UNIVERSITI PUTRA MALAYSIA

MEDIAN POLISH TECHNIQUES FOR ANALYSING PAIRED DATA

IDRIS AJOGE

FS 2017 3

MEDIAN POLISH TECHNIQUES FOR ANALYSING PAIRED DATA

By

IDRIS AJOGE

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfilment of the Requirements for the Degree of
Master of Science

January 2017

DEDICATIONS

I would like to dedicate this dissertion work to

- *Almighty Allah, the benevolent and the merciful.*
- *My respectful parents who have taught me a lot on the persistency in life.*
- *My beloved wife for all her contribution, patience and understandings throughout my studies. She incredibly supported me and made it all possible for me.*
- *My chidren for their patience and endurance and their love have always been my greatest inspiration.*
- *My late younger brother Aminu S. Ajoge, whom I lost during the course of my study*

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

MEDIAN POLISH TECHNIQUES FOR ANALYSING PAIRED DATA

By

IDRIS AJOGE

January 2017

Chairman: Mohd Bakri Adam, PhD
Faculty: Science

Median polish is used as a data analysis technique for examining the significance of various factors in single or multi-way models. The main goal of this research is to analyze paired data using median polish technique in order to get information from the data such as difference between the paired data through column and row effects. Median polish algorithm is useful in removing any noise in the data by computing medians for various coordinates on the data set. In this research, we will focus on paired data. Some effects such as overall, rows and column were determined using median polish algorithm. In this research, one-way and two-way median polish has been expanded for pairwise scenario. The pairwise median polish criterion addresses the fairness of declaring a difference between a paired data. Paired data involves collection of data prior to the treatment and compare it with after the treatment. We extend the analysis of paired data for the case of missing values. Later, exercising of comparison values in a two-way median polish for paired data was implemented to verifying the association between rows and columns effects. In addition, to determine whether there is need for transformation of data or not. Pairwise median polish model is successfully employed in the analysing the comparison and verification of the difference between paired data of grain yields in classification of contingency table. For the two-way median polish for paired data, comparison values calculated shows there is no association between rows and columns effects and transformation of data is not required in this study. The median polish provides simple estimation of main effects for paired data as well as various factor effects. The findings also have shown that there is a difference in grand effects of both after treatment data without missing values and imputed values using paired median polish procedure.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Sarjana Sains

KAEDAH PENGILAP MEDIAN UNTUK MENGANALISIS DATA BERPASANGAN

Oleh

IDRIS AJOGE

Januari 2017

Pengerusi: Mohd Bakri Adam, PhD
Fakulti: Sains

Penggilapan median digunakan sebagai teknik analisis data untuk memeriksa kepentingan pelbagai faktor di dalam model sehala ataupun model pelbagai hala. Matlamat utama bagi kajian ini adalah untuk menganalisis data berpasangan dengan menggunakan teknik penggilapan median untuk mendapatkan maklumat daripada data seperti perbezaan antara data berpasangan melalui kesan pada lajur dan baris. Algoritma penggilapan median sangat berkesan bagi menghapuskan mana-mana gangguan di dalam data dengan pengiraan pelbagai median koordinat pada set data. Kajian ini memberi tumpuan kepada data berpasangan. Beberapa kesan seperti kesan keseluruhan, kesan pada baris dan lajur ditentukan menggunakan algoritma penggilapan median. Bagi kajian ini, penggilapan median satu hala dan dua hala telah diperluaskan kepada senario berpasangan. Kriteria penggilap median berpasangan dapat memberikan pembuktian di dalam perbezaan antara data berpasangan. Data berpasangan telah melibatkan pengumpulan data bagi sebelum rawatan untuk di bandingkan dengan data selepas rawatan. Kajian ini juga telah memperluaskan analisis data berpasangan kepada kes yang melibatkan data yang tidak lengkap. Kemudian, perbandingan nilai di dalam penggilapan median dua hala bagi data berpasang telah dilaksanakan untuk pengesahan hubungan antara kesan pada baris dan kesan ada lajur. Di samping itu, kajian ini juga menentukan sama ada terdapat keperluan untuk sesuatu data itu di ubahsuai mahupun tidak. Model penggilapan median berpasangan telah berjaya di dalam penganalisaan perbandingan dan pengesahan perbezaan antara data berpasangan daripada hasil bijirin dengan pengelasan jadual kontingensi. Bagi penggilapan median dua hala untuk data berpasang, nilai perbandingan yang dikira menunjukkan ianya tiada kaitan di antara kesan pada baris dengan kesan pada lajur dan pengubahsuaian data tidak diperlukan di dalam kajian ini. Penggilapan median telah menyediakan penganggaran yang mudah untuk kesan yang utama bagi data berpasangan dan juga pelbagai kesan faktor. Dapatan daripada kajian ini juga menunjukkan bahawa terdapat perbezaan di dalam kesan yang besar daripada kedua-dua data yang lengkap selepas kehilagan data selepas rawatan bersama nilai-nilai yang ditambah menggunakan prosedur penggilapan median berpasangan.

