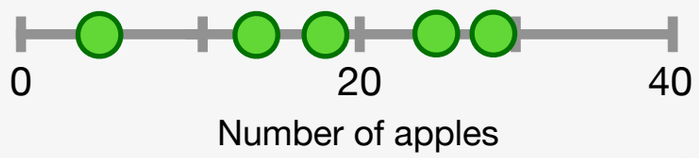
Appendix B: The Mean, Variance, and Standard Deviation

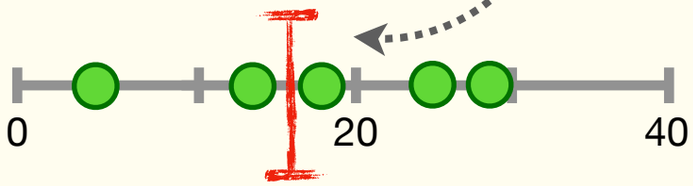
14 Instead of calculating **Population Parameters**, we *estimate* them from a relatively small number of measurements.



15 *Estimating* the **Population Mean** is super easy: we just calculate the average of the measurements we collected…

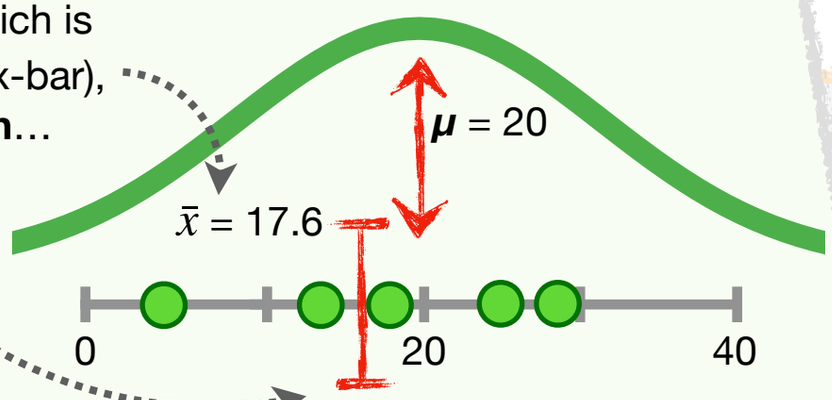
…and when we do the math, we get **17.6**.

$$\textbf{Estimated Mean} = \frac{\text{Sum of Measurements}}{\text{Number of Measurements}} = \frac{3 + 13 + 19 + 24 + 29}{5} = 17.6$$



16 **NOTE:** The **Estimated Mean**, which is often denoted with the symbol $\bar{x}$ (x-bar), is also called the **Sample Mean**…

…and due to the relatively small number of measurements used to calculate the **Estimated Mean**, it's different from the **Population Mean**.



A lot of **Statistics** is dedicated to quantifying and compensating for the differences between **Population Parameters**, like the **Mean** and **Variance**, and their *estimated* counterparts.

17 Now that we have an **Estimated Mean**, we can calculate an **Estimated Variance** and **Standard Deviation**. However, we have to compensate for the fact that we only have an **Estimated Mean**, which will almost certainly be different from the **Population Mean**.

18 Thus, when we calculate the **Estimated Variance** and **Standard Deviation** using the **Estimated Mean**…

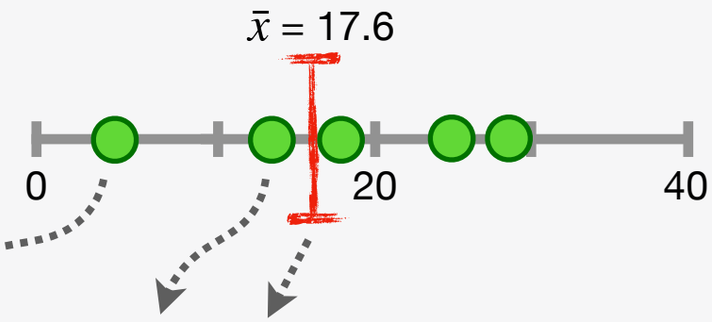
…we compensate for the difference between the **Population Mean** and the **Estimated Mean** by dividing by number of measurements minus **1**, ***n* - 1**, rather than ***n***.

$$\textbf{Estimated Variance} = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$\textbf{Estimated Standard Deviation} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Appendix B: The Mean, Variance, and Standard Deviation

19 Now when we plug the data into the equation for the **Estimated Variance**…



NOTE: If we had divided by ***n***, instead of ***n* - 1**, we would have gotten **81.4**, which is a significant *underestimate* of the true **Population Variance**, **100**.

$$\textbf{Estimated Variance} = \frac{\sum (x - \bar{x})^2}{n - 1} = \frac{(3 - 17.6)^2 + (13 - 17.6)^2 + (19 - 17.6)^2 + (24 - 17.6)^2 + (29 - 17.6)^2}{5 - 1} = 101.8$$

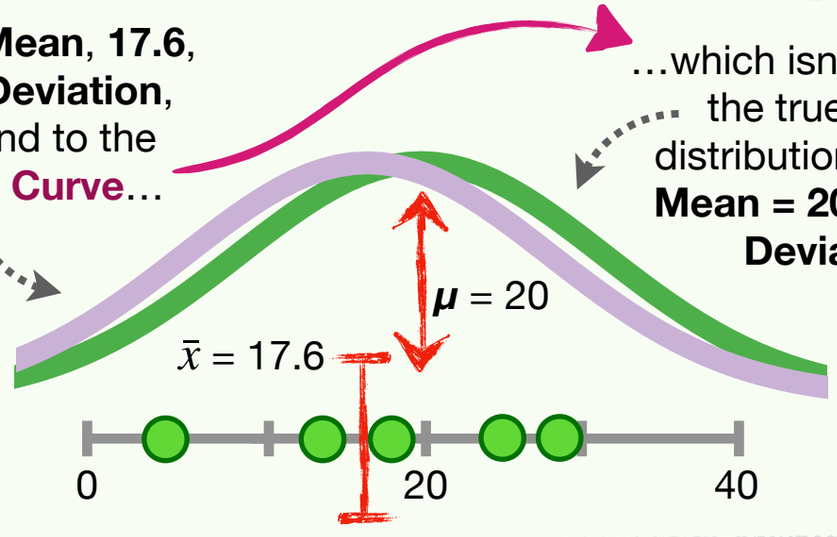
…we get **101.8**, which is a pretty good estimate of the **Population Variance**, which, as we saw earlier, is **100**.

20 Lastly, the **Estimated Standard Deviation** is just the square root of the **Estimated Variance**…

…so, in this example, the **Estimated Standard Deviation** is **10.1**. Again, this is relatively close to the **Population** value we calculated earlier.

$$\textbf{Estimated Standard Deviation} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\textbf{Estimated Variance}} = \sqrt{101.8} = 10.1$$

21 The **Estimated Mean**, **17.6**, and **Standard Deviation**, **10.1**, correspond to the **purple Normal Curve**…



…which isn't too far off from the true **Population** distribution in **green**, with **Mean = 20** and **Standard Deviation = 10**.

TRIPLE BAM!!!

Appendix C:

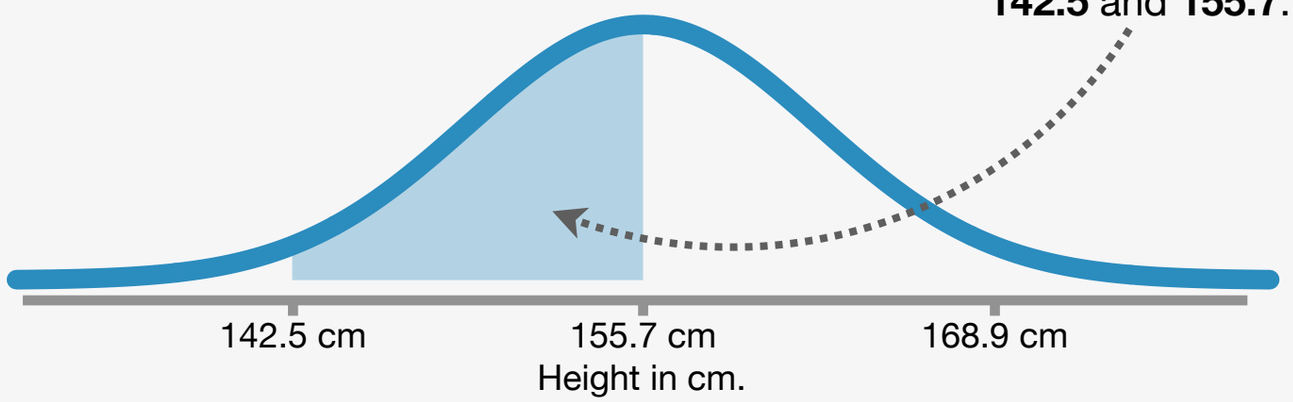
Computer Commands for Calculating Probabilities with Continuous Probability Distributions

Appendix C: Computer Commands for Calculating Probabilities with Continuous Probability Distributions

1 Given this **Normal Distribution**, with **Mean = 155.7** and **Standard Deviation = 6.6**…

…let's talk about ways we can get a computer to calculate the area under the curve between **142.5** and **155.7**.

However, before we get into the specific commands that we can use in **Google Sheets**, **Microsoft Excel**, and ***R***, we need to talk about **Cumulative Distribution Functions**.

