Malaysian Journal of Mathematical Sciences 8(1): 109-116 (2014)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

Statistics Commonly Used in Research Reports: Issues and Challenges

[1*]Foo Kien Kheng, [1]Nuralina Azlan and [2]Noraini Idris

[1]Department of Mathematics,
Faculty of Computer Science and Mathematics,
Universiti Teknologi MARA, Negeri Sembilan Campus,
72000 Kuala Pilah, Negeri Sembilan, Malaysia

[2] Department of Mathematics and Science Education,
Faculty of Education, University of Malaya,
50603, Kuala Lumpur, Malaysia

E-mail: kienkheng@yahoo.com

*Corresponding author

ABSTRACT

Traditional descriptive measures like mean and variance together with the ubiquitous *p*-value in hypothesis testing are still widely analyzed and reported by researchers. Unfortunately literature has shown that researchers fail to exploit the full potential of statistics to unlock more meaningful information from their data. This study is aimed at looking into this issue in a Malaysian context and determining the type of research designs and statistics used in reporting findings based on an analysis of over 220 papers published in a conference proceeding organized by a local university. Frequency count, tabulations and graphical representations were used to analyse the data. Findings indicated that majority of the papers employed quantitative design with descriptive statistics as their major reporting technique. Among the 128 papers using inferential statistics, 35% of them used univariate tests while only 12% resorted to multivariate tests. Two faculties i.e. Social Sciences & Humanities and Science & Technology are major contributors to the high percentage of usage of statistical-related reports. Furthermore, only 7 papers reported their findings using confidence interval (CI) while 23 papers used effect size (ES) and none reported power of tests. This paper concludes with a short discussion on pertinent issues raised and provided some suggestions with regards to additional reporting measures other than *p*-value.

Keywords: research reports, descriptive statistics, inferential, confidence interval, effect size and power.

1. INTRODUCTION

Doing statistics is basically different from doing mathematics. Doing statistics is about how the learner deals with uncertainties and missing information whereas doing mathematics is about deriving a unique answer. Mathematics does not necessarily require a context to understand it. This cannot be said of Statistics for it requires a context to frame the problem meaningfully. Due to this, students who can think well mathematically may not be so comfortable with probability and statistics. Students and even teachers find topics such as probability to be comparatively difficult to understand.

A study was carried out by Foo (2011) to determine the perception of Malaysian and Singaporean undergraduate and postgraduate students concerning the understanding of some selected statistical terms used in introductory statistics classes. It was found that less than 30% of the one hundred respondents indicated that they had a good understanding of the terms such as inductive inference, deductive inference, Bayesian interpretation, frequentist interpretation, posterior probability, statistical signal and noise, strength of evidence, Cohen d and eta square. Qualitative analysis of interview data collected from 9 of these respondents, indicated that there was clearly a lack of understanding that seemed to stem from inadequate exposure to these concepts. Findings from this study attributed the students' difficulty in learning statistics to lack of statistical understanding, misconception and anxiety. The scenario in many statistics classes today do not differ much across countries. Statistics class usually begins with the teacher devoting a major portion of class time to communicating facts and formulas, a short question and answer session and finally an exercise with exam-oriented questions emphasizing procedures and calculations. This teaching approach sadly inhibits students conceptual development whereby learning inferential concepts is far more difficult than that of descriptive terms and concepts. Weak statistics foundation will impact their learning in later courses. In addition, students who come to class with well-grounded misconceptions face even greater problems in learning statistical inference.

Anderson-Cook and Dorai Raj (2003) came to the same conclusion concerning students' difficulties with understanding inferential statistics. They found that in frequentist hypothesis testing students usually faced problems in selecting correct statistical tests, identifying the 'correct procedures and calculation, interpretation of a critical value or *p*-value'.

In this respect, without a good grounding in the basics, hypothesis testing will just be an act of memorizing facts with shallow and isolated understanding. Hence, denying themself of knowing the important relationships that exist between them (delMas et al, 1999).

On the subject of reporting results of frequentist hypothesis testing, Pituch (2004) analyzed twelve introductory statistics textbooks published between 1998 and 2001. His analysis found that concepts such as effect size, confidence intervals, power and a prior sample size determination, and statistical assumptions were discussed in these texts but they failed to present an integrated reporting guide on how informative measures like effect size, confidence intervals and power could be presented to enhance the quality of the findings.

