Advances in Social Sciences Research Journal – Vol.3, No.7
Publication Date: July. 25, 2016
DoI:10.14738/assrj.37.2125.
Grice, J. W., Cota, L. D., Barrett, P. T., Wuensch, K. L. & Poteat, G M. (2016). A Simple and Transparent Alternative to Logistic Regression. *Advances in Social Sciences Research Journal,* 3*(7)* 147-165.

A Simple and Transparent Alternative to Logistic Regression

James W. Grice
Oklahoma State University

Lisa D. Cota
Oklahoma State University

Paul T. Barrett
Cognadev UK Ltd.

Karl L. Wuensch
East Carolina University

G. Michael Poteat
East Carolina University

Abstract

Observation oriented modeling was compared to logistic regression in the re-analysis of data from a published study. In the original study students' decisions to stop or continue controversial animal research was predicted from gender, the type of research being conducted, and measures of ethical attitudes. With regard to classification accuracy, results from the observation oriented modeling analyses compared favorably to results obtained from logistic regression and two other methods of classification. Prediction profiles created from the observation oriented modeling analyses were moreover parsimonious and transparent, and the analyses themselves were free of assumptions. Unique features of the observation oriented modeling approach were discussed.

Keywords: Observation Oriented Modeling, Logistic Regression, Prediction

A SIMPLE AND TRANSPARENT ALTERNATIVE TO LOGISTIC REGRESSION

Observation oriented modeling is a novel way to both conceptualize and analyze data within the social and life sciences. While it is a relatively new methodology, its flexibility and promise can be seen in the diverse topics of study to which it has been applied: for example, human cognition (Valentine & Buchanan, 2013), language comprehension (Gatobu et al., 2016) terror management theory (Grice et al., 2012), spirituality and religiosity (Anderson & Grice, 2014), honeybee and horse behavior (Craig et al., 2015; Dinges et al., 2013), marital conflict (Brown & Grice, 2012), and assessment in higher education (Comer, et al., 2014). In observation oriented modeling researchers are encouraged to create an integrated model that elucidates the causal structure of a natural system and then to evaluate the accuracy of the model at the level of the persons or entities in the study. On a technical level observation oriented modeling is most similar to nonparametric methods in that analyses are largely free of the routine assumptions of normality, equal population variances, homoscedasticity, etc. that accompany parametric techniques. Graphical aids also play a central role in the analyses as data are treated largely "as is" without making unwarranted assumptions about their structure, particularly the assumption of continuity (e.g., treating a 9-point rating scale as continuous; see Michell, 2011).

Analyses conducted within the framework of observation oriented modeling are consequently simple, transparent and intuitive.

A situation often confronted by social and life scientists is the prediction of a dichotomous variable from other dichotomous or non-dichotomous variables. Wuensch and Poteat (1998), for example, were interested in predicting students' decisions to stop or continue institutional support for research being conducted on cats. The fictitious research was described in a vignette and it required the cats to be beheaded after being subjected to brain surgery. The students were also randomly assigned to five different groups which differed with regard to the type of research being conducted: cosmetic, theoretical, meat production (via a growth hormone), veterinary, and medical. Lastly, students completed the Ethics Position Questionnaire (EPQ) as a measure of idealistic and relativistic attitudes. The EPQ is comprised of twenty 9-point Likert-type items that are averaged to form an idealism and a relativism subscale.

Given the dichotomous decision outcome (stop vs. continue), Wuensch and Poteat used logistic regression to analyze their data. Specifically, they regressed the dichotomous decision variable onto the gender, group (dummy coded), idealism, and relativism variables. Multiple R^2 was statistically significant [$\chi^2(7, N = 315) = 87.51, p < .001$] for the model, and specific results revealed statistically significant regression weights (p's < .05) for the gender, idealism, and relativism variables and for two of the dummy coded group variables (theory and meat). In the context of the complete regression model, women, individuals high in idealism, and individuals low in relativism were found to be more likely to stop the research than their respective counterparts. Individuals in the meat and theory groups were also found to be more likely to stop the research than individuals in the medical group (the reference group for the dummy coding). Interactions among the predictor variables were not explored.

Approaching Wuensch and Poteat's data[1] from the perspective of observation oriented modeling, the general goal is to maximize the correct classification of students as either stopping or continuing the research based on their gender, group membership, and EPQ scale scores. In logistic regression the goal is to maximize multiple R^2 and the absolute values of the regression weights while minimizing p-values. Of course the classification accuracy of a logistic regression model can also be assessed by first computing a probability of group membership (stop or continue) for each person on the dichotomous outcome variable:

$$p_{group} = \frac{1}{1 + e^{-(\beta_0 + \beta_x)}}$$

where β_0 and β_1 equal the maximum likelihood estimates from regressing the outcome variable onto a single predictor. For Wuensch and Poteat's analysis multiple regression weights were included in the formula, and if the probability value was equal to or greater than .50 for a particular student, then that person was assigned to the second group (continue the research); otherwise, the student was assigned to the first group (stop the research). These assigned values were then compared to the actual decisions, and the number of correct matches tallied, converted to a percentage and reported as classification accuracy. Overall, 71.75% of the 315 students were classified correctly using the regression-based combination of gender, group membership, and EPQ scales. Although employing different methods, this process of classification is considered as primary in observation oriented modeling.

If no interactions are expected, the analysis begins with the examination of individual predictor variables. In observation oriented modeling variables are referred to as "orderings" to emphasize the importance of the original units of observation and how those units might be organized into patterns. For these data the dichotomous decision is designated as the "target" ordering, and the predictor variables are designated as the "conforming" orderings. Beginning with gender, which is also dichotomous, a binary Procrustes rotation algorithm (see Grice, 2011) is used to rotate this ordering to conformity with the decision target ordering in order to classify the observations.[2] The results are reported in the multigram in Figure 1, and as can be seen a majority (n = 140, 70.00%) of the 200 women chose to stop the research, whereas a majority (n = 68, 59.13%) of the 115 men chose to continue the research. It can also be seen in the multigram that the students were correctly (grey bars) or incorrectly (white bars) classified based on their gender. The total percentage of students correctly classified is referred to as the Percent Correct Classification (PCC) index, and it is here equal to 66.03% (208 / 315 * 100). For dichotomous variables, the PCC index will typically equal the classification accuracy from a logistic regression analysis with one predictor.

An optional randomization test (Edgington & Onghena, 2007; Manly, 1997; Winch & Campbell, 1976) can be used to aid the evaluation of the PCC value. The test is conducted by randomly pairing the target and conforming observations across the participants, performing the binary Procrustes rotation, and then computing the PCC index. This process is repeated 1000 times (as determined by the researcher) and the PCC indices recorded. The number of PCC values from the randomized data that equal or exceed the observed PCC index are tallied and converted to a proportion. This frequency probability is reported as a chance-value, or c-value, and was zero for the PCC index of 66.03% (c-value < .001, 1000 randomized trials). The chance-value is here reported as "c-value < .001" rather than "c-value = 0" because with additional randomized trials a result of at least 66.03% will eventually be observed.

The group predictor variable is analyzed in similar fashion. Recall the participants were randomly assigned to five conditions differentiated by the reason for the controversial feline research. As can be seen in Figure 2, majorities of students in the cosmetic, theory, meat, and veterinary groups voted to stop the research (range = 59% to 63%), whereas a slight majority of the participants in the medical group (54%) decided to continue the research (PCC = 60.95%, c-value = .06). The structure of the group ordering was therefore simplified by combining the veterinary, meat, theory, and cosmetic conditions. The dichotomous group ordering was then rotated to conformity with the decision ordering. The PCC index was again 60.95% (c-value = .01), and a majority (63%) of the participants in the combined non-medical condition voted to stop the researcher, whereas a slight majority (54%) of participants in the medical condition voted to continue the research.

