SOLVING THE STATISTICAL TEST SELECTION PROBLEM IN HYPOTHESIS TESTING

Kien-Kheng, Foo[1], Nuralina Azlan[2], Noraini Idris[3] and Siti Noor Dina Ahmad[4]

[1] *Institute of Graduate Studies, University of Malaya, Jalan Universiti, 50603, Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur*

[2] *Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Cawangan Negeri Sembilan, 70300, Kampus Seremban 3, Negeri Sembilan*

[3] *Pejabat Timbalan Naib Canselor (Penyelidikan & Inovasi), Universiti Pendidikan Sultan Idris, 35900, Tanjong Malim, Perak*

[4] *Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA Cawangan Negeri Sembilan, 72000, Kampus Kuala Pilah, Negeri Sembilan*

e-mail: kienkheng@yahoo.com
e-mail: noridris@upsi.edu.my

Abstract

Hypothesis testing is widely regarded as a valuable tool for data analysis but it is conceptually difficult to understand and even more difficult to carry out the procedure. Even so, interpretation of the results is often littered with mistakes and misuses. This paper proposes an approach through the use of STestMAP. It is a tool resembling a concept map designed to assist researchers in identifying relevant univariate statistical tests to match their research objectives. An earlier study in the efficacy of this approach as a tool to acquire selection skills has shown STestMAP provides a flexible structure in assisting users to successfully identify the statistical test but further research is needed to confirm the usefulness of STestMAP.

Keywords*: Statistical Tests, Test Selection Problem, Hypothesis Testing*.

1. Introduction

Null Hypothesis Significance Testing (NHST) or better known as Hypothesis Testing is one of the many quantitative analytic tools available to researchers today. It is both versatile and reliable when used correctly but if used inappropriately it can misinterpret the findings. There is a lot of subjective judgment involved in statistical inference which is why statistical thinking and reasoning are of utmost importance for assisting in making relevant interpretations and decisions. One of the greatest statisticians of our time Tukey had this to say, "If data analysis is to be well-done, much of it must be a matter of judgment, and 'theory', whether statistical or non-statistical, will have to guide, not command" (Tukey, 1962, pg10). Good judgment is dependent upon strong statistical foundation, experience and practice. All these factors compounded the difficulties young researchers are facing presently. Furthermore, misconceptions arising from weak knowledge and 'incomplete' understanding of basic statistical concepts and their inter-relationships put up more barriers, hence making NHST difficult to handle (Gardner and Hudson, 1999; Quilici and Mayer, 1996).

Anderson-Cook and Dorai Raj (2003) reiterated that null hypothesis testing is an important topic in statistics from the conceptual point of view but students often found difficulties in the choosing the correct tests, understanding the correct procedures and being able to calculate as well as interpret the critical value or p-value. Without a good grounding in the basics, hypothesis testing will just be an act of memorizing facts with shallow and isolated understanding denying itself of knowing the important relationships that exist between them

(Nayak and Hazra, 2011; delMas, Garfield and Chance, 1999). Chance and Garfield (2002) commented that teaching to answer individual and isolated problems as class exercise does not serve the purpose of meaningful understanding. Instructions should encourage students to look at the structure of the hypothesis testing process as a whole or in other words, see the big picture. Commenting on statistical inference their study found that students learn to apply procedures when directed but then after the course, are lost as to where to begin when presented with novel questions. It seems that students may have learn how to use different kinds of statistical tests but a major concern is to learn when those procedures can be used (Gardner and Hudson, 1999; Quilici and Mayer, 1996; Schau and Mattern, 1999; Nayak and Hazra, 2011; Regoniel, 2015).

What needed to be done now is a tangible effort on the part of statistics educators to find ways in helping learners to see the coherent picture rather than bits of information here and there. Without a statistical scaffold in place to help them construct the required knowledge, NHST would be at best, just a set of rules to follow.

2. Researchers' Difficulties with NHST Procedure

What makes test selection difficult originates from the complex interplay of reasoning, thinking and decision making students have to do concurrently as they move from one step to the other of the NHST Procedure (Foo, 2011; Regoniel, 2015). Being well-organized in their thinking process would be a great help. For example, students must start off the procedure by firstly deciding on two important questions: what type of analysis to use and what type of data has been collected.

