# STANDARD ERROR OF THE SLOPE

Fundamentals of

**PROGRAM EVALUATION**

JESSE LECY

# STANDARD ERROR
# OF THE SLOPE

We want to end up here.
Need to work backwards.

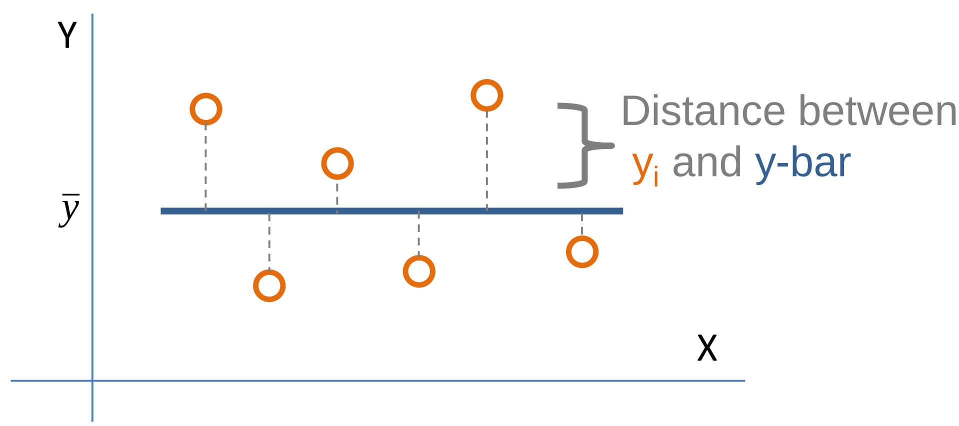
This is the one formula
you need to remember.

$$\text{Standard Error of the Slope} \approx \frac{\text{residual}}{\text{sample size} \cdot \text{variance X}}$$

$$SE_{b_1} = \frac{sd_e}{\sqrt{(n-1) \cdot \mathrm{var}(x)}}$$

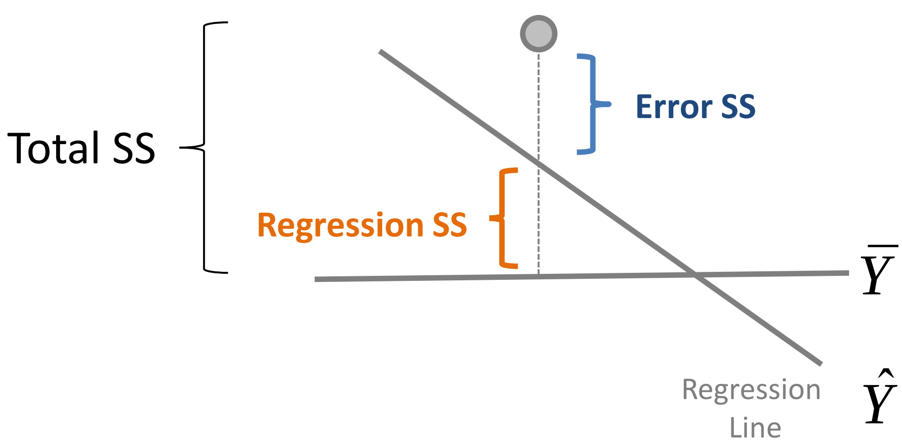
# THE VARIANCE CALCULATION

$$\mathrm{var}(y) = \frac{\sum \overbrace{(y_i - \bar{y})}^{\text{Distance between } y_i \text{ and y-bar}}{}^2}{n-1}$$




