2004

Tests of homogeneity of several location and scale populations, and analysis of paired count data with zero-inflation and over-dispersion.

Xing Jiang
University of Windsor

Tests of Homogeneity of Several Location and Scale Populations, and Analysis of Paired Count Data with Zero-Inflation and Over-Dispersion

by

Xing Jiang

A Dissertation
Submitted to the Faculty of Graduate Studies and Research
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada
2004

Your file Votre référence
ISBN: 0-612-96372-1
Our file Notre référence
ISBN: 0-612-96372-1

Abstract

This thesis consists of two parts, refereed as Part I and Part II.

Part I: Testing homogeneity of several location-scale populations.

The widely used method for testing homogeneity of several normal populations is to test the equality of means based on the assumption that the variances among different groups are same. But in practice, we often get data which are different not only in means but also in variances.

Singh (1986) tests the homogeneity of several normal populations simultaneously regarding commonality of means and variances based on a method by Fisher (1950). However, this problem arises not only in normal populations but also in other populations. In this thesis, I extend Fisher's method to location-scale models in general. The location-scale models encompass all two parameter mean-variance models, such as the normal, negative binomial and beta-binomial models. Two test statistics are developed, one of which is based on the combination of two likelihood ratio statistics and the other is based on the combination of two score test statistics. Theoretical and empirical properties of these procedures are studied and applied to real life data analysis problems.

Part II: Analysis of paired count data with zero-inflation and over-dispersion.

Data in the form of paired counts (pre-treatment and post-treatment counts) arise in many fields such as biomedical, toxicology, epidemiology and so on. Poisson and binomial models are the most widely used models for these data. Frequently encountered problems in these data are the presence of extra-zeros and extra-dispersion and,

the possible correlation between the pre-treatment and post-treatment count.

In this thesis I developed methods of analysis for two different sets of paired count data, one of the data set is obtained from an experiment on premature ventricular contractions (PVC)(Berry, 1987) and the other set is a dental epidemiology data representing decayed, missing and filled teeth (DMFT) index (Böhning, Dietz, Schlattmann, Mendonca and Kirchner, 1999). I, then, study properties of these methods and anslyse the PVC data and the DMFT index data.

Acknowledgements

I wish to express my deep appreciation to my supervisor Professor S. R. Paul for his guidance and encouragement in preparation of this dissertation. I would like to thank Dr. R. J. Cook, Dr. M. Hlynka, Dr. A. Hussein and Dr. Y. Aneja for their useful comments and suggestions which greatly improved this dissertation.

I sincerely thank the Department of Mathematics and Statistics, and the University of Windsor, for giving me financial support in terms of graduate teaching assistantships, tuition scholarships and appointing me as sessional instructor. The financial support from Dr. S. R. Paul through his NSERC research grant throughout my 5-years graduate studies is gratefully acknowledged.

I would like to thank my professors Dr. S. R. Paul, Dr. K. Y. Fung, Dr. M. Hlynka and Dr. C. S. Wong for their readiness to help and stimulating discussions during graduate course work in Statistics.

I would like to thank my parents for their guidance and support throughout my life.

I would like to thank my wife and my son for their love, support and understanding.

Contents

List of Tables

Chapter 1

Introduction

When data are obtained from several different groups in an experiment, a very common statistical inference problem is to test if these data come from the same population. This problem can arise in many different areas. For example, a corn field is divided into several parts, each part is treated with a different fertiliser to see if these fetilisers have different effects; a teacher practices different teaching methods on different groups of students in her class to see if these methods yield different results; a doctor treats patients with different medicines to see if the treatment effect is same or not and so on. When we test this problem, the Fisher analysis of variance technique is widely used, by which we test the equality of means based on the assumption that the variances among different groups are same.

Hovewer, in practice, we often get data which are different not only in means but also in variances. Snedecor and Cochran (1967, pp 324) observed that an application of different treatments to otherwise homogeneous experimental units often results in groups that are different not only in means but also in variances. Thus, testing homogeneity of several populations in terms of means and variances is of considerable interest. The usual practice for testing homogeneity of several populations in terms of means and variances is first to test for the equality of variances and once this assumption is found to be tenable then to test the equality of means. Fisher (1950) suggested combining several independent tests. We quote (Fisher, 1950, pp 99)

"When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are on the whole lower than would often have been obtained by chance. It is sometimes desired, taking account only of these probabilities, and not of the detailed composition of the data from which they are derived, which may be of very different kinds, to obtain a single test of the significance of the aggregate, based on the product of the probabilities individually observed."

Assume that we wish to test a null hypothesis $H_0 : \theta \in \Theta_0$, where Θ_0 is a subset of a parameter space Θ. Suppose we have available p independent tests for testing H_0. We wish to combine these p tests into an overall test for H_0. Several methods of combining independent tests, including a method by Fisher (1950), are available. None of these procedures are uniformly most powerful. However, Littell and Folks (1971) have compared Fisher's method with three other well-known methods via exact Bahadur relative efficiency, and have found that Fisher's method is always at least as efficient as the other three methods and Littell and Folks (1973) have shown that Fisher's method is the most efficient.

Singh (1986) uses Fisher's method for testing simultaneously the equality of means and the equality of variances of several normal populations. Singh uses a test statistic which is the combination of two independent likelihood ratio statistics. However, this problem arises not only in normal populations but also in other populations such as an over-dispersed Poisson model, namely the negative binomial model and an over-dispersed binomial model, namely the beta-binomial model. Both models are widely used for count data with over-dispersion in many fields such as public health, toxicology, epidemiology, sociology, psychology, engineering, agriculture and so on. Also, this problem arises in many widely used lifetime models, such as, the Weibull or extreme-value models.

In this thesis, we extend Fisher's method to location-scale models in general. Two test statistics are developed, one of which is based on the combination of two likelihood ratio statistics and the other is based on the combination of two score test statistics. Under the general location-scale setup asymptotic independence is established for the two likelihood ratio statistics as well as for the two score test statistics. Then, by applying the general results, we obtain specific test statistics for testing homogeneity of several normal (μ, σ^2) populations, several negative binomial (m, c) populations, several beta-binomial (π, ϕ) populations and several Weibull (ψ, ϕ) populations . In the normal case exact independence of the two likelihood ratio statistics is shown by Singh (1986). In this thesis, we show exact independence of the two score test statistics. In all four cases simulations are conducted to compare the two procedures. We conclude that Fisher's method of combining two statistics, even when they are only asymptotically independent, does, in general, perform well for testing homogeneity of several populations in terms of the means and the variances.

However, the score test statistics have simple forms, are easy to calculate, and have uniformly good level properties. Therefore Fisher's method based on combining two score test statistics might be the method of choice.

Another problem considered in this thesis is the analysis of data in the form of paired counts. Data in the form of counts arise in many fields such as biomedical, toxicology, epidemiology and so on. Poisson and binomial models are most widely used models for count data. However, a Poisson model and a binomial model may not fit count data well. Frequently encountered problems in these data are the presence of more zeros than what can be expected and the presence of over-dispersion, which lead to a failure of the variance-mean relation of a Poisson model and a binomial model. In practice, the paired counts data are obtained before and after an experiment and the extra zeros may occur in different ways. For example, the data on premature ventricular contractions (PVC), given as paired counts by Berry (1987) for before and after drug administration, only have extra zeros after the drug administration, while the DMFT index data (Böhning et al., 1999), which have the form of (DMFT1, DMFT2) as paired count data for pre-treatment and after-treatment, have extra zeros, in most situations, as the common pair of (0, 0). In this thesis, we develop two different procedures to analyse data in the form of paired counts with zero-inflation and over-dispersion.

The data on premature ventricular contractions (PVC), originally given as counts by Berry (1987), are analysed by Farewell and Sprott (1988) as proportions. Conditional on the total count before and after drug administration, a binomial distribution is introduced. However, a binomial model may fail to fit a set of data in the form of proportions either because of the presence of zero-inflation (Farewell and Sprott, 1988) or because of the presence of over-dispersion (Deng and Paul, 2000). In this study, we use a zero-inflated beta-binomial model to develop procedures for testing for treatment effect. Based on this model, we can analyse the treatment effect through two parameters, namely, the zero-inflation parameter and the proportion parameter. Note that the zero-inflation parameter represents the proportion of cure and the proportion parameter represents the effect of the treatment on the uncured population. Therefore to determine treatment effect we can (i) estimate the zero-inflation parameter, the proportion of cure, and test whether the uncured population had any improvement of their prevailing condition as a result of the treatment or (ii) test the overall effect of the treatment. Results of a small simulation experiment, to study small sample behavior of a score test and a likelihood ratio test, are reported and the PVC data are analysed.

In biomedical and dental epidemiological experiments, data arise in the form of pre-treatment and post-treatment counts. The DMFT data, a dental epidemiology data set, are presented by (Böhning et al., 1999) for a prospective study of caries prevention of school-children from an urban area of Belo Horizonte (Brazil). To study treatment effects, Böhning et al. (1999) use a zero-inflated Poisson regression model (ZIPR) with the log function as the link and the pre-treatment count as the one of the covariates to delete the baseline effect. We introduce a zero-inflated bivariate Poisson regression model (ZIBPR) with a log-linear link for the ratio of the two mean parameters of the bivariate Poisson distribution and jointly model pre-treatment and post-treatment counts. We develop the EM-algorithm (Dempster et al., 1977) to obtain the maximum likelihood estimates of the parameters of the ZIBPR model. Further, we obtain exact the Fisher information matrix of the parameters of the ZIBPR model and develop a procedure for testing treatment effects of the method. A model selection procedure is used to decide on an appropriate model. For the DMFT index data, based on the model selected, we arrive at a ranking of the treatment effects which coincides with that from a simple analysis of treatment effects.

This thesis consists of two parts. Part I, including Chapter 3, Chapter 4 and Chapter 5, developes procedures for testing homogeneity of several location-scale populations in general. We compare our precedure with the procedure proposed by Singh (1986) for the normal case and apply the general method to several non-normal cases. Part II, including Chapter 6 and Chapter 7, analyses the treatment effects of paired count data with zero-inflation and over-dispersion. We develop two procedures, one of which is illustrated by the PVC data (Berry, 1987) and the other is illustrated by the DMFT data (Böhning et al., 1999).

In Chapter 2, we review some basic concepts and large sample hypothesis testing procedures such as the likelihood ratio test and the $C(\alpha)$ test. We also give review Fisher's method for combining several independent test statistics, the EM-algorithm and orthogonal transformations for parameters.

In Chapter 3, we extend Fisher's method to location-scale models in general. Two test statistics are developed, one of which is based on the combination of two likelihood ratio statistics and the other is based on the combination of two score test statistics. Under the general location-scale setup, asymptotic independence is established for the two likelihood ratio statistics as well as for the two score test statistics.

In Chapter 4, we use the general result of Chapter 3 to test the homogeneity of several normal populations based on combining the score test statistics and compare our method with the procedure proposed by Singh (1986). In the normal case, exact independence of the two likelihood ratio statistics is shown by Singh (1986). In Chapter 4, we show exact independence of the two score test statistics. Some simulations are conducted to compare the two procedures.

In Chapter 5, by applying the general results, we obtain two procedures for testing homogeneity of some non-normal populations. Here we consider two over-dispersed discrete models, namely the negative binomial model and the beta-binomial model. We also consider a widely used lifetime model, namely the Weibull or extreme-value model. In all three cases, simulations are conducted to compare the two procedures.

In Chapter 6, we develop score tests to test for treatment effect in the PVC data based on a zero-inflated beta-binomial model. Results of a small simulation experiment, to study small sample behavior of a score test and a likelihood ratio test, are reported and the PVC data are analysed.

In Chapter 7, a zero-inflated bivariate Poisson model is proposed to analyse the DMFT index data. We develop an EM-algorithm to obtain the maximum likelihood estimates and a procedure for testing treatment effects based on a zero-inflated bivariate Poisson regression model. We illustrate our procedures by the DMFT index data of Böhning et al. (1999).

In Chapter 8, a summary of this thesis and some discussions of future research are given.

Chapter 2

Some Preliminaries and Review

2.1 $\sqrt{n}$ consistent estimators

Let $\hat{\theta}_n, n = 1, 2, ...,$ be a sequence of estimators of θ. If the quantity $|\hat{\theta}_n - \theta| = O(n^{1/2})$ in probability as $n \to \infty$, then the estimator $\hat{\theta}_n$ is called a $\sqrt{n}$ consistent estimator of θ.

$\sqrt{n}$ consistent estimators were first suggested by Neyman (1959) for constructing the $C(\alpha)$ test. Also noted by Moran (1970).

If $\hat{\theta}_n$ is a sequence of maximum likelihood estimates of θ, then by the asymptotic properties of maximum likelihood estimators, it can be showed that maximum likelihood estimator is $\sqrt{n}$ consistent.

2.2 Likelihood ratio test

Suppose $X_1, X_2, ..., X_n$ be a random sample of size n from the population of X, which has a distribution function $f(X, \lambda)$, where $\lambda = (\theta, \phi)'$ with

$\theta = (\theta_1, \theta_2, ..., \theta_p)'$ and $\phi = (\phi_1, \phi_2, ..., \phi_s)'$.

Then the likelihood can be given as $L(X_1, X_2, ..., X_n, \lambda)$. It is of interest to test the null hypothesis $H_0 : \theta = \theta_0 = (\theta_{10}, \theta_{20}, ..., \theta_{p0})'$ against $H_1 : \theta \neq \theta_0$ treating $\phi = (\phi_1, \phi_2, ..., \phi_s)'$ as the nuisance parameter.

The likelihood ratio for testing H_0 is defined as

$$\Lambda = \frac{L(X_1, X_2, ..., X_n, \theta_0, \hat{\phi})}{L(X_1, X_2, ..., X_n, \tilde{\theta}_0, \tilde{\phi})}.$$

Let log refer to the base e logarithm. Then, the log-likelihood ratio statistic is given by

$$LR = -2 \log \Lambda = 2(l_1 - l_0),$$

where $l_0 = \log L(X_1, X_2, ..., X_n, \theta_0, \hat{\phi})$ is the maximized log-likelihood under H_0 with $\hat{\phi}$ as the maximum likelihood estimate of ϕ under H_0 and $l_1 = \log L(X_1, X_2, ..., X_n, \tilde{\theta}_0, \tilde{\phi})$ is the maximized log-likelihood under the alternative hypothesis with $\tilde{\theta}_0$ and $\tilde{\phi}$ as the maximum likelihood estimates of θ and ϕ under H_1 respectively. Under the null hypothesis H_0, for a large n, the statistic LR is distributed approximately as a chi-square with p degrees of freedom.

2.3 Score test

Let $l = l(\theta, \phi; y)$ be the log-likelihood for data $y = (y_1, ..., y_n)$ with parameters $\theta = (\theta_1, ..., \theta_p)'$ and $\phi = (\phi_1, ..., \phi_s)'$, where θ is the parameter of interest and ϕ is the nuisance parameter. Suppose we wish to test

$H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$, Further, let

$$\psi = \left.\frac{\partial l}{\partial \theta}\right|_{\theta=\theta_0} = \left.\left[\frac{\partial l}{\partial \theta_1}, ..., \frac{\partial l}{\partial \theta_p}\right]'\right|_{\theta=\theta_0},$$

$$\gamma = \left.\frac{\partial l}{\partial \phi}\right|_{\theta=\theta_0} = \left.\left[\frac{\partial l}{\partial \phi_1}, ..., \frac{\partial l}{\partial \phi_s}\right]'\right|_{\theta=\theta_0},$$

$$I_{\theta\theta} = E\left(-\left.\frac{\partial^2 l}{\partial\theta\partial\theta'}\right|_{\theta=\theta_0}\right),$$

$$I_{\theta\phi} = E\left(-\left.\frac{\partial^2 l}{\partial\theta\partial\phi'}\right|_{\theta=\theta_0}\right),$$

and

$$I_{\phi\phi} = E\left(-\left.\frac{\partial^2 l}{\partial\phi\partial\phi'}\right|_{\theta=\theta_0}\right).$$

Now, define $S = \frac{\partial l}{\partial \theta} - B\frac{\partial l}{\partial \phi}$, where $B = I_{\theta\phi}I_{\phi\phi}^{-1}$ is the partial regression coefficient matrix obtained by regressing $\frac{\partial l}{\partial \theta}$ on $\frac{\partial l}{\partial \phi}$. The dispersion matrix of S is

$I_{\theta\theta\cdot\phi} = I_{\theta\theta} - I_{\theta\phi}I_{\phi\phi}^{-1}I_{\phi\theta}$.

Then, it can be shown (Neyman, 1959) that asymptotically, as $n \to \infty$

$S'I_{\theta\theta\cdot\phi}^{-1}S \sim \chi^2_{(p)}$.

If $\phi = (\phi_1, ..., \phi_s)'$ in S and $I_{\theta\theta\cdot\phi}$ is replaced by some $\sqrt{n}$-consistent estimator $\tilde{\phi} = (\tilde{\phi}_1, ..., \tilde{\phi}_s)'$, then, asymptotically, as $n \to \infty$,

$$\tilde{S}'\tilde{I}_{\theta\theta\cdot\phi}^{-1}\tilde{S} \sim \chi^2_{(p)},$$

where $\tilde{S}$ and $\tilde{I}_{\theta\theta\cdot\phi}$ are obtained by replacing ϕ by $\tilde{\phi}$ in S and $I_{\theta\theta\cdot\phi}$. This is Neyman's $C(\alpha)$ test. Further, let $\hat{\phi}$ be the maximum likelihood estimate of ϕ under H_0, and $\hat{\psi}$ and $\hat{I}_{\theta\theta\cdot\phi}$ be the estimate values obtained by replacing ϕ by $\hat{\phi}$ in ψ and

$I_{\theta\theta\cdot\phi}$ respectively, then $\tilde{S}$ reduced to $\hat{\psi}$. The $C(\alpha)$ statistic then reduces to $\hat{\psi}'\hat{I}^{-1}_{\theta\theta\cdot\phi}\hat{\psi}$. Asymptotically, as $n \to \infty$, $\hat{\psi}'\hat{I}^{-1}_{\theta\theta\cdot\phi}\hat{\psi} \sim \chi^2_{(p)}$. This is Rao's score test (Rao, 1947).

The score test is a special case of the more general $C(\alpha)$ test in which the nuisance parameters are replaced by maximum likelihood estimates. The score test is particularly appealing as it requires estimates of the parameters only under the null hypothesis, and often produces a statistic which is simple to calculate. For more discussion on the choice of $C(\alpha)$ or score tests see Breslow (1990) and Paul and Banerjee (1998).

2.4 Fisher's method of combining independent tests

Assume that we wish to test a null hypothesis $H_0 : \theta \in \Theta_0$, where Θ_0 is a subset of a parameter space Θ. Suppose we have available p independent tests for testing H_0. We wish to combine these p tests into an overall test for H_0.

Let $T^{(1)}, \cdots, T^{(p)}$ be p independent sequences of test statistics for testing H_0. We wish to combine $T^{(1)}, \cdots, T^{(p)}$ into an overall statistic T_n. Then, Fisher's method of combining the independent tests $T^{(1)}, \cdots, T^{(p)}$ is given by $T_n^{(F)} = -2\log\prod_i L^{(i)}$, where $L^{(i)} = 1 - F^{(i)}\left(T^{(i)}\right)$ and $F^{(i)}(t) = P_0\{T^{(i)} < t\}$ is the null cumulative distribution function of $T^{(i)}$. Then $L^{(i)}$, $i = 1, 2, ..., p$, are independently and uniformly distributed over (0, 1) for $\theta \in \Theta_0$. Further, for $\theta \in \Theta_0$ the quantity $-2\log L^{(i)}$ has a chi-square distribution with 2 degrees of freedom, and hence, the quantity $T_n^{(F)} = -2\log\prod_i L^{(i)} = -2\sum_i \log L^{(i)}$ has a chi-square distribution with $2p$ degree of freedom under H_0. A large value of $T_n^{(F)}$ indicates evidence against the null hypothesis.

Several methods of combining independent tests, including a method by Fisher (1950), are available. None of these procedures are uniformly most powerful. However, Littell and Folks (1971) have compared Fisher's method with three other well-known methods via exact Bahadur relative efficiency, and have found that Fisher's method is always at least as efficient as the other three methods and Littell and Folks (1973) have shown that Fisher's method is the most efficient.

2.5 Orthogonal parameter and orthogonal transformation

Let $l = l(\theta, \phi; y)$ be the log-likelihood for data $y = (y_1, ..., y_n)$ with parameters $\theta = (\theta_1, ..., \theta_p)' \in \Theta$ and $\phi = (\phi_1, ..., \phi_s)' \in \Phi$, where θ is the parameter of interest and ϕ is the nuisance parameter, and Θ and Φ are the corresponding parameter spaces.

Orthogonality is defined with respect to the expected Fisher information matrix. We define θ to be orthogonal to ϕ if the elements of the information matrix satisfy

$$i_{\theta_{k_1}\phi_{k_2}} = E\left(\frac{\partial l}{\partial \theta_{k_1}} \frac{\partial l}{\partial \phi_{k_2}}\right) = E\left(-\frac{\partial l^2}{\partial \theta_{k_1} \partial \phi_{k_2}}\right) = 0,$$

for $k_1 = 1, ..., p$ and $k_2 = 1, ..., s$. If this holds for all $\theta \in \Theta$ and all $\phi \in \Phi$, then θ and ϕ are globally orthogonal.

Under global orthogonality, the scores $\frac{\partial l}{\partial \theta}$ and $\frac{\partial l}{\partial \phi}$ are uncorrelated and $E\left(-\frac{\partial l^2}{\partial \theta \partial \phi'}\right) = 0$.

As noted in Cox and Reid (1987), it is not in general possible to find an orthogonal parameter. Cox and Reid (1987) give a special case in which a scalar parameter ψ is orthogonal to the other parameters $\lambda_1, \lambda_2, ..., \lambda_s$.

Suppose $(\psi, \phi_1, \phi_2, ..., \phi_s)$ is the original parameter for which log-likelihood function is $l(\psi, \phi) = l(\psi, \phi_1, \phi_2, ..., \phi_s)$. Further, suppose that we have the transformation

$$\phi_1 = \phi_1(\psi, \lambda),\ \phi_2 = \phi_2(\psi, \lambda), ..., \phi_s = \phi_s(\psi, \lambda),$$

where $\lambda = (\lambda_1, \lambda_2, ..., \lambda_s)'$.

The dependence of ϕ on ψ and λ can be determined by the following partial differential equations

$$\sum_{k_1=1}^{s} i_{\phi_{k_1}\phi_{k_2}} \frac{\partial \phi_{k_1}}{\partial \psi} = -i_{\psi\phi_{k_2}}, k_2 = 1, ..., s, \tag{2.1}$$

where $i_{\phi_{k_1}\phi_{k_2}} = E\left(\frac{\partial l}{\partial \phi_{k_1}} \frac{\partial l}{\partial \phi_{k_2}}\right)$ and $i_{\psi,\phi_{k_2}} = E\left(\frac{\partial l}{\partial \psi} \frac{\partial l}{\partial \phi_{k_2}}\right)$.

According to the theorem of Frobenius in differential geometry (Boothby, 1975, page 159), the equations in (2.1) have a general solution. A special case of this transformation occurs if we have only one scale parameter ψ and one location parameter ϕ. So we have the transformation $(\psi, \phi) = (\psi, \phi(\psi, \lambda))$, such that ψ and λ are orthogonal to each other. We can get this transformation through the partial differential equation

$$i_{\phi\phi} \frac{\partial \phi}{\partial \psi} = -i_{\psi\phi}. \tag{2.2}$$

2.6 Bivariate Poisson distribution

Suppose that variables $Z_i, i = 0, 1, 2$ are independent Poisson random variables with parameters λ_0, λ_1 and λ_2 respectively. Let

$X = Z_0 + Z_1$ and $Y = Z_0 + Z_2$,

then (X, Y) is distributed as bivariate Poisson distribution with the probability function

$$Pr(X = x, Y = y) = \exp(-\lambda_1 - \lambda_2 - \lambda_0) \sum_{i=0}^{\min\{x,y\}} \frac{\lambda_1^{x-i}\lambda_2^{y-i}\lambda_0^i}{(x-i)!(y-i)!i!},$$

and we have $E(X) = \lambda_1 + \lambda_0$, $E(Y) = \lambda_2 + \lambda_0$ and $Cov(X, Y) = \lambda_0$.

More details of the bivariate Poisson distribution can be found in Holgate, (1964); Irwin, (1963); Paul and Ho, (1989); Kocherlakota and Kocherlakota, (1992), and Karilis and Ntzoufras, (1998).

2.7 EM algorithm

The EM-algorithm is a general iterative method to obtain maximum likelihood estimates in incomplete data situations. It was first proposed by Hartley (1958) and was generalized by Dempster, Laird and Rubin (1977). Let $y \in \mathbb{R}^n$ denote a vector of observed data and $z \in \mathbb{R}^m$ a vector of unobservable data. Then the complete data are given by (y, z). Furthermore, let $f(y, z; \theta)$ denote the joint density of the complete data depending on an unknown parameter vector θ.

Then the maximum likelihood estimate of θ can be obtained iteratively by the EM-algorithm using an E-step and an M-step. If $\theta^{(0)}$ denotes a starting value for θ, the $(p + 1)$th cycle of the EM-algorithm consists of the following two steps for $p = 0, 1, ...$

E(xpectation)-step:

Compute the expectation $M(\theta|\theta^{(p)})$,

where $M(\theta|\theta^{(p)}) = E[\log f(y, z; \theta)|y; \theta^{(p)}] = \int \log[f(y, z; \theta)]k(z|y; \theta^{(p)})dz$.

Here $k(z|y; \theta^{(p)})$ is the conditional density of the unobservable data z, given the observed data y and the value $\theta^{(p)}$, which is the estimate of θ in the pth cycle.

M(aximizing)-step:

Determine $\theta^{(p+1)}$ by maximizing $M(\theta|\theta^{(p)})$ with respect to θ.

The iterations are stopped according to a termination criterion, e.g., if $|\theta^{(p)} - \theta^{(p+1)}|/|\theta^{(p)}| < \varepsilon$ is satisfied.

Part I

Testing Homogeneity of Several Location-Scale Populations

Chapter 3

Tests of Homogeneity of Several Location-scale Populations : The General Results

3.1 Introduction

The widely used Fisher analysis of variance technique tests the equality of means based on the assumption that the variances among different groups are homogeneous. But in practice, we often get data which are different not only in means but also in variances. Snedecor and Cochran (1967, p 324) observe that an application of different treatments to otherwise homogeneous experimental units often results in groups that are different not only in means but also in variances. Thus, testing homogeneity of several populations in terms of means and variances is of considerable interest. The usual practice for testing homogeneity of several populations in terms of means and variances is first to test for the equality of variances and once this assumption is found to be tenable, the equality of means is tested.

For testing simultaneously the equality of means and the equality of variances of several normal (μ, σ^2) populations, Singh (1986) uses a test statistic based on the combination of two independent likelihood ratio statistics. Singh's procedure is based on a method by Fisher (1950) for combining two or more independent test statistics to test a general hypothesis.

The purpose of this chapter is to extend Fisher's method to location-scale models in general. Two test statistics are developed, one of which is based on the combination of two likelihood ratio statistics and the other is based on the combination of two score test statistics. Under the general location-scale setup, asymptotic independence is established for the two log-likelihood ratio statistics as well as for the two score test statistics.

In Section 3.2, we extend Singh's procedure for the likelihood ratio procedure to the general location-scale model and show asymptotic independence. In Section 3.3, we derive the score test statistics. Fisher's procedure for combining two score statistics and the asymptotic independence of the two score statistics are given in Section 3.4. A conclusion is given in Section 3.5.

3.2 Fisher's procedure for combining two log-likelihood ratio test statistics and their asymptotic independence

Consider a location-scale family of distributions $f(x, \psi, \phi)$, where ψ is the location parameter and ϕ is the scale parameter. Suppose we obtain data $x_{i1}, x_{i2}, ..., x_{in_i}$ from the ith, $i = 1, ..., k$, population with parameters ψ_i and ϕ_i. Then, the log-likelihood can be written as

$$l = \sum_{i=1}^{k} l_i, \quad \text{where } l_i = \sum_{j=1}^{n_i} \log f(x_{ij}, \psi_i, \phi_i), i = 1, ..., k.$$

Now, let $\Psi = (\psi_1, ..., \psi_k)'$, $\Phi = (\phi_1, \ldots, \phi_k)'$. Define the parameter spaces

$\Theta = \{(\Psi, \Phi) \mid \psi_i \text{ and } \phi_i, i = 1, \ldots, k \text{ are unspecified } \}$,

$\Theta_0 = \{(\Psi, \Phi) \mid \psi_i = \psi, \phi_i = \phi, i = 1, \ldots, k, \text{ where } \psi \text{ and } \phi \text{ are unspecified } \}$,

$\Theta_1 = \{(\Psi, \Phi) \mid \phi_i = \phi, i = 1, \ldots k, \text{ where } \Psi \text{ and } \phi \text{ are unspecified } \}$.

Suppose we wish to test

$H_0 : \psi_i = \psi, \phi_i = \phi, i = 1, \ldots, k$, where ψ and ϕ are unspecified against H_1 : at least two ψ's or two ϕ's are not same.

Then the test by Fisher's method is the combination of two independent tests corresponding to the following hypotheses:

H_0': $\psi_i = \psi, \phi_i = \phi, i = 1, \ldots, k$, where ψ and ϕ are unspecified against H_1': at least two ψ's are not same and $\phi_i = \phi, i = 1, \ldots, k$, where ϕ are unspecified.

and

H_0'': $\phi_i = \phi, i = 1, \ldots, k$, where λ are unspecified against H_1'': at least two ϕ's are not same.

Let LR be the log-likelihood ratio statistic for testing H_0 against H_1. Similarly, let LR_1 and LR_2 be the log-likelihood ratio statistics for testing H_0' against H_1' and H_0'' against H_1'' respectively. Further, let $\hat{l}_0, \hat{l}_1, \hat{l}_2$ denote the estimated values of log-likelihood function under Θ_0, Θ_1 and Θ respectively. Then, the log-likelihood ratio statistics for testing H_0 against H_1, H_0' against H_1' and H_0'' against H_1'' are

$LR = 2(\hat{l}_2 - \hat{l}_0)$,

$LR_1 = 2(\hat{l}_1 - \hat{l}_0)$

and

$LR_2 = 2(\hat{l}_2 - \hat{l}_1)$

respectively.

Now, $LR = 2(\hat{l}_2 - \hat{l}_0) = LR_2 + LR_1$. Asymptotically, as $n_i \to \infty, i = 1, \cdots, k$, $LR \sim \chi^2_{2(k-1)}$ under Θ_0, $LR_1 \sim \chi^2_{k-1}$ under Θ_0 and $LR_2 \sim \chi^2_{k-1}$ under Θ_1.

Since the parameter space $\Theta_1 \supset \Theta_0$, we can conclude that all of the above asymptotic results also hold under Θ_0 and by using the Cochran's Theorem (Cochran, 1934), the two statistics LR_1 and LR_2 are asymptotically independent.

Let $L_1(t_1) = Pr(LR_1 \geq t_1 \mid H_0')$ and $L_2(t_2) = Pr(LR_2 \geq t_2 \mid H_0'')$. Further, let M_1 be the test statistic of Fisher's procedure for combining two log-likelihood ratio test statistics.

Since LR_1 and LR_2 are asymptotically independently distributed, then, following Fisher's method,

$$M_1 = -2\log[L_1(LR_1)L_2(LR_2)]$$

is approximately distributed as χ^2_4. Thus, we reject H_0 in favor of H_1, if $M_1 \geq \chi^2_4(\alpha)$, where $\chi^2_4(\alpha)$ is the $100\alpha\%$ point of the χ^2 distribution with 4 degrees of freedom.

3.3 The derivations of the score test statistics S_1 and S_2

As in Section 3.2, to develop Fisher's procedure for combining two score test statistics, we need to obtain the test statistic S_1 for testing hypothesis H_0' against H_1' and the test statistic S_2 for testing hypothesis H_0'' against H_1''. However, the score test statistics S_1 and S_2 may not be independent or asymptotically independent in general. To obtain score test statistics S_1 and S_2, which are asymptotcally independent in general, we need to transform the parameters $(\psi_i,\ \phi_i)$, $i = 1, ..., k$, into a set of orthogonal parameters $(\psi_i,\ \lambda_i)$, $i = 1, ..., k$ (Cox and Reid, 1987, p3). Let $(\psi_i,\ \phi_i = \phi_i(\psi_i, \lambda_i))$, $i = 1, ..., k$, be such transformations which satisfy

$$i_{\psi_i\phi_i} + i_{\phi_i\phi_i}\frac{\partial \phi_i}{\partial \psi_i} = 0, i = 1, \ldots, k, \tag{3.1}$$

where $i_{\psi_i\phi_i} = E\left(-\frac{\partial^2 l_i}{\partial \psi_i \partial \phi_i}\right)$ and $i_{\phi_i\phi_i} = E\left(-\frac{\partial^2 l_i}{\partial \phi_i^2}\right)$.

Then the hypotheses given above can be expressed in terms of the orthogonal parameters as

H_0 : $\psi_i = \psi, \lambda_i = \lambda, i = 1, \ldots, k$, where ψ and λ are unspecified against H_1 : at least two ψ's or two λ's are not same.

Then the test by Fisher's method is the combination of two independent tests corresponding to the following hypotheses:

H_0': $\psi_i = \psi, \lambda_i = \lambda, i = 1, \ldots, k$, where ψ and λ are unspecified against H_1': at least two ψ's are not same and $\lambda_i = \lambda, i = 1, \ldots, k$, where λ are unspecified.

and

H_0'': $\lambda_i = \lambda, i = 1, \ldots, k$, where λ are unspecified against H_1'': at least two λ's are not same.

3.3.1 The derivation of score test statistic S_1 for testing H_0' vs H_1'

We now derive the score test statistic S_1 for testing H_0' against H_1'. For convenience we write the log-likelihood in terms of the orthogonal parameters as

$$l^* = \sum_{i=1}^{k} l_i^*, \text{ where }, l_i^* = \sum_{j=1}^{n_i} \log f(x_{ij}, \psi_i, \phi_i(\psi_i, \lambda_i)), i = 1, ..., k.$$

Reparameterize $\psi_i, i = 1, ..., k$, under H_1', by $\psi_i = \psi + \alpha_i$ with $\alpha_k = 0$.

Let

$\alpha' = (\alpha_1, ..., \alpha_{k-1})$ and $\omega_1' = (\psi, \lambda)$.

Then testing H_0' is equivalent to testing $\alpha = 0$ with ω_1 being treated as a nuisance parameter.

