African Journal of Agricultural Research Vol. 6(2), pp. 423-431, 18 January, 2011
Available online at http://www.academicjournals.org/AJAR
DOI: 10.5897/AJAR10.621
ISSN 1991-637X ©2011 Academic Journals

Full Length Research Paper

Statistical Mix: Sequential statistical analysis approach to legitimate statistical techniques in agricultural extension, education and rural development

Iraj Malakmohammadi

Department of Agricultural Extension Education, University of Tehran, Karaj, Iran. E-mail: amalek@ut.ac.ir.

Accepted 23 September, 2010

Amalgamation of extensive theoretical models, a considerable number of variables (observed and latent), and a considerable number of uncontrollable intervening variables require behavioral (casual) researchers to have sufficient statistical reasoning and thinking competencies to structure a core 'Statistical Mix' embodied with statistical procedures appropriately to test their research hypothesis legitimately. Preliminary survey investigations and content analysis of PhD dissertations ($N_{1=}$ 65) and M.Sc. thesis (n=275 out of N_2=653) in agricultural extension education; rural development; and agricultural development in Iran showed that the applied statistical procedures were rarely eligible and appropriate. The main purpose of this article is to develop a strategic statistical road map to exploratory strategic programming, versus the common "if-then" conditional imperative programming in statistical decision trees. Hermeneutic approach is applied to study statistical thinking models and statistical devices linkages to demonstrate their applicability in educational and behavioral research. This statistical strategic road map, consisting of 50 sequential steps in four phases, is developed through an in-depth investigation of statistical resources to assist graduate and post graduate students, as well as statistics educators worldwide to develop their statistical reasoning and thinking ability and generate or come up with more realistic research outcomes.

Key words: Statistical Mix, statistical procedure, behavioral research, statistical reasoning, graduate and post graduate research.

INTRODUCTION

The main problem in conducting M.Sc. theses, Ph.D. dissertations or, even, onward academic or institutional research is that researchers either do not realize their crucial role in choosing, applying appropriate statistical tests and the process of choosing and applying the right test for the right situation (variable combination), or they are not fully aware of the coherence of required consecutive statistical procedures as a core 'Statistical Mix'. Consequently, choosing and applying the statistical tests seems to become a mere copy-and-paste process for which no initial arguments and intellectual reasoning are offered.

Wild and Pfannkuch (2002) indicated, every problem an applied statistician works on is embedded in a larger problem, the "real problem". Surrounding this larger problem is a body of "context" knowledge. Statistical investigation is used to expand the context knowledge-base. Its goal is learning in order to produce an improved context-matter understanding. Learning (in general) is much more than collecting information, it involves synthesizing the new ideas and information with existing ideas and information. The "statistical thinking" that interests us consists of the generic thinking processes which should take place where statistical methodology meets a real-world problem. This idea is even more emphasized by Ramanathan (2002) and Whitney (2005) when they stressed that "the sole purpose of statistics is to manipulate and analyze the data". Therefore, "many statistics educators and researchers today agree that there should be greater focus on the "big ideas" of statistics (that is, data, distribution, trend, variability, model, association, samples, sampling, and inference] but, after one topic has been studied, there is little mention of it again, and students fail to see how the big ideas actually provided a foundation for course content and that they underlie statistical reasoning.

Based on Wild and Pfannkuch (2002) who compared four statistical thinking models that is, JT (Jonesand, G. and Thornton, L), BF (Ben-Zvi, D. and Friedlaner, A.), WP (Wild, J. and Pfannkuch, M.), and HS (Hoerl, W. and Snee, D.), these models are intended for use as thinking tools to improve processes either within business (HS) or in the classroom (JT). The HS model shows people how to do statistical thinking and act as a guide for people to follow when they are in the process of problem solving. The JT model shows teachers the level at which their students are working in terms of their thinking capability and hence acts as an indicator for teachers to follow when planning learning task. But, as indicated by Windish and Diener-West (2006) and Govindrajulue (2004), there are a few, if any, references to the use of sequential statistics in the literature. Although choosing the right statistical test for a particular set of data appears to be an overwhelming task, to Wheater and Cook (2000), particularly if such decisions are rendered after the data are collected, what is overwhelming really, is the sequences and placements (array) of statistical tests to understand their role and mission in the first place. Wheater and Cook (2000) believe that the investigator is definitely responsible for the choice of statistical methods used. Therefore, the researcher must be able to use statistics effectively to organize, evaluate, and analyze the data (Whitney, 2005) and to apply the proper statistical tests.

To ease the dilemma, it is helpful to identify the statistical test, as stated by Hoffman (2004), which is a procedure for deciding whether an assertion (e.g., a hypothesis) about a quantitative feature of a population is true or false. There are a few cautionary steps to follow in selecting a statistical test in educational research; firstly, because of the high number of variables involved and secondly, because of the involvement of a considerable number of latent (unobserved) (Vermunt and Magidson, 2003), hidden (Moyulsky, 1995), and discrete variables. To avoid type I and type II errors, a statistical test can be applied when it is robust (that is, strong), eligible (that is, the right test in the right place), appropriate (that is, at the right position regarding the research purpose), and suitable to the research design. Also, the statistical test should be powerful (that is, the probability that a test will produce a significant difference at a given significance level).

