University of Alberta

Sequential Tests for the Comparison of Several Treatments with Normal Response

by

Ming Jie Yang

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of **Master of Science**

in

Statistics

Department of Mathematical & Statistical Sciences

Edmonton, Alberta
Fall 2008

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-47446-4
Our file Notre référence
ISBN: 978-0-494-47446-4

Abstract

The comparison of several treatments is an important statistical problem especially in clinical trials. In this thesis, we use methods of sequential analysis to test the statistical mean difference of several treatments. Since sequential analysis could stop a trial earlier than a fixed sample test, ethical and economic are the main reasons for using sequential analysis.

This thesis compares several treatments that have independent normal responses with unknown σ^2. The methods used in the thesis are group sequential analysis and fully sequential analysis. Monte Carlo simulations are performed to carry out the sample size calculation while fixing the Type I error and the power. The main purse of the thesis is to find a relationship among these methods and the relationship between fixed-sample size and sequential maximal sample size. An application is presented later in the thesis to show the benefits of using sequential analysis.

Contents

Chapter 1

Introduction

The comparison of several treatments is an important statistical problem especially in clinical trials. Let J be the number of treatments, and suppose observations of each treatment are from normal distributions with mean $\mu_j,\ j = 1, \ldots, J$, and variance σ^2. The null hypothesis can be written as $H_0 :\ \mu_1 = \mu_2 = \ldots = \mu_J$, and it is tested against the alternative hypothesis $H_a :$ not all means are equal. Various methods are used to test the equality or statistical difference of several treatments: sequential analysis is one of them. There are two types of sequential analysis: (1) fully sequential analysis, in which data is analyzed after every new observation is allocated in the data set until the hypothesis selected, and (2) group sequential analysis, in which interim analyses are performed. Since sequential analysis could stop a trial earlier than a fixed sample size test, the main reasons for using sequential analysis are ethical and economic.

Group sequential analysis can be performed in two different ways. One is the original group sequential analysis, which is designed to fix the number of analyses K in the design stage and requires approximately even observations in each group. The original sequential tests can get the overall Type I error (selecting H_a when H_0 is true) very close or equal to the required Type I error α. But, as group size (sample size for each treatment at each stage) becomes more and more uneven, the original sequential tests cannot guarantee Type I error will be equal or close to α. In this case, another method is needed. Flexible group sequential analysis uses the error spending approach (see Lan & DeMets (1983) or Jennison & Turnbull (2000) Ch. 7. for details) to deal with unpredicted group size and guarantee Type I error exactly equal to α. This method does not require the number of analyses K to be fixed in advance and also does not require a roughly equal group size for each group. Since the original group sequential analysis is the common method usually used by statisticians in analyzing data, it is more valuable to use the original group sequential analysis in this thesis than the flexible method.

This thesis compares several treatments that have independent normal responses with unknown σ^2. In sequential analysis, the sample size needed to reach the required power is an important issue. Monte Carlo simulations are performed to carry out the sample size calculation. Different methods

of sequential analysis are evaluated while fixing the Type I error α and the required power (selecting H_a when H_a is true) at different levels of treatment differences with unknown σ^2. The main purpose of the thesis is to find a relationship between fixed-sample size and sequential maximal sample size, and to compare the maximal sample sizes and average stopping times of different methods. Fixed-sample size is the number of observations needed to obtain the required power and type I error when analyzing the whole data set at one time. The maximal sample size is the necessary maximum sample size to obtain the required power and type I error when using sequential methods. In sequential analysis, the sample size at which decision is made is a random variable, and these methods are evaluated using the average stopping time, the expected sample size when reaching the conclusion (reject H_0). The author is also interested in comparing the effectiveness of fully sequential tests and group sequential tests.

Chapter two compares two treatments. The fixed-sample size calculation will be presented first. In group sequential tests, the theory of O'Brien & Flemming (OBF) (OBF (1979)) and Pocock (Pocock (1977)) tests will be introduced under known and unknown σ^2. Rao test (Gombay & Hussein (2006)) will be used in the fully sequential analysis. In the simulation, patients are allocated to one treatment with allocation rate of 0.5; Hence, the sample size for each treatment in each stage is not exactly the same. The power and the Type I error are fixed to 0.9 and 0.05, respectively. The

simulation is based on $10,000$ replicates to get the maximal sample sizes and average sample sizes at mean differences $\delta = 0.1,\ 0.2,\ 0.5$, and 0.9. For group sequential tests, the number of analyses K are set to $K = 5$ and $K = 20$. The simulation algorithm and results are given as well.

Chapter three introduces a multi-treatment comparison. The author presents an approximate method (Chow, Shao, and Wang (2003)) and an exact method (Guenther(1977)) to carry out fixed-sample size calculations. OBF and Pocock tests are still used for group sequential tests under known and unknown σ^2 (Jennison & Turnbull (2000)). The sequential F-test (Siegmund (1980)) and the Rao tests 1 and 2 (Gombay & Serban (2007)) are used in fully sequential analysis. In the simulation, patients in each treatment in each stage are exactly the same for both group sequential and fully sequential analyses. The power and the Type I error are fixed to 0.8 and 0.05, respectively. The simulation is based on $5,000$ replicates to get the maximal sample sizes and average sample sizes at mean differences $\delta = 0.2,\ 0.4,\ 0.6$, and 0.8. Here, the definition of δ is different from that in Chapter two (see Ch. 3. for details). For group sequential tests, the number of analyses are set to $K = 5$ and $K = 10$. The simulation algorithm and results are also given. Finally, the author presents an application to show the benefits of using sequential methods. This application compares the equality of three treatments for an orthodontic clinical trial. The results show that all sequential methods could stop the trial earlier and get

the same conclusion.

We will be using the following notations.

n_f=fixed-sample size for each treatment,

N_f=fixed-sample size for all treatments added,

n_0=the maximal sample size for each treatment,

N_0=the maximal sample size for all treatments added,

$avst$=the average stopping time for each treatment,

and $AVST$=the average stopping time for all treatments added.

In other words, let J be the number of treatments, then $N_f = n_f * J$, $AVST = avst * J$, and $N_0 = n_0 * J$, if the number of treatments is greater than two. The relationships of the above formulas are not satisfied when the number of treatments is two because in two-treatment comparison, sample sizes for each treatment are not exactly the same in this thesis.

Chapter 2

The Comparison of Two Treatments with Normal Response

This chapter is to review the work by Gombay & Hussein (2006) for two-treatment comparison.

2.1 Fixed-Sample Size Calculation

For two treatments A and B, n patients are allocated to each treatment. Let X_{Ai} and X_{Bi} be the responses of patients receiving the two treatments, respectively, $i = 1, \ldots, n$. Suppose X_{Ai} are normally distributed with mean μ_A and variance σ^2, which we write $X_{Ai} \sim N(\mu_A, \sigma^2)$, $i = 1, \ldots, n$. Like-

wise, suppose $X_{Bi} \sim N(\mu_B, \sigma^2)$, and all observations are independent.

The null hypothesis is $H_0 : \mu_A = \mu_B$, and the alternative hypothesis is $H_a : \mu_A \neq \mu_B$ or $H_a : \mu_A - \mu_B = \theta = \pm\delta$. A Type I error α is the probability to reject H_0 when H_0 is true. Power $1-\beta$ is the probability to reject H_0 when H_a is true, where β is a Type II error (selecting H_0 when H_a is true). If we want to detect the difference at $|\mu_A - \mu_B| = \delta$ with given type I error α and power $1-\beta$,

α =P(reject H_0 when H_0 is true)= $P_{\theta=0}(|\bar{X}_A - \bar{X}_B| \geq c)$

$$= P\Big(\frac{|\bar{X}_A - \bar{X}_B|}{\sqrt{2\sigma^2/n}} \geq \frac{c}{\sqrt{2\sigma^2/n}}\Big)$$

$$\Rightarrow \frac{c}{\sqrt{2\sigma^2/n}} = \Phi^{-1}(1-\alpha/2), \tag{2.1.1}$$

$1-\beta$ =P(reject H_0 when H_a is true)= $P_{\theta=\delta}(|\bar{X}_A - \bar{X}_B| \geq c)$

$$= P\Big(\frac{|\bar{X}_A - \bar{X}_B| - \delta}{\sqrt{2\sigma^2/n}} \geq \frac{c-\delta}{\sqrt{2\sigma^2/n}}\Big)$$

$$\Rightarrow \frac{c-\delta}{\sqrt{2\sigma^2/n}} = -\Phi^{-1}(1-\beta), \tag{2.1.2}$$

where Φ denotes the standard normal cumulative distribution function (cdf). From the formulas (2.1.1) and (2.1.2), the necessary sample size is

$$n_f(\alpha, \beta, \delta, \sigma^2) = \frac{[\Phi^{-1}(1-\frac{\alpha}{2}) + \Phi^{-1}(1-\beta)]^2 2\sigma^2}{\delta^2}. \tag{2.1.3}$$

Under H_0, the statistic has standard normal distribution. In other words,

$$Z = \frac{1}{\sqrt{2n\sigma^2}}\left(\sum_{i=1}^{n} X_{A_i} - \sum_{i=1}^{n} X_{B_i}\right) \sim N(0,1) \text{ and } n = n_f.$$

Rejection rule of fixed-sample test: reject H_0 if $|Z| > \Phi^{-1}(1-\alpha/2)$; otherwise, fail to reject H_0.

2.2 Group Sequential analysis

Group sequential analysis has good features such as early stopping, and the data is analyzed at intervals rather than after every new observation (Jennison & Turnbull (2000)).

For the original group sequential tests, the most important group sequential analysis came from Pocock (1977) and OBF (1979) tests. In a two-treatment comparison, choose the number of interim analyses K and a group size m. In each interval or stage k, $k = 1, 2, 3, \ldots, K$, m new observations are allocated in treatment A, and another m new observations are allocated in treatment B. Note that group size m and K are determined by the power requirement. The analysis proceeds on the accumulating data after $2m$ new observations are allocated in the two treatments. A standard-

ized statistic Z_k $(k = 1, 2, \ldots, K)$ is calculated using the first k groups of observations. If Z_k is greater than or equal to the critical value c_k, which is calculated based on the type I error α, H_0 is rejected after the k^{th} analysis and the analysis stops at this stage. If Z_k is less than c_k for all k, H_0 is accepted after the last analysis.

We will introduce some concepts of Pocock and OBF sequential analysis methods under known and unknown σ^2.

2.2.1 Pocock and OBF Tests – When σ^2 is Known

We first focus on the simplest case when the number of patients is the same for each treatment in each stage.

Pocock Test

For two treatments A and B, m patients are allocated to each treatment in each stage, and the analysis will proceed at each stage for the accumulated data set. K, the maximum number of stages, is chosen before the patients are assigned in the sequential study. Let X_{Ai} and X_{Bi} be the responses of subjects receiving the two treatments, $i = 1, \ldots,$ and X_{Ai} and X_{Bi} are normally distributed with same variance σ^2 and means μ_A and μ_B, respectively, that is $X_{Ai} \sim N(\mu_A, \sigma^2)$ and $X_{Bi} \sim N(\mu_B, \sigma^2)$, and the observations are independent.

As Jennison & Turnbull (2000) explain "Pocock adopted the idea of a 'significance test' at a constant nominal significance level to analyze the accumulating data at a relatively small number of times over the course of a study". Hence, the critical value $c_k = C_p(K, \alpha)$ is a constant across all k intervals when K and α are fixed. The standardized statistic after each group of observations is

$$Z_k = \frac{1}{\sqrt{2mk\sigma^2}}\Big(\sum_{i=1}^{mk} X_{A_i} - \sum_{i=1}^{mk} X_{B_i}\Big), \quad k = 1, \ldots, K. \tag{2.2.1}$$

The Value of $C_P(K, \alpha)$ is calculated using the following formula to obtain the required type I error α,

$$P_{\delta=0}\{|Z_k| \geq c_k \text{ for some } k = 1, ..., K\} = \alpha. \tag{2.2.2}$$

The test rule of Pocock test:

after group $k = 1, ..., K-1$:

$$\text{if } |Z_k| \geq C_P(K, \alpha) \Rightarrow \text{stop, reject } H_0,$$
$$\text{otherwise} \Rightarrow \text{continue to group } k+1;$$

after group K:

$$\text{if } |Z_K| \geq C_P(K, \alpha) \Rightarrow \text{stop, reject } H_0,$$
$$\text{otherwise} \Rightarrow \text{stop, accept } H_0.$$

Constants $C_p(K, \alpha)$ are displayed in Table (4.1) in the appendix for the Type I error $\alpha = 0.01,\ 0.05$, and 0.1 (Jennison & Turnbull (2000) or Pocock (1977)).

When testing the statistical difference of two treatments, one of the important properties of Pocock test is the great opportunity for an early stop.

OBF Test

Instead of using a 'significance test' at a constant nominal significance level to analyze the accumulating data, OBF test uses increasing nominal significance level as the study progress to analyze the accumulating data. Hence, one of the main properties of OBF test is that it is more difficult to reject H_0 at early analyses but easier later on. $c_k = C_B(K, \alpha)\sqrt{K/k}$ is decreasing with k when K and α are fixed, and the value of $C_B(K, \alpha)$ is calculated using equation (2.2.2) to obtain the required type I error α.

The test rule of OBF test:

after group $k = 1, ..., K-1$:

$$\text{if } |Z_k| \geq C_B(K, \alpha)\sqrt{K/k} \Rightarrow \text{stop, reject } H_0,$$
$$\text{otherwise} \Rightarrow \text{continue to group } k+1;$$

after group K:

$$\text{if } |Z_K| \geq C_B(K, \alpha) \Rightarrow \text{stop, reject } H_0,$$

otherwise $\Rightarrow$ stop, accept H_0.

Values of $C_B(K, \alpha)$ which ensure an overall Type I error probability α (0.01, 0.05, and 0.1) are provided in the appendix Table (4.2) (Jennison & Turnbull (2000) or OBF (1979)).

2.2.2 Group Sequential t-tests – When σ^2 is unknown

Now consider two-treatment comparison when σ^2 is unknown. Let X_{Ai} and X_{Bi} be the responses of subjects receiving the two treatments, $i = 1, \ldots,$ and X_{Ai} and X_{Bi} are normally distributed, ie. $X_{Ai} \sim N(\mu_A, \sigma^2)$ and $X_{Bi} \sim N(\mu_B, \sigma^2)$, $i = 1, \ldots,$ and the observations are independent. We are interested in testing the null hypothesis $H_0 : \mu_A = \mu_B$, and the alternative hypothesis $H_a : \mu_A \neq \mu_B$ with type I error α and power $1 - \beta$ at specific difference $\delta = |\mu_A - \mu_B|$. Some definitions, such as m, the group size, are the same as the previous definitions. The t-statistic after each group of observations is

$$T_k = \frac{\sum_{i=1}^{mk} X_{Ai} - \sum_{i=1}^{mk} X_{Bi}}{\sqrt{2mks_k^2}}, \quad k = 1, \ldots, K, \tag{2.2.3}$$

where

$$s_k^2 = \frac{\sum_{i=1}^{mk}(X_{Ai} - \bar{X}_A^{(k)})^2 + \sum_{i=1}^{mk}(X_{Bi} - \bar{X}_B^{(k)})^2}{2(mk-1)}. \tag{2.2.4}$$

$\bar{X}_A^{(k)}$ and $\bar{X}_B^{(k)}$ are the means of $X_{A1}, \ldots, X_{A,mk}$ and $X_{B1}, \ldots, X_{B,mk}$, respectively. T_k has a marginal t_{2mk-2} distribution, which means T_k has t-distribution with degree of freedom $2mk - 2$.

To test the null hypothesis, we need to specify the critical values for each group analysis. Pocock (1977) suggested using the two-sided significance levels, $2(1 - \Phi(c_k))$, defined for the Z-statistics but apply these to the t-statistics $T_1, \ldots, T_K$. Denote $t_{\upsilon,q}$ the upper q quantile of a t-distribution on υ degree of freedom, which means $P(T > t_{\upsilon,q}) = q$ when $T \sim t_\upsilon$. When σ^2 is known, the two-sided significance levels is determined by $P(|Z_k| \geq c_k) = 2(1 - \Phi(c_k))$. Hence, when σ^2 is unknown, we modify the critical values c_k using the t-distribution so that we get $P(|T_k| \geq t_{2mk-2,1-\Phi(c_k)}) = 2(1 - \Phi(c_k))$.

The test rule:

we reject H_0 at analysis k if

$$|T_k| \geq t_{2mk-2,1-\Phi(c_k)},$$

and we accept H_0 if H_0 has not been rejected at analysis K, where c_k is the critical values defined by Pocock or OBF.

2.3 Fully Sequential Analysis—Sequential Rao Test

The sequential Rao test was first suggested by Gombay (2002). The test of two-treatment comparison is proposed by Gombay and Hussein (2006). Data is analyzed after every new observation. For the two treatments A and B, let X_{Ai} and X_{Bj} be the independent responses, and assume that X_{Ai} and X_{Bj} have $N(\mu_A, \sigma^2)$ and $N(\mu_B, \sigma^2)$ distributions, respectively, $i = 1, \ldots,$ and $j = 1, \ldots$. Patients are allocated to treatment A with probability u and patients are allocated to treatment B with probability $1 - u$. At stage k, we have n patients in treatment A, and n' patients in treatment B, and $k = n + n'$. Then n is a random variable that has binomial(k, u) distribution. Denote N_0 the maximal sample size (or truncation point) for the total of the two treatments, which means the test should stop if N_0 observations have been obtained.