ACKNOWLEDGEMENTS

First, I would like to thank and most gratefully pay my gratitude to Almighty Allah, the Most Gracious, the Most Merciful for bestowing His countless blessings all the way through my life.

I want to express my profound appreciation and cordial respect to my honorable supervisor, Dr Mohd Bakri Bin Adam whose persistent supervision and unrestricted support has made this research work possible. His genuine guidance and sincere assistance has enabled me to learn to carry out research and materialize my thesis. I am also deeply thankful for his ever-ready support, which has always been a source of immense encouragement all through my research work. He took all the patience to ask me good questions to help me think through my research problems, either philosophical or computational. I feel truly fortunate and highly privileged to have worked under his priceless supervision, guidance and attention.

My special thanks is for my co-supervisor Dr. Anwar Fitrianto, It has been a pleasure to work with him. With his encouragement, constant guidance and moral assistance, I was able to finish this my research work. He was always there to meet and talk about my ideas. My appreciation goes to my examiners, Dr Lee Lai Soon (Chairman) and Dr Mahendran Shitan (Internal Examiner) for their professional guidance, not only in this thesis but also with career decisions and imparting valuable knowledge of learning to me.

I feel obligated to pay great debt of thanks to my dear family especially my wife, my mum and my children whose moral, patience, emotional, intellectual and numerous sacrifices enabled me to finish this programme. I am entirely grateful to: Prof. M.M. Barau, Dr A.Y. Sadiq, Dr Suleiman B. Sa'ad, Mall. Shaibu Ireyi Aroke, Chief Samuel E. Onyeamachi, Alh. & Mrs Abdulraheem Dada Momoh, Dr Siaka Abdulrahaman, Dr David Gbenyi, Hajia Maimunat Ibrahim, Late Engr. Sani Dahiru Buba (may his soul rest in perfect peace), Late Aminu S. Ajoge, Engr. E. Akainji and Qasim H. N., for their roles in completion of this programme and helped me to meet this challenge with the help of their moral & financial supports, and continuous encouragement.

I am also indebted to: Mr & Mrs Muyiwa S. Olaoye, Hajia F. N. Mohammed, Mr. David Adekoya, Mr. Gadzama M. A, Mall. Yusufu Sani, Alh. Salau S. Ajoge, Yusuf S. Ajoge, Dr. Samuel Adinoyi Ojo (MAUTECH), Mall. Maksha Abdu Ibrahim, Dr Simon Stephen, Abdulwahab Ibrahim, Dr Dauda M. Belel, Sani Lawan (Kadpoly), Dr Aisha Sha'ffi, Hajia Maryam Ahmed, Abdullahi A. Musa, Lawan B. Mohd, Dr Sirajo Lawan Bichi, Buba Ahmed Tarajo, Ibrahim B. Babura, Dr. Abdulazeez Sikiru Adeyinka, Alh. Halilu Gidado and my other family members and friends who contributed positively to the completion of this programme.

