TYPE I ERROR AND POWER RATES OF ROBUST METHODS WITH VARIABLE TRIMMED MEAN

ZAHAYU MD YUSOF

UNIVERSITI SAINS MALAYSIA

2010

TYPE I ERROR AND POWER RATES OF ROBUST METHODS WITH VARIABLE TRIMMED MEAN

by

ZAHAYU MD YUSOF

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

UNIVERSITI SAINS MALAYSIA

2010

ACKNOWLEDGEMENTS

This thesis arose in part out of years of research that has been done since I came to ‘Robust Statistics Computing Laboratory’ at USM. By that time, I have worked with a great number of people whose contribution in assorted ways to the research, and the making of the thesis, deserved special mention. It is a pleasure to convey my gratitude and humble acknowledgment to them all.

First and foremost, I offer my sincerest gratitude to my supervisor, Associate Professor Dr Abdul Rahman Othman, who has supported me throughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. I attribute the level of my PhD to his encouragement and effort, and without him this thesis, too, would not have been completed or written. One simply could not wish for a better or friendlier supervisor.

I gratefully acknowledge Associate Professor Dr Sharipah Soaad for her advice, supervision, and crucial contribution, which made her a backbone of this research and to this thesis. I would also like to gratefully acknowledge the support of some very special individuals. They helped me immensely by giving me encouragement and friendship. I gratefully thank Masnita for her constructive comments on this thesis. Many thanks go in particular to Nora, Sin Yin, K.Nazihah, Aishah, Sue, Rini and Seou Ling, for giving me such a pleasant time when working together with them since I knew them at USM.

Words fail me to express my appreciation to my beloved husband Muhammad Syahril whose dedication, love and persistent confidence in me, has taken the load off my shoulders. I owe him for being unselfishly letting his intelligence, passions, and ambitions collide with mine. My parents, parents-in-law, and my wonderful sons, Muhammad Harriz Anwar and Muhammad Hadi Asyraf deserve special mention for their inseparable support and prayers.

Finally, I would like to thank everybody who was important to the successful realization of this thesis, as well as expressing my apology that I could not mention them personally.

TABLES OF CONTENTS

CHAPTER THREE: METHODOLOGY

CHAPTER FIVE: CONCLUSION

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
F_t	A statistical method for testing the equality of central tendency measures
LMS_n	A scale estimator
MAD_n	Median absolute deviation about the median
T_1	A statistical method for testing the equality of central tendency measures
T_n	A scale estimator

LIST OF PUBLICATIONS

Md Yusof, Z., Othman, A.R. & Syed Yahaya, S.S. (2007). Statistik terpangkas F_t menggunakan kriteria pemangkasan berdasarkan penganggar skala teguh MAD_n. *Menemui Matematik*, **29**(1), 21 – 28.

Yaacob, C. R, Othman, A. R. & Md Yusof, Z. (2007). Some computationally simple robust central tendency measures. *Menemui Matematik*, **29**(1), 1 – 8.

Md Yusof, Z., Othman, A. R., & Syed Yahaya, S. S. (2007). Statistik T_1 yang diubahsuai menggunakan kriteria pemangkasan berdasarkan penganggar skala teguh MAD_n. [Modified T_1 statistic using trimming criteria based on robust scale estimator MAD_n]. In D. Mohamad, R. Ismail, M. Awang, K.I. Othman, A. M. Shahidin, A. Nik Taib, N. Salim, N. H. Sarmin (Eds.), *Prosiding Simposium Kebangsaan Sains Matematik Ke 15: Sains Matematik Pemangkin Pembangunan Modal Insan* [*Proceedings of the Fifteenth National Mathematical Sciences Symposium: Mathematical Sciences the Catalyst of Human Capital Development*] (pp. 629-634). Shah Alam, Malaysia: Universiti Teknologi Mara.

Md Yusof, Z., Othman, A.R. & Syed Yahaya, S.S. (2007). Type I error rates of trimmed *F* statistic. *In Proceedings of the 56th Session of the International Statistical Institute (ISI 2007),* 22 - 29 August 2007 (CD ROM). Lisbon Portugal.

Md Yusof, Z., Othman, A.R. & Syed Yahaya, S.S. (2007). Type I error rates of trimmed *F* statistic for equal population variances. *In Proceedings of the 9th Islamic Countries Conference on Statistical Sciences 2007 (ICCS IX) (CD ROM)*, 12 - 14 December 2007, Shah Alam Malaysia.

Md Yusof, Z., Othman, A.R. & Syed Yahaya, S.S. (2008). Type I error rates of T_1 statistic. *International Journal of Mathematical Sciences and Engineering Applications (IJMSEA)*, **2**(4), 305 - 312.

Md Yusof, Z., Othman, A.R. & Syed Yahaya, S.S. (2008). Detecting outliers by using trimming criterion based on robust scale estimators with SAS procedure. *Borneo Science*, **23**, 21 - 26.

