4-2012

A New Non-Parametric Test for Umbrella Alternatives

Gilbert Ngwa Muma
Western Michigan University, gilbert.muma@yahoo.com

A NEW NON-PARAMETRIC TEST FOR UMBRELLA ALTERNATIVES

by

Gilbert Ngwa Muma

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics
Advisor: Jeffrey Terpstra, Ph.D.

Western Michigan University
Kalamazoo, Michigan
April 2012

THE GRADUATE COLLEGE
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN

Date 12/02/2011

WE HEREBY APPROVE THE DISSERTATION SUBMITTED BY

Gilbert N Muma

ENTITLED New Nonparametric Test for Umbrella Alternatives

AS PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF Doctor of Philosophy

Statistics
(Department)

Dr Jeffrey Terpstra
Dissertation Review Committee Chair

Statistics
(Program)

Dr Magdalena Niewiadomska-Bugaj
Dissertation Review Committee Member

Dr Joseph W. McKean
Dissertation Review Committee Member

Dr Gary Neidert
Dissertation Review Committee Member

APPROVED

Dean of The Graduate College

Date April 2012

A NEW NON-PARAMETRIC TEST FOR UMBRELLA ALTERNATIVES

Gilbert Ngwa Muma, Ph.D

Western Michigan University, 2012

The Mann-Whitney statistics have commonly been used as a building block for many tests involving ordered and umbrella alternatives. Such tests are based on pair wise information that is obtained from all C_2^k pairs of samples. This dissertation introduces a new nonparametric test for testing umbrella alternatives in a completely randomized one way design. This test is based on information obtained from a subset of the C_3^k trios of samples. Unlike most existing tests for umbrella alternatives, the new test lays emphasis on the importance of testing across the peak of the umbrella, thereby rendering the new test more efficient. The test has the flexibility of testing other patterned alternatives such as the monotone ordering of location parameters (increasing or decreasing). The mean and variance of the test statistic are derived under the null hypothesis with the extensive details of the derivation included in the write-up. I also present a simplified mean and variance result that is in a practical form. Based on the derivation of the asymptotic distribution, the standardized test statistic corresponding to the new test converges in distribution to a standard normal distribution. Some numerical examples involving clinical data are analyzed. A simulation study compared power estimates of the new tests under different sample sizes and location parameters to those of seven other existing tests. The new test generally competed with all other existing tests and performed better than all the other tests in many scenarios.

ACKNOWLEDGMENTS

The printed pages of this dissertation hold far more than the culmination of years of study. These pages also reflect the relationships with many generous and inspiring people I have met since beginning my graduate work. The list is long, but I cherish each contribution to my development as a statistician:

It gives me immense pleasure to acknowledge my committee chair Dr Jeffrey Terpstra who tirelessly helped me prepare this dissertation.

To my committee members, Dr Magdalena Niewiadomska-Bugaj, Dr Joseph W. McKean and Dr Gary Neidert for their encouraging words, thoughtful criticism, time and attention during busy semesters.

To all my professors who gave me the statistical background and knowledge to build on: Dr Joshua D. Naranjo, Dr Rajib Paul, Dr Mike Stoline, Dr Gerald L. Sievers and Dr Jung Chao Wang.

To my mum and dad who are not here to witness the product of their hard work. They have been an inspiration for me throughout this journey.

To my wonderful wife who has supported and encouraged me all the way, I am proud to be your husband.

Last but not least, to my brothers and sisters who have stood by me all along, I am proud to be your brother.

Gilbert Ngwa Muma

TABLE OF CONTENTS

Table of Contents — Continued

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

In many settings a dose response-relationship need not be monotonic in the dosage and may follow an umbrella trend instead. In in-vitro mutagenicity assays, for example, experimental organisms may not survive the toxic side effects of high doses of the test agent, thereby actually reducing the number of organisms at risk of mutation and leading to a downturn (i.e umbrella pattern) in the dose-response curve. The data in Table 1 is taken from Table 6.10, of Hollander and Wolfe(1999). Plates containing Salmonella bacteria of strain TA98 were exposed to various doses of Acid Red 114. The tabled observations are the numbers of visible revertant colonies on 12 plates in the study. Suppose a researcher wants to test the hypothesis H_0 (no dose/treatment effect) against the alternative that the peak of the dose response curve for Salmonella bacteria of strain TA98 exposure to Acid Red 114 occurs at dosage level $333\mu g/ml$.

Table 1: Salmonella Bacteria Strain

Dose$\mu g/ml$					
0	100	333	1000	3333	10,000
22	60	98	60	22	23
23	59	78	82	44	21
35	54	50	59	33	25

The mean plot of this data shown in Figure 6 indicates an umbrella pattern across the groups from the first group to the last group.

1.1. Ordered and Umbrella Alternatives

Dose-response studies are frequently used to assess the relative treatment effects of increasing or decreasing dose levels of a substance in animal experiments or clinical trials. A researcher comparing multiple treatments (or samples) in a one way setting is often able to rank the treatments according to the order of magnitude of the

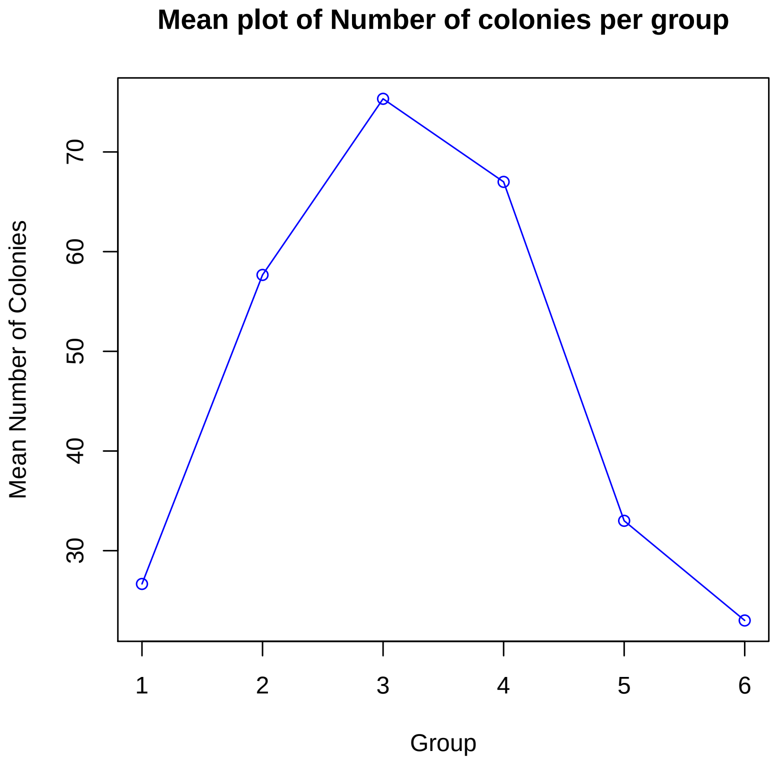

Figure 1: Mean Plot of Salmonella Bacteria Strain by Group

effect of each treatment prior to testing. The guess of this order may be based on the experience of the researcher. Within the one-way analysis of variance setting, the researcher is often concerned with detecting deviations from the null hypothesis of no treatment effect. Particular deviations of interest have included the omnibus alternative (i.e., there is a treatment effect), the ordered alternative (i.e., there is monotone treatment effect) and the umbrella alternative (i.e., there is a monotone alternative that is subject to change indirection). The conventional one way analysis of variance is not a good test for testing the ordered and umbrella alternatives. This is because the F-test is independent of the order in which the group means occur (Jonckheere, 1954). For instance, consider the effect of a drug, which is typically increasing up to a certain point p, and then it decreases. The interest of the study may be on finding the change-point group (i.e. the group where an inversion of trend of the variable under study is observed). A change point is not merely a maximum (or a minimum), but a further requirement is that the trend is monotonically increasing before group p and monotonically decreasing afterwards.

The need for understanding the order of the magnitudes of the effects of treatments has led to a rise in research and the development of distribution free tests for testing homogeneity against ordered alternatives of treatments (no treatment effects). Such testing procedures are generally based on ranks (Basso and Salmaso, 2011). Wolfe (2006) and Millen and Wolfe (2005) did an extensive review of such tests. Two main versions of tests for testing patterns have been discussed in many papers: one based on average ranks after ranking the combined data and one based on pairwise rank statistics formed by ranking only within each of the pairs of samples (Hettmansperger and Norton, 1987). That is, the tests are all based on pairwise information that is obtained from all C_2^k pairs of samples. However, tests based on ranking within each of the combination of a trio of samples has not been investigated.

Several tests have been constructed to test patterned alternatives against homogeneity of the k-samples ($H_0 : \mu_1 = \cdots = \mu_k$). The alternative varies depending on the goal of a researcher. The most common alternatives that have been studied for k-samples ($k \geq 3$) have been;

$$H_a : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \; (\mu_1 < \mu_k), \quad (1)$$

$$H_a : \mu_1 \geq \mu_2 \geq \cdots \geq \mu_k \; (\mu_1 < \mu_k), \quad (2)$$

$$H_a : \mu_1 \leq \cdots \leq \mu_{p-1} \leq \mu_p \geq \mu_{p+1} \geq \cdots \geq \mu_k. \quad (3)$$

(atleast one strict inequality)

The monotone trend and umbrella trend of the location parameters have been the most common patterns studied in the area of patterned alternatives. However, most studies involving patterned ordered alternatives have been centered around the monotone trend. There is still need for a substancial amount of effort in studying the umbrella alternative. For the sake of clarity, throughout this dissertation X_{ij} denotes the j^{th} observation in the i^{th} sample.

1.2. Literature Review

The Mann-Whitney test statistic has been the framework of many tests involving ordered alternatives. Many tests for testing umbrella alternatives have been an extension of some test for testing monotone trends.

The tests proposed by Jonckheere (1954), Mack and Wolfe (1981) and Bhat (2009) are all an extension of the Mann-Whitney test statistic. Jonckheere (1954) proposed the following test for testing $H_a : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \ (\mu_1 < \mu_k)$:

$$T = \sum_{i<j} U_{ij} \tag{4}$$

where

$$U_{ij} = \sum_{l=1}^{n_i} \sum_{k=1}^{n_j} I(X_{il} < X_{jk})$$

Notice, this is basically a combination of the 2-sample Wilcoxon which compares two samples at a time. Terpstra and Magel (2003) proposed the following test for testing the same hypothesis $H_a : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \ (\mu_1 < \mu_k)$:

$$TM = \sum_{j_1=1}^{n_1} \cdots \sum_{j_k=1}^{n_k} I(X_{1j_1} < X_{2j_2} \cdots < X_{kj_k}) \tag{5}$$

This test simultaneously tests all the groups. When $k = 2$ (two groups) this tests becomes a two sample Wilcoxon test.

Terpstra *et. al.* (2011) introduced another test for testing the monotone trend of the location parameters ($H_a : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \ (\mu_1 < \mu_k)$). This test makes use of Spearman's correlation coefficient.

$$T = \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} r_s(X_{1i_1}, X_{2i_2}, \ldots, X_{ki_k}), \tag{6}$$

where $r_s(X_{1i_1}, X_{2i_2} \ldots X_{ki_k})$ denotes the Spearman rank correlation coefficient based on the following data: $\{(1, X_{1i_1}), (2, X_{2i_2}), \ldots (k, X_{ki_k})\}$. This test proposed by Terpstra *et. al.* (2011) has an advantage over many existing tests that test monotone trends in that it provides an intuitive summary measure for the degree of association between the response variable and the treatment groups.

So much work has been carried out in developing new tests for testing ordered alternatives (monotonic treatment effects) in comparison to the few that are out the for testing umbrella alternatives.

Mack and Wolfe (1981) proposed the following statistic to assess the umbrella alternative in equation (3) when the changepoint group p is known prior to testing. That is,

$$A_p = \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} U_{ij} + \sum_{i=p}^{k-1} \sum_{j=i+1}^{k} U_{ji}, \tag{7}$$

where U_{ij} is the Mann-Whitney statistic between the i^{th} and the j^{th} samples.

Mack and Wolfe's test is an extension of the Jonckheere Terpstra test. It is based on pairwise rank statistics formed by ranking only within each of the pairs of samples. Basso and Salaso(2011) and Hettmansperger and Norton (1987) pointed out that no comparisons are made in the Mack-Wolfe test between samples preceding the known peak and those following it. Hettmansperger and Norton (1987) stated the absence of across-the-peak comparison as a problem.

Bhat (2009) developed the following k-sample rank test for testing umbrella alternatives.

$$A = \sum_{i=1}^{p-1} a_i U_{i,i+1} + \sum_{i=p}^{k-1} a_i U_{i+1,i} \tag{8}$$

where $a_1, \ldots, a_{k-1}$ are real constants to be chosen suitably and U_{ij} is the two sample Mann-Whitney U-statistic for the i^{th} and j^{th} samples. This test compares successive groups. For instance,if $p = 4$ and $k = 6$, then $A = U_{12}+U_{23}+U_{34}+U_{54}+U_{65}$. The fact that Bhat's test does not test across groups can cause very misleading results. If a pattern has multiple peaks, Bhat's test would test significant if any peak is tested indicating the presence of an umbrella pattern which is not the case. Consider the following data and it's mean plot.

Table 2: Simulated Data

(Group 1)	(Group 2)	(Group 3)	(Group 4)	(Group 5)
-0.2645469	5.3564944	0.498346	1.6867112	-0.3965792
1.1782195	4.7889641	0.7456039	0.6361922	0.240107
-1.3325297	4.9146766	0.2423505	2.762041	1.2912549

Using Bhat's test to test for a peak at Group4 ($p = 4$), the test will give the following results;

A=25, Mean= 18, Variance= 16.5, Test Statistic = 1.7233, **p value = 0.0424**

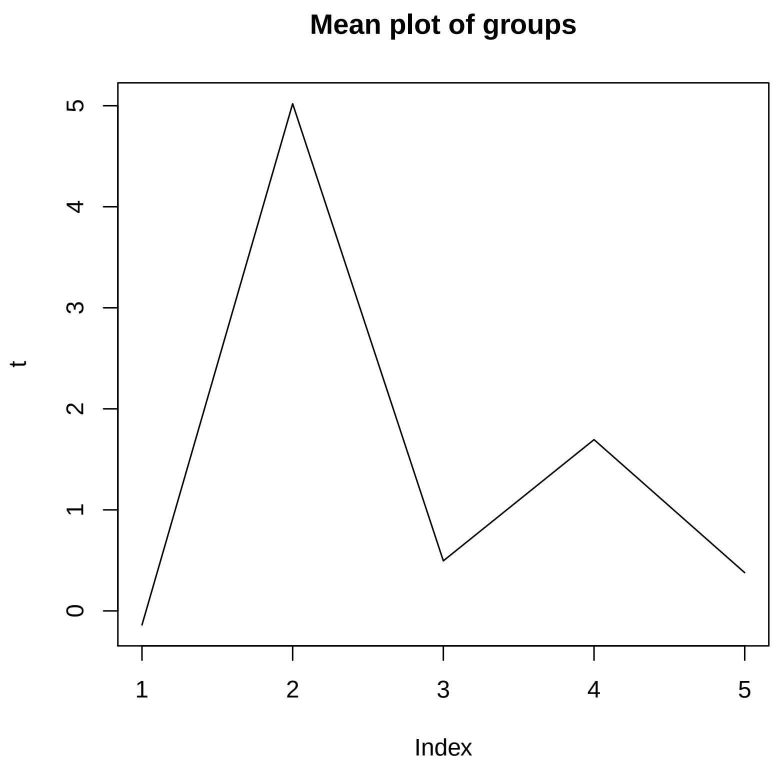

Figure 2: Mean Plot of Data in Table 2

The results of Bhat's test indicates an umbrella pattern with a peak at group 4 which is far from reality. This is the set back of not testing across all groups. Secondly, like the Mack Wolfe test, Bhat's test does not test across the peak. However, the Mack and Wolfe's test rejects H_o. This disproves the conclusion made in Bhat (2009) that Bhat's test is preferable to the test proposed by Mack and Wolfe for testing umbrella alternatives. As noticed, Bhat's test indicates the presence of peaks, regardless of whether the peak is the highest peak or the lowest peak.

Salman (2010) proposed the following test for umbrella alternatives

$$T_A = \sum_{i_1=1}^{p-1} \sum_{i_3=p+1}^{k} A_{i_1,p,i_3} \tag{9}$$

where

$$A_{i_1,p,i_3} = \sum_{j_1=1}^{n_{i1}} \sum_{j_2=1}^{n_p} \sum_{j_3=1}^{n_{i3}} I(X_{i_1 j_1} < X_{p j_2} > X_{i_3 j_3})$$

Salman does compare across the peak in his test but does not test the increasing monotone alternative on the left of the peak and the decreasing monotone alternative on the right of the peak. Salman's statistic is formed by ranking within three groups at a time.

Hettmansperger and Norton (1987) constructed a test based on average ranks after ranking the combined data. Their test statistic is stated as

$$V = \sum_{j=1}^{k} \lambda_j (c_j - \overline{c}_w) \overline{R}_j \tag{10}$$

where the set of constants $c_1 \dots c_k$ specifies the pattern to be detected, $\lambda_j = n_j/N$, $N = \sum n_j$, $\bar{c}_w = \sum \lambda_j c_j$ and $\overline{R_j}$ is the average rank of the j^th group.

Basso and Salmaso (2011) introduced a permutation test for testing umbrella alternates. Their test is suitable for small sample sizes but the computation can be very time consuming for larger sample sizes. Just like the Mack and Wolfe's test, the pemutation test does not compare across the peak.

A new distribution free test for testing umbrella alternatives in the case where the peak is known prior to testing is introduced in this dissertation. The flexibility of this test in terms of how it can test other patterned alternatives is discussed. Unlike other tests for testing umbrella alternatives that compare only two groups at a time ($X_{i_1 j_1} < X_{i_2 j_2}$) this new test compares 3 groups at a time e.g. ($X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}$). That is, the new test being proposed is based on information obtained from the subset of the C_3^k trios of samples. Therefore there should be a minimum of three groups for this test to be applied.

The test takes into consideration the importance of testing across the peak as well as both sides of the peak. A detailed description of how the test is carried out is discussed. Some special cases in which the peak appears in varying positions are also discussed. The second part of the research derives the mean and variance of the test. Simplified formulas for the mean and variances of the test for both the balanced and unbalanced cases. An extensive proof of the mean and variance of the test is outlined. In the third chapter the asymptotic distribution of the test is derived, with the help of theorems that relate to multisample U-statistics and other asymptotic theorems. Some real life data are analyzed and their results are compared to other existing tests for umbrella alternatives. An extensive simulation study is carried out in which I compare powers from results of the proposed test to some selected existing tests.

1.3. New Statistics

1.3.1. Notations Used Throughout the Dissertation

p refers to the peak group

k is the number of groups

n_p is the sample size of the peak group

n_i is the sample size of the nth group

$S_a^b(c)$ is a set of c-tuplets, where each tuplet corresponds to c ordered integers ranging from a to b. There are C_c^{b-a+1} elements in this set.

For example:

$S_1^4(3) = \{(1,2,3),(1,2,4),(1,3,4),(2,3,4)\}$

$S_1^4(5) = \emptyset$. (null set)

1.3.2. Hypothesis and Assumptions

This dissertation introduces a new class of nonparametric tests of homogeneity between the k treatment groups. That is,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_{p-1} = \mu_p = \mu_{p+1} = \cdots = \mu_k \tag{11}$$

against the umbrella alternative in equation (3)

Assumptions

- The N random variables $X_{i,j}$, $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots n_i$ are mutually independent

- For each fixed $i \in \{1, 2, \ldots, k\}$, the n_i random variables $X_{1,1}, X_{1,2} \ldots X_{k,n_i}$ are a random sample from a continuous distribution with distribution function F.

1.3.3. Design

Let $\{X_{ij}\}$, $i = 1, 2, ...k$, $j = 1, 2, ..., n_i$, denote the $N = \sum_{i=1}^{k} n_i$ random variables corresponding to k random samples from a completely randomized design (CRD). The cumulative distribution function (CDF) for X_{ij} is denoted as $F_i(x); i = 1, 2, ...k$. Thus the general design is a one way layout as follows:

g_1	g_2	g_3	.	.	g_p	.	.	g_k
X_{11}	X_{21}	X_{31}	.	.	X_{p1}	.	.	X_{k1}
X_{12}	X_{22}	X_{32}	.	.	X_{p2}	.	.	X_{k2}
X_{13}	X_{23}	X_{33}	.	.	X_{p3}	.	.	X_{k3}
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
X_{1n_1}	X_{2n_2}	X_{3n_3}	.	.	X_{pn_p}	.	.	X_{kn_k}

The groups may have equal or unequal sample sizes (balanced or unbalanced).

1.3.4. New Test

This dissertation is proposing the following test statistic:

$$T = T_L + T_A + T_R, \tag{12}$$

This is a general case, where

$$T_L = \sum_{i \in S_1^p(3)} L_{i_1,i_2,i_3} \quad \text{and } L_{i_1,i_2,i_3} = \sum_{j_1=1}^{n_{i1}} \sum_{j_2=1}^{n_{i2}} \sum_{j_3=1}^{n_{i3}} I(X_{i_1 j_1} \leq X_{i_2 j_2} \leq X_{i_3 j_3})$$

$$T_A = \sum_{i_1=1}^{p-1} \sum_{i_3=p+1}^{k} A_{i_1,p,i_3} \text{ and } A_{i_1,p,i_3} = \sum_{j_1=1}^{n_{i1}} \sum_{j_2=1}^{n_p} \sum_{j_3=1}^{n_{i3}} I(X_{i_1 j_1} \leq X_{p j_2} \geq X_{i_3 j_3})$$

$$T_R = \sum_{i \in S_p^k(3)} R_{i_1,i_2,i_3} \quad \text{and } R_{i_1,i_2,i_3} = \sum_{j_1=1}^{n_{i1}} \sum_{j_2=1}^{n_{i2}} \sum_{j_3=1}^{p_{i3}} I(X_{i_1 j_1} \geq X_{i_2 j_2} \geq X_{i_3 j_3})$$

In this dissertation, the indicator is 1 only if there is atleast 1 strict inequality.

1.3.5. Flexibility of New Test

The proposed test is very flexible in that it can be used to test multiple ordered patterns. In addition to the umbrella alternative stated above, this test has the ability to test the following types of alternative hypothesis:

- $H_a : \mu_1 < \mu_2 > \mu_3$

  This hypothesis is testing an umbrella pattern with peak at 2 with three groups. The corresponding test statistic for this hypothesis would be,

$$T = T_A. \tag{13}$$

When $p = 2$ and $k = 3$, the new test is the same as the test proposed by Salman (2010).

- $H_a : \mu_1 < \mu_2 < \cdots < \mu_k$

  This hypothesis is the same as that being tested by the Jonckheere-Terpstra test. This hypothesis is testing for a peak at the last group (i.e. $p = k$). The corresponding test for this hypothesis is

$$T = T_L. \tag{14}$$

- $H_a : \mu_1 > \mu_2 > \cdots > \mu_k$

  In this case $p = 1$. This is also the hypothesis being tested by the Jonckheere-Terpstra test. The corresponding test for this special case is,

$$T = T_R. \tag{15}$$

- $H_a : \mu_1 < \mu_2 > \mu_3 > \cdots > \mu_k$

  The corresponding test for this hypothesis is

$$T = T_A + T_R. \tag{16}$$

- $H_a : \mu_1 < \cdots < \mu_{k-2} < \mu_{k-1} > \mu_k$

The corresponding test for this is

$$T = T_L + T_A. \tag{17}$$

CHAPTER 2. EXAMPLES

2.1. General Example

In this section an illustration of how T is calculated is presented. Consider the following unbalanced one-way table with six groups ($n_1 = n_2 = n_3, n_4 = n_5 = n_6 = 1$). suppose we want to test for an umbrella alternative with the peak at group 3.

Group	1	2	3	4	5	6
	a	c	e	g	i	j
	b	d	f			

The following derivations detail the calculations of T for this special case. Note that there are atleast two groups on the left of the peak which means, the monotone increase on the left (T_L) has to be tested, and there are atleast two groups on the right of the peak as well which means the monotone derease on the right (T_R) has to be tested. There is atleast one group on both sides of the peak which means there needs to be a test across the peak (T_A).

$$
\begin{aligned}
T &= T_L + T_A + T_R \\
&= L_{123} + A_{134} + A_{234} + A_{135} + A_{235} + A_{136} + A_{236} + R_{345} + R_{346} + R_{356} + R_{456} \\
T_L &= L_{123} \\
L_{123} &= I(a \leq c \leq e) + I(a \leq c \leq f) + I(a \leq d \leq e) + I(a \leq d \leq f) \\
&\quad + I(b \leq c \leq e) + I(b \leq c \leq f) + I(b \leq d \leq e) + I(b \leq d \leq f)
\end{aligned}
$$

$$
\begin{aligned}
T_A &= A_{134} + A_{234} + A_{135} + A_{235} + A_{136} + A_{236} \\
A_{134} &= I(a \le e \ge g) + I(a \le f \ge g) + I(b \le e \ge g) + I(b \le f \ge g) \\
A_{234} &= I(c \le e \ge g) + I(c \le f \ge g) + I(d \le e \ge g) + I(d \le f \ge g) \\
A_{135} &= I(a \le e \ge i) + I(a \le f \ge i) + I(b \le e \ge i) + I(b \le f \ge i) \\
A_{235} &= I(c \le e \ge i) + I(c \le f \ge i) + I(d \le e \ge i) + I(d \le f \ge j) \\
A_{136} &= I(a \le e \ge j) + I(a \le f \ge i) + I(b \le e \ge j) + I(b \le f \ge j) \\
A_{236} &= I(c \le e \ge j) + I(c \le f \ge j) + I(d \le e \ge j) + I(d \le f \ge j)
\end{aligned}
$$

$$
\begin{aligned}
T_R &= R_{345} + R_{346} + R_{356} + R_{456} \\
R_{345} &= I(e \ge g \ge i) + I(f \ge g \ge i) \\
R_{346} &= I(e \ge g \ge j) + I(f \ge g \ge j) \\
R_{356} &= I(e \ge i \ge j) + I(f \ge i \ge j) \\
R_{456} &= I(g \ge i \ge j)
\end{aligned}
$$

T is the sum of all the indicators in T_L, $T_A$and T_R.

