Drug Development and Industrial Pharmacy, 22(9&10), 891-908 (1996)



# Helping Analytical Scientists Apply Statistics

Kwan R. Lee and Paul R. McAllister

*Statistical Sciences Department, Research and Development, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406*

## ABSTRACT

*The development and operating characterization of an analytical method or assay requires extensive collaboration between analytical scientists and statisticians. Minimal performance requirements, validation reporting and user expectations of analytical methods rely heavily on statistical concepts, calculations and terminology. This paper provides a descriptive and illustrative recitation of commonly used, and often misunderstood, statistical principles. The goal is to provide the analytical scientist with clearer understanding and practical guidance in the use of statistics. Percent relative standard deviation was used for a measure of precision through the sections dealing with nested experiments and variance components estimation, calibration experiments and sample size estimation. A comprehensive treatment of limit of detection and quantitation was given and the case when the calibration experiment has a nested error structure was also discussed.*

*Key Words: Calibration experiments; Limit of detection and quantitation; Nested error structure; Nested experiments; Relative standard deviation.*

## INTRODUCTION

A principle of extreme importance to the analytical scientist is that of precision. Scientists frequently use relative standard deviation (RSD) as a primary measure of precision for their analytical tests, bioassays, etc. RSD is the standard deviation expressed as a fraction of the mean. It is often expressed as a percent of the mean and is sometimes called the coefficient of variation (CV). Statistical questions and issues are often couched in terms of RSD or CV. The following questions are common in relation to the use of statistics for estimating and improving precision.

1. How do I use RSD to summarize the precision of an assay?
2. How do I calculate and use RSD when my precision study has two or more sources of variation—such as between days and within days, or solutions and injections, etc.?

3. How many replicates do I need to obtain an RSD of 5% or less?
4. How many concentrations should I use in determining the calibration line?
5. How many tests should be run for each unknown sample?
6. I want to have the assayed concentration at 100% of the nominal to be within ±1% of the nominal value. How many replicate tests do I need? How many different concentrations do I need to include in the calibration curve?
7. Can I use the calibration experiment to calculate the limit of detection and limit of quantitation?
8. When determining the standard curve, several determinations are usually made at each concentration. Does this change the way I calculate the standard curve?

In this note we attempt to give practical answers to the above questions based on our consulting experiences with SmithKline Beecham. Our approach is to give the general equations that describe certain relationships and then provide tables of results for ranges that appear frequently in practice. In this way, the theoretical results are translated immediately into practicable recommendations.

In addition, we present some standard statistical calculations in somewhat nonstandard formats because we believe the nonstandard presentation is more intuitive. In this area we have included estimation of variance components, precision of calibration lines, sample size calculations, and limit of detection and limit of quantitation. Finally, we consider the analysis of a calibration experiment in light of a possible lack of randomization in the order in which standards were run.

## TERMS AND DEFINITIONS

### Relative Standard Deviation

The relative standard deviation is intended to provide a quantification of variation that is independent of the units of measure. Having a quantification of variation that is independent of the units of measure allows for comparison between different situations. Thus, based on RSDs, one is able to pose a question such as "Which assay has smaller variation, the assay for final product concentration (with units milligrams/milliliter) or the assay for water content (with units % $H_2O$)?"

The unitless nature of RSD makes it almost a standard component for characterizing the performance of an assay. In many instances, assay developers are held to certain maximum values for the RSD of their assay. Furthermore, the RSD typically serves as an indicator, to those who may submit samples to be assayed, of the quality of the assay results. In this case, quality of the assay results generally means, "If I have the same samples reassayed at another point in time, how closely will the two sets of assay results agree?" For these reasons, it is important to fully understand the computation of RSD and its interpretation.

Definition 1: Relative Standard Deviation (RSD)

If $y_1, y_2, \ldots, y_n$ are independently obtained as a random sample from a population with mean $\mu$ and variance $\sigma^2$, the RSD of individual values, $y$, is given by $100\%(\sigma/\mu)$ and is estimated by $100\%(s/\bar{y})$. In this notation, $\bar{y}$ is the sample average and $s$ is the sample standard deviation, with $(n - 1)$ degrees of freedom ($df$).

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{1}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{(n-1)}} \tag{2}$$

$$\mathrm{RSD}(y) = \frac{100\% \cdot s}{\bar{y}} \tag{3}$$

We write RSD($y$) to indicate we are referring to the RSD of individual assay values.

There are several important points to note regarding definition 1. First, the data are represented by the values $y_1, \ldots, y_n$. The assumption is that the data are $n$ repeats of exactly the same assay steps on a single uniform sample of material appropriate to the assay. By exactly the same assay steps we mean that each value was obtained by executing the same series of operations, independently, $n$ times. For example, suppose that the assay involves preparation of the sample and then injection onto a column. Each value, $y_i$ should have been obtained by separately and independently preparing a sample from scratch and injecting that preparation onto the column. An example that would violate this principle would be if three preparations were made and four injections performed from each preparation. In this case, the 12 values would not all be obtained by executing the same series of steps. In fact, this second example re-

quires a nested analysis of variance, as described in the next section.

The value of $\mu$ in the definition is the "true value" of this uniform sample of material—that is, the value that would be obtained if every molecule in the sample could be unerringly identified and counted. The value of $\sigma^2$ is the "true" variation of the assay process. It is the value of $s^2$ we would expect to obtain if $n$ became very large (hundreds or thousands of repeats). Thus we would speak of the "true" RSD, given by (100%* $\sigma/\mu$) and its estimated value based on $n$ repeats of the assay as (100%* $s/\bar{y}$).

Finally, RSD($y$) measures the variation associated with individual assay results, obtained by repeating the same series of operations as those used to obtain the $y_i$, $i = 1, \ldots, n$. To the extent that actual real-world samples are analyzed by this same series of operations, this RSD will be an indicator of assay performance to the submitter of samples. In many assays, results for real-world samples may be obtained by repeating the series of operations several times and averaging the individual values. The person submitting the sample never sees the individual values but only the average of these several repeats. In such a case, the RSD($y$) for individual values is no longer a good indicator of assay quality as far as the submitter of samples is concerned. Rather, the assay developer must also calculate the RSD of the average of several results. We introduce the notation RSD($\bar{y}_k$) to indicate the RSD of the average of $k$ results. Fortunately, there is a straightforward relationship between RSD($y$) and RSD($\bar{y}_k$), as follows:

$$\mathrm{RSD}(\bar{y}_k) = \frac{\mathrm{RSD}(y)}{\sqrt{k}} \tag{4}$$

## Relative Error of Estimation

Reporting RSD as a measure of assay variation takes no account of the number of repeats, $n$, used to estimate the average and standard deviation. A more informative measure of variation would be one that includes a mechanism for accounting for sample size. The concept of a confidence interval is just such a mechanism.

A confidence interval is simply two numbers that are expected to include the true value of some quantity with some specified probability or degree of confidence. The difference between the two values is the width of the interval. The half-width of a confidence interval for many statistics is formed by multiplying the standard deviation of the statistic times a factor called a $t$ value. The $t$ value is a function of both the sample size used to estimate the standard deviation and the degree of confidence desired. The $t$ value decreases with increasing sample size, and the $t$ value increases with the degree of confidence desired. Thus, the width of the confidence interval incorporates both the absolute value of the variation and the sample size upon which the standard deviation is based.

In order to generalize the concept of confidence interval half-width to a relative measure, we define the relative error of estimation (REE) to be the relative half-width of a 95% confidence interval. Notationally, we write:

$$\mathrm{REE}(y) = t_{n-1,\ .05(2)} \cdot \mathrm{RSD}(y) \tag{5}$$

The value of $t_{(n-1),\ .05}$ is found in tables of the two-tailed $t$ distribution. A short version of such a table is included here as Table 1.

