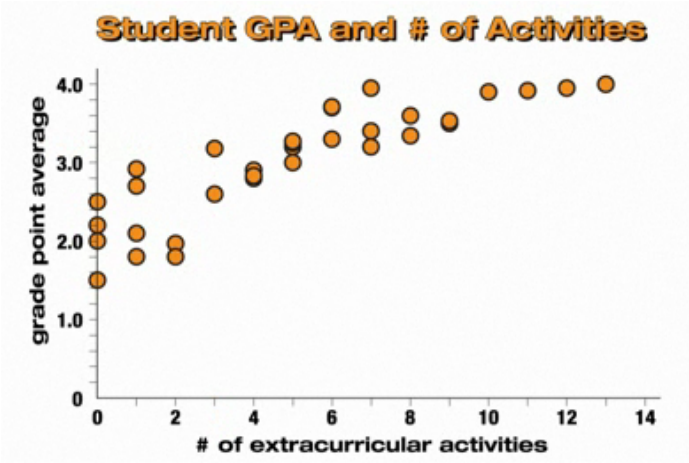
# Interpreting the Correlation Coefficient

## A Closer Look at the Video

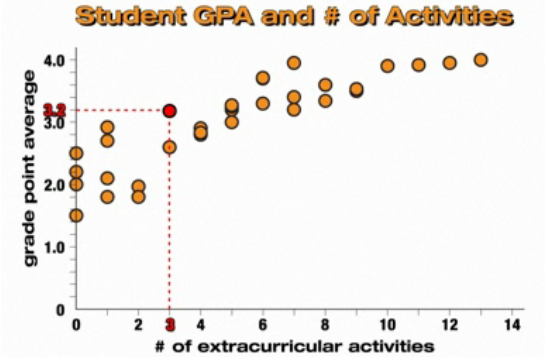
After watching the *Interpreting the Correlation Coefficient* video, make sense of the mathematics by reading through the problem situation and solution. Use the comments and questions in bold to help you understand correlation.

**Problem:** *Nate has been helping his friend, Maddie, conduct background research for a series she is working on for the local newspaper. Maddie is looking into the correlation between extra-curricular activities and the grade point average, or GPA, of high school students. She is thinking of making tomorrow's headline read, "Doing More Activities Will Increase Your GPA." Help Maddie decide if this statement is true by investigating the correlation coefficient for the data set below.*




**Maddie has been citing a study that includes this scatter-plot. What do the x-axis and y-axis represent in this graph?**

The x-axis shows the number of extra-curricular activities the student participates in, and the y-axis shows the grade point average for that student. For example, the red dot represents a student who participates in three activities and has a GPA of 3.2.



# Interpreting the Correlation Coefficient

## A Closer Look at the Video

**By glancing at the scatter plot, we can see that there seems to be a trend. What is the trend?**

The trend shows that the more activities a student participates in, the higher his or her GPA. The scatter-plot allows us to visualize the relationship between two variables. Maddie needs to know the strength of the relationship.

**What is a correlation coefficient?**

The correlation coefficient is a single number that quantifies the strength of a linear relationship between two variables, in this case, extra-curricular activities and GPA.

**How can we find the correlation coefficient for a data set?**

We can find the correlation coefficient by using a formula, shown below, which requires using the mean and standard deviation of each of the two data sets. The correlation coefficient is denoted by r.

$$r = \frac{\sum_{i-1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)s_x s_y}$$

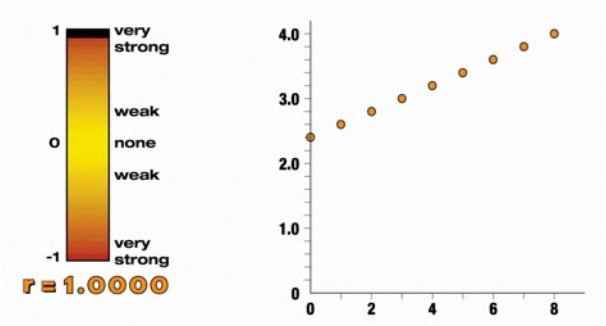
We could calculate the correlation coefficient by hand, but most people determine the correlation coefficient using a calculator or computer. We're not going to focus on why this formula works, but rather understand what the correlation coefficient means.

