# INTER-LABORATORY TESTING

ELWYN LEWIS (ASQ) CQT, CQE, CQA, CQMgr

## INTRODUCTION

Whilst reviewing the data received from a recent Inter-laboratory study conducted by the South African Bureau of Standards (SABS), it became evident that the method used for determining outliers seemed a bit vague. Little or no input towards identifying the probable cause for the outlier could be extracted from the reports. If the outlying laboratory wishes to correct the faults then the information supplied by the inter-laboratory test is insufficient. Surely there must be a more meaningful method of analysing the results from an Inter-laboratory study and based on this belief I have since undertaken a study of 'some' Inter-laboratory testing articles, I have also combined other proven and accepted techniques commonly used in industrial fields other than chemical.

This document does not supply all the answers but should help towards initiating a lateral thinking towards solving the problem. I would like to share what I have done, to date, with all interested people as well as laboratories interested in Inter-laboratory testing programmes.

## THE South African Bureau of Standards Approach

Using genuine data from Grant Wernimont's article, "Design and Interpretation of Inter-Laboratory Studies of Test Methods - Analytical Chemistry 1951", I have compiled an example to help illustrate the dilemma.

The test method used is the Eberstadt method for determination of Acetyl in cellulose acetate. A single observation from each of the eight laboratories involved has been selected and analysed using the same method as the SABS uses.

| Laboratory | Result |
|---|---|
| 1 | 39.16 |
| 2 | 38.96 |
| 3 | 39.24 |
| 4 | 39.26 |
| 5 | 39.08 |
| 6 | 39.00 |
| 7 | 38.84 |
| 8 | 38.85 |
| Average | 39.049 |
| R for Method | 0.3 |
| ♦ | 0.1638 |
| No of Outliers | 0 |

From the above tabulation it can be seen that no outliers are detected.

In addition to the SABS method we may also include the Dixon's test for outliers, which is quiet common :

For the Higher values:

$$r_{11} = \frac{39.24 - 39.26}{38.84 - 39.26} = 0.0476$$

For the Lower values :

$$r_{11} = \frac{38.85 - 38.84}{39.24 - 38.84} = 0.025$$

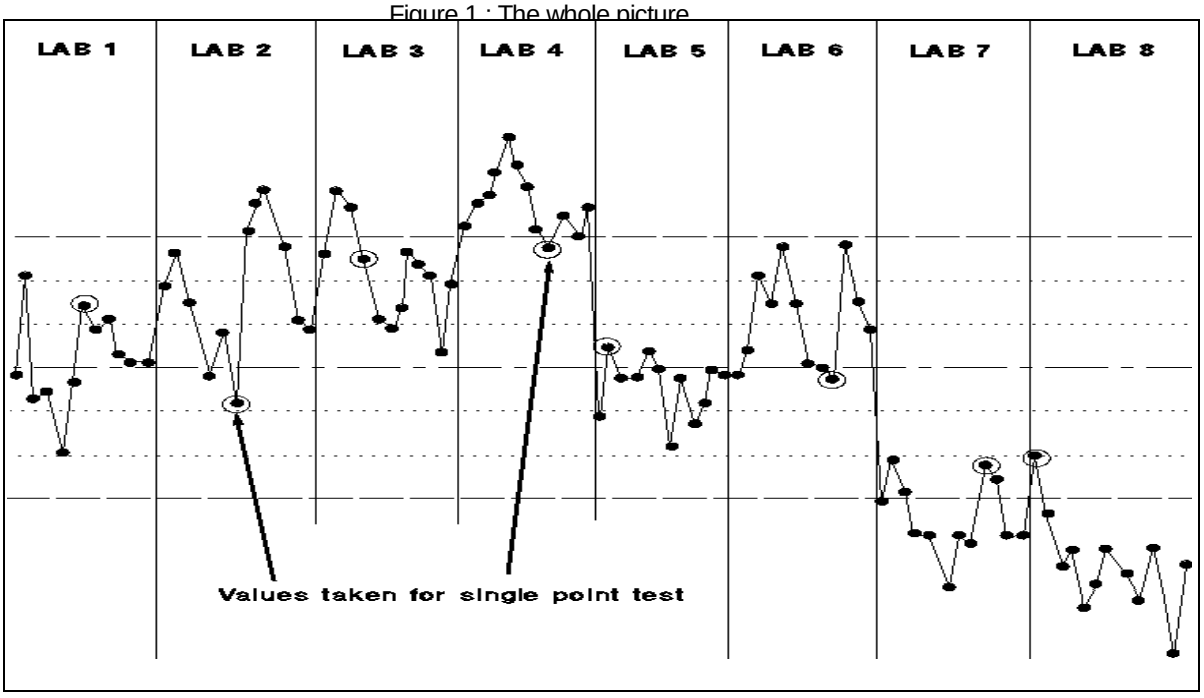
The Dixon value in the tables (for n = 8) at a confidence level of 95% = 0.554