Next they must be able to determine the purpose/s of their research. In a simple univariate or bivariate analysis alone, these purposes can be roughly categorized as:

- Describing a single population.
- Comparing two populations.
- Comparing more than two populations.
- Analyzing the relationship between two populations.
- Analyzing the relationship for more than two populations.

Subsequently students are required to identify the type of data that has been collected. Data are classified as quantitative data (scale, interval, ratio), ranked data (ordinal) and qualitative data (categorical measures, nominal). All the parameters to be tested in hypothesis testing, is determined based on this information. Henceforth the sample characteristics like the number of groups to be compared, their sampling distribution and the issue of independence of observations must be addressed accordingly. When the sampling distribution is known, probability can be calculated but for sampling distribution that is unknown then different test procedure (distribution-free tests) should be used (Foo, 2011).

Apart from what has been discussed earlier, it is important to bear in mind that not all statistical tests are robust to violations of the assumptions. These assumptions are important as many statistical methods/techniques make certain assumptions about the type of data to be used and how they are collected. The common violations that many are guilty of include not checking for normality, homogeneity of variances, independence and sample size to name a few. Thus if the generated output of a statistical analysis is to be considered correct, the assumptions have to be (Nayak and Hazra, 2011; Sheskin, 2001). Take for example the issue of normality. Normally-distributed data are very rare. For most statistical test procedures it is enough if the data are approximately normal. When the sample size is large, one should not look at only the assumption tests for normality but also at the actual departure from normality before deciding whether the data is normal or otherwise. After careful deliberations on these issues only then will one be confident of selecting an appropriate statistical test that matches the objective of the study.

3. Finding an Approach to Statistical Test Selection Procedure

It is relevant to begin by appreciating the NHST procedure in selecting an appropriate statistical test. To get a feel of this multifaceted routine, one can use the ubiquitous 6-step procedure for illustration purposes as illustrated below.

Step 1: Write down the null and alternative hypotheses
Step 2: Set the level of significance (alpha) after giving considerations to the Type II error.
Step 3: Choose and calculate the test statistics
Step 4: Find the critical region
Step 5: Make a decision to reject or not to reject the null hypothesis
Step 6: Interpret the result with reference to the research question.

Each of these steps is problematic to a differing degree. Some are easier to surmount as compared to the others. Steps 1, 2, 4 and 5 are much easier to master and with the advent of user-friendly statistical packages many of the mechanical processes utilized in these steps can be handled well by statistical packages like MS EXCEL, SPSS and Minitab (StatPac, 2010; Field, 2013). The more challenging processes are found in Step 3 and 6 where they require a fairly complex interplay of understanding, reasoning, decision making and sound judgment. Step 6 is littered with many hurdles if the researchers carry with them misconceptions related to the interpretation of p-value, the fallacies of proving a hypothesis and replicability issues (Foo, 2011). In this review the authors examine step 3 more closely as this concerns the choice of the test statistics in relation to the objectives of the research.

Mastery of the selection skill is very much influenced by the mastery of basic statistical selection concepts and interrelationships inherent in the hypothesis testing process. This implies a need to have both instrumental and relational understanding (Skemp, 1971). Getting a good grip of the selection procedure that is 'knowing the how', would be an important step towards 'knowing the how and the why'.

For this purpose this paper seeks to review an approach called STestMAP procedure that is designed to assist beginner researchers with selecting univariate statistical tests when doing hypothesis testing. The proposed solution to mastering the NHST routine is through the usage of a 'network' mapping called STestMAP- A Roadmap to Statistical Tests.

4. Development of STestMAP (Statistical Test through Multimedia Procedure)

A proposed solution to the selection problem is to use a mapping technique called STestMAP (Foo, Leong and Lau, 2005). This is a visual tool for seeing the concepts and relationships in the context of the statistical test selection procedure. The mapping model is envisaged as a cognitive network of related ideas. STestMAP consists of 6 roadmaps much like concept maps impressed upon a tabular template (see Figure 1). Each of these maps consists of a network of concepts and items that the user has to first of all understand and then make appropriate choices to steer their way through the different paths being mapped out.

