Short Communication

Open Access CrossMark

Correct use of percent coefficient of variation (%CV) formula for log-transformed data

Abstract

The coefficient of variation (CV) is a unit less measure typically used to evaluate the variability of a population relative to its standard deviation and is normally presented as a percentage.1 When considering the percent coefficient of variation (%CV) for log-transformed data, we have discovered the incorrect application of the standard %CV form in obtaining the %CV for log-transformed data. Upon review of various journals, we have noted the formula for the %CV for log-transformed data was not being applied correctly. This communication provides a framework from which the correct mathematical formula for the %CV can be applied to log-transformed data.

Keywords: coefficient of variation, log-transformation, variances, statistical technique

Volume 6 Issue 4 - 2017

Jesse A Canchola, Shaowu Tang, Pari Hemyari, Ellen Paxinos, Ed Marins
Roche Molecular Systems, Inc., USA

Correspondence: Jesse A Canchola, Roche Molecular Systems, Inc., 4300 Hacienda Drive, Pleasanton, CA 94588, USA, Email jesse.canchola@roche.com

Received: October 30, 2017 | **Published:** November 16, 2017

Abbreviations: CV, coefficient of variation; %CV, CV x 100%

Introduction

i. The percent coefficient of variation, %CV, is a unit less measure of variation and can be considered as a "relative standard deviation" since it is defined as the standard deviation divided by the mean multiplied by 100 percent:

$$\%CV = 100\% \cdot \frac{\sigma}{\mu} \quad (1)$$

This formula (1) holds true for non-transformed data. The %CV calculation will be different mathematically depending on the mean and variance of the transformation.

If the untransformed %CV is used on log-normal data, the resulting %CV will be too small and give an overly optimistic, but incorrect, view of the performance of the measured device.

For example, Hatzakis et al.,[1] Table 1, showed an assessment of inter-instrument, inter-operator, inter-day, inter-run, intra-run and total variability of the Aptima HIV-1 Quant Dx in various HIV-1 RNA concentrations. In Table 1, below, we recreate their total SD and %CV columns (the latter for which they use Formula (1), and calculate the correct log-normal %CV from Formula (7) below. From the Table 1, it can be seen that using the incorrect %CV formula for log normally distributed data will give abnormally smaller %CVs.

Table 1 Recreation of portions of Table 5 from Hatzakis et al.,[1] and the correct calculation of lognormal %CV

				Formula (1)	Formula (7)
		Log normal	Log normal	Published incorrect	Correct
Level	N	Mean	Total SD	%CV	%CV
5.00E+01	41	1.66	0.144	8.67	34.1
1.00E+02	74	1.82	0.18	9.91	43.3
1.00E+03	81	2.75	0.112	4.08	26.2
1.00E+04	81	3.81	0.067	1.77	15.5
1.00E+05	81	4.96	0.067	1.35	15.5
1.00E+06	78	6	0.055	0.92	12.7
1.00E+07	81	6.89	0.062	0.9	14.3

To estimate variances of transformations of raw values, we use a statistical technique called the method of moments. Table 2 shows the variances standard deviations and %CVs for the untransformed and log-transformation one may consider.

The formula has been published previously in Nelson.[2] The next section derives the correct percent coefficient of variation formula for the log-transformation in Table 2.

Table 2 Variances, SDS and %CV of log-transformation

Transformation	$\mathbf{Var}\left[\mathbf{f}(\mathbf{x})\right]$	$\mathbf{D}\left[\mathbf{f}(\mathbf{x})\right]$	%CV %CV
None: X	$Var(x)$	$\sqrt{Var(x)}$	$\%CV=100\%\cdot\frac{\sigma}{\mu}$
$Log:\ log_{10}X\ or\ ln(X)$	$\frac{Var(x)}{(\ln(10).E(x))^2}$	$\frac{\sqrt{Var(x)}}{(\ln(10).E(x))}$	$\%CV(Y)=100\%\cdot\sqrt{10^{\ln(10)\sigma_{\log}^2}-1}$

σ =standard deviation; μ =mean; ln(•)=natural logarithm; σ_{log} =standard deviation of the log-transformed data; $E(x)$ is the expected value of x.