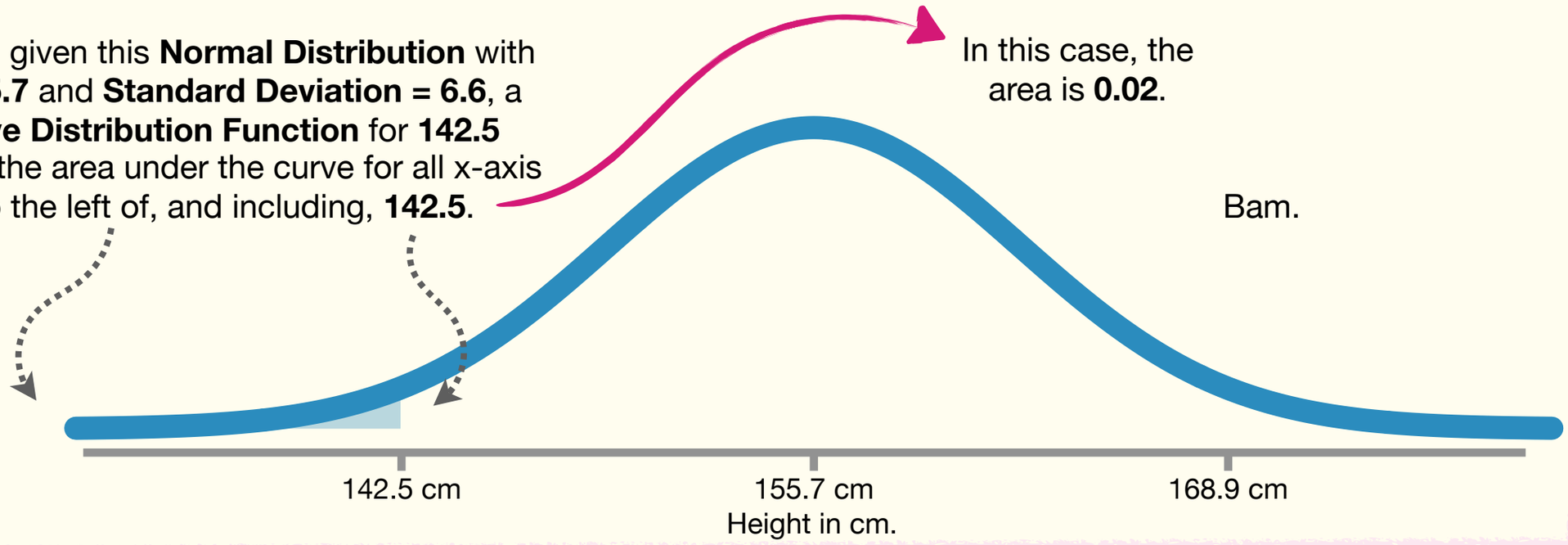


2 A **Cumulative Distribution Function** (**CDF**) simply tells us the area under the curve up to a specific point.

For example, given this **Normal Distribution** with **Mean = 155.7** and **Standard Deviation = 6.6**, a **Cumulative Distribution Function** for **142.5** would tell us the area under the curve for all x-axis values to the left of, and including, **142.5**.

In this case, the area is **0.02**.

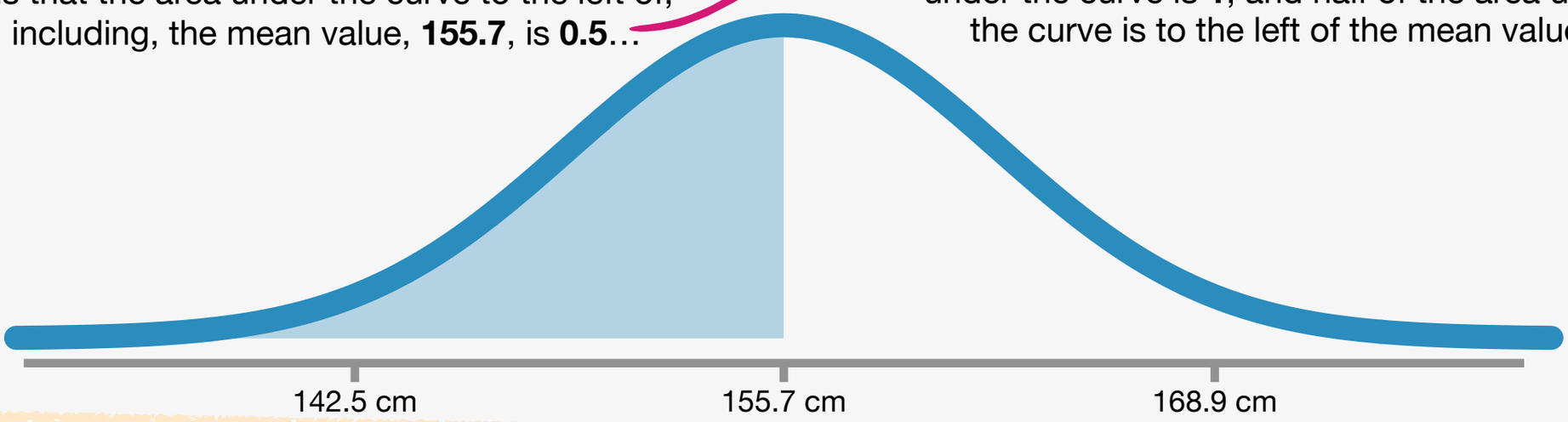
Bam.



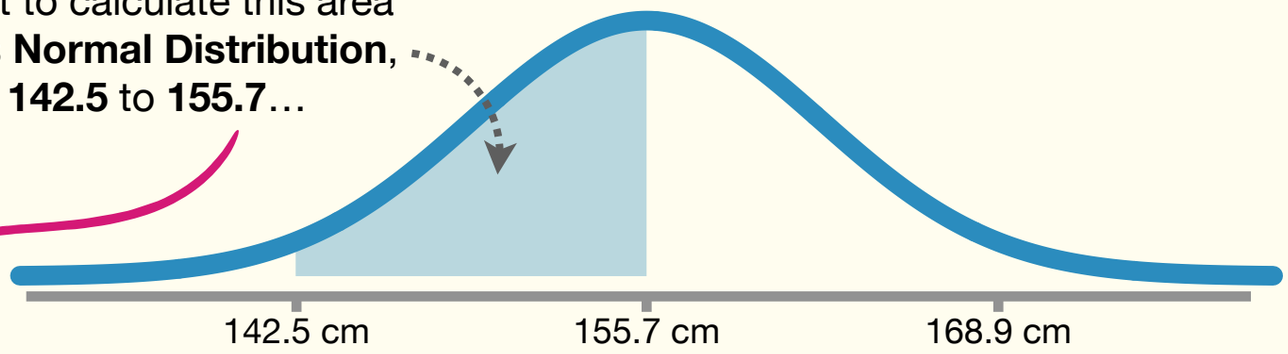
Appendix C: Computer Commands for Calculating Probabilities with Continuous Probability Distributions

3 Likewise, the **Cumulative Distribution Function** tells us that the area under the curve to the left of, and including, the mean value, **155.7**, is **0.5**…

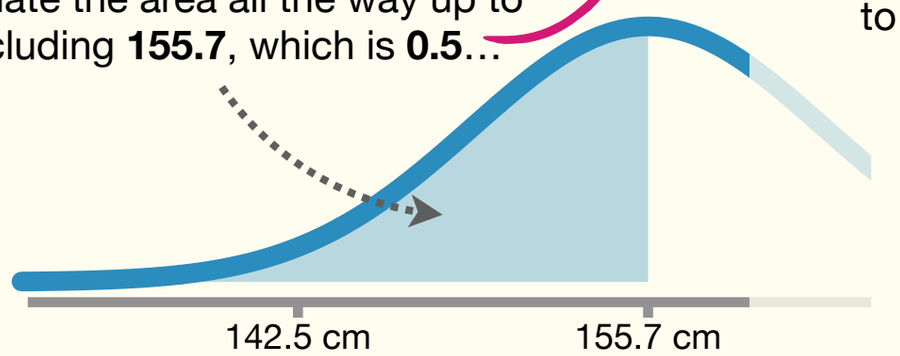
…and that makes sense because the total area under the curve is **1**, and half of the area under the curve is to the left of the mean value.



4 If we want to calculate this area under this **Normal Distribution**, from **142.5** to **155.7**…



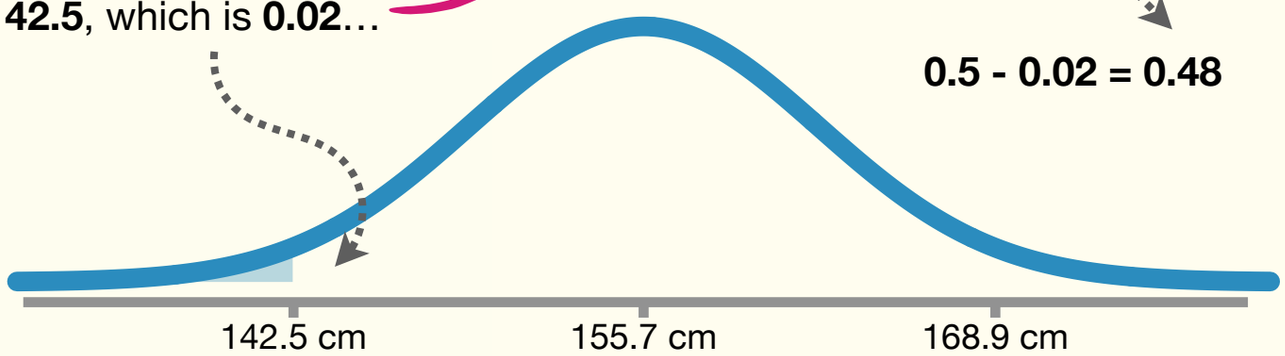
…then we start by using a **Cumulative Distribution Function** to calculate the area all the way up to and including **155.7**, which is **0.5**…



…and we subtract the area under the curve up to **142.5**, which is **0.02**…

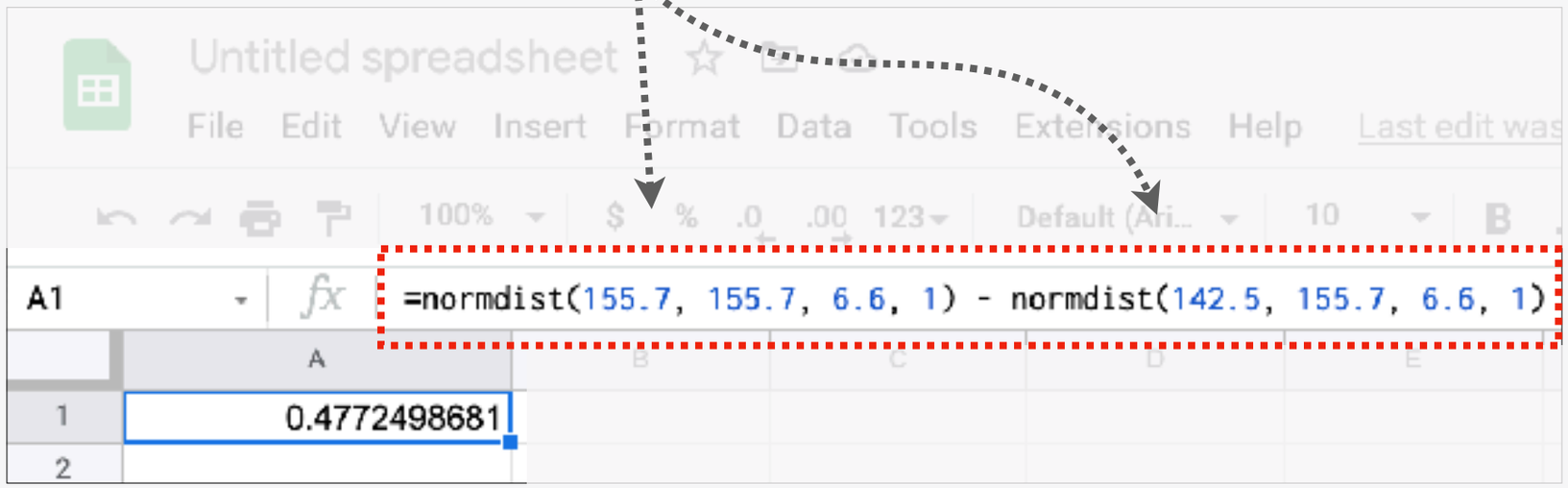
…and when we do the math, we get **0.48**.