The power of the test is equal to the probability of rejecting the null hypothesis when it is untrue that is, making the correct decision. It is 1 minus the probability of a type II error. The true differences between the populations compared, the sample size, and the significance level chosen affect the power of a statistical test. According to Dixon (2003), in whatever test you choose, it is important to think about and justify the choice. That justification can be as simple as "I did not see any complications in the data".

Watt and Berg (2002) stressed that the choice of the correct statistical test depends on the definition of the variables, particularly upon their level of measurement. It also depends on the research design used and the nature of the hypotheses: are they comparative or related; is there more than one independent variable?

There are remarkable numbers of latent variables involved in educational research along with explicit variables; many independent variables are not independent of one another in the real situation; and, more importantly, many researchers do not really know how to select and apply appropriate statistical tests to handle these variables accurately.

Watt and Berg (2002) suggest in terms of answers to the following six questions be given in order to identify the correct statistical test:

1. How many independent variables covary with the dependent variable?
2. At what level of measurement is the independent variable?
3. What is the level of measurement of the dependent variable?
4. Are the observations independent or dependent?
5. Are you comparing populations to populations, a sample to a population, or two or more samples?
6. Is the hypothesis being tested a comparison or a relationship?

Six complementary questions are proposed in this article to deal with more complex research designs as:

7. How many variables, explicit and latent, are actually involved in the research?
8. How can interfering (unwanted) variables be identified and eliminated from the study?
9. Are any premade latent variables already identified to be measured and/or being identified and measured as combined new variable/s?
10. Is there a set of more than one dependent variable being predicted from a set of more than one independent variable?
11. Are respondents or variables being grouped?
12. Is any hypothetical model being tested?

The main purpose of this article is to develop a strategic statistical road map to exploratory strategic programming, versus the common "if-then" conditional imperative programming in statistical decision trees, to sift non-relevant interfering variables through, retain sustained variables that reasonably contribute to the research pattern or model, and array of strategic statistical procedures to test research hypothesis. Eventually, to resolve the statistical anxiety, delusion, and misapplication of statistic tests among graduate and post graduate students and assist them worldwide to develop their statistical thinking and reasoning ability to generate more realistic outcomes from their professional research work.

MATERIALS AND METHODS

From the view point of pure statistics, all statistical techniques are separately valid, eligible and applicable in isolated relevant situations, to process the whole research issue but not the issue as a whole. However, from the scope of applied statistics, a 'Statistical Mix' is unquestionable to explore more fitting models, to the real situation. Especially when a complex applied casual research in techno-behavioral science that usually combines soft, semi-soft and hard technologies (such as in agricultural development among producers who are willing or not willing to apply research findings, new technologies, updated policies and strategies) is designed and implemented on the ground. However, preliminary investigation of many M.Sc. theses, Ph.D. dissertations and even research projects clearly reveals the application of popular isolated statistical techniques (as they are taught) rather than a core 'Statistical Mix' (as needed in real situation).

For the purpose of this research, a survey investigation and content analysis were employed to assess the eligibility of statistical procedures applied in considerable number of M.Sc. theses and PhD dissertations. On the other hand, reviewing the statistics literature along with the statistical thinking models led the researcher to develop an assessment tool as "a 50 steps sequential statistical analysis approach (SSAA) to Statistical Mix ", consisting of four consecutive phases as: descriptive, analytical, inferential, and modeling along with their relevant steps, firstly, to evaluate students' statistical performances (JT model) and secondly, to show their statistical thinking (HS model) throughout their thesis regarding the statistical thinking models already discussed in this article.

This 'Statistical Mix' is developed by the author based on exploratory strategic programming rather than the common "if-then" imperative conditional strategy that was, commonly, applied in developing decision trees in statistical books. Thereafter, this instrument will be implemented to evaluate the legitimacy of the statistical tests applied in Ph.D. dissertations (N_1=58) and M.Sc. thesis (n=275 out of N_2=653) from all comprehensive universities (N_3=7) offering agricultural extension education, rural development, and agricultural development majors in Iran.

An interpretive and analytical study is also conducted to provide foreground of conceptual frameworks within the context of the statistics "big ideas" (Ben-Zvi and Garfield, 2004). In addition, hermeneutic approach is applied to study statistical devices linkages and to demonstrate their applicability in educational and behavioral research to establish proposed 'Statistical Mix'.

RESULTS AND DISCUSSION

According to Tabachnick and Fidell (2007), since each statistical technique has specific assumptions, therefore, before applying any technique, or sometimes even before choosing a technique, it should be determined how the data fits some very basic assumptions underlying most of the multivariate statistics. Moreover, each statistical technique has some limitation along with its advantages. Therefore no statistical test can replace the other under the very same circumstances. For instance, while Hill and Lewicki (2007) identified multiple regression as "a seductive technique as plug in as many predictor variables as you can think of and usually at least a few of them will come out significant". Eventually, many difficulties tend to arise when there are more than five independent variables in a multiple regression equation. One of the most frequent is the problem of two or more independent variables being highly correlated to one another. This is called multicollinearity. If a correlation coefficient matrix with all the independent variables indicates correlations of 0.75 or higher, then there may be a problem with multicollinearity.

