

Calculating a covariance matrix involves several steps. Let's break it down with a simple example.

### Step 1: Collect Data

Assume we have a dataset with two variables,  $X$  and  $Y$ . Here's a small dataset:

Observation	$X$	$Y$
1	2	3
2	4	5
3	6	7
4	8	9

### Step 2: Calculate the Means

Calculate the mean of each variable.

$$\text{Mean of } X = \frac{2 + 4 + 6 + 8}{4} = 5$$

$$\text{Mean of } Y = \frac{3 + 5 + 7 + 9}{4} = 6$$

### Step 3: Calculate Deviations from the Mean

For each observation, calculate the deviation from the mean for  $X$  and  $Y$ .

Observation	$X$	$Y$	$X - \text{Mean of } X$	$Y - \text{Mean of } Y$
1	2	3	$2 - 5 = -3$	$3 - 6 = -3$
2	4	5	$4 - 5 = -1$	$5 - 6 = -1$
3	6	7	$6 - 5 = 1$	$7 - 6 = 1$
4	8	9	$8 - 5 = 3$	$9 - 6 = 3$

## Step 4: Calculate the Covariance

Now, we can calculate the covariance for  $X$  and  $Y$ .

$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

Where  $n$  is the number of observations.

1. Calculate the products of the deviations:

Observation	$(X - \bar{X})(Y - \bar{Y})$
1	$(-3) \cdot (-3) = 9$
2	$(-1) \cdot (-1) = 1$
3	$1 \cdot 1 = 1$
4	$3 \cdot 3 = 9$

2. Sum the products:

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 9 + 1 + 1 + 9 = 20$$

3. Divide by  $n - 1$ :

$$\text{Cov}(X, Y) = \frac{\downarrow 20}{4-1} = \frac{20}{3} \approx 6.67$$

## Step 5: Calculate Variances

Now, calculate the variances for  $X$  and  $Y$ :

Variance of  $X$ :

$$\text{Var}(X) = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

Observation	$(X - \bar{X})^2$
1	$(-3)^2 = 9$
2	$(-1)^2 = 1$
3	$1^2 = 1$
4	$3^2 = 9$

$$\sum (X_i - \bar{X})^2 = 9 + 1 + 1 + 9 = 20$$

$$\text{Var}(X) = \frac{20}{3} \approx 6.67$$

Variance of  $Y$ :

$$\text{Var}(Y) = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$$

Observation	$(Y - \bar{Y})^2$
1	$(-3)^2 = 9$
2	$(-1)^2 = 1$
3	$1^2 = 1$
4	$3^2 = 9$

$$\sum (Y_i - \bar{Y})^2 = 9 + 1 + 1 + 9 = 20$$

$$\text{Var}(Y) = \frac{20}{3} \approx 6.67$$

### Step 6: Construct the Covariance Matrix

The covariance matrix is structured as follows:

$$\begin{bmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Var}(Y) \end{bmatrix}$$

Substituting our results:

$$\text{Covariance Matrix} = \begin{bmatrix} \frac{20}{3} & \frac{20}{3} \\ \frac{20}{3} & \frac{20}{3} \end{bmatrix} \approx \begin{bmatrix} 6.67 & 6.67 \\ 6.67 & 6.67 \end{bmatrix}$$

### Conclusion

The covariance matrix for the dataset is:

$$\begin{bmatrix} 6.67 & 6.67 \\ 6.67 & 6.67 \end{bmatrix}$$

This matrix indicates the variance of each variable along the diagonal and the covariance between the variables off the diagonal.