The gender and group orderings are strictly categorical whereas the remaining idealism and relativism orderings are assumed to represent interval scales, at least in the context of the logistic regression analysis.[3] In observation oriented modeling this assumption is not necessary, but the numbers assigned to these orderings are considered to represent at least simple orders. Consequently, a Threshold Analysis in the Observation Oriented Modeling (OOM; Grice, 2016) software can be used to efficiently identify cut-points (or thresholds) for the idealism and relativism orderings that maximize their respective PCC indices when conformed to the dichotomous decision ordering. Identifying such thresholds will simplify the interpretation of results and will also facilitate the combining of orderings (variables) in subsequent analyses.

Beginning with the idealism ordering, the threshold analysis works systematically from the lowest to highest observed scores. A dichotomous ordering is first created by defining the lowest value as one unit (viz., 1.70) and the remaining values as the second unit (viz., >1.70 to 9.00). This newly formed dichotomous ordering is then rotated to conformity with the decision ordering and the PCC index recorded. A second dichotomous ordering is then created from the second-lowest observed value (viz., 1.70 to 3.20; >3.20 to 9.00) and rotated to conformity with the decision ordering. The PCC index is again recorded. This process of (1) creating a dichotomous ordering based on the next highest idealism value, (2) rotating to conformity with the decision ordering, and (3) recording the PCC is continued until the highest observed value is reached. For the idealism ordering this process generated 60 PCC indices which were then plotted in Figure 3. As can be seen, the PCC indices increase from the 1.70 cut-point to a maximum for the 5.80 cut-point value. They then decrease unevenly until the highest cut-point value (8.90) is reached. The uneven decrease is due to the lack of a crisp threshold upon which to dichotomize the idealism scores. This fact can be understood in two ways. First, the multigram in Figure 4 shows the full scale idealism ordering conformed to the decision ordering. As can be seen, clear majorities of students with relatively low idealism scores (< 5.30) decided to continue the research, whereas clear majorities of students with high idealism scores (> 6.70) decided to stop the research. The ability to discriminate between those who voted to stop or continue the research is less clear, however, for those idealism scores near the median of the distribution (>5.20 and <6.80). In other words, a crisp threshold is not visible. Second, the PCC plots in Figure 5 show idealized results based on four distributions with clear thresholds. These distributions were created with the same number of scale values as the idealism scores, but the data were generated and manipulated to possess clear thresholds with PCC values of 100%. The simulated data represent normal and uniform distributions with thresholds at the 50th and 75th percentile. As can be seen, each plot shows a clear peak in the PCC indices which reveals the threshold value. A peak can be seen in the PCC plot for the idealism scores in Figure 3, but it is not as well defined as the ideal cases presented in Figure 5.

Using 5.80 as the threshold value, the idealism ordering was dichotomized and rotated to conformity with the decision ordering. The multigram in Figure 6 shows a clear pattern for students with relatively low (1.7 – 5.80) and high (5.9 – 9.00) idealism scores. Whereas as a majority of the low idealism students (66.33%) voted to continue the research, a larger majority of the high idealism participants (70.97%) chose to stop the research. The overall PCC index (69.52%, c-value < .01) indicates that classifying the participants solely on the basis of their idealism scores yields results slightly superior to those obtained from the classification based on gender above (66.03%).

The Threshold Analysis for the relativism ordering (variable) produced an odd shaped PCC plot (see Figure 7) without a clear peak. Nonetheless, the threshold was identified as 6.5 on the scale (range = 1 to 9), and the resulting multigram in Figure 8 shows that while a majority of students (64.49%) at or below the threshold value decided to stop the research, a slim majority (51.49%) above the threshold value decided to continue. The PCC index was equal to 60.32 (c-value = .01) and lower in magnitude than the result obtained for idealism (69.52%).

ROC CURVES AND CUT-SCORE OPTIMIZATION

The Threshold Analysis is similar to a Receiver Operating Characteristic (ROC) curve, and a comparison between the two methods is instructive. Originally used during World War II in conjunction with radar to assist radar operators in discriminating between friendly and enemy ships, ROC curves are used today in medicine (Grzybowski & Younger, 1997; Choi, 1998) and machine learning (Bradley, 1997; Kubat, Holte, & Matwin, 1998), among other applications.

Similar to the observation oriented modeling approach, in an ROC curve each recorded data value of the presumed continuous variable can be considered as a possible threshold or demarcation point for dichotomizing the observations for comparison to the binary variable.[4] Classification accuracy is calculated as a ratio of the sum of the true positives and true negatives to total sample size. Determining the best demarcation point is a matter of optimizing sensitivity, defined as the true positive rate, and specificity, defined as the true negative rate. The goal, simply stated, is to be as accurate as possible while limiting the number of false positives (akin to a Type I error) and false negatives (Type II error). This information can be graphed in the form of a curve and a diagonal reference line. Figure 9 shows an ideal curve and reference line. The area under the curve (AUC) relative to the reference line represents the probability that a randomly chosen positive case will be ranked higher than a randomly chosen negative case (assuming a positive condition is coded higher than a negative condition; Fawcett, 2005). Higher AUC values are therefore desirable.

The ROC curve for idealism in Figure 9 yielded an AUC value equal to .74. Interestingly, the SPSS software reported the following warning with this analysis: "The test result variable(s): Idealism has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased." Nonetheless, similar to the process of evaluating the PCC plot in Figure 3, the goal is to identify the highest peak in the curve relative to the diagonal reference line. In this case the ideal demarcation point was determined through trial and error to be 6.55, which yielded sensitivity equal to .76 and specificity equal to .59. The overall accuracy was equal to 68.25% and was slightly lower than the value obtained from the Threshold Analysis (69.52%).

The ROC curve for the relativism scale is presented in Figure 9 as well with an AUC value equal to .60. Similar to the corresponding PCC plot in Figure 7, the curve itself is much less pronounced and less smooth than the curve for idealism.[5] Moreover, at one point the curve actually bows across the reference line, meaning there is a chance that a randomly chosen positive case will be ranked lower than a randomly chosen negative case. Nonetheless, while using a threshold value of 6.35 provides the greatest overall accuracy (60.0%) because of high sensitivity, it does so at the expense of having the highest false-positive rate. Using 6.25 might be slightly better overall, even if the classification accuracy is slightly lower (.59), if for no other reason than a better false-positive rate. With either cut-point, the accuracy values from the ROC analysis for the relativism scores were similar to the PCC index from the Threshold Analysis (60.32%).

It is also instructive to compare the results of the Threshold Analyses for the idealism and relativism scales to those obtained from a cut-score optimization procedure. The algorithm for this procedure was developed by Barrett (see Barrett, 2016 for an example application) and written in Statistica Basic within the Statistica software package. Like the Threshold Analysis in the OOM software, it works iteratively through every possible value of the prediction variable while making no assumptions about the variable's quantitative nature. Figure 10 presents the results of the cut-score optimization routine for the idealism scale. If the goal is to optimize accuracy, the peak of the predicted accuracy line is located at the cut-score of 5.9 with an accuracy value equal to 69.52% (sensitivity = 0.82, specificity = 0.51). Figure 11 presents the results of the cut-score optimization routine for the relativism scale. The cut-score that optimizes accuracy is 6.6 on this scale with a total predictive accuracy equal to 60.32% (sensitivity = 0.41, specificity = 0.74). With both scales, the predictive accuracy values were identical to those determined from the Threshold Analysis.

