# THE ANALYSIS OF MULTIVARIATE, LONGITUDINAL CATEGORICAL DATA BY LOG-MULTILINEAR MODELS

BY

CAROLYN JANE ANDERSON

B.A., University of California, Berkeley, 1982
M.S., University of Illinois at Urbana-Champaign, 1986

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1993

Urbana, Illinois

# THE ANALYSIS OF MULTIVARIATE, LONGITUDINAL CATEGORICAL DATA BY LOG-MULTILINEAR MODELS

Carolyn Jane Anderson, Ph.D.
Department of Psychology
University of Illinois at Urbana-Champaign, 1993
Stanley Wasserman, Advisor

The models developed in this thesis are designed to analyze the nature of the interactions present in 3–mode, multivariate, longitudinal categorical data. They are extensions of loglinear models that provide graphical representations of interactions or associations among discrete variables. The models are generalizations of association models for 2–way tables and are similar to canonical correlation models and their generalizations.

Basic concepts regarding the modeling and analysis of categorical data are reviewed, as well as association and correlation models for 2–way tables. The association and correlation models are inherently limited because they rely on 2–way singular value decompositions. Various strategies for using these models to analyze 3– or higher–way tables and existing generalizations of these models are reviewed. To demonstrate both the usefulness and the limitations of these approaches, they are applied to a multivariate, longitudinal data set from a study in developmental psychology (Kramer & Gottman, 1992).

To overcome the limitations and problems associated with the existing models, a new class of models is developed. This class consists of extensions of loglinear models for 3–way tables in which particular combinations of the 2– and higher–way interactions are decomposed by Tucker's 3–mode principal components model (Tucker, 1963, 1964, 1966;

Kroonenberg, 1983). Tucker's 3–mode model, which is a generalization of 2–way singular value decomposition to 3–way matrices, is quite useful for analyzing longitudinal data.

Typical of longitudinal data, the data analyzed in this study contain missing observations, and the standard sampling assumption of independent, Poisson variables is not valid. Three different strategies for dealing with missing data are employed. The violations of the sampling assumptions invalidate statistical tests and affect estimates of standard errors. Residual analyses are performed to ascertain the degree to which violations of the sampling assumptions affect the results of analyses.

All of the models and techniques reviewed and used to analyze the data in this study are compared theoretically and empirically. Equivalences among them are pointed out. There is a remarkable similarity of results from fitting the various models to the data, but the new models yield the most parsimonious representations.

# Acknowledgments

I would like to thank my committee members for their comments and suggestions both on my proposed study and on this thesis. Ulf Böckenholt and Lawrence Hubert discussed ideas and directed me to relevant literature. Ulf Böckenholt also challenged me to think critically about the models. Laurie Kramer provided the substantive psychological problem that motivated this work and allowed me to use to her data. I am grateful to Jeffrey Tanaka for counseling me to put bounds on this project and reminding me that I should have fun with it.

I would especially like to thank Stanley Wasserman, my advisor, for his continual guidance, encouragement and support throughout my graduate studies, and in particular, for the opportunities and resources to pursue this work. I would also like to acknowledge research support provided by a grant awarded to Dr. Wasserman from the Research Board of the University of Illinois.

This work benefited from the excellent instruction and stimulating intellectual environment provided by the quantitative division of the Department of Psychology at the University of Illinois. I am indebted to the Michael Birnbaum and Elke Weber whose teaching and advise influenced me greatly. I would also like to thank Earle Heffley for computing and programming assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# Chapter 1

# INTRODUCTION AND OVERVIEW

In social and behavioral science research, data are often discrete and consist of repeated measures on the same subjects under different conditions or over time. The development of techniques and models to analyze and represent the relationship (dependency, interaction, or association) among categorical variables has been and continues to be an extremely active area of quantitative research. However, the existing methods and the newer developments are not fully adequate for analyzing multivariate, longitudinal categorical data. The present work was motivated by the desire to adequately analyze one such data set from a study in developmental psychology.

## 1.1 The Type of Data

The specific data set studied here is a 3-mode, cross-classification of the behaviors exhibited by children playing with their best friends. The data are from a study by Kramer (Kramer & Gottman, 1992) that examined the adjustment of firstborn children to the birth of a sibling. The three modes of the "peer play" data are a classification of the children into groups, the type or quality of the behavior, and the time point or occasion when the behavior was observed. The group classification is based on three characteristics of the children that were deemed important. The three binary variables are age (i.e., younger versus older), gender, and the adjustment or acceptance of their sibling (i.e., high versus low), which define $2^3 = 8$ groups. The qualities of play that were recorded were chosen for substantive reasons. The play qualities can be thought of as the "response" variable. The same children were observed on 5 different occasions, 2 before and 3 after the birth of their sibling (except that some of the children were not observed on the third occasion). Missing observations resulted in zeros that are missing at random in the data for the third occasion. Besides repeated measures across occasions, repeated observations were made on each occasion. The study and data are described in more detail in Chapter 2 where preliminary analyses of it are presented.

While the models developed here are designed for the peer play data, they can also be used to analyze other multivariate categorical data. The type of data that can be analyzed by the models and methods presented here consist of observations on either three discrete variables, or more than three discrete variables where the variables are

classified into three different types or modes. The peer play data set is an example of the latter type of data, because the 5 variables, age, gender, adjustment, play qualities and occasions, make up the 3 modes: groups, play, and time. While various models are reviewed and developed in this thesis with the intention of analyzing longitudinal data, they can also be used to analyze data where none of the three modes (or variables) corresponds to time.

## 1.2 Existing Methods

Loglinear modeling is one of the standard and most well-known modeling techniques for analyzing multivariate, categorical data. Loglinear models readily yield information about which interactions are present, but they are not as useful for examining and studying the nature of the interactions. Van der Heijden (1987, p.63) describe the difficulty with using loglinear models to study the nature of the associations between (or among) variables as follows:

> It is not common practice to interpret individual $u$-terms ... there is no underlying empirical scale for a dependent variable. The 'dependent variable' here is the logarithm of a frequency. Furthermore, the number of $u$-terms that must be interpreted often becomes very large, especially when there are higher-order interactions, or when the number of categories is large.

The large number of parameters of loglinear models (i.e., "$u$-terms") is a less troublesome problem than the interpretation of individual $u$-terms. In general, many of the $u$-terms

are "nuisance" parameters (e.g., the constant and "main" effect terms), which are not of interest to the researcher, and thus do not have to be interpreted. Hence, the interpretation problem only applies to the terms representing the highest order and substantively interesting interactions.

Two general classes of models, canonical correlation models (Gilula & Haberman, 1986, 1988; Goodman, 1985, 1986, 1991) and log-multiplicative or "association" models (Goodman, 1985, 1986, 1991; Pannekoek, 1985), are much better than loglinear models at describing the relationships in data. Unlike loglinear models, which do not impose any specific structure on the interactions, correlation and association models hypothesize specific structures for the interactions. In association and correlation models, the interactions or dependencies between variables are decomposed into components or dimensions. The categories of the variables are assigned scale values on each component (in some optimal way), and a measure of the degree to which the variables are related on these dimensions is calculated. Mathematically, interactions are represented in the models by the sum over a fixed number of dimensions of the product of the scale values assigned to categories and parameters that measure the "relatedness" of the variables (i.e., the strength of the interaction). Typically, the products are "bilinear" terms; that is, there are usually only two variables, and for each category of one variable, the terms are a linear function of the scale values assigned to the categories of the other variable (and vice versa). Only a small number of dimensions (2 or 3) is usually needed to adequately represent or approximate an interaction. The estimated scale values are plotted to provide a "picture" of the relationship among the categories of variables.

Association and correlation models have similar characteristics, but are also different in important ways. The models in the two classes are based on different definitions of interaction. The canonical correlation models belong to a more general class of models referred to as "additive" models. In these models, interactions are defined and measured in terms of correlations between variables. This class also includes such techniques as correspondence analysis (Hill, 1974; Greenacre, 1984, 1988; Nishisato, 1980) and latent class models (Goodman, 1987a, 1987b; McCuthcheon, 1987)). The association models belong to a more general class of models referred to as "multiplicative" models. In these models, interactions are defined and measured in terms of odds or odds ratios. This class of models also includes loglinear models (Agresti, 1983, 1990; Bishop, Fienberg and Holland, 1975; Fienberg, 1985; Goodman, 1978) and multidimensional extensions of log-multiplicative models (Goodman, 1985, 1986, 1987a, 1991; Becker, 1989a; Escoufier, 1988; Choulakian, 1988a). Basic concepts and models, such as definitions and measures of interaction, properties of additive and multiplicative models, and the basic association and correlation model for 2–way tables, are reviewed in Chapter 3.

Association and correlation models were originally developed to analyze the relationship between the two variables of a 2–way cross-classification. The models are most readily applicable to 2–way tables, because the decomposition is essentially a singular value decomposition of some 2–way matrix. The mathematical properties of 2–way decompositions are well known (Eckart & Young, 1936; Franc, 1989; Denis & Dhorne, 1989; d'Aubigny & Polit, 1989). The matrices that are decomposed contain information about the dependency or association between the rows and columns of the tables. A matrix to

be decomposed could contain standardized residuals or estimated interaction parameters from a loglinear model (Becker, 1990a; Greenacre, 1984; Gilula & Haberman, 1986, 1988; van der Heijden, de Falugerolles, & de Leeuw, 1989).

A loglinear modeling approach to analyzing categorical data can incorporate more than two variables into a model by simply adding more terms. Extending correlation and association models to three or more variables is much more complex than this simple addition. With a 2–way table, there is only one possible interaction, and therefore, only one interaction to study. The number of possible interactions present in a table increases with the dimensionality of the table. For example, a 3-way table has four possible interactions: three 2–way interactions and one 3-way interaction. In the case of multiway tables, it becomes necessary to identify which of the possible interactions are "important" and which of the interactions are "interesting." Important interactions are those that are present in the data. To adequately model the data, terms representing the important interactions need to be included. Interesting interactions, which may or may not be important, are those that researchers want to examine and study. Typically, interesting interactions that are important are those that are decomposed. Whether an interaction is interesting depends on substantive considerations, and whether an interaction is important depends on the data.

Various strategies and generalizations have been proposed to extend the association and correlation models for 2–way tables to 3– and higher–way arrays. These proposals either reduce the data to essentially a 2–way problem (reduction approach) or explicitly introduce more terms into the model (model generalization approach). In the reduction

approach, where the task is reduced to a 2–way problem, multiway tables are rearranged into 2–way arrays, which can then be analyzed by standard correlation and/or association models for 2–way tables. Information regarding the structure within the rows and/or columns can be incorporated into the analysis by placing linear or order restrictions on the scale values assigned to rows and/or columns. The reduction approach is discussed in detail in Chapter 4, where limitations of this approach for analyzing higher–way tables are noted.

The model generalization approach to 3– and higher-way tables, which relies on adding more terms to the models, focuses on generalizing the singular value decomposition aspect of the models. The models in one sub-class of this approach still rely on 2–way decompositions. In these models, either more bilinear terms are included to represent additional 2–way interactions or bilinear terms are used to represent the combined effect of 2– and higher–way interactions. The models in another sub-class of the "model generalization approach" utilize 3–way decompositions and represent higher–way interactions by trilinear terms, which are the products of three scale values and a measure of association. The mathematical properties of 3–way decompositions are more complex and are less well known than the properties of the singular value decomposition of 2–way arrays (Coppi & Bolasco, 1989). The two sub-classes of model generalizations and various 3–way decompositions are reviewed and discussed in Chapter 5. The drawbacks of these techniques for the peer play data are identified.

The models that include extra bilinear and/or trilinear terms are more general than those in the "reduction approach". When only bilinear terms are included in the model

generalizations, many of the models in the reduction approach are special cases of some generalized decomposition model. When the generalized models contain 3–way decompositions (i.e., trilinear terms), there is a clear distinction between the approaches. This distinction between the reduction methods and the model generalizations highlights the limitations of the reduction approach for representing and analyzing 3–factor associations. There are qualitative and interpretative similarities and differences between the models from these two approaches. By distinguishing between the models in the reduction and model generalization approaches, the qualitative and interpretative differences among the models are more easily understood, and the advantages of each of them are identified. The best aspects of the various models are incorporated into the models designed for the peer play data.

## 1.3 New Models

Additive and multiplicative models for analyzing categorical data were developed relatively independently of each other. Only within the last ten years has the relationship between the models from the two approaches been a topic of research and the complementary nature of the models considered (Goodman, 1985, 1986, 1991; van der Heijden & de Leeuw, 1985, 1989; van der Heijden, 1987; van der Heijden & Worsley, 1988; van der Heijden, de Falugerolles & de Leeuw, 1989). For example, van der Heijden (1987) and van der Heijden and de Leeuw (1989) proposed analyzing longitudinal categorical data by using correspondence analysis to study standardized residuals from loglinear models

fit to the data. For this procedure to yield useful information, the loglinear models must fail to adequately fit the data and the residuals must contain the substantively interesting interactions. Rather than analyzing residuals from some model, the approach taken here is to design models that fit the data *and* provide "pictures" of the interesting interactions.

In Chapter 6, a new family of models is developed that incorporates Tucker's 3-mode components model (Tucker, 1963, 1964, 1966; Kroonenberg, 1983, 1984) into a multiplicative model. These models are extensions of loglinear models for 3–way tables and are generalizations of the association model for 2–way tables. Multiplicative models were chosen rather than additive models, because the general properties of multiplicative models are more desirable than those of additive models for analyzing count data (in particular, the peer play data). Tucker's 3–mode components model was chosen as the 3–way decomposition for a number of reasons. The primary reason for this choice is that the mathematical structure of Tucker's model makes it is very useful for analyzing longitudinal data (e.g., Tucker, 1963; Kroonenberg, 1983; Kroonenberg, Lammers & Stoop, 1985; Iacobucci, 1989). Change is represented in terms of underlying components for time or occasions. The other 2 modes of the data, groups (or individuals) and variables, are also represented by components, and the relationship among the 3-modes of the data is modeled by the components for the 3-modes. Complex interactions can be represented in relatively simple forms with Tucker's 3-mode model.

Since data from longitudinal studies often contain missing values, as does the peer play data, as well as zeros missing at random (or structural zeros), three strategies for dealing with missing data are described and illustrated. Two of them involve deleting

parts of the data, while the third uses all of the data available. In the third method, certain parameters are not estimated and the fitted values for the cells that contain structural zeros are set equal to zero. For the peer play data, the different methods of dealing with missing data result in only minor differences in the graphical representations and interpretations of the associations in the data.

The primary intent in this thesis is to demonstrate that the models based on the mathematical structure proposed in Chapter 6 are useful for analyzing multivariate, longitudinal data, and that they yield the most parsimonious representations of the peer play data. For simplicity, the observations are assumed to be independent, Poisson random variables, and except for two cases, maximum likelihood estimation is used to fit models to the peer play data in Chapters 2 — 6. While the observations on the children are probably not independent, violations of the independence assumption primarily effect variance estimates (Agresti, 1990; McCullagh & Nelder, 1989; Cox, 1983); therefore, estimates of variances and standard errors are not computed. Likelihood ratio statistics (i.e., $G^2$) are not asymptotically distributed as chi-squared random variables; however, $G^2$ can still be used as an index of fit, which is the approach taken by van der Heijden (1987) in the loglinear modeling stage of his correspondence analysis of longitudinal data. The likelihood ratio statistic and the ratio of $G^2$ to the degrees of freedom ($df$) are used in this thesis as indices of fit. In addition, the residuals from models that fit well in terms of $G^2/df$ are analyzed to determine whether the models fail to fit the data in some systematic way and to examine whether violations of the sampling assumptions are a serious problem. The residual analyses are used to ascertain whether the sampling

assumptions of independence and Poisson distributions are an adequate approximation to the truth.

The methods for analyzing multivariate categorical data that are reviewed and presented throughout this thesis and the proposed model generalizations are all applied to the peer play data. In the final chapter, Chapter 7, the results from the various analyses are compared. The comparison reveals that it is advantageous to include a 3–way decomposition to represent a 3-factor association, and that two of the new models proposed in Chapter 6 yield the most parsimonious representations (descriptions) of the data.

# Chapter 2

# THE DATA AND PRELIMINARY ANALYSES

The peer play data set is an example of a 3-mode, cross-classification of count or frequency data in which one of the modes is time. The analysis of these data and others like them is the primary motivation for the review and development of the models and techniques in Chapters 3 - 6. In Section 2.1, pertinent details about the study, collection, and pre-processing of the peer play data are described. In Section 2.2, data analytic problems associated with this data set (and others like it) and the approaches taken here to deal with them are discussed. In Section 2.3, preliminary analyses of the data using loglinear models are presented. These analyses provide useful information, but they fail to give readily interpretable descriptions of the nature of the associations and interactions in the data.

## 2.1 The Peer Play Study

The peer play data are from part of a larger study by Laurie Kramer (1992) on the development of sibling relationships. The study examined the adaptation and coping of firstborn children to the birth of a second child. One of the premises of the study is that children use their relationships with their friends to help them cope with the change in their family's structure and the stress of becoming a sibling. The peer play data was collected (in part) to test this conjecture.

In Section 2.1.1, the collection and pre-processing of the data are described and in Section 2.1.2, more specific research questions to be answered.

### 2.1.1 The Data

The subjects were thirty firstborn, three to five year old children whose families were expecting a second child. An audio-recording was made of each child playing with their best friend. The children played in a room by themselves for 30 minutes on five different occasions. The taped play constitutes the "raw" data, which was coded into the 3-mode, cross-classification of counts analyzed here. The three types or modes of the peer play data are described in the following three sections, where the details of the coding and pre-processing of the raw data are briefly described.

#### 2.1.1.1 Mode 1: Groups

For the peer play portion of the study, three attributes of the children were deemed important for evaluating the role of children's friendships in making the transition to becoming a sibling. Obviously, a measure of the acceptance or adjustment of the first-born to the second born child is needed. Sibling acceptance was measured on the fifth occasion, and the children with scores less than the median were classified as exhibiting "low" sibling acceptance and those with score greater than the median were classified as exhibiting "high" sibling acceptance.

In addition to sibling acceptance, the age and gender of a child could influence the children's play. Due to normal development and change, the behavior of a three year old can be quite different from that of a five year old child. Children younger than the median age, which equals 45.5 months or 3.79 years, were classified as "younger", and those older than the median were classified as "older". With respect to gender, the play of males and females tend to differ even at young ages, and the differences tend to become greater as the children age (i.e., the play of 3 year old boys and girls are more alike than that of 5 year old boys and girls). There are 18 females and 12 males.

While gender is inherently a nominal, binary variable, the other two attributes (age and sibling acceptance) are continuous variables that are coded as binary variables. The three variables thus define $2^3 = 8$ types or groups of children. For convenience, the groups are denoted by three letters where the first letter indicates gender, the second, age, and the third, adjustment. For example, the group classification FYL consists of

the younger females who exhibit low adjustment, and MOH consists of the older males who exhibit high adjustment. The groups are composed of different numbers of children. The groups FOH and MYL each contain 6 children, FOL contains 5, FYH contains 4, FYL and MOH each contain 3, MYH contains 2, and MOL only contains 1 child.

#### 2.1.1.2 Mode 2: Play Qualities

Each 30 minute play session was divided into three 10 minute blocks, and each 10 minute block was coded separately. The first step in coding the children's play was to mark which of 44 different behavior "codes" occurred during a block. The codes consist of both positive and negative types of play behaviors, communication, and affect. On the basis of substantive considerations, the 44 codes were classified by the researcher into 11 mutually exclusive and exhaustive categories or play qualities. The 11 qualities are

1. sustained communication (sus),
2. coordinated or successful gossip (gos),
3. coordinated or positive play (pos),
4. excitement (exc),
5. amity (amt),
6. shared or successful fantasy (fan),
7. unsustained communication (uns),
8. uncoordinated or poor play (poor),

9. negative emotions (nem),

10. conflict (fight), and

11. prohibitions (prh).

For example, the play category "negative emotions" consists of the three behavior codes "anger", "sadness" and "whining". As another example, the play category "excitement" consists of the four codes "joy", "laughter", "humor" and "positive teasing". The play categories consist of different numbers of codes. More detailed descriptions of the coding of the data, the 44 behavior codes and the 11 qualities can be found in Kramer and Gottman (1992).

Given the coding of the raw data, an "event" is a behavior type or code, which is an example of a play quality. Each "event" is counted only once. For each child, on each block, the number of the codes observed that are classified as a particular quality equals the frequency of that play quality for that child on that block. The frequencies of the qualities are summed over blocks and children within groups to obtain the frequencies of play qualities for each group on each occasion. Thus, the play mode consists of eleven categories, each of which corresponds to a quality of play.

#### 2.1.1.3 Mode 3: Occasions

The play sessions occurred over an 8 month period, at approximately two month intervals. The first session was 3 months prior to the expected date of the second child's birth, and the last session was 5 to 6 months after the birth (i.e., observations were made at 3 and

1 month before and at 1, 3 and 5 months after the birth). Except for five children who were absent from the third session (i.e., 1 month after the birth), all of the subjects were observed playing with their best friend on 5 occasions.

Table 2.1 contains the frequencies of the play codes exhibited by the 30 subjects cross-classified by occasions, groups and play qualities. The asterisks, "$*$", next to the groups FYL, FOL, FOH, MYH and MOL on the third occasion (i.e., +1 mth) indicate that a child was absent from the group on that session. Since there is only one child in the MOL group and this child was absent on the third session, Table 2.1 contains 11 zeros missing at random, which are denoted in the table by ".".

### 2.1.2 Expectations and Research Questions

The primary purpose of analyzing the peer play data is to investigate the relationship between peer play and the development of the firstborn's relationship with their new sibling. Assuming that the children use their friendships to help them cope and adapt to becoming a sibling, the quality and type of play they exhibit with their friends should be related to the measured sibling acceptance, as well as age and possibly gender. More specifically, the data are analyzed to answer the following questions:

1. Does the quality of play before the second child's birth predict acceptance by the firstborn child of the second child?

2. How does play change over time? What is stable and what changes?

**Table 2.1**: Frequencies of the play codes cross-classified by groups, play qualities and occasions ($N = 30$).

| Time | Group | sus | gos | pos | exc | amt | fan | uns | poor | nem | fight | prh |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FYL | 27 | 9 | 6 | 5 | 2 | 5 | 36 | 4 | 4 | 4 | 8 |
| | FYH | 45 | 17 | 6 | 12 | 18 | 11 | 51 | 12 | 21 | 23 | 11 |
| | FOL | 59 | 14 | 13 | 21 | 14 | 7 | 50 | 14 | 8 | 13 | 11 |
| −3 mth | FOH | 69 | 20 | 18 | 22 | 20 | 14 | 48 | 8 | 6 | 20 | 14 |
| | MYL | 51 | 13 | 10 | 14 | 18 | 9 | 77 | 19 | 13 | 17 | 14 |
| | MYH | 21 | 5 | 6 | 7 | 6 | 6 | 21 | 4 | 6 | 13 | 3 |
| | MOL | 10 | 0 | 3 | 5 | 2 | 0 | 8 | 5 | 6 | 10 | 3 |
| | MOH | 33 | 6 | 14 | 15 | 7 | 6 | 29 | 1 | 1 | 2 | 8 |
| | FYL | 30 | 7 | 8 | 10 | 2 | 7 | 24 | 2 | 2 | 7 | 5 |
| | FYH | 47 | 10 | 8 | 5 | 14 | 7 | 36 | 10 | 9 | 13 | 10 |
| | FOL | 56 | 11 | 16 | 17 | 15 | 9 | 31 | 11 | 3 | 10 | 10 |
| −1 mth | FOH | 63 | 18 | 17 | 19 | 18 | 10 | 44 | 4 | 4 | 8 | 13 |
| | MYL | 59 | 15 | 19 | 14 | 18 | 7 | 58 | 16 | 17 | 23 | 15 |
| | MYH | 25 | 5 | 3 | 4 | 7 | 5 | 24 | 5 | 6 | 8 | 5 |
| | MOL | 11 | 0 | 0 | 0 | 3 | 0 | 7 | 6 | 6 | 13 | 3 |
| | MOH | 34 | 8 | 9 | 5 | 6 | 5 | 31 | 6 | 0 | 5 | 7 |
| | FYL∗ | 19 | 5 | 8 | 5 | 8 | 0 | 23 | 4 | 9 | 13 | 6 |
| | FYH | 42 | 15 | 17 | 18 | 18 | 4 | 28 | 15 | 7 | 12 | 10 |
| | FOL∗ | 43 | 8 | 10 | 15 | 15 | 8 | 40 | 9 | 4 | 7 | 11 |
| +1 mth | FOH∗ | 60 | 16 | 16 | 17 | 17 | 8 | 41 | 18 | 5 | 14 | 11 |
| | MYL | 72 | 11 | 14 | 34 | 17 | 12 | 53 | 8 | 10 | 21 | 14 |
| | MYH∗ | 12 | 5 | 0 | 5 | 5 | 2 | 12 | 4 | 2 | 1 | 3 |
| | MOL∗ | . | . | . | . | . | . | . | . | . | . | . |
| | MOH | 32 | 8 | 9 | 10 | 12 | 7 | 14 | 4 | 3 | 3 | 7 |
| | FYL | 29 | 9 | 10 | 19 | 11 | 6 | 24 | 8 | 3 | 8 | 5 |
| | FYH | 46 | 10 | 11 | 10 | 18 | 8 | 39 | 11 | 9 | 13 | 10 |
| | FOL | 59 | 23 | 12 | 24 | 25 | 11 | 43 | 14 | 5 | 13 | 15 |
| +3 mth | FOH | 65 | 20 | 13 | 26 | 21 | 14 | 57 | 8 | 4 | 12 | 13 |
| | MYL | 61 | 9 | 14 | 14 | 8 | 11 | 55 | 13 | 7 | 24 | 16 |
| | MYH | 23 | 7 | 5 | 8 | 9 | 3 | 25 | 2 | 6 | 6 | 4 |
| | MOL | 9 | 3 | 2 | 2 | 6 | 1 | 10 | 5 | 2 | 4 | 3 |
| | MOH | 34 | 4 | 8 | 1 | 10 | 4 | 20 | 6 | 0 | 11 | 6 |
| | FYL | 29 | 5 | 10 | 9 | 5 | 7 | 29 | 12 | 5 | 4 | 8 |
| | FYH | 51 | 15 | 11 | 20 | 18 | 9 | 34 | 7 | 7 | 7 | 9 |
| | FOL | 55 | 15 | 12 | 12 | 22 | 10 | 61 | 13 | 8 | 19 | 11 |
| +5 mth | FOH | 63 | 17 | 15 | 20 | 20 | 12 | 36 | 11 | 2 | 11 | 10 |
| | MYL | 64 | 16 | 14 | 24 | 14 | 11 | 61 | 17 | 8 | 16 | 17 |
| | MYH | 26 | 6 | 7 | 6 | 8 | 5 | 31 | 4 | 3 | 10 | 5 |
| | MOL | 12 | 2 | 2 | 1 | 5 | 0 | 13 | 3 | 4 | 8 | 3 |
| | MOH | 29 | 4 | 3 | 8 | 5 | 5 | 20 | 1 | 1 | 3 | 4 |

3. Are some types of play disrupted after the second child's birth? If so, for which children?

4. Are there differences between the groups with respect to play qualities on different occasions? Do the qualities exhibited by children differ over time in some systematic way according to the children's age, gender, and/or sibling acceptance?

The first question can be answered by studying the data from the first occasion ($-3$ months) and the second occasion ($-1$ month). Standard analyses of partial association, which are presented later in this chapter, can be used to answer this question; however, these analyses do not yield descriptions of the relationship between play qualities and sibling adjustment. If there is a relationship, then we want to know which qualities tend to be exhibited by those who show high and low sibling acceptance. In Chapter 3, association and correlation models for 2-way tables are used to describe the relationship for the first occasion.

To answer the remaining questions, the data from all five occasions need to be analyzed. The children are observed over time to investigate whether their play changes in some systematic way. Before the birth of the second child, the children are anticipating a change in their family structure, and after the birth, they experience change in their family life. One expectation regarding change is that the play will be the most disrupted immediately after the birth (i.e., $+1$ month). If play is disrupted, then we will want to identify which play qualities are most affected for which children. Since all of the children have all undergone the same "treatment" (i.e., become a sibling), another expectation is

that the children will be more similar to each other at the end of the study than they were at the beginning.

The most substantively interesting and important question is how the children differ with respect to play qualities over time. The same children were observed over time to see whether the children with the higher sibling adjustment scores tend to exhibit different qualities in their play at certain time points and/or throughout the period under study than those with lower sibling adjustment scores. We expect the older children to show different play qualities and to be better able to adapt to the changes in their lives than the younger children. Gender may also be important. A representation and description of the pattern of play qualities exhibited by the children classified into the different groups over time is needed to confirm these expectations and to study the relationship between peer play and sibling acceptance. In sum, we need to simultaneously represent the children, play qualities, occasions, and the relationship among these three aspects of the data.

## 2.2 Data Analytic Problems

Two major problems must be dealt with in the analyses of the peer play data: missing data and violations of sampling assumptions. The specific problems and concerns associated with each of them and the approaches taken to deal with them are discussed in Sections 2.2.1 and 2.2.2.

### 2.2.1 Missing Data

Three approaches are used to deal with missing data. The most "wasteful" approach is to use the data from just the 25 children who were present for all 5 sessions (i.e., $N = 25$). The data from the 25 child sub-sample are given in Table 2.2. The second method uses more of the available data without having to deal with the zeros missing at random. For the second method, the data analyzed include all of the data except the data from the only child in the MOL classification (i.e., $N = 29$, the $(7 \times 11 \times 5)$ subtable of Table 2.1 that does not include MOL). The third method uses all of the data available (i.e., $N = 30$). In the third method, the eleven missing observations for the third time period are treated as structural zeros and are fit perfectly by the models.

The majority of the analyses presented in this thesis were performed on the data from all three samples (i.e., the 25 and 29 child sub-samples and the full 30 child sample). The results of standard analyses of association and loglinear modeling of each of these three versions of the data are presented Section 2.3. In Chapter 3, only the data from the first occasion are analyzed. Since there are no missing observations on the first occasion, the frequencies are from all 30 children (i.e., the sub-table of Table 2.1 consisting of the first 8 rows).

For both practical and theoretical reasons, only the results from the analyses of the 25 child sub-sample are presented and discussed in detail for all of the techniques described and illustrated in Chapters 4 – 6. There is one practical reason for concentrating on a sub-sample of the data. Many different models can be fit to the data. The models

**Table 2.2**: Frequencies of the play codes cross-classified by groups, play qualities and occasions ($N = 25$).

| Time | Group | sus | gos | pos | exc | amt | fan | uns | poor | nem | fight | prh |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FYL | 15 | 7 | 4 | 3 | 1 | 4 | 21 | 2 | 3 | 4 | 5 |
| | FYH | 45 | 17 | 6 | 12 | 18 | 11 | 51 | 12 | 21 | 23 | 11 |
| | FOL | 49 | 10 | 11 | 17 | 10 | 5 | 37 | 8 | 7 | 8 | 8 |
| −3 mth | FOH | 58 | 16 | 16 | 20 | 16 | 12 | 43 | 6 | 6 | 19 | 11 |
| | MYL | 51 | 13 | 10 | 14 | 18 | 9 | 77 | 19 | 13 | 17 | 14 |
| | MYH | 11 | 3 | 4 | 6 | 5 | 3 | 9 | 2 | 6 | 11 | 3 |
| | MOH | 33 | 6 | 14 | 15 | 7 | 6 | 29 | 1 | 1 | 2 | 8 |
| | FYL | 20 | 3 | 6 | 5 | 1 | 6 | 16 | 0 | 1 | 5 | 3 |
| | FYH | 47 | 10 | 8 | 5 | 14 | 7 | 36 | 10 | 9 | 13 | 10 |
| | FOL | 44 | 7 | 13 | 13 | 14 | 7 | 26 | 9 | 3 | 8 | 8 |
| −1 mth | FOH | 52 | 13 | 15 | 16 | 12 | 7 | 37 | 4 | 4 | 6 | 10 |
| | MYL | 59 | 15 | 19 | 14 | 18 | 7 | 58 | 16 | 17 | 23 | 15 |
| | MYH | 12 | 0 | 0 | 2 | 4 | 3 | 14 | 2 | 6 | 8 | 3 |
| | MOH | 34 | 8 | 9 | 5 | 6 | 5 | 31 | 6 | 0 | 5 | 7 |
| | FYL | 19 | 5 | 8 | 5 | 8 | 0 | 23 | 4 | 9 | 13 | 6 |
| | FYH | 42 | 15 | 17 | 18 | 18 | 4 | 28 | 15 | 7 | 12 | 10 |
| | FOL | 43 | 8 | 10 | 15 | 15 | 8 | 40 | 9 | 4 | 7 | 11 |
| +1 mth | FOH | 60 | 16 | 16 | 17 | 17 | 8 | 41 | 18 | 5 | 14 | 11 |
| | MYL | 72 | 11 | 14 | 34 | 17 | 12 | 53 | 8 | 10 | 21 | 14 |
| | MYH | 12 | 5 | 0 | 5 | 5 | 2 | 12 | 4 | 2 | 1 | 3 |
| | MOH | 32 | 8 | 9 | 10 | 12 | 7 | 14 | 4 | 3 | 3 | 7 |
| | FYL | 20 | 5 | 6 | 14 | 8 | 3 | 13 | 3 | 3 | 6 | 3 |
| | FYH | 46 | 10 | 11 | 10 | 18 | 8 | 39 | 11 | 9 | 13 | 10 |
| | FOL | 47 | 19 | 11 | 20 | 21 | 9 | 34 | 7 | 3 | 8 | 12 |
| +3 mth | FOH | 55 | 18 | 9 | 22 | 17 | 14 | 43 | 7 | 4 | 9 | 12 |
| | MYL | 61 | 9 | 14 | 14 | 8 | 11 | 55 | 13 | 7 | 24 | 16 |
| | MYH | 11 | 5 | 2 | 7 | 2 | 2 | 13 | 2 | 1 | 2 | 2 |
| | MOH | 34 | 4 | 8 | 1 | 10 | 4 | 20 | 6 | 0 | 11 | 6 |
| | FYL | 21 | 3 | 6 | 5 | 5 | 5 | 20 | 6 | 5 | 3 | 6 |
| | FYH | 51 | 15 | 11 | 20 | 18 | 9 | 34 | 7 | 7 | 7 | 9 |
| | FOL | 44 | 12 | 9 | 8 | 16 | 9 | 48 | 10 | 8 | 13 | 8 |
| +5 mth | FOH | 55 | 16 | 14 | 18 | 17 | 10 | 34 | 9 | 2 | 9 | 10 |
| | MYL | 64 | 16 | 14 | 24 | 14 | 11 | 61 | 17 | 8 | 16 | 17 |
| | MYH | 13 | 3 | 4 | 5 | 5 | 3 | 13 | 2 | 3 | 10 | 3 |
| | MOH | 29 | 4 | 3 | 8 | 5 | 5 | 20 | 1 | 1 | 3 | 4 |

can be compared and contrasted in terms of goodness of fit, parsimony, substantive interpretations, and so forth. To make valid and "fair" comparisons, the same data should be used for all such studies. Since no observations are available for MOL at the third session, the full peer play data set ($N = 30$) contains the zeros missing at random. Unfortunately, one of the pre-existing programs that was used to fit some of the models presented in Chapter 4 cannot handle the such zeros. Therefore, the group MOL was dropped from the analyses presented in Chapters 4 – 6, which leaves the data from a sub-sample of 29 children.

A substantive or theoretical reason for using the 25 child sub-sample instead of the 29 child sub-sample to compare the various models is that the most dramatic changes and/or disruptions in the peer play are expected to occur on the third occasion, the first session after the birth. Since all of the missing data in the 29 child sub-sample are from the third session, differences between the third session and the other sessions could reflect changes due to the arrival of the second child and/or group composition. If most of the changes do indeed occur at this session, then including data from the children who were absent from this session would confound changes due to the birth of the second child and differences due to group composition. Including the data from the children with missing observations could mask or smooth out the data so that actual changes are less detectable. For these reasons, the 25 child sub-sample was chosen for the purpose of comparing and contrasting all of the different models.

While the results from the analyses of the 25 child sub-sample are presented and discussed in Chapters 4 – 6, most of the analyses were also performed on the 29 and 30

child samples. Differences due to group composition turn out to be minimal, as suggested by the results of the analyses presented later in this chapter. The topic of missing data is discussed further in Chapter 6 where the effect of missing observations on the results of the analyses using the new models are examined and the new models are fit to the data containing the zeros that are missing at random.

### 2.2.2 Violation of Sampling Assumptions

The second problem is more elusive and harder to solve satisfactorily. To fit models to the data using maximum likelihood estimation, the frequencies are assumed to be independent, Poisson random variables. The children within the groups are assumed to have a common distribution of frequencies of the eleven play qualities for each occasion (i.e., all of the children within a group classification have the same Poisson parameter for each play quality). The validity of the sampling assumptions of independence, a Poisson distribution, and homogeneity of children within groups is questionable for the peer play data, as well as for other longitudinal data sets.

Since each child contributes many responses during each session and the same children are observed at five different times, the validity of the assumption of independent observations is in doubt. Dependency among observations typically leads to over-dispersion, where the observed variance is greater than the variance predicted by some model (Agresti, 1990; McCullagh & Nelder, 1989). Over-dispersion can also be caused by heterogeneity of individuals classified into the same groups.

The dependencies in the data and heterogeneity of children within groups invalidate statistical tests. While the likelihood ratio statistic $G^2$ and Pearson's chi-squared statistic $X^2$ should not be compared to chi-squared distributions, they can still be used as measures or indices of fit and to compare the fit between models (van der Heijden and de Leeuw, 1985, 1987; van der Heijden and Worsley, 1988).

Dependency usually has minimal effects on the actual values of estimated model parameters (McCullagh & Nelder, 1989; and others), but is more likely to affect estimates of standard errors of model parameters. Since the direction of the effect on the standard errors in not predictable (under or over estimate), standard errors are not computed here to avoid the "temptation" of interpreting them.

If the sampling assumptions are grossly or even moderately violated, then the residuals should exhibit particular structures. Extensive residual analyses are preformed in Chapter 6 to look for evidence that the sampling assumptions are violated. Analyses are performed to look for over-dispersion, correlations between residuals, and relationships between estimated means and variances that are expected for Poisson variables. The results of the residuals analyses indicate that for the peer play data, the sampling assumptions of independence, Poisson distribution and homogeneity among children within groups are reasonable approximations to the actual sampling process underlying the data.

## 2.3 Preliminary Analyses

Standard analyses of the peer play data are presented in this section to provide an initial impression of the complexity of the associations in the data. These analyses provide useful information about which interactions are present in the data and the relative strength of these interactions; however, they do not provide good descriptions of the nature of the relationships.

Three different analyses are presented in Sections 2.3.1 – 2.3.3. In Section 2.3.1, analyses of partial and marginal association of the data on the first occasion are presented to show that each of the classification variables (i.e., gender, age and sibling acceptance) are related to play qualities and that the 3–way interaction among them (i.e., the group classification variable) should be included in further analyses. In Section 2.3.2, analyses of the partial and marginal association of the data from all 5 occasions are presented, and in Section 2.3.3, the results of loglinear modeling of the data are presented.

### 2.3.1 Analysis of Association: First Occasion

The data from the first occasion is analyzed here to determine whether any of the classification variables are not important and can be disregarded in further analyses. The results of these analyses also indicate whether sibling acceptance, which was measured at the end of the study, and play qualities exhibited 3 months prior to the birth of the second child are related.

The data from the first occasion consist of the $(2 \times 2 \times 2 \times 11)$ cross-classification of frequencies according gender (G), age (A), sibling acceptance (S) and play qualities (P). Statistics for assessing the presence of partial and marginal associations in this 4–way table are reported in Table 2.3. The data analyzed here consist of all of the available data from the first occasion (i.e., $N = 30$). The $\Delta G^2$'s, which are reported in the third column of the table, indicate whether there is partial association between (or among) the variables listed in the first column. The $\Delta G^2$'s are differences between likelihood ratio statistics of nested loglinear models fit to the 4–way table where the only difference between the models is that one includes the effect and the other does not. In the sixth column, likelihood ratio statistics for the marginal associations are presented. Since $\Delta G^2$ and $G^2$ should not be compared to chi-squared distributions, the ratio of the statistic to its corresponding degrees of freedom are computed as an index of fit. Evidence for the presence of an association is indicated by larger ratios.

The statistics for the partial and marginal associations are very similar. The highest order interaction(s) with large $\Delta G^2/df$ and $G^2/df$'s are those that are of interest. The lower order effects are not particularly interesting and do not need to be interpreted; however, the lower order effects should be included in any additional models fit to the data.

