NONPARAMETRIC ANOVA USING KERNEL METHODS

By

SU CHEN

Bachelor of Science in Statistics
Wuhan University of Technology
Wuhan, Hubei, China
2007

Master of Science in Statistics
Oklahoma State University
Stillwater, OK, USA
2009

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2013

NONPARAMETRIC ANOVA USING KERNEL METHODS

Dissertation Approved:

Dr. Ibrahim Ahmad

Dissertation Advisor

Dr. Mark Payton

Dr. Lan Zhu

Dr. Tim Krehbiel

Outside Committee Member

Name: Su Chen

Date of Degree: July, 2013

Title of Study: NONPARAMETRIC ANOVA USING KERNEL METHODS

Major Field: Statistics

Abstract:

A new approach to one-way and two-way analysis of variance from the nonparametric view point is introduced and studied. It is nonparametric in the sense that no distributional format assumed and the testing pertain to location and scale parameters. In contrast to the rank transformed approach, the new approach uses the measurement responses along with the highly recognized kernel density estimation, and thus called "kernel transformed" approach. Firstly, a novel kernel transformed approach to test the homogeneity of scale parameters of a group of populations with unknown distributions is provided. When homogeneity of scales is not rejected, we proceed to develop a one-way ANOVA for testing equality of location parameters and a generalized way that handles the two-way layout with interaction. The proposed methods are asymptotically F-distributed. Simulation is used to compare the empirical significance level and the empirical power of our technique with the usual parametric approach. It is demonstrated that in the Normal Case, our method is very close to the standard ANOVA. While for other distributions that are heavy tailed (Cauchy) or skewed (Log-normal) our method has better empirical significance level close to the nominal level α and the empirical power of our technique are far superior to that of the standard ANOVA.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 One-way Analysis of Variance

Analysis of Variance (ANOVA) is a process of analyzing the differences in means (or medians; or distributions) among several groups. Both parametric and nonparametric methods have been developed in the literature. The classical analysis of variance, which is a parametric method, is usually called F-test or variance ratio test. It is called 'parametric' test as its hypothesis is about population parameters, namely the mean and standard deviation. Compared to parametric methods, nonparametric methods do not make any assumptions about the distribution, therefore it usually does not make hypothesis about the parameter, like the mean, but rather about the population distributions or locations instead.

1.1.1 Parametric One-way ANOVA

Suppose $\{X_{ij}\}$ are independent random variables sampling from K populations (or groups), where $i = 1, 2, \cdots, K$, $j = 1, 2, \cdots, n_i$. The usual parametric ANOVA aims to test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$ versus $H_a : \mu_i \neq \mu_j$ for any $i \neq j$, where μ_i is the mean of i^{th} population, i.e. $\mu_i = E(X_{ij})$. The usual parametric test, i.e. F-test, relies on the assumptions of independence of observations, normality of the distribution and constancy of variance. Thus, before implementing the analysis of variance, Levene's test [28] and/or Bartlett's test([5], [37]) are usually applied to test the homogeneity of variances. The Levene's test statistics is given by

$$W = \frac{N-K}{K-1} \frac{\sum_{i=1}^{K} n_i (\bar{z}_{i.} - \bar{z}_{..})^2}{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{i.})^2}, \tag{1.1}$$

where

$$\begin{aligned}
N &= \sum_{i=1}^{K} n_i, \\
z_{ij} &= \begin{cases} \left| x_{ij} - \bar{x}_{i.} \right| & \text{if } \bar{x}_{i.} \text{ is a mean of i-th group,} \\ \left| x_{ij} - \tilde{x}_{i.} \right| & \text{if } \tilde{x}_{i.} \text{ is a median of i-th group,} \end{cases} \\
\bar{z}_{i.} &= \frac{1}{n_i} \sum_{j=1}^{n_i} z_{ij}, \text{ and} \\
\bar{z}_{..} &= \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n_i} z_{ij}.
\end{aligned}$$

It is known that Levene's test statistic follows F distribution with degrees of freedom $K-1$ and $N-K$. So we reject the null hypothesis $H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_K$ if test statistics $W > F(\alpha, K-1, N-K)$, where α is the significance level.

Barlett's test is an alternative homogeneity of variances test to Levene's test. Its test statistic is given by

$$T = \frac{(N-K)ln(S_p^2) - \sum_i (n_i - 1) ln(S_i^2)}{1 + \frac{1}{3(K-1)} \left(\sum_i (\frac{1}{n_i - 1}) \frac{1}{N-K} \right)}, \tag{1.2}$$

where, S_i is the sample variance of the i^{th} group and $S_p = \frac{\sum_i (n_i-1) S_i^2}{N-K}$ is the pooled estimate of variance. The test statistic is shown to have approximately a χ^2_{K-1} distribution. Thus the null hypothesis is rejected if the test statistics $T > \chi^2_{K-1,\alpha}$.

Both of these tests are used to detect if the K groups of samples come from populations with an equal variance. Bartlett's test is very sensitive to the departures from normality, while Levene's test does not have the requirement of normality. However, both of them are homogeneity of variance tests. Most of the software packages perform both of them before a comparison of means via classical ANOVA test.

If the homogeneity of variances test is not rejected, then it is appropriate to perform the analysis of variance tests. In the parametric analysis of variance, the total variability (total sum of squared deviations from the mean) from the samples are partitioned into parts. For example, in the one-way ANOVA, the sum of squares is partitioned into two parts: sum of square within the groups (SSW) and sum of square between the groups (SSB). SSW is also called error or residual sum of square, which is measured by $SSW = \sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2$. SSB is usually called explained sum of square, which is given by $SSB = \sum_i n_i(\bar{x}_{i.} - \bar{x}_{..})^2$. The F-test statistic is constructed by taking the ratio of the two sum of squares with the adjustment of the corresponding degrees of freedom. Large ratio indicates large differences between the groups. Thus, the F-test statistic can be rewritten as: [34]

$$F = \frac{MSB}{MSW} = \frac{SSB/(K-1)}{SSW/(N-K)} = \frac{\sum_{i=1}^{K} n_i(\bar{x}_{i.} - \bar{x}_{..})^2/(K-1)}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2/(N-K)}, \tag{1.3}$$

where, $\bar{x}_{i.} = \frac{1}{n_i}\sum_j x_{ij}$ and $\bar{x}_{..} = \frac{1}{N}\sum_i \sum_j x_{ij}$. It follows F-distribution with degrees of freedom $K-1$ and $N-K$. Thus the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$ is rejected when $F > F(\alpha, K-1, N-K)$.

1.1.2 Rank Transformed Nonparametric One-way ANOVA

Conover and Iman [9] proposed the rank transformation procedure as a ‘bridge’ between parametric and nonparametric statistics. Generally speaking, the rank transformation procedure is simply carried out by replacing the numerical raw data with their ranks and applying usual parametric methods, such as ANOVA, regression, discriminant analysis, cluster analysis and so on. This approach includes a class of nonparametric tests for one or more independent samples, such as the Wilcoxon signed rank test for one sample, the Wilcoxon-Mann-Whitney test for two independent samples, the Kruskal-Wallis test for one-way ANOVA and Friedman test for one-way analysis of variance applied to a complete block design.

Consider the K independent random samples $\{X_{ij}\}$ in Section 1.1.1. The rank transformation procedure is implemented as follows: let $R(X_{ij})$ be the rank of X_{ij} in the combined set of data and $R_i = \sum_{j=1}^{n_i} R(X_{ij})$. Then the rank transformed F test statistic is given by

$$F_R = \frac{\left[\sum_{i=1}^{K} \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}\right]/(K-1)}{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(R(X_{ij}) - \frac{R_i}{n_i}\right)^2/(N-K)}, \tag{1.4}$$

which approximately follows a F-distribution [34]. Another rank based nonparametric one-way ANOVA is the well known Kruskal & Wallis H test [23]. The H test statistic, with the correction for ties, is given by

$$H = \frac{\sum_{i=1}^{K} \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4}}{\left(\sum_{i=1}^{K}\sum_{j=1}^{n_i} R^2(X_{ij}) - \frac{N(N+1)^2}{4}\right)/(N-1)}. \tag{1.5}$$

Interestingly, the rank-transformed F test is equivalent to Kruskal-Wallis H test since $F_R = \frac{H/(K-1)}{(N-1-H)/(N-K)}$, which can be easily verified through some elementary algebra. H test statistic is shown to be approximated by a chi-square distribution with degrees of freedom $K-1$ when the sample size is large. The exact distribution of H if there is no tie, is given by Iman, Quade and Alexander [20]. Hence there are two ways to obtain the critical values for the rank transformed F test or Kruskal & Wallis H test: one is to use the F-distribution tables for F_R; the other is to use the chi-square tables for F_R as a function of H. Iman and Davenport [19] compared the two approximations and showed that F approximation is preferred to chi-square in general.

1.2 Two-way Analysis of Variance

1.2.1 Parametric Two-way ANOVA

Let $\{X_{ijk}\}$ be the random variable denoting the response of k^{th} replicate receiving the i^{th} level of treatment A and j^{th} level of treatment B. Consider the two-way layout:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \tag{1.6}$$

where, μ, α_i, β_j, γ_{ij} are the general mean, the effect of the i^{th} level of factor A (i.e. row effect), the effect of the j^{th} level of factor B (i.e. column effect) and the interaction between i^{th}-row and j^{th}-column, respectively; e_{ijk} are independent and identically distributed (i.i.d.) $N(0, \sigma^2)$ for all $k = 1, 2, \cdots, n_{ij}$, $i = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, c$. The hypotheses for testing the row, column and interaction of row and column effects are as follows:

$$\begin{aligned} H_A &: \text{ all } \alpha_i\text{=0,(test for main effect of factor A)} \\ H_B &: \text{ all } \beta_j\text{=0,(test for main effect of factor B)} \\ H_{AB} &: \text{ all } \gamma_{ij}\text{=0(test for interaction).} \end{aligned}$$

Without loss of generality assume that $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$ for all i, j. Similar to the parametric one-way ANOVA, if the homogeneity of variance assumption holds, the total variability from the samples is decomposed into four parts: sum of squares for factor A (SSA), sum of squares for factor B (SSB), sum of squares for the interaction of factor A and B(SSAB) and sum of squares for the error (SSE), where

$$\begin{aligned} SSA &= \sum_{i=1}^{r} n_{i.}(\bar{x}_{i..} - \bar{x}_{...})^2, \\ SSB &= \sum_{j=1}^{c} n_{.j}(\bar{x}_{.j.} - \bar{x}_{...})^2, \\ SSAB &= \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2, \\ SSE &= \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{ij.})^2, \end{aligned}$$

and

$$\begin{aligned}
n_{i.} &= \sum_j n_{ij}, n_{.j} = \sum_i n_{ij}, \sum_i \sum_j n_{ij} = N, \\
\bar{x}_{ij.} &= \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}, \\
\bar{x}_{i..} &= \frac{1}{n_{i.}} \sum_{j=1}^{c} \sum_{k=1}^{n_{ij}} x_{ijk}, \\
\bar{x}_{.j.} &= \frac{1}{n_{.j}} \sum_{i=1}^{r} \sum_{k=1}^{n_{ij}} x_{ijk}, \\
\bar{x}_{...} &= \frac{1}{N} \sum_{i=1}^{r} \sum_{j=1}^{c} \sum_{k=1}^{n_{ij}} x_{ijk}.
\end{aligned}$$

The degrees of freedom for SSA, SSB, SSAB and SSE, are $r-1$, $c-1$, $(r-1)(c-1)$ and $N-rc$ respectively. Thus, the F-test statistic for null hypotheses H_A : all α_i=0 is given by $F_A = \frac{SSA/(r-1)}{SSE/(N-rc)}$, which follows a F-distribution with degrees of freedom $r-1$ and $N-rc$. Similarly, the test statistic for testing the main effect B is $F_B = \frac{SSB/(c-1)}{SSE/(N-rc)}$, which follows a F-distribution with degrees of freedom $c-1$ and $N-rc$. And the F-test statistic for interaction H_{AB} : all γ_{ij}=0 is given by $F_{AB} = \frac{SSAB/((r-1)(c-1))}{SSE/(N-rc)}$, which follows an F-distribution with degrees of freedom $(r-1)(c-1)$ and $N-rc$. The details are discussed in [34].

1.2.2 Rank Transformed Nonparametric Two-way ANOVA

Let X_{ijk}, $i = 1, 2, \cdots, r$, $j = 1, 2, \cdots, c$, $k = 1, 2, \cdots, n$, be independent random variables such that X_{ijk} has the continuous distribution function F_{ij}. Note that $\{X_{ijk}\}$, as defined in Section (1.2.1), is the random variable denoting the response of k^{th} replicate in the (i, j) cell, and can be written in the two-way layout (1.6) as well. We want to test the column effects, row effects and their interaction, just as in the parametric two-way ANOVA. However, in the nonparametric case, the errors e_{ijk} are neither assumed to be normal nor to have homogeneous variance. Thus, for instance, to test the column effect, the null and alternative hypotheses are:

$$
\begin{aligned}
H_0 &: F_{ij} = F_i, \text{ for } j = 1, 2, \cdots, c, \\
H_1 &: F_{ij} \neq F_i, \text{ for at least one } j = 1, 2, \cdots, c.
\end{aligned} \tag{1.7}
$$

This is equivalent to the test:

$$
\begin{aligned}
H_0 &: X_{ijk} = \mu + \alpha_i, \text{ for } j = 1, 2, \cdots, c, \\
H_1 &: X_{ijk} = \mu + \alpha_i + \delta_{ij} \text{ for all } i = 1, 2, \cdots, r,\ j = 1, 2, \cdots, c,
\end{aligned}
$$

or

$$
\begin{aligned}
H_0 &: \delta_{ij} = 0, \text{ for all } i = 1, 2, \cdots, r,\ j = 1, 2, \cdots, c, \\
H_1 &: \delta_{ij} \neq 0 \text{ for some i, j},
\end{aligned} \tag{1.8}
$$

where,

$$
\begin{aligned}
\delta_{ij} &= \beta_j + \gamma_{ij}, \\
\sum_i \alpha_i &= \sum_j \delta_{ij} = 0,
\end{aligned}
$$

and β_j is the effect of j^{th} level of Factor B as describe in equation(1.6).

In general, two rank-based approaches are commonly used to test either (1.7) or (1.8): Hora & Conover rank-score transformed test and Akritas rank transformed test. Let R_{ijk} be the rank of X_{ijk} among all $\{X_{ijk}\}$ data. Define the score of X_{ijk} by $a(R_{ijk})$, where $a(\cdot)$ is the rank score function as in [16]. Set $S_{i..} = \sum_j \sum_k a(R_{ijk})$, $S_{.j.} = \sum_i \sum_k a(R_{ijk})$, $S_{ij.} = \sum_k a(R_{ijk})$ and $\bar{S} = \frac{1}{c} \sum_i \sum_j \sum_k a(R_{ijk})$. The rank-score transformed test statistic proposed by Hora & Conover [18] is given by

$$
W_N = \frac{\sum_j (S_{.j.} - \bar{S})^2 / (n(c-1))}{\sum_i \sum_j \sum_k (a(R_{ijk}) - S_{ij.}/n)^2 / (c(n-1))}, \tag{1.9}
$$

which converges weakly to χ^2_{c-1}, as $n \to \infty$, under the null hypotheses. Under a sequence of Pitman alternatives, the limiting distribution of W_N is normal [40].

Rinaman [33] proposed a similar test statistic with its limiting distribution χ^2_{c-1} as well, although R_{ijk} is redefined as the rank of X_{ijk} in row i instead of the whole samples and a new score function is used.

Unlike Hora & Conover's rank-score transformed approach, Akritas [4] adjusted the rank transformed data for heteroscedasticity by defining $z_{ijk} = \frac{\hat{G}(X_{ijk})}{\hat{\sigma}_i}$, where $\hat{G}(X_{ijk}) = R_{ijk}/(rcn)$, $\hat{\sigma}_i = \frac{1}{nc}\sum_j \sum_k \left(\hat{G}(X_{ijk}) - \hat{AG}_{i..}\right)^2$, $\hat{AG}_{i..} = \frac{1}{nc}\sum_j \sum_k \hat{G}(X_{ijk})$ and R_{ijk} is the same as defined in Hora & Conover's method. The rank transformed test statistic proposed by Akritas [4] is as follows:

$$W_A = \frac{n\sum_i \sum_j (\bar{z}_{ij.} - \bar{z}_{i..})^2/(r(c-1))}{\sum_i \sum_j \sum_k (z_{ijk} - \bar{z}_{ij.})^2/(rc(n-1))}, \tag{1.10}$$

which asymptotically follows central $\chi^2_{r(c-1)}/(r(c-1))$ distribution under H_0. The major advantage of Akritas rank transformed approach is not the adjustment of heteroscedasticity ([4] showed that it is not necessary for Hora and Conover statistic to be adjusted for heteroscedasticity), but rather the simple extension to unbalanced data. Suppose X_{ijk} are independent random variables, where $i = 1, 2, \cdots, r$, $j = 1, 2, \cdots, c_i$, $k = 1, 2, \cdots, n_{ij}$. The Akritas test statistic for unbalanced data is almost the same as the balanced one, and it is given by

$$W_A^U = \frac{\sum_i \sum_j n_{ij}(\bar{z}_{ij.} - \bar{z}_{i..})^2/(C-r)}{\sum_i \sum_j \sum_k (z_{ijk} - \bar{z}_{ij.})^2/(N-C)}, \tag{1.11}$$

where, $N = \sum_i \sum_j n_{ij}$ and $C = \sum_i c_i$. It is shown that W_A^U asymptotically follows a central $\chi^2_{C-r}/(C-r)$ distribution under H_0.

The common limitation of the Hora & Conover's W_N test, Akritas W_A test and other rank transformed two-way ANOVA test such as Lemmer & Stoker [27], de Kroon & van der Lann [11] is that they either ignore the test of interaction ($H_0 : \gamma_{ij} = 0$ for all i, j), or fail to test the main effect in the presence of interaction. Blair, Sawilowsky and Higgins [7] verified the fact that the Hora & Conover rank-score transform test is robust when testing for main effects in the absence of interaction; however, the test may produce inflated Type I error rates when testing for main effect in the presence of interaction.

The test of interaction in rank transformed based nonparametric two-way ANOVA challenges researchers when they try to extend one-way ANOVA to the two-way case. Until recently, there are several methods to test the interaction, i.e. $H_0 : \gamma_{ij} = 0$ for all i, j. Patel and Hoel [31] defined their own interaction and derived the nonparametric test on a special case. Bhapkar and Gore [6] introduced a nonparametric test based on the U-statistics under the orthogonality assumption. [6] formed quadruplets of cells, say, x_{ij},$x_{i'j}$,$x_{ij'}$,$x_{i'j'}$ from the $r \times c$ matrix $\mathbf{X}$, and defined the function $\phi(t)$ to be 1 if $t > 0$, 1/2 if $t = 0$ and zero otherwise. Suppose each quadruplet have $(n_{ij}, n_{i'j}, n_{ij'}, n_{i'j'})$ quadruplet of observations. Let $V_{i,i',j,j'} = \sum_a \sum_b \sum_c \sum_d \phi(x_{ija} - x_{i'jb} - x_{ij'c} + x_{i'j'd})$, $U_{i,i',j,j'} = V_{i,i',j,j'}/n_{ij}n_{i'j}n_{ij'}n_{i'j'}$, $W_{ij} = \sum_{i \neq i'}^{r} \sum_{j \neq j'}^{c} U_{i,i',j,j'}$, $N = \sum_{i,j} n_{ij}$ and $p_{ij} = n_{ij}/N$. By assuming $p_{ij} = a_i b_j$, and defining $W_{i.} = \frac{\sum b_j W_{ij}}{\sum_j b_j}$, $W_{.j} = \frac{\sum a_i W_{ij}}{\sum_i a_i}$ and $W_{..} = \sum_{i,j} a_i b_j W_{ij}$, the test statistic proposed by Bhapkar and Gore [6] can be written as

$$T = \frac{N}{r^2 c^2 (\hat{\eta}(F) - 1/4)} \sum_i \sum_j a_i b_j (W_{ij} - W_{i.} - W_{.j} + W_{..})^2, \qquad (1.12)$$

where, $\hat{\eta}(F)$ is a consistent estimator of a nuisance parameter defined in ([6], equation (2.3)). It is shown that T follows asymptotically a central chi-square distribution with $(r-1)(c-1)$ degrees of freedom. However, this test statistic relies on unknown nuisance parameter, which might reduce the power. Hartlaub, Dean and Wolfe [17] proposed a rank transformed based test for interaction in two-way ANOVA with only one observation per cell. Unfortunately, two-way layouts with replication in cells are more frequent in practice. Gao and Alvo [15] proposed a rank transformed method which combines the row rankings and column rankings, while other literature just consider either one of the rankings or the rankings based on the whole data set. Suppose r_{ijk} is the rank of X_{ijk} with respect to the i^{th} row and c_{ijk} is the rank of X_{ijk}

with respect to the j^{th} column, where $i = 1, 2, \cdots, r$, $j = 1, 2, \cdots, c$, $k = 1, 2, \cdots, n$. Define

$$\begin{aligned} S_n(i,j) &= \frac{1}{nc+1}\sum_k r_{ijk}, \\ T_n(i,j) &= \frac{1}{nr+1}\sum_k c_{ijk}, \\ S_n &= (S_n(1,1), S_n(1,2), \cdots, S_n(r,c))', \\ T_n &= (T_n(1,1), T_n(1,2), \cdots, T_n(r,c))', \\ \Sigma_1 &= \lim_{n\to\infty} \frac{1}{n} var(S_n), \\ \Sigma_2 &= \lim_{n\to\infty} \frac{1}{n} var(T_n), \\ \Sigma_{12} &= \lim_{n\to\infty} \frac{1}{n} cov(S_n, T_n). \end{aligned}$$

Then the statistic Gao and Alvo [15] proposed is given by:

$$W = \frac{1}{n}(\mathbf{AS_n} + \mathbf{BT_n})'(\mathbf{A\hat{\Sigma}_1A'} + \mathbf{2A\hat{\Sigma}_{12}B'} + \mathbf{B\hat{\Sigma}_2B'})^{-}(\mathbf{AS_n} + \mathbf{BT_n}), \qquad (1.13)$$

where $\mathbf{A} = \mathbf{J_r} \bigotimes (-\frac{\mathbf{1}}{\mathbf{r}}\mathbf{I_c}) + \mathbf{I_r} \bigotimes \mathbf{I_c}$, $\mathbf{B} = \mathbf{I_r} \bigotimes (\mathbf{I_c} - \frac{\mathbf{1}}{\mathbf{c}}\mathbf{J_c})$, $\hat{\Sigma}_1$, $\hat{\Sigma}_2$, $\hat{\Sigma}_{12}$ are the corresponding consistent estimates of Σ_1, Σ_2 and Σ_{12}. It is shown that W follows a central chi-squared distribution with $(r-1)(c-1)$ degrees of freedom as $n \to \infty$. Actually, Gao and Alvo's test statistics can be easily extended to unbalanced designs.

1.3 Limitations

The limitations for parametric ANOVA, no matter one-way or two-way layout, are very obvious. It assumes normality and homogeneity of variances. None of the data in our real world comes from an exact normal distribution. By the robustness of F-test, it is still reliable to perform the test when the sample size is large and the population distribution is not too far from normal. However, these assumptions sometimes are too strict for the data in real research settings. In these circumstances, nonparametric techniques should be applied instead. Until now, almost all the nonparametric

ANOVA tests are based on rank (score) transformed technique. The major drawback of using the ranks, rather than the raw data, is that it loses information. The rank keeps the order of the raw data, but it ignores the magnitude of the differences among the data. There could be two sets of interval data with exactly the same rank, but with totally different means, variances or distributions. If we throw away the raw data by analyzing the ranks instead, no difference will be detected for these two groups of data sets. Unfortunately, none of the rank transformed techniques can compensate this loss. Moreover, the majority of the literature works in nonparametric ANOVA try to express and interpret their models in the same way as the parametric models, even when it is inappropriate to do so. For example, $x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$ is the two-way layout in the parametric ANOVA. In the nonparametric ANOVA, x_{ijk} does not have to come from normal distribution any more, but the literature that promote the rank transformed approaches still interpret μ as the grand mean and $\mu + \alpha_i + \beta_j + \gamma_{ij}$ as the mean of cell (i, j), even if for some distributions, the first moment does not exist.

1.4 Kernel Density Estimate

1.4.1 Kernel Estimate of Probability Density Function $f(x)$

Let $F(x)$ be the cumulative distribution function (CDF) of the random variable X, where X is a random variable with probability density function (pdf) $f(x)$. From the definition that $f(x) = \frac{d}{dx}F(x)$, an obvious estimate of $f(x)$ is

$$\hat{f}(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}, \tag{1.14}$$

where $F_n(x)$ is empirical cumulative distribution function defined by

$$F_n(x) = \frac{\text{the number of } X_i\text{'s such that } X_i \leq x}{n}. \tag{1.15}$$

If we define a uniform kernel function by

$$K(z) = \begin{cases} 1/2 & \text{if } |z| \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1.16}$$

then by substituting (1.15) into (1.14), we can rewrite (1.14) as

$$\begin{aligned} \hat{f}(x) &= \frac{\text{the number of } X_i\text{'s fall within } [x-h, x+h]}{2nh} \\ &= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right). \end{aligned} \tag{1.17}$$

Equation (1.17) is a "naive" kernel estimator of $f(x)$ if $K(\cdot)$ is defined in (1.16). In general, we refer to $K(\cdot)$ as a kernel function and to h as a smoothing parameter (or alternative a bandwidth or window width). The kernel function is not limited to equation (1.16), but rather has many other possible choices, such as standard normal kernel, i.e.

$$K(z) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}z^2}. \tag{1.18}$$

It is shown in [29] that the kernel estimator $\hat{f}(x)$ defined in (1.17) with any general non-negative bounded kernel function $K(\cdot)$ satisfying 3 conditions, i.e.

(a) $\int K(z)dz = 1$,

(b) $K(z) = K(-z)$,

(c) $\int z^2 K(z)dz = \kappa_2 > 0$,

is a consistent estimator of $f(x)$. The literature for the kernel estimate is very rich including the books by Wand and Jones, Silverman, Bowman and Azzalini, and Scott ([41], [36], [35], [8]) among others.

1.4.2 Kernel Estimate of $\int f^2(x)dx$

With the kernel estimate of $f(x)$ in Section (1.4.1), $\int f^2(x)dx$ can be estimated by Ahmad [1],

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

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

1.5 Organization of the Dissertation

In Chapter 2, a kernel based nonparametric one-way ANOVA test is proposed. In Section 2.1, the one-way kernel based nonparametric scale test is derived to test the homogeneity of scale parameters among groups. In Section 2.2, the kernel based test statistic for nonparametric one-way ANOVA with homogeneous scale parameter among groups is constructed and its limiting distribution is studied as well. In Section 2.3, the powers of the kernel based nonparametric scale and location test are investigated through simulation and compared to the corresponding parametric tests. The kernel based nonparametric one-way ANOVA test for shape parameters are proposed in Section 2.4.

In Chapter 3, the kernel based nonparametric one-way ANOVA test to the two-way layout is extended. Section 3.1 has two subsections. In Section 3.1.1, the kernel based nonparametric ANOVA test for main effects in locations is derived under the assumption of homogeneous scale among cells. In Section 3.2.2, the kernel based nonparametric ANOVA test for interactions in locations is studied under the assumption of homogeneous scale among cells. In Section 3.3, the powers of the kernel based nonparametric two-way ANOVA test of main effects and interactions in locations are investigated via simulation study and compared with the corresponding parametric two-way ANOVA tests.

In Chapter 4, the kernel based nonparametric ANOVA model is applied to the nonparametric policy analysis. In Section 4.1, a brief introduction of policy analysis is given. In Section 4.2, Stock's nonparametric policy analysis model is introduced and the limitations of stock's model are summarized. In Section 4.3, a new nonparametric policy analysis model is proposed by extending our nonparametric ANOVA results in Section 2.2.

Finally, Section 5.1 summarizes the contributions of this dissertation and the major findings of the simulation studies. Section 5.2 outlines some possible future research topics followed by the dissertation.

CHAPTER 2

One-way Kernel Based Nonparametric ANOVA

Assume X_{ij} comes from a distribution with probability density function (p.d.f.) $f_i(x)$, where $i = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, n_i$. Define μ_i and σ_i to be the location and scale parameter of $f_i(x)$. That is to say, the following equation holds:

$$f_i(x) = \frac{1}{\sigma_i} f_0\left(\frac{x - \mu_i}{\sigma_i}\right), \tag{2.1}$$

where, $f_0(\cdot)$ is a base density. Thus, we have

$$\begin{aligned}\int x f_i^2(x)dx &= \int \frac{y\sigma_i + \mu_i}{\sigma_i} f_0^2(y)dy \\ &= \int y f_0^2(y)dy + \frac{\mu_i}{\sigma_i}\int f_0^2(y)dy,\end{aligned}$$

which implies that

$$\begin{aligned}\mu_i &= \frac{\sigma_i(\int x f_i^2(x)dx - \int y f_0^2(y)dy)}{\int f_0^2(y)dy} \\ &= \frac{\sigma_i \int x f_i^2(x)dx}{\int f_0^2(y)dy} - \frac{\sigma_i \int y f_0^2(y)dy}{\int f_0^2(y)dy}.\end{aligned} \tag{2.2}$$

By Ahmad and Amezziane [2], the scale parameters can be written as:

$$\sigma_i = \frac{\int_{-\infty}^{\infty} f_0^2(x)dx}{\int_{-\infty}^{\infty} f_i^2(x)dx}. \tag{2.3}$$

2.1 Kernel Based Nonparametric Test for Scale Parameters

Before introducing the ANOVA test, the homogeneity of scale parameters needs to be verified. Suppose that σ_1, σ_2, $\cdots$, σ_K are the scale parameters of the K populations. The hypotheses to be tested are H_0: $\sigma_1 = \sigma_2 = \ldots = \sigma_K$ versus H_1: $\sigma_i \neq \sigma_j$ for some $i \neq j$.

Let $R_0 = \int_{-\infty}^{\infty} f_0^2(x)dx$ and $R_i = \int_{-\infty}^{\infty} f_i^2(x)dx$, then equation(2.3) can be written as $R_i = R_0/\sigma_i$. Thus, the null hypothesis of equal scale parameters becomes H_0: $R_1 = R_2 = \ldots = R_K$ against H_1: $R_i \neq R_j$ for some $i \neq j$. Consider the nonparametric kernel estimates of R_i, denoted as $\hat{R}_i$, where

$$\hat{R}_i = \frac{1}{n_i(n_i-1)h_i} \sum \sum_{j_1 \neq j_2} K\left(\frac{X_{ij_1} - X_{ij_2}}{h_i}\right). \tag{2.4}$$

Lemma 2.1 *If for any $i = 1, 2, \cdots, K$, $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ as $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} f_i^3(x)dx < \infty$ and if $f_i(\cdot)$ is twice differentiable, then*

$$\sqrt{n_i}(\hat{R}_i - R_i) \xrightarrow{d} N(0, \upsilon_i^2), \tag{2.5}$$

as $\min_i n_i \to \infty$, where, $\upsilon_i^2 = 4\{\int f_i^3(x)dx - (\int f_i^2(x)dx)^2\}$.

Proof: See the proof of Theorem 2.2 in [1].

■

The test statistic for the scale test is defined as

$$S_1 = \sum_{i=1}^{K} \frac{n_i(\hat{R}_i - \hat{R}.)^2}{\hat{\upsilon_i}^2}, \tag{2.6}$$

where

$$\hat{R}. = \frac{\sum_{i=1}^{K} n_i \hat{R}_i / \hat{\upsilon_i}^2}{\sum_{i=1}^{K} n_i / \hat{\upsilon_i}^2}, \tag{2.7}$$

and $\hat{\upsilon_i}^2$ is a consistent estimate of υ_i^2. To obtain the asymptotic distribution of S_1, another two auxiliary variables are defined as follows:

$$S_1^0 = \sum_{i=1}^{K} \frac{N\lambda_i(\hat{R}_i - \bar{R})^2}{\upsilon_i^2} - N\left[\sum_{i=1}^{K} \frac{\lambda_i}{\upsilon_i^2}\right](R^* - \bar{R})^2, \tag{2.8}$$

and

$$S_1^{00} = \sum_{i=1}^{K} \frac{n_i(\hat{R}_i - \bar{R})^2}{\hat{\upsilon_i}^2} - \sum_{i=1}^{K} \frac{n_i(\hat{R}. - \bar{R})^2}{\hat{\upsilon_i}^2}, \tag{2.9}$$

where

$$\begin{aligned} R^* &= \frac{\sum_{i=1}^K \lambda_i \hat{R}_i/\upsilon_i^2}{\sum_{i=1}^K \lambda_i/\upsilon_i^2}, \\ \bar{R} &= \frac{\sum_{i=1}^K \lambda_i R_i/\upsilon_i^2}{\sum_{i=1}^K \lambda_i/\upsilon_i^2}, \\ \lambda_i &= \lim_{n_i\to\infty} \frac{n_i}{N}. \end{aligned}$$

Lemma 2.2 *Let* $N = \sum_{i=1}^K n_i$. *If* $\lambda_i = \lim_{n_i\to\infty} \frac{n_i}{N}$ *and* $\hat{\upsilon_i}^2 \xrightarrow{p} \upsilon_i^2$, *then*

(i) $S_1^0 - S_1^{00} \xrightarrow{p} 0$,

(ii) $S_1^{00} - S_1 \xrightarrow{p} 0$,

as $\min_i n_i \to \infty$.

Proof: They can be proved directly by applying Slutsky Theorems [10].