Primary investigations of M.Sc. theses and Ph.D. dissertations as supervisor, co-supervisor, external examiner, and researcher, revealed major challenges concerning selection and application of legitimate statistical techniques. This investigation showed that in majority of the cases, graduate and post graduate students are seduced (due to their lack of statistical literacy, reasoning and/or thinking) by widely applied statistic decision trees and/or statistical tables, and do not consider the authors' guidelines appropriately. Therefore, unintentionally, apply inappropriate tests.

What was found commonly neglected in graduate researches under this study is that all groups involved in their studies were taken as independent groups, while potentially, all or some of them have been dependent in their nature.

However, in some cases, as having one DV and two or more IVs with independent groups and with ordinal or interval nature of DV, suitable statistical tests are lacking. Also, when there is one interval IV and one DV with interval and nominal nature in one case, and ordinal or interval nature in another case, correlation and non-parametric correlation are recommended, respectively, but this does not sound quite right because of very rare conditions that correlation may imply causation as explained by Huck (2009).

In numerous cases of hypotheses testing when the "p-value" was significant, then the researcher usually has not cared that some other tests would give a smaller (more significant) *p*-value. If the *p*-value is not significant, then the researcher usually considers whether there is a better test (Dixon, 2003). This is often true with graduate students whose hypotheses were mostly rejected; they usually try some other statistical tests to possibly change their results (Type I error). Likewise, when the study reaches a conclusion of "no statistically significant difference", it should not necessarily be concluded that the treatment was ineffective. Otherwise, a Type II error happens, as was the case in many thesis and dissertations. Consequently, the power of the statistical tests (the probability of rejecting null hypothesis when it is false) is questioned due to the fact that for a fixed Type I error rate (α) the goal of constructing and testing a hypothesis is to maximize *Power* (Anderson-Cook and Dorai-Raj, 2003).

Regarding misapplication of statistical procedures, the following four consecutive phases were developed along with their components to build up SSAA. Each phase is composed of a few steps through which general and specific criteria for selecting and applying statistical tests are being discussed as follows.

Descriptive phase

Variable mining and measurement

This entails listing of variables involved in the study and measuring them after a scrutiny of some general research notions as: research problem and research question (Bruin, 2006; Marion, 2004); the goal of the analysis (Tabachnick and Fidell, 2007); nature of the data, research design, kind of research (Moyulsky, 1995; Wadsworth, 2005; Dinove, 2008); kind of variables (Wheater and Cook, 2000); and variable mathematical nature (nominal, ordinal, interval, or ratio) (Healey, 2005; Windish and Dinner-West, 2006; Kaminsky, 2008); and finally, number of variables (Tabachnick and Fidell, 2007).

Variable sorting out techniques

There are a few variable sorting out techniques to come up with the optimum IVs prior to involving all variables in the hypotheses-testing process. The following are some procedures that are already implemented by the author in different projects:

Reliability: Applying this technique makes it possible for a researcher to eliminate variables with low Cronbach's alpha (Ferrando, 2009); Kudar Richardson (Rudner and Schafer, 2001), and recently, ordinal Theta coefficient (Zumbo et al., 2007).

Coefficient of variability (CV): Coefficient of variation (Calvine, 2004) is recommended in this article for the purpose of consistency and accountability measurement as well as setting priority or even ranking variables. By applying CV, the researcher can select the most consistent variables with the lowest risk, and leave out the least consistent variables from the study.

Correlation matrix: Variables with statistically significant and higher correlation coefficient may be more legitimate and subject to further investigations in the research process. Therefore, variables with low or no significant correlation coefficient in the matrix can be eliminated.

Analytical phase

Variable refinery

To isolate the sensitive cases and exclude them from the main study, personal characteristics of the respondents are being tested against each one of the dependent variables (DV) (Malakmohammadi, 2008). By applying this technique, the researcher can extract as few appropriate independent variables as he/she should, due to the limited capacity of inferential statistical techniques (that is, regression, path analysis and structural equation modeling), to enhance research reliability and create favorable environment to applying appropriate statistical tests.

Variable reduction

Following the above technique and developing from what Thompson (2004) and, Tabachnick and Fidell (2007) explained about Exploratory Factor Analysis (EFA), this technique is implemented in SSAA as a converter to group numerous single variables into few "super-variables" or "factor/s", and explicit relevant discrete or continuous latent variables in the study on which the subject differs.

