THE ROBUSTNESS OF *H* STATISTIC WITH HINGE ESTIMATORS AS THE LOCATION MEASURES

NUR FARAIDAH MUHAMMAD DI

MASTER OF SCIENCE (STATISTICS)
UNIVERSITI UTARA MALAYSIA
2014

Permission to Use

In presenting this thesis in fulfilment 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 the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their 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

Dalam menguji kesamaan sukatan lokasi, ujian klasik seperti ujian-*t* dan analisis varians (*ANOVA*) masih lagi di antara prosedur yang biasa dipilih. Prestasi prosedur ini adalah terbaik jika andaian kenormalan data dan kehomogenan varians dipenuhi. Sebarang ketidakpatuhan andaian boleh menjejaskan hasil ujian klasik tersebut. Walau bagaimanapun, dalam kehidupan sebenar, andaian ini sering tidak dipatuhi, oleh yang demikian prosedur teguh menjadi pilihan. Kajian ini mencadangkan dua prosedur teguh dengan mengintegrasikan statistik H bersama min terpangkas suai menggunakan penganggar engsel, *HQ* dan HQ_1. Prosedur yang dicadangkan masing-masing ditandai sebagai $H\hat{T}_{HQ}$ dan $H\hat{T}_{HQ_1}$. Statistik H dikenali dengan keupayaannya mengawal kadar Ralat Jenis I manakala $\hat{T}_{HQ}$ dan $\hat{T}_{HQ_1}$ pula adalah penganggar lokasi teguh. Kaedah min terpangkas suai memangkas data menggunakan kaedah pemangkasan asimetrik, dengan hujung taburan dipangkas berdasarkan ciri-ciri taburan tersebut. Untuk mengkaji prestasi (keteguhan) prosedur, beberapa pemboleh ubah dipergunakan untuk menghasilkan keadaan yang dapat menyerlahkan kekuatan dan kelemahan prosedur. Pemboleh ubah tersebut adalah jumlah pemangkasan, bilangan kumpulan, saiz sampel yang seimbang dan tidak seimbang, jenis taburan, keheterogenan varians dan sifat pasangan. Kaedah Bootstrap telah digunakan untuk menguji hipotesis disebabkan taburan statistik H adalah tidak diketahui. Integrasi antara statistik H dan min terpangkas suai menghasilkan prosedur teguh yang mampu menangani masalah ketidakpatuhan andaian. Hasil kajian menunjukkan prestasi prosedur yang dicadangkan adalah terbaik dalam mengawal kadar Ralat Jenis I dengan jumlah pangkasan yang berbeza; prestasi $H\hat{T}_{HQ}$ adalah terbaik dengan pangkasan 20%, manakala 15% adalah terbaik untuk $H\hat{T}_{HQ_1}$. Di samping itu, kedua-dua prosedur terbukti lebih teguh jika dibandingkan dengan ujian parametrik klasik (ujian-*t* dan *ANOVA*) dan ujian tak berparameter (Mann Whitney dan Kruskal-Wallis).

Kata kunci: Statistik teguh, Statistik *H*, Penganggar engsel, Kadar ralat jenis I, Pemangkasan asimetrik.

Abstract

In testing the equality of location measures, the classical tests such as *t*-test and analysis of variance (ANOVA) are still among the most commonly chosen procedures. These procedures perform best if the assumptions of normality of data and homogeneity of variances are fulfilled. Any violation of these assumptions could jeopardize the result of such classical tests. However, in real life, these assumptions are often violated, and therefore, robust procedures may be preferable. This study proposed two robust procedures by integrating *H* statistic with adaptive trimmed mean using hinge estimators, *HQ* and HQ_1. The proposed procedures are denoted as $H\hat{T}_{HQ}$ and $H\hat{T}_{HQ_1}$, respectively. *H* statistic is known for its ability to control Type I error rates while $\hat{T}_{HQ}$ and $\hat{T}_{HQ_1}$ are the robust location estimators. The method of adaptive trimmed mean trims data using asymmetric trimming technique, where the tail of the distribution is trimmed based on the characteristic of that particular distribution. To investigate on the performance (robustness) of the procedures, several variables were manipulated to create conditions which are known to highlight its strengths and weaknesses. Such variables are the amount of trimming, number of groups, balanced and unbalanced sample sizes, type of distributions, variances heterogeneity and nature of pairings. Bootstrap method was used to test the hypothesis since the distribution of *H* statistic is unknown. The integration between *H* statistic and adaptive trimmed mean produced robust procedures that are capable of addressing the problem of violations of the assumptions. The findings showed that the proposed procedures performed best in terms of controlling the Type I error rate with different trimming amounts; the $H\hat{T}_{HQ}$ performed best with 20% trimming, while 15% was best for the $H\hat{T}_{HQ_1}$. In addition, both procedures were also proven to be more robust than the classical tests of parameteric (*t*-test and ANOVA) and non-parametric (Mann Whitney and Kruskal-Wallis).