Due to the above two calculated values being smaller than the table value it can be deduced that there are no apparent outliers in the data.

THE WHOLE PICTURE

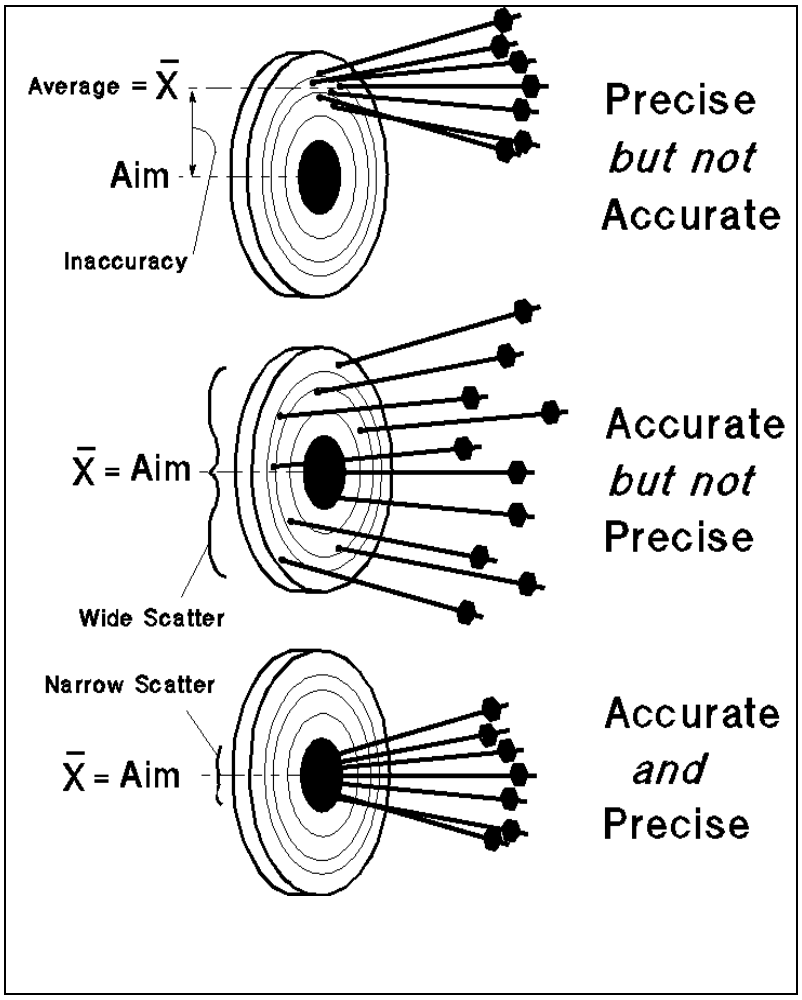
Let us now take a look at the whole picture and see from where our single samples are taken. The values used in the above single point test are circled.

Figure 1 : The whole picture




It can be easily seen that single points do not display the whole picture and erroneous decisions can be made. Once the data is plotted, just a glance at the chart

shows clearly that there is in fact a difference between laboratories. Before we go any deeper into the subject let us first determine the difference between accuracy (bias) and precision. The easiest way to explain is with the use of the target analogy:




**Accuracy**: The accuracy of a measuring process is defined as the extent to which the average of a series of repeat measurements made on a single unit of product differs from the true value. In most cases the difference can be assigned to the system being out of calibration.

