# Variance and Covariance Exercises (Pre-Reading for Data Science MSc)

1. State the formula for the variance of a variable $x$ and explain what each component means.

$$Var(x) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

- $x_i \rightarrow$ the i-th data point
- $\bar{x} \rightarrow$ the sample mean
- $(x_i - \bar{x})^2 \rightarrow$ the squared deviation
- $\sum_{i=1}^{n} (x_i - \bar{x})^2 =$ sum of sq. dev.
- $\frac{1}{n-1} \rightarrow$ scaling factor for the sample variance

2. Compute the variance of the dataset $x = (2, 4, 6, 8)$ using the formula.

$$Var(x) = \frac{1}{3} \sum_{i=1}^{n} (x_i - 5)^2$$

$$= \frac{1}{3}(9 + 1 + 1 + 9) = \frac{20}{3}$$

3. State the formula for the covariance between two variables $x$ and $y$.

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

4. Compute the covariance of $x = (1, 2, 3)$ and $y = (2, 4, 5)$.

$$\mathrm{Cov}(x, y) = \frac{1}{2} \sum_{i=1}^{n} (x_i - 2)\left(y_i - \frac{11}{3}\right)$$

$$= \frac{1}{2}\left( 5/3 + 0 + 4/3 \right) = 1.5$$

5. Write down the general form of the covariance matrix for three variables $X_1, X_2, X_3$.

$$\text{Matrix } A = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \mathrm{Cov}(X_1, X_3) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \mathrm{Cov}(X_2, X_3) \\ \mathrm{Cov}(X_3, X_1) & \mathrm{Cov}(X_3, X_2) & \mathrm{Var}(X_3) \end{bmatrix}$$

6. For the covariance matrix

$$\Sigma = \begin{bmatrix} 4 & -2 \\ -2 & 3 \end{bmatrix},$$

state the variances and covariances represented.

$\mathrm{Var}(X_1) = 4$  $\mathrm{Cov}(X_1, X_2) = -2$

$\mathrm{Var}(X_2) = 3$

7. State three key properties of covariance matrices.

1. Square ($n \times n$ for $X_n$ points)
2. Symmetric
3. Diagonal entries = variances
4. Always PSD

8. Define what it means for a matrix to be positive semidefinite (PSD).

Matrix A is PSD if for all vectors z:

$$z^\top A z \geq 0$$

9. For $z = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and

$$\Sigma = \begin{bmatrix} 4 & -2 \\ -2 & 3 \end{bmatrix},$$

compute $z^\top \Sigma z$ to check PSD property.

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 4 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \end{bmatrix} = 8 \geq 0$$