■

Theorem 2.3 *Under the null hypothesis, if for any* $i = 1, 2, \ldots, K$, $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ *as* $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} f_i^3(x)dx < \infty$ *and if* $f_i(\cdot)$ *is twice differentiable, then* S_1 *is asymptotically* χ^2 *with degrees of freedom* $K-1$. *In general (under the alternative),* S_1 *is asymptotically non-central* $\chi^2(K-1)$ *with non-centrality parameter:*

$$\psi_1 = \frac{1}{2}\mu^{(\mathbf{1})'}\mathbf{B_1}\mu^{(\mathbf{1})}, \tag{2.10}$$

where $\mu^{(\mathbf{1})} = (\mu_1^{(1)}, \mu_2^{(1)}, \ldots, \mu_K^{(1)})$,

$$\mu_i^{(1)} = -\frac{\sqrt{\lambda_i}(2d_i + \frac{\sum_{i=1}^K \lambda_i d_i/\upsilon_i^2}{\sum_{i=1}^K \lambda_i/\upsilon_i^2}) \int_{-\infty}^{\infty} f_0^2(x)dx}{2\sqrt{\int f_i^3(x)dx - (\int_{-\infty}^{\infty} f_i^2(x)dx)^2}}, \tag{2.11}$$

$$\mathbf{B_1} = \begin{bmatrix} 1 - \frac{\lambda_1/\upsilon_1^2}{\sum_{i=1}^K \lambda_i/\upsilon_i^2} & -\frac{(\sqrt{\lambda_1}/\upsilon_1)(\sqrt{\lambda_2}/\upsilon_2)}{\sum_{i=1}^K \lambda_i/\upsilon_i^2} & \cdots & -\frac{(\sqrt{\lambda_1}/\upsilon_1)(\sqrt{\lambda_K}/\upsilon_K)}{\sum_{i=1}^K \lambda_i/\upsilon_i^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(\sqrt{\lambda_1}/\upsilon_1)(\sqrt{\lambda_K}/\upsilon_K)}{\sum_{i=1}^K \lambda_i/\upsilon_i^2} & -\frac{(\sqrt{\lambda_2}/\upsilon_2)(\sqrt{\lambda_K}/\upsilon_K)}{\sum_{i=1}^K \lambda_i/\upsilon_i^2} & \cdots & 1 - \frac{\lambda_K/\upsilon_K^2}{\sum_{i=1}^K \lambda_i/\upsilon_i^2} \end{bmatrix}, \tag{2.12}$$

$\lambda_i = \lim_{n_i\to\infty} \frac{n_i}{\sum_{i=1}^K n_i}$ *and* d_i *is such that* $\sigma_i = 1 + \frac{d_i}{\sqrt{\sum_{i=1}^K n_i}}$.

Proof: Let $N = \sum_{i=1}^{K} n_i$, and then $\lambda_i \simeq \frac{n_i}{N}$. Set $T_i^{(1)} = \sqrt{\lambda_i N}(\hat{R}_i - \bar{R})/\upsilon_i$, and then under H_0, $T_i^{(1)} \overset{a}{\sim} N(0,1)$ as $N \to \infty$ by Lemma (2.1).

Note that

$$\begin{aligned} S_1^0 &= \sum_{i=1}^{K} \frac{N\lambda_i(\hat{R}_i - \bar{R})^2}{\upsilon_i^2} - N\Big[\sum_{i=1}^{K} \frac{\lambda_i}{\upsilon_i^2}\Big](R^* - \bar{R})^2 \\ &= \sum_{i=1}^{K} \frac{N\lambda_i(\hat{R}_i - \bar{R})^2}{\upsilon_i^2} - N\Big[\sum_{i=1}^{K} \frac{\lambda_i}{\upsilon_i^2}\Big]\left(\frac{\sum_i \lambda_i \hat{R}_i/\upsilon_i^2}{\sum_i \lambda_i/\upsilon_i^2} - \frac{\sum_i \lambda_i R_i/\upsilon_i^2}{\sum_i \lambda_i/\upsilon_i^2}\right)^2 \\ &= \sum_{i=1}^{K} \frac{N\lambda_i(\hat{R}_i - \bar{R})^2}{\upsilon_i^2} - \frac{N}{\sum_{i=1}^{K} \frac{\lambda_i}{\upsilon_i^2}}\left(\sum_{i=1}^{K} \frac{\lambda_i(\hat{R}_i - R_i)}{\upsilon_i^2}\right)^2 \\ &= \sum_{i=1}^{K} \frac{N\lambda_i(\hat{R}_i - \bar{R})^2}{\upsilon_i^2} - \sum_{i=1}^{K}\sum_{j=1}^{K} \frac{\sqrt{N\lambda_i}(\hat{R}_i - \bar{R})}{\upsilon_i} \frac{\sqrt{N\lambda_j}(\hat{R}_j - \bar{R})}{\upsilon_j} \frac{(\sqrt{\lambda_i}/\upsilon_i)(\sqrt{\lambda_j}/\upsilon_j)}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2}, \end{aligned} \tag{2.13}$$

which can be written as a quadratic form, i.e. $S_1^0 = \mathbf{U_1}'\mathbf{B_1}\mathbf{U_1}$, where,

$$\mathbf{U_1} = (T_1^{(1)}, T_2^{(1)}, \cdots, T_K^{(1)})', \tag{2.14}$$

and

$$\mathbf{B_1} = \begin{bmatrix} 1 - \frac{\lambda_1/\upsilon_1^2}{\sum_{i=1}^{K}\lambda_i/\upsilon_i^2} & -\frac{(\sqrt{\lambda_1}/\upsilon_1)(\sqrt{\lambda_2}/\upsilon_2)}{\sum_{i=1}^{K}\lambda_i/\upsilon_i^2} & \cdots & -\frac{(\sqrt{\lambda_1}/\upsilon_1)(\sqrt{\lambda_K}/\upsilon_K)}{\sum_{i=1}^{K}\lambda_i/\upsilon_i^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(\sqrt{\lambda_1}/\upsilon_1)(\sqrt{\lambda_K}/\upsilon_K)}{\sum_{i=1}^{K}\lambda_i/\upsilon_i^2} & -\frac{(\sqrt{\lambda_2}/\upsilon_2)(\sqrt{\lambda_K}/\upsilon_K)}{\sum_{i=1}^{K}\lambda_i/\upsilon_i^2} & \cdots & 1 - \frac{\lambda_K/\upsilon_K^2}{\sum_{i=1}^{K}\lambda_i/\upsilon_i^2} \end{bmatrix}. \tag{2.15}$$

It can be easily shown that $\mathbf{B}_1$ is symmetric and idempotent. Thus, we obtain that

$$\begin{aligned} rank(\mathbf{B_1}) &= tr(\mathbf{B_1}) \\ &= tr(\mathbf{I} - \mathbf{a_1}\mathbf{a_1}') \\ &= tr(\mathbf{I}) - tr(\mathbf{a_1}\mathbf{a_1}') \\ &= K - tr(\mathbf{a_1}'\mathbf{a_1}) \\ &= K - 1, \end{aligned} \tag{2.16}$$

where

$$\mathbf{a_1} = \begin{pmatrix} \sqrt{\lambda_1}/\upsilon_1 \\ \sqrt{\lambda_2}/\upsilon_2 \\ \vdots \\ \sqrt{\lambda_K}/\upsilon_K \end{pmatrix}. \tag{2.17}$$

$\mathbf{U_1}$ follows approximately multivariate normal with mean $\mathbf{0}$ and variance $\mathbf{I}$, since $\{T_i^{(1)}\}$ is independent and asymptotically distributed univariate standard normal distribution. Therefore, S_1 is asymptotically $\chi^2(K-1)$ under H_0.

Under the alternative, since d_i is chosen such that $\sigma_i = 1 + \frac{d_i}{\sqrt{N}}$, then we have

$$\begin{aligned} R_i &= \frac{R_0}{\sigma_i} = \frac{R_0}{1 + d_i/\sqrt{N}} \\ &= R_0(1 - \frac{d_i}{\sqrt{N}} + o(\frac{1}{\sqrt{N}})) \\ &= R_0 - R_0\frac{d_i}{\sqrt{N}} + o(\frac{1}{\sqrt{N}}). \end{aligned}$$

Thus, $\bar{R}$ can be written as:

$$\begin{aligned} \bar{R} &= \frac{\sum_{i=1}^{K} \lambda_i R_i/\upsilon_i^2}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2} \\ &= \frac{\sum_{i=1}^{K} \lambda_i (R_0 - R_0\frac{d_i}{\sqrt{N}})/\upsilon_i^2}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2} + o(\frac{1}{\sqrt{N}}) \\ &= R_0 - \frac{R_0}{\sqrt{N}}\frac{\sum_{i=1}^{K} \lambda_i d_i/\upsilon_i^2}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2} + o(\frac{1}{\sqrt{N}}). \end{aligned}$$

Under H_1, $T_i^{(1)} = \sqrt{N\lambda_i}(\hat{R}_i - \bar{R})/\upsilon_i = \frac{\sqrt{N\lambda_i}(\hat{R}_i - R_i)}{\upsilon_i} + \frac{\sqrt{N\lambda_i}(R_i - \bar{R})}{\upsilon_i}$. By Lemma (2.1), the first term $\frac{\sqrt{N\lambda_i}(\hat{R}_i - R_i)}{\upsilon_i}$ is approximately standard normally distribution, and the second term's limit is given by

$$
\begin{aligned}
\mu_i^{(1)} &= \lim_{N\to\infty} \frac{\sqrt{N\lambda_i}(R_i - \bar{R})}{\upsilon_i} \\
&= \lim_{N\to\infty} \frac{\sqrt{N\lambda_i}(\int_{-\infty}^{\infty} f_i^2(x)dx - R_0 - \frac{R_0}{\sqrt{N}} \frac{\sum_{i=1}^{K} \lambda_i d_i/\upsilon_i^2}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2})}{\upsilon_i} \\
&= \lim_{N\to\infty} \frac{\sqrt{N\lambda_i}(\frac{1}{\sigma_i^2} \int_{-\infty}^{\infty} f_0^2(x)dx - \int_{-\infty}^{\infty} f_0^2(x)dx(1 + \frac{1}{\sqrt{N}} \frac{\sum_{i=1}^{I} \lambda_i d_i/\upsilon_i^2}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2}))}{\upsilon_i} \\
&= \lim_{N\to\infty} \frac{\sqrt{N\lambda_i}(\frac{1}{(1+d_i/\sqrt{N})^2} - 1 - \frac{1}{\sqrt{N}} \frac{\sum_{i=1}^{K} \lambda_i d_i/\upsilon_i^2}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2}) \int_{-\infty}^{\infty} f_0^2(x)dx}{\upsilon_i} \\
&= \lim_{N\to\infty} \frac{\sqrt{N\lambda_i}(1 - 2\frac{d_i}{\sqrt{N}} - 1 - \frac{1}{\sqrt{N}} \frac{\sum_{i=1}^{K} \lambda_i d_i/\upsilon_i^2}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2}) \int_{-\infty}^{\infty} f_0^2(x)dx}{2\sqrt{\int f_i^3(x)dx - (\int_{-\infty}^{\infty} f_i^2(x)dx)^2}} \\
&= -\frac{\sqrt{\lambda_i}(2d_i + \frac{\sum_{i=1}^{K} \lambda_i d_i/\upsilon_i^2}{\sum_{i=1}^{K} \lambda_i/\upsilon_i^2}) \int_{-\infty}^{\infty} f_0^2(x)dx}{2\sqrt{\int f_i^3(x)dx - (\int_{-\infty}^{\infty} f_i^2(x)dx)^2}}.
\end{aligned}
\tag{2.18}
$$

Therefore, in general $T_i^{(1)}$ is approximately normally distribution with mean $\mu_i^{(1)}$ given in equation(2.17) and variance 1. This implies that $S_1^0 = \mathbf{U_1}'\mathbf{B_1}\mathbf{U_1}$ is asymptotically noncentral $\chi^2(K-1)$ with non-centrality parameter $\psi_1 = \frac{1}{2}\mu^{(\mathbf{1})'}\mathbf{B_1}\mu^{(\mathbf{1})}$.

■

2.2 One-way ANOVA: Kernel Based Nonparametric Test for Location Parameters with Equal Scale Parameter

Under the assumption of homogeneity of variance, i.e. $\sigma_i = \sigma$ for all i. Thus, equation (2.2) becomes:

$$\mu_i = \frac{\sigma \int x f_i^2(x)dx}{\int f_0^2(y)dy} - \frac{\sigma \int y f_0^2(y)dy}{\int f_0^2(y)dy}.$$

Let $V_i = \int x f_i^2(x)dx$, $c_1 = \frac{\sigma}{\int f_0^2(y)dy}$ and $c_2 = -\frac{\sigma \int y f_0^2(y)dy}{\int f_0^2(y)dy}$, thus $\mu_i = c_1 V_i + c_2$. Therefore, the null hypothesis of equal location parameter H_0: $\mu_1 = \mu_2 = \ldots = \mu_K$ becomes H_0: $V_1 = V_2 = \ldots = V_K$ against H_1: $V_i \neq V_j$ for some $i \neq j$. Consider the nonparametric kernel estimates of V_i, denoted as $\hat{V}_i$, where

$$\hat{V}_i = \frac{1}{n_i(n_i-1)h_i} \sum \sum_{j_1 \neq j_2} \left(\frac{X_{ij_1} + X_{ij_2}}{2}\right) K\left(\frac{X_{ij_1} - X_{ij_2}}{h_i}\right). \tag{2.19}$$

Lemma 2.4 *If for any* $i = 1, 2, \cdots, K$, $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ *as* $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} x^2 f_i^3(x)dx < \infty$ *and if* $f_i(\cdot)$ *is twice differentiable, then*

$$\sqrt{n_i}(\hat{V}_i - V_i) \xrightarrow{d} N(0, \omega_i^2), \tag{2.20}$$

as $\min_i n_i \to \infty$, *where,* $\omega_i^2 = 4\{\int x^2 f_i^3(x)dx - (\int x f_i^2(x)dx)^2\}$.

Proof: The proof follows the lines used to prove Theorem 2.1 in [2]. We include it for completeness.

Let $\varphi(X_{ij_1}, X_{ij_2}) = \left(\frac{X_{ij_1}+X_{ij_2}}{2h_i}\right) K\left(\frac{X_{ij_1}-X_{ij_2}}{h_i}\right)$, and $\hat{\mu_i} = c_1 \hat{V}_i + c_2$. Then $\hat{\mu_i}$ is a U-statistics with mean

$$
\begin{aligned}
E(\hat{\mu_i}) &= c_1 E(\hat{V_i}) + c_2 \\
&= c_1 E(\frac{X_{i1}}{h_i} K\left(\frac{X_{i1} - X_{i2}}{h_i}\right)) + c_2 \\
&= c_1 \frac{1}{h_i} \int\int xK\left(\frac{x-y}{h_i}\right) f_i(x) f_i(y) dxdy + c_2 \\
&= c_1 \int\int xK(u) f_i(x) f_i(x + uh_i) dxdu + c_2 \\
&= c_1 \int\int xK(u) f_i(x) \big(f_i(x) + f_i^{(1)}(x) uh_i + o(h_i^2) dxdu\big) + c_2 \\
&= c_1 \int K(u) du \int x f_i^2(x) dx + c_2 + O(h_i) \\
&= c_1 \int x f_i^2(x) dx + c_2 + O(h_i) \\
&\simeq \mu_i, \qquad (2.21)
\end{aligned}
$$

and variance written as

$$
Var(\hat{\mu_i}) = c_1^2 \left(\frac{4}{n_i} cov(\varphi(X_{i1}, X_{i2}), \varphi(X_{i1}, X_{i3})) + \frac{2}{n_i(n_i - 1)} var(\varphi(X_{i1}, X_{i2})) \right).
$$

It can easily be shown that $var(\varphi(X_{i1}, X_{i2})) = O(h_i^{-1})$ and since $1/(n_i h_i) = o(1)$, the second term of $Var(\hat{\mu_i})$ in the parentheses can be neglected. Then the variance of $\hat{\mu_i}$ is dominated by $\frac{1}{n_i}\varpi_i^2$, where

$$
\begin{aligned}
\varpi_i^2 &= 4c_1^2 cov(\varphi(X_{i1}, X_{i2}), \varphi(X_{i1}, X_{i3})) \\
&= 4c_1^2 [E(\varphi(X_{i1}, X_{i2})\varphi(X_{i1}, X_{i3})) - E(\varphi(X_{i1}, X_{i2}))E(\varphi(X_{i1}, X_{i3}))] \\
&= 4c_1^2 \Big[\frac{1}{4h_i^2}\int\int\int (x+y)(x+z)K(\frac{x-y}{h_i})K(\frac{x-z}{h_i})f_i(x)f_i(y)f_i(z)dxdydz \\
&\quad -\Big(E\Big(\frac{X_{i1}}{h_i}K\Big(\frac{X_{i1}-X_{i2}}{h_i}\Big)\Big)\Big)^2\Big] \\
&= c_1^2\Big[\int\int\int (2x+uh_i)(2x+vh_i)K(u)K(v)f_i(x)f_i(x+uh_i)f_i(x+vh_i)dxdudv \\
&\quad -4\Big(\frac{1}{h_i}\int\int xK\Big(\frac{x-y}{h_i}\Big)f_i(x)f_i(y)dxdy\Big)^2\Big] \\
&= c_1^2\Big[\int\int\int (4x^2+2uxh_i+2vxh_i+uvh_i^2)K(u)K(v)f_i(x)\big(f_i(x)+f_i^{(1)}(x)uh_i+o(h_i^2)\big) \\
&\quad \big(f_i(x)+f_i^{(1)}(x)vh_i+o(h_i^2)\big)dxdudv - 4\Big(\int\int xK(u)f_i(x)f_i(x+uh_i)dxdu\Big)^2\Big] \\
&= c_1^2\Big[4\int x^2 f_i^3(x)dx + o(h_i) - 4\Big(\int\int xK(u)f_i(x)\Big(f_i(x)+uh_if_i^2(x)+o(h_i)\Big)dxdu\Big)^2\Big] \\
&= 4c_1^2\Big[\int x^2 f_i^3(x)dx - \Big(\int xf_i^2(x)dx\Big)^2\Big] + o(h_i) \qquad (2.22)
\end{aligned}
$$

By central limit theorem of U-statistics (See Koroljuk and Borovskich [22], pp. 128-129), we have $\sqrt{n_i}(\hat{\mu}_i - \mu_i) \xrightarrow{d} N(0, \varpi_i^2)$. Thus, we obtain that

$$\sqrt{n_i}(c_1\hat{V}_i + c_2 - (c_1V_i + c_2)) = \sqrt{n_i}c_1(\hat{V}_i - V_i) \xrightarrow{d} N(0, \varpi_i^2). \qquad (2.23)$$

Let $\omega_i^2 = \frac{\varpi_i^2}{c_1^2}$. Then we have $\sqrt{n_i}(\hat{V}_i - V_i) \xrightarrow{d} N(0, \omega_i^2)$, where

$$\omega_i^2 = 4\Big\{\int x^2 f_i^3(x)dx - \Big(\int xf_i^2(x)dx\Big)^2\Big\}. \qquad (2.24)$$

■

Define the sum of squares between as

$$SSB = \sum_{i=1}^{K} \frac{n_i(\hat{V}_i - \hat{V}.)^2}{\hat{\omega}_i{}^2}, \qquad (2.25)$$

where

$$\hat{V}. = \frac{\sum_{i=1}^{K} n_i\hat{V}_i/\hat{\omega}_i{}^2}{\sum_{i=1}^{K} n_i/\hat{\omega}_i{}^2}, \qquad (2.26)$$

and $\hat{\omega_i}^2$ is a consistent estimate of ω_i^2. To obtain the asymptotic distribution of SSB, another two auxiliary variables are defined as follows:

$$S_2^0 = \sum_{i=1}^{K} \frac{N\lambda_i(\hat{V}_i - \bar{V})^2}{\omega_i^2} - N\left[\sum_{i=1}^{I} \frac{\lambda_i}{\omega_i^2}\right](V^* - \bar{V})^2, \tag{2.27}$$

and

$$S_2^{00} = \sum_{i=1}^{K} \frac{n_i(\hat{V}_i - \bar{V})^2}{\hat{\omega_i}^2} - \sum_{i=1}^{K} \frac{n_i(\hat{V}. - \bar{V})^2}{\hat{\omega_i}^2}, \tag{2.28}$$

where

$$\begin{aligned} V^* &= \frac{\sum_{i=1}^{K} \lambda_i \hat{V}_i/\omega_i^2}{\sum_{i=1}^{K} \lambda_i/\omega_i^2}, \\ \bar{V} &= \frac{\sum_{i=1}^{K} \lambda_i V_i/\omega_i^2}{\sum_{i=1}^{K} \lambda_i/\omega_i^2}. \end{aligned}$$

Lemma 2.5 *Let* $N = \sum_{i=1}^{K} n_i$. *If* $\lambda_i = \lim_{\min n_i \to \infty} \frac{n_i}{N}$ *and* $\hat{\omega}_i^2 \xrightarrow{p} \omega_i^2$, *then*

(i) $S_2^0 - S_2^{00} \xrightarrow{p} 0$,

(ii) $S_2^{00} - SSB \xrightarrow{p} 0$,

as $\min_i n_i \to \infty$.

Proof: They can be proved directly by applying Slutsky Theorem [10].

■

Theorem 2.6 *Under the null hypothesis, if for any* $i = 1, 2, \ldots, K$, $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ *as* $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} x^2 f_i^3(x)dx < \infty$ *and if* $f_i(\cdot)$ *is twice differentiable, then SSB is asymptotically* $\chi^2(K-1)$. *In general (under the alternative), SSB is asymptotically non-central* $\chi^2(K-1)$ *with non-centrality parameter:*

$$\psi_2 = \frac{1}{2}\mu^{(\mathbf{2})'}\mathbf{B_2}\mu^{(\mathbf{2})}, \tag{2.29}$$

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

$$\mu_i^{(2)} = \frac{\sqrt{\lambda_i}(e_i - \frac{\sum_{i=1}^K \lambda_i e_i/\omega_i^2}{\sum_{i=1}^K \lambda_i/\omega_i^2}) \int f^2(y)dy}{\sigma\omega_i}, \tag{2.30}$$

$$\mathbf{B_2} = \begin{bmatrix} 1 - \frac{\lambda_1/\omega_1^2}{\sum_{i=1}^K \lambda_i/\omega_i^2} & -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_2}/\omega_2)}{\sum_{i=1}^K \lambda_i/\omega_i^2} & \cdots & -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^K \lambda_i/\omega_i^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^K \lambda_i/\omega_i^2} & -\frac{(\sqrt{\lambda_2}/\omega_2)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^K \lambda_i/\omega_i^2} & \cdots & 1 - \frac{\lambda_K/\omega_K^2}{\sum_{i=1}^K \lambda_i/\omega_i^2} \end{bmatrix}, \tag{2.31}$$

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

Proof: Let $N = \sum_{i=1}^K n_i$, and then $\lambda_i \simeq \frac{n_i}{N}$. Set $T_i^{(2)} = \sqrt{\lambda_i N}(\hat{V}_i - \bar{V})/\omega_i$, and then under H_0, $T_i^{(2)} \overset{a}{\sim} N(0,1)$ as $N \to \infty$ by Lemma (2.4). Note that

$$\begin{aligned} S_2^0 &= \sum_{i=1}^K \frac{N\lambda_i(\hat{V}_i - \bar{V})^2}{\omega_i^2} - N\left[\sum_{i=1}^K \frac{\lambda_i}{\omega_i^2}\right](V^* - \bar{V})^2 \\ &= \sum_{i=1}^K \frac{N\lambda_i(\hat{V}_i - \bar{V})^2}{\omega_i^2} - N\left[\sum_{i=1}^K \frac{\lambda_i}{\omega_i^2}\right]\left(\frac{\sum_i \lambda_i \hat{V}_i/\omega_i^2}{\sum_i \lambda_i/\omega_i^2} - \frac{\sum_i \lambda_i V_i/\omega_i^2}{\sum_i \lambda_i/\omega_i^2}\right)^2 \\ &= \sum_{i=1}^K \frac{N\lambda_i(\hat{V}_i - \bar{V})^2}{\omega_i^2} - \frac{N}{\sum_{i=1}^K \frac{\lambda_i}{\omega_i^2}}\left(\sum_{i=1}^K \frac{\lambda_i(\hat{V}_i - V_i)}{\omega_i^2}\right)^2 \\ &= \sum_{i=1}^K \frac{N\lambda_i(\hat{V}_i - \bar{V})^2}{\omega_i^2} - \sum_{i=1}^K \sum_{j=1}^K \frac{\sqrt{N\lambda_i}(\hat{V}_i - \bar{V})}{\omega_i} \frac{\sqrt{N\lambda_j}(\hat{V}_j - \bar{V})}{\omega_j} \frac{(\sqrt{\lambda_i}/\omega_i)(\sqrt{\lambda_j}/\omega_j)}{\sum_{i=1}^K \lambda_i/\omega_i^2}, \end{aligned} \tag{2.32}$$

which can be written as a quadratic form, i.e. $S_2^0 = \mathbf{U_2}'\mathbf{B_2}\mathbf{U_2}$, where,

$$\mathbf{U_2} = (T_1^{(2)}, T_2^{(2)}, \cdots, T_K^{(2)})', \tag{2.33}$$

and

$$\mathbf{B_2} = \begin{bmatrix} 1 - \frac{\lambda_1/\omega_1^2}{\sum_{i=1}^K \lambda_i/\omega_i^2} & -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_2}/\omega_2)}{\sum_{i=1}^K \lambda_i/\omega_i^2} & \cdots & -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^K \lambda_i/\omega_i^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(\sqrt{\lambda_1}/\omega_1)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^K \lambda_i/\omega_i^2} & -\frac{(\sqrt{\lambda_2}/\omega_2)(\sqrt{\lambda_K}/\omega_K)}{\sum_{i=1}^K \lambda_i/\omega_i^2} & \cdots & 1 - \frac{\lambda_K/\omega_K^2}{\sum_{i=1}^K \lambda_i/\omega_i^2} \end{bmatrix}. \tag{2.34}$$

It can be easily shown that $\mathbf{B_2}$ is symmetric and idempotent. Thus, we obtain that

$$\begin{aligned} rank(\mathbf{B_2}) &= tr(\mathbf{B_2}) \\ &= tr(\mathbf{I} - \mathbf{a_2}\mathbf{a_2}') \\ &= tr(\mathbf{I}) - tr(\mathbf{a_2}\mathbf{a_2}') \\ &= K - tr(\mathbf{a_2}'\mathbf{a_2}) \\ &= K - 1, \end{aligned} \tag{2.35}$$

where

$$\mathbf{a_2} = \begin{pmatrix} \sqrt{\lambda_1}/\omega_1 \\ \sqrt{\lambda_2}/\omega_2 \\ \vdots \\ \sqrt{\lambda_K}/\omega_K \end{pmatrix}. \tag{2.36}$$

$\mathbf{U_2}$ follows approximately multivariate normal with mean $\mathbf{0}$ and variance $\mathbf{I}$, since $T_i^{(2)}$'s independently follow uni-variate standard normal distribution. Therefore, S_2^0 is asymptotically χ^2 with degrees of freedom $K-1$ under H_0.

Under the alternative, since e_i is chosen such that $\mu_i = 1 + \frac{e_i}{\sqrt{N}}$, then we have

$$\begin{aligned} V_i &= \frac{\mu_i}{c_1} - \frac{c_2}{c_1} \\ &= \frac{1-c_2}{c_1} + \frac{e_i}{c_1\sqrt{N}}. \end{aligned}$$

Thus, $\bar{V}$ can be written as:

$$\begin{aligned} \bar{V} &= \frac{\sum_{i=1}^K \lambda_i V_i/\omega_i^2}{\sum_{i=1}^K \lambda_i/\omega_i^2} \\ &= \sum_{i=1}^K \frac{\lambda_i}{\omega_i^2}(\frac{1-c_2}{c_1} + \frac{e_i}{c_1\sqrt{N}})/\sum_{i=1}^K \frac{\lambda_i}{\omega_i^2} \\ &= \frac{1-c_2}{c_1} + \frac{1}{c_1\sqrt{N}}\frac{\sum_{i=1}^K \lambda_i e_i/\omega_i^2}{\sum_{i=1}^K \lambda_i/\omega_i^2}. \end{aligned}$$

Under H_1, $T_i^{(2)} = \sqrt{N\lambda_i}(\hat{V}_i - \bar{V})/\omega_i = \frac{\sqrt{N\lambda_i}(\hat{V}_i - V_i)}{\omega_i} + \frac{\sqrt{N\lambda_i}(V_i - \bar{V})}{\omega_i}$. By Lemma (2.4), the first term $\frac{\sqrt{N\lambda_i}(\hat{V}_i - V_i)}{\omega_i}$ is approximately standard normally distribution. And the second term's limit is

$$
\begin{aligned}
\mu &= \lim_{N\to\infty} \frac{\sqrt{N\lambda_i}(V_i - \bar{V})}{\omega_i} \\
&= \lim_{N\to\infty} \frac{\sqrt{N\lambda_i}\frac{1}{c_1\sqrt{N}}(e_i - \frac{\sum_{i=1}^{K}\lambda_i e_i/\omega_i^2}{\sum_{i=1}^{K}\lambda_i/\omega_i^2})}{\omega_i} \\
&= \frac{\sqrt{\lambda_i}(e_i - \frac{\sum_{i=1}^{K}\lambda_i e_i/\omega_i^2}{\sum_{i=1}^{K}\lambda_i/\omega_i^2})}{c_1\omega_i} \\
&= \frac{\sqrt{\lambda_i}(e_i - \frac{\sum_{i=1}^{K}\lambda_i e_i/\omega_i^2}{\sum_{i=1}^{K}\lambda_i/\omega_i^2})\int f^2(y)dy}{\sigma\omega_i}, \qquad (2.37)
\end{aligned}
$$

where ω_i is given by the square root of equation(2.24). Therefore, in general $T_i^{(2)}$ is approximately normally distribution with mean $\mu^{(2)}$ given in equation (2.37) and variance 1. This implies that $S_2^0 = \mathbf{U^{(2)}}'\mathbf{B_2}\mathbf{U^{(2)}}$ is asymptotically non-central $\chi^2(K-1)$ with non-centrality parameter $\psi_2 = \frac{1}{2}\mu^{\mathbf{(2)}'}\mathbf{B_2}\mu^{\mathbf{(2)}}$.

By Lemma (2.5), SSB converges in probability to S_2^0. Therefore, SSB follows asymptotically χ^2 with degrees of freedom $K-1$ under null hypothesis and asymptotically non-central $\chi^2(K-1)$ with non-centrality parameter $\psi_2 = \frac{1}{2}\mu^{\mathbf{(2)}'}\mathbf{B_2}\mu^{\mathbf{(2)}}$. ■

Let $A_{ij_1j_2} = \frac{1}{h_i}\left(\frac{X_{ij_1}+X_{ij_2}}{2}\right)K\left(\frac{X_{ij_1}-X_{ij_2}}{h_i}\right)$. Then, $\hat{V}_i$ can be rewritten as:

$$
\hat{V}_i = \frac{1}{n_i(n_i-1)}\sum\sum_{j_1\neq j_2} A_{ij_1j_2}. \qquad (2.38)
$$

Define

$$
S_3^0 = \sum_{i=1}^{K}\sum\sum_{j_1\neq j_2}^{n_i}(A_{ij_1j_2} - \hat{V}_i)^2/{\omega_i}^2, \qquad (2.39)
$$

and the sum of squares within as

$$
SSW = \frac{\sum_{i=1}^{K}\sum\sum_{j_1\neq j_2}^{n_i}(A_{ij_1j_2} - \hat{V}_i)^2/{\hat{\omega}_i}^2}{C_w}, \qquad (2.40)
$$

where

$$
C_w = \begin{cases} \frac{n}{2}-1 & \text{if } n_i = n \text{ for all i,} \\ c_0 & \text{otherwise,} \end{cases} \qquad (2.41)
$$

and $c_0 = \sum_{i=1}^{d} \pi_i/d$, $\pi_1, \pi_2, \cdots, \pi_d$ are the eigenvalues of $\mathbf{B}_3$ with

$$\mathbf{B}_3 = \begin{bmatrix} (\frac{n_1}{2}-1)(\mathbf{I}_{n_1} - \frac{1}{n_1}\mathbf{J}_{n_1}) & 0 & \cdots & 0 \\ 0 & (\frac{n_2}{2}-1)(\mathbf{I}_{n_2} - \frac{1}{n_2}\mathbf{J}_{n_2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\frac{n_K}{2}-1)(\mathbf{I}_{n_K} - \frac{1}{n_K}\mathbf{J}_{n_K}) \end{bmatrix}. \tag{2.42}$$

Lemma 2.7 *Let* $N = \sum_{i=1}^{K} n_i$*. If* $\lambda_i = \lim_{n_i \to \infty} \frac{n_i}{N}$ *and* $\hat{\omega}_i{}^2 \xrightarrow{p} \omega_i^2$*, then* $S_3^0 - C_w SSW \xrightarrow{p} 0$ *as* $\min_i n_i \to \infty$*.*

Proof: This can be proved directly by applying Slutsky Theorem [10].

