

Applied Data Analytics

Statistics — Measures for bivariate data

Covariances

Hans-Martin von Gaudecker and Aapo Stenhammar

(Co)variance

Variance:

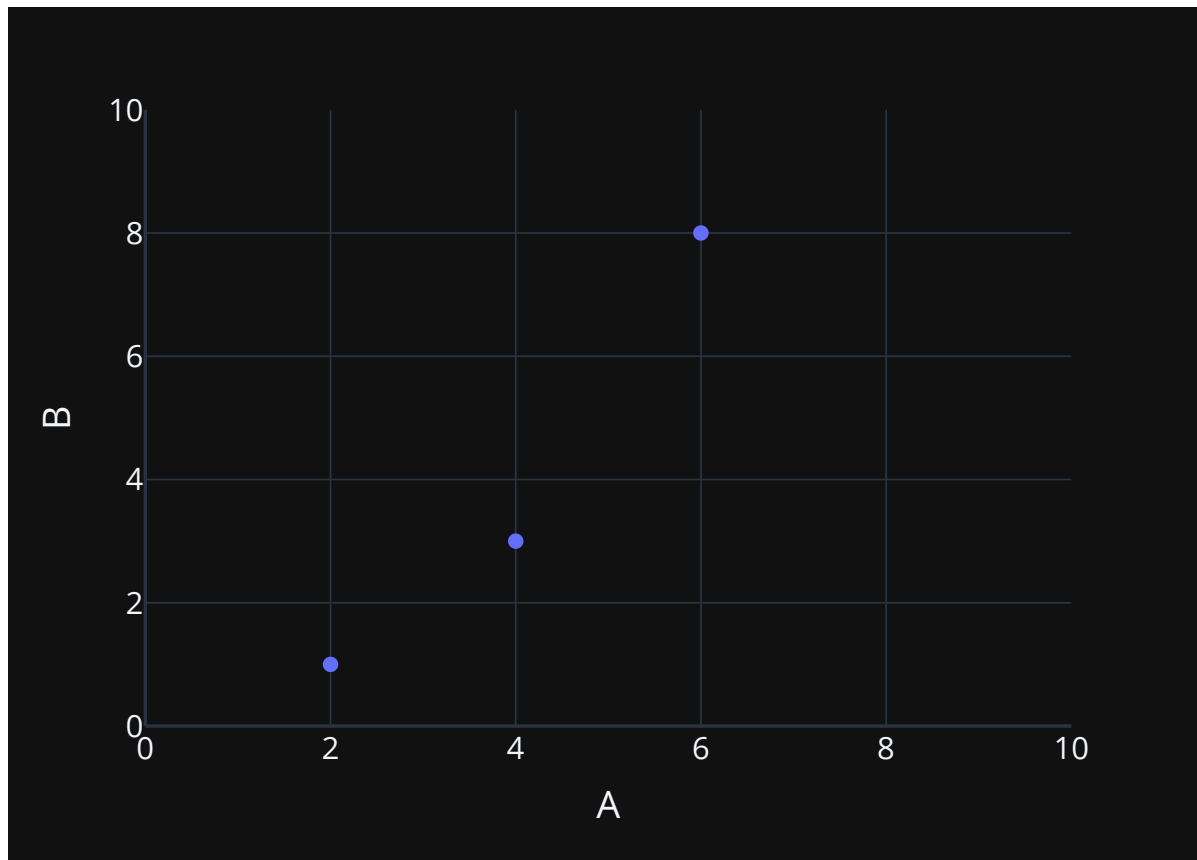
$$\begin{aligned}s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (x_i - \bar{x})\end{aligned}$$

Covariance:

