Examination of a variable ordering index in linear regression models: An assessment of the relative Pratt index in Likert data.

By
Charles O. Ochieng & Bruno D. Zumbo

Department of Educational Counseling Psychology and Special Education, University of British Columbia,
Vancouver BC,
V6T 1Z4

Paper presentation at the First Annual Bob Conry conference on Measurement Evaluation and Research and Methodology. The University of British Columbia, Coach House Green College February 9th 2001

For a detailed copy of the paper and further enquiries please e-mail Charles Ochieng at cochieng@interchange.ubc.ca

Outline of the study

- Background to the problem
- Description of relative Pratt index
- Rationale of the study
- Research question
- Method
- Results
- Conclusion
- Implications for research

Background to the problem

- Often, researchers are concerned by significance and relative importance of variables in linear regression models for meaningful interpretation of results. Several indices have been proposed in determining relative importance of predictor variables in regression models among which are relative Pratt index, Beta coefficients β, *t*-values, commonality components, semi-partial correlation sr_j and partial correlation pr_j. Selection and ordering of variables in a regression model are therefore dependent on the relative importance of the variable compared to the others.

  Relative importance of variable X_j in a regression equation is determined by the proportion of the variance in the criterion variable Y accounted for by X_j (Kruskall, 1987; Pratt, 1987; Bring, 1994, 1996; Thomas, Hughes, & Zumbo, 1998) where X_j is a predictor variable in the regression model. Pedhazur (1982) and Darlington (1990) suggest that beta weights and partial correlations should not be used in determining relative importance due to luck of proportionality and additive properties in relation to the variance in the criterion variable, instead semi-partial correlations may be used.
- Because of its additive property and simplicity in interpretation, the relative Pratt index has an advantage over the semi-partial correlation and lends itself as the most promising and prudent index to use in determining relative importance and ordering of variables (Huberty, 1994; Thomas & Zumbo, 1996; Thomas, Hughes, & Zumbo, 1998).

Description of the relative Pratt index

- Relative Pratt index is a geometric extension based on the axiomatic derivation in the 2-variable case (Pratt, 1987) of the product of the simple correlation and beta coefficient of a variable X_j from the standardized regression equation of the form:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_j X_j + \beta_p X_p.$$

- The relative Pratt index d_j (Thomas, Hughes & Zumbo, 1998) is then computed as follows;

$$d_j = \beta_j r_j / R^2.$$

The resulting quotient d_j is the proportion of variance R^2 in the criterion variable accounted for by the predictor variable X_j. Thus, the Pratt index is used in the variance partitioning of R^2 for each predictor. For a regression model the sum of the relative Pratt index of the predictors add up to one.
- Relative Pratt index can be used together with beta weights to identify any suppressor variables that may be present in a regression equation (see Thomas, Hughes, & Zumbo, 1998). This is derived from negative valued Pratt index. Only relative Pratt index of the value $d_j > 1/2p$ where p is the number of predictors in the regression equation, have meaningful relative importance and is usually interpreted.

Rationale of the study

- While relative Pratt index has been endorsed as an appropriate measure of relative importance of predictor variables in linear regression models, little is known of the effect of Likert data on its accuracy and performance under different conditions of Likert scales and response distribution.

Research question

The basic assumption in linear regression models is that predictors and criterion variable s are continuous and normally distributed. This may not be the case in real data encountered in practice in social science research. Research questions are as follows;

- How stable are Pratt indices under non-continuous and non-normal conditions?
- What are the effects of (1) distribution of response patterns, (2) correlation matrices, (3) variable combination of continuous and Likert data and (4) number of Likert scale points, on the accuracy and consistency of relative Pratt index?
- To answer the research questions, the following method was used.

Method

- Simulated data was used in the study. Because the study sort to determine the accuracy (bias) of relative Pratt index under the stated conditions, a population of 500 000 was generated using three correlation matrices depicting low moderate and high inter-item correlation among predictors and criterion for each condition (Stevens, 1986). Three types of response distributions were simulated namely, (1) equal interval symmetric distribution (2) unequal interval and threshold, positively skewed distribution (3) unequal interval and threshold, negatively skewed distribution.
- For variable combination of continuous and Likert data, three types of data condition of X and Y were simulated namely, (1) Y Likert, and X continuous (2) Y continuous and X Likert (3) Y Likert and X Likert. For each of the three conditions, eight Likert scale categories ranging from 2-scale points to 9-scale points were generated. The study was a 3x3x3x8 factorial design.