The $\Delta G^2$ for the most complex partial association, $PGAS$, is relatively large compared to the degrees of freedom, which suggests that there is a 4–way interaction between the child classification variables and the play qualities on the first occasion. Since sibling acceptance is one of the classification variables, there is a relationship between sibling

**Table 2.3**: Partial and marginal associations in the 4–way cross-classification of frequencies from the first session ($N = 30$).

| Effect | Partial Association | | | Marginal Association | | |
|---|---|---|---|---|---|---|
| | $df$ | $\Delta G^2$ | $\Delta G^2/df$ | $df$ | $G^2$ | $G^2/df$ |
| Play | 10 | 612.20 | 61.22 | | | |
| Gender | 1 | 64.25 | 64.25 | | | |
| Age | 1 | 0.81 | .81 | | | |
| Sibling | 1 | 3.14 | 3.14 | | | |
| $PG$ | 10 | 8.62 | .86 | 10 | 7.86 | .79 |
| $PA$ | 10 | 33.20 | 3.32 | 10 | 32.94 | 3.29 |
| $PS$ | 10 | 15.73 | 1.57 | 10 | 18.09 | 1.81 |
| $GA$ | 1 | 79.72 | 79.72 | 1 | 87.33 | 87.33 |
| $GS$ | 1 | 29.40 | 29.40 | 1 | 39.61 | 39.61 |
| $AS$ | 1 | 5.30 | 5.30 | 1 | 16.02 | 16.02 |
| $PGA$ | 10 | 6.14 | .61 | 10 | 4.41 | .44 |
| $PGS$ | 10 | 26.10 | 2.61 | 10 | 20.08 | 2.01 |
| $PAS$ | 10 | 38.26 | 3.83 | 10 | 22.58 | 2.26 |
| $GAS$ | 1 | 91.30 | 91.30 | 1 | 96.16 | 96.16 |
| $PGAS$ | 10 | 14.47 | 1.45 | | | |

adjustment and peer play, which answers one of the questions listed in Section 2.1.2. However, this relationship also depends on the age and gender of the child.

The $GAS$ interaction needs to be "controlled" for and taken into account in any further analyses for two reasons. Besides the fact that there is a 4–way interaction, which includes $GAS$, the $GAS$ margin corresponds to the cross-classification of the children into groups. This margin is strongly affected by the number of children who fall into each group (which is not substantively interesting, especially given the relatively small sample of children in this study). Since there is a 4–way association and the $GAS$ margin is affected by the sampling design of the study, in all of the following analyses, gender, age and sibling acceptance are treated as a single variable, "groups", with 8 categories.

The analysis in this section is extremely limited. It only reveals that there is a relationship between peer play and the child classification variables on the first occasion. It does not describe the nature of the relationship. Furthermore, it only tells us about just one of the five time points. Of substantive interest is how patterns of play behavior of the first born child change with the birth of a second child.

### 2.3.2 Analysis of Association: All 5 Occasions

The analyses presented in this section incorporate the data from all five sessions. The three child attribute variables are treated as a single variable. Thus, the data analyzed in this section consist of the 3–way, Group $\times$ Play $\times$ Time cross-classification of counts. The frequencies from the 25, 29 and 30 child samples are all analyzed here to demonstrate

that the results are basically the same regardless of the specific sample, but that there may be subtle differences.

Tables 2.4, 2.5 and 2.6 contain statistics for assessing the presence of partial and marginal associations between/among groups, play qualities, and occasions in the 3–way tables for 25, 29 and 30 samples, respectively. Differences between likelihood ratio statistics of nested loglinear models fit to the 3–way tables are given in the third column of each of the tables. The $\Delta G^2$'s indicates whether there is a partial association between/among the variables listed in the first column. The likelihood ratio statistics, $G^2$, in the sixth column pertain to marginal associations.

The basic pattern of associations is the same in each the three data sets, and is basically the same for the partial and marginal associations. The indices for the $GPT$ partial association (i.e. for $N = 25, 29$ and 30, $\Delta G^2/df = 1.16$, 1.10 and 1.10, respectively) are large enough to suggest that there is a 3–way association in the data, but the indices are also small enough to throw some doubt on this conclusion, especially since the sampling assumptions used to compute the $G^2$'s may not be valid. It may be the case that the strength of the $GPT$ association is relatively small and the sample size is not large enough to definitively indicate whether there is a 3–way association in the "population". Given this uncertainty, the conclusion that the highest way interaction is present is tentative.

While the indices for the $GPT$ association suggest that there may be a 3–way association, the indices for the $GP$ and $GT$ partial and marginal associations are large enough to indicate that these 2–way interactions are present in the data. The indices for the $PT$ partial and marginal associations are small enough to safely draw the conclusion that

**Table 2.4**: Partial and marginal associations in the 3–way cross-classification ($N = 25$).

| Effect | Partial Association | | | Marginal Association | | |
|---|---|---|---|---|---|---|
| | *df* | $\Delta G^2$ | $G^2/df$ | *df* | $G^2$ | $G^2/df$ |
| Play | 10 | 2456.46 | 245.65 | | | |
| Time | 4 | 10.14 | 2.54 | | | |
| Group | 6 | 1110.24 | 185.04 | | | |
| $PT$ | 40 | 42.13 | 1.05 | 40 | 41.22 | 1.03 |
| $GP$ | 60 | 126.90 | 2.12 | 60 | 126.02 | 2.10 |
| $GT$ | 24 | 37.88 | 1.58 | 24 | 37.00 | 1.54 |
| $GPT$ | 240 | 278.07 | 1.16 | | | |

**Table 2.5**: Partial and marginal associations in the 3–way cross-classification ($N = 29$).

| Effect | Partial Association | | | Marginal Association | | |
|---|---|---|---|---|---|---|
| | *df* | $\Delta G^2$ | $G^2/df$ | *df* | $G^2$ | $G^2/df$ |
| Play | 10 | 2818.87 | 281.89 | | | |
| Time | 4 | 19.03 | 4.76 | | | |
| Group | 6 | 816.36 | 136.06 | | | |
| $PT$ | 40 | 38.79 | .97 | 40 | 37.98 | .95 |
| $GP$ | 60 | 127.91 | 2.13 | 60 | 127.08 | 2.12 |
| $GT$ | 24 | 62.30 | 2.60 | 24 | 61.49 | 2.56 |
| $GPT$ | 240 | 264.10 | 1.10 | | | |

**Table 2.6**: Partial and marginal associations in the 3–way cross-classification ($N = 30$).

| Effect | Partial Association | | | Marginal Association | | |
|---|---|---|---|---|---|---|
| | *df* | $\Delta G^2$ | $G^2/df$ | *df* | $G^2$ | $G^2/df$ |
| Play | 10 | 2862.77 | 286.28 | | | |
| Time | 4 | 18.92 | 4.73 | | | |
| Group | 7 | 1270.41 | 181.49 | | | |
| $PT$ | 40 | 40.90 | 1.02 | 40 | 39.82 | 1.00 |
| $GP$ | 70 | 196.46 | 2.81 | 70 | 195.40 | 2.79 |
| $GT$ | 27 | 62.95 | 2.33 | 27 | 62.05 | 2.30 |
| $GPT$ | 270 | 297.82 | 1.10 | | | |

neither the $PT$ marginal nor partial association exists in the data. The $GP$ association definitely warrants further study, while the $GT$ association is not particularly interesting. The $GT$ margin depends on the number of children in each group on each occasion. This margin reflects the different number of children in the groups and for $N = 29$ and 30, the missing observations from the third occasion.

In sum, the analyses discussed here indicate that there is definitely a $GP$ interaction and there is probably a 3–way association among groups, play qualities and occasions; however, the analyses do not provide descriptions and representations of the nature of this relationship.

### 2.3.3 Loglinear Modeling

The analyses in the previous section were based on loglinear models. Rather than examining differences of $G^2$'s, the loglinear models for the 3–way tables will now be examined to provide further insight into the associations in the data and the limited usefulness of loglinear models for describing the interactions. The fit statistics for loglinear models fit to the 3–way, Group × Play × Time tables for $N = 25$, 29 and 30 are reported in Tables 2.7, 2.8 and 2.9, respectively.

Like the previous analyses, the pattern of results is basically the same for $N = 25$, 29 and 30. The main difference between them, which can also be seen in the analyses in the last section, is that for $N = 29$ and 30, the models fit slightly better. This result is the opposite of what is usually observed when more data is included. When more observations are included, models tend to fit worse, because of increased power to reject

the null hypothesis. With the peer play data, including the children with missing data may have smoothed out some of the effects, particularly if there are differences between the third occasion and the others. If this conjecture is true, then the data from the 25 child sub-sample should be analyzed.

If, as concluded in the previous section, there is a 3–way association is correct, then a saturated loglinear model is needed for the peer play data. With the saturated model, there is no reduction or summary of the data into a smaller set of numbers, which can be easily examined to "sort out" the nature associations. There are as many 3–way interaction terms as there are cells in the table.

If there is no 3–way association, then there should be a non-saturated model that fits the data. The two best, non-saturated loglinear models are the $(GP, GT)$ and $(GT, GP, PT)$ models. Even with these two models, characterizing the nature of the interactions by examining the estimated interaction terms is difficult. For example, with the simpler model, $(GP, GT)$, there are two sets of interaction parameters, one set for each of the 2–way interactions. Each set consists of a relatively large number of terms. The $GT$ interaction is not substantively interesting and the $u$-terms for it do not need be examined. But for the 25 and 29 child sub-samples, there are 77 $u$-terms for the $GP$ interaction, and for the 30 child sample, there are 88 parameters. Meaningful interpretations of the $u$-terms for the $GP$ interaction is difficult.

In sum, analyses based on loglinear models indicate which interactions are present in the data. For the peer play data, there are definitely $GP$ and $GT$ associations. There is probably a 3–way, $GPT$ interaction. While the change in play over time is substantively

**Table 2.7**: Loglinear models fit to the 3–way table of frequencies ($N = 25$).

| Model | | *df* | $G^2$ | $G^2/df$ | $X^2$ |
|---|---|---|---|---|---|
| Complete Independence | $(G, P, T)$ | 364 | 483.27 | 1.33 | 467.25 |
| Joint Independence | $(GT, P)$ | 340 | 446.24 | 1.31 | 422.96 |
| | $(G, PT)$ | 324 | 441.96 | 1.36 | 421.20 |
| | $(GP, T)$ | 304 | 357.25 | 1.18 | 341.79 |
| Conditional Independence | $(GT, PT)$ | 300 | 404.97 | 1.35 | 377.89 |
| | $(GP, GT)$ | 280 | 320.20 | 1.14 | 302.56 |
| | $(GP, PT)$ | 264 | 315.95 | 1.20 | 299.76 |
| No 3-way Interaction | $(GT, GP, PT)$ | 240 | 278.07 | 1.16 | 260.60 |

**Table 2.8**: Loglinear models fit to the 3–way table of frequencies ($N = 29$).

| Model | | *df* | $G^2$ | $G^2/df$ | $X^2$ |
|---|---|---|---|---|---|
| Complete Independence | $(G, P, T)$ | 364 | 491.61 | 1.35 | 469.64 |
| Joint Independence | $(GT, P)$ | 340 | 430.07 | 1.26 | 407.68 |
| | $(G, PT)$ | 324 | 453.50 | 1.40 | 421.20 |
| | $(GP, T)$ | 304 | 364.45 | 1.20 | 354.80 |
| Conditional Independence | $(GT, PT)$ | 300 | 392.01 | 1.31 | 366.66 |
| | $(GP, GT)$ | 280 | 302.90 | 1.08 | 294.21 |
| | $(GP, PT)$ | 264 | 326.40 | 1.24 | 314.28 |
| No 3-way Interaction | $(GT, GP, PT)$ | 240 | 264.10 | 1.10 | 254.44 |

**Table 2.9**: Loglinear models fit to the 3–way table of frequencies ($N = 30$).

| Model | | *df* | $G^2$ | $G^2/df$ | $X^2$ |
|---|---|---|---|---|---|
| Complete Independence | $(G, P, T)$ | 407 | 596.25 | 1.46 | 576.00 |
| Joint Independence | $(GT, P)$ | 380 | 534.14 | 1.41 | 511.87 |
| | $(G, PT)$ | 367 | 556.32 | 1.52 | 528.21 |
| | $(GP, T)$ | 337 | 400.84 | 1.19 | 386.61 |
| Conditional Independence | $(GT, PT)$ | 340 | 494.27 | 1.45 | 466.33 |
| | $(GP, GT)$ | 310 | 338.72 | 1.09 | 325.80 |
| | $(GP, PT)$ | 297 | 360.77 | 1.21 | 343.77 |
| No 3-way Interaction | $(GT, GP, PT)$ | 270 | 297.82 | 1.10 | 284.35 |

interesting, there does not appear to be a partial or marginal $PT$ association. However, if there is a 3–way association, then we should be examining the $GPT$ interaction to see how play changes over time for the different groups.

The $GT$ association is present in the data, but it is not substantively interesting. Based on considerations regarding the sampling and processing of the raw data, the $GT$ margin should be fit perfectly by any model used to analyze the peer play data. The $GT$ margin is affected by the missing data and the number of children in each group. Furthermore, since we are interested in comparing the relative frequencies of play qualities between groups and at the same and different time points, the number of frequencies per occasion per group should be "controlled for" by treating the total number per group per occasion as fixed.

With loglinear models it is difficult to see what the interactions "look" like. The loglinear models and their fits provide a reference against which to evaluate the models discussed in the following chapters. In the next chapter, we first review basic concepts regarding the measurement and modeling of interactions in categorical data, and then present simple models for 2–way tables that provide graphical representations of interactions.

# Chapter 3

# BASIC CONCEPTS AND MODELS

Additive and multiplicative models for categorical data are reviewed in this chapter. Different definitions of interaction (or "no interaction") are associated with the models in each of these classes. The differences are reflected in the general forms of the models and the basic properties associated with them. As an example, consider a 3–way contingency table with three categorical variables $A$, $B$ and $C$. Suppose that there is "no 3–way interaction," but that there are 2–way interactions between all pairs of variables. Let $P_{ijk}$ be the probability of observing an object or individual at level $i$ on variable $A$, level $j$ on $B$, and level $k$ on $C$. The no 3–way interaction model using an additive definition is

$$ {}_aP_{ijk} = P_{i..}P_{.j.}P_{..k} + P_{i..}\Delta^{BC}_{jk} + P_{.j.}\Delta^{AC}_{ik} + P_{..k}\Delta^{AB}_{ij} \tag{3.1} $$

where $P_{ij.} = \sum_k P_{ijk}$, $P_{i..} = \sum_{jk} P_{ijk}$, and $\Delta_{ij}^{AB} = P_{ij.} - P_{i..}P_{.j.}$ (Goodman, 1987a; Darroch, 1974). The first term on the right-hand side of this equation, $P_{i..}P_{.j.}P_{..k}$, is the prediction under the independence model. The terms $\Delta_{jk}^{BC}$, $\Delta_{ik}^{AC}$, and $\Delta_{ij}^{AB}$ represent the 2–way "interactions" or associations between pairs of variables. These interactions are multiplied (weighted) by the marginal probabilities $P_{i..}$, $P_{.j.}$, and $P_{..k}$, respectively. The $\Delta$'s measure the departures from independence in the 2–way marginal tables. Thus, the additive model states that the cell probabilities are a linear combination of the associations (departures from independence) between the rows and columns of the 2–way marginal tables.

The no 3–way interaction model using a multiplicative definition is

$$ {}_mP_{ijk} = \tau\tau_i^A\tau_j^B\tau_k^C\tau_{ij}^{AB}\tau_{ik}^{AC}\tau_{jk}^{BC} \tag{3.2} $$

where $\tau$ is a constant, $\tau_i^A$, $\tau_j^B$, and $\tau_k^C$ represent the main effects, $\tau_{ij}^{AB}$, $\tau_{ik}^{AC}$, and $\tau_{jk}^{BC}$ represent the 2–way interactions between pairs of variables. Since the logarithms of the cell probabilities are a linear function of log–scaled effects, the model is frequently written as

$$ \ln(P_{ijk}) = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} \tag{3.3} $$

which corresponds to the standard, no 3-factor interaction loglinear model for a 3–way table (Agresti, 1990; Bishop, Fienberg, & Holland, 1975; Fienberg, 1980; Wickens, 1989), which is [AB][AC][BC] in Fienberg's (1985) notation or $(AB, AC, BC)$ in Agresti's nota-

tion (Agresti, 1984, 1990). Model 3.3 and simpler versions of it were fit to the play data in the previous chapter.

In section 3.1, basic concepts about measures of association and properties of additive and multiplicative models are reviewed. In section 3.2, multidimensional row–column association and correlation models for 2–way tables are presented, fit to the Group × Play sub-table for the first session, and compared both theoretically and empirically. In the final section, 3.3, the material presented in this chapter is briefly summarized. Applying this material to the play data leads to a preference for odds ratios and log-models (multiplicative models and generalizations of them) for further analyses of this data.

## 3.1 Basic Concepts

Additive models have the general form

probability = (prediction under complete independence) + (interaction)

while multiplicative models have the general form

probability = (prediction under complete independence)(interaction).

Since modeling and measurement are integrally related, the definitions of interaction or dependency among variables and how these ideas are measured in additive and multiplicative models are examined in section 3.1.1. Multiplicative and additive models use different definitions of dependency or interaction. In section 3.1.2, the differences in

the general properties of additive and multiplicative models for contingency tables are reviewed.

### 3.1.1 Measures of Association

Following Goodman (1991), two basic measures of dependency (interaction, nonindependence, association) between discrete variables are presented: odds ratios and correlations. Odds ratios and the correlation coefficient are discussed in Sections 3.1.1.1 and 3.1.1.2, respectively. Other measures that are functions or extensions of these two measures are also described. The measures presented here, which are just a subset of the many possible indices that have been proposed for categorical variables, are relevant and useful for the problems considered in this thesis (for others, see Goodman & Kruskal, 1979; Krippendorff, 1986; Read & Cressie, 1988; Wickens, 1989; Agresti, 1990).

Since additive and multiplicative models differ in terms of how interaction is defined, the models differ in terms of how they measure the strength of the dependency between discrete variables. In Section 3.1.1.3, the relationship between odds ratios and multiplicative and additive models is described. Each type of model has associated with it a natural, global index that measures the total dependency in a table. In Section 3.1.1.4, global indices of interaction are reviewed.

#### 3.1.1.1 Odds Ratios

In a 2–way, $(I \times J)$ table, given that an observation is in row $i$, the odds that it is in column $j$ versus $j'$ equals

$$\theta_{jj'(i)} = \frac{P_{ij}/P_{i.}}{P_{ij'}/P_{i.}} = \frac{P_{ij}}{P_{ij'}}$$

The odds $\theta_{ii'(j)}$ are similarly defined. If the row and column variables are completely independent, then $P_{ij} = P_{i.}P_{.j}$, which implies that odds equal the ratio of their respective marginal probabilities (i.e., $\theta_{ii'(j)} = P_{i.}/P_{i'.}$ and $\theta_{jj'(i)} = P_{.j}/P_{.j'}$). Independence also implies that the ratio of odds in any $2 \times 2$ sub-table, often called the "cross product ratio," equals one; that is,

$$\Theta_{ii',jj'} = \frac{\theta_{jj'(i)}}{\theta_{jj'(i')}} = \frac{\theta_{ii'(j)}}{\theta_{ii'(j')}} = \frac{P_{ij}P_{i'j'}}{P_{i'j}P_{ij'}} = 1$$

Neither estimates of odds nor odds ratios, which are computed from empirical frequencies or probabilities, depend on the marginal probabilities or on the number of observations in a table.

Odds ratios and functions of them can also be used to measure associations in higher–way tables. For instance, if there is no 3–way association in a 3–way table, then the ratio of odds ratios of each of the $2 \times 2 \times 2$ sub-tables equals 1; that is,

$$\Theta_{ii',jj',kk'} = \frac{\Theta_{ii',jj'(k)}}{\Theta_{ii',jj'(k')}} = \frac{P_{ijk}P_{i'j'k}P_{i'jk'}P_{ij'k'}}{P_{i'jk}P_{ij'k}P_{ijk'}P_{i'j'k'}} = 1$$

where $\Theta_{ii',jj'(k)}$ and $\Theta_{ii',jj'(k')}$ are odds ratios of $2 \times 2$ sub-tables of variables $A$ and $B$ for levels $k$ and $k'$, respectively, of variable $C$. If complete independence holds in a 3–way table, then all of the odd ratios of $2 \times 2$ sub-tables for constant levels of the third variable will equal one (i.e., $\Theta_{ii',jj'(k)} = \Theta_{ii',kk'(j)} = \Theta_{jj',kk'(i)} = 1$ for all $i$, $i'$, $j$, $j'$, $k$, and $k'$).

Odds ratios take on values from 0 to $\infty$. The departure of an odds ratio from 1 indicates the presence of an association. The more that an odds ratio deviates from 1, the stronger the association. The strength of an association for $\Theta = x$ is the same as that for $\Theta = 1/x$; that is, odds ratios are multiplicatively symmetric around 1. Rather than using odds ratios, logarithms of odds ratios (log–odds ratios) are frequently used to measure association. Logarithms of odds ratios, which range from $-\infty$ to $\infty$, are additively symmetric around the value 0, which corresponds to "no association." The strength of an association for $\ln \Theta = x$ is the same as that for $\ln \Theta = -x$. Since neither odds ratios nor log–odds ratios depend on the total number of observations in a table (or sub-table) or on marginal probabilities, these measures are ideal for comparing the strength of associations across tables and/or sub-tables.

Other measures of association in contingency tables have been proposed that are functions of odds ratios (Agresti, 1990; Fienberg, 1985; Goodman, 1979; Wickens, 1989). Often these measures are transformations of odds ratios that restrict the range of the measure to the interval $-1$ to 1, such as Yule's coefficient of association, which equals $Q = (\Theta-1)/(\Theta+1)$, and Yule's coefficient of colligation, which equals $Y = (\sqrt{\Theta}-1)/(\sqrt{\Theta}+1)$. While these and other measures, which are functions of odds ratios, have slightly different

interpretations and uses, they all measure the same basic thing: associations in tables using odds ratios.

#### 3.1.1.2 Correlation Coefficient

The second basic measure of dependency is the correlation coefficient. In a $2 \times 2$ table, the correlation between the rows and columns is

$$\rho = \frac{(P_{11}P_{22} - P_{12}P_{21})}{(P_{1.}P_{2.}P_{.1}P_{.2})^{1/2}}$$

(Bishop, Fienberg, and Holland, 1975; Goodman and Kruskal, 1979; Goodman, 1991). This quantity is the product moment (Pearson) correlation between the row and column variables where the scores $-1$ and $+1$ have been assigned to the two rows and to the two columns. If the row and column variables are independent (i.e., $P_{ij} = P_{i.}P_{.j}$), then $\rho = 0$. If there is complete dependence, $P_{12} = P_{21} = 0$ (or $P_{11} = P_{22} = 0$), then $\rho = +1$ (or $-1$).

A straightforward generalization of $\rho$ to an $(I \times J)$ table is to compute the correlation between the row and column variables where scores have been assigned to the each of the $I$ rows and the $J$ columns in some way. This generalization is related to Pearson's $X^2$ for testing independence (Bishop, Fienberg & Holland, 1975; Goodman, 1991). Pearson's $X^2$ is discussed in more detail later in this section. The relationship between the correlation $\rho$ and $X^2$ is discussed later in this chapter in connection to the $RC(M)$ correlation model.

Correlations possess two properties that are theoretically undesirable for measuring interactions or dependencies in the play quality data. Unlike odds ratios, correlations

depend on the margins of the table, as well as the strength of the interaction. Since there are different numbers of children in each group and at the third time period some of the children were absent, the marginal distributions reflect these differences. When analyzing the play data, we want to be able to compare the associations across time between play qualities and groups, and compare associations across groups between play and time. If correlations are used to measure interactions, then the comparisons reflect differences in the marginal distributions as well as the strength of associations. Odds and odds ratios are preferable to correlations, because they measure the dependency, independent of the margins. The second undesirable property of correlations is that they do not readily generalize to 3– and higher–way tables. Correlations are computed between *pairs* of variables. Odds ratios are not limited to measuring 2–way associations and can be used to measure the association *among* three or more variables.

#### 3.1.1.3 Multiplicative and Additive Models

Some researchers prefer multiplicative models over additive models because multiplicative models are readily interpretable in terms of odds ratios (Clogg, 1986,1989; Goodman, 1987; and others). Odds ratios, which are multiplicative functions of cell probabilities, can be expressed as a product of the parameters of multiplicative models, or equivalently, log–odds ratios, which are linear functions of the logarithms of cell probabilities, can be expressed as a linear function of the parameters of loglinear models. For example, the

$(AB, AC)$ loglinear model, which is

$$\ln P_{ijk} = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC}$$

implies that the odds ratios in all of the possible $2 \times 2 \times 2$ sub-tables equal one (i.e., $\Theta_{ii',jj',kk'} = 1$), and that the odds ratios in all $2 \times 2$ sub-tables of variables $B \times C$ for constant levels of $A$ equal one (i.e., $\Theta_{jj',kk'(i)} = 1$). The odds ratios and log–odds ratios of the $2 \times 2$ sub-tables of variables $A$ and $B$ for constant levels of $C$ are simple functions of the interaction parameters as follows:

$$\ln \Theta_{ii',jj'(k)} = u_{ij}^{AB} + u_{i'j'}^{AB} - u_{i'j}^{AB} - u_{ij'}^{AB}$$

Note that this log odds ratio does not depend on $k$. The odds and log-odds ratios of the $2 \times 2$ sub-tables of $A$ and $C$ for constant levels of $B$ are similarly represented.

Multiplicative (loglinear) models represent the associations in contingency tables as measured by odds ratios computed on sub-tables of the entire table. Extensions of these models decompose odds ratios into various effects, such as an overall association, row and/or column effects, and row–column interactions (Goodman, 1979, 1985). Multiplicative models define "interaction" and "no interaction" in terms of odds ratios computed on $2 \times 2$ sub-tables. An "interaction" means that particular odds ratios do not all equal one (or $\ln \Theta \neq 0$), and "no interaction" means that particular odds ratios equal one (or $\ln \Theta = 0$).

Unlike in multiplicative models, the odds ratios of sub-tables are not simple functions of the interaction parameters of additive models. The statement that the interaction parameters representing the interaction between two variables equals zero (e.g., $\Delta_{ij}^{AB} = 0$) is not equivalent to the statement that all of the odds ratios of the $2 \times 2$ sub-tables of the two variables for each level of the third variable equal one (e.g., $\Theta_{ii',jj'(k)} = 1$). Additive models define interactions in terms of the associations in the 2–way *marginal* tables of higher–way contingency tables. For example, using the additive definition of interaction, whether there is an "interaction" between variables $A$ and $B$ in the 3–way table depends on whether or not variables $A$ and $B$ are independent in the $AB$ marginal table (i.e., $P_{ij.} = P_{i..}P_{.j.}$), which implies that $\Delta_{ij}^{AB} = 0$ and the odds ratios in the marginal tables equal one. Even the relationship between the additive model's interaction parameters and the odds ratios in the marginal table is relatively complex,

$$\Theta_{ii',jj'} = \frac{(\Delta_{ij}^{AB} + P_{i..}P_{.j.})(\Delta_{i'j'}^{AB} + P_{i'..}P_{.j'.})}{(\Delta_{i'j}^{AB} + P_{i'..}P_{.j.})(\Delta_{ij'}^{AB} + P_{i..}P_{.j'.})}$$

which involves marginal probabilities, $P_{i..}$ and $P_{.j.}$, as well as, the interaction parameters.

#### 3.1.1.4 Global Indices of Association

Global indices of the total dependency in a table can be computed using both multiplicative or additive models. One possibility is to compute a function of the estimated interaction parameters of saturated models. Goodman (1991) uses this approach and defines global indices as the root mean square of the interaction parameters. A slightly

different approach is taken here. Rather than using saturated models and interaction parameters, "no interaction" models and residuals from these models are used to compute global measures. The indices presented here are two common statistics, the likelihood-ratio statistic $G^2$ and Pearson's $X^2$ statistic for testing independence, which can also be used to test other models.

Both $G^2$ and $X^2$ are sums of standardized residuals, but they differ with respect to how the residuals are computed and standardized. The residuals used in computing $G^2$ equal the difference between the logarithms of the observed and predicted probabilities. These residuals are standardized by multiplying them by the observed probabilities. The formula for the likelihood–ratio statistic $G^2$ is

$$G^2 = 2N \sum_{cells} P \ln(P/\hat{P})$$

where $N$ is the total number of observations in the table, $P$ is an observed probability, and $\hat{P}$ is the predicted probability from some model. The residuals used in computing $X^2$ equal the difference between observed and predicted probabilities and are standardized by dividing them by the square root of their predicted value. Pearson's $X^2$ equals the sum of the squares of these standardized residuals; that is,

$$X^2 = N \sum_{cells} \frac{(P - \hat{P})^2}{\hat{P}}$$

When the predictions are from the model of complete independence, $G^2$ and $X^2$ measure how much dependency is present in a table. If there is no association, then the predictions of both the additive and multiplicative models of complete independence will equal the observed probabilities and $G^2 = X^2 = 0$. If there is some relationship between the rows and columns, then $G^2$ and $X^2$ will be greater than zero; however, both $G^2$ and $X^2$ also depend on the total number of observations in a table. By dividing $G^2$ and $X^2$ by $N$, the resulting quantities measure just the association in a table and can be thought of as "distances" from independence (or from some other model). The quantity $G^2/2N$ is the Kullback–Leibler measure (Gilula, Krieger & Ritov, 1988), and the quantity $X^2/N$ has been called a "Pearsonian distance" (Gilula, et. al. 1988), a "chi–square distance," the "total interia" of a table (Greenacre, 1984; van der Heijden, de Falugerolles & de Leeuw, 1989; and many others from the correspondence analysis literature).

Additive models focus on the departures from independence as measured by Pearson's $X^2$. The additive model orthogonally decomposes the lack of fit or the departure from independence into amounts due to various interactions (Darroch, 1974). For instance, the departure of each cell in a 3–way table from independence can be partitioned as follows

$$\frac{(P_{ijk} - P_{i..}P_{.j.}P_{..k})}{P_{i..}P_{.j.}P_{..k}} = \frac{(P_{ij.} - P_{i..}P_{.j.})}{P_{i..}P_{.j.}} + \frac{(P_{i.k} - P_{i..}P_{..k})}{P_{i..}P_{..k}} + \frac{(P_{.jk} - P_{.j.}P_{..k})}{P_{.j.}P_{..k}} + \frac{(P_{ijk} -_a \hat{P}_{ijk})}{P_{i..}P_{.j.}P_{..k}}$$

where ${}_a\hat{P}_{ijk}$ is the prediction from the additive, no 3–way interaction model (equation 3.1). If equation 3.1 fits perfectly, then the last term on the right side of this equation equals zero. The additive model also provides a global partitioning of the total association

or departure from independence in the table as measured by Pearson's $X^2$. This global partitioning is given by

$$\frac{X^2}{N} = \sum_{i,j} \frac{(P_{ij.} - P_{i..}P_{.j.})^2}{P_{i..}P_{.j.}} + \sum_{i,k} \frac{(P_{i.k} - P_{i..}P_{..k})^2}{P_{i..}P_{..k}} + \sum_{j,k} \frac{(P_{.jk} - P_{.j.}P_{..k})^2}{P_{.j.}P_{..k}} + \sum_{i,j,k} \frac{(P_{ijk} -_a \hat{P}_{ijk})^2}{P_{i..}P_{.j.}P_{..k}}$$

(Darroch, 1974; Bishop, Fienberg and Holland, 1975). The first three terms on the right side of this equation represent the contributions of each of the 2–way interactions, and the fourth term represents the contribution of the 3–way interaction (the lack of fit of equation 3.1). Multiplicative models partition Pearson's $X^2$ only in the case of complete independence where $u_{ij}^{AB} = u_{ik}^{AC} = u_{jk}^{BC} = u_{ijk}^{ABC} = 0$ (Bishop, Fienberg, and Holland, 1975; Darroch, 1974).

The orthogonal partitioning property of $X^2$ associated with additive models has been given as an advantage of additive models over multiplicative models (Lancaster, 1969; Darroch, 1974; and others from the correspondence literature). While additive models partition $X^2$, multiplicative models partition the likelihood–ratio statistic $G^2$ into components due to various effects. The departure of each cell in a 3–way table from the model of independence can be decomposed into parts due to each of the 2–way interactions and the 3–way interaction; that is,

$$\ln(P_{ijk}/_m P_{ijk}) = u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC}$$

where ${}_mP_{ijk}$ is the fitted values from the $(A, B, C)$ loglinear model and the $u$-terms are from the saturated loglinear model. The multiplicative model also gives a global partitioning of $G^2$ into components due to the various interactions as follows:

$$\frac{G^2}{2N} = \sum_{i,j} P_{ij.} u_{ij}^{AB} + \sum_{i,k} P_{i.k} u_{ik}^{AC} + \sum_{j,k} P_{.jk} u_{jk}^{BC} + \sum_{i,j,k} P_{ijk} u_{ijk}^{ABC}$$

The additive model does not lead to any such partitioning of $G^2$. Under certain regularity conditions, both $G^2$ and $X^2$ are asymptotically equivalent and have chi-squared distributions (Read & Cressie, 1988). While $G^2$ and $X^2$ often yield very similar values when computed for the same table, there are cases when they are substantially different and one measure may be preferable to the other (Fienberg, 1985; Read & Cressie, 1988; Wickens, 1989). Pearson's $X^2$ tends to give too much weight to small cells (i.e., less than 5). An additional advantage of $G^2$ over $X^2$ is that the differences between $G^2$ statistics for nested models are asymptotically chi-squared distributed.

In summary, odds ratios readily extend to higher–way tables and can measure the interaction among 3 or more variables, while correlations are limited to measuring the interaction between pairs of variables. Odds ratios do not depend on the margins of tables or on the total number of observations, while correlations depend on both of these. For further discussion of the differences between these measures, and examples of tables that have the same odds ratios but different correlations and visa versa, see Goodman (1991). Multiplicative (loglinear) models define interactions in terms of odds ratios computed on sub-tables and partition $G^2$. Additive models define interactions in terms of correlations

between pairs of variables and partition $X^2$. Based on these properties, odds ratios and multiplicative models are preferable to correlations and additive models for analyzing the play quality data.

### 3.1.2 General Properties of Additive and Multiplicative Models

Three types of properties that differentiate additive and multiplicative models are reviewed here. Properties of "independence", "existence", and "invariance" are defined and described in Sections 3.1.2.1, 3.1.2.2, and 3.1.2.3, respectively.

#### 3.1.2.1 Independence

Properties of complete, joint, and conditional independence are distinct from *models* of complete, joint, and conditional independence. The *properties* are defined with respect to data and specify that a certain relationship holds among the cells of a table. Additive and multiplicative *models* of complete, joint, and conditional independence refer to particular cases of additive and multiplicative models and specify which interactions are in the models. To facilitate an explication of this point, consider a 3–way table containing probabilities $\{p_{ijk}\}$ where $i$, $j$, and $k$ index variables $A$, $B$ and $C$, respectively. Since the concepts of joint and conditional independence are defined for three or more variables, a 3–way table is needed to illustrate the above statement. The concepts extend to four or more variables in a straightforward manner.

The property of complete independence means that there is no dependency between any of the variables; that is, $p_{ijk} = p_{i..}p_{.j.}p_{..k}$ for all $i$, $j$, and $k$. The property of joint

independence between $A$ and $BC$ (for instance) can be thought of as complete independence between the variables $A$ and $BC$, where the variable $BC$ has $JK$ categories, each of which corresponds to a combination of variables $B$ and $C$. If a table is characterized by joint independence, then the equation $p_{ijk} = p_{i..}p_{.jk}$ holds for all $i$, $j$, and $k$. The property of conditional independence between $A$ and $B$ given $C$ (for instance) means that for each level of variable $C$, the variables $A$ and $B$ are completely independent. In other words, $p_{ijk} = p_{i.k}p_{.jk}/p_{..k}$ for all $i$, $j$, and $k$.

Models of complete independence contain only terms representing the "main" effects of variables. Models of joint and conditional independence contain interaction terms in addition to main effect terms. Models of joint independence between $A$ and $BC$ (for instance) contain one additional term for the $BC$ interaction. Models of conditional independence between $A$ and $B$ given $C$ (for instance) contain two additional terms, one for the $AC$ interaction and one for the $BC$ interaction.

Regardless of the number of variables, additive and multiplicative models of complete independence are equivalent (Darroch, 1974). In other words, the models yield the same predictions and have the same number of parameters. Furthermore, additive and multiplicative models of complete independence imply the property of complete independence. Since joint independence can be thought of as a complete independence for two "variables" where one "variable" is actually a combination of variables (e.g., $BC$), the properties and models of joint independence behave in the same way that they do for complete independence.

Additive and multiplicative models of conditional independence differ. The multiplicative model implies the property of conditional independence, but the additive model does not (Darroch, 1974; Goodman, 1987a, 1987b). For example, the $(AC, BC)$ multiplicative model of conditional independence,

$$_{m}P_{ijk} = \tau\tau_i^A\tau_j^B\tau_k^C\tau_{ik}^{AC}\tau_{jk}^{BC}$$

will perfectly fit a 3–way table characterized by the property of conditional independence where variables $A$ and $B$ are independent given variable $C$. The analogous additive model of conditional independence,

$$_{a}P_{ijk} = P_{i..}P_{.j.}P_{..k} + P_{.j.}\Delta_{ik}^{AC} + P_{i..}\Delta_{jk}^{BC}$$

will *not* fit a table characterized by the property of conditional independence, and in general, neither will the additive, no 3–way interaction model. The additive model implies only that variables $A$ and $B$ are independent in the marginal table (i.e., $P_{ij.} = P_{i..}P_{.j.}$) (Darroch, 1974). Marginal independence does not imply conditional independence or vice versa, which allows for Simpson's paradox. In case of Simpson's paradox, the relationship between variables in the marginal table are in the opposite direction as the relationship between the variables in the sub-tables for a given level of the third variable (Bishop, Fienberg & Holland, 1975; Agresti, 1990; Wickens, 1989). Relative to multiplicative models, additive models tend to result in less parsimonious representations and conclu-

sions regarding the associations in tables characterized by the property of conditional independence. This lack of parsimony leads some researchers to prefer multiplicative models over additive ones.

For the play quality data, one goal of the analysis is to find the most parsimonious yet adequate and accurate representation of the interactions. The preliminary analyses reported in Chapter 2 favor multiplicative models over additive models. Specifically, there is some evidence that "play" (P) and "time" (T) are independent given "group" (G), and the conditional model loglinear model $(GT, GP)$ might be adequate. Even if this model is not adequate, the slightly more complex, no 3-factor interaction loglinear model might fit. If additive models were fit to this table, it is unlikely that a model without a 3–way interaction would fit it.

#### 3.1.2.2 Existence

Conditional independence is not the only property where the additive and multiplicative models differ. Given any consistent set of marginal probabilities (i.e., the marginal probabilities sum to 1), multiplicative models will always be defined, but for certain sets of consistent marginal probabilities, the additive model will not (Darroch, 1974; Goodman, 1978, 1985; Aldrich & Nelson, 1984). This lack of an additive model is the most serious drawback of additive models. Linear probability models, which are examples of additive models, can yield predictions of probabilities outside the range of 0 to 1.

#### 3.1.2.3 Invariance

Sub-table invariance and amalgamation invariance are two additional properties that Darroch (1974) uses to differentiate additive and multiplicative models. A model is *sub-table invariant* if it holds simultaneously for the full table and for every sub-table. This property is characteristic of multiplicative but not additive models. In contrast, amalgamation invariance is characteristic of additive but not multiplicative models. A model is said to possess *amalgamation invariance* if the predicted probabilities from the model fit to a collapsed table equal the sum of the predictions from the model fit to the uncollapsed table (Darroch, 1974; Goodman, 1979). Different researchers using multiplicative models to analyze different tables based on the same data (e.g., collapsed versus uncollapsed or collapsed in different ways) could produce different results and possibly conflicting conclusions.

With respect to the play quality data, sub-table invariance is definitely a desirable property while amalgamation invariance is not important. One of the proposed ways to deal with the zeros missing at random for the group MOL at the third time point is to delete the group from the analyses involving all five sessions. If a model fits the full table including MOL, then it should also fit the sub-table without the group MOL. In other words, the basic results of the analyses of the sub-table without the missing data and those of the full table should be the same. Amalgamation invariance is not important for the play data, because we do not want to collapse the table. Of substantive interest are comparisons between different groups, time periods, and play qualities.

In summary, both additive and multiplicative models are equivalent with respect to complete and joint independence, but differ with respect to conditional independence. Multiplicative models tend to be simpler and more parsimonious than additive models, are always defined, and are sub-table invariant. Additive models are not always defined, but are invariant with respect to the amalgamation of categories. Given these general properties and those reviewed in subsection 3.1.1 on measures of association, multiplicative models are preferable to additive models and are more appropriate for analyzing the play quality data.

## 3.2 Models for 2–Way Tables

In subsections 3.2.1 and 3.2.2, association and correlation models for 2–way tables are presented. Correlation and association models are additive and multiplicative models, respectively, in which the interaction terms are decomposed into components; therefore, these models possess the properties described in the previous section. Since the association model is a multiplicative model, this model is theoretically preferable to the correlation model, which is an additive model. In subsection 3.2.3, more specific theoretical comparisons of the association and correlation models are made, and an empirical comparison made using the play data.

### 3.2.1 Association Models

In Section 3.2.1.1, the association model is presented, and in Section 3.2.1.2, fitting the model to data and the estimation of the model parameters are described. After discussing the interpretation of the model in Section 3.2.1.3, in Section 3.2.1.4, the model is used to analyze the peer play data from the first occasion .

#### 3.2.1.1 The Model

The association model is a multidimensional extension of the saturated loglinear model for a 2–way table where the interaction between the rows and columns is decomposed into components (Gilula & Haberman, 1986, 1988; Goodman 1981b, 1985, 1986, 1987b, 1991; Wickens, 1989; Agresti, 1990). Let $P_{ij}$ be the probability of observing an object or individual at level $i$ on variable $A$ (rows) and level $j$ on variable $B$ (columns) where $i = 1, \ldots, I$ and $j = 1, \ldots, J$. The multidimensional row–column association model (or the $RC(M)$ association model, for short) is

$$P_{ij} = \alpha_i \beta_j \exp(\sum_{m=1}^{M} \phi_m \mu_{im} \nu_{jm}) \tag{3.4}$$

or in log-scale

$$\ln(P_{ij}) = u + u_i^A + u_j^B + \sum_{m=1}^{M} \phi_m \mu_{im} \nu_{jm} \tag{3.5}$$

where $\mu_{im}$ and $\nu_{jm}$ are scale values on dimension $m$ of row $i$ and column $j$, respectively, $\phi_m$ is a measure of the intrinsic association between the rows and columns on dimension $m$, and $M \leq \min(I, J)-1$. The degrees of freedom for this model equal $(I-M-1)(J-M-1)$.