I also wish to thank the staff of Ministry of Agricultural and Farm Extension Adamawa state of Nigeria for providing the data I used in this thesis. Finally, I am thankful to all my teachers and friends of the Department of Mathematics, UPM in Malaysia and Staff of Statistics department, Federal polytechnic Mubi, Adamawa state of Nigeria for their physical and moral support.

I certify that a Thesis Examination Committee has met on 23 January 2017 to conduct the final examination of Idris Ajoge on his thesis entitled "Median Polish Techniques for Analysing Paired Data" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

Lee Lai Soon, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Mahendran a/l S.Shitan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Fadhilah binti Yusof, PhD
Associate Professor
Universiti Teknologi Malaysia
Malaysia
(External Examiner)

NOR AINI AB. SHUKOR, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 22 March 2017

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science.
The members of the Supervisory Committee were as follows:

Mohd Bakri Adam, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Anwar Fitrianto, PhD
Senior Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature:____________________ Date:____________________

Name and Matric No: Ajoge Idris, GS 39830

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________________
Name of Chairman of Supervisory Committee
Mohd Bakri Adam, PhD

Signature: ______________________________
Name of Member of Supervisory Committee
Anwar Fitrianto, PhD

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EDA	Exploratory Data Analysis
PC	Paired Comparison
PMP	Pairwise Median Polish
MPP	Median Polish Procedure
CM	Column Median
RM	Row Median
CE	Column Effect
RE	Row Effect
OE	Overall Effect
MCAR	Missing Completely at Random
MAR	Missing at Random
MNAR	Missing Not at Random

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Tukey (1977) defines exploratory data analysis (EDA) as an approach for analyzing data, which employs a variety of procedures to maximize insight into a data set, extract important variables, detect outliers and determine optimal factor settings. In addition, Tukey (1977) promotes EDA to statisticians to explore the data, and possibly formulate hypotheses that could lead to the analysis of data in any experiment.

On the other hand, EDA is an approach for analyzing data sets, which summarize the main characteristics, often with visual methods which foreseen what the data can tell us beyond the formal modeling or hypothesis testing task (Hoaglin et al., 1986).

More so, Hoaglin et al. (1986) describe EDA as not a mere collection of techniques rather as a philosophy to which data set is dissected, what we look out for, how its being looked at and how we make the interpretation. Consequently, EDA heavily uses tables and graphics as techniques in displaying statistical information about a specific set of data. One of the main role of EDA is to explore data by making use of tables that give us the unparalleled power to use the data. This help in revealing its structural details, and always ready to gain some new often-unsuspected insight into the data.

1.2 Median polish

The median polish is an EDA technique introduce by Tukey (1977) which finds a simple model for data in a two-way layout table of the form grand, row and column effects. On the other hand, median polish is a technique from EDA, which can help in understanding the analysis of data base on a simple model. Therefore, median polish is a data analysis technique for examining the significance of various factors in a single or multi-way model.

Similarly, median polish algorithm is useful for removing any noise in the data by computing the medians for various coordinates on the data sets. Given a contingency table, the iteration of the median polish produces row and column effects which are the main factors based on medians in an additive model, assuming no interaction among factors. For details See (Tukey, 1977; Velleman and Hoaglin, 1981; Emerson and Stoto, 1982). The advantage of this approach is not limited to the estimation of main effects only but quantifies the magnitude of the main effect as the partial interaction of row and column effects.

Siegel (1983) emphasizes that median polish method iteratively applies the median operator to the medians of all rows and columns while Fink (1988) describes median polish fits as an additive model by operating on the data from the table by subtracting medians along each dimension of the cell. Each factor has a certain number of values it can contain. Each factor need not to have the same level. The cell is the product of levels in each factor on a contingency table.

An efficient way to estimate main effects as a partial multiplicative interaction between row and column is by median polish method. Main effects can be estimated through other EDA techniques. The main difference between median polish approach and other approaches to main-effect modeling is that the median polish approach defines the main effect as a special form of multiplicative interaction between rows and columns.This makes median polish approach as reliable, robust, and simple statistical procedure than other EDA techniques (Hoaglin et al., 1986).

