International Journal of Quantitative and Qualitative Research Methods

Vol.9, No.1, pp.1-13, 2021

ISSN 2056-3620(Print)

ISSN 2056-3639(Online)

AN ASSOCIATION RULE GENERAL ANALYTIC SYSTEM (ARGAS) FOR HYPOTHESIS TESTING IN QUALITATIVE AND QUANTITATIVE RESEARCH

Frederick Parente
Postal Address: 1212 Brookview Road, Towson, Maryland, 21286
Email: fparente@towson.edu

John-Christopher Finley
Email: jfinle2@students.towson.edu

Christopher Magalis
Towson University
Email: cmagalis@towson.edu

ABSTRACT: *This paper describes an Association Rule General Analytic System (ARGAS) as an alternative to the General Linear Model (GLM) for hypothesis testing. We illustrate how the ARGAS can be used to analyze both qualitative and quantitative research data. The advantages of the ARGAS approach derives from the fact that it is designed to analyze words or numbers that are converted into words. Unlike the GLM, it does not have any distributional assumptions. Association rule calculations are well-developed and there are a variety of computer software applications available that expedite the computations. The purpose of this study is to illustrate how the ARGAS can be applied and how to interpret the results.*

KEYWORDS*:* ARGAS, GLM, hypothesis testing, association rule analysis, pattern recognition, quantitative, qualitative

INTRODUCTION

The General Linear Model (GLM) is a commonly used data analytic system that can accommodate a variety of different research designs and data (Christensen, 2011; Nelder & Baker, 1972). One category of quantitative research involves predicting one set of variables from others, and another examines mean differences among groups. Although the GLM is the most ubiquitous data analytic system for testing hypotheses with numerical data, it is not without limitations. One problem with the GLM is that it is not appropriate for analyzing words or text that are the typical units of measures in qualitative research. The GLM is also based on restrictive assumptions, e.g. homogeneity of variance, normality of distributions, independence of error terms etc. Unfortunately, these assumptions are not always evaluated before applying one of the variants of the GLM. Although there are transformation procedures available for modifying data to meet these assumptions, they are not commonly used nor do they always achieve their goal (Hutcheson & Sofroniou, 1999). Further, many GLM applications are designed to detect differences among arithmetic means of different

International Journal of Quantitative and Qualitative Research Methods
Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

conditions in an experiment. However, the means are hypothetical values which may not actually exist in the data and may be severely affected by extreme scores. Given these issues, it is reasonable to pursue alternative methods of data analyses that can analyze quantitative and qualitative research data and that obviate the distributional limitations of the GLM.

We propose an Association Rule General Analytic System (ARGAS) that provides an alternative to the GLM for analyzing a variety of different research designs. The ARGAS is specifically designed to analyze words and text rather than numerical data. It can, however, analyze numerical data that is first converted into words. It does not have any distributional assumptions and it does not examine differences between hypothetical mean values. It does identify specific patterns in the data, called rules, which associate the variables (Han, Pei, & Kamber, 2011) and it does test the significance of the rules. Like the GLM, the purpose of the ARGAS is to assess the association between a set of predictors (called "antecedents") and a set of outcomes (called "consequents"). The goal of the current study is to illustrate how the ARGAS approach can be used as a data analytic tool for hypothesis testing. The analysis fills a gap in the data science methodology by providing an analytic model that can be adapted to most numerically based quantitative research designs and to most word based qualitative research projects. As a prelude to this discussion, we begin with a basic description of association rule analysis which is the foundation of the ARGAS.

LITERATURE /THEORETICAL UNDERPINNINGS

The concept of Association Rules is often explained via a "market basket analysis" example, which is used by many retailers (Agrawal & Srikant, 1994; Webb, 2003). Each person's purchases can be conceptualized as a vector of items which are amalgamated into a larger data base of words that represent the consumers' collective purchases. This larger data structure can be analyzed in aggregate to extract association rules that derive from the frequency with which two or more items are purchased together. This type of analysis allows the retailer to predict what products will be purchased along with others. Though, initially intended for marketing, association rules have been recently used for testing hypotheses in qualitative research (Parente, Finley and Magalis, 2020). However, there have been few other hypothesis testing applications of association rule analysis outside of the marketing parlance.