The improvements and enhancement of STestMAP have been progressively carried out through the feedback from participants of many STestMAP workshops, presentations and seminars throughout the years from 2006 till 2011.

In the process of learning inferential statistics, a complete expert map may enhance the ability to apply statistics by giving a 'whole picture' and providing an overview of all the decisions that must be made in choosing an appropriate test (Gardner and Hudson, 1999; Schau and Mattern, 1997). Notwithstanding its weaknesses, the mapping model possesses favorable qualities that encourage both instrumental and relational understanding.

STESTMAP

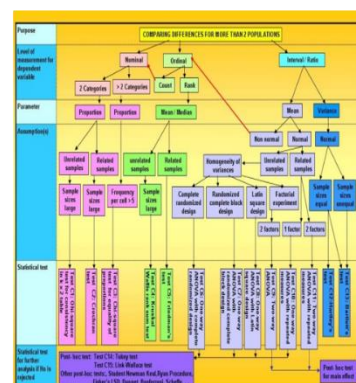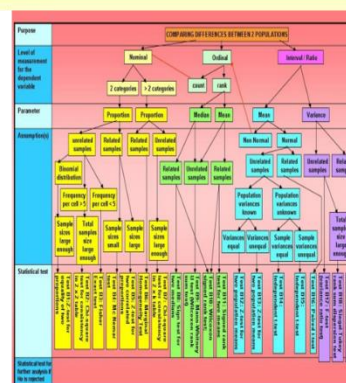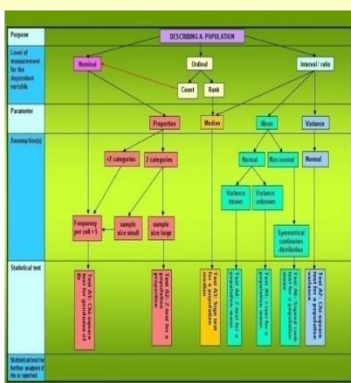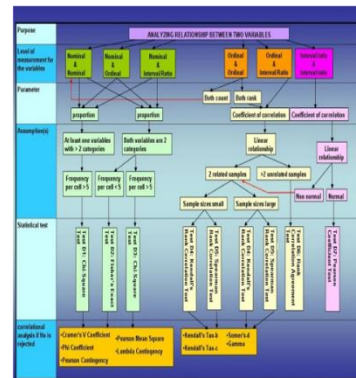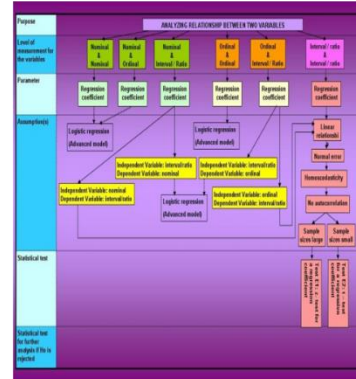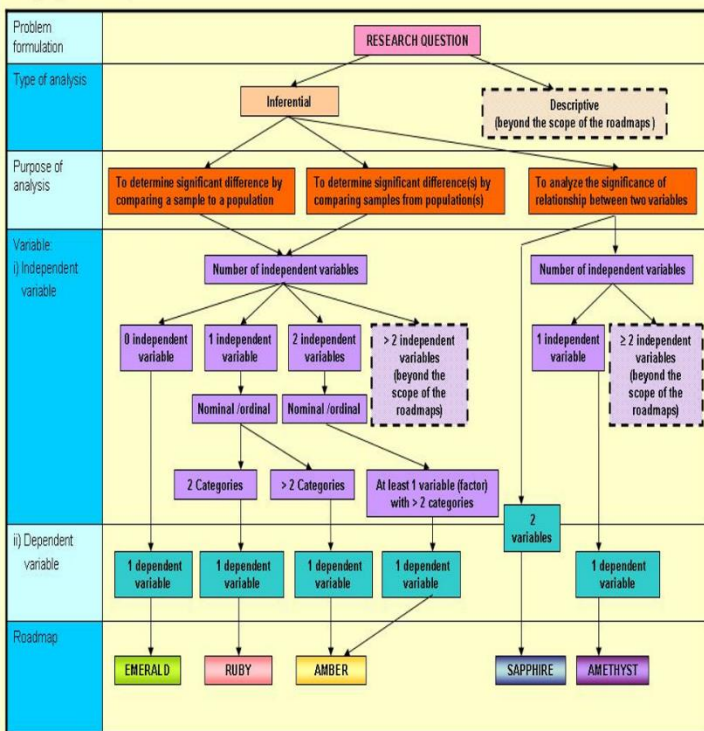Figure 1: The Six Main Maps with 49 commonly-used univariate and bivariate Statistical Tests. (Foo, Leong and Lau, 2005).