RALAT JENIS I DAN KUASA UJIAN BAGI KAEDAH TEGUH MENGGUNAKAN PEMBOLEHUBAH MIN TERPANGKAS

ABSTRAK

Kesan ketaknormalan data serta masalah heteroskedastisiti terhadap statistik min terpangkas T_1 dan F_t diselidiki menggunakan dua kaedah pemangkasan iaitu min terpangkas secara automatik yang dicadangkan dan pemangkasan biasa menggunakan amaun tetap. Ini merupakan masalah tipikal bagi pengujian ukuran kecenderungan memusat. Bagi setiap ujian statistik, tiga prosedur pemangkasan automatik menggunakan penganggar skala yang berbeza, MAD_n, T_n, dan LMS_n, dan prosedur min terpangkas tetap diuji untuk keteguhan melalui kadar Ralat Jenis I dan kuasa ujian. Untuk mengenal pasti keteguhan setiap prosedur, beberapa pembolehubah dimanipulasi seperti bilangan kumpulan, saiz sampel seimbang dan sebaliknya, keheterogenan varians, pasangan bagi saiz sampel dan varians kumpulan, dan jenis taburan. Dapatan menunjukkan bahawa, apabila pemangkasan automatik menggunakan LMS_n diapplikasikan pada statistik F_t di bawah pengaruh taburan yang sangat terpencong bersama kes varians homogen, prestasi Ralat Jenis I adalah sangat menyakinkan. Bagi taburan berekor normal dan varians heterogen, statistik T_1 menggunakan pemangkasan automatik T_n menunjukkan prestasi yang memuaskan. Merujuk kepada kuasa ujian, pemangkasan automatik mampu menghasilkan kadar kuasa ujian yang tinggi terutamanya apabila keadaan saiz sampel tidak sama dan varians homogen. Dengan menggunakan pemangkasan automatik, prestasi statistik F_t adalah lebih baik dari statistik T_1 berdasarkan kadar Ralat Jenis I dan kuasa ujian secara serentak apabila varians homogen dan sampel saiz tidak sama. Untuk keadaan lain, min terpangkas tetap masih cenderung digunakan.

Katakunci: Statistik teguh, ralat Jenis I, kuasa ujian, penganggar skala teguh, taburan terpencong

TYPE I ERROR AND POWER RATES OF ROBUST METHODS WITH VARIABLE TRIMMED MEAN

ABSTRACT

The effects of nonnormality and heteroscedasticity on the T_1 and trimmed F (F_t) test statistics were investigated using two methods of trimming namely the proposed automatic trimmed mean and the typical fixed amount of trimming. These are typical problems in any test of equality of central tendency measure. For each test statistic, three automatic trimming procedures using different scale estimators MAD_n, T_n, and LMS_n, and a fixed trimmed mean procedure were examined for their robustness via Type I error and power rates. To identify the robustness of each procedure, several variables were manipulated such as number of groups, balanced and unbalanced sample sizes, variance heterogeneity, pairing of group variances and group sample sizes, and types of distributions. The findings show that when automatic trimming using LMS_n was applied on F_t statistic under the condition of extremely skewed distribution with homogeneous variance cases, the performance of Type I error is very convincing. For normal-tailed distributions and heterogeneous variances, the T_1 statistic with automatic trimming using T_n performed reasonably well. With regard to power, the automatic trimming is able to produce high power rates especially for the conditions of unequal sample sizes and homogeneous variances. By means of automatic trimming, the performance of F_t statistic is better than the T_1 statistic simultaneously in terms of Type I error and power rates for homogeneous variances and unequal sample sizes. For other conditions, fixed trimmed mean is still favorable.

Keywords: Robust statistics, Type I error, power, robust scale estimators, skewed distributions

CHAPTER 1

BACKGROUND

1.1 Introduction

Analysis of variance (ANOVA) is the most commonly used statistical method for locating treatment effects in the one-way independent group design. However, ANOVA can be adversely affected by two general problems, namely nonnormality and heteroscedasticity. When these two problems arise simultaneously, Type I error rates are usually inflated resulting in spurious rejection of null hypotheses and reduction in the power of the test statistics.

The usual group means and variances are greatly influenced by the presence of outliers in the score distribution. Reduction in the power to detect differences between groups occurs because of the standard error for the usual mean can become seriously inflated when the underlying distribution is heavy-tailed (Lix & Keselman, 1998). In addition, the classical least squares estimators can be highly inefficient when assumptions of normality are not fulfilled.

One way to overcome the problems of controlling Type I error rates is by using robust statistics. Hence, by substituting robust measures of location and scale such as trimmed means and Winsorized variances in place of the usual means and variances respectively, tests that are insensitive to the combined effects of nonnormality and variance heterogeneity can be obtained (Lix & Keselman, 1998). Wilcox, Keselman &

Kowalchuk (1998) stated that one is able to obtain test statistics that do not suffer losses in power due to nonnormality by using trimmed means and variances based on Winsorized sum of squares.

