}_1 = 0$$

Which means that

$$(1d)\quad \hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1$$

The estimate for $\hat{\beta}_1$

Working with condition (2), multiple both sides by -1/2

$$(2a)\quad \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1)x_i = 0$$

Substitute $\bar{y}$ - $\bar{x}$ $\hat{\beta}_1$ for $\hat{\beta}_0$ (from condition 1b)

$$(2b)\quad \sum_{i=1}^{n}(y_i - (\bar{y} - \bar{x}\hat{\beta}) - x_i\hat{\beta}_1)x_i = 0$$

Collect like terms

$$(2c)\quad \sum_{i=1}^{n}((y_i - \bar{y}) - (x_i - \bar{x})\hat{\beta}_1)x_i = 0$$

Expand the terms in the summation and complete the square, and because $\hat{\beta}_1$ is a constant, you can bring it outside the summation

$$(2d)\quad \sum_{i=1}^{n}(y_i - \bar{y})x_i - \hat{\beta}_1\sum_{i=1}^{n}\ (x_i - \bar{x})x_i = 0$$

Recognize two facts:

$$\sum_{i=1}^{n}(y_i - \bar{y})x_i = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$$

$$\sum_{i=1}^{n}(x_i - \bar{x})x_i = \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Substitute these values into (2d)

$$(2e)\quad \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) - \hat{\beta}_1\sum_{i=1}^{n}\ (x_i - \bar{x})^2 = 0$$

Bringing the second term to the right hand side

$$(2f)\quad \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \hat{\beta}_1\sum_{i=1}^{n}\ (x_i - \bar{x})^2$$

Then solve for $\hat{\beta}_1$

$$(2g)\qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Some useful properties of OLS estimates:

1. From (1c) above, note that $\bar{y} = \hat{\beta}_1 + \bar{x}\hat{\beta}_1$. The OLS models fits means of X through the means of y. OLS is sometimes referred to as a mean regression.

2. From (1a) above, note that $\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1\right) = 0$ and note further that $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - x_i\hat{\beta}_1$. Therefore $\sum_{i=1}^{n}\left(\hat{\varepsilon}_i\right) = 0$ which indicates that the sample mean of $\hat{\varepsilon}$ is equal to zero, or $\bar{\hat{\varepsilon}} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varepsilon}_i\right) = 0$

3. From (2a) above, recall that $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - x_i\hat{\beta}_1$ so (2a) can be written as $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - x_i\hat{\beta}_1)x_i = 0 = \sum_{i=1}^{n}\hat{\varepsilon}_i x_i$. Therefore, by construction, the optimal choices of $\hat{\beta}_0$ and $\hat{\beta}_1$ are such that x_i and $\hat{\varepsilon}_i$ are uncorrelated.

4. Looking at the OLS estimate in (2g), divide the numerator and denominator by (n-1)

$$(2h) \qquad \hat{\beta}_1 = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Notice that the numerator in (2h) is $\hat{\sigma}_{xy}$ and then denominator is $\hat{\sigma}_x^2$. Recognize also that $\hat{\rho}_{xy} = \hat{\sigma}_{xy}/(\hat{\sigma}_x\hat{\sigma}_y)$, so

$$\hat{\beta}_1 = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2} = \frac{\hat{\sigma}_{xy}\hat{\sigma}_y}{\hat{\sigma}_x\hat{\sigma}_x\hat{\sigma}_y} = \left(\frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x\hat{\sigma}_y}\right)\left(\frac{\hat{\sigma}_y}{\hat{\sigma}_x}\right) = \frac{\hat{\rho}_{xy}\hat{\sigma}_y}{\hat{\sigma}_x}$$

If one knows the variances and correlations coefficients, one can easily estimate the OLS value for $\hat{\beta}_1$

Deriving the R^2

Given the basic regression model: $y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$

Predicted outcome: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Estimated residual: $\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - x_i\hat{\beta}_1$

By construction: (1) $y_i = \hat{y}_i + \hat{\varepsilon}_i$

Take the average of equation (1) over all observations, then

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i + \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i = \bar{\hat{y}} + \bar{\hat{\varepsilon}}$$

Remember that the sample average of $\bar{\hat{\varepsilon}}$ is zero, so $\bar{y} = \bar{\hat{y}}$ (the sample mean of y equals the sample mean of predicted y).

