Variance and covariance

Last modified: 2026-09-28 23:45:35 (PDT)

Deviation, error, and noise

Definition 0.1 (Deviation)**.** A **deviation** is the difference between a value and a reference value. For any quantity z and reference value r, the deviation of z from r is:

$$z - r$$

Remark. In probability and statistics, "deviation" usually means deviation from a population mean.
See: Wikipedia: Deviation (statistics)

Example 0.1 (Deviation from a reference weight)**.** A newborn weighing 3,200 g deviates from a reference weight of 3,500 g by $3200 - 3500 = -300$ g.

Definition 0.2 (Deviation from a population or subpopulation mean)**.** The **deviation from the mean** of a random variable Y is its deviation from its expectation:

$$e(Y) \stackrel{\text{def}}{=} Y - \mathrm{E}[Y]$$

For a realized observation y:

$$e(y) \stackrel{\text{def}}{=} y - \mathrm{E}[Y]$$

Remark. Other sources often call this quantity an **error** or **noise term**. In regression settings, the reference mean is often conditional on covariates: $e(y_i) \stackrel{\text{def}}{=} y_i - \mathrm{E}[Y_i \mid X_i = x_i]$. These notes prefer "deviation" for this mean-deviation quantity; "error" and "noise" are common aliases. The Morrison Lab's regression notes use "residual" (defined in their Linear regression chapter) for deviations from fitted values, write $e(\cdot)$ for these model/data

deviations, and reserve $\varepsilon(\cdot)$ for estimator-to-estimand deviations (see Estimation).
See:

- Wikipedia: Errors and residuals
- Wikipedia: Deviation (statistics)
- Wikipedia: Linear regression — Notation and terminology

Example 0.2 (Deviation of a die roll from its mean)**.** A fair die roll Y has $\mathrm{E}[Y] = 3.5$ (computed on the expectation page), so a roll of $y = 5$ has deviation $e(5) = 5 - 3.5 = 1.5$, and a roll of $y = 2$ has deviation $e(2) = 2 - 3.5 = -1.5$.

Variance and related characteristics

Definition 0.3 (Variance)**.** The **variance** of a random variable X with $\mathrm{E}[X^2] < \infty$ is the expectation of the squared deviation from the mean; that is:

$$\mathrm{Var}(X) \stackrel{\text{def}}{=} \mathrm{E}[[e(X)]^2]$$

Theorem 0.1 (Variance as expected squared deviation from the mean)**.**

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2]$$

Proof

Proof. Substituting the definition of $e(X)$ from Definition 0.2 into Definition 0.3:

$$\mathrm{Var}(X) \stackrel{\text{def}}{=} \mathrm{E}[[e(X)]^2] = \mathrm{E}[(X - \mathrm{E}[X])^2].$$

□

Remark. This is the more commonly seen form of the variance definition, and the canonical starting point in most treatments; it's given here as a theorem rather than the primary definition to keep $e(X)$ as the definition's single building block, matching the notation of Definition 0.2.

Theorem 0.2 (Simplified expression for variance)**.**

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$$

Proof

Proof. By linearity of expectation, with $\mathrm{E}[X]$ treated as a constant:

$$\begin{aligned}
\mathrm{Var}(X) &\overset{\text{def}}{=} \mathrm{E}[[e(X)]^2] && \text{(definition of variance)}\\
&= \mathrm{E}\big[(X - \mathrm{E}[X])^2\big] && \text{(definition of deviation from mean)}\\
&= \mathrm{E}\Big[X^2 - 2X\,\mathrm{E}[X] + (\mathrm{E}[X])^2\Big] && \text{(expand binomial square)}\\
&= \mathrm{E}\big[X^2\big] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + (\mathrm{E}[X])^2 && \text{(linearity of expectation; } \mathrm{E}[X] \text{ is a constant)}\\
&= \mathrm{E}\big[X^2\big] - (\mathrm{E}[X])^2 && \text{(combine like terms)}
\end{aligned}$$

□

Example 0.3 (Variance of a Bernoulli random variable)**.** Let $X \sim \mathrm{Ber}(\pi)$. Then $\mathrm{E}[X] = \pi$ and $\mathrm{E}[X^2] = \pi$ (both computed on the expectation page), so:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\big[X^2\big] - (\mathrm{E}[X])^2 && \text{(simplified expression for variance)}\\
&= \pi - \pi^2 && \text{(substitute)}\\
&= \pi(1-\pi) && \text{(factor)}
\end{aligned}$$

For a fair coin ($\pi = 1/2$), $\mathrm{Var}(X) = 1/4$.

Definition 0.4 (Conditional variance)**.** The **conditional variance** of Y given $X = x$ is the variance of Y under its conditional distribution given $X = x$:

$$\mathrm{Var}(Y \mid X = x) \overset{\text{def}}{=} \mathrm{E}[(Y - \mathrm{E}[Y \mid X = x])^2 \mid X = x]$$

Evaluating this function of x at X gives the random variable $\mathrm{Var}(Y \mid X)$, just as the conditional expectation function $\mathrm{E}[Y \mid X]$ comes from $\mathrm{E}[Y \mid X = x]$.

Example 0.4 (Conditional variance of a binary outcome)**.** Let (X, Y) have the joint PMF $\mathrm{P}(X = 0, Y = 0) = 0.2$, $\mathrm{P}(X = 0, Y = 1) = 0.3$, $\mathrm{P}(X = 1, Y = 0) = 0.1$, $\mathrm{P}(X = 1, Y = 1) = 0.4$ (the table in the expectation page's exercise). Given $X = 0$, Y is binary with $\mathrm{P}(Y = 1 \mid X = 0) = 0.3/0.5 = 0.6$, so $\mathrm{E}[Y \mid X = 0] = 0.6$, and by Definition 0.4:

$$\begin{aligned}
\mathrm{Var}(Y \mid X = 0) &= (0 - 0.6)^2 \cdot 0.4 + (1 - 0.6)^2 \cdot 0.6 && \text{(definition of conditional variance)}\\
&= 0.144 + 0.096 && \text{(multiply)}\\
&= 0.24 && \text{(add)}
\end{aligned}$$

The same steps with $\mathrm{P}(Y = 1 \mid X = 1) = 0.4/0.5 = 0.8$ give $\mathrm{Var}(Y \mid X = 1) = (0.2)^2 \cdot 0.8 + (0.8)^2 \cdot 0.2 = 0.16$.