Trimmed mean is a good measure of location because the standard error of the trimmed mean is less affected by departures from normality. This is due to the fact that the extreme values or outliers are removed (Lix & Keselman, 1998). According to Gross (1976), the Winsorized variance is a consistent estimator of the variance of the corresponding trimmed mean. Furthermore, the trimmed mean and Winsorized variance are intuitively appealing because of their computational simplicity and good theoretical properties (Wilcox, 1995).

In recent years, numerous methods are being studied in terms of finding better methods for controlling the rates of Type I error in the one-way independent group designs (Babu, Padmanabhan & Puri, 1999; Othman, Keselman, Padmanabhan, Wilcox & Fradette, 2004; Wilcox & Keselman, 2003).

1.2 Robust Statistics

There are several definitions of robust statistics that have been found in the literature and these unfortunately lead to the inconsistency of its meaning. Most of the definitions are based on the objective of the particular study by different researchers (Huber, 1981).

A statistical method is considered robust if the inferences are not seriously invalidated by the violation of such assumptions, for instance nonnormality and variance heterogeneity (Scheffe, 1959). Huber (1981) defined robustness as a situation which is not sensitive to small changes in assumptions while Brownlee (1965) reported slight effects on a procedure when appreciable departures from the assumptions were observed.

The theory of robust statistics deals with deviations from the assumptions on the model and is concerned with the construction of statistical procedures which is still reliable and reasonably efficient in a neighborhood of the model (Ronchetti, 2006). Hampel, Ronchetti, Rousseeuw and Stahel (1986), stated that in a broad informal sense, robust statistics is a body of knowledge, partly formalized into “theories of robustness” relating to deviations from idealized assumptions in statistics. As mentioned by Hoel, Port and Stone (1971), a test that is reliable under rather strong modifications of the assumptions on which it was based is said to be robust. Hence in this thesis, a statistical method is considered robust when it has estimators which cannot be influenced by the deviations from the given assumptions when hypothesis testing is being conducted.

Robust statistics has widely been used for many years now. Ronchetti (2006) reported that research in robust statistics has been conducted since 40 years ago and this area of research is still being actively studied today. In Ronchetti’s (2006) quick search in the Current Index of Statistics, 1617 papers on robust statistics were found between 1987 and 2001 in statistics journals and related fields.

To date, there are several new procedures that were developed to deal with group trimmed means. One of which is the modified *MOM-H* statistic introduced by Wilcox & Keselman (2003) which used modified one-step *M*-estimator *(MOM)* as the central tendency measure in their work on the *H* statistic. Essentially, *MOM* is variable trimmed mean with trimming carried out automatically. This method was proven to have good control of Type I error rates when comparing for the differences between distributions. Motivated by the good performance of this procedures, in this study we propose a modification of T_1 statistic developed by Babu *et al*. (1999) with automatic trimming strategy based on trimming criteria using robust scale estimators, MAD_n, T_n and LMS_n (Rousseeuw & Croux, 1993).

The other new procedure is a modified trimmed *F* statistic (F_t statistic) based on a priori determined symmetric or asymmetric trimming strategies introduced by Keselman, Wilcox, Lix, Algina and Fradette (2007). This method was also proven to have good control of Type I error rates when comparing for the differences between distributions. In our study, we change the a priori trimming strategies to automatic trimming. Again, the automatic trimming was based upon the three robust scale estimators mentioned earlier.

The original T_1 and F_t statistics used fixed trimming percentage of 15% symmetric trimming in order to calculate the trimmed means. Unlike the original, we proposed automatic trimming. No fixed trimming percentage is needed.

1.3 Trimming

Two approaches that may be considered by researchers faced with data that appear to violate the ANOVA assumptions are (i) to apply a transformation to the data and proceed with use of the *F* test or (ii) to select an alternative test procedure which is insensitive (i.e., robust) to assumption violations.

1.3.1 Purpose of trimming

When data are not normal and variances are heterogeneous, it is often possible to transform the data so that the new scores approximate normality and equality of variances. For example, when dealing with skewed distributions, two general suggestions are to take the square root or logarithms of every observation. Often these transformations produce data that are nearly normal. In some circumstances, the same transformations also achieve equality of variances (Maxwell & Delaney, 2004). Transforming data from designed experiments is an old and valuable tool (Carroll, 1982). Most researchers would wish to transform data if such was necessary to obtain a normal distribution. Upon transformation, standard analyses will often be performed.

However, there are some issues that should be kept in mind when applying transformation. First, transformation of data indicates that an attempt to avoid making inferences about the mean of the original score. This will lead to complex issues of interpretation, since the conclusions which are drawn must be based on the transformed scores, not the original observations (Lix, Keselman & Keselman, 1996). Thus, the interpretation of the results may also be less clear (Maxwell & Delaney, 2004). For example, most individuals find it difficult to understand the mean value of the square root of their original observations. Second, the complex transformations (i.e. Box-Cox transformation) do not remove the effects of outliers. That is, outliers remain and can inflate the sample variance and also lower the power by a substantial amount. Third, if each observation is transformed in the same manner, situations arise where the distribution of the observed scores remains skewed (Wilcox, 2002). Fourth, there is the problem of finding the correct transformation. Even though, there are a variety of transformations which may be applied to a set of data (Oshima & Algina, 1992), depending on the particular type and degree of assumption violation that is thought to be present in the data, this may not always be a simple solution (Lix *et al*., 1996). Also, it is difficult to find a transformation that will simultaneously deal with asymmetric data distributions and variance heterogeneity (Keselman *et al*., 2007).

