Magruder Check Sample Statistics

Report two and only two results on each analysis attempted. Each of the results reported must be done on different days. Collaborators are requested to use official AOAC methods or other standardized methods, and to indicate the method used on the report sheet. Three or four digit codes are used to designate the methods according to a code sheet supplied each participant upon annual subscription to the program. A single typewritten copy of the report sheet is forwarded to the Check Sample Committee chairman, due no later than the 10th of the month following receipt of the sample.

The Committee Chairman examines each report sheet received to see that duplicate results are reported, that the collaborator number is properly entered on the sheet, and that the results have been coded as to method in the units digit column of the code. (ALL RESULTS NOT CODED AS THE METHOD WHEN RECEIVED ARE ASSIGNED A "99" IN THE LAST TWO POSITIONS OF THE METHOD CODE NUMBER, IMPLYING A "MISCELLANEOUS" METHOD FOR THE PAIR OF DETERMINATIONS. ALL RESULTS ENTERED AS "MISCELLANEOUS" ARE NOT INCLUDED IN ANY STATISTICAL TREATMENT OR SUMMARY.) The Committee Chairman scans the reports to see that no gross errors are present, and sends the reports to Mr. Bob Coelho, Statistical Consultant, for computerized statistical treatment. The computer program is identical to that employed in the Magruder Fertilizer Series, for which Mr. Coelho is also Statistical Consultant.

Grand averages (arithmetic mean) and estimates of standard deviations are computed for each method code.

Outliers among analytical results are always a concern. Check Sample analyses are first screened for extreme values by an estimation procedure that is not affected by outliers. It accomplishes this by counting the number of results by a given method of analysis and by ranking results in the order of their magnitude. Counting in from each end of the ranking by sixteen percent of the count leaves sixty-eight percent at the center. The center sixty-eight percent is the fraction of normally distributed results that lies between one-sigma boundaries on each side of the mean. Consequently the values at the count boundaries are two standard deviations apart. One-half of the difference between them estimates the standard deviation of the set of analyses without any influence from outliers. The mean of the central sixty-eight percent is also entirely free of outlier effect.

Using these estimates of mean and standard deviation, the preliminary screening of outliers is at 3.5-sigma control limits in order to avoid over-censoring laboratory analyses because the count-boundary estimate, although not affected by outliers, is based on only part of the available information. Next, all of the screened analyses are used to calculate a mean, standard deviation, and average range of duplicates. These are the basis for control limits at 3.0-sigma to identify outlier analyses and ranges not identified at the wider control limits of the preliminary screening.

Outliers found by this first pass through the full-scale statistical calculations are omitted from a second pass that produces the summary statistics that are reported each month.

The precision within laboratories is reported as the average range because it relates directly to a laboratory's difference between duplicates, the range being defined as the absolute value of the difference. An average range is larger by a factor of 1.128 than the standard deviation which can be calculated from the same data. The upper 99.7 percent control limit for ranges of duplicates is the average range multiplied by 3.267. (The lower control limit is zero).

Outlier analyses and ranges cause a "flag" to appear in a report next to the related average of a laboratory's duplicate analyses.

Gross outliers found by the preliminary screening of analytical results are omitted from all statistical calculations. Borderline outliers found by Pass 1 and omitted from Pass 2 calculations of an analytical method's summary statistics are, however, included in the calculation of a laboratory's performance report.

In addition to these data summaries, participants in the Program receive a confidential report on their laboratory's performance. To help assess this report, a brief description follows of (a) statistical terms used, and (b) the basis for the numbers and letters that are on a laboratory's current report.

The performance reports show the relative bias, precision, and accuracy of the several methods rated for that laboratory. Bias is a measure of positive or negative deviation from the mean. It is usually a consistent effect - too low or too high. Precision describes the repeatability of results. Accuracy as used on the report card is a combination of bias and precision. The size of both the bias and the precision determines accuracy. The relationship is:

$$accuracy = \sqrt{bias^2 + precision^2}$$

A statistical technique called normalizing provides a way to evaluate a laboratory's performance when several check samples are analyzed by several methods. The performance ratings cover individual methods and the combination of these methods in a composite evaluation.

The general report lists, for each method, the mean and standard deviation of results from all consistent laboratories for the current check sample. Each result is judged on how well it agrees with the mean of all analyses by a given method. This is done by dividing the difference between the result and the mean by the standard deviation. The quotient is called a normalized value, index, or Z value. The Z values tell how well results agree with mean values.

Each performance card is based on use of normalized values from a laboratory's current report and from its reports on past check samples to calculate bias, precision and accuracy. This is done first for whichever individual methods were reported on the current sample and two past samples within the last twelve months. A method not used on the current sample is not evaluated. Then the individual evaluations are collected into a composite evaluation of performance.

