5-2020

Nonparametric Approach to Multivariate Analysis of Variance (Manova) Using Density-Based Kernel Methods

Hend Aljobaily

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

NONPARAMETRIC APPROACH TO MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA) USING DENSITY-BASED KERNEL METHODS

A Dissertation Submitted in Partial Fulfillment
of the Requirement for the Degree of
Doctor of Philosophy

Hend Aljobaily

College of Educational and Behavioral Sciences
Department of Applied Statistics and Research Methods

May 2020

This Dissertation by: Hend Aljobaily

Entitled: *Nonparametric Approach to Multivariate Analysis of Variance (MANOVA) Using Density-Based Kernel Methods*

has been approved as meeting the requirement for the Degree of Doctoral of Philosophy in College of Education and Behavioral Sciences in Department of Applied Statistics and Research Methods

Accepted by the Doctoral Committee

____

Han Yu, Ph.D., Research Advisor

____

Trent Lalonde, Ph.D., Committee Member

____

William Merchant, Ph.D., Committee Member

____

George Thomas, Ph.D., Faculty Representative

Date of Dissertation Defense ______________________________

Accepted by the Graduate School

____

Cindy Wesley, Ph.D.
Interim Associate Provost and Dean
The Graduate School and International Admissions

ABSTRACT

Aljobaily, Hend. *Nonparametric Approach to Multivariate Analysis of Variance (MANOVA) Using Density-Based Kernel Methods.* Published Doctor of Philosophy dissertation, University of Northern Colorado, 2020.

Density estimation and nonparametric methods have become increasingly important over the last three decades. These methods are useful in analyzing modern data which includes many variables with numerous observations. It is common to see datasets with hundreds of variables and millions of observations. Examples proliferate in fields such as geological exploration, speech recognition, and biological and medical research. Therefore, there is an increasing demand for tools that can detect and summarize the multivariate structure in complex data. Only nonparametric methods are able to do so when little is known about the data generating process. The term nonparametric does not necessarily mean that models lack parameters but that the number of parameters is infinite and not fixed. The functional forms of its parametric counterparts are known up to only a finite number of parameters. Kernel method is one of the most prominent and useful nonparametric approaches in statistics, machine learning, and econometrics. In fact, virtually all nonparametric algorithms are asymptotically kernel methods. Kernel analysis allows for transformation of high-dimensional data to low-dimensional statistical problems via kernel functions. Density estimation is now recognized as a useful tool with univariate and bivariate data. The goal of this study was to demonstrate that it is also a powerful tool for the analysis of high-dimensional data. The asymptotic aspects as well as the application of the nonparametric methods applied to multivariate data were the focus

in this research which eventually leads to the research on Gaussian processes (GP), or more generally, random fields (RF).

In this dissertation, a novel multivariate nonparametric approach was proposed to more strongly smooth raw data, reducing the dimension of the solution to a handful of interesting parameters. The proposed approach employed methods that exploited kernel density estimation (KDE) which can be applied to hypothesis testing of the equality of location parameters in the one-way layout as well as the main effects and interaction effect in the two-way layout. First, multivariate kernel-based tests were developed to conduct multivariate analysis of variance (MANOVA) for testing hypotheses against distinct group means in various settings, including one and two-way with interaction settings. Then, the asymptotic properties of the proposed methods were investigated and the asymptotic distributions were derived. Next, simulations were conducted to investigate the small-sample behavior of the proposed nonparametric kernel-based test statistics for the one and two-way layout. Then, comparisons were made between the proposed nonparametric kernel-based methods and the traditional parametric counterparts for one-way and two-way layout. Finally, the proposed nonparametric kernel-based methods were applied to a real image dataset. The results of this dissertation showed that the proposed nonparametric kernel-based methods have greater power than the counterpart parametric methods, i.e. MANOVA, in detecting differences between groups in multivariate settings when the underlying distribution of the data is not normal.

ACKNOWLEDGMENTS

There is no way to thank and acknowledge all the people who have supported me throughout my life and education. I am the person I am today because of all of the individuals who believed in and supported me.

Foremost, I would like to thank my advisor, Dr. Han Yu, for the support, encouragement, patience, and immense knowledge you have provided through the dissertation process. Thank you for being understanding whenever I was stressing out and for always finding the time to meet with me and calm me down.

I would like to extend my sincere thanks to members of my committee, Dr. Trent Lalonde, Dr. William Merchant, and Dr. George Thomas, for the time that you have invested into reviewing this dissertation. Your guidance has been tremendously valuable. I am so grateful to have had the chance to work with each of you through my tenure at the University of Northern Colorado.

I must also thank all of my professor in the department of Applied Statistics and Research Methods for their guidance throughout my graduate years. Your guidance and encouragement has opened the door to this path. I would also like to express my gratitude to Dr. Su Chen, an assistant professor at the University of Memphis, for her insightful recommendations, especially with the coding of this dissertation.

I am also grateful to my classmates in the Applied Statistics and Research Methods program at the University of Northern Colorado. There have been so many discussions and conversations that I will never forget. A special thanks goes to Kofi Wagya and Lauren Matheny for supporting and believing in me, and for insightful comments and recommendations.

I am also grateful for the Administrative Assistant of the Applied Statistics and Research Methods program, Keyleigh Gurney for her overwhelming support throughout my journey at the University of Northern Colorado. I am thankful for all the time you spent listening to me as I voiced my struggles and complaints. You always listened and encouraged me.

I am immensely grateful to the unwaivering and unending support of my family. I cannot begin to express my thanks to my amazing parents, who raised me to appreciate education and taught me that the sky is the limit. Although my parent are far away in Saudi Arabia, they never fail to be encouraging and supportive. My engineer sister, Sara, thank you for inspiring and encouraging me and for being the supportive and proud sister you are. My brothers, Bader and Yousef, thank you for your humor and support and for always bragging about your Ph.D. sister to everyone you know. Words cannot express how much all of you have impacted my life. I love you so much and I would never choose another family.

To the most amazing sister anyone could ask for, Nouf, the completion of this dissertation would not have been possible without your support and encouragement. I am deeply thankful for all of the nights you stayed up late with me to support me during the making of this dissertation as well as for your support throughout my entire educational journey. You were the person next to me when I chose the topic of this dissertation and through all my highs and lows. For that, I love you and will be forever thankful for having you as my sister.

Last, but certainly not least, to my partner and the love of my life, Michael, thank you for your unconditional support and limitless patience. Thank you for putting up with all my late nights writing and coding this dissertation. Thank you for all the countless hours you put into editing and reviewing the English language, grammar, and sentence structure of this dissertation. This dissertation would not have happened without your support and edits. You are my biggest fan and supporter in this journey and in my life.

Finally, I would like to say that words cannot express how much I love you and appreciate having you in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Conducting research with high-dimensional data usually results in the violation of the normality assumption of most statistical tests which negatively influences the results of the research. Thus, comprehensive and innovative new approaches are needed to address this problem and offer new solutions. If the issue of violating normality is not taken into consideration when creating and building research design, the final results might not be accurate or valid, and the outcome of the study may not be reliable and useful. Thus, studying and researching the concept of “high-dimensional data” can greatly benefit fields that utilize inferential statistical models. High-dimensional data refers to datasets with staggeringly high numbers of dimensions, known as *features* or *variables* (Finney, 1977). When dealing with high-dimensional data, the problem of violating the normality assumption can arise due to the increase in dimensions as the number of observations increase (Chen & Xia, 2019). Another problem that arises with high-dimensional data has been labeled “the curse of dimensionality.” In general, this problem occurs since when an increase in the dimensionality of data causes the volume of the space to increase rapidly, causing the data to become sparse. When this happens, problems with statistical significance and reliability arise (Bellman, 1957).

Motivation

The motivation of this research comes from the limitation of the currently used inferential nonparametric statistical approaches when dealing with complex datasets, especially when the purpose of the study is to compare groups in multivariate data settings. Nonparametric approaches are used extensively across many fields and disciplines when the population data has an unknown distribution, violating normality assumption. For example, when studying images of breast cancer cells in patients, the images are constructed as multivariate data with many variables such as wavelength, electron energy, and particle mass (Geladi & Grahn, 2006). Since image data are considered complex, high-dimensional data, traditional parametric approaches cannot be applied. Another example of a situation in which a multivariate nonparametric approach is commonly used, when the population has unknown distribution, is when studying water quality with the goal of studying and modeling the differences between water indexes where the normality assumption is violated. This data is considered multivariate and, because of the violation of normality assumption, the traditional parametric approach is not appropriate. The inability to use the traditional parametric methods with high-dimensional data requires the use of alternative nonparametric statistical approaches with the ability to account for a high-dimensional data structure.

Many old, high-dimensional methods rely heavily on parametric models, which assume that the underlying distribution of the data has a finite number of parameters. When these assumptions are met, accurate results can be expected. However, with the increasing complexity of data, the results obtained from the parametric-based methods can lead to inaccurate and misleading results (Liu, 2010). To deal with these challenges, nonparametric methods with an infinite number of parameters are appropriate and powerful enough to use on most modern high-dimensional data. Kernel method is an example of a nonparametric approach that can be applied in such cases. Kernel method is considered an appropriate method to use because it does not assume that the underlying distribution of the data has a finite number of parameters, i.e. parametric distribution. This dissertation is focused on studying the differences between groups in a multivariate framework with high-dimensional data, thus the kernel-based nonparametric multivariate analysis of variance method is most appropriate for use in this study.

Kernel Method

Kernel method is a nonparametric technique that is widely used for density estimation. It allows for the estimation of density with minimal assumptions about the model's functional form of the distribution. Kernel method also has many advantages over the traditional histogram approach for statistical modeling. It has better scaling with dimensionality than the histogram approach by creating a kernel at each dimension. Additionally, it eliminates the biasness in the histogram caused by the cut-off points that lack strong substantive justification. The kernel method solves the memory problem by reducing computing time. It does so by eliminating the computation normally involved in the "training" phase by only requiring storage of the training set.

In machine learning, kernel methods are used for pattern recognition. The general task of pattern recognition is to find and study the types of relationships among variables in a dataset such as clusters, principal components, correlations, and classifications. The kernel method allows for the transformation of data into a higher dimension that has a clear dividing margin between classes of data. Kernel methods use kernel functions to enable them to operate in a high-dimensional space by computing the inner products of all pairs of data in the feature space (Theodoridis & Koutroumbas, 2009). Kernel functions have been used for the analysis of sequence, graph, text, and image data. The kernel method can be used to turn any linear model into a more flexible model by simply replacing its features (predictors) with a kernel function (Kaluza, 2016). Support vector machines (SVM), Gaussian processes, principal components analysis (PCA), canonical correlation analysis, ridge regression, and clustering have algorithms which make them capable of operating with kernels.

Multivariate Analysis of Variance

Analysis of variance (ANOVA) was introduced by the statistician and evolutionary biologist, Ronald Fisher, in 1918. The ANOVA is based on the law of total variance, where the observed variation is partitioned into components that are attributed to different sources of variation in a given dataset (Fisher, 1918). ANOVA is a statistical method used to test differences between two or more group means. It makes inferences about group means by analyzing variation between and within groups (Lane, 2013). ANOVA can be extended to what is known as multivariate analysis of variance (MANOVA). MANOVA can be thought of as an ANOVA with several dependent variables. In other words, ANOVA tests for the difference in means between two or more groups while MANOVA tests for the difference in mean vectors between two or more groups (Rencher & Christensen, 2012).

Purpose

The limitation of parametric MANOVA is clear. It assumes normality which is an unrealistic assumption when applied to most real-world data. In practice, data do not come from an exact normal distribution. Thus, this assumption is often too strict for the data in real research settings. In such circumstances, nonparametric techniques should be applied instead. Until now, almost all the nonparametric MANOVA tests are based on rank-transformed scores which use ranks of observations instead of the original observation. One of the major drawbacks of using ranks instead of original observations is the loss of information. In addition, these ranked techniques can only be applied in a one-way setting when one independent variable is of interest. Thus, the purpose of this dissertation was to introduce and demonstrate a new approach for analyzing multivariate data using nonparametric multivariate analysis of variance with kernel methods. This provides a new, useful technique that can be applied by researchers and practitioners to compare groups in multivariate settings when the underlying distribution of the data is not normal.

Research Questions

In order to develop an appropriate nonparametric multivariate analysis of variance (MANOVA) using kernel methods, this dissertation addressed the following questions:

Q1 How can a kernel density estimator be constructed for non-Gaussian multivariate data?

Q2 How can hypotheses be tested using multivariate data when kernel methods are used within the one-way MANOVA technique?

Q3 How can the main effect hypothesis be tested using multivariate data when kernel methods are used within the two-way MANOVA technique?

Q4 How can the interaction effect hypothesis be tested using multivariate data when kernel methods are used within the two-way MANOVA technique?

Q5 How do Type I error rate and power of the proposed kernel-based one-way MANOVA test behave compared to the parametric one-way MANOVA?

Q6 How do Type I error rate and power of the proposed kernel-based test for the interaction term in the two-way MANOVA behave compared to the interaction term in the parametric one-way MANOVA?

To provide a comprehensive answer to the research questions of interest, they were divided into two phases. The first phase focuses on answering the first four research questions which are addressed theoretically through mathematical derivations and proofs provided in Chapter III. The second phase answers the fifth and sixth research questions through simulations of Type I error rate and power. The conditions of these simulations are provided in Chapter III and the results are provided in Chapter IV. It was crucial to develop the theoretical portion of this dissertation before implementing the empirical component of the study.

Organization of the Dissertation

In Chapter II, a comprehensive review of the analysis of the variance and the development of nonparametric methods through history is provided along with an in-depth discussion of kernel methods. In Chapter III, a nonparametric one-way kernel-based MANOVA test with a homogeneous variance-covariance matrix among groups is constructed and its limiting distribution is studied. In this chapter, the nonparametric kernel-based MANOVA test is extended to the two-way layout for both the main and interaction effects. In Chapter IV, simulation studies are used to evaluate the performance of the proposed methods in comparison to traditional parametric MANOVA, using the type I error and power. The second section of Chapter IV contains real-data application of the proposed kernel-based MANOVA tests using an image dataset. Chapter V consists of discussion, limitations, and the direction of possible future work pertaining to the topics and methods discussed throughout this dissertation.

CHAPTER II

LITERATURE REVIEW

This chapter is dedicated to reviewing the fundamental of parametric and nonparameric approches and the literature on multivariate data as well as various nonparametric approaches currently used to analyze multivariate data. The first section gives a general introduction of the parametric analysis and the traditional parametric techniques used in the field. It also includes two subsections that provide a discussion about the traditional parametric analyses available for use in analyzing univariate analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA). The second section introduces the idea of nonparametric analysis and the traditional nonparamteric techniques used in the field. It includes two subsections that have a discussion about the traditional nonparametric analyses available foe use in analyzing univariate analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA).

The third section gives a general introduction to the idea of nonparametric analysis and the traditional nonparametric methods and techniques used in the field as well as providing a discussion about the traditional nonparametric analyses available to use. Within this section, there are four subsections that are layed out in the order they were originally developed: Binomial-Based Nonparametric Tests, Rank-Based Nonparametric Tests, Empirical Cumulative Density Function (ECEF)-Based Tests, and Density-Based Nonparametric Tests. The Binomial-Based Nonparametric Tests subsection provides a summary of various nonparametric tests used to analyze data using a single dichotomous variable for scoring. The Rank-Based Nonparametric Tests subsection discusses the idea of nonparametric tests that are based on ranks of the original data using the order information of the data as a polytomous scoring. It also discusses the traditional nonparametric alternatives to ANOVA and MANOVA such as the Kruskal Wallis (KW) test and Multivariate Kruskal Wallis (MKW) test, and their limitations. The Empirical Cumulative Density Function (ECDF)-Based Nonparametric Tests subsection provides background information pertaining to the techniques used in these ECDF-Based tests. Finally, the Density-Based Nonparametric Tests subsection discusses the concept of using density as the data generating function providing background information for the second section of this chapter. Each subsection provides examples of various developed methods, each attempting to fix the limitations of the previous technique. Additionally, each subsection provides detailed background information to widen understanding of the multivariate data setting and plausible methods to fix its problems and limitations.

The fourth section introduces kernel analysis and its use as a nonparametric multivariate approach. This section also discusses the use of kernel approach when working with nonlinear data structural. Within this section there are two subsections: Classification Using Kernels and Analysis of Variance using kernels. The Classification Using Kernels subsection provides detailed background information about an important topic in “big data” and machine leaning, which is classification. In this subsection, there are summaries of novel methods developed recently that use kernels performing classification. The Analysis of Variance Using Kernels subsection discusses the use of kernel function or “kernel trick” to provide an alternative nonparametric method to the traditional parametric analysis of variance (ANOVA). This approach does not have restrictive assumptions such as normality and homogeneity of the variance. Relaxing these assumptions makes it possible to analyze non-normal and non-linear data. It also uses mapping to represent data in a higher dimensional space which makes the interpretation of the result easier. The kernel approach discussed in this section is used and extended to develop the proposed method of this dissertation, multivariate analysis of the variance when dealing with non-normal and non-linear data.

The final section of this chapter discusses the limitations of the traditional parametric analysis of the variance in a multivariate setting (MANOVA). It also discusses the limitations of the traditional nonparametric method which is based on rank-transformed techniques. Many recently developed methods and techniques do not provide appropriate approaches to analyze multivariate data. These limitations give a strong rationale for the purpose of this dissertation which is developing an extension for the kernel method that can be implemented when dealing with multivariate data to perform a nonparametric MANOVA.

Parametric Methods

Statistical methods that use distributional assumptions that are known up to a finite number of parameters are called parametric methods. They are a branch of statistics that assumes the sample data comes from a population that follows a distribution based on a finite, fixed set of parameters such as the mean and standard deviation in the normal distribution (Altman & Bland, 2009). Many popular statistical methods are parametric such as t-test, z-test, Pearson's correlation, and ANOVA. Parametric models usually have strong assumptions about a given population including normality and homogeneity of the variance. When these assumptions are confirmed, parametric methods will produce accurate and precise estimates which result in a high statistical power. However, when the assumptions are violated, parametric methods have a greater chance of failing: thus, they are not statistically robust methods (Altman & Bland, 2009). Additionally, when the functional form of parametric model's distribution is misspecified, biased estimates occur.

Parametric Univariate Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) is a process of analyzing the differences in means in three or more groups by examining the variation between and within these groups in a sample data. The traditional parametric analysis of variance uses an F-test or a variance ratio test.

Let X_{ij} be independent random variables sampled from I populations, where $i = 1,2,...,I,\ j = 1,2,...,n_i$. The usual parametric ANOVA aims to test the hypothesis $H_0 : \mu_1 = \mu_2 = ... = \mu_I$ versus $H_a : \mu_i \neq \mu_j$ for any $i \neq j$, where μ_i is the mean of ith population. The usual parametric test, F-test, relies on assuming the independence of observations, normality of the distribution, and constancy of variance. It also relies on the concept of total variability, i.e. total sum of squared deviations from the mean. For example, in the one-way ANOVA, the sum of squares is partitioned into two parts: sum of squares within the groups (SSW) and sum of squares between the groups (SSB). The F-test statistic is constructed by calculating the ratio of the two sums of squares with their corresponding degrees of freedom - a large ratio indicates large differences between the groups. Therefore, the F-test statistic can be rewritten as (Scheffé, 1959):

$$F = \frac{MSB}{MSW} = \frac{SSB/(I-1)}{SSB/(N-I)} = \frac{\sum_{i=1}^{I} n_i(\bar{x}_{i.} - \bar{x}_{..})^2/(I-1)}{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2/(N-I)}, \tag{2.1}$$

where, $\bar{x}_{i.} = \frac{1}{n_i}\sum_i x_{ij}$ and $\bar{x}_{..} = \frac{1}{N}\sum_i\sum_j x_{ij}$. The above test statistic follows an F-distribution with degrees of freedom $I-1$ and $N-I$. Thus, the null hypothesis $H_0 : \mu_1 = \mu_2 = ... = \mu_I$ is to be rejected when $F > F(\alpha, I-1, N-I)$ for the significance level of α.

Parametric Multivariate Analysis of Variance (MANOVA)

Multivariate Analysis of Variance (MANOVA) is an extension of the univariate analysis of variance (ANOVA). Statistical differences among groups are being examined by univariate ANOVA while MANOVA test is an extension of the ANOVA test that takes into account multiple (two or more) continuous dependent variables. In statistics, MANOVA is a method for comparing multivariate sample means. It compares whether the linear combination of dependent variables differs among groups. In this way, it tests whether the independent grouping variable simultaneously explains a statistically significant amount of variance in the dependent variable. Usually, a MANOVA test is followed by separate ANOVA tests for the individual dependent variables to identify significant difference between groups.

In the case of multivariate, the null hypothesis that the population means are equal to one another for all p variables is tested . Just like AVONA, MANOVA relies on the concept of total variability, i.e. total sum of squared deviations from the mean. However, unlike the univariate case, there is not just one method (i.e. the F-test) that can be used to form a test statistic. Instead, there are several methods which are the following (Bray & Maxwell, 1985):

1- Wilks' lambda which can be written as,

$$U = \prod_{i=1}^{I} \frac{1}{1+\lambda_i}, \tag{2.2}$$

2- Pillai-Bartlett trace which can be written as,

$$V = \sum_{i=1}^{I} \frac{\lambda_i}{1+\lambda_i}, \tag{2.3}$$

3- Roy's greatest characteristic root which can be written as,

$$GCR = \frac{\lambda_i}{1+\lambda_i}, \tag{2.4}$$

4- Hotelling-Lawley trace which can be written as,

$$T = \sum_{i=1}^{I} \lambda_i, \tag{2.5}$$

where $\lambda_i = \frac{SSB}{SSW}$.

Nonparametric Methods

Nonparametric statistics is a branch of statistics that has recently been gaining in popularity. The term nonparametric does not mean that models completely lack parameters but that the number of parameters is infinite and not fixed as is the case in parametric methods (Dickhaus, 2018). Also, the nonparametric model structure is data driven; this means it is not specified in advance but is instead determined from data. Thus, nonparametric methods can be used when less is known about the studied distribution (Dickhaus, 2018). One of the advantages of nonparametric methods is that they make fewer assumptions than parametric methods. Because of this, nonparametric methods tend to be more robust than parametric methods. Another advantage of using nonparametric methods is simplicity. In certain cases, such as when rapid results are needed, nonparametric methods are used even when the use of parametric methods would be justified. Due to the simplicity and robustness of nonparametric methods, they are seen to be more appropriate and to eliminate misunderstanding (Corder & Foreman, 2014). When parametric tests are more appropriate, nonparametric tests have less power. This means a larger sample size might be required to draw conclusions with the same degree of confidence as parametric methods would provide (Corder & Foreman, 2014).

Nonparametric Univariate
Analysis of Variance

When compared to parametric methods, nonparametric methods make fewer assumptions about parameters such as means and standard deviations. The Kruskal-Wallis test is a non-parametric method used to compare k independent samples. It is equivalent to the parametric one-way ANOVA with the use of ranks instead of the original data. The Kruskal-Wallis test statistic can be formulated as (Daniel, 1990):

$$T = \frac{(N-1)(S_t^2 - C)}{S_r^2 - C}, \tag{2.6}$$

where $S_t^2 = \sum_{i=1}^{I} \frac{R_i^2}{n_i}$, $S_r^2 = \sum_{i=1}^{N} r_i j^2$, $C = \frac{N(N+1)^2}{4}$, $R = \sum r_i$, and r_i is the ranked data point.

In a multi-group experimental design, the nonparametric Kruskal-Wallis test provides a more powerful alternative to the parametric one-way ANOVA when the assumption of normality is violated (May & Johnson, 1997). The Kruskal-Wallis test has fewer assumptions with the main assumption being that observations in each group are identically and independently distributed. This assumption is significant because if the original observations are identically distributed, the Kruskal-Wallis test can be interpreted as testing for a difference between medians instead of means.

Nonparametric Multivariate Analysis of Variance

In a multi-group experimental design with a univariate dependent variable and the assumption of non-normality, the nonparametric Kruskal-Wallis test is more appropriate than the parametric ANOVA. For a multivariate setting where multiple dependent variables are of interest, Puri and Sen (1969) proposed a generalization of the nonparametric Kruskal-Wallis test that is potentially equivalent to the one-way multivariate analysis of variance (MANOVA) when the dependent variables are measured on at least an ordinal scale (Puri & Sen, 1969). Large sample theory suggests that the multivariate Kruskal-Wallis test statistic is approximately distributed as chi-squared. However, when small samples are to be analyzed, randomization theory needs to be used to tabulate the exact distribution (Puri & Sen, 1969).

Historical Development of Nonparametric Analysis

Nonparametric statistical analysis is an advancing approach that has been used widely in many fields. The idea of nonparametric statistical analysis started in the early 1700s and experienced a renaissance in the1940s. Since then, the approach has developed into a coherent group of modern statistical procedures that can handle data with high dimensionality and complexity (Liu & McKean, 2016).

Binomial-Based Nonparametric Tests

Binomial distribution has been used for statistical analysis for more than 300 years. The sign test developed by the British physician Arbuthnott in 1710 was the first binomial-based test (Liu & McKean, 2016). In 1713, Jacques Bernoulli developed

Bernoulli trials which is another binomial-based statistical calculation which are considered a breakthrough in the history of probability theory (Liu & McKean, 2016). Today, binomial procedures remain one of the easiest and most useful statistical procedures (Hollander, Chicken, & Wolfe, 2014).

The binomial test of significance is a probabilistic used to examine the distribution of a single dichotomous variable in small samples. It involves testing the difference between a sample proportion and a given proportion using the following equation (Siegel, 1956):

$$P(Y=C)=\binom{N}{C}p^{C}(1-p)^{N-C}, \tag{2.7}$$

where $\binom{N}{C}$ is the number of possible combinations, P is the probability if obtaining the outcome of interest, and C is the outcome of interest. When the sample size is greater than 25, normal approximation of the binomial test of significance is to be utilized.

The sign test, proposed by Arbuthnott in 1710, tests the median of a continuous population which counts the number of observations greater than the null hypothesized value of the median. The null and alternative distributions of the sign test statistic are both binomial (Daniel, 1990).

The limitations of the binomial-based tests are obvious including the assumption of a dichotomous-type distribution, meaning that the variable of interest is dichotomous in nature. Also, they assume that the variable of interest has only two possible values (outcomes) that are mutually exclusive and mutually exhaustive in all cases being considered.

Rank-Based Nonparametric Tests

The modern era for nonparametric tests began with the work of Wilcoxon in 1945. Wilcoxon proposed the nonparametric Wilcoxon signed-rank test for the median of a symmetric population formulated based on the sum of signed ranks instead of the original data points (Daniel, 1990):

$$W = \sum_{i=1}^{N} [sgn(x_{2,i} - x_{1,i}) R_i] \tag{2.8}$$

and the nonparametric Wilcoxon rank sum test for the difference in population medians. The test statistic can be written as such (Hollander & Wolfe, 1999):

$$W = \sum_{j=1}^{n} R_j \tag{2.9}$$

where R_j are the ranks of each compared group.

In 1947, Mann and Whitney showed that the rank sum test is equivalent to the sign test proposed by Wilcoxon when applied to the pairwise differences across the two samples (Daniel, 1990). Also, in 1949, Tukey proposed the Walsh averages which are equivalent to the signed rank test when applied to the pairwise averages from the sample (Liu & McKean, 2016). After that, in 1948, Pitman introduced the idea of efficiency for hypothesis testing using the nonparametric approach (Liu & McKean, 2016). Following that, in 1956 and 1960 Hodges and Lehmann analyzed the efficiency of several rank tests and compared them to different least squares tests such as the t-test and F-test (Liu & McKean, 2016). They proved that the efficiency of the Wilcoxon tests relative to the t-tests is 0.955 for normal distribution and never less than 0.864; it can also be arbitrarily large for heavy tailed model distributions (Daniel, 1990). These results made the nonparametric rank-based tests low-power alternatives to the parametric t-tests.

The next major step in the evolution of nonparametric rank-based methods was made by Hodges and Lehmann in 1963. Hodges and Lehmann developed estimators based on rank test statistics called R-estimates. They showed that these estimators inherit the efficiency of the rank tests they are derived from. Since the Wilcoxon test statistics and the L_1 norm are connected, the Hodges-Lehmann estimate of location is the median of the pairwise averages, and the estimate for the difference in locations is the median of the pairwise differences across the two samples (Liu & McKean, 2016). By the mid-1960s, rank-based tests and estimates for location models, including the one-way layout, were available and had the same efficiency properties as the previous rank-based tests. Around the same time, in 1964, robustness was introduced by Huber and again by Hampel in 1974 (Liu & McKean, 2016). The fundamental tools of robustness are the influence function and break down point which Wilcoxon R-estimates have these in addition to the desirable efficiency properties mentioned above (Liu & McKean, 2016). In 1967, Hájek and Šidák published a seminal work on the development of rank-based tests followed by a second edition in 1999 which extends much of the theory and includes material on R-estimates (Liu & McKean, 2016).

Hence, during the 1960s, nonparametric rank-based tests and rank-based estimates for location models were well understood and provided excellent alternatives to parametric methods such as the t-tests and F-tests. Unfortunately the rank-based techniques have not extended easily to the two-way layout, especially with interaction terms.

The next major step in the development of the rank-based methods was the nonparametric rank-based linear regression in one-way and two-way layout for testing regression parameters in a rank-based way. These techniques were developed by Jurečková between 1969 and 1971 when he developed the asymptotic theory of the nonparametric rank-based tests. They were later developed by Jaeckel in 1972 when he provided a nonparametric rank-based dispersion function that, when minimized, produced R-estimates (Liu & McKean, 2016). Finally, in 1975, McKean developed rank-based tests as a nonparametric alternative for the linear model. McKean also improved the theory of the asymptotic distribution of these nonparametric rank-based linear model tests.

Various tests and techniques have since been developed and extended based on the nonparametric rank-based tests mentioned above. All the traditional nonparametric rank-based tests and techniques mentioned above are essential in the field of nonparametric statistics and have been crucial to its development.

Although the rank-based nonparametric tests show evidence of high power and robustness when compared to their corresponding parametric tests, there are still many limitations. Some of these limitations include the loss of information when ranking the data instead of using the original observations as well as when discarding the observations when the difference between groups is zero in many rank-based tests. Additionally, rank-based nonparametric tests have limited application for the two-way layout especially in the presence of interaction terms.

Empirical Cumulative Density Function (ECDF)-Based Nonparametric Tests

An empirical distribution function is the distribution function associated with the empirical measure of a sample (Shorack & Wellner, 1986). It is considered as an estimate of the cumulative distribution function that generated the points in the sample. This cumulative distribution function is a step function that jumps up by $1/n$ at each n data points. At any specified value of the measured variable, the value of the empirical distribution function is calculated as the fraction of observations of the measured variable that are less than or equal to the specified value. According to the Glivenko–Cantelli theorem (Tucker, 1959), the empirical distribution function converges to the data's underlying distribution; thus, it has a probability of 1.

Let $(X_1, \ldots, X_n)$ be independent and identically distributed random variables with the cumulative distribution function, $F(t)$. When this is the case, the basic empirical distribution function is defined as follows (Vaart, 1998):

$$\hat{F}_n(t) = \frac{\text{number of elements in the sample} \leq t}{n} = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq t}, \tag{2.10}$$

or as (Coles, 2001):

$$\hat{F}_n(t) = \frac{1}{n+1}\sum_{i=1}^{n} 1_{X_{i \leq t}}, \tag{2.11}$$

where 1_A is the indicator of event A. For a fixed t, the indicator $1_{X_{i \leq t}}$ is a Bernoulli random variable with the parameter $p = F(t)$; hence $n\hat{F}_n(t)$ is a binomial random variable with mean $nF(t)$ and variance $nF(t)(1 - F(t))$, which implies that $\hat{F}_n(t)$ is an unbiased estimator for $F(t)$.

Since the ratio $\frac{n+1}{n}$ approaches 1 as n goes to infinity, the asymptotic properties of the two definitions that are given above are the same. Hence, by the strong law of large numbers, the estimator $\hat{F}_n(t)$ converges to F(t) as $n \to \infty$ for every value of t (Vaart, 1998):

$$\hat{F}_n(t) \stackrel{a.s.}{\to} F(t) \tag{2.12}$$

therefore, the estimator $\hat{F}_n(t)$ is consistent. The Glivenko–Cantelli theorem states that the convergence happens uniformly over t as such (Vaart, 1998):

$$||\hat{F}_n(t) - F(t)||_\infty \equiv sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F(t)| \stackrel{a.s.}{\to} 0. \tag{2.13}$$

This result led to what is known as the Kolmogorov–Smirnov statistic for testing the goodness-of-fit between the empirical distribution $\hat{F}_n(t)$ and the assumed true cumulative distribution function F which is equivalent to the sup-norm of $||\hat{F}_n(t) - F(t)||_\infty$. Another relative result is the Cramér–von Mises statistic which is equivalent to the L^2-norm of $||\hat{F}_n(t) - F(t)||_\infty$ (Vaart, 1998).

Additionally, it was proven by the central limit theorem that the asymptotic distribution of $\hat{F}_n(t)$ is normal with the standard $\sqrt{n}$ rate of convergence (Vaart, 1998). This result can be written as such:

$$\sqrt{n}(\hat{F}_n(t) - F(t)) \stackrel{d}{\to} N(0, F(t)(1 - F(t))). \tag{2.14}$$

In 1952, Donsker extended this result to what is known as the Donsker's theorem. The Donsker's theorem states that the empirical process, $\sqrt{n}(\hat{F}_n - F)$, converges in distribution in the Skorokhod space $D[-\infty,+\infty]$ to the mean-zero Gaussian process $G_F = B \circ F$, where B is the standard Brownian bridge (Vaart, 1998). The uniform rate of convergence in Donsker's theorem can be quantified by the result known as the Hungarian embedding as (Vaart, 1998):

$$\limsup_{n\to\infty} \frac{\sqrt{n}}{\ln^2 n} \left\| \sqrt{n}\left(\hat{F}_n - F\right) - G_{F,n} \right\|_\infty < \infty, \quad \text{a.s.} \tag{2.15}$$

Additionally, the rate of convergence of $\sqrt{n}\left(\hat{F}_n - F\right)$ can also be quantified using the Dvoretzky-Kiefer-Wolfowitz inequality provides bound on the tail probabilities of $\sqrt{n}\left\|\left(\hat{F}_n - F\right)\right\|_\infty$ as (Vaart, 1998):

$$\Pr\left(\sqrt{n}\left\|\hat{F}_n - F\right\|_\infty > z\right) \leq 2e^{-2z^2}. \tag{2.16}$$

Also, Kolmogorov has shown that if the cumulative distribution function F is continuous, then $\sqrt{n}\left\|\hat{F}_n - F\right\|_\infty$ converges in distribution to $\|B\|_\infty$, which has a Kolmogorov distribution that does not depend on F (Vaart, 1998).

One of the limitations of the ECDF-based nonparametric tests is that, with many ECDF-based test, the distribution it is tested against needs to be specified. Using a permutation test can be one solution to this limitation. Permutation tests, also called randomization tests, are one type of nonparametric test. They were proposed in the early twentieth century and have only recently become popular due to the availability of modern, inexpensive computational power (Nichols & Holmes, 2001). Permutation tests are a type of statistical test used for testing significance in which the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic on the observed data points. The theory behind the permutation tests has evolved since the works of Ronald Fisher and E. J. G. Pitman in the 1930s (Berry, Johnston, & Mielke, 2014).

To illustrate the basic idea of a permutation test, assume we have random variables X_A and X_B for each individual from two groups, A and B, with sample means of $\bar{x}_A$ and $\bar{x}_B$. Assume also that we want to know whether or not X_A and X_B come from the same distribution. Let n_A and n_B be the sample obtained from each group. First, the permutation test is applied to determine whether the observed difference between the sample means is large enough to reject the null hypothesis, H_0, and that the data drawn from group A and B are coming from the same distribution. After this, the desired test statistic, T, is calculated. Next, the mean difference of group A and B is calculated and recorded for every possible division of the pooled values into two groups of size n_A and n_B. The set of these calculated differences is the exact distribution of all possible differences under the null hypothesis.

The next step is calculating the p-value of the chosen test statistic. The one-sided p-value of the test is calculated as the proportion of sampled permutations where the difference in means is greater than or equal to T. The two-sided p-value of the test is calculated as the proportion of sampled permutations in which the absolute difference was greater than or equal to $|T|$.

Permutation tests are intuitive and easy to compute when the computation power is available. However, an important assumption behind a permutation test is that the observations are exchangeable and randomly assigned under the null hypothesis. One disadvantage to this assumption is that tests of difference in location require equal variance which is not always the case with real-world data. One solution to this limitation is to use a bootstrap-based tests; however, bootstrap tests are not considered exact tests as permutation tests (Good, 2005).

Density-Based Nonparametric Tests

In probability and statistics, density estimation is the process of constructing an estimate of the underlying density function using observed data. Many techniques are used for density estimation, including the kernel density estimator, also known as the

Parzen estimator. The kernel density estimation (KDE) is among the most commonly used methods of the density-based nonparametric approach.

Nonparametric Kernel Analysis

In recent years, due to the increasing availability of “high-dimensional data” and the consequent need to solve increasingly complex multivariate problems, there is a growing interest in nonparametric statistical methods, especially Kernel analysis. Kernel methods are a class of algorithms used in machine learning for pattern analysis. The purpose of pattern analysis is to find and study clusters, rankings, principal components, correlations, and classifications in a dataset. Kernel methods use kernel functions that enable them to operate in a high-dimensional way, avoiding the curse of dimensionality, by simply computing the inner products between all pairs of data in the feature space. Using kernels is often computationally cheaper than the explicit computation of the coordinates.

Kernel functions have been used for sequence, graphs, text, images, and vector data. Any linear model can be turned into a non-linear model by applying the “kernel trick” to the model (Aizerman, Braverman, & Rozonoer, 1964; Boser, Guyon, & Vapnik, 1992). By applying the “kernel trick” using some kernel function, $k(x,y)$, that can be expressed as an inner product, the data is transferred from its original space, $\mathscr{X}$, to an inner dot-product space, $\mathscr{K}$. In this operation, the mapping of features (predictors) that are needed to obtain linear learning algorithms to learn nonlinear functions or decision boundaries, is performed implicitly, $k(x,y) = \langle \phi(x), \phi(y) \rangle$ (See Figure 2.1). Choosing the right kernel function is crucial to the success of the linear model in the feature space (Howley & Madden, 2005; Alizadeh & Ebadzadeh, 2011). Many kernel algorithms are based on convex optimization or eigen problems where their statistical properties are analyzed using statistical learning theory, such as Rademacher complexity (Diaf, Boufama, & Benlamri, 2012).

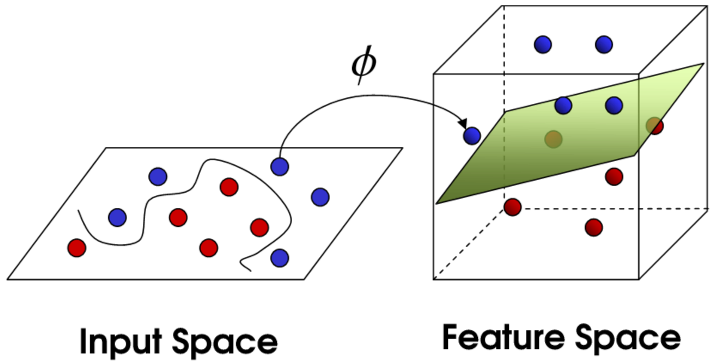

Figure 2.1. Kernel Trick

Classification Using Kernels

Classification plays an important role in machine learning and statistics generally. Kernel methods are a class of algorithms for pattern analysis that can be used to solve classification problems. Kernel mapping is a nonlinear method that has gained a lot of attention from researchers in the field of pattern recognition and statistical machine learning. A Kernel-based approach is a better choice when dealing with a nonlinear classification model.

Diaf et al. (2012) proposed a competitive nonlinear classification approach based on the nonparametric version of Fisher's discriminant analysis called kernel nonparametric discriminant analysis (KNPDA). The idea of the proposed KNPDA approach is to perform the same concept of nonparametric discriminant analysis (NDA) in the kernel feature space, $\mathscr{F}$. Since the idea of linear discrimination still applies, even in $\mathscr{F}$, and because $\mathscr{F}$ cannot always be explicitly computed, a new objective expression that makes use of data samples in terms of only inner-dot products (or Gram matrix K) based on the"kernel trick" has been derived (Diaf et al., 2012). The mapping of the input data from its original input space into a higher dimension Hilbert space $\mathscr{H}$ is achieved through a mapping function Φ expressed as follows:

$$\Phi : \begin{cases} \mathbb{R}^2 \rightarrow \mathbb{R}^m \quad (n < m \leq \infty) \\ \mathscr{X} \rightarrow \mathscr{F} \\ X \longmapsto \phi(X) \end{cases} \tag{2.17}$$

In the linear case, nonparametric Fisher's discriminants are computed by maximizing $S_W^{-1} S_B$. The same criterion can be expressed in $\mathscr{F}$ in terms of mapped training patterns Φ as:

$$J\left(w_\Phi\right) = \frac{w_\Phi^T S_B^\Phi w_\Phi}{w_\Phi^T S_W^\Phi w_\Phi} \tag{2.18}$$

where any solution of $w_\Phi \in \mathscr{F}$ lies in the span of all mapped training samples in $\mathscr{F}$ based on the theory of reproducing kernels (Mika, Rätsch, Weston, Schölkopf, & Müller, 1999). S_W^Φ and S_B^Φ can be defined as follows:

$$S_W^\Phi = \frac{1}{N} \sum_{i=1}^{C} \sum_{l=1}^{N_i} \left(\phi(x_l^i) - \mu_i^\Phi\right) \left(\phi(x_l^i) - \mu_i^\Phi\right)^T, \tag{2.19}$$

$$S_B^\Phi = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1, j\neq i}^{C} \sum_{l=1}^{N_i} \omega^\Phi(i,j,l)(\phi(x_l^i) - m_j(\phi(x_l^i)))(\phi(x_l^i) - m_j(\phi(x_l^i)))^T \tag{2.20}$$

where $\omega^\Phi(i,j,l)$ is a weighting function that plays an important role in preserving the boundary structure between mapped classes by assigning 0.5 (or close to 0.5) for samples near the classification boundary and zero to samples away from the classification boundary. $m_j^\Phi\left(\phi\left(x_l^i\right)\right)$ are the mean vectors what are used to represent nonparametric global information about each class in $\mathscr{F}$, and can be defined as follows:

$$m_j^\Phi\left(\phi\left(x_l^i\right)\right) = \frac{1}{k} \sum_{p=1}^{k} nn\left(\phi\left(x_l^i\right), j, p\right) = NN_{il}^j 1_{\frac{1}{k}}, \tag{2.21}$$

where each NN_{il}^J is an $(L^\Phi \times k)$ matrix that holds the k-nearest neighbors of $\phi\left(x_l^i\right)$ from class j in $\mathscr{F}$, and $1_{\frac{1}{k}}$ is a $(k \times 1)$ vector with all elements equal $\frac{1}{k}$.

For any data sample x_s to be projected into a one-dimensional eigenspace using the best eigenvector α corresponding to the largest eigenvalue λ_1, the following equation is used:

$$w_{\Phi}^{T}\,\phi\left(x_s\right) = \alpha^T k\left(X, x_s\right) = \sum_{i=1}^{N} \alpha_i k(x_i, x_s). \tag{2.22}$$

Generally, to compute the non-linear discriminant features $(\acute{L})$, the top eigenvalues of $A^{-1}B$ will be calculated. The, $\alpha_1, ..., \alpha_L$ are chosen to create the eigenspace E containing all the training samples (Diaf et al., 2012). Each training observation x_i is then projected into E as a point e based on the following equation:

$$e = (\alpha_1, \ldots, \alpha_{\acute{L}})^T k(X, x_i). \tag{2.23}$$

In eigenspace-based techniques, recognizing a new sample is simply done by seeking the most similar sample in the produced eigenspace E using the Euclidean metric. When classifying a test sample x_t, the sample is projected into E and is then classified based on its best matching training sample in E.

