Technical

Proficiency program for rubber testing

By Kimberly O'Farrell and Scott Granish
Collaborative Testing Services

Proficiency testing is an incredible tool that has multiple uses and benefits for many industries.

Collaborative Testing Services traces its roots to an earlier proficiency testing program at the National Bureau of Standards, now known as NIST. Initially three programs (paper, container board and rubber) were offered. As we celebrate our 50th year as a proficiency testing provider, we now offer eight industry programs as well as a forensics program covering 10 disciplines.

TECHNICAL NOTEBOOK
Edited by John Dick

Executive summary

The Collaborative Testing Services Rubber Proficiency Testing Program is used by participants to satisfy a broad spectrum of internal quality assurance needs, regulatory requirements, as well as maintenance of accreditation status. Through proficiency testing, participants are able to validate test methods and measurement techniques, generate precision statements, investigate causes of systematic error/bias, assess performance against known values and determine uniformity of testing compared to other participants' performance.

To do so, the PT program must be able to gather and analyze a large amount of data, and distill the analysis into reports that present results that are easily understood and actionable. By utilizing comparative statistics, we assess reproducibility and consensus values from all participants to determine the best value for that property. Coupling these statistical models with a two-sample analysis approach allows for the use of a bi-variate analysis technique. This elevates the analysis level to evaluate the measurement performance between two samples, and the results can be represented by graphing each participant's data onto an ellipse plot.

Using a two-sample analysis allows us to access precision and consistency factors. Additionally, by analyzing the grand means and standard deviations, we are able to calculate the comparative performance value (CPV) to evaluate each participant's performance over time, as represented in a trend chart. The presentation of this data using the control ellipse and trend chart analysis allows for participants to easily evaluate and act on a complex statistical analysis of testing data.

The rubber program includes offerings for tensile, hardness, compression set, density and ODR/MDR/RPA testing. By participating in proficiency testing, laboratories can compare their results to known values, giving them a reassurance that their instrument and procedures are producing reliable results for their product.

Proficiency testing builds trust with new clients by assuring them that a sound quality system is maintained that meets both parties' needs. By working together, CTS can identify laboratories' strengths and weaknesses, improve testing competency and strengthen methods for the industry community.

Experimental

The test design comprises many key components, such as comparative statistics, multivariate control charts, two-sample plots generation of a control ellipse and interpretation of the control ellipse. Each key component is explained in further detail.

Comparative statistics

The CTS industry interlaboratory program utilizes comparative statistics. Reproducibility is the key principle in comparative statistics, as it evaluates two or more laboratories agreeing on a value for a given property. The most agreed upon values in this case is designated the grand mean.

In this example, the grand mean would be the mean for all participants' data for a sample set. It is important to note that with consensus statistics, the grand means are calculated best estimates of the true value, based on the reported data. CTS analyzes the difference between the laboratory mean and grand mean, along with the variation around the grand mean, to evaluate how close the result is to the consensus value as shown in **Fig 1**.

Through the additional use of a two-sample analysis (Sample A and Sample B in this case), comparative statistics are further used to determine testing consistency with one another. When testing, it is expected that Sample A and Sample B will produce similar results, provided they comprise the same or similar material types. If the results for Sample A and Sample B are analyzed simultaneously, further conclusions about a laboratory's precision and consistency then can be drawn.

Trend charts

Continual monitoring of an ongoing PT scheme is an important component of any quality system. The calculation of a unitless comparative performance value (CPV) allows for chart results over multiple testing rounds. This allows for more than a year's worth of testing to be shown, allows labs to track performance over time and, in some cases, identify an impending problem before it can significantly affect the lab.

Inconsistency in testing between samples, consistent testing above or below the grand means or movement away from the consensus in subsequent cycles are problems that can be identified using the trend charts.

Two-sample plots

Most of the CTS industry interlaboratory tests utilize a two-sample plot analysis. To keep the explanation simple, refer to Sample A and Sample B. Participants submit their data for Sample A and Sample B, and properties—such as lab means, grand means and between-lab standard deviation—are calculated for each sample.

Another statistic, called the critical value for h, is calculated for each sample as well. When a participant's lab mean exceeds the critical value for h, that data is removed from the statistical analysis. Each participant's data is then plotted onto a scatter plot and is represented by a single point.