The null hypothesis is $H_0 : \mu_A = \mu_B$, and the alternative hypothesis is $H_a : \mu_A \neq \mu_B$ with σ^2 unknown.

Sequential Rao Test:

we reject H_0 at stage k if $(k/N_0)^{1/2}(R_k^*)^{1/2} \geqslant CV(\alpha)$, $k = 2, 3, \ldots, N_0$; otherwise fail to reject H_0, where α is a level α of the test, and R_k^* is defined by (2.3.1) and (2.3.2),

$$R_k^* = \frac{(n\sum_{i=1}^{n} X_{Ai} - n'\sum_{j=1}^{n'} X_{Bj})^2}{(n+n')nn'\hat{\sigma_k}^2}, \tag{2.3.1}$$

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^{n} X_{Ai}^2 + \sum_{j=1}^{n'} X_{Bj}^2}{n+n'} - \Big(\frac{\sum_{i=1}^{n} X_{Ai} + \sum_{j=1}^{n'} X_{Bj}}{n+n'}\Big)^2. \tag{2.3.2}$$

It was proven in Gombay and Hussein (2006) that $(k/N_0)^{1/2}(R_k^*)^{1/2}$, $k = 1, \ldots, N_0$, is well approximated by a Brownian motion $B(t)$, $0 \leq t \leq 1$. So, the critical value $CV(\alpha)$ can be approximated using the well known distribution (see for example Csörgő and Révész (1981))

$$1 - P\big[\sup_{0<t<1}|B(t)| > y\big] = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\Big(\frac{-\pi^2(2k+1)^2}{8y^2}\Big). \tag{2.3.3}$$

The sum term of (2.3.3) is from zero to infinity, but $k = 5$ is sufficient for the calculations. Let y on the right hand side of the equation be a sequence from the ranging 1.5 to 3 by 0.001, and calculate the right hand side of the equation and get a sequence of $1 - p$. The left hand side of the equation $1 - p$ is equal to $1 - \alpha$. Choose the critical value y, which is corresponding to the value of α. Table (2.1) shows the critical values $CV(\alpha)$ for different levels of significance ($\alpha = 0.01, 0.05,$ and 0.1).

Table 2.1: Critical value for different type I error

α	$CV(\alpha)$
0.01	2.807
0.05	2.241
0.10	1.960

2.4 Simulation Procedure and Results

2.4.1 Simulation

In the simulation, let the outcomes of treatment A be distributed as $N(\mu_A, \sigma^2)$ and the outcomes of treatment B be distributed as $N(\mu_B, \sigma^2)$. The allocation rate is $u = 0.5$ for the three sequential tests, Pocock test, OBF test, and sequential Rao test. Denote $AVST$ the average stopping time (or Average Sample Number) for two treatments added and $\delta = |\mu_A - \mu_B|$ the mean difference. The power and the Type I error are fixed to 0.9 and 0.05, respectively. The mean differences are set to be 0.1, 0.2, 0.5, and 0.9. For Pocock and OBF tests, we consider $K = 5$ and 20. Through the simulation, we can get N_0 and $AVST$ and also get the empirical size $\widehat{\alpha}$ at the given maximum sample size N_0. $\frac{N_0}{N_f}$ is calculated for the purpose of comparing the relationship between the fixed-sample size N_f and the maximum sample size N_0.

In the Pocock and OBF simulation, the group sizes are slightly different

for treatment A and treatment B since we set the allocation rate u to 0.5; hence, in the simulation, T_k in (2.2.3) and s_k^2 in (2.2.4) cannot be used. Let us take a group sequential test with five groups of observations ($K = 5$) as an example. Suppose we set $N_0 = 100$. Then for each group analysis stage, we need additional 100/5=20 patients together in treatments A and B. The number of additional patients allocated to treatment A is binomial$(20, 0.5)$. In each group analysis stage, let p be the number of additional patients observed to treatment A, q be the number of additional patients observed to treatment B , and $p + q = 20$ should be satisfied. At stage k, we have n patients in treatment A and m patients in treatment B. The t-statistic after each group of observations is

$$T_k = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{(m+n)s_k^2/mn}}, \tag{2.4.1}$$

where

$$s_k^2 = \frac{1}{n+m-2}(\sum_{i=1}^{n}(X_{Ai} - \bar{X}_A)^2 + \sum_{i=1}^{m}(X_{Bi} - \bar{X}_B)^2). \tag{2.4.2}$$

$\bar{X}_A$ and $\bar{X}_B$ are the means of $X_{A1}, \ldots, X_{A,n}$ and $X_{B1}, \ldots, X_{B,m}$, respectively. T_k has a marginal t_{n+m-2} distribution, which means T_k has t-distribution with degree of freedom $m + n - 2$. Note that formulas (2.2.3) and (2.2.4) are the special cases of (2.4.1) and (2.4.2), respectively.

The test rule:

we reject H_0 at analysis k if

$$|T_k| \geq t_{m+n-2,1-\Phi(c_k)}, \tag{2.4.3}$$

and we accept H_0 if H_0 has not been rejected at analysis K, where c_k is the critical values defined by Pocock or OBF.

2.4.2 Algorithm

This simulation involved generating observations for the two treatments A and B with normal responses, and mean vector of the two treatments is $\mu = (\mu_A, \mu_B)^t$. We set $\mu = (0, \mu_B)^t$, where μ_B is determined by the value of δ and $\delta = |\mu_A - \mu_B|$. When simulating the responses of the two treatments, we simply set $\mu_B = \delta$ and $\sigma^2 = 1$.

Group Sequential Analysis

1. Set $K = 5$ or $K = 20$, and the initial value N_0 which is equal to the fixed-sample size N_f. N_f is used as starting point for the calculation of N_0.

2. Calculate the number of patients for each stage using N_0/K, and if N_0/K is not an integer, then the number of patients in each stage is

the smallest integer not less than the corresponding N_0/K. Denote the number of patients for each stage by nts. Hence, in software R, the program language is written as $nts = ceiling(N_0/K)$.

3. Generate patients for each treatment at each stage (starting at $k = 1$). With probability (allocation rate) 0.5, p of nts patients are allocated to treatment A, and q ($q = nts - p$) patients are allocated to treatment B. Hence, p is distributed as binomial $(nts, 0.5)$. In software R, generate p using $p = rbinom(1, nts, 0.5)$, and then $q = nts - p$. Also, the responses of treatments A and B have $N(0, 1)$, and $N(\mu_B, 1)$ distributions, respectively. At stage k, n patients are allocated in treatment A, and m patients are allocated in treatment B.

4. Calculate s_k^2 and T_k according to the equations (2.4.2) and (2.4.1).

5. Compare $|T_k|$ with $t_k = t_{m+n-2,1-\Phi(c_k)}$ according the equation (2.4.3) when $k \leq K$. If $|T_k| \geq t_k$, reject H_0; otherwise, continue to assign another nts patients for the two treatments, and redo steps 3, 4 and 5 for stage $k + 1$.

6. Calculate the simulated power based on 10000 replicates. If the power is far below the required power 0.9, increase N_0, and find the maximal value of N_0 until the simulated the power is just below 0.9. Calculate $AVST$. Record the values of the final N_0, power and average stopping time $AVST$ as $n1$, $pw1$ and $En1$. Continue to increase N_0, and find

the minimal value of N_0 until the simulated power is just over 0.9. Calculate $AVST$. Record the values of the final N_0, power and $AVST$ as $n2$, $pw2$ and $En2$.

7. Calculate N_0 and $AVST$ using the linear interpolation method (This method can be found in any numerical analysis text book, eg. Burden & Faires (2001)). According to this method,

$$\frac{n2 - N_0}{n2 - n1} = \frac{pw2 - 0.9}{pw2 - pw1}, \; and \; \frac{En2 - AVST}{En2 - En1} = \frac{pw2 - 0.9}{pw2 - pw1}.$$

Hence, $N_0 = (0.9 - pw2)(n2 - n1)/(pw2 - pw1) + n2$ and $AVST = (0.9 - pw2)(En2 - En1)/(pw2 - pw1) + En2$.

8. Simulate $\hat{\alpha}$. The empirical size $\hat{\alpha}$ is simulated using the maximal sample size N_0 which is calculated in step 7. Redo steps 2, 3, 4, and 5. But, in step 3, the distribution for the treatment B should be changed to treatment B $\sim N(0, 1)$. Then calculate the $\hat{\alpha}$.

Sequential Rao test

1. Set initial N_0 which is equal to the fixed-sample size N_f.

2. Generate an observation from $U(0, 1)$, the uniform distribution on (0,1). Let $u_k \sim U(0, 1)$. If $u_k > 0.5$, allocate this patient to treatment A; otherwise, allocate this patient to treatment B. Again, treatments A and B have $N(0, 1)$ and $N(\mu_B, 1)$ distributions, respectively.

3. Once there is at least one patient in each treatment, calculate $\hat{\sigma}_k^2$ and R_k^* according to the equations (2.3.2) and (2.3.1).

4. Compare $R_k = (k/N_0)^{1/2}(R_k^*)^{1/2}$ with the critical value $CV(\alpha)$ (see table (3.1)). If $R_k \geq CV(\alpha)$, reject H_0; otherwise, continue to include another patient according to step 2, and redo steps 3 and 4.

5. Same procedures as steps 6, 7, and 8 in group sequential analysis.

2.4.3 Results

The simulation results are shown in Table (2.2). From Table (2.2), we can compare the three different methods based on the maximal sample size N_0, the ratio of N_0 and N_f, and the average stopping time $AVST$ under the fixed power (0.9) and the level of significance ($\alpha = 0.05$). RT denotes the ratio of N_0 and N_f, and RT_x means the ratio of N_0 and N_f using x method.

- Since $RT_{OBF} < RT_{Pocock}$, the maximal stopping point N_0 of OBF test is smaller than that of Pocock test. For given the same δ, RT of $K = 5$ is smaller than RT of $K = 20$, but $AVST$ of $K = 5$ is greater than $AVST$ of $K = 20$ for both OBF and Pocock tests.

- Pocock test has the highest value of RT (only except for $\delta = 0.9$ of Rao test, in which RT is just slightly higher than RT of Pocock),

Table 2.2: Simulation results for comparison of two normally distributed two-treatment. Set $\alpha = 0.05, 1 - \beta = 0.9$, allocation rate $u = 0.5$

K	$\hat{\alpha}$	δ	$AVST$	N_f	N_0	$RT = \frac{N_0}{N_f}$
OBF						
20	0.046	0.1	3036.26	4204	4513.00	1.07
5	0.050	0.1	3117.76	4204	4250.18	1.01
20	0.046	0.2	766.21	1052	1138.18	1.08
5	0.049	0.2	798.06	1052	1087.47	1.03
20	0.046	0.5	125.77	170	185.00	1.09
5	0.048	0.5	128.91	170	175.04	1.03
5	0.053	0.9	42.02	52	55.97	1.08
Pocock						
20	0.045	0.1	2808.57	4204	5633.48	1.34
5	0.049	0.1	2901.23	4204	5124.92	1.22
20	0.049	0.2	699.93	1052	1391.85	1.32
5	0.051	0.2	726.88	1052	1276.56	1.21
20	0.051	0.5	116.77	170	228.81	1.35
5	0.047	0.5	119.24	170	207.99	1.22
5	0.051	0.9	39.51	52	66.69	1.28
Rao Test						
n_0	0.051	0.1	2922.54	4204	4465.58	1.06
n_0	0.050	0.2	728.18	1052	1127.32	1.07
n_0	0.049	0.5	123.76	170	188.12	1.10
n_0	0.075	0.9	40.42	52	67.07	1.29

which means this test needs the largest maximal stopping point N_0, but Pocock has its advantage, in which it has the smallest average stopping time among these three methods for given same δ.

- OBF test has relatively small value of RT compared with Pocock test for each δ, small value of RT compared with Rao test when δ is high, and similar RT with Rao test when δ is small. But, OBF has the highest average stopping time among these three methods.

- The average stopping time of Rao test is between that of OBF and Pocock tests. RT of Rao is getting larger as δ increases .

- In terms of empirical significance levels, $\hat{\alpha}$ of OBF, Pocock, and Rao tests are all approximately equal to 0.05, only except for $\delta = 0.9$ of Rao test.

Chapter 3

The Comparison of Multi-Treatment with Normal Response

The following methods focus on simultaneous comparison of means of J univariate normal distributions, where J is the number of treatment arms.

3.1 Fixed-Sample Size Calculation

3.1.1 The Approximate Method

We are interested in the simultaneous comparison of the means of J univariate normal distributions. For treatments $1, 2, \ldots, J$, we allocate n patients to each treatment. Let X_{ji} be the i^{th} subject from the j^{th} treatment arm,

$i = 1, \ldots, n,\ j = 1, \ldots, J.$

The approximate method to calculate fixed-sample size for power requirement can be found in some undergraduate text books, e.g. by Chow, Shao, and Wang (2003). Let us consider the multiple-sample one-way ANOVA test first. The one-way analysis of variance model is

$$x_{ji} = \mu_j + \varepsilon_{ji},$$

where μ_j is the fixed effect of the j^{th} treatment and ε_{ji} is a random error in observing x_{ji}. It is assumed that $\varepsilon_{ji} \sim (0, \sigma^2)$. So, the sum squares of error is

$$SSE = \sum_{j=1}^{J} \sum_{i=1}^{n} (x_{ji} - \bar{x}_{j.})^2,$$

and the sum squares of treatments is

$$SST = \sum_{j=1}^{J} (\bar{x}_{j.} - \bar{x}_{..})^2,$$

where

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ji} \quad and \quad \bar{x}_{..} = \frac{1}{J} \sum_{j=1}^{J} \bar{x}_{j.}$$

For simultaneous comparison, the hypotheses of interest are

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_J$$

$$H_\alpha : \mu_i \neq \mu_j \quad \textit{for some } i \textit{ and } j,\ 1 \leq i < j \leq J. \tag{3.1.1}$$

The null hypothesis H_0 is rejected at the α level of significance if

$$F_A = \frac{nSST/(J-1)}{SSE/[J(n-1)]} > F_{\alpha,J-1,J(n-1)},$$

where $F_{\alpha,J-1,J(n-1)}$ is the α upper quantile of an F-distribution with $J-1$ and $J(n-1)$ degrees of freedom. Under the alternative hypothesis, the power of this test is

$$\begin{aligned} P(F_A > F_{\alpha,J-1,J(n-1)}) &= P\Big(\frac{nSST/(J-1)}{SSE/[J(n-1)]} > F_{\alpha,J-1,J(n-1)}\Big) \\ &\overset{(1)}{\approx} P\Big(\frac{nSST/(J-1)}{SSE/[J(n-1)]} > \chi^2_{\alpha,J-1}/(J-1)\Big) \\ &\overset{(2)}{\approx} P\Big(nSST/(J-1) > \sigma^2\chi^2_{\alpha,J-1}/(J-1)\Big) \\ &= P(nSST/\sigma^2 > \chi^2_{\alpha,J-1}), \end{aligned} \tag{3.1.2}$$

in which, (1) uses the approximation $(J-1)F_{\alpha,J-1,J(n-1)} \approx \chi^2_{\alpha,J-1}$ when $J(n-1)$ is large and $\chi^2_{\alpha,J-1}$ is the upper α quantile for a χ^2 distribution with $J-1$ degrees of freedom, and (2) uses the fact that $SSE/[J(n-1)]$ is approximately σ^2. So, under the alternative hypothesis, $nSST/\sigma^2$ is a non-central χ^2 distribution with $J-1$ degrees of freedom and non-centrality

parameter $\lambda = n\Delta$, where

$$\Delta = \frac{1}{\sigma^2}\sum_{j=1}^{J}(\mu_j - \bar{\mu})^2, \qquad \bar{\mu} = \frac{1}{J}\sum_{j=1}^{J}\mu_j. \tag{3.1.3}$$

So, in order to achieve power $1-\beta$, the sample size needed can be obtained by solving

$$\chi^2_{J-1}(\chi^2_{\alpha,J-1}|\lambda) = \beta,$$

where $\chi^2_{J-1}(\cdot|\lambda)$ is the cumulative distribution function of the non-central χ^2 distribution with $J-1$ degrees of freedom and non-centrality parameter λ. Different power, different significant level and different number of treatment groups will give different values of λ. The values of λ for different power (0.80 and 0.90) with different significant level (0.01 and 0.05) for different number of treatment groups ($J = 1, ..., J = 20$) are listed in Table (4.3) in the appendix (Chow, Shao, and Wang (2003)). Once Δ given, we can obtain λ from Table (4.3) in the appendix, and the required sample size as

$$n = n_f = \lambda/\Delta. \tag{3.1.4}$$

3.1.2 The Exact method

When comparing several treatments, the power of the tests about the means depends on a noncentral F distribution, which is determined by the degrees of freedom but also by the noncentrality parameter. Several

tables and graphs are available for calculating the power of the test. The most troublesome in evaluating the power is the noncentrality parameter which is a function of the unknown parameter σ. It is usual to calculate the power and to solve for the sample-size based on differences between hypothesized and alternative values of means selected as multiples of σ. This choice eliminates σ from the noncentrality parameter. An alternative way to eliminate σ is proposed by Guenther (1977), and is, perhaps, more intuitively appealing. In his paper, the form of the alternative hypothesis gives one or more means as the quantile of order p of a distribution with another mean. In our thesis, we choose the method of Guenther (1977). The details and explanation are as follows.

