NONPARAMETRIC ADAPTIVE DISTRIBUTION-FREE PROCEDURE FOR CROSSOVER DESIGN WITH REPEATED MEASURES

by

AFSHAN BOODHWANI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2017

Acknowledgements

A very special gratitude goes out to my advisor Dr. Shan Sun Mitchell for countless hours of discussion and feedback. Through the course of the last three years she has not only been a thesis supervisor, but a life mentor and a friend, I will always treasure that. I also wish to express a sincere appreciation to the committee members for their insightful questions and feedback which incented me to widen my research from various perspectives. I would like to thank my friends for always challenging me and for accepting nothing less than excellence from me.

To my life-coaches, my parents Nazneen and Aslam Sadruddin: because I owe it all to you. Many Thanks!

My eternal cheerleader, my husband Aleem Boodhwani, thank you for providing me with emotional and moral support through the years and for always being encouraging and enthusiastic. I am also grateful to my mother and father-in-law, ZarinTaj and Mansoorali Boodhwani: always keen to know what I was doing and how I was proceeding, although it is likely that they never grasped what it was all about! I will cherish your celebrations of joy whenever a significant momentous was reached and also just your general affection. I am also grateful to my other family members and friends who have supported me along the way.

And finally, last but by no means least, I need to thank my little boy: expected to be born in July 2017. Considering he isn't even born yet it is hard to imagine how he can contribute to my thesis, but in fact he helps put it all into perspective for me: nations can be renewed and generational change is possible. He reminds me daily of all that is yet to be.

Thanks for all your encouragement!

April 19, 2017

Abstract

NONPARAMETRIC ADAPTIVE DISTRIBUTION-FREE PROCEDURE FOR CROSSOVER DESIGN WITH REPEATED MEASURES

Afshan Boodhwani, PhD

The University of Texas at Arlington, 2017

Supervising Professor: Shan Sun Mitchell

We propose to apply adaptive nonparametric procedures (Hill, Padmanabhan, & Puri, 1988) on 2x2 crossover design with repeated measures. We will derive the test-statistics (based on function of ranks) and find their asymptotic distributions. These test-statistics will be used to test (a) equality of carryover effects; (b) equality of direct treatment effects; (c) equality of carryover effects over time (repeated measures); and (d) equality of direct treatment effects over time (repeated measures), as suggested by Johnson and Grender (Johnson & Grender, 1993). We will be testing these hypotheses using modified versions of the test statistics derived by Johnson and Grender (Johnson & Grender, 1993) and Brunner et al. (Brunner, Domhof, & Langer, 2002) tailored to the underlying distribution of the data. In addition, we provide examples to illustrate the new methods.

The methods proposed extend the methods developed by Sun (Sun, 1997) for c-sample problems.

TABLE OF CONTENTS

List of Illustrations

List of Tables

Chapter 1

Introduction

Sample size is often a cause of concern during the initial phase of clinical trials. A method commonly used to handle this issue is the crossover design. A crossover design is a repeated measure type experiment where each subject is used to compare the types of treatments.

Extensive research has been done on crossover designs using the traditional normal distribution models by several authors such as Grizzle, Wallenstein and Fisher, Hills and Armitage, and Brown (Johnson & Grender, 1993). Moreover, works from Koch (Koch, 1972), Taulbee (Taulbee, 1982), Johnson and Grender (Johnson & Grender, 1993) and Brunner et al. (Brunner, Domhof, & Langer, 2002) discuss nonparametric methods for 2x2 crossover designs. The nonparametric methods discussed use the ranks of the observation for analysis for the data. In this paper we propose to instead use functions of ranks, called scores. The impetus for working with function (general scores) of ranks was provided by the work of Hogg et al. (Hogg, Fisher, & Randles, 1975), which specifies scores that are tailored to the given shape of the data. Extensive Monte Carlo studies (Hill, Padmanabhan, & Puri, 1988) have established the supremacy of these procedures over (i) the usual (nonadaptive) procedure of always working only with the ranks and the resulting statistics, such as Wilcoxon, and (ii) the usual parametric procedure (based on the assumption of normality of the underlying distribution function). Hogg et al. developed their procedures only in the context of hypothesis testing in the two-sample and one-sample problems. But the results of Puri (Puri M. L., 1965) and Puri and Sen (Puri & Sen, 1971) ensure that their nice properties also extend to more general situations. Sun (Sun, 1997) revisited this problem and extended it to resolve the practical problem of handling ties. Sun showed, by evaluating relative efficiency of the test statistics based on adaptive

nonparametric procedures to that of the standard nonparametric methods, the improvement is substantial.

Since the cross-over design with the repeated measurements longitudinal data is much more complicated than the c-sample problems in Sun (Sun, 1997), for example the nature of the dependency among periods and blocks as well as the interaction issues, hence more sophisticated technicalities are required. That is precisely what this paper plans to achieve, as well as establish the supremacy of using adaptive nonparametric procedure in 2x2 crossover designs with repeated measure, as compared to both the traditional parametric methods, and rank-based nonparametric methods.

Chapter 2

The 2x2 Crossover Design with Repeated Measures

Consider its simplest form, a two-treatment, two-period (2x2) crossover design. The subjects are randomly assigned to group 1, which receives treatment A followed by treatment B, or group 2, which receives the treatments in reverses order. A major advantage of this is that each subject serves as their own control, providing a better efficiency with a smaller sample size. Of course this is with the assumption that one treatment does not alter the subject's ability to conduct the second treatment. For example, crossover design may not be appropriate if testing for a drug that cures a disease. In addition to that we need to allow for a wash out period between the two treatments, so the lingering effects of one does not impact the other treatment's effect. This is lingering effect is known as the carry-over effect. A 2x2 crossover design is with p repeated measures is illustrated in Figure 2-1.

Group (i)	Sub	Period I			Period II		
		1	…	p	1	…	p
Sequence 1 (AB)	1	Y_{1111}	…	Y_{111p}	Y_{1211}	…	Y_{121p}
	2	Y_{1121}	…	Y_{112p}	Y_{1221}	…	Y_{122p}
	…	Y_{11k1}	…	Y_{11kp}	Y_{12k1}	…	Y_{12kp}
	n_1	Y_{11n_11}	…	Y_{11n_1p}	Y_{12n_11}	…	Y_{12n_1p}
Sequence 2 (BA)	1	Y_{2111}	…	Y_{211p}	Y_{2211}	…	Y_{221p}
	2	Y_{2121}	…	Y_{212p}	Y_{2221}	…	Y_{222p}
	…	Y_{21k1}	…	Y_{21kp}	Y_{22k1}	…	Y_{22kp}
	n_2	Y_{21n_21}	…	Y_{21n_2p}	Y_{22n_21}	…	Y_{22n_2p}

Figure 2-1: 2x2 Crossover Design with p Repeated Measures.

Section 2.1: The Model

Let y_{ijks} represent an observation from the i^{th} group, in the j^{th} period for the k^{th} subject's s^{th} repeated measure. So $i = 1,2$ groups, $j = 1,2$ periods, $k = 1, \dots, n_i$ ($n_1 + n_2 = n$) subject, $s = 1, \dots, p$ repeated measures, v is either treatment A or treatment B, and v' is the other treatment. Then the model for the 2x2 crossover design with p repeated measures can be expressed in terms of the given fixed effects:

$$E[Y_{ijks}] = \varphi_{ijs} = \mu_s + \pi_{js} + \tau_{vs} + h\lambda_{v's} \qquad (2\text{-}1)$$

where μ_s represents the general mean for the s^{th} repeated measure; π_{js} is the period effect of the j^{th} period in the s^{th} repeated measure; τ_{vs} is the direct treatment effect of the v^{th} treatment in the s^{th} repeated measure and $\lambda_{v's}$ is the carryover effect of v' treatment in the s^{th} repeated measure; and h is an indicator variable such that $h = 0$ if $j = 1$ and $h = 1$ if $j = 2$.

This model can also be expressed in a vector form:

$$E[\boldsymbol{Y}_{ijk}] = \boldsymbol{\varphi}_{ij} = (\varphi_{ij1} \quad \cdots \quad \varphi_{ijp})' = \boldsymbol{\mu} + \boldsymbol{\pi}_j + \boldsymbol{\tau}_v + h\boldsymbol{\lambda}_{v'} \qquad (2\text{-}2)$$

where $\boldsymbol{\mu} = (\mu_1 \quad \cdots \quad \mu_p)'$, $\boldsymbol{\pi}_j = (\pi_{j1} \quad \cdots \quad \pi_{jp})'$, $\boldsymbol{\tau}_v = (\tau_{v1} \quad \cdots \quad \tau_{vp})'$ and $\boldsymbol{\lambda}_{v'} = (\lambda_{v'1} \quad \cdots \quad \lambda_{v'p})'$. Figure 2-2 shows the main effects model for the 2x2 crossover design. Notice since the carry-over effect measures an effect the first treatment may have on the second treatment, this effect can only be observed in period II. This also implies that before we proceed with the analysis of direct treatment effects, a test must be conducted to establish the equality of carry-over effects for the two treatments. Thus, in the following sections we define a set of hypotheses, which must be tested in order. The results from one test may impact the validity of the next test and therefore caution must be taken when the results are interpreted.

Period / Group (Sequence)	Period I ($j = 1$)	Period II ($j = 2$)
	$s = 1, 2, \ldots p$	$s = 1, 2, \ldots p$
$i = 1$ **(AB)**	$\varphi_{11s} = \mu_s + \pi_{1s} + \tau_{As}$	$\varphi_{12s} = \mu_s + \pi_{2s} + \tau_{Bs} + \lambda_{As}$
$i = 2$ **(BA)**	$\varphi_{21s} = \mu_s + \pi_{1s} + \tau_{Bs}$	$\varphi_{22s} = \mu_s + \pi_{2s} + \tau_{As} + \lambda_{Bs}$

Figure 2-2 Main Effects Model for 2x2 Crossover Design For Fixed Repeated Measure.

Section 2.2: Hypotheses

To study a 2x2 crossover design with repeated measures there are six particular hypotheses that might be of interest. These hypotheses are identified below, however the order of testing the hypothesis is important since some tests are based on specific results from the previous tests.

Hypothesis 1: Testing the Equality of Carry-Over Effects

$$H_0^1: \lambda_{As} = \lambda_{Bs}\,, \qquad s = 1, 2, \ldots, p \tag{2-3}$$

In order to proceed to the next hypothesis, we must establish the equality of carryover effects thus, Grizzle (Grizzle, 1965) advocates using a higher significance level for this hypothesis, such as 0.10 or 0.15. In the event that the first test was rejected, some have suggested to rerun the experiment with a longer washout period, or others advocate using just the data from the first period to analyze the treatment effects. Koch et al. (Koch, Gitomer, Skalland, & Stokes, 1983) discuss the interpretation of direct treatment effects in the presence of significant carry-over effects.

Hypothesis 2: Testing the Equality of Direct Treatment Effect When Carry-Over Effects are Equal

$$H_0^2: \tau_{As} = \tau_{Bs}\,, \qquad s = 1,2,\ldots,p \qquad \textit{(2-4)}$$

Hypothesis 3: Testing the Equality of Carry-Over Effects over Time

$$H_0^3: \lambda_1 = \lambda_2 = \cdots = \lambda_p\,, \qquad \textit{where } \lambda_s = \lambda_{As} - \lambda_{Bs};\ \ s = 1,2,\ldots,p \qquad \textit{(2-5)}$$

Hypothesis 4: Testing the Equality of Direct Treatment Effect over Time when Carry-Over Effect over Time is Equal

If the equality of carry-over effect over time (H_0^3) has been established, then we can further test the data to determine the equality of direct treatment effect over time with the hypothesis:

$$H_0^4: \tau_1 = \tau_2 = \ \ldots = \tau_p\ , \qquad \textit{where } \tau_s = \tau_{As} - \tau_{Bs};\ \ s = 1,2,\ldots,p \qquad \textit{(2-6)}$$

Hypothesis 5: Testing the Average Response for Carry-Over Effects

Johnson and Grender (Johnson & Grender, 1993) indicate that the next two tests are more powerful than the previous ones since they are carried out on the average responses over the *p* repeated measures. However, these are only meaningful if we can ascertain that the effects being tested do not interact with time. Thus, if $H_0^3: \lambda_1 = \lambda_2 = \cdots = \lambda_p$ holds then we can proceed to test the equality of carry-over effects in average responses. The hypothesis:

$$H_0^5: \sum_{s=1}^{p} \lambda_{As} = \sum_{s=1}^{p} \lambda_{Bs} \qquad \textit{(2-7)}$$

Notice that this is a simple *c*-sample problem as discussed by Sun (Sun, 1997). Sun showed, that the results of Puri and Sen (Puri & Sen, 1971) can be extended, with some modification, to rounded-off data. The work of Sun (Sun, 1997) has already shown that in such cases adaptive procedures outrank the traditional non-adaptive nonparametric

methods, and parametric methods. However, for the sake of completeness we will replicate the results using the examples identified in this paper.

Hypothesis 6: Testing the Average Response for Direct Treatment Effects when Average Response for Carry-Over Effects are Equal

Much like the previous hypothesis, the following proves to be a two-sample problem. However, once again we must note that this test is only valid if:

$H_0^4: \tau_1 = \tau_2 = \ ... = \ \tau_p$ is not rejected, indicating there is no time by direct treatment effect, and

$H_0^5: \sum_{s=1}^{p} \lambda_{As} = \sum_{s=1}^{p} \lambda_{Bs}$ is not rejected supporting that the carry-over effect for the average responses is equal.

Then the null hypothesis becomes:

$$H_0^6: \sum_{s=1}^{p} \tau_{As} = \sum_{s=1}^{p} \tau_{Bs}$$

(2-8)

In Chapter 4 and Chapter 5 we discuss two different approaches to testing these hypotheses. Method 1 is proposed by Johnson and Grender (Johnson & Grender, 1993), and Method 2 is based on the work by Brunner et al. (Brunner, Domhof, & Langer, 2002). Both these methods rely on the rank of the data, however we have modified it to account for the adaptive procedures (Chapter 3) using score functions based on the underlying distribution. We will later illustrate the advantage of this method using several examples in Chapter 6.

Chapter 3

Nonparametric Adaptive Procedures

In this chapter we present the adaptive procedure for two-sample problem as defined by Hogg et al. (Hogg, Fisher, & Randles, 1975). Let X be a continuous random variable with cumulative distribution function (cdf) $F(x)$ then we assume that the marginal cdf of X_{ij} for $i = 1,2$ and fixed j would only vary from location shifts θ_i i.e. let $X_1 = (X_{11}, \dots, X_{1n_1})$ and $X_2 = (X_{21}, \dots, X_{2n_2})$ be two random samples with continuous-type distribution with unknown cdf $F_1(x) = F(x - \theta_1)$ and $F_2(x) = F(x - \theta_2)$ respectively. In order to test the hypothesis:

$$H_0: \theta_1 - \theta_2 = 0 \qquad \text{Vs.} \qquad H_a: \theta_1 - \theta_2 > 0,$$

Hogg et al. (Hogg, Fisher, & Randles, 1975) proposed using the test statistic is based on ranks of the observations in the form of $\sum_{i=1}^{n_1} a(R_{1i})$, where R_{1i} denoted the rank of X_{1i} among all $n = n_1 + n_2$ observations, and $a(1), a(2), \dots, a(n)$ denote scores which satisfy $a(1) \leq a(2) \leq \dots \leq a(n)$ with $a(1) \neq a(n)$. In the case where ties occurred, Sun (Sun, 1997) used the average rank scores. Since $F(x)$ is unknown, the traditional nonparametric (non-adaptive) method used the Mann-Whitney-Wilcoxon test, where $a(i) = i$ for $i = 1,2, \dots, n$. This has been established as the locally most powerful test in detecting shifts in a logistic distribution, and has good power properties for most underlying distributions specifically for moderate- to heavy-tailed error distributions, which are fairly symmetric in nature. However, this is not always the case. In many cases the distribution is either light-tailed and symmetric or skewed. Therefore, if we could detect the shape of the underlying distribution, we could improve the power of the test by using appropriate scores.

Section 3.1: Shape of the Underlying Distribution

Since the distribution of $F(x)$ is unknown, we need to rely on the given data to determine the shape of the underlying distribution. This shape can be described by its procession of symmetry, its tendency to skew, and Kurtosis, the weights of the tails of the distribution. These characteristics are described in the following sections.

3.1.1 Skewness

In statistics, skewness is defined as the measure of asymmetry of a distribution. The distribution can be either be symmetric (skewness =0), positively skewed, or negatively skewed. If the left tail (tail at small end of the distribution) is more pronounced than the right tail (tail at the large end of the distribution), the function is said to have negative or left skewness. If the reverse is true, it has positive or right skewness. If the two are equal, it has zero skewness. This is illustrated in Figure 3-1.

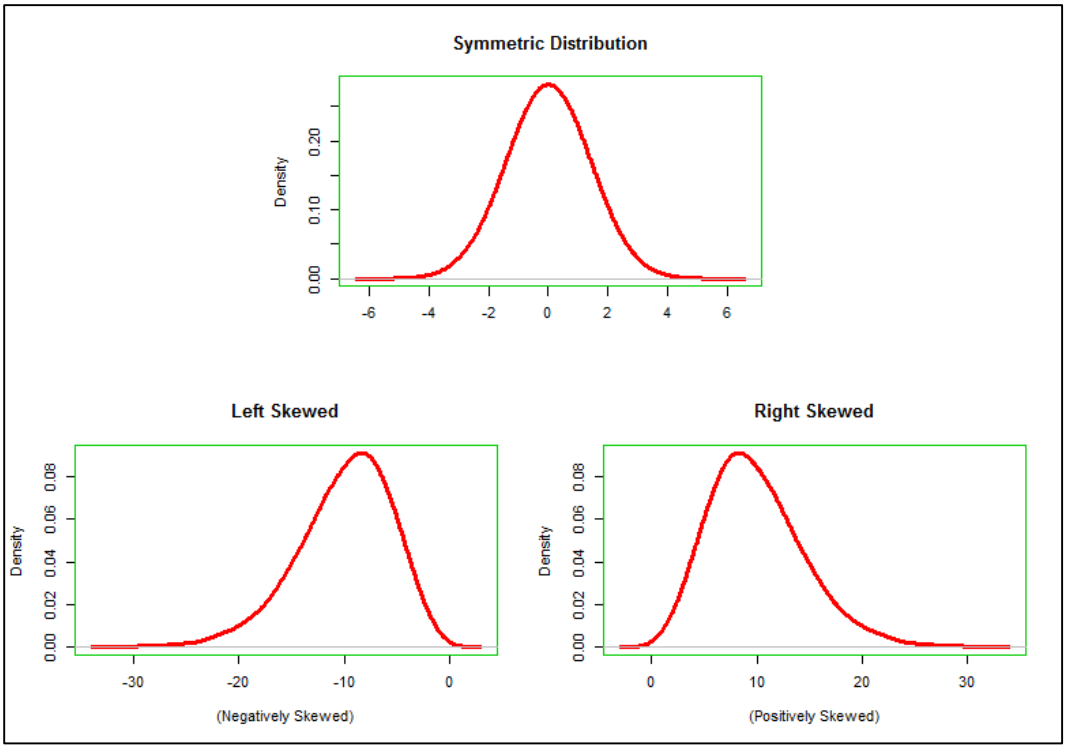

Figure 3-1: Different Types of Skewness

3.1.2 **Kurtosis**

Another measurement that helps define the shape of the underlying distribution is the measure of Kurtosis. It measures the heaviness of the tails as compared to the tails of a normal distribution, so any distribution with similar Kurtosis is said to have a moderate-tailed or mesokurtic distribution. A distribution with longer or fatter tails than a normal distribution is said to be heavy-tailed or leptokurtic. On the other hand, if the tails are thinner or shorter the distribution is said to be light-tailed or platykurtic. The three types of Kurtosis are illustrated in Figure 3-2.

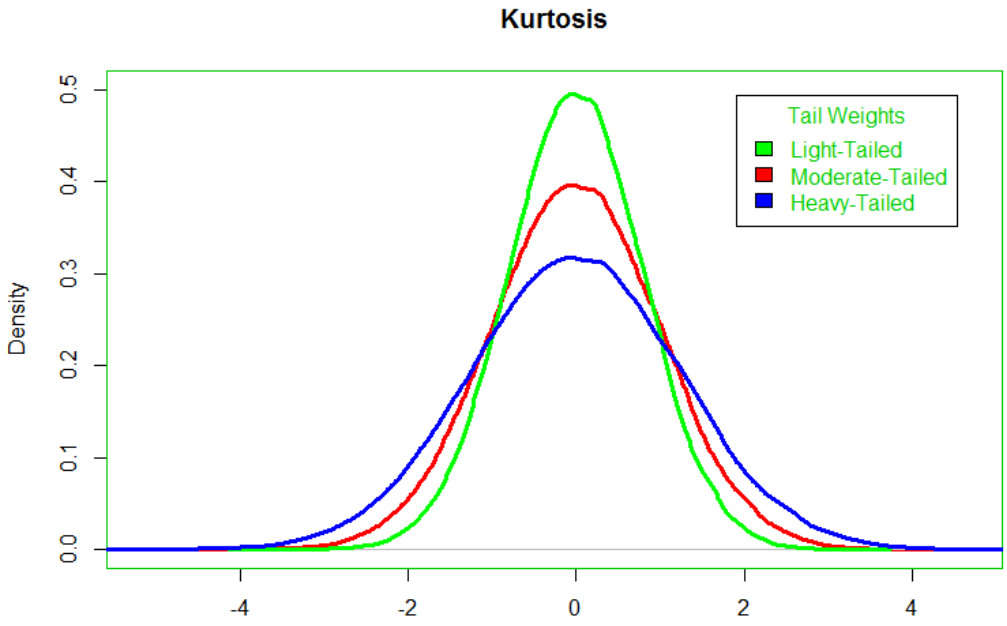

Figure 3-2: Illustration of Different Tail Weights

Section 3.2: Score Indicators

In order to help determine the shape of the underlying distribution, we modify the score indicators (Q_1, Q_2) described by Hogg et al. (Hogg, Fisher, & Randles, 1975) to obtain $(\bar{Q}_1, \bar{Q}_2)$ as defined by Hill et al. (Hill, Padmanabhan, & Puri, 1988).