Definition 0.5 (Homoskedasticity)**.** A random variable Y is **homoskedastic** with respect to a random variable (or vector) X if the conditional variance of Y given $X = x$ is the same for every x:

$$\mathrm{Var}(Y \mid X = x) = \sigma^2 \quad \text{for all } x$$

for some constant σ^2.

Example 0.5 (A homoskedastic total)**.** Flip two fair coins, let X_1 indicate heads on the first, and let Y be the total number of heads. Given $X_1 = 0$, Y is 0 or 1 with probability $1/2$ each; given $X_1 = 1$, Y is 1 or 2 with probability $1/2$ each. Either way, Y is $1/2$ away from its conditional mean with probability 1, so $\mathrm{Var}(Y \mid X_1 = 0) = \mathrm{Var}(Y \mid X_1 = 1) = 1/4$, and Y is homoskedastic with respect to X_1, with $\sigma^2 = 1/4$.

Definition 0.6 (Heteroskedasticity)**.** A random variable Y is **heteroskedastic** with respect to X if it is not homoskedastic with respect to X: that is, if $\mathrm{Var}(Y \mid X = x)$ differs between some values of x.

Example 0.6 (A heteroskedastic binary outcome)**.** In Example 0.4, $\mathrm{Var}(Y \mid X = 0) = 0.24 \neq 0.16 = \mathrm{Var}(Y \mid X = 1)$, so Y is heteroskedastic with respect to X. More generally, a binary outcome with $\mathrm{P}(Y = 1 \mid X = x) = \pi(x)$ has conditional variance $\pi(x)(1 - \pi(x))$, so it is homoskedastic only if $\pi(x)$ takes at most two values, π and $1 - \pi$, for some π.

Theorem 0.3 (Law of total variance)**.** *For random variables X and Y with $\mathrm{E}[Y^2] < \infty$:*

$$\mathrm{Var}(Y) = \mathrm{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathrm{E}[Y \mid X])$$

Proof

Proof. Write $\mu \overset{\text{def}}{=} \mathrm{E}[Y]$ and $m(X) \overset{\text{def}}{=} \mathrm{E}[Y \mid X]$. Adding and subtracting $m(X)$ inside the deviation and expanding the square:

$$
\begin{aligned}
\operatorname{Var}(Y) &= \mathrm{E}\big[(Y-\mu)^2\big] && \text{(definition of variance)}\\
&= \mathrm{E}\big[([Y-m(X)]+[m(X)-\mu])^2\big] && \text{(add and subtract } m(X)\text{)}\\
&= \mathrm{E}\Big[(Y-m(X))^2\Big] + \mathrm{E}\Big[(m(X)-\mu)^2\Big] + 2\,\mathrm{E}[[Y-m(X)][m(X)-\mu]] && \text{(expand the square; linearity}
\end{aligned}
$$

The three terms are handled separately, each using the law of iterated expectations, applied to a function of X and Y (conditional expectation of a function). The third term also uses two facts about conditional expectations: a function of X factors out of a conditional expectation given X, and conditional expectation is linear.
First term:

$$
\begin{aligned}
\mathrm{E}\Big[(Y-m(X))^2\Big] &= \mathrm{E}\Big[\mathrm{E}\Big[(Y-m(X))^2 \mid X\Big]\Big] && \text{(law of iterated expectations)}\\
&= \mathrm{E}[\operatorname{Var}(Y \mid X)] && \text{(definition of conditional variance)}
\end{aligned}
$$

Second term: by the law of iterated expectations, $\mathrm{E}[m(X)] = \mathrm{E}[\mathrm{E}[Y \mid X]] = \mu$, so:

$$
\begin{aligned}
\mathrm{E}\Big[(m(X)-\mu)^2\Big] &= \mathrm{E}\Big[(m(X)-\mathrm{E}[m(X)])^2\Big] && (\mu = \mathrm{E}[m(X)])\\
&= \operatorname{Var}(m(X)) && \text{(definition of variance)}\\
&= \operatorname{Var}(\mathrm{E}[Y \mid X]) && \text{(definition of } m(X)\text{)}
\end{aligned}
$$

Third term:

$$
\begin{aligned}
\mathrm{E}[[Y-m(X)][m(X)-\mu]] &= \mathrm{E}[\mathrm{E}[[Y-m(X)][m(X)-\mu] \mid X]] && \text{(law of iterated expectations)}\\
&= \mathrm{E}[[m(X)-\mu]\cdot\mathrm{E}[Y-m(X) \mid X]] && \text{(a function of } X \text{ factors out, with } g(\\
&= \mathrm{E}[[m(X)-\mu]\cdot[\mathrm{E}[Y \mid X]-\mathrm{E}[m(X) \mid X]]] && \text{(linearity of conditional expectation)}\\
&= \mathrm{E}[[m(X)-\mu]\cdot[\mathrm{E}[Y \mid X]-m(X)]] && (\,\mathrm{E}[g(X) \mid X] = g(X), \text{ with } g = m)\\
&= \mathrm{E}[[m(X)-\mu]\cdot 0] && (\,\mathrm{E}[Y \mid X] = m(X))\\
&= 0 && \text{(expectation of a constant)}
\end{aligned}
$$

Substituting the three terms:

$$
\begin{aligned}
\operatorname{Var}(Y) &= \mathrm{E}[\operatorname{Var}(Y \mid X)] + \operatorname{Var}(\mathrm{E}[Y \mid X]) + 2\cdot 0 && \text{(substitute the three terms)}\\
&= \mathrm{E}[\operatorname{Var}(Y \mid X)] + \operatorname{Var}(\mathrm{E}[Y \mid X]) && \text{(simplify)}
\end{aligned}
$$

□

Remark. Alternate names include: the **conditional variance formula**, **Eve's law**, and the **variance decomposition formula**.