**What can we say about the range of values of the correlation coefficient? What do the values in this range mean?**

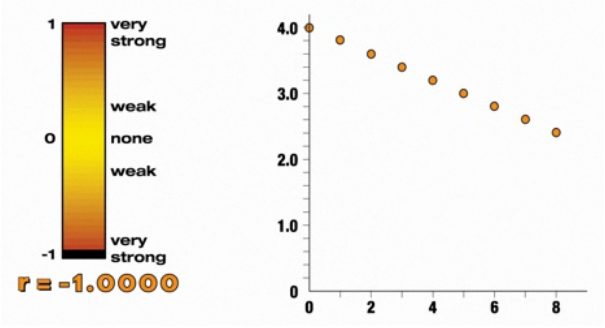
A correlation coefficient is always a value between negative one and positive one and indicates the strength of the association between two variables. The strongest correlation occurs when this number is equal to negative one or positive one.

**What does a correlation coefficient of positive one or negative one mean?**

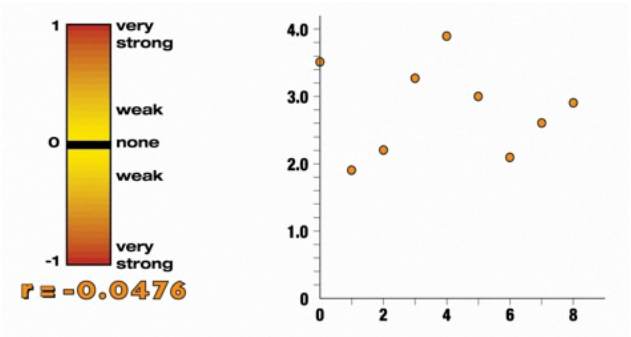
Positive one indicates that the data points all lie on a line with a positive slope; as one variable increases the other does too.



Negative one indicates that the data points lie on a straight line with a negative slope; as one variable increase the other decreases. Both of these scatter plots reflect strong associations.




**What would a weak association look like?**
Notice as the correlation coefficient moves away from one or negative one and *r* approaches 0, the data does not have a visible pattern or trend.




**For our data, the correlation coefficient turned out to be 0.82. Do we have a strong or weak correlation?**
Our correlation coefficient of 0.82 signals a moderately strong relationship between the number of extra-curricular activities and grade point averages. Different industries or fields have different thresholds for what they consider a strong relationship. For example, in the field of psychology, 0.82 is considered a strong association, but it might be considered to be a moderate association in pharmacology.

**Now, Maddie needs to write her headline. One possibility is "Participating in More Extra-Curricular Activities Causes Higher GPAs." What might be the result of this headline?**

If that is the headline, some parents will rush to sign their children up for more extra-curricular activities thinking their GPAs are going to skyrocket. But this would be assuming a causation not a correlation.

**Maddie is worried that she may be jumping to conclusions. This goes back to her original question: Does a strong, 0.82 correlation mean that participating in more activities causes a higher GPA?**
A strong correlation does not necessarily mean participating in more activities *causes* a higher GPA.

**What are three possible theories for what the strong correlation does indicate?**

- *Theory one:* Participating in many extra-curricular activities causes a high GPA.
- *Theory two:* High GPA causes a student to participate in many extra-curricular activities.
- *Theory three:* Neither of these causes the other. There may be many factors influencing both variables. These are called **lurking variables**.

**What are some possible lurking variables in our situation?**
Possible lurking variables might be a strong work ethic, or an income level that provides for coaches and tutors, or a social group that encourages activities and study, or maybe other things that we haven't thought of.

**So again, just because two variables have a strong correlation doesn't mean that one must cause the other. What can we gain from knowing that a strong correlation exists between variables?**
A strong correlation allows us to make predictions and indicates that further investigation is worthwhile.

**To summarize, what is the correlation coefficient, and what are three key points to remember about the correlation coefficient?**
The correlation coefficient is a number that quantifies the strength of a linear relationship between two variables.

- **It is a value between -1 and 1.**
- **The closer the value is to -1 or 1, the stronger the linear relationship.**
- **Strong correlation does not mean causation.**