Let $s_1 = \left.\frac{\partial l^*}{\partial \alpha}\right|_{\alpha=0}$,

and

$$A_1 = E\left(-\left.\frac{\partial^2 l^*}{\partial\alpha\partial\alpha'}\right|_{\alpha=0}\right), \; C_1 = E\left(-\left.\frac{\partial^2 l^*}{\partial\alpha\partial\omega_1'}\right|_{\alpha=0}\right), \; D_1 = E\left(-\left.\frac{\partial^2 l^*}{\partial\omega\partial\omega_1'}\right|_{\alpha=0}\right).$$

If we use the maximum likelihood estimate $\hat{\omega}$ of the nuisance parameter ω under the null hypothesis H_0' in s_1, A_1, C_1 and D_1, then the score test for testing H_0' against H_1' is

$$S_1 = \hat{s}_1'(\hat{A}_1 - \hat{C}_1\hat{D}_1^{-1}\hat{C}_1')^{-1}\hat{s}_1.$$

Note that A_1 can be simplified as

$$A_1 = diag\left[E\left(-\left.\frac{\partial^2 l^*}{\partial\alpha_1^2}\right|_{\alpha=0}\right), E\left(-\left.\frac{\partial^2 l^*}{\partial\alpha_2^2}\right|_{\alpha=0}\right), \cdots, E\left(-\left.\frac{\partial^2 l^*}{\partial\alpha_{k-1}^2}\right|_{\alpha=0}\right)\right].$$

Further, note that the transformatin of $(\psi_1, \psi_2, ..., \psi_k)$ to (α, ψ) is only a linear transformation. So the parameter (α, ψ) is still orthogonal with λ. Now, based on the orthogonality of the parameter (α, ψ) to λ, the off-diagonal elements of the matrix D_1 and all elements of the second column of the $(k-1) \times 2$ matrix C_1 are zero. We can simplify D_1 and C_1 as

$$C_1 = \begin{pmatrix} E\left(-\frac{\partial^2 l^*}{\partial \alpha_1 \partial \psi}\Big|_{\alpha=0}\right) & 0 \\ E\left(-\frac{\partial^2 l^*}{\partial \alpha_2 \partial \psi}\Big|_{\alpha=0}\right) & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ E\left(-\frac{\partial^2 l^*}{\partial \alpha_{k-1} \partial \psi}\Big|_{\alpha=0}\right) & 0 \end{pmatrix},$$

and

$$D_1 = \begin{pmatrix} E\left(-\frac{\partial^2 l^*}{\partial \psi^2}\Big|_{\alpha=0}\right) & 0 \\ 0 & E\left(-\frac{\partial^2 l^*}{\partial \lambda^2}\Big|_{\alpha=0}\right) \end{pmatrix}.$$

Note that

$$E\left(-\frac{\partial^2 l^*}{\partial \alpha_i \partial \psi}\Big|_{\alpha=0}\right) = E\left(-\frac{\partial^2 l_i^*}{\partial \psi^2}\Big|_{\alpha=0}\right), i = 1, ..., k-1,$$

and

$$E\left(-\frac{\partial^2 l^*}{\partial \psi^2}\Big|_{\alpha=0}\right) = \sum_{i=1}^{k} E\left(-\frac{\partial^2 l_i^*}{\partial \psi^2}\Big|_{\alpha=0}\right).$$

Then the inverse of $A_1 - C_1 D_1^{-1} C_1'$ is

$$A_1^{-1} + \frac{\mathbf{1}\mathbf{1}'}{E\left(-\frac{\partial^2 l_k^*}{\partial \psi^2}\Big|_{\alpha=0}\right)},$$

where $\mathbf{1}_{(k-1)\times 1} = (1, 1, \cdots, 1)'$.

Further, we have $\left.\frac{\partial l^*}{\partial \alpha_i}\right|_{\alpha=0} = \left.\frac{\partial l_i^*}{\partial \psi}\right|_{\alpha=0}$, $i = 1, ..., k-1$, and $\left.\frac{\partial l^*}{\partial \psi}\right|_{\alpha=0} = \sum_{i=1}^{k} \left.\frac{\partial l_i^*}{\partial \psi}\right|_{\alpha=0}$. Note that when we use the maximum likelihood estimate $\hat{\omega}_1$ of the nuisance parameter ω_1 in $\left.\frac{\partial l^*}{\partial \psi}\right|_{\alpha=0}$ under H_0', the estimated value of $\left.\frac{\partial l^*}{\partial \psi}\right|_{\alpha=0}$ is 0. From above results, we obtain

$$S_1 = \sum_{i=1}^{k} \frac{\hat{s}_{1i}^2}{\hat{v}_{1i}},$$

where $\hat{s}_{1i}$ and $\hat{v}_{1i}$, $i = 1, ..., k$, are the estimated values of $\left.\frac{\partial l_i^*}{\partial \psi}\right|_{H_0'}$ and $E\left(-\left.\frac{\partial^2 l_i^*}{\partial \psi^2}\right|_{H_0'}\right)$, $i = 1, ..., k$, respectively.

3.3.2 The derivation of score test statistic S_2 for testing H_0'' vs H_1''

For the derivation of the score test statistic S_2 for testing H_0'' against H_1'' we reparameterize $\lambda_i, i = 1, ..., k$, by $\lambda_i = \lambda + \beta_i$ with $\beta_k = 0$. Let $\beta' = (\beta_1, ..., \beta_{k-1})$ and $\omega_2' = (\psi_1, \psi_2, ..., \psi_k, \lambda)$. Then testing H_0'' is equivalent to testing $\beta = 0$ with ω_2 being treated as a nuisance parameter.

Further, let $s_2 = \left.\frac{\partial l^*}{\partial \beta}\right|_{\beta=0}$,

and

$$A_2 = E\left(-\left.\frac{\partial^2 l^*}{\partial \beta \partial \beta'}\right|_{\beta=0}\right), \; C_2 = E\left(-\left.\frac{\partial^2 l^*}{\partial \beta \partial \omega_2'}\right|_{\beta=0}\right), \; D_2 = E\left(-\left.\frac{\partial^2 l^*}{\partial \omega_2 \partial \omega_2'}\right|_{\beta=0}\right).$$

If we use the maximum likelihood estimate $\hat{\omega}_2$ of the nuisance parameter ω_2 under the null hypothesis H_0'' in s_2, A_2, C_2 and D_2, then the score test for testing H_0'' against H_1'' is

$$S_2 = \hat{s}_2'(\hat{A}_2 - \hat{C}_2\hat{D}_2^{-1}\hat{C}_2')^{-1}\hat{s}_2.$$

Note that A_2 can be simplified as

$$A_2 = diag\left[E\left(-\frac{\partial^2 l^*}{\partial \beta_1^2}\bigg|_{\beta=0}\right), E\left(-\frac{\partial^2 l^*}{\partial \beta_2^2}\bigg|_{\beta=0}\right), \cdots, E\left(-\frac{\partial^2 l^*}{\partial \beta_{k-1}^2}\bigg|_{\beta=0}\right)\right].$$

Further, as in Section 3.3.1, the transformation of $(\lambda_1, \lambda_2, ..., \lambda_k)$ to (β, λ) is only a linear transformation. So the parameter (β, λ) is still orthogonal with $(\psi_1, \psi_2, ..., \psi_k)$. Now, we can obtain the off-diagonal elements of the matrix D_2 and all elements of the first k columns of the $(k-1) \times (k+1)$ matrix C_2 are zero. D_2 and C_2 can be simplified as

$$C_2 = \begin{pmatrix} \overbrace{0 \cdots 0}^{k} & E\left(-\frac{\partial^2 l^*}{\partial \beta_1 \partial \lambda}\bigg|_{\beta=0}\right) \\ \overbrace{0 \cdots 0}^{k} & E\left(-\frac{\partial^2 l^*}{\partial \beta_2 \partial \lambda}\bigg|_{\beta=0}\right) \\ \cdot & \cdot \\ \cdot & \cdot \\ \overbrace{0 \cdots 0}^{k} & E\left(-\frac{\partial^2 l^*}{\partial \beta_{k-1} \partial \lambda}\bigg|_{\beta=0}\right) \end{pmatrix},$$

and

$$D_2 = \begin{pmatrix} E\left(-\frac{\partial^2 l^*}{\partial \Psi \partial \Psi'}\bigg|_{\beta=0}\right) & 0 \\ 0 & E\left(-\frac{\partial^2 l^*}{\partial \lambda^2}\bigg|_{\beta=0}\right) \end{pmatrix}.$$

Note that

$$E\left(-\frac{\partial^2 l^*}{\partial \beta_i \partial \lambda}\bigg|_{\beta=0}\right) = E\left(-\frac{\partial^2 l_i^*}{\partial \lambda^2}\bigg|_{\beta=0}\right), i = 1, ..., k-1,$$

and

$$E\left(-\frac{\partial^2 l_i^*}{\partial\lambda^2}\bigg|_{\beta=0}\right)=\sum_{i=1}^{k}E\left(-\frac{\partial^2 l_i^*}{\partial\lambda^2}\bigg|_{\beta=0}\right).$$

Then the inverse of $A_2 - C_2 D_2^{-1} C_2'$ is

$$A_2^{-1}+\frac{\mathbf{1}\mathbf{1}'}{E\left(-\frac{\partial^2 l_k^*}{\partial\lambda^2}\Big|_{\beta=0}\right)},$$

where $\mathbf{1}_{(k-1)\times 1}=(1,1,\cdots,1)'$.

Further, we have $\frac{\partial l^*}{\partial\beta_i}\Big|_{\beta=0}=\frac{\partial l_i^*}{\partial\lambda}\Big|_{\beta=0}$, $i=1,...,k-1$, and $\frac{\partial l^*}{\partial\lambda}\Big|_{\beta=0}=\sum_{i=1}^{k}\frac{\partial l_i^*}{\partial\lambda}\Big|_{\beta=0}$. Note that when we use the maximum likelihood estimate $\hat{\omega}_2$ of the nuisance parameter ω_2 in $\frac{\partial l^*}{\partial\lambda}\Big|_{\beta=0}$ under H_0'', the estimated value of $\frac{\partial l^*}{\partial\lambda}\Big|_{\beta=0}$ is 0. From above results, we obtain

$$S_2=\sum_{i=1}^{k}\frac{\hat{s}_{2i}^2}{\hat{v}_{2i}},$$

where $\hat{s}_{2i}$ and $\hat{v}_{2i}$, $i=1,...,k$, are the estimated values of $\frac{\partial l_i^*}{\partial\lambda}\Big|_{H_0''}$ and $E\left(-\frac{\partial^2 l_i^*}{\partial\lambda^2}\Big|_{H_0''}\right)$, $i=1,...,k$, respectively.

3.4 Asymptotic independence of the two score test statistics S_1 and S_2, and Fisher's procedure for combining two score test statistics

Theorem 1 Under H_0, asymptotically, as $n_i\to\infty, i=1,...,k$, the two statistics S_1 and S_2 are independent.

Let S be the score test statistic for testing H_0 against H_1. As in the proof of the indepedence of LR_1 and LR_2 in Section 3.2, if we can prove that under H_0, asymptotically $S = S_1 + S_2$ and the distributions of S, S_1 and S_2 are $\chi^2_{2(k-1)}, \chi^2_{k-1}$ and χ^2_{k-1}, then by using the Cochran's Theorem (Cochran, 1934), the asymptotically independence of S_1 and S_2 can be established.

Proof:

For this we first need to derive the score test for testing H_0 against H_1. Now, reparameterize ψ_i and λ_i, $i = 1, ..., k$, under H_1, by $\psi_i = \psi + \alpha_i, i = 1, ..., k$, with $\alpha_k = 0$ and $\lambda_i = \lambda + \beta_i, i = 1, ..., k$, with $\beta_k = 0$. Then testing H_0 is equivalent to testing $\alpha = 0$ and $\beta = 0$ with $\omega = (\psi, \lambda)'$ as nuisance parameters.

The score test statistic for testing H_0 against H_1 is obtained in Appendix A.1. Now we have

$$S = S_{01} + S_{02},$$

where $S_{01} = \sum_{i=1}^{k} \frac{\hat{s}^2_{01i}}{\hat{v}_{01i}}$ and $S_{02} = \sum_{i=1}^{k} \frac{\hat{s}^2_{02i}}{\hat{v}_{02i}}$ with the estimated values of $s_{01i} = \left.\frac{\partial l_i^*}{\partial \psi}\right|_{H_0}$, $v_{01i} = E\left(-\left.\frac{\partial^2 l_i^*}{\partial \psi^2}\right|_{H_0}\right)$, $s_{02i} = \left.\frac{\partial l_i^*}{\partial \lambda}\right|_{H_0}$ and $v_{02i}=E\left(-\left.\frac{\partial^2 l_i^*}{\partial \lambda^2}\right|_{H_0}\right)$, $i = 1, ..., k$, respectively.

Now, asymptotically, as $n_i \to \infty, i = 1, \ldots, k$, the distributions of S, S_1 and S_2 are $\chi^2_{2(k-1)}, \chi^2_{k-1}$ and χ^2_{k-1} under H_0, H_0' and H_0'' respectively. Note that the two null hypotheses H_0 and H_0' are the same. It is then obvious that $S_{01} = S_1$. Further, S_2 is obtained by using the maximum likelihood estimates $\hat{\psi}_i, i = 1, 2, \ldots, k$ and $\hat{\lambda}$ under H_0'' and S_{02} is obtained by using the maximum likelihood estimates $\hat{\psi}$ and $\hat{\lambda}$ under H_0. Since, under some regularity conditions, the maximum likelihood estimates are consistent, then, asymptotically, as $n_i \to \infty, i = 1, \cdots, k$, under H_0, the estimates $\hat{\psi}_i, i = 1, ..., k$ and $\hat{\psi}$ all converge to ψ and $\hat{\lambda}$ converges to λ. Thus, S_2 and S_{02} are asymptotically equivalent. So asymptotically, we have

$S = S_1 + S_2.$

Therefore S_1 and S_2 are asymptotically independent.

Note that if the original parameters are orthogonal then the parameters Λ and Φ are identical, and S_1 and S_2 involve the parameters Ψ and Φ instead of Ψ and Λ. However, in the situations where the original parameters of the distribution are not orthogonal, we need to express the score test statistics in terms of the estimates of the original parameters as

$$\hat{s}_{1i} = \left(\frac{\partial l_i}{\partial \psi_i} - \frac{\partial l_i}{\partial \phi_i}\frac{i_{\psi_i\phi_i}}{i_{\phi_i\phi_i}}\right)\Bigg|_{H_0', \psi=\hat{\psi}, \phi=\hat{\phi}},$$

$$\hat{v}_{1i} = \left(i_{\psi_i\psi_i} - \frac{i^2_{\psi_i\phi_i}}{i_{\phi_i\phi_i}}\right)\Bigg|_{H_0', \psi=\hat{\psi}, \phi=\hat{\phi}},$$

$$\hat{s}_{2i} = \left(\frac{\partial l_i}{\partial \phi_i}\right)\Bigg|_{H_0'', \Psi=\hat{\Psi}, \phi=\hat{\phi}}$$

and

$$\hat{v}_{2i} = i_{\phi_i\phi_i}\big|_{H_0'', \Psi=\hat{\Psi}, \phi=\hat{\phi}}.$$

The details of derivations of $\hat{s}_{1i}$, $\hat{v}_{1i}$, $\hat{s}_{2i}$ and $\hat{v}_{2i}$ are given in Appendix A.2. Through those expressions, we can calculate the score test statistics S_1 and S_2 without solving the partial differential equations (3.1).

Let M_2 be the test statistic of Fisher's procedure for combining two score test statistics. Now, according to the Fisher's method, let

$L_1(t_1) = Pr(S_1 \geq t_1 \mid H_0')$ and $L_2(t_2) = Pr(S_2 \geq t_2 \mid H_0'')$.

Then it follows that

$$M_2 = -2\log[L_1(S_1)L_2(S_2)]$$

is approximately distributed as χ^2_4.

3.5 Conclusion

Singh (1986) develops a procedure for testing homogeneity of several normal populations based on combining two separate independent likelihood ratio test statistics using a method proposed by Fisher (1950). We extended Fisher's method to test homogeneity of several location-scale populations using two likelihood ratio statistics as well as two score test statistics. Asymptotic independence of the two likelihood ratio statistics and also of the two score test statistics have been established.

Chapter 4

Tests of the Homogeneity of Several Normal Populations

4.1 Introduction

The purpose of this chapter is to use the general result of Chapter 3 to test the homogeneity of the several normal populations based on combining the score test statistics and to compare our procedure with the procedure proposed by Singh (1986). Singh (1986) uses Fisher's method for testing simultaneously the equality of means and the equality of variances of several normal populations. For the case of several normal populations, exact independence of the two likelihood ratio statistics is shown by Singh (1986). In Chapter 3, we show, in general, the asymptotic independence of the two log-likelihood ratio statistics and also of the two score test statistics.

In this chapter, we develop a procedure to test the homogeneity of the several normal populations based on combining two score test statistics and we show exact independence of the two score test statistics. Simulations are also conducted to compare the procedure based on the score tests with that developed by Singh (1986). We also include the four other large sample tests in the simulation comparison. These are (i) the log-likelihood ratio test for simultaneously testing the equality of means and the equality of the variances; (ii) the score test for simultaneously testing the equality of means and the equality of the variances; (iii) the ordinary log-likelihood ratio procedure in which we first test the equality of the variances by using a log-likelihood ratio statistics and once this hypothesis is not rejected we test for the equality of the means by using a log-likelihood ratio statistic; (iv) the ordinary score test procedure in which we first test the equality of the variances by using a score test statistic and once this hypothesis is not rejected we test for the equality of the means by using a score test statistic.

In Section 4.2 we first review the procedure based on the likelihood ratio tests developed by Singh (1986) for testing the equality of the means and the equality of the variances of several normal populations, and then we develop a procedure based on score tests. In Section 4.3, we prove exact independence of the two score test statistics. Simulations are conducted in Section 4.4. The conclusions are given in Section 4.5.

4.2 Homogeneity of several normal $N(\mu, \sigma^2)$ populations

4.2.1 The likelihood ratio procedure

Singh (1986) applies Fisher's method to test homogeneity of several normal populations. Let $N(\mu_i, \sigma_i^2)$, $i = 1, 2, ..., k$, denote the ith normal distribution.

Then, testing homogeneity of the k normal populations implies testing

$H_0 : \mu_i = \mu, \sigma_i^2 = \sigma^2$, for all i against H_1: at least two μ's or two σ^2's are unequal, where μ and σ^2 are unspecified.

To test the above hypothesis the usual practice is to test the following two hypotheses separately.

$H_0' : \mu_i = \mu, \sigma_i^2 = \sigma^2$, for all i against H_1': at least two μ's are unequal and $\sigma_i^2 = \sigma^2$, for all i, where μ and σ^2 are unspecified,

and

H_0'': $\sigma_i^2 = \sigma^2$, for all i against H_1'': at least two σ^2's are unequal, where σ^2 are unspecified.

Let $x_{i1}, ..., x_{in_i}$ be the sample from $N(\mu_i, \sigma_i^2)$. Further, let $n = \sum_{i=1}^k n_i$, $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij}/n_i$, $\bar{x} = \sum_{i=1}^k n_i\bar{x}_i/n$, $\hat{\sigma_i}^2 = \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2/n_i$, $\hat{\sigma_0^2} = \sum_{i=1}^k \sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2/n$, $\hat{\sigma}^2 = \sum_{i=1}^k n_i\hat{\sigma_i}^2/n$, $s_i^2 = n_i\hat{\sigma}_i^2/(n_i - 1)$, $s^2 = n\hat{\sigma}^2/(n-k)$.

Then, the likelihood ratio statistic for testing H_0' against H_1' is

$$T_1 = (n-k)\{\sum_{i=1}^k n_i(\bar{x}_i - \bar{x})^2\}/\{(k-1)\sum_{i=1}^k n_i\hat{\sigma_i^2}\}$$

and the log-likelihood ratio statistic for testing H_0'' against H_1'' is

$$V = n\log\hat{\sigma^2} - \sum_{i=1}^k n_i \log\hat{\sigma_i^2}.$$

Under H_0', T_1 has an exact $F(k-1, n-k)$ distribution. Under H_0'', V is asymptotically distributed as $\chi_{k-1}^2(\alpha)$. Using a Bartlett correction, a modified likelihood ratio statistic is

$$T_2 = C^{-1}\{v\log s^2 - \textstyle\sum_{i=1}^k v_i \log s_i^2\},$$

where $v_i = n_i - 1$, $i = 1, ..., k$, $v = n - k$ and $C = 1 + \{\sum_{i=1}^k v_i^{-1} - v^{-1}\}/3(k-1)$.

It is well known that the Bartlett-corrected likelihood ratio statistic approximates better to the χ^2_{k-1} distribution than its uncorrected counterpart.

It can be shown that the statistics T_1 and T_2 are independently distributed (Singh, 1986). Now, define

$L_1(t_1) = Pr(T_1 \geq t_1 \mid H_0')$

and

$L_2(t_2) = Pr(T_2 \geq t_2 \mid H_0'')$.

Further, let NM_1 be the test statistic of Fisher's procedure for combining statistics T_1 and T_2. Since T_1 and T_2 are independently distributed, then following Fisher's method (Singh, 1986)

$NM_1 = -2\log[G_1(T_1)G_2(T_2)]$

is approximately distributed as $\chi^2(4)$. Thus, we reject H_0 in favor of H_1, if $NM_1 \geq \chi^2_4(\alpha)$, where $\chi^2_4(\alpha)$ is the $100\alpha\%$ point of the χ^2 distribution with 4 degrees of freedom.

4.2.2 The score test procedure

Let $x_{i1}, ..., x_{in_i}$ be the sample from $N(\mu_i, \sigma_i^2)$, $i = 1, ..., k$. As in section 4.2.1, the hypothesis in which we are interested is

$H_0 : \mu_i = \mu, \sigma_i^2 = \sigma^2$, for all i against H_1: at least two μ's or two σ^2's are unequal, where μ and σ^2 are unspecified.

Then, following Singh (1986), we have two hypothesis:

$H_0' : \mu_i = \mu, \sigma_i^2 = \sigma^2$, for all i against H_1': at least two μ's are unequal and $\sigma_i^2 = \sigma^2$, for all i, where μ and σ^2 are unspecified.

and

H_0'': $\sigma_i^2 = \sigma^2$, for all i against H_1'': at least two σ^2's are unequal, where σ^2 are unspecified.

The same notations n, $\bar{x}_i$, $\bar{x}$, $\hat{\sigma}_i{}^2$, $\hat{\sigma}_0^2$ and $\hat{\sigma}^2$ of Section 4.2.1 are used here. Then following the general results in Chapter 3, the score test statistic for testing H_0' against H_1' is

$$NS_1 = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{\hat{\sigma}_0^2}$$

and the score test statistic for testing H_0'' against H_1'' is

$$NS_2 = \frac{\sum_{i=1}^{k} n_i(\hat{\sigma}_i^2 - \hat{\sigma}^2)^2}{2(\hat{\sigma}^2)^2}.$$

Using the properties of the Dirichlet distribution, it can be shown that the two statistics NS_1 and NS_2 are exactly independent. The proof is given in Section 4.3.

Now, define

$L_1(t_1) = Pr(NS_1 \geq t_1 \mid H_0')$

and

$L_2(t_2) = Pr(NS_2 \geq t_2 \mid H_0'')$.

Further, let NM_2 be the test statistic of Fisher's procedure for combining statistics NS_1 and NS_2. Since NS_1 and NS_2 are independently distributed, then following Fisher's method

$NM_2 = -2\log[L_1(NS_1)L_2(NS_2)]$

is approximately distributed as $\chi^2(4)$. Thus, we reject H_0 in favor of H_1, if $NM_2 \geq \chi_4^2(\alpha)$, where $\chi_4^2(\alpha)$ is the $100\alpha\%$ point of the χ^2 distribution with 4 degrees of freedom.

4.3 Proof of exact independence of NS_1 and NS_2

For convenience we write NS_1 and NS_2 as

$$NS_1 = n/(1+\frac{1}{\frac{\sum_{i=1}^{k} n_i(\bar{x}_i-\bar{x})^2}{\sum_{i=1}^{k} u_i^2}}) = n/\left(1+\frac{1}{ST1}\right), \tag{4.3.1}$$

and

$$NS_2 = \sum_{i=1}^{k} \frac{n^2}{2n_i}\left(\frac{u_i^2}{\sum_{i=1}^{k} u_i^2}\right)^2 - \frac{n}{2} = ST2 - \frac{n}{2}, \tag{4.3.2}$$

where $u_i^2 = \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$, $i = 1,2,...,k$. It now suffices to prove that ST1 and ST2 are independently distributed. To prove this we use the following lemma .

Lemma (Hogg and Craig, 1995, p. 187): Let $x_1, x_2, ..., x_k$ be independent random variables, each being $\Gamma(\alpha_i, 1)$ with

$$\begin{aligned} y_i &= \frac{x_i}{x_1+x_2+\cdots+x_k}, \quad i = 1,2,\cdots,k-1, \\ y_k &= x_1+x_2+\cdots+x_k. \end{aligned}$$

Then

(1) $(y_1, y_2, ..., y_{k-1}) \sim$ Dirichlet Distribution with parameter $(\alpha_1, \alpha_2, \cdots, \alpha_k)$,

(2) y_k has gamma distribution $\sim \Gamma(\sum_{i=1}^{k}\alpha_i, 1)$,

(3) y_k is independent of $(y_1, y_2, ..., y_{k-1})$.

Using the above Lemma we only need to prove that each component of ST2 , namely, $\frac{u_i^2}{\sum_{i=1}^{k} u_i^2}$, $i = 1,2,...,k-1$, is independent of ST1.

From the property of the normal distribution of the x_{ij} 's, we know that under H_0, $2u_i^2/\sigma^2$, $i = 1,2,...,k-1$ are distributed as $\Gamma\left(\frac{n_i-1}{2}, 1\right)$. Further, under H_0,

$\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$ has distribution as $\sigma^2\chi^2(k-1)$ and also it is independent with $u_i^2, i = 1, 2, \ldots, k$. Now, let

$$\begin{aligned} y_i &= \frac{2u_i^2}{2u_1^2 + 2u_2^2 +, \cdots, +2u_k^2} = \frac{u_i^2}{\sum_{i=1}^{k} u_i^2}, i = 1, 2, \cdots, k-1, \\ y_k &= 2u_1^2 + 2u_2^2 +, \cdots, +2u_k^2 = 2\sum_{i=1}^{k} u_i^2. \end{aligned}$$

Then, from the property (3) of lemma and the above results we see that each of the random quantities $\sum_{i=1}^{k} u_i^2$, $\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$ and $\frac{u_i^2}{\sum_{i=1}^{k} u_i^2}$, $i = 1, 2, ..., k-1$ are independent of each other and hence ST1=$\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2 / \sum_{i=1}^{k} u_i^2$ and $\frac{u_i^2}{\sum_{i=1}^{k} u_i^2}$ are indpendently distributed from which the independence of NS_1 and NS_2 is established under H_0.

4.4 Simulation

A simulation study was conducted to compare the performance, in terms of size and power, of the statistic NM_1 based on the likelihood ratio statistics and the statistic NM_2 based on the score test statistics for testing homogeneity of several normal populations. In the comparison we have also included four other the log-likelihood ratio statistics and the score test statistics, given in what follows, for simultaneously testing the equality of means and the equality of variances of several normal populations. Using the notations in Section 4.2, these procedures are

(i) the log-likelihood ratio statistic (LR)

$$LR = \sum_{i=1}^{k} n_i \log\left(\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2}\right)$$

for simultaneously testing the equality of the means and the equality of the variances,

(ii) the score test statistic (S)

$$S = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2}{\hat{\sigma}_0^2} + \frac{\sum_{i=1}^{k} n_i[(\bar{x}_i - \bar{x})^2 + \hat{\sigma}_i^2 - \hat{\sigma}_0^2]^2}{2\left(\hat{\sigma}_0^2\right)^2}.$$

for simultaneously testing the equality of the means and the equality of the variances,

(iii) the ordinary log-likelihood ratio procedure (LRO) in which we first test the equality of the variances by using the log-likelihood ratio statistic T_2 and once this hypothesis is not rejected we test for the equality of the means using the log-likelihood ratio statistic T_1, and

(iv) the ordinary score test procedure (SO) in which we first test the equality of the variances by using the score test NS_1 and once this hypothesis is not rejected we test for the equality of the means using the score test statistic NS_2

We have considered K=2, 3 and 4 populations, two nominal levels $\alpha = 0.05$ and $\alpha = 0.10$ and equal sample sizes from each population. Results for k=3 and k=4 are similar. So, we give results for only k=2 and k=3. For calculating empirical size we generated samples from $N(0, 1)$ populations. For calculating empirical power we generated samples from $N(\mu, \sigma^2)$ populations for values of μ and σ^2 given in Table 4.1 to Table 4.4. Each simulation experiment was based on 10,000 samples. Results of the simulations are presented in Table 4.1 to Table 4.4. However, those results for LRO and SO, in general show, either extremely conservative or liberal behavior. So, we omitted simulation results for these procedures in the chapter.

Results in Table 4.1 to Table 4.4 show that the likelihood ratio test statistic (LR) for simultaneously testing the equality of means and variances of several normal populations shows liberal behavior. The corresponding score test statistic (S) shows conservative behavior for small sample sizes. The statistic NM_1 maintains level well in all situations studied here. The statistic NM_2 shows some conservative behavior for k=2 and small sample sizes. Otherwise it maintains level well. Power of both of these statistics are similar, although the statistic NM_1 shows slightly better power properties than the statistic NM_2 for small k (k=2) and small samples, because the later is conservative in these situations. However, for larger k, NM_2 has some edge over NM_1. It seems for large sample sizes that the statistic S will perform as well as the statistics NM_1 and NM_2.

4.5 Conclusion

Singh (1986) developed a procedure for testing homogeneity of several normal populations based on combining two separate independent likelihood ratio test statistics using a method proposed by Fisher (1950). We have developed procedures for testing homogeneity of several normal populations based on combining two separate independent score test statistics using Fisher's method. Exact independence of the two score test statistics have been established in the normal case. Compared with the ordinary method to test the homogeneity of several normal distribution, we see that Fisher's method works well irrespective of whether we combine two likelihood ratio test statistics or two score test statistics.

Table 4.1: Empirical power(%) of different statistics for testing homogeneity of K = 2 populations when data are simulated from the normal distributions $N(\mu_i, \sigma_i^2), i = 1, 2$ based on 10,000 simulations; $\alpha = 0.05$

Sample size	Test Statistic	(μ_1, μ_2) (σ_1^2, σ_2^2)					
		(0.0,0.0) (1.0,1.0)	(0.0,0.32) (1.0,1.40)	(0.0,0.64) (1.0,1.75)	(0.0,1.2) (1.0,3.5)	(0.0,2.0) (1.0,4.0)	(0.0,3.0) (1.0, 5.0)
5	NM_1	4.58	6.57	11.29	26.41	49.93	76.31
	LR	10.83	13.81	21.05	43.44	67.70	87.75
	NM_2	3.28	4.83	9.24	21.71	45.96	74.05
	S	0.91	1.46	3.26	7.86	22.38	44.03
7	NM_1	5.38	8.24	16.14	42.71	71.61	92.60
	LR	9.66	13.77	23.76	54.86	80.10	95.68
	NM_2	4.14	7.02	14.38	37.97	69.51	91.89
	S	2.25	3.91	8.33	23.56	53.50	82.70
9	NM_1	5.02	9.23	20.44	56.77	85.00	98.11
	LR	7.92	13.52	26.71	64.63	89.09	98.82
	NM_2	4.30	8.13	19.19	53.39	84.13	97.98
	S	2.65	5.63	13.95	42.84	76.05	95.96
11	NM_1	5.14	10.23	24.35	66.72	91.90	99.64
	LR	7.38	13.61	29.64	72.37	94.02	99.77
	NM_2	4.68	9.49	23.15	64.32	91.54	99.63
	S	3.35	7.45	18.72	56.88	87.67	99.21
13	NM_1	4.89	11.37	29.0	75.84	96.27	99.89
	LR	6.71	14.40	33.94	79.70	97.21	99.93
	NM_2	4.45	10.67	28.31	74.21	96.24	99.89
	S	3.55	8.54	24.00	68.61	94.35	99.79
15	NM_1	4.93	12.73	33.84	82.91	98.12	99.96
	LR	6.49	15.70	38.23	85.28	98.48	99.97
	NM_2	4.61	12.22	33.06	81.77	98.07	99.95
	S	3.74	10.34	29.49	78.24	97.41	99.93
20	NM_1	4.97	16.02	44.26	93.00	99.76	100.0
	LR	6.10	18.09	47.30	93.71	99.81	100.0
	NM_2	4.69	15.53	43.81	92.75	99.76	100.0
	S	4.09	14.02	40.60	91.02	99.64	100.0

Table 4.2: Empirical power(%) of different statistics for testing homogeneity of K = 2 populations when data are simulated from the normal distributions $N(\mu_i, \sigma_i^2), i = 1, 2$ based on 10,000 simulations; $\alpha = 0.10$

Sample size	Test Statistic	(μ_1, μ_2) (σ_1^2, σ_2^2)					
		(0.0,0.0) (1.0,1.0)	(0.0,0.32) (1.0,1.40)	(0.0,0.64) (1.0,1.75)	(0.0,1.2) (1.0,3.5)	(0.0,2.0) (1.0,4.0)	(0.0,3.0) (1.0, 5.0)
5	NM_1	9.77	12.59	19.56	41.04	66.22	87.16
	LR	18.23	22.58	32.01	57.68	78.46	93.68
	NM_2	7.95	10.78	17.91	37.57	64.62	87.14
	S	4.54	6.61	12.02	25.93	50.76	76.92
7	NM_1	10.74	15.28	25.86	58.09	82.65	96.58
	LR	16.58	22.21	34.91	67.50	88.14	98.01
	NM_2	9.71	14.38	25.10	56.01	82.23	96.50
	S	7.85	11.55	21.36	48.89	77.01	94.75
9	NM_1	9.94	16.85	31.60	69.82	91.97	99.30
	LR	14.25	22.46	38.76	75.94	94.46	99.50
	NM_2	9.33	16.20	31.18	68.78	92.01	99.28
	S	8.07	14.03	27.75	64.53	89.57	98.92
11	NM_1	10.35	17.66	35.90	78.34	96.30	99.88
	LR	13.46	22.35	41.69	82.26	97.29	99.93
	NM_2	9.92	17.30	35.64	77.82	96.34	99.89
	S	8.80	15.62	33.40	74.70	95.18	99.83
13	NM_1	9.91	19.46	41.89	85.34	98.36	100.0
	LR	12.63	23.8	46.63	87.83	98.70	100.0
	NM_2	9.67	19.26	41.75	85.04	98.36	100.0
	S	8.76	17.85	39.40	83.09	97.98	99.96
15	NM_1	10.19	21.41	47.13	90.00	99.16	100.0
	LR	12.48	25.19	51.34	91.68	99.33	100.0
	NM_2	10.06	21.18	46.94	89.74	99.18	100.0
	S	9.10	20.10	45.33	88.47	99.00	100.0
20	NM_1	9.67	25.80	57.36	96.45	99.91	100.0
	LR	11.57	28.41	60.17	96.91	99.92	100.0
	N_2	9.49	25.68	57.45	96.47	99.91	100.0
	S	9.38	24.66	55.68	95.84	99.88	100.0

Table 4.3: Empirical power(%) of different statistics for testing homogeneity of K = 3 populations when data are simulated from the normal distributions $N(\mu_i, \sigma_i^2), i = 1, 2, 3$ based on 10,000 simulations; $\alpha = 0.05$

Sample size	Test Statistic	(μ_1, μ_2, μ_3) $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$					
		(0.0,0.0,0.0) (1.0,1.0,1.0)	(0.0,0.08,0.32) (1.0,1.10,1.40)	(0.0,0.32,0.64) (1.0,1.2,1.75)	(0.0,0.64,1.2) (1.0,1.75,3.5)	(0.0,1.0,2.0) (1.0,2.0,4.0)	(0.0,1.5,3.0) (1.0,2.5,5.0)
5	NM_1	4.98	6.29	9.45	20.89	38.66	63.42
	LR	12.66	15.39	20.61	38.30	59.54	81.42
	NM_2	4.83	6.29	9.72	22.09	40.06	63.75
	S	2.19	3.07	5.27	12.93	25.18	43.68
7	NM_1	5.06	7.30	12.67	32.83	59.25	85.36
	LR	10.32	13.47	21.07	45.97	71.15	91.83
	NM_2	4.96	7.34	13.01	33.75	59.55	84.83
	S	3.13	4.89	9.35	25.16	45.55	69.57
9	NM_1	4.92	8.11	15.92	44.77	74.87	95.02
	LR	8.70	13.34	22.92	55.03	82.71	97.19
	NM_2	4.79	8.13	16.33	45.22	74.69	94.58
	S	3.50	6.48	13.14	36.39	63.61	86.87
11	NM_1	4.99	8.93	19.13	55.84	85.48	98.54
	LR	7.76	12.86	24.46	62.99	89.19	99.21
	NM_2	5.01	9.18	19.69	55.91	85.16	98.37
	S	3.81	7.48	15.57	46.59	76.13	95.31
13	NM_1	5.11	10.12	22.78	65.23	91.93	99.59
	LR	7.28	13.49	28.02	72.05	94.14	99.83
	NM_2	5.01	10.31	23.44	65.07	91.56	99.52
	S	3.86	8.72	20.43	58.75	85.89	98.48
15	NM_1	5.04	10.82	26.04	73.91	95.56	99.86
	LR	7.00	13.52	31.00	78.52	96.73	99.91
	NM_2	4.99	10.85	26.83	73.87	95.35	99.83
	S	3.95	9.72	23.79	67.39	92.14	99.54
20	NM_1	5.14	13.49	34.95	87.11	99.23	100.0
	LR	6.34	15.90	38.76	89.33	99.40	100.0
	NM_2	5.04	13.74	35.74	86.78	99.20	100.0
	S	4.01	12.88	33.14	85.18	98.63	100.0

Table 4.4: Empirical power(%) of different statistics for testing homogeneity of K=3 populations when data are simulated from the normal distributions $N(\mu_i, \sigma_i^2), i = 1, 2, 3$ based on 10,000 simulations; $\alpha = 0.10$

Sample size	Test Statistic	(μ_1, μ_2, μ_3) $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$					
		(0.0,0.0,0.0) (1.0,1.0,1.0)	(0.0,0.08,0.32) (1.0,1.10,1.40)	(0.0,0.32,0.64) (1.0,1.2,1.75)	(0.0,0.64,1.2) (1.0,1.75,3.5)	(0.0,1.0,2.0) (1.0,2.0,4.0)	(0.0,1.5,3.0) (1.0,2.5,5.0)
5	NM_1	9.82	12.09	16.97	33.14	53.96	77.43
	LR	20.94	24.19	31.41	51.83	71.40	89.13
	NM_2	10.12	12.61	17.95	35.21	55.59	77.86
	S	6.37	8.74	12.68	26.05	43.54	63.68
7	NM_1	10.12	13.62	21.46	46.87	72.77	92.60
	LR	17.31	21.82	31.53	59.16	81.47	96.18
	NM_2	10.07	13.96	22.19	47.98	73.24	92.30
	S	7.98	11.13	18.34	39.68	62.65	83.81
9	NM_1	9.89	14.89	25.69	58.83	85.00	97.87
	LR	15.96	21.55	33.82	68.30	89.85	98.94
	NM_2	9.90	15.17	26.49	59.41	85.00	97.70
	S	8.18	13.19	23.28	52.29	78.10	94.49
11	NM_1	10.07	16.16	29.82	69.03	92.21	99.46
	LR	14.13	21.05	35.94	74.73	94.32	99.72
	NM_2	10.22	16.39	30.69	69.17	92.10	99.40
	S	8.34	14.03	26.82	62.20	87.25	98.58
13	NM_1	10.11	17.50	34.36	76.97	96.00	99.86
	LR	13.53	21.67	40.86	81.64	97.14	99.94
	NM_2	10.09	17.83	35.17	77.09	95.94	99.86
	S	8.80	16.10	32.80	72.02	93.45	99.58
15	NM_1	10.14	18.57	38.22	83.39	97.98	99.97
	LR	12.87	22.64	43.54	86.49	98.51	99.98
	NM_2	10.11	19.15	39.14	83.55	97.87	99.95
	S	8.73	17.42	36.68	79.61	96.52	99.87
20	NM_1	9.99	22.09	47.52	92.97	99.74	100.0
	LR	12.13	25.05	51.54	94.33	99.79	100.0
	NM_2	10.02	22.53	48.33	92.88	99.69	100.0
	S	8.96	21.62	46.28	90.89	99.42	100.0

Chapter 5

Tests of the Homogeneity of Several Non-normal Populations

5.1 Introduction

In Chapter 3, we obtain general results for testing homogeneity of several location-scale populations. In Chapter 4, we obtain and compare two statistics for testing the homogeneity of the several normal populations. In this chapter, by applying the general results, we obtain two procedures for testing homogeneity of some non-normal populations. Here we consider two over-dispersed discrete models, namely the negative binomial model and the beta-binomial model. We also consider a widely used lifetime model, namely the Weibull or extreme-value model. In all three cases simulations are conducted to compare the two procedures. We omit the details of derivation of the log-likelihood ratio statistics for these three models. They are easy to obtain but have complicated expressions involving estimates of the parameters under the alternative hypotheses. We denote the log-likelihood ratio based statistics, analogous to the statistic NM_1 discussed in Chapter 4, for testing homogeneity of negative binomial, beta-binomial and Weibull populations by NBM_1, BBM_1 and

WBM_1 respectively.