Latent variable measurement

Multiple regression analysis is highlighted in SSAA due to its capability to identify and measure latent variable/s in the study through a mathematical model. Of course, IVs (predictors) involved in predicting latent DVs (indicator/s) can be latent construct (factor) as the outcome of exploratory factor analysis, or simple variables. Either one of these should be specified prior to regression analysis. Notably, the scale or mathematical nature (Healey, 2005), of criterion variable is worthy of consideration in choosing the right regression model. That is, when criterion variable is nominal dichotomous; Logistic Regression (LR), when it is ordinal (discrete); Ordinal Logistic Regression (OLR) (Conne, 2006; Hilbe, 2009), and when it is quantitative (continuous); Ordinary Regression (OR) suits the model.

Although it is stated that multiple regression is a seductive technique: "plug in" as many predictor variables as you can think of and usually at least a few of them will come out significant (Statistica, 2008), but, to Palmer (MND), it is possible that the independent variables could obscure each other's effects. To prevent this situation, SSAA is considering multiple regression (in either forms), as another converter technique with dual simultaneous role to be applied after EFA. The first role deals with the limitation of regression analysis in embedding numerous variables in the equation. In this case, super-variables (the explicated latent variables through EFA) will be entered in the equation to measure variation of a latent predictor that could not be measured directly before. And, MRA, in its' second role eliminates those variables with no significant impact on the predictor variable. What remains will be utilized next in the SSAA inferential phase.

Inferential phase

As indicated by Bruin (2006), to enable one infer from his/her population data, procedures that use significance tests must be employed. Rationales behind inferential phase to help the applicants choose appropriate statistical tests are as follows.

Variables, data, and groups

Variables (independent/dependent) (Hill and Lewicki, 2007; Tabachnick and Fidell, 2007), or exogenous/ endogenous (Streiner, 2005); matched or paired data (Kaminsky, 2008); Kind of samples being compared (independent/dependent) and; number of groups being compared (one, two, or more than two).

Hypotheses testing (choosing the legitimate statistical technique)

A hypothesis is a statement that describes or explains a relationship between or among variables (Graveter and Forzano, 2008). Also, a statistical hypothesis test is defined by Lehmann and Romano (2005) as a method of making statistical decisions using experimental data. If there is no hypothesis, then there is no statistical test. P-value (Calvine, 2004); effect size (Denis, 2003; McCloskey, 2008; Graveter and Forzano, 2008); sample size (NN, 2007); central limit theorem (McDonald, 2008); number of independent hypotheses or multiple comparisons (Moyulsky, 1995; Wadsworth, 2005); paired or unpaired (Moyulsky, 1995); parametric/nonparametric (Motulsky, 1995; Dixon, 2003; Dinove, 2008; McDonald, 2008) are detected as major criteria for testing a hypothesis and considered in SSAA to choosing eligible statistical test.

Structure or model phase

To Bartholomew (1998), a model is:

1) An abstraction of the real world in which the relevant relations between the real elements are replaced by similar relations between mathematical entities.
2) A set of assumptions about the relationship between the parts of the system. Its adequacy is judged by the success with which it can predict the effects of changes in the system.

Structural Equation Modeling (SEM)

Haenlein and Kaplan (2004), referring to Gefen et al. (2000), named regression analysis as a first-generation technique, which analyzes only one layer of relationships among multiple independent and dependent variables. At the same time, they recommended SEM as a second-generation technique that allows simultaneous modeling of relationships among multiple independent and dependent constructs. Raykov and Markoulides (2006) observed that SEM enables researchers to readily develop, estimate, and test complex multivariable models as well as to study both direct and indirect effects of variables involved in a given model. The combination of direct and indirect effects makes up the total effect of an explanatory variable on a dependent variable.

Garson (2008) believes that “SEM grows out of and serves purposes similar to multiple regression, but in a more powerful way, which takes into account the modeling of interactions, nonlinearities, correlated independents, measurement error, correlated error terms, multiple latent independents each measured by multiple indicators, and one or more latent dependents, also each with multiple indicators. SEM may be used as a more powerful alternative to multiple regression, path, factor and time series analyses, as well as analysis of covariance". This technique combines factor analysis, canonical correlation, and multiple regressions to evaluate whether the model provides reasonable fit to the data and the contribution of each of the IVs to the DVs (Tabachnick and Fidell, 2007). While Garson (2008) views SEM as a confirmatory rather than an exploratory procedure, Raykov and Markoulides (2006) consider four types of SEM: path analysis model, confirmatory factor analysis model (CFA), structural regression model, and latent change model. Having the capacity of testing modeling hypotheses, SEM is installed in SSAA to develop structures or models considering the following applications of path analysis and CFA.

Path analysis (PA): To Streiner (2005), path analysis is an extension of multiple regression therefore, it goes beyond regression to allow the analysis of more complicated models. Although, despite its previous name of “causal modeling,” Streiner does not believe in path analysis as to establish causality or even to determine whether a specific model is correct; rather, it can only determine whether the data are consistent with the model. However, it is extremely powerful for examining complex models and for comparing different models to determine which one best fits the data.

To Salkind (2008), path analysis basically examines the direct relationships through the postulation of some theoretical relationships between variables and then tests to see if the direction of these relationships is substantiated by the data.

