Covariance and Correlation

Hayk Aprikyan, Hayk Tarkhanyan

Covariance

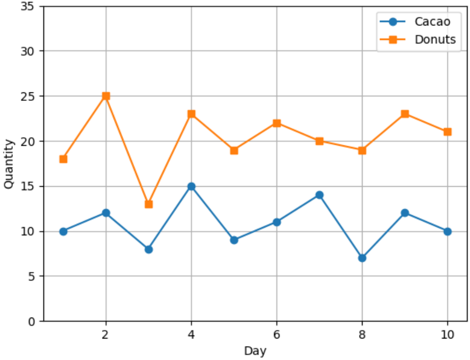
In the table below, we have the number of customers of Ponchikanots who bought cacao and those who bought donuts, for 10 consecutive days.

Day	Cacao	Donuts
1	10	18
2	12	25
3	8	13
4	15	23
5	9	19
6	11	22
7	14	20
8	7	19
9	12	23
10	10	21

Covariance

In the table below, we have the number of customers of Ponchikanots who bought cacao and those who bought donuts, for 10 consecutive days.

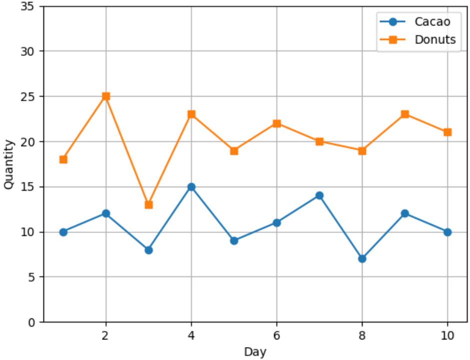
Day	Cacao	Donuts
1	10	18
2	12	25
3	8	13
4	15	23
5	9	19
6	11	22
7	14	20
8	7	19
9	12	23
10	10	21



Covariance

In the table below, we have the number of customers of Ponchikanots who bought cacao and those who bought donuts, for 10 consecutive days.

Day	Cacao	Donuts
1	10	18
2	12	25
3	8	13
4	15	23
5	9	19
6	11	22
7	14	20
8	7	19
9	12	23
10	10	21



Does there seem to be any relation between the number of cacao and donuts sold?

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold, do more donuts get sold as well?

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold, do more donuts get sold as well?

What does "more cacao" mean?

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold *compared to the average*, do more donuts get sold as well (again, compared to the average)?

What does "more cacao" mean?

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold *compared to the average*, do more donuts get sold as well (again, compared to the average)?

What does "more cacao" mean?

Day	Cacao	Donuts
1	10	18
2	12	25
3	8	13
4	15	23
5	9	19
6	11	22
7	14	20
8	7	19
9	12	23
10	10	21

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold *compared to the average*, do more donuts get sold as well (again, compared to the average)?

What does "more cacao" mean?

Day	Cacao - average	Donuts - average
1	$10 - 10.8$	$18 - 20.3$
2	$12 - 10.8$	$25 - 20.3$
3	$8 - 10.8$	$13 - 20.3$
4	$15 - 10.8$	$23 - 20.3$
5	$9 - 10.8$	$19 - 20.3$
6	$11 - 10.8$	$22 - 20.3$
7	$14 - 10.8$	$20 - 20.3$
8	$7 - 10.8$	$19 - 20.3$
9	$12 - 10.8$	$23 - 20.3$
10	$10 - 10.8$	$21 - 20.3$

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold *compared to the average*, do more donuts get sold as well (again, compared to the average)?

What does "more cacao" mean?

Day	Cacao - average	Donuts - average
1	-0.8	-2.3
2	1.2	4.7
3	-2.8	-7.3
4	4.2	2.7
5	-1.8	-1.3
6	0.2	1.7
7	3.2	-0.3
8	-3.8	-1.3
9	1.2	2.7
10	-0.8	0.7

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold *compared to the average*, do more donuts get sold as well (again, compared to the average)?

What does "more cacao" mean?

Day	Cacao - average	Donuts - average	Product
1	−0.8	−2.3	1.84
2	1.2	4.7	5.64
3	−2.8	−7.3	20.44
4	4.2	2.7	11.34
5	−1.8	−1.3	2.34
6	0.2	1.7	0.34
7	3.2	−0.3	−0.96
8	−3.8	−1.3	4.94
9	1.2	2.7	3.24
10	−0.8	0.7	−0.56

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold *compared to the average*, do more donuts get sold as well (again, compared to the average)?

What does "more cacao" mean?

Day	Cacao - average	Donuts - average	Product
1	-0.8	-2.3	1.84
2	1.2	4.7	5.64
3	-2.8	-7.3	20.44
4	4.2	2.7	11.34
5	-1.8	-1.3	2.34
6	0.2	1.7	0.34
7	3.2	-0.3	-0.96
8	-3.8	-1.3	4.94
9	1.2	2.7	3.24
10	-0.8	0.7	-0.56

Covariance

- In other words, what happens to the number of donuts sold when the number of cacao sold increases/decreases?
- When more cacao is sold *compared to the average*, do more donuts get sold as well (again, compared to the average)?