Example 0.7 (Decomposing the variance of a binary outcome)**.** Continuing Example 0.4, $\mathrm{P}(X=0)=\mathrm{P}(X=1)=0.5$, $\mathrm{E}[Y \mid X=0]=0.6$, and $\mathrm{E}[Y \mid X=1]=0.8$, so $\mathrm{E}[\mathrm{E}[Y \mid X]]=0.7$, and by Theorem 0.3:

$$\begin{aligned}
\mathrm{E}[\mathrm{Var}(Y \mid X)] &= 0.5 \cdot 0.24 + 0.5 \cdot 0.16 = 0.20 && \text{(average the conditional variances)} \\
\mathrm{Var}(\mathrm{E}[Y \mid X]) &= 0.5 \cdot (0.6-0.7)^2 + 0.5 \cdot (0.8-0.7)^2 = 0.01 && \text{(variance of the conditional means)} \\
\mathrm{Var}(Y) &= 0.20 + 0.01 = 0.21 && \text{(law of total variance)}
\end{aligned}$$

As a check, Y is Bernoulli with $\mathrm{P}(Y=1)=0.3+0.4=0.7$, so by Example 0.3, $\mathrm{Var}(Y)=0.7 \cdot 0.3=0.21$.

Definition 0.7 (Precision)**.** The **precision** of a random variable X, often denoted $\tau(X)$, τ_X, or shorthanded as τ, is the inverse of that random variable's variance (for $\mathrm{Var}(X)>0$); that is:

$$\tau(X) \stackrel{\text{def}}{=} (\mathrm{Var}(X))^{-1}$$

Definition 0.8 (Standard deviation)**.** The **standard deviation** of a random variable X is the square root of the variance of X:

$$\mathrm{SD}(X) \stackrel{\text{def}}{=} \sqrt{\mathrm{Var}(X)}$$

Remark. The standard deviation is on the same scale as X itself (unlike the variance, which is on the scale of X's square), which is why it is often the preferred measure of spread when reporting results.

Example 0.8 (Precision and standard deviation of a fair coin flip)**.** In Example 0.3, a fair coin flip has $\mathrm{Var}(X)=1/4$, so its precision is $\tau(X)=1/(1/4)=4$ and its standard deviation is $\mathrm{SD}(X)=\sqrt{1/4}=1/2$.

Covariance

Definition 0.9 (Covariance)**.** The **covariance** of two random variables X and Y with $\mathrm{E}[X^2]<\infty$ and $\mathrm{E}[Y^2]<\infty$ is:

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

Theorem 0.4 (Alternative formula for covariance)**.**

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

Proof

Proof. By linearity of expectation, analogous to the proof of the simplified expression for variance:

$$\begin{aligned}
\operatorname{Cov}(X, Y) &\overset{\text{def}}{=} \operatorname{E}[(X - \operatorname{E}[X])(Y - \operatorname{E}[Y])] && \text{(definition of covariance)} \\
&= \operatorname{E}[XY - X\operatorname{E}[Y] - Y\operatorname{E}[X] + \operatorname{E}[X]\operatorname{E}[Y]] && \text{(expand the product)} \\
&= \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y] - \operatorname{E}[Y]\operatorname{E}[X] + \operatorname{E}[X]\operatorname{E}[Y] && \text{(linearity of expectation; } \operatorname{E}[X], \operatorname{E}[Y] \text{ are constan} \\
&= \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y] && \text{(combine like terms)}
\end{aligned}$$

□

Example 0.9 (Covariance of a binary exposure and outcome)**.** For the joint PMF in Example 0.4, $\operatorname{E}[XY] = \operatorname{P}(X = 1, Y = 1) = 0.4$, $\operatorname{E}[X] = 0.5$, and $\operatorname{E}[Y] = 0.7$, so:

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y] && \text{(alternative formula for covariance)} \\
&= 0.4 - 0.5 \cdot 0.7 && \text{(substitute)} \\
&= 0.05 && \text{(evaluate)}
\end{aligned}$$

Theorem 0.5 (Independent random variables have zero covariance)**.** *If X and Y are independent, each discrete or continuous, with defined expectations, then:*

$$\operatorname{E}[XY] = \operatorname{E}[X]\operatorname{E}[Y]$$

If also $\operatorname{E}[X^2] < \infty$ and $\operatorname{E}[Y^2] < \infty$, then $\operatorname{Cov}(X, Y) = 0$.

Proof

Proof. Write f_X and f_Y for the PMFs or densities of X and Y, with reference measures μ_X and μ_Y as in the joint-distribution form of Fubini–Tonelli (counting measure for a discrete variable, Lebesgue measure for a continuous one). Because X and Y are independent, $f_{X,Y}(x, y) = f_X(x)\, f_Y(y)$ is their joint PMF, density, or density-mass function (the factorization in the notes to the definition of independence). First, with $h(x, y) = |x||y| \geq 0$ (condition (a)):

$$
\begin{aligned}
\mathrm{E}[|XY|] &= \int \left(\int |x||y| \, f_X(x) \, f_Y(y) \, d\mu_Y(y) \right) d\mu_X(x) && \text{(joint-distribution form of Fubini–Tonelli, condition (a)} \\
&= \int |x| \, f_X(x) \left(\int |y| \, f_Y(y) \, d\mu_Y(y) \right) d\mu_X(x) && (|x| \, f_X(x) \text{ does not depend on } y) \\
&= \int |x| \, f_X(x) \cdot \mathrm{E}[|Y|] \, d\mu_X(x) && \text{(LOTUS for } |Y|) \\
&= \mathrm{E}[|X|] \cdot \mathrm{E}[|Y|] && \text{(LOTUS for } |X|)
\end{aligned}
$$

which is finite, because X and Y have defined expectations. So condition (b) holds for $h(x, y) = xy$, and the same steps without the absolute values give:

$$
\begin{aligned}
\mathrm{E}[XY] &= \int \left(\int xy \, f_X(x) \, f_Y(y) \, d\mu_Y(y) \right) d\mu_X(x) && \text{(joint-distribution form of Fubini–Tonelli, condition (b))} \\
&= \int x \, f_X(x) \left(\int y \, f_Y(y) \, d\mu_Y(y) \right) d\mu_X(x) && (x \, f_X(x) \text{ does not depend on } y) \\
&= \int x \, f_X(x) \cdot \mathrm{E}[Y] \, d\mu_X(x) && \text{(definition of expectation)} \\
&= \mathrm{E}[X] \cdot \mathrm{E}[Y] && \text{(definition of expectation)}
\end{aligned}
$$

