Statistics Bootcamp Day 5

20 September 2019

Welcome to bootcamp!

Our goals:

1. Increase students' understanding of and confidence with basic statistical concepts.
2. Build students' programming intuition and data management skills.
3. Encourage collaboration and camaraderie among the graduate student cohort.

Overview of the week

~~Monday: mindset, descriptive & inferential statistics, summary statistics, and Stata workshop~~

~~Tuesday: graphing, exponents/logarithms, sampling distributions, and statistical significance~~

~~Wednesday: probability basics, file structure and data workflow~~

~~Thursday: variable types, functions, lines of best fit, prediction equations~~

Friday: matrix algebra basics, reading calculus

Today’s learning objectives

...understand what a matrix and a vector are

...and how to multiply matrices with vectors

...be able to represent a prediction equation in matrix notation

...read the notation of, understand, and interpret basic calculus relevant to a statistics context (e.g. limits, derivatives, integrals)

On a notecard...

Write down the general form of a linear regression equation for an outcome Y with two predictor variables.

On a notecard...

Write down the general form of a linear regression equation for an outcome Y with two predictor variables.

$$Y = a + bX + cZ + \epsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon$$

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon}$$

$$\hat{y}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

Why do some of these equations have “e” at the end and others do not?

$$\text{Happiness} = \beta_0 + \beta_1\text{traffic} + \varepsilon$$

	happy	traffic
1	23	1
2	25	1
3	31	0

Goal of OLS: Find values of β_0 and β_1 that minimize the sum of squared errors.

Happiness = $\beta_0 + \beta_1$traffic + ε

$$\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix} = \beta_0 + \beta_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

Goal of OLS: Find values of β_0 and β_1 that minimize the sum of squared errors.

Happiness = $\beta_0 + \beta_1$traffic + ε

$$\underbrace{\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix}}_{\text{vector}} = \underbrace{\beta_0}_{\text{scalar}} + \underbrace{\beta_1}_{\text{scalar}} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\text{vector}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}_{\text{vector}}$$

Goal of OLS: Find values of β_0 and β_1 that minimize the sum of squared errors.

Happiness = $\beta_0 + \beta_1$traffic + ε

$$\underbrace{\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix}}_{\text{vector}} = \underbrace{\beta_0}_{\text{scalar}} + \underbrace{\beta_1}_{\text{scalar}} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\text{vector}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}_{\text{vector}}$$

Goal of OLS: Find values of β_0 and β_1 that minimize the sum of squared errors.

$\boxed{\text{scalar * vector} \rightarrow \text{“distribute”}}$

$$Happiness = \beta_0 + \beta_1 traffic + \varepsilon$$

```
. reg happy traffic

      Source |       SS           df       MS      Number of obs   =         3
-------------+----------------------------------   F(1, 1)         =     16.33
       Model |  32.6666667         1  32.6666667   Prob > F        =    0.1544
    Residual |           2         1           2   R-squared       =    0.9423
-------------+----------------------------------   Adj R-squared   =    0.8846
       Total |  34.6666667         2  17.3333333   Root MSE        =    1.4142

------------------------------------------------------------------------------
       happy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     traffic |         -7   1.732051    -4.04   0.154    -29.00779    15.00779
       _cons |         31   1.414214    21.92   0.029     13.03071    48.96929
------------------------------------------------------------------------------
```

Happiness = $\beta_0 + \beta_1$traffic + ε

$$\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix} = 31 - 7 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

scalar + **vector** → add straight across

vector + **vector** → add straight across

$$\text{Happiness} = \beta_0 + \beta_1\text{traffic} + \beta_2\text{dogs} + \varepsilon$$

	happy	traffic	dogs
1	23	1	5
2	25	1	6
3	31	0	1

Goal of OLS: Find values of β_0 and β_1 and β_2 that minimize the sum of squared errors.

$$\text{Happiness} = \beta_0 + \beta_1\text{traffic} + \beta_2\text{dogs} + \varepsilon$$

$$\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix} = \beta_0 + \beta_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \beta_2 \begin{bmatrix} 5 \\ 6 \\ 1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

Goal of OLS: Find values of β_0 and β_1 and β_2 that minimize the sum of squared errors.

$$\text{Happiness} = \beta_0 + \beta_1\text{traffic} + \beta_2\text{dogs} + \varepsilon$$

```
. reg happy traffic dogs

      Source |       SS           df       MS      Number of obs   =         3
-------------+----------------------------------   F(2, 0)         =         .
       Model |  34.6666667         2  17.3333333   Prob > F        =         .
    Residual |           0         0           .   R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =         .
       Total |  34.6666667         2  17.3333333   Root MSE        =         0

------------------------------------------------------------------------------
       happy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     traffic |        -16          .        .       .            .           .
        dogs |          2          .        .       .            .           .
       _cons |         29          .        .       .            .           .
------------------------------------------------------------------------------
```

Happiness = $\beta_0 + \beta_1$traffic + $\beta_2$dogs + ε

$$\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix} = 29 - 16 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ 6 \\ 1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

Happiness = $\beta_0 + \beta_1$traffic + $\beta_2$dogs + ε

$$\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix} = 29 - 16 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ 6 \\ 1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

$$\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix} = 29 + \begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -16 \\ 2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

$$\text{Happiness} = \beta_0 + \beta_1\text{traffic} + \beta_2\text{dogs} + \varepsilon$$

$$\underset{\text{vector}}{\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix}} = \underset{\text{scalar}}{29} - \underset{\text{scalar}}{16} \underset{\text{vector}}{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} + \underset{\text{scalar}}{2} \underset{\text{vector}}{\begin{bmatrix} 5 \\ 6 \\ 1 \end{bmatrix}} + \underset{\text{vector}}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}$$

$$\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix} = 29 + \begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -16 \\ 2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