**Precision**: The precision of a measuring process is the extent to which the system repeats the results when making repeat measurements on the same unit of product.

ASTMDEFINITION: PRECISION

> The precision of a measuring process, and hence the stated precision of the test method from which the process is generated, is a generic concept related to the closeness of agreement between test results obtained under prescribed like conditions from the measurement process being evaluated. The measurement process must be in a state of statistical control; else the precision of the process has no meaning. The greater the dispersion the poorer the precision.

ASTM DEFINITION: BIAS (ACCURACY RELATING TO A MEASURING PROCESS)

> The bias of a measurement process is a generic concept related to a consistent or systematic difference between a set of test results from the process and an accepted reference value of the property being measured. The measurement process must be in a state of statistical control; else the bias of the process has no meaning.

ASTM DEFINITION: STATISTICAL CONTROL

> The measurement process is in a state of statistical control when the test results obtained vary in a predictable manner, showing no assignable trends, cycles, abrupt changes, excess scatter, or other unpredictable variations as determined by the application of appropriate statistical methods. The endurance of a state of statistical control is not a simple

matter, but it may be helped by the use of control charts.
Before we can make any assumptions on the measuring processes precision we must be confident that there are no abnormal influences in our process and the only way that we can determine this is through the application of the Statistical Process Control (SPC) to our measuring processes.

It stands to reason that by taking a single result from a laboratory's measuring process, without knowledge of the precision and accuracy, can not only be misleading, but can lead us to incorrect conclusions and have negative effects on our laboratories and analysts. It is the responsibility of the respective laboratory management to ensure that all the inputs into the measuring process are capable

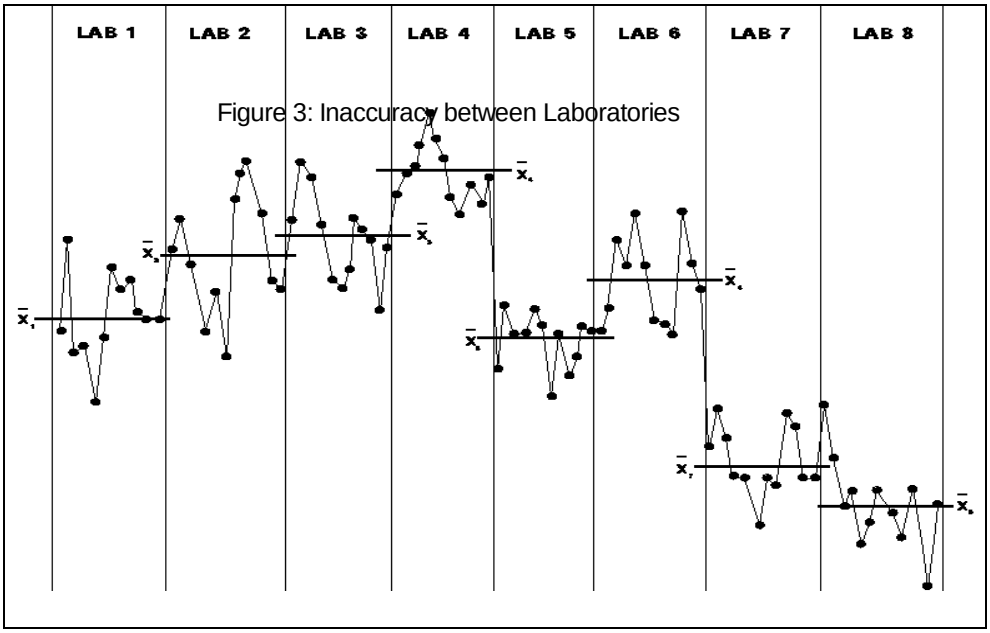


Figure 3: Inaccuracy between Laboratories

Now if we return to our whole picture, as seen in figure 3, it may be easier to answer the question, 'Do we have an accuracy problem or do we have a precision problem?'