Of course (5), gives the REE for individual assay results. If the assay in question is usually performed as the average of several repeats, then the REE for averages can be easily defined:

$$\mathrm{REE}(\bar{y}_k) = t_{n-1,\ .05(2)} \cdot \mathrm{RSD}(\bar{y}_k) \tag{6}$$

***Table 1***

*Upper Percentage Points for the* t *Distribution*

| *df* | $t_{df,10(2)}$ | $t_{df,05(2)}$ | $t_{df,0.02(2)}$ | $t_{df,01(2)}$ |
|---|---|---|---|---|
| 1 | 6.314 | 12.706 | 31.821 | 63.357 |
| 2 | 2.920 | 4.303 | 6.965 | 9.925 |
| 3 | 2.353 | 3.182 | 4.541 | 5.841 |
| 4 | 2.132 | 2.776 | 3.747 | 4.604 |
| 5 | 2.015 | 2.571 | 3.365 | 4.032 |
| 6 | 1.943 | 2.447 | 3.143 | 3.707 |
| 7 | 1.895 | 2.365 | 2.998 | 3.499 |
| 8 | 1.860 | 2.306 | 2.896 | 3.355 |
| 9 | 1.833 | 2.262 | 2.821 | 3.250 |
| 10 | 1.812 | 2.228 | 2.764 | 3.169 |
| 11 | 1.796 | 2.201 | 2.718 | 3.106 |
| 12 | 1.782 | 2.179 | 2.681 | 3.055 |
| 13 | 1.771 | 2.160 | 2.650 | 3.012 |
| 14 | 1.761 | 2.145 | 2.624 | 2.977 |
| 15 | 1.753 | 2.131 | 2.602 | 2.947 |
| 16 | 1.746 | 2.120 | 2.583 | 2.921 |
| 17 | 1.740 | 2.110 | 2.567 | 2.898 |
| 18 | 1.734 | 2.101 | 2.552 | 2.878 |
| 19 | 1.729 | 2.093 | 2.539 | 2.861 |
| 20 | 1.725 | 2.086 | 2.528 | 2.845 |
| >20 | 1.75 | 2.0 | 2.5 | 3.0 |

*Note.* The values in this table are values for $t_{df,\alpha(2)}$. They are two-tailed critical values corresponding to total tail probability $\alpha$. To find the one-tailed critical values use: $t_{df\alpha(1)} = t_{df,2\alpha(2)}$.

The REE($y$) is valuable in helping to interpret any particular reported analytical result. We know with 95% confidence that the true value of that sample is within ±REE($y$) percent of the reported value. We suspect that many individuals mistakenly ascribe this interpretation to the RSD($y$).

### Maximum Relative Difference

Investigators submitting samples for analysis may have difficulty applying the RSD and REE directly to the interpretation of their experimental work. An investigator submitting samples might typically be interested in answering a question similar to: "How different might the results be if I submit two identical samples to this assay?" This question is relevant because the investigator needs to know when the differences observed between different samples are large enough to be attributed to causes other than the random variation in the assay. To answer this question, one needs the fact that the standard deviation of a difference is equal to the square root of 2 multiplied by the standard deviation of the quantities being differenced. Specifically, with 95% confidence, the maximum relative difference (MRD) between two identical samples due only to assay variance would be:

$$\text{Max. rel. difference for 95\% confidence} = \text{MRD}(y) = (\sqrt{2})\text{REE}(y) \quad (7)$$

### Numerical Example

The concepts described in the preceding sections are illustrated here with a numerical example. Suppose that an assay has been repeated completely independently 10 times on the same stock of uniform material. The 10 results are as follows:

12.3, 13.2, 14.2, 11.4, 12.5, 13.8, 11.7, 12.9, 13.2, 13.5

The following summary statistics are readily calculated:

$n = 10 \quad \bar{y} = 12.87 \quad s = 0.89697$

With the basic statistics above, the RSD, REE, and MRD are computed as follows:

For individual results: RSD($y$) = 7.0%; REE($y$) = 15.8%; MRD($y$) = 22.3%

For averages of 3 results: RSD($\bar{y}_3$) = 4.0%; REE($\bar{y}_3$) = 9.1%; MRD($\bar{y}_3$) = 12.9%

## NESTED EXPERIMENTS

### Components of Variance and Precision of Assay

In the performance of many assays, some steps in the assay procedure are nested within other steps. For example, high-performance liquid chromatographic (HPLC) assay may involve the preparation of several standard solutions and multiple injections from each standard solution. In this case, the injections are nested within the standard solutions. Each injection results in a peak area and subsequent concentration estimate. The final reported sample concentration is the average over all injections from all of the standard solutions prepared. Notice that if there are $n$ preparations and $k$ injections for each preparation, the $nk$ values cannot be considered completely independent. One would expect that results of injections from the same solution would be more alike than results from different solutions. This situation is called *intraclass correlation* among the injections. The difficulty caused is that when calculating REE or other summary statistics, ($nk$ - 1) cannot be used as the degrees of freedom. Using this value implies more confidence than is actually justified by the experiment. An example will help to clarify this issue.

Suppose five separate solutions of a drug substance were prepared at 100% of the nominal concentration, injected in triplicate, and the peak area/concentration ratio determined from an HPLC assay. The results are tabulated in Table 2. A proper statistical model for such an experiment is the one-way nested analysis of variance model given as:

$$y_{ij} = \mu + b_i + w_{ij}, \qquad i = 1,2,\ldots,n; \quad j = 1,2,\ldots,k \quad (8)$$

In Eq. (8), $y_{ij}$ are the observed measurements, $\mu$ is the true value of the common stock of material being assayed in this experiment, $b_i$ is a random component for between solutions variation, and $w_{ij}$ is another random component for within-solution variation. The terms $b_i$ are assumed to behave as if they are a random sample of values from a distribution having zero mean and a constant variance $\sigma_b^2$. Likewise, the terms $w_{ij}$ are assumed to behave as if they are a random sample of values from a distribution having zero mean and a constant variance $\sigma_w^2$. In other words, $\sigma_b$ is the standard deviation for solution-to-solution variation and $\sigma_w$ is the standard deviation for injection-to-injection variation. The components $b$ and $w$ are also assumed to be statistically independent.

*Table 2*

*Data for Triplicate Injections of Five Separate Solutions with RSD*

| | Injection 1 | Injection 2 | Injection 3 | Average | SD | RSD |
|---|---|---|---|---|---|---|
| Sol. 1 | 436580 | 437329 | 437881 | 437263 | 653 | 0.15% |
| Sol. 2 | 438901 | 435310 | 439396 | 437869 | 2230 | 0.51% |
| Sol. 3 | 440581 | 443031 | 439993 | 441202 | 1611 | 0.37% |
| Sol. 4 | 434588 | 435079 | 437780 | 435816 | 1719 | 0.39% |
| Sol. 5 | 440268 | 437124 | 440759 | 439384 | 1972 | 0.45% |
| Mean | 438184 | 437575 | 439162 | 438307 | 801 | 0.55% |
| SD | 2285 | 2877 | 1170 | | | |
| RSD | 0.58% | 0.74% | 0.30% | | | |

Under these assumptions, the variance of a single observed value $y$ is given by $\sigma_b^2 + \sigma_w^2$, which is usually called the total variance.

If an average value is calculated using $n$ solutions and $k$ injections per solution, the variance of such an average, $\bar{y}_{..}$ is:

$$V(\bar{y}_{..}) = \frac{\sigma_b^2}{n} + \frac{\sigma_w^2}{nk} \tag{9}$$

If we use statistical software to analyze such data, the analysis of variance presented in Table 3 would be typical (at least the first three or four columns of the table). In the next section we show how to perform these calculations in a spreadsheet format. Note that the total RSD is less than 1%. This total RSD is that which would apply to a single injection from one solution. The total variation is composed of nearly equal contributions from between and within solutions. The RSDs of the between and within components of the model are computed relative to the grand average of all the data.

The headings of Table 2 show the typical spreadsheet fields used commonly in analytical labs to show the precision of their assay results. The RSDs are computed separately for every solution and for each injection. The RSD of 0.55% of the bottom right-most corner is based on all 15 determinations. The individual RSDs are each based on only a few values and do not provide a concise summary measure of variation. The advantage of the analysis presented in Table 3 is that a single overall measure of variation is calculated and the contribution of each identified source of variation is explicitly estimated.

