Drug Development and Industrial Pharmacy, 25(1), 37–43 (1999)

RESEARCH PAPER

Estimation of Relative SD

Per Broberg

Astra Draco, S-22100 Lund, Sweden

ABSTRACT

The relative SD is a measure that is now in common use in the pharmaceutical industry, for example, in pharmaceutical development and analytical chemistry, pharmacology, and quality control. This article comments on the usual point estimate of the true value; presents seven interval estimation methods, easily implemented on a desktop computer; and compares their individual merits and drawbacks. Finally, the question of determining an adequate sample size is addressed, and a possible solution is indicated.
Key Words: *RSD; CV; Interval estimation; Bootstrap; Sample size.*

INTRODUCTION

The RSD is a dimensionless measure of variation, commonly defined as the SD over the mean value expressed in percent. The term coefficient of variation is also used but sometimes denotes the decimal version of RSD. These are the definitions used in this article. The definitions tacitly assume that data are positive.

RSD is the inverse of the signal-to-noise ratio and hence might be called the noise-to-signal ratio. The RSD is particularly useful when the SD increases with the mean in a linear fashion. It is also in this situation that we could consider the logarithmic transform to stabilize variances.

In measuring content uniformity, the RSD has been widely used and researchers have used some methods of interval estimation of the parameter (1). This article compares two of these methods with five others.

Unless otherwise stated, we consider independent and normally distributed observations , a sample $x_1, \ldots, x_n$ following a $N(\mu, \sigma^2)$ distribution.

It is worth noting that there are actually two versions of the RSD, one true value that we want to estimate and then the sample RSD, which forms our estimate. To be specific, $s/\bar{x}$ estimates the true population value σ/μ, say. (Here $s^2 = \sum(x_i - \bar{x})^2/(n - 1)$, the usual sample variance.) It is all too common not to distinguish between the two entities.

We repeatedly use maximum likelihood, meaning that we choose the parameter estimate that gives the observations the highest likelihood. We consider one-sided intervals, except in the penultimate section, but the methods are easily modified to produce two-sided ones.

ESTIMATION

Point Estimate

The estimate $s/\bar{x}$ is a strange one that is neither unbiased nor a maximum likelihood estimate (MLE). However, for any reasonable sample sizes it is very close to the MLE,

$$\frac{s}{\bar{x}}\sqrt{\frac{n-1}{n}}$$

and it has good properties like consistency, meaning that the estimate will approach the true value as the sample size increases. If normal variates are at hand, one may argue, however, that the MLE is the estimate of choice.

The rest of this section deals with interval estimation, or more precisely, how to calculate the upper limit of a one-sided 95% confidence interval.

Method M

A suggested method for interval estimation based on the chi-square distribution is

$$\left(0, \frac{1}{\sqrt{\chi^2_{0.05}(n-1)\left(\frac{\bar{x}^2}{(n-1)s^2}+\frac{1}{n}\right)-1}} \times 100\right)$$

This is the method M from Bohidar and Bohidar (1), and the above gives an upper 95% confidence limit for the RSD assuming normal distribution. Note that one has to take the 5% percentile of the chi-square distribution with $n-1$ degrees of freedom. The main reference (1) recommends this method for batch-to-batch release purposes in production and manufacturing because it is claimed to be conservative.

Method B

Method B from the above reference follows from application of Fieller's theorem. It produces the interval

$$(0{,}100 \times (\bar{x}^2 - t^2_{0.05}s^2/n)(\bar{x}s + \sqrt{\bar{x}^2 s^2 - (\bar{x}^2 - t^2_{0.05}s^2/n)(s^2 - t^2_{0.05}s^2/2(n-1))}))$$

where $t_{0.05}$ is 5% percentile of the t distribution with $n-1$ degrees of freedom. The main reference (1) recommends this method for formulation research and development, because it is less conservative than the other preferred method M.