COMBINING ORDERINGS

Wuensch and Poteat regressed the decision variable onto the gender, idealism, relativism, and type of research (dummy coded) variables. With all of the variables combined into a composite, the model explained 24.26% of the variance in the decision variable and accurately classified 71.75% students in the sample. Orderings (variables) can similarly be examined in combined form in observation oriented modeling. A straightforward approach involves crossing the orderings, much like creating a contingency table from different variables. Crossing all four dichotomous versions of the orderings (gender, group, idealism, and relativism) from Wuensch and Poteat's study, for example, yields a new ordering comprised of sixteen groups. Rotating this ordering to conformity with the decision to stop or continue the research yielded a high and unusual PCC index (73.02%, c-value < .001); however, with so many groups created from the crossed orderings, the results are difficult to interpret on a conceptual or pragmatic level.

A simpler and more parsimonious way forward is to begin with the ordering yielding the highest individual PCC index and then to test combinations of orderings in an effort to maximize the PCC. From the threshold analysis above the idealism ordering resulted in the highest PCC index (69.52%), and simple combinations of orderings revealed that crossing gender, idealism, and relativism improved the PCC index to 72.06% (c-value < .001). The multigram from this analysis in Figure 12 shows that majorities of female students who were relatively high in idealism voted to stop the research, while majorities of women low in idealism decided to continue the research. The results for male students were more complex. While those low in idealism predominantly voted to continue the research, those high in idealism and low in relativism decided to stop the research. A majority of men who were high in both idealism and relativism voted to continue the research.

Yet another way to combine the orderings that might also yield a more parsimonious prediction model is to use logical analyses. Ragin (2010) refers to such an approach as Qualitative Comparative Analysis (see also Fiss, 2011), and in observation oriented modeling it is referred to as Logical Hypothesis Testing (Grice, 2011, pp. 201-206). As the latter title implies the goal of the analysis is to form logical combinations of the dichotomous predictor orderings that conform accurately to the decision ordering. These combinations can be created manually or automatically using specific software, and they can include a variety of logical operations (e.g., conjunction, disjunction, and logical implication). Using the automated functions in the OOM software, the following logical statement was found to yield a high (71.11%) and unusual (c-value < .001) PCC index:

$$\text{Stop Research} \equiv ((\text{Female} \lor \text{Non-Medical Research}) \land \text{High Idealism}_{5.9\text{-}9.0})$$

Specifically, female students or students in the non-medical research groups who were also high in idealism chose to stop the research with greater frequency (73.66%) than those students who did not fit this profile (33.03%).

Finally, more nuanced combinations of the predictor orderings can be examined by blending the Procustes rotation, Threshold Analysis, and Logical Hypothesis Testing methods. For instance, different threshold values for the idealism and relativism scores can be determined separately for men and women. The logical analyses above assumed the threshold values to be equal for men and women. Examination of the PCC plots for women revealed that a clear threshold could not be determined for the relativism scores. A threshold was established, however, for idealism for the women, and separate thresholds were established for men on both orderings. With these gender-specific dichotomized orderings in hand (excluding relativism for women), logical combinations were then examined in an effort to maximize their conformity with the decision ordering. Results revealed the following logical statement with a high and unusual PCC (73.02%, c-value < .001):

$$\text{Stop Research} \equiv (\text{High Idealism}_{6.0\text{-}9.0} \wedge \text{Female}) \vee (\text{Low Relativism}_{1.0\text{–}5.1} \wedge \text{Male})$$

High idealism women or low relativism men were classified (or predicted) as deciding to stop the controversial research, whereas all other individuals were classified as deciding to continue the research. In this manner, very specific and interesting profiles can be constructed and examined when observation oriented modeling is adopted.

DISCUSSION

The re-analysis of Wuensch and Poteat's (1998) data showed that observation oriented modeling is an intriguing alternative to logistic regression. In terms of overall classification accuracy, several prediction profiles yielded classification accuracy values (PCCs of 71.11% and 73.02%) nearly as high or higher than the equivalent value resulting from the regression model (71.75%). Because the goal of the analysis was to identify robust patterns within the data, several of the observation oriented analyses yielded predictive profiles in the form of logical statements – similar to decision trees from CART analyses (Brieman et al., 1984) – rather than equations. Multiple regression equations express functions relating combinations of predictor variables to an outcome variable, and the goal for most social and life scientists is to maximize the multiple R^2, the proportion of variance explained for the outcome variable. It is interesting this goal does not necessarily match the goal of maximizing the classification accuracy from the logistic regression. Indeed, for Wuensch and Poteat's data all but the gender and idealism predictors can be dropped from the equation without lowering the classification accuracy. In other words, even though most of the other variables were statistically significant and naturally increased the multiple R^2 (from .20 to .24) when included in the regression equation, they did not add to the function's accuracy in classifying participants' decisions to stop or continue the research. The observation oriented modeling analyses made this fact clear as conforming the participants' decisions to the idealism ordering alone yielded a PCC index (69.52%) nearly as high as 71.75%.

Analyses combining the variables (referred to as orderings) in the context of observation oriented modeling took three forms. First, the dichotomized predictor orderings were simply crossed and then conformed to the dichotomous decision outcome. Results revealed that by crossing gender, idealism, and relativism the PCC index could be improved to 72.06% and the pattern of findings interpreted. Second, logical combinations of the dichotomized predictors were explored in an effort to maximize the PCC index. The following predictive logical statement (PCC = 71.11%) was discovered:

$$\text{Stop Research} \equiv ((\text{Female} \vee \text{Non-Medical Research}) \wedge \text{High Idealism}_{5.9\text{-}9.0})$$

Third, permitting different thresholds between genders, the idealism and relativism scores were dichotomized separately for men and women, and then logical combinations of the dichotomized predictors were again examined. This analysis yielded the highest PCC index (73.02%) and the following nuanced logical statement:

$$\text{Stop Research} \equiv (\text{High Idealism}_{6.0\text{-}9.0} \wedge \text{Female}) \vee (\text{Low Relativism}_{1.0\text{–}5.1} \wedge \text{Male})$$

Those deciding to stop the research were characterized as idealistic females or absolutist males.

Beyond the arduous task of exploring higher-order interactions there is simply no equivalent means for developing such profiles in logistic regression, and as shown above statistically significant predictors may not even be necessary for achieving a high classification accuracy. Harris (1993, 2001) has long advocated interpreting regression weights in linear regression, MANOVA, and CANONA as well as factor score coefficients in factor analysis as a means for developing multivariate profiles (see particularly, Harris, Harris, & Bochner, 1982, and Grice & Harris, 2001). It is via these types of coefficients that truly multivariate information is extracted from the analysis. While there is no clear correspondence between the profiles generated from the observation oriented analysis and logistic regression, it is nonetheless in the spirit of Harris' argument that the logical profiles above were constructed and evaluated.

The development of the profiles required the identification of demarcation or "threshold" points for the idealism and relativism scores. In the realm of traditional inferential statistics where the most common goals are to model means, maximize variance explained, and estimate population parameters, dichotomizing is almost universally frowned upon. The primary reason is the general loss of statistical power; i.e., the lowering of the probability of correctly rejecting a false null hypothesis (Fitzsimons, 2008; MacCallum et al., 2011). In observation oriented modeling, by contrast, the goal is to establish an inference to best explanation based on the patterns within the observations. In other words, the goal is abduction (Haig, 2005, 2008, 2014). By seeking patterns and using assumption-free randomization tests (i.e., c-values), concepts associated with null hypothesis significance testing, such as statistical power and Type I, II, and III errors, may be avoided entirely (see Grice, 2014, 2015). Assessing the legitimacy of dichotomizing observations from the Threshold Analyses above would thus switch from a statistical to a theoretical concern. For instance, an explanatory model describing how each student's judgment of the sensitive animal research concludes in a dichotomously-structured decision and behavior (i.e., marking "continue" or "stop" on the questionnaire) would need to be developed. If idealism is truly a continuity along which people can be ordered, is there also a point along the continuity at which every person will "toggle" between stopping or continuing the research? If such a threshold point is not expected, exactly how does a continuous quality cause a dichotomously-structured decision? The point here is to notice that we are constrained, in a sense, by the dichotomous effect. In observation modeling, it is generally argued that effects and causes must conform to one another, and the ultimate goal is therefore to build an explanatory model showing how a dichotomous effect can result from a continuous cause. The only way this seems possible and reasonable for Wuensch and Poteat's study is to reduce the cause to a dichotomy.