The total *variation* in y, or the Sum of Squared Total (SST) is defined as

$$(2) \quad SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

This is nothing more than a statement about how much movement there is in y in your sample. Noting that $y_i = \hat{y}_i + \hat{\varepsilon}_i$ and $\bar{y} = \bar{\hat{y}}$, substitute these values into SST and complete the square

$$(4) \quad SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i + \hat{\varepsilon}_i - \bar{\hat{y}})^2 = \sum_{i=1}^{n}[(\hat{y}_i - \bar{\hat{y}})^2 + \hat{\varepsilon}_i^2 + 2\hat{\varepsilon}_i(\hat{y}_i - \bar{\hat{y}}_i)]$$

$$= \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 + \sum_{i=1}^{n}\hat{\varepsilon}_i^2 + 2\sum_{i=1}^{n}\hat{\varepsilon}_i(\hat{y}_i - \bar{\hat{y}})$$

Focus on the third term in the equality. Note a few things. First, since $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ and $\bar{\hat{y}} = \bar{y}$ and $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}$, then it is easy to show that $(\hat{y}_i - \bar{\hat{y}}) = \hat{\beta}_1(x_i - \bar{x})$. Substitute this value into the third term

$$(5) \quad 2\sum_{i=1}^{n}\hat{\varepsilon}_i(\hat{y}_i - \bar{\hat{y}}) = 2\sum_{i=1}^{n}\hat{\varepsilon}_i\hat{\beta}_1(x_i - \bar{x}) = 2\hat{\beta}_1\sum_{i=1}^{n}\hat{\varepsilon}_i(x_i - \bar{x})$$

In equation (5), we can take $\hat{\beta}_1$ outside the summation because it is the same value over all i. Look at the notes for "Deriving the OLS Estimates for the Bivariate Regression Model". On the final page, we note some useful properties of the OLS estimates.

Condition 3 states that by construction, $\sum_{i=1}^{n}\hat{\varepsilon}_i x_i = 0$ which means that equation (5) above is by construction, equal to zero. Therefore, equation (4) reduces to

$$(6) \qquad SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 + \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

The SST or the total variation in y has two separate parts. The first is

$$(7) \quad SSM = \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2$$

Where SSM is defined as the sum of squared model. This is a measure of the variation in the predicted value in Y.

The final term in equation (6) should look very familiar; it is none other than the objective function or, the sum of squared residuals (SSR).

$$(8) \quad SSR = \sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

Therefore, what we have demonstrates is that

(9) SST = SSM + SSR

…or the actual variation in y (SST) is a function of two components. The first is the variation predicted by the model (SSM), while the second is the variation that we cannot predict (SSR).

Dividing both sides of (9) by SST, note that

$$1 = \frac{SSM}{SST} + \frac{SSR}{SST}$$

Or alternatively

$$(10) \qquad R^2 = \frac{SSM}{SST} = 1 - \frac{SSR}{SST}$$

The R^2 measures what fraction of the variation in y is explained by the regression model. Since SST = SSM + SSR, by construction $0 \le R^2 \le 1$.

Just a note about the textbook. The author calls the term $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ the SSE or sum of squared explained. PLEASE NOTE: The textbook definition of R^2 is $R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$ where the author defines SSE as sum of squared estimated. I do not like this abbreviation for $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$.
Our definition SSM matches much better with the STATA prints out – SST is sum of squared total, SSM is sum of squared model and SSR is sum of squared residuals – so we will use these abbreviations.