Now, by using the definition of $m_j^{\Phi}(\phi\left(x_l^i\right))$, the between-class scatter matrix of the mapped data S_B^{Φ} can be rewritten as:

$$S_B^{\Phi} = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1, j\neq i}^{C}\sum_{l=1}^{N_i} \omega^{\Phi}\left(i, j, l\right)\left(\phi(x_l^i) - NN_{il}^{j} 1_{\frac{1}{k}}\right)\left(\phi(x_l^i) - NN_{il}^{j} 1_{\frac{1}{k}}\right)^T, \tag{2.24}$$

which makes B equal to the following:

$$B = \tfrac{1}{N}\textstyle\sum_{i=1}^{C}\sum_{j=1, j\neq i}^{C}\sum_{l=1}^{N_i} \omega^{\Phi}\left(i, j, l\right)\phi(X)^T \underbrace{\left(\phi\left(x_l^i\right) - NN_{il}^{j} 1_{\frac{1}{k}}\right)}_{P1}\underbrace{\left(\phi\left(x_l^i\right) - NN_{il}^{j} 1_{\frac{1}{k}}\right)^T}_{P2}\phi\left(X\right), \tag{2.25}$$

Consider the left part of the dot product $(P1)$ along with the definitions of kernel matrix and kernel function $K = \left\langle \phi\left(X\right), \phi\left(X\right)\right\rangle$ and $k\left(x_i, x_j\right) = K_{ij} = \left\langle \phi\left(x_i\right), \phi\left(x_j\right)\right\rangle$,

$$\phi(X)^T \left(\phi\left(x_l^i\right) - NN_{il}^j 1_{\frac{1}{k}}\right) = \phi(X)^T \phi\left(x_l^i\right) - \phi(X)^T NN_{il}^j 1_{\frac{1}{k}} = K_l^i - NNK_{il}^j 1_{\frac{1}{k}} \quad (2.26)$$

where K_l^i is an $(N \times 1)$ kernel vector of data sample $x_l \in X_i$ and $NNK_{il}^j \subset K$ which is an $(N \times k)$ matrix holding the kernel vectors of NN_{il}^j. Hence, we are able to represent B as,

$$B = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1, j\neq i}^{C}\sum_{l=1}^{N_i} \omega^{\Phi}\left(i, j, l\right)\left(K_l^i - NNK_{il}^j 1_{\frac{1}{k}}\right)\left(K_l^i - NNK_{il}^j 1_{\frac{1}{k}}\right)^T, \quad (2.27)$$

with $\omega^{\Phi}\left(i, j, l\right)$ as,

$$\omega^{\Phi}\left(i, j, l\right) = \frac{min\left\{d^{\alpha}\left(\phi(x_l^i),\, NN_{il}^j\left[k\right]\right), d^{\alpha}\left(\phi(x_l^i),\, NN_{il}^j\left[k\right]\right)\right\}}{d^{\alpha}\left(\phi(x_l^i),\, NN_{il}^j\left[k\right]\right) + d^{\alpha}\left(\phi(x_l^i),\, NN_{il}^j\left[k\right]\right)} \quad (2.28)$$

where $d\left(\phi\left(p\right), \phi\left(q\right)\right)$ is the Euclidean distance between two mapped points, $\phi\left(p\right)$ and $\phi\left(q\right)$, in $\mathscr{F}$. The kernel-based formula of d that computes the distance in $\mathscr{F}$ using the original points p and q in $\mathscr{X}$; $d_{\Phi}(p, q)$, can simply be derived as follows (Wang, Yunde, Changbo, & Turk, 2004):

$$d_{\Phi}^2\left(x_p, x_q\right) = k\left(x_p,\, x_q\right) - 2k\left(x_p,\, x_q\right) + k\left(x_p,\, x_q\right). \quad (2.29)$$

The aftermath can show that the weighting function ω^{Φ}, can be written in terms of the kernel matrix K, as,

$$\omega^{\Phi}\left(i, j, l\right) = \frac{min\left\{\left(K_{pp} - 2k_{pq} + K_{qq}\right)^{\alpha}, \left(K_{pp} - 2k_{pr} + K_{rr}\right)^{\alpha}\right\}}{\left(K_{pp} - 2k_{pq} + K_{qq}\right)^{\alpha} + \left(K_{pp} - 2k_{pr} + K_{rr}\right)^{\alpha}}, \quad (2.30)$$

where p,q, and r are indices to data samples in input space $\mathscr{X}$ corresponding to $\phi\left(x_l^i\right)$, $NN_{il}^i[k]$, and $NN_{il}^j[k]$ respectively.

Since $\sum_{i=1}^{N}\phi(x_i)-\mu^{\Phi}=\phi(X)-\mu^{\Phi}1$, the within-class scatter matric S_W^{Φ} of the mapped data can be written as,

$$S_W^{\Phi}=\frac{1}{N}\sum_{i=1}^{C}\phi(X_i)\,\phi(X_i)^T-\frac{1}{N_i}\phi(X_i)\,\phi(X_i)^T=\frac{1}{N}\sum_{i=1}^{C}\phi(X_i)\left(I-1_{\frac{1}{N_i}}\right)\phi(X_i)^T, \tag{2.31}$$

where I is an N_i-squared identity matrix and $1_{\frac{1}{N_i}}$ is an N_i-squared matric with all entries equal $1/N_i$.

Analysis of Variance Using Kernels

The analysis of variance (ANOVA) models play a vital role in analyzing the effect of categorical factors on a response variable. They have been applied in analyzing data from a wide range of fields such as biology, psychology, business, and sociology. The main idea of ANOVA is to decompose the variability in the response variable according to the effect of different factors. The existing literature on ANOVA can be categorized into two divisions: parametric and nonparametric techniques. The parametric tests, i.e. the traditional F-test, rely on the assumptions of homoscedasticity and normality of the errors. The existing nonparametric ANOVA methods are either based on rank transformed techniques or performed purely by simulations. Also, none of the literature on nonparametric two-way ANOVA has provided methods with theoretical support to test models with main and interaction effects, as in the case with parametric ANOVA test. In 2013, Chen has proposed a novel distribution-free ANOVA test that provides a nonparametric analog of traditional F-test for both one-way and two-way layout. These test statistics are not based on rank transformed techniques. Rather "kernel-transformed" technique. The kernel estimate of the function $f(x)$ was estimated by Ahmed (1982):

$$\int\hat{f}^2(x)\,dx=\frac{1}{n^2h^2}\sum\sum_{i\neq j}\bar{K}\left(\frac{X_i-X_j}{h}\right), \tag{2.32}$$

where $\bar{K}(z) = \int K(u)K(z-u)\,du$, which is also a kernel function.

Under the assumption of homogeneity of variance, i.e. $\sigma_i = \sigma$ for all i. The F-test statistic of a kernel-based nonparametric test for location parameters in a one-way layout is defined as such:

$$F_l = \frac{MSB}{MSW} = \frac{SSB/(K-1)}{SSW/df_w} = \frac{\sum_{i=1}^{K} \frac{n_i\left(\hat{V}_i - \hat{V}_.\right)^2}{\omega_i^2} / (K-1)}{\left(\sum_{i=1}^{K} \sum\sum_{j_1 \neq j_2}^{n_i} \frac{\left(A_{ij_1j_2} - \hat{V}_i\right)^2}{\omega_i^2}\right) / df_w}, \tag{2.33}$$

where

$$\hat{V}_i = \frac{1}{n_i(n_i-1)h_i} \sum\sum_{j_1 \neq j_2} \left(\frac{X_{ij_1} + X_{ij_2}}{2}\right) K\left(\frac{X_{ij_1} - X_{ij_2}}{h_i}\right), \tag{2.34}$$

$$\hat{V}_. = \frac{\sum_{i=1}^{K} n_i \hat{V}_i / \omega_i^2}{\sum_{i=1}^{K} n_i / \omega_i^2}, \tag{2.35}$$

$$\omega_i^2 = 4\left\{\int x^2 f_i^3(x)dx - \left(\int x f_i^2(x)\,dx\right)^2\right\}, \tag{2.36}$$

$$A_{ij_1j_2} = \frac{1}{h_i}\left(\frac{X_{ij_1} + X_{ij_2}}{2}\right) K\left(\frac{X_{ij_1} - X_{ij_2}}{h_i}\right), \text{ and} \tag{2.37}$$

$$df_w = \begin{cases} K(n-1) & \text{if } n_i = n \text{ for all } i, \\ d & \text{otherwise} \end{cases}, \tag{2.38}$$

where d is the number of eigenvalues of B_3 given by:

$$B_3 = \begin{bmatrix} (\frac{n_1}{2}-1)(I_{n_1} - \frac{1}{n_1}J_{n_1}) & 0 & \dots & 0 \\ 0 & (\frac{n_2}{2}-1)(I_{n_2} - \frac{1}{n_2}J_{n_2}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\frac{n_K}{2}-1)(I_{n_K} - \frac{1}{n_K}J_{n_K}) \end{bmatrix}. \tag{2.39}$$

Under the null hypothesis, $H_0 : V_1 = V_2 = ... = V_K$, $F_l = MSB/MSW$ follows asymptotically an F distribution with degrees of freedom $K-1$ and $K(n-1)$ for balanced data and an F distribution with degrees of freedom $K-1$ and d for unbalanced data.

Under the alternative hypothesis, $F_l = MSB/MSW$ has an asymptotic non-central F distribution with degrees of freedom $K-1$ and $K(n-1)$ for balanced data. It follows an asymptotic non-central F distribution with degrees of freedom $K-1$ and d for unbalanced data (Chen, 2013), with the non-centrality parameter defined as,

$$\psi_2 = \frac{1}{2}\mu^{(2)'}B_2\mu^(2), \tag{2.40}$$

where $\mu^{(2)} = \left(\mu_1^{(2)},\ \mu_2^{(2)},\ \ldots, \mu_K^{(2)}\right)$,

$$\mu_1^{(2)} = \frac{\sqrt{\lambda_i}\left(e_i - \frac{\sum_{i=1}^{K}\lambda_i e_i/\omega_i^2}{\sum_{i=1}^{K}\lambda_i/\omega_i^2}\right)}{\sigma\omega_i}, \tag{2.41}$$

$$B_2 = \begin{bmatrix} 1 - \frac{\lambda_1/\omega_1^2}{\sum_{i=1}^{K}\lambda_i/\omega_i^2} & -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_2}/\omega_2)}{\sum_{i=1}^{K}\lambda_i/\omega_i^2} & \cdots & -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^{K}\lambda_i/\omega_i^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^{K}\lambda_i/\omega_i^2} & -\frac{(\sqrt{\lambda_2}/\omega_2)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^{K}\lambda_i/\omega_i^2} & \cdots & 1 - \frac{\lambda_K/\omega_K^2}{\sum_{i=1}^{K}\lambda_i/\omega_i^2} \end{bmatrix}, \tag{2.42}$$

where $\lambda_i = \lim_{n_i\to\infty}\frac{n_i}{\sum_{i=1}^{K}n_i}$ and e_i is such that $\mu_i = 1 + \frac{e_i}{\sqrt{\sum_{i=1}^{K}n_i}}$.

A simulation study showed that the kernel-based nonparametric one-way ANOVA test is almost as powerful as the parametric (traditional) one-way ANOVA test when the samples come from normal distributions and it outperforms the parametric one-way ANOVA test when the samples come from Cauchy or Lognormal distributions (Chen, 2013).

For the two-way ANOVA layout, there are two different potential tests; the test for the main effect and the test for the interaction effect. For the main effect, the F-test statistic of the kernel-based nonparametric test for location parameters of the row effect is defined as,

$$F_{R_l} = \frac{MSR}{MSW} = \frac{SSR/(r-1)}{SSW/df_{w2}} = \frac{\sum_{i=1}^{r} \hat{m}_i \left(\hat{V}_{i.} - \hat{V}_{..}\right)^2 /(r-1)}{\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1 \neq k_2}\left(A_{ijk_1k_2} - \hat{V}_{ij}\right)^2 / \widehat{\omega}_{ij}^2\right)/df_{w_2}}, \quad (2.43)$$

where $\hat{V}_{.j} = \frac{\sum_i \hat{m}_{ij}\hat{V}_{ij}}{\hat{m}_{.j}}$ and $\hat{m}_{.j} = \sum_i \hat{m}_{ij}$. Thus, the F-test statistic of kernel based nonparametric test for location parameters of the column effect is defined as follows:

$$F_{C_l} = \frac{MSC}{MSW} = \frac{SSC/(c-1)}{SSW/df_w} = \frac{\sum_{i=1}^{c} \hat{m}_i \left(\hat{V}_{.j} - \hat{V}_{..}\right)^2 /(c-1)}{\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1 \neq k_2}\left(A_{ijk_1k_2} - \hat{V}_{ij}\right)^2 / \widehat{\omega}_{ij}^2\right)/df_{w_2}}, \quad (2.44)$$

where

$$\sum\sum_{k_1 \neq k_2} \frac{\left(A_{ijk_1k_2} - \hat{V}_{ij}\right)^2}{\omega_{ij}^2} = \left(\frac{n_{ij}}{2} - 1\right)\left[\sum_{k=1}^{n_{ij}} \frac{\left(2\varphi\left(X_{ijk_1}\right) - V_{ij}\right)^2}{\omega_{ij}^2} - \frac{n_{ij}\left(\hat{V}_{ij} - V_{ij}\right)^2}{\omega_{ij}^2}\right], \quad (2.45)$$

$$\hat{V}_{ij} = \frac{1}{n_{ij}\left(n_{ij}-1\right)h_{ij}} \sum\sum_{k_1 \neq k_2} \left(\frac{X_{ijk_1} + X_{ijk_2}}{2}\right) K\left(\frac{X_{ijk_1} - X_{ijk_2}}{h_{ij}}\right), \quad (2.46)$$

$$\hat{V}_{i.} = \frac{\sum_j \hat{m}_{ij}\hat{V}_{ij}}{\hat{m}_{ij}}, \hat{V}_{..} = \frac{\sum_i \sum_j \hat{m}_{ij}\hat{V}_{ij}}{\hat{m}_{..}},\ \hat{m}_{i.} = \sum_j \hat{m}_{ij},\ \hat{m}_{ij} = \frac{n_{ij}}{\hat{\omega}_{ij}^2}, \quad (2.47)$$

$$\hat{\omega}_{ij}^2 = 4\left\{\int x^2 f_{ij}^3(x)dx - \left(\int x f_{ij}^2(x)\,dx\right)^2\right\}, \text{ and} \quad (2.48)$$

$$df_{w_2} = \begin{cases} rc\,(n-1) \text{ if } n_i = n \text{ for all } i, \\ d_2 \text{ otherwise} \end{cases}, \quad (2.49)$$

where d_2 is the number of eigenvalues of B_6 given by

$$B_6 = \begin{bmatrix} \left(\frac{n_{11}}{2}-1\right)\left(I_{n_{11}} - \frac{1}{n_{11}}J_{n_{11}}\right) & 0 & \cdots & 0 \\ 0 & \left(\frac{n_2}{2}-1\right)\left(I_{n_2} - \frac{1}{n_2}J_{n_2}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\frac{n_K}{2}-1\right)\left(I_{n_{rc}} - \frac{1}{n_{rc}}J_{n_{rc}}\right) \end{bmatrix}. \quad (2.50)$$

Under the null hypothesis $H_0 : \bar{V}_{i.} = \bar{V}_{..}$ for all i, $F_{R_l} = MSR/MSW$ follows asymptotically an F distribution with degrees of freedom $r-1$ and $rc(n-1)$for balanced data and an F distribution with degrees of freedom $r-1$ and d_2 for unbalanced data. Under the alternative hypothesis, $F_{R_l} = MSR/MSW$ follows asymptotically a non-central F distribution with degrees of freedom $(r-1)(c-1)$ and $rc(n-1)$ for balanced data, and a non-central F distribution with degrees of freedom $(r-1)(c-1)$ and d_2 for unbalanced data (Chen, 2013).

Similarly, $F_{C_l} = MSC/MSW$ follows asymptotically an F distribution with degrees of freedom $c-1$ and $rc(n-1)$ for balanced data and an F distribution with degrees of freedom $c-1$ and d_2 for unbalanced data (Chen, 2013).

For the interaction effect, the F-test statistic of the kernel-based nonparametric test for location parameters of the interaction effect is defined as,

$$F_{I_l} = \frac{MSI}{MSW} = \frac{SSI/(r-1)(c-1)}{SSW/df_{w2}} = \frac{\sum_{i=1}^{r} \hat{m}_i \left(\hat{V}_{ij} - V_{i.}^{*} - V_{.j}^{*} - V_{..}^{*}\right)^2 / (r-1)(c-1)}{\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1 \neq k_2}\left(A_{ijk_1k_2} - \hat{V}_{ij}\right)^2 / \widehat{\omega}_{ij}^2\right)/df_{w_2}}, \quad (2.51)$$

where $V_{i.}^{*} = \frac{\sum_j m_{ij}\hat{V}_{ij}}{m_{i.}}$ and $V_{..}^{*} = \frac{\sum_i \sum_j m_{ij}\hat{V}_{ij}}{m_{..}}$.

Under the null hypothesis $H_0 : V_{ij} - \bar{V}_{i.} - \bar{V}_{.j} + \bar{V}_{..} = 0$ for all i and j, $F_{I_l} = MSI/MSW$ follows asymptotically an F distribution with degrees of freedom $(r-1)(c-1)$ and $rc(n-1)$ for balanced data and an F distribution with degrees of freedom $(r-1)(c-1)$ and d_2 for unbalanced data. Under the alternative hypothesis, $F_{I_l} = MSI/MSW$ follows asymptotically a non-central F distribution with degrees of freedom $(r-1)(c-1)$ and $rc(n-1)$ for balanced data, and a non-central F distribution with degrees of freedom $(r-1)(c-1)$ and d_2 for unbalanced data.

A simulation study showed that the kernel based nonparametric two-way ANOVA test of interaction and test of main effects are less powerful than the nonparametric one-way ANOVA test for the same cell size when the samples come from a normal distribution. However, the kernel based nonparametric two-way ANOVA test of interaction and main effects are more powerful than the nonparametric one-way ANOVA test for the same cell size when the samples come from Cauchy and Lognormal distributions. Also, the kernel based nonparametric ANOVA test is more powerful than the parametric (traditional) ANOVA for non-normal data, especially with strongly skewed data (Chen, 2013).

Limitations

Although MANOVA has considerable advantages when compared to multiple separate analyses of variance, the limitations for parametric MANOVA, whether it is one or two-way layout, are clear. It assumes multivariate normality and homogeneity of covariance matrices of each group. In practice, it is uncommon for data to come from a normal distribution which sometimes make these assumptions too strict to be met in real research settings. Also, often MANOVA requires a large sample size for some complicated models since the number of cases in each category must be larger than the number of dependent variables.

In these circumstances, nonparametric techniques should be applied instead. Much of the published work about nonparametric MANOVA tests provide ranked-based techniques. One major drawback of using the ranks, rather than the raw data, is the loss of information. The rank keeps the order of the raw data while ignoring the magnitude of the differences within the data. For instance, we can have two datasets of continuous data that have the exact same rank but have totally different means, variances, and/or distributions. If only the ranked datasets are analyzed, instead of the raw data, there will be no difference detected between these two datasets. Unfortunately, none of the rank transformed techniques can compensate for the loss of information. In addition, the majority of literature in nonparametric MANOVA tries to express and interpret their models in the same way as parametric models even when it is inappropriate to do so. For example, $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$ is the two-way with interaction term layout in the parametric MANOVA. In the nonparametric MANOVA, the dependent variables, y_{ijk}, do not have to come from a normal distribution as in the parametric model. However, much of the literature that uses the rank transformed approaches still interpret μ as the grand mean even if, for some distributions, the first moment does not exist.

So far, important work has been done to solve the problems faced when performing the Kruskal-Wallis test for more than one factor. For example, a novel method, described earlier in this chapter, was proposed by Chen (2013) to develop a one and two-way nonparametric ANOVA using kernels. However, no extension for this method has been developed for handling multivariate data to perform a nonparametric MANOVA using kernels. Developing such a method will be the focus of this dissertation.

CHAPTER III

METHODOLOGY

This chapter is dedicated to constructing theoretical derivations of the proposed nonparametric kernel-based MANOVA tests, examining their asymptotic distributions when using multivariate data. The proposed nonparametric methods are based on a similar framework as the traditional parametric MANOVA, obtaining sum of squares to allow derivation of the statistical tests' asymptotic distributions.

This chapter includes six sections that reveal the methodology used in this study. First, a overview of the Reproducing Kernel Hilbert Space theory, which provides the theory behind the methodology of this dissertation, is provided. Second, the multivariate kernel density estimation technique is provided along with a discussion of the selected bandwidth matrix. Third, the nonparametric kernel-based one-way MANOVA test derivations are presented to provide a theoretical understanding of the proposed methods. Fourth, the nonparametric kernel-based two-way MANOVA tests derivations for the main and interaction effects are presented to provide a theoretical understanding of the proposed methods. Fifth, a discussion of the techniques used to evaluate the performance of the proposed nonparametric methods is given. Finally, a detailed explanation of the data simulation schemes and conditions used in this study is provided.

Reproducing Kernel Hilbert Space (RKHS)

In machine learning, kernel methods are a class of algorithms usually used for pattern analysis. Pattern analysis is used to obtain and study the types of relations exist in datasets. For many of these algorithms, kernel methods require only a kernel function specified by the user, i.e., a similarity function for pairs of raw data points. Kernel methods utilize kernel functions which enable them to operate in a high-dimensional feature space without the need to compute the coordinates of the data in that space. Instead, they only compute the inner products between all pairs of data in the feature space which is often computationally cheaper than the complex computation of the coordinates. This approach is called the "kernel trick." The kernel trick allows any non-linear model to be turned into a linear model by replacing its features (predictors) with a kernel function. All kernel algorithms are based on convex optimization or eigen problems and are statistically well-founded.

Definition 3.1. *Let X be a non-empty set. Then a function $k: X \times X \to \mathbb{R}$ is called a kernel on X if there exists a $\mathbb{R}$-Hilbert space H and a map $\phi: X \to H$ such that for all $x,\ x' \in X$ we have,*

$$k\left(x,\ x'\right) = \left\langle \phi(x'),\ \phi(x) \right\rangle \ .$$

We call ϕ a feature map and H a feature space of k (Steinwart & Christmann, 2008).

To further clarify Definition 3.1, let X be an arbitrary set. It can be said that H is a reproducing kernel Hilbert space (RKHS) on X over x$\mathbb{F}$ if the following is true (Paulsen & Raghupathi, 2016):

1. H is a vector subset of $f(x,\ \mathbb{F})$,
2. H is a vector subset of $f(x,\ \mathbb{F})$, is endowed with an inner product, $\langle .,. \rangle$, making it into a Hilbert space,
3. For every $y \in X$, the linear evaluation functional, $E_y : H \to \mathbb{F}$, defined by $E_y(f) = f(y)$, is bounded.

If H is an RKHS on X, then since every bounded linear functional is given by the inner product with a unique vector in H, there exists a unique vector, $k_y \in H$, for every $y \in X$, such that for every $f \in H,\ f(y) = \langle f, k_y \rangle$. This allows the reproducing kernel of H to be definde as a function $k: X \times X \to \mathbb{R}$ by $k(x,y) = k_y(x) = \left\langle k_y, k_x \right\rangle$.

From the above definitions, it is easily observed that $k: X \times X \to \mathbb{R}$ is both symmetric and a positive definite, i.e.,

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) > 0, \tag{3.1}$$

for any $n \in \mathbb{N}$, $x_1, \ldots, x_n \in X$ and $c_1, \ldots, c_n \in \mathbb{R}$ (Riesz, 1907).

Kernel Density Estimation

Multivariate Kernel Estimation

A basic multivariate kernel estimator of $f(x)$ can be written as such (Wand & Jones, 1994)

$$\hat{f}(x;H) = \frac{1}{n|\boldsymbol{H}|^{\frac{1}{2}}} \sum_{i=1}^{n} K(\boldsymbol{H}^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{X}_i)), \tag{3.2}$$

where $\boldsymbol{H}$ is defined as a bandwidth matrix.

Any kernel function that satisfies that conditions below can be used for K. For example, one of the most common kernel functions to use is the Gaussian (normal) kernel which can be written such that

$$K(z) = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}z'z}, \tag{3.3}$$

which satisfies

1. $\int K(z)\,dz = 1$
2. $K(z) = K(-z)$
3. $\int z'zK(z)\,dz = k_2 > 0$

and is a consistent estimator of $f(x)$.

Bandwidth Selection

The choice of bandwidth is crucial to the kernel density estimation (KDE) (Wand & Jones, 1995). Various bandwidth selection techniques for KDE have been developed in the past three decades. According to Scott (1992) and Bowman and Azzani (1997), the bandwidth matrix for multivariate data can be written using the generalized Scott's rule of thumb such that,

$$\hat{\boldsymbol{H}} = n^{\frac{-1}{p+4}}\hat{\boldsymbol{\Sigma}}^{\frac{1}{2}}, \tag{3.4}$$

where n is the sample size and p is the dimension of the data.

Estimation of $\boldsymbol{\mu_i}$

Assuming homogeneity of the variance-covariance matrix, $\boldsymbol{\Sigma}$, and $\boldsymbol{X_{ij}}$ comes from a distribution with probability density function (p.d.f.) $f_i(x)$ where $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, n_i$, with p dimension. That is to say, the following equation holds true:

$$f_i(\boldsymbol{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} f_0\left((\boldsymbol{x} - \boldsymbol{\mu_i})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu_i})\right), \tag{3.5}$$

where $f_0(.)$ is a base density function. Using the substitution technique where $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{\mu_i}$, we obtain

$$\begin{aligned}
\int \boldsymbol{x} f_i^2(\boldsymbol{x}) d\boldsymbol{x} &= \int \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y} \\
&= \int \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \boldsymbol{y} f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y} + \int \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \boldsymbol{\mu}_i f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y} \\
&= \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \int \boldsymbol{y} f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y} + \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \int \boldsymbol{\mu}_i f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y}. \qquad (3.6) \\
\boldsymbol{\mu}_i &= \frac{|\boldsymbol{\Sigma}|^{\frac{1}{2}} \int \boldsymbol{x} f_i^2(\boldsymbol{x}) d\boldsymbol{x} - \int \boldsymbol{y} f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y}}{\int f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y}} \\
&= \frac{1}{\int f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} \int \boldsymbol{x} f_i^2(\boldsymbol{x}) d\boldsymbol{x} \\
&- \frac{1}{\int f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y}} \int \boldsymbol{y} f_0^2(\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}) d\boldsymbol{y}. \qquad (3.7)
\end{aligned}$$

$\int f^2(\boldsymbol{x}) d\boldsymbol{x}$ is used instead of $\int f(\boldsymbol{x}) d\boldsymbol{x}$ to obtain a biased estimator of $\boldsymbol{\mu}_i$ as suggested by Ahmad (1982).

Nonparametric Kernel-Based One-Way MANOVA

Consider the multivariate one-way layout

$$\boldsymbol{Y}_{ij} = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{ij}. \tag{3.8}$$

Under the assumptions of homogeneity of variance-covariance where

$$\boldsymbol{\mu}_i = \frac{1}{\int f_0^2(\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})d\boldsymbol{y}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\int \boldsymbol{x} f_i^2(\boldsymbol{x})d\boldsymbol{x} - \frac{1}{\int f_0^2(\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})d\boldsymbol{y}}\int \boldsymbol{y} f_0^2(\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})d\boldsymbol{y}.$$

Let $\boldsymbol{Q}_i = \int \boldsymbol{x} f_i^2(\boldsymbol{x})d\boldsymbol{x}$, $c_1 = \frac{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}{\int f_0^2(\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})d\boldsymbol{y}}$, and $\boldsymbol{c}_2 = -\frac{1}{\int f_0^2(\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})d\boldsymbol{y}}\int \boldsymbol{y} f_0^2(\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y})d\boldsymbol{y}$. Thus, $\boldsymbol{\mu}_i = c_1\boldsymbol{Q}_i + \boldsymbol{c}_2$. Therefore, the tested null hypothesis can be written as $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_I$ versus the alternative hypothesis, $H_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$ for some $i \neq j$.

Now, consider the estimate of $\boldsymbol{Q}_i$ based on the extension of Ahmad's univariate estimation of location parameter (1982):

$$\hat{\boldsymbol{Q}}_i = \frac{1}{n_i(n_i-1)|\boldsymbol{H}_i|^{\frac{1}{2}}}\sum\sum_{j_1\neq j_2}\frac{1}{2}\left(\boldsymbol{X}_{ij_1} + \boldsymbol{X}_{ij_2}\right)K\left(\boldsymbol{H_i}^{\frac{-1}{2}}(\boldsymbol{X}_{ij_1} - \boldsymbol{X}_{ij_2})\right). \tag{3.9}$$

Lemma 3.1. *If for any* $i = 1,\ldots,I$, $n_i|\boldsymbol{H}_i|^4 \to 0,\ n_i|\boldsymbol{H}_i| \to \infty$ *as* $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty}\boldsymbol{x}^2 f_i^3(\boldsymbol{x})d\boldsymbol{x} < \infty$ *and if* $f_i(.)$ *is twice differentiable, then*

$$\sqrt{n_i}\left(\hat{\boldsymbol{Q}}_i - \boldsymbol{Q}_i\right) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\omega}_i^2), \tag{3.10}$$

as $\min_i n_i \to \infty$, *where,* $\boldsymbol{\omega}_i^2 = 4\left\{\int \boldsymbol{x}^2 f_i^3(\boldsymbol{x})d\boldsymbol{x} - \left(\int \boldsymbol{x} f_i^2(\boldsymbol{x})d\boldsymbol{x}\right)^2\right\}$.

Proof. Let $\varphi\left(\boldsymbol{X}_{ij_1}, \boldsymbol{X}_{ij_2}\right) = \frac{1}{2}\left(\boldsymbol{X}_{ij_1} + \boldsymbol{X}_{ij_2}\right)K\left(\boldsymbol{H_i}^{\frac{-1}{2}}(\boldsymbol{X}_{ij_1} - \boldsymbol{X}_{ij_2})\right)$ and $\hat{\mu}_i = c_1\hat{\boldsymbol{Q}}_i + \boldsymbol{c}_2$. Now, $\hat{\mu}_i$ is a U-statistics with mean

$$E(\hat{\boldsymbol{\mu}}_i) = c_1E(\hat{\boldsymbol{Q}}_i) + \boldsymbol{c}_2$$

$$= c_1 E\left(\frac{1}{|\boldsymbol{H}_i|^{1/2}}\boldsymbol{X}_{i1}K(\boldsymbol{H}_i^{\frac{-1}{2}}(\boldsymbol{X}_{i1}-\boldsymbol{X}_{i2}))\right)+\boldsymbol{c}_2$$

$$= c_1\frac{1}{|\boldsymbol{H}_i|^{1/2}}\iint \boldsymbol{x}K(\boldsymbol{H}_i^{-1/2}(\boldsymbol{x}-\boldsymbol{y}))f_i(\boldsymbol{x})f_i(\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y}+\boldsymbol{c}_2$$

$$= c_1\iint \boldsymbol{x}K(\boldsymbol{u})f_i(\boldsymbol{x})f_i(\boldsymbol{x}+\boldsymbol{H}_i^{1/2}\boldsymbol{u})d\boldsymbol{x}d\boldsymbol{u}+\boldsymbol{c}_2$$

$$= c_1\iint \boldsymbol{x}K(\boldsymbol{u})f_i(\boldsymbol{x})(f_i(\boldsymbol{x})+f_i^{(1)}(\boldsymbol{x})\boldsymbol{u}\boldsymbol{H}_i^{\frac{1}{2}}+o(|\boldsymbol{H}|)d\boldsymbol{x}d\boldsymbol{u})+\boldsymbol{c}_2$$

$$= c_1\int K(\boldsymbol{u})d\boldsymbol{u}\int \boldsymbol{x}f_i^2(\boldsymbol{x})d\boldsymbol{x}+O(|\boldsymbol{H}_i|)+\boldsymbol{c}_2$$

$$= c_1\int \boldsymbol{x}f_i^2(\boldsymbol{x})d\boldsymbol{x}+\boldsymbol{c}_2+O(|\boldsymbol{H}_i|)$$

$$\simeq \boldsymbol{\mu}_i, \tag{3.11}$$

and variance

$$Var(\hat{\mu}_i) = c_1^2\left(\frac{4}{n_i}cov(\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i2}),\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i3}))+\frac{2}{n_i(n_i-1)}var(\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i2}))\right). \tag{3.12}$$

It can easily be shown that $var(\varphi(\boldsymbol{X}_{ij_1},\boldsymbol{X}_{ij_2})) = O(|\boldsymbol{H}_i^{-1/2}|)$ and, since $1/(n_i|\boldsymbol{H}_i^{1/2}|) = o(1)$, the second term of $Var(\hat{\boldsymbol{\mu}}_i)$ can be ignored. Then, $Var(\hat{\boldsymbol{\mu}}_i)$ is dominated by $1/n_i\varpi_i^2$, where

$$
\begin{aligned}
\varpi_i^2 &= 4c_1^2 cov(\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i2}),\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i3}))\\
&= 4c_1^2[E(\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i2}),\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i3})) - E(\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i2}),\varphi(\boldsymbol{X}_{i1},\boldsymbol{X}_{i3}))]\\
&= 4c_1^2\left[\frac{1}{4|\boldsymbol{H}_i|}\iiint(\boldsymbol{x}+\boldsymbol{y})(\boldsymbol{x}+z)K(\boldsymbol{H}_i^{-1/2}(\boldsymbol{x}-\boldsymbol{y}))K(\boldsymbol{H}_i^{-1/2}(\boldsymbol{x}-z))f_i(\boldsymbol{x})f_i(\boldsymbol{y})f_i(z)d\boldsymbol{x}d\boldsymbol{y}dz\right.\\
&\left.-\left(E\left(\frac{\boldsymbol{X}_{i1}}{\boldsymbol{H}_i^{-1/2}}K(\boldsymbol{H}_i^{-1/2}(\boldsymbol{X}_{i1}-\boldsymbol{X}_{i2}))\right)\right)^2\right]\\
&= c_1^2\left[\iiint(2\boldsymbol{x}+\boldsymbol{H}_i^{1/2}\boldsymbol{u})(2\boldsymbol{x}+\boldsymbol{H}_i^{1/2}\boldsymbol{v})K(\boldsymbol{u})K(\boldsymbol{v})f_i(\boldsymbol{x}+\boldsymbol{H}_i^{1/2}\boldsymbol{u})f_i(\boldsymbol{x}+\boldsymbol{H}_i^{1/2}\boldsymbol{v})d\boldsymbol{x}d\boldsymbol{u}d\boldsymbol{v}\right.\\
&\left.-4\left(\frac{1}{|\boldsymbol{H}|_i}\iint \boldsymbol{x}K(\boldsymbol{H}_i^{-1/2}(\boldsymbol{x}-\boldsymbol{y}))f_i(\boldsymbol{x})f_i(\boldsymbol{y})d\boldsymbol{x}d\boldsymbol{y}\right)^2\right]\\
&= c_1^2\left[\iiint(4\boldsymbol{x}^T\boldsymbol{x}+2\boldsymbol{u}^T\boldsymbol{H}_i^{1/2}\boldsymbol{x}+2\boldsymbol{v}^T\boldsymbol{H}_i^{1/2}\boldsymbol{x}+\boldsymbol{u}^T\boldsymbol{H}_i^2\boldsymbol{v})K(\boldsymbol{u})K(\boldsymbol{v})f_i(\boldsymbol{x})(f_i(\boldsymbol{x})+f_i^{(1)}(\boldsymbol{x})\boldsymbol{H}_i^{1/2}\boldsymbol{u}\right.\\
&\left.+o(|\boldsymbol{H}_i|)(f_i(\boldsymbol{x})+f_i^{(1)}(\boldsymbol{x})\boldsymbol{H}_i\boldsymbol{v}+o(|\boldsymbol{H}_i|))d\boldsymbol{x}d\boldsymbol{u}d\boldsymbol{v}-4\left(\iint \boldsymbol{x}K(\boldsymbol{u})f_i(\boldsymbol{x})f_i(\boldsymbol{x}+\boldsymbol{H}_i^{1/2}\boldsymbol{u})d\boldsymbol{x}d\boldsymbol{u}\right)^2\right]\\
&= c_1^2\left[4\int \boldsymbol{x}^T\boldsymbol{x}f_i^3(\boldsymbol{x})d\boldsymbol{x}+o(|\boldsymbol{H}_i|)-4\left(\iint \boldsymbol{x}K(\boldsymbol{u})f_i(\boldsymbol{x})(f_i(\boldsymbol{x})+\boldsymbol{H}_i^{1/2}\boldsymbol{u}f_i^2(\boldsymbol{x})+o(|\boldsymbol{H}_i^{1/2}|))d\boldsymbol{x}d\boldsymbol{u}\right)^2\right]\\
&= 4c_1^2\left[\int \boldsymbol{x}^T\boldsymbol{x}f_i^3(\boldsymbol{x})d\boldsymbol{x}-\left(\int \boldsymbol{x}f_i^2(\boldsymbol{x})d\boldsymbol{x}\right)^2\right]+o(|\boldsymbol{H}_i|). \qquad (3.13)
\end{aligned}
$$

By using the Delta method and the central limit theorem of U-statistics (Koroljuk & Borovskich, 1994), we obtain $\sqrt{n_i}(\hat{\boldsymbol{\mu}}_i-\boldsymbol{\mu}_i)\xrightarrow{D}N_p(\boldsymbol{0},\varpi_i^2\boldsymbol{I})$. Thus, we obtain

$$\sqrt{n_i}(c_1\hat{\boldsymbol{Q}}_i+\boldsymbol{c}_2-(c_1\boldsymbol{Q}_i+\boldsymbol{c}_2)) = \sqrt{n_i}c_1(\hat{\boldsymbol{Q}}_i-\boldsymbol{Q}_i)\xrightarrow{D}N_p(\boldsymbol{0},\boldsymbol{\varpi}_i^2) \qquad (3.14)$$

Let $\boldsymbol{\omega}_i^2=\frac{1}{c_1^2}\boldsymbol{\varpi}_i^2$. Then, we have $\sqrt{n_i}(\hat{\boldsymbol{Q}}_i-\boldsymbol{Q}_i)\xrightarrow{D}N_p(\boldsymbol{0},\boldsymbol{\omega}_i^2)$, where

$$\boldsymbol{\omega}_i^2 = 4\left\{\int \boldsymbol{x}^T\boldsymbol{x}f_i^3(\boldsymbol{x})d\boldsymbol{x}-\left(\int \boldsymbol{x}f_i^2(\boldsymbol{x})d\boldsymbol{x}\right)^2\right\} \qquad (3.15)$$

Define the between sum of squares as

$$\boldsymbol{SSB} = \sum_{i=1}^{I} n_i \hat{\boldsymbol{\omega}}_i^{-2} \left(\hat{\boldsymbol{Q}}_i - \hat{\boldsymbol{Q}}_{.}\right)\left(\hat{\boldsymbol{Q}}_i - \hat{\boldsymbol{Q}}_{.}\right)^T, \tag{3.16}$$

where

$$\hat{\boldsymbol{Q}}_{.} = \sum_{i=1}^{I} \frac{1}{n_i} \hat{\boldsymbol{\omega}}_i^2 \sum_{i=1}^{I} n_i \hat{\boldsymbol{Q}}_i \hat{\boldsymbol{\omega}}_i^{-2}$$

and $\hat{\boldsymbol{\omega}}_i^2$ is a consistent estimate of $\boldsymbol{\omega}_i^2$. To obtain the asymptotic distribution of $\boldsymbol{SSB}$, two additional auxiliary variables need to be defined as:

$$\boldsymbol{S}_1^0 = \sum_{i=1}^{I} N\lambda_i \boldsymbol{\omega}_i^{-2} \left(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}\right)\left(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}\right)^T - N\left[\sum_{i=1}^{I} \lambda_i \boldsymbol{\omega}_i^{-2}\right]\left(\boldsymbol{Q}^* - \bar{\boldsymbol{Q}}\right)\left(\boldsymbol{Q}^* - \bar{\boldsymbol{Q}}\right)^T, \tag{3.17}$$

and

$$\boldsymbol{S}_1^{00} = \sum_{i=1}^{I} n_i \hat{\boldsymbol{\omega}}_i^{-2} \left(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}\right)\left(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}\right)^T - \sum_{i=1}^{I} n_i \hat{\boldsymbol{\omega}}_i^{-2} \left(\hat{\boldsymbol{Q}}_{.} - \bar{\boldsymbol{Q}}\right)\left(\hat{\boldsymbol{Q}}_{.} - \bar{\boldsymbol{Q}}\right)^T \tag{3.18}$$

where

$$\boldsymbol{Q}^* = \sum_{i=1}^{I} \frac{1}{\lambda_i} \boldsymbol{\omega}_i^2 \sum_{i=1}^{I} \lambda_i \boldsymbol{\omega}_i^{-2} \hat{\boldsymbol{Q}}_i^{-1}$$

$$\bar{\boldsymbol{Q}} = \sum_{i=1}^{I} \frac{1}{\lambda_i} \boldsymbol{\omega}_i^2 \sum_{i=1}^{I} \lambda_i \boldsymbol{\omega}_i^{-2} \boldsymbol{Q}_i^{-1}.$$

Lemma 3.2. *Let* $N = \sum_{i=1}^{I} n_i$. *If* $\lambda_i = \lim_{\min_{n_i} \to \infty} \frac{n_i}{N}$ *and* $\hat{\boldsymbol{\omega}}_i^2 \xrightarrow{P} \boldsymbol{\omega}_i^2$, *then*

1. $\boldsymbol{S}_1^0 - \boldsymbol{S}_1^{00} \xrightarrow{P} \boldsymbol{0}$,
2. $\boldsymbol{S}_1^{00} - \boldsymbol{SSB} \xrightarrow{P} \boldsymbol{0}$

as $\min_i n_i \to \infty$.

Proof. Both can be proved directly by applying Slutsky's Theorem (DasGupta, 2008).

Theorem 3.3. *Under the null hypothesis, if for any* $i = 1, 2, \ldots, I$, $n_i |\boldsymbol{H}_i|^4 \to 0, n_i |\boldsymbol{H}_i| \to \infty$ *as* $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} \boldsymbol{x}^2 f_i^3(\boldsymbol{x}) dx < \infty$ *and if* $f_i(.)$ *is twice* *differentiable, then* ***SSB*** *is asymptotically* $W_p(I-1,\boldsymbol{I})$*. Under the alternative,* ***SSB*** *is asymptotically non-central* $W_p(I-1,\boldsymbol{I})$ *with non-centrality matrix:*

$$\boldsymbol{\Delta}_1 = \frac{1}{2}\boldsymbol{\mu}^{(2)}\mathbf{B}_1\boldsymbol{\mu}^{(2)^T}, \tag{3.19}$$

where $\boldsymbol{\mu}^{(2)} = \left(\boldsymbol{\mu}_1^{(2)}, \boldsymbol{\mu}_2^{(2)}, \ldots, \boldsymbol{\mu}_I^{(2)}\right)$,

$$\boldsymbol{\mu}_i^{(2)} = \frac{\sqrt{\lambda_i}}{\sum \boldsymbol{\omega}_i}\left(\boldsymbol{e}_i - \frac{\sum_{i=1}^I \lambda_i e_i/\boldsymbol{\omega}_i^2}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2}\right)\int f^2(\boldsymbol{y})d\boldsymbol{y} \tag{3.20}$$

$$\mathbf{B_1} = \begin{bmatrix} 1-\frac{\lambda_1/\boldsymbol{\omega}_1^2}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2} & -\frac{(\sqrt{\lambda_1}/\boldsymbol{\omega}_1)(\sqrt{\lambda_2}/\boldsymbol{\omega}_2)}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2} & \cdots & -\frac{(\sqrt{\lambda_1}/\boldsymbol{\omega}_1)(\sqrt{\lambda_I}/\boldsymbol{\omega}_I)}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(\sqrt{\lambda_1}/\boldsymbol{\omega}_1)(\sqrt{\lambda_I}/\boldsymbol{\omega}_I)}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2} & -\frac{(\sqrt{\lambda_2}/\boldsymbol{\omega}_2)(\sqrt{\lambda_I}/\boldsymbol{\omega}_I)}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2} & \cdots & 1-\frac{\lambda_I/\boldsymbol{\omega}_I^2}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2} \end{bmatrix}, \tag{3.21}$$

where $\lambda_i = \lim_{n_i\to\infty} \frac{n_i}{\sum_{i=1}^I n_i}$ *and* $\boldsymbol{e}_i$ *is such that* $\boldsymbol{\mu}_i = 1 + \frac{1}{\sqrt{\sum_{i=1}^I n_i}}\boldsymbol{e}_i$.