It is worth noting as well that logistic regression essentially reduces continuities or a continuous function to a dichotomous prediction, as shown in the classification formula provided above. An important difference between the Threshold Analysis and logistic regression is that the former permits the researcher to readily examine exactly where the demarcation point is located on the presumed continuous scale, similar to the approach undertaken in the ROC curve and cut-score optimization analyses. All three methods yielded similar or identical threshold values for the relativism and idealism scales, thus supporting the efficacy of the Threshold Analysis. Most importantly, these methods encourage researchers to pay attention to how their observations are actually structured and to stay as close to the data as possible…which is the basis for the "observation oriented" moniker. With regard to Breiman's (2001) two cultures of stochastic and algorithmic statistical modeling (see also

Woodside, 2013), observation oriented modeling is most similar to the latter. Rather than presume a stochastic process or linear model in a bid to explain the maximum amount of variance in an outcome variable, the goal is instead to create a model that accurately matches and explains the patterns within a given set of observations. In brief, the goal is to maximize accuracy. The visual tools (the PCC plots and multigrams for the analyses herein) and the PCC index also insure that results are readily interpretable and conveyable to both experts and lay persons alike, thus avoiding the need for esoteric and arbitrary conventions to interpret effect sizes (e.g., d = .20 represents a small effect).

Regardless of the agreement among the ROC, cut-point, and threshold analyses, an obvious concern is the stability of the thresholds and the profiles from the re-analysis of Wuensch and Poteat's data. In brief, how stable are the specific threshold values, the profiles, and the resulting PCC indices across samples? As with the traditional p-value from null hypothesis significance testing, the c-value from the observation oriented analyses offers absolutely no information regarding the probability of replicating a magnitude discovered in a study (Cohen, 1990; Maraun & Gabriel, 2010). Consequently, the only way to answer the question of stability across independent samples is to collect such data and conduct the analyses again. As has been made abundantly clear to psychologists in recent years (Open Science Collaboration, 2015; Yong, 2012), there is simply no substitute for exact replication

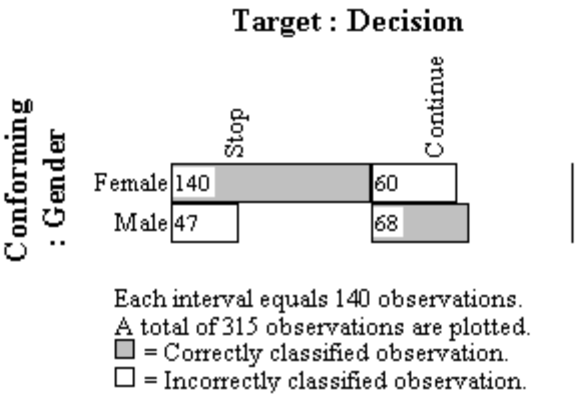


Figure 1. Multigram showing results for rotating the gender ordering to the research decision ordering

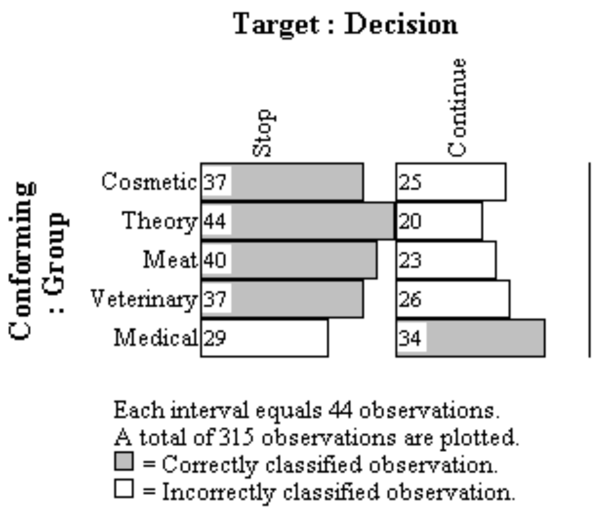


Figure 2. Multigram showing results for the type of research rotated to conformity with the decision regarding the research

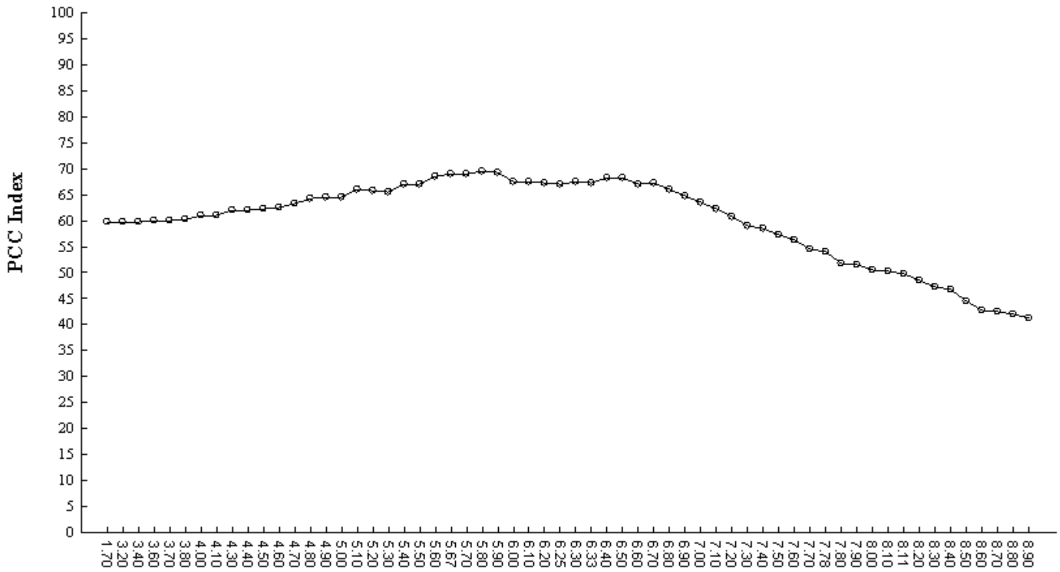


Figure 3. Plot of PCC indices from Threshold Analysis for idealism scores

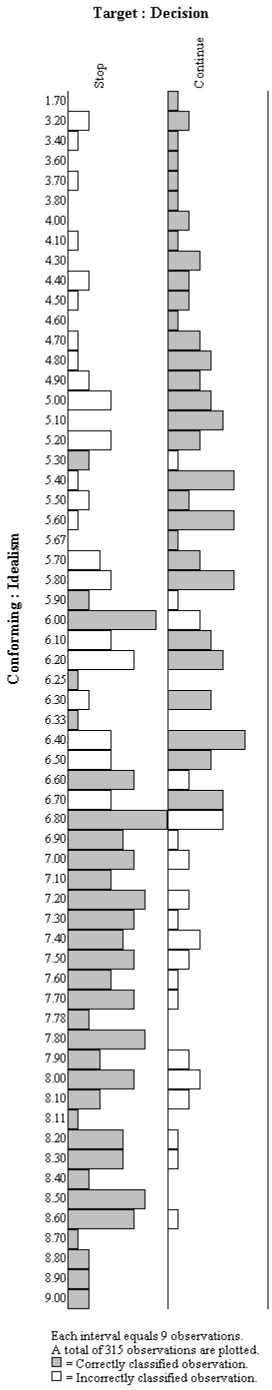


Figure 4. Multigram showing results for rotating idealism to conformity with the decision to stop or continue support for the research