Suppose we are interested in testing the hypotheses of (3.1.1) and seeking the power when some means are equal to quantiles of order p of distributions with other means. Tables and graphs are available for the calculation of the power (eg. Odeh and Fox (1975)), and we need

$$\phi = \Big[\frac{n}{J}\sum_{j=1}^{J}(\mu_j - \bar{\mu})^2\Big]^{1/2}/\sigma, \tag{3.1.5}$$

where n is the sample size for each group, and $n = n_1 = n_2 = \ldots = n_J$.

However, our goal is to fix the power first and then try to find the sample-size. The following example explains the method. Suppose we have three

treatments, $J = 3$, and we want to test

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{ against } H_a : \text{ not all means are equal}$$

with $\alpha = 0.05$. We seek the minimum n at $1 - \beta = 0.8$ when two means are equal and the third mean is the quantile of order 0.20 for the first two distributions. Then, $\mu_1 = \mu_2$, $(\mu_3 - \mu_1)/\sigma = Z_{0.20}$, where Z_p is the upper quantile p of a standard normal distribution. So, $\bar{\mu} = \mu_1 + Z_{0.20}(\sigma/3)$, and $\sum_{j=1}^{3}(\mu_j - \bar{\mu})^2 = \frac{2}{3}Z_{0.20}^2\sigma^2$. Solving (3.1.5) for n, we get,

$$n = \left[\sigma^2 / \sum_{j=1}^{J}(\mu_j - \bar{\mu})^2\right] J\phi^2,$$

which for our problem reduces to $n = \frac{9\phi^2}{2Z_{0.20}^2} = 6.353\phi^2$. We note that the degrees of freedom for the denominator of the F ratio is $\nu_2 = nJ - J$, and then we have $n = (\nu_2/J) + 1$. The degrees of freedom for the nominator of the F ratio is $\nu_1 = J - 1$. For given n and power, we read ϕ using Lehmer (1944) Tables. If we choose $\nu_2 = \infty$, $n = \infty$, we read $\phi = 1.792$ and compute $n = 20.4 < \infty$. With $\nu_2 = 80$, $n = 27.6$, we read $\phi = 1.83$ and compute $n = 21.27 < 27.6$.With $\nu_2 = 60$, $n = 21$, we read $\phi = 1.838$, and compute $n = 21.46 > 21$. Hence $21.27 \leq n \leq 21.46$, and n=22 is the minimum sample size for each treatment. Hence, $n_f = 22$ for this case.

3.2 Group Sequential Analysis

Now, we are interested in comparing the means of J univariate normal distributions using group sequential analysis, where $J \geq 3$ (see Jennison & Turnbull (2000)). Independent observations are available from each arm $j = 1, \ldots, J$, and assumed to be distributed with $N(\mu_j, \sigma^2)$. The group size is g. For $j = 1, \ldots, J$ and $k = 1, \ldots, K$, denote $\bar{X}_{jk}$, the sample mean of the $n_k = gk$ responses from treatment arm j available at stage k. Also, denote s_k^2, the pooled within-arms estimate of σ^2 available at stage k.

3.2.1 Group Sequential Chi-Squared Tests–When σ^2 is known

When σ^2 is known, we can consider group sequential test based on monitoring successive χ^2 at each stage k to test the hypothesis of homogeneity of J normal means, $H_0 : \mu_1 = \mu_2 = \ldots = \mu_J$. The statistic is

$$S_k = \frac{n_k}{\sigma^2} \sum_{j=1}^{J} (\bar{X}_{jk} - \bar{X}_{.k})^2, \quad k = 1, \ldots, K. \tag{3.2.1}$$

Here, $n_k = gk$, the cumulative sample size on each arm, and

$$\bar{X}_{.k} = \frac{1}{J} \sum_{j=1}^{J} \bar{X}_{jk}$$

is the overall mean at stage k.

S_k has χ^2 distribution with $J-1$ degrees of freedom and noncentrality parameter

$$n_k \sum_{j=1}^{J} (\mu_j - \bar{\mu})^2 / \sigma^2,$$

where $\bar{\mu} = \frac{1}{J} \sum_{j=1}^{J} \mu_j$. We reject H_0 if $S_k \geq c_k$, at some k, $k = 1, \ldots, K$, and accept H_0 if $S_K < c_K$ at the final stage K. $c_1, \ldots, c_K$ are critical values to satisfy a specified Type I error probability α.

For Pocock test, with the constant nominal significance levels, set $c_k = C_P(p, K, \alpha)$, for $k = 1, \ldots, K$, where $p = J - 1$. For OBF test, set $c_k = (K/k) C_B(p, K, \alpha)$, for $k = 1, \ldots, K$. Values of $C_P(p, K, \alpha)$ and $C_B(p, K, \alpha)$ for $\alpha = 0.05$, $K = 1, \ldots, 10$, and $J = 2, \ldots, 5$ are shown in Table (4.4) in the appendix (Jennison & Turnbull (1991)).

3.2.2 Group Sequential F-Tests–When σ^2 is unknown

When σ^2 is unknown, σ^2 in (3.2.1) is replaced with its current estimate s_k^2. Then we can monitor the F-statistics

$$F_k = \frac{n_k}{(J-1)s_k^2} \sum_{j=1}^{J} (\bar{X}_{jk} - \bar{X}_{.k})^2, \quad k = 1, \ldots, K, \tag{3.2.2}$$

where

$$s_k^2 = \frac{1}{J(n_k - 1)} \sum_{j=1}^{J} \sum_{i=1}^{n_k} (X_{ji} - \bar{X}_j)^2, \tag{3.2.3}$$

where $\bar{X}_j^2 = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ji}$. Under the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_J$, F_k has an F distribution with degrees of freedom on $J-1$ and $J(n_k-1)$. H_0 is rejected if $F_k \geq c_k$ at some stage k, $1 \leq k \leq K$, and accepted if $F_K < c_K$ at the final stage K. The critical values, $c_1, c_2, \ldots, c_K$, can be approximately obtained by using the significance level approach introduced in section (2.2.2). Then the values of c_k satisfy

$$P(F_{J-1,J(n_k-1)} \geq c_k) = \alpha_k^{'}, \quad k = 1, \ldots, K, \tag{3.2.4}$$

where F_{ν_1,ν_2} denotes an F-distribution with degrees of freedom on ν_1 and ν_2 and $\alpha_k^{'}$ is the nominal significance level at stage k of a group sequential χ^2 test. The values $\alpha_k^{'}$ $(k = 1, \ldots, K)$ can be found using constants $C_P(p, K, \alpha)$ or $C_B(p, K, \alpha)$ in Table (4.4).

3.3 Fully sequential Analysis

Assume there are J groups (treatments), and the observations are made sequentially on vectors $X_k = (X_{1k}, X_{2k}, \ldots, X_{Jk})^t$, $k \geq 1$, where X_{jk} is the k^{th} observation from group j. The observations are independent and

normally distributed, i.e.

$$X_{jk} \ \ iid \ \ N(\mu_j, \sigma^2), \ for \ all \ k \geq 1, \ j = 1, \ldots, J.$$

We are interested in testing (3.1.1) again. The following two sections will introduce the methods to test the hypotheses. One method is sequential F-test using log likelihood ratio statistic, which is proposed by Siegmund (1980). The other method is using Rao's statistic, which involves two tests, Rao test 1 and Rao test 2 proposed by Gombay (2002) (or see Gombay & Serban (2007) for details).

3.3.1 Sequential F-test

Denote group sample means as $\bar{X}_{j.} = \frac{1}{k}\sum_{i=1}^{k} X_{ji}$, and the overall sample mean as $\bar{X}_{..} = \frac{1}{kJ}\sum_{j=1}^{J}\sum_{i=1}^{k} X_{ji}$, for $j = 1, \ldots, J$ based on k observations in each group. The log likelihood ratio statistic for testing the equal treatment means is

$$L_k = \frac{kJ}{2} log\{1 + \frac{J-1}{J(k-1)} F_k\}, \tag{3.3.1}$$

where

$$F_k = \frac{k\sum_{j=1}^{J}(\bar{X}_{j.} - \bar{X}_{..})^2/(J-1)}{\sum_{j=1}^{J}\sum_{i=1}^{k}(X_{ji} - \bar{X}_{j.})^2/(J(k-1))}. \tag{3.3.2}$$

F_k has an F-distribution with degrees of freedom $J-1$ and $J(k-1)$ under H_0 and the proof is as follows.

Proof:

Under H_0, $\mu_1 = \mu_2 = \ldots = \mu_J = \mu$,

as

$$X_{jk} \ \ iid \ \ N(\mu, \sigma^2), \ for \ all \ k \geq 1, \ j = 1, \ldots, J,$$

we have

$$\frac{1}{\sigma^2} \sum_{j=1}^{J} \sum_{i=1}^{k} (X_{ji} - \bar{X}_{j.})^2 \sim \chi^2_{J(k-1)}. \tag{3.3.3}$$

Also, as

$$\bar{X}_{j.} \ \ iid \ \ N(\mu, \sigma^2/k), \ for \ all \ k \geq 1, \ j = 1, \ldots, J,$$

we get

$$\frac{1}{\sigma^2/k} \sum_{j=1}^{J} (\bar{X}_{j.} - \bar{X}_{..})^2 \sim \chi^2_{J-1}. \tag{3.3.4}$$

Since $\sum_{j=1}^{J} \sum_{i=1}^{k} (X_{ji} - \bar{X}_{j.})^2$ in (3.3.3) and $\sum_{j=1}^{J} (\bar{X}_{j.} - \bar{X}_{..})^2$ in (3.3.4) are independent,

$$F_k = \frac{(3.3.4)/J - 1}{(3.3.3)/J(k-1)} = \frac{k \sum_{j=1}^{J} (\bar{X}_{j.} - \bar{X}_{..})^2/(J-1)}{\sum_{j=1}^{J} \sum_{i=1}^{k} (X_{ji} - \bar{X}_{j.})^2/(J(k-1))}.$$

Hence,

$$F_k \sim F^{J-1}_{J(k-1)}.$$

$\|$

It is convenient to use the following definition to do the test. Define $\bar{\mu} = \frac{1}{J} \sum_{j=1}^{J} \mu_j$ and $\alpha_j = \bar{\mu} - \mu_j$. Then $E(X_{jk}) = \bar{\mu} + \alpha_j$, with $\sum_{j=1}^{J} \alpha_j = 0$.

Then, the expression of the hypotheses can be equivalently written as

$$H_0 : \sum_{j=1}^{J} \alpha_j^2 = 0, \quad against \quad H_A : \sum_{j=1}^{J} \alpha_j^2 > 0. \tag{3.3.5}$$

Sequential F-test: Given integers $k_0 < n_0$ and constants $0 < c \leq a$, stop sampling at min(T,n_0), where

$$T = inf\{k : k \geq k_0, L_k > a\},$$

and reject H_0 if either $T \leq n_0$ and $L_{n_0} > c$.

The role of c is that if no rejection happened until the final stage n_0 using the fully sequential analysis (ie. using criterion $L_{n_0} \leq a$), then we add a fixed-sample test at the end of analysis using c as the critical value. Note, that $c \leq a$. In order to be consistent with the other sequential analyses, we set $c = a$.

Then, **the sequential F-test** becomes:
given integers $k_0 < n_0$ and constants $0 < a$, stop sampling at min(T,n_0), where

$$T = inf\{k : k \geq k_0, L_k > a\},$$

and reject H_0 if $T \leq n_0$.

According to Siegmund (1980), the power of the test depends on

$$\delta = \frac{1}{\sigma}(\sum_{j=1}^{J} \alpha_j^2)^{1/2}, \tag{3.3.6}$$

which is the square root of Δ defined in (3.1.3)

Then, the power function of the test is given by

$$P_\delta(T \leq n_0).$$

It is very difficult to get an exact value of the power function for $J \geq 3$. Siegmund (1980) provided an approximate method to calculate the constant a to obtain a level of significance α, and it is

$$\alpha \approx P_0(L_{k_0} \geq a) + P_0(k_0 < T \leq n_0). \tag{3.3.7}$$

The first term on the right-hand side of equation (3.3.7) can be obtained directly from tables of F-distribution, and it can be written as

$$P_0(L_{k_0} \geq a) = P_0\big(F_{k_0} \geq [e^{\frac{2a}{k_0 J}} - 1]\frac{J(k_0 - 1)}{J - 1}\big), \tag{3.3.8}$$

where F_{k_0} is distributed as an F-distribution with degrees of freedom $J-1$ and $J(k_0 - 1)$ under the null hypothesis. The proof of equation (3.3.8) is as follows.

Proof:

From equation (3.3.1), we get

$$L_{k_0} = \frac{k_0 J}{2} log\{1 + \frac{J-1}{J(k_0-1)} F_{k_0}\}.$$

Then,

$$\begin{aligned} P_0(L_{k_0} \geq a) &= P_0\Big(\frac{k_0 J}{2} log\{1 + \frac{J-1}{J(k_0-1)} F_{k_0}\} \geq a\Big) \\ &= P_0\Big(F_{k_0} \geq [e^{\frac{2a}{k_0 J}} - 1]\frac{J(k_0-1)}{J-1}\Big). \end{aligned} \tag{3.3.9}$$

From Siegmund (1980), the second term on the right-hand side of equation (3.3.7) is approximated as

$$\begin{aligned} &P_0(k_0 < T \leq n_0) \approx 2e^{-a}\Big(\frac{a}{J}\Big)^{(J-1)/2}\{\Gamma(\frac{J-1}{2})\}^{-1} \\ &\times \int_{l_1}^{l_2} x^{J-2} \upsilon_J(x)\Big(1 + \frac{x^2}{J}\Big)^{1/2}\{log(1 + \frac{x^2}{J})\}^{\frac{1-J}{2}} dx, \end{aligned} \tag{3.3.10}$$

where

$$\upsilon_J(x) \approx exp\{-0.583x(1 + \frac{x^2}{J})^{-1}\},$$

$$l_1 = \sqrt{J\{exp(\frac{2a}{n_0 J}) - 1\}}, \quad l_2 = \sqrt{J\{exp(\frac{2a}{k_0 J}) - 1\}},$$

and $\Gamma(\cdot)$ is the gamma function defined by

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt.$$

3.3.2 Tests Based on Rao's Statistic

Let $X_1, X_2, \ldots, X_k$ sequentially come from a distribution with density function $f(x, \theta, \eta)$, where k is the k^{th} observation, θ is the parameter of interest known under H_0 with dimension $d \geq 0$, and η is the nuisance parameter with dimension $p \geq 0$. Gombay (2002) proposed two sequential tests, Rao test 1 and Rao test 2, based on Rao's statistic. The hypotheses are

$$H_0 : \ \theta = \theta_0, \ \eta \ unknown, \ \ against \ \ H_A : \ \theta \neq \theta_0, \ \eta \ unknown. \quad (3.3.11)$$

The efficient score vector is defined as

$$V_k(\xi) = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \nabla_\xi log f(X_i, \xi), \quad (3.3.12)$$

where $\xi = (\theta, \eta)$, and ∇_ξ denotes the vector of partial derivatives. Then, Rao's statistic can be defined as

$$R_k(\xi) = V_k(\xi) I^{-1}(\xi) V_k^t(\xi),$$

where $I(\xi) = -E_\xi(\frac{\partial^2}{\partial_{\xi_i}\partial_{\xi_j}} log f(X;\xi))$ is the $(d+p)\cdot(d+p)$ information matrix. So, we can partition this matrix on $\xi = (\theta, \eta)$ as

$$I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}.$$

The inverse of I is

$$I^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix}.$$

To deal with the nuisance parameter, we need to replace it by its maximum likelihood estimator under H_0, that is the solution of the equation

$$\sum_{i=1}^{k} \nabla_\eta log f(X_i; \theta_0, \eta) = 0.$$

Then, the efficient score vector becomes d-dimensional vector

$$V_k(\theta, \hat{\eta_k}) = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \nabla_\theta log f(X_i; \theta_0, \hat{\eta_k}),$$

and the Rao's statistic becomes

$$R_k(\theta_0, \hat{\eta_k}) = V_k(\theta_0, \hat{\eta_k}) I^{11}(\theta_0, \hat{\eta_k}) V_k^t(\theta_0, \hat{\eta_k}). \tag{3.3.13}$$

TEST 1. Stop and conclude that H_0 is not supported by the data at the first k when

$$T_1(k) = \left(\frac{k}{n_0} R_k(\theta_0, \hat{\eta}_k)\right)^{1/2} > C_1(\alpha, d). \tag{3.3.14}$$

Fail to reject H_0 if it is not rejected by $k = n_0$.

Here, n_0 is the truncation point or maximal sample size for each treatment arm.

The critical value $C_1(\alpha, d)$ can be obtained by solving

$$1 - \alpha = \sum_{k=1}^{\infty} \frac{j_{\nu,k}^{\nu-1}}{2^{\nu-1} J_{\nu+1}(j_{\nu,k}) \nu!} exp\left(-\frac{j_{\nu,k}^2}{2C_1^2}\right),$$

where $\nu = d/2 - 1$, $J_\nu(x)$ is the Bessel function defined by

$$J_\nu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k (x/2)^{\nu+2k}}{k!(\nu + k)!},$$

and $0 < j_{\nu,1} < j_{\nu,2} < \ldots$ are positive zeros of $J_\nu(\cdot)$. Values of $C_1(\alpha, d)$ for different d and different levels of significance α (0.10, 0.05, and 0.01) are shown in Table (4.5) in the appendix (Gombay & Serban (2007)).

