The Concept of Degrees of Freedom

Many laboratory analysts find difficulty in understanding the concept of degrees of freedom (v). But in fact, it is not difficult to apprehend it at all.

Recall the formula for sample variance which is the square of standard deviation, namely:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$Sample\,Variance = s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

In order for us to compute s^2, we first need to know $\bar{x}$ and this sample mean value controls any variation in the n sample values. Hence, we can say that only $(n\text{-}1)$ of the sample values are *free* to vary. That is, there are $(n\text{-}1)$ degrees of freedom.

This concept can be illustrated as follows.

Suppose we have analyzed a sample with 5 replicates and obtained a mean of 100. Since we know n =5 and $\bar{x}$ =100, we can be sure that $\sum_{i=1}^{n} x_i = 500$

because $\sum_{i=1}^{n} x_i / n = \bar{x}$. So, once we know four of the values, the fifth one will not be *free* to vary because the sum must add up to 500.

The same argument goes for the linear regression equation taking the form of $y = a + bx$ where a = y-intercept and b = gradient.

In this case, we are restricted by two constants, i.e. the y-intercept and the gradient of the linear curve. Hence, when there are n (x,y) points plotted on the calibration curve, its degrees of freedom should be $(n\text{-}2)$, instead.