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

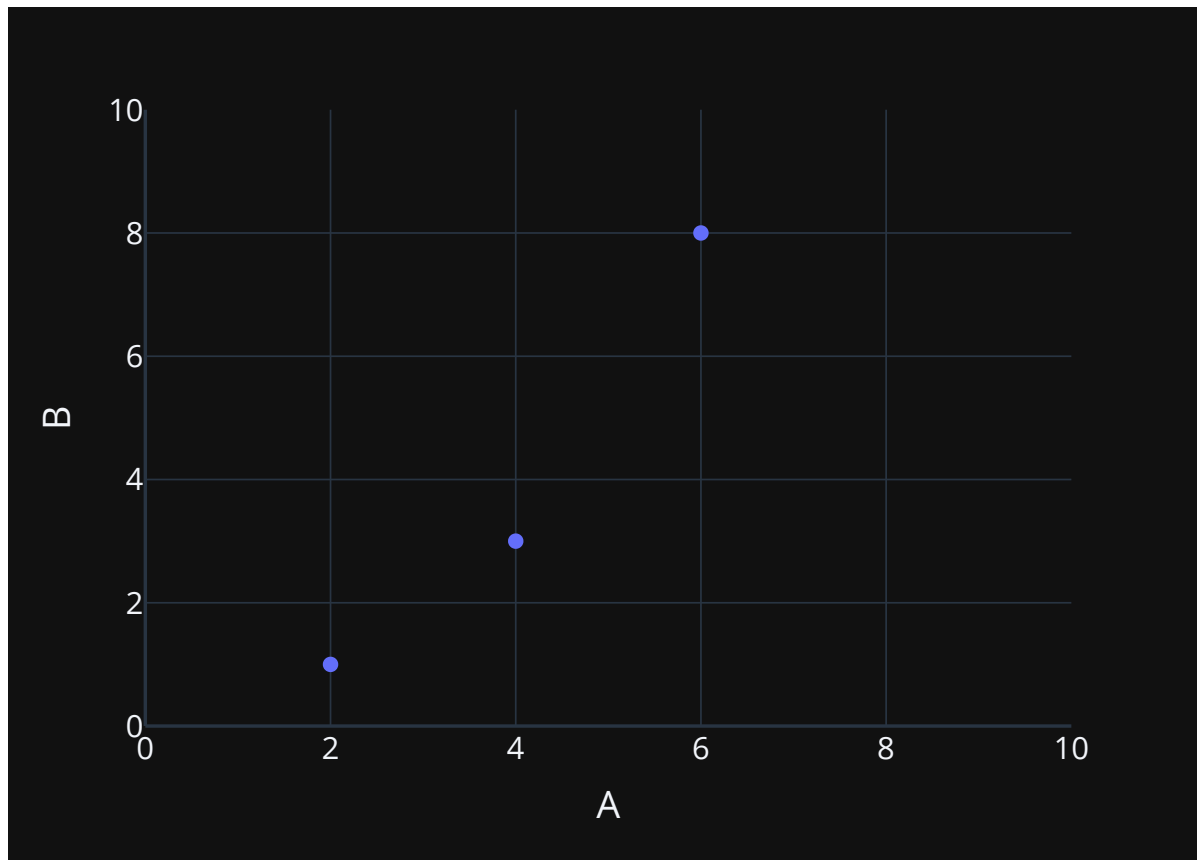
Example

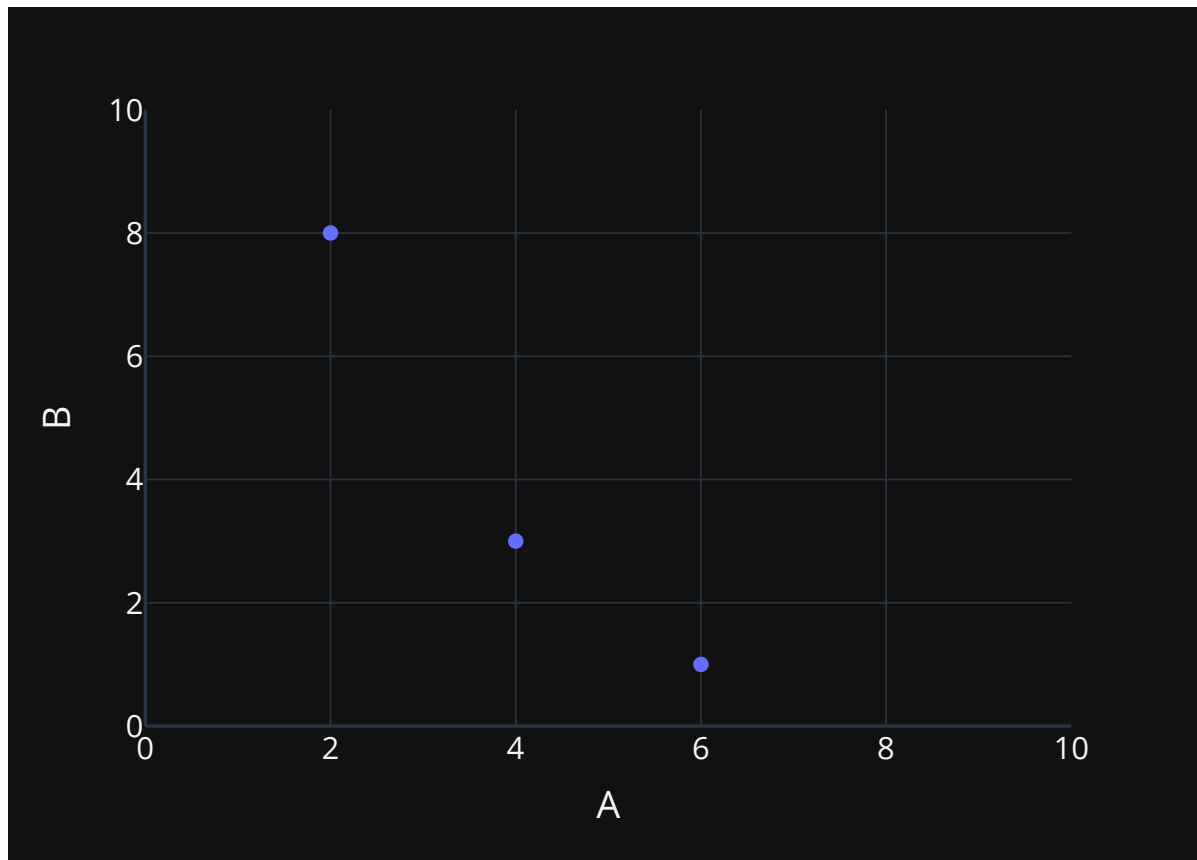
A	B
2	1
4	3
6	8



Positive covariance

A	B	A - 4	B - 4	(A - 4)(B - 4)
2	1	-2	-3	6
4	3	0	-1	0
6	8	2	4	8





Negative covariance

A	B	A - 4	B - 4	(A - 4)(B - 4)
2	8	-2	4	-8
4	3	0	-1	0
6	1	2	-3	-6

Magnitude is difficult to interpret

- Covariance is not standardized
- Multiply **A** by 10, covariance goes up by a factor of 10
- Check whether positive, negative, or zero
- Conceptually important, also as a building block!