■

Theorem 2.8 *For any* $i = 1, 2, \ldots, K$*, if* $n_i h_i^4 \to 0$*,* $n_i h_i \to \infty$ *as* $\min_i n_i \to \infty$*,* $\int_{-\infty}^{\infty} x^2 f_i^3(x)dx < \infty$ *and if* $f_i(\cdot)$ *is twice differentiable, then then SSW is asymptotically* χ^2 *with degrees of freedom* df_w*, where*

$$df_w = \begin{cases} K(n-1) & \text{if } n_i = n \text{ for all } i, \\ d & \text{otherwise,} \end{cases} \tag{2.43}$$

where d is the number of eigenvalues of $\mathbf{B}_3$ *given by equation (2.42).*

Proof: By the Hajek projection [16], $A_{ij_1j_2}$ can be decomposed into the sum of conditional expected values and a residual as follows:

$$A_{ij_1j_2} = E(A_{ij_1j_2}|X_{ij_1}) + E(A_{ij_1j_2}|X_{ij_2}) + O_p(n_i). \tag{2.44}$$

Set $\varphi(X_{ij_1}) = E(A_{ij_1j_2}|X_{ij_1})$ and $\varphi(X_{ij_2}) = E(A_{ij_1j_2}|X_{ij_2})$, thus

$$
\begin{aligned}
\hat{V}_i &= \frac{1}{n_i(n_i-1)} \sum\sum_{j_1 \neq j_2} A_{ij_1j_2} \\
&\approx \frac{1}{n_i(n_i-1)} \sum\sum_{j_1 \neq j_2} (\varphi(X_{ij_1}) + \varphi(X_{ij_2})) \qquad (2.45) \\
&= \frac{1}{n_i(n_i-1)} \Big(\sum_{j_1}\sum_{j_2} (\varphi(X_{ij_1}) + \varphi(X_{ij_2})) - \sum\sum_{j_1=j_2} (\varphi(X_{ij_1}) + \varphi(X_{ij_2})) \Big). \\
&= \frac{1}{n_i(n_i-1)} \Big(2n_i \sum_{j_1} \varphi(X_{ij_1}) - 2\sum_{j_1} \varphi(X_{ij_1}) \Big). \\
&= \frac{1}{n_i} \sum_{j_1}^{n_i} 2\varphi(X_{ij_1}) \qquad (2.46)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\sum\sum_{j_1 \neq j_2} \frac{(A_{ij_1j_2} - \hat{V}_i)^2}{\omega_i^2} &\approx \sum\sum_{j_1 \neq j_2} \frac{(\varphi(X_{ij_1}) + \varphi(X_{ij_2}) - \hat{V}_i)^2}{\omega_i^2} \\
&= \sum_{j_1}\sum_{j_2} \frac{(\varphi(X_{ij_1}) - \frac{1}{n_i}\sum\varphi(X_{ij_1}) + \varphi(X_{ij_2}) - \frac{1}{n_i}\sum\varphi(X_{ij_2}))^2}{\omega_i^2} \\
&- \sum_{j_1} \frac{\big(2\varphi(X_{ij_1}) - \frac{1}{n_i}\sum_{j_1=1}^{n_i} 2\varphi(X_{ij_1})\big)^2}{\omega_i^2} \\
&= 2n_i \sum_{j_1} \frac{(\varphi(X_{ij_1}) - \frac{1}{n_i}\sum\varphi(X_{ij_1}))^2}{\omega_i^2} - \sum_{j_1} \frac{\big(2\varphi(X_{ij_1}) - \hat{V}_i\big)^2}{\omega_i^2} \\
&= \frac{n_i}{2} \sum_{j_1} \frac{(2\varphi(X_{ij_1}) - \hat{V}_i)^2}{\omega_i^2} - \sum_{j_1} \frac{\big(2\varphi(X_{ij_1}) - \hat{V}_i\big)^2}{\omega_i^2} \\
&= (\frac{n_i}{2} - 1) \sum_{j_1} \frac{(2\varphi(X_{ij_1}) - \hat{V}_i)^2}{\omega_i^2} \\
&= (\frac{n_i}{2} - 1) \Big[\sum_{j=1}^{n_i} \frac{(2\varphi(X_{ij_1}) - V_i)^2}{\omega_i^2} - \frac{n_i(\hat{V}_i - V_i)^2}{\omega_i^2} \Big], \qquad (2.47)
\end{aligned}
$$

since $\sum_{j_1=1}^{n_i} (2\varphi(X_{ij_1}) - V_i)^2 = \sum_{j_1=1}^{n_i} (2\varphi(X_{ij_1}) - \hat{V}_i)^2 + n_i(\hat{V}_i - V_i)^2$.

Let $H_{ij} = \frac{2\varphi(X_{ij}) - V_i}{\omega_i}$ for $j = 1, 2, \cdots, n_i$ and $\mathbf{H}_i = (H_{i1}, H_{i2}, \cdots, H_{in_i})'$ for $i =$

$1, 2, \cdots, K$. Thus, equation (2.47) can be rewritten in matrix form as

$$
\begin{aligned}
& (\frac{n_i}{2} - 1)\Big[\sum_{j=1}^{n_i} \frac{(2\varphi(X_{ij_1}) - V_i)^2}{\omega_i^2} - \frac{n_i(\hat{V}_i - V_i)^2}{\omega_i^2}\Big] \\
= & (\frac{n_i}{2} - 1)\Big[\mathbf{H}_i'\mathbf{H}_i - \mathbf{H}_i'\frac{1}{n_i}\mathbf{J}_{n_i}\mathbf{H}_i\Big] \\
= & \mathbf{H}_i'(\frac{n_i}{2} - 1)(\mathbf{I} - \frac{1}{n_i}\mathbf{J}_{n_i})\mathbf{H}_i. \qquad (2.48)
\end{aligned}
$$

Let $\mathbf{H} = (\mathbf{H}_1', \mathbf{H}_2', \cdots, \mathbf{H}_K')'$. Therefore, S_3^0 can be written in the matrix form as

$$
\begin{aligned}
S_3^0 &= \sum_{i=1}^{K}\sum\sum_{j_1 \neq j_2}^{n_i} (A_{ij_1j_2} - \hat{V}_i)^2/{\omega_i}^2 \\
&= \sum_{i=1}^{K} \mathbf{H}_i'(\frac{n_i}{2} - 1)(\mathbf{I} - \frac{1}{n_i}\mathbf{J}_{n_i})\mathbf{H}_i \\
&= \mathbf{H}'\begin{bmatrix} (\frac{n_1}{2} - 1)(\mathbf{I}_{n_1} - \frac{1}{n_1}\mathbf{J}_{n_1}) & 0 & \cdots & 0 \\ 0 & (\frac{n_2}{2} - 1)(\mathbf{I}_{n_2} - \frac{1}{n_2}\mathbf{J}_{n_2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\frac{n_K}{2} - 1)(\mathbf{I}_{n_K} - \frac{1}{n_K}\mathbf{J}_{n_K}) \end{bmatrix}\mathbf{H} \\
&= \mathbf{H}'\mathbf{B}_3\mathbf{H}. \qquad (2.49)
\end{aligned}
$$

Now, we need to show that $\mathbf{H}$ asymptotically follows multivariate normal distribution.

Note that $E(2\varphi(X_{ij_1})) = E(\frac{1}{n_i}\sum_{j_1=1}^{n_i} 2\varphi(X_{ij_1})) = E(\hat{V}_i) \approx V_i$ since $\hat{\mu}_i$ is asymptotically unbiased by equation (2.21). Also $Var(2\varphi(X_{ij_1})) = \frac{1}{n_i}\sum_{j_1=1}^{n_i} Var(2\varphi(X_{ij_1})) = n_i Var(\frac{1}{n_i}\sum_{j_1=1}^{n_i} 2\varphi(X_{ij_1})) = n_i Var(\hat{V}_i) = \omega_i^2$. By the central limit theorem of U-statistics, $H_{ij} = \frac{2\varphi(X_{ij}) - V_i}{\omega_i}$ is distributed asymptotically normal with mean 0 and variance 1. Since the H_{ij}'s are independent, then $\mathbf{H}$ follows asymptotically multivariate normal distribution with mean $\mathbf{0}$ and variance $\mathbf{I}$.

(i) If $n_i = n$ for all i, then it is easy to verify that $\mathbf{B}_3/(\frac{n}{2} - 1)$ is a symmetric and idempotent matrix with rank $\sum_{i=1}^{I} n_i - K = N - K = K(n-1)$. Therefore, $S_3^0/(\frac{n}{2} - 1)$ is asymptotically χ^2 with degrees of freedom $N - K = K(n-1)$. By

Lemma (2.7), the sum of square within SSW is asymptotically χ^2 with degrees of freedom $K(n-1)$.

(ii) If $n_i \neq n_j$ for some $i \neq j$, $\mathbf{B}_3$ is symmetric, although not idempotent. Thus, there exists $\mathbf{H}'\mathbf{B}_3\mathbf{H} = \sum_{i=1}^{d} \pi_i z_i^2$, where $\pi_1, \pi_2, \cdots, \pi_d$ are the eigenvalues of $\mathbf{B}_3$, $z_i \sim N(0,1)$ and are independent. Let $c_0 = \sum_{i=1}^{d} \pi_i/d$, then by [42], $S_3^0/c_0 = \mathbf{H}'\mathbf{B}_3\mathbf{H}/c_0 \dot{\sim} \chi_d^2$. By Lemma (2.7), the sum of squares within (SSW) is asymptotically χ^2 with degrees of freedom d.

■

Define the F-test statistics of kernel based nonparametric test for location parameters as:

$$\begin{aligned} F_l &= \frac{MSB}{MSW} = \frac{SSB/(K-1)}{SSW/df_w} \\ &= \frac{\sum_{i=1}^{K} \frac{n_i(\hat{V}_i - \hat{V}.)^2}{\hat{\omega}_i{}^2}/(K-1)}{\left(\sum_{i=1}^{K} \sum\sum_{j_1 \neq j_2}^{n_i} (A_{ij_1j_2} - \hat{V}_i)^2/\hat{\omega}_i{}^2\right)/df_w}, \end{aligned} \tag{2.50}$$

where df_w is given in equation (2.43).

Theorem 2.9 *If for any* $i = 1, 2, \ldots, K$, $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ *as* $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} x^2 f_i^3(x)dx < \infty$ *and if* $f_i(\cdot)$ *is twice differentiable, then under null hypothesis,* F_l *in equation (2.50) follows asymptotically F distribution with degrees of freedom* $K-1$ *and* df_w*. Under the alternative,* F_l *follows asymptotically non-central* $F(K-1, df_w)$ *with non-centrality parameter* ψ_2 *described in equation (2.29).*

Proof: Theorem (2.6) shows that SSB follows asymptotically χ^2 with degrees of freedom $K-1$ under null hypothesis and asymptotically non-central $\chi^2(K-1)$ under the alternative. Furthermore, Theorem (2.8) implies that the sum of squares within SSW is asymptotically χ^2 with degrees of freedom $K(n-1)$ for balanced data and χ^2 with degrees of freedom d for unbalanced data, where d is the number of eigenvalues of $\mathbf{B}_3$ in equation (2.49). In order to show $F_l = \frac{MSB}{MSW}$ follows asymptotically F distribution under null hypothesis and non-central F distribution under alternative, we just need to show SSB and SSW are asymptotically independent as $\min_i n_i \to \infty$. In theorem (2.6), S_2^0, which converges in probability to SSB, is written as a quadratic form $S_2^0 = \mathbf{U_2}'\mathbf{B_2}\mathbf{U_2}$. Note that under the null hypothesis

$$T_i^{(2)} \simeq \sqrt{n_i}(\hat{V}_i - \bar{V})/\omega_i = \frac{1}{\sqrt{n_i}}\mathbf{H}_i'\mathbf{j}_{n_i}. \tag{2.51}$$

Hence, S_2^0 can also be written as

$$\begin{aligned}
S_2^0 &= (\frac{1}{\sqrt{n_1}}\mathbf{H}_1'\mathbf{j}_{n_1}, \cdots, \frac{1}{\sqrt{n_K}}\mathbf{H}_K'\mathbf{j}_{n_K})\mathbf{B_2}\begin{pmatrix} \frac{1}{\sqrt{n_1}}\mathbf{H}_1'\mathbf{j}_{n_1} \\ \vdots \\ \frac{1}{\sqrt{n_K}}\mathbf{H}_K'\mathbf{j}_{n_K} \end{pmatrix} \\
&= \mathbf{H}'\begin{bmatrix} \frac{1}{\sqrt{n_1}}\mathbf{j}_{n_1}' & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_2}}\mathbf{j}_{n_2}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_K}}\mathbf{j}_{n_K}' \end{bmatrix}' \mathbf{B}_2 \begin{bmatrix} \frac{1}{\sqrt{n_1}}\mathbf{j}_{n_1}' & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_2}}\mathbf{j}_{n_2}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_K}}\mathbf{j}_{n_K}' \end{bmatrix}\mathbf{H} \\
&\stackrel{def}{=} \mathbf{H}'\mathbf{B}_4\mathbf{H}.
\end{aligned} \tag{2.52}$$

Recall from Theorem (2.8) that $S_3^0 = \mathbf{H}'\mathbf{B}_3\mathbf{H}$. And it is easy to check that

$$\mathbf{B}_3\mathbf{B}_4 = \mathbf{0} * \mathbf{B}_2 \begin{bmatrix} \frac{1}{\sqrt{n_1}}\mathbf{j}_{n_1}' & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_2}}\mathbf{j}_{n_2}' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_K}}\mathbf{j}_{n_K}' \end{bmatrix} = \mathbf{0}. \tag{2.53}$$

Thus, S_2^0 and S_3^0 are independent. By Lemma (2.5) and Lemma (2.7), SSB and SSW are asymptotically independent under null hypothesis H_0: $V_1 = V_2 = \ldots = V_K$. Hence, under null hypothesis, $F_l = \frac{MSB}{MSW}$ in equation (2.50) follows asymptotically F distribution with degrees of freedom $K-1$ and $K(n-1)$ for balanced data and $F(K-1, d)$ for unbalanced data. Under the alternative, $F_l = \frac{MSB}{MSW}$ follows asymptotically non-central $F(K-1, K(n-1))$ for balanced data, and non-central $F(K-1, d)$ for unbalanced data, with non-centrality parameter ψ_2 described in equation (2.29).

■

2.3 Simulation Study for Evaluating the Power of Kernel-based Nonparametric One-way ANOVA

In this section, powers of the kernel-based nonparametric scale tests and location tests, i.e. ANOVA tests, are evaluated via simulation. To better demonstrate the properties of kernel-based nonparametric ANOVA tests compared with the traditional parametric ANOVA tests, the performances of both tests on data from various distributions are studied.

2.3.1 Simulation Study for Scale Tests

As stated in Section 1.1.1, Levene's test for testing homogeneity of variance is not sensitive to the departure of normality, as Bartlett's is. So we compared the performance of Levene's test with our new proposed nonparametric scale test in 3 distinctive cases listed in Table 2.1.

Table 2.1: Three Cases

	Distribution
Case I	Normal Distribution
Case II	Cauchy Distribution
Case III	Three-parameter Lognormal Distribution

In case study I, groups of data coming from Normal distributions with different variances are tested with Levene's test (called Parametric test later) and kernel-based nonparametric scale test (called Nonparametric test later). In case study II, groups of data coming from Cauchy distributions with different scale parameters are tested as in Case I. It is known that Cauchy distribution is a heavy-tailed distribution compared to normal distribution. The histogram of Cauchy distribution with location parameter 10 and scale parameter 10 is given in Figure 2.1. As shown in Figure 2.1, the data from Cauchy (10,10) ranges from around -15000 to 20000, which are far more spread than normal distribution. In case study III, groups of data come from Lognormal distributions with different scale parameters are tested as in Case I. Lognormal distributions are very important in finance and economics. For example, it is often used to characterize stock prices in financial modeling, including pricing and hedging. A lognormal distribution is a right skewed distribution. Figure 2.2 gives the histogram of three-parameter lognormal distributions. Figure 2.2(a) shows a lognormal distribution with location parameter 16, scale parameter 4 and shape parameter 3, while Figure 2.2(b) shows a lognormal distribution with location parameter 6, scale parameter 2 and shape parameter 1. Comparing Figure 2.2(a) with 2.2(b), it is obvious that the larger the shape parameter is, the stronger the skewness is. Our goal of investigating these three cases is to study the situations that the kernel based nonparametric scale test outperforms the Levene's test. Since in Section 2.1 we showed that kernel based nonparametric scale tests are asymptotic F test, another goal of simulation study in this section is to find the sample size we need to reach an appropriate power.

To evaluate the actual Type I error rate and the power of the Levene's and nonparametric scale tests proposed in Section 2.1 in the 3 cases in Table 2.1, we test the homogeneity of scale parameters of three groups, i.e. $K = 3$ for each case. To obtain the actual Type I error while setting the significant level $\alpha = 0.05$, we follow the steps below:

(1) Randomly generate 3 groups of data with balanced sample size n from N(3,1) for Case I, Cauchy(10, 10) for Case II and Lognormal(16, 4, 3) for Case III.

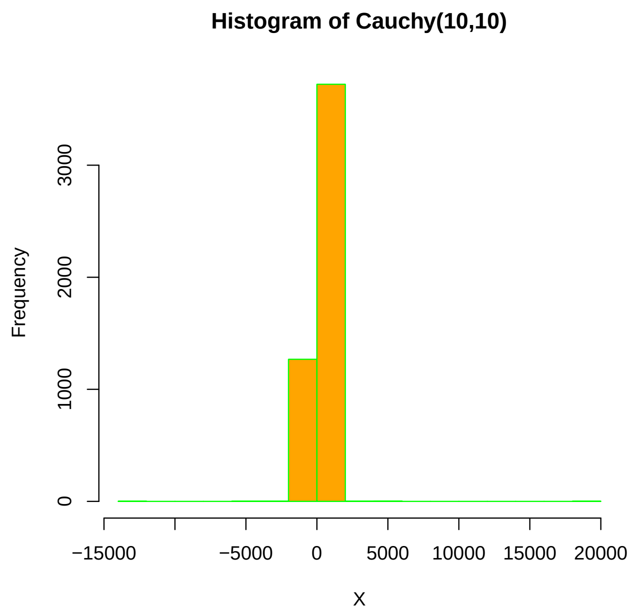
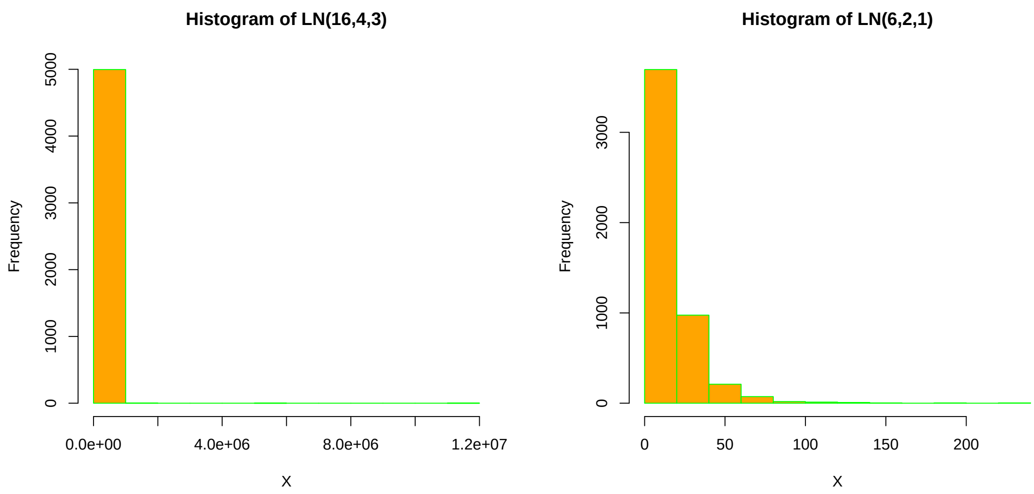


Figure 2.1: Histogram of Cauchy(location=10, scale=10)



(a) Lognormal(16,4,3)

(b) Lognormal(6,2,1)

Figure 2.2: Histogram of Lognormal Distribution

(2) Apply the Levene's test and our scale test separately. Record the test result as 1 or 0. 1 means "reject the null hypothesis" and 0 means fail to reject.

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

(4) Repeat (1)-(3) for sample size $n = 15, 20, 25, 30, 35, 40, 45, 50, 55, 60$.

The procedure to calculate the empirical power is very similar except generating 3 groups of data from 3 different distributions for each case in Step (1). The distribution types and parameters assigned to each group in each case are listed in Table 2.2.

Table 2.2: Evaluate the Power of Scale Tests in 3 Cases

Case	**Group1**	**Group2**	**Group3**
I^a	N(3, 0.25)	N(3, 1)	N(3, 1.75)
II^a	Cauchy(10, 2)	Cauchy(10, 10)	Cauchy(10, 20)
III^a	Lognormal(16, 2, 3)	Lognormal(16, 4, 3)	Lognormal(16, 6, 3)

As shown in Table 2.2, in Case I^a, all the three groups are from normal with the same mean 3, but standard deviation 0.25, 1 and 1.75. The side-by-side boxplot of the three groups in Case I^a is given in Figure 2.3. In Figure 2.3, the plot with the box painted in red is the boxplot for group 1, which is generated from $N(3, 0.25)$. The plot painted in green is the boxplot for group 2, which is generated from $N(3, 1)$. And the plot painted in blue is the boxplot for group 3, which is generated from $N(3, 1.75)$. It is not hard to tell that the three groups have the same central tendency, but different variabilities.

The side-by-side boxplot of the three groups in Case II^a is given in Figure 2.4. In Figure 2.4, the plot with the box painted in red is the boxplot for group 1, which is generated from $Cauchy(10, 2)$. The plot painted in green is the boxplot for group 2, which is generated from $Cauchy(10, 10)$. And the plot painted in blue is the boxplot

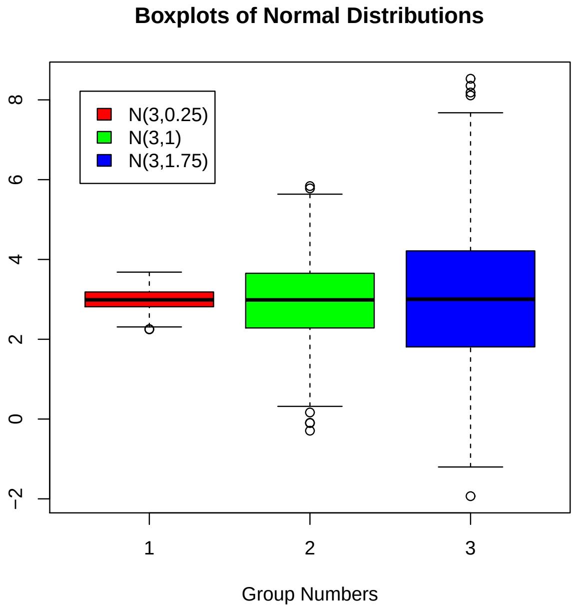
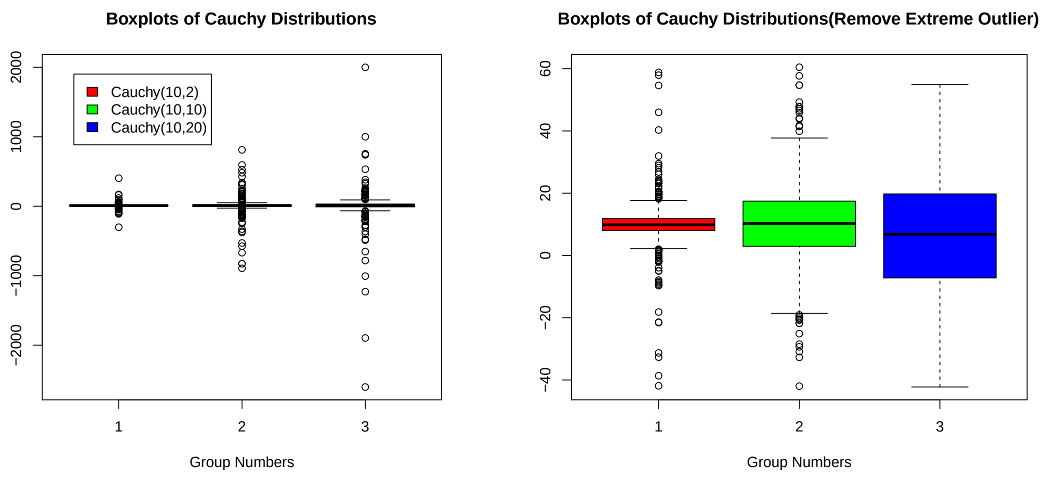


Figure 2.3: Side-by-Side Boxplot for the 3 Groups in Case I^a: Normal Distributions

for group 3, which is generated from $Cauchy(10, 20)$. The color of the boxes in Figure 2.4(a) is invisible, since Cauchy distribution has very fat tails. Figure 2.4(b) shows the boxplot of the three groups after removing the extreme outliers, i.e. the points that are smaller than the first quartile subtract 3 times inter-quartile range or bigger than the third quartile plus 3 times inter-quartile range. It is clear to see in Figure 2.4(b) that the three groups have similar central tendency, but distinct spreadness.

The side-by-side boxplot of the three groups in Case III^a is given in Figure 2.5. In Figure 2.5, the plot with the box painted in red is the boxplot for group 1, which is generated from lognormal distribution with location parameter 16, scale parameter 2 and shape parameter 3, i.e. $Lognormal(16, 2, 3)$. The plot painted in green is the boxplot for group 2, which is generated from $Lognormal(16, 4, 3)$. And the plot painted in blue is the boxplot for group 3, which is generated from $Lognormal(16, 6, 3)$. Similar to Figure 2.4(a), the color of the boxes in Figure 2.5(a) is invisible, since lognormal distribution is strongly right skewed. Figure 2.5(b) shows the boxplot of the three



(a) Boxplot for All Data (b) Boxplot Without Extreme Outliers

Figure 2.4: Side-by-Side Boxplot for the 3 Groups in Case II^a: Cauchy Distributions

groups after removing the extreme outliers as what we did in Figure 2.4(b). It is not easy to tell in Figure 2.5(b) whether the three groups have similar central tendency since the distributions are strongly right skewed.

By following the similar steps as calculating the actual Type I error rates, empirical powers can be evaluated for each case and a series of sample sizes. The simulation results are listed in Table 2.3 - Table 2.5.

Table 2.3 shows that the actual Type I errors for both tests, either parametric and nonparametric, are around 0.05, the significant level. As we expected, in Case I^a, Levene's test performs a little bit better when the sample size is less than 35. When the samples come from normal distribution, parametric tests are always the best choice for small sample size. However, as shown in the Table 2.3, our test is as good as the parametric test even in the normal case. The power of our nonparametric test is around 90% when the sample size is only 20.

Table 2.4 shows that Levene's test suffers from inflated Type I error rates when the data come from fat-tailed distributions. Whereas, our kernel based nonparametric

Table 2.3: Power for the Scale Test: Case I^a (Normal Distribution)

Sample Size	Parametric Test		Nonparametric Test	
	Power	**Type I**	**Power**	**Type I**
15	0.8002	0.0599	0.7890	0.0620
20	0.9221	0.0586	0.8982	0.0587
25	0.9737	0.0568	0.9601	0.0508
30	0.9913	0.0553	0.9837	0.0610
35	0.9980	0.0516	0.9939	0.0567
40	0.9991	0.0554	0.9979	0.0629
45	0.9998	0.0546	0.9988	0.0546
50	1.0000	0.0532	0.9997	0.0597
55	1.0000	0.0533	0.9999	0.0613
60	1.0000	0.0502	1.0000	0.0619

[1] * Note: Given significance level: $\alpha = 0.05$.

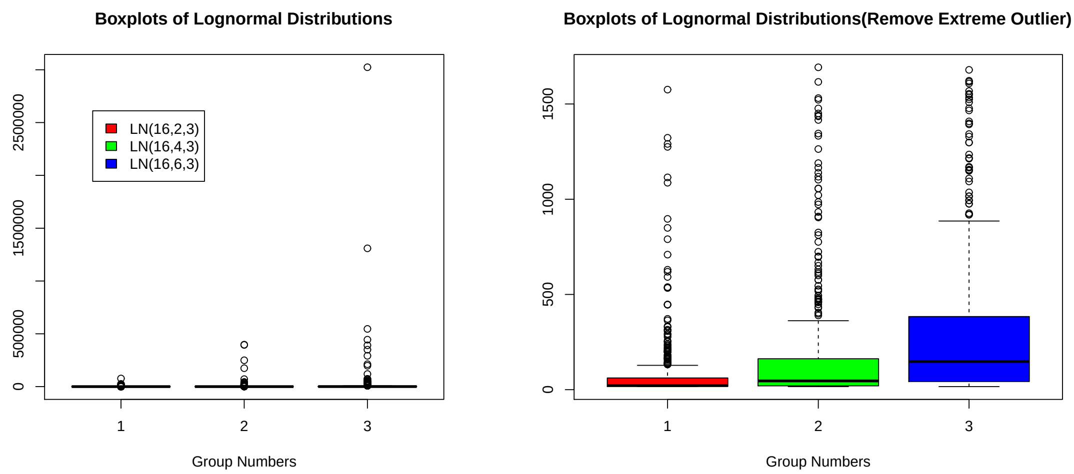
[2] * Note: The Type I in the table means Actual Type I Error.

Table 2.4: Power for the Scale Test: Case II^a (Cauchy Distribution)

Sample Size	**Parametric Test**		**Nonparametric Test**	
	Power	**Type I**	**Power**	**Type I**
15	0.6743	0.2994	0.7317	0.0456
20	0.7028	0.2847	0.8220	0.0597
25	0.7155	0.2795	0.9159	0.0602
30	0.7306	0.2758	0.9523	0.0569
35	0.7506	0.2731	0.9734	0.0476
40	0.7633	0.2704	0.9805	0.0410
45	0.7748	0.2798	0.9897	0.0377
50	0.7799	0.2646	0.9933	0.0339
55	0.7852	0.2659	0.9967	0.0339
60	0.7928	0.2697	0.9991	0.0329

[1] * Note: Given significance level: $\alpha = 0.05$.

[2] * Note: The Type I in the table means Actual Type I Error.



(a) Boxplot for All Data (b) Boxplot Without Extreme Outliers

Figure 2.5: Side-by-Side Boxplot for the 3 Groups in Case III^a: Lognormal Distributions

tests' actual type I error rates are close to $\alpha = 0.05$ in Case II^a, though it tends to be a little bit conservative when sample size is large ($n > 40$). However, this little conservative does not affect the powerfulness of our test when sample size is large. The power of the nonparametric test is 98% or above when sample size is beyond 40. Although Levene's test is robust to the departure from normality, it losses power very quickly when the distribution has thick tails. The kernel based nonparametric scale test significantly outperforms Levene's test for the heavy-tailed underlying distribution, such as Cauchy distribution.

Table 2.5 shows that Levene's test severely suffers from inflated Type I errors when the data come from strongly skewed distributions. The actually type I error of Levene's test is up to 0.6 when the nominal Type I error is just 0.05. The kernel based nonparametric test's actual type I errors are closer to $\alpha = 0.05$ compare the ones of Levene's test in Case III^a. Moreover, as the sample size increases, the problem of inflated Type I errors becomes weaker and weaker. When the sample size is

Table 2.5: Power for the Scale Test: Case III^a (Lognormal Distribution)

Sample Size	Parametric Test		Nonparametric Test	
	Power	**Type I**	**Power**	**Type I**
15	0.8671	0.5874	0.8458	0.1445
20	0.8792	0.5659	0.9192	0.1016
25	0.9021	0.5740	0.9591	0.0851
30	0.9130	0.5518	0.9799	0.0790
35	0.9196	0.5527	0.9927	0.0668
40	0.9304	0.5526	0.9971	0.0598
45	0.9381	0.5480	0.9991	0.0583
50	0.9424	0.5448	0.9995	0.0596
55	0.9477	0.5344	0.9998	0.0590
60	0.9524	0.5354	0.9998	0.0568

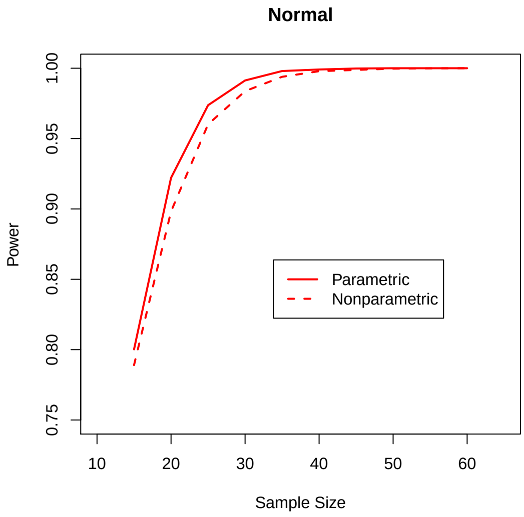
[1] * Note: Given significance level: $\alpha = 0.05$.

[2] * Note: The Type I in the table means Actual Type I Error.

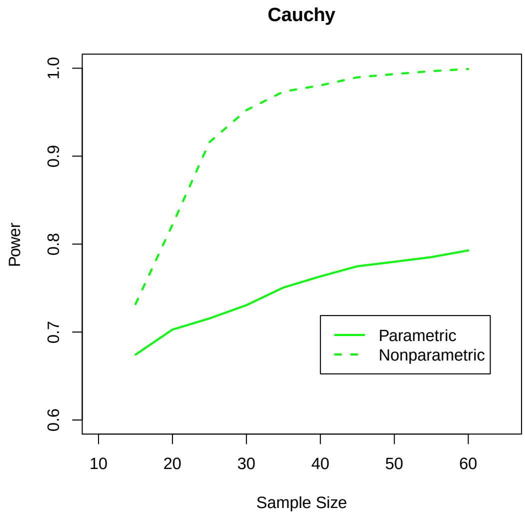
40 or above, the actual type I error of the test is very close to 0.05. Table 2.5 also demonstrates that the nonparametric test is very powerful. When the sample size is 30, the power is over 90%. To conclude, the kernel based nonparametric scale test significantly outperforms Levene's test for the strongly skewed underlying distribution, such as Lognormal distribution.

Figure 2.6 demonstrates the power of the parametric and the new nonparametric test with respect to sample size when the underlying distribution of the samples are Normal, Cauchy and Lognormal distribution. In Figure 2.6, the solid line represents the power of the parametric scale test (i.e. Levene's test), while the dashed line represents the power of our kernel based nonparametric test. The red line (solid and dashed) represents the power of test in Case I^a when the underlying distribution is Normal distribution. The green line (solid and dashed) represents the power of test in Case II^a when the underlying distribution is Cauchy distribution. And the blue line (solid and dashed) represents the power of test in Case III^a when the underlying distribution is Lognormal distribution.