Tangent Method

An outline follows of how to construct a 95% confidence interval for RSD based on the fact that the sample mean and SD are independent for normal deviates (see, e.g., reference 2, p. 271). Calculate one-sided confidence intervals I_1 and I_2 for μ and σ, respectively, both on the level $\sqrt{0.95}$. We use the t interval for the expected value and the chi-square interval for the SD. Then we find the upper limit by taking the lower limit of I_1 and the upper limit of I_2. Thus, we have the 95% interval defined by

$$(0{,}100 \times \max(I_2)/\min(I_1))$$

ML Method

Another method is based on the asymptotics for MLEs. The parameter in the normal distribution are in vector notation $\theta = (\mu, \sigma^2)$. The information matrix is then (3, p. 256)

$$j(\theta) = \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{pmatrix}$$

The population RSD may be written as $g(\theta) = \sigma/\mu = \sigma^2/\mu\ 1/\sqrt{\sigma^2}$. Furthermore,

$$\frac{\partial}{\partial\mu} g(\theta) = \frac{-\sigma}{\mu^2}$$

and

$$\frac{\partial}{\partial\sigma^2} g(\theta) = \frac{1}{2\mu\sigma}$$

Let $\hat{\theta}$ denote the estimate of θ. The asymptotic theory entails that $g(\hat{\theta}) - g(\theta) \approx N(0, g'(\theta)i(\theta)^{-1}g'(\theta)^T))$, where i denotes the expected information matrix, prime differentiation, and T transpose.

We use the information matrix in the estimation of variance. The inverse of that matrix equals

$$j(\theta)^{-1} = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}$$

The variance may then be written as

$$\frac{\sigma^4}{\mu^4 n} + \frac{\sigma^2}{2\mu^2 n}$$

which we estimate by inserting the MLEs $\bar{x}$ and $s_n^2 = (n-1)s^2/n$. To be consistent, one ought to use the MLE also of RSD. Thus, we arrive at the bound

$$\left(0{,}100 \times \left[\frac{s_n}{\bar{x}} + z_{0.95}\sqrt{\frac{s_n^4}{\bar{x}^4 \mathrm{n}} + \frac{s_n^2}{2\bar{x}^2 n}}\right]\right)$$

where $z_{0.95}$ is the 95% percentile of the standard normal distribution (and as is well known equals 1.64).

For normal data it is well known that s_n is biased with the approximate expectation $\sigma(1 - 3/4n)$ (3). This could be used for bias correction, as could the approximation of the expectation of $1/\bar{x}$, which could be written as $1/\mu + \sigma^2/\mu^3$, if the fourth moment is small compared with μ^5, i.e., $3CV^4/\mu$ is small. However, bias correction tends to increase variability and should be used with care.

G Method

Yet another method is given by Eq. (3.15) (4, p. 91), which suggests the variance approximation

$$Var[RSD] = \frac{Var(s)}{\mu^2} + \frac{\sigma^2}{\mu^4} \times \frac{\sigma^2}{n}$$

basically from the Gauss approximation formulae.

The problem then is to give the variance of the sample SD. However, this could to a reasonable degree of accuracy be approximated by $\sigma^2/2(n - 1)$, using a Taylor expansion. And finally

$$Var[RSD] = \frac{\sigma^2}{\mu^2}\left[\frac{1}{2(n-1)} + \frac{\sigma^2}{\mu^2 n}\right]$$

which is very close to the MLE-based method. And we arrive at the interval

$$\left(0{,}100 \times \left\{\frac{s}{\bar{x}} + z_{0.95}\sqrt{\frac{s^2}{\bar{x}^2}\left[\frac{1}{2(n-1)} + \frac{s^2}{\bar{x}^2 n}\right]}\right\}\right)$$

Parametric Bootstrap

Furthermore, we shall consider the parametric bootstrap, which uses resampling to calculate the confidence interval (see, e.g., 5,6). In that method we take the parameter MLE $\hat{\theta} = (\bar{x}, s_n^2)$ and simulate new samples from the distribution $N(\bar{x}, s_n^2)$ and find the 95% percentile of the estimated RSD from those artificial samples. If the underlying distribution is indeed the normal distribution, this method could perform well. The method produces the interval $(0, RSD^*_{0.95})$, where $RSD^*_{0.95}$ is the 95% percentile in the artificial sample.