There is an alternative to the data analysis of this situation which is important to mention. Our objection to using all 15 data values as if they were a single random sample was that, within solutions, the $y_{ij}$ are not independent. However, we did assume that the solutions were prepared independently and so the group averages $\bar{y}_{i.}$, $i = 1, 2, \ldots, 5$ are independent. Hence, the five group averages can be treated as five independent samples from the population of averages of three injections. If the assay will always be performed by using three injections of every preparation and averaging these three values, then there is no loss of information in

*Table 3*

*Analysis of Variance Table for the Data in Table 2*

| | | Analysis of Variance | | | | | |
|---|---|---|---|---|---|---|---|
| Source of Variation | *df* | Sum of Squares | Mean Squares | Variance | Percent Variation | SD | RSD |
| Between solutions | 4 | 51078395 | 12769599 | 3267243 | 52% | 1808 | 0.41% |
| Within solutions | 10 | 29678695 | 2967869 | 2967869 | 48% | 1723 | 0.39% |
| Total | 14 | 80757089 | 5768364 | 6235113 | 100% | 2497 | 0.57% |

characterizing assay performance in terms of these averages. The statistical calculations are handled exactly as the example in Terms and Definitions, Numerical Example, with $n = 5$. The SD of the five averages is 2063 and this standard deviation has 4 degrees of freedom. Therefore, the appropriate $t$ value is $t_{4,.05(2)} = 2.78$. The statistics are as follows:

$$\text{RSD}(y) = (100\% \cdot 2063/438307) = 0.47\%$$
$$\text{REE}(y) = 2.78 \cdot \text{RSD}(y) = 1.31\%$$
$$\text{MRD}(y) = (\sqrt{2}) \cdot \text{REE}(y) = 1.85\%$$

We note that if one mistakenly treats all 15 observations as if they were independent, the resulting RSD($y$) would be 0.55%, but the REE($y$) would be 1.18%. The REE is misleadingly small and gives a false sense of precision for the assay.

## Sample Sizes Versus Precision of Final Assay

We now turn to the third question posed in the introduction, "How many solutions and replicates do I need to make the RSD of the final assay less than 5%?" Equation (9) is instructive for developing a strategy to answer this question. As a first step, Eq. (9) is rewritten in terms of RSDs as follows:

$$\text{RSD}(\bar{y}_{..}) = \sqrt{\frac{\text{RSD}_b^2}{n} + \frac{\text{RSD}_w^2}{nk}} \tag{10}$$

Table 4 was created from Eq. (10) using the RSDs in Table 3. This example illustrates how the RSD of the average changes with $n$ and $k$, when the component RSDs are nearly equal. Table 5 is a similar example,

***Table 4***

*RSD of Assay Results Reported as Averages Using* n *Solutions and* k *Injections per Solution ($RSD_b$ = 0.41% and $RSD_w$ = 0.39%)*

| Number of Injections per Solution ($k$) | Number of solutions ($n$) | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 0.56% | 0.40% | 0.33% | 0.28% | 0.25% |
| 2 | 0.49% | 0.35% | 0.28% | 0.25% | 0.22% |
| 3 | 0.47% | 0.33% | 0.27% | 0.23% | 0.21% |
| 4 | 0.45% | 0.32% | 0.26% | 0.23% | 0.20% |
| 5 | 0.44% | 0.32% | 0.26% | 0.22% | 0.20% |
| 6 | 0.44% | 0.31% | 0.25% | 0.22% | 0.20% |
| 8 | 0.43% | 0.30% | 0.25% | 0.22% | 0.19% |
| 10 | 0.43% | 0.30% | 0.25% | 0.21% | 0.19% |

***Table 5***

*RSD of Assay Results Reported as Averages Using* n *Solutions and* k *Injections per Solution ($RSD_b$ = 8% and $RSD_w$ = 2%)*

| Number of Injections per Solution ($k$) | Number of solutions ($n$) | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 8.25% | 5.83% | 4.76% | 4.12% | 3.69% |
| 2 | 8.12% | 5.74% | 4.69% | 4.06% | 3.63% |
| 3 | 8.08% | 5.72% | 4.67% | 4.04% | 3.61% |
| 4 | 8.06% | 5.70% | 4.65% | 4.03% | 3.61% |
| 5 | 8.05% | 5.69% | 4.65% | 4.02% | 3.60% |
| 6 | 8.04% | 5.69% | 4.64% | 4.02% | 3.60% |
| 8 | 8.03% | 5.68% | 4.64% | 4.02% | 3.59% |
| 10 | 8.02% | 5.67% | 4.63% | 4.01% | 3.59% |

but assuming unequal RSDs for the component variation.

Both tables provide useful tools for understanding approaches to achieving a "final goal" in terms of some specified RSD. Suppose for Table 4 our goal was to achieve 0.5% RSD. We can see that using just one solution and one injection will not achieve that goal. The most efficient increase is to use two solutions and only one injection per solution. For the case shown in Table 5 with $RSD_b = 8\%$ and $RSD_w = 2\%$, in order to achieve a goal of 5%, at least 3 solutions are necessary, but increasing the number of replicates for each solution does not have much benefit. Such a table can be created in a spreadsheet and "what if" analysis readily performed.

## Additional Components of Variance

### Design to Separate Sources of Variation

During assay development, RSD($y$) obtained from repeating the same series of operations characterizes the "best case" variation of the assay. The repeats are often obtained by the same analyst (often the developer of the method), using just one instrument, and repeated close together in time. In some contexts, RSD($y$) obtained under these conditions is called the *repeatability of the method*. Implicit in this naming is the recognition that repeated analysis of the same sample by different analysts, or by using different equipment, or spaced over longer time intervals, will not produce as close agreement of the $n$ results. That is, by allowing these other sources of variation to operate, the variation will naturally increase.

An important component of assay development is to understand the amount of variation contributed by these other sources. In order to estimate separately the contributions to variation from each of several sources, carefully designed data collection must be undertaken. Following data collection, statistical analysis of the data will provide estimates of variance contributed from each source separately. The estimates of variance contributed are often called *components of variance*. The statistician will speak of a "variance components analysis" to describe the data collection design and statistical analysis.

The most common and straightforward data collection scheme for estimating variance components is called a *nested* or *hierarchical design*. In this arrangement, sources of variation are isolated through a series of splittings or branches, such that whenever a split occurs, everything above a given split is identical.

As an example, suppose that a certain assay involves sample preparation and then injection onto a column. The assay developer is interested in understanding how much variation will be contributed by having different, well-trained analysts perform the assay. The developer also wants to know how much of the assay variation is due to the preparation phase and how much variation is contributed by the instrument and column. A possible data collection design might be diagrammed as presented in Fig. 1.

Note that there are five analysts involved in the plan. Each analyst is asked to make two independent preparations of material from a common stock. From each preparation two injections are made onto the column. From this arrangement a total of 20 assay results will be obtained. It is important to note that these 20 results

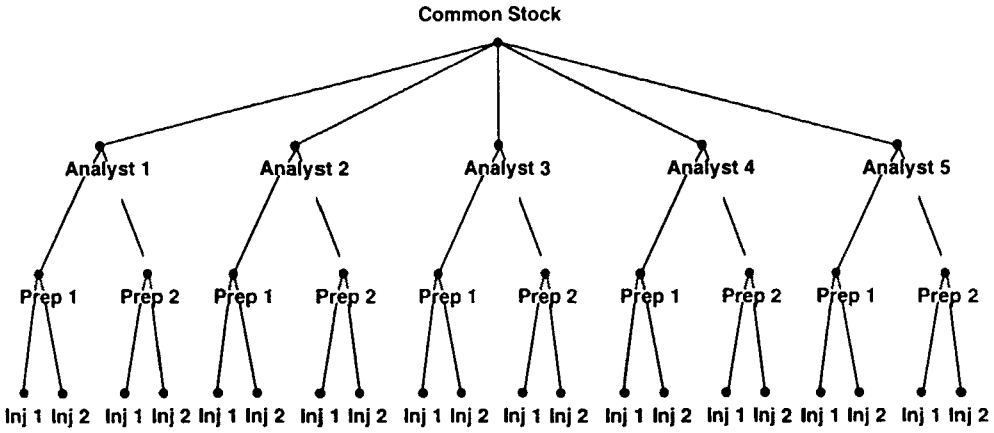



**Figure 1.** Nested diagram showing three factors: analyst, preparation, and injection.

are not equivalent with respect to the operations used to obtain them. Therefore, it would be inappropriate to simply calculate the average and standard deviation and form a single RSD.