It becomes quite clear that there is in fact a problem with the accuracies of the various Laboratories. If we remove the inaccuracies, by aligning the averages in one line, then we can develop a picture of the various precisions. Is there one laboratory that looks as if there is excessive variation?

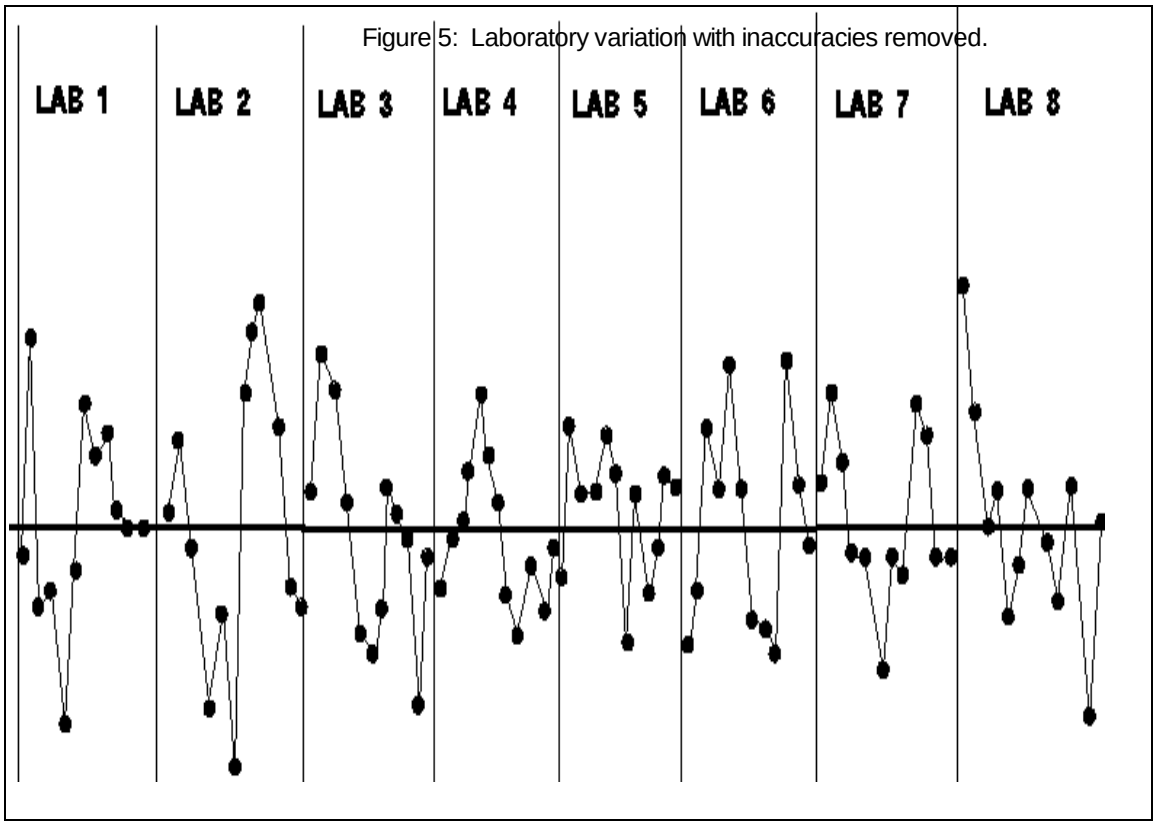
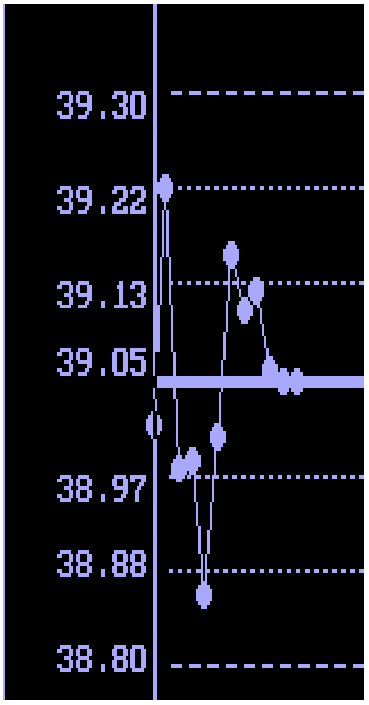


