Laboratory Technology for the Mining Industry

This paper is reprinted from the February 1998 issue of the *Alchemist Digest*.

PRECISION CONTROLS

The absence or presence of a bias can not be proven if the bias is smaller than the uncertainty limits of the data. This means that, in order for us to test the accuracy of a measurement system to a specified level, we must first ensure that it produces data of adequate precision. The precision of a system is measured by the standard deviation of the data. This can be calculated using data sets incorporating replicate measurements.

Replicate insertion - While it is not economical to repeat every measurement several times, the inclusion of duplicate measurements is, since not every measurement need be duplicated in large sample sets. Given enough duplicate data, a reliable estimate of the standard deviation of the measurement system, and thus its precision, can be calculated. This is accomplished with the use of a simple calculation as follows:

Variance $v = s^2 = \Sigma R^2/2N$
R = the difference between duplicates
N = the number of duplicate pairs.
s = the standard deviation

There are three concerns that need to be addressed to ensure that the estimate of the standard deviation "s" calculated with this method is reliable. First, N should be at least 20. In addition, the concentrations of the samples from which the duplicates were selected should not differ by orders of magnitude (10 ppm to 500 ppm for example, is probably too large a concentration range - try calculating the variance for two ranges in this case: <100 ppm and > 100 to 500 ppm). The third concern is that the duplicates must not exhibit a bias between the duplicate sets - the sum of the differences between the individual duplicate pairs (with the sign taken into account) must be close to zero.

An example of this will help to illustrate the description of the calculation. Assume an industrial plant has an environmental permit to release its treated process stream into the municipal sewage system as long as the level of copper did not exceed 2.0 mg/L. In order to ensure compliance, the effluent stream was monitored on an hourly basis and the sampling and Cu analysis were performed in duplicate.

TABLE 1

[Cu] in waste stream, mg/L

Hour	Dup1	Dup2	R	R^2	Hour	Dup1	Dup2	R	R^2
01	1.01	1.20	.19	.0361	13	.97	.95	-.05	.0025
02	1.12	.95	-.17	.0289	14	1.01	1.11	-.10	.0100
03	.98	.97	-.01	.0001	15	.85	1.02	.17	.0289
04	.99	.91	-.08	.0064	16	1.04	.98	-.06	.0036
05	.99	1.13	.14	.0196	17	1.15	1.04	-.11	.0121
06	1.02	1.02	0	0	18	.98	.92	-.06	.0036
07	.97	.97	0	0	19	.96	.97	.01	.0001
08	1.02	1.14	.12	.0144	20	1.03	1.08	.05	.0025
09	1.12	1.03	-.09	.0081	21	1.05	1.05	0	0
10	.96	.95	-.01	.0001	22	.99	1.00	.01	.0001
11	1.18	1.05	-.13	.0169	23	.95	1.12	.17	.0289
12	1.10	1.12	.02	.0004	24	1.12	1.09	-.03	.0009

The results obtained over a 24 hour period are found in Table 1. From this data the values for ΣR and ΣR^2 must be determined, N of course is 24. The "s" is calculated as below: ΣR = -.020 (very close to 0 so no bias is evident)
ΣR^2 = 0.2242
$v = s^2 = \Sigma R^2/2N$ = .2242/48 = .004671
s = .0683 mg/L
mean [Cu] = 1.028 mg/L

There are two advantages associated with using duplicates to estimate the precision of a process. First, it is a realistic estimate of the precision since it is based on actual samples and not on repeated measurements on a standard sample that may not have the variability in it that is found in the real samples going through the process. Second, the overall variance obtained reflects the sum of the variances introduced at each step of the process after the inclusion of the duplicates.

Range charts - The routine inclusion of duplicates into measurement systems also permits a visual monitoring of the precision with range charts. There are three different types of range charts in common use - range control charts, range performance charts and range ratio charts.

Range control charts are established by accumulating the results of 20 or more sets of duplicates. The mean (R) of the absolute values of the difference (R) between each of the 20 duplicate pairs is calculated. The value (R) is then plotted as well as lines at 0.845 (R), 2.51 (R) and 3.27 (R). Once the chart has been set up the R values from duplicates obtained in routine work are plotted on it. Statistically, 50% of the values plotted should lie below the 0.845 (R) line, only 5% above the 2.51 (R) line and no values above the 3.27 (R) line. Since the difference between duplicate pairs is concentration-dependent (larger at the 500 ppm range for example, than at the 10 ppm range), it may be necessary to construct charts for different concentration ranges. Charts plotted for different concentration ranges are referred to as **subspan range control charts**. In all cases, the data is plotted in order of measurement sequence.

