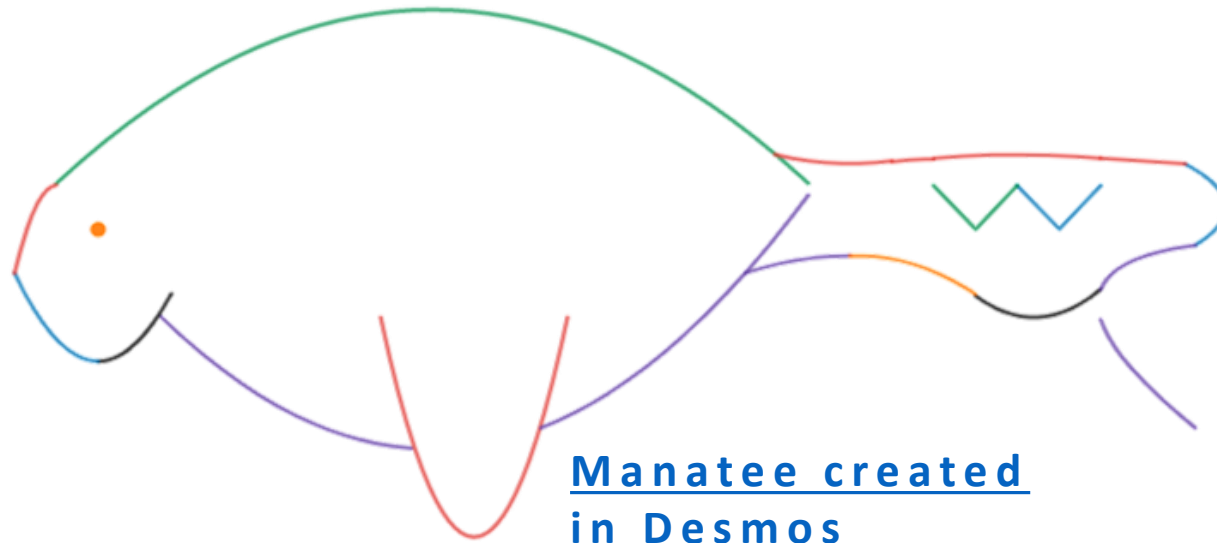


Lesson 3.2 – r and r^2

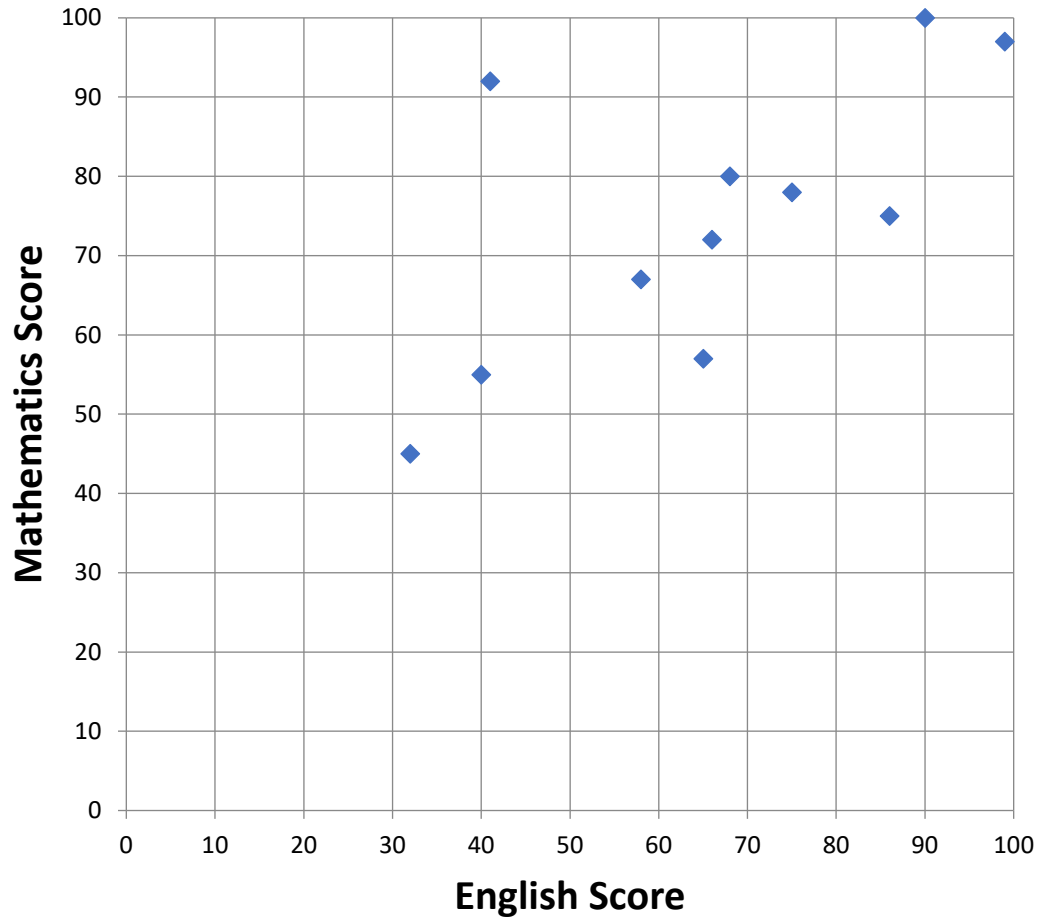
The measures of correlation



Victoria Amarasiri
November 11, 2020

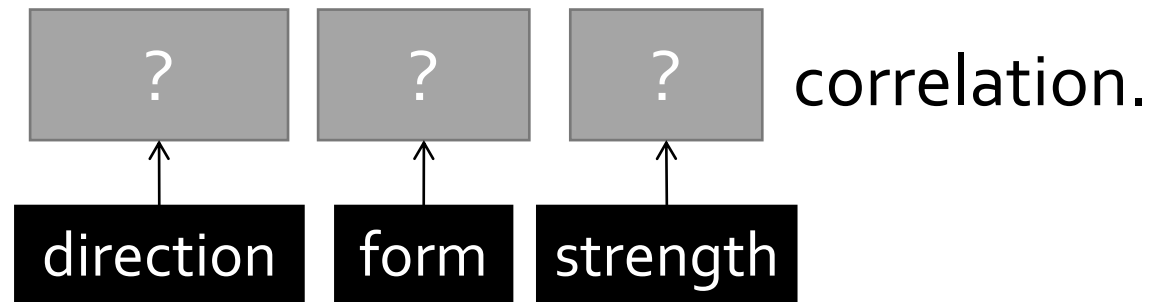
Statistics and Probability

While you are waiting for class to start...



Identify the direction, form and strength of this scatterplot. Put your answer in chat.