Proof. Let $N = \sum_{i=1}^I n_i$, then $\lambda_i \simeq \frac{n_i}{N}$. Set $\boldsymbol{T}_i^{(2)} = \sqrt{\lambda_i N}/\boldsymbol{\omega}_i(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}})$, then under H_0, $\boldsymbol{T}_i^{(2)} \sim N_p(\mathbf{0},\boldsymbol{I})$ as $N\to\infty$ by Lemma (3.1). Note that

$$\begin{aligned} \boldsymbol{S}_1^0 =& \sum_{i=1}^I N\lambda_i\boldsymbol{\omega}_i^{-2}\left(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}\right)\left(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}\right)^T - N\left[\sum_{i=1}^I \lambda_i\boldsymbol{\omega}_i^{-2}\right]\left(\boldsymbol{Q}^* - \bar{\boldsymbol{Q}}\right)\left(\boldsymbol{Q}^* - \bar{\boldsymbol{Q}}\right)^T \\ =& \sum_{i=1}^I N\lambda_i\boldsymbol{\omega}_i^{-2}\left(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}\right)\left(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}\right)^T \\ & -N\left[\sum_{i=1}^I \lambda_i\boldsymbol{\omega}_i^{-2}\right]\left(\frac{1}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2}\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2\hat{\boldsymbol{Q}}_i - \frac{1}{\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2}\sum_{i=1}^I \lambda_i/\boldsymbol{\omega}_i^2\boldsymbol{Q}_i\right) \end{aligned}$$

$$\left(\frac{1}{\sum_{i=1}^{I}\lambda_i/\boldsymbol{\omega}_i^2}\sum_{i=1}^{I}\lambda_i/\boldsymbol{\omega}_i^2\hat{\boldsymbol{Q}}_i - \frac{1}{\sum_{i=1}^{I}\lambda_i/\boldsymbol{\omega}_i^2}\sum_{i=1}^{I}\lambda_i/\boldsymbol{\omega}_i^2\boldsymbol{Q}_i\right)^T$$

$$=\sum_{i=1}^{I}N\lambda_i\boldsymbol{\omega}_i^{-2}\left(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}\right)\left(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}\right)^T-\frac{N}{\sum_{i=1}^{I}\frac{\lambda_i}{\boldsymbol{\omega}_i^2}}\left(\sum_{i=1}^{I}\frac{\lambda_i}{\boldsymbol{\omega}_i^2}\right)^2\left(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}\right)\left(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}\right)^T$$

$$=\sum_{i=1}^{I}\frac{N\lambda_i}{\boldsymbol{\omega}_i^2}\left(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}\right)\left(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}\right)^T-\sum_{i=1}^{I}\sum_{j=1}^{I}\frac{\sqrt{N\lambda_i}}{\boldsymbol{\omega}_i}\left(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}\right)\frac{\sqrt{N\lambda_j}}{\boldsymbol{\omega}_j}$$

$$\left(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}\right)^T\frac{(\sqrt{\lambda_i}/\boldsymbol{\omega}_i)(\sqrt{\lambda_j}/\boldsymbol{\omega}_j)}{\sum_{i=1}^{I}\lambda_i/\boldsymbol{\omega}_i^2}, \tag{3.22}$$

which can be written in a quadratic form such that $\boldsymbol{S}_1^0 = \mathbf{U}_1^T\mathbf{B}_1\mathbf{U}_1$ where

$$\mathbf{U}_1 = (\boldsymbol{T}_1^{(2)}, \boldsymbol{T}_2^{(2)}, \ldots, \boldsymbol{T}_I^{(2)})^T. \tag{3.23}$$

Since $\mathbf{B}_1$ is symmetric and idempotent, we obtain

$$\begin{aligned} rank(\mathbf{B}_1) &= tr(\mathbf{B}_1) \\ &= tr(\boldsymbol{I} - \boldsymbol{a}_2\boldsymbol{a}_2^T) \\ &= tr(\boldsymbol{I}) - tr(\boldsymbol{a}_2\boldsymbol{a}_2^T) \\ &= tr(\boldsymbol{J}) - tr(\boldsymbol{a}_2\boldsymbol{a}_2^T) \\ &= I-1, \end{aligned} \tag{3.24}$$

where

$$\boldsymbol{a}_2 = \begin{pmatrix} \sqrt{\lambda_1}\boldsymbol{\omega}_1^{-1} \\ \sqrt{\lambda_2}\boldsymbol{\omega}_2^{-1} \\ \vdots \\ \sqrt{\lambda_I}\boldsymbol{\omega}^{-1}{}_I \end{pmatrix}. \tag{3.25}$$

$\mathbf{U}_1$ approximately follows a multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}$. Therefore, $\boldsymbol{S}_1^0$ is asymptotically distributed as W_p with degrees of freedom $I-1$ and scale matrix $\boldsymbol{I}$ under H_0.

Under the alternative, since $\boldsymbol{e}_i$ is chosen such that $\boldsymbol{\mu}_i = \mathbf{1} + \frac{1}{\sqrt{N}}\boldsymbol{e}_i$, then we have

$$\begin{aligned} \boldsymbol{Q}_i &= \frac{1}{c_1}\boldsymbol{\mu}_i - \frac{1}{c_1}\boldsymbol{c}_2 \\ &= \frac{1}{c_1}1 - \boldsymbol{c}_2 + \frac{1}{c_1\sqrt{N}}\boldsymbol{e}_i. \end{aligned} \tag{3.26}$$

Thus, $\bar{\boldsymbol{Q}}$ can be written as such:

$$\begin{aligned} \bar{\boldsymbol{Q}} &= \sum^{I_{i=1}} \frac{1}{\lambda_i}\boldsymbol{\omega}_i^2 \sum_{i=1}^{I} \lambda_i \boldsymbol{\omega}_i^{-2}\boldsymbol{Q}_i \\ &= \frac{\sum_{i=1}^{I} \frac{\lambda_i}{\boldsymbol{\omega}_i^2}\left(\frac{1-\boldsymbol{c}_2}{c_1} + \frac{\boldsymbol{e}_i}{c_1\sqrt{N}}\right)}{\sum_{i=1}^{I}\frac{\lambda_i}{\boldsymbol{\omega}_i^2}} \\ &= \frac{1}{c_1}1 - \boldsymbol{c}_2 + \frac{1}{c_1\sqrt{N}}\sum_{i=1}^{I}\lambda_i\boldsymbol{e}_i\boldsymbol{\omega}_i^{-2}\frac{1}{\sum_{i=1}^{I}\lambda_i\boldsymbol{\omega}_i^{-2}} \ . \end{aligned} \tag{3.27}$$

Under H_1, $\boldsymbol{T}_i^{(2)} = \frac{\sqrt{N}\lambda_i}{\boldsymbol{\omega}_i}(\hat{\boldsymbol{Q}}_i - \bar{\boldsymbol{Q}}) = \frac{\sqrt{N}\lambda_i}{\boldsymbol{\omega}_i}(\hat{\boldsymbol{Q}}_i - \boldsymbol{Q}_i) + \frac{\sqrt{N}\lambda_i}{\boldsymbol{\omega}_i}(\boldsymbol{Q}_i - \bar{\boldsymbol{Q}})$. By Lemma (3.1), the first term, $\frac{\sqrt{N}\lambda_i}{\omega_i}(\hat{\boldsymbol{Q}}_i - \boldsymbol{Q}_i)$, is approximately distributed as a standard normal. The second term's limit is

$$\begin{aligned} \boldsymbol{\mu} &= \lim_{N\to\infty}\frac{\sqrt{N}\lambda_i}{\boldsymbol{\omega}_i}\left(\boldsymbol{Q}_i - \bar{\boldsymbol{Q}}\right) \\ &= \lim_{N\to\infty}\frac{\sqrt{N}\lambda_i\frac{1}{c_1\sqrt{N}}}{\boldsymbol{\omega}_i}\left(\boldsymbol{e}_i - \sum_{i=1}^{I}\lambda_i\boldsymbol{e}_i\boldsymbol{\omega}_i^{-2}\frac{1}{\sum_{i=1}^{I}\lambda_i\boldsymbol{\omega}_i^{-2}}\right) \\ &= \frac{\sqrt{\lambda_i}\left(\boldsymbol{e}_i - \sum_{i=1}^{I}\lambda_i\boldsymbol{e}_i\boldsymbol{\omega}_i^{-2}\frac{1}{\sum_{i=1}^{I}\lambda_i\boldsymbol{\omega}_i^{-2}}\right)}{c_1\boldsymbol{\omega}_i} \\ &= |\boldsymbol{\Sigma}|^{-1/2}\frac{\sqrt{\lambda_i}}{\boldsymbol{\omega}_i}\left(\boldsymbol{e}_i - \sum_{i=1}^{I}\lambda_i\boldsymbol{e}_i\boldsymbol{\omega}_i^{-2}\frac{1}{\sum_{i=1}^{I}\lambda_i\boldsymbol{\omega}_i^{-2}}\right)\int f^2(\boldsymbol{y}\boldsymbol{\Sigma}^{-1}\boldsymbol{y})d\boldsymbol{y}, \end{aligned} \tag{3.28}$$

where $\boldsymbol{\omega}_i$ is given by the square root of equation (3.15). Thus, generally $\boldsymbol{T}_i^{(2)}$ has an approximately normally distribution with mean $\boldsymbol{\mu}^{(2)}$ given in Equation (3.19) and variance $\boldsymbol{I}$, which implies that $\boldsymbol{S}_1^0 = \mathbf{U^{(2)^T}B_1U^{(2)}}$ is asymptotically non-central $W_p(I-1)$ with scale matrix $\boldsymbol{I}$ and non-centrality matrix $\boldsymbol{\Delta}_1 = \frac{1}{2}\boldsymbol{\mu}^{(2)}\mathbf{B}_1\boldsymbol{\mu}^{(2)^T}$.

By Lemma (3.2), $\boldsymbol{SSB}$ converges in probability to $\boldsymbol{S}_1^0$. Therefore, $\boldsymbol{SSB}$ has asymptotic W_p distribution with $I-1$ degrees of freedom and scale matrix $\boldsymbol{I}$ under the null hypothesis and asymptotic non-central W_p distribution with $I-1$ degrees of freedom and scale matrix $\boldsymbol{I}$, with non-centrality matrix, $\boldsymbol{\Delta}_1 = \frac{1}{2}\boldsymbol{\mu}^{(2)}\mathbf{B_1}\boldsymbol{\mu}^{(2)^T}$.

Let $\boldsymbol{A}_{ij_1j_2} = \frac{1}{|H|^{1/2}}\frac{1}{2}\left(\boldsymbol{X}_{ij_1}+\boldsymbol{X}_{ij_2}\right)K(H^{-1/2}(X_{ij_1}-X_{ij_2}))$. Then, $\hat{\boldsymbol{Q}}_i$ can be written as:

$$\hat{\boldsymbol{Q}}_i = \frac{1}{n_i(n_i-1)}\sum\sum_{j_1\neq j_2}\boldsymbol{A}_{ij_1j_2}. \tag{3.29}$$

Define $\boldsymbol{S}_2^0$ as

$$\boldsymbol{S}_2^0 = \sum_{i=1}^{I}\sum\sum_{j_1\neq j_2}^{n_i}\boldsymbol{\omega}_i^{-2}\left(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i\right)\left(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i\right)^T, \tag{3.30}$$

then, the within sum of squares can be written as:

$$\boldsymbol{SSW} = \sum_{i=1}^{I}\sum\sum_{j_1\neq j_2}^{n_i}\frac{1}{C_w}\boldsymbol{\omega}_i^{-2}\left(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i\right)\left(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i\right)^T, \tag{3.31}$$

where

$$C_w = \begin{cases} \frac{n}{2}-1 \; if \; n_i = n \text{ for all } i, \\ c_0 \text{ otherwise,} \end{cases} \tag{3.32}$$

and $c_0 = \sum_{i=1}^{d}\frac{\pi_i}{d}$, $\pi_1.\pi_2,\ldots,\pi_d$ are the eigenvalues of $\mathbf{B}_2$, where

$$\mathbf{B}_2 = \begin{bmatrix} \left(\frac{n_1}{2}-1\right)\left(\boldsymbol{I}_{n_1}-\frac{1}{n_1}\boldsymbol{J}_{n_1}\right) & 0 & \cdots & 0 \\ 0 & \left(\frac{n_2}{2}-1\right)\left(\boldsymbol{I}_{n_2}-\frac{1}{n_2}\boldsymbol{J}_{n_2}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\frac{n_I}{2}-1\right)\left(\boldsymbol{I}_{n_I}-\frac{1}{n_I}\boldsymbol{J}_{n_I}\right) \end{bmatrix}. \tag{3.33}$$

Lemma 3.4. *Let $N=\sum_{i=1}^{I}n_i$. If $\lambda_i = \lim_{n_i\to\infty}\frac{n_i}{N}$ and $\hat{\boldsymbol{\omega}}_i^2 \xrightarrow{P} \boldsymbol{\omega}_i^2$, then $\boldsymbol{S}_2^0 - C_w\boldsymbol{SSW} \xrightarrow{P} 0$ as $\min_i n_i \to \infty$.*

Proof. This can be proved directly by applying Slutsky's Theorem (DasGupta, 2008).

Theorem 3.5. *For any* $i = 1,2,\ldots,I, n_i|\boldsymbol{H}_i^4| \to 0, n_i|\boldsymbol{H}_i| \to \infty$ *as* $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} \boldsymbol{x}^2 f_i^3(\boldsymbol{x}) d\boldsymbol{x} < \infty$ *and if* $f_i(.)$ *is twice differentiable, then* $\boldsymbol{SSW}$ *is asymptotically* $W_p(I-1,\boldsymbol{I})$. *Under the alternative,* $\boldsymbol{SSW}$ *is asymptotically distributed as* W_p *with* df_w *degrees of freedom and scale matrix* $\boldsymbol{I}$, *where*

$$df_w = \begin{cases} I(n-1), \text{ if } \mathrm{n_i} = \mathrm{n} \text{ for all } i \\ d, \text{ otherwise} \end{cases}, \tag{3.34}$$

where d is the number of eigenvalues of $\mathbf{B_2}$ *given in Equation (3.33).*

Proof. By the Hajek projection (Hajek, 1986), $\boldsymbol{A}_{ij_1j_2}$ can be decomposed into the sum of conditional expected values and a residual as shown below:

$$\boldsymbol{A}_{ij_1j_2} = E\left(\boldsymbol{A}_{ij_1j_2} \mid \boldsymbol{X}_{ij_1}\right) + E\left(\boldsymbol{A}_{ij_1j_2} \mid \boldsymbol{X}_{ij_2}\right) + O_p(n_i). \tag{3.35}$$

Set $\varphi\left(\boldsymbol{X}_{ij_1}\right) = E\left(\boldsymbol{A}_{ij_1j_2} \mid \boldsymbol{X}_{ij_1}\right)$ and $\varphi\left(\boldsymbol{X}_{ij_2}\right) = E\left(\boldsymbol{A}_{ij_1j_2} \mid \boldsymbol{X}_{ij_2}\right)$, therefore,

$$\begin{aligned}
\hat{\boldsymbol{Q}}_i &= \frac{1}{n_i(n_i-1)} \sum\sum_{j_1 \neq j_2} \boldsymbol{A}_{ij_1j_2} \\
&\approx \frac{1}{n_i(n_i-1)} \sum\sum_{j_1 \neq j_2} \left(\varphi\left(\boldsymbol{X}_{ij_1}\right) + \varphi\left(\boldsymbol{X}_{ij_2}\right)\right) \\
&= \frac{1}{n_i(n_i-1)} \left(\sum_{j_1}\sum_{j_2} \left(\varphi\left(\boldsymbol{X}_{ij_1}\right) + \varphi\left(\boldsymbol{X}_{ij_2}\right)\right) - \sum\sum_{j_1 \neq j_2} \left(\varphi\left(\boldsymbol{X}_{ij_1}\right) + \varphi\left(\boldsymbol{X}_{ij_2}\right)\right)\right) \\
&= \frac{1}{n_i(n_i-1)} \left(2n_i \sum_{j_1} \varphi\left(\boldsymbol{X}_{ij_1}\right) - 2\sum_{j_1} \varphi\left(\boldsymbol{X}_{ij_1}\right)\right) \\
&= \frac{1}{n_i} \sum_{j_1}^{n_i} 2\varphi\left(\boldsymbol{X}_{ij_1}\right).
\end{aligned} \tag{3.36}$$

Thus,

$$\sum\sum_{j_1\neq j_2}\boldsymbol{\omega}_i^{-2}\left(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i\right)\left(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i\right)^T \approx \sum\sum_{j_1\neq j_2}\boldsymbol{\omega}_i^{-2}\left(\varphi\left(\boldsymbol{X}_{ij_1}\right)+\varphi\left(\boldsymbol{X}_{ij_2}\right)-\hat{\boldsymbol{Q}}_i\right)^2$$
$$=\sum\sum_{j_1\neq j_2}\boldsymbol{\omega}_i^{-2}(\varphi(\boldsymbol{X}_{ij_1})+\frac{1}{n_i}\sum\varphi(\boldsymbol{X}_{ij_1})+\varphi(\boldsymbol{X}_{ij_2})-\frac{1}{n_i}\sum\varphi(\boldsymbol{X}_{ij_2}))^2-\sum_{j_1}\boldsymbol{\omega}_i^{-2}(2\varphi(\boldsymbol{X}_{ij_1})$$
$$-\frac{1}{n_i}\sum_{j_1=1}^{n_i}2\varphi(\boldsymbol{X}_{ij_2}))^2=2n_i\sum_{j_1}\omega_i^{-2}\boldsymbol{I}\left(\varphi\left(\boldsymbol{X}_{ij_1}\right)-\frac{1}{n_i}\sum\varphi\left(\boldsymbol{X}_{ij_1}\right)\right)^2-\sum_{j_1}\boldsymbol{\omega}_i^{-2}\left(2\varphi\left(\boldsymbol{X}_{ij_1}\right)-\hat{\boldsymbol{Q}}_i\right)^2$$
$$=\frac{n_i}{2}\sum_{j_1}\boldsymbol{\omega}_i^{-2}\left(2\varphi\left(\boldsymbol{X}_{ij_1}\right)-\hat{\boldsymbol{Q}}_i\right)^2-\sum_{j_1}\boldsymbol{\omega}_i^{-2}\left(2\varphi\left(\boldsymbol{X}_{ij_1}\right)-\hat{\boldsymbol{Q}}_i\right)^2=\left(\frac{n_i}{2}-1\right)\sum_{j_1}\boldsymbol{\omega}_i^{-2}\left(2\varphi\left(\boldsymbol{X}_{ij_1}\right)-\hat{\boldsymbol{Q}}_i\right)^2$$
$$=\left(\frac{n_i}{2}-1\right)\left[\sum_{j_1=1}^{n_i}\boldsymbol{\omega}_i^{-2}\left(2\varphi\left(\boldsymbol{X}_{ij_1}\right)-\boldsymbol{Q}_i\right)^2-\boldsymbol{\omega}_i^{-2}n_i\left(\hat{\boldsymbol{Q}}_i-\boldsymbol{Q}_i\right)^2\right], \tag{3.37}$$

because $\sum_{j_1=1}^{n_i}\left(2\varphi\left(\boldsymbol{X}_{ij_1}\right)-\boldsymbol{Q}_i\right)^2=\sum_{j_1=1}^{n_i}\left(2\varphi\left(\boldsymbol{X}_{ij_1}\right)-\hat{\boldsymbol{Q}}_i\right)^2+n_i\left(\hat{\boldsymbol{Q}}_i-\boldsymbol{Q}_i\right)^2$.

Now let $\mathbf{H}_{\mathrm{ij}}=\boldsymbol{\omega}_i^{-1}2\varphi\left(\boldsymbol{X}_{ij}\right)-\boldsymbol{Q}_i$ for $j=1,2,\ldots,n_i$ and $\mathbf{H}_{\mathrm{i}}=\left(\mathbf{H}_{\mathrm{i}1},\mathbf{H}_{\mathrm{i}2},\ldots,\mathbf{H}_{\mathrm{in_i}}\right)^{\mathrm{T}}$ for $i=1,2,\ldots,I$. Thus, equation (3.37) can be written in a matrix form such that

$$\left(\frac{n_i}{2}-1\right)\left[\sum_{j_1=1}^{n_i}\boldsymbol{\omega}_i^{-2}\left(2\varphi\left(\boldsymbol{X}_{ij_1}\right)-\boldsymbol{Q}_i\right)^2-\boldsymbol{\omega}_i^{-2}n_i\left(\hat{\boldsymbol{Q}}_i-\boldsymbol{Q}_i\right)^2\right]$$

$$=\left(\frac{n_i}{2}-1\right)\left[\mathbf{H}_{\mathrm{i}}^{\mathrm{T}}\mathbf{H}_{\mathrm{i}}-\mathbf{H}_{\mathrm{i}}^{\mathrm{T}}\frac{1}{n_i}J_{n_i}\mathbf{H}_{\mathrm{i}}\right]$$
$$=\mathbf{H}_{\mathrm{i}}^{\mathrm{T}}\left(\frac{n_i}{2}-1\right)\left(\boldsymbol{I}-\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\mathbf{H}_{\mathrm{i}}\ . \tag{3.38}$$

Let $\mathbf{H}=(\mathbf{H}_1^T,\mathbf{H}_2^T,\ldots,\mathbf{H}_I^T)^{\mathrm{T}}$. Hence, $\boldsymbol{S}_2^0$ can be written in a matrix form as

$$\boldsymbol{S}_2^0=\sum_{i=1}^{I}\sum\sum_{j_1\neq j_2}^{n_i}\boldsymbol{\omega}_i^{-2}\left(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i\right)^2$$

$$=\sum_{i=1}^{I}\mathbf{H}_i^{\mathrm{T}}\left(\frac{n_i}{2}-1\right)\left(\boldsymbol{I}-\frac{1}{n_i}\boldsymbol{J}_{n_i}\right)\mathbf{H}_i$$

$$= \mathbf{H}^T \begin{bmatrix} (\frac{n_1}{2}-1)(\boldsymbol{I}_{n_1}-\frac{1}{n_1}\boldsymbol{J}_{n_1}) & 0 & \dots & 0 \\ 0 & (\frac{n_2}{2}-1)(\boldsymbol{I}_{n_2}-\frac{1}{n_2}\boldsymbol{J}_{n_2}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\frac{n_I}{2}-1)(\boldsymbol{I}_{n_I}-\frac{1}{n_I}\boldsymbol{J}_{n_I}) \end{bmatrix} \mathbf{H}$$

$$= \mathbf{H}^{\mathrm{T}}\mathbf{B}_2\mathbf{H}. \tag{3.39}$$

Now, we need to show that $\mathbf{H}$ is asymptotically distributed as a multivariate normal.

Note that $E(2\varphi(\boldsymbol{X}_{ij_1})) = E(\frac{1}{n_i}\sum_{j_1=1}^{n_i} 2\varphi(\boldsymbol{X}_{ij_1})) = E(\hat{\boldsymbol{Q}}_i) \to \boldsymbol{Q}_i$ since $\hat{\boldsymbol{\mu}}_i$ is asymptotically unbiased by equation (3.11). Additionally,
$Var(2\varphi(\boldsymbol{X}_{ij_1})) = \frac{1}{n_i}\sum_{j_1=1}^{n_i} Var(2\varphi(\boldsymbol{X}_{ij_1})) = n_i Var(\frac{1}{n_i}\sum_{j_1=1}^{n_i} 2\varphi(\boldsymbol{X}_{ij_1})) = n_i Var(\hat{\boldsymbol{Q}}_i) = \boldsymbol{\omega}_i^2$.

By the central limit theorem of U-statistics, $\mathrm{H}_{ij} = \frac{1}{\boldsymbol{\omega}_i} 2\varphi(\boldsymbol{X}_{ij_1}) - \boldsymbol{Q}_i$ has asymptotically normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}$. Since H_i's are independent, then H has asymptotically multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}$.

1. If $n_i = n$ for all i, then it is easy to verify that $\frac{1}{(\frac{n}{2}-1)}\mathbf{B}_2$ is a symmetric and idempotent matrix with rank $\sum_{i=1}^{I} n_i - I = N - I = I(n-1)$. Thus, $\frac{1}{(\frac{n}{2}-1)}\boldsymbol{S}_2^0$ is asymptotically distributed as a W_p with $I(n-1)$ degrees of freedom and scale matrix $\boldsymbol{I}$. By using Lemma (3.4), the within sum of squares, $\boldsymbol{SSW}$, is asymptotically distributed as a W_p with $I(n-1)$ degrees of freedom and scale matrix $\boldsymbol{I}$.
2. If $n_i \neq n$ for some $i \neq j$, $\mathbf{B}_2$ is symmetric but not idempotent. So, there exists $\mathbf{H}^{\mathrm{T}}\mathbf{B}_2\mathbf{H} = \sum_{i=1}^{d} \pi_i z_i^2$, where $\pi_1, \pi_2, \dots, \pi_d$ are the eigenvalues of $\mathbf{B}_2$ and $z_i \sim N(\mathbf{0}, \boldsymbol{I})$ which are independent. Let $c_0 = \sum_{i=1}^{d} \pi_i / d$, then using (Yuan & Bentler, 2010), $\boldsymbol{S}_2^0 / c_0 = \mathbf{H}^{\mathrm{T}}\mathbf{B}_2\mathbf{H}/c_0 \mathrel{\dot{\sim}} W_p(d, \boldsymbol{I})$. Thus. By using Lemma (3.4), the within sum of squares, $\boldsymbol{SSW}$, is asymptotically distributed as a W_p with d degrees of freedom and scale matrix $\boldsymbol{I}$.

Define the Λ-test statistics of kernel based nonparametric test for one-way MANOVA as

$$\Lambda_{\mathrm{k}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSB}+\boldsymbol{SSW}|} = \frac{1}{1+\left|\frac{\boldsymbol{SSB}}{\boldsymbol{SSW}}\right|}$$

$$= \frac{1}{1+\left|\frac{\sum_{i=1}^{I}\hat{\boldsymbol{\omega}}_i^{-2} n_i(\hat{\boldsymbol{Q}}_i-\hat{\boldsymbol{Q}}_.)(\hat{\boldsymbol{Q}}_i-\hat{\boldsymbol{Q}}_.)^T}{\sum_{i=1}^{I}\sum\sum_{j_1\neq j_2}^{n_i}\boldsymbol{\omega}_i^{-2}(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i)(\boldsymbol{A}_{ij_1j_2}-\hat{\boldsymbol{Q}}_i)^T}\right|}. \tag{3.40}$$

Note that $\boldsymbol{SSB} \dot{\sim} W_p(I-1,\boldsymbol{I})$ and $\boldsymbol{SSW} \dot{\sim} W_p(df_w,\boldsymbol{I})$, where df_w is given in Equation (3.33).

Theorem 3.6. *If for any* $i=1,2,\ldots,I, n_i|\boldsymbol{H}_i^4| \to 0,\ n_i|\boldsymbol{H}_i| \to \infty$ *as* $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty}\boldsymbol{x}^2 f_i^3(\boldsymbol{x})d\boldsymbol{x} < \infty$ *and if* $f_i(.)$ *is twice differentiable, then under the null hypothesis,* Λ_{k} *in Equation (3.40) has a function of asymptotic F distribution with degrees of freedom* df_1 *and* df_2*. Under the alternative,* Λ_{k} *has asymptotically non-central* $F(df_1,df_2)$*with non-centrality matrix* $\boldsymbol{\Delta}_1$ *given in Equation (3.19).*

Proof. Theorem (3.3) shows that $\boldsymbol{SSB}$ follows asymptotically W_p with degrees of freedom $I-1$ and scale matrix $\boldsymbol{I}$ under the null hypothesis and asymptotically non-central $W_p(I-1,\boldsymbol{I})$ under the alternative. Furthermore, Theorem (3.5) implies that the within sum of squares, $\boldsymbol{SSW}$, is asymptotically W_p with degrees of freedom $I(n-1)$ and scale matric $\boldsymbol{I}$ for balanced data and W_p with degrees of freedom d and scale matrix $\boldsymbol{I}$ for unbalanced data, where d is the number of eigenvalues of $\mathbf{B}_2$ in Equation (3.33). In order to show that Λ_{k} follows asymptotically F distribution under the null hypothesis and non-central F distribution under the alternative hypothesis, we need to show that $\boldsymbol{SSB}$ and $\boldsymbol{SSW}$ are asymptotically independent as $\min_i n_i \to \infty$. In theorem (3.3), $\boldsymbol{S}_1^0$, which converges in probability to $\boldsymbol{SSB}$, is written in a quadratic form as $\boldsymbol{S}_1^0 = \mathbf{U}_1^{\mathrm{T}}\mathbf{B}_1\mathbf{U}_1$. Note that under the null hypothesis

$$\boldsymbol{T}_i^{(2)} \simeq \sqrt{n_i}\boldsymbol{\omega}_i^{-1}(\hat{\boldsymbol{Q}}_i-\bar{\boldsymbol{Q}}) = \frac{1}{\sqrt{n_i}}\mathrm{H}_{\mathrm{i}}^{\mathrm{T}}\boldsymbol{J}_{n_i}. \tag{3.41}$$

Thus, $\boldsymbol{S}_1^0$ can be written as

$$\boldsymbol{S}_1^0 = \left(\frac{1}{\sqrt{n_1}}\mathbf{H}_1^{\mathrm{T}}\boldsymbol{j}_{n_1},\ \frac{1}{\sqrt{n_2}}\mathbf{H}_2^{\mathrm{T}}\boldsymbol{j}_{n_2},\ldots,\frac{1}{\sqrt{n_I}}\mathbf{H}_I^{\mathrm{T}}\boldsymbol{j}_{n_I}\right)\mathbf{B}_1\begin{pmatrix}\frac{1}{\sqrt{n_1}}\mathbf{H}_1^{\mathrm{T}}\boldsymbol{j}_{n_1}\\ \vdots\\ \frac{1}{\sqrt{n_I}}\mathbf{H}_I^{\mathrm{T}}\boldsymbol{j}_{n_I}\end{pmatrix}$$

$$= \mathbf{H}^{\mathrm{T}}\begin{bmatrix}\frac{1}{\sqrt{n_1}}\boldsymbol{j}_{n_1}^T & 0 & \ldots & 0\\ 0 & \frac{1}{\sqrt{n_2}}\boldsymbol{j}_{n_2}^T & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \frac{1}{\sqrt{n_I}}\boldsymbol{j}_{n_I}^T\end{bmatrix}^T \mathbf{B}_1\begin{bmatrix}\frac{1}{\sqrt{n_1}}\boldsymbol{j}_{n_1}^T & 0 & \ldots & 0\\ 0 & \frac{1}{\sqrt{n_2}}\boldsymbol{j}_{n_2}^T & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \frac{1}{\sqrt{n_I}}\boldsymbol{j}_{n_I}^T\end{bmatrix}\mathbf{H}$$

$$= \mathbf{H}^{\mathrm{T}}\mathbf{B}_3\mathbf{H}. \tag{3.42}$$

Recall from Theorem (3.5) that $\boldsymbol{S}_2^0 = \mathbf{H}^{\mathrm{T}}\mathbf{B}_2\mathbf{H}$. Now, it is a straightforward process to check that

$$\mathbf{B}_2\mathbf{B}_3 = \mathbf{0}\times\mathbf{B}_1\begin{bmatrix}\frac{1}{\sqrt{n_1}}\boldsymbol{j}_{n_1}^T & 0 & \ldots & 0\\ 0 & \frac{1}{\sqrt{n_2}}\boldsymbol{j}_{n_2}^T & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \frac{1}{\sqrt{n_I}}\boldsymbol{j}_{n_I}^T\end{bmatrix} = \mathbf{0}. \tag{3.43}$$

Thus, $\boldsymbol{S}_1^0$ and $\boldsymbol{S}_2^0$ are independent. By Lemma (3.2) and Lemma(3.4), ***SSB*** and ***SSW*** are asymptotically independent under null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_I$.

According to Mardia, Kent, and Bibby (1979) Λ_{k} can be related to the F-distribution.

Let

$$a_1 = df_w + \frac{(I-1)-p-1}{2}, \tag{3.44}$$

$$b_1 = \begin{cases}\sqrt{\frac{p^2(I-1)^2-4}{p^2+(I-1)^2-5}}\ ,\ p^2+(I-1)^2-5>0\\ 1,\text{otherwise}\end{cases}, \tag{3.45}$$

and

$$c_1 = \frac{p(I-1)}{2} - 1. \tag{3.46}$$

Now, let $df_1 = p(I-1)$ and $df_2 = a_1 b_1 - c_1$. Thus, the F approximation is

$$F = \frac{1-(\Lambda_{\text{k}})^{1/b_1}}{(\Lambda_{\text{k}})^{1/b_1}} \times \frac{df_2}{df_1}. \tag{3.47}$$

Therefore, under the null hypothesis, Λ_{k} in Equation (3.40) follows asymptotically F distribution with degrees of freedom df_1 and df_2. Under the alternative, Λ_{k} follows asymptotically non-central F distribution with degrees of freedom df_1 and df_2 with non-centrality parameter $\boldsymbol{\Delta}_1$ described in Equation (3.19).

Nonparametric Kernel-Based Two-Way MANOVA

Assuming homogeneity of the variance-covariance matrix, $\boldsymbol{\Sigma}$, and $\boldsymbol{X_{ij}}$ comes from a distribution with probability density function (p.d.f.) $f_i(x)$, where $i = 1,2,\ldots,I$, $j = 1,2,\ldots,J$, and $k = 1,2,\ldots,n_{ij}$ with p dimension. That is to say, the following equation holds:

$$f_{ij}(\boldsymbol{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} f_{00}\left((\boldsymbol{x}-\boldsymbol{\mu_{ij}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})\right) \tag{3.48}$$

Under the assumptions of homogeneity of variance-covariance where

$$\boldsymbol{\mu}_{ij} = \frac{1}{\int f_{00}^2(\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y})d\boldsymbol{y}}|\boldsymbol{\Sigma}|^{\frac{1}{2}} \int \boldsymbol{x} f_{ij}^2(\boldsymbol{x})d\boldsymbol{x} - \frac{1}{\int f_{00}^2(\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y})d\boldsymbol{y}} \int \boldsymbol{y} f_{00}^2(\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y})d\boldsymbol{y}. \tag{3.49}$$

Let $\boldsymbol{Q}_{ij} = \int \boldsymbol{x} f_{ij}^2(\boldsymbol{x})d\boldsymbol{x}$, $C_1^0 = \frac{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}{\int f_{00}^2(\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y})d\boldsymbol{y}}$, and $\boldsymbol{C}_2^0 = \frac{1}{\int f_{00}^2(\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y})d\boldsymbol{y}} \int \boldsymbol{y} f_{00}^2(\boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y})d\boldsymbol{y}$.

Then, $\boldsymbol{\mu}_{ij}$ can be written as

$$\boldsymbol{\mu}_{ij} = C_1^0\boldsymbol{Q}_{ij} + \boldsymbol{C}_2^0. \tag{3.50}$$

Now, consider the two-way layout

$$\boldsymbol{X}_{ijk} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} + \boldsymbol{\varepsilon}_{ijk}, \tag{3.51}$$

where $\boldsymbol{X}_{ijk}$ is the vector of observation and has $p \times 1$ dimension, $\boldsymbol{\mu}$ is the overall mean or location and has $p \times 1$ dimension, $\boldsymbol{\alpha}_i$ is the ith row effect, $\boldsymbol{\beta}_j$ is the jth column effect, and $\boldsymbol{\gamma}_{ij}$ is interaction effect of ith row and jth column. The decomposition is not unique. Therefore, the following restrictions are imposed:

$$\sum_i \lambda_{ij}\boldsymbol{\omega}_{ij}^{-2}\boldsymbol{\alpha}_i = \mathbf{0}, \tag{3.52}$$

$$\sum_j \lambda_{ij}\boldsymbol{\omega}_{ij}^{-2}\boldsymbol{\beta}_j = \mathbf{0}, \tag{3.53}$$

$\sum_i \lambda_{ij}\boldsymbol{\omega}_{ij}^{-2}\boldsymbol{\gamma}_{ij} = \sum_j \lambda_{ij}\boldsymbol{\omega}_{ij}^{-2}\boldsymbol{\gamma}_{ij} = \mathbf{0}$, where $\lambda_{ij} = \lim_{\min_{i,j}} n_{ij} \to \infty \frac{n_{ij}}{N}$ and $N = \sum_{i,j} n_{ij}$. Thus, we can conclude that

$$\boldsymbol{\mu} + \boldsymbol{\alpha}_i = \sum_j \lambda_{ij}\boldsymbol{\mu}_{ij}\boldsymbol{\omega}_{ij}^{-2} \sum_j 1/\lambda_{ij}\boldsymbol{\omega}_{ij}^2, \tag{3.54}$$

$$\boldsymbol{\mu} + \boldsymbol{\beta}_j = \sum_i \lambda_{ij}\boldsymbol{\mu}_{ij}\boldsymbol{\omega}_{ij}^{-2} \sum_i 1/\lambda_{ij}\boldsymbol{\omega}_{ij}^2, \tag{3.55}$$

$$\boldsymbol{\mu} = \sum_i \sum_j \lambda_{ij}\boldsymbol{\mu}_{ij}\boldsymbol{\omega}_{ij}^{-2} \sum_i \sum_j 1/\lambda_{ij}\boldsymbol{\omega}_{ij}^2. \tag{3.56}$$

Thus, using equation (3.50) in equations (3.54 - 3.56), we obtain

$$\boldsymbol{\mu} + \boldsymbol{\alpha}_i = C_1^0 \bar{\boldsymbol{Q}}_{i.} + \boldsymbol{C}_2^0, \tag{3.57}$$

$$\boldsymbol{\mu} + \boldsymbol{\beta}_j = C_1^0 \bar{\boldsymbol{Q}}_{.j} + \boldsymbol{C}_2^0, \tag{3.58}$$

$$\boldsymbol{\mu} = C_1^0 \bar{\boldsymbol{Q}}_{..} + \boldsymbol{C}_2^0, \tag{3.59}$$

where

$$\bar{\boldsymbol{Q}}_{i.} = \sum_j \frac{m_{ij}}{m_{i.}} \boldsymbol{Q}_{ij},$$

$$\bar{\boldsymbol{Q}}_{.j} = \sum_j \frac{m_{ij}}{m_{j.}} \boldsymbol{Q}_{ij},$$

$$\bar{\boldsymbol{Q}}_{..} = \sum_i \sum_j \frac{m_{ij}}{m_{..}} \boldsymbol{Q}_{ij},$$

$m_{ij} = N\lambda_{ij}\boldsymbol{\omega}_{ij}^{-2}, m_{i.} = \sum_j m_{ij},\ m_j = \sum_i m_{ij}$, and $m_{..} = \sum_i \sum_j m_{ij}$. Thus, the following can be written as

$$\boldsymbol{\alpha}_i = C_1^0(\bar{\boldsymbol{Q}}_{i.} - \bar{\boldsymbol{Q}}_{..}), \tag{3.60}$$

$$\boldsymbol{\beta}_j = C_1^0(\bar{\boldsymbol{Q}}_{.j} - \bar{\boldsymbol{Q}}_{..}), \tag{3.61}$$

$$\boldsymbol{\gamma}_{ij} = \boldsymbol{\mu}_{ij} - (\boldsymbol{\mu}_{.} + \boldsymbol{\alpha}_i) - (\boldsymbol{\mu} - \boldsymbol{\beta}_j) + \boldsymbol{\mu}_{.}$$

$$= C_1^0(\boldsymbol{Q}_{ij} - \bar{\boldsymbol{Q}}_{i.} - \bar{\boldsymbol{Q}}_{.j} + \bar{\boldsymbol{Q}}_{..}). \tag{3.62}$$

Hence, the null hypothesis for testing the row effect can be written as $H_0 : \boldsymbol{\alpha}_i = \mathbf{0}$ versus the alternative hypothesis $H_1 : \boldsymbol{\alpha}_i \neq \mathbf{0}$ for some i. Also, the null hypothesis for testing the column effect can be written as $H_0 : \boldsymbol{\beta}_j = \mathbf{0}$ versus the alternative hypothesis $H_1 : \boldsymbol{\beta}_j \neq \mathbf{0}$ for some j. Similarly, the hypothesis for testing the interaction effect of the rows and columns can be written as $H_0 : \boldsymbol{\gamma}_{ij} = \mathbf{0}$ versus the alternative hypothesis $H_1 : \boldsymbol{\gamma}_{ij} \neq \mathbf{0}$ for some i and j.

Main Effect

Let's consider the nonparametric kernel estimate of $\boldsymbol{Q}_{ij}$, $\hat{\boldsymbol{Q}}_{ij}$, is defined as

$$\hat{\boldsymbol{Q}}_{ij} = \frac{1}{n_{ij}(n_{ij}-1)|\boldsymbol{H}_{ij}|^{\frac{1}{2}}} \sum\sum_{k_1 \neq k_2} \frac{1}{2}(\boldsymbol{X}_{ijk_1} + \boldsymbol{X}_{ijk_2}) K(\boldsymbol{H}_{ij}^{-1/2}(\boldsymbol{X}_{ijk_1} - \boldsymbol{X}_{ijk_2})). \tag{3.63}$$

Lemma 3.7. *If for any* $i = 1, \ldots, I,\ j = 1, \ldots, J,\ n_i|\boldsymbol{H}_{ij}^4| \to 0,\ n_i|\boldsymbol{H}_{ij}| \to \infty$ *as* $\min_{i,j} n_{ij} \to \infty$, $\int_{-\infty}^{\infty} \boldsymbol{u} f^2(\boldsymbol{u}) d\boldsymbol{u} < \infty$ *and if* $\int \boldsymbol{u}^2 f(\boldsymbol{u}) d\boldsymbol{u} < \infty$, *then*

$$\sqrt{n_{ij}}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\omega}_{ij}^2), \tag{3.64}$$

as $\min_{i,j} n_{ij} \to \infty$, *where* $\boldsymbol{\omega}_{ij}^2 = 4\left\{\int \boldsymbol{x}^2 f_{ij}^3(\boldsymbol{x}) d\boldsymbol{x} - \left(\int \boldsymbol{x} f_{ij}^2(\boldsymbol{x}) d\boldsymbol{x}\right)^2\right\}$.

Proof. The proof is similar to Lemma (3.7), showed previously in this chapter.

In order to test $H_0 : \boldsymbol{\gamma}_{ij} = \mathbf{0}$ for all i, define the Row Sum of Squares (***SSR***) as

$$\boldsymbol{SSR} = \sum_{i=1}^{I} \hat{m}_{i.}(\hat{\boldsymbol{Q}}_{i.} - \hat{\boldsymbol{Q}}_{..})(\hat{\boldsymbol{Q}}_{i.} - \hat{\boldsymbol{Q}}_{..})^T, \tag{3.65}$$

where $\hat{\boldsymbol{Q}}_{i.} = \sum_j \frac{\hat{m}_{ij}}{\hat{m}_{i.}}\hat{\boldsymbol{Q}}_{ij}, \hat{\boldsymbol{Q}}_{..} = \sum_i \sum_j \frac{\hat{m}_{ij}}{\hat{m}_{..}}\hat{\boldsymbol{Q}}_{ij}, \hat{m}_{i.} = \sum_j \hat{m}_{ij},\ \hat{m}_{ij} = n_{ij}\hat{\boldsymbol{\omega}}_{ij}^{-2}$, and $\hat{\boldsymbol{\omega}}_{ij}^2$ is a consistent estimate of $\boldsymbol{\omega}_{ij}^2$.

To obtain the asymptotic distribution of ***SSR***, an additional auxiliary variable needs to be defined as

$$\boldsymbol{S}_R^0 = \sum_{i=1}^{I} m_{i.}(\boldsymbol{Q}_{i.}^* - \boldsymbol{Q}_{..}^*)(\boldsymbol{Q}_{i.}^* - \boldsymbol{Q}_{..}^*)^T, \tag{3.66}$$

where

$$\boldsymbol{Q}_{i.}^* = \sum_j \frac{m_{ij}}{m_{i.}}\hat{\boldsymbol{Q}}_{ij}$$

$$\boldsymbol{Q}_{..}^* = \sum_i \sum_j \frac{m_{ij}}{m_{..}}\hat{\boldsymbol{Q}}_{ij}.$$

Lemma 3.8. *Let* $N = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}$. *If* $\lambda_{ij} = \lim_{\min n_{ij}} \rightarrow \infty \frac{n_{ij}}{N}$ *and* $\hat{\boldsymbol{\omega}}_{ij}^2 \xrightarrow{p} \boldsymbol{\omega}_{ij}^2$, *then* $\boldsymbol{S}_R^0 - \boldsymbol{SSR} \xrightarrow{P} 0$ *as* $\min_{i,j} n_{ij} \rightarrow \infty$.

Proof. This can be proved directly by applying Slutsky Theorem (DasGupta, 2008).