In section 5.2, we deal with several negative binomial (m, c) populations. In section 5.3, we deal with several beta-binomial (π, ϕ) populations. Section 5.4 is devoted to the several Weibull (ψ, ϕ) populations. The chapter ends with a concluding section 5.5.

5.2 Homogeneity of several negative binomial $NB(m, c)$ populations

5.2.1 Fisher's procedure for combining two score test statistics

Now let $x_{i1}, ..., x_{in_i}$ be a sample from the negative binomial distribution $NB(m_i, c_i)$, $i = 1, ..., k$, with probability mass function

$$Pr(X = x) = \frac{\Gamma(x + c^{-1})}{x!\Gamma(c^{-1})} \left(\frac{cm}{1 + cm}\right)^x \left(\frac{1}{1 + cm}\right)^{c^{-1}},$$

where m is the mean and c is the dispersion parameter. Note that the mean and variance of X are m and $m(1+cm)$. Thus, homogeneity of the $NB(m_i, c_i)$ populations, $i = 1, ..., k$, implies $m_i = m$ and $c_i = c$ for all $i = 1, ..., k$. Let c_0 and c_1 be the common value of $c_i, i = 1, ..., k$, under $H_0' : m_i = m, c_i = c$, for all i and $H_0'' : c_i = c$, for all i respectively. Then again following the results in section 3.2, the score test statistic for testing H_0' against H_1': at least two m's are unequal and $c_i = c$, for all i is

$$NBS_1 = \sum_{i=1}^{k} \frac{n_i(\bar{x}_i - \bar{x})^2}{\bar{x}(1 + \hat{c}_0\bar{x})},$$

where $\hat{c}_0$ is the maximum likelihood estimate of c_0 obtained by solving the maximum likelihood estimating equation

$$n \log(1 + c_0 \bar{x}) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \sum_{l=1}^{x_{ij}} \frac{c_0}{1 + c_0(l-1)}.$$

The score test statistic testing H_0'' against H_1'': at least two c_i's are unequal is

$$NBS_2 = \sum_{i=1}^{k} \frac{d_i^2}{n_i \lambda_i},$$

where

$$d_i = \hat{c}_1^{-2} \sum_{j=1}^{n_i} \left[\log(1 + \hat{c}_1 \bar{x}_i) - \hat{c}_1 \sum_{l=1}^{x_{ij}} \frac{1}{1 + \hat{c}_1(l-1)} \right],$$

$$q_i = \frac{\hat{c}_1 \bar{x}_i}{1 + \hat{c}_1 \bar{x}_i},$$

$$\lambda_i = \hat{c}_1^{-4} \sum_{j=1}^{\infty} \frac{j!(\hat{c}_1 q_j)^{j+1}}{(j+1) \prod_{l=1}^{j} (1 + l\hat{c}_1)},$$

and $\hat{c}_1$ is the maximum likelihood estimate of c_1 obtained by solving the maximum likelihood estimating equation

$$\sum_{i=1}^{k} n_i \log(1 + c_1 \bar{x}_i) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \sum_{l=1}^{x_{ij}} \frac{c_1}{1 + c_1(l-1)}.$$

These score tests have also been obtained by Barnwal and Paul (1988). From the general proof in Section 3.2 it is obvious that, asymptotically, as $n_i \to \infty, i = 1, ..., k$, the statistics NBS_1 and NBS_2 are independent. We denote the statistic obtained by combining the score test statistics NBS_1 and NBS_2 by NBM_2.

5.2.2 Simulation

In the simulation study we considered K=2, 3 and 4 populations, two nominal levels $\alpha = 0.05$ and $\alpha = 0.10$ and equal sample sizes from each population. Each simulation experiment was based on 10,000 samples.

For calculating empirical size, we generated samples from NB (m,c) populations with equal m's and equal c's. Unequal m's and unequal c's were considered for power calculations. Here also results for k=3 and k=4 are similar. So, we give results for only k=2 and k=3 with $\alpha = 0.05$ and $\alpha = 0.10$ respectively. Results of the simulations for k=2 with $\alpha = 0.05$ and $\alpha = 0.10$ are presented in Table 5.1 and Table 5.2, and those for k=3 are presented in Table 5.3 and Table 5.4. In the simulation study we have also considered other values of m. The empirical level and power results are similar to those presented in Table 5.1 to Table 5.4. So, we omit them here.

Results in Table 5.1 to Table 5.4 show that the statistic NBM_1 is in general liberal, whereas the statistic NBM_2 maintains level well. Power of the statistic NBM_1 is in general larger than that of the statistic NBM_2. This is not surprising as the statistic NBM_1 is in general liberal.

Further, we have extended the simulation experiment to study size adjusted power properties of these two statistics. The empirical 95% quantiles derived from the corresponding size simulation have been used to ensure that each test had approximately the nominal size of 0.05. Empirical quantiles were calculated based on 40,000 replications and empirical power calculations were based on 10,000 replications. In Table 5.5, we provide empirical power values for $k = 2$, $m_1 = m_2 = 2.0$, $c_1 = c_2 = 0.05$, $n_1 = n_2 = 10, 15, 20, 40$ and for different combinations of the unequal $m's$ and unequal $c's$. Results in Table 5.5 show that both the size adjusted statistics NBM_1 and NBM_2 have similar power.

5.2.3 Example

Example 1 (McCaughran & Arnold, 1976). The data in Data sets, Table D.1 refer to counts of embryonic dearths in a control group and two treatment groups. Analysis of the data, based on the NB model, gives $NBLR_1 = 3.259$, $NBLR_2 = 0.016$, $NBS_1 = 3.01$ and $NBS_2 = 0.22$. From these the values of NBM_1 and NBM_2 are 3.275 and 3.023 with p-values .513 and .554 respectively. Neither of these procedures reject the hull hypothesis of homogeneity of the two groups.

Table 5.1: Empirical power(%) of different statistics for testing homogeneity of K=2 negative binomial populations when data are simulated from $NB(m_i, c_i)$, $i = 1, 2$; based on 10,000 replications; $\alpha = 0.05$

Sample size	Test Statistic	(m_1, m_2) (c_1, c_2)					
		(2,2) (0.05,0.05)	(2,2.5) (0.05,0.10)	(2,3.0) (0.05,0.15)	(2,3.5) (0.05,0.2)	(2,4.0) (0.05,0.25)	(2,4.5) (0.05,0.3)
10	NBM_1	7.25	11.60	22.33	35.44	48.83	61.24
	NBM_2	4.23	6.70	14.01	24.38	35.45	46.64
15	NBM_1	7.53	14.18	29.38	48.10	65.91	79.01
	NBM_2	4.65	9.40	22.42	39.86	57.49	71.47
20	NBM_1	7.09	15.00	35.57	59.52	78.28	89.59
	NBM_2	4.47	10.94	29.34	53.08	73.29	86.18
30	NBM_1	6.27	18.45	49.94	78.88	92.93	97.94
	NBM_2	4.94	16.35	46.16	75.6	91.67	97.40
40	NBM_1	6.75	23.76	62.80	89.45	97.92	99.73
	NBM_2	4.75	21.01	60.01	87.98	97.50	99.64
50	NBM_1	5.52	27.20	72.92	95.25	99.47	99.96
	NBM_2	4.49	25.23	70.74	94.69	99.35	99.96

Table 5.2: Empirical power(%) of different statistics for testing homogeneity of K=2 negative binomial populations when data are simulated from $NB(m_i, c_i)$, $i = 1, 2$; based on 10,000 replications; $\alpha = 0.10$

Sample size	Test Statistic	(m_1, m_2) (c_1, c_2)					
		(2,2) (0.05,0.05)	(2,2.5) (0.05,0.10)	(2,3.0) (0.05,0.15)	(2,3.5) (0.05,0.2)	(2,4.0) (0.05,0.25)	(2,4.5) (0.05,0.3)
10	NBM_1	13.29	19.73	33.04	48.81	63.02	73.45
	NBM_2	8.90	13.96	25.73	39.96	53.71	65.65
15	NBM_1	13.67	22.30	40.94	61.65	77.46	87.39
	NBM_2	9.66	17.41	35.51	55.55	72.71	84.14
20	NBM_1	12.82	23.64	48.53	71.95	86.90	94.74
	NBM_2	9.43	19.83	43.81	67.96	84.41	93.40
30	NBM_1	11.52	28.54	62.87	87.11	96.26	99.14
	NBM_2	10.03	26.38	59.69	85.01	95.7	98.87
40	NBM_1	11.77	34.51	74.48	94.42	99.07	99.92
	NBM_2	9.70	31.85	73.07	93.71	98.93	99.92
50	NBM_1	10.79	39.09	82.65	97.77	99.8	99.99
	NBM_2	9.16	36.85	81.25	97.64	99.7	100.0

Table 5.3: Empirical power(%) of different statistics for testing homogeneity of K= 3 negative binomial populations when data are simulated from $NB(m_i, c_i)$, $i = 1, 2, 3$; based on 10,000 replications; $\alpha = 0.05$

Sample size	Test Statistic	(m_1, m_2, m_3) (c_1, c_2, c_3)				
		(2.0, 2.0, 2.0)	(2.0, 2.25, 2.5)	(2.0, 2.5, 3.0)	(2.0, 2.75, 3.5)	(2.0, 3.25, 4.5)
		(0.05,0.05,0.05)	(0.05,0.075,0.1)	(0.05,0.1,.15)	(0.05,0.125,0.2)	(0.05,0.175,0.3)
10	NBM_1	7.53	10.47	18.09	28.78	52.32
	NBM_2	4.66	6.07	10.45	17.80	35.64
15	NBM_1	7.97	12.80	23.58	39.98	70.37
	NBM_2	4.94	7.87	17.40	30.64	60.01
20	NBM_1	7.57	12.71	29.63	51.43	83.80
	NBM_2	4.80	9.01	23.10	43.62	78.03
30	NBM_1	6.13	15.2	41.84	70.36	96.29
	NBM_2	5.63	14.55	39.90	69.14	95.49
40	NBM_1	6.10	17.41	51.97	82.63	99.19
	NBM_2	4.95	15.32	48.16	79.92	98.94
50	NBM_1	5.78	21.16	62.84	91.40	99.88
	NBM_2	4.74	21.12	63.83	91.34	99.83

Table 5.4: Empirical power(%) of different statistics for testing homogeneity of K= 3 negative binomial populations when data are simulated from $NB(m_i, c_i)$, $i = 1, 2, 3$; based on 10,000 replications; $\alpha = 0.10$

Sample size	Test Statistic	(m_1, m_2, m_3) (c_1, c_2, c_3)				
		(2.0, 2.0, 2.0) (0.05,0.05,0.05)	(2.0, 2.25, 2.5) (0.05,0.075,0.1)	(2.0, 2.5, 3.0) (0.05,0.1,.15)	(2.0, 2.75, 3.5) (0.05,0.125,0.2)	(2.0, 3.25, 4.5) (0.05,0.175,0.3)
10	NBM_1	14.06	18.38	28.85	41.66	65.97
	NBM_2	8.90	12.25	20.31	30.37	53.52
15	NBM_1	15.04	20.91	35.04	53.11	81.27
	NBM_2	9.92	15.30	28.24	45.40	74.60
20	NBM_1	13.24	21.20	42.54	64.31	90.45
	NBM_2	9.57	16.52	36.12	58.56	87.26
30	NBM_1	11.88	24.84	54.89	80.83	98.2
	NBM_2	9.63	21.5	51.25	77.61	97.59
40	NBM_1	11.65	27.67	64.58	89.85	99.70
	NBM_2	10.00	25.11	61.82	88.25	99.60
50	NBM_1	11.23	31.78	73.91	95.4	99.97
	NBM_2	10.11	30.01	72.17	94.72	99.97

Table 5.5: Size adjusted empirical power(%) of the statistics NBM_1 and NBM_2 for testing homogeneity of K=2 negative binomial populations when data are simulated from $NB(m_i, c_i)$, $i = 1, 2$; empirical quantiles based on 40,000 replications; empirical size based on 10,000 replications; $\alpha = 0.05$

Sample size	Test Statistic	(m_1, m_2) (c_1, c_2)					
		(2,2) (0.05,0.05)	(2,2.5) (0.05,0.10)	(2,3.0) (0.05,0.15)	(2,3.5) (0.05,0.2)	(2,4.0) (0.05,0.25)	(2,4.5) (0.05,0.3)
10	NBM_1	4.84	6.24	9.99	15.43	22.46	30.76
	NBM_2	5.11	6.58	10.94	17.40	25.17	33.59
15	NBM_1	5.10	7.58	13.99	24.29	36.55	49.06
	NBM_2	5.03	7.46	14.11	24.72	37.25	50.35
20	NBM_1	4.89	7.95	17.51	32.49	48.61	63.72
	NBM_2	4.94	8.19	17.98	32.97	49.22	64.27
40	NBM_1	4.66	12.12	35.41	63.43	83.45	93.81
	NBM_2	4.75	12.16	35.24	63.43	83.60	93.80

5.3 Homogeneity of several beta-binomial $BB(\pi, \phi)$ populations

5.3.1 Fisher's procedure for combining two score test statistics

Now let $x_{i1}, ..., x_{in_i}$ be a sample from the beta binomial distribution $BB(\pi_i, \phi_i), i = 1, ..., k$, with probability mass function

$$Pr(X = x|m) = \begin{pmatrix} m \\ x \end{pmatrix} \frac{\prod_{r=0}^{x-1}(\pi(1-\phi) + r\phi) \prod_{r=0}^{m-x-1}((1-\pi)(1-\phi) + r\phi)}{\prod_{r=0}^{m-1}(1-\phi+r\phi)},$$

where π is the proportion parameter and ϕ is the dispersion parameter. Note that the mean and variance of X are $m\pi$ and $m\pi(1-\pi)(1+(m-1)\phi)$ respectively. Thus, testing the equality of means and equality of variances of the $BB(\pi_i, \phi_i)$ populations, $i = 1, ..., k$, is equivalent to testing $\pi_i = \pi$ and $\phi_i = \phi$ for all $i = 1, ..., k$.

Now, from the general results in Section 3.2, we obtain the score test statistic for testing

$H_0' : \pi_i = \pi, \phi_i = \phi$, for all i against H_1' : at least two π's are unequal and $\phi_i = \phi$, for all i as

$$BBS_1 = \sum_{i=1}^{k} \frac{s_{1i}^2}{v_{1i}},$$

where $s_{1i} = l_{i1} - l_{i2}l_{i12}/l_{i11}$, $v_{1i} = l_{i11} - l_{i12}^2/l_{i22}$

$$l_{i1} = \sum_{j=1}^{n_i} \left[\sum_{r=1}^{x_{ij}} \frac{(1-\hat{\phi})}{\hat{\pi}(1-\hat{\phi}) + (r-1)\hat{\phi}} - \sum_{r=1}^{m_{ij}-x_{ij}} \frac{(1-\hat{\phi})}{(1-\hat{\pi})(1-\hat{\phi}) + (r-1)\hat{\phi}} \right]$$

$$l_{i2} = \sum_{j=1}^{n_i} \left[\sum_{r=1}^{x_{ij}} \frac{(r-1) - \hat{\pi}}{\hat{\pi}(1-\hat{\phi}) + (r-1)\hat{\phi}} + \sum_{r=1}^{m_{ij}-x_{ij}} \frac{(r-1) - (1-\hat{\pi})}{(1-\hat{\pi})(1-\hat{\phi}) + (r-1)\hat{\phi}} \right.$$

$$-\sum_{r=1}^{m_{ij}} \frac{r-2}{1-\hat{\phi}+(r-1)\hat{\phi}}\Bigg],$$

$$l_{i11} = (1-\hat{\phi})^2 \sum_{j=1}^{n_i} \left[\sum_{r=1}^{m_{ij}} \frac{P(x_{ij} \geq r)}{\{\hat{\pi}(1-\hat{\phi})+(r-1)\hat{\phi}\}^2} + \sum_{r=1}^{m_{ij}} \frac{P(x_{ij} \leq m_{ij}-r)}{\{(1-\hat{\pi})(1-\hat{\phi})+(r-1)\hat{\phi}\}^2}\right],$$

$$l_{i12} = \sum_{j=1}^{n_i} \Bigg[-\frac{\hat{\pi}(1-\hat{\phi})}{\hat{\phi}} \sum_{r=1}^{m_{ij}} \frac{P(x_{ij} \geq r)}{\{\hat{\pi}(1-\hat{\phi})+(r-1)\hat{\phi}\}^2} +$$

$$\frac{(1-\hat{\pi})(1-\hat{\phi})}{\hat{\phi}} \sum_{r=1}^{m_{ij}} \frac{P(x_{ij} \leq m_{ij}-r)}{\{(1-\hat{\pi})(1-\hat{\phi})+(r-1)\hat{\phi}\}^2}\Bigg],$$

$$l_{i22} = \frac{1}{\hat{\phi}^2} \sum_{j=1}^{n_i} \Bigg[\hat{\pi}^2 \sum_{r=1}^{m_{ij}} \frac{P(x_{ij} \geq r)}{\{\hat{\pi}(1-\hat{\phi})+(r-1)\hat{\phi}\}^2}$$

$$+(1-\hat{\pi})^2 \sum_{r=1}^{m_{ij}} \frac{P(x_{ij} \leq m_{ij}-r)}{\{(1-\hat{\pi})(1-\hat{\phi})+(r-1)\hat{\phi}\}^2} - \sum_{r=1}^{m_{ij}} \frac{1}{\{1-\hat{\phi}+(r-1)\hat{\phi}\}^2}\Bigg]$$

and $\hat{\pi}$ and $\hat{\phi}$ are the maximum likelihood estimates of π and ϕ under H_0', obtained by solving the maximum likelihood estimating equations

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\sum_{r=1}^{x_{ij}} \frac{(1-\phi)}{\pi(1-\phi)+(r-1)\phi} - \sum_{r=1}^{m_{ij}-x_{ij}} \frac{(1-\phi)}{(1-\pi)(1-\phi)+(r-1)\phi} = 0$$

and

$$\sum_{j=1}^{k}\sum_{j=1}^{n_i}\sum_{r=1}^{x_{ij}} \frac{(r-1)-\pi}{\pi(1-\phi)+(r-1)\phi} + \sum_{r=1}^{m_{ij}-x_{ij}} \frac{(r-1)-(1-\pi)}{(1-\pi)(1-\phi)+(r-1)\phi}$$

$$- \sum_{r=1}^{m_{ij}} \frac{r-2}{1-\phi+(r-1)\phi} = 0,$$

simultaneously.

Similarly, the score test statistic for testing

$H_0'' : \phi_i = \phi$, for all i against H_1'' : at least two ϕ''s are unequal , for all i

is

$$BBS_2 = \sum_{i=1}^{k} \frac{s_{2i}^2}{v_{2i}},$$

where

$$s_{2i} = \sum_{j=1}^{n_i} \left[\sum_{r=1}^{x_{ij}} \frac{(r-1)-\hat{\pi}_i}{\hat{\pi}_i(1-\hat{\phi}')+(r-1)\hat{\phi}'} + \sum_{r=1}^{m_{ij}-x_{ij}} \frac{(r-1)-(1-\hat{\pi}_i)}{(1-\hat{\pi}_i)(1-\hat{\phi}')+(r-1)\hat{\phi}'} - \sum_{r=1}^{m_{ij}} \frac{r-2}{1-\hat{\phi}'+(r-1)\hat{\phi}'} \right],$$

$$v_{2i} = \frac{1}{\hat{\phi}'^2} \sum_{j=1}^{n_i} \left[\hat{\pi}_i^2 \sum_{r=1}^{m_{ij}} \frac{P(x_{ij} \geq r)}{\{\hat{\pi}_i(1-\hat{\phi}')+(r-1)\hat{\phi}'\}^2} + (1-\hat{\pi}_i)^2 \sum_{r=1}^{m_{ij}} \frac{P(x_{ij} \leq m_{ij}-r)}{\{(1-\hat{\pi}_i)(1-\hat{\phi}')+(r-1)\hat{\phi}'\}^2} - \sum_{r=1}^{m_{ij}} \frac{1}{\{1-\hat{\phi}'+(r-1)\hat{\phi}'\}^2} \right]$$

and $\hat{\pi}_i, i = 1, ..., k$ and $\hat{\phi}'$ are the maximum likelihood estimates of $\pi_i, i = 1, ..., k$ and ϕ' under H_0'', obtained by solving the maximum likelihood estimating equations

$$\sum_{j=1}^{n_i} \sum_{r=1}^{x_{ij}} \frac{(1-\phi')}{\pi_i(1-\phi')+(r-1)\phi'} - \sum_{r=1}^{m_{ij}-x_{ij}} \frac{(1-\phi')}{(1-\pi_i)(1-\phi')+(r-1)\phi'} = 0, i = 1, ..., k,$$

and

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \sum_{r=1}^{x_{ij}} \frac{(r-1)-\pi_i}{\pi_i(1-\phi')+(r-1)\phi'} + \sum_{r=1}^{m_{ij}-x_{ij}} \frac{(r-1)-(1-\pi_i)}{(1-\pi_i)(1-\phi')+(r-1)\phi'} - \sum_{r=1}^{m_{ij}} \frac{r-2}{1-\phi'+(r-1)\phi'} = 0$$

simultaneously.

Again, from the general results in Section 3.2, it is obvious that, asymptotically, as $n_i \to \infty, i = 1, ..., k$, the statistics BBS_1 and BBS_2 are independent. We denote the statistic obtained by combining the score test statistics BBS_1 and BBS_2 by BBM_2.

5.3.2 Simulation

In the simulation study we considered K=2, 3 and 4 populations, two nominal levels $\alpha = 0.05$ and $\alpha = 0.10$ and equal sample sizes from each population. Each simulation experiment was based on 10,000 samples.

For size calculation in the beta-binomial case, we generated samples from $BB(\pi, \phi)$ populations with equal π's and equal ϕ's. Unequal π's and unequal ϕ's were considered for power calculations. The beta-binomial index m was generated from a discrete uniform (1, 16) distribution, as in many toxicological data sets m varies from 1 to 16(see data in Data sets, Table D.2). Here also results for k=3 and k=4 are similar. So, we give results for only k=2 and k=3 with $\alpha = 0.05$ and $\alpha = 0.10$ respectively. Results of the simulations for k=2 of $\alpha = 0.05$ and $\alpha = 0.10$ are presented in Table 5.5 and 5.6 and those for k=3 of $\alpha = 0.05$ and $\alpha = 0.10$ are presented in Table 5.7 and 5.8.

According to results in Table 5.5 to Table 5.8, the statistic BBM_1 shows some conservative behavior for small sample sizes ($n \leq 15$); otherwise it holds level well. The statistic BBM_2 maintains level well in all situations studied here. Power properties of both the statistics are similar.

5.3.3 Example

Example 2. (Paul, 1982). The data in Data sets, Table D.2 refer to live foetuses in a litter affected by treatment, and the number of live foetuses, for each of k=4 doses groups: control(C), low dose(L), medium dose(M), and high dose(H). Analysis of the data, based on the BB model, gives $BBLR_1 = 10.89$, $BBLR_2 = 2.865$, $BBS_1 = 11.62$ and $BBS_2 = 2.38$. From these the values of BBM_1 and BBM_2 are 10.56 and 10.855 with p-values 0.032 and 0.0285 respectively. Both procedures reject the hull hypothesis of homogeneity of the four groups.

Table 5.6: Empirical power(%) of different statistics for testing homogeneity of K=2 beta binomial populations when data are simulated from $BB(m_i,\ \pi_i,\ \phi_i)$, $i = 1, 2$; based on 10,000 replications; $\alpha = 0.05$

Sample size	Test Statistic	(π_1, π_2) / (ϕ_1, ϕ_2)					
		(0.30,0.30)	(0.30,0.35)	(0.30,0.40)	(0.30,0.45)	(0.30,0.50)	(0.30,0.55)
		(0.10,0.10)	(0.10,0.12)	(0.10,0.14)	(0.10,0.16)	(0.10,0.18)	(0.10,0.20)
10	BBM_1	2.50	3.72	7.63	15.31	25.53	39.06
	BBM_2	3.59	5.06	9.38	15.92	25.23	36.82
15	BBM_1	3.62	6.37	14.18	27.67	44.05	63.14
	BBM_2	4.62	7.02	14.09	27.08	43.19	59.48
20	BBM_1	4.51	8.39	18.45	36.54	56.95	76.86
	BBM_2	4.66	8.17	18.24	35.52	55.21	75.00
30	BBM_1	5.23	11.42	28.98	55.34	79.27	93.16
	BBM_2	4.88	11.02	27.63	53.80	78.41	93.09
40	BBM_1	5.19	13.69	36.50	67.03	89.86	97.82
	BBM_2	4.94	12.91	36.10	67.91	88.57	97.74
50	BBM_1	5.48	14.91	44.5	77.40	95.03	99.40
	BBM_2	4.82	14.87	44.84	78.51	95.23	99.42

Table 5.7: Empirical power(%) of different statistics for testing homogeneity of K=2 beta binomial populations when data are simulated from $BB(m_i,\ \pi_i,\ \phi_i)$, $i = 1, 2$; based on 10,000 replications; $\alpha = 0.10$

Sample size	Test Statistic	(π_1, π_2) (ϕ_1, ϕ_2)					
		(0.30,0.30) (0.10,0.10)	(0.30,0.35) (0.10,0.12)	(0.30,0.40) (0.10,0.14)	(0.30,0.45) (0.10,0.16)	(0.30,0.50) (0.10,0.18)	(0.30,0.55) (0.10,0.20)
10	BBM_1	6.61	8.38	15.03	25.53	38.21	52.76
	BBM_2	8.57	10.85	18.01	27.65	39.52	52.10
15	BBM_1	8.07	12.72	23.18	40.32	58.21	75.11
	BBM_2	10.15	14.04	24.20	40.25	57.62	73.67
20	BBM_1	9.51	14.95	29.30	49.63	69.56	85.61
	BBM_2	9.88	14.83	29.85	49.59	69.13	85.06
30	BBM_1	10.86	18.87	38.11	63.31	84.31	94.5
	BBM_2	10.04	18.24	37.54	62.91	83.06	94.27
40	BBM_1	10.48	19.39	42.05	68.40	87.38	96.62
	BBM_2	10.20	19.11	40.56	66.89	86.94	96.72
50	BBM_1	10.29	22.18	49.30	77.89	94.24	99.06
	BBM_2	10.17	21.56	49.30	79.31	93.95	98.94

Table 5.8: Empirical power(%) of different statistics for testing homogeneity of K=3 beta binomial populations when data are simulated from $BB(m_i,\ \pi_i,\ \phi_i)$, $i = 1, 2, 3$; based on 10,000 replications; $\alpha = 0.05$

Sample size	Test Statistic	(π_1, π_2, π_3) (ϕ_1, ϕ_2, ϕ_3)					
		(0.30,0.30,0.30) (0.10,0.10,0.10)	(0.30,0.32,0.35) (0.10,0.11,0.12)	(0.30,0.35,0.40) (0.10,0.12,0.14)	(0.30,0.37,0.45) (0.10,0.13,0.16)	(0.30,0.40,0.50) (0.10,0.14,0.18)	(0.30,0.42,0.55) (0.10,0.15,0.20)
10	BBM_1	2.24	2.96	6.19	11.46	20.49	32.11
	BBM_2	4.12	5.13	8.04	13.38	21.18	31.77
15	BBM_1	3.13	5.38	10.3	20.90	35.82	52.78
	BBM_2	4.49	6.73	11.67	20.81	33.84	50.44
20	BBM_1	4.04	7.32	14.74	28.52	47.94	67.96
	BBM_2	4.74	7.04	13.91	28.13	45.23	64.09
30	BBM_1	4.83	9.22	21.39	42.52	66.45	85.60
	BBM_2	5.11	9.04	20.67	40.89	65.16	84.10
40	BBM_1	5.04	10.57	29.16	57.68	82.50	95.62
	BBM_2	4.73	10.17	27.52	56.81	82.79	95.51
50	BBM_1	4.83	11.71	34.18	67.03	90.44	98.23
	BBM_2	5.06	12.77	35.54	67.27	90.46	98.55

Table 5.9: Empirical power(%) of different statistics for testing homogeneity of K=3 beta binomial populations when data are simulated from $BB(m_i,\ \pi_i,\ \phi_i)$, $i = 1, 2, 3$; based on 10,000 replications; $\alpha = 0.10$

Sample size	Test Statistic	(π_1, π_2, π_3) (ϕ_1, ϕ_2, ϕ_3)					
		(0.30,0.30,0.30)	(0.30,0.32,0.35)	(0.30,0.35,0.40)	(0.30,0.37,0.45)	(0.30,0.40,0.50)	(0.30,0.42,0.55)
		(0.10,0.10,0.10)	(0.10,0.11,0.12)	(0.10,0.12,0.14)	(0.10,0.13,0.16)	(0.10,0.14,0.18)	(0.10,0.15,0.20)
10	BBM_1	5.63	7.15	11.84	20.39	32.89	46.33
	BBM_2	8.81	11.13	16.28	20.39	34.35	47.03
15	BBM_1	7.51	10.59	18.72	32.68	49.84	66.18
	BBM_2	9.70	12.70	21.22	33.41	47.42	64.55
20	BBM_1	8.80	13.53	24.75	41.22	61.58	78.86
	BBM_2	9.46	13.55	23.36	40.83	59.90	76.33
30	BBM_1	10.06	16.25	32.83	55.87	77.16	92.19
	BBM_2	10.08	15.23	32.20	54.29	76.48	91.04
40	BBM_1	10.68	18.12	41.87	69.69	89.98	98.05
	BBM_2	9.95	17.84	40.09	69.46	89.96	97.86

5.4 Homogeneity of several Weibull $WB(\psi, \phi)$ populations

5.4.1 Fisher's procedure for combining two score test statistics

Now let $x_{i1}, ..., x_{in_i}$ be a sample from the Weibull distribution $WB(\psi_i, \phi_i), i = 1, ..., k$, with probability density function

$$Pr(X = x) = \left(\frac{\psi}{\phi}\right)\left(\frac{x}{\phi}\right)^{\psi-1} \exp\left[-\left(\frac{x}{\phi}\right)^{\psi}\right].$$

Note that the mean and variance of X are $\phi\Gamma\left(1+\frac{1}{\psi}\right)$ and $\phi^2\left[\Gamma\left(1+\frac{2}{\psi}\right) - \Gamma^2\left(1+\frac{1}{\psi}\right)\right]$ respectively. Thus, testing the equality of means and equality of variances of the $WB(\psi_i, \phi_i)$ populations, $i = 1, ..., k$, is equivalent to testing $\psi_i = \psi$ and $\phi_i = \phi$ for all $i = 1, ..., k$.

Now, from the general results in Section 3.2, we obtain the score test statistic for testing

$H_0' : \psi_i = \psi, \phi_i = \phi$, for all i against H_1' : at least two ψ's are unequal and $\phi_i = \phi$, for all i, as

$$WBS_1 = \sum_{i=1}^{k} \frac{s_{1i}^2}{v_{1i}},$$

where

$$s_{1i} = l_{i1} - l_{i2}l_{i12}/l_{i11},$$

$$v_{1i} = l_{i11} - l_{i12}^2/l_{i22},$$

$$v_{1i} = l_{i11} - l_{i12}^2/l_{i22},$$

with $l_{i1} = \frac{n_i}{\hat{\psi}} + \sum_{j=1}^{n_i} \log\left(\frac{x_{ij}}{\hat{\phi}}\right) - \sum_{j=1}^{n_i}\left(\frac{x_{ij}}{\hat{\phi}}\right)^{\hat{\psi}} \log\left(\frac{x_{ij}}{\hat{\phi}}\right),$

$$l_{i2} = \frac{n_i\hat{\psi}}{\hat{\phi}} + \sum_{j=1}^{n_i} \hat{\psi} x_{ij}^{\hat{\psi}} \hat{\phi}^{-(\hat{\psi}+1)},$$

$$l_{i11} = \frac{n_i[\pi^2/6 + (1-\gamma)^2]}{\hat{\psi}^2},$$
$$l_{i12} = -\frac{n_i(1-\gamma)}{\hat{\phi}},$$
$$l_{i22} = n_i \left(\frac{\hat{\psi}}{\hat{\phi}}\right)^2,$$

where γ is Euler's constant, and $\hat{\psi}$ and $\hat{\phi}$ are the maximum likelihood estimates (m.l.e) of ψ and ϕ under H_0' . The m.l.e of ψ is obtained by solving the maximum likelihood estimating equation

$$\frac{n}{\hat{\psi}} + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log x_{ij} - \frac{n \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} \log x_{ij}}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}} = 0,$$

and $\hat{\phi}$ can be obtained by

$$\hat{\phi} = \left(\frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^{\hat{\psi}}}{n}\right)^{1/\hat{\psi}}.$$

Similarly, the score test statistic for testing

$H_0'' : \phi_i = \phi$, for all i against H_1'' : at least two ϕ''s are unequal, for all i is

$$WBS_2 = \sum_{i=1}^{k} \frac{s_{2i}^2}{v_{2i}},$$

where

$$s_{2i} = \frac{n_i\hat{\psi}_i}{\hat{\phi}} + \sum_{j=1}^{n_i} \hat{\psi}_i x_{ij}^{\hat{\psi}_i} \hat{\phi}^{-(\hat{\psi}_i+1)},$$
$$v_{2i} = n_i \left(\frac{\hat{\psi}_i}{\hat{\phi}}\right)^2,$$

and $\hat{\psi}_i, i = 1, ..., k$ and $\hat{\phi}'$ are the maximum likelihood estimates of $\psi_i, i = 1, ..., k$ and ϕ' under H_0'', obtained by solving the maximum likelihood estimating equations

$$\frac{n_i}{\hat{\psi}} + \sum_{j=1}^{n_i} \log\left(\frac{x_{ij}}{\hat{\phi}}\right) - \sum_{j=1}^{n_i} \left(\frac{x_{ij}}{\hat{\phi}}\right)^{\hat{\psi}} \log\left(\frac{x_{ij}}{\hat{\phi}}\right) = 0,\ i = 1, ..., k,$$

and

$$\frac{\sum_{i=1}^{k} n_i\psi_i}{\phi} + \sum_{i=1}^{k}\sum_{j=1}^{n_i} \psi_i x_{ij}^{\psi_i} \phi^{-(\psi_i+1)} = 0,$$

simultaneously.