2.2. Numerical Example 1

Consider the data in Table 1. Although the mean plot for this data indicates $p = 3$, the H_a is tested at $p = 4$ $(1000/ml)$ so that the results can be compared to the results analyzed by Hollander and Wolfe (1999).

$$H_o \; : \; \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$

$$H_a \; : \; \mu_1 < \mu_2 < \mu_3 < \mu_4 > \mu_5 > \mu_6$$

Based on the hypothesis,

$$p = 4,\ k = 6 \text{ and } n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = 0$$

$$\begin{aligned}
T &= T_L + T_A + T_R \\
T &= L_{123} + L_{124} + L_{134} + L_{234} + A_{145} + A_{146} + A_{245} + A_{246} + A_{345} + A_{346} + R_{456} \\
T &= 18 + 24 + 12 + 3 + 27 + 27 + 24 + 24 + 12 + 12 + 21 \\
&= 204
\end{aligned}$$

Using the formulas from chapter 3, (i.e. equation (19) and equation (21)) the mean and the variance for this example is given as follows

Mean

$$\begin{aligned}
E(T) &= \frac{3^3}{6}\left(\binom{4}{3} + 2(4-1)(6-4) + \binom{6-4+1}{3}\right) \\
&= \frac{27}{6}(4 + 12 + 1) \\
&= 76.5
\end{aligned}$$

Variance

$$Var(T) =$$

$$\begin{aligned}
& \frac{2}{360}\left\{39\binom{4}{5}3^5+\binom{4}{4}(48(3^5)+27(3^4))+\binom{4}{3}(9(3^5)+12(3^4)+4(3^3))\right\} \\
+ \quad & \frac{2}{360}\left\{39\binom{6-4}{5}3^5+\binom{6-4}{4}(48(3^5)+27(3^4))+\binom{6-4}{3}(9(3^5)+12(3^4)+4(3^3))\right\} \\
+ \quad & \frac{(4-1)(6-4)3^3}{45}\left\{(6-1)(3^2)+4(4-1)(6-4)3^2+\frac{5(6-1)3+2(3)}{4}+1\right\} \\
+ \quad & 2\frac{(6-4)}{360}\left\{48\binom{4-1}{3}3^5+\binom{4-1}{2}(40(3^5)+10(3^4))\right\} \\
+ \quad & 2\frac{(4-1)}{360}\left\{48\binom{6-4}{3}3^5+\binom{6-4}{2}(40(3^5)+10(3^4))\right\} \\
+ \quad & 2\binom{4-1}{2}\binom{6-4}{2}\frac{3^5}{45}
\end{aligned}$$

$$\begin{aligned}
= \quad & \frac{2}{360}(0+13851+13068)+\frac{2}{360}(0+0+3267)+\frac{162}{45}(45+216+20.25+1) \\
+ \quad & \frac{4}{360}(11664+31590)+\frac{6}{360}(0+10530)+\frac{1458}{45} \\
= \quad & 149.55+18.15+1016.1+480.6+175.5+32.4 \\
= \quad & 1872.3
\end{aligned}$$

Note that

$$\binom{4}{5} = 0, \binom{2}{5} = 0, \binom{2}{4} = 0 \text{ and } \binom{2}{3} = 0$$

$$z = \frac{204 - 76.5}{\sqrt{1872.3}}$$

$$= 2.9466$$

$$\text{p value} = 0.0016$$

The data is analyzed using the test statistics in equations (7), (9), (8), (10) and the Kruskal Wallis test.

Table 3: Example 1 Results for Other Umbrella Tests

Test	T	Mean	Variance	Test Statistic	pvalue
LAR	204	76.5	1872.3	2.9466	0.0016
MW	69	40.5	96.75	2.8975	0.0019
Sal	126	54	1016.1	2.2587	0.0120
Bhat	35	22.5	17.25	3.0096	0.0013
HN	1.6759	0.7500	0.1613	2.3057	0.0106
KW				13.4121	0.0198

Although the plot of the data indicates the peak at group 3, all the tests for umbrella alternatives in Table 3 are consistent. They all reject H_0 in favor of H_a.

2.3. Numerical Example 2

Anogenital distance is the distance from the anus to the genitalia, the base of the penis or vagina. Consider the following data taken from Bradstreet (1992).

Suppose a researcher wants to test for a dose related change in anogenital distance . This researcher expects anogenital distance to decrease with an increase in dose concentration. Forty pregnant female rats were assigned randomly to one of four treatments (10 rats/treatment) including a Vehicle Control and graded oral doses (50, 100, 200 mg/kg) of an investigational compound. Dosing was performed once daily on Days 16 through 19 of gestation. On Day 20, the animals were sacrificed. The anogenital distance and sex of each fetus was recorded. Average anogenital distances for each sex in a litter are shown in the data below.

a) For males, is there a dose related change in anogenital distance?

Table 4: Anogenital Distance

Dose		Litter									
(mg/kg)	Sex	1	2	3	4	5	6	7	8	9	10
0	M	2.66	2.66	2.34	2.34	2.68	2.44	2.43	2.79	2.80	2.61
0	F	1.11	1.18	1.13	1.09	1.26	1.10	1.15	1.27	1.21	1.17
50	M	1.76	1.56	2.28	1.74	1.95	1.82	1.91	1.83	2.24	1.91
50	F	0.98	1.06	1.40	1.01	1.14	1.02	1.13	1.20	1.27	1.14
100	M	1.58	1.56	1.48	1.55	1.51	1.92	1.48	1.48	1.72	1.52
100	F	1.11	1.10	1.18	1.12	1.13	1.32	1.20	1.05	1.28	1.23
200	M	1.28	1.30	1.24	1.30	1.26	1.42	1.20	1.20	1.39	(a)
200	F	1.01	1.08	1.01	1.06	1.12	1.38	1.11	1.15	1.27	(a)

(a) All fetuses found dead

Given the goal of the study, we want to test the following hypothesis,

$$H_o \ : \ \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_a \ : \ \mu_1 > \mu_2 > \mu_3 > \mu 4$$

Based on the hypothesis, T is the same as in Equation (15).

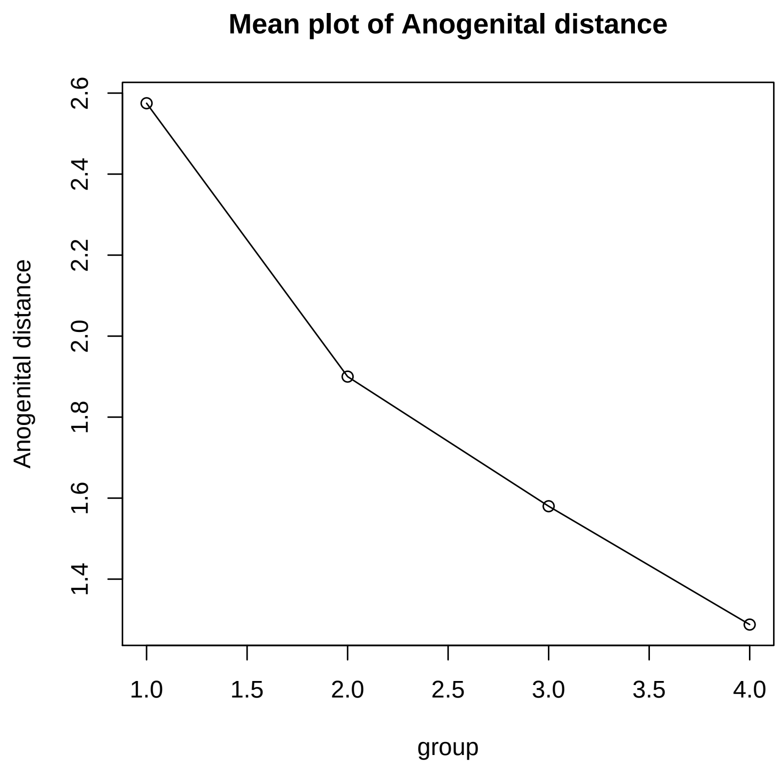

Figure 3: Mean Plot of Anogenital Distance

Note that $p = 1$, and $k = 4$. Therefore

$$T = T_R = \sum_{i \in S_1^4(3)} R_{i_1 i_2 i_3},$$

$$\begin{aligned} & where \\ & R_{i_1 i_2 i_3} = \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} (X_{i_1 j_1} > X_{i_2 j_2} > X_{i_3 j_3}) \end{aligned}$$

$$S_1^4(3) = \{(1,2,3),(1,2,4),(1,3,4),(2,3,4)\}\text{, and}$$

$$
\begin{aligned}
T = T_R &= R_{123} + R_{124} + R_{134} + R_{234} \\
&= 910 + 900 + 900 + 819 \\
&= 3529
\end{aligned}
$$

Using the formulas from chapter 3, (i.e. Equation (18) and Equation (20)) the mean and the variance for this example is given as follows

Mean

$$
\begin{aligned}
\sum_{i \in S_1^4(3)} n_{i_1} n_{i_2} n_{i_3} \frac{1}{3!} &= \frac{1}{3!}(n_1 n_2 n_3 + n_1 n_2 n_4 + n_1 n_3 n_4 + n_2 n_3 n_4) \\
&= \frac{1}{3!}(10.10.10 + 10.10.9 + 10.10.10.9 + 10.10.9) \\
&= 616.6667
\end{aligned}
$$

Variance

$$
Var(T) = \qquad Var(T_R) = \frac{2}{360}(39r_1 + r_2 + r_3), \text{ where}
$$

$$
r_1 = \sum_{i \in S_1^4(5)} n_{i_1} n_{i_2} n_{i_3} n_{i_4} n_{i_5} = 0
$$

$$
\begin{aligned}
r_2 &= \sum_{i \in S_1^4(4)} n_{i_1} n_{i_2} n_{i_3} n_{i_4} (9(n_{i_1} + n_{i_4}) + 15(n_{i_2} + n_{i_3}) + 27) \\
&= n_1 n_2 n_3 n_4 (9(n_1 + n_4) + 15(n_2 + n_3) + 27) \\
&= 10 \times 10 \times 10 \times 9 \times (9(10 + 9) + 15(10 + 10) + 27) \\
&= 4482000, \text{and}
\end{aligned}
$$

$$
\begin{aligned}
r_3 &= \sum_{i \in S_1^4(3)} n_{i_1} n_{i_2} n_{i_3} (4 n_{i_1} n_{i_2} + 4 n_{i_2} n_{i_3} + n_{i_1} n_{i_3} + 5(n_{i_1} + n_{i_3}) + 2 n_{i_2} + 4) \\
r_3 &= 10.10.10(4.10.10 + 4.10.10 + 10.10 + 5(10 + 10) + 2.10 + 4) \\
&= 10.10.9(4.10.10 + 4.10.9 + 10.9 + 5(10 + 9) + 2.10 + 4) \\
&= 10.10.9(4.10.10 + 4.10.9 + 10.9 + 5(10 + 9) + 2.10 + 4) \\
&= 10.10.9(4.10.10 + 4.10.9 + 10.9 + 5(10 + 9) + 2.10 + 4) \\
&= 1024000 + 872100 + 872100 + 872100 \\
&= 3640300 \ .
\end{aligned}
$$

thus

$$
Var(T) = \frac{2}{360}(39(0) + 4482000 + 36403000) = 45123.89
$$

$$
Z = \frac{3529 - 616.6667}{\sqrt{45123.89}} = 13.71,
$$

and

p value = 0.0000 .

Table 5: Example 2a: Dose-Response in Males

Test	T	Mean	Variance	Test Statistic	pvalue
LAR	3529	616.6667	45123.89	13.71	0.0000
MW	560.5	285	1591.667	6.91	0.0000
Bhat	281	145	183.333	10.0443	0.0000
HN	-0.0244	-1.2229	0.0429	5.7878	0.0000
KW				34.3467	0.0000

b) For females, is there a dose related change in anogenital distance?

Table 6: Example 2b: Dose-Response in Females

Test	T	Mean	Variance	Test Statistic	pvalue
LAR	802	616.6667	45123.89	0.8725	0.1915
MW	316.5	285	1591.667	0.7896	0.2149
Bhat	164	145	183.333	1.4032	0.0803
HN	-1.0581	-1.2229	0.0429	0.7960	0.2130
KW				2.1810	0.5357

The results in Table 5 and Table 6 differ between males and females. Based on the results, H_o is rejected for males and is not rejected for females at $\alpha = 0$. The results are also consistent for all the tests. This indicates there is a dose-response relationship in males but none in females.

2.4. Numerical Example 3

This data is from Pedersen *et. al.* (2008). This was a study carried out to describe a Doppler waveform index representing the hepatic vein flow velocity pattern and to examine its relationship to the degree of hepatic fibrosis. The study examined the a consecutive series of patients who underwent percutaneous liver needle biopsy and sonographic examination that included recording of hepatic vein Doppler waveform 5 days before the biopsy. The study comprised 66 patients (37 females and 29 males).

After completion of the study, all tracings were evaluated blindly without knowledge about the pathology report, in one session, and hepatic vein waveform index (HVWI) was calculated as (maximum velocity - minimum velocity)/(maximum velocity). Two pathologists blinded to the Doppler findings examined the biopsy specimens and graded fibrosis. In case of disagreement, they both re-examined the specimens to reach a consensus. Fibrosis was graded as: no fibrosis (group 1); mild fibrosis (group

2); moderate fibrosis (group 3); severe fibrosis (group 4) and cirrhosis (group 5).

In order to test the relationship between degree of fibrosis and HVWI, suppose the study wants to test the hypothesis that there is an umbrella pattern with peak at group 2 (mild fibrosis).

Table 7: Hepatic Vein Waveform Index

HVWI				
No fibrosis	Mild fibrosis	Moderate fibrosis	Severe fibrosis	Cirrhosis
1.7917	1.7124	1.5570	1.6726	1.4182
1.7090	1.6694	1.4017	1.2893	1.4017
1.6661	1.6033	1.3587	0.9950	1.3025
1.6099	1.5934	1.3256	0.4959	1.2000
1.5669	1.5438	1.1702		1.1372
1.5537	1.4612	0.8165		0.9289
1.5008	1.3620			0.6743
1.3884	1.2860			0.6711
1.3884	1.2860			0.6248
1.3289	1.2595			0.6280
1.2496	1.2198			0.5983
0.9950	1.2033			0.5354
0.2511	1.1702			0.4694
0.0594	1.1306			0.4397
	0.9983			0.4430
	0.9983			0.4033
	0.8496			0.3967
	0.7504			0.3769
	0.6413			0.2511
	0.6017			
	0.5554			
	0.5587			
	0.1388			

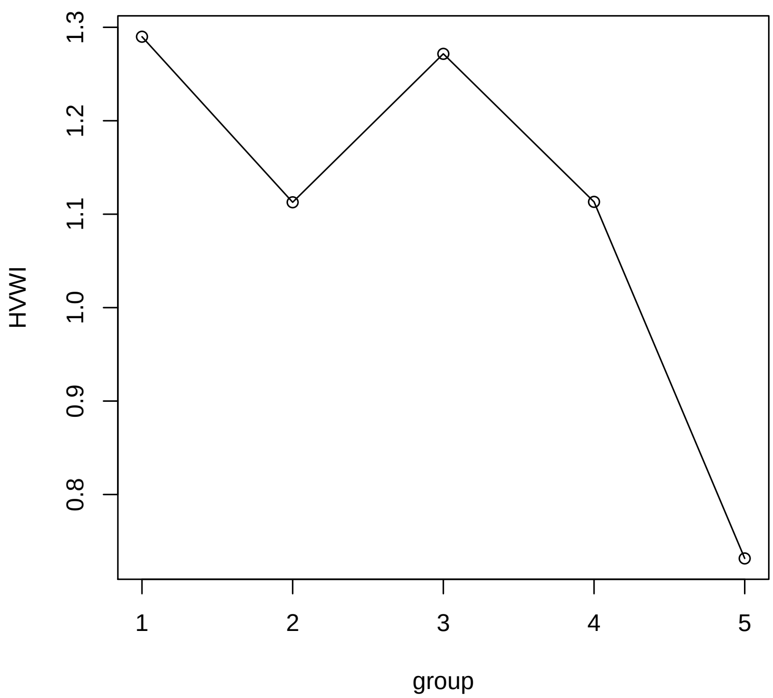

Figure 4: Mean Plot of Data in Table 7

$$
\begin{aligned}
H_o &: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \\
H_a &: \mu_1 < \mu_2 > \mu_3 > \mu_4 > \mu_5 \\
& \quad p = 2, k = 5 \\
& \quad n_1 = 14, n_2 = 23, n_3 = 6, n_4 = 4, n_5 = 19
\end{aligned}
$$

$$
\begin{aligned}
T &= T_L + T_A + T_R \\
&= 0 + T_A + T_R \\
&= (A_{123} + A_{124} + A_{125}) + (R_{234} + R_{235} + R_{245} + R_{345}) \\
&= (418 + 280 + 1808) + (115 + 828 + 509 + 176) \\
&= 4144
\end{aligned}
$$

Mean

$$E(T) = E(T_A) + E(T_R)$$

$$\begin{aligned}
E(T) &= \sum_{i_1=1}^{1}\sum_{i_3=3}^{5} n_{i_1} n_2 n_{i_3} \frac{1}{3} + \sum_{i=S_2^5(3)} n_{i_1} n_{i_2} n_{i_3} \frac{1}{3!} \\
&= \frac{n_1 n_2 n_3}{3} + \frac{n_1 n_2 n_4}{3} + \frac{n_1 n_2 n_5}{3} + \frac{n_2 n_3 n_4}{6} + \frac{n_2 n_3 n_5}{6} + \frac{n_2 n_4 n_5}{6} + \frac{n_3 n_4 n_5}{6} \\
&= \frac{14 \times 23 \times 6}{3} + \frac{14 \times 23 \times 4}{3} + \frac{14 \times 23 \times 19}{3} + \frac{23 \times 6 \times 4}{6} \\
&\quad + \frac{23 \times 6 \times 19}{6} + \frac{23 \times 4 \times 19}{6} + \frac{6 \times 4 \times 19}{6} \\
&= 3112.6667 + 896.3333 \\
&= 4009
\end{aligned}$$

Variance

$$Var(T) = Var(T_A) + Var(T_R) + 2Cov(T_A, T_R)$$

$$Var(T_A) = \frac{n_1^* n_2 n_3^*}{45}\left(n_2(n_1^* + n_3^*) + 4n_1^* n_3^* + \frac{5(n_1^* + n_3^*) + 2n_2}{4} + 1\right)$$

$$n_1^* = \sum_{i=1}^{1} n_i = n_1 = 14$$

$$n_3^* = \sum_{i=3}^{5} n_i = n_3 + n_4 + n_5 = 6 + 4 + 19 = 29$$

$$n_p = n_2 = 23$$

$$\begin{aligned} Var(T_A) &= \frac{14 \times 23 \times 29}{45}\left(23(14+29) + 4 \times 14 \times 29 + \frac{5(14+29) + 2(23)}{4} + 1\right) \\ &= \frac{9338}{45}(989 + 1624 + 65.25 + 1) \\ &= 555974.1444 \end{aligned}$$

$$Var(T_R) = \frac{2}{360}(39r_1 + r_2 + r_3)$$

$$r_1 = \sum_{i \in S_2^5(5)} n_{i_1} n_{i_2} n_{i_3} n_{i_4} n_{i_5} = 0,$$

$$r_2 = \sum_{i \in S_2^5(4)} n_{i_1} n_{i_2} n_{i_3} n_{i_4} (9(n_{i_1} + n_{i_4}) + 15(n_{i_2} + n_{i_3}) + 27)$$

$$= n_2 n_3 n_4 n_5 (9(n_2 + n_5) + 15(n_3 + n_4) + 27)$$

$$= 23.6.4.19(9(23 + 19) + 15(6 + 4) + 27)$$

$$= 10488(378 + 150 + 27)$$

$$= 5820840$$

$$\begin{aligned}
r_3 &= \sum_{i \in S_2^5(3)} n_{i_1} n_{i_2} n_{i_3} (4n_{i_1} n_{i_2} + 4n_{i_2} n_{i_3} + n_{i_1} n_{i_3} + 5(n_{i_1} + n_{i_3}) + 2n_{i_2} + 4) \\
&= n_2 n_3 n_4 (4n_2 n_3 + 4n_3 n_4 + n_2 n_4 + 5(n_2 + n_4) + 2n_3 + 4) \\
&+ n_2 n_3 n_5 (4n_2 n_3 + 4n_3 n_5 + n_2 n_5 + 5(n_2 + n_5) + 2n_3 + 4) \\
&+ n_2 n_4 n_5 (4n_2 n_4 + 4n_4 n_5 + n_2 n_5 + 5(n_2 + n_5) + 2n_4 + 4) \\
&+ n_3 n_4 n_5 (4n_3 n_4 + 4n_4 n_5 + n_3 n_5 + 5(n_3 + n_5) + 2n_4 + 4) \\
&= 23 \times 6 \times 4(4 \times 23 \times 6 + 4 \times 6 \times 4 + 23 \times 4 + 5(23 + 4) + 2 \times 6 + 4) \\
&+ 23 \times 6 \times 19(4 \times 23 \times 6 + 4 \times 6 \times 19 + 23 \times 19 + 5(23 + 19) + 2 \times 6 + 4) \\
&+ 23 \times 4 \times 19(4 \times 23 \times 4 + 4 \times 4 \times 19 + 23 \times 19 + 5(23 + 19) + 2 \times 4 + 4) \\
&+ 6 \times 4 \times 19(4 \times 6 \times 4 + 4 \times 4 \times 19 + 6 \times 19 + 5(6 + 19) + 2 \times 4 + 4) \\
&= 491832 + 4381362 + 2326588 + 296856 \\
&= 7496638
\end{aligned}$$

Therefore,

$$\begin{aligned}
Var(T_R) &= \frac{2}{360} (39(0) + 5820840 + 7496638) \\
&= 73985.9889
\end{aligned}$$

$$Cov(T_A, T_R) = \frac{n_2 n_l}{360}\left\{48 \sum_{i \in S_3^5(3)} n_{i_1} n_{i_2} n_{i_3} + \sum_{i \in S_3^5(2)} n_{i_1} n_{i_2}(8n_2 + 16n_{i_1} + 16n_{i_2} + 10)\right\}$$

$$n_l = n_1 + \cdots + n_{p-1} = n_1 = 14$$

$$\sum_{S_3^5(3)} n_{i_1} n_{i_2} n_{i_3} = n_3 n_4 n_5$$

$$= 6 \times 4 \times 19$$

$$= 456$$

$$\sum_{i \in S_3^5(2)} n_{i_1} n_{i_2}(8n_p + 16n_{i_1} + 16n_{i_2} + 10)$$

$$\begin{aligned}
&= n_3 n_4(8n_p + 16n_3 + 16n_4 + 10) \\
&+ n_3 n_5(8n_p + 16n_3 + 16n_5 + 10) \\
&+ n_4 n_5(8n_p + 16n_4 + 16n_5 + 10)
\end{aligned}$$

$$\begin{aligned}
&= 6 \times 4(8 \times 23 + 16 \times 6 + 16 \times 4 + 10) \\
&+ 6 \times 19(8 \times 23 + 16 \times 6 + 16 \times 19 + 10) \\
&+ 4 \times 19(8 \times 23 + 16 \times 4 + +16 \times 19 + 10) \\
&= 8496 + 67716 + 42712 \\
&= 118924
\end{aligned}$$

$$Cov(T_A, T_R) = \frac{23 \times 14}{360}(48 \times 456 + 118924)$$

$$= 125948.511$$

$$Var(T) = 555974.1444 + 73985.9889 + 2 \times 125948.511$$
$$= 881857.1553$$

The test statistics can thus be calulated as,

$$Z = \frac{4144 - 4009}{\sqrt{881857.1553}} = 0.1438$$

Table 8: Example 3: Results of Other Umbrella Tests

Test	T	Mean	Variance	Test Statistic	pvalue
LAR	4144	4009	881857.2	0.1438	0.4428
MW	707	601.5	6018.917	1.3599	0.0869
Sal	2506	3112.7	555974.1	-0.8136	0.7921
Bhat	232.5	280	1692.667	-1.1545	0.8759
HN	-0.9781	-1.5999	0.0477	2.8461	0.0022
KW				15.0037	0.0047

The results in Table 8 are contradictory. Although the mean plot in Figure 4, does not indicate a peak at group 2, HN rejects H_0 at $\alpha = 0.05$ (i.e. indicating peak is at group 2) while MW is marginally significant (p value=0.0869). The LAR, Sal and Bhat test however do not rejects H_0.

CHAPTER 3. THEORY

The first part of this chapter describes the distributional properties of T under the null hypothesis. The expectation and variance of T under the null hypothesis are stated along with a few worked results. The asymptotic null distribution of T is also stated. The second part of the chapter gives detailed proofs of the expectation, variance and the asymptotic distribution of T. The derivation of a few preliminary results are also included. The fundamental elements of the proofs deal with probabilities such as $P(X_1 < X_2 < X_3)$, $P(X_1 < X_2 > X_3)$ and $P(X_1 < X_2 < X_3,\ X_1 < X_4 < X_5)$. In the derivation of the asymptotic distribution of T, this treats T as a multisample U-statistic and makes use of classical theorems for multisample U-statistics.

3.1. Main Results Under the Null Hypothesis

3.1.1. T is Distribution Free

Under H_0 all observations (i.e. $\{X_{ij}\}$) are IID F, where F is a continuous distribution function. Thus, the possibilities of ties between the Xs have zero probability. Moreover, this implies that T is distribution free under H_o. See, for example, Terpstra and Magel (2003).

3.1.2. Expectation of T

The mean of T under H_0 is presented in Theorem 3.1.

Theorem 3.1: Under H_0, the expectation of T is given by,

$$E(T) = \sum_{i \in S_1^p(3)} n_{i_1} n_{i_2} n_{i_3} \frac{1}{3!} + \sum_{i_1=1}^{p-1} \sum_{i_3=p+1}^{k} n_{i_1} n_p n_{i_3} \frac{1}{3} + \sum_{i \in S_p^k(3)} n_{i_1} n_{i_2} n_{i_3} \frac{1}{3!}. \tag{18}$$

See Section 3.2.2 for the detailed proof of this result. That said, when $n_1 = n_2 = \cdots = n_k = n$, equation (18) reduces to,

$$E(T) = \frac{n^3}{6}\left(\binom{p}{3} + 2(p-1)(k-p) + \binom{k-p+1}{3}\right), \tag{19}$$

where $\binom{n}{x}$ denotes the number of ways to choose x objects from n objects.

