DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

TESTING NON-ADDITIVITY IN STATISTICAL MODELS

by

Aylin ALIN

September, 2006

İZMİR

TESTING NON-ADDITIVITY IN STATISTICAL MODELS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics Program

by
Aylin ALIN

September, 2006
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"TESTING NON-ADDITIVITY IN STATISTICAL MODELS"** completed by **AYLİN ALIN** under supervision of **PROF. DR. SERDAR KURT** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Serdar KURT

Supervisor

Assist. Prof. Dr. Ali Rıza FİRUZAN

Committee Member

Assist. Prof. Dr. Ali Kemal ŞEHİRLİOĞLU

Committee Member

Prof. Dr. Öztaş AYHAN

Jury Member

Assist. Prof. Dr. Cenk ÖZLER

Jury Member

Prof.Dr. Cahit HELVACI
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I am greatly indebted to my adviser Serdar KURT for accepting me to the department and giving me the chance to apply to the Ph.D degree. After such a long time since my Ms.c degree, he made my dream come true. He has always supported and believed in me. He has helped me in both academic and “real life” aspect of my graduate study. No thanks are enough to express my appreciations to him for everything he has done for me.

I would like to acknowledge my dissertation committee members, Ali Rıza FİRUZAN and Ali Kemal ŞEHİRLİOĞLU for their constructive comments and suggestions.

I would also like to thank Robert Boik for helping and guiding me through the development of my first paper which is also second chapter of this thesis. One special thank should go to Fernando Tusell. He has not only been a good friend to me but also he has been very generous for helping me and guiding me through my first paper. I would like to express my appreciations and thanks to Ayanendranath Basu. If it weren’t for him I would not know if my ideas about penalized power-divergence estimators and test statistics would work. He never hesitated to review my work and giving me advices even in his very busy schedule. It has been an honor and privilege to meet such persons with such academic achievements and still very modest.

Because of their endless love and undoubted belief in me since my childhood, my parents have helped me pursuit this degree for at least 25 years. They never stopped to believe in me and never had a doubt about me to accomplish things that I intend to. Getting my Ph.D degree will be the best gift that I could give them even I owe them so much. At last but not least, I would like to thank to my dearest husband Tufan ALIN. He has always been so patient and supportive during my graduate study and never stopped to believe in me even in the times that I had some doubts about my self. His love and support has been my encouragement and strength.

Aylin ALIN

TESTING NON-ADDITIVITY IN STATISTICAL MODELS

ABSTRACT

In this thesis, testing non-additivity (interaction) in two-way ANOVA tables, and contingency tables are studied. For two-way ANOVA tables, methods designed especially for testing interaction when there is only one observation (no replication) per cell are the focus, whereas log-linear models are considered for contingency tables.

Cressie & Read (1984) developed the family of power-divergence measures. The family of minimum power-divergence estimators is obtained by minimizing these measures for unknown parameter. Cressie & Pardo (2000, 2002) also developed unified approach to model selection problem in nested log-linear models with test statistics based on power-divergence measures. However, the weight they put on empty cells is the problem with power-divergence measures which affects the performances of minimum power-divergence estimators and power-divergence test statistics. Basu & Basu (1998) developed the family of penalized power-divergence measures as a solution for this problem. In this thesis, simulation study has been performed to compare efficiency and robustness properties of ordinary and penalized minimum power-divergence estimators for log-linear independence model in 2 x 2 contingency tables. The new families of penalized power-divergence test statistics have also been proposed to over come with the problem of weight that power-divergence test statistics put on empty cells. Ordinary power-divergence test statistics developed by Cressie & Pardo (2000, 2002) and proposed penalized power-divergence test statistics have been compared by simulation study in terms of exact size and power properties for testing nested log-linear models in 2 x 2 and 2 x 2 x 2 contingency tables.

Keywords: ANOVA, Contingency tables, Interaction, Non-additivity, Penalization, Power-divergence measures, Power-divergence test statistics.

İSTATİSTİKSEL MODELLERDE TOPLANAMAZLIK TESTİ

ÖZ

Bu tezde, iki yönlü ANOVA tablolarında etkileşim testi ve olumsallık tablolarında etkileşim testi çalışılmaktadır. İki yönlü ANOVA tablolarında her bir hücrede tek gözlem olduğu durumdaki etkileşim testi metotları çalışılırken, olumsallık tablolarında logaritmik-doğrusal modeller üzerinde durulmaktadır.

Cressie ve Read (1984) güç-sapma ölçümleri adı altında bir sapma ailesi geliştirmişlerdir. Güç-sapma kestiricileri ailesi bu güç sapma ölçümlerini bilinmeyen parametreye göre minimize ederek elde edilmektedir. Cressie ve Pardo (2000, 2002) iç içe geçmiş logaritmik-doğrusal modellerde model seçme problemine güç-sapma ölçümlerine dayanan test istatistikleri ile yeni bir yaklaşım geliştirmişlerdir. Ancak, güç-sapma ölçümlerinin problemi olumsallık tablolarındaki boş hücrelere atadıkları ağırlıklardır ki bu da hem güç-sapma kestiricilerinin hem de güç-sapma test istatistiklerinin performansını etkilemektedir. Basu ve Basu (1998) bu probleme bir çözüm olarak cezalandırılmış güç-sapma ölçümlerini geliştirmişlerdir. Bu tezde, normal güç-sapma kestiricileri ile cezalandırılmış güç sapma kestiricilerinin 2 x 2 lik olumsallık tablolarında bağımsız logaritmik-doğrusal model için etkinlik ve dirençlilik özelliklerini karşılaştırmak amacı ile benzetim çalışması yapılmıştır. Ayrıca, güç-sapma test istatistikleri ailesinin olumsallık tablolarındaki boş hücrelere koyduğu ağırlığa bir çözüm olarak cezalandırılmış güç-sapma test istatistikleri önerilmiştir. Cressie ve Pardo (2000, 2002) tarafından geliştirilen test istatistikleri ile önerilen cezalandırılmış güç-sapma test istatistiklerinin I. tip hata ve güç değerleri 2 x 2 lik ve 2 x 2 x 2 lik olumsallık tablolarında iç içe geçmiş logaritmik doğrusal modellerin testi için benzetim çalışmaları yolu ile karşılaştırılmıştır.

Anahtar sözcükler: ANOVA, Cezalandırma, Güç-sapma ölçümleri, Etkileşim, Olumsallık tabloları, Toplanamazlık, Güç-sapma test istatistikleri.

CONTENTS

Page

CHAPTER ONE
INTRODUCTION

Many studies in the social or biomedical sciences are carried out to investigate the size of some effect, for example opinions on various issues, the treatment effect in a controlled clinical trial or the risk of exposure to some substance in an observational study. It is usual to compare groups of subjects, such as those receiving active treatment with those getting placebo or those exposed with those not exposed or opinions of male and females on various subjects. This kind of comparison is often made using a statistical technique, such as t-test, a Pearson's chi-square test, ANOVA or ANCOVA. If these techniques or any other statistical techniques are applied to the groups as a whole then only one estimate of the size effect will be obtained. This means that there is an implicit assumption that the effect size is constant across different types of patients or subjects. In practice, this assumption may not be sensible. For instance, women may respond to a drug treatment differently than men do or opinions of women may differ depending on their education. The statistical term used for such heterogeneity of effect is "interaction". Interaction is used in Statistics to refer to situations where the effect of an explanatory variable on response is not the same at all levels of other explanatory variables. If there is interaction between two factors, model of observations include interaction term and is called "non-additive model" which makes interaction and non-additivity equivalent in terms of meaning.

Testing interaction effect is important in all sciences including the social and biomedical science applications and is used in most of the medical fields, such as pharmacology, cardiology or psychology. If researchers want to know whether the usage of a drug produces different effects on blood pressure according to gender or whether opinions of women and men on various subjects differ according to education or if they want to know whether the effects of the propofol on left ventricular pressure differs for normal and hypercholesterolemic isolated rabbit hearts as Oztekin, Kalkan, Ozzeybek, Tuncok, Guven, & Elar (2000) did, then they should test interaction between those variables.

Factorial designs are most efficient experimental design when there are effects of two or more factors to be studied. Since all possible combinations of factor levels are investigated in this design, it is possible to have interaction effect between factors. When there are two different factors to study, there will be a two-way table. In each cell of this table, there will be values of dependent variable for corresponding combination of factor levels (treatment). Each value of dependent variable is obtained by repeating the experiment on different subject for the corresponding treatment. Interaction between factors is usually investigated by repeating experiment. But, researchers may not be so lucky to have replicates. Deciding about significant interaction between two factors depends on the number of the replicates in each cell of two-way table and the structure of interaction in the model of observations. When there are replicates for each treatment, testing significant interaction between two factors can be accomplished by two-way ANOVA method. However, when there is only one observation per cell, using two-way ANOVA method for interaction test is not possible. But, there are other tests available for this. These tests will be discussed in Chapter 2.

A contingency table or cross-classified table is a table of counts and it basically deals with qualitative categories. It is called according to number of variables used for classifying subjects. For instance; it is called as “two-way contingency table” if there are two variables to classify subjects, whereas it is called as “three-way contingency table” if there are three variables to classify the subjects. Each variable corresponds to the one dimension of the table. Contingency tables are very useful in most of the applied sciences. A variety of biological and social science data come in the form of contingency tables. According to Wang & Zhang (1998), articles in Chinese medical journals tend to use t-tests and contingency tables more than other statistical methods. Tansey (1996) states that contingency tables are very useful tools in the strategy and human research management areas to analyze expected patterns produced by various combinations of certain variable levels. Using a contingency table, a researcher may wish to find out if interested qualitative characteristics are associated with one another, or independent of each other. In this sense, analysis of contingency tables has its counterpart such as the absence or presence of interaction

terms in the analysis of variance. “A number of ad hoc proposals have been suggested to take advantage of the analogies between these two types of analyses, but these efforts resulted only in limited success” (Ku, Varner, & Kullback , 1971). Darroch (1962) compared the interactions in contingency tables with interactions in the analysis of variance. However, he mentioned that “interactions in contingency tables enjoy only a few of the fortuitously simple properties of interactions in the analysis of variance”.

“In contingency table analysis the ANOVA-like models are used to describe the structural relationship among the variables corresponding to the dimensions of the table” (Fienberg, 1980, p:3). Log-linear modeling is one of these techniques. This technique is a discrete multivariate technique describing the association patterns among categorical variables by modeling expected cell counts (or equivalently cell probabilities) in contingency tables without distinguishing between independent and response variables. As understood from the name, log-linear models are linear in their logarithms. When a distinction is made between independent variables and response variables, log-linear models can be converted into logit or linear logistic response models. “Log-linear models are analogous to ANOVA models with unequal numbers of observations” (Christensen, 1990, p: 48). However, there is a much closer analogy between linear logistic models and the usual ANOVA or regression models. The use of log-linear modeling is more appropriate in situations where applied researchers are interested in the various pairwise and higher order associations (interactions) among set of independent categorical variables. Lee & Viele (2001) used log-linear models for modeling data from train waybills particularly focusing on revealing the relationships between cargo volume and origin, destination and commodity type. “Log-linear models are also used in survival analysis as accelerated failure time models” (Christensen, 2000). Among other categorical data techniques to analyze contingency tables, such as; probit or logit modeling, it is only focused on log-linear modeling for this thesis.

When there are different models available for the data set, there is a problem of determining the best model for the corresponding data. The method is needed that

aids in the selection of the interaction term / terms to be included in the fitted model. Fienberg (1970), Goodman (1970), Ku et al. (1971), Ku & Kullback (1974) are among others who suggest different approaches to the model selection problem. Later, Cressie & Read (1984) developed unified approach to model testing by the family of power-divergence test statistics. These statistics are based on the power-divergence measures which are also included in the family of disparity measures defined by Lindsay (1994). The family of power-divergence test statistics includes the well known test statistics such as likelihood ratio or Pearson's chi-square test statistics. A problem with the family of power-divergence measures, especially for small sample sizes, is disproportionately large weight that these measures put on empty cells. As a solution for this problem, Basu & Basu (1998) considered the empty cell penalty for the family of power divergence measures. By using these penalized power-divergence measures in the family of test statistics (which will be called as "the family of ordinary power-divergence test statistics" to not to confuse with the ones defined in this thesis), the family of penalized power-divergence test statistics have been defined in this thesis as a new solution to the model testing for small sample sizes in the contingency tables with empty cells.

Organization of this thesis is as follows: After studying the methods for testing interaction in two-way ANOVA tables with no replication case in Chapter 2, log-linear models will be described in Chapter 3. In Chapter 4, ordinary and penalized power-divergence measures will be described, and families of ordinary and penalized power-divergence estimators will be compared in terms of efficiency and robustness properties in two-way contingency tables. Chapter 4 will also include simulation studies performed to examine the size and power properties of the family of ordinary power-divergence test statistics and proposed family of penalized power-divergence test statistics in two-way and three-way contingency tables. Finally, conclusion and potential further research will be discussed in Chapter5.

CHAPTER TWO
TESTING INTERACTION IN TWO-WAY ANOVA TABLES

After giving the linear model for fixed two factors in the subsequent section, the regular two-way ANOVA test statistic for interaction with n > 1 observations in each cell will be given in Section 2.2. Then, tests designed for testing interaction with one observation per cell will be described in Section 2.3, starting with the tests designed for the models with interaction terms that depend on both or any of the main effects and followed by the tests designed for the models in which interaction terms do not depend on any of the main effects. Tests studied in these subsections are applicable when the interaction structure is specificied. There are other tests that do not need specified interaction structure. They will be reviewed as well in a separate subsection. Each subsection in Section 2.3 will be followed by an appropriate example, and finally additional example about measurement reliability will be given in Section 2.4.

2.1 Model

Let us have two factors called factor A and factor B. In two-way ANOVA with equal number of observations per cell, observations can be described by the following linear model including interaction effect of two factors;

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \\ k = 1, \ldots, n \end{cases} \tag{2.1}$$

where

Y_{ijk}: k-th observation at i-th level of factor A and j-th level of factor B

μ : Overall mean

α_i : Main effect of i-th level of factor A

β_j : Main effect of j-th level of factor B

τ_{ij} : Interaction effect of i-th level of factor A and j-th level of factor B

ε_{ijk}: Error on the k-th observation of i-th level of factor A and j-th level of factor B

Factors are assumed to be fixed and errors are assumed to have $N(0, \sigma^2)$. The following restrictions are imposed on main effects and interaction effects:

$$\sum_{i=1}^{a} \alpha_i = 0, \sum_{j=1}^{b} \beta_j = 0, \sum_{i=1}^{a} \tau_{ij} = 0, \sum_{j=1}^{b} \tau_{ij} = 0.$$

These restrictions on parameters and assumptions about the distribution of error will be valid for the rest of this Chapter.

The least square and maximum likelihood estimators of these parameters are;

$$\hat{\mu} = \overline{Y}_{...} \qquad \hat{\alpha}_i = \overline{Y}_{i..} - \overline{Y}_{...} \qquad \hat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...} \qquad \hat{\tau}_{ij} = \overline{Y}_{ij.} - (\overline{Y}_{...} + \hat{\alpha}_i + \hat{\beta}_j)$$

where

$$\overline{Y}_{ij.} = \frac{\sum_{k=1}^{n} Y_{ijk}}{n}, \ \overline{Y}_{.j.} = \frac{\sum_{i=1}^{a} \overline{Y}_{ij.}}{a}, \ \overline{Y}_{i..} = \frac{\sum_{j=1}^{b} \overline{Y}_{ij.}}{b}, \ \overline{Y}_{...} = \frac{\sum_{i=1}^{a} \overline{Y}_{i..}}{a}, \ \overline{Y}_{...} = \frac{\sum_{j=1}^{b} \overline{Y}_{.j.}}{b}.$$

When there is no replication, i.e., n = 1,

$$\overline{Y}_{ij.} = Y_{ij}, \ \overline{Y}_{i..} = \overline{Y}_{i.}, \ \overline{Y}_{.j.} = \overline{Y}_{.j}, \ \overline{Y}_{...} = \overline{Y}_{..}$$

The fitted form of the model given in Eq.(2.1) is,

$$\hat{Y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\tau}_{ij} \tag{2.2}$$

If there is no interaction between factors, i.e. if all of the interactions are equal to zero ($\tau_{ij} = 0$ for all i and j), then model given with Eq.(2.1) reduces to Eq.(2.3) and is called "Additive model". This model will be tested against the model given with Eq.(2.1), i.e., $H_0 : \tau_{ij} = 0$ for all i,j.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \tag{2.3}$$

2.2 Testing Interaction When There are n > 1 Observations per Cell

When there are n > 1 observations in each cell, as in model (2.1), usual test statistics for $H_0 : \tau_{ij} = 0$ is (Montgomery, 1991, p: 206)

$$F = \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2 \Big/ (a-1)(b-1)}{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \overline{Y}_{ij.})^2 \Big/ ab(n-1)} \tag{2.4}.$$

Under the null hypothesis H_0, this statistic has an F-distribution with [(a-1)(b-1), ab(n-1)] degrees of freedom. Nominator of this test statistic corresponds to the mean squares of interaction while denominator corresponds to mean square error (MSE).

2.3 Testing Interaction When There is One Observation per Cell

"When there is only one observation per cell, the error variance σ^2 is not estimable if interaction is modeled without imposing any structure (other than rows and columns summing to zero); that is, to estimate σ^2, one must model interaction

using fewer than (a-1)(b-1) functionally independent parameters" (Boik, by personal communication, 27 August 2004). "Otherwise, the two factor interaction effect τ_{ij} and the experimental error can not be separated" (Montgomery, 1991, p: 219). One way to solve this problem is to use sum of squares for interaction to estimate error variance σ^2 which causes another problem of being unable to use conventional linear model theory to check for interaction.

There are tests designed especially for the case of no replication. These tests can be categorized according to how the interaction structure is defined in the model. At following subsections, these tests will be discussed.

2.3.1 Models with Interaction Terms that Depend on Both or Any of The Main Effects

"Tukey (1949) was the first to propose a test to determine if there is interaction in two-way classification model with one observation per cell" (Johnson & Graybill, 1972). "In 1949, he developed a test for non-additivity, deriving a sum of squares with one degree of freedom" (Jazairi, 1993, p: 365). It should be pointed out that, in his paper Tukey (1949) did not propose a model with any particular structure of interaction (i.e. no specific functional form of τ_{ij}), but it appears that his test can be taken as a test of $H_0 : \lambda = 0$ in the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j + \varepsilon_{ij} \quad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \end{cases} \tag{2.5}$$

with the usual assumptions of normality of ε_{ij} and the restrictions on α_i, β_j and $\tau_{ij} = \lambda\alpha_i\beta_j$. To apply Tukey's one degree of freedom test, at least one of the a or b should be greater than two. Table 2.1 gives the analysis of variance table for Tukey (1949)'s one-degree of freedom test for non-additivity.

Table 2.1 The analysis of variance table for Tukey's one degree of freedom test for non-additivity

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Factor A	$SS_A = b\sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})^2$	a-1	$\frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$
Factor B	$SS_B = a\sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2$	b-1	$\frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$
Non-additivity	$SS_{AB} = \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}(\overline{Y}_{i.} - \overline{Y}_{..})(\overline{Y}_{.j} - \overline{Y}_{..})\right)^2}{\sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})^2 \sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2}$	1	SS_{AB}	$\frac{MS_{AB}}{MS_E}$
Error	$SS_E = SS_T - SS_{AB} - SS_A - SS_B$	(a-1)(b-1)-1	$\frac{SS_E}{(a-1)(b-1)-1}$	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{..})^2$	ab-1		

The analysis of variance shown in Table 2.1 yields Tukey's non-additivity test statistic $F = \frac{MS_{AB}}{MS_E}$ which has an F-distribution with [1, (a-1)(b-1)-1] degrees of freedom under the hypothesis $H_0 : \lambda = 0$.

Another model for testing interaction with one observation per cell, proposed by Mandel (1961), includes Tukey's one degree of freedom test as a special case. Mandel (1961) defined interaction depending on row effects or column effects. His interaction can be specified as follows;

$$\tau_{ij} = \theta_i\beta_j \rightarrow \text{interaction depending on column effect} \qquad (2.6)$$

$$\tau_{ij} = \Re_j\alpha_i \rightarrow \text{interaction depending on row effect} \qquad (2.7)$$

The model including $\theta_i\beta_j$ as an interaction effect term is given in Eq.(2.8)

$$Y_{ij} = \mu + \alpha_i + \beta_j + \theta_i\beta_j + \varepsilon_{ij} \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \end{cases} \tag{2.8}$$

with the usual assumption of normality of ε_{ij} and following restrictions;

$$\sum_{i=1}^{a} \alpha_i = 0, \sum_{j=1}^{b} \beta_j = 0, \sum_{i=1}^{a} \theta_i = 0, \sum_{i=1}^{a} \tau_{ij} = 0, \sum_{j=1}^{b} \tau_{ij} = 0.$$

Mandel (1961) suggested that useful class of non-additivity situations is obtained by assuming that for any given row, the quantity $E(Y_{ij} - \overline{Y}_{.j})$ is a linear function of β_j as given by Eq.(2.9)

$$E(Y_{ij} - \overline{Y}_{.j}) = \alpha_i + \theta_i\beta_j \tag{2.9}$$

where $E(Y_{ij}) = \mu + \alpha_i + \beta_j + \theta_i\beta_j$ and $E(\overline{Y}_{.j}) = \mu + \beta_j$. By defining Eq.(2.8) as given in Eq.(2.10)

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\phi_i - 1)\beta_j + \varepsilon_{ij} \tag{2.10}$$

where $\theta_i = \phi_i - 1$ and taking $\mu_i = \mu + \alpha_i$, Mandel (1961) obtained the linear form in Eq.(2.11) called "rows-linear model" which means that two-way classification table represents a bundle of "a" straight lines, corresponding to the "a" rows and differing from each other in both parameters μ_i and ϕ_i. In other words, rows-linear model implies that there is a linear effect on Y_{ij} from the column main effects, but this linear effect changes from row to row.

$$Y_{ij} = \mu_i + \phi_i\beta_j + \varepsilon_{ij} \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \end{cases} \tag{2.11}$$

where

$$\bar{\phi} = \frac{\sum_{i=1}^{a} \phi_i}{a} = \frac{\sum_{i=1}^{a} (1 + \theta_i)}{a} = 1.$$

Mandel suggested that testing non-additivity is equal to test H_0: $\phi_i = 1$ for all i = 1, . . ., a where ϕ_i is the slope of the line plotted for the values corresponding to i-th row of factor A versus β_j for j = 1, . . ., b. If all $\phi_i = 1$, in other words if all of the lines corresponding to the rows are parallel, the model reduces to the simple additive model. This can be seen from Eq.(2.10) and Eq.(2.11). The adequacy of the rows-linear model suggests that a plot of each row of the table versus the column averages $\bar{Y}_{.j}$ will be a scatter of points around a straight line. The analysis of variance table for the Mandel (1961)'s rows-linear model is given by Table 2.2.

Table 2.2 The analysis of variance table for Mandel's test for non-additivity

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Mean	$ab\bar{Y}_{..}^2$	1		
Factor A	$SS_A = b\sum_{i=1}^{a} (\bar{Y}_{i.} - \bar{Y}_{..})^2$	a-1	$\frac{SS_A}{a-1}$	$\frac{MS_A}{MS_{Res}}$
Factor B	$SS_B = a\sum_{j=1}^{b} (\bar{Y}_{.j} - \bar{Y}_{..})^2$	b-1	$\frac{SS_B}{b-1}$	$\frac{MS_B}{MS_{Res}}$
Slopes	$SS_{slopes} = \sum_{i=1}^{a} (\hat{\phi}_i - 1)^2 \sum_{j=1}^{b} (\bar{Y}_{.j} - \bar{Y}_{..})^2$ where $\hat{\phi}_i = \frac{\sum_{j=1}^{b} Y_{ij}(\bar{Y}_{.j} - \bar{Y}_{..})}{\sum_{j=1}^{b} (\bar{Y}_{.j} - \bar{Y}_{..})^2}$	a-1	$\frac{SS_{slopes}}{a-1}$	$\frac{MS_{slopes}}{MS_{Res}}$
Residual	$SS_{Res} = \sum_{i=1}^{a}\sum_{j=1}^{b} \left((Y_{ij} - \bar{Y}_{i.}) - \hat{\phi}_i(\bar{Y}_{.j} - \bar{Y}_{..})\right)^2$	(a-1)(b-2)	$\frac{SS_{Res}}{(a-1)(b-2)}$	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2$	ab		

What the analysis of Table 2.2 has accomplished is a breakdown of the (a-1)(b-1) degrees of freedom for interaction into two portions: (a-1) degrees of freedom for slopes and (a-1)(b-2) degrees of freedom for random error. The analysis given by Table 2.2 can be used for testing H_0: $\phi_i = 1$ for all i = 1, . . ., a (i.e., testing additivity) by comparing the mean squares of slopes and mean squares of residuals by the F-test with [(a-1), (a-1)(b-2)] degrees of freedom.

Mandel (1961) used Tukey (1961)'s test to show whether there is concurrence of the bundle of lines by defining the Tukey's sum of squares for interaction in the following manner:

$$SS_{AB} = \frac{\left(\sum_{i=1}^{a} \hat{\phi}_i (\overline{Y}_{i.} - \overline{Y}_{..})\right)^2}{\sum_{i=1}^{a} (\overline{Y}_{i.} - Y_{..})^2} \sum_{j=1}^{b} (\overline{Y}_{.j} - \overline{Y}_{..})^2 \qquad (2.12)$$

where

$$\hat{\phi}_i = \frac{\sum_{j=1}^{b} Y_{ij} (\overline{Y}_{.j} - \overline{Y}_{..})}{\sum_{j=1}^{b} (\overline{Y}_{.j} - \overline{Y}_{..})^2}.$$

This expression suggests that the single degree of freedom proposed by Tukey reflects the linear regression of $\hat{\phi}_i$ on $\hat{\alpha}_i$, or in terms of the parameters the linear regression of ϕ_i on α_i as given by Eq.(2.13) meaning that the slopes change linearly with rows.

$$\phi_i - 1 = \omega \alpha_i + \delta_i \ , \ i = 1 \ , \ \ldots \ , \ a \qquad (2.13)$$

The analysis of variance given in Table 2.2 can be extended to include the information contained in Eq. (2.13) as given in Table 2.3.

Table 2.3 Components of slopes

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Slopes	$SS_{slopes} = \sum_{i=1}^{a}(\hat{\phi}_i - 1)^2 \sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2$	a-1		
Regression of ϕ on μ (concurrence)	$SS_{conc} = \frac{\left(\sum_{i=1}^{a}\hat{\phi}_i(\overline{Y}_{i.} - \overline{Y}_{..})\right)^2}{\sum_{i=1}^{a}(\overline{Y}_{i.} - Y_{..})^2}\sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2$	1	$SS_{conc.}$	$F = \frac{MS_{conc.}}{MS_{nonconc.}}$
Deviations from regression (Non-concurrence)	$SS_{nonconc} = \sum_{i=1}^{a}\hat{\delta}_i^2 \sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2$ where, $\hat{\delta}_i = (\hat{\phi}_i - 1) - \left[\frac{\sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})\hat{\phi}_i}{\sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})^2}\right](\overline{Y}_{i.} - \overline{Y}_{..})$	a-2	$\frac{SS_{nonconc.}}{a-2}$	

Table 2.3 enables us to determine the magnitude of the mean square for slopes due to concurrence. Concurrence exists when $\hat{\phi}_i$ tends to increase systematically as $\hat{\alpha}_i$ increases. This can be confirmed by a plot of $\hat{\phi}_i$ on $\overline{Y}_{i.}$. Distinct linear trend in this plot is definite indication of a tendency for concurrence of all lines in the bundle. The point of concurrence lies on the 45 degree-line of the Y_{ij} versus $\hat{\beta}_j$ plot and has coordinate as given by following equation.

$$\beta_0 = \mu - \frac{1}{\omega} \tag{2.14}$$

Concurrence of the lines can be checked by $F = \frac{MS_{conc.}}{MS_{nonconc.}}$ and $F = \frac{MS_{nonconc}}{MS_{Res}}$. If the data originate from a bundle of concurrent lines, the single degree of freedom denoted "Regression of ϕ on μ" in Table 2.3 will reflect the entire row column

interaction. In that case, it is expected to have insignificant $F = \frac{MS_{nonconc}}{MS_{Res}}$ with degrees of freedom (a-2) and (a-1)(b-2), and significant $F = \frac{MS_{conc.}}{MS_{nonconc.}}$ with degrees of freedom 1 and (a-2). If we get both $F = \frac{MS_{conc.}}{MS_{nonconc.}}$ and $F = \frac{MS_{nonconc}}{MS_{Res}}$ significant, we can still conclude that we have a bundle of concurrent lines but there still remains some unexplained sources of variability. Having concurrent bundle of lines also means to have significant interaction from Tukey's one degree of freedom test.