We used contingency table because it is the common basis for investigating the effects of treatments on the grain crops. The rows and columns of contingency table corresponds to two different categorical attributes. Each cell of the contingency table contains a numerical entry which reflects a measurement under the combination of the categorical attributes corresponding to the cell. Data for analysis often come in the form of a two-way table, involving values for each combination of several levels of two different variables.

Barbará and Wu (2003) describe median polish as a simple yet robust method to perform exploratory data analysis which is resistant to holes in the contingency table, but it may require much iterations through the data.

Median polish for analysing a one-way or two-way table is a resistant technique which isolated large disturbances in a small number of cells which will not affect the common value, row and column effects and instead will reflected in the residuals. For this reason, we can adopt median polish for EDA instead of corresponding analysis by means that form the basis for classical analysis of varaince (Hoaglin et al., 1983). This part provides a brief review of the analysis based on means and compares with the analysis of the median polish that uses medians.

1.3 Problem statement

Median polish for fitting simple models to two-way table is a relatively recent development which is one of EDA methods that is resistant to outliers as described by Tukey (1977).

Median polish model is used for determining the effect of rows and columns in a two-way table using medians instead of means.

Median polish methods is used by various authors in analysing data in a various fields, among them are: Velleman & Hoaglin (1981) use median polish method for testing interaction in a two-way layout with several observations per cell. Hoaglin et al. (1983) estimates the row and column effects by method of median polish in square combining table using means.

Shitan and Vazifedan (2011) apply two-way median polish with covariate to demonstrate the hour (covariate) spent by police squad on patrol has association with the rate of reduction in car theft.

Fitrianto et al. (2014) describe the relationship between students' performance in relation with the to the attendance in the classroom taking seven course as sample to run median polish algorithm. The method is good but could not estimate the difference between the students scores and the attendance in a contigency table.

We use median polish technique in analysing paired data which can be found in agriculure, medicine, engineering and so on. From the literature, to the best of my knowledge, no work on median polish for analysing paired data is found. Therefore, there is the need to handle paired data using median polish techniques to get information from the paired data. The information involve revealing the difference between the paired data through column and row effects.

Median polish methods in EDA can be applied in solving missing paired data. The missing paired data method in EDA is useful to infer the structure in the data and interprete the contribution of each variable. Paired data with missing values can arise for a variety of reasons, including the inability or unwillingness of participants to meet appointments for evaluation in some studies.

Median polish is resistant to holes in a table. If cells were empty in the table, the fit produced by median polish would coincide with the data given originally. Missing values produce holes in the table and holes distort a median polish analysis. Hence, a method of median polish model is employed to analyse the process of missing data on a contingency table (Barbará and Wu, 2003).

The impact of missing data on quantitative research can introduce potential bias in parameter estimation and weaken the generalizability of the results (Rubin, 1996). Secondly, ignoring cases with missing data leads to the loss of information which in turn decreases statistical power and increases standard errors Peng et al. (2006).

There are various ways of analysing missing paired data like multiple imputation which was developed by Rubin (1976) as a method for averaging the outcomes across multiple imputed data sets.

These methods are good to handle missing paired data which does not add any bias, but it does decrease the power of the analysis by decreasing the effective sample size (Rubin and Little, 2002). However, none of these methods use median polish to analyse missing data. Hence, we use median polish method to analyse missing data in a contingency table. Since we have a median polish model for the paired data, finding a value for a missing paired data point is usually easy. we can fit the data to the model and evaluate the fit result at the missing point (Camacho, 2010).

More so, median polish method is prefer to traditional statistics for the following reasons:

- median polish method is simpler and easier.
- median polish violate the assumptions required by traditional statistics
- median polish is more robust.
- median polish command can work with unbalanced design.
- it is resistant to outliers.
- uses median instead of mean.