The score indicators (Q_1, Q_2) as defined by Hogg et al. (Hogg, Fisher, & Randles, 1975), can be used identify the amount of skewness and level of tail weights of the underlying distribution, respectively. Firstly, assume that the underlying distribution is symmetric, then an appropriate measure of tail weight is:

$$Q_2 = \frac{\bar{U}_{.05} - \bar{L}_{.05}}{\bar{U}_{.5} - \bar{L}_{.5}}$$

(3-1)

where $\bar{U}_{.05}$, $\bar{U}_{.5}$ are, respectively, the averages of the largest 5%, and 50% of the order statistics of the combined sample, and $\bar{L}_{.05}$, $\bar{L}_{.5}$ are, respectively, the averages of the smallest 5%, and 50% of the order statistics of the combined sample. However, Hill et al. (Hill, Padmanabhan, & Puri, 1988) indicate that while this measure is a good indicator of tail weight when $(\theta_1 - \theta_2)$ is close to zero, it may indicate the wrong test statistic if the shift is large. For testing purposes, this may not be a big issue, since most tests detect a large shift with a high probability. However, an inappropriate test statistic may lead to a large confidence interval, which is a serious problem in terms of estimation. Thus, we shall work with the weighted average of Q_2 values based on the individual samples (Hill, Padmanabhan, & Puri, 1988):

$$\bar{Q}_2 = \frac{(n_1 Q_{2,1} + n_2 Q_{2,2})}{(n_1 + n_2)}$$

(3-2)

where $Q_{2,i}$ is the value of Q_2 as defined by (3-2) restricted to the i^{th} sample (group). Note $\bar{Q}_2$ is unaffected by the actual value of $(\theta_1 - \theta_2)$ and therefore performs better in indicating the tail-weights.

Similarly, we can use the indicator function Q_1 to study the skewness of $F(x)$. The function as studied by Fisher, and reported by Hogg (Hogg, Fisher, & Randles, 1975) can be written as:

$$Q_1 = \frac{\bar{U}_{.05} - \bar{M}_{.5}}{\bar{M}_{.5} - \bar{L}_{.05}}$$

(3-3)

where $\bar{U}_{.05}$, $\bar{M}_{.5}$ and $\bar{L}_{.05}$ are, respectively, the averages of the top 5%, middle 50% and bottom 5%, of the order statistics of the combined sample. However, for reasons explained earlier, we once again consider the weighted average of Q_1 values based on the $Q_{1,i}$ values as defined by (3-3) for each i^{th} sample considered separately (Hill, Padmanabhan, & Puri, 1988):

$$\bar{Q}_1 = \frac{(n_1 Q_{1,1} + n_2 Q_{1,2})}{(n_1 + n_2)}.$$

(3-4)

Using the values obtained from these score indicators, we can identify the shape of $F(x)$ and select an appropriate score functions tailored for that particular distribution. The benchmark for selecting the appropriate score function is given in Table 3-1, as studied by Hogg (Hogg, Fisher, & Randles, 1975) and modified by Hill (Hill, Padmanabhan, & Puri, 1988).

Table 3-1: Benchmark for Selector Functions

Benchmark	Distribution Indicated	Score Selected
$\bar{Q}_2 > 3.8$	Heavy-tailed symmetric	φ_W
$½ \le \bar{Q}_1 \le 2$, and $2.24 \le \bar{Q}_2 \le 3.8$	Moderate-tailed symmetric	φ_W
$½ \le \bar{Q}_1 \le 2$, and $\bar{Q}_2 < 2.24$	Light-tailed symmetric	φ_{ML}
$\bar{Q}_1 < ½$, and $\bar{Q}_2 < 3.8$	Skewed Left	φ_{SL}
$\bar{Q}_1 > 2$, and $\bar{Q}_2 < 3.8$	Skewed Right	φ_{SR}

Furthermore, the corresponding score functions used based on the score indicators are as follows:

Light-tailed symmetric:
$$\varphi_{ML}(u) = \begin{cases} u - \frac{1}{4} & 0 < u \le {}^1\!/_4 \\ 0 & {}^1\!/_4 < u \le {}^3\!/_4 \\ u - \frac{3}{4} & {}^3\!/_4 < u < 1 \end{cases}$$

Moderate or heavy-tailed symmetric:
$$\varphi_W(u) = u \qquad 0 < u < 1$$

Skewed left:
$$\varphi_{SL}(u) = \begin{cases} u - \frac{1}{2} & 0 < u \le {}^1\!/_2 \\ 0 & {}^1\!/_2 < u < 1 \end{cases}$$

Skewed right:
$$\varphi_{SR}(u) = \begin{cases} 0 & 0 < u \le {}^1\!/_2 \\ u - \frac{1}{2} & {}^1\!/_2 < u < 1 \end{cases}$$

(3-5)

For each of the hypothesis indicated in Section 2.2: Hypotheses, we will alter (i) the test statistic used by Johnson and Grender (Johnson & Grender, 1993); and (ii) the test statistic used by Brunner et al. (Brunner, Domhof, & Langer, 2002) , to incorporate the

score functions, and illustrate the supremacy of using adaptive procedures in 2x2 crossover designs with repeated measures using multiple examples.

For each of the hypothesis indicated in Section 2.2, we will alter (i) the test statistic used by Johnson and Grender (Johnson & Grender, 1993); and (ii) the test statistic used by Brunner et al. (Brunner, Domhof, & Langer, 2002) , to incorporate the score functions, and illustrate the supremacy of using adaptive procedures in 2x2 crossover designs with repeated measures using multiple examples.

Chapter 4

Nonparametric Adaptive Procedures Tailored to Johnson and Grender (J&G) Method

For each of the six statistics, there are two different methods for calculating the nonparametric test statistic based on rank. The first method of calculating the test statistic for the given hypothesis was inspired by the methodology described by Johnson and Grender (Johnson & Grender, 1993). In the following sections, the technical formulation for each of the hypothesis is explored followed by the test statistic for the hypothesis.

Section 4.1 Statistic Measures for the Hypotheses

4.1.1 Testing the Equality of Carry-Over Effects

Consider the null-hypothesis of equality of carry-over effect:

$$\mathrm{H}_0^1\colon \lambda_{\mathrm{As}} = \lambda_{\mathrm{Bs}}\,, \qquad \mathrm{s} = 1, 2, \ldots, \mathrm{p} \tag{2-3}$$

Looking back at the main effect model in Figure 2-1, note that this is the same as testing for the equality of the sum of the effects over the two periods. In other words, we can rewrite the null hypothesis as:

$$H_0^1\colon \varphi_{11s} + \varphi_{12s} = \varphi_{21s} + \varphi_{22s}\,, \qquad s = 1, 2, \ldots, p$$

$$H_0^1\colon 2\mu_s + (\pi_{1s} + \pi_{2s}) + (\tau_{As} + \tau_{Bs}) + \lambda_{As} = 2\mu_s + (\pi_{1s} + \pi_{2s}) + (\tau_{As} + \tau_{Bs}) + \lambda_{Bs}\,, \quad s = 1, 2, \ldots, p$$

$$H_0^1\colon \lambda_{As} = \lambda_{Bs}, \qquad s = 1, 2, \ldots, p \tag{4-1}$$

Therefore, it is more beneficial to define a new measure that uses the sum of the observations over the two periods. More precisely measurements for testing the carry-over effects are given by:

$$X_{iks} = Y_{i1ks} + Y_{i2ks} \tag{4-2}$$

4.1.2 Testing the Equality of Direct Treatment Effect When Carry-Over Effects are Equal

After the equality of the carry over effects has been established we can proceed to test the hypothesis of equal direct treatment effect.

$$H_0^2: \tau_{As} = \tau_{Bs}, \qquad s = 1,2,\dots,p \tag{2-4}$$

Under the null hypothesis of equal treatment effects, the within-subject difference satisfy the same model in the two groups. Similar to the preceding hypothesis, the hypothesis of equal direct treatment effects can be written as the difference of effects over the two periods. So,

$$H_0^2: \varphi_{11s} - \varphi_{12s} = \varphi_{21s} - \varphi_{22s}, \qquad s = 1,2,\dots,p$$

$$H_0^2: (\pi_{1s} - \pi_{2s}) + (\tau_{As} - \tau_{Bs}) - \lambda_{As} = (\pi_{1s} - \pi_{2s}) + (\tau_{Bs} - \tau_{As}) - \lambda_{Bs}, \quad s = 1,2,\dots,p$$

If the carry-over effects are equal then $\lambda_{As} = \lambda_{Bs}$, which implies the null hypothesis can be simplified to

$$H_0^2: \tau_{As} = \tau_{Bs}, \qquad s = 1,2,\dots,p \tag{4-3}$$

Thus, we define the new generated measure for this test as follows:

$$X_{iks} = Y_{i1ks} - Y_{i2ks} \tag{4-4}$$

4.1.3 Testing the Equality of Carry-Over Effect over Time

Another test that could be important, particularly when the responses for one repeated measure seem predominately larger or smaller than the others, is the test of carry-over or treatment effect over time.

$$H_0^3: \lambda_1 = \lambda_2 = \cdots = \lambda_p, \qquad \text{where } \lambda_s = \lambda_{As} - \lambda_{Bs};\ \ s = 1,2,\dots,p \tag{2-5}$$

Then the new measurements can be generated by:

Let, $Y'_{i1km} = Y_{ijkm} - Y_{ijk(m+1)}$, $m = 1,2,\dots,p-1$, for fixed *i, j*

Then $X_{ikm} = Y'_{i1km} + Y'_{i2km}$

(4-5)

4.1.4 Testing the Equality of Direct Treatment Effect over Time when Carry-Over Effect over Time is Equal

If the equality of carry-over effect over time (H_0^3) has been established, and the equality of direct treatment effect (H_0^2) has been rejected then we can further test the data to determine the equality of direct treatment effect over time with the hypothesis:

$$H_0^4: \tau_1 = \tau_2 = \ ... = \ \tau_p \ , \qquad \text{where } \tau_s = \tau_{As} - \tau_{Bs};\ s = 1, 2, \ldots, p \qquad (2\text{-}6)$$

The new measurements for the test can be generated as in the previous test with a slight variation.

Let, $Y'_{i1km} = Y_{ijkm} - Y_{ijk(m+1)}$, $m = 1, 2, \ldots, p-1$, for fixed *i, j*

Then $X_{ikm} = Y'_{i1km} - Y'_{i2km}$

(4-6)

4.1.5 Testing the Average Response for Carry-Over Effects

Recall, the next two hypotheses are only meaningful if we can ascertain that the effects being tested do not interact with time. Thus, if $H_0^3: \lambda_1 = \lambda_2 = \cdots = \lambda_p$ holds then we can proceed to test the equality of carry-over effects in average responses. The hypothesis:

$$H_0^5: \sum_{s=1}^{p} \lambda_{As} = \sum_{s=1}^{p} \lambda_{Bs}$$

(2-7)

can be tested by using a measure that sums the responses over time and period. So,

$$X_{ik} = \sum_{s=1}^{p} Y_{i1ks} + \sum_{s=1}^{p} Y_{i2ks}.$$

(4-7)

Notice that once we take the average of the responses, the generated measure is no longer dependent on the index s. Thus, the model has been reduced to a simple two-sample

problem. The work of Sun (Sun, 1997) has already shown that in such cases adaptive procedures outrank the traditional non-adaptive nonparametric methods, and parametric methods. However, for the sake of completeness we will replicate the results using the examples identified in this paper.

4.1.6 Testing the Average Response for Direct Treatment Effects when Average Response for Carry-Over Effects are Equal

Much like the previous hypothesis, the following proves to be a two-sample problem. However, once again we must note that this test is only valid if:

- $H_0^4: \tau_1 = \tau_2 = \ ... = \ \tau_p$ is not rejected, indicating there is no time by direct treatment effect, and
- $H_0^5: \sum_{s=1}^{p} \lambda_{As} = \sum_{s=1}^{p} \lambda_{Bs}$ is not rejected supporting that the carry-over effect for the average responses is equal.

Then the null hypothesis becomes:

$$H_0^6: \sum_{s=1}^{p} \tau_{As} = \sum_{s=1}^{p} \tau_{Bs} \qquad \textit{(2-8)}$$

This can be tested by using a measure generated as follows:

$$X_{ik} = \sum_{s=1}^{p} Y_{i1ks} - \sum_{s=1}^{p} Y_{i2ks}. \qquad \textit{(4-8)}$$

Note once again that this results in a two-sample problem.

Section 4.2: Test Statistics

The method described by Johnson and Grender (Johnson & Grender, 1993) for calculating the test statistics for the hypothesis described above utilized the rank of the new generated measures (i.e. the sums or differences of measures across the two periods), the method described here uses the score functions instead to develop an adaptive nonparametric procedure for 2x2 crossover designs with repeated measures. Recall that the same statistic is utilized in the first four hypotheses with the exception that the generated measures, X_{iks}, differ in each case. Where $i = 1,2$ groups, $k = 1, \dots, n_i$ ($n_1 + n_2 = n$) subject, and $s = 1, \dots, p$ repeated measures. Whereas the last two hypotheses are c-sample problems that implement the works of Sun (Sun, 1997).

4.2.1 Test Statistic for Hypotheses 1-4

While the generated measures for each of the hypothesis differs, the formula for calculating the test statistic for hypotheses 1-4 remains the same. We will first define the test statistic, W, used to test these hypotheses as described by Johnson and Grender (Johnson & Grender, 1993), and later propose a modified test statistic, W^a, which tailors to the underlying distribution based on adaptive nonparametric procedures of Hoggs et al. (Hogg, Fisher, & Randles, 1975).

In order to determine the W −test statistic used by Johnson and Grender (Johnson & Grender, 1993), let R_{iks}denote the rank for each X_{iks} across all n observations. Then the sample mean of the ranks is given by:

$$\bar{R}_{i.s} = \sum_{k=1}^{n_i} \frac{R_{iks}}{n_i}$$

(4-9)

Thus, the test statistic for a multivariate rank sum test can be described as:

$$W = (n-1)\sum_{i=1}^{2} n_i \boldsymbol{U}_i' \boldsymbol{V}^{-1} \boldsymbol{U}_i$$

where, $\boldsymbol{U} = \begin{bmatrix} \bar{R}_{i.1} - m & \bar{R}_{i.2} - m & \cdots & \bar{R}_{i.p} - m \end{bmatrix}$

And covariance matrix

$$\boldsymbol{V} = \sum_{i=1}^{2}\sum_{k=1}^{n_i} \left(\boldsymbol{R}_{ik} - m\mathbf{1}_{\boldsymbol{p}}\right)\left(\boldsymbol{R}_{ik} - m\mathbf{1}_{\boldsymbol{p}}\right)'$$

(4-10)

where $m = \frac{1}{2}(n+1)$ and $\mathbf{1_p}$ is a p-dimensional vector of ones. However, in order to account for the shape of the underlying distribution, we utilize the score functions rather than ranks. Recall $a(R_{iks})$ is the score function of the rank of X_{iks} over the n observations. Moreover, let $\bar{a}(R_{i.s})$ be the sample mean of the new ranks (using score functions) among all n_i subjects in the i^{th}group for the s^{th}repeated measure. This can be written as:

$$\bar{a}(R_{i.s}) = \sum_{k=1}^{n_i} \frac{a(R_{iks})}{n_i}$$

Thus, the modified test statistic, W^a, is given by:

$$W^a = (n-1)\sum_{i=1}^{2} n_i \boldsymbol{U_i^{a\prime}} \boldsymbol{V^{a-1}} \boldsymbol{U_i^a}$$

where,

$$\boldsymbol{U^a} = \begin{bmatrix} \bar{a}(R_{i.1}) - m & \bar{a}(R_{i.2}) - m & \cdots & \bar{a}(R_{i.p}) - m \end{bmatrix}$$

$$\boldsymbol{V^a} = \sum_{i=1}^{2}\sum_{k=1}^{n_i} \left(\boldsymbol{a}(\boldsymbol{R}_{ik}) - m\mathbf{1}_{\boldsymbol{p}}\right)\left(\boldsymbol{a}(\boldsymbol{R}_{ik}) - m\mathbf{1}_{\boldsymbol{p}}\right)'$$

(4-11)

in which m= median of the function of ranks and $\mathbf{1_p}$ is a p-dimensional vector of ones.

Under the hypothesis of equal distribution of ranks in the two groups, W is distributed approximately as the chi-square distribution with 2 degrees of freedom if n is large.

If n is small, the statistical significance of W can be determined using the permutation distribution corresponding to $\frac{n!}{\prod_{i=1}^{2} n_i!}$

Since the only modification is using a linear or truncated function of the rank rather than the rank itself, the distribution of W^a is the same as the distribution of W under the null hypothesis.

4.2.2 **Test Statistic for Hypotheses 5-6**

For the remaining two hypotheses of a 2x2 crossover design with repeated measures, which tested average response over each effect, we refer to the simple c-sample problem as discussed by Sun (Sun, 1997). Sun showed, that the results of Puri and Sen (Puri & Sen, 1971) can be extended, with some modification, to rounded-off data. So let $\tilde{X}_{ik}$ be the measurements X_{ik} rounded-off to the nearest integer. Then using the test statistic provided in by Sun, developed by Puri and Sen we have:

$$S_c = \frac{(N-1)\sum_i n_i((\tilde{S}_i/n_i) - \bar{a}_N)^2}{\sum_i(\tilde{a}_N(i) - \bar{a}_N)^2}$$

(4-12)

where $\tilde{a}_N(i)$ denotes the scores obtained after applying the average scores method and $\tilde{R}_{ik}$denotes the rank of $\tilde{X}_{ik}$in the combined sample of size *N.* Then $\tilde{S}_i = \sum_{k=1}^{n_1} \tilde{a}_N(\tilde{R}_{ik})$, and $\bar{a}_N$ is the average of the modified scores. S_c has asymptotically a chi-square distribution with $(c-1)$ degrees of freedom.

Chapter 5

Nonparametric Adaptive Procedures Tailored to F1-LD-F1 Method

Another way of looking at the hypotheses described in Section 2.2 is a modification of the methodology described by Brunner et al. (Brunner, Domhof, & Langer, 2002). Note that after attaining the new generated measures using sums or differences of measure across the two periods for testing carry-over or direct treatment effects respectively, we can utilize the F1-LD-F1 method. An important difference in the F1-LD-F1 method is that R_{iks} is the rank of X_{iks}, among all $N = p \cdot \sum_{i=1}^{g} n_i$ observations, as opposed to R_{iks} being the rank for each X_{iks} across $n = \sum_{i=1}^{g} n_i$ observations in the J&G method.

Another important difference is that the F1-LD-F1 model contains three hypotheses tests within its design. The three tests are:

- Test for group effects
- Test for time effects
- Test for interaction between group and time effects

In this chapter, we first present the hypotheses and their corresponding test statistics in F1-LD-F1 model (Brunner, Domhof, & Langer, 2002), and then we show that the formulations of hypotheses in the 2x2 crossover design with repeated measures are equivalent to those described in the F1-LD-F1, using either the sum or differences of the measures across the two periods.

Section 5.1: F1-LD-F1 Model

First we explain the F1-LD-F1 model and the hypotheses associated with this model. The F1-LD-F1 model is a nonparametric marginal model that makes use solely of the independence structure of observations in order to determine from the design which

marginal distributions of the observed random vectors are identical. The hypotheses are then tested using values that can be obtained from the marginal distributions.

In this model, $i = 1, \dots, g$ groups each consisting of $k = 1, \dots, n_i$ subjects are observed on $s = 1, \dots, p$ occasions, where the measurements $X_{ik1}, \dots, X_{ikp}$ are examined. The vectors $\boldsymbol{X}_{ik} = (X_{ik1}, \dots, X_{ikp})'$ are assumed independent, however the components of each of these vectors can be dependent on one another. The observations of different subjects within the same group are considered as replications of the experiment, as a way of modeling the independence structure. Hence, it is reasonable to assume that the common distribution functions of the vectors $\boldsymbol{X}_{ik}$ are identical, i.e. they do not depend on the index k. The observations X_{iks} and the marginal distribution F_{is} of this experiment design are shown in Figure 5-1.

Groups	Subjects	Data		Marginal Distribution
		Vectors	Time	Time
			s=1 ... s=p	s=1 ... s=p
Group 1	$k=1$: : $k=n_1$	X_{11} : : X_{1n_1}	$X_{111} \dots X_{11p}$: : : : : : $X_{1n_11} \dots X_{1n_1p}$	$F_{11} \dots F_{1p}$
Group 2	$k=1$: : $k=n_2$	X_{21} : : X_{2n_2}	$X_{211} \dots X_{21p}$: : : : : : $X_{2n_21} \dots X_{2n_2p}$	$F_{21} \dots F_{2p}$

Figure 5-1: Observations and distributions for a two group F1-LD-F1 model

Next we will discuss the three hypotheses associated with the F1-LD-F1 design and the corresponding test statistic. Later in Section 5.2, we will connect these hypotheses

of F1-LD-F1 model to the hypothesis of 2x2 crossover design with repeated measures as described in Section 2.2.

