Why Divide Covariance by Standard Deviations?
Intuition Behind Correlation

Dr. Ratnesh Prasad Srivastava
GGV, Bilaspur

The Core Formula

The correlation coefficient (r) is defined as:

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

This transforms covariance into a standardized measure. Below we explain **why** this division is essential.

1. The Problem with Raw Covariance

Covariance suffers from two critical limitations:

A. Unit Dependency

- **Problem**: Covariance retains original units of X and Y.
- **Example**:

$$\text{If } X : \text{hours}, \quad Y : \text{test points}$$
$$\text{Cov}(X, Y) : \text{hour} \times \text{points}$$

- **Consequence**: Cannot compare across datasets (e.g., height-weight vs. income-age).

B. Scale Sensitivity

- **Problem**: Covariance scales with variable magnitudes.
- **Example**:

$$\text{If } Y_{\text{new}} = 2Y,$$
$$\text{Cov}(X, Y_{\text{new}}) = 2\text{Cov}(X, Y)$$

- **Consequence**: Artificial inflation/deflation of relationship strength.

2. The Solution: Standardization

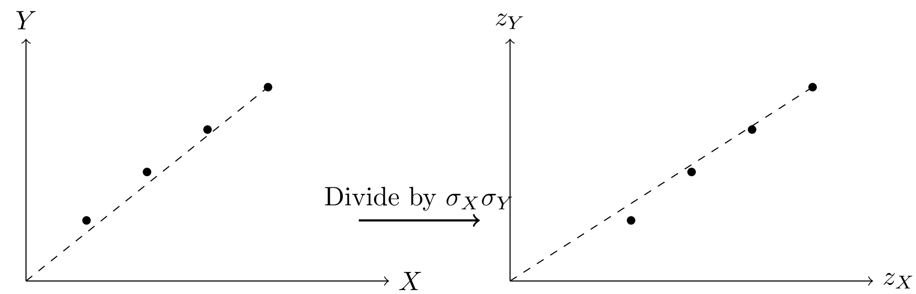
Dividing by $\sigma_X \sigma_Y$ solves both problems.

Mathematical Intuition

$$r = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \text{Cov}\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right)$$

- $\frac{X-\mu_X}{\sigma_X} = z_X$ (standardized X: mean 0, SD 1)
- $\frac{Y-\mu_Y}{\sigma_Y} = z_Y$ (standardized Y: mean 0, SD 1)
- **Key Insight**: Correlation = Covariance of *z-scores*.

Geometric Intuition



Original Data (Covariance = 1.25)   Standardized Data (Correlation = 1.0)

- Standardization rotates and scales the data to a unitless frame.
- Correlation measures alignment in this normalized space.

3. Why $\sigma_X \sigma_Y$? The Mathematical Guarantee

The Cauchy-Schwarz Inequality ensures:

$$\left[\sum (X_i - \bar{X})(Y_i - \bar{Y})\right]^2 \leq \left[\sum (X_i - \bar{X})^2\right] \left[\sum (Y_i - \bar{Y})^2\right]$$

Dividing both sides by $(n-1)^2$:

$$[\text{Cov}(X,Y)]^2 \leq \sigma_X^2 \sigma_Y^2$$

Thus:

$$-1 \leq \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} \leq 1$$

Consequence: r is bounded between -1 and 1, providing:

- Universal strength interpretation (e.g., $r = 0.8$ always means "strong").
- Direction clarity (sign indicates positive/negative relationship).

4. Example: Test Scores vs. Study Hours

Student	Hours (X)	Score (Y)
1	2	50
2	4	70
3	6	80
4	8	85
5	10	95

Calculations

$$\text{Cov}(X, Y) = \frac{1}{4}\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 52.5$$

$$\sigma_X = \sqrt{\frac{\sum(X_i - \bar{X})^2}{4}} = \sqrt{10} \approx 3.16$$

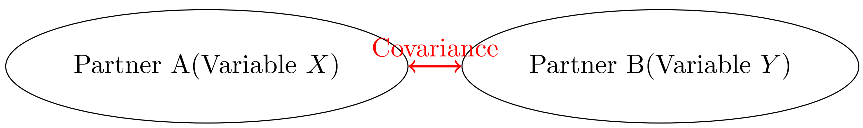
$$\sigma_Y = \sqrt{\frac{\sum(Y_i - \bar{Y})^2}{4}} = \sqrt{292.5} \approx 17.10$$

$$r = \frac{52.5}{3.16 \times 17.10} \approx 0.97$$

Interpretation

- Raw covariance (52.5) is large but uninterpretable alone.
- Correlation (0.97) reveals a **strong, nearly perfect linear relationship**.
- This value is comparable across contexts (e.g., height-weight, temperature-sales).

5. Real-World Analogy: Dance Partners



$$\text{Correlation} = \frac{\text{Synchronization}}{\text{(A's range)} \times \text{(B's range)}}$$

- **Covariance**: Raw synchronization between partners.
- **Standard Deviations**: Measure each partner's natural movement range.
- **Correlation**: Adjusts for partners' differing styles (e.g., hyperactive vs. calm).

Key Takeaways

1. **Unitless**: r is independent of measurement scales.
2. **Scale Invariant**: Multiplying X or Y by constants leaves r unchanged.
3. **Universal Interpretation**:
   - $|r| = 1$: Perfect linear relationship
   - $|r| > 0.8$: Strong relationship
   - $|r| < 0.3$: Weak relationship
4. **Foundational**: Forms basis for regression, PCA, and hypothesis testing.