Rightly Transforming Right-Skewed Data

Linh To

Abstract

Data transformations often facilitate regression analysis, yet many commonly used transformations make hypothesis testing misleading because the results depend on the measurement units of the data. This paper establishes a five-way equivalence that fully characterizes the set of transformations for which the conclusions are independent of measurement units. Central to these results are the concepts of scale equivariance (that any scaling of the data can be reversed by appropriately rescaling the estimator) and scale invariance (that any scaling of the data does not affect the estimator). The equivalence result demonstrates that desirable properties such as scale-equivariant coefficient estimates, scale-invariant t-statistics, and scale-invariant semi-elasticities arise if and only if the transformation is a logarithmic or a power function. Power transformations thus provide a natural extension of logarithmic transformations that both preserves the essential feature of obtaining unit-independent estimates for unitless quantities of interest and is defined at zero. Popular alternatives that approximate the shape of the logarithmic function at large values, such as adding a small positive constant before applying a logarithmic transformation or the inverse hyperbolic sine transformation, can result in arbitrarily large semi-elasticity estimates and can change sign and statistical significance depending on the choice of measurement units, which we highlight both theoretically and empirically.