5.1.1 Test for Group Effects

The group effect in an experiment design with repeated measures corresponds to the main effect or the average treatment effect for an experimental group over all time points. Recall the F1-LD-F1 test can be generalized for experiments with more than two groups of subjects, with measurements taken at multiple time points. So the k^{th} subject in the i^{th} group is observed p times and the results are arranged in the vector form $\boldsymbol{X_{ik}} = (X_{ik1}, \dots, X_{ikp})', k = 1, \dots, n_i, i = 1, \dots, g$, which are assumed to be independent. As defined earlier, the marginal distribution of the functions of X_{iks} are denoted by $F_{is}(x)$.

Thus in order to understand the technical formulation of the hypothesis of no group effects and its test statistic as presented by Brunner et al. (Brunner, Domhof, & Langer, 2002), consider the mean $\bar{F}_{i.} = \frac{1}{p}\,\mathbf{1}_p'\,\boldsymbol{F_i}$, where $\frac{1}{p}\,\mathbf{1}_p' = (\frac{1}{p}, \dots, \frac{1}{p})$ and $\boldsymbol{F_i} = (F_{i1}, \dots, F_{ip})'$. The equality of the means $\bar{F}_{1.} = \cdots = \bar{F}_{g.}$ can then be formulated using the centering matrix $\boldsymbol{P}_g = \boldsymbol{I}_g - \frac{1}{g}\boldsymbol{J}_g$, where $\boldsymbol{I}_g$ is an gxg identity matrix, and $\boldsymbol{J}_g$ is an gxg matrix of ones. So contrast matrix, C_A can be derived by taking the Kronecker-product $C_A = \boldsymbol{P}_g \otimes \frac{1}{p}\,\mathbf{1}_p'$. Thus the test for *no group effect* can be written as:

$$H_0^F(A): C_A\boldsymbol{F} = \left(\boldsymbol{P}_g \otimes \frac{1}{p}\,\mathbf{1}_p'\right)\boldsymbol{F} = \begin{bmatrix} \bar{F}_{1.} - \bar{F}_{..} \\ \vdots \\ \bar{F}_{g.} - \bar{F}_{..} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

(5-1)

where $\bar{F}_{..} = \frac{1}{gp}\sum_{i=1}^{g}\sum_{s=1}^{p} F_{is}$. Traditionally, F1-LD-F1 produces three test-statistics, a Wald Type Statistic (WTS), an ANOVA Type Statistic (ATS) and a modified ANOVA Type Statistic (mod ATS). However, for medium and small size statistics the ATS is preferable

to achieve better approximation. Thus, here the concentration is on ATS. The ATS test statistic defined by Brunner et al. (Brunner, Domhof, & Langer, 2002) is:

$$F_n(A) = \frac{g}{(g-1)\sum_{i=1}^{g} \frac{\hat{\sigma}_i^{\,2}}{n_i}} \sum_{i=1}^{g} (\bar{R}_{i..} - \bar{R}_{...})^2$$

(5-2)

where R_{iks} is the rank of X_{iks} over all $N = p \cdot \sum_{i=1}^{g} n_i$ observations. Thus, the means over all groups, respectively, are denoted by:

$$\bar{R}_{ik.} = \frac{1}{p} \sum_{s=1}^{p} R_{iks}$$

$$\bar{R}_{i..} = \frac{1}{n_i} \sum_{k=1}^{n_i} \bar{R}_{ik.}$$

$$\bar{R}_{...} = \frac{1}{g} \sum_{i=1}^{g} \bar{R}_{i..}$$

and

$$\hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\bar{R}_{ik.} - \bar{R}_{i..})^2$$

(5-3)

Moreover, under the null hypothesis $H_0^F(A)$, the distribution of $F_n(A)$ can be approximated by the central $F(\hat{f}_A, \hat{f}_0)$-distribution. Such that the degrees of freedom are given by:

$$\hat{f}_A = \frac{(\mathrm{g}-1)^2}{1 + g(g-2)\left[\frac{\sum_{i=1}^{g}\left(\frac{\hat{\sigma}_i^{\,2}}{n_i}\right)^2}{\left(\sum_{i=1}^{g}\frac{\hat{\sigma}_i^{\,2}}{n_i}\right)^2}\right]}$$

and

$$\hat{f}_0 = \frac{(\sum_{i=1}^{g} \frac{\hat{\sigma}_i^{\,2}}{n_i})^2}{\left(\sum_{i=1}^{g} \frac{1}{\mathrm{n_i} - 1}\left(\left(\frac{\hat{\sigma}_i^{\,2}}{n_i}\right)^2\right)\right)}$$

(5-4)

The test statistics obtained, after modifying to account for the adaptive procedure is:

$$F_n^a(A) = \frac{g}{(g-1)\sum_{i=1}^{g} \frac{\hat{\sigma}_i^{a\,2}}{n_i}} \sum_{i=1}^{g} (\bar{a}(R_{i\cdot\cdot}) - \bar{a}(R_{\cdot\cdot\cdot}))^2$$

(5-5)

where $a(R_{iks})$ is the score function of the rank of X_{iks} over all $N = p \cdot \sum_{i=1}^{g} n_i$ observations. Thus, the means over all groups, respectively, are denoted by:

$$\bar{a}(R_{ik\cdot}) = \frac{1}{p} \sum_{s=1}^{p} a(R_{iks})$$

$$\bar{a}(R_{i\cdot\cdot}) = \frac{1}{n_i} \sum_{k=1}^{n_i} \bar{a}(R_{ik\cdot})$$

$$\bar{a}(R_{\cdot\cdot\cdot}) = \frac{1}{g} \sum_{i=1}^{g} \bar{a}(R_{i\cdot\cdot})$$

and

$$\hat{\sigma}_i^{a\,2} = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (\bar{a}(R_{ik\cdot}) - \bar{a}(R_{i\cdot\cdot}))^2$$

(5-6)

Moreover, under the null hypothesis $H_0^F(A)$, since the only change in the test statistic results from applying a function of ranks rather than the rank itself, the distribution of $F_n^a(A)$ can be approximated by the central $F(\hat{f}_A^a, \hat{f}_0^a)$-distribution. The proof is similar to

one provided by Brunner et al. (Brunner, Domhof, & Langer, 2002). Such that the degrees of freedom are given by:

$$\hat{f}_A^a = \frac{(g-1)^2}{1 + g(g-2)\left[\frac{\sum_{i=1}^{g}\left(\frac{\hat{\sigma}_i^{a\,2}}{n_i}\right)^2}{(\sum_{i=1}^{g}\frac{\hat{\sigma}_i^{a\,2}}{n_i})^2}\right]}$$

and

$$\hat{f}_0^a = \frac{(\sum_{i=1}^{g}\frac{\hat{\sigma}_i^{a\,2}}{n_i})^2}{\left(\sum_{i=1}^{g}\frac{1}{\mathrm{n_i}-1}\left(\left(\frac{\hat{\sigma}_i^{a\,2}}{n_i}\right)^2\right)\right)}$$

(5-7)

5.1.2 **Test for Time effects**

Another question of interest in a F1-LD-F1 model is the investigation of separate time effects within each group. The hypothesis of no time effect can be written as:

$$H_0^F(T)\colon C_p\boldsymbol{F} = \left(\frac{1}{g}\,\mathbf{1}_g' \otimes \boldsymbol{P_p}\right)\boldsymbol{F} = \begin{bmatrix}\bar{F}_{\cdot 1} - \bar{F}_{\cdot\cdot}\\ \vdots \\ \bar{F}_{\cdot p} - \bar{F}_{\cdot\cdot}\end{bmatrix} = \begin{bmatrix}0\\ \vdots \\ 0\end{bmatrix} = \mathbf{0}$$

(5-8)

where $\boldsymbol{C_p} = \left(\frac{1}{g}\,\mathbf{1}_g' \otimes \boldsymbol{P}_p\right)$, $\boldsymbol{P}_p = \boldsymbol{I}_p - \frac{1}{p}\boldsymbol{J}_p$, where $\boldsymbol{I}_p$ is a $p$xp identity matrix, and $\boldsymbol{J}_p$ is an pxp matrix of ones is the centering matrix and $\bar{F}_{\cdot\cdot} = \frac{1}{gp}\sum_{i=1}^{g}\sum_{s=1}^{p}F_{is}$.

Next in order to calculate the test statistic for each group, we calculate the vectors of midranks, and their means within each group denoted by:

$$\boldsymbol{R_{ik}} = \left(R_{ik1}, \dots, R_{ikp}\right)', \qquad i = 1, \dots, g;\ k = 1, \dots, n_i,$$

$$\bar{\boldsymbol{R}}_{\boldsymbol{i}\cdot} = \frac{1}{n_i}\sum_{k=1}^{n_i}\boldsymbol{R_{ik}},$$

$$\bar{\boldsymbol{R}}_{\cdot\cdot} = \frac{1}{g}\sum_{i=1}^{g}\bar{\boldsymbol{R}}_{\boldsymbol{i}\cdot},$$

$$\bar{R}_{i.s} = \frac{1}{n_i}\sum_{k=1}^{n_i} R_{iks}$$

$$\tilde{R}_{..s} = \frac{1}{g}\sum_{i=1}^{g} \bar{R}_{i.s} \ ,$$

$$\tilde{R}_{...} = \frac{1}{p}\sum_{s=1}^{p} \bar{R}_{..s} \ .$$

(5-9)

Then based on the works of Brunner et al. (Brunner, Domhof, & Langer, 2002), the covariance matrix needed for the average time effect can be estimated by:

$$\widehat{\boldsymbol{S}}_p = \frac{1}{g^2}\sum_{i=1}^{g} \widehat{\boldsymbol{V}}_i \ ,$$

$$\widehat{\boldsymbol{V}}_i = \frac{n}{N^2 n_i (n_i - 1)}\sum_{k=1}^{n_i} (\boldsymbol{R_{ik}} - \overline{\boldsymbol{R}}_{\boldsymbol{i.}})(\boldsymbol{R_{ik}} - \overline{\boldsymbol{R}}_{\boldsymbol{i.}})'$$

(5-10)

where $n = \sum_{i=1}^{g} n_i$ is the total number of subjects, and $N = np$ is the total number of observations in the experiment. Then again, using the ATS test-statistic we have:

$$F_n(T) = \frac{n}{N^2 tr(\boldsymbol{C_p}\widehat{\boldsymbol{S}}_{\boldsymbol{p}})}\sum_{s=1}^{p} (\widetilde{\boldsymbol{R}}_{..s} - \widetilde{\boldsymbol{R}}_{...})^2 \ .$$

(5-11)

The distribution of this test statistic can be approximated by the central $F(\hat{f}_T, \infty)$-distribution. Such that:

$$\hat{f}_T = \frac{\left[tr(\boldsymbol{C_p}\widehat{\boldsymbol{S}}_{\boldsymbol{p}})\right]^2}{tr(\boldsymbol{C_p}\widehat{\boldsymbol{S}}_{\boldsymbol{p}}\boldsymbol{C_p}\widehat{\boldsymbol{S}}_{\boldsymbol{p}})}$$

(5-12)

We now look at the modified version of this test statistic by applying the score functions in order to account for the shape of the underlying distribution. Let $a(R_{iks})$ is the

score function of the rank of X_{iks} over all $N = p \cdot \sum_{i=1}^{g} n_i$ observations as described in Chapter 3 then:

$$\boldsymbol{a}(\boldsymbol{R_{ik}}) = \left(a(R_{ik1}), \dots, a\left(R_{ikp}\right)\right)' , \qquad i = 1, \dots, g;\ k = 1, \dots, n_i ,$$

$$\overline{\boldsymbol{a}}(\boldsymbol{R_{i.}}) = \frac{1}{n_i}\sum_{k=1}^{n_i} \boldsymbol{a}(\boldsymbol{R_{ik}}) ,$$

$$\bar{a}(R_{i.s}) = \frac{1}{n_i}\sum_{k=1}^{n_i} a(R_{iks}) ,$$

$$\tilde{a}(R_{..s}) = \frac{1}{g}\sum_{i=1}^{g} \bar{a}(R_{i.s}) ,$$

$$\tilde{a}(R_{...}) = \frac{1}{p}\sum_{s=1}^{p} \tilde{a}(R_{..s}) .$$

(5-13)

Moreover, the new covariance matrix can be estimated by:

$$\widehat{\boldsymbol{S}}_p^a = \frac{1}{g^2}\sum_{i=1}^{g} \widehat{\boldsymbol{V}}_i^a ,$$

$$\widehat{\boldsymbol{V}}_i^a = \frac{n}{N^2 n_i (n_i - 1)}\sum_{k=1}^{n_i} (\boldsymbol{a}(\boldsymbol{R_{ik}}) - \overline{\boldsymbol{a}}(\boldsymbol{R_{i.}}))(\boldsymbol{a}(\boldsymbol{R_{ik}}) - \overline{\boldsymbol{a}}(\boldsymbol{R_{i.}}))'.$$

(5-14)

where $n = \sum_{i=1}^{g} n_i$ is the total number of subjects, and $N = np$ is the total number of observations in the experiment. Then again, using the ATS test-statistic we have:

$$F_n^a(T) = \frac{n}{N^2 tr(\boldsymbol{C_p}\widehat{\boldsymbol{S}}_p^a)}\sum_{s=1}^{p} (\tilde{a}(R_{..s}) - \tilde{a}(R_{...}))^2.$$

(5-15)

The distribution of this test statistic can be approximated by the central $F(\hat{f}_T^a, \infty)$-distribution. Such that:

$$\hat{f}_T^a = \left[tr(\boldsymbol{C_p}\widehat{\boldsymbol{S}}_p^a)\right]^2 \Big/ tr(\boldsymbol{C_p}\widehat{\boldsymbol{S}}_p^a \boldsymbol{C_p}\widehat{\boldsymbol{S}}_p^a)$$

(5-16)

5.1.3 Test for Interaction between Group and Time Effects

Yet another question of investigation is whether the profiles of the time curves for the treatment groups are different. This question is particularly important as we use the F1-LD-F1 model to test the hypotheses of a 2x2 crossover design, since certain assumptions cannot be made if the interaction is significant.

The hypothesis of no interaction between group and time can be written as:

$$H_0^F(AT)\colon C_{AT}\boldsymbol{F} = \left(\boldsymbol{P_g} \otimes \boldsymbol{P_p}\right)\boldsymbol{F} = \begin{bmatrix} \bar{F}_{11} - \bar{F}_{1\cdot} - \bar{F}_{\cdot 1} + \bar{F}_{\cdot\cdot} \\ \vdots \\ \bar{F}_{gp} - \bar{F}_{g\cdot} - \bar{F}_{\cdot p} + \bar{F}_{\cdot\cdot} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

(5-17)

where $\boldsymbol{P}_b = \boldsymbol{I}_b - \frac{1}{b}\boldsymbol{J}_b$, where $\boldsymbol{I}_b$ is a bxb identity matrix, and $\boldsymbol{J}_b$ is a bxb matrix of ones for any arbitrary natural number b. Once again, in order to test the interaction hypothesis for small and medium sample sizes, the ATS is preferred. To obtain the quadratic form for this hypothesis, Brunner et al. (Brunner, Domhof, & Langer, 2002) utilize the matrix $\boldsymbol{T}_{AT} = \boldsymbol{P}_g \otimes \boldsymbol{P}_p$, so that the test statistic becomes:

$$F_n(AT) = \frac{n}{N^2 tr(\boldsymbol{T}_{AT}\widehat{\boldsymbol{V}}_{\boldsymbol{n}})}\sum_{i=1}^{g}\sum_{s=1}^{p}(\bar{\boldsymbol{R}}_{i.s} - \bar{\boldsymbol{R}}_{i..} - \bar{\boldsymbol{R}}_{..s} + \bar{\boldsymbol{R}}_{...})^2 \ ,$$

(5-18)

where the midranks and their means are defined as in *(5-8)*, and $\widehat{\boldsymbol{V}}_{\boldsymbol{n}}$ in *(5-9)* . The distribution of $F_n(AT)$ under $H_0^F(AT)$ can be approximated by a central $F(\hat{f}_{AT}, \infty)$-distribution. Such that:

$$\hat{f}_{AT} = \frac{\left[tr(\boldsymbol{T}_{AT}\widehat{\boldsymbol{V}}_{\boldsymbol{n}})\right]^2}{tr(\boldsymbol{T}_{AT}\widehat{\boldsymbol{V}}_{\boldsymbol{n}}\boldsymbol{T}_{AT}\widehat{\boldsymbol{V}}_{\boldsymbol{n}})}.$$

(5-19)

Next, we look at the test statistic for $H_0^F(AT)$ tailored to the shape of the underlying distribution given by:

$$F_n^a(AT) = \frac{n}{N^2 tr(\boldsymbol{T}_{AT}\widehat{\boldsymbol{V}}_{\boldsymbol{n}}^{\boldsymbol{a}})}\sum_{i=1}^{g}\sum_{s=1}^{p}(\bar{a}(R_{i.s}) - \bar{a}(R_{i..}) - \tilde{a}(R_{..s}) + \tilde{a}(R_{...}))^2 \ ,$$

(5-20)

where the midranks of the score functions and their means are defined as in (5-12) and $\widehat{\boldsymbol{V}}_{\boldsymbol{n}}^{\boldsymbol{a}}$ in (5-13) . The distribution of $F_n^a(AT)$ under $H_0^F(AT)$ can be approximated by a central $F(\hat{f}_{AT}^a, \infty)$-distribution. Such that:

$$\hat{f}_{AT}^a = \frac{\left[tr(\boldsymbol{T}_{AT}\widehat{\boldsymbol{V}}_{\boldsymbol{n}}^{\boldsymbol{a}})\right]^2}{tr(\boldsymbol{T}_{AT}\widehat{\boldsymbol{V}}_{\boldsymbol{n}}^{\boldsymbol{a}}\boldsymbol{T}_{AT}\widehat{\boldsymbol{V}}_{\boldsymbol{n}}^{\boldsymbol{a}})}.$$

(5-21)

Note that $H_0^F(AT)$ is equivalent to testing if for all monotone transformation of the observations, the expectation can be decomposed into a total effect, group effect and time effect. Importantly, that an additive decomposition exists for every monotone transformation.

Since the validity of the nonparametric hypothesis is invariant under arbitrary monotone transformations of the observation, the equivalence of certain hypothesis can only hold in certain cases. Importantly, if $H_0^F(AT)$ holds then in a two-way location model such that: $X_{iks} \sim F_{\boldsymbol{is}}(x - \mu_{\mathrm{is}}), i = 1, \dots, g; s = 1, \dots., p,$ the following hypothesis are equivalent:

$$H_0^{\mu}(A): \left(\boldsymbol{P}_a \otimes \tfrac{1}{p}\,\mathbf{1}_p'\right)\boldsymbol{\mu} = \mathbf{0} \ \Leftrightarrow \ H_0^F(A): \left(\boldsymbol{P}_g \otimes \tfrac{1}{p}\,\mathbf{1}_p'\right)\boldsymbol{F} = \mathbf{0} \qquad \text{and}$$

$$H_0^{\mu}(T): \left(\frac{1}{g}\,\mathbf{1}_g' \otimes \boldsymbol{P}_p\right)\boldsymbol{\mu} = \mathbf{0} \ \Leftrightarrow \ H_0^F(T): \left(\frac{1}{g}\,\mathbf{1}_g' \otimes \boldsymbol{P}_p\right)\boldsymbol{F} = \mathbf{0}$$

where $\boldsymbol{F} = (F_{11}, \dots, F_{gp})'$ and $\boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{gp})'$ (Brunner, Domhof, & Langer, 2002). This is particularly useful in Section 5.2 as we connect the F1-LD-F1 hypotheses to the 2x2 crossover design hypotheses described in Section 2.2.

Section 5.2: Interpreting Hypotheses in 2x2 Cross-Over Design using F1-LD-F1 Hypotheses

The hypotheses for the F1-LD-F1 design can be used to test the hypotheses identified for 2x2 crossover designs with repeated measures in Section 2.2. The important step is to ensure that in order to test the carryover effects in a crossover design, we must use the generated measures as described in (4-2), whereas to test the direct treatment effects in a crossover design we utilize the generated measures described in *(4.5)*.

5.2.1 Testing the Equality of Carry-Over Effects and Carry-Over Effects over Time in the Crossover Design

Recall that before the direct treatment effect can be tested, the equality of carry-over effects must be established. Moreover, the suggestion by Grizzle (Grizzle, 1965), an alpha value of 0.10 or 0.15 is recommended.

Proposition 5-1: *Assume no interaction effect in the F1-LD-F1 model with the generated measure as described in (4-2), then*

a) *The hypothesis of no carry-over effect in a crossover design described in (2.3) is equivalent to testing for no group effect in the F1-LD-F1 design described (5-1)*
b) *The hypothesis of no carry-over effect over time in a crossover design described in (2.5) is equivalent to testing for no time effect in the F1-LD-F1 design described in (5-8) .*

Proof:

In order to test for the equality of carry-over effects in a 2x2 crossover design with repeated measures, we will utilize new generated measure described in Chapter 4.

$$\mathrm{X_{iks}} = \mathrm{Y_{i1ks}} + \mathrm{Y_{i2ks}} \tag{4-2}$$

Bruner et al. (Brunner, Domhof, & Langer, 2002) recommend a more general model when looking at the nonparametric models. With the assumption that:

$$X_{iks} \sim F_{is}\,, \qquad i = 1{,}2\,;\; s = 1, 2, \ldots, p$$

where $\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \end{pmatrix} = (F_{11} \quad \cdots \quad F_{1p} \quad F_{21} \quad \cdots \quad F_{2p})'$ = is a vector of the marginal distributions, and that the vectors $\boldsymbol{X}_{ik} = (X_{ik1} \quad \cdots \quad X_{ikp})'$ are independent. Under the assumption that the difference in the marginal distribution for each group must be a location shift.

$$F_{1s}(x) = F_{\boldsymbol{s}}(x - \lambda_{\mathrm{As}})$$
$$F_{2s}(x) = F_{\boldsymbol{s}}(x - \lambda_{Bs})\,, \qquad s = 1, 2, \ldots, p$$

(5-22)

Thus, we have a technical formulation for the hypothesis of equality of carryover effects in the 2x2 crossover design, which is equivalent to the group effect in the F1-LD-F1 design with the new measure, given that the interaction effect is not significant. Note that if the interaction test in F1-LD-F1 design is significant then this method cannot be used to test the 2x2 crossover hypotheses as the assumption that the difference in the marginal distribution for each group is a location shift is violated.