For decades these measures are still not reported extensively as compared to the ubiquitous *p*-value in journals and theses. Steiger and Fouladi (1997) hypothesized four probable reasons behind this phenomenon i.e. tradition, pragmatism, ignorance and lack of availability of computer software. A survey of the present day texts revealed that there are only limited works written specifically to enlighten new researchers on how to report these measures. Works by Schmidt and Hunter (1997); Steiger and Fouladi (1997) and Cumming and Finch (2001) represent some of the materials made available.

2. METHODS

This study is quantitative in nature. Its objectives include determining and analyzing the different research designs and types of statistics commonly used by researchers in reporting their findings. To achieve these objectives, a sample of 224 research papers from a conference organized by Research Management Institute, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia held from 14-15 March 2009 was selected for analysis. This 'Conference on Scientific and Social Research', provided a platform for researchers to share their findings with colleagues from branches all over Malaysia. UiTM the largest university in the country, has 24 faculties grouped under three major groupings, namely: Social Sciences and Humanities (7 faculties), Science and Technology (13 faculties) and Management and Business (4 faculties). Apart from these groupings, the conference also received papers from the Academy of Languages (APB), UiTM, Centre for Islamic Thought & Understanding (CITU), UiTM and other UiTM academic centers grouped under 'Miscellaneous'. Data analysis

was organized and presented according to these groupings. The analysis used only descriptive statistics where frequency count, percentage and cross-tab tables represent the major reporting measures. The trend among variables of interest was noted for further analysis and interpretation.

3. FINDINGS

TABLE 1: Distribution on the Usage of Research Designs used in reports among Groupings

	Quantitative	Qualitative	Mixed
Social Science and Humanities	25	12	11
Science and Technology	97	5	9
Management and Business	37	2	5
Islamic Studies	2	1	0
Academy of Language Studies	8	1	4
Miscellaneous	3	1	1
TOTAL	**172**	**22**	**30**

Table 1 shows the distribution of research designs employed by each grouping. All the groupings used quantitative design but in varying degree. Results show that as high as 76.8% of all the papers used quantitative methods while qualitative design and mixed methods totaled 23.2%. The highest number of papers using quantitative design comes from Science and Technology followed by Management and Business while Social Sciences and Humanities is more inclined towards qualitative or mixed methods.

TABLE 2: Distribution on the type of statistics used in reports among Groupings

	Descriptive only	Inferential only	Both
Social Science and Humanities	18	1	18
Science and Technology	61	3	26
Management and Business	12	2	29
Islamic Studies	1	0	1
Academy of Language Studies	5	2	4
Miscellaneous	2	0	3
TOTAL	**99**	**8**	**81**

Table 2 illustrates the level of usage of descriptive and inferential statistics out of the 202 papers that used quantitative design approach. This total included 30 papers from the 'mixed methods' category that employed some form of statistical analysis. The remaining 22 papers did not have a clear methodology description; thus they were excluded. Science & Technology grouping reported a sizable amount of descriptive statistics-61 papers and a further 26 papers used both descriptive and basic inferential analysis. The Management and Business on the other hand used both descriptive and inferential statistics. Interestingly enough, the difference in the usage of the type of statistics among the faculties is probably due to the design of their studies.

TABLE 3: Frequency distribution on measures like CI, ES and Power used in reports among Groupings

	Confidence Interval (CI)		Effect Size (ES)		Power of test	
	N	%	N	%	N	%
Social Science and Humanities	2	0.89	3	1.34	0	0
Science and Technology	5	2.23	6	2.68	0	0
Management and Business	0	0.00	13	5.80	0	0
Islamic Studies	0	0.00	0	0.00	0	0
Academy of Language Studies	0	0.00	0	0.00	0	0
Miscellaneous	0	0.00	1	0.45	0	0
TOTAL	**7**	**3.13**	**23**	**10.27**	**0**	**0**

Finally, in frequentist hypothesis testing, confidence interval, effect size and power of tests are important measures for inclusion in the planning of any quantitative studies. Looking at Table 3, there is very low incidence of reports using the measures. CI is found in 7 out of the 224 papers, a mere 3.13% while ES achieved 10.27% representing 23 papers out of the grand total. There is no paper at all reporting power of test. Among the three measures, ES is a more popular measure accounting for 5.80% among the researchers in Management & Business and 2.68% in Science & Technology. This reflects the importance of ES as a supplementary measure for these researchers as compared to the other two measures. Confidence Interval comes next in popularity with the Science & Technology grouping. Sadly in all the papers using inferential statistics, none reported power of test.

