Covariance and Fitting of Exponential Curve

Sagarika Srivastava
Sidra Shaikh
Sidra Solkar

September 25, 2024

Abstract

Covariance measures how two variables change together. In exponential curve fitting, we estimate parameters (like a and b) to best fit the data. The covariance between these parameters shows how changes in one affect the other. High covariance suggests they are more dependent, which can make the fit less stable. Low covariance means they vary more independently, leading to a more robust fit.This work is jointly done by Sagarika Srivastava, Sidra Solkar and Sidra Shaikh, Rizvi College of Engineering, Mumbai.

Keywords: Correlation, Curve Fitting, Variance

Covariance

Definition

Covariance is a statistical measure that indicates the extent to which two random variables change together. If the variables tend to increase or decrease together, the covariance is positive. If one variable increases while the other decreases, the covariance is negative. A covariance of zero indicates no relationship between the variables.

Covariance Theorem and Formula

The covariance between two random variables X and Y is given by:

$$\text{Cov}(X, Y) = \mathbb{E}[(X] - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Alternatively, this can be expressed as:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Where:

- $\mathbb{E}[X]$ is the expected value (mean) of X,

- $\mathbb{E}[Y]$ is the expected value (mean) of Y,

- $\mathbb{E}[XY]$ is the expected value of the product of X and Y.

Example 01:

Find the covariance for the following data:

Solution:

$x = \{6, 5, 3, 4, 2\}$, $y = \{12, 10, 8, 6, 4\}$, and $N = 5$ Step 1: Calculate Means

$$\bar{X} = \frac{6 + 5 + 3 + 4 + 2}{5} = \frac{20}{5} = 4$$

$$\bar{Y} = \frac{12 + 10 + 8 + 6 + 4}{5} = \frac{40}{5} = 8$$

Step 2: Sample Covariance Formula Using the sample covariance formula:

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{N - 1}$$

$$= \frac{(6-4)(12-8) + (5-4)(10-8) + (3-4)(8-8) + (4-4)(6-8) + (2-4)(4-8)}{5-1}$$

$$= \frac{8 + 2 + 0 + 0 + 6}{4}$$

$$= \frac{16}{4} = 4$$

Step 3: Population Covariance Formula Using the population covariance formula:

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{N}$$

$$= \frac{(6-4)(12-8) + (5-4)(10-8) + (3-4)(8-8) + (4-4)(6-8) + (2-4)(4-8)}{5}$$

$$= \frac{8 + 2 + 0 + 0 + 6}{5}$$

$$= \frac{16}{5} = 3.2$$

Example 02:

The standard deviation of X is 15, and the standard deviation of Y is 20. The correlation between X and Y is 5. Find the covariance of X and Y.

Solution:

Given Data

$$\sigma_X = 15, \quad \sigma_Y = 20, \quad \rho(X, Y) = 5$$

Formula The formula for the correlation between X and Y is:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Substitute the Values Substituting the given values into the formula:

$$5 = \frac{\text{Cov}(X, Y)}{15 \times 20}$$

Solve for Covariance

$$\text{Cov}(X, Y) = 5 \times 15 \times 20 = 1500$$

Thus, the covariance of X and Y is:

$$\boxed{1500}$$

Example 03:

Compute the covariance for the following data set:

Solution:

$$X = \{10, 12, 15, 18, 20\}, \quad Y = \{8, 10, 12, 14, 16\}$$

Step 1: Calculate the mean of X and Y

$$\bar{X} = \frac{10 + 12 + 15 + 18 + 20}{5} = \frac{75}{5} = 15$$

$$\bar{Y} = \frac{8 + 10 + 12 + 14 + 16}{5} = \frac{60}{5} = 12$$

Step 2: Compute the deviations from the mean and their products

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
10	8	-5	-4	20
12	10	-3	-2	6
15	12	0	0	0
18	14	3	2	6
20	16	5	4	20

Step 3: Sum the products

$$\sum (X - \bar{X})(Y - \bar{Y}) = 20 + 6 + 0 + 6 + 20 = 52$$

Step 4: Calculate the covariance
For $n = 5$, the sample covariance is:

$$\text{Cov}(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n - 1} = \frac{52}{5 - 1} = \frac{52}{4} = 13$$

The covariance of the given data set is:

$$\boxed{13}$$

Since the covariance is positive, this indicates that X and Y tend to increase together.