In the simulation study, 1000 resamples were generated for each of the 500 original samples.

Studentized Bootstrap

We now try to move beyond normal observations. One way to improve on traditional methods has in the last few years been to implement the Studentized bootstrap (6). It is the vague similarity to Student's *t*-test that has given it its name. This methods means resampling

$$t^* = (g(\hat{\theta}^*)) - g(\tilde{\theta})/\sigma\ (g(\hat{\theta}^*))$$

where the asterisk denotes a resampled quantity and tilde a value that comes from the original sample. The trick is to find an estimate of the SD. In this case we are lucky; we just build on the Gauss approximation formula above but now we include a covariance term to recognize the fact that observations may be nonnormal and thus $\bar{x}$ and may be dependent (2,4, p. 92). The formula may now be written as

$$Var\left(\frac{s}{\bar{x}}\right) = \frac{\sigma^2}{2\mu^2(n-1)} - 2Cov(s, \bar{x})\frac{\sigma}{\mu^3} + \frac{\sigma^4}{n\mu^4}$$

but with a linearization $\mathrm{Cov}[s, \bar{x}]\sigma = Cov[s^2, \bar{x}]/2$. And after some tedious calculations, one finds that

$$\mathrm{Cov}[s^2,\bar{x}] = \frac{1}{n(n-1)}\mathrm{Cov}\left\{\sum x_i, \sum x_i^2 - n\bar{x}^2\right\} = \frac{\{E(x^3) + \mu^3 - \mu\sigma^2\}}{n}$$

But because we are mainly concerned with normal deviates, we shall disregard the covariance term and then obtain the same variance estimate as in the G method.

The one-sided confidence limit may now be written as

$$\left(0{,}100 \times \left[\frac{s_n}{\bar{x}} - t^*_{0.05}\sqrt{\frac{s_n^4}{\bar{x}^4 n} + \frac{s_n^2}{2\bar{x}^2 n}}\right]\right)$$

where $t^*_{0.05}$ is the 5% percentile of the bootstrap distribution of t^*.

In the simulation, study 1000 resamples were generated for each of the 500 original samples.

Simulation Results

First let us look at a very peaked normal distribution, where the sample RSD ought to have a rather stable behavior.

Normal deviates with $\mu = 100$, $\sigma = 2$ (i.e., RSD = 2%) were generated, and the upper 95% confidence limit was calculated for various methods. A total of 500

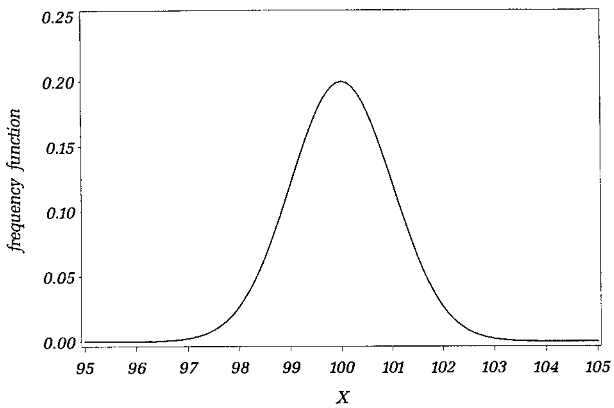

Figure 1. The normal distribution with $\mu = 100$ and $\sigma = 2$.

samples each consisting of 10 observations were generated. In the simulated data, the mean of the 500 RSD values was 1.97 and the upper 95% quantile was 2.74. The RSD based on all observations was 2.02% (Fig. 1, Table 1).

