

14.2: What do Two Quantitative Variables Look Like?

Let's begin learning about two quantitative variables with a scenario about chocolate and happiness.

Chocolate and Happiness Scenario

Let's imagine that a person's supply of chocolate has a *causal influence* on their level of happiness. This means that having chocolate is the reason why someone is happy. Let's further imagine that the more chocolate you have the more happy you will be, and the less chocolate you have, the less happy you will be. (What tends to blow my mind is that many causal relationships can also go the other way; what if I eat chocolate every time that I'm happy? That means that being happy is the reason why I might eat chocolate!) Finally, because we suspect happiness is caused by lots of other things in a person's life, we anticipate that the relationship between chocolate supply and happiness won't be perfect. What do these assumptions mean for how the data should look?

Our first step is to collect some imaginary data from 100 people. We walk around and ask the first 100 people we meet to answer two questions:

1. How much chocolate do you have today?, and
2. How happy are you today?

For convenience, both the scales will go from 0 to 100. For the chocolate scale, 0 means no chocolate, 100 means eating chocolate constantly from the moment they woke up. Any other number is somewhere in between. For the happiness scale, 0 means no happiness, 100 means all of the happiness, and in between means some amount in between.

We asked each participants our two questions so there are two scores for each participant, one for their chocolate supply and one for their level of happiness. Although they both are on a 0-10 scale, they are different questions, so finding the means of each won't necessarily help us if chocolate predicts happiness. Table 14.2.1 show data from the first 10 imaginary participants, who all had very low ratings on our chocolate scale and on our happiness scale.

Table 14.2.1- Chocolate and Happiness Scores

| Participant | Chocolate | Happiness |
|-------------|-----------|-----------|
| A | 1 | 1 |
| B | 1 | 1 |
| C | 2 | 2 |
| D | 2 | 4 |
| E | 4 | 5 |
| F | 4 | 5 |
| G | 7 | 5 |
| H | 8 | 5 |
| I | 8 | 6 |
| J | 9 | 6 |

Look at Table 14.2.1. Do you notice any relationship between amount of chocolate and level of happiness? We can see that there is variance in chocolate supply across the 10 participants. We can see that there is variance in happiness across the 10 participants. Variance means that the numbers have some change in them, they are not all the same, some of them are big, some are small.

In particular, does it look like participants with more chocolate tend to score higher on the happiness scale? What does this have to do with variance? Well, it means there is a relationship between the variance in chocolate supply, and the variance in happiness levels. The two measures seem to vary together. When we have two measures that vary together, they are like a happy couple who share their variance. Chocolate and happiness seem to co-vary (meaning that the two variables seem to vary together). This is what co-variance refers to, the idea that the pattern of varying numbers in one measure is shared by the pattern of varying numbers in another measure.

To make this co-variance even more clear, let's plot all of the data in a graph. Think back to the [chapter on graphs](#), then answer the following question:

? Exercise 14.2.1

Which type of graph is used when we have two quantitative variables and want to see them both combined?

Answer

A scatterplot plots one quantitative variable on the x-axis and the other quantitative variable on the y-axis. This makes a "scatter" of dots in which scores that are low on both variables are towards the bottom left, and scores that are high on both variables are towards the top right.

Graphing Two Quantitative Variables

The scatter plot in Figure 14.2.1 shows 100 dots for each participant. You might be wondering, why are there only 100 dots for the data. Didn't we collect 100 measures for chocolate, and 100 measures for happiness, shouldn't there be 200 dots? Nope. Each dot is for one participant, there are 100 participant, so there are 100 dots. Each dot is placed so that it represents *both* measurements for each participant. In other words, each dot has two coordinates, an x-coordinate for chocolate, and a y-coordinate for happiness.