Then using the marginal distributions for the new measure, the nonparametric hypothesis becomes:

$$H_0^1 \colon F_{1s} = F_{2s}, \qquad s = 1, \ldots, p$$

(5-23)

A contrast matrix can be formulated as:

$$\boldsymbol{C} = \begin{bmatrix} 1 & & 0 & -1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & -1 \end{bmatrix}_{p\mathrm{x}2p}$$

So the hypothesis in terms of the contrast matrix becomes:

$$H_0^1: \boldsymbol{CF} = \begin{bmatrix} 1 & & 0 & -1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & -1 \end{bmatrix}_{px2p} \begin{bmatrix} F_{11} \\ \vdots \\ F_{1p} \\ F_{21} \\ \vdots \\ F_{2p} \end{bmatrix}_{2px1} = \begin{bmatrix} F_{11} - F_{21} \\ \vdots \\ F_{1p} - F_{2p} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}.$$

(5-24)

Note if there is no interaction effect in the F1-LD-F1 model then this is analogous to:

$$H_0^1: \boldsymbol{C\lambda} = \begin{bmatrix} 1 & & 0 & -1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & -1 \end{bmatrix}_{px2p} \begin{bmatrix} \lambda_{A1} \\ \vdots \\ \lambda_{Ap} \\ \lambda_{B1} \\ \vdots \\ \lambda_{Bp} \end{bmatrix}_{2px1} = \begin{bmatrix} \lambda_{A1} - \lambda_{B1} \\ \vdots \\ \lambda_{Ap} - \lambda_{Bp} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}.$$

(5-25)

Moreover the time effect for this model provides a test for testing the hypothesis of equality of carry-over effects over time, given there is no interaction effect.

$$H_0^3: \lambda_1 = \lambda_2 = \cdots = \lambda_p \,, \qquad \text{where } \lambda_s = \lambda_{As} - \lambda_{Bs}; \quad s = 1, 2, \ldots, p$$

$$\Leftrightarrow H_0^3(T): C_p\boldsymbol{F} = \left(\frac{1}{g}\, \mathbf{1}_g' \otimes \boldsymbol{P_p}\right)\boldsymbol{F} = \begin{bmatrix} \bar{F}_{.1} - \bar{F}_{..} \\ \vdots \\ \bar{F}_{.p} - \bar{F}_{..} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}.$$

(5-26)

If the carry-over effect is not significant, we can proceed to test the direct treatment effects. In the event that carry-over is significant, as indicated earlier, the experimenter is advised to either rerun the experiment with a longer washout period, or others use just the data from the first period to analyze the treatment effects.

5.2.2 **Testing the Equality of Direct Treatment Effects when Carry-Over are Equal and Direct Treatment Effects over Time when Carry-Over are Equal**

Once the equality of carry over effects has been established, direct treatment effects may be tested. The procedure is similar to the preceding section.

Proposition 5-2: *Assume no interaction effect in the F1-LD-F1 model with the generated measure as described in*

$$\mathrm{X_{iks}} = \mathrm{Y_{i1ks}} - \mathrm{Y_{i2ks}} \qquad \textit{(4-5),}$$

then

a) *The hypothesis of no direct treatment effect in a crossover design described in (2.4) is equivalent to testing for no group effect in the F1-LD-F1 design described in (5-1) .*
b) *The hypothesis of no direct treatment effect over time in a crossover design described in (2.6) is equivalent to testing for no time effect in the F1-LD-F1 design described in (5-8).*

Proof:

To measure the direct treatment effect in a crossover design, we can utilize a F1-LD-F1 model with the different set of generated measures described by:

$$\mathrm{X_{iks}} = \mathrm{Y_{i1ks}} - \mathrm{Y_{i2ks}} \qquad (4\text{-}5)$$

Using the model described by Brunner et al. (Brunner, Domhof, & Langer, 2002), given that:

$$X_{iks} \sim F_{is}\ , \qquad i = 1{,}2\ ;\ s\ =\ 1, 2, \ldots, p$$

where $\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_1 \\ \boldsymbol{F}_2 \end{pmatrix} = (F_{11} \quad \cdots \quad F_{1p} \quad F_{21} \quad \cdots \quad F_{2p})'$ is a vector of the marginal distributions, and that the vectors $\boldsymbol{X}_{ik} = (X_{ik1} \quad \cdots \quad X_{ikp})'$ are independent. Such that:

$$F_{1s}(x) = F_s\left(x - \left(\tau_{As} - \tau_{Bs}\right)\right)$$

$$F_{2s}(x) = F_s\left(x - \left(\tau_{Bs} - \tau_{As}\right)\right), \qquad s = 1,2,\dots,p$$

(5-27)

If the interaction effect is not significant then using these generated measures, the hypothesis can be written as:

$$H_0^2: F_{1s} = F_{2s}\ , \qquad s = 1,2,\dots,p$$

(5-28)

The technical formulation for testing the equality of direct treatment effect, when carry-over effect is equal, is analogous to the first test. Moreover, similar to the carry-over effects, this model can be used to test the equality of direct treatment effect (*group effects*) and the equality of direct treatment effect over time (*time effect*) if there is no interaction effect.

5.2.3 *Testing Average Response for Carry-Over Effects and Average Response for Direct Treatment Effects* when Carry-Over Effects are Equal

With the assumption that the effect being tested is equal over time, we can conduct the following two tests using the methods discussed by Sun (Sun, 1997). These tests can also be formulated using the marginal distributions. Consider the hypothesis of equality of carry-over effects in average responses. The hypothesis as described in (2-7):

$$H_0^5: \sum_{s=1}^{p} \lambda_{As} = \sum_{s=1}^{p} \lambda_{Bs}$$

can be tested by using a measure that sums the responses over time and period. So, as described in (4-7) the new generated measure is given by

$$X_{ik} = \sum_{s=1}^{p} Y_{i1ks} + \sum_{s=1}^{p} Y_{i2ks}.$$

Such that $X_{ik} \sim F_i\ , \qquad i = 1,2$

With the assumption

$$F_1(x) = F\left(x - \sum_{s=1}^{p} \lambda_{As}\right)$$

$$F_2(x) = F\left(x - \sum_{s=1}^{p} \lambda_{Bs}\right)$$

(5-29)

So under the assumption that the carry-over effect is the same for the average responses, our null hypothesis becomes:

$$H_0^5\colon\ F_1 = F_2$$

(5-30)

However, this is a c-sample problem instead of an F1-LD-F1 model, so we can use the test statistic described in (4-12).

Testing the next hypothesis of equality of average responses for direct treatment effects when average responses for carry over effects is similar in formulation with the exception that the generated measures take the difference of the sum of observations over time across the two periods. Recall that the hypothesis from (2-8)

$$H_0^6\colon \sum_{s=1}^{p} \tau_{As} = \sum_{s=1}^{p} \tau_{Bs}$$

This can be tested by using a measure generated as follows:

$$\mathrm{X_{ik}} = \sum_{s=1}^{p} \mathrm{Y_{i1ks}} - \sum_{s=1}^{p} \mathrm{Y_{i2ks}}. \qquad (4\text{-}8)$$

As noted earlier, this is a two-sample problem that can be tested using the statistic S_c described in (4-12).

Chapter 6

Numerical Example

In this chapter we will use multiple examples to demonstrate the use of the test statistics derived earlier based on score functions, and compare this to the traditional nonparametric methods. By means of the examples we will demonstrate the supremacy of these adaptive nonparametric methods over the traditional rank-based methods.

Section 6.1: Collagen bits (BITS) Example

As an example of the two methods described previously, consider the numerical example provided by Johnson and Grender (Johnson & Grender, 1993). The example utilizes data from a 2x2 crossover design for an experiment developed to evaluate the presence of collagen bits (BITS) in a 1:8 concentration of Optipranolol suspended in Murocel. There were a total of 12 participants, 7 in the first group which received treatments in the order of BITS/No BITS and 5 in the second group which received treatment in the reverse order i.e. No BITS/BITS. There was a 7-day washout period between the two treatments in each group. The response variable (Y_{ijks}) considered was the difference in intraocular pressure from baseline, this was measured on each eye and the resulting data are presented in Table 6-1

Table 6-1: Change in Right (R) and Left (L) Eye Pressure

Sequence	Sub	Period I		Period II	
		R	L	R	L
BITS/ NO BITS	1	2	3	6	6
	2	1	1	5	7
	3	4	2	4	2
	4	8	5	6	2
	5	6	6	6	7

Sequence	Sub	Period I		Period II	
		R	L	R	L
	6	5	2	5	5
	7	−1	0	4	4
NO BITS/ BITS	8	4	6	7	6
	9	7	6	1	2
	10	6	8	-2	-2
	11	5	8	5	2
	12	5	5	4	2

Therefore, based on the model described in Section 2.1

$$E[Y_{ijks}] = \varphi_{ijs} = \mu_s + \pi_{js} + \tau_{vs} + h\lambda_{v's} \quad (2\text{-}1)$$

the experiment has two groups or sequences, so $i = 1,2$, with $n = n_1 + n_2 = 7 + 5 = 12$ subjects, moreover it's a two period design thus $j = 1,2$ periods; $s = 1,2$ repeated measures since measurements are taken from the right eye and left eye for each subject; and v corresponds to either treatment BITS or NO BITS, and v' is the other treatment. With the given model, we then proceed to test the hypotheses associated with 2x2 crossover design as described in Section 2.2.

6.1.1 Adaptive Nonparametric Procedures Tailored to J&G Method

In order to test each of the six hypotheses indicated in Section 2.2 using the nonparametric adaptive J&G method, first the underlying distribution of the data must be determined. Then the hypotheses can be tested and compared to the results to those obtained from the traditional nonparametric rank-based method proposed by Johnson and Grender (Johnson & Grender, 1993). Recall caution is taken to conduct the six hypotheses in order, since the results from one may impact the validity of the next test.

Testing the Equality of Carry-Over Effects

From the previous sections we know that the null-hypothesis of equality of carry-over effect can be written as:

$$H_0^1: \lambda_{As} = \lambda_{Bs}, \qquad s = 1, 2, \ldots, p \tag{2-3}$$

In order to test this hypothesis we will use the measure defined by:

$$X_{iks} = Y_{i1ks} + Y_{i2ks} \tag{4-2}$$

So for each repeated measure, s, we use the sum of the within-subject measures across the two periods, the new measures generated for the two repeated measures, and their corresponding ranks are illustrated in Table 6-2. Next we need to identify the shape of the underlying distribution, in order to apply the appropriate score function. For the right eye the selector functions are:

$$\bar{Q}_1 = \frac{(n_1 Q_{1,1} + n_2 Q_{1,2})}{(n_1 + n_2)} = \frac{(7 * 0.941 + 5 * 0.4)}{(7 + 5)} = 0.716 .$$

And,

$$\bar{Q}_2 = \frac{(n_1 Q_{2,1} + n_2 Q_{2,2})}{(n_1 + n_2)} = \frac{(7 * 2.316 + 5 * 2.333)}{(7 + 5)} = 2.323 .$$

Thus, the distribution is moderate-tailed and therefore the score function used is

$$\varphi_W(u) = u \qquad 0 < u < 1$$

where $u = \frac{rank\ of\ the\ measure\ over\ the\ combined\ sample}{(n_1+n_2)+1}$. Similarly for the left eye we obtain selector functions $(\bar{Q}_1, \bar{Q}_2) = (1.646, 2.233)$ indicating a light-tailed distribution, therefore the score function used is :

$$\varphi_{ML}(u) = \begin{cases} u - \frac{1}{4} & 0 < u \leq 1/4 \\ 0 & 1/4 < u \leq 3/4 \\ u - \frac{3}{4} & 3/4 < u < 1 \end{cases}$$

Based on the adaptive ranks, as shown in Table 6-2 we can then calculate the test statistic, $W^a = 0.548$ with a p-value 0.76. There is not enough evidence to reject the null hypothesis, indicating carryover effects are equal for the two treatments. However, when the test was conducted using traditional nonparametric rank-based method, we obtained $\chi^2 = .77$, p-value=0.68 asymptotically, or p-value=0.75 using 792 permutations for the W-statistic observed. Even with a higher cut-off of $\alpha = 0.15$ as suggested by Grizzle (Grizzle, 1965), we fail to reject the null hypothesis in both cases.

Table 6-2: Generated Measure for Testing Carry-Over Effects (Sum over Period), Corresponding Ranks, and Adaptive Ranks of the Right and Left Eye Change in Pressure based on (3-5)

Sequence	Sub	Sum		Sum Ranks		Adaptive Ranks	
		R	L	R	L	R	L
BITS/ NO BITS	1	8	9	5	9	0.385	0
	2	6	8	3	7.5	0.231	0
	3	8	4	5	1.5	0.385	-0.018
	4	14	7	12	5	0.923	0
	5	12	13	11	12	0.846	0.030
	6	10	7	8.5	5	0.654	0
	7	3	4	1	1.5	0.077	-0.018
NO BITS/ BITS	8	11	12	10	11	0.769	0.009
	9	8	8	5	7.5	0.385	0
	10	4	6	2	3	0.154	0
	11	10	10	8.5	10	0.654	0
	12	9	7	7	5	0.538	0

Testing the Equality of Direct Treatment Effects when Carry-Over Effects are Equal

After the equality of carry-over effects has been established, we proceed to test the direct treatment effects. The hypothesis from Section 2.2 can be written as:

$$H_0^2: \tau_{As} = \tau_{Bs}, \qquad s = 1,2,\ldots,p \tag{2-4}$$

Then using the generated measures as defined in Section 4.1, we have:

$$X_{iks} = Y_{i1ks} - Y_{i2ks} \tag{4-4}$$

Based on these new generated measures, the selector functions $(\bar{Q}_1, \bar{Q}_2)$ are $(0.676, 3.531)$ for the right eye, and $(1.393, 2.362)$ for the left eye, indicating a moderate-tailed distribution in each case. The new generated measures, their ranks and the adaptive ranks are illustrated in Table 6-3.

Table 6-3: Generated Measure for Testing Direct Treatment Effects (Differences over Period), Corresponding Ranks, and Adaptive Ranks of the Right and Left Eye Change in Pressure

Sequence	Sub	Difference		Diff Ranks		Adaptive Ranks	
		R	L	R	L	R	L
BITS/ NO BITS	1	-4	-3	2.5	3.5	0.192	0.269
	2	-4	-6	2.5	1	0.192	0.077
	3	0	0	6.5	6.5	0.500	0.5
	4	2	3	10	8.5	0.769	0.654
	5	0	-1	6.5	5	0.500	0.385
	6	0	-3	6.5	3.5	0.500	0.269
	7	-5	-4	1	2	0.077	0.154
NO BITS/ BITS	8	-3	0	4	6.5	0.308	0.5
	9	6	4	11	10	0.846	0.769
	10	8	10	12	12	0.923	0.923
	11	0	6	6.5	11	0.500	0.846
	12	-1	3	9	8.5	0.692	0.654

Since the distribution for both the right and the left eye are moderate-tailed, the test statistic under the adaptive procedure is similar to the test statistic for the rank-based analysis, as expected. $W^a = W = 7.12$ with a p-value of 0.03. The exact test based on the permutation distribution resulted in a p-value of 0.01. Hence, there is sufficient evidence to reject the null hypothesis, indicating there is a difference in the two treatment effects.

Testing the Equality of Carry-Over Effect over Time

In order to test the next two hypotheses, the procedure is the same except the new measures are differences between two repeated measures. Thus, from Section 2.2 the test for carry-over effect over time can be written as:

$$H_0^3: \lambda_1 = \lambda_2 = \cdots = \lambda_p, \qquad \text{where } \lambda_s = \lambda_{As} - \lambda_{Bs}; \quad s = 1, 2, \ldots, p \tag{2-5}$$

and can be measured by taking the sum across period of the difference in measures of the right and the left eye in each period. Notice in this particular case, we now have a c −sample problem, and so we can utilize the test statistic S_c as described by Sun (Sun, 1997). The generated measures, their corresponding ranks, and the adaptive ranks are summarized in Table 6-4.

The underlying distribution has $(\bar{Q}_1, \bar{Q}_2)$ values of $(1.050, 2.833)$ indicating a moderate-tailed distribution. Thus, $S_c = 0.172$ with a p-value of 0.678, which is the same as the p-value obtained under the two-sample Wilcoxon rank-sum test. Therefore we conclude there is insufficient evidence to reject the null hypothesis, supporting that the carry-over effect are equal for each repeated measure.

Table 6-4: Generated Measure for Testing Carry-Over Effects over Time, Corresponding Ranks, and Adaptive Ranks of the Difference (R-L) between Right and Left Eye Change in Pressure based on (3-5)

Sequence	Sub	Period I	Period II	Sum	Sum Rank	Adaptive Rank
		(R-L)	(R-L)			
BITS/ NO BITS	1	-1	0	-1	4.5	0.346
	2	0	-2	-2	1.5	0.115
	3	2	2	4	11	0.846
	4	3	4	7	12	0.923
	5	0	-1	-1	4.5	0.346
	6	3	0	3	10	0.769
	7	-1	0	-1	4.5	0.346
NO BITS/ BITS	8	-2	1	-1	4.5	0.346
	9	1	-1	0	7.5	0.577
	10	-2	0	-2	1.5	0.115
	11	-3	3	0	7.5	0.577
	12	0	2	2	9	0.692

Testing the Equality of Direct Treatment Effect over Time when Carry-Over Effect over Time is Equal

After establishing the equality of carry-over effect over time, we can proceed to the test direct treatment effect over time as described earlier by the hypothesis:

$$H_0^4: \tau_1 = \tau_2 = \ ... = \ \tau_p \ , \qquad \text{where } \tau_s = \tau_{As} - \tau_{Bs};\ s \ = \ 1, 2, \ldots, p \qquad (2\text{-}6)$$

Using the difference over period of the difference in each repeated measure, we obtain the new generated measure, once again obtaining a two-sample problem. The new measures and their corresponding ranks, and adaptive ranks are shown in Table 6-5. The underlying distribution was moderate-tailed since on $(\bar{Q}_1, \bar{Q}_2) = (1.465, 3.037)$. Appropriate score function was selected accordingly and the resulting test statistic obtained was $S_c = 3.879$ *with* a p-value of 0.049. As expected due to the moderate-tailed shape of the

underlying distribution, the p-value for the Wilcoxon rank-sum test is the same, indicating direct treatment effect over time is marginally significant.

Table 6-5: Generated Measure for Testing Direct Treatment Effects over Time, Corresponding Ranks, and Adaptive Ranks based on (3-5)

Sequence	Sub	Period I (R-L)	Period II (R-L)	Difference	Diff Rank	Adaptive Rank
BITS/ NO BITS	1	-1	0	-1	6	0.462
	2	0	-2	2	10.5	0.808
	3	2	2	0	8	0.615
	4	3	4	-1	6	0.462
	5	0	-1	1	9	0.692
	6	3	0	3	12	0.923
	7	-1	0	-1	6	0.462
NO BITS/ BITS	8	-2	1	-3	2	0.154
	9	1	-1	2	10.5	0.808
	10	-2	0	-2	3.5	0.269
	11	-3	3	-6	1	0.077
	12	0	2	-2	3.5	0.269

Testing the Average Responses

Under the assumption that the effects do not vary over time, we can continue testing the equality of average responses for carry-over effects (hypothesis described by (2-7)) and the equality of average response for direct treatment effects when average responses for carry-over effects are equal (hypothesis described in (2-8)). Note that both these tests are two-sample problems, thus we employ the testing procedures of Sun (Sun, 1997). Moreover, in this example the direct treatment over time was marginally significant with $S_c = 3.879$ with a p-value of 0.049 therefore caution must be taken when interpreting the results of the test for average responses for direct treatment effects.

As discussed in Section 4.1, the generated measure for these hypotheses are the sum (or differences) over the two periods for the sum of responses in each repeated measure (R+L), represented in Table 6-6 along with their ranks and adaptive ranks based on (3-5). First, we test the average responses for carry-over effects. With $(\bar{Q}_1, \bar{Q}_2) = (0.905, 3.056)$ the indicator functions suggest that the underlying distribution is moderate-tailed. Using the appropriate score function, we obtain $S_c = 0.007$ with a p-value of 0.935. Under the Wilcoxon rank-sum test a similar p-value was obtained indicating there is insufficient evidence to reject the null hypothesis of equal average responses for carry-over effects.

Thus, we continue with the next hypothesis testing the average responses of direct treatment effects. Note as mentioned earlier, since direct treatment effects over time was marginally significant based on the results of hypothesis H_0^4, caution must be taken when interpreting the results from this test. The generated measures for this test are obtained by taking the difference across periods of the sum of responses from all repeated measures (R+L) as shown in Table 6-6. The underlying distribution for these generated measures is light-tailed as indicated by $(\bar{Q}_1, \bar{Q}_2) = (1.287, 2.139)$. Based on that we attain the test-statistic $S_c = 3.004$ with a p-value of 0.083, which fails to reject the null hypothesis that average responses for direct treatment effects are equal. Under the Wilcoxon rank-sum test the p-value was 0.0344 supporting the alternate hypothesis that the average response for direct treatment effects are not equal. However, based on the work of Sun (Sun, 1997), we know the supremacy of adaptive procedures over the rank-based nonparametric procedures in a two-sample problem indicating that based on the evidence we cannot conclude that the average response for direct treatment effects are not equal.

