

Knowing (*ESS*) and Not Knowing (*D*)

Paul R. Yarnold, Ph.D.

Optimal Data Analysis, LLC

Taken together the *ESS* and *D* statistics elucidate a mystery that began with the discovery of statistical analysis.

Until recently it was common knowledge that: "...no statistical measure based on the data table will help to decide how much of the complete explanation of (the class variable) we have found. ... This means that part of the story told by the variables omitted from the table may be, in fact, much more interesting and significant than the part we have found, but there is no way of telling whether this is so."¹ However, this is no longer apropos because of two new test statistics: *ESS* and *D*.

ESS is a chance-corrected (0 = the level of predictive accuracy expected by chance) and maximum-corrected (100 = perfect, errorless prediction) index of predictive accuracy: a measure of the magnitude of prediction and control that is possible over a phenomenon on the basis of the information available for the application.²

It is intuitively obvious that the "error" in prediction—an index of magnitude of failure to predict and control the phenomenon—may be computed as $100 - ESS$. However, novometric theory² shows that a myopic, unidimensional consideration of overall accuracy is incorrect. Rather, it is also crucial to consider the role of parsimony in evaluating the efficacy of statistical models. The *D* statistic integrates the concepts of accuracy and parsimony: *D* indexes the number of equivalent effects that are needed to obtain an ideal model yielding perfect accuracy

with maximum possible parsimony for the application.²

In the optimal data analysis ("maximum-accuracy") statistical paradigm the *ESS* statistic is conceptualized as a normed measure of how much is known in a given application, and *D* as a normed measure of how much more remains to be discovered in order to reap a theoretically ideal model for the application.

References

¹Gilbert N (1993). *Analyzing tabular data: Loglinear and logistic models for social researchers*. London, England: UCL Press (pp. 107-108).

²Yarnold PR, Soltysik RC (In Review). *Maximizing predictive accuracy*. Chicago, IL: ODA Books.

Author Notes

The study analyzed de-individuated data and was exempt from Institutional Review Board review. No conflict of interest was reported.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., #163
Chicago, IL 60646
USA