Association rules are the joint probabilities that predict the co-occurrence of *antecedent* and *consequent* events. Antecedent variables are similar to independent or predictor variables in GLM analyses. Consequent events are analogous to dependent or outcome variables. Association rules are the probabilities that describe how one or more antecedent(s) affects one or more consequent(s). Computing dependent joint probabilities can be a complex process. For simple rules, the process is straightforward, however, more complex rules that involve several variables usually require specialized computer software (e.g., Webb, 2007).
Conventional association rule analyses provide a variety of measures of association between the antecedent and consequent variables. Most of these statistics are hybrids of the

dependent conditional probability computations mentioned above (Balcázar & Dogbey, 2013; Parente & Finley, 2018). The various measures are different ways to assess the associative relationships among a set of variables. Perhaps the simplest measure is the number of rules that the analysis generates. The more rules – the stronger the associative strength between the antecedents and consequents. Other measures include *Support* which is the proportion of cases in the data that include one or all of the antecedent and consequent values under consideration. *Coverage* is a measure of how extensively a given item occurs in the antecedent portion of the rule. *Confidence* is a conditional probability that a person experiences event B given that the he or she also experiences event A. The *Lift* measure is an index of the predictive value of the rule relative to using no rule at all. *Leverage* measures the extent to which two or more items are recalled together versus what would be expected if the items were recalled independently.

METHODOLOGY

We next describe how the ARGAS can be used to analyze qualitative and quantitative research. We begin with brief examples of qualitative studies where the units of measure were words that participants generated to describe their experience in different situations. The words may be derived from existing media, for example, words used in political candidates' speeches or tweets. Regardless of the source, association rules that derive from the analysis of these words or phrases are then scrutinized to capture the essence of the participant's personal experience. Because ARGAS is applicable for qualitative and quantitative research, we provide examples of its use in both domains.

RESULTS/FINDINGS

Qualitative Research Applications

Qualitative research involves analyzing words or text that describes human experience. Association rule analysis allows the researcher to test hypotheses regarding those experiences. For example, Magalis (2020) investigated differences among words that students chose to describe their emotional experience with math versus statistics. He hypothesized that math and statistics anxieties were different phenomenon. To test this hypothesis, he had groups of students solve either a simple quadratic equation or a standard deviation problem. After completing the solution, each student generated words that described their unique experience. They had also filled out a Math Anxiety Scale and a Statistics Anxiety Scale. Quantitative analysis of the scaling data did not show any significant difference between the math versus statistics problem groups. However, association rule analysis of the words showed that two rules distinguished the groups. Students in the statistics problem condition consistently generated the words "anxious" and "confusion" to describe their experience with the standard deviation problem. The math problem data analysis did not generate any rules. These findings suggest that statistics anxiety differs from math anxiety. It is, perhaps, a unique form of anxiety that is predisposed by feelings of confusion. It is interesting to note that none of the scaling questionnaire items mentioned confusion although this word was consistently generated to describe the students' experience with the statistics problem.

Parente and Finley (2018) investigated the use of association rules as a neuropsychological assessment tool. If association rules are indices of organizational structure and if memory is related to organization (Tulving, 1966), then the number of association rules that derive from participants' recall should correlate with measures of recall. To test this hypothesis, participants with brain injury learned a list of 12 unrelated nouns over 12 study and test trials. Each participants' word recall data was then subjected to association rule analysis to determine which of the derived measures (number of rules, confidence, lift, leverage etc.) would correlate with different measures of recall. The results indicated that recall was directly related to the number of association rules and the confidence measure. This finding suggests that the association rule measures can be used as an index of a person's ability to organize information in memory. The number of rules is an index of the associative structure of the words in memory. The confidence measure is an index of the strength of those associations.