1.4 Objectives of the study

The main objectives of this study is to propose median polish techniques for analysing paired data. The specific objectives are:

1. to propose a new appproach to handle paired data using median polish.
2. to propose a method of handle incomplete paired data using median polish.

3. to propose procedure of algorithm to handle paired data with missing values.

1.5 Significance of the study

The findings of this study will contribute to the development of validated and reliable median polish technique in assessing the difference between a paired data. Further, median polish with covariate method can also assist in estimating the relationship between paired data. The information gathered from the application of these methods can be used to assess and evaluate treatment data for future occurrences. Finally, it is hoped that this study would be helpful for future related researches.

1.6 Scope of the study

The aim of this study is to find an improved way of handling paired data of grain yields using median polish technique. The data is collected as a secondary data from ministry of Agriculture and farm extension in Adamawa state of Nigeria base on the grain yields from the field. The data were fit into the model and analyse using R-programming. We assess the effect of treatment on paired data using median polish on the data. Median polish algorithm is applied to the data to estimate the grand, row and column effects. Comparison values were computed to measure the association between rows and column effects. The value of slope b computed from comparison values indicates no need of transformation because it is close to zero. The median polish method is propose to handle incomplete and missing data in a contingency table.

1.7 Overview of the study

This thesis contain six chapters: Introduction, Literature review, Research methodolgy, Application of one-way median polish, Application of two-way median polish and conclusion and recommendations for future studies. The details of the chapters are as follows:

Chapter 1 provides an overview of the thesis such as background of the research study, discussion on median polish, problem statement, objectives of the study, scope and signficance of the study.

Chapter 2 presents the literature reviews on exploratory data analysis and median polish. We discuss about paired data and paired comparison. The review on missing paired data and how to handle missing paired data using median polish technique in contingency table.

Chapter 3 presents the methodology applied in this study. Research framework including median polish concepts, application of one-way and two-way way median polish, resistant lines for paired data using median polish approach, dependent pairwise data using median polish and missing data concepts and imputation.

Chapter 4 presents the computational results and discussions on application of one-way median polish on paired data. The descriptive data analysis is carried out to verify the relationship and difference between the paired data using median polish technique. We also discuss about the procedure of imputing missing values.

Chapter 5 provides explanation on the results and discussions of application of two-way median polish on paired data. The computation of comparison values to determine whether the paired data require transformation or not.

Chapter 6 presents summary, contibution, conclusion and recommendations for future research.

BIBLIOGRAPHY

Allison, P. D. (2002). Missing data: Quantitative applications in the social sciences. *British Journal of Mathematical and Statistical Psychology*, 55(1):193–196.

Barbará, D. and Wu, X. (2003). An approximate median polish algorithm for large multidimensional data sets. *Knowledge and Information Systems*, 5(4):416–438.

Bennett, D. A. (2001). How can i deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25(5):464–469.

Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, 76(374):296–311.

Bradley, E. L. and Blackwood, L. G. (1989). Comparing paired data: a simultaneous test for means and variances. *The American Statistician*, 43(4):234–235.

Camacho, J. (2010). Missing-data theory in the context of exploratory data analysis. *Chemometrics and Intelligent Laboratory Systems*, 103(1):8–18.

Davis, H. (1963). The method of pairwise comparisons. *Griffin Press, London, UK*.

Durrant, G. B. and Skinner, C. (2006). Using missing data methods to correct for measurement error in a distribution function. *Survey Methodology*, 32(1):25.

Emerson, J. D. and Stoto, M. A. (1982). Exploratory methods for choosing power transformations. *Journal of the American Statistical Association*, 77(377):103–108.

Enders, C., Dietz, S., Montague, M., and Dixon, J. (2006). Modern alternatives for dealing with missing data in special education research. *Advances in learning and behavioral disorders*, 19:101–130.

Fink, A. (1988). How to polish off median polish. *SIAM Journal on Scientific and Statistical Computing*, 9(5):932–940.

Fitrianto, A., Wijayanto, H., Rana, M., and Cheong, Y. (2014). Median polish for final grades of MTH3000 and MTH4000 level courses. *Applied Mathematical Sciences*, 8(126):629–6302.