0.5 - 0.02 = 0.48



Appendix C: Computer Commands for Calculating Probabilities with Continuous Probability Distributions

5 Now, if we want to do those calculations with **Google Sheets** or **Microsoft Excel**, we use the **NORMDIST()** function.



6 The **NORMDIST()** function takes **4** arguments:

```
normdist( x-axis value, mean, standard deviation, use CDF)
```

…the x-axis value that we want to calculate the area under the curve to the left of, and including; in our example, this means we set this to either **155.7** or **142.5**…

…the **Mean** of the **Normal Distribution**, which in this example is **155.7**…

…the **Standard Deviation**, which in this example is **6.6**…

…and either **0** or **1**, depending on whether or not we want to use the **Cumulative Distribution Function** (**CDF**). In this example, we set it to **1** because we want to use the **CDF**.

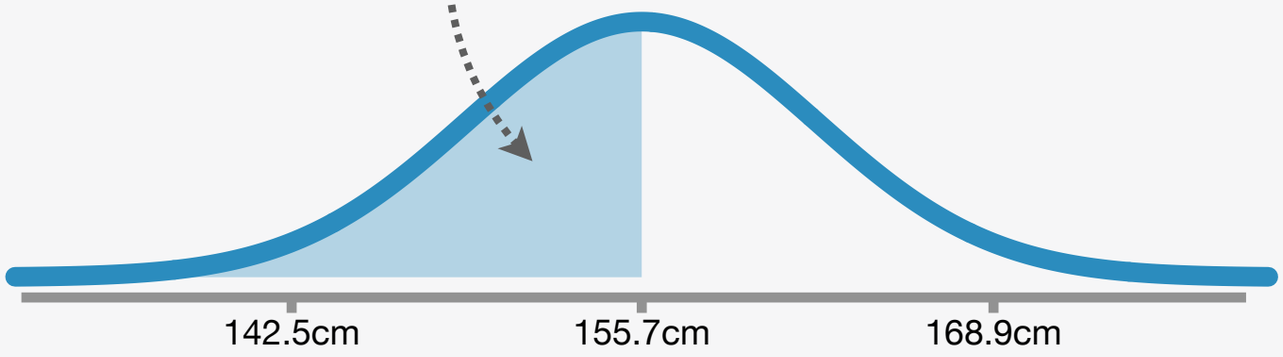
Appendix C: Computer Commands for Calculating Probabilities with Continuous Probability Distributions

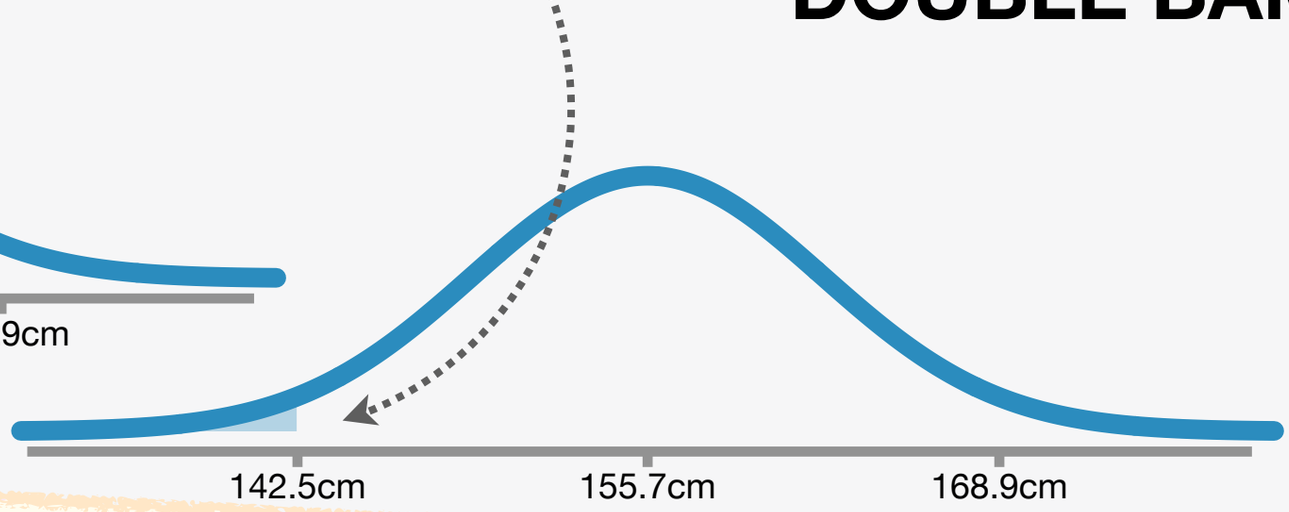
7 Putting everything together, we get **0.5 - 0.02 = 0.48**.

Gentle Reminder about the arguments for the **NORMDIST()** function:

```
normdist( x-axis value, mean, standard deviation, use CDF)
```

```
normdist( 155.7, 155.7, 6.6, 1) - normdist( 142.5, 155.7, 6.6, 1) = 0.48
```





DOUBLE BAM!!

8 In the programming language called ***R***, we can get the same result using the **pnorm()** function, which is just like **NORMDIST()**, except we don't need to specify that we want to use a **CDF**.

TRIPLE BAM!!!

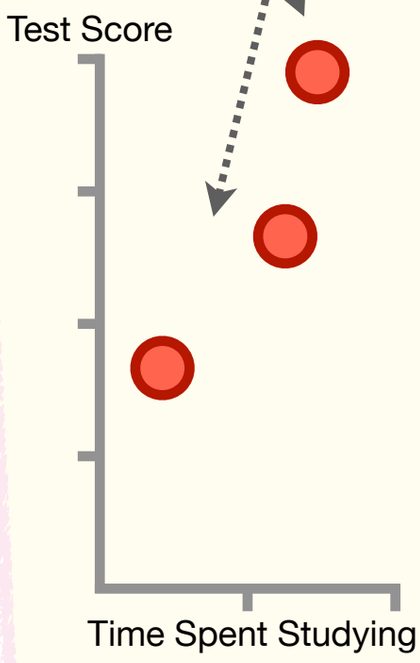
```
pnorm(155.7, mean=155.7, sd=6.6) - pnorm(142.5, mean=155.7, sd=6.6)

                                 = 0.48
```

Appendix D:

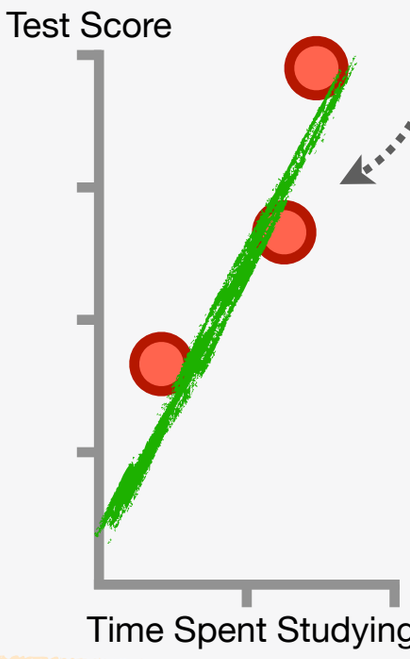
The Main Ideas of Derivatives

Appendix D: The Main Ideas of Derivatives

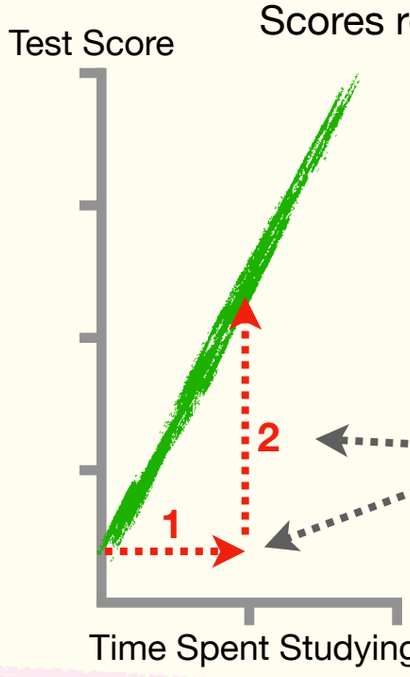
1 Imagine we collected Test Scores and Time Spent Studying from **3** people…



2 …and we fit a **straight line** to the data.



