A linear regression approach to check bias between methods – Part I

Linear regression is used to establish a relationship between two variables. In analytical chemistry, linear regression is commonly used in the construction of calibration curve for analytical instruments in, for example, gas and liquid chromatographic and many other spectrophotometric analyses. In these cases, we work on a linear relationship within a certain range between the analytical responses and the concentration of the analyte of interest.

The general equation which describes a fitted straight line can be written as

$$y = a + bx$$

where b is the gradient of the line and a, its intercept with the y-axis. The method of least-squares linear regression is used to establish the values of a and b. The 'best-fit' line obtained from least-squares linear regression is the line which minimizes the sum of the squared differences between the experimental (observed) and fitted values for y.

The signed difference between an observed value (y) and a fitted value ($\hat{y}$) is known as a residual, i.e. ($y_i - \hat{y}_i$). The common form of regression is of y on x, which means an assumption that the x values are known exactly (no uncertainty) and the only error occurs in the measurement of y.

In this article, we want to compare the results of two analytical methods for any bias between them by using the linear regression approach. *This approach can also be applied to evaluate two sampling methods for any sampling bias, one of which is a reference or well established method.*

The linear regression approach is safe as long as the variance of the independent (x) variable is somewhat smaller than that of the dependent (y) variable, because the basic assumption of regression (invariant x-values) cannot be violated. Hence, one of the analytical methods studied must be sound, such as standard or reference method with good precision.

We can study the bias between two analytical methods by using the approach modelled through linear regression in *paired* experiments. In this case, we use both methods to analyze the same set of test materials and then to compare their results.

There are actually four possible outcomes of such experiments, as shown graphically in Figures 1 – 4 below.

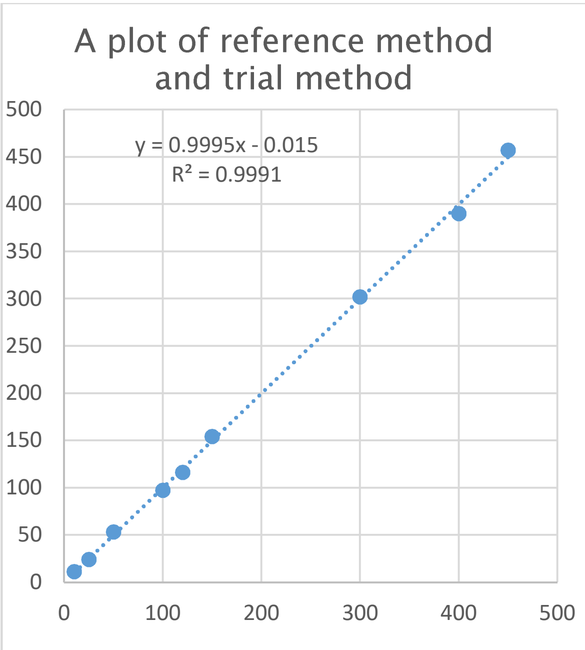

Figure 1: If both methods gave almost identical results (i.e., apart from random measurement variations), we should get a trend of results following the ideal model $y = x$ where gradient b=1 and y–intercept, a = 0.