**Variance:** square the distances, add them up, divide by n-1

# PARTITIONING THE VARIANCE OF Y

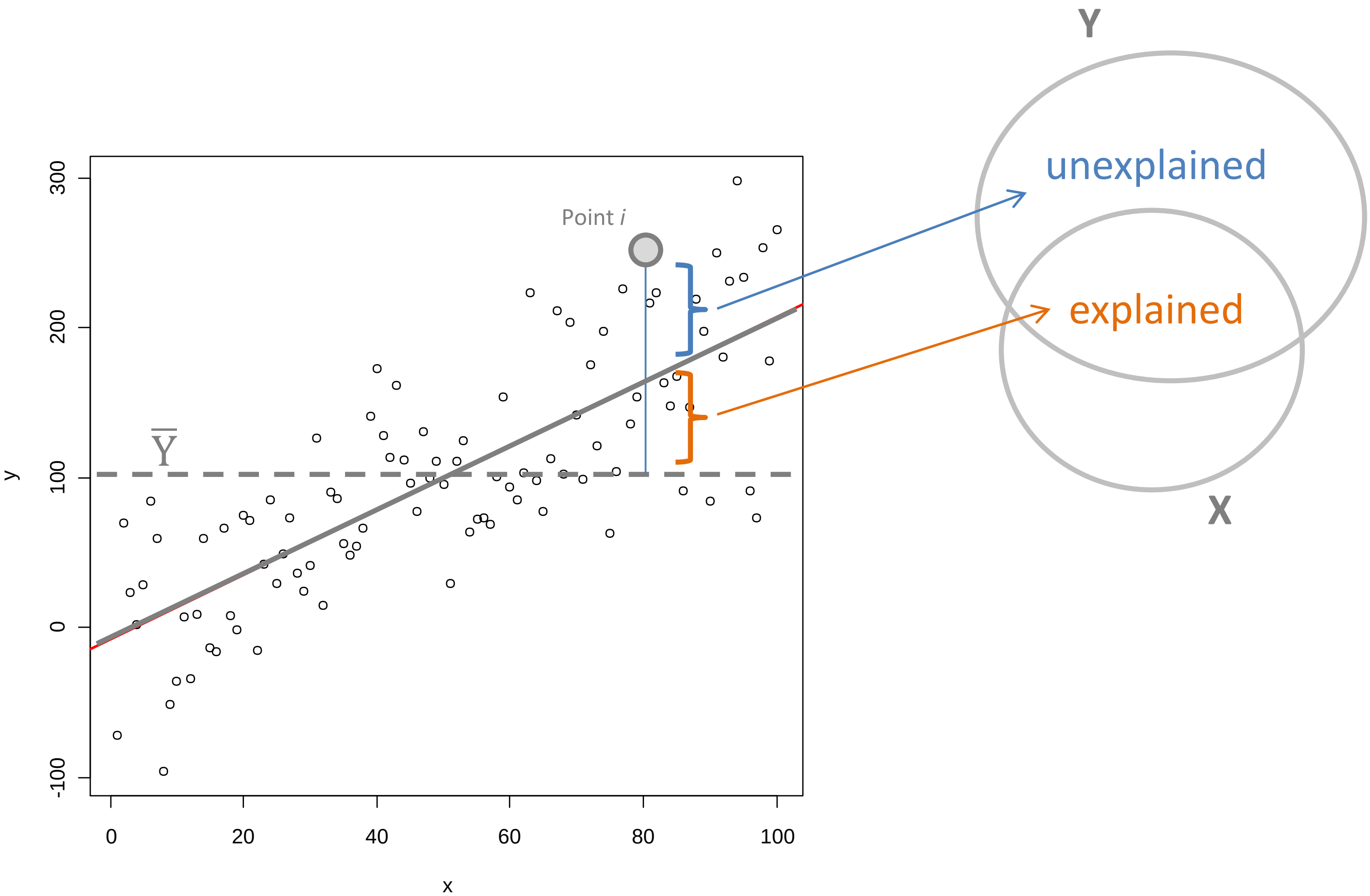



$$TotalSS = \sum (y_i - \bar{y})^2$$

$$RegressionSS = \sum (\hat{y}_i - \bar{y})^2$$

$$ErrorSS = \sum (y_i - \hat{y}_i)^2$$

$$TSS = RSS + ESS$$




$$\textit{Unexplained}:\quad Y_i - \hat{Y}_i = e_i$$

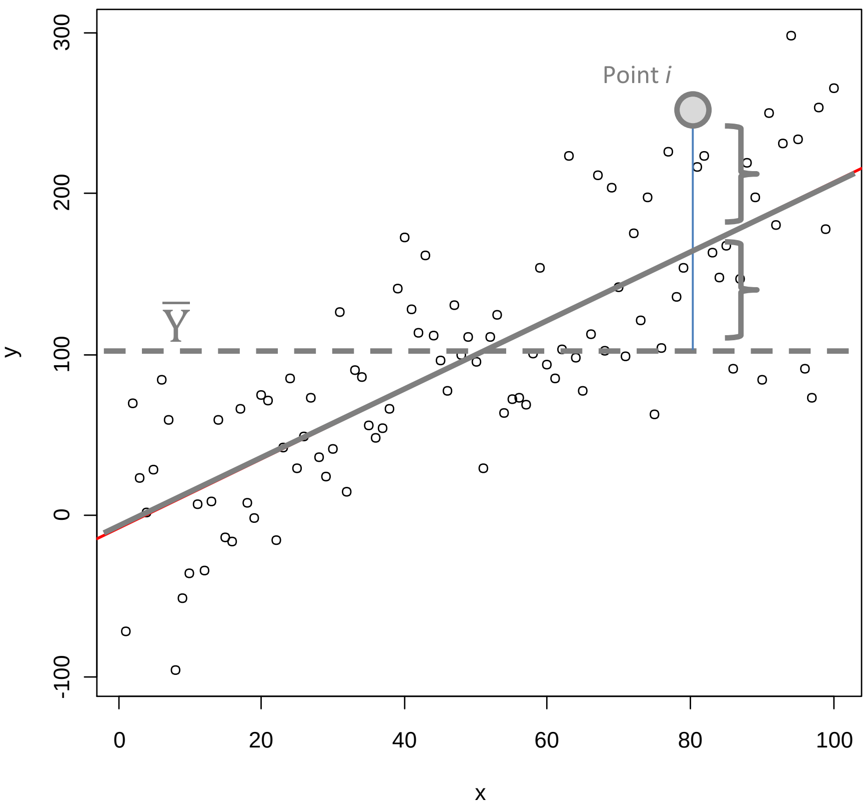
$$\textit{Explained}:\quad \hat{Y}_i - \bar{Y}$$

## Two parts of the variance of Y

The Venn diagram is a simplified representation of the regression model. In our regression, the explained portion of the variance of outcome will always be the distance from the mean to the predicted value of Y (which always falls on the regression line), and the unexplained portion is the distance between the regression line and the actual data point, also called the residual or the error *e*.

# STANDARD ERROR OF THE SLOPE

# Standard Error in Regression




$$Y_i - \hat{Y}_i = \boldsymbol{e_i}$$

$$Y_i - \bar{Y}$$

$$SSE = \sum e_i^2$$

**Sum of Squared Error Terms**

$$\hat{\sigma}_\varepsilon^2 = \frac{SSE}{n-2}$$

**Variance of the residual**

$$SE_{b_1} = \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\sum (x_i - \bar{x})^2}}$$

**Standard error of the slope**

The standard error of the slope is one of the most important concepts in regression because it determines the size of the confidence interval and thus the statistical significance of our study. We want standard errors to be as small as possible. We see here that the size of the standard error will be directly related to the amount of unexplained variance we have in our model, the residual *e*. The important thing to note is the unexplained portion shows up in the numerator of the standard error. As a result, the size of the standard error will be proportional to the amount of unexplained variance (plus a couple of other considerations to be covered later).

# STANDARD ERROR IN REGRESSION

$$SE_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

The size of the standard error of the mean is driven by the variance of the variable, and the sample size.

$$\mathrm{var}(x) = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \Rightarrow$$

$$\boxed{(n-1)\cdot \mathrm{var}(x)} = \boxed{\sum (x_i - \bar{x})^2}$$

We can write the formula for the standard error of the slope in a couple of ways. I prefer the top because it is explicit about sample size and var(x).

$$SE_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)\cdot \mathrm{var}(x)}}$$

$$\Updownarrow$$

$$SE_{b_1} = \frac{s_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}}$$

Similarly, the standard error of the slope is a function of the variance of the residual (the amount of unexplained variance in the outcome), the sample size (n-1), AND the variance of the explanatory variable.