Confirmatory factor analysis CFA: CFA is usually employed to examine patterns of interrelationships among several latent constructs. According to Raykov

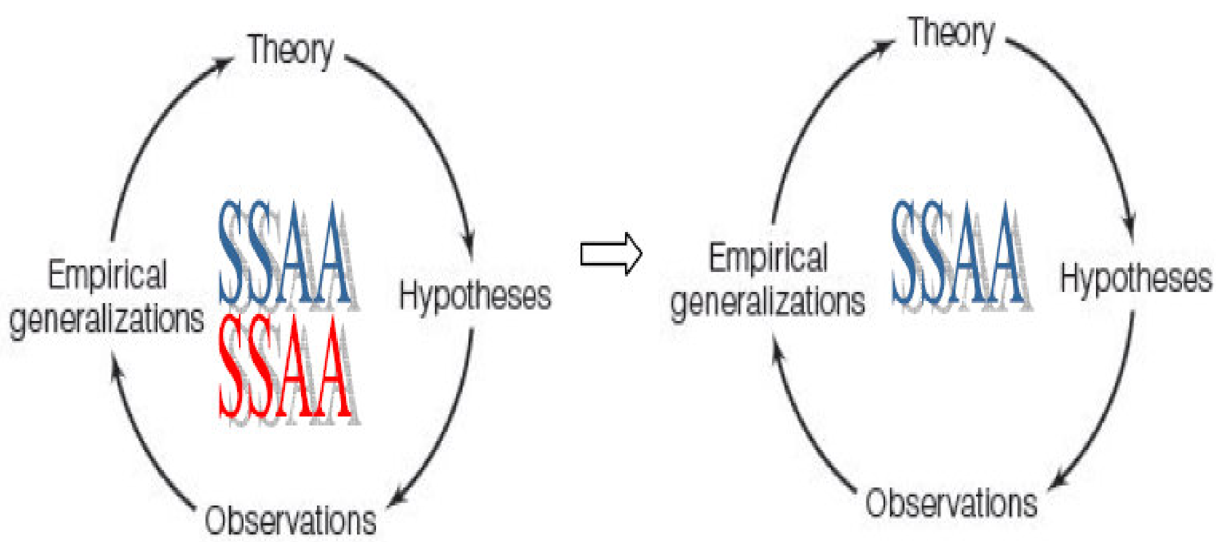

Figure 1. Wheel of science (adapted from Healey, 2005).

and Markoulides (2006), "no specific directional relationships are assumed between constructs, only that they are potentially correlated to one another". The starting point of CFA is a very demanding one, requiring that the complete details of a proposed model be specified before it is fitted to the data." The latter statement by Raykov and Maroulides was more clearly explained by Stapleton (1997) when he described CFA as "a theory-testing model as opposed to a theory-generating method like EFC.

Conclusion

Concerning the obstacles in classic methods of teaching, learning and applying statistics, this is an evidence-based policy making article to provide 'Statistical Mix' as a practical guide for graduate and post graduate students and researchers in general and researchers of casual (techno-scientific applied and developmental research) that is, agricultural and environmental development. Accordingly, conclusions drawn from what is already discussed and applied in this article, mainly are dealing with the hermeneutic and not that much with the data-based stage of the original research, although, fully data supported research article will be revealed shortly after this. These conclusions are classified in two categories as follows.

The wheel of science

Bases on the findings and misapplications of statistical tests in MSc. theses and PhD dissertations and also taking off from what Healey (2005) has revealed about the wheel of science being made of four initial parts (theory, hypotheses, observations, and empirical generalization) and using these to respond to an aforementioned problem, the author prepared this analytical article by focusing on the proposed 'Statistical Mix' to assist researchers in the social, educational, and behavioral fields to, first, purify and, secondly, promote their research findings. As shown in Figure 1, while SSAA is central to the model to manage the circulation of the wheel of science, a qualitative or quantitative variable refining process is done at every two initial stages of the science wheel to identify the appropriate and eliminate those invalid variables that are not even directly measurable (latent variables) in the process of generating knowledge. Another aim is to identify variables that have undesirable impacts or effects on the process and modify these according to certain criterion.

Statistical mix: Sequential statistical analysis approach (SSAA)

The bottom line of a realistic and truly applicable causal research is a justifiable model. The model would not be rationally justified without going through SSAA, that is, firstly, a variable is sustained in the SEM process when it is found to be valid (justified to be measured), consistent (very low coefficient of variability), and accountable (eligible to involve in modeling process) through the VRP (Malakmohammadi, 2008).

Complimenting what is already provided in highly consumed statistical decision trees which are not actually concerned in sequence or hierarchical placement of statistical tests, Table 1 shows SSAA as an additive "sorting-out technique" to come up with much more useful subset of the IVs which "best" explains the DVs. Firstly, this will diminish misconceptions among young

Table 1. Statistical Mix: Sequential statistical analysis approach (SSAA).