3.1.3. Variance of T

The variance of T under H_0 is presented in Theorem 3.2. In what follows we make use of the following quatities:

$$\begin{aligned}
l_1 &= 39 \sum_{i \in S_1^p(5)} n_{i_1} n_{i_2} n_{i_3} n_{i_4} n_{i_5}, \\
l_2 &= \sum_{i \in S_1^p(4)} n_{i_1} n_{i_2} n_{i_3} n_{i_4} (9(n_{i_1} + n_{i_4}) + 15(n_{i_2} + n_{i_3}) + 27), \\
l_3 &= \sum_{i \in S_1^p(3)} n_{i_1} n_{i_2} n_{i_3} (4n_{i_1} n_{i_2} + 4n_{i_2} n_{i_3} + n_{i_1} n_{i_3} + 5(n_{i_1} + n_{i_3}) + 2n_{i_2} + 4), \\
r_1 &= 39 \sum_{i \in S_p^k(5)} n_{i_1} n_{i_2} n_{i_3} n_{i_4} n_{i_5}, \\
r_2 &= \sum_{i \in S_p^k(4)} n_{i_1} n_{i_2} n_{i_3} n_{i_4} (9(n_{i_1} + n_{i_4}) + 15(n_{i_2} + n_{i_3}) + 27), \\
r_3 &= \sum_{i \in S_p^k(3)} n_{i_1} n_{i_2} n_{i_3} (4n_{i_1} n_{i_2} + 4n_{i_2} n_{i_3} + n_{i_1} n_{i_3} + 5(n_{i_1} + n_{i_3}) + 2n_{i_2} + 4),
\end{aligned}$$

$$
\begin{aligned}
n_l &= n_1 + \cdots + n_{p-1}, \\
n_r &= n_{p+1} + \cdots + n_k, \\
a &= n_p(n_l + n_r) + 4n_l n_r + \frac{5(n_l + n_r) + 2n_p}{4} + 1, \\
u_1 &= 48 \sum_{i \in S_1^{p-1}(3)} n_{i_1} n_{i_2} n_{i_3}, \\
u_2 &= \sum_{i \in S_1^{p-1}(2)} n_{i_1} n_{i_2} (8n_p + 16n_{i_1} + 16n_{i_2} + 10), \\
v_1 &= 48 \sum_{i \in S_{p+1}^{k}(3)} n_{i_1} n_{i_2} n_{i_3}, \\
v_2 &= \sum_{i \in S_{p+1}^{k}(2)} n_{i_1} n_{i_2} (8n_p + 16n_{i_1} + 16n_{i_2} + 10), \\
w_1 &= \sum_{i \in S_1^{p-1}(2)} n_{i_1} n_{i_2}, \text{and} \\
w_2 &= \sum_{i \in S_{p+1}^{k}(2)} n_{i_1} n_{i_2}
\end{aligned}
$$

Recall that

$$\sum_{i \in S_a^b(c)} = 0 \quad \text{if} \quad b - a + 1 < c$$

.

Theorem 3.2: Under H_0, the variance of T is given by,

$$Var(T) = \frac{2}{360}(l_1 + l_2 + l_3 + r_1 + r_2 + r_3) + \frac{n_p}{45}(n_l n_r a + w_1 w_2)$$

$$+\frac{n_p}{360}(n_r(u_1 + u_2) + n_l(v_1 + v_2)). \tag{20}$$

Section 3.2.3 contains the proof of this result. That said, when $n_1 = n_2 = \cdots = n_k = n$, equation (20) reduces to,

$$\begin{aligned} V(T) = {} & \frac{2}{360}\left\{39\binom{p}{5}n^5 + \binom{p}{4}(48n^5 + 27n^4) + \binom{p}{3}(9n^5 + 12n^4 + 4n^3)\right\} \\ + & \frac{2}{360}\left\{39\binom{k-p}{5}n^5 + \binom{k-p}{4}(48n^5 + 27n^4) + \binom{k-p}{3}(9n^5 + 12n^4 + 4n^3)\right\} \\ + & \frac{(p-1)(k-p)n^3}{45}\left\{(k-1)n^2 + 4(p-1)(k-p)n^2 + \frac{5(k-1)n + 2n}{4} + 1\right\} \\ + & 2\frac{(k-p)}{360}\left\{48\binom{p-1}{3}n^5 + \binom{p-1}{2}(40n^5 + 10n^4)\right\} \\ + & 2\frac{(p-1)}{360}\left\{48\binom{k-p}{3}n^5 + \binom{k-p}{2}(40n^5 + 10n^4)\right\} \\ + & 2\binom{p-1}{2}\binom{k-p}{2}\frac{n^5}{45}. \end{aligned} \tag{21}$$

3.1.4. Examples of Computed Means and Variances

Table 9 contains the calculated means and variances of some randomly selected sample sizes and peaks. In the case where $n_1 = n_2 = \cdots = n_k = n$, the mean when $p = 1$ is equal to the mean when $p = k$ and the variance when $p = 1$ is equal to the variance when $p = k$.

Table 9: Some Computed Means and Variances

n_1	n_2	n_3	n_4	n_5	n_6	Peak	Mean	Variance
2	2	2	-	-	-	1	1.3333	2.8444
2	2	2	-	-	-	2	2.6667	5.5111
2	2	2	-	-	-	3	1.3333	2.8444
3	3	3	-	-	-	1	4.5	18.1500
5	5	5	-	-	-	2	41.6667	461.1111
10	10	10	-	-	-	1	166.6667	5688.889
10	10	10	-	-	-	2	333.3333	14022.22
2	2	2	2	-	-	1	5.3333	22.3111
3	3	3	3	-	-	3	22.5	211.95
7	7	7	7	-	-	3	285.8333	13561.84
9	9	9	9	-	-	4	486	30354.75
4	4	4	4	4	-	1	106.6667	2476.089
4	4	4	4	4	-	2	106.6667	2779.022
11	11	11	11	11	-	4	2218.333	411375.1
9	9	9	9	9	9	3	2065.5	428830.2
1	2	3	-	-	-	1	1	2.1
2	3	7	-	-	-	2	14	90.3
19	14	9	-	-	-	3	399	25416.3
5	10	11	12	-	-	2	603.3333	45728.22
2	3	1	3	-	-	1	6.5	30.75
1	6	7	9	-	-	3	154	4240.6
4	8	13	20	-	-	3	1109.333	114460.1
3	2	4	1	3	-	4	17.3333	199.4222
6	14	16	18	19	-	1	4767.333	1287338
6	7	9	12	14	-	2	1211	156730
17	6	8	16	2	-	4	1229.333	126704.6
16	6	11	3	19	-	5	1728.833	221884
2	2	5	6	3	2	6	107.3333	2641.511
10	2	4	5	13	14	1	1373	156319.4
7	3	6	4	13	5	3	641.3333	62408.16

3.1.5. Asymptotic Distribution of T

Next consider the asymptotic distribution of T. T can essentially be viewed as a linear combination of U-statistics. It follows from Theorem 4.5.1 of Koroljuk and Borovskich that T is asymptotically normal. This result is expressed in Theorem 3.3

Theorem 3.3: Under H_0,

$$Z = \frac{T - E(T)}{\sqrt{V(T)}} \xrightarrow{d} N(0,1) \tag{22}$$

as $n = min(n_1, n_2, \ldots n_k) \to \infty$ provided, for $i = 1, 2, \ldots, k$, that $\frac{n_i}{N} \to \lambda_i \in (0,1)$, where

$$N = \sum_{i=1}^{k} n_i.$$

3.2. Proofs

3.2.1. Preliminary Results

This section presents some important results which will be useful in deriving the expectation and the variance of T. To begin, the following two results are used for the expection of T:

$P(X_1 < X_2 < X_3) = \frac{1}{6}$ and

$P(X_1 < X_2 > X_3) = \frac{1}{3}$

For example,

$$\begin{aligned}
P(X_1 < X_2 < X_3) &= \int_{-\infty}^{\infty}\int_{x_1}^{\infty}\int_{x_2}^{\infty} f(x_1)f(x_2)f(x_3)dx_3dx_2dx_1 \\
&= \int_{-\infty}^{\infty}\int_{x_1}^{\infty} f(x_1)f(x_2)\left(\int_{x_2}^{\infty} f(x_3)dx_3\right)dx_2dx_1 \\
&= \int_{-\infty}^{\infty} f(x_1)\left(\int_{x_1}^{\infty} f(x_2)\left(1-F(x_2)\right)dx_2\right)dx_1
\end{aligned}$$

Next, let $u = 1 - F(x_2)$ so $du = -f(x_2)dx_2$ and do a u-substitution to get,

$$\begin{aligned}
P(X_1 < X_2 < X_3) &= \int_{-\infty}^{\infty} f(x_1)\left(-\int_{1-F(x_1)}^{1-F(\infty)} udu\right)dx_1 \\
&= \int_{-\infty}^{\infty} f(x_1)\left(\int_{0}^{1-F(x_1)} udu\right)dx_1 \\
&= \int_{-\infty}^{\infty} f(x_1)\left(\frac{1}{2}u^2\Big|_0^{1-F(x_1)}\right)dx_1 \\
&= \int_{-\infty}^{\infty} f(x_1)\frac{1}{2}(1-F(x_1))^2dx_1.
\end{aligned}$$

Now let $u = 1 - F(x_1)$ so $du = -f(x_1)dx_1$ and do another u-substitution to get,

$$\begin{aligned}
P(X_1 < X_2 < X_3) &= -\int_{1}^{0} \frac{1}{2}u^2du \\
&= \frac{1}{2}\int_{0}^{1} u^2\partial u = \frac{1}{2}.\frac{1}{3}u^3\Big|_0^1 \\
&= \frac{1}{2}.\frac{1}{3}(1) = \frac{1}{3!} = \frac{1}{6}.
\end{aligned}$$

Following the same logics as the previous derivation we get,

$$
\begin{aligned}
P\left(X_1 < X_2 > X_3\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{x_2}\int_{-\infty}^{x_2} f(x_1)f(x_2)f(x_3)dx_2dx_1dx_3 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{x_2} f(x_2)f(x_1)\left(\int_{-\infty}^{x_2} f(x_3)dx_3\right)dx_2dx_1 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{x_2} f(x_2)f(x_1)\left(F(x_2) - F(-\infty)\right)dx_2dx_1 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{x_2} f(x_2)f(x_1)F(x_2)dx_2dx_1 \\
&= \int_{-\infty}^{\infty} f(x_2)F(x_2)\left(\int_{-\infty}^{x_2} f(x_1)dx_1\right)dx_2 \\
&= \int_{-\infty}^{\infty} f(x_2)F(x_2)\left(F(x_2)\right)dx_2 \\
&= \int_{-\infty}^{\infty} f(x_2)F^2(x_2)dx_2 \\
&= \int_0^1 u^2du = \frac{1}{3}u^3\Big|_0^1 = \frac{1}{3}.
\end{aligned}
$$

Next we consider results needed for the variance of T. Covariance results such as $Cov(I(X_1 < X_2 < X_3)I(X_1 < X_4 < X_5))$ will be required.

Note first that,

$$Cov(I(X_1 < X_2 < X_3)I(X_1 < X_4 < X_5)) =$$
$$E\left[I(X_1 < X_2 < X_3)I(X_1 < X_4 < X_5)\right] - E\left[I(X_1 < X_2 < X_3)\right]E\left[I(X_1 < X_4 < X_5)\right] =$$
$$P(X_1 < X_2 < X_3, X_1 < X_4 < X_5) - P(X_1 < X_2 < X_3)P(X_1 < X_4 < X_5).$$

The marginal expectations have already been discussed. Thus, all that is needed is the various joint expectations. These are given below.

$$E(I(X_1 < X_2 < X_3), I(X_1 < X_4 < X_5)) = \frac{1}{20}$$

$$E(I(X_1 < X_2 < X_3), I(X_2 < X_4 < X_5)) = \frac{1}{40}$$

$$E(I(X_1 < X_3 < X_4), I(X_2 < X_3 < X_5)) = \frac{1}{30}$$

$$E(I(X_1 < X_4 < X_5), I(X_2 < X_3 < X_4)) = \frac{1}{40}$$

$$E(I(X_1 < X_2 < X_3), I(X_3 < X_4 < X_5)) = \frac{1}{120}$$

$$E(I(X_1 < X_2 < X_5), I(X_3 < X_4 < X_5)) = \frac{1}{20}$$

$$E(I(X_1 < X_2 < X_3), I(X_1 < X_2 < X_4)) = \frac{1}{12}$$

$$E(I(X_1 < X_2 < X_3), I(X_1 < X_3 < X_4)) = \frac{1}{24}$$

$$E(I(X_1 < X_2 < X_4), I(X_1 < X_3 < X_4)) = \frac{1}{12}$$

$$E(I(X_1 < X_2 < X_3), I(X_2 < X_3 < X_4)) = \frac{1}{24}$$

$$E(I(X_1 < X_2 < X_4), I(X_2 < X_3 < X_4)) = \frac{1}{24}$$

$$E(I(X_1 < X_3 < X_4), I(X_2 < X_3 < X_4)) = \frac{1}{12}$$

$$E(I(X_1 < X_3 > X_4), I(X_2 < X_3 > X_5)) = \frac{1}{5}$$

$$E(I(X_1 < X_2 > X_3), I(X_1 < X_2 > X_4)) = \frac{1}{4}$$

$$E(I(X_1 < X_3 > X_4), I(X_2 < X_3 > X_4)) = \frac{1}{4}$$

$$E(I(X_1 < X_2 < X_3), I(X_1 < X_4 > X_5)) = \frac{3}{40}$$

$$E(I(X_1 < X_2 < X_3), I(X_2 < X_4 > X_5)) = \frac{7}{120}$$

$$E(I(X_1 < X_2 < X_3), I(X_3 < X_4 > X_5)) = \frac{1}{30}$$

$$E(I(X_1 < X_2 < X_4), I(X_3 < X_4 > X_5)) = \frac{1}{10}$$

$$E(I(X_1 < X_2 < X_3), I(X_1 < X_3 > X_4)) = \frac{1}{8}$$

Let us only consider the first expectation, which we can write as

$$E\left[I(X_{i_1} < X_{i_2} < X_{i_3})I(X_{i'_1} < X_{i'_2} < X_{i'_3})\right],$$

where $i_2 \neq i'_2$ and $i_3 \neq i'_3$. Since this is equivalent to $P(X_{i_1} < X_{i_2} < X_{i_3}), X_{i'_1} < X_{i'_2} < X_{i'_3})$ we get,

$$\begin{aligned}
& P\left(X_{i_1} < X_{i_2} < X_{i_3} \cap X_{i_1} < X_{i'_2} < X_{i'_3}\right) \\
= & \int_{-\infty}^{\infty}\int_{x_{i_1}}^{\infty}\int_{x_{i_2}}^{\infty}\int_{x_{i'_1}}^{\infty}\int_{x_{i'_2}}^{\infty} f(x_{i_1})f(x_{i_2})f(x_{i_3})f(x_{i'_2})f(x_{i'_3})dx_{i_1}dx_{i_2}dx_{i_3}dx_{i'_2}dx_{i'_3} \\
= & \int_{-\infty}^{\infty}\int_{x_{i_1}}^{\infty}\int_{x_{i_2}}^{\infty}\int_{x_{i'_1}}^{\infty} f(x_{i_1})f(x_{i_2})f(x_{i_3})f(x_{i'_2})dx_{i_1}dx_{i_2}dx_{i_3}dx_{i'_2}\left[\int_{x_{i'_2}}^{\infty} f(x_{i'_3})dx_{i'_3}\right] \\
= & \int_{-\infty}^{\infty}\int_{x_{i_1}}^{\infty}\int_{x_{i_2}}^{\infty}\int_{x_{i'_1}}^{\infty} f(x_{i_1})f(x_{i_2})f(x_{i_3})f(x_{i'_2})(1 - F(x_{i'_2}))dx_{i_1}dx_{i_2}dx_{i_3}dx_{i'_2} \\
= & \int_{-\infty}^{\infty}\int_{x_{i_1}}^{\infty}\int_{x_{i_2}}^{\infty} f(x_{i_1})f(x_{i_2})f(x_{i_3})dx_{i_1}dx_{i_2}dx_{i_3}\left[\int_{x_{i_1}}^{\infty} f(x_{i'_2})(1 - F(x_{i'_2}))dx_{i'_2}\right] \\
= & \int_{-\infty}^{\infty}\int_{x_{i_1}}^{\infty}\int_{x_{i_2}}^{\infty} f(x_{i_1})f(x_{i_2})f(x_{i_3})\frac{1}{2}(1 - F(x_{i_1}))^2 dx_{i_1}dx_{i_2}dx_{i_3} \\
= & \int_{-\infty}^{\infty}\int_{x_{i_1}}^{\infty} f(x_{i_1})f(x_{i_2})\frac{1}{2}(1 - F(x_{i_1}))^2 dx_{i_1}dx_{i_2}\left[\int_{x_{i_2}}^{\infty} f(x_{i_3})dx_{i_3}\right]
\end{aligned}$$

$$= \int_{-\infty}^{\infty}\int_{x_{i_1}}^{\infty} f(x_{i_1})f(x_{i_2})\frac{1}{2}(1-F(x_{i_1}))^2(1-F(x_{i_2}))dx_{i_1}dx_{i_2}$$

$$= \int_{-\infty}^{\infty} f(x_{i_1})\frac{1}{2}(1-F(x_{i_1}))^2 dx_{i_1}\left[\int_{x_{i_1}}^{\infty} f(x_{i_2})(1-F(x_{i_2}))dx_{i_2}\right]$$

$$= \int_{-\infty}^{\infty} f(x_{i_1})\frac{1}{2}(1-F(x_{i_1}))^2\frac{1}{2}(1-F(x_{i_1}))^2 dx_{i_1}$$

$$= \int_{-\infty}^{\infty} f(x_{i_1})\frac{1}{4}(1-F(x_{i_1}))^4 dx_{i_1}$$

$$= \frac{1}{4}\int_{1}^{0} -u^4 du = \frac{1}{20}$$

Now, from a previous result, we know that

$$P\left(X_{i_1} \leq X_{i_2} \leq X_{i_3}\right) = \frac{1}{6} \text{ and}$$

$$P\left(X_{i_1} \leq X_{i'_2} \leq X_{i'_3}\right) = \frac{1}{6}$$

Therefore, it follows that,

$$Cov(I(X_{i_1} \leq X_{i_2} \leq X_{i_3}), I(X_{i_1} \leq X_{i'_2} \leq X_{i'_3})) = \frac{1}{20} - \frac{1}{6}\frac{1}{6} = \frac{1}{45}.$$

Finally, we conclude this section with some more examples of some results that will be of use. They easily follow from the fact that $u = F(x)$ has a uniform distribution on the interval $(0, 1)$.

$$E\left((1 - F(x))^4\right) = \frac{1}{5}$$

$$E\left((1 - F(x))^2(1 - F^2(x))\right) = \frac{3}{10}$$

$$E\left((1 - F(x))^2\right) = \frac{1}{3}$$

$$E\left((1 - F^2(x))^2\right) = \frac{8}{15}$$

$$E\left((1 - F^2(x))\right) = \frac{2}{3}$$

$$E\left(F^2(x)(1 - F(x))^2\right) = \frac{1}{30}$$

$$E\left(F^3(x)(1 - F(x))\right) = \frac{1}{20}$$

$$E\left(F(x) - F^2(x)\right) = \frac{1}{6}$$

$$E\left(F^4(x)\right) = \frac{1}{5}$$

3.2.2. Proof of Expectation of T

The expectation of T can be expressed as

$$E(T) = E(T_L) + E(T_A) + E(T_R), \text{where}$$

$$E(T_L) = E\left(\sum_{i \in S_1^p(3)} L_{i_1,i_2,i_3}\right) = \sum_{i \in S_1^p(3)} \sum_{j_1=1}^{n_{i1}} \sum_{j_2=1}^{n_{i2}} \sum_{j_3=1}^{n_{i3}} E\left(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3})\right)$$

$$= \sum_{i \in S_1^p(3)} n_{i_1} n_{i_2} n_{i_3} \frac{1}{3!},$$

$$E(T_A) = E\left(\sum_{i_1=1}^{p-1} \sum_{i_3=p+1}^{k} A_{i_1,p,i_3}\right) = \sum_{i_1=1}^{p-1} \sum_{i_3=p+1}^{k} \sum_{j_1=1}^{n_{i1}} \sum_{j_2=1}^{n_p} \sum_{j_3=1}^{n_{i3}} E\left(I(X_{i_1 j_1} < X_{p j_2} > X_{i_3 j_3})\right)$$

$$= \sum_{i_1=1}^{p-1} \sum_{i_3=p+1}^{k} n_{i_1} n_p n_{i_3} \frac{1}{3}, \text{and}$$

$$E(T_R) = E\left(\sum_{i \in S_p^k(3)} R_{i_1,i_2,i_3}\right) = \sum_{i \in S_p^k(3)} \sum_{j_1=1}^{n_{i1}} \sum_{j_2=1}^{n_{i2}} \sum_{j_3=1}^{n_{i3}} E\left(I(X_{i_1 j_1} > X_{i_2 j_2} > X_{i_3 j_3})\right)$$

$$= \sum_{i \in S_p^k(3)} n_{i_1} n_{i_2} n_{i_3} \frac{1}{3!}.$$

3.2.3. Proof of Variance of T

To begin, note that

$$\begin{aligned} Var(T) &= V(T_L + T_A + T_R) \\ &= V(T_L) + V(T_A) + V(T_R) + 2Cov(T_L, T_A) + 2Cov(T_A, T_R) \\ &+ 2Cov(T_L, T_R) \end{aligned}$$

In what follows, we treat each of the six quantities individually.

In this section, each component of the $Var(T)$ is derived seperately.

Variance of T_L

Here,

$$\begin{aligned} V(T_L) &= Cov(T_L, T_L) = Cov\left(\sum_{i \in S_1^p(3)} L_{i_1 i_2 i_3}, \sum_{i' \in S_1^p(3)} L_{i_1 i_2 i_3}\right) \\ &= \sum_{i \in S_1^p(3)} Var(L_{i_1 i_2 i_3}) + 2 \sum_{i \in S_1^p(3)} \sum_{i' \in S_1^p(3)} Cov(L_{i_1 i_2 i_3}, L_{i'_1 i'_2 i'_3}) \qquad (23) \end{aligned}$$

In what follows we solve for $Var(L_{i_1 i_2 i_3})$ and $Cov(L_{i_1 i_2 i_3}, L_{i'_1 i'_2 i'_3})$

It is important to note that $i_1 < i_2 < i_3$ for i and $i'_1 < i'_2 < i'_3$ for i'.

In order to derive the variance of T_L, it is important to first identify all the possible cases for i and i'. The list below shows there are basically 14 cases. For example, in case 3, $i_2 = i_{\prime 1}$ and all other subscripts are distinct. That is, there is only one tied subscript and it occurs in the second position of i and the first position of i'. We note that the case where $i = i'$ is not included as this is the same as $i_2 = i'$. Hence, the factor of 2 for the covariance terms.

Case	i_1	i_2	i_3	i'_1	i'_2	i'_3	*Ties*
1	$-$	$-$	$-$	$-$	$-$	$-$	0
2	x	$-$	$-$	x	$-$	$-$	1
3	$-$	x	$-$	x	$-$	$-$	1
4	$-$	x	$-$	$-$	x	$-$	1
5	$-$	x	$-$	$-$	$-$	x	1
6	$-$	$-$	x	x	$-$	$-$	1
7	$-$	$-$	x	$-$	$-$	x	1
8	x	x	$-$	x	x	$-$	2
9	x	$-$	x	x	x	$-$	2
10	x	$-$	x	x	$-$	x	2
11	$-$	x	x	x	x	$-$	2
12	$-$	x	x	x	$-$	x	2
13	$-$	x	x	$-$	x	x	2
14	x	x	x	x	x	x	3

The covariances for each of these cases of ties are derived and summed below to obtain the variance of T_L. Note that case 14 corresponds to $Var(L_{i_1 i_2 i_3})$, while cases 1 through 13 pertains to $Cov(L_{i_1 i_2 i_3}, L_{i'_1 i'_2 i'_3})$.

Case 1

Under case 1, all of the covariance terms are equal to zero since there are no tied subscripts. That is, there are six distinct groups. Under Case 1, $i \neq i'$. For example,

$Cov(L_{123}, L_{456})$ falls under this case. Therefore $Cov(T_{i_1 i_2 i_3}, T_{i'_1 i'_2 i'_3}) = 0$.

Case 2

Under Case 2, $i_1 = i'_1$, $i_2 \neq i'_2$ and $i_3 \neq i'_3$. For example , $Cov(T_{L_{123}}, T_{L_{145}})$ falls under this case. To begin, note that

$$\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i_1 i'_2 i'_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), \sum_{j'_1=1}^{n_{i_1}} \sum_{j'_2=1}^{n_{i'_2}} \sum_{j'_3=1}^{n_{i'_3}} I(X_{i_1 j'_1} < X_{i'_2 j'_2} < X_{i'_3 j'_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} \sum_{j'_1=1}^{n_{i_1}} \sum_{j'_2=1}^{n_{i'_2}} \sum_{j'_3=1}^{n_{i'_3}} Cov\left(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i'_2 j'_2} < X_{i'_3 j'_3})\right)
\end{aligned}$$

Now, since $i_1 = i'_1$, $i_2 \neq i'_2$ and $i_3 \neq i'_3$ it follows that j_1 may or may not be equal to j'_1. Hence, we get the following 2 subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

2.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i'_2 j'_2} < X_{i'_3 j'_3})) = 0$

2.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j_1} < X_{i'_2 j'_2} < X_{i'_3 j'_3}) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$

Lastly multiplying 2.2 by the corresponding number of terms leads to the desired result is thus;

$$Cov(L_{i_1 i_2 i_3}, L_{i_1 i'_2 i'_3}) = n_{i_1} n_{i_2} n_{i_3} n_{i'_2} n_{i'_3} \left(-\frac{1}{360}\right)$$

Case 3

Under Case 3, $i_1 \neq i'_1$, $i_2 = i'_1$ and $i_3 \neq i'_3$. For example , $Cov(L_{123}, L_{245})$ falls under this case. To begin, note that

$$\begin{aligned} & Cov(L_{i_1 i_2 i_3}, L_{i_2 i'_2 i'_3}) \\ = \; & Cov\left(\sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), \sum_{j'_1=1}^{n_{i_2}} \sum_{j'_2=1}^{n_{i'_2}} \sum_{j'_3=1}^{n_{i'_3}} I(X_{i_2 j'_2} < X_{i'_2 j'_2} < X_{i'_3 j'_3})\right) \\ = \; & \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} \sum_{j'_1=1}^{n_{i_1}} \sum_{j'_2=1}^{n_{i'_2}} \sum_{j'_3=1}^{n_{i'_3}} Cov\left(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_2 j'_2} < X_{i'_2 j'_2} < X_{i'_3 j'_3})\right) \end{aligned}$$

Now, since $i_1 \neq i'_1$, $i_2 = i'_1$ and $i_3 \neq i'_3$ it follows that j_2 may or may not be equal to j'_2. Hence, we get the following 2 subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	x	-	x	-	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

3.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_2 j'_2} < X_{i'_2 j'_2} < X_{i'_3 j'_3})) = 0$

3.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_2 j_2} < X_{i'_2 j'_2} < X_{i'_3 j'_3}) = \frac{3}{120} - \frac{1}{3!}\frac{1}{3!} = -\frac{1}{360}$

Multiplying 3.2 by the corresponding number of terms leads to the desired result is thus;

$$Cov(L_{i_1 i_2 i_3}, L_{i_2 i'_2 i'_3}) = n_{i_1} n_{i_2} n_{i_3} n_{i'_2} n_{i'_3} \left(\frac{1}{45} \right)$$

Case 4

Under Case 4, $i_1 \neq i'_1$, $i_2 = i'_2$ and $i_3 \neq i'_3$. For example, $Cov(L_{134}, L_{235})$ falls under this case. To begin, note that

$$\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i'_1 i_2 i'_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), \sum_{j'_1=1}^{n_{i'_1}} \sum_{j'_2=1}^{n_{i_2}} \sum_{j'_3=1}^{n_{i'_3}} I(X_{i'_1 j'_1} < X_{i_2 j'_2} < X_{i'_3 j'_3}) \right) \\
&= \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} \sum_{j'_1=1}^{n_{i_1}} \sum_{j'_2=1}^{n_{i_2}} \sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i'_1 j'_1} < X_{i_2 j'_2} < X_{i'_3 j'_3})).
\end{aligned}$$