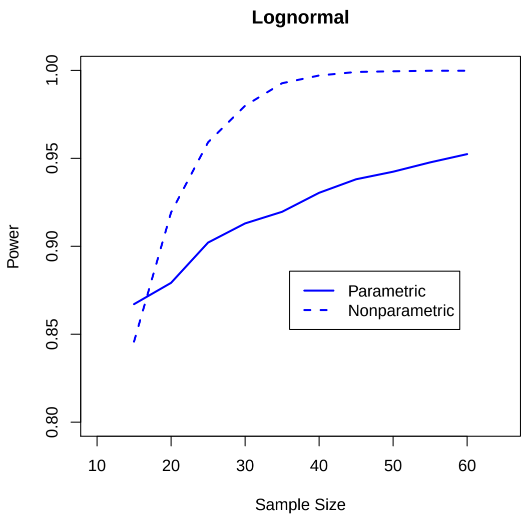
Figure 2.6(a) compares the power of parametric and nonparametric scale test on the three groups in Case I^a. It shows that the powers of both tests increase sharply to around 95% when sample size goes from 15 to 25 in Case I^a. Parametric scale test performs a little bit better than the nonparametric scale test we proposed for small sample size in Case I^a, which confirms what we withdrawn from Table 2.3. Figure 2.6(b) compares the power of parametric and nonparametric scale test on the three groups in Case II^a. It is shown in Figure 2.6(b) that the power of the parameter test does not increase as much as the nonparametric test as the sample size increases. The power of nonparametric test has a steep increase especially when sample size goes from 15 to 25 in Case II^a. Figure 2.6(c) compares the power of parametric and nonparametric scale test on the three groups in Case III^a. It is indicated that the power of parametric test is higher than the nonparametric test when the sample size



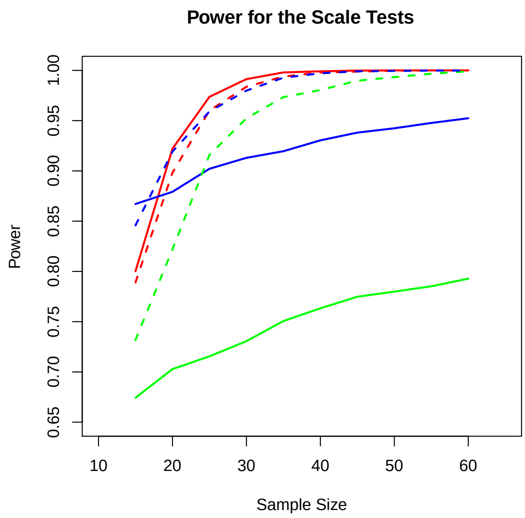
(a) Case I^a: Normal



(b) Case II^a: Cauchy



(c) Case III^a: Lognormal



(d) Case I^a,II^a, III^a

Figure 2.6: (a) Power of the parametric and nonparametric scale test on the three groups in Case I^a; (b) Power of the parametric and nonparametric scale test on the three groups in Case II^a; (c) Power of the parametric and nonparametric scale test on the three groups in Case III^a; (d) Power of the parametric and nonparametric scale test on the three groups in Case I^a, II^a and III^a.

is very small, then the relationship is reversed when the sample size goes beyond 16 or so. Like in Case II^a, the power of nonparametric test increases way faster than the parametric test in Case III^a. Figure 2.6(d) combines Figure (2.6(a))(2.6(b))(2.6(c)) in one graph, so it is easier to compare the powers of the either test throughout the 3 cases. The relationship of three solid lines infers that the power of the parametric scale test is far more severely hurt by fat tails or extreme outliers than the skewness does. The relationship of three dashed lines infers that the power of the nonparametric scale test is not sensitive to the fat tails or the skewness. As long as the sample size is large, it is a very powerful test.

2.3.2 Simulation Study for the One-way ANOVA Tests

In this section we will evaluate the performance of our new proposed nonparametric location test, i.e. kernel based nonparametric one-way ANOVA test, and compare with the traditional parametric F test. Like the performance study of scale test, the power of nonparametric and parametric ANOVA tests in 3 distinctive cases listed in Table 2.1 are studies as well. To evaluate the actual Type I error rate and the power of the parametric F test and nonparametric location tests proposed in section 2.2 in the 3 cases in Table 2.1, we test the equality of location parameters of three groups, i.e. $K = 3$ for each case. To obtain the actual Type I error while setting the significant level $\alpha = 0.05$, we follow the steps below:

(1) Randomly generate 3 groups of data with balanced sample size n from N(3,1) for Case I, Cauchy(10, 2) for Case II and Lognormal(6, 2, 1) for Case III.

(2) Apply the F test and the location test separately. Keep the test result as 1 or 0. 1 means "reject the null hypothesis" and 0 means fail to reject.

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

(4) Repeat (1) -(3) for sample size $n = 15, 20, 25, 30, 35, 40, 45, 50, 55, 60$.

The procedure to calculate the empirical power is very similar except generating 3 groups of data from 3 different distributions for each case in Step (1). The distribution types and parameters assigned to each group in each case are listed in Table 2.6.

Table 2.6: Evaluate the Power of ANOVA Tests in 3 Cases

Case	**Group1**	**Group2**	**Group3**
I^b	N(2.5, 1)	N(3, 1)	N(3.5, 1)
II^b	Cauchy(2, 2)	Cauchy(10, 2)	Cauchy(20, 2)
III^b	Lognormal(2, 2, 1)	Lognormal(6, 2, 1)	Lognormal(10, 2, 1)

As shown in Table 2.6, in Case I^b, all the three groups are from normal with the same standard deviation 1, but mean 2.5, 3 and 3.5. The side-by-side boxplot of the three groups in Case I^b is given in Figure 2.7. In Figure 2.7, the plot with the box painted in red is the boxplot for group 1, which is generated from $N(2.5, 1)$. The plot painted in green is the boxplot for group 2, which is generated from $N(3, 1)$. And the plot painted in blue is the boxplot for group 3, which is generated from $N(3.5, 1)$. It is not hard to tell that the three groups have the same variability, but different centralities.

The side-by-side boxplot of the three groups in Case II^b is given in Figure 2.8. In Figure 2.8, the plot with the box painted in red is the boxplot for group 1, which is generated from $Cauchy(2, 2)$. The plot painted in green is the boxplot for group 2, which is generated from $Cauchy(10, 2)$. And the plot painted in blue is the boxplot for group 3, which is generated from $Cauchy(20, 2)$. The color of the boxes in Figure 2.8(a) is invisible, since Cauchy distribution has very fat tails. Figure 2.8(b) shows the boxplot of the three groups when kicking out the extreme outliers, i.e. the points that are smaller than the first quartile subtract 3 times inter-quartile range or bigger than the third quartile plus 3 times inter-quartile range. It is clear to see in Figure

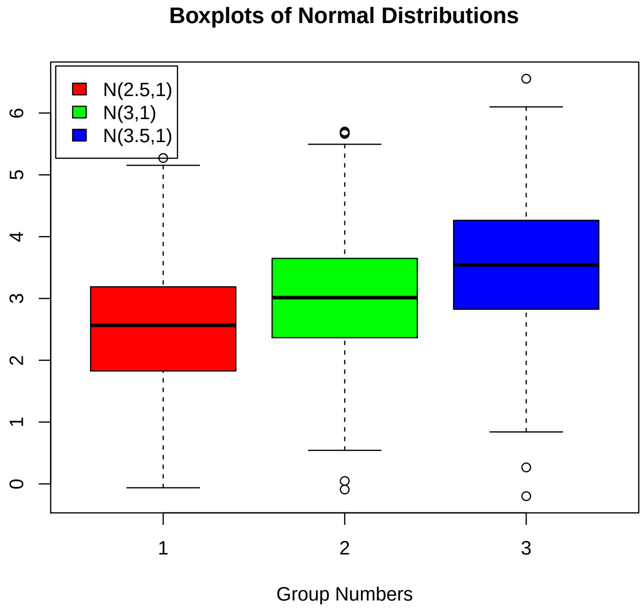


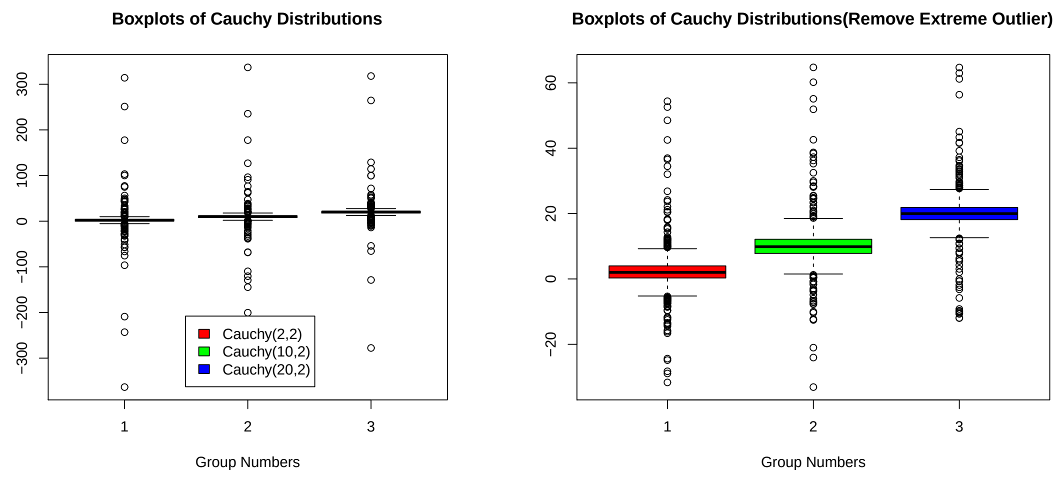
Figure 2.7: Side-by-Side Boxplot for the 3 Groups in Case I^b: Normal Distributions

2.8(b) that the three groups have the same variability, but different centralities.

The side-by-side boxplot of the three groups in Case III^b is given in Figure 2.9. In Figure 2.9, the plot with the box painted in red is the boxplot for group 1, which is generated from lognormal distribution with location parameter 16, scale parameter 2 and shape parameter 3, i.e. $Lognormal(2, 2, 1)$. The plot painted in green is the boxplot for group 2, which is generated from $Lognormal(6, 2, 1)$. And the plot painted in blue is the boxplot for group 3, which is generated from $Lognormal(10, 2, 1)$. Although the distributions are strongly right skewed, the difference in central measurements is still visible from Figure 2.9.

Similar to the scale test, empirical powers of the location or ANOVA test can be evaluated for each case and a series of sample sizes. The simulation results are listed in Table 2.7 - Table 2.9.

Table 2.7 shows that the actual Type I errors for both tests, either parametric and nonparametric, are around 0.05, the significance level. As we expected, in Case



(a) Boxplot for All Data (b) Boxplot Without Extreme Outliers

Figure 2.8: Side-by-Side Boxplot for the 3 Groups in Case II^b: Cauchy Distributions

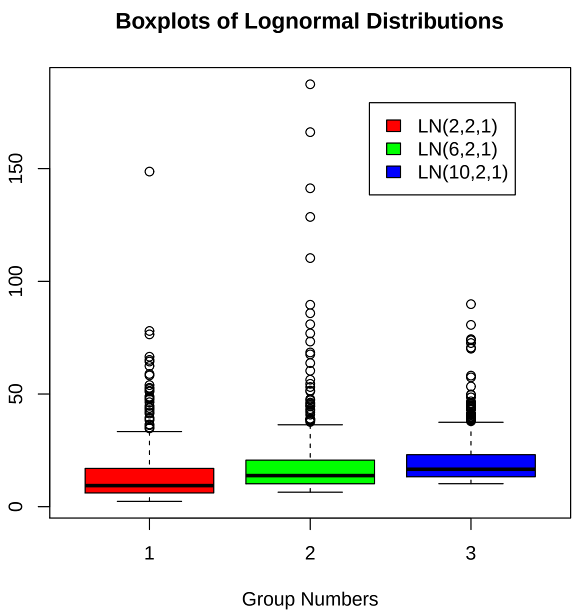


Figure 2.9: Side-by-Side Boxplot for the 3 Groups in Case III^b: Lognormal Distributions

Table 2.7: Power for the ANOVA Test: Case I^b (Normal Distribution)

Sample Size	**Parametric Test**		**Nonparametric Test**	
	Power	**Type I**	**Power**	**Type I**
15	0.6511	0.0501	0.6161	0.0471
20	0.7980	0.0515	0.7719	0.0472
25	0.8820	0.0468	0.8622	0.0461
30	0.9325	0.0495	0.9252	0.0487
35	0.9676	0.0463	0.9590	0.0457
40	0.9822	0.0503	0.9800	0.0484
45	0.9914	0.0495	0.9874	0.0469
50	0.9961	0.0505	0.9946	0.0491
55	0.9978	0.0499	0.9972	0.0455
60	0.9991	0.0543	0.9991	0.0508

I, parametric F test performs a little bit better when the sample size is less than 40. When the samples come from normal distribution, parametric tests are always the best choice for small sample size. However, as shown in the Table 2.7, our test is as good as the parametric test even in the normal case. The power of our nonparametric ANOVA test is around 92% when the sample size is 30.

Table 2.8: Power for the ANOVA Test: Case II^b(Cauchy Distribution)

Sample Size	**Parametric Test**		**Nonparametric Test**	
	Power	**Type I**	**Power**	**Type I**
15	0.6224	0.0177	0.7798	0.0110
20	0.6308	0.0177	0.7838	0.0156
25	0.6357	0.0160	0.8004	0.0157
30	0.6208	0.0184	0.8181	0.0155
35	0.6358	0.0181	0.8404	0.0141
40	0.6236	0.0177	0.8555	0.0136
45	0.6407	0.0199	0.8719	0.0165
50	0.6390	0.0171	0.8780	0.0156
55	0.6442	0.0171	0.8887	0.0166
60	0.641	0.0169	0.8962	0.0188

[1] * Note: Given significance level: $\alpha = 0.05$.

[2] * Note: The Type I in the table means Actual Type I Error.

Table 2.8 shows that both parametric and nonparametric test in Case II^b are equally conservative since the actual Type I errors for both tests are around 0.01-0.02, which is less than the significance level $\alpha = 0.05$. It is known that conservative tests to some degree reduce the power. However, our kernel based nonparametric test still get pretty decent power compared to the parametric F test, around 86% when the sample size is 40. Moreover, the power of the our nonparametric test in Case II^b increases as the sample size increases, although not as fast as in Case I^b and III^b. The parameter test does not benefit from the growth of sample size. So the kernel based nonparametric ANOVA test significantly outperforms parametric ANOVA test for the heavy-tailed underlying distribution, such as Cauchy distribution.

Table 2.9: Power for the ANOVA Test: Case III^b (Lognormal Distribution)

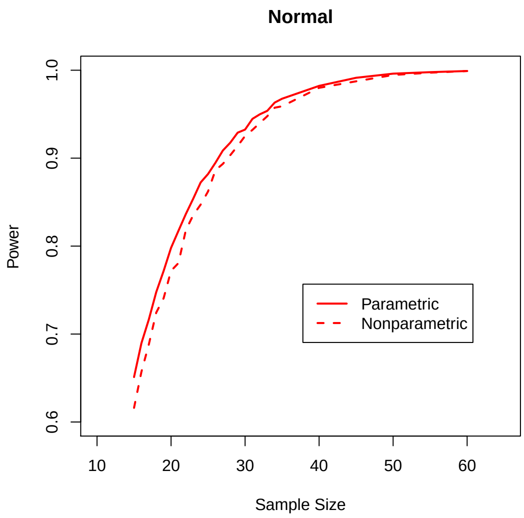
Sample Size	**Parametric Test**		**Nonparametric Test**	
	Power	**Type I**	**Power**	**Type I**
15	0.3202	0.0349	0.3898	0.0413
20	0.3913	0.0361	0.4854	0.0377
25	0.4482	0.0392	0.5866	0.0379
30	0.4848	0.0408	0.6741	0.0399
35	0.5407	0.0387	0.7434	0.0425
40	0.5945	0.0403	0.8093	0.0433
45	0.6340	0.0407	0.8562	0.0426
50	0.6671	0.0442	0.8917	0.0420
55	0.6987	0.0415	0.9203	0.0457
60	0.7336	0.0406	0.9423	0.0456

Table 2.9 shows that the parametric and nonparametric F test in Case III^b are a little bit conservative since the actual Type I errors for both tests are around 0.04. But the kernel based nonparametric test tends to be less conservative as sample size increases. Furthermore, as the sample size increases, the power of the nonparametric test grows faster than the parametric test. Table 2.9 also demonstrates that the nonparametric test is very powerful. When the sample size is 40, the power of the nonparametric test is over 80%, while the power of the parametric test is below 60%. To conclude, the kernel based nonparametric ANOVA test significantly outperforms

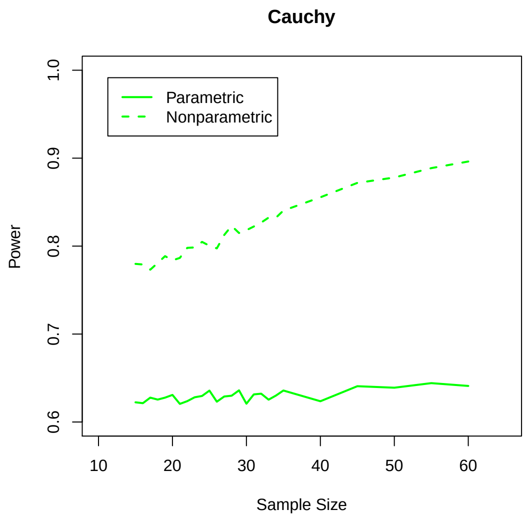
tradition F test for the strongly skewed underlying distribution, such as lognormal distribution.

Figure 2.10 demonstrates the power of the parametric and nonparametric test with respect to sample size when the underlying distribution of the samples are normal, cauchy and lognormal distribution. In Figure 2.10, the solid line represents the power of the parametric location test (i.e. traditional F test), while the dashed line represents the power of our kernel based nonparametric ANOVA test. The red line (solid and dashed) represents the power of test in Case I^b when the underlying distribution is normal distribution. The green line (solid and dashed) represents the power of test in Case II^b when the underlying distribution is Cauchy distribution. And the blue line (solid and dashed) represents the power of test in Case III^b when the underlying distribution is lognormal distribution.

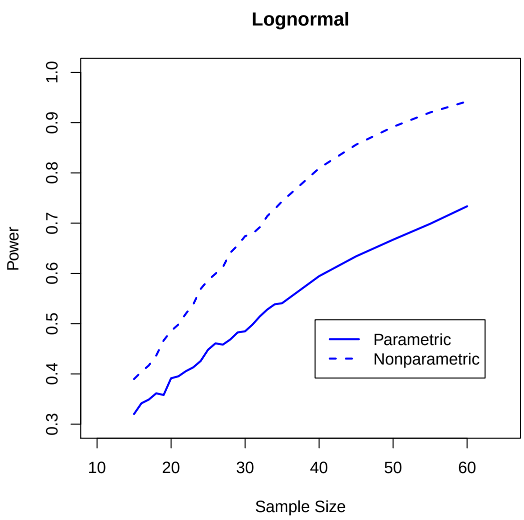
Figure 2.10(a) compares the power of parametric and nonparametric ANOVA test on the three groups in Case I^b. It shows that the powers of both tests increase sharply to around 95% when sample size goes from 15 to 30 in Case I^b. Parametric ANOVA test performs a little bit better than the kernel based nonparametric ANOVA test we proposed in Case I^b, which is consistent with what we concluded from Table 2.7. Figure 2.10(b) compares the power of parametric and nonparametric scale test on the three groups in Case II^b. It is shown in Figure 2.10(b) that the power of our nonparametric ANOVA test grows as the sample size increases, while the power of parametric ANOVA test almost keeps constant in Case II^b. Figure 2.10(c) compares the power of parametric and nonparametric scale test on the three groups in Case III^b. It is indicated that the power of nonparametric test increase way faster than the parametric test in Case III^b. Figure 2.10(d) combines Figure (2.10(a))(2.10(b))(2.10(c)) in one graph, so it is easier to compare the powers of the either test throughout the 3 cases. The relationship of three solid lines infers that the power of the parametric ANOVA test is far more severely hurt by fat tails or extreme outliers than the skewness does.



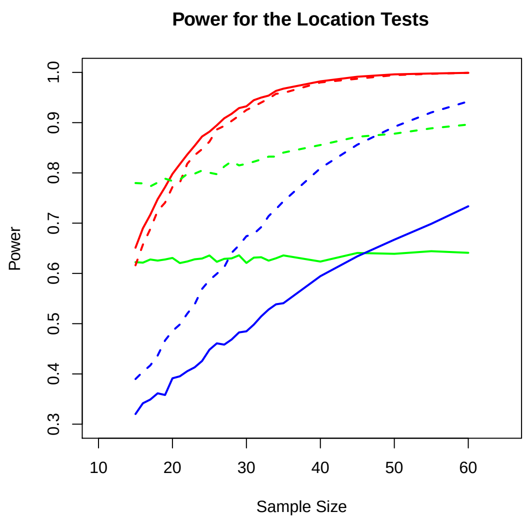
(a) Case I^b: Normal



(b) Case II^b: Cauchy



(c) Case III^b: Lognormal



(d) Case I^b, II^b, III^b

Figure 2.10: (a) Power of the parametric and nonparametric ANOVA test on the three groups in Case I^b; (b) Power of the parametric and nonparametric ANOVA test on the three groups in Case II^b; (c) Power of the parametric and nonparametric ANOVA test on the three groups in Case III^b; (d) Power of the parametric and nonparametric ANOVA test on the three groups in Case I^b, II^b and III^b.

If the data come from a skewed distribution, large sample size leads to better power. However, if the data come from a fat-tailed distribution, large sample size would not make any improvement in power. The relationship of three dashed lines infers that the power of the nonparametric one-way ANOVA test is not sensitive to the fat tails or the skewness. Larger sample size always helps.

2.4 One-way Kernel Based Nonparametric Test for Shape Parameters

The parameter θ is called shape parameter if it satisfies $f_\theta(x) = \theta x^{\theta-1} f_{1s}(x^\theta)$, where f_{1s} is the base density. Thus, if $f_i(x) = \theta_i x^{\theta_i - 1} f_{1s}(x^{\theta_i})$ for $i = 1, 2, \cdots, K$, then

$$\begin{aligned} \int x f_i^2(x)dx &= \theta_i^2 \int x^{2\theta_i - 1} f_{1s}^2(x^{\theta_i})dx \\ &= \theta_i \int y f_{1s}^2(y)dy. \end{aligned} \tag{2.54}$$

Let $W = \int y f_{1s}^2(y)dy$, then $V_i = \theta_i W$. Thus, testing H_0: $\theta_1 = \theta_2 = \cdots = \theta_K$ versus H_a: $\theta_i \neq \theta_j$ for any $i \neq j$ is equivalent to test H_0: $V_1 = V_2 = \cdots = V_K$ versus H_a: $V_i \neq V_j$. Note that the kernel estimate of V_i is given in equation (2.19). Therefore, the test statistic for testing shape parameters is exactly the same as the location parameters.

CHAPTER 3

Two-way Kernel Based Nonparametric ANOVA

Assume X_{ijk} comes from a distribution with probability density function $f_{ij}(x)$, where $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, c$ and $k = 1, 2, \ldots, n_{ij}$. Define μ_{ij} and σ_{ij} as the location and scale parameter of $f_{ij}(x)$. That is to say, the following equation holds.

$$f_{ij}(x) = \frac{1}{\sigma_{ij}} f_{00}\left(\frac{x - \mu_{ij}}{\sigma_{ij}}\right), \tag{3.1}$$

where f_{00} is the base density.

In Ahmad and Amezziane [2], the location and scale parameters can be written as:

$$\mu_{ij} = \frac{\sigma_{ij}(\int_{-\infty}^{\infty} x f_{ij}^2(x)dx - \int_{-\infty}^{\infty} x f_{00}^2(x)dx)}{\int_{-\infty}^{\infty} f_{00}^2(x)dx}, \tag{3.2}$$

and

$$\sigma_{ij} = \frac{\int_{-\infty}^{\infty} f_{00}^2(x)dx}{\int_{-\infty}^{\infty} f_{ij}^2(x)dx}. \tag{3.3}$$

We can test the homogeneity of scale parameters among the rc cells by using the test statistics proposed in Chapter 2.

3.1 Two-way Kernel Based Nonparametric Test for Location Parameters with Equal Scale Parameter

Under the assumption of homogeneity of scale parameters, i.e. $\sigma_{ij} = \sigma$ for all i and j, we have

$$\mu_{ij} = \frac{\sigma(\int_{-\infty}^{\infty} x f_{ij}^2(x)dx - \int_{-\infty}^{\infty} x f_{00}^2(x)dx)}{\int_{-\infty}^{\infty} f_{00}^2(x)dx}. \tag{3.4}$$

Define $V_{ij} = \int_{-\infty}^{\infty} x f_{ij}^2(x)dx$, $C_1^0 = \frac{\sigma}{\int_{-\infty}^{\infty} f_{00}^2(x)dx}$ and $C_2^0 = -\frac{\int_{-\infty}^{\infty} x f_{00}^2(x)dx}{\int_{-\infty}^{\infty} f_{00}^2(x)dx}$. Then, μ_{ij} can be rewritten as

$$\mu_{ij} = C_1^0 V_{ij} + C_2^0. \tag{3.5}$$

Consider the two way layout: $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, where μ is the overall location, α_i is the i^{th} row effect, β_j is the j^{th} column effect and γ_{ij} is the interaction effect of i^{th} row and j^{th} column. The decomposition is not unique, so we impose the following restrictions: $\sum_i \frac{\lambda_{ij}}{\omega_{ij}^2}\alpha_i = 0$, $\sum_j \frac{\lambda_{ij}}{\omega_{ij}^2}\beta_j = 0$ and $\sum_i \frac{\lambda_{ij}}{\omega_{ij}^2}\gamma_{ij} = \sum_j \frac{\lambda_{ij}}{\omega_{ij}^2}\gamma_{ij} = 0$, where $\lambda_{ij} = \lim_{\min_{i,j} n_{ij}\to\infty} \dfrac{n_{ij}}{N}$ and $N = \sum_{i,j} n_{ij}$. Thus, we can infer that

$$\mu + \alpha_i = \frac{\sum_j \lambda_{ij}\mu_{ij}/\omega_{ij}^2}{\sum_j \lambda_{ij}/\omega_{ij}^2}, \quad (3.6)$$

$$\mu + \beta_j = \frac{\sum_i \lambda_{ij}\mu_{ij}/\omega_{ij}^2}{\sum_i \lambda_{ij}/\omega_{ij}^2}, \quad (3.7)$$

$$\mu = \frac{\sum_i\sum_j \lambda_{ij}\mu_{ij}/\omega_{ij}^2}{\sum_i\sum_j \lambda_{ij}/\omega_{ij}^2}. \quad (3.8)$$

By plugging (3.5) into equations (3.6-3.8), we obtain

$$\mu + \alpha_i = C_1^0 \bar{V}_{i.} + C_2^0, \quad (3.9)$$

$$\mu + \beta_j = C_1^0 \bar{V}_{.j} + C_2^0, \quad (3.10)$$

$$\mu = C_1^0 \bar{V}_{..} + C_2^0, \quad (3.11)$$

where

$$\bar{V}_{i.} = \frac{\sum_j m_{ij}V_{ij}}{m_{i.}},$$

$$\bar{V}_{.j} = \frac{\sum_i m_{ij}V_{ij}}{m_{.j}},$$

$$\bar{V}_{..} = \frac{\sum_i\sum_j m_{ij}V_{ij}}{m_{..}},$$

$m_{ij} = N\lambda_{ij}/\omega_{ij}^2$, $m_{i.} = \sum_j m_{ij}$, $m_{.j} = \sum_i m_{ij}$, and $m_{..} = \sum_i\sum_j m_{ij}$. Thus, by some algebra, we have

$$\alpha_i = C_1^0(\bar{V}_{i.} - \bar{V}_{..}), \quad (3.12)$$

$$\beta_j = C_1^0(\bar{V}_{.j} - \bar{V}_{..}), \quad (3.13)$$

$$\begin{aligned}\gamma_{ij} &= \mu_{ij} - (\mu + \alpha_i) - (\mu + \beta_j) + \mu. \\ &= C_1^0(V_{ij} - \bar{V}_{i.} - \bar{V}_{.j} + \bar{V}_{..}). \quad (3.14)\end{aligned}$$

Thus, the hypothesis for testing homogeneous row effects, i.e. $H_0 : \alpha_i = 0$ for all i versus $H_1 : \alpha_i \neq 0$ for some i, becomes $H_0 : \bar{V}_{i.} = \bar{V}_{..}$ for all i versus $H_1 : \bar{V}_{i.} \neq \bar{V}_{..}$ for some i. Similarly, the hypothesis for testing homogeneous column effects, i.e. $H_0 : \beta_j = 0$ for all j versus $H_1 : \beta_j \neq 0$ for some j, becomes $H_0 : \bar{V}_{.j} = \bar{V}_{..}$ for all j versus $H_1 : \bar{V}_{.j} \neq \bar{V}_{..}$ for some j, and the hypothesis for testing the homogeneous interaction of the row and column effects, i.e. $H_0 : \gamma_{ij} = 0$ for all i, j versus $H_1 : \gamma_{ij} \neq 0$ for some i, j, becomes $H_0 : V_{ij} - \bar{V}_{i.} - \bar{V}_{.j} + \bar{V}_{..} = 0$ for all i, j versus $H_1 : V_{ij} - \bar{V}_{i.} - \bar{V}_{.j} + \bar{V}_{..} \neq 0$ for some i, j.

3.1.1 Kernel Based Nonparametric Test for Main Effects

Consider the nonparametric kernel estimate of V_{ij}, denoted as $\hat{V}_{ij}$, where

$$\hat{V}_{ij} = \frac{1}{n_{ij}(n_{ij}-1)h_{ij}} \sum \sum_{k_1 \neq k_2} \left(\frac{X_{ijk_1} + X_{ijk_2}}{2} \right) K \left(\frac{X_{ijk_1} - X_{ijk_2}}{h_{ij}} \right). \tag{3.15}$$

Lemma 3.1 *If for any $i = 1, 2, \cdots, r$, $j = 1, 2, \cdots, c$, $n_{ij}h_{ij} \to \infty$ as $\min_{i,j} n_{ij} \to \infty$, $\int u f^2(u) du < \infty$ and $\int u^2 f(u) du < \infty$, then*

$$\sqrt{n_{ij}}(\hat{V}_{ij} - V_{ij}) \xrightarrow{d} N(0, \omega_{ij}^2), \tag{3.16}$$

as $\min_{i,j} n_{ij} \to \infty$, where, $\omega_{ij}^2 = 4\{\int x^2 f_{ij}^3(x) dx - (\int x f_{ij}^2(x) dx)^2\}$.

Proof: The proof is similar to Lemma 2.4.

■

Consider the test statistic for row effect first. To test $H_0 : \bar{V}_{i.} = \bar{V}_{..}$ for all i, define the Row Sum of Squares (SSR) as

$$SSR = \sum_{i=1}^{r} \hat{m}_{i.} \left(\hat{V}_{i.} - \hat{V}_{..} \right)^2, \tag{3.17}$$

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

To obtain the asymptotic distribution of SSR, another auxiliary variable is defined as follows:

$$S_R^0 = \sum_{i=1}^{r} m_{i.} \left(V_{i.}^* - V_{..}^* \right)^2, \tag{3.18}$$

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

Lemma 3.2 *Let* $N = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}$. *If* $\lambda_{ij} = \lim_{\min n_{ij}\to\infty} \frac{n_{ij}}{N}$ *and* $\hat{\omega}_{ij}^2 \xrightarrow{p} \omega_{ij}^2$, *then* $S_R^0 - SSR \xrightarrow{p} 0$, *as* $\min_{i,j} n_{i,j} \to \infty$.

Proof: This can be proved directly by applying Slutsky Theorem [10].

■

Theorem 3.3 *Under the null hypothesis,* $H_0 : \bar{V}_{i.} = \bar{V}_{..}$, *if for any* $i = 1, 2, \cdots, r$ *and* $j = 1, 2, \cdots, c$, $n_{ij}h_{ij}^4 \to 0$, $n_{ij}h_{ij} \to \infty$ *as* $\min_{i,j} n_{ij} \to \infty$, *and if* $\int x^2 f_{ij}^3(x)dx < \infty$, *then SSR is asymptotically* $\chi^2(r-1)$.