What does "more cacao" mean?

Day	Cacao - average	Donuts - average	Product
1	-0.8	-2.3	1.84
2	1.2	4.7	5.64
3	-2.8	-7.3	20.44
4	4.2	2.7	11.34
5	-1.8	-1.3	2.34
6	0.2	1.7	0.34
7	3.2	-0.3	-0.96
8	-3.8	-1.3	4.94
9	1.2	2.7	3.24
10	-0.8	0.7	-0.56

} average!

Covariance

Now in general, if X and Y are two random variables, and we want to measure how much (on average) a change in X affects a change in Y, we can do the following:

Covariance

Now in general, if X and Y are two random variables, and we want to measure how much (on average) a change in X affects a change in Y, we can do the following:

- Take X and Y

$$X \qquad\qquad Y$$

Covariance

Now in general, if X and Y are two random variables, and we want to measure how much (on average) a change in X affects a change in Y, we can do the following:

- Take X and Y
- Subtract their means to see how much they deviate from the average

$$X - \mathbb{E}[X] \qquad Y - \mathbb{E}[Y]$$

Covariance

Now in general, if X and Y are two random variables, and we want to measure how much (on average) a change in X affects a change in Y, we can do the following:

- Take X and Y
- Subtract their means to see how much they deviate from the average
- Multiply those deviations
  (so that pos $\times$ pos $=$ pos, neg $\times$ neg $=$ pos, and pos $\times$ neg $=$ neg)

$$(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])$$

Covariance

Now in general, if X and Y are two random variables, and we want to measure how much (on average) a change in X affects a change in Y, we can do the following:

- Take X and Y
- Subtract their means to see how much they deviate from the average
- Multiply those deviations
  (so that pos $\times$ pos $=$ pos, neg $\times$ neg $=$ pos, and pos $\times$ neg $=$ neg)
- Take the average of those products

$$\mathbb{E}\big[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\big]$$

Covariance

Now in general, if X and Y are two random variables, and we want to measure how much (on average) a change in X affects a change in Y, we can do the following:

- Take X and Y
- Subtract their means to see how much they deviate from the average
- Multiply those deviations
  (so that pos $\times$ pos $=$ pos, neg $\times$ neg $=$ pos, and pos $\times$ neg $=$ neg)
- Take the average of those products

$$\mathbb{E}\big[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\big]$$

Covariance

Now in general, if X and Y are two random variables, and we want to measure how much (on average) a change in X affects a change in Y, we can do the following:

- Take X and Y
- Subtract their means to see how much they deviate from the average
- Multiply those deviations
  (so that pos $\times$ pos $=$ pos, neg $\times$ neg $=$ pos, and pos $\times$ neg $=$ neg)
- Take the average of those products

$$\mathbb{E}\left[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\right]$$

This number is called the **covariance** between X and Y, and it measures how much the changes in X and Y are related to each other.

Covariance

Definition

The *covariance* between random variables X and Y is defined by

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

Definition

The *covariance* between random variables X and Y is defined by

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

Can we get a simpler formula for covariance?

Covariance

Definition

The *covariance* between random variables X and Y is defined by

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

Can we get a simpler formula for covariance?

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

Covariance

Definition

The *covariance* between random variables X and Y is defined by

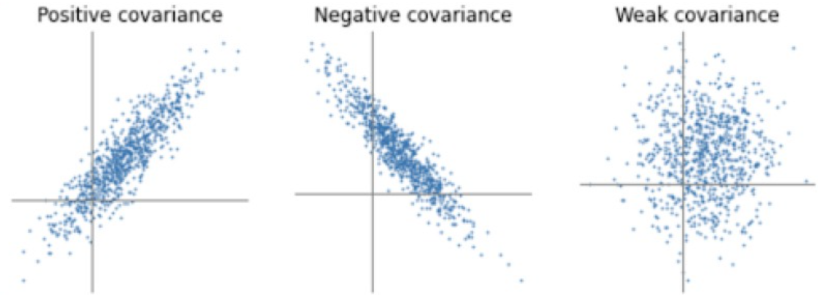
$$\text{Cov}[X, Y] = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]$$

Can we get a simpler formula for covariance?

