MODIFICATION OF S_1 STATISTIC WITH HODGES-LEHMANN AS THE CENTRAL TENDENCY MEASURE

LEE PING YIN

MASTER OF SCIENCE (STATISTICS)
UNIVERSITI UTARA MALAYSIA
2018

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)

LEE PING YIN

calon untuk Ijazah
(candidate for the degree of) **MASTER OF SCIENCE (STATISTICS)**

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

"MODIFICATION OF S1 STATISTIC WITH HODGES-LEHMANN AS THE CENTRAL TENDENCY MEASURE"

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 : ***08 Mei 2018.***
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: ***May 08, 2018.***

Pengerusi Viva: *(Chairman for VIVA)*	**Assoc. Prof. Dr. Nazihah Ahmad**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Assoc. Prof. Dr. Zulkifley Mohamed**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Assoc. Prof. Dr. Zahayu Md Yusof**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Prof. Dr. Sharipah Soaad Syed Yahaya**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Nor Aishah Ahad**	Tandatangan *(Signature)*

Tarikh:
(Date) ***May 08 2018***

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University's 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

Andaian kenormalan dan kehomogenan varians adalah merupakan perkara penting bagi prosedur parametrik seperti dalam pengujian kesamaan kecendurangan memusat. Sebarang ketidakpatuhan andaian tersebut boleh meningkatkan kadar Ralat Jenis I yang serius, yang akan mengakibatkan penolakan hipotesis nol yang tidak betul. Prosedur parametric seperti *ANOVA* dan ujian-*t* sangat bergantung pada andaian yang sukar ditemui dalam data sebenar. Sebaliknya, prosedur tak berparameter tidak bergantung pada taburan data tetapi prosedur tersebut kurang kuasanya. Untuk mengatasi isu yang dinyatakan, prosedur teguh adalah dicadangkan. Statistik S_1 adalah salah satu prosedur teguh yang menggunakan median sebagai parameter lokasi untuk menguji kesamaan kecenderungan memusat di antara kumpulan, dan ia membabitkan data asal tanpa perlu memangkas atau mentransformasi data untuk mencapai kenormalan. Kajian terdahulu terhadap S_1 menunjukkan kekurangan keteguhan dalam beberapa keadaan di bawah reka bentuk seimbang. Oleh itu, objektif kajian ini adalah menambahbaik statistik S_1 asal dengan menggantikan median kepada penganggar *Hodges-Lehmann*. Penggantian juga dilakukan terhadap penganggar skala menggunakan varians bagi penganggar *Hodges-Lehmann* serta beberapa penganggar skala teguh yang lain. Bagi memeriksa kekuatan dan kelemahan prosedur yang dicadangkan dalam mengawal Ralat Jenis I, beberapa pemboleh seperti jenis taburan, bilangan kumpulan, saiz kumpulan yang seimbang dan tidak seimbang, varians yang sama dan tidak sama, dan sifat pasangan telah dimanipulasikan. Hasil kajian menunjukkan kesemua prosedur yang dicadangkan adalah teguh merentasi semua keadaan bagi setiap kes kumpulan. Selain itu, tiga prosedur yang dicadangkan iaitu $S_1(MAD_n)$, $S_1(T_n)$ dan $S_1(S_n)$ menunjuk prestasi yang lebih baik berbanding prosedur S_1 asal di bawah taburan pencong yang ekstrem. Secara keseluruhan, prosedur yang dicadangkan menunjukkan keupayaannya mengawal peningkatan Ralat Jenis I. Oleh yang demikian, objektif kajian ini telah tercapai apabila tiga daripada prosedur yang dicadangkan menunjukkan peningkatan keteguhan di bawah taburan terpencong.