Looking at the two-sample plot, the participant's lab mean for Sample A is plotted using the x-axis, while the participant's lab mean for Sample B is plotted using the y-axis. Where the data points intersect is representative of a participant's result. **Fig. 2** shows the concept of a two-sample plot, which often is attributed to W.J. Youden.[1]

The two-sample plot is a simple but incredibly effective tool for interlaboratory comparisons, as well as in-laboratory comparison. By comparing the two samples this way, one can visually assess the nature of the errors (systematic or inconsistent), help determine the robustness of the test method and possibly determine if there are any other influences on the participants' results.

Multivariate control charts

Once the results are plotted, a 95-percent control ellipse is drawn, not to be confused with a 95-percent confidence interval. The 95-percent control ellipse involves analyzing two samples simultaneously, whereas the 95-percent confidence interval involves analyzing single samples.

The 95-percent control ellipse is drawn such that 95 percent of the time, a randomly selected laboratory is included inside the ellipse. CTS does calculate a 99-percent ellipse, but it is not drawn on the plot. Laboratories that fall between the 95-percent and 99-percent ellipse are assigned a warning "*" to review testing procedures, but data are still included in the analysis. Laboratories that fall outside the 99-percent control ellipse are excluded, and new grand means and between-lab standard deviation are calculated.

It is not uncommon for analysts to track systematic variations of a property over time for a product. Similar to that same principle, the multivariate control chart tracks performance of samples or variables against control limits over time. The 95-percent control ellipse found in the CTS reports is the exact mathematical transformation of a T2 multivariate control chart.

In the 1930s and 1940s, an American mathematical statistician, Harold Hotelling, developed the T^2 statistic.[2] The revolutionary work done by Hotelling laid the

See **Rubber**, *page 21*

O'Farrell Granish

The authors

Kimberly O'Farrell is an industry program analyst with Collaborative Testing Services for the plastics and rubber programs. She has collaborated with laboratories worldwide and provided support for their proficiency testing needs.

O'Farrell earned a bachelor's in forensic science, with a concentration in forensic biology. She is active in ASTM and the ACS Rubber Division.

Scott Granish is director of industry programs with CTS. He has been with CTS since late 2003, and has managed the color, paper, plastics, rubber and wine programs. He currently oversees the eight industry programs at CTS and continues to run the wine program.

Granish earned a bachelor's in biology and holds professional memberships in ASTM and the ACS Rubber Division.

Fig. 1: Data table showing participant's evaluation to the consensus results.

WebCode	Test Cycle	Sample Code	(Lab Mean	- Grand Mean)	÷ Btwn Lab Std Dev	= CPV	Data Flag (if assigned)
92NF4N	102	X43	14.82	14.93	0.98	-0.12	75 of 83 labs included
		X44	14.96	15.07	0.93	-0.12	

Testing Date: 2nd Qtr 2017 Sample X43: PP Sample X44: PP

Rubber

Continued from page 20

foundation for statistics during a crucial time in World War II.

Generation of control ellipses

In 1901, Karl Pearson developed the principal components analysis. The principal components analysis is a method for removing and condensing the inter-relationships among a large number of correlated variables. One of the variables is treated as independent, while the other variable is dependent. The relationship between the two variables would result in a line of best fit, also known as the orthogonal regression line.[3]

A few years later, Walter Shewhart[4], also known as the "Father of Quality Control," recommended the ellipse as a quality control tool in the 1930s. Shewhart's main focus was to solve problems involved in the control of quality in manufactured products.

In 1956, J. Edward Jackson wrote Quality Control Methods[5] that focused on outlining the need for simultaneous control of related variables. Jackson discusses that oftentimes variables are correlated, which results in a tipped ellipse where the major axis coincides with the orthogonal regression line between the two variables. Through Hotelling's T^2 control chart, Sample A and Sample B are represented by a single number. The disadvantage of the T^2 control chart is if an out-of-control condition exists, one must review the original data to determine the origin.

Considering CTS focuses on the participant's individual result, instead of multiple results, this does not pose a problem in the industry proficiency program, as the use of the control ellipse overcomes this difficulty. Jackson concludes his work by indicating the results obtained by each method are identical. The T^2 control chart summarizes the condition of process, which is represented by a single number, while the control ellipse indicates the nature of the out-of-control conditions.

The CTS industry proficiency programs utilize a combination of both these approaches to create a 95-percent control ellipse. This is done because 3 sigma limits are generally too broad and insensitive to detect errors, and 2 sigma limits are often too narrow. CTS's goal is to reduce the possibility of rejecting a result and declaring it out of control when the result should actually be included in the analysis.