From the above arrangement, three separate sources of variation may be identified. These are: variation due to the differences in analysts, variation due to differences in preparation by the same analyst, and variation due to differences caused by the instrument and column, when the same preparation is injected. A model for this situation is as follows:

$$y_{ijk} = \mu + a_i + p_{ij} + w_{ijk} \quad (11)$$

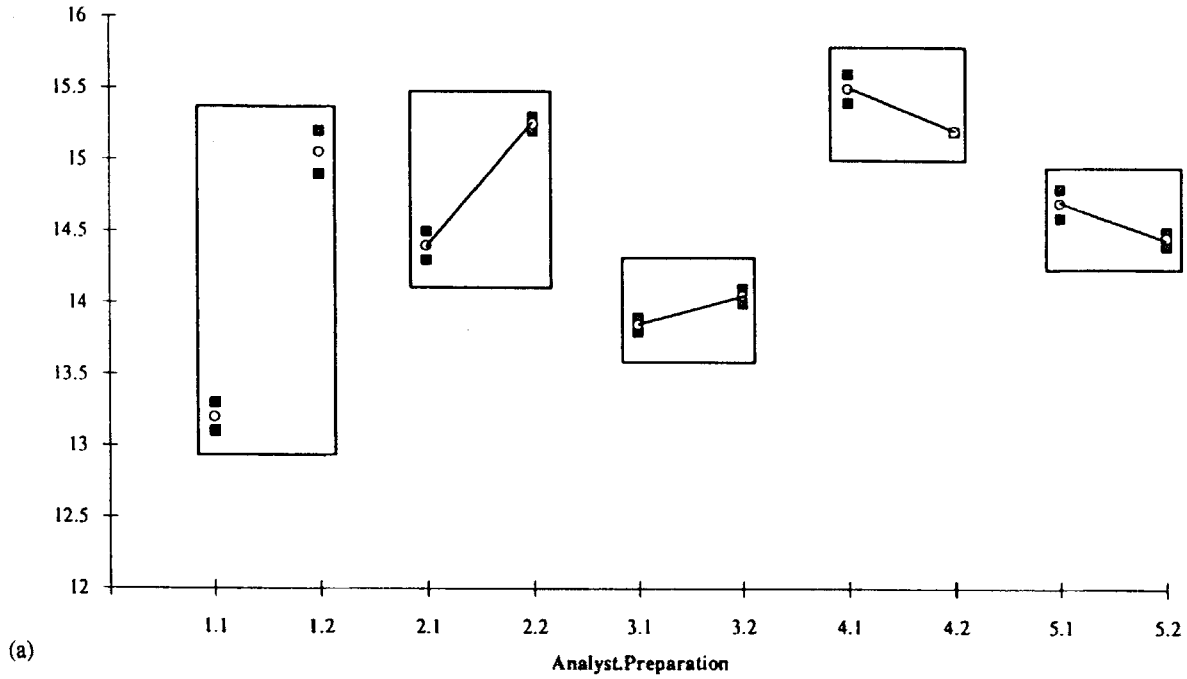
In Eq. (11), $y_{ijk}$ are the observed measurements; $\mu$ is the true value of the common stock of material being assayed in this experiment, $a_i$ is a random component for between analysts variation, $p_{ij}$ is a random component for between-preparation variation, and $w_{ijk}$ is a random component for within-preparation variation. The terms $a_i$ are assumed to behave as if they are a random sample of values from a distribution having zero mean and constant variance $\sigma_a^2$. Likewise, the terms $p_{ij}$ and $w_{ijk}$ are assumed to behave as if they are random samples of values from distributions having zero means and a constant variances $\sigma_p^2$ and $\sigma_w^2$, respectively. All three components are assumed to be statistically independent.

When we contemplate the assay in routine use, with potentially many analysts performing the work and preparing and injecting each sample, all three of these sources of variation will be in operation. From the point of view of the submitter of samples, it is the sum total, $\sigma_a^2 + \sigma_p^2 + \sigma_w^2$, of these three variance components that characterizes the precision of the assay.

## How to Estimate the Variance Components

In this section, the calculations for estimating variance components from a nested scheme are illustrated. The calculations are illustrated using the same design as shown in the preceding subsection. The reader should note that the pattern of calculations being followed may be generalized to two sources of variation and to more than three sources of variation. The generalization will hold as long as the number of branchings at any one level of the design is the same number every time within that level of the design. The number of branchings may differ between levels. Furthermore, the data must be complete. That is, for every branch of the design, there must be a data value. If these conditions are not met, a computer program specifically intended to estimate variance components from a nested design is required.

Before proceeding to calculations, it is usually a good idea to examine the data graphically. The dot frequency plot shown as Fig. 2(a) is often a convenient way to communicate the results of the nested experiment. The




(a)

**Figure 2.** (a) Dot frequency plot. (b) Data in spreadsheet format.

(b)

Spreadsheet for the Computation of Variance Components

| | A | B | C | D | E | F | G | H | J |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | Injections | | | Preparations | |
| 2 | Analyst | Preparation | Injection 1 | Injection 2 | Av | Var | | Av | Var |
| 3 | 1 | 1 | 13.3 | 13.1 | 13.2 | 0.02 | | | |
| 4 | | 2 | 15.2 | 14.9 | 15.05 | 0.045 | | | |
| 5 | | | | | | | | 14.125 | 1.71125 |
| 6 | | | | | | | | | |
| 7 | 2 | 1 | 14.5 | 14.3 | 14.4 | 0.02 | | | |
| 8 | | 2 | 15.2 | 15.3 | 15.25 | 0.005 | | | |
| 9 | | | | | | | | 14.825 | 0.36125 |
| 10 | | | | | | | | | |
| 11 | 3 | 1 | 13.8 | 13.9 | 13.85 | 0.005 | | | |
| 12 | | 2 | 14.1 | 14.0 | 14.05 | 0.005 | | | |
| 13 | | | | | | | | 13.95 | 0.02 |
| 14 | | | | | | | | | |
| 15 | 4 | 1 | 15.4 | 15.6 | 15.5 | 0.02 | | | |
| 16 | | 2 | 15.2 | 15.2 | 15.2 | 0 | | | |
| 17 | | | | | | | | 15.35 | 0.045 |
| 18 | | | | | | | | | |
| 19 | 5 | 1 | 14.6 | 14.8 | 14.7 | 0.02 | | | |
| 20 | | 2 | 14.4 | 14.5 | 14.45 | 0.005 | | | |
| 21 | | | | | | | | 14.575 | 0.03125 |
| 22 | | | | | | | | | |
| 23 | | | | | Av→ | 0.014 | | 14.565 | 0.43375 |
| 24 | | | | | Var→ | | | 0.3139 | |
| 25 | | | | | | | | | |
| 26 | | | | | | | | | |

**Figure 2.** Continued

vertical axis of the plot corresponds to values of the assay response. The horizontal axis is defined by creating all combinations of the variance source levels except the source at the bottom of the hierarchy. Data are enclosed in rectangles to highlight the variation from each of the sources.

We set out the data from this experiment in a spreadsheet format [Fig. 2(b)]. From this format various averages and variances are calculated as intermediates to the final variance component estimates.

Some notes about the spreadsheet format should be kept in mind:

1. The averages in column E are the averages of the two injections from the same preparation. For example, the average in E3 is the average of the values in C3 and D3.
2. The variances in column F are the variances of the two injections from the same preparation. For example, the variance in F3 is the variance of the values in C3 and D3.
3. The averages in column H are the averages of the two injection averages from the same analyst. For example, the average in H5 is the average of the values in E3 and E4.
4. The variances in column J are the variances of the two injection averages from the same analyst. For example, the variance in J5 is the variance of the values in E3 and E4.
5. The averages at the bottom of the table in row 23 are simple arithmetic averages of the numbers in their respective columns. For instance, the value in F23 is the sum of the 10 variances in column F divided by 10.
6. The variance in H24 is the variance of the five numbers in column H (H5, H9, H13, H17, H21).