# THE INTUITIVE STANDARD ERROR

$$SE_{b_1} = \sqrt{\frac{\sigma_\varepsilon^2}{\sum (x_i - \bar{x})^2}}$$

## NOTE:

$$\mathrm{var}(x) = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \Rightarrow$$

$$(n-1) \cdot \mathrm{var}(x) = \sum (x_i - \bar{x})^2$$

## THUS:

$$SE_{b_1} = \sqrt{\frac{\sigma_\varepsilon^2}{(n-1)\,\mathrm{var}(x)}}$$

## THEREFORE:

$$SE_{b_1} \approx \frac{residual_y}{sample\, size \cdot \mathrm{var}(x)}$$

$$SE_{b_1} = \sqrt{\frac{\hat{\sigma}_{\varepsilon}^2}{\sum (x_i - \bar{x})^2}}$$

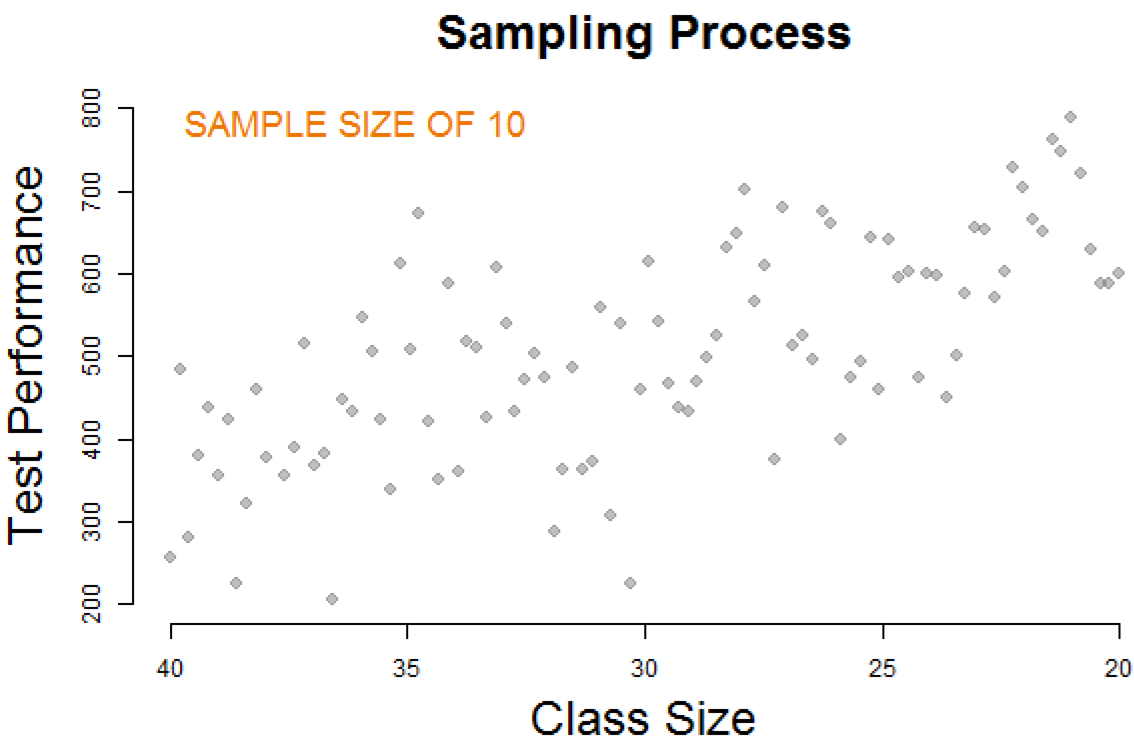
### Standard error of the Slope

Don't get too caught up with the math. The formula for the standard error of a regression coefficient is actually quite simple when you break it down. There are three moving parts – three things that can affect the size of the standard error. The portion of unexplained variance of the dependent variable (the residual), the sample size of the regression, and the amount of variance in the variable X associated with the regression slope.