Proof that $\hat{\beta}_1$ is an Unbiased Estimate

Recall the definition for $\hat{\beta}_1$

$$(1) \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Recalling the properties of summations, note that numerator can be written as

$$(2) \qquad \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n} y_i(x_i - \bar{x})$$

Note further that the true relationship between y and y is given by the population regression line

$$(3) \qquad y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$$

Using (2) and substituting the true value for y into the model,

$$(4) \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} y_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i + \varepsilon_i)(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Break apart the terms in the numerator

$$(5) \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}\beta_0(x_i - \bar{x}) + \sum_{i=1}^{n}\beta_1 x_i(x_i - \bar{x}) + \sum_{i=1}^{n} \varepsilon_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

We can simplify the terms in (5) using the properties of summations:

In the first term in the numerator, note that β_0 is a constant and can be pulled outside the summation. As a result, we have the summation of a deviation from a mean, which equals zero

$$\sum_{i=1}^{n}\beta_0(x_i - \bar{x}) = \beta_0\sum_{i-1}^{n}(x_i - \bar{x}) = \beta_0(0) = 0$$

In the second term in the numerator, β_1 is a constant and it can pulled outside the summation. Recall also that $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i(x_i - \bar{x})$ so

$$\sum_{I=1}^{n}\beta_1 x_i(x_i-\bar{x})=\beta_1\sum_{i=1}^{n}x_i(x_i-\bar{x})=\beta_1\sum_{i=1}^{n}(x_i-\bar{x})^2$$

The first term in the numerator drops out, the second term reduces to β_1 and therefore, we can write the OLS estimate for $\hat{\beta}_1$ as

$$(6)\qquad \hat{\beta}_1=\beta_1+\frac{\sum_{i=1}^{n}\varepsilon_i(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

WE WILL BE USING THIS CHARACTERIZATION OF THE OLS ESTIMATE FOR $\hat{\beta}_1$ A LOT THIS SEMESTER. PLEASE UNDERSTAND HOW WE GOT TO THIS POINT.

Equation (6) points out two important things. First, the estimate for $\hat{\beta}_1$ is a function of the 'truth' that is the true value of β_1. Likewise, the estimated value for $\hat{\beta}_1$ is a function of the n people who were selected for this sample. The true source of randomness in the model is therefore the unknown residual ε_i. As a result, the properties of $\hat{\beta}_1$ will be a function of the properties we assume about ε_i. We typically make four assumptions about ε_i

Three assumptions about the residual ε_i

1) $E(\varepsilon_i) = E(\varepsilon_i|x_i) = 0$
2) $V(\varepsilon_i) = V(\varepsilon_i|x_i) = \sigma^2_\varepsilon$
3) $Cov(\varepsilon_i, \varepsilon_j) = 0$ for all $i\neq j$

The first assumption says that on average, the expected error is zero and that this expectation does not depend on the value of x. The second assumption says that the errors are "homoskedastic" or they have the same variance. Assumption (3) states that errors are not correlated across observations. The second and third assumptions will be relaxed throughout the semester.

Assumption (1) is the killer. If (1) is true, the model has very nice properties, if it false, the model is useless.

Assumption (1) states that ε and x are independent. This says that the realization of x conveys no information about the likely value of ε and therefore, the conditional expectation $E(\varepsilon_i|x_i)$ provides the same information as the unconditional expectation $E(\varepsilon_i)$.

Recall that $\text{cov}(x_i,\varepsilon_i)=E(x_i\varepsilon_i)-E(x_i)E(\varepsilon_i)$. Because $E(\varepsilon_i) = E(\varepsilon_i|x_i) = 0$ then the second term drops out and $\text{cov}(x_i,\varepsilon_i)=E(x_i\varepsilon_i)$. Let's work with the right hand side of this term. $E(x_i\varepsilon_i)=E(\varepsilon_i\mid x_i)x_i=E(\varepsilon_i)x_i=0$ and hence $\text{cov}(x_i,\varepsilon_i)=0$. In essence by conditioning on x, we "fix" this value and $E(x_i\varepsilon_i)$ becomes $E(\varepsilon_i)x_i$ which equals zero by assumption (1).