Finally, by the alternative formula for covariance:

$$
\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{E}[XY] - \mathrm{E}[X] \, \mathrm{E}[Y] && \text{(alternative formula for covariance)} \\
&= \mathrm{E}[X] \, \mathrm{E}[Y] - \mathrm{E}[X] \, \mathrm{E}[Y] && (\mathrm{E}[XY] = \mathrm{E}[X] \, \mathrm{E}[Y]) \\
&= 0 && \text{(subtract)}
\end{aligned}
$$

□

Example 0.10 (Two independent coin flips)**.** For the independent coin flips X_1 and X_2 of the independence page's example, $X_1 X_2 = 1$ only when both flips are heads, so:

$$
\begin{aligned}
\mathrm{E}[X_1 X_2] &= \mathrm{P}(X_1 = 1, X_2 = 1) && (X_1 X_2 \text{ is the indicator of two heads}) \\
&= \tfrac{1}{4} && (\text{each outcome has probability } \tfrac{1}{4}) \\
&= \tfrac{1}{2} \cdot \tfrac{1}{2} && \text{(factor)} \\
&= \mathrm{E}[X_1] \, \mathrm{E}[X_2] && (\text{each flip is } \mathrm{Ber}(1/2))
\end{aligned}
$$

so $\mathrm{Cov}(X_1, X_2) = 0$, as Theorem 0.5 requires.

Example 0.11 (Zero covariance without independence)**.** The converse of Theorem 0.5 is false. Let X take the values -1, 0, and 1 with probability $1/3$ each, and let $Y = X^2$.

Then $\mathrm{E}[X] = (-1+0+1)/3 = 0$, and $XY = X^3 = X$, so:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] && \text{(alternative formula for covariance)} \\
&= \mathrm{E}[X] - \mathrm{E}[X]\,\mathrm{E}[Y] && (XY = X^3 = X \text{ on } \{-1,0,1\}) \\
&= 0 - 0 \cdot \mathrm{E}[Y] && (\mathrm{E}[X] = 0) \\
&= 0 && \text{(multiply)}
\end{aligned}$$

But X and Y are not independent: Y is a function of X, and $\mathrm{P}(X = 0, Y = 0) = \mathrm{P}(X = 0) = \frac{1}{3}$, while $\mathrm{P}(X = 0)\ \mathrm{P}(Y = 0) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$.

Definition 0.10 (Correlation)**.** The **correlation** of two random variables X and Y with finite, positive variances is their covariance divided by the product of their standard deviations:

$$\mathrm{Cor}(X,Y) \stackrel{\text{def}}{=} \frac{\mathrm{Cov}(X,Y)}{\mathrm{SD}(X)\ \mathrm{SD}(Y)}$$

Remark. Dividing by the standard deviations removes the units of X and Y. This population correlation is a property of a joint distribution; the sample (Pearson) correlation coefficient computed from data estimates it.

Example 0.12 (Correlation of a binary exposure and outcome)**.** In Example 0.9, $\mathrm{Cov}(X,Y) = 0.05$, where $X \sim \mathrm{Ber}(0.5)$ has $\mathrm{Var}(X) = 0.25$ (Example 0.3) and Y has $\mathrm{Var}(Y) = 0.21$ (Example 0.7), so:

$$\begin{aligned}
\mathrm{Cor}(X,Y) &= \frac{\mathrm{Cov}(X,Y)}{\mathrm{SD}(X)\ \mathrm{SD}(Y)} && \text{(definition of correlation)} \\
&= \frac{0.05}{\sqrt{0.25} \cdot \sqrt{0.21}} && (\text{substitute; } \mathrm{SD}(\cdot) = \sqrt{\mathrm{Var}(\cdot)}) \\
&\approx \frac{0.05}{0.5 \cdot 0.458} && \text{(evaluate the square roots)} \\
&\approx 0.218 && \text{(divide)}
\end{aligned}$$

Definition 0.11 (Uncorrelated random variables)**.** Random variables X and Y with $\mathrm{E}[X^2] < \infty$ and $\mathrm{E}[Y^2] < \infty$ are **uncorrelated** when their covariance is 0:

$$\mathrm{Cov}(X,Y) = 0$$

Remark. Theorem 0.5 says independent random variables are uncorrelated, and Example 0.11 shows the converse fails.

Example 0.13 (Correlated and uncorrelated pairs)**.** In Example 0.11, $\mathrm{Cov}(X, Y) = 0$, so X and $Y = X^2$ are uncorrelated. In Example 0.9, $\mathrm{Cov}(X, Y) = 0.05 \neq 0$, so the binary exposure and outcome are not uncorrelated.

Corollary 0.1 (Uncorrelated means zero correlation)**.** *If X and Y have finite, positive variances, then X and Y are uncorrelated if and only if* $\mathrm{Cor}(X, Y) = 0$.

Proof

Proof. By Definition 0.10, $\mathrm{Cor}(X, Y) = \mathrm{Cov}(X, Y)/(\mathrm{SD}(X)\,\mathrm{SD}(Y))$, and $\mathrm{SD}(X)\,\mathrm{SD}(Y) > 0$ because both variances are positive, so $\mathrm{Cor}(X, Y) = 0$ if and only if $\mathrm{Cov}(X, Y) = 0$, which is Definition 0.11. □

Definition 0.12 (Conditional covariance)**.** The **conditional covariance** of Y and Z given $X = x$ is their covariance under their conditional distribution given $X = x$:

$$\mathrm{Cov}(Y, Z \mid X = x) \stackrel{\text{def}}{=} \mathrm{E}[(Y - \mathrm{E}[Y \mid X = x])(Z - \mathrm{E}[Z \mid X = x]) \mid X = x]$$

Evaluating this function of x at X gives the random variable $\mathrm{Cov}(Y, Z \mid X)$.