A three-sample basis for the "running measure" of laboratory performance is used so that evaluations will be sensitive to any change in performance. Thus, a rating above or below normal for a result by a particular method will influence the ratings for the next two reports that show that method. (NOTE: This feature was actually turned off several years ago at the request of the AAFCO committee. We now calculate the statistics based on a single sample only.)

Five or more laboratories must use a method for it to be evaluated.

Letters that accompany evaluations of bias, precision and accuracy help to indicate performance. The number of laboratories whose evaluations are within 99.7 percent confidence limits is divided into thirds. The best one-third is marked A; the next best one-third is marked B; and the remaining one-third is marked C. Evaluations that are outside the 99.7 percent confidence limits are marked D.

Miscellaneous methods have code numbers which end in “99”. These are not evaluated. Nor are gross outliers evaluated.

The Z values provided by the confidential performance report card can be used to answer many questions about a laboratory’s general performance. When a method used to analyze a sample has been evaluated, on can compare (a) results on previous samples analyzed by the same method, (b) results by different methods for the same analysis, and (c) methods used for different types of analyses.

A laboratory’s standing among other laboratories is ranked according to the magnitude of its composite bias, precision and accuracy, beginning with a rank of 1 for best performance, corresponding to the smallest normalized value, Z.

Given the following example of normalized values from three check samples analyzed in duplicate by one analytical method, the calculation of accuracy, bias and precision is made in that order as show here:

$$
\begin{aligned}
accuracy &= \sqrt{\frac{1.38^2 + 1.91^2 + 2.57^2 + 2.11^2 + 1.94^2 + 2.59^2}{6}} \\
&= \sqrt{\frac{27.0812}{6}} \\
&= 2.124
\end{aligned}
$$

$$
\begin{aligned}
bias &= \frac{1.38 + 1.91 + 2.57 + 2.11 + 1.94 + 2.59}{6} \\
&= \frac{12.5}{6} \\
&= 2.083
\end{aligned}
$$

$$
\begin{aligned}
precision &= \sqrt{\frac{27.0812}{6} - \left(\frac{12.50}{6}\right)^2} \\
&= \sqrt{0.1733} \\
&= 0.416
\end{aligned}
$$

$$
and\, 2.124 = \sqrt{2.083^2 + 0.416^2}
$$

A laboratory's composite bias, precision and accuracy are obtained by including in the above calculations its normalized values for all the analytical methods included on its performance card.

There has been some confusion concerning the "precision" ranking on a lab report card. Some have thought that the closer the two replicate results are for a particular method, the better the precision should be on the report card. This is definitely **not** the case. The precision on the report card measures how precise the labs results are to the grand average over the course of three samples. The precision within a lab is reported as the average range because it relates directly to a laboratory's difference between duplicates, the range being defined as the absolute value of the difference. This precision is used primarily for screening for outliers as a first pass in the statistical treatment of the data.

In the following example, the results are shown of an actual lab with very good replicate precision, but with a poor overall precision. In the second set of data, only the second set of results was changed on two samples to match the grand average for those samples. While the two replicates were now quite far apart, the overall precision improved considerably.

If a lab does not run replicate results as required by the check sample program (several days apart), but instead reports a second result that is the same as, or very close to the first results, they may actually be hurting their performance ranking instead of helping it.

Actual Results

									Z" Values	
Sample	1st Result	2nd Result	Grand Average	Standard Deviation	1st Result - Average	2nd Result - Average	1st Diff / Std Dev	2nd Diff / Std Dev	Avg. Index	Avg. Index ** 2
9531	15.48	15.46	15.6172	0.23891	-0.1372	-0.1572	-0.57427	-0.65799	-0.61613	0.379618
9530	27.2	27.22	26.5297	0.27147	0.6703	0.6903	2.469149	2.542822	2.505986	6.279965
9529	49.46	49.47	48.8295	0.31024	0.6305	0.6405	2.032298	2.064531	2.048414	4.196

1.3128 Normalized Bias
1.9022 Accuracy Index
1.3767 Precision

Modified Results

Sample	1st Result	2nd Result	Grand Average	Standard Deviation	1st Result - Average	2nd Result - Average	1st Diff / Std Dev	2nd Diff / Std Dev	Avg. Index	Avg. Index ** 2
9531	15.48	15.46	15.6172	0.23891	-0.1372	-0.1572	-0.57427	-0.65799	-0.61613	0.379618
9530	27.2	**26.53**	26.5297	0.27147	0.6703	0.0003	2.469149	0.001105	1.235127	1.525539
9529	49.46	**48.83**	48.8295	0.31024	0.6305	0.0005	2.032298	0.001612	1.016955	1.034197

0.5453 Normalized Bias
0.9898 Accuracy Index
0.8261 Precision