The Impact of Measurement Variability Heterogeneity on Calibration and Control Charting

By Thomas J. Bzik

Q Does measurement system heterogeneity impact statistical analysis?

A Yes, this heterogeneity can meaningfully impact all basic statistical analysis techniques. Basic statistical techniques commonly use the simplifying assumptions of variance homogeneity and normality. Variance homogeneity is the expectation that all sample measurements, a priori, are determined with the same degree of random uncertainty (originate from a distribution with the same standard deviation/variance). Variance heterogeneity is the lack of variance homogeneity.

CALIBRATION

Modern analytical instruments rarely directly measure in the scale of interest. They instead measure a physical property that can be related to the scale of interest through calibration. In a typical calibration analysis, "known" levels of a working standard are measured on the system to be calibrated. Least squares regression analysis is often used with such data to establish calibration. This model allows future measurements on the instrument measurement scale to be mathematically converted back into the more useful scale of the standards. Calibration is opaque to the consumers of the instrument data since they only see measurements expressed in the units of interest. Any deficiencies in a calibration modeling process are already baked into measured results.

Least squares calibration is sensitive to the presence of measurement system heterogeneity. Figure 1 illustrates a linear calibration model for an increasing pattern of measurement system variance heterogeneity. Standard least squares regression analysis assumes variance homogeneity and fits the calibration model, assuming every point is equally reliably measured. This causes the points with smaller uncertainty (left side) to be relatively underweighted in model fitting and points with greater uncertainty to be relatively overweighted in model fitting if variance heterogeneity is unaddressed in the regression analysis. Where variability is smallest, the least squares calibration model may not even pass through the data or statistically near enough to the data at that level, relatively speaking. A weighted least squares analysis or a variant regression approach is required to obtain a reasonable estimation of calibration in this scenario.

A relatively routine measurement situation where such heterogenous measurement variation is likely to apply is in trace-level measurement systems. Here, measurements are frequently at part per million (ppm), part per billion (ppb) levels, with ppb systems being somewhat more likely to evidence a

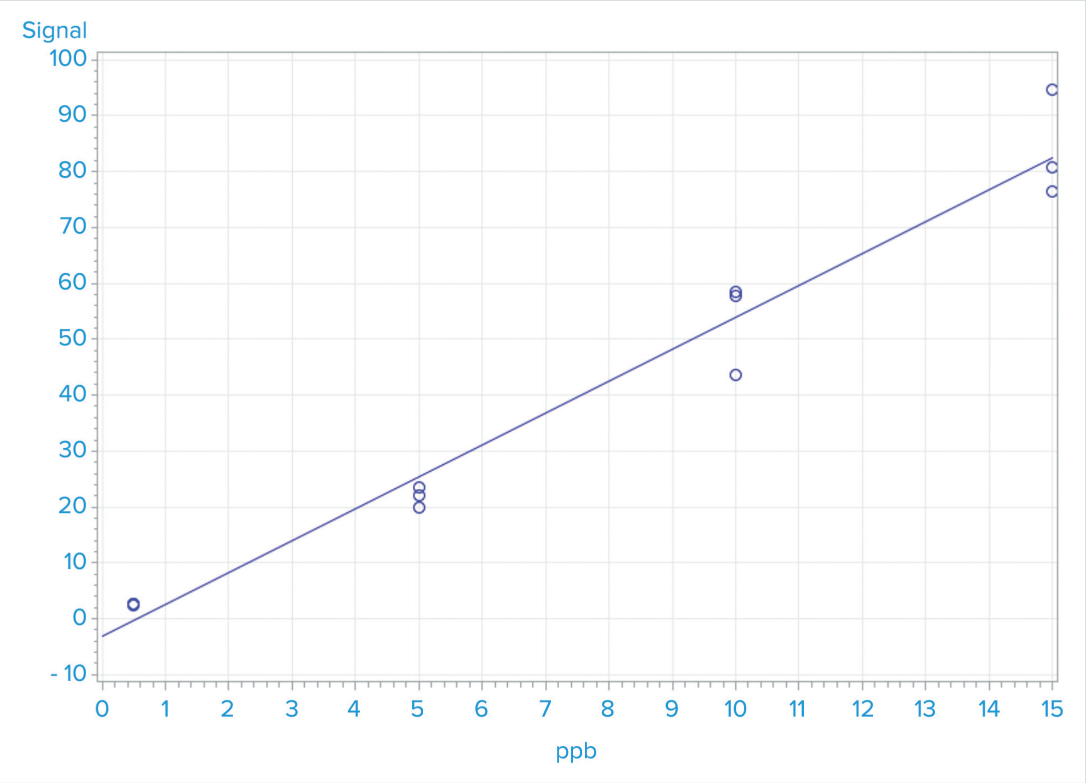

Figure 1 — Calibration Using Least Squares Regression Analysis in the Presence of Variance Heterogeneity

stronger version of variance heterogeneity than ppm systems. A typical pattern elicits an increasing pattern of variability as estimated concentration increases.

