SUPTECH WORKSHOP III

Background Session II – Measures of statistical association

Measures of association – motivation

We are often interested in modeling **relationships**:

- is A related to B?
- does A cause B?
- do A and B usually coincide?
- what implications does the occurrence of A have on B? (conditional probability)

Probabilistic independence: Two events A and B are independent ($A \perp B$) if and only if their joint probability equals the product of their probabilities.

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) \iff \mathrm{P}(A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \mathrm{P}(A \mid B)$$

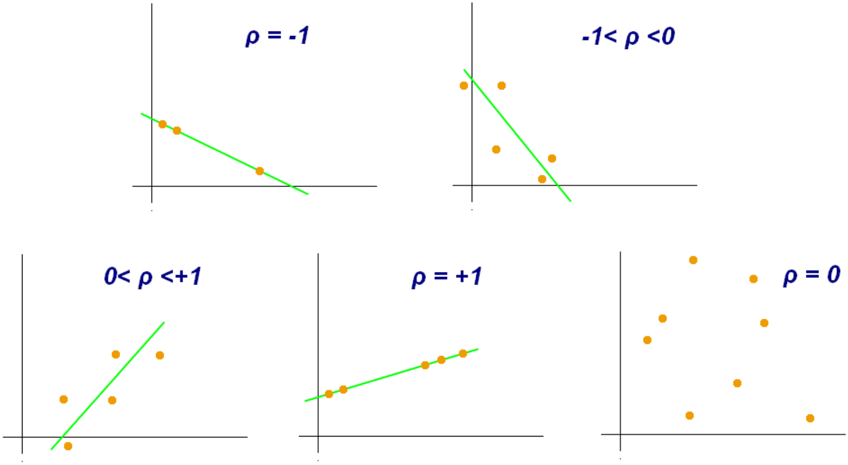
Independence and correlation

Pearson product-moment correlation coefficient:

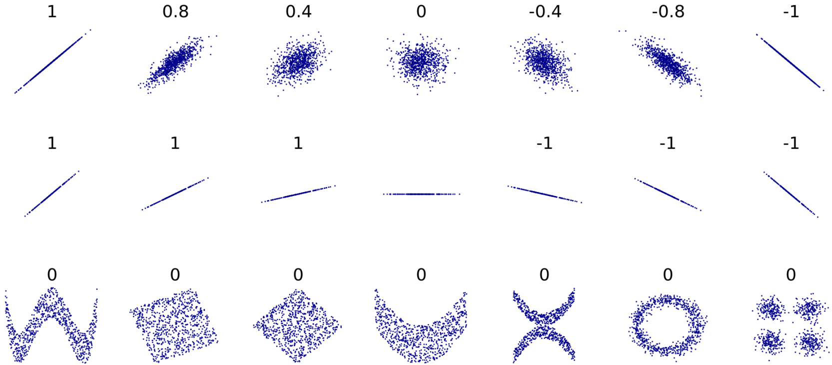
$$\rho_{X,Y} = \operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

where E is the expected value operator, cov means covariance, and corr is a the correlation coefficient.

Pearson correlation coefficient



Pearson correlation coefficient



Correlation vs causality

For any two correlated events, A and B, the different possible relationships include:

- A causes B (direct causation);
- B causes A (reverse causation);
- A and B are consequences of a common cause, but do not cause each other;
- A and B both cause C, which is (explicitly or implicitly) conditioned on;
- A causes B and B causes A (bidirectional or cyclic causation);
- A causes C which causes B (indirect causation);
- There is no connection between A and B; the correlation is a coincidence.

Correlation vs causality

Thus there can be **no conclusion** made regarding the existence or the direction of a cause-and-effect relationship only from the fact that A and B are correlated.

Determining whether there is an actual cause-and-effect relationship requires further investigation, even when the relationship between A and B is statistically significant, a large effect size is observed, or a large part of the variance is explained.

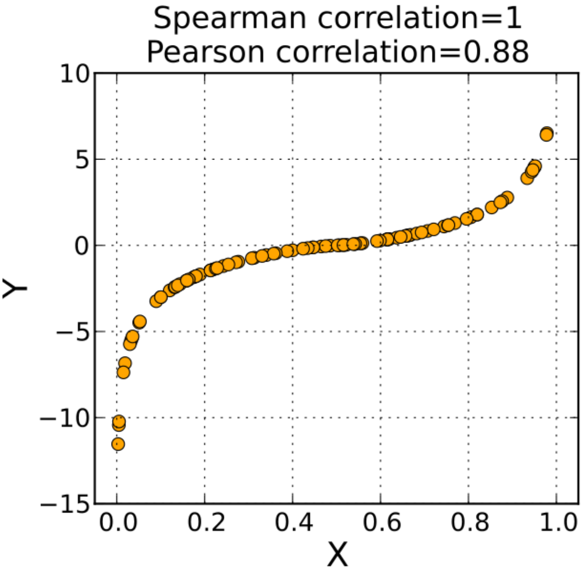
Spearman's correlation coefficient

Spearman's rank correlation coefficient is a **nonparametric** measure of rank correlation .
It assesses how well the relationship between two variables can be described using a **monotonic** function.

