Stony Brook University

The Power of Detecting a Mixture in the Treatment Groups through Analysis of Dose Response Data

A Dissertation Presented

by

Xiawei Tu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

August 2011

Stony Brook University

The Graduate School

Xiawei Tu

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation

Nancy R. Mendell – Dissertation Advisor
Professor, Department of Applied Mathematics and Statistics

Stephen J. Finch – Chairperson of Defense
Professor, Department of Applied Mathematics and Statistics

Wei Zhu
Deputy Chair, Professor, Department of Applied Mathematics and Statistics

Deborah Levy
Director, Associate Professor of Psychology,
Psychology Research Laboratory in McLean Hospital

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

The Power of Detecting a Mixture in the Treatment Groups through Analysis of Dose Response Data

by

Xiawei Tu

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2011

In many clinical experiments, some subjects are unaffected by the treatment. This so-called non-response phenomenon has attracted the attention of many researchers in recent years. This dissertation focuses on detecting the association between the dose level and the observed values in the case where there is a mixture in treatment groups. That is, there is a linear relation between dose and response in a fraction of the observations and the shift in mean increases as the dose level increases.

We investigate the Likelihood Ratio Test (LRT) in the context of normal mixture models. We do this based on critical values for the LRT obtained through simulation. We customize the

general Expected-Maximization (EM) Algorithm to our situation in order to obtain Maximum Likelihood Estimates (MLE) of the parameter values under the alternative. We note that as we expected MLE of the parameters are close to the true parameter values and the mean square error decreases as the sample size increases.

For the power study we also conduct LRT, Spearman's correlation test, and Simple linear regression test on each simulated sample. The power of three tests is compared and the McNemar's test is conducted to test the difference between tests. Overall, the power of the LRT is greater than Spearman's test and Simple linear regression test, although the three tests are not powerful with small shifted proportion and small shifted mean. We conclude that the LRT is very powerful in those cases where the mixing proportion is greater than 0.5 and there is a linear dose response relationship with slope greater than or equal to 0.3 standard deviation units. At the same time, the Simple linear regression test works almost as well as the LRT in those cases where the power is greater than 0.5.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I would like to express my great gratitude to my esteemed advisor, Professor Nancy R. Mendell, for her constantly support, help, and invaluable guidance. Her kindness and help made my life in Stony Brook University enjoyable and unforgettable.

I would like to thank my committee members, Professor Deborah Levy, Professor Stephen J. Finch, and Professor Wei Zhu, for their helpful suggestions. Especially, thanks for Professor Deborah Levy's support and academic training on my research.

I am thankful to the department of Applied Math and Statistics, for the academic and facility support during my study.

Finally, I would like to thank my family members and friends for the care and the support over years. My life is much more meaningful with all of them.

Chapter 1

Introduction and Literature Review

1.1 Introduction

In many clinical experiments, some subjects are unaffected by the treatment. This so-called non-response phenomenon has attracted the attention of many researchers in recent years. A mixture model has traditionally been proposed to describe the distribution of responses in treatment groups for such experiments. The distributions in the affected subjects are represented by the same distribution as the unaffected subjects, but with shifts in the means.

Several investigators have considered various special cases of group comparison in the presence of mixtures in treatment groups. In particular, Good (1979) [1] considered the situation where one is comparing a treatment group to a control group and one has a mixture in the treatment group only. The mixture consists of a fraction of subjects with the same distribution as the controls and the remaining fraction with the same distribution but a shift in the mean. The hypotheses Good considered were:

$$H_0: G_2(x) = G_1(x) \qquad (1.1)$$

$$H_1: G_2(x) = p\, G_1(x - \Delta) + (1 - p) G_1(x) \qquad (1.2)$$

Here $0 < p \leq 1$ (where p is the proportion of responders), $\Delta \neq 0$ (where Δ is the shifted mean of responders), and $G_1(x)$ and $G_2(x)$ denote the cumulative density function (c.d.f.) for the control and treatment groups respectively. Good assumed that $G_1(x)$ is normally distributed. Thus under the alternative hypothesis $G_2(x)$ is a mixture of two normal distributions. Under the alternative hypothesis, the difference between the average observed in treated and control groups will lead to unbiased estimates of the difference between the treatment mean and the control mean. Good proposed a randomization test statistic for this alternative and suggested that the Wilcoxon Rank Sum test is not effective for this situation. The test statistic $V(\theta)$ proposed by Good took the mixture into account and was defined as follows:

$$V(\theta) = \theta\left(\frac{1}{n} + \frac{1}{m}\right)^{-1}(\bar{X} - \bar{Y})^2 + (1-\theta){S_y}^2 \qquad (1.3)$$

Here $0 \leq \theta \leq 1$ (θ is the proportion of the difference in sample means in the test statistic), $\bar{X}$, n ($\bar{Y}$, m) denote the mean, and the size of the control (treatment) sample respectively, and ${S_y}^2$ denotes the sum of squares of deviations from the mean of the treatment group. Good suggested that one calculate $V(0.67)$ as a compromise when there is no advance information either of underlying distributions or the expected proportion of responders in the treatment group.

This dissertation is the extension of Good's work considering the situation where there is a dose-response phenomenon in the treatment groups. The Likelihood Ratio Test is conducted to detect the relation between dose-level variable and the observed value in each group. The

purpose of this study is to compare the power of the Likelihood Ratio Test (LRT), Spearman's Correlation Test, and the Simple Linear Regression Test. The power will be assessed and compared in several different scenarios by mixing proportion, shifted mean, sample size of observed mixture populations.

A comprehensive literature review of the topic is presented in the following section. It includes the methods that have been used to solve the mixture problem and the outline of the problem in this dissertation. Chapter 2 explores the Likelihood Ratio Test and the EM algorithm that is used for finding MLEs for dose-response mixture models. Simulation results are presented with parameter estimates. We investigate the null distribution of the LRT statistics in Chapter 3. The power study is discussed in Chapter 4. In addition, we compare the power of the three tests for each simulated sample. Finally, Chapter 5 contains the conclusions and proposed future studies.

1.2 Literature Review

1.2.1 Normal Mixture Model

Based on the mixture model from Good's paper, several researchers investigated many methods to test whether the means of the control and the treatment group are equal in the presence of non-responders in the treatment group. Boos and Brownie (1986) [2] explored this method and claimed that Good's test appeared to be effective, but also noted that Wilcoxon Rank Sum test and the ordinary two sample t-test are almost as powerful as Good's method. Moreover, the latter are much easier to use and also easier to interpret. Boos and Brownie also noted that when the proportion of responders, p, is 0.6 or more, the Wilcoxon test will generally be a good choice for detecting a treatment effect.

Johnson et al. (1987) [3] obtained rank tests that are locally most powerful with respect to changes in the mixing proportion based on Good's hypothesis. They considered two special cases for the test, a mixture of normal distributions and a mixture of uniform distributions. The empirical power studies and asymptotic efficiencies were constructed and compared with Wilcoxon and normal scores tests.

Conover et al. (1988) [4] proposed two two-parameter models for testing the hypothesis of no treatment effect against the alternative that a subset of the treated patients will show an improvement. The Lehmann alternatives [5] are used in both models to keep the range of measurements the same for treated and control patients. They developed the locally most powerful rank tests for each model and each parameter.

Lo et al. (2002) [6] considered a modification of Good's model and proposed that a two-sample permutation test be applied to the likelihood ratio statistic. They compared the working memory scores of relatives of schizophrenia patients to normal controls. The alternative was that the distribution of the scores in relatives of schizophrenia patients was a mixture of an exponential and a normal distribution whereas the distribution of the scores in the normal controls was a single exponential distribution with the same parameter value as that of relatives. The null hypothesis was that the distribution of the working memory scores was an exponential distribution with equal means in the two groups. They proposed a likelihood ratio test, calculated the critical value of the likelihood ratio test statistic and the p-values. Lo et al. also did a modest power study of the likelihood ratio test for a mixture of normal distributions in treatment groups and a single normal distribution in the control group, as well as for a mixture of exponential distributions in treatment groups and a single exponential distribution in the control group. Although they did the power study for small sample sizes ($n = 50, 75$), they obtained good power for a mixing proportion of 0.5 for normal mixture distributions in treatment groups as well as for a mixing proportion of 0.5 and 0.7 for exponential mixture distributions in treatment groups but not for mixing proportion of 0.9 in both situations. Additionally they did not compare the power of the likelihood ratio test to other test statistics.

McMahon et al. (2005) [7] proposed more powerful two-sample tests for the differences in repeated measures of adverse effects in psychiatric trials. They use Kendall's τ_b [8] as a summary measure of within-participant trends in adverse events, in conjunction with a weighted modification of a rank test proposed by Conover and Salsburg. A power study was conducted to compare the proposed analysis for repeated measures Analysis of Variance (ANOVA) using mixed models and the alternate tests for treatment differences in τ_b trend scores.

1.2.2 Dose-response Model

Based on the mixture model used in the non-response phenomenon, researchers considered dose-response models in which the proportion of non-responders changes at different dose amounts in different treatment groups. The dose-response relationship, or exposure-response relationship, describes the change in effect on an organism caused by differing levels of exposure, or doses, to a stressor after a certain exposure time. This may apply to individuals or to populations. Studying dose-response, and developing dose-response models with mixture in the treatment groups, is central to determining the relationship between response proportion and levels of dosages for drugs or other relationships among substances to which humans or other organisms are exposed. These conclusions are often the basis for public policy.

Boos and Brownie (1991) [9] investigated a model combining a logistic regression on dose for the probability that an animal will "respond" to treatment with a linear regression on dose for the mean of the responders. They described the maximum likelihood estimation by Expectation and Maximization algorithm (EM algorithm) and used likelihood ratio tests to distinguish between the full model and meaningful reduced-parameter versions.

Razzaghi and Kodell (2000) [10] investigated a mixture dose-response model on additional risk in laboratory animals, defined as the excess risk over the background risk due to an added dose. They derived an upper confidence limit on additional risk using the asymptotic distribution of the likelihood ratio statistic. The EM algorithm was used to find the maximum likelihood estimates of model parameters and was extended to derive the estimates when the model is subject to a specified level of added risk.

Luo et al. (2004) [11] developed three score tests for hypothesis testing in a dose-response framework and showed that increased power is possible by using a mixture model where both the logarithm of the response rate and the response mean are linear functions of the dose level. They used permutation tests to control the type I error. The power properties of the tests showed that the proposed score tests have good performance.

Many applied papers are based on the mixture model with a proportion of non-responders in treatment group. Levin and Bowman (1986) [12] focused on the relationship between behavioral effects of chronic exposure to low concentrations of halothane and the development in rats. Rice (1990) [13] explored the relationship between behavioral impairment on spatial discrimination reversal task and the periods of development of monkeys. Cox et al. (1993) [14] proposed the dose-response model to determine the selectivity of lead-induced changes in learning and to explore the nature of the underlying error patterns contributing to any learning deficits.

1.2.3 The Expectation-Maximization Algorithm

The Expectation-Maximization (EM) iterative algorithm is a broadly applicable statistical technique for maximizing complex likelihoods and handling the incomplete data problem. At each iteration step of the algorithm, two steps are performed: (i) the E-Step, consisting of projecting an appropriate function containing the augmented data on the space of the original, incomplete data, and (ii) the M-Step, consisting of maximizing the function. The name EM algorithm was coined by Dempster, Laird, and Rubin in their fundamental paper [15]. This paper presents the EM algorithm as an effective way to compute the maximum-likelihood estimates when the observations can be viewed as incomplete data.

The use of the EM algorithm for parameter estimation is stated as follows. Suppose we want to estimate the weights or proportions of a fixed number of fully known distributions. The EM approach introduces unobserved indicators with the goal of simplifying the likelihood. The weights are estimated by the maximum-likelihood method. Assume that a sample $x_1, x_2, \ldots, x_n$ comes from the mixture

$$f(x, w) = \sum_{j=1}^{k} w_j f_j(x) \qquad (1.4)$$

where the weights $0 \leq w_j \leq 1$ are unknown and constitute a k -dimensional vector $(w_1, \ldots, w_{k-1})$ with $w_k = 1 - w_1 - \cdots - w_{k-1}$; the class-densities $f_j(x)$ are fully specified.

Even in this simplest case when the only parameters are the weights, w, the log-likelihood assumes a quite complicated form,

$$\sum_{i=1}^{n} \log f(x_i, w) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{k} w_j f_j(x_i)\right) \qquad (1.5)$$

The derivatives with respect to w_j lead to a system of equations that is not solvable in a closed form. However, that is where the EM algorithm comes in. Upon applying the EM algorithm, we first augment the data $x = (x_1, x_2, \dots, x_n)$ by an "unobservable" matrix with *n* rows and *k* columns having entries $z = (z_{ij}, i = 1,2, \dots, n; j = 1,2, \dots, k)$. The values z_{ij} are indicators, defined as

$$z_{ij} = \begin{cases} 1, & x_i \in f_j \\ 0, & otherwise \end{cases} \qquad (1.6)$$

for j=1, 2, …, k and i=1, 2, …, n. The unobservable matrix z describes where the i^{th} observation x_i comes from. Note that each row of z contains only one value equal to 1 and $(k-1)$ values equal to 0. With the augmented data, $x = (y, z)$, the complete likelihood takes quite a simple form,

$$\prod_{i=1}^{n} \prod_{j=1}^{k} (w_j f_j(x_i))^{z_{ij}} \qquad (1.7)$$

The complete log-likelihood is

$$\log L_c(w) = \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \log w_j + C \qquad (1.8)$$

where $C = \sum_i \sum_j z_{ij} \log f_j(x_i)$ is independent of w.

Next we assume that the m^{th} iteration of weights is $w^{(m)}$ is already obtained. The m^{th} E-Step is,

$$E_{w^{(m)}}\left(z_{ij}|x\right) = P_{w^{(m)}}\left(z_{ij} = 1|x\right) = {z_{ij}}^{k} \qquad (1.9)$$

where ${z_{ij}}^{(m)}$ is the posterior probability of the ith observation coming from the jth mixture-component, f_j, in the iterative step m.

$${z_{ij}}^{(m)} = \frac{{w_i}^{(m)} f_j(x_i)}{f\left(x_i, w^{(m)}\right)} \qquad (1.10)$$

Since the $\log L_c(w)$ is linear in the z_{ij}'s, $Q(w, w^{(m)})$ is simply $\sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij} \log w_j + C$. The subsequent M-Step is simple: $Q(w, w^{(m)})$ is maximized by,

$$w_j^{(m+1)} = \frac{\sum_{i=1}^{n} z_{ij}^{(m)}}{n} \qquad (1.11)$$

In many applications, the component for individual observations is unknown or missing, and thus the mixture models can be interpreted as describing an incomplete data situation. Several applications are provided, including examples of grouped, truncated data, finite mixture models, variance component estimation, and factor analysis.

