Derivation of Correlation Coefficient

Data Science and A.I. Lecture Series

Bindeshwar Singh Kushwaha

PostNetwork Academy

Problem Statement

Objective: Derive the formula for the correlation coefficient $r(X, Y)$:

$$r(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X \sigma_Y}.$$

Problem Statement

Objective: Derive the formula for the correlation coefficient $r(X, Y)$:

$$r(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X \sigma_Y}.$$

Definitions:

- σ_X^2: Variance of X.

Problem Statement

Objective: Derive the formula for the correlation coefficient $r(X, Y)$:

$$r(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X \sigma_Y}.$$

Definitions:

- σ_X^2: Variance of X.
- σ_Y^2: Variance of Y.

Problem Statement

Objective: Derive the formula for the correlation coefficient $r(X, Y)$:

$$r(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X \sigma_Y}.$$

Definitions:

- σ_X^2: Variance of X.
- σ_Y^2: Variance of Y.
- σ_{X-Y}^2: Variance of $Z = X - Y$.

Problem Statement

Objective: Derive the formula for the correlation coefficient $r(X, Y)$:

$$r(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X\sigma_Y}.$$

Definitions:

- σ_X^2: Variance of X.
- σ_Y^2: Variance of Y.
- σ_{X-Y}^2: Variance of $Z = X - Y$.
- Covariance between X and Y: $\text{Cov}(X, Y)$.

Problem Statement

Objective: Derive the formula for the correlation coefficient $r(X, Y)$:

$$r(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X \sigma_Y}.$$

Definitions:

- σ_X^2: Variance of X.
- σ_Y^2: Variance of Y.
- σ_{X-Y}^2: Variance of $Z = X - Y$.
- Covariance between X and Y: $\text{Cov}(X, Y)$.

Key Idea: Use the relationship between variance and covariance to expand σ_{X-Y}^2 and relate it to the correlation coefficient $r(X, Y)$.

Problem Statement

Objective: Derive the formula for the correlation coefficient $r(X, Y)$:

$$r(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X \sigma_Y}.$$

Definitions:

- σ_X^2: Variance of X.
- σ_Y^2: Variance of Y.
- σ_{X-Y}^2: Variance of $Z = X - Y$.
- Covariance between X and Y: $\text{Cov}(X, Y)$.

Key Idea: Use the relationship between variance and covariance to expand σ_{X-Y}^2 and relate it to the correlation coefficient $r(X, Y)$.

Step 1: Variance of $Z = X - Y$

Define $Z = X - Y$. The variance of Z is:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)^2.$$

Step 1: Variance of $Z = X - Y$

Define $Z = X - Y$. The variance of Z is:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)^2.$$

Where:

- $z_i = x_i - y_i$: Difference between corresponding values of X and Y.

Step 1: Variance of $Z = X - Y$

Define $Z = X - Y$. The variance of Z is:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)^2.$$

Where:

- $z_i = x_i - y_i$: Difference between corresponding values of X and Y.
- $\overline{Z} = \overline{X} - \overline{Y}$: Mean of Z, obtained as the difference of the means of X and Y.

Step 1: Variance of $Z = X - Y$

Define $Z = X - Y$. The variance of Z is:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)^2.$$

Where:

- $z_i = x_i - y_i$: Difference between corresponding values of X and Y.
- $\overline{Z} = \overline{X} - \overline{Y}$: Mean of Z, obtained as the difference of the means of X and Y.

Substitute $z_i = x_i - y_i$:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left\{(x_i - \overline{X}) - (y_i - \overline{Y})\right\}^2.$$

Step 1: Variance of $Z = X - Y$

Define $Z = X - Y$. The variance of Z is:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - \overline{Z}\right)^2.$$

Where:

- $z_i = x_i - y_i$: Difference between corresponding values of X and Y.
- $\overline{Z} = \overline{X} - \overline{Y}$: Mean of Z, obtained as the difference of the means of X and Y.

Substitute $z_i = x_i - y_i$:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left\{(x_i - \overline{X}) - (y_i - \overline{Y})\right\}^2.$$

Step 2: Expanding the Variance

Expand the squared term inside the summation:

$$\sigma_{X-Y}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \overline{X})^2 + (y_i - \overline{Y})^2 - 2(x_i - \overline{X})(y_i - \overline{Y})\right].$$

Step 2: Expanding the Variance

Expand the squared term inside the summation:

$$\sigma_{X-Y}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \overline{X})^2 + (y_i - \overline{Y})^2 - 2(x_i - \overline{X})(y_i - \overline{Y})\right].$$

This gives three components:

1. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \sigma_X^2$, the variance of X.