An illustration using a case study on how STestMAP can be used in choosing the correct test should suffice to see how this approach works.

The Consumer Association in a certain city was concerned about the accuracy of the information given on the bottle label produced by a certain mineral water bottling plant. The label stated "net weight of 150ml". A random sample of 30 bottles was selected and the volume of the contents measured. Can the association conclude, based on the data shown below, that the product label was misleading?

Table 1: Content of mineral water measured in ml.

151	153	149	147	152	155
148	149	151	153	154	147
150	146	152	155	156	148
149	147	152	159	153	155
144	149	150	153	152	151

Table 1 shows an example as it is commonly found in many introductory statistics books. In a normal class situation the 'hint' for the choice of a statistical test depends on which topic is being taught. If a student who has just finished learning about one-sample tests and this question is given as a practice, he/she will naturally select tests related to that section. This is actually an unrealistic situation. In doing research, researchers are not given such clear-cut classes of tests to choose from. They do not realize the difficulty of identifying the relevant tests till they embark on a study where the choice of the tests depends very much on the objectives of that study. In this respect, STestMAP provides an alternative to looking for such a test based on the research proposal. For this particular example, the flow of the NHST procedure starts with the Bridging Roadmap.

5. Bridging Roadmap

The procedure in STestMAP starts with the Bridging Map. This particular map functions as a bridge between the researcher's purpose and objectives to the maps where the relevant tests are located. Study the Bridging Map (see Figure 2) closely and familiarize with the different concepts, relationships and routes. Notice the flowchart is superimposed onto a table in an organized structure. The map has been designed to be read both vertically and horizontally. This main map is organized under six categories of terms i.e. Problem formulation, type of analysis, purpose of analysis, variable and roadmap. If one were to read horizontally one will find all the related terms under this particular category. Take for example: Type of analysis. This route asks the user to determine if the research objective is dealing with inferential statistics or descriptive statistics. In a way users will learn that when doing statistical analysis they need to know if they are dealing with inferencing or just merely describing their data set. By reading the map vertically, it starts with the term 'Inferential' and it flows down to other concepts such as purpose and variable linking them in a coherent manner leading to the identification of the next map to use.
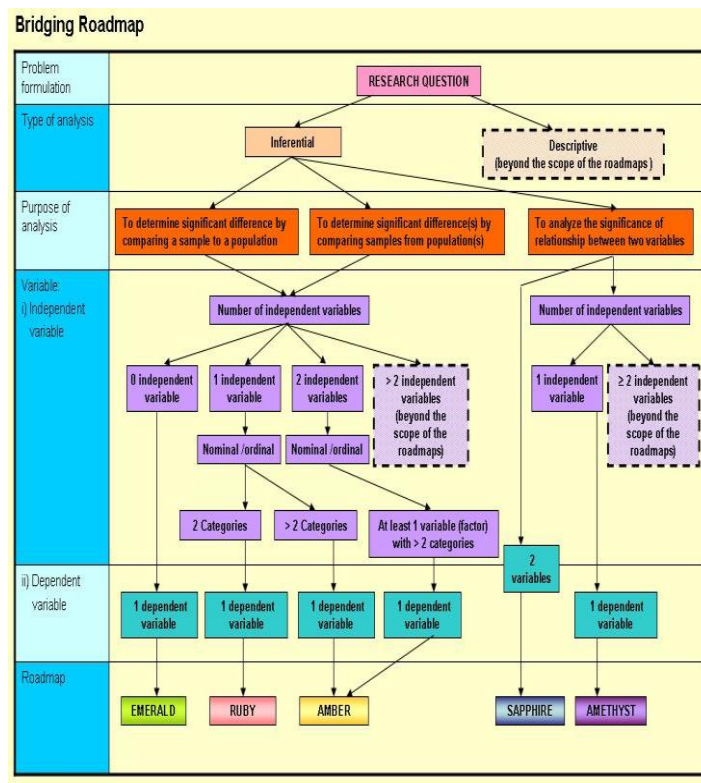