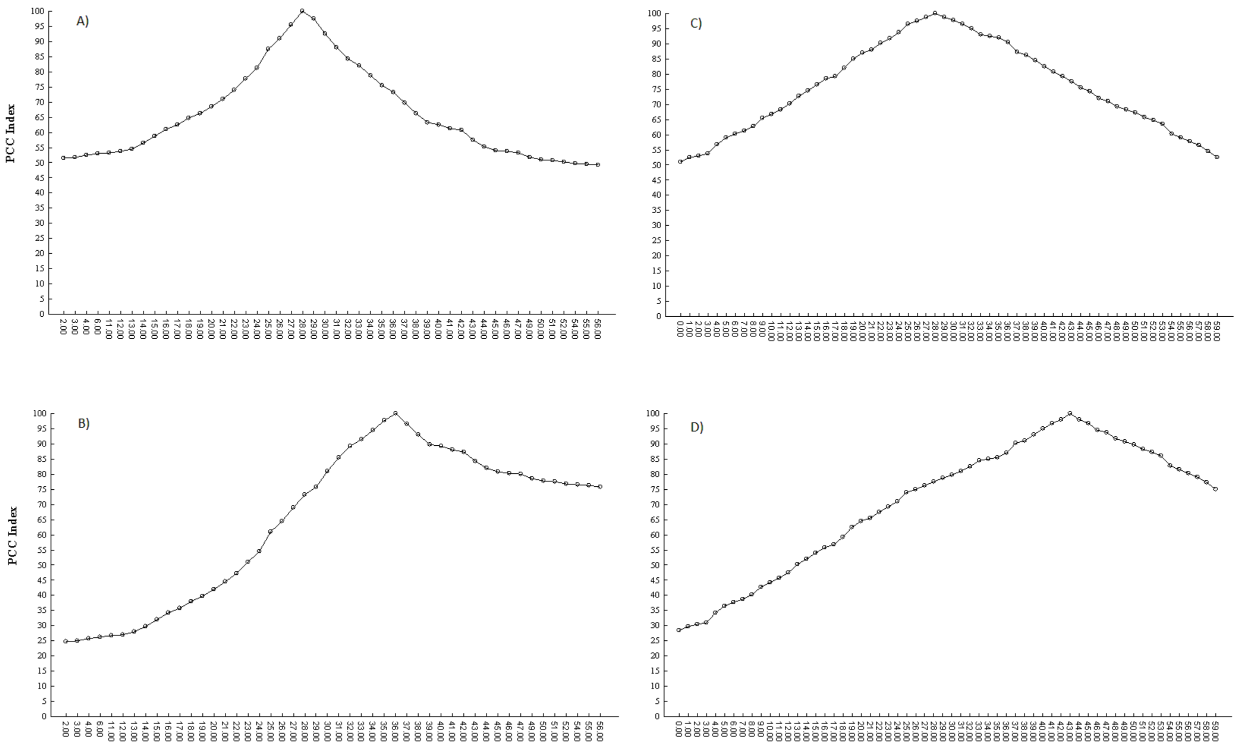


Figure 5. Plots of PCC indices for idealized data: A) normal distribution, 50th percentile threshold; B) normal distribution, 75th percentile threshold; C) uniform distribution, 50th percentile threshold; D) uniform distribution, 75th percentile threshold

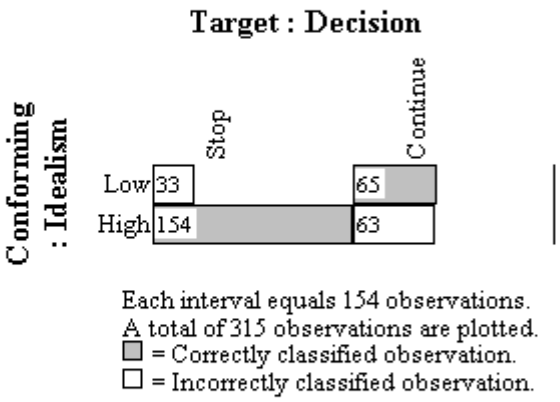


Figure 6. Multigram showing results for rotating the dichotomized idealism ordering to conformity with the research decision ordering

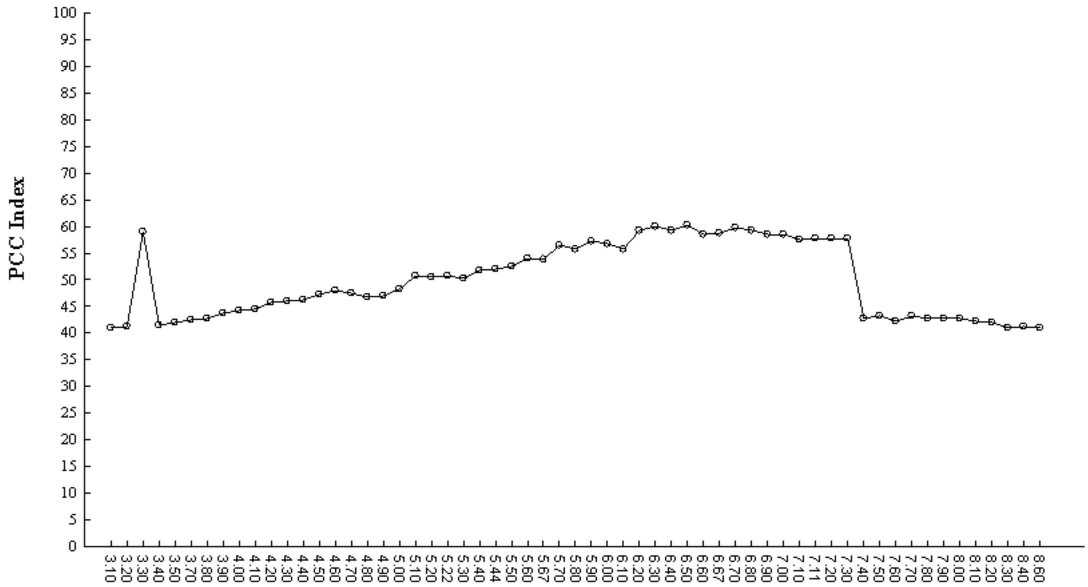


Figure 7. Plot of PCC indices from Threshold Analysis for idealism scores.