Aitkin et al. (1980) [16] applied the EM algorithm to normal mixtures for outliers, in both single-sample and regression problems. The study considers the two component mixture model with unequal means and variances as well as regression models. The EM algorithm is very simply programmed, converges rapidly on the three examples considered, and can provide the ML estimates of the parameters, including the proportion of each component in the mixture, the asymptotic covariance matrix, and the maximized log-likelihood function.

S. N. Rai et al. (1993) [17] proposed an improved EM algorithm for situations in which the maximization of the "complete data" likelihood function does not have a closed-form solution. Self-consistency of the modified EM algorithm was established. Application to carcinogenicity experiments was illustrated in the paper, and the results of a simulation study

comparing the original and modified versions indicate that use of the proposed modification can lead to significant computational savings.

R. C. Jansen et al. (1993) [18] embedded the finite mixture model within the general framework of generalized linear models (GLMs). Implementation of the proposed EM algorithm was readily done in statistical packages with facilities for GLMs. In this paper, a practical example was presented where a generalized linear finite mixture model of ten Weibull distributions was adopted. The example was concerned with the flow cytometric measurement of the DNA content of spermatids in a mutant mouse, which shows non-disjunction of specific chromosomes during meiosis.

1.2.4 The Likelihood Ratio Test

The likelihood-ratio test (LRT) is a statistical test for comparing a more complex model to a simpler model. If so, the additional parameters of the more complex model are often used in subsequent analyses. The LRT is only valid if used to compare hierarchically-nested models. That is, the more complex model must differ from the simple model only by the addition of one or more parameters. Adding additional parameters will always result in a higher likelihood score. However, there comes a point when adding additional parameters is no longer justified in terms of significant improvement in fit of a model to a particular data set. The LRT provides one objective criterion for selecting among possible models.

E. V. Nordheim et la (1984) [19] examined the unusual performance of likelihood methods for genetic linkage models with unknown phase. The likelihood ratio test for homogeneity of linkage among several groups is often too conservative. The magnitudes of the differences between nominal and real probability levels of the LRT test were computed for various combinations of number of groups and sample sizes. An example from forest genetics was discussed at the end of this paper.

J. T. Kent (1982) [20] constructed the distribution of the likelihood ratio statistic when the data do not come from the parametric model, but when the 'nearest' member of the parametric family still satisfies the null hypothesis. In this paper, the likelihood ratio statistic no longer follows an asymptotic chi-squared distribution, and an alternative statistic based on the union-intersection approach was proposed.

M. V. Matz and R. Nielsen (2005) [21] described a likelihood ratio test that can be used to test if a sampled sequence is a member of an a priori specified species. Investigation of the performance of the test was conducted by using coalescence simulations, as well as using the real data from butterflies and frogs representing two kinds of challenge for DNA barcoding: extremely low and extremely high levels of sequence variability.

McLachlan, G. J (2000) [27] explored that the asymptotic distribution of LRT statistics is the combination of chi-square distribution with the degree of freedom 1 and chi-square distribution with the degree of freedom 2 under regularity conditions that do not hold for the normal mixture model. Based on this research, the asymptotic distribution of the LRT statistics for the normal mixture case can be treated as the combination of two chi-square distributions instead of one single chi-square distribution.

Based on the asymptotic distribution of LRT statistics in the mixture case, investigators use simulation to obtain the empirical null distribution for power studies. Bootstrap and permutation test are the most common method for the power study based on the null distribution. Hall (1987) [28] described a method for constructing likelihood-based confidence regions for a vector parameter, using the bootstrap and nonparametric density estimation. By permutation test, Lo (2000) [29] et al. reported simulation studies of two cases where they tested a single normal versus a two-component normal mixture and a two-component normal mixture versus a three-component normal mixture.

1.2.5 The Classic Statistical Tests

Single Variable Linear Regression

Single variable linear regression fits a straight line through the set of points in such a way that minimizes the sum of squared residuals of the model (that is, vertical distances between the points of the data set and the fitted line).

In the presence of mixture with dose-response model, a simple method for testing the null hypothesis of no dose (or treatment) effect is based on linear regression, that is, regressing the response on the dose level and using the t statistics to test if the slope equals 0. In the linear regression process, the relation between the mean response function and dose level variable in a simple linear fashion is linked as follows:

$$\mu_i = \mu(x_i) = (1-\pi)\mu + \pi\beta x_i = A + Bx_i \qquad (1.14)$$

The model regards each observation as a random departure from the central value $\mu(x_i)$, say, $y_{ij} = \mu(x_i) + \varepsilon_i$, where $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, n_i$, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. In the model, the regression parameters, A and B, represent the Y-intercept and slope, respectively, of the straight-line response. B is the parameter of interest. If $B = 0$, then the mean response does not vary in a

linear fashion and no linear trend with x exists. In this sense, the model is used for simple trend analysis, by testing H_0: $B = 0$.

The least-squares estimation is a popular method to obtain the estimator of parameters. The closed forms are:

$$\hat{B} = \frac{\sum_{i=0}^{I}\sum_{j=1}^{n_i}(x_i - \bar{x})(y_{ij} - \bar{y})}{\sum_{i=0}^{I}(x_i - \bar{x})^2} \qquad (1.15)$$

$$\hat{A} = \bar{y} - \hat{B}\bar{x} \qquad (1.16)$$

where $\bar{x} = \sum_{i=0}^{I} x_i / I$ and $\bar{y} = \sum_{i=0}^{I}\sum_{j=1}^{n_i} y_{ij} / \sum_{i=0}^{I} n_i$ are the sample means of dose level and response variable. Also, the estimated standard error of the slope estimator is:

$$se\left[\hat{B}\right] = \frac{\hat{\sigma}}{\{\sum_{i=0}^{I}(x_i - \bar{x})^2\}^{1/2}} \qquad (1.17)$$

where $\hat{\sigma}$ is the root mean square error :

$$\hat{\sigma} = \{\frac{1}{\sum_{i=0}^{I} n_i - 2}\sum_{i=0}^{I}\sum_{j=1}^{n_i}(y_{ij} - \hat{A} - \hat{B}x_i)^2\}^{1/2} \qquad (1.18)$$

To conduct a trend under this simple linear model, the t-statistics is calculated, $t = \hat{B}/se[\hat{B}]$. The null reference distribution of t-statistics is $t(\sum_{i=0}^{I} n_i - 2)$.The H_0 is rejected in favor of a increasing trend, $H_a: B \geq 0$, if t exceeds the upper-α critical point $t_\alpha(\sum_{i=0}^{I} n_i - 2)$.

Spearman's rank correlation

Spearman's rank correlation is also used when one is interested in an order association between two measurement variables. The purpose is to test whether the two measurement variables covary, or whether as one variable increases the other variable tends to increase or decrease. Spearman's rank correlation is the non-parametric alternative to correlation — it assesses how well an arbitrary monotonic function could describe the relationship between two variables, without making any other assumptions about the particular nature of the relationship between the variables. Spearman's rank correlation is also used when one or both of the variables consist of ranks. The raw values are converted to ranks, and the differences, d_i, between the ranks of each observation on the two variables are calculated. The steps for calculating Spearman's correlation coefficient are as follows:

(1) Rank the original values in each group in ascending order, respectively.

(2) Calculate the difference between the ranks of corresponding values x_i and y_i.

(3) If there are no tied ranks, then ρ is given by:

$$\rho = 1 - \frac{6 \sum {d_{ij}}^2}{N(N^2 - 1)} \tag{1.19}$$

where ${d_{ij}}^2 = R_{x_{ij}} - R_{y_{ij}}$ is the difference between the ranks of corresponding values $R_{x_{ij}}$ and $R_{y_{ij}}$, and N is the number of values in each data set (same for both sets). If tied ranks exist, the classic Pearson's correlation coefficient between ranks has to be used instead of the above formula:

$$\rho = \frac{N \sum_{i=0}^{I} \sum_{j=1}^{n_i} R_{x_i} R_{y_{ij}} - (\sum_{i=0}^{I} n_i R_{x_i})(\sum_{i=0}^{I} \sum_{j=1}^{n_i} R_{y_{ij}})}{\sqrt{n(\sum_{i=0}^{I} n_i {R_{x_i}}^2) - (\sum_{i=0}^{I} n_i R_{x_i})^2} \sqrt{N(\sum_{i=0}^{I} \sum_{j=1}^{n_i} {R_{y_{ij}}}^2) - (\sum_{i=0}^{I} \sum_{j=1}^{n_i} R_{y_{ij}})^2}} \tag{1.20}$$

For large samples, the test statistic $\frac{\rho}{\sqrt{(1-\rho^2)/(N-2)}}$ approximately follows a t distribution with $N - 2$ degrees of freedom under the null hypothesis. Thus a t test might be used to test the null hypothesis of no correlation between the ranks of two variables.

1.3 Summary of the Dissertation

In this thesis, we consider the situation where we test the relation between qualitative explanatory variable X which represents the dose level and continuous dependent variable Y which represents the observed values in each group. Our purpose is to compare the power of Likelihood Ratio Test with Single Variable Linear Regression Test and Spearman's Correlation Test with the presence of normal mixture in the treatment groups. Our model is a mixture model like that proposed by Good (1979) but modified to allow a dose-dependent effect in a subset of responders. We customize the EM algorithm to our specific situation and generate the empirical null distribution for power study. After the power comparison, we find out that the LRT is powerful than other two tests, especially for large samples with large slope and shifted proportion.

Chapter 2 The Problem and Methods

2.1 The problem and the study design

2.1.1 The problem

Assume that $y_{01}, \cdots, y_{0n_0}$ are independent and identically distributed (I.I.D.) as the control group, and that $y_{i1}, \cdots, y_{in_i}, i = 1, 2, \ldots, \mathrm{I}$ are I.I.D. as treatment groups with distribution function F. We try to test the hypothesis.

$$H_0: F(y; X = x_i) = \varphi\left(\frac{y - \mu_0}{\sigma}\right)$$

$$H_1: \begin{cases} (control) & F(y; X = 0) = \varphi\left(\frac{y - \mu}{\sigma}\right) \\ (treatment) & F(y; X = x_i) = (1 - \pi) * \varphi\left(\frac{y - \mu}{\sigma}\right) + \pi * \varphi\left(\frac{y - \mu - x_i * \beta}{\sigma}\right) \end{cases}$$

(2.1)

where:

(1) π ($0 < \pi < 1$) is the proportion of responders given $x_i > 0$;

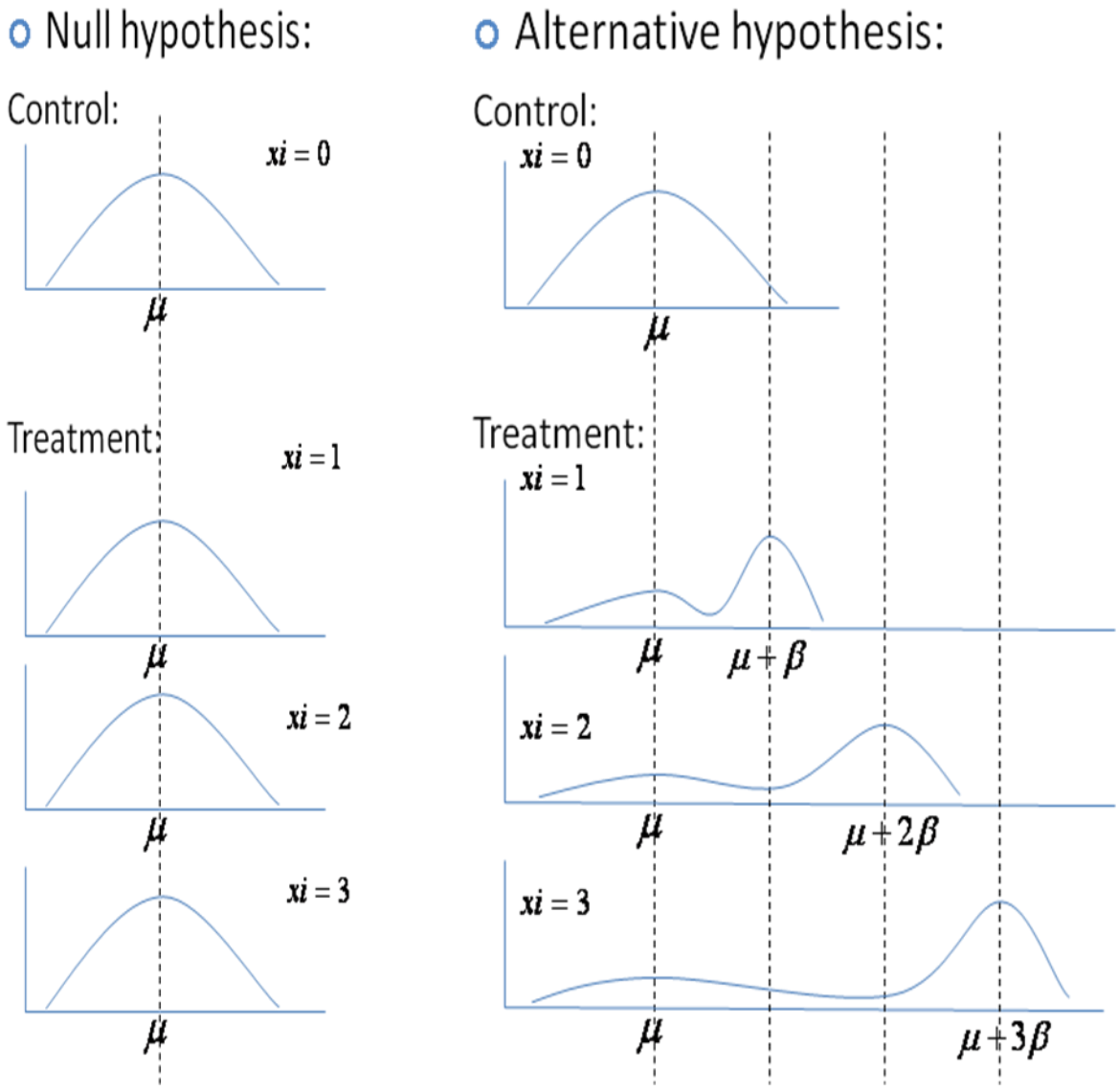
(2) μ is the mean for non-responders;

(3) σ is a scale parameter assumed common to the distributions for responders and non-responders;

(4) x_i is a given dose.

(5) $\varphi(x)$ is the standard normal distribution.

Figure 2.1 --- An example of the null hypothesis and the alternative hypothesis with the presence of mixture in the treatment groups.