Keywords: Robust statistics, *H* statistic, Hinge estimator, Type I error rates, Asymmetric trimming.

Acknowledgement

بسم الله الرحمن الرحيم

Alhamdulillah, praised to Allah S. W. T for giving me this opportunity to complete this thesis. This journey gave me priceless experiences, helped me to know myself better and improve myself in so many aspects.

This thesis would not have been materialized without the support of several individuals. First, I am deeply indebted to my supervisor Associate Prof. Dr. Sharipah Soaad Syed Yahya for her patient, support, guidance and encouragement regarding my work. She has been most helpful throughout my entire study. I also like to thanks my co supervisor Dr. Suhaida Abdullah who gave me advices and critical review on my works.

I also like to take this opportunity to express my gratitude to those who always support me from behind, especially my father; Muhammad Di Ishak, my siblings; Wijyawati, Rosmanizam, Noorfarizan and my beloved sister, Allahyarhamah Norfaidatul Akma. Without support from them, I will never have the courage to complete this work.

Last but not least, my appreciation goes to Universiti Utara Malaysia and the individuals and parties that involves directly or indirectly throughout this study.

Thank you very much.

Table of Contents

CHAPTER THREE

CHAPTER FOUR

List of Tables

List of Figures

List of Appendices

Glossary of Terms

Analysis of Variance (ANOVA): A statistical technique which helps in making inference whether three or more samples might come from populations having the same mean.

Bias: A systematic (not random) deviation from the true value.

Bootstrapping: Bootstrapping is sampling with replacement from observed data to estimate the variability in a statistic of interest.

Classical Tests: In statistical inference procedures (hypothesis tests and confidence intervals), classical tests are those that incorporate assumptions about population parameters.

Estimator: A statistic, measure, or model, applied to a sample, intended to estimate some parameter of the population that the sample came from.

Heteroscedasticity: Heteroscedasticity generally means unequal variation of data, e.g. unequal variance.

Homoscedasticity: Homoscedasticity generally means equal variation of data, e.g. equal variance.

Kruskal-Wallis Test: Nonparametric test for finding if three or more independent samples come from populations having the same distribution.

Kurtosis: Kurtosis measures the "heaviness of the tails" of a distribution (in compared to a normal distribution).

Mann Whitney Test: Nonparametric test for finding if two independent samples come from populations having the same distribution.

Mean: For a population or a sample, the mean is the arithmetic average of all values. The mean is a measure of central tendency or location.

Nonparametric Tests: In statistical inference procedures (hypothesis tests and confidence intervals), nonparametric procedures are those that are relatively free of assumptions about population parameters.

Normal Distribution: The normal distribution is a probability density which is bell-shaped, symmetrical, and single peaked. The mean, median and mode coincide and lie at the centre of the distribution.

Robustness: The robustness of a statistical method is its insensitivity to departures from classical test assumptions.

Skewness: Skewness measures the lack of symmetry of a probability distribution.

***t*-test:** A statistical technique which helps in making inference whether two samples might come from populations having the same mean.

Trimmed Mean: The trimmed mean is computed by sorting all the N discarding the percentages of the smallest and percentages of the largest values, and computing the mean of the remaining values.

Type I Error: Type I error is the error of rejecting the null hypothesis when it is true.

List of Abbreviations

ANOVA	Analysis of Variance
H	Robust test to measure the equality of central tendency
MOM	One-step *M*-estimator
SAS	Statistical Analysis Software
SAS/IML	Statistical Analysis Software/ Interactive Matrix Language
SPSS	Statistical Package for Social Science

CHAPTER ONE
INTRODUCTION

1.1 Introduction

In recent years, there have been extensive studies regarding the test on the equality of central tendency measures in terms of their robustness. However, researchers in the field of social sciences, economics, and business for example are still attached to the classical tests which are available in all the statistical packages in the market. When testing the differences between groups, *t*-test and Analysis of Variance (ANOVA) will be the most commonly chosen methods. ANOVA and *t*-test have several assumptions that need to be fulfilled before the procedure can be applied. The main assumptions are such that the data should be normally distributed and variances are homogeneous. Unaware to most of them, these tests are unreliable and produce misleading results when there are any violations in the assumptions (Mantalos, 2010).

1.2 Problem Statement

As mentioned in previous subsection, ANOVA and *t*-test have underlying assumptions that need to be fulfilled. In real situation, these assumptions are often violated and to obtain the ideal data which satisfy all the assumptions are hardly achieved (Wu, 2007). Figure 1.1 and Figure 1.2 depicts the example of violations of the classical assumptions in real life. Figure 1.1 is the distribution of the ship's service time before privatization with mean 14.48 hours. Figure 1.2 is the distribution of the ship's service time after privatization with mean 15.95 hours. To test the equality of central tendency measures between the two groups, *t*-test is an

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*, 321-339.