What does this scatter plot say? Do you agree?



What does this scatter plot say to you?

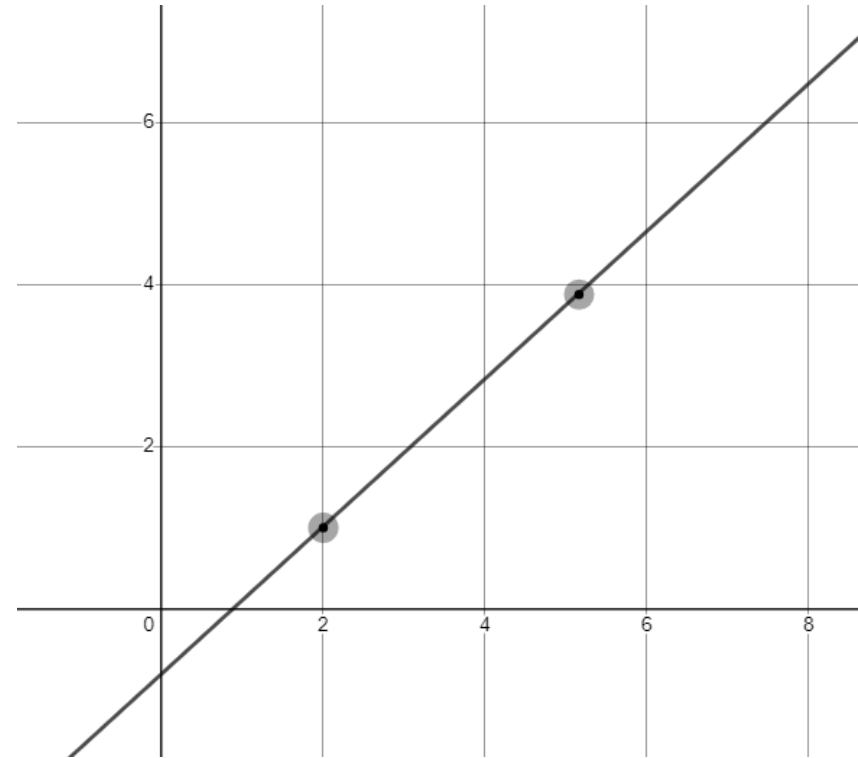
Sometimes it is easy and sometimes it is difficult to “eyeball” the strength of correlation. Today we will be learning about measures of correlation that we can use to make sure we get this right.