Because of all of these drawbacks especially the interpretation issues, e.g. square root of the mean and log of the mean, we will ignore transformation and consider a robust method involving trimming.

The robust method involving trimming is another alternative method to deal with nonnormal distribution. This robust test will control the actual Type I error rate close to the nominal level of significance, even when the data do not conform to the test's derivational assumptions, and will maintain actual statistical power close to theoretical power, as well (Lix *et al.*, 1996). The literature so far suggest that this robust test is generally superior to the classical ANOVA *F* test and alternative test statistics (e.g., Welch) in the majority of assumption violation situations (see Levy, 1978; Tomarken & Serlin, 1986).

Methodology researchers consider ways to improve the performance of alternative procedures when the data are nonnormal (Lix *et al.*, 1996). Wilcox (1995) has suggested that trimming, or discarding outliers from a data set prior to analysis, can lead to improve performance, both in terms of Type I error control and power. Trimming is the most popular robust based method when dealing with skewed data. Naturally, trimming is a very drastic way of dealing with extreme observations. However, removing a small set of observations in a relatively large sample should not change the results in a major way (Rodrigues & Rubia, 2006).

The key factors in trimming are the amount of trimming and how the trimming is specifically conducted. There are two common methods in trimming, symmetric and asymmetric trimming. In symmetric trimming, equal amount of trimming is applied on both tails of the distribution. In asymmetric trimming, the process of trimming is either conducted on one-tail or on both tails with unequal amounts. In order to avoid loss of information, trimming need to be conducted with care. Before trimming could be performed, the amount of trimming has to be determined first, usually by fixing the amount of trimming (predetermined). In our study, we are going to depart from trimming with fixed amount to automated trimming.

1.3.2 Trimmed mean

Trimming will definitely get rid of outliers but how do we address the question of outliers? Usually outliers are causes of nonnormality and heterogeneity. Even so, if we are looking at the differences between groups, the presence of a few outliers in one group will definitely lead to rejection of the null hypothesis. How do we deal with this rejection? This rejection should not be taken at face value. Further analysis will now be done on these outliers in order to determine their inclusion or exclusion in the study. In our study, the question of outliers does not arise because our study conditions do not involve them. Our study conditions are variance heterogeneity, pairing of group variances and group sample sizes, types of distributions, balanced and unbalanced sample sizes and number of groups.

Trimmed mean is a central tendency measure that summarizes data when trimming is carried out. By using the trimmed means, the effect of the tails of the distribution is reduced by their removal based on the trimming percentage that has to be stated in advanced (predetermined amount). The common trimmed mean used the fixed amount of trimming method. It needs the fix amount of trimming percentage and tight down with this amount of trimming. By using this method, amounts such as 10% or 20% of the observations from a distribution will be trimmed from both tails. In the case of a light-tailed distribution or the normal distribution, it may be desirable to trim a few observations or none at all. There is extensive literature regarding this trimming method that uses the fixed amount of symmetric trimming. Among them are Lee and Fung (1985), Keselman, Wilcox, Othman and Fradette (2002), and Wilcox (2003).

If we have skewed distributions then the amounts of trimming on both tails should be different. More should be trimmed from the skewed tail. However, if the fixed symmetric trimming is used, regardless of the shape of the tails, the trimming is done symmetrically as set. A research by Keselman *et al*. (2007) used asymmetric trimming and in particular, applying hinge estimators proposed by Reed and Stark (1996) to determine the suitable amount of trimming on each tail of a distribution. However, their method still used fixed trimming percentages.

The trimmed mean is not so robust because the breakdown point of trimmed mean is just as much as the percentage of trimming and this shows that trimmed mean cannot withstand large numbers of extreme value. Wilcox, Keselman, Muska nad Cribbie (2000) in their study stated that when comparing trimmed means versus means with actual data, the power of the trimmed mean procedure was observed to be greatly increased. They also discovered that there was improved control over the probability of a Type I error.

The question that always remains unanswered is "How can we determine the best percentage of trimming that would ensure good Type I error control and reasonable power?" A probable answer lies in trimming carried out for the calculation of modified one-step *M* - estimators (*MOMs*). Here trimming is based upon a trimming criterion that relies upon a robust scale estimator known as MAD_n (Wilcox & Keselman, 2002). With this method of trimming we do not have to fix the amount of trimming required. The criterion will identify how many extreme values need to be removed from the distribution.