Theorem 3.9. *Under the null hypothesis, if for any* $i = 1, 2, \ldots, I$, *and* $j = 1, 2, \ldots, J$, $n_{ij}|\boldsymbol{H}_{ij}|^4 \rightarrow 0$, $n_{ij}|\boldsymbol{H}_{ij}| \rightarrow \infty$ *as* $\min_{i,j} n_{ij} \rightarrow \infty$, *and if* $\int_{-\infty}^{\infty} \boldsymbol{x}^2 f_{ij}^3(\boldsymbol{x})d\boldsymbol{x} < \infty$, *then* ***SSR*** *is asymptotically distributed as* $W_p(I-1, \boldsymbol{I})$

Proof. Set $\boldsymbol{T}_{ij}^{(2)} = \sqrt{\lambda_{ij}N}\boldsymbol{\omega}_{ij}^{-1}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})$, then under H_0, $\boldsymbol{T}_{ij}^{(2)} \sim N(\mathbf{0}, \boldsymbol{I})$ as $N \rightarrow \infty$ by Lemma (3.7). Note that under the null hypothesis,

$$\begin{aligned}
\boldsymbol{S}_R^0 &= \sum_{i=1}^{I} m_{i.}(\boldsymbol{Q}_{i.}^* - \boldsymbol{Q}_{..}^*)(\boldsymbol{Q}_{i.}^* - \boldsymbol{Q}_{..}^*)^T \\
&= \sum_{i=1}^{I} m_{i.}\Big((\boldsymbol{Q}_{i.}^* - \bar{\boldsymbol{Q}}_{i.}) - (\boldsymbol{Q}_{..}^* - \bar{\boldsymbol{Q}}_{..})\Big)^2 \\
&= \sum_{i=1}^{I} m_{i.}(\boldsymbol{Q}_{i.}^* - \bar{\boldsymbol{Q}}_{i.})^2 + m_{..}(\boldsymbol{Q}_{..}^* - \bar{\boldsymbol{Q}}_{..})^2 - 2(\boldsymbol{Q}_{..}^* - \bar{\boldsymbol{Q}}_{..})\sum_{i=1}^{I} m_{i.}\sum_{j}\frac{m_{ij}}{m_{i.}}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij}) \\
&= \sum_{i=1}^{I} m_{i.}(\boldsymbol{Q}_{i.}^* - \bar{\boldsymbol{Q}}_{i.})^2 - m_{..}(\boldsymbol{Q}_{..}^* - \bar{\boldsymbol{Q}}_{..})^2 \\
&= \sum_{i=1}^{I}\frac{1}{m_{i.}}\left(\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})\right)^2 - \frac{1}{m_{..}}\left(\sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})\right)^2 \\
&= \sum_{i=1}^{I}\sum_{j_1=1}^{J}\sum_{j_2=1}^{J}\frac{N}{m_{i.}}\frac{\sqrt{N\lambda_{i_1 j_1}}}{\boldsymbol{\omega}_{i_1 j_1}}(\hat{\boldsymbol{Q}}_{i_1 j_1} - \boldsymbol{Q}_{i_1 j_1})\frac{\sqrt{N\lambda_{i_2 j_2}}}{\boldsymbol{\omega}_{i_2 j_2}}(\hat{\boldsymbol{Q}}_{i_2 j_2} - \boldsymbol{Q}_{i_2 j_2})\frac{\sqrt{\lambda_{i_1 j_1}}}{\boldsymbol{\omega}_{i_1 j_1}}\frac{\sqrt{\lambda_{i_2 j_2}}}{\boldsymbol{\omega}_{i_2 j_2}},
\end{aligned} \tag{3.67}$$

which can be written in two quadratic forms. The first term can be written in the form $\mathbf{U}_2^{\mathrm{T}}\mathbf{M}^{(1)}\mathbf{U}_2$, and the second term can be written in the form $\mathbf{U}_2^{\mathrm{T}}\mathbf{M}^{(2)}\mathbf{U}_2$, where $\mathbf{U}_2 = (\mathbf{T}_{11}^{(2)}, \mathbf{T}_{12}^{(2)}, \ldots, \mathbf{T}_{IJ}^{(2)})^{\mathrm{T}}$,

$$\mathbf{M}^{(1)} = \begin{bmatrix} \frac{N}{m_{1.}}\mathbf{M}_{11} & 0 & \ldots & 0 \\ 0 & \frac{N}{m_{2.}}\mathbf{M}_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{N}{m_{I.}}\mathbf{M}_{II} \end{bmatrix}, \tag{3.68}$$

$$\mathbf{M}^{(2)} = \begin{bmatrix} \frac{N}{m_{..}}\mathbf{M}_{11} & \frac{N}{m_{..}}\mathbf{M}_{12} & \ldots & \frac{N}{m_{..}}\mathbf{M}_{1I} \\ \frac{N}{m_{..}}\mathbf{M}_{21} & \frac{N}{m_{..}}\mathbf{M}_{22} & \ldots & \frac{N}{m_{..}}\mathbf{M}_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{N}{m_{..}}\mathbf{M}_{I1} & \frac{N}{m_{..}}\mathbf{M}_{I2} & \ldots & \frac{N}{m_{..}}\mathbf{M}_{II} \end{bmatrix}, \tag{3.69}$$

and

$$\mathbf{M}_{\mathrm{ij}} = \begin{bmatrix} \sqrt{\lambda_{i1}}\sqrt{\lambda_{j1}}\boldsymbol{\omega}_{i1}^{-1}\boldsymbol{\omega}_{j1}^{-1} & \sqrt{\lambda_{i1}}\sqrt{\lambda_{j2}}\boldsymbol{\omega}_{i1}^{-1}\boldsymbol{\omega}_{j2}^{-1} & \dots & \sqrt{\lambda_{i1}}\sqrt{\lambda_{jJ}}\boldsymbol{\omega}_{i1}^{-1}\boldsymbol{\omega}_{jJ}^{-1} \\ \sqrt{\lambda_{i2}}\sqrt{\lambda_{j1}}\boldsymbol{\omega}_{i2}^{-1}\boldsymbol{\omega}_{j1}^{-1} & \sqrt{\lambda_{i2}}\sqrt{\lambda_{j2}}\boldsymbol{\omega}_{i2}^{-1}\boldsymbol{\omega}_{j2}^{-1} & \dots & \sqrt{\lambda_{i2}}\sqrt{\lambda_{jJ}}\boldsymbol{\omega}_{i2}^{-1}\boldsymbol{\omega}_{jJ}^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_{iJ}}\sqrt{\lambda_{j1}}\boldsymbol{\omega}_{iJ}^{-1}\boldsymbol{\omega}_{j1}^{-1} & \sqrt{\lambda_{iJ}}\sqrt{\lambda_{j2}}\boldsymbol{\omega}_{iJ}^{-1}\boldsymbol{\omega}_{jJ}^{-1} & \dots & \sqrt{\lambda_{iJ}}\sqrt{\lambda_{jJ}}\boldsymbol{\omega}_{iJ}^{-1}\boldsymbol{\omega}_{jJ}^{-1} \end{bmatrix}, \quad (3.70)$$

for all $i = 1,2,\dots,I,\ j = 1,2,..J$. From that we can conclude that $\boldsymbol{S}_R^0$ can be written in a quadratic form such as

$$\begin{aligned} \boldsymbol{S}_R^0 &= \mathbf{U}_2^{\mathrm{T}}\mathbf{M}^{(1)}\mathbf{U}_2 - \mathbf{U}_2^{\mathrm{T}}\mathbf{M}^{(2)}\mathbf{U}_2 \\ &= \mathbf{U}_2^{\mathrm{T}} \begin{bmatrix} \left(\frac{N}{m_{1.}} - \frac{N}{m_{..}}\right)\mathbf{M}_{11} & -\frac{N}{m_{..}}\mathbf{M}_{12} & \dots & -\frac{N}{m_{..}}\mathbf{M}_{1\mathrm{I}} \\ -\frac{N}{m_{..}}\mathbf{M}_{21} & \left(\frac{N}{m_{2.}} - \frac{N}{m_{..}}\right)\mathbf{M}_{22} & \dots & -\frac{N}{m_{..}}\mathbf{M}_{2\mathrm{I}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{N}{m_{..}}\mathbf{M}_{\mathrm{I}1} & -\frac{N}{m_{..}}\mathbf{M}_{\mathrm{I}2} & \dots & \left(\frac{N}{m_{I.}} - \frac{N}{m_{..}}\right)\mathbf{M}_{\mathrm{II}} \end{bmatrix} \mathbf{U}_2 \\ &= \mathbf{U}_2^{\mathrm{T}}\mathbf{B}_4\mathbf{U}_2. \end{aligned} \quad (3.71)$$

Since $\mathbf{B}_4$ is symmetric and idempotent, we obtain

$$\begin{aligned} rank(\mathbf{B}_4) &= tr(\mathbf{B}_4) \\ &= \sum_{i=1}^{I} \frac{1}{m_{i.}} \sum_{i=1}^{I} m_{ij} - \frac{1}{m_{..}} \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij} \\ &= I - 1. \end{aligned} \quad (3.72)$$

$\mathbf{U}_2$ approximately follows a multivariate normal distribution with mean $\mathbf{0}$ and variance $\mathbf{I}$, since $\boldsymbol{T}_{ij}^{(2)}$'s independently follow univariate standard normal distributions. Therefore, $\boldsymbol{S}_R^0$ is asymptotically W_p with degrees of freedom $I-1$ and scale matrix $\boldsymbol{I}$ under H_0. Hence, by Lemma (3.8), $\boldsymbol{SSR}$ asymptotically W_p with degrees of freedom $I-1$ and sclae matrix $\boldsymbol{I}$.

Now, Let $\boldsymbol{A}_{ijk_1k_2} = \frac{1}{|\boldsymbol{H}_i|^{\frac{1}{2}}} \frac{1}{2}(\boldsymbol{X}_{ijk_1} + \boldsymbol{X}_{ijk_2})K(\boldsymbol{H}_i^{-1/2}(\boldsymbol{X}_{ijk_1} - \boldsymbol{X}_{ijk_2}))$. Then, $\hat{\boldsymbol{Q}}_i$ can be written as:

$$\hat{\boldsymbol{Q}}_i = \frac{1}{n_{ij}(n_{ij}-1)} \sum \sum_{k_1 \neq k_2} \boldsymbol{A}_{ijk_1k_2}. \tag{3.73}$$

Then, the within sum of squares can be written as:

$$\boldsymbol{SSW} = \frac{1}{C_{w_2}} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum \sum_{k_1 \neq k_2} \boldsymbol{\omega}_{ij}^{-2} (\boldsymbol{A}_{ijk_1k_2} - \hat{\boldsymbol{Q}}_{ij})(\boldsymbol{A}_{ijk_1k_2} - \hat{\boldsymbol{Q}}_{ij})^T \tag{3.74}$$

where

$$C_{w_2} = \begin{cases} \frac{n}{2} - 1, \text{ if } \mathrm{n_{ij}} = \mathrm{n} \text{ for all } i \\ c_0^{(2)}, \text{ otherwise} \end{cases}, \tag{3.75}$$

$c_0^{(2)} = \sum_{i=1}^{d_2} \frac{\pi_i^{(2)}}{d_2}$, $\pi_1^{(2)}.\pi_2^{(2)}, \ldots, \pi_d^{(2)}$ are the eigenvalues of $\mathbf{B}_5$, where

$$\mathbf{B}_5 = \begin{bmatrix} (\frac{n_{11}}{2}-1)(\boldsymbol{I}_{n_1} - \frac{1}{n_{11}}\mathrm{J}_{n_{11}}) & 0 & \cdots & 0 \\ 0 & (\frac{n_{12}}{2}-1)(\boldsymbol{I}_{n_{12}} - \frac{1}{n_{12}}\mathrm{J}_{n_{12}}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\frac{n_{IJ}}{2}-1)(\boldsymbol{I}_{n_{IJ}} - \frac{1}{n_{IJ}}\mathrm{J}_{n_{IJ}}) \end{bmatrix}. \tag{3.76}$$

To obtain the asymptotic distribution of $\boldsymbol{SSW}$, an additional auxiliary variable is defined as

$$\boldsymbol{S}_W^0 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum \sum_{k_1 \neq k_2} \boldsymbol{\omega}_{ij}^{-2} (\boldsymbol{A}_{ijk_1k_2} - \hat{\boldsymbol{Q}}_{ij})(\boldsymbol{A}_{ijk_1k_2} - \hat{\boldsymbol{Q}}_{ij})^T. \tag{3.77}$$

Lemma 3.10. *Let $N = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$. If $\lambda_{ij} = \lim_{\min\, n_{ij} \to \infty} \frac{n_{ij}}{N}$ and $\hat{\boldsymbol{\omega}}_{ij}^2 \xrightarrow{p} \boldsymbol{\omega}_{ij}^2$, then $\boldsymbol{S}_W^0 - C_{w_2}\boldsymbol{SSW} \xrightarrow{p} 0$ as $\min_{i,j} n_{ij} \to \infty$.*

Proof. This can be proven by applying Slutsky Theorem (DasGupta, 2008).

Theorem 3.11. *For any $i = 1, 2, \ldots, I$,and $j = 1, 2, \ldots, J, n_{ij}|\boldsymbol{H}_{ij}^4| \to 0$, $n_{ij}|\boldsymbol{H}_{ij}|$ $\to \infty$ as $\min_{i,j} n_{ij} \to \infty, \int_{-\infty}^{\infty} \boldsymbol{x}^2 f_{ij}^3(\boldsymbol{x}) d\boldsymbol{x} < \infty$, then $\boldsymbol{SSW}$ is asymptotically distributed as W_p*

with df_{w_2} degrees of freedom and sclae matrix $\boldsymbol{I}$, where

$$df_{w_2} = \begin{cases} IJ(n-1), \text{ if } \mathrm{n_{ij} = n \text{ for all }} i \\ d_2, \text{ otherwise} \end{cases}, \tag{3.78}$$

where d_2 is the number of eigenvalues of $\mathbf{B}_5$ given in Equation (3.76).

Proof. By the Hajek projection (Hajek, 1986), $\boldsymbol{A}_{ijk_1k_2}$ can be decomposed into the sum of conditional expected values and a residual as follows:

$$\boldsymbol{A}_{ijk_1k_2} = E(\boldsymbol{A}_{ijk_1k_2} | \boldsymbol{X}_{ijk_1}) + E(\boldsymbol{A}_{ijk_1k_2} | \boldsymbol{X}_{ijk_2}) + O(n_{ij}). \tag{3.79}$$

Set $\varphi(\boldsymbol{X}_{ijk_1}) = E(\boldsymbol{A}_{ijk_1k_2} | \boldsymbol{X}_{ijk_1})$ and $\varphi(\boldsymbol{X}_{ijk_2}) = E(\boldsymbol{A}_{ijk_1k_2} | \boldsymbol{X}_{ijk_2})$, therefore,

$$\begin{aligned}
\hat{\boldsymbol{Q}}_{ij} &= \frac{1}{n_{ij}(n_{ij}-1)} \sum\sum_{k_1 \neq k_2} \boldsymbol{A}_{ijk_1k_2} \\
&\approx \frac{1}{n_{ij}(n_{ij}-1)} \sum\sum_{k_1 \neq k_2} (\varphi(\boldsymbol{X}_{ijk_1}) + \varphi(\boldsymbol{X}_{ijk_2})) \\
&= \frac{1}{n_{ij}(n_{ij}-1)} \left(\sum_{k_1}\sum_{k_2} (\varphi(\boldsymbol{X}_{ijk_1}) + \varphi(\boldsymbol{X}_{ijk_2})) - \sum\sum_{k_1 \neq k_2} (\varphi(\boldsymbol{X}_{ijk_1}) + \varphi(\boldsymbol{X}_{ijk_2})) \right) \\
&= \frac{1}{n_{ij}(n_{ij}-1)} \left(2n_{ij} \sum_{k_1} \varphi(\boldsymbol{X}_{ijk_1}) - 2 \sum_{k_1} \varphi(\boldsymbol{X}_{ijk_1}) \right) \\
&= \frac{1}{n_{ij}} \sum_{k_1}^{n_{ij}} 2\varphi(\boldsymbol{X}_{ijk_1}).
\end{aligned} \tag{3.80}$$

Thus,

$$
\begin{aligned}
\sum\sum_{k_1\neq k_2} \boldsymbol{\omega}_{ij}^{-2}(\boldsymbol{A}_{ijk_1k_2}-\hat{\boldsymbol{Q}}_{ij})^2 &\approx \sum\sum_{k_1\neq k_2} \boldsymbol{\omega}_{ij}^{-2}(\varphi(\boldsymbol{X}_{ijk_1})+\varphi(\boldsymbol{X}_{ijk_2})-\hat{\boldsymbol{Q}}_{ij})^2 \\
&= \sum_{k_1}\sum_{k_2}\boldsymbol{\omega}_{ij}^{-2}\left(\varphi(\boldsymbol{X}_{ijk_1})+\frac{1}{n_{ij}}\sum\varphi(\boldsymbol{X}_{ijk_1})+\varphi(\boldsymbol{X}_{ijk_2})-\frac{1}{n_{ij}}\sum\varphi(\boldsymbol{X}_{ijk_2})\right)^2 \\
&-\sum_{k_1}\boldsymbol{\omega}_{ij}^{-2}\left(2\varphi(\boldsymbol{X}_{ijk_1})-\frac{1}{n_{ij}}\sum_{k_1=1}^{n_{ij}}2\varphi(\boldsymbol{X}_{ijk_1})\right)^2 \\
&= 2n_{ij}\sum_{k_1}\boldsymbol{\omega}_{ij}^{-2}\left(\varphi(\boldsymbol{X}_{ijk_1})-\frac{1}{n_{ij}}\sum\varphi(\boldsymbol{X}_{ijk_1})\right)^2-\sum_{k_1}\boldsymbol{\omega}_{ij}^{-2}(2\varphi(\boldsymbol{X}_{ijk_1})-\hat{\boldsymbol{Q}}_{ij})^2 \\
&= \frac{n_{ij}}{2}\sum_{k_1}\boldsymbol{\omega}_{ij}^{-2}(2\varphi(\boldsymbol{X}_{ijk_1})-\hat{Q}_{ij})^2-\sum_{k_1}\boldsymbol{\omega}_{ij}^{-2}(2\varphi(\boldsymbol{X}_{ijk_1})-\hat{\boldsymbol{Q}}_{ij})^2 \\
&= \left(\frac{n_{ij}}{2}-1\right)\sum_{k_1}\boldsymbol{\omega}_{ij}^{-2}(2\varphi(\boldsymbol{X}_{ijk_1})-\hat{\boldsymbol{Q}}_{ij})^2 \\
&= \left(\frac{n_{ij}}{2}-1\right)\left[\sum_{k_1=1}^{n_{ij}}\boldsymbol{\omega}_{ij}^{-2}(2\varphi(\boldsymbol{X}_{ijk_1})-\boldsymbol{Q}_{ij})^2-\boldsymbol{\omega}_{ij}^{-2}n_{ij}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})^2\right]. \qquad (3.81)
\end{aligned}
$$

Now let $\mathbf{H}_{\mathrm{ij}}=\boldsymbol{\omega}_{ij}^{-1}2\varphi(\boldsymbol{X}_{ijk})-\boldsymbol{Q}_{ij}$ for $j=1,2,\ldots,n_{ij}$ and $\mathbf{H}_{ij}=(\mathbf{H}_{ij_1},\mathbf{H}_{ij_2},\ldots,\mathbf{H}_{ij_{n_{ij}}})^{\mathrm{T}}$ for $i=1,2,\ldots,I$, and $j=1,2,\ldots,J$. Thus, equation (3.81) can be written in a matrix form as

$$
\begin{aligned}
&\left(\frac{n_{ij}}{2}-1\right)\left[\sum_{k=1}^{n_{ij}}\boldsymbol{\omega}_{ij}^{-2}(2\varphi(\boldsymbol{X}_{ijk_1})-\boldsymbol{Q}_{ij})^2-\boldsymbol{\omega}_{ij}^{-2}n_{ij}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})^2\right] \\
&= \left(\frac{n_{ij}}{2}-1\right)\left[\mathbf{H}_{\mathrm{ij}}^{\mathrm{T}}\mathbf{H}_{\mathrm{ij}}-\mathbf{H}_{\mathrm{ij}}^{\mathrm{T}}\frac{1}{n_{ij}}\boldsymbol{J}_{n_{ij}}\mathbf{H}_{\mathrm{ij}}\right] \\
&= \mathbf{H}_{\mathrm{ij}}^{\mathrm{T}}\left(\frac{n_{ij}}{2}-1\right)\left(I-\frac{1}{n_{ij}}\boldsymbol{J}_{n_{ij}}\right)\mathbf{H}_{\mathrm{ij}}. \qquad (3.82)
\end{aligned}
$$

Let $\mathbf{H} = \left(\mathbf{H}_{11}^{\mathrm{T}}, \mathbf{H}_{12}^{\mathrm{T}}, \ldots, \mathbf{H}_{\mathrm{IJ}}^{\mathrm{T}}\right)^{\mathrm{T}}$. Hence, $\boldsymbol{S}_W^0$ can be written in a matrix form as

$$\begin{aligned}\boldsymbol{S}_W^0 &= \sum_{i=1}^{I}\sum_{j=1}^{J}\sum\sum_{k_1\neq k_2} \boldsymbol{\omega}_{ij}^{-2}(\boldsymbol{A}_{ijk_1k_2} - \hat{\boldsymbol{Q}}_{ij})(\boldsymbol{A}_{ijk_1k_2} - \hat{\boldsymbol{Q}}_{ij})^T \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J} \mathbf{H}_{\mathrm{ij}}^{\mathrm{T}} \left(\frac{n_{ij}}{2} - 1\right)\left(I - \frac{1}{n_{ij}}\,\mathrm{J}_{n_{ij}}\right)\mathbf{H}_{\mathrm{ij}} \\ &= \mathbf{H}^{\mathrm{T}}\mathbf{B}_5\mathbf{H}. \end{aligned} \tag{3.83}$$

Now, we need to show that $\mathbf{H}$ is asymptotically distributed as multivariate normal. Note that $E(2\varphi(\boldsymbol{X}_{ijk_1})) = E\left(\frac{1}{n_{ij}}\sum_{k_1=1}^{n_{ij}} 2\varphi(\boldsymbol{X}_{ijk_1})\right) = E(\hat{\boldsymbol{Q}}_{ij}) \approx \boldsymbol{Q}_{ij}$ since $\hat{\boldsymbol{\mu}}_{ij}$ is asymptotically unbiased by Lemma (3.10). Additionally, $Var(2\varphi(\boldsymbol{X}_{ijk_1})) = \frac{1}{n_{ij}}\sum_{k_1=1}^{n_{ij}} Var(2\varphi(\boldsymbol{X}_{ijk_1})) = n_{ij}Var(\frac{1}{n_{ij}}\sum_{k_1=1}^{n_{ij}} 2\varphi(\boldsymbol{X}_{ijk_1})) = n_{ij}Var(\hat{\boldsymbol{Q}}_{ij}) = \boldsymbol{\omega}_{ij}^2$. By the central limit theorem of U-statistics, $\mathrm{H}_{ijk} = \frac{1}{\boldsymbol{\omega}_{ij}} 2\varphi(\boldsymbol{X}_{ijk_1}) - \boldsymbol{Q}_{ij}$ has asymptotically multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}$. Since H_{ijk}'s are independent, then H has asymptotically multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}$.

1. If $n_{ij} = n$ for all i and j, then it is easy to verify that $\frac{1}{(\frac{n}{2}-1)}\boldsymbol{B}_5$ is a symmetric and idempotent matrix with rank $\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} - IJ = N - IJ = IJ\,(n-1)$. Thus, $\frac{1}{(\frac{n}{2}-1)}\boldsymbol{S}_W^0$ is asymptotically distributed as a W_p with $IJ(n-1)$ degrees of freedom. By using Lemma (3.10), the within sum of squares, $\boldsymbol{SSW}$, is asymptotically distributed as a W_p with $IJ(n-1)$ degrees of freedom and scale matrix $\boldsymbol{I}$.
2. If $n_{ij} \neq n_{i^T j^T}$ for some $i \neq i^T$ or $j \neq j^T$, $\mathbf{B}_5$ is symmetric but not idempotent. So, there exists $\mathbf{H}^{\mathrm{T}}\mathbf{B}_2\mathbf{H} = \sum_{i=1}^{d_2} \pi_i^{(2)} z^T z_i$, where $\pi_1^{(2)}, \pi_2^{(2)}, \ldots, \pi_{d_2}^{(2)}$ are the eigenvalues of B_5 and $z_i \sim N(\mathbf{0}, \boldsymbol{I})$ which are independent. Let $c_0^{(2)} = \sum_{i=1}^{d_2} \pi_i^{(2)}/d_2$, then using (Yuan & Bentler, 2010), $\boldsymbol{S}_W^0/c_0 = \mathbf{H}^{\mathrm{T}}\mathbf{B}_2\mathbf{H}/c_0^{(2)} \;\dot{\sim}\; W_p(d_2)$. Thus, by using Lemma (3.10), the within sum of squares, $\boldsymbol{SSW}$, is asymptotically distributed as W_p with d_2 degrees of freedom and scale matrix $\boldsymbol{I}$.

Define the Λ-test statistics of the kernel-based nonparametric test for the row effect in two-way MANOVA as

$$\Lambda_{\mathrm{k^R}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSR}+\boldsymbol{SSW}|} = \frac{1}{1+\left|\frac{\boldsymbol{SSR}}{\boldsymbol{SSW}}\right|}$$

$$= \frac{1}{1+\left|\frac{\sum_{i=1}^{I}\hat{m}_{i.}(\hat{\boldsymbol{Q}}_{i.}-\hat{\boldsymbol{Q}}_{..})(\hat{\boldsymbol{Q}}_{i.}-\hat{\boldsymbol{Q}}_{..})^T}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum\sum_{k_1\neq k_2}\hat{\boldsymbol{\omega}}_{ij}^2(\boldsymbol{A}_{ijk_1k_2}-\hat{\boldsymbol{Q}}_{ij})(\boldsymbol{A}_{ijk_1k_2}-\hat{\boldsymbol{Q}}_{ij})^T}\right|}. \tag{3.84}$$

Note that $\boldsymbol{SSR} \mathrel{\dot\sim} W_p(I-1,\boldsymbol{I})$ and $\boldsymbol{SSW} \mathrel{\dot\sim} W_p(df_{w_2},\boldsymbol{I})$, where df_{w_2} is given in Equation (3.78).

Theorem 3.12. *If for any $i = 1,2,\ldots,I,$ and $j = 1,2,\ldots,J$, $n_{ij}|\boldsymbol{H}_{ij}^4| \to 0$, $n_{ij}|\boldsymbol{H}_{ij}| \to \infty$ as $\min\limits_{i,j} n_{ij} \to \infty$, $\int_{-\infty}^{\infty}\boldsymbol{x}^2 f_{ij}^3(\boldsymbol{x})d\boldsymbol{x} < \infty$, then under the null hypothesis, $\Lambda_{\mathrm{k^R}}$ in Equation (3.84) has asymptotically F distribution with degrees of freedom df_1 and df_2.*

Proof. Theorem (3.9) shows that $\boldsymbol{SSR}$ follows asymptotically W_p with degrees of freedom $I-1$ and scale matrix $\boldsymbol{I}$ under null hypothesis and asymptotically non-central $W_p(I-1,\boldsymbol{I})$ under the alternative. Furthermore, Theorem (3.11) implies that the within sum of squares, $\boldsymbol{SSW}$, is asymptotically W_p with degrees of freedom $IJ(n-1)$ and scale matrix $\boldsymbol{I}$ for balanced data and W_p with degrees of freedom d_2 and scale matrix $\boldsymbol{I}$ for unbalanced data where d_2 is the number of eigenvalues of $\mathbf{B}_5$ in Equation (3.76). In order to show that $\Lambda_{\mathrm{k^R}}$ follows asymptotically F distribution under the null hypothesis and non-central F distribution under the alternative hypothesis, we need to show that $\boldsymbol{SSR}$ and $\boldsymbol{SSW}$ are asymptotically independent as $\min\limits_{i,j} n_{ij} \to \infty$. In Lemma (3.8), $\boldsymbol{S}_R^0$, which converges in probability to $\boldsymbol{SSR}$, is written in a quadratic form as $\boldsymbol{S}_R^0 = \mathbf{U}_2^{\mathrm{T}}\mathbf{B}_4\mathbf{U}_2$. Note that under the null hypothesis

$$\boldsymbol{T}_{ij}^{(2)} \simeq \sqrt{n_{ij}}\boldsymbol{\omega}_{ij}^{-1}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij}) = \frac{1}{\sqrt{n_{ij}}}\mathbf{H}_{\mathrm{ij}}^{\mathrm{T}} j_{n_{ij}}. \tag{3.85}$$

Thus, $\boldsymbol{S}_R^0$ can also be written as

$$\boldsymbol{S}_R^0 = \left(\frac{1}{\sqrt{n_{11}}}\mathbf{H}_{11}^{\mathrm{T}}\boldsymbol{j}_{n_{11}}, \ldots, \frac{1}{\sqrt{n_{IJ}}}\mathbf{H}_{IJ}^{\mathrm{T}}\boldsymbol{j}_{n_{IJ}}\right)\mathbf{B}_4 \begin{pmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{H}^{\mathrm{T}}{}_{11}\boldsymbol{j}_{n_{11}} \\ \vdots \\ \frac{1}{\sqrt{n_{IJ}}}\mathbf{H}_{IJ}^{\mathrm{T}}\boldsymbol{j}_{n_{IJ}} \end{pmatrix}$$

$$= \mathbf{H}^{\mathrm{T}} \begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\boldsymbol{j}_{n_{11}}^T & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\boldsymbol{j}_{n_{12}}^T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\sqrt{n_{IJ}}}\boldsymbol{j}_{n_{IJ}}^T \end{bmatrix}^T \mathbf{B}_4 \begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\boldsymbol{j}_{n_{11}}^T & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\boldsymbol{j}_{n_{12}}^T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\sqrt{n_{IJ}}}\boldsymbol{j}_{n_{IJ}}^T \end{bmatrix} \mathbf{H}$$

$$= \mathbf{H}^{\mathrm{T}}\mathbf{B}_6\mathbf{H}. \tag{3.86}$$

Recall from Theorem (3.11) that $\boldsymbol{S}_W^0 = \mathbf{H}^{\mathrm{T}}\mathbf{B}_5\mathbf{H}$. Now, it is a straightforward process to verify that

$$\mathbf{B}_5\mathbf{B}_6 = \mathbf{0} \times \mathbf{B}_4 \begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\boldsymbol{j}_{n_{11}}^T & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\boldsymbol{j}_{n_{12}}^T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\sqrt{n_{IJ}}}\boldsymbol{j}_{n_{IJ}}^T \end{bmatrix} = \mathbf{0}. \tag{3.87}$$

Thus, $\boldsymbol{S}_R^0$ and $\boldsymbol{S}_W^0$ are independent. By Lemma (3.8) and Lemma (3.10), ***SSR*** and ***SSW*** are asymptotically independent under null hypothesis $H_0 : \boldsymbol{\alpha}_i = \mathbf{0}$.

According to Mardia, Kent, and Bibby (1979) Λ_{kR} can be related to the F-distribution.

Let

$$a_R = df_{w_2} + \frac{(I-1)-p-1}{2}, \tag{3.88}$$

$$b_R = \begin{cases} \sqrt{\frac{p^2(I-1)^2-4}{p^2+(I-1)^2-5}}\ , \ p^2+(I-1)^2-5>0 \\ 1, \ \text{otherwise} \end{cases}, \tag{3.89}$$

and

$$c_R = \frac{p(I-1)}{2} - 1. \tag{3.90}$$

Now, let $df_1 = p(I-1)$ and $df_2 = a_R b_R - c_R$. Thus, the F approximation is

$$F = \frac{1-(\Lambda_{\mathrm{kR}})^{1/b_R}}{(\Lambda_{\mathrm{kR}})^{1/b_R}} \times \frac{df_2}{df_1}. \tag{3.91}$$

Therefore, under the null hypothesis, Λ_{kR} in Equation (3.84) follows asymptotically F distribution with df_1 and df_2 degrees of freedom.

Testing the column effect, $H_0 : \boldsymbol{\beta}_j = \mathbf{0}$, can be done in a similar manner to testing the row effect showed above. First, define the Column Sum of Squares (***SSC***) as

$$\boldsymbol{SSC} = \sum_{j=1}^{J} \hat{m}_{.j}(\hat{\boldsymbol{Q}}_{.j} - \hat{\boldsymbol{Q}}_{..})(\hat{\boldsymbol{Q}}_{.j} - \hat{\boldsymbol{Q}}_{..})^T, \tag{3.92}$$

where $\hat{\boldsymbol{Q}}_{.j} = \sum_i \frac{\hat{m}_{ij}}{\hat{m}_{.j}} \hat{\boldsymbol{Q}}_{ij}$ and $\hat{m}_{.j} = \sum_i \hat{m}_{ij}$. Then, the Λ-test statistics of kernel based nonparametric test for the column effect in two-way MANOVA as

$$\Lambda_{\mathrm{kC}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSC}+\boldsymbol{SSW}|} = \frac{1}{1+\left|\frac{\boldsymbol{SSC}}{\boldsymbol{SSW}}\right|}$$

$$= \frac{1}{1+\left|\frac{\sum_{i=1}^{I}\hat{m}_{i.}(\hat{\boldsymbol{Q}}_{i.}-\hat{\boldsymbol{Q}}_{..})(\hat{\boldsymbol{Q}}_{i.}-\hat{\boldsymbol{Q}}_{..})^T}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum\sum_{k_1\neq k_2}\hat{\boldsymbol{\omega}}_{ij}^{-2}(\boldsymbol{A}_{ijk_1k_2}-\hat{\boldsymbol{Q}}_{ij})(\boldsymbol{A}_{ijk_1k_2}-\hat{\boldsymbol{Q}}_{ij})^T}\right|}. \tag{3.93}$$

Note that $\boldsymbol{SSC} \dot{\sim} W_p(J-1,\boldsymbol{I})$ and $\boldsymbol{SSW} \dot{\sim} W_p(df_{w_2},\boldsymbol{I})$, where df_{w_2} is given in Equation (3.78).

Theorem 3.13. *If for any $i = 1,2,\ldots,I$, and $j = 1,2,\ldots,J$, $n_{ij}|\boldsymbol{H}_{ij}^4| \to 0$, $n_{ij}|\boldsymbol{H}_{ij}| \to \infty$ as $\min_{i,j} n_{ij} \to \infty$, $\int_{-\infty}^{\infty}\boldsymbol{x}^2 f_{ij}^3(\boldsymbol{x})d\boldsymbol{x} < \infty$, then under the null hypothesis, Λ_{kC} in Equation (3.93) has asymptotically F distribution with df_1 and df_2 degrees of freedom.*

Proof. This can be proven in a similar manner as in Theorem (3.12) with replacing I's to J's.

Interaction Effect

In order to test the interaction effect of the two-way MANOVA, $\boldsymbol{\gamma}_{ij} = \mathbf{0}$ for all i and j, define the Interaction Sum of Squares ($\boldsymbol{SSI}$) as

$$\boldsymbol{SSI} = \sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}(\boldsymbol{Q}_{ij} - \bar{\boldsymbol{Q}}_{i.} - \bar{\boldsymbol{Q}}_{.j} + \bar{\boldsymbol{Q}}_{..})(\boldsymbol{Q}_{ij} - \bar{\boldsymbol{Q}}_{i.} - \bar{\boldsymbol{Q}}_{.j} + \bar{\boldsymbol{Q}}_{..})^{T}. \tag{3.94}$$

To obtain the asymptotic distribution of $\boldsymbol{SSI}$, additional auxiliary variable needs to be defined as:

$$\boldsymbol{S}_{I}^{0} = \sum_{i=1}^{I}\sum_{j=1}^{J} \hat{m}_{ij}(\boldsymbol{Q}_{ij} - \bar{\boldsymbol{Q}}_{i.} - \bar{\boldsymbol{Q}}_{.j} + \bar{\boldsymbol{Q}}_{..})(\boldsymbol{Q}_{ij} - \bar{\boldsymbol{Q}}_{i.} - \bar{\boldsymbol{Q}}_{.j} + \bar{\boldsymbol{Q}}_{..})^{T}, \tag{3.95}$$

where $\boldsymbol{Q}^{*}_{.j} = \sum_i \frac{m_{ij}}{m_{.j}} \hat{\boldsymbol{Q}}_{ij}$.

Lemma 3.14. *Let* $N = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}$. *If* $\lambda_{ij} = \lim_{\min\, n_{ij}\to\infty} \frac{n_{ij}}{N}$ *and* $\hat{\boldsymbol{\omega}}_{ij}^{2} \xrightarrow{p} \boldsymbol{\omega}_{ij}^{2}$, *then* $\boldsymbol{S}_{I}^{0} - \boldsymbol{SSI} \xrightarrow{P} 0$, as $\min_{ij}\, n_{ij} \to \infty$.

Proof. This can be proven by applying Slutsky Theorem (DasGupta, 2008).

Theorem 3.15. *Under the null hypothesis, if for any* $i = 1,2,\ldots,I$, *and* $j = 1,2,\ldots,J$, $n_{ij}|\boldsymbol{H}_{ij}^{4}| \to 0$, $n_{ij}|\boldsymbol{H}_{ij}| \to \infty$ *as* $\min_{i,j}\, n_{ij} \to \infty$, *and if* $\int_{-\infty}^{\infty} x^2 f_{ij}^{3}(x)dx < \infty$, *then* $\boldsymbol{SSI}$ *is asymptotically* $W_p((I-1)(J-1))$.

Proof. Set $\boldsymbol{T}_{ij}^{(2)} = \sqrt{\lambda_{ij}N}\boldsymbol{\omega}_{ij}^{-1}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})$, then under H_0, $\boldsymbol{T}_{ij}^{(2)} \sim N(\mathbf{0},\boldsymbol{I})$ as $N \to \infty$ by Lemma (3.1). Note that under the null hypothesis,

$$
\begin{aligned}
\boldsymbol{S}_I^0 &= \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{i.}^{*} - \boldsymbol{Q}_{.j}^{*} + \boldsymbol{Q}_{..}^{*})^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}((\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij}) - (\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.}) - (\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j}) + (\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..}))^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}((\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})^2 - (\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})^2 - (\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j})^2 + (\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..})^2) \\
&- \ 2(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.}) - 2(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j}) + 2(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..}) \\
&+ \ 2(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j}) - 2(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..}) - 2(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j})(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..}) \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})^2 + \sum_{i=1}^{I} m_{i.}(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})^2 + \sum_{j=1}^{J} m_{.j}(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j})^2 + m_{..}(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..})^2 \\
&- \ 2\sum_{i=1}^{I} m_{i.}(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})^2 - 2\sum_{j=1}^{J} m_{.j}(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j})^2 + 2m_{..}(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..})^2 \\
&+ \ 2\sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j}) - 2m_{..}(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..})^2 - 2m_{..}(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..})^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij} - \boldsymbol{Q}_{ij})^2 - \sum_{i=1}^{I} m_{i.}(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})^2 - \sum_{j=1}^{J} m_{.j}(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j})^2 + m_{..}(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..})^2 \\
&+ \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j}) - 2m_{..}(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..})^2. \qquad (3.96)
\end{aligned}
$$

Let assume that $m_{ij} = \frac{1}{m_{..}}(m_{i.})(m_{.j})$, then

$$
\sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\boldsymbol{Q}_{i.}^{*} - \bar{\boldsymbol{Q}}_{i.})(\boldsymbol{Q}_{.j}^{*} - \bar{\boldsymbol{Q}}_{.j})^T = 2m_{..}(\boldsymbol{Q}_{..}^{*} - \bar{\boldsymbol{Q}}_{..})^2. \qquad (3.97)
$$

Thus,

$$
\begin{aligned}
\boldsymbol{S}_I^0 &= \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})^2 - \sum_{i=1}^{I} m_{i.}(\boldsymbol{Q}^*_{i.}-\bar{\boldsymbol{Q}}_{i.})^2 - \sum_{j=1}^{J} m_{.j}(\boldsymbol{Q}^*_{.j}-\bar{\boldsymbol{Q}}_{.j}{}^2 + m_{..}(\boldsymbol{Q}^*_{..}-\bar{\boldsymbol{Q}}_{..})^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})^2 - \sum_{i=1}^{I}\frac{1}{m_{i.}}\left(\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})\right)^2 - \sum_{j=1}^{J}\frac{1}{m_{.j}}\left(\sum_{i=1}^{I} m_{ij}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})\right)^2 \\
&+ \frac{1}{m_{..}}\left(\sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})\right)^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J} N\lambda_{ij}\boldsymbol{\omega}_{ij}^{-2}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})^2 \\
&- \sum_{i=1}^{I}\sum_{j_1=1}^{J}\sum_{j_2=1}^{J}\frac{N}{m_{i.}}\frac{\sqrt{N\lambda_{ij_1}}}{\boldsymbol{\omega}_{ij_1}}(\hat{\boldsymbol{Q}}_{ij_1}-\boldsymbol{Q}_{ij_1})\frac{\sqrt{N\lambda_{ij_2}}}{\boldsymbol{\omega}_{ij_2}}(\hat{\boldsymbol{Q}}_{ij_2}-\boldsymbol{Q}_{ij_2})\frac{\sqrt{\lambda_{ij_1}}}{\boldsymbol{\omega}_{ij_1}}\frac{\sqrt{\lambda_{ij_2}}}{\boldsymbol{\omega}_{ij_2}} \\
&- \sum_{i_1=1}^{I}\sum_{i_2=1}^{I}\sum_{j_2=1}^{J}\frac{N}{m_{.j}}\frac{\sqrt{N\lambda_{i_1j}}}{\boldsymbol{\omega}_{i_1j}}(\hat{\boldsymbol{Q}}_{i_1j}-\boldsymbol{Q}_{i_1j})\frac{\sqrt{N\lambda_{i_2j}}}{\boldsymbol{\omega}_{i_2j}}(\hat{\boldsymbol{Q}}_{i_2j}-\boldsymbol{Q}_{i_2j})\frac{\sqrt{\lambda_{i_1j}}}{\boldsymbol{\omega}_{i_1j}}\frac{\sqrt{\lambda_{i_2j}}}{\boldsymbol{\omega}_{i_2j}} \\
&+ \frac{N}{m_{..}}\sum_{i_1=1}^{I}\sum_{i_2=1}^{I}\sum_{j_1=1}^{J}\sum_{j_2=1}^{J}\frac{\sqrt{N\lambda_{i_1j_1}}}{\boldsymbol{\omega}_{i_1j_1}}(\hat{\boldsymbol{Q}}_{i_1j_1}-\boldsymbol{Q}_{i_1j_1})\frac{\sqrt{N\lambda_{i_2j_2}}}{\boldsymbol{\omega}_{i_2j_2}}(\hat{\boldsymbol{Q}}_{i_2j_2}-\boldsymbol{Q}_{i_2j_2})\frac{\sqrt{\lambda_{i_1j_1}}}{\boldsymbol{\omega}_{i_1j_1}}\frac{\sqrt{\lambda_{i_2j_2}}}{\boldsymbol{\omega}_{i_2j_2}},
\end{aligned}
\quad (3.98)
$$

which can be written in two quadratic forms. The second and the forth terms in (3.98) can be written in the form $\mathbf{U}_2^{\mathrm{T}}\mathbf{M}^{(1)}\mathbf{U}_2$ and $\mathbf{U}_2^{\mathrm{T}}\mathbf{M}^{(2)}\mathbf{U}_2$ and the third term can be written in the form $\mathbf{U}_2^{\mathrm{T}}\mathbf{N}^{(1)}\mathbf{U}_2$, where $\mathbf{U_2} = (\mathbf{T}_{11}^{(2)}, \mathbf{T}_{12}^{(2)}, \ldots, \mathbf{T}_{IJ}^{(2)})^T$ and

$$
\mathbf{N}^{(1)} = \begin{bmatrix} \boldsymbol{N}_{11} & \boldsymbol{N}_{12} & \ldots & \boldsymbol{N}_{1\mathrm{I}} \\ \boldsymbol{N}_{21} & \boldsymbol{N}_{22} & \ldots & \boldsymbol{N}_{2\mathrm{I}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{N}_{\mathrm{I}1} & \boldsymbol{N}_{\mathrm{I}2} & \ldots & \boldsymbol{N}_{\mathrm{II}} \end{bmatrix}, \quad (3.99)
$$

and

$$
\mathbf{N}_{\mathrm{ij}} = \begin{bmatrix} \frac{N}{m_{.1}}\sqrt{\lambda_{i1}}\sqrt{\lambda_{j1}}\boldsymbol{\omega}_{i1}^{-1}\boldsymbol{\omega}_{j1}^{-1} & 0 & \ldots & 0 \\ 0 & \frac{N}{m_{.2}}\sqrt{\lambda_{i2}}\sqrt{\lambda_{j2}}\boldsymbol{\omega}_{i2}^{-1}\boldsymbol{\omega}_{j2}^{-1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{N}{m_{.J}}\sqrt{\lambda_{iJ}}\sqrt{\lambda_{jJ}}\boldsymbol{\omega}_{iJ}^{-1}\boldsymbol{\omega}_{jJ}^{-1} \end{bmatrix}, \quad (3.100)
$$

for all $i = 1,2,\ldots,I,\ j = 1,2,\ldots,J$. From this, we can conclude that $\boldsymbol{S}_I^0$ can be written in a quadratic form such as

$$\begin{aligned}\boldsymbol{S}_I^0 &= \mathbf{U}_2^{\mathrm{T}}\mathbf{U}_2 - \mathbf{U}_2^{\mathrm{T}}\mathbf{M}^{(1)}\mathbf{U}_2 - \mathbf{U}_2^{\mathrm{T}}\mathbf{N}^{(1)}\mathbf{U}_2 + \mathbf{U}_2^{\mathrm{T}}\mathbf{M}^{(2)}\mathbf{U}_2 \\ &= \mathbf{U}_2^{\mathrm{T}}(\mathbf{I}_{\mathrm{IJ}} - \mathbf{M}^{(1)} - \mathbf{N}^{(1)} + \mathbf{M}^{(2)})\mathbf{U}_2 \\ &= \mathbf{U}_2^{\mathrm{T}}\mathbf{B}_7\mathbf{U}_2. \end{aligned} \tag{3.101}$$

where

$$\mathbf{B}_7 = \begin{bmatrix} \mathrm{I_J} - (\frac{N}{m_{1.}} - \frac{N}{m_{..}})\mathbf{M}_{11} - \mathbf{N}_{11} & -\mathbf{N}_{12} + \frac{N}{m_{..}}\mathbf{M}_{12} & \cdots & -\mathbf{N}_{1\mathrm{I}} + \frac{N}{m_{..}}\mathbf{M}_{1\mathrm{I}} \\ -\mathbf{N}_{12} + \frac{N}{m_{..}}\mathbf{M}_{12} & \mathrm{I_J} - (\frac{N}{m_{2.}} - \frac{N}{m_{..}})\mathbf{M}_{22} - \mathbf{N}_{22} & \cdots & -\mathbf{N}_{2\mathrm{I}} + \frac{N}{m_{..}}\mathbf{M}_{2\mathrm{I}} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{N}_{1\mathrm{I}} + \frac{N}{m_{..}}\mathbf{M}_{1\mathrm{J}} & -\mathbf{N}_{11} + \frac{N}{m_{..}}\mathbf{M}_{2\mathrm{I}} & \cdots & \mathrm{I_J} - (\frac{N}{m_{I.}} - \frac{N}{m_{..}})\mathbf{M}_{\mathrm{JJ}} - \mathbf{N}_{\mathrm{JJ}} \end{bmatrix}. \tag{3.102}$$

Since $\mathbf{B}_7$ is symmetric and idempotent, we obtain

$$\begin{aligned} rank(\mathbf{B}_7) &= tr(\mathbf{B}_7) \\ &= IJ - \sum_{i=1}^{I}\frac{1}{m_{i.}}\sum_{j=1}^{J} m_{ij} - \sum_{j=1}^{J}\frac{1}{m_{.j}}\sum_{i=1}^{I} m_{ij} + \frac{1}{m_{..}}\sum_{i=1}^{I}\sum_{j=1}^{J} m_{ij} \\ &= IJ - I - J + 1 \\ &= (I-1)(J-1). \end{aligned} \tag{3.103}$$

$\mathbf{U}_2$ approximately follows a multivariate normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{I}$, since $\boldsymbol{T}_{ij}^{(2)}$'s independently follow a multivariate standard normal distribution. Therefore, $\boldsymbol{S}_I^0$ is asymptotically W_p with degrees of freedom $(I-1)(J-1)$ and scale matrix $\boldsymbol{I}$ under H_0. Hence, by Lemma (3.14), ***SSI*** asymptotically W_p with degrees of freedom $(I-1)(J-1)$ and scale matrix $\boldsymbol{I}$.