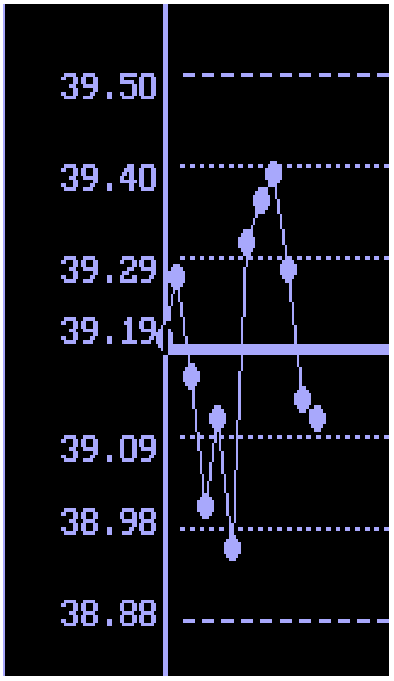

Figure 5: Laboratory variation with inaccuracies removed.

Figure 5 seems to indicate that there is no significant difference between the laboratories variation. We can verify this by viewing the individual charts of all eight laboratories below. Notice that none of the Laboratories are displaying any unpredictable variation.

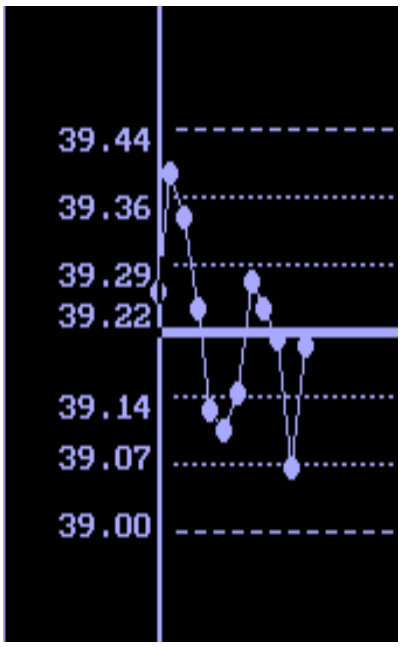
An and R chart as well as an x-bar and sigma chart annexure 1 and 2 respectively serve to confirm that there is no significant difference between the laboratories precisions. These charts also highlight the accuracy differences.



