

Statistics Notes

Percentage differences, symmetry and natural logarithms

Tim J Cole, Douglas G Altman

**Population, Policy and Practice Programme
Great Ormond Street Institute of Child Health**

University College London

30 Guilford Street

London WC1N 1EH

Tim J Cole, *professor*

Centre for Statistics in Medicine

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences

University of Oxford

Oxford OX3 7LD

Douglas G Altman, *professor*

Correspondence to: T J Cole

tim.cole@ucl.ac.uk

Despite its wide use in statistics, the logarithmic transformation can make non-statisticians uncomfortable.¹⁻⁴ This is a shame as logarithms have very useful properties, including a secret not widely known even among statisticians.

The two familiar forms of logarithm are common logs, to base 10, and natural logs, to base e .¹ Here we focus on natural logs (or “ln” for short), which have the following “natural” interpretation: if we take two numbers a and b , then the difference between their logs, $\ln(a) - \ln(b)$, is the fractional difference between a and b . And multiplied by 100, i.e. $100 \ln(a) - 100 \ln(b)$, it is the percentage difference between a and b .⁵

This value is slightly different from the conventional percentage difference, but it avoids the problems of asymmetry and non-additivity described earlier.⁶ We showed in the previous note that on average women are 7.7% shorter than men, yet men are 8.4% taller than women, based on mean heights of 177.3 cm and 163.6 cm respectively.⁶ The log-based percentage difference between them is

$$100 \ln(177.3) - 100 \ln(163.6) = 8.04\%$$

so men are 8.04% taller than women, and women are 8.04% shorter than men. Swapping the two changes the direction of the percentage difference but not its value – unlike the conventional percentage difference it is a *symmetric* percentage difference.

A previous Statistics Note⁴ gave data on biceps skinfold thickness in patients with Crohn's disease or coeliac disease. The group means of the log transformed data were 1.44 and 1.14 [log units](#), a difference of 0.30 (95% CI [-0.11](#) to 0.71). This difference, multiplied by 100, can be viewed as a symmetric mean percentage difference of 30% (95% CI [-11%](#) to 71%). So on average, there is a 30% difference in biceps skinfold between the Crohn's and coeliac patients. This is a simple and convenient way to summarise log transformed data that avoids the need for antilogs.¹

The standard deviation of natural log transformed data is also problematic - what does it mean? By the same token it is a fractional standard deviation, similar to the standard deviation divided by the mean, i.e. a form of coefficient of variation. Multiplying by 100 converts the log standard deviation to a coefficient of variation in percentage units.

In the biceps skinfold example, the standard deviations of the log values in the two groups were 0.49 and 0.52 respectively, equivalent to "coefficients of variation" of 49% and 52%. For comparison the conventional coefficients of variation were rather larger, 51% and 56% respectively. This difference indicates that the data were closer to a log normal than a normal distribution.⁴

Regression analyses with log transformed outcomes are easier to report using the 100 ln transformation. Hyppönen et al.⁷ used the transformation to demonstrate nonlinearity in the relationship between serum IgE concentration and vitamin D status. They reported that compared to the reference group with 25(OH)D between [100-100](#) and [1225](#) nmol/l, and adjusted for 11 other factors, IgE was 29% higher (95% CI 9% to 48%) for participants with 25(OH)D <25 nmol/l, and 56% higher (95% CI 17% to 95%) for those with 25(OH)D >135 nmol/l.

As another example, neonatologists measure weight gain in units of g per kg per day. This fractional growth rate, thousandths per day, can be estimated from the regression coefficient of 1000 ln(weight) on age. In the same way that 100 ln(height) measures height in symmetric percentage units, 1000 ln(weight) measures weight in symmetric thousandths of units.⁵

To summarise, the 100 ln transformation leads to comparisons on a modified percentage scale that is both symmetric and additive. Differences between group means are in percentage units; standard deviations are coefficients of variation in percentage units, and regression coefficients are also in percentage units. Reference⁵ has further examples. The approach is generally useful for exploring relationships with positive valued continuous outcome data, as are common in biochemistry and anthropometry.

References

1. Bland JM, Altman DG. Logarithms. *BMJ* 1996;312:700.

2. Bland JM, Altman DG. Transforming data. *BMJ* 1996;312:770.
3. Bland JM, Altman DG. Transformations, means, and confidence intervals. *BMJ* 1996;312:1079.
4. Bland JM, Altman DG. The use of transformation when comparing two means. *BMJ* 1996;312:1153.
5. Cole TJ. Sympercents: symmetric percentage differences on the 100 log_e scale simplify the presentation of log transformed data. *Stat Med* 2000;19:3109-25.
6. Cole TJ, Altman DG. What is a percentage difference? *BMJ* 2017.
7. Hyppönen E, Berry DJ, Wjst M, Power C. Serum 25-hydroxyvitamin D and IgE - a significant but nonlinear relationship. *Allergy* 2009;64:613-20.