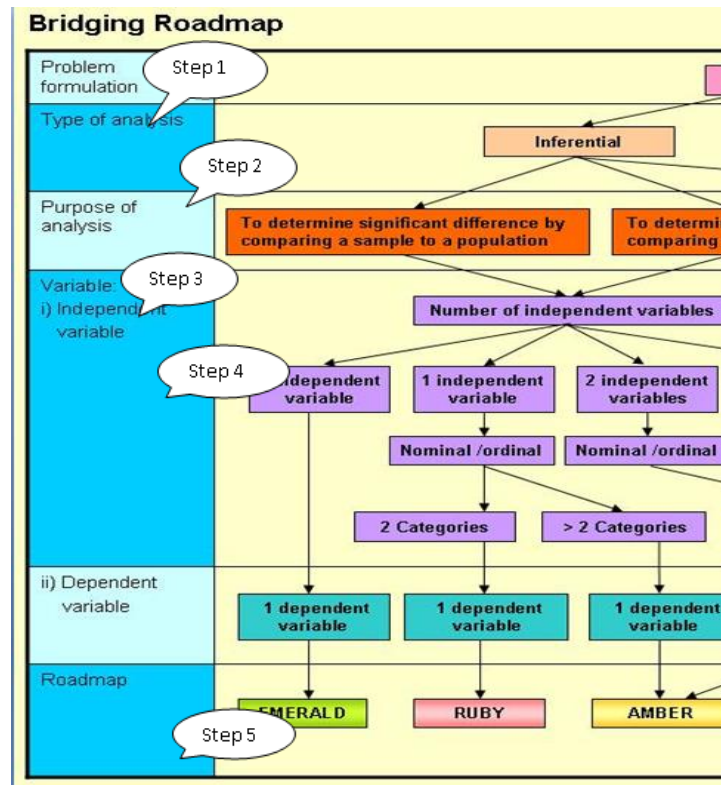Figure 2: Bridging Map (Foo, Leong and Lau, 2005)

Figure 3: The Procedure of Selection (Foo, Leong and Lau, 2005).

With reference to the case study, the search starts by focusing on the left column (Figure 3) that provides the guide for the users to get around this map.

Problem Formulation: The focus here is about the research problem and its objective. In this case study, from the information given, it is obvious that the objective of this research is to test whether the average volume of the contents is 150ml or not? To do that, two hypotheses are formulated i.e. null and alternative hypotheses (see Step 1 of Figure 3).

$H_0: \mu = 150$
$H_1: \mu < 150$

Type of Analysis: Is inferential statistics or descriptive statistics appropriate to be used to address the objective? From the objective, this analysis calls for 'inferential statistics' (see Step 2 of Figure 3).

Purpose of analysis: This step is concerned with the choice between 'Describing a population' or 'comparing between 2 or more populations'. As there is only one population of interest (i.e., all the mineral water bottles produced by this company at the time of inspection) the user can conclude that the purpose is 'to describe a population' (see Step 3 of Figure 3).

Variable: This part requires the user to identify the variable in the research question, how many of them and differentiate them into independent or dependent variable (s). As this map caters for univariate analysis, the search for the appropriate tests is limited to one dependent variable and at most 2 independent variables. In this case study, there is no independent variable for there is no need split up the population of bottles into subpopulations for comparison. However there is only one dependent variable though (i.e. "the content of mineral water bottle") (see Step 4 of Figure 3).
  - Independent variable - 0

- Dependent variable - 1

Following the 5-step path in Figure 3, it leads to the 'Emerald' roadmap (see Step 5 of Figure 3).