Now, since $i_1 \neq i'_1$, $i_2 = i'_2$ and $i_3 \neq i'_3$, it follows that j_2 may or may not be equal to j'_2 . Hence, we get the following 2 subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	x	-	-	x	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

4.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i'_1 j'_1} < X_{i_2 j'_2} < X_{i'_3 j'_3})) = 0,$

4.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i'_1 j'_1} < X_{i_2 j_2} < X_{i'_3 j'_3})) = \frac{4}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{180},$

Lastly, multiplying 4.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$Cov(L_{i_1i_2i_3}, L_{i'_1i_2i'_3} = n_{i_1}n_{i_2}n_{i_3}n_{i'_1}n_{i'_3}\left(\frac{1}{180}\right)$$

Case 5

Under Case 5, $i_1 \neq i'_1$, $i_2 = i'_3$ and $i_3 \neq i'_3$. For example, $Cov(L_{145}, L_{234})$ falls under this case. To begin, note that

$$\begin{aligned} &Cov(L_{i_1i_2i_3}, L_{i'_1i'_2i_2}) \\ &= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i'_1j'_1} < X_{i'_2j'_2} < X_{i_2j'_2})\right) \\ &= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i'_1,j'_1} < X_{i'_2j'_2} < X_{i_2j'_2})). \end{aligned}$$

Now, since $i_1 \neq i'_1$, $i_2 = i'_3$ and $i_3 \neq i'_3$, it follow that j_2 may or may not be equal to j'_3 Hence, we get the following 2 subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	x	-	-	-	x	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

5.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i'_1j'_1} < X_{i'_2j'_2} < X_{i_2j'_2})) = 0,$

5.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i'_1j'_1} < X_{i'_2j'_2} < X_{i_2j_2})) = \frac{3}{120} - \frac{1}{3!}\frac{1}{3!} = -\frac{1}{360},$

Lastly, multiplying 5.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$L_{i_1i_2i_3}, L_{i'_1,i'_2,i_2}) = n_{i_1}n_{i_2}n_{i_3}n_{i'_1}n_{i'_2}\left(-\frac{1}{360}\right).$$

Case 6

Under Case 6, $i_1 \neq i'_1$, $i_2 \neq i'_2$ and $i_3 = i'_1$. For example, $Cov(L_{123}, L_{345})$ falls under this case. To begin, note that

$$\begin{aligned}
&Cov(L_{i_1i_2i_3}, L_{i_3,i'_2,i'_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_3j'_3} < X_{i'_2j'_2} < X_{i'_3j'_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_3j'_3} < X_{i'_2j'_2} < X_{i'_3j'_3})).
\end{aligned}$$

Now, since$i_1 \neq i'_1$, $i_2 \neq i'_2$ and $i_3 = i'_1$, it follows that j_3 may or may not be equal to j'_1. Hence, we get the following two subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	-	x	x	-	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

6.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_3j'_3} < X_{i'_2j'_2} < X_{i'_3j'_3})) = 0,$

6.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_3j_3} < X_{i'_2j'_2} < X_{i'_3j'_3})) = \frac{1}{120} - \frac{1}{3!}\frac{1}{3!} = -\frac{7}{360}$

Lastly, multiplying 6.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$Cov(L_{i_1i_2i_3}, L_{i_3,i'_2,i'_3}) = n_{i_1}n_{i_2}n_{i_3}n_{i'_2}n_{i'_3}\left(-\frac{7}{360}\right)$$

Case 7

Under Case 7, $i_1 \neq i'_1$, $i_2 \neq i'_2$ and $i_3 = i'_3$. For example, $Cov(L_{123}, L_{345})$ falls under this case. To begin, note that

$$
\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i'_1, i'_2, i_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i'_1 j'_1} < X_{i'_2 j'_2} < X_{i_3 j'_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i'_1 j'_1} < X_{i'_2 j'_2} < X_{i_3 j'_3})).
\end{aligned}
$$

Now, since $i_1 \neq i'_1$, $i_2 \neq i'_2$ and $i_3 = i'_3$, it follows that j_3 may or may not be equal to j'_3 . Hence, we get the following two subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	-	x	-	-	x	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

7.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i'_1 j'_1} < X_{i'_2 j'_2} < X_{i_3 j'_3})) = 0,$

7.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i'_1 j'_1} < X_{i'_2 j'_2} < X_{i_3 j_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$

Lastly, multiplying 7.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$Cov(L_{i_1 i_2 i_3}, L_{i'_1, i'_2, i_3}) = n_{i_1} n_{i_2} n_{i_3} n_{i'_1} n_{i'_2} \left(\frac{1}{45}\right).$$

Case 8

Under Case 8, $i_1 = i'_1$ and $i_2 = i'_2$. For example, $Cov(L_{123}, L_{124})$ falls under this case.

To begin, note that

$$
\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i_1 i_2 i'_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_1 j'_1} < X_{i_2 j'_2} < X_{i'_3 j'_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i_2 j'_2} < X_{i'_3 j'_3})).
\end{aligned}
$$

Now, since $i_1 = i'_1$ and $i_2 = i'_2$, it follow that j_1 may or may not be equal to j'_1 and j_2 may or may not be equal to j'_2. Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1
3	-	x	-	-	x	-	1
4	x	x	-	x	x	-	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

8.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i_2 j'_2} < X_{i'_3 j'_3})) = 0,$

8.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j_1} < X_{i_2 j'_2} < X_{i'_3 j'_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45},$

8.3 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i_2 j_2} < X_{i'_3 j'_3})) = \frac{4}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{180},$ and

8.4 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i'_3 j'_3})) = \frac{2}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{18}.$

Lastly, multiplying 8.2 through 8.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$
\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i_1 i_2 i'_3}) \\
&= n_{i_1} n_{i_2}(n_{i_2} - 1) n_{i_3} n_{i'_3}\left(\frac{1}{45}\right) + n_{i_1}(n_{i_1} - 1) n_{i_2} n_{i_3} n_{i'_3}\left(\frac{1}{180}\right) + n_{i_1} n_{i_2} n_{i_3} n_{i'_3}\left(\frac{1}{18}\right).
\end{aligned}
$$

Case 9

Under Case 9, $i_1 = i'_1$, $i_2 \neq i'_2$ and $i_3 = i'_2$. For example, $Cov(L_{123}, L_{134})$ falls under this case. To begin, note that

$$\begin{aligned} & Cov(L_{i_1 i_2 i_3}, L_{i_1 i_2 i'_3}) \\ = \;& Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_1 j'_1} < X_{i_3 j'_3} < X_{i'_3 j'_3})\right) \\ = \;& \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i_3 j'_3} < X_{i'_3 j'_3})). \end{aligned}$$

Now, since $i_1 = i'_1$, $i_2 \neq i'_2$ and $i_3 = i'_2$, it follow that j_1 may or may not be equal to j'_1 and j_3 may or may not be equal to j'_2. Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1
3	-	-	x	-	x	-	1
4	x	-	x	x	x	-	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

9.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i_3 j'_3} < X_{i'_3 j'_3})) = 0$,

9.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j_1} < X_{i_3 j'_3} < X_{i'_3 j'_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$,

9.3 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i_3 j_3} < X_{i'_3 j'_3})) = \frac{3}{120} - \frac{1}{3!}\frac{1}{3!} = -\frac{1}{360}$, and

9.4 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j_1} < X_{i_3 j_3} < X_{i_3 j'_3})) = \frac{1}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{72}$.

Lastly, multiplying 9.2 through 9.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$
\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i_1 i_2 i_3'}) \\
&= n_{i_1} n_{i_2} n_{i_3}(n_{i_3}-1) n_{i_3'} \left(\frac{1}{45}\right) + n_{i_1}(n_{i_1}-1) n_{i_2} n_{i_3} n_{i_3'} \left(-\frac{1}{360}\right) + n_{i_1} n_{i_2} n_{i_3} n_{i_3'} \left(\frac{1}{72}\right).
\end{aligned}
$$

Case 10

Under Case 10, $i_1 = i_1'$, $i_2 \neq i_2'$ and $i_3 = i_3'$. For example, $Cov(L_{124}, L_{134})$ falls under this case. To begin, note that

$$
\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i_1, i_2', i_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), \sum_{j_1'=1}^{n_{i_1}} \sum_{j_2'=1}^{n_{i_2}} \sum_{j_3'=1}^{n_{i_3'}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} \sum_{j_1'=1}^{n_{i_1}} \sum_{j_2'=1}^{n_{i_2}} \sum_{j_3'=1}^{n_{i_3'}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3})).
\end{aligned}
$$

Now, since $i_1 = i_1'$, $i_2 \neq i_2'$ and $i_3 = i_3'$, it follows that j_1 may or may not be equal to j_1' and j_3 may or may not be equal to j_3'. Hence, we get the following four subcases.

	j_1	j_2	j_3	j_1'	j_2'	j_3'	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1
3	-	-	x	-	-	x	1
4	x	-	x	x	-	x	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

10.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j_1'} < X_{i_2' j_2'} < X_{i_3 j_3'})) = 0,$

10.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1} < X_{i'_2j'_2} < X_{i_3j'_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$,

10.3 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j'_1} < X_{i'_2j'_2} < X_{i_3j_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$, and

10.4 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1} < X_{i'_2j'_2} < X_{i_3j_3})) = \frac{2}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{18}$.

Lastly, multiplying 10.2 through 10.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$\begin{aligned} & Cov(L_{i_1i_2i_3}, L_{i_1i_2i'_3}) \\ = \;& n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)n_{i'_2}\left(\frac{1}{45}\right) + n_{i_1}(n_{i_1}-1)n_{i_2}n_{i_3}n_{i'_2}\left(\frac{1}{45}\right) + n_{i_1}n_{i_2}n_{i_3}n_{i'_2}\left(\frac{1}{18}\right). \end{aligned}$$

Case 11

Under Case 11, $i_1 \neq i'_1$, $i_2 = i'_1$ and $i_3 = i'_2$. For example, $Cov(L_{123}, L_{234})$ falls under this case. To begin, note that

$$\begin{aligned} & Cov(L_{i_1i_2i_3}, L_{i_2,i_3,i'_3}) \\ = \;& Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_2j'_2} < X_{i_3j'_3} < X_{i'_3j'_3})\right) \\ = \;& \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j'_2} < X_{i_3j'_3} < X_{i'_3j'_3})). \end{aligned}$$

Now, since $i_1 \neq i'_1$, $i_2 = i'_1$ and $i_3 = i'_2$, it follows that j_2 may or may not be equal to j'_1 and j_3 may or may not be equal to j'_2. Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	x	-	x	-	-	1
3	-	-	x	-	x	-	1
4	-	x	x	x	x	-	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

11.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j'_1} < X_{i_2j'_2} < X_{i'_3j'_3})) = 0,$

11.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j_2} < X_{i_3j'_3} < X_{i'_3j'_3})) = \frac{3}{120} - \frac{1}{3!}\frac{1}{3!} = -\frac{1}{360},$

11.3 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j'_2} < X_{i_3j_3} < X_{i'_3j'_3})) = \frac{3}{120} - \frac{1}{3!}\frac{1}{3!} = -\frac{1}{360},$ and

11.4 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j_2} < X_{i_3j_3} < X_{i_3j'_3})) = \frac{1}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{72}.$

Lastly, multiplying 11.2 through 11.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$
\begin{aligned}
&Cov(L_{i_1i_2i_3}, L_{i_2,i_3,i'_3}) \\
&= n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)n_{i'_3}\left(-\frac{1}{360}\right) + n_{i_1}n_{i_2}(n_{i_2}-1)n_{i_3}n_{i'_3}\left(-\frac{1}{360}\right) + n_{i_1}n_{i_2}n_{i_3}n_{i'_3}\left(\frac{1}{72}\right).
\end{aligned}
$$

Case 12

Under Case 12, $i_1 \neq i'_1$, $i_2 = i'_1$ and $i_3 = i'_3$. For example, $Cov(L_{123}, L_{124})$ falls under this case. To begin, note that

$$
\begin{aligned}
&Cov(L_{i_1i_2i_3}, L_{i_2,i'_2,i_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_2j'_2} < X_{i'_2j'_2} < X_{i_3j'_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j'_2} < X_{i'_2j'_2} < X_{i_3j'_3})).
\end{aligned}
$$

Now, since $i_1 \neq i'_1$, $i_2 = i'_1$ and $i_3 = i'_3$, it follows that j_2 may or may not be equal to j'_1 and j_3 may or may not be equal to j'_3. Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	x	-	x	-	-	1
3	-	-	x	-	-	x	1
4	-	x	x	x	-	x	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

12.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j'_2} < X_{i'_2j'_2} < X_{i_3j'_3})) = 0,$

12.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j_2} < X_{i'_2j'_2} < X_{i_3j'_3})) = \frac{3}{120} - \frac{1}{3!}\frac{1}{3!} = -\frac{1}{360},$

12.3 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j'_2} < X_{i'_2j'_2} < X_{i_3j_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45},$ and

12.4 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j_2} < X_{i'_2j'_2} < X_{i_3j_3})) = \frac{1}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{72}.$

Lastly, multiplying 12.2 through 12.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$\begin{aligned} &Cov(L_{i_1i_2i_3}, L_{i_2,i'_2,i_3}) \\ &= n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)n_{i'_2}\left(-\frac{1}{360}\right) + n_{i_1}n_{i_2}(n_{i_2}-1)n_{i_3}n_{i'_2}\left(\frac{1}{45}\right) + n_{i_1}n_{i_2}n_{i_3}n_{i'_2}\left(\frac{1}{72}\right). \end{aligned}$$

Case 13

Under Case 13, $i_1 \neq i'_1$, $i_2 = i'_2$ and $i_3 = i'_3$. For example, $Cov(L_{134}, L_{234})$ falls under this case. To begin, note that

$$
\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i_1',i_2,i_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), \sum_{j_1'=1}^{n_{i_1}}\sum_{j_2'=1}^{n_{i_2}}\sum_{j_3'=1}^{n_{i_3'}} I(X_{i_1'j_1'} < X_{i_2j_2'} < X_{i_3j_3'})\right) \\
&= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j_1'=1}^{n_{i_1}}\sum_{j_2'=1}^{n_{i_2}}\sum_{j_3'=1}^{n_{i_3'}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1'j_1'} < X_{i_2j_2'} < X_{i_3j_3'})).
\end{aligned}
$$

Now, since $i_1 \neq i_1'$, $i_2 = i_2'$ and $i_3 = i_3'$, it follows that j_2 may or may not be equal to j_2' and j_3 may or may not be equal to j_3'. Hence, we get the following four subcases.

	j_1	j_2	j_3	j_1'	j_2'	j_3'	Ties
1	-	-	-	-	-	-	0
2	-	x	-	-	x	-	1
3	-	-	x	-	-	x	1
4	-	x	x	-	x	x	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

13.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1'j_1'} < X_{i_2j_2'} < X_{i_3j_3'})) = 0$,

13.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1'j_1'} < X_{i_2j_2} < X_{i_3j_3'})) = \frac{4}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{180}$,

13.3 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1'j_1'} < X_{i_2j_2'} < X_{i_3j_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$, and

13.4 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1'j_1'} < X_{i_2j_2} < X_{i_3j_3})) = \frac{2}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{18}$.

Lastly, multiplying 13.2 through 13.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$
\begin{aligned}
&Cov(L_{i_1 i_2 i_3}, L_{i_1',i_2,i_3}) \\
&= n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)n_{i_1'}\left(\frac{1}{180}\right) + n_{i_1}n_{i_2}(n_{i_2}-1)n_{i_3}n_{i_1'}\left(\frac{1}{45}\right) + n_{i_1}n_{i_2}n_{i_3}n_{i_1'}\left(\frac{1}{18}\right).
\end{aligned}
$$

Case 14

Under Case 14, $i_1 = i'_1$, $i_2 = i'_2$ and $i_3 = i'_3$. For example, $Cov(L_{123}, L_{123})$ falls under this case. To begin, note that

$$\begin{aligned} &Cov(L_{i_1 i_2 i_3}, L_{i_1,i_2,i_3}) \\ &= \ Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_1 j'_1} < X_{i_2 j'_2} < X_{i_3 j'_3})\right) \\ &= \ \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i_2 j'_2} < X_{i_3 j'_3})). \end{aligned}$$

Now, since $i_1 = i'_1$, $i_2 = i'_2$ and $i_3 = i'_3$, it follow that j_1 may or may not be equal to j'_1, j_2 may or may not be equal to j'_2 and j_3 may or may not be equal to j'_3. Hence, we get the following eight subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1
3	-	x	-	-	x	-	1
4	-	-	x	-	-	x	1
5	x	x	-	x	x	-	2
6	x	-	x	x	-	x	2
7	-	x	x	-	x	x	2
8	x	x	x	x	x	x	3

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

14.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j'_1} < X_{i_2 j'_2} < X_{i_3 j'_3}) = 0$,

14.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i_1 j_1} < X_{i_2 j'_2} < X_{i_3 j'_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$,

14.3 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1'} < X_{i_2j_2} < X_{i_3j_3'})) = \frac{4}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{180}$,

14.4 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1'} < X_{i_2j_2'} < X_{i_3j_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$

14.5 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3'})) = \frac{2}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{18}$,

14.6 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1'} < X_{i_2j_2'} < X_{i_3j_3}) = \frac{3}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{18}$,

14.7 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1} < X_{i_2j_2'} < X_{i_3j_3'})) = \frac{2}{24} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{18}$, and

14.8 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1,j_1} < X_{i_2,j_2} < X_{i_3,j_3})$

Lastly, multiplying 14.2 through 14.8 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$
\begin{aligned}
& Cov(L_{i_1i_2i_3}, L_{i_1i_2i_3}) \\
& = \; n_{i_1}n_{i_2}(n_{i_2}-1)\,n_{i_3}(n_{i_3}-1)\left(\frac{1}{45}\right) + n_{i_1}(n_{i_1}-1)\,n_{i_2}n_{i_3}(n_{i_3}-1)\left(\frac{1}{180}\right) \\
& + \; n_{i_1}(n_{i_1}-1)\,n_{i_2}(n_{i_2}-1)\,n_{i_3}\left(\frac{1}{45}\right) + n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)\left(\frac{1}{18}\right) \\
& + \; n_{i_1}n_{i_2}(n_{i_2}-1)\,n_{i_3}\left(\frac{1}{18}\right) + n_{i_1}(n_{i_1}-1)\,n_{i_2}n_{i_3}\left(\frac{1}{18}\right) + n_{i_1}n_{i_2}n_{i_3}\left(\frac{5}{36}\right)
\end{aligned}
$$

To summarize, the variance and covariance results for each of the 14 cases are given below. The variance of T_L is obtained by simply summing these quantities according to equation (23).

$$2: \quad n_{i_1}n_{i_2}n_{i_3}n_{i'_2}n_{i'_3}\left(\frac{1}{45}\right)$$

$$3: \quad n_{i_1}n_{i_2}n_{i_3}n_{i'_2}n_{i'_3}\left(-\frac{1}{360}\right)$$

$$4: \quad n_{i_1}n_{i_2}n_{i_3}n_{i'_1}n_{i'_3}\left(\frac{1}{180}\right)$$

$$5: \quad n_{i_1}n_{i_2}n_{i_3}n_{i'_1}n_{i'_2}\left(-\frac{1}{360}\right)$$

$$6: \quad n_{i_1}n_{i_2}n_{i_3}n_{i'_2}n_{i'_3}\left(-\frac{7}{360}\right)$$

$$7: \quad n_{i_1}n_{i_2}n_{i_3}n_{i'_1}n_{i'_2}\left(\frac{1}{45}\right)$$

$$8: \quad n_{i_1}n_{i_2}(n_{i_2}-1)n_{i_3}n_{i'_3}\left(\frac{1}{45}\right)+n_{i_1}(n_{i_1}-1)n_{i_2}n_{i_3}n_{i'_3}\left(\frac{1}{180}\right)+n_{i_1}n_{i_2}n_{i_3}n_{i'_3}\left(\frac{1}{18}\right)$$

$$9: \quad n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)n_{i'_3}\left(\frac{1}{45}\right)+n_{i_1}(n_{i_1}-1)n_{i_2}n_{i_3}n_{i'_3}\left(-\frac{1}{360}\right)+n_{i_1}n_{i_2}n_{i_3}n_{i'_3}\left(\frac{1}{72}\right)$$

$$10: \quad n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)n_{i'_2}\left(\frac{1}{45}\right)+n_{i_1}(n_{i_1}-1)n_{i_2}n_{i_3}n_{i'_2}\left(\frac{1}{45}\right)+n_{i_1}n_{i_2}n_{i_3}n_{i'_2}\left(\frac{1}{18}\right)$$

$$11: \quad n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)n_{i'_3}\left(-\frac{1}{360}\right)+n_{i_1}n_{i_2}(n_{i_2}-1)n_{i_3}n_{i'_3}\left(-\frac{1}{360}\right)+n_{i_1}n_{i_2}n_{i_3}n_{i'_3}\left(\frac{1}{72}\right)$$

$$12: \quad n_{i_1}n_{i_2}n_{i_3}(n_{i_3}-1)n_{i'_2}\left(-\frac{1}{360}\right)+n_{i_1}n_{i_2}(n_{i_2}-1)n_{i_3}n_{i'_2}\left(\frac{1}{45}\right)+n_{i_1}n_{i_2}n_{i_3}n_{i'_2}\left(\frac{1}{72}\right)$$

$$13: \quad n_{i_1} n_{i_2} n_{i_3} (n_{i_3} - 1) n_{i'_1} \left(\frac{1}{180}\right) + n_{i_1} n_{i_2} (n_{i_2} - 1) n_{i_3} n_{i'_1} \left(\frac{1}{45}\right) + n_{i_1} n_{i_2} n_{i_3} n_{i'_1} \left(\frac{1}{18}\right)$$

$$14: \quad n_{i_1} \left(n_{i_1} - 1\right) n_{i_2} \left(n_{i_2} - 1\right) n_{i_3} \left(n_{i_3} - 1\right) (0)$$

$$+ \quad n_{i_1} n_{i_2} \left(n_{i_2} - 1\right) n_{i_3} \left(n_{i_3} - 1\right) \left(\frac{1}{45}\right)$$

$$+ \quad n_{i_1} \left(n_{i_1} - 1\right) n_{i_2} n_{i_3} \left(n_{i_3} - 1\right) \left(\frac{1}{180}\right)$$

$$+ \quad n_{i_1} \left(n_{i_1} - 1\right) n_{i_2} \left(n_{i_2} - 1\right) n_{i_3} \left(\frac{1}{45}\right)$$

$$+ \quad n_{i_1} n_{i_2} n_{i_3} \left(n_{i_3} - 1\right) \left(\frac{1}{18}\right)$$

$$+ \quad n_{i_1} n_{i_2} \left(n_{i_2} - 1\right) n_{i_3} \left(\frac{1}{18}\right)$$

$$+ \quad n_{i_1} \left(n_{i_1} - 1\right) n_{i_2} n_{i_3} \left(\frac{1}{18}\right)$$

$$+ \quad n_{i_1} n_{i_2} n_{i_3} \left(\frac{5}{36}\right)$$

With that said, after very extensive algebraic manipulation, the result for the variance of T_L can be simplified for the balanced design (i.e. $n_1 = n_2 = \cdots = n_k = n$).

It can be given by

$$Var(T_L) \quad = \quad \frac{2}{360}\left\{39\begin{pmatrix} p \\ 5 \end{pmatrix} n^5 + \begin{pmatrix} p \\ 4 \end{pmatrix}(48n^5 + 27n^4) + \begin{pmatrix} p \\ 3 \end{pmatrix}(9n^5 + 12n^4 + 4n^3)\right\}.$$

Variance of T_R

The variance of T_R follows from the variance of T_L. For instance, recall that the variance of T_L is a function of p and $n_1, n_2, \ldots, n_p$, say $v(p; n_i, n_2, \ldots, n_p)$. So, by essentially reversing the order of the groups, we can obtain the variance of T_R from the variance of T_L. That is , $V(T_R) = V(k - p + 1; n_k, n_{k-1}, \ldots, n_p)$. Note that this results is also important from a programming point of view.

Variance of T_A

Next note that

$$V(T_A) = Cov(T_A, T_A) = Cov(\sum_{i_1=1}^{p-1} \sum_{i_3=p+1}^{k} A_{i_1 p i_3}, \sum_{i'_1=1}^{p-1} \sum_{i'_3=p+1}^{k} A_{i'_1 p i'_3}) =$$

$$\sum_{i_1=1}^{p-1} \sum_{i_3=p+1}^{k} Var(A_{i_1,p,i_3}) + 2\sum_{i \neq i'}^{p-1} Cov(A_{i_1 p i_3}, A_{i'_1 p i'_3}) \qquad (24)$$

Once more, recall that $i_1 < p < i_3$ for i and $i'_1 < p < i'_3$ for i'.

The table below shows all the possible ties between groups:

	i_1	i_p	i_3	i'_1	i'_p	i'_3	$Ties$
1	$-$	x	$-$	$-$	x	$-$	1
2	x	x	$-$	x	x	$-$	2
3	$-$	x	x	$-$	x	x	2
4	x	x	x	x	x	x	3

Case 4 is the variance component $Var(A_{i_1pi_3})$.

Case 1

Under Case 1, $i_1 \neq i'_1$, $i_p = i'_p$ and $i_3 \neq i'_3$. For example, $Cov(A_{134}, A_{235})$ falls under this case. To begin, note that

$$\begin{aligned}
&Cov(A_{i_1pi_3}, A_{i_1pi'_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_p}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{pj_2} > X_{i_3j_3})), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_p}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_1j'_1} < X_{pj'_2} > X_{i_3j'_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_p}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_p}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{pj_2}.X_{i_3j_3}), I(X_{i_1j'_1} < X_{pj'_2} > X_{i_3j'_3})).
\end{aligned}$$

Now, since $i_1 \neq i'_1$, $i_p = i'_p$ and $i_3 \neq i'_3$, it follows that j_1 may or may not be equal to j'_1 . Hence, we get the following two subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	x	-	-	x	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

1.1 $Cov(I(X_{i_1j_1} < X_{pj_2} > X_{i_3j_3})), I(X_{i_1j'_1} < X_{pj'_2} > X_{i'_3j'_3})) = 0,$

1.2 $Cov(I(X_{i_1j_1} < X_{pj_2} > X_{i_3j_3}, I(X_{i_1j'_1} < X_{pj'_2} > X_{i'_3j'_3})) = \frac{24}{120} - \frac{1}{3}\frac{1}{3} = \frac{4}{45}$.