A key result we will use time and time again throughout the semester is that if we maintain assumption (1) and we see $E(\varepsilon_i x_i)$, this reduces to $E(\varepsilon_i)x_i$ which will equal zero.

As we will see, if the value of x conveys information about ε then the model is sunk. We will go over this in detail about two dozen times throughout the semester.

Let's also work with condition (2) a little. This states that the variance of ε_i is the same whether we know x or not. Recall the definition of variance $Var(\varepsilon_i) = E[(\varepsilon_i - E(\varepsilon_i))^2]$. Because $E(\varepsilon_i) = 0$, the definition of the variance reduces $Var(\varepsilon_i) = E[\varepsilon_i^2] = \sigma_\varepsilon^2$. Therefore, any time we see a $E[\varepsilon_i^2]$ this means σ_ε^2.

In the derivations below, we will also see a lot of terms that are $E[\varepsilon_i^2 x_i^2]$. Given assumption (2), $E[\varepsilon_i^2 x_i^2] = E[\varepsilon_i^2 \mid x_i^2] = E[\varepsilon_i^2]x_i^2 = \sigma_i^2 x_i^2$

Therefore, a key result we will use time and time again throughout the semester -- if we maintain assumption (2) and we see $E[\varepsilon_i^2 x_i^2]$ this reduces to $E[\varepsilon_i^2]x_i^2]$ which equal $\sigma_i^2 x_i^2$.

For now let's concentrate on the case if (1) is true and see what that buys us.

We have established that $\hat{\beta}_1$ is a random variable. Any time you have a random variable, the first two questions you need to ask are a) what is the expected value and b) what is the variance. In this section, we will produce E[$\hat{\beta}_1$]

First start with the definition of $\hat{\beta}_1$ from in equation (6) and take the expectation

$$(7) \qquad E[\hat{\beta}_1] = E\left[\beta_1 + \frac{\sum_{i=1}^{n}\varepsilon_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] = E[\beta_1] + E\left[\frac{\sum_{i=1}^{n}\varepsilon_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] = \beta_1 + \frac{E\left[\sum_{i=1}^{n}\varepsilon_i(x_i - \bar{x})\right]}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

There is a lot going on in equation (7). First note that E[a+b]=E[a]+E[b] so we can break apart the two big terms in the expectation. Second, note that the true value β_1 is a fixed constant there E[β_1]= β_1. Note also that because we assume x is "fixed" then $\sum_{i=1}^{n}(x_i - \bar{x})^2$ is not random and it too can be brought outside the expectation.

Therefore, the properties of E[$\hat{\beta}_1$] will be driven by the expectation $E\left[\sum_{i=1}^{n}\varepsilon_i(x_i - \bar{x})^2\right]$.

Let's work with this term. First, write out the terms in the summation under the expectation

(8) $E\left[\sum_{i=1}^{n}\varepsilon_i(x_i-\bar{x})^2\right]=E\left[\varepsilon_1(x_1-\bar{x})\right]+E\left[\varepsilon_2(x_2-\bar{x})\right]+E\left[\varepsilon_3(x_3-\bar{x})\right]+\cdots E\left[\varepsilon_n(x_n-\bar{x})\right]$