Pilot study data for a study by St. Pierre and Parente (2018) investigated word choices by "mock Jurors" who judged whether or not a hypothetical defendant was innocent of his/her alleged crime. The researchers were interested in word choices that described the physical and psychological features of a person that influenced a jurors' judgment of innocence. Association rule analysis was then applied to the participants' word choices which generated ten rules. Psychological characteristics included: *gentle, friendly, sweet, victim, graceful, smart, shy, and strong*. *Symmetrical face* was the only physical characteristic identified. The students also used the word *irrational* that suggested their perception of innocence was arbitrary unique to each individual.

Quantitative Applications

The above examples illustrate how the ARGAS can be used to analyze non-numerical data that is typical in qualitative research. However, the analysis can also be applied to words or phrases that represent numerical measures (e.g, above-the- median or below-the-median). Yet, there is little research using association rules with numerical data that have been transformed in this manner (Agrawal, Imieliński, & Swami, 1993; Imberman & Domanski, 2001). The following paragraphs explain how the ARGAS can be used to analyze numerical data. Table 1 presents a breakdown of several common research paradigms. The left side of the table identifies GLM analyses that are appropriate for use in that paradigm. The right side of the table presents the alternative ARGAS analysis that would be appropriate once the numerical data has been transformed into words. In the following simplified examples, the authors transformed the numerical data into words that labeled the scores as either above-the-median or below-the-median. However, a more specific breakdown (e.g., quartiles) may be more appropriate with different data sets.

International Journal of Quantitative and Qualitative Research Methods
Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

Comparison of ARGAS and the GLM

		GLM				ARGAS	
Canonical Analysis	Multiple Predictors		Multiple Outcomes		Multiple Antecedents		Multiple Consequents
Multiple Prediction	Multiple Predictors		1 Outcome		Multiple Antecedents		1 Consequent
Multivariate Group Prediction	2 or more groups		2 or more outcomes		2 or more group labels		2 or more consequents
Univariate group comparison	2 or more groups		1 outcome		2 or more group labels		1 consequent

Table 1. GLM vs. ARGAS Approaches to Hypothesis Testing

We begin with an example of how ARGAS can be used to analyze canonical data structures in which there are multiple antecedent and multiple consequent variables. We next provide an example of the use of ARGAS as an alternative to conventional multiple regression analysis in which there are multiple antecedents and a single consequent. We then discuss how these two analyses can be modified to accommodate multivariate and univariate group comparison designs. In each of these examples we provide a comparison of the ARGAS with the corresponding GLM procedure regarding interpretation of the results.

Canonical Analysis

Canonical analysis assesses the relationship between a set of numerical predictors and a set of numerical outcome variables. The research hypothesis is that one or more of the predictors will predict one or more of the outcomes. The analysis provides an overall measure of association (canonical correlation) as well as indices of the value of each predictor and outcomes (called loadings). In essence, the canonical correlation is a Pearson product moment correlation between the linear combination of the predictor variables and a corresponding linear combination of outcome measures. Specifically, the set of predictor scores is collapsed into a weighted combination (called a variate) which is correlated with the outcome variate. The canonical correlation is, therefore, a bivariate correlation between the two variate scores. A significant canonical correlation indicates that the set of predictors and the set of outcomes are significantly correlated. The analysis also yields weighting coefficients (loadings) that are indices of the extent to which each individual variable in the predictor and outcome sets contributes to its variate. Another set of weights (called cross loadings) are indices of how well the same variable predicts the opposite variate.

To illustrate this process, we performed a canonical analysis on a public domain data set that assessed the correlative relationship among several student/teacher evaluations. Each of 50

International Journal of Quantitative and Qualitative Research Methods

Vol.9, No.1, pp.1-13, 2021

ISSN 2056-3620(Print)

ISSN 2056-3639(Online)

universities provided their average ratings for six student evaluation questions. The purpose of the study was to assess the relationship between the set of predictors that included: (1) the students' perception of the quality of the exams, (2) the average grade in the course, (3) enrollment in the course, and (4) the perceived knowledge of the instructor. The set of outcomes included: (1) the perceived teaching quality and (2) the overall evaluation of the course. Half of the variables met the distributional assumptions of the GLM, whereas the remaining variables did not, even after transforming the variables.