Happiness = $\beta_0 + \beta_1$traffic + $\beta_2$dogs + ε

$$\underbrace{\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix}}_{\text{vector}} = \underbrace{29}_{\text{scalar}} - \underbrace{16}_{\text{scalar}} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\text{vector}} + \underbrace{2}_{\text{scalar}} \underbrace{\begin{bmatrix} 5 \\ 6 \\ 1 \end{bmatrix}}_{\text{vector}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}_{\text{vector}}$$

$$\underbrace{\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix}}_{\text{vector}} = \underbrace{29}_{\text{scalar}} + \underbrace{\begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix}}_{\text{MATRIX}} \cdot \underbrace{\begin{bmatrix} -16 \\ 2 \end{bmatrix}}_{\text{vector}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}_{\text{vector}}$$

Matrix-by-vector multiplication

$$\underbrace{\begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix}}_{\text{MATRIX}} \cdot \underbrace{\begin{bmatrix} -16 \\ 2 \end{bmatrix}}_{\text{vector}}$$

Matrix-by-vector multiplication

$$\begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -16 \\ 2 \end{bmatrix}$$

MATRIX | vector

3 x 2 | 2 x 1

These must match!

Matrix-by-vector multiplication

$$\begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -16 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot -16 + 5 \cdot 2 \\ 1 \cdot -16 + 6 \cdot 2 \\ 0 \cdot -16 + 1 \cdot 2 \end{bmatrix}$$

vector

MATRIX

vector

3 x 2

2 x 1

3 x 1

These must match!

Matrix-by-vector multiplication

$$\begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -16 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \cdot -16 + 5 \cdot 2 \\ 1 \cdot -16 + 6 \cdot 2 \\ 0 \cdot -16 + 1 \cdot 2 \end{bmatrix}$$

vector

MATRIX

vector

3 x 2

2 x 1

3 x 1

These must match!

matrix * **vector** → go across rows in first matrix/vector and down columns in the second matrix/vector, first multiplying then adding

Happiness = $\beta_0 + \beta_1$traffic + $\beta_2$dogs + ε

$$\underbrace{\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix}}_{\text{vector}} = \underbrace{29}_{\text{scalar}} + \underbrace{\begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix}}_{\text{MATRIX}} \cdot \underbrace{\begin{bmatrix} -16 \\ 2 \end{bmatrix}}_{\text{vector}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}}_{\text{vector}}$$

Happiness = $\beta_0 + \beta_1$traffic + $\beta_2$dogs + ε

$$\begin{bmatrix} 23 \\ 25 \\ 31 \end{bmatrix} = 29 + \begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -16 \\ 2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}$$

vector

scalar

MATRIX

vector

vector

$$Y_i = \beta_0 + \mathbf{X_i}\boldsymbol{\beta} + \epsilon_i$$

General OLS regression equation in matrix form:

$$Y_i = \beta_0 + \mathbf{X_i}\boldsymbol{\beta} + \epsilon_i$$

The “i” subscript indicates that these are vectors or matrices, **not scalars**

The X is bold to indicate a **matrix**

The beta is bold to indicate it is a **vector** of multiple coefficients

It comes after the X in order for the matrix multiplication to work properly

Practice: Matrices

scalar * vector → "distribute"

scalar + vector → add straight across

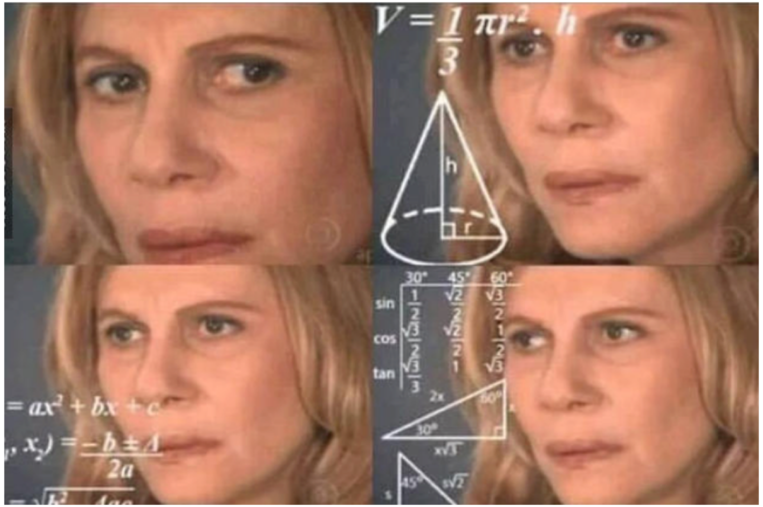
vector + vector → add straight across

matrix * vector → go across rows in first matrix/vector and down columns in the second matrix/vector, first multiplying then adding

Goal of OLS: Find βs to minimize the sum of squared errors

There must be a better way to minimize the sum of squared errors than trial and error?

YES! CALCULUS!