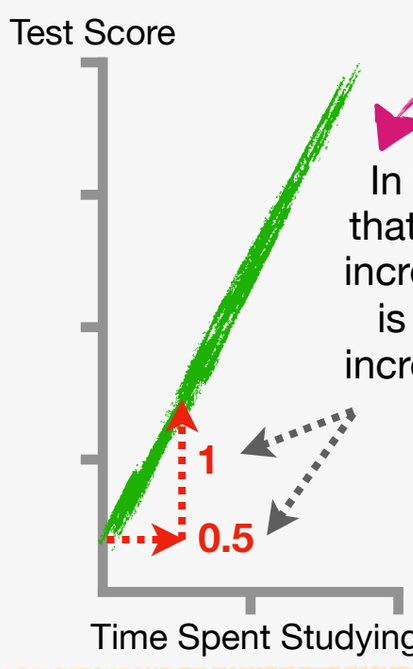
3 One way to understand the relationship between Test Scores and Time Spent Studying is to look at changes in Test Scores relative to changes in Time Spend Studying.



In this case, we see that for every unit increase in Time, there is a two-unit increase in Test Score.

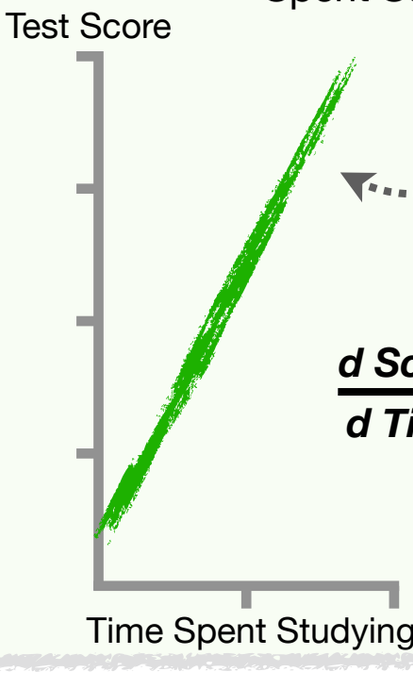
In other words, we can say we go up **2** for every **1** we go over.

4 **NOTE:** The relationship "**2** up for every **1** over" holds even if we only go over **1/2** unit.



In this case, we see that for every half-unit increase in Time, there is a **2 x 0.5 = 1** unit increase in Test Score.

5 Because the "**2** up for every **1** over" relationship holds no matter how small a value for Time Spent Studying, we say that the **Derivative** of this **straight line**…



…is the change in Score (***d Score***) relative to the change in Time (***d Time***), which is is **2**.

$$\frac{d\ Score}{d\ Time} = 2$$

Now let's talk about how the **Derivative** is related to the **straight line**.

Appendix D: The Main Ideas of Derivatives

6 Here is the equation for the **straight line**.

The slope is **2**…

Test Score

Score = 1/2 + (2 x Time)

…and the y-axis intercept is **1/2**.

Time Spent Studying

7 Thus, we see that the slope, **2**, is equal to the **Derivative**, **2**, and both tell us to go "**2** up for every **1** over."

Test Score

Score = 1/2 + (2 x Time)

$$\frac{d\ Score}{d\ Time} = 2$$

Time Spent Studying

8 Now let's stop talking about Studying and talk about Eating. This line has a slope of **3**, and, regardless of how small a value for Time Spent Eating, our Fullness goes up **3** times that amount.

Fullness

Thus, the **Derivative**, the change in Fullness relative to the change in Time, is **3**.

$$\frac{d\ Fullness}{d\ Time} = 3$$

Time Spent Eating

9 When the slope is **0**, and the y-axis value never changes, regardless of the x-axis value…

…then the **Derivative**, the change in the y-axis value relative to the change in the x-axis value, is **0**.

Height of the Empire State Building

$$\frac{d\ Height}{d\ Occupants} = 0$$

Number of Occupants

10 When the straight line is vertical, and the x-axis value never changes, then the **Derivative** is undefined. This is because it's impossible to measure the change in the y-axis value relative to the change in the x-axis value if the x-axis value never changes.

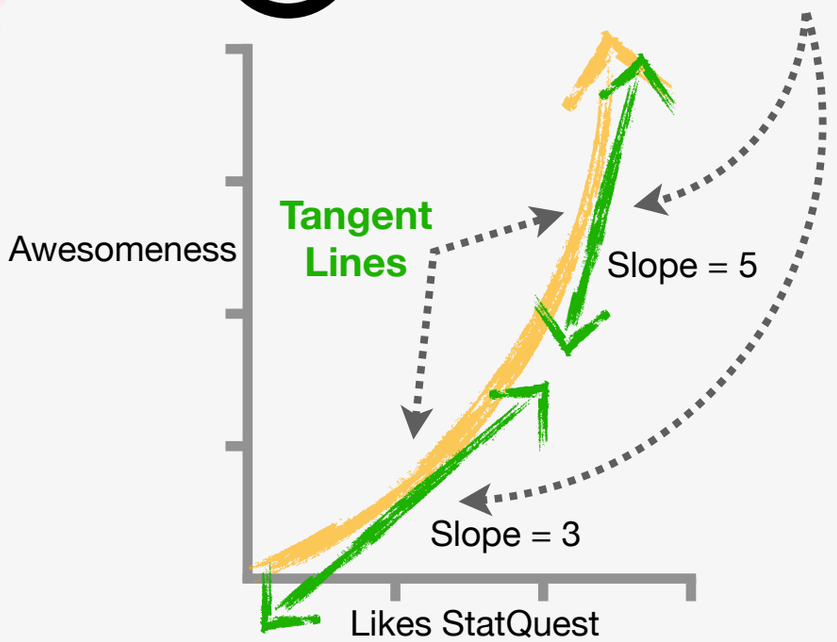
y-axis

x-axis

Appendix D: The Main Ideas of Derivatives

11 Lastly, when we have a **curved line** instead of a straight line…

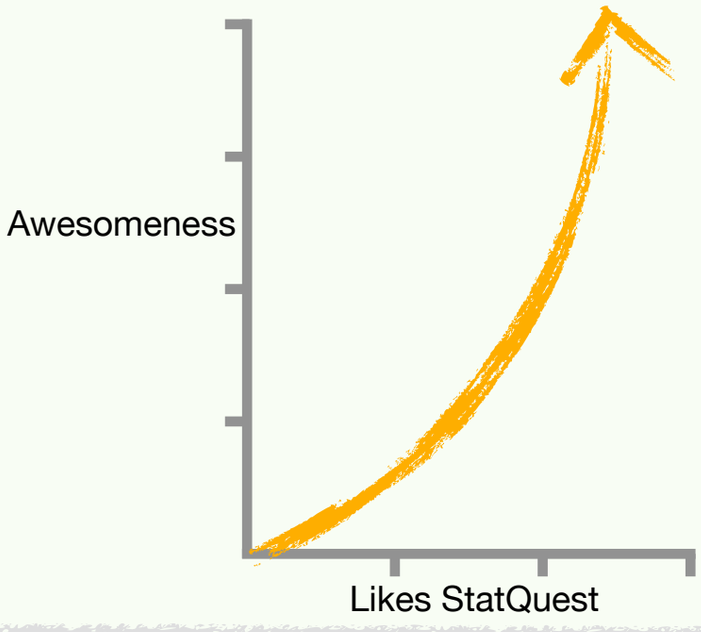
12 …the **Derivative** is the slope of any **straight line** that just barely touches the **curved line** at a single point.



Terminology Alert!!!

A straight line that touches the curve at a single point is called a **tangent line**.

13 Unfortunately, the **Derivatives** of **curved lines** are not as easy to determine as they are for **straight lines**.



However, the good news is that in machine learning, **99%** of the time we can find the **Derivative** of a curved line using **The Power Rule** (See **Appendix E**) and **The Chain Rule** (See **Appendix F**).

BAM!!!

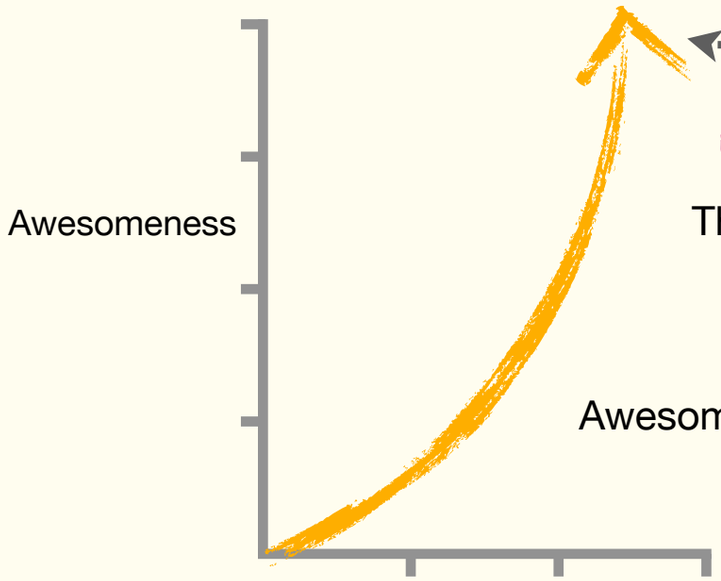
Appendix E:

The Power Rule

NOTE: This appendix assumes that you're already familiar with the concept of a derivative (**Appendix D**).

Appendix E: The Power Rule

1 In this example, we have a **parabola** that represents the relationship between Awesomeness and Likes StatQuest.



This is the equation for the **parabola**:

$$\text{Awesomeness} = (\text{Likes StatQuest})^2$$

2 We can calculate the derivative, the change in Awesomeness with respect to the change in Likes StatQuest…

$$\frac{d}{d\,\text{Likes StatQuest}}\,\text{Awesomeness}$$

…by first plugging in the equation for Awesomeness…

$$\frac{d}{d\,\text{Likes StatQuest}}\,(\text{Likes StatQuest})^2$$

…and then applying **The Power Rule**.