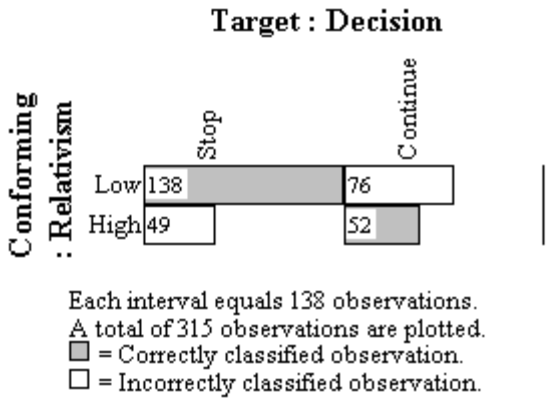


Figure 8. Multigram showing results for rotating the dichotomized relativism ordering to conformity with the research decision ordering

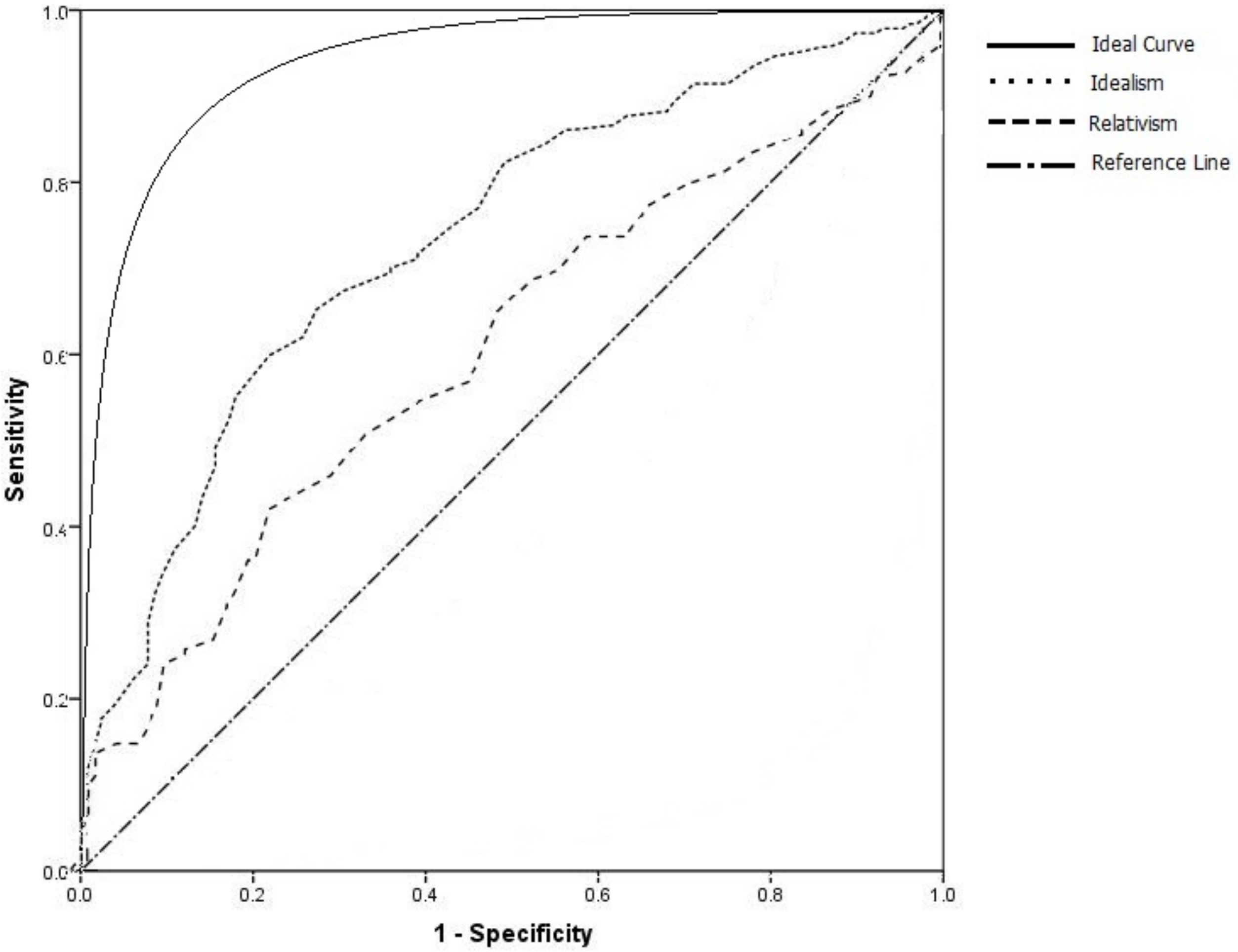
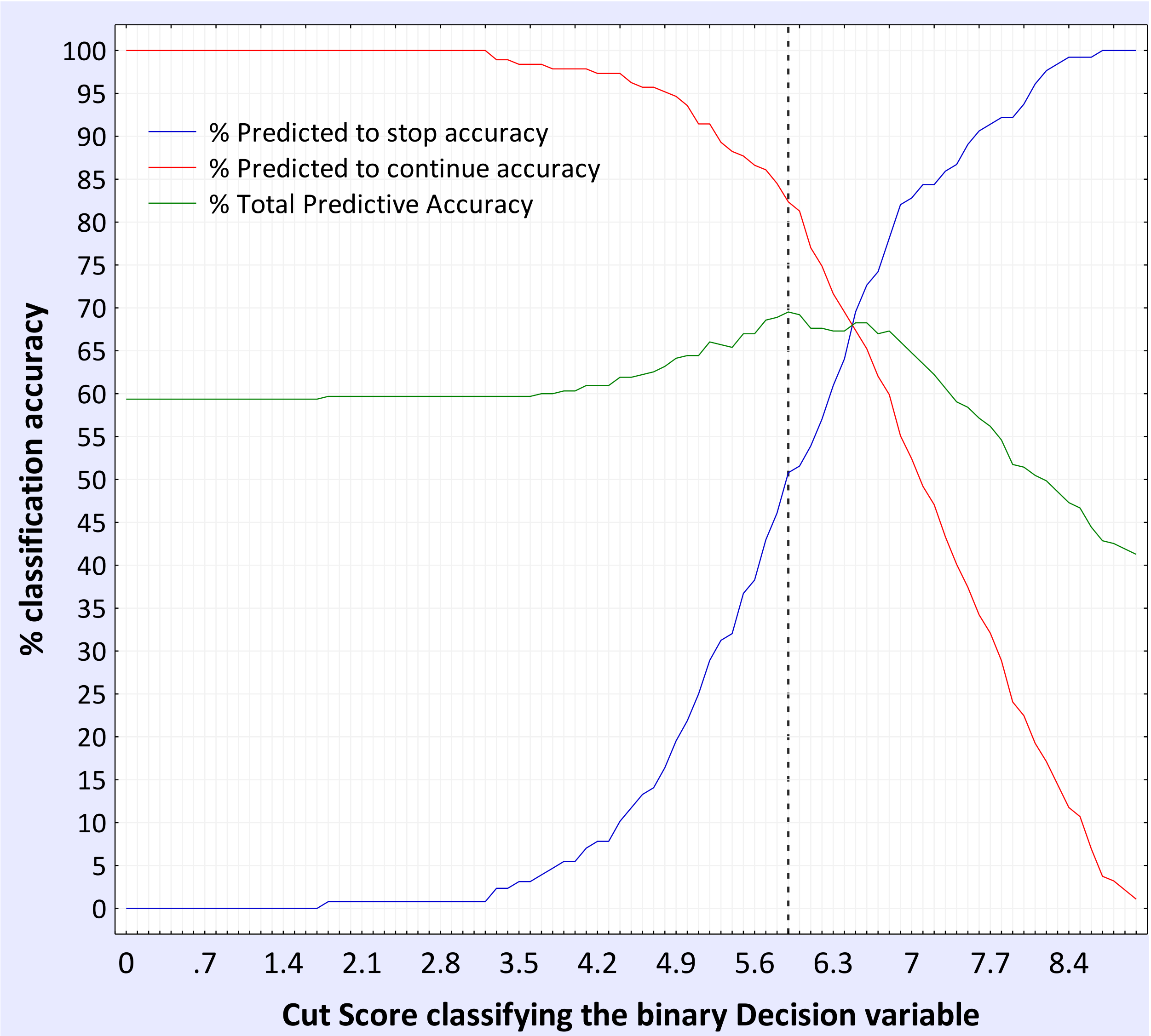
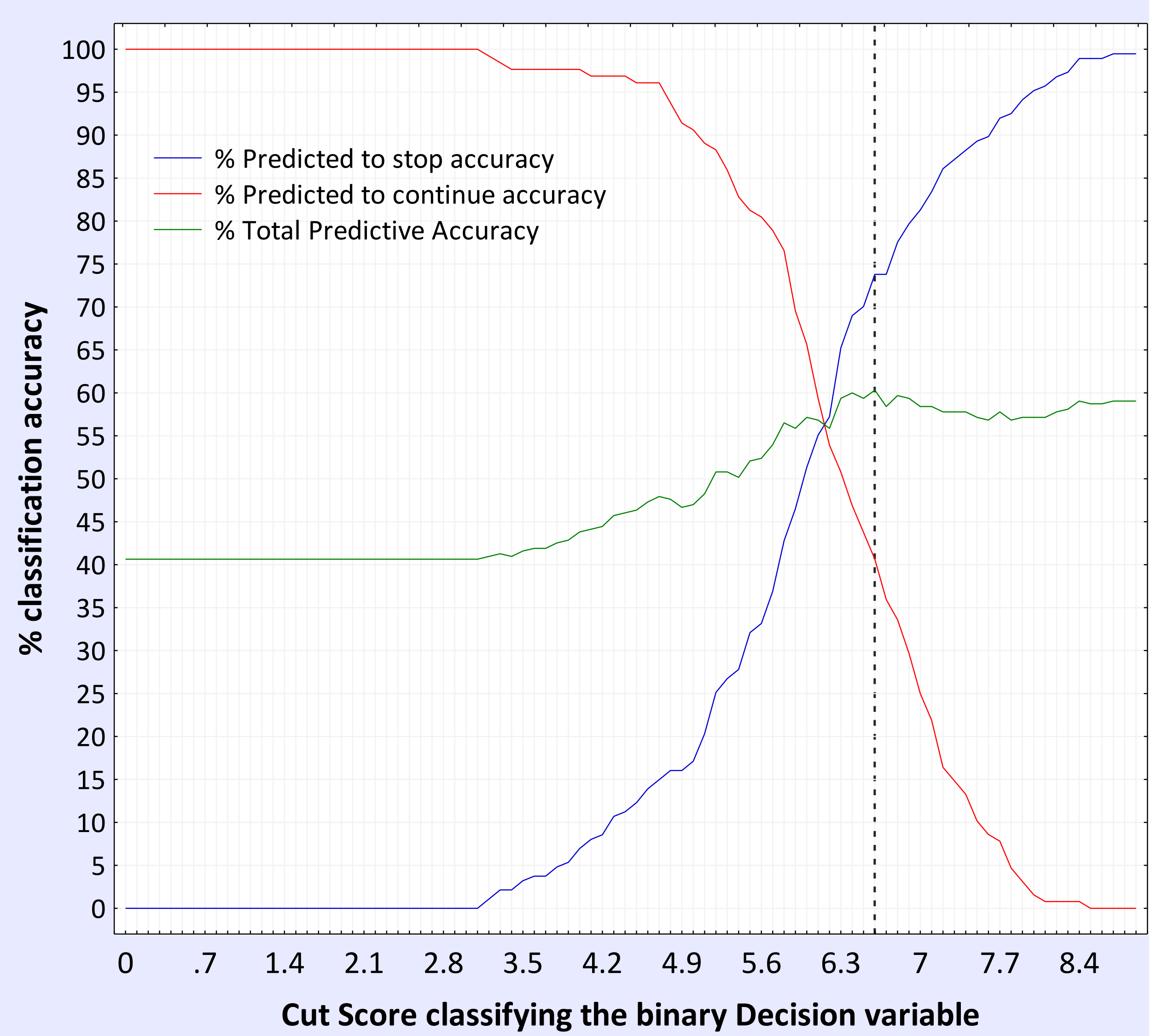