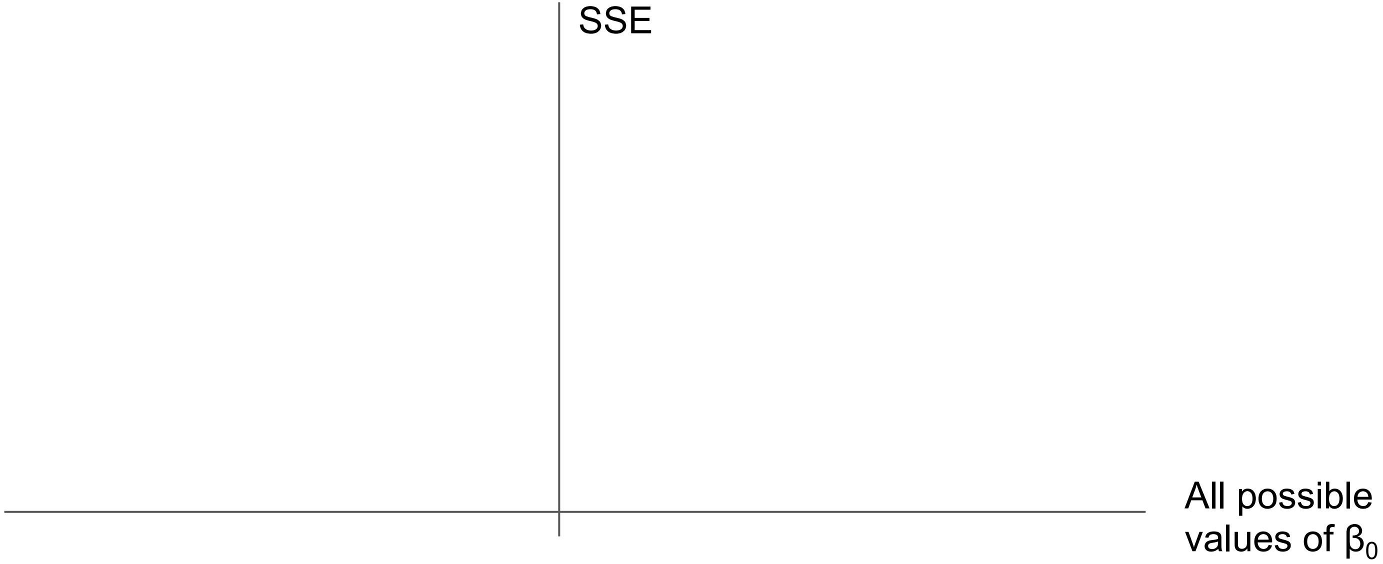
Minimizing with respect to the sum of squared errors:

$$Y = \beta_0 + e$$

We need to find the value of β_0 that will give us the smallest sum of squared errors (SSE).

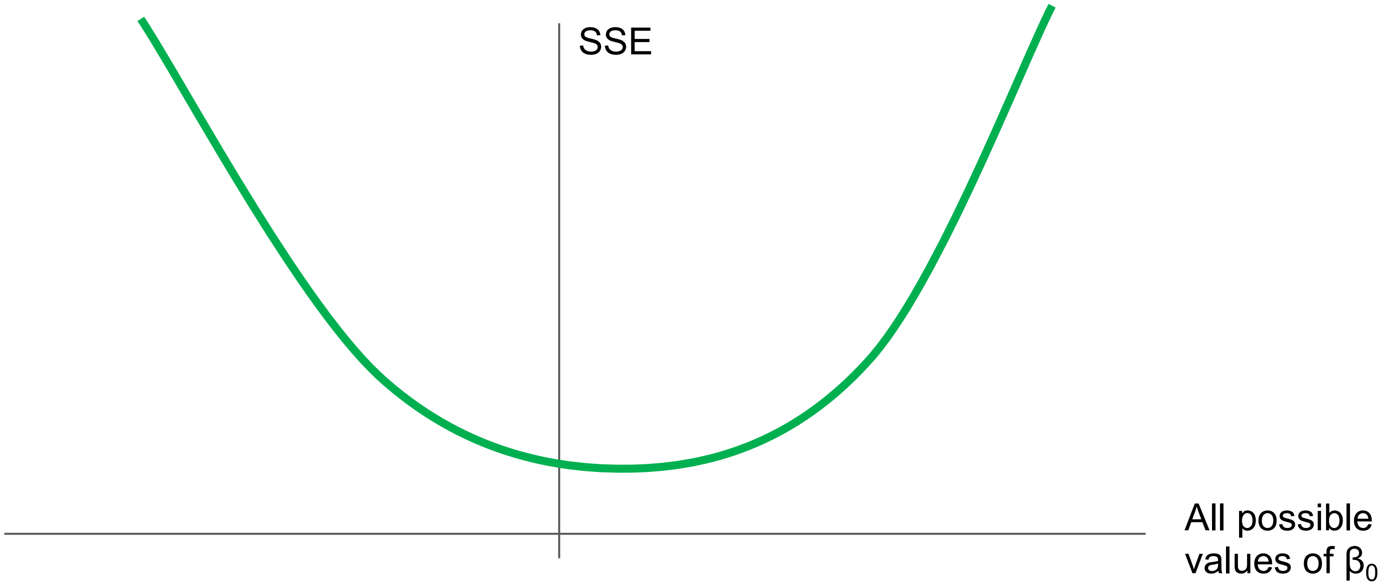
Minimizing with respect to the sum of squared errors: $Y = \beta_0 + e$

We need to find the value of β_0 that will give us the smallest sum of squared errors (SSE).