3 **The Power Rule** tells us to multiply Likes StatQuest by the power, which, in this case, is **2**…

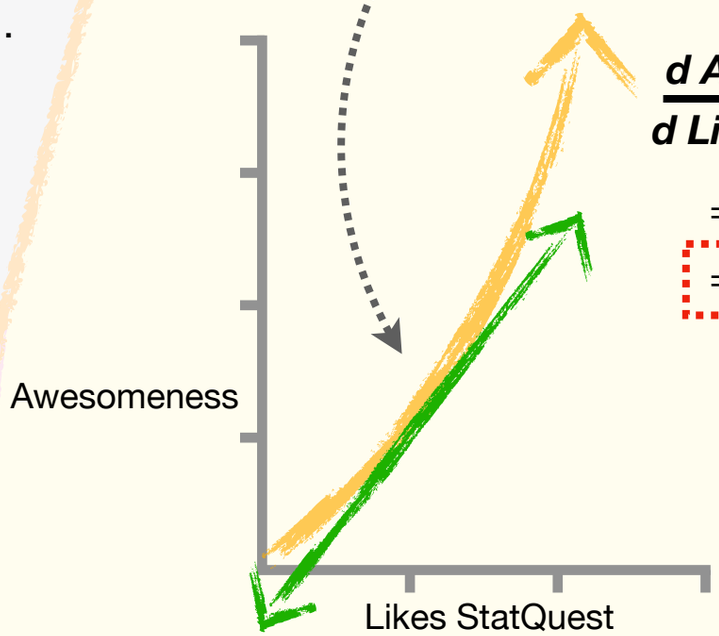
…and raise Likes StatQuest by the original power, **2**, minus **1**…

$$\frac{d}{d\,\text{Likes StatQuest}}\,(\text{Likes StatQuest})^2 = 2 \times \text{Likes StatQuest}^{2-1}$$

$$= 2 \times \text{Likes StatQuest}$$

…and since **2-1 = 1**, the derivative of Awesomeness with respect to Likes StatQuest is **2** times Likes StatQuest.

4 For example, when Likes StatQuest **= 1**, the derivative, the slope of the **tangent line** is **2**.



$$\frac{d\,\text{Awesomeness}}{d\,\text{Likes StatQuest}} =$$

$$= 2 \times \text{Likes StatQuest}$$

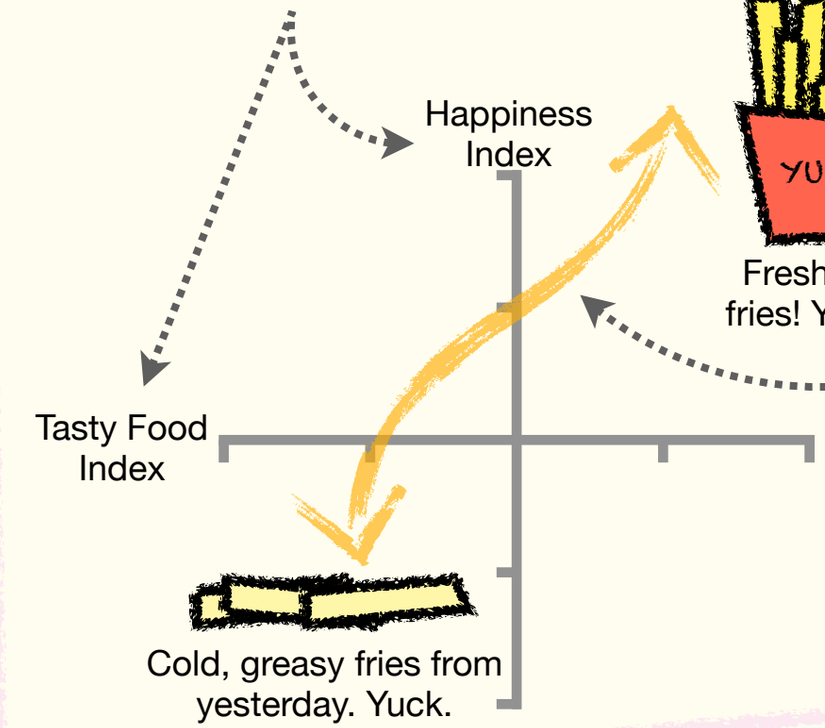
$$= 2 \times 1 = 2$$

BAM!!!

Now let's look at a fancier example!!!

Appendix E: The Power Rule

5 Here we have a graph of how Happy people are relative to how Tasty the food is…



…and this is the equation for the **squiggle**:

$$Happy = 1 + Tasty^3$$

6 We can calculate the derivative, the change in Happiness with respect to the change in Tastiness…

$$\frac{d\ Happy}{d\ Tasty} = \frac{d}{d\ Tasty} Happy$$

…by plugging in the equation for Happy…

$$\frac{d}{d\ Tasty}(1 + Tasty^3)$$

…and taking the derivative of each term in the equation.

$$\frac{d}{d\ Tasty}(1 + Tasty^3) = \frac{d}{d\ Tasty}1 + \frac{d}{d\ Tasty}Tasty^3$$

7 The constant value, **1**, doesn't change, regardless of the value for Tasty, so the derivative with respect to Tasty is **0**.

$$\frac{d}{d\ Tasty}1 = 0$$

The Power Rule tells us to multiply Tasty by the power, which, in this case is **3**…

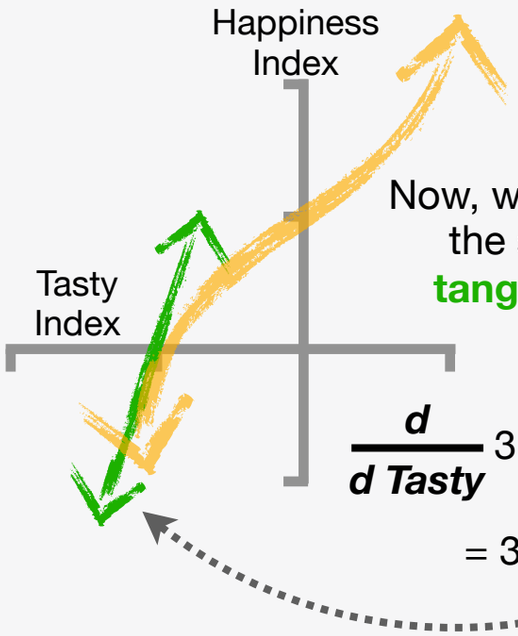
…and raise Tasty by the original power, **3**, minus **1**.

$$\frac{d}{d\ Tasty}Tasty^3 = 3 \times Tasty^{3-1}$$
$$= 3 \times Tasty^2$$

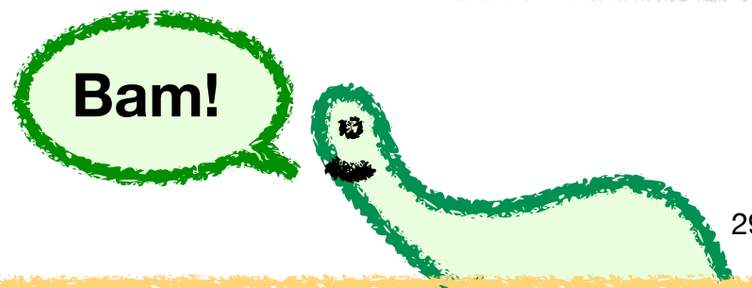
8 Lastly, we recombine both terms to get the final derivative.

$$\frac{d}{d\ Tasty}Happy = 0 + 3 \times Tasty^2$$
$$= 3 \times Tasty^2$$

9 Now, when Tasty **= -1**, the slope of the **tangent line** is **3**.



$$\frac{d}{d\ Tasty}3 \times Tasty^2$$
$$= 3 \times -1^2 = 3$$



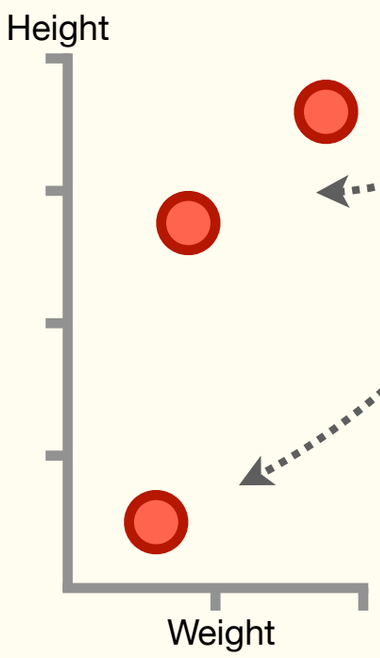
Appendix F:

The Chain Rule!!!

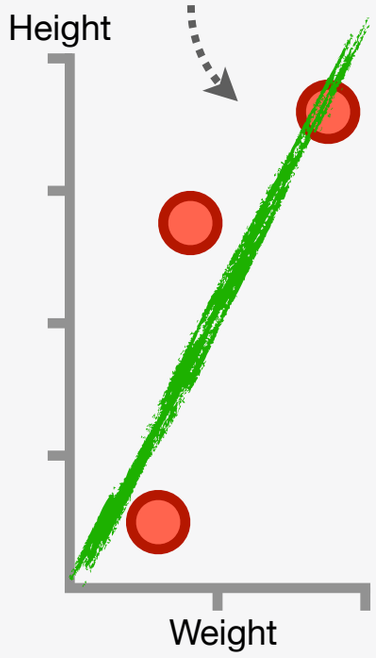
NOTE: This appendix assumes that you're already familiar with the concept of a derivative (**Appendix D**) and **The Power Rule** (**Appendix E**).

Appendix F: The Chain Rule

1 Here we have Weight and Height measurements from **3** people…

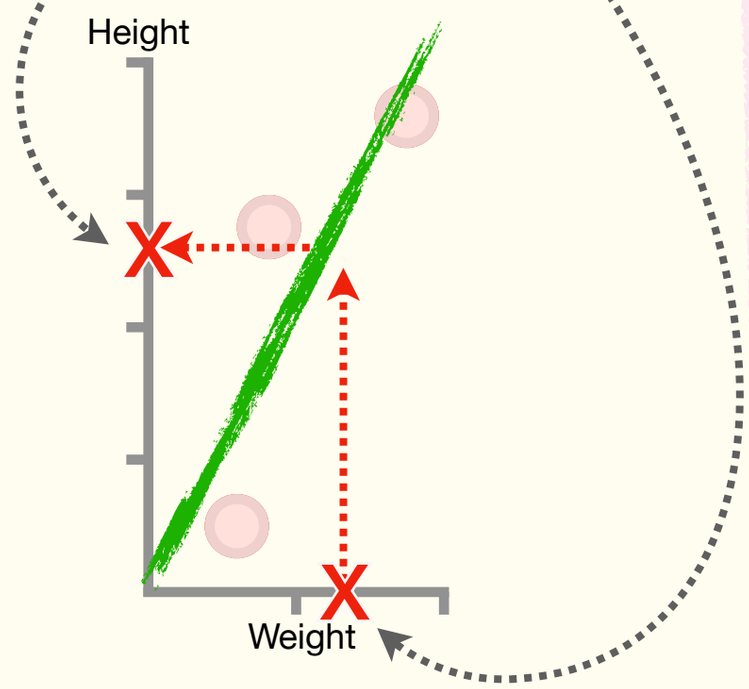


2 …and we fit a line to the data.