We are interested in fitting the normal mixture model to generate data and the main purposes of the research are following:

(1) To investigate the power of Likelihood Ratio Test on detecting the relation between X and Y upon estimating all the parameters in the normal mixture model.

(2) To compare the power of Likelihood Ratio Test, Spearman's Rank Correlation Test, and Simple Linear Regression Test on each simulated sample and to determine the most powerful test.

2.1.2 The design

In our study design, we assume that there is one control group which is a single standard normal distribution and there are three treatment groups where the difference in dose between treatment groups equals to the dose in the first treatment group. Thus, we consider the dose values X as 0, 1, 2, and 3. We consider the case where we have an equal number of individuals in each group, i.e. $n_i = n \; for \; i = 0, 1, 2, 3.$

Thus, the distribution for each group under the alternative hypothesis can be summarized as follows:

$$F(y|x_0) \equiv \varphi\left(\frac{y-\mu}{\sigma}\right)$$

$$F(y|x_1) = \pi\varphi\left(\frac{y-\mu-\beta}{\sigma}\right) + (1-\pi)\,\varphi\left(\frac{y-\mu}{\sigma}\right)$$

$$F(y|x_2) = \pi\varphi\left(\frac{y-\mu-2*\beta}{\sigma}\right) + (1-\pi)\,\varphi\left(\frac{y-\mu}{\sigma}\right)$$

$$F(y|x_3) = \pi\varphi\left(\frac{y-\mu-3*\beta}{\sigma}\right) + (1-\pi)\,\varphi\left(\frac{y-\mu}{\sigma}\right) \qquad (2.2)$$

In the simulation study, we generate samples under null hypothesis and under alternative hypothesis as stated above.

2.1.3 The Simulation

The simulation techniques provide empirical estimation of the sampling distribution of the parameters of interest that could not be achieved from a single study and enable the estimation of accuracy measures, such as the bias in the estimates of interest, as the truth is known. Hence, in our study, we simulate one data set as the control group and three data sets as treatment groups with different parameters. The data is generated as following:

(1) For the control group of size *n* (X=0) we generate all observations following a standard normal distribution using Box-Muller Transformation (1958) [30].

(2) For three treatment groups (with X=1, 2 or 3), each also of size *n*,, we generate an indicator which follows a standard normal distribution for each observed value. If the indicator is less than π, then the observed value is generated following a standard normal distribution, otherwise, it is generated following a normal distribution with a shifted mean $X * \beta$, $X = 1, 2, 3$.

Simulation studies usually examine the properties of one or more statistical methods in several scenarios defined by values of various factors such as sample size and proportion of censoring. These factors are generally examined in a fully factorial arrangement. The number of scenarios to be investigated and the methods for evaluation must be determined and justifications for these choices provided in the protocol. The scenarios investigated should aim to reflect the most common circumstances and if possible cover the range of plausible parameter values. In our situation the parameter settings considered under the alterative hypothesis are

$\mu = 0$

$\sigma = 1$

$\beta = 0.1, 0.3, 0.5$

$\pi = 0.1, 0.3, 0.5.$

The parameter settings that we consider under the null hypothesis is $\mu = 0, \sigma = 1, \beta = 0,$ $\pi = 0.$ The sample sizes that we consider are $n = 15, 30, and\ 50\ (or\ N = 60, 120, and\ 200).$

2.2 The Likelihood Ratio Test (LRT)

In the scenario of mixture model, the LRT test can be described as follows. Under the null hypothesis, the likelihood for an observation from the control group $(x_i = 0)$ is $\varphi\left(\frac{y-\mu}{\sigma}\right)$, which follows the distribution of $N(0,1)$. The likelihood for an observation from the treatment group

$(x_i > 0)$ is $F(y) = \pi\varphi\left(\frac{y-\mu-\beta*x_i}{\sigma}\right) + (1-\pi)\,\varphi\left(\frac{y-\mu}{\sigma}\right)$. Therefore, the likelihood for all observations under the null hypothesis is

$$L_0 = \prod_{control} F(y)\ \prod_{treatment} F(y) = \prod_{j=1}^{n_0} \varphi\left(\frac{y_{0j}-\mu}{\sigma}\right)\ \prod_{i=1}^{I}\prod_{j=1}^{n_i} \varphi\left(\frac{y_{ij}-\mu}{\sigma}\right) =$$
$$\prod_{i=0}^{I}\prod_{j=1}^{n_i} \varphi\left(\frac{y_{ij}-\mu}{\sigma}\right) = (2\pi\sigma^2)^{\frac{N}{2}} exp\left\{\frac{-\sum_{i=0}^{I}\sum_{j=1}^{n_i}(y_{ij}-\mu)^2}{2\sigma^2}\right\} \qquad (2.3)$$

The likelihood for all observations under the alternative hypothesis is

$$L_1 = \prod_{control} \varphi\left(\frac{y-\mu}{\sigma}\right) \prod_{treatment} F(y)$$
$$= \prod\nolimits_{j=1}^{n_0} \varphi\left(\frac{y_{0j}-\mu}{\sigma}\right) \prod\nolimits_{i=1}^{I}\prod\nolimits_{j=1}^{n_i} \pi\varphi\left(\frac{y_{ij}-\mu-\beta x_i}{\sigma}\right) + (1-\pi)\,\varphi\left(\frac{y_{ij}-\mu}{\sigma}\right)$$
$$= (2\pi\sigma^2)^{\frac{N}{2}} exp\left\{\frac{-\sum_{j=1}^{n_0}(y_{0j}-\mu)^2}{2\sigma^2}\right\} \prod\nolimits_{i=1}^{I}\prod\nolimits_{j=1}^{n_i}\left[\pi exp\left\{\frac{-(y_{ij}-\mu-\beta x_i)^2}{2\sigma^2}\right\}\right.$$
$$\left. + (1-\pi) exp\left\{\frac{-(y_{ij}-\mu)^2}{2\sigma^2}\right\}\right] \qquad (2.4)$$

We calculate $G^2 = -2ln\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \frac{L_0}{L_1}$ is the likelihood-ratio test of the null hypothesis — that the control and the treatment groups have the same single distribution — versus the alternative hypothesis — that there is a dose-response mixture distribution in the treatment groups.

2.3 Maximum Likelihood Estimation

2.3.1 The Expectation Maximization (EM) Algorithm

The EM Algorithm is widely used for the computation of the maximum likelihood estimates. The EM algorithm produces a sequence of monotonically increasing values with many desirable properties, such as the simplicity and generality of the associated theory. The detailed description of the convergence properties of the EM algorithm is given by Wu [22]. In many applications, the population component for individual observations is missing; hence the mixture models can be interpreted as solving the problem for an incomplete data situation.

Suppose that an observable y is represented as n observations $y = (y_1, y_2, \ldots, y_n)$, which is the incomplete data collected from a k-component normal mixture. Let $v_1 = (y_1, z_1),\ v_2 = (y_2, z_2), \ldots, v_n = (y_n, z_n)$ be the hypothetical complete data, where $z_{j1}, z_{j2}, \ldots, z_{jn}$ is an unobserved indicator vector whose components are all zero except for one equal to unity indicating the subpopulation from which the *j*th observation arises. Suppose further that there is a function linking the complete the data $v_1, v_2, \ldots, v_n$ with the incomplete data $y_1, y_2, \ldots, y_n$ such that $y = h(v)$. Since the parameters cannot be obtained through the maximization of the complete data likelihood function, the MLEs are obtained by maximizing the expectation of the incomplete data log likelihood function by manipulating the relationship between the complete data likelihood and incomplete data likelihood.

Suppose that the likelihood function of $y' = (y_1,\ y_2,\ \ldots,\ y_n)$ and $v' = (v_1,\ v_2, \ldots, v_n)$ are

$$G(y|\theta) = \prod_{i=1}^{n} g(y_i;\theta) = \prod_{i=1}^{n} \sum_{j=1}^{k} \pi_j\, f_j(y_i;\ \mu_j, \sigma_j, \beta_j) \qquad (2.5)$$

and

$$F(v|\theta) = \prod_{i=1}^{n} \prod_{j=1}^{k} \pi_j^{z_{ij}}\, f_j(y_i;\ \mu_j, \sigma_j, \beta_j)^{z_{ij}} \qquad (2.6)$$

respectively, where $\theta = (\pi,\ \mu, \beta, \sigma)$ is the unknown parameters and z_{ij} equals 1 if the *j*th observation arises from component *m* and 0 otherwise. The conditional likelihood function of v given y is given by

$$K(v|y,\theta) = \frac{F(v|\theta)}{G(y|\theta)} = \prod_{i=1}^{n} \frac{\prod_{j=1}^{k} \pi_j^{z_{ij}}\, f_j\left(y_i;\ \mu_j, \sigma_j, \beta_j\right)^{z_{ij}}}{g\,(y_i;\theta)} \qquad (2.7)$$

and the conditional expectation of the incomplete data log likelihood function, $\log G(y|\theta)$, is

$$L(\theta, \theta^t) = \ Q(\theta, \theta^t) - H(\theta, \theta^t) \qquad (2.8)$$

for fixed value θ^0, where $Q(\theta, \theta^0) = E\left(L_{com}(Z, Y, \theta|Y, \theta^0)\right)$, $\theta^0 = E\ (\log K(yv, \theta)v, \theta^0)$.

By applying Jensen's inequality to the convex function, it can be shown that the change in *H(.)* is negative, which indicates that $H(\theta, \theta^{t+1}) \leq H(\theta, \theta^{t})$, and the equality holds if the $K(v|y, \theta^{t+1}) = K(v|y, \theta^{t})$. Thus the maximizing $H(\theta, \theta^{t})$ is equivalent to maximizing $Q(\theta, \theta^{t})$.

The EM algorithm proceeds iteratively in two steps, the expectation (E) step and the maximization (M) step. Based on initial values for the parameters, the conditional expectation of the complete-data log likelihood function is calculated by giving the observed data. Suppose that θ^{t} is calculated at the *t*th iteration, then the approximation, θ^{t+1}, at the next iteration is obtained after the E step, calculation of $Q(\theta, \theta^{t})$, and an M step, determination of θ^{t+1} through the maximization of $Q(\theta, \theta^{t})$. The E step is completed simply by replacing z_{ij} by its conditional expectation given by y_i by using the current fit θ^{t} for θ.

Deriving the term of equation with respect to π, μ, σ, and β and rearranging the likelihood equations in the M step, we can get the likelihood equations. For the case in this thesis, likelihood equations and parameter estimates are given in detail in the following section.

2.3.2 The MLE based on the EM Algorithm

The estimation of the parameter $\theta = (\pi, \beta, \mu, \sigma)$ can be viewed as a missing data problem since the subpopulation from which individual observations are taken is not known. Suppose that $y_{01}, \cdots, y_{0n_0}$ are independent and identically distributed (I.I.D.) as control group and that $y_{i1}, \ldots, y_{in_i}, i = 1, 2, \ldots, \mathrm{I}$ are I.I.D. as treatment groups. Then (y_{ij}, z_{ij}), $j = 1, 2, \ldots, \mathrm{n_i}$, $i = 0, 1, 2, \ldots, \mathrm{I}, N = \mathrm{n_0} + \mathrm{n_1} + \mathrm{n_2} + \cdots + \mathrm{n_I}$, can be complete data, where y_{ij} is the observed quantitative measure and $z_{ij} = 1$ if y_{ij} is from the second component distribution in i^{th} treatment group; otherwise $z_{ij} = 0$. Based on the likelihood of observations, the complete log likelihood is:

$$
\begin{aligned}
\log L_c = \log \{ & \prod_{j=1}^{n_0} \varphi\left(\frac{y_{0j} - \mu}{\sigma}\right) \prod_{i=1}^{I} \prod_{j=1}^{n_i} \varphi\left(\frac{y_{ij} - \mu - \beta x_i}{\sigma}\right)^{z_{ij}} \varphi\left(\frac{y_{ij} - \mu}{\sigma}\right)^{1-z_{ij}} \pi^{z_{ij}} (1 \\
& - \pi)^{1-z_{ij}} \} \\
& = N \log \frac{1}{\sqrt{2\pi}} - \mathrm{N} \log \sigma + \log(1-\pi) \sum_{\mathrm{i}=1}^{\mathrm{I}} \sum_{\mathrm{j}=1}^{\mathrm{n_i}} (1 - z_{ij}) \\
& + \log \pi \sum_{\mathrm{i}=1}^{\mathrm{I}} \sum_{\mathrm{j}=1}^{\mathrm{n_i}} z_{ij} \\
& - \frac{1}{2\sigma^2} \sum_{\mathrm{i}=1}^{\mathrm{I}} \sum_{\mathrm{j}=1}^{\mathrm{n_i}} \left[(1 - z_{ij})(y_{ij} - \mu)^2 + z_{ij}(y_{ij} - \mu - \beta x_i)^2 \right] \\
& - \frac{1}{2\sigma^2} \sum_{\mathrm{j}=1}^{\mathrm{n_0}} (y_{0j} - \mu)^2
\end{aligned}
$$

(2.9)

The E-step and M-step of the algorithm correspond to calculating updated estimates $\theta^{t+1} = (\pi^{t+1}, \beta^{t+1}, \mu^{t+1}, \sigma^{t+1})$ given current values $\theta^t = (\pi^t, \beta^t, \mu^t, \sigma^t)$ by choosing θ^{t+1} as the

value of θ that maximizes $E(\log L_c(Z,Y,\theta|Y,\theta^t))$. Here $E(\log L_c(Z,Y,\theta|Y,\theta^t))$ is $\log L_c$ with z_{ij} replaced by the conditional expectation:

$$E(z_{ij}|y_{ij},\theta) = \frac{\pi\varphi\left(\frac{y_{ij}-\mu-\beta x_i}{\sigma}\right)}{\pi\varphi\left(\frac{y_{ij}-\mu-\beta x_i}{\sigma}\right)+(1-\pi)\,\varphi\left(\frac{y_{ij}-\mu}{\sigma}\right)} \qquad (2.10)$$

${z_{ij}}^t$ is the conditional expectation obtained at the t^{th} iteration. The maximum likelihood estimate of parameters are calculated as following,

First to estimate π, let $\frac{\partial E(\log L_c(Z,Y,\theta|Y,\theta^t))}{\partial \pi} = 0$, then:

$$-\frac{1}{1-\hat{\pi}}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(1-{z_{ij}}^t)+\frac{1}{\hat{\pi}}\sum_{i=1}^{I}\sum_{j=1}^{n_i}{z_{ij}}^t = 0 \qquad (2.11)$$

Hence

$$\hat{\pi} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i}{z_{ij}}^t}{N-n_0} \qquad (2.12)$$