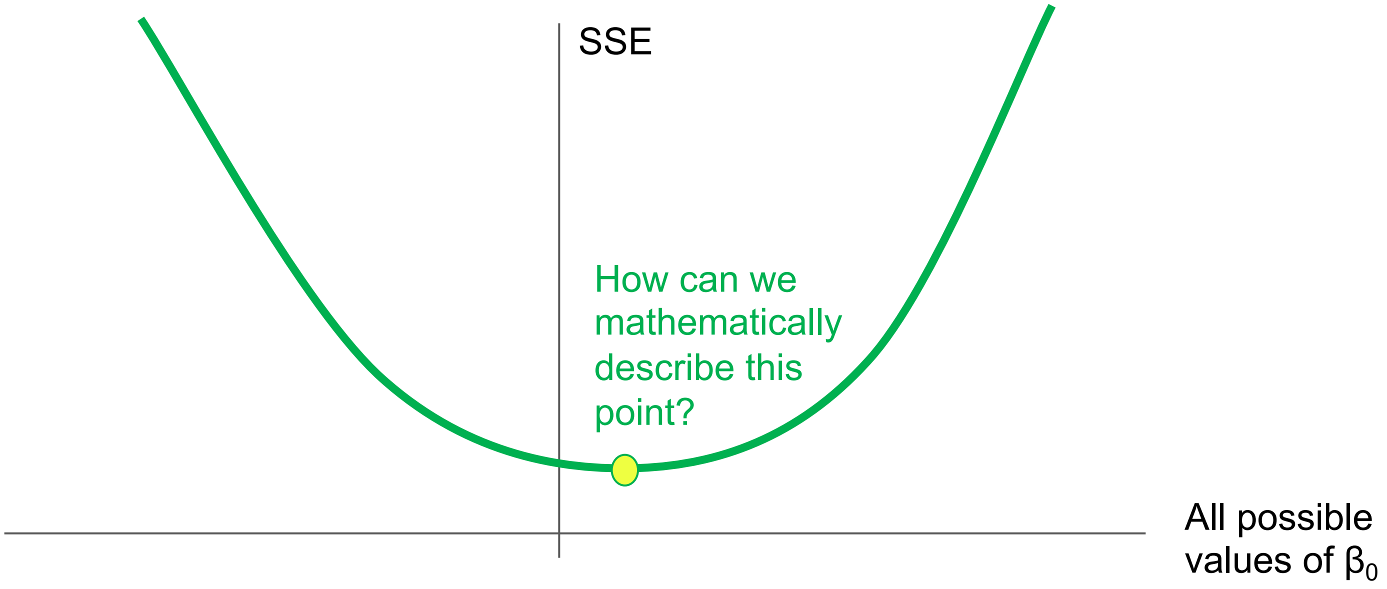
Minimizing with respect to the sum of squared errors: $Y = \beta_0 + e$

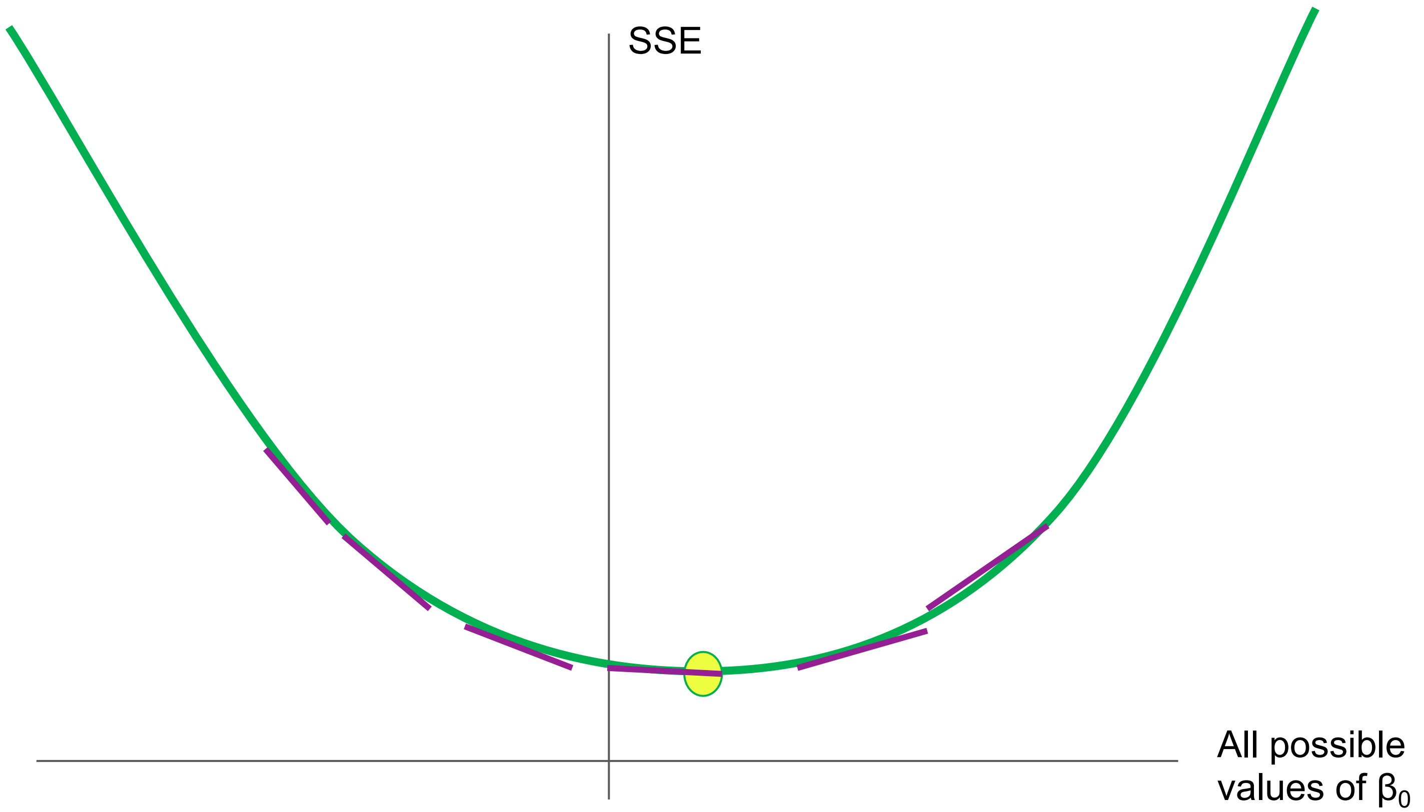
We need to find the value of β_0 that will give us the smallest sum of squared errors (SSE).