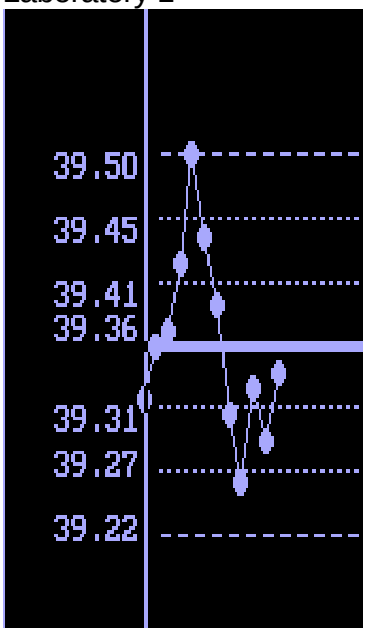
Laboratory 1



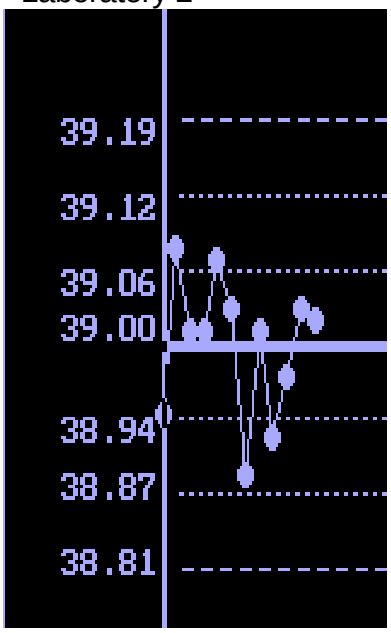
Laboratory 2



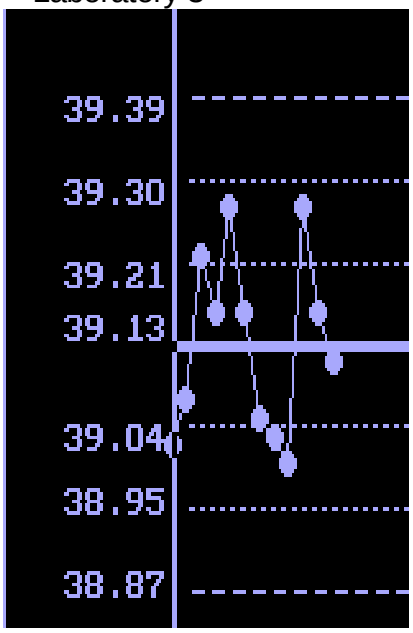
Laboratory 3



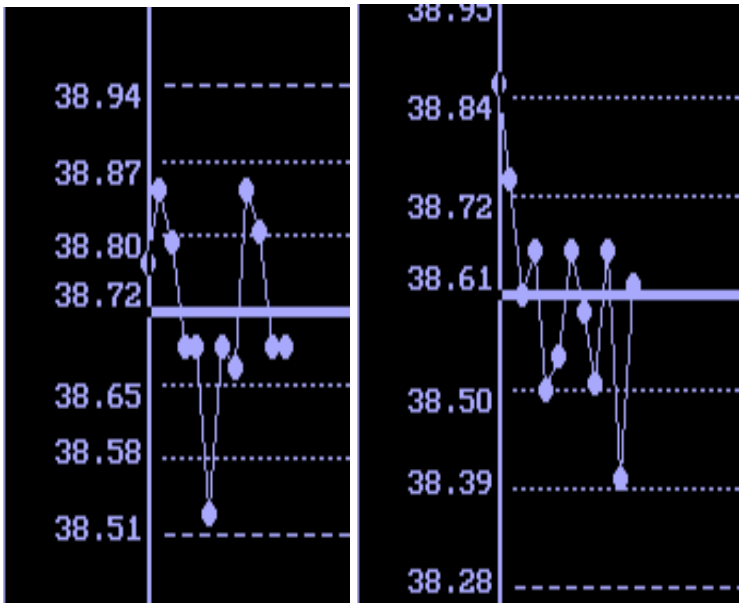
Laboratory 4



Laboratory 5



Laboratory 6




Laboratory 7 Laboratory 8

A DIFFERENT APPROACH

Inter-laboratory Precision

A useful alternative to single sample tests is to determine each laboratory's precision and accuracy before comparing Inter-laboratory results. In other words, each laboratory must first prove their capability to be both accurate and precise, by repeating the test method at least ten times. These results must be plotted on a control chart in order to analyse the chart for assignable causes of variation.

Only after each Laboratory has proven it's precision, can a test for accuracy between Laboratories be performed.

The Calculation for the maximum overall variation (reproducibility) between the eight laboratories can be done as follows:

$$\blacklozenge_{TOTAL} = \sqrt{(\blacklozenge_{Lab1}^2 + \blacklozenge_{Lab2}^2 + \blacklozenge_{Lab3}^2 + \ldots + \blacklozenge_{Lab8}^2)}$$

To calculate sigma ($\blacklozenge_{Lab}$) for each laboratory we use the formula:

$$\blacklozenge_{Lab} = MR/d_2$$

This formula is used to calculate the upper and lower control limits for the individual charts, as well as the total Inter-laboratory precision as seen above. The MR is the ***average*** moving range. The MR is the positive difference between two successive sequential values. The $d_2$ value is a factor that is found in the tables and is dependent on the subgroup sample size, which in this case is two (n = 2). The $d_2$ value for n = 2 is 1.128

The MR must not be confused with the overall Range "R" of the subgroup. In our example there are 12 values per laboratory and this overall range is calculated by subtracting the highest value from the lowest value in the subgroup of 12. The ***average*** overall range $R_x$ is used in the accuracy calculation as shown on page 9.