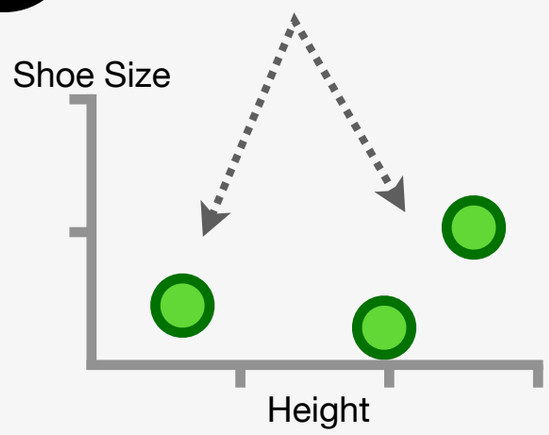


3 Now, if someone tells us that they weigh this much…

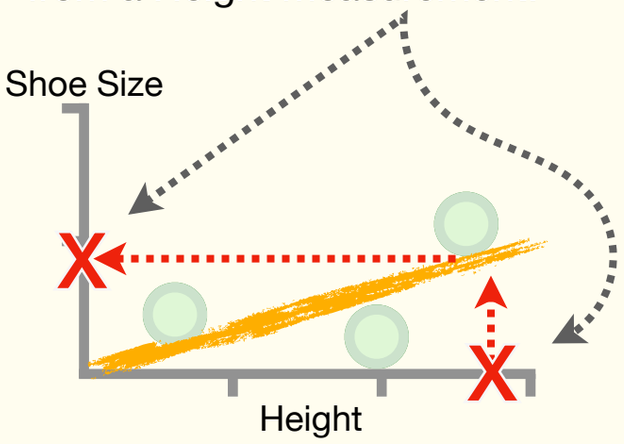
…then we can use the **green fitted line** to predict that they're this tall.

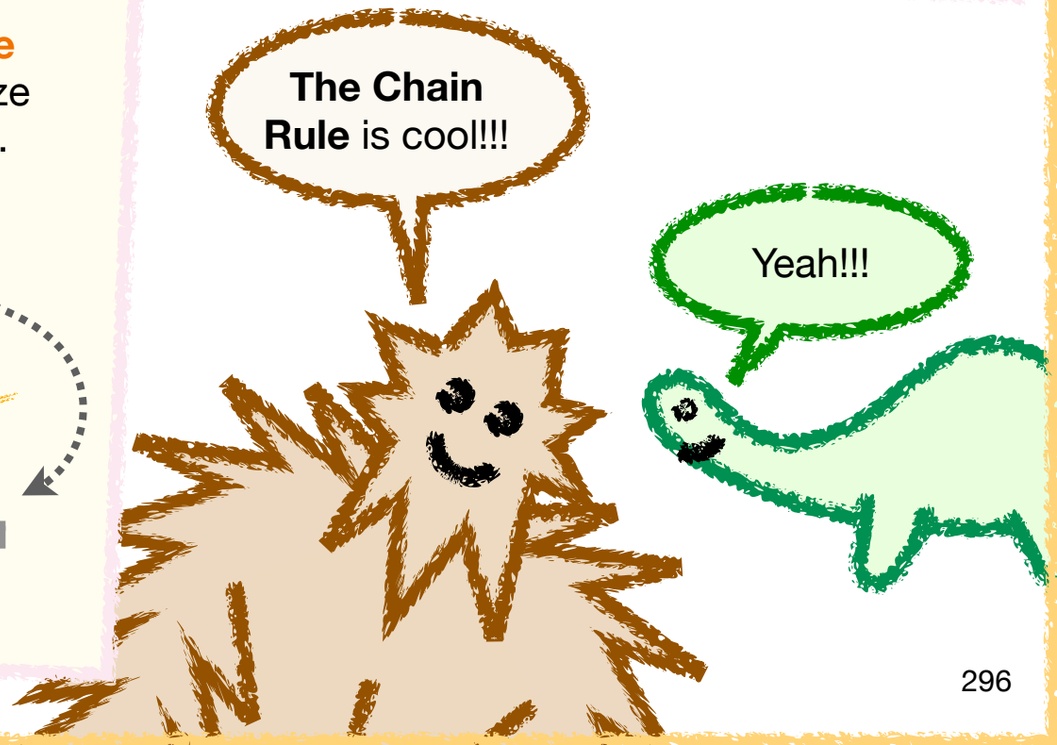


4 Here we have Height and Shoe Size measurements…



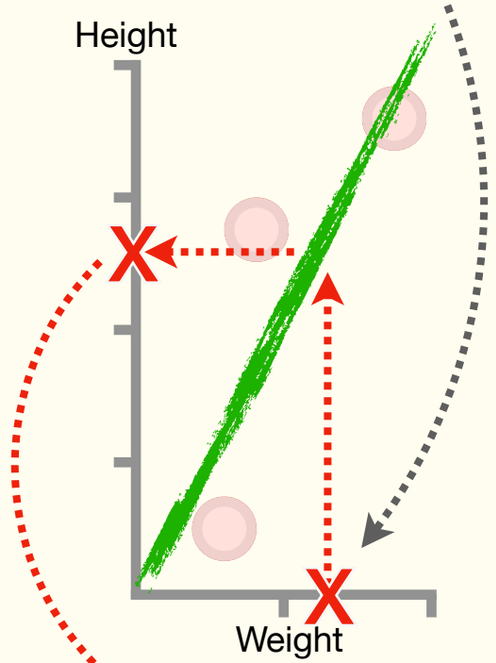
5 …and we can use an **orange fitted line** to predict Shoe Size from a Height measurement.





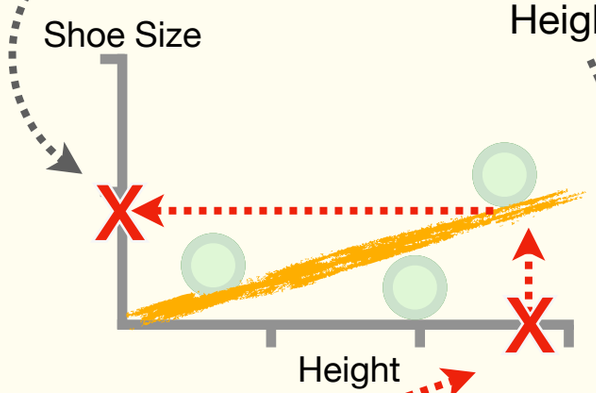
Appendix F: The Chain Rule

6 Now, if someone tells us that they weight this much…

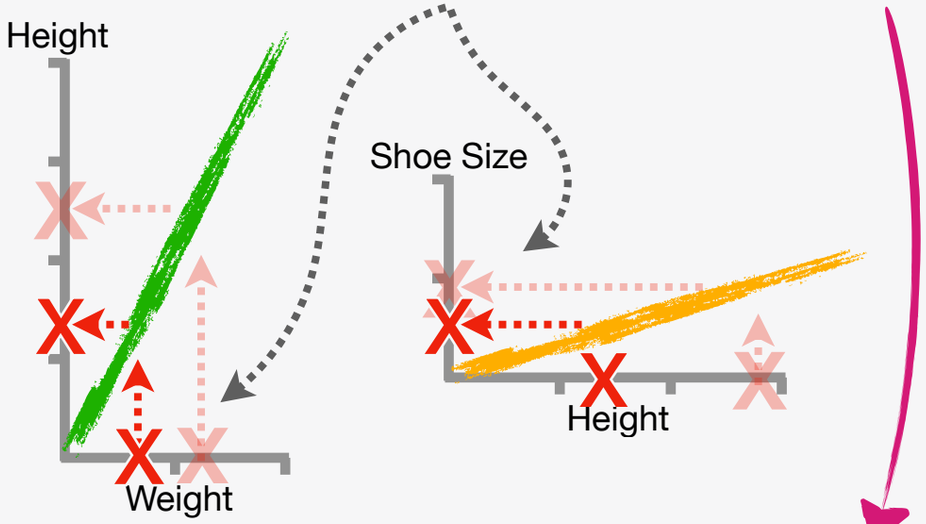


…then we can predict that this is their Shoe Size…

…because Weight and Shoe Size are connected by Height…

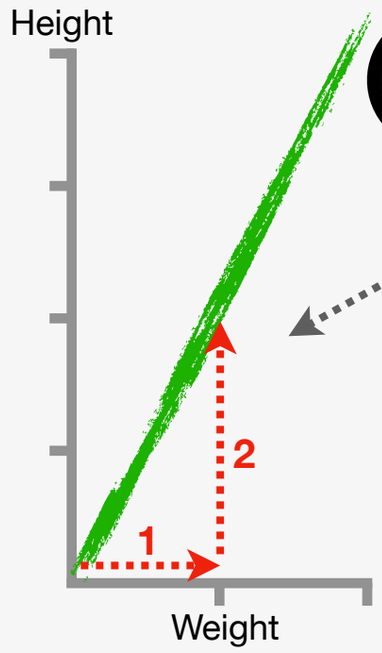


7 …and if we change the value for Weight, like make it smaller, then the value for Shoe Size changes. In this case, Shoe Size gets smaller, too.



Now, if we want to quantify how much Shoe Size changes when we change Weight, we need to calculate the derivative for Shoe Size with respect to Weight.