The method M is based on an approximation that is rather sensitive, but it works this time, even if its SD is on the high end. The MLE underestimates the limit slightly at this sample size but comes close and gives a good SD. The G method is very similar to the MLE. Method B rates with the two previous ones. The parametric bootstrap does not do well at all because it gives too short an interval. The Tangent method does not take into account the functional form of RSD and that takes a toll, but it is not so bad after all. The Studentized bootstrap gives good coverage but is a bit volatile.

Now, in the next run the samples consisted of 20 observations each. The mean RSD was 1.99% and the corresponding 95% quantile 2.51%. The RSD based on all 10,000 observations was 2.01%. Table 2 summarizes the results.

Method M is a strong contender with these data, with good coverage and a reasonable STD. MLE is not doing very well but is not a disaster either, with a too low coverage. The same goes for G method. Method B is as good. The parametric bootstrap is a bit overoptimistic, giving

Table 1

Normal Distribution (n = 10)

Method	Mean of 500 Samples (%)	Percentage of Samples Covering True Value	SD of 500 Samples (%)
Method M	3.25	96.6	0.73
MLE	2.81	90	0.63
G	2.74	88.4	0.61
Method B	2.83	90.4	0.63
Parametric bootstrap	2.44	78.8	0.55
Tangent	3.65	98.4	0.83
Studentized bootstrap	3.32	93.8	1.02

Table 2

Normal Distribution (n = 20)

Method	Mean of 500 Samples (%)	Percentage of Samples Covering True Value	SD of 500 Samples (%)
Method M	2.72	95.6	0.44
MLE	2.54	88.8	0.41
G	2.52	88.6	0.41
Method B	2.54	89.4	0.41
Parametric bootstrap	2.38	83.2	0.39
Tangent	2.93	97.8	0.48
Studentized bootstrap	2.73	93.2	0.51

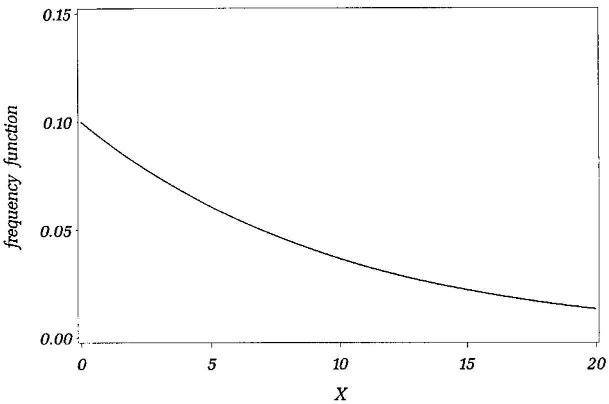

Figure 2. The exponential distribution with expectation 10.

too short an interval. The Tangent method performs rather well this time, whereas the Studentized bootstrap shows the most dramatic improvement as the sample size rises, giving good coverage.

Now, the idea is to look at a more flat distribution, where the sample RSD may be expected to be more volatile and where the sample mean and variance are no longer independent (Fig. 2).

Next, exponentially distributed observations were generated with an expectation of 10, which also equals the SD. The theoretical RSD is then 100%. A total of 500 samples each consisting of 10 observations were generated, and the average sample RSD was 94% and its 95% quantile equaled 134%. The RSD based on all 5000 observations was 102%. Table 3 summarizes how well the methods fared.

The extremely low coverage of method M is due to numerical problems; it tries to calculate the square root of negative numbers on occasion. Despite the fact that it has a low variation, method B is a total disaster: it misses the true value! Again, MLE does rather well. The Tangent method really does badly here. The Studentized bootstrap has good coverage but is a bit volatile.

Then 20 observations were generated for the 100 samples, and the results came out as follows. The average RSD was 96%, and the corresponding 95% percentile 131%. The RSD based on all 2000 observations was 101% (Table 4).