Results and discussion

Interpretation of control ellipse

Once the control ellipse is generated, the task of interpreting the results remains. In 1974, John Mandel and Theodore Lashof conducted research to develop a more general method of interpreting the control ellipse. When two materials were close to one another, the orientation of the ellipse was close to 45 degrees, resulting in an additive model. With the additive model, it was noted that estimates of laboratory biases and random error components could be calculated.

When two materials are broadly different, the resulting angle of the ellipse diverges from 45 degrees and/or the shape of the ellipse becomes similar to that of a circle. In these cases, additional information is required to interpret the Youden plot.[6]

In an ideal world, each testing laboratory would be perfectly centered on the crosshairs. However, every laboratory has some error in their measurements. Instrument bias, operator error and differences in test conditions are just a few examples of what can introduce error to a lab's results. A component of this error is usually consistent across subsequent test results. This fact is how good data tends to form an additive model, which is represented by a narrow ellipse that is aligned on a 45-degree angle similar to that in **Fig 3**.

These ellipses provide useful information to labs that fall outside their boundaries and are looking to take corrective action. If a lab falls outside the ellipse but along its major axis, that is a clear indication that there is a consistent source of error that the lab needs to address. If a lab falls outside the ellipse but along its minor axis, that indicates that there is an inconsistency in testing; i.e., some source of error was introduced or changed between testing the two samples.

Summary

Through proficiency testing, an individual laboratory can evaluate its performance against pre-established criteria. Proficiency testing can be used for validating methods, technical training, measurement techniques, and investigating causes of systematic errors/bias. Most importantly, it is an essential tool to ensure competence in testing a laboratory's product as well as an essential tool for any ongoing quality assurance program in a lab.

Generations of complex statistical data does not serve the labs if it is not presented in a clearly understandable and actionable format. A collaborative relationship between the PT provider and the technical leaders in the participating laboratories can greatly benefit the industry.

References

1. Youden, W.J. (1972). Graphical Diagnosis of Interlaboratory Test Results. Journal of Quality Technology, 4(1), 29–33.
2. Mason R., Young J. (2001). Implementing Multivariate Statistical Process Control Using Hotelling's T^2 Statistic. Quality Progress. April:71-73.
3. Pearson, K. (1901). On lines and planes of closed fit to systems of points in space. Philosophical Magazine 2:559-572.
4. Shewhart, W. (1931). An Economic Control of Quality of Manufactured Product, D. VanNostrand Co., New York.
5. Jackson J. (1956). Quality Control Methods. Industrial Quality Control. January: 4-8.
6. Mandel, J., and Lashof, T.W. (1974). Interpretation and Generalization of Youden's Two-Sample Diagram. Journal of Quality Technology, 6(1), 22–36.

Fig. 2: Example of a two-sample plot attributed to Youden.

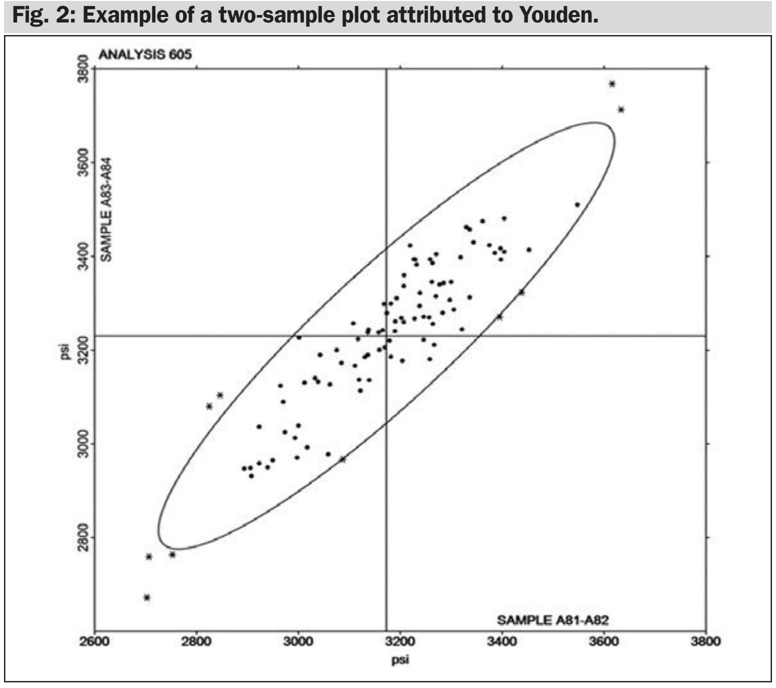

Fig. 3: Control ellipse mapping inconsistency in testing and systematic errors.