Now, define the Λ-test statistics of kernel based nonparametric test for the interaction effect in two-way MANOVA as:

$$\Lambda_{\mathrm{kI}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSI}+\boldsymbol{SSW}|} = \frac{1}{1+\left|\frac{\boldsymbol{SSI}}{\boldsymbol{SSW}}\right|}$$

$$= \frac{1}{1+\left|\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\hat{m}_{ij}(\boldsymbol{Q}_{ij}-\bar{\boldsymbol{Q}}_{i.}-\bar{\boldsymbol{Q}}_{.j}+\bar{\boldsymbol{Q}}_{..})(\boldsymbol{Q}_{ij}-\bar{\boldsymbol{Q}}_{i.}-\bar{\boldsymbol{Q}}_{.j}+\bar{\boldsymbol{Q}}_{..})^{T}}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum\sum_{k_1\neq k_2}\hat{\boldsymbol{\omega}}_{ij}^{2}(\boldsymbol{A}_{ijk_1k_2}-\hat{\boldsymbol{Q}}_{ij})(\boldsymbol{A}_{ijk_1k_2}-\hat{\boldsymbol{Q}}_{ij})^{T}}\right|}. \quad (3.104)$$

Note that $\boldsymbol{SSI} \dot{\sim} W_p((I-1)(J-1),\boldsymbol{I})$ and $\boldsymbol{SSW} \dot{\sim} W_p(df_{w_2},\boldsymbol{I})$ where df_{w_2} is given in Equation (3.78).

Theorem 3.16. *If for any $i = 1,2,\dots,I$, and $j = 1,2,\dots,J$, $n_{ij}|\boldsymbol{H}_{ij}^{4}| \to 0$, $n_{ij}|\boldsymbol{H}_{ij}| \to \infty$ as $\min_{i,j} n_{ij} \to \infty$, $\int_{-\infty}^{\infty}\boldsymbol{x}^2 f_{ij}^{3}(\boldsymbol{x})d\boldsymbol{x} < \infty$, then under the null hypothesis, Λ_{kI} in Equation (3.104) has asymptotically F distribution with degrees of freedom df_1 and df_2.*

Proof. Theorem (3.15) shows that $\boldsymbol{SSI}$ follows asymptotically W_p with degrees of freedom $(I-1)(J-1)$ and $\boldsymbol{I}$ scale matrix under null hypothesis and asymptotically non-central $W_p((I-1)(J-1),\boldsymbol{I})$ under the alternative. Furthermore, Theorem (3.11) implies that the within sum of squares, $\boldsymbol{SSW}$, is asymptotically W_p with degrees of freedom $IJ(n-1)$ and scale matrix $\boldsymbol{I}$ for balanced data and W_p with degrees of freedom d_2 and scale matrix $\boldsymbol{I}$ for unbalanced data, where d_2 is the number of eigenvalues of $\mathbf{B}_5$ in Equation (3.76). In order to show that Kernel$-\Lambda_{\mathrm{I}}$ follows asymptotically F distribution under the null hypothesis and non-central F distribution under the alternative hypothesis, we need to show that $\boldsymbol{SSI}$ and $\boldsymbol{SSW}$ are asymptotically independent as $\min_{i,j} n_{ij} \to \infty$. In Lemma (3.14), $\boldsymbol{S}_I^0$, which converges in probability to $\boldsymbol{SSI}$, is written in a quadratic form as $\boldsymbol{S}_I^0 = \mathbf{U}_2^{\mathrm{T}}\mathbf{B}_7\mathbf{U}_2$. Note that under the null hypothesis

$$\boldsymbol{T}_{ij}^{(2)} \simeq \frac{\sqrt{n_{ij}}(\hat{\boldsymbol{Q}}_{ij}-\boldsymbol{Q}_{ij})}{\boldsymbol{\omega}_{ij}} = \frac{1}{\sqrt{n_{ij}}}\mathrm{H}_{\mathrm{ij}}^{\mathrm{T}}\boldsymbol{j}_{n_{ij}}. \quad (3.105)$$

Thus, $\boldsymbol{S}_I^0$ can also be written as

$$\boldsymbol{S}_I^0 = \left(\frac{1}{\sqrt{n_{11}}} \mathbf{H}_{11}^{\mathrm{T}} \boldsymbol{j}_{n_{11}}, \ \ldots, \frac{1}{\sqrt{n_{IJ}}} \mathbf{H}_{IJ}^{\mathrm{T}} \boldsymbol{j}_{n_{IJ}} \right) \mathbf{B}_7 \begin{pmatrix} \frac{1}{\sqrt{n_{11}}} \mathbf{H}_{11}^{\mathrm{T}} \boldsymbol{j}_{n_{11}} \\ \vdots \\ \frac{1}{\sqrt{n_{IJ}}} \mathbf{H}_{IJ}^{\mathrm{T}} \boldsymbol{j}_{n_{IJ}} \end{pmatrix}$$

$$= \mathbf{H}^{\mathrm{T}} \begin{bmatrix} \frac{1}{\sqrt{n_{11}}} \mathbf{j}_{n_{11}}^T & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}} \boldsymbol{j}_{n_{12}}^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{IJ}}} \boldsymbol{j}_{n_{IJ}}^T \end{bmatrix}^T \mathbf{B}_7 \begin{bmatrix} \frac{1}{\sqrt{n_{11}}} \boldsymbol{j}_{n_{11}}^T & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}} \boldsymbol{j}_{n_{12}}^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{IJ}}} \boldsymbol{j}_{n_{IJ}}^T \end{bmatrix} \mathbf{H}$$

$$= \mathbf{H}^{\mathrm{T}} \mathbf{B}_8 \mathbf{H}. \tag{3.106}$$

Recall from Theorem (3.11) that $\boldsymbol{S}_W^0 = \mathbf{H}^{\mathrm{T}} \mathbf{B}_5 \mathbf{H}$. Now, it is a straightforward process to verify that

$$\mathbf{B}_5 \mathbf{B}_8 = \mathbf{0} \times \mathbf{B}_7 \begin{bmatrix} \frac{1}{\sqrt{n_{11}}} \boldsymbol{j}_{n_{11}}^T & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}} \boldsymbol{j}_{n_{12}}^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{IJ}}} \boldsymbol{j}_{n_{IJ}}^T \end{bmatrix} = \mathbf{0}. \tag{3.107}$$

Thus, $\boldsymbol{S}_I^0$ and $\boldsymbol{S}_W^0$ are independent. By Lemma (3.14) and Lemma (3.10), $\boldsymbol{SSI}$ and $\boldsymbol{SSW}$ are asymptotically independent under null hypothesis $H_0 : \boldsymbol{\gamma}_{ij} = \mathbf{0}$.

According to Mardia, Kent, and Bibby (1979) Λ_{kI} can be related to the F-distribution.

Let

$$a_I = df_{w_2} + \frac{((I-1)(J-1)) - p - 1}{2}, \tag{3.108}$$

$$b_I = \begin{cases} \sqrt{\frac{p^2((I-1)(J-1))^2 - 4}{p^2 + ((I-1)(J-1))^2 - 5}}, & p^2 + ((I-1)(J-1))^2 - 5 > 0 \\ 1, \text{otherwise} \end{cases}, \tag{3.109}$$

and

$$c_I = \frac{p((I-1)(J-1))}{2} - 1. \tag{3.110}$$

Now, let $df_1 = p((I-1)(J-1))$ and $df_2 = a_I b_I - c_I$. Thus, the F approximation is

$$F = \frac{1-(\Lambda_{\mathrm{kI}})^{1/b_I}}{(\Lambda_{\mathrm{kI}})^{1/b_I}} \times \frac{df_2}{df_1}. \tag{3.111}$$

Therefore, under null hypothesis, Λ_{kI} in Equation (3.103) follows asymptotically F distribution with df_1 and df_2 degrees of freedom.

Method Evaluation

The asymptotic behavior of the proposed nonparametric kernel-based tests was first studied theoretically as shown in this chapter by providing the proofs constructed. To evaluate the performance of the proposed nonparametric one-way kernel-based MANOVA test, Type I error rate and power are obtained then compared to the corresponding parametric one-way MANOVA test. To evaluate the performance of the interaction effect test in the proposed nonparametric two-way kernel-based MANOVA test, Type I error rate and power are obtained then compared to the corresponding interaction effect test in the parametric two-way MANOVA test. To estimate Type I error rate and power for the proposed nonparametric kernel-based MANOVA test for both one and two-way layout, a simulation study using small sample sizes was conducted. Additionally, a real dataset was used as an exemplar dataset. The developed nonparametric technique for comparing groups within a multivariate setting is practical because it allows for the adoption of the technique by applied researchers and practitioners.

Simulation Study

The simulation for this study was conducted using the 3.5.5 version of R program for statistical computing (R Core Team, 2019). Various values of p (the dimension) and n (the sample size) were used. The dimensions of the datasets used in simulation are $p = 2$, $p = 4$, and $p = 6$, representing common multivariate data dimensions, often seen in practice. Sample sizes of datasets used in simulation are $n = 10$, $n = 30$, $n = 50$, $n = 70$, and $n = 100$, representing multiple ranges of samples that may be seen in practice. Data used in the simulation study were randomly generated from multivariate normal, multivariate Cauchy, and multivariate exponential distribution for both the proposed nonparametric kernel-based test and their corresponding parametric tests.

Simulation Conditions for Evaluating Type I Error and Power for the Nonparametric Kernel-Based One-Way MANOVA

In this section, the scheme of parameters used in the simulation to evaluate the performance of the proposed nonparametric kernel-based one-way MANOVA test, and compare it to the traditional parametric one-way MANOVA test is shown. For each dimension, the Type I error rate and power of the proposed nonparametric and parametric MANOVA tests were evaluated using three different distributions each with three cases showed in Tables 3.1 - 3.6. To evaluate the actual Type I error rate and the power of the parametric one-way MANOVA test and nonparametric kernel-based one-way MAOVA test proposed in this chapter, the equality of the location parameters, $\boldsymbol{\mu}_i$, of three groups, i.e. $I = 3$ for each case is tested. To obtain the actual Type I error at the significant level $\alpha = 0.05$, we follow the steps below for $p = 2, 4, 6$:

(1) Randomly generate I groups of data with balanced sample size n from multivariate normal, multivariate Cauchy, and multivariate exponential distribution with rate $\mathbf{1}_p$ as described in Table 3.1, Table 3.3, and Table 3.5, for $p = 2, 4, 6$, respectively.

(2) Apply the traditional parametric Wilks test and the proposed nonparametric test separately. Keep the test result as 1 or 0 —1 when we “reject the null hypothesis” and 0 when we “fail to reject.”

(3) Repeat (1) and (2) 10,000 times and count the percentage of rejections.

(4) Repeat (1) -(3) for sample size $n = 10, 30, 50, 70, 100$.

(5) Repeat (1) -(4) for dimension $p = 2, 4, 6$.

The procedure to estimate the empirical power is similar to the one used to estimate Type I error rate except we generate three groups of data from the three distributions with different means for each case in Step (1). The distribution types and parameters assigned to each group in each case are listed in Table 3.2, Table 3.4, and Table 3.6.

For the case when we have two dependent variables, $p = 2$, under the null hypothesis, the conditions for calculating Type I error are shown in Table 3.1, where

$\mathbf{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ represents the case of no correlation between dependent variables,

$\mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$ represents the case of low correlation between dependent variables, and

$\mathbf{\Sigma}_3 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$ represent the case of low correlation between dependent variables.

Table 3.1.

Conditions for Calculating Type I Error for $p = 2$

Distribuion	Case	Group 1	Group 2	Group 3
	Case I	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_1)$
Multivariate Normal Distribution	Case II	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_2)$
	Case III	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_N, \mathbf{\Sigma}_3)$
	Case I	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_1)$
Multivariate Cauchy Distribution	Case II	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_2)$
	Case III	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_C, \mathbf{\Sigma}_3)$
	Case I	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_1)$
Multivariate Exponential Distribution	Case II	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_2)$
	Case III	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_E, \mathbf{\Sigma}_3)$

where $\boldsymbol{\mu}_N = [1 \quad 2]^T$, $\boldsymbol{\mu}_C = [5 \quad 6]^T$, $\boldsymbol{\mu}_E = [4 \quad 3]^T$.

Under the alternative hypothesis, the conditions for calculating power showed in Table 3.2, the same variance-covariance matrices used for calculating Type I error are used with changes in each mean vector.

Table 3.2.

Condition for Calculating Power for $p = 2$

Distribuion	Case	Group 1	Group 2	Group 3
Multivariate Normal Distribution	Case I	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_3)$
Multivariate Cauchy Distribution	Case I	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_3)$
Multivariate Exponential Distribution	Case I	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_{N^1} = [0.5 \quad 1.5]^T$, $\boldsymbol{\mu}_{N^2} = [1 \quad 2]^T$, $\boldsymbol{\mu}_{N^3} = [1.5 \quad 2.5]^T$,$\boldsymbol{\mu}_{C^1} = [4.5 \quad 5.5]^T$, $\boldsymbol{\mu}_{C^2} = [5 \quad 6]^T$, $\boldsymbol{\mu}_{C^3} = [5.5 \quad 6.5]^T$, $\boldsymbol{\mu}_{E^1} = [3.5 \quad 2.5]^T$, $\boldsymbol{\mu}_{E^2} = [4 \quad 3]^T$, and $\boldsymbol{\mu}_{E^3} = [4.5 \quad 3.5]^T$.

When we have four dependent variables, $p = 4$, under the null hypothesis, the conditions for calculating Type I error are shown in Table 3.3, where

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.2 & 0.3 & 0.1 \\ 0.2 & 1 & 0.1 & 0.2 \\ 0.3 & 0.1 & 1 & 0.3 \\ 0.1 & 0.2 & 0.3 & 1 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & 0.7 & 0.9 & 0.8 \\ 0.7 & 1 & 0.8 & 0.7 \\ 0.9 & 0.8 & 1 & 0.9 \\ 0.8 & 0.7 & 0.9 & 1 \end{bmatrix},$$

representing the case of no correlation, low correlation, and high correlation among dependent variables, respectively.

Table 3.3.

Condition for Calculating Type I Error for $p = 4$

Distribuion	Case	Group 1	Group 2	Group 3
Multivariate Normal Distribution	Case I	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_1)$
	Case II	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_2)$
	Case III	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_3)$
Multivariate Cauchy Distribution	Case I	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_1)$
	Case II	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_2)$
	Case III	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_3)$
Multivariate Exponential Distribution	Case I	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_1)$
	Case II	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_2)$
	Case III	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_N = [1 \quad 2 \quad 3 \quad 4]^T$, $\boldsymbol{\mu}_C = [4 \quad 5 \quad 6 \quad 7]^T$, and $\boldsymbol{\mu}_E = [2 \quad 4 \quad 3 \quad 5]^T$.

Under the alternative hypothesis, the conditions for calculating power are shown in Table 3.4, and the same variance-covariance matrices used for calculating Type I error are used with changes in each mean vector.

Table 3.4.

Condition for Calculating Power for $p = 4$

Distribuion	Case	Group 1	Group 2	Group 3
Multivariate Normal Distribution	Case I	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_3)$
Multivariate Cauchy Distribution	Case I	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_3)$
Multivariate Exponential Distribution	Case I	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_{N^1} = [0.5 \quad 1.5 \quad 2.5 \quad 3.5]^T$, $\boldsymbol{\mu}_{N^2} = [1 \quad 2 \quad 3 \quad 4]^T$, $\boldsymbol{\mu}_{N^3} = [1.5 \quad 2.5 \quad 3.5 \quad 4.5]^T$,

$\boldsymbol{\mu}_{C^1} = [3.5 \quad 4.5 \quad 5.5 \quad 6.5]^T$, $\boldsymbol{\mu}_{C^2} = [4 \quad 5 \quad 6 \quad 7]^T$, $\boldsymbol{\mu}_{C^3} = [4.5 \quad 5.5 \quad 6.5 \quad 7.5]^T$,
$\boldsymbol{\mu}_{E^1} = [1.5 \quad 3.5 \quad 2.5 \quad 4.5]^T$, $\boldsymbol{\mu}_{E^2} = [2 \quad 4 \quad 3 \quad 5]^T$, and $\boldsymbol{\mu}_{E^3} = [2.5 \quad 4.5 \quad 3.5 \quad 5.5]^T$.

When we have six dependent variables, $p = 6$, under the null hypothesis, the conditions for calculating Type I error are shown in Table 3.5, where

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 \\ 0.1 & 1 & 0.1 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.1 & 1 & 0.1 & 0.2 & 0.3 \\ 0.3 & 0.2 & 0.1 & 1 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.2 & 0.1 & 1 & 0.3 \\ 0.1 & 0.1 & 0.1 & 0.2 & 0.3 & 1 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & 0.7 & 0.8 & 0.9 & 0.8 & 0.7 \\ 0.7 & 1 & 0.7 & 0.8 & 0.9 & 0.7 \\ 0.8 & 0.7 & 1 & 0.8 & 0.9 & 0.7 \\ 0.9 & 0.8 & 0.8 & 1 & 0.8 & 0.9 \\ 0.8 & 0.9 & 0.9 & 0.8 & 1 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.9 & 0.7 & 1 \end{bmatrix},$$

representing the case of no correlation, low correlation, and high correlation among dependent variables, respectively.

Table 3.5.

Condition for Calculating Type I Error for $p = 6$

Distribuion	Case	Group 1	Group 2	Group 3
Multivariate Normal Distribution	Case I	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_1)$
	Case II	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_2)$
	Case III	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_3)$
Multivariate Cauchy Distribution	Case I	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_1)$
	Case II	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_2)$
	Case III	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_3)$
Multivariate Exponential Distribution	Case I	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_1)$
	Case II	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_2)$
	Case III	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_E, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_N = [1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6]^T$, $\boldsymbol{\mu}_C = [3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8]^T$, $\boldsymbol{\mu}_E = [4 \quad 3 \quad 5 \quad 6 \quad 4 \quad 5]^T$.

Under the alternative hypothesis, the conditions for calculating power showed in Table 3.6, the same variance-covariance matrices used for calculating Type I error are used with changes in each mean vector.

Table 3.6.

Conditions for Calculating Power for $p = 6$

Distribuion	Case	Group 1	Group 2	Group 3
Multivariate Normal Distribution	Case I	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_1)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_2)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVN(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_3)$	$MVN(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_3)$
Multivariate Cauchy Distribution	Case I	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_1)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_2)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVC(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_3)$	$MVC(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_3)$
Multivariate Exponential Distribution	Case I	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_1)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_1)$
	Case II	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_2)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_2)$
	Case III	$MVEXP(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_3)$	$MVEXP(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_{N^1} = [0.5 \quad 1.5 \quad 2.5 \quad 3.5 \quad 4.5 \quad 5.5]^T$, $\boldsymbol{\mu}_{N^2} = [1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6]^T$,
$\boldsymbol{\mu}_{N^3} = [1.5 \quad 2.5 \quad 3.5 \quad 4.5 \quad 5.5 \quad 6.5]^T$, $\boldsymbol{\mu}_{C^1} = [2.5 \quad 3.5 \quad 4.5 \quad 5.5 \quad 6.5 \quad 7.5]^T$,
$\boldsymbol{\mu}_{C^2} = [3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8]^T$, $\boldsymbol{\mu}_{C^3} = [3.5 \quad 4.5 \quad 5.5 \quad 6.5 \quad 7.5 \quad 8.5]^T$,
$\boldsymbol{\mu}_{E^1} = [3.5 \quad 2.5 \quad 4.5 \quad 5.5 \quad 3.5 \quad 4.5]^T$, $\boldsymbol{\mu}_{E^2} = [4 \quad 3 \quad 5 \quad 6 \quad 4 \quad 5]^T$, and
$\boldsymbol{\mu}_{E^3} = [4.5 \quad 3.5 \quad 5.5 \quad 6.5 \quad 4.5 \quad 5.5]^T$.

Simulation Conditions for Evaluating Type I Error and Power for the Interaction Effect in the Nonparametric Kernel -Based Two-Way MANOVA

In this section, the scheme of parameters used in the simulation to evaluate the performance of the interaction effect in the proposed nonparametric kernel-based two-way MANOVA test, and compare it to the interaction effect in the traditional parametric two-way MANOVA test is shown. For each dimension, the Type I error rate and power of the proposed nonparametric and parametric MANOVA tests was evaluated using three different distributions: Normal, Cauchy, and Exponential with rate $\mathbf{1}_p$ showed in Tables 3.7 - 3.12. Consider a study with two factors, Factor A (Row) and Factor B (Column);

each factor has three levels, i.e. $r = 3$ and $c = 3$. Consider the multivariate two-way layout, $\boldsymbol{\mu}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij}$. Let the overall location $\boldsymbol{\mu} = \mathbf{3}_p$, row effect $\boldsymbol{C}_1^T \boldsymbol{\alpha}_i = \boldsymbol{C}_1^T(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$, and column effect $\boldsymbol{C}_2^T \boldsymbol{\beta}_j = \boldsymbol{C}_2^T(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$, where $\boldsymbol{C}_1 = (-1, 0, 1)$ and $\boldsymbol{C}_2 = (-1, 0, 1)$.

To evaluate the actual Type I error rate and the power of the interaction effect in the parametric two-way MANOVA test and the interaction effect in the nonparametric kernel-based two-way MANOVA test proposed in this chapter, the equality of the interaction effect, $\boldsymbol{\gamma}_{ij}$ is tested. To obtain the actual Type I error at the significant level $\alpha = 0.05$, we follow the steps below for $p = 2, 4, 6$:

(1) Randomly generate $r \times c$ groups of data with balanced sample size n from multivariate normal, multivariate Cauchy, and multivariate exponential distribution with rate $\mathbf{1}_p$ as described in Table 3.7, Table 3.9, and Table 3.11 for $p = 2, 4, 6$, respectively. Note that the location parameters in Table 3.7, Table 3.9, and Table 3.11 are determined by letting the interaction effect $\boldsymbol{\gamma} = \mathbf{0}$ in addition to the $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ described above for all cases.

(2) Apply the traditional parametric F test and the proposed nonparametric test separately. Keep the test result as 1 or 0 —1 when we "reject the null hypothesis" and 0 when we "fail to reject."

(3) Repeat (1) and (2) 10,000 times and count the percentage of rejections.

(4) Repeat (1) - (3) for sample size $n = 10, 30, 50, 70, 100$.

(5) Repeat (1) - (4) for dimension $p = 2, 4, 6$.

The procedure to estimate the empirical power is similar to the one used to estimate Type I error rate except we generate nine groups of data from the three distributions with different means for each case in Step (1). The distribution types and parameters assigned to each group in each case are listed in Table 3.8, Table 3.10, and Table 3.12 for $p = 2, 4, 6$, respectively. Additionally, we have to let the contrasts for the interaction effect

$$\boldsymbol{C}_1^T\boldsymbol{\gamma}_{ij} = \begin{bmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\gamma}_{ij} \text{ for the multivariate normal distribution cases,}$$

$$\boldsymbol{C}_2^T\boldsymbol{\gamma}_{ij} = \begin{bmatrix} -1.5 & -1.5 & 3 \\ 0 & 1.5 & -1.5 \\ 1.5 & 0 & -1.5 \end{bmatrix} \boldsymbol{\gamma}_{ij} \text{ for the multivariate Cauchy distribution cases, and}$$

$$\boldsymbol{C}_3^T\boldsymbol{\gamma}_{ij} = \begin{bmatrix} -1 & 0.5 & 0.5 \\ 1 & 0 & -1 \\ 0 & -0.5 & 0.5 \end{bmatrix} \boldsymbol{\gamma}_{ij} \text{ for the multivariate exponential distribution cases.}$$

For the case when we have two dependent variables, $p = 2$, under the null hypothesis, the conditions for calculating Type I error are shown in Table 3.7, where

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ represents the case of no correlation between dependent variables,}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \text{ represents the case of low correlation between dependent variables, and}$$

$$\boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \text{ represents the case of low correlation between dependent variables.}$$

Table 3.7.

Conditions for Calculating Type I Error for Interaction Effect for $p = 2$

Distribution	Case			Factor B		
				Level 1	Level 2	Level 3
			Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$
	Case I		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_1)$
			Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$
Multivariate Normal	Case II		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_2)$
			Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$
	Case III		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_3)$
			Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$
	Case I		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_1)$
		Factor A	Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$
Multivariate Cauchy	Case II		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_2)$
			Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$
	Case III		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_3)$
			Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$
	Case I		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_1)$
			Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$
Multivariate Exponential	Case II		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_2)$
			Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$
	Case III		Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_1 = [1 \quad 1]^T$, $\boldsymbol{\mu}_2 = [2 \quad 2]^T$, $\boldsymbol{\mu}_3 = [3 \quad 3]^T$, $\boldsymbol{\mu}_4 = [4 \quad 4]^T$, and $\boldsymbol{\mu}_5 = [5 \quad 5]^T$.

Under the alternative hypothesis, the conditions for calculating power showed in Table 3.8, the same variance-covariance matrices used for calculating Type I error are used with changes in each mean vector.

Table 3.8.

Condition for Calculating Power for Interaction Effect for $p = 2$

Distribution	Case			Factor B		
				Level 1	Level 2	Level 3
Multivariate Normal	Case I	Factor A	Level 1	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^5}, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^5}, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^5}, \boldsymbol{\Sigma}_3)$
Multivariate Cauchy	Case I		Level 1	$(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_{C^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^6}, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^7}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^8}, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_{C^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^6}, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^7}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^8}, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_{C^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^6}, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^7}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^8}, \boldsymbol{\Sigma}_3)$
Multivariate Exponential	Case I		Level 1	$(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^5}, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^5}, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^5}, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_{N^1} = [1.5 \quad 1.5]^T$, $\boldsymbol{\mu}_{N^2} = [3 \quad 3]^T$, $\boldsymbol{\mu}_{N^3} = [3.5 \quad 3.5]^T$,$\boldsymbol{\mu}_{N^4} = [4 \quad 4]^T$, $\boldsymbol{\mu}_{N^5} = [5 \quad 5]^T$, $\boldsymbol{\mu}_{C^1} = [-0.5 \quad -0.5]^T$, $\boldsymbol{\mu}_{C^2} = [0.5 \quad 0.5]^T$, $\boldsymbol{\mu}_{C^3} = [6 \quad 6]^T$, $\boldsymbol{\mu}_{C^4} = [2 \quad 2]^T$, $\boldsymbol{\mu}_{C^5} = [4.5 \quad 4.5]^T$, $\boldsymbol{\mu}_{C^6} = [2.5 \quad 2.5]^T$, $\boldsymbol{\mu}_{C^7} = [4 \quad 4]^T$, $\boldsymbol{\mu}_{C^8} = [3.5 \quad 3.5]^T$, $\boldsymbol{\mu}_{E^1} = [0 \quad 0]^T$, $\boldsymbol{\mu}_{E^2} = [2.5 \quad 2.5]^T$, $\boldsymbol{\mu}_{E^3} = [3.5 \quad 3.5]^T$,$\boldsymbol{\mu}_{E^4} = [3 \quad 3]^T$, and $\boldsymbol{\mu}_{E^5} = [5.5 \quad 5.5]^T$.

When we have four dependent variables, $p = 4$, under the null hypothesis, the conditions for calculating Type I error are shown in Table 3.9, where

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.2 & 0.3 & 0.1 \\ 0.2 & 1 & 0.1 & 0.2 \\ 0.3 & 0.1 & 1 & 0.3 \\ 0.1 & 0.2 & 0.3 & 1 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & 0.7 & 0.9 & 0.8 \\ 0.7 & 1 & 0.8 & 0.7 \\ 0.9 & 0.8 & 1 & 0.9 \\ 0.8 & 0.7 & 0.9 & 1 \end{bmatrix},$$

representing the case of no correlation, low correlation, and high correlation among dependent variables, respectively.

Table 3.9.

Conditions for Calculating Type I Error for Interaction Effect for $p = 4$

Distribution	Case			Factor B		
				Level 1	Level 2	Level 3
Multivariate Normal	Case I	Factor A	Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_3)$
Multivariate Cauchy	Case I		Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_3)$
Multivariate Exponential	Case I		Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4,\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5,\boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_1 = [1 \quad 1 \quad 1 \quad 1]^T$, $\boldsymbol{\mu}_2 = [2 \quad 2 \quad 2 \quad 2]^T$, $\boldsymbol{\mu}_3 = [3 \quad 3 \quad 3 \quad 3]^T$, $\boldsymbol{\mu}_4 = [4 \quad 4 \quad 4 \quad 4]^T$, and $\boldsymbol{\mu}_5 = [5 \quad 5 \quad 5 \quad 5]^T$.

Under the alternative hypothesis, the conditions for calculating power showed in Table 3.10, the same variance-covariance matrices used for calculating type I error are used with changes in each mean vector.

Table 3.10.

Condition for Calculating Power for Interaction Effect for $p = 4$

Distribution	Case			Factor B		
				Level 1	Level 2	Level 3
Multivariate Normal	Case I	Factor A	Level 1	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^5}, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^5}, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_{N^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^3}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{N^2}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^5}, \boldsymbol{\Sigma}_3)$
Multivariate Cauchy	Case I		Level 1	$(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_{C^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^6}, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^7}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^8}, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_{C^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^6}, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^7}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^8}, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_{C^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^2}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^3}, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_{C^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^6}, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{C^5}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^7}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^8}, \boldsymbol{\Sigma}_3)$
Multivariate Exponential	Case I		Level 1	$(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^5}, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^5}, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_{E^1}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^2}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{E^4}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^3}, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^5}, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_{N^1} = [1.5 \quad 1.5 \quad 1.5 \quad 1.5]^T$, $\boldsymbol{\mu}_{N^2} = [3 \quad 3 \quad 3 \quad 3]^T$, $\boldsymbol{\mu}_{N^3} = [3.5 \quad 3.5 \quad 3.5 \quad 3.5]^T$, $\boldsymbol{\mu}_{N^4} = [4 \quad 4 \quad 4 \quad 4]^T$, $\boldsymbol{\mu}_{N^5} = [5 \quad 5 \quad 5 \quad 5]^T$, $\boldsymbol{\mu}_{C^1} = [-0.5 \quad -0.5 \quad 0.5 \quad -0.5]^T$, $\boldsymbol{\mu}_{C^2} = [0.5 \quad 0.5 \quad 0.5 \quad 0.5]^T$, $\boldsymbol{\mu}_{C^3} = [6 \quad 6 \quad 6 \quad 6]^T$, $\boldsymbol{\mu}_{C^4} = [2 \quad 2 \quad 2 \quad 2]^T$,

$\boldsymbol{\mu}_{C^5} = [4.5 \quad 4.5 \quad 4.5 \quad 4.5]^T$, $\boldsymbol{\mu}_{C^6} = [2.5 \quad 2.5 \quad 2.5 \quad 2.5]^T$, $\boldsymbol{\mu}_{C^7} = [4 \quad 4 \quad 4 \quad 4]^T$, $\boldsymbol{\mu}_{C^8} = [3.5 \quad 3.5 \quad 3.5 \quad 3.5]^T$, $\boldsymbol{\mu}_{E^1} = [0 \quad 0 \quad 0 \quad 0]^T$, $\boldsymbol{\mu}_{E^2} = [2.5 \quad 2.5 \quad 2.5 \quad 2.5]^T$, $\boldsymbol{\mu}_{E^3} = [3.5 \quad 3.5 \quad 3.5 \quad 3.5]^T$, $\boldsymbol{\mu}_{E^4} = [3 \quad 3 \quad 3 \quad 3]^T$, and $\boldsymbol{\mu}_{E^5} = [5.5 \quad 5.5 \quad 5.5 \quad 5.5]^T$.

For the case when we have six dependent variables, $p = 6$, under the null hypothesis, the conditions for calculating Type I error are shown in Table 3.11, where

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 \\ 0.1 & 1 & 0.1 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.1 & 1 & 0.1 & 0.2 & 0.3 \\ 0.3 & 0.2 & 0.1 & 1 & 0.1 & 0.2 \\ 0.2 & 0.3 & 0.2 & 0.1 & 1 & 0.3 \\ 0.1 & 0.1 & 0.1 & 0.2 & 0.3 & 1 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma}_3 = \begin{bmatrix} 1 & 0.7 & 0.8 & 0.9 & 0.8 & 0.7 \\ 0.7 & 1 & 0.7 & 0.8 & 0.9 & 0.7 \\ 0.8 & 0.7 & 1 & 0.8 & 0.9 & 0.7 \\ 0.9 & 0.8 & 0.8 & 1 & 0.8 & 0.9 \\ 0.8 & 0.9 & 0.9 & 0.8 & 1 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.9 & 0.7 & 1 \end{bmatrix}.$$

Representing the case of no correlation, low correlation, and high correlation among dependent variables, respectively.

Table 3.11.

Conditions for Calculating Type I Error for Interaction Effect for $p = 6$

Distribution	Case			Factor B		
				Level 1	Level 2	Level 3
Multivariate Normal	Case I	Factor A	Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_3)$
Multivariate Cauchy	Case I		Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_3)$
Multivariate Exponential	Case I		Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_1)$
	Case II		Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_2)$
	Case III		Level 1	$(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$
			Level 2	$(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_1 = [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]^T$, $\boldsymbol{\mu}_2 = [2 \quad 2 \quad 2 \quad 2 \quad 2 \quad 2]^T$, $\boldsymbol{\mu}_3 = [3 \quad 3 \quad 3 \quad 3 \quad 3 \quad 3]^T$, $\boldsymbol{\mu}_4 = [4 \quad 4 \quad 4 \quad 4 \quad 4 \quad 4]^T$, and $\boldsymbol{\mu}_5 = [5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5]^T$.

Under the alternative hypothesis, the conditions for calculating power showed in Table 3.12, the same variance-covariance matrices used for calculating Type I error are used with changes in each mean vector.

Table 3.12.

Condition for Calculating Power for Interaction Effect for $p = 6$

Distribution	Case			Factor B Level 1	Factor B Level 2	Factor B Level 3
			Level 1	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^2},\boldsymbol{\Sigma}_1)$
	Case I		Level 2	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^3},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^4},\boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{N^2},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^4},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{N^5},\boldsymbol{\Sigma}_1)$
			Level 1	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^2},\boldsymbol{\Sigma}_2)$
Multivariate Normal	Case II		Level 2	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^3},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^4},\boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{N^2},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^4},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{N^5},\boldsymbol{\Sigma}_2)$
			Level 1	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^2},\boldsymbol{\Sigma}_3)$
	Case III		Level 2	$(\boldsymbol{\mu}_{N^1},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^3},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^4},\boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{N^2},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^4},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{N^5},\boldsymbol{\Sigma}_3)$
			Level 1	$(\boldsymbol{\mu}_{C^1},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^2},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^3},\boldsymbol{\Sigma}_1)$
	Case I		Level 2	$(\boldsymbol{\mu}_{C^4},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^5},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^6},\boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{C^5},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^7},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{C^8},\boldsymbol{\Sigma}_1)$
		Factor A	Level 1	$(\boldsymbol{\mu}_{C^1},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^2},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^3},\boldsymbol{\Sigma}_2)$
Multivariate Cauchy	Case II		Level 2	$(\boldsymbol{\mu}_{C^4},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^5},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^6},\boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{C^5},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^7},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{C^8},\boldsymbol{\Sigma}_2)$
			Level 1	$(\boldsymbol{\mu}_{C^1},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^2},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^3},\boldsymbol{\Sigma}_3)$
	Case III		Level 2	$(\boldsymbol{\mu}_{C^4},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^5},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^6},\boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{C^5},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^7},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{C^8},\boldsymbol{\Sigma}_3)$
			Level 1	$(\boldsymbol{\mu}_{E^1},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^2},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^3},\boldsymbol{\Sigma}_1)$
	Case I		Level 2	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_1)$
			Level 3	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^3},\boldsymbol{\Sigma}_1)$	$(\boldsymbol{\mu}_{E^5},\boldsymbol{\Sigma}_1)$
			Level 1	$(\boldsymbol{\mu}_{E^1},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^2},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^3},\boldsymbol{\Sigma}_2)$
Multivariate Exponential	Case II		Level 2	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_2)$
			Level 3	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^3},\boldsymbol{\Sigma}_2)$	$(\boldsymbol{\mu}_{E^5},\boldsymbol{\Sigma}_2)$
			Level 1	$(\boldsymbol{\mu}_{E^1},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^2},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^3},\boldsymbol{\Sigma}_3)$
	Case III		Level 2	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_3)$
			Level 3	$(\boldsymbol{\mu}_{E^4},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^3},\boldsymbol{\Sigma}_3)$	$(\boldsymbol{\mu}_{E^5},\boldsymbol{\Sigma}_3)$

where $\boldsymbol{\mu}_{N^1} = [1.5 \quad 1.5 \quad 1.5 \quad 1.5 \quad 1.5 \quad 1.5]^T$, $\boldsymbol{\mu}_{N^2} = [3 \quad 3 \quad 3 \quad 3 \quad 3 \quad 3]^T$,

$\boldsymbol{\mu}_{N^3} = [3.5 \quad 3.5 \quad 3.5 \quad 3.5 \quad 3.5 \quad 3.5]^T$, $\boldsymbol{\mu}_{N^4} = [4 \quad 4 \quad 4 \quad 4 \quad 4 \quad 4]^T$,

$\boldsymbol{\mu}_{N^5} = [5 \quad 5 \quad 5 \quad 5 \quad 5 \quad 5]^T$, $\boldsymbol{\mu}_{C^1} = [-0.5 \quad -0.5 \quad 0.5 \quad -0.5 \quad 0.5 \quad -0.5]^T$,

$\boldsymbol{\mu}_{C^2} = [0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad 0.5]^T$, $\boldsymbol{\mu}_{C^3} = [6 \quad 6 \quad 6 \quad 6 \quad 6 \quad 6]^T$,

$\boldsymbol{\mu}_{C^4} = [2 \quad 2 \quad 2 \quad 2 \quad 2 \quad 2]^T$, $\boldsymbol{\mu}_{C^5} = [4.5 \quad 4.5 \quad 4.5 \quad 4.5 \quad 4.5 \quad 4.5]^T$,

$\boldsymbol{\mu}_{C^6} = [2.5 \quad 2.5 \quad 2.5 \quad 2.5 \quad 2.5 \quad 2.5]^T$, $\boldsymbol{\mu}_{C^7} = [4 \quad 4 \quad 4 \quad 4 \quad 4 \quad 4]^T$,

$\boldsymbol{\mu}_{C^8} = [3.5 \quad 3.5 \quad 3.5 \quad 3.5 \quad 3.5 \quad 3.5]^T$, $\boldsymbol{\mu}_{E^1} = [0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]^T$,

$\boldsymbol{\mu}_{E^2} = [2.5 \quad 2.5 \quad 2.5 \quad 2.5 \quad 2.5 \quad 2.5]^T$, $\boldsymbol{\mu}_{E^3} = [3.5 \quad 3.5 \quad 3.5 \quad 3.5 \quad 3.5 \quad 3.5]^T$, $\boldsymbol{\mu}_{E^4} = [3 \quad 3 \quad 3 \quad 3 \quad 3 \quad 3]^T$, and $\boldsymbol{\mu}_{E^5} = [5.5 \quad 5.5 \quad 5.5 \quad 5.5 \quad 5.5 \quad 5.5]^T$.

CHAPTER IV

RESULTS

This chapter presents and interprets the results of the simulation procedure described in Chapter III in order to evaluate the performance of the proposed nonparametric kernel-based multivariate analysis of the variance (MANOVA). The main purpose of this simulation was to evaluate Type I error rate and power of the proposed nonparametric kernel-based MANOVA tests for the one-way layout and the interaction term in the two-way layout developed in Chapter III of this dissertation, then compare it to the traditional parametric MANOVA for the one-way layout and the interaction term in the two-way layout. Additionally, this chapter briefly discusses the research questions of this dissertation provided in Chapter I.

Summary of Results

Answers to Research Questions

The first four research questions given in Chapter I were addressed in Chapter III; they are also briefly discussed below for adoption by researchers and applied practitioners. More details about each of the research questions are discussed in Chapter III.

Q1 How can a kernel density estimator be constructed for non-Gaussian multivariate data?

Answer to Question 1. In order to construct a multivariate kernel density estimator (KDE) for non-normal and non-linear multivariate data, three steps need to be taken which are summarized below

Step 1. A basic multivariate kernel estimator of $f(x)$ is used as described by Wand and Jones (1994):

$$\hat{f}_i(x;H) = \frac{1}{n|\boldsymbol{H}|^{\frac{1}{2}}} \sum_{i=1}^{n} K(\boldsymbol{H}^{-1/2}(x - X_i)) \tag{4.1}$$

Step 2. Find K which is defined as a kernel function. K can be any kernel function determined by the researcher or by the data. In this dissertation, the Gaussian (normal) kernel is used which can be written such that

$$K(z) = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}z'z}. \tag{4.2}$$

Step 3. Calculate $\boldsymbol{H}$ is defined as a bandwidth matrix which can be determined by the researcher or driven from the data. In this dissertation, the generalized Scott's rule of thumb is used, such as

$$\hat{\boldsymbol{H}} = n^{\frac{-1}{p+4}} \hat{\Sigma}^{\frac{1}{2}}, \tag{4.3}$$

where $\hat{\Sigma}$ is the sample variance-covariance matrix obtained from the data.

Q2 How can hypotheses be tested using multivariate data when kernel methods are used within the one-way MANOVA technique?

Answer to Question 2. In order to test hypotheses for the kernel-based one-way MANOVA, three steps need to be taken which are summarized below

Step 1. The null hypothesis to test the equality of the group means needs to be written using kernel density estimation (KDE) method as $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_I$, where

$$\hat{\boldsymbol{Q}}_i = \frac{1}{n_i(n_i-1)|\boldsymbol{H}_i|^{\frac{1}{2}}} \sum \sum_{j_1 \neq j_2} \frac{1}{2} \left(\boldsymbol{X}_{ij_1} + \boldsymbol{X}_{ij_2}\right) K\left(\boldsymbol{H_i}^{-1/2}(\boldsymbol{X}_{ij_1} - \boldsymbol{X}_{ij_2})\right). \tag{4.4}$$

Step 2. It was shown in Chapter III that the between sum of squares of the nonparametric kernel-based one-way MANOVA, ***SSB***, follows a Wishart distribution with $(I-1)$ degrees of freedom and $\boldsymbol{I}$ scale matrix, where I is the number of groups of the factor. It was also shown that the within sum of squares of the nonparametric kernel-based one-way MANOVA, $\boldsymbol{SSW}$ follows Wishart distributions with df_w degrees of freedom and $\boldsymbol{I}$ scale matrix. More details about $\boldsymbol{SSB}$ and $\boldsymbol{SSW}$ and their degrees of freedom are provided in Chapter III.

Step 3. It was also shown in Chapter III that the proposed nonparametric kernel-based one-way MANOVA test follows an F distribution using Wilks' lambda approximation, where

$$\Lambda_{\mathrm{k}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSB}+\boldsymbol{SSW}|} = \frac{1}{1+\left|\frac{\boldsymbol{SSB}}{\boldsymbol{SSW}}\right|}, \tag{4.5}$$

and

$$F = \frac{1-(\Lambda_{\mathrm{k}})^{1/b_1}}{(\Lambda_{\mathrm{k}})^{1/b_1}} \times \frac{df_2}{df_1} \dot{\sim} F(df_1, df_2), \tag{4.6}$$

where df_1 and df_2 are given in detail in Chapter III.

Q3 How can the main effect hypothesis be tested using multivariate data when kernel methods are used within the two-way MANOVA technique?