Geffen, G., Bradshaw, J., and Nettleton, N. (1973). Attention and hemispheric differences in reaction time during simultaneous audio-visual tasks. *The Quarterly Journal of Experimental Psychology*, 25(3):404–412.

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, England.

Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., and Schafer, J. L. (1997). Analysis with missing data in prevention research. *The Science of Prevention: Methodological Advances From Alcohol and Substance Abuse Research*, 1:325–366.

Hinloopen, E. and Nijkamp, P. (1990). Qualitative multiple criteria choice analysis. *Quality & Quantity*, 24(1):37–56.

Hoaglin, D. C., Iglewicz, B., and Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396):991–999.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). *Understanding Robust and Exploratory Data Analysis*. John Wiley & Sons, New York, USA.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (2011). *Exploring Data Tables, Trends, and Shapes*, volume 101. John Wiley & Sons, New Jersey, USA.

Kemperman, J. (1984). Least absolute value and median polish. *Lecture Notes-Monograph Series: Inequalities in Statistics & Probability, 84–103*.

Keyes, K. M. and Li, G. (2010). A multiphase method for estimating cohort effects in age-period contingency table data. *Annals of Epidemiology*, 20(10):779–785.

Kim, J. K. and Shao, J. (2013). *Statistical Methods for Handling Incomplete Data*. CRC Press, New York, USA.

Koczkodaj, W. and Orłowski, M. (1999). Computing a consistent approximation to a generalized pairwise comparisons matrix. *Computers & Mathematics with Applications*, 37(3):79–85.

Little, R. J. and Rubin, D. B. (1983). On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician*, 37(3):218–220.

McNeil, D. (1992). On graphing paired data. *The American Statistician*, 46(4):307–311.

Menz, H. B. (2005). Analysis of paired data in physical therapy research: time to stop double-dipping? *Journal of Orthopaedic & Sports Physical Therapy*, 35(8):477–478.

Peng, C. Y. J., Harwell, M., Liou, S. M., and Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. *Real Data Analysis, 4, 31–78*.

Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation*, 7(4):353–383.

Royston, P. (2005). Multiple imputation of missing values: update. *Stata Journal*, 5(2):188–205.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473–489.

Rubin, D. B. and Little, R. J. (2002). Statistical analysis with missing data. *John Wiley & Sons, Hoboken, USA*.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall, New York, USA.

Shitan, M. and Vazifedan, T. (2011). *Exploratory Data Analysis for Amost Everyone*. Universiti Putra Malaysia Press, Serdang, Malaysia.

Siegel, A. F. (1983). Low median and least absolute residual analysis of two-way tables. *Journal of the American Statistical Association*, 78(382):371–374.

Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., Refulio ZRodriguez, N. F., Walker, J. B., Moore, M. J., and Carlsward, B. S. (2011). Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany*, 98(4):704–730.

Sun, Y. and Genton, M. G. (2012). Functional median polish. *Journal of Agricultural, Biological, and Environmental Statistics*, 17(3):354–376.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286.

Tukey, J. W. (1977). Exploratory data analysis. *Addison -Wesley, Reading, UK*.

Underwood, A. (1992). Beyond baci: the detection of environmental impacts on populations in the real, but variable, world. *Journal of Experimental Marine Biology and Ecology*, 161(2):145–178.

Van Buuren, S. and Oudshoorn, K. (1999). Flexible multivariate imputation by mice. *TNO Prevention Center: Leiden, The Netherlands*.

Velleman, P. F. and Hoaglin, D. C. (1981). *Applications, Basics, and Computing of Exploratory Data Analysis*. Duxbury Press, Boston, USA.

Westerhuis, J. A., van Velzen, E. J., Hoefsloot, H. C., and Smilde, A. K. (2010). Multivariate paired data analysis: multilevel plsda versus oplsda. *Metabolomics*, 6(1):119–128.

Yan, W. (2013). Biplot analysis of incomplete two-way data. *Crop Science*, 53(1):48–57.