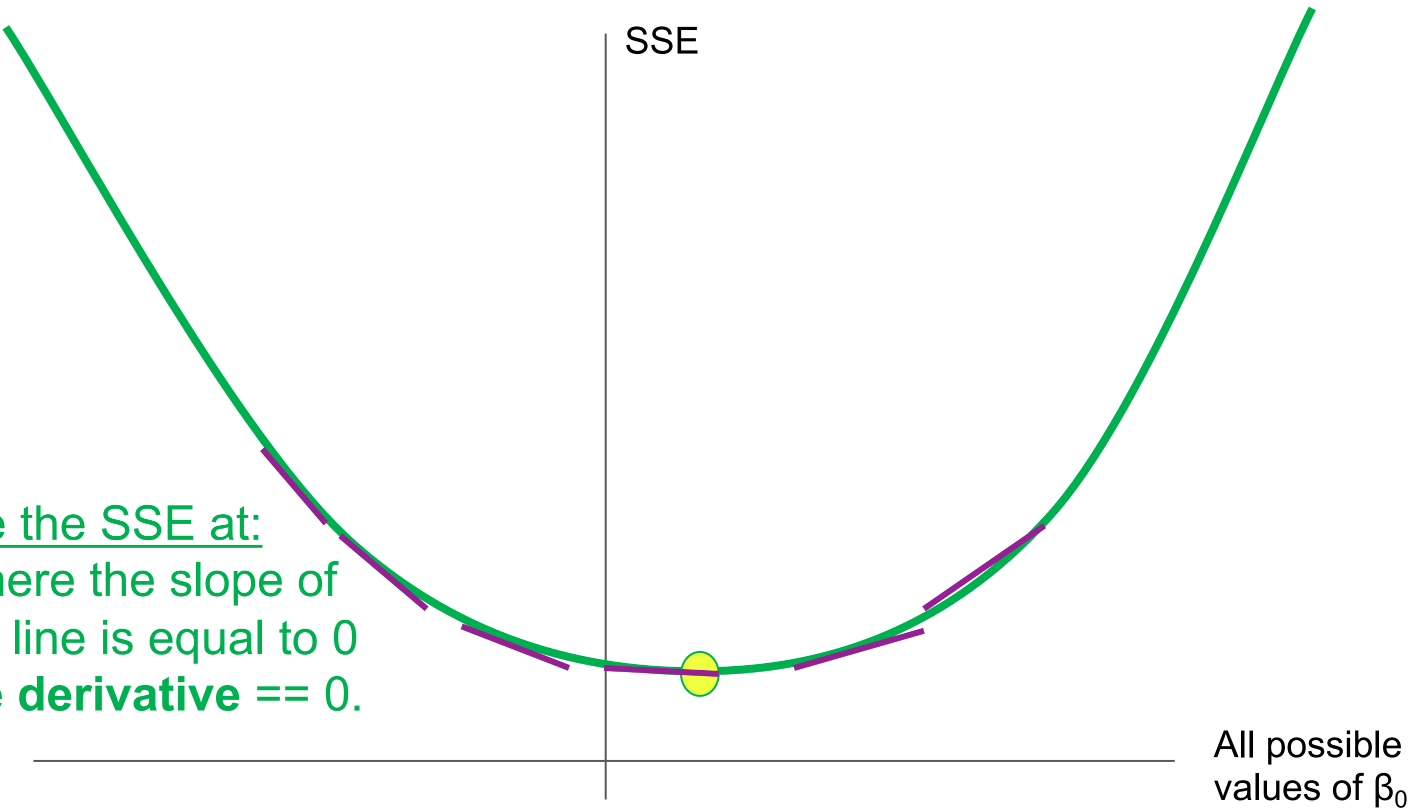
Minimizing with respect to the sum of squared errors: $Y = \beta_0 + e$

We need to find the value of β_0 that will give us the smallest sum of squared errors (SSE).




SSE
All possible
values of β_0

We minimize the SSE at:
The point where the slope of the [tangent] line is equal to 0
→ where the **derivative** == 0.

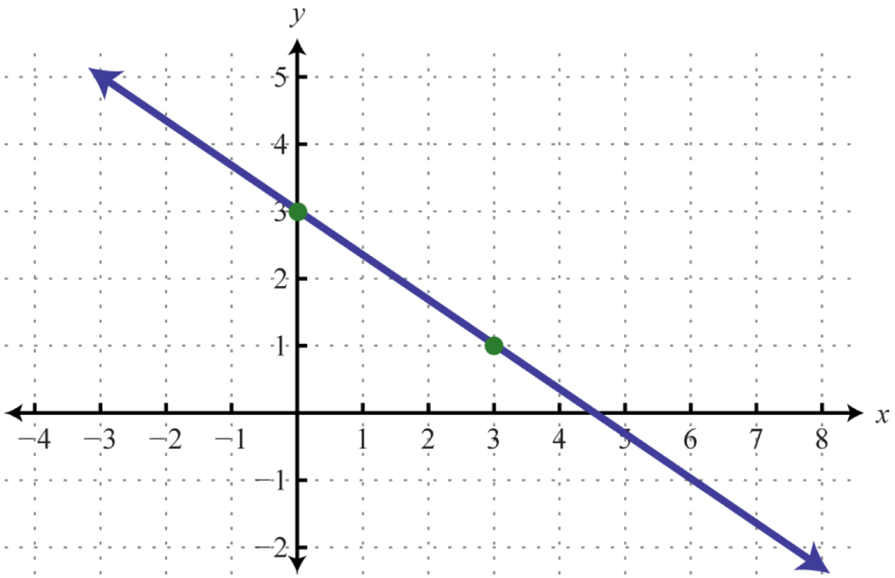


What is a derivative?