A- Initial phase

Variable mining and measurement process

1- Identify research problem/s.
2- Specify research question/s.
3- Articulate research objective/s.
4- Review related research and literature (RRRL).
5- Provide theoretical contingency table (TCT) to show resources, issues, and their frequencies.
6- Select high frequency issues (HFI) in TCT.
7- Design theoretical framework (TF) embodying HFIs and their realized relations.
8- Configure specific research method and materials (RMM) to investigate and test TF.
9- Construct research instrument to collect data.

Variable reduction (refinery)

10- Look at the validity of a measure (Turpin, 2004).
11-Test the reliability of RI (pilot research instrument). Check the reliability of a measure (Turpin, 2004), calculate Cronbach's α coefficient.
12- Eliminate items with low reliability coefficient, if applicable, from RI.
13- Define research population (RP) and sampling procedure.
14- Collect the data (Statistica, 2008).
15- Assess each variable separately first (obtain measures of central tendency and dispersion; frequency distributions; graphs); is the variable normally distributed? (Statistica, 2008).
16- Calculate the variables' coefficient of variability (CV).
17- Eliminate variables with higher CV, if applicable.
18- Identify sustained variables in the study.
19- Develop a correlation matrix .
20- Eliminate variables with no significant correlation coefficient, if applicable.
21- State refinery hypotheses (test variables against sample's characteristics).
22- Eliminate variables highly affected by sample characteristics, if applicable.

Variable or respondent grouping process

23- Apply R-type Exploratory Factor Analysis (EFA) (either orthogonal type if factors are not correlated, or oblique type if factors are correlated) to find out hypothetical factors (latent variable/s) and variables capable of building factor/s (Thompson, 2004).
24- Eliminate variables with lower than 1 eigen value.
25- Identify new grouped (factor/s) variables (basically latent).
26- Compare factor analysis output with theoretical model to identify compatible variables.
27- Design conceptual model by compatible variables.

B- Intermediate (inferential) phase

- Hypotheses development

29- State the research hypothesis (RH) (Statistica, 2008).
30- State the null hypothesis (NH) (Statistica, 2008).
31- Assess the relationship of each independent variable, one at a time, with the dependent variable (calculate the correlation coefficient; obtain a scatter plot); are the two variables linearly related (Turpin, 2004; Statistica, 2008)?are responding groups independent?

Variable and group identification

32- Identify variables' nature (scale) and role (IV/DV) and groups' essence (two/more than two and independent/dependent).

Hypotheses testing (choosing appropriate statistical test)

33- Choose specific P value/s to test the null hypotheses by appropriate statistical tests
Choose appropriate statistical test for prediction and/or comparison based on the information at http://www.ats.ucla.edu/stat/stata/ado/analysis/ 1, corresponding to the null hypotheses in the study.
34- Test the null hypotheses.

Table 1. Count'd

C- Advanced (Modeling) phase
Regression (multiple and multivariable)
35- Design regression analysis (multiple and/or multiple) to measure indicator/s and latent variable/s.
36- Identify independent variable/s and dependent variable in the hypothetical multiple regression equation to choose the right regression model ((that is, Logistic Regression for nominal dichotomous DV, Ordinal Logistic Regression for ordinal or discrete DV (Conne, 2006; Hilbe, 2009), and Ordinary Regression basically for continuous DV)).
37- Calculate appropriate statistic suitable to the regression model (that is, F for quantitative regression model) to realize the significance of the equation as a whole.
38- Eliminate equation/s with no significant F value (for the whole regression equation).
39- Calculate and examine appropriate measures of association and tests of statistical significance for each coefficient (Statistica 2008).
40- Eliminate predictors with no significant R value (when F value for the equation is significant).
41- Regress each explanatory variable against a constant and the remaining explanatory variables. There should be k–1 values for VIF. If any of them is high, then MC is indicated. It can be concluded that the higher VIF or the lower the tolerance index, the higher the variance of $\hat{\beta}$, and the greater the chance of finding βi insignificant, which means that severe MC effects are present. A general rule is that the VIF should not exceed 10 (Belsley et al., 1980). MC might still be present and hence the next step is to regress each explanatory variable against all the other right hand side variables and compute the tolerance ($1–R^2$) and VIF (Ramanathan, 2002; Ramathan, 2008).
42- Reject or accept the research hypothesis (Turpin, 2004).
43- Eliminate variables with insignificant coefficients, but one at a time to find the superior model (Ramanathan, 2008).
Structural Equation Modeling (SEM)
44- Apply SEM utilizing EQS, LISREL or M*plus* to figure out the final contingency framework (Raykov and Markoulides, 2006).
Path analysis
45- Explain the practical implications of findings for further investigation through path analysis as threshold for SEM.
46- Apply path analysis (PA) and draw the path diagram (casual model).
47-Compare the PA outcome with conceptual research framework to argue challenges.
Confirmatory Factor Analysis (CFA)
48- Identify and explain endogenous and exogenous variables.
49- State the SEM hypotheses.
50- Apply confirmatory factor analysis (CFA) (Thompson, 2004) to test the revised hypothetical model (theoretical model or final path analysis model) and revise the modeling hypothesis based on the CFA outcome (if necessary) for the closest possible arrangements to the real situation and make a final decision about the research contingency model.

researchers and graduate students who are after accurate application of statistical methods, and secondly, this will lower their stat phobia by leading them towards a 50-strategic-sequential-statistic-roadmap for choosing and applying appropriate statistical tests, interpreting their findings, and implementing scientific analysis more realistically in a creative research enterprise.