The variance estimates can now be calculated from the information in the spreadsheet. The variance for injections, $\sigma_I^2$, is estimated by averaging the 10 estimates obtained from each pair of injections. This value is in cell F23, and so $\hat{\sigma}_I^2 = 0.0145$. The variance for preparations is taken from column J, the average variance of the preparation averages. The only complication is that this variance also includes a contribution from the injection variance, which must be subtracted out as follows:

$$\hat{\sigma}_P^2 = 0.43375 - \frac{\hat{\sigma}_I^2}{2} = 0.43375 - \frac{0.0145}{2} = 0.4265$$

Finally, the analyst variance is estimated using the variation among the averages for each analyst. But, again, these averages contain both preparation and injection variation. These contributions must be subtracted in order to obtain an estimate of only analyst variance as follows:

$$\hat{\sigma}_A^2 = 0.31393 - \frac{\hat{\sigma}_P^2}{5} - \frac{\hat{\sigma}_I^2}{5.2}$$

$$= 0.31393 - \frac{0.4265}{5} - \frac{0.0145}{10} = 0.22718$$

The results of the calculations are reported in a format similar to Table 6.

## SAMPLE SIZES IN THE CALIBRATION EXPERIMENT

### Setup of Calibration Experiment

The calibration experiment is designed to describe the relationship between known amounts of an analyte of interest and the response observed from the assay or instrument. There are $n$ different known amounts of analyte symbolized as $x_i$, and their corresponding assay responses (e.g., area under a chromatographic peak) are denoted as $y_i$. The range of values for the $x_i$ is chosen to be one that will cover the range of samples likely to be encountered in practice. The preparations containing the known amounts of analyte are assumed to have been prepared independently of one another. In particular, these samples are not generated by a process of serial dilutions. The simplest case is when the relationship between the $y_i$ and $x_i$ is a straight line. The principles of linear least squares regression can then be applied to help us evaluate the calibration line statistically.

*Table 6*

*Reporting Variance Component Estimates*

| Source of Variation | Degrees of Freedom | Variance Estimate | % of Total Variance | RSD |
|---|---|---|---|---|
| Analyst | 4 | 0.2272 | 34% | 3.3% |
| Preparation | 8 | 0.4265 | 64% | 4.5% |
| Injection | 16 | 0.0145 | 2% | 0.8% |
| Total | 4 | 0.6682 | 100% | 5.6% |

## Relative Error of Estimation for Calibration

The statistical model for a straight-line regression can be written as below. The terms $\varepsilon_i$ arise due to variation in the preparation, instrument, etc., and are assumed to be independent with mean 0 and common variance $\sigma^2$.

$$y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \varepsilon_i, \quad i = 1, 2, \ldots, n \quad (12)$$

The analysis of variance presented as Table 7 summarizes the results of least squares estimation and decomposition of sums of squares.

Very often, the quality of the regression fit for the calibration line is evaluated by examining the value of $R^2$, the multiple correlation coefficient. In terms of the above table, $R^2$ is defined as the ratio of slope sum of squares divided by the Total sum of squares (SSR/SST). That is, $R^2$ is the proportion of total variation accounted for the regression line. However, using $R^2$ as the sole critierion to judge the fit suffers from a similar drawback as that described for the RSD; no account is taken of the sample sizes used. We can always obtain an $R^2$ of exactly 1, by using only $n = 2$ analyte amounts to fit our line.

A better method for assessing the fit is to consider the slope estimate and its statistical properties. The slope is a measure of the sensitivity of the assay to detect small differences in analyte concentration. Since the slope estimate is calculated from the data, it has a variance and a standard deviation (usually called the standard error of the slope). Analogous to RSD($y$), we can form RSD($\hat{\beta}_1$) as follows:

$$\text{RSD}(\hat{\beta}_1) = \frac{\sqrt{s^2/S_{xx}}}{\hat{\beta}_1}$$

Hence the REE of the slope estimate is given by:

$$\text{REE}(\hat{\beta}_1) = \frac{t_{(n-2),.05(2)} \cdot s}{\hat{\beta}_1\sqrt{S_{xx}}} \quad (13)$$

Criteria for judging the quality of the calibration line might be more meaningfully expressed in terms of REE($\hat{\beta}_1$), using values such as 5% or 1% as maximums.

## Planning the Calibration Experiment

We turn our attention now to question 4 of the Introduction: "How many different values, $x_i$, should be used to determine the calibration line?" In answering this question, our interest is in achieving the best fit of the calibration data to the straight line. This is equivalent to saying that the variation surroundnig the line at any given amount of analyte $x_0$, is minimized. Therefore, we need to consider the variance of predicted assay responses, denoted by $\hat{y}_0$. The variance is given by:

$$\text{Var}(\hat{y}) = \sigma^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right) \quad (14)$$

Suppose we designed a calibration range so that the midpoint of the range is of most interest (e.g., 100% of the nominal concentration at the center) and the range of the design is wide enough to make $S_{xx}$ large. Near the point of interest, $x_0$ will be close to $\bar{x}$ and $x_0 - \bar{x}$ will be negligible compared to $S_{xx}$ (let's call this condition the *designed range condition*). Under this reasonably favorable condition, the variance in Eq. (14) reduces to the familiar equation:

$$\text{Var}(\hat{y}_0) = \frac{\sigma^2}{n}$$

Using the regression equation, the predicted value $\hat{y}_0$, near the center of the $x$ range is $\hat{\beta}_0$. Hence the RSD of $\hat{y}_0$ is given by:

$$\text{RSD}(\hat{y}_0) = \frac{\sigma/\sqrt{n}}{\beta_0} = \frac{\sigma}{\sqrt{n}\beta_0}$$

***Table 7***

*Analysis of Variance Table for the Straight-Line Regression*

| Source | *df* | Sums of Squares | Mean Squares |
|---|---|---|---|
| Slope/regression | 1 | $\text{SSR} = k\hat{\beta}_1^2 S_{xx}$ | $s^2_b = \text{SSR}$ |
| Residual/error | $n - 2$ | $\text{SSE} = \Sigma_i(y_i - \hat{y}_i)^2$ | $s^2 = \text{SSE}/(n - 2)$ |
| Total | $n - 1$ | $\text{SST} = \Sigma_i(y_i - \bar{y}_i)^2$ | |

*Note.* $S_{xx} = \Sigma_i(x_i - \bar{x})^2$, $\hat{\beta}_0 = \bar{y}$, $\hat{\beta}_1 = \Sigma_i (x_i - \bar{x})(y_i - \bar{y})/S_{xx}$ and $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 (x_i - \bar{x})$.

This can be estimated by:

$$\hat{RSD}(\hat{y}_0) = \frac{s}{\sqrt{n\bar{y}}}$$

The relative error of estimation is given by:

$$REE(\hat{y}_0) = t_{n-2,.05(2)} \frac{s}{\sqrt{n\bar{y}}} \quad (15)$$

Frequently RSD($y$) is available from a pilot study and REE can also be estimated as:

$$REE(\hat{y}_0) = t_{n-2,.05(2)} \frac{RSD(y)}{\sqrt{n}} \quad (16)$$

This equation is useful for planning in advance how many concentrations, $n$, are needed to make the REE of the resultant prediction near the center of the design range less than a given goal.

Table 8 was created using Eq. (16). If the current RSD of the instrument readings ($y$) is 5%, then we can achieve less than 5% (4.59% to be exact) REE of the predicted readings with 9 or more different concentrations in the calibration experiment and we can claim we are 95% confident that our prediction around the center is within ±5% of the actual average of the predictions at the given concentration.

## Relative Error of Estimation for Reported Concentration

In the actual use of the calibration line, we are interested in the estimation of $x$ (concentration) for given $y$ (instrument reading), that is, inverse prediction of concentration. We observe a new $\bar{y}_0$, the mean of $m$ new observations known to have risen from a single sample with concentration $x_0$, and we wish to predict the concentration and to construct confidence limits and relative error of estimation for this prediction. The point estimate of $x_0$, say $\hat{x}_0$, is the concentration reported to the submitter of the sample and is easy to calculate. Replace $\hat{y}$ by $\bar{y}_0$ in the prediction equation $\hat{y} = \bar{y} + \hat{\beta}_1(x - \bar{x})$ and solve the equation in terms of $x$. Then $\hat{x}_0$ is given by:

$$\hat{x}_0 = \bar{x} + (\bar{y}_0 - \bar{y})/\hat{\beta}_1 \quad (17)$$

Useful approximate confidence limits for $x_0$ are given as (Brownlee, 1965):

$$\hat{x}_0 \pm \frac{t_{n-2,.05(2)} \cdot s}{|\hat{\beta}_1|} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(\hat{x}_0 - \bar{x})^2}{S_{xx}}} \quad (18)$$

This is valid if REE($\hat{\beta}_1$) ($\sqrt{g}$ in Brownlee's work) is less than about 30%, which is easily achieved in most calibration situations.