Again, from the proofs in Section 3.2, it is obvious that, asymptotically, as $n_i \to \infty, i = 1, ..., k$, the statistics WBS_1 and WBS_2 are independent. We denote the statistic obtained by combining the score test statistics WBS_1 and WBS_2 by WBM_2.

5.4.2 Simulation

In the simulation study we considered K=2, 3 and 4 populations, two nominal levels $\alpha = 0.05$ and $\alpha = 0.10$ and equal sample sizes from each population. Each simulation experiment was based on 10,000 samples.

In the Weibull distribution case, for calculating empirical size, we generated samples from WB (ψ,ϕ) populations with equal ψ's and equal ϕ's. Unequal ψ's and unequal ϕ's were considered for power calculations. Here also Results for k=3 and k=4 are similar. So, we give results for only k=2 and k=3 with $\alpha = 0.05$ and $\alpha = 0.10$ respectively. Results of the simulations for k=2 of $\alpha = 0.05$ and $\alpha = 0.10$ are presented in Table 5.9 and Table 5.10. Those for k=3 are $\alpha = 0.05$ and $\alpha = 0.10$ presented in Table 5.11 and Table 5.12.

According to the results in Table 5.9 to Table 5.12, the statistic WBM_1 is in general liberal, whereas the statistic WBM_2 maintains level well except for small sample sizes ($n \leq 15$), where it shows some conservative behavior; otherwise it holds level well. The power of the statistic WBM_1 is in general larger than that of the statistic WBM_2. This is not surprising as the statistic WBM_1 is in general liberal.

Table 5.10: Empirical power(%) of different statistics for testing homogeneity of K=2 Weibull populations when data are simulated from $WB(\psi_i, \phi_i)$, $i = 1, 2$; based on 10,000 replications; $\alpha = 0.05$

Sample size	Test Statistic	(ψ_1, ψ_2) (ϕ_1, ϕ_2)					
		(1.2,1.2) (3.2,3.2)	(1.2,1.4) (3.2,3.4)	(1.2,1.6) (3.2,3.6)	(1.2,1.8) (3.2,3.8)	(1.2,2.0) (3.2,4.0)	(1.2,2.2) (3.2,4.2)
10	WBM_1	7.35	9.14	13.11	19.2	26.88	35.20
	WBM_2	2.81	4.02	7.23	11.93	18.54	26.66
15	WBM_1	6.75	9.05	15.6	24.75	36.67	49.91
	WBM_2	3.85	5.92	11.73	20.58	31.93	44.88
20	WBM_1	6.65	9.84	17.84	30.79	46.27	61.73
	WBM_2	4.37	7.07	14.91	27.72	43.68	59.92
30	WBM_1	5.92	10.74	24.01	43.03	62.76	78.66
	WBM_2	4.45	9.13	22.13	41.74	62.63	78.77
40	WBM_1	5.82	11.58	29.59	54.39	75.95	89.79
	WBM_2	4.77	10.51	29.03	54.43	76.07	90.06
50	WBM_1	5.82	13.58	35.8	63.84	84.68	95.36
	WBM_2	4.42	12.78	35.74	64.56	85.28	95.71

Table 5.11: Empirical power(%) of different statistics for testing homogeneity of K=2 Weibull populations when data are simulated from $WB(\psi_i, \phi_i)$, $i = 1, 2$; based on 10,000 replications; $\alpha = 0.10$

Sample size	Test Statistic	(ψ_1, ψ_2) (ϕ_1, ϕ_2)					
		(1.2,1.2) (3.2,3.2)	(1.2,1.4) (3.2,3.4)	(1.2,1.6) (3.2,3.6)	(1.2,1.8) (3.2,3.8)	(1.2,2.0) (3.2,4.0)	(1.2,2.2) (3.2,4.2)
10	WBM_1	13.61	15.94	21.35	29.02	38.15	48.02
	WBM_2	7.42	9.57	14.39	21.55	30.42	40.23
15	WBM_1	12.6	16.05	24.77	36.07	49.63	62.38
	WBM_2	8.46	11.88	20.18	31.39	45.17	58.81
20	WBM_1	12.27	16.69	27.67	43.04	59.15	73.32
	WBM_2	9.01	12.92	24.42	40.1	57.15	71.9
30	WBM_1	11.59	18.32	35.14	55.97	73.95	86.95
	WBM_2	8.96	16.09	32.73	54.48	72.90	86.91
40	WBM_1	11.13	19.62	42.03	66.73	84.43	94.39
	WBM_2	9.24	17.64	40.82	66.09	84.56	94.46
50	WBM_1	11.05	22.17	48.09	74.92	91.12	97.8
	WBM_2	9.05	20.78	47.77	74.67	91.36	97.9

Table 5.12: Empirical power(%) of different statistics for testing homogeneity of K=3 Weibull populations when data are simulated from $WB(\psi_i, \phi_i)$, $i = 1, 2, 3$; based on 10,000 replications; $\alpha = 0.05$

Sample size	Test Statistic	(ψ_1, ψ_2, ψ_3) (ϕ_1, ϕ_2, ϕ_3)					
		(0.30,0.30,0.30) (0.10,0.10,0.10)	(0.30,0.32,0.35) (0.10,0.11,0.12)	(0.30,0.35,0.40) (0.10,0.12,0.14)	(0.30,0.37,0.45) (0.10,0.13,0.16)	(0.30,0.40,0.50) (0.10,0.14,0.18)	(0.30,0.42,0.55) (0.10,0.15,0.20)
10	WBM_1	7.87	9.41	12.19	16.8	22.91	30.66
	WBM_2	2.46	3.39	5.52	9.18	14.48	21.31
15	WBM_1	7.42	8.95	13.38	20.35	29.81	40.05
	WBM_2	3.68	4.86	8.51	15.59	24.91	36.07
20	WBM_1	6.67	8.77	14.39	24.30	37.09	51.84
	WBM_2	4.05	5.53	11.24	21.33	34.43	50.24
30	WBM_1	6.26	9.3	18.89	34.41	52.9	70.16
	WBM_2	4.29	7.37	16.62	33.59	53.54	71.58
40	WBM_1	5.97	10.66	23.15	44.2	66.04	83.32
	WBM_2	4.48	9.03	22.58	45.66	67.83	85.23
50	WBM_1	5.70	11.74	28.96	54.58	77.04	91.31
	WBM_2	4.54	10.09	28.88	56.3	79.09	92.73

Table 5.13: Empirical power(%) of different statistics for testing homogeneity of K=3 Weibull populations when data are simulated from $WB(\psi_i, \phi_i)$, $i = 1, 2, 3$; based on 10,000 replications; $\alpha = 0.10$

Sample size	Test Statistic	(ψ_1, ψ_2, ψ_3) (ϕ_1, ϕ_2, ϕ_3)					
		(0.30,0.30,0.30) (0.10,0.10,0.10)	(0.30,0.32,0.35) (0.10,0.11,0.12)	(0.30,0.35,0.40) (0.10,0.12,0.14)	(0.30,0.37,0.45) (0.10,0.13,0.16)	(0.30,0.40,0.50) (0.10,0.14,0.18)	(0.30,0.42,0.55) (0.10,0.15,0.20)
10	WBM_1	14.61	16.13	20.41	26.67	34.57	43.09
	WBM_2	6.86	8.06	11.88	17.36	24.94	33.65
15	WBM_1	13.31	15.81	21.65	31.01	41.35	53.40
	WBM_2	7.67	10.0	16.04	25.17	36.57	49.05
20	WBM_1	12.52	15.66	23.76	35.76	50.17	64.48
	WBM_2	7.99	10.74	19.43	31.73	47.42	63.28
30	WBM_1	11.65	16.55	29.12	46.73	65.48	80.11
	WBM_2	8.46	13.41	26.17	45.6	65.49	80.7
40	WBM_1	11.64	17.68	34.36	56.86	76.78	89.82
	WBM_2	8.90	15.62	33.56	56.7	78.1	91.21
50	WBM_1	11.00	19.39	40.66	66.26	85.42	95.26
	WBM_2	8.94	17.35	39.93	67.31	86.61	95.97

5.5 Discussion and conclusion

Singh (1986) develops a procedure for testing homogeneity of several normal populations based on combining two separate independent likelihood ratio test statistics using a method proposed by Fisher (1950). In Chapter 3, we extended Fisher's method to test homogeneity of several location-scale populations using two likelihood ratio statistics as well as two score test statistics. Asymptotic independence of the two likelihood ratio statistics and also of the two score test statistics have been established. The problem of testing for the homogeneity of several populations, in terms of the means and the variances, arises not only in normal populations but also in other populations. That is why we included two important over-dispersed discrete distributions and also the Weibull distribution in our development of theory and simulation comparison in this chapter.

The statistics based on combining two score tests hold level in all situations investigated here. The statistics based on combining two likelihood ratio statistics hold level in general, although they show either liberal or conservative behavior in some situations, particularly for small sample sizes. We conclude that Fisher's method of combining two statistics, even when they are only asymptotically independent, does perform well for testing homogeneity of several populations in terms of the means and the variances. However, the score test statistics have simple forms, are easy to calculate, because they do not require estimates of the parameters under the alternative hypotheses and have uniformly good level properties. Therefore Fisher's method based on combining two score test statistics might be the method of choice.

Part II

Analysis of Paired Count Data with Zero-Inflation and Over-Dispersion

Chapter 6

Test of Treatment Effect in Pre-drug and Post-drug Count Data with Zero-inflation and Over-dispersion

6.1 Introduction

Data in the form of pre-treatment and post-treatment counts, such as premature heart beats, tumor cells, epileptic seizures, etc., arise in numerous applications. The purpose of this chapter is to present a procedure for testing no treatment effect in these data sets. As an example we consider the data given in Data sets, Table D.3. The data on premature ventricular contractions (PVC) originally given as counts by Berry (1987) are analysed by Farewell and Sprott (1988) as proportions. The data pertain to twelve patients who experienced frequent premature ventricular contractions (PVCs) and were administered a drug with antiarrhythmic properties. One-minute EKG recordings were taken before and after drug administration. The PVCs were counted on both recordings. The observations occur as paired data (x_i, y_i), which are the pre-drug and post-drug count, respectively, for the ith patient. Assume that x_i is a Poisson variate with mean λ_i and that for patients who are not cured y_i is independently Poisson with mean $\beta\lambda_i$. In order to eliminate the "incidental" nuisance parameters λ_i, one for each uncured subject, Farewell and Sprott (1988) use the conditional distribution of y_i given $m_i = x_i + y_i$, which is

$$f(y_i; p|m_i) = \begin{pmatrix} m_i \\ y_i \end{pmatrix} p^{y_i}(1-p)^{m_i - y_i}, y_i = 0, 1, ..., m_i,$$

where $p = \beta\lambda_i/(\lambda_i + \beta\lambda_i) = \beta/(1+\beta)$. A binomial model may fail to fit a set of data in the form of proportions either because of the presence of zero-inflation or because of the presence of over-dispersion. Let ω be the probability of cure implying that $y_i = 0$. Then, the distribution of y_i, conditional on m_i can be written as a mixture model (the zero-inflated binomial model).

$$Pr(y_i|m_i) = \begin{cases} \omega + (1-\omega)f(0; p|m_i) & \text{if } y_i = 0 \\ (1-\omega)f(y_i; p|m_i) & \text{if } y_i > 0. \end{cases} \tag{6.1.1}$$

Using a score test based on this model Deng and Paul (2000) find significant zero inflation in the PVC data. An over-dispersed model such as the beta-binomial model with probability parameter π and dispersion parameter ϕ having probability function

$$f(y_i; \pi, \phi|m_i) = \begin{pmatrix} m_i \\ y_i \end{pmatrix} \frac{\prod_{r=0}^{y_i-1}(\pi(1-\phi)+r\phi)\prod_{r=0}^{m_i-y_i-1}((1-\pi)(1-\phi)+r\phi)}{\prod_{r=0}^{m_i-1}(1-\phi+r\phi)}$$

may fit the data as well or better than the zero-inflated binomial model. Again, Deng and Paul (2000) use a score test developed by Dean (1992) to show that there is significant over-dispersion in the PVC data. They, in fact, fitted the binomial, the zero-inflated binomial and the beta-binomial model to the PVC data and concluded that among these models the zero-inflated binomial model fits the data best. However, they argued that the PVC data and other similar data may contain both zero-inflation and over-dispersion. As a zero-inflated over-dispersed model one can consider the zero-inflated beta-binomial model which is given as

$$Pr(y_i|m_i) = \begin{cases} \omega + (1-\omega)f(0;\pi,\phi|m_i) & \text{if } y_i = 0 \\ (1-\omega)f(y_i;\pi,\phi|m_i) & \text{if } y_i > 0. \end{cases} \tag{6.1.2}$$

The parameter ω is the zero-inflation parameter and the parameter ϕ is the intraclass correlation parameter. The zero-inflation parameter can take negative values provided $-\frac{f(0;\pi,\phi|m_i)}{1-f(0;\pi,\phi|m_i)} \leq \omega < 1$. Note that if $\omega > 0$, then $P(Y=0) > f(0;\pi,\phi|m_i)$ and if $\omega < 0$, then $P(Y=0) < f(0;\pi,\phi|m_i)$. While the former indicates existence of too many zeros (zero inflation), the latter indicates that there exist too few zeros (zero deflation) in the data. Further, the intraclass parameter ϕ also may assume positive as well as negative values provided $max(\frac{-1}{m_i-1}) < \phi < 1$ (Prentice, 1986). In the limit as $\phi \to 0$ the zero-inflated beta-binomial model converges to the zero-inflated binomial model.

Thus, the zero-inflated beta-binomial model is the most flexible model for the analysis of data similar to the PVC data. Farewell and Sprott (1988) alluded to such a model.

In this chapter we use this model to develop procedures for testing for treatment effect. As one can see that treatment can affect two parameters, namely, the zero-inflation parameter ω and the parameter π. Note that the parameter ω represents the proportion of cure and the parameter π represents the effect of the treatment on the uncured population. Therefore to determine treatment effect one can (i) estimate ω, the proportion of cure and test whether the uncured population had any improvement of their prevailing condition as a result of the treatment or (ii) test the overall effect of the treatment. Note that $\omega = 0$ indicates that the treatment fails to cure the disease while $\pi = 1/2$ indicates that the treatment had no effect on the uncured population. Therefore, we develop tests: (i) of $H_0 : \pi = 1/2$ against $H_1 : \pi \neq 1/2$ treating ω and ϕ as nuisance parameters and (ii) of $H_0' : \pi = 1/2$, $\omega = 0$ against $H_1' : \pi \neq 1/2$ or $\omega \neq 0$ treating ϕ as a nuisance parameter. In particular we develop score tests and likelihood ratio tests.

The score tests and the likelihood ratio tests are developed in Section 6.2. Some simulations are carried out in Section 6.3 to study level and power properties of the score and the likelihood ratio tests. In Section 6.4 we analyse the PVC data. A discussion is given in Section 6.5.

6.2 Test for no treatment effect

6.2.1 The maximum likelihood estimates

We now give maximum likelihood estimates of the parameters under different hypotheses as these will be used in the score and the likelihood ratio statistics. Let $y_i, i = 1, ..., n$, be a sample of independent observations from the zero-inflated beta-binomial model (6.1.2). Then, the log-likelihood can be written as

$$l(\gamma, \pi, \phi; y) = \sum_{i=1}^{n} l_i(\gamma, \pi, \phi; y_i)$$

$$= \sum_{i=1}^{n}\{-\log(1+\gamma) + I_{\{y_i=0\}}\log(\gamma + f_{0i}) + I_{\{y_i>0\}}\log f_{y_i}\},$$

where $f_{y_i} = \begin{pmatrix} m_i \\ y_i \end{pmatrix} \frac{\prod_{r=0}^{y_i-1}(\pi(1-\phi)+r\phi)\prod_{r=0}^{m_i-y_i-1}((1-\pi)(1-\phi)+r\phi)}{\prod_{r=0}^{m_i-1}(1-\phi+r\phi)}$, $\gamma = \frac{\omega}{1-\omega}$ and $f_{0i} = \frac{\prod_{r=0}^{m_i-1}((1-\pi)(1-\phi)+r\phi)}{\prod_{r=0}^{m_i-1}(1-\phi+r\phi)}$. Note that for convenience we have reparameterized ω into γ. Thus $\omega = 0$ implies $\gamma = 0$.

Now, let $l_{0i} = \log(f_{0i})$, $l_{y_i} = \log(f_{y_i})$. Further, let $l'_{0i(\pi)} = \frac{\partial l_{0i}}{\partial \pi}$, $l'_{0i(\phi)} = \frac{\partial l_{0i}}{\partial \phi}$, $l'_{y_i(\pi)} = \frac{\partial l_{y_i}}{\partial \pi}$ and $l'_{y_i(\phi)} = \frac{\partial l_{y_i}}{\partial \phi}$. Explicit expressions for these terms are given in the Appendix B. Then the maximum likelihood estimates of the parameters γ, π and ϕ are obtained by solving the estimating equations

$$\sum_{i=1}^{n}\{I_{\{y_i=0\}}\frac{f_{0i}l'_{0i(\pi)}}{\gamma + f_{0i}} + I_{\{y_i>0\}}l'_{y_i(\pi)}\} = 0,$$

$$\sum_{i=1}^{n}\{I_{\{y_i=0\}}\frac{f_{0i}l'_{0i(\phi)}}{\gamma + f_{0i}} + I_{\{y_i>0\}}l'_{y_i(\phi)}\} = 0$$

and

$$\sum_{i=1}^{n}\{I_{\{y_i=0\}}\frac{1+\gamma}{\gamma + f_{0i}} - 1\} = 0,$$

simultaneously. These are the estimates under the general alternative in which none of the parameters are specified. We denote these by $\tilde{\gamma}$, $\tilde{\pi}$ and $\tilde{\phi}$. Further, under the null hypothesis $H_0 : \pi = 1/2$, the maximum likelihood estimates of γ and ϕ are obtained by solving the estimating equations

$$\sum_{i=1}^{n}\{I_{\{y_i=0\}}\frac{f_{0i}l'_{0i(\phi)}}{\gamma + f_{0i}} + I_{\{y_i>0\}}l'_{y_i(\phi)}\}|_{\pi=1/2} = 0$$

and

$$\sum_{i=1}^{n}\{I_{\{y_i=0\}}\frac{1+\gamma}{\gamma + f_{0i}} - 1\}|_{\pi=1/2} = 0$$

simultaneously and the maximum likelihood estimator $\hat{\phi}'$ of the nuisance parameter ϕ under the null hypothesis $H_0 : \pi = 1/2, \gamma = 0$ is obtained by solving the estimating equation

$$\sum_{i=1}^{n}\{I_{\{y_i=0\}}l'_{0i(\phi)} + I_{\{y_i>0\}}l'_{y_i(\phi)}\}|_{\pi=1/2} = 0.$$

6.2.2 The score tests

The score test (Rao, 1947) is a special case of the more general $C(\alpha)$ test (Neyman, 1966) in which the nuisance parameters are replaced by maximum likelihood estimates which are $\sqrt{N}$ (N=number of observations used in estimating the parameters) consistent estimates. The score test is particularly appealing as it often maintains, at least approximately, a preassigned level of significance (see Bartoo and Puri, 1967). Further, it requires estimates of the parameters only under the null hypothesis, and often produces a statistic which is simple to calculate. For more discussion on the choice of $C(\alpha)$ or score tests see Barnwal and Paul (1988), Breslow (1990), and Paul and Banerjee (1998).

We want to obtain score tests for $H_0 : \pi = 1/2$ against $H_1 : \pi \neq 1/2$ when γ and ϕ are treated as nuisance parameters and for $H_0' : \pi = 1/2$, $\gamma = 0$ against $H_1' : \pi \neq 1/2$ or $\gamma \neq 0$ when ϕ is treated as a nuisance parameter.

Derivation of the score tests are quite involved. So, here we give the results relegating the proof to the Appendix B. The score test statistic for testing $H_0 : \pi = 1/2$ against $H_1 : \pi \neq 1/2$ is

$$S_1 = \hat{\Psi}_1^2/\hat{V}_1^2,$$

where $\hat{\Psi}_1 = \Psi_1(\hat{\phi}, \hat{\gamma})|_{\pi=1/2}$ and $\hat{V}_1 = V_1(\hat{\phi}, \hat{\gamma})|_{\pi=1/2}$

with

$$\Psi_1 = \frac{\partial l}{\partial \pi} = \sum_{i=1}^n \{I_{\{y_i=0\}} \frac{f_{0i} l'_{0i(\pi)}}{\gamma + f_{0i}} + I_{\{y_i>0\}} l'_{y_i(\pi)}\}$$

and

$$V_1^2 = I_{\pi\pi} - \frac{I_{\phi\phi} I_{\pi\gamma}^2 + I_{\gamma\gamma} I_{\pi\phi}^2 - 2 I_{\pi\gamma} I_{\pi\phi} I_{\gamma\phi}}{I_{\phi\phi} I_{\gamma\gamma} - I_{\phi\gamma}^2},$$

where $I_{\pi\pi} = \{\sum_{i=1}^n \frac{(\gamma + f_{0i}) f_{0i} l''_{0i(\pi\pi)} - \gamma f_{0i} (l'_{0i(\pi)})^2}{(1+\gamma)(\gamma + f_{0i})} + l''_{\pi\pi}\}$, $I_{\pi\gamma} = \frac{1}{1+\gamma} \sum_{i=1}^n \{\frac{f_{0i} l'_{0i(\pi)}}{\gamma + f_{0i}}\}$, $I_{\pi\phi} = \{\sum_{i=1}^n \frac{(\gamma + f_{0i}) f_{0i} l''_{0i(\pi\phi)} - \gamma f_{0i} l'_{0i(\pi)} l'_{0i(\phi)}}{(1+\gamma)(\gamma + f_{0i})} + l''_{\pi\phi}\}$, $I_{\phi\gamma} = \frac{1}{1+\gamma} \sum_{i=1}^n \{\frac{f_{0i} l'_{0i(\phi)}}{\gamma + f_{0i}}\}$, $I_{\phi\phi} = \{\sum_{i=1}^n \frac{(\gamma + f_{0i}) f_{0i} l''_{0i(\phi\phi)} - \gamma f_{0i} (l'_{0i(\phi)})^2}{(1+\gamma)(\gamma + f_{0i})} + l''_{\phi\phi}\}$, $I_{\gamma\gamma} = \sum_{i=1}^n \{-\frac{1}{(1+\gamma)^2} + \frac{1}{(1+\gamma)(\gamma + f_{0i})}\}$. The quantities $l''_{0i(\pi\pi)}$, $l''_{0i(\phi\phi)}$, $l''_{0i(\pi\phi)}$, $l''_{\pi\pi}$, $l''_{\pi\phi}$, $l''_{\phi\phi}$ used above are given in the Appendix B.

The statistic S_1, asymptotically, as $n \to \infty$, has a $\chi^2(1)$ distribution. Note that in this score test we use the maximum likelihood estimates of γ and ϕ under the null hypothesis $H_0 : \pi = 1/2$.

The score test for $H'_0 : \pi = 1/2, \gamma = 0$ against $H'_1 : \pi \neq 1/2$ or $\gamma \neq 0$ is

$$S_2 = \hat{\Psi}'_2 \hat{V}_2^{-1} \hat{\Psi}_2,$$

with $\hat{\Psi}_2 = \Psi_2(\hat{\phi}')$ and $\hat{V}_2 = V_2(\hat{\phi}')$, where $\Psi_2 = \left(\frac{\partial l}{\partial \pi}, \frac{\partial l}{\partial \gamma}\right)' \Big|_{H'_0}$,

with

$$\frac{\partial l}{\partial \pi}\Big|_{H'_0} = \sum_{i=1}^n \{I_{\{y_i=0\}} l'_{0i(\pi)} + I_{\{y_i>0\}} l'_{y_i(\pi)}\}|_{\pi=1/2}$$

and

$$\frac{\partial l}{\partial \gamma}\Big|_{H'_0} = -n + \sum_{i=1}^n \frac{I_{\{y_i=0\}}}{f_{0i}}\Big|_{\pi=1/2}$$

and

$$V_2 = I_{11} - I_{12}I_{22}^{-1}I_{21}, \text{ with } I_{11} = \begin{pmatrix} J_{\pi\pi} & J_{\pi\gamma} \\ J_{\pi\gamma} & J_{\gamma\gamma} \end{pmatrix}, I_{21} = I_{12}' = (J_{\phi\pi}, J_{\phi\gamma}), I_{22} = J_{\phi\phi}.$$

The quantities $J_{\pi\pi}$, $J_{\pi\gamma}$, $J_{\gamma\gamma}$, $J_{\phi\pi}$, $J_{\phi\gamma}$ and $J_{\phi\phi}$ respectively are $I_{\pi\pi}$, $I_{\pi\gamma}$, $I_{\gamma\gamma}$, $I_{\phi\pi}$, $I_{\phi\gamma}$ and $I_{\phi\phi}$ defined earlier by replacing π by 1/2, γ by 0.

The statistic S_2, asymptotically, as $n \to \infty$, has a $\chi^2(2)$ distribution. Note that in this score test we use the maximum likelihood estimate of ϕ under the null hypothesis $H_0' : \pi = 1/2, \gamma = 0$.

6.2.3 The log-likelihood ratio tests

The likelihood ratio statistic for testing $H_0 : \pi = 1/2$ against $H_1 : \pi \neq 1/2$ is

LR1= $2(\ \hat{l}(\tilde{\gamma}, \tilde{\pi}, \tilde{\phi}; y)$-$\hat{l}(\hat{\gamma}, .5, \hat{\phi}; y))$

and that for testing $H_0' : \pi = 1/2, \gamma = 0$ against $H_1' : \pi \neq 1/2$ or $\gamma \neq 0$ is

LR2= $2(\ \hat{l}(\tilde{\gamma}, \tilde{\pi}, \tilde{\phi}; y)$-$\hat{l}(0, .5, \hat{\phi}'; y))$,

where $\hat{l}(0, .5, \hat{\phi}'; y))$ is the maximized log-likelihood under the null hypothesis $H_0' : \pi = 1/2, \gamma = 0$, $\hat{l}(\hat{\omega}, .5, \hat{\phi}; y))$ is the maximized log-likelihood under the null hypothesis $H_1 : \pi = 1/2$ and $\hat{l}(\tilde{\omega}, \tilde{\pi}, \tilde{\phi}; y)$ is the maximized log-likelihood under the alternative hypothesis $H_1 : \pi \neq 1/2$ or $H_1' : \pi \neq 1/2$ or $\gamma \neq 0$.

Asymptotically, as $n \to \infty$, the distribution of $LR1$ is $\chi^2(1)$ and that of $LR2$ is $\chi^2(2)$.

6.3 Simulation

We now report results of a simulation study conducted to examine the empirical size and power of the score test statistic S_1 and the likelihood ratio statistic $LR1$. Samples of size n=12 and n=24 were considered. For n=12, the sample size configuration m_i, i= 1,..., 12, considered were the total PVC counts 11, 11, 17, 22, 9, 6, 5, 14, 9, 7, 22, 51 in the data in Table D.3. For n=24 the sample size configuration considered were 11, 11, 17, 22, 9, 6, 5, 14, 9, 7, 22, 51, 11, 11, 17, 22, 9, 6, 5, 14, 9, 7, 22, 51. That is, we just doubled the data considered for $n = 12$. Empirical size and power of the test statistics S_1 and $LR1$ were calculated using data from the zero-inflated beta-binomial distribution with $\pi = .2, .4, .46, .50, .54, .6, .8$, $\omega = .05, .10, .20$ and $\phi = .1, .2$. Each simulation experiment was based on 10,000 simulations. Empirical size and power results of the test statistics S_1 and $LR1$ with $\phi = 0.10$ are presented in Table 6.1 and those with $\phi = 0.20$ are given in Table 6.2. Note, the entries in column 8 with $\pi = .5$ of each of Table 6.1 and Table 6.2 represent empirical levels.

Both the statistics S_1 and $LR1$ hold level well and they both show excellent power property. In the important range $\pi < .5$ power of the score test statistic S_1 is slightly better than the likelihood ratio statistic $LR1$ and for $\pi > .5$ power of the statistic $LR1$ is slightly better than the statistic S_1. Note that $\pi < .5$ indicates positive treatment effect, whereas $\pi > .5$ indicates negative treatment effect. Sample size also seems to have an effect on power. For example, power with $n = 24$ is larger than that with $n = 12$. Power also seems to be a decreasing function of ϕ. For example, powers of S_1 and $LR1$ with $\alpha = 0.05$, $\omega = .2$, $\pi = .6$ and $\phi = .1$ are .182 and .246 respectively. With the same values of α, ω, π and with $\phi = .2$, powers of S_1 and $LR1$ are only

.1191 and .1798 respectively. The power would be largest when ϕ is smallest, that is, when we have a zero-inflated binomial model.

Our very limited simulation study (the results are not given here) revealed similar properties of the statistics S_2 and $LR2$ as those of the statistics S_1 and $LR1$. The statistics S_2 and $LR2$, however, showed some conservative behavior.

Either the score tests or the likelihood ratio tests can be used for testing the presence of treatment effect. The score tests, however, may be preferable because they use estimates of the parameters only under the null hypothesis and in the important range $\pi < .5$, the power of the score test statistic S_1 is slightly better than the likelihood ratio statistic $LR1$.

6.4 Analysis of the PVC data

We now test for treatment effect in the PVC data. For this we fit three models to the data, namely the beta-binomial model with $\pi = 1/2$ and unknown parameter ϕ, the zero-inflated beta-binomial model with $\pi = 1/2$ and unknown parameters γ and ϕ, and the zero-inflated beta-binomial model with unknown parameters π, γ and ϕ. For the PVC data we obtain $\hat{\phi}' = 0.71$, $\hat{l}(0, .5, \hat{\phi}'; y) = -25.275$, $\hat{\gamma} = 1.35$, $\hat{\phi} = 0.122$, $\hat{l}(\hat{\gamma}, .5, \hat{\phi}; y) = -19.462$ and $\tilde{\gamma} = 1.262$ $\tilde{\pi} = 0.336$, $\tilde{\phi} = 0.084$, $\hat{l}(\tilde{\gamma}, \tilde{\pi}, \tilde{\phi}; y) = -18.03$. From these maximized log-likelihoods we obtain $LR1 = 2.873$ and $LR2 = 11.62$. Further, the values of the score test statistics S_1 and S_2 are 2.59 and 8.358 respectively.

To test whether the uncured population had any improvement of their prevailing condition as a result of the treatment, the p-values of the LR test and the score test are 0.09 and 0.108 respectively. The conclusion from the likelihood ratio test is essentially the same as that from the score test. The tests show some evidence of the effect of the treatment, though not highly significant. That is, treatment has improved the prevailing condition of the uncured population. Also note that treatment has resulted in a significant proportion of cure ($\hat{\omega}$=0.575). To test the overall effect of the treatment, the p-values of the LR test and the score test are 0.003 and 0.015 respectively. Both tests show a highly significant treatment effect for the whole population.

6.5 Discussion

Berry (1987) used a paired t-test after logarithmically transforming the pre-drug and the post-drug counts. For example, the y-data were transformed to $z = \log(y + c)$, where c is to be determined so that a function $g_0(c)$ given in equation (5) of Berry (1987) is minimum with respect to c. He concluded that there is a significant treatment effect (p-value=.001). Note that Berry's method uses the pre-drug (x) and the post-drug (y) counts. As such, his test based on these data is an unconditional test. Also, his method cannot estimate the zero-inflation and the over-dispersion parameters. Our method, based on the zero-inflated beta- binomial model, is a conditional (conditional on $x + y = m$) approach. Our method enables us not only to test for over-all treatment effect, but also to test for effect of the treatment on the uncured population. In addition, our model facilitates estimation of the proportion of cure and the amount of over-dispersion present in the data. Note that the test of the hypothesis $H_0' : \pi = 1/2, \gamma = 0$ brings out the same conclusion regarding the treatment effect (p-values being 0.003 and 0.015 based the likelihood ratio test and the score test respectively) as that of Berry (the p-value for his test statistic being .001). Our recommendation, however, is not to ignore Berry's method. We agree with his statement "Researchers should learn as much as possible from their data. This includes looking at the data in various ways" (see discussion in Berry, 1987).

Table 6.1: Empirical power(%) of the score test statistic S_1 and the likelihood ratio statistic $LR1$ for testing no treatment effect when data are simulated from the zero-inflated beta-binomial distribution with $\phi = 0.10$ and different values of π and ω. The column under $\pi = .5$ represents the empirical level of the statistics S_1 and $LR1$. Empirical level and power results are based on 10,000 simulations.

Sample size	α	ω	Test statistic	π						
				.20	.40	.46	.50	.54	.60	. 80
12	.05	.05	S_1	98.04	29.32	8.58	4.42	7.60	28.41	97.20
			LR_1	96.83	28.11	7.84	4.86	9.19	31.97	99.12
		.10	S_1	97.25	29.10	8.37	4.58	6.87	23.90	91.52
			LR_1	96.21	26.71	8.37	5.23	9.17	29.97	97.38
		.20	S_1	94.74	26.56	8.87	4.58	6.08	18.21	76.36
			LR_1	93.51	25.83	8.38	4.96	7.46	24.55	92.53
	.10	.05	S_1	99.28	43.52	17.03	10.11	15.98	44.24	99.34
			LR_1	98.76	40.27	14.91	9.49	16.74	45.88	99.70
		.10	S_1	99.04	43.41	15.94	9.78	14.34	39.58	97.79
			LR_1	98.40	39.47	14.74	10.38	16.24	43.34	99.15
		.20	S_1	97.98	41.07	16.86	10.84	13.60	34.00	92.50
			LR_1	97.01	37.61	14.96	10.02	14.39	37.00	96.94
24	.05	.05	S_1	100.0	55.33	13.64	4.69	12.81	53.94	99.98
			LR_1	99.99	54.25	12.69	5.47	15.39	59.51	99.99
		.10	S_1	99.97	55.16	13.54	4.81	11.51	48.65	99.93
			LR_1	99.97	53.74	13.24	5.46	13.83	54.32	100.0
		.20	S_1	99.95	50.39	13.24	5.04	9.72	42.83	99.35
			LR_1	99.92	50.90	12.74	5.70	10.98	49.42	99.93
	.10	.05	S_1	100.0	68.52	22.95	10.10	22.02	68.03	100.0
			LR_1	99.99	67.28	21.31	10.65	24.74	71.69	100.0
		.10	S_1	99.99	68.25	22.64	10.16	20.30	64.65	100.0
			LR_1	100.0	66.31	22.07	10.96	23.01	67.05	100.0
		.20	S_1	99.97	64.31	21.87	10.64	18.53	58.80	99.80
			LR_1	99.98	63.61	21.38	10.67	18.29	62.37	100.0

Table 6.2: Empirical power(%) of the score test statistic S_1 and the likelihood ratio statistic $LR1$ for testing no treatment effect when data are simulated from the zero-inflated beta-binomial distribution with $\phi = 0.20$ and different values of π and ω. The column under $\pi = .5$ represents the empirical level of the statistics S_1 and $LR1$. Empirical level and power results are based on 10,000 simulations.