The model is saturated and fits perfectly for $M = \min(I, J) - 1$, but it often provides an adequate fit for $M < \min(I, J) - 1$, such as $M = 1$ or 2 (for examples, see Becker & Clogg, 1989a; Gilula & Haberman, 1988; Goodman, 1979, 1981a, 1981b, 1985, 1986, 1987b, 1991; Wickens, 1989).

#### 3.2.1.2 Estimation

When the observations in a table are random samples from a multinomial, product multinomial, or independent Poisson distributions, the parameters in equation 3.4 or 3.5 can be estimated by maximum likelihood (MLE), which involves an iterative method (Goodman, 1985, 1986; Gilula & Haberman, 1988). Regardless of distributional assumptions, the parameters can always be estimated by least squares (LSE) by finding the singular value decomposition of a matrix (Gilula, 1982; Escoufier & Junca, 1986; Escoufier, 1988), which minimizes the quantity $(\sum_i \sum_j (p_{ij} - P_{ij})^2 g_i h_j)$, where $\{p_{ij}\}$ are observed probabilities, $\{P_{ij}\}$ are fitted values from equation 3.4, and $g_i$ and $h_j$ are row and columns weights that are defined later. The major advantages of MLE over LSE are that both the fit of particular models and the difference in fit between the two models (where one model is a special case of the other) can be statistically tested. An additional advantage is that asymptotic standard errors of model parameters can be estimated. The major advantages of LSE over MLE are that the process of fitting models to tables is computationally efficient and straightforward, and it is not necessary to make distributional assumptions. Given the availability of computing resources and advances in computing power, the large computational demands that are associated with MLE are less problem-

atic than they once were. Least squares estimation is well suited for exploratory and/or descriptive purposes. An example can be found in Becker (1990b), where LSE is used to approximate possible models, and then the LSE parameters are used as starting values in the MLE iterative routine. The use of LSE to obtain starting values for parameters in MLE routines is a standard practice, because they are consistent and efficient.

To estimate the parameters in equation 3.4 or 3.5, the following identification constraints on the location and scale of the row and column scores are imposed:

$$\begin{aligned} \sum_i \mu_{im} g_i &= \sum_j \nu_{jm} h_j = 0 \\ \sum_i \mu_{im} \mu_{im'} g_i &= \sum_j \nu_{jm} \nu_{jm'} h_j = \delta_{mm'} \end{aligned} \tag{3.6}$$

where $\delta_{mm'} = 1$ for $m = m'$ and $\delta_{mm'} = 0$ for $m \neq m'$, and $g_i$ and $h_j$ are known weights for row $i$ and column $j$, respectively. These constraints are more general than those typically given. The two most frequently used systems of weights are marginal probabilities (i.e., $g_i = P_{i.}$ and $h_j = P_{.j}$) (Goodman 1981a, 1981b, 1985, 1986, 1987, 1991; Becker & Clogg, 1989a; Wickens, 1989), and unit or uniform weights (i.e., $g_i = h_j = 1$ or $g_i = 1/I$ and $h_j = 1/J$) (Goodman, 1979, 1991; Becker & Clogg, 1989a; Gilula & Haberman, 1988; Wickens, 1989).

The constraints are arbitrary with respect to the fit of the model (i.e., predicted cell probabilities and predicted odds ratios are the same regardless of the weights used), but not with respect to the parameter estimates (Becker & Clogg, 1989a). Different weighting schemes can lead to different interpretations of data and of the model itself.

The choice of weights obviously affects the size of the intrinsic association parameter, and the measured dependency in the table (Goodman, 1991). The weights also can affect the relative spacing and the order of the estimated row and column scores[1] (Becker & Clogg, 1989a). The further the marginal probabilities deviate from a uniform distribution, the greater the likelihood that the rank orders and relative spacing of scale values will be different under uniform (or unit) weighting schemes and marginal probabilities.

In the MLE iteration routines used to obtain estimates of the parameters in equation 3.4 or 3.5, a generalized singular value decomposition is used to impose the scaling and orthogonality restrictions on the scale values (i.e., $\sum_i \mu_{im}\mu_{im'}g_i = \sum_j \nu_{jm}\nu_{jm'} = \delta_{mm'}$) (Becker, 1990; Becker & Clogg, 1989b). Generalized singular value decompositions are not necessarily monotonic transformations, so the rank order of scale values under alternative weight systems can differ. The dependence of the rank order of scale values on the weights used in the identification constraints would seem to be an especially unappealing and troubling feature of the $RC(M)$ association model, particularly when the scale values are given substantive interpretations. However, the choice of weights should be made on the basis of substantive considerations about the data and desired interpretational aspects of the model.

When marginal probabilities are used as weights, inferences about interactions are confounded with the observed marginal probabilities in the sense that the relative order and spacing of the scale values and the size of the intrinsic association parameter de-

[1]Wickens (1989) erroneously states that the standardization affects the relative spacing but not the order of category scale values.

pend on the observed marginal probabilities, as well as, on the nature of the interaction (Becker & Clogg, 1989a; Goodman, 1991). If the marginal probabilities do not represent an intrinsic marginal distribution of some larger population, as in the case when they are determined by an experimenter, then uniform or unit weights should be used (Wickens, 1989). If the variables are samples from some population and the sample marginal probabilities equal (or approximate) those in the population, then marginal probabilities should be used (Wickens, 1989). When the weights are probability measures, the restrictions can be interpreted as fixing the means and variances to 0 and 1, respectively (Becker & Clogg, 1989b; Wickens, 1990), and the quantity $\sum_i \sum_j \mu_{im}\nu_{jm}P_{ij}$ can be interpreted as the correlation between the rows and columns on dimension $m$.

For the peer play data, unit weights should be used in the identification constraints. Unit weights allow comparisons to be made across time and/or groups without confounding the marginal distributions and the interactions in the table. Furthermore, neither the "play" (P), "group" (G), "time" (T), nor the "group" by "time" (TG) marginal distributions reflect underlying population distributions. These margins are determined by the design of the study. The same children are in each group at each time point, and there is no reason to assume that the distribution of children across groups reflects the distribution in the population. Twelve families with boys and 18 with girls were selected for the study, and the classification of the children into younger or older age groups and into high or low sibling adjustment is based on the median value of these variables for the children in the study. The distribution of children between the different groups affects the G margin. The missing data for the third occasion means that there are fewer

observations or frequencies at this time point (i.e., the T margin is affected by the fact that five children were not observed at the third session). The P margin is also affected by the design of the study in that different numbers of the 44 recorded behaviors are collapsed into the eleven play categories, which are interpreted as representing qualities of play. Given these characteristics of the design, unit weights are the logical and most appropriate choice.

#### 3.2.1.3 Interpretation

In terms of the parameters of the $RC(M)$ association model, the log-odds ratios are given by

$$\ln \Theta_{ii',jj'} = \sum_{m=1}^{M} \phi_m(\mu_{im} - \mu_{i'm})(\nu_{jm} - \nu_{j'm})$$

The parameter $\phi_m$ measures the association between the rows and columns on dimension $m$ for a unit change in both the row and column scale values. When $M = 1$, $\phi$ is the log–odds ratio corresponding to a unit change in both the row and column scale values. The scale values and $\phi_m$ represent the dependency in the table, which can also be seen by expressing the interaction parameters $u_{ij}^{AB}$ from the saturated loglinear model as

$$\begin{aligned} u_{ij}^{AB} &= \ln\left(P_{ij}/(\alpha_i\beta_j)\right) \\ &= \sum_{m=1}^{M} \phi_m \mu_{im} \nu_{jm} \end{aligned}$$

where $M = \min(I, J) - 1$. The $RC(M)$ association model can be thought of as decomposing the total association in a table as measured by the likelihood ratio statistic $G^2$

for testing independence or the Kullback–Leibler measure, $G^2/2N$, into $M$ orthogonal components; that is,

$$\begin{aligned} \frac{G^2}{2N} &= \sum_{i=1}^{I}\sum_{j=1}^{J} P_{ij} u_{ij}^{AB} \\ &= \sum_{m=1}^{M} \phi_m \rho_m \end{aligned}$$

where $\rho_m = \sum_i \sum_j \mu_{im}\nu_{jm}P_{ij}$, which is the product moment (Pearson) correlation between the row and column scores on dimension $m$. The association parameter itself equals a weighted sum of the product of the row and column scale values with the weights equal the logarithms of the cell probabilities,

$$\phi_m = \sum_{i=1}^{I}\sum_{j=1}^{J} \mu_{im}\nu_{jm} \ln(P_{ij}) \tag{3.7}$$

(Goodman, 1985, 1986, 1991), which is similar to a product moment correlation coefficient.

The row and column scores can be plotted to provide a visual display of the relationship among the categories of each variable and the interaction between the variables. Observed patterns in these plots can suggest simpler models, such as equal spacing between categories or equivalent scores for some categories. Both row and column scores are often plotted in the same figure to provide a picture of the association between the variables. Goodman (1986) suggests plotting normalized scores: $\phi_m^{\gamma}\mu_{im}$ and $\phi_m^{\delta}\nu_{jm}$ where $\gamma+\delta = 1$, which yields $(\phi_m)^{1/2}\mu_{im}$ and $(\phi_m)^{1/2}\nu_{jm}$ as a special case. In these plots, the in-

ner product between a row and column point equals a cell's contribution to the measured dependency (lack of independence) between the variables (i.e., $\sum_m \phi_m \mu_{im} \nu_{jm}$). Since the scale of the row and column scores depend on the weights used in the identification restrictions and on the chosen normalization, the absolute distances between points should *not* be interpreted; however, relative distances between points are informative about the relationship between the categories of a variable and the relationship between the variables. A thorough discussion of the interpretation and geometric properties of such plots can be found in Goodman (1986). Examples where such plots are used can be found in Becker (1989a, 1990b) and Clogg, Eliason, and Wahl (1990).

When categorical variables are ordinal, assigning numerical scores to the categories is reasonable and natural. Ordered variables are sometimes just continuous variables that have been discretized or measured on a discrete scale. For example, in the play quality data set, age, which is inherently a continuous variable, is treated as a discrete variable with children classified into one of two categories: younger or older. It is often reasonable to assume that ordered variables are discretized, continuous variables from normal distributions. The $RC(1)$ association model is implied by any joint distribution of two continuous variables where each of the variables can be separately transformed to a normal distribution (Goodman, 1981a, 1985). The parameters of the $RC(1)$ model can be used to estimate the parameters of the underlying continuous distribution. Furthermore, the association model can be used to compute integrals of the bivariate normal distribution (Becker, 1989; Goodman, 1981, 1985, 1991; Lancaster, 1957; Rom & Sarkar, 1990; Wang, 1987).

With ordinal and/or nominal variables, the $RC(M)$ association model can be interpreted in terms of latent, normal variables (Whittaker, 1989; Lauritzen & Wermuth, 1989). Bartholomew's (1980, 1987) latent (normal) variable model for two observed categorical variables and one latent variables is a equivalent to the $RC(1)$ association model. This interpretation arises from the relationship between Bartholomew's model and the probability density function of normal variables, which is a special case of the $RC(1)$ association model. An alternative and more general interpretation of the $RC(M)$ association model states that if the $RC(M^*)$ association model fits an $(I \times J)$ table of variables $A$ and $B$, then there are $M^*$ latent normal variables $Z_m$ such that $A$ and $B$ are independent given $Z_1, \ldots, Z_{M^*}$; $Z_m$ is independent of $Z_{m'}$ given $(a_i, b_j)$ for all $m \neq m'$; and $Z_m|(a_i, b_j) \sim N(\mu_{im} + \nu_{jm}, (1/\phi_m))$, which is the usual ANOVA model except that the "dependent" or response variables are unobserved (Whittaker, 1989).

While scaling ordered variables is quite natural, scaling nominal variables might seem unjustified or even nonsensical; however, analyzing the association between two nominal variables by the $RC(M)$ association model can be though of as a summary or description of the relationship. Variables do not have to be ordinal for concepts such as entropy, uncertainty, and information to be applicable (Krippendorff, 1986). The $RC(M)$ association model has an interpretation in terms of such information theoretic concepts. Gilula, Krieger, & Ritov (1988) show that the $RC(1)$ association model maximizes the information (negative entropy) given the marginal distributions (i.e., $P_{i\cdot}$ and $P_{\cdot j}$), the row and column scores, and the correlation between them. They state that the intrinsic association parameter $\phi_1$ can be interpreted as a measure of uncertainty. In the multidi-

mensional case where $M \geq 2$, the $\phi_m$ parameters for $m = 1, 2, \ldots M$ can be interpreted as measures of the information or the reduction in uncertainty gained from the knowledge of the scores on each dimension given the scores on previous dimensions. The $RC(M)$ association model can be interpreted as decomposing $G^2$, and $G^2$ is proportional to the increase in information from knowing $P_{ij}$ instead of just $P_{i.}$ and $P_{.j}$. More specifically,

$$\begin{aligned} G^2/2N &= \sum_i \sum_j P_{ij} \ln\left(\frac{P_{ij}}{\hat{P}_{ij(0)}}\right) \\ &= \sum_i \sum_j P_{ij} \ln\left(\frac{\hat{P}_{ij(M^*)}}{\hat{P}_{ij(0)}}\right) + \sum_i \sum_j P_{ij} \ln\left(\frac{P_{ij}}{\hat{P}_{ij(M^*)}}\right) \end{aligned}$$

where $\hat{P}_{ij(0)}$ is the prediction from the model of independence (i.e., $M = 0$) and $\hat{P}_{ij(M^*)}$ is the prediction from the $RC(M^*)$ association model. The term $\sum_{i,j} P_{ij} \ln(\hat{P}_{ij(M^*)}/\hat{P}_{ij(0)})$ represents the amount of entropy explained (information transmitted or gained) by the association model, and the term $\sum_{i,j} P_{ij} \ln(P_{ij}/\hat{P}_{ij(M^*)})$ represents the lack of fit of the association model.

#### 3.2.1.4 The Peer Play Data: the first session

As an example, the $RC(M)$ association model was fit to the Group $\times$ Play sub-table of the first occasion ($k = 1$) using the data from all $N = 30$ children, which is the first $8 \times 11$ sub-matrix in Table 2.1. The fit statistics for $M = 0, 1$, and 2 are reported in Table 3.1.

Given the large value of the fit statistics relative to the degrees of freedom for $M = 0$, the variables "group" and "play" are certainly not independent at $k = 1$. Both the

**Table 3.1**: Fit of the $RC(M)$ association model to the Group $\times$ Play sub-table for $k = 1$ ($N = 30$).

| $M$ | $df$ | $G^2$ | $G^2/df$ | $X^2$ | $X^2/df$ | ($p$-value) |
|---|---|---|---|---|---|---|
| 0 | 70 | 128.58 | 1.84 | 121.90 | 1.74 | .000 |
| 1 | 54 | 60.19 | 1.11 | 55.64 | 1.03 | .412 |
| 2 | 40 | 34.16 | .85 | 29.39 | .73 | .892 |

**Table 3.2**: Analysis of Association of the Group $\times$ Play sub-table for $k = 1$ ($N = 30$).

| Component | Models | $df$ | $G^2$ | $G^2/df$ | ($p$-value) | Percent |
|---|---|---|---|---|---|---|
| $m = 1$ | $(G, P)$ - $RC(1)$ | 16 | 68.39 | 4.27 | .000 | 53.19% |
| $m = 2$ | $RC(1)$ - $RC(2)$ | 14 | 26.03 | 1.86 | .026 | 20.24% |
| unexplained by $RC(2)$ | $RC(2)$ | 40 | 34.16 | .85 | .730 | 26.57% |
| total | $(G, P)$ | 70 | 128.58 | 1.84 | .000 | 100.00% |

uni- and two dimensional models fit the data much better than the independence model. Since the sampling assumption of independent observations is not valid, the $p$-values are suspect and should not be used to decide between the one and two dimensional models. We need to consider whether the second dimension of the $RC(2)$ association model is really needed or whether the single dimension of the $RC(1)$ association model is sufficient to explain the dependency in the table. If only one dimension is needed, then the $RC(2)$ association model overfits the data and the second dimension may not be not very stable. To help examine the relative contribution or the "importance" of each of the two dimensions with respect to accounting for the dependency in the table, an analysis of association table (Goodman, 1991, and others), which is analogous to an ANOVA table, is presented in Table 3.2.

In the first column of Table 3.2 are the source or component effects. The third and fourth columns contain the degrees of freedom and likelihood ratio statistics, respectively, for each of the effects. Each of the $G^2$'s is either the fit statistic for a specific model or the difference between likelihood ratio statistics for two models. The model(s) used to compute $G^2$ are reported in the second column. The last column contains the percent of the total dependency (non-independence) accounted for by each source or component. The single dimension of the $RC(1)$ association model accounts for 53.19% of the association in the data, which leaves 46.81% of the total dependency in the Group $\times$ Play sub-table unexplained. The second dimension of the $RC(2)$ association model accounts for an additional 20.24%, which leaves 26.57% of the total dependency unexplained. For $m = 2$, $G^2/df$ is relatively large (i.e., $G^2/df = 1.86$), which indicates that the second dimension adds a "significant" amount of information. However, the estimated scale values from both the $RC(1)$ and $RC(2)$ association models are examined and compared to see if one of the models has a more sensible interpretation than the other.

To estimate the parameters of the $RC(M)$ association models, unit weights were used in the identification constraints for the reasons previously discussed. For the $RC(1)$ model, the estimated association parameter $\hat{\phi}$ equals 3.663, and the scale values ($\hat{\mu}_i$ and $\hat{\nu}_j$) are plotted in Figure 3.1. Since there is only a single dimension, the scale values are not normalized (re-scaled) by $\hat{\phi}$. The play qualities are distributed such that the more immature or "bad" qualities tend to have negative scale values while the more mature or "good" qualities have positive scores. The quality of negative emotion (nem) is the most extreme of the "bad" qualities followed by the qualities of fighting (fight) and poor

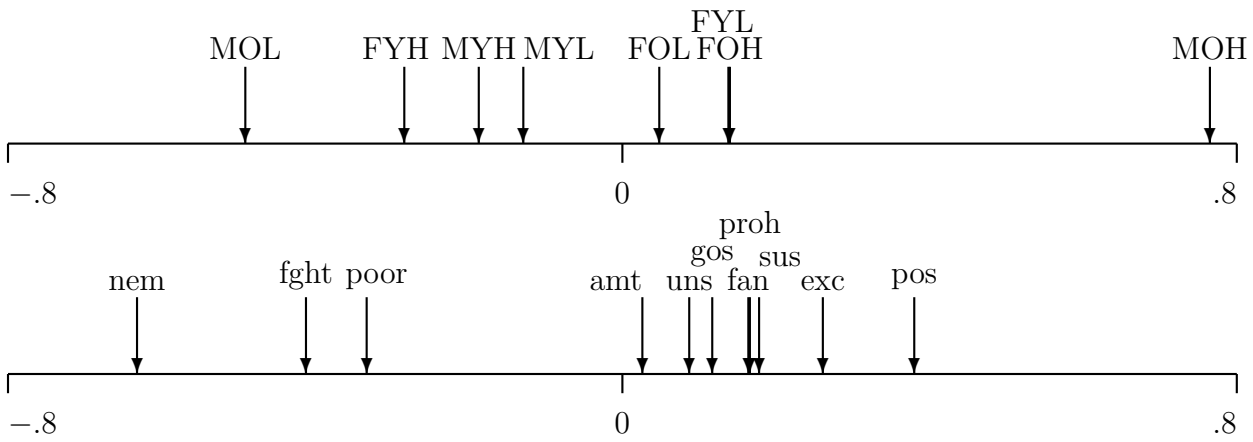



**Figure 3.1**: Estimated scale values for groups ($\hat{\mu}_i$) and play qualities ($\hat{\nu}_j$) from the $RC(1)$ association model with unit weights.

play (poor). Of the "good" qualities, positive play (pos) is the most extreme, followed by excitability (exc).

The group scale seems to be primarily an age "factor" where the younger children tend have negative scores and the older children tend to have positive scores. The two exceptions to this interpretation are the older males who showed low sibling acceptance (MOL), and the younger females who showed low sibling acceptance (FYL). The MOL group contains only one child and the results may not be representative of MOL children in the population (i.e., generalizations to the population should not be made). The observation that the MOL child shows more immature behaviors and fewer mature ones may be related to the low sibling acceptance/adjustment (i.e., this kid has problems).

The close proximity of the FYL group and the FOH groups is more difficult to "explain away." One possibility is that the groups FOL, FYL and FOH are all relatively close to zero and may not be significantly different from zero. Estimates of the standard errors of the scale values would be useful, but because of the violation of the sampling assumption, these were not even estimated to avoid the temptation of examining them. Another possibility is that the one dimensional solution is not sufficient and the second dimension of the $RC(2)$ association model is needed.

Overall, Figure 3.1 shows that the children with positive scale values (MOH, FOH, FOL, and FYL) tend to exhibit more mature qualities and fewer immature ones than expected if the variables "group" and "play" were independent. Likewise, the children with negative scale values (MYL, MYH, FYL, and MOL) tend to exhibit more immature qualities and fewer mature ones than expected if "group" and "play" were independent.

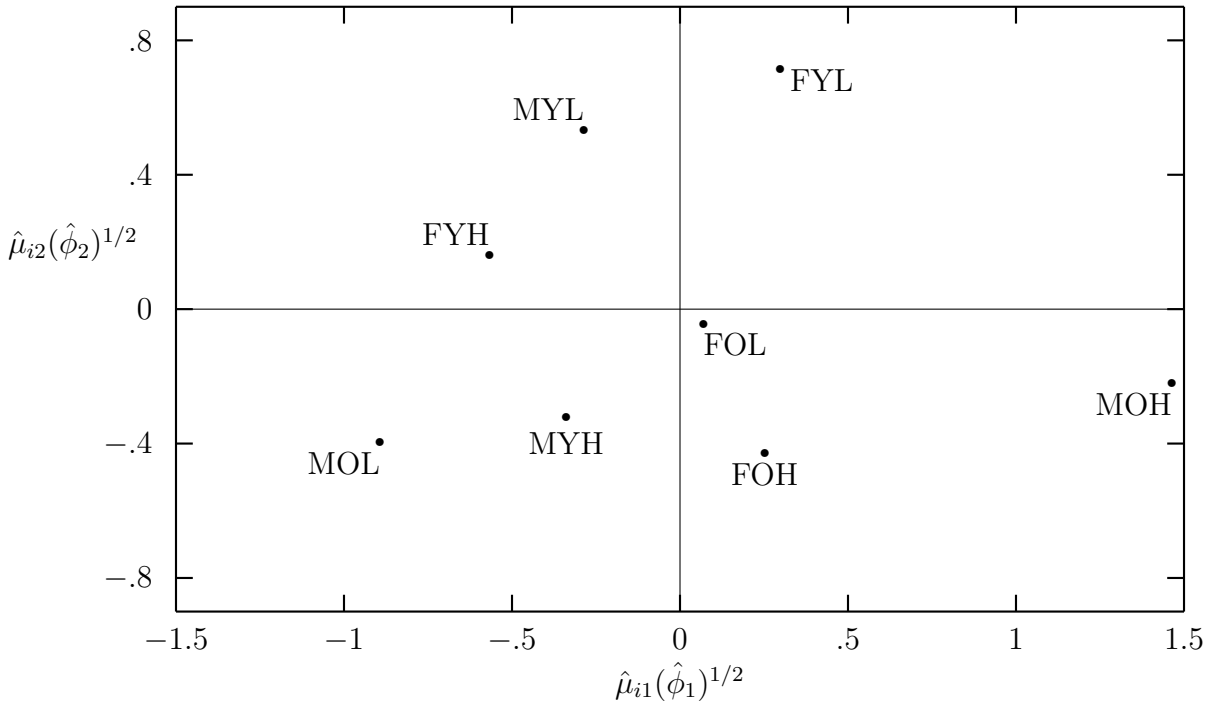



**Figure 3.2**: Estimated group scale values of the $RC(2)$ association model with unit weights fit to the Group × Play sub-table from the first occasion ($N = 30$).

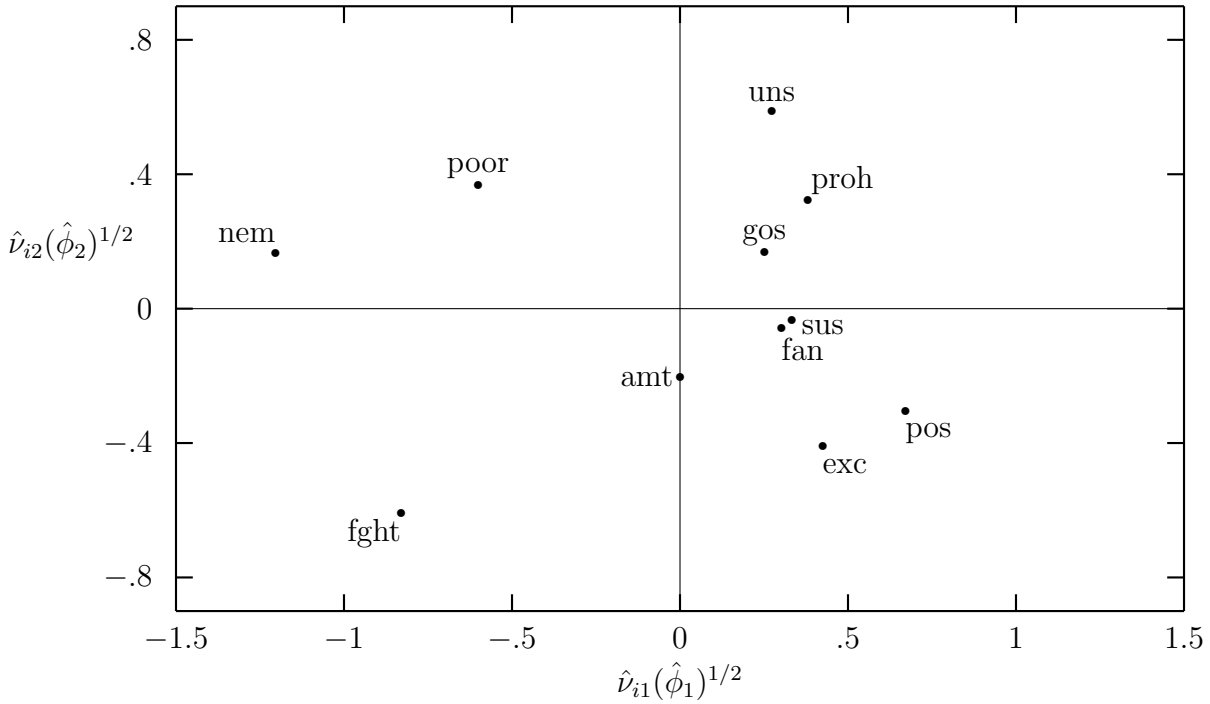



**Figure 3.3**: Estimated play scale values of the $RC(2)$ association model with unit weights fit to the Group × Play sub-table from the first occasion ($N = 30$).

The estimated parameters of the $RC(2)$ association model help explain the close proximity of the FYL and FOH on the first dimension of the $RC(1)$ association model. The estimated group and play scale values for the $RC(2)$ association model are plotted in Figures 3.2 and 3.3, respectively. The scale values were normalized by multiplying them by $\sqrt{\hat{\phi}_m}$, where $\hat{\phi}_1 = 3.615$ and $\hat{\phi}_2 = 1.314$. The scale values on the first dimension of $RC(1)$ and $RC(2)$ association models can differ, but the order and relative spacing of the groups and play qualities along the first dimension are very similar. The second dimension primarily contrasts the younger children with low sibling acceptance (FYL and MYL) against older children and younger males with high sibling acceptance (FOH, MOH, MOL, and MYH). In the two dimensional space, the second dimension pulls the FYL group away from the older females. With respect to the play qualities, the second dimension primarily contrasts unsustained play (uns) and fighting (fight).

While a reasonable interpretation can be made of both dimensions, whether the second dimension *should* be interpreted is another question. The second dimension pulls FYL away from FOL and FOH, but it could be unstable and due to chance or quirks in the data. It may not reflect anything "real" in terms of the population of children from which the subjects were members.

Taken together, Figures 3.2 and 3.3 represent the interaction between the groups and play qualities. Suppose that the scale values for each of the variables have been drawn in the same plot (i.e., a "joint" plot), and consider vectors drawn from the origin (0,0) to the points representing the categories of each of the variables. The inner products between vectors for particular groups and play qualities equal the interaction for the

corresponding cell in the Group × Play table. If the vectors for categories of the same variables point in a similar direction, then these categories are similar with respect to their interactions. If the vector for a group points in the same direction as a vector for a play quality, then that group tends to exhibit more of the quality than would be expected if the two variables were independent. Alternatively, if the vector for a group points in the opposite direction as the vector for a play quality, that group tends to exhibit less of the quality than would be expected if the two variables were independent. For example, since the vectors for the groups FOL, FOH, and MOH point in the same general direction as positive, excite, fantasy, and sustain, these children exhibited more of these qualities than the children in the other groups and more than expected if group and play were independent. Likewise, the child in the MOL group showed more fighting and negative emotion than other children and more than expected under independence.

### 3.2.2 Correlation Models

The correlation model is presented in Section 3.2.2.1, and the model is used to analyze the peer play data from the first session in Section 3.2.2.2.

#### 3.2.2.1 The Correlation Model

The correlation model for 2–way tables is an extension of the saturated, linear or additive model for 2–way tables. Deleting all of the terms associated with variable $C$, indexed by $k$, from the additive model given in equation 3.1 yields the independence model for

2–way tables. The saturated additive model for 2–way tables can be written

$$P_{ij} = P_{i.}P_{.j} + \Delta_{ij}^{AB}$$

where $\Delta_{ij}^{AB} = P_{ij} - P_{i.}P_{.j}$, which represents the interaction between variables $A$ and $B$. This model will perfectly fit any 2–way table. The correlation model is a multidimensional generalization of this saturated additive model where the interaction terms are decomposed into components as follows:

$$\Delta_{ij}^{AB} = P_{i.}P_{.j}(\sum_{m=1}^{M} \lambda_m x_{im} y_{jm})$$

where $x_{im}$ and $y_{jm}$ are the scale values or "canonical scores" on dimension $m$ of row $i$ and column $j$, respectively, $\lambda_m$ is the (canonical) correlation coefficient for dimension $m$, and $M \leq \min(I, J) - 1$ (Gilula & Haberman, 1986, 1988; Goodman, 1985, 1986, 1987b; Wickens, 1989). With this decomposition, the correlation model is

$$P_{ij} = P_{i.}P_{.j}(1 + \sum_{m=1}^{M} \lambda_m x_{im} y_{jm}) \tag{3.8}$$

(Goodman, 1985, 1986, 1991; Gilula & Haberman, 1986, 1988; Greenacre, 1984; Wickens, 1989). This model is saturated and fits perfectly when $M = \min(I, J) - 1$, but it often provides an adequate fit for small values of $M$, such as $M = 1, 2$, or 3 (for examples, see Gilula & Haberman, 1986, 1988; Goodman, 1985, 1986, 1987b; Greenacre, 1984).

To estimate the parameters in the above model, constraints analogous to those used to estimate the $RC(M)$ association model are used (i.e., replace the $\mu$'s and $\nu$'s in 3.6 by $x$'s and $y$'s, respectively). With the correlation model, the weights are almost always set equal to the marginal probabilities: $g_i = P_{i.}$ and $h_j = P_{.j}$ (Goodman, 1985, 1986; Gilula & Haberman, 1986, 1988). The derivation and interpretation of the model depends critically on this choice[2] (For more on the effect of weights used in the identifications constraints on the estimates of the parameters, see Goldstein, 1987, and Healy & Goldstein, 1976).

For each dimension $m$, $\lambda_m$ is the correlation between the row and column scores (i.e., $\lambda_m = \sum_i \sum_j x_{im} y_{jm} P_{ij}$). These scores are uncorrelated across dimensions (i.e., $\sum_i x_{im} x_{im'} P_{i.} = \sum_j y_{jm} y_{jm'} P_{.j} = \sum_i \sum_j x_{im'} y_{jm} P_{ij} = 0$ for $m \neq m'$). Assuming that the $\lambda_m$'s are ordered from largest to smallest ($\lambda_1 \geq \ldots \geq \lambda_M$), each $\lambda_m$ is the largest possible correlation between the row and column scale values on dimension $m$, such that $\{x_{im}\}$ and $\{x_{im'}\}$ are uncorrelated, and $\{y_{jm}\}$ and $\{y_{jm'}\}$ are uncorrelated for all $m < m'$ (Greenacre, 1984; Goodman, 1985, 1986). Assuming that the rows and columns are stochastically ordered, the correlation coefficient can also be interpreted as a measure of stochastic order extremity (Gilula, Krieger, & Ritov, 1988). The sum over $m$ of the squared correlation coefficients, where $M = \min(I, J) - 1$, is proportional to $X^2$

$$\frac{1}{N} X^2 = \sum_{m=1}^{M} \lambda_m^2$$

[2]One exception is "asymmetric correspondence analysis" (Lauro & D'Ambra, 1984; D'Ambra & Lauro, 1989; Lauro, 1989) where the categories of the response or dependent variable have unit weights and the categories of the other variable are weighted by their marginal probabilities. This weighting scheme leads to a decomposition of Goodman-Kruskal $\tau$ rather than $X^2$.

In other words, the goodness of fit statistic $X^2$ can be partitioned into $M = \min(I, J) - 1$ components (Goodman, 1985, 1986). In the correspondence analysis literature, this sum is referred to as the "total inertia" in a table.

Equation 3.8 has been referred to as the canonical correlation model (Gilula & Haberman, 1986, 1988), the maximum correlation model (Wickens, 1989), and when $M = 1$, the $RC$ correlation model (Goodman, 1985, 1986). As with the association model, the notation $RC(M)$ is used here to indicate that it is a two–way (row–column) model with $M$ dimensions. The parameters of the $RC(M)$ correlation model can be estimated either by maximum likelihood, which involves iterative procedures, or least squares, which involves finding the singular value decomposition of some matrix.

Equation 3.8 is the mathematical model underlying the technique of correspondence analysis (Greenacre, 1984; Goodman, 1985, 1986). Least squares estimation is always used in correspondence analysis (CA), which is often presented as a geometric, exploratory or descriptive technique rather than as a model of data. A number of different approaches and problems encountered in analyzing categorical data have all led to the technique of correspondence analysis. The technique has been referred to as canonical correlation analysis of categorical data, dual (or optimal) scaling, reciprocal averaging, and simultaneous linear regressions (Nishisato, 1980; Greenacre, 1984). Greenacre (1984) gives a history of the development of CA and a description of the different approaches. Examining the different approaches that lead to correspondence analysis provides a better understanding of uses and interpretations of the $RC(M)$ correlation model. Since the

mathematics of CA and the different problems that motivated the development of CA are fairly well known, they are reviewed in Appendix A rather than in this chapter.

While the $RC(M)$ association model is related to models for latent continuous variables, the $RC(M)$ correlation model is related to models for latent categorical variables. Two such models are the latent class model (Gilula, 1984; Goodman, 1987a, 1987b; McCutcheon, 1987) and the latent budget model (de Leeuw & van der Heijden, 1988; van der Heijden, Mooijaart & de Leeuw, 1989; de Leeuw, van der Heijden & Verboon, 1989). These two models are equivalent to each other. They yield the same predictions and have the same number of degrees of freedom (van der Heijden, Mooijaart & de Leeuw, 1989). The only difference between them is their parameterizations; hence, only the latent class model for 2–way tables is presented here. The latent class model is

$$P_{ij} = \sum_{l=1}^{L} \pi_l \pi_{i|l} \pi_{j|l}$$

where $L$ equals the number of latent classes, $\pi_l$ equals the probability of latent class $l$, and $\pi_{i|l}$ and $\pi_{j|l}$ equal the conditional probabilities that $A = a_i$ and $B = b_j$, respectively, given latent class $l$. Since the parameters of the model are probabilities, they must be non-negative and sum to one (i.e., $\sum_k \pi_l = \sum_i \pi_{i|l} = \sum_j \pi_{j|l} = 1$).

Under certain circumstances the correlation model is equivalent to the latent class model (Gilula, 1984; Goodman, 1987a, 1987b; de Leeuw & van der Heijden, 1991). The $RC(1)$ correlation model always implies the rank two ($L = M + 1 = 2$) latent class model and visa versa (de Leeuw & van der Heijden, 1991). Gilula (1984) give the relationship

between the parameters of these two models. For $M \geq 2$ ($L \geq 3$), the latent class model always implies the $RC(M)$ correlation model, but the $RC(M)$ correlation model does not imply the rank $M+1$ latent class model (de Leeuw & van der Heijden, 1991). Gilula (1984) gives the conditions under which the $RC(M)$ correlation model is equivalent to the latent class model.

Unlike the $RC(M)$ association model, the $RC(M)$ correlation model is not defined for all probabilities matrices (Goodman, 1985; de Leeuw & van der Heijden, 1991). The $RC(M^*)$ correlation model is defined and yields fitted values in the range of 0 to 1 whenever

$$-1 \leq \sum_{m=1}^{M^*} \lambda_m x_{im} y_{jm} \leq \frac{(1 - P_{i.}P_{.j})}{P_{i.}P_{.j}} \tag{3.9}$$

where $1 \leq M^* < \min(I, J) - 1$. For $M^* = 1$, the model yields non-negative fitted values whenever

$$\frac{-1}{\mid \lambda \mid} \leq x_i y_j \tag{3.10}$$

(Goodman, 1985). In tables where there is a relatively strong relationship between the variables, the estimated correlations $\lambda_1, \ldots, \lambda_M^*$ will tend to be "large." The larger the correlations, the more likely it is that condition 3.9 (or 3.10) will be violated and the model not defined. The $RC(M)$ correlation will also tend to yield improper solutions when some of the cells have small frequencies.

#### 3.2.2.2 The Peer Play Data: the first session

As an example, the $RC(M)$ correlation model was fit to the Group $\times$ Play sub-table analyzed previously. The $RC(M)$ correlation model could not be fit to the table using unrestricted maximum likelihood estimation (MLE), because the model is not defined for this table. The least squares estimates of the parameters of the $RC(M)$ correlation model show that the above conditions (3.9) were violated such that the model yields negative fitted values for some cells. For example, using the least squares estimates for MOH ($i = 8$) and negative emotion ($j = 9$), $\tilde{x}_8\tilde{y}_9 = -5.206$ and $1/\tilde{\lambda} = -1/.2172 = -4.604$, which violates condition 3.10. The tilde's indicate that these are least squares estimates rather than maximum likelihood estimates. For $M = 2$, $\sum_{m=1}^{2} \tilde{\lambda}_m \tilde{x}_{8,m}\tilde{y}_{9,m} = -1.125$, which violates condition 3.9. The $RC(M)$ correlation model also yields negative fitted values for the Group $\times$ Play sub-tables for $N = 25$ and 29, which do not include MOL, and for the sub-table that does not include both MOL and MOH.

Since the $RC(M)$ correlation model is not defined for the Group $\times$ Play sub-table or even variations of it, when unrestricted MLE is used to fit the model, the solutions will not converge. An alternative in such cases is to use restricted MLE where the fitted values are constrained to be non-negative (Gilula & Haberman, 1986; Goodman, 1987b). With restricted MLE, the model will not fit the data as well as it could have if the fitted values could be negative. Rather than resorting to restricted MLE or using an improper solution, a better strategy is to use the $RC(M)$ association model instead of the correlation model.

**Table 3.3**: Correspondence analysis of the Group × Play sub-table from the first occasion ($N = 30$).

| $m$ | $\lambda_m$ | $\lambda_m^2$ | $X^2$ | Percent | Cumulative |
|---|---|---|---|---|---|
| 1 | .2172 | .0472 | 63.540 | 52.13% | 52.13% |
| 2 | .1435 | .0206 | 27.750 | 22.77% | 74.90% |
| 3 | .1057 | .0112 | 15.040 | 12.34% | 87.24% |
| 4 | .0736 | .0054 | 7.302 | 5.99% | 93.23% |
| 5 | .0540 | .0029 | 3.923 | 3.22% | 96.45% |
| 6 | .0498 | .0025 | 3.343 | 2.74% | 99.19% |
| 7 | .0272 | .0007 | 0.998 | .82% | 100.00% |
| total | | .0905 | 121.896 | | |

For the sake of comparison, the results of the least squares fit of the $RC(M)$ correlation model (i.e., correspondence analysis) to the Group × Play sub-table for $k = 1$ and $N = 30$ are reported in Table 3.3. The first and second dimensions account for 52.12% and 22.77%, respectively, of the total dependency in the table as measured by Pearson's $X^2$ for testing independence. This is approximately the same as that for the first and second dimensions of the association model, which accounted for 52.19% and 20.24%, respectively, of the dependency as measured by $G^2$ for testing independence.

The standard practice in correspondence analysis is to plot the scale values from the first two dimensions. The plot is presented and interpreted as a graphical representation of the dependency in the table (for more details, see Appendix A). The reason that two dimensions are usually examined is that two dimensional plots are easier to look at than higher-dimensional ones. Equally rare are one dimensional plots. For the above example, the justification likely to be given for choosing two dimensions would be that the first

two dimensions account for 74.90% of the total association in the table, and this amount is "large" enough.