The canonical correlation analysis revealed a significant Rsqr (.782) between the predictor and outcome sets. The standardized coefficients for the individual predictors indicated that two variables i.e., perceived exam quality and perceived knowledge of the instructor significantly predicted the outcomes i.e., overall evaluation and teaching competence. Redundancy analysis indicated that the predictors accounted for 55% of the variance in the outcome set and about 30% of the variance in the predictor set.

The same data were then evaluated to derive association rules that related the same antecedents (predictors) with the consequents (outcomes). The individual variables were first split at their respective medians and the data for each variable were recoded into words that identified each value as either above or below-the-median. For example, the word "OVAbove" indicated that the numerical score value for the *Overall* evaluation variable was above-the-median for that variable. The word "ExmBelow" indicated that the numerical value for perceived Exam Quality was below-the-median for that variable. It is reasonable to use a more specific transformation of the numeric variables (e.g., quartiles instead of a median split); however, dichotomizing at the median was used in this example and those that follow to simplify the explanation of the ARGAS process.

The analysis began with a random segregation of the data into training and verification samples of equal size. The software (Magnum Opus; Webb, 2007) was configured to search for the rules with the highest lift values and to select only those that were significant ($p < .05$). This analysis identified six rules that replicated in the training and verification samples. A comparison of the results of the GLM and ARGAS analyses is presented in Table 2, below:

		Canonical Correlation					ARGAS		
Model Fit		R = .782					Six Significant Rules		
Significant Predictors		Exams Knowledge					Knowledge Exams Enrollment		

Table 2. Comparison of canonical correlation and ARGAS analysis results.

Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

Six binary rules (one antecedent and one consequent) related the antecedents to the Overall variable. High Perceived Knowledge, (1) and High Enrollment, (2) predicted a high Overall rating. Low Perceived Knowledge, (3) and Low Enrollment, (4) ratings predicted low Overall ratings. The analysis also indicated a direct relationship between Exam Quality and Teaching Competence. High Exam Quality predicted High Teaching Competence, (5). Low Exam Quality predicted low Teaching Competence, (6).

These results are generally in agreement with the canonical correlation analysis. Both analyses revealed a significant overall relationship between the predictor/outcome variables or, alternatively, the antecedent/consequent words. The canonical correlation analysis relied on the correlation between the numerical variate sets as a measure of an overall relationship, whereas the overall measure of association in the ARGAS analysis was the number of significant rules that associated the antecedent with the consequent variables. Each analysis also showed significant relationships between the Exam Quality and Perceived Knowledge predictors (antecedents) and the Overall Quality and Teaching Quality outcomes (consequents).

The results, however, differ in several ways. First, the results of the ARGAS analysis indicate that Enrollment was also a significant predictor of the consequent variables which was not identified in the canonical analysis. Second, several of the numerical variables violated the assumptions of the GLM whereas the ARGAS analysis was not constrained by these assumptions. Third, the canonical analysis assumes a linear relationship between the predictors and the outcomes whereas the ARGAS does not make this assumption.

Multiple Regression

The analyses described above are designed for canonical data structures in which multiple predictors are related to multiple outcomes. There are, however, several modifications of the canonical model that are appropriate with other types of experimental designs. For example, Multiple Regression Analysis relates multiple predictors to a single outcome. By analogy, the ARGAS procedure with multiple antecedents and a single consequent would also be appropriate for this type of analysis. The number of significant association rules between the antecedents and consequents would reflect the overall relationship among the variables. This statistic is analogous to the multiple R-square, which is the percent of variance in the outcome measure that can be accounted for by the predictors. The beta weights for each predictor in the multiple regression analysis are indices of the predictive utility of that variable. The lift values for the antecedent variables in the ARGAS analysis are indices of the value of that variable as a predictor.