Next step involves the usage of the Emerald roadmap. This map requires a few more steps to determine the kind of statistical test to use.
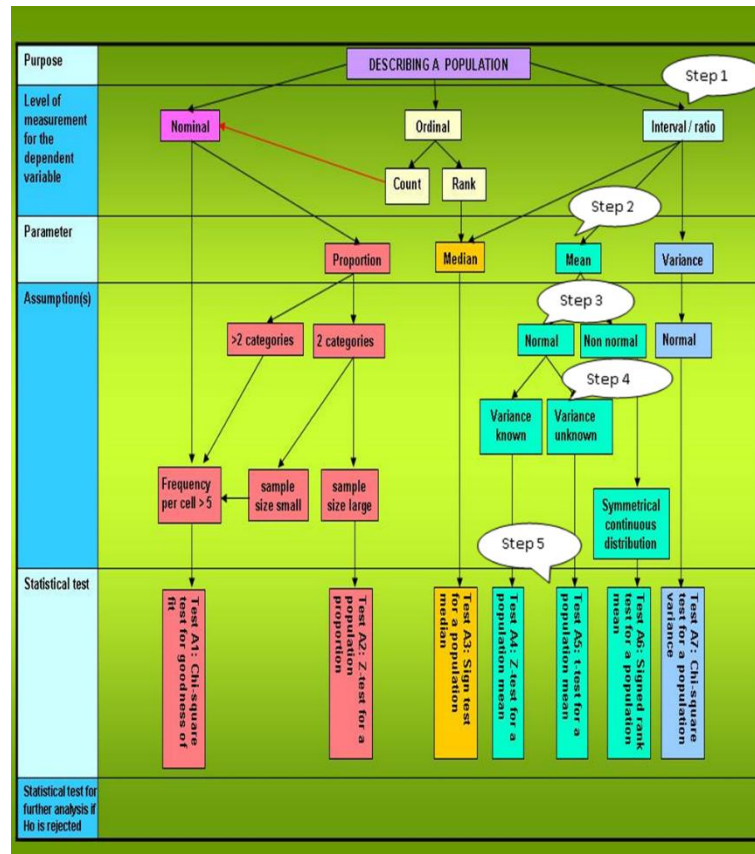

Figure 4. Emerald roadmap (Foo, Leong and Lau, 2005) for ' Describing a Population'.

Purpose: Previous steps had helped to establish that the case study is about 'Describing a population'. This map is suitable when there is only 0/1 independent variable and 1 dependent variable.

Level of Measurement for the dependent variable: Notice that the dependent variable is 'content of the bottle'. This is measured in milliliter and it is a ratio data. Choose the 'Interval/ratio' branch located at the top right section of the map. (Step 1)

Parameter: Parameters are measures of the population. Examples of commonly-used parameters are like mean, median, proportion, variance, correlation coefficient. In our case, it is appropriate to select mean (μ) as our parameter for the purpose of making generalization about the average content volume of all the bottles found in the population of study. (Step 2)

Assumptions: The assumptions of tests are important (Sheskin, 2001). The conditions under which certain tests can be employed must be determined now. Some of the assumptions are implicitly stated in the research design but some need to be checked before these tests are allowed. Two assumptions will be considered in our case study, namely normality of population and the population variance. There are many techniques for checking normality which are

usually discussed in standard statistics textbooks. Once it is established that the population is normal, proceed to look into the variance.

- Population is considered normal   (Step 3)
- Variance unknown   (Step 4)

All of the pertinent information (Steps 1-4) leads to Test A5: t-test for a population mean (Step 5). Once the test is identified, it is quite straightforward to analyze the data using one of the commercially available statistical packages like SAS, Stata, Minitab or SPSS (StatPac, 2010; Field, 2013).