Example 0.14 (Conditional covariance of a variable with itself)**.** Taking $Z = Y$ in Definition 0.12 gives the conditional variance: $\mathrm{Cov}(Y, Y \mid X = x) = \mathrm{Var}(Y \mid X = x)$. In Example 0.4, for example, $\mathrm{Cov}(Y, Y \mid X = 0) = 0.24$.

Theorem 0.6 (Law of total covariance)**.** *For random variables X, Y, and Z with $\mathrm{E}[Y^2] < \infty$ and $\mathrm{E}[Z^2] < \infty$:*

$$\mathrm{Cov}(Y, Z) = \mathrm{E}[\mathrm{Cov}(Y, Z \mid X)] + \mathrm{Cov}(\mathrm{E}[Y \mid X], \mathrm{E}[Z \mid X])$$

Proof

Proof. The proof follows the proof of the law of total variance, which is the special case $Z = Y$. Write $m_Y(X) \stackrel{\text{def}}{=} \mathrm{E}[Y \mid X]$, $m_Z(X) \stackrel{\text{def}}{=} \mathrm{E}[Z \mid X]$, $\mu_Y \stackrel{\text{def}}{=} \mathrm{E}[Y]$, and $\mu_Z \stackrel{\text{def}}{=} \mathrm{E}[Z]$. Adding and subtracting $m_Y(X)$ and $m_Z(X)$ inside the deviations:

$$
\begin{aligned}
\operatorname{Cov}(Y,Z) &= \mathrm{E}[(Y-\mu_Y)(Z-\mu_Z)] && \text{(definition of covariance)}\\
&= \mathrm{E}[([Y-m_Y(X)]+[m_Y(X)-\mu_Y])([Z-m_Z(X)]+[m_Z(X)-\mu_Z])] && \text{(add and subtract)}\\
&= \mathrm{E}[[Y-m_Y(X)][Z-m_Z(X)]] + \mathrm{E}[[Y-m_Y(X)][m_Z(X)-\mu_Z]]\\
&\quad + \mathrm{E}[[m_Y(X)-\mu_Y][Z-m_Z(X)]] + \mathrm{E}[[m_Y(X)-\mu_Y][m_Z(X)-\mu_Z]] && \text{(expand the product; linea}
\end{aligned}
$$

The two middle terms are 0, by the same steps as the cross term in the proof of Theorem 0.3: condition on X by the law of iterated expectations, factor out the function of X (pull-out property), and use $\mathrm{E}[Y-m_Y(X)\mid X]=0$ (or $\mathrm{E}[Z-m_Z(X)\mid X]=0$), which follows from linearity of conditional expectation. For the first term:

$$
\begin{aligned}
\mathrm{E}[[Y-m_Y(X)][Z-m_Z(X)]] &= \mathrm{E}[\mathrm{E}[[Y-m_Y(X)][Z-m_Z(X)]\mid X]] && \text{(law of iterated expectations)}\\
&= \mathrm{E}[\operatorname{Cov}(Y,Z\mid X)] && \text{(definition of conditional covariance)}
\end{aligned}
$$

For the last term, the law of iterated expectations gives $\mathrm{E}[m_Y(X)]=\mu_Y$ and $\mathrm{E}[m_Z(X)]=\mu_Z$, so:

$$
\begin{aligned}
\mathrm{E}[[m_Y(X)-\mu_Y][m_Z(X)-\mu_Z]] &= \mathrm{E}[[m_Y(X)-\mathrm{E}[m_Y(X)]][m_Z(X)-\mathrm{E}[m_Z(X)]]] && \text{(substitute the means)}\\
&= \operatorname{Cov}(m_Y(X),m_Z(X)) && \text{(definition of covariance)}\\
&= \operatorname{Cov}(\mathrm{E}[Y\mid X],\mathrm{E}[Z\mid X]) && \text{(definitions of } m_Y, m_Z)
\end{aligned}
$$

Adding the four terms gives the result. □

Remark. Alternate names include: the **covariance decomposition formula** and the **conditional covariance formula**.

Lemma 0.1 (The covariance of a variable with itself is its variance)**.** *For any random variable X with $\mathrm{E}[X^2]<\infty$:*

$$\operatorname{Cov}(X,X)=\operatorname{Var}(X)$$

Proof

Proof. By the alternative formula for covariance:

$$
\begin{aligned}
\operatorname{Cov}(X, X) &= \mathrm{E}[XX] - \mathrm{E}[X]\,\mathrm{E}[X] && \text{(alternative formula for covariance, with } Y = X\text{)} \\
&= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 && (XX = X^2) \\
&= \operatorname{Var}(X) && \text{(simplified expression for variance)}
\end{aligned}
$$

□

Definition 0.13 (Variance/covariance of a $p \times 1$ random vector)**.** For a $p \times 1$ dimensional random vector $\tilde{X}$,

$$
\begin{aligned}
\operatorname{Var}\big(\tilde{X}\big) &\stackrel{\text{def}}{=} \operatorname{Cov}\big(\tilde{X}\big) \\
&\stackrel{\text{def}}{=} \mathrm{E}\Big[\big(\tilde{X} - \mathrm{E}\,\tilde{X}\big)\big(\tilde{X} - \mathrm{E}\,\tilde{X}\big)^\top\Big]
\end{aligned}
$$

Theorem 0.7 (Elements of the variance-covariance matrix are pairwise covariances)**.** *For a $p \times 1$ random vector $\tilde{X} = (X_1, \ldots, X_p)^\top$, the (i, j)-th element of $\operatorname{Var}\big(\tilde{X}\big)$ is $\operatorname{Cov}(X_i, X_j)$:*

$$
\operatorname{Var}\big(\tilde{X}\big) = \begin{pmatrix}
\operatorname{Var}(X_1) & \operatorname{Cov}(X_1, X_2) & \cdots & \operatorname{Cov}\big(X_1, X_p\big) \\
\operatorname{Cov}(X_2, X_1) & \operatorname{Var}(X_2) & \cdots & \operatorname{Cov}\big(X_2, X_p\big) \\
\vdots & \vdots & \ddots & \vdots \\
\operatorname{Cov}\big(X_p, X_1\big) & \operatorname{Cov}\big(X_p, X_2\big) & \cdots & \operatorname{Var}\big(X_p\big)
\end{pmatrix}
$$