Other than MAD_n, Rousseeuw and Croux (1993) have demonstrated that robust scale estimators T_n and LMS_n can also be used successfully as trimming criteria for *MOMs* based procedures. Hence this study will examine the viability of the usage of these three trimming criteria in variable trimmed means based procedures.

1.4 T_1 Statistic

Types of distributions and homogeneity of variances are two important aspects that need to be taken into consideration before we proceed with the testing of the equality of central tendency measures using robust statistics. If the type of distribution is unknown and cannot be assumed as normally distributed, Babu *et al*. (1999) suggested the use of their T_1 statistic to compare the differences between distributions. They applied this statistic when the distributions are tested symmetric. This procedure used 15% symmetric trimming with trimmed mean as the central tendency measure.

1.5 Trimmed *F* Statistic, F_t

Lee and Fung (1985) introduced a statistical procedure that is able to handle problems with sample locations when nonnormality occurs but the homogeneity of variances assumption still applies. This statistic is known as trimmed *F* statistic. We denote it as F_t. They also suggested that this new statistic be used as an alternative to the classical *F* method involving one-way independent group design. By using the 15% symmetric trimming, this procedure would give reasonable results for various types of distributions. Furthermore, this procedure is easy to compute.

1.6 Scale Estimators

A scale measure is a quantity that explains the dispersion of a distribution. The value of breakdown point is one of the main factors to be considered when we look for a scale estimator (Wilcox, 2005a). Rousseeuw and Croux (1993) proposed several scale estimators with high breakdown point such as S_n, T_n and LMS_n. A breakdown point refers to the quantitative description of the effect of a small change in the underlying distribution F in changing the distribution of an estimate (Wilcox, 2005a). Another important feature for a robust scale estimator is the bounded influence function. In general, an influence function measures the change in the function due to small amount of contamination at the point of the observation x.

Syed Yahaya, Othman and Keselman (2004a, 2004b) identified four scale estimators with highest breakdown point and bounded influence function that were capable of maintaining the robustness of the S_1 statistic. The scale estimators are Q_n, S_n, T_n (Rousseeuw & Croux, 1993) and the well known scale estimator, MAD_n. The S_1 statistic which uses median as the central tendency measure was discovered by Babu *et al*. (1999) to test for differences between distributions. This flexible statistic dealt with asymmetric distributions and heteroscedasticity settings satisfactorily. The S_1 statistic works with the original data without having to trim or transform the data to achieve symmetry. Syed Yahaya *et al*. (2004a, 2004b) observed that the combination of the S_1 statistic with the aforementioned estimators produced good Type I error rates. The combination of the S_1 method with the scale estimator T_n produced a very promising procedure in robust statistics.

The scale estimator $LMS_{n,}$ is found to have influence function and efficiency which equals to *MAD*. However, LMS_n can be used under asymmetric distributions as well (Rousseeuw & Leroy, 1987; Grubel, 1988).

Motivated by the good performance of the scale estimators MAD_n and T_n in controlling Type I error rates in Syed Yahaya *et al*. (2004a, 2004b) and the good review of LMS_n by Rousseeuw and Leroy (1987), and Grubel (1988), we chose the three robust scale estimators, MAD_n, T_n and LMS_n as the criteria for choosing sample values (trimming criterion), and used the values to calculate T_1 and F_t under skewed distributions.

1.7 Objective of the Study

The main objective of this study is to examine the operating conditions that would result in good Type I error rates and power for the following new procedures:

1. T_1 with variable trimmed means derived using MAD_n.
2. T_1 with variable trimmed means derived using T_n.
3. T_1 with variable trimmed means derived using LMS_n.
4. F_t with variable trimmed means derived using MAD_n.
5. F_t with variable trimmed means derived using T_n.
6. F_t with variable trimmed means derived using LMS_n.

The secondary objective is to compare 1 – 6 against the original T_1 and F_t, both with 15% symmetric trimmed means. In doing so, this study should be able to

1. determine if 1 – 6 are improvements over the original T_1 and F_t.
2. recommend the best procedure for extreme conditions.

1.8 Significance of the Study

Experimental design methodology depends on the assumptions of normality and homogeneity of variances, but these assumptions are rarely fulfilled in the real world. Researchers need alternative methods when these situations arise. This study contributes to the development of robust statistics that uses trimming strategy in its test statistic or in its procedures. Robust statistics with trimming were designed to handle violation of assumptions such as normality and variance homogeneity. The usual trimming strategy normally requires fixed amount of trimming which have to be stated in advanced. In our proposed method, this is not the case. The proposed strategy trims data automatically based on the shape of the distribution. By using this strategy, researchers do not have to worry about how much trimming should be done to achieve good Type I error and high power rates. This study will also naturally want to determine whether the proposed trimming strategy will improve the performance of the T_1 and F_t statistics.