Under the "designed range condition" described in the previous section, Eq. (18) reduces to:

$$\hat{x}_0 \pm \frac{t_{n-2,.05(2)} \cdot s}{|\hat{\beta}_1|} \sqrt{\frac{1}{m} + \frac{1}{n}}$$

Or equivalently, REE($\hat{x}_0$) is given by:

$$REE(\hat{x}_0) = \frac{t_{n-2,.05(2)} \cdot s}{|\hat{\beta}_1|\hat{x}_0} \sqrt{\frac{1}{m} + \frac{1}{n}} \quad (19)$$

***Table 8***

*Relative Error of Estimation for Prediction*

| Number of concentrations | RSD($y$) 1.00% | 2.00% | 3.00% | 4.00% | 5.00% |
|---|---|---|---|---|---|
| 5 | 1.56% | 3.13% | 4.69% | 6.25% | 7.82% |
| 6 | 1.29% | 2.58% | 3.87% | 5.17% | 6.46% |
| 7 | 1.12% | 2.24% | 3.37% | 4.49% | 5.61% |
| 8 | 1.00% | 2.01% | 3.01% | 4.02% | 5.02% |
| 9 | 0.92% | 1.83% | 2.75% | 3.67% | **4.59%** |
| 10 | 0.85% | 1.70% | 2.55% | 3.40% | 4.25% |
| 11 | 0.79% | 1.59% | 2.38% | 3.18% | 3.97% |
| 12 | 0.75% | 1.50% | 2.25% | 2.99% | 3.74% |
| 13 | 0.71% | 1.42% | 2.13% | 2.84% | 3.55% |
| 14 | 0.68% | 1.35% | 2.03% | 2.71% | 3.38% |
| 15 | 0.65% | 1.30% | 1.94% | 2.59% | 3.24% |

*Note.* $\alpha = 0.05$.

In terms of REE of the slope estimate [see Eq. (14)]

$$\mathrm{REE}(\hat{x}_0) = \mathrm{REE}(\hat{\beta}_1)\sqrt{\frac{1}{m} + \frac{1}{n}}\frac{\sqrt{S_{xx}}}{\hat{x}_0} \qquad (20)$$

The last factor in Eq. (20), $\sqrt{S_{xx}}/\hat{x}_0$, is a scale factor, which depends on the actual experiment. Other than that, the relative error of estimation for the reported concentrations $\hat{x}_0$ depends mostly on the REE of the slope, the number of concentrations, $n$, and the number of new tests, $m$. Table 9 was constructed using Eq. (20) and assuming the scale factor to be 1. To use Table 9 for any particular situation, the values need to be multiplied by the appropriate scale factor for that situation.

## LIMIT OF DETECTION AND LIMIT OF QUANTITATION

### Overview

The limit of detection (LOD) and limit of quantitation (LOQ) of an assay are important performance characteristics that are fundamentally important to the interpretation of assay results. Both characteristics are sometimes lumped under the general heading of *sensitivity* of the assay. In this context, sensitivity refers to the ability of the assay to reliably detect differences as the concentration of analyte approaches zero.

There are two experimental approaches to the determination of LOD and LOQ. The first approach involves repeated assay of a sample with zero analyte. The crucial aspect of this experiment is the ability to create a sample matrix that is identical to the "usual" samples except that none of the analyte of interest is present. In the absence of this ability, one may perform replicate assays at a series of spiked concentrations into the sample matrix and estimate the LOD and LOQ through the use of regression analysis. Both methods are described below.

The calibration experiment itself may be useful in helping to define the LOD and LOQ of the assay. The key condition that must be satisfied is that the matrix in which the analyte of interest is placed for the calibration experiment must be exactly the same matrix as will be present in real samples containing the analyte. If the calibration experiment is performed using a matrix that is much "cleaner" than actual samples, then the calibration experiment will not give a reliable determination of LOD or LOQ. In this situation, the methodology described here can be applied using samples in which the matrix is like actual samples, but with known amounts of the analyte "spiked" into the matrix. This implies that the LOD and LOQ may be a function of the sample matrix as much as of the assay format and concentration of analyte.

### Definition

In our readings of various texts, the scientific literature, and discussions with analytical scientists, the definition of LOD is fairly consistently stated in a manner equivalent to that below, Definition 2. However, a consensus definition of LOQ is not so easily found. Definition 3 takes a purely statistical view of the problem. Other definitions appear to take a more generic view of LOQ as the lowest amount of analyte that can be reliably determined.

In this view, "reliably determined" means that the assay meets both precision and accuracy requirements

*Table 9*

*REE of Reported Concentrations, Assuming REE* $(\hat{\beta}_1) = 5\%$

| Number of Tests ($m$) | Number of Concentrations ($n$) | | | | |
|---|---|---|---|---|---|
| | 4 | 6 | 8 | 10 | 12 |
| 1 | 5.59% | 5.40% | 5.30% | 5.24% | 5.20% |
| 2 | 4.33% | 4.08% | 3.95% | 3.87% | 3.82% |
| 3 | 3.82% | 3.54% | 3.39% | 3.29% | 3.23% |
| 4 | 3.54% | 3.23% | 3.06% | 2.96% | 2.89% |
| 5 | 3.35% | 3.03% | 2.85% | 2.74% | 2.66% |
| 6 | 3.23% | 2.89% | 2.70% | 2.58% | 2.50% |
| 7 | 3.13% | 2.78% | 2.59% | 2.46% | 2.38% |
| 8 | 3.06% | 2.70% | 2.50% | 2.37% | 2.28% |
| 9 | 3.00% | 2.64% | 2.43% | 2.30% | 2.20% |
| 10 | 2.96% | 2.58% | 2.37% | 2.24% | 2.14% |

*Note.* $\mathrm{REE}(\beta_1) = 5\%$; scale factor = 1.

down to concentrations equal to the LOQ; but that at concentrations lower than the LOQ, one or both of these requirements fails. As an example, we quote Hokanson (1994): "The LOQ is the lowest amount of an analyte in a sample that can be quantitatively determined with precision and accuracy under the stated experimental conditions." Using this more generic regime, the LOQ may be assessed by determining percent recovery values for various decreasing amounts of analyte, and by calculating RSD over varying concentrations. Criteria are then established for acceptable recoveries and RSDs. This approach may not lead to the same estimate of LOQ as that given by the statistical definition. In our opinion, all of these definitions should be applied and the most conservative LOQ reported as the limit of quantitation for the method. Regardless of the strategy adopted, we agree emphatically with Rodbard (1978), who states:

> More than one definition and method of calculation has been available—and undoubtedly will continue to be. The methods presented here do not eliminate this ambiguity, and are not intended as a unique, inviolate, and universal definition. However, *it is essential to indicate the exact formulas used and to state all assumptions explicitly.*

The simple use of a term, such as LOD or LOQ, does not convey precise meaning. As scientists, we must pay as close attention to our use of language as we do to our use of reagants and instruments. We use the following definitions:

> **Definition 2: Limit of Detection (LOD)**
> The LOD is that amount of analyte, say $x_{LOD}$ such that the probability is small (usually $\leq .01$) that the corresponding assay response $y_{LOD}$ could have arisen from a sample with zero analyte.

> **Definition 3: Limit of Quantitation (LOQ)**
> The LOQ is that amount of analyte, say $x_{LOQ}$ such that the probability is small (usually $\leq .01$) that the corresponding assay response $y_{LOQ}$ could have arisen from a sample with $x_{LOD}$ analyte.

The definitions imply that for LOD, we must consider the distribution of response values that might arise from samples whose true analyte amount is zero. Likewise, the definition of LOQ implies that we consider the distribution of possible response values arising from samples whose true analyte amount is $x_{LOQ}$. The above definitions of LOD and LOQ were made originally by Hubaux and Vos (1970).

