Introduction to correlation

Introduction the correlation coefficient equation

Here we will consider the Pearson product-moment correlation coefficient

- Several equations are to follow
- Will walk through them step by step
- Explain each step verbally
- Relate the equation back to previous material

You will not need to memorise any of the equations you see here, but you should be able to recognise and understand the equation for the correlation coefficient.

How is the correlation coefficient defined?

The covariance between two variables X and Y, Cov(X, Y), divided by the standard deviation of X times the standard deviation of Y,

$$Correlation = \frac{Cov(X, Y)}{StDev(X) \times StDev(Y)}.$$

But what is Cov(X, Y), exactly?

$$Cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

Let's break this down a bit further!

What is the covariance?

What is the covariance?

$$Cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

Assume ‘x’ is bird length and ‘y’ is bird mass. How would we calculate this, explained verbally?

1. For each bird, calculate its length minus mean bird length
2. For each bird, calculate its mass minus mean bird mass
3. Multiply the two steps above together
4. Do the above for all birds, and add up all the values
5. Divide what you added up by the total number of birds

This is the numerator part of the correlation coefficient

Back to the definition of the correlation

$$Correlation = \frac{Cov(X, Y)}{StDev(X) \times StDev(Y)}.$$

Now we can expand that covariance

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{StDev(X) \times StDev(Y)}.$$

We already know the formula for standard devation

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

The Pearson's correlation coefficient

$$r = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}\sqrt{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}}.$$