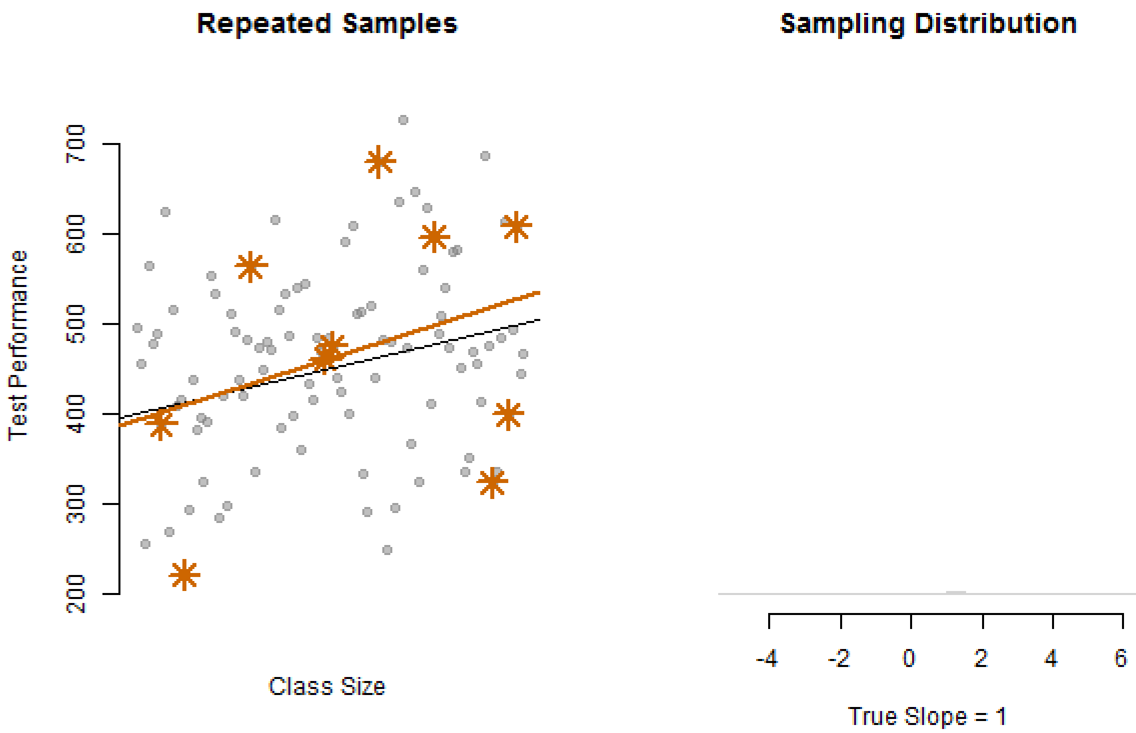
**Standard Error of the Slope** $\approx \dfrac{\text{residual}}{\text{sample size} \cdot \text{variance X}}$

