ROBUST PERCENTILE BOOTSTRAP TEST WITH MODIFIED ONE-STEP *M*-ESTIMATOR (*MOM*): AN ALTERNATIVE MODERN STATISTICAL ANALYSIS

NURUL HANIS BINTI HARUN

MASTER OF SCIENCE (STATISTICS)
UNIVERSITI UTARA MALAYSIA
2015

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

NURUL HANIS HARUN

calon untuk Ijazah MASTER
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"ROBUST PERCENTILE BOOTSTRAP TEST WITH MODIFIED ONE-STEP M-ESTIMATOR (MOM): AN ALTERNATIVE MODERN STATISTICAL ANALYSIS"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***29 September 2014.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: ***September 29, 2014.***

Pengerusi Viva: *(Chairman for VIVA)*	Assoc. Prof. Dr. Haslinda Ibrahim	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	Dr. Norhashidah Awang	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	Assoc. Prof. Dr. Sharipah Soaad Syed Yahaya	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	Dr. Zahayu Md Yusof	Tandatangan *(Signature)*

Tarikh:
(Date) ***September 29, 2014***

Permission to Use

In presenting this thesis in fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or, in her absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences

UUM College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

Abstrak

Kenormalan dan homoskedastisiti merupakan dua andaian utama yang perlu dipenuhi apabila berurusan dengan ujian-ujian parameter klasik untuk perbandingan kumpulan. Pelanggaran mana-mana andaian tersebut akan menyebabkan keputusan ujian menjadi tidak sah. Walau bagaimanapun, pada realitinya, kedua-dua andaian tersebut sukar dicapai. Untuk mengatasi masalah tersebut, kajian ini mencadangkan pengubahsuaian satu kaedah yang dikenali sebagai ujian Bootstrap Berparameter dengan menggantikan min sebenar, $\bar{X}$ dengan ukuran lokasi yang sangat teguh iaitu penganggar-M satu-langkah terubahsuai (*MOM*). (*MOM*) merupakan min terpangkas tidak simetri. Penggantian ini akan menjadikan ujian Bootstrap Berparameter lebih teguh untuk perbandingan kumpulan. Dalam kajian ini, kriteria pemangkasan untuk *MOM* menggunakan dua pengganggar skala yang amat teguh iaitu MAD_n dan T_n. Satu kajian simulasi telah dijalankan untuk mengkaji prestasi kaedah yang dicadangkan berdasarkan kadar Ralat Jenis I. Untuk mengenal pasti kekuatan dan kelemahan kaedah, lima pembolehubah iaitu: bilangan kumpulan, saiz sampel seimbang dan tak seimbang, jenis taburan, keheterogenan varians, dan sifat pasangan bagi saiz sampel dan varians kumpulan dimanipulasi untuk menghasilkan pelbagai keadaan yang biasanya wujud dalam kehidupan sebenar. Prestasi kaedah yang dicadangkan kemudiannya dibandingkan dengan ujian parameter klasik dan ujian tidak berparameter yang paling kerap digunakan untuk dua (ujian-*t* tidak bersandar dan ujian *Mann Whitney* masing-masing) dan lebih daripada dua kumpulan tidak bersandar (ANOVA dan ujian *Kruskal Wallis* masing-masing). Dapatan kajian menunjukkan bahawa, untuk dua kumpulan, ujian Bootstrap Berparameter yang teguh menunjukkan prestasi yang baik di bawah keadaan varians heterogen dengan taburan normal atau taburan terpencong. Manakala untuk lebih daripada dua kumpulan, ujian tersebut menjana pengawalan Ralat Jenis I yang baik di bawah varians heterogen dan taburan terpencong. Dalam perbandingan dengan kaedah parameter klasik dan keadah tidak berparameter, ujian yang dicadangkan menunjukkan prestasi yang lebih baik di bawah taburan terpencong dan varians heterogen. Prestasi setiap prosedur juga ditunjukkan dengan menggunakan data sebenar. Secara umumnya, prestasi Ralat Jenis I bagi ujian yang dicadangkan adalah sangat menyakinkan walaupun andaian kenormalan dan homoskedastisiti dilanggar.

Kata kunci: Titik kegagalan, Heterogen, Taburan terpencong, Ralat Jenis I.