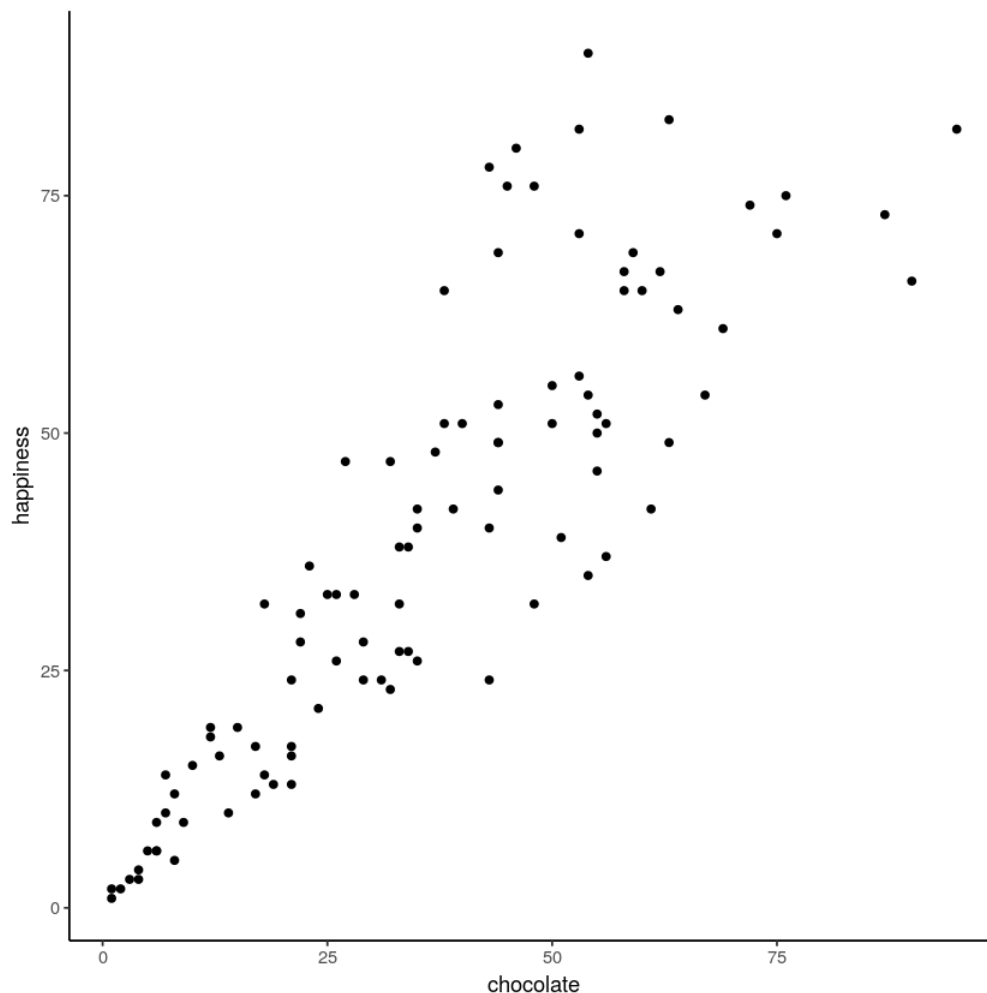


Figure 14.2.1: 100 Participants' Amount of Chocolate and Amount Happiness. (CC-BY-SA, [Matthew J. C. Crump](#) via [Answer Questions with Data](#))

What do the dots mean? The dot all the way on the bottom left is a participant who had close to 0 chocolate and close to zero happiness. You can look at any dot, then draw a straight line down to the x-axis: that will tell you how much chocolate that participant has. You can draw a straight line left to the y-axis: that will tell you how much happiness the participant has.

Positive, Negative, Curvilinear, and No Relationship

Now that we are looking at the scatter plot, we can see many things. It looks like the dots show a relationship between chocolate supply and happiness. Happiness is lower for people with smaller supplies of chocolate, and higher for people with larger supplies of chocolate. It looks like the more chocolate you have the happier you will be, and vice-versa. This kind of relationship is called a *positive correlation*. In this sense, "positive" doesn't mean "good." Instead, it means "going in the same direction."

Seeing as we are in the business of imagining data, let's imagine some more. We've already imagined what data would look like if larger chocolate supplies increase happiness. What do you imagine the scatter plot would look like if the relationship was reversed, and larger chocolate supplies decreased happiness? This means that the more chocolate you might have, the less happy you are. This is called a *negative correlation* because as one variable increase, the other variable decreases. Again, "negative" isn't saying that anything is bad, but that the variables are going in the opposite direction. Instead of a general trend going up and to the right, a negative correlation goes starts at the upper left of the graph and has a trend down to the lower right.

What do you imagine the scatter plot would look like if people were least happy with no chocolate and with lots of chocolate? This would be a curvilinear (curved line) relationship, and is described as a reverse-U because the general trend looks like an upside down U.

What do you imagine the scatter plot would look like if there was no relationship, and the amount of chocolate that you have doesn't do anything to your happiness? That kind of scatter plot just looks like a bunch of dots, scattered randomly.

Correlations

Now that we know that two quantitative variables can be related, can we infer a statistical relationship? Yes! That's what a Pearson's r (or Pearson's correlation) measures. A positive Pearson's r means that there is a positive linear relationship; as one variable increase, the other variable also increases. If you calculated a correlation with the sample of 10 participants from Table 14.2.1, you'd get $r(8)=0.86$, $p<.05$. The calculated r is positive and the p -value is less than 0.05, so we can say that these 10 participants show a positive linear relationship between chocolate and happiness.

A negative Pearson's r means that there is a negative linear relationship; as one variable increases, the other variable decreases. For example, the more you miss class (number of absences), the lower your final grade percentage; as absences increase, grades decrease.

The tricky part is when Pearson's r is close to zero. This could show that there is no relationship between the two quantitative variables, OR it could mean that there is a curvilinear relationship. Pearson's r only looks for linear relationships (how close to a straight line all of the dots form), so it can't tell the difference between a random splatter of dots and clearly curved line. That's why looking at the scatter plot is so important.

Cause and Effect

We are wading into the idea that one variable causes the change in another variable (having more chocolate causes more happiness), but Pearson's r only measures linear relationships, it can't tell you what causes happiness. Maybe something else causes both happiness and chocolate supply to increase? Dr. MO always have more chocolate around the house at the end of October, but she also loves costumes, so maybe Halloween causes an increase in chocolate supply and an increase in happiness. Fair warning: we will find patterns that look like one thing is causing another, even when that one thing DOES NOT CAUSE the other thing. Hang in there.

Contributors and Attributions

- Matthew J. C. Crump (Brooklyn College of CUNY)
-

Dr. MO (Taft College)

This page titled 14.2: What do Two Quantitative Variables Look Like? is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Matthew J. C. Crump via source content that was edited to the style and standards of the LibreTexts platform.

- 3.1: If something caused something else to change, what would that look like? by Matthew J. C. Crump is licensed CC BY-SA 4.0.
Original source: <https://www.crumplab.com/statistics/>.