Katakunci: Statistik S_1, *Hodges-Lehmann*, penganggar skala teguh, ralat Jenis I, taburan terpesong

Abstract

Normality and variance homogeneity assumptions are usually the main concern of parametric procedures such as in testing the equality of central tendency measures. Violation of these assumptions can seriously inflate the Type I error rates, which will cause spurious rejection of null hypotheses. Parametric procedures such as ANOVA and *t*-test rely heavily on the assumptions which are hardly encountered in real data. Alternatively, nonparametric procedures do not rely on the distribution of the data, but the procedures are less powerful. In order to overcome the aforementioned issues, robust procedures are recommended. S_1 statistic is one of the robust procedures which uses median as the location parameter to test the equality of central tendency measures among groups, and it deals with the original data without having to trim or transform the data to attain normality. Previous works on S_1 showed lack of robustness in some of the conditions under balanced design. Hence, the objective of this study is to improve the original S_1 statistic by substituting median with Hodges-Lehmann estimator. The substitution was also done on the scale estimator using the variance of Hodges-Lehmann as well as several robust scale estimators. To examine the strengths and weaknesses of the proposed procedures, some variables like types of distributions, number of groups, balanced and unbalanced group sizes, equal and unequal variances, and the nature of pairings were manipulated. The findings show that all proposed procedures are robust across all conditions for every group case. Besides, three proposed procedures namely $S_1(MAD_n)$, $S_1(T_n)$ and $S_1(S_n)$ show better performance than the original S_1 procedure under extremely skewed distribution. Overall, the proposed procedures illustrate the ability in controlling the inflation of Type I error. Hence, the objective of this study has been achieved as the three proposed procedures show improvement in robustness under skewed distributions.

Keywords: S_1 statistic, Hodges-Lehmann, robust scale estimators, Type I error, skewed distributions.

Acknowledgement

First of all, I would like to thank God for giving me the chance to complete the thesis which I have spent five years in studying Master of Sciences (Statistics) as a part time student. This is truly a blessing to me. Besides, I would like to extend my appreciation to my supervisor, Associate Professor Dr. Sharipah Soaad Syed Yahya and co-supervisor, Dr Aishah Ahad who have given continuous guidance, patience and support to me. They always be there for me whenever I face difficulty in my journey of writing my thesis and generating data using statistical computer software. I appreciate their help so much. In addition, I would like to thank Universiti Utara Malaysia (UUM) too for approving my master study's application and few staffs in Awang Had Salleh Graduate School who assisted me in the process of submission.

I am deeply grateful to my family, my fiance, Chan Jin Swan and my best undergraduate roommate, Nurull Salmi Md Dazali, my Indonesia friend, Fera, my colleague, Wong Sock Leng and my mentor, Wern Lu who encourage me throughout this study by giving infinite motivation. Due to all the support I get, I manage to complete my study. I would like to give my grateful appreciations to all of them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA	Analysis of Variance
HL	Hodges-Lehmann Estimator
MAD_n	Median Absolute Deviation about the Median
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 most research, hypothesis testing has been used as a method of decision making with the help of primary and secondary data that can be obtained from sources such as observations, experiments, journals, articles, reference books and many other sources. The researchers are required to identify the statement of null hypothesis which is usually corresponds to a situation of equality or "no difference" and it is assumed as true hypothesis until receiving an evidence that shows otherwise. Alternative hypothesis is known as the negation of null hypothesis (Sullivan, 2004). Due to the statistical nature of a test, two types of error are determined, Type I error and Type II error. Type I error occurred in the situation where by the null hypothesis is rejected when it is true. In contrast, Type II error existed when the null hypothesis is failed to reject when it is false. There is an inverse relationship between the two errors such that an increase in Type I error will decrease Type II error and vice versa. Furthermore, when Type II error increases, the statistical power of a test will decrease, causing less detection of a test effect. Thus, these two errors need to be in control. A good statistical procedure should be able to control the errors. However, working with Type I error is easier than Type II error as the earlier is usually set in advance by the researcher while the latter is harder to know as it requires estimating the distribution of the alternative hypothesis (Ramsey, 2001).

In order to achieve a good test, we need an appropriate procedure which is able to control Type I error rate and increase the power at the same time. We do not want to

The contents of the thesis is for internal user only

REFERENCES

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

Bickel, P. J. (1965). On some robust estimates of location. *The Annals of Mathematical Statistics*, *36*(3), 847-858.

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

Boos, D. D. (1982). A test for asymmetry associated with the Hodges-Lehmann estimator. *Journal of the American Statistical Association*, *77*(379), 647-651.

Boos, D. D. and Monahan, J. F. (1986). Bootstrap methods using prior information. *Biometrika*, *73*(1), 77-83.

Box, G. E. P. (1953). Non-normality and tests on variance. *Biometrika*, *40*, 318-355.