TEST 2. Stop and conclude that H_0 is not supported by the data at

Table 3.1: Critical value $C_2(\alpha, d, n_0)$ for different n_0 and d when α is fixed to 0.05.

d	$n_0 = 50$	$n_0 = 100$	$n_0 = 200$	$n_0 = 500$
2	3.490	3.540	3.585	3.633
3	3.830	3.880	3.920	3.970
4	4.105	4.155	4.200	4.249

the first k when

$$T_2(k) = (R_k(\theta_0, \hat{\eta}_k))^{1/2} > C_2(\alpha, d, n_0). \tag{3.3.15}$$

Fail to reject H_0 if it is not rejected by $k = n_0$.

The critical value $C_2(\alpha, d, n_0)$ can be obtained by using a result of Vostrikova (1981). Then, $C_2(\alpha, d, n_0)$ can be obtained by solving

$$\alpha = \frac{C_2^d exp(-C_2^2/2)}{2^{d/2}\Gamma(d/2)}\{log(n_0)(1 - \frac{d}{C_2^2}) + \frac{4}{C_2^2} + O(\frac{1}{C_2^4})\}. \tag{3.3.16}$$

The critical values $C_2(\alpha, d, n_0)$ used in the later simulation study are calculated using Maple software. Table (3.1) gives some critical values for different n_0 and d when α is fixed to 0.05.

Comparison of Three Treatments.

We are interested in comparing the means of three treatments ($J = 3$). Vec-

tors $X_k = (X_{1k}, X_{2k}, X_{3k})^t$ be observed, $k \geq 1$, and $X_{jk} \sim \ iid\ N(\mu_j, \sigma^2)$. The log likelihood function for the first k observations is

$$l_k = -\frac{3}{2}log2\pi - \frac{3}{2}log\sigma^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{3}\sum_{i=1}^{k}(X_{ji} - \mu_j)^2.$$

Using the following parametrization

$$\theta = (\theta_1, \theta_2) = \Big(\frac{\mu_1 + \mu_2 - 2\mu_3}{2\sigma^2\sqrt{6}}, \frac{\mu_1 - \mu_2}{2\sigma^2\sqrt{2}}\Big), \quad \eta = (\eta_1, \eta_2) = \Big(\frac{\mu_3}{2\sigma^2}, \frac{-1}{2\sigma^2}\Big),$$

we get $d = 2$ and $p = 2$. The hypotheses of (3.1.1) can be equivalently written as

$$H_0 : \theta = (0,0),\ \eta\ unknown,\ against\ H_a :\ \theta \neq (0,0),\ \eta\ unknown$$

The Rao's statistic (3.3.13) becomes

$$R_k = \Big(\frac{\sum_{i=1}^{k}(X_{1i} + X_{2i} - 2X_{3i})}{\hat{\sigma}_k\sqrt{6k}}\Big)^2 + \Big(\frac{\sum_{i=1}^{k}(X_{1i} - X_{2i})}{\hat{\sigma}_k\sqrt{2k}}\Big)^2, \qquad (3.3.17)$$

where

$$\hat{\sigma}_k^2 = \frac{1}{3k}\sum_{i=1}^{k}(X_{1i}^2 + X_{2i}^2 + X_{3i}^2) - \Big(\frac{1}{3k}\sum_{i=1}^{k}(X_{1i} + X_{2i} + X_{3i})\Big)^2. \qquad (3.3.18)$$

Comparison of Four Treatments.

When $J = 4$, Vectors $X_k = (X_{1k}, X_{2k}, X_{3k}, X_{4k})^t$ be observed, $k \geq 1$, and

$X_{jk} \sim iid\ N(\mu_j, \sigma^2)$. The log likelihood function for the first k observations is

$$l_k = -2log2\pi - 2log\sigma^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{4}\sum_{i=1}^{k}(X_{ji} - \mu_j)^2.$$

The null hypothesis of interest is $\theta_0 = (0,0,0)$, $d = 3$ and $p = 2$ using the following parametrization

$$\theta = (\theta_1, \theta_2, \theta_3) = \left(\frac{\mu_1 - \mu_2}{2\sigma^2\sqrt{2}}, \frac{\mu_3 - \mu_4}{2\sigma^2\sqrt{2}}, \frac{(\mu_1 + \mu_2) - (\mu_3 + \mu_4)}{4\sigma^2}\right)$$

$$\text{and } \eta = (\eta_1, \eta_2) = \left(\frac{\mu_3 + \mu_4}{2\sigma^2\sqrt{2}}, -\frac{1}{2\sigma^2}\right).$$

The Rao's statistic (3.3.13) becomes

$$R_k = \left[\frac{\sum_{i=1}^{k}(X_{1i} - X_{2i})}{\hat{\sigma}_k\sqrt{2k}}\right]^2 + \left[\frac{\sum_{i=1}^{k}(X_{3i} - X_{4i})}{\hat{\sigma}_k\sqrt{2k}}\right]^2 + \left[\frac{\sum_{i=1}^{k}(X_{1i} + X_{2i} - X_{3i} - X_{4i})}{2\hat{\sigma}_k\sqrt{k}}\right]^2, \tag{3.3.19}$$

where $\hat{\sigma}^2$, the estimator of σ^2 is

$$\hat{\sigma}_k^2 = \frac{1}{4k}\sum_{i=1}^{k}(X_{1i}^2 + X_{2i}^2 + X_{3i}^2 + X_{4i}^2) - \left(\frac{1}{4k}\sum_{i=1}^{k}(X_{1i} + X_{2i} + X_{3i} + X_{4i})\right)^2. \tag{3.3.20}$$

3.4 Simulation Process and Results

3.4.1 Simulation

This simulation is based on theories of group sequential analysis (F-test on Pocock and OBF tests) and fully sequential analysis (Sequential F-test, Rao test 1 and Rao test 2). The Type I error is fixed to 0.05. The power is fixed to 0.8 at treatments differences δ (δ = 0.2, 0.4, 0.6, and 0.8), where δ is defined in equation (3.3.6). The study is to simulate the maximal stopping points n_0 and the average stopping time *avst* for three- and four-treatment with normally distributed outcomes (unknown σ^2), and to evaluate the ratio of n_0 and fixed-sample size n_f (RT) for future reference.

Since the condition for calculating the fixed-sample size is when the alternative hypothesis is given in quantiles (see section (3.1.2)), we first need to translate the values of δ into the values of quantile. Then, we will present the procedures for carrying out the fixed-sample size using one example in three-treatment comparison and one example in four-treatment comparison, and setting the parameters of normal distributions of all treatments in the simulation .

For three treatments, we simply assume $\mu_A = \mu_B = 0$, so the mean vector can be written as $\mu = (0, 0, \mu_C)^t$. The alternative hypothesis is to test the statistical difference of the three treatments where first two means are

equal and the third mean is the quantile of order p for the first two distributions. Then, $\mu_A = \mu_B$, $(\mu_C - \mu_A)/\sigma = Z_p$. Hence, $\mu = (\mu_A, \mu_B, \mu_C)^t$ can also be written as $(0, 0, Z_p\sigma)^t$, ie.

$$\mu_C = Z_p\sigma. \tag{3.4.1}$$

According to the formula (3.3.6),

$$\delta = \frac{1}{\sigma}\left[(\frac{\mu_C}{3})^2 + (\frac{\mu_C}{3})^2 + (\frac{2\mu_C}{3})^2\right]^{1/2} = \sqrt{\frac{2}{3}}(\frac{\mu_C}{\sigma}) = \sqrt{\frac{2}{3}}Z_p.$$

Hence,

$$Z_p = \frac{\delta}{\sqrt{\frac{2}{3}}}. \tag{3.4.2}$$

Let take $\delta = 0.4$ as an example. From (3.4.2), we get $Z_p = \frac{0.4}{\sqrt{\frac{2}{3}}} = 0.4899$. Since Z_p is the upper p quantile of a standard normal distribution, we get $p = 0.3121$, which means the third mean is the quantile of order 0.3121 for the first two distributions. To carry out the fixed-sample size, the procedure is similar to the example shown in section (3.1.2), and we get $60.75 < n_f < 63.34$ while power and Tpye I error are fixed to 0.8 and 0.05, respectively. Since n_f is relative large (when $n_f > 50$, say), the approximate method (see section (3.1.1)) can be used. From Table (4.3), we read $\lambda = 9.64$. According to (3.1.3), we get $\Delta = \frac{2}{3}Z_p^2 = 0.16$. From equation (3.1.4), we get $n_f = 9.64/0.16 = 60.25 \approx 61$, and $N_f = 3n_f = 183$. To

simulate the responses of the three treatments, we simply set $\sigma^2 = 1$, and using (3.4.1), we get $\mu_C = Z_p = 0.4899$. Hence, in the simulation, for $\delta = 0.4$, treatment A and B have $N(0,1)$ distribution, and treatment C has $N(0.4899, 1)$ distribution.

For four treatments, we simply assume $\mu_A = \mu_B = \mu_C = 0$, so the mean vector can be written as $\mu = (0, 0, 0, \mu_D)^t$. The alternative hypothesis is to test the statistical difference of the four treatments where first three means are equal and the fourth mean is the quantile of order p for the first three distributions. Then, $\mu_A = \mu_B = \mu_C$, $(\mu_D - \mu_A)/\sigma = Z_p$. Hence, $\mu = (\mu_A, \mu_B, \mu_C, \mu_D)^t$ can also be written as $(0, 0, 0, Z_p\sigma)^t$, ie.

$$\mu_D = Z_p\sigma. \tag{3.4.3}$$

According to the formula (3.3.6),

$$\delta = \frac{1}{\sigma}\left[(\frac{\mu_D}{4})^2 + (\frac{\mu_D}{4})^2 + (\frac{\mu_D}{4})^2 + (\frac{3\mu_D}{4})^2\right]^{1/2} = \sqrt{\frac{3}{4}}(\frac{\mu_D}{\sigma}) = \sqrt{\frac{3}{4}}Z_p.$$

Hence,

$$Z_p = \frac{\delta}{\sqrt{\frac{3}{4}}}. \tag{3.4.4}$$

Let take $\delta = 0.4$ as an example again. From (3.4.4), we get $Z_p = \frac{0.4}{\sqrt{\frac{3}{4}}} = 0.46188$ and $p = 0.3221$, which means the fourth mean is the quantile

Table 3.2: Fixed-sample sizes for different δ values when $\alpha = 0.05$, *power* = 0.8. N_f is the fixed sample size for the total treatments ($N_f = J * n_f$)

δ	N_f-three treatments	N_f-four treatments
0.2	726	1092
0.4	183	276
0.6	84	124
0.8	48	72

of order 0.3221 for the first three distributions. To carry out the fixed-sample size, the procedure is similar to the example shown in section (3.1.2), and we get $n_f > 50$ while power and Tpye I error are fixed to 0.8 and 0.05, respectively. Then, the approximate method (see section (3.1.1)) can be used. From Table (4.3), we read $\lambda = 10.91$. According to (3.1.3), we get $\Delta = \frac{3}{4}Z_p^2 \approx 0.16$. From equation (3.1.4), we get $n_f = 10.91/0.16 = 68.19 \approx 69$, and $N_f = 4n_f = 276$. To simulate the responses of the four treatments, we simply set $\sigma^2 = 1$, and using (3.4.3), we get $\mu_D = Z_p = 0.46188$. Hence, in the simulation, for $\delta = 0.4$, treatment A, B, and C have $N(0,1)$ distribution, and treatment D has $N(0.46188, 1)$ distribution.

Fixing the level of significance α to 0.05 and power to 0.8, the values of fixed-sample size are shown in Table (3.2) for different δ values.

For Pocock and OBF tests, we use $K = 5$ and $K = 10$ to compare different number of interim analyses K. For sequential F-test, we choose k_0, the starting point, approximately one fifth of the maximal stopping point n_0. In each case, we carry out the simulated experiment to evaluate the maximal stopping point n_0 and the average stopping time *avst*. Each case in these simulations is based on 5000 replicates.

3.4.2 Algorithm

This simulation involved generating three and four treatments. For three-treatment comparison ($J = 3$), treatments A, B, and C are normally distributed, in which the mean vector of the three treatments is $\mu = (\mu_A, \mu_B, \mu_C)^t$. We set $\mu = (0, 0, \mu_C)^t$, where μ_C is determined by the value of δ (see equation (3.3.6) or see the example in section (3.4.1)). For four treatments comparison ($J = 4$), treatments A, B, C, and D are normally distributed with mean vector $\mu = (\mu_A, \mu_B, \mu_C, \mu_D)^t = (0, 0, 0, \mu_D)^t$, where μ_D is determined by the value of δ (see equation (3.3.6) or see the example in section (3.4.1)). In the simulation, we simply set $\sigma^2 = 1$.

Group Sequential Analysis

1. Set $K = 5$ or $K = 10$, and the initial n_0 which is equal to the fixed-

sample size n_f. n_f is used as the starting point for the calculation of n_0.

2. Calculate group size $g = \frac{n_0}{K}$, if $\frac{n_0}{K}$ is not integer, then set g to the smallest integer not less than the corresponding $\frac{n_0}{K}$.

3. For $J = 3$, generate g patients for each treatment at each stage, and treatments A and B are distributed with $N(0, 1)$, and treatment C has distribution $N(\mu_C, 1)$. For $J = 4$, generate g patients for each treatment at each stage, and treatments A, B, and C have $N(0, 1)$ distribution, and treatment D has $N(\mu_D, 1)$ distribution.

4. Calculate s_k^2 and F_k according to the equations (3.2.3) and (3.2.2).

5. Compare F_k with c_k according to the equation (3.2.4) when $k \leq K$. If $F_k \geq c_k$, reject H_0; otherwise, continue to include another g patients according to step 2, and redo steps 3, 4 and 5.

6. Calculate the simulated power based on 5000 replicates. If the power is far below the required power 0.8, increase n_0, and find the maximal value of n_0 until the simulated the power is just below 0.8. Calculate *avst*. Record the values of the final n_0, power and *avst* as $n1$, $pw1$ and $En1$. Continue to increase n_0, and find the minimal value of n_0 until the simulated power is just over 0.8. Calculate *avst*. Record the values of the final n_0, power and *avst* as $n2$, $pw2$ and $En2$.

7. Calculate n_0 and $avst$ using the linear interpolation method. So, $n_0 = (0.8 - pw2)(n2 - n1)/(pw2 - pw1) + n2$ and $avst = (0.8 - pw2)(En2 - En1)/(pw2 - pw1) + En2$.

8. Simulate $\hat{\alpha}$. The empirical size $\hat{\alpha}$ is simulated using the maximal sample size n_0 which was calculated in step 7. Redo steps 2, 3, 4, and 5. But, in step 3, the distribution for the treatment C should be changed to treatment C $\sim N(0,1)$ for $J = 3$, or the distribution for treatment D should be changed to $N(0,1)$ for $J = 4$. Then calculate the $\hat{\alpha}$.

Sequential F-test

1. Set k_0 a little bit more than one fifth of n_f and the initial n_0 which is equal to the fixed-sample size n_f. Using n_f as the starting point for the calculation of n_0.

2. Calculate the constant a according to the equation (3.3.7), (3.3.8), and (3.3.10).

3. Generate k_0 patients for each treatment, and treatments A and B have $N(0,1)$ distribution, and treatment C has $N(\mu_C, 1)$ distribution for $J = 3$. Or, generate k_0 patients for each treatment, and treatments A, B, and C have $N(0,1)$ distribution, and treatment D has $N(\mu_D, 1)$ distribution for $J = 4$.

4. Calculate L_k according to the equation (3.3.1).

5. Compare L_k with a when $k \leq n_0$. If $L_k \geq a$, reject H_0; otherwise, continue to generate another patient for each treatment, and redo steps 4 and 5.

6. Same procedures as steps 6, 7,and 8 in group sequential analysis.

Rao tests

1. Set initial n_0 which is equal to the fixed-sample size n_f.

2. Generate one patient for each treatment, and treatments A and B have $N(0,1)$ distribution, and treatment C has $N(\mu_C, 1)$ distribution for $J = 3$. Or, generate one patient for each treatment, and treatments A, B, and C have $N(0,1)$ distribution, and treatment D has $N(\mu_D, 1)$ distribution for $J = 4$.

3. Calculate $\hat{\sigma}_k^2$ and R_k according to the equations (3.3.18 and 3.3.20) and (3.3.17 and 3.3.19) for $J = 3$ and $J = 4$, respectivly.

4. Compute $T_1(k)$ *or* $T_2(k)$ according to the equation (3.3.14) and (3.3.15). Compare $T_1(k)$ *or* $T_2(k)$ with the critical values (Rao test 1: see Table (4.5), Rao Test 2: solving the critical value according to the equation (3.3.16)) when $k \leq n_0$. If $T_i(k) \geq c_i(k)$ ($i = 1$ *or* 2, ie. $c_1(k)$ is the critical value for Rao test 1, and $c_2(k)$ is the critical value for Rao test 2.), reject H_0; otherwise, continue to generate another patient for each treatment, and redo steps 3 and 4.

5. Same procedures as steps 6, 7,and 8 in group sequential analysis.

3.4.3 Results

Three-Treatment Comparison.

The simulation results are shown in Table (3.3). From Table (3.3), we can compare the five different methods based on the maximal sample size N_0, the ratio of N_0 and N_f (RT), and the average stopping time $AVST$ under the same power (0.8) and the level of significance ($\alpha = 0.05$). $AVST_x$ denotes the average stopping time for all treatments added using method x.

