

Why self-report 'Likert' scale data should not be log-transformed.

Nearly 10 years ago, Nevill (1997) wrote an editorial expelling the virtues of taking logarithmic transformations when analysing performance variables recorded on the ratio scale. Most performance variables reported in sports journals such as the Journal of Sports Sciences (JSS) are invariably recorded on the ratio scale, e.g. maximum oxygen uptake ($\text{VO}_2 \text{ max}$), power (W) and strength (N). All such performance measurements have an important characteristic in common, they all have a natural zero point i.e., they can never take negative values. It is because such variables are bounded (by zero) that their frequency distributions have a tendency to be positively skewed, a characteristic that can be naturally overcome by taking logarithms. A further advantage of taking logarithms includes the overcoming of heteroscedastic errors (i.e., the error increases in proportion to the size of the variable) when regressing a ratio variable as the response variable using other ratio variable(s) as the predictor variable(s). Recently, however, a number of authors submitting articles to JSS using self-report 'likert' scale data have also advocated the use of logarithmic transformations to overcome skewness and non-normality observed in their data.

Sport and exercise psychologists typically use self-report data with assessments being made on likert type scales. The participant is asked to rate a feeling, or estimate an attitude on an arbitrary scale on which numbers are designed to relate to the strength of feelings. For example, participants report feelings on the Brunel Mood Scale (BRUMS: Terry, Lane, Lane, & Keohane, 1999) from 'not at all' (0) to 'very much so' (4). To illustrate the arbitrary relationship between the number and the statement it describes, the item 'Nervous' is on the BRUMS and the Positive and Negative Affect Schedule (PANAS: Watson, Clark, & Tellegen, 1988). On the PANAS, participants rate Nervous from 'not at all' to 'very much so' on a 1 to 5 scale. A participant who does not feel nervous could report a value of either 0 or 1 to the item Nervous depending on the scale being administered. Clearly, most of these self-report 'likert' scales are recorded on an interval rather than a ratio scale. As most authors are well aware, interval data has no fixed "zero" point, i.e. the scale could go, for example, from -2 to +2, from 0 to 4 or from 1 to 5 (as described above). All three scales will result in identical conclusions when analysed using parametric or non-parametric statistical tests. However, if the authors try to take logs of the scale -2 to +2, or 0 to 4, they will not get the same conclusions (compared with 1 to 5). Indeed the authors will have great difficulty in log-transforming zero or negative values on their measurement scales. It is because likert-type data, being interval and not ratio scale data, that the log transformation is totally inappropriate (only appropriate for true ratio scale data).

A second issue with interval data is whether the values ascribed to an attribute accurately reflect the numerical difference. Arguably, the scale is theoretically a continuous construct and values between data points are equal. Evidence indicates that this is not always the case. Lane and Terry (2000) observed that the difference between reporting not at all (0) and 'A little' (1) to depressed mood items in the BRUMS (Terry et al., 1999) is much larger than the difference between reporting a value of somewhat (2), moderately (3) or very much so (4). Lane and Terry have given considerable scrutiny to how participants report depressed mood items on the BRUMS (see Lane, 2004). Caution must be placed on whether Likert scale estimates represent true differences between values.

Based on the above observations, we suggest that researchers adopting, analysing and reporting self-report Likert style data, should take great caution when analysing their data using parametric methods. Adopting non-parametric methods are more likely to accommodate the rank style differences between discrete values of a Likert scale, i.e., Likert scale data should be treated as "interval" scale with great caution but certainly, not as ratio scale data, when parametric methods such as logarithmic transformations would be totally inappropriate.

Alan Nevill and Andrew Lane
University of Wolverhampton

- Lane, A. M. (2004). Measures of emotions and coping in sport. *In Coping and Emotion in Sport*. Pp255-271. Editors Lavalley, D., Thatcher, J., & Jones, M. Nova Science, NY.
- Lane, A. M., & Terry, P. C. (2000). The nature of mood: Development of a conceptual model with a focus on depression. *Journal of Applied Sport Psychology*, **12**, 16-33.
- Nevill, A.M. (1997) "Why the analysis of performance variables recorded on a ratio scale will invariably benefit from a log transformation". [Editorial] *Journal of Sports Sciences*, **15**, 457-458.
- Terry, P. C., Lane, A. M., Lane, H. J., & Keohane, L. (1999). Development and validation of a mood measure for adolescents. *Journal of Sports Sciences*, **17**, 861-872.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, **54**, 1063-1070.