Abstract

Normality and homoscedasticity are two main assumptions that must be fulfilled when dealing with classical parametric tests for comparing groups. Any violation of the assumptions will cause the results to be invalid. However, in reality, these assumptions are hardly achieved. To overcome such problem, this study proposed to modify a method known as Parametric Bootstrap test by substituting the usual mean, $\bar{X}$ with a highly robust location measure, modified one step M-estimator (*MOM*). *MOM* is an asymmetric trimmed mean. The substitution will make the Parametric Bootstrap test more robust for comparing groups. For this study, the trimming criteria for *MOM* employed two highly robust scale estimators namely MAD_n and T_n. A simulation study was conducted to investigate on the performance of the proposed method based on Type I error rates. To highlight the strength and weakness of the method, five variables: number of groups, balanced and unbalanced sample sizes, types of distributions, variances heterogeneity and nature of pairings of sample sizes and group variances were manipulated to create various conditions which are common to real life situations.The performance of the proposed method was then compared with the most frequently used parametric and non parametric tests for two (independent sample *t*-test and Mann Whitney respectively) and more than two independent groups (ANOVA and Kruskal Wallis respectively). The finding of this study indicated that, for two groups, the robust Parametric Bootstrap test performed reasonably well under the conditions of heterogeneous variances with normal or skewed distributions. While for more than two groups, the test generate good Type I error control under heterogeneous variances and skewed distributions. In comparison with the parametric and non parametric methods, the proposed test outperforms its counterparts under non-normal distribution and heterogeneous variances. The performance of each procedure was also demonstrated using real data. In general, the performance of Type I error for the proposed test is very convincing even when the assumptions of normality and homoscedasticity are violated.

Keywords: Breakdown point, Heterogeneity, Skewed distributions, Type I error.

Acknowledgement

This thesis would not been possible without guidance and help of several individuals who contributed their assistance in the preparation and completion of this study. It gives me great pleasure to acknowledge their support.

First and foremost, I would like to express the deepest appreciation and gratitude to my supervisor, Dr. Zahayu Md Yusof for her valuable support and guidance throughout this study. I could not have imagined having a better advisor and supporter for my master study in University Utara Malaysia.

I am deeply grateful to my parents, Harun Hamid and Sharifah Abu Bakar for their love, inspiration, patience and support. I dedicated this work to my parents. I would also like to thank to my aunt and my brother for their support.

Lastly, I would like to thank everybody who had directly or indirectly helped me during this research.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

ANOVA	Analysis of variance
Parametric Bootstrap	A statistical method for testing the equality of central tendency
MAD_n	Median absolute deviation about median
T_n	A scale estimator

List of Publications

Md Yusof, Z., Harun, N. H., Syed Yahaya, S. S. & Abdullah, S. (2013). A modified parametric bootstrap: an alternative to classical parametric test. *In proceeding of the World Conference on Integration of Knowledge 2013*, 25 – 26 November, Langkawi, Malaysia.

Harun, N. H, Md Yusof, Z. (2013). Testing the Equality of Central Tendency using Robust Parametric Bootstrap Test with *MOM* Estimator for Two Groups Case. *In proceeding of the 1st Innovation and Analytics Conference and Exhibition 2013*, 29 December, Universiti Utara Malaysia, Malaysia.

Harun, N. H, Md Yusof, Z. (2014). Robust Parametric Bootstrap Test with *MOM* Estimator: An Alternative to Independent Sample *t*-Test. In proceeding of the *3rd International Conference on Quantitative Sciences and Its Applications 2014*, 12 – 14 August, Langkawi, Malaysia.

CHAPTER ONE

BACKGROUND

1.1 Introduction

Statistics encompasses a wide variety of activities, ideas and results that can handle the situations involving uncertainties. Statistics consists of two basic statistical analysis namely descriptive statistics and inferential statistics. Recording and summarizing a data set is the main purpose of descriptive statistics whereas inferential statistics involves drawing conclusion and making decisions. There are extensive studies in testing equality of central tendency measures in inferential statistics using statistical method in order to make inferences based on obtained results. Basically, classical parametric tests such as analysis of variance (ANOVA) and independent sample *t*-test are often used in testing the central tendency measure by researchers rather than other methods since the aforementioned methods provide a good control of Type I error and generally more powerful than other methods when all the assumptions are fulfilled (Wilcox & Keselman, 2010).