Proof: Set $T_{ij}^{(3)} = \sqrt{N\lambda_{ij}}(\hat{V}_{ij} - V_{ij})/\omega_{ij}$. Then $T_{ij}^{(3)} \overset{a}{\sim} N(0,1)$ as $N \to \infty$ by Lemma (3.1). Note that, under the null hypothesis,

$$
\begin{aligned}
S_R^0 &= \sum_{i=1}^{r} m_{i.}\Big(V_{i.}^* - V_{..}^*\Big)^2 \\
&= \sum_{i=1}^{r} m_{i.}\Big((V_{i.}^* - \bar{V}_{i.}) - (V_{..}^* - \bar{V}_{..})\Big)^2 \\
&= \sum_{i=1}^{r} m_{i.}\Big((V_{i.}^* - \bar{V}_{i.})^2 + (V_{..}^* - \bar{V}_{..})^2 - 2(V_{i.}^* - \bar{V}_{i.})(V_{..}^* - \bar{V}_{..})\Big) \\
&= \sum_{i=1}^{r} m_{i.}(V_{i.}^* - \bar{V}_{i.})^2 + m_{..}(V_{..}^* - \bar{V}_{..})^2 - 2(V_{..}^* - \bar{V}_{..})\sum_{i=1}^{r} m_{i.}\frac{\sum_j m_{ij}(\hat{V}_{ij} - V_{ij})}{m_{i.}} \\
&= \sum_{i=1}^{r} m_{i.}(V_{i.}^* - \bar{V}_{i.})^2 - m_{..}(V_{..}^* - \bar{V}_{..})^2 \\
&= \sum_{i=1}^{r} \frac{1}{m_{i.}}\Big(\sum_{j=1}^{c} m_{ij}(\hat{V}_{ij} - V_{ij})\Big)^2 - \frac{1}{m_{..}}\Big(\sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(\hat{V}_{ij} - V_{ij})\Big)^2 \\
&= \sum_{i=1}^{r}\sum_{j_1=1}^{c}\sum_{j_2=1}^{c} \frac{N}{m_{i.}} \frac{\sqrt{N\lambda_{ij_1}}}{\omega_{ij_1}}(\hat{V}_{ij_1} - V_{ij_1})\frac{\sqrt{N\lambda_{ij_2}}}{\omega_{ij_2}}(\hat{V}_{ij_2} - V_{ij_2})\frac{\sqrt{\lambda_{ij_1}}}{\omega_{ij_1}}\frac{\sqrt{\lambda_{ij_2}}}{\omega_{ij_2}} \qquad (3.19) \\
&- \frac{N}{m_{..}}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}\sum_{j_1=1}^{c}\sum_{j_2=1}^{c} \frac{\sqrt{N\lambda_{i_1j_1}}}{\omega_{i_1j_1}}(\hat{V}_{i_1j_1} - V_{i_1j_1})\frac{\sqrt{N\lambda_{i_2j_2}}}{\omega_{i_2j_2}}(\hat{V}_{i_2j_2} - V_{i_2j_2})\frac{\sqrt{\lambda_{i_1j_1}}}{\omega_{i_1j_1}}\frac{\sqrt{\lambda_{i_2j_2}}}{\omega_{i_2j_2}}.
\end{aligned}
$$

Let $\mathbf{U_3} = (T_{11}^{(3)}, T_{12}^{(3)}, \cdots, T_{rc}^{(3)})$, then the first term of (3.19) can written as a quadratic form $\mathbf{U_3}'\mathbf{M^{(1)}}\mathbf{U_3}$, and the second term of (3.19) can written as a quadratic form $\mathbf{U_3}'\mathbf{M^{(2)}}\mathbf{U_3}$, where

$$
\mathbf{M^{(1)}} = \begin{bmatrix} \frac{N}{m_{1.}}\mathbf{M_{11}} & 0 & \cdots & 0 \\ 0 & \frac{N}{m_{2.}}\mathbf{M_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{N}{m_{r.}}\mathbf{M_{rr}} \end{bmatrix}, \qquad (3.20)
$$

$$
\mathbf{M^{(2)}} = \begin{bmatrix} \frac{N}{m_{..}}\mathbf{M_{11}} & \frac{N}{m_{..}}\mathbf{M_{12}} & \cdots & \frac{N}{m_{..}}\mathbf{M_{1r}} \\ \frac{N}{m_{..}}\mathbf{M_{21}} & \frac{N}{m_{..}}\mathbf{M_{22}} & \cdots & \frac{N}{m_{..}}\mathbf{M_{2r}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{N}{m_{..}}\mathbf{M_{r1}} & \frac{N}{m_{..}}\mathbf{M_{r2}} & \cdots & \frac{N}{m_{..}}\mathbf{M_{rr}} \end{bmatrix}, \qquad (3.21)
$$

and

$$\mathbf{M_{ij}} = \begin{bmatrix} \frac{\sqrt{\lambda_{i1}}\sqrt{\lambda_{j1}}}{\omega_{i1}\omega_{j1}} & \frac{\sqrt{\lambda_{i1}}\sqrt{\lambda_{j2}}}{\omega_{i1}\omega_{j2}} & \cdots & \frac{\sqrt{\lambda_{i1}}\sqrt{\lambda_{jc}}}{\omega_{i1}\omega_{jc}} \\ \frac{\sqrt{\lambda_{i2}}\sqrt{\lambda_{j1}}}{\omega_{i2}\omega_{j1}} & \frac{\sqrt{\lambda_{i2}}\sqrt{\lambda_{j2}}}{\omega_{i2}\omega_{j2}} & \cdots & \frac{\sqrt{\lambda_{i2}}\sqrt{\lambda_{jc}}}{\omega_{i2}\omega_{jc}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sqrt{\lambda_{ic}}\sqrt{\lambda_{j1}}}{\omega_{ic}\omega_{j1}} & \frac{\sqrt{\lambda_{ic}}\sqrt{\lambda_{j2}}}{\omega_{ic}\omega_{j2}} & \cdots & \frac{\sqrt{\lambda_{ic}}\sqrt{\lambda_{jc}}}{\omega_{ic}\omega_{jc}} \end{bmatrix} \tag{3.22}$$

for all $i = 1, 2, \cdots, r$, $j = 1, 2, \cdots, c$. Thus, S_R^0 can be rewritten by the following quadratic form:

$$\begin{aligned} S_R^0 &= \mathbf{U_3}'\mathbf{M^{(1)}}\mathbf{U_3} - \mathbf{U_3}'\mathbf{M^{(2)}}\mathbf{U_3} \\ &= \mathbf{U_3}' \begin{bmatrix} (\frac{N}{m_{1.}} - \frac{N}{m_{..}})\mathbf{M_{11}} & -\frac{N}{m_{..}}\mathbf{M_{12}} & \cdots & -\frac{N}{m_{..}}\mathbf{M_{1r}} \\ -\frac{N}{m_{..}}\mathbf{M_{21}} & (\frac{N}{m_{2.}} - \frac{N}{m_{..}})\mathbf{M_{22}} & \cdots & -\frac{N}{m_{..}}\mathbf{M_{2r}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{N}{m_{..}}\mathbf{M_{r1}} & -\frac{N}{m_{..}}\mathbf{M_{r2}} & \cdots & (\frac{N}{m_{r.}} - \frac{N}{m_{..}})\mathbf{M_{rr}} \end{bmatrix} \mathbf{U_3} \\ &\overset{def}{=} \mathbf{U_3}'\mathbf{B_5}\mathbf{U_3}. \end{aligned} \tag{3.23}$$

It can easily be shown that $\mathbf{B_5}$ is symmetric and idempotent. Thus, we obtain that

$$\begin{aligned} rank(\mathbf{B_5}) &= trace(\mathbf{B_5}) \\ &= \sum_{i=1}^{r} \frac{1}{m_{i.}} \sum_{j=1}^{c} m_{ij} - \frac{1}{m_{..}} \sum_{i=1}^{r} \sum_{j=1}^{c} m_{ij} \\ &= r - 1 \end{aligned} \tag{3.24}$$

$\mathbf{U_3}$ follows approximately multivariate normal with mean $\mathbf{0}$ and variance $\mathbf{I}$, since $T_{ij}^{(3)}$'s independently follow univariate standard normal distribution. Therefore, S_R^0 is asymptotically χ^2 with degrees of freedom $r-1$ under H_0. By Lemma (3.2), SSR is asymptotically χ^2 with degrees of freedom $r-1$ under H_0.

■

Let $A_{ijk_1k_2} = \frac{1}{h_{ij}}\left(\frac{X_{ijk_1}+X_{ijk_2}}{2}\right)K\left(\frac{X_{ijk_1}-X_{ijk_2}}{h_{ij}}\right)$. Then $\hat{V}_{ij}$ can be rewritten as:

$$\hat{V}_{ij} = \frac{1}{n_{ij}(n_{ij}-1)} \sum\sum_{k_1 \neq k_2} A_{ijk_1k_2}. \tag{3.25}$$

And the Within cells Sum of Squares (SSW) is given by

$$SSW = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1\neq k_2}(A_{ijk_1k_2} - \hat{V}_{ij})^2/\hat{\omega}_{ij}^2}{C_{w_2}}, \tag{3.26}$$

where

$$C_{w_2} = \begin{cases} \frac{n}{2} - 1 & \text{if } n_i = n \text{ for all i}, \\ c_0^{(2)} & \text{otherwise}, \end{cases} \tag{3.27}$$

and $c_0^{(2)} = \sum_{i=1}^{d_2} \pi_i^{(2)}/d_2$, and $\pi_1^{(2)}, \pi_2^{(2)}, \cdots, \pi_{d_2}^{(2)}$ are all the eigenvalues of $\mathbf{B}_6$, where

$$\mathbf{B}_6 = \begin{bmatrix} (\frac{n_{11}}{2}-1)(\mathbf{I}_{n_{11}} - \frac{1}{n_{11}}\mathbf{J}_{n_{11}}) & 0 & \cdots & 0 \\ 0 & (\frac{n_{12}}{2}-1)(\mathbf{I}_{n_{12}} - \frac{1}{n_{12}}\mathbf{J}_{n_{12}}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\frac{n_{rc}}{2}-1)(\mathbf{I}_{n_{rc}} - \frac{1}{n_{rc}}\mathbf{J}_{n_{rc}}) \end{bmatrix}. \tag{3.28}$$

To obtain the asymptotic distribution of SSW, another auxiliary variable is defined as follows:

$$S_W^0 = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1\neq k_2}(A_{ijk_1k_2} - \hat{V}_{ij})^2/\omega_{ij}^2. \tag{3.29}$$

Lemma 3.4 *Let* $N = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}$. *If* $\lambda_{ij} = \lim_{\min n_{ij}\to\infty} \frac{n_{ij}}{N}$ *and* $\hat{\omega}_{ij}^2 \xrightarrow{p} \omega_{ij}^2$, *then* $S_W^0 - C_{w_2}SSW \xrightarrow{p} 0$, *as* $\min_{i,j} n_{i,j} \to \infty$.

Proof: They can be proved directly by applying Slutsky Theorem [10]. ■

Theorem 3.5 *For any* $i = 1, 2, \cdots, r$ *and* $j = 1, 2, \cdots, c$, *if* $n_{ij}h_{ij}^4 \to 0$, $n_{ij}h_{ij} \to \infty$ *as* $\min_{i,j} n_{ij} \to \infty$, *and if* $\int x^2 f_{ij}^3(x)dx < \infty$, *then SSW follows asymptotically* χ^2 *with degrees of freedom* df_{w_2}, *where,*

$$df_{w_2} = \begin{cases} rc(n-1) & \text{if } n_i = n \text{ for all } i, \\ d_2 & \text{otherwise}. \end{cases} \tag{3.30}$$

Proof: By Hajek projection [16], $A_{ijk_1k_2}$ can be decomposed into the sum of conditional expected values and a residual as follows:

$$A_{ijk_1k_2} = E(A_{ijk_1k_2}|X_{ijk_1}) + E(A_{ijk_1k_2}|X_{ijk_2}) + O_p(n_i). \tag{3.31}$$

Set $\varphi(X_{ijk_1}) = E(A_{ijk_1k_2}|X_{ijk_1})$ and $\varphi(X_{ijk_2}) = E(A_{ijk_1k_2}|X_{ijk_2})$, thus

$$\begin{aligned}
\hat{V}_{ij} &= \frac{1}{n_{ij}(n_{ij}-1)} \sum\sum_{k_1 \neq k_2} A_{ijk_1k_2} \\
&\approx \frac{1}{n_{ij}(n_{ij}-1)} \sum\sum_{k_1 \neq k_2} (\varphi(X_{ijk_1}) + \varphi(X_{ijk_2})) \qquad (3.32) \\
&= \frac{1}{n_{ij}(n_{ij}-1)} \Big(\sum_{k_1}\sum_{k_2} (\varphi(X_{ijk_1}) + \varphi(X_{ijk_2})) - \sum\sum_{k_1 = k_2} (\varphi(X_{ijk_1}) + \varphi(X_{ijk_2})) \Big). \\
&= \frac{1}{n_{ij}(n_{ij}-1)} \Big(2n_{ij} \sum_{k_1} \varphi(X_{ijk_1}) - 2 \sum_{k_1} \varphi(X_{ijk_1}) \Big). \\
&= \frac{1}{n_{ij}} \sum_{k_1}^{n_{ij}} 2\varphi(X_{ijk_1}) \qquad (3.33)
\end{aligned}$$

Hence,

$$\begin{aligned}
\sum\sum_{k_1 \neq k_2} \frac{(A_{ijk_1k_2} - \hat{V}_{ij})^2}{\omega_{ij}^2} &\approx \sum\sum_{k_1 \neq k_2} \frac{(\varphi(X_{ijk_1}) + \varphi(X_{ijk_2}) - \hat{V}_{ij})^2}{\omega_{ij}^2} \\
&= \sum_{k_1}\sum_{k_2} \frac{(\varphi(X_{ijk_1}) - \frac{1}{n_{ij}} \sum \varphi(X_{ijk_1}) + \varphi(X_{ijk_2}) - \frac{1}{n_{ij}} \sum \varphi(X_{ijk_2}))^2}{\omega_{ij}^2} \\
&- \sum_{k_1} \frac{\big(2\varphi(X_{ijk_1}) - \frac{1}{n_{ij}} \sum_{k_1=1}^{n_{ij}} 2\varphi(X_{ijk_1})\big)^2}{\omega_{ij}^2} \\
&= 2n_{ij} \sum_{k_1} \frac{(\varphi(X_{ijk_1}) - \frac{1}{n_{ij}} \sum \varphi(X_{ijk_1}))^2}{\omega_{ij}^2} - \sum_{k_1} \frac{\big(2\varphi(X_{ijk_1}) - \hat{V}_{ij}\big)^2}{\omega_{ij}^2} \\
&= \frac{n_{ij}}{2} \sum_{k_1} \frac{(2\varphi(X_{ijk_1}) - \hat{V}_{ij})^2}{\omega_{ij}^2} - \sum_{k_1} \frac{\big(2\varphi(X_{ijk_1}) - \hat{V}_{ij}\big)^2}{\omega_{ij}^2} \\
&= (\frac{n_{ij}}{2} - 1) \sum_{k_1} \frac{(2\varphi(X_{ijk_1}) - \hat{V}_{ij})^2}{\omega_{ij}^2} \\
&= (\frac{n_{ij}}{2} - 1) \Big[\sum_{k=1}^{n_{ij}} \frac{(2\varphi(X_{ijk_1}) - V_{ij})^2}{\omega_{ij}^2} - \frac{n_{ij}(\hat{V}_{ij} - V_{ij})^2}{\omega_{ij}^2} \Big]. \qquad (3.34)
\end{aligned}$$

Let $H_{ijk} = \frac{2\varphi(X_{ijk})-V_{ij}}{\omega_{ij}}$ for $k = 1, 2, \cdots, n_{ij}$ and $\mathbf{H}_{ij} = (H_{ij1}, H_{ij2}, \cdots, H_{ijn_{ij}})'$ for $i = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, c$. Thus, equation(3.34) can be rewritten in matrix form as

$$\begin{aligned}
& (\frac{n_{ij}}{2} - 1)\Big[\sum_{j=1}^{n_{ij}} \frac{(2\varphi(X_{ijk_1}) - V_{ij})^2}{\omega_{ij}^2} - \frac{n_{ij}(\hat{V}_{ij} - V_{ij})^2}{\omega_{ij}^2}\Big] \\
= & (\frac{n_{ij}}{2} - 1)\Big[\mathbf{H}'_{ij}\mathbf{H}_{ij} - \mathbf{H}'_{ij}\frac{1}{n_{ij}}\mathbf{J}_{n_{ij}}\mathbf{H}_{ij}\Big] \\
= & \mathbf{H}'_{ij}(\frac{n_{ij}}{2} - 1)(\mathbf{I} - \frac{1}{n_{ij}}\mathbf{J}_{n_{ij}})\mathbf{H}_{ij}. \qquad (3.35)
\end{aligned}$$

Let $\mathbf{H} = (\mathbf{H}'_{11}, \mathbf{H}'_{12}, \cdots, \mathbf{H}'_{rc})'$. Therefore, S_W^0 can be written in the matrix form as

$$\begin{aligned}
S_W^0 &= \sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1 \neq k_2}^{n_{ij}} (A_{ijk_1k_2} - \hat{V}_{ij})^2/\omega_{ij}{}^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{c} \mathbf{H}'_{ij}(\frac{n_{ij}}{2} - 1)(\mathbf{I} - \frac{1}{n_{ij}}\mathbf{J}_{n_{ij}})\mathbf{H}_{ij} \\
&= \mathbf{H}'\mathbf{B}_6\mathbf{H}, \qquad (3.36)
\end{aligned}$$

where, $\mathbf{B}_6$ is given in equation (3.28).

Now we need to show that $\mathbf{H}$ follows asymptotically multivariate normal distribution. Note that $E(2\varphi(X_{ijk_1})) = E(\frac{1}{n_{ij}}\sum_{k_1=1}^{n_{ij}} 2\varphi(X_{ijk_1})) = E(\hat{V}_{ij}) \approx V_{ij}$ since $\hat{\mu}_{ij}$ is asymptotically unbiased by Lemma (3.1). Also $Var(2\varphi(X_{ijk_1})) = \frac{1}{n_{ij}}\sum_{k_1=1}^{n_{ij}} Var(2\varphi(X_{ijk_1})) = n_{ij}Var(\frac{1}{n_{ij}}\sum_{k_1=1}^{n_{ij}} 2\varphi(X_{ijk_1})) = n_{ij}Var(\hat{V}_{ij}) = \omega_{ij}^2$. By the central limit theorem of U-statistics, $H_{ijk} = \frac{2\varphi(X_{ijk})-V_{ij}}{\omega_{ij}}$ is distributed asymptotically normal with mean 0 and variance 1. Since the H_{ijk}'s are independent, then $\mathbf{H}$ follows asymptotically multivariate normal distribution with mean $\mathbf{0}$ and variance $\mathbf{I}$.

(i) If $n_{ij} = n$ for all i and j, then it is easy to verify that $\mathbf{B}_6/(\frac{n}{2} - 1)$ is a symmetric and idempotent matrix with rank $\sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} - rc = N - rc = rc(n-1)$. Therefore, $S_W^0/(\frac{n}{2} - 1)$ is asymptotically χ^2 with degrees of freedom $N - rc = rc(n-1)$. By Lemma(3.4), the sum of square within SSW is asymptotically χ^2 with degrees of freedom $rc(n-1)$.

(ii) If $n_{ij} \neq n_{i'j'}$ for some $i \neq i'$ or $j \neq j'$, $\mathbf{B}_6$ is symmetric, although not idempotent. Thus, there exists $\mathbf{H}'\mathbf{B}_6\mathbf{H} = \sum_{i=1}^{d_2} \pi_i^{(2)} z_i^2$, where $\pi_1^{(2)}, \pi_2^{(2)}, \cdots, \pi_{d_2}^{(2)}$ are the eigenvalues of $\mathbf{B}_6$, $z_i \sim N(0,1)$ and are independent. Let $c_0^{(2)} = \sum_{i=1}^{d_2} \pi_i^{(2)}/d_2$, then by [42], $S_3^0/c_0^{(2)} = \mathbf{H}'\mathbf{B}_6\mathbf{H}/c_0^{(2)} \dot{\sim} \chi^2_{d_2}$. By Lemma (3.4), the sum of square within SSW is asymptotically χ^2 with degrees of freedom d_2, where d_2 is given in Theorem (3.5).

■

Define the F-test statistics of kernel based nonparametric test for location parameters of the row effect as:

$$\begin{aligned} F_{R_l} &= \frac{MSR}{MSW} = \frac{SSR/(r-1)}{SSW/df_{w_2}} \\ &= \frac{\sum_{i=1}^{r} \hat{m}_{i.}\left(\hat{V}_{i.} - \hat{V}_{..}\right)^2/(r-1)}{\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1\neq k_2}(A_{ijk_1k_2} - \hat{V}_{ij})^2/\hat{\omega}_{ij}^2\right)/df_{w_2}}, \end{aligned} \tag{3.37}$$

where df_{w_2} is given in equation (3.30).

Theorem 3.6 *If for any $i = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, c$, $n_{ij}h_{ij}^4 \to 0$, $n_{ij}h_{ij} \to \infty$ as $\min_{i,j} n_{ij} \to \infty$, and if $\int x^2 f_{ij}^3(x)dx < \infty$, then under null hypothesis, F_{R_l} in equation (3.37) follows asymptotically F distribution with degrees of freedom $r-1$ and df_{w_2}.*

Proof: Theorem (3.3) shows that SSR follows asymptotically χ^2 with degrees of freedom $r-1$ under null hypothesis and asymptotically non-central $\chi^2(r-1)$ under the alternative. Furthermore, Theorem (3.5) implies that SSW is asymptotically χ^2 with degrees of freedom $rc(n-1)$ for balanced data and χ^2 with degrees of freedom d_2 for unbalanced data, where d_2 is the number of eigenvalues of $\mathbf{B}_6$ in equation (3.28). In order to show $F_{R_l} = \frac{MSR}{MSW}$ follows asymptotically F distribution under null hypothesis and non-central F distribution under alternative, we just need to show SSR and SSW are asymptotically independent as $\min_{i,j} n_{ij} \to \infty$.

In Lemma (3.2), S_R^0, which converges in probability to SSR, is written as a quadratic form $S_R^0 = \mathbf{U_3}'\mathbf{B_5}\mathbf{U_3}$. Note that

$$T_{ij}^{(3)} \simeq \sqrt{n_{ij}}(\hat{V}_{ij} - V_{ij})/\omega_{ij} = \frac{1}{\sqrt{n_{ij}}}\mathbf{H}'_{ij}\mathbf{j}_{n_{ij}}. \tag{3.38}$$

Hence, S_R^0 can also be written as

$$\begin{aligned}
S_R^0 &= \left(\frac{1}{\sqrt{n_{11}}}\mathbf{H}'_{11}\mathbf{j}_{n_{11}}, \cdots, \frac{1}{\sqrt{n_{rc}}}\mathbf{H}'_{rc}\mathbf{j}_{n_{rc}}\right)\mathbf{B_5}\begin{pmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{H}'_{11}\mathbf{j}_{n_{11}} \\ \vdots \\ \frac{1}{\sqrt{n_{rc}}}\mathbf{H}'_{rc}\mathbf{j}_{n_{rc}} \end{pmatrix} \\
&= \mathbf{H}'\begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{j}'_{n_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\mathbf{j}'_{n_{12}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{rc}}}\mathbf{j}'_{n_{rc}} \end{bmatrix}' \mathbf{B}_5 \begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{j}'_{n_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\mathbf{j}'_{n_{12}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{rc}}}\mathbf{j}'_{n_{rc}} \end{bmatrix}\mathbf{H} \\
&\stackrel{def}{=} \mathbf{H}'\mathbf{B}_7\mathbf{H}.
\end{aligned} \tag{3.39}$$

Recall from theorem (3.5) that $S_W^0 = \mathbf{H}'\mathbf{B}_6\mathbf{H}$. And it is easy to check that

$$\mathbf{B}_6\mathbf{B}_7 = \mathbf{0} * \mathbf{B}_5 \begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{j}'_{n_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\mathbf{j}'_{n_{12}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{rc}}}\mathbf{j}'_{n_{rc}} \end{bmatrix} = \mathbf{0}. \tag{3.40}$$

Thus, S_R^0 and S_W^0 are independent. By Lemma (3.2) and Lemma (3.4), SSR and SSW are asymptotically independent under null hypothesis $\bar{V}_{i.} = \bar{V}_{..}$ for all i. Hence, under null hypothesis, $F_{R_l} = \frac{MSR}{MSW}$ in equation (3.37) follows asymptotically F distribution with degrees of freedom $r-1$ and $rc(n-1)$ for balanced data and $F(r-1, d_2)$ for unbalanced data.

■

Similarly, to test the column effect, i.e. $H_0 : \bar{V}_{.j} = \bar{V}_{..}$ for all j, define the Column Sum of Squares (SSC) as

$$SSC = \sum_{j=1}^{c} \hat{m}_{.j}\left(\hat{V}_{.j} - \hat{V}_{..}\right)^2, \tag{3.41}$$

where $\hat{V}_{.j} = \frac{\sum_i \hat{m}_{ij}\hat{V}_{ij}}{\hat{m}_{.j}}$ and $\hat{m}_{.j} = \sum_i \hat{m}_{ij}$. Then, the F-test statistics of kernel based nonparametric test for location parameters of the column effect is given by

$$\begin{aligned} F_{C_l} &= \frac{MSC}{MSW} = \frac{SSC/(c-1)}{SSW/df_{w_2}} \\ &= \frac{\sum_{j=1}^{c} \hat{m}_{.j}\left(\hat{V}_{.j} - \hat{V}_{..}\right)^2/(c-1)}{\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1\neq k_2}(A_{ijk_1k_2} - \hat{V}_{ij})^2/\hat{\omega}_{ij}^2\right)/df_{w_2}}, \end{aligned} \quad (3.42)$$

where df_{w_2} is given by equation (3.30).

Theorem 3.7 *If for any $i = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, c$, $n_{ij}h_{ij}^4 \to 0$, $n_{ij}h_{ij} \to \infty$ as $\min_{i,j} n_{ij} \to \infty$, and if $\int x^2 f_{ij}^3(x)dx < \infty$, then under null hypothesis, F_{C_l} in equation(3.42) follows asymptotically F distribution with degrees of freedom $c-1$ and df_{w_2}.*

Proof: Similar to Theorem (3.6).

■

3.1.2 Kernel Based Nonparametric Test for Interactions of Row and Column Effects

In order to test the interaction effects, i.e. $H_0 : V_{ij} - \bar{V}_{i.} - \bar{V}_{.j} + \bar{V}_{..} = 0$ for all i and j, define the Interaction Sum of Squares (SSI) as

$$SSI = \sum_{i=1}^{r}\sum_{j=1}^{c} \hat{m}_{ij}\left(\hat{V}_{ij} - \hat{V}_{i.} - \hat{V}_{.j} + \hat{V}_{..}\right)^2. \quad (3.43)$$

To obtain the asymptotic distribution of SSI, another auxiliary variable is defined as follows:

$$S_I^0 = \sum_{i=1}^{r}\sum_{j=1}^{c} \hat{m}_{ij}\left(\hat{V}_{ij} - V_{i.}^* - V_{.j}^* + V_{..}^*\right)^2, \quad (3.44)$$

where $V_{.j}^* = \frac{\sum_i m_{ij}\hat{V}_{ij}}{m_{.j}}$. Note that $V_{i.}^*$ and $V_{..}^*$ are defined in Section 3.1.1.

Lemma 3.8 *Let* $N = \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}$*. If* $\lambda_{ij} = \lim_{\min n_{ij}\to\infty} \frac{n_{ij}}{N}$ *and* $\hat{\omega}_{ij}^2 \xrightarrow{p} \omega_{ij}^2$*, then* $S_I^0 - SSI \xrightarrow{p} 0$*, as* $\min_{i,j} n_{i,j} \to \infty$*.*

Proof: This can be proved directly by applying Slutsky Theorem [10].

■

Theorem 3.9 *Under the null hypothesis,* $H_0 : V_{ij} - \bar{V}_{i.} - \bar{V}_{.j} + \bar{V}_{..} = 0$ *for all* i *and* j*, if for any* $i = 1, 2, \cdots, r$ *and* $j = 1, 2, \cdots, c$*,* $n_{ij}h_{ij}^4 \to 0$*,* $n_{ij}h_{ij} \to \infty$ *as* $\min_{i,j} n_{ij} \to \infty$*, and if* $\int x^2 f_{ij}^3(x)dx < \infty$*, then* SSI *is asymptotically* $\chi^2\big((r-1)(c-1)\big)$*.*

Proof: Set $T_{ij}^{(3)} = \sqrt{N\lambda_{ij}}(\hat{V}_{ij} - V_{ij})/\omega_{ij}$. Then $T_{ij}^{(3)} \overset{a}{\sim} N(0,1)$ as $N \to \infty$ by Lemma (3.1). Note that, under the null hypothesis,

$$
\begin{aligned}
S_I^0 &= \sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}\Big(\hat{V}_{ij} - V_{i.}^* - V_{.j}^* + V_{..}^*\Big)^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}\Big((\hat{V}_{ij} - V_{ij}) - (V_{i.}^* - \bar{V}_{i.}) - (V_{.j}^* - \bar{V}_{.j}) + (V_{..}^* - \bar{V}_{..})\Big)^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}\Big((\hat{V}_{ij} - V_{ij})^2 + (V_{i.}^* - \bar{V}_{i.})^2 + (V_{.j}^* - \bar{V}_{.j})^2 + (V_{..}^* - \bar{V}_{..})^2 \\
&- 2(\hat{V}_{ij} - V_{ij})(V_{i.}^* - \bar{V}_{i.}) - 2(\hat{V}_{ij} - V_{ij})(V_{.j}^* - \bar{V}_{.j}) + 2(\hat{V}_{ij} - V_{ij})(V_{..}^* - \bar{V}_{..}) \\
&+ 2(V_{i.}^* - \bar{V}_{i.})(V_{.j}^* - \bar{V}_{.j}) - 2(V_{i.}^* - \bar{V}_{i.})(V_{..}^* - \bar{V}_{..}) - 2(V_{.j}^* - \bar{V}_{.j})(V_{..}^* - \bar{V}_{..})\Big) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(\hat{V}_{ij} - V_{ij})^2 + \sum_{i=1}^{r} m_{i.}(V_{i.}^* - \bar{V}_{i.})^2 + \sum_{j=1}^{c} m_{.j}(V_{.j}^* - \bar{V}_{.j})^2 + m_{..}(V_{..}^* - \bar{V}_{..})^2 \\
&- 2\sum_{i=1}^{r} m_{i.}(V_{i.}^* - \bar{V}_{i.})^2 - 2\sum_{j=1}^{c} m_{.j}(V_{.j}^* - \bar{V}_{.j})^2 + 2m_{..}(V_{..}^* - \bar{V}_{..})^2 \\
&+ 2\sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(V_{i.}^* - \bar{V}_{i.})(V_{.j}^* - \bar{V}_{.j}) - 2m_{..}(V_{..}^* - \bar{V}_{..})^2 - 2m_{..}(V_{..}^* - \bar{V}_{..})^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(\hat{V}_{ij} - V_{ij})^2 - \sum_{i=1}^{r} m_{i.}(V_{i.}^* - \bar{V}_{i.})^2 - \sum_{j=1}^{c} m_{.j}(V_{.j}^* - \bar{V}_{.j})^2 + m_{..}(V_{..}^* - \bar{V}_{..})^2 \\
&+ \sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(V_{i.}^* - \bar{V}_{i.})(V_{.j}^* - \bar{V}_{.j}) - 2m_{..}(V_{..}^* - \bar{V}_{..})^2. \qquad (3.45)
\end{aligned}
$$

If we assume $m_{ij} = \frac{(m_{i.})(m_{.j})}{m_{..}}$, then

$$
\sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(V_{i.}^* - \bar{V}_{i.})(V_{.j}^* - \bar{V}_{.j}) = 2m_{..}(V_{..}^* - \bar{V}_{..})^2.
$$

Hence,

$$
\begin{aligned}
S_I^0 &= \sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(\hat{V}_{ij}-V_{ij})^2 - \sum_{i=1}^{r} m_{i.}(V_{i.}^{*}-\bar{V}_{i.})^2 - \sum_{j=1}^{c} m_{.j}(V_{.j}^{*}-\bar{V}_{.j})^2 + m_{..}(V_{..}^{*}-\bar{V}_{..})^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(\hat{V}_{ij}-V_{ij})^2 - \sum_{i=1}^{r}\frac{1}{m_{i.}}\Big(\sum_{j=1}^{c} m_{ij}(\hat{V}_{ij}-V_{ij})\Big)^2 - \sum_{j=1}^{c}\frac{1}{m_{.j}}\Big(\sum_{i=1}^{r} m_{ij}(\hat{V}_{ij}-V_{ij})\Big)^2 \\
&+ \frac{1}{m_{..}}\Big(\sum_{i=1}^{r}\sum_{j=1}^{c} m_{ij}(\hat{V}_{ij}-V_{ij})\Big)^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{N\lambda_{ij}(\hat{V}_{ij}-V_{ij})^2}{\omega_{ij}^2} \qquad (3.46) \\
&- \sum_{i=1}^{r}\sum_{j_1=1}^{c}\sum_{j_2=1}^{c} \frac{N}{m_{i.}}\frac{\sqrt{N\lambda_{ij_1}}}{\omega_{ij_1}}(\hat{V}_{ij_1}-V_{ij_1})\frac{\sqrt{N\lambda_{ij_2}}}{\omega_{ij_2}}(\hat{V}_{ij_2}-V_{ij_2})\frac{\sqrt{\lambda_{ij_1}}}{\omega_{ij_1}}\frac{\sqrt{\lambda_{ij_2}}}{\omega_{ij_2}} \\
&- \sum_{i_1=1}^{r}\sum_{i_2=1}^{r}\sum_{j=1}^{c} \frac{N}{m_{.j}}\frac{\sqrt{N\lambda_{i_1j}}}{\omega_{i_1j}}(\hat{V}_{i_1j}-V_{i_1j})\frac{\sqrt{N\lambda_{i_2j}}}{\omega_{i_2j}}(\hat{V}_{i_2j}-V_{i_2j})\frac{\sqrt{\lambda_{i_1j}}}{\omega_{i_1j}}\frac{\sqrt{\lambda_{i_2j}}}{\omega_{i_2j}} \\
&+ \frac{N}{m_{..}}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}\sum_{j_1=1}^{c}\sum_{j_2=1}^{c} \frac{\sqrt{N\lambda_{i_1j_1}}}{\omega_{i_1j_1}}(\hat{V}_{i_1j_1}-V_{i_1j_1})\frac{\sqrt{N\lambda_{i_2j_2}}}{\omega_{i_2j_2}}(\hat{V}_{i_2j_2}-V_{i_2j_2})\frac{\sqrt{\lambda_{i_1j_1}}}{\omega_{i_1j_1}}\frac{\sqrt{\lambda_{i_2j_2}}}{\omega_{i_2j_2}}.
\end{aligned}
$$

Let $\mathbf{U_3} = (T_{11}^{(3)}, T_{12}^{(3)}, \cdots, T_{rc}^{(3)})$, then the second term of (3.46) can written as a quadratic form $\mathbf{U_3}'\mathbf{M^{(1)}}\mathbf{U_3}$, the third term of (3.46) as $\mathbf{U_3}'\mathbf{N^{(1)}}\mathbf{U_3}$ and the fourth term of (3.46) as $\mathbf{U_3}'\mathbf{M^{(2)}}\mathbf{U_3}$, where $\mathbf{M^{(1)}}$ and $\mathbf{M^{(2)}}$ are defined in equation (3.20) and (3.21) respectively,

$$
\mathbf{N^{(1)}} = \begin{bmatrix} \mathbf{N_{11}} & \mathbf{N_{12}} & \cdots & \mathbf{N_{1r}} \\ \mathbf{N_{21}} & \mathbf{N_{22}} & \cdots & \mathbf{N_{2r}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{N_{r1}} & \mathbf{N_{r2}} & \cdots & \mathbf{N_{rr}} \end{bmatrix}, \qquad (3.47)
$$

and

$$
\mathbf{N_{ij}} = \begin{bmatrix} \frac{N}{m_{.1}}\frac{\sqrt{\lambda_{i1}}\sqrt{\lambda_{j1}}}{\omega_{i1}\omega_{j1}} & 0 & \cdots & 0 \\ 0 & \frac{N}{m_{.2}}\frac{\sqrt{\lambda_{i2}}\sqrt{\lambda_{j2}}}{\omega_{i2}\omega_{j2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{N}{m_{.c}}\frac{\sqrt{\lambda_{ic}}\sqrt{\lambda_{jc}}}{\omega_{ic}\omega_{jc}} \end{bmatrix} \qquad (3.48)
$$

for all $i = 1, 2, \cdots, r,\ j = 1, 2, \cdots, c$. Thus, S_I^0 can be rewritten by the following quadratic form:

$$\begin{aligned} S_I^0 &= \mathbf{U_3}'\mathbf{U_3} - \mathbf{U_3}'\mathbf{M^{(1)}}\mathbf{U_3} - \mathbf{U_3}'\mathbf{N^{(1)}}\mathbf{U_3} + \mathbf{U_3}'\mathbf{M^{(2)}}\mathbf{U_3} \\ &= \mathbf{U_3}'(\mathbf{I_{rc}} - \mathbf{M^{(1)}} - \mathbf{N^{(1)}} + \mathbf{M^{(2)}})\mathbf{U_3} \\ &= \mathbf{U_3}'\mathbf{B_8}\mathbf{U_3}, \end{aligned} \tag{3.49}$$

where $\mathbf{B_8}$ is given by

$$\mathbf{B_8} = \begin{bmatrix} \mathbf{I_c} - (\frac{N}{m_{1.}} + \frac{N}{m_{..}})\mathbf{M_{11}} - \mathbf{N_{11}} & -\mathbf{N_{12}} + \frac{N}{m_{..}}\mathbf{M_{12}} & \cdots & -\mathbf{N_{1r}} + \frac{N}{m_{..}}\mathbf{M_{1r}} \\ -\mathbf{N_{12}} + \frac{N}{m_{..}}\mathbf{M_{12}} & \mathbf{I_c} - (\frac{N}{m_{2.}} + \frac{N}{m_{..}})\mathbf{M_{22}} - \mathbf{N_{22}} & \cdots & -\mathbf{N_{2r}} + \frac{N}{m_{..}}\mathbf{M_{2r}} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{N_{1r}} + \frac{N}{m_{..}}\mathbf{M_{1r}} & -\mathbf{N_{2r}} + \frac{N}{m_{..}}\mathbf{M_{2r}} & \cdots & \mathbf{I_c} - (\frac{N}{m_{r.}} + \frac{N}{m_{..}})\mathbf{M_{rr}} - \mathbf{N_{rr}} \end{bmatrix} \tag{3.50}$$

It can easily be shown that $\mathbf{B_8}$ is symmetric and idempotent. Thus, we obtain that

$$\begin{aligned} rank(\mathbf{B_8}) &= trace(\mathbf{B_8}) \\ &= rc - \sum_{i=1}^{r} \frac{1}{m_{i.}} \sum_{j=1}^{c} m_{ij} - \sum_{j=1}^{c} \frac{1}{m_{.j}} \sum_{i=1}^{c} m_{ij} + \frac{1}{m_{..}} \sum_{i=1}^{r} \sum_{j=1}^{c} m_{ij} \\ &= rc - r - c + 1 \\ &= (r-1)(c-1). \end{aligned} \tag{3.51}$$

$\mathbf{U_3}$ follows approximately multivariate normal with mean $\mathbf{0}$ and variance $\mathbf{I}$, since $T_{ij}^{(3)}$'s independently follow univariate standard normal distribution. Therefore, S_I^0 is asymptotically χ^2 with degrees of freedom $(r-1)(c-1)$ under H_0. By Lemma (3.8), SSI is asymptotically χ^2 with degrees of freedom $(r-1)(c-1)$ under H_0.