Lastly, multiplying 1.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$Cov(A_{i_1i_2i_3}, A_{i_1pi'_3}) = n_{i_1}n_pn_{i_3}n_{i'_1}n_{i'_3}\left(\frac{4}{45}\right).$$

Case 2

Under Case 2, $i_1 = i'_1$, $i_p = i'_p$ and $i_3 \neq i'_3$. For example, $Cov(A_{123}, A_{124})$ falls under this case. To begin, note that

$$\begin{aligned} &Cov(A_{i_1pi_3}, A_{i_1,p,i'_3}) \\ &= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_p}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3})), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_p}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_1j'_1} < X_{pj'_p} > X_{i'_3j'_3})\right) \\ &= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_p}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_p}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1j'_1} < X_{pj'_p} > X_{i'_3j'_3})). \end{aligned}$$

Now, since $i_1 = i'_1$, $i_p = i'_p$ and $i_3 \neq i'_3$, it follows that j_1 may or may not be equal to j'_1 and j_2 may or may not be equal to j'_2 . Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1
3	-	x	-	-	x	-	1
4	x	x	-	x	x	-	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

2.1 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1j'_1} < X_{pj'_p} > X_{i'_3j'_3})) = 0$,

2.2 $Cov(I(X_{i_1j_1} < X_{pj_2} > X_{i_3j_3}), I(X_{i_1j_1} < X_{pj'_p} > X_{i'_3j'_3})) = \frac{16}{120} - \frac{1}{3}\frac{1}{3} = \frac{1}{45}$

2.3 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1j'_1} < X_{pj_p} > X_{i'_3j'_3})) = \frac{24}{120} - \frac{1}{3}\frac{1}{3} = \frac{4}{45}$

2.4 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1j_1} < X_{pj_p} > X_{i'_3j'_3})) = \frac{6}{24} - \frac{1}{3}\frac{1}{3} = \frac{5}{36}$

Lastly, multiplying 2.2 through 2.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$\begin{aligned} & Cov(A_{i_1i_2i_3}, A_{i_1pi'_3}) \\ = \; & n_{i_1}n_p(n_p-1)n_{i_3}n_{i'_3}\left(\frac{1}{45}\right) + n_{i_1}(n_{i_1}-1)n_pn_{i_3}n_{i'_3}\left(\frac{4}{45}\right) + n_{i_1}n_pn_{i_3}n_{i'_3}\left(\frac{5}{36}\right). \end{aligned}$$

Case 3

Under Case 3, $i_1 \neq i'_1$, $i_p = i'_p$ and $i_3 = i'_3$. For example, $Cov(A_{123}, A_{124})$ falls under this case. To begin, note that

$$\begin{aligned} & Cov(A_{i_1pi_3}, A_{i'_1,p,i_3}) \\ = \; & Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_p}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3})), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_p}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i'_1j'_1} < X_{pj'_p} > X_{i_3j'_3})\right) \\ = \; & \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_p}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_p}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i'_1j'_1} < X_{pj'_p} > X_{i_3j'_3})). \end{aligned}$$

Now, since $i_1 \neq i'_1$, $i_p = i'_p$ and $i_3 = i'_3$, it follows that j_p may or may not be equal to j'_p and j_3 may or may not be equal to j'_3 . Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	x	-	-	x	-	1
3	-	-	x	-	-	x	1
4	-	x	x	-	x	x	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

3.1 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i'_1j'_1} < X_{pj'_p} > X_{i_3j'_3})) = 0,$

3.2 $Cov(I(X_{i_1j_1} < X_{pj_2} > X_{i_3j_3}), I(X_{i'_1j'_1} < X_{pj_p} > X_{i_3j'_3})) = \frac{24}{120} - \frac{1}{3}\frac{1}{3} = \frac{4}{45}$

3.3 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i'_1j'_1} < X_{pj'_p} > X_{i_3j_3})) = \frac{16}{120} - \frac{1}{3}\frac{1}{3} = \frac{1}{45}$

3.4 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i'_1j'_1} < X_{pj_p} > X_{i_3j_3})) = \frac{6}{24} - \frac{1}{3}\frac{1}{3} = \frac{5}{36}$

Lastly, multiplying 3.2 through 3.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$\begin{aligned} & Cov(A_{i_1i_2i_3}, A_{i'_1,p,i_3}) \\ = \quad & n_{i_1}n_pn_{i_3}(n_{i_3}-1)n_{i'_1}\left(\frac{4}{45}\right) + n_{i_1}n_p(n_p-1)n_{i_3}n_{i'_1}\left(\frac{1}{45}\right) + n_{i_1}n_pn_{i_3}n_{i'_1}\left(\frac{5}{36}\right). \end{aligned}$$

Case 4

Under Case 4, $i_1 = i'_1$, $i_p = i'_p$ and $i_3 = i'_3$. For example, $Cov(A_{123}, A_{123})$ falls under this case. To begin, note that

$$\begin{aligned} & Cov(A_{i_1pi_3}, A_{i_1,p,i_3}) \\ = \quad & Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_p}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3})), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_p}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_1j'_1} < X_{pj'_p} > X_{i_3j'_3})\right) \\ = \quad & \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_p}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_p}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1j'_1} < X_{pj'_p} > X_{i_3j'_3})). \end{aligned}$$

Now, since $i_1 = i'_1$, $i_p = i'_p$ and $i_3 = i'_3$, it follows that j_1 may or may not be equal to j'_1, j_p may or may not be equal to j'_p and j_3 may or may not be equal to j'_3 . Hence, we get the following eight subcases.

	j_1	j_2	j_3	j_1'	j_2'	j_3'	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1
3	-	x	-	-	x	-	1
4	-	-	x	-	-	x	1
5	x	x	-	x	x	-	2
6	x	-	x	x	-	x	2
7	-	x	x	-	x	x	2
8	x	x	x	x	x	x	3

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

4.1 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1j_1'} < X_{pj_p'} > X_{i_3j_3'})) = 0,$

4.2 $Cov(I(X_{i_1j_1} < X_{pj_2} > X_{i_3j_3}), I(X_{i_1j_1} < X_{pj_p'} > X_{i_3j_3'})) = \frac{16}{120} - \frac{1}{3}\frac{1}{3} = \frac{1}{45}$

4.3 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1j_1'} < X_{pj_p} > X_{i_3j_3'})) = \frac{24}{120} - \frac{1}{3}\frac{1}{3} = \frac{4}{45}$

4.4 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1,j_1'} < X_{p,j_p'} > X_{i_3,j_3})) = \frac{16}{120} - \frac{1}{3}\frac{1}{3} = \frac{1}{45}$

4.5 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1,j_1} < X_{p,j_p} > X_{i_3,j_3'})) = \frac{6}{24} - \frac{1}{3}\frac{1}{3} = \frac{5}{36}$

4.6 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1,j_1} < X_{p,j_p'} > X_{i_3,j_3})) = \frac{4}{24} - \frac{1}{3}\frac{1}{3} = \frac{1}{18}$

4.7 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1,j_1'} < X_{p,j_p} > X_{i_3,j_3})) = \frac{6}{24} - \frac{1}{3}\frac{1}{3} = \frac{5}{36}$

4.8 $Cov(I(X_{i_1j_1} < X_{pj_p} > X_{i_3j_3}), I(X_{i_1,j_1} < X_{p,j_p} > X_{i_3,j_3})) = \frac{2}{6} - \frac{1}{3}\frac{1}{3} = \frac{2}{9}.$

Lastly, multiplying 4.2 through 4.8 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$
\begin{aligned}
& Cov(A_{i_1 i_2 i_3}, A_{i_1' p i_3}) \\
= \quad & n_{i_1} n_p (n_p - 1) n_{i_3} (n_{i_3} - 1) \left(\frac{1}{45}\right) + n_{i_1}(n_{i_1} - 1) n_p n_{i_3} (n_{i_3} - 1) \left(\frac{4}{45}\right) \\
+ \quad & n_{i_1}(n_{i_1} - 1) n_p (n_p - 1) n_{i_3} \left(\frac{1}{45}\right) + n_{i_1} n_p n_{i_3} (n_{i_3} - 1) \left(\frac{5}{36}\right) \\
+ \quad & n_{i_1} n_p (n_p - 1) n_{i_3} \left(\frac{1}{18}\right) + n_{i_1}(n_{i_1} - 1) n_p n_{i_3} \left(\frac{5}{36}\right) \\
+ \quad & n_{i_1} n_p n_{i_3} \left(\frac{2}{9}\right)
\end{aligned}
$$

To summarize, the variance and covariance results for each of the 14 cases are given below. The variance of T_A is obtained by simply summing these quantities according to equation (24)

$$1 \; : \; n_{i_1} n_p n_{i_3} n_{i_1'} n_{i_3'} \left(\frac{4}{45}\right)$$

$$2 \; : \; n_{i_1} n_p (n_p - 1) n_{i_3} n_{i_3'} \left(\frac{1}{45}\right) + n_{i_1}(n_{i_1} - 1) n_p n_{i_3} n_{i_3'} \left(\frac{4}{45}\right) + n_{i_1} n_p n_{i_3} n_{i_3'} \left(\frac{5}{36}\right)$$

$$3 \; : \; n_{i_1} n_p n_{i_3} (n_{i_3} - 1) n_{i_1'} \left(\frac{4}{45}\right) + n_{i_1} n_p (n_p - 1) n_{i_3} n_{i_1'} \left(\frac{1}{45}\right) + n_{i_1} n_p n_{i_3} n_{i_1'} \left(\frac{5}{36}\right)$$

$$4 \ : \ n_{i_1}n_p(n_p-1)n_{i_3}(n_{i_3}-1)\left(\frac{1}{45}\right)+n_{i_1}(n_{i_1}-1)n_pn_{i_3}(n_{i_3}-1)\left(\frac{4}{45}\right)$$

$$+n_{i_1}(n_{i_1}-1)n_p(n_p-1)n_{i_3}\left(\frac{1}{45}\right)+n_{i_1}n_pn_{i_3}(n_{i_3}-1)\left(\frac{5}{36}\right)$$

$$+n_{i_1}n_p(n_p-1)n_{i_3}\left(\frac{1}{18}\right)+n_{i_1}(n_{i_1}-1)n_pn_{i_3}\left(\frac{5}{36}\right)$$

$$+n_{i_1}n_pn_{i_3}\left(\frac{2}{9}\right)$$

With that said, after very extensive algebraic manipulation, the result for the variance of T_A can be simplified for the balanced design (i.e. $n_1 = n_2 = \cdots = n_k = n$). It can be given by

$$Var(T_A) = \frac{(p-1)(k-p)n^3}{45}\left\{(k-1)n^2+4(p-1)(k-p)n^2+\frac{5(k-1)n+2n}{4}+1\right\}$$

Covariance of T_L and T_A

Note first that

$$Cov(T_L, T_A) = Cov(\sum_{i=S_1^p(3)} L_{i_1i_2i_3}, \sum_{i'_1=1}^{p-1}\sum_{i'_3=p+1}^{k} A_{i'_1pi'_3})$$

$$= \sum_{i=S_1^p(3)}\sum_{i'_1=1}^{p-1}\sum_{i'_3=p+1}^{k} Cov(L_{i_1i_2i_3}, A_{i'_1pi'_3}) \quad (25)$$

There are the following possibilities of ties between groups:

	i_1	i_2	i_3	i'_1	p	i'_3	$Ties$
1	−	−	−	−	−	−	0
2	x	−	−	x	−	−	1
3	−	x	−	x	−	−	1
4	−	−	x	x	−	−	1
5	−	−	x	−	x	−	1
6	x	−	x	x	x	−	2
7	−	x	x	x	x	−	2

Case 2

Under Case 2, $i_1 = i'_1$, $i_2 \neq p$ and $i_3 \neq i'_3$. For example, $Cov(L_{123}, A_{145})$ falls under this case. To begin, note that

$$\begin{aligned} & Cov(L_{i_1 i_2 i_3}, A_{i_1 p i'_3}) \\ = \; & Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1,j_1} < X_{i_2,j_2} < X_{i_3,j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i'_1,j'_1} < X_{p,j'_p} > X_{i'_3,j'_3})\right) \\ = \; & \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{i_3 j_3}), I(X_{i'_1 j'_1} < X_{p j'_p} > X_{i'_3 j'_3})). \end{aligned}$$

Now, since $i_1 = i'_1$, $i_2 \neq p$ and $i_3 \neq i'_3$, it follows that j_1 may or may not be equal to j'_1 . Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_p	j'_3	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

2.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1'} < X_{pj_2'} < X_{i_3'j_3'})) = 0,$

2.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_1j_1} < X_{pj_p'} > X_{i_3'j_3'}) = \frac{9}{120} - \frac{1}{3!}\frac{1}{3} = \frac{7}{360},$

Lastly, multiplying 2.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$Cov(L_{i_1i_2i_3}, A_{i_1pi_3'}) = n_{i_1}n_{i_2}n_{i_3}n_pn_{i_3'}\left(\frac{7}{360}\right).$$

Case 3

Under Case 3, $i_1 \neq i_1'$, $i_2 = i_1'$ and $i_3 \neq i_3'$. For example, $Cov(L_{123}, A_{245})$ falls under this case. To begin, note that

$$\begin{aligned} &Cov(L_{i_1i_2i_3}, A_{i_2pi_3'}) \\ &= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), \sum_{j_1'=1}^{n_{i_1}}\sum_{j_2'=1}^{n_{i_2}}\sum_{j_3'=1}^{n_{i_3'}} I(X_{i_2j_1'} < X_{pj_p'} > X_{i_3'j_3'})\right) \\ &= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j_1'=1}^{n_{i_1}}\sum_{j_2'=1}^{n_{i_2}}\sum_{j_3'=1}^{n_{i_3'}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j_1'} < X_{pj_p'} > X_{i_3'j_3'})). \end{aligned}$$

Now, since $i_1 \neq i_1'$, $i_2 = i_1'$ and $i_3 \neq i_3'$, it follows that j_2 may or may not be equal to j_1' . Hence, we get the following four subcases.

	j_1	j_2	j_3	j_1'	j_p'	j_3'	Ties
1	-	-	-	-	-	-	0
2	-	x	-	x	-	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

3.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j_1'} < X_{pj_p'} > X_{i_3'j_3'})) = 0,$

3.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_2j_2} < X_{pj_p'} > X_{i_3'j_3'})) = \frac{7}{120} - \frac{1}{3!}\frac{1}{3} = \frac{1}{360},$

Lastly, multiplying 3.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$Cov(L_{i_1i_2i_3}, A_{i_2pi'_3}) = n_{i_1}n_{i_2}n_{i_3}n_pn_{i'_3}\left(\frac{1}{360}\right).$$

Case 4

Under Case 4, $i_1 \neq i'_1$, $i_2 \neq i'_2$ and $i_3 = i'_1$. For example, $Cov(L_{123}, A_{345})$ falls under this case. To begin, note that

$$\begin{aligned} &Cov(L_{i_1i_2i_3}, A_{i_3,p,i_3}) \\ &= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_3j'_1} < X_{pj'_p} > X_{i'_3j'_3})\right) \\ &= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_3j'_1} < X_{pj'_p} > X_{i'_3j'_3})). \end{aligned}$$

Now, since $i_1 \neq i'_1$, $i_2 \neq i'_2$ and $i_3 = i'_1$, it follows that j_3 may or may not be equal to j'_1 . Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_p	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	-	x	x	-	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

4.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_3j'_1} < X_{pj'_p} > X_{i'_3j'_3})) = 0,$

4.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{i_3j_3}), I(X_{i_3j_3} < X_{pj'_p} > X_{i'_3j'_3})) = \frac{4}{120} - \frac{1}{3!}\frac{1}{3} = -\frac{1}{45},$

Lastly, multiplying 4.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$Cov(L_{i_1i_2i_3}, A_{i_2pi'_3}) = n_{i_1}n_{i_2}n_{i_3}n_pn_{i'_3}\left(-\frac{1}{45}\right).$$

Case 5

Under Case 5, $i_1 \neq i'_1$, $i_2 \neq i'_2$ and $i_3 = p$. For example, $Cov(L_{124}, A_{345})$ falls under this case. To begin, note that

$$\begin{aligned} &Cov(L_{i_1i_2i_3}, A_{i'_1pi'_3}) \\ &= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i'_1j'_1} < X_{pj'_p} > X_{i'_3j'_3})\right) \\ &= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), I(X_{i'_1j'_1} < X_{pj'_p} > X_{i'_3j'_3})). \end{aligned}$$

Now, since $i_1 \neq i'_1$, $i_2 \neq i'_2$ and $i_3 = p$, it follows that j_3 may or may not be equal to j'_p . Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_p	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	-	x	-	x	-	1

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

5.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), I(X_{i_1j'_1} < X_{pj'_p} > X_{i'_3j'_3})) = 0,$

5.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), I(X_{i_1j'_1} < X_{pj_p} > X_{i'_3j'_3})) = \frac{12}{120} - \frac{1}{3!}\frac{1}{3} = \frac{2}{45},$

Lastly, multiplying 5.2 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$Cov(L_{i_1i_2i_3}, A_{i'_1pi'_3}) = n_{i_1}n_{i_2}n_pn_{i'_1}n_{i'_3}\left(\frac{2}{45}\right).$$

Case 6

Under Case 6, $i_1 = i'_1$ and $i_3 = p$. For example, $Cov(L_{123}, A_{134})$ falls under this case. To begin, note that

$$\begin{aligned} &Cov(L_{i_1 i_2 i_3}, A_{i_1 p i'_3}) \\ &= Cov\left(\sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_p}), \sum_{j'_1=1}^{n_{i_1}} \sum_{j'_2=1}^{n_{i_2}} \sum_{j'_3=1}^{n_{i'_3}} I(X_{i_3 j'_1} < X_{p j'_p} > X_{i'_3 j'_3}) \right) \\ &= \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} \sum_{j'_1=1}^{n_{i_1}} \sum_{j'_2=1}^{n_{i_2}} \sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_p}), I(X_{i_3 j'_1} < X_{p j'_p} > X_{i'_3 j'_3})). \end{aligned}$$

Now, since $i_1 = i'_1$ and $i_3 = p.$, it follows that j_1 may or may not be equal to j'_1 and j_3 may or may not be equal to j_p. Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_p	j'_3	Ties
1	-	-	-	-	-	-	0
2	x	-	-	x	-	-	1
3	-	-	x	-	x	-	1
4	x	-	x	x	x	-	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

6.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_p}), I(X_{i_1 j'_3} < X_{p j'_p} > X_{i'_3 j'_3})) = 0,$

6.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_p}), I(X_{i_1 j_1} < X_{p, j'_p} > X_{i'_3 j'_3})) = \frac{9}{120} - \frac{1}{3!}\frac{1}{3} = \frac{7}{360},$

6.3 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_p}), I(X_{i_1 j'_1} < X_{p j_p} > X_{i'_3 j'_3})) = \frac{12}{120} - \frac{1}{3!}\frac{1}{3} = \frac{2}{45}$

6.4 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_p}), I(X_{i_1 j_1} < X_{i_p j_p} > X_{i'_3 j'_3})) = \frac{3}{24} - \frac{1}{3!}\frac{1}{3} = \frac{5}{72}$

Lastly, multiplying 6.2 through 6.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$\begin{aligned}
&Cov(L_{i_1i_2i_3}, A_{i_1pi'_3}) \\
&= n_{i_1}n_{i_2}n_p(n_p-1)n_{i'_3}\left(\frac{7}{360}\right) + n_{i_1}(n_{i_1}-1)n_{i_2}n_pn_{i'_3}\left(\frac{2}{45}\right) + n_{i_1}n_{i_2}n_pn_{i'_3}\left(\frac{5}{72}\right).
\end{aligned}$$

Case 7

Under Case 7, $i_1 \neq i'_1$, $i_2 = i'_1$ and $i_3 = i'_3$. For example, $Cov(L_{123}, A_{234})$ falls under this case. To begin, note that

$$\begin{aligned}
&Cov(L_{i_1i_2i_3}, A_{i_2,p,i'_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), \sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{i_2j'_2} < X_{pj'_p} > X_{i'_3j'_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}}\sum_{j'_1=1}^{n_{i_1}}\sum_{j'_2=1}^{n_{i_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), I(X_{i_2j'_2} < X_{pj'_p} > X_{i'_3j'_3})).
\end{aligned}$$

Now, since $i_2 = i'_1$ and $i_3 = p$., it follows that j_2 may or may not be equal to j'_1 and j_3 may or may not be equal to j_p. Hence, we get the following four subcases.

	j_1	j_2	j_3	j'_1	j'_p	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	x	-	x	-	-	1
3	-	-	x	-	x	-	1
4	-	x	x	x	x	-	2

Recall that an x denotes a tied subscript. Next, using our previously derived covariance results we obtain the following:

7.1 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), I(X_{i_2j'_2} < X_{pj'_p} > X_{i'_3j'_3})) = 0$,

7.2 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), I(X_{i_2j_2} < X_{pj'_p} > X_{i'_3j'_3})) = \frac{7}{120} - \frac{1}{3!}\frac{1}{3} = \frac{1}{360}$,

7.3 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), I(X_{i_2j'_2} < X_{pj_p} > X_{i'_3j'_3})) = \frac{12}{120} - \frac{1}{3!}\frac{1}{3} = \frac{2}{45}$

7.4 $Cov(I(X_{i_1j_1} < X_{i_2j_2} < X_{pj_p}), I(X_{i_2j_2} < X_{pj_p} > X_{i'_3j'_3})) = \frac{3}{24} - \frac{1}{3!}\frac{1}{3} = \frac{5}{72}$

Lastly, multiplying 7.2 through 7.4 by the corresponding number of terms, and then summing, leads to the desired result. That is,

$$\begin{aligned} & Cov(L_{i_1i_2i_3}, A_{i_1pi'_3}) \\ = \;& n_{i_1}n_{i_2}n_p(np-1)n_{i'_3}\left(\frac{1}{360}\right) + n_{i_1}n_{i_2}(n_{i_2}-1)n_pn_{i'_3}\left(\frac{2}{45}\right) + n_{i_1}n_{i_2}n_pn_{i'_3}\left(\frac{5}{72}\right). \end{aligned}$$

To summarize, the variance and covariance results for each of the 14 cases are given below. The covariance of T_L and T_A is obtained by simply summing these quantities according to equation (25).

Note that $i_1 < i_2 < i_3 \in T_L$ and $i'_1 < i'_p < i'_3 \in T_A$.

$$2 \;:\; n_{i_1}n_{i_2}n_{i_3}n_pn_{i'_3}\left(\frac{7}{360}\right)$$

$$3 \;:\; n_{i_1}n_{i_2}n_{i_3}n_pn_{i'_3}\left(\frac{1}{360}\right)$$

$$4 \;:\; n_{i_1}n_{i_2}n_{i_3}n_pn_{i'_3}\left(-\frac{1}{45}\right)$$

$$5 \;:\; n_{i_1}n_{i_2}n_pn_{i'_1}n_{i'_3}\left(\frac{2}{45}\right)$$

$$6 \;:\; n_{i_1}n_{i_2}n_p(n_p-1)n_{i'_3}\left(\frac{7}{360}\right) + n_{i_1}(n_{i_1}-1)n_{i_2}n_pn_{i'_3}\left(\frac{2}{45}\right) + n_{i_1}n_{i_2}n_pn_{i'_3}\left(\frac{5}{72}\right)$$

$$7 \;:\; n_{i_1}n_{i_2}n_p(np-1)n_{i'_3}\left(\frac{1}{360}\right) + n_{i_1}n_{i_2}(n_{i_2}-1)n_pn_{i'_3}\left(\frac{2}{45}\right) + n_{i_1}n_{i_2}n_pn_{i'_3}\left(\frac{5}{72}\right)$$

With that said, after very extensive algebraic manipulation, the result for the covariance of T_L and T_A can be simplified for the balanced design (i.e. $n_1 = n_2 = \cdots = n_k = n$). It can be given by

$$Cov(T_L, T_A) = \frac{(k-p)}{360}\left\{48\begin{pmatrix} p-1 \\ 3 \end{pmatrix} n^5 + \begin{pmatrix} p-1 \\ 2 \end{pmatrix}(40n^5 + 10n^4)\right\}$$

Covariance of T_L and T_R

Note first that

$$\begin{aligned} Cov(T_L, T_R) &= Cov\left(\sum_{i \in S_1^p(3)} L_{i_1 i_2 i_3}, \sum_{i' \in S_p^k(3)} R_{i'_1 i'_2 i'_3}\right) \\ &= \sum_{i \in S_1^p(3)} \sum_{i' \in S_p^k(3)} Cov(L_{i_1 i_2 i_3}, R_{i'_1 i'_2 i'_3}), \end{aligned} \tag{26}$$

where $i_1 < i_2 < i_3$ in i and $i'_1 < i'_2 < i'_3$ in i'. In what follows we solve for $Cov(L_{i_1 i_2 i_3}, R_{i'_1 i'_2 i'_3})$ for the different cases of i and i'. The table below shows that there are only two cases.

	i_1	i_2	i_3	i'_1	i'_2	i'_3	Ties
1	-	-	-	-	-	-	0
2	-	-	x	x	-	-	1

For example, in case 2, $i_3 = i'_1 = p$ and all other subscripts are distinct. The covariances for each of these cases are now calculated.

Case 1

Under case 1, the covariance terms are equal to zero since there are no tied subscripts. That is, since there are six distinct groups, $L_{i_1 i_2 i_3}$ and $R_{i'_1 i'_2 i'_3}$ are independent.

Case 2

Under case 2, $i_3 = i'_1 = p$. For example, $Cov(L_{123}, R_{345})$ falls under this case. To begin, note that

$$\begin{aligned}
&Cov(L_{i_1 i_2 p}, R_{p i'_2 i'_3}) \\
&= Cov\left(\sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_p} I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_3}), \sum_{j'_1=1}^{n_p}\sum_{j'_2=1}^{n_{i'_2}}\sum_{j'_3=1}^{n_{i'_3}} I(X_{p j'_1} > X_{i'_2 j'_2} > X_{i'_3 j'_3})\right) \\
&= \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_p}\sum_{j'_1=1}^{n_p}\sum_{j'_2=1}^{n_{i'_2}}\sum_{j'_3=1}^{n_{i'_3}} Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_3}), I(X_{p j'_1} > X_{i'_2 j'_2} > X_{i'_3 j'_3})).
\end{aligned}$$

Now, since $i_3 = i'_1 = p$, it follow that j_3 may or may not be equal to j'_1. Hence, we get the following two subcases.

	j_1	j_2	j_3	j'_1	j'_2	j'_3	Ties
1	-	-	-	-	-	-	0
2	-	-	x	x	-	-	1

One more, x denotes the tied subscript. Next, using our previously derived covariance results we obtain the following:

2.1 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_3}), I(X_{p j'_1} > X_{i'_2 j'_2} > X_{i'_3 j'_3})) = 0$ and

2.2 $Cov(I(X_{i_1 j_1} < X_{i_2 j_2} < X_{p j_3}), I(X_{p j_3} > X_{i'_2 j'_2} > X_{i'_3 j'_3})) = \frac{6}{120} - \frac{1}{3!}\frac{1}{3!} = \frac{1}{45}$.

Lastly, multiplying 2.2 by the corresponding number of terms leads to the desired result. That is,

$$Cov(L_{i_1 i_2 p}, R_{p i'_2 i'_3}) = n_{i_1} n_{i_2} n_p n_{i'_2} n_{i'_3} \left(\frac{1}{45}\right).$$

Hence, the covariance between T_L and T_R is obtained by simply substituting this value (and 0) back into (26). For the special case where the design is balance (i.e.