The data set that was used in the canonical analysis above was reanalyzed using multiple regression procedures. The multiple regression and the corresponding ARGAS analysis included all of the antecedents except the Overall variable as predictors of the Overall

International Journal of Quantitative and Qualitative Research Methods
Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

variable. Although the multiple R-square of .755 for the regression analysis was significant, $F(5,44) = 27.184$, $p < .05$, some of the variables did not pass tests of normality. Only two of the variables, Teaching Competence and Instructor Knowledge, were significant predictors. Table 3 compares the results from the two analytic models. The ARGAS analysis began by randomly selecting half of the cases to be used as a verification sample. The remaining cases were used as a training sample. The ARGAS analysis was then performed on both data sets to determine which rules replicated. The analysis generally indicated a significant and replicable relationship between the antecedent variables and the Overall consequent variable. Three significant rules replicated in the training and verification sample.

		Multiple Regression					ARGAS		
Model Fit		Rsqr = .755					Three Significant Rules		
Significant Predicotrs		Perceived Knowledge (a) – Teaching Competence (b)					Perceived Knowledge (a) Teaching Competence (b) Enrollment (c)		
Effect Size		Beta (a) = .62 (b) = .33					Lift (a) = 1.4 (b) = 1.8 (c) = 1.6		

Table 3. Comparison of multiple regression and ARGAS data analyses results

Two significant and replicable rules derived from the multiple regression analysis revealed that the Overall evaluation was directly related to Teaching Competence, and Perceived Knowledge. The beta weights for the regression model indicated that Perceived Knowledge was a better predictor than Teaching Competence. The ARGAS analysis also identified a significant relationship between the antecedents and the consequent words however; the Lift values suggest that Teaching Competence was the best individual predictor. In addition, the ARGAS analysis indicated that that Enrollment was also a significant antecedent.

Group Comparison Analyses

The ARGAS analysis can also be used for comparing groups. As an example of this procedure, data were collected from three groups of participants who were asked to memorize a list of 12 words after hearing them presented with different procedures. One group (Rehearsal) heard each word repeated twice as the list was read to them. Another group (Control) heard the words presented once. A third group (Imagery) was asked to form a mental image of the words as they heard them presented once. Each person's recall of the words was tested immediately and again after a half hour delay. The purpose of the

Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

experiment was to assess the effect of rehearsal and mental imagery on immediate and delayed memory relative to a control condition that received neither.

With three independent groups and two dependent measures, conventional Multivariate Analysis of Variance procedures computed on these data indicated a significant overall difference among the groups Wilks' Lambda = .184, $F(4,60) = 184.5$, $p < .05$, $\eta^2 = .297$, power = 1.0. However, the assumption of equality of variance was violated for both dependent variables. Indeed, all but one of the participants in the imagery group recalled all twelve items immediately and ten of 12 after the delay interval. All but 2 participants in the Rehearsal group recalled all 12 items as well. Average recall for the Rehearsal and Imagery conditions was higher than occurred in the Control condition but with much less variability. Participants in the Imagery condition produced slightly better recall relative to the Rehearsal condition and much better recall relative to the Control.

The ARGAS analysis of the same data began with splitting the larger sample in half producing a training and verification sample. The immediate and delayed recall data were split at their respective medians and the ARGAS was applied to the words that identified the data points as either above-the median or below-the-median for that consequent variable (immediate and delayed recall). The antecedent variable consisted of words that identified group membership (Rehearsal, Control, and Imagery).

The results of an overall ARGAS analysis produced four significant rules that associated the antecedent (rehearsal, control, imagery) and consequent variables (immediate and delayed recall). Each of these rules was significant in both the training and verification data sets. These results are presented in Table 5, below:

Group		**Rule**
Control		Below Median for Immediate Recall
Control		Below Median for Delayed Recall
Imagery		Above Median for Immediate Recall
Imagery		Above Median for Delayed Recall

Table 5. ARGAS rules that predict relationships among antecedent and consequent variables

The fact that there were four significant and replicable rules indicate an overall difference among the groups, i.e., the group membership antecedent predicted the consequent variables (immediate or delayed recall). The individual rules showed that Imagery produced an immediate improvement in recall and that the effect persisted after a delay. The rules also indicated a significant decrease in recall in the control condition for both immediate and delayed recall. A separate two group analyses that compared the Imagery and Rehearsal conditions did not produce any significant and replicated rules that differentiated these conditions. The results of the ARGAS analysis indicated that mental imagery and rehearsal produced significantly better immediate and delayed recall relative to doing nothing at all.