The estimated scale values for the groups and play qualities on the first two dimensions are plotted in Figures 3.4 and 3.5. The same normalization is used here as was used in Figures 3.2 and 3.3, except here the scale values were multiplied by $(\tilde{\lambda}_m)^{1/2}$. The relative "importance" of the dimension in terms of how much of the interia (dependency) in the table is accounted for by the dimension is reflected by the size of the normalized scale values on the dimensions. The range of the first dimension is approximately twice as large as that of the second dimension, as expected from the fact that the first dimension accounts for approximately twice as much of the total interia as the second dimension (i.e., 52.13% versus 22.77%). The relative distance of points from the origin reflect the contribution of the category to the lack of independence in the table.

With respect to groups, the scale values from the $RC(2)$ correlation and association models have nearly the same order and relative spacing on the first dimension, but show more differences on the second dimension. In the correlation model, MOL is has the smallest (largest negative) value, and the groups MYH, FOH, and MOH have scale values that are approximately half the size of the MOL scale value. In the $RC(2)$ association model, the scale value for FOH is the smallest, and the scale values for the groups MOL, MOH, and MYH are only slightly smaller.

With respect to the play qualities, the most noticeable differences between the scale values of the correlation and association models involve the qualities amity (amt), unsustain (uns), gossip (gos), and prohibit (proh). On the first dimension, the correlation

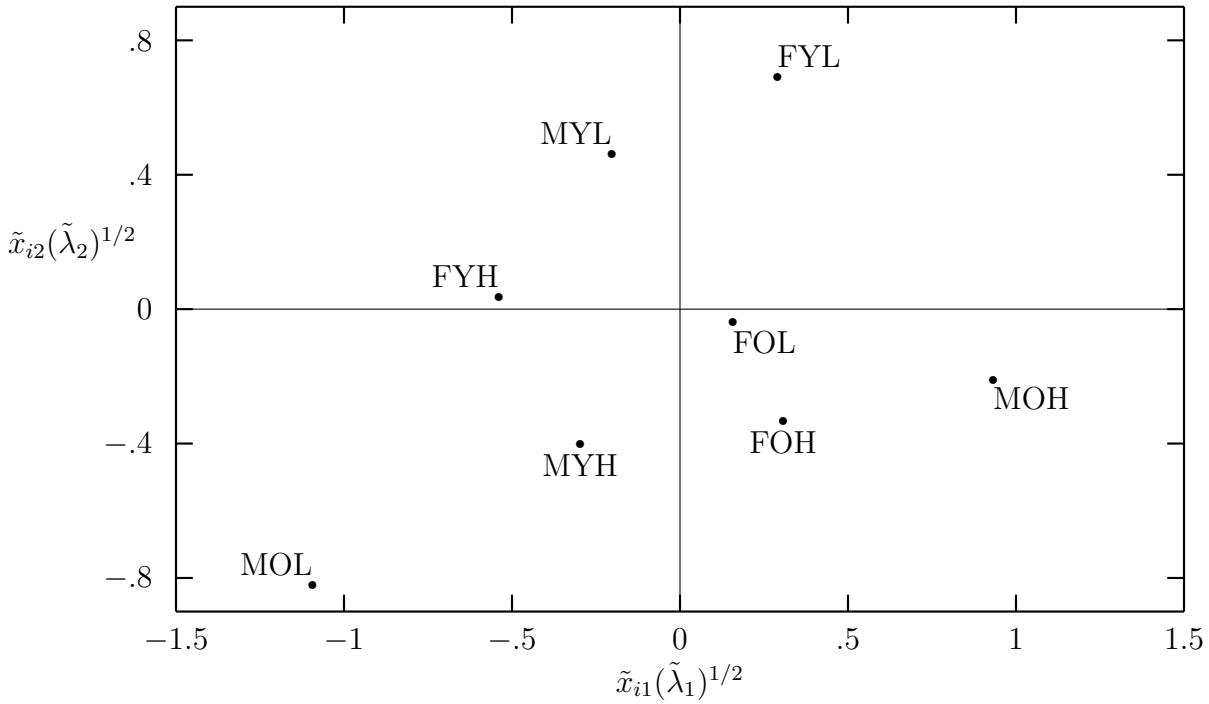



**Figure 3.4**: Group scale values on the first two dimensions of the correspondence analysis of Group × Play sub-table from the first occasion ($N = 30$).

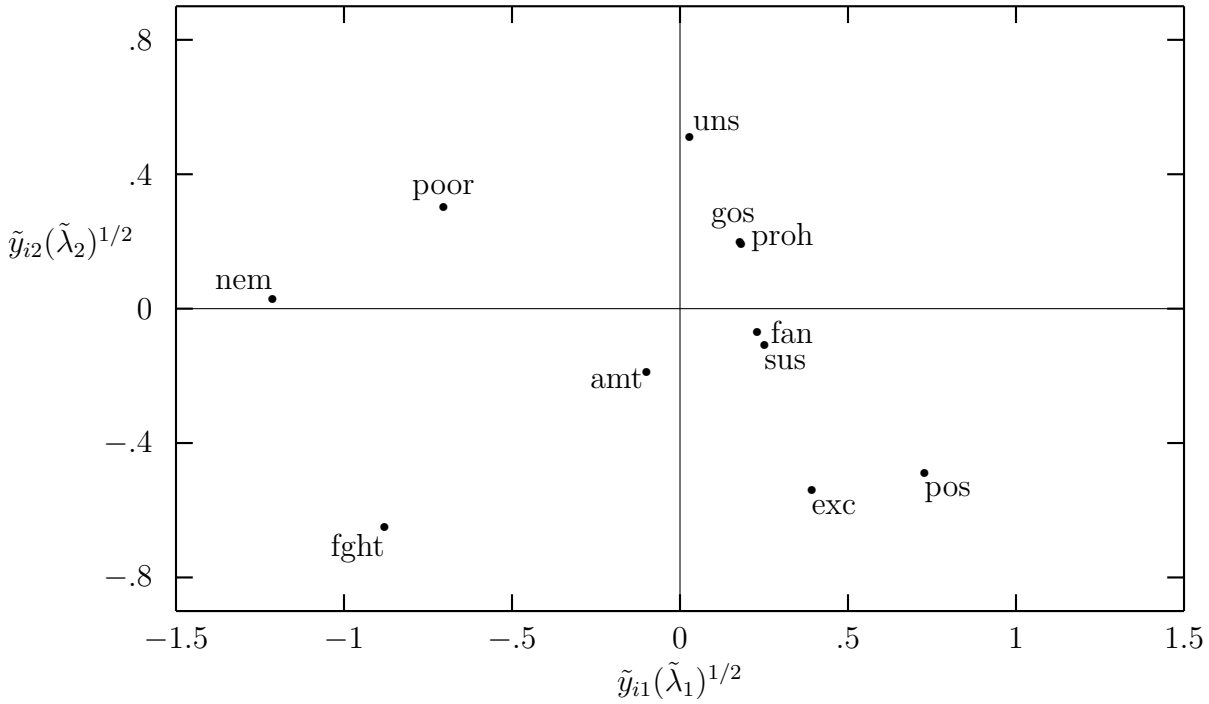



**Figure 3.5**: Play scale values on the first two dimensions of the correspondence analysis of Group × Play sub-table from the first occasion ($N = 30$).

model scale values for amity and unsustained play are closer to zero and relatively further from gossip, prohibit, fantasy and sustained play than the corresponding association model scale values. The correlation model scale values for gossip and prohibit on both of the dimensions are nearly equal while the association model scale values for prohibit are larger on both dimensions than the values for gossip.

The $RC(M)$ correlation model scale values for the groups and play qualities are very similar to those from the $RC(M)$ association models. This need not be the case. The scale values from the $RC(M)$ association and correlation models are expected to differ, because the two models use different definitions of interaction, and different weights in the identification constraints required to estimate the model parameters (i.e., unit weights for the association model and marginal probabilities for the correlation model). With respect to weights, the further that the marginal probabilities deviate from a uniform distribution, the greater the possibility that there will be differences between the scale values from the association and correlation models. For example, the group MOL has a much smaller marginal probability than any of the other groups (i.e., the marginal probability for MOL equals .04, while the marginal probabilities for the other groups range from .07 to .19), and one of the most noticeable differences between the scale values from the two models involved this group. Although not presented here, the parameters of the $RC(2)$ association model in which the marginal probabilities are used instead of unit weights to identify the parameters were plotted. These plots were in between Figures 3.2 and 3.3 and Figures 3.4 and 3.5 in terms of similarity. The reason for the observed

similarity between the estimated scale values of the $RC(M)$ association and correlation models is discussed further in the next section.

### 3.2.3 Comparison of Association and Correlation Models

For 2–way tables, the $RC(M)$ correlation and association models are compared here both theoretically and empirically. The empirical comparison is made with respect to the models fit to the data in the two previous sections. In the case of complete independence, the models are equivalent and $\phi_m = \lambda_m = 0$. If the association between the rows and columns is "small" or the variables are "close" to independent, then the correlation and association models are approximations of each other (Goodman, 1985, 1986; van der Heijden, de Falugerolles & de Leeuw, 1989). As far as determining how small is "small," Escoufier (van der Heijden, de Falugerolles & de Leeuw, 1989) showed that if $z = \sum_m \lambda_m x_{im} y_{jm}$ is small relative to 1, then $\log(1 + z) \approx z$, which implies that $\log(1 + \sum_m \lambda_m x_{im} y_{jm}) \approx \sum_m \lambda_m x_{im} y_{jm} \approx \sum_m \phi_m \mu_{im} \nu_{jm}$. Another way to see this result involves applying the power series expansion, $e^x = 1 + x + x^2/2! + x^3/3! + \ldots$, to the $RC(M)$ association model, which yields

$$P_{ij} = \alpha_i \beta_j (1 + \sum_m \phi_m \mu_{im} \nu_{jm} + \frac{(\sum_m \phi_m \mu_{im} \nu_{jm})^2}{2!} + \ldots)$$

(Wickens, 1989). For "small" values of $\sum_m \phi_m \mu_{im} \nu_{jm}$, only the first one or two terms of the expansion are needed for the power series expansion to be a good approximation of the $RC(M)$ association model, in which case $\sum_m \phi_m \mu_{im} \nu_{jm} \approx \sum_m \lambda_m x_{im} y_{jm}$. In

the example of the Group × Play sub-table presented in the two previous sections, the interaction in this table is "small" enough for the $RC(M)$ association and correlation models to roughly approximate each other, which partially explains why the scale values from the two models were so similar.

While the $RC(M)$ association and correlation models are approximations of each other when the dependency between the variables is small, it should be emphasized that the estimated parameters from the two different type of models can be very different and yield different pictures of the relationship among categories of each variable and of the relationship between the row and column variables (Clogg, 1986; Goodman, 1985, 1991). In the extreme case of perfect association between rows and columns (complete mutual dependence where all the conditional probabilities $P_{i|j}$ and $P_{j|i}$ equal 0 or 1), the $RC(1)$ association model fits perfectly with $\phi_1 \to \infty$, but the $RC(M)$ correlation model requires $M = \min(I, J) - 1$ (Gilula, Krieger & Ritov, 1988). In cases of very strong association, the $RC(M)$ association model provides a more parsimonious description of the relationship between the rows and columns of a contingency table than the $RC(M)$ correlation model, provided that the latter exists.

Since the correlation model is an additive model, it is not defined for all possible tables with consistent marginal probabilities (i.e., marginal probabilities that sum to 1). This problem was encountered in the previous section when attempts to fit it to the Group × Play sub-table for $k = 1$ and $N = 25, 29$, and 30. Even when both of the groups MOL and MOH were deleted, the $RC(M)$ correlation model still did not yield a proper solution. Condition 3.9 was violated. When condition 3.10 is violated or almost violated,

the association and correlation models can give quite different results (Goodman, 1985); however, there are no practical guidelines for assessing when this condition is "almost violated." In the example, the conditions were violated, but the models still gave similar results. This probably results from the fact that the amount of the dependency was not too "large." Restricted MLE could be used to fit the $RC(M)$ correlation model in such cases, but the fit will be worse than the fit of the $RC(M)$ association model. When least squares estimation is used, statistical tests cannot be performed and standard errors of the parameters cannot be estimated. Since the $RC(M)$ association model is a multiplicative model, it always exists for any consistent table and negative fitted values will never be a problem for this model.

If a contingency table arises from either an underlying discretized joint bivariate normal distribution or a joint bivariate distribution where separate transformations of the row and columns scores result in a bivariate normal distribution, the $RC(1)$ association and correlation models are approximately equal; namely, $\mu_i \approx x_i$, $\nu_j \approx y_j$, and $\phi \approx \lambda$ (Goodman, 1981a, 1985). While the models approximate each other under these circumstances, Goodman (1985) also states that the association model provides a better approximation of the underlying distribution than the correlation model. Even when a bivariate normal distribution does not underlie the data, $\phi$ is still a meaningful measure of "intrinsic association" in the underlying distribution, but the correlation between $\mu_i$ and $\nu_j$ is not as meaningful (Goodman, 1985). Since there is no reason to suppose that a bivariate normal distribution underlies the Group $\times$ Play sub-table, the $RC(M)$ association model is preferred to the correlation model.

## 3.3 Summary

In the first section, the ways that multiplicative and additive models define and measure interactions in contingency tables were reviewed. Some of the general properties of these models where also described. Multiplicative and additive models are equivalent with respect to complete and joint independence, but differ with respect to conditional independence. Properties that are characteristic of multiplicative models, but not additive models, include interactions defined as odds ratios computed on sub-tables, partitions of $G^2$, sub-table invariance, and always defined (i.e., no negative fitted values). Properties that are characteristic of additive models, but not multiplicative models, include interactions defined as correlations between pairs of variables, orthogonal partitions of $X^2$, and amalgamation invariance. Considering these properties, multiplicative models are preferable to additive models for analyzing the peer play data.

In section 3.2, the $RC(M)$ association and correlation models were presented as extensions of saturated multiplicative and additive models, respectively, where the interaction terms are decomposed into the sum of bilinear terms. These models, which represent the dependency between rows and columns in contingency tables, were fit to the Group $\times$ Play sub-table for $k = 1$ and $N = 30$ and the results were compared.

Theoretical and empirical reasons for preferring the $RC(M)$ association over the correlation model were discussed. Since loglinear models were used in the preliminary analyses of the play data in Chapter 2, a natural approach is to use extensions of them to further analyze the interactions.

Considering the desired properties of measurements of interactions, preference is given to the $RC(M)$ association model, which models odds ratios. Desired analyses of the data include comparisons of interactions across groups of children and across time. Since odds and odds ratios do not depend on the margins of tables, these quantities are better measures of the strength of interactions than correlations, which do depend on the margins. The measured dependency should not differ because there are different numbers of children in the groups and/or different numbers of observations at different times. Unlike correlations, odds ratios readily generalize to higher-way tables. The peer play data set is actually the 3–way, Group $\times$ Play $\times$ Time table. The dependency *among* the ways or variables of this table needs to be measured and represented. With correlations, we would be limited to examining dependency between pairs of variables in the margins of the full table.

The fundamental interpretation of the $RC(M)$ association model does not depend on the choice of the weights in the identification constraints required to estimate the parameters of the model, but the basic interpretation of the $RC(M)$ correlation model does depend on them. With the $RC(M)$ correlation model, marginal probabilities must be used in the identification constraints; however, with the $RC(M)$ association model, unit, uniform, marginal probabilities, or any other scheme may be used. The choice of unit weights with the $RC(M)$ association model for the play data allows comparisons of the dependency across time and/or groups without confounding the effects of the margins and the dependency in the table.

Another reason for preferring the $RC(M)$ association model over the correlation model is the property of sub-table invariance, which is a characteristic of a multiplicative models. Sub-table invariance is desirable whenever there is missing data. The basic results of the analyses should not depend on whether the whole table is analyzed of just part of the table.

The row–column models presented in this chapter are most readily applicable to 2–way tables, which is why only a 2–way sub-table of the play quality data was analyzed here. In the case of 2–way tables where there is a strong association, the $RC(M)$ association model may require fewer dimensions than the $RC(M)$ correlation model to adequately represent the relationship between the rows and columns. While the association in the 2–way Group $\times$ Play sub-table does not appear to be strong enough to make a difference in terms of parsimony, extensions of multiplicative models for higher-way tables (e.g., generalizations of $RC(M)$ association models) are likely to be more parsimous than extensions of additive models (e.g., generalizations of $RC(M)$ correlation models). The multiplicative models of conditional independence and no 3–way interaction may be sufficient to model the peer play data.

Given the properties of models reviewed in this chapter, multiplicative models (i.e., loglinear models), extensions of them (i.e., the $RC(M)$ association model), and generalizations of them are definitely better for analyzing the peer play data than additive models, including extensions and generalizations of them (i.e., the $RC(M)$ correlation model). Therefore, in the following chapters, attention is focused primarily on the $RC(M)$ association model and generalizations of it to higher–way tables.

# Chapter 4

# TWO-WAY REDUCTION METHODS

In the previous chapter, the $RC(M)$ association model was presented as an extension of a loglinear model for 2–way tables; however, the peer play data consist of a 3–way table. One approach to analyzing higher–way tables, such as the peer play data, is described in this chapter. Models are presented that reduce the problem to that of analyzing 2–way tables. Multiway tables are rearranged into 2–way tables, which can be analyzed by the $RC(M)$ association model. While the $RC(M)$ correlation model can also be used to analyze re-coded tables, only the association model is explicitly considered here based on the reasons discussed in Chapter 3 for preferring multiplicative models to additive models.

Marginal and joint tables are described and the analyses of such tables by $RC(M)$ association models are presented in Sections 4.1 and 4.2, respectively. Studying both

marginal and joint tables was proposed by van der Heijden and de Leeuw (van der Heijden, 1987; van der Heijden & de Leeuw, 1987, 1989) as the basis of a procedure for the analysis of longitudinal, categorical data. While their approach uses correspondence analysis, the $RC(M)$ association model is used here. In Section 4.3, conditional tables and their analyses are discussed. In Sections 4.1, 4.2 and 4.3, $RC(M)$ association models are fit to the peer play data formatted as 2–way matrices to illustrate both the usefulness and the limitations of these reduction methods for the analysis of longitudinal categorical data.

The data analyzed in this chapter and the next two chapters are from the sub-sample of children who were present at all five sessions (i.e., $N = 25$ subjects, frequencies in Table 2.2). This choice was made for practical and theoretical reasons. While only the results of the analyses of the 25 child sub-sample are reported in detail here and in the 2 next chapter, most of the analyses were also performed on the 29 and 30 child samples. In Chapter 6, the issue of missing data and its effect on the analyses is examined. It is worth noting here that the same basic "story" is told regardless of which sub-sample of the data is analyzed.

The last section of this chapter, Section 4.4, contains a discussion of the inadequacy of 2–way decompositions for representing 3–way interactions. The marginal, joint and conditional reduction strategies presented in this chapter are summarized and compared. To deal with some of the drawbacks and limitations of the reduction techniques, various constraints and restrictions on model parameters can be imposed. However, with such restrictions, the terms representing interactions are no longer just simple 2–way decom-

positions, but are similar to particular cases of model generalizations discussed in the following chapters.

## 4.1 Marginal Analyses

In Section 4.1.1, the marginal approach is described, and the implications of fitting $RC(M)$ association models marginal tables and the limitations of the method are discussed. In Section 4.1.2, the marginal analyses of the peer play data are presented.

### 4.1.1 The Marginal Approach

The "marginal approach," which is the simplest 2–way reduction method, consists of analyzing the 2–way marginal tables of a multiway table. Separate $RC(M)$ association models can be fit to each of the 2–way marginal tables. This approach results in as many models as there are 2–way margins. The number of 2–way marginal tables that can be formed from an $n$–way table is $\binom{n}{2}$. For example, a 3–way, $(I \times J \times K)$ contingency table of variables $A$, $B$, and $C$ has three 2–way marginal tables: an $(I \times J)$ table containing the probabilities $P_{ij.}$, an $(I \times K)$ table containing $P_{i.k}$, and a $(J \times K)$ table containing $P_{.jk}$. Tables containing probabilities are denoted by $\mathbf{P}$ and those containing frequencies are denoted by $\mathbf{F}$. A specific marginal table is denoted by (for example) $\mathbf{P}_{AB+}$, where the subscripts indicate which 2–way margin is contained in the table (i.e., $AB$) and which variable has been summed over.

The marginal method is severely limited in that only 2–way marginal associations are analyzed. Three– and higher–way associations are lost by reducing a multiway table to a set of 2–way marginal tables. Even when only 2–way associations are present in a multiway table (i.e., 3– and higher–way associations are not present), a marginal analysis may not yield the most parsimonious representation possible. Analyzing a set of marginal tables yields multiple sets of scales values for the categories of each of the variables such that different scale values are assigned to the same categories depending on which 2–way margin is analyzed.

### 4.1.2 Marginal Analysis of the Peer Play Data

The marginal approach is illustrated here by applying it to the peer play data. With the peer play data, there are two interesting 2–way interactions, the group by play quality ($GP$) and the play quality by time ($PT$) interactions. Since the $PT$ partial and marginal associations are not important (see Table 2.4, $df = 40$, $G^2 = 40.90$ and $G^2 = 39.82$, respectively) but the $GP$ partial and marginal associations are important ($df = 60$, $G^2 = 126.90$ and $G^2 = 126.02$, respectively), $RC(M)$ association model is fit only to the $GP$ margin.

The results from fitting the $RC(M)$ association model for $M = 0, 1$, and 2 to the $GP$ marginal table $\mathbf{F}_{GP+}$ are reported in Table 4.1. Both the one and two dimensional models appear to fit the data quite well ($G^2 < df$). For $M = 1$, the estimated intrinsic association equals 1.761, while for $M = 2$, the estimated values are $\hat{\phi}_1 = 1.821$ and $\hat{\phi}_2 = .000$. Since $\hat{\phi}_2$ equals zero to at least three decimal places, the second dimension

**Table 4.1**: Fit of the $RC(M)$ association model to the $\mathbf{F}_{GP+}$ marginal table ($N = 25$).

| $M$ | $df$ | $G^2$ | $G^2/df$ | (p-value) | $X^2$ | (p-value) |
|---|---|---|---|---|---|---|
| 0 | 60 | 126.07 | 2.10 | (.000) | 122.50 | (.000) |
| 1 | 45 | 43.68 | .97 | (.528) | 43.46 | (.537) |
| 2 | 32 | 19.58 | .61 | (.958) | 19.53 | (.959) |

does not appear to be very important and it is assumed that this two dimensional model "overfits" the data.

The estimated group ($\hat{\mu}_i$) and play quality ($\hat{\nu}_j$) scale values of the $RC(1)$ association model are plotted in Figure 4.1. The group scale is primarily an age dimension. All of the groups containing younger children have negative scale values and all of the groups containing the older children have positive ones. The most extreme groups are MOH (positive) and MYH (negative). The scale values for three groups MYL, FYL, and FYH are all very close to each other, while the rest of the groups are fairly well spread apart. The scale values for the play qualities are ordered from immature or "bad" to mature or "good." Of the "bad" qualities, negative emotion (nem), fight, and poor play have negative scale values while unsustain (uns) and prohibit (prh) have small positive values relatively close to zero. All of the "good" qualities have positive values. The qualities with the largest negative and positive values are negative emotion and positive play, respectively. On substantive grounds, successful fantasy play (fan) and sustained play (sus) should be close to each other. This expectation is met with the scale values for fantasy and sustained play being closer to each other than to any of the other qualities. Overall, Figure 4.1 shows that the nature of the dependency between groups and play

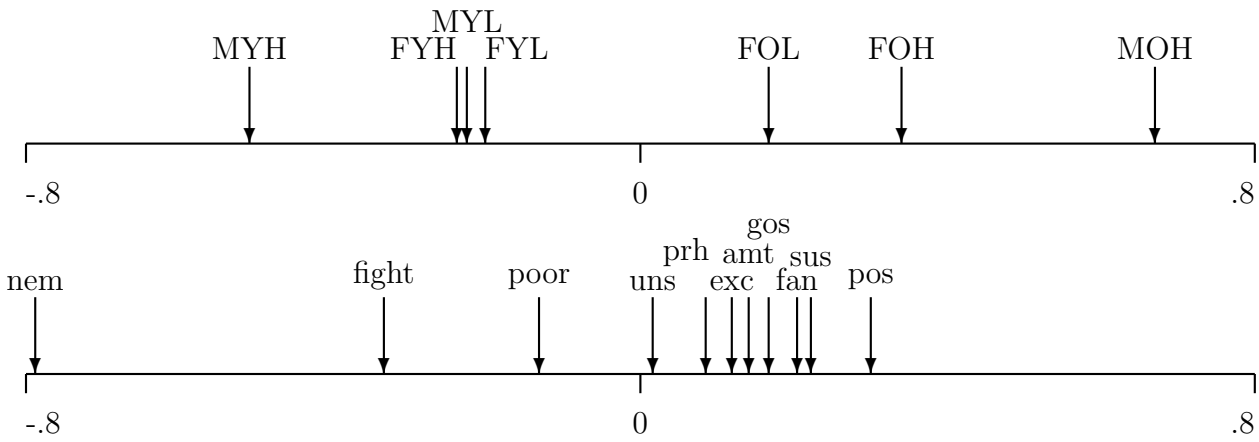



**Figure 4.1**: Estimated scale values for groups and play qualities of the $RC(1)$ association model with unit weights fit to the $\mathbf{F}_{GP+}$ marginal table ($N = 25$).

qualities in the marginal table is such that the younger children tend to exhibit more of the immature qualities, in particular, negative emotion and fighting, while the older children tend to exhibit more of the mature qualities, especially positive play.

The interpretation of the Group$\times$Play margin is substantively nice and neat; however, this analysis does not provide any information about the effect of time and completely ignores the 3–way interaction between groups, play, and time. The next method of rearranging multiway tables to 2–way tables does not suffer this limitation.

## 4.2 Joint Analyses

Rearranging multiway tables into "joint" tables and the application of $RC(M)$ association models to such tables is described in the next section, Section 4.2.1. In Section 4.2.2, van der Heijden and de Leeuw's (1987) approach for analyzing longitudinal data by analyzing both marginal and joint tables is reviewed and modified for the peer play data. The method is illustrated in Section 4.2.3 where the joint analysis of the peer play data is presented. Analyzing joint tables with linear restrictions are imposed on model parameters is an alternative to analyzing both marginal and joint tables. Linear restrictions are discussed in the last section, Section 4.2.4.

### 4.2.1 The Joint Approach

The "joint" method for rearranging multiway tables into 2–way tables treats multiple variables as single polytomous variables. In this method, the variables must first be split into two types. The combinations of the categories of the variables in one type become the rows of the 2–way, rearranged or "flattened" table, and the combinations of the categories of the variables in the other type become the columns. When combinations of the categories of two or more variables are treated as the categories of a single polytomous variable, the variables are said to be "interactively" or "factorialy" coded.

For 3–way tables, an alternative way to describe the construction of these tables only starts by considering the data in the form of a 3–way matrix or cube. The data are rearranged to form a 2–way matrix by horizontally or vertically concatenating or

"stacking" 2–way matrices or "slices" of the 3–way matrix, which leads some researchers to refer to these tables as "stacked" tables (e.g., van der Heijden, 1987; van der Heijden & de Leeuw, 1989). In this thesis, the term "joint" is preferred to "stacked," because it emphasizes that if the rows and columns of such a table are independent, then the variables in the rows are *jointly* independent of the variables in the columns. A specific table will be denoted by (for example) ${}_A\mathbf{P}_{BC}$, where the subscript preceding $\mathbf{P}$ indicates which variable(s) corresponds to the rows and the subscript following $\mathbf{P}$ indicates which variables correspond to the columns.

In the example of the 3–way, $(I \times J \times K)$ table of variables $A$, $B$, and $C$, one of the three possible joint tables is the $(I \times JK)$ table of $A$ versus $BC$,

$$ {}_A\mathbf{P}_{BC} = [\mathbf{p}_{11}\ \mathbf{p}_{12}\ \cdots \mathbf{p}_{JK}\ ] \tag{4.1} $$

where each $\mathbf{p}_{jk}$ is an $(I \times 1)$ vector containing elements $\{P_{ijk}, i = 1, \ldots, I\}$. The other two possible tables are the $(J \times IK)$ table of $B$ versus $AC$, ${}_B\mathbf{P}_{AC}$, and the $(K \times IJ)$ table of $C$ versus $AB$, ${}_C\mathbf{P}_{AB}$. The transposes of joint tables are not distinct, since, for example, ${}_C\mathbf{P}'_{AB} =_{AB} \mathbf{P}_C$, which has no effect on analyses.

No information is lost by interactively coding variables. The entries in joint tables are the probabilities or frequencies of the original multiway table (e.g., $P_{ijk}$ or $F_{ijk}$). The distinct 2–way tables that can be formed by interactively coding variables of a multiway table differ only with respect to how the elements are arranged. The interactions between the row and column variables are emphasized over the interactions between the variables

within the rows or columns (i.e., the variables that have been interactively coded). If the $RC(M)$ association model is applied to the table ${}_A\mathbf{P}_{BC}$ defined in equation 4.1, then the cell probabilities are modeled as

$$P_{ijk} = \alpha_i \beta_{jk} \exp(\sum_{m=1}^{M} \phi_m \mu_{im}^{A} \nu_{(jk)m}^{BC})$$

or, in terms of a log-bilinear model,

$$\ln(P_{ijk}) = u + u_i^A + u_j^B + u_k^C + u_{jk}^{BC} + \sum_{m=1}^{M} \phi_m \mu_{im}^{A} \nu_{(jk)m}^{BC} \tag{4.2}$$

where $M \leq \min(I, JK) - 1$. Superscripts have been added to the scale values to emphasize which variables or combination of variables they represent, and parentheses are placed around the categories of the variables interactively coded.

Model 4.2 can be thought of as decomposing the lack-of-fit of the joint independence model $(A, BC)$. The combined effects of the $AB$, $AC$ and $ABC$ interactions are represented by sum of the bilinear terms, and the $BC$ interaction is taken into account by $u_{jk}^{BC}$ in model 4.2. The effect of the $BC$ interaction is not reflected in the row or column scale values of the bilinear terms, and the $BC$ margin is fit perfectly in these models.

The identification constraints used to estimate the parameters of the joint $RC(M)$ association model are the same as those used for the simple $RC(M)$ association model, except that the weights and scale values for interactively coded variables have extra subscripts. For example, the constraints used to estimate the parameters in model 4.2

are

$$\begin{aligned} \sum_i \mu^A_{im} g_i &= \sum_j \sum_k \nu^{BC}_{(jk)m} h_{(jk)} = 0 \\ \sum_i \mu^A_{im} \mu^A_{im'} g_i &= \sum_j \sum_k \nu^{BC}_{(jk)m} \nu^{BC}_{(jk)m'} h_{(jk)} = \delta_{mm'} \end{aligned} \tag{4.3}$$

where the weights $g_i$ and $h_{(jk)}$ are known and fixed.

### 4.2.2 Longitudinal Data

When one mode of a 3–mode table corresponds to time, not all three ways of forming joint tables are worth analyzing. Van der Heijden (1987) and van der Heijden and de Leeuw (1989) describe an approach for analyzing longitudinal categorical data by correspondence analysis that involves analyzing particular joint and marginal tables. The two joint tables of interest are the ones where time (occasions) and variables are interactively coded and time and individuals (or groups) are interactively coded. The two marginal tables of interest are the two margins involving the time. In their approach, they first use loglinear models to identify which interactions are important (i.e., those present in the data), and then use correspondence analysis to analyze them. This basic approach is applied here to the peer play data, but rather than correspondence analysis or the $RC(M)$ correlation model, the $RC(M)$ association model is used to analyze the rearranged data. This approach is similar to "generalized" correspondence analysis (van der Heijden & Meijerink, 1988) or the "combined approach" of contingency table analysis where correspondence analysis is used to study the interactions of loglinear models (van

der Heijden & de Leeuw, 1985; van der Heijden & Worsley, 1988; van der Heijden, de Falugerolles & de Leeuw, 1989).

Before describing the details of the approach for analyzing longitudinal categorical data by fitting the $RC(M)$ models to marginal and joint tables, two slight differences between the peer play data and the data considered by van der Heijden (1987) should be noted. First, van der Heijden's (1987) data consist of a 3–way matrix (cube) where rows correspond to individuals, columns correspond to categories of variable(s), and layers correspond to occasions. The entries in this matrix are 0/1 indicator variables. Specifically, if individual $i$ on occasion $k$ falls into the category of the variable corresponding to column $j$, then the $(i, j, k)$ cell equals 1; otherwise, it equals 0. In the peer play data, the rows correspond to classifications or groups of children rather than to individuals. This is important for the statistical modeling approach, because the model degrees of freedom do not depend on the sample size, which means that asymptotic theory is applicable when samples are "large" (and the standard sampling assumptions are met). The second difference is that the entries in the 3–way table in the peer play data are frequencies or counts rather than 0/1 indicator variables.

In van der Heijden and de Leeuw's approach to analyzing longitudinal data, two joint tables, referred to as the "BROAD" and "LONG" matrix, are analyzed. The "BROAD" matrix corresponds to the joint table where variables and occasions are interactively coded (van der Heijden, 1987; van der Heijden & de Leeuw, 1989). In the case of the peer play data for the 25 child sub-sample, this is the $(7 \times 55)$ table ${}_{G}\mathbf{F}_{PT}$ where the rows correspond to groups of children and the columns correspond to the play qualities on each

occasion. When the $RC(M)$ association model is fit to ${}_G\mathbf{F}_{PT}$, the two 2–way interactions involving groups (i.e., $GP$ and $GT$), and the 3–way interaction (i.e., $GPT$) are all simultaneous represented by the scale values (i.e., $u_{ij}^{GP} + u_{ik}^{GT} + u_{ijk}^{GPT} \approx \sum_{m=1}^{M} \phi_m \mu_{im}^{G} \nu_{jkm}^{PT}$). The scale values in the bilinear terms represent the departures from the $(GT, P)$ joint independence model, and show the relationship between groups and the combination of play qualities and occasions. Fitting the $RC(M)$ association model to the "BROAD" matrix does not analyze the structure between the play qualities and occasions. To study the structure within interactively coded variables, van der Heijden and de Leeuw suggest analyzing the corresponding marginal table, which in this case is the $\mathbf{F}_{+PT}$ marginal table. Since occasions and play qualities show no evidence of marginal or even partial association, there is no reason to analyze the Play × Time margin.

For the peer play data, the analysis of the "BROAD" matrix is not very useful or informative. Due to the design of the study and for reasons previously given, the Group × Time margin should be fit perfectly and should not be included in the interaction effects that are decomposed and represented by the scale values. If the $RC(M)$ association model was fit to the ${}_G\mathbf{F}_{PT}$ table, it would be very likely that three or more dimensions would be needed to adequately fit the data[1]. The task of interpreting the three or more sets of scale values for the 7 rows (groups) and the 55 columns (11 play qualities at each of the 5 sessions) could be quite difficult.

[1]In the correspondence analysis of ${}_G\mathbf{F}_{PT}$, the first one, two, three and four dimensions account cumulatively for $25.55\%, 47.91\%, 66.85\%$ and $80.79\%$, respectively, of the $X^2$ for testing joint independence $(GT, P)$, which is $X^2 = 421.23$ with $df = 324$

The second joint table is the "LONG" table where the rows correspond to groups at each occasion and the columns correspond to the play qualities (van der Heijden, 1987; van der Heijden and de Leeuw, 1989). For the $N = 25$ sub-sample, fitting the $RC(M)$ association model to the $(35 \times 11)$ table ${}_{TG}\mathbf{F}_{P}$ analyzes the departures from the $(GT, P)$ joint independence model. The model decomposes the sum of the $u_{ij}^{GP}, u_{jk}^{PT}$, and $u_{ijk}^{GPT}$ interactions. Since groups and occasions are interactively coded, a different scale value is estimated for each group on each occasion; therefore, group differences and differences due to time are confounded. To supplement an analysis of a "LONG" table, van der Heijden and de Leeuw suggest also analyzing the individuals by time margin (van der Heijden, 1987; van der Heijden and de Leeuw, 1989). For the peer play data, this corresponds to the Group $\times$ Time margin, $\mathbf{F}_{G+T}$, which as previously mentioned is not an interesting analysis.

### 4.2.3 Joint Analysis of the Peer Play Data

The $RC(M)$ association model was fit to the "LONG" table ${}_{TG}\mathbf{F}_{P}$ for $N = 25$ (i.e., Table 2.2). The fit statistics for $M = 0$ and 1 are reported in Table 4.2. The model with $M = 1$ fits quite well ($G^2 < df$) and accounts for $281.21/446.24 \times 100 = 63.02\%$ of the interaction between the rows and columns.

Unit weights were used in the identification constraints to estimate the parameters of the model (i.e., $g_{ik} = h_j = 1$). The estimated intrinsic association parameter is $\hat{\phi} = 4.826$. The estimated scale values for the play qualities and the combinations of the groups and observation sessions are plotted in Figures 4.2 and 4.3, respectively. The play quality

**Table 4.2**: Fit of the $RC(M)$ association model to the ${}_{TG}\mathbf{F}_{P}$ joint table, $N = 25$.

| $M$ | $df$ | $G^2$ | $G^2/df$ | (p-value) | $X^2$ | (p-value) |
|---|---|---|---|---|---|---|
| 0 | 340 | 446.24 | 1.31 | (.000) | 422.96 | (.001) |
| 1 | 297 | 281.21 | .95 | (.736) | 255.30 | (.962) |

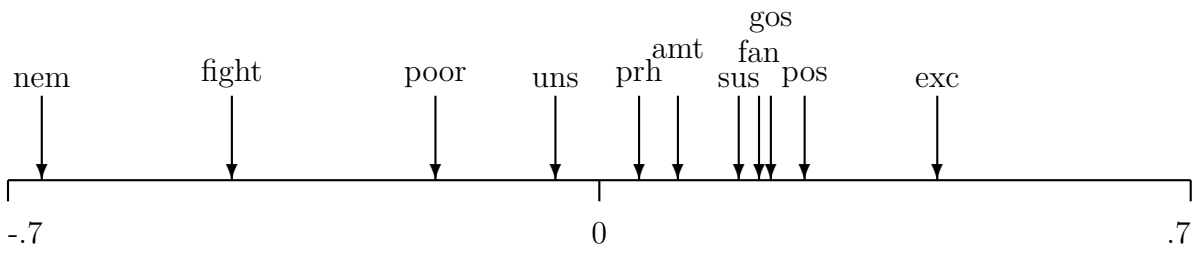



**Figure 4.2**: Estimated scale values of the play qualities from the joint $RC(1)$ association model with unit weights fit to ${}_{TG}\mathbf{F}_{P}$ ($N = 25$).

scale values are again ordered in terms of immature versus mature qualities. The most noticeable difference with respect to previous analyses is that excitability (exc), rather than positive play (pos), is assigned the largest positive scale value. The scale values for the rows (Group × Time combinations) are more difficult to interpret.

The scale values for the group–time combinations could have been plotted with groups on the horizontal axis. Since the groups have no inherent order, a reasonable way to place them along the horizontal axis is by order of their mean scale values (mean across the five occasions). However, since there is a natural order for occasions, the scale values in

Figure 4.3 are plotted versus time with a different symbol for each group classification. The specific mapping of groups to symbols is given in the figure. To aid interpretation, the variable age is coded by the shading of the symbol; namely, the solid symbols ● , ■ and ♣ correspond to the older children and the open or unshaded symbols △, ♡, □ and ○ correspond to the younger children. Sibling acceptance is coded such that the symbols with straight lines (i.e., △, ■, □) correspond to the low sibling acceptance groups, and the symbols with the curved lines (i.e., ♡, ♣, ○ , ● ) correspond to the high acceptance groups.

Similar to the previous analyses of the groups, the older children tend to have positive scale values and the younger children tend to have negative ones. The nature of the interaction involving sibling acceptance emerges from this analysis. At 1 and 3 months before and after the birth of the children's siblings, there is no consistent pattern of the scale values for those who show higher versus lower sibling acceptance. However, at 5 months, the children who showed higher acceptance of their new brother or sister have positive scale values and those who showed lower acceptance have negative scale values. The only exception is for MYH, which contains younger children. At 5 months after birth, the peer play behavior of the children who exhibit lower acceptance of their new sibling tend to have more negative or immature qualities, and the behavior of those who exhibit higher adjustment tends to have more good or mature qualities. Recall that sibling acceptance was measured at the 5 months after their sibling's birth. An implication of this analysis is that the overall quality of peer play at earlier time points is

not very predictive (indicative) of a child's adjustment to their new sibling at 5 months after the sibling's birth.

Examining the scale values in Figure 4.3 reveals a number of other interesting features in the data. The pattern of the scale values at 3 months before the birth ($-3$) is relatively similar to the pattern at 1 month before the birth ($-1$). At $-3$ and $-1$, all of the groups containing older children have positive scale values, and with the exception of FYL at $-1$, all of the groups containing younger children have negative scale values. At the 1 month after birth ($+1$), a rather different pattern is observed. The five groups FYH, FOL, FOH, MYL, and MYH have approximately the same scale value, which is relatively small and close to zero. The model of dependence at $+1$ is due to the behavior of children in the groups MOH and FYL, the latter of which contains only one child. By the fifth session a new pattern seems to emerge with the higher acceptance groups having positive scale values while the lower acceptance groups have negative values. Overall, the pattern of peer play exhibited by the groups before the siblings' births, which seems to be primarily related to age, is disrupted immediately following the birth and a new pattern is established by the fifth session, which seems to be primarily related to the adjustment of the children to their new siblings and secondarily to age.