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

The most important part about covariance is its **sign.**

Covariance



Positive covariance
Negative covariance
Weak covariance

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

- If X and Y are *independent*, we have: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, so the last term becomes zero:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] \qquad \textit{(if } X, Y \textit{ are independent)}$$

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

- If X and Y are *independent*, we have: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, so the last term becomes zero:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] \qquad \textit{(if } X, Y \textit{ are independent)}$$

- If Y is totally dependent on X, e.g. when $Y = 2X$, then:

  $\text{Var}[X+Y] =$

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

- If X and Y are *independent*, we have: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, so the last term becomes zero:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] \qquad \textit{(if } X, Y \textit{ are independent)}$$

- If Y is totally dependent on X, e.g. when $Y = 2X$, then:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

- If X and Y are *independent*, we have: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, so the last term becomes zero:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] \qquad \textit{(if } X, Y \textit{ are independent)}$$

- If Y is totally dependent on X, e.g. when $Y = 2X$, then:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[2X] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

- If X and Y are *independent*, we have: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, so the last term becomes zero:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \qquad \textit{(if } X, Y \textit{ are independent)}$$

- If Y is totally dependent on X, e.g. when $Y = 2X$, then:

$$\text{Var}[X+Y] = \text{Var}[X]+4\cdot\text{Var}[X]+2\cdot(\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

- If X and Y are *independent*, we have: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, so the last term becomes zero:

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] \qquad \textit{(if } X, Y \textit{ are independent)}$$

- If Y is totally dependent on X, e.g. when $Y = 2X$, then:

$$\text{Var}[X+Y] = 5 \cdot \text{Var}[X] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

- If X and Y are *independent*, we have: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, so the last term becomes zero:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \qquad \textit{(if } X, Y \textit{ are independent)}$$

- If Y is totally dependent on X, e.g. when $Y = 2X$, then:

$$\text{Var}[X + Y] = 5 \cdot \text{Var}[X] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]) = 9 \cdot \text{Var}[X]$$

Covariance – alternative definition (optional)

Intuitively, we want covariance to measure the "dependence" between two random variables.

Then why do we define covariance as $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$?

Recall:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

- If X and Y are *independent*, we have: $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$, so the last term becomes zero:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \qquad \textit{(if } X, Y \textit{ are independent)}$$

- If Y is totally dependent on X, e.g. when $Y = 2X$, then:

$$\text{Var}[X + Y] = 5 \cdot \text{Var}[X] + 2 \cdot (\mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]) = 9 \cdot \text{Var}[X]$$

so the last term is pretty large.

Covariance – alternative definition (optional)

In other words, you have

- $\text{Var}[X+Y]$ is far from being equal to $\text{Var}[X]+\text{Var}[Y]$ when X and Y are highly dependent,
- $\text{Var}[X+Y]$ is exactly equal to $\text{Var}[X]+\text{Var}[Y]$ when X and Y are independent.

Covariance – alternative definition (optional)

In other words, you have

- $\text{Var}[X + Y]$ is far from being equal to $\text{Var}[X] + \text{Var}[Y]$ when X and Y are highly dependent,
- $\text{Var}[X + Y]$ is exactly equal to $\text{Var}[X] + \text{Var}[Y]$ when X and Y are independent.

So the **difference** between

$$\text{Var}[X + Y] - \Big(\text{Var}[X] + \text{Var}[Y]\Big)$$

is a good measure of how much X and Y are dependent.

Covariance – alternative definition (optional)

In other words, you have

- $\text{Var}[X+Y]$ is far from being equal to $\text{Var}[X] + \text{Var}[Y]$ when X and Y are highly dependent,
- $\text{Var}[X+Y]$ is exactly equal to $\text{Var}[X] + \text{Var}[Y]$ when X and Y are independent.

So the **difference** between

$$\text{Var}[X+Y] - \Big(\text{Var}[X] + \text{Var}[Y]\Big)$$

is a good measure of how much X and Y are dependent.

But this difference is exactly equal to $\text{Cov}[X, Y]$ (times 2)!

Covariance

Properties

1. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$

Covariance

Properties

1. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
2. $\text{Cov}[a \cdot X, Y] = a \cdot \text{Cov}[X, Y]$

Covariance

Properties

1. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
2. $\text{Cov}[a \cdot X, Y] = a \cdot \text{Cov}[X, Y]$
3. $\text{Cov}[X + Z, Y] = \text{Cov}[X, Y] + \text{Cov}[Z, Y]$

Covariance

Properties

1. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
2. $\text{Cov}[a \cdot X, Y] = a \cdot \text{Cov}[X, Y]$
3. $\text{Cov}[X + Z, Y] = \text{Cov}[X, Y] + \text{Cov}[Z, Y]$
4. $\text{Var}\left[X_1 + X_2\right] = \text{Var}[X_1] + \text{Var}[X_2] + 2 \cdot \text{Cov}[X_1, X_2]$

Covariance

Properties

1. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
2. $\text{Cov}[a \cdot X, Y] = a \cdot \text{Cov}[X, Y]$
3. $\text{Cov}[X + Z, Y] = \text{Cov}[X, Y] + \text{Cov}[Z, Y]$
4. $\text{Var}\,[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] + 2 \cdot \text{Cov}[X_1, X_2]$
5. $\text{Cov}[X, X] = \text{Var}[X]$

Covariance

Properties

1. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
2. $\text{Cov}[a \cdot X, Y] = a \cdot \text{Cov}[X, Y]$
3. $\text{Cov}[X + Z, Y] = \text{Cov}[X, Y] + \text{Cov}[Z, Y]$
4. $\text{Var}\left[X_1 + X_2\right] = \text{Var}[X_1] + \text{Var}[X_2] + 2 \cdot \text{Cov}[X_1, X_2]$
5. $\text{Cov}[X, X] = \text{Var}[X]$
6. And of course, if X and Y are independent, $\text{Cov}[X, Y] = 0$ (but the converse isn't true).