The columns-linear model implies that there is a linear effect on Y_{ij} from the row main effects, but this linear effect changes from column to column. If the researchers suspect that the model is column linear, then they should change the indices in all formulation regarding slopes and concurrence. For example, indices in the summations should be replaced with j instead of i and with i instead of j in slope equation given in Table 2.2 and all other respective changes should be made in other formulations.

Tukey's one degree of freedom test is applicable when at least a or b is bigger than 2, whereas Mandel's rows-linear model test is applicable when b is bigger than 2 or in the columns- linear case when a is bigger than 2.

Example 1

Propofol is an intravenous anesthetic agent with favorable pharmacokinetic properties and some results indicate that it induces a depression in myocardial contractility resulting in a decrease in arterial blood pressure and cardiac output. Therefore, in patients with coronary artery disease, the magnitude of the arterial pressure reduction is a cause of concern. It is suggested that propofol has cardio protective effect on myocardial ischemia. According to evidence, it is possible that hypercholesterolemia may increase production of oxygen free radicals and lead to

cell injury. Oztekin et al. (2000) performed a study to compare the effects of propofol on cardiac contractile force in normal and hypercholesterolemic isolated rabbit hearts. Both groups of rabbits were given 25, 50, and 100 μM propofol by infusion and their left ventricular pressures (LVP) were measured every ten minutes for an hour. They worked with 20 rabbits with normal hearts and 19 rabbits with hypercholesterolemic heart and compared the effects of different concentrations of drug on LVP. They did not mention any interaction effect between heart type and propofol concentrations on LVP. Hence, it has been decided to determine if any interaction exists between these two factors as a part of this thesis. Since the number of rabbits is not equal for each combination of heart type and propofol dosage, three of them have been randomly chosen from each to be able to get a balanced design, and two-way ANOVA method has been applied on percentage of the change in LVP after one hour. Results obtained by using SPSS are given by Table 2.4.

Table 2.4 Two way ANOVA table for heart type and propofol with three observations per cell

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	p-value
H_Type	2782.083	1	2782.083	19.190	.001[a]
Dosage	5612.480	2	2806.240	19.357	.000[a]
H_Type*Dosage	12782.590	2	6391.295	44.085	.000[a]
Error	1739.705	12	144.975		
Total	22916.858	17			

[a] Significant at 0.01 probability level

According to Table 2.4, heart type (H_Type) and propofol concentrations (Dosage) have significant effect on LVP with p-value < 0.01. These conclusions support the results obtained by Oztekin et al. (2000). Interaction effect is also significant between these two factors with p-value < 0.01.

To see that whether the inference about interaction will change in the case of no replication per cell, two observations from each combination have been randomly discarded and ended with just one observation per cell. Table 2.5 gives the final observations which will be analyzed.

Table 2.5 Percentage of change of LVP measures

	25μM Propofol	50μM Propofol	100μM Propofol
Rabbits with normal heart	-28.64	-25.54	-34.31
Rabbits with hypercholesterolemic heart	-14.99	-66.71	30.86

It should be noted that in the case of no interaction, all of the lines in Figure 2.1 are expected to be parallel. So, it can be said that interaction exists between propofol concentrations and heart type.

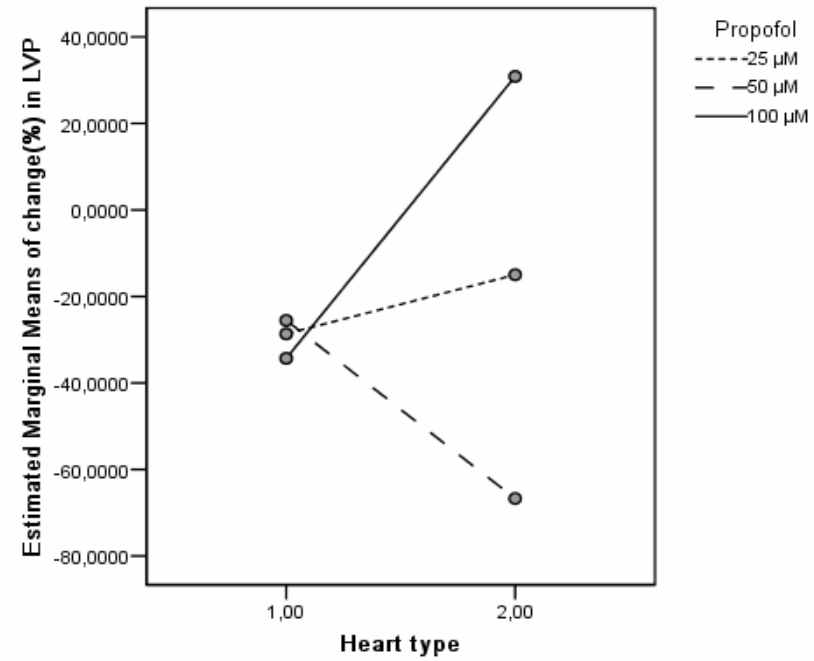

Figure 2.1 Plot of percentage of change in LVP versus heart type for each propofol concentration

Structure of this interaction can be a function of product of two main effects since all the lines intercept. So, Tukey's one degree of freedom test of non-additivity would be the right test to test interaction. Results of this test given in Table 2.6 support this assumption with significant p-value = 0.023 for F = 736.212. So, it is concluded that effect of propofol concentrations on percentage of change in LVP values differ for rabbits with normal heart and hypercholesterolemic heart. "Several in vitro studies have showed that propofol decreased contractile function of left ventricle in a dosage dependent fashion"(Oztekin et al., 2000). Figure 2.1 suggests that while 100 μM

propofol caused more reduction of left ventricle pressure at normal hearts, the same dosage of propofol has caused increment of left ventricle pressure at hypercholesterolemic hearts. This is the implication of cardio protective effect of 100 μM propofol level at hypercholesterolemic hearts.

Table 2.6 The analysis of variance table for Tukey's one degree of freedom test

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	p-value
Hearttype	236.254	1	236.254		
Propofol	1977.875	2	988.940		
Non-additivity	2824.033	1	2824.033	736.212	0.023[a]
Error	3.840	1	3.840		
Total	5041.920	5			

[a] Significant at 0.05 probability level

Example 2

"Insulin-like growth factors (IGFs) are multifunctional peptides that regulate cell proliferation and differentiation. The main source of IGF-I in the serum is liver, but many other tissues synthesize it and are sensitive to its tropic action. "IGF-I appears to influence neuronal structure and functions throughout the life span. It has been shown to have the ability to preserve nerve cell function and promote nerve growth in experimental studies" (PDR health, 2005). In the circulation, IGF-I binds mainly to the main IGF binding protein, IGFBP-3 both of which are dependent on growth hormone, but are also affected by age, sex, and nutritional status.

> Renehan, Zwahlen, Minder, O'Dwyer, Shalet, & Egger. (2004) state that "*Concentrations of both of these peptides vary greatly between individuals, which might effect the distribution of cancer risk in population. Results of early studies on risk of prostate, breast, colorectal and lung cancer suggested that high circulating IGF-I concentrations are associated with an increased risk of cancer, where as high IGFBP-3 concentrations are associated with a decreased risk"*.

IGF-I and IGFBP-3 concentrations can be measured easily in blood, and according to Renehan et al. (2004) they might be useful in assessment of cancer risk.

Freake, Govoni, Guda, Huang, & Zinn (2001) emphasize the role of the both thyroid hormone (triiodo-L-thyronine, T3) and zinc in growth and development. "The T3 receptor is thought to require zinc to adopt its biologically active confirmation. Some of the effects of zinc deficiency , therefore, may be due to loss of zinc from the T3 receptor and impairment of T3 action" (Freake et al., 2001). This possibility was investigated in growing rats by examining the effects of hypothyroidism and hyperthyroidism in zinc-deficient, pair fed and control rats by Freake et al. (2001). Because both T3 and zinc are known to regulate growth, they tried to determine the effect of both of T3 and zinc on IGF-I and IGFBP-3 as well as the effects on other things such as the growth hormone, total food consumed, final body weight and heart weight. According to their analysis, both zinc and T3 have significant effects on IGF-I, while just zinc appears to have significant effect on IGFBP-3. Both hypothyroidism and hyperthyroidism reduced serum IGF-I and a greater reduction due to hyperthyroidism was apparent in zinc-deficient rats. They used 3 or 4 rats for each combination of nine factor levels. They included the means of each measurement taken from the rats in the paper. So, these means were used for two things in this thesis: 1) To determine if there is any interaction between T3 and zinc on IGF-I and IGFBP-3, 2) To see if the same conclusion about the main effects will be reached as Freake et al.(2001). Below the mean values of IGF-I for each combination of zinc and T3 levels are given.

Table 2.7 Mean values of IGF-I

	Zinc deficient	Pair-fed	Control
Hypothyroid	509	756	943
Euthyroid	949	1585	1912
Hyperthyroid	298	935	1507

Figure 2.2 suggests that interaction exists between T3 and zinc. There also seems a linear effect of zinc status on IGF-I values. Interaction between T3 and zinc may be due to the differences of the slopes of the lines corresponding to the levels of T3 (thyroid). This is the indication of rows-linear model.

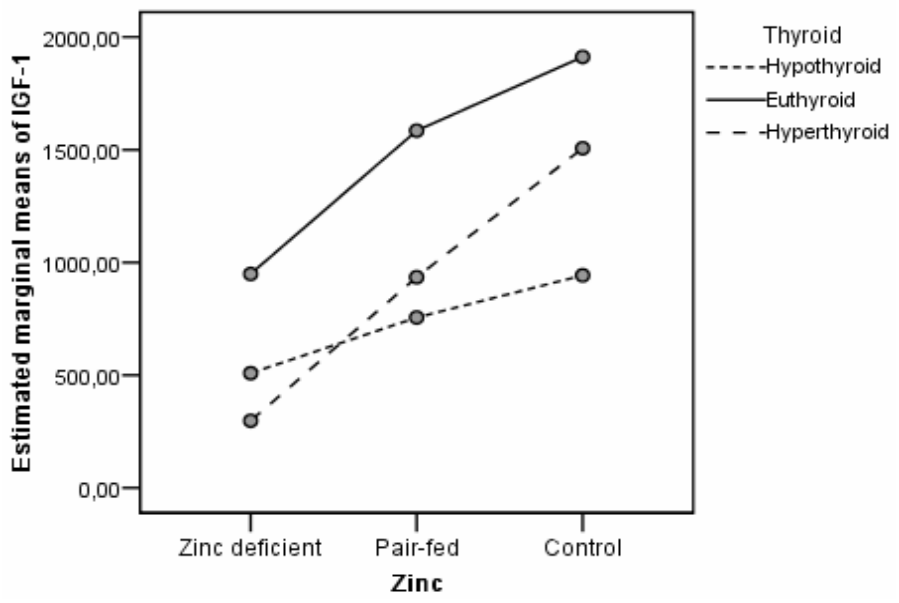

Figure 2.2 Plot of means of IGF-I versus the levels of zinc for the levels of thyroid

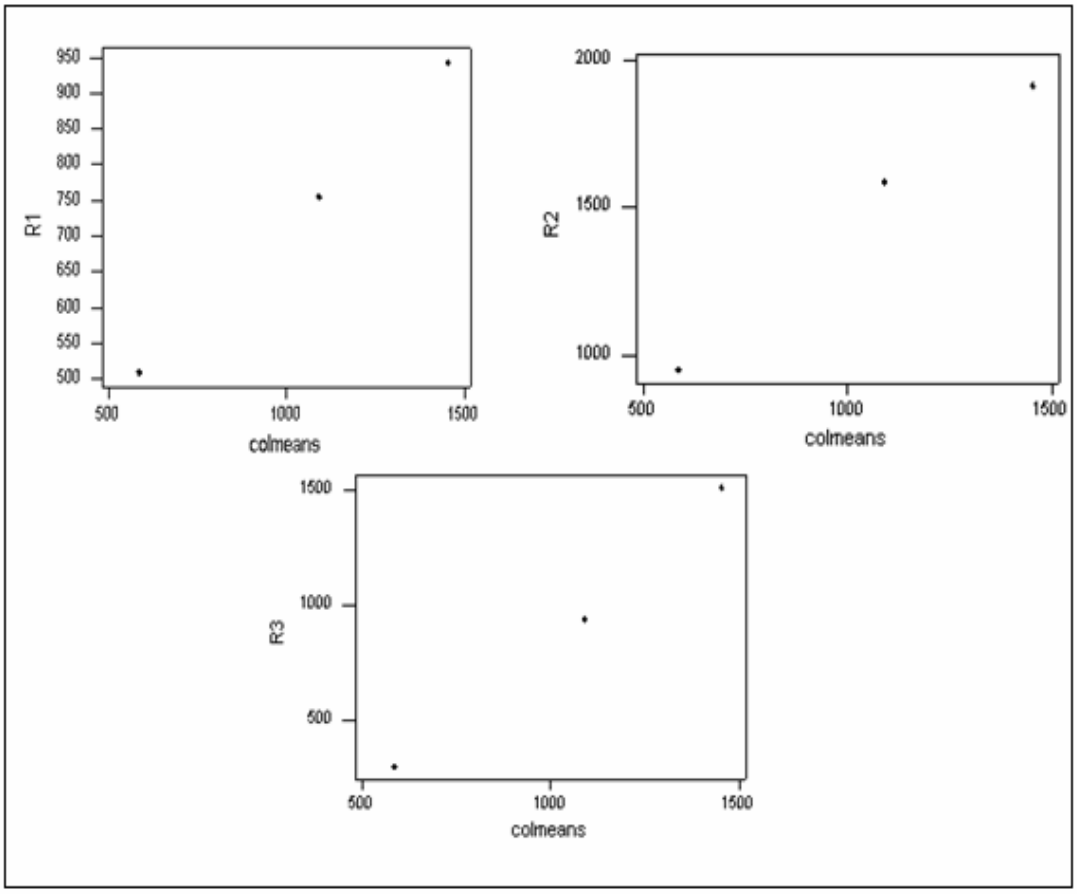

Figure 2.3 Plot of each row of the table versus the column averages $\overline{Y}_{.j}$

Figure 2.3 implies that the assumption about rows-linear model may be true. Table 2.8 gives the slopes of the lines plotted for the values corresponding to i-th row of T3 versus β_j for j =1,2,3, and suggests that the three slope values are not identical. Since the standard deviation between these slopes, 0.454, is appreciably larger than the standard errors of the individual slopes, it is concluded that these differences are not random. So, it is logical to test interaction by Mandel's rows-linear model.

Table 2.8 The slopes of the lines

Row	Slope	Standard error of slope
1	0.499	0.008
2	1.118	0.097
3	1.383	0.087

Based on Table 2.9, there seems to exist interaction between T3 and zinc on IGF-I, and the differences in the slopes of rows are the cause of this interaction. p-value = 0.041 for $F = \frac{MS_{slopes}}{MS_{Res}} = 23.261$ provides evidence for this conclusion at %5 significance level. Since p-value = 0.025 for $F = \frac{MS_{nonconc}}{MS_{Res}} = 37.476$ is significant at %5 significance level, it can be concluded that these rows do not concur. Insignificant p-value = 0.709 for $F = \frac{MS_{Conc}}{MS_{Noncon}} = 0.241$ also supports the non-concurrence of rows.

Table 2.9 The analysis of variance table for Mandel's test for non-additivity

Source of variation	Sum of squares	Degrees of freedom	Mean Square	F	p-value
Mean	9805248.000	1			
Thyroid (T3)	911344.900	2	455672.400	135.162	0.007
Zinc	1142337.000	2	571168.400	169.421	0.005
Slopes	156841.200	2	78420.590	23.261	0.041[a]
Concurrence	30496.020	1	30496.020	0.241	0.709[b]
Non-concurrence	126345.100	1	126345.100	37.476	0.025[a]
Residual	6742.606	2	3371.303		
Total	12022513.710				

[a] Statistically significant at 0.05 probability level
[b] Not statistically significant

If Tukey's one degree of freedom test had been applied on these data, insignificant interaction would have gotten from the analysis. This conclusion can be made even without applying the Tukey's test since concurrence also means linear effect of rows and linear effect of columns on all Y_{ij}. This is equivalent defining the interaction as T3 x zinc. Even though interaction between zinc and T3 is because of the difference in the slopes of each row, these rows do not seem to concur. Significant main effects for T3 and zinc with p-values 0.007 and 0.005, respectively support the results obtained by Freake et al. (2001).

When the mean values of the IGFBP-3 given with Table 2.10 are evaluated, neither rows-linear nor columns-linear model seems to be appropriate. Tukey's one degree of freedom test also results with insignificant interaction. But, with p-value = 0.040, zinc has significant main effect on IGFBP-3 which also supports the results obtained by Freake et al. (2001).

Table 2.10 Mean values of IGFBP-3

	Zinc deficient	Pair-fed	Control
Hypothyroid	104	112	154
Euthyroid	107	142	206
Hyperthyroid	60	115	145

2.3.2. Models with Interaction Terms that Do Not Depend Any of The Main Effects

"The structure of interaction is very important, since misspecification will decrease the power of test, even if the magnitude of existing non-additivity is large" (Boik, 1993a). Unfortunately, it is really hard to know in advance how the structure of interaction will be shown. Tukey's one degree of freedom test and Mandel's rows-linear model test depend on the known function of row and column effects. However, for many situations, the interaction effect may not be a function of row and column effects but a function of two unknown parameters as given by Eq.(2.15)

$$Y_{ij} = \mu + \alpha_i + \beta_j + \lambda\varphi_i\upsilon_j + \varepsilon_{ij} \quad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \end{cases} \tag{2.15}$$

with the normality assumption of error and following restrictions.

$$\sum_{i=1}^{a}\alpha_i = 0,\ \sum_{j=1}^{b}\beta_j = 0,\ \sum_{i=1}^{a}\varphi_i = 0,\ \sum_{j=1}^{b}\upsilon_j = 0,\ \sum_{i=1}^{a}\varphi_i^2 = 1,\ \sum_{j=1}^{b}\upsilon_j^2 = 1$$

As in the model given in Eq.(2.5), for this model testing interaction means testing H_0: $\lambda = 0$.

If the model in Eq.(2.5) is the correct one, then the methods based on the Tukey's one degree of freedom test will undoubtedly have good power, but they may have very poor power if the model given in Eq.(2.15) is the real model. On the other hand, if a good test were available for the model (2.15), it would definitely be more powerful. "Hence, it may be more desirable to use the model given by Eq.(2.15) instead of Eq.(2.5) unless one is very certain that the model (2.5) applies" (Johnson & Graybill, 1972).

The model in Eq.(2.15) has been discussed by Mandel (1971). Sticking to the notations used in this thesis so far, Mandel (1971)'s model can be defined as;

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{k=1}^{r}\lambda_k\varphi_{ki}\upsilon_{kj} + \varepsilon_{ij} \quad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \end{cases} \tag{2.16}$$

This new model is based on the singular value decomposition of the interaction matrix $\boldsymbol{\tau} = \{\tau_{ij}\}$ where $\tau_{ij} = \sum_{k=1}^{r}\lambda_k\varphi_{ki}\upsilon_{kj}$. As seen in (2.16), k goes through r which is the rank of the interaction matrix $\boldsymbol{\tau}$. $\boldsymbol{\varphi} = \{\varphi_{ki}\}$ and $\boldsymbol{\upsilon} = \{\upsilon_{kj}\}$ are orthonormal matrices of order a x r and b x r respectively; and λ_k^2 for k = 1, . . ., r are the nonzero eigen values of $\boldsymbol{\tau}'\boldsymbol{\tau}$ ordered from the largest to smallest. Because the rank of the interaction is not known in practice, "Mandel replaced r by r_a which is the hypothesized rank of the interaction under H_a" (Boik, 1993a). .The resulting interaction is

$$\tau_{ij} = \sum_{k=1}^{r_a} \lambda_k \varphi_{ki} \upsilon_{kj} \qquad (2.17)$$

If the data is additive, i.e. there is no significant interaction, the interaction has rank zero. Having zero rank means there is no eigen value different from zero. So, H_0: r = 0 equals to test H_0: $\lambda_k = 0$ for all k's. Snee (1982) has indicated that frequently only one of $\lambda\varphi\upsilon$ term is needed to describe the interaction (i.e., λ_k^2 is not significantly different from zero for $k \geq 2$). Then the model gets equal to Eq.(2.15).

The model given by Eq.(2.15) contains special cases such as; Mandel (1961)'s rows-linear (υ_j proportional to β_j) and columns-linear models (φ_i proportional to α_i), given by Eq.(2.18) and Eq.(2.19), respectively and Tukey (1949)'s model (υ_j proportional to β_j and φ_i proportional to α_i), which is also referred as the concurrent model by Mandel (1961).

Rows-linear model:

$$Y_{ij} = \mu_i + (1 + \lambda\varphi_i / \beta)\beta_j + \varepsilon_{ij} \text{ with } \upsilon_j = \beta_j / \beta \qquad (2.18)$$

Columns-linear model:

$$Y_{ij} = \mu_j + (1 + \lambda\upsilon_j / \alpha)\alpha_i + \varepsilon_{ij} \text{ with } \varphi_i = \alpha_i / \alpha \qquad (2.19)$$

Concurrent model :

$$Y_{ij} = Z_0 + \lambda(\mu_i - Z_0)(\mu_j - Z_0) / \alpha\beta + \varepsilon_{ij} \text{ with } \varphi_i = \alpha_i / \alpha \text{ and } \upsilon_j = \beta_j / \beta \qquad (2.20)$$

where

$$\beta^2 = \sum_{j=1}^{b} \beta_j^2 \,,\ \alpha^2 = \sum_{i=1}^{a} \alpha_i^2 \text{ and } Z_0 = \mu - \frac{\alpha\beta}{\lambda}.$$

Snee (1982) used the model given with Eq.(2.16) to separate interaction effect and non-homogenous variance in two-way tables with one observation per cell.

Johnson & Graybill (1972) also studied the model in Eq.(2.15). The basic difference of their paper from previous studies is that they derive the distributional properties for the test of no-interaction hypothesis. They derive the following maximum likelihood estimators of the parameters in the model (2.15) instead of least squares estimators and then find the likelihood ratio test of the hypothesis H_0: $\lambda = 0$. This test procedure is the same as testing H_0: $r = 0$ versus H_a: $r_a = 1$ in Mandel (1971)'s paper because there is supposed to be one eigen value of $\boldsymbol{\tau}'\boldsymbol{\tau}$ different from zero to test.

$$\hat{\alpha}_i = \overline{Y}_{i.} - \overline{Y}_{..},\ \hat{\beta}_j = \overline{Y}_{.j} - \overline{Y}_{..}$$

$\hat{\lambda}^2 = l_1$ i.e. the largest root of $\mathbf{Z}'\mathbf{Z}$ where $\mathbf{Z}$ is the a x b residual matrix of $\{e_{ij}\}$ with $e_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}$

$\hat{\boldsymbol{\varphi}} = \{\hat{\varphi}_i\}$ is a normalized characteristic vector of $\mathbf{Z}\mathbf{Z}'$ corresponding to root l_1

$\hat{\boldsymbol{\upsilon}} = \{\hat{\upsilon}_j\}$ is a normalized characteristic vector of $\mathbf{Z}'\mathbf{Z}$ corresponding to root l_1

The likelihood ratio (LR) test statistics suggested by Johnson & Graybill (1972) is;

$$\frac{l_1}{l_1 + l_2 + \ldots + l_p} \tag{2.21}$$

where l_i's are the non-zero (characteristics) roots of $\mathbf{Z}'\mathbf{Z}$ among p of which l_1 is the largest one for $p = \min(a-1, b-1)$. H_0: $\lambda = 0$ is rejected in favour of H_a: $\lambda \neq 0$ if $\frac{l_1}{l_1 + l_2 + \ldots + l_p} > K$ for the given Type I error α, where K is the critical point for

the distribution of the LR test statistics. Critical points for the distribution of $\frac{l_1}{l_1 + l_2 + \ldots + l_p}$, for $p \geq 2$, can be found in Johnson & Graybill (1972).

One should note that Johnson & Graybill (1972) gave the critical values with the assumption that $b \leq a$, i.e. $p = b-1$. If researchers have experimental design with $a \leq b$, then they may redesign the table as $b \leq a$ so they will come up with $p = b-1$ or they may look at the critical values table by interchanging the places of rows and columns. Johnson & Graybill (1972) studied the relationship of the model (2.15) with Tukey's one degree of freedom test. They suggested that if the interaction is indeed a function of the two sets of treatments, then Tukey's one degree of freedom test and the others which are outgrowths of the Tukey's test are good tests for showing significant interaction, and in these cases the best way to analyze the data may be to find a transformation that will restore additivity. However, if the interaction is not a function of the row and column factor effects, then Tukey's one degree of freedom test and other related ones may not be the right tests for showing interaction, and transformations may not be suitable, either. In these cases, the test statistic given by Eq.(2.21) would be preferred.

Example

Following table gives the mean values of growth hormone for the levels of zinc and T3 obtained by Freake et al.(2001).

Table 2.11 Mean values of growth hormone

	Zinc deficient	Pair-fed	Control
Hypothyroid	17.6	10.6	3.0
Euthyroid	5.6	5.4	16.2
Hyperthyroid	8.9	4.8	4.2

Figure 2.4 indicates interaction between zinc and T3 on the mean values of growth hormone. Zinc deficient rats with hypothyroid seem to have larger growth hormone values where as zinc deficient rats with euthyroid seem to have lower growth hormone values. This interaction could be detected neither by rows- linear

model nor by Tukey's one degree of freedom test. So, if it is assumed that interaction does not depend any of the main effects, Johnson & Graybill's likelihood ratio test can be applied.

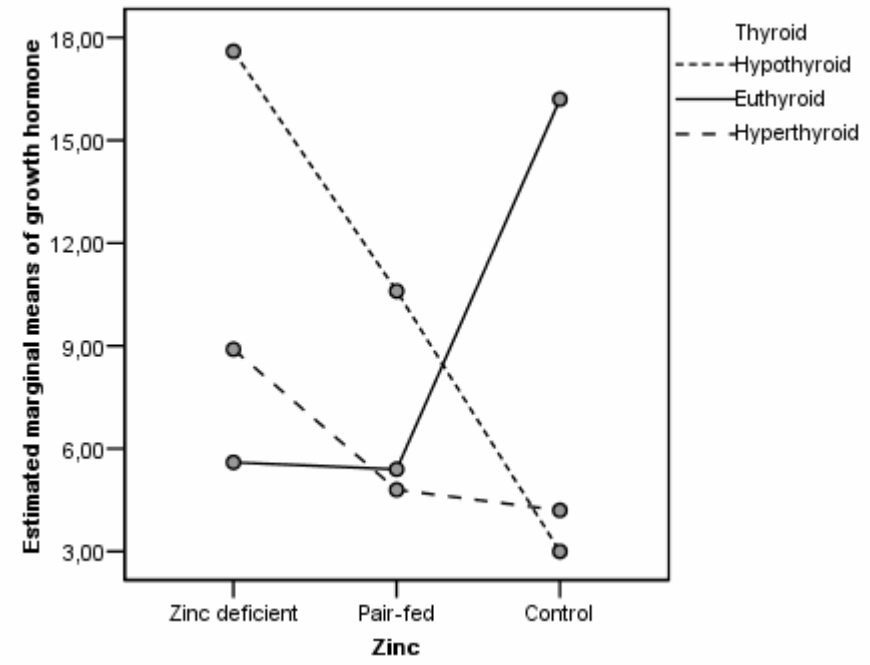

Figure 2.4 Plot of mean values of growth hormone versus the levels of Zinc for the levels of T3

Residual matrix for the mean values of growth hormone is given below.

$$\mathbf{Z} = \begin{bmatrix} 4.978 & 1.744 & -6.722 \\ -5.689 & -2.122 & 7.811 \\ 0.711 & 0.378 & -1.089 \end{bmatrix}$$

The sample eigen values of $\mathbf{Z'Z}$ are 172.703 and 0.021. So, Johnson & Graybill's likelihood ratio test is equal to

$$\frac{l_1}{l_1 + \ldots + l_{b-1}} = \frac{172.703}{172.203 + 0.021} = 0.9998$$

The 5% critical value for this test is 0. 9994 which also provides significant evidence for interaction since test statistics is bigger than critical value. So, there exists interaction between zinc and T3 on the growth hormone values.