Proof

Proof. Let $\mu_i = \mathrm{E}[X_i]$ for $i = 1, \ldots, p$, so $\mathrm{E}\,\tilde{X} = (\mu_1, \ldots, \mu_p)^\top$. By Definition 0.13:

$$
\begin{aligned}
\operatorname{Var}\big(\tilde{X}\big) &= \operatorname{E}\Big[\big(\tilde{X} - \operatorname{E}\tilde{X}\big)\big(\tilde{X} - \operatorname{E}\tilde{X}\big)^\top\Big] \\
&= \operatorname{E}\left[\begin{pmatrix} X_1 - \mu_1 \\ \vdots \\ X_p - \mu_p \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 & \cdots & X_p - \mu_p \end{pmatrix}\right] \\
&= \operatorname{E}\left[\begin{pmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & \cdots & (X_p - \mu_p)(X_p - \mu_p) \end{pmatrix}\right] \\
&= \begin{pmatrix} \operatorname{E}[(X_1 - \mu_1)(X_1 - \mu_1)] & \cdots & \operatorname{E}\big[(X_1 - \mu_1)(X_p - \mu_p)\big] \\ \vdots & \ddots & \vdots \\ \operatorname{E}\big[(X_p - \mu_p)(X_1 - \mu_1)\big] & \cdots & \operatorname{E}\big[(X_p - \mu_p)(X_p - \mu_p)\big] \end{pmatrix} \\
&= \begin{pmatrix} \operatorname{Cov}(X_1, X_1) & \cdots & \operatorname{Cov}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \operatorname{Cov}(X_p, X_1) & \cdots & \operatorname{Cov}(X_p, X_p) \end{pmatrix} \\
&= \begin{pmatrix} \operatorname{Var}(X_1) & \cdots & \operatorname{Cov}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \operatorname{Cov}(X_p, X_1) & \cdots & \operatorname{Var}(X_p) \end{pmatrix}
\end{aligned}
$$

where:

- the step from the third to fourth line uses the expectation of a random matrix,
- the step from the fourth to fifth line uses Definition 0.9, and
- the last step uses Lemma 0.1.

□

Theorem 0.8 (Alternate expression for variance of a random vector)**.**

$$\operatorname{Var}\big(\tilde{X}\big) = \operatorname{E}\big[\tilde{X}\tilde{X}^\top\big] - \big(\operatorname{E}\tilde{X}\big)\big(\operatorname{E}\tilde{X}\big)^\top$$

Proof

Proof.

$$
\begin{aligned}
\operatorname{Var}\big(\tilde{X}\big) &= \operatorname{E}\Big[\big(\tilde{X} - \operatorname{E}\tilde{X}\big)\big(\tilde{X} - \operatorname{E}\tilde{X}\big)^\top\Big] && \text{(definition)} \\
&= \operatorname{E}\big[\tilde{X}\tilde{X}^\top - \tilde{X}\big(\operatorname{E}\tilde{X}\big)^\top - \big(\operatorname{E}\tilde{X}\big)\tilde{X}^\top + \big(\operatorname{E}\tilde{X}\big)\big(\operatorname{E}\tilde{X}\big)^\top\big] && \text{(expand the product)} \\
&= \operatorname{E}\big[\tilde{X}\tilde{X}^\top\big] - \big(\operatorname{E}\tilde{X}\big)\big(\operatorname{E}\tilde{X}\big)^\top - \big(\operatorname{E}\tilde{X}\big)\big(\operatorname{E}\tilde{X}\big)^\top + \big(\operatorname{E}\tilde{X}\big)\big(\operatorname{E}\tilde{X}\big)^\top && \text{(linearity, element-wise; } \operatorname{E}\tilde{X} \text{ is cons} \\
&= \operatorname{E}\big[\tilde{X}\tilde{X}^\top\big] - \big(\operatorname{E}\tilde{X}\big)\big(\operatorname{E}\tilde{X}\big)^\top && \text{(combine like terms)}
\end{aligned}
$$

□

Theorem 0.9 (Variance of a linear combination)**.** *For any vector of random variables $\tilde{X} = (X_1, \dots, X_n)$ and corresponding vector of constants $\tilde{a} = (a_1, \dots, a_n)$, the variance of their linear combination is:*

$$\begin{aligned}\operatorname{Var}\big(\tilde{a} \cdot \tilde{X}\big) &= \operatorname{Var}\left(\sum_{i=1}^{n} a_i X_i\right) \\ &= \tilde{a}^\top \operatorname{Var}\big(\tilde{X}\big)\tilde{a} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \operatorname{Cov}(X_i, X_j)\end{aligned}$$

Proof

Proof. Treat $\tilde{a}$ and $\tilde{X}$ as $n \times 1$ column vectors, so $\tilde{a} \cdot \tilde{X} = \tilde{a}^\top \tilde{X} = \sum_{i=1}^{n} a_i X_i$, a scalar. By linearity of expectation, $\operatorname{E}\big[\tilde{a}^\top \tilde{X}\big] = \tilde{a}^\top \operatorname{E}\tilde{X}$, so:

$$\begin{aligned}\operatorname{Var}\big(\tilde{a}^\top \tilde{X}\big) &= \operatorname{E}\Big[\big(\tilde{a}^\top \tilde{X} - \tilde{a}^\top \operatorname{E}\tilde{X}\big)^2\Big] && \text{(definition of variance)} \\ &= \operatorname{E}\Big[\big(\tilde{a}^\top \big(\tilde{X} - \operatorname{E}\tilde{X}\big)\big)^2\Big] && \text{(factor out } \tilde{a}^\top\text{)} \\ &= \operatorname{E}\Big[\tilde{a}^\top \big(\tilde{X} - \operatorname{E}\tilde{X}\big)\big(\tilde{X} - \operatorname{E}\tilde{X}\big)^\top \tilde{a}\Big] && \text{(a scalar equals its transpose, so } s^2 = s\,s^\top\text{)} \\ &= \tilde{a}^\top \operatorname{E}\Big[\big(\tilde{X} - \operatorname{E}\tilde{X}\big)\big(\tilde{X} - \operatorname{E}\tilde{X}\big)^\top\Big]\, \tilde{a} && \text{(linearity of expectation, element-wise)} \\ &= \tilde{a}^\top \operatorname{Var}\big(\tilde{X}\big)\tilde{a} && \text{(variance of a random vector)} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \operatorname{Cov}(X_i, X_j) && \text{(expand the quadratic form, using the elements of } \operatorname{Var}\big(\tilde{X}\big)\text{)}\end{aligned}$$