■

Define the F-test statistics of kernel based nonparametric test for location parameters of interaction effects as:

$$\begin{aligned} F_{I_l} &= \frac{MSI}{MSW} = \frac{SSI/(r-1)(c-1)}{SSW/df_{w_2}} \\ &= \frac{\sum_{i=1}^{r}\sum_{j=1}^{c}\hat{m}_{ij}\left(\hat{V}_{ij} - V_{i.}^{*} - V_{.j}^{*} + V_{..}^{*}\right)^2/(r-1)(c-1)}{\left(\sum_{i=1}^{r}\sum_{j=1}^{c}\sum\sum_{k_1\neq k_2}(A_{ijk_1k_2} - \hat{V}_{ij})^2/\hat{\omega}_{ij}^2\right)/df_{w_2}}, \end{aligned} \tag{3.52}$$

where df_{w_2} is given by equation (3.30).

Theorem 3.10 *If for any $i = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, c$, $n_{ij}h \to \infty$ as $\min_{i,j} n_{ij} \to \infty$, $\int uf^2(u)du < \infty$ and $\int u^2 f(u)du < \infty$, then under null hypothesis, F_{I_l} in equation (3.52) follows asymptotically F distribution with degrees of freedom $(r-1)(c-1)$ and df_{w_2}.*

Proof: Theorem (3.9) shows that SSI follows asymptotically χ^2 with degrees of freedom $(r-1)(c-1)$ under null hypothesis and asymptotically non-central $\chi^2((r-1)(c-1))$ under the alternative. Furthermore, Theorem (3.5) implies that SSW is asymptotically χ^2 with degrees of freedom $rc(n-1)$ for balanced data and χ^2 with degrees of freedom d_r for unbalanced data, where d_r is the number of eigenvalues of $\mathbf{B}_6$ in equation (3.29). In order to show $F_{I_l} = \frac{MSI}{MSW}$ follows asymptotically F distribution under null hypothesis and non-central F distribution under alternative, we just need to show SSI and SSW are asymptotically independent as $\min_{i,j} n_{ij} \to \infty$.

In Lemma (3.8), S_I^0, which converges in probability to SSI, is written as a quadratic form $S_I^0 = \mathbf{U_3}'\mathbf{B_8}\mathbf{U_3}$. Note that

$$T_{ij}^{(3)} \simeq \sqrt{n_{ij}}(\hat{V}_{ij} - V_{ij})/\omega_{ij} = \frac{1}{\sqrt{n_{ij}}}\mathbf{H}'_{ij}\mathbf{j}_{n_{ij}}. \tag{3.53}$$

Hence, S_I^0 can also be written as

$$\begin{aligned}
S_I^0 &= (\frac{1}{\sqrt{n_{11}}}\mathbf{H}'_{11}\mathbf{j}_{n_{11}}, \cdots, \frac{1}{\sqrt{n_{rc}}}\mathbf{H}'_{rc}\mathbf{j}_{n_{rc}})\mathbf{B_8}\begin{pmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{H}'_{11}\mathbf{j}_{n_{11}} \\ \vdots \\ \frac{1}{\sqrt{n_{rc}}}\mathbf{H}'_{rc}\mathbf{j}_{n_{rc}} \end{pmatrix} \\
&= \mathbf{H}'\begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{j}'_{n_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\mathbf{j}'_{n_{12}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{rc}}}\mathbf{j}'_{n_{rc}} \end{bmatrix}' \mathbf{B}_8 \begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{j}'_{n_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\mathbf{j}'_{n_{12}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{rc}}}\mathbf{j}'_{n_{rc}} \end{bmatrix} \mathbf{H} \\
&\stackrel{def}{=} \mathbf{H}'\mathbf{B}_9\mathbf{H}. \qquad (3.54)
\end{aligned}$$

Recall from theorem (3.5) that $S_W^0 = \mathbf{H}'\mathbf{B}_6\mathbf{H}$. And it is easy to check that

$$\mathbf{B}_6\mathbf{B}_9 = \mathbf{0} * \mathbf{B}_8 \begin{bmatrix} \frac{1}{\sqrt{n_{11}}}\mathbf{j}'_{n_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{n_{12}}}\mathbf{j}'_{n_{12}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{n_{rc}}}\mathbf{j}'_{n_{rc}} \end{bmatrix} = \mathbf{0}. \qquad (3.55)$$

Thus, S_I^0 and S_W^0 are independent. By Lemma (3.8) and Lemma (3.4), SSI and SSW are asymptotically independent under null hypothesis $H_0 : V_{ij} - \bar{V}_{i.} - \bar{V}_{.j} + \bar{V}_{..} = 0$ for all i and j. Hence, under null hypothesis, $F_{I_l} = \frac{MSI}{MSW}$ in equation (3.52) follows asymptotically F distribution with degrees of freedom $(r-1)(c-1)$ and $rc(n-1)$ for balanced data and $F((r-1)(c-1), d_2)$ for unbalanced data. Under the alternative, $F_{I_l} = \frac{SSI}{SSW}$ follows asymptotically non-central $F((r-1)(c-1), rc(n-1))$ for balanced data, and non-central $F((r-1)(c-1), d_2)$ for unbalanced data,

■

3.2 Simulation Study for Evaluating the Power of Kernel-based Nonparametric Two-way ANOVA

In this section, powers of the kernel-based nonparametric two-way ANOVA tests, including the test for interactions and main effects, are evaluated through simulation. To better demonstrate the properties of kernel-based nonparametric ANOVA tests compared with the traditional parametric ANOVA tests, the performances of both tests for interaction and row effects based on data from the three distributions described in Table 2.1 are studied.

3.2.1 Simulation Study of the Test for Interaction

The objective of this section is to study the Type I error rate and power of the test of interaction proposed in section 3.1.2 and to compare it with the parametric two-way ANOVA test in 3 distinctive cases: Normal, Cauchy and Lognormal, given in Table 2.1. Consider an experiment with two treatments, Factor A (Row) and Factor B (Column). Each factor has three levels, i.e. $r = 3$ and $c = 3$ in Section 3.1. Consider two-way layout $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ in Section 3.1. Let the overall location $\mu = 3$, row effect $\alpha = (\alpha_1, \alpha_2, \alpha_3) = (-1, 0, 1)$ and column effect $\beta = (\beta_1, \beta_2, \beta_3) = (-1, 0, 1)$.

To obtain the actual Type I error rate when the significant level is set to be 0.05, we follow the steps below:

(1) Randomly generate 9 groups of data (considered as the observations of the response variable) with balanced sample size n from distributions listed in Table 3.1 for Case I, II and III. Note that the location and scale parameters in Table 3.1 are determined by letting the interaction $\gamma = \mathbf{0}$, in addition to the μ, α and β described above for all the three cases.

(2) Apply the parametric ANOVA test for interaction and the kernel based nonparametric ANOVA test for interaction separately. Record the test result as 1

or 0. 1 means "reject the null hypothesis" and 0 means fail to reject.

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

(4) Repeat (1)-(3) for sample size $n = 15, 20, 25, 30, 35, 40, 45, 50, 55, 60$.

Table 3.1: Evaluate the Type I Error Rate of Tests for Interaction in 3 Cases

			Factor B		
			Level 1	Level 2	Level 3
Case I^c	**Factor A**	Level 1	N(1,1)	N(2,1)	N(3,1)
		Level 2	N(2,1)	N(3,1)	N(4,1)
		Level 3	N(3,1)	N(4,1)	N(5,1)
Case II^c		Level 1	Cauchy(1,1)	Cauchy(2,1)	Cauchy(3,1)
		Level 2	Cauchy(2,1)	Cauchy(3,1)	Cauchy(4,1)
		Level 3	Cauchy(3,1)	Cauchy(4,1)	Cauchy(5,1)
Case III^c		Level 1	LN(1,0.5,1)	LN(2,0.5,1)	LN(3,0.5,1)
		Level 2	LN(2,0.5,1)	LN(3,0.5,1)	LN(4,0.5,1)
		Level 3	LN(3,0.5,1)	LN(4,0.5,1)	LN(5,0.5,1)

The procedure of calculating the empirical power is very similar except letting the interaction

$$\gamma_{\mathbf{N}} = \begin{bmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.56}$$

for Case I^c,

$$\gamma_{\mathbf{C}} = \begin{bmatrix} -1.5 & -1.5 & 3 \\ 0 & 1.5 & -1.5 \\ 1.5 & 0 & -1.5 \end{bmatrix} \tag{3.57}$$

for Case II^c, and

$$\gamma_{\mathbf{L}} = \begin{bmatrix} -1 & 0.5 & 0.5 \\ 1 & 0 & -1 \\ 0 & -0.5 & 0.5 \end{bmatrix} \tag{3.58}$$

for Case III^c. Thus, for each case, we randomly generate 9 groups of data with balanced sample size n from distributions listed in Table 3.2, rather than Table 3.1, in Step (1).

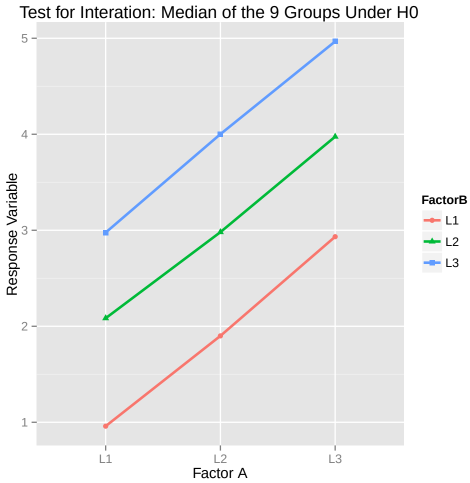
Table 3.2: Evaluate the Power of Tests for Interaction in 3 Cases

			Factor B		
			Level 1	Level 2	Level 3
Case I^c	**Factor A**	Level 1	N(1.5,1)	N(1.5,1)	N(3,1)
		Level 2	N(1.5,1)	N(3.5,1)	N(4,1)
		Level 3	N(3,1)	N(4,1)	N(5,1)
Case II^c		Level 1	Cauchy(-0.5,1)	Cauchy(0.5,1)	Cauchy(6,1)
		Level 2	Cauchy(2,1)	Cauchy(4.5,1)	Cauchy(2.5,1)
		Level 3	Cauchy(4.5,1)	Cauchy(4,1)	Cauchy(3.5,1)
Case III^c		Level 1	LN(0,0.5,1)	LN(2.5,0.5,1)	LN(3.5,0.5,1)
		Level 2	LN(3,0.5,1)	LN(3,0.5,1)	LN(3,0.5,1)
		Level 3	LN(3,0.5,1)	LN(3.5,0.5,1)	LN(5.5,0.5,1)

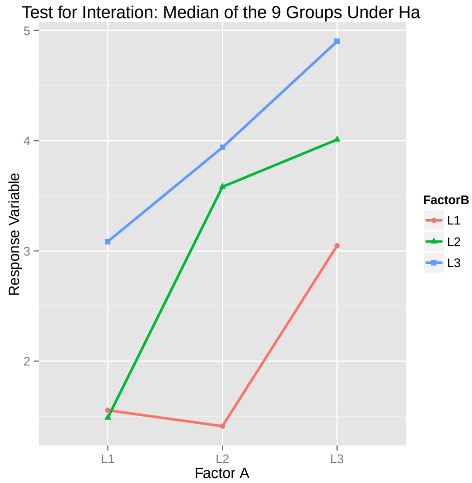
As illustrated in previous paragraphs, Table 3.1 shows the distributions of the 9 cells under the null hypothesis, i.e. no interaction. For instance, in Case I^c, the distribution of the response variable in cell $(1,1)$, when Factor A is set at the first level and Factor B is set at the first level as well, is Normal distribution with mean 1 and standard deviation 1, denoted as $N(1,1)$. In Case II^c, the distribution of the response variable in cell $(2,3)$, when Factor A is set at the second level and Factor B is set at the third level, is Cauchy distribution with location parameter 4 and scale parameter 1, denoted as $Cauchy(4, 1)$. In Case III^c, the distribution of the response variable in cell $(3, 1)$, when Factor A is set at the third level and Factor B is set at the first level, is three-parameter Lognormal distribution with location parameter 3, scale parameter 0.5, and shape parameter 1, denoted as $LN(3, 0.5, 1)$.

Table 3.2 shows the distributions of the 9 cells under the alternative hypothesis, i.e. there exist interactions which are given in equation (3.56), (3.57) and (3.58) for Case I^c, Case II^c and Case III^c respectively. For instance, in Case I^c, the distribution of the response variable in cell $(1, 1)$, when Factor A is set at the first level and Factor B is set at the first level as well, is Normal distribution with mean 1.5 and standard deviation 1, denoted as $N(1.5, 1)$. In Case II^c, the distribution of the response variable in cell $(2, 3)$, when Factor A is set at the second level and Factor B is set at the third level, is Cauchy distribution with location parameter 2.5 and scale parameter 1, denoted as $Cauchy(2.5, 1)$. In Case III^c, the distribution of the response variable in cell $(3, 1)$, when Factor A is set at the third level and Factor B is set at the first level, is three-parameter Lognormal distribution with location parameter 3, scale parameter 0.5, and shape parameter 1, denoted as $LN(3, 0.5, 1)$.

To clearly illustrate the interactions in locations of the distributions under the null hypothesis (in Table 3.1) and alternative hypothesis (in Table 3.2), Figure 3.1-Figure 3.3 plot the median of the distribution in the 9 cells for each case in Table 3.1 and Table 3.2. Figure 3.1(a) shows the median of the 9 cells in the Case I^c of Table 3.1. The horizontal axis represents the levels of factor A, and the vertical axis represents the median of the response variable generated from the distribution in Case I^c of Table 3.1. The colored lines represent the levels of factor B. The red line represents the level 1 of factor B. The green line represents the level 2 of factor B. And the blue line represents the level 3 of factor B. The three lines are parallel, which indicates no interaction between factor A and factor B. Figure 3.1(b) shows the median of the 9 cells in the Case I^c of Table 3.2. The three lines are obviously not parallel, which

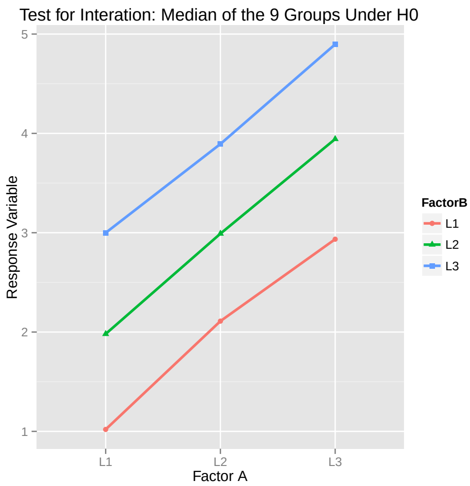


(a) Under H_0: $\gamma = \mathbf{0}$

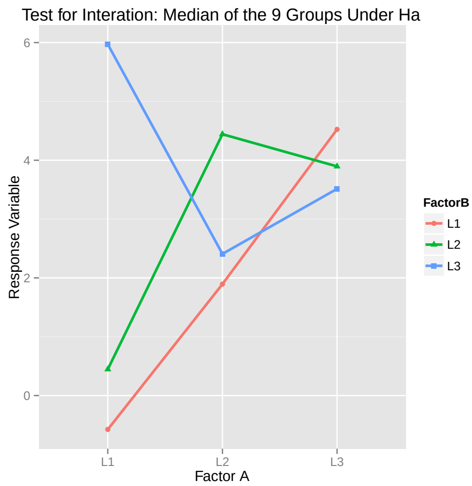


(b) Under H_a: $\gamma = \gamma_{\mathbf{N}}$

Figure 3.1: Median of the 9 groups in Case I^c: Normal Distributions

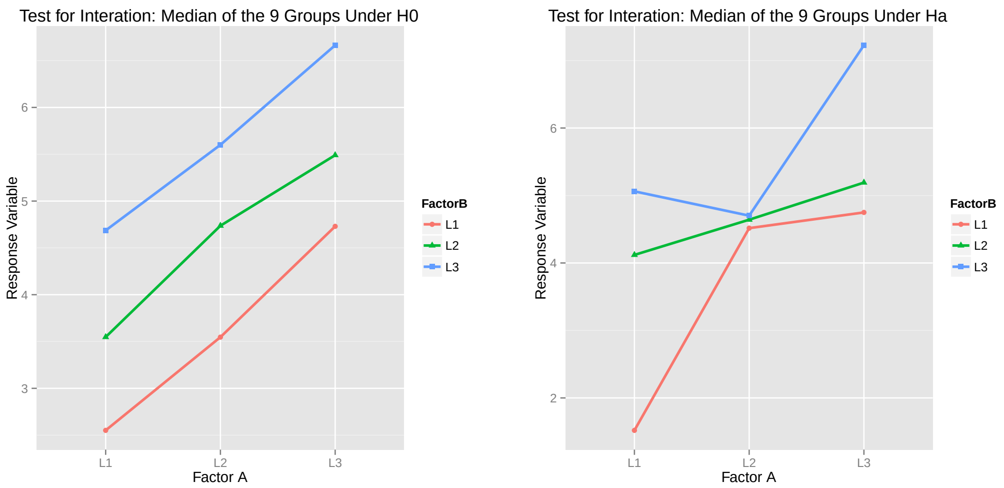


(a) Under H_0: $\gamma = \mathbf{0}$



(b) Under H_a: $\gamma = \gamma_{\mathbf{C}}$

Figure 3.2: Median of the 9 groups in Case II^c: Cauchy Distributions



(a) Under H_0: $\gamma = \mathbf{0}$ (b) Under H_a: $\gamma = \gamma_{\mathbf{L}}$

Figure 3.3: Median of the 9 groups in Case III^c: Lognormal Distributions

indicates that there exists interaction between the two factors in Case I^c of Table 3.2. Figure 3.2(a) and figure 3.3(a) show the median of the 9 cells in the Case II^c and Case III^c of Table 3.1. It is not hard to tell that three lines in Figure 3.2(a) and figure 3.3(a) are almost parallel, which confirms that there is no interaction between factor A and factor B in the Case II^c and Case III^c of Table 3.1. Figure 3.2(b) and figure 3.3(b) show the median of the 9 cells in the Case II^c and Case III^c of Table 3.2 respectively. The unparallelled lines in Figure 3.2(b) and Figure 3.3(b) verify that there exist interactions between factor A and factor B in the Case II^c and Case III^c of Table 3.2.

The simulation results for the test of interaction via the kernel based nonparametric two-way ANOVA test and parametric two-way ANOVA test are given in Table 3.3-Table 3.5.

Table 3.3 lists the actual Type I error rates and the empirical powers for the test of interaction in Case I^c: Normal Case. It is shown that the actual Type I error rates for both parametric and nonparametric test of interaction are around 0.05, the

Table 3.3: Power for Test of Interactions: Case I^c (Normal Distribution)

Sample Size	Parametric Test		Nonparametric Test	
	Power	Type I	Power	Type I
15	0.8761	0.0489	0.2370	0.0462
20	0.9587	0.0489	0.3460	0.0548
25	0.9892	0.0496	0.3940	0.0463
30	0.9975	0.0532	0.4443	0.0519
35	0.9989	0.0495	0.5038	0.0464
40	0.9999	0.0556	0.5384	0.0454
45	0.9999	0.0507	0.5792	0.0476
50	1	0.0483	0.6133	0.0510
55	1	0.0473	0.6406	0.0427
60	1	0.0504	0.6702	0.0469

[1] * Note: Given significance level: $\alpha = 0.05$.

[2] * Note: The Type I in the table means Actual Type I Error.

significance level. As we expected, in Normal Case, parametric F test performs better than the kernel based nonparametric test of interaction. Moreover, the kernel based nonparametric two-way ANOVA test is less powerful than nonparametric one-way ANOVA. Although the power of the nonparametric test increases as the sample size grows, the power of the nonparametric test of interaction is only 67% even when the sample size is up to 60.

Table 3.4: Power for Test of Interactions: Case II^c (Cauchy Distribution)

Sample Size	**Parametric Test**		**Nonparametric Test**	
	Power	**Type I**	**Power**	**Type I**
15	0.2490	0.0183	0.6056	0.0133
20	0.2475	0.0162	0.7781	0.0139
25	0.2493	0.0164	0.9591	0.0173
30	0.2498	0.0161	0.9863	0.0177
35	0.2595	0.0156	0.9923	0.0267
40	0.2588	0.0180	0.9987	0.0265
45	0.2592	0.0165	0.9995	0.0331
50	0.2582	0.0140	0.9999	0.0394
55	0.2559	0.0156	0.9998	0.0409
60	0.2527	0.0143	1	0.0501

[1] $*$ Note: Given significance level: $\alpha = 0.05$.

[2] $*$ Note: The Type I in the table means Actual Type I Error.

Table 3.4 lists the actual Type I error rates and the empirical powers for the test of interaction in Case II^c: Cauchy Case. It is shown that the actual Type I error rates of the parametric test are around 0.01 when the significance level is 0.05, which infers that the parametric test of interaction is conservative in Case II^c. Whereas, the actual Type I error rates of our nonparametric test are around 0.01 when the cell sample size is small, and get closer and closer to 0.05 as the sample size grows. As we expected, in the Cauchy Case, the kernel based nonparametric test performs way better than the parametric F test of interaction. The power of nonparametric test of interaction is 60.56% when the cell sample size is only 15. Moreover, the power of nonparametric test quickly increases to 1 as the sample size rises. The power of parametric test of interaction is only 24.9% when the sample size is 15 and does not increase as the sample size grows.

Table 3.5: Power for Test of Interactions: Case III^c (Lognormal Distribution)

Sample Size	**Parametric Test**		**Nonparametric Test**	
	Power	**Type I**	**Power**	**Type I**
15	0.4526	0.0431	0.2854	0.0169
20	0.5683	0.0413	0.5114	0.0250
25	0.6511	0.0436	0.6238	0.0331
30	0.7226	0.0419	0.7317	0.0461
35	0.7783	0.0435	0.8481	0.0509
40	0.8306	0.0438	0.9517	0.0518
45	0.8699	0.0461	0.9549	0.0553
50	0.8974	0.0442	0.9617	0.0525
55	0.9196	0.0473	0.9630	0.0518
60	0.9335	0.0443	0.9672	0.0528

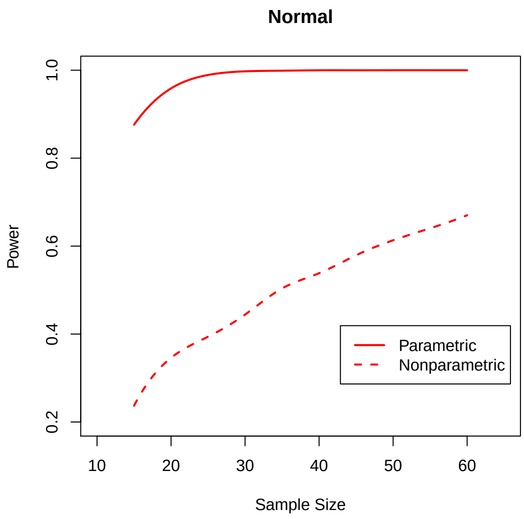
[1] $*$ Note: Given significance level: $\alpha = 0.05$.

[2] $*$ Note: The Type I in the table means Actual Type I Error.

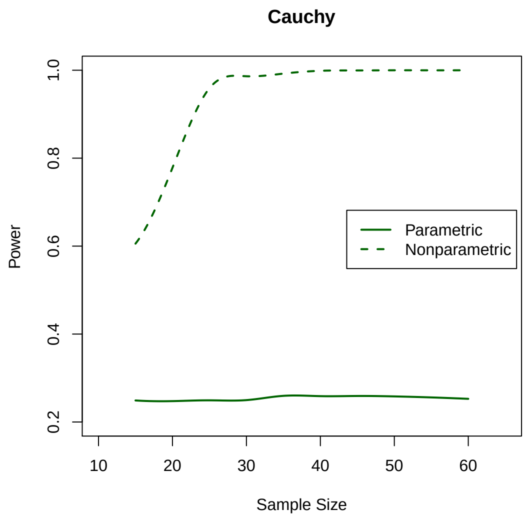
Table 3.5 lists the actual Type I error rates and the empirical powers for the test of interaction in Case III^c: Lognormal Case. It is shown that the actual Type I error rates for the parametric test are between 0.04 and 0.05 when the significance level is 0.05, which infers that the parametric test of interaction is little bit conservative in Case III^c. Whereas, the actual Type I error rates of our nonparametric test are around 0.02 when the cell sample size is small, and get closer and closer to 0.05 as the sample size grows. When the sample size is around 30-35, the actual Type I error rates of the kernel based nonparametric test of interaction is around 0.05, the significance level. Since the nonparametric test is more conservative than the parametric test when the sample size is small, the parametric test has higher power than the nonparametric test. However, when the sample size is 30 or above, the nonparametric test is more powerful than the parametric test as we expected. When the sample size is 40, the power of the parametric test is only 83.6%, while the power of the nonparametric test is 95.17%.

Figure 3.4 demonstrates the power of the parametric and nonparametric test of interaction with respect to sample size when the underlying distributions of the samples are Normal, Cauchy and Lognormal respectively. In Figure 3.4, the solid line represents the power of the parametric test of interaction, while the dashed line represents the power of our kernel based nonparametric test of interaction in Section 3.1.2. The red line (solid and dashed) represents the power of test in Case I^c when the underlying distribution is Normal distribution. The green line (solid and dashed) represents the power of test in Case II^c when the underlying distribution is Cauchy distribution. And the blue line (solid and dashed) represents the power of test in Case III^c when the underlying distribution is Lognormal distribution.

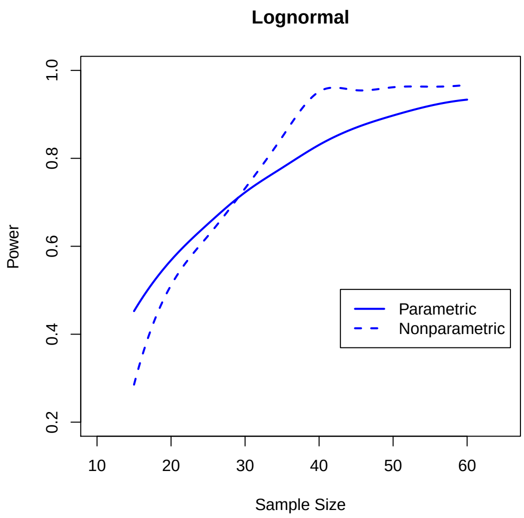
Figure 3.4(a) compares the power of parametric and the kernel based nonparametric test of interaction in Case I^c. It shows that the power of the parametric test increases sharply to around 95% when sample size goes from 15 to 20 in Case I^c.



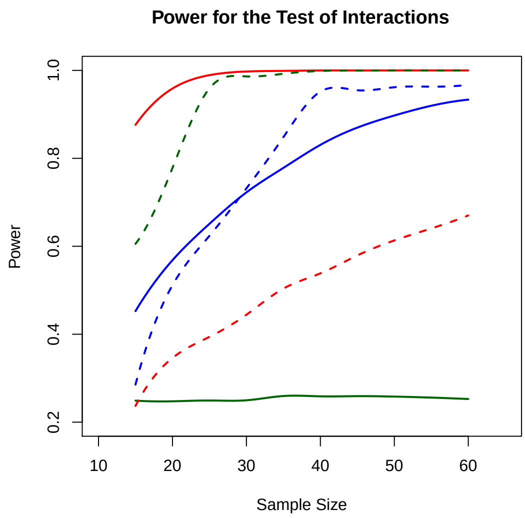
(a) Case I^c: Normal



(b) Case II^c: Cauchy



(c) Case III^c: Lognormal



(d) Case I^c,II^c, III^c

Figure 3.4: (a) Power of the parametric and nonparametric test of interaction on the 9 cells in Case I^c; (b) Power of the parametric and nonparametric test of interaction on the 9 cells in Case II^c; (c) Power of the parametric and nonparametric test of interaction on the 9 cells in Case III^c; (d) Power of the parametric and nonparametric test of interaction on the 9 cells in Case I^c, II^c and III^c.

Compare to the parametric test of interaction, the new nonparametric test is much less powerful than the traditional parametric test, which is consistent with what we concluded from Table 3.3. Fortunately, the new nonparametric test increases its power as the sample size grows. Although in Figure 3.4(a), it is shown that there is still a gap in power between the parametric and our nonparametric test, the gap tends to reduce as the sample size go beyond 60. Figure 3.4(b) compares the power of parametric and the kernel based nonparametric test of interaction in Case II^c. It is shown in Figure 3.4(b) that the power of our nonparametric ANOVA test grows as the sample size increases, while the power of parametric ANOVA test almost keeps constant in Case II^c. Moreover, it is easy to tell from Figure 3.4(b) that the power of the new nonparametric test of interaction in Case II^c is much higher than the one of the parametric test, even when the sample size is 15. Figure 3.4(c) compares the power of parametric and nonparametric test of interaction in Case III^c. In Figure 3.4(c), it is demonstrated that the blue solid line is above the blue dashed line when sample size is less than 35, which infers that the parametric test outperforms our nonparametric test when the sample size is less than 35 in Case III^c. However, this relationship flips as the sample size goes beyond 35, which indicates that the new parametric test of interaction outperforms the traditional parametric test. Figure 3.4(d) combines Figure (3.4(a))(3.4(b))(3.4(c)) in one graph, so it is easier to compare the powers of the either test throughout the 3 cases. The relationship of three solid lines infers that the power of the parametric ANOVA test is far more severely hurt by fat tails or extreme outliers than the skewness does. If the data comes from a skewed distribution, large sample size leads to better power. However, if the data come from a fat-tailed distribution, large sample size would not make any improvement in power. The relationship of three dashed lines infers that nonparametric test of interaction for the two-way ANOVA is not as powerful as the nonparametric one-way ANOVA if the underlying distribution is normal. However, larger sample size always helps.