Table 6-6: Generated Measure for Testing Average Responses, Corresponding Ranks, and Adaptive Ranks of the Sum (R+L) of Right and Left Eye Change in Pressure based on (3-5)

Seq	Sub	Period I (R+L)	Period II (R+L)	Sum	Sum Ranks	Sum Adaptive Ranks	Diff	Diff Ranks	Diff Adaptive Rank
BITS/ NO BITS	1	5	12	17	7.5	0.577	-7	3	0
	2	2	12	14	4	0.308	-10	1	-0.030
	3	6	6	12	3	0.231	0	7	0
	4	13	8	21	10	0.769	5	9	0
	5	12	13	25	12	0.923	-1	6	0
	6	7	10	17	7.5	0.577	-3	4.5	0
	7	-1	8	7	1	0.077	-9	2	-0.009
NO BITS/ BITS	8	10	13	23	11	0.846	-3	4.5	0
	9	13	3	16	5.5	0.423	10	11	0.009
	10	14	-4	10	2	0.154	18	12	0.030
	11	13	7	20	9	0.692	6	10	0
	12	10	6	16	5.5	0.423	4	8	0

6.1.2 **Adaptive Nonparametric Procedures Tailored to F1-LD-F1 Design**

Another way of testing the hypotheses for a 2x2 crossover design with repeated measures was described in Chapter 5. In this section, we will work through the BITS example using the adaptive nonparametric procedures tailored to the F1-LD-F1 design and compare the results to those obtained by the traditional rank-based F1-LD-F1 design. However, since hypotheses 5 and 6 of the crossover design for testing the equality of

average responses are in fact two-sample problems, those will not be repeated in this section.

Testing the Equality of Carry-Over Effects and Carry-Over Effects over Time in the Crossover Design

The first F1-LD-F1 design we will look at tests the carry-over effects, and carry-over effect over time in the crossover design. The generated measures used for these are described by:

$$\mathrm{X_{iks}} = \mathrm{Y_{i1ks}} + \mathrm{Y_{i2ks}} \tag{4-2}$$

Recall in a F1-LD-F1 design we consider the R_{iks} as the rank of X_{iks}, among all $N = p \cdot \sum_{i=1}^{a} n_i$ observations. Thus, in order to determine the shape of the underlying distribution, we consider the combined sample of all $N = 2(7+5) = 24$ observations. Based on the combined sample, we determine that the underlying distribution is moderate-tailed for these generated measures. The $\bar{Q}_1$ and $\bar{Q}_2$ values are 0.994 and 2.605 respectively.

Thus, using the appropriate score function from (3-5), we first consider the interaction effect between group and time. The interaction effect was not significant with a test-statistic of 0.648 and a p-value of 0.517. This implies that we can use the tailored F1-LD-F1 design to test the crossover hypotheses. The group effect in the F1-LD-F1 design is not significant with a test-statistic of -0.385 and a p-value of 0.700 indicating that carry-over effects in the crossover design are equal. Finally, consider the time effect in the F1-LD-F1 design, which has a test statistic of 1.031 and a p-value of 0.303 suggesting that the carry-over over time in the crossover design is equal. Since the underlying distribution was moderate-tailed, the p-values obtained from the rank-based F1-LD-F1 model were exactly the same in each case, as one might expect.

Testing the Equality of Direct Treatment Effects when Carry-Over are Equal and Direct Treatment Effects over Time when Carry-Over are Equal

After the equality of carry-over effects has been established, we continue to test the direct treatment effects. Using the generated measure given by:

$$X_{iks} = Y_{i1ks} - Y_{i2ks} \tag{4-5}$$

we can use the F1-LD-F1 design to test for the direct treatment effects and direct treatment effects over time in the crossover design. The shape of the underlying distribution for the new generate measures is light-tailed based on score indicator values of $(0.824, 1.934)$. Thus, the results from the tailored F1-LD-F1 method mirror those obtained by the rank-based F1-LD-F1 method. The interaction effect between group and time is not significant with test-statistic of 0.574, which results in a p-value of 0.566. Therefore by Proposition 5-2, we can use the F1-LD-F1 model to test the crossover hypotheses.

Next consider the group effect which has a test-statistic of -2.132 and a p-value of 0.033 which implies that the direct treatment effect is not equal. However, the time effect has a test-statistic of -0.318 with a p-value of 0.750 indicating direct treatment effects over time are equal. Note with the J&G method, the direct treatment effects over time was marginally significant. Moreover, if we look at the results from the rank-based F1-LD-F1 model, the interaction effect had a test statistic of 1.865 with a p-value of 0.062. Based on the argument provided by Grizzle (Grizzle, 1965), since we want the interaction effect to be insignificant, a higher alpha value is recommended. Thus indicating that we cannot use the rank-based F1-LD-F1 model to test the crossover hypotheses for direct treatment effects.

Section 6.2 Sleep Apnea Effect on Blood Pressure Example

Our next example is in fact the inspiration for this research. An experiment was conducted by Alex et al. (Alex, et al., 2014) in order to better understand the relation between severity of sleep apnea and arterial blood pressure (BP), and to develop new metrics to analyze the effects of sleep apnea on BP. In this study, 26 healthy subjects were asked to do a series of breath hold experiments in a repeatable and controlled environment to simulate sleep apnea. BP waveforms were continuously monitored and the features extracted for study were: Pulse pressure, Slope of systolic/diastolic trend and Area under waveform (Area), in addition to the traditionally used metrics of Systolic, Diastolic and Mean arterial blood pressures (MAP). These features are illustrated in Figure 6-1.

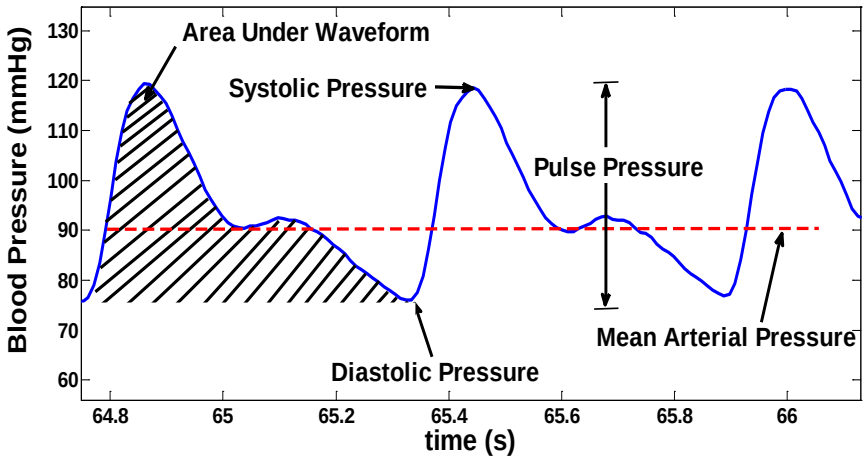

Figure 6-1 Different Blood Pressure Features Extracted From the Waveform for the Experiment (Alex R. M., 2010)

Further, to measure the effects of sleep apnea severity, two separate experimental protocols were implemented with different inter-breath hold intervals. Moreover, the study also measured the effect of different postures on blood pressure thus the subjects' posture varied between sitting and supine position. In our study however, we focus on the slope of

systolic trend, and slope of diastolic trend measures obtained in the supine position (Figure 6-2).

Figure 6-2 Data Collected While the Subject is in Supine Position (Alex R. M., 2010)

In this experiment a baseline blood pressure was measured while the subject was in a normal breathing, or resting stage. Then the subject is asked to hold their breath for as long as they can, followed by a timed normal breathing period. This process is repeated five times. There are two protocols designed to account for the severity of sleep apnea. Protocol A, which corresponds to a less severe apnea, allows a normal breathing time of 90 second between each breath hold, and a pre and post experiment normal breathing rest of 60 seconds. In Protocol B, to reflect a more severe apnea, the interval between breath holds is reduced to 30 seconds, while the pre and post experiment normal breathing time remains constant at 60 seconds. Each subject follows both protocols, although which protocol is implemented first is randomly selected. The two protocols are demonstrated in Figure 6-3.

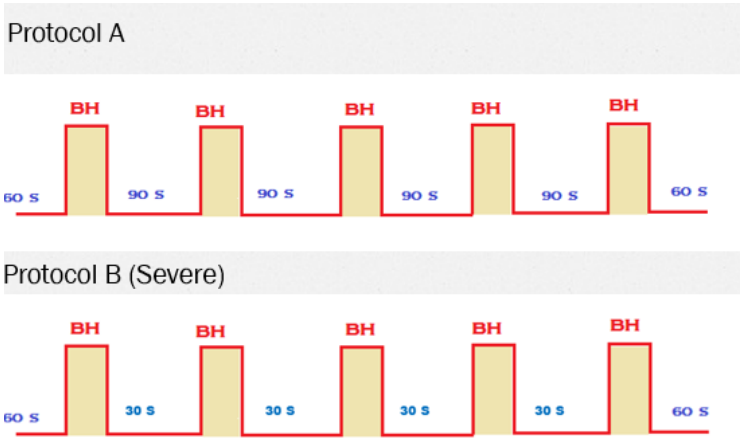

Figure 6-3 Two Breath Hold (BH) Protocols Simulating Sleep Apnea (Protocol B is More Severe Form of Sleep Apnea Compared to Protocol A) (Alex R. M., 2010)

Due to the nature of the experiment, to analyze the effects of the severity of sleep apnea on BP (Protocol A versus Protocol B), a 2x2 crossover design with repeated measures is utilized, with $i = 1{,}2$ groups, and $s = 1, \ldots, 5$ repeated measures. Moreover, since the normality assumption was not satisfied, a nonparametric analysis was recommended. In order to test the hypotheses of a crossover design we employed and compared the adaptive and rank-based methods for both the J&G design and the F1-LD-F1 design, where the first step was to determine the shape of the underlying distribution for each generated measure. The results are summarized in Table 6-7. Note that the J&G design looks at the shape of the underlying distribution for each repeated measure separately, while the F1-LD-F1 design looks at the overall shape of the underlying distribution for all $N = p \cdot \sum_{i=1}^{a} n_i$ observations.

Table 6-7 Shape of Underlying Distribution for Slope of Systolic Trend and Slope of Diastolic Trend

Hypotheses		**Systolic Slope**	**Diastolic Slope**
Hypothesis 1:	s=1	Moderate Tailed	Moderate Tailed
Equal Carry-Over effects	s=2	Light Tailed	Moderate Tailed
	s=3	Right Skewed	Moderate Tailed
	s=4	Right Skewed	Moderate Tailed
	s=5	Right Skewed	Moderate Tailed
	overall	**Moderate Tailed**	**Moderate Tailed**
Hypothesis 2:	s=1	Right Skewed	Left Skewed
Equal Direct Treatment effects	s=2	Moderate Tailed	Moderate Tailed
When Carry-Over Effect is Equal	s=3	Moderate Tailed	Moderate Tailed
	s=4	Right Skewed	Light Tailed
	s=5	Light Tailed	Moderate Tailed
	overall	**Moderate Tailed**	**Moderate Tailed**
Hypothesis 3:	m=1	Moderate Tailed	Heavy Tailed
Equal Carry-Over Effect Over Time	m=2	Moderate Tailed	Right Skewed
	m=3	Light Tailed	Light Tailed
	m=4	Moderate Tailed	Moderate Tailed
Hypothesis 4:	m=1	Moderate Tailed	Moderate Tailed
Equal Direct Treatment Effect Over Time	m=2	Moderate Tailed	Moderate Tailed
When Carry-Over Effect Over Time is Equal	m=3	Moderate Tailed	Left Skewed
	m=4	Moderate Tailed	Moderate Tailed
Hypothesis 5: Equal Average Response for Carry-over Effects		Moderate Tailed	Light Tailed
Hypothesis 6: Equal Average Response for Direct-Treatment Effects		Moderate Tailed	Moderate Tailed

Based on this information, we were able to select the appropriate score functions and compare the four methods of testing the crossover hypotheses. However, the F1-LD-F1 considers all $N = p \cdot \sum_{i=1}^{g} n_i$ observations together. Thus, for an F1-LD-F1 model we

have a large sample size. Using the recommendation by Brunner et al. (Brunner, Domhof, & Langer, 2002) we use the Wald Type Statistic (WTS) rather than the ANOVA Type Statistic (ATS) used earlier. This is summarized in Table 6-8 for the slope of systolic trend and in Table 6-9 for the slope of diastolic trend, along with the result obtained from the J&G method.

Table 6-8 Test Statistics and P-Values for Analysis of Slope of Systolic Trend

Testing the Slope of Systolic Trend	**J&G**		**Adaptive J&G**		**F1-LD-F1 (WTS)**		**Adaptive F1-LD-F1 (WTS)**	
	W	p-value	W^a	p-value	Q_n	p-value	Q_n^a	p-value
Hypothesis 1	3.178	0.204	16.861	<0.001	0.321	0.571	0.321	0.571
Hypothesis 2	5.811	0.055	20.639	<0.001	2.806	0.094	2.806	0.094
Hypothesis 3	6.067	0.048	4.281	0.117	6.762	0.149	6.762	0.149
Hypothesis 4	1.655	0.437	1.655	0.437	5.814	0.213	5.814	0.213
Hypothesis 5	90	0.623	1.492	0.222				
Hypothesis 6	59	0.286	1.432	0.232				
Interaction Effect ($X_{iks} = Y_{i1ks} + Y_{i2ks}$**)**					7.413	0.116	7.413	0.116
Interaction Effect ($X_{iks} = Y_{i1ks} - Y_{i2ks}$**)**					5.387	0.250	5.387	0.250

Table 6-9 Test Statistics and P-Values for Analysis of Slope of Diastolic Trend

Testing the Slope of Diastolic Trend	**J&G**		**Adaptive J&G**		**F1-LD-F1**		**Adaptive F1-LD-F1**	
	W	p-value	W^a	p-value	Q_n	p-value	Q_n^a	p-value
Hypothesis 1	6.006	0.050	6.006	0.050	0.010	0.920	0.010	0.920
Hypothesis 2	4.400	0.111	14.001	<0.001	0.470	0.493	0.470	0.493
Hypothesis 3	5.994	0.050	15.400	<0.001	1.274	0.281	1.274	0.281
Hypothesis 4	4.716	0.095	14.862	<0.001	0.108	0.966	0.108	0.966
Hypothesis 5	80	>0.999	1.553	0.213				
Hypothesis 6	76	0.856	0.049	0.826				
Interaction Effect ($X_{iks} = Y_{i1ks} + Y_{i2ks}$**)**					0.840	0.477	0.840	0.477
Interaction Effect ($X_{iks} = Y_{i1ks} - Y_{i2ks}$**)**					1.099	0.351	1.099	0.351

Notice from Table 6-8 that the adaptive J&G method indicates that there is a significant difference in carry-over effect (hypothesis 1) for the slope of systolic trend which the rank-based J&G method was unable to detect. This is further validated by looking at the slope of the diastolic trend from Table 6-9 which also suggests a significant carry-over effect. Recall based on Grizzle's (Grizzle, 1965) recommendation we use an alpha value of .15 to test the carryover effect. This is an important distinction since a significant carry-over effect impacts the validity of the remaining hypotheses.

Moreover, the F1-LD-F1 (WTS) for the slope of systolic trend indicates than the interaction effect is significant and thus, this model cannot be used to test the hypotheses of the crossover design. However, this is not the case for the slope of diastolic trend in Table 6-9. In fact here the results contradict those obtained by the J&G method. Since the J&G method analyzes the shape of the distribution for each of the repeated measures separately, rather than the F1-LD-F1 method which combines all observations across all repeated measures, in our opinion the former method is more reliable.

Chapter 7

Conclusion and Discussions

The objective of the research was to tailor existing methods for nonparametric analysis of 2x2 crossover designs to the shape of the underlying distribution of the data given. Hence, existing methods developed by Johnson and Grender (Johnson & Grender, 1993) and Brunner et al. (Brunner, Domhof, & Langer, 2002) were modified using adaptive procedures as described by Hogg et al. (Hogg, Fisher, & Randles, 1975) and revised by Hill et al. (Hill, Padmanabhan, & Puri, 1988). This idea was supported by the work of Sun (Sun, 1997) who showed that in a C-sample problem adaptive nonparametric method is better since the lengths of the simultaneous confidence intervals for adaptive nonparametric methods are shorter compared to those of non-adaptive nonparametric and parametric methods, and the relative asymptotic efficiency is better compared to non-adaptive nonparametric methods. We extended this idea from a C-sample to a 2x2 crossover design. Some observations are highlighted in the following sections.

Section 7.1 Comparing Adaptive to Non-adaptive Methods

It was observed that for moderate-tailed and heavy–tailed symmetric distributions, there was no difference in the adaptive or non-adaptive methods. However, when the distribution was skewed or light-tailed, differences in the test-statistic were detected. This difference was particularly important in the sleep apnea example described in Section 6.2 when looking at the slope of systolic trend. In this example, the traditional nonparametric method of Johnson and Grender failed to detect reject the null hypothesis of equal carry-over effects, however with the adaptive modification the null hypothesis was rejected indicated that the experiment should either be redone, or we should only use the data from period I to examine the direct treatment effects. This conclusion was also supported by the

results from the slope of diastolic trend where both the adaptive and non-adaptive J&G method resulted in rejecting the null hypothesis of equal carry-over effects. In our opinion since both the systolic slope and diastolic slope were from the same pool of subjects, the results observed for the hypothesis of equal carry-over effects should be similar.

Similarly, when looking at the F1-LD-F1 design, the collagen bits example in Section 6.1 was interesting. While the traditional F1-LD-F1 method could not be used to test the direct treatment effects due to signification group and time interaction, when modified using adaptive procedure to account for the light-tailed symmetric shape of the underlying distribution, it was observed that that the interaction was not significant and thus, the model could be used to test the 2x2 crossover hypotheses. However, with large sample sizes or repeated measures, such as in the case of sleep apnea example in Section 6.2, it was more difficult to apply adaptive procedure to the traditional F1-LD-F1 design, as discussed in Section 7.2.

Section 7.2 Comparing J&G Method to F1-LD-F1 Method

Two methods described in this paper are the J&G method (Johnson & Grender, 1993) and the F1-LD-F1 method (Brunner, Domhof, & Langer, 2002). A key difference in the two methods is how the data is ranked, while R_{iks} is the rank for each X_{iks} across $n = \sum_{i=1}^{g} n_i$ observations for each s^{th}measure in the J&G method, the F1-LD-F1 method combines all p repeated measures so that R_{iks} is the rank of X_{iks}, among all $N = p \cdot \sum_{i=1}^{g} n_i$ observations. However, by combining observations from all repeated measures, the F1-LD-F1 method fails to account for the attributes of the underlying distribution for each repeated measure. This can particularly be observed in Table 6-7 looking hypothesis 1 in the case of the slope of systolic trend where most repeated measures are not moderate-tailed symmetric but when we look at all N observations together, the data seems to be

moderate-tailed. This may be due to the fact that we have 5 repeated measures thus a large sample size N.

Moreover, while the J&G method is specifically designed to test the hypotheses for crossover design, the F1-LD-F1 method is sometimes restricted by the propositions defined in Section 5.2 when the interaction term is significant. Therefore, in our opinion the former method is more reliable.

Section 7.3 Limitations

Although based on the work of Sun (Sun, 1997) and the numerical examples studied for this research both indicate that the nonparametric adaptive procedure for crossover design with repeated measures is better than the traditional nonparametric methods for crossover design with repeated measures, however the research has certain limitations.

The examples used in each case were preexisting datasets provided to us, therefore the true results for the hypotheses were unknown. We could extend the study to create simulated data sets were predetermined results for the six hypotheses tested and then identify which of the four methods is best at replicating those results.

Another issue faced was the sample size issue, especially with the F1-LD-F1 case. For the F1-LD-F1 method we proposed to apply the adaptive procedure for small sample sizes, however this is often not the case if we have multiple repeated measures since the F1-LD-F1 combines the observations across all repeated measures.

Finally, as with the case of any statistical model, the test statistic and generated measures provided here are specifically to test the six hypotheses described in Section 2.2. However, it is upon the experimenter of determine the validity of the hypotheses in

real-world terms, and determine the validity of the 2x2 crossover design model for their experiment.