International Journal of Quantitative and Qualitative Research Methods
Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

DISCUSSION

The purpose of the present study was to demonstrate the use of association rule modeling as a general data analytic system for hypothesis testing. In the qualitative domain, the analysis defines rules for identifying the structure of verbal information that describes a persons' unique life experience. It may therefore be especially valuable as a mixed method technique. In the quantitative domain, the ARGAS may serve as a useful alternative analysis when the distributional assumptions of the GLM are not met (Weathington, Cunningham, & Pittenger, 2010). The corpus of this paper provided examples of how these analyses would progress. We continue below with a discussion of advantage/disadvantages of the ARGAS and considerations when using it.

Implications for Research and Practice

ARGAS is not intended as a replacement for the GLM. Our purpose is to present an alternative analytic system that can be used in situations where the assumptions of the GLM are not met or where the data are text or words. Perhaps the biggest advantage of the ARGAS is that it is as a general analytic system that can be used to test hypotheses in both qualitative and quantitative research paradigms. It is also appropriate for analyzing both univariate and multivariate data. Further, because ARGAS is specifically designed for analyzing word/text data; it is ideal for qualitative, exploratory, or mixed method studies. Perhaps the biggest disadvantage of the ARGAS is that many researchers will be unfamiliar with association rule modeling. There is a dearth of literature that explains how to interpret the various statistics that derive from this type of analysis. Software for doing the analysis is obscure (e.g., SAS, SPSS Modeler, BigML.com; KH Coder download | SourceForge.net) and may require a substantial investment of time to master the packages. We therefore address several interpretative and practical issues when using the ARGAS below.

Data Preparation

Transforming numbers into words. We used the median split procedure to create the word transformations for the various numeric variables. Because there is very little published literature that uses association rule modeling with numerical data, we can only speculate about other splitting procedures that may provide better interpretation of the effects. For example, transforming the numeric variables into quartiles or even deciles may produce a clearer picture of the results. It is therefore necessary to investigate this issue in future research.

Interpretation

Overall measure of association. Our experience has been that the number of significant rules is the best index of the strength of the relationship between the antecedent and consequent variables. Even one significant rule that replicates in the verification sample, establishes a relationship although more significant rules indicate a stronger overall relationship. The

International Journal of Quantitative and Qualitative Research Methods
Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

individual rules provide information about specific associations that exist within the antecedent and consequent variable sets. Additional research will be necessary to identify how many rules constitute a small, moderate, or large effect.

Association symmetry. The ARGAS often generates *asymmetric rules* which may be difficult to interpret. For example, although a rule may associate above-the-median scores for an antecedent with above-the-median values for a consequent, the reverse relationship may not be significant. These asymmetrical rules suggest that the relationship between the variables exists but only in one end of the data distributions and not the other. Asymmetric relationships generally occur in situations where the original numeric data distributions are highly skewed.

Rule complexity. Association rules may also contain multiple significant antecedents and multiple consequents. These relationships usually occur when the variables interact with one another. For example, in a study of concerns about the COVID-19 pandemic the words senior and diabetes might significantly associate with the words hospitalization and death. Both the antecedent and consequent sides of the rule contain two items which, together, form a complex association rule.

Alternative measures of sensitivity. We have used the Lift measure as our primary index of association strength however; there are other measures such as *Support*, *Confidence*, *Leverage*, and *Coverage* that may also be useful for interpreting the results. For example, Parente & Finley, (2018) have found that the support and confidence measures were most useful. However, the lift measure is relatively easy to interpret. A value of 1.0 indicates that using the rule is no better than not using the rule. Associations with the largest Lift values are typically the most replicable.

Significance testing

Testing significance of association rules can be accomplished in several ways. For example, because the measure of lift is a ratio of proportions; Lift = (P(x & y)/P(x)P(y), one can simply compute a Z test comparing the numerator and denominator proportions. Another approach is to cross tabulate the antecedent and consequent words for significant rules filling in the frequencies for the various cells and computing a chi square statistic on the table. However, the cross tabulation procedure is, perhaps, only appropriate for simple rules. Never-the-less, cross tabulation statistics are generally familiar to most researchers. A third method for testing significance involves replication of significant rules using verification sampling.