Consider one of these expectations $E\left[\varepsilon_i(x_i-\bar{x})\right]$ for any i. Break this term apart to read $E[\varepsilon_i x_i]+E[\varepsilon_i\bar{x}]$. Note assumption (1) above states that $E(\varepsilon_i|x_i) = 0$. Looking at the first term of $E[\varepsilon_i x_i]+E[\varepsilon_i\bar{x}]$, we can easily write it as

$$E[\varepsilon_i x_i]+E[\varepsilon_i \mid x_i]=0$$

Therefore, if assumption (1) is correct, this term should be zero. The second term in $E[\varepsilon_i x_i]+E[\varepsilon_i\bar{x}]$ requires the definition of $\bar{x}$ which is

$$\bar{x}=\frac{1}{n}\left(x_1+x_2+.....x_n\right)$$

And substituting this in generates

$$E[\varepsilon_i\bar{x}]=E\left[\frac{\varepsilon_i x_1}{n}+\frac{\varepsilon_i x_2}{n}+..\frac{\varepsilon_i x_i}{n}+..\frac{\varepsilon_i x_n}{n}\right]=E\left[\frac{\varepsilon_i x_1}{n}\right]+E\left[\frac{\varepsilon_i x_2}{n}\right]+...E\left[\frac{\varepsilon_i x_n}{n}\right]$$

Note that each term $E\left[\frac{\varepsilon_i x_j}{n}\right]$ for $j\neq i$ is 0 and the term $E\left[\frac{\varepsilon_i x_i}{n}\right]$ must be equal to zero for the same arguments above. As a result, the far right hand term in equation (7) is zero and therefore,

(9) $E[\hat{\beta}_1] = \beta_1$

The estimate $\hat{\beta}_1$ is an unbiased estimate of β_1, that is, if one were to draw a large number of samples at random, estimate $\hat{\beta}_1$ each time, the average of all these estimates would be the true value β_1

Please note --- an unbiased estimate does not mean you have the correct estimate – it simply means that you used a procedure that on average will give you the correct answer.

Here is another way to think about how the correlation between x and ε would get you into trouble

From equation (6), divide the numerator and denominator of the right hand term by (n-1)

(10) $$\hat{\beta}_1 = \beta_1+\frac{\frac{1}{n-1}\sum_{i=1}^{n}\varepsilon_i(x_i-\bar{x})}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \beta_1+\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i-\bar{\varepsilon})(x_i-\bar{x})}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Notice also in the final term in (10), we use the fact that the numerator can be written as

$$\sum_{i=1}^{n} \varepsilon_i (x_i - \overline{x}) = \sum_{i=1}^{n} (\varepsilon_i - \overline{\varepsilon})(x_i - \overline{x})$$

The numerator is the nothing more than the sample correlation between x_i and the ACTUAL error term ε_i. The denominator is the sample variance in x.

$$\hat{\sigma}_{x\varepsilon} = \frac{1}{n-1} \sum_{i=1}^{n} (\varepsilon_i - \overline{\varepsilon})(x_i - \overline{x})$$

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

And therefore, the estimate $\hat{\beta}_1$ can be written as

$$(11) \quad \hat{\beta}_1 = \beta_1 + \frac{\hat{\sigma}_{x\varepsilon}}{\hat{\sigma}_x^2}$$

Notice that if in a sample $\hat{\sigma}_{x\varepsilon} = 0$, then $\hat{\beta}_1 = \beta_1$. However, if $\hat{\sigma}_{x\varepsilon} > 0$ then by construction $\hat{\beta}_1 > \beta_1$ whereas if $\hat{\sigma}_{x\varepsilon} < 0$ then $\hat{\beta}_1 < \beta_1$

The Variance of $\hat{\beta}_1$

Demonstrating the $\text{var}(\hat{\beta}_1)$ is the most detailed and complicated derivation we will do all semester. In the end, it is a lot of algebra but it simply exploits the properties of definitions and expectations we have already used.