$n_1 = n_2 = \cdots = n_k = n$) we obtain

$$Cov(T_L, T_R) \;=\; \binom{p-1}{2}\binom{k-p}{2}\left(\frac{n^5}{45}\right).$$

3.2.4. Asymptotic Distribution of T

This section begins by generalizing the idea of U-statistics to the case in which we have more than one random sample. Suppose that for $i = 1, 2, \ldots k$, $X_{i1}, \ldots, X_{in_i}$, is an iid sample from F_i. In other words we have k random samples, each potentially from a different distribution. Recall, n_i is the sample size of the ith distribution. We may define the statistical functional;

$$\theta = E\phi(X_{11}, \ldots, X_{1a_1}; X_{21}, \ldots, X_{2a_2}; \cdots ; X_{k1}, \ldots X_{ka_k}). \tag{27}$$

Notice that the function in equation (27) has $a_1 + a_2 + \cdots + a_k$ arguments. The first a_1 of them can be permuted without changing the value of ϕ, the next a_2 of them can be permuted without changing the value of ϕ, etc. In other words, there are k distinct blocks of arguments of ϕ. In the case of T, $a_1 = a_2 = \cdots = a_k = 1$ (because we are taking one element of each sample at a time) in which case we obtain the U-statistic corresponding to the equation (27) as

$$U_N = \frac{1}{n_1 \cdots n_k} \sum_{j_1=1}^{n_1} \cdots \sum_{j_k=1}^{n_k} \phi(X_{1j_1}, \ldots X_{kj_k}). \tag{28}$$

Theorem 3.4:

For i=1, 2, . . . , k, consider

$$\phi_i(x) = E\left\{\phi(X_{11}, \ldots, X_{k1}) - \theta | X_{i1} = x\right\}, \text{and} \tag{29}$$

$$\sigma_i^2 = Var\left(\phi_i\right). \tag{30}$$

Suppose that

$$E(\phi^2(X_{11}, \ldots, X_{k1})) < \infty \text{ and } \max_{1 \leq i \leq k} \sigma_i^2 > 0.$$

Furthermore, suppose that there exist constants $\lambda_1, \lambda_2, \ldots \lambda_k$, in the interval (0,1) such that $n_i/N \to \lambda_i$ for all i. It therefore follows from Theorem 4.5.1 of Koroljuk and Borovskich, that

$$\frac{(U_N - \theta)}{\sqrt{Var(U_N)}} \to N(0, 1), \tag{31}$$

where $N(0, 1)$ is the standard normal distribution.

Theorem 3.5

For $i = 1, 2, \ldots, k$, let $X_{i1}, X_{i2}, \ldots, X_{in_i}$ denote k independent random samples from distribution F_i. Let $S = \{(i_1, i_2, i_3) : 1 \leq i_1 < i_2 < i_3 \leq k\}$ and let $S_1 \subset S$ where $S_1 \neq \emptyset$. For $i \in S_1$, define

$$T_i = \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} t_i(X_{i_1}, X_{i_2}, X_{i_3}) \text{ and}$$

$$\mu_i = E(t_i(X_{i_1}, X_{i_2}, X_{i_3})).$$

Finally, let

$$T = \sum_{i \in S_1} T_i \quad \text{and} \quad N = n_1 + n_2 + \cdots + n_k.$$

Assume the following:

A1 For $i \in S$, $t_i(X_{i_1}, X_{i_2}, X_{i_3})$ satisfies the kernel assumptions of the multi-sample U-statistic central limit theorem for 3 samples,

A2

$$\lim_{N \to \infty} \frac{n_i}{N} = \lambda_i \in (0,1) \text{ for } i \in \{1, 2, \ldots, k\},$$

A3

$$\lim_{N \to \infty} \frac{1}{N^5} V(T) = v \quad (> 0),$$

A4

$$\phi(X_1, X_2, \ldots, X_k) = \sum_{i \in S_1} \lambda_{i_1} \lambda_{i_2} \lambda_{i_3} t_i(X_{i_1} X_{i_2} X_{i_3})$$

satisfies the kernel assumptions of the multi-sample U-statistic central limit theorem for k samples.

Then, under A1-A4,

$$Z = \frac{T - E(T)}{\sqrt{Var(T)}} \xrightarrow{D} N(0,1). \tag{32}$$

Proof

Let $N = n_1 + n_2 + \ldots n_k$ and $N^* = n_1 n_2 \cdots n_k$. To begin, note that,

$$
\begin{aligned}
T &= \sum_{i \in S_1} T_i \\
&= \sum_{i \in S_1} \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} t_i(X_{i_i j_1}, X_{i_2 j_2}, X_{i_3 j_3}) \\
&= \sum_{i \in S_1} \frac{1}{n_{i_4} n_{i_5} \cdots n_{i_k}} \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} \sum_{j_4=1}^{n_{i_4}} \cdots \sum_{j_k=1}^{n_{i_k}} t_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}) \\
&= \frac{1}{N^*} \sum_{i \in S_1} n_{i_1} n_{i_2} n_{i_3} \sum_{j_1=1}^{n_{i_1}} \cdots \sum_{j_k=1}^{n_{i_k}} t_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}) \\
&= \frac{N^3}{N^*} \sum_{i \in S_1} \left(\frac{n_{i_1} n_{i_2} n_{i_3}}{N^3} \right) \sum_{j_1=1}^{n_{i_1}} \cdots \sum_{j_k=1}^{n_{i_k}} t_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}) \\
&= \frac{N^3}{N^*} \sum_{i \in S_1} \left(\frac{n_{i_1} n_{i_2} n_{i_3}}{N^3} - \lambda_{i_1} \lambda_{i_2} \lambda_{i_3} + \lambda_{i_1} \lambda_{i_2} \lambda_{i_3} \right) \left(\sum_{j_1=1}^{n_{i_1}} \cdots \sum_{j_k=1}^{n_{i_k}} t_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}) \right) \\
&= \frac{N^3}{N^*} \sum_{i \in S_1} \left(\frac{n_{i_1} n_{i_2} n_{i_3}}{N^3} - \lambda_{i_1} \lambda_{i_2} \lambda_{i_3} \right) \sum_{j1=1}^{n_{i_1}} \cdots \sum_{jk=1}^{n_{i_k}} t_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}) + \\
&\quad \frac{N^3}{N^*} \sum_{i \in S_1} \lambda_{i_1} \lambda_{i_2} \lambda_{i_3} \sum_{j_1=1}^{n_{i_1}} \cdots \sum_{j_k=1}^{n_{i_k}} t_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}) \\
&= N^3 \sum_{i \in S_1} \left(\frac{n_{i_1} n_{i_2} n_{i_3}}{N^3} - \lambda_{i_1} \lambda_{i_2} \lambda_{i_3} \right) \left(\frac{1}{n_{i_1} n_{i_2} n_{i_3}} \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=2}^{n_{i_2}} \sum_{j_3=3}^{n_{i_3}} t_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}) \right) + \\
&\quad \frac{N^3}{N^*} \sum_{j_1=1}^{n_{i_1}} \cdots \sum_{j_k=1}^{n_{i_k}} \left(\sum_{i \in S_1} t_i(X_{1 j_1}, X_{2 j_2}, \ldots, X_{k j_k}) \right),
\end{aligned}
$$

where, for the second term, we basically have

$$t_i(X_{1j_1}, X_{2j_2}, \ldots, X_{kj_k}) = t_i(X_{i_1j_1}, X_{i_2j_2}, X_{i_3j_3}).$$

$$= N^3 \left(\sum_{i \in S_1} \left(\frac{n_{i_1} n_{i_2} n_{i_3}}{N^3} - \lambda_{i_1}\lambda_{i_2}\lambda_{i_3} \right) \overline{T_i} \right) + N^3 \left(\frac{1}{N^*} \sum_{j_1=1}^{n_{i_1}} \cdots \sum_{j_k=1}^{n_{i_k}} \phi(X_{1j_1}, X_{2j_2}, \ldots, X_{kj_k}) \right),$$

where

$$\phi(X_1, X_2, \ldots, X_k) = \sum_{i \in S_1} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3} t_i(X_{i_1}, X_{i_2}, X_{i_3})$$

$$= N^3(T_1 + \overline{T_2}).$$

Next, let

$$\begin{aligned} T_N &= \frac{1}{N^{3-\frac{1}{2}}}(T - E(T)) \\ &= \frac{1}{N^{3-\frac{1}{2}}} N^3(T_1 + \overline{T_2} - E(T_1) - E(\overline{T_2})) \\ &= \sqrt{N}(T_1 - E(T_1)) + \sqrt{N}(\overline{T_2} - E(\overline{T_2})) \\ &= T_{N1} + T_{N2}. \end{aligned}$$

It follows that,

$$
\begin{aligned}
Z &= \frac{T - E(T)}{\sqrt{V(T)}} \\
&= \frac{\frac{1}{N^{3-\frac{1}{2}}}(T - E(T))}{\frac{1}{N^{3-\frac{1}{2}}}\sqrt{V(T)}} \\
&= \frac{T_{N1} + T_{N2}}{\sqrt{\frac{1}{N^5}V(T)}} \\
&= \frac{T_{N1}}{\sqrt{\frac{1}{N^5}V(T)}} + \frac{T_{N2}}{\sqrt{\frac{1}{N^5}V(T)}} \\
&= Z_1 + Z_2.
\end{aligned}
$$

Consider Z_1 first.

$$
\begin{aligned}
Z_1 &= \frac{T_{N1}}{\sqrt{\frac{1}{N^5}V(T)}} \\
&= \frac{\sqrt{N}(T_1 - E(T_1))}{\sqrt{\frac{1}{N^5}V(T)}} \\
&= \frac{\sum_{i\in S_1}\left(\frac{n_{i_1}n_{i_2}n_{i_3}}{N^3} - \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\right)\left(\sqrt{\frac{N}{N_{3i}}}\right)\left(\sqrt{N_{3i}}(\overline{T_i} - E(\overline{T_i}))\right)}{\sqrt{\frac{1}{N^5}V(T)}},
\end{aligned}
$$

where $N_{3i} = n_{i_1} + n_{i_2} + n_{i_3}$

$$
= \frac{\sum_{i\in S_1} a_i b_i c_i}{\sqrt{\frac{1}{N^5}V(T)}}.
$$

By A2, $\lim_{N \to \infty} a_i = 0$.

Also by A2, $\lim_{N \to \infty} b_i = \lim_{N \to \infty} \sqrt{\dfrac{N}{n_{i_1} + n_{i_2} + n_{i_3}}} = \sqrt{\dfrac{1}{\lambda_{i_1} + \lambda_{i_2} + \lambda_{i_3}}}$ $(< \infty)$ $\forall i$.

Next, note that $\sqrt{N_{3i}}(\overline{T_i} - E(\overline{T_i})) = \sqrt{N_{3i} V(\overline{T_i})} \left(\dfrac{\overline{T_i} - \mu_i}{\sqrt{V(\overline{T_i})}} \right) = 0_p(1) \forall i$ by A1.

That is $\overline{T_i}$ satisfies the assumptions corresponding to the multi-sample U-statistic central limit theorem. Thus,

$$\frac{\overline{T_i} - \mu_i}{\sqrt{V(\overline{T_i})}} \xrightarrow{D} N(0,1) \text{ and } N_{3i} V(\overline{T_i}) \to \sum_{l=1}^{3} \frac{1}{\lambda_{i_l}} \xi_{i_l},$$

where this quantity is essentially the variance of the corresponding projection. Hence, the numerator of Z_1 is $o_p(1)$. It now follows from A3, that Z_1 is also $o_p(1)$.

Consider Z_2 next.

$$\begin{aligned}
Z_2 &= \frac{T_{N2}}{\sqrt{\frac{1}{N^5}V(T)}} \\
&= \frac{\sqrt{N}(\overline{T_2} - E(\overline{T_2}))}{\sqrt{\frac{1}{N^5}V(T)}} \\
&= \left(\frac{\sqrt{N}\sqrt{V(\overline{T_2})}}{\sqrt{\frac{1}{N^5}V(T)}}\right)\left(\frac{\overline{T_2} - E(\overline{T_2})}{\sqrt{V(\overline{T_2})}}\right) \\
&= \left(\frac{\sqrt{NV(\overline{T_2})}}{\sqrt{\frac{1}{N^5}V(T)}}\right)\left(\frac{\overline{T_2} - E(\overline{T_2})}{\sqrt{V(\overline{T_2})}}\right).
\end{aligned}$$

Now, $$\frac{\overline{T_2} - E(\overline{T_2})}{\sqrt{V(\overline{T_2})}} \xrightarrow{D} N(0,1) \text{ by A4.}$$

Thus, if we can show

$$\frac{NV(\overline{T_2})}{\frac{1}{N^5}V(T)} \to 1,$$

then it follows that $Z_2 \xrightarrow{D} N(0,1)$. It will then follow from the result on Z_1 that $Z = Z_1 + Z_2 \xrightarrow{D} N(0,1)$.

Now,

$$
\begin{aligned}
\frac{NV(\overline{T_2})}{\frac{1}{N^5}V(T)} &= \frac{V(\sqrt{N}\overline{T_2})}{V(\frac{1}{N^{\frac{5}{2}}}T)} = \frac{V(\sqrt{N}(\overline{T_2}-E(\overline{T_2})))}{V(\frac{1}{N^{\frac{5}{2}}}(T-E(T)))} = \frac{V(T_{N2})}{V(T_N)} = \frac{V(T_N-T_{N_1})}{V(T_N)} \\
&= \frac{V(T_N)+V(T_{N1})+-2Cov(T_N,T_{N1})}{V(T_N)} = 1+\frac{V(T_{N1})}{V(T_N)} - \frac{2Cov(T_N,T_{N1})}{V(T_N)}.
\end{aligned}
$$

Note first that,

$$
0 \leq \left|\frac{-2Cov(T_N,T_{N_1})}{V(T_N)}\right| = \frac{2\left|Cov(T_N,T_{N1})\right|}{V(T_N)} \leq \frac{2\sqrt{V(T_N)V(T_{N1})}}{V(T_N)}
$$

By A3, we know that ;

$$
V(T_N) = \frac{1}{N^5}V(T) \to \upsilon\ (>0).
$$

Hence, if we can show $V(T_{N1}) \to 0$, it will follow that

$$
\frac{2Cov(T_N,T_{N1})}{V(T_N)} \text{ and } \frac{V(T_{N1})}{V(T_N)}
$$

both go to 0 as $N \to \infty$.

This in turn implies that $\frac{NV(\overline{T_2})}{\frac{1}{N^5}V(T)} \to 1$as $N \to \infty$.

With that said, note that,

$$
\begin{aligned}
V(T_{N1}) &= V(\sqrt{N}(T_1-E(T_1))) \\
&= V\left(\sum_{i\in S_1} \underbrace{\left(\frac{n_{i_1}n_{i_2}n_{i_3}}{N^3}-\lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\right)}_{a_i} \underbrace{\sqrt{\frac{N}{N_{3i}}}}_{b_i} \underbrace{\sqrt{N_{3i}}(\overline{T_i}-E(\overline{T_i}))}_{c_i}\right) \\
&= V\left(\sum_{i\in S_1} a_i b_i c_i\right)
\end{aligned}
$$

$$= \quad Cov\left(\sum_{i\in S_1} a_i b_i c_i, \sum_{j\in S_1} a_j b_j c_j\right)$$

$$= \quad \sum_{i\in S_1}\sum_{j\in S_1} Cov\left(a_i b_i c_i, a_j b_j c_j\right)$$

$$= \quad \sum_{i\in S_1}\sum_{j\in S_1} a_i a_j b_i b_j Cov(c_i, c_j)$$

$$\leq \quad \sum_{i\in S_1}\sum_{j\in S_1} |a_i a_j b_i b_j| \sqrt{V(c_i)V(c_j)}$$

Now, we have already assumed that t_i satisfies the assumptions of the multi-sample U-statistic central limit theorem (i.e A1). Hence, $V(c_i) \to \upsilon_i$ ($< \infty$), for some υ_i. Furthermore, $a_i \to 0\ \forall i$ and $b_i \to \beta_i$ ($< \infty$). Since $\sum_{i\in S_1}\sum_{j\in S_1}$ is a finite sum, it follows that $V(T_{N1}) \to 0$. This completes the proof of the theorem.

Theorem : Let T be the test statistic defined in (12) . Then, under H_0,

$$Z = \frac{T - E[T]}{\sqrt{V(T)}} \xrightarrow{D} N(0,1).$$

We will now apply the conditions A1 to A4 to T.

Proof of Asymptotic Distribution of T

To begin, we show that T belongs to the class of statistics defined in Theorem 3.5. To that end, let S_1 denote the set of tri-tuplets corresponding to the terms in T. That is, $S_1 = \left\{i : i = (i_1, i_2, i_3), i \in S_1^p(3) \cup (S_1^{p-1}(1) \times \{p\} \times S_{p+1}^k(1)) \cup S_p^k(3)\right\}$.

Next, for $i \in S_1$, let

$$T_i = \sum_{j_1=1}^{n_{i_1}}\sum_{j_2=1}^{n_{i_2}}\sum_{j_3=1}^{n_{i_3}} t_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}),$$

where

$$t_i(X_{i_1}, X_{i_2}, X_{i_3}) = \begin{cases} I(X_{i_1} < X_{i_2} < X_{i_3}),\ i \in S_1^p(3) \\ I(X_{i_1} < X_{i_2} > X_{i_3}),\ i \in S_1^{p-1}(1) \times \{p\} \times S_{p+1}^k(1) \\ I(X_{i_1} > X_{i_2} > X_{i_3}),\ i \in S_p^k(3). \end{cases}$$

Thus, the test statistic can be written as

$$T = \sum_{i \in S_1} T_i = \sum_{i \in S_1} \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=1}^{n_{i_2}} \sum_{j_3=1}^{n_{i_3}} \phi_i(X_{i_1 j_1}, X_{i_2 j_2}, X_{i_3 j_3}).$$

In addition, note that under H_0

$$E[T] = \theta = \sum_{i \in S_1} E(T_i) = \sum_{i \in S_1} \sum_{j_1=1}^{n_{i_1}} \sum_{j_2=2}^{n_{i_2}} \sum_{j_3=3}^{n_{i_3}} \theta_i = \sum_{i \in S_1} n_{i_1} n_{i_2} n_{i_3} \theta_i \text{ where}$$

$$\theta_i = \begin{cases} \frac{1}{6},\ i \in S_1^p(3) \\ \frac{1}{3},\ i \in S_1^{p-1}(1) \times \{p\} \times S_{p+1}^k(1) \\ \frac{1}{6},\ i \in S_p^k(3). \end{cases}$$

Thus, to prove the results of this theorem it suffices to show that the conditions of Theorem 3.5 are satisfied.

We begin with condition A2.

It has already been assumed that $\frac{n_i}{N} \to \lambda_i \in (0,1)$ as $N \to \infty$.

Therefore, A2 is automatically satisfied.

Next, consider condition A1. Since t_i is an indicator function for all $i \in S_1$, it follows immediately that $E[t_i^2] < \infty \; \forall \; i \in S_1$. Now let

$$\begin{aligned}
f_1(x) &= E\left(I(X_{i_1} < X_{i_2} < X_{i_3}) | X_{i_1} = x\right) - \frac{1}{6} = \frac{1}{2}(1 - F(x))^2 - \frac{1}{6} \\
&= \frac{1}{2}F^2(x) - F(x) + \frac{1}{3}, \\
f_2(x) &= E\left(I(X_{i_1} < X_{i_2} < X_{i_3}) | X_{i_2} = x\right) - \frac{1}{6} = (1 - F(x))\, F(x) - \frac{1}{6} \\
&= -F^2(x) + F(x) - \frac{1}{6}, \\
f_3(x) &= E\left(I(X_{i_1} < X_{i_2} < X_{i_3}) | X_{i_3} = x\right) - \frac{1}{6} = \frac{1}{2}F^2(x) - \frac{1}{6} \\
&= \frac{1}{2}F^2(x) + 0F(x) - \frac{1}{6}, \\
f_4(x) &= E\left(I(X_{i_1} < X_{i_2} > X_{i_3}) | X_{i_1} = x\right) - \frac{1}{3} = \frac{1}{2}(1 - F^2(x)) - \frac{1}{3} \\
&= -\frac{1}{2}F^2(x) + 0F(x) + \frac{1}{6},
\end{aligned}$$

$$
\begin{aligned}
f_5(x) &= E\left(I(X_{i_1} < X_{i_2} > X_{i_3})|X_{i_2} = x\right) - \frac{1}{3} = F^2(x) - \frac{1}{3} \\
&= F^2(x) + 0F(x) - \frac{1}{3}, \\
f_6(x) &= E\left(I(X_{i_1} < X_{i_2} > X_{i_3})|X_{i_3} = x\right) - \frac{1}{3} = \frac{1}{2}(1 - F^2(x)) - \frac{1}{3} \\
&= -\frac{1}{2}F^2(x) + 0F(x) + \frac{1}{6}, \\
f_7(x) &= E\left(I(X_{i_1} > X_{i_2} > X_{i_3})|X_{i_1} = x\right) - \frac{1}{6} = \frac{1}{2}F^2(x) - \frac{1}{6} \\
&= \frac{1}{2}F^2(x) + 0F(x) - \frac{1}{6}, \\
f_8(x) &= E\left(I(X_{i_1} > X_{i_2} > X_{i_3})|X_{i_2} = x\right) - \frac{1}{6} = \left(1 - F(x)\right)F(x) - \frac{1}{6} \\
&= -F^2(x) + F(x) - \frac{1}{6}, \\
f_9(x) &= E\left(I(X_{i_1} > X_{i_2} > X_{i_3})|X_{i_3} = x\right) - \frac{1}{6} = \frac{1}{2}(1 - F(x))^2 - \frac{1}{6} \\
&= \frac{1}{2}F^2(x) - F(x) + \frac{1}{3}.
\end{aligned}
$$

It is obvious from these results that $f_i(x)$ takes the quadratic form $a_iF^2(x)+b_iF(x)+c_i$. Furthermore there are no cases for which $a_i = b_i = 0$. It follows that $V(f_i(x)) > 0$ for all i. Hence, the maximum variance condition corresponding to the multi-sample

U-statistic theorem is satisfied in all cases. This shows that A1 is satisfied. Condition A1 is thus satisfied.

Next, consider condition A4. This condition requires us to show that the conditions of the multi-sample U-statistic theorem are satisfied for the kernel,

$$t(X_1, X_2, \ldots, X_k) = \sum_{i \in S_1} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3} t_i(X_{i_1}, X_{i_2}, X_{i_3}).$$

Once more, t is just a linear combination of indicator functions. Hence, it follows that $E[t^2] < \infty$. To show that the maximum variance condition is satisfied we consider the following cases: $T = T_R$, $T = T_A + T_R$, $T = T_L + T_A + T_R$, $T = T_L + T_A$, and $T = T_L$. Actually, it turns out that the last two cases follow from the first two cases. Hence, only the first three cases need to be considered. The derivations are similar for all three cases, so we only show the details corresponding to $T = T_A + T_R$.
To this end, we consider

$$\begin{aligned} f(x) &= E\left[t(X_1, X_2, \ldots, X_k) - E[t] | X_1 = x\right] \\ &= E\left[\sum_{i \in S_1} (\lambda_{i_1}\lambda_{i_2}\lambda_{i_3} t_i(X_{i_1}, X_{i_2}, X_{i_3}) - \theta_i) | X_1 = x\right] \\ &= \sum_{i \in S_1} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3} \left(-\frac{1}{2}F^2(x) + 0F(x) + \frac{1}{6}\right) \\ &= \left(-\frac{1}{2}\sum_{i \in S_{11}} (\lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\right) F^2(x) + \left(\frac{1}{6}\sum_{i \in S_{11}} (\lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\right) \\ &= aF^2(x) + bF(x) + c \ (say), \end{aligned}$$

where S_{11}, is the subset of S_1, where $1 \in i$ and the terms correspond to the T_A only. That is, the conditional expectations are 0 for the T_R terms. Once more, Since

$\lambda_i \in (0,1)\ \forall\ i$ and $f(x)$ has the form $aF^2(x) + bF(x) + c$, where a and b are not both equal to 0, it follows that $V(f(x)) > 0$. Hence, the "maximum varianc" condition is satisfied for this case. As previously mentioned, the other remaining cases are argued in a similar fashion.

Lastly, we show condition A3. The variance of the general case of T can be used to prove that

$$\frac{1}{N^5}V(T) \quad \to \quad K > 0.$$

Recall that,

$$\begin{aligned}\frac{1}{N^5}V(T) \quad &= \quad \frac{1}{N^5}\left(V(T_L) + V(T_A) + V(T_A) + 2Cov(T_L, T_A) + 2Cov(T_A, T_R)\right.\\ &\quad \left. +2Cov(T_L, T_R)\right)\end{aligned}$$

Consider $V(T_L)$ first. As $N \to \infty$,

$$\frac{1}{N^5}Var(T_L) \quad \to \quad \frac{2}{360}(39\lambda_{l_1} + \lambda_{l_2} + \lambda_{l_3})$$

where

$$\lambda_{l_1} = \sum_{i \in S_1^p(5)} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\lambda_{i_4}\lambda_{i_5}$$

$$\lambda_{l_2} = \sum_{i \in S_1^p(4)} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\lambda_{i_4}(9(\lambda_{i_1} + \lambda_{i_4}) + 15(\lambda_{i_2} + \lambda_{i_3}))$$

$$\lambda_{l_3} = \sum_{i \in S_1^p(3)} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}(4\lambda_{i_1}\lambda_{i_2} + 4\lambda_{i_2}\lambda_{i_3} + \lambda_{i_1}\lambda_{i_3})$$

where $\lambda_{l_1}, \lambda_{l_2}$ and λ_{l_3} are all positive constants.

Therefore,

$$\frac{1}{N^5}V(T_L) \rightarrow K_L > 0.$$

Consider $V(T_R)$ next. As $N \rightarrow \infty$,

$$\frac{1}{N^5}Var(T_R) \rightarrow \frac{2}{360}(39\lambda_{r_1} + \lambda_{r_2} + \lambda_{r_3})$$

where

$$\lambda_{r_1} = \sum_{i \in S_p^k(5)} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\lambda_{i_4}\lambda_{i_5}$$

$$\lambda_{r_2} = \sum_{i \in S_p^k(4)} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\lambda_{i_4}(9(\lambda_{i_1} + \lambda_{i_4}) + 15(\lambda_{i_2} + \lambda_{i_3}))$$

$$\lambda_{r_3} = \sum_{i \in S_p^k(3)} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}(4\lambda_{i_1}\lambda_{i_2} + 4\lambda_{i_2}\lambda_{i_3} + \lambda_{i_1}\lambda_{i_3})$$

where $\lambda_{r_1}, \lambda_{r_2}$ and λ_{r_3} are all positive constants.

Therefore,

$$\frac{1}{N^5}V(T_R) \rightarrow K_R > 0.$$

Next, consider $V(T_A)$ as $N \rightarrow \infty$. Here,

$$\frac{1}{N^5}V(T_A) \rightarrow \frac{1}{45}\left(\lambda_{i_1}^2\lambda_p^2\lambda_{i_3} + \lambda_{i_1}\lambda_p^2\lambda_{i_3}^2 + \lambda_{i_1}^2\lambda_p\lambda_{i_3}^2\right)$$

where λ_{i_1}, λ_p and λ_{i_3} are all positive constants.