Differences in the relative variability or spread of the scale values between occasions can be interpreted as reflecting the strength of the association between play qualities and groups on each occasion. Since the variability of the group scale values is the largest at the 3 months before birth, the strongest association between the group classifications and play qualities occurs at 3 months before the siblings' births. The next largest variance

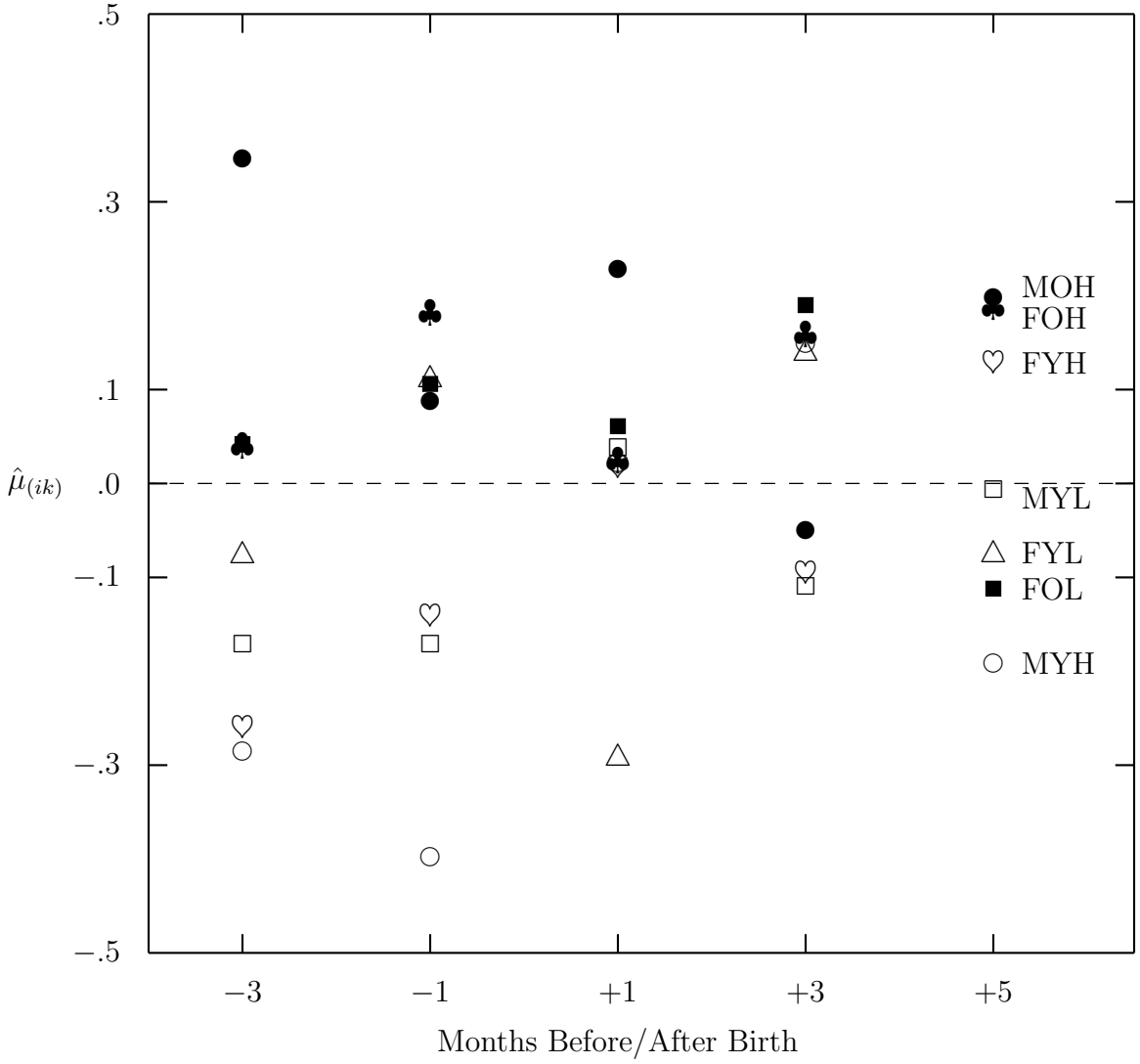



**Figure 4.3**: Estimated group scale values for each occasion ($\hat{\mu}_{(ik)}$) from the joint $RC(1)$ association model fit to ${}_{TG}\mathbf{F}_P$ ($N = 25$).

and second strongest association occurs at one month before birth. The variance of the scale values is the smallest at +1, which implies that the strength of the interaction is the smallest at this time point. At the fourth and fifth sessions, the overall variability of the scale values is larger than it is at +1, but does not reach the same level of variability exhibited at the first session.

The variance of the scale values across groups can also be compared to make statements about which groups show more change/stability. For example, the scale values for the FOH and MYL groups do not vary much across time (i.e., a line connecting the ♣'s or the □'s in Figure 4.3 is relatively flat) relative to the scale values for the groups FYL and MYH, which exhibit a lot of variability (i.e., a line connecting the △'s or the ○ 's is very jagged).

Caution is warranted to guard against "over-interpreting" the scale values of the joint model. The relative variability of scale values across time of each group, which was just interpreted as reflecting the change/stability of the relationship between children and qualities of their peer play, may be due partially to differences between the groups in terms of the number of children in each group. The groups with more children seem to be more stable (e.g., FOH and MYL have 5 and 6 children, respectively) and more likely to reflect general tendencies in the population. The groups with only one or two children (e.g., MYH and FYL have 1 and 2 children, respectively) are more likely to be idiosyncratic. Another reason for caution is that the model may be overfitting the data (van der Heijden & Worsley, 1988). This topic is further discussed in the Section 4.3 where the data are analyzed in a slightly different way.

### 4.2.4 Linear Restrictions

The simple application of the $RC(M)$ association model to joint tables ignores the structure within interactively coded rows and/or columns. The primary suggestion given by van der Heijden and de Leeuw for studying the structure within the interactively coded rows or columns is to analyze the appropriate marginal table (van der Heijden, 1987; van der Heijden & de Leeuw, 1989). Given the design of the peer play study, this approach is not useful here. Fitting the $RC(M)$ association model to the $\mathbf{F}_{G+T}$ marginal table is not an interesting or meaningful analysis. In general, analyzing joint tables and supplementing these analyses by analyzing marginal tables involves separate, non-integrated analyses. It ignores parts of the data to examine other parts. In the peer play data, the marginal and partial associations between groups and play are similar, which is expected given that there is some evidence that time and play are conditionally independent given groups (i.e., $(GT, GP)$ nearly fits the data; see Chapter 2). In other 3–way tables with variables A, B and C, if neither of the $(AB, AC)$ nor $(AB, BC)$ conditional independence model holds, not only can the pattern of partial and marginal associations between A and B be different, it can even be in opposite directions (Bishop, Fienberg, & Holland, 1975; Agresti, 1990). Analyzing joint and marginal tables does not simultaneously assign scale values to groups, play qualities and occasions, and represent the interrelationship among these three modes of the data.

An alternative method to study the structure within interactively coded rows and/or columns is to perform what van der Heijden and de Leeuw call a "constrained" analysis.

Regardless of whether the data in a 2–way table are longitudinal or are from a multi-dimensional table that has been reduced to a 2–way format, the structure within rows and/or columns can be investigated by placing linear restrictions on the scale values. Linear restrictions allow researchers to constrain scale values to equal linear functions of external and/or internal variables. External variables or "natural scores" are known quantities or numbers that have been assigned to the categories based on previous knowledge or some other substantive consideration (Gilula & Haberman, 1988). For example, in the peer play data, the mean age or sibling acceptance score of the children within groups could be used as external variables. Internal variables are unknown quantities that are estimated from the data. For example, when the $RC(M)$ association model is fit to the data, the scale values assigned to the rows and columns are internal variables. Linear restrictions can be used to impose a variety of different constraints on the row and/or column scale values, including constraining some scale values to be equivalent or forcing equal spacing between the scale values of ordered categories (Goodman, 1985, 1986, 1991; Gilula & Haberman, 1988).

With least squares estimation, linear restrictions are easily imposed on the scale values using Rao's work on principal component analysis (Rao, 1964; Böckenholt & Böckenholt, 1990; Sabatier, Lebreton & Chessel, 1989; Israëls, 1987). Böckenholt & Böckenholt (1990) give a clear and concise presentation of how to impose linear restrictions on row and/or column scale values in correspondence analysis. This method could be used to impose linear restrictions on the scale values in $RC(M)$ association models when the parameters are estimated by least squares. Gilula and Haberman (1988) incorporate

linear restrictions into the maximum likelihood estimation of the $RC(M)$ association (and correlation) model. Linearly restricting model parameters, Gilula and Haberman (1988) show how to fit various models, including the models of no 3–factor interaction, conditional independence, and symmetry. Takane, Yanai, and Mayekawa (1991) discuss two alternative, but equivalent, ways of imposing linear restrictions on row and/or column scale values in correspondence analysis. Although they are primarily concerned with correspondence analysis and least squares estimation, their results apply to maximum likelihood estimation of $RC(M)$ association models.

With interactively coded variables, certain restrictions have implications regarding the presence or absence of particular interactions in multiway tables. As an example, consider the joint $RC(M)$ association model in equation 4.2. The column structure can be investigated by writing $\nu^{BC}_{(jk)m}$ as a linear function of the effects of $B$ and $C$, just like in ANOVA where the "dependent" variable (in this case, $\nu^{BC}_{(jk)m}$) is written as a linear function of main and interaction effects of the "independent" variables (in this case, variables $B$ and $C$). Special cases that are especially useful with interactively coded variables are

$$\nu^{BC}_{(jk)m} = \nu^{B}_{jm} + \nu^{C}_{km} \tag{4.4}$$

$$\nu^{BC}_{(jk)m} = \nu^{B}_{jm} \tag{4.5}$$

$$\nu^{BC}_{(jk)m} = \nu^{C}_{km} \tag{4.6}$$

where $\nu^B_{jm}$ represents the effect of $B$ in the $AB$ association, and $\nu^C_{km}$ represents the effect of $C$ in the $AC$ association. The effects $\nu^B_{jm}$ and $\nu^C_{km}$ are internal variables and are estimated by fitting the model to the data. If there are natural scores, say $b_j$, assigned to the categories of variable $B$, then $\nu^B_{jm}$ in 4.4 and 4.5 can be replaced by $u^B_m b_j$, where $u^B_m$ is an unknown constant that must be estimated.

The "additive influence" restriction (equation 4.4), the "$B$ only" restriction (equation 4.5), and the "$C$ only" restriction (equation 4.6) all imply that there is no 3–way association (i.e., $\Theta_{ii',jj',kk'} = 1$ or $u^{ABC}_{ijk} = 0$). The "$B$ only" restriction additionally implies that there is no $AC$ interaction (i.e., $\Theta_{ii',kk'(j)} = 1$ or $u^{AC}_{ik} = 0$) and the bilinear term represents the $AB$ interaction (i.e., $\Theta_{ii',jj'(k)}$ or $u^{AB}_{ij}$). Likewise, the "$C$ only" restriction implies that there is no $AB$ interaction, and the bilinear term represents the $AC$ interaction.

With the peer play data, linear restrictions might be useful to impose equality constraints on particular rows and/or columns to see whether, for example, differences between the scale values of the qualities sustain and fantasy are attributable to chance or reflect some "real" difference. The interpretation given of the interactively coded groups and occasions Figure 4.3 may have over emphasized the observed differences in the scale values. It could be useful to restrict the scale values for particular groups to be equal across occasions and/or to restrict the scale values for some groups to be equal on particular occasions. However, unless such restrictions are determined *a priori* or are otherwise based on substantive theoretical reasons, imposing such restrictions could be "capitalizing on chance".

The linear restrictions that isolate particular interactions, such as those given in equations 4.4, 4.5, and 4.6 are not expected to be very useful for the peer play data. When there are higher–way interactions present in the data, models with such restrictions will not adequately fit the data. Although the size of the interaction may be relatively small, there is some evidence of an interaction between groups, play qualities and occasions in the peer play data (i.e., the $(GT, GP)$ and $(GT, GP, PT)$ loglinear models do not quite fit the data).

Without linear restrictions, the plots of scale values from fitting the $RC(M)$ association models to joint tables are likely to be complex and difficult to interpret. Analyzing both marginal and joint tables involves completely separate analyses. If the nature of marginal and partial associations are different, such analyses could be misleading.

Analyzing joint and marginal tables is not without merit. The analysis of the peer play data in this section did provide some insight into the nature of the associations in the data. However, collapsing the 3–mode table into a 2–way $\mathbf{F}_{GP+}$ marginal table ignored interesting information about the mode that frequencies were collapsed across (i.e., time). Rearranging the 3–mode table into a 2–way ${}_{TG}\mathbf{F}_{P}$ (or ${}_{G}\mathbf{F}_{PT}$) joint table forced occasions to be interactively coded with groups (or play qualities). Ideally, it is desirable to treat the 3 modes as different entities in the model. The next reduction method to be discussed allows groups, play qualities and occasions to have separate and distinct roles in the model.

## 4.3 Conditional Analyses

The conditional method is described in Section 4.3.1 and is illustrated by analyzing the peer play data in Section 4.3.2.

### 4.3.1 The Conditional Approach

The "conditional" method consists of fitting $RC(M)$ association models to each of the 2–way tables for each level of a third variable. In the case of an $(I \times J \times K)$ table of variables $A$, $B$ and $C$, there are three distinct ways of rearranging the data by conditioning on one of the variables: (1) $I$ tables of variables $B$ by $C$, (2) $J$ tables of variables $A$ by $C$, and (3) $K$ tables of variables $A$ by $B$. A set of tables contains all of the cell frequencies or probabilities from the multiway table. For example, when variable $C$ is the conditioning variable, the $K$ $(I \times J)$ tables contain the frequencies $F_{ijk}$ or the probabilities $P_{ijk}$ .

When conditional tables contain frequencies, they are denoted by (for example) $\mathbf{F}_{AB|1}$, $\mathbf{F}_{AB|2}, \ldots, \mathbf{F}_{AB|K}$ where each table $\mathbf{F}_{AB|k}$ contains the frequencies $F_{ijk}$. When conditional tables contain relative frequencies or probabilities, they are denoted by (for example) $\mathbf{P}_{AB|1}$, $\mathbf{P}_{AB|2}$, $\ldots$, $\mathbf{P}_{AB|K}$, where each $\mathbf{P}_{AB|k}$ is an $(I \times J)$ table containing elements $P_{ij|k} = P_{ijk}/P_{..k}$. No information is lost by rearranging a multiway table into a set of conditional tables. Since each table is analyzed separately, each of the conditional tables, which corresponds to one of the categories of the conditioning variable, is given equal "weight" in a conditional analysis.

If the $RC(M)$ association model is fit to each of the $\mathbf{P}_{AB|k}$ tables, then the conditional probabilities are given by

$$P_{ij|k} = \alpha^*_{i(k)}\beta^*_{j(k)} \exp(\sum_{m=1}^{M_k} \phi_{m(k)}\mu_{im(k)}\nu_{jm(k)}) \tag{4.7}$$

where $M_k \leq \min(I, J) - 1$. Multiplying both sides of 4.7 by $P_{..k}$, yields the equation for the unconditional cell probabilities:

$$P_{ijk} = \alpha_{i(k)}\beta_{j(k)} \exp(\sum_{m=1}^{M_k} \phi_{m(k)}\mu_{im(k)}\nu_{jm(k)}) \tag{4.8}$$

where $\alpha_{i(k)}\beta_{j(k)} = P_{..k}\alpha^*_{i(k)}\beta^*_{j(k)}$. The parameters on the right hand side of equations 4.7 and 4.8 are all subscripted by the letter $k$ to indicate that they can differ across the $K$ levels of variable $C$. While the number of dimensions needed to adequately represent the interaction between $A$ and $B$ can differ across levels of $C$, to simplify the discussion, $M_k$ is assumed to be constant across $C$ (i.e., $M_k = M$).

Becker & Clogg (1989a, 1989b) refer to 4.8 as the "RC(M)–Group association" model where the conditioning variable is a group variable. They present this model as a way to examine group differences in the association of two discrete variables. Since the conditioning variable need not be a group variable, the model in equations 4.7 and 4.8 is referred to here as the *conditional* $RC(M)$ association model. The conditional $RC(M)$ association model decomposes the lack-of-fit of a multiplicative (loglinear) model of conditional independence. In the above example, equation 4.8 decomposes the departures

of the cell probabilities from the loglinear model of the conditional independence of variables $A$ and $B$ given variable $C$ (i.e., $(AC, BC)$). The model can also be thought of as decomposing the sum of the interaction effects $u_{ij}^{AB} + u_{ijk}^{ABC}$, where $u_{ij}^{AB}$ and $u_{ijk}^{ABC}$ are from the saturated loglinear model. The bilinear terms in equations 4.7 and 4.8 represent the combined effect of the $AB$ and the $ABC$ interactions. The terms $\alpha_{i(k)}^{*}$ and $\beta_{j(k)}^{*}$ in equation 4.7, and $\alpha_{i(k)}$ and $\beta_{j(k)}$ in equation 4.8 account for the $AC$ and $BC$ interactions and ensure that these margins are fit perfectly. Only the categories of variables $A$ and $B$ are assigned scale values and these are allowed to depend on the level of the conditioning variable.

The identification constraints used to estimate the parameters of the conditional $RC(M)$models are the same as those in equation 3.6 for simple $RC(M)$ association models, except that the scale values and weights are now also indexed by the conditioning variable. The constraints for model 4.8 are

$$\begin{aligned} \sum_i \mu_{im(k)} g_{ik} &= \sum_j \nu_{jm(k)} h_{jk} = 0 \\ \sum_i \mu_{im(k)} \mu_{im'(k)} g_{ik} &= \sum_j \nu_{jm(k)} \nu_{jm'(k)} h_{jk} = \delta_{mm'} \end{aligned} \tag{4.9}$$

The constraints are within levels of the conditioning variable and not between them (i.e., no summations across $k$), which means that the scale values are centered and scaled for each category of the conditioning variable. With respect to the conditional $RC(M)$ association model, Becker & Clogg (1989a) consider unit weights, marginal probabilities (i.e., $P_{i.k}$ and $P_{.jk}$), and average marginal probabilities (i.e., $\sum_k w_k P_{i.k}$ and $\sum_k w_k P_{.jk}$,

where $\sum_k w_k = 1$) as possible choices for $g_{ik}$ and $h_{jk}$. Average marginal probabilities include $P_{i..}$ and $P_{.j.}$ as special cases. The conditional probabilities $P_{i.|k}$ and $P_{.j|k}$, which are the weights used in conditional correlation models or correspondence analysis (Israëls, 1987), could also be used as weights, but this choice for the association model (to the author's knowledge) has not been used in practice.

The choice of weights should depend on substantive considerations. If comparisons are made across tables, then the weights should be independent of the conditioning variable (Becker & Clogg, 1989a). For example, when each table corresponds to a different group, then unit or average marginal probabilities would be better choices than (bivariate) marginal probabilities or conditional probabilities. If the weights depend on the table, then the estimated scale values reflect the different weights as well as any differences in the structure (association) across the tables. Another consideration in choosing weights is whether certain margins are fixed by design.

### 4.3.2 Conditional Analysis of the Peer Play Data

For the peer play data, the interaction between groups and play qualities will be analyzed conditioning on occasions (time). The departures from the $(GT, PT)$ loglinear model, which does not fit the data ($df = 300, G^2 = 404.98$), will be studied. The model that is fit to the data is

$$\ln(F_{ijk}) = u_k + u^G_{i(k)} + u^P_{j(k)} + \sum_m \phi_{m(k)} \mu_{im(k)} \nu_{jm(k)}$$

or in terms more similar to standard loglinear models,

$$\ln(F_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{it}^{GT} + u^{PT} + \sum_m \phi_{m(k)}\mu_{im(k)}\nu_{jm(k)} \tag{4.10}$$

where $u_k = u + u_k^T$, $u_{i(k)}^G = u_i^G + u_{it}^{GT}$, and $u_{j(k)}^P = u_j^P + u_{it}^{PT}$. Both the $GT$ and the $PT$ margins will be fit perfectly. The sum of the interactions $u_{ij}^{GP} + u_{ijk}^{GPT}$ are decomposed and represented by the scale values and the intrinsic association parameters. To estimate the parameters of model 4.10, unit weights are used in the identification constraints for the reasons previously discussed.

As described so far, the conditional approach does not lead to parsimonious representations of the data, but rather to five separate, unconnected analyses — one for each $\mathbf{F}_{GP|k}$ table for $k = 1, \ldots, 5$. The analyses can be connected and simpler models fit by placing homogeneity constraints on the parameters across levels of the conditioning variable. Special cases of the conditional $RC(M)$model where some or all of the parameters in the bilinear terms are constrained to be constant across levels of the conditioning variable can be used to answer questions about whether the structure, strength, or both the structure and strength of the relationship between two variables differs across levels of the third variable. Placing homogeneity constraints on the measures of association (i.e., $\phi_{m(k)} = \phi_m$) is equivalent to assuming that the strength of the relationship is constant across levels of the third variable. Homogeneity constraints on the scale values (i.e., $\mu_{im(k)} = \mu_m$ and/or $\nu_{jm(k)} = \nu_{jm}$ ) can be interpreted as assuming that the structure of the relationship is the same across levels of the conditioning variable.

Conditional $RC(M)$ association models for $M = 0$ and 1 with various combinations of homogeneity constraints on the model parameters were fit to $\mathbf{F}_{GP|k=1}, \ldots, \mathbf{F}_{GP|k=5}$. The models with homogeneity constraints were fit using Eliason's (1989) program *Categorical Data Analysis System* (CDAS). In Table 4.3, the fit statistics for six models are reported. The overall fit statistic for each model is reported at the bottom of each section (next to "total"), and is broken down into the component statistics corresponding to each of the five, 2–way sub-tables. The models are arranged in Table 4.3 according to their degrees of freedom, such that the first model is the most parsimonious (i.e., conditional independence of groups and play qualities given occasions) and the last is the least parsimonious (i.e., a separate $RC(1)$ association model for each $\mathbf{F}_{GP|k}$). In all of the models, the $GT$ and $PT$ margins are fit perfectly.

The $(GT, PT)$ conditional independence model ($M = 0$, no restrictions) does not fit the data, but examining the fit statistics for each of the five sub-tables reveals that at $k = 5$, the model fits, and groups and play qualities appear to be independent. At $k = 1$ and 2, groups and play qualities are definitely dependent, and at $k = 3$ and 4, there is some evidence that they are dependent. These statistics suggest that the strength of the dependency may decrease over time.

The last model reported in Table 4.3 corresponds to fitting a separate $RC(1)$ association model to each $\mathbf{F}_{GP|k}$ table. Overall, this model fits the data well ($G^2 = 204.83$ with $df = 225$), and it fits each of the individual sub-tables well ($G^2 < df$ or $G^2 \approx df$). Since the table corresponding to $k = 5$ did not show any dependency, the model overfits

the data. Even though this model is not very parsimonious, it does illustrate that a one dimensional conditional model is adequate for the data.

The more interesting models are those with homogeneity constraints. The second model in Table 4.3, which has homogeneity constraints on all the parameters in the bilinear terms across time (i.e., $\phi_{(k)} = \phi$, $\mu_{i(k)} = \mu_i$, and $\nu_{j(k)} = \nu_j$), is an improvement over the independence model; however, the fit relative to the degrees of freedom is not good ($G^2 = 322.08$ with $df = 285$). Since the strength of the relationship between groups and play qualities appears to change over time, the next model fit includes a different $\phi_{(k)}$ parameter for each occasion. This model, the third one in Table 4.3, fits the data better ($G^2 = 307.80$ with $df = 281$) and may be adequate for the data. The estimated intrinsic association parameters for the third model are $\hat{\phi}_{(1)} = 2.714, \hat{\phi}_{(2)} = 2.479, \hat{\phi}_{(3)} = 0.588, \hat{\phi}_{(4)} = 0.977$, and $\hat{\phi}_{(5)} = 0.929$. These support the earlier conjecture from the joint analysis that the strength of the relationship decreases from the first through the third session, reaches its lowest value at the third session (1 month after the birth of the children's siblings), and increases slightly from the fourth to fifth session.

The nature of the relationship between groups and play qualities measured by the estimated $\phi_{(k)}$ parameters of the third model is represented in Figure 4.4, where the group and play quality scale values are plotted. The patterns of the scale values of the groups and play quality are similar to the scale values from the analysis of the $\mathbf{F}_{GP+}$ margin and the $RC(1)$ association model fit to $\mathbf{F}_{GP|1}$ for $N = 30$ reported in Chapter 3. The older children have larger scale values than those younger, and the younger and older males with high sibling adjustment (MYH and MOH) have the largest negative and positive

**Table 4.3**: Fit of unrestricted and restricted conditional $RC(M)$ association models with unit weights to $\mathbf{F}_{GP|1}, \ldots, \mathbf{F}_{GP|5}$ ($N = 25$).

| $M$ | Restrictions | Time | $df$ | $G^2$ | $G^2/df$ | (p-value) | $X^2$ |
|---|---|---|---|---|---|---|---|
| 0 | none | 1 | 60 | 101.91 | 1.70 | (.006) | 98.99 |
| | | 2 | 60 | 89.98 | 1.50 | (.007) | 76.59 |
| | | 3 | 60 | 77.94 | 1.30 | (.060) | 72.71 |
| | | 4 | 60 | 75.15 | 1.25 | (.091) | 75.15 |
| | | 5 | 60 | 54.46 | .91 | (.678) | 54.46 |
| | | total | 300 | 404.98 | 1.35 | (.000) | 377.90 |
| 1 | $\phi_t = \phi$ | 1 | | 65.60 | | | 63.48 |
| | $\mu_{i(k)} = \mu_i$ | 2 | | 62.47 | | | 51.00 |
| | $\nu_{j(k)} = \nu_j$ | 3 | | 71.45 | | | 66.18 |
| | | 4 | | 75.47 | | | 73.04 |
| | | 5 | | 47.09 | | | 46.37 |
| | | total | 285 | 322.08 | 1.13 | (.064) | 300.07 |
| 1 | $\mu_{i(k)} = \mu_i$ | 1 | | 51.13 | | | 50.41 |
| | $\nu_{j(k)} = \nu_j$ | 2 | | 55.29 | | | 45.87 |
| | | 3 | | 75.71 | | | 71.43 |
| | | 4 | | 76.18 | | | 70.93 |
| | | 5 | | 49.48 | | | 49.17 |
| | | total | 281 | 307.80 | 1.10 | (.130) | 287.82 |
| 1 | $\nu_{j(k)} = \nu_j$ | 1 | | 54.42 | | | 51.71 |
| | | 2 | | 48.33 | | | 38.23 |
| | | 3 | | 58.97 | | | 38.39 |
| | | 4 | | 58.73 | | | 52.06 |
| | | 5 | | 33.10 | | | 32.45 |
| | | total | 261 | 251.55 | .96 | (.652) | 226.67 |
| 1 | $\mu_{i(k)} = \mu_i$ | 1 | | 47.56 | | | 46.84 |
| | | 2 | | 58.50 | | | 48.50 |
| | | 3 | | 66.34 | | | 61.50 |
| | | 4 | | 65.76 | | | 63.32 |
| | | 5 | | 43.51 | | | 42.93 |
| | | total | 245 | 281.87 | 1.15 | (.053) | 263.09 |
| 1 | none | 1 | 45 | 43.06 | .96 | (.554) | 42.60 |
| | | 2 | 45 | 46.52 | 1.03 | (.410) | 38.23 |
| | | 3 | 45 | 46.47 | 1.03 | (.412) | 42.64 |
| | | 4 | 45 | 38.05 | .85 | (.759) | 35.35 |
| | | 5 | 45 | 30.74 | .68 | (.948) | 26.65 |
| | | total | 225 | 204.83 | .91 | (.829) | 188.47 |

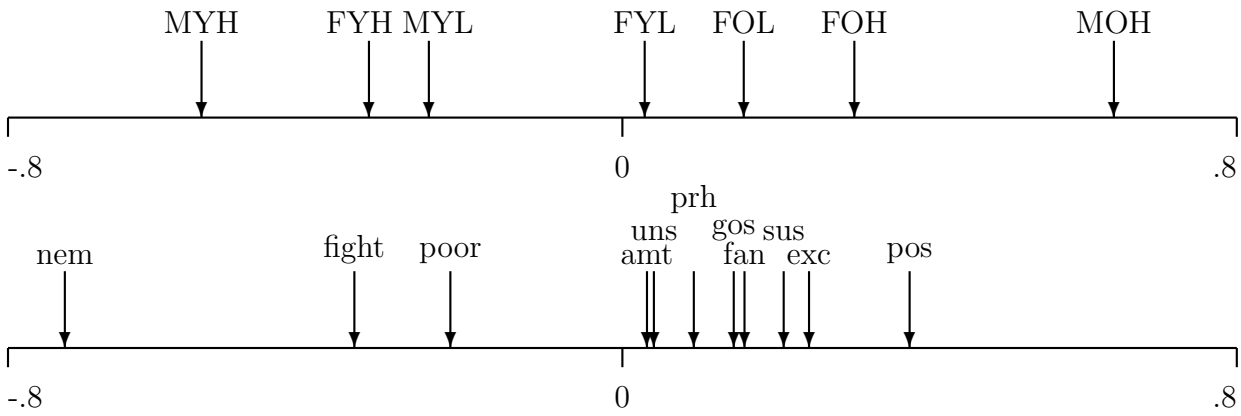



**Figure 4.4**: Estimated scale values of the $RC(1)$ association model fit to the conditional tables $\mathbf{F}_{GP|1}, \ldots, \mathbf{F}_{GP|5}$ with homogeneity constraints on the scale values for groups and play qualities (i.e., $\mu_{i(k)} = \mu_i$), $\nu_{j(k)} = \nu_j$).

scale values, respectively. The play qualities are ordered from mature or "bad" (i.e., negative emotion, fight, and poor play) to the mature or "good." Positive play has the largest positive scale value.

The component fit statistics of the third model for the $\mathbf{F}_{GP|1}$ and $\mathbf{F}_{GP|2}$ sub-tables are considerably smaller than the statistics of the first and second models, but the component fit statistics of the first, second, and third models for the $\mathbf{F}_{GP|3}$ and $\mathbf{F}_{GP|4}$ sub-tables are approximately the same. The overall improvement in fit of the third model relative to the first and second models arises primarily from the improved fit of the $\mathbf{F}_{GP|1}$ and $\mathbf{F}_{GP|2}$ sub-tables. The fourth model in Table 4.3, relaxes the homogeneity constraint on

the group scale values but retains the constraint on the play qualities (i.e., $\nu_{j(k)} = \nu_j$). A $\phi_{(k)}$ parameter and a set of seven group scale values are estimated for each of the five occasions. As expected, the fourth model fits better than the third model, which constrains both play qualities and groups. Most of the improvement in the overall fit of the model primarily comes from the improved fit of the sub-tables at $k = 3$ and 4.

The fourth model is interesting since it is functionally and empirically very similar to the $RC(1)$ association model fit to the ${}_{TG}\mathbf{F}_P$ joint table. The joint model is

$$\ln(F_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \phi\mu_{(ik)}^{GT}\nu_j^P \tag{4.11}$$

and the conditional model is

$$\ln(F_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + u_{jk}^{PT} + \phi_{(k)}\mu_{i(k)}\nu_j \tag{4.12}$$

Equations 4.11 and 4.12 are very similar. The major difference between them is that the conditional model includes the interaction terms $u_{jk}^{PT}$, which ensures that the $PT$ margin is fit perfectly, while the joint model does not include these interaction terms. While the bilinear terms in the models are similar, they are not identical, and the models are not nested. Since the joint model does not include $u_{jk}^{PT}$ parameters and has only one $\phi$ parameter, it has more degrees of freedom and is the more parsimonious of the two models. Empirically, the $PT$ association is not very important (see Chapter 2); therefore, the bilinear terms in the joint model are primarily representing the combined

effect of the $GP$ and $GPT$ interactions, which is the interaction effect represented by the bilinear terms in the conditional model. Consistent with this observation, the models are approximately equivalent in terms of goodness-of-fit. The ratio of $G^2$ to $df$ for the conditional model equals $251.55/261 = .964$, which is approximately equal to the ratio for the joint model $281.21/297 = .945$. The estimated scale values for the play qualities from the conditional model are plotted in Figure 4.5, and they are nearly identical to those from the joint model, which are plotted in Figure 4.2. Given the functional and empirical similarities between the joint and conditional models, if $\phi\mu_{ik}^{GT} \approx \phi_{(k)}\mu_{i(k)}$, then the models are basically equivalent. To compare the estimated group scale values of the conditional and joint models, the product of the estimated intrinsic association parameters and group scale values from the conditional model (i.e., $\hat{\phi}_{(k)}\hat{\mu}_{i(k)}$), which should be proportional to the $\hat{\mu}_{(ik)}^{GT}$ scale values from the joint model, are plotted in Figure 4.6. Figures 4.3 and 4.6 are very similar with only minor differences.

The fifth model in Table 4.3, which imposes homogeneity constraints only on the groups, does not fit. The model uses many degrees of freedom to estimate 5 sets of 11 scale values for the play qualities. The fact that the model does not fit supports the conjecture that the joint $RC(1)$ association model fit to ${}_{G}\mathbf{F}_{PT}$ would not fit the data. The fifth conditional model should fit better than the joint $RC(1)$ association model, because the conditional model includes the extra 2–way interaction parameters $u_{ij}^{GT}$, which are important. The joint model decomposes the $GT$ interaction in addition to the $GP$ and the $GPT$ interactions.

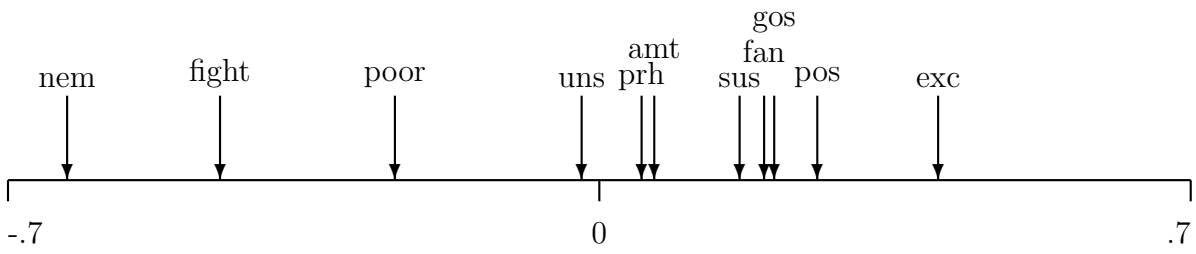



**Figure 4.5**: Estimated play quality scale values of the conditional $RC(1)$ association model with homogeneity constraints on the play quality scale values ($\nu_{j(k)} = \nu_j$).

The conditional analysis provides further support and confirmation of the general conclusions made from the analyses of joint tables, including the suspicions that the joint $RC(1)$ association model overfits the data. The conditional $RC(1)$ association model with homogeneity constraints on both the groups and play qualities (the third model in Table 4.3) fits the data fairly well, and an argument could be made that the model with constraints only on the play qualities (the fourth model) overfits the data. Since the fourth model is essentially equivalent to the joint $RC(1)$ association model fit to ${}_{TG}\mathbf{F}_P$, the joint model probably also overfits the data. Even the third conditional model may not be the most parsimonious model that adequately fits the data. The third model, which is

$$\ln(F_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{ij}^{GT} + u_{jk}^{PT} + \phi_{(k)}\mu_i\nu_j \tag{4.13}$$

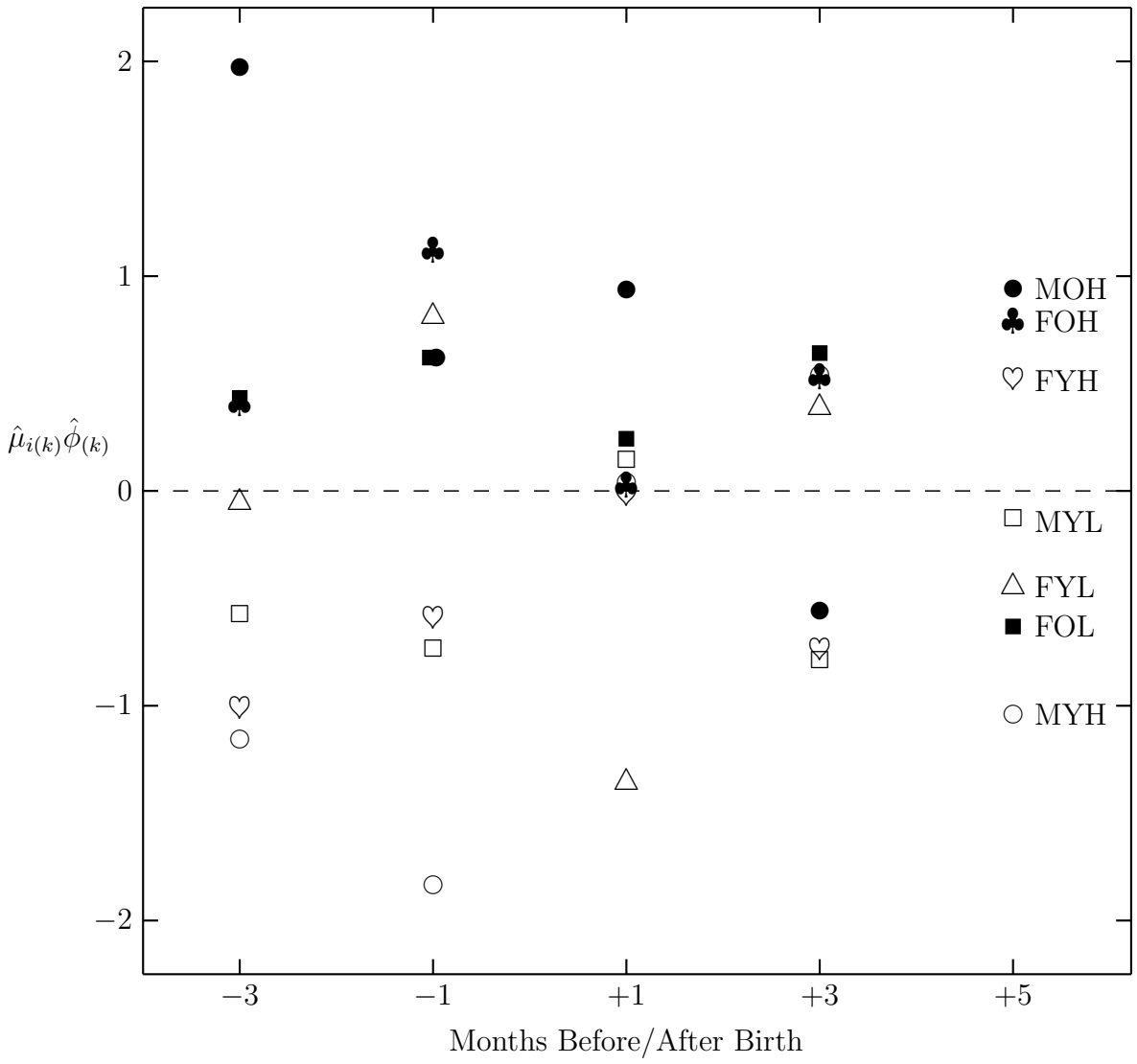



**Figure 4.6**: Estimated group scale values multiplied by intrinsic association parameters (i.e., $\hat{\mu}_{i(k)}\hat{\phi}_{(k)}$) from the conditional $RC(1)$ association model with homogeneity constraints on the play quality scale values (i.e., $\nu_{j(k)} = \nu_j$).

contains interactions terms for the $PT$ association. The $PT$ partial and marginal associations are not important. There is a more parsimonious model that does not include the $u_{jk}^{PT}$ terms and fits the data. This model, which is a special case of the model generalizations developed in Chapter 6, is presented and discussed in Chapter 6.

In general, the conditional reduction strategy readily handles cases where there are three variables in which one has a different role than the other two. With four or more variables, some of the variables can be interactively coded such that there are three "variables" where some or all of the "variables" are combinations of two or more of the simple (single) variables. The treatment of the classification variables age, gender and sibling acceptance is an example of this. To use a conditional model, the variables need to be classified into either two or three types, where one of the types has a slightly different role in the study.

## 4.4 Summary: Inadequacy of 2–Way Decompositions

Marginal, conditional and joint tables are three ways of arranging frequencies (or probabilities) from 3– or higher–way cross-classifications into 2–way formats such that they can be analyzed by $RC(M)$ association models. Reducing a multiway table to a 2–way array emphasizes certain aspects of the data, often at the expense of suppressing or completely losing other aspects. Therefore, the way that a multiway table is "flattened" into a 2–way array must be guided by the particular aspects of the data that a researcher

wants to analyze, by specific hypotheses or theories regarding the underlying structure of the data, and by any natural structure present in the data.

For the peer play data, the association between groups and play qualities is of particular interest; therefore, the $\mathbf{F}_{GP}$ marginal table, the ${}_{TG}\mathbf{F}_{P}$ joint table, and the $\mathbf{F}_{GP|1}, \ldots, \mathbf{F}_{GP|5}$ conditional tables were all analyzed in this chapter. Each of these analyses contributed to our knowledge and understanding of the nature of the dependencies in the data, but each focused on slightly different aspects of the data.

Marginal tables contain frequencies collapsed across at least one way of the higher–way table. Analyses of marginal tables focus on 2–way marginal associations rather than partial associations. Rearranging multiway tables into marginal tables results in the loss of the 3– and higher–way associations, which in the case of the peer play data, was the effect of time. The conditional and joint methods retain information about the higher–way associations. The bilinear terms in the conditional and joint $RC(M)$models simultaneously represent bivariate and higher–way associations; however, not all of the bivariate associations are decomposed in these models. The conditional method focuses on the association between two variables for each level of a third variable, and the departure of the data from a multiplicative model of conditional independence is decomposed. The specific conditional independence model depends on which variable is the conditioning variable. The joint method focuses on the association between two "variables" where the "variables" consist of interactively coded variables of the multiway table. The joint method decomposes the departure of the data from a model of joint independence where the specific joint independence model depends on which variables are interactively coded.