Appendix A

Data From the Sleep Apnea Example

Table 0-1 Data for Slope of Systolic Trend

Sequence	Sub	ST1 2	ST1 4	ST1 6	ST1 8	ST1 10	ST2 2	ST2 4	ST2 6	ST2 8	ST2 10
1	3	1.433365	0.95523	1.084191	0.605368	0.672902	1.119691	1.899888	1.589624	1.344086	1.33907
1	6	0.818107	0.679277	0.731157	0.664597	0.660964	0.586591	0.907461	0.654896	0.6315	0.553009
1	11	0.990327	0.932311	1.280855	0.85872	0.891577	0.972024	1.592012	1.276553	1.308612	1.256946
1	14	0.31216	0.787068	0.649242	0.401476	0.908363	1.965116	1.697505	-0.10073	0.795615	1.344323
1	15	1.743167	0.75088	0.579732	1.118014	0.564022	1.236704	1.203729	1.117942	1.130481	0.585753
1	16	0.893415	0.842711	1.189994	1.206263	0.875676	0.82586	0.496797	1.377594	1.128223	0.655106
1	17	0.870137	1.351827	1.066114	3.286514	0.60759	1.193422	1.494326	1.073147	0.846728	0.880282
1	18	1.318665	1.020203	1.060041	0.892884	1.180401	1.006404	1.688725	0.722106	1.323583	1.246562
1	25	1.232255	0.773087	0.934827	1.051683	0.802462	1.081932	0.997791	0.959038	0.84885	0.736026
1	27	1.167128	1.046115	0.879355	0.919296	1.291751	0.852619	1.195883	0.963292	1.54538	0.711665
2	1	2.089123	1.898565	2.504945	2.108077	1.893189	2.229742	2.2044	2.263624	2.806488	2.880644
2	2	1.679709	1.08767	2.325147	1.327482	1.698273	1.264595	2.036816	1.02914	2.426875	1.856125
2	4	0.292335	0.423303	0.57364	0.424304	0.57748	0.475057	0.475961	0.441884	0.512591	0.459488
2	5	0.940403	0.897161	0.330365	1.389076	0.920277	1.156629	1.091258	1.191778	1.321212	0.254425
2	7	0.832518	1.261772	1.318729	1.249405	1.13381	1.100204	1.749151	0.951884	0.734926	0.993977
2	8	1.197632	1.648106	1.397585	2.073675	1.829748	1.620038	0.998026	1.498683	1.8528	1.042208
2	10	1.129965	0.860305	1.057365	0.879373	0.893259	0.833983	0.796109	0.748023	0.8292	0.493605
2	12	0.218515	0.368614	0.801166	0.34679	0.359231	0.323118	0.221543	0.415916	0.494428	0.974632
2	13	0.572927	0.724654	0.604085	0.690043	0.520954	-0.08529	0.244068	0.711165	0.497568	0.706826
2	19	2.406304	1.164706	0.825318	0.68063	0.595274	1.021506	0.075173	1.965716	1.054371	0.7937
2	20	0.769469	1.00983	0.641886	0.906909	1.569029	0.988003	1.076628	0.719185	1.413388	1.319075
2	21	0.792039	1.243799	0.911348	1.051952	1.040248	0.340992	1.3182	0.186366	1.02299	0.471149
2	22	0.772357	1.369351	1.495304	0.975168	1.20853	1.052675	0.936809	1.07896	1.033591	1.342287
2	23	0.48357	0.623172	0.77097	0.970364	0.543473	0.898355	0.259419	1.265781	0.484595	0.715913
2	24	0.523627	0.612846	0.890227	1.579735	0.72517	0.547783	0.627419	0.697304	0.994643	0.56835
2	26	1.189519	1.125358	0.694809	1.000747	1.0286	0.628104	2.865978	0.888764	0.700785	1.983465

Table 0-2 Data for Slope of Diastolic Trend

Sequence	Sub	ST1 2	ST1 4	ST1 6	ST1 8	ST1 10	ST2 2	ST2 4	ST2 6	ST2 8	ST2 10
1	3	0.864481	0.67684	0.973713	0.527332	0.501993	0.511162	1.118513	0.873574	0.829553	0.686144
1	6	0.769912	0.621037	0.779812	0.64816	0.568255	0.505068	0.834901	0.613598	0.556842	0.509019
1	11	0.570306	0.710074	0.992727	0.64734	0.70188	0.77003	1.321779	1.019563	1.027622	0.824891
1	14	0.214157	0.441403	0.294442	-0.09783	0.617368	0.926002	2.781038	-0.05036	0.4572	1.334781
1	15	1.006665	0.66745	0.635346	0.523069	0.42319	1.528611	0.905746	0.858616	0.59785	0.655727
1	16	0.603037	0.615479	0.90965	0.867676	0.506823	0.533866	0.374708	0.997666	0.80608	0.461008
1	17	0.133866	0.544955	0.567776	1.165346	-0.0816	0.546329	1.014713	0.761755	0.480093	0.49544
1	18	1.021984	0.763382	0.629403	0.639445	1.130142	0.698041	1.323654	0.584152	1.11138	1.063217
1	25	0.770766	0.448815	0.653497	0.993116	0.942696	1.131153	-0.16983	0.611798	0.347395	0.762023
1	27	0.899263	1.368214	0.883983	0.782694	0.671354	0.520781	0.900055	0.811662	1.262795	0.527873
2	1	1.632993	1.501814	1.359796	1.781435	2.071422	1.399204	1.35171	1.929044	1.594386	0.982429
2	2	1.027177	1.510289	0.723758	1.801988	1.30308	1.388089	0.819657	1.815525	1.024468	1.264564
2	4	0.46246	0.425371	0.370501	0.348924	0.296053	0.208774	0.460243	0.545083	0.338422	0.350911
2	5	0.910099	1.076205	0.81753	1.095711	0.345058	0.941531	0.708682	0.493737	1.282727	0.803095
2	7	1.040946	1.426316	0.768181	0.55603	0.945327	0.80507	0.813237	1.046944	1.36009	1.133375
2	8	0.981108	0.429257	0.59193	1.050676	0.658543	0.30407	0.934275	0.763624	1.394695	1.045013
2	10	0.691895	0.644469	0.718641	0.667507	0.438443	1.003316	0.79055	0.915331	0.772949	0.750898
2	12	0.478709	0.221744	0.614518	0.508209	0.637141	0.288278	0.322798	0.773002	0.306478	0.332395
2	13	-0.17058	0.199631	0.516286	0.331825	0.674984	0.641015	0.634812	0.584807	0.616918	0.456759
2	19	0.412634	-0.11268	1.122479	1.058303	0.731629	3.032795	1.296722	0.519653	0.443891	0.422281
2	20	0.725685	1.206938	0.921576	1.379203	1.614151	0.76774	1.00819	0.626575	0.941792	1.595279
2	21	0.273543	0.66132	0.414323	0.275868	0.256879	0.663599	0.498614	0.790166	0.813069	0.829395
2	22	0.623917	0.689017	1.055969	0.967249	1.188081	0.360704	1.038124	1.057447	0.891281	1.100303
2	23	0.053076	0.771813	0.489834	0.653148	0.786846	0.485048	0.584847	0.633409	1.174143	0.627201
2	24	0.435291	0.591408	0.528147	0.772167	0.42021	0.581063	0.527332	0.555248	1.25568	0.535433
2	26	-0.46347	2.425572	0.242121	0.620174	0.99172	-0.52161	0.717597	0.169036	0.688308	0.240981