Precursors

Remember Algebra?

For any two distinct points, there is a unique line containing them.

How would we describe the linear relationship for this distribution model?



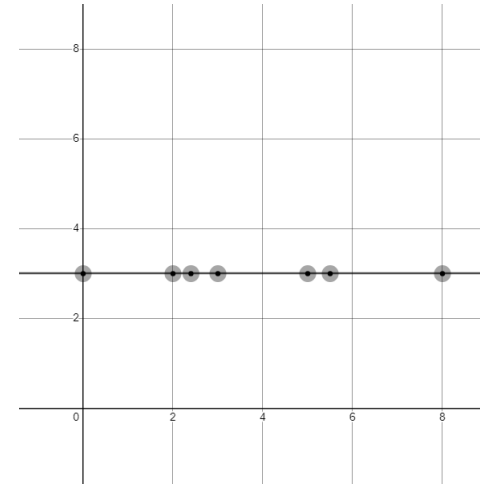
We shouldn't be assessing many distributions with only two points, but a simple regression line may help us to better understand the measures of correlation.

Precursors

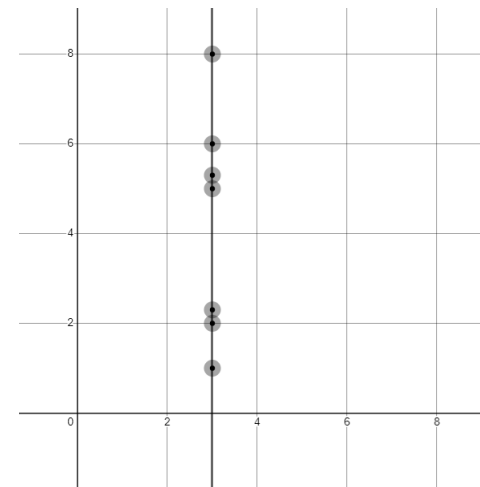
Remember Algebra!

If we have a vertical or horizontal line, the slope is either 0 or *undefined*:

- If $\text{slope}(m) = 0, y = b$; A horizontal line.
- If $\text{slope}(m) = \frac{1}{0}, x = b$; A vertical line.



$$y = b; m=0$$



$$x = b; m \text{ is undefined}$$

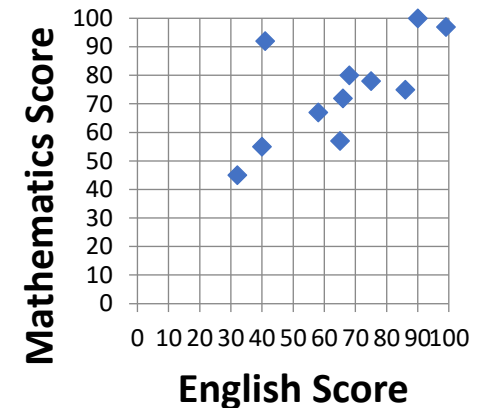
Understand that a 0 or undefined slope will result in an undefined measure of correlation.

Precursors

Remember!

Scatterplots help us to understand the **relationship** between two quantitative variables measured on the same individuals.

- – The variable that explains or influences changes in a response variable. It goes on the x-axis.
- – The variable that measures the result of a study. It goes on the y-axis.



In many studies, the goal is to show that changes in one or more explanatory variables explain changes in a response variable. Sometimes the goal is to show that explanatory variables cause changes...

...however correlation imply direct causation.

Today we will calculate and interpret the correlation coefficient (r) – a measure that helps us better describe the linear relationship between explanatory and response variables.