The methods differ with respect to uniqueness and complexity. There is basically just one way to reduce a multiway table to a set of marginal tables, but there are multiple ways of reducing a multi–way table to a joint table or to a set of conditional tables. To form a set of conditional tables, a decision must be made regarding which variable will be the conditioning variable. To form a joint table, a decision must be made regarding how to split the variables into two sets where one set becomes the rows and the other becomes the columns. Since the marginal and conditional methods result in sets of tables, sets of $RC(M)$ association models are needed to analyze them (assuming, of course, that all of the marginal tables are interesting and that no homogeneity constraints are placed on the model parameters across the different tables). Since the joint method results in a single table, only one $RC(M)$ model is needed.

With respect to the peer play data, the $\mathbf{F}_{GP+}$ marginal table, the ${}_{TG}\mathbf{F}_{P}$ joint table, and the $\mathbf{F}_{GP|1}, \ldots, \mathbf{F}_{GP|5}$ conditional tables provide interesting and informative ways of viewing and analyzing the data. The observed overlap and similarity of the results from these analyses was due to the pattern of associations and dependencies in the data, and to the theoretical relationship among the various models. Specifically, the nature of the partial association between groups and play qualities is the same as the marginal association (to the extent that the $(GT, GP)$ loglinear model of conditional independence fits the data). While this conditional independence model was judged to be inadequate, it picks up a large part of the structure that is present in the data. The lack of a 2–way interaction between play qualities and occasions (marginal and partial $PT$ association) and the functional similarity between the joint model fit to ${}_{TG}\mathbf{F}_{P}$ and the conditional

model fit to $\mathbf{F}_{GP|1}, \ldots, \mathbf{F}_{GP|5}$ with homogeneity restrictions on the scale values for the play qualities led to the similar results for these two models. The theoretical relationship between marginal, conditional and joint correspondence analyses is discussed in Takane, Yanai, and Mayekawa (1991). Takane, et al. (1991) show that the correspondence analysis of marginal and conditional tables is equivalent to the analysis of a particular joint table with certain linear restrictions on the row and column scale values. While they only explicitly consider correspondence analysis, the relationships also hold for $RC(M)$ association models.

Depending on the structure and associations present in the data, restrictions can be placed on the parameters of the $RC(M)$models to yield simpler, more parsimonious models and representations of the associations. With conditional models, homogeneity restrictions can be placed on the measures of association, the row scale values, and/or the column scale values. When tables contain interactively coded variables, linear restrictions can be placed on the scale values of these variables.

The simple application of the $RC(M)$ association model to marginal, joint and/or conditional tables does not simultaneously assign scale values to groups, play qualities, and occasions such that the dependencies among them can be graphically represented. The conditional and joint $RC(M)$ association models represent the combined effect of 2– and 3–way interactions with a 2–way decomposition. Without any simplifying restrictions, these models can be quite un-parsimonious and un-interpretable (Nelder, 1989). To make the models more parsimonious, various constraints and restrictions on the parameters can be imposed. In the case of joint $RC(M)$ models, if there are higher–way

interactions, models with additive (linear) restrictions such as those described in Section 4.2.4 will not fit. Conditional models with homogeneity constraints (described in Section 4.3) and joint models with additive restrictions resemble generalizations of the $RC(M)$ association model to higher–way tables. The bilinear terms in these models are not just simple 2–way, singular value decompositions. An example of one such model, which is of particular interest and is discussed further in Chapter 6, is equation 4.13. This and other models with linear and/or homogeneity constraints and restrictions on the parameters are special cases of some of the generalizations of the $RC(M)$ association model to higher–way tables.

The 2–way reduction methods can handle a variety of situations, but they all suffer from the same basic limitation. The models discussed in this chapter are basically just 2–way $RC(M)$ models and are limited to analyzing associations in terms of 2–factor interactions. When all of the 2–way interactions are decomposed by marginal analyses, the 3– and higher–way associations are lost by reducing a multiway table to just 2–ways. When conditional or joint reduction strategies are used, the higher–way associations are not lost, but the decompositions are 2–factor (singular value) decompositions, rather than 3– or higher–way decompositions. When there are three classes or modes of variables, each of which has a distinct role, 3–way decompositions in which scale values are assigned to the categories or levels of each mode would be desirable. A 3–way decomposition would permit an examination of the associations among the three modes. When a conditional or joint model is fit to a multiway table, at least one of the 2–way associations is not analyzed. The models in the next chapter do not suffer these limitations.

Other ways of rearranging higher–way tables into 2–way arrays, such as Burt matrices or composite tables, were not considered here primarily because they consist of concatenations of marginal tables and do not retain information about higher–way interactions. Burt matrices and composite tables are useful ways of rearranging data so that correlation models and correspondence analysis can be used to analyzed multiway tables (Israëls, 1987; Carroll & Green, 1988, Wasserman, Faust & Galaskiewicz, 1990; Giffi, 1990). With loglinear models, there is no need to collapse tables into 2–way formats to analyze and model the data. Generalizations of $RC(M)$ association models to higher-way tables, which are discussed in the following chapters, are extensions of loglinear models; therefore, such models can be used to analyze and model multiway tables without collapsing the data into 2–way formats. There is no need to suppress parts of the data to analyze other parts.

# Chapter 5

# MODEL GENERALIZATIONS

The models discussed in this chapter are generalizations of $RC(M)$ models to 3– and higher–way tables. These models include extra and/or more complex terms to represent the additional interactions present in multiway tables. Analyses of the peer play data will illustrate some of these models.

With 3– and higher–way tables, there are multiple 2–way interactions, each of which can be analyzed by separate 2-way decompositions. These "no 3–factor interaction" models include extra bilinear terms to represent each 2–way interaction. Simplifying constraints can be placed on the parameters of the additional bilinear terms to yield more parsimonious models. For example, the scale values for some or all of the variables can be restricted to be equal for different decompositions (bilinear terms). Unlike conditional and joint analyses, all of the 2–way interactions can be decomposed. When there are no 3– or higher–way interactions, adding extra bilinear terms can be satisfactory. The generalizations of the $RC(M)$ association model to multiway tables where only bivariate

interactions are important are reviewed in Section 5.1. These generalizations will be referred to as "bivariate interaction, log-bilinear" models.

Models containing 3–way interactions or combinations of 2– and 3–way interactions can be also be analyzed by 2–way decompositions. The conditional and joint models discussed in Chapter 4 are examples of this type of generalization. The higher–way effects are represented in the model by simple or complex bilinear terms, depending on whether simplifying restrictions are imposed on model parameters. Rather than analyzing 3–way associations by 2–factor decompositions, 3–way interactions can be analyzed by 3–factor decompositions. In the model generalizations that use 3–factor decompositions, the categories of each of the three variables or modes are assigned scale values. Only a few generalizations have been proposed that use 3–factor decompositions.

In Section 5.2, generalizations of $RC(M)$ association and correlation models that represent 3– or higher–way interactions by introducing more complex terms are reviewed. These generalizations will be referred to generically as "higher–way interaction, multilinear" models. Specific models depend on the highest order of interaction represented, the order of the decomposition, and whether the model is an additive (or linear) or log-additive model. For example, the joint $RC(M)$ association model described in the previous chapter is a "3–way interaction, log-bilinear model." Generalizations of correspondence analysis are included in Section 5.2, since such models suggest additional generalizations of the $RC(M)$ association model. Such further generalizations are described and developed in Chapter 6. To illustrate the potential usefulness of 3–way decompositions, a modification of one of the generalizations of correspondence analysis is

used to analyze the peer play data. This analysis is presented in Section 5.3. In the final section of this chapter, Section 5.4, the inadequacy of the models and methods reviewed in this chapter are briefly summarized.

## 5.1 Bivariate Interaction, Log-Bilinear Models

When only bivariate or first order interactions are present in a multi–way table, no–3–way interaction loglinear models are adequate (e.g., for a 3–way table, equation 3.3). The association model generalizations that are discussed in this section are no–3–way interaction loglinear models in which the interaction terms (e.g., $u_{ij}^{AB}$, $u_{ik}^{AC}$ and $u_{jk}^{BC}$ in equation 3.3) are decomposed by 2–way singular value decompositions (Becker, 1989a, 1990b). This family of no 3–way interaction association models (or "bivariate interaction, log-bilinear models"), which were described by Becker (1989a), is presented and various special cases of it are discussed in Section 5.1.1. Some of the special cases were proposed by others as extensions of loglinear models or as generalizations of the $RC(M)$ association model.

Applications of these models are relatively rare; therefore, in Section 5.1.2, a subset of the possible model generalizations are fit to the peer play data to illustrate how to fit and interpret these models.

### 5.1.1 'No 3–Way Interaction' Association Models

In this section, generalizations of the $RC(M)$ association model to higher–way tables are described where 2–way interactions are the highest order interaction modeled. In Section 5.1.1.1, a very general model proposed by Becker (1989) is presented. In Section 5.1.1.2 special cases of the general model are discussed, and in Section 5.1.1.3, generalizations to 4– and higher–way tables are described.

#### 5.1.1.1 Becker's General Model

Becker's no 3–way interaction association model is

$$\ln(P_{ijk}) = u^o_{ijk} + \sum_{m=1}^{M_1} \phi_m^{AB} \mu_{1im} \nu_{1jm} + \sum_{m=1}^{M_2} \phi_m^{AC} \mu_{2im} \eta_{2km} + \sum_{m=1}^{M_3} \phi_m^{BC} \nu_{3jm} \eta_{3km} \qquad (5.1)$$

where $u^o_{ijk} = u + u^A_i + u^B_j + u^C_k$, $M_1 \leq \min(I, J) - 1$, $M_2 \leq \min(I, K) - 1$, and $M_3 \leq \min(J, K) - 1$ (Becker, 1989a). The parameters $\phi_m^{AB}$, $\phi_m^{AC}$ and $\phi_m^{BC}$ measure the intrinsic association on dimension $m$ between $A$ and $B$, $A$ and $C$, and $B$ and $C$, respectively. The subscripts 1, 2, and 3 on the scale values indicate specific interactions; namely, 1 refers to the $AB$ interaction, 2 refers to $AC$, and 3 refers to $BC$. When $M_1 = \min(I, J) - 1$, $M_2 = \min(I, K) - 1$, and $M_3 = \min(J, K) - 1$, equation 5.1 is equivalent to the no 3–factor interaction loglinear model $(AB, AC, BC)$ (equation 3.3). When $M_1 = M_2 = M_3 = 1$, equation 5.1 is equivalent to the log–multiplicative model discussed by Pannekoek (1985) and one of the possible generalizations of the $RC(M)$ association model mentioned by Goodman (1986).

To estimate the parameters in 5.1, identification constraints on the main effect terms and the scale values are required. The constraints on the main effect terms are the ones typically used for standard loglinear models (e.g., $\sum_i u_i^A = \sum_j u_j^B = \sum_k u_k^C = 0$, or $u_1^A = u_1^B = u_1^C = 0$), and the constraints on the scale values in the bilinear terms,

$$\sum_i \mu_{1im} h_i^A = \sum_i \mu_{2im} h_i^A = \sum_j \nu_{1jm} h_j^B = \ldots = \sum_k \eta_{3km} h_k^C = 0 \tag{5.2}$$

$$\sum_i \mu_{1im}\mu_{1im'} h_i^A = \sum_i \mu_{2im}\mu_{2im'} h_i^A = \sum_j \nu_{1jm}\nu_{1jm'} h_j^B = \ldots = \sum_k \eta_{3km}\eta_{3km'} h_k^C = \delta_{mm'}$$

where $h_i^A$, $h_j^B$, and $h_k^C$ are fixed and known weights. When $h_i^A = h_j^B = h_k^C = 1$, the constraints are equivalent to those used by Becker (1989a). The more general constraints are given here, because in particular applications, other weighting schemes, such as marginal probabilities (i.e., $P_{i..}$, $P_{.j.}$, $P_{..k}$) are sometimes desirable.

Becker (1990b) shows that singular value decompositions of matrices of estimated interaction parameters from loglinear models yield approximations of the scale values. Such approximate solutions provide a way to explore various possible special cases of 5.1, and also provide good starting values for the iterative routines that compute the MLE of the model parameters.

The interpretation of model 5.1 is similar to the $RC(M)$ association model for 2–way tables in a number of respects. The models represent log–odds ratios as simple functions of the intrinsic association parameters and the scale values. For example, model 5.1 decomposes the conditional odds ratio of the $(2 \times 2)$ subtables of $A$ and $B$ for given levels

of $C$ as follows

$$\ln(\Theta_{ii',jj'(k)}) = \sum_{m=1}^{M_1} \phi_m^{AB}(\mu_{1im} - \mu_{1i'm})(\nu_{1jm} - \nu_{1j'm})$$

Each of these odds ratios are equivalent for different values of $k$ (i.e., $\Theta_{ii',jj'(k)} = \Theta_{ii',jj'(k')}$ for all $k, k'$), since the model does not contain a 3–way association. The conditional odds ratios $\Theta_{ii',kk'(j)}$ and $\Theta_{jj',kk'(i)}$ are similarly represented.

If a particular bivariate interaction (partial association) is not present in a table, then the corresponding intrinsic association parameter and the conditional log–odds ratios both equal zero. For example, if there is no $BC$ interaction (i.e., $u_{jk}^{BC} = 0$ for all $j$ and $k$), then $\phi_m^{BC} = 0$ for all $m$, and $\ln(\Theta_{jj',kk'(i)}) = 0$ for all $j$, $j'$, $k$, $k'$, and $i$. In this case, the conditional independence loglinear model $(AB, AC)$ will fit the data and the $AB$ and $AC$ partial associations will equal the marginal associations. Rather than using a generalization of the $RC(M)$ association model, such as equation 5.1, the $AB$ and $AC$ associations can be analyzed simply by fitting $RC(M)$ association models to the marginal tables $\mathbf{F}_{AB+}$ and $\mathbf{F}_{A+C}$. This strategy is especially useful when programs for fitting the more complex models are not available. If conditional independence does not hold or if certain simplifying constraints are desired on the scale values in equation 5.1, this strategy will not work.

#### 5.1.1.2 Special Cases of the General Model

Possible simplifying constraints that can be placed on the parameters include homogeneity or consistency constraints. Consistency constraints specify that the scale values for the categories of the variables are the same or consistent regardless of the interaction (i.e., $\mu_{1im} = \mu_{2im}$, $\nu_{1jm} = \nu_{3jm}$, and $\eta_{2km} = \eta_{3km}$). Equation 5.1 with consistency constraints yields a more parsimonious model. This special case is

$$\ln(P_{ijk}) = u^o_{ijk} + \sum_{m=1}^{M} (\phi^{AB}_m \mu_{im}\nu_{jm} + \phi^{AC}_m \mu_{im}\eta_{km} + \phi^{BC}_m \nu_{jm}\eta_{km}) \tag{5.3}$$

Except for two differences in the identification constraints, this model is equivalent to one of Choulakian's (1988a, 1988b) generalizations of the $RC(M)$ association model to 3–way tables. Choulakian uses the marginal sums as weights (i.e., $h^A_i = P_{i..}$, $h^B_j = P_{.j.}$, $h^C_k = P_{..k}$), and does not impose an orthogonality constraint on the scale values. The orthogonality restriction is not required to identify and estimate the scale values in equation 5.3; however, in some applications, it may be desirable to constrain them to be orthogonal. In general, the model with the orthogonality restriction will not fit a table as well as the model without the restriction. At best, it will be the same. Choulakian's original 1988a paper contains mistakes and omissions that are not corrected in the 1988b errata; hence, the correct and full set of maximum likelihood equations and an algorithm for fitting this model (equation 5.3) are reported in Appendix B.

The model with consistency constraints (equation 5.3) is a particularly interesting for a number of reasons. Besides being a more parsimonious model than equation 5.1,

it can be interpreted in terms of latent, normal variables as was the $RC(M)$ association model for 2–way tables given in Chapter 3. The only difference is that instead of two observed variables, there are now three. With respect to Bartholomew's (1980, 1987) latent (normal) variable model for categorical variables with three observed variables and either 1 or 2 latent normal variables, equation 5.3 with $M = 1$ (i.e., a consistent, unidimensional, bivariate log-bilinear model) can be used to estimate the underlying correlation matrix and the factor structure.[1]

The other latent variable model described in Chapter 3 can also be generalized. Specifically, if equation 5.3 with $M = M^*$ fits a 3–way table of variables $A$, $B$, and $C$, then there are $M^*$ latent normal variables $\{Z_m\}$ such that the observed variables $A$, $B$ and $C$ are independent given the latent variables $Z_1, \ldots, Z_{M^*}$; for all $m \neq m'$, $Z_m$ is independent of $Z_{m'}$ given $(a_i, b_j, c_k)$; and $Z_m|(a_i, b_j, c_k)$ is normally distributed with mean equal to $(\mu_{im} + \nu_{jm} + \eta_{km})$ and variance equal to a function of $\phi_m^{AB}$, $\phi_m^{AC}$ and $\phi_m^{BC}$ (Whittaker, 1989).

Just as the $RC(M)$ association model is implied when the data in a 2–way table arise from a discretized bivariate normal distribution (Goodman, 1981a, 1985, 1991), equation 5.3 with $M = 1$ is implied when the data in a table arise from a discretized trivariate normal distribution or from a joint distribution that can be transformed to a trivariate normal by applying separate transformations to each of the variables. The intrinsic association parameters for pairs of variables equal the corresponding elements of

[1]For example, one can to recover the correlation matrix and underlying structure of a discretized, trivariate normal distribution with 1 or 2 underlying factors by fitting equation 5.3 with $M = 1$ to the data.

the inverse the correlation matrix, which are proportional to the partial correlations. The correlations between variables in the underlying distribution can be estimated by computing the correlations between pairs of estimated scale values (e.g., $\hat{\rho}^{AB} = \sum_i \sum_j \hat{\mu}_i \hat{\nu}_j P_{ij}$).

The relationship between the $RC(1)$ association model and the bivariate normal distribution generalizes to the $p$-variate case for $p > 3$. This can be seen by noting similarity between the probability density function for a multivariate normal distribution and a unidimensional association model with consistency constraints on the scale values that includes a bilinear term for each pair of variables whose correlation is not equal to zero. This relationship can be used to compute integrals of multivariate normal distributions (Rom & Sarkar, 1990), as well as those of bivariate normal distributions (Becker, 1989b; Goodman, 1981a, 1985, 1991; Lancaster, 1957; Rom & Sarkar, 1990; Wang, 1987) .

#### 5.1.1.3 Generalizations to 4– and Higher–Way Tables

The bivariate interaction log-bilinear model, with or without consistency constraints, can easily be extended to 4– and higher–way tables simply by including more bilinear terms for the additional 2–way interactions. For example, the model with consistency constraints for a 4–way table of variables $A$, $B$, $C$, and $D$, which are indexed by $i$, $j$, $k$, and $l$, respectively, is

$$\ln(P_{ijkl}) = u^o_{ijkl} + \sum_{m=1}^{M} \quad (\phi^{AB}_m \mu^A_{im} \mu^B_{jm} + \phi^{AC}_m \mu^A_{im} \mu^C_{km} + \phi^{AD}_m \mu^A_{im} \mu^D_{lm} + \phi^{BC}_m \mu^B_{jm} \mu^C_{km} + \phi^{BD}_m \mu^B_{jm} \mu^D_{lm} + \phi^{CD}_m \mu^C_{km} \mu^D_{lm})$$

where $u^o_{ijkl} = u + u^A_i + u^B_j + u^C_k + u^D_l$. An example of such an extension can be found in Teeuwen (1988) who fits the special case of this model with $M = 1$, and the marginal probabilities as weights (i.e., $h^A_i = P_{i...}$, $h^B_j = P_{.j..}$, $h^C_k = P_{..k.}$, and $h^D_l = P_{...l}$) to the frequencies in a 4–way table of ordinal categorical variables.

### 5.1.2 Modeling the Peer Play Data

As an illustration of how to use these models to analyze data, 2–way interaction, log-bilinear models with the consistency constraint on the scale values were fit to the peer play data. Two sets of models were fit to the data from the 25 child sub-sample. In one set, equation 5.1 for different numbers of dimensions were fit. Specifically,

$$\ln(F_{ijk}) = u + u^G_i + u^P_j + u^T_k + \sum_{m=1}^{M} (\phi^{GP}_m \mu_{im}\nu_{jm} + \phi^{GT}_m \mu_{im}\eta_{km} + \phi^{PT}_m \nu_{jm}\eta_{km}) \qquad (5.4)$$

was fit for $M = 0, 1$, and 2 non-orthogonal, and 2 orthogonal dimensions. Since model 5.4 includes bilinear terms for each of the 2–way interactions, none of the 2–way margins were fit perfectly. However, due to the design of the study, the GT margin should be fit. The second set of models fit to the data included $u^{GT}_{ik}$ terms to ensure that the GT margin is fit. The second set consists of

$$\ln(F_{ijk}) = u + u^G_i + u^P_j + u^T_k + u^{GT}_{ik} + \sum_{m=1}^{M} (\phi^{GP}_m \mu_{im}\nu_{jm} + \phi^{PT}_m \nu_{jm}\eta_{km}) \qquad (5.5)$$

for $M = 0, 1, 2, 3, 4$, and 5 non-orthogonal dimensions, and for $M = 2$ orthogonal dimensions. Only the $GP$ and $PT$ partial associations are decomposed and analyzed by equation 5.5. Both of these models were fit by a *FORTRAN* program written to fit equations 5.4 and 5.5. The *FORTRAN* program is given in Appendix C, and the maximum likelihood equations and the algorithm implemented in the program for these models and other possible bivariate interaction, log-bilinear models are given in Appendix B.

The goodness-of-fit of the models to the data are evaluated in Section 5.1.2.1 and the results of the two "best" fitting models are presented in Sections 5.1.2.2 and 5.1.2.3.

#### 5.1.2.1 Model Evaluation: Goodness-of-Fit

The fit statistics for models 5.4 and 5.5 are reported in the upper and lower sections, respectively, of Table 5.1. The models in the top half of the table are equivalent to the joint $RC(M)$ association model fit to ${}_{TG}\mathbf{F}_P$ where the additive restriction given in equation 4.4 is imposed on the row scale values (i.e., $\mu_{ikm}^{GT} = \mu_{im}^{G} + \mu_{km}^{T}$ where $\mu_{im}^{G} = \mu_{im}$ and $\mu_{km}^{T} = \eta_{km}$). The first column of the table indicates the basic model, either $(G, P, T)$ or $(GT, P)$, which refer to equations 5.4 and 5.5, respectively. In the second column, the number of dimensions is specified, and for $M > 1$, whether the dimensions are orthogonal. When the dimensions are restricted, the fit of the model also depends on the weights, which are reported in the fourth column. The fourth column is blank for the models where the weights in the identification constraints are arbitrary, since they do not affect the fit of the model.

**Table 5.1**: Bivariate interaction, log-bilinear models fit to the 3-way table ($N = 25$).

| Model | Dimensions | Weights | *df* | $G^2$ | $G^2/df$ | p-value | $X^2$ |
|---|---|---|---|---|---|---|---|
| $(G, P, T)$ | 0 (loglinear) | | 364 | 483.27 | 1.328 | .000 | 467.25 |
| $(G, P, T)$ | 1 | | 344 | 385.18 | 1.120 | .062 | 365.62 |
| $(G, P, T)$ | 2 orthogonal | marg prob | 327 | 354.75 | 1.085 | .162 | 340.25 |
| $(G, P, T)$ | | unit | 327 | 352.21 | 1.077 | .140 | 339.88 |
| $(G, P, T)$ | 2 non-orthogonal | | 324 | 348.90 | 1.077 | .163 | 334.30 |
| $(GT, P)$ | 0 (loglinear) | | 340 | 446.24 | 1.312 | .000 | 422.96 |
| $(GT, P)$ | 1 | | 321 | 349.81 | 1.090 | .129 | 328.09 |
| $(GT, P)$ | 2 orthogonal | marg prob | 305 | 322.52 | 1.057 | .235 | 305.86 |
| $(GT, P)$ | | unit | 305 | 323.12 | 1.059 | .228 | 305.97 |
| $(GT, P)$ | 2 non-orthogonal | | 302 | 319.67 | 1.059 | .232 | 302.84 |
| $(GT, P)$ | 3 non-orthogonal | | 283 | 304.53 | 1.080 | .181 | 285.95 |
| $(GT, P)$ | 4 non-orthogonal | | 264 | 294.07 | 1.114 | .098 | 275.11 |
| $(GT, P)$ | 5 non-orthogonal | | 245 | 286.72 | 1.170 | .035 | 270.19 |

The best possible fit that can potentially be achieved by either equation 5.4 or 5.5 is the same as that of the no 3–way interaction loglinear model $(GT, GP, PT)$ (i.e., $G^2 = 278.07$ and $X^2 = 260.60$, $df = 240$; see table 2.7), which is the lower bound of the fit statistics for these bivariate interaction log-bilinear models. To achieve a fit as good as that of the $(GT, GP, PT)$ loglinear model, the number of dimensions needed is so large that the degrees of freedom for equations 5.4 or 5.5 would be smaller than the degrees of freedom for the $(GT, GP, PT)$ loglinear model (i.e, $df < 240$).

The largest improvements in fit relative to the degrees of freedom occur when the first dimension is added (i.e., $(G, P, T)$ versus $(G, P, T)$ with $M = 1$, and $(GT, P)$ versus $(GT, P)$ with $M = 1$). The fit statistics for equation 5.4 with $M = 1$ indicate that this model marginally fits the data (i.e., $G^2 = 385.18$ and $df = 344$, or $G^2/df = 1.120$), and the models with $M = 2$ appear to fit reasonably well (i.e., $G^2/df = 1.085$, and

1.077). The models in the top half of the table are not considered any further, because they do not fit the GT margin perfectly. The GT association, which is not substantively interesting, is important for substantive and design reasons.

The model $(GT, P)$ with $M = 1$ dimension is the simplest one that provides a reasonable fit. The fit relative to degrees of freedom is slightly better for the 2 dimensional models. Both the 1 and 2 dimensional models are examined in more detail below in Sections 5.1.2.2 and 5.1.2.3, respectively. Since there is only a slight difference between the orthogonal and non-orthogonal models in terms of goodness-of-fit and estimated parameter values, the results from the model with 2 orthonormal dimensions are presented here.

#### 5.1.2.2 The One Dimensional Model

The estimated scale values of equation 5.5 for $M = 1$ and unit weights are plotted in Figure 5.1. The group, play quality, and occasion scale values are plotted in the top, middle, and bottom parts, respectively, of the figure. The group and occasion scale values are multiplied by their respective intrinsic association parameters, $\hat{\phi}^{GP} = 1.754$ and $\hat{\phi}^{PT} = .502$, to reflect the difference in the relative size or importance of the GP and the PT interactions. The fact that $\hat{\phi}^{GP}$ is three times larger than $\hat{\phi}^{PT}$ indicates that the GP interaction is more than three times stronger or larger than the PT association.

Taken together, the top and middle scales in Figure 5.1 represent the GP interaction, and the bottom and middle scales represent the TP interaction. The relative order and spacing of the scales values for the groups and play qualities are very similar to the order

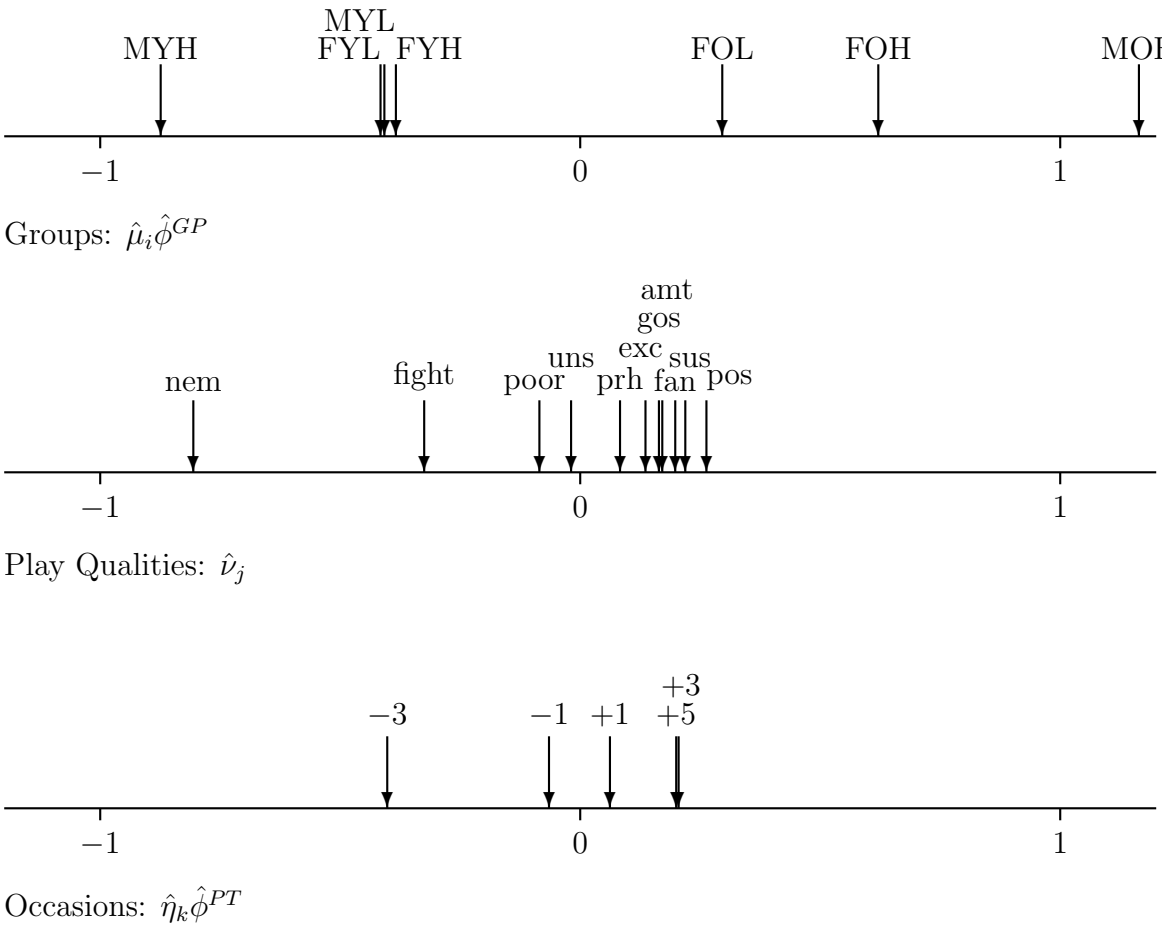



**Figure 5.1**: Estimated scale values from the fit of the $(GT, P)$ bivariate interaction, log-bilinear model with $M = 1$ dimension and unit weights.

and spacing of the scale values in Figure 4.1, which are from the $RC(1)$ association model fit to the $\mathbf{F}_{GP+}$ marginal table. This result is expected, because the conditional loglinear model $(GT, GP)$ comes close to providing an adequate fit, which implies that the GP partial and marginal associations are similar. The $(GT, GP)$ model picks up a large part of the structure (association) in the data.

The relative order of the scale values for occasions is the same as that of the $\phi_{(k)}$ parameters from the conditional $RC(1)$ association model fit to $\{\mathbf{F}_{GP|1}, \ldots, \mathbf{F}_{GP|5}\}$ with homogeneity constraints on the row (groups) and column (play qualities) scale values; however, the interpretation of the parameters from equation 5.5 is different from the interpretation of the parameters of the conditional association model. The $\hat{\phi}_{(k)}$'s from the conditional $RC(1)$ association model indicate that the strength of the relationship between groups and play qualities decreases over time. Since the $\hat{\phi}_{(k)}$'s are different for the different sessions, there is a 3–way interaction. The scale values for occasions in Figure 5.1 (i.e., $\hat{\eta}_k$) indicate that more of the immature qualities were observed at the earlier sessions while more of the mature qualities were observed at the later sessions. The model does not include a 3–way interaction, and only the bivariate interactions between groups and play (GP) and between play and occasions (PT) are analyzed. The conditional association model analyzes the combined effects of the GP and the GPT interactions, and fits the PT margin perfectly.

#### 5.1.2.3 The Two Dimensional Model

The bivariate interaction log-bilinear model with 2 orthonormal dimensions will now be presented. In the scale value plots for this model, the estimated scale values are multiplied by the square root of the appropriate intrinsic association parameter. The estimated intrinsic association parameters for the GP interaction are $\hat{\phi}_1^{GP} = 1.824$ and $\hat{\phi}_2^{GP} = .545$, and the estimates for the PT interaction are $\hat{\phi}_1^{PT} = .495$ and $\hat{\phi}_2^{PT} = .259$. Since there are now two dimensions, the scale value plots are more complex, but they reveal some interesting aspects about the associations in the data.

While the PT interaction is not very important in terms of the overall fit of the model, the scale values for occasions are rather interesting and suggest that there may be some systematic variation between occasions and play qualities. In Figure 5.2, the scale values for occasions are plotted against time, measured in terms of months before and after the birth of the children's siblings. The first dimension ($m = 1$) appears to be a "gain" factor or a linear component in which the scale values increases across time and level off by the last session. The scale values for the occasions immediately preceding and following the birth (i.e., $-1$ and $+1$ month) are close to zero. This is the same pattern that was obtained with the unidimensional model, and it is the stronger or more important component in the PT association (i.e., $\hat{\phi}_1^{PT} > \hat{\phi}_2^{PT}$). The second dimension ($m = 2$) appears to be a "change" dimension or a quadratic component. The scale values for the sessions immediately preceding and following the birth are the largest values,

negative and positive, respectively, while the scale values for the other three sessions are all close to zero.

The nature of the change in play qualities across occasions is represented in Figure 5.3, which is a joint plot of the scale values for occasion and play qualities. The configuration of the scale values for the play qualities is such that first dimension orders the scale values in terms of maturity ("bad" versus "good" qualities), and the second dimension primarily contrasts the quality unsustain (uns) versus the two qualities amity (amt) and gossip (gos). The relative order and spacing of the play qualities on the first dimension is nearly identical to that of the unidimensional model. At the first session ($-3$ months), the children's play is characterized by more negative emotion (nem) and fighting with fewer positive or mature qualities than would be expected if there was no partial association between play and occasion. The opposite pattern is observed for the last two sessions. Just before the birth, the children's peer play tends to be more unsustained (uns) and the children tend not to show amity (amt) or engage successfully in gossip (gos). The qualities that characterize the children's peer play at the session one month after the birth is the opposite of the pattern observed one month before the birth.

Figure 5.4 is the joint plot of the scale values for the groups and play qualities. Since the $GP$ association is approximately three times larger than the PT interaction, the configuration of points in Figure 5.4 is larger and more spread out than the points in Figure 5.3. The relative ordering and spacing of the groups on the first dimension are very similar to the ordering and spacing of the groups in the one-dimensional model. Taking into account the second dimension, the nature of the interaction between gender, age, and

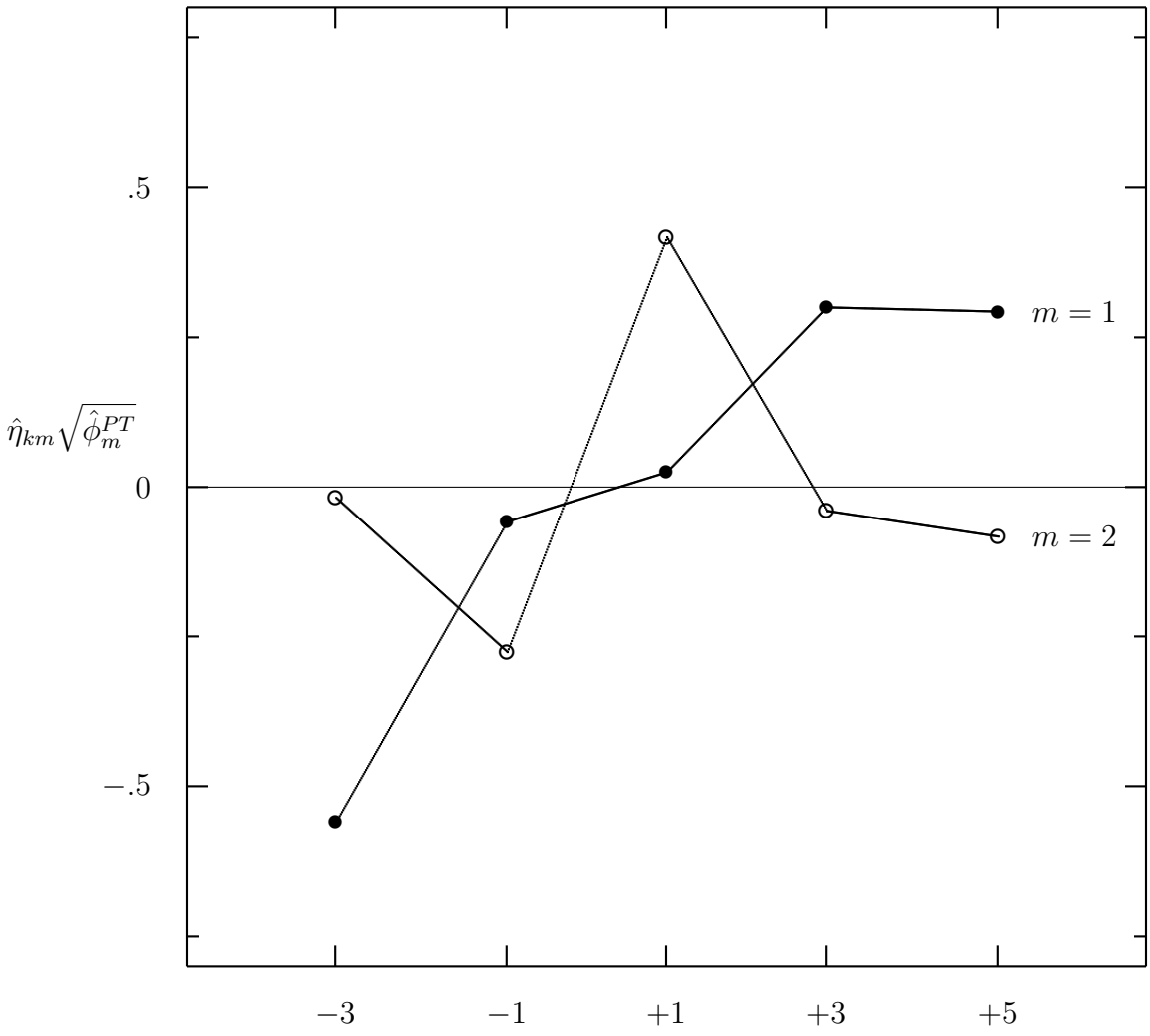



**Figure 5.2**: Scale values for occasions from the $(GT, P)$ bivariate interaction, log-bilinear model with 2 orthonormal dimensions.

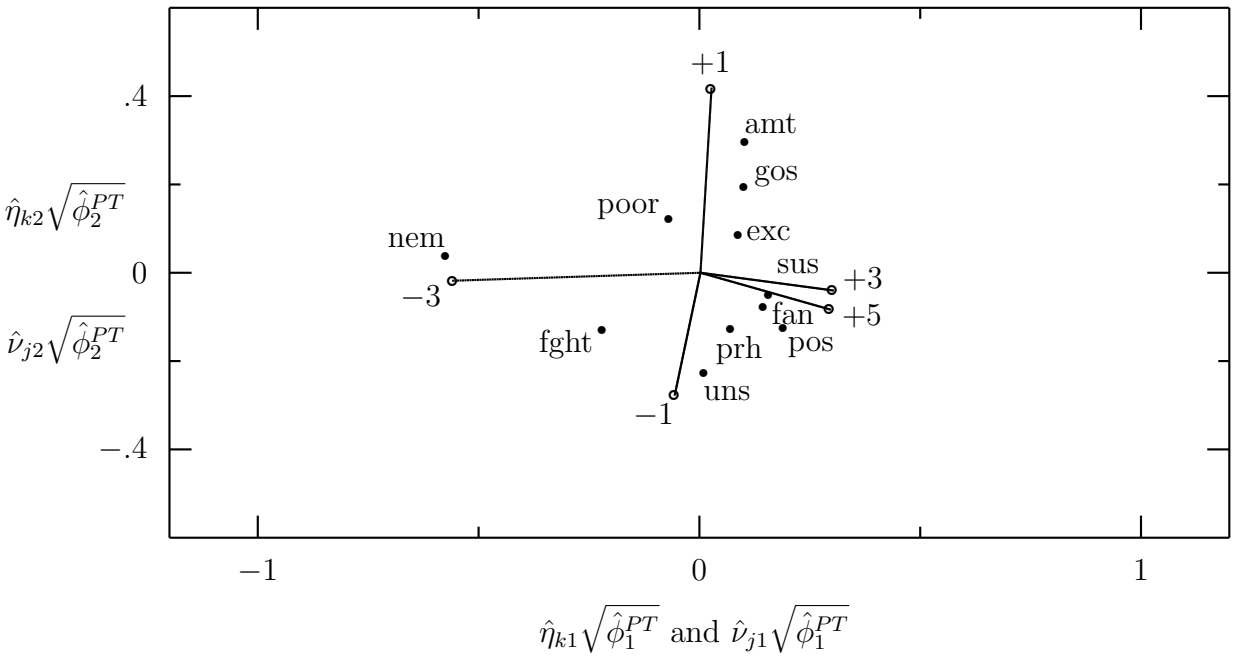



**Figure 5.3**: Joint plot of scale values for occasions and play qualities from the $(GT, P)$ bivariate interaction, log-bilinear model with 2 orthonormal dimensions ($N = 25$).