ANOVA is used to determine the mean equality for more than two groups while independent samples *t*-test is used to determine the mean equality for two independent groups. However, a characteristic of these procedures is the fact that making inference depends on certain assumptions that need to be fulfilled. There are three main assumptions that need to be fulfilled before making inference on the classical parametric test such as: (a) collecting data from independents groups, (b) normally distributed data and (c) variances in the groups are equal (homoscedasticity). However, the specific interest of this study is to focus only for

The contents of the thesis is for internal user only

REFERENCES

Babu, G. J., Padmanabhan, A. R., & Puri, M. L. (1999). Robust one-way ANOVA under possibly non-regular conditions. *Biometrical Journal*, *41*(3), 321-339.

Box, G. E. P. (1953).Non-normality and tests on variances. *Biometrika,* 10, 318-335.

Box, G. E. P. (1954). Some theorem on quadratics forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way model. *Annals of Mathematical Statistics*, 25, 290-302.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology,* 31, 144-152.

Cribbie, R. A., Fiksenbaum, L., Keselman, H. J., & Wilcox, R. R. (2012). Effect of non-normality on test statistics for one-way independent groups designs. *British Journal of Mathematical and Statistical Psychology,* 65, 56-73.

Daniel, W. W. (1990). *Applied Nonparametric Statistics*. Boston: PWS-Kent.

Efron, B. (1979). Bootstrap methods: Another look at the Jacknife. *Annals of Statistics*, 7, 1-26.

Efron, B. &Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall Inc.

Erceg-Hurn, D. M., &Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, 63, 591-601.

Guo, J. H. &Luh, W. M. (2000). An invertible transformation two-sample trimmed *t*-statistics under heterogeneity and nonnormality. *Statistics & Probability letters*, 49, 1-7.

Hampel, F. R. (1974).The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69, 383-393.

Hampel, F. R. (2001).*Robust statistics: A brief introduction and overview*. Invited talk in the Symposium “Robust Statistics and Fuzzy Techniques in Geodesy and GIS” held in ETH Zurich, Mar 12-16, 2001.

Hoaglin, D. C. (1985). Summarizing shape numerically: The *g*- and-*h* distributions. In D. Hoaglin, F. Mosteller, and J. Tukey (eds.), *Exploring data tables, trends, and shapes*. New York: Wiley.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Analysis of Mathematical Statistics,* 38, 33-101.

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown.*Biometrika,* 38, 324-329.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., Kowalchuk, R. K., Lowman, L. L., Petoskey, M. D., Keselman, J. C., & Levin, J. R. (1998). Statistical practices of Educational Researchers: An analysis of their ANOVA, MANOVA and ANCOVA analyses. *Review of Educational Research, 68*(3), 350-386.

Keselman, H. J., Kowalchuk, R. K., Algina, J., Lix, L. M., & Wilcox, R. R. (2000). Testing treatment effects in repeated measures designs: Trimmed means and bootstrapping. *British Journal of Mathematical and Statistical Psychology,* 53, 175-191.

Keselman, H. J., Wilcox, R. R., &Lix, L. M. (2003). A generally robust approach to hypothesis testing in independent and correlated group designs. *Psychophysiology,* 40, 586-596.

Keselman, H. J., Wilcox, R. R., Algina, J., Fradette, K., Othman, A. R. (2004). A power comparison of robust test statistics based on adaptive estimators. *Journal of Modern Applied Statistical Methods*. *3*(1): 27-38.

Keselman, H. J., Wilcox, R. R., Lix, L. M., Algina, J. &Fradette, K. H. (2007). Adaptive robust estimation and testing.*British Journal of Mathematical and Statistical Psychology*, 60, 267-293.

Keselman, H. J., Wilcox, R. R., Othman, A. R., &Fradette, K. (2002). Trimming, transforming statistics, and bootstrapping: circumventing the biasing effects of heteroscedasticity and nonnormality. *Journal of Modern Applied Statistical Methods*, 1, 288-309.

Kohr, R. L. & Games, P. A, (1974). Robustness of the analysis of variance, the Welch procedure, and a Box procedure to heterogenous variances. *Journal of Experimental Education,* 43, 61-69.

Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics and Data Analysis*, 51, 5731-5742.

Levene, H. (1960). Robust testes for equality of variances. *In Contributions to Probability and Statistics* (I. Olkin, ed.) 278-292. Stanford University Press, CA.

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of location equality under heteroscedasticity and non-normality. *Educational and Psychological Measurement.* Math. Proc. *Cambridge Philosophical Society,* 115, 335-363.

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance "*F*" test. *Review of Educational Research,* 66, 579-619.

Manly, B. F. J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology* (*2nd Ed.*). London: Chapman & Hall.

Md Yusof, Z. (2009). Type I error and power rates of robust methods with variable trimmed mean. Unpublished Ph.D. thesis, Universiti Sains Malaysia.

