Independence of Origin and Scale in Karl Pearson's Correlation Coefficient

Data Science and A.I. Lecture Series

Bindeshwar Singh Kushwaha

PostNetwork Academy

Definition of Correlation Coefficient

- The correlation coefficient $r(X, Y)$ is defined as:

Definition of Correlation Coefficient

- The correlation coefficient $r(X, Y)$ is defined as:
- $$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

Definition of Correlation Coefficient

- The correlation coefficient $r(X, Y)$ is defined as:
- $$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$
- **Covariance:**

$$\mathrm{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})$$

Definition of Correlation Coefficient

- The correlation coefficient $r(X, Y)$ is defined as:
- $$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$
- **Covariance:**
$$\mathrm{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})$$
- **Variance of X:**
$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

Definition of Correlation Coefficient

- The correlation coefficient $r(X, Y)$ is defined as:
- $$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}.$$
- **Covariance:**
$$\text{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})$$
- **Variance of X:**
$$\text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$$
- **Variance of Y:**
$$\text{Var}(Y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{Y})^2$$

Transforming Variables

- We apply the transformations:

$$U = \frac{X - a}{h}, \quad V = \frac{Y - b}{k}, \quad h > 0, \quad k > 0.$$

Transforming Variables

- We apply the transformations:

$$U = \frac{X - a}{h}, \quad V = \frac{Y - b}{k}, \quad h > 0, \quad k > 0.$$

- Substitute into the transformed variables:

$$x_i = a + hu_i, \quad y_i = b + kv_i.$$

Transforming Variables

- We apply the transformations:

$$U = \frac{X - a}{h}, \quad V = \frac{Y - b}{k}, \quad h > 0, \quad k > 0.$$

- Substitute into the transformed variables:

$$x_i = a + hu_i, \quad y_i = b + kv_i.$$

- Transformed means:

$$\bar{X} = a + h\bar{U}, \quad \bar{Y} = b + k\bar{V}.$$

Deviations from the Mean

The deviations from the mean are:

- For x_i:

$$x_i - \bar{X} = h(u_i - \bar{U}),$$

Deviations from the Mean

The deviations from the mean are:

- For x_i:

$$x_i - \bar{X} = h(u_i - \bar{U}),$$

- For y_i:

$$y_i - \bar{Y} = k(v_i - \bar{V}).$$

Deviations from the Mean

The deviations from the mean are:

- For x_i:
$$x_i - \bar{X} = h(u_i - \bar{U}),$$
- For y_i:
$$y_i - \bar{Y} = k(v_i - \bar{V}).$$
- Substitute these into the covariance formula:
$$\text{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y}),$$

Deviations from the Mean

The deviations from the mean are:

- For x_i:

$$x_i - \bar{X} = h(u_i - \bar{U}),$$

- For y_i:

$$y_i - \bar{Y} = k(v_i - \bar{V}).$$

- Substitute these into the covariance formula:

$$\mathrm{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y}),$$

-

$$\mathrm{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} \left[h(u_i - \bar{U}) \cdot k(v_i - \bar{V}) \right].$$

Simplified Covariance

Simplify the covariance:

- Covariance of X and Y:

$$\mathrm{Cov}(X, Y) = hk \cdot \frac{1}{n} \sum_{i=1}^{n} (u_i - \bar{U})(v_i - \bar{V}),$$

$$\mathrm{Cov}(X, Y) = hk \cdot \mathrm{Cov}(U, V).$$

Simplified Covariance

Simplify the covariance:

- Covariance of X and Y:

$$\mathrm{Cov}(X, Y) = hk \cdot \frac{1}{n} \sum_{i=1}^{n} (u_i - \bar{U})(v_i - \bar{V}),$$

$$\mathrm{Cov}(X, Y) = hk \cdot \mathrm{Cov}(U, V).$$

- Variance of X:

$$\mathrm{Var}(X) = h^2 \cdot \mathrm{Var}(U),$$

Simplified Covariance

Simplify the covariance:

- Covariance of X and Y:

$$\mathrm{Cov}(X, Y) = hk \cdot \frac{1}{n} \sum_{i=1}^{n} (u_i - \bar{U})(v_i - \bar{V}),$$

$$\mathrm{Cov}(X, Y) = hk \cdot \mathrm{Cov}(U, V).$$

- Variance of X:

$$\mathrm{Var}(X) = h^2 \cdot \mathrm{Var}(U),$$

- Variance of Y:

$$\mathrm{Var}(Y) = k^2 \cdot \mathrm{Var}(V).$$

Substituting into Correlation Coefficient

- Substitute into $r(X, Y)$:

$$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

Substituting into Correlation Coefficient

- Substitute into $r(X, Y)$:

$$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

- Substitute $\mathrm{Cov}(X, Y) = hk \cdot \mathrm{Cov}(U, V)$,

Substituting into Correlation Coefficient

- Substitute into $r(X, Y)$:

$$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

- Substitute $\mathrm{Cov}(X, Y) = hk \cdot \mathrm{Cov}(U, V)$,
- Substitute $\mathrm{Var}(X) = h^2 \cdot \mathrm{Var}(U)$,

