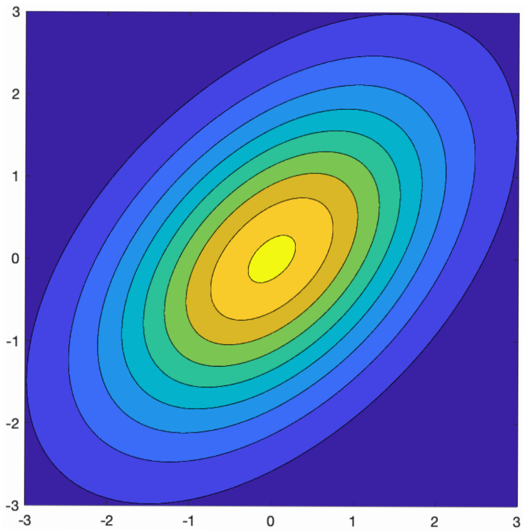
Correlation Coefficient

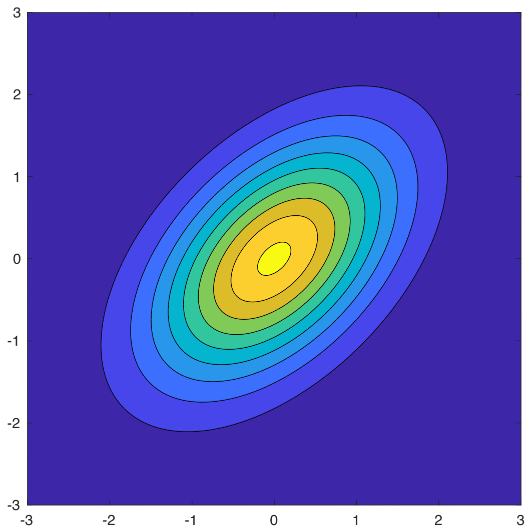
- The **covariance** $\mathrm{Cov}[X,Y] = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$ captures the (average) linear relationship between $X-\mathbb{E}[X]$ and $Y-\mathbb{E}[Y]$.
  → Sensitive to the scale of X and Y.

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



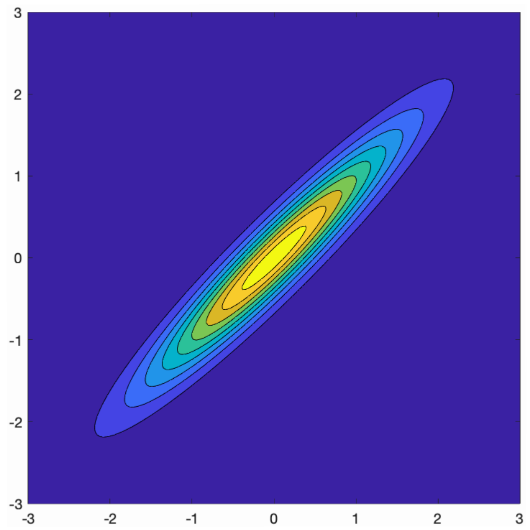
$\rho_{X,Y} = 0.5$

$\mathrm{Cov}[X,Y]=0.5$



$\rho_{X,Y} = 0.5$

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



$\rho_{X,Y} = 0.95$

- The **correlation coefficient** $\rho_{X,Y}$ is a scale-invariant measure of the (average) linear relationship.

$$\boxed{\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}}$$

Correlation Coefficient Properties:

→ $-1 \le \rho_{X,Y} \le +1$

→ $\rho_{X,Y} = +1$ if and only if $Y = aX + b$ for some $a > 0$ and b.

→ $\rho_{X,Y} = -1$ if and only if $Y = aX + b$ for some $a < 0$ and b.

→ $\mathrm{Cov}[X,Y] = \rho_{X,Y}\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}$

→ X and Y are $\underbrace{\text{uncorrelated}}_{\mathrm{Cov}[X,Y]=0}$ if and only if $\rho_{X,Y} = 0$.

→ If $U = aX + b$ and $V = cY + d$, then $\rho_{U,V} = \mathrm{sign}(ac)\,\rho_{X,Y}$.

$$\mathrm{sign}(z) = \begin{cases} +1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases}$$

- Intuition: The closer $|\rho_{X,Y}|$ is to 1, the better a line explains the relationship between X and Y.

- **Example**: $Var[X] = 4$, $Var[Y] = 1$, $\boxed{X \text{ and } Y \text{ are independent.}}$
  $V = X + Y$, $W = -2X + 3Y$ $\Rightarrow Cov[X, Y] = 0$

→ Calculate $\rho_{X,V}$.

$$\rho_{X,V} = \frac{Cov[X,V]}{\sqrt{Var[X]\,Var[V]}} = \frac{4}{\sqrt{4 \cdot 5}} = \frac{2}{\sqrt{5}} \approx 0.894$$

$$Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2ab \underbrace{\cancel{Cov[X,Y]}}_{0}$$

$$Var[V] = Var[X + Y] = 1^2 \cdot 4 + 1^2 \cdot 1 = 5$$

$$Cov[aX + bY, dX + eY] = ad\,Var[X] + be\,Var[Y] + (ae + bd) \underbrace{\cancel{Cov[X,Y]}}_{0}$$

$$Cov[X,V] = Cov[X, X+Y] = 1 \cdot 1 \cdot 4 + 0 \cdot 1 \cdot 1 = 4$$

→ Calculate $\rho_{V,W}$.

$$\rho_{V,W} = \frac{Cov[V,W]}{\sqrt{Var[V]\,Var[W]}} = \frac{-5}{\sqrt{5 \cdot 25}} = \frac{-1}{\sqrt{5}} \approx -0.447$$

$$Var[W] = (-2)^2 \cdot 4 + 3^2 \cdot 1 = 16 + 9 = 25$$

$$Cov[V, W] = Cov[X + Y, -2X + 3Y] = 1 \cdot (-2) \cdot 4 + 1 \cdot 3 \cdot 1 = -8 + 3 = -5$$