For our example the total Inter-laboratory precision =

$$\sigma_{TOTAL} = \sqrt{(0.083^2 + 0.103^2 + 0.073^2 + 0.047^2 + 0.063^2 + 0.087^2 + 0.072^2 + 0.112^2)}$$

$$\sigma_{TOTAL} = 0.233$$

Therefore with 95.45% confidence our total inter-laboratory precision =

$$\pm 2\,\sigma_{TOTAL} = \pm 0.466$$

It is interesting to note that the $4\,\sigma_{TOTAL}$ of 0.932 is significantly bigger than the reproducibility of 0.6 ($\pm$ 0.3) allowed.

It is of utmost importance to understand that the calculated precision is only a theoretical value until such time as the inaccuracies have been eliminated.

Inter-Laboratory Accuracy

When we perform the test for accuracy it is ideal to have used a sample with a known value for the Inter-laboratory testing. Of course the laboratories involved should not be aware of the actual value beforehand. The following formula applies:

$$\text{Mean (xbar)} - \text{Actual} \pm 3R/d_2\sqrt{n}$$

In our example we do not know what the actual value is so we will substitute the grand average (❍) of the eight laboratory averages. So our formula for this exercise looks something like:

$$\text{Grand Average } (\mu) - \text{Xbar} \pm 3\,\mathbf{R_x}/d_2\sqrt{n}$$

Remember : the $R_x$ is the average of the eight individual laboratory ranges 'R' which are simply calculated by finding the highest value in each laboratory subgroup of twelve and subtracting it from the lowest value. The $d_2$ value is a factor that is found in the tables and is dependent on the subgroup sample size.

Let us first calculate: $3\,\mathbf{R_x}/d_2\sqrt{n}$

$$= (3*0.319)/(3.258*\sqrt{12})$$

= 0.085

Create a table showing the eight laboratories " **μ - xbar**:" and the comparison of the result to our $3/d_2/n$ value of 0.085.

| | | | |
|---|---|---|---|
| Lab 1 | 0.015 * < $3/d_2/n$ | **Pass** | |
| Lab 2 | 0.155 * > $3/d_2/n$ | **Fail** | |
| Lab 3 | 0.185 * > $3/d_2/n$ | **Fail** | |
| Lab 4 | 0.325 * > $3/d_2/n$ | **Fail** | |
| Lab 5 | -0.035 * < $3/d_2/n$ | **Pass** | |
| Lab 6 | 0.095 * > $3/d_2/n$ | **Fail** | |
| Lab 7 | -0.315 * > $3/d_2/n$ | **Fail** | |
| Lab 8 | -0.425 * > $3/d_2/n$ | **Fail** | |

If we refer to the **Xbar** and σ chart (annexure 2) you will notice that on the graph the only Laboratories between the control limits are laboratories 1 and 5 which are the same laboratories which passed our test for accuracy.

## CONCLUSION

Surely what has been extracted from the data is more meaningful than simply identifying a laboratory as an outlier. This example shows us what erroneous conclusions were made by the SABS and what risks are attached to single point sampling, especially when no information about the measuring process accuracy and precision is known.

With the help of some very simple tools and techniques we have extracted enough information to help solve the problem. We are now in a position to search for reasons that could explain why the accuracy's of laboratories 2, 3, 4 and 6 were significantly higher and why laboratories 7 and 8 were significantly lower. Was it, perhaps, due to different calibration conditions? This approach of using charts can only tell us what sort of problem exists, either accuracy or precision, it is up to us to find the reasons thereof and either remove them, fix them or compensate for them.

**Biography** - Elwyn Lewis

ANNEXURE 1

ANNEXURE 2