

Simple Statistics for Laboratory Data Analysis

James August, CQA
Director of Quality Management
American Biltrite, Inc. Tape Products Division
Moorestown, NJ

Introduction

The use of statistics in the daily operation of test laboratories tends to be very formal. Six Sigma for Design, Design of Experiments, Gauge R&R, and ANOVA may be used to help design and analyze product and process characteristics in research labs. The use of statistical approaches can provide important preventive actions since many of the common SPC (Statistical Process Control) tools can detect small changes to your process and thus predict changes to your future results. As well, statistical tools are excellent for monitoring the effectiveness of corrective actions.

In traditional production test analysis, one set of test results is reported. The set may include some replications and all data and/or data averages are reported. In a statistical analysis the same data set is subjected to an analysis for validity; trends and significance are reported along with the data. The statistical approach requires a little more effort, but provides considerably more information.

Only a statistical interpretation will yield consistently valid evaluations of test data. But often, the rapid turnaround required in the production quality environment prohibits the use of lengthy, repetitive test processes and analyses. The concept of statistical thinking is embodied in the question: "How do these results compare with the results that were expected from this process?"

And, there are several approaches that can help answer this question including the use of simple techniques that don't require major investment in time and energy. Two such tools for analyzing laboratory test data are Chauvenet's Criterion for one-time tests and the X-mR Chart for variable trend analysis.

In the production of pressure sensitive tapes, laboratory evaluation of the coating quality is part of the production process. Performed in real time, it is necessary for technicians to be able to make rapid evaluations of test data. Testing frequently encompasses peel adhesions, thicknesses and coat weights. Because we are dealing with a semi-continuous web process, it is not uncommon to periodically evaluate three or more samples from across the web in order to determine process behavior. Unfortunately, there are two problems with this: the multiple samples are not independent and the sample size may be too small to interpret unambiguously. The inherent statistical problems with small, multiple specimen, samples from a web have been addressed by Frost (1), Roisum (2) and others.

Statistics of measurements

A distribution describes the frequency with which ranges of test results occur. The frequency distribution that results from the natural variation of outputs is the normal distribution. The normal distribution is characterized by having the most frequently occurring result centered, being symmetric, and approaching zero frequency asymptotically

on both sides of the mean. This shape is also referred to as the bell curve as shown in Figure 1.

The central limit theorem states that the distribution of means of samples drawn from a population with a stable distribution will be normal even if the original population's distribution is not normal. Thus, the use of normal statistics is common in process control methodologies.

The Variance is a measure of the spread of a distribution. The Standard Deviation is the square root of the variance and is symbolized by the Greek letter sigma: σ . In normally distributed results, 99.7% of the outcomes will fall within plus or minus three standard deviations of the mean. These bands of probability are frequently labeled as Zone A (the bands nearest the mean between zero and one sigma), Zone B (the bands between one and two sigma on either side of the mean) and Zone C (the bands between two and three sigma on either side of the mean) as shown in Figure 2. As a process becomes more consistent and repeatable with less variation, the value of the standard deviation becomes smaller.

Control limits are the limits of variation that we agree to accept as "natural" based on our understanding of the variability of the system. Many manufacturing processes commonly have control limits set at $\pm 3\sigma$. This has been good enough in the past but no longer satisfies many customers. Today's good processes will have standard deviations sufficiently small that no result will occur outside $\pm 4\sigma$ or $\pm 5\sigma$. And, there is a strong drive to take high volume processes to the $\pm 6\sigma$ level where we expect inconsistent outcomes only a few times in one million events.

A prime consideration when looking at test results is the stability of the measurement system. Measured outcomes will have a variance that reflects the sum of the variances of the independent processes working to create the outcome: that is, then variance in test results is the sum of the variance in the process output and the variance in the test method. Understanding the significance of any test measure, or any set of test measures, depends on knowing the population distribution from which the results were taken. So, a good analysis of test results will be based on solid expectations for a mean outcome as well as a firm understanding of the variances of the contributing processes. Such a basis allows identification of results that are outside expectations and indicative of a special event.

Empirical Rules

In several contemporary laboratories, decisions are being made on the two-out-of-three rule. There are two versions of this rule-of-thumb. The first says that if from a set of three readings, one is significantly different from the other two, it should not be considered in a determination of the sample mean. This rule is clearly subjective and can lead to different evaluations from person to person and time to time. The second variant of this rule allows that if a data point is over a specification limit, then perform a recheck; if the recheck does not verify the out-of-spec condition then perform a third test and accept the results of two-out-of-three. This rule is easy to apply but may lead to the acceptance of an out-of-control process and its product.