Therefore, $\frac{1}{N^5}V(T_A) \rightarrow K_A > 0$

Next, consider $Cov(T_L, T_A)$.

$$\frac{1}{N^5}Cov(T_L, T_A) \rightarrow \frac{1}{360}(48\lambda_{b_1} + 8\lambda_{b_2} + 16\lambda_{b_3} + 16\lambda_{b_4})$$

where

$$\lambda_{b_1} = \sum_{i \in S_1^{p-1}(3)} \sum_{i'=p+1}^{k} \lambda_{i_1}\lambda_{i_2}\lambda_{i_3}\lambda_p\lambda_{i'_3}$$

$$\lambda_{b_2} = \sum_{i \in S_1^{p-1}(2)} \sum_{i'=p+1}^{k} \lambda_{i_1}\lambda_{i_2}\lambda_p^2\lambda_{i'_3}$$

$$\lambda_{b_3} = \sum_{i \in S_1^{p-1}(2)} \sum_{i'=p+1}^{k} \lambda_{i_1}^2\lambda_{i_2}\lambda_p\lambda_{i'_3}$$

$$\lambda_{b_4} = \sum_{i \in S_1^{p-1}(2)} \sum_{i'=p+1}^{k} \lambda_{i_1}\lambda_{i_2}^2\lambda_p\lambda_{i'_3}$$

where $\lambda_{b_1}, \lambda_{b_2}, \lambda_{b_3}$ and λ_{b_4} are all positive constants.

Therefore, $\frac{1}{N^5}Cov(T_L, T_A) \rightarrow K_{LA} > 0$

Next, consider $Cov(T_A, T_R)$.

$$\frac{1}{N^5}Cov(T_A, T_R) \rightarrow \frac{1}{360}(48\lambda_{c_1} + 8\lambda_{c_2} + 16\lambda_{c_3} + 16\lambda_{c_4})$$

where

$$\lambda_{c_1} = \sum_{i=1}^{p-1} \sum_{i \in S_{p+1}^k(3)} \lambda_{i_1}\lambda_p\lambda_{i'_1}\lambda_{i'_2}\lambda_{i'_3}$$

$$\lambda_{c_2} = \sum_{i=1}^{p-1} \sum_{i \in S_{p+1}^k(2)} \lambda_{i_1}\lambda_p^2\lambda_{i'_2}\lambda_{i'_3}$$

$$\lambda_{c_3} = \sum_{i=1}^{p-1} \sum_{i \in S_{p+1}^k(2)} \lambda_{i_1}\lambda_p\lambda_{i'_2}\lambda_{i'_3}^2$$

$$\lambda_{c_4} = \sum_{i=1}^{p-1} \sum_{i \in S_{p+1}^k(2)} \lambda_{i_1}\lambda_p\lambda_{i'_2}^2\lambda_{i'_3}$$

where $\lambda_{c_1}, \lambda_{c_2}, \lambda_{c_3}$ and λ_{c_4} are all positive constants.

Therefore, $\frac{1}{N^5} Cov(T_A, T_R) \to K_{AR} > 0$.

Lastly, consider $Cov(T_L, T_R)$.

$$\frac{1}{N^5} Cov(T_R, T_L) \to \sum_{i \in S_1^p(3)} \sum_{i' \in S_p^k(3)} \lambda_{i_1}\lambda_{i_2}\lambda_{i_p}\lambda_{i'_2}\lambda_{i'_3}\frac{1}{45} \text{ for } i_3 = i'_1 = i_p.$$

Therefore,

$$\frac{1}{N^5} Cov(T_L, T_R) \to K_{LR} > 0.$$

Since all the components of the variance of T converge to a positive constant as $N \to \infty$, we can therefore conclude that condition A3 holds. This completes the proof of the theorem.

CHAPTER 4. A FINITE SAMPLE SIMULATION STUDY

4.1. Study Design

This chapter presents the results of a finite sample simulation study which compares the estimated powers of seven tests; namely, the new test being proposed in this dissertation (LAR), Mack-Wolfe's test (MW), the test proposed by Salman (SAL), Bhat's test, two tests proposed by Hettmansperger and Norton (for one equally spaced location parameters (HNEQ) and the optimal test (HNOPT)) and the Kruskal-Wallis test.

The simulation study focused around location parameters. Thus the model is given by $X_{ij} = \mu_i + \epsilon_{ij}$, where ϵ_{ij} are IID according to some continuous CDF F and μ_i is a location parameter. Note that μ_i does not necessarily represent the mean of X_{ij}. For the exponential distribution, μ_i represents a scale parameter. The null hypothesis for all the tests is that of equal population location parameters as defined in equation (11). The alternative hypothesis being tested by LAR, MW, SAl, Bhat, HNEQ and HNOPT is the umbrella hypothesis in Equation (3), while the alternative hypothesis being tested by the KW test is differences in group means. All tests are based on a significance level of $\alpha = 0.05$. The estimated power values (i.e. the proportion of times H_0 is rejected) are derived from 10,000 simulations. The relative powers were studied under the null and location shift alternatives. The study considered two different distributions, namely, the normal and exponential distributions. The simulation study also requires values for $k; n_1, n_2, \ldots n_k; \mu_1, \mu_2, \ldots, \mu_k$. Theoretically, there is an infinite number of k, sample size, location parameter, and distribution configurations. This particular simulation study considered $k = 3, 4, 5, 6$ and sample sizes for the k groups were chosen for balanced designs and randomly (with replacement from $1, 2, \ldots, 25$) for the unbalanced design. The location configurations were chosen to reflect either an umbrella patterns or a monotone trend. The simulation study investigated equally and unequally spaced location parameters. The study also looked at the effect of the wideness of the space between location parameters.

4.2. Simulation Results

4.2.1. Simulating under $N(\mu, 1)$

Equal variances are assumed for the results in Tables 10, 12, 14, 15 and 16. Table 10 shows the estimated Type I error rates. Notice that the table does not contain the HNOPT test. The reason being $c_i = c \ \forall i$ under H_o. This causes the HNOPT test, Equation (10) to be degenerate at zero under the null hypothesis. With 10,000 simulations and $\alpha = 0.05$, we see that each test in Table 10 with the exception of LAR and SAL which are slightly liberal, exhibits a power of approximately 0.05. Consequently, the tests: MW, Bhat, and HNEQ are equivalent in terms of the Type I error rates. KW is most conservative compared to all the tests. Figure 5 shows the distribution of T when simulated for three groups where $n_1 = n_2 = n_3$. It can be seen from Figure 5 that the distribution of T is right skewed which explains the reason for LAR being liberal.

4.2.2. Simulating under $N(\mu_i, 1)$

Recall from chapter 1 that the SAL test is a special case of the LAR test when $k = 3$ and $p = 2$. Power estimates of the LAR test and SAL test will be the same under this circumstance. Bhat's test also becomes a special case of the MW test when $k = 3$ and $p = 2$. Thus power estimates of MW test and Bhat test will be the same under this condition. The HNEQ and HNOPT are exactly the same test when the spacing of the location parameter between the groups are equal and will yield the same power estimates under this condition.

Terpstra *et. al.* (2011) note that the HNOPT test can be viewed as a benchmark for comparison purposes. This is because the HNOPT requires the proper specification of the spacings between location parameters, which we know in this context. This explains why the HNOPT test is typically the most powerful test among the seven tests being considered. However the power estimates of the other six tests are comparable.

The simulation study examined power estimates for a range of scenarios. The outcome of the simulation study depicted some differences in power under certain scenarios. Generally, the power estimates of all the tests increased with $\mu_{i+1} - \mu_i$ (left of peak) or $\mu_i - \mu_{i+1}$ (right of peak) and k. The power estimates were always close to 1 for any $\mu_{i+1} - \mu_i > \sigma^2$ (left of peak) or $\mu_i - \mu_{i+1} > \sigma^2$ (right of peak) . The LAR generally performed better than all the other tests when the peak of the umbrella was very shallow, i.e. when the differences in location parameters between adjacent groups is very small. The study also showed that when the differences between the location parameters of adjacent groups are all less than σ^2, the LAR performed far better than all the other tests including HNOPT. It should be noted that LAR generally performed better than all other tests including the HNOPT at small sample sizes and small k. The power estimates of the LAR test were higher than those of the KW test for almost every case. There were no evident differences in the power estimates of the balanced and unbalanced cases.

One of the methods that is proposed to correct the liberal nature of the test statistic is the gamma approximation .

$$\text{The shape parameter of } T \text{ is given by } \frac{\mu^2}{\sigma^2}$$
$$\text{while the scale parameter is given by } \frac{\sigma^2}{\mu}.$$

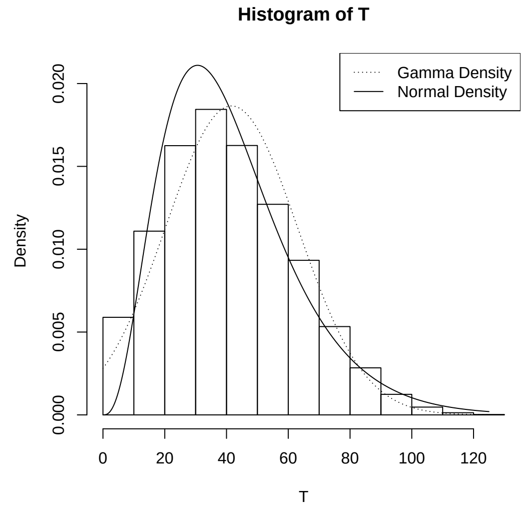

Figure 5: Graph Showing Density of T

After the gamma approximation, simulation results under the null distribution showed T as conservative. These results are shown in table 11.

Lloyd (2005) also states that it is naive to compare power values of a liberal test to that of a conservative test. Since the LAR and SAL are liberal while the other tests are conservative, it is there not fair in this lines to compare them. Lloyd (2005) proposed methods for adjusting for size in order to achieve a fair comparison of power estimates between liberal and conservative tests. New power estimates that are adjusted for size were computed using the original simulated power values. The adjusted power is given by;

$$R(\delta) = \Phi\left(\Phi^{-1}(\beta) - \Phi^{-1}(\alpha) + \Phi(\alpha^*)\right)$$

where

$\beta =$ the estimated power value,

$\alpha =$ the simulated power under the null,

$\alpha^* =$ the $\alpha-$level used in simulation.

Table 13 and Table 17 show the estimates of power values adjusted for size for the normal and exponential distributions respectively. After adjusting for size, LAR still generally performed better than the other tests for small differences between the groups and for cases where there is a gradual increases and then a sudden sharp increase toward the peak. Contrary to the results of the raw power values, LAR did not perform better for equally spaced location parameters.

Table 10: Estimated α Levels for $k = 3, 4, 5$ under the Null Distribution

n_1	n_2	n_3	n_4	n_5	μ_1	μ_2	μ_3	μ_4	μ_5	Peak	LAR	MW	SAL	Bhat	HNEQ	KW
2	2	2			0	0	0			2	0.0697	0.0697	0.0697	0.0697	0.0697	0.0000
3	3	3			3	3	3			2	0.0569	0.0462	0.0569	0.0462	0.0462	0.0092
3	3	3			0	0	0			2	0.0581	0.0467	0.0581	0.0467	0.0467	0.0105
5	5	5			0	0	0			2	0.0661	0.0480	0.0661	0.0480	0.0480	0.0441
10	10	10			0	0	0			2	0.0581	0.0484	0.0581	0.0484	0.0484	0.0490
15	15	15			0	0	0			2	0.0597	0.0510	0.0597	0.0510	0.0510	0.0444
3	5	3			0	0	0			2	0.0634	0.0378	0.0634	0.0378	0.0378	0.0379
7	5	2			0	0	0			2	0.0639	0.0589	0.0639	0.0589	0.0523	0.0304
9	11	12			0	0	0			2	0.0563	0.0509	0.0563	0.0509	0.0483	0.0477
9	8	17			0	0	0			2	0.0561	0.0477	0.0561	0.0477	0.0470	0.0451
5	5	5	5		0	0	0	0		2	0.0621	0.0488	0.0588	0.0415	0.0511	0.0369
10	10	10	10		0	0	0	0		2	0.0573	0.0522	0.0565	0.0508	0.0511	0.0461
15	15	15	15		0	0	0	0		2	0.0557	0.0477	0.0560	0.0474	0.0494	0.0471
5	5	5	5		0	0	0	0		3	0.0604	0.0497	0.0610	0.0451	0.0513	0.0373
7	3	7	9		0	0	0	0		3	0.0559	0.0462	0.0581	0.0478	0.0497	0.0398
4	4	4	4	4	0	0	0	0	0	2	0.0627	0.0491	0.0610	0.0416	0.0496	0.0285
5	5	5	5	5	0	0	0	0	0	2	0.0586	0.0485	0.0577	0.0512	0.0538	0.0352
5	5	5	5	5	0	0	0	0	0	3	0.0609	0.0507	0.0600	0.0561	0.0490	0.0337
10	10	10	10	10	0	0	0	0	0	3	0.0603	0.0493	0.0605	0.0526	0.0498	0.0457
10	10	10	10	10	0	0	0	0	0	4	0.0592	0.0493	0.0593	0.0530	0.0508	0.0480
15	15	15	15	15	0	0	0	0	0	3	0.0557	0.0510	0.0551	0.0473	0.0526	0.0505
5	10	5			0	0	0			2	0.0573	0.0500	0.0573	0.0500	0.0500	0.0425
17	10	19			0	0	0			2	0.0536	0.0443	0.0536	0.0443	0.0438	0.0451
18	13	10	15		0	0	0	0		2	0.0557	0.0470	0.0554	0.0499	0.0500	0.0477
19	19	20	15		0	0	0			3	0.0578	0.0509	0.0585	0.0513	0.0488	0.0466
21	22	23	20	15	0	0	0	0	0	3	0.0551	0.0475	0.0531	0.0493	0.0481	0.0456
25	25	25	25	25	0	0	0	0	0	3	0.0532	0.0488	0.0522	0.0494	0.0490	0.0462
15	10	15	10	15	0	0	0	0	0	5	0.0643	0.0520	NA	0.0511	0.0511	0.0423
15	20	15	20	15	0	0	0	0	0	5	0.0555	0.0492	NA	0.0464	0.0486	0.0454

Table 11: Simulating Under the Null after Gamma Approximation

n_1	n_2	n_3	n_4	n_5	Peak	LAR	MW	SAL	Bhat	HNEQ	KW
2	2	2			2	0.0662	0.0662	0.0662	0.0662	0.0662	0.0000
3	3	3			2	0.0493	0.0493	0.0606	0.0493	0.0493	0.0105
5	5	5			2	0.0400	0.0486	0.0682	0.0486	0.0486	0.0446
10	10	10			2	0.0423	0.0472	0.0581	0.0472	0.0472	0.0436
15	15	15			2	0.0413	0.0475	0.0549	0.0475	0.0475	0.0475
3	5	3			2	0.0429	0.0389	0.0660	0.0389	0.0389	0.0414
7	5	2			2	0.0377	0.0548	0.0636	0.0548	0.0501	0.0305
9	11	12			2	0.0487	0.0586	0.0634	0.0586	0.0565	0.0476
9	8	17			2	0.0412	0.0481	0.0568	0.0481	0.0484	0.0463
5	10	5			2	0.0422	0.0506	0.0569	0.0506	0.0506	0.0451
17	10	19			2	0.0419	0.0493	0.0580	0.0493	0.0492	0.0476
5	5	5	5		2	0.0425	0.0514	0.0637	0.0463	0.0543	0.0391
10	10	10	10		2	0.0437	0.0525	0.0572	0.0518	0.0494	0.0437
15	15	15	15		2	0.0424	0.0493	0.0535	0.0485	0.0500	0.0467
7	3	7	9		3	0.0420	0.0531	0.0617	0.0517	0.0463	0.0377
18	13	10	15		2	0.0414	0.0518	0.0554	0.0537	0.0491	0.0504
19	19	20	15		3	0.0415	0.0472	0.0555	0.0471	0.0476	0.0459
4	4	4	4	4	2	0.0414	0.0513	0.0609	0.0392	0.0499	0.0285
5	5	5	5	5	2	0.0461	0.0505	0.0622	0.0548	0.0514	0.0369
5	5	5	5	5	3	0.0391	0.0489	0.0593	0.0510	0.0471	0.0372
10	10	10	10	10	3	0.0436	0.0480	0.0577	0.0507	0.0487	0.0447
15	15	15	15	15	3	0.0405	0.0448	0.0539	0.0455	0.0450	0.0482

Table 12: Estimated Power Values for $k = 3, 4$ under $N(\mu_i, 1)$

n_1	n_2	n_3	n_4	n_5	n_6	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	Peak	LAR	MW	SAL	Bhat	HNEQ	HNOPT	KW
3	3	3				0.0	3.0	0.0				2	0.979	0.970	0.979	0.970	0.970	0.970	0.169
3	3	3				1.0	2.0	0.0				2	0.548	0.509	0.548	0.509	0.509	0.646	0.122
3	3	3				0.0	1.0	0.0				2	0.376	0.336	0.376	0.336	0.336	0.336	0.036
5	5	5				0.0	3.0	0.0				2	1.000	0.999	1.000	0.999	0.999	0.999	0.983
5	5	5				0.0	1.0	0.0				2	0.559	0.515	0.559	0.515	0.515	0.515	0.253
5	5	5				0.0	0.3	0.0				2	0.134	0.108	0.134	0.108	0.108	0.108	0.055
5	5	5				0.0	0.5	0.0				2	0.253	0.210	0.253	0.210	0.210	0.210	0.088
5	5	5				0.0	0.5	0.3				2	0.178	0.150	0.178	0.150	0.150	0.172	0.079
5	5	5				1.0	2.0	1.0				2	0.565	0.512	0.565	0.512	0.512	0.512	0.245
10	10	10				1.0	2.0	1.0				2	0.796	0.786	0.796	0.786	0.786	0.786	0.535
10	10	10				1.0	2.0	0.0				2	0.950	0.970	0.950	0.970	0.970	0.995	0.959
10	10	10				0.3	1.0	0.5				2	0.472	0.451	0.472	0.451	0.451	0.493	0.258
10	10	10				2.0	2.3	0.0				2	0.648	0.850	0.648	0.850	0.850	1.000	0.997
10	10	10				0.0	3.0	1.0				2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15	15	15				0.3	0.7	0.0				2	0.511	0.502	0.511	0.502	0.502	0.572	0.339
15	15	15				0.0	0.3	0.0				2	0.206	0.196	0.206	0.196	0.196	0.196	0.095
15	15	15				3.0	4.0	2.0				2	0.991	0.997	0.991	0.997	0.997	1.000	0.997
15	15	15				1.0	2.0	1.0				2	0.918	0.915	0.918	0.915	0.915	0.915	0.758
5	5	5	5			0.2	1.0	0.1	0.2			2	0.449	0.356	0.467	0.331	0.304	0.448	0.154
5	5	5	5			0.5	1.0	2.0	1.5			3	0.575	0.600	0.450	0.485	0.718	0.756	0.362
5	5	5	5			0.3	0.5	2.0	0.5			3	0.896	0.851	0.882	0.816	0.799	0.889	0.518
5	5	5	5			0.5	2.0	0.1	0.0			2	0.930	0.897	0.918	0.838	0.881	0.952	0.666
5	5	5	5			1.0	3.0	2.0	1.0			2	0.938	0.949	0.920	0.943	0.877	0.963	0.715
5	5	5	5			1.0	1.1	3.0	2.9			3	0.484	0.596	0.320	0.377	0.864	0.990	0.862

Table 13: Power Values Adjusted for Size $R(\delta)$ under $N(\mu_i, 1)$

n_1	n_2	n_3	n_4	n_5	μ_1	μ_2	μ_3	μ_4	μ_5	Peak	LAR	MW	SAL	Bhat	HNEQ	KW
3	3	3			0.0	0.0	0.0			2	0.0581	0.0467	0.0581	0.0467	0.0467	0.0105
3	3	3			0.0	3.0	0.0			2	0.9750	0.9722	0.9750	0.9722	0.9722	0.3840
3	3	3			0.0	1.0	0.0			2	0.3483	0.3481	0.3483	0.3481	0.3481	0.1280
5	5	5			0.0	0.0	0.0			2	0.0661	0.0480	0.0661	0.0480	0.0480	0.0441
5	5	5			0.0	0.5	0.0			2	0.2106	0.2157	0.2106	0.2157	0.2157	0.0980
5	5	5			1.0	2.0	1.0			2	0.5097	0.5198	0.5097	0.5198	0.5198	0.2643
10	10	10			0.0	0.0	0.0			2	0.0581	0.0484	0.0581	0.0484	0.0484	0.0490
10	10	10			1.0	2.0	1.0			2	0.7744	0.7906	0.7744	0.7906	0.7906	0.5389
10	10	10			1.0	2.0	0.0			2	0.9419	0.9711	0.9419	0.9711	0.9711	0.9598
5	5	5	5		0.0	0.0	0.0	0.0		2	0.0621	0.0488	0.0588	0.0415	0.0511	0.0369
5	5	5	5		0.2	1.0	0.1	0.2		2	0.4069	0.3604	0.4354	0.3637	0.3003	0.1904
5	5	5	5		0.5	2.0	0.1	0.0		2	0.9144	0.8991	0.9052	0.8588	0.8789	0.7163
10	10	10	10		0.0	0.0	0.0	0.0		2	0.0573	0.0522	0.0565	0.0508	0.0511	0.0461
10	10	10	10		0.0	0.5	0.1	0.0		2	0.3284	0.3140	0.3240	0.3206	0.2649	0.1417
10	10	10	10		0.0	1.0	0.0	0.0		2	0.8093	0.7643	0.8174	0.7687	0.6584	0.5317
10	10	10	10	10	0.0	0.0	0.0	0.0	0.0	3	0.0603	0.0493	0.0605	0.0526	0.0498	0.0457
10	10	10	10	10	0.0	0.3	1.0	0.3	0.0	3	0.7655	0.7930	0.7457	0.7665	0.7536	0.4370
10	10	10	10	10	0.0	0.1	1.0	0.1	0.0	3	0.8099	0.7890	0.8062	0.7594	0.7107	0.4621
3	3	10			0.0	0.0	0.0			2	0.0609	0.0516	0.0609	0.0516	0.0504	0.0341
3	3	10			0.0	0.1	0.0			2	0.0659	0.0702	0.0659	0.0702	0.0664	0.0507
3	3	3			1.0	2.7	2.0			2	0.4255	0.3612	0.4255	0.3612	0.3026	0.3442
6	3	5	4		0.0	0.0	0.0	0.0		2	0.0652	0.0577	0.0652	0.0576	0.0574	0.0337
6	3	5	4		0.0	5.0	3.0	2.5		2	0.9845	0.9592	0.9886	0.9570	0.2778	0.9995
6	3	5	4		0.0	2.0	0.1	0.0		2	0.8804	0.7779	0.8914	0.7439	0.5654	0.3891
7	6	9	12		0.0	0.0	0.0	0.0		2	0.0580	0.0466	0.0582	0.0462	0.0490	0.0427
7	6	9	12		0.0	2.0	0.5	0.0		2	0.9914	0.9807	0.9855	0.9594	0.9412	0.9035
7	6	9	12		0.0	0.9	0.1	0.0		2	0.5760	0.4956	0.5635	0.4442	0.4274	0.2829
3	6	4	3		0.0	0.0	0.0	0.0		3	0.0668	0.0583	0.0672	0.0425	0.0510	0.0297
3	6	4	3		0.5	1.0	2.0	0.0		3	0.7194	0.6961	0.7139	0.7072	0.5352	0.4477
3	6	4	3		0.0	0.1	0.4	0.0		3	0.1543	0.1535	0.1549	0.1553	0.1475	0.0730

Table 14: Estimated Power Values for $k = 4, 5, 6$ under $N(\mu_i, 1)$

n_1	n_2	n_3	n_4	n_5	n_6	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	Peak	LAR	MW	SAL	Bhat	HNEQ	HNOPT	KW
3	3	3	3	3		1.0	1.1	1.2	1.1	1.0		3	0.096	0.087	0.102	0.088	0.083	0.083	0.023
5	5	5	5	5		1.0	1.1	1.2	1.1	1.0		3	0.104	0.098	0.099	0.100	0.099	0.099	0.039
10	10	10	10	10		1.0	1.1	1.2	1.1	1.0		3	0.127	0.126	0.121	0.130	0.130	0.130	0.060
5	5	5	5	5		1.0	1.1	4.0	1.1	1.0		3	1.000	0.998	1.000	0.962	0.978	1.000	0.960
10	10	10	10			1.0	2.0	1.0	0.0			2	0.973	0.992	0.919	0.978	0.995	0.995	0.941
10	10	10	10			0.0	0.5	0.1	0.0			2	0.353	0.322	0.346	0.323	0.268	0.334	0.133
10	10	10	10			1.3	1.5	1.0	0.5			2	0.444	0.546	0.317	0.424	0.683	0.742	0.410
10	10	10	10			0.0	1.0	0.0	0.0			2	0.827	0.771	0.833	0.771	0.662	0.839	0.516
10	10	10	10	10		1.0	2.0	3.0	2.0	1.0		3	0.996	1.000	0.990	0.999	1.000	1.000	0.984
10	10	10	10	10		1.0	1.1	4.0	1.1	1.0		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	10	10	10	10		0.0	0.3	2.0	0.3	0.0		3	1.000	1.000	1.000	0.998	0.997	1.000	0.983
10	10	10	10	10		0.0	0.3	1.0	0.3	0.0		3	0.793	0.791	0.775	0.774	0.753	0.795	0.420
10	10	10	10	10		0.0	0.1	3.0	0.1	0.0		3	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	10	10	10	10		0.0	0.1	1.0	0.1	0.0		3	0.834	0.787	0.831	0.767	0.710	0.828	0.445
10	10	10	10	10		0.0	0.1	0.5	0.1	0.0		3	0.366	0.345	0.353	0.341	0.318	0.354	0.120
5	5	5	5	5		1.0	1.0	4.0	1.0	1.0		3	1.000	0.998	1.000	0.943	0.968	1.000	0.968
10	10	10	10	10		1.0	1.0	4.0	1.0	1.0		3	1.000	1.000	1.000	0.999	1.000	1.000	1.000
5	5	5	5	5		1.0	1.1	1.2	1.3	1.0		4	0.153	0.129	0.141	0.136	0.108	0.136	0.048
10	10	10	10	10		1.0	1.1	1.2	1.3	1.0		4	0.193	0.180	0.179	0.185	0.142	0.198	0.075
10	10	10	10	10	10	0.0	0.1	0.5	0.4	0.1	0.0	3	0.347	0.384	0.296	0.328	0.344	0.435	0.141
10	10	10	10	10	10	0.0	0.1	0.5	0.2	0.1	0.0	3	0.363	0.338	0.341	0.322	0.285	0.350	0.111
10	10	10	10	10	10	0.0	0.1	1.0	0.2	0.1	0.0	3	0.837	0.776	0.827	0.757	0.640	0.827	0.405
10	10	10	10	10	10	0.0	0.1	0.5	0.2	0.1	0.0	3	0.365	0.344	0.337	0.330	0.291	0.357	0.115