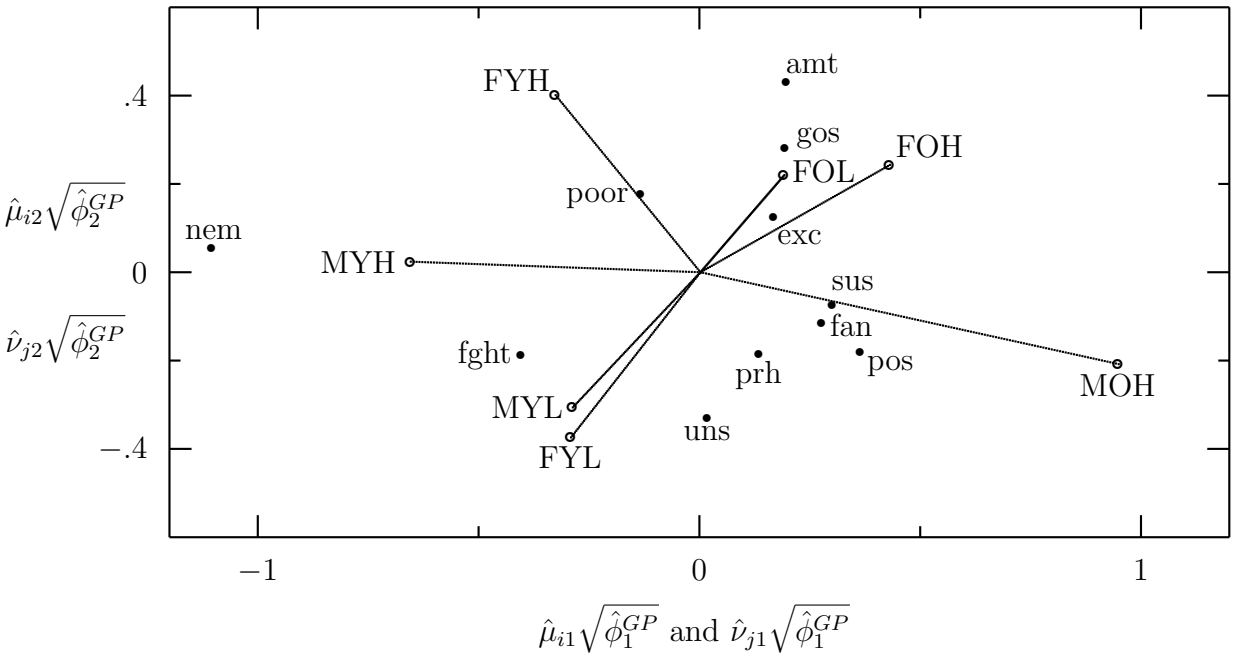



**Figure 5.4**: Joint plot of scale values for group and play qualities from the $(GT, P)$ bivariate interaction, log-bilinear model with 2 orthonormal dimensions ($N = 25$).

sibling acceptance with respect to play qualities can be discerned by noting which vectors connecting the origin to the points representing groups point in similar directions. The vectors for the groups MYL and FYL point in the same direction, which is a different direction (and quadrant of the plot) than the vectors for the groups MYH and FYH point. This implies that for the younger children, given sibling acceptance, there is not much difference between the play qualities exhibited by the males and females. For the older children, the vectors for FOL and FOH point in the same general direction, which is a different direction (and quadrant) than the vector for MOH. There is a difference between the genders for the older children, but no difference between the high and low sibling acceptance groups.

The nature of the interaction between groups and play qualities is represented in Figure 5.4. The qualities that are *relatively* closer to particular groups tend to be exhibited more in those groups compared to the other groups and qualities than would be expected if there was no GP partial association. For example, FOL and FOH tend to show more amity (amt), excitement (exc), and gossip (gos) and less negative emotion (nem), conflict (fght), and unsustained behavior (uns) than the other groups. Since the vector for MOH lies in the same direction as sustain (sus), successful fantasy (fan), and positive play (pos), these are play qualities that are more strongly associated with the older males who show high sibling acceptance.

While the analyses using the bivariate interaction, log-bilinear model are interesting and useful, they are not sufficient. The bivariate models do not represent or analyze the 3–way interaction. There certainly appears to be a 3–way association in the peer

play data. The results from the analyses in this section could be misleading; further, the models could have missed some interesting associations in the data. Since the models discussed in this section cannot represent changes over time in the relationship between play qualities and groups, attention is now turned to models that do analyze higher–way interactions. These models are reviewed and discussed in the next section.

## 5.2 Higher–Way Interaction Models

Two basic methods of incorporating higher–way interactions into association or correlation models have been proposed. In the simpler case, higher–way interactions are represented by bilinear terms. The conditional and joint $RC(M)$ models are special cases of these models. Two–way decompositions can lead to complex graphical representations, which may be difficult to interpret. When associations among three or more variables are analyzed by 2–way decompositions, scale values are assigned to combinations of the categories of (at least) some of the variables. With a 3–way decomposition, the categories of each variable or mode are assigned scale values, and the associations among them represented. In the peer play data where the three modes correspond to groups, play qualities and occasions, the serial dependencies and the associations between groups and play qualities can be simultaneously analyzed.

The properties of 3– and higher–way decompositions of arrays (not necessarily contingency tables) are more complex and are less well known than the 2–way, singular value decomposition of matrices (Coppi & Bolasco, 1989). Generalizing the decomposition of

2–way matrices to 3– and higher–way arrays is not straightforward, because not all of the properties of 2–way decompositions can be simultaneously extended (Kroonenberg, 1989; Kruskal, 1989; Franc, 1989; Denis & Dhorne, 1989; D'Aubigney & Polit, 1989). Several generalizations have been proposed that differ with respect to which aspects of 2–way decompositions are retained. The two most well known generalizations are Tucker's 3–mode model (Tucker, 1963, 1964, 1966; Kroonenberg, 1983, 1984; Bloxom, 1984; Law, Synder, Hattie, & McDonald, 1984) and the canonical decomposition model, CANDECOMP (Carroll & Chang, 1970; Kruskal, 1984), which is equivalent to the parallel factors model, PARAFAC (Harshman & Lundy, 1984; Kruskal, 1984). In subsection 5.2.1, 3–way decompositions are discussed with the emphasis on Tucker's 3–mode model.

In Sections 5.2.2 and 5.2.3, association and correlation models, respectively, that include 3– and higher–way interactions are reviewed. Generalizations of correlation models, in particular, generalizations of correspondence analysis, are included in section 5.2.3. One of these generalizations is based on the use of Tucker's 3–mode model as the 3–way decomposition.

### 5.2.1 Three–Way Decomposition Models

Two 3–way decomposition models, Tucker's 3–mode model, which was labeled "Tucker3" by Kroonenberg and de Leeuw (1980), and the canonical decomposition model (CANDECOMP) are described here. These models have been used to analyze continuous data, and in a limited number of instances, to analyze categorical data as well (e.g., Kroonenberg, 1983, 1989; Choulakian, 1988a, 1988b; Iacobucci, 1989). These analyses have been pri-

marily exploratory and descriptive. Of particular interest are the applications of Tucker's 3–mode model for the analysis of longitudinal data (Kroonenberg, 1983; Kroonenberg, Lammers & Stoop, 1985; Iacobucci, 1989). Tucker (1963) presents his 3–mode model as a way to generalize factor analysis to situations where the data are cross-classified into 3–ways or modes. He describes its use to analyze longitudinal, repeated measures data. Other researchers have recognized the potential of Tucker's 3–mode model as a model for the measurement of change and the analysis of longitudinal data (e.g., Kroonenberg, 1983; Visser, 1982), including the analysis of developmental data (Bentler, 1973).

There have been relatively few applications of Tucker's 3–mode model. This is due partially to the computational demands and problems associated with the estimation of the parameters of the model, and to the complexity of the model. Recent increases in computational power and the development of algorithms for estimating the model (Kroonenberg & de Leeuw, 1980; Kroonenberg, 1983; Kiers, Kroonenberg & ten Berge, 1992) have helped solve the first problem. There is still a lack of familiarity and understanding of the model by researchers. With many of the computational problems solved, it will be easier for researchers to use the model, which should lead to more applications of the model. Increased applications will, in turn, lead to a great familiarity and understanding of the model by researchers.

In Section 5.2.1.1, Tucker's 3–mode principal components model (or "Tucker3", for short) is presented, and in Section 5.2.1.2, special cases of it are presented.

#### 5.2.1.1 Tucker's 3–Mode Model

Tucker3 is presented here as a generalization of 2–way singular value decomposition (SVD) or standard principal components analysis, rather than as a factor analytic or statistical model. Three equivalent ways of presenting Tucker's 3–mode model are given here, because each way of expressing the model helps to point out particular properties and interpretations of the model. First, the model is presented in terms of individual elements of a 3–way table, next it is expressed in matrix form, and lastly, the 3–way decomposition is presented in terms of vectors . After reviewing the properties of Tucker3, various plots of scale values are described. These plots are useful representations and summaries of the data.

Let $x_{ijk}$ represent an element of a 3–way array where $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K$. The observed values $x_{ijk}$ are decomposed into $R$, $S$, and $T$ components of three "ideal" or "intrinsic" modes as follows

$$x_{ijk} = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} g_{rst} a_{ir} b_{js} c_{kt} \tag{5.6}$$

where $a_{ir}$, $b_{js}$, $c_{ks}$ are scores or loadings of the $i$, $j$, and $k$ levels of the "extrinsic" or observed variables on the $r$, $s$, and $t$, respectively, intrinsic modes or components, and $g_{rst}$ is the $(r, s, t)$ element of the "core matrix" (Tucker, 1963, 1964, 1966; Kroonenberg, 1983, 1989). The vectors of scores are orthonormal (i.e., $\mathbf{a}_r'\mathbf{a}_{r'} = \delta_{rr'}$, $\mathbf{b}_s'\mathbf{b}_{s'} = \delta_{ss'}$, and $\mathbf{c}_t'\mathbf{c}_{t'} = \delta_{tt'}$). The $g_{rst}$'s reflect the structure or relationship among the components. Tucker3 differs from 2–way SVD in that the singular values or core elements (i.e., $g_{rst}$)

are 3 dimensional and some of them can be negative (Kroonenberg, 1983, 1989). In the 2–way case, the arbitrary constraint that the singular values are non-negative is typically imposed. The Tucker3 decomposition always exists and is complete for $R = I$, $S = J$, and $T = K$ (i.e., equality holds in equation 5.6, $x_{ijk} = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} g_{rst} a_{ir} b_{js} c_{kt}$). Lower-rank approximations of the 3–way table can be found for $R < I$, $S < J$, and/or $T < K$ (i.e, $x_{ijk} \approx \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} g_{rst} a_{ir} b_{js} c_{kt}$).

With respect to longitudinal data, Tucker's model analyzes serial dependencies in terms of the intrinsic modes or components. When Tucker (1963) introduced equation 5.6 as a possible generalization of factor analysis to 3–way classifications, he illustrated the meaning and interpretation of the decomposition with a fictitious example where the three modes corresponded to individuals, variables, and occasions. This interpretation is extended here to the peer play data set where the first, second and third modes correspond to groups, play qualities and occasions, respectively. The decomposition can be expressed as

$$
\begin{aligned}
x_{ijk} &= \sum_{s} b_{js} \sum_{r} a_{ir} \sum_{t} g_{rst} c_{kt} && (5.7) \\
&= \sum_{s} b_{js} \sum_{r} g^{*}_{rsk} a_{ir} \\
&= \sum_{s} g^{**}_{isk} b_{js} && (5.8) \\
& && (5.9)
\end{aligned}
$$

where $g^{*}_{rsk} = \sum_{t} g_{rst} c_{kt}$ and $g^{**}_{isk} = \sum_{r} g^{*}_{rsk} a_{ir}$. The $a_{ir}$'s correspond to the group components and relate the $I$ observed groups to the $R$ ideal or intrinsic groups. The $b_{js}$'s

correspond to the play quality components and relate the $J$ observed qualities to the $S$ ideal or intrinsic play qualities. The $c_{kt}$'s are interpreted as components of change and/or stability that relate the observed $K$ occasions to the $T$ intrinsic occasions. The $g_{rst}$'s give the relationship among the components. The quantity $g^{*}_{rsk}$ is interpreted as the scale value of intrinsic group $r$ at time point $K$, and the quantity $g^{**}_{isk}$ is the scale value or score of the $i$th observed group on the $k$th occasion on play component $s$. The $g^{**}_{isk}$'s show how the children's behavior changes over time. This is the desired theoretical conception regarding change for this data set; that is, the behavior of the children is considered to change and not the play qualities.

In matrix form, the Tucker3 decomposition is

$$ {}_{A}\mathbf{X}_{BC} = \mathbf{A}({}_{A}\mathbf{G}_{BC})(\mathbf{B}' \otimes \mathbf{C}') \tag{5.10} $$

where $\otimes$ is the Kronecker product, ${}_{A}\mathbf{X}_{BC}$ is the 3–way table written as an $(I \times JK)$ matrix, ${}_{A}\mathbf{G}_{BC}$ is the 3–way core matrix written as an $(R \times ST)$ matrix, $\mathbf{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_R\}$, $\mathbf{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_S\}$, and $\mathbf{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_T\}$. Analogous equations can also be written for the two other ways of arranging the elements of the 3–way table into 2–way matrices, ${}_{B}\mathbf{X}_{AC}$ and ${}_{C}\mathbf{X}_{AB}$.

When equality holds in equation 5.6 (i.e., the decomposition is complete with $R = I$, $S = J$ and $T = K$, or the decomposition is exact for $R < I$, $S < j$ and/or $T < K$), then the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are the eigenvectors or canonical solutions of the matrix products $({}_{A}\mathbf{X}_{BC})({}_{BC}\mathbf{X}_{A})$, $({}_{B}\mathbf{X}_{AC})({}_{AC}\mathbf{X}_{B})$, and $({}_{C}\mathbf{X}_{AB})({}_{AB}\mathbf{X}_{C})$, respectively. If the de-

composition is not complete, then $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are not canonical solutions, and the eigenvectors of the matrix products $({}_A\mathbf{X}_{BC})({}_{BC}\mathbf{X}_A)$, $({}_B\mathbf{X}_{AC})({}_{AC}\mathbf{X}_B)$, and $({}_C\mathbf{X}_{AB})({}_{AB}\mathbf{X}_C)$ are not least squares estimates of the components. The "best" least squares approximation (i.e., $\sum_i \sum_j \sum_k (x_{ijk} - \hat{x}_{ijk})^2$ is minimized for $\hat{x}_{ijk} = \sum_r \sum_s \sum_t \hat{g}_{rst}\hat{a}_{ir}\hat{b}_{js}\hat{c}_{kt}$) always exists for $R < I$, $S < J$, and/or $T < K$ (Kroonenberg, 1983, 1989). Alternating least squares algorithms for computing estimates are described by Kroonenberg and de Leeuw (1980), Kroonenberg (1983), and Kroonenberg and Brouwer (1985a, 1985b). Unlike 2–way SVD, the approximate solutions of Tucker3 are not nested (Kroonenberg, 1983; d'Aubigny & Polit, 1989). For example, the least squares solution for $R = S = T = 2$ is not necessarily contained in the one for $R = S = 2$ and $T = 3$.

A third way of expressing Tucker3 is in terms of vectors and "tensors," which are Kronecker or "outer" products of vectors:

$$\text{vec}(X) = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} g_{rst}(\mathbf{a}_r \otimes \mathbf{b}_s \otimes \mathbf{c}_t) \tag{5.11}$$

where $\text{vec}(X)$ is an $(IJK \times 1)$ vector with elements $x_{ijk}$ arranged such that the index $k$ changes the fastest and $i$ the slowest. The tensors are not necessarily orthogonal; that is, $(\mathbf{a}_r \otimes \mathbf{b}_s \otimes \mathbf{c}_t)$ is not necessarily orthogonal to $(\mathbf{a}_{r'} \otimes \mathbf{b}_{s'} \otimes \mathbf{c}_{t'})$ for $r \neq r'$, $s \neq s'$, and/or $t \neq t'$ (Franc, 1989). A minimal decomposition is one that completely decomposes or approximates the data to a certain level (i.e., goodness-of-fit) into the smallest possible number of tensors. In the case of 2–way SVDs, orthogonal tensor decompositions are minimal; however, for 3– and higher–way decompositions, orthogonal tensor decomposi-

tions are not necessarily minimal (Franc, 1989; Denis & Dhorne, 1989). If it is desired to reduce or summarize the information in a 3–way array into the smallest number of tensors or components as possible, then non-orthogonal tensor decompositions are preferable.

The relationship among the levels or objects of each mode can be represented graphically by plotting the scale values. There is a rotational indeterminacy with Tucker3 decompositions, just as there is for 2–way SVDs. The configuration of points representing the objects of each mode is the same for all orthogonal rotations of the axes.

Joint plots where scale values of objects from two of the modes are plotted in the same space show the relationship between the two modes on the components of the third mode. For example, suppose that for each component $s$ of the variable (or mode) $B$, the inner products between $A$ and $C$ are computed, and the singular value decomposition of the inner product matrix for each $s$ is found. Assuming $T \leq R$,

$$\begin{aligned} \mathbf{A}\mathbf{G}_s\mathbf{C}' &= \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{V}'_s \\ &= (\mathbf{U}_s\mathbf{\Lambda}_s^{\gamma})(\mathbf{\Lambda}_s^{\delta}\mathbf{V}'_s) \\ &= \mathbf{A}_\mathbf{s}^*\mathbf{C}_\mathbf{s}^{*\prime} \end{aligned}$$

where $\mathbf{G}_s$ is the $(R \times T)$ matrix of $g_{rst}$ for fixed $s$, the columns of $\mathbf{U}_s$ and $\mathbf{V}_s$ are left and right (orthogonal) singular vectors, respectively, $\mathbf{\Lambda}_s$ is the $(T \times T)$ diagonal matrix of singular values, and $\gamma + \delta = 1$. In all of the examples in this thesis, $\gamma = \delta = 1/2$. The columns of $\mathbf{A}_s^*$ and $\mathbf{C}_s^*$ are the values that are plotted in the same figure. Numerous examples of such plots can be found in Kroonenberg (1983), and examples of joint plots

from analyses of the peer play data are provided later in this chapter and in the next chapter.

Since one of the modes of the peer play data is time, plots of the inner products between groups and occasions are also interesting and useful. For example, if $A$, $B$, and $C$ correspond to groups, play qualities, and occasions, respectively, then the rows of the $(I \times K)$ matrix of inner products $\mathbf{AG}_s\mathbf{C} = \{\sum_r \sum_t a_{is} g_{rst} c_{kt}\}$, which correspond to the groups, can be plotted against the columns, which correspond to the occasions. These inner products are the $g^{**}_{isk}$'s in equation 5.9, which are interpretable as group scale values at each time point on the play components. Such figures show how the groups change across time on play component $s$. Examples of such plots are given later.

#### 5.2.1.2 Special Cases of Tucker3

A special case of Tucker3 where only two of the three ways of an array are reduced is useful in situations where the third (non-reduced) way or mode corresponds to time or to a group variable that has relatively few levels (Kroonenberg, 1983; Kroonenberg & Brouwer, 1985a). Since one of the three modes is not reduced, this decomposition is often referred to as an "extended" 3–mode analysis or "Tucker2." For example, for $R < I$, $S < J$, and $T = K$, Tucker2 is

$$x_{ijk} = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{K} g_{rst} a_{ir} b_{js} c_{tk} = \sum_{r=1}^{R} \sum_{s=1}^{S} g^{*}_{rsk} a_{ir} b_{js} \tag{5.12}$$

where $g^*_{rsk} = \sum_t g_{rst} c_{tk}$. The properties associated with Tucker2 are essentially the same as those for Tucker3. By not reducing the third mode, comparisons of the relationship between the other two modes of the array can be made for each level of the third mode. This aspect of Tucker2 is similar to the conditional $RC(M)$ association model, but the structural hypotheses that can be examined by the two models are somewhat different. With a conditional $RC(M)$ model, equality of the strength and/or the structure across time can be investigated by placing homogeneity constraints on the intrinsic association parameters and/or the scale values. With Tucker2, strength, structure and/or the relationship among the components are the same or different across time. In this sense, Tucker2 (as well as Tucker3) is a more general model than the conditional approach.

The other major 3–way decomposition, CANDECOMP or PARAFAC (Carroll & Chang, 1970; Kruskal, 1984, 1989; Harshman & Lundy, 1984; Kruskal, Harshman & Lundy, 1989; Lundy, Harshman & Kruskal, 1989; Denis & Dhorne, 1989), is generally given as

$$x_{ijk} = \sum_{m=1}^{M} a^*_{im} b^*_{jm} c^*_{km} \tag{5.13}$$

but can also be written as

$$x_{ijk} = \sum_{m=1}^{M} g_m a_{im} b_{jm} c_{km} \tag{5.14}$$

where $a^*_{im} = g_m^{1/3} a_{im}$, $b^*_{jm} = g_m^{1/3} b_{jm}$, and $c^*_{km} = g_m^{1/3} c_{km}$. From equation 5.14, it can be seen that CANDECOMP is a special case of Tucker3 where $R = S = T = M$ and the dimensions are uncorrelated (i.e., $g_{rst} = 0$ for $r \neq s \neq t$). Given the specific diagonal structure imposed on the core matrix in Tucker3, CANDECOMP is sometimes referred

to as a "diagonal decomposition" (Denis & Dhorne, 1989). For $R = S = T = 1$, Tucker3 and CANDECOMP are equivalent.

CANDECOMP does not possess the same properties as Tucker3 (Kroonenberg, 1983, 1989; Kruskal, 1984, 1989; Harshman & Lundy, 1984; Kruskal, Harshman & Lundy, 1989; Lundy, Harshman & Kruskal, 1989; Denis & Dhorne, 1989). It retains the properties of Tucker3 regarding the non-nesting of approximate solutions, and the non-orthogonality of the tensors. Unlike Tucker3, the vectors of scale values are not necessarily orthogonal (e.g., $\mathbf{a_m}'\mathbf{a_{m'}}$ does not necessarily equal 0 for $m \neq m'$); however, there is no rotational indeterminacy of the axes. Degenerate solutions where dimensions are highly negatively correlated are a problem for CANDECOMP (Kruskal, Harshman & Lundy, 1989; Lundy, Harshman & Kruskal, 1989). Furthermore, the complete and/or approximate decomposition of a 3–way array by CANDECOMP does not always exist (Kroonenberg, 1983, 1989; Denis & Dhorne, 1989). Even when a decomposition does exist, the number of dimensions needed to completely decompose an array by CANDECOMP is unknown and difficult to determine (Kruskal, 1989; Kroonenberg, 1989). There is no algorithm for computing the number of dimensions, which Kruskal (1989) discusses as one possible definition of the rank of a 3–way array.

Other possible trilinear decompositions, such as Denis & Dhorne's (1989) "rocket form" and Yoshizawa's (1988) singular value decomposition of multiarray data (or "nested ANOVA"), are not considered here, because they do not necessarily reduce or summarize the information in multiway tables in a way that is readily interpretable. These decompositions emphasize complete decompositions and retain the properties of orthogonal

tensors and nested solutions, which are characteristics of 2–way singular value decompositions.

### 5.2.2 Association Model Generalizations

Becker (1989) proposed the following general model for modeling associations in 3–way contingency tables:

$$\ln(P_{ijk}) = u^o_{ijk} + \sum_{m=1}^{M_1} \phi^{AB(C)}_{m(k)} \mu^{A(C)}_{im(k)} \nu^{B(C)}_{jm(k)} + \sum_{m=1}^{M_2} \phi^{AC(B)}_{m(j)} \mu^{A(B)}_{im(j)} \eta^{C(B)}_{km(j)} + \sum_{m=1}^{M_3} \phi^{BC(A)}_{m(i)} \nu^{B(A)}_{jm(i)} \eta^{C(A)}_{km(i)} \tag{5.15}$$

where $u^o = u + u^A_i + u^B_j + u^C_k$, $M_1 \leq \min(I, J) - 1$, $M_2 \leq \min(I, K) - 1$, and $M_3 \leq \min(J, K) - 1$. Obviously, this model is too general to be useful in practice; however, this general model defines a family of association models. Possible simplifying constraints include consistency or homogeneity constraints on the intrinsic association parameters and/or the scale values, setting some of the intrinsic association parameters to zero, and using only a small number of dimensions (e.g., unidimensional, $M_1 = M_2 = M_3 = 1$). The choice of simplifying constraints determines which 2–way and 3–way interactions are represented and the form of the representation.

Special cases of 5.15 include various loglinear models, all of the association models for 3–way tables that have been discussed so far, and other potentially interesting models. The conditional and joint $RC(M)$ association models (equations 4.8 and 4.2, respectively), and the no 3–way association model (equation 5.1) are all special cases

of 5.15. The models in the family defined by equation 5.15 include hierarchical and non-hierarchical models.

With respect to modeling 3–way associations, Becker (1989a) also proposes a model in which a trilinear term is used to represent the 3–way interaction; namely,

$$\ln(P_{ijk}) = u^{o}_{ijk} + u^{AB}_{ij} + u^{AC}_{ik} + u^{BC}_{jk} + \phi^{ABC}\mu^{A}_{i}\nu^{B}_{j}\eta^{C}_{k} \tag{5.16}$$

As with other association models, this model represents the log–odds ratios in terms of the product of intrinsic association parameters and differences between scales values,

$$\ln(\Theta_{ii',jj',kk'}) = \phi^{ABC}(\mu^{A}_{i} - \mu^{A}_{i'})(\nu^{B}_{j} - \nu^{B}_{j'})(\eta^{C}_{k} - \eta^{C}_{k'}) \tag{5.17}$$

While Becker (1989) mentions the possibility of using such a model, he does not provide any examples. Others have also pointed out this or similar possibilities (e.g., Goodman, 1986; Gower, 1989), but few have developed or used models that include trilinear terms.

One exception is Choulakian's (1988a, 1988b) generalization of the $RC(M)$ association model. This model includes both bilinear terms, which represent the 2–way interactions, and trilinear terms, which represent the 3–way interaction. The scale values are constrained to be the same (consistent) across the different interactions. In terms of the notation used in this thesis, Choulakian's (1988a) generalization is

$$\ln(P_{ijk}) = u^{o}_{ijk} + \sum_{m=1}^{M}(\phi^{AB}_{m}\mu_{im}\nu_{jm} + \phi^{AC}_{m}\mu_{im}\eta_{km} + \phi^{BC}_{m}\nu_{jm}\eta_{km} + \phi^{ABC}_{m}\mu_{im}\nu_{jm}\eta_{km}) \tag{5.18}$$

This model yields an equation for the log–odds ratios of $2 \times 2 \times 2$ subtables similar to equation 5.17, except that it is multidimensional. However, even when $M = 1$, the representation of the conditional odds ratios given by equation 5.18 is more complex than that given by equation 5.16. The conditional odds ratio based on model 5.16 is

$$\ln(\Theta_{ii',jj'(k)}) = \phi^{ABC}\eta_k^C(\mu_i^A - \mu_{i'}^A)(\nu_j^B - \nu_{j'}^B) \tag{5.19}$$

while the same odds ratio based on model 5.18 for $M = 1$ is

$$\ln(\Theta_{ii',jj'(k)}) = (\phi_1^{AB} + \phi_1^{ABC}\eta_{k1}^C)(\mu_{i1}^A - \mu_{i'i}^A)(\nu_{j1}^B - \nu_{j'1}^B) \tag{5.20}$$

With Choulakian's model, the conditional odds ratios are a function of both the 2– and 3–way association parameters.

Model 5.18 is different from 5.16 in other ways. Model 5.18 will not necessarily fit the 2–way margins, while 5.16 will fit all of them. This aspect of Choulakian's (1988a) model is undesirable in cases where any of the 2–way margins are fixed by design. Choulakian's model can be altered by adding interaction terms to equation 5.18 such that certain margins are fit perfectly and the intrinsic association parameters corresponding to these set equal to zero. For example, to fit the $AB$ margin, interaction terms $u_{ij}^{AB}$ would be added to equation 5.18, and the $\phi_m^{AB}$ parameters fixed at zero.

When all 2–way margins are fit, Choulakian's model only includes trilinear terms and the conditional odds are given by a multidimensional extension of equation 5.17. While

the model can be altered to fit particular margins, the model has other undesirable characteristics. Unless all of the 2–way margins are fit by including $u$-terms for each of them, both bilinear and trilinear terms are included in the model. The 3–way decomposition is a "diagonal" or CANDECOMP decomposition (Carroll & Chang, 1970; Kruskal, 1984, 1989; Harshman & Lundy, 1984; Denis & Dhorne, 1989), which has the properties and problems discussed in section 5.2.1. The same number of dimensions or components are estimated for each of the variables. For some data sets, fewer components may be required for one or two of the variables. The problems associated with CANDECOMP decompositions are likely to arise with equation 5.18.

The above considerations along with the possibility of using Tucker's 3–mode principal components model for the 3–way decomposition are the primary motivations for the association model generalizations developed in Chapter 6.

### 5.2.3 Correlation Model Generalizations

Only two correlation model generalizations have been explicitly proposed for analyzing higher–way interactions in contingency tables. Both of these are generalizations of correspondence analysis (CA) to 3–way tables. These proposals retain different aspects of CA of 2–way tables. Choulakian's (1988a) generalization retains the dual scaling or reciprocal averaging aspect of CA. In his generalization, the scale values assigned to the categories of each variable are functions of the scales values assigned to the other two variables. He presents two possible sets of transition formulas, instead of just one as in simple CA. This model is the correlation model counterpart of the association model

given in equation 5.18; that is,

$$P_{ijk} = P_{i..}P_{.j.}P_{..k}[1 + \sum_{m=1}^{M}(\lambda_m^{AB}x_{im}y_{jm} + \lambda_m^{AC}x_{im}z_{km} + \lambda_m^{BC}y_{jm}z_{km} + \lambda_m^{ABC}x_{im}y_{jm}z_{km})] \tag{5.21}$$

Since the decomposition has the same form as that in equation 5.18, the same problems and criticisms discussed with respect to the association model also apply to this model.

Equation 5.21 is an extension of a saturated additive model. It has the general additive form

$$\text{probability} = \text{prediction} + \sum(\text{interaction})$$

where "prediction" is the prediction from complete independence and the sum is across the various 2– and 3–way interactions effects, which are represented by (multidimensional) bilinear and trilinear terms. Equation 5.21 has the additional undesirable properties discussed in Chapter 3 associated with additive probability models. This model can yield negative fitted values and may be more complex than necessary. If a table is characterized by conditional independence or there is no 3–way interaction defined in terms of odds ratios, relatively simpler association models will fit such a table. For example, if a table is characterized by conditional independence such that $P_{ijk} = P_{ij.}P_{i.k}/P_{i..}$, equation 5.21 would in general require non-zero values for the $\lambda_m^{BC}$ and $\lambda_m^{ABC}$ parameters, while for the analogous association model, $\phi_m^{BC} = \phi_m^{ABC} = 0$.

Two other generalizations of CA were proposed by Kroonenberg (1989). These generalizations retain the interpretation of profiles in terms of chi-squared distances. In the 2–way case, there is only one possible way to define profiles: profiles are vectors corre-

sponding to rows (or columns). In the 3–way case, profiles can be defined as either vectors or matrices. "Fibers" are vectors of frequencies or probabilities for fixed levels of two of the three variables (e.g., $\mathbf{p}_{jk}$ is an $(I \times 1)$ vector containing elements $\{P_{ijk} | i = 1, \ldots, I; j$ and $k$ fixed$\}$). "Slices" are matrices of frequencies for fixed levels of one the three variables (e.g., $\mathbf{P}_{ABk}$ is an $(I \times J)$ matrix containing elements $\{P_{ijk} | i = 1, \ldots, I; j = 1, \ldots, J;$ and $k$ fixed$\}$). The geometric development of the generalization parallels that for simple CA, except that the $X^2$–terms are decomposed by a 3–factor decomposition, instead of the 2–way singular value decomposition. The two 3–way decompositions proposed by Kroonenberg are Tucker's 3–mode principal components model and CANDECOMP.

The generalization for slices leads to decomposing the $X^2$–terms for testing complete independence,

$$\frac{(P_{ijk} - P_{i..}P_{.j.}P_{..k})}{\sqrt{P_{i..}P_{.j.}P_{..k}}} = \sum_r \sum_s \sum_t \lambda_{rst} a_{ir} b_{js} c_{kt}$$

and the one for fibers results in a decomposition of the $X^2$–terms for testing joint independence,

$$\frac{(P_{ijk} - P_{ij.}P_{..k})}{\sqrt{P_{ij.}P_{..k}}} = \sum_r \sum_s \sum_t \lambda_{rst} a_{ir} b_{js} c_{kt}$$

In the 3–mode CA of slices and fibers, the sum of the squared $\lambda_{rst}$'s equal the part of Pearson's $X^2$ for testing complete or joint independence, respectively, accounted for by the 3–way decomposition,

$$\begin{aligned} X^2(\text{model}) &= \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} \lambda_{rst}^2 + \sum_{r=r^*}^{I} \sum_{s=s^*}^{J} \sum_{t=t^*}^{K} \lambda_{rst}^2 \\ &= X^2(\text{fit}) + X^2(\text{residual}) \end{aligned}$$

where $X^2$(model) is the $X^2$ for testing complete or joint independence, $r^* = R + 1$, $s^* = S + 1$, and $t^* = T + 1$. In the 3–mode CA of the slices, the values that are plotted are $x_{ir} = a_{ir}/\sqrt{P_{i..}}$, $y_{js} = b_{js}/\sqrt{P_{.j.}}$, and $z_{kt} = c_{kt}/\sqrt{P_{..k}}$.

If one is willing to give up the interpretation of chi-squared distances, then the departures from other loglinear models for 3–way tables can also be analyzed. Three-mode decompositions of standardized residuals from other loglinear models can be used complementary to loglinear analysis to study the interactions present in the data. While Kroonenberg (1989) does not provide any examples of his 3–mode correspondence analysis, he does an give an example of the complementary use of loglinear models and 3–mode components analyses of residuals (Kroonenberg, 1983). This approach is a 3–mode analogue of "generalized correspondence analysis" (van der Heijden & Meijerink, 1988) or the "combined approach" (van der Heijden & de Leeuw, 1985; van der Heijden & Worsley, 1988; van der Heijden, de Falugerolles & de Leeuw, 1989). In these methods, loglinear models are used to determine which interactions are present in the data. To study the interactions that are present, the interactions are removed from the loglinear model so that their effects are present in the residuals. These standardized residuals are appropriately arranged into a 2–way matrix, which is analyzed by correspondence analysis, or a 3-way matrix, which is analyzed by Tucker3.

Rather than studying interactions by decomposing residuals of loglinear models that do *not* fit, a more direct and straightforward approach is to explicitly include Tucker3 decompositions in the models. The implicit model assumed in Kroonenberg's (1989) generalizations of CA and in the combined approach of loglinear modeling and 3–mode

component analysis of residuals is

$$\text{frequency} = \text{prediction} + \text{interaction} + \text{error}$$

or

$$\text{frequency} = \text{fitted value} + \text{error}$$

where "prediction" is from a multiplicative (loglinear) model, the "interaction" is the 3–mode term, which represents the combined effects of the interactions accounted for by the 3–mode terms, and "error" is the discrepancies from the model. The "fitted value" equals the predicted frequency from the loglinear model plus the interaction. This model is not a "pure" additive model, but is a "mixed" model in that it contains an additive and a multiplicative part. The presence of an additive part can lead to negative fitted values. This approach is "mixed" in the sense that the "prediction" is estimated by maximum likelihood where a statistical model is assumed, while the "interaction" part is estimated by least squares where a particular structural model for the interactions are assumed. In the next chapter, association model generalizations that include 3–mode terms in loglinear models are introduced that do not have these undesirable characteristics and problems.

## 5.3 Exploratory 3–Mode Analysis of the Peer Play Data

While the combined approach has problems, it is useful for exploratory purposes. Analyzing standardized residuals by Tucker3 yields approximations to the association models generalizations for the peer play data, which are presented in the next chapter. They also provide good starting values for the algorithm used to compute the maximum likelihood estimation of the parameters of the association model generalizations. As an illustration of the method and to provide examples to compare with the association model generalizations presented in the next chapter, the analysis of standardized residuals of the $(GT, P)$ loglinear model is presented here.

Tucker3 decompositions of standardized residuals for various numbers of components are presented and discussed in Section 5.3.1. The results from two of the decompositions are presented in Sections 5.3.2 and 5.3.3. These results are compared to those from the reduction methods from the previous chapter, and in the next chapter, they are compared to the results of the new model generalizations of the $RC(M)$ association model.

### 5.3.1 *TUCKALS3* Decompositions

The joint independence model $(GT, P)$ was chosen, because the $GT$ margin is not "interesting", and should be fit perfectly. Kroonenberg and Brower's (1985a, 1985b) programs "*TUCKALS3* " and "*TUCKALS2* " were used to compute the least squares estimates of Tucker's 3–mode model and the extended version of the model of the standardized

residuals or the "chi-terms" (i.e., $e_{ijk} = (f_{ijk} - \hat{f}_{ijk})/\sqrt{\hat{f}_{ijk}}$ where $\hat{f}_{ijk} = f_{i+k}f_{+j+}/f_{+++}$, the fitted value from the $(GT, P)$ loglinear model). The results from *TUCKALS3* are reported in Table 5.2. Each row in the table corresponds to a different decomposition. The models are arranged in order of complexity such that the ones at the top are simpler (fewer components) than those at the bottom (more components). The horizontal lines separate the decompositions into blocks that have the same number of total components. The first three columns specify the number of components for each of the modes. The next two columns contain the $X^2$(Fit) and $X^2$(Residual), which sum to $X^2$ for the $(GT, P)$ loglinear model, "$X^2_{(GT,P)}$". The last two columns, labeled "Fit" and "Residual", contain the percentages of $X^2_{(GT,P)}$ accounted for and not accounted for by the decompositions. In the present context, "Fit" refers to the part of the standardized residuals from $(GT, P)$ loglinear model accounted for by Tucker3, while "Residual" refers to residuals of residuals (i.e., the part of the $e_{ijk}$'s not accounted for by the decomposition).

The first row of the table, where no components are estimated or "000", corresponds to the $(GT, P)$ loglinear model. The decomposition in the second row has only one component for each of the modes or "111." It is equivalent to CANDECOMP and to Tucker2 where the extended mode is either group, play or time and only 1 component is estimated for the other 2 modes. The 111 decomposition accounts for 24.51% of $X^2_{(GT,P)}$, which leaves 75.49% unexplained by the model. The improvement in fit from adding one component for each mode is the largest for the 111 decomposition.

There are no models where two of the modes have one component and the third has 2 or more components (i.e., $11T$, $1S1$, $R11$), because they are equivalent to the 111

**Table 5.2**: *TUCKALS3* decompositions of the chi-terms from the $(GT, P)$ loglinear model.

| Components | | | $X^2$ | | Percent of $X^2_{(GT,P)}$ | |
|---|---|---|---|---|---|---|
| Group | Play | Time | Fit | Residual | Fit | Residual |
| 0 | 0 | 0 | 0 | 422.96 | 0.00 | 100.00 |
| 1 | 1 | 1 | 103.66 | 319.30 | 24.51 | 75.49 |
| 1 | 2 | 2 | 112.35 | 310.61 | 26.56 | 73.44 |
| 2 | 1 | 2 | 130.22 | 292.74 | 30.08 | 69.21 |
| 2 | 2 | 1 | 128.14 | 294.82 | 28.93 | 71.07 |
| 2 | 2 | 2 | 158.79 | 264.17 | 37.54 | 62.46 |
| 2 | 2 | 3 | 165.80 | 257.16 | 39.20 | 60.80 |
| 2 | 3 | 2 | 166.35 | 256.61 | 39.33 | 60.67 |
| 3 | 2 | 2 | 160.61 | 262.35 | 37.97 | 62.03 |
| 3 | 3 | 2 | 190.23 | 232.73 | 44.98 | 55.02 |
| 3 | 2 | 3 | 184.88 | 238.08 | 43.71 | 56.29 |
| 2 | 3 | 3 | 180.07 | 242.89 | 42.57 | 57.43 |
| 3 | 3 | 3 | 214.29 | 208.67 | 50.66 | 49.33 |
| 4 | 4 | 4 | 268.39 | 154.57 | 63.45 | 36.55 |
| 5 | 5 | 5 | 330.78 | 92.18 | 78.21 | 21.79 |
| 6 | 6 | 5 | 367.76 | 55.20 | 86.95 | 13.05 |
| 7 | 7 | 5 | 400.55 | 22.41 | 94.70 | 5.05 |
| 7 | 8 | 5 | 410.56 | 12.40 | 97.07 | 2.93 |
| 7 | 9 | 5 | 418.66 | 4.30 | 98.98 | 1.02 |
| 7 | 10 | 5 | 422.96 | 0 | 100.00 | 0.00 |

decomposition. To demonstrate this equivalence, consider the $R11$ decomposition,

$$e_{ijk} = \sum_{r=1}^{R} g^*_{r11} a^*_{ir} b_j c_k$$

which can be re-written as

$$\begin{aligned} e_{ijk} &= a^{**}_{i11} b_j c_k \\ &= g a_i b_j c_k \end{aligned}$$

where $a^{**}_{i11} = \sum_{r=1}^{R} (g^*_{r11} a_{ir})$, $a_i = a^{**}_{i11} / \sqrt{\sum_i (a^{**}_{i11})^2}$, and $g = \sqrt{\sum_i (a^{**}_{i11})^2}$. The last line is the 111 decomposition where $a_i$, $b_j$, and $c_k$ are the scale values for groups, play qualities, and occasions, respectively, and $g$ is the core element that reflects the strength of the relationship among the three modes.