Next to estimate μ and β, let $\frac{\partial E(\log L_c(Z,Y,\theta|Y,\theta^t))}{\partial \mu} = 0$, then:

$$-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left[2\left(y_{ij}-\hat{\mu}\right)\left(1-{z_{ij}}^t\right)(-1)+2(y_{ij}-\hat{\mu}\right.\right.$$
$$\left.\left.-\hat{\beta}x_i){z_{ij}}^t(-1)\right]-\frac{1}{2\sigma^2}\sum_{j=1}^{n_0}2\left(y_{0j}-\hat{\mu}\right)(-1)\right\}=0 \quad (2.13)$$

Also, let $\frac{\partial E(\log L_c(Z,Y,\theta|Y,\theta^t))}{\partial \beta} = 0$, then:

$$\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left[2(y_{ij}-\hat{\mu}-\hat{\beta}x_i){z_{ij}}^t(-1)\right]\right\}=0 \quad (2.14)$$

Substituting $\hat{\beta}$ by $\hat{\mu}$ into (2.22), we obtain:

$$\hat{\mu}=\frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i}[y_{ij}(1-{z_{ij}}^t)]+\sum_{i=1}^{n_0}y_{0j}}{N-\sum_{i=1}^{I}\sum_{j=1}^{n_i}{z_{ij}}^t} \quad (2.15)$$

Substituting $\hat{\mu}$ by (2.24) into (2.23), we obtain

$$\hat{\beta} = \frac{N\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}{z_{ij}}^t - \left(\sum_{i=1}^{I}\sum_{j=1}^{n_i} {z_{ij}}^t\right)\left(\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}\right) + \left(\sum_{i=1}^{I}\sum_{j=1}^{n_i} {z_{ij}}^t\right)\left(\sum_{i=1}^{n_0} y_{0j}\right)}{N\sum_{i=1}^{I}\sum_{j=1}^{n_i} x_i{z_{ij}}^t - \left(\sum_{i=1}^{I}\sum_{j=1}^{n_i} {z_{ij}}^t\right)\left(\sum_{i=1}^{I}\sum_{j=1}^{n_i} x_i{z_{ij}}^t\right)}$$

(2.16)

Finally to estimate of σ^2, let $\frac{\partial E(\log L_c(Z,Y,\theta|Y,\theta^t))}{\partial \sigma^2} = 0$.

$$-\frac{N}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3}\sum_{i=1}^{I}\sum_{j=1}^{n_i}[(y_{ij}-\hat{\mu})^2\left(1-{z_{ij}}^t\right) + {z_{ij}}^t(y_{ij}-\hat{\mu}-\hat{\beta}x_i)^2] + \frac{1}{\hat{\sigma}^3}\sum_{j=1}^{n_0}(y_{0j}-\hat{\mu})^2 = 0$$

(2.17)

or equivalently

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i}[(y_{ij}-\hat{\mu})^2\left(1-{z_{ij}}^t\right) + {z_{ij}}^t(y_{ij}-\hat{\mu}-\hat{\beta}x_i)^2] + \sum_{j=1}^{n_0}(y_{0j}-\hat{\mu})^2}{N}$$

(2.18)

2.3.3 Selection of Starting Values for the EM algorithm and Calculation of the Likelihood

It is very common that the likelihood functions of mixture models have multiple roots. A wide range of random starting points should be used to search for all possible local maxima while implementing the EM algorithm. If the number of starting points is not sufficient, the final MLE might be trapped in the local maximum. On the other hand, it is computationally expensive to use lots of starting points. Fowlkes [23] suggested the use of the point of an inflection in a Q-Q (Quantile-Quantile) plot of the Quantile of a two-component normal mixture versus the Quantile of a single normal as a starting value for the mixing proportion. Thode et al. [24] suggested using 1/4, ¾, 1/n and (n-1)/n as starting values for the mixing proportions. A set of good initial values for the unknown parameter vector is important to the calculation of the global maximum.

For the null hypothesized model, we first select the initial values for the unknown parameter $\theta = (\pi,\ \mu, \beta, \sigma)$ (Lo. [6]), and then calculate the Maximum Likelihood Estimates and the observed global maximum log likelihood using the EM algorithm, The algorithm to calculate Maximum Likelihood Estimates and the global maximum log likelihood is as follows:

(1) Generate a random value π from a uniform distribution on the interval (0,1), and assign it as the initial value of the mixing proportion for the null hypothesis model, $\hat{\pi}^0 = \pi$. Let $y_{11}, y_{12,\ldots} y_{1n}, y_{21}, y_{22,\ldots} y_{2n}, y_{31}, y_{32,\ldots} y_{3n}, y_{41}, y_{42,\ldots} y_{4n}$ be a sorted four-group random sample with sample size of each group of n. Given the value of π, the two components in each group has different number of observations, namely, $n_{i1} = [n * \pi]$, and $n_{i2} = n - n_{i1}$, $i = 0, 1, 2, 3$. Hence, the initial estimates based on the random generated proportion are computed as follows

$$\hat{\mu}^{(0)} = \frac{\sum_{i=1}^{4} \sum_{j=1}^{n_{i1}} y_{ij}}{4 * n_{i1}} \qquad (2.19)$$

$$\hat{\beta}^{(0)} = \frac{\sum_{i=1}^{3} \frac{\frac{\sum_{j=1}^{n_{i2}} y_{ij}}{n_{i2}} - \frac{\sum_{j=1}^{n_{i1}} y_{ij}}{n_{i1}}}{i}}{3} \qquad (2.20)$$

$$\sigma^{2(0)} = \frac{\sum_{i=0}^{4} \frac{\sum_{j=1}^{n_{i1}} (y_{ij} - \hat{\mu}^{(0)})^2}{n_{i1}} + \frac{\sum_{j=1}^{n_{i2}} (y_{ij} - \hat{\mu}^{(0)} - i * \hat{\beta}^{(0)})^2}{n_{i2}}}{8} \qquad (2.21)$$

(2) Then, 150 random values of the mixing proportions are generated from a uniform distribution on the interval (0, 1), and the values of the other parameters

are computed corresponding to each of these mixing proportions using the EM algorithm. The log likelihood of each of these 150 random starting points is calculated. We try to find the maximum of the 150 log likelihood results.

(3) When the maximum of the first two steps is found, the corresponding solution of the likelihood equations are the Maximum Likelihood Estimates of the parameters.

The completed procedure is presented in the following figure. It shows how the Maximum Likelihood Estimates are calculated using the EM algorithm. Based on the log likelihood obtained under the null and alternative models, the power of the LRT can be measured by comparing the observed value to the critical value. The null distribution and the power study are investigated in the following chapters.

Figure 2.2 Flow chart of the computation of the Maximum Likelihood Estimates for a random sample.

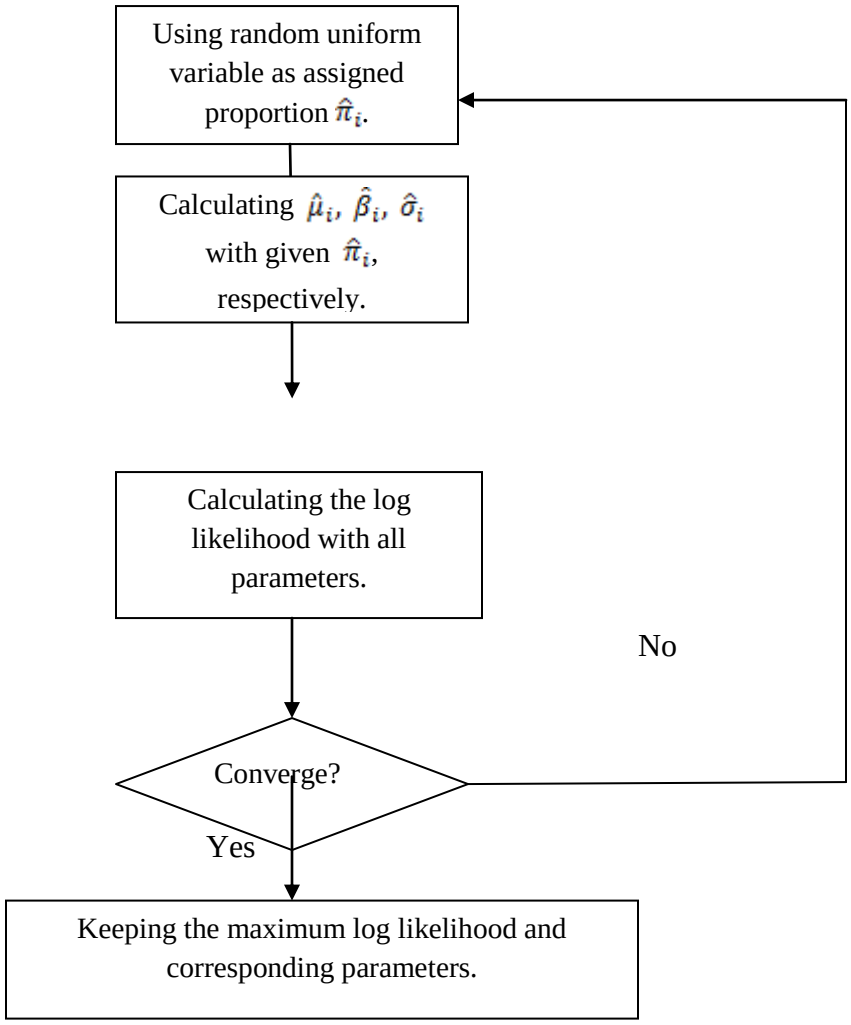


Table 2.1 The difference between maximum LRT statistics for selected numbers of Random Starting Points (RSPs) under the null and the alternative hypothesis for normal mixture models.

Sample Size	Number of RSPs	Under H_0	Under H_1						Average under H_1
			π=0.3		π=0.5		π=0.8		
			β=0.5	β=1	β=0.5	β=1	β=0.5	β=1	
N=60	1	0	0	0	0	0	0	0	0
	10	0.072680	0.01083	0.017211	0.109727	0.283031	0.002180	0.122301	0.090881
	40	0.045160	0.004588	0.019367	0.00632	0.000401	0.002809	0.001317	0.005800
	60	0	0	0	0	0.065545	0	0.006088	0.011938
	80	0	0.00097	0.005662	5.28E-07	0	0.000661	0	0.001216
	100	0	0.000102	0	0.00775	0.007369	1.17E-07	8.52E-09	0.00253
	150	0.001309	0	0.002730	0	0	0.003673	0	0.001067
	200	0.004238	0	0	0.00539	0.000314	0	0.000377	0.001015
N=120	1	0	0	0	0	0	0	0	0
	10	0.011504	0.053480	0.001431	0.001697	0.126705	1.31E-06	0.119730	0.050507
	40	0.00041	3.11E-05	0.013189	0.00035	0.007008	0.001498	0.045345	0.011238
	60	0	0	0.022129	4.31E-07	0	0	5.79E-09	0.003688
	80	0	0	0	0.00173	0	0	0	0.000289
	100	0	0	9.76E-06	4.69E-07	0	0.000108	0	1.99E-05
	150	0.002757	7.98E-05	0	0.00196	0	0.000364	1.13E-07	0.000401
	200	0	0.00023	0	0	0	8.304E-08	0.011418	0.001941
N=200	1	0	0	0	0	0	0	0	0
	10	0.002268	0.024714	0.270548	0.033095	0.386515	0.017788	4.58E-06	0.12111
	40	0.001358	0.002657	0.367722	0.025684	0.088161	2.64E-05	0.026992	0.08520
	60	9.65E-08	0.000183	0.29624	0.01450	0	0	4.06E-08	0.051822
	80	0	0	0.005594	0	0	0.00584	0.000655	0.002015
	100	0	0	0.062874	0	0	0	0	0.010479
	150	1.28E-09	0.000100	0.074427	0	0.070332	0.002245	0.004466	0.025262
	200	2.5E-06	8.14E-05	0	0	0	0.011235	0	0.001886

In our research, in order to find the appropriate number of random starting points to be used, we calculate different LRT statistics using different number of starting points. We calculate the average differences of LRT statistics when more random starting points are used for each sample size. When these differences become smaller and smaller, it means that the LRT statistics converges. Then the number of starting point is sufficient for our situation to find the global maximum of LRT statistics and its MLEs. From the table above, we can see that there is a big difference between using one random starting point and using multiple random starting points. Under the null hypothesis, the LRT statistics is quite stable and converge to the maximum when using 150 random starting points. Thus 150 random starting points are used to generate the LRT distribution under the null hypothesis. For the alternative hypothesis, the average differences are quite small between using 80 random starting points and 60 starting points. As more random starting points are used, the larger LRT statistics that we can obtain. However, the average differences of using more than 80 random starting points are not very significant. For the complexity of computation, 80 random starting points are used to generate the LRT distribution under the alternative hypothesis. Since the EM algorithm is customized to calculate the MLE for the mixture normal models, it is easier to use EM algorithm to calculate the estimates when there is a normal mixture in the distributions. Under the null hypothesis, most of the samples are single normal distribution, so it is relatively difficult for the EM algorithm to calculate the estimates. Thus more random starting points are used to calculate MLE under the null hypothesis than alternative hypothesis.

2.3.4 Simulated Results of MLE for Normal Mixture Model

The EM algorithm stated above is customized to fit the normal mixture model in the research. In order to verify the EM algorithm, we calculate the MLE for 1000 samples with different sample size and different parameters. Under the alternative hypothesis, 80 random starting points are used. The MLE, standard error of MLE, and the bias of MLE are shown in the table below. Three sample sizes $n = 15, 30, and\ 50$ are considered. Since we have four groups for each sample, the total number of observations are $N = 60, 120, and\ 200,$ respectively. For each sample size, we consider three different shifted proportion, $\pi = 0.3, 0.5, and\ 0.8$. Two possible shifted means are used, namely $\beta = 0.5\ and\ 1$. So there are eighteen scenarios for the combination of different parameters and different sample sizes. Table 2.2 shows the simulated mean MLEs with the standard error (in parentheses) under the normal mixture population based on 1,000 samples in each case. 150 random starting points are used to obtain the MLEs in this table.

Table2.2 Mean and standard error of the MLEs under the alternative hypothesis for different sample sizes (1000 random samples are used for each estimate).