1.9 Organization of the Thesis

Chapter 1 gives an introduction on the importance of the study and gives in depth explanation regarding the robust statistical methods. This chapter also presents a brief introduction to the methods proposed in this study, namely T_1 and F_t statistics. Details of these methods are presented in Chapter 2. Chapter 2 also discusses about the scale estimators and defines terminologies used throughout this study. Explanations about operating conditions that have been manipulated are found in Chapter 3. They are the number of groups, the sample sizes for balanced and unbalanced design, heterogeneity of variances, the nature of pairings of group sample sizes and group variances and type of distributions. This chapter further gives the design specifications and explains the generation of data used in this study. The results from the analyses of Type I error and power were presented in Chapter 4. We conclude our findings and propose suggestions for further studies in the last chapter.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The two sample *t*-test and the analysis of variance (ANOVA) are two common statistical methods used to locate treatment effects in a one-way independent group design. However, in using these two statistics, assumptions of normality and variance homogeneity need to be fulfilled. In real life applications, these conditions are rarely achieved and these will lead to inaccuracy in decision based on the testing procedure.

Departures from normality originate from two problems, i.e. skewness and the existence of outliers. These problems could be remedied by using transformation such as exponential, logarithm and others but sometimes, even after the transformation, problems with nonnormal data still occur. Simple transformations of the data such as by taking logarithm can reduce skewness but not for complex transformations such as the class of Box-Cox transformations (Wilcox & Keselman, 2003). However, problems due to the outliers are not eliminated. According to Wilcox and Keselman (2003), a simple transformation can alter skewed distributions to make them more symmetrical, but they still do not deal directly with outliers. They suggested using a trimming method when dealing directly with outliers.

The existence of outliers in a sample data will cause the probability of Type I error to be less than the nominal alpha level and concurrently lower the power of the test statistic. In the application of *t*-test, outliers can inflate the sample variance and simultaneously lower the value of the test (Wilcox & Keselman, 2003). Even when sampling from a perfectly symmetrical distribution, outliers can still cause the *t*-test to lose power when compared against modern methods. Modern methods here are methods that are based on robust measures of location (Wilcox & Keselman, 2003). According to Keselman, Lix and Kowalchuk (1998), the reduction in the power to detect differences between groups occurs because the usual population standard deviation is greatly influenced by the presence of the extreme observations in a distribution of scores.

The presence of outliers will inevitably lead to the observed scores being skewed. However, skewness itself can be an inherent property of several score distributions. It is also well known that skewness can also be a problem when we are trying to control the probability of Type I error. Type I error rates and the confidence intervals can be highly inaccurate when the data are skewed. For the normal distribution and any symmetric distribution, the skewness for the distributions is zero. When the data are skewed to the left, the skewness value is negative. This denotes that the left tail is longer than the right tail. When the data are skewed to the right, the skewness value will be positive. Many classical statistical tests depend on normality assumptions. When this assumption is not satisfied, the rate of Type I error and the power of the test conducted will be affected.

The sample mean is the most common estimator used in most statistical analyses. However, this estimator is very sensitive to the presence of outliers and skewness. One single outlier could easily influence this estimator, thus causing it to have a low breakdown point (Sawilowsky, 2002). In addition, the sample mean also has unbounded influence function, implying that a single contaminated observation may have a considerable effect on the estimate (Thomas, 2000). Under these conditions, any test that used the sample mean as the estimator will produce low power and distorted rates of Type I error. These include the *t*-test and ANOVA. Furthermore, the standard error of the usual mean can become seriously inflated when the underlying distribution is heavy-tailed. To address this problem, Wilcox and Keselman (2003) suggested using estimators of robust measures of location and rank-based methods. Some of these robust estimators are the *M*-estimator and trimmed mean.

The sample trimmed mean (will be referred to as "trimmed mean" throughout this thesis) is one of the estimators which are able to handle the problem of nonnormality due to skewness. When using this estimator, the smallest and the largest observations in the distribution will be trimmed, thus automatically discarding skewed data. By using the trimmed mean, high power, accurate probability coverage, relatively low standard errors, a negligible amount of bias and a good control over the probability of a Type I error can be achieved (Wilcox & Keselman, 2003).

There are two possibilities of estimating the trimmed mean, i.e. equal amount of trimming or symmetric trimming and unequal amount of trimming or asymmetric trimming. In symmetric trimming, the trimming is done equally on both sides of the distribution. While for asymmetric trimming, the trimming is done on only one side or unequally on both sides of the distribution. Othman, Keselman, Wilcox, Fradette and Padmanabhan (2002) in their study suggested that when the data are said to be skewed to the right, then in order to achieve robustness to nonnormality and greater sensitivity to detect effects, one should trim data just from the upper tail of the data distribution. Hogg (1974), Hertsgaard (1979), and Tiku (1980, 1982) suggested that the data should have different amounts of trimming percentages from the right and left tails of the distribution. Keselman *et al*. (2007) proposed a method called adaptive robust estimators to determine the number of observations to be trimmed from each tail of the distribution. By using this method, the total amount of trimming is determined a priori before making the decision whether to trim the data symmerically, asymmetrically or not to trim at all.