According to Tabachnick and Fidell (2007), "a critical issue in designing any study is whether there is adequate power". Undoubtedly, this measure should have been included prior to the start of the data collection process.

4. ISSUES AND CHALLENGES

The results have shown that majority of researchers in this particular university used quantitative design with descriptive statistics as their main analytical tool. A reasonable amount of inferential statistics is employed with the use of *p*-value taking center stage. There seems to be a lack of usage of other informative measures such as CI, ES and Power of Test or the more advanced measures associated with Bayesian statistics. This lacuna in reporting style could come about due to many factors as had been explained by Steiger and Fouladi (1997) and Foo (2011). In addition, Thompson (2001) noted, "It is conceivable that some researchers may not fully understand statistical methods that they (a) rarely read in the literature and (b) infrequently use in their own work" (p. 26).

Psychologists of late have renewed the call for changes to our entrenched practices in reporting empirical studies. According to Cumming and Finch (2001), "Reform of statistical practice in the social and behavioral sciences requires wider use of confidence intervals (CIs) and effect size measures. For decades, many advocates of statistical reform have recommended CIs as an alternative, or at least as a supplement, to *p* values". The American Psychological Association's (American Psychological Association, 2001) *Publication Manual* now calls CIs "the best reporting strategy" (p. 22).

Recommendations for an increased usage of these reporting techniques supported by brief discussion of the advantages of confidence interval over *p* value are few and far between. Till today there is an underlying "resistance" in the use of CIs and other reporting techniques probably not due to disagreements on their usefulness but on the lack of awareness and exposure to how these measures are to be reported. Foo (2011), Steiger & Fouladi (1997), Cumming & Finch, (2001) and Pituch (2004) among many had elaborated on this issue. It is apparent from their discussions that some of the basic reasons for the low usage of these measures are due to the following: a) lack of clear and comprehensive guidelines on how to present, interpret and report confidence interval, effect size and power analysis, (b) little encouragement from journal editors and

reviewers in making it compulsory for contributors of articles to report confidence interval, effect size and power of test besides *p*-value; (c) ineffectiveness of statistics and research method educators in explaining how these measures enhance the report; (d) traditionally, textbook authors had followed a fixed 'recipe' in writing the content for an introductory statistics book. For practical reasons they do not divert from the 'beaten path' to save time and effort. Consequently without doing a thorough analysis to know what present readers really need, authors are not in tune with the current needs of statistics education and seldom incorporate recent research findings and issues of importance to the statistics community.

It is especially important that readers critically evaluate their present statistical practices and are willing to make changes when changes are called for. Ultimately, one pertinent question we need to ask ourselves as educators and researchers is: 'Are we doing enough to be exemplary by reporting appropriate statistical measures other than the mean, standard deviation and *p*-value?'

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).

American Psychological Association. (2001). *Publication manual of the American Psychological Association*. Washington, DC: Author.

Cumming, G., and Finch S. (2001). A primer on the understanding use, and calculation of confidence intervals that are based on central and noncentral distribution. *Educational and Psychological Measurement.* **61**: 532-575.

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

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

Pituch, K. A. (2004). Textbook Presentations on Supplemental Hypothesis Testing Activities, Nonnormality, and the Concept of Mediation. *Understanding Statistics.* **3**(3): 135 - 150.

Schmidt, F. L. and Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. Harlow, Mulaik, S. A. and Steiger, J. H. (Ed.), *What if there were no significance tests?* London, UK: Erlbaum. 37-64.

Steiger, J. H. and Fouladi, R.T (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, and J. H. Steiger (Eds.), *What if there were no significance tests?* London, UK: Erlbaum. 251-257.

Tabachnick, B. G. and Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson.

Thompson, B. (2001). Significance, effect sizes, stepwise methods and other issues: Strong arguments move the field. *Journal of Experimental Education.* **69**: 80-93.