Essentially, the **rate of change** of a function at a given point.

What is a derivative?

Essentially, the **rate of change** of a function at a given point.



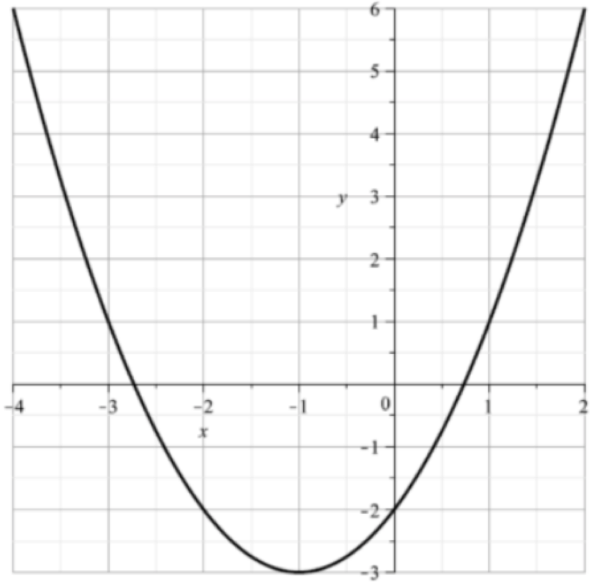
1. Write the equation of this function.

2. Where is the derivative of this function the largest?

3. Where is the derivative of this function the smallest?

What is a derivative?

Essentially, the **rate of change** of a function at a given point.

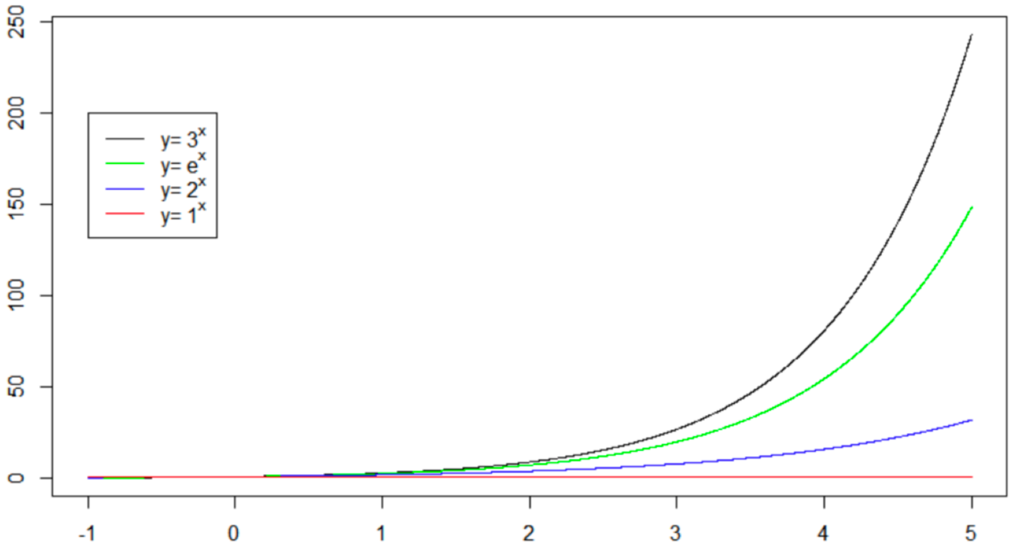


1. Where is the derivative of this function the largest?

2. Where is the derivative of this function the smallest?

What is a derivative?

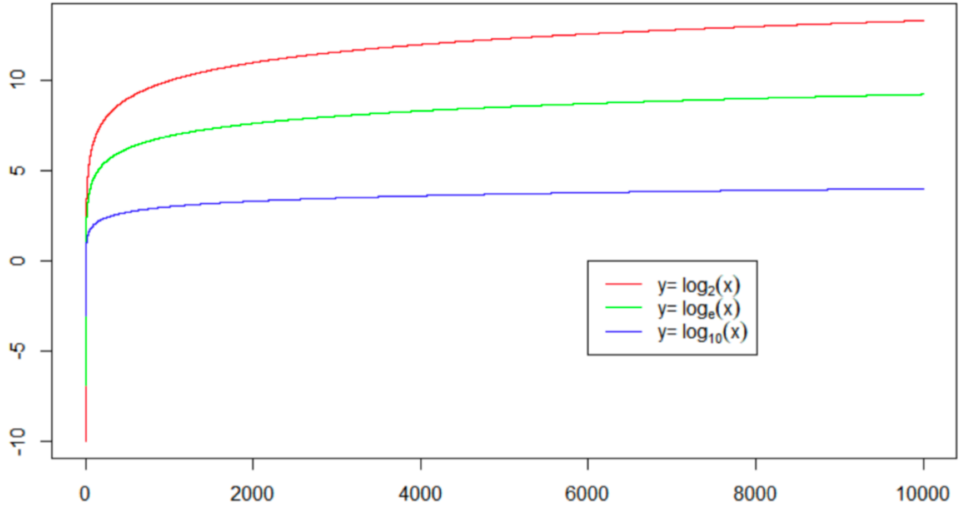
Essentially, the **rate of change** of a function at a given point.