$$SE_{b_1} = \frac{sd_e}{\sqrt{(n-1) \cdot \mathrm{var}(x)}}$$

# STANDARD ERROR AS A SAMPLING STATISTIC

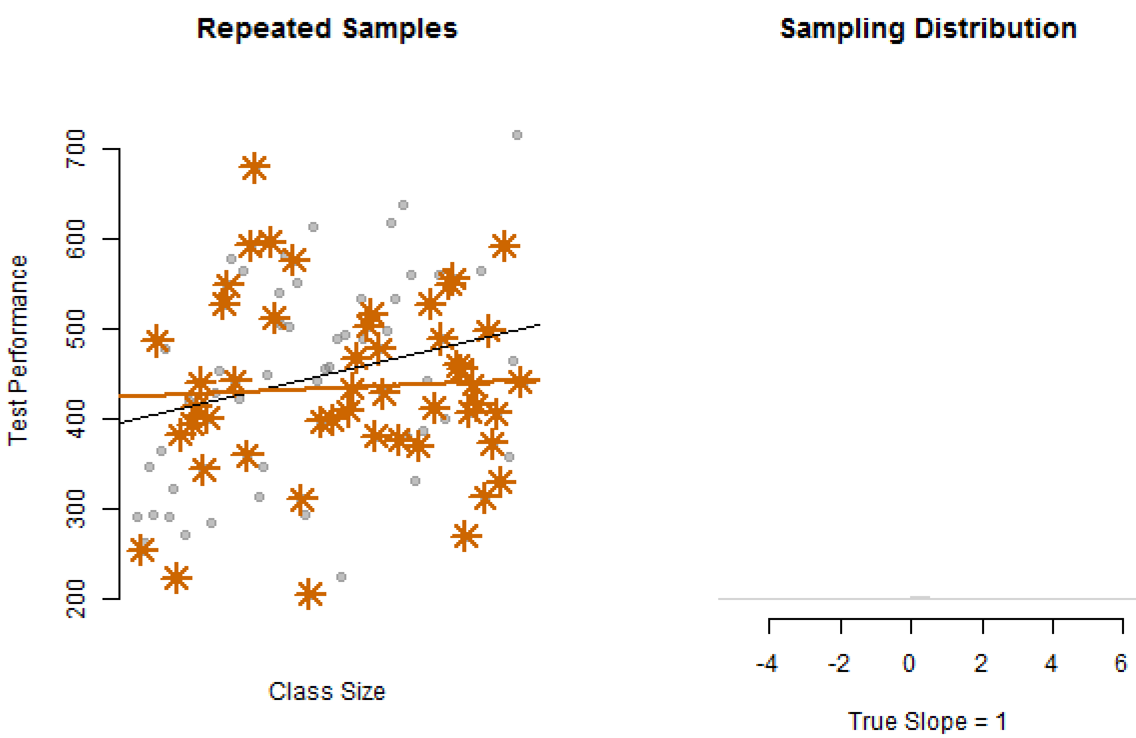


# STANDARD ERROR OF THE SLOPE



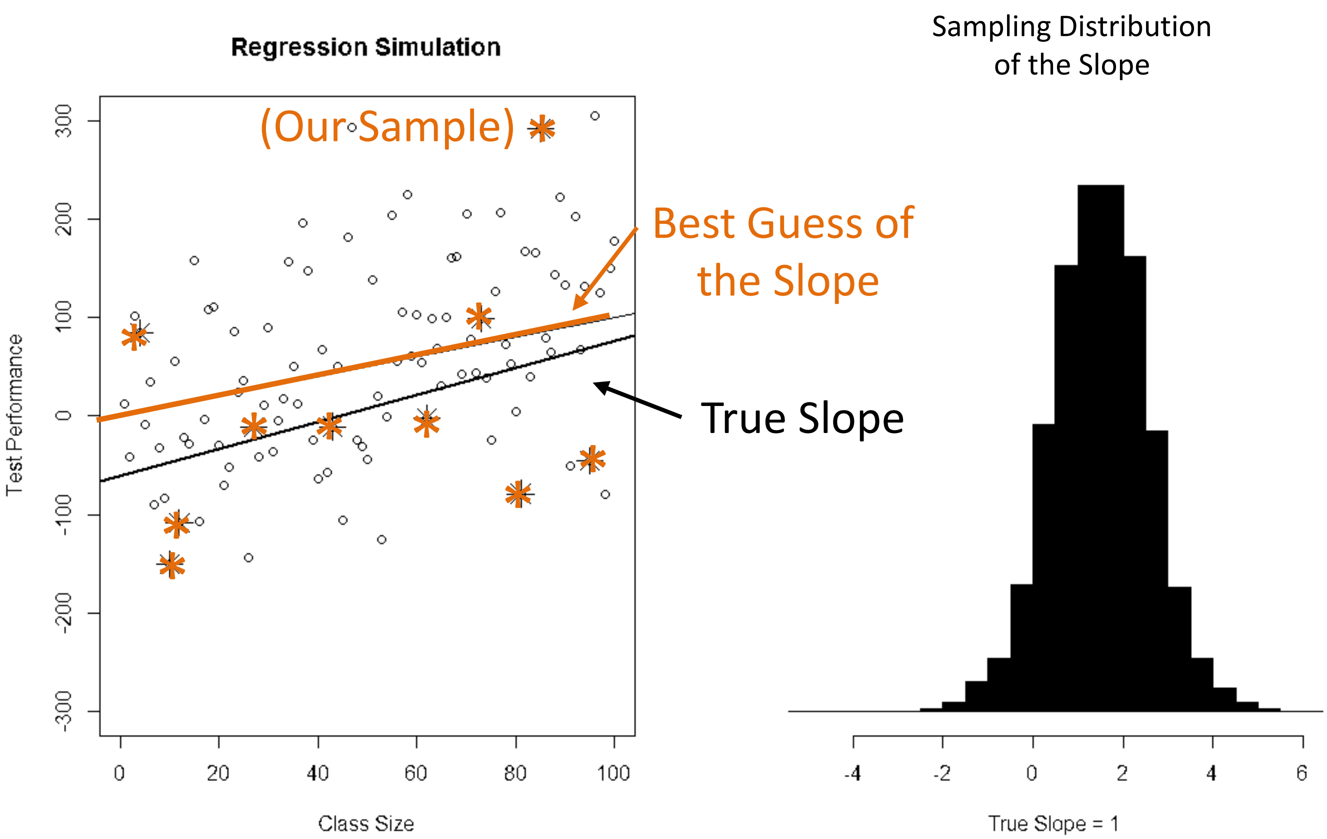

## SAMPLE SIZE = 10

# STANDARD ERROR OF THE SLOPE




## SAMPLE SIZE = 50

# STANDARD ERROR OF THE SLOPE




$$SE_{b_1} = \sqrt{\frac{\sigma_\varepsilon^2}{\sum (x_i - \bar{x})^2}}$$

**"AVERAGE ERROR" (OF THE SLOPE ESTIMATE)**

# TRANSLATING CONCEPTS

Standard Error of the Mean

$$SE_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

Standard Error of the Slope

$$SE_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)\cdot \mathrm{var}(x)}}$$

X ← Source of Variance → Y

Increase sample size ← Reduction of St. Error → (1) Increase sample size
(2) Explain more variance of Y
(i.e. add controls)
(3) Increase variance of X

# STANDARD DEVIATION VERSUS STANDARD ERROR

The standard deviation is, **how far the data is from the mean**, on average.

~

The standard error is, **how far our best guest is from 'the truth**', on average.

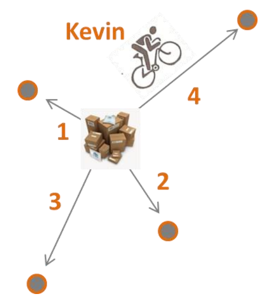
