COVARIENCE.

$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

→ n → no: of data points
→ $n-1$ → degree of freedom (using sample data rather than population)
→ X_i & Y_i are individual datapoints of variables X & Y respectively.
→ $\bar{X}$ & $\bar{Y}$ are the means of variables X & Y respectively.

$x = [100, 200, 137, 457, 890, 345]$
$y = [345, 678, 123, 56, 890, 500, 120, 230]$

Step-1 → Truncate y to make its length same as x.

$y = [345, 678, 123, 56, 890, 500]$

Step-2 → Mean of x & y.

$$\bar{x} = \frac{100+200+137+457+890+345}{6} = 354.8333$$

$$\bar{y} = \frac{345+678+123+56+890+500}{6} = 432$$

$\bar{x} = 354.8333$

$\bar{y} = 432.$

Step-3 →

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
100	345	-254.8333	-87	22170.4971
200	678	-154.8333	246	-38088.9918
137	123	-217.8333	-309	67310.4897
457	56	102.1667	-376	-38414.6792
890	890	535.1667	458	245106.349
345	500	-9.8333	68	-668.6644
				257415

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

$$= \frac{257415}{6-1} = \frac{257415}{5}$$

$$= 51483.0001$$

CORRELATION

$$Corr(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} \rightarrow \text{Covariance}$$

σ_x ↙ Std of x

σ_y ↘ Std of y.

Step-1 → Truncate y_i to make its length same as x_i.

Step-2 → find $\bar{x}$ & $\bar{y}$

$$\bar{x} = 354.8333$$
$$\bar{y} = 432.$$

Step-3 → find σ_x & σ_y.

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
100	345	-254.8333	-87	64940.0107	7569
200	678	-154.8333	246	23973.3507	60516
137	123	-217.8333	-309	47451.3465	95481
457	56	102.1667	-376	10438.0345	141376
890	890	535.1667	458	286403.3467	209764
345	500	-9.8333	68	96.6937	4624
				433302.832	519330

$$\sigma_x = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N}} = \sqrt{\frac{433302.832}{6}}$$

$$= 268.7324$$

$$\sigma_y = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{N}} = \sqrt{\frac{519330}{6}}$$

$$= 294.2023.$$

Step-4 ⇒ find covariance

$$cov(x,y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N}$$

(in case of correlation we take N instead (n-1)).

$$= \frac{257415}{6} = 42902.5$$

Step-5 ⇒ find Correlation.

$$corr(x,y) = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

$$= \frac{42902.5}{268.7324 \times 294.2023}$$

$$= \frac{42902.5}{79061.6902}$$

$$= 0.54264587$$