Baguio, C. B. (2008). Trimmed mean as an adaptive robust estimator of a location parameter for Weibull Distribution. *World Academy of Science, Engineering and Technology*, 681-686.

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

Erceg-hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods. *American Psychologist*, 591-601.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics, 14*(4), 1453-1462.

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, 909-923.

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

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, *59*(4), 434-446.

Keselman, H. J., Wilcox, R. R., Algina, J., Othman, A. R., & Fradette, K. (in press). A comparative study of robust tests for spread: asymmetric trimming strategies. *British Journal of Mathematical and Statistical Psychology, 60,* 267-293.

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

Keselman, H. J., Algina, J., & Fradette, K. (2005). Robust confidence intervals for effect size in the two-group case. *Journal of Modern Applied Statistical methods, 4*(2), 353-371.

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

Keselman, H. J., Wilcox, R. R., Othman, A. R., & Fradette, K. (2002). Trimming, transforming statistics, and bootstrapping: Circumventing the biasing effects of heteroscedasticity and non-nomality. *Journal of Modern Applied Statistical Methods*, *1*(2), 288–399.

Keselman, H. J., Wilcox, R. R., Taylor, J., & Kowalchuk, R. K. (2000). Tests for mean equality that do not require homogeneity of variances: Do they really work? *Communications in Statistics - Simulation and Computation, 29*(3), 875–895.

Keselman, H. J., Lix, L. M., & Kowalchuk, R. K. (1998). Multiple comparison procedures for trimmed means. *Psychological Methods, 3*, 123-141.

Khan, A., & Rayner, G. D. (2003). Robustness to non-normality of common tests for the many-sample location problem. *Journal of Applied Mathematics and Decision Sciences, 7*(4), 187-206.

Luh, W. -M., & Guo, J. -H. (1999). A powerful transformation trimmed mean method for one-way fixed effects ANOVA model under non-normality and inequality of variances. *British Journal of Mathematical and Statistical Pscycology*, 303-320.

Mantalos, P. (2010). *Three different measures of sample skewness and kurtosis and their effects on the Jargue-Bera test for normality*. Discussion paper, Jönköping International Business School.

Md. Nor, N. G., Ohta, K., & Mohd Yusoff, M. F. (2003). The impact of privatization on efficiency: Empirical evidence from a Malaysian Port. *Journal of the Eastern Asia Society for Transportation Studies, 5*, 2883-2892.

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

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*, *57*(2), 215-234.

Reed, J. F., & Stark, D. B. (1996). Hinge estimators of location: Robust to asymmetry. *Computer Methods and Programs in Biomedicine*, 11-17.

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

Schrader, R. M., & Hettmansperger, T. P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika*, *67*(1), 93-101.

See Zar, J. H. (1996). *Biostatistical Analysis (3rd ed.)*. Upper Saddle River, NJ: Prentice-Hall.

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

Syed Yahaya, S. S. (2005). *Robust statistical procedures for testing the equality of central tendency parameters under skewed distribution*. (Unpublished doctoral dissertation). Universiti Sains Malaysia, Pulau Pinang.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th ed.).* Boston: Allyn and Bacon.

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing (2nd ed.).* San Diego: Academic Press.

Wilcox, R. R., & Tian, T. S. (2011). Measuring effect size: A robust heteroscedastic approach for two or more groups. *Journal of Applied Statistics, 33*, 1359-1368.

Wilcox, R. R. (2010). *Fundamentals of modern statistical methods: Substantially improving power and accuracy (2nd ed.).* New York: Springer.

Wilcox, R. R. (2005). Trimmed means. *Encyclopedia of Statistics in Behavioral Science*, *4*, 2066-2067.

Wilcox, R. R., & Keselman, H. J. (2003). Repeated measures ANOVA based on a modified one-step *M*-Estimator. *Journal of British Mathematical and Statistical Psychology*, *56*(1), 15–26.

Wilcox, R. R., Keselman, H. J., & Kowalchuk, R. K. (1998). Can test for treatment group equality be improved?: The bootstrap and trimmed means conjecture. *British Journal of Mathematical and Statistical Psychology, 51,* 123-134.

Wu, P. -C. (2007). Modern one-way ANOVA F methods: Trimmed means, one step M-estimators and bootstrap methods. *Journal of Quantitative Research*, 155-172.

Yaacob, C. R., Syed Yahaya, S. S., & Othman, A. R. (2006). *The dual role of two scale estimators*. Paper presented at the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, The Gurney Hotel, Pulau Pinang, Malaysia.

Yu, C. -H. (2012). *Parametric tests*. Retrieved January 4, 2012, from Creative-wisdom: http://www.creative-wisdom.com/teaching/WBI/parametric_test.shtml.