- First, let us look at the group sequential analysis, Pocock and OBF tests. Since $RT_{OBF} < RT_{Pocock}$, the maximal stopping point n_0 of OBF test is smaller than that of Pocock test. For given the same δ, RT of $K = 5$ is smaller than RT of $K = 10$ for both OBF and Pocock tests. When fixing K, for large δ, $AVST_{OBF}$ is similar with $AVST_{Pocock}$, and for small δ, $AVST_{OBF}$ is greater than $AVST_{Pocock}$.

- For fully sequential analysis, $RT_{Rao1} < RT_{SeqF} < RT_{Rao2}$, but $AVST_{SeqF} < AVST_{Rao1} < AVST_{Rao2}$.

- For the five methods, OBF test has the smallest RT, which means OBF test needs the smallest maximum sample size, but relatively high $AVST$ especially for small δ. Sequential F-test has the smallest

Table 3.3: Simulation results for comparison of normally distributed three-treatment. Set $\alpha = 0.05, 1-\beta = 0.8$. K is the number of interim analyses, N_0 is the maximal stopping point, and AVST is the average stopping time for all treatments. $K_0 = 3 * k_0$

K	$\hat{\alpha}$	δ	$AVST$	N_0	$RT = \frac{N_0}{N_f}$	K_0
OBF						
5	0.050	0.2	606.7	734.8	1.012	
10	0.051	0.2	588.5	746.0	1.028	
5	0.052	0.4	154.3	185.8	1.015	
10	0.054	0.4	148.6	187.4	1.024	
5	0.059	0.6	70.6	84.9	1.011	
10	0.050	0.6	68.1	85.9	1.023	
5	0.051	0.8	41.5	49.8	1.038	
Pocock						
5	0.049	0.2	592.8	886.5	1.221	
10	0.055	0.2	581.6	930.4	1.282	
5	0.050	0.4	150.4	225.1	1.230	
10	0.049	0.4	150.5	240.0	1.306	
5	0.054	0.6	69.1	102.3	1.218	
10	0.057	0.6	69.0	109.8	1.307	
5	0.053	0.8	40.9	59.3	1.24	
Sequential F-test						
n_0	0.056	0.2	571.9	963.0	1.326	180
n_0	0.045	0.4	146.3	246.0	1.344	45
n_0	0.042	0.6	70.3	118.7	1.413	21
n_0	0.036	0.8	41.7	69.6	1.451	15
Rao Test 1						
n_0	0.045	0.2	591.5	782.6	1.078	
n_0	0.039	0.4	152.8	200.8	1.097	
n_0	0.033	0.6	71.7	93.6	1.114	
n_0	0.035	0.8	43.4	56.0	1.167	
Rao Test 2						
n_0	0.022	0.2	754.1	1237.2	1.704	
n_0	0.018	0.4	194.0	311.8	1.704	
n_0	0.014	0.6	91.0	141.2	1.681	
n_0	0.011	0.8	54.5	82.5	1.719	

$AVST$ especially for small δ. For large δ, $AVST$ of OBF, Pocock, and Sequential F-test are similar. Rao test 2 has the highest RT and $AVST$, and we note that $AVST$ of Rao test 2 exceeds the fixed-sample size N_f.

- Values of $\hat{\alpha}$ of Rao test 2 are far less than 0.05.

Four-Treatment Comparison.

The simulation results are shown in Table (3.4). From Table (3.4), we can compare the five different methods based on the maximal sample size N_0, the ratio of N_0 and N_f, and the average stopping time $AVST$ under the same power (0.8) and the level of significance ($\alpha = 0.05$).

- For group sequential analysis, since $RT_{OBF} < RT_{Pocock}$, the maximal stopping point N_0 of OBF test is smaller than that of Pocock test. For given the same δ, RT of $k = 5$ is smaller than RT of $K = 10$ for both OBF and Pocock tests. For every given δ, $AVST_{OBF}$ is greater than $AVST_{Pocock}$, which means Pocock test tends to stop earlier when comparing with OBF test.

- For fully sequential analysis, $RT_{Rao1} < RT_{SeqF} < RT_{Rao2}$. Rao test 2 has the largest $AVST$ and RT, and $AVST$ of Sequential F-test is similar to $AVST$ of Rao test 1.

- For the five methods, OBF has the smallest RT, which means OBF

Table 3.4: Simulation results for comparison of normally distributed four-treatment. Set $\alpha = 0.05, 1-\beta = 0.8$. K is the number of interim analyses, n_0 is the maximal stopping point , and $AVST$ is the average stopping time for all treatments. $K_0 = 4 * k_0$.

K	$\hat{\alpha}$	δ	$AVST$	N_0	$RT = \frac{N_0}{N_f}$	K_0
OBF						
5	0.054	0.2	936.3	1112	1.018	
10	0.049	0.2	905.2	1125	1.031	
5	0.057	0.4	237.6	281.5	1.020	
10	0.052	0.4	227.2	282.7	1.024	
5	0.051	0.6	108.1	127.7	1.029	
10	0.049	0.6	104.5	130.2	1.050	
5	0.054	0.8	62.7	74.1	1.029	
Pocock						
5	0.053	0.2	888.4	1311.3	1.201	
10	0.050	0.2	881.3	1377.5	1.260	
5	0.048	0.4	222.2	355.5	1.216	
10	0.048	0.4	226.3	352.8	1.278	
5	0.049	0.6	104.4	153.5	1.238	
10	0.047	0.6	104.0	160.0	1.290	
5	0.051	0.8	60.8	88.2	1.225	
Sequential F-test						
n_0	0.045	0.2	890.1	1489.1	1.364	240
n_0	0.044	0.4	233.5	389.5	1.411	60
n_0	0.040	0.6	107.3	180.3	1.454	28
n_0	0.030	0.8	62.9	104.0	1.445	20
Rao Test 1						
n_0	0.051	0.2	881.2	1144.5	1.048	
n_0	0.042	0.4	231.9	300.3	1.088	
n_0	0.039	0.6	107.1	137.4	1.108	
n_0	0.038	0.8	64.1	82.0	1.138	
Rao Test 2						
n_0	0.021	0.2	1122.0	1800.0	1.648	
n_0	0.018	0.4	290.5	462.9	1.677	
n_0	0.013	0.6	133.0	204.5	1.649	
n_0	0.011	0.8	79.5	119.9	1.666	

needs the smallest maximal sample size, but relatively high $AVST$ especially for small δ. Pocock test has the smallest $AVST$ (except when $\delta = 0.2,\ K = 5$), and relatively reasonable maximal sample size. Rao test 1 also has reasonable $AVST$ and RT. For Sequential F-test, the $AVST$ is reasonable but a little bit higher in RT. Rao test 2 has high $AVST$ and RT, and we note that $AVST$ of Rao test 2 exceeds the fixed-sample size N_f.

- Values of $\hat{\alpha}$ of Rao test 2 are far less than 0.05.

3.5 Application to a Three-Treatment Comparison

The treatment of palatal expansion might produce a reduction in nasal resistance. An orthodontic clinical trial is to test the difference on total nasal volume among control, Hyrax expansion (traditional), and bone-anchored expansion groups, in which control group subjects did not start treatment for 12 months from induction and served as an untreated control group. The study has begun at University of Alberta since January 2008. Patients were recruited from the Graduate Orthodontic Clinic patient pool, and specific airway dimension measures were performed at the University of Alberta in Graduate Orthodontic Studies using the Eccovision Acoustic Rhinometer (Hood Laboratories, Pembroke, MA). Nasal airway dimen-

Table 3.5: Data set of total nasal volume (TV)(cm^3) for $n_0 = 50$ and three treatment arms (C-control group, T-traditional group, B–bone-anchored expansion group)

Entry	C	T	B	Entry	C	T	B
1	4.460	7.905	5.287	2	5.253	7.608	2.765
3	2.480	4.580	6.164	4	2.642	8.458	4.124
5	4.497	6.028	6.410	6	5.334	4.963	5.581
7	5.237	6.237	3.522	8	3.359	3.166	9.526
9	5.445	6.653	2.337	10	7.835	5.095	1.303
11	3.761	5.561	6.827	12	5.487	6.874	4.730
13	5.573	6.254	7.344	14	4.343	2.301	6.537
15	3.904	6.836	4.075	16	4.608	5.836	8.811
17	5.126	5.547	8.471	18	5.042	11.504	5.722
19	4.954	4.894	6.399	20	7.332	5.832	0.952
21	5.725	7.579	3.900	22	4.593	4.127	5.744
23	5.310	8.290	6.506	24	2.260	8.440	3.888
25	3.699	7.559	7.572	26	3.836	4.621	4.179
27	0.214	2.979	6.001	28	5.819	4.503	3.305
29	4.565	7.983	0.853	30	4.540	5.651	6.766
31	3.594	4.908	2.144	32	5.903	8.919	5.080
33	4.519	6.001	6.269	34	2.668	8.321	3.568
35	3.859	7.237	5.575	36	5.508	7.696	4.533
37	4.280	8.220	7.433	38	5.532	7.966	4.283
39	5.693	6.925	4.576	40	6.196	7.543	5.383
41	4.570	8.392	6.015	42	4.087	4.258	1.327
43	1.330	6.883	6.211	44	5.962	7.421	6.124
45	3.099	6.965	4.426	46	3.435	8.219	3.172
47	2.381	4.125	5.768	48	5.476	7.315	8.099
49	3.825	5.638	10.807	50	5.159	8.520	5.698

Table 3.6: The statistic, cumulative number of patients for the three arms (CNP), and critical values for each k for OBF and Pocock tests with $K = 10$

k	statistics	CNP	Critical Value–OBF	Critical Value–Pocock
1	5.655870	15	1323.782500	6.883226
2	1.759055	30	31.321578	5.459769
3	1.967216	45	14.121313	5.124035
4	2.204977	60	9.368310	4.974410
5	4.653000	75	7.099825	4.889776
6	5.041698	90	5.749483	4.835353
7	8.120567	105	4.845108	4.797418
8	11.053025	120	4.193484	4.769465
9	14.343509	135	3.699992	4.748012
10	15.852972	150	3.312473	4.731029

sions were measured for all treatment subjects immediately following the expansion. Dimensions were measured for each nostril and the two sides of the nose were combined, providing total nasal volume (TV) (cm^3). The study had planed to recruit 50 patients per group (total 150 subjects), and recruited total of 15 patients so far. Based on this partial data set, Giseon Heo simulated the entire data set, which is for the purpose of an illustration. The actual data set will be analyzed as the more observations are gathered. The data set is shown in Table (3.5). Assume that there were three patients at each time, and they were randomly assigned to one of three treatments. We use five sequential methods which introduced in Chapter 3 to show the procedures of sequential tests and to analyze the data set. We want to test the group difference with type I error $\alpha = 0.05$. For group sequential tests, we set $K = 10$.

Table 3.7: The statistic (stat) for each k for Rao test 1 and Rao test 2 and Sequential F-test.

k	stat–Rao1	stat–Rao2	stat–Seq-F	CNP
1	0.2449490	1.732051	0.000000	3
2	0.4416139	2.208069	0.000000	6
3	0.4628016	1.889380	0.000000	9
4	0.7232810	2.557184	0.000000	12
5	0.8531453	2.697882	0.000000	15
6	0.9060183	2.615449	0.000000	18
7	1.0368113	2.770995	4.778933	21
8	0.8731454	2.182863	2.655801	24
9	0.9296627	2.191236	2.643473	27
10	0.8316709	1.859673	1.837251	30
11	0.8899017	1.897277	1.905772	33
12	1.0195607	2.081170	2.307403	36
13	1.0443467	2.048132	2.219025	39
14	0.8887649	1.679608	1.460147	42
15	1.0753224	1.963261	2.014770	45
16	1.1194413	1.978911	2.042548	48
17	1.1710191	2.008281	2.100805	51
18	1.3962047	2.327008	2.853049	54
19	1.4352800	2.328332	2.848266	57
20	1.3128142	2.075741	2.235620	60
21	1.4469134	2.232636	2.596461	63
22	1.4330218	2.160362	2.420199	66
23	1.5957683	2.352832	2.885264	69
24	1.8932349	2.732649	3.941826	72
25	2.0717430	2.929887	4.558264	75
26	2.1355409	2.961462	4.651853	78
27	2.1889596	2.978797	4.698968	81
28	2.1455903	2.867166	4.325615	84
29	2.2969992	3.016108	4.804246	87
30	2.3682514	3.057399	4.934801	90
31	2.4485812	3.109701	5.105410	93
32	2.6163824	3.270478	5.670103	96
33	2.7121840	3.338468	5.912097	99
34	2.9880378	3.623528	7.027682	102

The conclusion of all five methods of sequential analysis is that we reject the null hypothesis and conclude that there are statistical different nasal volumes among these three groups.

For group sequential analyses, in Table (3.6), we list the statistics, cumulative number of patients for the three arms (CNP), and critical values for each k for OBF and Pocock tests with $K = 10$. From the table, the critical values for OBF test are getting smaller as k increasing. The critical values for Pocock test is not a constant for each k because we use t-distribution and the same significant levels as standard normal distribution to carry out the critical values. As the degree of freedom is getting larger, a t-distribution tends to be a normal distribution; hence, c_k for Pocock test tends to stabilize as the degree of freedom increases. From Table (3.6), the analysis of OBF test stops at $k = 7$, which means 105 of 150 patients have been accrued. For Pocock tests, the analysis stops at $k = 6$, which means 90 of 150 patients have been accrued. The first two graphs in Figure (3.5.1) are the graphs for monitoring the OBF and Pocock statistics for the three treatment arms. It shows the same results as above.

For fully sequential analysis, in Table (3.7), we list k, the statistics, and CNP for Rao test 1, Rao test 2, and Sequential F-test. When fixing $\alpha = 0.05$ and $1 - \beta = 0.8$, the critical values (CV) for Rao test 1, Rao

test 2, and Sequential F-test are 2.695, 3.490, and 5.543, repectively. The table shows that the analyses of Rao test 1, Rao test 2, and Sequential F-test stop at $k = 33$, $k = 34$, and $k = 32$, respectively, ie. the conclusion is made right after 99 of 150, 102 of 150, and 96 of 150 patients have been accrued, respectively. Also, it is easy to get the same results from the last three graphs in Figure (3.5.1).

With type I error is equal to 0.05, using any tests of sequential analysis can stop the trial earlier than n_0 to get the final conclusion; hence, this example shows sequential analysis could stop a trial as soon as one treatment is significantly different with other treatments. Among the five methods, Pocock test is the test with the smallest number of patients accrued, and OBF is the the test with the most number of patients accrued.

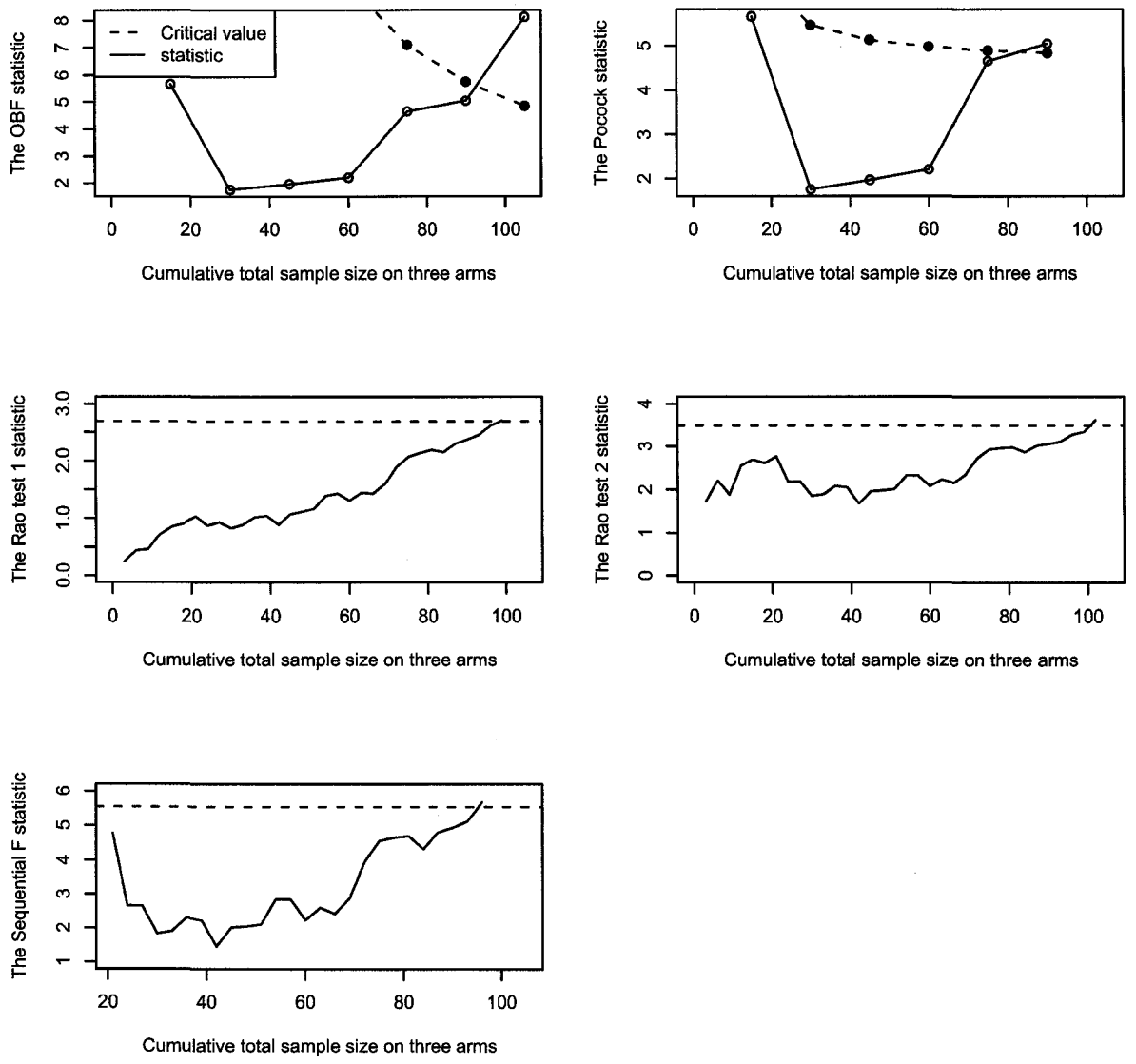

Figure 3.5.1: Monitoring the statistic for the three treatment arms

Chapter 4

Appendix

4.1 Tables

1. Table (4.1): Pocock constants for two-sided tests with overall significance level α
2. Table (4.2): OBF constants for two-sided tests with overall significance level α
3. Table (4.3): λ values satisfying the power condition
4. Table (4.4): Pocock and OBF constants for repeated χ^2-tests of homogeneity of J normal means
5. Table (4.5): Critical values for Rao test 1