In the early 1980s, a mentor provided lab technicians another rule of thumb for determining if one value in a set of five was an outlier – a point that was most likely due to a special cause

and therefore could be eliminated from statistical calculations. The rule was: if the difference between any one point and the mean of the set was more than four times the difference between the means of the set calculated with the point and a new mean calculated without the point, then that point is an outlier and can be discarded from the set. This was a good empirical rule for quick use in the days before electronic calculators, "Statistical-Packs" and laptops. The effect of any outlier on the average is approximately $1/j$ of its distance from the mean, where j is the number of elements in the set. Most likely, the justification for this rule-of-thumb test is rooted both in normal statistics and in Chauvenet's Criteria.

Chauvenet's Criterion

Occasionally one or more measurements in any set may exhibit considerable variance from the mean. How do we evaluate a small, independent set of test results such as we might get from periodic adhesion testing of the continuous coating process when we have no basis for an expected outcome? Measurements such as 11.2, 9.3 and 9.4 oz/inch are not unreasonable for multiple test results from one section of a web. However, it is not clear what should be accepted as the average for this process. Is the value of 11.2 within the expected range or should it be discarded from the data before computing the mean? Clearly, the blind application of the two-out-of-three rule could easily lead to an erroneous decision.

These questions arise often in the sparse data environment of new developments and a reasonable decision cannot be made without understanding the distribution of expected results from the sampled population. Some objective criterion is needed to evaluate the data: it is here that Chauvenet's Criteria may be used.

Chauvenet's Criteria states that an observation should be discarded if the probability of its occurrence is less than or equal to $1/2*j$ where j is the number of observations. That is, the probability of an outlier being included in any data set is less than $(1/2j)*j$ or 50%. Thus, of a set of five readings, one value may be discarded as an outlier if its probability of being in the parent population is less than $1/(2*5)$ or 10%. From the normal distribution table, this corresponds to +/- 1.65 standard deviations from the mean. Similarly, from a set of ten readings, values outside +/- 1.95 standard deviations are outliers under the criterion.

Examination of a data set may show more than one point which should be discarded. Chauvenet's Criteria is based on the assumption of a normally distributed population from which all values are taken. Thus, unlike the Q-test, his criteria may be applied more than one time on the same data set. (It has been argued that the criteria may not be applied to a set of four or fewer data points. Thus, the criteria should be applied only once to a data set of five values [4].)

Now consider the following set of test results, expanding the original three points to five:
Case a) 11.2, 9.3, 9.4, 10.2 and 10.5 ounce/inch. The sample mean is 10.1 ounce/inch with a sample standard deviation of 0.82. Thus, there are no outliers in this set of data as all of the values lie within the +/- 1.65 standard deviation range (between 8.75 and 11.45 ounce/inch).

Equally, it is possible that additional tests give results like case b) 11.2, 9.3, 9.4, 9.2 and 9.3 ounce/inch; these results are significantly different from Case a). Now, the sample mean is 9.66 with a sample standard deviation of 0.86. Chauvenet's Criterion suggests an upper limit of 11.1; thus, the value of 11.2 is an outlier and should be discarded from the data set. The mean and standard deviation for the sample are recomputed using the four accepted values yielding a true mean of 9.28 with a sample standard deviation of 0.10.

How different is our understanding of those first three data points when we also can see additional information. Chauvenet's criteria has found applications in astronomy, nuclear technology, geology, epidemiology, molecular biology, radiology and many other fields of physical science. It provides a simple tool for evaluating single sets of data for which there was (1) insufficient data to establish reliable three sigma limits and (2) no clear expectation for a particular result.

But who was Chauvenet?

[This sidebar information taken primarily from *Edited Appleton's Encyclopedia, Copyright © 2001 Virtualology (3)*]

William Chauvenet was born in Milford, Pennsylvania, May 24, 1820. He studied in Philadelphia and graduated from Yale in 1840. He served in the US Navy on the United States steamer "Mississippi," and in 1841 succeeded to the Chair of Mathematics at the Naval Asylum in Philadelphia.

The "Shore School for Midshipmen at the Naval Asylum" in Philadelphia was an informal and poorly equipped program in 1842. This quickly changed when Professor Chauvenet arrived and took charge. A tougher course of study was enacted and new instruments were obtained for the student's studies. However, in 1842, the course of study allowed for midshipmen at Philadelphia was only eight months long (one academic year). Chauvenet worked for the next three years to expand the curriculum from one year to two.