The construction of range performance charts allows one to avoid the necessity to use two or more subspan range control charts. To develop the range performance chart, the mean ranges R are obtained from duplicate data in the low, medium and high concentration ranges expected to be encountered in routine work. These values of R are then plotted against concentration. Control lines are plotted on the chart in a manner similar to those on the range control chart. Subsequent range values obtained from duplicates run as part of the routine work are plotted on the chart and the same statistical controls apply in this instance as in the previous case - 50% of the values plotted should lie below the 0.845R line, only 5% above the 2.51R line and no values above the 3.27R line. The major disadvantage is that the axis used for sequence of measurement in the range control chart is used to plot concentration and the sequence of measurement information is lost.

If it is important to keep the measurement sequence information and still accommodate large concentration ranges on one chart, the range ratio control chart can be used. In developing this type of control chart, a large number of samples (50 or more) run in duplicate and covering the appropriate concentration range are selected and used to construct a range performance chart as described above (excluding the control lines). From this plot, the expected range R_c for any concentration covered by the chart can be obtained. When duplicate pairs are run as part of the routine process, the difference between the measurements is calculated as R_o and ratioed to the expected range for the mean concentration of the duplicate pair R_c obtained from the plot. This gives the range ratio R_r ($R_r = R_o/R_c$) which is plotted on a graph with horizontal lines drawn at 1.0, 2.51 and 3.27 (no units). The range ratios are plotted in order of measurement sequence and the control rules are the same as in the previous two cases.

As an example, assume a laboratory was performing work for a company involved in drilling off a possible gold deposit. A duplicate sample was included in every batch of 20 samples. After 1000 samples had been processed there were 50 duplicate pairs. These were used to generate a least squares line for the value of versus the mean concentration of the duplicate pairs. Given that the equation of the line (in the format y = mx +b) is

$= 9.04R_c - 140$

R_c is the expected difference between duplicate results is the mean of the duplicate results.

With this equation we can now construct a range performance chart with the control lines being the correct multiple of the range line. In addition, using this graph, we can predict what the range is expected to be for these samples at a given mean concentration of a duplicate pair. As an example, the duplicate results for one pair are 950 ppb gold and 760 ppb gold. The mean value of the duplicate pair is 855. Putting these values into the equation above (or using a plotted graph) shows us that the expected range is 110 ppb. The range observed is 950 - 760 = 190 ppb. This value would be plotted on one of the first two types of range charts discussed. The ratio 190/110 = 1.7 would be plotted on the third type of range chart. Its value of 1.7 is less than 2.51 which indicates the difference between duplicates is within the range expected for a sample with 855 ppb gold.

Now that we have seen how to construct these charts to help keep track of duplicate data, the question is how do we use them to assure ourselves that the measurement process is giving adequate precision. This is done with a set of simple rules. The lines on the charts at 2.512 and 3.267 are labelled UWL and UCL respectively. These stand for "Upper Warning Limit" and "Upper Control Limit". These control limits are used to determine if a process is in control by applying the following set of rules:

1. If R is within the warning limits, accept all data.
2. If R is outside the control limits (UCL), the system is out-of-control. Reject all data since last-known-to-be-in-control and take corrective action. Reestablish control before accepting data.
3. If R exceeds the UWL but is within the UCL accept the measurements tentatively. If R for the next duplicate pair is within UWL, accept all previous data. If it exceeds the UWL, reject all data taken since the system was last-known-to-be-in-control and take corrective action. Reestablish control before accepting data.
4. Reestablishment of control should be demonstrated by the results of three consecutive sets of measurements that are in control.

The confidence levels of the UWL and UCL are 95% and 99.7% respectively which means that a system that is producing data of adequate precision should not exceed these limits very often. If it does then either the limits are set unrealistically tight or there is something wrong with the system that should be corrected. There is a tendency to set the control limits to give results of more precision than is required for the intended use of the data. This should not be done because it means the corrective action and rejection of sets of measurements is more frequent and the measurement process is more expensive than is needed. Neither of these situations is more acceptable than is poor data.

Using the example above, we can demonstrate situations of control or lack thereof. Consider the data set below which is a continuation of the gold drilling program:

TABLE 2

Dup 1	Dup 2	Mean	R_o	R_c	R_r
950	760	855	190	110	1.73
620	450	535	170	74.7	2.28
50	120	85	70	24.9	2.81*
375	270	322	105	51.1	2.05
75	50	62	25	22.4	1.11
100	370	235	270	41.4	6.57*
890	730	810	160	105	1.52

We have two data points (marked *) where R_r is outside the UWL (greater than 2.51). In the first instance, 2.81 > 2.512 but the next two data points are within UWL so the system is still in control (Rule 3). In the instance where R_r is 6.57 however, this value exceeds the UCL and the system is out of control (Rule 2). As a consequence, the data set it belonged with (20 samples) and the next one must be of suspect precision. According to the rules above, these two data sets must be rejected, the cause of the problem identified, corrective action taken and the samples in those two sets rerun. The duplicate in the set that gave R_r = 1.11 demonstrated that the system was still in control up to that point.