					Mean MLEs (SE)			
	μ	σ	π	β	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\pi}$	$\hat{\beta}$
n=60	0	1	0.3	0.5	-0.015 (0.05)	0.941 (0.05)	0.308 (0.05)	0.593 (0.08)
				1	-0.017 (0.05)	0.974 (0.06)	0.318 (0.03)	1.021 (0.06)
			0.5	0.5	0.029 (0.05)	0.936 (0.05)	0.401 (0.05)	0.622 (0.06)
				1	0.010 (0.05)	0.962 (0.05)	0.48 (0.02)	1.027 (0.03)
			0.8	0.5	0.09 (0.05)	0.907 (0.05)	0.574 (0.05)	0.597 (0.04)
				1	0.069 (0.05)	0.985 (0.05)	0.737 (0.03)	1.007 (0.03)
n=120	0	1	0.3	0.5	0.005 (0.03)	0.985 (0.04)	0.300 (0.04)	0.581 (0.006)
				1	-0.015 (0.03)	0.978 (0.04)	0.314 (0.02)	1.009 (0.04)
			0.5	0.5	0.025 (0.04)	0.954 (0.04)	0.420 (0.04)	0.587 (0.04)
				1	0.006 (0.03)	0.966 (0.03)	0.480 (0.02)	1.028 (0.02)
			0.8	0.5	0.065 (0.04)	0.914 (0.03)	0.622 (0.04)	0.576 (0.04)
				1	0.050 (0.04)	0.977 (0.04)	0.743 (0.02)	1.010 (0.02)
n=200	0	1	0.3	0.5	0.002 (0.02)	0.981 (0.03)	0.284 (0.03)	0.570 (0.05)
				1	0.008 (0.02)	0.976 (0.03)	0.307 (0.01)	1.008 (0.02)
			0.5	0.5	0.02 (0.03)	0.960 (0.03)	0.424 (0.03)	0.570 (0.03)
				1	0.012 (0.02)	0.998 (0.03)	0.483 (0.01)	1.016 (0.02)
			0.8	0.5	0.037 (0.03)	0.979 (0.03)	0.762 (0.01)	1.008 (0.01)
				1	0.057 (0.01)	0.937 (0.01)	0.651 (0.02)	0.558 (0.01)

The table above shows the MLEs under the different combinations of parameters. In each scenario, we calculate the MLE of each sample and record the average of the 1000 samples. The standard error is calculated at the same time in order to show the variance of the MLEs. We can see that the average MLEs are quit close to the true value which means that the customized EM algorithm for normal mixture model works well and the number of random starting points is sufficient.

For the estimation of mean of non-responders μ, we can see that the estimates are all quite close to zero which is the true value of the parameter of each case. However, when the shifted proportion is large, the estimates are always bigger than those with small shifted proportion. So the estimate is biased when it comes to large shifted proportion. For the estimation of the common variance σ, the estimates are always very close with the true value of the parameter which is one. Under different scenarios as it shown in the table above, the estimates of σ are stable and variant within the small range. For the estimation of shifted proportion π, the estimates are more precise with the larger proportion. Since the larger shifted proportion makes the mixture part of the distribution more obvious, the EM algorithm is better for the calculation of the shifted proportion. For the estimate of the slope β, there are two true values are considered, namely 0.5 and 1. The estimates of the slope are more precise when the true value of the parameter is larger. Overall, the estimate of the parameter value 1 is better than those of the parameter value 0.5. The reason is that the larger slope makes the mixture part more obvious which is the same reason that why the estimate of larger shifted proportion is more accurate.

Overall, the MLE estimates are quite close to the true parameter value with relatively small standard errors

Chapter 3

The Null Distribution of the Likelihood Ratio Test Statistics

3.1 Introduction

In the chapter, we explore the null distribution for the LRT statistics and try to find the asymptotic null distribution and its fitted distribution. In section 3.2, we show the empirical distribution by simulating the null model with different sample sizes. We compare the empirical distribution with chi-square with 1 degree of freedom and a chi-square distribution with 2 degree of freedom. In section 3.3, we fit the null distribution as the combination of a proportion of zero LRT statistics and another proportion of chi-square distributed LRTs. The method suggested by Wilson and Hilferty [22] is used to transform non-zero LRTs into a standard normal distribution, thus it is much easier to find the cumulative density function of the LRT statistics. Finally, the type I error rate is calculated with different sample sizes to assess the validity of the proposed fitted distribution.

3.2 Empirical Null Distribution of LRT Statistics

The empirical null distribution of the LRT statistics was obtained through simulation. We compared the theoretical asymptotic distribution with the empirical null distribution found from the simulation. The 90^{th}, 95^{th}, and 99^{th} percentile of the empirical null distribution of the LRT statistics and corresponding chi-square values with same confidence intervals are computed.

We simulate samples with three different sample size which is showed in the table following. Each null distribution is based on 10,000 simulated samples from the null distribution. The mean and variance of the distribution is calculated, and the selected percentiles of LRT statistics are reported.

Table 3.1 Summary of empirical null distribution of LRT statistics under the null hypothesis of normal mixture models. (10,000 samples for each sample size)

	Sample size	Mean	variance	Percentile of LRT		
				90^{th}	95^{th}	99^{th}
Empirical Null	n=60	0.7698	1.9797	2.4172	3.6420	6.5375
	n=120	0.8376	2.1383	2.6021	3.8550	6.7275
	n=200	0.8524	2.2066	2.6059	3.8185	6.8715
x_1^2		1	2	2.7060	3.8410	6.6350
x_2^2		2	4	4.6050	5.9910	9.2100

Note 1. The number of observations per x values $n_x = n / 4$ for each sample size n. Groups have values of $x = 0,1,2, and\ 3$.

From the table above, we can see that the mean of the LRT statistics becomes larger as the sample size becomes larger. The variance also becomes bigger with large sample size, however, the trend is not quite obvious. The table shows selected percentiles of LRT statistics as well as the corresponding chi-square statistics. We can see that, for the 90^{th} percentile statistics, the critical values are all less than the chi-square with degree of freedom of one. For the 95^{th} percentile statistics, the critical values are less than the chi-square with degree of freedom of one except that the statistics for sample size of 120 is a little larger than the corresponding chi-square statistics. For the 99^{th} percentile statistics, the critical value of sample size of 120 and 200 are between chi-square with degree of freedom of one and the chi-square statistics with degree of freedom of two. However, they are a lot less than the chi-square statistics with degree of freedom of two. For the summary of the table, we can see that the asymptotic distribution of $0.5\chi_1^2 + 0.5\chi_2^2$ suggested by the work of McLachlan, G. J. (2000) does not hold here since most of the LRT percentiles are less than the corresponding chi-square statistics with degree of freedom of one.

3.3 Fitted Null Distribution of the LRT Statistics

In this section, we attempt to fit the empirical null distribution into combination of chi-square distribution for each sample size. We transform the empirical null distribution so as to consider the proposition that the asymptotic distribution is a mixture of chi square with 0 degrees

of freedom and chi square with some positive number of degrees of freedom. We consider the proposition then that the null distribution of the LRT for a given value of n is of the form $p\chi_0^2 + (1-p)\chi_v^2$. Then, upon applying the result of Wilson and Hilerfy (1931), the distribution of the cube root of the statistic would be a mixture of a fraction of zero with a proportion of *p* and an approximately normal distribution. We then estimate the corresponding x^2 mixture distribution $p\chi_0^2 + (1-p)\chi_v^2$ by first conditioning on whether the value of the statistic is zero or not as follows:

(1)We estimate the *p* with the observed fraction of values equal to 0.0;

(2) Conditional on the statistic having a non-zero value (note that the probability that a x^2 distribution has the value of 0 is zero), we estimate the mean, M by $\bar{x}$, where $\bar{x}$ is the average of the non-zero LRT statistics; and

(3) we estimate the standard deviation, with s, the sample estimate based on the non zero values. The fitted estimate of the cumulative probability is calculated as

$$\Pr(lrt < L) = 0 \; for \; L < 0$$

$$= p \; for \; L = 0$$

$$= p + (1-p)\emptyset\left(\frac{(L)^{1/3} - M}{s}\right) \qquad (3.1)$$

where $\emptyset$ is the cumulative distribution function of an $N(0,1)$ variable.

Equivalently the P-value associate with an observed LRT statistics can be approximated by

$$P = \Pr(lrt \geq L) = (1-p)\left(1 - \emptyset\left(\frac{(L)^{1/3} - M}{s}\right)\right)$$

$$for\ L > 0. \qquad (3.2)$$

Similarly the critical value for an α level test is calculated as

$$M + \emptyset^{-1}(Z_{\alpha*}) * S \quad where \quad \alpha* = {}^{\alpha}/_{(1-p)} \qquad (3.3)$$

Table 3.2 shows the selected critical values based on the fitted distribution.

Table 3.2 Summary of empirical null distribution of LRT statistics under the null hypothesis and the corresponding fitted distribution.

	Sample size	Mean	variance	Percentile of LRT			Prob(0's)	Cube root of non-zeroes	
				90th	95th	99th		mean	SE
Empirical Null	n=60	0.7698	1.9797	2.4172	3.642	6.5375	0.3634	0.8811	0.4373
	n=120	0.8376	2.1383	2.6021	3.855	6.7275	0.3151	0.8782	0.4414
	n=200	0.8524	2.2066	2.6059	3.8185	6.8715	0.3131	0.8852	0.4414
Fitted Distribution	n=60			2.3059	3.3751	6.0523			
	n=120			2.4264	3.5148	6.2452			
	n=200			2.4652	3.5636	6.3100			

The 90^{th} percentile, 95^{th} percentile, and 99^{th} percentile critical values are calculated, respectively, based on the fitted distribution and are showed in the table compared with the empirical critical value. From the table above, we can see that the critical values of the fitted distribution become larger as the sample size increase, which is the same trend of the empirical distribution. There are certain differences between the critical values of the fitted distribution and those of empirical values. We would expect that the critical values of a good fitted distribution would be closer to the empirical values as the sample size increase. However, in the table above, the differences between the critical values of the fitted distribution and those of empirical distribution do not change as the sample size increases. The differences remain around 0.5 for all sample sizes of 99^{th} percentile level. The differences remain around 0.2 for all sample sizes for the 95^{th} percentile and 90^{th} percentile. Thus, we conclude that the fitted distribution does not fit the empirical distribution well. We explore the type I error for the empirical distribution and the fitted distribution in the followings. The normality test for the cube root of non-zero LRT statistics is also provided to verify our conclusion, which is showed in Table 3.3.

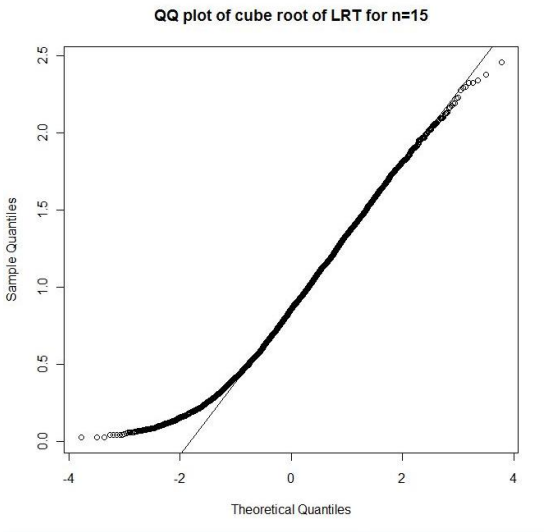
Table 3.3 Summary of Type I error for empirical and fitted distribution with different critical values. 1000 sample under the null distribution for each sample size are used.

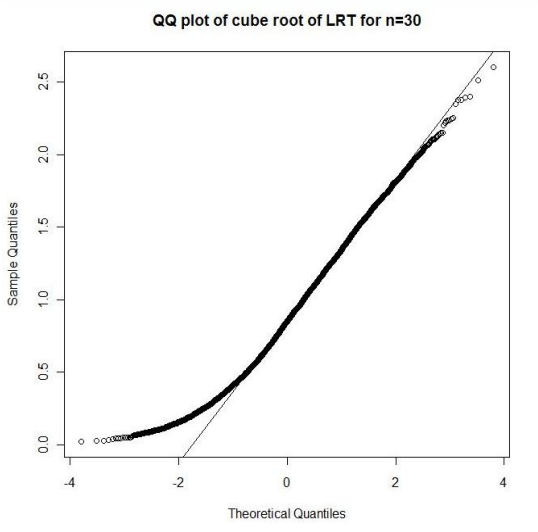
Sample Size		90^{th} percentile critical value	95^{th} percentile critical value	99^{th} percentile critical value
n=60	Empirical Distribution	0.097	0.047	0.008
	Fitted Distribution	0.104	0.055	0.011
n=120	Empirical Distribution	0.089	0.043	0.007
	Fitted Distribution	0.101	0.056	0.009
n=200	Empirical Distribution	0.102	0.057	0.005
	Fitted Distribution	0.109	0.062	0.008

We random generate 1000 samples under the null hypothesis in order to calculate the type I error for the empirical null distribution and the fitted null distribution, respectively. From the table above, we can see that the type I error which is calculated based on the empirical distribution is relatively smaller than that calculated based on the fitted distribution. At the same time, the table shows that the type I error of the fitted distribution are close to the significance level of the critical value, which shows that the fitted distribution works well. On the other hand, most of the type I error calculated based on the empirical distribution are smaller than the significance level of the critical value, which shows that the empirical null distribution based on 10,000 random samples works fairly well.

We conduct normality check on the cube root of the LRT statistics. We apply the Kolmogorov-Smirnov normality test to each sample size and construct the QQ plot for each case, which is showed in Figure 3.1. It looks like the fitted null works well even though it departs significantly from normality. The problem is that it is specific to this design (i.e. with equal number of observations at each dose and an equal number of untreated patients).

Figure 3.1 QQ plot for normality check of the non-zero LRTs. (n=60, 120, and 200)





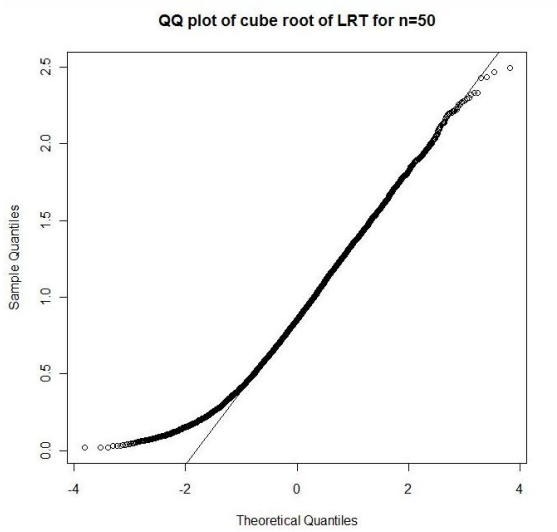


Table 3.4 – Kolmogorov-Smirnov(K-S) normality test for different sample sizes (n=60, 120, and 200)

Sample Size	D statistics	P-value
n=60	0.0395	< 2.2e-16
n=120	0.0452	< 2.2e-16
n=200	0.0363	< 2.2e-16

From the table above, we can see that, for each sample size, the p-values of the K-S test are extremely small. Thus, we reject the null hypothesis to conclude that the cube roots of the non-zero LRT statistics are not normally distributed. The fitted distribution obtained above is calculated by transforming the cube root of the non-zero LRT statistics into normally distributed variables. Although the cube roots of the LRT statistics are not normally distributed, the approximation fits the empirical distribution well.