- Where is the derivative of this function the largest?
- Where is the derivative of this function the smallest?

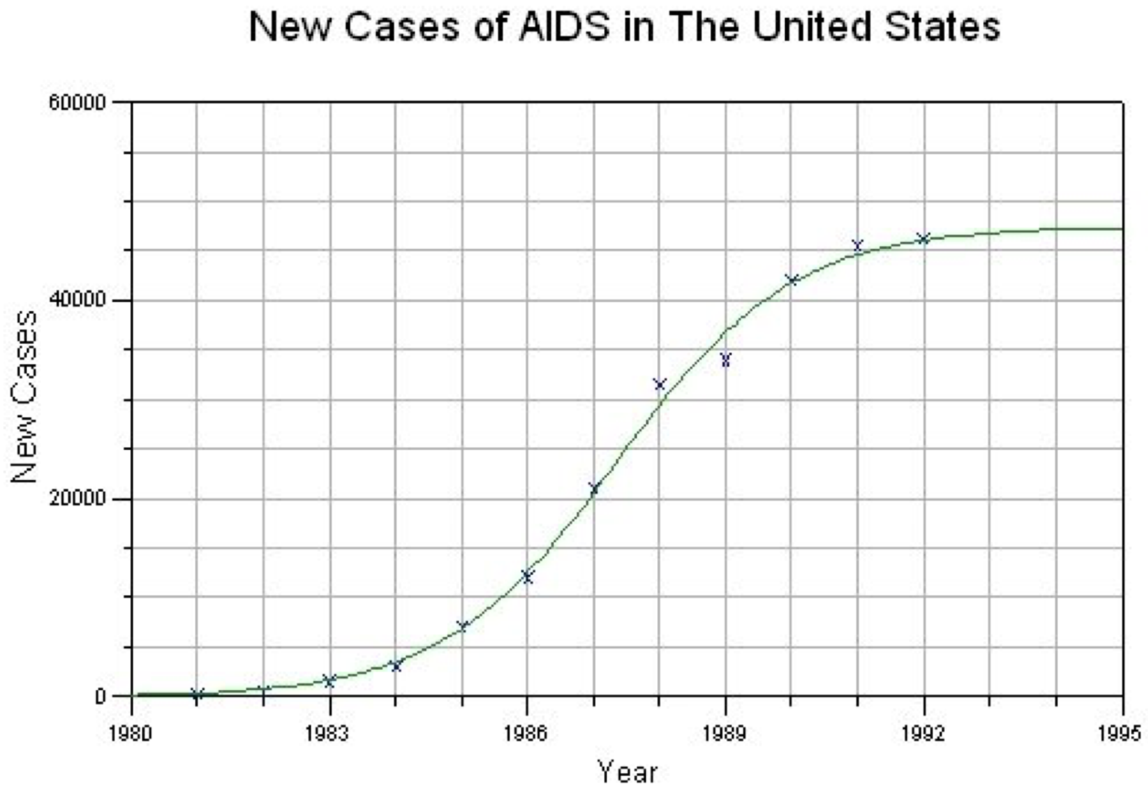
What is a derivative?

Essentially, the **rate of change** of a function at a given point.



1. Where is the derivative of this function the largest?

2. Where is the derivative of this function the smallest?



1. Where is the derivative of this function the largest?

2. Where is the derivative of this function the smallest?

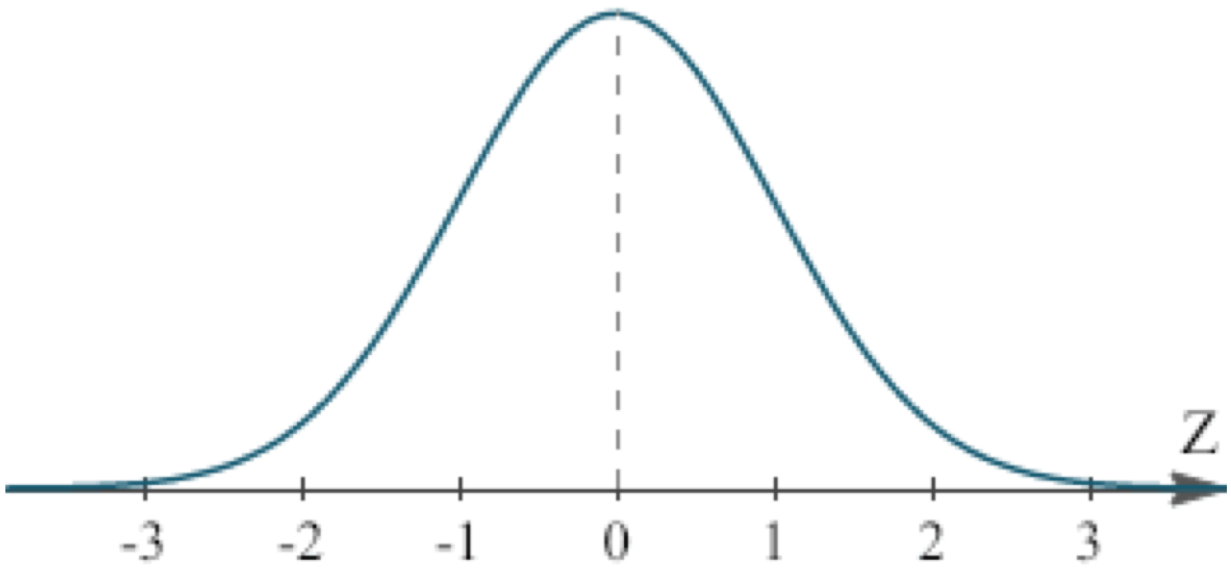
3. What does this tell us substantively?

