One situation where covariances show up is when we try to calculate the variance of a sum of random variables.

So let us look at the variance of the sum of two random variables, X1 and X2.

If the two random variables are independent, then we know that the variance of the sum is the sum of the variances.

Let us now look at what happens in the case where we may have dependence.

By definition, the variance is the expected value of the difference of the random variable we're interested in from its expected value, squared.

And now we rearrange terms here and write what is inside the expectation as follows.

We put together X1 with the term minus the expected value of X1 and then X2 together with negative the expected value of X2.

So now we have the square of the sum of two terms.

We expand the quadratic to obtain expected value of the square of the first term plus the square of the second term plus 2 times a cross term.

And what do we have here?

The expected value of the first term is just the variance of X1.

The expected value of this second term is just the variance of X2.

And finally, the cross term, the expected value of it, we recognize that it is the same as the covariance of X1 with X2.

And we also have this factor of 2 up here.

So this is the general form for the variance of the sum of two random variables.

In the case of independence, the covariance is 0, and we just have the sum of the two variances.

But when the random variables are dependent, it is possible that the covariance will be non-zero, and we have one additional term.

Let us now not generalize this calculation.

Here is for reference and comparison the formula for the case where we add two random variables.

But now let us look at the variance of the sum of many of them.

To keep the calculation simple, we're going to assume that the means are zero.

But the final conclusion will also be valid for the case of non-zero means.

Since we have assumed zero means, the variance is the same as the expected value of the square of the random variable involved, which is this one.

And now we expand this quadratic to obtain the expected value of: we will have a bunch of terms of this, where i ranges from 1 up to n.

And then we will have a bunch of cross terms of the form Xi, Xj.

And we obtain one cross term for each choice of i from 1 to n and for each choice of j from 1 to n, as long as i is different from j.

So overall here, this sum will have n squared minus n terms.

Now, we use linearity to move the expectation inside the summation.

And so from here, we obtain the sum of the expected value of Xi squared, which is the same as the variance of Xi, since we assumed zero means.

And similarly here, we're going to get this double sum over i's that are different from j of the expected value of Xi, Xj.

And in the case of 0 means again, this is the same as the covariance of Xi with Xj.

And so we have obtained this general formula that gives us the variance of a sum of random variables.

If the random variables have 0 covariances, then the variance of the sum is the sum of the variances.

And this happens in particular when the random variables are independent.

For the general case, where we may have dependencies and non-zero variances, then the variance of the sum involves also all the possible covariances between the different random variables.

And let me finally add that this formula is also valid for the general case where we do not assume that the means

are zero.

And the derivation is very similar, except that there's a few more symbols that are floating around.