Notably, although the goal in EFA is to represent those things that are related to one another by a more general term such as a factor (Salkind 2008), or to drive just a few responses underlying structure called "factors" (Tabachnick and Fidell, 2007), it is assigned a dual role technique in SSAA to discover latent variables (factors) or "super-variables" in the study from one hand, and to eliminate variables that do not significantly contribute (correlate) to building up super-variable from the other hand.

Moreover, multiple regression analysis (MRA), in either forms of conventional, logistic, and ordinal logistic, is considered in SSAA as another statistical technique with dual simultaneous role to play after EFA. The first role deals with the limitation of regression analysis in feeding numerous variables in the equation. In this case, factors (the explicated latent variables through EFA), are being entered in the equation along with other explicit variables to measure variation of a latent predictor that could not be measured directly before. MRA, in its' second role to play, eliminates those variables with no significant impact on the predictor variable.

It can be concluded, however, that path analysis, sequentially, infrastructures, rather than prerequisite, SEM, as does regression (multiple and multi-variable) analysis to path analysis.

Taking advantage of the model development approach, researchers can test the hypothesis of factors and factor loading through confirmatory factor analysis (CFA) as the threshold stage for SEM as indicated in Table 1.

REFERENCES

Anderson-Cook, M, Dorai-Raj S (2003). Making the Concepts of Power and Sample Size Relevant and Accessible to Students in Introductory Statistics Courses using Applets. J. Statistics Educ., [Online], 11(3).

Bartholomew J (1998). Social Process. London School of Economics and Political Science. Data and Sources of Collinearity. New York: John Wiley & Sons.

Belsley DA, Kuh E, Welsch RE (1980). Regression Diagnostics. New York: Wiley.

Ben-Zvi D, Garfield J (2004). The Challenge of Developing Statistical Literacy, Reasoning and Thinking. Springer Science + Business Media, Inc. Netherlands. 978-1-4020-2277-7

Bruin J (2006). Newtest: Command to Compute New Test. UCLA: Academic Technology Services, Statistical Consulting Group. http://www.ats.ucla.edu/stat/stata/ado/analysis/.

Calvin D (2004). Choosing and Using Statistics, a Biologists Guide. Blackwell Publishing Company. MA. USA.

Conne A (2006). Logistic Regression Models for Ordinal Response Variables. Sage Publications Inc. Thousand Oaks, Cl, USA., Pp. 10-25

Denis D (2003). Alternatives to Null Hypothesis Significance Testing. J. Theory. Sci., 4-1 The International Consortium for the Advancement of Academic Publication, Athabasca, Canada.

Dinove I (2008). Choose the Right Test. From http://www..stat.ucla.edu Accessed in May 2009.

Dixon P (2003). Choosing Statistical Methods. Iowa State University, USA. From http://www.bcb.iastate.edu/faculty/Dixon_CV.pdf. Accessed in March 2009.

Ferrando P (2009). A General Factor-Analytic Procedure for Assessing Response Bias in Questionnaire Measures. Structural Equation Modeling: A Multidisciplinary J., 16(2):364–381. DOI: 10.1080/10705510902751374

Garson D (2008). 'Structural Equation Modeling' from Statnotes: Topics in Multivariate Analysis. North Carolina State University, Retrieved May 25, 2009.http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Govindrajulue Z (2004). Sequential Statistics. World Scientific Publishing Inc. NJ, USA. p. 1

Graveter F, Forzano L (2008). Research Methods for the Behavioral Science. Third ed. PP:19-22. Thomson/Wadsworth, ISBN0495091456, 9780495091455.

Haenlein M, Kaplan A (2004). A Beginner's Guide to Partial Least Squares *Analysis*. Lawrence Erlbaum Associates Inc. Understanding Stat., 3(4): 283-297.

Healey J (2005). A Tool for Social Research. 7th ed. Wasdworth, Thomson Learning Academic Resource Center, CA, USA.

Hill T, Lewicki P (2007). Statistics Methods and Applications. StatSoft, Tusla, OK.

Hilbe J (2009) Logistic Regression Models. Chapman & Hall/CRC Press, Texts in Statistical Science. NY.USA. PP: 353-3 Hill, T. & Lewicki, P. (2007). STATISTICS Methods and Applications. StatSoft, Tulsa, OK. USA.

Hoffman H (2004). Internet Glossary of Statistical Terms in: Statistics Explained . www.animatedsoftware.com/statglos/statglos.htm)

Huck S (2009). Statistical Misconceptions. Taylor and Francis Group, NY, USA. pp. 43.