- A linear regression model was then fit for each condition using three predictors X_1, X_2, and X_3. The resulting R^2 and relative Pratt indices associated with each predictor were computed and compared to those of the baseline condition with Y continuous and X continuous, and equal interval as well as normal distribution.

Bias in Pratt index was computed as follows;

$Pratt\ Bias\Delta = Pratt_{model} - Pratt_{cont}$

Percent bias was also computed as follows:

$Percent\ bias = Pratt\ bias / Pratt_{cont} \times 100$.

Percent bias was used as a unit of Analysis.

<u>Results</u>

1. Graphs were plotted for the percent bias against the number of Likert scale point categories for each variable combination and type of response pattern within each condition of correlation matrix.
2. To provide further insight on the effects of the four independent variables on percent bias of relative Pratt index, a response surface model was fit using the response surface methodology (Box & Draper 1987;Khuri & Cornell, 1987; Zumbo & Harwell, 1999)

<u>Findings</u>

- From the response surface model, within low and moderate matrices there were significant main effects for category at variable combination 2 (Y continuous, X Likert) and significant main effects for response pattern at variable combination 3 (Y Likert and X Likert). However, for high correlation matrix there were significant main effects for categories in the three variable combinations. While there were significant differences in percent bias between categories of Likert scales, the order of the variables remained unchanged across the independent variables. Percent bias diminished with increase in scale points from 2-point Likert scale and leveled off asymptotically at 4-point Likert scale.

<u>Conclusion</u>

- Relative Pratt index remains relatively robust in terms of variable ordering of relative importance under the stated conditions of types correlation matrix, type of response pattern distribution, number of Likert scale points and variable combination of continuous and Likert data. Percent bias was relatively large at the 2-point scale in all the three response patterns for equal and unequal thresholds and intervals.
- While bias occurred at each Likert scale point, relative ordering of the variables was not affected. Thus, Likert scaling does not impact the ordering of variables and therefore relative importance, as measured by relative Pratt index.
- Bias in relative Pratt index reduced with increase in Likert scale points, but remained almost constant after the 4-point Likert scale.

Implication for research

- Researchers may opt for relative Pratt index due to its additive properties in variance partitioning, ease of interpretation, low bias in 4-point Likert scales or higher, and consistency across Likert scales that are frequently encountered in social science research.
- Due to its additive property in the variance partition of the criterion variable, relative Pratt index may now replace or supplement other measures of relative importance such as communality analysis measures that are cumbersome in terms of meaningful interpretation. This is clearly evident with more predictors in the model. Relative Pratt index may also be used together with standardized regression weights to identify any suppressor variables and multicollinearity cases that may be present in the regression model. This is inferred from negative values that may occur in the computation of Pratt index. The suppressor variables can then be analyzed separately.

References:

Box, G.E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York: John Wiley & Sons, Inc..

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician*, *48*, 209-213.

Bring, J. (1996). A geometric approach to compare variables in a regresion model. *The American Statistician*, *50*, 57-62.

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

Khuri, A. I., Cornell, J.A. (1987). *Response surfaces: Designs and analyses*. New York: Marcel Dekker, Inc..

Kruskal, W. (1987). Relative importance by averaging over orderings. *American Statistician*, *41*, 6-10.

Pedhazur, E. J. (1982). *Multiple Regression in Behavioural Research* (2nd ed.), New York: Holt, Reinhart & Winston.

Pratt, J.W. (1987). Dividing the indivisible: using simple symmetry to partition variance explained," In T. Pukkila an S. Puntanen (Eds.), *Proceedings of the Second International Conference in Statistics*, (pp. 245-260). Tampere, Findland: University of Tampere.

Stevens,J. (1986). Applied multivariate statistics for the social sciences. Hillsdale, New Jersey.Lawrence Erlbaum Associates, Publishers.

Thomas, D. R., Hughes, E., & Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, *45*, 253-275.

Thomas, D. R., & Zumbo, B. D. (1996). Using a measure of variable importance to investigate the standardization of discriminant coefficients. *Journal of Educational & Behavioral Statistics, 21*, 110-130.

Zumbo, B. D., & Harwell, M. R. (1999). *The methodology of methodological research: Analyzing the results of simulation experiments*. (Paper No. ESQBS-99-2). Prince George, B. C.: University of Northern British Columbia. Edgeworth Laboratory for Quantitative Behavioral Science.