Verification sampling. All of the ARGAS analyses reported here included a training and verification sample. Simple logic dictates that verification is desirable in any data analysis. Specifically, if an effect does not replicate with a verification sample that is randomly selected from the same population that spawned the training sample, then the reliability of the effect is questionable. We suggest generating enough data so that the data set can be split

Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

into at least a training sample and a verification sample. With the GLM, each sample should include enough data to ensure adequate power. In our experience, an adequate sample has at least twice as many cases as variables and is seldom less than 50. Whenever possible, the original sample should be subdivided into more than one verification sample to identify rules that consistently replicate.

Future Research

We do not view the ARGAS as a replacement for the GLM. Our primary purpose is to illustrate through example how association rules can be used to analyze both quantitative and qualitative research data. Although there are a number of methodological issues that remain to be sorted, we assert that the ARGAS is ready for application. Future research should focus on issues such as: power estimation and sample size, relative efficiency of measures like lift, leverage, etc., alternative scaling procedures for transforming numerical data into words, and use of ARGAS for exploratory investigations.

References

Agrawal, R., Imieliński, T., and Swami, A. (1993 June). Mining association rules between sets of items in large databases. In *Acm sigmod record* (Vol. 22, No. 2, pp. 207-216). ACM.

Agrawal, R., and Srikant, R. (1994, September).Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).

Balcázar, J. L., and Dogbey, F. (2013, October) Evaluation of association rule quality measures through feature extraction. In *International Symposium on Intelligent Data Analysis* (pp. 68-79). Springer, Berlin, Heidelberg.

Christensen, R. (2011) *Plane Answers to Complex Questions: The Theory of Linear Models,* Springer Science & Business Media.

Han, J., Pei, J., and Kamber, M. (2011) *Data mining: concepts and techniques*. Elsevier.

Hutcheson, G. D., and Sofroniou, N. (1999) *The multivariate social scientist: Introductory statistics using generalized linear models*, Sage.Thousand Oaks, CA

Imberman, S., and Domanski, B. (2001) Finding association rules from quantitative data using data booleanization. *AMCIS 2001 Proceedings*, 72.

Magalis, C. (2020) Differences in computational anxiety for students doing math versus statistics problems. Unpublished data, Towson, University.

Nelder, J. A., and Baker, R. J. (1972).*Generalized linear models,*. John Wiley & Sons, Inc..

Nesbit, R. J. and Bourme, V. J., (2018) Statistics Anxiety Rating Scale (STARS) use in psychology students: A review and analysis with an undergraduate sample. *Psychology Teaching Review*, 24(2), 101-110.

Parente, F., and Finley, J. C. (2018) Using association rules to measure subjective organization after acquired brain injury. *NeuroRehabilitation*, *42*(1), 9-15.

Parente, R.Finley, J. C., Magalis, C. (2020) A Quantitative Analysis For Qualitative

Vol.9, No.1, pp.1-13, 2021
ISSN 2056-3620(Print)
ISSN 2056-3639(Online)

Research. SAGE Advance. Preprint. https://doi.org/10.31124/advance.13238558.v1

Parente, R; Finley, J C; Magalis, C (2020) An association rule general analytic system for hypothesis testing and data mining applications, SAGE Advance Preprints. https://doi.org/10.31124/advance.7336763.v1

Suinn, R. and Winston, E. H. (2003). The math anxiety rating scale, a brief version: psychometric data. *Psychological Reports*, Feb:92(1) 167-73.

Tulving, E., (1966) Subjective organization and effects of repetition in multi-trial free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 5(2), 193-197.

Weathington, B. L., Cunningham, C. J., & Pittenger, D. J. (2010) *Research Methods for the Behavioral and Social Sciences*. John Wiley & Sons.

Webb, A. R. (2003) *Statistical Pattern Recognition*. John Wiley & Sons.