Comparison of two methods for regression analysis for log-normal exposure data

Sara Gustavsson, *Sahlgrenska Academy, Gothenburg, Sweden*
Sandra Johannesson, *Sahlgrenska Academy, Gothenburg, Sweden*
Gerd Sallsten, *Sahlgrenska Academy, Gothenburg, Sweden*
Eva M Andersson, *Sahlgrenska Academy, Gothenburg, Sweden*

Background and Aims: Log-normally distributed variables are common in exposure data. One characteristic of the log-normal distribution is that the variance increases with the mean. Often it is of interest to analyze which background variables (e.g. time spent in traffic) that are associated with the personal exposure (e.g. particles). Often we want to estimate the absolute effect (of e.g. time spent in traffic on exposure) but by log-transforming data we instead estimate the relative effect.
We propose a new regression method for log-normal data, using a maximum likelihood approach (ML) by which we can account for the heteroscedasticity. The aim of this study was to compare this new ML-approach with ordinary Least-Squares regression (LS).
Methods: A large simulation study was performed, using a model in which expected personal exposure (MuY) is a linear function of the outdoor particle levels (x): MuY=4.804+0.574·*x* In the model the variance (in personal exposure) increases with the exposure level. The bias of the estimated coefficients (b0 and b1) was assessed by comparing the estimated values to those in the model.
Results: Both ML and LS showed small bias for both b0 and b1 ($< 2 \cdot 10^{-4}$). However, the standard errors were larger for LS (about 54% larger for SE(b0) and 10% for SE(b1)). As a result the power of LS will be lower, especially regarding test of the intercept. The power for detecting a small effect was estimated by simulating the situation where beta1=0.10 (rather than 0.574). We found the power to be 77% (LS) and 80% (ML).
Conclusions: Both ML and LS give unbiased estimates of the regression parameter, but the larger standard errors for LS-regression means that the confidence intervals will be wider and it will be harder to detect small effects. Therefore ML-regression seems more suitable for estimating linear relationships in log-normally distributed data.