Measuring Linear Association: The Correlation Coefficient (r)

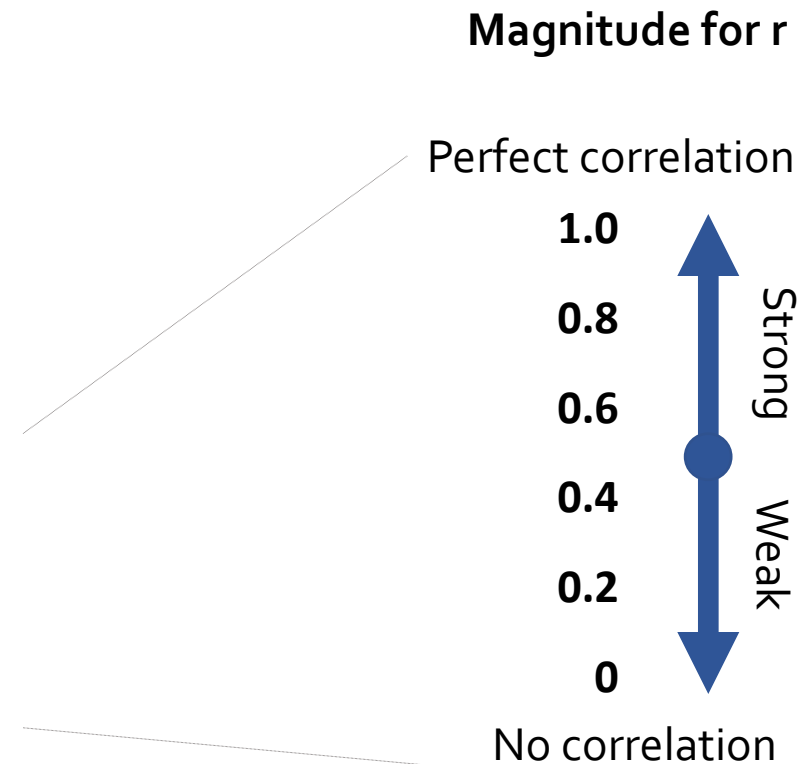
The correlation coefficient r is a **measure** of the **strength** and **direction** of a **linear relationship between two quantitative variables**.

- **Sign**

- + positive association
(as x increases, y increases)

- negative association
(as x increases, y decreases).

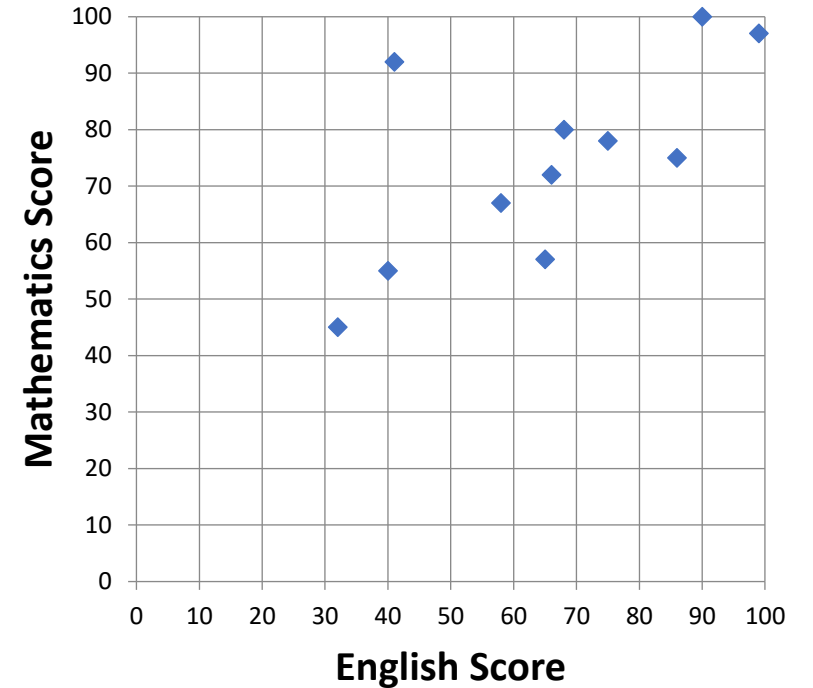
- **Magnitude** is $0 \leq r \leq 1$, where
 $r = 0$ shows no correlation
and $r = 1$ shows perfect correlation.