Answer to Question 3. In order to test hypotheses for the main effects in the kernel-based two-way MANOVA, three steps need to be taken which are summarized below

Step 1a. The null hypothesis to test the equality of the row main effect needs to be written using the kernel density estimation (KDE) method while accounting for the two-way decomposition restriction described in Chapter III, as $H_0 : \boldsymbol{\alpha}_i = \boldsymbol{0}$, where

$$\hat{\boldsymbol{Q}}_{ij} = \frac{1}{n_{ij}(n_{ij}-1)|\boldsymbol{H}_{ij}^{1/2}|} \sum \sum_{k_1 \neq k_2} \frac{1}{2} (\boldsymbol{X}_{ijk_1} + \boldsymbol{X}_{ijk_2}) K(\boldsymbol{H}_{ij}^{-1/2} (\boldsymbol{X}_{ijk_1} - \boldsymbol{X}_{ijk_2})). \tag{4.7}$$

Step 1b. The null hypothesis to test the equality of the column main effect needs to be written using the kernel density estimation (KDE) method while accounting for the two-way decomposition restriction described in Chapter III, as $H_0 : \boldsymbol{\beta}_j = \mathbf{0}$, where

$$\hat{\boldsymbol{Q}}_{ij} = \frac{1}{n_{ij}(n_{ij}-1)|\boldsymbol{H}_{ij}^{1/2}|} \sum \sum_{k_1 \neq k_2} \frac{1}{2} (\boldsymbol{X}_{ijk_1} + \boldsymbol{X}_{ijk_2}) K(\boldsymbol{H}_{ij}^{-1/2} (\boldsymbol{X}_{ijk_1} - \boldsymbol{X}_{ijk_2})). \tag{4.8}$$

Step 2a. Chapter III showed also that the row sum of squares of the nonparametric kernel-based two-way MANOVA, $\boldsymbol{SSR}$, follows a Wishart distribution with $(I-1)$ degrees of freedom and $\boldsymbol{I}$ scale matrix, where I is the number of groups of the row effect factor. It was also shown that the within sum of squares of the nonparametric kernel-based two-way MANOVA, $\boldsymbol{SSW}$ follows a Wishart distribution with df_{w_2} degrees of freedom and $\boldsymbol{I}$ scale matrix. More details about $\boldsymbol{SSR}$ and $\boldsymbol{SSW}$ and their degrees of freedom are provided in Chapter III.

Step 2b. It was shown in Chapter III that the column sum of squares of the nonparametric kernel-based two-way MANOVA, $\boldsymbol{SSC}$, follows Wishart distributions with $(J-1)$ degrees of freedom and $\boldsymbol{I}$ scale matrix, where J is the number of groups of the column effect factor. It was also shown that the within sum of squares of the nonparametric kernel-based two-way MANOVA, $\boldsymbol{SSW}$ follows Wishart distributions with df_{w_2} degrees of freedom and $\boldsymbol{I}$ scale matrix. More details about $\boldsymbol{SSC}$ and $\boldsymbol{SSW}$ and their degrees of freedom are provided in Chapter III.

Step 3a. Chapter III also showed that the row effect in the proposed nonparametric kernel-based two-way MANOVA test follows an F distribution using Wilks' lambda approximation, where

$$\Lambda_{\mathrm{kR}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSR} + \boldsymbol{SSW}|} = \frac{1}{1 + \left|\frac{\boldsymbol{SSR}}{\boldsymbol{SSW}}\right|} \tag{4.9}$$

and

$$F = \frac{1 - (\Lambda_{\mathrm{kR}})^{1/b_R}}{(\Lambda_{\mathrm{kR}})^{1/b_R}} \times \frac{df_2}{df_1} \dot{\sim} F(df_1, df_2), \tag{4.10}$$

where df_1 and df_2 are given in detail in Chapter III.

Step 3b. It was also shown in Chapter III that the column effect in the proposed nonparametric kernel-based two-way MANOVA test follows an F distribution using Wilks' lambda approximation, where

$$\Lambda_{\mathrm{kC}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSC}+\boldsymbol{SSW}|} = \frac{1}{1+\left|\frac{\boldsymbol{SSC}}{\boldsymbol{SSW}}\right|} \tag{4.11}$$

and

$$F = \frac{1-(\Lambda_{\mathrm{kC}})^{1/b_C}}{(\Lambda_{\mathrm{kC}})^{1/b_C}} \times \frac{df_2}{df_1} \dot{\sim} F(df_1, df_2), \tag{4.12}$$

where df_1 and df_2 are given in detail in Chapter III.

Q4 How can the interaction effect hypothesis be tested using multivariate data when kernel methods are used within the two-way MANOVA technique?

Answer to Question 4. In order to test hypotheses for the interaction effect in the kernel-based two-way MANOVA, three steps need to be taken which are summarized below

Step 1. The null hypothesis to test the equality of the interaction effect needs to be written using the kernel density estimation (KDE) method, while accounting for the two-way decomposition restriction described in Chapter III, as $H_0 : \boldsymbol{\gamma}_{ij} = \mathbf{0}$.

Step 2. It was shown in Chapter III that the interaction sum of squares of the nonparametric kernel-based two-way MANOVA, $\boldsymbol{SSI}$, follows a Wishart distribution with $((I-1)(J-1))$ degrees of freedom and $\boldsymbol{I}$ scale matrix, where I is the number of groups of the row effect factor and J is the number of groups of the column effect factor. It was also shown that the within sum of squares of the nonparametric kernel-based two-way MANOVA, $\boldsymbol{SSW}$, follows a Wishart distribution with df_{w_2} degrees of freedom and $\boldsymbol{I}$ scale matrix. More details about $\boldsymbol{SSI}$ and $\boldsymbol{SSW}$ as well as their degrees of freedom are provided in Chapter III.

Step 3. It was also shown in Chapter III that the interaction effect in the proposed nonparametric kernel-based two-way MANOVA test follows an F distribution using

Wilks' lambda approximation, where

$$\Lambda_{\mathrm{kI}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSI}+\boldsymbol{SSW}|} = \frac{1}{1+\left|\frac{\boldsymbol{SSI}}{\boldsymbol{SSW}}\right|}, \tag{4.13}$$

and

$$F = \frac{1-(\Lambda_{\mathrm{kI}})^{1/b_I}}{(\Lambda_{\mathrm{kI}})^{1/b_I}} \times \frac{df_2}{df_1} \dot{\sim} F(df_1, df_2), \tag{4.14}$$

where df_1 and df_2 are given in detail in Chapter III.

In order to answer the fifth and sixth research question given in Chapter I, an in depth simulation study needs to be conducted. Details about the simulations are provided below.

Simulation Study for Evaluating Type I Error and Power for the Nonparametric Kernel-Based One-Way MANOVA

In this section, the performance of the Type I error and power of the proposed nonparametric kernel-based one-way MANOVA is evaluated and compared to the performance of the Type I error and power of the traditional parametric one-way MANOVA.

Q5 How do Type I error rate and power of the proposed kernel-based one-way MANOVA test behave compared to the parametric one-way MANOVA?

Answer to Question 5. In order order to answer this question, multiple simulations were conducted using different dimensions, sample sizes, and variance-covariance matrices representing no correlation, low correlation, and high correlation among variables. The simulation study was conducting using different distributions: multivariate normal, multivariate Cauchy, and multivariate exponential distribution. Conditions for this simulation study is shown in Tables 3.1 - 3.6 in Chapter III. In this simulation study, the Type I error and power of the proposed nonparametric kernel-based one-way MANOVA is evaluated and compared to the performance of the

Type I error and power of the traditional parametric one-way MANOVA. The simulation results are provided in Table 4.1 - Table 4.3.

Table 4.1.

Type I Error and Power for the Kernel-Based vs. Parametric Tests: One-Way MANOVA for Multivariate Normal Distribution

	n_i	Case I				Case II				Case III			
		Parametric		Nonparametric		Parametric		Nonparametric		Parametric		Nonparametric	
		Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power
	10	0.048(4.8%)	0.633(63.3%)	0.053(5.3%)	0.422(42.2%)	0.044(4.4%)	0.629(62.9%)	0.049(4.9%)	0.418(41.8%)	0.046(4.6%)	0.599(59.9%)	0.051(5.1%)	0.379(37.9%)
	30	0.047(4.7%)	0.994(99.4%)	0.051(5.9%)	0.771(77.1%)	0.048(4.8%)	0.986(98.6%)	0.052(5.2%)	0.741(74.1%)	0.043(4.3%)	0.908(90.8%)	0.054(5.4%)	0.613(61.3%)
$p=2$	50	0.051(5.1%)	0.998(99.8%)	0.053(5.3%)	0.837(80.9%)	0.050(4.6%)	0.997(99.7%)	0.046(4.6%)	0.809(83.7%)	0.049(4.9%)	0.994(99.4%)	0.052(5.2%)	0.765(76.5%)
	70	0.053(5.3%)	0.999(99.9%)	0.049(4.9%)	0.888(87.4%)	0.046(4.6%)	0.999(99.9%)	0.050(5.0%)	0.874(88.8%)	0.047(4.7%)	0.999(99.9%)	0.044(4.4%)	0.832(83.2%)
	100	0.045(4.5%)	1.000(100.0%)	0.052(5.2%)	0.953(95.3%)	0.051(5.1%)	1.000(100.0%)	0.054(5.4%)	0.932(93.2%)	0.053(5.3%)	1.000(100.0%)	0.043(4.3%)	0.917(91.7%)
	10	0.052(5.2%)	0.634(63.4%)	0.053(5.3%)	0.594(69.4%)	0.049(4.9%)	0.634(63.4%)	0.051(5.1%)	0.538(53.8%)	0.053(5.3%)	0.615(61.5%)	0.048(4.8%)	0.402(40.2%)
	30	0.052(5.2%)	0.995(99.5%)	0.052(5.2%)	0.768(76.8%)	0.050(5.0%)	0.986(98.6%)	0.046(4.6%)	0.756(75.6%)	0.048(4.8%)	0.923(92.3%)	0.051(5.1%)	0.732(73.2%)
$p=4$	50	0.050(5.0%)	0.999(99.9%)	0.047(4.7%)	0.824(82.4%)	0.050(5.0%)	0.999(99.9%)	0.048(4.8%)	0.817(81.7%)	0.051(5.1%)	0.996(99.6%)	0.046(4.6%)	0.815(81.5%)
	70	0.048(4.8%)	1.000(100.0%)	0.049(4.9%)	0.893(89.3%)	0.053(5.3%)	1.000(100.0%)	0.046(4.6%)	0.869(86.9%)	0.050(5.0%)	0.999(99.9%)	0.048(4.8%)	0.869(86.9%)
	100	0.051(5.1%)	1.000(100.0%)	0.053(5.3%)	0.952(95.2%)	0.049(4.9%)	1.000(100.0%)	0.051(5.1%)	0.948(94.8%)	0.051(5.1%)	1.000(100.0%)	0.051(5.1%)	0.945(94.5%)
	10	0.051(5.1%)	0.650(65.0%)	0.048(4.8%)	0.644(64.4%)	0.047(4.7%)	0.643(64.3%)	0.052(5.2%)	0.641(64.1%)	0.049(4.9%)	0.634(63.4%)	0.047(4.7%)	0.632(63.2%)
	30	0.047(4.7%)	0.997(99.7%)	0.051(5.1%)	0.761(76.1%)	0.048(4.8%)	0.998(99.8%)	0.046(4.6%)	0.759(75.9%)	0.052(5.2%)	0.956(95.6%)	0.051(5.1%)	0.753(75.3%)
$p=6$	50	0.052(5.2%)	1.000(100.0%)	0.053(5.3%)	0.825(82.5%)	0.052(5.2%)	1.000(100.0%)	0.046(4.6%)	0.800(80.0%)	0.046(4.6%)	0.998(99.8%)	0.052(5.2%)	0.799(79.9%)
	70	0.048(4.8%)	1.000(100.0%)	0.047(4.7%)	0.886(88.6%)	0.051(5.1%)	1.000(100.0%)	0.051(5.1%)	0.872(87.2%)	0.053(5.3%)	1.000(100.0%)	0.045(4.5%)	0.866(86.6%)
	100	0.053(5.3%)	1.000(100.0%)	0.049(4.9%)	0.962(96.2%)	0.049(4.9%)	1.000(100.0%)	0.048(4.8%)	0.948(94.8%)	0.051(5.1%)	1.000(100.0%)	0.044(4.4%)	0.945(94.5%)

Table 4.1 shows that the actual Type I errors for both the proposed nonparametric kernel-based and the parametric one-way MANOVA tests for balanced design, when the underlying distribution of the data is multivariate normal, were around 0.05. As expected, the results generally performed better for the parametric MANOVA tests than the proposed nonparametric kernel-based tests. In addition, the power of both parametric and nonparametric tests increased as the sample size and dimension increased. Also, it was observed that as the correlation in the variance-covariance matrices increased, the power of both tests slightly decreased.

For Case I, Case II, and Case III, when $p = 2$, the parametric one-way MANOVA test reached a power of 100% when the sample size reaches 100. However, the proposed nonparametric kernel-based one-way MANOVA only reached a power of around 95%, 93%, and 91% when the sample size reached 100 for Case I, Case II, and Case III, respectively. When $p = 4$ and the sample size reached 70, the parametric one-way MANOVA tests reach a power of 100% or close to 100% for Case I, Case II, and Case III.

However, when the sample size reaches 100, the proposed nonparametric kernel-based one-way MANOVA only reached a power of around 95% for Case I and Case II, and 94% for Case III. When $p = 6$ and the sample size reached 50, the parametric one-way MANOVA tests reach a power of 100% or close to 100% for Case I, Case II, and Case III. However, the proposed nonparametric kernel-based one-way MANOVA only reaches a power of around 96%, 95%, and 94% when the sample size reaches 100 for Case I, Case II, and Case III, respectively.

Figure 4.1 shows the power plots for the one-way MANOVA when the underlying distribution of the data is multivariate normal for $p = 2,4,6$. The proposed nonparametric kernel-based tests are represented as solid lines while the parametric tests represented as dashed lines in Figure 4.1. In general, the traditional parametric one-way MANOVA test had a higher power than the proposed nonparametric approach when the data come from multivariate normal. Additionally, as the sample size increased the parametric test reached a higher power faster than the proposed nonparametric kernel-based test. It can also be observed that the lower the correlation in the variance-covariance matrix, the higher the power.

Table 4.2 shows that the Type I errors were around 0.02 for both the proposed nonparametric kernel-based and parametric one-way MANOVA tests for balanced design when the underlying distribution of the data is multivariate Cauchy. As expected, the results generally performed better for the proposed nonparametric kernel-based tests than the parametric MANOVA tests. In addition, the power of both parametric and nonparametric tests increased as the sample size and dimension increased. Also, it was observed that as the correlation in the variance-covariance matrices increased, the power of both tests slightly decreased.

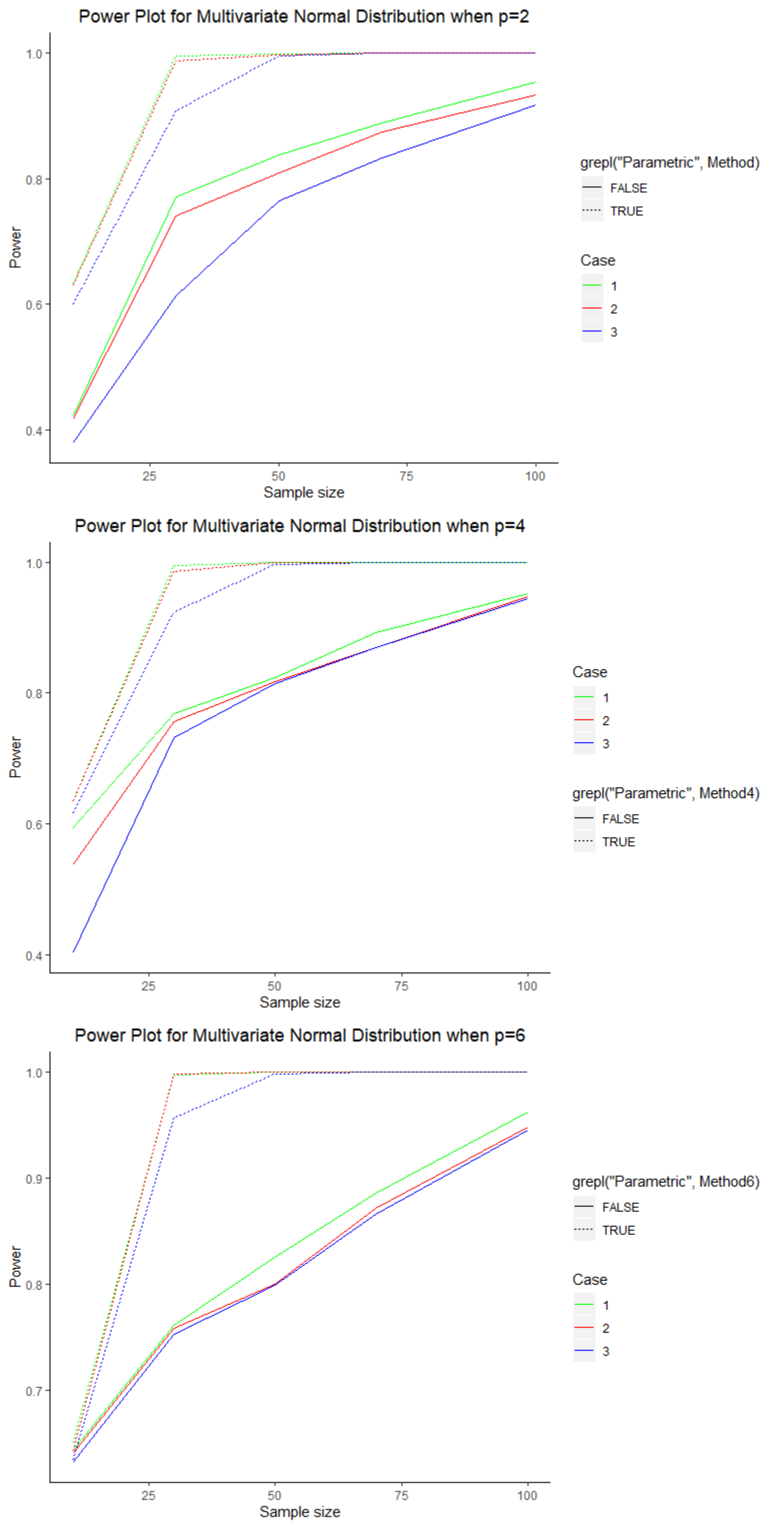

Figure 4.1. Power vs. Sample Size for the One-Way MANOVA for Multivariate Normal Distribution

Table 4.2.

Type I Error and Power for the Kernel-Based vs. Parametric Tests: One-Way MANOVA for Multivariate Cauchy Distribution

	n_i	Case I				Case II				Case III			
		Parametric		Nonparametric		Parametric		Nonparametric		Parametric		Nonparametric	
		Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power
	10	0.018(1.8%)	0.296(29.6%)	0.023(2.3%)	0.752(75.2%)	0.015(1.5%)	0.191(19.1%)	0.022(2.2%)	0.745(74.5%)	0.016(1.6%)	0.170(17.0%)	0.021(2.1%)	0.729(72.9%)
	30	0.018(1.8%)	0.548(54.8%)	0.022(2.2%)	0.817(81.7%)	0.016(1.6%)	0.422(42.2%)	0.017(1.7%)	0.809(80.9%)	0.012(1.2%)	0.403(40.3%)	0.012(1.2%)	0.779(77.9%)
$p=2$	50	0.019(1.9%)	0.657(65.7%)	0.024(2.4%)	0.863(86.3%)	0.013(1.3%)	0.649(64.9%)	0.014(1.4%)	0.861(86.1%)	0.017(1.7%)	0.558(55.8%)	0.018(1.8%)	0.854(85.4%)
	70	0.017(1.7%)	0.801(80.1%)	0.016(1.6%)	0.879(87.9%)	0.014(1.4%)	0.776(77.6%)	0.018(1.8%)	0.876(87.6%)	0.012(1.2%)	0.728(72.8%)	0.013(1.3%)	0.873(87.3%)
	100	0.012(1.2%)	0.939(93.9%)	0.019(1.9%)	0.963(96.3%)	0.011(1.1%)	0.930(93.0%)	0.012(1.2%)	0.954(95.4%)	0.015(1.5%)	0.923(92.3%)	0.011(1.1%)	0.948(94.8%)
	10	0.016(1.6%)	0.384(38.4%)	0.018(1.8%)	0.767(76.7%)	0.015(1.5%)	0.299(29.9%)	0.012(1.2%)	0.763(76.3%)	0.012(1.2%)	0.273(27.3%)	0.021(2.1%)	0.738(73.8%)
	30	0.017(1.7%)	0.562(56.2%)	0.015(1.5%)	0.830(83.0%)	0.016(1.6%)	0.457(45.7%)	0.016(1.6%)	0.818(81.8%)	0.014(1.4%)	0.438(43.8%)	0.012(1.2%)	0.807(80.7%)
$p=4$	50	0.014(1.4%)	0.692(69.2%)	0.015(1.5%)	0.881(88.1%)	0.014(1.4%)	0.677(67.7%)	0.014(1.4%)	0.879(87.9%)	0.019(1.9%)	0.618(61.8%)	0.018(1.8%)	0.856(85.6%)
	70	0.017(1.7%)	0.827(82.7%)	0.021(2.1%)	0.897(89.7%)	0.012(1.2%)	0.801(80.1%)	0.015(1.5%)	0.888(88.8%)	0.011(1.1%)	0.756(75.6%)	0.013(1.3%)	0.885(88.5%)
	100	0.013(1.3%)	0.943(94.3%)	0.018(1.8%)	0.962(96.2%)	0.015(1.5%)	0.936(93.6%)	0.017(1.7%)	0.959(95.9%)	0.013(1.3%)	0.933(93.3%)	0.011(1.1%)	0.947(94.7%)
	10	0.013(1.3%)	0.408(40.8%)	0.017(1.7%)	0.759(75.9%)	0.015(1.5%)	0.346(34.6%)	0.015(1.5%)	0.754(75.4%)	0.011(1.1%)	0.297(29.7%)	0.022(2.2%)	0.741(74.1%)
	30	0.016(1.6%)	0.589(58.9%)	0.016(1.6%)	0.834(83.4%)	0.016(1.6%)	0.512(51.2%)	0.016(1.6%)	0.821(82.1%)	0.015(1.5%)	0.482(48.2%)	0.021(2.1%)	0.819(81.9%)
$p=6$	50	0.018(1.8%)	0.699(69.9%)	0.022(2.2%)	0.878(87.8%)	0.012(1.2%)	0.696(69.6%)	0.017(1.7%)	0.876(87.6%)	0.017(1.7%)	0.647(64.7%)	0.018(1.8%)	0.868(86.8%)
	70	0.014(1.4%)	0.853(85.3%)	0.017(1.7%)	0.899(89.9%)	0.014(1.4%)	0.839(83.9%)	0.019(1.9%)	0.890(89.0%)	0.014(1.4%)	0.792(79.2%)	0.019(1.9%)	0.876(87.6%)
	100	0.014(1.4%)	0.969(96.9%)	0.019(1.9%)	0.963(96.3%)	0.017(1.7%)	0.968(96.8%)	0.021(2.1%)	0.961(96.1%)	0.012(1.2%)	0.968(96.8%)	0.016(1.6%)	0.954(95.4%)

In Table 4.2, when $p = 2$, the parametric one-way MANOVA test reached a power of 94%, 93%, and 92% when the sample size reached 100 for Case I, Case II, and Case III. However, the proposed nonparametric kernel-based one-way MANOVA reached a power of around 96% for Case I, and 95% for Case II and Case III. When $p = 4$, the parametric one-way MANOVA test only reached a power of 94% when the sample size reached 100 for Case I, and 93% for Case II and Case III. However, the proposed nonparametric kernel-based one-way MANOVA reached a power of around 96% for Case I, and 95% for Case II and Case III. When $p = 6$, the parametric one-way MANOVA test only reached a power of 94%, 93%, and 92% when the sample size reached 100 for Case I, Case II, and Case III, respectively. However, the proposed nonparametric kernel-based one-way MANOVA reached a power of around 96% for Case I and Case II, and 95% for Case III.

Figure 4.2 shows the power plots for the one-way MANOVA when the underlying distribution of the data is multivariate Cauchy for $p = 2, 4, 6$. The proposed nonparametric kernel-based tests are represented as solid lines while the parametric tests represented as dashed lines in Figure 4.2. In general, the proposed nonparametric kernel-based one-way MANOVA test had a higher power than the traditional parametric approach when the data come from multivariate Cauchy distribution. Additionally, as the sample size increased, the nonparametric test had a higher power for smaller sample sizes than the parametric test. Also, the the proposed nonparametric kernel-based test reached a higher power faster than the parametric test. It can also be seen that the lower the correlation in the variance-covariance matrix, the higher the power.

Table 4.3.

Type I Error and Power for the Kernel-Based vs. Parametric Tests: One-Way MANOVA for Multivariate Exponential Distribution

	n_i	Case I				Case II				Case III			
		Parametric		Nonparametric		Parametric		Nonparametric		Parametric		Nonparametric	
		Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power
	10	0.038(3.8%)	0.332(33.2%)	0.034(3.4%)	0.685(68.5%)	0.041(4.1%)	0.326(32.6%)	0.032(3.2%)	0.682(68.2%)	0.039(3.9%)	0.309(30.9%)	0.035(3.5%)	0.679(67.9%)
	30	0.042(4.2%)	0.567(56.7%)	0.037(3.7%)	0.799(79.9%)	0.041(4.1%)	0.535(53.5%)	0.035(3.5%)	0.791(79.1%)	0.040(4.0%)	0.528(52.8%)	0.029(2.9%)	0.788(78.8%)
$p=2$	50	0.039(3.9%)	0.658(65.8%)	0.044(4.4%)	0.861(86.1%)	0.041(4.1%)	0.647(64.7%)	0.037(3.7%)	0.858(85.8%)	0.037(3.7%)	0.643(64.3%)	0.041(4.1%)	0.853(85.3%)
	70	0.041(4.1%)	0.793(79.3%)	0.035(3.5%)	0.889(87.9%)	0.039(3.9%)	0.787(78.7%)	0.041(4.1%)	0.877(87.7%)	0.044(4.4%)	0.772(77.2%)	0.038(3.8%)	0.870(87.0%)
	100	0.044(4.4%)	0.874(87.4%)	0.039(3.9%)	0.967(96.3%)	0.036(3.6%)	0.872(87.2%)	0.038(3.8%)	0.958(95.8%)	0.041(4.1%)	0.868(86.8%)	0.036(3.6%)	0.949(94.9%)
	10	0.041(4.1%)	0.353(35.3%)	0.035(3.5%)	0.700(70.0%)	0.042(4.2%)	0.336(33.6%)	0.040(4.0%)	0.689(68.9%)	0.035(3.5%)	0.314(31.4%)	0.041(4.1%)	0.673(67.3%)
	30	0.044(4.4%)	0.574(57.4%)	0.036(3.6%)	0.785(78.5%)	0.041(4.1%)	0.555(55.5%)	0.041(4.1%)	0.783(78.3%)	0.040(4.0%)	0.537(53.7%)	0.038(3.8%)	0.779(77.9%)
$p=4$	50	0.037(3.7%)	0.689(68.9%)	0.039(3.9%)	0.905(90.5%)	0.038(3.8%)	0.672(67.2%)	0.044(4.4%)	0.892(89.2%)	0.037(3.7%)	0.663(66.3%)	0.038(3.8%)	0.864(86.4%)
	70	0.040(4.0%)	0.799(79.9%)	0.041(4.1%)	0.956(95.6%)	0.037(3.7%)	0.793(79.3%)	0.042(4.2%)	0.936(93.6%)	0.042(4.2%)	0.792(79.2%)	0.040(4.0%)	0.910(91.0%)
	100	0.039(3.9%)	0.901(90.1%)	0.039(3.9%)	0.974(97.4%)	0.043(4.3%)	0.891(89.1%)	0.041(4.1%)	0.972(97.2%)	0.038(3.8%)	0.885(88.5%)	0.043(4.3%)	0.958(95.8%)
	10	0.039(3.9%)	0.367(36.7%)	0.031(3.1%)	0.682(68.2%)	0.043(4.3%)	0.354(35.4%)	0.038(3.8%)	0.657(65.7%)	0.036(3.6%)	0.327(32.7%)	0.044(4.4%)	0.653(65.3%)
	30	0.042(4.2%)	0.615(61.5%)	0.034(3.4%)	0.786(78.6%)	0.037(3.7%)	0.590(59.0%)	0.042(4.2%)	0.762(76.2%)	0.035(3.5%)	0.564(56.4%)	0.037(3.7%)	0.759(75.9%)
$p=6$	50	0.044(4.4%)	0.725(72.5%)	0.039(3.9%)	0.923(92.3%)	0.043(4.3%)	0.698(69.8%)	0.045(4.5%)	0.879(87.9%)	0.041(4.1%)	0.679(67.9%)	0.039(3.9%)	0.877(87.7%)
	70	0.039(3.9%)	0.814(81.4%)	0.043(4.3%)	0.957(95.7%)	0.039(3.9%)	0.811(81.1%)	0.037(3.7%)	0.947(94.7%)	0.044(4.4%)	0.807(80.7%)	0.036(3.6%)	0.942(94.2%)
	100	0.038(3.8%)	0.945(94.5%)	0.036(3.6%)	0.986(98.6%)	0.042(4.2%)	0.942(94.2%)	0.041(4.1%)	0.979(979%)	0.037(3.7%)	0.938(93.8%)	0.042(4.2%)	0.971(97.1%)

Table 4.3 shows that the Type I errors were around 0.04 for both the proposed nonparametric kernel-based and parametric one-way MANOVA tests for balanced design when the underlying distribution of the data is multivariate exponential. As expected, the results generally performed better for the proposed nonparametric kernel-based tests than the parametric MANOVA tests. In addition, the power of both parametric and nonparametric tests increased as the sample size and dimension increased. Also, it was observed that as the correlation in the variance-covariance matrices increased, the power of both tests slightly decreased.

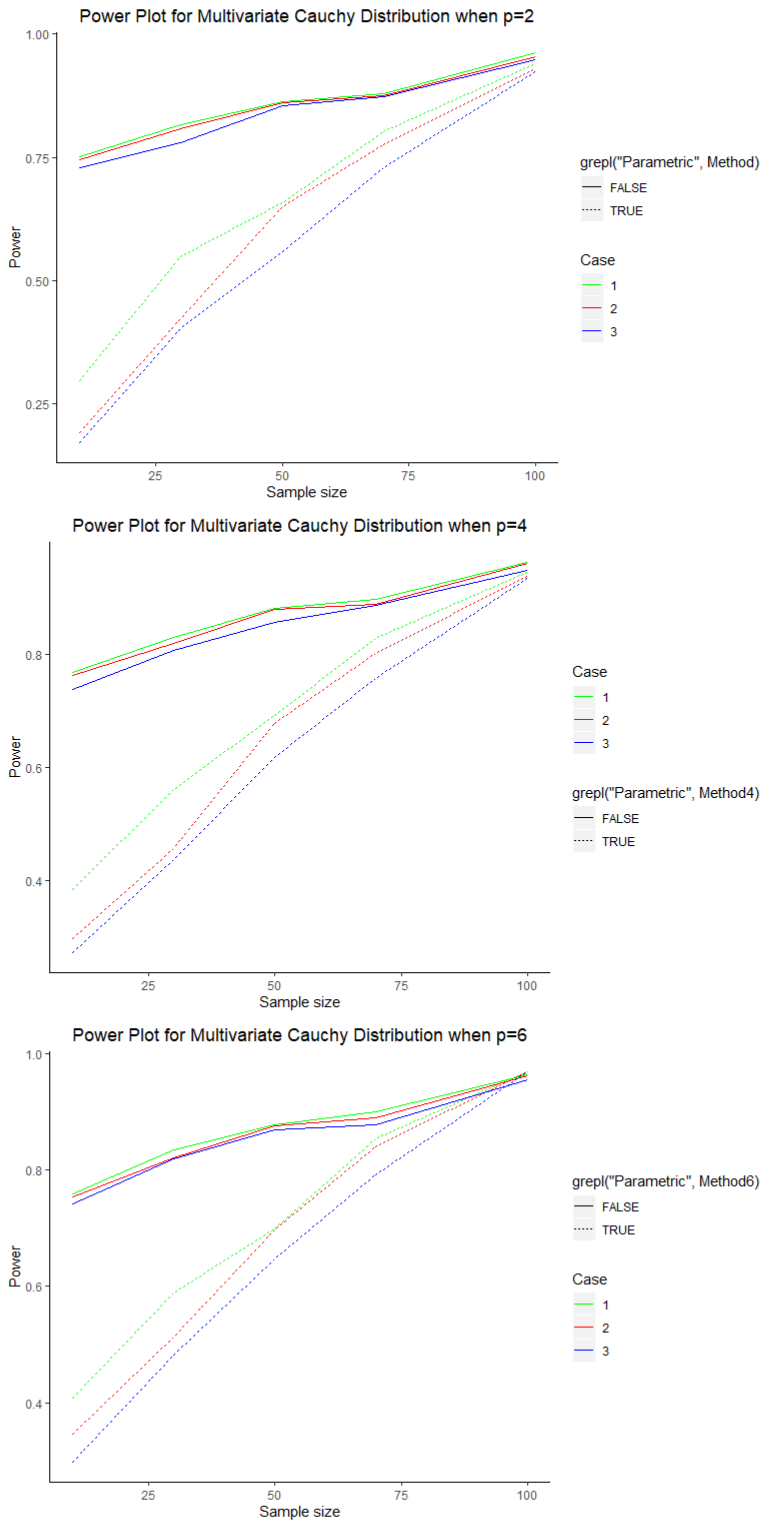

Figure 4.2. Power vs. Sample Size for the One-Way MANOVA for Multivariate Cauchy Distribution

In Table 4.3, when $p = 2$, the parametric one-way MANOVA test only reached a power of 87% when the sample size reached 100 for Case I, Case II, and Case III. However, the proposed nonparametric kernel-based one-way MANOVA reached a power of 97%, 96%, and 95% for Case I, Case II, and Case III, respectively. When $p = 4$, the parametric one-way MANOVA test only reached a power of 90%, 89%, and 88% when the sample size reached 100 for Case I, Case II, and Case III, respectively. However, the proposed nonparametric kernel-based one-way MANOVA reached a power of around 97% for Case I, and Case II, and 96% for Case III. When $p = 6$, the parametric one-way MANOVA test only reached a power of 95% for Case I, and 94% for Case II and Case III when the sample size reached 100. However, the proposed nonparametric kernel-based one-way MANOVA reached a power of around 99%, 98%, and 97% for Case I, Case II, and Case III, respectively.

Figure 4.3 shows the power plots for the one-way MANOVA when the underlying distribution of the data is multivariate exponential for $p = 2, 4, 6$. The proposed nonparametric kernel-based tests are represented as solid lines while the parametric tests represented as dashed lines in Figure 4.3. In general, the proposed nonparametric kernel-based one-way MANOVA test had a higher power than the traditional parametric approach when the data come from a multivariate exponential distribution. Additionally, as the sample size increased, the nonparametric test had a higher power for smaller sample sizes than the parametric test. Also, the the proposed nonparametric kernel-based test reached a higher power faster than the parametric test. It can also be observed that the lower the correlation in the variance-covariance matrix, the higher the power.

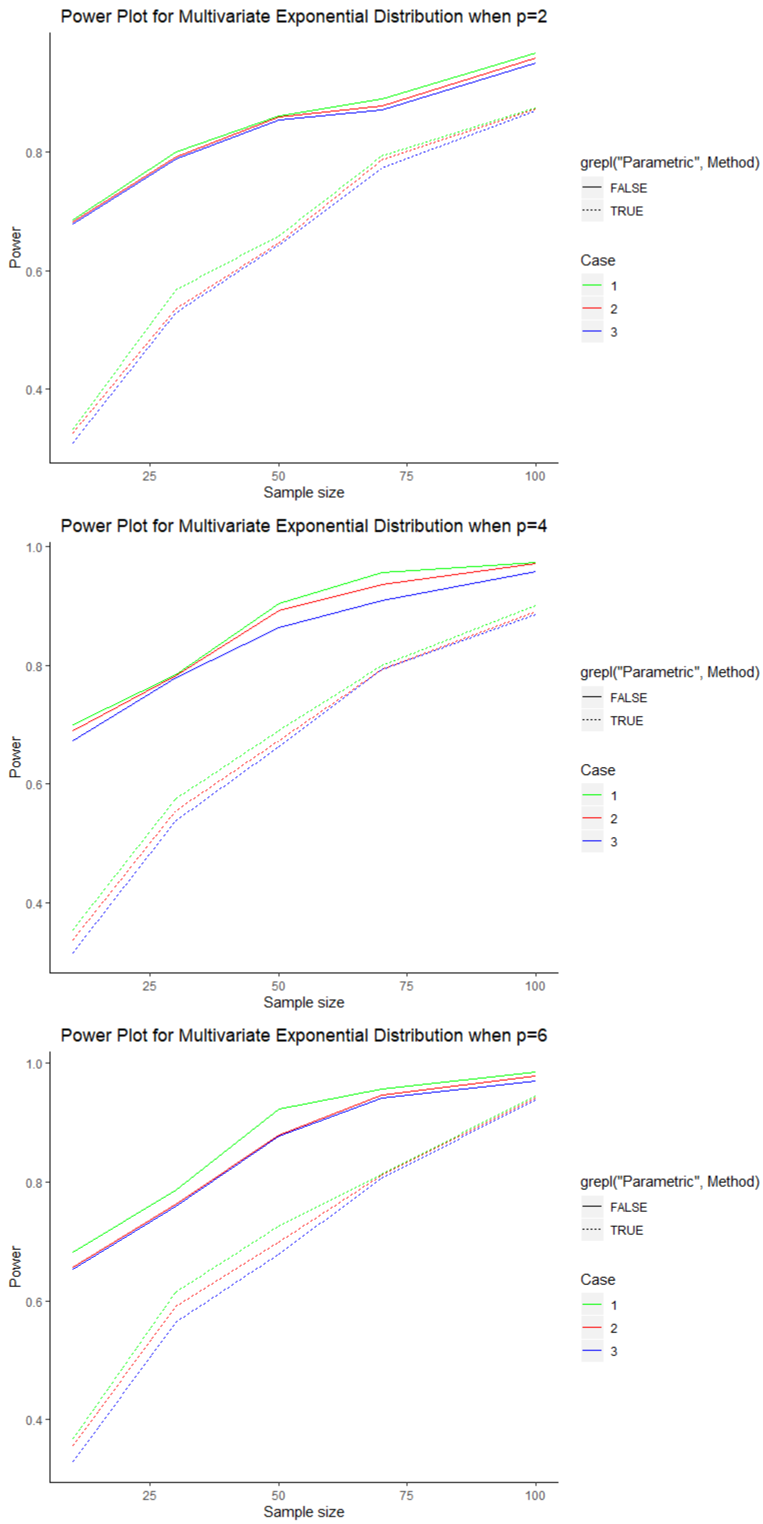

Figure 4.3. Power vs. Sample Size for the One-Way MANOVA for Multivariate Exponential Distribution

Simulation Study for Evaluating Type I Error and Power for the Interaction Effect in the Nonparametric Kernel-Based Two-Way MANOVA

In this section, the performance of the Type I error and power of the interaction effect of the proposed nonparametric kernel-based two-way MANOVA is evaluated and compared to the performance of the Type I error and power of the interaction effect of the traditional parametric two-way MANOVA.

Q6 How do Type I error rate and power of the proposed kernel-based test for the interaction term in the two-way MANOVA behave compared to the interaction term in the parametric one-way MANOVA?

Answer to Question 6. In order to answer this question, multiple simulations were conducted using different dimensions, sample sizes, and variance-covariance matrices representing no correlation, low correlation, and high correlation among variables. The simulation study was conducting using the following distributions: multivariate normal, multivariate Cauchy, and multivariate exponential distribution. Conditions for this simulation study are shown in Tables 3.6 - 3.12 in Chapter III. In this simulation study, the Type I error and power of the interaction effect of the proposed nonparametric kernel-based two-way MANOVA are evaluated and compared to the interaction effect of the traditional parametric two-way MANOVA. The simulation results are provided in Table 4.3 - Table 4.6.

Table 4.4 shows that the Type I errors are around 0.05 for the interaction effect of both the proposed nonparametric kernel-based tests and the parametric two-way MANOVA tests for balanced design when the underlying distribution of the data is multivariate normal. As expected, the results generally performed better for the parametric MANOVA tests than the proposed nonparametric kernel-based tests. In addition, the power of both parametric and nonparametric tests increased as the sample size and dimension increased. Also, it was observed that as the correlation in the variance-covariance matrices increased, the power of both tests slightly decreased.

Table 4.4.

Type I Error and Power for the Kernel-Based vs. Parametric Tests: Interaction Effect in Two-Way MANOVA for Multivariate Normal Distribution

	n_i	Case I				Case II				Case III			
		Parametric		Nonparametric		Parametric		Nonparametric		Parametric		Nonparametric	
		Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power
	10	0.052(5.2%)	0.895(89.5%)	0.046(4.6%)	0.247(24.7%)	0.041(4.1%)	0.879(87.9%)	0.051(5.1%)	0.239(23.9%)	0.049(4.9%)	0.854(85.4%)	0.048(4.8%)	0.218(21.9%)
	30	0.048(4.8%)	0.998(99.8%)	0.054(5.4%)	0.464(46.4%)	0.044(4.4%)	0.996(99.6%)	0.047(4.7%)	0.444(44.4%)	0.046(4.6%)	0.987(98.7%)	0.049(4.9%)	0.427(42.7%)
$p = 2$	50	0.050(5.0%)	1.000(100.0%)	0.051(5.1%)	0.672(67.2%)	0.053(5.3%)	1.000(100.0%)	0.048(4.8%)	0.665(66.5%)	0.051(5.1%)	0.999(99.9%)	0.053(5.3%)	0.657(65.7%)
	70	0.044(4.4%)	1.000(100.0%)	0.051(5.1%)	0.716(71.6%)	0.047(4.7%)	1.000(100.0%)	0.052(5.2%)	0.695(69.5%)	0.051(5.1%)	1.000(100.0%)	0.056(5.6%)	0.684(68.4%)
	100	0.045(4.5%)	1.000(100.0%)	0.042(4.2%)	0.891(89.1%)	0.052(5.2%)	1.000(100.0%)	0.051(5.1%)	0.886(88.6%)	0.048(4.8%)	1.000(100.0%)	0.047(4.7%)	0.878(87.8%)
	10	0.046(4.6%)	0.906(90.6%)	0.051(5.1%)	0.285(28.5%)	0.051(5.1%)	0.893(89.3%)	0.052(5.2%)	0.254(25.4%)	0.042(4.2%)	0.881(88.1%)	0.055(5.5%)	0.222(22.2%)
	30	0.047(4.7%)	0.999(99.9%)	0.047(4.7%)	0.512(51.2%)	0.054(5.4%)	0.997(99.7%)	0.046(4.6%)	0.498(49.8%)	0.049(4.9%)	0.993(99.3%)	0.048(4.8%)	0.481(48.1%)
$p = 4$	50	0.052(5.2%)	1.000(100.0%)	0.049(4.9%)	0.696(69.6%)	0.043(4.3%)	1.000(100.0%)	0.048(4.8%)	0.683(68.3%)	0.054(5.4%)	0.999(99.9%)	0.047(4.7%)	0.679(67.9%)
	70	0.051(5.1%)	1.000(100.0%)	0.052(5.2%)	0.731(73.1%)	0.046(4.6%)	1.000(100.0%)	0.053(5.3%)	0.729(72.9%)	0.048(4.8%)	1.000(100.0%)	0.052(5.2%)	0.718(71.8%)
	100	0.045(4.5%)	1.000(100.0%)	0.045(4.5%)	0.914(91.4%)	0.048(4.8%)	1.000(100.0%)	0.051(5.1%)	0.908(90.8%)	0.052(5.2%)	1.000(100.0%)	0.049(4.9%)	0.894(89.4%)
	10	0.051(5.1%)	0.914(90.6%)	0.047(4.7%)	0.300(30.0%)	0.047(4.7%)	0.909(90.9%)	0.052(5.2%)	0.295(30.0%)	0.053(5.3%)	0.897(98.7%)	0.051(5.1%)	0.267(26.7%)
	30	0.051(5.1%)	0.999(99.9%)	0.053(5.3%)	0.537(53.7%)	0.044(4.4%)	0.998(99.8%)	0.046(4.6%)	0.526(53.7%)	0.048(4.8%)	0.996(99.6%)	0.046(4.6%)	0.419(41.9%)
$p = 6$	50	0.046(4.6%)	1.000(100.0%)	0.051(5.1%)	0.711(7.11%)	0.052(5.2%)	1.000(100.0%)	0.048(4.8%)	0.699(7.11%)	0.052(5.2%)	1.000(100.0%)	0.045(4.5%)	0.651(65.1%)
	70	0.044(4.4%)	1.000(100.0%)	0.046(4.6%)	0.765(76.5%)	0.048(4.8%)	1.000(100.0%)	0.053(5.3%)	0.752(76.5%)	0.046(4.6%)	1.000(100.0%)	0.047(4.7%)	0.749(74.9%)
	100	0.052(5.2%)	1.000(100.0%)	0.048(4.8%)	0.934(93.4%)	0.046(4.6%)	1.000(100.0%)	0.051(5.1%)	0.929(93.4%)	0.045(4.5%)	1.000(100.0%)	0.053(5.3%)	0.918(91.8%)

When $p = 2$ and the sample size researched 50, the parametric test for the interaction effect in the two-way MANOVA reached a power of 100% for Case I, Case II, and Case III. However, the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA reached a power of around 89% for Case I and Case II and 88% for Case III when the sample size reached 100. When $p = 4$, the parametric test for the interaction effect in the two-way MANOVA almost reached a power of 100% when the sample size reached 30 for Case I, Case II, and Case III. On the other hand, the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA only reached a power of around 91% for Case I and Case II, and 89% for Case III. When $p = 6$, the parametric test for the interaction effect in the two-way MANOVA reached a power of 100% or close to 100% when the sample size reached 30 for Case I, Case II, and Case III. However, the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA only reached a power of around 93% for Case I and Case II, and 93% for Case III when the sample size reaches 100.