3.2.2 Simulation Study of the Test for Main Effect

If the test of interaction fails to reject, which means that there is no interaction detected, then main effects will be tested. The purpose of this section is to study the Type I error rate and power of the test of main effect proposed in Section 3.1.1 and to compare with the parametric two-way ANOVA test in 3 distinctive cases: Normal, Cauchy and Lognormal, given in Table 2.1. Since testing row effect is exactly the same as testing the column effect, without loss of generality, we just illustrate the simulation results of the test of row effect in this section. As in Section 3.2.1, we consider the same experiment with two treatments, Factor A (Row) and Factor B (Column). Each factor has three levels, i.e. $r = 3$ and $c = 3$. In the two-way layout, we set the overall location $\mu = 3$, and the interaction $\gamma = \mathbf{0}$. The column effect is set to be $\beta_{\mathbf{N}} = (-0.25, 0, 0.25)$ for the Normal case, $\beta_{\mathbf{CL}} = (-2, 1, 1)$ for the Cauchy case and Lognormal case.

To obtain the actual Type I error rate when the significant level is set to be 0.05, we follow the steps below:

(1) Randomly generate 9 groups of data (considered as the observations of the response variable) with balanced sample size n from distributions listed in Table 3.6 for Case I, II and III. Note that the location and scale parameters in Table 3.6 are determined by letting the row effect $\alpha = (\alpha_1, \alpha_2, \alpha_3) = \mathbf{0}$, in addition to the μ, $\beta_{\mathbf{N}}$ (or $\beta_{\mathbf{CL}}$) and γ described above for all the three cases.

(2) Apply the parametric two-way ANOVA test for row effect and the kernel based nonparametric two-way ANOVA test for row effect separately. Record the test result as 1 or 0. 1 means "reject the null hypothesis" and 0 means fail to reject.

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

(4) Repeat (1)-(3) for sample size $n = 15, 20, 25, 30, 35, 40, 45, 50, 55, 60$.

The procedure of calculating the empirical power is very similar except letting the row effect $\alpha_{\mathbf{N}} = (0, -0.25, 0.25)$ for the normal case (Case I^d), $\alpha_{\mathbf{LC}} = (0.75, -1.5, 0.75)$ for the Cauchy case (Case II^d) and Lognormal case(Case III^d). Thus, for each case, we randomly generate 9 groups of data with balanced sample size n from distributions listed in Table 3.7, rather than Table 3.6, in Step (1).

Table 3.6: Evaluate the Type I Error Rate of Tests for Row Effect in 3 Cases

			Factor B		
			Level 1	Level 2	Level 3
Case I^d	**Factor A**	Level 1	N(2.75,1)	N(3,1)	N(3.25,1)
		Level 2	N(2.75,1)	N(3,1)	N(3.25,1)
		Level 3	N(2.75,1)	N(3,1)	N(3.25,1)
Case II^d		Level 1	Cauchy(1,1)	Cauchy(4,1)	Cauchy(4,1)
		Level 2	Cauchy(1,1)	Cauchy(4,1)	Cauchy(4,1)
		Level 3	Cauchy(1,1)	Cauchy(4,1)	Cauchy(4,1)
Case III^d		Level 1	LN(1,1,1)	LN(4,1,1)	LN(4,1,1)
		Level 2	LN(1,1,1)	LN(4,1,1)	LN(4,1,1)
		Level 3	LN(1,1,1)	LN(4,1,1)	LN(4,1,1)

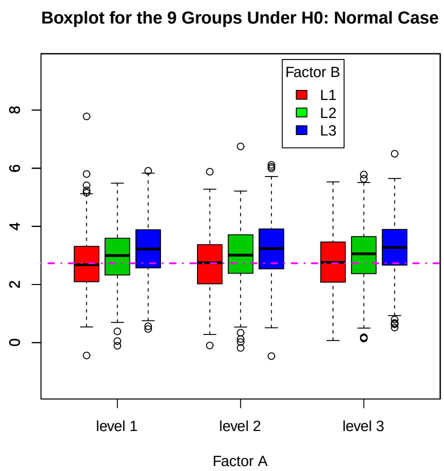
As illustrated in previous paragraphs, Table 3.6 shows the distributions of the 9 cells under the null hypothesis, i.e. no row effect. For instance, in Case I^d, the distribution of the response variable in cell $(1, 1)$, when Factor A is set at the first level and Factor B is set at the first level as well, is Normal distribution with mean 2.75 and standard deviation 1, denoted as $N(2.75, 1)$. In Case II^d, the distribution of the response variable in cell $(2, 3)$, when Factor A is set at the second level and Factor B is set at the third level, is Cauchy distribution with location parameter 4 and scale parameter 1, denoted as $Cauchy(4, 1)$. In Case III^d, the distribution of the response variable in cell $(3, 1)$, when Factor A is set at the third level and Factor B is set at the first level, is three-parameter Lognormal distribution with location parameter 1, scale parameter 1, and shape parameter 1, denoted as $LN(1, 1, 1)$.

Table 3.7 shows the distributions of the 9 cells under the alternative hypothesis, i.e. there exist row effects. For instance, in Case I^d, the distribution of the response variable in cell $(2, 1)$, when Factor A is set at the second level and Factor B is set at the second level as well, is Normal distribution with mean 2.75 and standard deviation 1, denoted as $N(2.75, 1)$. In Case II^d, the distribution of the response variable in cell $(3, 3)$, when Factor A is set at the third level and Factor B is set at the third level, is Cauchy distribution with location parameter 4.75 and scale parameter 1, denoted as $Cauchy(4.75, 1)$. In Case III^d, the distribution of the response variable in cell $(3, 1)$, when Factor A is set at the third level and Factor B is set at the first level, is three-parameter Lognormal distribution with location parameter 1.75, scale parameter 1, and shape parameter 1, denoted as $LN(1.75, 1, 1)$.

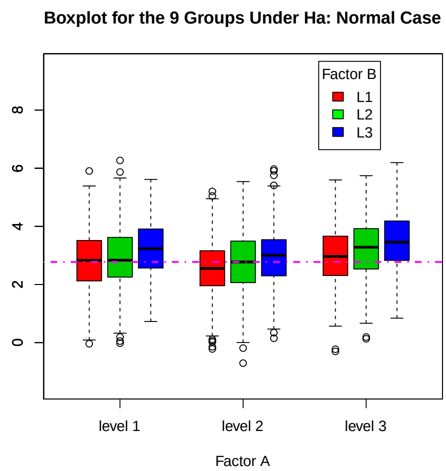
Table 3.7: Evaluate the Power of Tests for Row Effect in 3 Cases

			Factor B		
			Level 1	Level 2	Level 3
Case I^d	**Factor A**	Level 1	N(2.75,1)	N(3,1)	N(3.25,1)
		Level 2	N(2.5,1)	N(2.75,1)	N(3,1)
		Level 3	N(3,1)	N(3.25,1)	N(3.5,1)
Case II^d		Level 1	Cauchy(1.75,1)	Cauchy(4.75,1)	Cauchy(4.75,1)
		Level 2	Cauchy(-0.5,1)	Cauchy(2.5,1)	Cauchy(2.5,1)
		Level 3	Cauchy(1.75,1)	Cauchy(4.75,1)	Cauchy(4.75,1)
Case III^d		Level 1	LN(1.75,1,1)	LN(4.75,1,1)	LN(4.75,1,1)
		Level 2	LN(-0.5,1,1)	LN(2.5,1,1)	LN(2.5,1,1)
		Level 3	LN(1.75,1,1)	LN(4.75,1,1)	LN(4.75,1,1)

Before comparing the simulation results of the parametric and nonparametric tests, the row effects in locations of the distributions under the null hypothesis (in Table 3.6) and alternative hypothesis (in Table 3.7) are illustrated by the side-by-side boxplots in Figure 3.5-Figure 3.12. Figure 3.5(a), Figure 3.7(a), Figure 3.8(a) and Figure 3.11(a) show the side-by-side boxplots of the 9 cells in the Case I^d, II^d and III^d respectively in Table 3.6. Figure 3.5(b), Figure 3.7(b), Figure 3.8(b) and Figure 3.11(b) show the side-by-side boxplots of the 9 cells in the Case I^d, II^d and III^d respectively in Table 3.7. The horizontal axis represents the levels of factor A, and the vertical axis represents the response variable generated from the distributions in Table 3.6 or Table 3.7. The colored lines represent the levels of factor B. The red box represents the level 1 of factor B. The green box represents the level 2 of factor B. And the blue box represents the level 3 of factor B. Thus, the boxplot in red box and located at level 2 on the horizontal axis is the plot for the cell when factor A is at level 2 and factor B is at level 1. Figure 3.5 shows the side-by-side boxplot of the 9 cells in Normal case under null hypothesis and alternative hypothesis.



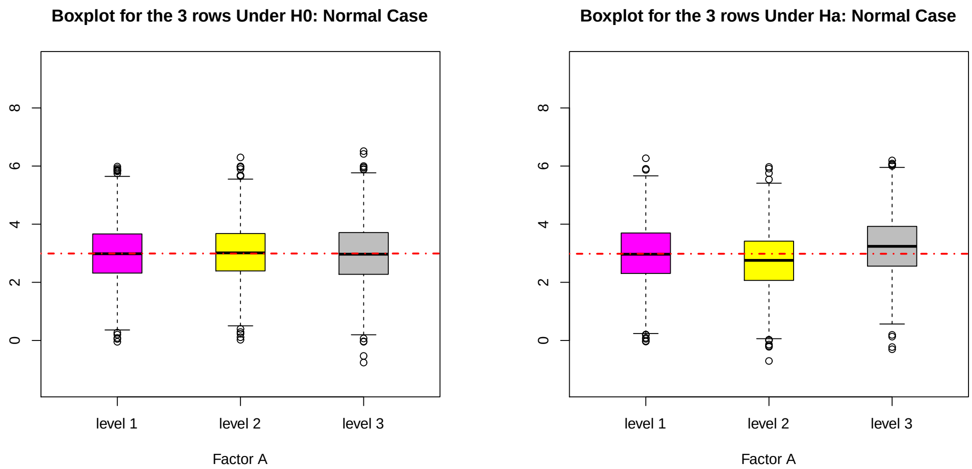
(a) Under H_0: $\alpha = \mathbf{0}$



(b) Under H_a: $\alpha = \alpha_{\mathbf{N}}$

Figure 3.5: Side-by-Side Boxplot for the 9 Cells in Case I^d: Normal Distributions

The pink horizontal line in Figure 3.5 serves as a benchmark, which indicates the median of the samples from cell (1,1). This benchmark line, also used in Figure 3.6, Figure 3.11 and Figure 3.12, aims to help notifying the tiny differences in the median of boxplots under the alternative hypothesis. The boxplot in Figure 3.5(a) indicates that column effects do exist since the edge line (quartiles) and the middle line (median) of red, green and blue box are not in a line. Figure 3.5(a) also indicates that there is no row effect among the 9 cells in Case I^d of Table 3.6, since all the three red boxes lie on the same line and so do the green boxes and blue boxes. Figure 3.5(b) infers that there exists row effect among the 9 cells in Case I^d of Table 3.7, since the three red boxes are not on a line and so do the green boxes and blue boxes.

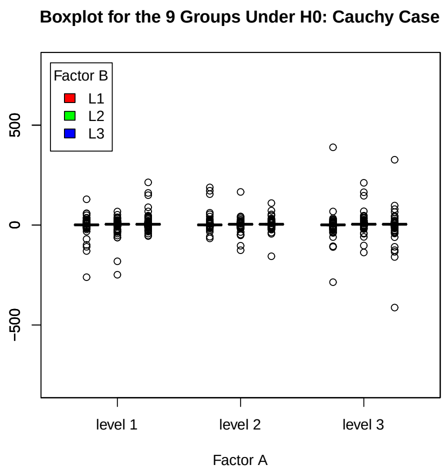


(a) Under H_0: $\alpha = \mathbf{0}$

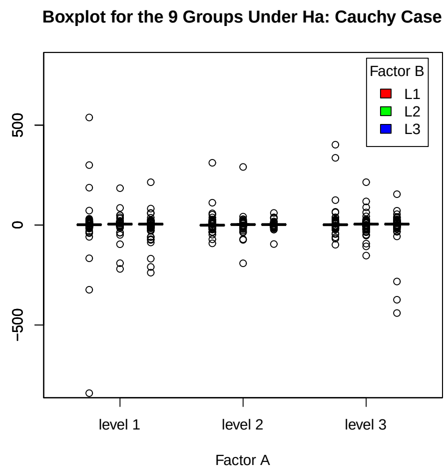
(b) Under H_a: $\alpha = \alpha_{\mathbf{N}}$

Figure 3.6: Side-by-Side Boxplot for the 3 Rows in Case I^d: Normal Distributions

To have a better view of the row effect under the null and alternative hypothesis, cells from the same row are combined for each case both under null and alternative hypothesis. For example, cell (1,1), (1,2) and (1,3) are the samples when factor A is at the first level, and thus they are combined into one set of data. In this case, there will be 3 sets of data for each case, one from row 1 when the factor A is at level 1, one from row 2 when the factor A is at level 2 and one from row 3 when the factor A is at level 3. Figure 3.6, Figure 3.9, Figure 3.10 and Figure 3.12 show the side-by-side boxplots of the three combined rows for the three cases in Table 3.6 and Table 3.7. Figure 3.6(a) shows the side-by-side boxplots of the 3 combined rows in the Case I^d under the null hypothesis $\alpha = \mathbf{0}$. Figure 3.6(b) shows the side-by-side boxplots of the 3 combined rows in the Case I^d under the alternative hypothesis $\alpha = \alpha_{\mathbf{N}}$. Like Figure 3.5, the horizontal axis in Figure 3.6 represents the levels of factor A, and the vertical axis represents the response variable. The deep pink box represents the level 1 of factor A. The yellow box represents the level 2 of factor A. And the gray box represents the level 3 of factor A. The red line in Figure 3.6 is a benchmark, which points to the median of the samples when factor A is at the first level. The relationship among the locations of the boxes and the benchmark in Figure 3.6(a) and Figure 3.6(b) indicates that there is no row effects in Case I^d of Table 3.6, and there exist some row effects in Case I^d of Table 3.7.



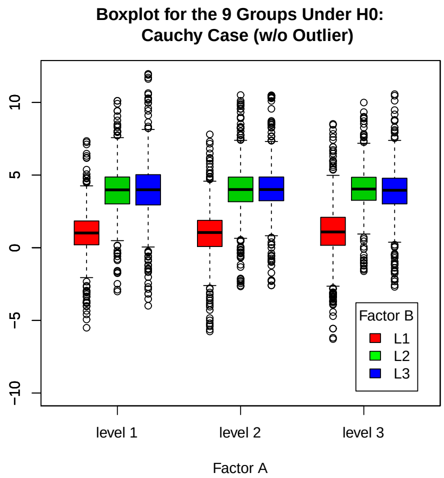
(a) Under H_0: $\alpha = \mathbf{0}$



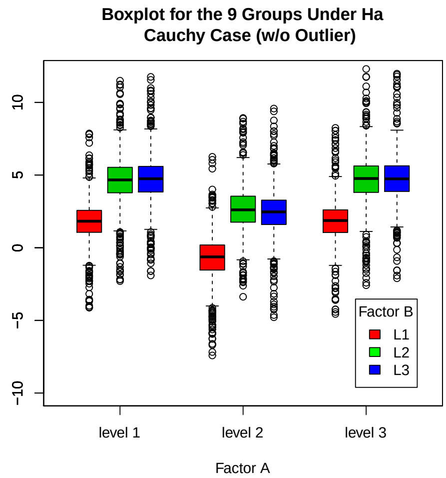
(b) Under H_a: $\alpha = \alpha_{\mathbf{CL}}$

Figure 3.7: Side-by-Side Boxplot for the 9 Cells in Case II^d: Cauchy Distributions

Similar to Figure 3.5, Figure 3.7 shows the side-by-side boxplot of the 9 cells in the Cauchy Case under null hypothesis and alternative hypothesis. Since Cauchy distribution has fat tails, it is hard to tell the color and the middle line of the boxplots. Hence, Figure 3.8 plots the truncated 9 cells in Cauchy case under null hypothesis and alternative hypothesis by removing the extreme outliers. The relationship among the locations of the boxes and the benchmark in Figure 3.8(a) and Figure 3.8(b) indicate that there is no row effect in Case II^d of Table 3.6, and there exist some row effects in Case II^d of Table 3.7.



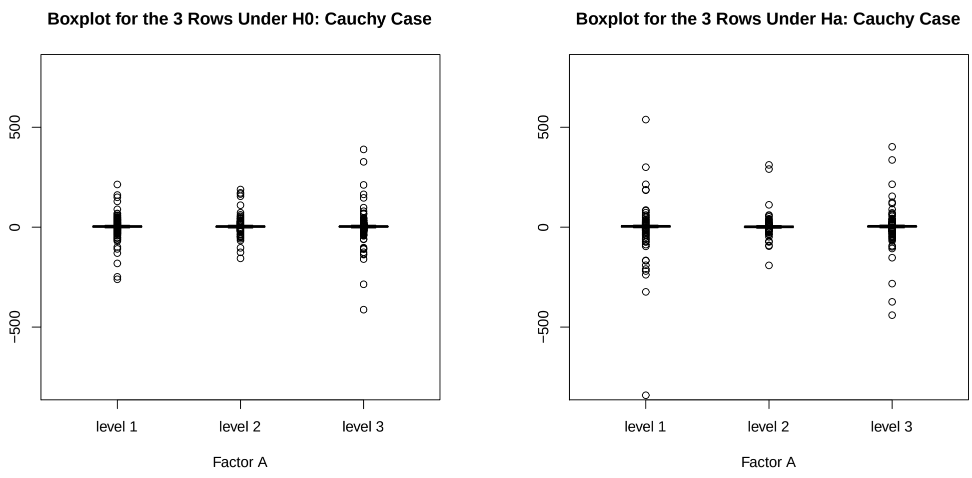
(a) Under H_0: $\alpha = \mathbf{0}$



(b) Under H_a: $\alpha = \alpha_{\mathbf{CL}}$

Figure 3.8: Side-by-Side Boxplot (w/o extreme outliers) for the 9 Cells in Case II^d: Cauchy Distributions

Like Figure 3.6, Figure 3.9(a) shows the side-by-side boxplots of the 3 combined rows in the Case II^d under the null hypothesis $\alpha = \mathbf{0}$. Figure 3.9(b) shows the side-by-side boxplots of the 3 combined rows in the Case II^d under the alternative hypothesis $\alpha = \alpha_{\mathbf{N}}$. Due to the heavy tails of the Cauchy distribution, it is difficult to tell the relative locations of the boxes. Figure 3.10 plots the truncated 3 combined rows in Cauchy Case under null hypothesis and alternative hypothesis by removing the extreme outliers. The relationship among the locations of the boxes in Figure 3.10(a) and Figure 3.10(b) confirms that there is no row effect in Case II^d of Table



(a) Under H_0: $\alpha = \mathbf{0}$

(b) Under H_a: $\alpha = \alpha_{\mathbf{CL}}$

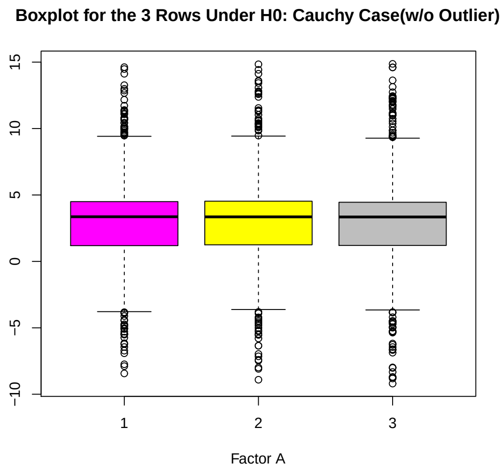
Figure 3.9: Side-by-Side Boxplot for the 3 Rows in Case II^d: Cauchy Distributions

3.6, and there exist some row effects in Case II^d of Table 3.7.

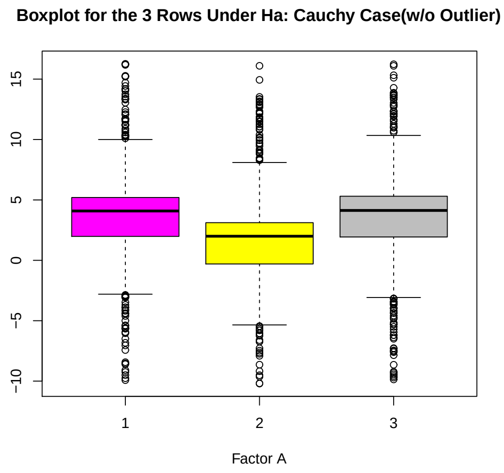
Similar to Figure 3.5, Figure 3.11 shows the side-by-side boxplot of the 9 cells in the lognormal case under null hypothesis and alternative hypothesis. The related locations of the boxes and benchmark in Figure 3.11(a) and Figure 3.11(b) indicate that there is no row effect in Case III^d of Table 3.6, and there exist some row effects in Case III^d of Table 3.7. Like Figure 3.6, Figure 3.12(a) shows the side-by-side boxplots of the 3 combined rows in the Case III^d under the null hypothesis $\alpha = \mathbf{0}$. Figure 3.12(b) shows the side-by-side boxplots of the 3 combined rows in the Case III^d under the alternative hypothesis $\alpha = \alpha_{\mathbf{CL}}$. The relationship among the locations of the boxes in Figure 3.12(a) and Figure 3.12(b) confirms that there is no row effect in Case III^d of Table 3.6, and there exist some row effects in Case III^d of Table 3.7.

The simulation results for the test of row effects via the kernel based nonparametric two-way ANOVA test and parametric two-way ANOVA test are given in Table 3.8-Table 3.10.

Table 3.8 lists the actual Type I error rates and the empirical powers for the test of

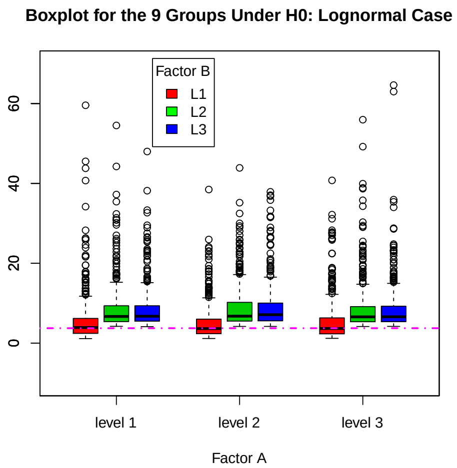


(a) Under H_0: $\alpha = \mathbf{0}$

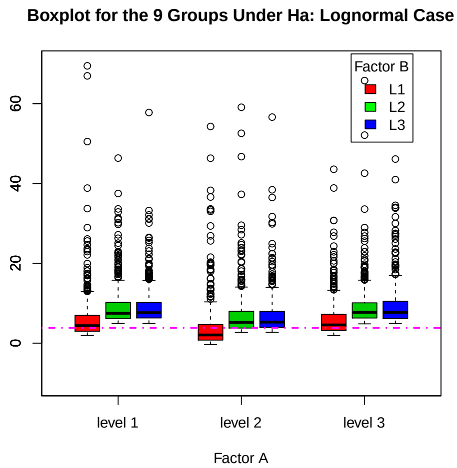


(b) Under H_a: $\alpha = \alpha_{\mathbf{CL}}$

Figure 3.10: Side-by-Side Boxplot (w/o extreme outliers) for the 3 Rows in Case II^d: Cauchy Distributions



(a) Under H_0: $\alpha = \mathbf{0}$



(b) Under H_a: $\alpha = \alpha_{\mathbf{CL}}$

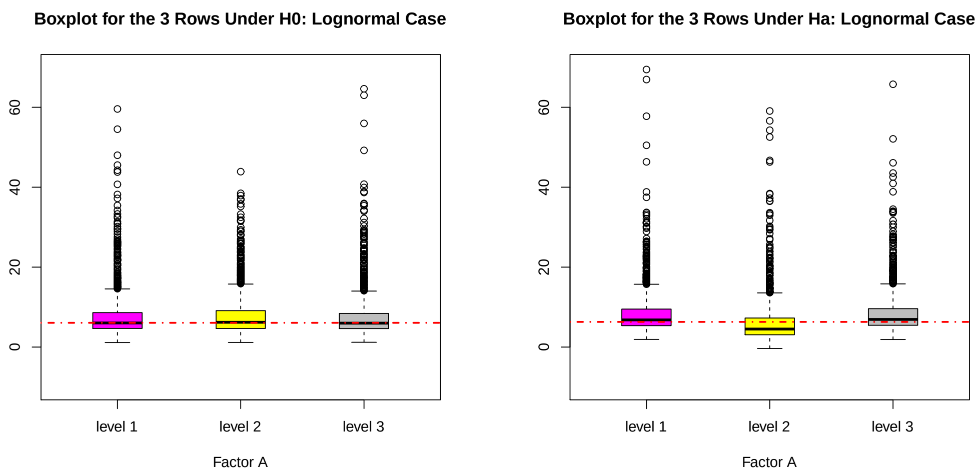
Figure 3.11: Side-by-Side Boxplot for the 9 Cells in Case III^d: Lognormal Distributions

Table 3.8: Power for Test of Row Effect: Case I^d (Normal Distribution)

Sample Size	Parametric Test		Nonparametric Test	
	Power	Type I	Power	Type I
15	0.5408	0.0517	0.2329	0.0128
20	0.6774	0.0469	0.3427	0.0118
25	0.7812	0.0530	0.4450	0.0088
30	0.8582	0.0505	0.5512	0.0067
35	0.9076	0.0541	0.6353	0.0072
40	0.9430	0.0487	0.7236	0.0113
45	0.9644	0.0492	0.7855	0.0103
50	0.9820	0.0473	0.8456	0.0067
55	0.9870	0.0528	0.8807	0.0063
60	0.9930	0.0474	0.9220	0.0065

[1] $*$ Note: Given significance level: $\alpha = 0.05$.

[2] $*$ Note: The Type I in the table means Actual Type I Error.



(a) Under H_0: $\alpha = \mathbf{0}$ (b) Under H_a: $\alpha = \alpha_{\mathbf{CL}}$

Figure 3.12: Side-by-Side Boxplot for the 3 Rows in Case III^d: Lognormal Distributions

row effects in Case I^d: Normal Case. It is shown that the actual Type I error rates for the parametric test of row effects are around 0.05, the significance level. The actual Type I error rates for the new nonparametric test of row effects are between 0.006 and 0.02, which infers that the new nonparametric test of row effects is very conservative in Case I^d. As we expected, in Normal Case, the parametric F test performs better than the kernel based nonparametric test of row effects. The new nonparametric test was hurt by its conservativeness in Type I error rates. Fortunately, the power of the nonparametric test increases as the sample size grows. When the sample size is 60, the power of the nonparametric test of interaction is 92.2%, which looks promising.

Table 3.9 lists the actual Type I error rates and the empirical powers for the test of row effects in Case II^d: Cauchy Case. It is shown that the actual Type I error rates of the parametric test are around 0.02 while the significance level is 0.05. This fact infers that the parametric test of row effects is conservative in Case II^d. Whereas, the actual Type I error rates of the new nonparametric test are around

Table 3.9: Power for Test of Row Effect: Case II^d (Cauchy Distribution)

Sample Size	**Parametric Test**		**Nonparametric Test**	
	Power	**Type I**	**Power**	**Type I**
15	0.1772	0.0186	0.7040	0.0578
20	0.1743	0.0180	0.8611	0.0548
25	0.1766	0.0200	0.9398	0.0459
30	0.1812	0.0180	0.9783	0.0507
35	0.1783	0.0184	0.9917	0.0480
40	0.1770	0.0201	0.9969	0.0454
45	0.1806	0.0169	0.9985	0.0437
50	0.1774	0.0187	0.9999	0.0472
55	0.1768	0.0195	1	0.0447
60	0.1808	0.0202	1	0.0428

[1] * Note: Given significance level: $\alpha = 0.05$.

[2] * Note: The Type I in the table means Actual Type I Error.

0.05, the significant level. Without any surprise, in the Cauchy Case, the kernel based nonparametric test performs much better than the parametric F test of row effect. The power of the new nonparametric test of row effects is 70.4% when the cell sample size is only 15. Moreover, the power of nonparametric test of row effect quickly increases to 1 as the sample size rises. The power of the parametric test of row effects is only 17.72% when the sample size is 15 and does not increase as the sample size grows.

Table 3.10: Power for Test of Row Effect: Case III^d (Lognormal Distribution)

Sample Size	**Parametric Test**		**Nonparametric Test**	
	Power	**Type I**	**Power**	**Type I**
15	0.5328	0.0443	0.7194	0.0467
20	0.6434	0.0382	0.7951	0.0512
25	0.7213	0.0429	0.8693	0.0519
30	0.7802	0.0409	0.9209	0.0502
35	0.8295	0.0433	0.9509	0.0489
40	0.8665	0.0455	0.9656	0.0466
45	0.8997	0.0479	0.9800	0.0474
50	0.9240	0.0419	0.9845	0.0450
55	0.9376	0.0443	0.9888	0.0439
60	0.9511	0.0461	0.9933	0.0447

[1] $*$ Note: Given significance level: $\alpha = 0.05$.

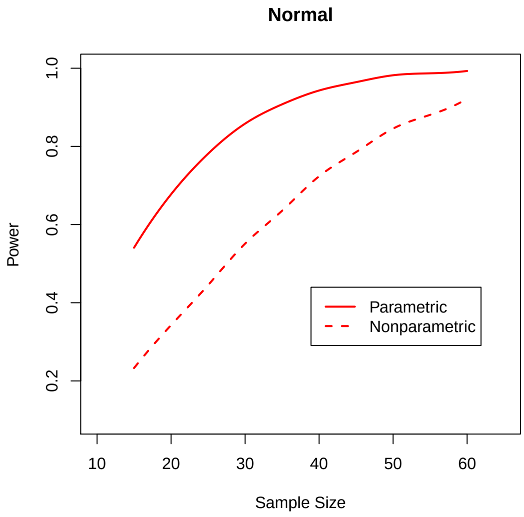
[2] $*$ Note: The Type I in the table means Actual Type I Error.

Table 3.10 lists the actual Type I error rates and the empirical powers for the test of row effects in Case III^d: Lognormal Case. It is shown that the actual Type

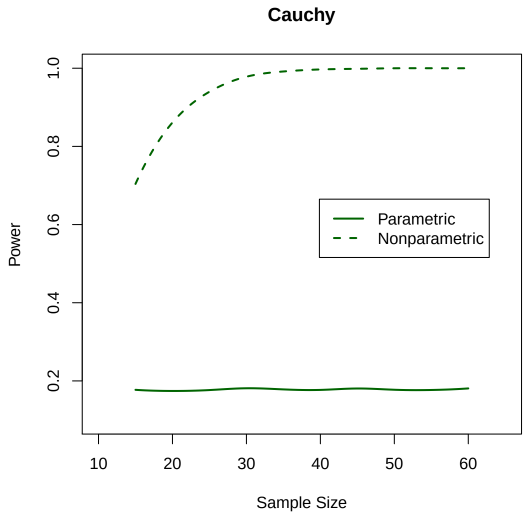
I error rates for the parametric test are around 0.04 when the significance level is 0.05, which infers that the parametric test of row effects is little bit conservative in Case III^d. Whereas, the actual Type I error rates of the new nonparametric test are around 0.05, the significance level. Same as in the Cauchy Case, in the Lognormal Case, the kernel based nonparametric test performs much better than the parametric F test of row effects. The power of nonparametric test of row effects is 71.94% when the cell sample size is only 15. Moreover, the power of the new nonparametric test of row effects quickly increases to 95% when the sample size rises up to 35. The power of the parametric test of row effects is only 53.28% when the sample size is 15 and only increases to 82.95% as the sample size grows to 35.

Figure 3.13 demonstrates the power of the parametric and nonparametric test of row effects with respect to sample size when the underlying distributions of the samples are Normal, Cauchy and Lognormal respectively. In Figure 3.13, the solid line represents the power of the parametric test of row effects, while the dashed line represents the power of the kernel based nonparametric test of row effects in Section 3.1.1. The red line (solid and dashed) represents the power of the tests in Case I^d when the underlying distribution is Normal distribution. The green line (solid and dashed) represents the power of the tests in Case II^d when the underlying distribution is Cauchy distribution. And the blue line (solid and dashed) represents the power of the tests in Case III^d when the underlying distribution is Lognormal distribution.