Chapter 4

Power Study of the Likelihood Ratio Test in the Presence of Mixtures in the Treatment Groups

4.1 Single Variable Linear Regression

In this chapter, we conduct a power comparison of three methods which are LRT, Spearman's correlation, and single variable linear regression for each sample that we generate. We also derive the approximate variance for single variable linear regression which is used when calculating the approximate power of single variable linear regression.

In this research, we consider the situation where mean response has a linear relationship with dose in a fraction rather than all of the observations in the treatment groups. That is, shift in mean from the control value is dose related. Thus there is an overall linear increase in mean response with each dose, however the increase depends on two parameters (1) the mixing proportion or the proportion of responders in each treatment group and (2) the slope or the increase in mean response per unit increase in dose. Additionally, because of the mixture, the variance of the observations in each treatment group increases as the dose level increases as well.

One of the most important assumptions of linear regression analysis is that the residual variance is a constant. For this reason, we derive the variance in each treatment group and develop the expressions for the approximate power for the linear regression test.

Assume that Y is the dependent variable which is the observed value in each group. X is the dependent variable which is the dose level for each group. For the control group, x=0, and for the treatment group, x=1,2,3, respectively. Thus, the variance of the dependent variable based on the dose level X is

$$Var(Y|X) = E[(Y|X)^2] + [E(Y|X)]^2 \quad . \qquad (4.1)$$

Furthermore we can easily show that $E(Y|X) = \pi X\beta$. Here E(Y|X) denotes the expected value of the independent variable based on the dose level. Thus, the main point of derivation is to derive an expression for the $E[(Y|X)^2]$. We divide the observations in one group into two parts, the shifted proportion and the un-shifted proportion, for calculation, respectively. Thus,

$$E[(Y|X)^2] = (1-\pi) * E[(Y_1|X)^2] + \pi * E[(Y_2|X)^2] \qquad (4.2)$$

where Y_1 denotes the value the response given a subject is in the group having the same mean as the control group e, and Y_2 denotes the values of the response given a subject is in the group with the shifted mean.

The calculation of $E[(Y_1|X)^2]$ is simple in our case, since

$$E[(Y_1|X)^2] = Var\ (Y_1|X) + [E(Y_1|X)]^2 = \sigma^2 + [E(Y_1|X)]^2 \quad (4.3)$$

For the calculation of $E[(Y_2|X)^2]$, we transform the Y_2 into Y_1. Thus,

$$E[(Y_2|X)^2] = E[(Y_1 - \beta|X)^2]$$

$$= E\left[\left({Y_1}^2 - 2 * \beta * Y_1 + \beta^2|X\right)\right]$$

$$= E\left({Y_1}^2|X\right) + 2 * \beta * E(Y_1) + E(\beta^2|X)$$

$$= Var(Y_1|X) + [E(Y_1|X)]^2 + X^2 * \beta^2$$

$$= \sigma^2 + X^2 * \beta^2 \quad (4.4)$$

Based on the derivation above, we obtain that

$$E[(Y|X)^2] = \sigma^2 + \pi * X^2 * \beta^2 \qquad (4.5)$$

Substituting the quantity of equations above into equation (4.1), we obtain the variance of the observed values Y in each group with dose level of X

$$Var(Y|X) = \sigma^2 + \pi * (1 - \pi) * X^2 * \beta^2 \qquad (4.6)$$

Also if we redefine $\beta^* = \beta/\sigma$, we have

$$Var(Y|X) = \sigma^2(1 + \pi * (1 - \pi) * X^2 * \beta^{*2}) \qquad (4.7)$$

Based on the equation (4.7), we obtain the conditional variance of Y each group by substituting the corresponding X.

Table 4.1 Variance of the observed value Y of each group with corresponding dose level X.

	Dose Level (X)	Variance of Y
Control Group	X=0	σ^2
Treatment Group 1	X=1	$\sigma^2(1+\pi*(1-\pi)*\beta^{*^2})$
Treatment Group 2	X=2	$\sigma^2(1+\pi*(1-\pi)*4*\beta^{*^2})$
Treatment Group 3	X=3	$\sigma^2(1+\pi*(1-\pi)*9*\beta^{*^2})$

For the approximate power calculation for the linear regression, we use the average of the variances of the four groups, which obtained from the table above, as the common variance of observed value Y conditioning on the dose level X. Thus, in the case where X=0,1,2 and 3 with equal frequency , the approximate within group variance is

$$Var(Y) = \frac{[Var(Y|X=0)+Var(Y|X=1)+Var(Y|X=2)+Var(Y|X=3)]}{4}$$

$$= \sigma^2(1+\frac{7}{2}*\pi*(1-\pi)*\beta^{*^2}) \qquad (4.8)$$

Thus, the properties of single variable linear regression customized to the mixture model are

$$E(\hat{B}) = \pi*\beta \qquad (4.9)$$

$$Var(\hat{B}) \cong \frac{Var(Y)}{\sum_{i=0}^{I} n_i * (x_i - \bar{x})^2} \qquad (4.10)$$

In this research, we consider four groups with same number of observations, thus, we have the common sample size n. The possible values of X are 0, 1, 2, 3. So the denominator of the equation (4.10) can be simplified. By substituting the common variance at the same time, we have

$$Var(\hat{B}) \cong \frac{\sigma^2(1 + \frac{7}{2} * \pi * (1 - \pi) * \beta^{*^2})}{5 * n} \qquad (4.11)$$

Based on the linear regression, we have

$$\frac{\hat{B} - E(\hat{B})}{\sqrt{Var(\hat{B})}} \sim N(0,1) \qquad (4.12)$$

where $E(\hat{B})$ and $Var(\hat{B})$ are the expected value slope and the variance of the slope of linear regression, respectively.

Thus, the approximate power of linear regression can be calculated as

$$Prob\left(z > \frac{Z_{\alpha/2} * SE(\hat{B}) - E(\hat{B})}{SE(\hat{B})}\right) \qquad (4.13)$$

which is probability that we reject the null hypothesis that there is a linear relationship between the observed values Y and the dose level X. Here, $Z_{\alpha/2}$ represents the two-sided critical value of the standard normal distribution given the significant level of α.

4.2 Result of Power comparison

In this research, we evaluate the performance of the Likelihood Ratio Test (LRT) by comparing its power with Spearman's Correlation Test, and Simple Linear Regression Test. The power of the LRT is based on using the critical values obtained from the simulated distribution under the null hypothesis given in the Chapter 3. For the power of Simple Linear Regression, we calculate the observed power and the approximate power (based on the results of the previous section, respectively.

For the data simulation, we consider the group sizes of 15, 30, and 50, which means the corresponding total sample sizes considered are 60, 120, and 200. The parameters that we consider are listed as follows:

$$\pi = 0.1,02, 0.5$$

$$\beta = 0.1, 0.3, 0.5$$

For each combination of parameter values and sample size, we generate 1,000 random samples under the alternative hypothesis and calculate the rate of rejection as the power of the test. The tables below shows the power of the three tests calculated from each sample under different significance levels.

Our interest lies in determining whether the LRT has essentially the same power distribution as Spearman's Correlation Test and the Simple Linear Regression Test. Since we execute the three tests for each simulated sample, the power comparison between the LRT and the Spearman's Correlation, as well as the comparison between LRT and Simple Linear Regression Test, are matched pair design. Thus, we apply the McNemar's Test to compare the power.

McNemar's test assesses the significance of the difference between two correlated proportions, such as might be found in the case where the two proportions are based on the same sample of subjects or on matched-pair samples. In our study, the correlated proportions are the proportion of rejections for a given sample size and set of parameter values. We conduct McNemar's test on LRT and Spearman's correlation test, and LRT and Simple linear regression test, respectively.

Table 4.2 Power comparison of LRT, Spearman's test, and Simple Linear Regression with significance level of 0.1.

	π	β	LRT	Spearman's Test	Linear Regression (Observed power)	Linear regression (Approximate Power)	Difference (Observed Power and Approximate Power)
n=60	0.1	0.1	**0.133**	0.114	0.116	0.060	0.056
		0.3	**0.143**	0.11**	0.112*	0.082	0.030
	0.2	0.3	**0.205**	0.139**	0.137**	0.128	0.009
	0.5	0.5	**0.7287**	0.579**	0.607**	0.624	0.017
n=120	0.1	0.1	**0.117**	0.101	0.114	0.064	0.050
		0.3	**0.189**	0.12**	0.121**	0.100	0.021
	0.2	0.3	**0.277**	0.166**	0.182**	0.177	0.005
	0.5	0.5	**0.9199**	0.852**	0.867**	0.870	0.003
n=200	0.1	0.1	**0.119**	0.103	0.117	0.068	0.049
		0.3	**0.19**	0.129**	0.13**	0.120	0.010
	0.2	0.3	**0.36**	0.251**	0.268**	0.236	0.032
	0.5	0.5	**0.979**	0.957**	0.972	0.974	0.002

Note: Significantly different in power compared to LRT using McNemar's Test (* 0.05; ** 0.01)
Approximate power of linear regression is calculated using the expressions in section 4.3.3.

Table 4.3 Power comparison of LRT, Spearman's test, and Simple Linear Regression with significance level of 0.05.

	π	β	LRT	Spearman's Test	Linear Regression (Observed Power)	linear regression (Approximate Power)	Difference (Observed Power and Approximate Power)
n=60	0.1	0.1	**0.056**	0.048	0.056	0.030	0.026
		0.3	**0.075**	0.046**	0.057	0.044	0.013
	0.2	0.3	**0.118**	0.08**	0.084**	0.073	0.011
	0.5	0.5	**0.6306**	0.451**	0.476**	0.500	0.024
n=120	0.1	0.1	**0.055**	0.052	0.056	0.033	0.023
		0.3	**0.094**	0.051**	0.057**	0.055	0.002
	0.2	0.3	**0.175**	0.100**	0.109**	0.107	0.002
	0.5	0.5	**0.867**	0.773**	0.801**	0.792	0.009
n=200	0.1	0.1	**0.057**	0.056	0.057	0.036	0.021
		0.3	**0.118**	0.077**	0.083*	0.068	0.015
	0.2	0.3	**0.249**	0.157**	0.169**	0.151	0.018
	0.5	0.5	**0.951**	0.925**	0.938	0.948	0.010

Note: Significantly different in power compared to LRT using McNemar's Test (* 0.05; ** 0.01);
Approximate power of linear regression is calculated using the expressions in section 4.3.3.

Table 4.4 Power comparison of LRT, Spearman's test, and Simple Linear Regression with significance level o f 0.01.

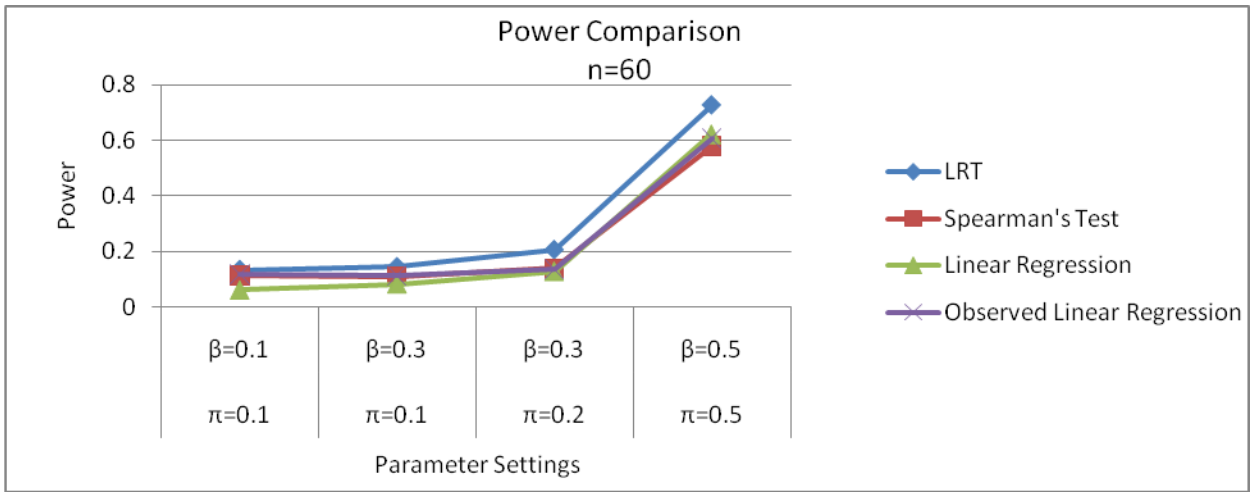
	π	β	LRT	Spearman's Test	Linear Regression (Observed Power)	linear regression (Approximate Power)	Difference (Observed Power and Approximate Power)
n=60	0.1	0.1	**0.013**	0.007	0.008	0.006	0.002
		0.3	**0.019**	0.013	0.013	0.010	0.003
	0.2	0.3	**0.044**	0.019**	0.02**	0.019	0.001
	0.5	0.5	**0.3654**	0.209**	0.24**	0.269	0.025
n=120	0.1	0.1	**0.011**	0.008	0.011	0.006	0.005
		0.3	**0.02**	0.011	0.01*	0.013	0.003
	0.2	0.3	**0.061**	0.027**	0.03**	0.032	0.002
	0.5	0.5	**0.767**	0.465**	0.582**	0.578	0.004
n=200	0.1	0.1	**0.01**	0.013	0.013	0.008	0.005
		0.3	**0.029**	0.017*	0.02	0.018	0.002
	0.2	0.3	**0.089**	0.047**	0.045**	0.049	0.004
	0.5	0.5	**0.881**	0.782**	0.837**	0.842	0.005

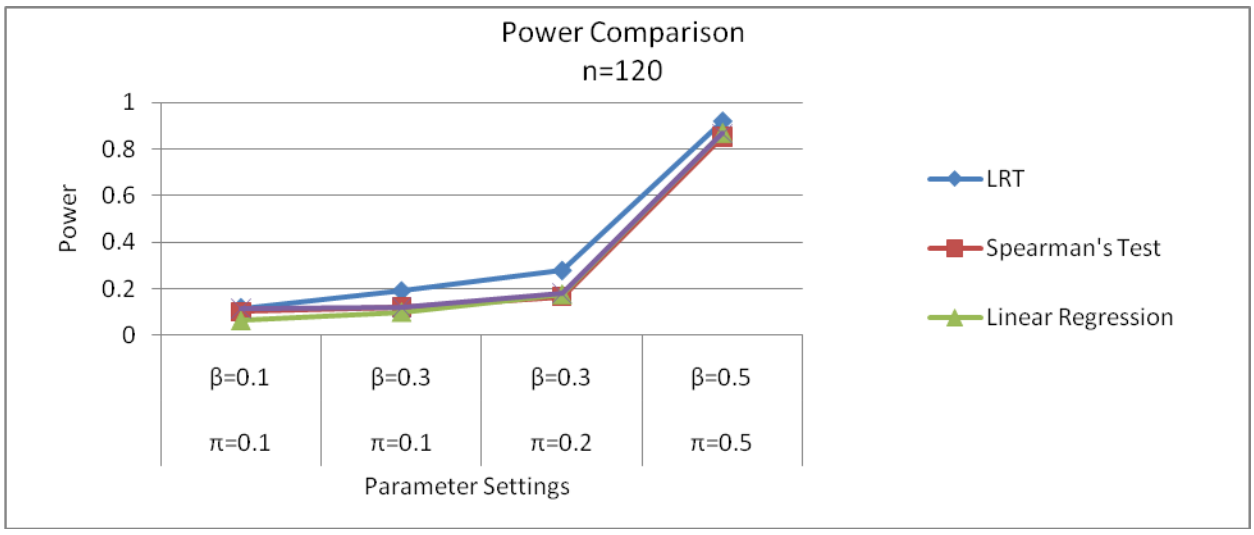
Note: Significantly different in power compared to LRT using McNemar's Test (* 0.05; ** 0.01);
Approximate power of linear regression is calculated using the expressions in section 4.3.3.