Figure 4.4 shows the power plots for the interaction effect of the two-way MANOVA when the underlying distribution of the data is multivariate normal distribution for $p = 2, 4, 6$. The proposed nonparametric kernel-based tests are represented as solid lines while the parametric tests are represented as dashed lines in Figure 4.4. In general, the traditional parametric interaction effect test in the two-way MANOVA had a higher power than the proposed nonparametric approach when the data come from a multivariate normal distribution. Additionally, as the sample size increased, the parametric test reached a higher power faster than the proposed nonparametric kernel-based test. It can also be observed that the lower the correlation in the variance-covariance matrix, the higher the power.

Table 4.5.

Type I Error and Power for the Kernel-Based vs. Parametric Tests: Interaction Effect in Two-Way MANOVA for Multivariate Cauchy Distribution

	n_i	Case I				Case II				Case III			
		Parametric		Nonparametric		Parametric		Nonparametric		Parametric		Nonparametric	
		Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power
	10	0.016(1.6%)	0.249(24.9%)	0.013(1.3%)	0.612(61.2%)	0.022(2.2%)	0.224(22.9%)	0.018(1.8%)	0.608(60.8%)	0.023(2.3%)	0.218(21.8%)	0.011(1.1%)	0.586(58.6%)
	30	0.019(1.9%)	0.264(26.4%)	0.027(2.7%)	0.927(92.7%)	0.016(1.6%)	0.238(23.8%)	0.016(1.6%)	0.914(91.4%)	0.019(1.9%)	0.222(22.2%)	0.024(2.4%)	0.899(89.9%)
$p = 2$	50	0.017(1.7%)	0.378(37.8%)	0.026(2.6%)	0.956(95.6%)	0.015(1.5%)	0.363(36.3%)	0.022(2.2%)	0.949(94.9%)	0.013(1.3%)	0.356(35.6%)	0.016(1.6%)	0.936(93.6%)
	70	0.021(2.1%)	0.532(53.2%)	0.017(1.7%)	0.983(98.3%)	0.013(1.3%)	0.498(49.8%)	0.019(1.9%)	0.978(97.8%)	0.021(2.1%)	0.484(48.4%)	0.018(1.8%)	0.973(97.3%)
	100	0.014(1.4%)	0.611(61.1%)	0.024(2.4%)	0.999(99.9%)	0.019(1.9%)	0.581(58.1%)	0.023(2.3%)	0.999(99.9%)	0.017(1.7%)	0.576(57.6%)	0.023(2.3%)	0.998(94.8%)
	10	0.019(1.9%)	0.299(29.9%)	0.025(2.5%)	0.634(63.4%)	0.022(2.2%)	0.281(28.1%)	0.022(2.2%)	0.631(63.1%)	0.017(1.7%)	0.269(26.9%)	0.017(1.7%)	0.619(61.9%)
	30	0.014(1.4%)	0.314(31.4%)	0.016(1.6%)	0.936(93.6%)	0.014(1.4%)	0.299(29.9%)	0.019(1.9%)	0.929(92.9%)	0.013(1.3%)	0.283(28.3%)	0.026(2.6%)	0.896(89.6%)
$p = 4$	50	0.022(2.2%)	0.406(40.6%)	0.031(3.1%)	0.967(96.7%)	0.018(1.8%)	0.384(38.4%)	0.014(1.4%)	0.958(95.8%)	0.023(2.3%)	0.351(35.1%)	0.018(1.8%)	0.943(94.3%)
	70	0.021(2.1%)	0.612(61.2%)	0.019(1.9%)	0.999(99.9%)	0.014(1.4%)	0.608(60.8%)	0.023(2.3%)	0.998(99.8%)	0.019(1.9%)	0.583(58.3%)	0.016(1.6%)	0.992(99.2%)
	100	0.016(1.6%)	0.735(73.5%)	0.017(1.7%)	1.000(100.0%)	0.016(1.6%)	0.728(72.8%)	0.023(2.3%)	0.999(99.9%)	0.021(2.1%)	0.716(71.6%)	0.012(1.2%)	0.997(99.7%)
	10	0.021(2.1%)	0.311(31.1%)	0.021(2.1%)	0.657(65.7%)	0.021(2.1%)	0.308(30.8%)	0.013(1.3%)	0.643(64.3%)	0.013(1.3%)	0.295(29.5%)	0.022(2.2%)	0.626(62.6%)
	30	0.022(2.1%)	0.352(35.2%)	0.019(1.9%)	0.952(95.2%)	0.019(1.9%)	0.339(33.9%)	0.016(1.6%)	0.947(94.7%)	0.022(2.2%)	0.317(31.7%)	0.023(2.3%)	0.933(93.3%)
$p = 6$	50	0.029(2.9%)	0.427(42.7%)	0.016(1.6%)	0.981(98.1%)	0.023(2.3%)	0.414(41.4%)	0.023(2.3%)	0.979(97.9%)	0.024(2.4%)	0.399(39.9%)	0.015(1.5%)	0.968(96.8%)
	70	0.013(1.3%)	0.649(64.9%)	0.024(2.4%)	1.000(100.0%)	0.013(1.3%)	0.633(63.3%)	0.019(1.9%)	0.999(99.9%)	0.016(1.6%)	0.627(62.7%)	0.012(1.2%)	0.998(99.8%)
	100	0.014(1.4%)	0.777(77.7%)	0.021(2.1%)	1.000(100.0%)	0.017(1.7%)	0.764(76.4%)	0.021(2.1%)	1.000(100.0%)	0.018(1.8%)	0.755(75.5%)	0.013(1.3%)	0.999(99.9%)

Table 4.5 shows that the Type I errors are around 0.02 for the interaction effect of both the proposed nonparametric kernel-based and the parametric two-way MANOVA tests for balanced design when the underlying distribution of the data is multivariate Cauchy. As expected, the results generally performed better for the proposed nonparametric kernel-based tests than the parametric MANOVA tests. In addition, the power of both parametric and nonparametric tests increased as the sample size and dimension increased. Also, it is observed that as the correlation in the variance-covariance matrices increased, the power of both tests slightly decreased.

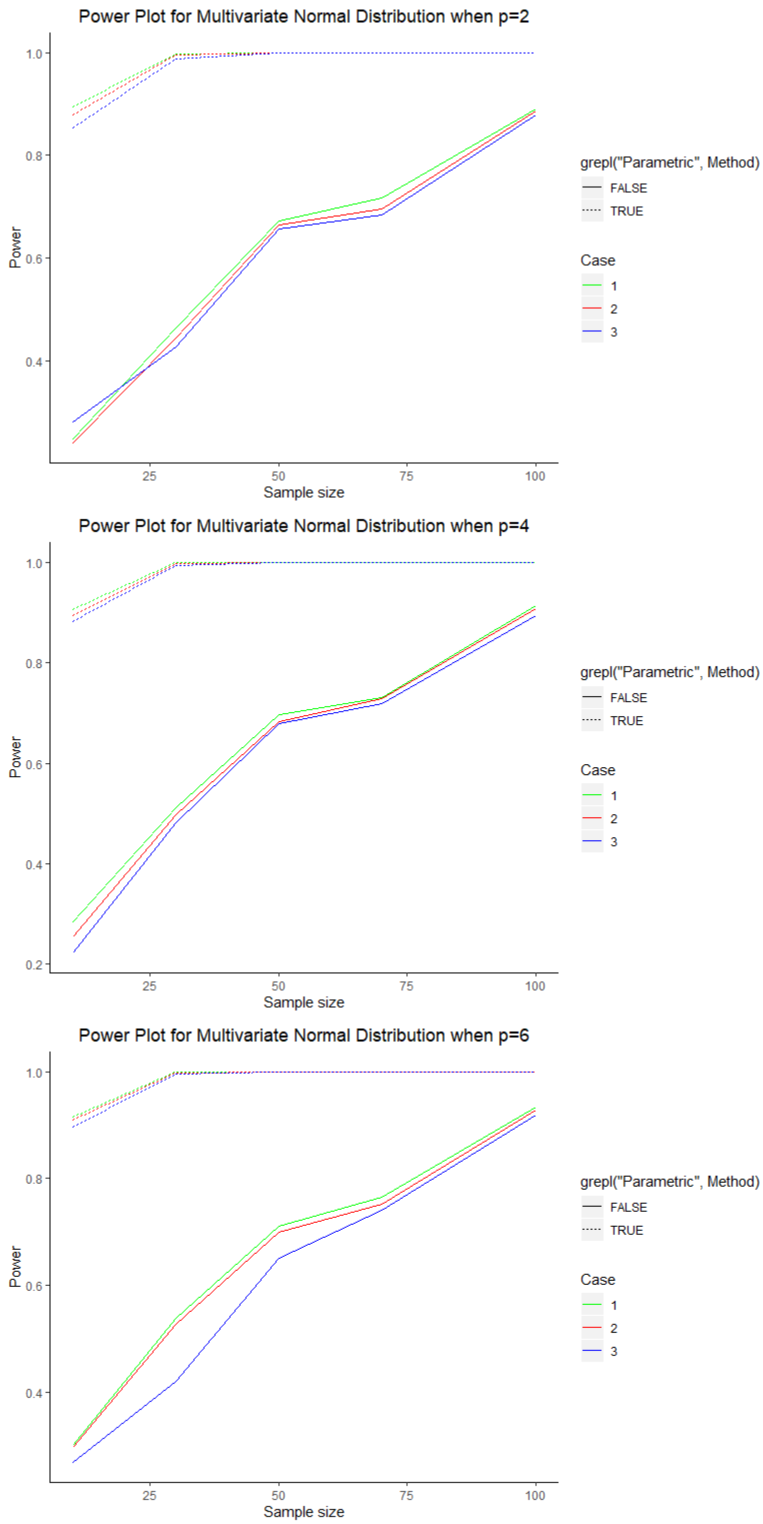

Figure 4.4. Power vs. Sample Size for the Interaction Effect of the Two-Way MANOVA for Multivariate Normal Distribution

In Table 4.5, when $p = 2$, the parametric test for the interaction effect in the two-way MANOVA only reached a power of 72%, 71%, and 69% when the sample size reached 100 for Case I, Case II, and Case III, respectively. However, theproposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA reached a power of around 99%, 98%, and 97% for Case I, Case II, and Case III, respectively. When $p = 4$, the parametric test for the interaction effect in the two-way MANOVA only reaches a power of 76% for Case I and 75% for Case II and Case III when the sample size reached 100. However, the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA reached a power of around 100% for Case I, and Case II, and Case III. When $p = 6$, the parametric test for the interaction effect in the two-way MANOVA only reached a power of 78% for Case I, and 76% for Case II and Case III when the sample size reached 100. However, the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA reached a power of almost 100% for Case I, Case II, and Case III when the sample size reached 70.

Figure 4.5 shows the power plots for the interaction effect of the two-way MANOVA when the underlying distribution of the data is multivariate Cauchy for $p = 2,4,6$. The proposed nonparametric kernel-based tests are represented as solid lines while the parametric tests are represented as dashed lines in Figure 4.5. In general, the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA had a higher power than the traditional parametric approach when the data come from a multivariate Cauchy distribution. Additionally, as the sample size increased, the nonparametric test had a higher power for smaller sample sizes than the parametric test. Also, the proposed nonparametric kernel-based test reached a higher power faster than the parametric test. It can also be seen that the lower the correlation in the variance-covariance matrix, the higher the power.

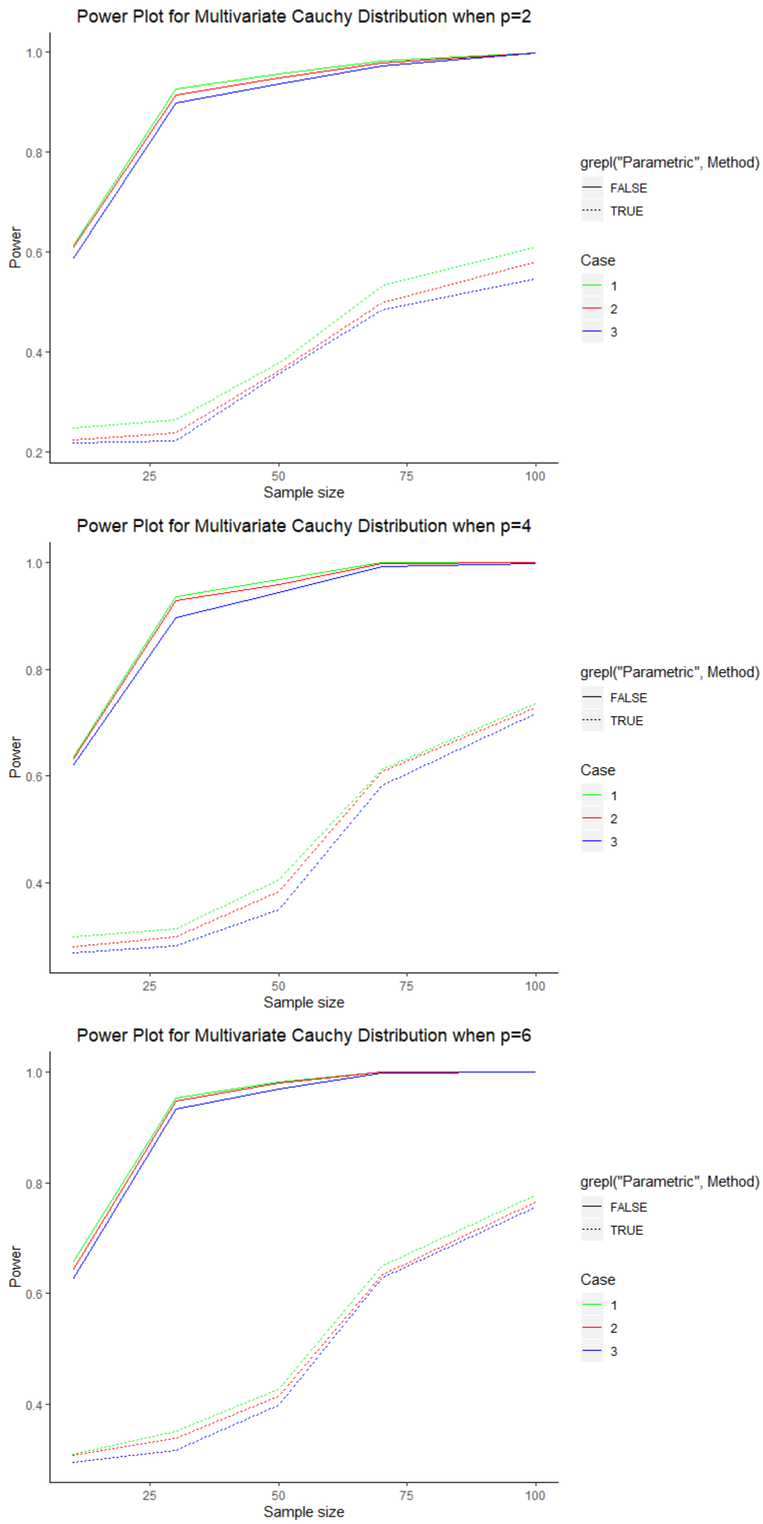

Figure 4.5. Power vs. Sample Size for the Interaction Effect of the Two-Way MANOVA for Multivariate Cauchy Distribution

Table 4.6.

Type I Error and Power for the Kernel-Based vs. Parametric Tests: Interaction Effect in Two-Way MANOVA for Multivariate Exponential Distribution

	n_i	Case I				Case II				Case III			
		Parametric		Nonparametric		Parametric		Nonparametric		Parametric		Nonparametric	
		Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power	Type I Error	Power
	10	0.042(4.2%)	0.285(38.5%)	0.035(3.5%)	0.655(65.5%)	0.041(4.1%)	0.262(26.2%)	0.038(3.8%)	0.642(64.2%)	0.039(3.9%)	0.248(24.8%)	0.036(3.6%)	0.636(63.6%)
	30	0.034(3.4%)	0.348(34.8%)	0.041(4.1%)	0.872(87.2%)	0.032(3.2%)	0.335(33.5%)	0.041(4.1%)	0.867(86.7%)	0.040(4.0%)	0.316(31.6%)	0.038(3.8%)	0.865(86.5%)
$p=2$	50	0.039(3.9%)	0.502(50.2%)	0.037(3.7%)	0.904(90.4%)	0.037(3.7%)	0.488(48.8%)	0.043(4.3%)	0.883(88.3%)	0.037(3.7%)	0.472(47.2%)	0.041(4.1%)	0.876(87.6%)
	70	0.035(3.5%)	0.642(64.2%)	0.039(3.9%)	0.959(95.9%)	0.031(3.1%)	0.631(63.1%)	0.041(4.1%)	0.945(94.5%)	0.044(4.4%)	0.619(61.9%)	0.034(3.4%)	0.941(94.1%)
	100	0.041(4.1%)	0.719(71.9%)	0.042(4.2%)	0.987(98.7%)	0.040(4.0%)	0.708(70.8%)	0.033(3.3%)	0.979(97.9%)	0.041(4.1%)	0.686(68.6%)	0.044(4.4%)	0.977(97.7%)
	10	0.036(3.6%)	0.352(35.2%)	0.032(3.2%)	0.675(67.5%)	0.035(3.5%)	0.347(34.7%)	0.042(4.2%)	0.668(66.8%)	0.039(3.9%)	0.319(31.9%)	0.041(4.1%)	0.644(64.4%)
	30	0.037(3.7%)	0.371(37.1%)	0.036(3.6%)	0.899(89.9%)	0.038(3.8%)	0.367(36.7%)	0.034(3.4%)	0.882(88.2%)	0.040(4.0%)	0.352(35.2%)	0.036(3.6%)	0.869(86.9%)
$p=4$	50	0.041(4.1%)	0.536(53.6%)	0.039(3.9%)	0.918(91.8%)	0.040(4.0%)	0.522(52.2%)	0.041(4.1%)	0.911(91.1%)	0.037(3.7%)	0.518(51.8%)	0.037(3.7%)	0.907(90.7%)
	70	0.042(4.2%)	0.666(66.6%)	0.035(3.5%)	0.974(97.4%)	0.041(4.1%)	0.648(64.8%)	0.037(3.7%)	0.970(97.0%)	0.034(3.4%)	0.636(63.6%)	0.036(3.6%)	0.968(96.8%)
	100	0.038(3.8%)	0.756(75.6%)	0.041(4.1%)	0.999(99.9%)	0.036(3.6%)	0.751(75.1%)	0.036(3.6%)	0.999(99.9%)	0.041(4.1%)	0.746(74.6%)	0.042(4.2%)	0.998(99.8%)
	10	0.038(3.8%)	0.384(38.4%)	0.029(2.9%)	0.691(69.1%)	0.041(4.1%)	0.371(37.1%)	0.042(4.2%)	0.682(68.2%)	0.035(3.5%)	0.353(35.3%)	0.042(4.2%)	0.666(66.6%)
	30	0.042(4.2%)	0.403(40.3%)	0.031(3.1%)	0.911(91.1%)	0.042(4.2%)	0.392(39.2%)	0.036(3.6%)	0.895(89.5%)	0.037(3.7%)	0.378(37.8%)	0.034(3.4%)	0.885(88.5%)
$p=6$	50	0.038(3.8%)	0.572(57.2%)	0.037(3.7%)	0.937(93.7%)	0.037(3.7%)	0.549(54.9%)	0.040(4.0%)	0.927(92.7%)	0.037(3.7%)	0.536(53.6%)	0.033(3.3%)	0.916(91.6%)
	70	0.041(4.1%)	0.713(71.3%)	0.039(3.9%)	0.993(99.3%)	0.036(3.6%)	0.695(69.5%)	0.039(3.9%)	0.984(98.4%)	0.043(4.3%)	0.683(68.3%)	0.037(3.7%)	0.982(98.2%)
	100	0.036(3.6%)	0.772(77.2%)	0.042(4.2%)	1.000(100.0%)	0.038(3.8%)	0.763(76.3%)	0.039(3.9%)	1.000(100.0%)	0.038(3.8%)	0.751(75.1%)	0.044(4.4%)	0.999(99.9%)

Table 4.6 shows that the actual Type I errors are around 0.04 for the interaction effect of both the proposed nonparametric kernel-based and parametric two-way MANOVA tests for balanced design when the underlying distribution of the data is multivariate exponential. As expected, the results generally performed better for the proposed nonparametric kernel-based tests than the parametric MANOVA tests. In addition, the power of both parametric and nonparametric tests increased as the sample size increased and dimension increased. Also, it was observed that as the correlation in the variance-covariance matrices increased, the power of both tests slightly decreased.

When $p = 2$, the parametric test for the interaction effect in the two-way MANOVA only reached a power of 72%, 71%, and 69% when the sample size reached 100 for Case I, Case II, and Case III, respectively. However, the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA reached a power of around 99% for Case I and around 98% for Case II and Case III. When $p = 4$, the parametric test for the interaction effect in the two-way MANOVA only reached a power of 76% for Case I and 75% for Case II and Case III when the sample size reached 100. However, the pproposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA reached a power of around 99% for Case I, and Case II, and Case III. When $p = 6$, the parametric test for the interaction effect in the two-way MANOVA only reached a power of 77%, 76%, and 75% for Case I, Case II, and Case III, respectively, when the sample size reached 100. However,proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA reached a power of almost 100% for Case I, Case II, and Case III.

Figure 4.6 shows the power plots for the interaction effect of the two-way MANOVA when the underlying distribution of the data is multivariate exponential for $p = 2, 4, 6$. The proposed nonparametric kernel-based tests are represented as solid lines while the parametric tests are represented as dashed lines in Figure 4.6. In general, the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA had a higher power than the traditional parametric approach when the data come from a multivariate exponential distribution. Additionally, as the sample size increased, the nonparametric test had a higher power for smaller sample sizes than the parametric test. Also, the the proposed nonparametric kernel-based test reached a higher power faster than the parametric test. Moreover, it can be seen that the lower the correlation in the variance-covariance matrix, the higher the power.

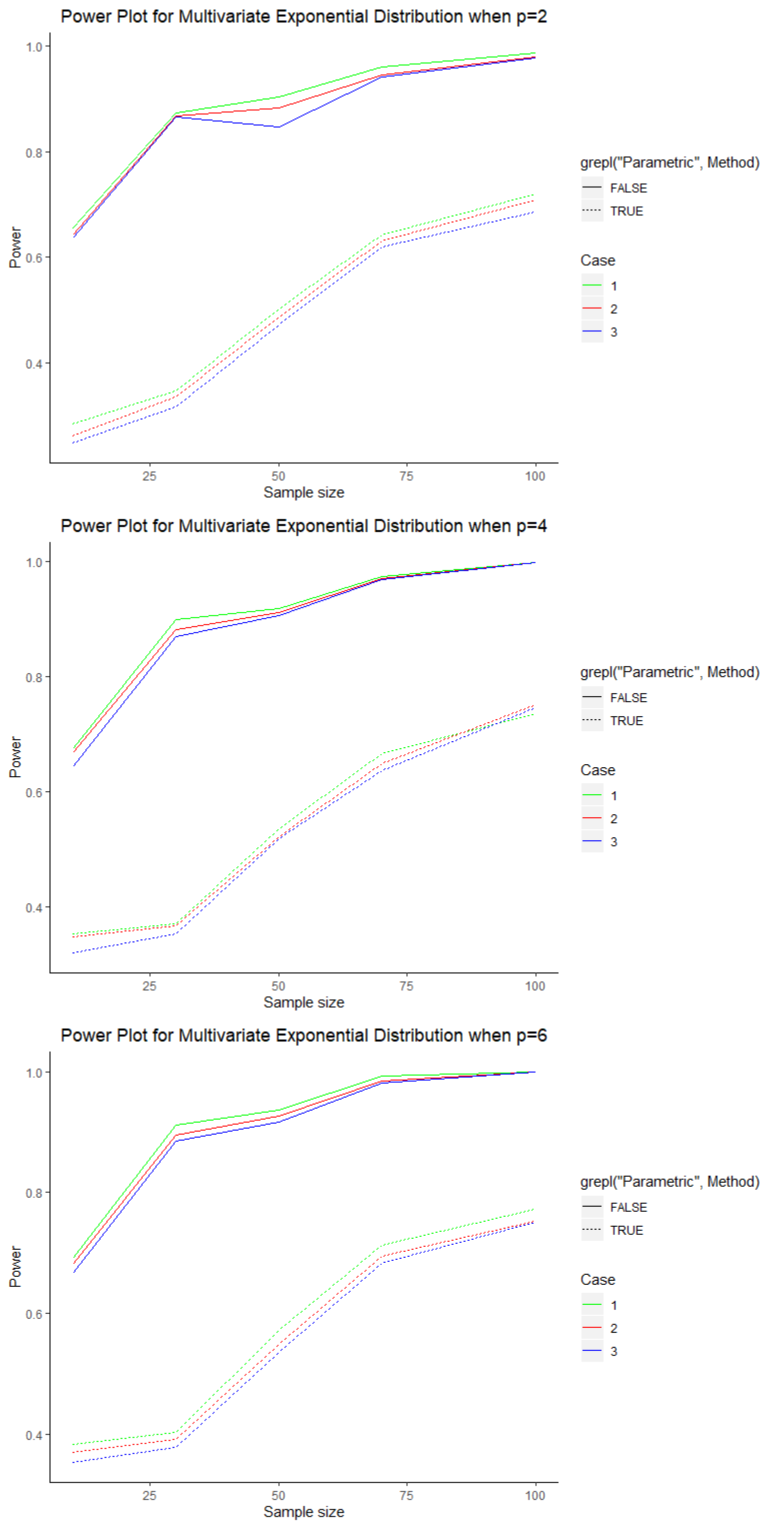

Figure 4.6. Power vs. Sample Size vs. Sample for the Interaction Effect of the Two-Way MANOVA for Multivariate Exponential Distribution

Summary of Simulation Study for the Type I Error and Power of the Proposed Nonparametric Kernel-Based MANOVA

Testing hypotheses for MANOVA is challenging when the data come from non-normal or non-linear distributions. The proposed nonparametric kernel-based MANOVA tests offer a new way to approach this challenge. In order to evaluate the performance of the proposed nonparametric kernel-based MANOVA tests, different dimensions, sample sizes, and variance-covariance matrices were considered.

Kernel-based tests were developed in this dissertation for testing the hypotheses for main and interaction effects for one and two-way MANOVA. Each test was compared to the traditional parametric MANOVA counterpart, Wilks Lambda. To evaluate the performance of the proposed nonparametric kernel-based one-way MANOVA, the Type I error and power were estimated via simulation and compared to the Type I error and power of the parametric one-way MANOVA. Then, to evaluate the performance of the proposed nonparametric kernel-based test interaction effect in the two-way MANOVA, the Type I error and power were estimated via simulation and compared to the Type I error and power of the interaction effect of the parametric two-way MANOVA.

The results of the simulation study showed that the proposed nonparametric kernel-based MANOVA tests performed better than the parametric MANOVA tests when the data are non-normal and non-linear. For example, the proposed nonparametric method for one-way MANOVA performed relatively well, i.e. power of 75% - 80%, with a sample size of 30 for each group when the data come from non-normal distribution. Also, the proposed nonparametric method for the interaction effect in the two-way MANOVA performed relatively well, i.e. power of 80%, with a sample size of 30 for each group when the data come from non-normal distribution. Additionally, the results from the simulation study showed that the power of the proposed nonparametric kernel-based MANOVA tests increased with sample size and dimension as expected. It was also observed that there was a slight decrease in power with higher correlation in the variance-covariance matrices. This could be due to the fact that when the data are more correlated, the statistical model tends to be more complicated, making statistical tests less powerful. Overall, when applied to non-normal data,the proposed nonparametric kernel-based MANOVA test performed as well as the the parametric MANOVA test for normal data regarding power.

The results of the simulation study showed that the proposed nonparametric kernel-based MANOVA tests and the parametric MANOVA tests have a Type I error rate of around 5% when the the underlying distribution of the data is normal distribution and significance level is $\alpha = 0.05$. Also, the results of the simulation study showed that the proposed nonparametric kernel-based MANOVA tests and the parametric MANOVA tests have a Type I error rate of around 2% when the the underlying distribution of the data is Cauchy and the significance level is $\alpha = 0.05$. The results of the simulation study showed that the proposed nonparametric kernel-based MANOVA tests and the parametric MANOVA tests have a Type I error rate of around 4% when the the underlying distribution of the data is exponential and the significance level is $\alpha = 0.05$. It can be concluded that the proposed nonparametric kernel-based MANOVA tests for non-normal data performed as well as the the parametric MANOVA tests for normal data regarding Type I error rate.

Real Data Application

In this section, an analysis using the proposed nonpramateric kernel-based one-way MANOVA is applied to provide an example for researchers and practitioners on how to use the proposed method on real-world data. Image data is one of many uses that the proposed nonparametric method can applied to. More details are provided in the example showen below.

Breast Cancer Cells

Introduction

Breast cancer is one of the leading causes of death in cancer-related diseases. It affects more than 10% of women worldwide (Siegel, Miller, & Jemal, 2017). Studies have shown that early diagnosis and treatment can significantly prevent the disease's progression and reduce its morbidity rates (Smith, Cokkinides, & Eyre, 2005). Thus, according to the National Breast Cancer Foundation, women are recommended to perform a regular self-checking and to receive a routine ultrasound and mammography screenings (2015). Screenings provide histology images which are essential in the early diagnosis of breast cancer (Aresta et al., 2019). During the evaluation of these images, pathologists search for signs of cancer on microscopic portions of the tissue. This procedure allows pathologists to differentiate between the malignant tissues and the benign (non-malignant) tissue to determine the proper treatment (Aresta et al., 2019).

Purpose

Distinguishing between samples of normal, benign, and malignant breast tissues brings crucial changes in the treatment of patients (Aresta et al., 2019). For instance, benign lesions can usually be treated clinically without the need for a surgical intervention while malignant lesions are usually treated through surgical intervention (Aresta et al., 2019). Thus, the purpose of analyzing this data is to determine if there is a difference in normal, benign, and malignant tissues by using histological images.

Data Description

The image dataset used was obtained from the 2018 BACH: Grand Challenge on Breast Cancer Histology Images. There are three different groups of breast tissue: Normal, Benign, and Malignant. Each group contains tissue patches with information about the radius, texture perimeter, area, and smoothness. Each group include 100 breast tissues. Samples of the tissues obtained from various groups of breast cancer is shown in Figure 4.7:

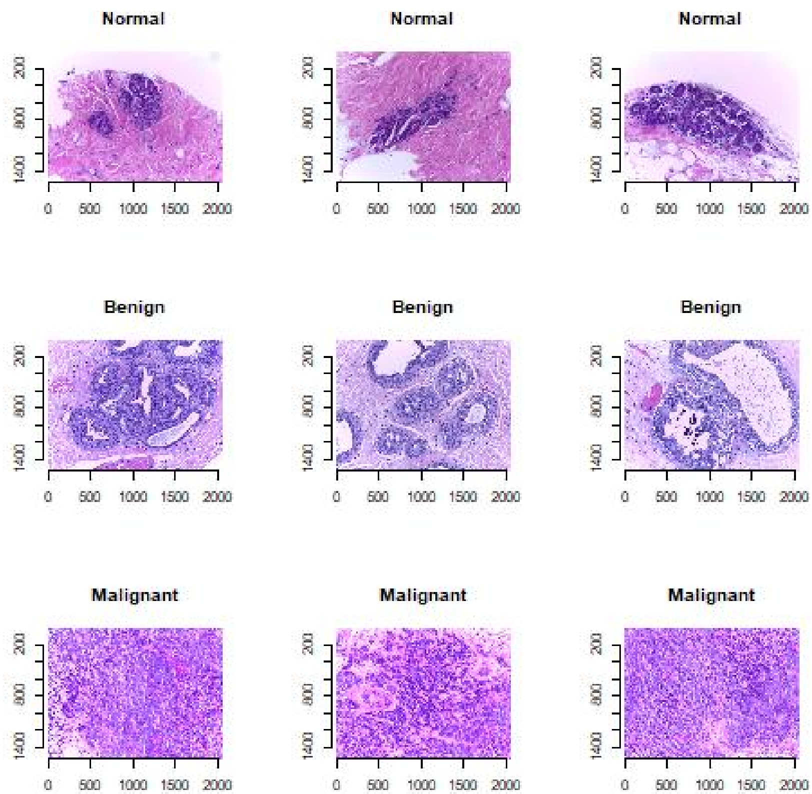

Figure 4.7. Sample of Breast Cancer Tissues from Different Groups

Data Analysis

The data were analyzed using the proposed one-way kernel-based MANOVA since the image dataset used does not have a normal distribution, thereby, violating a major assumption of the parametric MANOVA, which is the normality of the data.

The mean of each group was calculated using

$$\hat{\boldsymbol{Q}}_i = \frac{1}{n_i(n_i-1)|\boldsymbol{H}_i|^{\frac{1}{2}}} \sum\sum_{j_1 \neq j_2} \frac{1}{2}\left(\boldsymbol{X}_{ij_1} + \boldsymbol{X}_{ij_2}\right) K\left(\boldsymbol{H_i}^{\frac{-1}{2}}\left(\boldsymbol{X}_{ij_1} - \boldsymbol{X}_{ij_2}\right)\right). \tag{4.15}$$

Then, the $\boldsymbol{SSB}$ and $\boldsymbol{SSW}$ was calculated to obtain the Kernel$-\Lambda$ using

$$\Lambda_{\mathrm{k}} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SST}|} = \frac{|\boldsymbol{SSW}|}{|\boldsymbol{SSB}+\boldsymbol{SSW}|} = \frac{1}{1+\left|\frac{\boldsymbol{SSB}}{\boldsymbol{SSW}}\right|}. \tag{4.16}$$

Finally, the p-value of the F distribution approximation was obtained from

$$F = \frac{1-(\Lambda_{\mathrm{k}})^{1/b}}{(\Lambda_{\mathrm{k}})^{1}/b} \times \frac{df_2}{df_1} \dot{\sim} F(df_1, df_2), \tag{4.17}$$

where $df_1 = 10$ and $df_2 = 586$.

Results & Conclusion

The results obtained from performing the proposed one-way kernel-based MANOVA above show a rejection of the null hypothesis (p-value= 0.002) indicating a significant difference between breast tissues. Therefore, we can conclude that there is a significant difference between normal, benign, and malignant breast tissue patches using the radius, texture perimeter, area, and smoothness of lesions in tissues. Using the proposed nonparametric kernel-based method over the traditional parametric method gave the advantage of being able to compare image data that are non-normal. Thus, the proposed nonparametric kernel-based one-way MANOVA was more appropriate to use than the parametric MANOVA.

CHAPTER V

CONCLUSION

Nonparametric kernel-based multivariate analysis of variance (MANOVA) tests for non-normal multivariate data were developed in this dissertation. Kernel method was adopted within the MANOVA framework in order to provide a more efficient and appropriate tests than parametric MANOVA tests for the hypotheses testing procedure when the underlying distribution of the data is not normal.

The nonparametric MANOVA tests developed in this dissertation focused on the Wilks' lambda statistic, which follows asymptotically an F distribution. The centrality or non-centrality of the developed F distribution depends on whether the distribution of the kernel-based Wilks' lambda statistic is considered under the null or alternative hypothesis. Under the null hypothesis, this statistic follows an asymptotic central F distribution; however, it follows a non-central F distribution under the alternative hypothesis. Meaning, a central F ditribution has a mean of 0 while the a non-central F distribution has a shifted mean that is different than 0.

The objective of the proposed method presented in this dissertation was to help applied researchers and practitioners who design studies using real data to make valid decisions when detecting difference among groups with multivariate response. The contribution of the proposed nonparametric kernel-based MANOVA tests in this dissertation is that it allows the underlying distribution of the data to be non-normal and non-linear. The review of the literature on analysis of the variance and nonparametric approaches in Chapter II indicated that no test had been previously published that uses kernel density estimation (KDE) in the process of testing hypotheses of one-way and two-way MANOVA. This gap in the literature negatively affected the nonparametric statistics field in that the currently developed techniques were not developed to account for multivariate settings. In Chapter III, the proposed nonparametric tests were developed and proven theoretically. In Chapter IV, the performance of the proposed nonparametric tests was evaluated using type I error and power. Additionally, a real-data application was illustrated in Chapter IV using breast cancer image data.

The nonparametric kernel-based technique proposed in this dissertation is unique in that it is the first developed analysis of the variance technique that uses KDE, or kernel methods in general, to test hypotheses of the one and two-way MANOVA in order to detect difference between groups, main effects, or interaction effects for non-normal or non-linear multivariate data. In previously developed kernel-based techniques, multivariate settings were not taken into consideration which is not realistic given that we live in an era of “high-dimensional” data. The nonparametric kernel-based method proposed in this dissertation, however, allows researchers and practitioners to use multivariate data with non-normal distribution and still be able to test for difference between groups of factors. The developed nonparametric MANOVA method improves upon the other previously developed methods in that the procedure allows for testing main effects and interaction effects in the two-way layout for multivariate data while some other nonparametric tests, such as Multivariate Kruskal Wallis test, only allow the use of the one-way layout which presents a great limitation since most studies have multiple factors to test their group differences simultaneously as well as their interaction effect.

Thus, the nonparametric method developed and proposed in this dissertation has three major advantages over previously developed difference detecting methods. First, the current nonparametric kernel-based analysis of variance method considers only univariate settings while the proposed nonparametric kernel-based technique presented in this dissertation consider multivariate settings which allows for high-dimensional data analysis. Second, previously developed multivariate nonparametric approaches for testing group difference, i.e. Multivariate Kruskal Wallis, do not allow for testing more than one factor at a time. On the other hand, the nonparametric kernel-based method proposed in this dissertation allows two factors to be tested as well as their interaction effect. Finally, the proposed kernel-based method uses a nonparametric approach which allows for non-normal and non-linear data relaxing the normality assumption of the parametric methods which is very restrictive since modern data do not often follow a normal distribution.

The first research question addressed the process of constructing a kernel density estimator for multivariate data with a non-normal and non-linear underlying distribution. This question was answered theoretically in Chapter III, and the main steps used to answer this question, which lead to the results being applied to real data, are summarized in Chapter IV. Normal kernel function and Scott's rule of thumb bandwidth matrix were used in the development of the nonparametric kernel-based MANOVA proposed in this dissertation.

The second research question addressed the process of testing the hypotheses of the proposed nonparametric kernel-based one-way MANOVA. This question was answered theoretically in Chapter III and the main steps used to answer this question which lead to the results of applying the methods to real data are summarized in Chapter IV. The hypotheses regarding testing the group means were written using the kernel function estimate derived in Chapter III. Then, the asymptotic distribution of the proposed nonparametric kernel-based one-way MANOVA was driven to lead to evaluating the performance of the proposed method.

The third research question addressed the process of testing the hypotheses of the main effects of the proposed nonparametric kernel-based two-way MANOVA. This question was answered theoretically in Chapter III, and the main steps used to answer this question, which lead to the results of applying the methods to real data, are summarized in Chapter IV. The hypotheses regarding testing the group main effects were written using the kernel function estimate derived in Chapter III. Then, the asymptotic distribution of the main effects of the proposed nonparametric kernel-based two-way MANOVA was driven to evaluate the performance of the proposed method.

The fourth research question addressed the process of testing the hypotheses of the proposed nonparametric kernel-based for the interaction effect in the two-way MANOVA. This question was answered theoretically in Chapter III, and the main steps used to answer this question, which lead to the results of applying the methods to real data, are summarized in Chapter IV. The hypotheses regarding testing the group main effects were written using the kernel function estimate derived in Chapter III. Then, the asymptotic distribution of the proposed nonparametric kernel-based test for the interaction effect in the two-way MANOVA was driven to evaluate the performance of the proposed method.

The simulation study was used to answer the last two research questions regarding the performance of the proposed nonparametric methods using the type I error and power. Various conditions using different distributions, sample sizes, dimensions, and variance-covariance matrices were used in the simulation study showen in Chapter III. The results of the simulation study was showed in Chapter IV. The simulation conditions scheme was developed by accounting for a number of possible data scenarios to ensure that the results of the simulation study are generalizable to real data analysis; hence, the methodology could be adopted by researchers in different fields when using real data. Accounting for multiple possible cases resulted in increased generalizability of the methodology when evaluating the performance of the theoretically developed methods when dealing with real data. This helped provide helpful guidelines for researchers and practitioners regarding the situations that might arise while analyzing real data when the methods do not perform as well as theorized.

The performance of the proposed methodology was tested and evaluated in a simulation study as well as in a real-data application using a breast cancer image dataset consisting of three groups of breast tissues. The performance of the developed testing techniques and the conditions that would effect their performance in application were tabulated and discussed in Chapter IV.

The simulation study evaluated the power of the proposed nonparametric kernel-based MANOVA methods. It showed that the proposed nonparametric method for one-way and two-way MANOVA had a high power even with relatively small sample sizes when the data come from non-normal distributions. It was also shown that the proposed nonparametric kernel-based MANOVA tests perform as well as the the parametric MANOVA tests in the case of normal data. The results of the simulation study showed that the proposed nonparametric kernel-based MANOVA tests performed better when the data has non-Gaussian, in terms of power, as expected. Generally, statistical tests with high power have a higher probability and capability of detecting the difference between groups, if any exists. Thus, in practice, when the normality assumption of the parametric MANOVA is violated, it is best to use the proposed nonparametric kernel-based MANOVA instead. Additionally, the results from the simulation study showed that the power of the proposed nonparametric kernel-based MANOVA tests increased with sample size and dimension as expected. Larger sample sizes resulted in a higher probability of rejecting the null hypothesis which resulted in an increased power. Thus, the proposed nonparametric kernel-based MANOVA tests have a greater ability to detect difference in group means with larger sample sizes. The tables in Chapter IV provide more details regarding the appropriate sample sizes and desired power. It was also observed that there was a slight decrease in power with higher correlation in the variance-covariance matrices. This could be due to the fact that when the data are more correlated the statistical model tends to be more complicated and therefore statistical tests have less power. Although there is a decrease in power in the proposed kernel-based MANOVA when the data are correlated, it is still advantageous and more appropriate to use over the parametric MANOVA when the normality assumption is violated.

Moreover, the results of the simulation study looked at the Type I error rate. Type I error is the probability of rejecting the null hypothesis when is it true. The results of the simulation study showed that the proposed nonparametric kernel-based MANOVA and parametric MANOVA tests have a Type I error rate of around 5% when the underlying distribution of the data is normal and the significance level is $\alpha = 0.05$. Also, the results of the simulation study showed that the proposed nonparametric kernel-based MANOVA and parametric MANOVA tests have a Type I error rate of around 2% when the the underlying distribution of the data is Cauchy distribution and the significance level is $\alpha = 0.05$. The results of the simulation study showed that the proposed nonparametric kernel-based MANOVA and parametric MANOVA tests have a Type I error rate of around 4% when the the underlying distribution of the data is exponential and the significance level is $\alpha = 0.05$. It can be concluded that the proposed nonparametric kernel-based MANOVA tests for non-normal data perform as well as the the parametric MANOVA tests for normal data, regarding Type I error rate.

Type I error rate is significantly lower (around 2%) when the data come from a multivariate Cauchy distribution kernel. This could be due to the fact that the kernel used in this study, i.e. Gaussian kernel, has light tails while the Cauchy distribution has heavy tails. Thus, it is hard to obtain a Type I error rate similar to other distributions (normal and exponential) unless the sample size is sufficiently large.

For the real-data application considered in this dissertation, the developed methods were applied to a breast cancer image dataset. However, these methods can be applied to any discipline or area of research as long as the model and hypothesis tests are correctly specified, the assumptions are met, and the sample sizes are large enough for the statistical tests to follow the asymptotic distributions they are meant to follow in line with the theoretical proofs.

Limitations

Although this dissertation investigated the hypothesis testing in a MANOVA framework for non-normal and non-linear data using various number of situation to account for possible real-data situations, there are still some limitations.

First, the main assumption of the homogeneity of the variance-covariance matrices is a limitation to this dissertation. Homogeneity (equality) of the variance-covariance matrices is not realistic in most real-data applications. However, this assumption was considered most useful for conducting the methodology of in this dissertation.

Furthermore, the simulation study in this dissertation included only the situations of the balanced design. The developed theoretical methodology in Chapter III included techniques for both balanced and unbalanced design. However, due to the time and computational restrains, only the balanced design is used in the simulation study for evaluating the proposed nonparametric methodology.

Only small sample size simulation study were conducted in this dissertation. Although the nonparamteric approach is used in high dimensional data in many fields of research, in this dissertation only small sample sizes, $n = 10, ..., 100$, are considered due to time and computational restrains.

Future Direction

In future research, an extension of this dissertation will be conducted in several areas, including, but not limited to, extending the methodology to assess the limitations addressed in this dissertation. In real-world application, groups or population do not have the same variance-covariance matrices. Thus, an extension of this dissertation will be conducted to allow for variation in the variance-covariance matrices in multivariate settings.

An extension of the simulation study in this dissertation will be conducted to account for unbalanced designs. This would enable researchers and practitioners to understand the evaluation of the proposed nonparametric methodology performance proposed in this dissertation when data are not equal for each group.