Table 4.1: Pocock tests: constants $C_P(K, \alpha)$ for the two-sided tests with K groups of observations and level of significance α

$C_P(K, \alpha)$			
K	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.20$
1	2.576	1.960	1.645
2	2.772	2.178	1.875
3	2.873	2.289	1.992
4	2.939	2.361	2.067
5	2.986	2.413	2.122
6	3.023	2.453	2.164
7	3.053	2.485	2.197
8	3.078	2.512	2.225
9	3.099	2.535	2.249
10	3.117	2.555	2.270
11	3.133	2.572	2.288
12	3.147	2.588	2.304
15	3.182	2.626	2.344
20	3.225	2.672	2.392

Table 4.2: O'Brien & Fleming tests: constants $C_B(K, \alpha)$ for the two-sided tests with K groups of observations and level of significance α

$C_B(K, \alpha)$			
K	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.20$
1	2.576	1.960	1.645
2	2.580	1.977	1.678
3	2.595	2.004	1.710
4	2.609	2.024	1.733
5	2.621	2.040	1.751
6	2.631	2.053	1.765
7	2.640	2.063	1.776
8	2.648	2.072	1.786
9	2.654	2.080	1.794
10	2.660	2.087	1.801
11	2.665	2.092	1.807
12	2.670	2.098	1.813
15	2.681	2.110	1.826
20	2.695	2.126	1.842

Table 4.3: λ values satifying $\chi^2_{J-1}(\chi^2_{\alpha,J-1}|\lambda) = \beta$

	$1-\beta = 0.80$		$1-\beta = 0.90$	
J	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
2	11.68	7.85	14.88	10.51
3	13.89	9.64	17.43	12.66
4	15.46	10.91	19.25	14.18
5	16.75	11.94	20.74	15.41
6	17.87	12.83	22.03	16.47
7	18.88	13.63	23.19	17.42
8	19.79	14.36	24.24	18.29
9	20.64	15.03	25.22	19.09
10	21.43	15.65	26.13	19.83
11	22.18	16.25	26.99	20.54
12	22.89	16.81	27.80	21.20
13	23.57	17.34	28.58	21.84
14	24.22	17.85	29.32	22.44
15	24.84	18.34	30.04	23.03
16	25.44	18.82	30.73	23.59
17	26.02	19.27	31.39	24.13
18	26.58	19.71	32.04	24.65
19	27.12	20.14	32.66	25.16
20	27.65	20.56	33.27	25.66

Table 4.4: Constants $C_P(p, K, \alpha)$ and $C_B(p, K, \alpha)$ for respectively, Pocock-type and O'Brien & Fleming-type repeated χ^2-tests of homogeneity of J normal means. Tests have K analyses, the χ^2 statistic at each analysis has $p = J - 1$ degrees of freedom and the the level of significance is $\alpha = 0.05$.

J	p	K	$C_P(p, K, \alpha)$	$C_B(p, K, \alpha)$
2	1	1	3.84	3.84
		2	4.74	3.91
		3	5.24	4.02
		4	5.58	4.10
		5	5.82	4.16
		6	6.02	4.21
		10	6.53	4.35
3	2	1	5.99	5.99
		2	7.08	6.02
		3	7.67	6.12
		4	8.06	6.20
		5	8.35	6.27
		6	8.58	6.33
		10	9.17	6.48
4	3	1	7.81	7.81
		2	9.04	7.83
		3	9.69	7.92
		4	10.13	7.99
		5	10.44	8.06
		6	10.69	8.11
		10	11.34	8.26
5	4	1	9.49	9.49
		2	10.82	9.50
		3	11.53	9.57
		4	12.00	9.64
		5	12.35	9.71
		6	12.62	9.77
		10	13.32	9.93

Table 4.5: Critical values $C_1(\alpha, d)$ for Rao test 1 of different values of d and different levels of significance α

	α		
d	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
2	2.419	2.695	3.242
3	2.751	3.023	3.562
4	3.023	3.294	3.827
5	3.260	3.530	4.059
6	3.474	3.743	4.269
7	3.669	3.938	4.461
8	3.851	4.119	4.640
10	4.183	4.450	4.968
12	4.482	4.748	5.264

4.2 Reference

1. Burden, R.L.& Faires, J.D.(2001). Numerical Analysis, 7th Edition, *Brooks/Cole.*

2. Chow, S.C., Shao, J., Wang, H.(2003). Sample Size Calculations in Clinical Research, *MarcelDekker.*

3. Csörgő, M. and Révész, P. (1981). Strong Applications in Probability and Statistics. *Acadamic Press, New York.*

4. Gombay, E. (2002). Parametric Sequential Tests in the Presence of Nuisance Parameters. *Theory stoch, Processes No* $1-2$, $107-118$.

5. Gombay, E. and Hussein, A.(2006). A Class of Sequential Tests for

Two-sample Composite Hypotheses, *The Canadian Journal of Statistics*, *Vol.*31.

6. Gombay, E. and Serban, D. (2007), Sequential Comparison of d Populations and Related Tables. *Communications in Statistics – Simulation and Computation* 36 : 55 – 72.

7. Guenther, W.C. (1977). Power and Sample Size Determination When the Alternative Hypotheses Are Given in Quantiles, *The American Statistician*, *Vol.* 31, *No.* 3.

8. Jennison, C & Turnbull B.W. (1991). Exact Calculations for Sequential t, χ^2 and F tests, *Biometrika*, 78, 133 – 141.

9. Jennison, C. & Turnbull, B.W. (2000). Group Sequential Methods with Applications to Clinical Trials, *Chapman & Hall/CRC*.

10. Lan, K.K.G and DeMets, D.L. (1983). Discrete Sequential Boundaries for Clinical trials. *Bimetrika*, 70, 659 – 663.

11. Lehmer, E. (1944). Inverse Table of Probabilities of Errors of the Second Kind, *Annals of Mathematical Statistics*, 15, 388 – 98.

12. Pocock, S. J. (1977). Group Sequential Methods in the Design and Analysis of Clinical Trials, *Biometrika* 64, 191 – 199.

13. Odeh, R.E., and Fox, M. (1975). Sample Size Choice, *New York : Marcel Dekker, Inc.*.

14. O'Brien, P.C. and Flemming, T.R.(1979). A Multiple Testing Procedure for Clinical Trials, *Biometric* 35, 549 – 556.

15. Siegmund, D. (1980). Sequential and F Tests and the Related Confidence Intervals, *Biometrika*, 67, 389 – 402.

16. Vostrikova,L.Y. (1981) Detection of a "Disorder" in a Wiener Process, *Theory of Probability and Its Applications Vol. XXVI* (1981) 356–362.

4.3 Glossary

$avst$	the average stopping time for each treatment
$AVST$	the average stopping time for all treatments
cdf	cumulative distribution function
J	number of treatment arms
K	number of groups (analyses) in a group sequential procedure
n_f	fixed-sample size for each treatment
N_f	fixed-sample size for all treatments
n_0	the maximal sample size for each treatment
N_0	the maximal sample size for all treatments
OBF	O'Brien and Fleming
RT	ratio of N_0 and N_f
X^t	transpose of vector X

$t_{\nu,\alpha}$	upper α quantile of a t-distribution with ν degrees of freedom
Z_p	upper p quantile of a standard normal distribution
α	Type I error probability
β	Type II error probability
Φ	standard normal cumulative distribution function
$\chi^2_{\alpha,\nu}$	upper α quantile of a χ^2 distribution with ν degrees of freedom
$\chi^2_\nu(\cdot\|\lambda)$	non-central cumulative χ^2 distribution with ν degrees of freedom

4.4 Code

1. Two-treatment Comparison

- **Simulate the maximum sample size or truncation point and average stopping point**

```
######### Group sizes are not equal, set lan=0.5############

#OBF

alpha=0.05;sigma=1;N=10000;lan=0.5; power=0.9

#1
 K=20;Cb=2.162;delta=0.1;uA=0;uB=0.1
#2
 K=5;Cb=2.040;delta=0.1;uA=0;uB=0.1
#3
 K=20;Cb=2.162;delta=0.2;uA=0;uB=0.2
#4
 K=5;Cb=2.040;delta=0.2;uA=0;uB=0.2
#5
 K=20;Cb=2.162;delta=0.5;uA=0;uB=0.5
#6
 K=5;Cb=2.040;delta=0.5;uA=0;uB=0.5
#7
 K=5;Cb=2.040;delta=0.9;uA=0;uB=0.9

nf=ceiling((qnorm(1-alpha/2)+qnorm(1-beta))^2*2*sigma^2/delta^2)*2
n0=nf
nts=ceiling(n0/K)  # treatment size
n0=nts*K

y=rep(0,N)
pw=0.8
j=0
while(pw<0.9){
```

```
  n0=n0+K*j
  nts=n0/K
  for (i in 1:N){
     AA=rnorm(n0*(K+1),uA,1);BB=rnorm(n0*(K+1),uB,1)
     count=1
     n=rbinom(1,nts,lan)
     m=nts-n
     A=AA[1:n]
     B=BB[1:m]
     s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2)
     T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n)))
   while(T<qt(1-pnorm(Cb*sqrt(K/count)),n+m-2,lower.tail = F) & count<=K){
     (nplus=rbinom(1,nts,lan))
     (n=n+nplus)
     (m=m+(nts-nplus))
     (A=AA[1:n])
     (B=BB[1:m])
     (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2))
     (T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n))))
     (count=count+1)
     }
 y[i]=count
 }
pw=1-table(y,exclude=1:K)/N
n1=nts*K
j=j+1
}
pw1=pw

q=rep(0,N)
for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En1=mean(q)*nts

y=rep(0,N)
j=1
while(pw>0.9){
  n0=n1-K*j
```

```
  nts=n0/K
  for (i in 1:N){
     AA=rnorm(n0*(K+1),uA,1);BB=rnorm(n0*(K+1),uB,1)
     count=1
     n=rbinom(1,nts,lan)
     m=nts-n
     A=AA[1:n]
     B=BB[1:m]
     s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2)
     T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n)))
   while(T<qt(1-pnorm(Cb*sqrt(K/count)),n+m-2,lower.tail = F) & count<=K){
     (nplus=rbinom(1,nts,lan))
     (n=n+nplus)
     (m=m+(nts-nplus))
     (A=AA[1:n])
     (B=BB[1:m])
     (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2))
     (T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n))))
     (count=count+1)
     }
 y[i]=count
 }
 pw=1-table(y,exclude=1:K)/N
 j=j+1
}
pw0=pw

for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En0=mean(q)*nts

nn0=(power-pw0)*(n1-n0)/(pw1-pw0)+n0 #the maximum sample size or
truncation point
En=(En1-En0)*(power-pw0)/(pw1-pw0)+En0  # average stopping point
nn0
En
```

#Pocock

```
alpha=0.05;sigma=1;N=10000;lan=0.5; power=0.9

#1
 K=20;Cp=2.672;delta=0.1;uA=0;uB=0.1
#2
  K=5;Cp=2.413;delta=0.1;uA=0;uB=0.1
#3
  K=20;Cp=2.672;delta=0.2;uA=0;uB=0.2
#4
  K=5;Cp=2.413;delta=0.2;uA=0;uB=0.2
#5
  K=20;Cp=2.672;delta=0.5;uA=0;uB=0.5
#6
  K=5;Cp=2.413;delta=0.5;uA=0;uB=0.5
#7
  K=5;Cp=2.413;delta=0.9;uA=0;uB=0.9

nf=ceiling((qnorm(1-alpha/2)+qnorm(1-beta))^2*2*sigma^2/delta^2)*2
n0=nf
nts=ceiling(n0/K)   # treatment size
n0=nts*K

y=rep(0,N)
pw=0.8
j=0
while(pw<0.9){
   n0=n0+K*j
   nts=n0/K
   for (i in 1:N){
      AA=rnorm(n0*(K+1),uA,1);BB=rnorm(n0*(K+1),uB,1)
      count=1
      n=rbinom(1,nts,lan)
      m=nts-n
      A=AA[1:n]
      B=BB[1:m]
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2)
      T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n)))
    while(T<qt(1-pnorm(Cp),n+m-2,lower.tail = F) & count<=K){
      (nplus=rbinom(1,nts,lan))
```

```
      (n=n+nplus)
      (m=m+(nts-nplus))
      (A=AA[1:n])
      (B=BB[1:m])
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2))
      (T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n))))
      (count=count+1)
      }
 y[i]=count
 }
pw=1-table(y,exclude=1:K)/N
n1=nts*K
j=j+1
}
pw1=pw

q=rep(0,N)
for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En1=mean(q)*nts

q=rep(0,N)
y=rep(0,N)
j=1
while(pw>0.9){
   n0=n1-K*j
   nts=n0/K
   for (i in 1:N){
      AA=rnorm(n0*(K+1),uA,1);BB=rnorm(n0*(K+1),uB,1)
      count=1
      n=rbinom(1,nts,lan)
      m=nts-n
      A=AA[1:n]
      B=BB[1:m]
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2)
      T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n)))
    while(T<qt(1-pnorm(Cp),n+m-2,lower.tail = F) & count<=K){
      (nplus=rbinom(1,nts,lan))
```

```
      (n=n+nplus)
      (m=m+(nts-nplus))
      (A=AA[1:n])
      (B=BB[1:m])
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2))
      (T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n))))
      (count=count+1)
      }
 y[i]=count
 }
 pw=1-table(y,exclude=1:K)/N
 j=j+1
}
pw0=pw

for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En0=mean(q)*nts
En0

nn0=(power-pw0)*(n1-n0)/(pw1-pw0)+n0 #the maximum sample size or
truncation point
En=(En1-En0)*(power-pw0)/(pw1-pw0)+En0  # average stopping point
nn0
En
```

Sequential Rao

```
alpha=0.05;sigma=1;lan=0.5;N=10000; power=0.9
#1
 delta=0.1;uA=0;uB=0.1
#2
 delta=0.2;uA=0;uB=0.2
#3
 delta=0.5;uA=0;uB=0.5
#4
 delta=0.9;uA=0;uB=0.9
```

```
nf=ceiling((qnorm(1-alpha/2)+qnorm(1-beta))^2*2*sigma^2/delta^2)*2
n0=nf

y=rep(0,N)
pw=0.8
j=0
while(pw<0.9){
   n0=nf+j
   for (i in 1:N){
     n=0;m=0
     AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1)
   while(n<=0 || m<=0) {
    if (runif(1, min=0, max=1)<lan) (n=n+1) else (m=m+1)
     }
    k=n+m
     A=AA[1:n]
     B=BB[1:m]
      s2=(sum(A^2)+sum(B^2))/k-((sum(A)+sum(B))/k)^2
      R= 1/k*(n*sum(A)-m*sum(B))^2/(n*m*s2)
      a=k/n0*R
      RR=sqrt(a)
   while(RR<2.24 & k<=n0){
     (if (runif(1, min=0, max=1)<lan) (n=n+1) else (m=m+1))
     (A=AA[1:n])
     (B=BB[1:m])
     (k=n+m)
     (s2=(sum(A^2)+sum(B^2))/k-((sum(A)+sum(B))/k)^2)
     (R= 1/k*(n*sum(A)-m*sum(B))^2/(n*m*s2))
     (a=k/n0*R)
     (RR=sqrt(a))
     }
 y[i]=k
}
pw=1-table(y,exclude=1:n0)/N
n1=n0
j=j+10  # when delta=0.9 j=j+2, delta=0.1-->j=j+50 delta=0.5-->j=j+10
}
pw1=pw

q=rep(0,N)
```

```
for (i in 1:N){
if (y[i]>n0) (q[i]=y[i]-1)
 else if (y[i]<=n0) (q[i]=y[i])
  }
En1=mean(q)

j=2 # when delta=0.9-->j=1 delta=0.1,j=0.2--> j=10 delta=0.5-->j=2
while(pw>0.9){
  n0=n1-j
  for (i in 1:N){
     n=0;m=0
     AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1)
   while(n<=0 || m<=0) {
    if (runif(1, min=0, max=1)<lan) (n=n+1) else (m=m+1)
     }
   k=n+m
    A=AA[1:n]
    B=BB[1:m]
     s2=(sum(A^2)+sum(B^2))/k-((sum(A)+sum(B))/k)^2
     R= 1/k*(n*sum(A)-m*sum(B))^2/(n*m*s2)
     a=k/n0*R
     RR=sqrt(a)
   while(RR<2.24 & k<=n0){
     (if (runif(1, min=0, max=1)<lan) (n=n+1) else (m=m+1))
     (A=AA[1:n])
     (B=BB[1:m])
     (k=n+m)
     (s2=(sum(A^2)+sum(B^2))/k-((sum(A)+sum(B))/k)^2)
     (R= 1/k*(n*sum(A)-m*sum(B))^2/(n*m*s2))
     (a=k/n0*R)
     (RR=sqrt(a))
     }
 y[i]=k
}
pw=1-table(y,exclude=1:n0)/N
n1=n0
j=j+2  # when delta=0.1 delta=0.2--> j=j+10 delta=0.9-->j=j+1 delta=0.5-->j=j+2
}
pw0=pw
```

```
q=rep(0,N)
for (i in 1:N){
if (y[i]>n0) (q[i]=y[i]-1)
 else if (y[i]<=n0) (q[i]=y[i])
  }
En0=mean(q)

nn0=(power-pw0)*(n1-n0)/(pw1-pw0)+n0 #the maximum sample size or
truncation point
En=(En1-En0)*(power-pw0)/(pw1-pw0)+En0  # average stopping point
nn0
En
```