Method M fails completely and gives a missing result for most samples. However, when it does come up with an answer, it covers the true value with a vengeance. The MLE is impressive here. The parametric bootstrap is surprisingly good. Again, the Studentized bootstrap shows a dramatic improvement as we take more observations.

Conclusions

Method M is very strong when it comes to normal data, but otherwise it is highly unreliable. The MLE-based method has a very sound overall performance, even for nonnormal data. Its intervals are sometimes a bit short, however. It is easily programmed in almost any language or even implemented in a spreadsheet program. Furthermore, one can readily adapt the method for the purpose of comparing two samples.

Method B has a problem with undercoverage for normal data and is quite disastrous for nonnormal data. The parametric bootstrap is a bit erratic. The Tangent method is simple, but may be too simple on occasion.

The theory says that the Studentized bootstrap should improve quickly as we increase sample size (7). However, the method gave a rather poor performance at the sample size of 10. If the sample is reasonably large and

Table 3

Exponential Distribution (n = 10)

Method	Mean of 500 Sample (%)	Percentage of Samples Covering True Value	SD of 500 Samples (%)
Method M	220	4	156
MLE	159	91	54
G	155	90	52
Method B	99	66	26
Parametric bootstrap	181	89	97
Tangent	360	96	18,310
Studentized bootstrap	193	96	105

Table 4

Exponential Distribution (n = 20)

Method	Mean of 500 Samples (%)	Percentage of Samples Covering True Value	SD of 500 Samples (%)
Method M	566	8	1324
MLE	141	93	33
G	139	92	33
Method B	118	84	18
Parametric bootstrap	149	92	47
Tangent	269	100	123
Studentized bootstrap	153	93	69

we have reason to believe that data are nonnormal, then the robust Studentized bootstrap is probably the method of choice. Maybe inclusion of a covariance term may improve its performance when we are certain that data are nonnormal. The method can easily be used to compare two samples. When it comes to programming, references 5 and 6 provide examples.

SAMPLE SIZE DETERMINATION

One of the most frequent questions posed to a statistician is ''How big a sample should we take?'' The statistician solves this problem by performing what is called a power calculation. This means that given a difference that is worth detecting, a probability with which we want to be able to detect this difference (the so-called power), at a fixed significance level and some distributional assumptions, the statistician will come up with the magic number. However, this number is sometimes forbiddingly large, and the final sample size is then determined by some process of negotiation.

Let us look at the technical aspect only. Assume that we observe normally distributed variables from two samples: $X_{11}, \ldots, X_{1v}$ and $X_{21}, \ldots, X_{2v}$. Consider the hypothesis $RSD_1 - RSD_2 = \Delta$. From now on we express the difference Δ in decimal notation. Using a modification of a well-known relation for sample size (4, p. 198) and the variance for the MLE, we have

$$n = \frac{\left\{ z_{1-\beta}\sqrt{\left(\frac{\sigma_2}{\mu_2}+\Delta\right)^4 + \left(\frac{\sigma_2}{\mu_2}\right)^4 + \left[\left(\frac{\sigma_2}{\mu_2}+\Delta\right)^2 + \left(\frac{\sigma_2}{\mu_2}\right)^2\right]\frac{1}{2}} + z_{1-\alpha/2}\sqrt{2\left(\frac{\sigma_2}{\mu_2}\right)^4 + \left(\frac{\sigma_2}{\mu_2}\right)^2} \right\}^2}{\Delta^2}$$

where z_a is the a percentile of the standardized normal distribution, α is the significance level of a two-sided test, and $1 - \beta$ is the power (β is the type II error rate) (4).

To take an example, let us consider a case where the significance level should be 5% and we require our test to have 80% power against the alternative $\Delta = 0.02$, given that $\sigma_2/\mu_2 = 0.02$. Since $z_{0.975} = 1.96$ and $z_{0.8} = 0.842$, we see that we need roughly 15 observations per group (we round off 15.6 downward).