8 Because this **straight line** goes up **2** units for every **1** unit over, the slope and derivative are **2**…

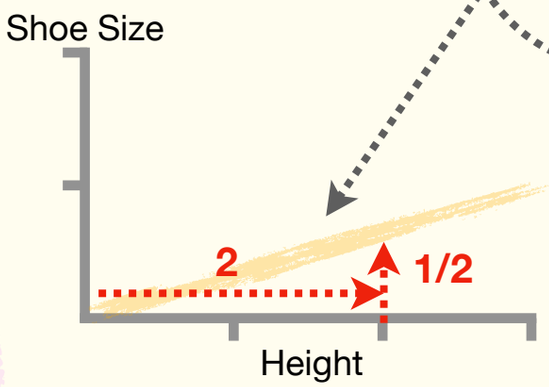


$$\frac{d\ Height}{d\ Weight} = 2$$

…and that means the equation for Height is:

$$Height = \frac{d\ Height}{d\ Weight} \times Weight = 2 \times Weight$$

9 Likewise, this straight line goes up **1/2** units for every **2** units over, so the slope and derivative are **1/4**…



$$\frac{d\ Size}{d\ Height} = \frac{1/2}{2} = \frac{1}{4}$$

…and that means the equation for Shoe Size is:

$$Shoe\ Size = \frac{d\ Size}{d\ Height} \times Height = \frac{1}{4} \times Height$$

Appendix F: The Chain Rule

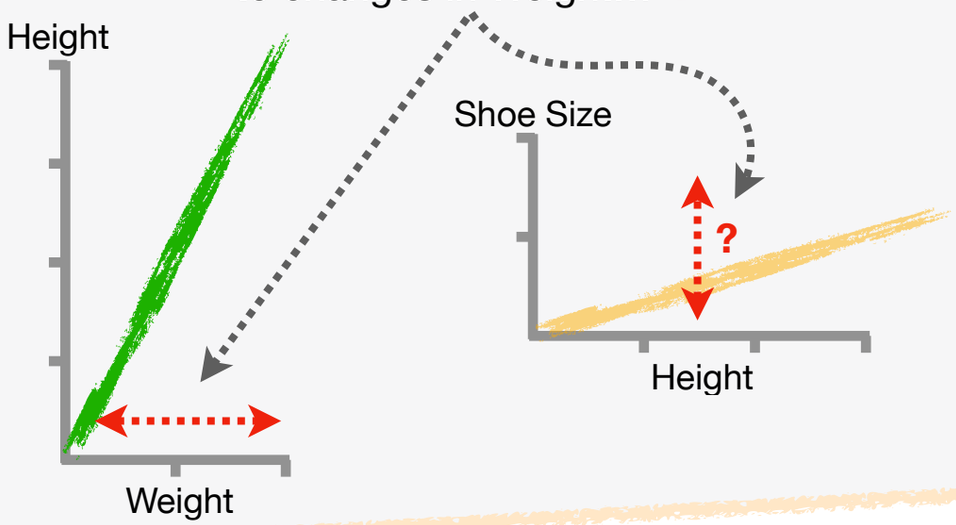
10 Now, because Weight can predict Height… …and Height can predict Shoe Size… …we can plug the equation for Height into the equation for Shoe Size.

$$\text{Height} = \frac{d\ \text{Height}}{d\ \text{Weight}} \times \text{Weight}$$

$$\text{Shoe Size} = \frac{d\ \text{Size}}{d\ \text{Height}} \times \text{Height}$$

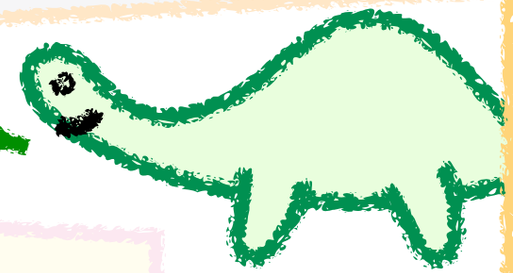
$$\text{Shoe Size} = \frac{d\ \text{Size}}{d\ \text{Height}} \times \frac{d\ \text{Height}}{d\ \text{Weight}} \times \text{Weight}$$

11 And if we want to determine how Shoe Size changes with respect to changes in Weight… …we solve for the derivative of Shoe Size with respect to Weight… …and, using **The Power Rule** (or realizing that when we change Weight, we multiply it by both derivatives to get the new Shoe Size), we end up with the derivative of Shoe Size with respect to Height multiplied by the derivative of Height with respect to Weight.



$$\frac{d\ \text{Size}}{d\ \text{Weight}} = \frac{d\ \text{Size}}{d\ \text{Height}} \times \frac{d\ \text{Height}}{d\ \text{Weight}}$$

In other words, because we can link the two equations with a common variable, in this case, Height, the derivative of the combined function is the product of the individual derivatives.



12 Finally, we can plug in values for the derivatives… …and we see that when Weight increases by **1** unit, Shoe Size increases by **1/2**.

Gentle Reminder:

$$\frac{d\ \text{Height}}{d\ \text{Weight}} = 2$$

$$\frac{d\ \text{Size}}{d\ \text{Height}} = \frac{1}{4}$$

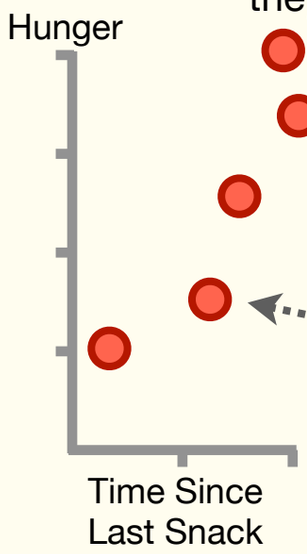
$$\frac{d\ \text{Size}}{d\ \text{Weight}} = \frac{d\ \text{Size}}{d\ \text{Height}} \times \frac{d\ \text{Height}}{d\ \text{Weight}}$$

$$= \frac{1}{4} \times 2 = \frac{1}{2}$$

BAM!!!

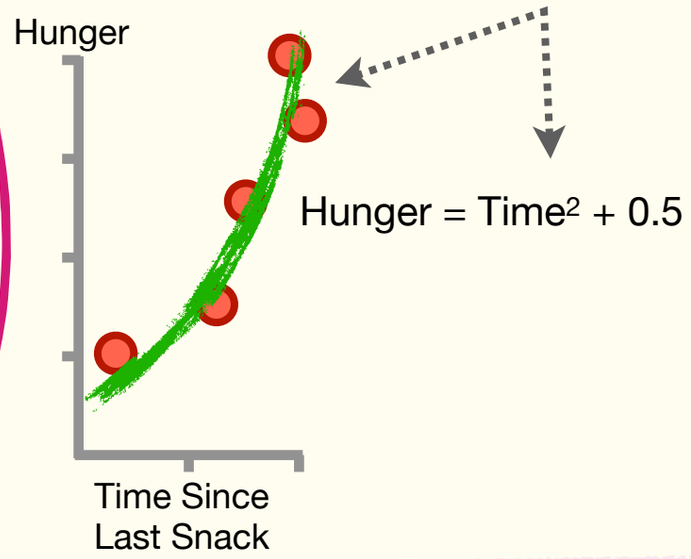
Appendix F: The Chain Rule, A More Complicated Example

1 Now imagine we measured how Hungry a bunch of people were and how long it had been since they last had a snack.

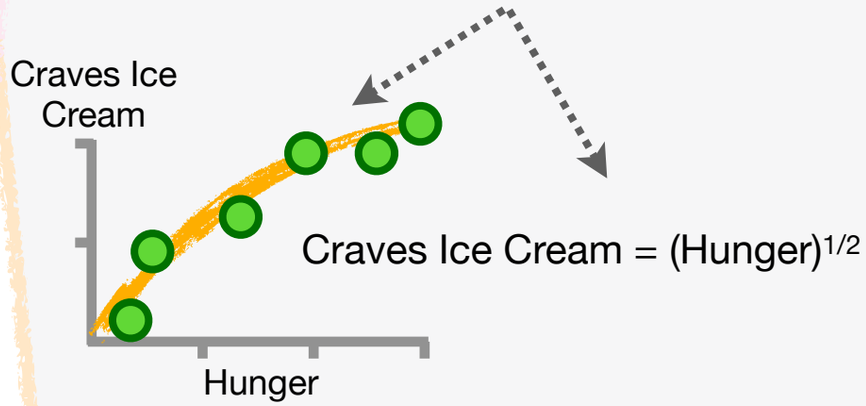


The more Time that had passed since their last snack, the hungrier they got!

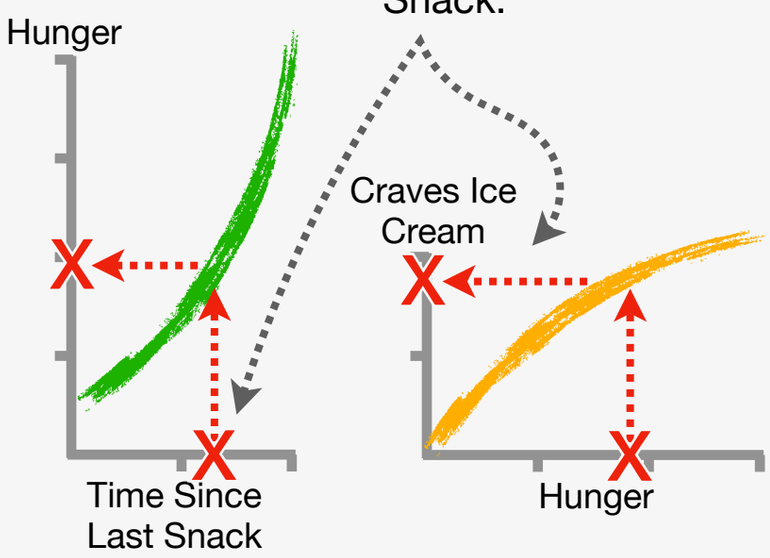
So, we fit a quadratic line with an intercept of **0.5** to the measurements to reflect the increasing rate of Hunger.