Figure 9. ROC curves comparing idealism and relativism (both presumed continuous) to the dichotomous decision to stop or continue the research. An ideal ROC curve is also shown for comparison



Note: The black dotted line is a reference line indicating the optimal cut-score at 5.9, with a total predictive accuracy of 69.52%

Figure 10. The cut-score maximization graph for idealism, classifying the class membership of the decision outcome (indicating prediction accuracy).



Note: The black dotted line is a reference line indicating the optimal cut-score at 6.6, with a total predictive accuracy of 60.32%

Figure 11. The cut-score maximization graph for relativism, classifying the class membership of the decision outcome (indicating prediction accuracy)

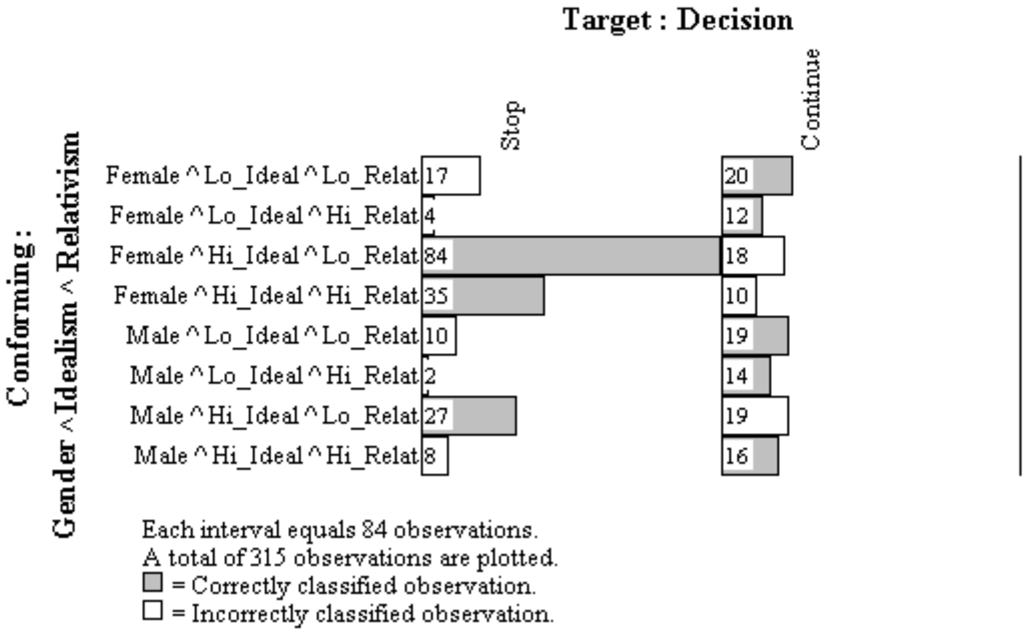


Figure 12. Multigram showing results for the crossed gender and dichotomized idealism and relativism orderings rotated to conformity with the decision ordering.

FOOTNOTES

1. The data from Wuensch and Poteat's (1998) study are available in SPSS format online at: http://core.ecu.edu/psyc/wuenschk/SPSS/Logistic.sav. The logistic regression results can be found at: (http://core.ecu.edu/psyc/wuenschk/MV/MultReg/Logistic-SPSS.pdf).
2. The OOM software may be freely downloaded from http://www.idiogrid.com/OOM. The analysis was conducted in the OOM software under the Build/Test Model analysis feature. The Conforming Only normalization option was chosen for the analyses in which the continuous or polychotomous causal orderings were rotated to conformity with the dichotomous effect ordering. Grice (2013, pp. 30-32) describes the two types of normalization and their impact on how observations are classified as correct or incorrect in the binary Procrustes analysis.
3. Wuensch and Poteat (1998) describe the EPQ scores as continuous, "Logistic regression was chosen over discriminant function analysis because we wanted to evaluate simultaneously the effects of two continuous predictors, one dichotomous predictor, and one qualitative predictor" (p. 145).
4. It is possible to use ROC curves to quantify an assumed continuous variable in terms of a multiple-class variable, although this procedure quickly becomes complicated depending on the number of classes in that variable (see Fawcett, 2006; Hand & Till, 2001; and Provost & Domingos, 2001).
5. For both the ROC and cut-score optimization analyses, the decision variable was reflected when compared to the relativism scale scores. It is reasonable to expect individuals high in relativism to decide to continue the researcher and those low in relativism to stop the research. The direction of this effect is opposite of what would be expected between the decision and idealism variables. The results of the logistic regression analysis were also consistent with these expectations.