Fitting of Exponential Curve

Definition

Exponential curve fitting involves finding an exponential function of the form:

$$y = ae^{bx}$$

that best fits a set of data points. It is used to model data that grows or decays at a constant rate, and is applied in fields such as biology, finance, and physics to identify patterns and trends.

Theorem and Formulae

Exponential curve fitting is a method used to find an exponential function that best approximates the relationship between two variables, typically represented as:

$$y = ae^{bx}$$

where a and b are constants, and e is the base of the natural logarithm. This method is commonly applied when data exhibits exponential growth or decay.

By transforming the exponential function into a linear form using logarithms, we can apply linear regression to estimate the parameters a and b.

Given a set of n data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we transform the equation $y = ae^{bx}$ by taking the natural logarithm of both sides:

$$\ln y = \ln a + bx$$

This becomes a linear equation:

$$Y = A + bx$$

where $Y = \ln y$ and $A = \ln a$. The goal is to find the best-fitting values for A and b using linear regression. The parameters are determined by:

$$b = \frac{n \sum x_i \ln y_i - \sum x_i \sum \ln y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$A = \frac{\sum \ln y_i - b \sum x_i}{n}$$

Once A is known, a can be found as:

$$a = e^A$$

Thus, the exponential function that best fits the data is:

$$y = ae^{bx}$$

Example 01:

Calculate the fitting exponential equation $y = ax^b$ using the least squares method for the following data:

Solution:

x	y
2	27.8
3	62.1
4	110
5	161

The curve to be fitted is $y = ax^b$.

Taking the logarithm on both sides, we get:

$$\log_{10} y = \log_{10} a + b \log_{10} x$$

Let $Y = \log_{10} y$, $A = \log_{10} a$, and $X = \log_{10} x$. This is now a linear equation in Y and X:

$$Y = A + bX$$

The corresponding normal equations are:

$$\sum Y = nA + b \sum X$$

$$\sum XY = A \sum X + b \sum X^2$$

The values are calculated using the following table:

x	y	$X = \log_{10}(x)$	$Y = \log_{10}(y)$	X^2	$X \cdot Y$
2	27.8	0.301	1.444	0.0906	0.4347
3	62.1	0.4771	1.7931	0.2276	0.8555
4	110	0.6021	2.0414	0.3625	1.229
5	161	0.699	2.2068	0.4886	1.5425
$\sum x = 14$	$\sum y = 360.9$	$\sum X = 2.0792$	$\sum Y = 7.4854$	$\sum X^2 = 1.1693$	$\sum X \cdot Y = 4.0618$

Substituting these values in the normal equations:

$$4A + 2.0792b = 7.4854$$

$$2.0792A + 1.1693b = 4.0618$$

Solving these equations, we get:

$$A = 0.868, \quad b = 1.9303$$

Thus, $a = \text{antilog}_{10}(A) = \text{antilog}_{10}(0.868) = 7.3791$.
Now, substituting these values in the equation $y = ax^b$, we get:

$$y = 7.3791 \cdot x^{1.9303}$$

Example 02:

Given Data: Calculate the fitting exponential equation $y = ax^b$ for the following data:

Solution:

x	y
1	2.7
2	7.4
3	20.1
4	54.6
5	148.4

Step 1: Transform the Data Taking the logarithm on both sides of the equation:

$$\log_{10} y = \log_{10} a + b \log_{10} x$$

Let $Y = \log_{10} y$, $A = \log_{10} a$, and $X = \log_{10} x$. The equation can now be expressed as:

$$Y = A + bX$$

Step 2: Prepare the Table We calculate X and Y values:

x	y	$X = \log_{10}(x)$	$Y = \log_{10}(y)$	X^2	XY
1	2.7	0	0.4314	0	0
2	7.4	0.3010	0.8686	0.0906	0.2618
3	20.1	0.4771	1.3032	0.2276	0.6200
4	54.6	0.6021	1.7370	0.3625	1.0435
5	148.4	0.6990	2.1716	0.4886	1.5195
$\sum x = 15$	$\sum y = 233.2$	$\sum X = 2.0792$	$\sum Y = 6.5118$	$\sum X^2 = 1.1693$	$\sum XY = 3.4448$

Step 3: Calculate Normal Equations The normal equations are:

$$\sum Y = nA + b \sum X$$

$$\sum XY = A \sum X + b \sum X^2$$

Substituting the sums into the equations: 1. For $n = 5$:

$$6.5118 = 5A + b(2.0792) \tag{1}$$

2. For the second equation:

$$3.4448 = A(2.0792) + b(1.1693) \tag{2}$$

Step 4: Solve the System of Equations From Equation (1):

$$5A + 2.0792b = 6.5118$$

From Equation (2):

$$2.0792A + 1.1693b = 3.4448$$

Using substitution or elimination methods, we can solve these equations to find A and b. After solving, we obtain:

$$A \approx 0.6692, \quad b \approx 1.4160$$

Step 5: Find a Using the relationship $a = 10^A$:

$$a \approx 10^{0.6692} \approx 4.688$$

Step 6: Write the Exponential Model Now substituting the values of a and b back into the exponential equation:

$$y \approx 4.688 \cdot x^{1.4160}$$

The fitted exponential curve is given by:

$$y \approx 4.688 \cdot x^{1.4160}$$

This model effectively captures the growth behavior of the dataset.

Example 03:

Fit an exponential curve of the type $Y = ae^{bX}$ from the following data:

x	y
1	5
2	10
4	30

Solution:

To fit the exponential curve $Y = ae^{bX}$, we first take the logarithm of both sides:

$$\log Y = \log a + bX$$

Let $u = \log Y$ and $A = \log a$. The equation can now be expressed as:

$$u = A + bX$$

The normal equations are:

$$\sum u = nA + B \sum x^2$$
$$\sum ux = A \sum x + B \sum x^2$$

Now, let's prepare the table for the computations:

x	y	$u = \log y$	x	x^2
1	5	0.6990	1	1
2	10	1.0000	2	4
4	30	1.4771	4	16
$\sum x = 7$	$\sum y = 45$	$\sum u = 3.1761$	$\sum x = 7$	$\sum x^2 = 21$

Now, substituting the values into the normal equations:
1. From the first equation:

$$3A + 7B = 3.1761 \tag{1}$$

2. From the second equation:

$$7A + 21B = 8.6075 \tag{2}$$

By solving these simultaneous equations, we get:

$$A \approx 0.4604 \quad \text{and} \quad B \approx 0.2564$$

Next, we calculate a:

$$a = \text{antilog}(A) = \text{antilog}(0.4604) \approx 2.8867$$

For b:

$$b = \frac{0.2564}{\log(2.71828)} \approx 0.5904$$

Thus, the curve of best fit is:

$$Y = 2.8867e^{0.5904X}$$

Conclusion

Covariance measures the relationship between two variables, indicating whether they move together or inversely. In curve fitting, covariance helps assess how well one variable predicts another, guiding the fit of mathematical models to data. Together, they are essential for understanding and modeling relationships in data, helping to identify trends and make accurate predictions.

[1–3]

References

[1] B.S. Grewal and J.S. Grewal. *Higher Engineering Mathematics*. Khanna, 2003.

[2] G.V. KUMBHOJKAR. *APPLIED MATHEMATICS III*. MUMBAI, 2nd. edition.

[3] T. Veerarajan. *Probability, Statistics and Random Processes*. Tata McGraw-Hill, India, 2008.

Team

Sagarika Srivastava 44 231P047
Sidra Shaikh 40 231P064
Sidra Solkar 43 231P087