Measuring Linear Association: The Correlation Coefficient

Properties of the correlation coefficient (r)

- ” **Outliers** have a **significant affect** on the value of r .
- ” The value of r makes **no distinction between explanatory and response variables**; the value of r remains the same when the two variables are interchanged.
- ” The value of r **is not affected when we change the measurement units** of the variables.



Measuring Linear Association: The Correlation Coefficient

Properties

• Outliers

• The value explains the value of the variable

• The value measures

Hey, students!

Go to student.desmos.com
and type in:

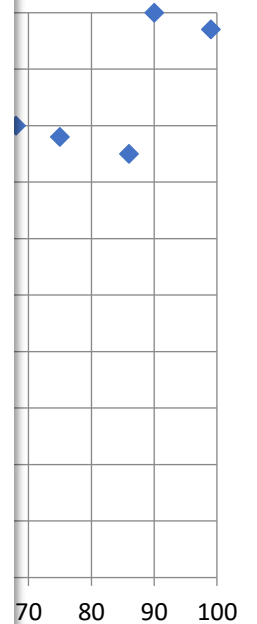
ADJ 7UU

You can also share this link with your students:

<https://student.desmos.com/join/adj7uu>



<https://student.desmos.com/join/adj7uu>



Why is the Manatee on the endangered species list?

These slow moving sea creatures lull and munch their days away in the warm waters off of the Florida coast.

They can weigh more than half a ton and yet they have a certain charm. They are highly curious and love interacting with people and things in their environment. They feed on vegetation in shallow coastal waters. They have no natural predators.

Yet a rising mortality rate has put the manatees in an endangered status.



<https://www.youtube.com/watch?v=ruSqRvCxi-s>

Start: 0:08

Stop: When reader is done.

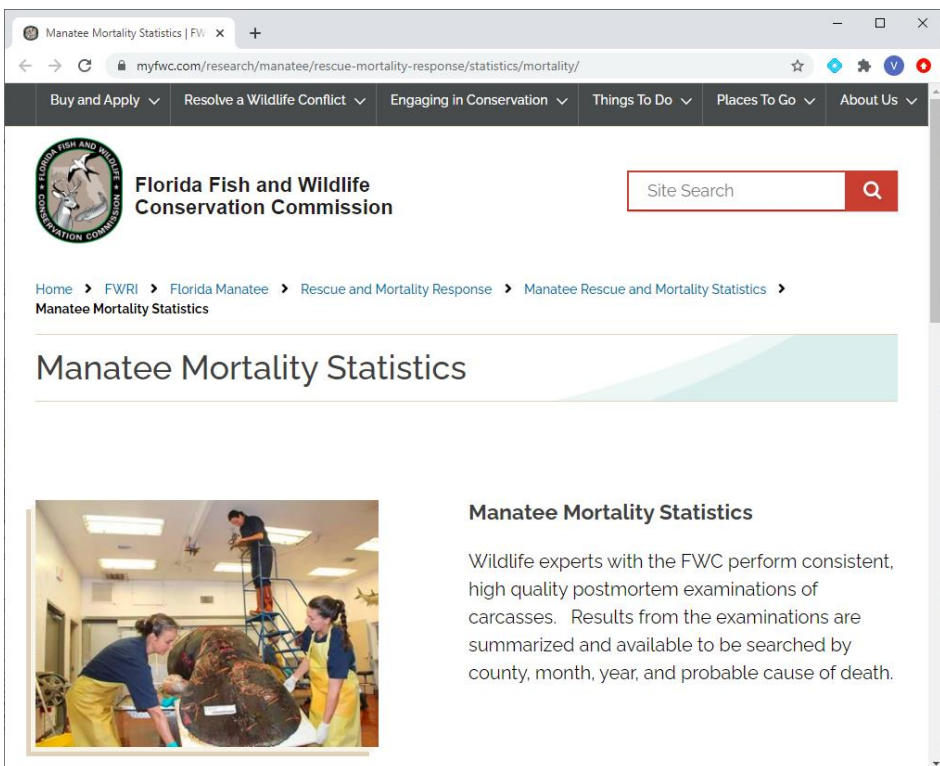
Sea Mermaids (Skip): 3:37

The Problem: 4:28

End: 5:11

The Florida Fish and Wildlife Conservation Commission tracks total manatee mortality in Florida from 1974-2004.