## Statistical Computations

### Strategy 1: Use of Repeat Assays with Zero Analyte

In some instances the analytical scientist will want to estimate the LOD separately from the calibration experiment. It may be that the sample matrix is complex and spiking studies are not easily performed, or the analyte of interest may not be readily available in a pure form. In this situation the LOD may be estimated by performing replicate analyses of samples with zero analyte. Based on the distribution of results from these repeats, the upper limit of responses from truly zero samples may be established.

To calculate the LOD, the average $\bar{y}_0$ and standard deviation, $s_0$, of the $n$ replicates are calculated. The 99% prediction interval for a single predicted assay response value is now given by:

$$y_{LOD} = \bar{y}_0 + t_{n-1,.01(1)} \cdot s_0 \cdot \sqrt{1 + \frac{1}{n}} \tag{21}$$

The LOD is the concentration of analyte $x_{LOD}$ corresponding to an assay response of $y_{LOD}$, as in Eq. (21). Equation (25), given later, may be used for this purpose. One-sided $t$ values with $n - 1$ $df$, $t_{n-1,.01(1)}$ are available from Table 1. This calculation corresponds to the minimal detectable concentration as given by Rodbard (1978).

The LOQ can be determined by solving the following equation for $y_{LOQ}$:

$$y_{LOD} = y_{LOQ} + t_{n-1,.01(1)} \cdot s_0 \cdot \sqrt{1 + \frac{1}{n}} \tag{22}$$

and then using Eq. (25), given later, to find the concentration of analyte $x_{LOQ}$ corresponding to $y_{LOQ}$. If the assay procedure calls for the average of $m$ replicates as the reported assay result, then the 1 terms in both Eqs. (21) and (22) are replaced by $1/m$.

### Strategy 2: Use of a Series of Spiked Samples and Regression Analysis

The spiking experiment is designed to describe the relationship between known spiked amounts of an analyte of interest and the response observed from the assay or instrument. There are $n$ different known spiked amounts of analyte symbolized as $x_i$, and their corresponding assay responses (e.g., area under a chromatographic peak) are denoted as $y_i$. The range of values for the $x_i$ is chosen to be one that is anticipated to be about

twice as large as the LOQ. The preparations containing the known amounts of analyte are assumed to have been prepared independently of one another. In particular, these samples are not generated by a process of serial dilutions. The simplest case is when the relationship between the $y_i$ and $x_i$ is a straight line. The principles of linear least squares regression can then be applied to help us evaluate the change in response with the change in spiked amount of analyte.

The statistical model for a straight line regression can be written as below. The terms $\varepsilon_i$ arise due to variation in the preparation, instrument, etc., and are assumed to be independent, with mean 0 and common variance $\sigma^2$.

$$y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

The analysis of variance presented in Table 10 summarizes the results of least squares estimation and decomposition of sums of squares.

A 99% prediction interval, around the regression line for a single future assay result is given as follows (see, e.g., Brownlee, (1965):

$$\hat{y}_0 \pm t_{n-2,.01(2)}\sqrt{s^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)} \tag{23}$$

For the determination of LOD, we set $x_0 = 0$ and calculate the upper one-sided prediction interval for the response at zero analyte as follows:

$$y_{LOD} = \hat{y}_0 + t_{n-2,.01(1)}\sqrt{s^2\left(1 + \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \tag{24}$$

In Eq. (24) the $t$ value is a one-tailed $t$ value, since we are only interested in values of the response greater than that predicted at zero analyte and $\hat{y}_0 = \hat{\beta}_0$. In practice, many investigators simply replace $t$ by a constant, such as 3 or 5. The actual value for the LOD is found by solving the estimated regression equation for the value of $x$ that predicts $y_{LOD}$:

$$x_{LOD} = (y_{LOD} - \hat{\beta}_0)/\hat{\beta}_1 \tag{25}$$

Equations (24) and (25) apply when the operating procedure for the analytical method is to report a single result for each unknown. If standard procedure is to perform replicate analyses for each sample and report the average of m replicates, then Eq. (24) is modified as follows:

$$y_{LOD} + t_{n-2,.01(2)}\sqrt{s^2\left(\frac{1}{m} + \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \tag{26}$$

The LOD is then determined as above.

The limit of quantitation is found by again starting with Eq. (24). In this case, we want the lower prediction limit for $y$ at $x_{LOQ}$ to correspond to $y_{LOD}$. That is:

$$y_{LOD} = y_{LOQ} - t_{n-2,.01(1)}\sqrt{s^2\left(1 + \frac{1}{n} + \frac{(x_{LOQ} - \bar{x})^2}{S_{xx}}\right)} \tag{27}$$

Because $y_{LOQ}$ also involves $x_{LOQ}$, the solution to Eq. (27) is somewhat messy. The exact solution is found by realizing that Eq. (27) is a quadratic equation in $x_{LOQ}$ and can, therefore, be solved by the usual quadratic equation.

In practice, if the prediction interval does not have much curvature in the region of the LOD, then the value for LOQ can be taken as twice the value of LOD. As with the LOD, if the standard assay procedure is to report the average of $m$ replicates, then Eq. (27) must be modified by replacing the 1 inside the brackets with the term $1/m$. A spreadsheet, shown in Table 11, was prepared utilizing Eqs. (24)–(27) to compute LOD and LOQ using regression analysis. Instead of solving the

***Table 10***

*Analysis of Variance Table for Least Squares*

| Source | *df* | Sums of Squares | Mean Squares |
|---|---|---|---|
| Slope/regression | 1 | $SSR = k\hat{\beta}_1^2 S_{xx}$ | $s_b^2 = SSR$ |
| Residual/error | $n - 2$ | $SSE = \Sigma_i(y_i - \hat{y}_i)^2$ | $s^2 = SSE/(n - 2)$ |
| Total | $n - 1$ | $SST = \Sigma_i(y_i - \bar{y}_i)^2$ | |

*Note.* $S_{xx} = \Sigma_i(x_i - \bar{x})^2$, $\hat{\beta}_0 = \bar{y}$, $\hat{\beta}_1 = \Sigma_i(x_i - \bar{x})(y_i - \bar{y})/S_{xx}$, and $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(x_i - \bar{x})$.

quadratic equation for LOQ, a simple "trial and error" routine using Eq. (27) was coded in the spreadsheet, and the results are presented in Table 11. A copy of a sample EXCEL spreadsheet is available by writing to the authors.

Figure 3 is an illustration of how these calculations work.

## CALIBRATION EXPERIMENT WITH NESTED ERROR STRUCTURE

Sometimes a calibration experiment is performed with replicate determinations of the analyte at each level of concentration. Similar to the experiment presented in the Nested Experiments section, individual determinations in the experiment may not be independent of each other since they are often obtained in a relatively short period of time at each given concentration. An appropriate model is a simple mixed linear model given as:

$$y_{ij} = \beta_0 + \beta_1 (x_i - \bar{x}) + b_i + w_{ij},$$
$$i = 1, \ldots, n; j = 1, \ldots, k \quad (28)$$

where $y_{ij}$ is the $j$th determination of an analyte at concentration $x_i$. The intercept $\beta_0$ and the slope $\beta_1$ need to be determined. The symbol $\bar{x}$ is the average of $x_i$'s. The term $b_i$ is an independent random variable with mean 0 and variance $\sigma_b^2$; $w_{ij}$ is an independent random variable with mean 0 and variance $\sigma_w^2$ and is independent of $b_i$.