2.3.3 Models with Unknown Interaction Structure

"The power function of testing non-additivity ought to depend on the magnitude of the existing non-additivity" (Boik, 1993a). Power functions of all of the discussed tests above depend on the magnitude of the non-additivity (say Λ) as well as how the specified interaction structure fit the existing non-additivity. A poor fit translates into low power even for the large magnitude of non-additivity. Srivastava & Carter (1980)'s analytic results suggest that the Johnson & Graybill's likelihood ratio test has lower power, if the alternative is misspecified (i.e. the interaction matrix has rank = 2 or more). Tusell (1990) suggests that in the case of rank one interaction Johnson & Graybill's likelihood ratio test handles best.

Because of the problems of having misspecified non-additivity, Tusell (1990) proposed an innovative test. His procedure consists of employing well-known likelihood ratio sphericity (LRS) test as a test of additivity. A vector of random variable, say Y, is usually said to have a spherical distribution if its covariance matrix $\sum \mathbf{Y}$ is of the form k**I**, i.e. if all variances are equal, where k is any constant and **I** is the identity matrix. Testing the H_0: $\sum \mathbf{Y} = \sigma^2 \mathbf{I_p}$ against the H_a: $\sum \mathbf{Y} \neq \sigma^2 \mathbf{I_p}$ is called "test for sphericity" with p = b-1 = min(a-1,b-1) with the assumption that $b \leq a$ (the roles of rows and columns interchangeable). The idea underlying the test proposed by Tusell (1990) is to perform successive linear operations on the rows and columns of $\mathbf{Y} = \{Y_{ij}\}$ so as to annihilate the μ, α_i and β_j parts of Y_{ij}. Tusell (1990) suggests that if interaction is not present, choosing adequately the linear operations it is ended up with centered and independent vectors with diagonal covariance matrix with equal variances. For $b \leq a$, his test statistics is

$$\Psi = \frac{|\mathbf{Z}'\mathbf{Z}|}{(\text{trace}(\mathbf{Z}'\mathbf{Z})/(b-1))^{b-1}} = \frac{l_1 l_2 \ldots l_{b-1}}{\left[(b-1)^{-1}\sum_{k=1}^{b-1} l_k\right]^{b-1}} \qquad (2.22)$$

where l_k's are non-zero eigen values of $\mathbf{Z}'\mathbf{Z}$. $\mathbf{Z} = \mathbf{X}'\mathbf{U_b}$ is (a-1) x (b-1) matrix for $\mathbf{X} = \mathbf{Y}'\mathbf{U_a}$. $\mathbf{U_a}$ and $\mathbf{U_b}$ can be any a x (a-1) and b x (b-1) orthonormal matrices of contrast coefficients. Ψ can be compared to the percentage points provided by Nagarsenker and Pillai (1975). Lower values of the Ψ lead to rejection of the hypothesis of no-interaction.

Tusell (1990) showed by simulation that the sphericity test would be a better choice if one is fairly confident that interaction, if it exists, is not nearly rank one. If this test is used in the case of rank one interaction, the result will be a decrease in power.

Boik (1993a) proposed another test, which is less subject to the model misspecification because it does not assume a particular structure in interaction terms. Boik's test called as the "Locally Best Invariant Test (LBI) of Additivity" is derived to maximize power for detecting alternatives in a neighborhood of non-additivity with zero magnitude. Boik (1993a) suggests that the LBI test, by construction, is the most powerful invariant test when magnitude of non-additivity is small. "One simple index of non-additivity magnitude is the squared scaled (by σ^2) norm of the interaction" (Boik, 1993a).

$$\Lambda = \frac{1}{\sigma^2}\sum_{k=1}^{r}\lambda_k \qquad (2.23)$$

Boik (1993b) showed that with alternative of $r_a = 1$, LBI test have essentially same power for large Λ as well as for small Λ even tough likelihood ratio test of Johnson & Graybill is especially designed for r = 1. When $r \geq 2$, LBI test is more powerful than Johnson & Graybill's likelihood ratio test. Power comparison of Johnson & Graybill's LR test, Tusell's LRS test and LBI test was performed by Boik (1993b).

"The test problem is invariant under scalar multiplication and orthogonal rotation of residual matrix" (Boik, 1993a). Consequently, invariant tests depend on the data only through the eigen values of $\frac{\mathbf{Z}'\mathbf{Z}}{\mathrm{tr}(\mathbf{Z}'\mathbf{Z})}$ which are defined as

$L_1 > L_2 > L_3 > \ldots > Lp > 0$ for

$$L_k = \frac{l_k}{l_1 + l_2 + \ldots + l_p} = \frac{\mathrm{eig}_k(\mathbf{Z}'\mathbf{Z})}{\mathrm{tr}(\mathbf{Z}'\mathbf{Z})} \qquad (2.24)$$

where

$\mathbf{Z} = \{e_{ij}\}$, $e_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}$ is obtained by fitting the additive model in (2.3),

$\mathrm{tr}(\mathbf{Z}'\mathbf{Z})$ = the trace of the $\mathbf{Z}'\mathbf{Z}$ matrix,

p = the rank of the residual matrix $\mathbf{Z}$ which equals to min (a-1, b-1)

l_k = k-th ordered eigen value of $\mathbf{Z}'\mathbf{Z}$, for k = 1 , . . . , p.

Power functions of invariant tests depend only on the non-zero eigen values of $\mathbf{Z}'\mathbf{Z}/\sigma^2$; i.e, $l_1/\sigma^2 \geq l_2/\sigma^2 \geq \ldots l_r/\sigma^2 > 0$. "That is, the power functions of the invariant tests depend on the magnitude of non-additivity but not on the direction of non-additivity" (Boik, 1993a). "Non-invariant tests require a priori models with which to model direction of non-additivity as well as magnitude" (Boik, 1993b).The Johnson & Graybill's LR test with the alternative H_a: $r_a = 1$ rejects the H_0 for large L_1 in Eq. (2.24) which is the same as Eq.(2.21). Thus, likelihood ratio test is invariant. Tusell's LRS test is also invariant. However as mentioned by Boik (1993b), Tukey's one degree of freedom test and Mandel's bundle of lines tests are non-invariant. "In the absence of prior information about the direction of non-additivity, invariant tests are attractive to non-invariant alternatives" (Boik, 1993b).

Boik (1993a) showed that if the coefficient of variation among the eigen values of $\mathbf{E(Z')E(Z)}$ is not too small, then LBI test of additivity will reject H_0 for small

$$\hat{\varepsilon}_{p,q} = \frac{[\mathrm{tr}(\mathbf{Z'Z})]^2}{p\,\mathrm{tr}[(\mathbf{Z'Z})^2]} = \left[p\sum_{k=1}^{p}(L_k - p^{-1})^2 + 1 \right]^{-1} \qquad (2.25)$$

where

p = min(a-1,b-1) with the restriction of being greater than or equal to 2,

q = max(a-1,b-1).

This test statistics is a monotonic decreasing function of the coefficient of variation among the eigen values of $\mathbf{Z'Z}$. A large coefficient of variation among the sample eigen values is evidence against additivity.

John (1971) and Sugiura (1972) showed that if row effects in the model (2.3) (with n = 1) are random and a ≥ b, then Eq. (2.25) is the LBI test of sphericity. So, test statistics given with Eq.(2.25) can be calculated by using SPSS. $\hat{\varepsilon}_{p,q}$ is equal to "Greenhouse-Geisser Epsilon" sphericity statistics, which is multiplied with the between-groups degrees of freedom to correct univariate F test with lack of sphericity, and can be calculated by "Repeated measures" statement under "General Linear Model" command in SPSS. But, data should be arranged as a ≥ b. p-value for $\hat{\varepsilon}_{p,q}$ can be calculated by an algorithm written by Boik with FORTRAN, and this algorithm can be obtained from the web page of Statlib (1989). It should be reminded that Johnson & Graybill's LR test, Boik's LBI test and Tusell's LRS tests are applicable when p = min (a-1,b-1) ≥ 2.

Example

If it is assumed that there is no prior information about the structure of the interaction between zinc and T3 on the mean values of growth hormone and LBI test and Tusell's LRS test are applied on these values, respectively, following results are obtained:

$$\hat{\varepsilon}_{p,q} = \frac{[\mathrm{tr}(\mathbf{Z}'\mathbf{Z})]^2}{p[\mathrm{tr}(\mathbf{Z}'\mathbf{Z})^2]} \Rightarrow \hat{\varepsilon}_{2,2} = \frac{(172.724)^2}{2x29826.4} = 0.50$$

$\hat{\varepsilon}_{2,2}$ statistic with p-value = 0.023 provides significant evidence for non-additivity at $\alpha = 0.05$ which is supported by large coefficient variation in the sample eigen values of $\mathbf{Z}'\mathbf{Z}$: 172.703, 0.021.

$$\frac{|\mathbf{Z}'\mathbf{Z}|}{(\mathrm{trace}(\mathbf{Z}'\mathbf{Z})/(b-1))^{b-1}} = \frac{l_1 l_2 \ldots l_{b-1}}{\left[(b-1)^{-1}\sum_{k=1}^{b-1} l_k\right]^{b-1}} = \frac{172.703x0.021}{[(1/2)x(172.703+0.021)]^2} = 0.002$$

p-value for Tusell's LRS test statistic is also less then 0.05 which is the indication of significant interaction.

2.4 Additional Example

Measurements are the end product of a measuring process which involves instrumentation that must be calibrated and kept in a state of proper calibration throughout the process. This is the view of measurement as a process. There is a second view, measurement as a relation, in which desired answer, say P, is obtained by establishing the relationship with measured characteristic, say M. Calibration experiment for the establisment of this relation can be done by measuring M for a series of samples for different values of P, covering all the range interested in. After

that, this relation can be visualized by a plot of M which is represented by the y-axis versus P which is represented by the x-axis. This curve is called “calibration curve”.

The usefulness of a measurement process depends largely on its stability both in time and space. Stability in time can be checked by control charts where as stability in space needs another experiment called “interlaboratory testing” which consists in sending materials to a number of laboratories and collecting, analyzing and interpreting the results obtained by these laboratories. Experiment given in Mandel (1991, p: 186) is an example to an interlaboratory studies. Aim in this experiment is to determine pentosans in wood pulp by measuring absorbency, a colorimetric result obtained on a distillate of the pulp, by means of a method known as the “orcinol”. Here P is the pentosan concentrations and M is the absorbency. Each laboratory received portions of each of nine materials corresponding to the different level of pentosan concentrations, and performed the test on each replicate. Original experiment includes three replicates on each material per laboratory. As in Mandel (1991, p.186), the mean of these replicates on each material per laboratory will be used to determine whether interaction exists between laboratory and pentosan concentrations on mean absorbency, i.e. if calibration curves corresponding to the laboratories in an interlaboratory test show differences from each other or effect of pentosan concentrations on absorbency is the same for each laboratory. Table 2.12 gives the results of this experiment.

Table 2.12 Mean of absorbance values

	A	B	C	D	E	F	G	H	I
Lab1	0.457	0.900	1.450	1.307	1.900	4.147	5.560	10.757	16.750
Lab2	0.410	0.833	1.120	1.253	1.960	4.103	5.260	9.980	16.083
Lab3	0.510	0.923	1.117	1.350	2.053	4.143	5.177	10.117	16.010
Lab4	0.383	0.947	1.137	1.290	2.043	4.207	5.200	10.723	16.770
Lab5	0.490	0.847	0.980	1.230	1.953	4.590	4.987	10.377	15.607
Lab6	0.413	0.893	1.113	1.307	1.993	3.897	4.877	9.587	14.940
Lab7	0.170	0.866	0.800	1.143	1.963	4.183	5.230	11.267	18.367

Assuming the linear effect of each material on absorbency values for each row, Mandel (1991, p.186) applied rows-linear model on this data which turned out to be concurrent at the point 1.78. With the point of concurrence to the left, it can be said

that there must be gradual widening of the bundle of straight lines from left to right, resulting in an increasing standard deviation between the laboratories. This is apparent in Figure 2.5. Since having significant concurrence also means to have significant interaction by Tukey's test, this test has not been performed on this data set. To see the results of invariant tests, these tests have been applied

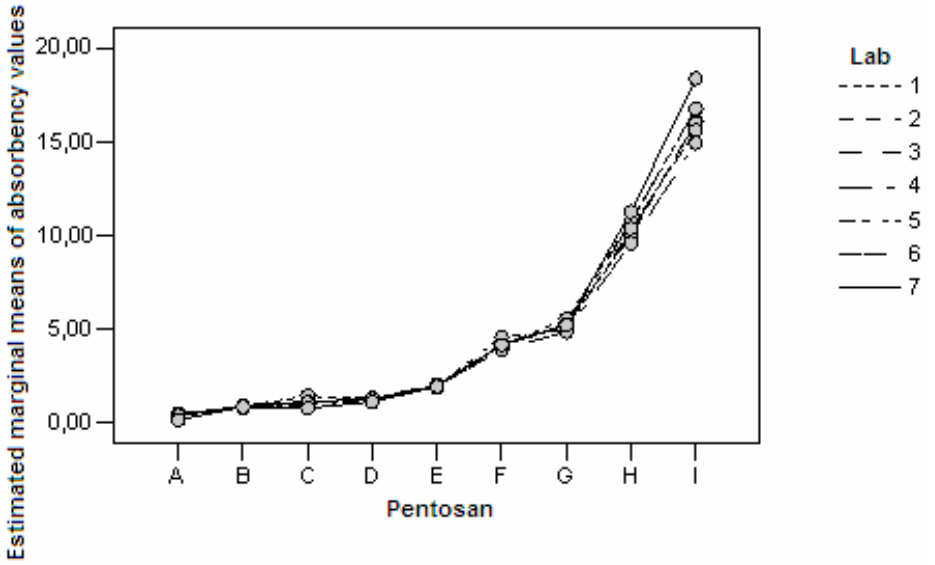

Figure 2.5 Plot of absorbency values of pentosans for each lab

Below is the residual matrix for Table 2.12.

$$\mathbf{Z} = \begin{bmatrix} -0.109 & -0.149 & 0.186 & -0.123 & -0.243 & -0.196 & 0.214 & 0.194 & 0.227 \\ 0.091 & 0.032 & 0.103 & 0.069 & 0.065 & 0.007 & 0.161 & -0.336 & -0.193 \\ 0.147 & 0.077 & 0.056 & 0.122 & 0.114 & 0.003 & 0.034 & -0.243 & -0.309 \\ -0.125 & -0.043 & -0.069 & -0.082 & -0.041 & -0.078 & -0.088 & 0.219 & 0.306 \\ 0.164 & 0.039 & -0.044 & 0.040 & 0.051 & 0.488 & -0.119 & 0.055 & -0.675 \\ 0.314 & 0.311 & 0.316 & 0.344 & 0.318 & 0.021 & -0.002 & -0.508 & -1.115 \\ -0.481 & -0.267 & -0.549 & -0.372 & -0.264 & -0.245 & -0.201 & 0.619 & 1.759 \end{bmatrix}$$

For the eigen values of the matrix $\mathbf{Z}'\mathbf{Z}$: 7.339, 0.438, 0.226, 0.070, 0.003, 0.001, Johnson & Graybill's LR test statistic is equal to

$$\frac{l_1}{l_1 + \ldots + l_{b-1}} = \frac{7.339}{7.339 + 0.438 + 0.226 + 0.070 + 0.003 + 0.001} = 0.9086$$

which indicates significant interaction at α=0.01, since it exceeds the critical value 0.5523.

LBI test statistic for this data set is

$$\hat{\varepsilon}_{p,q} = \frac{[\mathrm{tr}(\mathbf{Z}'\mathbf{Z})]^2}{p\,[\mathrm{tr}(\mathbf{Z}'\mathbf{Z})^2]} \Rightarrow \hat{\varepsilon}_{6,8} = \frac{(8.077)^2}{6\mathrm{x}54.107} = 0.201 .$$

p-value for this statistic is less than 0.01 which also provides significant evidence to conclude for significant interaction.

$$\frac{|\mathbf{Z}'\mathbf{Z}|}{(\mathrm{trace}(\mathbf{Z}'\mathbf{Z})/(b-1))^{b-1}} = \frac{l_1 l_2 \ldots l_{b-1}}{\left[(b-1)^{-1}\sum_{k=1}^{b-1} l_k\right]^{b-1}}$$

$$= \frac{7.339\mathrm{x}0.438\mathrm{x}0.226\mathrm{x}0.070\mathrm{x}0.003\mathrm{x}0.001}{[(1/6)\mathrm{x}(8.077)]^6}$$

$$= 3{,}18899\mathrm{E}-44$$

It is apparent that Tusell's LRS test statistic = 3,18899E-44 provides significant evidence for interaction.

Apart from the tests already applied by Mandel (1991, p: 186), all of the invariant tests indicate significant interaction between laboratory and pentosan concentrations on absorbency values which is the implication of non-stability of calibration curves. So, it can be concluded that relation with pentosan and absorbency values differ for each laboratory as shown in Figure 2.5.

CHAPTER THREE
LOG-LINEAR MODELS

This chapter will include the some information about contingency tables and sampling distribution for cell counts in contingency tables. After giving this information in Section 3.1 and 3.2, log-linear models will be described in Section 3.3. This final section will contain subsections describing the possible log-linear models for two-way and three-way contingency tables.

3.1 Contingency Tables

Contingency table also referred as cross-classified table basically is the table of counts. The units of a sampled population are cross-classified according to each of several categorical variables. Each variable represents one dimension of the table. If there are two variables, one variable represents row, whereas other represents columns. As mentioned in Chapter 1, these tables are called two-way, three-way or etc. according to number of variables. If a table is formed by two categorical variables each of which has two levels then this two-way contingency table is also called 2 x 2 table. Two-way and three-way contingency tables will be the focus of this thesis.

"Contingency table analysis is rooted in the turn-of the century work of Karl Pearson and George Udny Yule, who introduced the cross-product, or odds ratio, as a formal statistical tool" (Fienberg, 2000). "Many important figures in the history of statistics, including R. A. Fisher, Jerzy Neyman, William Cochran, and Maurice Bartlett, made important contributions to the categorical data literature" (Agresti, 1990, p: 1).

There is a relationship between the concept of independence and interaction in contingency tables. This relationship has been already mentioned by Bartlett (1935) who states that "the testing independence in a 2 x 2 table may be regarded as testing the interaction between two classifications". Interaction in contingency tables

received much more attention with his paper. Bartlett (1935) utilized Yule's cross-product ratio to define the notion of second order interaction in a 2 x 2 x 2 contingency table. Since Bartlett (1935)'s paper a very large number of papers have been published on testing interactions in contingency tables including the papers written by Norton (1945), Roy & Kastenbaum (1956), Darroch (1962, 1974), Plackett (1962), Good (1963), Goodman (1963, 1964a, 1964b, 1970), Bhapkar & Koch (1968), Grizzle, Starmer, & Koch (1969), Ku et al. (1971), Ku & Kullback (1974), and Dana & Gokhale (1983).

3.2 Sampling Distributions for Cell Counts

Let $S_1, \ldots, S_n$ be an independent and identically distributed sample of size $n \geq 1$. The statistic interested is $(N_1, \ldots, N_M)^T$. N_j, for $j = 1, \ldots, M$, denotes the cell count in the j-th cell of contingency table. Poisson, multinomial and product multinomial sampling distributions are mostly used sampling distributions for the statistic $(N, \ldots, N_M)^T$.

3.2.1 Poisson Sampling

A set of Poisson processes, one for each cell in the contingency table, is observed over a fixed period of time, with no prior knowledge regarding the total number of observations to be taken. Hence, for Poisson sampling the total $\sum_{j=1}^{M} N_j$ is random.

3.2.2 Multinomial Sampling

If total $\sum_{j=1}^{M} N_j$ is fixed then the sampling distribution of the statistic $(N_1, \ldots, N_M)^T$ becomes multinomial.

3.2.3 Product (Independent) Multinomial Sampling

For each category of the row variable, multinomial sample of fixed size is taken and each unit of the sample is classified according to its category for the column variable (the roles of rows and columns can be interchanged).

The study in this thesis is based on two assumptions:

1) The statistic $(N_1, \ldots, N_M)^T$ has multinomial sampling distribution with the probability function given in Eq.(3.1)

$$P(N_1 = n_1,\ldots,N_M = n_M) = \frac{n!}{n_1!\ldots n_M!} p_1(\theta)^{n1} \ldots p_M(\theta)^{nM} \qquad (3.1)$$

for n_j such that $\sum_{j=1}^{M} n_j = n$.

2) $p_j(\theta) = \frac{m_j(\theta)}{\sum_{j=1}^{M} m_j(\theta)}$, for j = 1, …, M, belongs to the general class of log-linear models.

3.3 Log-linear Models

Roy & Kastenbaum (1956)'s article forms the basis of log-linear model approach to contingency tables. The articles by Birch (1963), Darroch (1962), Good (1963), Goodman (1963, 1970), Gokhale (1972) led to an integrated theory and methodology for the contingency tables based on log-linear models. Lately the articles written by Christensen & Utts (1992), Christensen (2000), Cressie & Pardo (2000, 2002), Fienberg (2000), and Pardo & Pardo (2005) contribute to the literature of log-linear models. There are culminating series of books published just about log-linear models or included the log-linear models as separate chapters. The books by Plackett (1974),

Bishop, Fienberg, & Holland (1975), Fienberg (1980), Knoke & Burke (1980) Read & Cressie (1988), Christensen (1990), Agresti (1990) are some of them.

Log-linear models describe the association patterns among categorical variables by modeling expected cell counts (or equivalently cell probabilities) in contingency tables without distinguishing between independent (explanatory) and response variables. These models are linear in their logarithms.

3.3.1 Model in Matrix Form

If $p_j(\theta)$ belongs to the general class of log-linear models then it will be equal to

$$p_j(\theta) = \frac{m_j(\theta)}{\sum_{j=1}^{M} m_j(\theta)} = \frac{\exp(\mathbf{w_j^T \theta})}{\sum_{j=1}^{M} \exp(\mathbf{w_j^T \theta})}; j = 1, \dots, M \qquad (3.2)$$

with M_0 x 1 parameter vector of $\boldsymbol{\theta} = (\theta_1, ...\theta_{M_0})^T \in \boldsymbol{\Theta} \subseteq \Re^{M_0}$ and 1 x M_0 vector $\mathbf{w_j^T} = \left(w_{1(j)}, ..., w_{M_0(j)}\right)$ which forms the M x M_0 matrix $\mathbf{W} = (w_1, ..., w_M)^T$ assumed to have full column rank $M_0 \leq M\text{-}1$. Elements of this matrix is determined according to linear constraints on the parameter vector $\boldsymbol{\theta}$. Columns of $\mathbf{W}$ are linearly independent of the M x 1 column vector $(1, \dots, 1)^T$.

Matricial expression of the log-linear model given in Eq.(3.2) is

$$\mathbf{\log p(\theta^*) = D\theta^*} \qquad (3.3)$$

where $\mathbf{D} = (\mathbf{1}_{Mx1}, \mathbf{W}_{MxM_0})$ is a M x $(M_0 + 1)$ "design matrix" with rank M_0 and $\boldsymbol{\theta}^* = (u, \theta_1, ..., \theta_{M_0})^T$ is a $(M_0 + 1)$ x 1 column vector with

$$u = -\log\left(\sum_{j=1}^{M} \exp(\mathbf{w}_j^T \boldsymbol{\theta})\right). \tag{3.4}$$

Eq.(3.4) is overall mean effect parameter calculated as the normalizing constant to insure $\sum_{j=1}^{M} p_j(\theta) = 1$.

The log-linear model (3.3) can be expressed by

$$\mathbf{logm(\theta^{**})} = \mathbf{D\theta^{**}}$$

where $\mathbf{m(\theta^{**})} = n\,\mathbf{p(\theta^{*})}$, $\boldsymbol{\theta}^{**} = (u^{*}, \theta_1, ..., \theta_{M_0})^T$ with $u^{*} = u + \log n$ and

$$\mathbf{logm(\theta^{**})} = (\log m_1(\theta^{**}), ..., \log m_M(\theta^{**}))^T$$

The set $\mathbf{C(D_A)} = \{\mathbf{logp(\theta)}: \mathbf{logp(\theta)} = \mathbf{D_A\theta};\ \boldsymbol{\theta} \in \Re^{M_0}\}$ represents the class of log-linear models associated to the matrix $\mathbf{D_A}$. $\mathbf{C(D_A)}$ denotes the column space of the matrix $\mathbf{D_A}$ and consists of all possible values for $\mathbf{logp(\theta)}$ that satisfy the log-linear model. It can be observed that $\mathbf{logp(\theta)} \in \mathbf{C(D_A)}$ is equivalent to $\mathbf{logm(\theta)} \in \mathbf{C(D_A)}$ with $\mathbf{m(\theta)} = n\,\mathbf{p(\theta)}$ because

$$\mathbf{\log m(\theta)} = \mathbf{\log p(\theta)} + \mathbf{\log}\ n\ \mathbf{J_M}$$

where $\mathbf{J_M} \equiv (1, ..., 1)^T$. An important assumption for the purpose of normalization is that M x 1 vector $\mathbf{J_M} \in \mathbf{C(D_A)}$.

3.3.2 Models for Two-Way and Three-Way Contingency Tables

3.3.2.1 Models for Two-Way Contingency Tables

Suppose there are two variables X and Y with I and J levels, respectively. That is, there is a two-way contingency table with M = IJ cells. The full (saturated) model which gives the perfect fit to observed data for two-way contingency tables is given in Eq.(3.5).

$$\log p_{ij}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{12(ij)};\ i = 1, \ldots, I,\ j = 1, \ldots, J \tag{3.5}$$

The constraints on parameters are

$$\sum_{i=1}^{I}\theta_{1(i)} = \sum_{j=1}^{J}\theta_{2(j)} = \sum_{i=1}^{I}\theta_{12(ij)} = \sum_{j=1}^{J}\theta_{12(ij)} = 0. \tag{3.6}$$

$\theta_{12(ij)}$ is interaction parameter between the i-th level of variable X and j-th level of variable Y. $\theta_{1(i)}$ and $\theta_{2(j)}$ represent the main effects of the i-th level of variable X and j-th level of variable Y, respectively. The model without interaction term is called as "independence model" which is given in Eq.(3.7).

$$\log p_{ij}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)};\ i = 1, \ldots, I,\ j = 1, \ldots, J \tag{3.7}$$

The other models for two-way contingency table apart from the saturated model and independence model are given below.

$$\log p_{ij}(\boldsymbol{\theta}) = u + \theta_{1(i)} \text{ or } \log p_{ij}(\boldsymbol{\theta}) = u + \theta_{2(j)};\ i = 1, \ldots, I,\ j = 1, \ldots, J \tag{3.8}$$

$$\log p_{ij}(\boldsymbol{\theta}) = u;\ i = 1, \ldots, I,\ j = 1, \ldots, J \tag{3.9}$$

Let Θ_1, Θ_2, Θ_3, and Θ_4 denote the parameter spaces for the models given with equations (3.5), (3.7), (3.8), and (3.9), respectively. One of the important features of the analysis of contingency tables based on a log-linear model is that models (3.5), (3.7), (3.8), and (3.9) are nested with $\Theta_4 \subset \Theta_3 \subset \Theta_2 \subset \Theta_1 \subseteq \Re^{IJ-1}$ and $d_4 < d_3 < d_2 < d_1 \leq IJ\text{-}1$ where d_L = dimension of the parameter space Θ_L; L = 1,…, 4

3.3.2.2 Models for Three-Way Contingency Tables

Before getting into the models for three dimensional tables, "partial dependence" and "marginal dependence" terms should be explained. If the distribution of two variables is displayed at different levels of third variable, these two-way contingency tables are called "Partial tables". Partial dependence means the association of two variables when the other one is controlled. The two-way tables obtained by combining the partial tables are called "Marginal tables". These tables ignore the third variable rather than controlling it. In other words, these two-way contingency tables are obtained by collapsing over third variable. Marginal dependence means dependence of two variables when other variable is totally ignored by collapsing. Partial tables can exhibit quite different associations than marginal tables. In fact, it can be misleading to analyze only the marginal tables of a multi-way table.