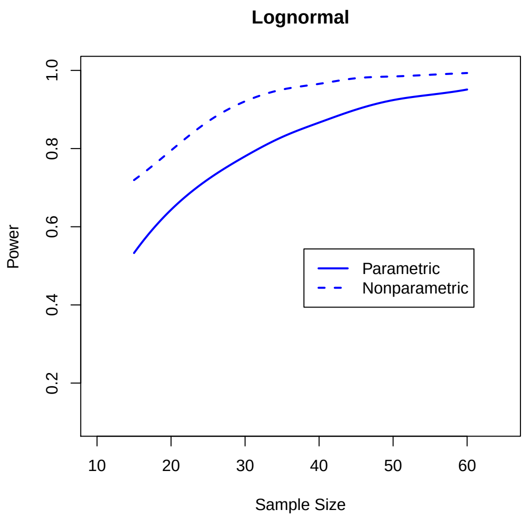
Figure 3.13(a) compares the power of parametric and the new nonparametric test of row effects in Case I^d. It shows that the power of the parametric test increases to around 90% when sample size goes up to 40 in Case I^d. Compare to the parametric test of row effects, the new nonparametric test is much less powerful than the traditional parametric test, which is consistent with what we concluded from Table 3.8. Fortunately, the new nonparametric test increases its power as the sample size grows. Although in Figure 3.13(a), it is shown that there is still a gap in power between



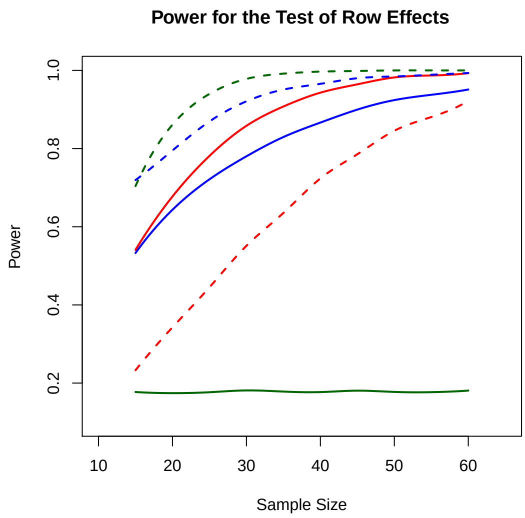
(a) Case I^d: Normal



(b) Case II^d: Cauchy



(c) Case III^d: Lognormal



(d) Case I^d,II^d, III^d

Figure 3.13: (a) Power of the parametric and nonparametric test of row effect on the 9 cells in Case I^d; (b) Power of the parametric and nonparametric test of row effect on the 9 cells in Case II^d; (c) Power of the parametric and nonparametric test of row effect on the 9 cells in Case III^d; (d) Power of the parametric and nonparametric test of row effect on the 9 cells in Case I^d, II^d and III^d.

the parametric and our nonparametric test, the gap tends to reduce as the sample size go beyond 60. Figure 3.13(b) compares the power of the parametric test with the kernel based nonparametric test of row effects in Case II^d. It is shown in Figure 3.13(b) that the power of the new nonparametric ANOVA test quickly grows to 1 as the sample size increases, whereas the power of the parametric ANOVA test almost keeps constant in Case II^d. Moreover, it is easy to tell from Figure 3.13(b) that the power of the new nonparametric test of row effects in Case II^d is much higher than the parametric test, even when the sample size is 15. Figure 3.13(c) compares the power of parametric and nonparametric test of row effect in Case III^d. In Figure 3.13(c), it is demonstrated that the blue dashed line is above the blue solid line even when the sample is as small as 15, which infers that the new nonparametric test outperforms the parametric test in Case III^d. Figure 3.13(d) combines Figure (3.13(a))(3.13(b))(3.13(c)) in one graph, so it is easier to compare the powers of the either test throughout the 3 cases. The relationship of three solid lines infers that the power of the parametric ANOVA test of row effects is far more severely hurt by fat tails or extreme outliers than the skewness does. If the data come from a skewed distribution, large sample size leads to better power. However, if the data come from a fat-tailed distribution, large sample size would not make any improvement in power. The relationship of three dashed lines infers that the new nonparametric test of row effects for the two-way ANOVA is not as powerful as the traditional parametric two-way ANOVA test of row effect, as well as, the nonparametric one-way ANOVA if the underlying distribution is Normal.

CHAPTER 4

Application to Policy Analysis

4.1 Introduction to Policy Analysis

Policy decisions are required to be made in corporations and/or most levels of government every year or even every day. Then policy analysis emerges to analyze policy-related information and provides policy decision makers with rational decisions. For instance, the U.S. Environmental Protection Agency (EPA) is planning new rules to regulate the interstate transport of sulfur dioxide (SO2) and nitrogen oxides (NOX) emitted from electric power generation facilities. Before the agency determining which newly proposed rules or neither of them should be taken, the economic impacts to regions, sectors and populations have to be assessed. The agency will choose the rule that produces significant benefits in terms of some variables reflecting improved health outcomes, and better environmental amenities and services. In order to make a rational decision, a policy analyst might not only be interested in the positive or negative relationships the dependent variable may have with the policy change, but also in estimating the mean benefit gained from a policy change, such as the mean change in house prices resulting from building a high-quality school or a highway nearby, or the mean change in consumption resulting from a change in income taxes. Thus, quantitative, rather than qualitative analytical techniques are most in need in policy analysis.

The quantitative policy analysis has its root in Harold Lasswell ([24], [25]) and has rapidly developed in social science and business since 1990s [14]. Statistical methods, such as, analysis of variance (ANOVA), regression analysis, are commonly used to evaluate the benefits or costs of proposed policies on dependent variables related to individual or social welfare. However, all parametric models, such as ANOVA, regression rely on the normality assumption. Deluchi & Bostrom [12] suggested to be wary of standard methods, such as t tests and parametric ANOVA when analyzing skewed-distribution data. It is known that the distributions of many economic and/or financial variables (income, wealth, prices, asset return) are right skewed with fat tails, since these variables never take values less than zero [32]. A more appropriate alternative for comparing the policy effects may be to use a nonparametric or distribution-free method.

4.2 Stock's Nonparametric Policy Analysis

Stock [38] proposed a nonparametric procedure to estimate the mean effect of certain policy interventions. Suppose Y is the dependent variable of interest, such as, the house price, and X is a p dimensional vector of independent variables, such as the size of the lot, the living area in the house, age of the house, before the policy of cleaning up a local hazardous waste site [39]. Let Y^* and X^* be the corresponding dependent and independent variables after the policy intervention. The policy benefit B is defined by the mean change of dependent variable Y after policy, i.e. $B = EY^* - EY$. Consider the semi-parametric regression model:

$$Y_i = g(X_i) + A'd_i + e_i, \tag{4.1}$$

where, A is a p dimensional vector of cell effects, $E(e_i|X_i, d_i) = 0$ and $E(e_i^2|X_i, d_i) = \sigma^2(x, d) < \infty$ for $i = 1, 2, \cdots, n$. By assuming the cell effects remain unchanged after the policy, the policy benefit becomes:

$$B = E^*g(X_i) - Eg(X_i), \tag{4.2}$$

where, $E^*[\cdot]$ is the expectation taken over X^*. To estimate B, Stock [38] firstly estimated A by borrowing the idea of ordinary least squares (OLS) estimator of A

in general linear models, then gave the kernel estimate of $g(X)$ by plugging in the estimated A for both cases: before and after policy intervention, denoted as $g_n(X)$ and $g_n(X^*)$. Thus, the policy benefit is estimated by

$$B_n = \frac{1}{n} \sum_{i=1}^{n} (g_n(X_i^*) - g_n(X_i)). \tag{4.3}$$

B_n is a consistent estimate of B. Details are described in [38]. See also Ahmad, Leelahanon and Li [3] for an extended semi-parametric approach.

Stock's policy analysis model is limited in several aspects. Firstly, Stock's model only provides an estimate of the gain or loss of the new policy. Decision makers may find it difficult to make a decision through only estimators, rather than a hypothesis test. It is also not very meaningful to interpret a difference in policy impacts when the difference is actually not statistically significant. Secondly, it can only evaluate one policy with two levels at a time. There are plenty of situations in which three or more policy effects need to be evaluated. For example, as in [21] one may need to evaluate the effects of three land uses: golf courses, a university, and a nitrogen plant, on the neighborhood home values in Lawrence, Kansas. Thirdly, Stock's model uses the average change of the dependent variable as a measurement of policy impacts. It is known that most of the economic variables are skewed, sometimes strongly skewed. The mean of samples from skewed distributions is not a good measurement of central tendency. It is very sensitive to extreme values. Thus, the average change of dependent variable should not be an appropriate policy benefit measurement. Finally, in Stock's model, $g(x)$ and A are estimated by directly utilizing the analogy format of the OLS estimator in generalized linear models without any theoretical verification. To sum up, the development of a more generalized and reliable policy analysis technique is a crucial task for many current policy analysts. In the next section, we will propose a hypothesis test of policy benefits in locations (such as, median) of the policy related variables.

4.3 A New Approach of Policy Analysis

As we discussed in the previous section, mean or average is not a good choice of measurements when the samples come from a skewed distribution. Therefore, the location of the samples, defined in Chapter 2, is considered instead. Suppose there are K policies that need to be evaluated. One of them is the old policy or current policy, and the other $K-1$ are newly proposed policies. In order to evaluate the efficiency of the new policies, the location effects of K policies on a dependent variable $\mathbf{X}$ and explanatory variable $\mathbf{U}$ are compared. Suppose (X_{ij}, U_{ij}) is the pair of observations from the j^{th} individual and under the i^{th} policy. Define μ_{i*} and σ_{i*} be the location and scale of the dependent variable X under the i^{th} policy. Then the efficiency of the K policies can be evaluated by performing a hypothesis test with the null hypothesis H_0: $\mu_{1*} = \mu_{2*} = \ldots = \mu_{K*}$, which means none of the new policies do any better or worse than the old one.

Assume that (X_{ij}, U_{ij}) comes from a joint distribution $f_i(x, u)$, where $i = 1, 2, \cdots, K$ and $j = 1, 2, \ldots, n_i$. By the definition of location parameter in Chapter 2, the following equation holds:

$$f_i(x,u) = \frac{1}{\sigma_{i*}} f_{0*}\left(\frac{x-\mu_{i*}}{\sigma_{i*}}, u\right), \tag{4.4}$$

where $f_{0*}(\cdot,\cdot)$ is a base density. Thus, we have

$$\begin{aligned}\int\int x f_i^2(x,u)dxdu &= \int\int (\frac{y\sigma_{i*}+\mu_{i*}}{\sigma_{i*}}) f_{0*}^2(y,u)dydu \\ &= \int y f_{0*}^2(y,u)dy + \frac{\mu_{i*}}{\sigma_{i*}}\int f_{0*}^2(y,u)dy,\end{aligned}$$

which implies that

$$\begin{aligned}\mu_{i*} &= \frac{\sigma_{i*}(\int\int x f_i^2(x,u)dxdu - \int\int y f_{0*}^2(y,u)dydu)}{\int\int f_{0*}^2(y,u)dydu} \\ &= \frac{\sigma_{i*}\int\int x f_i^2(x,u)dxdu}{\int\int f_{0*}^2(y,u)dydu} - \frac{\sigma_{i*}\int\int y f_{0*}^2(y,u)dydu}{\int\int f_{0*}^2(y,u)dydu}. \end{aligned}\tag{4.5}$$

Assume $\sigma_{i*} = \sigma_*$ for all $i = 1, 2, \cdots, K$, and then equation (4.5) becomes:

$$\mu_{i*} = \frac{\sigma_* \int\int x f_i^2(x) dx du}{\int\int f_{0*}^2(y,u) dy du} - \frac{\sigma_* \int\int y f_{0*}^2(y,u) dy du}{\int\int f_{0*}^2(y,u) dy du}. \quad (4.6)$$

Let $W_i = \int\int x f_i^2(x,u) dx du$. Hence, the hypothesis of equal location H_0: $\mu_{1*} = \mu_{2*} = \ldots = \mu_{K*}$ versus H_1: $\mu_{i*} \neq \mu_{j*}$ for some $i \neq j$ becomes H_0: $W_1 = W_2 = \ldots = W_K$ against H_1: $W_i \neq W_j$ for some $i \neq j$. Consider the nonparametric kernel estimate of W_i, denoted as $\hat{W}_i$, where

$$\hat{W}_i = \frac{1}{n_i(n_i-1)h_i^2} \sum\sum_{j_1 \neq j_2} \left(\frac{X_{ij_1} + X_{ij_2}}{2}\right) K\left(\frac{X_{ij_1} - X_{ij_2}}{h_i}\right) K\left(\frac{U_{ij_1} - U_{ij_2}}{h_i}\right). \quad (4.7)$$

Note that if $\sigma_{i*} \neq \sigma_*$ for some i, then the test statistic in Proposition 4.4 is to test H_0: $\mu_{1*}/\sigma_{1*} = \mu_{2*}/\sigma_{2*} = \ldots = \mu_{K*}/\sigma_{K*}$ vs H_a: $\mu_{i*}/\sigma_{i*} \neq \mu_{j*}/\sigma_{j*}$ for some $i \neq j$, instead of H_0: $\mu_{1*} = \mu_{2*} = \ldots = \mu_{K*}$ versus H_1: $\mu_{i*} \neq \mu_{j*}$ for some $i \neq j$.

Lemma 4.1 *If for any $i = 1, 2, \cdots, K$, $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ as $\min_i n_i \to \infty$, $\int\int x^2 f_i^3(x,u) dx du < \infty$ and if $f_i(x,u)$ is twice differentiable with respect to x and u, then*

$$\sqrt{n_i}(\hat{W}_i - W_i) \xrightarrow{d} N(0, \omega_{i*}^2), \quad (4.8)$$

as $\min_i n_i \to \infty$, where, $\omega_{i*}^2 = 4\{\int\int x^2 f_i^3(x,u) dx du - (\int\int x f_i^2(x,u) dx du)^2\}$.

Proof: Let $\varphi_*(X_{ij_1}, X_{ij_2}, U_{ij_1}, U_{ij_2}) = \left(\frac{X_{ij_1}+X_{ij_2}}{2h_i^2}\right) K\left(\frac{X_{ij_1}-X_{ij_2}}{h_i}\right) K\left(\frac{U_{ij_1}-U_{ij_2}}{h_i}\right)$, then $\hat{W}_i$

is a U-statistics with mean

$$
\begin{aligned}
E(\hat{W}_i) &= E(\frac{X_{i2}}{h_i^2} K\left(\frac{X_{i1}-X_{i2}}{h_i}\right) K\left(\frac{U_{i1}-U_{i2}}{h_i}\right)) \\
&= \frac{1}{h_i^2} \int\int\int\int x_2 K\left(\frac{x_1-x_2}{h_i}\right)\left(\frac{u_1-u_2}{h_i}\right) f_i(x_1,u_1) f_i(x_2,u_2) dx_1 dx_2 du_1 du_2 \\
&= \int\int\int\int x_2 K(u) K(v) f_i(x_2+uh_i, u_2+vh_i) f_i(x_2,u_2) dudvdx_2du_2 \\
&= \int\int\int\int x_2 K(u) K(v) \big[f_i(x_2,u_2) + \tfrac{\partial f_i(x,u)}{\partial x}\Big|_{(x_2,u_2)} uh_i \\
&\quad + \tfrac{\partial f_i(x,u)}{\partial u}\Big|_{(x_2,u_2)} vh_i + o(h_i)\big] f_i(x_2,u_2) dudvdx_2du_2 \\
&= \int\int\int\int x_2 K(u) K(v) f_i^2(x_2,u_2) dudvdx_2du_2 + o(h_i) \\
&= \int\int x_2 f_i^2(x_2,u_2) dx_2du_2 + o(h_i) \\
&\simeq W_i \qquad (4.9)
\end{aligned}
$$

and variance written as

$$
\begin{aligned}
Var(\hat{W}_i) &= \frac{4}{n_i} cov(\varphi_*(X_{ij_1}, X_{ij_2}, U_{ij_1}, U_{ij_2}), \varphi_*(X_{ij_1}, X_{ij_3}, U_{ij_1}, U_{ij_3})) \\
&\quad + \frac{2}{n_i(n_i-1)} var(\varphi_*(X_{ij_1}, X_{ij_2}, U_{ij_1}, U_{ij_2})).
\end{aligned}
$$

It can easily be shown that $var(\varphi(X_{ij_1}, X_{ij_2}, U_{ij_1}, U_{ij_2})) = O(h_i^{-1})$ and since $1/(n_ih_i) = o(1)$, the second term of $Var(\hat{W}_i)$ in the parentheses can be neglected. Then the variance of $\hat{W}_i$ is dominated by $\frac{1}{n_i}\omega_{i*}^2$, where

$$
\begin{aligned}
\omega_{i*}^2 &= 4cov(\varphi_*(X_{ij_1}, X_{ij_2}, U_{ij_1}, U_{ij_2}), \varphi_*(X_{ij_1}, X_{ij_3}, U_{ij_1}, U_{ij_3})) \\
&= 4[E(\varphi_*(X_{ij_1}, X_{ij_2}, U_{ij_1}, U_{ij_2})\varphi_*(X_{ij_1}, X_{ij_3}, U_{ij_1}, U_{ij_3})) \\
&\quad -E(\varphi_*(X_{ij_1}, X_{ij_2}, U_{ij_1}, U_{ij_2}))E(\varphi_*(X_{ij_1}, X_{ij_3}, U_{ij_1}, U_{ij_3}))] \\
&= 4\Big[\frac{1}{4h_i^4}\int\int\int\int\int\int (x_1+x_2)(x_1+x_3)K(\frac{x_1-x_2}{h_i})K(\frac{x_1-x_3}{h_i}) \\
&\quad K(\frac{u_1-u_2}{h_i})K(\frac{u_1-u_3}{h_i})f_i(x_1,u_1)f_i(x_2,u_2)f_i(x_3,u_3)dx_1dx_2dx_3du_1du_2du_3 \\
&\quad -\Big(\frac{1}{h_i^2}\int\int\int\int x_1K(\frac{x_1-x_2}{h_i})K(\frac{u_1-u_2}{h_i})f_i(x_1,u_1)f_i(x_2,u_2)dx_1dx_2du_1du_2\Big)^2\Big] \\
&= \int\int\int\int\int\int (2x_1+y_1h_i)(2x_1+y_2h_i)K(y_1)K(y_2)K(z_1)K(z_2) \\
&\quad f_i(x_1,u_1)f_i(x_1+y_1h_i,u_1+z_1h_i)f_i(x_1+y_2h_i,u_1+z_2h_i)dx_1dy_1dy_2du_1dz_1dz_2 \\
&\quad -4\Big(\frac{1}{h_i^2}\int\int\int\int x_1K(y_1)K(z_1)f_i(x_1,u_1)f_i(x_1+y_1h_i,u_1+z_1h_i)dx_1dy_1du_1dz_1\Big)^2 \\
&= \int\int\int\int\int\int (4x_1^2+2x_1y_2h_i+2x_1y_1h_i+y_1y_2h_i^2)K(y_1)K(y_2)K(z_1)K(z_2) \\
&\quad f_i(x_1,u_1)[f_i(x_1,u_1)+\tfrac{\partial f_i(x,u)}{\partial x}\Big|_{(x_1,u_1)}y_1h_i+\tfrac{\partial f_i(x,u)}{\partial u}\Big|_{(x_1,u_1)}z_1h_i+o(h_i)] \\
&\quad [f_i(x_1,u_1)+\tfrac{\partial f_i(x,u)}{\partial x}\Big|_{(x_2,u_2)}y_2h_i+\tfrac{\partial f_i(x,u)}{\partial u}\Big|_{(x_2,u_2)}z_2h_i+o(h_i)]dx_1dy_1dy_2du_1dz_1dz_2 \\
&\quad -4\Big(\frac{1}{h_i^2}\int\int\int\int x_1K(y_1)K(z_1)f_i(x_1,u_1)[f_i(x_1,u_1)+\tfrac{\partial f_i(x,u)}{\partial x}\Big|_{(x_1,u_1)}y_1h_i \\
&\quad +\tfrac{\partial f_i(x,u)}{\partial u}\Big|_{(x_1,u_1)}z_1h_i+o(h_i)]dx_1dy_1du_1dz_1\Big)^2 \\
&= \int\int\int\int\int\int 4x_1^2K(y_1)K(y_2)K(z_1)K(z_2)f_i^3(x_1,u_1)dx_1dy_1dy_2du_1dz_1dz_2 \\
&\quad -4\Big(\frac{1}{h_i^2}\int\int\int\int x_1K(y_1)K(z_1)f_i^2(x_1,u_1)dx_1dy_1du_1dz_1\Big)^2+O(h_i) \\
&\simeq 4\Big[\int\int x_1^2f_i^3(x_1,u_1)dx_1du_1-\Big(\int\int x_1f_i^2(x_1,u_1)dx_1du_1\Big)^2\Big] \qquad (4.10)
\end{aligned}
$$

By central limit theorem of U-statistics (See Koroljuk and Borovskich [22], pp. 128-129), we have $\sqrt{n_i}(\hat{W}_i - W_i) \xrightarrow{d} N(0, \omega_{i*}^2)$, where,

$$
\omega_{i*}^2 = 4\Big\{\int\int x^2f_i^3(x,u)dxdu - \Big(\int\int xf_i^2(x,u)dxdu\Big)^2\Big\}. \qquad (4.11)
$$

■

Define

$$\hat{W}. = \frac{\sum_{i=1}^{K} n_i \hat{W}_i / \hat{\omega}_{i*}^2}{\sum_{i=1}^{K} n_i / \hat{\omega}_{i*}^2}, \tag{4.12}$$

and the sum square between

$$SSB^* = \sum_{i=1}^{K} \frac{n_i(\hat{W}_i - \hat{W}.)^2}{\hat{\omega}_{i*}^2}. \tag{4.13}$$

Lemma 4.2 *Under the null hypothesis H_0: $W_1 = W_2 = \ldots = W_K$, if for any $i = 1, 2, \ldots, K$, $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ as $\min_i n_i \to \infty$, $\int\int x^2 f_i^3(x,u)dxdu < \infty$ and if $f_i(x,u)$ is twice differentiable with respect to x and u, then SSB^* is asymptotically $\chi^2(K-1)$. In general (under the alternative), SSB^* is asymptotically non-central $\chi^2(K-1)$ with non-centrality parameter:*

$$\psi_{2*} = \frac{1}{2}\mu_*^{(\mathbf{2})'}\mathbf{B_{2*}}\mu_*^{(\mathbf{2})}, \tag{4.14}$$

where $\mu_*^{(\mathbf{2})} = (\mu_{1*}^{(2)}, \mu_{2*}^{(2)}, \ldots, \mu_{K*}^{(2)})$,

$$\mu_{i*}^{(2)} = \frac{\sqrt{\lambda_i}(e_i - \frac{\sum_{i=1}^{K}\lambda_i e_i/\omega_{i*}^2}{\sum_{i=1}^{K}\lambda_i/\omega_{i*}^2})\int\int f^2(x,u)dxdu}{\sigma\omega_{i*}}, \tag{4.15}$$

$$\mathbf{B_{2*}} = \begin{bmatrix} 1 - \frac{\lambda_1/\omega_{1*}^2}{\sum_{i=1}^{K}\lambda_i/\omega_{i*}^2} & -\frac{(\sqrt{\lambda_1}/\omega_{1*})(\sqrt{\lambda_2}/\omega_{2*})}{\sum_{i=1}^{K}\lambda_i/\omega_{i*}^2} & \cdots & -\frac{(\sqrt{\lambda_1}/\omega_{1*})(\sqrt{\lambda_K}/\omega_{K*})}{\sum_{i=1}^{K}\lambda_i/\omega_{i*}^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{(\sqrt{\lambda_1}/\omega_{1*})(\sqrt{\lambda_K}/\omega_{K*})}{\sum_{i=1}^{K}\lambda_i/\omega_{i*}^2} & -\frac{(\sqrt{\lambda_2}/\omega_{2*})(\sqrt{\lambda_K}/\omega_{K*})}{\sum_{i=1}^{K}\lambda_i/\omega_{i*}^2} & \cdots & 1 - \frac{\lambda_K/\omega_{K*}^2}{\sum_{i=1}^{K}\lambda_i/\omega_{i*}^2} \end{bmatrix}, \tag{4.16}$$

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

Proof: Replace ω_i with ω_{i*}, and then follow the proof of Theorem (2.6).

■

Define

$$A_{ij_1j_2}^* = \left(\frac{X_{ij_1} + X_{ij_2}}{2h_i^2}\right)K\left(\frac{X_{ij_1} - X_{ij_2}}{h_i}\right)K\left(\frac{U_{ij_1} - U_{ij_2}}{h_i}\right), \tag{4.17}$$

and the Sum of Square Within

$$SSW^* = \frac{\sum_{i=1}^{K} \sum \sum_{j_1 \neq j_2}^{n_i} (A^*_{ij_1j_2} - \hat{W}_i)^2 / \hat{\omega}^2_{i*}}{C_w}, \tag{4.18}$$

where C_w is given in equation (2.41).

Lemma 4.3 *For any $i = 1, 2, \ldots, K$, if $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ as $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} x^2 f_i^3(x)dx < \infty$ and if $f_i(\cdot)$ is twice differentiable, then SSW^* is asymptotically χ^2 with degrees of freedom df_w, where df_w is given in equation (2.43).*

Proof: Replace ω_i with ω_{i*} and $A_{ij_1j_2}$ with $A^*_{ij_1j_2}$, and then follow the proof of Theorem (2.8). ■

Proposition 4.4 *If for any $i = 1, 2, \ldots, K$, $n_i h_i^4 \to 0$, $n_i h_i \to \infty$ as $\min_i n_i \to \infty$, $\int_{-\infty}^{\infty} x^2 f_i^3(x)dx < \infty$ and if $f_i(\cdot)$ is twice differentiable, then under null hypothesis, $F_l^* = \frac{MSB^*}{MSW^*} = \frac{SSB^*/(K-1)}{SSW^*/df_w}$ in equation (2.48) follows asymptotically F distribution with degrees of freedom $K - 1$ and df_w. Under the alternative, F_l^* follows asymptotically non-central $F(K - 1, df_w)$ with non-centrality parameter ψ_{2*} described in equation (4.14).*

Proof: Replace ω_i with ω_{i*} and $A_{ij_1j_2}$ with $A^*_{ij_1j_2}$, and then follow the proof of Theorem (2.9). ■

Proposition 4.4 illustrates a brand new hypothesis test to evaluate the impact of new policies on some dependent variable **X**. Frankly speaking, this is an analysis of covariance (ANCOVA) model. If the test fails to be rejected, it means that all the K policies have the same effect on the dependent variable, such as the house price. That's to say, none of the new policies does any better or worse than the old one. If the test is rejected, it means that there is at least one policy that has different effects on the dependent variable. Then the next interesting question is to find which policy(s) are significantly different and by how much. This leads to the multiple comparisons type problems. Until now, there is no literature work refers the nonparametric multiple comparisons. We will leave this question to our future work.

CHAPTER 5

Conclusions and Future Works

5.1 Conclusions

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 areas 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 are either based on rank transformed techniques or performed purely by simulations. What's worse, none of the literature work in nonparametric two-way ANOVA has provided methods with theoretical support to test the main effect and interaction defined in the traditional way as the parametric ANOVA test. We propose a novel distribution-free ANOVA test and provide a nonparametric analog of traditional F test for both one-way and two-way layout. These newly constructed test statistics are not based on rank transformed techniques, but rather our newly named "kernel transformed" technique. In addition to the nonparametric ANOVA test, we also propose the nonparametric scale test, which is considered as an nonparametric analog of homogeneity of variance test in the parametric case.

Simulation results in Section 2.3 and Section 3.2 show that: (i) The kernel based nonparametric scale test is almost as powerful as the Levene's test when the samples come from Normal distributions. It significantly outperforms Levene's test when the samples come from Cauchy or Lognormal distributions. (ii) The kernel based nonparametric one-way ANOVA test is almost as powerful as the parametric one-way ANOVA test when the samples come from Normal distributions and significantly outperforms parametric one-way ANOVA test when the samples come from Cauchy or Lognormal distributions. (iii) The kernel based nonparametric one-way ANOVA test is less powerful than the kernel based nonparametric scale test in any of the three cases. (iv) 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 Normal distribution. (v) The kernel based nonparametric two-way ANOVA test of interaction and test of 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. (vi) The kernel based nonparametric two-way ANOVA test of main effects is slightly more powerful than the nonparametric two-way ANOVA test of interaction in any of three cases comparing to the corresponding parametric tests.

Kernel based nonparametric ANOVA test is more powerful than the standard ANOVA for non-normal data, especially strongly skewed and fat-tailed data. Thus, it is highly recommended if the shape of the data severely departs from mound-shaped curve.

5.2 Future Work

For the future research, an extension to other experimental design models, such as incomplete block design, in which not all the treatments occur in every block, and Latin square design, can be considered. These are the designs that are more realistic in real-world applications than the complete randomized design. In addition, the multivariate analysis of variance (MANOVA) may be considered in the future when the response variable is not a single variable, but a vector of variables instead. Also, a random effect, rather than fixed effect, nonparametric analysis of variance can be studied in the future research.

BIBLIOGRAPHY

[1] I. A. Ahmad. On asymptotic properties of an estimate of a functional of a probability density. *Scandinavian Actuarial Journal*, 3:176–181, 1976.

[2] I. A. Ahmad and M. Amezziane. Estimation of location and scale parameters based on kernel functional estimators. In *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P Hettmansperger*, pages 1–14, 2011.

[3] I. A. Ahmad, S. Leelahanon, and Q. Li. Efficient estimation of a semiparametric partially linear varying coefficient model. *Annals of Statistics*, 33:258–283, 2005.

[4] M. G. Akritas. The rank transform method in some two-factor designs. *Journal of the American Statistical Association*, 85:73–78, 1990.

[5] M. S. Bartlett. Properties of sufficiency and statistical tests. In *Proceedings of the Royal Society of London. Series A*, 1937.

[6] V. P. Bhapkar and A. P. Gore. A nonparametric test for interaction in two-way layouts. *Sankhy: The Indian Journal of Statistics, Series A*, 36:261–272, 1974.

[7] R. C. Blair, S. S. Sawilowsky, and J. J. Higgins. Limitations of the rank transform statistic in tests for interactions. *Communications in Statistics - Simulation and Computation*, 16:1133–1145, 1987.

[8] A. W. Bowman and A. Azzalini. *Applied smoothing Techniques for Data Analysis*. Oxford University Press, 1997.

[9] W. J. Conover and R. L. Iman. Rank tranformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35:124–133, 1981.

[10] A. DasGupta. *Asymtotic Theory of Statistics and Probabilty.* Springer Science+Bussiness Media, LLC, 2008.

[11] J. de Kroon and P. van de Laan. Distribution-free test procedures in two-way layouts: A concept of rank-interaction. *Statistica Neerlandica*, 35:189–213, 1981.

[12] K. L. Delucchi and A. Bostrom. Methods for analysis of skewed data distributions in psychiatric clinical studies: working with many zero values. *American Journal of Psychiatry*, 161:1159–68, 2004.

[13] T. S. Ferguson. *A course in large sample theory.* Chapman and Hall, 1st edition, 1996.

[14] F. Fischer, G. J. Miller, and M. S. Sidney. *Handbook of Public Policy Analysis: Theory, Politics, and Methods.* Taylor & Francis Group, LLC, 2007.

[15] X. Gao and M. Alvo. A nonparametric test for interaction in two-way layouts. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 33:529–543, 2005.

[16] J. Hajek. Asymptotic normality of simple linear rank statistics under alternatives. *The Annals of Mathematical Statistics*, 39:325–346, 1968.

[17] B. A. Hartlaub, A. M. Dean, and D. A. Wolfe. Rank-based test procedures for interaction in the two-way layout with one observation per cell. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 27:863–874, 1999.

[18] S. C. Hora and W. J. Conover. The F statistic in the two-way layout with rank-score transformed data. *Journal of the American Statistical Association*, 79:668–673, 1984.

[19] R. L. Iman and J. M. Davenport. New approximations to the exact distribution of the Kruskal-Wallis test statistic. *Communications in Statistics-theory and methods*, 5:1335–1348, 1976.

[20] R. L. Iman, D. Quade, and D. Alexander. Exact probability levels for the Kruskal-Wallis test. In *Selected Tables in Mathematical Statistics*, volume 3. American Mathematical Society, Providence, R.I., 1975.

[21] S. Iwata, H. Murao, and W. Wang. Nonparametric assessment of the effects of neighborhood land uses on residential house values. *Advances in Econometrics : a Research Annual*, 14:229–257, 1999.

[22] V. S. Koroljuk and Y. V. Borovskich. *Theory of U-statistics*. Kluwer Acamedic Publisher, 1994.

[23] W. H. Kruskal and W. A. Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical of Association*, 47:583–621, 1952.

[24] H. D. Lasswell. The emerging conception of the policy sciences. *Policy Sciences*, 1:3–14, 1970.

[25] H. D. Lasswell. *A pre-view of policy sciences*. Elsevier, New York, 1971.

[26] A. J. Lee. *U-statistics: theory and practice*. Marcel Dekker, 1990.

[27] H. H. Lemmer and D. J. Stoker. A distribution-free analysis of variance for the two-way classification. *South African Statistical Journal*, 1:67–712, 1967.

[28] H. Levene. *Robust tests for equality of variances*. Stanford University Press, 1960.

[29] Q. Li and J. S. Racine. *Nonparametric Econometrics*. Princeton University Press, 2007.

[30] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical statistics*, 18:50–60, 1947.

[31] K. M. Patel and D. G. Hoel. A nonparametric test for interaction in factorial experiments. *Journal of the American Statistical Association*, 68:615–620, 1973.

[32] S. T. Rachev, C. Menn, and F. J. Fabozzi. *Fat-Tailed and Skewed Asset Return Distributions*. John Wiley & Sons, Inc., 2005.

[33] W. C. Rinaman Jr. On distribution-free rank tests for two-way layouts. *Journal of the American Statistical Association*, 78:655–659, 1983.

[34] H. Scheffe. *The Analysis of Variance*. John Wiley and Sons, Inc., 1st edition, 1959.

[35] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons Inc., New York, 1992.

[36] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.

[37] G. W. Snedecor and W. G. Cochran. *Statistical Methods*. Iowa State University, 7 edition, 1980.

[38] J. Stock. Nonparametric policy analysis. *Journal of the American Statistical Association*, 84:567–575, 1989.

[39] J. Stock. Nonparametric policy analysis: An application to estimating hazardous waste cleanup benefits. In *Nonparametric and Semiparametric Methods in*

Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics, W. Barnett, J. Powell and G. Tauchen (eds). Cambridge University Press, New York, 1991.

[40] G. L. Thompson and L. P. Ammann. Efficacies of rank-transform statistics in two-way models with no interaction. *Journal of the American Statistical Association*, 84:325–330, 1989.

[41] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman & Hall, New York, 1995.

[42] K-H. Yuan and P. M. Bentler. Two simple approximations to the distributions of quadratic forms. *British Journal of Mathematical and Statistical Psychology*, 63:273–291, 2010.

VITA

Su Chen

Candidate for the Degree of

Doctor of Philosophy

Dissertation: NONPARAMETRIC ANOVA USING KERNEL METHODS

Major Field: Statistics

Biographical:

Personal Data:
Born in Wuhan, Hubei Province, China on January 11, 1986.

Education:

Completed the requirements for the degree of Doctor of Philosophy with a major of Statistics, Oklahoma State University, Stillwater, Oklahoma, USA, in July 2013.

Obtained the M.S. degree with a major of Statistics from Oklahoma State University, Stillwater, Oklahoma, USA, in December 2009.

Received the B.S. degree with a major of Statistics from Wuhan University of Technology, Wuhan, Hubei Province, China, in June 2007.

Experience:

Employed as Graduate Teaching Assistant, Graduate Research Associate and Instructor at the Department of Statistics, Oklahoma State University, Stillwater, Oklahoma, USA, from August 2007 to July 2013.

Professional Membership:

Member of the American Statistical Association
Member of the Institute of Mathematical Statistics
Member of the International Chinese Statistical Association