

“A Statistical Guide for the Ethically Perplexed” (Chapter 4, Panter & Sterba, *Handbook of Ethics in Quantitative Methodology*, Routledge, 2011): Clarifying Disorientation Regarding the Etiology and Meaning of the Term *Optimal* as Used in the Optimal Data Analysis (ODA) Paradigm

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

The authors of Chapter 4 complain that use of the word “optimal” in the name Optimal Data Analysis (ODA) is unethical because it implies that alternative data analysis methods are less than optimal. The present *Errata* note addresses this misunderstanding.

Derived from the field of operations research, every statistical model developed in the ODA paradigm explicitly maximizes classification accuracy for the training sample.¹⁻⁴ Accuracy is the objective function which is maximized by ODA models, rather than, for example, variance or likelihood.^{5,6} Thus, “optimal” means that a model achieves its theoretically maximum possible level of accuracy. This is traditional nomenclature (“jargon”) in operations research.

What is meant by accuracy? Imagine a training sample with observations from class 0 or class 1. The statistical model classifies each

observation in the sample into one of these two class categories. If the classified category and the actual category for an observation are the same, then a point is scored. If the classified category and actual category are different, then no point is scored. For every sample the ODA model explicitly maximizes the number of points obtained. Observations may also be weighted, and the optimal—that is, most accurate—weighted solution is identified.

For every unique sample, variable geometry, and weighting scheme there exists an optimal solution. The optimal solution is the

most accurate solution that is theoretically possible to obtain for the sample. If any statistical model *explicitly* obtains the most accurate possible solution (optimal solution) for a given problem, then the model is said to be an optimal model. If a methodology explicitly obtains the theoretically most accurate possible solution, then the methodology is said to be optimal—by definition. It is the meaning of the word in this context. All methods which fail to explicitly prove optimality (maximum accuracy) are called suboptimal, by definition.

To be explicitly optimal, a method must be specifically formulated to find the most accurate solution. Finding an optimal model is an operations research problem requiring methods such as mathematical programming, linear algebra, integer programming, and so forth.⁷ Every different sample is a separate, unique optimization problem. For every statistical application the name of the most accurate possible model is the optimal solution, and there's only one way to explicitly achieve the optimal solution—by using operations research methods, including specifically-engineered software like UniODA¹ and CTA².

Remarkably it happens that it is possible to drape operations research methodologies in an exact statistical paradigm. The name of that paradigm is optimal data analysis.^{1,2}

References

¹Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC: APA Books.

²Yarnold PR, Soltysik RC (2010). Optimal data analysis: A general statistical analysis paradigm. *Optimal Data Analysis*, 1, 10-22.

³Soltysik RC, Yarnold PR (2010). Automated CTA software: Fundamental concepts and control commands. *Optimal Data Analysis*, 1, 144-160.

⁴Yarnold PR (2013). Initial use of hierarchically optimal classification tree analysis in medical research. *Optimal Data Analysis*, 2, 7-18.

⁵Grimm LG, Yarnold PR (Eds.). *Reading and Understanding Multivariate Statistics*. Washington, D.C.: APA Books, 1995.

⁶Grimm LG, Yarnold PR (Eds.). *Reading and Understanding More Multivariate Statistics*. Washington, D.C.: APA Books, 2000.

⁷Soltysik RC, Yarnold PR (2010). Two-group MultiODA: Mixed-integer linear programming solution with bounded *M*. *Optimal Data Analysis*, 1, 31-37.

Author Notes

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

eMail: Journal@OptimalDataAnalysis.com