A total of 500 simulations were performed where two samples each of size 15 were generated using the normal distributions $N(100,2^2)$ and $N(100,4^2)$, respectively. The Studentized bootstrap and the MLE method were applied, and in 83% of the cases the former method correctly rejected the null hypothesis of equal RSD, whereas the latter did so in 75% of the cases. Both methods come reasonably close to 80%. The Studentized bootstrap achieved a remarkably accurate rejection rate. However, had the sample size been 16, the MLE might have been closest to the target.

Power is not the only possible criterion for choosing sample size. Others are the length of the confidence interval and length of confidence interval given a coverage probability (8).

DISCUSSION

It is worthwhile making the choice of estimation method after some consideration, taking into account the

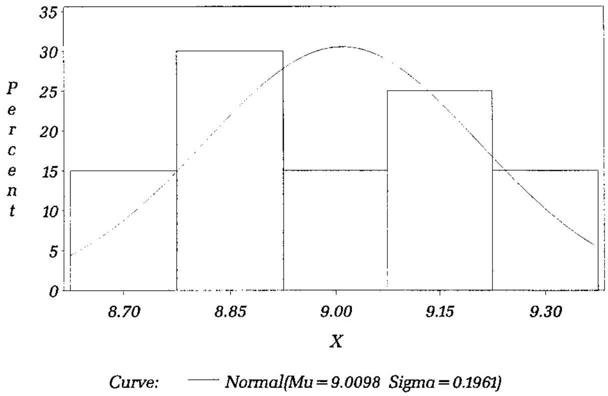

Figure 3. An empirical bimodal and the fitted normal distribution.

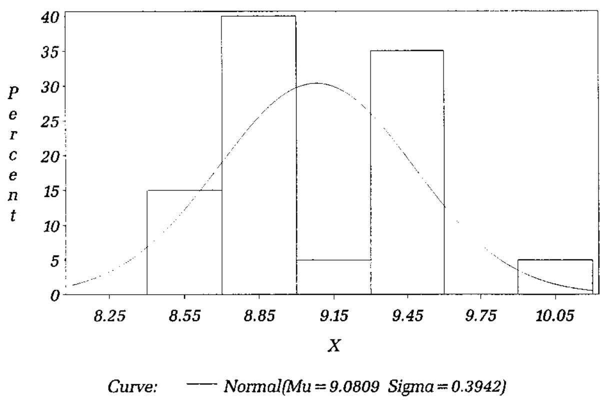

Figure 4. An empirical bimodal and the fitted normal distribution.

objective of the study, the sample size, and the likely distribution of one's data. The MLE method is robust and stands on secure theoretical footing and is easily programed. The Studentized bootstrap is probably one that should please regulatory authorities, because it does achieve the correct confidence level and is very unperturbed by whatever distribution is at hand. It has the drawback of requiring more programing, and it can give a rather poor performance at small samples of normal data compared with methods tailored after that particular distribution. These two methods, in my mind, stand out as most attractive.

It is rather common that the data I come across in my practice are nonnormal. Consider, as an example, the two sets of data presented in Figs. 3 and 4. I had to compare them with respect to RSD, and the natural thing to do was to put my faith in the Studentized bootstrap.

REFERENCES

1. N. R. Bohidar and N. R. Bohidar, Drug Dev. Indust. Pharm., 18(1), 21–37 (1992).
2. P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, 1977.
3. D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman & Hall, London, 1974.
4. P. Armitage and G. Berry, *Statistical Methods in Medical Research*, 3rd ed., Blackwell, Oxford, 1994.
5. B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, London, 1993.
6. J. S. U. Hjort, *Computer Intensive Statistical Methods Validation, Model Selection And Bootstrap*, Chapman & Hall, London, 1994.
7. P. G. Hall, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York, 1995.
8. O. Guilbaud, JSM 97 Proceedings, Biopharm, Anaheim, CA, August 10–14, 1997.