Md Yusof, Z., Abdullah, S., & Syed Yahaya, S. S. (2012a). Type I error rates of parametric, robust and nonparametric methods for two group cases. *World Applied Sciences Journal, 16*(12), 1815-1819.

Md Yusof, Z., Abdullah, S. & Syed Yahaya, S. S. (2012b). Testing the differences of student's scores between two groups. *Journal of Applied Sciences Research*, *8*(9), 4894-4899.

Md Yusof, Z., Harun, N. H., Syed Yahaya, S. S. & Abdullah, S. (2013). A modified parametric bootstrap: an alternative to classical test. *In proceeding of the World Conference on Integration of Knowledge 2013*, Langkawi, Malaysia.

Md Yusof, Z., Othman, A. R., and Syed Yahaya, S. S. (2010). Comparison of Type I error rates between T_1 and F_t statistics for unequal population variance using variable trimming. *Malaysian Journal of Mathematical Sciences*, *4*(2), 195-207.

Mehta, J. S., & Srinivasan, R. (1970). On the Behrens-Fisher problem. *Biometrika*, 57, 549-655.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletion*, 156-166.

Miller, R. L. & Brewer, J. D. (2003).*The A-Z Social Research*. London: SAGE Publications, Ltd.

Muhammad Di, N. F., Syed Yahaya, S. S., & Abdullah, S. (2014). Comparing groups using robust H statistic with adaptive trimmed mean. *Sains Malaysiana*, *43*(4), 643-648.

Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R. & Fradette, K. (2003). An improved Welch-James test statistic.*In proceeding of the Regional Conference on Integrating Technology in the Mathematical Sciences 2003*, Universiti Sains Malaysia, Pulau Pinang, Malaysia.

Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R. & Fradette, K. (2004). Comparing measures of the 'typical' score across treatment groups. *British Journal of Mathematical and Statistical Psychology*, 215-234.

Othman, A. R., Keselman, H. J., Wilcox, R. R., Fradette, K., & Padmanabhan, A. R. (2002). A test of symetri. *Journal of Modern Applied Statistical Methods*, 1, 310-315.

Rocke, D. M., Downs, G. W., &Rocke, A. J. (1982). Are robust estimator really necessary?. *Technometrics*, *24*(2), 95-101.

Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians and trimean. In D. Hoaglin, F. Mosteller& J. Tukey (Eds.). Understanding robust and exploratory data analysis, 297-336. New York: Wiley.

Rousseeuw, P. J., &Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88, 1273-1283.

SAS Institute Inc. (1999): *SAS/IML User's Guide version 8*. Cary. NC: SAS Institute Inc.

Scheffe, H. (1959). *The Analysis of Variance*. New York: Wiley.

Shapiro, S. S. &Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika*, 52, 591-611.

Staudte, R. G. & Sheather, S. J. (1990). *Robust estimation and testing*. New York: Wiley.

Syed Yahaya, S. S. (2005). Robust statistical procedures for testing the equality of central tendency parameters under skewed distributions.Unpublished Ph.D. thesis, Universiti Sains Malaysia.

Syed Yahaya, S. S., Othman, A. R., &Keselman, H. J. (2006). Comparing the "typical score" across independent groups based on different criteria for trimming. *Methodological Papers*, *3*(1), 49-62.

Welch, B. L. (1951). On the comparison of several means: An alternative approach. *Biometrika*, 38, 330-336.

Westfall, P. H. & Young, S. S. (1993). *Resampling-based Multiple Testing*. New York: Wiley.

Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing.* Academic Press, New York.

Wilcox, R. R. (2002). Understanding the practical advantages of modern ANOVA methods. *Journal of Clinical Child and Adolescent Psychology,* 31, 399-412.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd Ed.). San Diego, CA: Academic Press.

Wilcox, R. R., &Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, 8, 254-374.

Wilcox, R. R., & Keselman, H. J. (2010). Modern robust data analysis methods: Measures of central tendency. 1-43.

Wu, M., & Zuo, Y. (2009). Trimmed and Winsorized means based on a scaled deviation. *Journal of Statistical Planning and Inference*, *139*(2), 350-365.

Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2010).*Business research methods* (8th ed.). Thousand Oaks, CA: Thomson/South-Western.

Zimmerman, D. W. (2000). Statistical significance levels of nonparametric tests biased by heterogeneous variances of treatment groups. *Journal of General Psychology,* 127, 354-364.