PROBLEM SOLVED
MAKING SENSE OF MATHEMATICS

Interpreting Correlation Coefficient

Video Script

Scene		Full Transcript
1	**Nate:**	You have got to be kidding me! Eating fish can prevent crime? Oh hey, it's Nate. I've just been checking out these ridiculous headlines. Seriously, listen to this one, "Report claims ice cream sales grow due to increased shark attacks!" How about this one: "Studies show gum-chewing leads to higher math scores." It makes you wonder about the way some reporters use data for their stories. Can eating fish actually cause a reduction in crime? I've been helping my friend Maddie conduct background research for a series she's been working on for the local newspaper. She's looking into the correlation between extra curricular activities and the grade point average (or GPA) of high school students. Maddie was thinking of making tomorrows headline read: "Doing more activities will increase your GPA." She needs to know if this is a true statement. To answer Maddie's question, we're going to investigate the correlation coefficient. Let's get the scoop and get another, problem solved.
2	**Voice-Over Nate:**	Here's the newsroom at the paper where Maddie works. Soon it will be teaming with reporters and editors trying to make deadlines. Maddie has been citing a study that includes an interesting scatter plot. The X-axis shows the number of extra curricular activities a student participates in and the Y-axis show grade point average for that student. For example, this student participates in three activities and has a GPA of 3.2. Just by glancing at the plot, we can see that there seems to be a trend. The more activities for the student, the higher the GPA!
	Nate:	The scatter plot allows us to visualize the relationship between two variables. Maddie needs to know the strength of the relationship.
	Voice-Over Nate:	There's a single number, called the correlation coefficient that quantifies the strength of a linear relationship between two variables. In this case, extra curricular activities and GPA. We can find the correlation coefficient by using a formula, which requires the mean and standard deviation of each of the two data sets, number of activities and GPA. The correlation coefficient is denoted "r."
3	**Nate:**	We could calculate the correlation coefficient by hand but most people determine the correlation coefficient using a calculator or a computer. We're not going to focus on why this formula works, but rather understand what the correlation coefficient means.

MakingSense

Interpreting Correlation Coefficient

Video Script

		For this data our correlation coefficient turns out to be eighty-two hundredths. Our next task is to figure out exactly what this number means.
	Voice-Over Nate:	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. Positive one indicates all the data points lie on a line with a positive slope, like this! So as one variable increases, the other does too. While negative one indicates that they lie on a straight line with a negative slope, like this! As one variable increases, the other decreases. These scatter plots reflect strong associations. But what would a weak association look like? Notice as the correlation coefficient moves away from one or negative one and "r" approaches zero the data doesn't have a visible pattern or trend.
4	**Nate:**	I reminded Maddie, the closer the correlation coefficient is to positive or negative one, the stronger the association between the variables. So, where does that leave us?
5	**Voice-Over Nate:**	Our correlation coefficient of eighty-two hundredths signals a moderately strong relationship between the number of extra curricular activities and grade point averages.
	Nate:	In fact, different industries or fields have different thresholds for what they consider a strong relationship.
	Voice-over Nate:	For example, in the field of psychology, eighty-two hundredths is considered a strong correlation, but it might be considered a moderate correlation in Pharmacology.
6	**Nate:**	So, now Maddie needs to write her headline. One possibility is: "Participating in more extracurricular activities causes higher GPAs." If that is the headline, some parents will rush to sign their children up for more extra curricular activities, thinking their GPA's are going to skyrocket! Maddie is worried that she may be jumping to conclusions. This goes back to her original question, "Does a strong eighty-two hundredths correlation mean that participating in more activities causes a higher GPA? Not necessarily. There are three possible theories.
	Voice-Over Nate:	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 additional factors are called lurking variables.

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 we haven't thought of.

Nate: So again, just because two variables have a strong correlation, doesn't mean that one must cause the other. However, a strong correlation allows us to make predictions and indicates that further investigation is worthwhile.

Voice-Over Nate: To summarize, the correlation coefficient quantifies the strength of a linear relationship between two variables. It is a value between negative one and one and the closer the value is to negative one or one, the stronger the linear relationship. But remember, strong correlation does not mean causation.

Nate: So what will tomorrow's headlines read? Stop the presses! It will have to read, s"Correlation Does Not Imply Causation." Problem Solved!