William Chauvenet was very active in the movement that led to the establishment of the United States Naval Academy at Annapolis. His comprehensive two-year curriculum plan for the education of naval officers submitted to Secretary of the Navy, George Bancroft culminated in the establishment of the Naval School at Annapolis, MD in 1845.

Professor Chauvenet taught mathematics and astronomical navigation from the very first day of classes at this new Naval School, later teaching navigation and surveying.. He continued to work tirelessly to improve the school, equipping it's first astronomical observatory and developing the plan which extended the course of study from two years to four, a program that still exists today. In recognition of the significant changes that had occurred, the Naval School was renamed the U.S. Naval Academy in 1850. He headed the Naval Academy's Department of Mathematics 1845-1853. Professor Chauvenet became the head of the Department of Astronomy and Navigation in 1853. He was the first commandant at Annapolis, a position he held until 1859

In 1859 he was elected to the chair of mathematics in Washington University, St. Louis, Mo. Here also he gained the esteem and confidence of those with whom he was associated, and in 1862 he was chosen chancellor of the university.

By 1864 his health had begun to fail, and he spent some time in Wisconsin and Minnesota, but again resumed academic duties in 1865. He was obliged to resign the offices held by him in 1869, and then spent some time in travel, but without avail.



William Chauvenet (from the web - no attribution)

William Chauvenet died in St. Paul, Minn. on December 13, 1870. By the time of his death, Chauvenet was a renowned American scholar. He was a member of numerous scientific societies, and in 1859 general secretary of the American Association for the Advancement of Science, with which he had been connected since its first meeting. He was a founding member of the National Academy of Sciences, and at the time of his death its vice-president. He also was a member of both the American Academy of Arts and Sciences and the American Philosophical Society. Besides numerous contributions to the "American Journal of Science," "Proceedings of the American Association for the Advancement of Science," Gould's "Astronomical Journal," and the "Mathematical Monthly," he was the author of "Binomial Theorem and Logarithms for the Use of Midshipmen at the Naval School" (Philadelphia, 1843); "Treatise on Plane and Spherical Trigonometry" (1850); "Manual of Spherical and Practical Astronomy" (1863); and "Treatise of Elementary Geometry" (1870).

Chauvenet is buried in Bellefontaine Cemetery in St. Louis, MO but we remember him today for his contributions. Beginning in 1925, the Mathematics Association of America has given an annual math prize in his name. In 1969, Chauvenet Hall was dedicated at the U.S. Naval Academy. It houses the Department of Mathematics, as well as numerous classrooms and laboratories. *U.S.N.S. Chauvenet* was built by Upper Clyde Shipbuilders of Glasgow, Scotland in 1970; she was converted to a maritime school ship in 1996 is still working today.

X-mR Control Charts

But distributions may change with time and control charts offer a method to monitor changes to the distribution of results based on the trends of samples pulled from that population. Types of control charts based on sampling the parameters are the \bar{X} -R, \bar{X} -s and median-R charts. In the laboratory environment, however, it may be inconvenient, or even impossible, to obtain large samples of product or process data at regular intervals for analysis. One control charting technique that often will fill this gap is the X-moving Range (X-mR) chart.

The X-mR chart is a Shewhart-type chart: two graphs, one tracking the center of the distribution and one tracking the spread. Together this chart set shows the measure-to-measure variability of data. While it is common to use two consecutive readings to create the moving range values, it also is possible to roll several readings and improve the sensitivity of the chart. The central tendency chart uses individual values and is the same as the traditional

run chart except that control limits are estimated from the moving range data. X-mR chart constants for the calculation of control limits can be found in Smith (5) or any other basic SPC text. For $n=2$, $CL_x = \pm 2.66 \text{ mRbar}$ and $UCL_R = 3.27 \text{ mRbar}$ where mRbar is the average of the moving range values.

The obvious uses for the X-mR chart are in applications where there is insufficient time or information to collect enough samples to establish traditional control charts. Wheeler (6) makes a compelling case for the use of X-mR charts in many non-traditional applications such as inventories, shipping costs, reportable accidents, material costs, labor usage and even with monthly financials. Certainly, intermittent laboratory data will qualify for this approach.

