

## 3.2: Pearson's $r$

If Beyoncé was a statistician, she might look at these scatter plots and want to “put a number on it”. We think this is a good idea too. We’ve already learned how to create descriptive statistics for a single measure, like chocolate, or happiness (i.e., means, variances, etc.). Is it possible to create a descriptive statistic that summarized the relationship between two measures, all in one number? Can it be done? Karl Pearson to the rescue.

### Note

The stories about the invention of various statistics are very interesting, you can read more about them in the book, “The Lady Tasting Tea” (Salsburg 2001)

There’s a statistic for that, and Karl Pearson invented it. Everyone now calls it, “Pearson’s  $r$ ”. We will find out later that Karl Pearson was a big-wig editor at Biometrika in the 1930s. He took a hating to another big-wig statistician, Sir Ronald Fisher (who we learn about later), and they had some stats fights...why can’t we all just get along in statistics.

How does Pearson’s  $r$  work? Let’s look again at the first 10 subjects in our fake experiment:

subject	chocolate	happiness
1	1	1
2	2	2
3	2	3
4	3	3
5	3	3
6	5	5
7	4	6
8	5	5
9	9	5
10	6	9
Sums	40	42
Means	4	4.2

What could we do to these numbers to produce a single summary value that represents the relationship between the chocolate supply and happiness?

### The idea of co-variance

“Oh please no, don’t use the word variance again”. Yes, we’re doing it, we’re going to use the word variance again, and again, until it starts making sense. Remember what variance means about some numbers. It means the numbers have some change in them, they are not all the same, some of them are big, some are small. We can see that there is variance in chocolate supply across the 10 subjects. We can see that there is variance in happiness across the 10 subjects. We also saw in the scatter plot, that happiness increases as chocolate supply increases; which is a positive relationship, a positive correlation. 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 vary together don’t they? When we have two measures that vary together, they are like a happy couple who share their variance. 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.

**Co-variance** is **very, very, very, very** important. We suspect that the word co-variance is initially confusing, especially if you are not yet fully comfortable with the meaning of variance for a single measure. Nevertheless, we must proceed and use the idea of co-

variance over and over again to firmly implant it into your statistical mind (we already said, but redundancy works, it's a thing).

*Pro tip: Three-legged race is a metaphor for co-variance. Two people tie one leg to each other, then try to walk. It works when they co-vary their legs together (positive relationship). They can also co-vary in an unhelpful way, when one person tries to move forward exactly when the other person tries to move backward. This is still co-variance (negative relationship). Funny random walking happens when there is no co-variance. This means one person does whatever they want, and so does the other person. There is a lot of variance, but the variance is shared randomly, so it's just a bunch of legs moving around accomplishing nothing.*

*Pro tip #2: Successfully playing paddycake occurs when two people coordinate their actions so they have positively shared co-variance.*

---

This page titled [3.2: Pearson's r](#) 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.