Brownie, C and Boss, D. D. (1994). Type I error robustness of ANOVA and ANOVA on ranks when the number of treatments is large. *Biometrics*, *50*(2), 542-549.

Brunner, L. J. and Austin P. C. (2007). Inflation of Type I error in multiple regression when independent variables are measured with error. Retrieved from http://www.utstat.toronto.edu/~brunner/MeasurementError/MeasurementError3i.pdf

Chernick, M. R. (1999). *Bootstrap Methods: A Practitioner's Guide*. New Jersey, US: John Wiley & Sons, Inc.

Cui, H., He, X. and Ng, K. W. (2003). Asymptotic distributions of principal components based on robust dispersions. *Biometrika*, *90*(4), 953-966.

Dixon, W. J. and Yuen, K. K. (1974). Trimming and winsorization: a review. *Statistiche Hefte*, *2*, 157-170.

Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. *In A Festscherift for Erich L. Lehmann* (Brickel, P. J., Doksum, K. & Hidges, J. L., Jr., eds), 157-184.

Fisher, R. A. (1935). *The Design of Experiments*. New York, US: Hafner.

Geyer, C. J. (2006). Breakdown point theory notes. Retrieved from http://www.stat.umn.edu/geyer/5601/notes/break.pdf

Gibbons, J. D. and Chakraborti, S. (2003). *Nonparametric Statistical Inference*, 4th ed. Florida, US: CRC Press.

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

Hall, P. and Sheather, S. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society*, 381-391.

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., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York, US: Wiley.

Hodges, J. L. Jr. and Lehmann, E. L. (1963). Estimated of location based on rank tests. *The Annals of Mathematical Statistics*, *34*(2), 598-611.

Hogg, R. V. (1974). Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, *69*, 909-927.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73-101.

Huber, P. J. (1972). Robust statistics: a review. *The Annals Mathematical Statistics*, *43*(4), 1041-1067.

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

Kang, Y. and Harring, J. R. (2012). Investigating the impact of non-normality, effect size, and sample size on two-group comparison procedures: an empirical study. *Presented at the Annual Meeting of the American Educational Research Association (AERA),* SIG: Educational Statisticians, BC.

Kazempour, K. (1995). Impact of stratification imbalance on probability of Type I error. *The American Statistician*, *49*(2), 170-174.

Kellermann, A. P., Bellara, A. P., Gil, P. R. D., Nguyen, D., Kim, E. S., Chen, Y. and Kromrey, J. D. (2013). Variance heterogeneity and non-normality: how SAS PROC TEST can keep us honest. Retrieved from http://support.sas.com/resources/papers/proceedings13/228-2013.pdf

Keselman, H. J., Wilcox, R. R., Lix, L. M., Algina. J. and Fradette, K. (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. and 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.

Khan, A. and 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.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*, 583-621.

Kulinskaya, E., Staudte, R. G., and Gao, H. (2003). Power approximations in testing for unequal means in a one-way ANOVA weighted for unequal variances. *Communications in Statistics – Theory and Methods*, *32*, 2353-2371.

Lix, L. M. and Keselman, H. J. (1998). To trim or not to trim: tests of location equality under heteroscedasticity and nonnormality. *Educational and Psychological Measurement*, *58*, 409-429.

Lix, L. M., Keselman, J. C. and 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 Education Research*, *66*(4), 579-619.

MacGillivray, H. L. and Cannon, W. H. (2002). *Generalizations of the g-and-h distributions and their uses*. Unpublished Thesis.

Maxwell, S. E. and Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, 2nd ed. Mahwah, NJ, US: Lawrence Erlbaumn Associates Publishers.

Md. Yusof, Z., Othman, A. R. and Syed Yahaya, S. S. (2010). Comparison of Type I error rates between T1 and Ft statistics for unequal population variance using variable trimming. *Malaysian Journal of Mathematical Sciences*, *4*(2), 195-207.

Morgenthaler, S. (1992). Least-absolute-deviations fits for generalized linear models. *Biometrika*, *79*, 747-754.

Nachar, N. (2008). The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, *4*(1), 13-20.

Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R. and Fradette, K. (2004). Comparing measures of the "typical" score across treatment groups. *British Journal of Mathematical and Statistical Psychology, 57*(2), 215-234.

