M

MEMBER CENTER

DATA POINTS

The (Mis)Use of Ordinary Linear Squares Regression

Why ordinary linear regression should not be used to develop correlation between two test methods that claim to measure the same property.

By Alex T.C. Lau

Q Can ordinary linear regression (such as the linear trend function in Excel) be used to develop correlation between two test methods that claim to measure the same property?

A While this is tempting and is often performed by test method users that are non-statistical professionals, the short answer is: no.

Ordinary linear squares regression (OLR) is a model-fitting technique widely used in many different disciplines for many purposes. The most common use is to predict an outcome ($\hat{y}$) based on a specified or observed value of x by substituting the value x into the OLR model. The model is typically built using a finite number (n) of paired data (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), ... (x_n, y_n) obtained either from a specifically designed study, or from historical records (the latter is usually referred to as an observational study).

There are two other noteworthy uses of the model. One is process control, where a process characteristic (y) is controlled to a desired value by manipulation of a control variable (x) based on the predicted outcome $\hat{y}$ from the model. The other is instrument calibration, where the model relates instrument response (y) to reference standards chosen at specific x values.[1] In the case of instrument calibration, the model is used in reverse to *infer* the concentration in an unknown sample by substituting the observed instrument response y into the model and computing the value x.

We will limit our discussion to the common use of the correlation between two test methods that claim to measure the same property, which is to predict a value that would be obtained using method Y based on an actual test result x obtained using method X. The reasons that OLR should not be used to develop such a correlation are explained below.

OLR uses a least square technique to estimate the coefficients (b_0, b_1) of a linear function (Equation 1) that relates observed values of a response variable Y at various design values (x_1 to x_n) of a regressor or predictor X using the paired data set (x_1, y_1) to (x_n, y_n) collected from a study:

$$y = b_0 + b_1 x \quad [1]$$

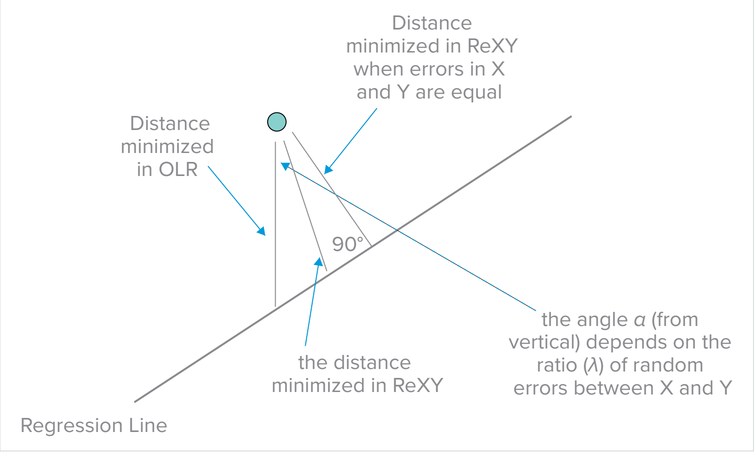
There are several assumptions associated with the OLR theory and the least square methodology. The two most important are:

1. The values x_i in the paired data are assumed to be error free (or negligible); and
2. Each observed value y_i is a random realization from a probability distribution (usually normal) for y at the exact value represented by x_i. The variance of this probability distribution is constant across all values of x.

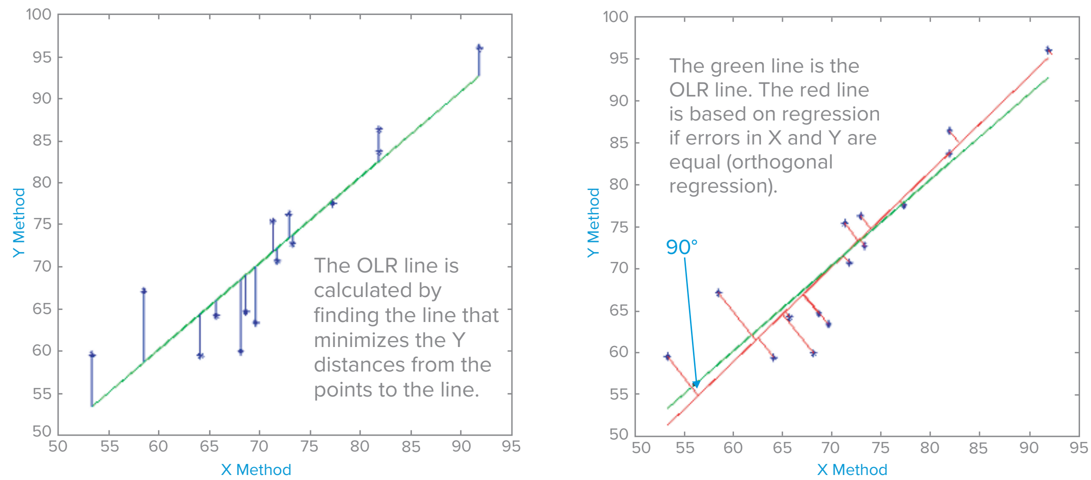
Based on these assumptions, the values b_0 and b_1 are estimated by seeking the line of "best fit" by minimizing the vertical squared distance between each paired observation to the line as illustrated in Figure 1.

