



RESEARCH METHODS & REPORTING

Statistics Notes: Percentage differences, symmetry, and natural logarithms

Tim J Cole *professor*¹, Douglas G Altman *professor*²

¹Population, Policy and Practice Programme, Great Ormond Street Institute of Child Health, University College London, London WC1N 1EH, UK;

²Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK

Despite its wide use in statistics, the logarithmic transformation can make non-statisticians uncomfortable.¹⁻⁴ This is a shame because 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—that is, $100 \times \ln(a) - 100 \times \ln(b)$ —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 in the accompanying Statistics Note.⁶ We show there 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 for men and 163.6 cm for women.⁶ The log based percentage difference between them is

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

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 Note gave data on biceps skinfold thickness in patients with Crohn's disease or coeliac disease.⁴ The group means of the log transformed data were 1.44 and 1.14 log units, a difference of 0.30 (95% confidence interval -0.11 to 0.71). This difference, multiplied by 100, can be viewed as a symmetric mean percentage difference of 30% (-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—that is, 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 $100 \times \ln$ transformation. Hyppönen et al used the transformation to demonstrate non-linearity in the relationship between serum IgE concentration and vitamin D status.⁷ They reported that, compared with the reference group with 25-hydroxyvitamin D (25(OH)D) blood level of 100–125 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 (17% to 95%) for those with 25(OH)D $\geq$ 135 nmol/L.

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

To summarise, the $100 \times \ln$ 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. Cole⁵ has further examples. The approach is generally useful for exploring relationships with positive valued continuous outcome data, as are common in biochemistry and anthropometry.

- 1 Bland JM, Altman DG. Statistics Notes: Logarithms. *BMJ* 1996;358:700. doi:10.1136/bmj.312.7032.700 pmid:8597743.
- 2 Bland JM, Altman DG. Statistics Notes: Transforming data. *BMJ* 1996;358:770. doi:10.1136/bmj.312.7033.770 pmid:8605469.
- 3 Bland JM, Altman DG. Statistics Notes: Transformations, means, and confidence intervals. *BMJ* 1996;358:1079. doi:10.1136/bmj.312.7038.1079 pmid:8616417.
- 4 Bland JM, Altman DG. Statistics Notes: The use of transformation when comparing two means. *BMJ* 1996;358:1153. doi:10.1136/bmj.312.7039.1153 pmid:8620137.
- 5 Cole TJ. Sympercents: symmetric percentage differences on the 100 log_e scale simplify the presentation of log transformed data. *Stat Med* 2000;358:3109-25. doi:10.1002/1097-0258(20001130)19:22<3109::AID-SIM558>3.0.CO;2-F pmid:11113946.
- 6 Cole TJ, Altman DG. Statistics Notes: What is a percentage difference? *BMJ* 2017;358:j3663. doi:10.1136/bmj.j3663.
- 7 Hyppönen E, Berry DJ, Wjst M, Power C. Serum 25-hydroxyvitamin D and IgE - a significant but nonlinear relationship. *Allergy* 2009;358:613-20. doi:10.1111/j.1398-9995.2008.01865.x pmid:19154546.

Published by the BMJ Publishing Group Limited. For permission to use (where not already granted under a licence) please go to <http://group.bmj.com/group/rights-licensing/permissions>