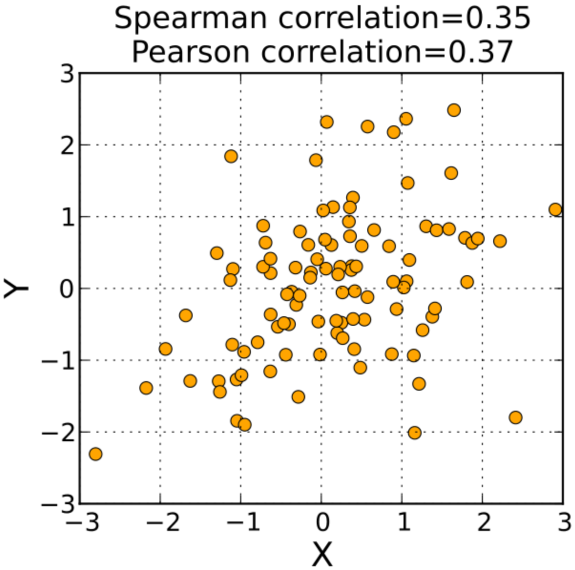
For a sample of size n, the scores X_i and Y_i are converted to ranks $\operatorname{rg} X_i, \operatorname{rg} Y_i$, and r_s is computed from:

$$r_s = \rho_{\operatorname{rg}_X, \operatorname{rg}_Y} = \frac{\operatorname{cov}(\operatorname{rg}_X, \operatorname{rg}_Y)}{\sigma_{\operatorname{rg}_X} \sigma_{\operatorname{rg}_Y}}$$

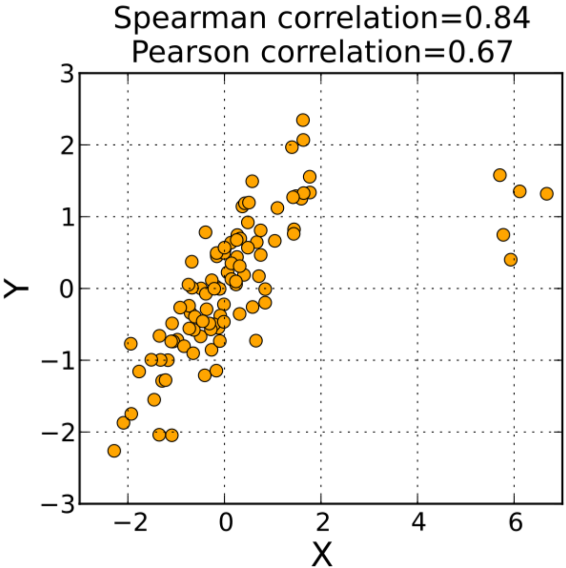
Spearman's correlation coefficient



Spearman’s correlation coefficient



Spearman's correlation coefficient



Partial correlation

Partial correlation measures the degree of association between two random variables, with the effect of a set of controlling random variables removed.

$$\rho_{XY\cdot Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$$

If we define the precision matrix $P = (p_{ij}) = \Omega^{-1}$, we have:

$$\rho_{X_i X_j \cdot \mathbf{V}\setminus\{X_i, X_j\}} = -\frac{p_{ij}}{\sqrt{p_{ii}p_{jj}}}$$

Cross-quantilogram (Han et al., 2016)

Denote $(p_t, s_t), t \in \mathbb{Z}$ a bivariate time series, with $F_{p,t}(\cdot)$ and $F_{s,t}(\cdot)$ the conditional distribution functions and $q_{p,t}(\alpha_p) = \inf\{x, F_{p,t}(x) \geq \alpha_p\}, q_{s,t}(\alpha_s) = \inf\{x, F_{s,t}(x) \geq \alpha_s\}$ the corresponding quantiles.

Given a lag $k \in \mathbb{Z}$, we want to analyze the dependence between the events $p_t \leq q_{p,t}(\alpha_p)$ and $s_{t-k} \leq q_{s,t}(\alpha_s)$. Denote $\psi_a(u) = I(u < 0) - a$, where I is an indicator function.
The cross-quantilogram for (α_s, α_p) at lag k is then defined as

$$\rho_{(\alpha_s,\alpha_p)}(k) = \frac{E\left[\psi_{\alpha_p}\left(p_t - q_{p,t}(\alpha_p)\right)\psi_{\alpha_s}\left(s_{t-k} - q_{s,t-k}(\alpha_s)\right)\right]}{\sqrt{E\left[\psi^2_{\alpha_p}(p_t - q_{p,t}(\alpha_p)\right]}\sqrt{E\left[\psi^2_{\alpha_s}(s_{t-k} - q_{s,t-k}(\alpha_s)\right]}}$$

Sources and further reading

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

https://en.wikipedia.org/wiki/Correlation_and_dependence

https://en.wikipedia.org/wiki/Correlation_does_not_imply_causation

https://en.wikipedia.org/wiki/Spearman's_rank_correlation_coefficient

https://en.wikipedia.org/wiki/Kendall_rank_correlation_coefficient

Han, H., Linton, O., Oka, T., & Whang, Y. J. (2016). The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series. Journal of Econometrics, 193(1), 251-270.