Suppose there are three variables X, Y, and Z with I, J, and K levels, respectively. That is, there is a three-way contingency table with M = IJK cells. The full (saturated) model for three-way contingency tables is given in Eq.(3.10).

$$\log p_{ijk}(\theta) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{12(ij)} + \theta_{13(ik)} + \theta_{23(jk)} + \theta_{123(ijk)} \quad (3.10)$$

$$i = 1,\ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K$$

Constraints on parameters given with Eq.(3.11) are

$$\sum_{i=1}^{I}\theta_{1(i)} = \sum_{j=1}^{J}\theta_{2(j)} = \sum_{k=1}^{K}\theta_{3(k)} = 0$$

$$\sum_{i=1}^{I}\theta_{12(ij)} = \sum_{j=1}^{J}\theta_{12(ij)} = \sum_{i=1}^{I}\theta_{13(ik)} = \sum_{k=1}^{K}\theta_{13(ik)} = \sum_{j=1}^{J}\theta_{23(jk)} = \sum_{k=1}^{K}\theta_{23(jk)} = 0 . \tag{3.11}$$

Interactions between variables of multi-way table can be the first order, the second order, etc. depending on the number of variables included in the table. For two-way contingency tables there exists the first order interaction, whereas for three-way contingency tables there exist the first and second order interactions. $\theta_{123(ijk)}$ is the second order interaction term. $\theta_{12(ij)}$, $\theta_{13(ik)}$, and $\theta_{23(jk)}$ are the first order interaction terms (partial dependence terms) between X and Y, X and Z, and Y and Z, respectively. $\theta_{1(i)}$, $\theta_{2(j)}$, and $\theta_{3(k)}$ represent the main effects of the i-th level of variable X, j-th level of variable Y, and k-th level of variable Z, respectively. The model without the second order interaction term is given below.

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{12(ij)} + \theta_{13(ik)} + \theta_{23(jk)}$$
$$i = 1,\ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K \tag{3.12}$$

If the last term, which is the partial dependence term between Y and Z, in Eq.(3.12) is discarded, the model of conditional independence of the variables Y and Z, given X is obtained.

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{12(ij)} + \theta_{13(ik)}$$
$$i = 1,\ldots,I,\ j = 1, \ldots, J,\ k = 1, \ldots, K \tag{3.13}$$

If Y and Z are independent in their corresponding partial table for the i-th level of X, then Y and Z are said to be conditionally independent at i-th level of X. Finally, Y and Z are conditionally independent, given X when they are conditionally independent at each level of X. The model (3.13) is one of the three models in which

only one pair of variables is conditionally independent. There are also the model of conditional independence of X and Y, given Z or the model of conditional independence of X and Z, given Y.

Among possible models for three-way contingency tables, there are three models in which only one pair of variables is conditionally dependent. These models are called as jointly independence models.

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{12(ij)} \qquad (3.14)$$

$$i = 1,\ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$$

In model (3.14), Z is jointly independent of X and Y which means Z and X are conditionally independent, given Y, and Z and Y are conditionally independent, given X. Similarly, Y could be jointly independent of X and Z, or X could be jointly independent of Y and Z.

In the model (3.12) all the partial dependence terms appear for each pair of variables, so no pair is conditionally independent. There is also a model in which no partial dependence terms appear. This is the model of mutual independence which means each pair of variables is conditionally independent and also marginally independent. This model is given in Eq.(3.15).

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} \qquad (3.15)$$

$$i = 1,\ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$$

The other models for three-way contingency table apart from the models given above are given with equations (3.16), (3.17) and (3.18).

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{2(j)}$$

or

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} + \theta_{3(k)} \tag{3.16}$$

or

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{2(j)} + \theta_{3(k)}$$

$$i = 1,\ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$$

$$\log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{1(i)} \text{ or } \log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{2(j)} \text{ or } \log p_{ijk}(\boldsymbol{\theta}) = u + \theta_{3(k)} \tag{3.17}$$

$$i = 1,\ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$$

$$\log p_{ijk}(\boldsymbol{\theta}) = u \tag{3.18}$$

$$i = 1,\ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$$

Let $\boldsymbol{\Theta_1}$, $\boldsymbol{\Theta_2}$, $\boldsymbol{\Theta_3}$, $\boldsymbol{\Theta_4}$, $\boldsymbol{\Theta_5}$, $\boldsymbol{\Theta_6}$, $\boldsymbol{\Theta_7}$, and $\boldsymbol{\Theta_8}$ denote the parameter spaces for the models given with equations (3.10), (3.12), (3.13),(3.14), (3.15), (3.16), (3.17) and (3.18), respectively. These models are nested with $\boldsymbol{\Theta_8} \subset \boldsymbol{\Theta_7} \subset \boldsymbol{\Theta_6} \subset \boldsymbol{\Theta_5} \subset \boldsymbol{\Theta_4} \subset \boldsymbol{\Theta_3} \subset \boldsymbol{\Theta_2} \subset \boldsymbol{\Theta_1} \subseteq \Re^{IJK-1}$ and $d_8 < d_7 < d_6 < d_5 < d_4 < d_3 < d_2 < d_1 \leq IJK\text{-}1$ where d_L = dimension of the parameter space $\boldsymbol{\Theta_L}$; L = 1, …, 8.

For the simulation studies presented in Chapter 4, the models given with equations (3.5) and (3.7) will be the focus for two-way contingency tables with the case of 2 x 2 tables while the models given with equations (3.14) and (3.15) will be the focus for three-way contingency tables with the case of 2 x 2 x 2 ones.

CHAPTER FOUR
ORDINARY AND PENALIZED POWER-DIVERGENCE MEASURES

The method is needed that aids in the selection of the interaction term / terms to be included in the fitted model. Fienberg (1970), Goodman (1970), Ku et al. (1971), Ku & Kullback (1974) are among others who suggest different approaches to the model selection problem. Later, Cressie & Read (1984) developed unified approach to model testing by the family of power-divergence test statistics. These statistics are based on the power-divergence measures which are included in the family of disparity measures defined by Lindsay (1994). Before describing the family of power-divergence test statistics, description of the families of ordinary and penalized power-divergence measures and estimators minimizing them will be given in Section 4.1. This section will also include a subsection for the simulation study performed to compare the small and moderate sample performances of ordinary and penalized minimum power-divergence estimators in terms of efficiency and robustness. After describing the families of ordinary and penalized test statistics in Section 4.2, simulation studies will be presented with separate subsection. These simulation studies have been performed to compare the size and power properties of both of these families of test statistics for testing nested log-linear models in two-way and three-way contingency tables. Finally, these test statistics will be illustrated with two medical examples at the end of Section 4.2.

4.1 Ordinary and Penalized Minimum Power-Divergence Estimators

4.1.1. The Family of Ordinary Minimum Power-Divergence Estimators

Since the parameter values are unknown in the class of log-linear models, they need to be estimated. “There are two fundamental-but potentially competing ideals in parametric estimation: efficiency when the model has been appropriately chosen and robustness when it has not” (Lindsay, 1994). When it comes to estimation, maximum likelihood estimator (MLE) is the mostly used estimator. “MLE is known to be efficient in regular models but it is also known to be non-robust” (Menéndez,

Morales, Pardo, & Vajda, 2001). "The efficiency as well as the non-robustness are resulting from specific properties of the logarithmic function used in the definition of the function" (Pardo, Pardo, & Zografos, 2001). By replacing the logarithmic function by other functions with appropriate properties, a new class of estimators such as minimum ϕ-divergence estimators can be obtained.

ϕ-divergence measure is a density based divergence and has been defined by Csiszár (1967) as follows;

$$D_\phi(\mathbf{p},\mathbf{q}) = \sum_{j=1}^{m} q_j \phi\left(\frac{p_j}{q_j}\right) ;\phi \in \mathbf{\Phi}^* \tag{4.1}$$

where $\mathbf{p} = (p_1, ..., p_m)^T$ and $\mathbf{q} = (q_1, ..., q_m)^T$ are discrete probability distributions. $\mathbf{\Phi}^*$ is the class of all convex functions $\phi : [0, \infty) \to \Re \cup \{\infty\}$, such that at x = 1, $\phi(1) = 0$ and $\phi''(1) > 0$, and at x = 0, $0\phi(0/0) = 0$ and $0\phi(p/0) = \lim_{u\to\infty} \phi(u)/u$. For every $\phi \in \mathbf{\Phi}^*$ that is differentiable at x = 1, the function $\psi(x) \equiv \phi(x) - \phi'(1)(x-1)$ also belongs to $\mathbf{\Phi}^*$. Then $\psi(x)$ and $\phi(x)$ are equivalent functions with ψ having additional property of $\psi'(1) = 0$.

$$\hat{\theta}_\phi \equiv \arg\min_{\theta\in\Theta} D_\phi(p,q) . \tag{4.2}$$

Asymptotic properties of minimum ϕ-divergence estimators given with Eq.(4.2) have been established by Lindsay (1994) and Morales, Pardo, & Vajda (1995). Based on the results established by them, asymptotic results for minimum ϕ-divergence estimators under log-linear model have been established by Cressie & Pardo (2000). According to their results,

$$n^{1/2}(\hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta_0}) \xrightarrow[n\to\infty]{L} N(0, (\mathbf{W}^T \mathbf{\Sigma}_{\mathbf{p}(\boldsymbol{\theta_0})} \mathbf{W})^{-1}) \tag{4.3}$$

where $\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)} = \mathbf{diag}(\mathbf{p}(\boldsymbol{\theta}_0)) - \mathbf{p}(\boldsymbol{\theta}_0)\mathbf{p}(\boldsymbol{\theta}_0)^{\mathbf{T}}$. "$\xrightarrow[n\to\infty]{L}$" denotes convergence in distribution. $\boldsymbol{\theta}_0$ is the true parameter vector and $\mathbf{W}^{\mathbf{T}}\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0)}\mathbf{W}$ is the Fisher information matrix. Eq.(4.3) means that every minimum ϕ-divergence estimator is asymptotically first-order efficient since its asymptotic variance reaches the asymptotically efficient bound, the inverse of the Fisher information matrix.

Through appropriate selection of the ϕ function, a large family of important divergences can be developed including the power-divergence family. Cressie & Read (1984) developed the power-divergence family with following ϕ function.

$$\phi_{(\lambda)}(x) = \begin{cases} \dfrac{1}{\lambda(\lambda+1)}(x^{\lambda+1} - x) & \text{for } \lambda \neq 0, \lambda \neq -1 \\ x \log x\text{-}(x\text{-}1) & \text{for } \lambda \to 0 \\ -\log x + (x\text{-}1) & \text{for } \lambda \to \text{-}1 \end{cases} \tag{4.4}$$

Since the function is not defined at λ = 0 and λ = -1, $\lim_{\lambda\to 0} \dfrac{1}{\lambda(\lambda+1)}(x^{\lambda+1} - x)$ and $\lim_{\lambda\to -1} \dfrac{1}{\lambda(\lambda+1)}(x^{\lambda+1} - x)$ are used. The function is subscripted with λ as it is indexed by λ.

The power-divergence measure between $\hat{\mathbf{p}}$ and $\mathbf{p}(\boldsymbol{\theta})$ can be defined by Eq.(4.5).

$$I_{\lambda}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \frac{1}{\lambda(\lambda+1)} \sum_{j=1}^{M} \hat{p}_j \left[\left(\frac{\hat{p}_j}{p_j(\theta)} \right)^{\lambda} - 1 \right] \tag{4.5}$$

As mentioned previously, the family of power-divergence measures is included in the family of disparities defined by Lindsay (1994). A "disparity" is a measure of discrepancy between a non-parametric density obtained from data and the model density. This measure assumes its minimum value zero when these two densities are

identical. The empirical density function can be used as a non-parametric density function. The observed relative frequency is defined to be the empirical density function. So, in the Eq.(4.5), $\hat{\mathbf{p}} = (\hat{p}_1,...,\hat{p}_M)^T$, with $\hat{p}_j = \frac{N_j}{n}$, represents the vector of observed relative frequencies while $\mathbf{p(\theta)} = (p_1(\theta), \ldots, p_M(\theta))^T$ represents the vector of probabilities obtained from the model.

As easily can be seen from Eq.(4.5), $I_0(\hat{p}, p(\theta)) = I_{-1}(p(\theta), \hat{p})$. In general, the estimates minimizing $D_\phi(\hat{\mathbf{p}}, \mathbf{p(\theta)})$ and $D_\phi(\mathbf{p(\theta)}, \hat{\mathbf{p}})$ coincide only if divergence measure is symmetric. Power-divergence family is symmetric about λ = -1/2, i.e. $I_{-1/2}(\hat{p}, p(\theta)) = I_{-1/2}(p(\theta), \hat{p})$. "This follows by observing that if λ and γ are such that $I_\lambda(\hat{p}, p(\theta)) = I\gamma(p(\theta), \hat{p})$, then $\gamma = -\lambda-1$, hence the only value for which $\gamma = \lambda$ is $\lambda = -1/2$" (Read & Cressie, 1998, p.112).

Harris & Basu (1997) have considered the family of power-divergence measures in the following form:

$$I_\lambda(\hat{\mathbf{p}}, \mathbf{p(\theta)}) = \sum_{j=1}^{M} \left[\frac{1}{\lambda(\lambda+1)} \hat{p}_j \left[\left(\frac{\hat{p}_j}{p_j(\theta)} \right)^\lambda - 1 \right] + \frac{p_j(\theta) - \hat{p}_j}{\lambda+1} \right]. \qquad (4.6)$$

The second component in Eq.(4.6) does not contribute anything to the measure and comes from using $\psi(x) \equiv \phi(x) - \phi'(1)(x-1)$ instead of $\phi(x)$. Since $\phi(x)$ and $\psi(x)$ are equivalent, Eq.(4.5) and Eq.(4.6) define the same divergence. From now on, the function $\psi(x)$ will be used instead of $\phi(x)$. That is, Eq.(4.6) will be used instead of Eq.(4.5). However, the function will still be defined as $\phi(x)$. The reason to prefer the symbol $\phi(x)$ over $\psi(x)$ is to make it remember that the function is included in ϕ-divergence measures.

The minimum power-divergence estimator of $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \Re^{M_0}$ as the value minimizing $I_\lambda(\hat{\mathbf{p}}, \mathbf{p(\theta)})$ for $\boldsymbol{\theta}$ is defined by

$$\hat{\boldsymbol{\theta}}_{\lambda}^{I} \equiv \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} I_{\lambda}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})). \tag{4.7}$$

The family of power-divergence measures given with Eq.(4.6) consists of well known statistics and estimators. For example:

♦ For $\lambda = 1$, $I_{\lambda}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$ is the Pearson's chi-square divergence.

$$I_{1}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \frac{1}{2}\sum_{j=1}^{M}\frac{(\hat{p}_j - p_j(\theta))^2}{p_j(\theta)} \tag{4.8}$$

♦ For $\lambda = -2$, $I_{\lambda}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$ is the Neyman's modified chi-square divergence.

$$I_{-2}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \frac{1}{2}\sum_{j=1}^{M}\frac{(\hat{p}_j - p_j(\theta))^2}{\hat{p}_j} \tag{4.9}$$

♦ For $\lambda = -1/2$, $I_{\lambda}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$ is equal to squared Hellinger distance. This measure is also called as Freeman-Tukey divergence by Read & Cressie (1988).

$$I_{-1/2}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = 2\sum_{j=1}^{M}\left[\sqrt{\hat{p}_j} - \sqrt{p_j(\theta)}\right]^2 \tag{4.10}$$

♦ For $\lambda \to 0$, $I_{\lambda}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$ is equal to Kullback-Leibler divergence

$$I_{0}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \sum_{j=1}^{M}\left[\hat{p}\log\frac{\hat{p}}{p(\theta)} + (p(\theta) - \hat{p})\right] \tag{4.11}$$

♦ For $\lambda \to -1$, $I_{\lambda}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$ is equal to discrimination information measure.

$$I_{-1}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \sum_{j=1}^{M}\left[p(\theta)\log\frac{p(\theta)}{\hat{p}} + (\hat{p} - p(\theta))\right] \tag{4.12}$$

Hence, the estimators, $\hat{\boldsymbol{\theta}}_{\lambda}^{I}$, minimizing equations (4.8), (4.9), (4.10), (4.11) and (4.12) are equal to minimum Pearson's chi-square estimator, minimum Neyman's modified chi-square estimator, minimum Hellinger distance estimator, minimum Kullback-Leibler divergence estimator which is also equivalent to MLE, and minimum discrimination information estimator, respectively. Cressie & Read (1984) defined a new estimator called minimum Cressie-Read divergence estimator for $\lambda = 2/3$.

Lindsay (1994) has developed "residual adjustment function (RAF)" as an alternative to "influence curve analysis" which is used to get robustness information. Lindsay (1994) states that "the influence curve is a very misleading measure of robustness in the case of minimum Hellinger distance estimator". The influence curve approach is also criticized by authors such as Beran (1977), Tamura & Boos (1986) and Simpson (1987) as being limited method in measuring robustness. Basu & Basu (1998) emphasizes that theoretical properties of the estimators are controlled by the form of RAF since the estimating equations are otherwise equivalent. According to Lindsay (1994), RAF can be obtained from the estimating functions for $\boldsymbol{\theta}$. So, for the family of ϕ-divergence measures, RAF can be obtained from the following estimating functions:

$$-\frac{\partial D_{\phi}(x)}{\partial \theta_i} = \sum_{j=1}^{M} A(x_j)\nabla_{\mathbf{p_j(\theta)}} \; ; \; i = 1, \ldots, M_0 \qquad (4.13)$$

with $\nabla \mathbf{p_j(\theta)} = \mathbf{p'_j(\theta)} = (\frac{\partial}{\partial \theta_1} p_j(\theta), \ldots, \frac{\partial}{\partial \theta_{M_0}} p_j(\theta))$ which is gradient of $p_j(\theta)$ for $\boldsymbol{\theta}$, and $A(x_j) = \phi'(x_j)x_j - \phi(x_j)$. According to Lindsay (1994), A(x) should be an increasing twice differentiable function on $[0,\infty)$ with $A(1) = 0$ and $A'(1) = 1$ to be called as RAF. Since $A'(x) = \phi''(x)x$, second order differentiability of ϕ, in addition to its strict convexity, implies that A(x) is a strictly increasing function of x on $[0,\infty)$. Following Lindsay (1994), A(x) can be redefined as $A(x) := A(x) - A(1)$ so that $A(1) = 0$ without changing the solutions to the estimating functions in Eq.(4.13). Finally,

since $A'(1) = \phi''(1) > 0$, A(x) can be rescaled so that it satisfies $A'(1) = 0$. Thus, it has been ended up with A(x) satisfying RAF assumptions. It should be noted that redefining A(x) as $A(x) \equiv A(x) - A(1)$ is equivalent to redefine $\phi(x)$ as $\psi(x) \equiv \phi(x) - \phi'(1)(x-1)$. For the family of power-divergence measures given with Eq.(4.6), RAF is equal to

$$A(x) = \frac{x^{\lambda+1}}{(\lambda+1)} - \frac{1}{\lambda+1}. \tag{4.14}$$

According to Lindsay (1994), the second derivative of the RAF at $x = 1, A''(1)$, is a measure of the trade off between robustness and efficiency for the family of ϕ-divergence measures, and large negative values of $A''(1)$ correspond to stronger robustness properties. Hence, for the family of power-divergence measures, minimum power-divergence estimators with large negative values of λ are more robust than other estimators, since $A''(1) = \lambda$. So, they perform better than MLE in terms of robustness. However, as it will be seen in simulation study results in Section 4.1.3, when it comes to small samples, MLE can perform better in terms of efficiency than many of the more robust estimators. According to Basu & Basu (1998), part of it can be for the manner in which disparities treat empty cells, especially for the values of λ close to -1. To illustrate the empty cell problem in the family of power-divergence measures, Basu & Basu (1998) have separated Eq.(4.7) as in Eq.(4.15).

$$\begin{aligned} I_\lambda(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = & \sum_{\substack{j=1 \\ \hat{p}_j \neq 0}}^{M} \left[\frac{1}{\lambda(\lambda+1)} \hat{p}_j \left[\left(\frac{\hat{p}_j}{p_j(\theta)} \right)^{\lambda} - 1 \right] + \frac{p_j(\theta) - \hat{p}_j}{\lambda+1} \right] \\ & + \frac{1}{\lambda+1} \sum_{\substack{j=1 \\ \hat{p}_j = 0}}^{M} p_j(\theta) \end{aligned} \tag{4.15}$$

Eq.(4.15) is comprised of two parts; the part for the non-empty cells and the part for the empty cells. The second component can become very large for the values of λ

close to -1. Basu & Basu (1998) proposed an empty cell penalty for the family of power-divergence measures to be able to cope with this problem.

4.1.2. The Family of Minimum Penalized Power-Divergence Estimators

By applying penalty on empty cells in Eq. (4.15), the family of penalized power-divergence measures can be obtained as given by Eq.(4.16).

$$P_\lambda^w(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = \sum_{\substack{j=1 \\ \hat{p}_j \neq 0}}^{M} \left[\frac{1}{\lambda(\lambda+1)} \hat{p}_j \left[\left(\frac{\hat{p}_j}{p_j(\theta)} \right)^\lambda - 1 \right] + \frac{p_j(\theta) - \hat{p}_j}{\lambda+1} \right] + w \sum_{\substack{j=1 \\ \hat{p}_j = 0}}^{M} p_j(\theta) \tag{4.16}$$

where w is the penalty put on empty cells. By increasing or decreasing this penalty, power-divergence measure can be made more or less sensitive to empty cells. As it can be seen, $I_0(\hat{p}, p(\theta)) = P_0^1(\hat{p}, p(\theta))$ and $I_1(\hat{p}, p(\theta)) = P_1^{0.5}(\hat{p}, p(\theta))$, i.e; when w = 1, the penalized Kullback-Leibler divergence puts the same weight on empty cells as ordinary Kullback-Leibler divergence does where as when w = 0.5, penalized Pearson's chi-square divergence puts the same weight as ordinary Pearson's chi-square divergence does.

The penalized minimum power-divergence estimator of $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \Re^{M_0}$ is defined as,

$$\hat{\boldsymbol{\theta}}_\lambda^{P^w} \equiv \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} P_\lambda^w(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) \tag{4.17}$$

For λ = -1, since A(0) = - ∞, $I_\lambda(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$ can not be defined even if there is only one empty cell (i.e. $\hat{p}_j = 0$). This can easily be seen from the following expression.

$$A(x_j) = \frac{(x_j)^{\lambda+1}}{(\lambda+1)} - \frac{1}{(\lambda+1)} = \frac{\left[\frac{\hat{p}_j}{p_j(\theta)}\right]^{\lambda+1}}{(\lambda+1)} - \frac{1}{(\lambda+1)}$$

This is true for all values of $\lambda \leq -1$. Whenever $\lambda < -1$, $\left[\frac{\hat{p}_j}{p_j(\theta)}\right]^{\lambda+1}$ will be equal to $\left[\frac{p_j(\theta)}{\hat{p}_j}\right]^{-(\lambda+1)}$. Hence, $A(0) = \infty$ for $\lambda < -1$. However, $P_w^{\lambda}(\hat{p}, p(\theta))$ can be defined for all values of λ irrespective of empty cells since empty cells component does not depend on λ.

Reweighting empty cells does not change the asymptotic properties of corresponding estimator. Hence, any ordinary estimator $\hat{\theta}_{(\lambda)}^{I(L)}$ and any penalized estimator $\hat{\theta}_{(\lambda)}^{P^{W(L)}}$ where L is the model under consideration are first- order efficient and asymptotically equivalent in the sense that they are asymptotically consistent and normal with the same minimum variance. Even though equivalence of all minimum power-divergence estimators, these asymptotic results give no information about the rate of convergence. To compare the first-order efficient estimators Rao (1961, 1962) introduces the concept of second-order efficiency to measure the amount of information up to terms of order $1/n^2$. Rao (1961, 1962) have proposed second-order efficiency measure following the Fisher's ideas about information measure. "Fisher considered information measure as a more intrinsic property of an estimator than its bias and variance, and suggested the choice of an estimator which preserves the maximum information" (Berkson, 1980). Rao (1961, 1962) have showed that in the case of multinomial distribution the MLE contains the maximum information in large samples. Efron (1975) and Ghosh & Subramanyam (1974) have emphasized the importance of second-order efficiency in statistical inference. Apart from the authors defending Rao's second-order efficiency approach to compare first-order efficient estimators, there are authors such as Berkson (1980), Read & Cressie (1988) who

disagree with Rao's criterion. As opposed to Rao (1961, 1962), Berkson (1980) have defended to compare estimators by using minimization of mean square error (MSE) criterion and claimed that minimum Pearson's chi-square estimator is better than MLE. More information and interesting discussions about Berkson ideas can be found in Berkson (1980)'s paper.

Read & Cressie (1988, p.157) have performed a study to compare first-order efficient estimators and they have demonstrated that the optimal choice of λ for power-divergence family is dependent on the unknown parameter θ, i.e., of the model considered. Basu & Basu (1998) have considered the model with a single parameter. They considered probabilities generated from Poisson and Geometric distributions, respectively. They have performed a comparison study of ordinary and penalized minimum power-divergence estimators for the values of λ (-1, 1] for n = 20 by Berkson (1980)'s MSE criterion. Based on the exact calculations, they have established that under moderate contaminations (ε = %10) the performance of MLE is worse than the estimates corresponding the large negative values of λ and in this case mean square errors for the penalized estimators are competitive or better than ordinary estimators. Later, Pardo & Pardo (2003) have tried to determine a member of power-divergence family that is more efficient than the others to estimate the unknown parameters for the sequence of nested models for three-way contingency tables. By using Berkson (1980)'s MSE criterion in the simulation study, they have showed that minimum Pearson's chi-square estimator is simultaneously the most robust and efficient estimator among members of the family of the ordinary minimum power-divergence estimators. In this thesis, the small and moderate sample efficiency and robustness properties of the ordinary and penalized minimum power-divergence estimators are compared by simulation study. The log-linear independence model in two-way contingency tables is the model under consideration. Following subsection includes the simulation study.

4.1.3. Simulation Study

In this simulation study, 2 x 2 contingency tables are considered. The aims are to study the efficiency and robustness properties of the ordinary minimum power-divergence estimators, and to compare the ordinary and penalized minimum power-divergence estimators for the penalty values of w = 1 and w = 0.5 under the log-linear independence model given with Eq.(3.7). The values of λ considered are among (-1, 1], Values arbitrarily chosen to imply fixed main effects are

$$\exp(\theta_{1(1)}) = \exp(\theta_{2(1)}) = \frac{1}{2} \tag{4.18}$$

with the constrains given with Eq.(3.6).

The matrix form of the model given with Eq.(3.7) is given with Eq.(4.19).

$$\mathbf{\log p(\theta^*) = D\theta^* \Rightarrow \log p(\theta^*) = [1, W]_{4x3}[\theta^*]_{3x1}} \tag{4.19}$$

$$\begin{bmatrix} \log p_1(\theta) \\ \log p_2(\theta) \\ \log p_3(\theta) \\ \log p_4(\theta) \end{bmatrix} = \begin{bmatrix} \log p_{11}(\theta) \\ \log p_{12}(\theta) \\ \log p_{21}(\theta) \\ \log p_{22}(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} u \\ \theta_{1(1)} \\ \theta_{2(1)} \end{bmatrix}$$

So, the log-linear model matrix (constraint matrix)

$$\mathbf{W} = \begin{bmatrix} \mathbf{w_1} & \mathbf{w_2} & \mathbf{w_3} & \mathbf{w_4} \end{bmatrix}^T = \begin{bmatrix} \mathbf{w_1^T} \\ \mathbf{w_2^T} \\ \mathbf{w_3^T} \\ \mathbf{w_4^T} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{bmatrix} \tag{4.20}$$

implies that

$$\exp(\theta_{1(2)}) = \exp(-\theta_{1(1)}) = 2$$

and (4.21)

$$\exp(\theta_{2(2)}) = \exp(-\theta_{2(1)}) = 2.$$

As given in Eq.(3.4), u is calculated with $u = -\log(\sum_{j=1}^{M} \exp(\mathbf{w}_j^T \boldsymbol{\theta}))$ to insure that $\sum_{j=1}^{M} p_j(\theta) = 1$. So for the 2 x 2 tables case, u is calculated as given with Eq.(4.22)

$$u = -\log(\sum_{j=1}^{4} \exp(\mathbf{w}_j^T \boldsymbol{\theta})) \tag{4.22}$$

By using the equations (4.18), (4.19), (4.21) and (4.22), model probability vector $\mathbf{p(\theta)}$ is obtained as in Eq.(4.23).

$$\mathbf{p(\theta)} = \begin{bmatrix} 0.04 & 0.16 & 0.16 & 0.64 \end{bmatrix}^T \tag{4.23}$$

To compare estimators, Berkson (1980)'s MSE criteria will be used. MSE values for parameters are calculated as follows:

$$MSE(\theta_{1(1)}) = \sum_{s=1}^{S} (\hat{\theta}^i_{\lambda[1(1)]} - \theta_{1(1)})^2 p_s(\theta) \tag{4.24}$$

$$MSE(\theta_{2(1)}) = \sum_{s=1}^{S} (\hat{\theta}^i_{\lambda[2(1)]} - \theta_{2(1)})^2 p_s(\theta) \tag{4.25}$$

for i= I(ordinary), P^w(Penalized). S is the total number of samples simulated, and $p_s(\theta)$ is the probability of the corresponding sample calculated as given in Eq.(3.1) with the cell probabilities given in Eq.(3.2).