Oshima, T. C. and Algina, J. (1992). Type I error rates for James's second-order test and Wilcox's Hm test under heteroscedasticity and non-normality. *British Journal of Mathematical and Statistical Psychology*, *42*, 255-263.

Ott, R. L. and Longnecker, M. T. (2010). *An Introduction to Statistical Methods and Data Analysis.* Belmont, CA: Brooks/Cole Cengage Learning.

Pitman, E. J. G. (1937a). Significance tests which may be applied to samples from any population. *Royal Statistical Society Supplement*, *4*(1-2), 119-130.

Pitman, E. J. G. (1937b). Significance tests which may be applied to samples from any population. *Royal Statistical Society Supplement*, *4*(2), 225-232.

Pitman, E. J. G. (1938). Significance tests which may be applied to samples from any population: the analysis of variance test. *Biometrika*, *29*(3-4): 332-335.

Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society*, *11*(1), 68-84.

Ramsey, J. B. (2001). *The Elements of Statistics: with Applications to Economics and the Social Sciences,* 1st ed. Boston, US:Cengage Learning.

Reed, J. F. (1998). Contributions to adaptive estimation. *Journal of Applied Statistics*, *25*(5), 651-669.

Ronchetti, E. M. (2006). The historical development of robust statistics. *In Proceedings of the 7th International Conference on Teaching Statistics (ICOTS-7).* Retrieved from https://iase-web.org/documents/papers/icots7/3B1_RONC.pdf

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

SAS Institute. (1989). *IML software: Usage and Reference, Version 6*, 1st ed. Cary, NC: SAS Institute.

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

Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures*, 5th ed. Florida, US: CRC Press.

Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods*, 7th ed. U.S.A, Ames:Iowa State University Press.

Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation and Testing*. New York, US: Wiley.

Stigler, S. M. (1977). Do robust estimators work with real data? *The Annals of Statistics*, *5*(6), 1055-1098.

Stigler, S. M. (2010). The changing history of robustness. *The American Statistician, 64*(4), 277-281.

Sullivan, M. III (2004). *Statistics: Informed Decisions Using Data*. Upper Saddle River, NJ: Pearson Education, Inc.

Syed Yahaya, S. S. (2005). Robust statistical procedures for testing the equality of central tendency parameters under skewed distributions. Unpublished Thesis.

Syed Yahaya, S. S., Md Yusof, Z. and Abdullah, S. (2011). A robust alternative to ANOVA. *Proceedings of the Second International Soft Science Conference 2011 (ISSC 2011)*. Ho Chi Minh City, Vietnam.

Syed Yahaya, S. S., Othman, A. R. and Keselman, H. J. (2004). Testing the equality of location parameters for skewed distributions using S_1 with high breakdown robust scale estimators. *Theory and Applications of Recent Robust Methods, Series: Statistics for industry and technology*, 319 – 328.

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

Teh, S. Y., Md Yusof, Z., Yaacob, C. R. & Othman, A. R. (2010). Performance of the traditional pooled variance t-test against the bootstrap procedure of difference between sample means. *Malaysian Journal of Mathematical Sciences*, *4*(1), 85-94.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. *In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (I. Olkin *et al*., eds), 448-485. Stanford University Press.

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

Wilcox, R. R. (1994). A one-way random effects model for trimmed means. *Psychometrika*, *59*(3), 289-306.

Wilcox, R. R. (1997). ANCOVA based on comparing a robust measure of location at empirically determined design points. *British Journal of Mathematical and Statistical Psychology*, *50*(1), 93-103.

Wilcox, R. R. and Keselman, H. J. (2002). Power analysis when comparing trimmed means. *Journal of Modern Applied Statistical Methods*, *1*(1), 24-31.

Wilcox, R. R. and Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, *8*, 254-274.

Wilcox, R. R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*, 3rd ed. San Diego, CA: Academic Press.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*, 80-83.

Winter, J. C. F. D. and Dodou, D. (2010). Five-point likert items: t test versus mann-whitney-wilcoxon. Retrieved from http://pareonline.net/pdf/v15n11.pdf

Yi, G. Y. and He, W. (2009). Median regression models for longitudinal data with dropouts. *Biometrics*, *65*(2), 618-625.

Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, *61*, 165-170.