Substituting into Correlation Coefficient

- Substitute into $r(X, Y)$:

$$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

- Substitute $\mathrm{Cov}(X, Y) = hk \cdot \mathrm{Cov}(U, V)$,
- Substitute $\mathrm{Var}(X) = h^2 \cdot \mathrm{Var}(U)$,
- Substitute $\mathrm{Var}(Y) = k^2 \cdot \mathrm{Var}(V)$,

Substituting into Correlation Coefficient

- Substitute into $r(X, Y)$:

$$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

- Substitute $\mathrm{Cov}(X, Y) = hk \cdot \mathrm{Cov}(U, V)$,
- Substitute $\mathrm{Var}(X) = h^2 \cdot \mathrm{Var}(U)$,
- Substitute $\mathrm{Var}(Y) = k^2 \cdot \mathrm{Var}(V)$,
- we get:

$$r(X, Y) = \frac{hk \cdot \mathrm{Cov}(U, V)}{\sqrt{h^2 \cdot \mathrm{Var}(U) \cdot k^2 \cdot \mathrm{Var}(V)}}.$$

Final Simplification

- Simplify the denominator:

$$r(X,Y) = \frac{hk \cdot \mathrm{Cov}(U,V)}{hk \cdot \sqrt{\mathrm{Var}(U) \cdot \mathrm{Var}(V)}}.$$

Final Simplification

- Simplify the denominator:

$$r(X, Y) = \frac{hk \cdot \mathrm{Cov}(U, V)}{hk \cdot \sqrt{\mathrm{Var}(U) \cdot \mathrm{Var}(V)}}.$$

- Cancel hk:

$$r(X, Y) = \frac{\mathrm{Cov}(U, V)}{\sqrt{\mathrm{Var}(U) \cdot \mathrm{Var}(V)}}.$$

Final Simplification

- Simplify the denominator:

$$r(X, Y) = \frac{hk \cdot \mathrm{Cov}(U, V)}{hk \cdot \sqrt{\mathrm{Var}(U) \cdot \mathrm{Var}(V)}}.$$

- Cancel hk:

$$r(X, Y) = \frac{\mathrm{Cov}(U, V)}{\sqrt{\mathrm{Var}(U) \cdot \mathrm{Var}(V)}}.$$

- Therefore:

$$r(X, Y) = r(U, V).$$

Final Simplification

- Simplify the denominator:

$$r(X, Y) = \frac{hk \cdot \mathrm{Cov}(U, V)}{hk \cdot \sqrt{\mathrm{Var}(U) \cdot \mathrm{Var}(V)}}.$$

- Cancel hk:

$$r(X, Y) = \frac{\mathrm{Cov}(U, V)}{\sqrt{\mathrm{Var}(U) \cdot \mathrm{Var}(V)}}.$$

- Therefore:

$$r(X, Y) = r(U, V).$$

- This proves that the correlation coefficient $r(X, Y)$ is independent of changes in origin and scale.

Conclusion

- The correlation coefficient $r(X, Y)$ remains unchanged under linear transformations:

$$U = \frac{X - a}{h}, \quad V = \frac{Y - b}{k}.$$

Conclusion

- The correlation coefficient $r(X, Y)$ remains unchanged under linear transformations:

$$U = \frac{X - a}{h}, \quad V = \frac{Y - b}{k}.$$

- This property demonstrates independence from:

Conclusion

- The correlation coefficient $r(X, Y)$ remains unchanged under linear transformations:

$$U = \frac{X - a}{h}, \quad V = \frac{Y - b}{k}.$$

- This property demonstrates independence from:
  - Changes in origin (a, b),

Conclusion

- The correlation coefficient $r(X, Y)$ remains unchanged under linear transformations:

$$U = \frac{X - a}{h}, \quad V = \frac{Y - b}{k}.$$

- This property demonstrates independence from:
  - Changes in origin (a, b),
  - Changes in scale (h, k).

Conclusion

- The correlation coefficient $r(X, Y)$ remains unchanged under linear transformations:

$$U = \frac{X - a}{h}, \quad V = \frac{Y - b}{k}.$$

- This property demonstrates independence from:
  - Changes in origin (a, b),
  - Changes in scale (h, k).
- This makes the correlation coefficient a robust measure of linear association.

Reach PostNetwork Academy

Website

PostNetwork Academy | www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Facebook Page

www.facebook.com/postnetworkacademy

LinkedIn Page

www.linkedin.com/company/postnetworkacademy

Thank You!