Even in processes where we have experience, there are occasions when we cannot estimate the outcome: new materials, new equipment or a new process paradigm. For example, coat weight data, shown in Table 1, was gathered from an experimental Meyer Rod coating application on a saturated paper backing:

Since this data was collected from an experimental run with a new adhesive in an unusual viscosity range, there is no comparable historical data on which to base estimated control limits. Yet there is a desire to establish guidelines rapidly to provide information to the machine operators. The X-mR chart as shown in Figure 3 allows control limits to be estimated with as few as 10 to 15 readings on this new system.

The charts can be examined for statistical indications of out-of-control situations as exemplified by specific patterns in the charted data. Any pattern that has a low probability of occurring naturally (usually less than about 0.5 %) is interpreted as indicating a change in the sampled distribution. The rules for the interpretation of X-mR control charts are similar to the rules for Xbar-R charts: any one point outside zone A; two of three consecutive in zone A, same side; four of five consecutive in zone B or beyond, same side; shifts; runs, mixtures, stratification, grouping, cycles, etc. As with other control charts, the limits were developed empirically and therefore are likely to be effective even if the population distribution is not normal. For the coat weight example above, the charts indicate a process in control, even at this early stage.

The X-mR chart provides statistically valid interpretation of trend data for new processes or processes for which data is sparse. For product development, it has advantages over pre-control charts in that no artificial product tolerance estimate is required to use the charts.

Summary

A statistical perspective is necessary because it yields consistently valid evaluations of test data. Two simple concepts for validating data and data trends are based on traditional statistical approaches and common sense: Chauvenet's Criteria and the X-mR chart. Use Chauvenet's Criteria to examine single sets of variable data to eliminate outliers; use the X-mR charts to monitor trends in sparse variable data.

As scientists, we should use, and provide people with, statistical tools that are easy to use. We should encourage the use of statistical approaches for trending key data and for analyzing one-time sets of values when dealing with numerical data. Only in this way will we be able to turn laboratory test data into useful information.

References

1. Frost, Paul Application Of SPC To Roll Products Third Ed
PJ Associates North Quincy, MA
2. Roisum, David R, PhD “Secrets of a Level Process and Product” January 2001
<http://roisum.com/documents/Level.pdf>
3. *Edited Appletons Encyclopedia, Copyright © 2001 VirtualologyTM*
Appleton's Cyclopedia of American Biography, edited by James Grant Wilson and John Fiske. Six volumes, New York: D. Appleton and Company
4. several websites provide guidance on the application of Chauvenet's Criterion
<http://www.wpi.edu/Academics/Depts/ME/Nuclear/Reactor/Labs/R-stat.html>
<http://instructor.physics.lsa.umich.edu/ip-labs/tutorials/errors/illerr.html>
http://www.chemistry.ohio-state.edu/~coe/chem541_au03/Chauvenet_coe.pdf
5. Smith, Gerald M Statistical Process Control and Quality Improvement Fourth Ed
2001 Prentice Hall Upper Saddle River, NJ
6. Wheeler, Donald J Understanding Variation the Key to Managing Chaos Second Ed
2000 SPC Press Knoxville, TN

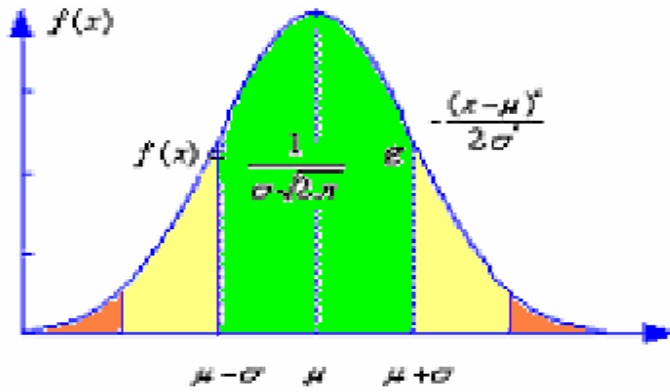


Figure 1. The normal distribution curve and its probability function.