Appendix B

R Functions and Codes

```
##############################################################################
##NewRank Function- Determines the shape of the underlying method using indicator function (Hill et al. 1988)
##and applies the appropriate score function to the ranks
##Developed by Afshan Boodhwani @2016
##############################################################################

NewRank=function(Sum,Sequence) {
  Sequence<-as.numeric(Sequence)
  Rank.Sum=rank(Sum)

  library(plyr)

  count=count(Sequence)
  n=1:length(count)

  for(i in 1:length(count)){
    n[i]=count$freq[i]
  }
  N=sum(n)

  #Sratify using Sequence
  Seq1.Sum=1:n[1]
  Seq2.Sum=1:n[2]
  j=1
  k=1

  for(i in 1:N) {
    if(Sequence[i]==1){
```

```
   Seq1.Sum[j]=Sum[i]
   j=j+1
  }
  if(Sequence[i]==2){
   Seq2.Sum[k]=Sum[i]
   k=k+1
  }
}

#Assigning Ranks Based on Sequence
Rank1.Sum=rank(Seq1.Sum)
Rank2.Sum=rank(Seq2.Sum)

#Calculating Q1
cutoff=quantile(Rank1.Sum, c(.95, .75,.5,.25, .05))

j1=0
j2=0
j3=0
j4=0
j5=0
U.05=0
L.05=0
U.5=0
L.5=0
M.5=0

#This will take care of duplicates
```

```
Sort.Rank=sort(Rank1.Sum, decreasing=TRUE)
Sort.Sum=sort(Seq1.Sum, decreasing=TRUE)

fifth.n1=ceiling(n[1]*.05)
fifty.n1= ceiling(n[1]*.5)

#Intermediate variables
for(i in 1:n[1]){
 if(Sort.Rank[i]>=cutoff[1] & j1 < fifth.n1){
  U.05=(U.05+Sort.Sum[i])
  j1=j1+1
 }

 if(Sort.Rank[i]<=cutoff[5] & j2 < fifth.n1){
  L.05=(L.05+Sort.Sum[i])
  j2=j2+1
 }

 if(Sort.Rank[i]>=cutoff[3] & j3 < fifty.n1){
  U.5=(U.5+Sort.Sum[i])
  j3=j3+1
 }

 if(Sort.Rank[i]<=cutoff[3] & j4 < fifty.n1){
  L.5=(L.5+Sort.Sum[i])
  j4=j4+1
 }
```

```
  if(cutoff[4]<Sort.Rank[i]& Sort.Rank[i]<=cutoff[2] & j5 < fifty.n1){
    M.5=(M.5+Sort.Sum[i])
    j5=j5+1
  }
}

U.05=U.05/j1
L.05=L.05/j2
U.5=U.5/j3
L.5=L.5/j4
M.5=M.5/j5

Q11=(U.05-M.5)/(M.5-L.05)
Q21=(U.05-L.05)/(U.5-L.5)

#Working with Seq2
#Calculating Q12
cutoff=quantile(Rank2.Sum, c(.95, .75,.5,.25, .05))

j1=0
j2=0
j3=0
j4=0
j5=0
U.05=0
L.05=0
U.5=0
L.5=0
```

```
M.5=0

#This will take care of duplicates
Sort.Rank=sort(Rank2.Sum, decreasing=TRUE)
Sort.Sum=sort(Seq2.Sum, decreasing=TRUE)

fifth.n2=ceiling(n[2]*.05)
fifty.n2= ceiling(n[2]*.5)

#Intermediate variables
for(i in 1:n[2]){
 if(Sort.Rank[i]>=cutoff[1] & j1 < fifth.n2){
  U.05=(U.05+Sort.Sum[i])
  j1=j1+1
 }

 if(Sort.Rank[i]<=cutoff[5] & j2 < fifth.n2){
  L.05=(L.05+Sort.Sum[i])
  j2=j2+1
 }

 if(Sort.Rank[i]>=cutoff[3] & j3 < fifty.n2){
  U.5=(U.5+Sort.Sum[i])
  j3=j3+1
 }

 if(Sort.Rank[i]<=cutoff[3] & j4 < fifty.n2){
  L.5=(L.5+Sort.Sum[i])
```

```
   j4=j4+1
  }

  if(cutoff[4]<=Sort.Rank[i] & Sort.Rank[i]<=cutoff[2] & j5 < fifty.n2){
    M.5=(M.5+Sort.Sum[i])
    j5=j5+1
  }
}

U.05=U.05/j1
L.05=L.05/j2
U.5=U.5/j3
L.5=L.5/j4
M.5=M.5/j5

Q12=(U.05-M.5)/(M.5-L.05)
Q22=(U.05-L.05)/(U.5-L.5)

###########Weighted Measures
Q1bar=((n[1]*Q11)+(n[2]*Q12))/N
Q1bar
Q2bar=((n[1]*Q21)+(n[2]*Q22))/N
Q2bar
#############Score Selection
Score="Unknown"
Desp="Unknown"
U=Rank.Sum/(N+1)
J.U=U
```

```
if(Q2bar > 3.8){
  Score="W"
  Desp="Heavy Tailed"
  J.U=U
} else if (.5 <= Q1bar & Q1bar <= 2 & 2.24 <= Q2bar & Q2bar<= 3.8){
  Score="W"
  Desp="Moderate Tailed"
  J.U=U
} else if (.5 <= Q1bar & Q1bar <= 2 & Q2bar < 2.24) {
  Score="ML"
  Desp="Light Tailed"
  for(i in 1:length(J.U)){
    if(0<U[i] & U[i]<.25){
      J.U[i]=-1*((U[i]-.25)^2)
    }  else if (.25<=U[i] & U[i]<=.75){
      J.U[i]=0
    }else if (.75 < U[i] & U[i] < 1){
      J.U[i] = (U[i]-.75)^2}
  }
} else if (Q1bar < .5 & Q2bar < 3.8){
  Score="SL"
  Desp="Left Skewed"
  for(i in 1:length(J.U)){
    if(0<U[i] & U[i]<= .5){
      J.U[i]=0
    }  else if (.5<U[i] & U[i]<1){
      J.U[i]=((U[i]-.5))}
```

```
    }
  } else if (Q1bar > 2 & Q2bar < 3.8){
    Score="SR"
    Desp="Right Skewed"
    for(i in 1:length(J.U)){
      if(0<U[i] & U[i]<= .5){
        J.U[i]=((U[i]-.5))
      } else if (.5<U[i] & U[i]<1){
        J.U[i]=0}
    }
  } else {
    Score="Error"
    Desp="Error"
    U=0
    J.U=0
  }

  output<-list(J.U,
       Desp,c(Q11,Q12,Q1bar), c(Q21,Q22,Q2bar))
names(output)<-c("Apadtive Ranks","Underlying Distribution","Q1","Q2")
names(output$Q1)<-c("Q11","Q12","Q1bar")
names(output$Q2)<-c("Q21","Q22","Q2bar")
output
  }
```

```
# R code for ADAPTIVE F1_LD_F1 macro
#
# Input:
#                 y: a vector of variable of interest
#                 group: a vector of group variable (factor level)
#                 time : a vector of time variable
#                 subject : a vector of independent subjects
#
# Optional Input:
#                 w.pat: pattern matrix of order group level x time level
#                 w.t : vector of order time level, pattern for interaction
#                 w.g : vector of order group level, group pattern
#                 time.name: name of the time vector. "Time" is set as default.
#                 group.name: name of the time vector. "Group" is set as default.
#           description: description of the output. Default is set to TRUE (show description)
#                 time.order: a vector of time levels specifying the order.
#                 group.order: a vector of group levels specifying the order.
#
f1.ld.f1.adap<- function(y, time, group, subject, w.pat=NULL, w.t=NULL, w.g=NULL,
time.name="Time", group.name="Group",
                 description=TRUE, time.order=NULL,
group.order=NULL,plot.RTE=TRUE,show.covariance=FALSE, order.warning=TRUE)
{
  #       For model description see Brunner et al. (2002)
  #       f1.ld.f1.adap Author: Afshan Boodhwani
  #                 Department of Mathematics, University of Texas at Arlington, Texas, USA
  #
  #       F1-LD-F1 Author: Mahbub Latif (mlatif@gwdg.de)
  #                 Department of Medical Statistics, Goettingen, Germany
```

```
#
#       Version:  01-01
#       Date: February 18, 2003
#
#      Editied by: Kimihiro Noguchi
#       Version:  01-02
#       Date: August 18, 2009
#
#      Editied by: Kimihiro Noguchi
#       Version:  01-03
#       Date: December 24, 2009
#
#   Key Variables:
#             time: time factor
#             t: number of levels of time
#             a: number of levels of group
#             N: total number of observations
#             ind: indicator of whether there exists a missing observation (0=Yes,1=No)
#             N.na: total number of missing observations
#             subject: total number of subject
#             rscore: ranks of the variable of interest
#             rankmean: mean rank for each level of time
#             Nobs: total number of observations for each level of time
#             RTE: relative treatment effects

#   check whether the input variables are entered correctly

var<-y
```

```
  if(is.null(var)||is.null(time)||is.null(group)||is.null(subject))
    stop("At least one of the input parameters (y, time, group, or subject) is not found.")

  sublen<-length(subject)
  varlen<-length(var)
  timlen<-length(time)
  grolen<-length(group)

  if((sublen!=varlen)||(sublen!=timlen)||(sublen!=grolen))
    stop("At least one of the input parameters (y, time, group, or subject) has a different length.")


##############################################################################
#######
  # The following are the helper functions for the main function
  # List of functions:
  # rte: outputs the relative treatment effect
  # case2x2: outputs statistics for 2 x 2 design
  # wald.test: outputs Wald-type test statistics
  # ANOVA.test: outputs ANOVA-type test statistics
  # Simple.time.test: outputs test statistics for time effect
  # pair.comp.test: outputs test statistics for paired comparison test statistics
  # pattern.group: outputs test statistics for patterned alternatives for group effects
  # df.p: calculates degrees of freedom for patterned alternatives
  # one: changes matrix to a vector
  # I: creates an identity matrix
  # J: creates a unit matrix
  # count.subj: counts the number of subjects in each level of group
  # vi: calculates the variance equation (8.18)
```

```
# V: calculates the block diagonal covariance matrix
# mean.factor: calculates the mean of each factor
# df: calculates the degrees of freedom
# tr: calculates the trace of the matrix

##############################################################################
# rte function for relative treatment effects
rte <- function(group, time, indx, rscore)
{
  a <- nlevels(group);
  t <- nlevels(time);
  tab <- t(matrix(mean.factor(rscore, group:time, indx), t, a))
  rankmean.g <- as.vector(apply(tab, 1, mean));
  rankmean.y <- as.vector(apply(tab, 2, mean));
  rankmean.gy <- mean.factor(rscore,group:time, indx)
  RankMean <- c(rankmean.g, rankmean.y, rankmean.gy);

  # no of observations per factors
  Nobs <- c(tapply(indx, group, sum), tapply(indx, time, sum), tapply(indx, group:time, sum))

  # rte
  RTE <- (1/sum(indx))*(RankMean - median(rscore))

  # output

  out <- data.frame(RankMeans=RankMean, Nobs=Nobs, RTE=RTE)
  levels(group) <- paste(group.name,glevel,sep="")
  levels(time) <- paste(time.name,tlevel,sep="")
  row.names(out)<-c(levels(group), levels(time), levels(group:time))
```

```
  return(out)
}

## Case 2 x 2 as described in 8.1.2.
case2x2 <- function(group, time, subj, rscore, ind)
{
  rscore <- rscore*ind
  rscore.s <- split(rscore, group)
  subj.s <- split(subj, group)
  ind.s <- split(ind, group)
  sigma2 <- rep(0,2)
  Un <- 0
  Un.const <- 0
  v.den <- 0
  UnT <- 0
  UnT.c <- 0
  vT <- 0
  UnAT <- 0

  for(i in 1:2)
  {
    junk <- t(sapply(split(as.vector(rscore.s[[i]]), as.vector(subj.s[[i]])), matrix))
    junk.i <- t(sapply(split(as.vector(ind.s[[i]]), as.vector(subj.s[[i]])), matrix))
    junk<-as.matrix(junk)
    junk.i<-as.matrix(junk.i)
    junk.m <- apply(junk,2,sum)/apply(junk.i,2,sum)
    junk.s <- apply(junk,1,sum)
```

```
  # Group effect
  Un <- Un + sum(junk.m)*(-1)^(i+1)
  sigma2[i] <- sum((junk.s - sum(junk.m))^2)/(nrow(junk)-1)
  Un.const <- Un.const + sigma2[i]/nrow(junk)
  v.den <- v.den+(sigma2[i]/nrow(junk))^2/(nrow(junk)-1)

  # Time
  junk.d <- apply(junk,1,diff)
  tau2 <- sum((junk.d - diff(junk.m))^2)/(nrow(junk)-1)
  UnT <- UnT - diff(junk.m)
  UnT.c <- UnT.c + tau2/nrow(junk)
  vT <- vT + (tau2/nrow(junk))^2/(nrow(junk)-1)

  # Interaction
  UnAT <- UnAT + diff(junk.m)*(-1)^i
}

# Group effect
Un <- Un/sqrt(Un.const)
v <- Un.const^2/v.den
if(!is.na(Un)&&(v > 0))
{
  pGN <- (pnorm(abs(Un),lower.tail=FALSE))*2
  pGT <- (pt(abs(Un), v,lower.tail=FALSE))*2
}
else
{
  pGN <- NA
```

```
  pGT <- NA
}
out <- data.frame(Statistics=Un, NN=pGN, DF=v, tt=pGT)

# Time effect
UnT <- UnT/sqrt(UnT.c)
vT <- UnT.c^2/vT
if(!is.na(UnT)&&(vT > 0))
{
  pTN <- (pnorm(abs(UnT),lower.tail=FALSE))*2
  pTT <- (pt(abs(UnT),vT,lower.tail=FALSE))*2
}
else
{
  pTN <- NA
  pTT <- NA
}
out <- rbind(out, c(UnT,pTN,vT,pTT))

# Interaction
UnAT <- UnAT/sqrt(UnT.c)
if(!is.na(UnAT)&&(vT > 0))
{
  pATN <- (pnorm(abs(UnAT),lower.tail=FALSE))*2
  pATT <- (pt(abs(UnAT),vT,lower.tail=FALSE))*2
}
else
{
```

```
    pATN <- NA
    pATT <- NA
  }
  out<- rbind(out, c(UnAT,pATN,vT,pATT))
  names(out) <- c("Statistic","p-value(N)","df","p-value(T)")
  row.names(out) <- c(group.name, time.name, paste(group.name,":",time.name,sep=""))
  return(list(case2x2=out))
}

# Wald test to test average group effect, average time effect, and global interaction effect
wald.test <- function(group, time, subject, rscore, ind, ni)
{
  n <- sum(ni);
  N <- sum(ind)
  a <- nlevels(group)
  t <- nlevels(time)
  V <- V(group, time, subject, rscore, ind, a, t, ni)$V
  R <- V(group, time, subject, rscore, ind, a, t, ni)$R

  # unconditional time mean
  tab <- t(matrix(mean.factor(rscore, group:time, ind), t, a))
  t.mean <- as.vector(apply(tab, 2, mean));

  # Average group effect
  p <- (R - 0.5)/N; # RTE group x time
  Pa <- I(a) - (1/a) * J(a)
  Pt <- I(t) - (1/t) * J(t)
  Pat <- kronecker(Pa, Pt)
```

```
# second last equation of page 134
cpg <- sqrt(n) * Pa %*% kronecker(I(a), (1/t)*t(one(t))) %*% p;

# last equation of page 134
Sigma <- kronecker(Pa,  (1/t)*t(one(t))) %*% V %*%  kronecker(Pa, (1/t)*one(t));

# equation (8.10)
cvc <- Pa %*% Sigma %*% Pa
Q.a <- t(cpg) %*% ginv(cvc) %*% cpg;

df.a <- tr(cvc%*%ginv(cvc))
if(!is.na(Q.a) && (Q.a > 0)) pval.a <- round(pchisq(Q.a, df.a,lower.tail=FALSE),Inf)
else pval.a <- NA;
A <- c(W=Q.a, df=df.a, pval=pval.a);

# Average time effect, eqn (8.9)
S <- kronecker((1/a)*t(one(a)), I(t)) %*% V %*% kronecker((1/a)*one(a), I(t))

cpt <- Pt %*% t.mean;

# equation (8.20)
cvc <- Pt %*% S %*% Pt
Q.t <- (n/N^2)*t(cpt) %*% ginv(cvc) %*% cpt;
df.t <- tr(cvc%*%ginv(cvc))
if(!is.na(Q.t) && (Q.t > 0)) pval.t <- round(pchisq(Q.t, df.t,lower.tail=FALSE),Inf)
else pval.t <- NA;
T <- c(Q.t, df.t, pval.t);
```

```
  # Global interaction effect
  Cat <- kronecker(Pa, Pt); # book notation of page 141

  # equation (8.26)
  cvc <- Cat %*% V %*% t(Cat)
  Q.at <- (n/N^2)*t(Cat %*% R) %*% ginv(cvc) %*% Cat %*% R;
  df.at <- tr(cvc%*%ginv(cvc))
  if(!is.na(Q.at) && (Q.at > 0)) pval.at <- round(pchisq(Q.at, df.at,lower.tail=FALSE), Inf)
  else pval.at <- NA;
  AT <- c(Q.at, df.at, pval.at);

  # results
  out.w <- rbind(A, T, AT);
  colnames(out.w) <- c("Statistic", "df", "p-value")
  rownames(out.w) <- c(group.name,time.name,paste(group.name,":",time.name,sep=""))
  out <- list(Wald.test=out.w);
}

# To test average group effect, average time effect, and global interaction effect
anova.test <- function(group, time, subject, rscore, ind, a, ni)
{
  group <- as.factor(group)
  time <- as.factor(time)
  t <- nlevels(time)

  t.mean <- apply(t(matrix(mean.factor(rscore, group:time, ind), t, a)),2,mean)
```

```
    n <- sum(ni);
    N <- sum(ind)

    V <- V(group, time, subject, rscore, ind, a, t, ni)$V
    R <- V(group, time, subject, rscore, ind, a, t, ni)$R

    # Average group effect
    p <- (R - 0.5)/N; # RTE corresponding group x time
    Pa <- I(a) - (1/a) * J(a); # centering matrix
    Pt <- I(t) - (1/t) * J(t); # centering matrix
    Pat <- kronecker(Pa, Pt);

    # second last equation of page 134
    cpg <- sqrt(n) * Pa %*% kronecker(I(a), (1/t)*t(one(t))) %*% p;
    Sigma <- kronecker(Pa,  (1/t)*t(one(t))) %*% V %*%  kronecker(Pa, (1/t)*one(t));

    # average group effect
    # equation (8.11)
    F.a <- t(cpg) %*% cpg/sum(diag(Pa %*% Sigma));
    # equation (5.7)
    df1.a <- sum(diag(Pa %*% Sigma))^2/sum(diag((Pa %*% Sigma) %*% (Pa %*% Sigma)));
    if((!is.na(F.a))&&(!is.na(df1.a))&&(F.a > 0)&&(df1.a > 0)) pval.a<-
pchisq(F.a*df1.a,df1.a,lower.tail=FALSE)
    else pval.a<-NA;
    A <- round(c(B=F.a, df=df1.a, pval=pval.a), Inf);

    # average time effect
    # equation (8.9)
    S <- kronecker((1/a)*t(one(a)), I(t)) %*% V %*% kronecker((1/a)*one(a), I(t))
```

```
    cpt <- Pt %*% t.mean;


    # equation (8.21)
    F.t <- (n/N^2) * (t(cpt) %*% cpt)/sum(diag(Pt %*% S));
    df1.t <- sum(diag(Pt %*% S))^2/sum(diag(Pt %*% S %*% Pt %*% S));
    if((!is.na(F.t))&&(!is.na(df1.t))&&(F.t > 0)&&(df1.t > 0)) pval.t<-
pchisq(F.t*df1.t,df1.t,lower.tail=FALSE)
    else pval.t<-NA;
    T <- round(c(F.t, df1.t, pval.t),Inf);


    # Global interaction effect
    F.at <- n * t(p) %*% Pat %*% p/sum(diag(Pat %*% V));
    df1.at <- sum(diag(Pat %*% V))^2/sum(diag(Pat %*% V %*% Pat %*% V));
    if((!is.na(F.at))&&(!is.na(df1.at))&&(F.at > 0)&&(df1.at > 0)) pval.at<-
pchisq(F.at*df1.at,df1.at,lower.tail=FALSE)
    else pval.at<-NA;
    AT <- round(c(F.at, df1.at, pval.at), Inf);
    out.box <- rbind(A, T, AT);
    colnames(out.box) <- c("Statistic", "df", "p-value")
    rownames(out.box) <- c(group.name,time.name,paste(group.name,":",time.name,sep=""))
    # modified Box-approximation
    df1 <- df(V, a, t, ni, ind)$df1
    df2 <- df(V, a, t, ni, ind)$df2
    if((!is.na(F.a)) && (!is.na(df1)) && (!is.na(df2)) && (F.a > 0) && (df1 > 0) && (df2 > 0))
pval.mb <- pf(F.a, df1, df2,lower.tail=FALSE)
    else pval.mb <- NA
    A <- rbind(round(c(B=F.a, df1=df1, df2=df2, pval=pval.mb), Inf));
    colnames(A) <- c("Statistic", "df1", "df2","p-value")
```

```
  rownames(A) <- c(group.name)
  out <- list(ANOVA.test=out.box, ANOVA.test.mod.Box=A);

  return(out);
}

# Simple time test to test time effect
simple.time.test <- function(name.group, a , t, ni, N, pat.dat, V1, R1)
{
  # simple time effect
  Pt <- I(t) - (1/t) * J(t);
  wald <- as.data.frame(matrix(0, a, 3));
  rownames(wald)<-name.group
  anova <- as.data.frame(matrix(0, a, 3));
  rownames(anova)<-name.group
  normal <- as.data.frame(matrix(0, a, 4));
  rownames(normal)<-name.group
  names(wald) <- c("Statistic", "df", "p-value");
  names(anova) <- c("Statistic", "df", "p-value");
  names(normal) <- c("Statistic", "p-value(N)", "df", "p-value(T)");
  n <- sum(ni); k <-1;

  for(i in 1:a)
  {
    V <- V1[k:(i*t), k:(i*t)];
    R <- R1[k:(i*t),1];
    k <- i*t + 1;
    cp <- Pt %*% R;
```

```
    Q.s <- round((n/N^2) * t(cp) %*% ginv(Pt %*% V %*% Pt) %*% cp, Inf);

    F.s <- round((n/N^2) * t(cp) %*% cp/sum(diag(Pt %*% V)), Inf);

    df <- tr((Pt %*% V %*% Pt)%*%ginv(Pt %*% V %*% Pt))

    df1 <- round(sum(diag(Pt %*% V))^2/sum(diag(Pt %*% V %*% Pt %*% V)), Inf);

    if((!is.na(Q.s)) && (!is.na(df)) && (Q.s > 0) && (df > 0)) pval <- round(pchisq(Q.s,
df,lower.tail=FALSE), Inf)

    else pval <- NA;

    if((!is.na(F.s)) && (!is.na(df1)) && (F.s > 0) && (df1 > 0)) pval1<-
round(pchisq(F.s*df1,df1,lower.tail=FALSE), Inf)

    else pval1 <- NA

    out <- c(Q.s, df, pval);

    out1 <- c(F.s, df1, pval1);

    wald[i,] <- out;

    anova[i,] <- out1;


    # pattern effect

    if(!is.null(pat.dat))

    {

      pi <- (R - 0.05)/N;

      s2 <- (ni[i]/n) * as.numeric(pat.dat[i,]) %*% Pt %*% V %*% Pt %*%
as.numeric(pat.dat[i,]);

      L <- round(sqrt(ni[i]/s2) * as.numeric(pat.dat[i,]) %*% Pt %*% pi, Inf);

      p.nor <- round((pnorm(L,lower.tail=FALSE)), Inf);

      df1 <- ni[i] -1;

      pval <- round((pt(L, df1,lower.tail=FALSE)), Inf);

      normal[i,] <- c(L, p.nor, df1, pval);

    }

  }
```

```
  sim.time.effect <- list(Wald.test.time=wald, ANOVA.test.time=anova);
  if(!is.null(pat.dat)) pat.time.effect <- list(pattern.time=normal)
  else pat.time.effect <- list(pattern.time=NULL)

  return(c(sim.time.effect, pat.time.effect));
}

# pairwise comparison test statistics
pair.comp.test <- function(data, ni, w, lev.grp)
{
  a <- nlevels(factor(data[,1]));
  t <- nlevels(factor(data[,2]));
  n <- sum(ni)
  N <- sum(data[,5])

  Pt <- I(t) - (1/t)*J(t)
  V11 <- V(data[,1], data[,2], data[,3], data[,4], data[,5], a, t, ni)$V;

  # arranging output
  out <- as.data.frame(matrix(0, 3*choose(a,2), 3));
  out.pat <- as.data.frame(matrix(0, choose(a,2), 4));
  names(out) <- c("Statistic", "df", "p-value");
  names(out.pat) <- c("Statistic", "p-value(N)", "df", "p-value(T)")
  Test <- rep(c(group.name,time.name,paste(group.name,":",time.name,sep="")), choose(a,2));
  Pairs <- rep(0, 3*choose(a ,2));


  k <- 1;
```

```
ll <-1
for(i in 1:(a-1))
{
  for(j in (i+1):a)
  {
    data.p <- data[data[,1]==i | data[,1]==j,]
    ni.p <- matrix(c(ni[i],ni[j]), 2,1);

    nn <- sum(ni.p)
    NN <- sum(data.p[,5]);

    gr <- data.p[,1]
    tm <- data.p[,2]
    subj <- data.p[,3]
    rs <- data.p[,4]
    ind <- data.p[,5]

    V <- V(as.numeric(gr), tm, subj, rs, ind, 2, t, ni.p)$V
    R <- V(as.numeric(gr), tm, subj, rs, ind, 2, t, ni.p)$R

    out[k:(k+2),] <- anova.test(gr, tm, subj, rs, ind, 2, ni.p)$ANOVA.test
    Pairs[k:(k+2)] <- paste(group.name,lev.grp[i], ":",group.name,lev.grp[j],sep="")
    k <- k + 3;

    # pattern interactions
    if(!is.null(w))
    {
      w <- matrix(w, t,1)
```

```
        sign <- t(w)%*%Pt%*%(V[1:t,1:t]+V[(t+1):(2*t),(t+1):(2*t)])%*%Pt%*%w
        out.pat[ll,1] <- sqrt(nn/sign)*t(w-mean(w))%*%(R[1:t,] - R[(t+1):(2*t),])/NN
        out.pat[ll,2] <- pnorm(out.pat[ll,1],lower.tail=FALSE)
        posA <- matrix(0, a, 1);
        posA[i,]<- 1; posA[j,] <- 1;
        CC <- t(w)%*%Pt
        M <- kronecker(diag(c(posA)),CC)
        S <- M%*%V11%*%t(M)
        lambda <- solve(diag(c(ni))-I(a))
        out.pat[ll,3] <- tr(S)^2/tr(S*S*lambda)
        out.pat[ll,4] <- pt(out.pat[ll,1],out.pat[ll,3],lower.tail=FALSE)
        row.names(out.pat)[ll] <- paste(group.name,lev.grp[i],":",group.name,lev.grp[j],sep="");
        ll <- ll + 1
       }
     }
   }

   out <- cbind(Pairs, Test, out);
   if(!is.null(w)) pair.comp<- list(pair.comparison=out,
pattern.pair.comparison=round(out.pat,Inf))
   else pair.comp <- list(pair.comparison=out, pattern.pair.comparison=NULL)
   return(pair.comp);
 }

 # patterned alternatives for group effects
 pattern.group <- function(group, time, subject, rscore, ind, a, t, ni, g.mean, w.g)
 {
   n <- sum(ni)
   N <- sum(ind)
```

```
  Pa <- I(a) - (1/a)*J(a)
  V <- V(group, time, subject, rscore, ind, a, t, ni)$V
  S <- kronecker(Pa, (1/t)*t(one(t)))%*%V%*%kronecker(Pa, (1/t)*one(t))
  w.g <- matrix(w.g, a, 1)
  lambda <- solve(diag(c(ni))-I(a))
  c <- Pa%*%w.g

  g <- (g.mean-.5)/N
  Kn <- sqrt(n)*t(w.g)%*%Pa%*%g
  sign <- t(c)%*%S%*%c

  df<- df.p(V, a, t, ni, ind, w.g)

  Ln <- Kn/sqrt(sign)
  pval.t <- pt(Ln, df,lower.tail=FALSE)
  pval.n <- pnorm(Ln,lower.tail=FALSE)
  out<-rbind(round(c(Ln=Ln, pval.N=pval.n, df=df, pval.t=pval.t),Inf))
  colnames(out) <- c("Statistic", "p-value(N)", "df", "p-value(T)")
  rownames(out) <- c(group.name)
  return(list(pattern.group=out))
}

# degrees of freedom calculatin for patterned alternative
df.p <- function(V, a, t, ni, ind, w)
{
  Pa <- I(a) - (1/a)*J(a)
  lambda <- solve(diag(c(ni)) - I(a))
  c <- (diag(c(t(w)%*%Pa)))^2
```

```
  m <- kronecker(I(a), (1/t)*t(one(t)))
  s <- m%*%V%*%t(m)
  x <- c*s
  df <- (tr(x))^2/tr((x%*%x)*lambda)
  return(df)
}

# one vector
one <- function(d) return(matrix(1, d, 1));

# Identity matrix
I <- function(d)
{
  junk <- rep(1, length=d);
  junk <- diag(junk);
  return(junk);
}

# Unit matrix
J <- function(d1, d2=d1) return(matrix(1,d1,d2));

# count the number of subjects in each level of group
count.subj <- function(group, subject)
{
  group <- as.factor(group)
  table <- table(group, subject);
  n <- matrix(0,nlevels(group), 1);
  for(i in 1:nlevels(group)) n[i] <- length(colnames(table)[table[i,] > 0]);
```

```
  return(n);
}

# variance Vi equation (8.18)
Vi <- function(data)
{
  time <- factor(data$time)
  subj <- factor(data$subj)
  srank <- data$rscore
  indx <- data$indx

  t <- nlevels(time)
  n0 <- nlevels(subj)

  # matrix of order subject x time, R_ik - R_i
  srank.m <- matrix(tapply(srank, subj:time, sum), t, n0)
  ind.m <- matrix(tapply(indx, subj:time, sum), t, n0)

  junk <- srank.m*ind.m;
  t.junk <- t(ind.m);

  R <- apply(junk,1,sum)/apply(ind.m, 1,sum);
  junk1 <- t(junk -R)

  V <- matrix(0, t, t);

  for(i in 1:t)
  {
```

```
    for(j in 1:t)
    {
      ls <- sum(t.junk[,i]);
      if(i==j) V[i,i] <- sum(t.junk[,i]*junk1[,i]^2)/(ls*(ls-1))
      else
      {
        lss <- sum(t.junk[,i]*t.junk[,j]);
        kss <- (ls-1)*(sum(t.junk[,j])-1) + lss - 1;
        V[i,j] <- sum(t.junk[,i]*t.junk[,j]*junk1[,i]*junk1[,j])/kss;
      }
    }
  }

  out <- list(R=R,V=V);
  return(out);
}

# block diagonal covariance matrix
V <- function(group, time, subj, rscore, indx, a, t, ni)
{
  group <- as.factor(group)
  subj <- as.factor(subj)
  time <- as.factor(time)
  data <- data.frame(group, time, subj, rscore, indx)

  n <- sum(ni);
  N <- sum(indx);
```

```
  # split rscore by group
  split.d <- split(data, data$group)

  # calculating covariance matrix for average group, time and global interaction effect
  V <- matrix(0, a*t, a*t);
  R <- matrix(0, a*t, 1);
  i <- 1; k <- 1;
  while(i <= a)
  {
    V[k:(i*t), k:(i*t)] <-  Vi(split.d[[i]])$V;
    R[k:(i*t), 1] <- Vi(split.d[[i]])$R;
    k <- i*t + 1;
    i <- i+1;
  }
  V <- V*n/N^2;
  out <- list(V=V, R=R);
}

# mean of each factor
mean.factor <- function(y, fact, indx)
{
  tab <- tapply(y, fact, sum)
  ind <- tapply(indx, fact, sum)
  return(c(tab/ind))
}

# degrees of freedom calculation
df <- function(V, a, t, ni, ind)
```

```
{
  Pa <- I(a) - (1/a)*J(a)
  Pt <- I(t) - (1/t)*J(t)
  c <- kronecker(Pa, (1/t)*t(one(t)))
  tt <- t(c)%*%ginv(c%*%t(c))%*%c
  tem1 <- tt%*%V;
  df1 <- (tr(tem1))^2/tr(tem1%*%tem1)
  dpr <- Pa*I(a)
  mat <- kronecker(I(a), (1/t)*t(one(t)))
  va <- mat%*%V%*%t(mat)
  lambda <- solve(diag(c(ni)) - I(a))
  tem1 <- (tr(dpr%*%va))^2
  tem2 <- tr(dpr%*%dpr%*%va%*%va%*%lambda)
  df2 <- tem1/tem2
  return(list(df1=df1, df2=df2))
}

# trace calculation
tr <- function(x) return(sum(diag(x)))

# end of helper functions

##############################################################################
# main function

library(MASS)
glevel <- unique(group)
tlevel <- unique(time)
slevel <- unique(subject)
```

```
 t <- length(tlevel)
 s <- length(slevel)
 a <- length(glevel)

 if((t*s)!=length(var))
  stop("Number of levels of subject (",s, ") times number of levels of time (",t,")
       is not equal to the total number of observations (",length(var),").",sep="")

 #    time order vector

 if(!is.null(time.order))
 {
  tlevel <- time.order
  tlevel2 <- unique(time)

  if(length(tlevel)!=length(tlevel2))  # if the levels of the order is different from the one in the
data
    stop("Length of the time.order vector (",length(tlevel), ")
         is not equal to the levels of time vector (",length(tlevel2),").",sep="")

  if(mean(sort(tlevel)==sort(tlevel2))!=1)     # if the elements in the time.order is different from
the time levels
    stop("Elements in the time.order vector is different from the levels specified in the time
vector.",sep="")
 }

 #    group order vector

 if(!is.null(group.order))
```

```
 {
   glevel <- group.order
   glevel2 <- unique(group)

   if(length(glevel)!=length(glevel2))# if the levels of the order is different from the one in the
data
     stop("Length of the group.order vector (",length(glevel), ")
         is not equal to the levels of group vector (",length(glevel2),").",sep="")

   if(mean(sort(glevel)==sort(glevel2))!=1)     # if the elements in the group.order is different
from the group levels
     stop("Elements in the group.order vector is different from the levels specified in the group
vector.",sep="")
 }

 #    sort data

 sortvector<-double(length(var))
 newtime<-double(length(var))
 newsubject<-double(length(var))
 newgroup<-double(length(var))

 for(i in 1:length(var))
 {
   row<-which(subject[i]==slevel)
   col<-which(time[i]==tlevel)
   newsubject[i]<-row
   newtime[i]<-col
   newgroup[i]<-which(group[i]==glevel)
```

```
  sortvector[((col-1)*s+row)]<-i
}

subject<-newsubject[sortvector]
var<-var[sortvector]
time<-newtime[sortvector]
group<-newgroup[sortvector]

#    sort again by group, and assign new subject numbers to subjects

grouptemp<-order(group[1:s])
groupplus<-(rep(c(0:(t-1)),e=s))*s
groupsort<-(rep(grouptemp,t))+groupplus

subject<-rep(c(1:s),t)
var<-var[groupsort]
time<-time[groupsort]
group<-group[groupsort]

#    organize data

group<-factor(group)
time<-factor(time)
subject<-factor(subject)

score <- var
N.na <- sum(is.na(score))
ind <- 1 - is.na(score)
```

```
  N <- sum(ind)
  rscore <- NewRank(score,group)$`Apadtive Ranks`*ind

  data <- cbind(group, time, subject, rscore, ind)
  ni <- count.subj(group, subject)
  n <- sum(ni); # number of subjects in the experiment
  model.name<-"F1 LD F1 Adaptive Model"

  if(description==TRUE)
  {
    cat(" Total number of observations: ",sum(ind),"\n")
    cat(" Total number of subjects:  " , n,"\n")
    cat(" Total number of missing observations: ",N.na,"\n")
    cat("\n Class level information ")
    cat("\n ---------------------- \n")
    cat(" Levels of", time.name, "(sub-plot factor time) : ", t,"\n")
    cat(" Levels of", group.name, "(whole-plot factor group) : ", a,"\n")
    cat("\n Abbreviations ")
    cat("\n ---------------------- \n")
    cat(" RankMeans = Rank means\n")
    cat(" Nobs = Number of observations\n")
    cat(" RTE = Relative treatment effect\n")
    cat(" case2x2 = tests for 2-by-2 design\n")
    cat(" Wald.test = Wald-type test statistic\n")
    cat(" ANOVA.test = ANOVA-type test statistic with Box approximation\n")
    cat(" ANOVA.test.mod.Box = modified ANOVA-type test statistic with Box
approximation\n")
    cat(" Wald.test.time = Wald-type test statistic for simple time effect\n")
    cat(" ANOVA.test.time = ANOVA-type test statistic for simple time effect\n")
```

```
  cat(" N = Standard Normal Distribution N(0,1)\n")
  cat(" T = Student's T distribution with respective degrees of freedom\n")
  if(!is.null(w.pat))
  {
   pattern.string<-c(w.pat)
  }
  else
  {
   pattern.string<-"no pattern specified"
  }
  if(!is.null(w.t))
  {
   pattern.string.t<-w.t
  }
  else
  {
   pattern.string.t<-"no pattern specified"
  }
  if(!is.null(w.g))
  {
   pattern.string.g<-w.g
  }
  else
  {
   pattern.string.g<-"no pattern specified"
  }


  cat(" pattern.time (time effects) = Test against patterned alternatives in time using normal
distribution (",pattern.string,")","\n")
```

```
   cat(" pair.comparison = Tests for pairwise comparisions (without specifying a pattern)","\n")

   cat(" pattern.pair.comparison = Test for pairwise comparisons with patterned alternatives in time (",pattern.string.t,")","\n")

   cat(" pattern.group (group effects) = Test against patterned alternatives in group (",pattern.string.g,")","\n")

   cat(" covariance = Covariance matrix","\n")

   cat(" Note: The description output above will disappear by setting description=FALSE in the input. See the help file for details.","\n\n")

  }


  if(order.warning==TRUE)

  {

   cat(" F1 LD F1 Adaptive Model ")

   cat("\n ---------------------- \n")

   cat(" Check that the order of the time and group levels are correct.\n")

   cat(" Time level:  " , paste(tlevel),"\n")

   cat(" Group level:  " , paste(glevel),"\n")

   cat(" If the order is not correct, specify the correct order in time.order or group.order.\n\n")

  }


  # unconditional group and time means

  tab <- t(matrix(mean.factor(rscore, group:time, ind), t, a))

  g.mean <- as.vector(apply(tab, 1, mean));

  t.mean <- as.vector(apply(tab, 2, mean));


  # covariance matrix

  V2 <- V(group, time, subject, rscore, ind, a, t, ni)$V;

  R <- V(group, time, subject, rscore, ind, a, t, ni)$R;
```

```
SING.COV <- FALSE
if(qr(V2)$rank < (t*a)) SING.COV <- TRUE
if(SING.COV)
{
  cat("\n Warning(s):\n")
  cat(" The covariance matrix is singular. \n")
}



sdat <- NULL
rte <- list(RTE=rte(group, time, ind, rscore))
rte.plot<-data.frame(rte)
namen.plot<-rownames(rte.plot)[(a+1):(a+t)]
namen.plot.g<-rownames(rte.plot)[1:a]
#rte <- data.frame(rte(group, time, ind, rscore))
### case2x2 is available only when there is no missing observation in the 2-by-2 design.
### otherwise, it returns NULL.
if(a==2 && t==2 && N.na==0)
{
  out2 <- case2x2(group, time, subject, rscore, ind)
}
else
{
  out2 <- list(case2x2=NULL)
}
wald.test.t <- wald.test(group, time, subject, rscore, ind, ni);
anova.test.t <- anova.test(group, time, subject, rscore, ind, a, ni);
out2 <- c(out2, wald.test.t, anova.test.t)
```

```
  simple.time.test.t <- simple.time.test(glevel, a, t, ni, N, w.pat, V2, R);
  pair.comp.t <- pair.comp.test(data, ni, w.t, glevel)
  out2 <- c(out2, simple.time.test.t,pair.comp.t)


  if(!is.null(w.g)) pattern.g <- pattern.group(group, time, subject, rscore, ind, a, t, ni, g.mean, w.g)
  else pattern.g <- NULL
  if (show.covariance == FALSE) {
    V2 <- NULL}
  old.rte <- data.frame(f1.ld.f1(y, time, group, subject, description=FALSE, plot.RTE=FALSE,
show.covariance=FALSE, order.warning=FALSE)$RTE)
    out <- c(sdat, old.rte, out2, pattern.g, list(covariance=V2), model.name=model.name)
  if (plot.RTE == TRUE) {
trad=f1.ld.f1(y, time, group, subject, description=FALSE, plot.RTE=TRUE,
show.covariance=FALSE, order.warning=FALSE)
  }


  return(out)
}
```

```
#####################################################################
## Johnson and Grender test statistic for 4 repeated measures
#####################################################################
JG.4<-function(R.iks,Sequence){
i=unique(Sequence)
 library("stringi")
s=nrow(R.iks) #repeated measures
n1=sum(stri_count_fixed(Sequence, 1))
n2=sum(stri_count_fixed(Sequence, 2))
n=c(n1,n2)

Rbar.i.s=cbind(c(sum(c(R.iks[1,1:n1]/n1)),sum(c(R.iks[2,1:n1]/n1)),sum(c(R.iks[3,1:n1]/n1)),su
m(c(R.iks[4,1:n1]/n1))),

c(sum(c(R.iks[1,(n1+1):(n1+n2)]/n2)),sum(c(R.iks[2,(n1+1):(n1+n2)]/n2)),sum(c(R.iks[3,(n1+1)
:(n1+n2)]/n2)),sum(c(R.iks[4,(n1+1):(n1+n2)]/n2))))

m=c(median(R.iks[1,1:(n1+n2)]),median(R.iks[2,1:(n1+n2)]),median(R.iks[3,1:(n1+n2)]),media
n(R.iks[4,1:(n1+n2)]))

##Matrix U
U=matrix(1,nrow=s,ncol=length(i))

for(k in (1:length(i))){
  for(j in (1:s)){
    U[j,k]=Rbar.i.s[j,k]-m[j]
  }
}

##Covariance Matrix
```

```
library(data.table)
V=(R.iks-m) %*% t(R.iks-m)
V.inv=solve(V)    #Matrix Inverse



#W Statistic
Mult=i
for(k in (1:length(i))){
  Mult[k]=t(U[,k])%*%V.inv%*%U[,k]
}

J=0
for(k in (1:length(Mult))){
  J=J+(n[k]*Mult[k])
}
W=(sum(n)-1)*J

##P-value using Chi-squared Distribution with 2 degrees of freedom (johnson and Grender)
output<-list(W, 1-pchisq(W,2))
names(output)<-c("Test statistic","p-value")
 output
}
```

```
#####################################################################
## Johnson and Grender test statistic for 5 repeated measures
#####################################################################
JG.5<-function(R.iks,Sequence){
  i=unique(Sequence)
  library("stringi")
  s=nrow(R.iks) #repeated measures
  n1=sum(stri_count_fixed(Sequence, 1))
  n2=sum(stri_count_fixed(Sequence, 2))
  n=c(n1,n2)

Rbar.i.s=cbind(c(sum(c(R.iks[1,1:n1]/n1)),sum(c(R.iks[2,1:n1]/n1)),sum(c(R.iks[3,1:n1]/n1)),su
m(c(R.iks[4,1:n1]/n1)),sum(c(R.iks[5,1:n1]/n1))),

c(sum(c(R.iks[1,(n1+1):(n1+n2)]/n2)),sum(c(R.iks[2,(n1+1):(n1+n2)]/n2)),sum(c(R.iks[3,(n1+1)
:(n1+n2)]/n2)),sum(c(R.iks[4,(n1+1):(n1+n2)]/n2)),sum(c(R.iks[5,(n1+1):(n1+n2)]/n2))))

m=c(median(R.iks[1,1:(n1+n2)]),median(R.iks[2,1:(n1+n2)]),median(R.iks[3,1:(n1+n2)]),media
n(R.iks[4,1:(n1+n2)]),median(R.iks[5,1:(n1+n2)]))

  ##Matrix U
  U=matrix(1,nrow=s,ncol=length(i))

  for(k in (1:length(i))){
    for(j in (1:s)){
      U[j,k]=Rbar.i.s[j,k]-m[j]
    }
  }

  ##Covariance Matrix
```

```
 library(data.table)
 V=(R.iks-m) %*% t(R.iks-m)
 V.inv=solve(V)    #Matrix Inverse

 #W Statistic
 Mult=i
 for(k in (1:length(i))){
   Mult[k]=t(U[,k])%*%V.inv%*%U[,k]
 }

 J=0
 for(k in (1:length(Mult))){
   J=J+(n[k]*Mult[k])
 }
 W=(sum(n)-1)*J

 ##P-value using Chi-squared Distribution with 2 degrees of freedom (johnson and Grender)
 output<-list(W, (1-pchisq(W,2)))
 names(output)<-c("Test statistic","p-value")
 output
}
```

```
#####################################################################
##Collagen BIT Example
#####################################################################
mydata=read.table("C:/Eye Example.csv",header=TRUE,sep=",")
attach(mydata)

#F1-Ld-F1 Measure

Measure.Carry <- c(R1+R2,L1+L2)
Measure.Direct <- c(R1-R2,L1-L2)

sequence <- c(Sequence,Sequence)
subject=c(Sub,Sub)

## F1-LD-F1 Test [rank-based, adaptive]
f1.ld.f1(Measure.f1ldf1,c(rep(1,12),rep(2,12)),sequence,subject, description = FALSE)
f1.ld.f1.adap(Measure.f1ldf1,c(rep(1,12),rep(2,12)),sequence,subject, description = FALSE)

##J&G Method
Seq=as.numeric(Sequence)

Sum.R=R1+R2
Sum.L=L1+L2

Diff.R=R1-R2
Diff.L=L1-L2
```

```
h01.R<-NewRank(Sum.R,Seq)$'Apadtive Ranks'
h01.L<-NewRank(Sum.L,Seq)$'Apadtive Ranks'
h02.R<-NewRank(Diff.R,Seq)$'Apadtive Ranks'
h02.L<-NewRank(Diff.L,Seq)$'Apadtive Ranks'

## Use the correct scores based on the hypothesis being tested
R.iks=rbind(h02.R,h02.L)

Sequence=Seq
i=unique(Sequence)
library("stringi")
s=nrow(R.iks) #repeated measures
n1=sum(stri_count_fixed(Sequence, i[1]))
n2=sum(stri_count_fixed(Sequence, i[2]))
n=c(n1,n2)

Rbar.i.s=cbind(c(sum(c(R.iks[1,1:n1]/n1)),sum(c(R.iks[2,1:n1]/n1))),
            c(sum(c(R.iks[1,(n1+1):(n1+n2)]/n2)),sum(c(R.iks[2,(n1+1):(n1+n2)]/n2))))

m=c(median(R.iks[1,1:(n1+n2)]),median(R.iks[2,1:(n1+n2)]))

##Matrix U
U=matrix(1,nrow=s,ncol=length(i))

for(k in (1:length(i))){
  for(j in (1:s)){
    U[j,k]=Rbar.i.s[j,k]-m[j]
  }
```

```
}

##Covariance Matrix
library(data.table)
V=(R.iks-m) %*% t(R.iks-m)
V.inv=solve(V)    #Matrix Inverse



#W Statistic
Mult=1:length(i)
for(k in (1:length(i))){
  Mult[k]=t(U[,k])%*%V.inv%*%U[,k]
}

J=0
for(k in (1:length(Mult))){
  J=J+(n[k]*Mult[k])
}
W=(sum(n)-1)*J

##P-value using Chi-squared Distribution with 2 degrees of freedom (johnson and Grender)
k<-list(W,1-pchisq(W,2))
names(k)<-c("W", "p-value")
k
c(names(h01.R)[1],names(h01.L)[1])
```

```
##############################################################################
###########
## Using Test Statistic developeped by Sun (1997) for c-sample problems
##############################################################################
###########

library(plyr)
count=count(Sequence)

n1=count[1,2]
n2=count[2,2]
N=n1+n2


Seq=as.numeric(Sequence)

###Input the correct generated measure based on the hypothesis being tested
Measure=(R1-L1)+(R2-L2)

Mes.tilda=round(Measure)

#Adaptive
Adp.Mes=NewRank(Mes.tilda,Sequence)$`Apadtive Ranks`
a.bar=mean(Adp.Mes)

S.tilda1=sum(Adp.Mes[1:n1])
S.tilda2=sum(Adp.Mes[(n1+1):N])

Numerator=(N-1)*((n1*((S.tilda1/n1)-a.bar)^2)+(n2*((S.tilda2/n2)-a.bar)^2))
```

```
Denominator=sum((Adp.Mes-a.bar)^2)


Test.statistic=Numerator/Denominator

c(Test.statistic,pchisq(Test.statistic,1,lower.tail=FALSE))


##Rank based method

Rank.Mes=rank(Measure)


##USING EXACT WILCOXON DISTRIBUTION

wilcox.test(Rank.Mes[1:n1],Rank.Mes[(n1+1):(n1+n2)],alternative="two.sided",conf.int=TRUE,
correct = FALSE)

wilcox.test(Measure~Seq,alternative="two.sided",conf.int=TRUE,correct = FALSE)
```

```
#########################################################################
###Slope of Systolic and Diastolic Trend Example
###J&G Method
#########################################################################
########Attach the correct file
##SysSlope Example
Mydata1=read.table("C:/ SysSlope.csv",header=TRUE,sep=",")

##DiaSlope
Mydata2=read.table("C:/ DiaSlope.csv",header=TRUE,sep=",")

attach(Mydata)
names(Mydata)



##Load NewRank function

##Rank Averages for each Sequence (fixed time point)
###Hyp 1 and 2 (use sum for carry over effects and diff for treatment effects)
R.iks=rbind(NewRank((ST1.2-ST2.2),Sequence)$`Apadtive Ranks`,
        NewRank((ST1.4-ST2.4),Sequence)$`Apadtive Ranks`,
        NewRank((ST1.6-ST2.6),Sequence)$`Apadtive Ranks`,
        NewRank((ST1.8-ST2.8),Sequence)$`Apadtive Ranks`,
        NewRank((ST1.10-ST2.10),Sequence)$`Apadtive Ranks`
         )

c((NewRank((ST1.2-ST2.2),Sequence))$`Underlying Distribution`,(NewRank((ST1.4-
ST2.4),Sequence))$`Underlying Distribution`,
```

```
  (NewRank((ST1.6-ST2.6),Sequence))$`Underlying Distribution`,(NewRank((ST1.8-
ST2.8),Sequence))$`Underlying Distribution`,(NewRank((ST1.10-
ST2.10),Sequence))$`Underlying Distribution`)


R.iks2=rbind(rank((ST1.2-ST2.2)),
          rank((ST1.4-ST2.4)),
          rank((ST1.6-ST2.6)),
          rank((ST1.8-ST2.8)),
          rank((ST1.10-ST2.10)))


JG.5(R.iks,Sequence)


###Hyp 3 and 4 (use sum for carry over effects over time and diff for trtment effects over time)
R.iks=rbind(NewRank(((ST1.2-ST1.4)-(ST2.2-ST2.4)),Sequence)$`Apadtive Ranks`,
            NewRank(((ST1.4-ST1.6)-(ST2.4-ST2.6)),Sequence)$`Apadtive Ranks`,
            NewRank(((ST1.6-ST1.8)-(ST2.6-ST2.8)),Sequence)$`Apadtive Ranks`,
            NewRank(((ST1.8-ST1.10)-(ST2.8-ST2.10)),Sequence)$`Apadtive Ranks`)


c((NewRank(((ST1.2-ST1.4)-(ST2.2-ST2.4)),Sequence))$`Underlying
Distribution`,(NewRank(((ST1.4-ST1.6)-(ST2.4-ST2.6)),Sequence))$`Underlying Distribution`,
  (NewRank(((ST1.6-ST1.8)-(ST2.6-ST2.8)),Sequence))$`Underlying
Distribution`,(NewRank(((ST1.8-ST1.10)-(ST2.8-ST2.10)),Sequence))$`Underlying
Distribution`)


R.iks=rbind(rank((ST1.2-ST1.4)-(ST2.2-ST2.4)),
            rank((ST1.4-ST1.6)-(ST2.4-ST2.6)),
            rank((ST1.6-ST1.8)-(ST2.6-ST2.8)),
            rank((ST1.8-ST1.10)-(ST2.8-ST2.10)))


JG.4(R.iks,Sequence)
```

```
#########################################################################
###Slope of Systolic and Diastolic Trend Example
###F1-LD-F1 Method
#########################################################################
#Variables defined
Carry<-c((ST1.2+ST2.2),(ST1.4+ST2.4),(ST1.6+ST2.6),(ST1.8+ST2.8),(ST1.10+ST2.10))
Direct<-c((ST1.2-ST2.2),(ST1.4-ST2.4),(ST1.6-ST2.6),(ST1.8-ST2.8),(ST1.10-ST2.10))

Measure.f1ldf1 <- Carry
sequence <- c(Sequence,Sequence,Sequence,Sequence,Sequence)
subject=c(Sub,Sub,Sub,Sub,Sub)
n=length(Sequence)
time=c(rep(1,n),rep(2,n),rep(3,n),rep(4,n),rep(5,n))

## Test
f1.ld.f1(Measure.f1ldf1,time,sequence,subject, description = FALSE)$Wald.test
f1.ld.f1.adap(Measure.f1ldf1,time,sequence,subject, description = FALSE)$Wald.test
```

```
#####################################################################
## Plots generated for the thesis write-up
#####################################################################

layout(matrix(c(0,1,1,0,2,2,3,3), 2, 4, byrow = TRUE))
 plot(density(rnorm(10000),bw=1), col='red', lwd=3, main="Symmetric Distribution", xlab="")
 plot(density(-x,bw=1), col='red', lwd=3, main="Left Skewed", xlab="(Negatively Skewed)")
 plot(density(x,bw=1), col='red', lwd=3, main="Right Skewed",xlab="(Positively Skewed)")

 par(mfrow=c(1,1))

 x=rnorm(100000)
 plot(density(x),col="red", lwd=3, ylim=c(0, 0.5),main="Kurtosis",xlab="")
 lines(density(x*.8),col="green",lwd=3)
 lines(density(x*1.25),col="blue",lwd=3)
 legend("topright", col="black",inset=.05, title="Tail Weights",
       c("Light-Tailed","Moderate-Tailed","Heavy-Tailed"), fill=c("green","red","blue"),
horiz=FALSE)
```

