



June 1975

Tom Swift and his electric regression analysis machine: 1973

J. Scott Armstrong

University of Pennsylvania, armstrong@wharton.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/marketing_papers

Recommended Citation

Armstrong, J. S. (1975). Tom Swift and his electric regression analysis machine: 1973. Retrieved from http://repository.upenn.edu/marketing_papers/37

Postprint version. Published in *Psychological Reports*, Volume 36, Number 3, June 1975, page 806. The author has asserted his/her right to include this material in *ScholarlyCommons@Penn*.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/marketing_papers/37
For more information, please contact libraryrepository@pobox.upenn.edu.

Tom Swift and his electric regression analysis machine: 1973

Abstract

Presents the "1973 Tom Swift Award for Data Abuse."

Comments

Postprint version. Published in *Psychological Reports*, Volume 36, Number 3, June 1975, page 806. The author has asserted his/her right to include this material in *ScholarlyCommons@Penn*.

Published in *Psychological Reports*, 36, 1975, 806.

Tom Swift and His Electric Regression Analysis Machine: 1973

J. Scott Armstrong

Tom Swift, who began his career with factor analysis (1967), is pleased to announce that the "1973 Tom Swift Award for Data Abuse" has been won by LeRoy Stone and James Brosseau. They originally (Stone, *et al.*, 1973) used 115 variables in a stepwise regression analysis to explain differences among 19 observations. They then claimed (Stone & Brosseau, 1973) to have tested the predictive validity of this model. This was done by regressing the 14 variables from the model on data from 18 new subjects. This "cross-validation" yielded a final model with six variables and an R^2 of 0.76.

They went beyond the call of duty by collecting all of these data since comparable results could have been obtained with random data (Ando & Kaufman, 1966; Armstrong, 1970). This may be shown also as follows. Assuming that there was *no* relationship, the calculated R^2 from 18 observations with 14 variables can be obtained from (Montgomery & Morrison, 1973):

$$R^2 = k/n$$

where k is the number of independent variables and n is the number of observations. Thus, expected R^2 would be 0.78 if all 14 variables were included, and close to this if only the best six variables were included. The R^2 obtained by Stone and Brosseau (1973) was, of course, less than this at 0.76. The evidence, then, did *not* support their conclusion that the model would "... quite accurately predict success of trainees ...".

References

- Ando, A. & Kaufman, G. M. (1966), "Evaluation of an ad hoc procedure for estimating parameters of some linear models," *Review of Economics and Statistics*, 48, 334-340.
- Armstrong, J. S. (1967), "The derivation of theory by means of factor analysis or Tom Swift and electric factor analysis machine," *American Statistician*, 21, 17-21.
- Armstrong, J. S. (1970), "How to avoid exploratory research," *Journal of Advertising Research*, 10, 27-30.
- Montgomery, D. B. & Morrison, D. G. (1973), "A note on adjusting R^2 ," *Journal of Finance*, 28, 1009-1013.
- Stone, L. A., Bassett, G. R., Brosseau, J. D., Demers, J., & Stiening, J. A. (1973), "Psychological test characteristics associated with training-success in a Medex (physician's extension) training program," *Psychological Reports*, 32, 231-234.
- Stone, L. & Brosseau, J. D. (1973), "Cross-validation of a system for predicting training success of Medex trainees," *Psychological Reports*, 33, 917-918.