Also, an extension of the methodology developed in this dissertation to other experimental design models, such as incomplete block design in which not all the treatments occur in every block and Latin square design, should be considered. These are the designs that are more realistic in real-data applications than the complete randomized design in multivariate settings. In addition, a random effect, rather than a fixed effect, nonparametric multivariate analysis of variance can be studied.

Additionally, an extension of the developed methodology will be considered to perform a kernel-based post-hoc test when significant difference among groups is present would be of interest.

An extension of this dissertation's simulation study will be conducted to evaluated the performance of the proposed nonparametric kernel-based test when high dimensional data is used. A dataset is considered a high-dimensional data when the data dimension p is larger than the sample size n.

Finally, developing an R package of the developed methodology of this dissertation will be part of the future work related to this dissertation. Developing an R package that can handle nonparametric MANOVA for one-way and two-way layout would enable many students, researchers, and practitioners to perform MANOVA tests for non-normal and non-linear data.

REFERENCES

Ahmad, I. A. (1982). Nonparametric estimation of the location and scale parameters based on density estimation. *Annals of the Institute of Statistical Mathematics*, 34, 39–53.

Aizerman, M. A., Braverman, E. A., & Rozonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, (p./pp. 821-837).

Alizadeh, M., & Ebadzadeh, M. M. (2011). Kernel evolution for support vector classification. *2011 IEEE Workshop on Evolving and Adaptive Intelligent Systems (EAIS)*. doi: 10.1109/eais.2011.5945924

Altman, D. G., & Bland, J. M. (2009). Parametric v non-parametric methods for data analysis. *Bmj, 338*(apr02 1), a3167-a3167. doi:10.1136/bmj.a3167

Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S. S., Safwan, M., Alex, V., . . . Aguiar, P. (2019). BACH: Grand challenge on breast cancer histology images. *Medical Image Analysis*, *56*, 122-139. doi:10.1016/j.media.2019.05.010

Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., . . . Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. *PloS One, 12*(6), e0177544. doi:10.1371/journal.pone.0177544

Bellman, & R. (1961). *Adaptive control processes: A guided tour*. US: Princeton University Press.

Bellman, R. (1957). *Dynamic programming*. Princeton: Princeton University Press.

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2014). *A chronicle of permutation statistical methods: 1920-2000, and beyond* (2014th ed.). Cham;New York;: Springer. doi:10.1007/978-3-319-02744-9

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory - COLT 92*. doi: 10.1145/130385.130401

Bowman, A. W., & Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations (Oxford science publications)*. Oxford University Press.

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance*. London;Newbury Park, [Calif.];: SAGE.

Chen, H., & Xia, Y. (2019). A Nonparametric Normality Test for High-dimensional Data.

Chen, S. (2013). *Nonparametric ANOVA using kernel methods* (Unpublished doctoral dissertation). Oklahoma State University, Stillwater, OK.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London;New York;: Springer.

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach* (Second;2nd; ed.). Hoboken, New Jersey: Wiley.

Daniel, W. W. (1990). *Applied nonparametric statistics* (2nd ed). Boston: PWS-KENT Pub.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. New York: Springer. doi:10.1007/978-0-387-75971-5

Diaf, A., Boufama, B., & Benlamri, R. (2012). A compound eigenspace for recognizing directed human activities. (pp. 122-129). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-31298-4_15

Dickhaus, T. (2018). *Theory of nonparametric tests.* Cham: Springer International Publishing.

Finney, D. (1977). Dimensions of Statistics. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 26*(3), 285-289. doi:10.2307/2346969

Fisher, R. A. (1918). The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Philosophical Transactions of the Royal Society of Edinburgh*, 52, 399–433.

Fondón, I., Sarmiento, A., García, A. I., Silvestre, M., Eloy, C., Polónia, A., & Aguiar, P. (2018). Automatic classification of tissue malignancy for breast carcinoma diagnosis. *Computers in Biology and Medicine, 96*, 41-51. doi:10.1016/j.compbiomed.2018.03.003

Geladi, P., & Grahn, H. (2006). *Multivariate image analysis*. Chichester: Wiley. doi: 10.1002/9780470027318.a8106

Good, P. I., & SpringerLink (Online service). (2005). *Permutation, parametric and bootstrap tests of hypotheses* (3rd ed.). New York: Springer. doi:10.1007/b138696

Hajek, J. (1968). Asymptotic Normality of Simple Linear Rank Statistics Under Alternatives. *The Annals of Mathematical Statistics, 39*(2), 325–346. doi: 10.1214/aoms/1177698394

Hollander, M., Chicken, E., & Wolfe, D. A. (2014). *Nonparametric statistical methods* (Third ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Hollander, M., & Wolfe, D. A. (1999). *Non parametric statistical methods* (Second ed.). New Jersey: John Wiley and Son.

Howley, T., & Madden, M. G. (2005). The genetic kernel support vector machine: Description and evaluation. *Artificial Intelligence Review, 24*(3), 379-395. doi:10.1007/s10462-005-9009-3

Kaluza, B. (2016). *Machine learning in java* (1st ed.). Birmingham, England: Packt Publishing.

Koroljuk, V. S., & Borovskich, Y. V. (1994). *Theory of U-statistics*. Dordrecht: Kluwer Academic.

Lane, D. M. (2013) Introduction to Statistics: An Interactive eBook.

Liu, H. (2010). *Nonparametric learning in high dimensions.* Carnegie Mellon University, ProQuest Dissertations Publishing.

Liu, R. Y., & McKean, J. W. (2016). *Robust rank-based and nonparametric methods: Michigan, USA, april 2015 : Selected, revised, and extended contributions*. Switzerland: Springer. doi:10.1007/978-3-319-39065-9

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. Academic Press.

May, W. L., & Johnson, W. D. (1997). A SASR macro for the multivariate extension of the kruskal-wallis test including multiple comparisons: Randomization and χ^2 criteria. *Computational Statistics & Data Analysis*, 26(2), 239-250. doi:10.1016/S0167-9473(97)82107-X

Mika, S., Rätsch, G., Weston, J., Schölkopf, B., & Müller, K.-R. (1999). Fisher discriminant analysis with kernels. In Hu, Y.-H., Larsen, J., Wilson, E., and Douglas, S., eds, *Neural Networks for Signal Processing*, IX, 41–48, IEEE.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping, 15*(1), 1-25. doi:10.1002/hbm.1058

Paulsen, V., & Raghupathi, M. (2016). *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces* (Cambridge Studies in Advanced Mathematics). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316219232

Puri, M. L., & Sen, P. K. (1969). A class of rank order tests for a general linear hypothesis. *The Annals of Mathematical Statistics*, 40(4), 1325-1343. doi:10.1214/aoms/1177697505

R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis, third edition* (3rd ed.). Hoboken: Wiley.

Riesz, F. Untersuchungen über Systeme integrierbarer Funktionen. *Mathematische Annalen, 69*, 449-497.

Scheffé, H. (1959). *The analysis of variance* (1st ed.). New Jersey: John Wiley and Sons, Inc.

Scott, D. W. (1992). *Multivariate density estimation: theory, practice, and visualization*. New Jersey: John Wiley & Sons, Inc.

Shorack, G.R.; Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. New York: Wiley.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: MacGraw-Hill.

Siegel, R. L., Miller, K. D., & Jemal, A. (2017). Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians, 67*(1), 7.

Smith, R. A., Cokkinides, V., & Eyre, H. J. (2006). American cancer society guidelines for the early detection of cancer, 2006. *CA: A Cancer Journal for Clinicians, 56*(1), 11-25. doi:10.3322/canjclin.56.1.11

Steinwart, I., Christmann, A., & SpringerLink (Online service). (2008). *Support vector machines* (1st ed.). New York: Springer. doi:10.1007/978-0-387-77242-4

Theodoridis, S., & Koutroumbas, K. (2009). *Pattern recognition* (4th ed.). Burlington, MA;London;: Academic Press.

Tucker, H. G. (1959). A generalization of the glivenko-cantelli theorem. *The Annals of Mathematical Statistics,* 30(3), 828-830. doi:10.1214/aoms/1177706212

Vaart, A. (1998). *Asymptotic Statistics* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511802256

Wand, M. P., & Jones, M. C. (1994). *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability (60). Boca Raton, FL, U.S.: Chapman & Hall.

Wang, Y., Yunde, J., Changbo, H., & Turk, M. (2004). Fisher non-negative matrix factorization for learning local features. *Proceedings of the 6th Asian Conference on Computer Vision, Jeju, Korea*, 27-30.

Wilimitis, D. (2018). *Kernel Trick*. Retrieved from https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f

Yuan, K. H., & Bentler, P. M. (2010). Two simple approximations to the distributions of quadratic forms. *The British journal of mathematical and statistical psychology*, *63*(Pt 2), 273–291. doi:10.1348/000711009X449771

APPENDIX A

R CODE

Data Generation

One-way MANOVA

```
#### Sample of How to Generate Data ####

### Type I Error ###

sigma <- matrix(c(1.0, 0,
                    0, 1.0), nrow = 2)

mu <- c(1,2)

  x <- data.frame(mvrnorm(n = 30, mu = mu, Sigma = sigma),
                  subjects = c(rep('1', 10),
                               rep('2', 10),
                               rep('3', 10)))

### Power ###

sigma <- matrix(c(1.0, 0,
                    1.0, 0), nrow = 2)

mu1 <- c(.5,1.5)

mu2 <- c(1,2)

mu3 <- c(1.5,2.5)

  x1 <- data.frame(mvrnorm(n = 10, mu = mu1, Sigma = sigma),
                   subjects = c(rep('1', 10)))

  x2 <- data.frame(mvrnorm(n = 10, mu = mu2, Sigma = sigma),
                   subjects = c(rep('2', 10)))

  x3 <- data.frame(mvrnorm(n = 10, mu = mu3, Sigma = sigma),
                   subjects = c(rep('3', 10)))

  x <- rbind(x1,x2,x3)
```

Interaction Effect of the Two-way MANOVA

```
#### Sample of How to Generate Data ####

### Type I Error ###
```

```
sigma <- matrix(c(1.0, 0,
                           0, 1.0), nrow = 2)
  x11 <- data.frame(mvrnorm(n = 10, mu = c(1,1), Sigma = sigma),
                   subjects = c(rep('11', 10)))
  x12 <- data.frame(mvrnorm(n = 10, mu = c(2,2), Sigma = sigma),
                   subjects = c(rep('12', 10)))
  x13 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                   subjects = c(rep('13', 10)))
  x21 <- data.frame(mvrnorm(n = 10, mu = c(2,2), Sigma = sigma),
                   subjects = c(rep('21', 10)))
  x22 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                   subjects = c(rep('22', 10)))
  x23 <- data.frame(mvrnorm(n = 10, mu = c(4,4), Sigma = sigma),
                   subjects = c(rep('23', 10)))
  x31 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                   subjects = c(rep('31', 10)))
  x32 <- data.frame(mvrnorm(n = 10, mu = c(4,4), Sigma = sigma),
                   subjects = c(rep('32', 10)))
  x33 <- data.frame(mvrnorm(n = 10, mu = c(5,5), Sigma = sigma),
                   subjects = c(rep('33', 10)))
  x <- rbind(x11,x12,x13,x21,x22,x23,x31,x32,x33)
### Power ###
sigma <- matrix(c(1.0, 0,
                           0, 1.0), nrow = 2)
  x11 <- data.frame(mvrnorm(n = 10, mu = c(1.5,1.5), Sigma = sigma),
                    subjects = c(rep('11', 10)))
  x12 <- data.frame(mvrnorm(n = 10, mu = c(1.5,1.5), Sigma = sigma),
```

```
                    subjects = c(rep('12', 10)))
  x13 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                    subjects = c(rep('13', 10)))
  x21 <- data.frame(mvrnorm(n = 10, mu = c(1.5,1.5), Sigma = sigma),
                    subjects = c(rep('21', 10)))
  x22 <- data.frame(mvrnorm(n = 10, mu = c(3.5,3.5), Sigma = sigma),
                    subjects = c(rep('22', 10)))
  x23 <- data.frame(mvrnorm(n = 10, mu = c(4,4), Sigma = sigma),
                    subjects = c(rep('23', 10)))
  x31 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                    subjects = c(rep('31', 10)))
  x32 <- data.frame(mvrnorm(n = 10, mu = c(4,4), Sigma = sigma),
                    subjects = c(rep('32', 10)))
  x33 <- data.frame(mvrnorm(n = 10, mu = c(5,5), Sigma = sigma),
                    subjects = c(rep('33', 10)))
  x <- rbind(x11,x12,x13,x21,x22,x23,x31,x32,x33)
```

Parametric MANOVA

One-way MANOVA

```
#### Sample of How to Generate Data ####
### Type I Error ###
sigma <- matrix(c(1.0, 0,
                    0, 1.0), nrow = 2)
mu <- c(1,2)
sim=10000
t1err=0
for (i in 1:sim){
  x <- data.frame(mvrnorm(n = 30, mu = mu, Sigma = sigma),
```

```
                  subjects = c(rep('1', 10),
                                    rep('2', 10),
                                    rep('3', 10)))
## p-value ##
  if (((summary(manova(as.matrix(x[,1:2])~x$subjects),'Wilks'))
  $stats[1,6])<= 0.05) (t1err=t1err+1)
}
cat("Type I error rate in percentage is", (t1err/sim)*100,"%")
### Power ###
sigma <- matrix(c(1.0, 0,
                    1.0, 0), nrow = 2)
mu1 <- c(.5,1.5)
mu2 <- c(1,2)
mu3 <- c(1.5,2.5)
sim=10000
t2err=0
for (i in 1:sim){
  x1 <- data.frame(mvrnorm(n = 10, mu = mu1, Sigma = sigma),
                   subjects = c(rep('1', 10)))
  x2 <- data.frame(mvrnorm(n = 10, mu = mu2, Sigma = sigma),
                   subjects = c(rep('2', 10)))
  x3 <- data.frame(mvrnorm(n = 10, mu = mu3, Sigma = sigma),
                   subjects = c(rep('3', 10)))
  x <- rbind(x1,x2,x3)
  if (((summary(manova(as.matrix(x[,1:2])~ x$subjects),'Wilks'))
  $stats[1,6]) > 0.05) (t2err=t2err+1)
}
```

```
cat("Power rate in percentage is",(1-(t2err/sim))*100,"%")
```

Interaction Effect of the Two-way MANOVA

```
### Type I Error ###

sigma <- matrix(c(1.0, 0,
                           0, 1.0), nrow = 2)

sim=10000

t1err=0

for (i in 1:sim){

  x11 <- data.frame(mvrnorm(n = 10, mu = c(1,1), Sigma = sigma),
                   subjects = c(rep('11', 10)))

  x12 <- data.frame(mvrnorm(n = 10, mu = c(2,2), Sigma = sigma),
                   subjects = c(rep('12', 10)))

  x13 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                   subjects = c(rep('13', 10)))

  x21 <- data.frame(mvrnorm(n = 10, mu = c(2,2), Sigma = sigma),
                   subjects = c(rep('21', 10)))

  x22 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                   subjects = c(rep('22', 10)))

  x23 <- data.frame(mvrnorm(n = 10, mu = c(4,4), Sigma = sigma),
                   subjects = c(rep('23', 10)))

  x31 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                   subjects = c(rep('31', 10)))

  x32 <- data.frame(mvrnorm(n = 10, mu = c(4,4), Sigma = sigma),
                   subjects = c(rep('32', 10)))

  x33 <- data.frame(mvrnorm(n = 10, mu = c(5,5), Sigma = sigma),
                   subjects = c(rep('33', 10)))

  x <- rbind(x11,x12,x13,x21,x22,x23,x31,x32,x33)
```

```
  ## p-value ##
  if (((summary(manova(as.matrix(x[,1:2])~x$subjects),'Wilks'))
        $stats[1,6])<= 0.05) (t1err=t1err+1)
}
cat("Type I error rate in percentage is", (t1err/sim)*100,"%")
### Power ###
rm(list = ls(all.names = TRUE))
sigma <- matrix(c(1.0, 0,
                  0, 1.0), nrow = 2)
sim=10000
t2err=0
for (i in 1:sim){
  x11 <- data.frame(mvrnorm(n = 10, mu = c(1.5,1.5), Sigma = sigma),
                    subjects = c(rep('11', 10)))
  x12 <- data.frame(mvrnorm(n = 10, mu = c(1.5,1.5), Sigma = sigma),
                    subjects = c(rep('12', 10)))
  x13 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                    subjects = c(rep('13', 10)))
  x21 <- data.frame(mvrnorm(n = 10, mu = c(1.5,1.5), Sigma = sigma),
                    subjects = c(rep('21', 10)))
  x22 <- data.frame(mvrnorm(n = 10, mu = c(3.5,3.5), Sigma = sigma),
                    subjects = c(rep('22', 10)))
  x23 <- data.frame(mvrnorm(n = 10, mu = c(4,4), Sigma = sigma),
                    subjects = c(rep('23', 10)))
  x31 <- data.frame(mvrnorm(n = 10, mu = c(3,3), Sigma = sigma),
                    subjects = c(rep('31', 10)))
  x32 <- data.frame(mvrnorm(n = 10, mu = c(4,4), Sigma = sigma),
```

```
                    subjects = c(rep('32', 10)))
  x33 <- data.frame(mvrnorm(n = 10, mu = c(5,5), Sigma = sigma),
                    subjects = c(rep('33', 10)))
  x <- rbind(x11,x12,x13,x21,x22,x23,x31,x32,x33)
  ## p-value ##
  if (((summary(manova(as.matrix(x[,1:2])~x$subjects),'Wilks'))
  $stats[1,6])<= 0.05) (t2err=t2err+1)
}
cat("Power rate in percentage is",(1-(t2err/sim))*100,"%")
```

Nonparametric MANOVA

One-way MANOVA

```
### Type I Error ###
sim=10000
t1err=0
I <- 3
H <- n[1]^(-1/(p+4))*(sigma^(1/2))
## Qhat ##
Q11 <- 1/((n[1]*(n[1]-1))*det(H^(1/2)))
Q21 <- 1/2*(x1[1,]+x1[-1,])*(dmvnorm((x1[1,]-x1[-1,])%*%(solve(H^(1/2)))))
Q31 <- matrix(c(sum(Q21[,1]), sum(Q21[,2])), nrow =1, ncol=p)
Qhat1 <- t(Q11*Q31)
Q12 <- 1/((n[2]*(n[2]-1))*det(H^(1/2)))
Q22 <- 1/2*(x2[1,]+x2[-1,])*(dmvnorm((x2[1,]-x2[-1,])%*%(solve(H^(1/2)))))
Q32 <- matrix(c(sum(Q22[,1]), sum(Q22[,2])), nrow =1, ncol=p)
Qhat2 <- t(Q12*Q32)
Q13 <- 1/((n[3]*(n[3]-1))*det(H^(1/2)))
Q23 <- 1/2*(x3[1,]+x3[-1,])*(dmvnorm((x3[1,]-x3[-1,])%*%(solve(H^(1/2)))))
```

```
Q33 <- matrix(c(sum(Q23[,1]), sum(Q23[,2])), nrow =1, ncol=p)
Qhat3 <- t(Q13*Q33)
## w2 ##
w21 <- 4*(((t(x1)%*%x1)*(mean((dmvnorm(x1)^3))))
-((t(x1)%*%x1)*(mean(dmvnorm(x1)^4))))
w22 <- 4*(((t(x2)%*%x2)*(mean((dmvnorm(x2)^3))))
-((t(x2)%*%x2)*(mean(dmvnorm(x2)^4))))
w23 <- 4*(((t(x3)%*%x3)*(mean((dmvnorm(x3)^3))))
-((t(x3)%*%x3)*(mean(dmvnorm(x3)^4))))
## Qdot ##
Qdot1 <- (1/(n[1]*det(w21)))* n[1]*(solve(w21)%*%Qhat1)
Qdot2 <- (1/(n[2]*det(w22)))* n[2]*(solve(w22)%*%Qhat2)
Qdot3 <- (1/(n[3]*det(w23)))* n[3]*(solve(w23)%*%Qhat3)
## SSB ##
SSB1 <- (n[1]*solve(w21))%*%((Qhat1-Qdot1)%*%t((Qhat1-Qdot1)))
SSB2 <- (n[2]*solve(w22))%*%((Qhat2-Qdot2)%*%t((Qhat2-Qdot2)))
SSB3 <- (n[3]*solve(w23))%*%((Qhat3-Qdot3)%*%t((Qhat1-Qdot3)))
SSB <- SSB1+SSB2+SSB3
## Aijk ##
A11 <- det(H^(1/2))
A21 <- 1/2*(x1[1,]+x1[-1,])*(dmvnorm((x1[1,]-x1[-1,])%*%(solve(H^(1/2)))))
A31 <- matrix(c(sum(A21[,1]), sum(A21[,2])), nrow =1, ncol=p)
A1 <- t(A11*A31)
A12 <- det(H^(1/2))
A22 <- 1/2*(x2[1,]+x2[-1,])*(dmvnorm((x2[1,]-x2[-1,])%*%(solve(H^(1/2)))))
A32 <- matrix(c(sum(A21[,1]), sum(A21[,2])), nrow =1, ncol=p)
A2 <- t(A12*A32)
```

```
A13 <- det(H^(1/2))
A23 <- 1/2*(x3[1,]+x3[-1,])*(dmvnorm((x3[1,]-x3[-1,])%*%(solve(H^(1/2)))))
A33 <- matrix(c(sum(A23[,1]), sum(A23[,2])), nrow =1, ncol=p)
A3 <- t(A13*A33)
## SSW ##
SSW1 <- (((n[1]/2)-1)*solve(w21))%*%((A1-Qhat1)%*%t((A1-Qhat1)))
SSW2 <- (((n[2]/2)-1)*solve(w22))%*%((A2-Qhat2)%*%t((A2-Qhat2)))
SSW3 <- (((n[3]/2)-1)*solve(w23))%*%((A3-Qhat3)%*%t((A3-Qhat3)))
SSW <- SSW1+SSW2+SSW3
## Test ##
a = (sum(n)-I)+(((I-1)-p-1)/2)
b <- if(p^2+((I-1)^2)-5 > 0)
{b=sqrt((p^2*((I-1)^2)-4)/(p^2*((I-1)^2)-5))} else 1
c = ((p*(I-1))/2)-1
df1 <- p*(I-1)
df2 <- (a*b)-c
Lambda <- 1/(1+(det(SSB/SSW)))
F1 <- (1-(Lambda^(1/b)))*df2
F2 <- (Lambda^(1/b))*df1
F <- F1/F2
Fvalue<-qf(.95,df1,df2)
if (F > Fvalue)  (t1err=t1err+1)
}
cat("Type I error rate in percentage is", (t1err/sim)*100,"%")
### Power ###
sim=10000
t2err=0
```

```
I <- 3
H <- n[1]^(-1/(p+4))*(sigma^(1/2))
## Qhat ##
Q11 <- 1/((n[1]*(n[1]-1))*det(H^(1/2)))
Q21 <- 1/2*(x1[1,]+x1[-1,])*(dmvnorm((x1[1,]-x1[-1,])%*%(solve(H^(1/2)))))
Q31 <- matrix(c(sum(Q21[,1]), sum(Q21[,2])), nrow =1, ncol=p)
Qhat1 <- t(Q11*Q31)
Q12 <- 1/((n[2]*(n[2]-1))*det(H^(1/2)))
Q22 <- 1/2*(x2[1,]+x2[-1,])*(dmvnorm((x2[1,]-x2[-1,])%*%(solve(H^(1/2)))))
Q32 <- matrix(c(sum(Q22[,1]), sum(Q22[,2])), nrow =1, ncol=p)
Qhat2 <- t(Q12*Q32)
Q13 <- 1/((n[3]*(n[3]-1))*det(H^(1/2)))
Q23 <- 1/2*(x3[1,]+x3[-1,])*(dmvnorm((x3[1,]-x3[-1,])%*%(solve(H^(1/2)))))
Q33 <- matrix(c(sum(Q23[,1]), sum(Q23[,2])), nrow =1, ncol=p)
Qhat3 <- t(Q13*Q33)
## w2 ##
w21 <- 4*(((t(x1)%*%x1)*(mean((dmvnorm(x1)^3))))
-((t(x1)%*%x1)*(mean(dmvnorm(x1)^4))))
w22 <- 4*(((t(x2)%*%x2)*(mean((dmvnorm(x2)^3))))
-((t(x2)%*%x2)*(mean(dmvnorm(x2)^4))))
w23 <- 4*(((t(x3)%*%x3)*(mean((dmvnorm(x3)^3))))
-((t(x3)%*%x3)*(mean(dmvnorm(x3)^4))))
## Qdot ##
Qdot1 <- (1/(n[1]*det(w21)))* n[1]*(solve(w21)%*%Qhat1)
Qdot2 <- (1/(n[2]*det(w22)))* n[2]*(solve(w22)%*%Qhat2)
Qdot3 <- (1/(n[3]*det(w23)))* n[3]*(solve(w23)%*%Qhat3)
## SSB ##
```

```
SSB1 <- (n[1]*solve(w21))%*%((Qhat1-Qdot1)%*%t((Qhat1-Qdot1)))

SSB2 <- (n[2]*solve(w22))%*%((Qhat2-Qdot2)%*%t((Qhat2-Qdot2)))

SSB3 <- (n[3]*solve(w23))%*%((Qhat3-Qdot3)%*%t((Qhat1-Qdot3)))

SSB <- SSB1+SSB2+SSB3

## Aijk ##

A11 <- det(H^(1/2))

A21 <- 1/2*(x1[1,]+x1[-1,])*(dmvnorm((x1[1,]-x1[-1,])%*%(solve(H^(1/2)))))

A31 <- matrix(c(sum(A21[,1]), sum(A21[,2])), nrow =1, ncol=p)

A1 <- t(A11*A31)

A12 <- det(H^(1/2))

A22 <- 1/2*(x2[1,]+x2[-1,])*(dmvnorm((x2[1,]-x2[-1,])%*%(solve(H^(1/2)))))

A32 <- matrix(c(sum(A21[,1]), sum(A21[,2])), nrow =1, ncol=p)

A2 <- t(A12*A32)

A13 <- det(H^(1/2))

A23 <- 1/2*(x3[1,]+x3[-1,])*(dmvnorm((x3[1,]-x3[-1,])%*%(solve(H^(1/2)))))

A33 <- matrix(c(sum(A23[,1]), sum(A23[,2])), nrow =1, ncol=p)

A3 <- t(A13*A33)

## SSW ##

SSW1 <- (((n[1]/2)-1)*solve(w21))%*%((A1-Qhat1)%*%t((A1-Qhat1)))

SSW2 <- (((n[2]/2)-1)*solve(w22))%*%((A2-Qhat2)%*%t((A2-Qhat2)))

SSW3 <- (((n[3]/2)-1)*solve(w23))%*%((A3-Qhat3)%*%t((A3-Qhat3)))

SSW <- SSW1+SSW2+SSW3

## Test ##

a = (sum(n)-I)+(((I-1)-p-1)/2)

b <- if(p^2+((I-1)^2)-5 > 0)

{b=sqrt((p^2*((I-1)^2)-4)/(p^2*((I-1)^2)-5))} else 1

c = ((p*(I-1))/2)-1
```

```
df1 <- p*(I-1)
df2 <- (a*b)-c
Lambda <- 1/(1+(det(SSB/SSW)))
F1 <- (1-(Lambda^(1/b)))*df2
F2 <- (Lambda^(1/b))*df1
F <- F1/F2
Fvalue<-qf(.95,df1,df2)
if (F > Fvalue)  (t2err=t2err+1)
}
cat("Type I error rate in percentage is", (1-(t2err/sim))*100,"%")
```

Interaction Effect of the Two-way MANOVA

```
### Type I Error ##
sim=10000
t1err=0
I <- 3
J <- 3
H <- n[1]^(-1/(p+4))*(sigma^(1/2))
## Qhat ##
Q11 <- 1/((n[1]*(n[1]-1))*det(H^(1/2)))
Q21 <- 1/2*(x1[1,]+x1[-1,])*(dmvnorm((x1[1,]-x1[-1,])%*%(solve(H^(1/2)))))
Q31 <- matrix(c(sum(Q21[,1]), sum(Q21[,2])), nrow =1, ncol=p)
Qhat1 <- t(Q11*Q31)
Q12 <- 1/((n[2]*(n[2]-1))*det(H^(1/2)))
Q22 <- 1/2*(x2[1,]+x2[-1,])*(dmvnorm((x2[1,]-x2[-1,])%*%(solve(H^(1/2)))))
Q32 <- matrix(c(sum(Q22[,1]), sum(Q22[,2])), nrow =1, ncol=p)
Qhat2 <- t(Q12*Q32)
Q13 <- 1/((n[3]*(n[3]-1))*det(H^(1/2)))
```

```
Q23 <- 1/2*(x3[1,]+x3[-1,])*(dmvnorm((x3[1,]-x3[-1,])%*%(solve(H^(1/2)))))

Q33 <- matrix(c(sum(Q23[,1]), sum(Q23[,2])), nrow =1, ncol=p)

Qhat3 <- t(Q13*Q33)

## w2 ##

w21 <- 4*(((t(x1)%*%x1)*(mean((dmvnorm(x1)^3))))

-((t(x1)%*%x1)*(mean(dmvnorm(x1)^4))))

w22 <- 4*(((t(x2)%*%x2)*(mean((dmvnorm(x2)^3))))

-((t(x2)%*%x2)*(mean(dmvnorm(x2)^4))))

w23 <- 4*(((t(x3)%*%x3)*(mean((dmvnorm(x3)^3))))

-((t(x3)%*%x3)*(mean(dmvnorm(x3)^4))))

## Qdot ##

Qdot1 <- (1/(n[1]*det(w21)))* n[1]*(solve(w21)%*%Qhat1)

Qdot2 <- (1/(n[2]*det(w22)))* n[2]*(solve(w22)%*%Qhat2)

Qdot3 <- (1/(n[3]*det(w23)))* n[3]*(solve(w23)%*%Qhat3)

## SSB ##

SSB1 <- (n[1]*solve(w21))%*%((Qhat1-Qdot1)%*%t((Qhat1-Qdot1)))

SSB2 <- (n[2]*solve(w22))%*%((Qhat2-Qdot2)%*%t((Qhat2-Qdot2)))

SSB3 <- (n[3]*solve(w23))%*%((Qhat3-Qdot3)%*%t((Qhat1-Qdot3)))

SSB <- SSB1+SSB2+SSB3

## Aijk ##

A11 <- det(H^(1/2))

A21 <- 1/2*(x1[1,]+x1[-1,])*(dmvnorm((x1[1,]-x1[-1,])%*%(solve(H^(1/2)))))

A31 <- matrix(c(sum(A21[,1]), sum(A21[,2])), nrow =1, ncol=p)

A1 <- t(A11*A31)

A12 <- det(H^(1/2))

A22 <- 1/2*(x2[1,]+x2[-1,])*(dmvnorm((x2[1,]-x2[-1,])%*%(solve(H^(1/2)))))

A32 <- matrix(c(sum(A21[,1]), sum(A21[,2])), nrow =1, ncol=p)
```

```
A2 <- t(A12*A32)

A13 <- det(H^(1/2))

A23 <- 1/2*(x3[1,]+x3[-1,])*(dmvnorm((x3[1,]-x3[-1,])%*%(solve(H^(1/2)))))

A33 <- matrix(c(sum(A23[,1]), sum(A23[,2])), nrow =1, ncol=p)

A3 <- t(A13*A33)

## SSW ##

SSW1 <- (((n[1]/2)-1)*solve(w21))%*%((A1-Qhat1)%*%t((A1-Qhat1)))

SSW2 <- (((n[2]/2)-1)*solve(w22))%*%((A2-Qhat2)%*%t((A2-Qhat2)))

SSW3 <- (((n[3]/2)-1)*solve(w23))%*%((A3-Qhat3)%*%t((A3-Qhat3)))

SSW <- SSW1+SSW2+SSW3

## Test ##

a = (I*J*(n[1]-1))+((((I-1)*(J-1))-p-1)/2)

b <- if(p^2+((I-1)^2)-5 > 0)

{b=sqrt((p^2*(((I-1)*(J-1))^2)-4)/(p^2*(((I-1)*(J-1))^2)-5))} else 1

c = ((p*((I-1)*(J-1)))/2)-1

df1 <- p*((I-1)*(J-1))

df2 <- (a*b)-c

Lambda <- 1/(1+(det(SSB/SSW)))

F1 <- (1-(Lambda^(1/b)))*df2

F2 <- (Lambda^(1/b))*df1

F <- F1/F2

Fvalue<-qf(.95,df1,df2)

if (F > Fvalue)  (t1err=t1err+1)

}

cat("Type I error rate in percentage is", (t1err/sim)*100,"%")


### Power ###
```

```
sim=10000

t2err=0

I <- 3

J <- 3

H <- n[1]^(-1/(p+4))*(sigma^(1/2))

## Qhat ##

Q11 <- 1/((n[1]*(n[1]-1))*det(H^(1/2)))

Q21 <- 1/2*(x1[1,]+x1[-1,])*(dmvnorm((x1[1,]-x1[-1,])%*%(solve(H^(1/2)))))

Q31 <- matrix(c(sum(Q21[,1]), sum(Q21[,2])), nrow =1, ncol=p)

Qhat1 <- t(Q11*Q31)

Q12 <- 1/((n[2]*(n[2]-1))*det(H^(1/2)))

Q22 <- 1/2*(x2[1,]+x2[-1,])*(dmvnorm((x2[1,]-x2[-1,])%*%(solve(H^(1/2)))))

Q32 <- matrix(c(sum(Q22[,1]), sum(Q22[,2])), nrow =1, ncol=p)

Qhat2 <- t(Q12*Q32)

Q13 <- 1/((n[3]*(n[3]-1))*det(H^(1/2)))

Q23 <- 1/2*(x3[1,]+x3[-1,])*(dmvnorm((x3[1,]-x3[-1,])%*%(solve(H^(1/2)))))

Q33 <- matrix(c(sum(Q23[,1]), sum(Q23[,2])), nrow =1, ncol=p)

Qhat3 <- t(Q13*Q33)

## w2 ##

w21 <- 4*(((t(x1)%*%x1)*(mean((dmvnorm(x1)^3))))

-((t(x1)%*%x1)*(mean(dmvnorm(x1)^4))))

w22 <- 4*(((t(x2)%*%x2)*(mean((dmvnorm(x2)^3))))

-((t(x2)%*%x2)*(mean(dmvnorm(x2)^4))))

w23 <- 4*(((t(x3)%*%x3)*(mean((dmvnorm(x3)^3))))

-((t(x3)%*%x3)*(mean(dmvnorm(x3)^4))))

## Qdot ##

Qdot1 <- (1/(n[1]*det(w21)))* n[1]*(solve(w21)%*%Qhat1)
```

```
Qdot2 <- (1/(n[2]*det(w22)))* n[2]*(solve(w22)%*%Qhat2)

Qdot3 <- (1/(n[3]*det(w23)))* n[3]*(solve(w23)%*%Qhat3)

## SSB ##

SSB1 <- (n[1]*solve(w21))%*%((Qhat1-Qdot1)%*%t((Qhat1-Qdot1)))

SSB2 <- (n[2]*solve(w22))%*%((Qhat2-Qdot2)%*%t((Qhat2-Qdot2)))

SSB3 <- (n[3]*solve(w23))%*%((Qhat3-Qdot3)%*%t((Qhat1-Qdot3)))

SSB <- SSB1+SSB2+SSB3

## Aijk ##

A11 <- det(H^(1/2))

A21 <- 1/2*(x1[1,]+x1[-1,])*(dmvnorm((x1[1,]-x1[-1,])%*%(solve(H^(1/2)))))

A31 <- matrix(c(sum(A21[,1]), sum(A21[,2])), nrow =1, ncol=p)

A1 <- t(A11*A31)

A12 <- det(H^(1/2))

A22 <- 1/2*(x2[1,]+x2[-1,])*(dmvnorm((x2[1,]-x2[-1,])%*%(solve(H^(1/2)))))

A32 <- matrix(c(sum(A21[,1]), sum(A21[,2])), nrow =1, ncol=p)

A2 <- t(A12*A32)

A13 <- det(H^(1/2))

A23 <- 1/2*(x3[1,]+x3[-1,])*(dmvnorm((x3[1,]-x3[-1,])%*%(solve(H^(1/2)))))

A33 <- matrix(c(sum(A23[,1]), sum(A23[,2])), nrow =1, ncol=p)

A3 <- t(A13*A33)

## SSW ##

SSW1 <- (((n[1]/2)-1)*solve(w21))%*%((A1-Qhat1)%*%t((A1-Qhat1)))

SSW2 <- (((n[2]/2)-1)*solve(w22))%*%((A2-Qhat2)%*%t((A2-Qhat2)))

SSW3 <- (((n[3]/2)-1)*solve(w23))%*%((A3-Qhat3)%*%t((A3-Qhat3)))

SSW <- SSW1+SSW2+SSW3

## Test ##

a = (sum(n)-I)+(((I-1)-p-1)/2)
```

```
b <- if(p^2+((I-1)^2)-5 > 0)
{b=sqrt((p^2*((I-1)^2)-4)/(p^2*((I-1)^2)-5))} else 1
c = ((p*(I-1))/2)-1
df1 <- p*(I-1)
df2 <- (a*b)-c
Lambda <- 1/(1+(det(SSB/SSW)))
F1 <- (1-(Lambda^(1/b)))*df2
F2 <- (Lambda^(1/b))*df1
F <- F1/F2
Fvalue<-qf(.95,df1,df2)
if (F > Fvalue)  (t2err=t2err+1)
}
cat("Type I error rate in percentage is", (1-(t2err/sim))*100,"%")

### Function ###
set.seed(1234567)
sim=10000
t1err=0
I <- 3
p <- 2
n<-c(20,20,20)
mu1=rep(1,p)
mu2=rep(1,p)
mu3=rep(1,p)
sigma <- matrix(c(1.0, 0,
                  0, 1.0), nrow = p)
groups<-rep(1:I, n)
```

```
x1=mvrnorm(n[1], mu1,sigma)
x2=mvrnorm(n[2], mu2,sigma)
x3=mvrnorm(n[3], mu3,sigma)
x=rbind(x1,x2,x3)
H <- n[1]^(-1/(p+4))*(sigma^(1/2))
## Qhat ##
Qhat <- function(x, groups,n,H){
  p = dim(x)[2]
  I=length(unique(groups));
  Q1<- matrix(0,ncol=I)
  Q2 <- matrix(0,nrow=p, ncol=I)
  Q3 <- matrix(0,nrow=p, ncol=I)
  Q4 <- matrix(0,nrow=p, ncol=I)
  for (i in 1:I){
    for (j in 1:n[i]){
    xobs<-x[groups==i,]
    n[i]<- length(xobs)
    Q1[i] <- 1/((n[i]*(n[i]-1))*det(H^(1/2)))
    sub <- t(sapply(1:nrow(xobs), function(x) if (x == 1) { xobs[x, ]
              - xobs[nrow(xobs), ] }  else {xobs[x, ] - xobs[x-1, ]}))
    add <- t(sapply(1:nrow(xobs), function(x) if (x == 1) { xobs[x, ]
              + xobs[nrow(xobs), ] }  else {xobs[x, ] + xobs[x-1, ]}))
    Q2 <- 1/2*add*(dmvnorm((sub)%*%(solve(H^(1/2)))))
    Q3[,i] <- t(t(sapply(1:p, function(x) (sum(Q2[,x])))))
    Q4[,i] <- Q1[i]*Q3[,i]
    }
  }
```

```
  return (Q4)
}
Omega <- function(x, groups){
  p = dim(x)[2]
  I=length(unique(groups));
  w2 <- matrix(0, nrow=I, ncol=p*p)
  for (i in 1:I){
    xobs<-x[groups==i,]
    w2[i,] <- 4*(((t(xobs)%*%xobs)*(mean((dmvnorm(xobs)^3))))
                  -((t(xobs)%*%xobs)*(mean(dmvnorm(xobs)^4))))
  }
  return (w2) ## Each p*p matrix is a row in this function
}
Qdot <- function(Qhat, Omega){
  p = dim(x)[2]
  I=length(unique(groups));
  temp1 <- matrix(0, nrow=I, ncol=p*p)
  temp2 <- matrix(0,nrow=p, ncol=I)
  temp <- matrix(0,nrow=p, ncol=I)
  for (i in 1:I){
    temp1 <- Omega(x, groups)
    temp2 <- matrix(temp1[i,], nrow=p)
    temp3 <- Qhat(x, groups,n,H)
    temp4 <- temp3[,i]
    temp[,i] <- (1/(n[i]*det(temp2)))* n[i]*(solve(temp2)%*%temp4)
  }
  return (temp)
```

```
}
## SSB ##
SSB1 <- function(Omega, Qhat, Qdot){
  p = dim(x)[2]
  I=length(unique(groups));
  temp <- matrix(0, nrow=I, ncol=p*p)
  for (i in 1:I){
    temp1 <- Omega(x, groups)
    temp2 <- matrix(temp1[i,], nrow=p)
    temp3 <- Qhat(x, groups,n,H)
    temp4 <- temp3[,i]
    temp5 <- Qdot(Qhat, Omega)
    temp6 <- temp5[,i]
    temp[i,] <- (n[i]*solve(temp2))%*%((temp4-temp6)%*%t((temp4-temp6)))
  }
  return (temp)
}
SSB <- function (SSB1){  ## t(SSB(SSB1))?
  p = dim(x)[2]
  SSB <- matrix(0, nrow=1, ncol=p*p)
  for (i in 1:(p*p)){
    SSB[,i] <- sum(SSB1(Omega, Qhat, Qdot)[,i])
  }
  return (matrix(SSB, nrow=p))
}
Aij <- function(x, groups,H){
  p = dim(x)[2]
```

```
  I=length(unique(groups));
  A1<- matrix(0,ncol=I)
  A2 <- matrix(0,nrow=p, ncol=I)
  A3 <- matrix(0,nrow=p, ncol=I)
  A4 <- matrix(0,nrow=p, ncol=I)
  for (i in 1:I){
      xobs<-x[groups==i,]
      A1[i] <- det(H^(1/2))
      sub <- t(sapply(1:nrow(xobs), function(x) if (x == 1) { xobs[x, ]
                - xobs[nrow(xobs), ] }  else {xobs[x, ] - xobs[x-1, ]}))
      add <- t(sapply(1:nrow(xobs), function(x) if (x == 1) { xobs[x, ]
                + xobs[nrow(xobs), ] }  else {xobs[x, ] + xobs[x-1, ]}))
      A2 <- 1/2*add*(dmvnorm((sub)%*%(solve(H^(1/2)))))
      A3[,i] <- t(t(sapply(1:p, function(x) (sum(A2[,x])))))
      A4[,i] <- A1[i]*A3[,i]
  }
  return (A4)
}
## SSW ##
SSW1 <- function(Omega, Qhat){
  p = dim(x)[2]
  I=length(unique(groups));
  temp <- matrix(0, nrow=I, ncol=p*p)
  for (i in 1:I){
    temp1 <- Omega(x, groups)
    temp2 <- matrix(temp1[i,], nrow=p)
    temp3 <- Qhat(x, groups,n,H)
```

```
    temp4 <- temp3[,i]
    temp5 <- Aij(x, groups,H)
    temp6 <- temp5[,i]
    temp[i,] <- (((n[i]/2)-1)*solve(temp2))%*%((temp6-temp4)%*%
                    t((temp6-temp4)))   }
  return (temp)
}
SSW <- function (SSW1){  ## t(SSW(SSW1))?
  p = dim(x)[2]
  SSW <- matrix(0, nrow=1, ncol=p*p)
  for (i in 1:(p*p)){
    SSW[,i] <- sum(SSW1(Omega, Qhat)[,i])
  }
  return (matrix(SSW, nrow=p))
}
## Kernel Test ##
KernelLambda1 <- function (x,groups,H){
  a = (sum(n)-I)+(((I-1)-p-1)/2)
  b <- if(p^2+((I-1)^2)-5 > 0) {b=sqrt((p^2*((I-1)^2)-4)/(p^2*((I-1)^2)-5))}
        else 1
  c = ((p*(I-1))/2)-1
  df1 <- p*(I-1)
  df2 <- (a*b)-c
  Lambda <- 1/(1+(det(SSB(SSB1)/SSW(SSW1))))
  F1 <- (1-(Lambda^(1/b)))*df2
  F2 <- (Lambda^(1/b))*df1
  F <- F1/F2
```

```
  ## p-value
  Fvalue<-qf(.95,df1,df2)
  return(Fvalue)
}
##  Test
set.seed(1234567)
t1err = 0
sim = 100
## if (F > Fvalue)  (t1err=t1err+1)
for (i in 1: sim){
  p <- 2
  n<-c(20,20,20)
  H <- n[1]^(-1/(p+4))*solve(sigma^(1/2))
  I <- 3
  mu1=rep(1,p)
  mu2=rep(1,p)
  mu3=rep(1,p)
  sigma <- matrix(c(1.0, 0,
                    0, 1.0), nrow = 2)
  groups<-rep(1:3, n)
  x1=mvrnorm(n[1], mu1,sigma)
  x2=mvrnorm(n[2], mu2,sigma)
  x3=mvrnorm(n[3], mu3,sigma)
  x=rbind(x1,x2,x3)
  if(KernelLambda1(x,groups,H)> F) {(t1err=t1err+1)}  }
cat("Type I error rate in percentage is", (t1err/sim)*100,"%")
```