The interpretation of regression model confidence intervals and all statistical significance interpretations are corrupted once variance heterogeneity becomes relatively meaningful and least squares is forced. As Figure 1 illustrates, the left end of the calibration can become displaced relative to the data. For example, unexpected negative intercepts or even unexpected negative estimates of concentration can be obtained from such a flawed calibration fitting scenario. Failure to address the heterogeneity of variation in calibration introduces systematic bias in calibration. In terms of the model parameter estimation for the increasing variance heterogeneity scenario illustrated, the slope estimate is unlikely to become significantly biased. However, the intercept often becomes significantly biased.[1,2] This introduces bias on the left side (lowest levels) of the calibration curve, and it will additionally corrupt estimation of detection limits. Unaddressed measurement variance heterogeneity has relatively meaningful impact on calibration.

SHEWHART INDIVIDUALS CONTROL CHARTING

The impact of measurement system variation on control charting, which implicitly assumes homogeneity of variation in its construction, is more complex. Assume, for discussion, that there are two sources of variation: that of the manufacturing process and that of the measurement process. Assume that the manufacturing process provides a normal distribution with some average and standard deviation σ_1. Also assume that the measurement system is unbiased and has some constant normally distributed error σ_2. In this homogeneity of measurement error variance scenario, normal theory is appropriate to apply to the pooled combination of these two error sources in control charting.

When heterogeneity of measurement variation is encountered, the most common scenario is that measurement variation increases, in some manner, with the magnitude of the property being quantified. With measurement variance heterogeneity, the resultant distribution of the data being control charted becomes nonnormal despite the same assumption of normality for the manufacturing variation and the assumption of normality, but with differential standard deviation, for each measured value. With such measurement variance heterogeneity, the lower

(Continued on page 48)

(*Data Points* — continued from page 47)

Table 1 — Out-of-Control Rates with a Shewhart Individuals Upper Control Limit When $n = 30$

Homogeneity ?	MSA % Total Std. Dev.	Out-of-Control %	Population Skew	Distribution
Yes	0% to 100%	0.38	0	Normal
No - Proportional	30%	0.55	0.15	Nonnormal
No - Proportional	50%	0.78	0.35	Nonnormal
No - Proportional	70%	0.92	0.44	Nonnormal
No - Sqrt(Prop)	70%	0.63	0.25	Nonnormal

tail of the observed distribution becomes relatively shorter on average, and the upper tail becomes relatively longer on average. If the relative range of measurement variance heterogeneity is small enough relative to manufacturing variation, this consequence may not be of practical significance.

Control charting issues occur when a manufacturing process is more tightly controlled than can be measured reliably (poor measurement system capability) and lack of measurement homogeneity is strongly anticipated. These are both encountered with some frequency in trace contamination measurement contexts in which the ultimate manufacturing goal is the lack of multiple possible trace-level contaminants. Some trace contaminants are not even anticipated but measured and "controlled" out of an abundance of caution. Other undescribed data measurement contexts may also be relevant.

Table 1 simultaneously illustrates the impact of multiple measurement variation scenarios and measurement capability scenarios for a sample size of 30. With homogeneity, the expected OOC% and population skew are not impacted by any result over the entire measurement capability range 0% to 100% (things that are impacted are not studied herein). Proportional heterogeneity and square root proportional heterogeneity, respectively, indicate that when measuring a value twice as large, it will have twice the measurement standard deviation and √2 times the measurement standard deviation. Additionally, in trace contamination scenario general practice, often only upper control limits are studied.

Normal theory risk normalized to the four tabled out-of-control heterogeneity percentages would be the risk equivalent to having built Individual control charts that only made use of 10 to 14 normally distributed observations. Measurement variance heterogeneity, with its induced skewness, even with imposed normality assumptions on all sampling distributions (process and measurement), adversely impacts calibration, detection limit estimation, and sometimes control limit estimation. ■

References

1. Ketkar, S.N., and Bzik, T.J., "Calibration of Analytical Instruments. Impact of Nonconstant Variance in Calibration Data," *Analytical Chemistry,* Vol. 72, No. 18, 2000, pp. 4672-4765.
2. Bzik, T. J., "Calibration," *Encyclopedia of Statistics in Quality and Reliability*, John Wiley & Sons, Hoboken, N.J., 2008.

Thomas J. Bzik is current chair of the committee on quality and statistics (E11), and he works as a statistical consultant.

Dean V. Neubauer, the Data Points column coordinator, is engineering fellow and chief statistician, Corning Inc., and the Data Points column coordinator. 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.