From the tables of power with different significance levels, we can see that the power of the LRT is greater than those of Simple Linear Regression and those of Spearman's Correlation Test. The power of the LRT is very low with the small shifted proportion and the small shifted mean are both small, since it is very difficult to detect the shifted part of the observation. However, the LRT turns out to be very powerful with larger parameter values (i.e. large mixing proportion or large value of slope) with each significance level. At the same time, the power of the LRT increases as the sample size becomes larger and larger. With the group size of 50, the power of LRT is 0.951 with the significance level of 0.05. Thus, we expect that the power will be higher with greater parameter values or with bigger sample size.

From the McNemar's test result above, we can see that the Spearman's correlation test and the linear regression test are significantly different from the LRT test wherever there is a large shift in mean and in mixing proportion. For the small parameter settings, the three tests do not work that well, since the power is low. However, for the larger parameter settings, the power of LRT is significantly higher than Spearman's correlation test and linear regression test which we can see from the McNemar's test result. The following figures show the trend of the power of the three tests with different parameter settings at different significance levels.

Figure 4.1 Power comparison at significance level of 0.1. (4 sets of parameter setting. 1000 samples are used for power calculation.)





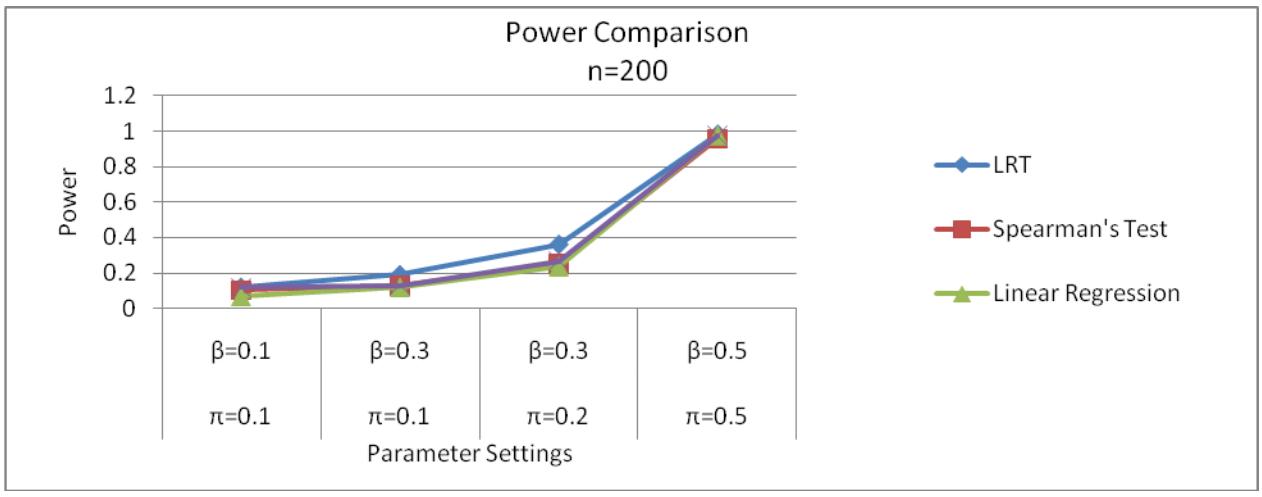
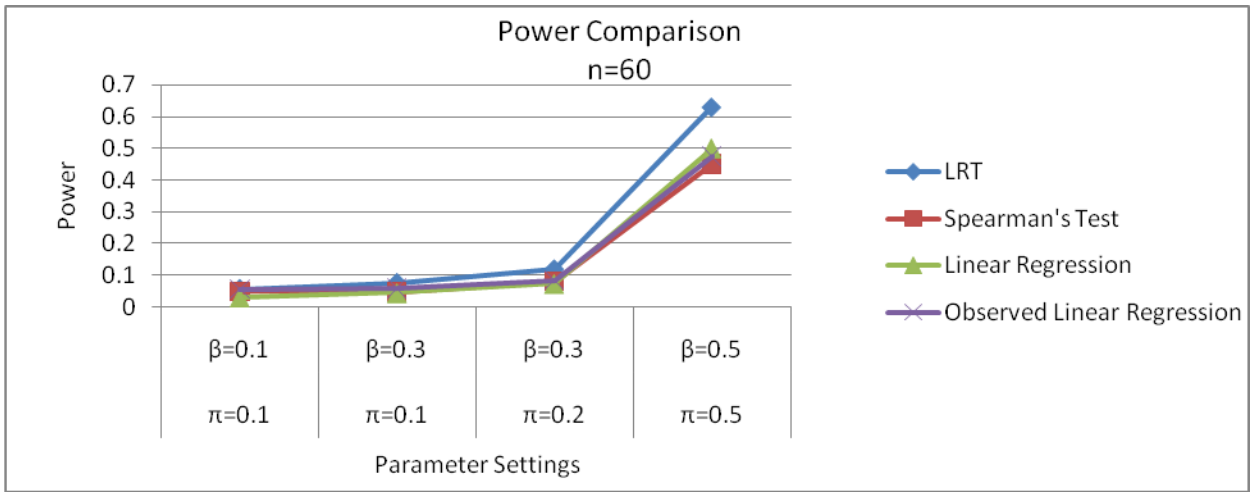
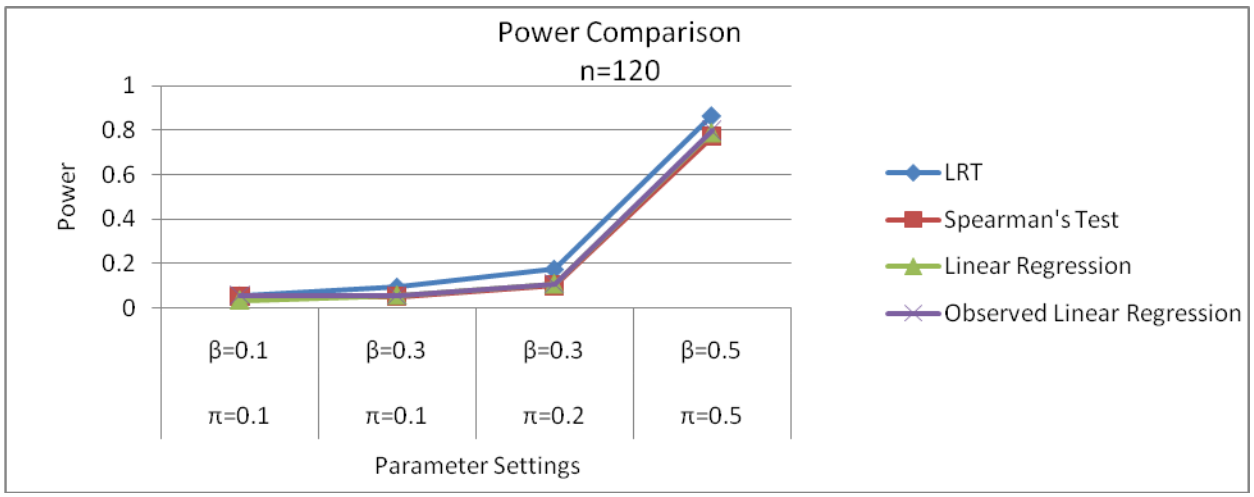


Figure 4.2 Power comparison at significance level of 0.05. (4 sets of parameter setting. 1000 samples are used for power calculation.)





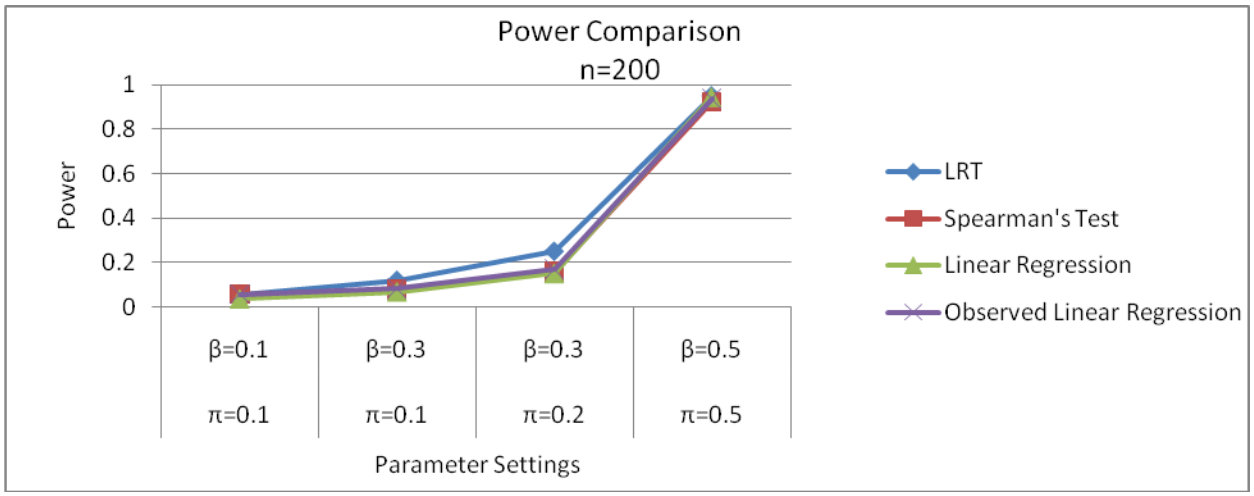
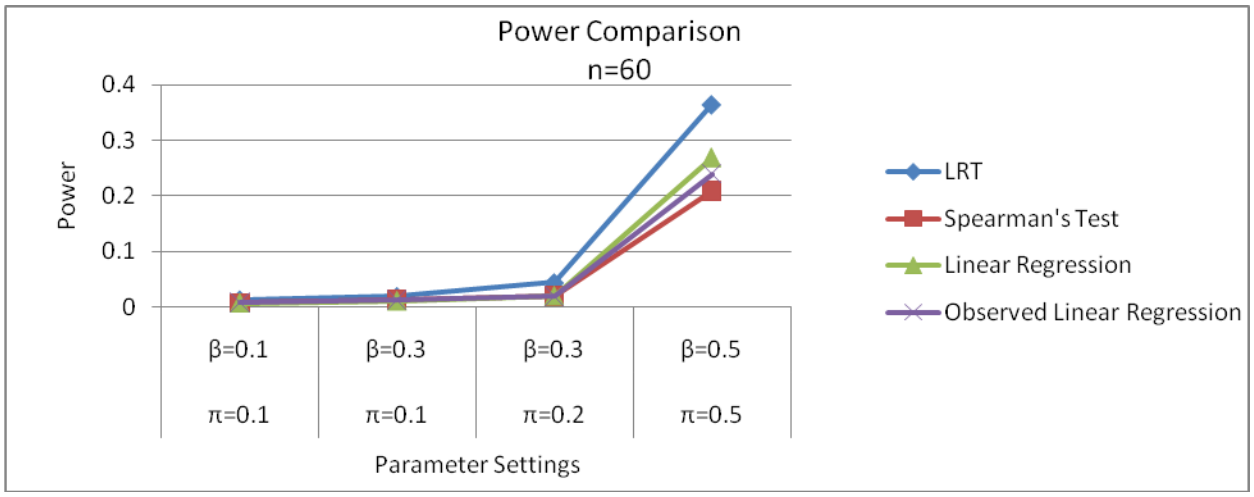
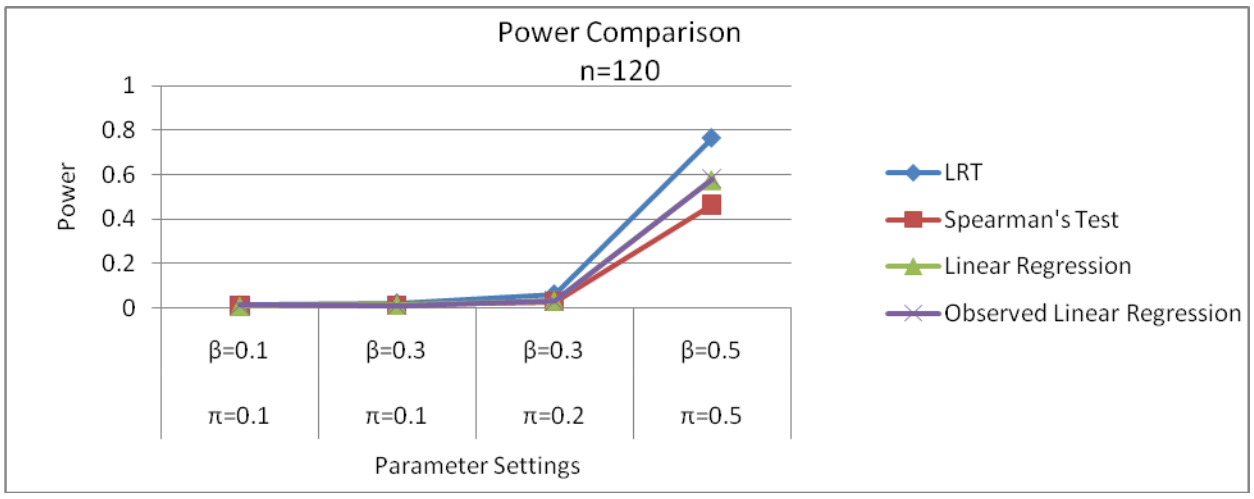
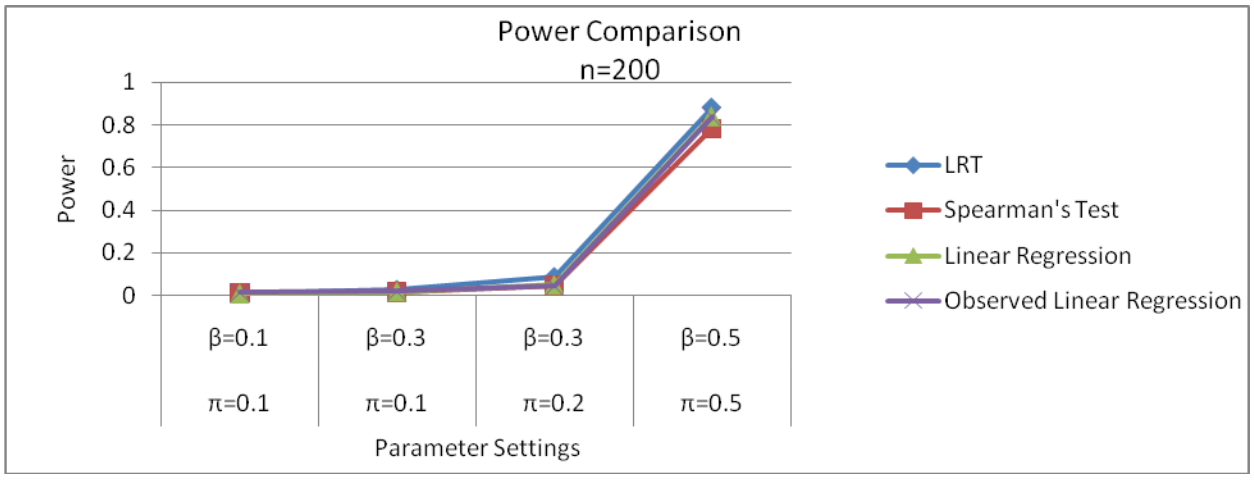