Covariance

Example

We toss 3 coins. Let X denote the number of heads and Y the number of tails. Let's compute $\text{Cov}[X, Y]$.

Example

We toss 3 coins. Let X denote the number of heads and Y the number of tails. Let's compute $\mathrm{Cov}[X, Y]$.

The possible values of (X, Y) are

Covariance

Example

We toss 3 coins. Let X denote the number of heads and Y the number of tails. Let's compute $\text{Cov}[X, Y]$.

The possible values of (X, Y) are $(0, 3)$, $(1, 2)$, $(2, 1)$, $(3, 0)$.

Covariance

Example

We toss 3 coins. Let X denote the number of heads and Y the number of tails. Let's compute $\mathrm{Cov}[X, Y]$.

The possible values of (X, Y) are $(0, 3)$, $(1, 2)$, $(2, 1)$, $(3, 0)$. For example,

$$\mathbb{P}[(X = 1) \text{ and } (Y = 2)] = \, ?$$

Covariance

Example

We toss 3 coins. Let X denote the number of heads and Y the number of tails. Let's compute $\text{Cov}[X, Y]$.

The possible values of (X, Y) are $(0, 3)$, $(1, 2)$, $(2, 1)$, $(3, 0)$. For example,

$$\mathbb{P}[(X = 1) \text{ and } (Y = 2)] = \,?$$

There are 3 possible subcases: `HTT`, `THT`, and `TTH`:

Covariance

Example

We toss 3 coins. Let X denote the number of heads and Y the number of tails. Let's compute $\mathrm{Cov}[X, Y]$.

The possible values of (X, Y) are $(0, 3)$, $(1, 2)$, $(2, 1)$, $(3, 0)$. For example,

$$\mathbb{P}[(X = 1) \text{ and } (Y = 2)] = ?$$

There are 3 possible subcases: HTT, THT, and TTH:

$$\mathbb{P}[\mathtt{HTT}] = \mathbb{P}[\mathtt{THT}] = \mathbb{P}[\mathtt{TTH}] = \frac{1}{8}$$

Covariance

Example

We toss 3 coins. Let X denote the number of heads and Y the number of tails. Let's compute $\mathrm{Cov}[X, Y]$.

The possible values of (X, Y) are $(0, 3)$, $(1, 2)$, $(2, 1)$, $(3, 0)$. For example,

$$\mathbb{P}[(X = 1) \text{ and } (Y = 2)] = ?$$

There are 3 possible subcases: HTT, THT, and TTH:

$$\mathbb{P}[\mathtt{HTT}] = \mathbb{P}[\mathtt{THT}] = \mathbb{P}[\mathtt{TTH}] = \frac{1}{8}$$

hence

$$\mathbb{P}[(X = 1) \text{ and } (Y = 2)] = \frac{3}{8}$$

Example

We toss 3 coins. Let X denote the number of heads and Y the number of tails. Let's compute $\text{Cov}[X, Y]$.

The possible values of (X, Y) are $(0, 3)$, $(1, 2)$, $(2, 1)$, $(3, 0)$. For example,

$$\mathbb{P}[(X = 1) \text{ and } (Y = 2)] = ?$$

There are 3 possible subcases: HTT, THT, and TTH:

$$\mathbb{P}[\mathtt{HTT}] = \mathbb{P}[\mathtt{THT}] = \mathbb{P}[\mathtt{TTH}] = \frac{1}{8}$$

hence

$$\mathbb{P}[(X = 1) \text{ and } (Y = 2)] = \frac{3}{8}$$

$$\mathbb{P}[(X = 2) \text{ and } (Y = 1)] = \frac{3}{8}$$

Example

X	Y	with probability
0	3	$1/8$
1	2	$3/8$
2	1	$3/8$
3	0	$1/8$

$$\mathbb{E}[X] = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5 = \mathbb{E}[Y]$$

Covariance

Example

$X - \mathbb{E}[X]$	$Y - \mathbb{E}[Y]$	with probability
-1.5	1.5	$1/8$
-0.5	0.5	$3/8$
0.5	-0.5	$3/8$
1.5	-1.5	$1/8$

$$\mathbb{E}[X] = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5 = \mathbb{E}[Y]$$

Now we can compute the covariance:

$$\mathbb{P}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = -2.25] = \frac{2}{8}$$
$$\mathbb{P}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = -0.25] = \frac{6}{8}$$

Example

$X - \mathbb{E}[X]$	$Y - \mathbb{E}[Y]$	with probability
-1.5	1.5	$1/8$
-0.5	0.5	$3/8$
0.5	-0.5	$3/8$
1.5	-1.5	$1/8$

$$\mathbb{E}[X] = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5 = \mathbb{E}[Y]$$

Now we can compute the covariance:

$$\mathbb{P}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = -2.25] = \frac{2}{8}$$
$$\mathbb{P}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = -0.25] = \frac{6}{8}$$

so $\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = -0.75$.