Step 2: Expanding the Variance

Expand the squared term inside the summation:

$$\sigma_{X-Y}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \overline{X})^2 + (y_i - \overline{Y})^2 - 2(x_i - \overline{X})(y_i - \overline{Y})\right].$$

This gives three components:

1. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \sigma_X^2$, the variance of X.
2. $\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y})^2 = \sigma_Y^2$, the variance of Y.

Step 2: Expanding the Variance

Expand the squared term inside the summation:

$$\sigma_{X-Y}^2 = \frac{1}{n} \sum_{i=1}^{n} \left[(x_i - \overline{X})^2 + (y_i - \overline{Y})^2 - 2(x_i - \overline{X})(y_i - \overline{Y}) \right].$$

This gives three components:

1. $\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{X})^2 = \sigma_X^2$, the variance of X.
2. $\frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{Y})^2 = \sigma_Y^2$, the variance of Y.
3. $\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y}) = \mathrm{Cov}(X, Y)$, the covariance between X and Y.

Step 2: Expanding the Variance

Expand the squared term inside the summation:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \overline{X})^2 + (y_i - \overline{Y})^2 - 2(x_i - \overline{X})(y_i - \overline{Y})\right].$$

This gives three components:

1. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \sigma^2_X$, the variance of X.
2. $\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y})^2 = \sigma^2_Y$, the variance of Y.
3. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y}) = \mathrm{Cov}(X, Y)$, the covariance between X and Y.

Substitute these into σ^2_{X-Y}:

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\,\mathrm{Cov}(X, Y).$$

Step 2: Expanding the Variance

Expand the squared term inside the summation:

$$\sigma^2_{X-Y} = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \overline{X})^2 + (y_i - \overline{Y})^2 - 2(x_i - \overline{X})(y_i - \overline{Y})\right].$$

This gives three components:

1. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \sigma^2_X$, the variance of X.
2. $\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y})^2 = \sigma^2_Y$, the variance of Y.
3. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y}) = \mathrm{Cov}(X, Y)$, the covariance between X and Y.

Substitute these into σ^2_{X-Y}:

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\,\mathrm{Cov}(X, Y).$$

Step 3: Correlation Coefficient

Recall the definition of the correlation coefficient:

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

Step 3: Correlation Coefficient

Recall the definition of the correlation coefficient:

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

From the variance expansion:

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}(X, Y).$$

Step 3: Correlation Coefficient

Recall the definition of the correlation coefficient:

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

From the variance expansion:

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}(X, Y).$$

Rearrange to express $\mathrm{Cov}(X, Y)$ in terms of σ_X^2, σ_Y^2, and σ_{X-Y}^2:

$$\mathrm{Cov}(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2}.$$

Step 3: Correlation Coefficient

Recall the definition of the correlation coefficient:

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

From the variance expansion:

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\,\mathrm{Cov}(X, Y).$$

Rearrange to express $\mathrm{Cov}(X, Y)$ in terms of σ^2_X, σ^2_Y, and σ^2_{X-Y}:

$$\mathrm{Cov}(X, Y) = \frac{\sigma^2_X + \sigma^2_Y - \sigma^2_{X-Y}}{2}.$$

Substitute into the formula for $r(X, Y)$:

$$r(X, Y) = \frac{\sigma^2_X + \sigma^2_Y - \sigma^2_{X-Y}}{2\sigma_X \sigma_Y}.$$

Step 3: Correlation Coefficient

Recall the definition of the correlation coefficient:

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

From the variance expansion:

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\,\mathrm{Cov}(X, Y).$$

Rearrange to express $\mathrm{Cov}(X, Y)$ in terms of σ^2_X, σ^2_Y, and σ^2_{X-Y}:

$$\mathrm{Cov}(X, Y) = \frac{\sigma^2_X + \sigma^2_Y - \sigma^2_{X-Y}}{2}.$$

Substitute into the formula for $r(X, Y)$:

$$r(X, Y) = \frac{\sigma^2_X + \sigma^2_Y - \sigma^2_{X-Y}}{2\sigma_X \sigma_Y}.$$

Conclusion: The formula for the correlation coefficient $r(X, Y)$ is derived successfully.

Step 3: Correlation Coefficient

Recall the definition of the correlation coefficient:

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

From the variance expansion:

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}(X, Y).$$

Rearrange to express $\mathrm{Cov}(X, Y)$ in terms of σ_X^2, σ_Y^2, and σ_{X-Y}^2:

$$\mathrm{Cov}(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2}.$$

Substitute into the formula for $r(X, Y)$:

$$r(X, Y) = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X \sigma_Y}.$$

Conclusion: The formula for the correlation coefficient $r(X, Y)$ is derived successfully.