References

Alex, R. M. (2010). Quantitative variation of blood pressure dynamics during sleep apnea. Biomedical Engineering, University of Texas at Arlington.

Alex, R., Manchikatla, S., Machiraju, K., Altuwaijri, E., Watenpaugh, D. E., Zhang, R., & Behbehani, K. (2014). Effect of apnea duration on apnea induced variations in cerebral blood flow velocity and arterial blood pressure. *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 270-273). IEEE.

Brunner, E., Domhof, S., & Langer, F. (2002). *Nonparametric Analysis of Longitudinal Data in Factorial Experiments.* John Wiley & Sons Inc.

Grizzle, J. (1965). The two-period change-over design and its use in clinical trials. *Biometrics, 21*, 467-480.

Grizzle, J. (1974). Corrections-The two-period change-over design and its use in clinical trials. *Biometrics, 30*, 727.

Hill, N., Padmanabhan, A., & Puri, M. (1988). Adaptive nonparametric procedures and applications. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 37*, 205-218.

Hogg, R., Fisher, D., & Randles, R. (1975). A two-sample adaptive distribution-free test. *Journal of American Statistical Association, 70*, 656-661.

Johnson, W. D., & Grender, J. M. (1993). Multivariate nonparametric analysis for the two-period crossover design with application in clinical trials. *Journal of Biopharmaceutical Statistics, 3(1)*, 1-12.

Koch, G. (1972). The use of non-parametric methods in statistical analysis of the two-period change-over design. *Biometrics, 28*, 577-584.

Koch, G., Gitomer, S., Skalland, L., & Stokes, M. (1983). Some non-parametric and categorical data analyses for a change-over design study and discussion of apparent carry-over effects. *Stat Med, 2(3)*, 397-412.

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *Journal of Statistical Software, 50*(12), 1-23. Retrieved from http://www.jstatsoft.org/v50/i12/

Puri, M. L. (1965). Some distribution-free K-sample rank tests of homogeneity against ordered alternatives. *Communications on Pure and Applied Mathematics, 18*, 51-63.

Puri, M., & Sen, P. (1971). *Nonparametric Methods in Multivariate Analysis.* New York: Wiley.

Sun, S. (1997). A class of adaptie distribution-free procedures. *Journal of Statistical Planning and Inference, 59*, 191-211.

Taulbee, J. (1982). A note on the use of nonparametric methods in statistical analysis of the two-period change-over design. *Biometrics, 38*, 1053-1055.

University of Illinois at Chicago. (2008). *Crossover Designs Under Subject Dropout.* ProQuest.

Biographical Information

Afshan Boodhwani (born Afshan Nayani) was born in Karachi, Pakistan on September 19, 1983, the daughter of Nazneen and Aslam Sadruddin. She received her Honors Degree Bachelor of Science in Mathematics and Actuarial Science from University of Toronto in Toronto, Canada. In 2012 she obtained her Masters of Science in Mathematics from Tarleton State University in Stephenville, Texas. She continued to stay at Tarleton State University afterwards to serve as faculty in the Department of Mathematics at Tarleton State University where she taught several undergraduate courses including, but not limited to, Introductory to Probability and Statistics, Mathematics for Business, and College Algebra.

After marrying her husband, Aleem Boodhwani in 2014, she continued to get her Doctor of Philosophy in Mathematics from University of Texas at Arlington, Arlington, Texas in 2017. During this time she conducted research under the NSF grant for "Arlington Undergraduate Research Based Achievement for STEM (AURAS)" Program, under the supervision of Dr. James Alvarez.

She was also involved with research at the LINK Lab at University of Texas at Arlington under the supervision of Dr. George Siemens and Dr. Catherine Spann on the NSF grant for “BIGDATA: Collaborative Research: F: COCOA: From Mining Massive Datasets to Designing Support for Coordinating Explanatory Coherence, Consensus, and Action”. Her other research projects with the LINK Lab include studying the impact of mindfulness in college students, and the impact of screen time on social and emotional development for student of grade 6-8 (based on data set obtained from DECD-Australia).