Sample size	α	ω	Test statistic	π						
				.20	.40	.46	.50	.54	.60	. 80
12	.05	.05	S_1	86.68	21.27	7.78	5.29	8.69	23.66	94.89
			LR_1	79.18	17.17	6.65	6.09	10.09	25.84	95.76
		.10	S_1	84.21	20.9	7.96	5.18	6.87	18.01	82.92
			LR_1	76.84	17.91	7.25	6.02	8.95	22.67	89.04
		.20	S_1	79.49	20.72	8.46	5.20	5.27	11.91	57.10
			LR_1	72.24	16.91	7.16	5.80	8.05	17.98	73.72
	.10	.05	S_1	92.74	32.58	14.58	11.65	16.43	36.50	97.75
			LR_1	87.54	25.91	12.35	11.37	17.59	37.92	98.21
		.10	S_1	90.98	31.83	14.8	10.97	14.19	30.55	91.78
			LR_1	85.86	26.83	12.74	11.82	16.20	33.72	94.13
		.20	S_1	88.18	32.11	15.63	11.08	13.21	24.92	77.56
			LR_1	82.10	26.62	13.47	11.61	14.74	28.45	84.80
24	.05	.05	S_1	99.16	39.32	11.00	5.67	10.42	38.48	99.69
			LR_1	98.22	31.24	8.46	5.37	12.32	40.61	99.89
		.10	S_1	98.53	37.78	11.06	5.11	8.41	32.27	98.40
			LR_1	97.94	32.91	8.80	5.52	10.63	35.12	98.68
		.20	S_1	97.62	34.52	10.39	5.32	7.63	26.37	92.97
			LR_1	96.41	31.80	10.15	5.84	9.45	31.17	96.73
	.10	.05	S_1	99.67	51.62	18.15	10.96	18.88	52.79	99.92
			LR_1	99.25	43.48	14.65	10.34	20.25	53.79	99.97
		.10	S_1	99.46	50.06	18.51	10.65	16.63	46.46	99.44
			LR_1	98.99	44.88	15.23	10.73	17.95	48.43	99.55
		.20	S_1	98.84	46.76	17.46	10.90	15.07	40.72	97.32
			LR_1	98.29	43.91	17.14	11.19	16.72	43.86	98.76

Chapter 7

Treatment Effect of DMFT Data Based on Zero-inflated Bivariate Poisson Regression Model

7.1 Introduction

In biomedical and dental epidemiological experiments data arise in the form of pre-treatment and post-treatment counts. For example, Böhning, Dietz, Schlattmann, Mendonca and Kirchner (1999) present dental epidemiology data of a prospective study of caries prevention of school-children from an urban area of Belo Horizonte (Brazil). The children were all 7 years of age at the beginning of the study. Dental status was measured by the decayed, missing and filled teeth (DMFT) index. Only the eight deciduous molars were considered, which implies that the smallest possible value of the DMFT index is 0 and the largest is 8. The prospective study was for a period of two years. The aim of the caries prevention study was to compare four methods, namely, oral health education, enrichment of the school diet with rice bran, mouthwash with 0.2% sodium fluoride solution and oral hygiene. Six schools took part in the study. Interventions were carried out according to the following scheme: School 1, oral health education; School 2, all four methods together; School 3, the control group; School 4, enrichment of the School diet with rice bran; School 5, mouthwash with 0.2% sodium fluoride solution; School 6, oral hygiene. The six treatments were randomized to the six schools, so that all children of a given school received the same treatment. 797 school children were examined both before and after the trial, their dental status evaluated and the DMFT index computed. The DMFT index data for the six treatments (schools), denoted as DMFT1, at the beginning of the study and those, denoted by DMFT2, at the end of the study are given in Böhning et al. (1999). Also given in Böhning et al. (1999) are information regarding the covariates Gender (Female, Male) and Ethnic group (Dark, White, Black).

To study treatment effects Böhning et al. (1999) use a zero-inflated Poisson regression model (ZIPR) of the DMFT2 data with School (School 1 to 6), Ethnic group (Dark, White, Black), Gender (Female, Male) and $\log(DMFT1 + 0.5)$ as covariates. In this chapter, we use a bivariate zero-inflated Poisson regression model (ZIBPR) for the paired data (DMFT1, DMFT2), with School, Ethnic group and Gender as covariates. The main difference between their modeling approach and ours is that they use $\log(DMFT1 + 0.5)$ in the ZIPR model as a covariate, whereas, we jointly model DMFT1 and DMFT2. We develop an EM-algorithm (Dempster et al., 1977) to obtain the maximum likelihood estimates of the parameters of the ZIBPR model. Further, we obtain the exact Fisher information matrix of the parameters of the ZIBPR model and develop a procedure for testing treatment effects. A model selection procedure is given to decide on an appropriate model. For the DMFT index data, based on the model selected, we arrival at a ranking of the treatment effects which coincides with that from a simple analysis of treatment effects.

In Section 7.2, we introduce the zero-inflated Poisson regression model (ZIPR) and zero-inflated bivariate Poisson regression model (ZIBPR). An EM-algorithm for obtaining the maximum likelihood estimates is developed in Section 7.3. In Section 7.4, we obtain the exact Fisher information matrix for model ZIBPR, which is given in Appendix C, and develop the procedure for testing treatment effects for the DMFT index data. Analysis of the DMFT index data and a comparision of the analysis by Böhning et al. (1999) are given in Section 7.5.

7.2 The zero-inflated Poisson and bivariate Poisson regression models

Let y represent the DMFT2 count. A commonly used model for y is the Poisson model

$$f(y, \lambda) = \exp(-\lambda)\lambda^y/y!. \tag{7.2.1}$$

In practice, however, a Poisson model may not fit count data of the type DMFT2, because of the presence of more zeros in the data than what can be expected under a Poisson model. A model that takes account of the extra zeros in the data is the zero-inflated Poisson model.

$$f_1(y, \lambda, \omega) = \begin{cases} \omega + (1-\omega)f(0, \lambda), & \text{If } y = 0, \\ (1-\omega)f(y, \lambda), & \text{If } y > 0, \end{cases} \tag{7.2.2}$$

where ω is the zero-inflation parameter. This model can be generalized by including covariates into the model. Note, our purpose is to test for the effects of the treatments after accounting for covariates including the base-line DMFT index. Suppose there are k treatments and p covariates, including the DMFT1 counts. Now, let x_1 be a $k \times 1$ vector of covariates representing the treatments and $\beta_1 = (\beta_{11}, \beta_{12}, ..., \beta_{1k})'$ be the corresponding $k \times 1$ regression parameters. Further, let x_2 be the $p \times 1$ vector of other covariates, such as Gender, Ethnic group, $\log(DMFT1 + 0.5)$ etc., and $\beta_2 = (\beta_{21}, \beta_{22}, ..., \beta_{2p})'$ be the corresponding $p \times 1$ vector of regression parameters. Then, model (7.2.2) can be written as

$$f_1(y, \lambda, \omega) = \begin{cases} \omega + (1-\omega) f(0, \lambda), & \text{if } y = 0, \\ (1-\omega) f(y, \lambda), & \text{if } y > 0, \end{cases} \tag{7.2.3}$$

with $\log \lambda = x_1'\beta_1 + x_2'\beta_2$. We denote this model as zero-inflated Poisson regression model (ZIPR). Note that the ZIPR model (7.2.3) is equivalent to the ZIPR model by Böhning et al. (1999) in which they introduce an intercept term. In our ZIPR model (7.2.3), $\beta_{11}, \beta_{12}, \ldots, \beta_{1k}$ are the effects of the k treatments. Then, testing for no effect of the jth treatment is equivalent to testing $H_0 : \beta_{1i} = 0$, i=1,..., k.

However, note that the data (DMFT1, DMFT2) are paired count data as these are obtained before and after application of a treatment. It may then be more appropriate to consider a bivariate zero-inflated Poisson model for the paired data (DMFT1, DMFT2). Denote (Y_1, Y_2) as the paired data (DMFT1, DMFT2). Then, the bivariate Poisson model for (Y_1, Y_2) (see Holgate, 1964; Irwin, 1963; Paul and Ho, 1989; Kocherlakota and Kocherlakota, 1992, and Karilis and Ntzoufras, 1998) can be written as

$$f_2(y_1, y_2 | \lambda_0, \lambda_1, \lambda_2) = \exp(-\lambda_1 - \lambda_2 - \lambda_0) \sum_{i=0}^{\min\{y_1, y_2\}} \frac{\lambda_1^{y_1 - i} \lambda_2^{y_2 - i} \lambda_0^i}{(y_1 - i)!(y_2 - i)!i!}, \tag{7.2.4}$$

where $E(Y_1) = \lambda_1 + \lambda_0$, $E(Y_2) = \lambda_2 + \lambda_0$ and $Cov(Y_1, Y_2) = \lambda_0$. Here we use a log-linear model for the ratio of the two mean parameters λ_1 and λ_2

$$\log(\lambda_2/\lambda_1) = x_1'\gamma_1 + \tilde{x}_2'\gamma_2, \qquad (7.2.5)$$

where, x_1 is a $k \times 1$ vector of covariates representing the treatments and γ_1 is the corresponding $k \times 1$ regression parameters, $\tilde{x}_2$ is the $(p-1) \times 1$ vector of covariates, such as, gender, ethnic group etc., and γ_2 is the corresponding $(p-1) \times 1$ vector of regression parameters. Note $\lambda_2 = \lambda_1 \exp(x_1'\gamma_1 + \tilde{x}_2'\gamma_2)$, so that λ_1 can be considered to be a parameter corresponding to the base-line counts DMFT1.

As in the Poisson regression model, a bivariate Poisson regression model may not fit paired count data of the type (DMFT1, DMFT2) because of the presence of more paired zeros in the data than can be expected under a bivariate Poisson regression model. A model that takes account of the extra zeros in the data is the zero-inflated bivariate Poisson regression model (ZIBPR). Let θ be the proportion of pairs of observations (y_1, y_2) having extra zeros. Then a bivariate zero-inflated Poisson regression model can be written as

$$f_3(y_1, y_2 |\, \theta, \lambda_0, \lambda_1, \lambda_2) = \begin{cases} \theta + (1-\theta) f_2(0, 0 |\, \lambda_0, \lambda_1, \lambda_2), & \text{if } (y_1, y_2) = (0, 0), \\ (1-\theta) f_2(y_1, y_2 |\, \lambda_0, \lambda_1, \lambda_2), & \text{if } y_1 > 0, y_2 > 0, \end{cases} \qquad (7.2.6)$$

with $\lambda_2 = \lambda_1 \exp(x_1'\gamma_1 + \tilde{x}_2'\gamma_2)$. Note that this model can be further generalized by introducing two additional zero-inflation parameters: one when zero inflation occurs for y_1 and not for y_2 and the other when zero inflation occurs for y_2 and not for y_1. To avoid complications we do not consider such a model. Note that under model (7.2.6) testing for no effect of the ith treatment is equivalent to testing $H_0 : \gamma_{1i} = 0$,

$i = 1, ..., k$.

7.3 Estimation of the parameters of the zero-inflated bivariate Poisson regression models.

Dempster et al. (1977) interpreted mixture data as incomplete data by regarding an observation on the mixture model as missing its component. The zero-inflated bivariate Poisson model can be interpreted as a mixture of a bivariate Poisson distribution $f_2(y_1, y_2 | \lambda_0, \lambda_1, \lambda_2)$ and a distribution with a point mass of one at (0, 0) with mixing probability θ.

Let (y_{1ij}, y_{2ij}) denote the $(DMFT1, DMFT2)$ index of the jth observation in the ith treatment, $j = 1, .., n_i$, and $i = 1, ..., k$. Now, the observation (0, 0) may come from a bivariate Poisson distribution or from a distribution with a point mass of one at (0, 0). Let

$$I_{ij} = \begin{cases} 1, & \text{if } (y_{1ij}, y_{2ij}) \text{ is observed from } f_2(y_1, y_2 | \lambda_0, \lambda_1, \lambda_2), \\ 0, & \text{otherwise.} \end{cases}$$

In the application of the EM algorithm we consider I_{ij} as missing data.

Further, the pair of random variables (Y_1, Y_2) has a bivariate Poisson distribution, if $Y_1 = Z_1 + Z_0$ and $Y_2 = Z_2 + Z_0$, where $Z_i, i = 0, 1, 2$ are independent Poisson random variables with parameters λ_0, λ_1 and λ_2 respectively (Kocherlakota and Kocherlakota, 1992). Thus, for $j = 1, ..., n_i$, $i = 1, ..., k$, we can write $y_{1ij} = z_{1ij} + z_{0ij}$ and $y_{2ij} = z_{2ij} + z_{0ij}$, and consider z_{0ij} as missing data.

In the application of the EM algorithm, the incomplete data consist of y_{1ij} and y_{2ij} and the corresponding complete data consist of $y_{1ij}, y_{2ij}, z_{0ij}$ and I_{ij}, $j = 1, ..., n_i$,

$i = 1, ..., k$. Note that under the complete data setup, $y_{1ij} - z_{0ij}$, $y_{2ij} - z_{0ij}$ and z_{0ij} are independent Poisson distributed with densities $f(y_{1ij} - z_{0ij}, \lambda_{1i})$, $f(y_{2ij} - z_{0ij}, \lambda_{2ij})$ and $f(z_{0ij}, \lambda_{0i})$ respectively. Therefore, the complete likelihood function is

$$L^c = \prod_{\{(i,j)|(y_{1ij}, y_{2ij})=(0,0)\}} \theta_{0i}^{(1-I_{ij})}[(1-\theta_{0i})f(y_{1ij}-z_{0ij},\lambda_{1i})f(y_{2ij}-z_{0ij},\lambda_{2ij})f(z_{0ij},\lambda_{0i})]^{I_{ij}}$$
$$\prod_{\{(i,j)|(y_{1ij}, y_{2ij})\neq(0,0)\}} [(1-\theta_{0i})f(y_{1ij}-z_{0ij},\lambda_{1i})f(y_{2ij}-z_{0ij},\lambda_{2ij})f(z_{0ij},\lambda_{0i})]^{I_{ij}},$$

where $\lambda_{2ij} = \lambda_{1i}\exp(x'_{1ij}\gamma_1 + \tilde{x}'_{2ij}\gamma_2)$ for $j = 1, ..., n_i$, $i = 1, ..., k$ and θ_{0i} is the zero-inflation parameter for ith treatment, $i = 1, ..., k$. Then, the complete data log-likelihood function is given by

$$\begin{aligned} l^c &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\{(1-I_{ij})\log(\theta_{0i}) + I_{ij}\log(1-\theta_{0i}) \\ &+ I_{ij}\log[f(y_{1ij}-z_{0ij},\lambda_{1i})f(y_{2ij}-z_{0ij},\lambda_{2ij})f(z_{0ij},\lambda_{0i})]\}. \end{aligned}$$

Let $x'_{ij} = (x'_{1ij}, \tilde{x}'_{2ij})$, $\lambda_1 = (\lambda_{11}, \lambda_{12}, ..., \lambda_{1k})'$, $\lambda_0 = (\lambda_{01}, \lambda_{02}, ..., \lambda_{0k})'$, $\theta_0 = (\theta_{01}, \theta_{02}, ..., \theta_{0k})'$, and $\gamma' = (\gamma'_1, \gamma'_2)$ with $\gamma'_1 = (\gamma_{11}, \gamma_{12}, ..., \gamma_{1k})$ and $\gamma'_2 = (\gamma_{21}, \gamma_{22}, ..., \gamma_{2(p-1)})$. The complete data log-likelihood l^c can be then written in a simplified form as

$$l^c = l_{\gamma,\lambda_1} + l_{\lambda_0} + l_{\theta_0},$$

where

$$\begin{aligned} l_{\gamma,\lambda_1} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} I_{ij}[(y_{1ij}-z_{0ji})\log(\lambda_{1i}) - \lambda_{1i} - \log(y_{1ij}-z_{0ij})! \\ &+ (y_{2ij}-z_{0ij})x'_{ij}\gamma - \lambda_{1i}\exp(x'_{ij}\gamma) - \log(y_{2ij}-z_{0ij})!], \\ l_{\lambda_0} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} I_{ij}[z_{0ij}\log\lambda_{0i} - \lambda_{0i} - \log(z_{0ij})!], \\ l_{\theta_0} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[I_{ij}\log(1-\theta_{0i}) + (1-I_{ij})\log(\theta_{0i})]. \end{aligned}$$

The maximum likelihood estimates of the parameters $\gamma_1, \gamma_2, \lambda_{1i}, \lambda_{0i}$ and θ_{0i}, $i = 1, ..., k$ can be found by using the EM algorithm. The E-step and M-step of the EM algorithm are described below.

E-step:

Calculate the expectations of the missing data z_{0ij} and I_{ij} conditional on the incomplete data (y_{1ij}, y_{2ij}), $j = 1, ..., n_i, i = 1, ..., k$, respectively as

$$\begin{aligned}
u_{ij} &= E(z_{0ij}|y_{1ij}, y_{2ij}) = \sum_{r=1}^{min\{y_{1ij},y_{2ij}\}} rPr[Z_{0ij} = r|Y_{1ij} = y_{1ij}, Y_{2ij} = y_{2ij}] \\
&= \sum_{r=1}^{min\{y_{1ij},y_{2ij}\}} r\frac{Pr[Z_{0ij} = r, Y_{1ij} = y_{1ij}, Y_{2ij} = y_{2ij}]}{Pr[Y_{1ij} = y_{1ij}, Y_{2ij} = y_{2ij}]} \\
&= \sum_{r=1}^{min\{y_{1ij},y_{2ij}\}} (1-\theta_{0i})\frac{rf(r,\lambda_{0i})f(y_{1ij}-r,\lambda_{1i})f(y_{2ij}-r,\lambda_{2ij})}{f_3(y_{1ij}, y_{2ij}|\, \theta_{0i}, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})} \\
&= \frac{\sum_{r=1}^{min\{y_{1ij},y_{2ij}\}} \lambda_{0i}(1-\theta_{0i})f(r-1,\lambda_{0i})f(y_{1ij}-r,\lambda_{1i})f(y_{2ij}-r,\lambda_{2ij})}{f_3(y_{1ij}, y_{2ij}|\, \theta_{0i}, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})} \\
&= \lambda_{0i}(1-\theta_{0i})\frac{\sum_{r=0}^{min\{y_{1ij}-1,y_{2ij}-1\}} f(y_{1ij}-1-r,\lambda_{1i})f(y_{2ij}-1-r,\lambda_{2ij})f(r,\lambda_{0i})}{f_3(y_{1ij}, y_{2ij}|\, \theta_{0i}, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})} \\
&= \lambda_{0i}(1-\theta_{0i})\frac{f_2(y_{1ij}-1, y_{2ij}-1|\, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})}{f_3(y_{1ij}, y_{2ij}|\, \theta_{0i}, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})},
\end{aligned}$$

and

$$\begin{aligned}
v_{ij} &= E(I_{ij}|y_{1ij}, y_{2ij}) = Pr[I_{ij} = 1|Y_{1ij} = y_{1ij}, Y_{2ij} = y_{2ij}] \\
&= \frac{Pr[I_{ij} = 1, Y_{1ij} = y_{1ij}, Y_{2ij} = y_{2ij}]}{Pr[Y_{1ij} = y_{1ij}, Y_{2ij} = y_{2ij}]} \\
&= (1-\theta_{0i})\frac{f_2(y_{1ij}, y_{2ij}|\, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})}{f_3(y_{1ij}, y_{2ij}|\, \theta_{0i}, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})}.
\end{aligned}$$

M-step:

Now, replacing z_{0ij} and I_{ij} by u_{ij} and v_{ij} respectively in l_{γ,λ_1}, l_{λ_0} and l_{θ_0}, we obtain $\hat{l}_{\gamma,\lambda_1}$, $\hat{l}_{\lambda_0}$ and $\hat{l}_{\theta_0}$. Note, to maximize $\hat{l}^c$ for given values of v_{ij} and u_{ij}, we only need to maximize $\hat{l}_{\lambda_0}$, $\hat{l}_{\theta_0}$ and $\hat{l}_{\gamma,\lambda_1}$ separately. Thus by maximizing $\hat{l}_{\lambda_0}$ we obtain

$$\hat{\lambda}_{0i} = \frac{\sum_{j=1}^{n_i} v_{ij} u_{ij}}{\sum_{j=1}^{n_i} v_{ij}}, i = 1, \ldots, k,$$

and by maximizing $\hat{l}_{\theta_0}$, we obtain

$$\hat{\theta}_{0i} = 1 - \textstyle\sum_{j=1}^{n_i} v_{ij}/n_i, i = 1, ..., k.$$

To find maximum likelihood estimates of γ and λ_{1i}, i=1, ..., k, we need to use the Newton-Raphson method. For this we first need to calculate the first derivatives of $\hat{l}_{\gamma,\lambda_1}$ with respect to γ and λ_1's, which are

$$\frac{\partial \hat{l}_{\gamma,\lambda_1}}{\partial \lambda_{1i}} = \sum_{j=1}^{n_i} v_{ij} \left[\frac{y_{1ij} + y_{2ij} - 2u_{ij}}{\lambda_{1i}} - 1 - \exp(-x_{ij}^{'}\gamma) \right] = U_{\lambda_{1i}}, i = 1, ..., k,$$

and

$$\frac{\partial \hat{l}_{\gamma,\lambda_1}}{\partial \gamma} = \sum_{j=1}^{n_i} v_{ij} \left[\frac{y_{2ij} - u_{ij}}{\lambda_{1i}} - \lambda_{1i} \exp(-x_{ij}^{'}\gamma) \right] x_{ij} = U_{\gamma},$$

where U_γ is $(k+p-1) \times 1$ vector and $U_{\lambda_1} = (U_{\lambda_{11}}, U_{\lambda_{12}}, ..., U_{\lambda_{1k}})'$. We then need to calculate the entries of the observed $(2k+p-1) \times (2k+p-1)$ information matrix

$$I^{obs}_{\gamma,\lambda_1} = \begin{pmatrix} I^{obs}_{\lambda_1\lambda_1} & I^{obs}_{\lambda_1\gamma} \\ I^{obs}_{\gamma\lambda_1} & I^{obs}_{\gamma\gamma} \end{pmatrix},$$

where $I^{obs}_{\lambda_1\lambda_1} = \left(-\frac{\partial \hat{l}^2_{\gamma,\lambda_1}}{\partial \lambda^2_{1i}} \right)_{k\times k}$, $I^{obs}_{\lambda_1\gamma} = I^{obs'}_{\gamma\lambda_1} = \left(-\frac{\partial \hat{l}^2_{\gamma,\lambda_1}}{\partial \lambda_{1i}\partial \gamma'} \right)_{k\times(k+p-1)}$

and $I^{obs}_{\gamma\gamma} = \left(-\frac{\partial \hat{l}^2_{\gamma,\lambda_1}}{\partial \gamma \partial \gamma'} \right)_{(k+p-1)\times(k+p-1)}$ with

$$-\frac{\partial \hat{l}^2_{\gamma,\lambda_1}}{\partial \lambda^2_{1i}} = \sum_{j=1}^{n_i} v_{ij} \left[\frac{y_{1ij} + y_{2ij} - 2u_{ij}}{\lambda^2_{1i}} \right],$$

$$-\frac{\partial \hat{l}^2_{\gamma,\lambda_1}}{\partial \lambda_{1i}\partial \gamma'} = \sum_{j=1}^{n_i} v_{ij} \left[\exp(-x_{ij}^{'}\gamma) \right] x_{ij},$$

$$-\frac{\partial \hat{l}^2_{\gamma,\lambda_1}}{\partial \gamma \partial \gamma'} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} v_{ij} \left[\lambda_{1i} \exp(-x_{ij}^{'}\gamma) \right] x_{ij} x_{ij}^{'}.$$

Then, given the values of $\lambda_1^{(s)}$, $\gamma^{(s)}$ and $(I^{obs}_{\lambda_1,\gamma})^{(s)}$ at the $s-$th step, the values of the parameter estimates at the $(s+1)-$th step are

$$\begin{pmatrix} \lambda_1^{(s+1)} \\ \gamma^{(s+1)} \end{pmatrix} = \begin{pmatrix} \lambda_1^{(s)} \\ \gamma^{(s)} \end{pmatrix} + \left(I^{(s)}\right)^{-1} \begin{pmatrix} U_{\lambda_1}^{(s)} \\ U_{\gamma}^{(s)} \end{pmatrix}.$$

To obtain the mle's of parameters $\gamma_1, \gamma_2, \lambda_{1i}, \lambda_{0i}$ and θ_{0i}, $i = 1, ..., k$, we need to iterate between the E-step and M-step until convergence.

7.4 Tests for treatment effects

Our interest is to test for the treatment effects after controlling for the effects of other covariates. Note, for the DMFT index data we have six treatments (schools) and 3 other covariates. Now the observed data log-likelihood is

$$l = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \log[f_3(y_{1ij}, y_{2ij}|\, \theta_{0i}, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})], \tag{7.4.1}$$

where $\lambda_{2ij} = \lambda_{1i}\exp(x'_{1ij}\gamma_1 + \tilde{x}'_{2ij}\gamma_2)$ for $j = 1, ..., n_i$ and $i = 1, ..., k$.

Let I be the expected information matrix for the parameters $\gamma, \lambda_1, \lambda_0$ and θ_0, obtained from the observed data log-likelihood (7.4.1). Note that there are $3k$ parameters $\lambda_{1i}, \lambda_{0i}$ and $\theta_{0i}, i = 1, ..., k$ in the zero-inflated bivariate Poisson distribution, k treatment parameters $\gamma_{1i}, i = 1, ..., k$ and $p-1$ regression parameters $\gamma_{2i}, i = 1, ..., p-1$ reprmeter of interest and $\phi' = (\gamma_2', \lambda_1', \lambda_0', \theta_0')$ is the nuisance parameter. Now partition I as

$$I = \begin{pmatrix} I_{\gamma_1\gamma_1} & I_{\gamma_1\phi} \\ I'_{\gamma_1\phi} & I_{\phi\phi} \end{pmatrix}.$$

Then, the approximate covariance matrix of $\hat{\gamma}_1$ is $I^{\gamma_1\gamma_1} = (I_{\gamma_1\gamma_1} - I_{\gamma_1\phi}I_{\phi\phi}^{-1}I'_{\gamma_1\phi})^{-1}$ (McCullagh and Nelder, 1989, Page 472), where $\hat{\gamma}_1$ is the maximum likelihood estimate of γ_1.

If we use the maximum likelihood estimates $\hat{\gamma}_1$ and $\hat{\phi} = (\hat{\gamma}_2, \hat{\lambda}_1, \hat{\lambda}_0, \hat{\theta}_0)$ of the parameters γ_1 and $\phi = (\gamma_2, \lambda_1, \lambda_0, \theta_0)$ in $I^{\gamma_1\gamma_1}$, we obtain $\hat{I}^{\gamma_1\gamma_1}$, the estimate of $I^{\gamma_1\gamma_1}$. Thus the asymptotic variance of $\hat{\gamma}_{1i}$, $i = 1, ..., k$, is given by the corresponding ith diagonal element of the matrix $\hat{I}^{\gamma_1\gamma_1}$. Then the effect of the ith treatment is tested by

$$Z_i = \hat{\gamma}_{1i}/\sqrt{var(\hat{\gamma}_{1i})}, \tag{7.4.2}$$

which is asymptotically distributed as $N(0, 1)$.

Now, denote the control group as c. Then, to test the effect of the ith group relative to that of the control group, we compare $\hat{\gamma}_{1i}$, $i \neq c$, with $\hat{\gamma}_{1c}$, for which we use

$$Z_{ic} = \frac{\hat{\gamma}_{1i} - \hat{\gamma}_{1c}}{\sqrt{var(\hat{\gamma}_{1i}) - 2cov(\hat{\gamma}_{1i}, \hat{\gamma}_{1c}) + var(\hat{\gamma}_{1c})}}, \tag{7.4.3}$$

where $cov(\hat{\gamma}_{1i}, \hat{\gamma}_{1c})$ is the $(i, c)-$entry of $\hat{I}^{\gamma_1\gamma_1}$, and $i = 1, 2, ..., k$ and $i \neq c$. The statistic Z_{ic} then is asymptotically distributed as $N(0, 1)$.

7.5 Analysis of the DMFT data

In this section, we deal with the analysis of the DMFT index data discussed earlier (for the data see Data sets, Table D.4). We first analyse the data using the ZIPR model (7.2.3) with log-likelihood.

$$l_0 = \sum_{i=1}^{6} \sum_{j=1}^{n_i} \log f_1(y_{2ij}, \lambda_{2ij}, \omega_i)$$

with $\log \lambda_{2ij} = x'_{1ij}\beta_1 + x'_{2ij}\beta_2$, for $j = 1, ..., n_i$ and $i = 1, ..., 6$. Let β_{21}, β_{22} and β_{20} be the components of β_2 corresponding to the covariates gender, ethnic and $\log(DMFT1 + 0.5)$ respectively.

We first test whether the zero-inflation parameter or $j = 1, ..., n_i$ and $i = 1, ..., 6$. Let β_{21}, β_{22} and β_{20} be the components of β_2 corresponding to the covariates gender, ethnic and $\log(DMFT1 + 0.5)$ respectively.

We first test whether the zero-inflation parameter ω varies from school to school. For this we consider two models:

Model ZI: Each school has different zero-inflation parameters $\omega_i, i = 1, ..., 6$, treatment effects $\beta_{1i}, i = 1, ..., 6$, and common regression parameter β_2.

Model ZII : Each school has different treatment effects $\beta_{1i}, i = 1, ..., 6$, and common zero-inflation parameter ω_0, common regression parameter β_2.

The maximized log-likelihoods along with the number of parameters estimated for the above two models are given in Table 7.1. Analysis of the results in Table 7.1 shows that zero-inflation parameters are not significantly different from school to school and Model ZII is the model of choice. Note that the values of log-likelihoods in Table 7.1 have some differences with those in Table 2 of Böhning et al. (1999, page 203). For example, the values of maximized log-likelihood for Model ZI and Model ZII are -1228.89 and -1232.02, and the corresponding log-likelihood values obtained by Böhning et al. (1999) are -1242.68 and -1246.89 respectively. This difference could be the result of the precision used in the calculation. We used double precision in our Fortran programming. However, the conclusion regarding the choice of the model remains the same.

We next check if we can eliminate any or both of the two covariates gender and ethnic group from the model. For this we consider the following sub-models of Model ZII:

Model $ZII1$: Each school has different treatment effects $\beta_{1i}, i = 1, ..., 6$, and common zero-inflation parameter ω_0 and common regression parameter β_{21} and β_{20} corresponding to ethnic group and $\log(DMFT1 + 0.5)$.

Model $ZII2$: Each school has different treatment effects $\beta_{1i}, i = 1, ..., 6$, and common zero-inflation parameter ω_0 and common regression parameter β_{22} and β_{20} corresponding to gender and $\log(DMFT1 + 0.5)$.

Model $ZII3$: Each school has different treatment effects $\beta_{1i}, i = 1, ..., 6$, and common zero-inflation parameter ω_0 and β_{20} corresponding to $\log(DMFT1 + 0.5)$.

The maximized log-likelihoods along with the number of parameters estimated for models $ZII1 - ZII3$ are also given in Table 7.1. Analyses of these results in Table 7.1 show that neither of the two covariates has significant effect. So, our final model is Model $ZII3$ with log-likelihood

$$l_0 = \sum_{i=1}^{6} \sum_{j=1}^{n_i} \log f_1(y_{2ij}, \lambda_{2ij}, \omega_0),$$

where, $\log \lambda_{2ij} = x'_{1ij}\beta_1 + \beta_{20} \log(y_{1ij} + 0.5)$, for $j = 1, ..., n_i$ and $i = 1, ..., 6$. The maximum likelihood estimates of the parameters of the ZII3 model together with their standard errors and other relevant quantities (test statistics) are given in Table 7.2.

Based on the Z-values in Table 7.2 the schools can be ranked, in terms of improvement in dental hygiene, from most significant improvement to the least significant improvement as School 2, School 5, School 1, School 6, School 3 and School 4. Now we compare all schools with the control group (School 3). The Z_c values to do these comparisons are also given in Table 7.2. It can be seen that the schools that significantly improved compared to School 3 are School 2, School 5 and School 1. These results coincide with those in Table 1 of Böhning et al. (1999). Note, their final model includes the covariates such as gender and ethnic. We do not include these covariates as they do not contribute significantly to the model fitting (see Table 7.1).

We now analyse the data using the ZIBPR model (7.2.6). Recall that the observed data log-likelihood for this model is

$$l = \sum_{i=1}^{6}\sum_{j=1}^{n_i}\log[f_3(y_{1ij}, y_{2ij}|\, \theta_{0i}, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})], \tag{7.5.1}$$

where $\lambda_{2ij} = \lambda_{1i}\exp(x'_{1ij}\gamma_1 + \tilde{x}'_{2ij}\gamma_2)$ for $j = 1, ..., n_i$ and $i = 1, ..., 6$. Let γ_{21} and γ_{22} be the components of γ_2 corresponding to the covariates gender and ethic respectively.

In what follows we fit the model (7.5.1) and a few sub-models to the DMFT index data. The models considered are:

Model I: Each school has different zero-inflation parameters $\theta_{0i}, i = 1, ..., 6$, covariance parameters $\lambda_{0i}, i = 1, ..., 6$, $\lambda_{1i}, i = 1, ..., 6$ parameters, treatment effect parameters γ_{1i}, $i = 1, ..., 6$ and common regression parameter γ_2.

Model II: Each school has different covariance parameters $\lambda_{0i}, i = 1, ..., 6$, $\lambda_{1i}, i = 1, ..., 6$ parameters, treatment effect parameters $\gamma_{1i}, i = 1, ..., 6$ and common zero-inflation parameter θ_{00} and common regression parameter γ_2.

Model III: Each school has different zero-inflation parameter $\theta_{0i}, i = 1, ..., 6$, $\lambda_{1i}, i = 1, ..., 6$ parameters, treatment effect parameters $\gamma_{1i}, i = 1, ..., 6$ and common covariance parameter λ_{00} and common regression parameter γ_2.

Model IV: Each school has different $\lambda_{1i}, i = 1, ..., 6$ parameters, treatment effect parameters $\gamma_{1i}, i = 1, ..., 6$ and common zero-inflation parameter θ_{00}, common covariance parameter λ_{00} and common regression parameter γ_2.

The maximized log-likelihoods along with the number of parameters estimated for the above four models are given in Table 7.3. Analysis of the results in Table 7.3 shows that Model I is the model of choice. Again, we next check if we can eliminate any or both of the two covariates gender and ethnic group from the model. For this we consider the following sub-models of Model I:

Model I1: Each school has different zero-inflation parameters $\theta_{0i}, i = 1, ..., 6$, covariance parameters $\lambda_{0i}, i = 1, ..., 6$, $\lambda_{1i}, i = 1, ..., 6$ parameters, treatment effect parameters $\gamma_{1i}, i = 1, ..., 6$ and common regression parameter γ_{21} corresponding to the covariate ethnic group.

Model I2: Each school has different zero-inflation parameters $\theta_{0i}, i = 1, ..., 6$, covariance parameters $\lambda_{0i}, i = 1, ..., 6$, $\lambda_{1i}, i = 1, ..., 6$ parameters, treatment effect parameters $\gamma_{1i}, i = 1, ..., 6$ and common regression parameter γ_{22} corresponding to the covariate gender.

Model I3: Each school has different zero-inflation parameters $\theta_{0i}, i = 1, ..., 6$, covariance parameters $\lambda_{0i}, i = 1, ..., 6$, $\lambda_{1i}, i = 1, ..., 6$ parameters, treatment effect parameters $\gamma_{1i}, i = 1, ..., 6$

The maximized log-likelihoods along with the number of parameters estimated for models I1-I3 are also given in Table 7.3. Analyses of these results in Table 7.3 show that neither of the two covariates has a significant effect. So, our final model is Model

I3 with

$$l = \sum_{i=1}^{6}\sum_{j=1}^{n_i} \log[f_3(y_{1ij}, y_{2ij}| \, \theta_{0i}, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})], \qquad (7.5.2)$$

where $\lambda_{2ij} = \lambda_{1i}\exp(x'_{1ij}\gamma_1)$ for $j = 1, ..., n_i$ and $i = 1, ..., 6$. The maximum likelihood estimates of the parameters of the Model I3 together with their standard errors and other relevant quantities (test statistics) are given in Table 7.4.

