Variance (Mathematical)

From a mathematical point of view, variance is a measure for the amount of spread one finds in observed data. When measuring any variable multiple times, the value for each observation may differ. The variation may be due to error, or to actual variation in the measured object. The former type of variation is observed for instance when one asks a class of school children to indicate a length of one meter with their hands: one will find that each estimated length has some deviation with regard to an exact meter. The latter type of variation one finds for instance in a taking the height of individuals in a group of people.

Expressed as a number, variance tells us how large the spread of the observations is, i.e. how much any result may deviate from the mean value found. The larger the variance the more imprecise the exact value of an observed feature was: e.g. most schoolchildren had no clue as to the distance a meter signifies. Of course when establishing the variance within a variable such as length of people, a larger variance does not indicate error in measurement but that, apparently, in the observed group length varies significantly.

The numerical expression of variance for N observations is defined as the mean of the total of the squared deviation of each observation from the mean value, m, for all observations. Thus:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m)^2$$

gives us a measure for how much observed values deviate on average from the mean value found.

Note that variance is thus measured in deviations from the mean, squared. This means variance is not a very 'natural' or human measure to us. To scale back to the order of magnitude that the observations were made in, one scales back to the standard deviation, which is the square root of ^{2}, so .

Variance and variants relate in the sense that we can count variants as data points. For instance, there is a certain spread on average in the amount of variants per document or witness. What does it mean if we find a witness that is extremely deviant in that amount? But we can also count variants into categories, at which point they may become genetically relevant, acting in similar way as DNA mutations cause variance and are tell tale of genetic provenance.

In other languages

DE: Varianz
FR: variance
IT: varianza

JZ