Violation of assumption 2 is a common problem associated with the instrument calibration application. The issues are discussed in the Bzik article referenced previously.

For correlation between two test methods, neither assumption is appropriate because of the random errors (method reproducibility) associated with the finite number of test results from each test method. The appropriate technique for developing such a



▼ Figure 1 — A geometric interpretation of differences between OLR and Re*XY*.



◀ Figure 2 — Re*XY* explained graphically.

correlation is known as Re*XY*, short for Regression with errors in *X* and *Y* variables (see Figure 2).

The fundamental difference between the outcome from Re*XY* versus OLR is the methodology used to arrive at b_0 and b_1. In Re*XY*, b_0 and b_1 for the line of best fit is obtained by minimizing the squared distance between each observation pair at an angle α to the vertical line as illustrated in Figure 1.

The angle α is determined based on the ratio $\lambda = (R_X / R_Y)^2$, where R_X and R_Y are the reproducibility of method X and Y evaluated at x_i and y_i respectively. This technique is essentially a weighted regression technique, where the relative weight each paired observation (w_i) has on the regression outcome is inversely proportionally to the joint uncertainties of each observation pair per Equation 2:

$$w_i = 1 / [(R_X)^2 + (R_Y)^2] \qquad [2]$$

At $\lambda = 0$ (i.e., the uncertainty of x_i is zero) for all x, angle α from vertical is zero, which is a special case that represents OLR. At $\lambda = 1$ for all x, i.e.: $R_X = R_Y$, the line of best fit is determined by minimizing the squared distance between each observation pair at a right angle (angle δ in Figure 2 = 90 deg) to the best fit line, which is also known as orthogonal regression. Re*XY* for constant λ across all *x* is also known as Deming regression.

As with all statistical studies, the most important considerations are to ensure both the design and conclusions are statistically sound and adequate. The D6708 standard, practice for statistical assessment and improvement of expected agreement between two test methods that purport to measure the same property of a material, provides detailed guidance on the design and computation to carry out this type of comparison study.

A main objective and prerequisite of D6708 use is to ascertain if there is adequate correlation between the results from method X and Y to permit use of one as a predictor of the other. If yes, statistical methodology is provided to determine if applying a simple linear bias correction to method X result (*x*) can "...further improve the expected agreement" between this bias-corrected *x* result, viewed as $\hat{y}$ (a predicted outcome from method Y), versus actual *y*. If there is strong evidence that a bias correction can statistically improve the agreement, then the parsimony principle is followed, whereby a simple correction is favored over a more complex one. The correlation methodology in this practice is method symmetric, meaning the correlation is an exact mathematical inverse if X and Y are switched, which is not the case with OLR.

The principal steps in practice D6708 are:

1. Assess sample set property variation adequacy relative to the precisions of the two test methods X and Y;
2. Assess correlation adequacy between test results from the two methods;
3. Assess if applying a simple linear bias correction (constant, proportional, or both) to method X results can result in a statistically visible improvement in agreement between method X (bias corrected) and Y results above the combined testing "noise" of both test methods;
4. Select the simplest and appropriate bias correction (if any);
5. Identify if there are sample matrix effects; and
6. Assess the residuals ($y - \hat{y}$) for normality.

Finally, note that D6708 is a statistical assessment of the outcome "closeness" between the two methods. It is not a technical assessment of the degree of similarity between the metrology techniques. Hence, results from the two test methods can be highly correlated, but the method measurement principles may be completely different, and therefore, are not the principal cause for correlation. As with all statistical models, usage should only be limited to within the sample space spanned by the study (i.e., do not extrapolate). ■

REFERENCES

1. Bzik, Thomas, "Measurement Variability Heterogeneity Impact on Calibration and Control Charting," Data Points, *ASTM Standardization News*, Nov./Dec. 2020, pp. 46-48.

Alex T.C. Lau, TCL Consulting, Whitby, Ontario, Canada, is an officer and member of the committee on petroleum products, liquid fuels, and lubricants (D02). An ASTM fellow, he is also a member of the committees on quality and statistics (E11), and accreditation and certification (E36).

Dean V. Neubauer, the Data Points column coordinator, is engineering fellow and chief statistician, Corning Inc. A member at large on the executive subcommittee of the committee on quality and statistics (E11), he is an ASTM International fellow and a past chair of the E11 committee.