One way to study the robustness of an estimator in multinomial models is to study the behavior of the mean square error for the parameters when the true distribution is contaminated (Pardo & Pardo, 2003). Hence, contamination is introduced for studying the robustness of the ordinary and penalized minimum power-divergence estimators. Contaminated probability vector is defined as

$$\mathbf{p(\theta)}^{*} = (p_{11}(\theta)^{*}, p_{12}(\theta)^{*}, p_{21}(\theta)^{*}, p_{22}(\theta)^{*})$$

with

$$p_{ij}(\theta)^{*} = (1-\varepsilon)p_{ij}(\theta)\,;\ \text{for all i, j}$$

except from

$$p_{ij}(\theta)^{*} = (1-\varepsilon)p_{11}(\theta) + \varepsilon\,;\ i = j = 1,$$

where ε is contamination value. Following Basu & Basu (1998), and Pardo & Pardo (2003), moderate contamination value of ε = %10 have been used. The contaminated probability vector is obtained as in Eq.(4.26).

$$\mathbf{p(\theta)}^{*} = \begin{bmatrix} 0.14 & 0.14 & 0.14 & 0.58 \end{bmatrix}^{T} \tag{4.26}$$

MSE values for parameters given with equations (4.24) and (4.25) are obtained using 500 contingency tables simulated from multinomial distributions with (n, **p(θ)**) and (n, $\mathbf{p(\theta)}^{*}$). **p(θ)** and $\mathbf{p(\theta)}^{*}$ are probability vectors given with equations (4.23) and (4.26), respectively. Since all of the minimum power-divergence estimators have the same asymptotic properties, the values for sample size "n" have been chosen as 25, 30, 35, 45, 55 to see their performances with small and moderate samples. n = 75 has also been included to verify that all estimators are asymptotically equivalent. All calculations including simulations have been performed by using Mathematica 5.0.

The results for ordinary and penalized minimum power-divergence estimators when the model holds (ε = 0) are given with Tables 4.1 – 4.3. The smallest MSE values are written bold.

Table 4.1 Mean square errors for estimates of the parameters of log-linear independence model for 2x2 contingency table by ordinary minimum power-divergence estimators without contamination (ε = 0.0, i.e., when the model holds)

		λ							
	n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
MSE for $\hat{\theta}^{I}_{(\lambda)1(1)}$	25	0.17740	**0.15016**	0.15550	0.23701	0.26943	0.32748	0.46333	1.22688
	35	0.07740	**0.06786**	0.06985	0.09619	0.10609	0.12349	0.16367	0.38474
	45	0.03327	**0.03031**	0.03090	0.03936	0.04258	0.04825	0.06095	0.13095
	55	0.01933	**0.01821**	0.01846	0.02126	0.02221	0.02379	0.02715	0.04386
	75	0.00773	**0.00741**	0.00749	0.00808	0.00820	0.00839	0.00873	0.01021
MSE for $\hat{\theta}^{I}_{(\lambda)2(1)}$	25	0.19187	**0.16144**	0.16732	0.25902	0.29531	0.36009	0.51165	1.39352
	35	0.07093	**0.06331**	0.06487	0.08694	0.09557	0.11091	0.14623	0.34299
	45	0.03467	**0.03154**	0.03217	0.04127	0.04480	0.05099	0.06506	0.14492
	55	0.01940	**0.01823**	0.01849	0.02148	0.02252	0.02430	0.02812	0.04644
	75	0.00802	**0.00759**	0.00770	0.00847	0.00863	0.00887	0.00934	0.01188

Table 4.2 Mean square errors for estimates of the parameters of log-linear independence model for 2x2 contingency table by penalized (w = 1) minimum power -divergence estimators without contamination (ε = 0.0, i.e., when the model holds)

		λ							
	n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
MSE for $\hat{\theta}^{P^1}_{(\lambda)1(1)}$	25	0.17740	**0.16703**	0.16992	0.18512	0.18693	0.18878	0.19073	0.19277
	35	0.07740	**0.07361**	0.07469	0.07997	0.08055	0.08116	0.08179	0.08241
	45	0.03327	**0.03196**	0.03231	0.03424	0.03446	0.03470	0.03494	0.03518
	55	0.01933	**0.01868**	0.01886	0.01983	0.01994	0.02005	0.02017	0.02030
	75	0.00773	**0.00745**	0.00752	0.00796	0.00801	0.00807	0.00813	0.00820
MSE for $\hat{\theta}^{P^1}_{(\lambda)2(1)}$	25	0.19187	**0.18125**	0.18414	0.20032	0.20235	0.20445	0.20661	0.20887
	35	0.07093	**0.06809**	0.06891	0.07292	0.07338	0.07387	0.07440	0.07496
	45	0.03467	**0.03345**	0.03379	0.03553	0.03573	0.03595	0.03617	0.03639
	55	0.01940	**0.01879**	0.01896	0.01984	0.01994	0.02005	0.02016	0.02027
	75	0.00802	**0.00765**	0.00775	0.00831	0.00838	0.00846	0.00853	0.00861

Table 4.3 Mean square errors for estimates of the parameters of log-linear independence model for 2x2 contingency table by penalized (w=0.5) minimum power-divergence estimators without contamination (ε = 0.0, i.e., when the model holds)

		λ							
	n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
MSE for $\hat{\theta}^{P^{0.5}}_{(\lambda)1(1)}$	25	0.15774	**0.15016**	0.15223	0.16353	0.16487	0.16622	0.16764	0.16906
	35	0.07076	**0.06785**	0.06868	0.07276	0.07320	0.07368	0.07416	0.07462
	45	0.03137	**0.03031**	0.03059	0.03217	0.03236	0.03255	0.03275	0.03295
	55	0.01879	**0.01821**	0.01837	0.01924	0.01935	0.01945	0.01956	0.01968
	75	0.00768	**0.00741**	0.00748	0.00791	0.00796	0.00802	0.00808	0.00815
MSE for $\hat{\theta}^{P^{0.5}}_{(\lambda)2(1)}$	25	0.16887	**0.16144**	0.16341	0.17517	0.17668	0.17822	0.17981	0.18142
	35	0.06540	**0.06331**	0.06392	0.06686	0.06718	0.06754	0.06791	0.06826
	45	0.03247	**0.03154**	0.03180	0.03314	0.03330	0.03347	0.03364	0.03382
	55	0.01876	**0.01823**	0.01838	0.01915	0.01924	0.01933	0.01943	0.01954
	75	0.00796	**0.00759**	0.00769	0.00825	0.00831	0.00839	0.00846	0.00854

Among ordinary minimum power-divergence estimators, minimum Pearson's chi-square estimator is the one with smallest MSE value for both of the parameter estimates for all sample sizes. This situation is same for the penalized minimum power-divergence estimators. Since the weight of empty cells increases with penalization value of w = 1 for minimum Pearson's chi-square estimator, MSE values increase. However, penalized Pearson's chi-square estimator, $\hat{\theta}^{P^1}_{(1)i(1)}$, is still the one with smallest MSE value among the all penalized estimators with w = 1. Since $I_1(\hat{p}, p(\theta)) = P_1^{0.5}(\hat{p}, p(\theta))$ and $I_0(\hat{p}, p(\theta)) = P_0^1(\hat{p}, p(\theta))$, $MSE(\hat{\theta}^{I}_{(1)i(1)}) = MSE(\hat{\theta}^{P^{0.5}}_{(1)i(1)})$ and $MSE(\hat{\theta}^{I}_{(0)i(1)}) = MSE(\hat{\theta}^{P^1}_{(0)i(1)})$ for i = 1, 2. Minimum Cressie - Read estimator ($\hat{\theta}^{I}_{(2/3)}$) can be a good opponent to minimum Pearson's chi-square estimator since it is always placed second while MLE ($\hat{\theta}^{I}_{(0)}$) is always placed third for both parameters. This order does not change for the penalized minimum power-divergence estimators for both penalization values. It is mentioned that the minimum power-divergence estimators with large negative values of λ are considered as robust estimators according to Lindsay (1994). However when the model holds ($\varepsilon = 0.0$) and as λ getting closer to -1, MSE values for corresponding estimators are getting very high compared to the minimum Pearson's chi-square estimator and MLE for small samples, especially for the sample size of 25. As mentioned previously, this is the trade of between small sample efficiency and robustness.

When penalized minimum power-divergence estimators are considered, the obtained MSE values for the robust minimum power-divergence estimators improve a lot. The reason to include sample size of 75 is to show that all estimators get asymptotically equivalent as n increases. Simulation results verify this. As n increases, MSE values are getting smaller and closer to each other.

From Table 4.4 – 4.6, it is possible to compare performances of ordinary and penalized minimum power-divergence estimators for two penalization values under moderate contamination ($\varepsilon = 10\%$).

Table 4.4 Mean square errors for estimates of the parameters of log-linear independence model for 2x2 contingency table by ordinary minimum power-divergence estimators with moderate contamination (ε = 10%)

		λ							
	n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
MSE for $\hat{\theta}^{I}_{(\lambda)1(1)}$	25	0.14908	0.17933	0.17075	0.13014	0.12672	0.12374	0.12143	**0.12073**
	35	0.07520	0.09329	0.08803	0.06446	0.06252	0.06080	0.05942	**0.05905**
	45	0.05471	0.06763	0.06396	0.04636	0.04469	0.04307	0.04153	**0.04009**
	55	0.03682	0.04639	0.04360	0.03087	0.02967	0.02850	0.02738	**0.02631**
	75	0.02259	0.02837	0.02669	0.01893	0.01817	0.01742	0.01668	**0.01597**
MSE for $\hat{\theta}^{I}_{(\lambda)2(1)}$	25	0.15207	0.17932	0.17169	0.13463	0.13147	0.12873	0.12663	**0.12652**
	35	0.08082	0.09873	0.09365	0.06956	0.06758	0.06613	0.06619	**0.07602**
	45	0.05091	0.06389	0.06015	0.04266	0.04100	0.03940	0.03788	**0.03646**
	55	0.04054	0.05011	0.04737	0.03439	0.03314	0.03191	0.03072	**0.02958**
	75	0.02258	0.02842	0.02672	0.01890	0.01815	0.01740	0.01667	**0.01596**

Table 4.5 Mean square errors for estimates of the parameters of log-linear independence model for 2x2 contingency table by penalized (w = 1) minimum power-divergence estimators with moderate contamination (ε = 10%)

		λ							
	n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
MSE for $\hat{\theta}^{P^1}_{(\lambda)1(1)}$	25	0.14908	0.17938	0.17079	0.12999	0.12649	0.12336	0.12072	**0.11862**
	35	0.07520	0.09331	0.08805	0.06445	0.06250	0.06074	0.05919	**0.05789**
	45	0.05471	0.06763	0.06396	0.04636	0.04469	0.04307	0.04152	**0.04008**
	55	0.03682	0.04639	0.04360	0.03087	0.02967	0.02850	0.02738	**0.02631**
	75	0.02259	0.02837	0.02669	0.01893	0.01817	0.01742	0.01668	**0.01597**
MSE for $\hat{\theta}^{P^1}_{(\lambda)2(1)}$	25	0.15207	0.17936	0.17172	0.13454	0.13132	0.12847	0.12608	**0.12419**
	35	0.08082	0.09873	0.09366	0.06941	0.06722	0.06518	0.06332	**0.06172**
	45	0.05091	0.06389	0.06015	0.04266	0.04100	0.03940	0.03788	**0.03645**
	55	0.04054	0.05011	0.04737	0.03439	0.03314	0.03191	0.03072	**0.02958**
	75	0.02258	0.02842	0.02672	0.01890	0.01815	0.01740	0.01667	**0.01596**

Table 4.6 Mean square errors for estimates of the parameters of log-linear independence model for 2x2 contingency table by penalized (w = 0.5) minimum power-divergence estimators with moderate contamination (ε = 10%)

		λ							
	n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
MSE for $\hat{\theta}^{P^{0.5}}_{(\lambda)1(1)}$	25	0.14901	0.17933	0.17073	0.12992	0.12642	0.12329	0.12065	**0.11855**
	35	0.07519	0.09329	0.08803	0.06445	0.06250	0.06074	0.05919	**0.05789**
	45	0.05471	0.06763	0.06396	0.04636	0.04469	0.04307	0.04152	**0.04008**
	55	0.03682	0.04639	0.04360	0.03087	0.02967	0.02850	0.02738	**0.02631**
	75	0.02259	0.02837	0.02669	0.01893	0.01817	0.01742	0.01668	**0.01597**
MSE for $\hat{\theta}^{P^{0.5}}_{(\lambda)2(1)}$	25	0.15203	0.17932	0.17168	0.13449	0.13128	0.12843	0.12603	**0.12414**
	35	0.08081	0.09873	0.09365	0.06938	0.06718	0.06511	0.06322	**0.06153**
	45	0.05091	0.06389	0.06015	0.04266	0.04100	0.03940	0.03788	**0.03645**
	55	0.04054	0.05011	0.04737	0.03439	0.03314	0.03191	0.03072	**0.02958**
	75	0.02258	0.02842	0.02672	0.01890	0.01815	0.01740	0.01667	**0.01596**

The results given with Table 4.4 verify the Lindsay's criterion for ordinary minimum power-divergence estimators with large negative values of λ being robust. As seen from the table, MSE decreases as λ decreases. The performance of ordinary MLE is worse than the ordinary estimators with large negative values of λ for both parameters. This is also expected since MLE is non-robust. Even penalization does not make difference. Although $\hat{\theta}^{P^{0.5}}_{(0)}$ has lower MSE values than $\hat{\theta}^{I}_{(0)}$, ordinary minimum power-divergence estimators with large negative values of λ still perform better than both ordinary and penalized MLEs. The robustness property of minimum power-divergence estimators with large negative values of λ is not compromised by empty cell penalty. Regarding the performances of penalization values, penalization value of w = 0.5 can be considered as the best for all values of λ since estimators have smallest MSE values for both parameters.

To conclude, Pearson's chi-square estimator ($\hat{\theta}^{I}_{(1)} = \hat{\theta}^{P^{0.5}}_{(1)}$) is the most efficient estimator among ordinary and penalized minimum power-divergence estimators. However, it is not the most robust estimator. On the contrary, it has the largest MSE values for all sample sizes under the moderate contamination (ε = 10%). Penalization causes improvement in the small sample efficiency performances of the robust ordinary estimators. Moreover, their robustness property improves with the empty cell penalty let alone compromised by it. Among all estimators, penalized minimum power-divergence estimator $\hat{\theta}^{P^{0.5}}_{(-0.9)}$ is the most robust one.

4.2. Ordinary and Penalized Power-Divergence Test Statistics

4.2.1 The Family of Ordinary Power-Divergence Test Statistics

The work in this section has been motivated by the work of Cressie, Pardo, & Pardo (2003). Let's define the following nested sequence of hypotheses,

$$H_L: \mathbf{p} = \mathbf{p}(\boldsymbol{\theta}); \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}_L; \quad L = 1, ..., T, \quad T \leq M_0 \leq M\text{-}1, \tag{4.27}$$

where $\mathbf{p(\theta)} = (p_1(\theta), \ldots, p_M(\theta))^T$ and $\boldsymbol{\theta} = (\theta_1, \ldots \theta_{M_0})^T \in \boldsymbol{\Theta}_\mathbf{L}$. $\boldsymbol{\Theta}_\mathbf{L}$ is the parameter space for the H_L such that $\boldsymbol{\Theta}_\mathbf{T} \subset \boldsymbol{\Theta}_\mathbf{T\text{-}1} \subset \ldots \subset \boldsymbol{\Theta}_\mathbf{1} \equiv \boldsymbol{\Theta} \subseteq \Re^{M_0}$; $M_0 \leq$ M-1and $\dim(\boldsymbol{\Theta}_\mathbf{L}) = d_L$ with

$$d_T < d_{T-1} < \ldots < d_1 = M_0. \tag{4.28}$$

Cressie et al. (2003)'s strategy is to test successively,

$$H_{null} = H_{L+1} \text{ against } H_{alt} = H_L; \quad L = 1, \ldots, T-1, \tag{4.29}$$

until the first L for which H_{L+1} is rejected as a null hypothesis. To test these nested hypotheses, Cressie & Pardo (2000, 2002) have suggested the following families of test statistics

$$T(O)^{(L)}_{\phi_1,\phi_2} = \frac{2n}{\phi_1''(1)} I_{\phi_1}(\mathbf{p}(\hat{\boldsymbol{\theta}}^{I(L)}_{\phi_2}), \mathbf{p}(\hat{\boldsymbol{\theta}}^{I(L+1)}_{\phi_2})) \tag{4.30}$$

$$S(O)^{(L)}_{\phi_1,\phi_2} = \frac{2n}{\phi_1''(1)} \{I_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}^{I(L+1)}_{\phi_2})) - I_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}^{I(L)}_{\phi_2}))\} \tag{4.31}$$

where ϕ_1 and ϕ_2 are convex functions, and $\hat{\boldsymbol{\theta}}^{I(L+1)}_{\phi_2}$ and $\hat{\boldsymbol{\theta}}^{I(L)}_{\phi_2}$ are defined with Eq.(4.7) under the models of H_{L+1} and H_L , respectively.

When $T(O)^{(L)}_{\phi_1,\phi_2} > c$, H_{null} is rejected, where c is specified so that the size of the test is α;

$$P(T(O)^{(l)}_{\phi_1,\phi_2} > c \,|\, H_{L+1}) = \alpha \; ; \; \alpha \in (0, 1). \tag{4.32}$$

Cressie & Pardo (2000, 2002) have shown that under multinomial sampling with probabilities belonging to a log-linear model and $H_{null} = H_{L+1}$, the test statistics

$T(O)^{(L)}_{\phi_1,\phi_2}$ converges in distribution to a chi-squared distribution with degrees of freedom $d_L - d_{L+1}$; L = 1, ..., T-1. Hence,

$$c = \chi^2_{d_L - d_{L+1}}(1-\alpha), \tag{4.33}$$

where $P(\chi^2_f \le \chi^2_f(p)) = p$. Cressie & Pardo (2000) have established that $T(O)^{(L)}_{\phi_1,\phi_2}$ and $S(O)^{(L)}_{\phi_1,\phi_2}$ are asymptotically equivalent under null hypothesis. That is, under $H_{null}: H_{L+1}$, $S(O)^{(L)}_{\phi_1,\phi_2}$ has chi-squared distribution with $d_L - d_{L+1}$ degrees of freedom; L = 1, ..., T-1.

The families of the test statistics given by Eq.(4.30) and Eq. (4.31) generalize most of the known test statistics. For example; when $\phi_1 \equiv \phi_2 \equiv \phi_{(0)}$, $T(O)^{(L)}_{\phi_1,\phi_2}$ and $S(O)^{(L)}_{\phi_1,\phi_2}$ are both equal to the usual likelihood ratio test statistic, when $\phi_1 \equiv \phi_{(1)}$ and $\phi_2 \equiv \phi_{(0)}$ $T(O)^{(L)}_{\phi_1,\phi_2}$ gives the classical Pearson's chi-square test statistic. When $\phi_1 \equiv \phi_{(1)}$ and $\phi_2 \equiv \phi_{(0)}$, $S(O)^{(L)}_{\phi_1,\phi_2}$ yields a statistic based on the difference of Pearson's chi-square statistics with maximum likelihood estimation used to obtain expected frequencies. (e.g., Agresti (1996, p. 197)). However, the non-negativity of $S(O)^{(L)}_{\phi_1,\phi_2}$ does not hold when $\phi_1 \neq \phi_2$ as mentioned by Cressie & Pardo (2000). Thus, for the case considered by Agresti, the difference of Pearson's chi-square statistics is not necessarily non-negative. It should be remembered that all $\phi_{(\lambda)}$s here and in the simulation study are actually $\psi_{(\lambda)}$.

On the contrary to H_{null}, theoretical results for $T(O)^{(L)}_{\phi_1,\phi_2}$ and $S(O)^{(L)}_{\phi_1,\phi_2}$ under alternative hypotheses are not easy to obtain except for an appropriately specified sequence of contiguous (Pitman) alternatives. Regarding these alternatives, Haberman (1974) was the first to study them. He has proved that the asymptotic distribution of the likelihood ratio and Pearson's test statistics is non-centrally chi-

squared with $d_L - d_{L+1}$ degrees of freedom. Oler (1985) and Fenech & Westfall (1988) have presented studies regarding these contiguous alternative hypotheses in multinomial populations with probabilities belonging to log-linear models.

Later, Cressie et al. (2003) have defined the sequence of contiguous alternative hypotheses;

$$H_{L+l,n} : \mathbf{p} = \mathbf{p}_n(\boldsymbol{\theta}) = \mathbf{p}(\boldsymbol{\theta}) + \frac{\mathbf{s}}{\sqrt{n}}; \boldsymbol{\theta} \in \boldsymbol{\Theta}_{\mathbf{L+1}}, n \geq n_0 > 0, \tag{4.34}$$

which is contiguous to the null hypothesis H_{L+1} where $\mathbf{s} \equiv (s_1,...,s_M)^T$ is a fixed M x 1 vector such that $\sum_{j=1}^{M} s_j = 0$ and "n" is the total count parameter of the multinomial distribution. For multinomial sampling with probabilities having a log-linear model, they have proved that under the sequence of contiguous alternative hypotheses given by (4.34) $T(O)^{(L)}_{\phi_1,\phi_2}$ has non-central chi-squared distribution with $d_L - d_{L+1}$ degrees of freedom. Non-centrality parameter of this distribution is ,

$$\mu = \mathbf{s}^T \mathrm{diag}(\mathbf{p}(\boldsymbol{\theta})^{-1/2})(\mathbf{A}_{(L)} - \mathbf{A}_{(L+1)})\mathrm{diag}(\mathbf{p}(\boldsymbol{\theta})^{-1/2})\mathbf{s}\,, \tag{4.35}$$

where

$$\mathbf{A}_{(i)} = \mathrm{diag}(\mathbf{p}(\boldsymbol{\theta})^{-1/2})\boldsymbol{\Sigma}_{P(\theta)}\mathbf{W}_{(i)}(\mathbf{W}_{(i)}^T\boldsymbol{\Sigma}_{P(\theta)}\mathbf{W}_{(i)})^{-1}\mathbf{W}_{(i)}^T\boldsymbol{\Sigma}_{P(\theta)}\mathrm{diag}(\mathbf{p}(\boldsymbol{\theta})^{-1/2})\,;$$
$$i = L, L+1.$$

$\mathbf{W_{(i)}} = (\mathbf{w_1},...,\mathbf{w_M})^T$ is M x M_0 log-linear model matrix of explanatory variables under the model H_i, i = L, L+1, and $\boldsymbol{\Sigma}_{P(\theta)} = \mathrm{diag}\,\mathbf{p}(\boldsymbol{\theta}) - \mathbf{p}(\boldsymbol{\theta})\mathbf{p}(\boldsymbol{\theta})^T$. So, as $n \to \infty$,

$$\pi_n \to P(\chi^2_{d_L - d_{L+l},\mu} > c)\,, \tag{4.36}$$

where $\pi_n \equiv P(T(O)^{(l)}_{\phi_1,\phi_2} > c \mid H_{L+l,n})$, i.e. the power of the $T(O)^{(L)}_{\phi_1,\phi_2}$, and c is given by (4.33).

In the test of a single null hypothesis, H_{L+1} : $\mathbf{p} = \mathbf{p}(\theta)$, non-centrality parameter given with Eq.(4.35) is equal to;

$$\mu = \mathbf{s}^T \mathrm{diag}(\mathbf{p}(\boldsymbol{\theta})^{-1/2})(\mathbf{I} - \mathbf{A}_{(L+1)})\mathrm{diag}(\mathbf{p}(\boldsymbol{\theta})^{-1/2})\mathbf{s}\,, \tag{4.37}$$

where **I** is the M x M identity matrix.

Based on the results established by Cressie & Pardo (2000), Haberman (1974), Oler (1985) and Fenech & Westfall (1988), there is no doubt about that $T(O)^{(L)}_{\phi_1,\phi_2}$ and $S(O)^{(L)}_{\phi_1,\phi_2}$ are asymptotically equivalent under the sequence of contiguous alternative hypotheses given by Eq.(4.34)

4.2.2 The Family of Penalized Power-Divergence Test Statistics

As mentioned before, Basu & Basu (1998) have proposed an empty cell penalty for the minimum power-divergence estimators in multinomial models to overcome the problem of disproportionately large weight put on empty cells by the family of ordinary power-divergence measures. In this thesis, the following two families of penalized power-divergence test statistics has been defined by using penalized power-divergence measure and penalized minimum power-divergence estimators instead of ordinary ones in Eq.(4.30) and Eq.(4.31) to test nested hypotheses given with Eq.(4.29).

$$T(P^W)^{(L)}_{\phi_1,\phi_2} = \frac{2n}{\phi_1''(1)} P^W_{\phi_1}(\mathbf{p}(\hat{\boldsymbol{\theta}}^{P^W(L)}_{\phi_2}), \mathbf{p}(\hat{\boldsymbol{\theta}}^{P^W(L+1)}_{\phi_2})) \tag{4.38}$$

$$S(P^W)^{(L)}_{\phi_1,\phi_2} = \frac{2n}{\phi_1''(1)} \{P^W_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}^{P^W(L+1)}_{\phi_2})) - P^W_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}^{P^W(L)}_{\phi_2}))\} \tag{4.39}$$

where $\hat{\boldsymbol{\theta}}_{\phi_2}^{P^W(L+1)}$ and $\hat{\boldsymbol{\theta}}_{\phi_2}^{P^W(L)}$ are defined by Eq.(4.17) under the models of H_{L+1} and H_L , respectively. Since the families of ordinary power-divergence test statistics given by Eq.(4.30), Eq.(4.31) and the families of penalized power-divergence test statistics given with Eq.(4.38), Eq.(4.39) differ only in empty cells, they have the same asymptotic distribution under the null hypothesis and contiguous sequence of alternative hypotheses. Hence, it is only focused on small sample properties of these test statistics in the simulation studies. Following subsection includes the simulation studies which are performed to compare these test statistics under the log-linear models in two-way and three-way contingency tables.

4.2.3 Simulation Studies

This subsection will contain two parts: 1) Simulation study for two-way contingency tables 2) Simulation study for three-way contingency tables. For two-way contingency tables, the properties of test statistics given with equations (4.30) and (4.38) will be compared, while the test statistics compared will be the ones given with equations (4.31) and (4.39) for three-way contingency tables. Multinomial sampling distribution with probabilities belonging to class of log-linear models will be assumed for both two-way and three-way contingency tables. All calculations have been done using Mathematica 5.0 except the asymptotic power values which are obtained by using Matlab 5.3.

4.2.3.1 Simulation Study for Two-Way Contingency Tables

Aim of this simulation study is to compare the size and power properties of the families of ordinary and penalized power-divergence test statistics given with Eq.(4.30) and Eq.(4.38) for testing nested log-linear models in two-way contingency tables. In simulation study, the case of 2 x 2 contingency tables has been considered. So, total number of cells in contingency tables, M, is equal to IJ = 4. The models given with Eq.(3.5) and Eq.(3.7) are the ones used for the simulation. So, the two models considered are:

$$H_1: p_{ij}(\theta) = \exp[u + \theta_{1(i)} + \theta_{2(j)} + \theta_{12(ij)}]; \quad i, j = 1,2$$

$$H_2: p_{ij}(\theta) = \exp[u + \theta_{1(i)} + \theta_{2(j)}]; \quad i, j = 1,2.$$

$H_{null}: H_2$ will be tested against $H_{alt}: H_1$ or $H_{contalt}: H_{1,n}$ Testing the H_2 against the H_1 means to decide if interaction parameter $\theta_{12(ij)}$ should be included in the model or not. Let

$$\mathbf{p_0} \in H_{null}, \mathbf{p_1} \in H_{alt}, \mathbf{p_{1,n}} \in H_{contalt}.$$

Here, $H_{contalt}: H_{1,n}$ represents the following sequence of contiguous alternatives.

$$H_{contalt}: H_{1,n}: \mathbf{p_{1,n}} = \mathbf{p_0} + \sqrt{\frac{35}{n}}(\mathbf{p_1} - \mathbf{p_0})$$

As sample size "n" increases, $\mathbf{p_{1,n}}$ converges to $\mathbf{p_0}$ at the rate of $n^{-1/2}$; that is $\{\mathbf{p_{1,n}}\}$ is a sequence of contiguous alternatives. As easily noticed, $\mathbf{p_1}$ is also used in the definition of contiguous alternative, and $\mathbf{p_{1,35}} = \mathbf{p_1}$.

