13.1: Variability and Covariance

Because we have two continuous variables, we will have two characteristics or scores on which people will vary. What we want to know is do people vary on the scores together. That is, as one score changes, does the other score also change in a predictable or consistent way? This notion of variables differing together is called covariance (the prefix "co" meaning "together").

Let's look at our formula for variance on a single variable:

$$s^2=\frac{\sum(X-M)^2}{N-1} \tag{13.1.1}$$

We use X to represent a person's score on the variable at hand, and M to represent the mean of that variable. The numerator of this formula is the Sum of Squares, which we have seen several times for various uses. Recall that squaring a value is just multiplying that value by itself. Thus, we can write the same equation as:

$$s^2=\frac{\sum((X-M)(X-M))}{N-1} \tag{13.1.2}$$

This is the same formula and works the same way as before, where we multiply the deviation score by itself (we square it) and then sum across squared deviations.

Now, let's look at the formula for covariance. In this formula, we will still use X to represent the score on one variable, and we will now use Y to represent the score on the second variable. The formula for covariance (cov_{XY} with the subscript XY to indicate covariance across the X and Y variables) is:

$$\operatorname{cov}_{XY}=\frac{\sum((X-M_X)(Y-M_Y))}{N-1} \tag{13.1.3}$$

As we can see, this is the exact same structure as the previous formula. Now, instead of multiplying the deviation score by itself on one variable, we take the deviation scores from a single person on each variable and multiply them together. We do this for each person (exactly the same as we did for variance) and then sum them to get our numerator. The numerator in this is called the Sum of Products.

$$SP=\sum((X-M_X)(Y-M_Y)) \tag{13.1.4}$$

We will calculate the sum of products using the same table we used to calculate the sum of squares. In fact, the table for sum of products is simply a sum of squares table for X, plus a sum of squares table for Y, with a final column of products, as shown below.

Table 13.1.1: Sum of Products table

X	$(X-M_X)$	$(X-M_X)^2$	Y	$(Y-M_Y)$	$(Y-M_Y)^2$	$(X-M_X)(Y-M_Y)$

This table works the same way that it did before (remember that the column headers tell you exactly what to do in that column). We list our raw data for the X and Y variables in the X and Y columns, respectively, then add them up so we can calculate the mean of each variable. We then take those means and subtract them from the appropriate raw score to get our deviation scores for each person on each variable, and the columns of deviation scores will both add up to zero. We will square our deviation scores for each variable to get the sum of squares for X and Y so that we can compute the variance and standard deviation of each (we will use the standard deviation in our equation below). Finally, we take the deviation score from each variable and multiply them together to get our product score. Summing this column will give us our sum of products. It is very important that you multiply the raw deviation scores from each variable, NOT the squared deviation scores.

Our sum of products will go into the numerator of our formula for covariance, and then we only have to divide by $N-1$ to get our covariance. Unlike the sum of squares, both our sum of products and our covariance can be positive, negative, or zero, and they will always match (e.g. if our sum of products is positive, our covariance will always be positive). A positive sum of products and covariance indicates that the two variables are related and move in the same direction. That is, as one variable goes up, the other will also go up, and vice versa. A negative sum of products and covariance means that the variables are related but move in opposite directions when they change, which is called an inverse relation. In an inverse relation, as one variable goes up, the other variable goes down. If the sum of products and covariance are zero, then that means that the variables are not related. As one variable goes up or down, the other variable does not change in a consistent or predictable way.

The previous paragraph brings us to an important definition about relations between variables. What we are looking for in a relation is a consistent or predictable pattern. That is, the variables change together, either in the same direction or opposite directions, in the same way each time. It doesn't matter if this relation is positive or negative, only that it is not zero. If there is no consistency in how the variables change within a person, then the relation is zero and does not exist. We will revisit this notion of direction vs zero relation later on.

Contributors and Attributions

- Foster et al. (University of Missouri-St. Louis, Rice University, & University of Houston, Downtown Campus)