From the Z-values in Table 7.4, we see that dental hygiene significantly improved for all schools. Again, based on these Z-values we can rank the schools, in terms of improvement in dental hygiene, from the most significant improvement to least significant improvement as School 1 , School 5 , School 2, School 3, School 4 and School 6. Note these rankings differ from those obtained by analysing the data using the ZIPR model. However, these rankings of the schools coincide with those that can be seen from the mean difference $\bar{y}_{1i} - \bar{y}_{2i}$, $i = 1, ..., 6$, where $\bar{y}_{1i}$ and $\bar{y}_{2i}$ are the means of the DMFT1 and DMFT2 respectively (see Table 7.5).

We note further that dental hygiene improved not only for the school children in which some treatments were applied, but also for the children in the control group. This finding coincides with that found by Böhning et al. (1999). Böhning et al. (1999) explain the improvement as "There are two possible explanations for this. One possibility is a trend in dental caries that has affected all the schools in the BELCAP study in a similar way. However, it could be that during the study, especially while the intervention phase was in progress, information from one school to another could have been passed over (spillover effect). Frequently meetings were held between the co-ordinators of the BELCAP study and the heads of the schools, to discuss matters concerning the execution of the programs. So, in this case a spillover effect cannot be completely excluded."

We now compare improvement of dental hygiene in School 1, School 2, School 4, School 5 and School 6 with School 3 (control group). The Z_c values are given in Table 7.4. It can be seen that dental hygiene did not improve significantly in the schools in which treatments were applied compared to that in School 3. Thus, it looks as though improvement in dental hygiene occured among the children of all schools mainly because of the awareness of dental hygiene as a result of the experiment.

Table 7.1: Maximized log-likelihoods under ZIPR models

Model	log-likelihood value	number of parameters in model
ZI	-1228.49	16
ZII	-1232.02	11
$ZII1$	-1232.024	10
$ZII2$	-1232.770	9
$ZII3$	-1232.773	8

Table 7.2: Parameter estimates of the ZIPR model ZII3 with standard errors

Parameter	Estimate	Standard error of β_1	Z-value	Z_c-value
School 1 (β_{11})	-0.369	0.095	-3.891	-2.597
School 2 (β_{12})	-0.462	0.098	-4.692	-3.291
School 3 (β_{13})	-0.136	0.087	-1.569	0.000
School 4 (β_{14})	-0.112	0.087	-1.281	0.287
School 5 (β_{15})	-0.372	0.090	-4.113	-2.703
School 6 (β_{16})	-0.216	0.091	-2.380	-0.875
$\log(DMFT1 + 0.5)(\beta_{20})$	0.733	0.040	18.342	.
ω_0	0.045	.	.	.

Table 7.3: Maximized log-likelihoods under the ZIPBR model

Model	log-likelihood value	number of parameters in model
I	-3016.21	27
II	-3026.94	22
III	-3031.92	22
IV	-3041.27	17
I1	-3016.30	26
I2	-3017.37	25
I3	-3017.39	24

Table 7.4: Effect estimates with standard error for DMFT index data based on model I3

Parameter	θ_0	λ_0	λ_1	γ_1	Standard error of γ_1	Z-value	Z_c-value
School 1 (γ_{11})	0.085	0.838	3.427	-1.052	0.106	-9.941	0.322
School 2 (γ_{12})	0.222	0.453	3.098	-0.927	0.104	-8.891	1.003
School 3 (γ_{13})	0.088	1.844	2.211	-1.111	0.151	-7.349	0.000
School 4 (γ_{14})	0.131	1.185	2.615	-0.706	0.103	-6.861	2.211
School 5 (γ_{15})	0.221	1.256	3.010	-1.247	0.134	-9.293	-0.674
School 6 (γ_{16})	0.143	1.219	2.196	-0.907	0.136	-6.626	1.001

Table 7.5: Averages of DMFT1, DMFT2 and their differences

average	DMFT1	DMFT2	DMFT1-DMFT2
School 1	3.90	1.86	2.04
School 2	2.76	1.31	1.45
School 3	3.70	2.35	1.35
School 4	3.30	2.15	1.15
School 5	3.32	1.65	1.67
School 6	2.93	1.81	1.12

Table 7.6: The Rank of treatment effect for different models according to Z-value.

Model	Böhning et al. (1999)	Rank in Table 7.2	Rank in Table 7.4	Rank for average (DMFT1-DMFT2)
School 1	3	3	1	1
School 2	1	1	3	3
School 3	6	5	4	4
School 4	5	6	5	5
School 5	2	2	2	2
School 6	4	4	6	6

Chapter 8

Summary and Future Research

This chapter summarizes the conclusions of this thesis and recommends some problems for future research.

8.1 Summary

This thesis consists of two parts. Part I, including Chapter 3, Chapter 4 and Chapter 5, developes procedures for testing homogeneity of several location-scale populations in general. We compare our procedure with the procedure proposed by Singh (1986) for the normal case and apply the general method to several non-normal cases. Part II, including Chapter 6 and Chapter 7, analyses the treatment effects of paired count data with zero-inflation and over-dispersion. We develop two procedures, one of which is illustrated by the PVC data (Berry, 1987) and the other is illustrated by the DMFT data (Böhning et al., 1999).

For testing simultaneously the equality of means and the equality of variances of several normal populations, Singh (1986) uses a test statistic based on the combination two independent likelihood ratio statistics. Singh's procedure is based on a method by Fisher (1950) for combining two or more independent test statistics to test a general hypothesis. We extend Fisher's method to location-scale models in general. Two test statistics are developed, one of which is based on the combination of two likelihood ratio statistics and the other is based on the combination of two score test statistics. Under the general location-scale setup, asymptotic independence is established for the two likelihood ratio statistics as well as for the two score test statistics. Then, by applying the general results, we obtain specific test statistics for testing homogeneity of several normal (μ, σ^2) populations, several negative binomial (m, c) populations, several beta-binomial (π, ϕ) populations and several Weibull (ψ, ϕ) populations . In the normal case exact independence of the two likelihood ratio statistics is shown by Singh (1986). In this thesis, we show exact independence of the two score test statistics. In all four cases simulations are conducted to compare the two procedures. We conclude that Fisher's method of combining two statistics, even when they are only asymptotically independent, does, in general, perform well for testing homogeneity of several populations in terms of the means and the variances. However, the score test statistics have simple forms, are easy to calculate, and have uniformly good level properties. Therefore Fisher's method based on combining two score test statistics might be the method of choice.

Another problem considered in this thesis is the analysis of data in the form of paired counts with zero-inflation and over-dispersion. As we point out before, Poisson and binomial models are most widely used models for count data. However, those model may not fit count data well, when the data exhibit zero-inflation and over-dispersion. In practice, the paired counts data are obtained before and after an experiment and the extra zeros may occur in different ways. For example, the PVC data, given as paired counts by Berry (1987) for before and after drug administration, only have extra zeros after the drug administration, while the DMFT index data (Böhning et al., 1999), which have the form of (DMFT1, DMFT2) as paired count data for pre-treatment and after-treatment, have extra zeros, in most situations, as the common pair of (0, 0).

For the PVC data, the score test statistic for testing for treatment effect in data is obtained based on a zero-inflated beta-binomial model, which allows us to analyse treatment effects while considering the effect of zero-inflation and over-dispersion. Results of a small simulation experiment, to study small sample behavior of a score test and a likelihood ratio test, are reported and the PVC data are analysed. Both the score tests and the log-likelihood ratio tests show good properties. Either the score tests or the log-likelihood ratio tests can be used for testing the presence of treatment effect. The score tests, however, may be preferable because they use estimates of the parameters only under the null hypothesis. For the DMFT data, we introduce a zero-inflated bivariate Poisson regression model (ZIBPR). We jointly model the pre-treatment and the post-treatment counts. A model selection procedure is given to decide on an appropriate model. For the DMFT index data, based on the model selected, we arrive at a ranking of the treatment effects which coincides with that from a simple analysis of treatment effects.

8.2 Future research

For univariate case, based on a Poisson model and a binomial model, we can obtain a zero-inflated Poisson and a zero-inflated binomial model. Further, based on an over-dispersion model such as a negative binomial model and a beta-binomial model, we can obtain zero-inflated negative binomial and zero-inflated beta-binomial model, which are widely used to fit the count data with zero-inflation and over-dispersion (Deng and Paul, 2000, and Hall, 2000). Therefore, for data in the form of paired counts with zero-inflation and over-dispersion, it would be of interest to develop zero-inflated bivariate Poisson model and zero-inflated bivariate binomial model as well as zero-inflated bivariate negative binomial and zero-inflated bivariate beta-binomial model to fit paired counts with varying zero-inflation and over-dispersion parameters. In this thesis, we analyse the PVC data based on the zero-inflated beta-binomial model and the DMFT index data based on the zero-inflated bivariate Poisson model. And also the bivariate Poisson model can be further generalized by introducing two additional zero-inflation parameters: one when zero inflation occurs for the pre-treatment count and not for the post-treatment count and the other when zero inflation occurs for post-treatment count and not for the pre-treatment count. The detail discussion is omitted here. In this section, we focus on testing the homogeneity in the presence of the nuisance parameters.

In Chapter 3, we extended Fisher's method to location-scale models in general. Under the general location-scale setup asymptotic independence was established for the two score test statistics by using the transformation of original parameters ψ and ϕ to the orthogonal parameters ψ and λ according the results of Cox and Ried (1987).

We found that:

(1) Compared with the score statistics which are obtained without using the orthogonal transformation, the score test statistics in Chapter 3 have simpler expressions. The reason is that $E\left(-\frac{\partial l^2}{\partial\psi\partial\lambda}\right) = 0$ and hence the maximum likelihood estimates $\hat{\psi}$ and $\hat{\lambda}$ of parameters ψ and λ are asymptotically independent. This simplifies the information matrix.

(2) Even though it may not be easy to get explicit solution of the partial differential equation (3.2.1), our score test statistics can be obtained in terms of the original parameters without solving such a partial differential equation.

Cox and Ried (1987) outline the properties of orthogonality of ψ and λ. We list some of them here, which are related to our problems of interest.

(1) The maximum likelihood estimates $\hat{\psi}$ and $\hat{\lambda}$ are asymptotically independent;

(2) The asymptotic standard error of $\hat{\psi}$ of ψ is the same irrespective of whether λ is treated as known or unknown;

(3) $\hat{\psi}_\lambda = \hat{\psi}(\lambda)$, the maximum likelihood estimate of ψ when λ is given, varies only slowly with λ.

According to these properties, we may conclude that the orthogonal nuisance parameters may have less effect on the score statistics than those obtained by using non-orthogonal nuisance parameters. It would be of interest to derive the score test statistic with the orthogonal nuisance parameters and compare it with the one without such transformation to see if we may gain some better properties to test a statistical hypothesis based on the orthogonal nuisance parameters.

Testing homogeneity in the presence of the nuisance parameters are widely discussed. In the following, we are interested in three cases based on score test statistics. We obtain the score test statistics for: (I) Testing homogeneity in presence of common nuisance parameters; (II) Testing the homogeneity against central mixture alternatives; (III) Testing homogeneity, in terms of departure from simple models, in generalized linear models. Conducting simulations and applying the above results are the interest of future study.

(I) Test of homogeneity in the presence of common nuisance parameters

Let Y_{ij} be the random variable for the observation j of group i, $j = 1, ..., n_i$, $i = 1, ..., k$ with $N = \sum_{i=1}^{k} n_i$. We assume that the probability density function of Y_{ij} is $f(y_{ij}; \psi_i, \phi)$ and $l_i = \sum_{j=1}^{n_i} \log f(y_{ij}; \psi_i, \phi)$ is the log-likelihood function for y_i's. The usual homogeneity hypothesis in the presence of the common nuisance parameters is $H_0 : \psi_1 = \psi_2 = ... = \psi_k = \psi$, where ϕ and ψ are unspecified vs $H_1 : \psi_i$, $i = 1, ..., k$ are not all same, where ϕ is unspecified.

Let $s_{nt} = \left(\frac{\partial l_1}{\partial \psi}, \frac{\partial l_2}{\partial \psi}, ..., \frac{\partial l_{k-1}}{\partial \psi} \right)' \Bigg|_{H_0}$ and

$$A_{nt} = diag\left(I_{\psi\psi}^{(1)}, I_{\psi\psi}^{(2)}, ..., I_{\psi\psi}^{(k-1)} \right),$$

$$C_{nt} = \begin{pmatrix} I_{\psi\psi}^{(1)} & I_{\psi\phi}^{(1)} \\ I_{\psi\psi}^{(2)} & I_{\psi\phi}^{(2)} \\ \vdots & \vdots \\ I_{\psi\psi}^{(k-1)} & I_{\psi\phi}^{(k-1)} \end{pmatrix},$$

and

$$D_{nt} = \begin{pmatrix} \sum_{i=1}^{k} I_{\psi\psi}^{(i)} & \sum_{i=1}^{k} i_{\psi\phi}^{(i)} \\ \sum_{i=1}^{k} I_{\psi\phi}^{(i)} & \sum_{i=1}^{k} I_{\phi\phi}^{(i)} \end{pmatrix},$$

where $I_{\psi\psi}^{(i)} = E\left(-\frac{\partial^2 l_i}{\partial\psi^2}\bigg|_{H_0}\right)$, $I_{\psi\phi}^{(i)} = E\left(-\frac{\partial^2 l_i}{\partial\psi\partial\phi}\bigg|_{H_0}\right)$ and $I_{\phi\phi}^{(i)} = E\left(-\frac{\partial^2 l_i}{\partial\phi^2}\bigg|_{H_0}\right)$, $i = 1, 2, ..., k$.

If we use maximum likelihood estimates $\hat{\psi}$ and $\hat{\phi}$ of the nuisance parameters ψ and ϕ in s_{nt}, A_{nt}, C_{nt} and D_{nt}, then the score test for testing H_0 against H_1 is

$$S_{nt} = \hat{s_{nt}}' \left(\hat{A_{nt}} - \hat{C_{nt}}\hat{D_{nt}}^{-1}\hat{C_{nt}}'\right)^{-1} \hat{s_{nt}}. \tag{8.2.1}$$

Now, we want to derive the homogeneity test statistic in the presence of a common orthogonal parameter. For this we need to transform the parameters $(\psi_i,\ \phi)$, $i = 1, ..., k$, into a set of orthogonal parameters $(\psi_i^*,\ \phi)$, $i = 1, ..., k$(Cox and Reid, 1987). Let $(\psi_i(\psi_i^*, \phi),\ \phi)$, $i = 1, ..., k$, be such a transformation which satisfies

$$I_{\psi_i\phi}^{(i)} + I_{\psi_i\psi_i}^{(i)} \frac{\partial\psi_i}{\partial\phi} = 0, i = 1, \ldots, k, \tag{8.2.2}$$

where $I_{\psi_i\phi}^{(i)} = E\left(-\frac{\partial^2 l_i}{\partial\psi_i\partial\phi}\right)$ and $I_{\psi_i\psi_i}^{(i)} = E\left(-\frac{\partial^2 l_i}{\partial\psi_i^2}\right)$. Then the above homogeneity hypothesis is equivalent to

$H_0 : \psi_1^* = \psi_2^* = ... = \psi_k^* = \psi^*$, with ψ^* and ϕ being unspecified vs $H_1 : \psi_i^*$, $i = 1, ..., k$, are not all the same and ϕ is unspecified.

As in Section 3.2, in terms of the original parameters (ψ, ϕ) under H_0, the score statistic S_t is given by

$$S_t = \sum_{i=1}^{k} \frac{\hat{s}_{ti}^{\,2}}{\hat{v}_{ti}}, \tag{8.2.3}$$

where $\hat{s}_{ti}$ and $\hat{v}_{ti}$, $i = 1, ..., k$, are estimated values of $\frac{\partial l_i}{\partial\psi}\bigg|_{H_0}$ and $I_{\psi\psi}^{(i)} = E\left(-\frac{\partial^2 l_i}{\partial\psi^2}\bigg|_{H_0}\right)$,

$i = 1, ..., k$, obtained by replacing the nuisance parameters ψ and ϕ by the corresponding maximum likelihood estimates $\hat{\psi}$ and $\hat{\phi}$ under H_0.

Both of the score test statistics S_t and S_{nt} can be used to test homogeneity in the presence of the common nuisance parameter ϕ. It would be of interest to compare performances of these score statistics, both theoretically and by simulation which will be subject of future investigation.

(II) Test of homogeneity against central mixture alternatives

Mixture distributions are widely used to obtain a over-dispersion family of sampling models. Test for mixtures are usually limited to a specific mixing. For example, by mixing the Poisson distribution with the gamma distribution, we obtain the negative binomial distribution and by mixing the binomial distribution with the beta distribution, we obtain the beta-binomial distribution. Liang (1987), Zelterman and Chen (1988) develop score test statistics based on the central mixture model as the alternative to test the homogeneity. This central mixture model is obtained by a general mixture without specifying any distribution. In this section, we develop a score test statistic by using orthogonal transformation to test homogeneity against central mixture alternative based on the model proposed by Liang (1987).

Let Y_{ij} be the random variable for observation j of group i, $j = 1, ..., n_i$, $i = 1, ..., k$, with $N = \sum_{i=1}^{k} n_i$. We assume that the probability density function of Y_{ij} is $f(y_{ij}; \psi_i, \phi)$. The usual homogeneity hypothesis in presence of the common nuisance parameter is

$H_0 : \psi_1 = \psi_2 = ... = \psi_k = \psi$, where ψ and ϕ are unspecified vs $H_1 : \psi_i$, $i = 1, ..., k$, are not all the same, where ϕ is unspecified

Assume that $\psi_i = \psi + \theta^{\frac{1}{2}} z_i$, $i = 1, ..., k$, where the z_i's are independently distributed from an unknown distribution F with zero mean and unit variance. Let

$$l_i(y_i; \theta, \psi, \phi) = \sum_{j=1}^{n_i} \log \left[\int f(y_{ij}; \psi + \theta^{\frac{1}{2}} z_i, \phi) dF(z_i) \right] \tag{8.2.4}$$

is the log-likelihood function based on the mixed model for the y_i's. Then, the hypothesis of homogeneity is equivalent to the hypothesis

$H_0 : \theta = 0$ vs $H_1 : \theta > 0$.

Under some regularity conditions, the score test statistic developed by Liang (1987) for testing H_0 is

$$S_{cnt} = \hat{s_{cnt}} / \hat{v_{cnt}}^{\frac{1}{2}}, \tag{8.2.5}$$

where

$$s_{cnt} = \sum_{i=1}^{k} \frac{\partial l_i}{\partial \theta} \bigg|_{H_0} = \frac{1}{2} \sum_{i=1}^{k} \left[\left(\frac{\partial l_i}{\partial \psi} \right)^2 + \frac{\partial^2 l_i}{\partial \psi^2} \right] \bigg|_{H_0},$$

and

$$v_{cnt} = i_{\theta\theta} - (i_{\theta\psi}, i_{\theta\phi}) \begin{pmatrix} i_{\psi\psi} & i_{\psi\phi} \\ i_{\phi\psi} & i_{\phi\phi} \end{pmatrix}^{-1} \begin{pmatrix} i_{\theta\psi} \\ i_{\theta\phi} \end{pmatrix},$$

with $i_{\theta\theta} = \sum_{i=1}^{k} E \left(\frac{\partial l_i}{\partial \theta} \bigg|_{H_0} \right)^2$, $i_{\theta\psi} = \sum_{i=1}^{k} E \left(\frac{\partial l_i}{\partial \psi} \bigg|_{H_0} \right)^2$, ..., etc, and $\hat{s}_{cnt}$ and $\hat{v}_{cnt}$ are the estimated values of s_{cnt} and v_{cnt} by replacing the nuisance parameters ψ and ϕ with the corresponding maximum likelihood estimates $\hat{\psi}$ and $\hat{\phi}$ under H_0.

Now, we develop the score test statistic by using the orthogonal nuisance parameters. For this we need to transform the parameters (θ, ψ, ϕ) into a set of orthogonal parameters (θ, ψ^*, ϕ^*), such that θ is orthogonal to (ψ^*, ϕ^*)(Cox and Reid, 1987). Let

$(\theta,\ \psi(\theta,\psi^*,\phi^*),\ \phi(\theta,\psi^*,\phi^*))$ be such a transformation which satisfies the equation

$$\begin{pmatrix} i_{\phi\phi} & i_{\phi\psi} \\ i_{\phi\psi} & i_{\psi\psi} \end{pmatrix} \begin{pmatrix} \frac{\partial\phi}{\partial\theta} \\ \frac{\partial\psi}{\partial\theta} \end{pmatrix} = - \begin{pmatrix} i_{\theta\phi} \\ i_{\theta\psi} \end{pmatrix}. \tag{8.2.6}$$

Now, in terms of the orthogonal parameters, we denote the likelihood function based on the mixed model for the y_i's as

$$l_i^* = \sum_{j=1}^{n_i} l^*(y_{ij};\theta,\psi^*,\phi^*) = \sum_{j=1}^{n_i} \log\left[\int f(y_{ij};\psi^* + \theta^{\frac{1}{2}} z_i, \phi^*) dF(z_i)\right]. \tag{8.2.7}$$

Let

$$s_{ct} = \sum_{i=1}^{k} \left.\frac{\partial l_i^*}{\partial\theta}\right|_{H_0} = \sum_{i=1}^{k} \left.\left\{ \frac{\partial l_i}{\partial\psi}\frac{\partial\psi}{\partial\theta} + \frac{\partial l_i}{\partial\phi}\frac{\partial\phi}{\partial\theta} + \frac{1}{2}\left[\left(\frac{\partial l_i}{\partial\psi}\right)^2 + \frac{\partial^2 l_i}{\partial\psi^2}\right]\right\}\right|_{H_0},$$

and

$$v_{ct} = i_{\theta\theta}^*|_{H_0} = \sum_{i=1}^{k} E\left(\left.\frac{\partial l_i^*}{\partial\theta}\right|_{H_0}\right)^2.$$

Note that $\frac{\partial\psi}{\partial\theta}$ and $\frac{\partial\phi}{\partial\theta}$ in the expressions for s_{ct} and v_{ct} can be expressed in terms of $i_{\phi\phi}, i_{\phi\psi}$,...etc from (8.2.6). So we can calculate the quantities s_{ct} and v_{ct} in terms of original parameters ψ and ϕ, without solving the partial differential equation (8.2.6). If we use maximum likelihood estimates $\hat{\psi}$ and $\hat{\phi}$ of the nuisance parameters ψ and ϕ in s_{ct} and v_{ct}, then the score test for testing H_0 against H_1 is

$$S_{ct} = \hat{s}_{ct}/\hat{v}_{ct}^{\frac{1}{2}}.$$

Again, both of the score test statistics S_{cnt} and S_{ct} can be used to test the homogeneity against central mixture alternatives. Comparing the performances theoretically and through a simulation study would be interesting.

(III)Test of homogeneity for generalized linear models

Mixed effects models based on a generalized linear model are widely used in many statistical studies. These models can be used to fit cluster data which may have interclass correlation within clusters. Jacqmin and Commenges (1995) develop a score test of homogeneity for mixed effect models. In this section, we develop a score test statistic by using an orthogonal transformation to test homogeneity based on the model proposed by Jacqmin and Commenges (1995).

Let Y_{ij} be the random variable for observation j of group i, $j = 1, ..., n_i$, $i = 1, ..., k$ with $N = \sum_{i=1}^{k} n_i$. We assume that the probability density function of Y_{ij} is $f(y_{ij}; \psi_i, \phi)$. Further, we assume that the probability density function of Y_{ij} is defined as:

$$f(y_{ij}; \theta_{ij}, \phi) = \exp\left[\frac{\theta_{ij} y_{ij} - g(\theta_{ij})}{\phi} + C(y_{ij}, \phi)\right] \tag{8.2.8}$$

The mean and variance of Y_{ij} are $\mu_{ij} = E(Y_{ij}) = g'(\theta_{ij})$ and $\sigma_{ij}^2 = var(Y_{ij}) = \phi g''(\theta_{ij})$. If θ_{ij} is a linear combination of the vector explanatory variables, then (8.2.8) specifies a generalized linear model , where θ_{ij} is the canonical parameter, ϕ is the dispersion parameter, and $(g')^{-1}$ is the canonical link (McCullagh and Nelder, 1989).

The mixed effects model considered by Jacqmin and Commenges (1995) is

$$\theta_{ij} = X_{ij}^T \beta + Z_{ij} \alpha_i, \tag{8.2.9}$$

where β denotes a $p \times 1$ vector of fixed effects with associated design vector X_{ij}, and α_i is the scalar random effect with associated covariate Z_{ij}.

Let $\alpha_i = \alpha + D^{\frac{1}{2}} v_i$, where the v_i's are independently and identically distributed with unspecified distribution F with zero mean and unit variance. We denote the log-likelihood function as

$$l_i = \sum_{j=1}^{n_i} l(y_{ij}; D, \beta, \alpha, \phi) = \sum_{j=1}^{n_i} \log\left[\int f(y_{ij}; \beta, \alpha + D^{\frac{1}{2}} v_i, \phi) dF(v_i)\right], \tag{8.2.10}$$

The hypothesis of homogeneity is $H_0 : D = 0$ vs $H_1 : D > 0$.

To derive score test statistic for H_0, Jacqmin and Commenges (1995) first assume that the dispersion parameter ϕ is known and (α, β) are considered as nuisance parameters. Following Liang (1987) and Chescher (1984), the score test statistic $H_S(\alpha, \beta, \phi)$ is obtained (for details, see Jacqmin and Commenges, 1995, page 1239). Further, when the parameter ϕ is unknown, a consistent estimate $\hat{\phi} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\hat{\mu_{ij}})^2/\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} g''(\hat{\theta}_{ij})\right)$ is used to replace ϕ in the statistic $H_S(\alpha, \beta, \phi)$. So $H_S(\alpha, \beta, \hat{\phi})$ is the statistic to test homogeneity when the parameter ϕ is unknown.

Now, in the following (IIIA) and (IIIB), we simplify the information matrix through an orthogonal transformation and obtain the exact variance of the score function. Further, we obtain score test statistics S_1 and S_2 corresponding to the statistics $H_S(\alpha, \beta, \phi)$ and $H_S(\alpha, \beta, \hat{\phi})$. A simulation study to compare the performance of these procedures would be interesting.

(IIIA) Score test of homogeneity when ϕ is known

For this we need to transform the parameters $(D,\ \beta,\ \alpha)$ into a set of orthogonal parameters $(D,\ \beta^*,\ \alpha^*)$, such that D is orthogonal to $(\beta^*,\ \alpha^*)$(Cox & Reid, 1987). Let $(D,\ \beta(D, \beta^*, \alpha^*),\ \alpha(D, \beta^*, \alpha^*))$ be such a transformation which satisfies the equation

$$\begin{pmatrix} i_{\alpha\alpha} & i_{\alpha\beta} \\ i'_{\alpha\beta} & i_{\beta\beta} \end{pmatrix} \begin{pmatrix} \dfrac{\partial \alpha}{\partial D} \\ \dfrac{\partial \beta}{\partial D} \end{pmatrix} = - \begin{pmatrix} i_{\alpha D} \\ i_{\beta D} \end{pmatrix}, \tag{8.2.11}$$

where $i_{\alpha\alpha} = \sum_{i=1}^{k} E\left(\dfrac{\partial l_i}{\partial \alpha}\right)^2$, $i_{\alpha\beta} = \sum_{i=1}^{k} E\left[\left(\dfrac{\partial l_i}{\partial \alpha}\right)\left(\dfrac{\partial l_i}{\partial \beta'}\right)\right]$,..., etc.

Now in terms of the orthogonal parameters, we denote the log-likelihood function based on the mixed model as

$$l_i^* = \sum_{j=1}^{n_i} l^*(y_{ij}; D, \beta^*, \alpha^*, \phi) = \sum_{j=1}^{n_i} \log \left[\int f(y_{ij}; \beta^*, \alpha^* + D^{\frac{1}{2}} v_i, \phi) dF(v_i) \right]. \quad (8.2.12)$$

Let

$$s_1 = \sum_{i=1}^k \left. \frac{\partial l_i^*}{\partial D} \right|_{H_0} = \sum_{i=1}^k \left. \left\{ \left(\frac{\partial l_i}{\partial \beta} \right)' \frac{\partial \beta}{\partial D} + \frac{\partial l_i}{\partial \alpha} \frac{\partial \alpha}{\partial D} + \frac{1}{2} \left[\left(\frac{\partial l_i}{\partial \alpha} \right)^2 + \frac{\partial^2 l_i}{\partial \alpha^2} \right] \right\} \right|_{H_0},$$

and

$$v_1 = i_{DD}^* |_{H_0} = \sum_{i=1}^k E \left(\left. \frac{\partial l_i^*}{\partial D} \right|_{H_0} \right)^2.$$

Note that $\frac{\partial \alpha}{\partial D}$ and $\frac{\partial \beta}{\partial D}$ in the expression of s_1 and v_1 can be expressed in terms of $i_{\alpha\alpha}, i_{\alpha\beta}$,...etc from (8.2.11). So we can express the quantities s_1 and v_1 in terms of the original parameters β and α without solving the partial differential equation (8.2.11). If we use maximum likelihood estimates $\hat{\beta}$ and $\hat{\alpha}$ of the nuisance parameters β and α in s_1 and v_1, then the score test for testing H_0 against H_1 is

$$S_1 = \hat{s}_1 / \hat{v}_1^{\frac{1}{2}}.$$

(IIIB). Score test of homogeneity when ϕ is unknown

For this we need to transform the parameters (D, β, α, ϕ) into a set of orthogonal parameters (D, β^*, α^*, ϕ^*), such that D is orthogonal to (β^*, α^*, ϕ^*)(Cox and Reid, 1987). Let

$$(D, \beta, \alpha, \phi) \longmapsto (D, \beta(D, \beta^*, \alpha^*, \phi^*), \alpha(D, \beta^*, \alpha^*, \phi^*), \phi(D, \beta^*, \alpha^*, \phi^*))$$

be such a transformation which satisfies the equation

$$\begin{pmatrix} i_{\alpha\alpha} & i_{\alpha\beta} & i_{\alpha\phi} \\ i'_{\alpha\beta} & i_{\beta\beta} & i_{\beta\phi} \\ i'_{\alpha\phi} & i'_{\beta\phi} & i_{\phi\phi} \end{pmatrix} \begin{pmatrix} \frac{\partial \alpha}{\partial D} \\ \frac{\partial \beta}{\partial D} \\ \frac{\partial \phi}{\partial D} \end{pmatrix} = - \begin{pmatrix} i_{\alpha D} \\ i_{\beta D} \\ i_{\phi D} \end{pmatrix}, \quad (8.2.13)$$

where $i_{\alpha\alpha} = \sum_{i=1}^{k} E\left(\frac{\partial l_i}{\partial \alpha}\right)^2$, $i_{\alpha\beta} = \sum_{i=1}^{k} E\left[\left(\frac{\partial l_i}{\partial \alpha}\right)\left(\frac{\partial l_i}{\partial \beta'}\right)\right]$,..., etc.

Now, in terms of the orthogonal parameters, we denote the log-likelihood function based on the mixed model as

$$l_i^* = \sum_{j=1}^{n_i} l^*(y_{ij}; D, \beta^*, \alpha^*, \phi^*) = \sum_{j=1}^{n_i} \log\left[\int f(y_{ij}; \beta^*, \alpha^* + D^{\frac{1}{2}} v_i, \phi^*) dF(v_i)\right]. \quad (8.2.14)$$

Let

$$\begin{aligned} s_2 &= \sum_{i=1}^{k} \left.\frac{\partial l_i^*}{\partial D}\right|_{H_0} \\ &= \sum_{i=1}^{k} \left.\left\{\left(\frac{\partial l_i}{\partial \beta}\right)' \frac{\partial \beta}{\partial D} + \frac{\partial l_i}{\partial \alpha}\frac{\partial \alpha}{\partial D} + \frac{\partial l_i}{\partial \phi}\frac{\partial \phi}{\partial D} + \frac{1}{2}\left[\left(\frac{\partial l_i}{\partial \alpha}\right)^2 + \frac{\partial^2 l_i}{\partial \alpha^2}\right]\right\}\right|_{H_0}, \end{aligned}$$

and

$$v_2 = i_{DD}^*|_{H_0} = \sum_{i=1}^{k} E\left(\left.\frac{\partial l_i^*}{\partial D}\right|_{H_0}\right)^2.$$

Again note that $\frac{\partial \alpha}{\partial D}, \frac{\partial \beta}{\partial D}$ and $\frac{\partial \phi}{\partial D}$ in the expressions of s_2 and v_2 can be expressed in terms of $i_{\alpha\alpha}, i_{\alpha\beta}$,..., etc from (8.2.13). So we can calculate the quantities s_2 and v_2 in terms of the original parameters β, α, ϕ without solving the partial differential equation (8.2.13). If we use maximum likelihood estimates $\hat{\beta}$, $\hat{\alpha}$ and $\hat{\phi}$ of the nuisance parameters β, α and ϕ in s_2 and v_2, then the score test for testing H_0 against H_1 is

$$S_2 = \hat{s}_2 / \hat{v}_2^{\frac{1}{2}}.$$

Table D.1: Counts of embryonic deaths in a control group and two treatment groups (McCaughran & Arnold, 1976, Table 6)

Number of	frequency		
deaths	control group	dose level 1	dose level 2
0	7	5	4
1	2	4	2
2	1	0	3
3	0	1	0
4	0	0	1

Table D.2: Toxicological data from Paul (1982)

Groups		
Control, C	(i)	1 1 4 0 0 0 0 0 1 0 2 0 5 2 1 2 0 0 1 0 0 0 0 3 2 4 0
	(ii)	12 7 6 6 7 8 10 7 8 6 11 7 8 9 2 7 9 7 11 10 4 8 10 12 8 7 1
Low dose, L	(i)	0 1 1 0 2 0 1 0 1 0 0 3 0 0 1 5 0 0 3
	(ii)	5 11 7 9 12 8 6 7 6 4 6 9 6 7 5 9 1 6 9
Medium dose, M	(i)	2 3 2 1 2 3 0 4 0 0 4 0 0 6 6 5 4 1 0 3 6
	(ii)	4 4 9 8 9 7 8 9 6 4 6 7 3 13 6 8 11 7 6 10 6
High dose, H	(i)	1 0 1 0 1 0 1 1 2 0 4 1 1 4 2 3 1
	(ii)	9 10 7 5 4 6 3 8 5 4 4 5 3 8 6 8 6

(i) Number of live foetuses affected by treatment. (ii) Total number of live foetuses.