Covariance

Example

$X - \mathbb{E}[X]$	$Y - \mathbb{E}[Y]$	with probability
-1.5	1.5	$1/8$
-0.5	0.5	$3/8$
0.5	-0.5	$3/8$
1.5	-1.5	$1/8$

$$\mathbb{E}[X] = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5 = \mathbb{E}[Y]$$

Now we can compute the covariance:

$$\mathbb{P}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = -2.25] = \frac{2}{8}$$
$$\mathbb{P}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = -0.25] = \frac{6}{8}$$

so $\text{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = -0.75$. So what?

Correlation

As you have seen, covariance can be any number, either large or small, and while

- its sign is important

Correlation

As you have seen, covariance can be any number, either large or small, and while

- its sign is important
- its magnitude is not very useful

Correlation

As you have seen, covariance can be any number, either large or small, and while

- its sign is important
- its magnitude is not very useful

Correlation

As you have seen, covariance can be any number, either large or small, and while

- its sign is important
- its magnitude is not very useful

What does this number $\text{Cov}[X, Y] = -0.75$ mean? Is that a large covariance or a small one?

Correlation

As you have seen, covariance can be any number, either large or small, and while

- its sign is important
- its magnitude is not very useful

What does this number $\text{Cov}[X, Y] = -0.75$ mean? Is that a large covariance or a small one?

To answer this question, we need to **normalize** the covariance.

Correlation

As you have seen, covariance can be any number, either large or small, and while

- its sign is important
- its magnitude is not very useful

What does this number $\text{Cov}[X, Y] = -0.75$ mean? Is that a large covariance or a small one?

To answer this question, we need to **normalize** the covariance.

Definition

For two random variables X and Y, the *Pearson correlation coefficient* or just *correlation* between them is defined as:

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]} \cdot \sqrt{\text{Var}[Y]}}$$

Correlation

As you have seen, covariance can be any number, either large or small, and while

- its sign is important
- its magnitude is not very useful

What does this number $\text{Cov}[X, Y] = -0.75$ mean? Is that a large covariance or a small one?

To answer this question, we need to **normalize** the covariance.

Definition

For two random variables X and Y, the *Pearson correlation coefficient* or just *correlation* between them is defined as:

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]} \cdot \sqrt{\text{Var}[Y]}}$$

The value of correlation is **always between** -1 **and** 1**.**

Correlation

Properties

1. $\rho_{X,Y} = \rho_{Y,X}$
2. $-1 \leq \rho_{X,Y} \leq 1$

Correlation

Properties

1. $\rho_{X,Y} = \rho_{Y,X}$
2. $-1 \leq \rho_{X,Y} \leq 1$

When does correlation take its maximum or minimum values, i.e. when is $\rho_{X,Y} = 1$ or -1?

Correlation

Properties

1. $\rho_{X,Y} = \rho_{Y,X}$
2. $-1 \leq \rho_{X,Y} \leq 1$

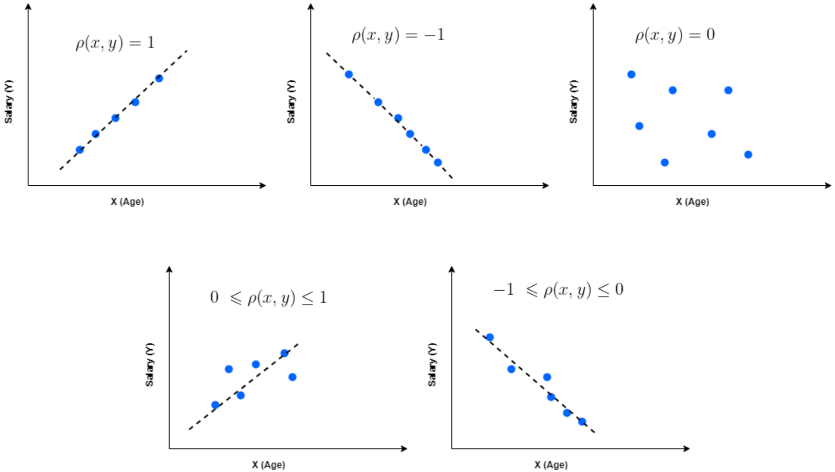
When does correlation take its maximum or minimum values, i.e. when is $\rho_{X,Y} = 1$ or -1?

