

## Manual: 8.6. Correlation

When two different tags have a relationship, they are called correlated. When we speak of correlation, we typically mean that the relationship is linear. This means that when one tag varies, the other tag varies proportionally to the first, i.e. the relationship is expressed by the constant of proportionality.

An example of two correlated tags is the temperature and pressure of a substance in a vessel. As long as the vessel retains its volume and no substance is allowed to enter or leave the vessel, the pressure and temperature are directly related to each other; as one increases, so does the other. These two tags are strongly related in this fashion and so they are highly correlated.

Another example of two correlated tags is the rotation rate and vibration amplitude. However this relationship depends on a multitude of other tags as well. While these two will tend to increase and decrease together, one might vary somewhat without the other due to other conditions. This relationship is thus weaker.

The strength of a correlation is measured by the correlation coefficient that is a number between -1 and 1. If the coefficient is 1, then the two tags will vary in exact mutual relationship. If the coefficient is -1, the same is true except that while one tag increases, the other decreases. If the coefficient is somewhere in the middle, the relationship is weaker. In real-life data, no two tags are ever correlated perfectly as there is always a natural random variation in any measurement; at the very least due to measurement uncertainty.

In general, it is not possible to interpret a single correlation coefficient in a precise manner because correlation depends on context. For example, if two tags are correlated with a coefficient of 0.758, what does this mean? By itself, this does not mean much. In comparison, correlation coefficients acquire meaning. If two tags have a coefficient of 0.758 and two others have a coefficient of 0.123, then we can say that the first pair is much more related than the second pair. What correlation coefficients are called high or low is a subjective decision of the observer.

In industrial data sets, we sometimes observe correlation coefficients above 0.95 between related tags and frequently get coefficients above 0.8.

We can compute a correlation matrix for a collection of tags in which we compute the correlation coefficient between every pair of tags. This matrix can then be analyzed to ask: What tags are most related to a particular tag? We could choose the tags that are most highly correlated to the tag of interest. We may, additionally, look at the correlation between the tags selected and throw out any tags that have high mutual correlation because this would indicate a (near) duplication of information.

The correlation matrix can form the basis for putting the tags into clusters. Generally, highly correlated tags are physically connected and thus would belong to a single cluster.