%CV for the log-normally distributed random variable (RV)

We show the derivation of the percent coefficient of variation (%CV) for a log-normally distributed random variable. The coefficient of variation for log-normally distributed random variable Y=ln(X) is estimated using the following formula:

$\%CV(Y) = 100\% \cdot \sqrt{e^{[\ln(10)]^2\sigma^2}-1}$ Or its equivalent $log_b(X) = \frac{log_c(X)}{log_c(b)}$

Where *ln* is the natural log and σ^2 is the variance. The derivation of the formulae follows.

Since the random variable X is log-normally distributed, then $Y=\ln(X)$ is distributed as a Normal probability distribution with mean μ and variance λ^2, that is, $Y\sim N(\mu,\lambda^2)$.

Now, the moment generating function for a Normal probability distribution is:[3]

$$M(t) = E(e^{tY}) = e^{\mu t+\frac{\lambda^2 t^2}{2}} \quad (2)$$

Therefore, it follows by substitution:

$$CV(Y)=\frac{SD(Y)}{E(Y)}=\frac{\sqrt{E(e^{2Y})-\left[E(e^Y)\right]^2}}{E(e^Y)}=\frac{\sqrt{M(2)-[M(1)]^2}}{M(1)}=\frac{\sqrt{e^{2\mu+2\lambda^2}-e^{2\mu+\lambda^2}}}{e^{\mu+\frac{\lambda^2}{2}}}=\sqrt{e^{\lambda^2}-1} \quad (3)$$

using the general statistical property that defines the variance as

$$Var(Y) = E\left[(Y-E[Y])^2\right] = E(Y^2)-[E(Y)]^2 \quad (4)$$

such that the standard deviation becomes

$$SD(Y) = \sqrt{E(Y^2)-[E(Y)]^2} \quad (5)$$

To simplify expression (5), above, we use the logarithm base change rule result[4] that shows

$log_b(X) = \frac{log_c(X)}{log_c(b)}$ for any logarithm base *b* and *c*. If *b*=10 and *c*=the "natural log base e"=*e*, then

$$\log_{10}(X)=\frac{\log_e(X)}{\log_e(10)}=\frac{\ln(X)}{\ln(10)}=\frac{Y}{\ln(10)}\sim N(\mu,\sigma^2) \quad (6)$$

since $Y=\ln(X)$ and, given that Y is distributed as a Normal probability distribution with mean μ and variance λ^2, that is, $Y\sim N(\mu,\lambda^2)$, this implies that $\lambda^2=[\ln(10)]^2\sigma^2$ [using the statistical property that $VAR(aX) = a^2 \bullet VarX$ where a is a constant and X is a random variable].

Next, substituting this result into the formula for the %CV involving λ and multiplying by 100% we obtain the final %CV expression:

$$\%CV(Y)=100\%\cdot\sqrt{e^{[\ln(10)]^2\sigma^2}-1}=100\%\cdot\sqrt{10^{\ln(10)\sigma^2}-1} \quad (7)$$

Conclusion

The authors have shown that it is easy for the researcher to be confused with respect to which is the correct formula to use for log-transformed data when calculating the percent coefficient of variation (%CV). When using the incorrect formula, the researcher may be faced with abnormally low %CV values. With that in mind, the authors have shown the correct formula to use for calculating %CV for log-transformed data.

Acknowledgements

The authors thank Enrique Marino, Merlin Njoya and Jeff Vaks for reviewing the earlier work and providing useful comments. This work is supported by Roche Molecular Systems, Inc.

Conflict of interest

The author declares no conflict of interest.

References

1. Hatzakis A, Papchristou H, Nair SJ, et al. Analytical characteristics and comparative evaluation of Aptima HIV-1 Quant Dx assay with Ampliprep/COBAS TaqMan HIV-1 test v2.0. *Virol J*. 2016;13(1):176.
2. Nelson W. *Applied Life Data Analysis*. USA: John Wiley & Sons Inc; 2003.
3. Walpole RE, Myers RH, Myers SL, et al. *Probability & Statistics for Engineers & Scientists*. 7th ed. USA: Prentice Hall; 2002. p. 186–190.
4. http://www.purplemath.com/modules/logrules5.htm

Citation: Canchola JA, Tang S, Hemyari P, et al. Correct use of percent coefficient of variation (%CV) formula for log-transformed data. *MOJ Proteomics Bioinform*. 2017;6(4):316–317. DOI: 10.15406/mojpb.2017.06.00200