The arbitrarily chosen values of main effects and interaction effect are

$$\exp[\theta_{1(1)}] = \exp[\theta_{2(1)}] = \frac{1}{2} \text{ and } \exp[\theta_{12(11)}] = \frac{2}{3} \tag{4.40}$$

with the linear constraints given by Eq.(3.6). The probability vector $\mathbf{p_0}$ is equal to the probability vector given with Eq.(4.23).

The matrix form of the model given with H_1 is as follows:

$$\begin{bmatrix} \log p_1(\theta) \\ \log p_2(\theta) \\ \log p_3(\theta) \\ \log p_4(\theta) \end{bmatrix} = \begin{bmatrix} \log p_{11}(\theta) \\ \log p_{12}(\theta) \\ \log p_{21}(\theta) \\ \log p_{22}(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} u \\ \theta_{1(1)} \\ \theta_{2(1)} \\ \theta_{12(11)} \end{bmatrix}. \quad (4.41)$$

So, the log-linear model matrix (constraint matrix) of explanatory variables for H_1

$$\mathbf{W_{(1)}} = \begin{bmatrix} \mathbf{w_1} & \mathbf{w_2} & \mathbf{w_3} & \mathbf{w_4} \end{bmatrix}^T = \begin{bmatrix} \mathbf{w_1^T} \\ \mathbf{w_2^T} \\ \mathbf{w_3^T} \\ \mathbf{w_4^T} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \quad (4.42)$$

implies that

$$\exp(\theta_{1(2)}) = \exp(-\theta_{1(1)}) = 2 \text{ , } \exp(\theta_{2(2)}) = \exp(-\theta_{2(1)}) = 2. \quad (4.43)$$

$$\exp(\theta_{12(12)}) = \exp(\theta_{12(21)}) = \exp(-\theta_{12(11)}) = \frac{3}{2} \text{, } \exp(\theta_{12(22)}) = \exp(\theta_{12(11)}) = \frac{2}{3},$$

Overall mean effect parameter u is calculated as given by Eq.(4.22). By using equations (4.22), (4.40), (4.41) and (4.43), probability vector $\mathbf{p_1}$ is obtained as follows:

$$\mathbf{p_1} = \begin{bmatrix} 0.03 & 0.22 & 0.22 & 0.53 \end{bmatrix}^T \quad (4.44)$$

By this simulation study, exact probabilities given with equations (4.45), (4.46) and (4.47) will be obtained.

$$\alpha_n^{T(O)^{(L)}} \equiv P\left(T(O)_{\phi_1,\phi_2}^{(1)} > c \middle| \mathbf{p_0}\right) \quad \alpha_n^{T(P^W)^{(L)}} \equiv P\left(T(P^W)_{\phi_1,\phi_2}^{(1)} > c \middle| \mathbf{p_0}\right) \quad (4.45)$$

$$\pi_n^{T(P^W)^{(L)}} \equiv P\left(T(P^W)_{\phi_1,\phi_2}^{(L)} > c \middle| \mathbf{p_1}\right) \quad \pi_n^{T(O)^{(L)}} \equiv P\left(T(O)_{\phi_1,\phi_2}^{(L)} > c \middle| \mathbf{p_1}\right) \quad (4.46)$$

$$\pi_n^{T(O)^{(L)}} \equiv P(T(O)^{(L)}_{\phi_1,\phi_2} > c \,|\, \mathbf{p_{1,n}}) \quad \pi_n^{T(P^w)^{(L)}} \equiv P(T(P^w)^{(L)}_{\phi_1,\phi_2} > c \,|\, \mathbf{p_{1,n}}) \qquad (4.47)$$

Sample sizes considered are n = 25, 35, 45, 75, 100. Penalization values (w) are chosen as 0.5 and 1. Different combinations of ϕ_1 and ϕ_2 have been considered for the ordinary and penalized power-divergence test statistics. These combinations are as follows:

For ordinary power-divergence test statistics:

- $\phi_1 \equiv \phi_{(\lambda_1)}$, λ_1 = -0.9, -0.8, -0.7, -0.6, -0.5, 0, 2/3, 1; $\phi_2 \equiv \phi_{(\lambda_2)}$, λ_2 = 0, 1
- $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$, λ = -0.9, -0.8, -0.7, -0.6, -0.5, 0, 2/3, 1

For penalized power-divergence test statistics:

♦ For w = 1

- $\phi_1 \equiv \phi_{(\lambda_1)}$, λ_1 = -0.9, -0.8, -0.7, -0.6, -0.5, 0, 2/3, 1; $\phi_2 \equiv \phi_{(\lambda_2)}$, λ_2 = 0
- $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$, λ = -0.9, -0.8, -0.7, -0.6, -0.5, 0, 2/3, 1

♦ For w = 0.5

- $\phi_1 \equiv \phi_{(\lambda_1)}$, λ_1 = -0.9, -0.8, -0.7, -0.6, -0.5, 0, 2/3, 1; $\phi_2 \equiv \phi_{(\lambda_2)}$, λ_2 = -0.9, 0
- $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$, λ = -0.9, -0.8, -0.7, -0.6, -0.5, 0, 2/3, 1

Even though MLE is the mostly known and used estimator, minimum Pearson's chi-square estimator has been found as the most efficient estimator in simulation studies in Subsection 4.1.3. Hence, these estimators have been compared by using them in ordinary test statistics with different functions of ϕ_1. As mentioned by Read & Cressie (1988, p.32), in general it seems quite reasonable to match ϕ_1 and ϕ_2.

Hence, the performances of all test statistics when $\phi_1 \equiv \phi_2$ have also been studied. Even though minimum Pearson's chi-square estimator is the most efficient one in 2 x 2 contingency tables, minimum penalized power-divergence estimator with λ = -0.9 and w = 0.5 has been found as the most robust estimator in 2 x 2 contingency tables. Therefore, $\phi_2 \equiv \phi_{(-0.9)}$ has also been used in $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ as estimating function apart from $\phi_2 \equiv \phi_{(0)}$ which gives MLE. Exact probability estimations given with equations (4.45), (4.46) and (4.47) have been obtained using 500 simulations from the multinomial distributions with (n, $\mathbf{p_0}$), (n, $\mathbf{p}_1$) and (n, $\mathbf{p_{1,n}}$).

As Cressie et. al. (2003), two basic criteria for a good performance have been used to compare test statistics:

1) Good exact power and size,

2) Good agreement of exact and asymptotic probabilities.

For the first criteria, H_{null} : $\mathbf{p_0} \in H_2$ against H_{alt} : $\mathbf{p_1} \in H_1$ have been considered. For the second, H_{null} : $\mathbf{p_0} \in H_2$ against H_{alt} : $\mathbf{p_{1,n}}$ have been used. To compare the test statistics, following function values have been calculated.

$$g_1(\lambda_1,\lambda_2) \equiv \left| AP_{C,n}(\lambda_1,\lambda_2) - SEP_{C,n}(\lambda_1,\lambda_2) \right| \tag{4.48}$$

$$g_2(\lambda_1,\lambda_2) \equiv (SEP_{F,n}(\lambda_1,\lambda_2) - STS_{F,n}(\lambda_1,\lambda_2))^{-1}. \tag{4.49}$$

$AP_{C,n}(\lambda_1,\lambda_2)$ is the asymptotic power under contiguous alternative, $SEP_{C,n}(\lambda_1,\lambda_2)$ is the simulated exact power under contiguous alternative, $SEP_{F,n}(\lambda_1,\lambda_2)$ is the simulated exact power under fixed alternative and $STS_{F,n}(\lambda_1,\lambda_2)$ is the simulated test size of the test statistic $T(i)^{(L)}_{\phi_{(\lambda_1)},\phi_{(\lambda_2)}}$ where i = (O) ordinary, (P^w) penalized. According to Fenech & Westfall (1988), and Cressie et

al. (2003), approximation to the asymptotic power of $T(i)^{(L)}_{\phi(\lambda_1),\phi(\lambda_2)}$ can be obtained by defining $\mathbf{p}_{1,n} \equiv \mathbf{p}_0 + \frac{1}{\sqrt{n}}\mathbf{s}$ where $\mathbf{s} = \sqrt{n}(\mathbf{p}_1 - \mathbf{p}_0)$ and then substituting into the definition of μ, and then finally μ into the right side of Eq.(4.36). Since there is only one null hypothesis to test, non-centrality parameter μ can be obtained by Eq.(4.35) or (4.37). Following Fenech & Westfall (1988), and Cressie et al. (2003), and using the log-linear model matrices of explanatory variables for H_1 and H_2 given with Eq.(4.42) and Eq.(4.20), respectively, approximated asymptotic powers of $T(i)^{(L)}_{\phi(\lambda_1),\phi(\lambda_2)}$; i = (O) ordinary, ($P^w$) penalized, for each sample sizes have been found as follows:

Table 4.7 Asymptotic powers of $T(i)^{(L)}_{\phi(\lambda_1),\phi(\lambda_2)}$

n	25	35	45	75	100
Asymptotic Power	0.3011	0.3981	0.4877	0.7020	0.8200

A statistic $T(i)^{(L)}_{\phi(\lambda_1),\phi(\lambda_2)}$ will be considered to be better than other statistic $T(i)^{(L)}_{\phi(\lambda_1^*),\phi(\lambda_2^*)}$ iff

$$g_1(\lambda_1,\lambda_2) < g_1(\lambda_1^*,\lambda_2^*) \text{ and } g_2(\lambda_1,\lambda_2) < g_2(\lambda_1^*,\lambda_2^*). \tag{4.50}$$

For these comparisons, only the statistics that satisfy the following inequality are considered.

$$\left|\text{logit}(1-\alpha_n^{T(i)^{(L)}}) - \text{logit}(1-\alpha)\right| \le e, \tag{4.51}$$

for i = (O) ordinary, (P^w) penalized. logit(p) is equal to $\ln\left(\frac{p}{1-p}\right)$. This inequality has been proposed by Dale (1986) and measures the closeness of nominal size and exact

size obtained from the simulation. The two probabilities are considered fairly close if they satisfy Eq.(4.51) with e = 0.7. For $\alpha = 0.05$, which is considered as nominal size in this study, e = 0.7 corresponds to

$$\alpha_n^{T(i)^{(L)}} \in [0.0254, 0.0959]. \tag{4.52}$$

The simulation results are given with Tables 4.8 - 4.25. Tables 4.8 - 4.13, Tables 4.14 - 4.19 and Tables 4.20 - 4.25 show the exact sizes, (exact power-asymptotic power) and (exact power - exact size)$^{-1}$ of $T(i)^{(1)}_{\phi(\lambda_1),\phi(\lambda_2)}$, respectively. Since $I_0(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = P_0^1(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$ and $I_1(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta})) = P_1^{0.5}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\theta}))$, $T(O)^{(1)}_{\phi_{(0)},\phi_{(0)}} = T(P^1)^{(1)}_{\phi_{(0)},\phi_{(0)}}$ and $T(O)^{(1)}_{\phi_{(1)},\phi_{(1)}} = T(P^{0.5})^{(1)}_{\phi_{(1)},\phi_{(1)}}$. Hence, the values in the tables corresponding to these test statistics are equal.

For n = 25, as seen in Table 4.8 and Table 4.9 the usual likelihood ratio ($T(O)^{(1)}_{\phi_{(0)},\phi_{(0)}}$), Pearson's chi-square ($T(O)^{(1)}_{\phi_{(1)},\phi_{(0)}}$) and Cressie-Read ($T(O)^{(1)}_{\phi_{(2/3)},\phi_{(0)}}$) test statistics are not even close to satisfying Eq.(4.52). For negative values of λ, especially large negative values, exact sizes of $T(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ are far behind satisfying Eq.(4.52). Penalization, especially w = 0.5, improves the exact sizes of all ordinary test statistics.

The results in Table 4.12 and Table 4.13 show the improvement in exact sizes of all ordinary test statistics by penalization. As will be noticed, improvement is much more obvious in the ordinary power-divergence test statistics with large negative values of λ such that the statistics that satisfy Eq.(4.52) correspond to $T(P^{0.5})^{(1)}_{\phi_{(-0.7)},\phi_{(-0.9)}}$, $T(P^{0.5})^{(1)}_{\phi_{(-0.8)},\phi_{(-0.9)}}$, $T(P^{0.5})^{(1)}_{\phi_{(-0.6)},\phi_{(-0.6)}}$, $T(P^{0.5})^{(1)}_{\phi_{(-0.8)},\phi_{(-0.8)}}$ and $T(P^{0.5})^{(1)}_{\phi_{(-0.9)},\phi_{(-0.9)}}$. It will be checked which one/ones of these test statistics satisfy the criterion given with Eq.(4.50).

Table 4.8 Exact sizes of $T(O)^{(1)}_{\phi_1,\phi_2}$ under null hypothesis H_{null}: $\mathbf{p_0} \in H_2$ with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2$

		λ_1							
n	λ_2	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0	0.1560	0.1360	0.1360	0.2200	0.2600	0.2960	0.3160	0.3480
	1	0.1820	0.0980	0.1120	0.2860	0.3160	0.3340	0.3700	0.3920
35	0	0.1260	0.1040	0.1040	0.1780	0.2020	0.2300	0.2560	0.2900
	1	0.1500	0.0720	0.0860	0.2320	0.2560	0.2800	0.3020	0.3340
45	0	0.0920	0.0640	0.0600	0.1580	0.1760	0.1920	0.2080	0.2160
	1	0.1040	0.0560	0.0600	0.1860	0.2080	0.2220	0.2320	0.2400
75	0	0.0740	0.0540	0.0540	0.1160	0.1180	0.1200	0.1240	0.1240
	1	0.0940	0.0420	0.0520	0.1320	0.1340	0.1400	0.1420	0.1500
100	0	0.0500	0.0440	0.0440	0.0540	0.0540	0.0540	0.0560	0.0560
	1	0.0520	0.0380	0.0440	0.0600	0.0620	0.0660	0.0700	0.0720

Table 4.9 Exact sizes of $T(O)^{(1)}_{\phi_1,\phi_2}$ under null hypothesis H_{null}: $\mathbf{p_0} \in H_2$ with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0.1560	0.0980	0.1080	0.1880	0.1960	0.2040	0.2220	0.2440
35	0.1260	0.0720	0.0840	0.1640	0.1720	0.1860	0.2020	0.2300
45	0.0920	0.0560	0.0580	0.1440	0.1560	0.1740	0.1880	0.1980
75	0.0740	0.0420	0.0520	0.1120	0.1140	0.1160	0.1180	0.1160
100	0.0500	0.0380	0.0440	0.0540	0.0540	0.0540	0.0520	0.0500

Table 4.10 Exact sizes of $T(P^1)^{(1)}_{\phi_1,\phi_2}$ under null hypothesis H_{null}: $\mathbf{p_0} \in H_2$ with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2 = 0$

	λ_1							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0.1560	0.1780	0.1720	0.1420	0.1460	0.1460	0.1460	0.1460
35	0.1260	0.1440	0.1360	0.1180	0.1200	0.1200	0.1240	0.1240
45	0.0920	0.1080	0.1000	0.0920	0.0940	0.0920	0.0920	0.0920
75	0.0740	0.0840	0.0800	0.0880	0.0880	0.0900	0.0940	0.0940
100	0.0500	0.0480	0.0460	0.0500	0.0500	0.0500	0.0520	0.0520

Table 4.11 Exact sizes of $T(P^1)^{(1)}_{\phi_1,\phi_2}$ under null hypothesis H_{null}:$\mathbf{p_0} \in H_2$ with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0.1560	0.1540	0.1560	0.1380	0.1320	0.1320	0.1300	0.1260
35	0.1260	0.1240	0.1280	0.1140	0.1100	0.1020	0.1000	0.1020
45	0.0920	0.1000	0.0980	0.0920	0.0840	0.0820	0.0800	0.0800
75	0.0740	0.0780	0.0780	0.0860	0.0840	0.0860	0.0880	0.0860
100	0.0500	0.0440	0.0460	0.0500	0.0500	0.0500	0.0480	0.0460

Table 4.12 Exact sizes of $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ under null hypothesis H_{null}: $\mathbf{p_0} \in H_2$ with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2$

		λ_1							
n	λ_2	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0	0.1020	0.1160	0.1160	0.1000	0.1040	0.1000	0.1000	0.1000
	-0.9	0.1260	0.1920	0.1740	0.0980	0.0980	0.0880	0.0860	0.0820
35	0	0.0780	0.0920	0.0840	0.0700	0.0720	0.0720	0.0760	0.0760
	-0.9	0.0940	0.1420	0.1300	0.0660	0.0620	0.0600	0.0540	0.0540
45	0	0.0480	0.0640	0.0560	0.0480	0.0500	0.0520	0.0520	0.0520
	-0.9	0.0580	0.1000	0.0880	0.0480	0.0480	0.0420	0.0400	0.0400
75	0	0.0380	0.0480	0.0440	0.0520	0.0520	0.0540	0.0580	0.0580
	-0.9	0.0520	0.0640	0.0620	0.0500	0.0500	0.0500	0.0520	0.0500
100	0	0.0440	0.0420	0.0400	0.0440	0.0440	0.0440	0.0460	0.0460
	-0.9	0.0460	0.0480	0.0480	0.0440	0.0440	0.0440	0.0440	0.0400

Table 4.13 Exact sizes of $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ under null hypothesis H_{null}: $\mathbf{p_0} \in H_2$ with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0.1020	0.0980	0.1000	0.0980	0.0880	0.0880	0.0860	0.0820
35	0.0780	0.0720	0.0760	0.0660	0.0620	0.0540	0.0520	0.0540
45	0.0480	0.0560	0.0540	0.0460	0.0440	0.0420	0.0400	0.0400
75	0.0380	0.0420	0.0420	0.0500	0.0480	0.0500	0.0520	0.0500
100	0.0440	0.0380	0.0400	0.0440	0.0440	0.0440	0.0420	0.0400

Table 4.14 (Exact power-Asymptotic power) of $T(O)^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$, with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2$

		λ_1							
n	λ_2	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0	0.0029	-0.0491	-0.0331	0.0549	0.0749	0.0929	0.1109	0.1229
	1	0.0169	-0.0691	-0.0471	0.0849	0.1049	0.1189	0.1409	0.1469
35	0	-0.0161	-0.0861	-0.0721	0.0319	0.0459	0.0499	0.0579	0.0579
	1	-0.0021	-0.1161	-0.0821	0.0679	0.0759	0.0759	0.0819	0.0819
45	0	-0.0197	-0.1177	-0.0997	0.0203	0.0243	0.0323	0.0363	0.0363
	1	-0.0057	-0.1617	-0.1077	0.0503	0.0543	0.0543	0.0623	0.0623
75	0	-0.0160	-0.0860	-0.0680	0.0160	0.0220	0.0260	0.0340	0.0440
	1	0.0060	-0.1220	-0.0760	0.0460	0.0480	0.0620	0.0660	0.0680
100	0	-0.0180	-0.0800	-0.0580	0.0040	0.0040	0.0040	0.0060	0.0080
	1	-0.0100	-0.0900	-0.0600	0.0080	0.0180	0.0220	0.0280	0.0320

Table 4.15 (Exact power-Asymptotic power) of $T(O)^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$, with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0.0029	-0.0691	-0.0491	0.0449	0.0569	0.0649	0.0789	0.0929
35	-0.0161	-0.1161	-0.0821	0.0319	0.0359	0.0439	0.0439	0.0479
45	-0.0197	-0.1617	-0.1117	0.0063	0.0143	0.0183	0.0183	0.0223
75	-0.0160	-0.1220	-0.0760	0.0100	0.0160	0.0220	0.0260	0.0260
100	-0.0180	-0.0900	-0.0640	0.0020	0.0040	0.0040	0.0040	0.0040

Table 4.16 (Exact power-Asymptotic power) of $T(P^1)^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$, with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2 = 0$

	λ_1							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0.0029	-0.0011	-0.0011	0.0009	0.0009	0.0009	0.0009	0.0009
35	-0.0161	-0.0321	-0.0301	-0.0141	-0.0101	-0.0081	-0.0001	-0.0001
45	-0.0197	-0.0497	-0.0477	0.0003	0.0043	0.0083	0.0123	0.0123
75	-0.0160	-0.0680	-0.0520	0.0160	0.0220	0.0260	0.0340	0.0440
100	-0.0180	-0.0800	-0.0580	0.0040	0.0040	0.0040	0.0060	0.0080

Table 4.17 (Exact power-Asymptotic power) of $T(P^1)^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$, with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0.0029	-0.0171	-0.0111	-0.0051	-0.0051	-0.0051	-0.0051	-0.0051
35	-0.0161	-0.0421	-0.0321	-0.0141	-0.0141	-0.0141	-0.0141	-0.0101
45	-0.0197	-0.0737	-0.0497	-0.0097	-0.0057	-0.0057	-0.0057	-0.0017
75	-0.0160	-0.0920	-0.0600	0.0100	0.0160	0.0220	0.0260	0.0260
100	-0.0180	-0.0900	-0.0640	0.0020	0.0040	0.0040	0.0040	0.0040

Table 4.18 (Exact power-Asymptotic power) of $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$, with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2$

		λ_1							
n	λ_2	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0	-0.0471	-0.0571	-0.0551	-0.0491	-0.0491	-0.0491	-0.0491	-0.0491
	-0.9	-0.0471	-0.0271	-0.0351	-0.0491	-0.0491	-0.0491	-0.0491	-0.0491
35	0	-0.0901	-0.1061	-0.1041	-0.0861	-0.0821	-0.0801	-0.0721	-0.0721
	-0.9	-0.0841	-0.0741	-0.0821	-0.0861	-0.0861	-0.0861	-0.0821	-0.0821
45	0	-0.1077	-0.1377	-0.1357	-0.0877	-0.0837	-0.0757	-0.0717	-0.0717
	-0.9	-0.0957	-0.0877	-0.0897	-0.0917	-0.0897	-0.0897	-0.0897	-0.0857
75	0	-0.0460	-0.0980	-0.0820	-0.0140	-0.0080	-0.0040	0.0040	0.0140
	-0.9	-0.0240	-0.0420	-0.0360	-0.0140	-0.0080	-0.0080	-0.0040	-0.0040
100	0	-0.0180	-0.0800	-0.0580	0.0040	0.0040	0.0040	0.0060	0.0080
	-0.9	-0.0080	-0.0240	-0.0160	0.0040	0.0040	0.0040	0.0040	0.0040

Table 4.19 (Exact power-Asymptotic power) of $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$, with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	-0.0471	-0.0691	-0.0631	-0.0491	-0.0491	-0.0491	-0.0491	-0.0491
35	-0.0901	-0.1161	-0.1061	-0.0861	-0.0861	-0.0861	-0.0861	-0.0821
45	-0.1077	-0.1617	-0.1377	-0.0977	-0.0937	-0.0897	-0.0897	-0.0857
75	-0.0460	-0.1220	-0.0900	-0.0200	-0.0140	-0.0080	-0.0040	-0.0040
100	-0.0180	-0.0900	-0.0640	0.0020	0.0040	0.0040	0.0040	0.0040

As seen in Table 4.18 and Table 4.19, $g_1(\lambda_1,\lambda_2)$ values of $T(P^{0.5})^{(1)}_{\phi_{(-0.7)},\phi_{(-0.9)}}$, $T(P^{0.5})^{(1)}_{\phi_{(-0.8)},\phi_{(-0.9)}}$ and $T(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for λ = -0.6, -0.7, -0.8, -0.9, are all equal. However, as easily seen in Tables 4.24 and 4.25, the only statistic that satisfies both of the inequalities in the criterion given with Eq.(4.50) is $T(P^{0.5})^{(1)}_{\phi_{(-0.9)},\phi_{(-0.9)}}$.

Table 4.20 (Exact power-Exact size)$^{-1}$ of $T(O)^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$ with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2$

		λ_1							
n	λ_2	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0	9.4340	12.8205	10.4167	10.2041	12.5000	16.6667	15.1515	20.8333
	1	10.6383	10.8696	9.8039	15.1515	17.8571	17.8571	20.0000	29.4118
35	0	3.9063	4.8077	4.5045	3.9683	4.1322	4.5872	5.0000	6.0241
	1	4.0650	4.7619	4.3478	4.2735	4.5872	5.1546	5.6180	6.8493
45	0	2.3148	2.6316	2.4876	2.5907	2.6882	2.7473	2.8571	2.9240
	1	2.3364	2.9070	2.5253	2.6455	2.7933	2.9070	2.9240	2.9940
75	0	1.6393	1.7182	1.6949	1.7065	1.7007	1.7065	1.6949	1.6611
	1	1.6667	1.7483	1.7007	1.6722	1.6556	1.6556	1.6447	1.6611
100	0	1.3333	1.3966	1.3699	1.3055	1.2987	1.2953	1.2987	1.2953
	1	1.3123	1.4045	1.3736	1.2953	1.2920	1.2887	1.2788	1.2723

Table 4.21 (Exact power-Exact size)$^{-1}$ of $T(O)^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$ with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	9.4340	10.8696	10.0000	8.1967	7.9365	7.9365	8.0645	7.9365
35	3.9063	4.7619	4.3103	3.7594	3.8168	3.9063	4.1667	4.6296
45	2.3148	2.9070	2.5253	2.5510	2.6042	2.7027	2.8090	2.8571
75	1.6393	1.7483	1.7065	1.7007	1.7007	1.6949	1.7007	1.6949
100	1.3333	1.4045	1.3812	1.3055	1.3055	1.3021	1.2953	1.2920

Table 4.22 (Exact power-Exact size)$^{-1}$ of $T(P^1)^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$ with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2 = 0$

	λ_1							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	9.4340	11.9048	11.6279	8.1967	8.4746	8.4746	8.4746	8.4746
35	3.9063	4.5045	4.3103	3.7594	3.7313	3.7037	3.6496	3.6496
45	2.3148	2.5907	2.5253	2.2422	2.2321	2.1930	2.1834	2.1834
75	1.6393	1.8116	1.7730	1.6287	1.6181	1.6234	1.6129	1.5823
100	1.3333	1.4045	1.3736	1.2987	1.2920	1.2887	1.2920	1.2887

Table 4.23 (Exact power-Exact size)$^{-1}$ of $T(P^1)^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$ with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	9.4340	11.6279	10.6383	7.9365	7.6923	7.6923	7.5758	7.3529
35	3.9063	4.3103	4.2017	3.7037	3.6496	3.5461	3.5211	3.4965
45	2.3148	2.7027	2.5253	2.2831	2.2321	2.2026	2.1930	2.1739
75	1.6393	1.8657	1.7857	1.6287	1.6181	1.6129	1.6181	1.6129
100	1.3333	1.4164	1.3850	1.2987	1.2987	1.2953	1.2887	1.2853

Table 4.24 (Exact power-Exact size)$^{-1}$ of $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$ with $\phi_1 \equiv \phi_{(\lambda_1)} \equiv \lambda_1$ and $\phi_2 \equiv \phi_{(\lambda_2)} \equiv \lambda_2$

		λ_1							
n	λ_2	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	0	8.7719	10.8696	10.8696	8.3333	8.6207	8.6207	8.6207	8.6207
	-0.9	10.6383	20.8333	16.1290	8.1967	8.1967	7.9365	7.8125	7.5758
35	0	4.3478	5.0000	4.7619	4.1322	4.0984	4.0650	4.0000	4.0000
	-0.9	4.5455	5.4945	5.3763	4.0650	4.0000	3.9683	3.8168	3.8168
45	0	2.4631	2.7778	2.7027	2.3810	2.3697	2.3364	2.3256	2.3256
	-0.9	2.4876	2.7322	2.6455	2.4038	2.3810	2.3474	2.3364	2.3148
75	0	1.5480	1.7007	1.6667	1.5385	1.5291	1.5337	1.5244	1.4970
	-0.9	1.5576	1.6129	1.5924	1.5385	1.5337	1.5244	1.5291	1.5244
100	0	1.3228	1.3928	1.3624	1.2887	1.2821	1.2788	1.2821	1.2788
	-0.9	1.3089	1.3298	1.3228	1.2887	1.2887	1.2853	1.2821	1.2755

Table 4.25 (Exact power-Exact size)$^{-1}$ of $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$ with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)} \equiv \lambda$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
25	8.7719	10.8696	10.0000	8.1967	7.9365	7.9365	7.8125	7.5758
35	4.3478	4.7619	4.6296	4.0650	4.0000	3.8760	3.8462	3.8168
45	2.4631	2.9070	2.7027	2.4155	2.3810	2.3474	2.3364	2.3148
75	1.5480	1.7483	1.6779	1.5385	1.5291	1.5244	1.5291	1.5244
100	1.3228	1.4045	1.3736	1.2887	1.2887	1.2853	1.2788	1.2755

For n = 35, while the likelihood ratio and Pearson's chi-square test statistics do not satisfy Eq.(4.52), $T(O)^{(1)}_{\phi_{(1)},\phi_{(1)}}$, $T(O)^{(1)}_{\phi_{(2/3)},\phi_{(1)}}$ and $T(O)^{(1)}_{\phi_{(2/3)},\phi_{(2/3)}}$ do. In addition to these ordinary test statistics, all of $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ for $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$, $T(P^{0.5})^{(1)}_{\phi_{(\lambda_1)},\phi_{(0)}}$ and most of the $T(P^{0.5})^{(1)}_{\phi_{(\lambda_1)},\phi_{(-0.9)}}$ have exact sizes within the interval [0.0254, 0.0959]. When the results given in Table 4.14 - 4.25 have been checked, the statistic among these that satisfies both inequalities given with Eq.(4.50) is again $T(P^{0.5})^{(1)}_{\phi_{(-0.9)},\phi_{(-0.9)}}$.