Table 15: Estimated Power Values for $k = 3, 4, 5$ under $N(\mu_i, 1)$

n_1	n_2	n_3	n_4	n_5	n_6	μ_1	μ_2	μ_3	μ_4	μ_5	Peak	LAR	MW	SAL	Bhat	HNEQ	HNOPT	KW
5	5	9				0.0	2.0	3.0			2	0.009	0.008	0.009	0.008	0.006	0.999	0.987
5	7	3				1.0	3.0	1.7			2	0.476	0.542	0.476	0.542	0.537	0.950	0.768
7	13	15				1.0	1.5	0.5			2	0.617	0.729	0.617	0.729	0.733	0.805	0.581
16	8	11				2.0	2.8	0.8			2	0.889	0.888	0.889	0.888	0.881	0.994	0.961
12	4	4				3.0	3.0	0.5			2	0.216	0.123	0.216	0.123	0.076	0.991	0.921
13	6	14				2.0	3.5	1.5			2	0.979	0.977	0.979	0.977	0.978	0.988	0.923
2	14	8				0.5	8.0	0.0			2	0.321	0.460	0.321	0.460	0.460	0.507	0.270
3	3	10				1.0	2.7	2.0			2	0.464	0.367	0.464	0.367	0.304	0.606	0.281
14	15	13				3.0	6.0	4.0			2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	5	15				0.0	5.0	0.0			2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	3	5	4			0.0	5.0	3.0	2.5		2	0.989	0.965	0.992	0.963	0.301	1.000	0.999
8	8	5	5			1.9	2.5	2.0	1.8		2	0.410	0.379	0.403	0.371	0.346	0.389	0.149
7	6	9	12			0.0	2.0	0.5	0.0		2	0.993	0.979	0.988	0.956	0.940	0.991	0.890
9	9	9	3			0.7	1.0	0.0	0.0		2	0.458	0.465	0.429	0.425	0.517	0.717	0.374
9	9	9	3			0.7	1.0	1.0	0.0		2	0.221	0.230	0.167	0.175	0.299	0.459	0.156
10	12	14	16			0.5	1.0	0.5	0.0		2	0.686	0.778	0.548	0.726	0.812	0.812	0.512
11	12	15	16			0.5	4.0	0.5	0.0		2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	6	4	3			0.5	1.0	2.0	0.0		3	0.766	0.722	0.762	0.680	0.539	0.808	0.355
5	10	11	12			0.0	1.0	0.9	0.1		3	0.381	0.486	0.334	0.432	0.461	0.820	0.515

Table 16: Estimated Power Values for $k = 3, 4, 5$ under $N(\mu_i, 1)$ for $p = 1$ and $p = k$

n_1	n_2	n_3	n_4	n_5	μ_1	μ_2	μ_3	μ_4	μ_5	Peak	LAR	MW	Bhat	HNEQ	HNOPT	KW
3	3	3			2	1.5	1.0			1	0.3378	0.3146	0.1993	0.2580	0.2580	0.0303
3	3	3			3	2.0	1.0			1	0.6987	0.7063	0.5284	0.6389	0.6389	0.1134
5	5	5			2	1.5	1.0			1	0.4214	0.4040	0.3629	0.4202	0.4202	0.1892
5	5	5			3	2.0	1.0			1	0.8619	0.8791	0.8379	0.8866	0.8866	0.6415
5	5	5			1.5	1.2	1.0			1	0.2039	0.1734	0.1597	0.1809	0.1886	0.0796
10	10	10			1.8	1.0	0.0			1	0.9694	0.9845	0.9767	0.9857	0.9855	0.9217
10	10	10			1.8	1.6	0.0			1	0.9061	0.9766	0.9456	0.9806	0.9935	0.9564
3	3	3			0	0.3	0.5			3	0.6241	0.6575	0.4554	0.5986	0.7406	0.1069
3	3	3			0	0.3	2.0			3	0.1771	0.1543	0.0904	0.1201	0.1410	0.0157
3	3	3			0	0.1	3.0			3	0.6614	0.7775	0.5030	0.7572	0.9683	0.1782
5	5	5			0.1	0.2	0.4			3	0.1357	0.1059	0.0995	0.1109	0.1185	0.0564
10	10	10			0.1	0.2	0.4			3	0.1751	0.1561	0.1476	0.1624	0.1637	0.0764
15	15	15			0.1	0.2	0.7			3	0.4340	0.4601	0.4332	0.4617	0.5140	0.2879
5	5	5	5		2	1.5	1.0	0.3		1	0.8305	0.8445	0.7145	0.8507	0.8497	0.4828
5	5	5	5		5	4.0	3.0	2.0		1	0.9958	0.9980	0.9793	0.9980	0.9980	0.9432
10	10	10	10		2	1.5	1.0	0.5		1	0.9461	0.9572	0.9164	0.9590	0.9590	0.7825
10	10	10	10		4	1.5	1.0	1.0		1	0.9981	0.9999	0.9984	1.0000	1.0000	1.0000
10	10	10	10		1.6	1.5	1.0	1.0		1	0.4525	0.4593	0.3375	0.4622	0.5141	0.2317
10	10	10	10		3	1.3	1.2	1.0		1	0.9634	0.9847	0.9726	0.9892	0.9988	0.9823
10	10	10	10		1	0.8	0.5	0.3		1	0.5215	0.5214	0.4606	0.5229	0.5229	0.2335
10	10	10	10	10	0.1	0.3	0.5	0.8	1.2	5	0.8167	0.8174	0.6927	0.8217	0.8328	0.4696
10	10	10	10	10	0.1	0.3	0.5	0.7	2	5	0.9834	0.9847	0.9725	0.9877	0.9983	0.9452
10	10	10	10	10	0.1	0.3	0.5	0.7	1	5	0.6714	0.6758	0.5445	0.6818	0.6856	0.3127
5	5	5	5	5	0.1	0.3	0.3	0.5	1	5	0.3883	0.3715	0.3313	0.3772	0.4212	0.1225
5	5	5	5	1	0.1	0.3	0.3	0.5	1	5	0.2310	0.1893	0.1643	0.1825	0.2066	0.0396
10	8	5	12	16	0.1	0.3	0.3	0.5	1	5	0.7035	0.7449	0.5287	0.7130	0.7672	0.4007
9	8	14	17	16	0.1	0.2	0.3	0.4	0.5	5	0.2795	0.2624	0.1887	0.2652	0.2652	0.0976

4.2.3. Simulating under $E(\theta_i)$

Note that the scale parameter for this exponential distribution is θ_i . Also, the mean of each group will be equal to θ_i while the variance is θ_i^2. LAR performed better than all other tests including HNOPT for equally spaced θ_i of groups. Generally, the LAR performed better than the other tests including the HNOPT when the differences between the θ_i's are small. The MW test performed better than the LAR when the differences in θ_i between the groups are unequal and very wide.

Table 17: Powers Adjusted for Size $R(\delta)$ under $E(\theta_i)$

n_1	n_2	n_3	n_4	n_5	μ_1	μ_2	μ_3	μ_4	μ_5	Peak	LAR	MW	SAL	Bhat	HNEQ	KW
5	5	5			0.0	0.0	0.0			2	0.0606	0.0485	0.0592	0.0432	0.0511	0.0398
5	5	5			1.0	2.0	1.0			2	0.3105	0.2796	0.3146	0.2985	0.2712	0.1376
5	5	5			1.0	1.5	1.0			2	0.1701	0.1533	0.1731	0.1667	0.1474	0.0834
5	5	5			0.1	0.4	0.2			2	0.4747	0.4450	0.4794	0.4669	0.4350	0.3053
5	5	5			0.1	0.3	0.2			2	0.3139	0.2971	0.3181	0.3164	0.2884	0.2209
10	10	10			0.0	0.0	0.0			2	0.0582	0.0479	0.0582	0.0479	0.0479	0.0469
10	10	10			0.1	0.4	0.2			2	0.7042	0.6892	0.7042	0.6892	0.6892	0.5792
10	10	10			1.0	2.0	1.0			2	0.4863	0.4553	0.4863	0.4553	0.4553	0.2443
10	10	10			1.0	3.0	1.0			2	0.7942	0.7456	0.7942	0.7456	0.7456	0.5081
5	5	5	5		0.0	0.0	0.0	0.0		3	0.0624	0.0506	0.0605	0.0461	0.0519	0.0412
5	5	5	5		0.1	0.3	0.2	0.1		2	0.4570	0.4802	0.4379	0.4787	0.4001	0.2347
5	5	5	5		1.0	1.5	3.2	1.0		3	0.5522	0.5136	0.5382	0.5108	0.4282	0.2379
5	5	5	5		1.0	1.1	2.0	1.0		3	0.5336	0.4969	0.5293	0.4924	0.4181	0.2290
10	10	10	10		0.0	0.0	0.0	0.0		2	0.0572	0.0530	0.0557	0.0528	0.0509	0.0458
10	10	10	10		1.0	1.0	2.0	1.0		3	0.8119	0.7661	0.8185	0.7622	0.6647	0.5432
10	10	10	10		1.0	2.0	1.0	0.0		2	0.9652	0.9893	0.9041	0.9725	0.9947	0.9442
5	5	5	5	5	0.0	0.0	0.0	0.0	0.0	3	0.0585	0.0492	0.0582	0.0526	0.0479	0.0340
5	5	5	5	5	1.0	1.0	1.5	2.0	1.0	4	0.5592	0.5463	0.5126	0.5012	0.4194	0.2498
5	5	5	5	5	1.0	1.1	1.5	2.0	1.0	4	0.5330	0.5270	0.4875	0.4973	0.4002	0.2327
5	5	5	5	5	1.0	1.0	1.4	1.1	1.0	3	0.1780	0.1723	0.1777	0.1621	0.1575	0.0714

Table 18: Estimated Power Values for $k = 3, 4$ under $E(\theta_i)$

n_1	n_2	n_3	n_4	n_5	n_6	μ_1	μ_2	μ_3	μ_4	μ_5	Peak	LAR	MW	SAL	Bhat	HNEQ	HNOPT	KW
5	5	5				1.0	2.0	1.0			2	0.3448	0.2747	0.3448	0.2747	0.2747	0.2747	0.1152
5	5	5				1.0	1.5	1.0			2	0.1953	0.1498	0.1953	0.1498	0.1498	0.1498	0.0680
5	5	5				1.0	1.2	1.0			2	0.1094	0.0811	0.1094	0.0811	0.0811	0.0811	0.0457
5	5	5				0.1	0.4	0.2			2	0.5126	0.4392	0.5126	0.4392	0.4392	0.5248	0.2685
5	5	5				0.1	0.3	0.2			2	0.3484	0.2920	0.3484	0.2920	0.2920	0.3975	0.1902
10	10	10				0.1	1.0	0.2			2	0.9910	0.9817	0.9910	0.9817	0.9817	0.9859	0.9430
10	10	10				0.1	0.4	0.2			2	0.7295	0.6818	0.7295	0.6818	0.6818	0.7874	0.5671
10	10	10				1.0	2.0	1.0			2	0.5161	0.4471	0.5161	0.4471	0.4471	0.4471	0.2347
10	10	10				1.0	3.0	1.0			2	0.8148	0.7389	0.8148	0.7389	0.7389	0.7389	0.4958
15	15	15				1.0	2.0	1.8			2	0.4839	0.5354	0.4839	0.5354	0.5354	0.8698	0.6694
15	15	15				1.0	2.0	0.0			2	0.9916	0.9972	0.9916	0.9972	0.9972	0.9999	0.9974
15	15	15				1.0	1.3	1.0			2	0.2555	0.2360	0.2555	0.2360	0.2360	0.2360	0.1127
15	15	15				1.0	3.0	1.0			2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996
5	5	5	5			0.1	0.3	0.2	0.1		2	0.5008	0.4825	0.4753	0.4632	0.4071	0.5304	0.2074
5	5	5	5			1.0	3.0	3.2	1.0		3	0.4337	0.4875	0.4001	0.4644	0.4124	0.6533	0.2866
5	5	5	5			1.0	1.5	3.2	1.0		3	0.5953	0.5159	0.5754	0.4952	0.4353	0.5319	0.2104
5	5	5	5			1.0	1.1	2.0	1.0		3	0.5770	0.4992	0.5666	0.4768	0.4252	0.5456	0.2021
5	5	5	5			1.0	1.0	2.0	1.0		3	0.5778	0.4813	0.5786	0.4496	0.4044	0.5556	0.2166
5	5	5	5			0.1	2.0	0.2	0.1		2	0.9901	0.9735	0.9875	0.9689	0.9327	0.9725	0.8133
5	5	5	5			0.1	1.0	0.2	0.1		2	0.9512	0.9089	0.9403	0.9044	0.8374	0.9033	0.6380
5	5	5	5			1.0	4.0	3.0	1.0		2	0.5850	0.6142	0.5508	0.5955	0.5292	0.7139	0.3562
5	5	5	5			1.0	3.0	4.0	1.0		3	0.5891	0.6221	0.5499	0.6060	0.5279	0.7279	0.3558
5	5	5	5			1.0	3.0	5.0	1.0		3	0.7161	0.7315	0.6805	0.7168	0.6361	0.7863	0.4334
5	5	5	5			1.0	1.0	2.0	1.0		3	0.5804	0.4920	0.5894	0.4656	0.4197	0.5630	0.2120
10	10	10	10			1.0	1.0	2.0	1.0		3	0.8292	0.7747	0.8321	0.7703	0.6679	0.8332	0.5264
10	10	10	10			1.0	2.0	1.0	0.0		2	0.9700	0.9901	0.9128	0.9741	0.9948	0.9948	0.9393
10	10	10	10			1.0	3.0	1.0	0.1		2	0.9850	0.9990	0.9276	0.9922	0.9999	0.9992	0.9994
15	15	15	15			1.0	2.0	3.0	1.0		3	1.0000	1.0000	0.9999	1.0000	0.9997	1.0000	0.9995
15	15	15	15			1.0	1.5	3.2	1.0		3	1.0000	1.0000	1.0000	1.0000	0.9999	1.0000	1.0000
15	15	15	15			1.0	2.0	3.0	1.0		3	1.0000	1.0000	0.9999	1.0000	0.9997	1.0000	0.9995

Table 19: Estimated Power Values for $k = 4, 5, 6$ under $E(\theta_i)$

n_1	n_2	n_3	n_4	n_5	n_6	μ_1	μ_2	μ_3	μ_4	μ_5		Peak	LAR	MW	SAL	Bhat	HNEQ	HNOPT	KW
5	5	5	5	5		1.0	2.0	3.0	2.0	1.0		3	0.4467	0.4977	0.3999	0.4863	0.5389	0.5389	0.1825
5	5	5	5	5		0.0	0.3	1.0	0.3	0.0		3	0.5391	0.5320	0.5100	0.5102	0.4957	0.5300	0.1668
5	5	5	5	5		1.0	2.0	2.5	2.0	1.0		3	0.3336	0.4033	0.2909	0.3867	0.4467	0.4767	0.1546
5	5	5	5	5		0.0	0.3	0.4	0.1	0.0		3	0.1705	0.1729	0.1549	0.1734	0.1741	0.1919	0.0556
5	5	5	5	5		0.0	0.3	0.5	0.3	0.0		3	0.2090	0.2202	0.1860	0.2209	0.2298	0.2340	0.0657
5	5	5	5	5		1.0	1.0	1.5	2.0	1.0		4	0.5895	0.5432	0.5424	0.5111	0.4113	0.5931	0.1962
5	5	5	5	5		1.0	1.1	1.5	2.0	1.0		4	0.5636	0.5239	0.5173	0.5072	0.3922	0.5627	0.1814
5	5	5	5	5		1.0	1.0	1.4	1.1	1.0		3	0.1989	0.1703	0.1978	0.1682	0.1526	0.1831	0.0499
5	5	5	5	5		1.0	1.5	2.0	1.5	1.0		3	0.2744	0.2845	0.2574	0.2763	0.2921	0.2921	0.0848
5	5	5	5	5		1.0	1.3	1.5	1.3	1.0		3	0.1562	0.1551	0.1466	0.1527	0.1598	0.1649	0.0509
10	10	10	10	10		0.1	0.2	0.5	0.2	0.1		3	0.9413	0.9550	0.9188	0.9541	0.9602	0.9430	0.7457
10	10	10	10	10		0.1	0.2	3.0	0.2	0.1		3	1.0000	1.0000	1.0000	0.9999	1.0000	1.0000	0.9975
5	5	5	5	5	5	0.1	0.2	0.2	0.4	1.0	0.0	5	0.5107	0.4263	0.5230	0.4468	0.2612	0.5258	0.1391
5	5	5	5	5	5	0.1	0.2	1.5	1.0	1.0	0.0	3	0.7249	0.7690	0.6580	0.7315	0.6281	0.8976	0.4600
5	5	5	5	5	5	0.0	0.2	2.0	0.4	0.2	0.0	3	0.9785	0.9462	0.9766	0.9099	0.8258	0.9741	0.5915
10	15	8				1.0	2.0	1.0				2	0.8602	0.8588	0.8602	0.8588	0.8580	0.8580	0.6487
10	15	8				1.5	2.0	1.0				2	0.6327	0.6173	0.6327	0.6173	0.6161	0.7176	0.4564
17	18	21				0.1	0.2	0.1				2	0.1053	0.0968	0.1053	0.0968	0.0958	0.0958	0.0533
14	17	11				0.5	0.8	0.6				2	0.1990	0.1909	0.1990	0.1909	0.1890	0.2020	0.0926
4	6	8	7			1.0	2.0	3.0	2.0			3	0.3814	0.3523	0.3213	0.2798	0.4198	0.4198	0.1454
13	12	9	20			1.0	2.0	2.5	2.0			3	0.3231	0.4250	0.2695	0.3572	0.5982	0.6459	0.3418
17	18	21	13			0.1	0.2	0.1	0.0			2	0.1293	0.1217	0.1152	0.1164	0.1273	0.1273	0.0610
17	18	21	13			0.0	0.5	0.3	0.0			2	0.4268	0.4395	0.3979	0.4417	0.3791	0.5009	0.2425
5	10	5	10	5		0.0	1.0	2.0	1.0	0.0		3	0.8659	0.9960	0.6885	0.9949	1.0000	1.0000	1.0000
20	23	21	24	25		0.1	0.1	0.2	1.0	0.1		4	0.9769	0.9571	0.9753	0.9645	0.8468	0.9816	0.8437
12	14	7	8	4		0.5	1.5	0.1	0.1	0.0		2	0.9716	0.9692	0.9613	0.9348	0.9478	0.9868	0.8547
5	20	5	20	5		0.0	0.5	1.0	0.5	0.0		3	0.4122	0.5381	0.3573	0.5393	0.5677	0.5677	0.2216
8	8	12	6	5		0.0	1.0	2.0	1.0	0.0		3	0.9941	0.9989	0.9855	0.9911	0.9995	0.9995	0.9650
9	11	14	16	3		0.0	0.1	0.3	0.2	0.0		3	0.1714	0.1609	0.1561	0.1465	0.1792	0.1894	0.0664
16	18	13	19	20	22	0.0	0.1	0.3	0.2	0.1	0.0	3	0.2276	0.2520	0.1952	0.2430	0.2260	0.2583	0.0886

CHAPTER 5. CONCLUSIONS

This dissertation has introduced a new nonparametric test for testing umbrella alternatives in a completely randomized design. This test is applied for cases where the peak of the umbrella pattern is known prior to testing. The test statistic is based on information from a trio of groups taken C_3^k at a time.

Unlike other tests, this test compares groups across the peak of the umbrella pattern. The importance of testing across the peak was first emphasized by Hettmansperger (1987). The results of this study has added further evidence of the importance of testing across the peak. Consider for example the data below.

Table 20: Simulated Data

x1	x2	x3	x4	x5
-0.1250049	2.554467	2.5158302	5.610838	0.6614217
1.1177840	4.050557	1.6159421	7.513144	0.6812125
1.7691099	1.544746	1.7757046	4.980371	-0.2921197
1.3769862	5.199789	0.9867563	5.354071	0.7294717

The mean plot for this data set is shown in the figure below.

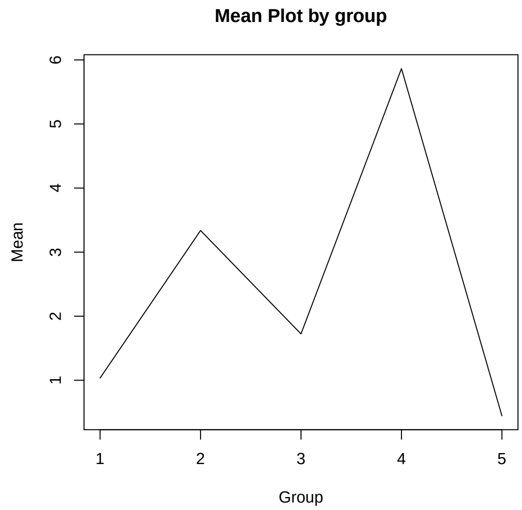

Figure 6: Mean Plot of Salmonella Bacteria Strain by Group

Table 21: Test Results p=2, k=5

Test	T	Mean	Variance	Test Statistic	pvalue
LAR	171	106.67	2779.02	1.22	0.1112
MW	77	56	158.67	1.67	0.0477
Sal	115	64	1190.4	1.48	0.0697
Bhat	44	32	37.33	1.96	0.0248
HN				1.48	0.0691

Table 22: Test Results p=4, k=5

Test	T	Mean	Variance	Test Statistic	pvalue
LAR	313	106.67	2779.02	3.91	0.0000
MW	93	56	158.67	2.94	0.0017
Sal	188	64	1190.4	3.59	0.0002
Bhat	50	32	37.33	2.95	0.0016
HN				1.70	0.0441

The results in Table 4.2 is a typical example of a situation where the Mack-Wolfe test and the Bhat's test can be very misleading. You would expect all the tests in Table 4.2 to reject H_o. However, only the LAR and Hettmansperger and Nortons test (HN) are reliable in this case. Based on the evidence just presented, the Mack-Wolfe test and Bhat's tests should not be recommended in situations where there is more than one peak in the umbrella pattern. This is because, the Mack-Wolfe test and Bhat's test would reject the null even at a peak that is not the highest in a case where there are multiple peaks.

The expectation and variance of the test statistics have been derived and expressed in the simplest possible form. The expectation and variance of the special case where all the group sample sizes are equal is also given. The expectation and variances are simplified to a form that makes it easy for the user to use manually for both balanced and unbalanced designs. This is one of the few tests involving ordered alternatives whose expectations and variance is simplified in such user friendly forms. Three numerical examples are solved manually and the results compared to other tests.

The asymptotic distribution of the of the test statistics is derived with the use of classical multi-sample U-statistics theorems and the detailed proof shown. The results of the derivation of the asymptotic distribution shows that the test statistics converges in distribution to a standard normal distribution.

Finally, a finite simulation study which compares LAR to other existing tests is presented in section 4. The LAR test was competing with all the other tests in terms of estimated powers. Based on the results of the power analysis, the LAR generally performed better than the other tests when the differences in location parameter between the adjacent groups was small. In such cases, the peak if very low making the entire pattern shallow. In such shallow umbrella patterns, the LAR is far more robust than all the other tests including the HNOPT. In situations where the slope (of either side of the peak) increased slowly and then had a sudden sharp jump, the LAR test outperformed all the other tests being compared to it. No tests with the exception of the HNOPT clearly stood out compared to the other tests when the differences in location parameter is considerably wide. The other tests with the exception of the KW test did better than the LAR in many cases where the peak is 1 or k.

These are the major points to take away from the study;

- LAR is best at detecting very minute changes in patterns between groups.
- LAR is best when the difference in means or other location or scale parameters are small (i.e. shallow umbrella).
- LAR performs better when there is a steady increase in the group location parameters and a sudden rise to the peak.
- The study found situations where the results of the other tests could be misleading. For instance, LAR is more efficient when the alternative contains multiple peaks or some groups being equal in location or scale parameters . Note that

the presence of outliers in the data set could lead to multiple peaks. In this line, the LAR is thus more robust to outliers than all the other tests.

After completing this study, the following recommendations came to mind.

- It would relevant to investigate the effectiveness of the proposed test (LAR) under equal variances and unequal variances.

- Since clinical research studies are very involved with repeated designs, designing a similar test for repeated designs would be of great benefit to the clinical research community.

- Many other patterns are of investigational interest to the research community, the exponential trend being one of them. The LAR test could be adjusted to serve such purposes.

REFERENCES

[1] A.R. Jonckheere,*A distribution-free k-sample test against ordered alternatives*, Biometrika **41** (1954), pp. 133-145.

[2] B.A. Millen and D.A. Wolfe, *A class of nonparametric tests for umbrella alternatives*, Journal of Statistical Research. **39** (2005),pp. 7-24

[3] A.S Salman, *A new nonparametric test for the umbrella alternative*, A Masters thesis submitted to the Department of Statistics, North Dakota University, (2010).

[4] C.J. Lloyd, *Estimating test power adjusted for size*, Journal of Statistical Computation and Simulation, **75:11** (2005),pp. 921-933

[5] D.A. Wolfe , *Nonparametric distribution-free procedures for order restricted altenatives*, In: Ahsanullah, M., Raquab, M.Z. (eds.) recent Developments in Order Random Variables. Nova Science, New York (2006).

[6] D. Basso and L. Salmaso,*A permutation test for umbrella alternatives*, Stat Comput, **21** (2011), pp. 45-54.

[7] G. A. Mack and D. A. Wolfe, *K-Sample Rank Tests for Umbrella Alternatives*, Journal of the American Statistical Association, **76** (1981), pp. 175-181.

[8] J.F. Pedersen, L.G. Madsen, A.L. Vibeke, O. Hamberg, T. Horn, B. Federspiel, and P. Bytzer, *A Doppler waveform index to characterize hepatic vein velocity pattern and evaluate hepatic fibrosis*, J. Clin. Ultrasound **36** (4) (2008), pp. 208-211.

[9] J. T. Terpstra, C. H. Chang and R. C. Magel, *On the use of Spearman's correlation coefficient for testing ordered alternatives*, Journal of Statistical Computation and Simulation **iFirst**, (2011), pp. 1-12.

[10] J. T. Terpstra and R. C. Magel,*A new nonparametric test for the ordered alternative problem*, Journal of Nonparametric Statistics **15** (3) (2003) , pp. 289-301.

[11] M. Hollander, and D. A. Wolfe, *Nonparametric Statistical Methods (second ed.). New York: John Wiley Sons Inc. A Wiley-Interscience Publication*, (1999).

[12] S.V. Bhat, *Simple K-Sample Rank Tests for Umbrella Alternatives* , Research Journal of Mathematics and Statistics **1**(1)(2009), pp. 27-29.

[13] T.E. Bradstreet, *"Favorite Data Sets from Early Phases of Drug Research - Part 2." Proceedings of the Section on Statistical Education of the American Statistical Association* , (1992).

[14] T.P. Hettmansperger and R.M. Norton, *Tests for patterned alternatives in k-sample problems*, J.Amer. Statist. Assoc.**82** (1987), pp. 292-299.

[15] V.S. Koroliuk and I.V. Borovskikh, *Theory of U-statistics*, Kluwer Academic Publishers Group, Dordrecht, (1994).