□

Corollary 0.2 (Variance of a sum of two random variables)**.** *For any two random variables X and Y and scalars a and b:*

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2(a \cdot b) \operatorname{Cov}(X, Y)$$

Proof

Proof. Apply Theorem 0.9 with $n = 2$, $X_1 = X$, and $X_2 = Y$:

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y)$$

Alternatively, by linearity of expectation:

$$\begin{aligned}
\operatorname{Var}(aX + bY) &\overset{\text{def}}{=} \mathrm{E}[(aX + bY - \mathrm{E}[aX + bY])^2] && \text{(definition of va} \\
&= \mathrm{E}[(a(X - \mathrm{E}[X]) + b(Y - \mathrm{E}[Y]))^2] && \text{(linearity of exp} \\
&= \mathrm{E}[a^2(X - \mathrm{E}[X])^2 + 2(a \cdot b)(X - \mathrm{E}[X])(Y - \mathrm{E}[Y]) + b^2(Y - \mathrm{E}[Y])^2] && \text{(expand the sq} \\
&= a^2 \mathrm{E}[(X - \mathrm{E}[X])^2] + 2(a \cdot b) \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] + b^2 \mathrm{E}[(Y - \mathrm{E}[Y])^2] && \text{(linearity of exp} \\
&= a^2 \operatorname{Var}(X) + 2(a \cdot b) \operatorname{Cov}(X, Y) + b^2 \operatorname{Var}(Y) && \text{(definitions of v}
\end{aligned}$$

□

Corollary 0.3 (Variance of a sum of independent random variables)**.** *If X and Y are independent, each discrete or continuous, with $\mathrm{E}[X^2] < \infty$ and $\mathrm{E}[Y^2] < \infty$, then:*

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$$

Proof

Proof. By Theorem 0.5, $\operatorname{Cov}(X, Y) = 0$. Applying Corollary 0.2 with $a = b = 1$:

$$\begin{aligned}
\operatorname{Var}(X + Y) &= 1^2 \operatorname{Var}(X) + 1^2 \operatorname{Var}(Y) + 2(1 \cdot 1) \operatorname{Cov}(X, Y) && \text{(variance of a sum of two random variables)} \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \operatorname{Cov}(X, Y) && \text{(simplify)} \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) && (\operatorname{Cov}(X, Y) = 0)
\end{aligned}$$

□

Remark. Corollary 0.2 is why two variables' covariance matters for combining them: the covariance term is what separates the variance of their sum from the sum of their variances, and Corollary 0.3 is the case where that term is 0.

Theorem 0.10 (A variance matrix is symmetric and positive semidefinite)**.** *For a $p \times 1$ random vector $\tilde{X} = (X_1, \ldots, X_p)^\top$ with $\mathrm{E}[X_i^2] < \infty$ for every i, $\operatorname{Var}\left(\tilde{X}\right)$ is symmetric and positive semidefinite: for every $p \times 1$ vector of constants $\tilde{a}$,*

$$\tilde{a}^\top \operatorname{Var}\big(\tilde{X}\big)\tilde{a} \geq 0$$

Proof

Proof. By Theorem 0.7, the (i,j)-th element of $\operatorname{Var}\big(\tilde{X}\big)$ is $\operatorname{Cov}\big(X_i, X_j\big)$, and by Definition 0.9, with $\mu_i = \operatorname{E}[X_i]$:

$$\begin{aligned}
\operatorname{Cov}\big(X_i, X_j\big) &= \operatorname{E}\big[(X_i - \mu_i)(X_j - \mu_j)\big] && \text{(definition of covariance)} \\
&= \operatorname{E}\big[(X_j - \mu_j)(X_i - \mu_i)\big] && \text{(multiplication of numbers is commutative)} \\
&= \operatorname{Cov}\big(X_j, X_i\big) && \text{(definition of covariance)}
\end{aligned}$$

so the (i,j)-th and (j,i)-th elements are equal, and $\operatorname{Var}\big(\tilde{X}\big)$ is symmetric.
For positive semidefiniteness, let $Y = \tilde{a}^\top \tilde{X}$. Then:

$$\begin{aligned}
\tilde{a}^\top \operatorname{Var}\big(\tilde{X}\big)\tilde{a} &= \operatorname{Var}(Y) && \text{(variance of a linear combination)} \\
&= \operatorname{E}\big[(Y - \operatorname{E}[Y])^2\big] && \text{(definition of variance)} \\
&\geq 0 && \big((Y - \operatorname{E}[Y])^2 \geq 0\big)
\end{aligned}$$

The first step is Theorem 0.9. The last step holds because a random variable that is never negative has a non-negative expectation: in the discrete and continuous cases of the definition of expectation, every term of the sum, or the integrand, is non-negative (for the general case, see Billingsley (1995)). □