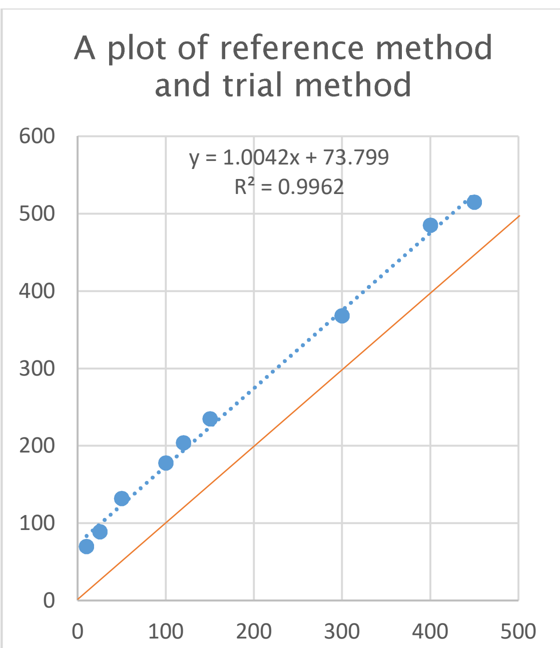

Figure 2: This dataset shows the gradient b was almost 1 but its y–intercept, $a \neq 0$. This type of bias is called ‘constant’ or ‘translational’ bias, and is quite commonly associated with baseline interference in analytical signals

Legends:

the regression line

the theoretical line of no bias

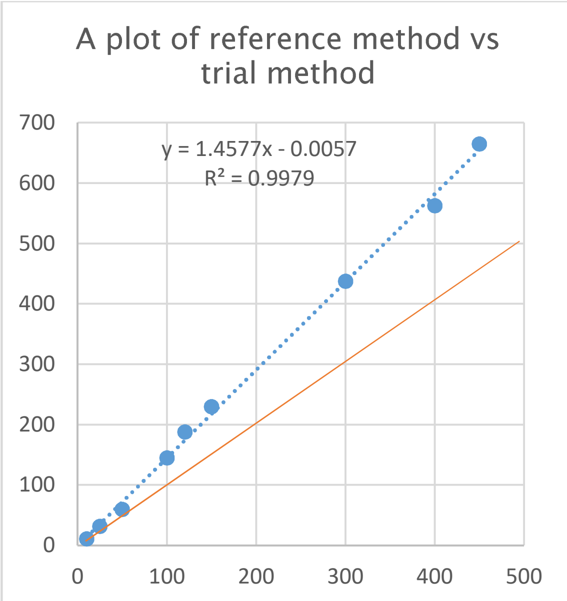

Figure 3: Another common type of bias is characterized by $b \neq 1$ and $a = 0$. In this case, the slope has departed from unity.

This style of bias is called 'rotational' or 'proportional' bias.

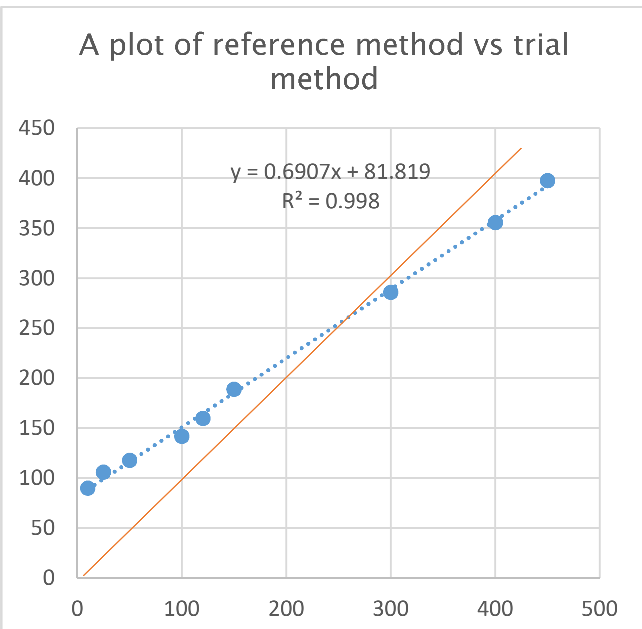

Figure 4: The fourth case shows both $a \neq 0$ and $b \neq 1$. This indicates that both translational and rotational types of bias are present between the methods.

To evaluate any significant difference between the two methods, we can carry out significance Student's t–test by testing the null hypothesis H_o: $a = 0$ *vs* alternate hypothesis H_1: $a \neq 0$, and, also H_o: $b = 1$ *vs* H_1: $b \neq 1$.

The application of such test will be appreciated by studying a worked example.