SRSM 2019 Abstract

Credibility, Prediction and Tolerance Intervals

Michael T. Brannick, University of South Florida

Hannah Rothstein, Baruch University

In meta-analysis, researchers typically estimate a summary statistic (typically an overall mean) as well as a between-studies or random-effects variance. It is increasingly common to place a bracket intended to capture a portion of the underlying distribution (credibility interval, tolerance interval) or a random observation from the distribution (prediction interval). In this talk, we will describe each of the three kinds of intervals and what they attempt to achieve. The credibility interval can be considered to a description of a certain portion of the population distribution, and the prediction and tolerance intervals attempt to capture either a random observation (prediction) or fraction of observations (tolerance) from the underlying population while accounting for error of estimation due to finite numbers of effect sizes (finite *k*). The conventional prediction and tolerance intervals have difficulties with small sample sizes (small *N*) and unequal sample sizes because they are based on the assumption of a random sample from a normal distribution where the sample variance represents the best estimate of the population variance. In a meta-analysis, the between studies variance used to estimate the population variance is obtained essentially by subtracting the within studies variance from the observed sample variance.

The purpose of the study was to compare different interval estimators in terms of how well they achieve their stated aims. Monte Carlo techniques were used to generate observed effect sizes from known underlying distributions. The observed distributions varied in the number of studies (k) and the average sample sizes (N) as well. We also varied the underlying mean and standard deviation of random effects. Conventional and bootstrap estimators of the tolerance intervals were calculated as were credibility and prediction intervals. Coverage rates were calculated across trials for each estimator. For small numbers of studies with small sample sizes, none of the estimators provide coverage near the nominal (95%) rate. With large sample sizes (large N, small k), however, the conventional tolerance and prediction intervals provided coverage near the nominal rate. The bootstrap tolerance interval provided near nominal coverage if the number of effect sizes was at least 30 even if the sample sizes were small.

Researchers can use the results of the study to help inform the choice of estimator. We also consider single-sided tolerance intervals and confidence intervals for quantiles of the underlying distribution in our discussion.