11-1-2003

Letters To The Editor

JMASM Editors

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Editors, JMASM (2003) "Letters To The Editor," *Journal of Modern Applied Statistical Methods*: Vol. 2 : Iss. 2 , Article 27.

DOI: 10.22237/jmasm/1067646420

Available at: <http://digitalcommons.wayne.edu/jmasm/vol2/iss2/27>

This Letter to the Editor is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Letters To The Editor

Additional Reflections On Significance Testing

Knapp, T.R. (2002). Some reflections on significance testing. *Journal of Modern Applied Statistical Methods*, 1(2), 240-242.

Knapp (2002) raised good points concerning significance testing. The bottom line, “If you have hypotheses to test ... [then] test them”, not only makes sense, but in fact is an argument that I have often made when consulting, although the closest I have come to this issue in the literature is an allusion (Berger, 2000, Section 2.1).

Yet, the support for this assertion, based on refuting the statement that confidence intervals provide identical or non-conflicting inferences with significance tests, might benefit from elaboration. In fact, the confidence interval constructed by Knapp (2002) is but one of several that could have been constructed. To argue that it is not the best among these is to argue that its discredit cannot serve as a simultaneous discredit to the class it purports to represent (albeit not very well).

It would be a simple matter to construct a confidence region as precisely the set of parameter values which, when serving as the null hypothesis, lead to significance tests that cannot be rejected. With this definition of a confidence set (which often, but not always, reduces to a confidence interval), it is a tautology that the confidence set cannot contradict the results of the significance test.

Why, then, should hypotheses be tested? Because it is problematic to base policy decisions, that affect the public, on apparent directions of effect when the study is conducted by a party with a vested interest in the outcome. Requiring statistical significance is one reasonable way to operationalize the need for a preponderance of evidence, and raise the hurdle, in such a case. If an alpha level were chosen strategically, perhaps based on safety, convenience, and cost in a medical study, then the results of the significance test of efficacy would correspond to the optimal decision. Clearly, there are other ways to raise the hurdle.

Vance W. Berger, Biometry Research Group,
National Cancer Institute. E-mail:
vb78c@nih.gov.

Reference

Berger, V.W. (2000). Pros and cons of permutation tests in clinical trials. *Statistics in Medicine* 19, 1319-1328.

Predictor Importance In Multiple Regression

Whittaker, T.A., Fouladi, R.T., & Williams, N.J. (2002). Determining predictor importance in multiple regression under varied correlational and distributional conditions. *Journal of Modern Applied Statistical Methods*, 1(2), 354-366.

William Kruskal may have been right in noting that the relative importance of predictors in a regression analysis is meaningful to researchers, but I'm not so sure it should always be the case.

My principal concerns about the Whittaker et al. (2002) article are these:

1. Multiple regression analysis is used for prediction and for causal analysis. When a user asks: “What are the most important variables in this regression?”, the answer depends upon the purpose of the analysis. Whittaker et al. failed to distinguish sufficiently between the two purposes and seem to advocate a “one size fits all” method for determining the relative importance of regressors (apparently Budescu's dominance analysis, perhaps augmented by the Johnson index).

2. I do not see the need for the Monte Carlo approach to the problem. In his text, Darlington (1990) provided a mathematical explanation for the equivalence with respect to rank-ordering of importance of their Methods 3 (the t statistics for the betas), 5 (the squared partials) and 6 (the squared semi-partial), along with two others (the p-values for the betas and the changes in R-square from the reduced model with the variable deleted to the full model with the variable

included). [They do cite Darlington (1968), but do not cite his later text.] On page 364 of their article they speculated as to why Methods 3, 5, and 6 “all performed identically”. Those three methods must perform identically.

As far as the other five methods are concerned, Method 1 (squared zero-order correlations, or unsquared zero-order correlations, for that matter) can be dismissed out of hand, because the other regressors are not statistically controlled.

I can't see any reason why anyone would ever use Method 2 (the betas). For one thing, the betas aren't restricted to the -1 to +1 range, so although they are standardized they are awkward to compare.

Method 4 (the beta-times-r products) has been criticized in the past (see, for example, Darlington, 1968). The fact that those products sum to the over-all R-square is a poor basis for variance partitioning and for the determination of relative importance (some of those products can even be negative--for suppressor variables).

That leaves Methods 7 (Budescu) and 8 (Johnson). I have no doubt that similar, non-Monte Carlo-based, arguments could be made regarding those methods for determining the relative importance of regressors, but even if such arguments were necessary, Whittaker et al. (2002) were interested in comparing all eight methods, not just those two.

3. Two different meanings of the word “dominance” was confusing. One of the meanings, “dominance analysis”, is associated with Budescu's method. The other meaning, identifying the “dominant predictor” (p. 358), was apparently the criterion for determining which methods were best. When I first read the article I thought that the Budescu method was

used as one of the methods AND as the goodness criterion, which would of course “stack the deck” in its favor. The confusion with the two meanings, however, remains.

I have a couple of other lesser concerns:

1. Their definition of “the dominant predictor” is a bit strange. In what sense is an independent variable that correlates .40 - .60 with the dependent variable dominant over other independent variables that correlate .30 with the dependent variable?

2. Their “Nursing Facility Consumer Satisfaction Survey” example is a poor example. The data are for seven-point Likert-type scales with ridiculously high means and there is an inherent regressor/regressand contamination since the three predictors are concerned with specific satisfactions and the dependent variable is over-all satisfaction.

Thomas R. Knapp, Professor Emeritus, University of Rochester & The Ohio State University.

Note: I would like to thank Richard Darlington for his very helpful suggestions regarding an earlier version of this critique.

References

Darlington, R.B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69(3), 161-182.

Darlington, R.B. (1990). *Regression and linear models*. New York: McGraw-Hill.