References

Anderson, J. & Grice, J. (2014). Toward an integrated model of spirituality. Journal of Psychology and Christianity, 33(1), 3-22.

Barrett, P.T. (2016). Hierarchical multiple linear regression: Interpreting deviation R-squares. Cognadev Technical Whitepaper #6. Retrieved from: http://www.pbarrett.net/cognadev/techseries/Cognadev_Technical_Report_6_Deviation_R-Square_17_June_16.pdf

Bradley, A. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30, 1145–1159.

Breiman, L. (2001). Statistical modeling: The two cultures. Statistical Science, 16(3), 199-231.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). Classification and regression trees. Belmont, CA: Wadsworth.

Brown, E. A. & Grice, J. W. (2012). Mediation analysis via Observation Oriented Modeling. International Journal of Sciences, 1, Nov., 1-42.

Choi, B. C. K. (1998). Slopes of a Receiver Operating Characteristic curve and likelihood ratios for a diagnostic test. American Journal of Epidemiology, 148, 1127-1132.

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45(12), 1304-1312. Comer, J., Bayles, M., Beier, C., Gelder, G., Miller, B., Wilber, G., Gordon, S.R., & Cota, L. D. (2014). Oklahoma State University Annual Assessment Report. Committee for the Assessment of General Education.

Craig, D. P. A., Varnon, C. A., Pollock, K. L., & Abramson, C. I. (2015). An assessment of horse (Equus ferus caballus) responding on fixed interval schedules of reinforcement: An individual analysis. Behavioural Processes, 120, 1-13.

Dinges, C., Avalos, A., Abramson, C., Craig, P., Austin, Z., Varnon, C., Nur Dal, F., Giray, T., & Wells, H. (2013). Aversive conditioning in honey bees (Apis mellifera anatolica): A comparison of drones and workers. The Journal of Experimental Biology, 216, 4124-4134. Doi: 10.1242/jeb.090100

Edgington E. & Onghena, P. (2007). Randomization tests (4th Ed.). London: Chapman & Hall.

Fawcett, T. (2005). An introduction to ROC analysis. Pattern Recognition Letters, 27, 861-874.

Fitzsimons, G. (2008). Death to dichotomizing. Journal of Consumer Research, 35, 5-8.

Fiss, P. C. (2011). Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research. Academy of Management Journal, 54, 393-420.

Gatobu, S. K., Arocha, J. F., & Hoffman-Goetz, L. (2016). Numeracy, health numeracy, and older immigrants' primary language: An observation-oriented exploration. Basic and Applied Social Psychology, doi:10.1080/01973533.2016.1197129

Grice, J. W. (2016). Observation Oriented Modeling Software Manual (Version 2). Stillwater, OK: Idiogrid Software.

Grice, J. W. (2015). From means and variances to patterns and persons. Frontiers in Psychology, 6:1007. doi: 10.3389/fpsyg.2015.01007.

Grice, J. W. (2014). Observation Oriented Modeling: Preparing students for the research in the 21st Century. Innovative Teaching, 3, 3. DOI 10.2466/05.08.IT.3.3

Grice, J. W., Barrett, P. T., Schlimgen, L. A., & Abramson, C. I. (2012) Toward a brighter future for psychology as an observation oriented science. Behavioral Sciences, 2(1), 1-22.

Grice, J. W. (2011). Observation oriented modeling: Analysis of cause in the behavioral sciences. New York, NY: Academic Press.

Grice, J. W., & Harris, R. J. (1998). A comparison of regression and loading weights for the computation of factor scores. Multivariate Behavioral Research, 33, 221-247.

Grzybowski, M., & Younger, J. G. (1997). Statistical Methodology: III. Receiver Operating Characteristic (ROC) Curves. Academic Emergency Medicine, 4, 818-826.

Haig, B.D. (2005). An abductive theory of scientific method. Psychological Methods, 10, 371–388.

Haig, B. (2008) Pre´cis of 'An abductive theory of scientific method'. Journal of Clinical Psychology, 64, 1019-1022. doi: 10.1002/jclp.20506

Haig, B. (2014). Investigating the psychological world. Cambridge, MA: MIT Press.

Hand, D.J., & Till, R.J., 2001. A simple generalization of the area under the ROC curve to multiple class classification problems. Machine Learning, 45,171–186.

Harris, R. J. (2001). A primer of multivariate statistics (3rd Ed.). Mahway, NJ: Lawrence Erlbaum.

Harris, R. J. (1993). Multivariate analysis of variance. In L. K. Edwards (Ed.) Applied analysis of variance in behavioral science, pp. 255-296. New York: Marcel Dekker, Inc.

Harris, M. B., Harris, R. J. and Bochner, S. (1982), Fat, Four-Eyed, and Female: Stereotypes of Obesity, Glasses, and Gender. Journal of Applied Social Psychology, 12, 503–516. doi: 10.1111/j.1559-1816.1982.tb00882.x

Kubat, M., Holte, R. C., & Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. Machine Learning, 30, 195-215.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. Psychological Methods, 7(1), 19-40.

Manly, B. (1997). Randomization, bootstrap, and Monte Carlo methods in biology (2nd Ed.). London: Chapman & Hall.

Maraun, M., & Gabriel, S. (2010). Killeen's (2005) prep coefficient: Logical and mathematical problems. Psychological Methods, 15(2), 182-191.

Michell, J. (2011). Qualitative research meets the ghost of Pythagoras. Theory & Psychology, 21(2), 241-259. DOI: 10.1177/0959354310391351

Open Science Collaboration (2015). Estimating the reproducibility of psychological science.

Science, 349, aac4716. doi:10.1126/science.aac4716

Provost, F., & Domingos, P. (2001). Well-trained PETs: Improving probability estimation trees. CeDER (Working Paper #IS-00-04). Accessed 7/5/2014 from http://citeseer.ist.psu.edu/viewdoc/download;jsessionid=BC0F691AA8F5AC6CA65C572627A52B73?doi=10.1.1.33.309&rep=rep1&type=pdf

Ragin, Charles C. (2010). User's Guide to Fuzzy-Set/Qualitative Comparative Analysis 2.0. Irvine, California: Department of Sociology, University of California.

Valentine, K., & Buchanan, E. (2013). JAM-boree: An application of observation oriented modelling to judgments of associative memory. Journal of Cognitive Psychology, 25(4), 400-422. doi:Winch, R. & Campbell, D. (1969). Proof? No. Evidence? Yes. The significance of tests of significance. The American Sociologist, 4, 140-143.

Woodside, A. G. (2013). Moving beyond multiple regression analysis to algorithms: Calling for adoption of a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. Journal of Business Research, 66, 463-472.

Wuensch, K. L. & Poteat, G. M. (1998). Evaluating the morality of animal research: Effects of ethical ideology, gender, and purpose. Journal of Social Behavior and Personality, 13, 139-150.

Yong, E. (2012). Bad Copy. Nature, 485, 298-300.