SCATTERPLOTS AND THE "R" VALUE

6:12 / 6:12 1.0x

SPEAKER: MICHAEL J. MAHOMETA, Ph.D.

Galton's first idea of showing a relationship between bivariate data,

with his "scattergraph," was pretty close to what

we call a scatterplot today.

He had counts that represented cases for each pairing of x and y data,

and the scatterplot does not.

It just has dots; now sometimes those dots

Help

are bigger or darker because of more points stacked up on the same location.

The great thing about the scatterplot - and Galton's use of it -

is that it follows our basic principle so far: graph the data.

Get a visual feel for what's going on.

In fact, that's all Galton had when he started - just a graph.

It wasn't until years later that the correlation coefficient

emerged as a standard metric of a linear relationship between two

numeric variables.

So how do we set up our graph?

And once that's done, how can we tie in what

we're seeing with our actual correlation value?

Just as always, let's start with a basic question.

Before we take a look at our actual data,

let's go over the idea of graph set-up.

We know that we have two numerical variables,

and we want to see a relationship between them.

So the scatterplot is a great choice to use

Help

for a graph.

But now we need to decide how to set up the scatterplot

- specifically what variable goes on what axis.

Now, we've all heard of the old saying "correlation does not imply causation."

And that's true.

Correlation is nothing more than a numerical value

for the relationship between two numerical variables.

To get actual causation, we would need to set up

a formal experiment where we manipulate some parameter.

But in showing a correlation, we want to keep in mind

the sense of FLOW to the data.

What variable makes sense to lead into another?

In the case of our question, it seems completely reasonable

that test anxiety would FLOW into test score.

That test anxiety DRIVES, in some way, test score.

Thinking about it the other way around doesn't really make sense.

Help

Does test score DRIVE test anxiety?

Typically, we like to put the variable that we are interested in -

the Dependent variable - on the y-axis.

In this case, I think test score makes a good candidate for the Dependent

variable.

It's the variable that we're really interested in.

So on the x-axis, we can put our test anxiety.

In general, this is good common practice.

Our Dependent variable on the y-axis, while on the other variable axis -

the Independent variable, the DRIVING variable - is placed on the x-axis.

Let's take a look and see how some possible - hypothetical -

data to our question might look and the corresponding correlation value.

We could have have a situation where as test anxiety increases then

test score decreases in a perfect pattern - a straight line.

If this is the case, then our correlation coefficient would be -1.0.

Negative because as one variable increases,

the other variable decreases.

Help

A 1.0 because there's actually a perfect correlation

relationship between the two variables.

What if we saw that as test anxiety increases then test score

still decreases, but there's a bit more play to the scatter.

Then our correlation would be about -0.8.

Here's a correlation of -0.5, and here's what a correlation of -0.2 looks like.

And here is what a correlation of zero looks like.

Now we can see the same relationships but going into the positive realm:

that when test anxiety increases, then test score also increases.

We see the same basic dispersion of our data,

but in the opposite - or positive - direction.

Remember that the correlation coefficient ranges from zero to 1.0

- closer to 1.0 meaning a higher relationship.

And the sign of the value indicating the direction:

positive for increases in the Dependent variable as the x -

or Independent variable - increases; and negatives for decreases

Help

in the y-variable as the x-variable increases.

With both positive and negative values taken into account,

our Pearson, or correlation coefficient, technically ranges from -1.0 to 1.0.

Remember also that the correlation is not

effected by the scale of the variable.

A relationship is a relationship, regardless of that scale.

Let's see what the actual data looked like for our question.

What would you picture if I told you the data showed a Pearson correlation

coefficient of -0.30.

Do you have an idea of what the scatter plot might look like?

Here's what the scatterplot actually looks like.

Did it match the picture in your head?

So to answer our question, it looks like test anxiety

is in fact related to the outcome score on an exam.

As anxiety increases, the outcome score, the performance, decreases.

Now the -0.30 isn't that big of a relationship,

but it's still a relationship.

So I'll use it every chance I get.

Remember: deep breaths - relax.

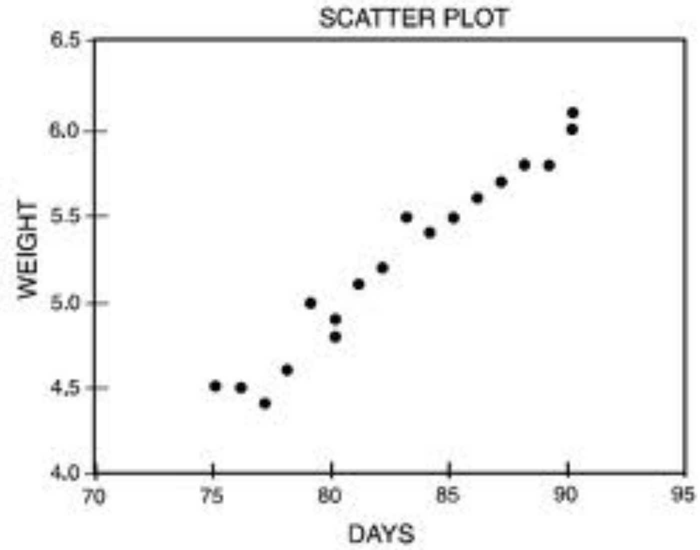
Decrease anxiety, increase performance.

Help

Comprehension Check

A researcher plotted the weights of each of his lab mice against their age (in days). This produced the following scatterplot:

Help



(4/4 points)

How would you describe the direction of the relationship shown in this scatterplot?

- Positive
- Negative
- There is no relationship

How would you describe the strength of the relationship in this scatterplot?

- ( ) Weak (the data points are widely scattered)
- ( ) Moderately strong (there is a clear pattern, but there are a few significant outliers)
- (•) Very strong (nearly all the data points sit on or near the linear trend line) ✔
- ( ) None (there is no visible relationship)

Help

Which of the following best describes the relationship between these two variables?

- ( ) As mice age, their growth rate slows down.
- (•) The mice get heavier as they age. ✔
- ( ) The mice do not appear to grow much during the first month of life.

Which would be the most likely value of the Pearson correlation coefficient, r, for this relationship between age and weight?

- ( ) -0.87
- ( ) 0.56
- (•) 0.93 ✔
- ( ) 1.00

Check | Show Answer

Below is a table of Pearson correlations (called a ***correlation matrix***) for income, amount of education, age, and intelligence.

	Education	Age	IQ
Income	0.65	0.41	0.27
Education		0.11	0.38
Age			-0.02

Help

(3/3 points)

Which variables have the strongest relationship?

- ○ Age and Income
- ○ Age and IQ
- ● Education and Income ✔
- ○ Education and IQ

Which variables appear to have almost no relationship at all?

- ○ Age and Education
- ● Age and IQ ✔
- ○ IQ and Income
- ○ There are no variables that are not related.

What would be the correlation of Education with Education?

- 0.00
- 0.50
- 1.00 ✔
- Cannot be answered

Help

Check | Show Answer

Terms of Service and Honor Code

Privacy Policy (Revised 4/16/2014)

Donate to edX

Jobs at edX

Google+

Help