To better understand the improvement in the small sample performance of the test statistics $T(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$,with large negative values of λ, due to penalty, histograms of the exact null distributions of $T(O)^{(1)}_{\phi_{(-0.8)},\phi_{(-0.8)}}$ and $T(P^{0.5})^{(1)}_{\phi_{(-0.8)},\phi_{(-0.8)}}$ are plotted for n= 25. The vertical dashed lines on the histograms correspond to the 5% critical point of χ^2_1 (i.e. $c = \chi^2_1(0.95) = 3.84$ as defined by Eq.(4.33)). The right hand tail area of this dashed line shows how well χ^2_1 density approximates.

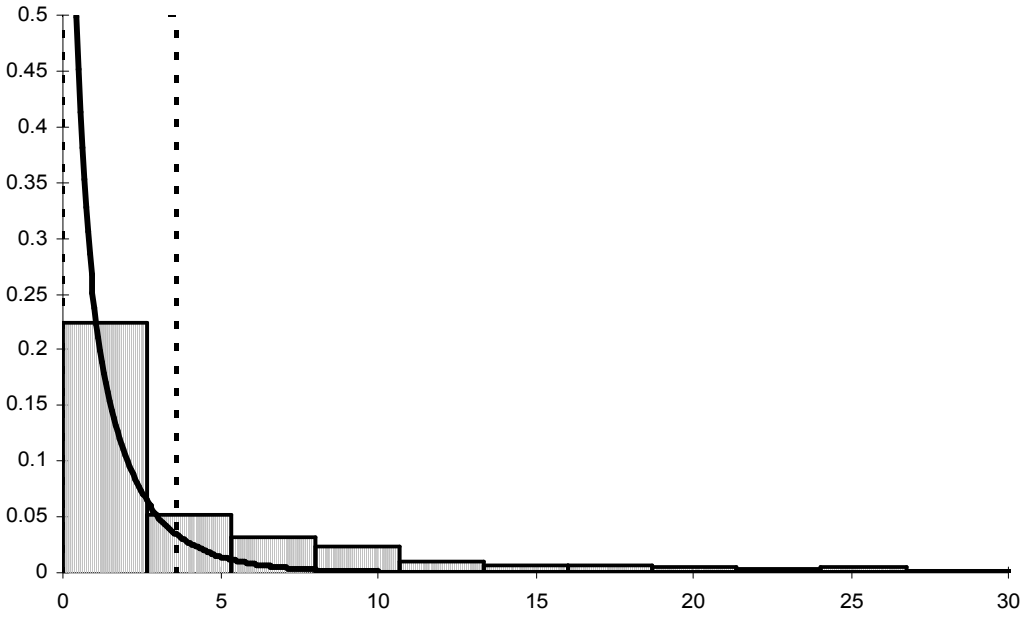

Figure 4.1 Histogram of the null distribution of $T(O)^{(1)}_{\phi_{(-0.8)},\phi_{(-0.8)}}$ with density function of χ^2_1 superimposed.

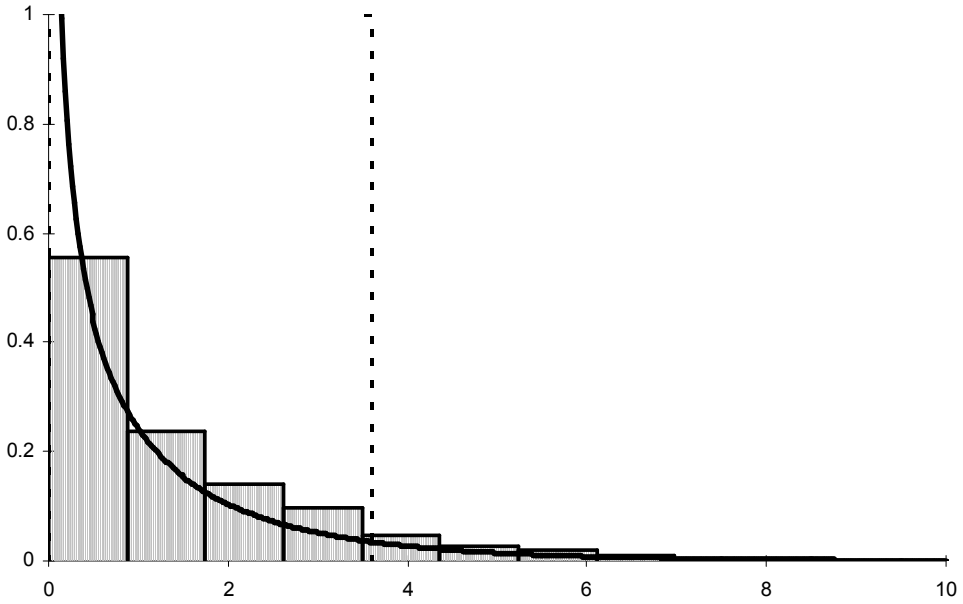

Figure 4.2 Histogram of the null distribution of $T(P^{0.5})^{(1)}_{\phi_{(-0.8)},\phi_{(-0.8)}}$ with density function of χ^2_1 superimposed.

Figure 4.1 shows the histogram of $T(O)^{(1)}_{\phi_{(-0.8)},\phi_{(-0.8)}}$, where the poor approximation to the χ^2_1 density is evident from the relatively heavy tail of the histogram behind the dashed line. On the other hand, the right tail area of the histogram given with Figure 4.2 shows the improvement in the distribution of $T(O)^{(1)}_{\phi_{(-0.8)},\phi_{(-0.8)}}$ by penalization value of w = 0.5.

It should be added that penalization can also cause increment in the exact sizes of some statistics when penalization value puts more weight on empty cells than ordinary test statistic. For instance, $T(P^1)^{(1)}_{\phi_{(1)},\phi_{(0)}}$ and $T(P^1)^{(1)}_{\phi_{(2/3)},\phi_{(0)}}$ given in Table 4.10 have larger exact sizes than their ordinary counterparts given in Table 4.8 because penalization value w = 1 put more weight on empty cells.

As mentioned before, increasing or decreasing the penalty value effects the sensitivity of the test statistics to empty cells. According to the results, w = 0.5 performs much better than w = 1 in small samples. However, this difference decreases as sample size gets larger and w = 1 starts to perform better. For n = 45, the

results in tables reveal that, more statistics have exact sizes within the interval [0.0254, 0.0959] with some statistics having closer exact sizes to nominal size 0.05; such as, $T(O)^{(1)}_{\phi_{(1)},\phi_{(1)}}$, $T(O)^{(1)}_{\phi_{(2/3)},\phi_{(2/3)}}$. Among the test statistics having exact sizes within that interval, penalized ones with w = 1 perform better according to the criteria given with Eq.(4.50). When ϕ_1 and ϕ_2 combinations have been compared in $T(P^1)^{(1)}_{\phi_1,\phi_2}$, it is seen that there is not much difference among $T(P^1)^{(1)}_{\phi_{(\lambda_1)},\phi_{(0)}}$ and $T(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for $\lambda_1 = \lambda$ = -0.9, -0.8, -0.7, -0.6, -0.5. However, large negative values of λ and λ_1 perform a little bit better.

The more the sample size gets larger, the more statistics satisfy the criterion given with Eq.(4.52) with more statistics having exact sizes closer to nominal size. For n = 75, the statistics except the ones $T(O)^{(1)}_{\phi_{(\lambda_1)},\phi_{(0)}}$, $T(O)^{(1)}_{\phi_{(\lambda_1)},\phi_{(1)}}$ and $T(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for $\lambda_1 = \lambda$ = -0.9, -0.8, -0.7, -0.6, -0.5 satisfy the criterion Eq.(4.52). Test statistics satisfying Eq.(4.52) are almost equivalent according to the criteria given with Eq.(4.50). But, penalized test statistics with negative values of λ perform better for all combinations of ϕ_1 and ϕ_2. It is mentioned that all of the ordinary and penalized power-divergence test statistics are asymptotically equivalent under null and contiguous alternative hypotheses. This is already seen for n = 100 in all tables in which all of the test statistics have exact sizes very close to nominal size 0.05. Moreover, there is not much difference among the test statistics in terms of the criteria given with Eq.(4.50), either. However, penalized test statistics with w = 0.5 seem slightly better among them.

In general, using the well known estimator MLE with $\phi_1 \equiv \phi_{(\lambda_1)}$ for negative values of λ_1 causes the increment in the exact sizes of $T(O)^{(1)}_{\phi_{(\lambda_1)},\phi_{(0)}}$ for the small and moderate samples. This is also true for Pearson's minimum chi-squared estimator. Moreover, using Pearson estimator with $\phi_{(2/3)}$ and $\phi_{(1)}$ make the exact sizes to decrease, whereas using this estimator with $\phi_{(0)}$ causes increment in small samples.

For small samples, $T(O)^{(1)}_{\phi_{(2/3)},\phi_{(1)}}$ performs better than $T(O)^{(1)}_{\phi_{(2/3)},\phi_{(0)}}$ (Cressie-Read statistic). But, as n gets larger, using the MLE or Pearson's minimum chi-squared estimator does not make any difference and $T(O)^{(1)}_{\phi_{(\lambda_1)},\phi_{(0)}}$ and $T(O)^{(1)}_{\phi_{(\lambda_1)},\phi_{(1)}}$ get asymptotically equivalent as all of power-divergence test statistics. Whether they satisfy (4.52) or not, all $T(i)^{(1)}_{\phi_1,\phi_2}$ for i = (O) ordinary, (P^w) penalized with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$, perform better for all sample sizes than the ones in which MLE, minimum Pearson's chi-squared estimator or $\hat{\theta}^{P^{0.5}}_{(-0.9)}$ used as an estimator.

4.2.3.2 Simulation Study for Three-Way Contingency Tables

In simulation study, the case of 2 x 2 x 2 contingency tables has been considered. So total number of cells in contingency tables, M, is equal to IJK = 8. Among the nested models, only the following two models have been considered.

$$H_1: p_{ijk}(\theta) = \exp[u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{12(ij)}]; \quad i, j, k = 1, 2$$

$$H_2: p_{ijk}(\theta) = \exp[u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)}]; \quad i, j, k = 1, 2$$

H_1 is jointly independence model given with Eq.(3.14), and H_2 is independence model given with Eq.(3.15). As in two-way contingency table case, H_{null}: H_2 will be tested against H_{alt}: H_1 or $H_{contalt}$: $H_{1,n}$. Let

$$\mathbf{p_0} \in H_{null}, \mathbf{p_1} \in H_{alt}, \mathbf{p_{1,n}} \in H_{contalt}.$$

Here, $H_{contalt}$: $H_{1,n}$ represents the following sequence of contiguous alternatives.

$$H_{contalt}: H_{1,n}: \mathbf{p_{1,n}} = \mathbf{p_0} + \sqrt{\frac{25}{n}}(\mathbf{p_1} - \mathbf{p_0})$$

Following Oler (1985), main effects and interaction effect are chosen as follows:

$$\exp[\theta_{1(1)}] = \exp[\theta_{2(1)}] = \exp[\theta_{3(1)}] = \frac{5}{6}, \exp[\theta_{12(11)}] = \frac{3}{4} \tag{4.53}$$

with the linear constraints given by Eq.(3.11). The matrix form of the model given with H_1 is given with Eq.(4.54).

$$\begin{bmatrix} \log p_1(\theta) \\ \log p_2(\theta) \\ \log p_3(\theta) \\ \log p_4(\theta) \\ \log p_5(\theta) \\ \log p_6(\theta) \\ \log p_7(\theta) \\ \log p_8(\theta) \end{bmatrix} = \begin{bmatrix} \log p_{111}(\theta) \\ \log p_{112}(\theta) \\ \log p_{121}(\theta) \\ \log p_{122}(\theta) \\ \log p_{211}(\theta) \\ \log p_{212}(\theta) \\ \log p_{221}(\theta) \\ \log p_{222}(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} u \\ \theta_{1(1)} \\ \theta_{2(1)} \\ \theta_{3(1)} \\ \theta_{12(11)} \end{bmatrix} \tag{4.54}$$

So, the log-linear model matrix (constraint matrix) of explanatory variables for H_1

$$\mathbf{W_{(1)}} = \begin{bmatrix} \mathbf{w_1^T} \\ \mathbf{w_2^T} \\ \mathbf{w_3^T} \\ \mathbf{w_4^T} \\ \mathbf{w_5^T} \\ \mathbf{w_6^T} \\ \mathbf{w_7^T} \\ \mathbf{w_8^T} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{bmatrix} \tag{4.55}$$

implies that

$$\exp[\theta_{1(2)}] = \exp[-\theta_{1(1)}] = \frac{6}{5}, \exp[\theta_{2(2)}] = \exp[-\theta_{2(1)}] = \frac{6}{5},$$

$$\exp[\theta_{3(2)}] = \exp[-\theta_{3(1)}] = \frac{6}{5}, \tag{4.56}$$

$$\exp[\theta_{12(12)}] = \exp[\theta_{12(21)}] = \exp[-\theta_{12(11)}] = \frac{4}{3},$$

$$\exp[\theta_{12(22)}]=\exp[\theta_{12(11)}] = \frac{3}{4}, \tag{4.57}$$

Overall mean effect parameter u is calculated as given by Eq.(4.58).

$$u = -\log(\sum_{j=1}^{4} \exp(\mathbf{w}_j^T \boldsymbol{\theta})) \tag{4.58}$$

By using equations (4.53), (4.54), (4.56), (4.57) and (4.58) model probability vector $\mathbf{p_1}$ is obtained as follows:

$$\mathbf{p_1} = [0.05 \quad 0.07 \quad 0.13. \quad 0.18 \quad 0.13 \quad 0.18 \quad 0.11 \quad 0.15]^T \tag{4.59}$$

The model H_2 does not include interaction term. Hence, constraint matrix $\mathbf{W_{(2)}}$ of H_2 will have three columns. That is, $\mathbf{W_{(2)}}$ will not have the last column which $\mathbf{W_{(1)}}$ given with Eq.(4.55) has. Parameter vector for model H_2 will be

$$\boldsymbol{\theta}^* = [u \quad \theta_{1(1)} \quad \theta_{2(1)} \quad \theta_{3(1)}]^T.$$

So, model probability vector $\mathbf{p_0}$ is obtained as follows:

$$\mathbf{p_0} = [0.07 \quad 0.10 \quad 0.10. \quad 0.14 \quad 0.10 \quad 0.14 \quad 0.14 \quad 0.21]^T \tag{4.60}$$

As in the simulation study for two-way contingency tables, exact probabilities given with equations (4.45), (4.46) and (4.47) will be obtained. However, in three-dimensional case, $S(O)^{(1)}_{\phi_1,\phi_2}$ and $S(P^w)^{(1)}_{\phi_1,\phi_2}$ are the families of test statistics compared. Sample sizes considered for this simulation study are n = 15, 25, 35. To show that $S(O)^{(1)}_{\phi_1,\phi_2}$ and $S(P^w)^{(1)}_{\phi_1,\phi_2}$ are asymptotically chi-squared with $d_1 - d_2$ degrees of

freedom, which equals to one in 2 x 2 x 2 case, under $H_{null}: H_2$, $\alpha_n^{S(O)^{(1)}}$ and $\alpha_n^{S(P^w)^{(1)}}$ for n = 45, 75 are also calculated. But these values are not included in the tables. They are used for plots. Penalization values, w, are chosen as 0.5 and 1. As mentioned before, the non-negativity of $S(O)^{(1)}_{\phi_1,\phi_2}$ does not hold when $\phi_1 \neq \phi_2$. Thus, only the combinations with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$ with λ = -0.9, -0.8, -0.7, -0.6, -0.5, 0, 2/3, 1 have been considered. The other important reason to chose $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$ is that this combination has performed better in two-way contingency tables. Exact probability estimations given with equations (4.45), (4.46) and (4.47) have been obtained using 1000 simulations from the multinomial distributions with (n, $\mathbf{p_0}$), (n, $\mathbf{p_1}$) and (n, $\mathbf{p_{1,n}}$). Probability vectors $\mathbf{p_0}$ and $\mathbf{p_1}$ are given with Eq.(4.60) and Eq.(4.59), respectively. Approximated asymptotic powers are found as using the same method described in Subsection 4.2.3.1. These are 0.1851, 0.2771 and 0.3663 for the sample sizes of 15, 25 and 35, respectively. Comparison criteria are same as two-way contingency tables case except interested family of test statistics. The simulation results are given in Tables 4.26 - 4.34. Tables 4.26 - 4.28, Tables 4.29 - 4.31 and Tables 4.32 - 4.34 show the exact size, (exact power - asymptotic power) and (exact power - exact size) values of $S(i)^{(1)}_{\phi(\lambda),\phi(\lambda)}$, respectively. Since $I_0(\hat{\mathbf{p}}, \mathbf{p(\theta)}) = P_0^1(\hat{\mathbf{p}}, \mathbf{p(\theta)})$ and $I_1(\hat{\mathbf{p}}, \mathbf{p(\theta)}) = P_1^{0.5}(\hat{\mathbf{p}}, \mathbf{p(\theta)})$, $S(O)^{(1)}_{\phi(0),\phi(0)} = S(P^1)^{(1)}_{\phi(0),\phi(0)}$ and $S(O)^{(1)}_{\phi(1),\phi(1)} = S(P^{0.5})^{(1)}_{\phi(1),\phi(1)}$. Thus, the values in the tables corresponding to these test statistics are equal.

For n = 15, the statistics having exact sizes within the interval [0.0254, 0.0959] (satisfying Eq.(4.52)) correspond to the $S(O)^{(1)}_{\phi(1),\phi(1)}$, $S(O)^{(1)}_{\phi(2/3),\phi(2/3)}$, $S(P^1)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for λ = -0.5, -0.6, -0.7, -0.8, -0.9 ,and $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for all of the values of λ. For negative values of λ, especially large negative values, exact sizes of $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ are far behind being within that interval. However, penalization improves the exact sizes for most the test statistics especially for large negative values of λ. Exact sizes of

$S(P^1)^{(1)}_{\phi_{(1)},\phi_{(1)}}$ and $S(P^1)^{(1)}_{\phi_{(2/3)},\phi_{(2/3)}}$ have increased since the weight put on empty cells in these test statistics increase with penalization value of w = 1. The usual likelihood ratio ($S(O)^{(1)}_{\phi_{(0)},\phi_{(0)}}$) does not satisfy Eq.(4.52), either. However, as seen from Table 4.27, penalization improves this statistic, too. For n = 25, all of the penalized power-divergence statistics and $S(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for λ = 0, 1, 2/3 satisfy Eq.(4.52). This is also the same for n = 35 with addition of two more statistics, $S(O)^{(1)}_{\phi_{(-0.5)},\phi_{(-0.5)}}$ and $S(O)^{(1)}_{\phi_{(-0.6)},\phi_{(-0.6)}}$, satisfying Eq.(4.52).

Table 4.26 Exact sizes of $S(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ under null hypothesis H_{null} : $\mathbf{p_0} \in H_2$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.1090	0.0560	0.0650	0.1520	0.1710	0.2050	0.2750	0.4050
25	0.0690	0.0520	0.0580	0.1110	0.1320	0.1670	0.2320	0.3390
35	0.0490	0.0370	0.0400	0.0710	0.0860	0.1050	0.1430	0.1790

Table 4.27 Exact sizes of $S(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ under null hypothesis H_{null} : $\mathbf{p_0} \in H_2$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.1090	0.1090	0.1070	0.0910	0.0880	0.0870	0.0840	0.0810
25	0.0690	0.0660	0.0670	0.0710	0.0690	0.0650	0.0670	0.0650
35	0.0490	0.0510	0.0500	0.0530	0.0530	0.0530	0.0530	0.0520

Table 4.28 Exact sizes of $S(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ under null hypothesis H_{null} : $\mathbf{p_0} \in H_2$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.0500	0.0560	0.0530	0.0410	0.0400	0.0370	0.0340	0.0340
25	0.0570	0.0520	0.0530	0.0510	0.0500	0.0490	0.0490	0.0470
35	0.0370	0.0370	0.0360	0.0410	0.0420	0.0420	0.0430	0.0430

Table 4.29 (Exact power-Asymptotic power) of $S(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.1489	0.0359	0.0609	0.2349	0.2569	0.3009	0.3349	0.4209
25	-0.0071	-0.0601	-0.0441	0.0509	0.0779	0.1119	0.1519	0.1989
35	-0.0103	-0.0573	-0.0403	-0.0013	-0.0023	-0.0013	-0.0013	-0.0023

Table 4.30 (Exact power-Asymptotic power) of $S(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.1489	0.1559	0.1519	0.0949	0.0919	0.0909	0.0839	0.0799
25	-0.0071	-0.0131	-0.0101	-0.0041	-0.0021	-0.0001	-0.0001	-0.0001
35	-0.0103	-0.0293	-0.0233	-0.0013	-0.0023	-0.0013	-0.0013	-0.0013

Table 4.31 (Exact power-Asymptotic power) of $S(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against $H_{contalt}$: $\mathbf{p_{1,n}}$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	0.0229	0.0359	0.0319	0.0079	0.0029	-0.0001	-0.0001	-0.0001
25	-0.0561	-0.0601	-0.0561	-0.0551	-0.0551	-0.0541	-0.0541	-0.0591
35	-0.0353	-0.0573	-0.0503	-0.0263	-0.0273	-0.0283	-0.0293	-0.0293

Table 4.32 (Exact power-Exact size)$^{-1}$ of $S(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	10.7527	12.0482	11.7647	8.0000	7.5188	6.9930	8.6207	10.0000
25	4.9751	6.0606	5.7143	4.6083	4.4843	4.5045	5.0761	7.2993
35	3.0581	3.6101	3.4364	3.0303	3.0120	3.0581	3.2362	3.4483

Table 4.33 (Exact power-Exact size)$^{-1}$ of $S(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	10.7527	10.6383	10.5263	12.3457	12.8205	12.6582	12.6582	12.5000
25	4.9751	5.0505	5.0000	4.9505	4.8544	4.7170	4.7619	4.7170
35	3.0581	3.4247	3.2787	3.0120	2.9940	3.0120	3.0030	2.9940

Table 4.34 (Exact power-Exact size)$^{-1}$ of $S(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$

	λ							
n	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
15	12.0482	12.0482	12.0482	11.4943	11.4943	11.2360	10.8696	10.8696
25	6.0976	6.0606	5.9524	5.8480	5.8140	5.7471	5.7471	5.8480
35	3.2573	3.6101	3.4722	3.2573	3.2680	3.2895	3.2895	3.2895

Regarding the departures of exact power of the ordinary test statistics from their asymptotic powers, penalization improves them especially for large negative values of λ. But it is not true for $S(P^1)^{(1)}_{\phi_{(1)},\phi_{(1)}}$ and $S(P^1)^{(1)}_{\phi_{(2/3)},\phi_{(2/3)}}$ since weight of empty cells has been increased with w = 1 over their ordinary counterparts. Among all the

test statistics that satisfy Eq.(4.52) for n = 15, the test statistics which satisfy criteria given with Eq.(4.50) are $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for λ = -0.8, -0.9. For n = 25 and 35, $S(P^{1})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ with large negative values of λ perform better among the test statistics satisfying Eq.(4.52).

To illustrate the improvement in the performance of the test statistics $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ for large negative values of λ due to empty cell penalty, the histograms of the exact null distributions of $S(O)^{(1)}_{\phi(-0.8),\phi(-0.8)}$ and $S(P^{0.5})^{(1)}_{\phi(-0.8),\phi(-0.8)}$ have been plotted for n = 15. As in Figure 4.1 and Figure 4.2, the interest is in the right hand tail area behind the dashed line which corresponds to 5% critical point of χ^2_1 which equals to 3.84. Figure 4.3 shows the histogram of $S(O)^{(1)}_{\phi(-0.8),\phi(-0.8)}$, where the poor approximation to the χ^2_1 density is evident from the relatively heavy tail of the statistic. On the other hand, the right tail of the histogram $S(P^{0.5})^{(1)}_{\phi(-0.8),\phi(-0.8)}$ is very well approximated by the χ^2_1 density (i.e. $c = \chi^2_1(0.95) = 3.84$).

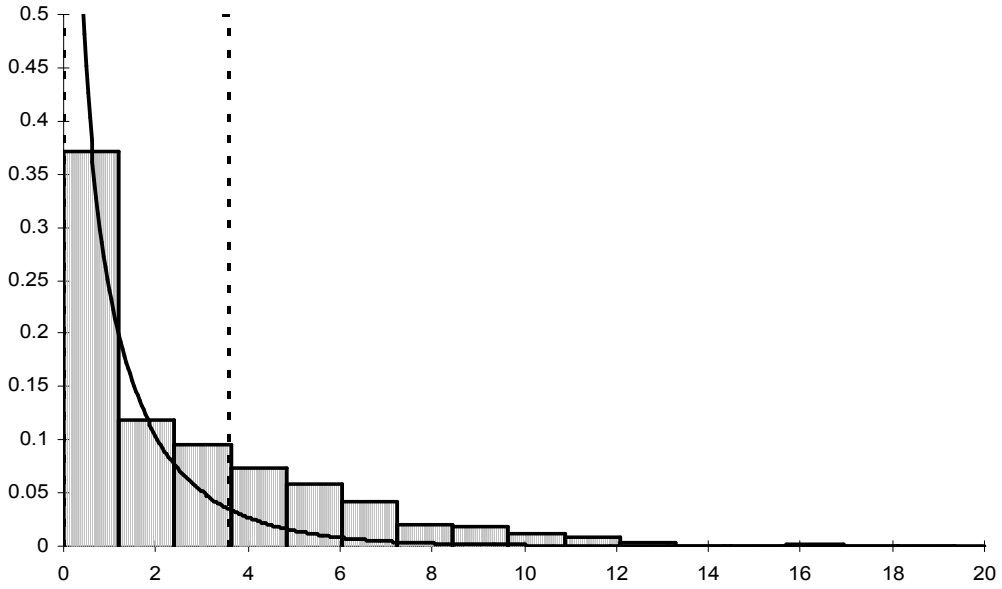

Figure 4.3 Histogram of the null distribution of $S(O)^{(1)}_{\phi(-0.8),\phi(-0.8)}$ with density function of χ^2_1 superimposed.

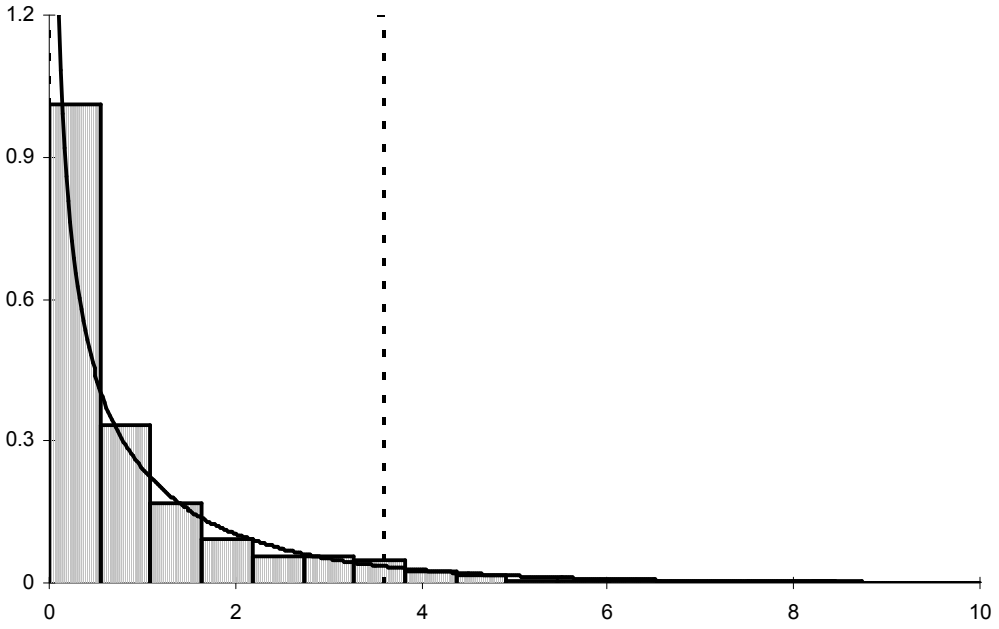

Figure 4.4 Histogram of the null distribution of $S(P^{0.5})^{(1)}_{\phi_{(-0.8)},\phi_{(-0.8)}}$ with density function of χ^2_1 superimposed.