Table D.3: The PVC counts for twelve patients one minute after administrating a drug with antiarrhythmic properties (Berry, 1987)

Patient	PVCs per minute		
number	Pre-drug(x_i)	Post-drug(y_i)	Total(m_i)
1	6	5	11
2	9	2	11
3	17	0	17
4	22	0	22
5	7	2	9
6	5	1	6
7	5	0	5
8	14	0	14
9	9	0	9
10	7	0	7
11	9	13	22
12	51	0	51

Table D.4: The DMFT index data (Böhning, Dietz, Schlattmann, Mendonca and Kirchner, 1999)

observation	DMFT1	DMFT2	Gender	Ethnic	School
1	6	3	1	3	1
2	2	1	0	3	1
3	1	0	1	3	1
4	7	2	1	2	1
5	3	3	0	2	1
6	0	0	0	3	1
7	4	3	1	2	1
8	2	1	1	2	1
9	8	2	0	1	1
10	3	3	1	2	1
11	0	0	0	3	1
12	5	2	1	1	1
13	2	2	0	2	1
14	5	2	0	1	1
15	3	2	0	2	1
16	3	1	0	2	1
17	2	1	0	2	1
18	2	3	1	2	1
19	2	1	1	2	1
20	0	0	0	1	1
21	8	1	0	2	1
22	2	2	0	2	1
23	5	1	0	1	1
24	3	3	1	1	1
25	5	1	1	3	1
26	4	3	1	3	1
27	6	4	1	2	1
28	2	1	0	1	1
29	8	1	0	1	1
30	5	2	1	1	1
31	4	3	0	1	1
32	4	0	1	1	1
33	1	1	0	2	1
34	1	1	1	2	1
35	6	5	1	2	1
36	4	4	1	1	1
37	6	4	0	2	1
38	0	0	0	2	1
39	7	0	0	2	1
40	5	2	1	2	1
41	0	1	0	2	1
42	0	2	0	2	1
43	8	5	0	1	1
44	6	4	1	2	1
45	7	5	1	2	1
46	5	4	0	3	1
47	8	3	1	2	1
48	1	2	0	2	1
49	4	0	1	2	1
50	4	2	0	3	1
51	4	1	1	2	1
52	5	4	1	2	1
53	4	3	1	3	1
54	5	1	1	1	1
55	6	2	1	1	1

observation	DMFT1	DMFT2	Gender	Ethnic	School
56	6	4	1	3	1
57	2	1	1	3	1
58	4	3	0	2	1
59	2	2	0	1	1
60	8	2	0	2	1
61	2	0	1	2	1
62	3	0	0	3	1
63	3	3	1	3	1
64	8	4	0	2	1
65	0	3	0	2	1
66	7	5	1	2	1
67	0	0	0	2	1
68	2	2	1	2	1
69	7	3	0	2	1
70	5	1	1	1	1
71	4	2	1	1	1
72	6	3	0	2	1
73	6	0	1	3	1
74	0	0	1	2	1
75	4	2	1	2	1
76	2	4	0	2	1
77	7	5	0	2	1
78	0	0	1	1	1
79	6	0	0	2	1
80	6	2	0	2	1
81	2	3	1	3	1
82	0	2	1	1	1
83	5	2	0	3	1
84	4	3	1	3	1
85	0	0	0	2	1
86	3	1	0	1	1
87	2	1	0	2	1
88	0	1	0	1	1
89	3	2	0	3	1
90	7	5	0	2	1
91	7	2	0	2	1
92	6	3	1	2	1
93	6	2	1	3	1
94	3	1	0	2	1
95	0	0	0	2	1
96	6	2	0	2	1
97	7	0	1	2	1
98	4	3	1	2	1
99	4	0	1	3	1
100	1	1	1	3	1
101	2	2	0	2	1
102	2	0	0	2	1
103	4	0	0	2	1
104	2	0	1	2	1
105	2	1	0	1	1
106	5	1	1	3	1
107	1	2	0	3	1
108	7	3	0	2	1
109	4	1	0	2	1
110	5	4	1	1	1
111	4	3	1	2	1
112	0	0	0	2	1
113	0	0	0	2	1
114	7	4	0	2	1
115	8	0	1	2	1
116	5	1	1	3	1
117	6	4	0	1	1
118	4	0	1	1	1
119	8	0	1	2	1
120	4	2	1	3	1
121	7	6	1	3	1
122	4	1	1	2	1
123	6	1	1	1	1
124	4	0	1	1	1
125	3	4	1	1	2

observation	DMFT1	DMFT2	Gender	Ethnic	School
126	6	2	1	1	2
127	0	0	1	3	2
128	3	1	0	2	2
129	6	2	0	1	2
130	6	0	0	1	2
131	2	2	1	1	2
132	3	4	0	2	2
133	5	3	0	1	2
134	4	4	1	2	2
135	5	3	1	2	2
136	7	2	1	1	2
137	0	0	1	2	2
138	6	0	0	2	2
139	5	0	0	2	2
140	3	0	1	2	2
141	6	4	0	2	2
142	1	0	1	1	2
143	4	6	0	3	2
144	2	0	0	1	2
145	8	4	1	1	2
146	4	3	0	2	2
147	2	1	1	3	2
148	1	1	0	1	2
149	3	2	0	2	2
150	0	0	1	2	2
151	4	1	1	1	2
152	0	0	1	3	2
153	3	4	1	2	2
154	2	1	0	1	2
155	5	4	0	1	2
156	1	0	1	3	2
157	2	0	0	2	2
158	2	1	1	2	2
159	0	0	1	2	2
160	0	1	0	2	2
161	0	0	0	1	2
162	4	0	1	3	2
163	8	0	0	2	2
164	4	1	0	2	2
165	6	1	0	1	2
166	4	0	0	2	2
167	1	0	1	1	2
168	0	0	1	2	2
169	7	3	1	3	2
170	6	1	1	1	2
171	5	2	0	1	2
172	2	1	1	2	2
173	3	1	0	1	2
174	2	2	0	1	2
175	0	0	0	2	2
176	8	0	0	2	2
177	0	0	0	1	2
178	0	1	1	1	2
179	0	0	1	2	2
180	0	0	1	1	2
181	2	0	1	2	2
182	0	1	1	1	2
183	2	2	1	1	2
184	2	1	0	1	2
185	3	1	0	1	2
186	0	0	0	1	2
187	0	0	0	3	2
188	2	0	0	2	2
189	1	1	1	1	2
190	6	0	1	3	2
191	7	5	1	1	2
192	0	0	0	2	2
193	0	0	0	1	2
194	2	1	0	3	2
195	2	2	0	3	2

observation	DMFT1	DMFT2	Gender	Ethnic	School
196	5	1	1	2	2
197	4	3	1	2	2
198	0	0	0	3	2
199	7	3	0	2	2
200	0	0	1	3	2
201	0	3	1	1	2
202	8	1	1	2	2
203	1	0	0	3	2
204	6	1	0	2	2
205	3	1	0	2	2
206	0	2	0	2	2
207	6	0	0	1	2
208	3	2	0	3	2
209	0	0	0	1	2
210	0	1	1	1	2
211	8	4	1	1	2
212	6	1	0	3	2
213	2	4	0	2	2
214	0	0	1	1	2
215	3	3	0	2	2
216	3	0	1	2	2
217	6	2	0	2	2
218	0	0	1	1	2
219	7	5	0	1	2
220	0	0	1	1	2
221	5	1	1	1	2
222	2	4	1	2	2
223	4	4	0	2	2
224	1	1	0	1	2
225	0	3	1	2	2
226	6	2	1	1	2
227	3	0	0	2	2
228	0	0	0	2	2
229	2	2	1	2	2
230	0	0	1	2	2
231	0	0	1	2	2
232	0	1	1	1	2
233	0	1	1	1	2
234	0	0	1	2	2
235	3	0	1	2	2
236	8	3	1	3	2
237	0	2	0	2	2
238	5	1	0	2	2
239	7	2	1	1	2
240	0	2	1	1	2
241	4	0	0	1	2
242	2	0	0	2	2
243	3	2	0	1	2
244	0	0	1	1	2
245	7	5	1	2	2
246	0	0	1	1	2
247	0	0	1	3	2
248	1	0	1	1	2
249	0	0	1	1	2
250	3	4	1	2	2
251	4	2	1	1	2
252	7	2	1	3	3
253	5	4	0	1	3
254	4	3	1	1	3
255	5	4	0	1	3
256	4	1	0	3	3
257	6	3	0	2	3
258	3	0	1	2	3
259	6	3	1	2	3
260	2	2	0	3	3
261	2	2	0	1	3
262	3	1	1	3	3
263	5	2	1	2	3
264	4	2	0	1	3
265	5	3	1	2	3

observation	DMFT1	DMFT2	Gender	Ethnic	School
266	5	1	1	2	3
267	3	2	1	1	3
268	1	0	0	1	3
269	4	2	0	1	3
270	4	1	1	3	3
271	2	2	1	1	3
272	4	1	1	1	3
273	8	3	1	1	3
274	4	3	1	3	3
275	2	1	1	1	3
276	1	2	0	3	3
277	5	4	1	2	3
278	0	0	0	3	3
279	3	0	0	1	3
280	5	3	1	1	3
281	0	0	0	1	3
282	6	4	0	2	3
283	3	2	1	2	3
284	3	3	1	2	3
285	6	3	1	2	3
286	5	3	1	1	3
287	5	4	0	1	3
288	1	0	1	2	3
289	0	0	0	2	3
290	6	5	1	1	3
291	2	3	0	3	3
292	0	0	1	3	3
293	0	0	1	3	3
294	8	4	1	2	3
295	4	5	1	2	3
296	7	1	1	2	3
297	6	6	1	1	3
298	6	5	1	2	3
299	4	5	0	3	3
300	8	6	1	2	3
301	8	6	1	1	3
302	1	1	0	2	3
303	4	2	1	3	3
304	7	4	0	1	3
305	1	1	0	1	3
306	1	2	0	2	3
307	7	5	1	2	3
308	3	0	0	2	3
309	6	2	0	1	3
310	2	3	0	2	3
311	0	0	0	3	3
312	5	3	1	3	3
313	2	2	1	1	3
314	0	0	1	1	3
315	0	0	1	2	3
316	5	4	1	3	3
317	3	3	1	1	3
318	2	2	0	1	3
319	0	0	1	1	3
320	4	1	1	1	3
321	1	0	0	1	3
322	2	0	0	1	3
323	8	6	1	3	3
324	2	1	1	1	3
325	0	1	1	1	3
326	5	5	1	1	3
327	7	6	0	2	3
328	1	1	0	2	3
329	0	0	0	2	3
330	8	6	0	3	3
331	0	1	1	1	3
332	5	3	1	1	3
333	4	1	1	2	3
334	4	5	0	1	3
335	3	0	0	2	3

observation	DMFT1	DMFT2	Gender	Ethnic	School
336	5	1	0	1	3
337	5	5	0	2	3
338	5	4	0	2	3
339	0	1	0	1	3
340	6	4	0	1	3
341	0	0	0	3	3
342	8	3	1	2	3
343	8	6	1	2	3
344	0	2	1	1	3
345	5	5	1	3	3
346	7	5	1	2	3
347	6	5	0	1	3
348	0	0	0	1	3
349	7	4	0	2	3
350	2	2	1	1	3
351	6	5	1	1	3
352	3	2	1	1	3
353	5	1	0	2	3
354	7	3	1	1	3
355	7	1	0	3	3
356	6	3	0	1	3
357	3	0	1	2	3
358	1	1	0	1	3
359	3	2	0	2	3
360	5	6	1	2	3
361	6	3	1	1	3
362	2	0	1	1	3
363	7	3	1	2	3
364	4	4	1	1	3
365	2	2	1	3	3
366	4	0	1	1	3
367	5	3	0	1	3
368	1	1	1	1	3
369	3	1	1	2	3
370	1	1	1	2	3
371	2	0	1	2	3
372	3	5	1	1	3
373	8	6	1	1	3
374	4	3	1	2	3
375	2	2	0	2	3
376	1	3	0	2	3
377	4	2	0	2	3
378	4	0	0	2	3
379	0	0	0	1	3
380	0	1	0	2	3
381	6	3	0	2	3
382	4	2	0	1	3
383	8	3	1	3	3
384	1	3	1	1	3
385	3	1	1	1	3
386	2	1	1	1	3
387	3	2	1	1	3
388	3	2	1	2	4
389	1	2	1	2	4
390	6	2	0	2	4
391	8	2	1	1	4
392	7	0	1	1	4
393	7	3	1	2	4
394	5	1	1	1	4
395	3	2	0	2	4
396	0	5	0	1	4
397	3	4	1	3	4
398	6	3	1	2	4
399	0	0	0	2	4
400	5	1	1	2	4
401	7	2	1	2	4
402	1	1	0	2	4
403	7	5	1	1	4
404	7	6	1	1	4
405	7	4	1	1	4

observation	DMFT1	DMFT2	Gender	Ethnic	School
406	6	4	0	2	4
407	2	3	0	2	4
408	5	3	0	2	4
409	0	0	1	3	4
410	0	1	1	3	4
411	0	2	0	3	4
412	0	0	1	1	4
413	7	5	1	2	4
414	4	3	0	1	4
415	0	1	0	2	4
416	1	1	1	3	4
417	1	0	0	3	4
418	1	1	1	2	4
419	0	0	0	1	4
420	0	0	1	2	4
421	2	5	1	2	4
422	4	1	1	2	4
423	5	3	1	1	4
424	6	3	0	1	4
425	0	0	0	1	4
426	3	1	0	2	4
427	2	2	0	1	4
428	1	2	0	3	4
429	0	0	0	1	4
430	3	2	0	2	4
431	4	3	1	1	4
432	7	5	1	1	4
433	3	1	1	2	4
434	4	4	0	2	4
435	0	0	0	1	4
436	5	3	0	2	4
437	1	1	0	2	4
438	7	5	1	2	4
439	4	3	1	1	4
440	8	2	1	1	4
441	0	0	0	2	4
442	8	6	0	2	4
443	3	0	0	1	4
444	4	4	0	3	4
445	1	2	0	2	4
446	0	1	0	2	4
447	8	5	1	3	4
448	8	3	1	1	4
449	1	0	1	1	4
450	2	4	0	2	4
451	6	5	1	1	4
452	0	0	1	2	4
453	8	6	1	2	4
454	5	0	0	1	4
455	1	1	0	1	4
456	0	2	0	1	4
457	1	0	1	1	4
458	2	3	0	2	4
459	4	3	0	2	4
460	0	0	1	1	4
461	2	4	1	2	4
462	0	3	0	2	4
463	4	3	0	1	4
464	6	4	0	1	4
465	4	2	1	1	4
466	7	1	1	1	4
467	0	0	0	1	4
468	3	1	0	1	4
469	0	4	0	1	4
470	3	3	1	1	4
471	8	5	0	3	4
472	2	2	1	1	4
473	6	3	0	2	4
474	8	5	0	1	4
475	3	1	0	1	4

observation	DMFT1	DMFT2	Gender	Ethnic	School
476	5	3	1	2	4
477	2	0	0	1	4
478	0	1	0	2	4
479	4	1	1	3	4
480	3	3	0	1	4
481	5	5	1	1	4
482	6	4	1	1	4
483	0	0	1	1	4
484	6	0	1	1	4
485	8	3	1	1	4
486	4	1	0	2	4
487	1	1	1	1	4
488	0	0	1	1	4
489	0	1	0	1	4
490	5	2	1	1	4
491	0	0	0	2	4
492	1	2	0	3	4
493	3	0	0	2	4
494	0	0	1	1	4
495	0	2	0	3	4
496	4	2	0	2	4
497	6	3	1	1	4
498	5	3	1	2	4
499	5	3	1	2	4
500	1	3	0	1	4
501	1	1	1	1	4
502	5	4	0	2	4
503	0	1	1	3	4
504	6	2	1	1	4
505	5	0	0	2	4
506	2	2	1	1	4
507	1	0	0	2	4
508	8	5	0	2	4
509	4	3	0	2	4
510	0	0	0	1	4
511	5	5	0	2	4
512	3	0	0	1	4
513	0	0	1	1	4
514	0	1	1	1	4
515	8	5	1	2	4
516	6	4	1	1	4
517	4	2	1	1	4
518	3	2	1	1	4
519	4	3	1	2	4
520	7	5	0	2	5
521	5	4	1	1	5
522	2	1	1	1	5
523	0	1	0	1	5
524	4	0	1	1	5
525	0	0	1	1	5
526	0	0	1	3	5
527	7	3	1	2	5
528	0	0	0	3	5
529	6	2	0	2	5
530	2	1	1	3	5
531	1	0	0	2	5
532	0	0	0	1	5
533	8	1	0	2	5
534	3	2	0	2	5
535	3	3	1	1	5
536	4	3	1	2	5
537	0	0	1	1	5
538	5	3	1	2	5
539	7	6	1	1	5
540	1	1	0	3	5
541	5	4	0	2	5
542	5	2	0	1	5
543	2	0	0	2	5
544	1	0	0	1	5
545	5	2	0	1	5

observation	DMFT1	DMFT2	Gender	Ethnic	School
546	1	1	1	2	5
547	6	3	1	2	5
548	3	3	1	1	5
549	6	0	0	3	5
550	7	3	0	2	5
551	6	3	1	2	5
552	4	4	1	2	5
553	3	4	0	1	5
554	7	4	0	1	5
555	8	5	1	2	5
556	0	0	1	1	5
557	3	2	1	2	5
558	1	0	0	3	5
559	4	3	0	1	5
560	4	2	1	2	5
561	4	4	0	2	5
562	5	2	1	2	5
563	0	0	0	3	5
564	5	1	1	1	5
565	3	1	1	1	5
566	8	2	0	2	5
567	1	3	1	2	5
568	0	0	0	2	5
569	8	2	1	2	5
570	0	2	0	1	5
571	0	0	0	2	5
572	2	1	0	2	5
573	1	2	1	2	5
574	5	3	0	2	5
575	5	4	1	1	5
576	0	0	0	2	5
577	0	0	0	2	5
578	5	0	1	1	5
579	0	0	1	1	5
580	7	5	1	1	5
581	2	2	0	2	5
582	4	3	0	1	5
583	0	0	0	1	5
584	0	0	1	1	5
585	3	1	0	2	5
586	0	0	1	2	5
587	7	0	1	2	5
588	6	5	1	2	5
589	4	4	1	2	5
590	1	0	0	1	5
591	5	1	0	3	5
592	5	1	0	2	5
593	5	4	0	2	5
594	5	5	1	2	5
595	7	1	0	2	5
596	0	0	0	1	5
597	4	1	0	1	5
598	8	5	0	3	5
599	0	0	1	2	5
600	8	2	0	1	5
601	0	0	1	3	5
602	0	0	1	2	5
603	4	2	1	2	5
604	1	0	1	3	5
605	6	2	0	1	5

observation	DMFT1	DMFT2	Gender	Ethnic	School
606	6	1	0	2	5
607	5	3	0	3	5
608	4	3	0	1	5
609	6	3	1	3	5
610	5	2	0	1	5
611	7	2	0	1	5
612	0	2	1	1	5
613	3	1	0	1	5
614	7	3	1	1	5
615	6	5	0	2	5
616	2	0	0	3	5
617	0	1	0	2	5
618	7	5	0	1	5
619	8	1	1	3	5
620	0	0	1	2	5
621	2	1	0	1	5
622	5	1	0	2	5
623	5	4	1	2	5
624	0	0	1	2	5
625	4	0	0	1	5
626	6	4	1	1	5
627	0	0	0	2	5
628	2	2	0	3	5
629	6	0	0	2	5
630	0	0	0	2	5
631	6	3	0	1	5
632	5	4	0	1	5
633	6	4	1	1	5
634	0	0	1	3	5
635	5	3	1	1	5
636	2	2	0	2	5
637	2	4	0	1	5
638	0	0	1	3	5
639	1	0	1	1	5
640	3	2	0	1	5
641	0	0	0	3	5
642	6	1	0	1	5
643	5	1	1	3	5
644	0	0	1	2	5
645	7	5	1	2	5
646	3	0	0	1	5
647	8	5	0	2	5
648	4	1	0	2	5
649	4	3	0	2	5
650	6	2	1	1	5
651	0	0	0	1	5
652	6	0	0	2	5
653	0	0	1	2	5
654	0	0	1	2	5
655	0	0	0	1	5
656	8	0	0	2	5
657	7	1	0	1	5
658	0	0	0	1	5
659	1	2	1	2	5
660	0	0	1	2	5
661	1	0	0	1	5
662	6	4	0	1	5
663	3	0	0	3	5
664	1	0	0	2	5
665	1	1	0	2	5
666	0	0	0	2	5
667	4	3	1	1	5
668	0	0	1	2	5
669	1	1	1	2	5
670	2	2	1	1	5
671	8	5	1	1	5
672	5	0	1	2	5
673	1	1	0	2	5
674	1	0	0	2	5
675	1	1	0	2	6

observation	DMFT1	DMFT2	Gender	Ethnic	School
676	2	3	0	2	6
677	0	0	0	2	6
678	4	0	0	1	6
679	2	0	0	2	6
680	0	0	0	1	6
681	4	4	0	1	6
682	6	5	1	1	6
683	2	0	1	2	6
684	1	0	1	3	6
685	4	1	1	2	6
686	6	6	0	1	6
687	3	3	0	2	6
688	0	1	0	1	6
689	6	2	1	2	6
690	5	3	0	2	6
691	8	5	1	2	6
692	0	0	1	2	6
693	2	1	1	2	6
694	7	6	0	2	6
695	2	0	1	2	6
696	4	2	1	2	6
697	7	5	1	2	6
698	7	5	1	1	6
699	1	0	0	1	6
700	1	6	0	2	6
701	1	0	0	2	6
702	3	1	1	1	6
703	4	1	0	1	6
704	0	0	1	3	6
705	5	5	1	2	6
706	6	3	1	2	6
707	4	1	0	2	6
708	0	1	1	2	6
709	7	6	0	2	6
710	3	4	1	2	6
711	2	1	0	2	6
712	2	0	0	1	6
713	4	0	0	2	6
714	1	0	0	1	6
715	0	1	0	2	6
716	1	1	0	2	6
717	2	2	1	2	6
718	0	0	0	2	6
719	2	2	0	2	6
720	2	1	1	1	6
721	5	2	0	2	6
722	2	3	0	2	6
723	4	4	1	2	6
724	3	0	1	1	6
725	4	2	0	2	6
726	4	4	0	2	6
727	3	1	0	2	6
728	2	1	1	2	6
729	1	1	1	2	6
730	0	0	0	2	6
731	7	0	0	3	6
732	3	1	0	3	6
733	0	1	1	2	6
734	2	2	1	2	6
735	0	2	1	3	6
736	6	5	1	2	6
737	0	0	1	1	6
738	5	6	0	2	6
739	8	6	0	2	6
740	3	2	0	1	6
741	1	0	0	2	6
742	5	3	1	1	6
743	7	4	0	1	6
744	2	1	1	1	6
745	6	4	1	2	6

observation	DMFT1	DMFT2	Gender	Ethnic	School
746	2	1	1	2	6
747	7	3	0	1	6
748	0	0	1	1	6
749	2	2	1	2	6
750	3	2	0	1	6
751	0	0	0	2	6
752	3	4	0	2	6
753	0	1	1	2	6
754	6	5	1	2	6
755	4	3	0	2	6
756	0	0	0	2	6
757	0	0	1	1	6
758	7	2	1	3	6
759	3	2	0	2	6
760	1	3	1	2	6
761	1	3	0	3	6
762	6	2	1	2	6
763	6	3	0	2	6
764	2	1	1	2	6
765	0	0	1	2	6
766	7	2	1	2	6
767	1	0	1	1	6
768	0	1	0	1	6
769	7	1	1	2	6
770	3	2	1	2	6
771	6	2	1	2	6
772	4	0	0	2	6
773	7	1	1	2	6
774	2	2	1	1	6
775	0	0	0	2	6
776	0	0	1	2	6
777	4	3	1	1	6
778	0	0	1	1	6
779	6	2	0	2	6
780	2	0	1	2	6
781	6	2	1	2	6
782	1	0	1	1	6
783	3	0	1	3	6
784	7	6	1	1	6
785	0	0	0	2	6
786	6	2	1	2	6
787	0	0	0	2	6
788	3	5	0	2	6
789	0	1	0	2	6
790	2	2	0	2	6
791	2	1	1	2	6
792	0	1	1	3	6
793	7	3	1	2	6
794	0	0	1	2	6
795	2	1	1	2	6
796	0	0	0	1	6
797	2	1	1	2	6

Appendix A

A.1. Derivation of the score statistic S

Reparameterize ψ_i and λ_i, $i = 1, ..., k$, under H_1, by $\psi_i = \psi + \alpha_i, i = 1, ..., k$, with $\alpha_k = 0$ and $\lambda_i = \lambda + \beta_i, i = 1, ..., k$, with $\beta_k = 0$. Then testing H_0 is equivalent to testing $\alpha = 0$ and $\beta = 0$ with $\omega = (\psi, \lambda)$ as nuisance parameters.

Now, let

$$s = \left.\begin{pmatrix} \dfrac{\partial l^*}{\partial \alpha} \\ \dfrac{\partial l^*}{\partial \beta} \end{pmatrix}\right|_{\alpha=0,\beta=0},$$

$$A = \begin{pmatrix} E\left(-\left.\dfrac{\partial^2 l^*}{\partial \alpha \partial \alpha'}\right|_{\alpha=0,\beta=0}\right) & E\left(-\left.\dfrac{\partial^2 l^*}{\partial \alpha \partial \beta'}\right|_{\alpha=0,\beta=0}\right) \\ \\ E\left(-\left.\dfrac{\partial^2 l^*}{\partial \beta \partial \alpha'}\right|_{\alpha=0,\beta=0}\right) & E\left(-\left.\dfrac{\partial^2 l^*}{\partial \beta \partial \beta'}\right|_{\alpha=0,\beta=0}\right) \end{pmatrix},$$

$$C = \begin{pmatrix} E\left(-\left.\dfrac{\partial^2 l^*}{\partial \alpha \partial \omega'}\right|_{\alpha=0,\beta=0}\right) \\ \\ E\left(-\left.\dfrac{\partial^2 l^*}{\partial \beta \partial \omega'}\right|_{\alpha=0,\beta=0}\right) \end{pmatrix},$$

and

$$D = E\left(-\frac{\partial^2 l^*}{\partial \omega \partial \omega'}\bigg|_{\alpha=0,\beta=0} \right).$$

If we use the maximum likelihood estimate $\hat{\omega}$ of the nuisance parameter ω under the null hypothesis H_0 in s, A, C and D, then the score test for testing H_0 against H_1 is

$$S = \hat{s}' \left(\hat{A} - \hat{C}\hat{D}^{-1}\hat{C}' \right)^{-1} \hat{s}.$$

As in Section 3.3, note that parameters (α, ψ) are orthogonal with parameters (β, λ). Then, by using the notations s_{0i}, A_{0i}, C_{0i}, $i = 1, 2$ and D, the expressions for s, A, D and C can be simplified as

$$A = \begin{pmatrix} A_{01} & 0 \\ 0 & A_{02} \end{pmatrix},$$

where,

$$A_{01} = diag\left[E\left(-\frac{\partial^2 l^*}{\partial \alpha_1^2}\bigg|_{\alpha=0,\beta=0} \right), E\left(-\frac{\partial^2 l^*}{\partial \alpha_2^2}\bigg|_{\alpha=0,\beta=0} \right), \cdots, E\left(-\frac{\partial^2 l^*}{\partial \alpha_{k-1}^2}\bigg|_{\alpha=0,\beta=0} \right) \right],$$

and

$$A_{02} = diag\left[E\left(-\frac{\partial^2 l^*}{\partial \beta_1^2}\bigg|_{\alpha=0,\beta=0} \right), E\left(-\frac{\partial^2 l^*}{\partial \beta_2^2}\bigg|_{\alpha=0,\beta=0} \right), \cdots, E\left(-\frac{\partial^2 l^*}{\partial \beta_{k-1}^2}\bigg|_{\alpha=0,\beta=0} \right) \right],$$

and

$$C = \begin{pmatrix} C_{01} \\ C_{02} \end{pmatrix},$$

where $C_{01} = \begin{pmatrix} E\left(-\frac{\partial^2 l^*}{\partial\alpha_1\partial\psi}\bigg|_{\alpha=0,\beta=0}\right) & 0 \\ E\left(-\frac{\partial^2 l^*}{\partial\alpha_2\partial\psi}\bigg|_{\alpha=0,\beta=0}\right) & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ E\left(-\frac{\partial^2 l^*}{\partial\alpha_{k-1}\partial\psi}\bigg|_{\alpha=0,\beta=0}\right) & 0 \end{pmatrix}$ and $C_{02} = \begin{pmatrix} 0 & E\left(-\frac{\partial^2 l^*}{\partial\beta_1\partial\lambda}\bigg|_{\alpha=0,\beta=0}\right) \\ 0 & E\left(-\frac{\partial^2 l^*}{\partial\beta_2\partial\lambda}\bigg|_{\alpha=0,\beta=0}\right) \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & E\left(-\frac{\partial^2 l^*}{\partial\beta_{k-1}\partial\lambda}\bigg|_{\alpha=0,\beta=0}\right) \end{pmatrix}$

and

$$D = \begin{pmatrix} E\left(-\frac{\partial^2 l^*}{\partial\psi^2}\bigg|_{\alpha=0,\beta=0}\right) & 0 \\ 0 & E\left(-\frac{\partial^2 l^*}{\partial\lambda^2}\bigg|_{\alpha=0,\beta=0}\right) \end{pmatrix}.$$

Then the matrix $A - CD^{-1}C'$ can be simplified as

$$A - CD^{-1}C = \begin{pmatrix} A_{01} - C_{01}C_{01}'/E\left(-\frac{\partial^2 l^*}{\partial\psi^2}\bigg|_{\alpha=0,\beta=0}\right) & 0 \\ 0 & A_{02} - C_{02}C_{02}'/E\left(-\frac{\partial^2 l^*}{\partial\lambda^2}\bigg|_{\alpha=0,\beta=0}\right) \end{pmatrix}.$$

Also let $s_{01} = \frac{\partial l^*}{\partial\alpha}\bigg|_{\alpha=0,\beta=0}$ and $s_{02} = \frac{\partial l^*}{\partial\beta}\bigg|_{\alpha=0,\beta=0}$. Then the score test statistic for testing H_0 against H_1 can be written as

$$S = S_{01} + S_{02},$$

where

$$S_{01} = \hat{s}_{01}'\left[\hat{A}_{01} - \hat{C}_{01}\hat{C}_{01}'/E\left(-\frac{\partial^2 l^*}{\partial\psi^2}\bigg|_{\alpha=0,\beta=0}\right)\right]^{-1}\hat{s}_{01}$$

and

$$S_{02} = \hat{s}_{02}' \left[\hat{A}_{02} - \hat{C}_{02}\hat{C}_{02}'/E\left(-\frac{\partial^2 l^*}{\partial \lambda^2}\bigg|_{\alpha=0,\beta=0} \right) \right]^{-1} \hat{s}_{02}.$$

Similarly as in the proofs for S_1 and S_2, we have

$$E\left(-\frac{\partial^2 l^*}{\partial \alpha_i \partial \psi}\bigg|_{\alpha=0,\beta=0} \right) = E\left(-\frac{\partial^2 l_i^*}{\partial \psi^2}\bigg|_{\alpha=0,\beta=0} \right), i = 1, ..., k-1,$$

$$E\left(-\frac{\partial^2 l^*}{\partial \beta_i \partial \lambda}\bigg|_{\alpha=0,\beta=0} \right) = E\left(-\frac{\partial^2 l_i^*}{\partial \lambda^2}\bigg|_{\alpha=0,\beta=0} \right), i = 1, ..., k-1,$$

$$E\left(-\frac{\partial^2 l^*}{\partial \psi^2}\bigg|_{\alpha=0,\beta=0} \right) = \sum_{i=1}^k E\left(-\frac{\partial^2 l_i^*}{\partial \psi^2}\bigg|_{\alpha=0,\beta=0} \right),$$

and

$$E\left(-\frac{\partial^2 l^*}{\partial \lambda^2}\bigg|_{\alpha=0,\beta=0} \right) = \sum_{i=1}^k E\left(-\frac{\partial^2 l_i^*}{\partial \lambda^2}\bigg|_{\alpha=0,\beta=0} \right).$$

We can obtain the score test statistic S for testing H_0 against H_1 as

$$S = S_{01} + S_{02},$$

where $S_{01} = \sum_{i=1}^k \frac{\hat{s}_{01i}^2}{\hat{v}_{01i}}$ and $S_{02} = \sum_{i=1}^k \frac{\hat{s}_{02i}^2}{\hat{v}_{02i}}$ with the estimated values of $s_{01i} = \frac{\partial l_i^*}{\partial \psi}\Big|_{H_0}$, $v_{01i} = E\left(-\frac{\partial^2 l_i^*}{\partial \psi^2}\Big|_{H_0} \right)$, $s_{02i} = \frac{\partial l_i^*}{\partial \lambda}\Big|_{H_0}$ and v_{02i}=$E\left(-\frac{\partial^2 l_i^*}{\partial \lambda^2}\Big|_{H_0} \right)$, $i = 1, ..., k$, respectively.

A.2. Derivations of the s_{1i}, v_{1i}, s_{2i} and v_{2i}, $i = 1, ..., k$, in terms of the original parameters

For this, first, we need to express the first derivatives of the log-likelihood function l_i^*, $i = 1, ..., k$, with respect to the orthogal parameters ψ_i and λ_i in terms of the original parameters ψ_i and ϕ_i of the log-likelihood function l_i, $i = 1, ..., k$. We have

$$\begin{aligned} \frac{\partial l_i^*}{\partial \psi_i} &= \frac{\partial l_i}{\partial \psi_i} + \frac{\partial l_i}{\partial \phi_i}\frac{\partial \phi_i}{\partial \psi_i}, i = 1, ..., k, \\ \frac{\partial l_i^*}{\partial \lambda_i} &= \frac{\partial l_i}{\partial \phi_i}\frac{\partial \phi_i}{\partial \lambda_i}, i = 1, ..., k. \end{aligned}$$

Note that $E\left(-\frac{\partial^2 l_i^*}{\partial \psi_i^2}\right) = E\left(\frac{\partial l_i^*}{\partial \psi_i}\right)^2$, $E\left(-\frac{\partial^2 l_i^*}{\partial \lambda_i^2}\right) = E\left(\frac{\partial l_i}{\partial \lambda_i}\right)^2$ and $E\left(-\frac{\partial^2 l_i^*}{\partial \psi_i \partial \lambda_i}\right) = E\left(\frac{\partial l_i}{\partial \psi_i}\frac{\partial l_i^*}{\partial \lambda_i}\right)$.

So we can obtain

$$\begin{aligned}
E\left(\frac{\partial l_i^*}{\partial \psi_i}\right)^2 &= E\left(\frac{\partial l_i}{\partial \psi_i} + \frac{\partial l_i}{\partial \phi_i}\frac{\partial \phi_i}{\partial \psi_i}\right)^2 \\
&= E\left(\frac{\partial l_i}{\partial \psi_i}\right)^2 + 2\frac{\partial \phi_i}{\partial \psi_i} E\left(\frac{\partial l_i}{\partial \psi_i}\frac{\partial l_i}{\partial \phi_i}\right) + \left(\frac{\partial \phi_i}{\partial \psi_i}\right)^2 E\left(\frac{\partial l_i}{\partial \psi_i}\right)^2 \\
&= i_{\psi_i\psi_i} + 2 i_{\psi_i\phi_i}\frac{\partial \phi_i}{\partial \psi_i} + \left(\frac{\partial \phi_i}{\partial \psi_i}\right)^2 i_{\phi_i\phi_i}, \\
E\left(\frac{\partial l_i^*}{\partial \lambda_i}\right)^2 &= \left(\frac{\partial \phi_i}{\partial \lambda_i}\right)^2 E\left(\frac{\partial l_i}{\partial \phi_i}^2\right) = \left(\frac{\partial \phi_i}{\partial \lambda_i}\right)^2 i_{\phi_i\phi_i}.
\end{aligned}$$

From (3.1), replacing $\frac{\partial \phi_i}{\partial \psi_i}$ by $-i_{\psi_i\phi}/i_{\phi_i\phi_i}$, we can obtain

$$s_{1i} = \left(\frac{\partial l_i}{\partial \psi_i} - \frac{\partial l_i}{\partial \phi_i}\frac{i_{\psi_i\phi_i}}{i_{\phi_i\phi_i}}\right),$$

$$v_{1i} = \left(i_{\psi_i\psi_i} - i^2_{\psi_i\phi_i}/i_{\phi_i\phi_i}\right),$$

$$s_{2i} = \left(\frac{\partial l_i}{\partial \phi_i}\right)\frac{\partial \phi_i}{\partial \lambda_i},$$

and

$$v_{2i} = \left(\frac{\partial \phi_i}{\partial \lambda_i}\right)^2 i_{\phi_i\phi_i}.$$

Note that, generally, we cannot find the quantities $\frac{\partial \phi_i}{\partial \lambda_i}$, $i = 1, ..., k$. However, when we calculate the score statistics S_2, the denominator and numerator have this common factor, which cancel out. Thus, the quantities s_{1i}, v_{1i}, s_{2i} and v_{2i}, $i = 1, ..., k$, in terms of the original parameters, are

$$s_{1i} = \left(\frac{\partial l_i}{\partial \psi_i} - \frac{\partial l_i}{\partial \phi_i}\frac{i_{\psi_i\phi_i}}{i_{\phi_i\phi_i}}\right),$$

$$v_{1i} = \left(i_{\psi_i\psi_i} - i^2_{\psi_i\phi_i}/i_{\phi_i\phi_i}\right),$$

$$s_{2i} = \frac{\partial l_i}{\partial \phi_i} \text{ and } v_{2i} = i_{\phi_i\phi_i}.$$

Appendix B: Derivation for score test

In what follows we derive the score test statistic S_1 for testing

$H_0 : \pi = 1/2$ against $H_1 : \pi \neq 1/2$ when γ and ϕ are treated as nuisance parameters.