Example 0.15 (Two variance matrices that differ only in sign)**.** Let $\tilde{X} = (X_1, X_2)^\top$ be equally likely to be each of the four points

$$D_1 = \{(-5, 1),\ (0, -1),\ (0, 1),\ (5, -1)\},$$

and let $\tilde{X}'$ be equally likely to be each of the four points

$$D_2 = \{(5, 1),\ (0, -1),\ (0, 1),\ (-5, -1)\}.$$

Both sets have first coordinates $\{-5, 0, 0, 5\}$ and second coordinates $\{1, -1, 1, -1\}$, so $\operatorname{E}[X_1] = \operatorname{E}[X_2] = 0$, $\operatorname{E}[X_1^2] = (25+0+0+25)/4 = 12.5$, and $\operatorname{E}\big[X_2^2\big] = (1+1+1+1)/4 = 1$, and likewise for $\tilde{X}'$. They differ in the products of the coordinates:

$$\begin{aligned}
\operatorname{E}[X_1 X_2] &= \frac{(-5)(1) + (0)(-1) + (0)(1) + (5)(-1)}{4} = -2.5, \\
\operatorname{E}[X_1' X_2'] &= \frac{(5)(1) + (0)(-1) + (0)(1) + (-5)(-1)}{4} = 2.5.
\end{aligned}$$

Since the means are 0, Theorem 0.7 and Theorem 0.4 give:

$$\mathrm{Var}\big(\tilde{X}\big) = \begin{pmatrix} 12.5 & -2.5 \\ -2.5 & 1 \end{pmatrix}, \qquad \mathrm{Var}\big(\tilde{X}'\big) = \begin{pmatrix} 12.5 & 2.5 \\ 2.5 & 1 \end{pmatrix}.$$

Both are symmetric. Theorem 0.9 with $\tilde{a} = (1,1)^\top$ and $\tilde{a} = (1,-1)^\top$ gives:

$$\mathrm{Var}(X_1 + X_2) = 12.5 + 1 + 2(-2.5) = 8.5, \quad \mathrm{Var}(X_1 - X_2) = 12.5 + 1 - 2(-2.5) = 18.5,$$
$$\mathrm{Var}(X_1' + X_2') = 12.5 + 1 + 2(2.5) = 18.5, \quad \mathrm{Var}(X_1' - X_2') = 12.5 + 1 - 2(2.5) = 8.5.$$

As a check, $X_1 + X_2$ takes the values $-4, -1, 1, 4$, each with probability $1/4$, so its mean is 0 and its variance is $(16 + 1 + 1 + 16)/4 = 8.5$.

Remark. The two sets of points have the same spread along each coordinate axis, so the diagonals of their variance matrices agree. The sign of the covariance says which diagonal direction, $(1,1)^\top$ or $(1,-1)^\top$, has the larger spread.

Example 0.16 (A variance matrix that is not positive definite)**.** Let X_1 have variance $\sigma^2 > 0$, and let $X_2 = X_1$. Every covariance in $\tilde{X} = (X_1, X_2)^\top$ is $\mathrm{Cov}(X_1, X_1) = \sigma^2$ (Lemma 0.1), so:

$$\mathrm{Var}\big(\tilde{X}\big) = \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

With $\tilde{a} = (1,-1)^\top$, $\tilde{a}^\top \mathrm{Var}\big(\tilde{X}\big)\tilde{a} = \mathrm{Var}(X_1 - X_2) = \mathrm{Var}(0) = 0$, so $\mathrm{Var}\big(\tilde{X}\big)$ is positive semidefinite but not positive definite (compare this example). If X_1 is continuous, $\tilde{X}$ has no joint density.

Corollary 0.4 (Independent components give a diagonal variance matrix)**.** *Let* $\tilde{X} = (X_1, \ldots, X_p)^\top$ *be a random vector whose components are each discrete or continuous, with:*

- $\mathrm{E}[X_i^2] < \infty$ *for every* i*, and*
- X_i *and* X_j *independent for every* $i \neq j$*.*

Then $\mathrm{Var}\big(\tilde{X}\big)$ *is diagonal, with diagonal elements* $\mathrm{Var}(X_1), \ldots, \mathrm{Var}\big(X_p\big)$*.*

Proof

Proof. By Theorem 0.7, the (i,j)-th element of $\mathrm{Var}\left(\tilde{X}\right)$ is $\mathrm{Cov}\left(X_i, X_j\right)$. For $i \neq j$, X_i and X_j are independent, so $\mathrm{Cov}\left(X_i, X_j\right) = 0$ by Theorem 0.5. The (i,i)-th element is $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$ by Lemma 0.1. □

Theorem 0.11 (Correlation lies between -1 and 1)**.** *If X and Y have finite, positive variances, then their correlation lies in $[-1, 1]$ (Casella and Berger 2002):*

$$-1 \leq \mathrm{Cor}(X, Y) \leq 1$$

Proof

Proof. Write $\sigma_X \overset{\text{def}}{=} \mathrm{SD}(X) > 0$ and $\sigma_Y \overset{\text{def}}{=} \mathrm{SD}(Y) > 0$, and take either sign $\pm$ throughout. A variance is the expectation of a squared deviation, a non-negative random variable, so it is non-negative. Applying Corollary 0.2 with $a = 1/\sigma_X$ and $b = \pm 1/\sigma_Y$:

$$\begin{aligned}
0 &\leq \mathrm{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) && \text{(a variance is non-negative)} \\
&= \frac{\mathrm{Var}(X)}{\sigma_X^2} + \frac{\mathrm{Var}(Y)}{\sigma_Y^2} \pm \frac{2\,\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} && \text{(variance of a sum of two random variables)} \\
&= 1 + 1 \pm \frac{2\,\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} && \text{(definition of standard deviation: } \sigma_X^2 = \mathrm{Var}(X),\ \sigma_Y^2 = \mathrm{Var}(Y)\text{)} \\
&= 2 \pm 2\,\mathrm{Cor}(X,Y) && \text{(definition of correlation)}
\end{aligned}$$

With the $+$ sign, $0 \leq 2 + 2\,\mathrm{Cor}(X,Y)$ gives $\mathrm{Cor}(X,Y) \geq -1$. With the $-$ sign, $0 \leq 2 - 2\,\mathrm{Cor}(X,Y)$ gives $\mathrm{Cor}(X,Y) \leq 1$. □

References

Billingsley, Patrick. 1995. *Probability and Measure*. 3rd ed. Wiley Series in Probability and Mathematical Statistics. Wiley.

Casella, George, and Roger Berger. 2002. *Statistical Inference*. 2nd ed. Cengage Learning. https://www.cengage.com/c/statistical-inference-2e-casella-berger/9780534243128/.