# The Pearson correlation – Testing its significance

The correlation coefficient is much used in sciences where relationships between variables subject to many influence factors are being studied. As $r^2$ of 0.5 might turn out to be highly significant in the case of the epidemiology of some disease, for the laboratory analytical chemist, the calibration model is usually a very good description of the relationship between variables (e.g. concentrations of standards and instrumental responses).

Hence, in general, when we have obtained the linear correlation coefficient, $r$ from two variables, we must then ask if this correlation is significant or not.

We can conduct a significance testing with the null hypothesis $Ho$ being no relationship between the variables, i.e. $r = 0$ and the alternative hypothesis $H_1$ being $r \neq 0$. We will use an alpha $\alpha$ level of 0.05 and a $t$-test statistic formula as below to test whether our results are significantly different from zero 0;

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where

$r$ is the Pearson correlation coefficient for the sample, and
$n$ is the sample size (i.e. the number of data sets)

By relooking at the $r$ value of the example mentioned in last blog on "The Pearson correlation coefficient", we have $r = 0.999$ and $n = 6$ and hence,

$$t = \frac{0.999\sqrt{6-2}}{\sqrt{1-0.999^2}} = \frac{1.998}{0.045} = 44.7$$

According to the $t$-table, the critical value for a two-tailed $t$ -test with 4 degrees of freedom at $\alpha = 0.05$ is 2.777. As our computed value $t = 44.7$ is very much larger than the critical value 2.777, we will reject the null hypothesis which states that the standard concentrations and the instrument intensities are unrelated. Similarly, the p-value calculated is $1.5\text{x}10^{-6}$ which is less than 0.05, indicating similar conclusion. In other words, the $r$-value and hence the correlation between these two variables were highly significant.