Kaminsky J (2008). Qualitative and Quantitative Analysis. Kwantlen Polytechnic University. BNS Nursing Program. Greater Vancouver, British Columbia, Canada.

Lehmann E, Romano J (2005). Testing Statistical Hypotheses (3rd ed.). New York: Springer. ISBN 0387988645. www.virtualcurriculum.com/N4120/LA9_Figure8.pdf. Accessed on March 2009. http://www.ats.ucla.edu/stat/sas/notes2/ Accessed in March, 2009.

Malakmohammadi I (2008). Variables Refinery Process to Ensure Research Unbiasedness (Validity) and Invariance (Reliability) in Agricultural Extension and Education. Am. J. Agric, Biolo. Sci., 3(1): 342-347.

Marion R (2004). The Whole World of Deduction Research Skills for New Scientists. The University of Texas Medical Branch, Texas, USA.

McDonald JH (2008). Handbook of Biological Statistics, Choosing a Statistical Test. Sparky House Publishing, Baltimore, Maryland. P.282-287.

Moyulsky H (1995). Intuitive Biostatics: Choosing a Statistical Test. Oxford University Press, Inc.

Motulsky H (1999). Analyzing Data with GraphPad Prism. GraphPad Software Inc., San Diego, CA.

McCloskey D (2008). The Cult of Statistical Significance. Ann Arbor: University of Michigan Press. ISBN 0472050079.

NN (2007). Statistical Test: More Complicated Discriminate. PHY310: Lecture 14. http://nngroup.physics.sunysb.edu/~mcgrew/phy310/lectures/phy310-lecture-14-2007.pdfnngroup.physics.sunysb.edu/~mcgrew/phy310/lectures/phy310-lecture-14-2007.pdf

Palmer MND (2009). Multiple Regression. Ordination Method for Ecologists. Oklahoma State University. http://ordination.okstate.edu/MULTIPLE.htm. Accessed Feb. 08,

Ramanathan R (2002). Introductory Econometrics with application. Fifth ed. Harcourt College Publishers. SENGAGE Learning. Florence, KY.

Ramanathan R (2008). More on Multicollinearity (MC) Variance Inflation Factor (VIF) http://www.econ.ucsd.edu/~rramanat/MoreonMC.pdf (accessed March 2008).

Raykov T, Marcoulides G (2006). A First Course in Structural Equation Modeling. Lawrence Erlbaum Association Inc. Publishers. NJ, USA. pp. 117-118.

Rudner LM, Schafer WD (2001). Reliability. ERIC Digest. Online http://www.ericdigests.org/2002-2/reliability.htm

Salkind N (2008). Statistics for People Who (Think They) Hate Statistics. Excel 2007 Edition. Sage Publication Inc. CA, USA. pp. 326

Stapleton C (1997). Basic Concepts and Procedures of Confirmatory Factor Analysis. Paper presented at the Annual Meeting of the Southwest Educational Research Association. Austin, TX, January 23-25.

Statistica (2008). Multiple Regression. Electronic Textbook. Statsoft, Inc. http://www.statsoft.com/textbook/stmulreg.html

Streiner D (2005). Research Methods in Psychiatry. Finding Our Way: An Introduction to Path Analysis. The Canadian J. Psych., 50: 115–122

Tabachnick B, Fidell L (2007). Using Multivariate Statistics. 5th Ed. Pearson. NJ. USA. ISBN-13: 9780205459384

Thompson B (2004). Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications. Washington, D.C.: American Psychological Association.

Turpin G (2004). Which test? A Clinical Psychologist's Online Guide to Selecting a Statistical Test. Center for Applied Social and Psychological Development. Canterbury Christ Church, UK.

Vermunt J, Magidson J (2003). Statistical Innovation. Encyclopedia of Social Science Research Methods, Sage Publications.

Wadsworth (2005). Choosing the Correct Statistical Test. Wadsworth Cengage Learning. Florence, KY. http://www.wadsworth.com/psychology_d/templates/student_resources/workshops/stat_workshp/chose_stat/chose_stat_01.html

Watt J, Berg S (2002). Research Methods for Communication Science. Amazon.com.

Wheater P, Cook P (2000). Using Statistics to Understand the Environment. Routledge, NY, USA.

Whitney (2005). Statistics: A Tool for Social Research. 7th Ed. Thomson Learning, Inc. Thomson Wadsworth. Belmont, CA 94002-3098, USA.

Windish D, Diener-West M (2006). A Clinician-educator's Roadmap to Choosing and Interpreting Statistical Tests. J. Gen. Int. Med., 21 (9, 8), 1009-1009. DOI 10.1111/j.1525-1497.2006.00390.x

Wild C, Pfannkuch M (2002). Statistical Thinking Models. The Sixth International Conference of Teaching Statistics. Cape Town, South Africa 7 – 12 July.

Zumbo B, Gadermann A, Zeisser C (2007). Ordinal Versions of Coefficient Alpha and Theta for Likert Scale. J. Modern Appl. Stat. Methods, 6(1): 21-29.