*Table 11*

*Spreadsheet Data for Computing LOD and LOQ from Fitted Line*

| Concentration | Peak Area | $\hat{y}$ | SE of Pred. $y$ | UPL |
|---|---|---|---|---|
| 0.022 | 4.634 | 3.6412 | 1.1721 | 3.6711 |
| 0.028 | 4.245 | 4.4797 | 1.1676 | 4.5096 |
| 0.044 | 5.948 | 6.7159 | 1.1566 | 6.7454 |
| 0.044 | 4.786 | 6.7159 | 1.1566 | 6.7454 |
| 0.055 | 9.628 | 8.2532 | 1.1501 | 8.2826 |
| 0.073 | 11.321 | 10.7689 | 1.1412 | 10.7980 |
| 0.088 | 12.865 | 12.8653 | 1.1356 | 12.8943 |
| 0.100 | 14.175 | 14.5424 | 1.1323 | 14.5713 |
| 0.130 | 18.456 | 18.7351 | 1.1286 | 18.7640 |
| 0.150 | 21.369 | 21.5303 | 1.1298 | 21.5592 |
| 0.160 | 24.245 | 22.9279 | 1.1315 | 22.9568 |
| 0.180 | 26.666 | 25.7231 | 1.1372 | 25.7521 |
| 0.260 | 35.186 | 36.9037 | 1.1880 | 36.9341 |
| 0.290 | 40.206 | 41.0965 | 1.2178 | 41.1276 |
| 0.350 | 50.651 | 49.4820 | 1.2930 | 49.5151 |
| $\bar{x}$ | 0.132 | | | |
| $n$ | 15 | | | |
| $S_{xx}$ | 0.143 | | | |
| $\beta_0$ | 0.566 | | | |
| $\beta_1$ | 139.759 | | | |
| $\hat{\sigma}$ | 1.093 | | | |
| | Given $x$ | SE | $\hat{y}$ | |
| Origin | 0.0000 | 1.1909 | 0.5665 | |
| At LOQ | 0.0446 | 1.1562 | 6.8000 | |
| $y_{LOD}$ | | 3.7227 | | |
| LOD | | 0.0226 | | |
| $y_{LOQ}$ | | 6.8000 | | |
| LOQ | | 0.0446 | | |
| Check | | 0.0130 | | |

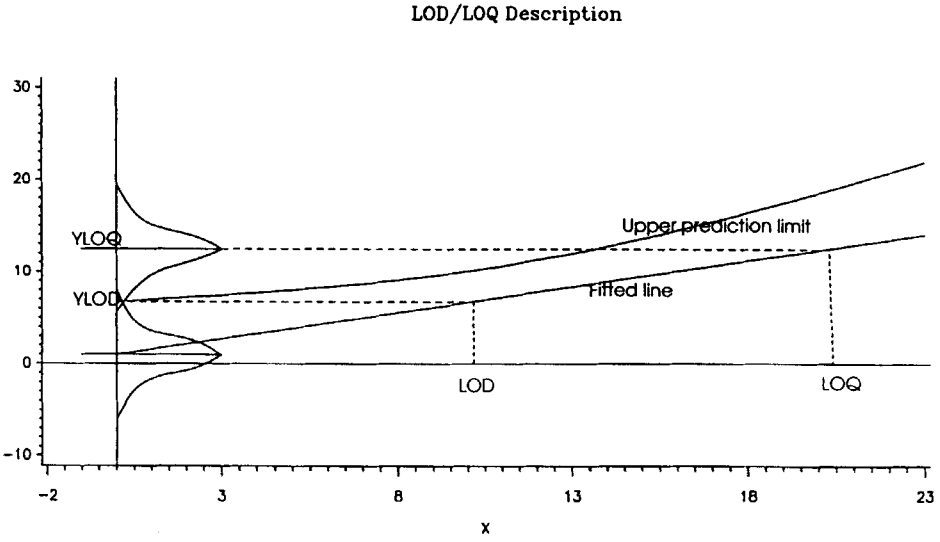



**Figure 3.** LOD/LOQ calculation from prediction intervals.

Lee and Gagnon (1996) showed that if the design is balanced (same number of determinations at each concentration), a practice of using the averages at each concentration (practice B) and performing the usual regression analysis with those $n$ averages is equivalent to the statistical analysis based on the model in Eq. (28) (practice A) except for the estimation of the variance components, $\sigma_b^2$ and $\sigma_w^2$. In other words, the analysis based only on the averages will not only have the same slope and intercept but will also have the correct statistical inference about the slope and the intercept. However, use of averages will not help us to estimate the individual variance components $\sigma_b^2$ and $\sigma_w^2$. As long as interest is primarily focused on fitting and using the calibration line, this is not a drawback.

The confidence interval for $\beta_1$ was also given in the same paper as:

$$\hat{\beta}_1 \pm t_{n-2,\alpha(2)} \frac{s}{\sqrt{S_{xx}}} \tag{29}$$

Equation (29) appears the same as in the section on Sample Sizes, under Relative Error. Note that the *df* are $(n - 2)$, not $(nk - 2)$. The calculating formulas for $\hat{\beta}_1$ and $s$ are given below for completeness:

$$\hat{\beta}_1 = \Sigma_i(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})/S_{xx}$$

$$s = \sqrt{\Sigma_i(\bar{y}_i - \hat{y}_i)^2/(n-2)} \tag{30}$$

where $S_{xx} = \Sigma_i(x_i - \bar{x})^2$; $\bar{y}_i = \Sigma_j y_{ij}/k$; $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(x_i - \bar{x})$; and $\hat{\beta}_0 = \bar{\bar{y}} = \Sigma y_{ij}/nk$.

The same study also showed that a common practice of treating all $nk$ determinations as if they are independent (practice C) should not be done since its confidence interval is erroneously narrow. A simulation study showed that with moderate intraclass correlation of 0.5 or more, the experimental coverage probability of the confidence interval based on practice C is 67% or less for a nominal confidence coefficient of 95%. One might ask: "what then is the utility of getting more determinations at each concentration?" It was shown in the same paper that:

$$\text{Var}(\hat{\beta}_1) = \frac{\left(\sigma_b^2 + \dfrac{\sigma_w^2}{k}\right)}{S_{xx}} \tag{31}$$

Hence, as $k$ increases, the precision of the slope estimate increases. In fact, $s^2$ from averages (practice B) is estimating the numerator of Eq. (31). We also expect the confidence interval is getting tighter as $k$ increases by making the averages $\bar{y}_i$ more precise and hence making $s^2$ smaller.

If the estimates of the individual variance components $\sigma_b^2$ and $\sigma_w^2$ are available, Eq. (31) should be helpful for planning how many determinations for each concentration should be made to achieve a desired precision for

the slope. The desired precision of the slope may be determined by considering the relationship between the precision of the slope and the percision of the final assayed concentration (see equation (4.9) and Table 8).

Other than the requirement that the averages are used instead of individual determinations, the results presented under Relative Error in the section on Sample Sizes apply without change.

## SUMMARY

We stated in the abstract that the goal of writing this paper was to provide the analytical scientist with a clear understanding and practical guidance in the use of statistics. We tried to achieve this by extending concepts behind the popular relative standard deviation (RSD) and hence by showing the proper use of the RSD. We extended the concept of RSD to the relative error of estimation (REE) and maximum relative difference (MRD). The REE is a proper multiple of RSD so that it can be used as the half-width of a confidence interval for the mean assay. We recommend the use of REE as a measure of precision instead of RSD to accompany the final assay results. MRD is literally the maximum possible relative difference of two assay results due to random variation in the assay alone and is also a constant multiple of RSD. MRD is useful whenever one wants to determine if an observed difference between two samples may be caused by something other than the variation of the assay.

We illustrated proper application of these statistics to some common experiments in assay development, including nested experiments and calibration experiments. For a nested experiment, rather than the typical use of a multiway table of RSDs, a better way of summarizing the results is to partition the total variation into its contributing sources. For a calibration experiment, usage of these statistics is illustrated for the estimation of slope, predicted response, inverse estimation of the concentration, and the estimation of LOD and LOQ. We also answered the question of sample sizes in the planning stages of these experiments. We believe this paper will help to promote proper use of more appropriate precision measures for assays typically performed by analytical scientists.

## REFERENCES

Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*, Wiley, New York.

Hokanson, G. C. (1994), A life cycle approach to the validation of analytical methods during pharmaceutical product development, part I: The initial method validation process, Pharm. Technol., 118–130.

Hubaux, A., and Vos G. (1970), Decision and detection limits for linear calibration curves, Anal. Chem., 42(8), 349–355.

Lee, K. R., and Gagnon, R. C. (1996), Calibration and stability analysis with a simple mixed linear model when the experiment is of a split-plot type, Drug Dev. Ind. Pharm., 22(2), 143–151.

Rodbard, D. (1978), Statistical estimation of the minimal detectable concentration ("sensitivity") for radioligand assays, Anal. Biochem., 90, 1–12.