Figure 4.3 Power comparison at significance level of 0.01. (4 sets of parameter setting. 1000 samples are used for power calculation.)







4.3 Mean Square Error (MSE) Calculation.

For the LRT and Single variable linear regression test, we obtain the estimate of the parameters and compare them in the following table.

Table 4.5 MSE of the parameters from LRT and Single variable linear regression.

	π	β	$Var(\hat{\beta})$	$Bias(\hat{\beta})$	$MSE(\hat{\beta})$	Observed $Var(\hat{B})$	Approximate $Var(\hat{B})$	$Bias(\hat{B})$	Observed $MSE(\hat{B})$	Approximate $MSE(\hat{B})$
n=60	0.1	0.1	0.3279	0.1089	0.3397	0.0142	0.0134	0.09	0.0223	0.0215
	0.1	0.3	0.2798	0.0527	0.2825	0.0140	0.0137	0.27	0.0869	0.0866
	0.2	0.3	0.2144	0.0611	0.2181	0.0134	0.0140	0.24	0.0710	0.0716
	0.5	0.5	0.0500	0.0121	0.0501	0.0166	0.0163	0.25	0.0791	0.0788
n=120	0.1	0.1	0.2445	0.0591	0.2480	0.0068	0.0067	0.09	0.0149	0.0148
	0.1	0.3	0.2362	0.0287	0.2370	0.0069	0.0069	0.27	0.0798	0.0798
	0.2	0.3	0.1620	0.0325	0.1630	0.0070	0.0070	0.24	0.0646	0.0646
	0.5	0.5	0.0306	0.0071	0.0307	0.0087	0.0081	0.25	0.0712	0.0706
n=200	0.1	0.1	0.2290	0.0923	0.3075	0.0043	0.0040	0.09	0.0124	0.0121
	0.1	0.3	0.1815	0.0321	0.1825	0.0042	0.0041	0.27	0.0771	0.0770
	0.2	0.3	0.1560	0.0445	0.1579	0.0043	0.0042	0.24	0.0619	0.0618
	0.5	0.5	0.0169	0.0006	0.0169	0.0050	0.0048	0.25	0.0675	0.0673

Note: $\hat{\beta}$ is the estimate slope in the LRT; $\hat{B}$ is the estimate of the slope in the single linear regression test. $E(B) = \pi * \beta$. π is the proportion of responders in the mixture model.

We see on comparing the precision of LRT and the single variable linear regression that, in particular, the variance of the MLE of $\hat{\beta}$ is greater than that of single variable linear regression. This is the case except for those situations where either π or β is large, i.e. $\pi = 0.5$ or $\beta = 0.5$. However, the MLE of the LRT has minimum MSE in several cases with large π or large β. As expected, the MSE of LRT decreases as the sample size become larger. We can see from the table that the approximate variance of single variable linear regression is quite close with the observed one. Thus, the approximate calculation of single variable linear regression can be a fairly accurate estimate.

Chapter 5

Discussion and Conclusions

In this study, three methods are considered to test for an ordered association between dose and response in the case where there is a mixture normal distribution in the treatment groups. These are the Likelihood ratio test, Spearman's correlation test, and the single variable linear regression test. We consider the case where we have four equal size groups of which one is the control group and the other three are treatment groups with the dosage increase equaling the dose of the first treatment group. The null hypothesis of the test is that there is no relation between the dose level and mean response in the treatment groups. The alternative hypothesis of the test is that there in a fraction of the treatment groups there linear relationship between the shifted means and the dose level. For each parameter setting, the power of the three tests is calculated and the McNemar's test is conducted to test if the rejection rates of the methods are the same.

For the Likelihood ratio test, we customized the EM algorithm to our case in order to find the MLE of the parameters values under the alternative. In developing the EM algorithm, we calculated the MLE for each sample using different numbers of random starting points. Finally, 150 and 80 random starting points are used to find the global MLEs for samples obtained under the null hypothesis and alternative hypothesis, respectively. The mean of the MLE and the mean

square error is calculated in chapter 2. These results show that the estimates calculated using the customized EM are close to the true parameters.

In order to calculate the power of the LRT, we investigate the distribution of the statistics under the null hypothesis. For each sample size, 10000 random samples are generated. It was observed that the distribution did not fit the proposed asymptotic null distribution of chi square with 2 df. Although the square root of non-zero LRT statistics do not follow a standard normal distribution, the attempts to fit the data to a mixture of chi square with 0 degrees of freedom and chi square with non-integer degrees of freedom work well.

We use the empirical null distribution for our power study. One thousand simulated samples are generated for each parameter setting under the alternative hypothesis. Overall, based on McNemar tests, the performance of the LRT is significantly better than Spearman's correlation test and the single variable linear regression test (t test of a slope equal to 0.0). With the large sample sizes, the Single variable linear regression test is less powerful but not much. At the same time, we derive the expression of approximate power of the single variable linear regression. The approximate power is quite close with the observed power of the single variable linear regression. Thus, if the sample size is large enough, we recommend that one calculated the approximate power of the single variable linear regression test serving as an estimate of the power of observed single variable linear regression. It also can be considered as an estimate of the power of LRT which is slight underestimate. In the treatment group, the power of the LRT

increases as the sample size increase. The sample with greater shifted proportion and greater shifted mean has a greater power when the LRT is conducted.

However, there are several limitations of this research, some of which could be seen as interesting directions for future research. The first one is that we consider only a sampling design where we have an equal number of observations for values of the predictor fixed at X=0, 1, 2, 3. This would be the case in some dose response studies. There are several possible alternative designs which could result in a different null distribution as well as different power values for the same total sample size. Examples are designs where there are only 3 groups, unequal spacing between doses, or unequal group sizes. Or we could have an observational study where the predictor variable X, is a random variable. This would be the case perhaps in a study where disease susceptibility is a function of some quantitative variable, time since onset say, in a subset of the population and unrelated to this factor in the remainder of the population. In practice one would probably use either a permutation test or the bootstrap LRT in order to determine whether one accepts or rejects the null hypothesis. We are assuming that our use of critical values obtained through simulation corresponds to the results one would obtain with the bootstrap. We have no way of knowing whether the permutation test would result in similar power. A study comparing the power of these two approaches (bootstrap and permutation test) would be of interest. For the consideration of parameter settings, we can explore the test on larger sample with greater π and β. For example, $\pi = 0.8$ and $\beta = 1.0$. In the real world, the proportion of responders might vary from different dose levels. Thus, it will be an interesting and realistic design where the proportion of responders is expressed as a variable of dose levels.

The alternative hypothesis in this research can be considered as the special case of switching regression which is proposed by Quandt and Ramsey (1978) [31]. Thus, it would be an interesting future work to conduct the hypothesis test in terms of switching regression. The null hypothesis and the alternative hypothesis test for the special case of the switching regression can be expressed as follows.

$$\begin{aligned} H_0&: \alpha + \beta * X \\ H_1&: \alpha + \beta * X \qquad with\ probability\ of\ \pi \\ &\quad \alpha \qquad\qquad with\ probability\ of\ (1-\pi) \end{aligned} \tag{5.1}$$

References

[1] Good, PI, (1979). Detection of a Treatment Effect When Not All Experimental Subjects Will Respond to Treatment. *Biometrics*, 1979; 35: 483-489.

[2] Boos, DD. Brownie, C. Testing for a Treatment Effect in the Presence of Non-responders. *Biometrics*, 1986; 42:191-197.

[3] Johnson, RA. Verrill, S. Moore, DH. 2-Sample Rank-test for Detecting that Occur in a Small Proportion of the Treated Population. *Biometrics*, 1987; 43:641-655.

[4] Conover, WJ. Salsburg, DS. Locally Most Powerful Tests for Detecting Treatment Effects When Only a Subset of Patients Can Be Expected to Respond to Treatment. *Biometrics*, 1988; 44:189-196.

[5] Salsburg, D. S. Alternative Hypotheses for the Effects of Drugs in Small-Scale Clinical Studies. Biometrics ,1986; 42: 671-674.

[6] Lo, YT. Matthvsse, S. Rubin, DB, et al. Permutation Tests for Detecting and Estimating Mixtures in Task Performance Within Groups. *Statistics in Medicine*, 2002; 21:1937-1953.

[7] McMahon, RP. Arndt, S. Conley, RR. More Powerful Two-sample Tests for Differences in Repeated Measures of Adverse Effects in Psychiatric Trials When Only Some Patients May be at Risk. *Statistics In Medicine*, 2005; 24:11-21.

[8] Arndt, S. Turvey, C. Coryell, WH. Dawson, JH. Leon, AC. Akiskal, HS. A Comparison of Statistics Used to Summarize Participant Course in Longitudinal And Repeated Measures Studies. *Journal of Psychiatric Research*, 2000, 34:105–113.

[9] Boos, DD. Brownie, C. Mixture-models for Continuous Data in Dose-response Studies When Some Animals Are Unaffected by Treatment. *Biometrics,* 1991. 47:1489-1504.

[10] Razzaghi, M. Kodell, R. Dose-response modeling for Developmental Neurotoxicity Data. *Environmental and Ecological Statistics,* 2000. 7:191-203.

[11] Luo, XH. Boos, DD. Tamura, RN. Score Tests for Dose Effect in the Presence of Non-responders. *Statistics in Medicine, 2004*; 23:3581-3591.

[12] Levin, ED. Bowman, RE. Behavioral-effects of Chronic Exposure to Low Concentrations of Halothane During Development in Rats. *Anesthesia and Analgesia,* 1986; 65: 653-659.

[13] Rice DC. Lead-induced Behavioral Impairment on a Spatial Discrimination Reversal Task in Monkeys Exposed During Different Periods of Development. *Toxicology and Applied Pharmacology,* 1990; 106:327-333.

[14] Cohn, J. Cox, C. Soryslechta, DA. The Effects of Lead-exposure on learning in a Multiple Repeated Acquisition and Performance Schedule. *Neurotoxicology*, 1993; 14: 329-346.

[15] Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society, 1977, B,* 39: 1–38.

[16] Murray Aitkin, Granville Tunnicliffe Wilson. Mixture Models, Outliers, and the EM Algorithm. *Technometrics*, 1980, 22, 3: 325-331.

[17] Rai S. N., Matthews D. E. Improving the EM Algorithm, *Biometrics,* 1993, 49, 2: 587-591.

[18] R. C. Jansen. *Maximum Likelihood in a Generalized Linear Finite Model by Using the EM Algorithm. Biometrics,* 1993, 49, 1: 227-231.

[19] Nordheim E. V. On the Performance of a Likelihood Ratio Test for Genetic Linkage. *Biometrics,* 1984, 40, 3:785-790.

[20] Kent J. T. Robust Properties of Likelihood Ratio Test. *Biometrika,* 1982, 69, 1: 19-27.

[21] Matz M. V., Nielsen R. A Likelihood Ratio Test for Species Membership Based on DNA Sequence Data. *Philosophical Transactions: Biological Sciences,* 2005, 360, 1462: 1969-1974.

[22] Wilson E., Hilferty M., The distribution of Chi-Square. *Proceedings of the National Academy of Sciences of the United States of America,* 1931, *12, 17:684-688.*

[23] Baker, G., Note on the Distribution of the Stan dard Deviations and Second Moments of Samples From a Gram Charlie Population, *Annals of Mathematical Statistics,* 1934, 5, 113-123.

[24] Moore, G., Transformations to Normality Using Fractional Powers of a Variable, *Journal of the American Statistical Association,* 1957, 52, 237-246.

[25] Sankaran, M., On the Noncentral Chi-square Distribution, *Biometrika,* 1972, 59, 235-237.

[26] Jensen, D., Solomon, H., A Gaussian Approximation to the Distribution of a Definite Quadratic Form, *Journal of the American Statistical Association,* 1962, 67, 898-902.

[27] McLachlan, G. J. and Peel, D. *Finite Mixture Models,* 2000. New York: Wiley.

[28] Hall P. On the Bootstrap and Likelihood-Based Confidence Regions , *Biometrika,* 1987, 74, 481-493.

[29] Lo, Y. Mendell, R, N. Rubin, D, B. Testing the Number of Components in a Normal Mixture, *Biometrika,* 2001, 88, 3, 767-778.

[30] Box, *G. E. P.* and Muller, *M. E.* A Note on the Generation of Random Normal Deviates, *The Annals of Mathematical Statistics,* 1958, 29, 2, 610–611.

[31] Quandt, R. E. and Ramsey, J. B. Estimating mixtures of normal distributions and switching regression, *Journal of American Statistical Association,* 1978, 73, 730-738.