If the distribution is skewed, the trimmed mean provides better estimates of the typical score than the usual mean. This is due to the fact that when a distribution is skewed, the trimmed mean does not estimate μ but rather some value (i.e. μ_t) that is typically closer to the bulk of the observations (Keselman *et al.*, 2004). Herron and Hillis (2000) stated that, for heavy-tailed distributions, the trimmed mean is less sensitive to the outliers and also have smaller standard errors than the usual mean. To avoid unnecessary loss of information due to trimming, if a distribution is highly skewed to the left, it seems more reasonable to trim more observations from the left tail of the distribution than from the right tail.

However, the trimmed mean suffers from at least two practical concerns which are (i) the proportion of data at the tails exceeds the percentage of adopted trimming and vice versa and (ii) the trimming is done unproportionately. In the latter case, the problem occurs when equal percentage of trimming (as in trimmed mean) on both tails is adopted on skewed distribution, whereas it would be more reasonable to trim more observations from the tail that is highly skewed. Note that these problems arise because of the amount of trimming have to be fixed in advance without examining the characteristics of the data. In many situations, researchers would want to use an adaptive trimmed mean, (i.e. asymmetric trimmed mean) in which the trimming proportion adapts itself to the characteristics of the distribution on the basis of the sample.

To avoid from trimming erroneously, the process needs to be done meticulously. In our proposed method of trimming, this problem can be avoided since the amount of trimming is determined by the characteristics of the sample data. This method utilizes characteristics of the observed data to determine whether data should be trimmed symmetrically, asymmetrically or not at all. The idea is that, good efficiency will be obtained when sampling from normal distributions as well as non-normal distributions by introducing flexibility into how much is trimmed.

Another problem which researchers always encountered when using the classical methods is heteroscedasticity. Some of the parametric methods that can handle this problem are those proposed by Welch (1961), James (1951) and Alexander and Govern (1994). Unfortunately, all of these methods have difficulty in dealing with problem of nonnormal data. Nonetheless, Abdullah, Syed Yahaya and Othman (2008) found that Alexander and Govern test which uses automatically trimmed mean as the central tendency measure in place of the usual mean is robust to skewed data when the trimming strategy was adopted.

Some researchers sought for alternatives in the non-parametric methods, such as Mann Whitney and Kruskall Wallis. However, these methods have low power (Wilcox, 1992). Even though non-parametric methods are distribution free, they are not assumptions free. Usually the distribution has to be symmetric. The alternative is to use a robust approach to deal with the problems of nonnormality and heteroscedasticity.

Robust statistics combine the virtues of both, the parametric and the non-parametric approach. In general, these statistics are used in handling the problem of the violation of the independence assumptions such as nonnormality and variance heterogeneity. In this study, we suggested two robust procedures, the T_1 statistic proposed by Babu *et al*. (1999) and the trimmed F statistic, F_t introduced by Lee and Fung (1985). Babu *et al*. (1999) suggested the use of T_1 statistic to compare the differences between distributions if the type of distribution is unknown and cannot be assumed as normally distributed. They applied this statistic with 15% symmetric trimmed mean as the central tendency measure when the distributions are tested symmetric. Trimmed F statistic is a statistical method that is able to handle problems with sample locations when nonnormality occurs but the homogeneity of variances assumption still applies.

In this study, we will look at the problems of nonnormality and variance heterogeneity, simultaneously. We will use these statistics with trimming strategies using robust scale estimators, T_n and LMS_n proposed by Rousseeuw and Croux (1993). In addition to these two estimators, we also consider one of the most popular estimators, MAD_n. We choose these estimators because of their high breakdown points and bounded influence functions. These strategies will trim extreme values without the need to state the trimming percentage in advanced.

There are a few terminologies that will be used throughout our study. We will discuss these terminologies briefly in the next sections prior to the in depth discussion of the proposed methods.

2.2 Trimming

Trimming is a method to eliminate outliers or extreme observations from each tail of a distribution. Determining the percentage of trimming must be made prior to the testing. In order to make this decision, efficiency is one factor to be considered. In this context, efficiency means achieving relatively small standard error when the trimming method is used. Trimming needs to be done cautiously. If the amount of trimming is too small, efficiency can be very poor when sampling is from heavy-tailed distribution, but if the amount is too large, efficiency will be very poor when we consider the sampling from a normal distribution (Keselman, Kowalchuk, Algina, Lix & Wilcox, 2000).

Trimming can be very beneficial in terms of efficiency and in achieving high power. Trimming can eliminate outliers and power might be increased substantially. This is a conclusion that follows almost immediately from a result derived by Laplace two centuries ago (Wilcox, 2005b). According to Wilcox (1998) trimming can be good or bad in terms of power, depending upon the criteria we adopt and the goals we hope to achieve. In Wilcox (2005b), it is stated that the median corresponds to the most extreme case in which all but one or two values are trimmed. He gave an example that if *n* is even, all but two observations are trimmed and if *n* is odd, all but one. Due to the extreme amount of trimming reflected by the usual sample median, the sample median will have a large standard error and low power relative to using the usual sample mean (Wilcox, 2005b).

