BSE 5643, Regression Analysis

Interpreting regression coefficients in a linear model whose outcome variable (y) has been log transformed

When constructing linear models, a continuous outcome y is sometimes log transformed (ln y) to ensure the tenability of assumptions that underlie the model. Such a linear model predicts or estimates means $\overline{y_k^*}$ of a set of log transformed outcomes:

$$\overline{y_k^*} = \frac{1}{n}\sum_{i=1}^{n}(\ln y_i)$$

The mean of log-transformed observations is related to the original observations' geometric mean.

To describe the model's inference on the mean of y in y's original (untransformed) units, apply the appropriate inverse transformation (exponentiation in this case) to the estimate of $\overline{y_k^*}$:

$$\exp \overline{y_k^*} = \exp\{\frac{1}{n}\sum_{i=1}^{n}(\ln y_i)\}$$

$$= \exp\{\frac{1}{n}(\ln y_1 + \ln y_2 + \ldots + \ln y_n)\}$$

$$= \exp\{\frac{1}{n}(\ln(y_1 * y_2 * \ldots * y_n)\}$$

$$= \exp\{\frac{1}{n}(\ln \prod_{i=1}^{n} y_i)\}$$

$= (\prod_{i=1}^{n} y_i)^{\frac{1}{n}}$, which is the geometric mean of the original (untransformed) values for y.

A log-transformed difference in means is related to a ratio of geometric means in the original (untransformed) units of measurement.

A regression coefficient generally estimates the difference in the mean outcome that is associated with a one unit difference in a covariate: $\hat{\beta} = E(\bar{y}_1 - \bar{y}_2)$. When regression coefficients are estimated using log transformed outcomes, we can interpret them in terms of y's original (untransformed) units by applying the inverse transformation:

$$\exp(\hat{\beta}) = \exp(\overline{y_1^*} - \overline{y_2^*}) = \frac{\exp \overline{y_1^*}}{\exp \overline{y_2^*}}$$

We know from the earlier algebra demonstration that this result represents a ratio of geometric means.

Therefore, regression coefficients that are obtained from linear models performed on log transformed outcomes are interpretable, once they are exponentiated, as ratios of geometric means.

Bland and Altman (1996) point out that the log transformation uniquely yields this concrete interpretation when the appropriate inverse transformation is applied to estimates from regression models. For this reason, the log transformation is more popular than, for example, the square root or reciprocal transformations.

Bland JM, Altman DG. Statistics notes: The use of transformation when comparing two means. BMJ 1996; 312:1153. Obtained February 16, 2012 from http://www.bmj.com/content/312/7039/1153.full

David M. Thompson

October 10, 2012