2 Likewise, we fit a square root function to the data that shows how Hunger is related to craving ice cream.



3 Now we want to see how Craves Ice Cream changes relative to Time Since Last Snack.



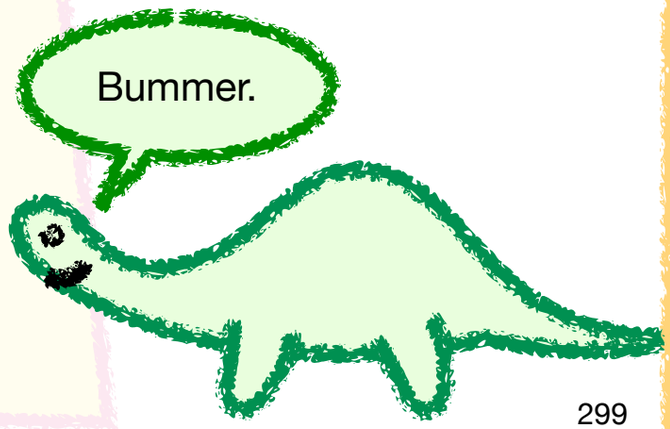
4 Unfortunately, when we plug the equation for Hunger into the equation for Craves Ice Cream…

Hunger = Time² + 0.5

Craves Ice Cream = (Hunger)^1/2

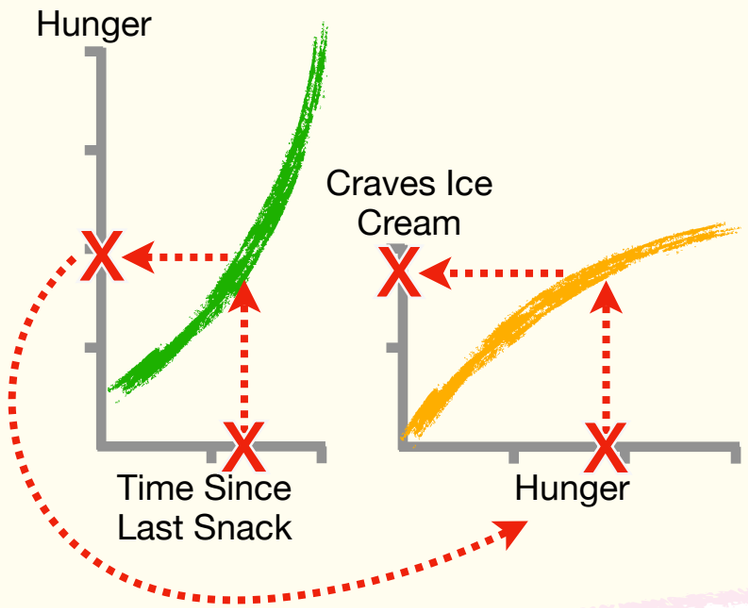
Craves Ice Cream = (Time² + 0.5)^1/2

…raising the sum by **1/2** makes it hard to find the derivative with **The Power Rule**.



Appendix F: The Chain Rule, A More Complicated Example

5 However, because Hunger links Time Since Last Snack to Craves Ice Cream, we can solve for the derivative using **The Chain Rule**!!!



6 **The Chain Rule** tells us that the derivative of Craves Ice Cream with respect to Time…

$$\frac{d\ Craves}{d\ Time} = \frac{d\ Craves}{d\ Hunger} \times \frac{d\ Hunger}{d\ Time}$$

…is the derivative of Craves Ice Cream with respect to Hunger…

…multiplied by the derivative of Hunger with respect to Time.

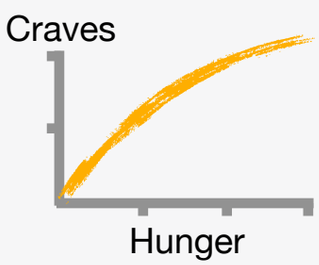
7 First, **The Power Rule** tells us that the derivative of Hunger with respect to Time is this equation:

$$Hunger = Time^2 + 0.5$$

$$\frac{d\ Hunger}{d\ Time} = 2 \times Time$$

8 Likewise, **The Power Rule** tells us that the derivative of Craves Ice Cream with respect to Hunger is this equation:

$$Craves\ Ice\ Cream = (Hunger)^{1/2}$$



$$\frac{d\ Craves}{d\ Hunger} = 1/2 \times Hunger^{-1/2}$$

$$= \frac{1}{2 \times Hunger^{1/2}}$$

9 Now we just plug the derivatives into **The Chain Rule**…

$$\frac{d\ Craves}{d\ Time} = \frac{d\ Craves}{d\ Hunger} \times \frac{d\ Hunger}{d\ Time}$$

$$= \frac{1}{2 \times Hunger^{1/2}} \times (2 \times Time)$$

$$= \frac{2 \times Time}{2 \times Hunger^{1/2}}$$

$$\frac{d\ Craves}{d\ Time} = \frac{Time}{Hunger^{1/2}}$$

…and we see that when there's a change in Time Since Last Snack, the change in Craves Ice Cream is equal to Time divided by the square root of Hunger.

10

NOTE: In this example, it was obvious that Hunger was the link between Time Since Last Snack and Craves Ice Cream, which made it easy to apply **The Chain Rule**.

However, usually we get both equations jammed together like this…

$$Craves\ Ice\ Cream = (Time^2 + 0.5)^{1/2}$$

…and it's not so obvious how **The Chain Rule** applies. So, we'll talk about how to deal with this next!!!

BAM!!!

Appendix F: The Chain Rule, When The Link Is Not Obvious

1 In the last part, we said that raising the sum by **1/2** makes it difficult to apply **The Power Rule** to this equation…

$$\text{Craves Ice Cream} = (\text{Time}^2 + 0.5)^{1/2}$$

…but there was an obvious way to link Time to Craves with Hunger, so we determined the derivative with **The Chain Rule**.

However, even when there's no obvious way to link equations, *we can create a link* so that we can still apply **The Chain Rule**.

2 First, let's create a link between Time and Craves Ice Cream called **Inside**, which is equal to the stuff inside the parentheses…

$$\text{Inside} = \text{Time}^2 + 0.5$$

…and that means Craves Ice Cream can be rewritten as the square root of the stuff **Inside**.

$$\text{Craves Ice Cream} = (\text{Inside})^{1/2}$$

3 Now that we've created **Inside**, the link between Time and Craves, we can apply **The Chain Rule** to solve for the derivative.

The Chain Rule tells us that the derivative of Craves with respect to Time…

$$\frac{d\ Craves}{d\ Time} = \frac{d\ Craves}{d\ Inside} \times \frac{d\ Inside}{d\ Time}$$

…is the derivative of Craves with respect to **Inside**…

…multiplied by the derivative of **Inside** with respect to Time.

4 Now we use **The Power Rule** to solve for the two derivatives…

$$\frac{d\ Craves}{d\ Inside} = \frac{d}{d\ Inside}(\text{Inside})^{1/2} = 1/2 \times \text{Inside}^{-1/2} = \frac{1}{2 \times \text{Inside}^{1/2}}$$

$$\frac{d\ Inside}{d\ Time} = \frac{d}{d\ Time}\text{Time}^2 + 0.5 = 2 \times \text{Time}$$

5 …and plug them into **The Chain Rule**…

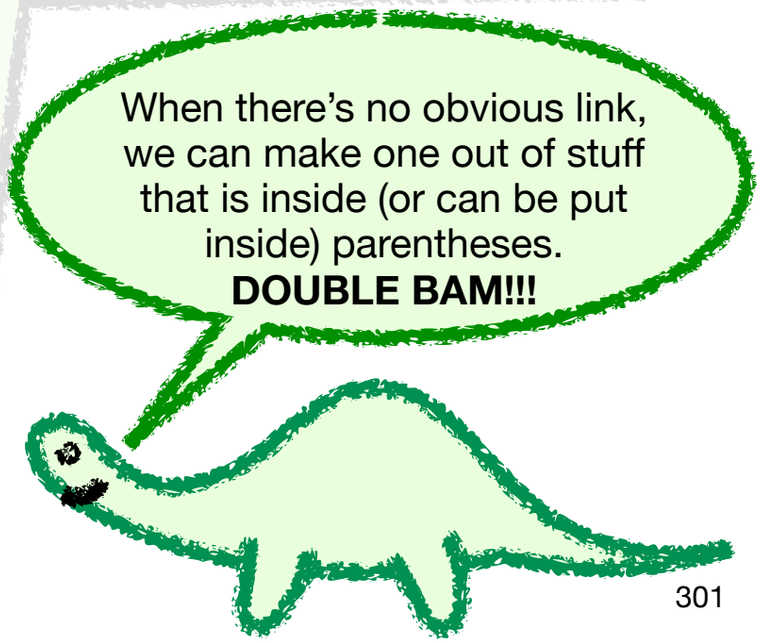
$$\frac{d\ Craves}{d\ Time} = \frac{d\ Craves}{d\ Inside} \times \frac{d\ Inside}{d\ Time}$$

$$\frac{d\ Craves}{d\ Time} = \frac{1}{2 \times \text{Inside}^{1/2}} \times (2 \times \text{Time})$$

$$= \frac{2 \times \text{Time}}{2 \times \text{Hunger}^{1/2}}$$

$$\frac{d\ Craves}{d\ Time} = \frac{\text{Time}}{\text{Hunger}^{1/2}}$$

…and just like when the link, Hunger, was obvious, when we created a link, **Inside**, we got the exact same result. **BAM!!!**



Acknowledgments

Acknowledgments

The idea for this book came from comments on my YouTube channel. I'll admit, when I first saw that people wanted a **StatQuest** book, I didn't think it would be possible because I didn't know how to explain things in writing the way I explained things with pictures. But once I created the **StatQuest** study guides, I realized that instead of *writing* a book, I could *draw* a book. And knowing that I could draw a book, I started to work on this one.

This book would not have been possible without the help of many, many people.

First, I'd like to thank all of the **Triple BAM** supporters on Patreon and YouTube: U-A Castle, J. Le, A. Izaki, Gabriel Robet, A. Doss, J. Gaynes, Adila, A. Takeh, J. Butt, M. Scola, Q95, Aluminum, S. Pancham, A. Cabrera, and N. Thomson.

I'd also like to thank my copy editor, **Wendy Spitzer**. She worked magic on this book by correcting a million errors, giving me invaluable feedback on readability, and making sure each concept was clearly explained. Also, I'd like to thank the technical editors, **Adila**, **Gabriel Robet**, **Mahmud Hasan**, **PhD**, **Ruizhe Ma**, and **Samuel Judge** who helped me with everything from the layout to making sure the math was done properly.

Lastly, I'd like to thank Will Falcon and the whole team at Grid.ai.

Index

Index