Jaccard Similarity

The Jaccard Similarity measures the similarity between two finite sets, A and B:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

In terms of True Positives (TP), False Positives (FP), and False Negatives (FN):

$$J = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

Spearman Rank Correlation Coefficient (ρ or r_s)

Spearman's rank correlation coefficient assesses monotonic relationships. If d_i is the difference between the ranks of corresponding values for n observations (with no tied ranks):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

When ties are present, the formula using Pearson correlation on the ranks (rg_X, rg_Y) is:

$$r_s = \frac{\text{cov}(rg_X, rg_Y)}{\sigma_{rg_X} \sigma_{rg_Y}}$$

where $\text{cov}(rg_X, rg_Y)$ is the covariance of the rank variables, and σ_{rg_X} and σ_{rg_Y} are their standard deviations.

Kendall's Tau Coefficient (τ)

Kendall's Tau coefficient measures ordinal association.

Kendall's Tau-a (does not account for ties)

$$\tau_A = \frac{n_c - n_d}{\frac{1}{2} n(n-1)}$$

Where:

- n_c: number of concordant pairs
- n_d: number of discordant pairs
- n: number of observations

Kendall's Tau-b (accounts for ties)

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_c + n_d + T_X)(n_c + n_d + T_Y)}}$$

Where:

- n_c: number of concordant pairs
- n_d: number of discordant pairs
- T_X: number of pairs tied only on the X variable
- T_Y: number of pairs tied only on the Y variable