To start, recall the following facts:

(a) The equation for the estimate of $\hat{\beta}_1$

$$(1) \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

(b) Recall also that the true underlying relationship between x and y is given by the equation

$$(2) \quad y_i = \beta_0 + x_i\beta_1 + \varepsilon_i$$

(c) To analyze some of the properties of $\hat{\beta}_1$, we substituted the true value for y_i, as defined by equation (2) into the estimate (1). This substitution leads to the following result:

$$(3) \qquad \hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}\varepsilon_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

By definition,

$$(4) \qquad Var(\hat{\beta}_1) = E[(\hat{\beta}_1 - E(\hat{\beta}_1))^2]$$

Previously, we demonstrated that $\hat{\beta}_1$ is an unbiased estimate or $E[\hat{\beta}_1]=\beta_1$ and therefore, substituting β_1 for $E[\hat{\beta}_1]$ in equation (4), the $Var(\hat{\beta}_1)$ is then

$$(5) \qquad Var(\hat{\beta}_1) = E[(\hat{\beta}_1 - \beta_1)^2]$$

Looking at equation (3), note that the difference $\hat{\beta}_1 - \beta_1$ is simply

$$(6) \qquad \hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n} \varepsilon_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} \varepsilon_i (x_i - \bar{x})}{SST_x} \qquad where\ SST_x = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

The variable SST_x is the sum of squared total for x and similar to the SST for y used in the construction of the R^2

Using the definition of the variance and equation (6)

$$(7) \quad Var(\hat{\beta}_1) = E[(\hat{\beta}_1 - \beta_1)^2] = E\left[\left(\frac{\sum_{i=1}^{n} \varepsilon_i (x_i - \bar{x})}{SST_x}\right)^2\right] = E\left[\left(\frac{\sum_{i=1}^{n} \varepsilon_i \tilde{x}_i}{SST_x}\right)^2\right]$$

Where $\tilde{x}_i = x_i - \bar{x}$. Because SST_x is a constant (x is considered fixed) we can bring it outside the summation. Therefore

$$(8) \quad Var(\hat{\beta}_1) = \frac{1}{SST_x^2} E\left[\left(\sum_{i=1}^{n} \varepsilon_i \tilde{x}_i\right)^2\right]$$

Let's work with the numerator in the far right hand term in equation (8). Complete the square on this term.

$$\left[\sum_{i=1}^{n} \varepsilon_i \tilde{x}_i\right]^2 = \left[\varepsilon_1 \tilde{x}_1 + \varepsilon_2 \tilde{x}_2 + ...\varepsilon_n \tilde{x}_n\right]^2 = [\varepsilon_1^2 \tilde{x}_1^2 + \varepsilon_2^2 \tilde{x}_2^2 + ...\varepsilon_n^2 \tilde{x}_n^2 + 2\varepsilon_1 \tilde{x}_1 \varepsilon_2 \tilde{x}_2 + 2\varepsilon_1 \tilde{x}_1 \varepsilon_3 \tilde{x}_3 + ...2\varepsilon_{n-1} \tilde{x}_{n-1} \varepsilon_n \tilde{x}_n]$$

Next, we take the expectation of this term

$$(10) \quad E\left[\sum_{i=1}^{n} \varepsilon_i \tilde{x}_i\right]^2 = E[\varepsilon_1^2 \tilde{x}_1^2 + \varepsilon_2^2 \tilde{x}_2^2 + ...\varepsilon_n^2 \tilde{x}_n^2 + 2\varepsilon_1 \tilde{x}_1 \varepsilon_2 \tilde{x}_2 + 2\varepsilon_1 \tilde{x}_1 \varepsilon_3 \tilde{x}_3 + ...2\varepsilon_{n-1} \tilde{x}_{n-1} \varepsilon_n \tilde{x}_n]$$

$$= E[\varepsilon_1^2 \tilde{x}_1^2] + E[\varepsilon_2^2 \tilde{x}_2^2] + ...E[\varepsilon_n^2 \tilde{x}_n^2] + E[2\varepsilon_1 \tilde{x}_1 \varepsilon_2 \tilde{x}_2] + E[2\varepsilon_1 \tilde{x}_1 \varepsilon_3 \tilde{x}_3] + ...E[2\varepsilon_{n-1} \tilde{x}_{n-1} \varepsilon_n \tilde{x}_n]$$