The statistic S_2 can be obtained following similar steps.

Consider the log-likelihood l given in Section 6.3.2. Now, Define

$$\Psi_1 = \frac{\partial l}{\partial \pi}\bigg|_{H_0}.$$

The asymptotic variance of Ψ_1 is

$$V_1^2 = I_{\pi\pi} - \frac{I_{\phi\phi} I_{\pi\gamma}^2 + I_{\gamma\gamma} I_{\pi\phi}^2 - 2 I_{\pi\gamma} I_{\pi\phi} I_{\gamma\phi}}{I_{\phi\phi} I_{\gamma\gamma} - I_{\phi\gamma}^2},$$

where $I_{\pi\pi} = E\{-\frac{\partial^2 l}{\partial \pi^2}\}\bigg|_{H_0}$, $I_{\pi\phi} = E\{-\frac{\partial^2 l}{\partial \pi \partial \phi}\}\bigg|_{H_0}$, $I_{\phi\phi} = E\{-\frac{\partial^2 l}{\partial \phi^2}\}\bigg|_{H_0}$, $I_{\pi\gamma} = E\{-\frac{\partial^2 l}{\partial \pi \partial \gamma}\}\bigg|_{H_0}$, $I_{\pi\phi} = E\{-\frac{\partial^2 l}{\partial \pi \partial \phi}\}\bigg|_{H_0}$ and $I_{\gamma\phi} = E\{-\frac{\partial^2 l}{\partial \gamma \partial \phi}\}\bigg|_{H_0}$.

Then, it can be shown that (Neyman, 1966) asymptotically, as $n \to \infty$, the distribution of $S_1 = {\Psi_1}^2/{V_1}^2$ is chi-square with 1 degree of freedom. If the nuisance parameters γ and ϕ are replaced by their maximum likelihood estimates $\hat{\gamma}$ and $\hat{\phi}$, which are $\sqrt{n}$-consistent, in Ψ_1 and V_1, then, asymptotically, as $n \to \infty$, $S_1 = \hat{\Psi_1}^2/\hat{V_1}^2$ is $\chi^2(1)$.

We now evaluate the score function Ψ_1 and the elements of the variance V_1^2, namely, the quantities $I_{\pi\pi}$, $I_{\pi\phi}$, etc.

Define the functions

$l_{0i} = \log(f_{0i})$ and $l_{y_i} = \log(f_{y_i})$.

Then, it can be seen that

$$l_{0i} = \sum_{r=1}^{m_i} \log[(1-\pi)(1-\phi) + (r-1)\phi] - \sum_{r=1}^{m_i} \log[1-\phi+(r-1)\phi],$$

and

$$\begin{aligned} l_{yi} &= \log \begin{pmatrix} m_i \\ y_i \end{pmatrix} + \sum_{r=1}^{y_i} \log[\pi(1-\phi)+(r-1)\phi] + \sum_{r=1}^{m_i-y_i} \log[(1-\pi)(1-\phi)+(r-1)\phi] \\ &\quad - \sum_{r=1}^{m_i} \log[1-\phi+(r-1)\phi]. \end{aligned}$$

Further, using these in the log-likelihood l and taking its derivative with respect to π we obtain

$$\Psi_1 = \frac{\partial l}{\partial \pi} = \sum_{i=1}^{n}\{I_{\{y_i=0\}} \frac{f_{0i}l'_{0i(\pi)}}{\gamma + f_{0i}} + I_{\{y_i>0\}} l'_{y_i(\pi)}\}.$$

Further, we see that

$$\begin{aligned} I_{\pi\pi} &= E\{-\frac{\partial^2 l}{\partial \pi^2}\} = E\{\sum_{i=1}^{n}(I_{\{y_i=0\}} \frac{-\frac{\partial^2 l_{0i}}{\partial \pi^2} f_{0i}(\gamma+f_{0i}) - \gamma f_{0i}(\frac{\partial l_{0i}}{\partial \pi})^2}{(\gamma+f_{0i})^2} - I_{\{y_i>0\}} \frac{\partial^2 l_{yi}}{\partial \pi^2})\}, \\ I_{\pi\phi} &= E\{-\frac{\partial^2 l}{\partial \pi \partial \phi}\} = E\{\sum_{i=1}^{n}(I_{\{y_i=0\}} \frac{-\frac{\partial^2 l_{0i}}{\partial \pi \partial \phi} f_{0i}(\gamma+f_{0i}) - \gamma f_{0i}\frac{\partial l_{0i}}{\partial \pi}\frac{\partial l_{0i}}{\partial \phi}}{(\gamma+f_{0i})^2} - I_{\{y_i>0\}} \frac{\partial^2 l_{yi}}{\partial \pi \partial \phi})\}, \\ I_{\phi\phi} &= E\{-\frac{\partial^2 l}{\partial \phi^2}\} = E\{\sum_{i=1}^{n}(I_{\{y_i=0\}} \frac{-\frac{\partial^2 l_{0i}}{\partial \phi^2} f_{0i}(\gamma+f_{0i}) - \gamma f_{0i}(\frac{\partial l_{0i}}{\partial \phi})^2}{(\gamma+f_{0i})^2} - I_{\{y_i>0\}} \frac{\partial^2 l_{yi}}{\partial \phi^2})\}, \end{aligned}$$

$$\begin{aligned} I_{\pi\gamma} &= E\{-\frac{\partial^2 l}{\partial \pi \partial \gamma}\} = E\{\sum_{i=1}^{n} I_{\{y_i=0\}} \frac{\frac{\partial l_{0i}}{\partial \pi} f_{0i}}{(\gamma+f_{0i})^2}\}, \\ I_{\phi\gamma} &= E\{-\frac{\partial^2 l}{\partial \phi \partial \gamma}\} = E\{\sum_{i=1}^{n} I_{\{y_i=0\}} \frac{\frac{\partial l_{0i}}{\partial \phi} f_{0i}}{(\gamma+f_{0i})^2}\}, \end{aligned}$$

and

$$I_{\gamma\gamma} = E\{-\frac{\partial^2 l}{\partial \gamma^2}\} = E\{\sum_{i=1}^{n} -\frac{1}{(1+\gamma)^2} + I_{\{y_i=0\}}\frac{1}{(\gamma + f_{0i})^2}\}.$$

To obtain the quantities $I_{\pi\pi}$, $I_{\pi\phi}$ etc. in closed form we need the quantities in what follows.

$$l'_{0i(\pi)} = \frac{\partial l_{0i}}{\partial \pi} = \sum_{r=1}^{m_i} \frac{-(1-\phi)}{(1-\pi)(1-\phi)+(r-1)\phi},$$

$$l'_{0i(\phi)} = \frac{\partial l_{0i}}{\partial \phi} = \sum_{r=1}^{m_i} \frac{(r-1)-(1-\pi)}{(1-\pi)(1-\phi)+(r-1)\phi} - \sum_{r=1}^{m_i} \frac{r-2}{1-\phi+(r-1)\phi},$$

$$l'_{y_i(\pi)} = \frac{\partial l_{y_i}}{\partial \pi} = \sum_{r=1}^{y_i} \frac{(1-\phi)}{\pi(1-\phi)+(r-1)\phi} - \sum_{r=1}^{m_i-y_i} \frac{(1-\phi)}{(1-\pi)(1-\phi)+(r-1)\phi},$$

$$\begin{aligned} l'_{y_i(\phi)} &= \frac{\partial l_{y_i}}{\partial \phi} = \sum_{r=1}^{y_i} \frac{(r-1)-\pi}{\pi(1-\phi)+(r-1)\phi} + \sum_{r=1}^{m_i-y_i} \frac{(r-1)-(1-\pi)}{(1-\pi)(1-\phi)+(r-1)\phi} \\ &\quad - \sum_{r=1}^{m_i} \frac{r-2}{(1-\phi+(r-1)\phi)}. \end{aligned}$$

$E(I_{\{y_i=0\}}) = Pr(y_i = 0|m_i) = \omega + (1-\omega)f_{0i} = \frac{\gamma + f_{0i}}{1+\gamma}$, $E(I_{\{y_i>0\}}) = Pr(y_i > 0|m_i) = 1 - \omega - (1-\omega)f_{0i} = \frac{1-f_{0i}}{1+\gamma}$ and $E(I_{\{y_i>0\}}g(y_i)) = (1-\omega)\sum_{y_i=1}^{m_i}(g(y_i)f_{y_i}) = \frac{1}{1+\gamma}\sum_{y_i=1}^{m_i}(g(y_i)f_{y_i})$, where $g(y_i)$ is a function of y_i.

$$l''_{0i(\pi\pi)} = E(\frac{-\partial^2 l_{0i}}{\partial \pi^2}) = \sum_{r=1}^{m_i} \frac{(1-\phi)^2}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2},$$

$$l''_{0i(\pi\phi)} = E(\frac{-\partial^2 l_{0i}}{\partial \pi \partial \phi}) = \sum_{r=1}^{m_i} \frac{-(r-1)}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2},$$

$$l''_{0i(\phi\phi)} = E(\frac{-\partial^2 l_{0i}}{\partial \phi^2}) = \sum_{r=1}^{m_i} \frac{\{(r-1)-(1-\pi)\}^2}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2} - \sum_{r=1}^{m_i} \frac{(r-2)^2}{\{(1-\phi+(r-1)\phi\}^2},$$

$$l''_{\pi\pi} = E\sum_{i=1}^{n}(I_{\{y_i>0\}}\frac{-\partial^2 l_{y_i}}{\partial\pi^2}) = \frac{1}{1+\gamma}\sum_{i=1}^{n}\sum_{y_i=1}^{m_i}[\frac{-\partial^2 l_{y_i}}{\partial\pi^2}]f_{y_i}$$

$$= \frac{(1-\phi)^2}{1+\gamma}\sum_{i=1}^{n}\sum_{y_i=1}^{m_i}[\sum_{r=1}^{y_i}\frac{1}{\{\pi(1-\phi)+(r-1)\phi\}^2}$$

$$+\sum_{r=1}^{m_i-y_i}\frac{1}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2}]f_{y_i}.$$

$$l''_{\pi\phi} = \sum_{i=1}^{n}E(I_{\{y_i>0\}}\frac{-\partial^2 l_{y_i}}{\partial\pi\partial\phi})$$

$$= \frac{1}{1+\gamma}\sum_{i=1}^{n}\sum_{y_i=1}^{m_i}[\sum_{r=1}^{y_i}\frac{r-1}{\{\pi(1-\phi)+(r-1)\phi\}^2} - \sum_{r=1}^{m_i-y_i}\frac{r-1}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2}]f_{y_i}$$

$$= \frac{1}{\phi(1+\gamma)}\sum_{i=1}^{n}\sum_{y_i=1}^{m_i}[\sum_{r=1}^{y_i}\frac{1}{\pi(1-\phi)+(r-1)\phi} - \sum_{r=1}^{m_i-y_i}\frac{1}{(1-\pi)(1-\phi)+(r-1)\phi}]f_{y_i}$$

$$- \frac{\pi(1-\phi)}{\phi(1+\gamma)}\sum_{i=1}^{n}\sum_{y_i=1}^{m_i}[\sum_{r=}^{y_i}\frac{1}{\{\pi(1-\phi)+(r-1)\phi\}^2}]f_{y_i}$$

$$+ \frac{(1-\pi)(1-\phi)}{\phi(1+\gamma)}\sum_{i=1}^{n}\sum_{y_i=1}^{m_i}[\sum_{r=1}^{m_i-y_i}\frac{1}{\{(1-\pi)(1-\phi)+(r-1)\phi)^2}\}]f_{y_i}.$$

Now it can be shown that

$$\sum_{y_i=1}^{m_i}[\sum_{r=1}^{y_i}\frac{1}{\{\pi(1-\phi)+(r-1)\phi\}^2}]f_{y_i} = \sum_{r=1}^{m_i}\frac{P(y_i \geq r)}{\{\pi(1-\phi)+(r-1)\phi\}^2}$$

and

$$\sum_{y_i=1}^{m_i}[\sum_{r=1}^{m_i-y_i}\frac{1}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2}]f_{y_i} = \sum_{r=1}^{m_i}\frac{P(y_i \leq m_i - r) - f_{oi}}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2}.$$

Further, from $E(\frac{\partial l}{\partial\pi}) = 0$ we obtain

$$E[\sum_{i=1}^{n} I_{\{y_i>0\}}l'_{y_i(\pi)}\}] = -E[\sum_{i=1}^{n}\{I_{\{y_i=0\}}\frac{f_{0i}l'_{0i(\pi)}}{\gamma+f_{0i}}] = -\frac{1}{1+\gamma}\sum_{i=1}^{n}f_{0i}l'_{0i(\pi)}.$$

Again,

$$E[\sum_{i=1}^{n} I_{\{y_i>0\}} l'_{y_i(\pi)}] = \frac{1}{1+\gamma}\sum_{i=1}^{n}\sum_{y_i=1}^{m_i} l'_{y_i(\pi)} f_{yi}$$
$$= \frac{1-\phi}{1+\gamma}\sum_{i=1}^{n}\sum_{y_i=1}^{m_i}[\sum_{r=1}^{y_i}\frac{1}{\pi(1-\phi)+(r-1)\phi}$$
$$- \sum_{r=1}^{m_i-y_i}\frac{1}{(1-\pi)(1-\phi)+(r-1)\phi}]f_{y_i}.$$

Using these results we obtain

$$l''_{\pi\pi} = \frac{(1-\phi)^2}{1+\gamma}\sum_{i=1}^{n}[\sum_{r=1}^{m_i}\frac{P(y_i \geq r)}{\{\pi(1-\phi)+(r-1)\phi\}^2} + \sum_{r=1}^{m_i}\frac{P(y_i \leq m_i - r) - f_{oi}}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2}]$$

and

$$l''_{\pi\phi} = \frac{-1}{\phi(1-\phi)(1+\gamma)}\sum_{i=1}^{n} f_{0i} l'_{0i(\pi)} - \frac{\pi(1-\phi)}{\phi(1+\gamma)}\sum_{i=1}^{n}\sum_{r=1}^{m_i}\frac{P(y_i \geq r)}{\{\pi(1-\phi)+(r-1)\phi\}^2}$$
$$+ \frac{(1-\pi)(1-\phi)}{\phi(1+\gamma)}\sum_{i=1}^{n}\sum_{r=1}^{m_i}\frac{P(y_i \leq m_i - r) - f_{oi}}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2}.$$

E$(\frac{\partial^2 l_{y_i}}{\partial\phi^2})$ is more easily obtained by defining $c = \frac{1-\phi}{\phi}$ and using the fact that $\frac{\partial^2 l_{y_i}}{\partial\phi^2} = \frac{\partial^2 l_{y_i}}{\partial c^2}(\frac{\partial c}{\partial\phi})^2 + \frac{\partial l_{y_i}}{\partial c}(\frac{\partial^2 c}{\partial\phi^2})$, where, in $\frac{\partial^2 l_{y_i}}{\partial c^2}$, c is replaced by $\frac{1-\phi}{\phi}$. Then, proceeding in the mannar in which $l''_{\pi\phi}$ and $l''_{\pi\phi}$ were obtained, it can be shown that

$$l''_{\phi\phi} = E\sum_{i=1}^{n}(I_{\{y_i>0\}}\frac{-\partial^2 l_{y_i}}{\partial\phi^2})$$
$$= \frac{1}{\phi^2(1+\gamma)}\sum_{i=1}^{n}[\pi^2\sum_{r=1}^{m_i}\frac{P(y_i \geq r)}{\{\pi(1-\phi)+(r-1)\phi\}^2}$$
$$+ (1-\pi)^2\sum_{r=1}^{m_i}\frac{P(y_i \leq m_i - r) - f_{0i}}{\{(1-\pi)(1-\phi)+(r-1)\phi\}^2}$$
$$- (1-f_{0i})\sum_{r=1}^{m_i}\frac{1}{\{1-\phi+(r-1)\phi\}^2}] - \frac{2}{\phi(1+\gamma)^2}\sum_{i=1}^{n} f_{0i} l_{0i}(\phi).$$

Using these results on the right hand side of the expressions for $I_{\pi\pi}$, $I_{\pi\phi}$ etc, after simplification we obtain

$$
\begin{aligned}
I_{\pi\pi} &= E\{-\frac{\partial^2 l}{\partial \pi^2}\} = \{\sum_{i=1}^{n} \frac{(\gamma + f_{0i}) f_{0i} l''_{0i(\pi\pi)} - \gamma f_{0i} (l'_{0i(\pi)})^2}{(1+\gamma)(\gamma + f_{0i})} + l''_{\pi\pi}\}, \\
I_{\pi\phi} &= E\{-\frac{\partial^2 l}{\partial \pi \partial \phi}\} = \{\sum_{i=1}^{n} \frac{(\gamma + f_{0i}) f_{0i} l''_{0i(\pi\phi)} - \gamma f_{0i} l'_{0i(\pi)} l'_{0i(\phi)}}{(1+\gamma)(\gamma + f_{0i})} + l''_{\pi\phi}\}, \\
I_{\phi\phi} &= E\{-\frac{\partial^2 l}{\partial \phi^2}\} = \{\sum_{i=1}^{n} \frac{(\gamma + f_{0i}) f_{0i} l''_{0i(\phi\phi)} - \gamma f_{0i} (l'_{0i(\phi)})^2}{(1+\gamma)(\gamma + f_{0i})} + l''_{\phi\phi}\}, \\
I_{\pi\gamma} &= E\{-\frac{\partial^2 l}{\partial \pi \partial \gamma}\} = \frac{1}{1+\gamma} \sum_{i=1}^{n} \{\frac{f_{0i} l'_{0i(\pi)}}{\gamma + f_{0i}}\}, \\
I_{\phi\gamma} &= E\{-\frac{\partial^2 l}{\partial \phi \partial \gamma}\} = \frac{1}{1+\gamma} \sum_{i=1}^{n} \{\frac{f_{0i} l'_{0i(\phi)}}{\gamma + f_{0i}}\}, \\
I_{\gamma\gamma} &= E\{-\frac{\partial^2 l}{\partial \gamma^2}\} = \sum_{i=1}^{n} \{-\frac{1}{(1+\gamma)^2} + \frac{1}{(1+\gamma)(\gamma + f_{0i})}\}.
\end{aligned}
$$

Appendix C: The expected Fisher information matrix of zero-inflated bivariate Poisson regression model

Consider the log-likelihood l in equation (7.4.1). For $j = 1, ..., n_i$, $i = 1, ..., k$, define the following quantities.

$$\begin{aligned}
A_{ij}^{(1)} &= y_{1ij} - (1-\theta_{0i})(\lambda_{1i} + \lambda_{0i}), \\
A_{ij}^{(2)} &= y_{2ij} - (1-\theta_{0i})(\lambda_{2ij} + \lambda_{0i}), \\
A_{ij}^{(3)} &= (1-\theta_{0i})\left[\frac{f_2(y_{1ij}-1, y_{2ij}-1 |\, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})}{f_3(y_{1ij}, y_{2ij} |\, \theta, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})} - 1\right], \\
A_{ij}^{(4)} &= \frac{\delta_{y_{1ij}, y_{2ij}}(0,0)}{f_3(y_{1ij}, y_{2ij} |\, \theta, \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})} - 1,
\end{aligned}$$

where $\delta_{y_{1ij}, y_{2ij}}(0,0) = \begin{cases} 1, & \text{if } (y_{1ij}, y_{2ij}) = (0,0), \\ 0, & \text{otherwise,} \end{cases}$

Now write $A_{ij} = (A_{ij}^{(1)}, A_{ij}^{(2)}, A_{ij}^{(3)}, A_{ij}^{(4)})'$. Further, for $i = 1, ...k$, $j = 1, ..., n_i$, let

$$C_{ij} = \begin{pmatrix} 0 & 0 & 1 & \theta_{0i} \\ \frac{1}{\lambda_{1i}} & 0 & -\frac{\lambda_{0i}}{\lambda_{1i}} & \theta_{0i} \\ 0 & \frac{1}{\lambda_{2ij}} & -\frac{\lambda_{0i}}{\lambda_{2ij}} & \theta_{0i} \\ 0 & 0 & 0 & \frac{1}{1-\theta_{0i}} \end{pmatrix}$$

be a 4×4 matrix. Now, let $C_{ij}^{(s)}$ be the sth row of the matrix C_{ij}, $s = 1, 2, 3, 4$.

Then, the first derivatives of l with respect to the parameters $\lambda_{0i}, \lambda_{1i}$, θ_{0i}, $i = 1, ..., k$, and γ are given by

$$\begin{aligned} \frac{\partial l}{\partial \lambda_{0i}} &= \sum_{j=1}^{n_i} C_{ij}^{(1)} A_{ij}, \\ \frac{\partial l}{\partial \lambda_{1i}} &= \sum_{j=1}^{n_i} [C_{ij}^{(2)} A_{ij} + \exp(x_{ij}^{'}\gamma) C_{ij}^{(3)} A_{ij}], \\ \frac{\partial l}{\partial \theta_{0i}} &= \sum_{j=1}^{n_i} [C_{ij}^{(4)} A_{ij}], \end{aligned}$$

and

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \lambda_{2ij} C_{ij}^{(3)} A_{ij} x_{ij}.$$

To find the expected Fisher information matrix we need to evaluate expected values of the second mixed partial derivatives, which can be expressed in terms of the product of the first derivatives as $E\left(-\frac{\partial l^2}{\partial \gamma \partial \gamma'}\right) = E\left[\left(\frac{\partial l}{\partial \gamma}\right)\left(\frac{\partial l}{\partial \gamma}\right)'\right]$, $E\left(-\frac{\partial l^2}{\partial \gamma \partial \lambda_{1i}}\right) = E\left[\left(\frac{\partial l}{\partial \gamma}\right)\left(\frac{\partial l}{\partial \lambda_{1i}}\right)\right]$, ..., etc.

Now, it can be seen that $E(A_{ij}) = 0$, $E(A_{ij}A_{i'j'}') = 0$, if $i \neq i'$ or $j \neq j'$. Denote $E(A_{ij}A_{ij}')$ by V_{ij}, which is a 4×4 symmetric matrix. It can be shown that

$$V_{ij}(1,1) = (1-\theta_{0i})(\lambda_{1i} + \lambda_{0i})[1 + \theta_{0i}(\lambda_{1i} + \lambda_{0i})],$$

$$V_{ij}(1,2) = (1-\theta_{0i})[\lambda_{0i} + \theta_{0}(\lambda_{1i} + \lambda_{0i})(\lambda_{2ij} + \lambda_{0i})]\ ,$$

$$V_{ij}(1,3) = (1-\theta_{0i}) + \theta_{0i}(1-\theta_{0i})(\lambda_{2ij} + \lambda_{0i}),$$

$$V_{ij}(1,4) = -(1-\theta_{0i})(\lambda_{1i} + \lambda_{0i}),$$

$$V_{ij}(2,2) = (1-\theta_{0i})(\lambda_{2ij} + \lambda_{0i})[1 + \theta_{0i}(\lambda_{1i} + \lambda_{0i})],$$

$$V_{ij}(2,3) = (1-\theta_{0i}) + \theta_{0i}(1-\theta_{0i})(\lambda_{2ij} + \lambda_{0i}),$$

$$V_{ij}(2,4) = -(1-\theta_{0i})(\lambda_{2ij} + \lambda_{0i}),$$

$V_{ij}(3,3) = (1-\theta_{0i})Q_{ij} - (1-\theta_{0i})^2,$

where $Q_{ij} = \sum_{y_1=1}^{\infty}\sum_{y_2=1}^{\infty} \frac{f_2^2(y_1-1, y_2-1|\ \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})}{f_2(y_1, y_2|\ \lambda_{0i}, \lambda_{1i}, \lambda_{2ij})},$

$V_{ij}(3,4) = -(1-\theta_{0i}),$

$V_{ij}(4,4) = \frac{(1-\theta_{0i})[1-\exp(-\lambda_{1i}-\lambda_{2ij}-\lambda_{0i})]}{\theta_{0i}+(1-\theta_{0i})\exp(-\lambda_{1i}-\lambda_{2ij}-\lambda_{0i})},$

Now, partition the $(4k+p-1)\times(4k+p-1)$ expected Fisher information matrix I as

$$I = \begin{pmatrix} I_{\gamma\gamma} & I_{\gamma\lambda_1} & I_{\gamma\lambda_0} & I_{\gamma\theta_0} \\ I'_{\gamma\lambda_1} & I_{\lambda_1\lambda_1} & I_{\lambda_1\lambda_0} & I_{\lambda_1\theta_0} \\ I'_{\gamma\lambda_0} & I'_{\lambda_1\lambda_0} & I_{\lambda_0\lambda_0} & I_{\lambda_0\theta_0} \\ I'_{\gamma\theta_0} & I'_{\lambda_1\theta_0} & I'_{\lambda_0\theta_0} & I_{\theta_0\theta_0} \end{pmatrix}.$$

Now, we obtain the elements of the partitioned matrices $I_{\gamma\gamma}$, $I_{\gamma\lambda_1}$, ..., etc. Using the above results we obtain the $(k+p-1)\times(k+p-1)$ matrix $I_{\gamma\gamma}$ as

$I_{\gamma\gamma} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}[\lambda_{2ij}^2 C_{ij}^{(3)} V_{ij} C_{ij}^{(3)'}]x_{ij}x'_{ij}.$

Note that each of the matrices $I_{\gamma\lambda_1}$, $I_{\gamma\lambda_0}$ and $I_{\gamma\theta_0}$ is of dimension $(k+p-1)\times k$. Now let $I_{\gamma\lambda_1}(i)$, $I_{\gamma\lambda_0}(i)$ and $I_{\gamma\theta_0}(i)$ be the ith column, $i = 1, ..., k$, of the matrices $I_{\gamma\lambda_1}$, $I_{\gamma\lambda_0}$ and $I_{\gamma\theta_0}$ respectively. Then it can be shown that

$I_{\gamma\lambda_1}(i) = \sum_{j=1}^{n_i}\lambda_{2ij}[C_{ij}^{(2)}V_{ij}C_{ij}^{(3)'} + \exp(x'_{ij}\gamma)C_{ij}^{(3)}V_{ij}C_{ij}^{(3)'}]x_{ij},$

$I_{\gamma\lambda_0}(i) = \sum_{j=1}^{n_i}\lambda_{2ij}[C_{ij}^{(1)}V_{ij}C_{ij}^{(3)'}]x_{ij},$

$I_{\gamma\theta_0}(i) = \sum_{j=1}^{n_i}\lambda_{2ij}[C_{ij}^{(3)}V_{ij}C_{ij}^{(4)'}]x_{ij}.$

Further, note each of the matrices $I_{\lambda_1\lambda_1}$ $I_{\theta_0\theta_0}$, $I_{\lambda_0\lambda_0}$, $I_{\lambda_1\lambda_0}$, $I_{\lambda_1\theta_0}$ and $I_{\lambda_0\theta_0}$ is a $k\times k$ diagonal matrix. Now let $I_{\lambda_1\lambda_1}(i,i)$, $I_{\theta_0\theta_0}(i,i)$, $I_{\lambda_0\lambda_0}(i,i)$, $I_{\lambda_1\lambda_0}(i,i)$, $I_{\lambda_1\theta_0}(i,i)$ and $I_{\lambda_0\theta_0}(i,i)$ be the (i,i)th element, $i = 1, ..., k$ of the matrices $I_{\theta_0\theta_0}$, $I_{\lambda_0\lambda_0}$, $I_{\lambda_1\lambda_0}$, $I_{\lambda_1\theta_0}$ and $I_{\lambda_0\theta_0}$ respectively. Then it can be shown that

$I_{\lambda_1\lambda_1}(i,i) = \sum_{j=1}^{n_i}[C_{ij}^{(2)}V_{ij}C_{ij}^{(2)'} + 2\exp(x_{ij}'\gamma)C_{ij}^{(2)}V_{ij}C_{ij}^{(3)'} + \exp(2x_{ij}'\gamma)C_{ij}^{(3)}V_{ij}C_{ij}^{(3)'}]$,

$I_{\theta_0\theta_0}(i,i) = \sum_{j=1}^{n_i}[C_{ij}^{(4)}V_{ij}C_{ij}^{(4)'}]$,

$I_{\lambda_0\lambda_0}(i,i) = \sum_{j=1}^{n_i}[C_{ij}^{(1)}V_{ij}C_{ij}^{(1)'}]$,

$I_{\lambda_1\lambda_0}(i,i) = \sum_{j=1}^{n_i}[C_{ij}^{(1)}V_{ij}C_{ij}^{(2)'} + \exp(x_{ij}'\gamma)C_{ij}^{(1)}V_{ij}C_{ij}^{(3)'}]$,

$I_{\lambda_1\theta_0}(i,i) = \sum_{j=1}^{n_i}[C_{ij}^{(2)}V_{ij}C_{ij}^{(4)'} + \exp(x_{ij}'\gamma)C_{ij}^{(3)}V_{ij}C_{ij}^{(4)'}]$,

$I_{\lambda_0\theta_0}(i,i) = \sum_{j=1}^{n_i}[C_{ij}^{(1)}V_{ij}C_{ij}^{(4)'}]$ respectively.

This completes evaluation of the elements of the matrix I.

Bibliography

[1] Böhning, D., Dietz, E., Schlattmann, P., Mendonca, L. and Kirchner, U.(1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society A*, **162**, 195-209.

[2] Böhning, D. and Seidel, W. (2003). Editorial: recent developments in mixture models. *Computational Statistics and Data Analysis*, **41**, 349-357.

[3] Barnwal, R.K. and Paul, S.R. (1988). Analysis of one-way layout of count data with negative binomial variation. *Biometrika*, **75**, 215-222.

[4] Breslow, N. (1990). Tests of hypotheses in over-dispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association*, **85**, 565-571.

[5] Bartoo, J.B. and Puri, P.S. (1967). On optimal asymptotic tasks of composite statistical hypothesis. *The Annals of Mathematical Statistics*, **38**, 1845-1852.

[6] Berry, D.A. (1987). Logarithmic transformations in ANOVA. *Biometrics*, **43**, 439-456.

[7] Chesher, A. (1984). Testing for neglected heterogeneity. *Econometrica*, **52**, 865-872.

[8] Cochran, W.G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Mathematical Proceedings of the Cambridge Philosophical Society*, **30**, 178-191.

[9] Cox, D.R. (1983). Some remarks on overdispersion. *Biometrika*, **70**, 1269-1274.

[10] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, London: Chapman and Hall.

[11] Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **49**, 1-39

[12] Hall, D.B.(2000). Zero-inflated Poisson and binomial regression with random effects: A Case Study. *Biometrics*, **56**, 1030-1039.

[13] Dean, C.B. (1992). Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*, **87**, 451-457

[14] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, **39**, 1-38.

[15] Deng, D. and Paul, S.R. (2000). Score tests for zero-inflation in generalized linear models. *The Canadian Journal of Statistics*, **28**, 563-570.

[16] Farewell, V.T. and Sprott, D.A. (1988). The use of a mixture model in the analysis of count data. *Biometrics*, **44**, 1191-1194.

[17] Fisher, R.A. (1950). *Statistical Methods for Research Workers.* 11th edition. London: Oliver and Boyd.

[18] Jacqmin, H. and Commenges, D. (1995). Tests of homogeneity for generalized linear models. *Journal of the American Statistical Association*, **90**, 1237-1246.

[19] Hartley, H. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, **14**, 1 and 2, 174-194.

[20] Hogg, R.V. and Craig, A. T. (1995). *Introduction to Mathematical and Statistics.* Fifth edition. Prentice-Hall, Inc.

[21] Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika*, **51**, 1 and 2, 241-245.

[22] Johnson, N.L., Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Distributions.* Second edition. New York: John Wiley & Sons, Inc.

[23] Karlis, D. and Ntzoufras, I. (2001). Analysis of sports data using bivariate Poisson models. Technical Report, Department of Statistics, Athens University of Economics and Business.

[24] Kocherlakota, S. and Kocherlakota, K. (2001). Regression in the bivariate Poisson distribution. *Communication Statistics*, **30(5)**, 815-825.

[25] Liang, K.Y. (1987). A locally most powerful test for homogeneity with many strata. *Biometrika*, **74(2)**, 259-64.

[26] Lambert, D.(1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1-14.

[27] Lawless, J.F. (1987a). Negative binomial regression models. *Canadian Journal of Statistics,* **15**, 209-226.

[28] Lawless, J.F. (1987b). Regression methods for Poisson process data. *Journal of the American Statistical Association*, **82**, 808-815.

[29] Li, C.-S., Lu, J.-C., Park, J., Kim, K., Brinkley, P.A. and Peterson, J.P. (1999). Multivariate zero-inflated Poisson models and their applications. *Technometrics*, **41**, 29-38.

[30] Littell, R.C. and Folks, J. L. (1971). Asymptotic optimality of Fisher's method of combining independent tests. *Journal of the American Statistical Association*, **66**, 802-806.

[31] Littell, R.C. and Folks, J. L. (1973). Asymptotic optimality of Fisher's method of combining independent tests II. *Journal of the American Statistical Association*, **68**, 193-194.

[32] McCaughran, D.A. and Arnold, D.W. (1976). Statistical models for numbers of implantation sites and embryonic deaths in mice. *Toxicology and applied pharmacology*, **38**, 325-333.

[33] McCullagh, P. and Nelder, J.A (1987). *Generalized Linear Models.* 2nd ed. London: Chapman and Hall.

[34] Moran, P.A.P. (1970). On asymptotically optimal tests of composite hypothesis. *Biometrika*, **57**, 47-55.

[35] Neyman, J. (1959). Optimal asymptotic tests for composite hypothesis. In *Probability and Statistics: The Harold Cramer Volume,* U. Grenander (ed). New York: Wiley, 213-214.

[36] Neyman, J. and Scott, E.L. (1966). On the use of $c(\alpha)$ optimal tests of composite hypothesis. In *Bulletin of the Institute of International Statistics*, **41**, 477-497.

[37] Paul, S.R. (1982). Analysis of proportions of affected foetuses in teratological experiments. *Biometrics*, **38**, 361-370.

[38] Paul, S.R. and Banerjee, T. (1998). Analysis of two-way layout of count data involving multiple counts in each cell. *Journal of the American Statistical Association*, **93**, 1419-1429.

[39] Paul, S.R. and Ho, I. (1989). Estimation in the bivariate Poisson distribution and hypothesis testing concerning independence. *Communications in Statistics: Theory and Methods*, **A18**, 1123-1133.

[40] Peterson. L. E. (1997). PIRLS: Poisson iteratively reweighted least squares computer program for additive, multiplicative, power, and non-linear models. *Journal of Statistical Software* , **2**, Issue 5.

[41] Prentice, R.L. (1986). Binary regression using an extended beta-binomial distribution with discussion on correlation induced by covariate measurement errors. *Journal of the American Statistical Association*, **81**, 321-327.

[42] Rao, C.R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, **44**, 50-57.

[43] Singh, N. (1986). A simple and asymptotically optimal test for the equality of normal populations. A pragmatic approach to one-way classification. *Journal of the American Statistical Association*, **81**, 703-704.

[44] Snedecor, G.W. and Cochran, W.G. (1967). *Statistical Methods.* 6th edition. Ames, Iowa. Iowa State University Press.

[45] Wang, K., Lee, A. H., Yau, K. W. and Carrivick, P.J.W. (2003). A bivariate zero-inflated Poisson regression model to analyse occupational injuries. *Accident Analysis & Prevention* **35**, 625-629.

[46] Zelterman, D. and Chen, C. (1988). Homogeneity tests against central mixture alternatives. *Journal of the American Statistical Association*, **83**, 179-182.

Vita Auctoris

Name: Xing Jiang

Date of Birth: 1966

Place of Birth: Heilongjiang, China

Education: B.S. in Mathematics, 1989, University of Science and Technology of China

M.S. in Statistics, 1992, University of Science and Technology of China

Ph.D. in Statistics, 2004, University of Windsor, Canada