- **Simulate the empirical size at the given maximal sample size**

```
############ Calculate type I error ##########
#OBF
alpha=0.05;beta=0.1;sigma=1;N=10000;lan=0.5

#1
 K=20;Cb=2.162;delta=0.1;uA=0;uB=0;n0=4513
#2
  K=5;Cb=2.040;delta=0.1;uA=0;uB=0;n0=4250
#3
  K=20;Cb=2.162;delta=0.2;uA=0;uB=0;n0=1138.2
#4
  K=5;Cb=2.040;delta=0.2;uA=0;uB=0;n0=1087.5
#5
  K=20;Cb=2.162;delta=0.5;uA=0;uB=0;n0=185
#6
  K=5;Cb=2.040;delta=0.5;uA=0;uB=0;n0=175
#7
  K=5;Cb=2.040;delta=0.9;uA=0;uB=0;n0=56

nts=ceiling(n0/K)
y=rep(0,N)
 for (i in 1:N){
      AA=rnorm(n0*(K+1),uA,1);BB=rnorm(n0*(K+1),uB,1)
      count=1
      n=rbinom(1,nts,lan)
```

```
      m=nts-n
      A=AA[1:n]
      B=BB[1:m]
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2)
      T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n)))
    while(T<qt(1-pnorm(Cb*sqrt(K/count)),n+m-2,lower.tail = F) &
 count<=K){
      (nplus=rbinom(1,nts,lan))
      (n=n+nplus)
      (m=m+(nts-nplus))
      (A=AA[1:n])
      (B=BB[1:m])
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2))
      (T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n))))
      (count=count+1)
      }
  y[i]=count
  }
typeI=1-table(y,exclude=1:K)/N
typeI

#Pocock
alpha=0.05;beta=0.1;sigma=1;N=10000;lan=0.5

#1
  K=20;Cp=2.672;delta=0.1;uA=0;uB=0;n0=5633.5
#2
  K=5;Cp=2.413;delta=0.1;uA=0;uB=0;n0=5124.9
#3
  K=20;Cp=2.672;delta=0.2;uA=0;uB=0;n0=1392
#4
  K=5;Cp=2.413;delta=0.2;uA=0;uB=0;n0=1276.6
#5
  K=20;Cp=2.672;delta=0.5;uA=0;uB=0;n0=228.8
#6
  K=5;Cp=2.413;delta=0.5;uA=0;uB=0;n0=208
#7
  K=5;Cp=2.413;delta=0.9;uA=0;uB=0;n0=67

   y=rep(0,N)
```

```
  nts=ceiling(n0/K)
  for (i in 1:N){
     AA=rnorm(n0*(K+1),uA,1);BB=rnorm(n0*(K+1),uB,1)
     count=1
     n=rbinom(1,nts,lan)
     m=nts-n
     A=AA[1:n]
     B=BB[1:m]
     s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2)
     T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n)))
   while(T<qt(1-pnorm(Cp),n+m-2,lower.tail = F) & count<=K){
     (nplus=rbinom(1,nts,lan))
     (n=n+nplus)
     (m=m+(nts-nplus))
     (A=AA[1:n])
     (B=BB[1:m])
     (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2))/(n+m-2))
     (T=abs((mean(A)-mean(B))/sqrt((m+n)*s2/(m*n))))
     (count=count+1)
     }
 y[i]=count
 }
typeI=1-table(y,exclude=1:K)/N
typeI
```

#Sequential Rao

```
alpha=0.05;beta=0.1;sigma=1;lan=0.5;N=10000
#1
 delta=0.1;uA=0;uB=0;n0=4465.6
#2
 delta=0.2;uA=0;uB=0;n0=1127.3
#3
 delta=0.5;uA=0;uB=0;n0=188.1
#4
 delta=0.9;uA=0;uB=0;n0=67.1

y=rep(0,N)
for (i in 1:N){
     n=0;m=0
```

```
    AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1)
  while(n<=0 || m<=0) {
   if (runif(1, min=0, max=1)<lan) (n=n+1) else (m=m+1)
    }
  k=n+m
   A=AA[1:n]
   B=BB[1:m]
    s2=(sum(A^2)+sum(B^2))/k-((sum(A)+sum(B))/k)^2
    R= 1/k*(n*sum(A)-m*sum(B))^2/(n*m*s2)
    a=k/n0*R
    RR=sqrt(a)
  while(RR<2.24 & k<=n0){
    (if (runif(1, min=0, max=1)<lan) (n=n+1) else (m=m+1))
    (A=AA[1:n])
    (B=BB[1:m])
    (k=n+m)
    (s2=(sum(A^2)+sum(B^2))/k-((sum(A)+sum(B))/k)^2)
    (R= 1/k*(n*sum(A)-m*sum(B))^2/(n*m*s2))
    (a=k/n0*R)
    (RR=sqrt(a))

    }
 y[i]=k
}
typeI=1-table(y,exclude=1:n0)/N
typeI
```

- **Calculate CV(alpha) for Sequential Rao**

```
#################### CV(alpha)  for Sequential Rao ###############

n=5;m=6
k=c(0:n);q=c(0:m)
r=seq(from=1.5, to=3, by=0.001)
N=length(r)
z=matrix(rep(0,N*4),ncol=4,byrow=T,dimnames = list(c(1:N),c("CV",
"k=5","k=6","diff")))
for (i in 1:N) {
 x=r[i]
```

```
 y=4/pi*sum((-1)^k/(2*k+1)*exp(-pi^2*(2*k+1)^2/(8*x^2)))
 p=4/pi*sum((-1)^q/(2*q+1)*exp(-pi^2*(2*q+1)^2/(8*x^2)))
 d=p-y
 z[i,]=c(x,y,p,d)
}
z
```

2. Three-treatment Comparison

- **Simulate the maximum sample size or truncation point and average stopping point**

```
alpha=0.05; sigma=1;N=5000; power=0.8;J=3

#OBF

#1)delta=0.8;
K=5;Cb=6.27;uA=0;uB=0; uC=0.9798;nf=48  # nf is sample size for all treatment
K=10;Cb=6.48;uA=0;uB=0; uC=0.9798;nf=48

#2)delta=0.6
K=5;Cb=6.27;uA=0;uB=0; uC=0.7348;nf=84
K=10;Cb=6.48;uA=0;uB=0; uC=0.7348;nf=84
K=1;Cb=5.99;uA=0;uB=0; uC=0.7348;nf=84



#3)delta=0.4
K=5;Cb=6.27;uA=0;uB=0; uC=0.4899;nf=183
K=10;Cb=6.48;uA=0;uB=0; uC=0.4899;nf=183

#4)delta=0.2
K=5;Cb=6.27;uA=0;uB=0; uC=0.2449;nf=726
K=10;Cb=6.48;uA=0;uB=0; uC=0.2449;nf=726

#### Find n1 and En1 (average stopping time) to obtain the maximal value of
power ####which is less than 0.8
n0=nf
nts=ceiling(n0/(K*J))   # group size
n0=nts*K*J
```

```
q=rep(0,N)
y=rep(0,N)
   for (i in 1:N){
      AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1);CC=rnorm(n0,uC,1)
      count=1
      m=1
      n=nts
      A=AA[1:nts]
      B=BB[1:nts]
      C=CC[1:nts]
      xbar=mean(c(A,B,C))
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(nts-1))
      F=(nts/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2)
    while(F<qf(1-pchisq(Cb*(K/count),J-1),J-1,J*(n-1),lower.tail = FALSE) &
count<=K){
      (m=m+1)
      (n=nts*m)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
      (count=count+1)
      }
 y[i]=count
 }
pw=1-table(y,exclude=1:K)/N
pw1=pw

q=rep(0,N)
for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En1=mean(q)*nts*J
```

```
#### Find n0 and En0 (average stopping time) to obtain the minimal value of
power ####which is greater than 0.8
n0 #set the value of n0 greater than n1
q=rep(0,N)
y=rep(0,N)
j=1

   nts=n1/(K*J)
   for (i in 1:N){
       AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1);CC=rnorm(n0,uC,1)
      count=1
      m=1
      n=nts
      A=AA[1:nts]
      B=BB[1:nts]
      C=CC[1:nts]
      xbar=mean(c(A,B,C))
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(nts-1))
      F=(nts/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2)
    while(F<qf(1-pchisq(Cb*(K/count),J-1),J-1,J*(n-1),lower.tail = FALSE) &
count<=K){
      (m=m+1)
      (n=nts*m)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
      (count=count+1)
      }
  y[i]=count
  }
 pw=1-table(y,exclude=1:K)/N
pw0=pw

for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
```

```
else if (y[i]<=K) (q[i]=y[i])
   }
En0=mean(q)*nts*J

nn0=(power-pw0)*(n1-n0)/(pw1-pw0)+n0 #the maximum sample size or
truncation point
En=(En1-En0)*(power-pw0)/(pw1-pw0)+En0
```

#Pocock

```
#1)delta=0.8;
K=5;Cp=8.35;uA=0;uB=0; uC=0.9798;nf=48  # nf is sample size for all treatment
K=10;Cp=9.17;uA=0;uB=0; uC=0.9798;nf=48

#2)delta=0.6
K=5;Cp=8.35;uA=0;uB=0; uC=0.7348;nf=84
K=10;Cp=9.17;uA=0;uB=0; uC=0.7348;nf=84

#3)delta=0.4
K=5;Cp=8.35;uA=0;uB=0; uC=0.4899;nf=183
K=10;Cp=9.17;uA=0;uB=0; uC=0.4899;nf=183

#4)delta=0.2
K=5;Cp=8.35;uA=0;uB=0; uC=0.2449;nf=726
K=10;Cp=9.17;uA=0;uB=0; uC=0.2449;nf=726

#### Find n1 and En1 (average stopping time) to obtain the maximal value of
power ####which is less than 0.8

n0=nf
nts=ceiling(n0/(K*J))   # group size
n0=nts*K*J

q=rep(0,N)
y=rep(0,N)
   for (i in 1:N){
      AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1);CC=rnorm(n0,uC,1)
      count=1
      m=1
      n=nts
```

```
      A=AA[1:nts]
      B=BB[1:nts]
      C=CC[1:nts]
      xbar=mean(c(A,B,C))
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(nts-1))
      F=(nts/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2)
    while(F<qf(1-pchisq(Cp,J-1),J-1,J*(n-1),lower.tail = FALSE) & count<=K){
      (m=m+1)
      (n=nts*m)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
      (count=count+1)
      }
  y[i]=count
  }
 pw=1-table(y,exclude=1:K)/N
 n1=nts*K*J
pw1=pw

q=rep(0,N)
for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En1=mean(q)*nts*J

#### Find n0 and En0 (average stopping time) to obtain the minimal value of
power ####which is greater than 0.8

n0 #set the value n0 greater than n1
q=rep(0,N)
y=rep(0,N)

   n0=n1-K*j*J
```

```
  nts=n0/(K*J)
  for (i in 1:N){
     AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1);CC=rnorm(n0,uC,1)
     count=1
     m=1
     n=nts
     A=AA[1:nts]
     B=BB[1:nts]
     C=CC[1:nts]
     xbar=mean(c(A,B,C))
     s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(nts-1))
     F=(nts/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2)
   while(F<qf(1-pchisq(Cp,J-1),J-1,J*(n-1),lower.tail = FALSE) & count<=K){
     (m=m+1)
     (n=nts*m)
     (A=AA[1:n])
     (B=BB[1:n])
     (C=CC[1:n])
     (xbar=mean(c(A,B,C)))
     (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
     (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
     (count=count+1)
     }
 y[i]=count
 }
pw=1-table(y,exclude=1:K)/N
pw0=pw

for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
  }
En0=mean(q)*nts*J

nn0=(power-pw0)*(n1-n0)/(pw1-pw0)+n0 #the maximum sample size or
truncation point
En=(En1-En0)*(power-pw0)/(pw1-pw0)+En0
```

Sequential F-test

```
############ Seq F ############
# calculating "a" for giving d (#of treatment) ,k0,n0

integrant<-function(x) {x^(d-2)*exp(-0.583*x*(1+x^2/d)^(-
1))*sqrt(1+x^2/d)*(log(1+x^2/d))^(0.5-d/2)}

d ; k0  ;   n0   # set d, k0 and n0 to some value

p1<-function(a) {1-pf((exp(2*a/(k0*d))-1)*d*(k0-1)/(d-1),d-1,d*(k0-1))}
p2<-function(a) {2*exp(-a)*(a/d)^(d/2-0.5)*(gamma(d/2-0.5))^(-
1)*integrate(integrant,lower=sqrt(d*(exp(2*a/(n0*d))-
1)),upper=sqrt(d*(exp(2*a/(k0*d))-1)))$value}

r=seq(from=5,to=6,by=0.001)
m=length(r)
z=matrix(rep(0,m*2),ncol=2,byrow=T)
for (i in 1:m) {
 a=r[i]
 p=p1(a)+p2(a)
 z[i,]=c(a,p)
}
z

# simulation

alpha=0.05;beta=0.2;N=5000; power=0.8; J=3

#1)delta=0.8;
uA=0;uB=0; uC=0.9798;nf=48  # nf is sample size for all treatments

#2)delta=0.6
uA=0;uB=0; uC=0.7348;nf=84

#3)delta=0.4
uA=0;uB=0; uC=0.4899;nf=183

#4)delta=0.2
uA=0;uB=0; uC=0.2449;nf=726
```

```
n0=ceiling(n0*3)/3  # n0 is sample size for each treatment

#### Find n0 and En0 (average stopping time) to obtain the minimal value of
power ####which is greater than 0.8

k0=
n0=
a=        # set k0 and n0 to what we want, and a is calculated based on k0 and n0.

N=5000
q=rep(0,N)
y=rep(0,N)

for (i in 1:N){
      AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
      n=k0
      A=AA[1:n]
      B=BB[1:n]
      C=CC[1:n]
      ma=mean(A)
      mb=mean(B)
      mc=mean(C)
      m=mean(c(ma,mb,mc))
      Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2)))
      while(Lk<=a & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (ma=mean(A))
      (mb=mean(B))
      (mc=mean(C))
      (m=mean(c(ma,mb,mc)))
      (Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2))))
      }
 y[i]=n
 }
```

```
pw=1-table(y,exclude=1:n0)/N;pw
pw1=pw

q=rep(0,N)
for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
  }
En1=mean(q)

#### Find n0 and En0 (average stopping time) to obtain the minimal value of
power ####which is greater than 0.8
n0= # set n0 greater than n1

for (i in 1:N){
     AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
     n=k0
     A=AA[1:n]
     B=BB[1:n]
     C=CC[1:n]
     ma=mean(A)
     mb=mean(B)
     mc=mean(C)
     m=mean(c(ma,mb,mc))
     Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2)))
     while(Lk<=a & n<=n0){
     (n=n+1)
     (A=AA[1:n])
     (B=BB[1:n])
     (C=CC[1:n])
     (ma=mean(A))
     (mb=mean(B))
     (mc=mean(C))
     (m=mean(c(ma,mb,mc)))
     (Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2))))
     }
 y[i]=n
 }
```

```
pw=1-table(y,exclude=1:n0)/N;pw
pw0=pw

q=rep(0,N)
for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En0=mean(q)

nn0=(power-pw0)*(n1-n0)/(pw1-pw0)+n0 #the maximum sample size or
truncation point
En=(En1-En0)*(power-pw0)/(pw1-pw0)+En0
nn0;En
```

Rao test 1

```
################ Rao test 1  ############
alpha=0.05;beta=0.2;N=5000;power=0.8

#1)delta=0.8;
uA=0;uB=0; uC=0.9798;nf=48  # nf is sample size for all treatments

#2)delta=0.6
uA=0;uB=0; uC=0.7348;nf=84

#3)delta=0.4
uA=0;uB=0; uC=0.4899;nf=183

#4)delta=0.2
uA=0;uB=0; uC=0.2449;nf=726

#### Find n0 and En0 (average stopping time) to obtain the minimal value of
power ####which is greater than 0.8

n0=ceiling(nf/3)   # per treatment
q=rep(0,N)
y=rep(0,N)

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

```
      AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
      n=1
      A=AA[1:n]
      B=BB[1:n]
      C=CC[1:n]
      sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2)
      R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2
      T1=((n/n0)*R)^(1/2)
      while(T1<=2.695 & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
      (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
      (T1=((n/n0)*R)^(1/2))
      }
 y[i]=n
 }
 pw=1-table(y,exclude=1:n0)/N
 n1=n0
pw1=pw
n1 # per treatment

for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En1=mean(q)