Whenever there is a **linear relationship** between X and Y:

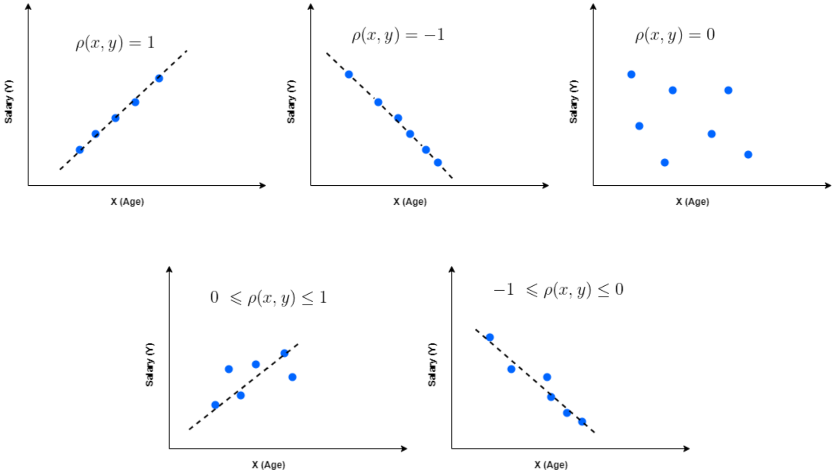
- $Y = X$
- or $Y = -X$
- or $Y = 2X + 3$

i.e. when $Y = aX + b$ for some constants a and b.

Correlation

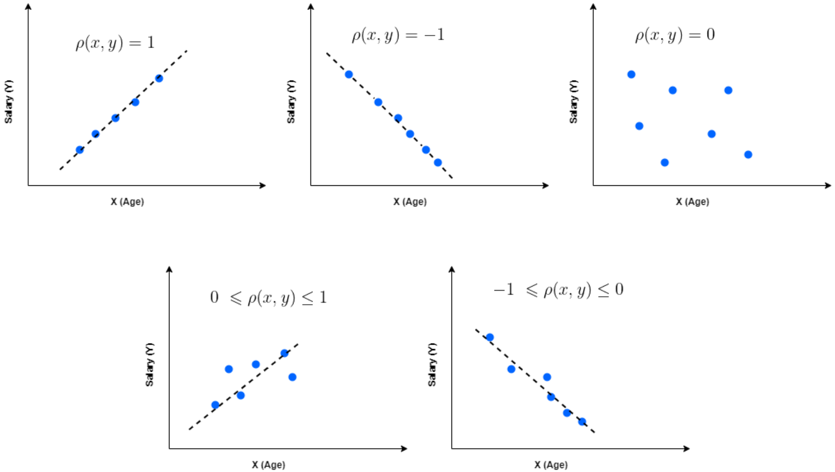


Correlation

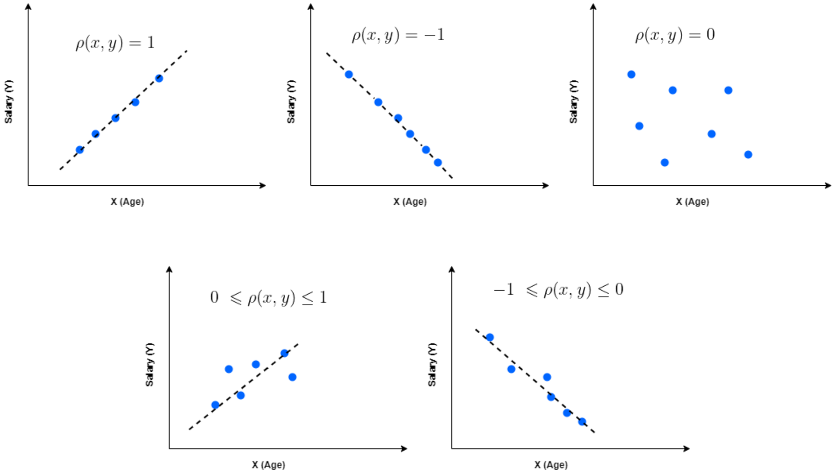


- Play with this correlation visualization!

Correlation



Correlation



- Play with this correlation visualization!

Correlation

In practice, instead of knowing the distributions of X and Y, we often have samples of their values, i.e. some data consisting of n rows and 2 columns:

X	Y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

Correlation

In practice, instead of knowing the distributions of X and Y, we often have samples of their values, i.e. some data consisting of n rows and 2 columns:

X	Y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

Then the sample correlation between X and Y is computed as:

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of X and Y.

Drawbacks of correlation

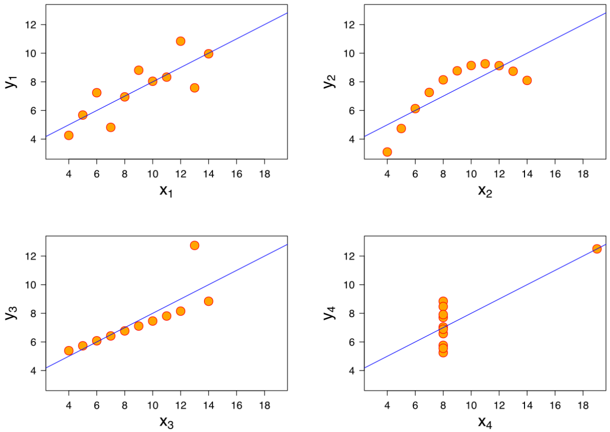
Warning 1

One number doesn't tell the whole story – it cannot capture the full relationship between two random variables.