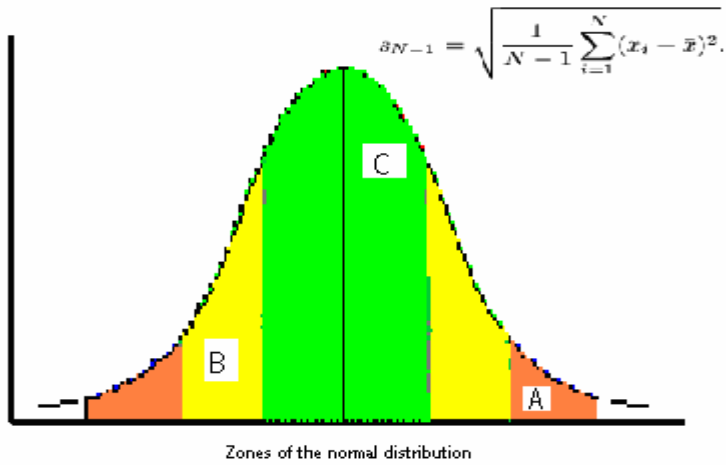


Figure 2. The zones of the normal distribution and the standard deviation function.

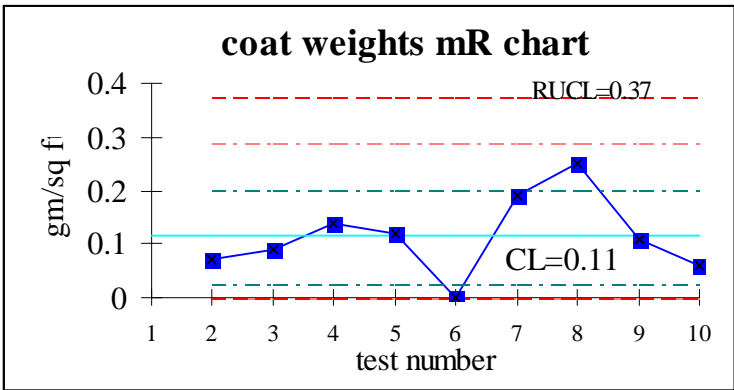
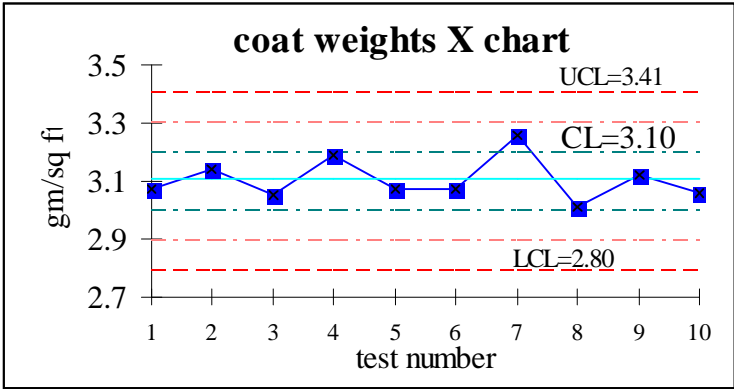


Figure 3. Example of X-mR charts for a new adhesive system.

reading number	Left Coat Weight	Center Coat Weight	Right Coat Weight	Reported Coat Weight	moving Range (2)
1	3.06	3.05	3.11	3.07	
2	3.09	3.15	3.18	3.14	0.07
3	2.96	3.15	3.04	3.05	0.09
4	3.29	3.18	3.09	3.19	0.14
5	3.07	3.07	3.06	3.07	0.12
6	3.12	3.07	3.02	3.07	0.00
7	3.27	3.26	3.25	3.26	0.19
8	3.00	3.06	2.97	3.01	0.25
9	3.14	3.03	3.18	3.12	0.11
10	2.99	3.06	3.14	3.06	0.06

Table 1. Table of Average coat weights and associated moving ranges.

Biographical data

James August, CQA **Director of Quality Management, American Biltrite Inc.**

James August is a quality professional with over twenty years experience in industrial and military/aerospace manufacturing. He holds a Bachelor of Engineering from SUNY Stony Brook, a Masters in Materials Science from the same university and has completed post graduate work at the Colorado School of Mines. Starting as an Engineer and Metallurgist, Mr. August's career evolved toward quality due to early experiences establishing manufacturing systems. Currently, Jim is Director of Quality Management for American Biltrite Tape Products Division where he led their recent transition to ISO 9001:2000.

James August rejoined ASQ in 1992 and is a Certified Quality Auditor since 1993. Mr. August is active in Quality New Jersey since 1991 and serves as the Administrator for their Performance Excellence Network. He has been a member of the QNJ Governor's Award for Performance Excellence Board of Examiners for the last nine years (Baldrige criteria), and a Senior Examiner for five years. He is a case study analyst for the ANSI/ASQ sponsored ISO 9000:2000 Product Support Initiative since 2001. James August also is a member of ASCD and a Registered Professional Development Provider in New Jersey.