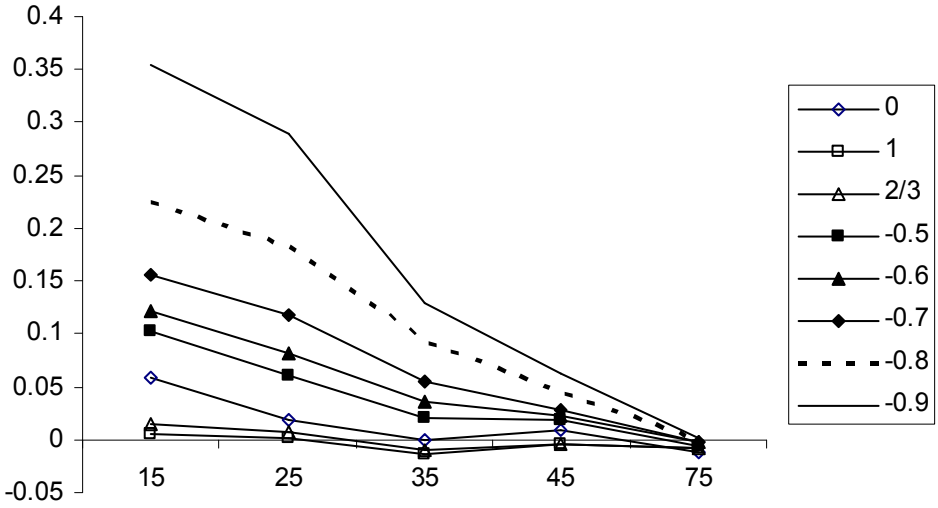

Figure 4.5 (Exact size - Nominal size of 0.05) for $S(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$

Since all of the ordinary power-divergence test statistics are all asymptotically equivalent, ordinary power-divergence test statistics with large negative values of λ get exact sizes very close to nominal size of 0.05 as n gets larger even though they behave entirely opposite for small samples. To illustrate this behavior, the departures of the exact sizes of $S(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ from the nominal size of $\alpha = 0.05$ are plotted

against the sample sizes of 15, 25, 35, 45, and 75. This plot is given with Figure 4.5. As n gets larger, all of the power-divergence test statistics satisfy Eq.(4.52) by getting closer to the nominal size $\alpha = 0.05$.

4.2.4 Examples

Example 1

Melanoma is a disease of the skin in which cancer cells are found in the cells that color the skin. Table 4.35 presents the Melanoma disease data taken from Christensen (1990, p. 343). However, only part of the data was taken to adapt to the 2 x 2 case. This table involves two factors related to the survival of male patients with stages 3 and 4 melanoma: Immunity and Remission.

Table 4.35 Melanoma data (Christensen 1990, p. 343)

	Remission	
Immunity	No	Yes
Yes	5	11
Unknown	0	6

H_{null}: H_2 against H_{alt}: H_1 have been tested on this data set using the $T(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$, $T(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ and $T(P^{1})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$. The models H_1 and H_2 are as follows.

$$H_1: p_{ij}(\theta) = \exp[u + \theta_{1(i)} + \theta_{2(j)} + \theta_{12(ij)}]; \quad i, j = 1, 2$$

$$H_2: p_{ij}(\theta) = \exp[u + \theta_{1(i)} + \theta_{2(j)}]; \quad i, j = 1, 2$$

So, the aim in this hypothesis testing is to decide if there is an interaction between having immunity and remission. Test statistics are given in Table 4.36. As seen from Table 4.36, $T(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ values for negative values of λ are very large compared to the $\lambda = 0, 1, 2/3.$, and they have a great tendency to reject the null hypothesis. As

seen in simulation results given in Subsection 4.2.3.1, this is an expected situation since their exact sizes are much bigger compared to nominal size of 0.05. But, as expected, penalization affected these values and they decreased.

Table 4.36 $T(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$, $T(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ and $T(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ with $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$ values of melanoma data for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$

	λ							
	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
$T(O)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$	3.7075	2.1780	2.5217	5.8424	6.6407	7.7249	9.2910	10.8468
$T(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$	3.7075	3.7754	3.7544	3.6672	3.6586	3.6497	3.6407	3.6314
$T(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$	2.1616	2.1780	2.1728	2.1523	2.1503	2.1483	2.1463	2.1442

The value of $T(O)^{(1)}_{\phi_{(0)},\phi_{(0)}}$ is greater than both of $T(O)^{(1)}_{\phi_{(1)},\phi_{(1)}}$ and $T(O)^{(1)}_{\phi_{(2/3)},\phi_{(2/3)}}$ because $T(O)^{(1)}_{\phi_{(0)},\phi_{(0)}}$ puts more weight on empty cells than both of $T(O)^{(1)}_{\phi_{(1)},\phi_{(1)}}$ and $T(O)^{(1)}_{\phi_{(2/3)},\phi_{(2/3)}}$. According to our simulation results for n = 25, $T(P^{0.5})^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for large negative values of λ performes better. Since n = 22 is very close to n = 25, $T(P^{0.5})^{(1)}_{\phi_{(-0.9)},\phi_{(-0.9)}}$ can be used as a test statistic to decide about if H_{null}: $\mathbf{p_0} \in H_2$ is rejected or not. According to the value of $T(P^{0.5})^{(1)}_{\phi_{(-0.9)},\phi_{(-0.9)}}$, there seems not to exist a relationship between immunity to melanoma disease and remission since $T(P^{0.5})^{(1)}_{\phi_{(-0.9)},\phi_{(-0.9)}} < c = \chi^2_1(0.95) = 3.84$.

Example 2

Table 4.37 presents the data from a study of response to thymosin in bronchogenic carcinoma (or lung cancer) patients. This data set has been taken from Freidlin & Gastwirth (1999). However, only part of the data has been used to adapt to the 2 x 2 x 2 case. There are three factors: Medicine, Response and Group.

Table 4.37 Thymosin data (Freidlin & Gastwirth, 1999)

	Failure		Success	
	Group 1	Group2	Group 1	Group2
Thymosin	0	0	9	8
Placebo	1	3	11	7

The following models have been considered for the thymosin data 4.37.

$$H_1: p_{ijk}(\theta) = \exp[u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)} + \theta_{12(ij)}]; \quad i, j, k = 1, 2$$

$$H_2: p_{ijk}(\theta) = \exp[u + \theta_{1(i)} + \theta_{2(j)} + \theta_{3(k)}]; \quad i, j, k = 1, 2$$

$S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$, $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ and $S(P^{1})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ are used for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$. Rejecting H_{null} means that there is an interaction between given medicine and response to the medicine. Results of test statistics are presented Table 4.38.

Table 4.38 $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$, $S(P^{1})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ and $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$ values of thymosin data for testing H_{null}: $\mathbf{p_0} \in H_2$ against H_{alt}: $\mathbf{p_1} \in H_1$

	λ							
	0	1	2/3	-0.5	-0.6	-0.7	-0.8	-0.9
$S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$	4.9309	3.3569	3.7349	6.6925	7.1800	7.6381	7.8297	7.3288
$S(P^{1})^{(1)}_{\phi(\lambda),\phi(\lambda)}$	4.9309	5.7479	5.5099	4.4074	4.2951	4.1809	4.0653	3.9485
$S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$	2.9224	3.3569	3.2279	2.6565	2.6007	2.5444	2.4877	2.4310

As seen from Table 4.38, $S(O)^{(1)}_{\phi(\lambda),\phi(\lambda)}$ values for negative values of λ are very high compared to the λ = 0, 1, 2/3. However, penalization affected these values and they decreased. The value of $S(O)^{(1)}_{\phi(0),\phi(0)}$ is greater than both of $S(O)^{(1)}_{\phi(1),\phi(1)}$ and $S(O)^{(1)}_{\phi(2/3),\phi(2/3)}$ as in Example 1. Except the values of $S(P^{0.5})^{(1)}_{\phi(\lambda),\phi(\lambda)}$, $S(O)^{(1)}_{\phi(1),\phi(1)}$ and $S(O)^{(1)}_{\phi(2/3),\phi(2/3)}$, all the values are greater than $c = \chi^2_1(0.95) = 3.84$. In simulation

study for 2 x 2 x 2 tables with n = 35, $S(P^1)^{(1)}_{\phi_{(\lambda)},\phi_{(\lambda)}}$ for large negative values of λ have performed better. Since n = 39 is very close to n = 35. $S(P^1)^{(1)}_{\phi_{(-0.9)},\phi_{(-0.9)}}$ can be used as a test statistic. Since the value of this test statistics is bigger than 3.84, H_{null}: $\mathbf{p_0} \in H_2$ is rejected. There seems to exist a relationship between medicine given to the patient and response to the medicine.

CHAPTER FIVE
CONCLUSION AND FURTHER RESEARCH

A statistical model can include the interaction term between variables. However, these models can differ if variables continuous or discrete, or if the design used for research is experimental or observational. Among these statistical models such as multiple linear regression, logistic regression etc., only two-way ANOVA and log-linear models are the focus for this thesis.

In an experimental design called as factorial design, continuous response variable can be modeled by general ANOVA model. If this model includes interaction between two variables, testing that if this interaction term should really be in the model is not a complicated thing when there are replicates. This can be accomplished by the usual two-way ANOVA method. However, the researcher may not be so lucky to have replicates for the experiment, and this makes testing interaction very complicated. In this thesis, some methods for testing interaction (non-additivity) when there is only one observation (no replication) have been studied and illustrated with examples in Chapter 2. Tukey's non-additivity test, Mandel's Row-linear model, Johnson & Graybill's likelihood ratio (LR) test, Boik's locally best invariant (LBI) test and Tusell's likelihood ratio test for sphericity (LRS) are the ones included in Chapter 2. Some of these tests depend on the known interaction structure while some of them do not. These methods are very easy to apply by just using Microsoft® Excel and Statistical package SPSS. Eigen values for LBI, LR and LRS tests can be obtained by mathematical software programs such as Matlab and Mathematica.

Log-linear models describe the association patterns among categorical variables by modeling expected cell counts (or equivalently cell probabilities) in contingency tables without distinguishing between independent (explanatory) and response variables. As in all statistical models, in log-linear models there exist unknown parameters to be estimated. Usually, maximum likelihood estimator (MLE) is used to estimate parameters in log-linear models. In the case of model fitting, the aim is to fit the model as parsimonious as possible. Hence, the researcher should decide if the

model should include the interaction term/terms or not. Likelihood ratio or Pearson's chi-square tests are the mostly used tests statistics for this purpose. Even though MLE and likelihood ratio or Pearson's chi-square test statistics are included in most of the statistical packages, they may not perform well when there are empty cells in contingency tables. There are other estimators and goodness of fit test statistics that may be better.

MLE is included in density based divergence estimators known as minimum power-divergence estimators. Cressie & Pardo (2000, 2002) developed the families of test statistics known as the family of power-divergence test statistics for goodness of fit testing in log-linear models by using minimum power-divergence estimators and power-divergence measures. Likelihood ratio test is included in these families. The problem with power-divergence measures, which also affected the performances of MLE as an estimator or likelihood ratio or other statistics as goodness of fit test statistics, is the weight that these measures put on empty cells in contingency tables. This affection is much more obvious for the values of λ closer to -1. The closer the value of λ to -1, the more weight put on empty cells. However, Basu & Basu (1998) developed empty cell penalty for these measures to cope with this problem. Penalized power-divergence estimators and test statistics have been studied in Chapter 4. Section 4.1 includes the simulation study to compare the efficiency and robustness properties of ordinary and penalized minimum power-divergence estimators. Since the efficiency and robustness properties of these families of estimators in log-linear models for 2 x 2 x 2 tables have already been studied by Pardo & Pardo (2003), these two families of estimators have been compared in log-linear independence model for 2 x 2 tables. The focus has been on small and moderate sample performances of these estimators since they are asymptotically equivalent. Even though the minimum power-divergence estimators with large negative values of λ are the most robust estimators, they perform really worse than other estimators in terms of efficiency. However, penalization improves their efficiency properties. This improvement is more remarkable at small samples. Moreover, their robustness property improves with the empty cell penalty let alone compromised by it. Simulation results also have showed that minimum Pearson's

chi-square estimator is the most efficient estimator among ordinary and penalized minimum power-divergence estimators. However, it is not the most robust estimator. Among all estimators, penalized minimum power-divergence estimator $\hat{\theta}^{P^{0.5}}_{(-0.9)}$ is the most robust one. Ordinary MLE and minimum Pearson's chi-square estimators are well known and mostly used estimators. However, when efficiency and robustness performances of penalized minimum power-divergence estimators with large negative values λ are considered, it would be reasonable to use them with the penalization value of 0.5 for small and moderate sample sizes in two-way contingency tables under the independence model.

Even though likelihood ratio and Pearson's chi-square test statistics are the mostly used test statistics for goodness of fit testing in log-linear models, in the case of empty cells they may not perform well as other power-divergence test statistics. In Section 4.2, the family of penalized power-divergence test statistics has been defined by using penalized power-divergence measure and penalized power-divergence estimators in the family of ordinary power-divergence test statistics. The size and power properties of ordinary and penalized power-divergence test statistics for testing nested log-linear models have been compared by simulation studies. 2 x 2 and 2 x 2 x 2 contingency tables have been considered. The reason to study three-way contingency table case apart from two-way tables is to see how the results change with the dimension of the table. According to the simulation results, it appears that for small and moderate samples in 2 x 2 tables, penalization has improved the size and power properties of ordinary power-divergence test statistics with negative values of λ closer to -1 such that they perform much better than all of the ordinary ones, including the likelihood ratio, Pearson's chi-square test statistics just as expected. Even though the families of ordinary and penalized power-divergence test statistics are asymptotically equivalent, the family of penalized power-divergence test statistics seems better for all sample sizes. Penalization value w = 0.5 is better for small samples, whereas w = 1 is a better choice for moderate samples. Matching ϕ_1 and ϕ_2 is definitely better for all sample sizes and for both of the family of test statistics than to use well known estimators MLE or Pearson's minimum chi-squared

estimator as ϕ_2. So, it would be the judicious decision to use $T(P^{0.5})^{(1)}_{\phi_1,\phi_2}$ for $\phi_1 \equiv \phi_2 \equiv \phi_{(\lambda)}$ with large negative values of λ for all sample sizes in 2 x 2 contingency tables. In 2 x 2 x 2 tables different families of ordinary and penalized power-divergence test statistic have been compared. But as in 2 x 2 tables, penalization improves the size and power properties of the ordinary power-divergence statistics in 2 x 2 x 2 contingency tables especially for large negative values of λ. It also appears that penalized power-divergence test statistics with large negative values of λ perform better than all of the ordinary ones, including the likelihood ratio test statistic. It seems that penalization value of w = 0.5 is better for n = 15, whereas w = 1 is a better choice for n = 25 and n = 35.

The focus was only on independence model in 2 x 2 for comparing ordinary and penalized minimum power-divergence estimators. However, performances of minimum power-divergence estimators can be studied for the tables with bigger dimensions and more models. This also applies to the test statistics. It is only focused on jointly independence and independence model for 2 x 2 x 2 tables. But there are other nested models available as given in Section 3.3. Size and power properties of ordinary and penalized power-divergence test statistics for testing these models can also be studied.

Standard log-linear models do not allow incorporation about ordered categories which can be encountered in lots of applied studies. One simple method of incorporating such information is to specify ordered factor scores assigned to the factor levels that are not truly quantitative. If it is assumed that these factor scores (quantitative levels) are known then there is no problem to consider log-linear models to model the expected cell frequencies (or equivalently cell probabilities). However, as mentioned by Christensen & Utts (1993) " it is often desirable to allow factor scores to be determined by the data instead of being preassigned". This would mean to consider the factor scores as parameters and to estimate them which would make the model no longer log-linear. Tukey (1949) is the first person considering models in which non-linear terms have been added (e.g. the model given with

Eq.(2.5)). As mentioned in Chapter 2, his study is about testing interaction in the two-way ANOVA model with no replication. There have been several extensions of this test to the models with different functions of interactions such as Mandel's row or column linear model. Later Milliken & Graybill (1970) extended these ideas to the general linear model and to the case where interaction is any known function of the block and treatment effects. The model given with Eq.(2.15) considered by Mandel (1971) and Johnson & Graybill (1972) is the model where the interaction may not be a function of the any of the treatment or block effects. All of these procedures can be called as two-stage tests because parameters are first estimated using an additive model, and then the estimates are treated as known constants for the second stage of the test. Christensen & Utts (1992) have presented a simple two-stage test procedure for log-linear and logit models similar to the ones for analysis of variance. They proposed likelihood ratio test as a test statistic. Later Pardo & Pardo (2005) presented the families of test statistics based on power-divergence measures and studied their size and power properties for testing interaction in log-linear models by using two-stage tests procedure as a method of testing interaction. This study, especially for small sample sizes, can be extended to include the family of penalized power-divergence test statistics defined in this thesis.

REFERENCES

Agresti, A. (1990). *Categorical data analysis.* NY: John Wiley & Sons.

Agresti, A. (1996). *An introduction to categorical data analysis.* NY: John Wiley & Sons.

Bartlett, M. S. (1935). Contingency table interactions. *Journal of the Royal Statistical Society*, *Supplement*, *2*, 248-252.

Basu, A., & Basu, S. (1998). Penalized minimum disparity methods for multinomial models. *Statistica Sinica*, *8*, 841-860.

Beran, R. (1977). Minimum Hellinger distance for parametric models. *Annals of Statistics, 5*, 445-463.

Berkson, J. (1980). Minimum chi-square, not maximum likelihood. *Annals of. Statistics*, *3*, 457-487.

Bhapkar, V. P., & Koch, G. G. (1968). Hypothesis of no ınteraction in multidimensional contingency tables. *Technometrıcs*, *10*, 107-123.

Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society*, *Series B (Methodological)*, *25*, 220-233.

Bishop, Y. M. M, Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. Cambride, MA: MIT Press.

Boik, R. J. (1993a). Testing additivity in two-way classifications with no replications: the locally best invariant test. *Journal of Applied Statistics*, *20*, 41-55.

Boik, R. J. (1993b). A comparison of three invariant tests of additivity in two-way classifications with no replications. *Computational Statistics & Data Analysis*, *15*, 411-424.

Christensen, R. (1990). *Log-linear models*. NY: Springer-Verlag.

Christensen, R., & Utts, J. (1992). Testing for nonadditivity in log-linear and logit models. *Journal of Statistical Planning and Inference*, *33*, 333-343.

Christensen, R. (2000). Linear and log-linear models. *Journal of the American Statistical Association*, *95*, *Vignettes*, 1290-1293.

Cressie, N., & Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, *46*, 440-464.

Cressie, N., & Pardo, L. (2000). Minimum ϕ-divergence estimator and hierarchical testing in log-linear models. *Statistica Sinica*, *10*, 867-884.

Cressie, N., & Pardo, L. (2002). Model checking in log-linear models using ϕ-divergences and MLEs. *Journal of Statistical Planning and Inference*,*103*, 437-453.

Cressie, N., Pardo, L. & Pardo, M.C (2003). Size and power considerations for testing loglinear models using ϕ-divergence test statistics. *Statistica Sinica*, *13*, 555-570.

Csiszár, I. (1967). Information type measures of difference of probability distributions and indirect observations. *Studia Scientarum Mathematicarum Hungarica, 2*, 105-113.

Dana, J. L., & Gokhale, D. V. (1983). On the hypothesis of no interaction on a linear scale in contingency tables. *Communications in Statistics: Theory and Methods*, *12*, 1225-1240.

Darroch, J. N. (1962). Interaction in multi-factor contingency tables. *Journal of the Royal Statistical Society*, *Series B*, *24*, 251-63.

Darroch, J. N. (1974). Multiplicative and additive ınteraction in contingency tables. *Bıometrıka*, *61*, 207-214

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Annals of Statistics*, *3*, 1189-1242.

Fenech, A.P., & Westfall P. H. (1988). The power function of conditional log-linear models. *Journal of the American Statistical Association*, *83*, 198-203.

Fienberg, S. E. (1970). The analysis of multidimensional contingency tables. *Ecology*, *51*, 419-433.

Fienberg, S. (1980). *The analysis of cross-classified categorical data* (2nd ed.). MA: MIT Press.

Freake, H. C., Govoni, K. E., Guda, K., Huang, C., & Zinn, S. A. (2001). Actions and interactions of thyroid hormone and zinc status in growing rats. *Journal of Nutrition* ,*131*, 1135-41.

Freidlin, B., & Gastwirth, J. L. (1999). Unconditional Versions of Several Tests Commonly Used in the Analysis of Contingency Tables. *Biometrics*, *55*, 264-267.

Ghosh, J.K.,& Subramanyam, K. (1974). Second order efficiency of maximum likelihood estimators. *Sankyā Series. A*, *36*, 325-358.

Gokhale, D. V. (1972). Analysis of log-linear models. *Journal of the Royal statistical Society*, *Series B*, *34*, 371-376.

Good, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *The Annals of Mathematical Statistics*, *34*, 911-934.

Goodman, L. A. (1963). On methods for comparing contingency tables. *Journal of the Royal statistical Society*, *Series A*, *126*, 94-108.

Goodman, L. A. (1964a). Simple methods for analyzing three-factor interaction in contingency tables. *Journal of the American Statistical Association*, *59*, 319-352.

Goodman, L. A. (1964b). Interactions in multidimensional contingency tables. *The Annals of Mathematical Statistics*, *35*, 632-646.

Goodman, L. A. (1970). The multivariate analysis of qualitative data: interactions among multiple classifications. *Journal of the American Statistical Association*, *65*, 226-56.

Grizzle, J. E., Starmer, F., & Koch, G.G. (1969). Analysis of categorical data by linear models. *Biometrics*, *25*, 489-504.

Haberman, S. J. (1974). *The Analysis of Frequency Data*. Chicago: University of Chicago Press.

Harris, I. R., & Basu, A. (1997). A generalized divergence measure, *Technical report*, *Applied Statistics Unit, Indian Statistical Institute, Calcutta 700 035, India.*

Jazairi, N. T. (1993). *Encyclopedia Of Statistical Science, Vol. 6.*, Canada: John Wıley & Sons.

John, S. (1971). Some optimal multivariate tests. *Biometrika*, *58*, 123-127.

Johnson, D. E. & Graybill, F. A. (1972). An analysis of a two-way model with interaction and no replication. *Journal of American Statistical Association*, *67*, 862-868.

Knoke, D. & Burke, P. J. (1980). *Log-linear models*. CA: SAGE Publications.

Ku, H. H., Varner, R. N., & Kullback, S. (1971). On the analysis of multidimensional contingency tables. *Journal of the American Statistical Association*, *66*, 55-64.

Ku, H. H.,& Kullback, S. (1974). Log-linear models in contingency table analysis. *The American Statistician*, *28*, 115-22.

Lee, H., & Viele, K. (2001). Log-linear models and goodness-of-fit statistics for train waybill data. *Journal of Transportation and Statistics*, *4*, 83-87.

Lindsay, B.G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Annals of Statistics*, *22*, 1081-1114.

Mandel, J. (1961). Non-additivity in two-way analysis of variance. *Journal of American Statistical Association*, *56*, 878-888.

Mandel, J. (1971). A new analysis of variance model for non-additive data. *Technometrics*, *13*, 1-18.

Mandel, J. (1991). *Evaluation and control of measurements*. NY: Marcel Dekker.

Menéndez, M., Morales, D., Pardo, L., & Vajda, I. (2001). Minimum divergence estimators based on grouped data. *Annals of the Institute of Statistical Mathematics*, *53*, 277-288.

Milliken G. A., & Graybill F. A. (1970). Extensions of the general linear hypothesis model. *Journal of American Statistical Association, 65*, 797-807.

Montgomery, D.C. (1991). *Design and analysis of variance* (3rd ed.). NY: John Wıley & Sons.

Morales, D., Pardo, L., & Vajda, I. (1995). Asymptotic divergence measures of estimates of discrete distributions. *Journal of Statistical. Planning and Inference, 48*, 347-369.

Nagarsenker, B.N., & Pillai, K. C. S. (1975) The distribution of sphericity test criterion. *Mimeograph Series No. 284, Dept. Of Statistics, Purdue University. Tables Reproduced in Kres, H*.

Norton, H. W. (1945). Calculation of chi-square from complex contingency tables. *Journal of American Statistical Association*, *40*, 251-258.

Oler, J. (1985). Noncentrality parameters in chi-squared goodness-of-fit analyses with an application to log-linear procedures. *Journal of the American Statistical Association*, *80*, 181-189.

Oztekin, S., Kalkan, S., Ozzeybek, D., Tuncok, Y., Guven, H., & Elar, Z. (2000). The effects of propofol on normal and hypercholesterolemic isolated rabbit heart. *General Pharmocology*, *35*, 65-70.

Pardo, L., Pardo, M.C. & Zografos, K. (2001). Minimum ϕ-divergence estimator for homogeneity in multinomial populations. *The Indian Journal of Statistics*, *63*, 72-92.

Pardo, L., & Pardo, M. D. C. (2003). Minimum power-divergence estimator in three-way contingency tables. *Journal of Statistical Computation Simulation*, *73*, 819-831.

Pardo, L., & Pardo, M. D. C. (2005). Nonadditivity in log-linear models using ϕ-divergences and MLEs. *Journal of Statistical Planning and Inference*, *127*, 237-252.

PDR health. (n.d.) Insulin-like growth factor 1. Retrieved February 21, 2005 from http://www.gettingwell.com/drug_info/nmdrugprofiles/nutsupdrugs.

Plackett, R. L. (1962). A Note on interactions in contingency tables. *Journal of the Royal statistical Society Ser. B*, *24*, 162-66.

Plackett, R. L. (1974). *The analysis of categorical data.* London: Charles Griffin.

Rao, C. R. (1961). Asymptotic efficiency and limiting information. *Procedings of. Fourth Berkeley Symposium. Mathematical Statistics and Probability 1,* 531-546.

Rao, C. R. (1962). Efficient estimates and optimum inference procedures in large samples (with discussion). *Journal of the Royal Statistical Society, Series B*, *24*, 46-72.

Reed, T. R. C., & Cressie, N. A. C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. NY: Springer-Verlag.

Renehan, A. G., Zwahlen, M., Minder, C., O'Dwyer, S. T., Shalet, S.M., & Egger, M. (2004). Insulin-like growth factor (IGF)-I, IGF binding protein 3 and cancer risk: systematic review and meta regression analysis. *The LANCET*; *363*, 1346-53.

Roy, S. N., & Kastenbaum, M. A. (1956). On the hypothesis of no "interaction" in a multi-way contingency tables. *The Annals of Mathematical Statistics*, *27*, 749-757.

Simpson, D.G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, *82*, 802-807.

Snee, R. D. (1982). Non-additivity in a two-way classification: is it interaction or non-homogenous variance? *Journal of American Statistical Association*, *77*, 515-519.

Srivastava, M. S.,& Carter, E.M. (1980). Asymptotic distribution of latent roots and applications. *Multivariate Statistical Analysis*, *In: R. P. Gupta (Ed.)*, 219-236.

Statlib at Carneigie Mellon University. (1989). p-value for LBI test statistics. Retrieved January 14, 2004 from http://lib.stat.cmu.edu/apstat/284.

Sugiura, N. (1972). Locally best invariant test for sphericity and the limiting distribution. *Annals of Mathematical Statistics*, *43*, 1312-1316.

Tamura, R. N., & Boos, D. D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association, 81*, 223-229.

Tansey, R., White, M., Long, R. G., & Smith, M. (1996). A comparison of log-linear modeling and logistic regression in management research. *Journal of Management*, *22*, 339-358.

Tukey, J.W. (1949). One degree of freedom for non-additivity. *Biometrics, 5*, 232-242.

Tusell, F. (1990). Testing for interaction in two-way ANOVA tables with no replication. *Computational Statistics & Data Analysis*, *10***,** 29-45.

Wang, Q., & Zhang, B. (1998). Research design and statistical methods in Chinese medical journals. *The Journal of the American Medical Association*, *280*, 283-285.