

Quiz: Dependence, Correlation, and Interaction

Student Quiz

Q1. Geometric interpretation Pearson

Explain the geometric interpretation of covariance and Pearson's correlation. What does this view reveal about the sign and strength of linear dependence?

Q2. Spearman versus Pearson

Unlike Pearson's correlation, Spearman's correlation is computed on ranks instead of raw values. How does this affect the types of relationships it can detect? What are the consequences of using ranks?

Q3. Slope of a linear model as a dependence metric

Suppose we want to assess dependence between two numeric features by regressing $X_1 = \theta_0 + \theta_1 X_2 + \varepsilon$. Is the estimated slope $\hat{\theta}_1$ a good measure of the linear dependence? Justify your answer briefly.

Q4. Dependence without correlation

Describe (verbally or via a sketch) a data pattern where X_1 and X_2 are clearly dependent, yet Pearson's ρ is approximately zero.

Q5. Mutual information and independence

The mutual information between X_1 and X_2 is defined as

$$MI(X_1, X_2) = \mathbb{E}_{p(x_1, x_2)} \left[\log \frac{p(x_1, x_2)}{p(x_1)p(x_2)} \right].$$

State the fundamental property that connects mutual information to the definition of statistical independence.

Q6. Estimating MI for numeric features

Outline the typical two-step approach used to estimate mutual information between two continuous variables in practice. Mention one major statistical challenge associated with the procedure.

Q7. Interaction vs. correlation or dependence

(a) Explain the Friedman–Popescu (2008) criterion that indicates a second-order interaction between two features x_j and x_k .

(b) Explain in one sentence why an interaction can be present even when x_j and x_k are statistically independent.

Suggested Answers

- Q1.**
- Each observation defines a rectangle with sides given by its deviations from the variable means. The product of these deviations gives a *signed area*: positive in quadrants I and III, negative in II and IV.
 - The *sample covariance* is the average of these signed areas. Dividing by $n - 1$ instead of n corrects for the bias introduced by using sample means, ensuring an unbiased estimator of the population covariance.
 - *Pearson's correlation* rescales the covariance by the product of the sample standard deviations. This preserves the sign and maps the result to $[-1, 1]$, yielding a scale-invariant measure of linear association.
- Q2.** Using ranks instead of raw values makes Spearman's correlation sensitive to *monotonic* relationships, not just linear ones. This allows it to detect associations where one variable consistently increases or decreases with the other, even if the relationship is nonlinear.
- **Detection of monotonic non-linear dependence:** Spearman can capture patterns such as logarithmic, exponential, or sigmoidal trends that Pearson's correlation would fail to detect.
 - **Robustness to outliers:** Because it only uses rank order, Spearman's correlation is less affected by extreme values compared to Pearson's correlation.
- Q3.** No. $\hat{\theta}_1$ is not scale-invariant. If either variable is rescaled (e.g., °C to °F), the slope changes. Therefore, it cannot be used as a universal indicator of dependence. By contrast, R^2 compares the model's sum of squared errors (SSE) to that of a constant-only model, effectively normalising the explained variance. This makes R^2 a unitless quantity that is invariant under linear rescaling of the variables, and thus an appropriate, scale-free measure of the strength of linear dependence.
- Q4.** Example: Data points arranged in a circle centered at the origin, or an X-pattern formed by two crossing diagonals. These patterns show dependence (structure) but have symmetric contributions to covariance, so $\rho \approx 0$.
- Q5.** $MI(X_1, X_2) = 0$ if and only if X_1 and X_2 are statistically independent, i.e., $p(x_1, x_2) = p(x_1)p(x_2)$. If the variables are independent, the ratio is always 1, so the log is zero and $MI = 0$. Any deviation—whether linear or nonlinear—yields positive contributions. Thus, mutual information detects *all types of dependence*, not just linear associations.
- Q6.** Step 1: Estimate the joint and marginal densities using histograms, kernel density estimation, or k -nearest neighbors.
 Step 2: Plug the estimated densities into the MI formula.
 Challenge: Estimation is sensitive to choice of bins or bandwidth (bias-variance tradeoff), and the method scales poorly in higher dimensions (curse of dimensionality).
- Q7.** (a) A second-order interaction between x_j and x_k is present if $\mathbb{E} \left[\left(\frac{\partial^2 f}{\partial x_j \partial x_k} \right)^2 \right] > 0$.
 (b) Because interaction refers to how the *effect* of one feature depends on another in the function $f(\cdot)$, it can occur even when the input variables are statistically independent.