#### Find n0 and En0 (average stopping time) to obtain the minimal value of
power ####which is greater than 0.8
n0= # set n0 greater than n1
q=rep(0,N)
y=rep(0,N)

       for (i in 1:N){
      AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
      n=1
      A=AA[1:n]
```

```
      B=BB[1:n]
      C=CC[1:n]
      sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2)
      R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2
      T1=((n/n0)*R)^(1/2)
     while(T1<=2.695 & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
      (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
      (T1=((n/n0)*R)^(1/2))
      }
 y[i]=n
 }
 pw=1-table(y,exclude=1:n0)/N
pw0=pw
pw0
n0  # per treatment

for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En0=mean(q)
En0

nn0=(power-pw0)*(n1-n0)/(pw1-pw0)+n0 #the maximum sample size or
truncation point
En=(En1-En0)*(power-pw0)/(pw1-pw0)+En0

nn0 # trancation point per treatment
En  # expect per treatment


# Rao test 2

######## Rao Test 2 ############
alpha=0.05;beta=0.2;N=5000;pwer=0.8
```

```
#1)delta=0.8;
uA=0;uB=0; uC=0.9798;nf=48  # nf is sample size for all treatments

#2)delta=0.6
uA=0;uB=0; uC=0.7348;nf=84

#3)delta=0.4
uA=0;uB=0; uC=0.4899;nf=183

#4)delta=0.2
uA=0;uB=0; uC=0.2449;nf=726

q=rep(0,N)
y=rep(0,N)

n0=nf
cv=   # calculated by Maple software

#### Find n0 and En0 (average stopping time) to obtain the minimal value of
power ####which is greater than 0.8

  for (i in 1:N){
     AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
     n=1
     A=AA[1:n]
     B=BB[1:n]
     C=CC[1:n]
     sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2)
     R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2
     T2=(R)^(1/2)
     while(T2<=cv & n<=n0){
     (n=n+1)
     (A=AA[1:n])
     (B=BB[1:n])
     (C=CC[1:n])
     (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
     (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
     (T2=(R)^(1/2))
     }
```

```
 y[i]=n
 }
 pw=1-table(y,exclude=1:n0)/N;pw
pw1=pw

q=rep(0,N)
for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En1=mean(q)

#### Find n0 and En0 (average stopping time) to obtain the minimal value of
power ####which is greater than 0.8
n0= # set n0 greater than n1

  for (i in 1:N){
     AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
     n=1
     A=AA[1:n]
     B=BB[1:n]
     C=CC[1:n]
     sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2)
     R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2
     T2=(R)^(1/2)
     while(T2<=3.538 & n<=n0){
     (n=n+1)
     (A=AA[1:n])
     (B=BB[1:n])
     (C=CC[1:n])
     (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
     (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
     (T2=(R)^(1/2))
     }
 y[i]=n
 }
 pw=1-table(y,exclude=1:n0)/N;pw
pw0=pw

q=rep(0,N)
```

```
for (i in 1:N){
if (y[i]>K) (q[i]=y[i]-1)
else if (y[i]<=K) (q[i]=y[i])
   }
En0=mean(q)
nn0=(power-pw0)*(n1-n0)/(pw1-pw0)+n0 #the maximum sample size or
truncation point
En=(En1-En0)*(power-pw0)/(pw1-pw0)+En0
nn0 # trancation point per treatment
En  # expect per treatment
```

- **Simulate the empirical size at the given maximal sample size**

```
### Calculating type I error
#OBF
#1)delta=0.8;
K=5;Cb=6.27;uA=0;uB=0; uC=0;n0=49.84  # nf is sample size for all treatment

#2)delta=0.6
K=5;Cb=6.27;uA=0;uB=0; uC=0;n0=84.85
K=10;Cb=6.48;uA=0;uB=0; uC=0;n0=85.89

#3)delta=0.4
K=5;Cb=6.27;uA=0;uB=0; uC=0;n0=185.79
K=10;Cb=6.48;uA=0;uB=0; uC=0;n0=187.39

#4)delta=0.2
K=5;Cb=6.27;uA=0;uB=0; uC=0;n0=734.79
K=10;Cb=6.48;uA=0;uB=0; uC=0;n0=746.04

nts=ceiling(n0/(K*J))
y=rep(0,N)
for (i in 1:N){
      AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1);CC=rnorm(n0,uC,1)
      count=1
      m=1
      n=nts
      A=AA[1:nts]
      B=BB[1:nts]
```

```
      C=CC[1:nts]
      xbar=mean(c(A,B,C))
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(nts-1))
      F=(nts/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2)
    while(F<qf(1-pchisq(Cb*(K/count),J-1),J-1,J*(n-1),lower.tail = FALSE) &
count<=K){
      (m=m+1)
      (n=nts*m)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
      (count=count+1)
      }
  y[i]=count
  }
 typeI=1-table(y,exclude=1:K)/N
typeI
```

Pocock

```
#1)delta=0.8;
K=5;Cp=8.35;uA=0;uB=0; uC=0;n0=59.34;  # nf is sample size for all treatment
K=10;Cp=9.17;uA=0;uB=0; uC=0;n0=65.16;

#2)delta=0.6
K=5;Cp=8.35;uA=0;uB=0; uC=0;n0=102.25
K=10;Cp=9.17;uA=0;uB=0; uC=0;n0=109.82

#3)delta=0.4
K=5;Cp=8.35;uA=0;uB=0; uC=0;n0=225.09
K=10;Cp=9.17;uA=0;uB=0; uC=0;n0=238.98

#4)delta=0.2
K=5;Cp=8.35;uA=0;uB=0; uC=0;n0=886.52
K=10;Cp=9.17;uA=0;uB=0; uC=0;n0=930.404
```

```
nts=ceiling(n0/(K*J))
y=rep(0,N)
for (i in 1:N){
      AA=rnorm(n0,uA,1);BB=rnorm(n0,uB,1);CC=rnorm(n0,uC,1)
      count=1
      m=1
      n=nts
      A=AA[1:nts]
      B=BB[1:nts]
      C=CC[1:nts]
      xbar=mean(c(A,B,C))
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(nts-1))
      F=(nts/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2)
    while(F<qf(1-pchisq(Cp,J-1),J-1,J*(n-1),lower.tail = FALSE) & count<=K){
      (m=m+1)
      (n=nts*m)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
      (count=count+1)
      }
 y[i]=count
 }
 typeI=1-table(y,exclude=1:K)/N
typeI
```

Sequential F-test

```
#1)delta=0.8;
uA=0;uB=0;uC=0;n0=23.215 # n0 is sample size for each treatment

#2)delta=0.6
uA=0;uB=0; uC=0;n0=39.55385
```

```
#3)delta=0.4
uA=0;uB=0; uC=0;n0=82


#4)delta=0.2
uA=0;uB=0; uC=0;n0=321.0154

#type I error
n0=ceiling(n0*3)/3  # n0 is sample size for each treatment

y=rep(0,N)
for (i in 1:N){
      AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
      n=k0
      A=AA[1:n]
      B=BB[1:n]
      C=CC[1:n]
      ma=mean(A)
      mb=mean(B)
      mc=mean(C)
      m=mean(c(ma,mb,mc))
      Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2)))
      while(Lk<=a & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (ma=mean(A))
      (mb=mean(B))
      (mc=mean(C))
      (m=mean(c(ma,mb,mc)))
      (Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2))))
      }
 y[i]=n
 }
 typeI=1-table(y,exclude=1:n0)/N
typeI
```

```
#Rao test 1
### calculating type I error
#1)delta=0.8;
uA=0;uB=0; uC=0;n0=18.67 # n0 is sample size for each treatment

#2)delta=0.6
uA=0;uB=0; uC=0;n0=31.215

#3)delta=0.4
uA=0;uB=0; uC=0;n0=66.93

#4)delta=0.2
uA=0;uB=0; uC=0;n0=260.85

n0=ceiling(n0*3)/3  # n0 is sample size for each treatment
y=rep(0,N)
for (i in 1:N){
      AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
      n=1
      A=AA[1:n]
      B=BB[1:n]
      C=CC[1:n]
      sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2)
      R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2
      T1=((n/n0)*R)^(1/2)
      while(T1<=2.695 & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
      (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
      (T1=((n/n0)*R)^(1/2))
      }
 y[i]=n
 }
 typeI=1-table(y,exclude=1:n0)/N
typeI

# Rao test 2
```

```
### calculating type I error
#1)delta=0.8;
uA=0;uB=0; uC=0;n0=27.51 ; cv=3.435# nf is sample size for all treatments

#2)delta=0.6
uA=0;uB=0; uC=0;n0=47.075;cv=3.483

#3)delta=0.4
uA=0;uB=0; uC=0;n0=103.94;cv=3.542

#4)delta=0.2
uA=0;uB=0; uC=0;n0=412.39;cv=3.624

n0=ceiling(n0)
y=rep(0,N)
for (i in 1:N){
      AA=rnorm(n0*2,uA,1);BB=rnorm(n0*2,uB,1);CC=rnorm(n0*2,uC,1)
      n=1
      A=AA[1:n]
      B=BB[1:n]
      C=CC[1:n]
      sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2)
      R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2
      T2=R^(1/2)
      while(T2<=cv & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
      (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
      (T2=R^(1/2))
      }
 y[i]=n
 }
 typeI=1-table(y,exclude=1:n0)/N
typeI
```

- **Calculate Critical value for Rao test 2 (Maple software)**

```
> f:=(n0,d)->-solve(0.05=c2^d*exp(-c2^2/2)/(2^(d/2)*GAMMA(d/2))*(ln(n0)*(1-
d/c2^2)+4/c2^2));
> f(n0,2)[3];   (input n0 to calculate CV)
```

3. Four-treatment Comparison

- **Program procedures are similar to three-treatment comparison.**

4. Application

```
cc<-read.table("F:/thesis_usb/thesis_usb/new data
set/simulatedforMing.txt",h=TRUE)
AA=cc[,2]   #AA-Volume.control
BB=cc[,3]   #BB-Volume.traditional
CC=cc[,4]   #CC-Volume.Bone
alpha=0.05; n0=50
win.graph()
par(mfrow=c(3,2))

# OBF
J=3;K=10;Cb=6.48;n0=50 #nts: sample size for each treatment
nts=50/K
      count=1
      m=1
      n=nts
      A=AA[1:nts]
      B=BB[1:nts]
      C=CC[1:nts]
      xbar=mean(c(A,B,C))
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(nts-1))
      F=(nts/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2)
    while(F<qf(1-pchisq(Cb*(K/count),J-1),J-1,J*(n-1),lower.tail = FALSE) &
count<=K){
      (m=m+1)
      (n=nts*m)
```

```
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
      (count=count+1)
      }
count
count*nts # stopping time

N=K
z=matrix(rep(0,N*3),ncol=3,byrow=T)
for (i in 1:N){
 (n=nts*i)
 (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
       (ss=n*J)
       (cv=qf(1-pchisq(Cb*(K/i),J-1),J-1,J*(n-1),lower.tail = FALSE))
       z[i,]=c(F,ss,cv)
   }
z
stat=z[1:count,1] # OBF statistic
ss=z[1:count,2]  # cumulative total sample size on three arms
cv=z[1:count,3]  # critical value
plot(ss,stat, type="l", lty=1,xlab="Cumulative total sample size on three
arms",ylab="The OBF statistic")
lines(ss,cv,lty=2)
points(ss,stat)
points(ss,cv,pch=19)
legend(x="topleft",legend=c("Critical value", "statistic"),lty=c(2,1))
#axis(1,at=c(0,15,30,45,60,75,90,105),labels=c(0,15,30,45,60,75,90,105))

######Pocock
```

```
J=3; K=10;Cp=9.17;n0=50
nts=n0/K
      count=1
      m=1
      n=nts
      A=AA[1:nts]
      B=BB[1:nts]
      C=CC[1:nts]
      xbar=mean(c(A,B,C))
      s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(nts-1))
      F=(nts/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2)
    while(F<qf(1-pchisq(Cp,J-1),J-1,J*(n-1),lower.tail = FALSE) & count<=K){
      (m=m+1)
      (n=nts*m)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
      (count=count+1)
      }
  count
count*nts


N=K
z=matrix(rep(0,N*3),ncol=3,byrow=T)
for (i in 1:N){
  (n=nts*i)
  (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (xbar=mean(c(A,B,C)))
      (s2=(sum((A-mean(A))^2)+sum((B-mean(B))^2)+sum((C-
mean(C))^2))/(J*(n-1)))
      (F=(n/((J-1)*s2))*((mean(A)-xbar)^2+(mean(B)-xbar)^2+(mean(C)-xbar)^2))
      (ss=n*J)
```

```
      (cv=qf(1-pchisq(Cp,J-1),J-1,J*(n-1),lower.tail = FALSE))
       (z[i,]=c(F,ss,cv))
   }
z
stat=z[1:count,1] # pocock statistic
ss=z[1:count,2]  # cumulative total sample size on three arms
cv=z[1:count,3]  # critical value
plot(ss,stat, type="l", lty=1,xlab="Cumulative total sample size on three
arms",ylab="The Pocock statistic")
lines(ss,cv,lty=2)
points(ss,stat)
points(ss,cv,pch=19)
legend(x="topleft",legend=c("Critical value", "statistic"),lty=c(2,1))
```

Rao test 1

```
cv=2.695;n0=50;J=3
      n=1
      A=AA[1:n]
      B=BB[1:n]
      C=CC[1:n]
      sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2)
      R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2
      T1=((n/n0)*R)^(1/2)
      while(T1<=cv & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
      (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
      (T1=((n/n0)*R)^(1/2))
      }
count=n
count
N=n0
z=matrix(rep(0,N*2),ncol=2,byrow=T)
for (i in 1:N){
      (n=i)
      (A=AA[1:n])
```

```
      (B=BB[1:n])
      (C=CC[1:n])
      (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
      (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
      (T1=((n/n0)*R)^(1/2))
      (ss=i*J)
      z[i,]=c(T1,ss)
}
z
stat=z[1:count,1] # Rao test 1 statistic
ss=z[1:count,2]  # cumulative total sample size on three arms

plot(ss,stat, type="l",xlim=c(0,100),ylim=c(0,4.5), lty=1,xlab="Cumulative total
sample size on three arms",ylab="The Rao test 1 statistic")
abline(h=cv,lty=2)
legend(x="topleft",legend=c("Critical value", "statistic"),lty=c(2,1))

##### Rao test 2
n0=50;cv=3.490;J=3
      n=1
      A=AA[1:n]
      B=BB[1:n]
      C=CC[1:n]
      sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2)
      R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2
      T2=(R)^(1/2)
      while(T2<=cv & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
      (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
      (T2=(R)^(1/2))
      }
count=n
count
N=n0
z=matrix(rep(0,N*2),ncol=2,byrow=T)
for (i in 1:N){
```

```
      (n=i)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (sigma=sqrt((1/(3*n))*sum(A^2+B^2+C^2)-((1/(3*n))*sum(A+B+C))^2))
      (R=(sum(A+B-2*C)/(sigma*sqrt(6*n)))^2+(sum(A-B)/(sigma*sqrt(2*n)))^2)
      (T2=(R)^(1/2))
      (ss=i*J)
      z[i,]=c(T2,ss)
}
z

stat=z[1:count,1] # Rao test 2 statistic
ss=z[1:count,2]  # cumulative total sample size on three arms

plot(ss,stat, type="l",xlim=c(0,105),ylim=c(0,5.5), lty=1,xlab="Cumulative total
sample size on three arms",ylab="The Rao test 2 statistic")
abline(h=cv,lty=2)
legend(x="topleft",legend=c("Critical value", "statistic"),lty=c(2,1))
```

Sequential F-Test

```
n0=50;k0=7;a=5.543;J=3 ;d=3 # a---cv
      n=k0
      A=AA[1:n]
      B=BB[1:n]
      C=CC[1:n]
      ma=mean(A)
      mb=mean(B)
      mc=mean(C)
      m=mean(c(ma,mb,mc))
      Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2)))
      while(Lk<=a & n<=n0){
      (n=n+1)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (ma=mean(A))
      (mb=mean(B))
```

```
      (mc=mean(C))
      (m=mean(c(ma,mb,mc)))
      (Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2))))
      }
n
count=n

N=n0
z=matrix(rep(0,N*2),ncol=2,byrow=T)
for (i in k0:N){
      (n=i)
      (A=AA[1:n])
      (B=BB[1:n])
      (C=CC[1:n])
      (ma=mean(A))
      (mb=mean(B))
      (mc=mean(C))
      (m=mean(c(ma,mb,mc)))
      (Lk=n*d/2*log(1+n*((ma-m)^2+(mb-m)^2+(mc-m)^2)/(sum((A-
ma)^2)+sum((B-mb)^2)+sum((C-mc)^2))))
      (ss=i*J)
      z[i,]=c(Lk,ss)
}
z
 stat=z[k0:count,1] # Sequential F statistic
ss=z[k0:count,2]  # cumulative total sample size on three arms

plot(ss,stat, type="l",xlim=c(k0*J,100),ylim=c(1,8), lty=1,xlab="Cumulative total
sample size on three arms",ylab="The Sequential F statistic")
abline(h=a,lty=2)
legend(x="topleft",legend=c("Critical value", "statistic"),lty=c(2,1))
 dev.copy2eps(file="paper.eps")
```