4. Based on this graph, what would we expect to happen in the year 2015? **Why?**

5. What does the total area under the curve represent? **Why might we want to know this?**

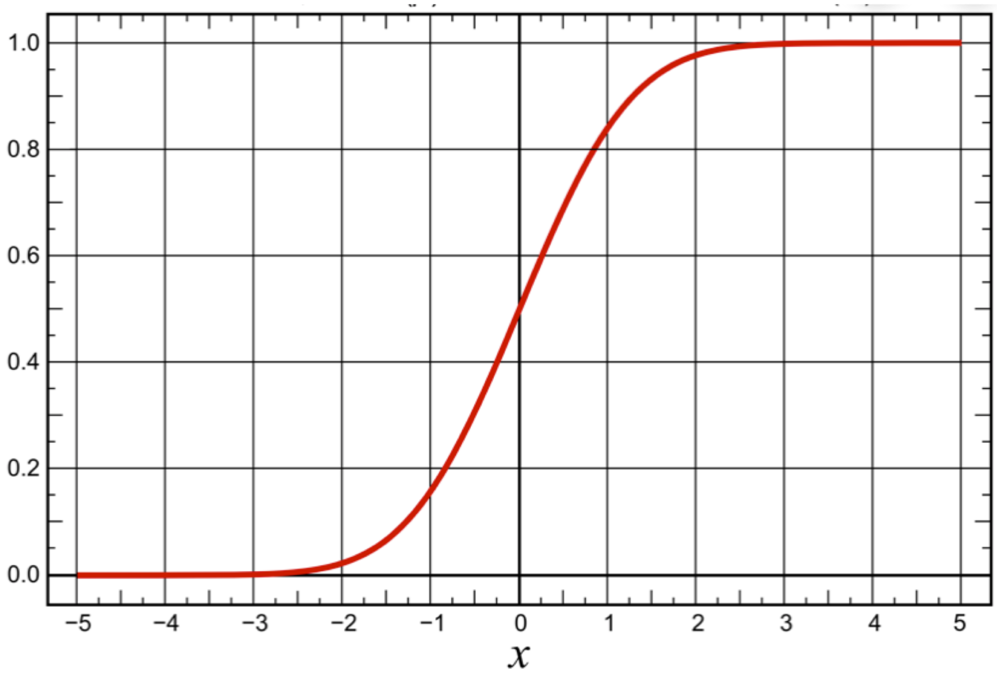
Key Calculus Terms:

- **Derivative**: The rate of change of a function
- **Limit**: The value a function approaches
- **Integral**: The area under a function



1. Talk about the limits of this function. What do they tell us?

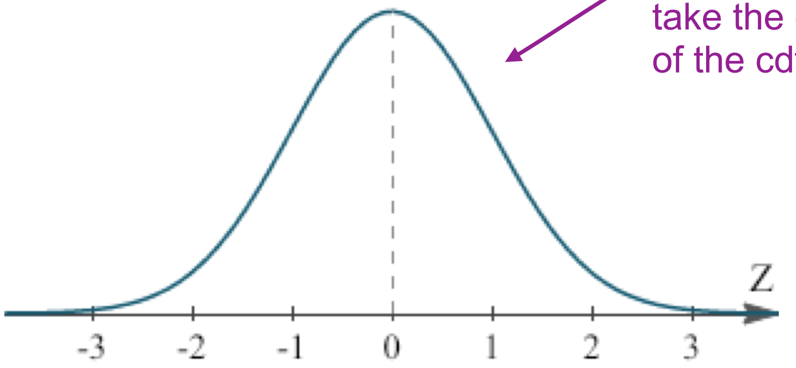
2. Talk about the integrals of this function (*assume that the total area under the curve is equal to 1 by definition*). What do they tell us?



1. Talk about the limits of this function. What do they tell us?

2. Talk about the derivatives of this function. What do they tell us?

Probability density function (pdf)

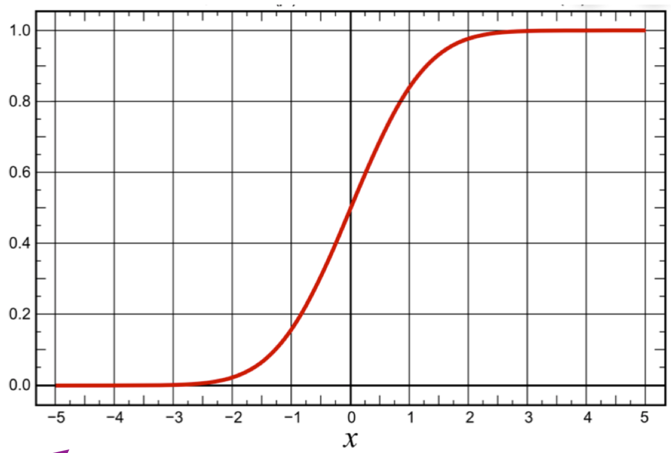


take the derivative of the cdf

take the integral of the pdf

- Area under the curve sums to 1
- Y-values represent the probability of getting that ***exact*** x-value

Cumulative density function (cdf)



- The limit of the function as $x \to \infty$ is 1
- Y-values represent the probability of getting that x-value ***or lower.***

Exit ticket

1. Fill out the post-bootcamp survey!