Let's look at the terms in equation (10). Consider $E[\varepsilon_j^2 \tilde{x}_j^2]$ for any j=1, 2…n. Recall above from assumption (2) that anytime we see $E[\varepsilon_j^2 \tilde{x}_j^2]$ this reduces to $E[\varepsilon_j^2]\tilde{x}_j^2 = \sigma_\varepsilon^2 \tilde{x}_j^2$ because $E[\varepsilon_i^2 \mid x_i] = E[\varepsilon_i^2]$. Note also that we established above that any time we see $E[\varepsilon_i^2]$ this equals $E[\varepsilon_i^2] = \sigma_\varepsilon^2$. Therefore, the first n terms in the second line of equation (10), $E[\varepsilon_j^2 \tilde{x}_j^2]$, equal $\sigma_\varepsilon^2 \tilde{x}_j^2$ for j=1,2,…n.

Next, consider the expectation of the cross terms $E[2\varepsilon_i \varepsilon_j \tilde{x}_i \tilde{x}_j]$. The 2 is a constant so it can be brought outside the expectation. By assumption, $\tilde{x}_i \tilde{x}_j$ are also constants so they can be brought outside the expectations as well. Therefore, $E[2\varepsilon_i\varepsilon_j\tilde{x}_i\tilde{x}_j] = 2\tilde{x}_i\tilde{x}_j E[\varepsilon_i\varepsilon_j]$. Recall above that we assumed that $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ and the definition of $\text{cov}(\varepsilon_i, \varepsilon_j)$ is $\text{cov}(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i\varepsilon_j] - E[\varepsilon_i]E[\varepsilon_j]$ and since $E[\varepsilon_i]=E[\varepsilon_j]=0$, $\text{cov}(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i\varepsilon_j] = 0$. Therefore, all the expectation of cross-terms in (10) are zero. Combining these results

$$(11) \quad Var(\hat{\beta}_1) = \frac{1}{SST_x^2} E\left[\left(\sum_{i=1}^{n} \varepsilon_i \tilde{x}_i\right)^2\right] = \frac{1}{SST_x^2}[\sigma_\varepsilon^2\tilde{x}_1^2 + \sigma_\varepsilon^2\tilde{x}_2^2 + \sigma_\varepsilon^2\tilde{x}_3^2 + ...\sigma_\varepsilon^2\tilde{x}_n^2]$$

We can reduce the numerator of 11,

$$(12) \quad [\sigma_\varepsilon^2\tilde{x}_1^2 + \sigma_\varepsilon^2\tilde{x}_2^2 + \sigma_\varepsilon^2\tilde{x}_3^2 + ...\sigma_\varepsilon^2\tilde{x}_n^2] = \sigma_\varepsilon^2 \sum_{i=1}^{N} \tilde{x}_i^2 = \sigma_\varepsilon^2 \sum_{i=1}^{N} (x_i - \tilde{x})^2 = \sigma_\varepsilon^2 SST_x$$

And therefore:

$$(13) \quad Var(\hat{\beta}_1) = \frac{1}{SST_x^2}\sigma_\varepsilon^2 SST_x = \frac{\sigma_\varepsilon^2}{SST_x} = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Notice that the definition of (13) includes a σ_ε^2 which is the Var(ε_i). Unfortunately, we do not know σ_ε^2 so it must be estimated

An unbiased estimate for σ_ε^2 is as follows

$$(14) \quad \hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n-k-1} = \frac{SSR}{n-k-1}$$

Where k is the number of x's included in the model. Thus in the simple bivariate model, k=1 and the degrees of freedom in the denominator is n-2.

The estimated variance of $\hat{\beta}_1$ is then

$$(15) \quad Est.Var(\hat{\beta}_1) = \frac{\hat{\sigma}_\varepsilon^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

As with all variances, the units of measure on (15) are in $\hat{\beta}_1$ squared units so we need to take the square root. The square root of this variance is typically called the "Standard error"

$$(16) \quad se(\hat{\beta}_1) = \frac{\hat{\sigma}_\varepsilon}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2}}$$