Drawbacks of correlation

Warning 1

One number doesn't tell the whole story – it cannot capture the full relationship between two random variables.



Drawbacks of correlation

Warning 2

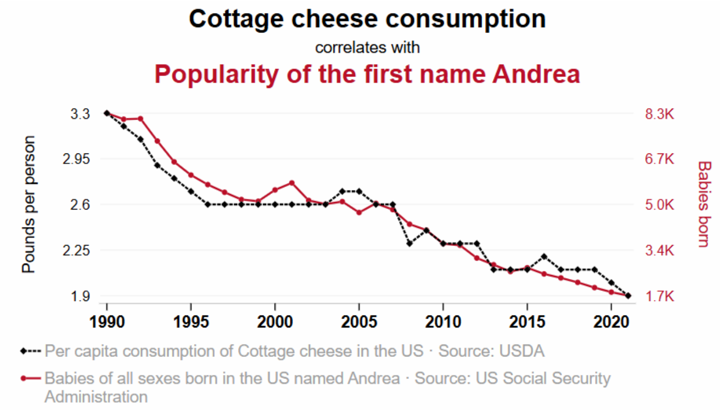
Correlation $\neq$ causation!
Just because two random variables have high correlation, it doesn't mean that one causes the other.

Drawbacks of correlation

Warning 2

Correlation $\neq$ causation!

Just because two random variables have high correlation, it doesn't mean that one causes the other.



Drawbacks of correlation

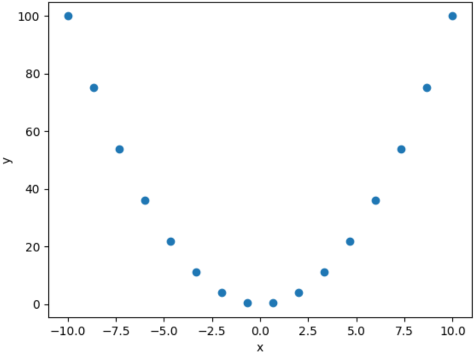
Warning 3

Correlation and covariation only measure **linear relationships** between random variables – two random variables can be very dependent in a non-linear way, but have low correlation.

Drawbacks of correlation

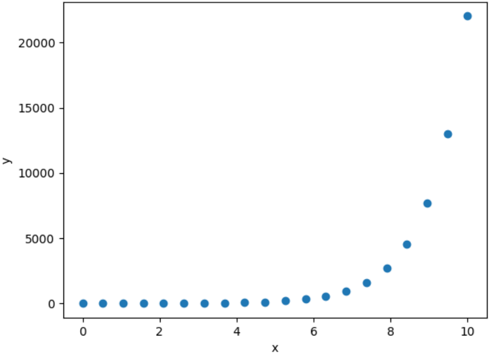
Warning 3

Correlation and covariation only measure **linear relationships** between random variables – two random variables can be very dependent in a non-linear way, but have low correlation.



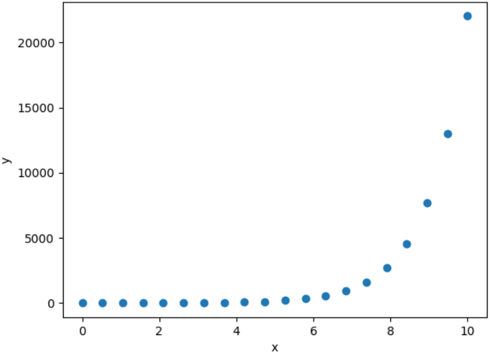
Here $Y = X^2$.
Yet, the correlation between X and Y is just 0!

Drawbacks of correlation



And here $Y = e^X$.
Yet, the correlation between X and Y is only ≈ 0.7!

Drawbacks of correlation

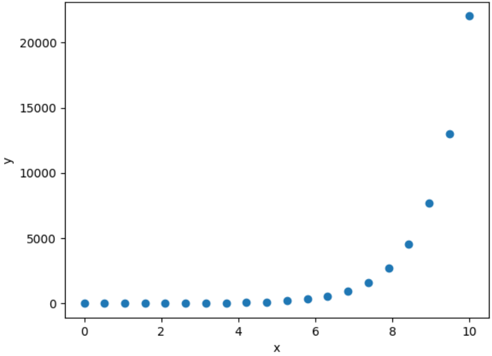


And here $Y = e^X$.
Yet, the correlation between X and Y is only ≈ 0.7!

So

- having high correlation $\Rightarrow$ there is some linear relationship
- having low correlation $\nRightarrow$ the two random variables are independent

Drawbacks of correlation



And here $Y = e^X$.
Yet, the correlation between X and Y is only ≈ 0.7!

So

- having high correlation $\Rightarrow$ there is some linear relationship
- having low correlation $\not\Rightarrow$ the two random variables are independent

Can we measure dependence between random variables in another way?