The Number of Manatee Deaths Per Year



<https://myfwc.com/research/manatee/rescue-mortality-response/statistics/mortality/>

1974	3	1990	47	2006	92
1975	6	1991	53	2007	73
1976	10	1992	38	2008	90
1977	13	1993	35	2009	97
1978	21	1994	49	2010	83
1979	24	1995	42	2011	88
1980	16	1996	60	2012	82
1981	24	1997	54	2013	73
1982	20	1998	66	2014	69
1983	15	1999	82	2015	86
1984	34	2000	78	2016	106
1985	34	2001	81	2017	111
1986	33	2002	95	2018	124
1987	39	2003	73	2019	85
1988	43	2004	69		
1989	50	2005	79		

What explanatory variables might we associate with this rise in manatee deaths over?

Proposed association: Manatee deaths to powerboat registrations.

Human related fatalities are a major contributor to the manatees endangered status.

Their inquisitive personality and their tendency to hangout just under the water's surface put them in the path of powerboat propellers. If you are not real careful when you start your motor you risk damaging these beautiful animals.



<https://www.youtube.com/watch?v=ruSqRvCxi-s>

Start: 4:28

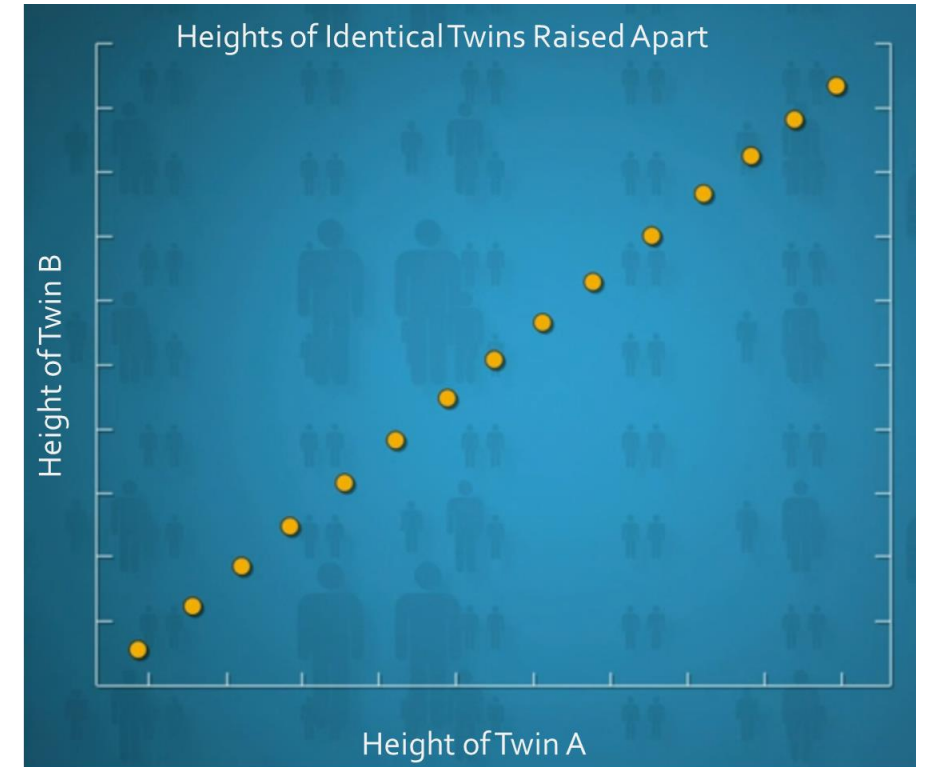
Stop: when reader is done.

End: 5:11

Let's use r to check the strength of the correlation of powerboats registered each year in this area of Florida, but which value should go on the x axis and which on the y?

Exit Ticket

1. What are all the possible values of the correlation coefficient r ?
2. In the scatter plot, the heights of identical twins raised apart are correlated. What is the correlation coefficient (r)?
3. Is the association causation?
4. Is it easy to guess how strong the correlation is by looking at a scatterplot? Explain.
5. On a scale from -1 to 1 , how strong is your understanding of the correlation coefficient (r)?



Thank you!