The 'truth' means different things depending upon what kind of standard error you are calculating.

# THE ROAD MAP

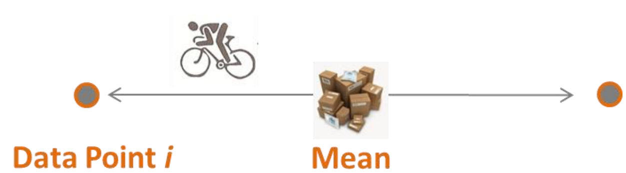
| | Of the **Mean**: | Of the **Slope**: |
|---|---|---|
| Sampling Variance: ↓ | $\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ (for x) | $\sigma_\varepsilon^2 = \frac{SSE}{n-2} = \frac{\sum e_i^2}{n-2}$ (using the residual) |
| Standard Deviation: ↓ | $\sigma_x = \sqrt{\sigma_x^2}$ | $\sigma_\varepsilon = \sqrt{\sigma_\varepsilon^2}$ |
| Standard Error: ↓ | $SE_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ | $SE_{b_1} = \sqrt{\frac{\sigma_\varepsilon^2}{\sum (x_i - \bar{x})^2}}$ |
| Confidence Interval | $\mu = \bar{x} \pm t \cdot SE_{\bar{x}}$ (of the mean) | $\beta_1 = b_1 \pm t \cdot SE_{b_1}$ (of the slope) |

# USEFUL METAPHORS

Variance

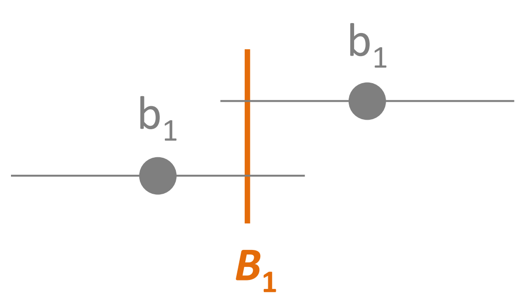



Standard Deviation




Standard Error

Confidence Interval



# What should be clear in my mind?

1. **We split the variance of Y into explained and unexplained portions with a trick, inserting the regression line y-hat.**

2. **The standard error of the slope is derived from the unexplained portion of Y, the residual.**