Spearman's Rank Correlation Coefficient

One way to measure whether

$$X \text{ increases} \quad \Rightarrow \quad Y \text{ increases}$$

is to rank the values of X and Y:

1. sort the values of X in increasing order and assign ranks to them (the smallest value gets rank 1, the second smallest gets rank 2, etc.)
2. do the same for Y
3. compute the Pearson correlation coefficient between these "ranks"

Spearman's Rank Correlation Coefficient

One way to measure whether

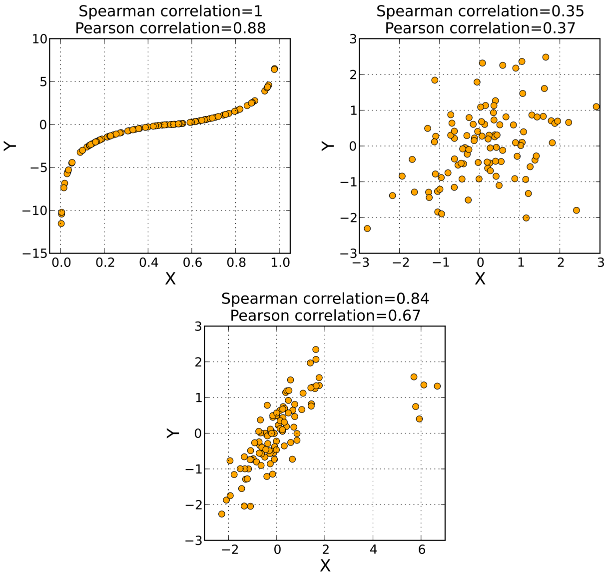
$$X \text{ increases} \quad \Rightarrow \quad Y \text{ increases}$$

is to rank the values of X and Y:

1. sort the values of X in increasing order and assign ranks to them (the smallest value gets rank 1, the second smallest gets rank 2, etc.)
2. do the same for Y
3. compute the Pearson correlation coefficient between these "ranks"

This is called *Spearman's rank correlation coefficient*, and it measures the **monotonic** relationship between two random variables, i.e. whether one variable tends to increase when the other increases.

Spearman's Rank Correlation Coefficient



Miss me?

Question

Have you noticed how this Cov$[X, Y]$ and $\rho_{X,Y}$ stuff looks like something from linear algebra?

Miss me?

Question

Have you noticed how this Cov$[X, Y]$ and $\rho_{X,Y}$ stuff looks like something from linear algebra?

That's because actually, Cov$[X, Y]$ is the "dot product" of the random variables X and Y, and $\rho_{X,Y}$ is the $\cos(\theta)$ between them!

Miss me?

Question

Have you noticed how this $\text{Cov}[X, Y]$ and $\rho_{X,Y}$ stuff looks like something from linear algebra?

That's because actually, $\text{Cov}[X, Y]$ is the "dot product" of the random variables X and Y, and $\rho_{X,Y}$ is the $\cos(\theta)$ between them!

- **Cauchy–Schwarz:** $|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]} \cdot \sqrt{\text{Var}[Y]}$

Miss me?

Question

Have you noticed how this $\text{Cov}[X, Y]$ and $\rho_{X,Y}$ stuff looks like something from linear algebra?

That's because actually, $\text{Cov}[X, Y]$ is the "dot product" of the random variables X and Y, and $\rho_{X,Y}$ is the $\cos(\theta)$ between them!

- **Cauchy–Schwarz:** $|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]} \cdot \sqrt{\text{Var}[Y]}$ (and equality holds if and only if X and Y are "on the same line", i.e. $Y = aX + b$)

Miss me?

Question

Have you noticed how this $\text{Cov}[X, Y]$ and $\rho_{X,Y}$ stuff looks like something from linear algebra?

That's because actually, $\text{Cov}[X, Y]$ is the "dot product" of the random variables X and Y, and $\rho_{X,Y}$ is the $\cos(\theta)$ between them!

- **Cauchy–Schwarz:** $|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]} \cdot \sqrt{\text{Var}[Y]}$ (and equality holds if and only if X and Y are "on the same line", i.e. $Y = aX + b$)
- **Orthogonal = Uncorrelated $\approx$ Independent:**

$$X \text{ and } Y \text{ independent} \implies \text{Cov}[X, Y] = 0$$

Miss me?

Question

Have you noticed how this $\text{Cov}[X, Y]$ and $\rho_{X,Y}$ stuff looks like something from linear algebra?

That's because actually, $\text{Cov}[X, Y]$ is the "dot product" of the random variables X and Y, and $\rho_{X,Y}$ is the $\cos(\theta)$ between them!

- **Cauchy–Schwarz:** $|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]} \cdot \sqrt{\text{Var}[Y]}$
  (and equality holds if and only if X and Y are "on the same line", i.e. $Y = aX + b$)
- **Orthogonal = Uncorrelated $\approx$ Independent:**

$$X \text{ and } Y \text{ independent } \implies \text{Cov}[X, Y] = 0$$

- **Triangle Inequality:** $\sqrt{\text{Var}[X + Y]} \leq \sqrt{\text{Var}[X]} + \sqrt{\text{Var}[Y]}$