This structured procedure although it resembles a 'recipe book' format, is designed to help users make connections between the six roadmaps. It will help users see the big picture of the possible relationships among the 49 tests made available in the maps. The layout of the maps is particularly useful to users who are able to analyze and synthesize the multi-faceted relationships to form new knowledge and understanding made explicit through the 'big picture'. A comparison of this nature cannot be made possible if one is to read about each of these tests from the textbooks. The main disadvantage of textbook is that the learning is linearly-structured and hierarchical in nature where contents are partitioned into discrete packages of knowledge by chapters and authors, more often than not, failed to make readers realize the relationships between these pockets of knowledge (Foo, 2011). More importantly, users of STestMAP can also see the connections among statistical concepts within each of the maps.

6. Conclusion

NHST is a difficult topic for researchers and research students alike. For many, the hypothesis testing routine is complex and arduous (Regoniel, 2015). Instructors are continuously exploring innovative teaching methods in solving selection problems faced by students. Statistical test selection in NHST remains a crucial skill for researchers to acquire as complete understanding of the basics will be useful in moving forward to doing multivariate analysis (Foo, 2011). It is important to design appropriate learning tools that have potential in helping students learn and understand statistical inference selection through techniques that suit the cognitive make-up of the students. A tool like STestMAP is comparatively new but it represents a promising approach to the mastery of selection skill.

The effectiveness of this approach has not been well-established as this is considered fairly new. Initial findings from a government-funded research on the effectiveness of STestMAP as a tool for the acquisition of test selection skill have shown that STestMAP provides a flexible structure in assisting users to successfully identify the statistical test (Tang et al., 2009). This study started out in 2008 and obtained its samples from Malaysia and Singapore. The findings showed that participants not only acquired procedural understanding of the NHST process but also employed metacognitive skills in their interaction with the various maps. Further research into STestMAP is highly recommended to realize and document its full potential in term of teaching and learning of inferential statistics for university students and researchers.

References

Anderson-Cook, C., M., and Dorai-Raj, S. (2003). Making the concepts of power and sample size relevant and accessible to students in introductory statistics courses using applets. *Journal of Statistics Education*, 11(3). Retrieved from ww2.amstat.org/publications/jse/v11n3/anderson-cook.html.

Chance, B. L., and Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Journal of Educational Statistics,* 1(2), 38–44.

delMas, R. C., Garfield, J. B., and Chance, B. (1999). Exploring the role of computer simulations in developing understanding of sampling distributions. *American Educational Research Association.*

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th edition.). London: Sage Publications Ltd.

Foo, K. K. (2011). Null Hypothesis Significance Testing: An Asian perspective. Pusat Penerbitan Universiti (UPENA), Universiti Teknologi MARA.

Foo, K. K., Leong, S. H., and Lau, U. H. (2005). STestMAP: Roadmaps to Statistical Tests. Pusat Penerbitan Universiti (UPENA), Universiti Teknologi MARA.

Gardner, P. L., and Hudson, I. (1999). University students' ability to apply Statistical procedures. *Journal of Statistics Education*, 7(1).

Nayak, B.K. and Hazra, A (2011) How to choose the right statistical test?. *Indian J Ophthalmol*, 59(2): 85–86. doi: 10.4103/0301-4738.77005

Quilici, J. L., and Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, 88(1), 144–161.

Regoniel, P. A. (2015). Statistical analysis: How to choose a statistical test. Retrieved from http://simplyeducate.me/2015/02/18/statistical-analysis-how-to-choose-a-statistical-test/

Schau, C., and Mattern, N. (1997). Assessing students' connected understanding of statistical relationships. *The Assessment Challenge in Statistics Education*, 91–104.

Sheskin, B. (2001). *Handbook of Parametric and NonParametric Statistical Procedures* (3rd edition). Florida: Chapman and Hall.

Skemp, R. (1971). *The Psychology of Learning Mathematics*. Penguin Books.

StatPac (2010). *StatPac user's guide*. Retrieved from https://statpac.com/manual/index.htm?turl=collinearitydiagnostics.htm.

Tang, H. E., Lau, U. H., and Foo, K. K. (2009). STestMAP as an alternative for significance test selection. Institute of Research, Development and Commercialization, Universiti Teknologi MARA.

Tukey, J.W. (1962). The future of data analysis. *Annals of Mathematical Statistics*, 33(1), 1-67.