Theory indicates that the more we trim, the more we can reduce problems due to skewness. Rocke, Downs and Rocke (1982) in their paper concluded that the best results were obtained with 20% – 25% symmetric trimming, while Othman *et al*. (2004) reported that one can achieve a slightly better Type I error control with a 15% symmetric trimming rather than a 20% symmetric trimming. Keselman, Othman, Wilcox and Fradette (2004) demonstrated that good control of Type I error can be achieved with only modest amounts of trimming, namely 15% or 10% from each tail of the distribution. For long-tailed symmetric distributions, Lee and Fung (1985) recommended the used of 15% symmetric trimming. According to the literature, the optimal fixed amount of symmetric trimming percentage is between 0% and 25%.

When sampling from a symmetric distribution, it is intuitively appealing to use symmetric trimming (Wilcox, 2003). Symmetric trimming trims the same number of observations at both ends of data and hence is quite efficient for symmetric distributions. However, this strategy becomes less efficient when there is even just a slight departure from symmetry, for example with one end containing outlying points (Wu & Zuo, 2009). Higher amount (i.e. more than 20%) of symmetric trimming should be used when sampling from a skewed distribution (Wilcox, 2003). Nevertheless if the amount of trimming is too high, this can result in lower power when sampling from a light tailed distribution (i.e. normal distribution) where outliers are relatively rare. While for heavy-tailed distributions, the power goes up as the amount of trimming increases, (Wilcox, 1995).

It has been a general practice that 90%, 95%, and 99% are typical choices to specify coverage probabilities. Nevertheless, as stated in Granger (1996), practical forecasters seem to prefer 50% intervals whereas academic writers focus almost exclusively on 95% intervals. It is noted that the larger the probability coverage, the wider the prediction interval, and vice versa. Relating to the trimming percentages, Wilcox (1998) stated that the more we trim, the less effect skewness had on the probability coverage. According to Wilcox (1996), a 20% trimming provide more accurate probability coverage of confidence intervals regarding differences between means when the distributions are skewed.

Nevertheless, when the sample size, *n* is small, the optimal amount of trimming is yet to be determined. The amount of trimming can also be arrived at empirically. However, it is difficult to do so. This is usually attempted when doing one-sided or asymmetric trimming. Othman *et al*. (2002) dealt with predetermined amount of trimming on one side. The recent study done by Keselman *et al.* (2007) also worked with fixed total amount of trimming for both sides of the distribution. They then identified the number of observations that should be trimmed from each tail by the characteristics of the sample data. However, the total number of trimmed data from the left and right tail of the distribution must be equal to the total amount of trimming that they determined earlier. The mismatch of the proportion of skewed data is still of practical concern if we use this method. Thus, in this study, we proposed a method of trimming without any fixed amount. The amount of trimming for both tails of the distribution is determined automatically using robust scale estimators, namely, MAD_n, T_n and LMS_n to get the sample values. We also compared this automatic method of trimming with the usual symmetric trimming. Specifically we chose 15% symmetric trimming for this purpose.

Essentially one does not trim a fixed amount of the data but only the skewed data. These trimming mechanisms will ensure that the problems of outliers and skewed data will be adequately addressed.

2.3 Type I Error

Hypothesis testing is the art of testing if variation between sample distributions can either be explained by chance or not. If we are to test two distributions to see if they vary in a meaningful way, we must be aware that the difference is not just by chance. Type I error is the error of rejecting the null hypothesis given that it is actually true. In other words, this is the error of accepting an alternative hypothesis when the results can be attributed to chance.

According to Steven (1990), a test statistic is robust if the actual level of significance is very close to the nominal level. The nominal level is the level set by the experimenter and is the percent of time one rejects falsely when the null hypothesis is true and all assumptions are met. While the actual level is the percent of time one rejects falsely if one or more of the assumptions are violated.

Type I error rejects an idea that should not have been rejected and also claims that two observations are different, when they are actually the same. It is also known as a ‘false positive’. A false positive usually means that a test claims something to be positive, when that is not the situation. The probability of a Type I error is designated by the Greek letter alpha (α) and is called the Type I error rate.

Conventionally Type I error is set at 0.05 or 0.01. This brings the meaning of there is only 5 or 1 in 100 chance that the variation that we obtained is due to chance. This is called the 'level of significance'. The significance levels need to be chosen attentively. For example, a 5% significance level is the rate to declare a result to be significant when there is actually no relationship in the population. The 5% value is also known as the rate of false alarms or false positives.

By convention, a procedure can be considered robust if it’s Type I error is between 0.5α and 1.5α (Bradley, 1978). Thus, when the nominal level is set at $\alpha =$ 0.05, the Type I error rate should be in between 0.025 and 0.075. Type I error rates are considered liberal when they are above the 0.075 limit while those below the 0.025 limit are considered conservative. However, Guo and Luh (2000) in their study regarded a test with 5% level of significance to be robust if its empirical Type I error rate does not exceed the 0.075 limit.