The next block of three rows, 122, 212 and 221, are special in that the core matrix is a 2–way matrix, which can be arbitrarily transformed to a diagonal matrix. For example, the 212 decomposition simplifies as follows:

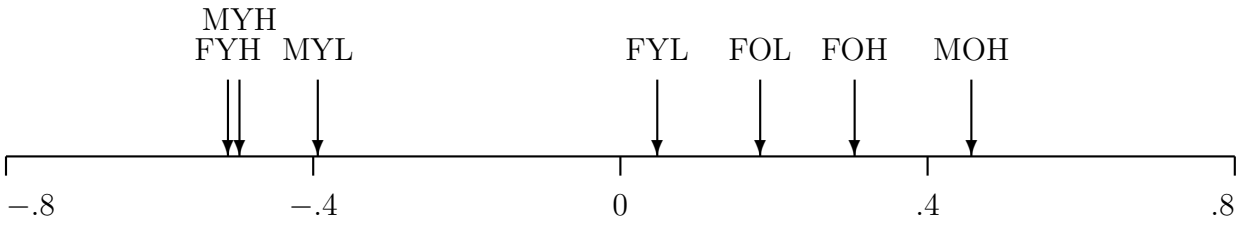
$$\begin{aligned} e_{ijk} &= \sum_{r=1}^{2} \sum_{t=1}^{2} g^*_{r1t} a^*_{ir} b_j c^*_{kt} \\ &= b_j (\sum_{r=1}^{2} \sum_{t=1}^{2} g^*_{r1t} a^*_{ir} c^*_{kt}) \\ &= b_j g^{**}_{ik} \end{aligned}$$

where $g^{**}_{ik} = \sum_{r=1}^{2} \sum_{t=1}^{2} (g^{*}_{r1t} a^{*}_{ir} c^{*}_{kt})$. Let $g_m$ equal the two non-zero singular values from the singular value decomposition of the $(7 \times 5)$ matrix of group-occasion scores (i.e., $\{g^{**}_{ik}\}$), and $\{a_{im}\}$ and $\{c_{km}\}$ are the corresponding left and right singular vectors. Substituting $\sum_{m=1}^{2} g_m a_{im} c_{km}$ for $g^{**}_{ik}$ in the above equation yields

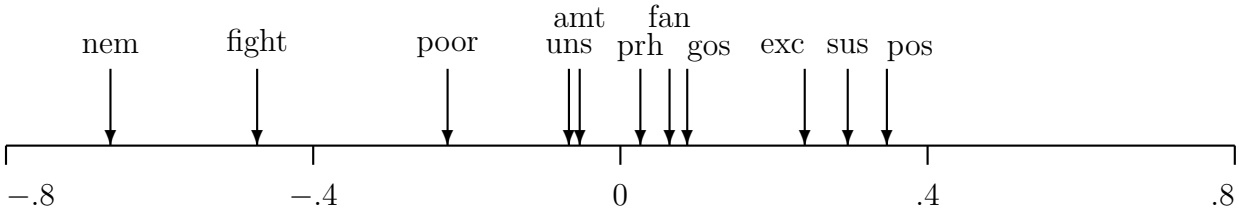
$$e_{ijk} = b_j \sum_{m=1}^{2} g_m a_{im} c_{km}$$

The 222 decomposition accounts for 37.54% of $X^2_{(GT,P)}$. The first model in the next block, 223, has the same fit statistics as the Tucker2 decomposition where time is the extended mode and two other modes are reduced to two components, which suggests that these two decompositions are equivalent. The decompositions with 3, 4, and 5 components per mode account for 50.66%, 63.45%, and 78.21%, respectively, of $X^2_{(GT,P)}$. As in standard principal components analyses, the increments in fit as more components are added become increasingly small. A complete decomposition of the standardized residuals is achieved by $R = I = 7$, $S = (J - 1) = 10$, and $T = K = 5$ components for groups, play qualities and occasions, respectively.

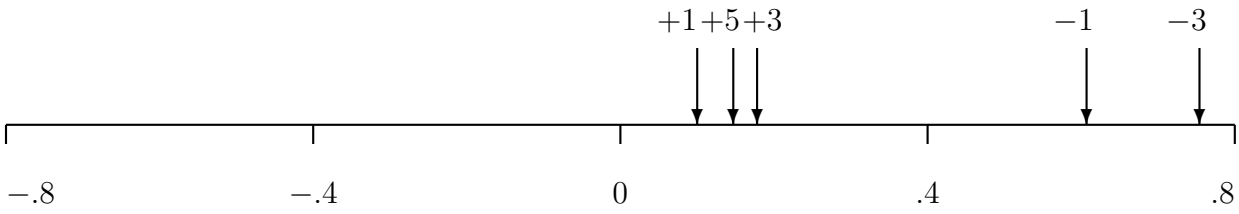
The 111 and 212 decompositions are examined in detail in Sections 5.3.2 and 5.3.3 both to illustrate the graphical methods and to compare them with models presented in the previous chapters and those presented in the next chapter. These decompositions may seem inadequate based on the percent of the $X^2_{(GT,P)}$ that they explain; however, the increments in fit for additional components are relatively small, and more complex de-




Groups: $a_i$




Play Qualities: $b_j$




Occasions: $c_k$

**Figure 5.5**: Scale values from the 111 *TUCKALS3* decomposition the standardized residuals from the $(GT, P)$ loglinear model.

compositions may be representing random, non-systematic variation in the data. Further justification for these simple decompositions are given in the next chapter.

### 5.3.2 The 111 Decomposition

Since Tucker's 3–mode model is being used here to analyze residuals rather than as a 3–mode generalization of correspondence analysis, the scale values $a_{ir}$, $b_{js}$ and $c_{kt}$ are plotted in all of the figures rather than $x_{ir}$, $y_{js}$, and $z_{kt}$, which are the scale values divided by the square root of their respective marginal probabilities (i.e., $P_{i..}$, $P_{.j.}$ and $P_{..k}$). Given the orthonormality identification constraint (i.e., $\mathbf{a}_r'\mathbf{a}_{r'} = \delta_{rr'}$, $\mathbf{b}_s'\mathbf{b}_{s'} = \delta_{ss'}$, and $\mathbf{c}_t'\mathbf{c}_{t'} = \delta_{tt'}$), the scale values are uniquely identified. Centering constraints on the scale values are not necessary, and the sums $\sum_i a_{is}$, $\sum_j b_{js}$, and $\sum_k c_{kt}$ will not necessarily equal 0.

The scale values for groups, play qualities and occasions from the 111 decomposition are given in Figure 5.5. The relative order of the groups and play qualities are similar to those from models previously examined. The older children have higher scale values than the younger ones, the group MOH has the largest positive scale value and the two groups FYH and MYH having the largest negative scale values. The play qualities are ordered from "bad" or immature to "good" or mature. The qualities with the largest negative scale value are negative emotion, followed by fight, and poor play. The qualities with the largest positive scale values are positive play, followed by sustain and excite.

The first occasion ($-3$ months) has the largest scale value, while the next largest value corresponds to the second occasion ($-1$ month). The three occasions following the

birth of the sibling are all relatively close together and have small, positive values. The scale values for occasions are all positive, which indicates that the relationship between groups and play qualities only changes in terms of the strength of the association and not in terms of the direction or nature of the relationship. Taken together, the scale values for the groups and play qualities show that the older children tend to exhibit more of the "good" qualities than the younger ones, who tend to exhibit more of the "bad" qualities. This relationship is strongest for the first session, slightly weaker for the second session, and is the weakest for the last three sessions.

While there are similarities with respect to previous models, there are also some differences. For example, the relative position of the group FYL is different from its position in Figure 5.1, which is the plot of the scale values from the $(GT, P)$ bivariate interaction, log-bilinear model with $M = 1$. With this model, FYL has a negative scale value, which is approximately equal to that of the groups MYL and FYH. In the Tucker3 decomposition, FYL has a small positive scale value (close to zero) and the group with the next closest scale value is FOL. This difference is due mainly to the fact that the *TUCKALS3* analysis represents the 3–way interaction, as well as the $GP$ and $PT$ 2–way interactions. In both of these models, one component or dimension is estimated for each mode, so that the models have the same number of scale values. The 3–mode decomposition accounts for more of the interactions in the data than the bivariate model, even though an additional parameter (i.e., $\phi^{TP}$) is estimated in the bivariate model. For the 3–mode decomposition, $X^2$(Residual) = 319.30, while for the bivariate model, $X^2 = 328.09$. The difference in fit can be explained by that fact there

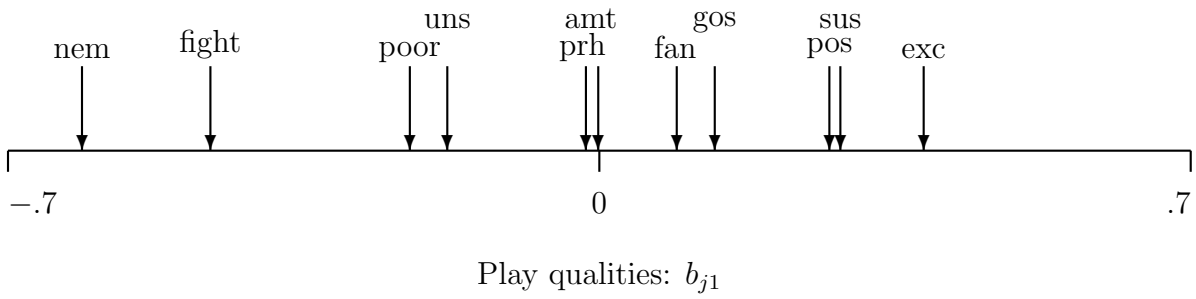



**Figure 5.6**: Play quality scale values from the 212 *TUCKALS3* decomposition of the standardized residuals from the $(GT, P)$ loglinear model.

appears to be a 3–way partial association (i.e., $G^2 = 278.07$ with $df = 240$), while there is very little evidence of a 2–way $PT$ partial (i.e., $G^2 = 42.13$ with $df = 40$) or marginal association (i.e., $G^2 = 41.22$ with $df = 40$).

### 5.3.3 The 212 Decomposition

The next decomposition presented here, which has two components for groups and time and one component for play qualities (i.e., 212), accounts for 30.08% of $X^2_{(GT,P)}$. As discussed above, the core matrix for this decomposition is diagonal, and the estimated values are $g_1 = 9.99$ and $g_2 = 5.52$. A number of useful and informative figures are presented and interpreted below for the 212 *TUCKALS3* analysis.

Figure 5.6 is a unidimensional plot of the scale values for the play qualities, and it represents the relationship among the qualities. As in previous analyses, the qualities are ordered such that immature or "bad" qualities have negative values while the more

mature or "good" qualities have positive values. The relative ordering of the qualities is very similar to the ordering from the joint $RC(1)$ association model fit to ${}_{TG}\mathbf{F}_{P}$ (see Figure 4.2) and the conditional $RC(1)$ association model with homogeneity constraints on the scale values for play qualities (see Figure 4.5). In Figure 5.6, the quality unsustain has a more extreme negative scale value and is closer to poor play than it is in Figures 4.2 and 4.5. Relative to these joint and conditional analyses, the quality sustain has a more extreme positive value and is closer to positive play, while the qualities fantasy and gossip have slightly smaller positive values.

The next figure, Figure 5.7, is a plot of the two components for occasions. In this figure, the scale values for occasions have been multiplied by the core elements so that the relative importance of the two dimensions is reflected in the plot. The relative order and spacing of the scale values on the first component ($t = 1$) are nearly identical to those from the 111 decomposition. The values on the first component decrease over time, and after the birth of the sibling, the values are close to zero. The second values on the second time component ($t = 2$) are close to zero on the first two occasions (before the birth) and are either negative or positive on the last three occasions. The third and fifth sessions have approximately the same positive scale values, while the fourth session has a negative value. Since $g_1$ is approximately twice as large as $g_2$, the largest values on the second component are approximately half the size of the largest values on the first component.

The next figure, Figure 5.8, is a joint plot of groups and occasions. The relationship among the groups, among occasions, and between groups and occasions are represented

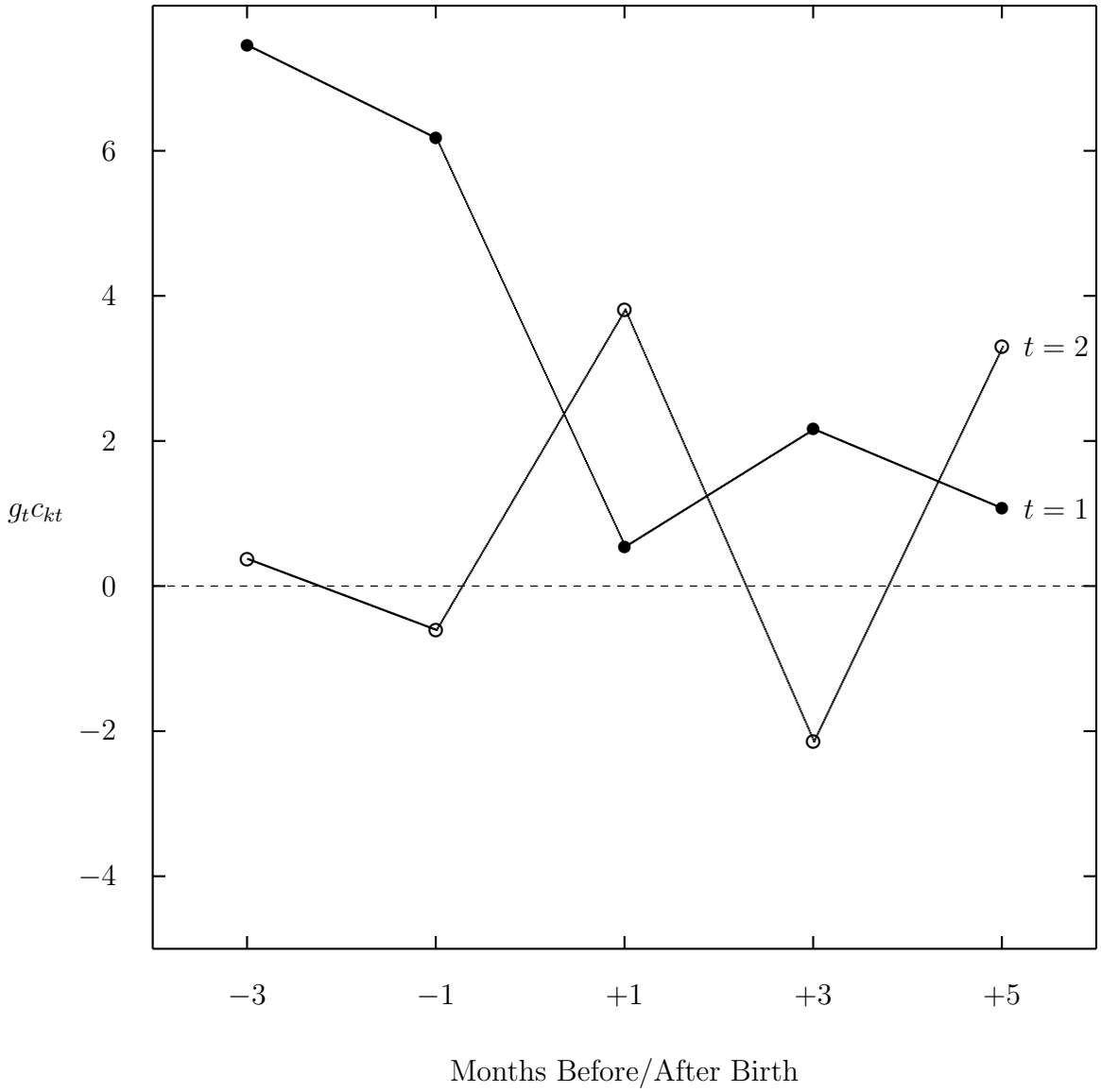



**Figure 5.7**: Scale values for occasions from the 212 *TUCKALS3* analysis of the standardized residuals from the $(GT, P)$ loglinear model

in this figure. The group and occasion scale values have been weighted by the square root of the appropriate core elements (i.e., $a_{im}\sqrt{g_{m1m}}$ and $c_{km}\sqrt{g_{m1m}}$ for $m = 1, 2$) to reflect the relative importance of the two components. Lines have been drawn from the origin to points representing occasions to help differentiate the scale values for occasions from those for groups.

The relative order and spacing of the groups and the occasions on the first dimension are nearly identical to that from the 111 *TUCKALS3* analysis. With respect to groups, the first component differentiates the groups according to age, and the second component mainly contrasts FYL (largest negative value) and MOH (largest positive value). With respect to occasions, the first dimension could be interpreted as a "before" component, and the second dimension could be interpreted as an "after" component. The first two occasions have large positive values on the first component. The second component mainly contrasts +3 months and the other two time points after the births of the siblings (+1 and +5 months).

The distances between points of the same mode are interpretable as reflecting the similarity between objects/levels within a mode. For example, since FYH and MYL are relatively close together, the children in these groups tend to exhibit similar qualities over time. Since MYH and MOH are far apart and are in opposite quadrants in the space, the children in these two groups are the most dissimilar over time in terms of the qualities observed in their play. With respect to occasions, the points corresponding to the first two sessions ($-3$ and $-1$ month) are quite close to each other, as well as the

third and fifth (+1 and +5 months), which implies that the occasions in these pairs are very similar.

The relationship between the modes is also represented in Figure 5.8. The absolute distances between points representing groups and occasions should not be interpreted, but the relative distances are meaningful. The directions of the vectors connecting the origin and points representing groups and occasions are also interpretable. The inner products between the vectors that correspond to groups and occasions equal scores for the combinations of groups and occasions on the play component (i.e., $\sum_m (a_{im}\sqrt{g_{m1m}})(c_{km}\sqrt{g_{m1m}}) = \sum_m g_{m1m}a_{im}c_{km} = g^{**}_{i1k}$). For example, relative to all of the other groups and all of the other occasions, the points representing the groups FYH, MYL and MYH are far from the points representing $-3$ months and $-1$ month. Since the vectors from the origin to FYH, MYL and MYH point in the opposite direction as the vectors representing $-3$ and $-1$, the groups FYH, MYL and MYH have negative scores at these time point. The children in these groups tend to exhibit more immature qualities and fewer mature ones relative to the other children and to other occasions.

Another way to examine the relationship between groups and occasions is to plot the group scores on occasions against time (i.e., $g^{**}_{ik}$ versus $k$). These scores are the inner products between the vectors in Figure 5.8 that correspond to the groups and the occasions. The group scores on occasions (i.e., inner products) are plotted in Figure 5.9. The symbols that were used in Figures 4.3 and 4.6 for the different groups are used in Figure 5.9. Recall that the variable age is coded by the shading of the symbol, such that the solid symbols (i.e., ●, ■, ♣) correspond to the older children, and the open

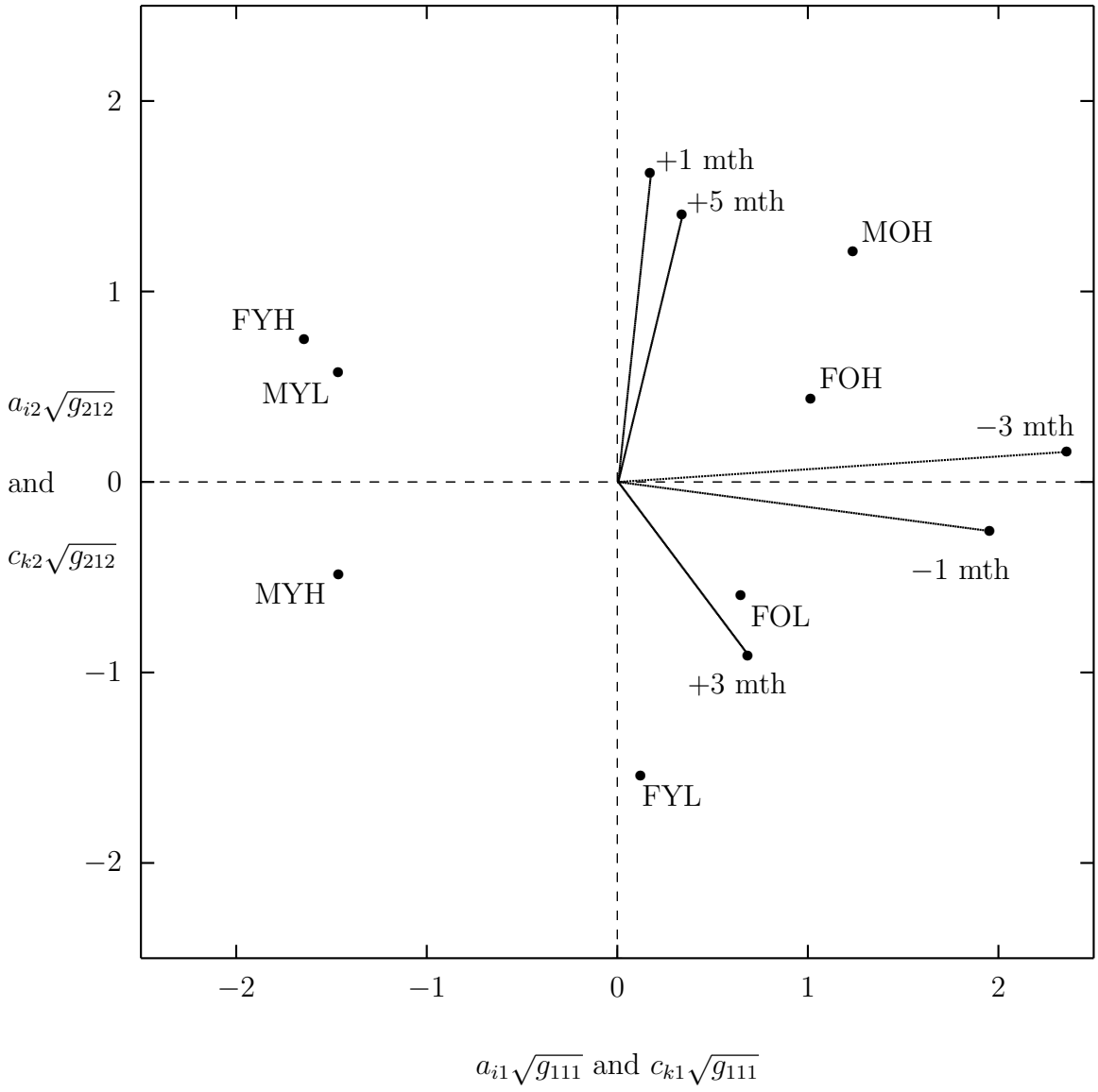



**Figure 5.8**: Joint plot of group and occasion scale values from the 212 *TUCKALS3* analysis of the standardized residuals from the $(GT, P)$ loglinear model

or unshaded symbols (i.e, △, ♡, □, ○ ) correspond to the younger children. Sibling acceptance is coded such that the symbols with straight lines (i.e., △, ■, □) correspond to the low sibling acceptance groups, and the symbols with the curved lines (i.e., ♡, ♣, ○ , ● ) correspond to the high acceptance groups.

Similarities and differences between groups over time can be studied using Figure 5.9. For example, the older children (shaded symbols) tend to have positive scores across occasions, while the younger children (open symbols) tend to have negative scores. The points corresponding to each group can be connected to see how a single group's scores change over time. The lines or "group profiles" can be used to compare groups. For example, the profile for FOH is relatively flat and smooth compared to the profiles for the other groups. The variability of the groups' scores can also be compared to see which group(s) show the most or least amount of change. The variance of FOH's scores, which equals $\sigma^2_{\text{FOH}} = 0.59$, is smaller than the variances for the other groups, which implies that the pattern of play qualities exhibited by the FOH children is relatively stable and changes the least. The scores for the FYL children have the largest variance (i.e., $\sigma^2_{\text{FYL}} = 3.75$), which implies the behavior of the FYL children changes more across time than the behavior of the other children.

Similarities and differences between time points across groups can also be studied. The first two occasions have the largest variances (i.e., $\sigma^2_{(-3\text{mth})} = 7.69$ and $\sigma^2_{(-1\text{mth})} = 5.29$), and the largest ranges (i.e., scores at $-3$ and $-1$ month range from $-3.76$ to 3.11, and $-3.40$ to 2.10, respectively), which means that the groups are more heterogeneous earlier in the study than they are the later time points.

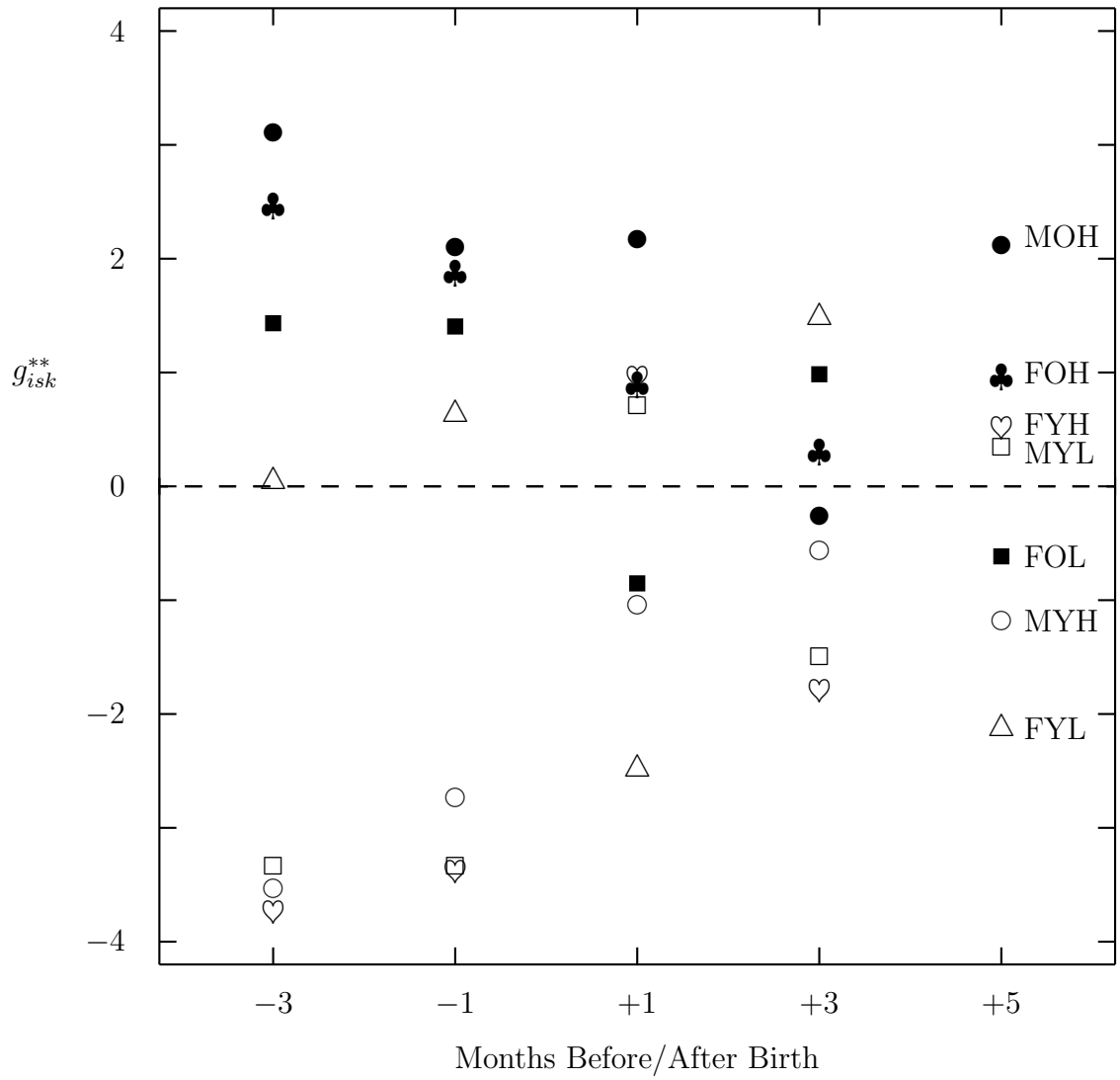



**Figure 5.9**: Group scores on occasions from the 212 *TUCKALS3* decomposition of the standardized residuals from the $(GT, P)$ loglinear model

The nature of the relationship among the three modes is given in Figures 5.6 and 5.9. The $g_{ik}^{**}$'s plotted in Figure 5.9 are scores for the groups on each occasion and the $b_j$'s plotted in Figure 5.6 are weights for the play qualities. Groups with positive scores tend to exhibit more of the "good" qualities, while groups with negative scores tend to exhibit more of the "bad" qualities. For example, since the group MOH has the largest positive score on all of the occasions except +3 months, the children in this group tend to exhibit more of the "good" qualities and fewer of the "bad" ones than the other children.

The plots of the play quality scale values and the group scores on the occasions from the 212 *TUCKALS3* decomposition (i.e., Figures 5.6 and 5.9), are very similar to the scale value plots from the joint $RC(1)$ association model fit to ${}_{TG}\mathbf{F}_P$ (i.e., Figures 4.2 and 4.3) and the conditional $RC(1)$ association model fit to $\{\mathbf{F}_{GP|k} | k = 1, \ldots, 5\}$ with homogeneity constraints on the scale values for the play qualities. The interpretation of the interaction among groups, play qualities, and occasions based on the 212 *TUCKALS3* decomposition is basically the same as the interpretation based on the joint and conditional $RC(1)$ association models. Theoretically, the joint $RC(M)$ association model and the *TUCKALS3* analysis decompose the $PT$, $GP$ and $GPT$ interactions, while the conditional $RC(M)$ association model decomposes the $GP$ and $GPT$ interactions.

Given the empirical similarity between the figures, all three analyses are capturing and representing the same basic underlying structure in the data. However, the *TUCKALS3* analysis is much simpler in that fewer scale values are estimated. In all three analyses, 11 scale values for play qualities are estimated, but in the conditional and joint models, a scale value is assigned to each group on each occasion for a total of 35 scale values,

while in the 3–mode decomposition, two scale values are assigned to each group and to each occasion for a total of 24 scale values.

## 5.4 Discussion

If there is no 3–way interaction, then the models presented in the section 5.1 are sufficient for analyzing the associations in 3–mode contingency tables. When there are higher–way associations, the methods and models presented in section 5.2 can be used.

The 3–mode models and analyses presented in section 5.2.3 are well suited for analyzing longitudinal data, such as the peer play data. The *TUCKALS3* analyses of the residuals from the $(GT, P)$ loglinear model were useful for detecting and representing the substantively interesting interactions in the peer play data. In terms of the number of parameters estimated, the 3–mode analyses of the peer play data were more parsimonious than the 2–way reduction methods presented in Chapter 4. However, the purpose of residual analyses is usually to detect whether a model adequately fits the data. If a model does not fit, then the results of the residual analyses are used to identify changes to the model to allow it to fit better.

A problem with using techniques such as correspondence analysis, generalized correspondence analysis, and *TUCKALS3* to study interactions in data is the lack of a basis for determining how many dimensions are needed. The number of dimensions chosen is generally based on the "percent of lack-of-fit" that is accounted for by the decomposition. For example, the 111 decomposition *only* accounts for 24.51% of $X^2_{(GT,P)}$, which leaves

75.49% unexplained. The mathematical model underlying the analysis consists of a multiplicative part (i.e., the $(GT, P)$ loglinear model) and an additive part (i.e., the 3–mode decomposition). However, since no statistical model was assumed, there is no basis for deciding whether the unexplained 75.49% is "large" or whether it could be attributed to chance.

There are additional problems with viewing the 3–mode residual analyses as "models" for the observed frequencies. This "model" could yield negative fitted values, because the interaction effects that are represented by the 3–way decomposition are added to the fitted values from a loglinear model. Furthermore, the "model" parameters are estimated in a stepwise manner such that the loglinear portion is estimated first by maximum likelihood, and the decomposition of the interactions is estimated second by least squares. The 3–mode model/residual analysis could easily be developed into a statistical model where all of the parameters are estimated simultaneously; however, a more logical choice for development is the extension of loglinear models where interactions terms are decomposed by Tucker's 3–mode model. Such models will yield non-negative fitted values, define interactions in terms of odds ratios, and represent them by relatively simple formulas involving scale values and measures of association. These extensions, which are also generalizations of the $RC(M)$ association model to 3–way tables, are presented in the next chapter.

# Chapter 6

# NEW MODEL GENERALIZATION

The models proposed and developed in this chapter are extensions of loglinear models for 3–way tables where either the 3–way interaction terms or some combination of 2–way and 3–way interaction terms of a loglinear model are approximated by Tucker's 3–mode principal components model. The models are also generalizations of the $RC(M)$ association model where instead of a 2–way singular value decomposition, Tucker's 3–mode model is used to represent the associations in the data.

The terms "variable" and "mode", and "category" and "level" are used interchangeably throughout this chapter. A 3–way table is a either a cross-classification of data into three variables or three modes where each mode consists of one or more (interactively coded) variables. In the models proposed here, the levels/categories of each of the modes/variables are assigned scale values. The models will be referred to as "log 3–mode" models. The term "log" indicates that the logarithm of fitted values is an additive function of the model parameters representing the "main" and "interaction" effects, thus

the fitted values are a multiplicative function of the "main" and "interaction" effects. The term "3–mode" indicates that the models contain a 3–way decomposition, which models the association among the 3 modes.

In Section 6.1, the models are presented in terms of a general framework. The specific steps involved in modeling data are described and illustrated using the peer play data. In Section 6.2, residual analyses that were performed to assess the appropriateness of various models are described, and the results are reported. In Section 6.3, the effect of missing data on the analyses and interpretation of the data is examined by fitting log 3–mode models to the data from the 29 child sub-sample and the full 30 child sample. The results are compared with those from the log 3–mode modeling of the 25 child sub-sample. In the final section, Section 6.4, the models and analyses presented in this chapter are briefly summarized.

## 6.1 Log 3–Mode Models

In Section 6.1.1, the model generalization is presented, along with a general framework for expressing the models developed here. Most of the models presented in the previous chapters can be expressed within this framework. In Section 6.1.2, the estimation of log 3-mode models is discussed. Since there are a number of possible generalizations, guidelines for model development and selection are presented in Section 6.1.3. The modeling process is illustrated in Section 6.1.4 where the peer play data are analyzed using log 3-mode models.

### 6.1.1 The Model

The models proposed here are extensions of saturated loglinear models for 3–way tables in which the 3–way interaction terms or some combination of the 2–way and 3–way interaction terms are approximated by Tucker's 3–mode principal components model. For example, the model in which the combined effects of all of the interaction terms are decomposed by Tucker's 3-mode model is

$$\ln(F_{ijk}) = u + u_i^A + u_j^B + u_k^C + \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T}\phi_{rst}\mu_{ir}\nu_{js}\eta_{kt} \tag{6.1}$$

where $u$ is a constant, $u_i^A$, $u_j^B$ and $u_k^C$ are main effect terms, $\mu_{ir}$, $\nu_{js}$ and $\eta_{kt}$ are the scale values for categories $i$, $j$ and $k$ of the variables $A$, $B$ and $C$ on components $r$, $s$ and $t$, respectively, and $\phi_{rst}$ is the intrinsic association parameter. All of the interactions in the data are represented by the sum over components of the product of the scale values and intrinsic association parameters. The parameter $\phi_{rst}$ measures the strength of the relationship (amount of association) among the 3 modes that is modeled by components $r$, $s$ and $t$.

Unlike saturated loglinear models, log 3-mode models will not necessarily provide an exact fit to a 3–way table. For example, the number of components in equation 6.1 that guarantees a complete decomposition of the interactions in the data is $R = I$, $S = J$ and $T = K$; however, when $R = I$, $S = J$ and $T = K$, equation 6.1 has more parameters than there are observations. When a model such as equation 6.1 is not saturated (i.e., $df > 0$), the sum of the $AB$, $AC$, $BC$ and $ABC$ interaction terms of the saturated

loglinear model are approximated by the decomposition (i.e., $(u_{ij}^{AB}+u_{ik}^{AC}+u_{jk}^{BC}+u_{ijk}^{ABC}) \approx \sum_r \sum_s \sum_t \phi_{rst}\mu_{ir}\nu_{js}\eta_{kt}$). In practice, log 3-mode models are useful when a relatively small number of components are needed to adequately model data.

Equation 6.1 is just one of the possible model generalizations. Other generalizations include 2–way interactions terms to ensure that particular 2–way margins are fit perfectly and/or to decompose a particular combination of interaction effects. A general expression for the family of model generalizations is

$$\ln(F_{ijk}) = u_{ijk}^{(0)} + u_{ijk}^{(1)} + u_{ijk}^{(1,2)} \tag{6.2}$$

where $u_{ijk}^{(0)} = (u + u_i^A + u_j^B + u_k^C)$, $u_{ijk}^{(1)}$ equals the first order interactions that are *not* decomposed, and $u_{ijk}^{(1,2)}$ equals the first and second order interactions that are decomposed. The margins of the table corresponding to the effects included in $u^{(0)}$ and $u^{(1)}$ are fit perfectly.

For a 3–way cross-classification of variables $A$, $B$ and $C$, the possible choices for $u^{(1)}$ and $u^{(1,2)}$ are given in table 6.1. Each row corresponds to a basic model. The entries in the first column are labels for the basic models and indicate the effects included in $u^{(0)}$ and $u^{(1)}$, and they also indicate the margins fit perfectly by the model. The effects included in $u^{(1)}$ and $u^{(1,2)}$ are listed in the second and third columns, respectively. The numbers listed in the last column correspond to equation numbers of centering constraints that are necessary to identify the scale values when $u^{(1,2)}$ is approximated by Tucker-3. The identification constraints are discussed in detail in Section 6.1.2.

**Table 6.1**: Basic models for 3–way tables

| Model | $u^{(1)}$ | $u^{(1,2)}$ | Constraints |
|---|---|---|---|
| $(A, B, C)$ | | $u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC}$ | |
| $(AB, C)$ | $u_{ij}^{AB}$ | $u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC}$ | 6.10 |
| $(AC, B)$ | $u_{ik}^{AC}$ | $u_{ij}^{AB} + u_{jk}^{BC} + u_{ijk}^{ABC}$ | 6.9 |
| $(A, BC)$ | $u_{jk}^{BC}$ | $u_{ij}^{AB} + u_{ik}^{AC} + u_{ijk}^{ABC}$ | 6.8 |
| $(AB, AC)$ | $u_{ij}^{AB} + u_{ik}^{AC}$ | $u_{jk}^{BC} + u_{ijk}^{ABC}$ | 6.9, 6.10 |
| $(AB, BC)$ | $u_{ij}^{AB} + u_{jk}^{BC}$ | $u_{ik}^{AC} + u_{ijk}^{ABC}$ | 6.8, 6.10 |
| $(AC, BC)$ | $u_{ik}^{AC} + u_{jk}^{BC}$ | $u_{ij}^{AB} + u_{ijk}^{ABC}$ | 6.8, 6.9 |
| $(AB, AC, BC)$ | $u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC}$ | $u_{ijk}^{ABC}$ | 6.8, 6.9, 6.10 |

The first model in the table, denoted by "$(A, B, C)$", corresponds to equation 6.1. This model fits all of the 1–way margins perfectly, and the sum of the first and second order interactions of the saturated loglinear model are approximated by Tucker's 3-mode decomposition model,

$$\begin{aligned} u_{ijk}^{(1,2)} &= u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC} \\ &\approx \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T} \phi_{rst}\mu_{ir}\nu_{js}\eta_{kt} \end{aligned}$$

Besides the full range of models developed in this chapter, most of the models discussed in the previous chapters can also be expressed using equation 6.2 and Table 6.1. For example, the joint $RC(M)$ association model fit to ${}_{AB}\mathbf{F}_C$ is equivalent to the $(AB, C)$ model where the effects included in $u^{(1,2)}$ are modeled by a 2–way, singular value decomposition (i.e., $u_{ijk}^{(1,2)} = u_{lk}^{(1,2)} = \sum_m \phi_m \mu_{lm}^{AB} \nu_{km}^{C}$ where $l = 1, 2, \ldots, IJ$). The conditional

$RC(M)$ association model fit to $\{\mathbf{F}_{AB|1}, \ldots, \mathbf{F}_{AB|K}\}$ is equivalent to the $(AC, BC)$ model where $u_{ijk}^{(1,2)} = \sum_m \phi_{m(k)} \mu_{im(k)} \nu_{jm(k)}$.

Since Tucker's 3-mode principal components model ("Tucker-3") is a promising model for longitudinal data and its potential usefulness was demonstrated in the last chapter, only Tucker-3 is explicitly considered here. The other two 3–way decomposition models, Tucker-2 and CANDECOMP, can be thought of as special cases of Tucker-3. For example, if one of the modes is not reduced in equation 6.1 (e.g, $T = K$), then the combined effects of the $AB$, $AC$, $BC$, and $ABC$ interactions are modeled by an extended 3-mode or Tucker-2 decomposition. Alternatively, if $R = S = T$ and $\phi_{rst} = 0$ for $r \neq s \neq t$, the decomposition has a diagonal form and is equivalent to CANDECOMP. In cases where diagonal decompositions are desired, the orthogonality constraint on the scale values should be relaxed, because it is not necessary to identify the parameters. Log 3-mode models with diagonal decompositions are likely to have the same problem of degenerate solutions that are associated with CANDECOMP (i.e., components with large negative correlations).

Log 3-mode models are basically multiplicative models; that is, they have the following general form

$$F_{ijk} = \exp(u) \exp(u_i^A) \exp(u_j^B) \ldots \exp(\sum_r \sum_s \sum_t \phi_{ir} \mu_{ir} \nu_{js} \eta_{tk})$$

or in terms of logarithms,

$$\ln(F_{ijk}) = u + u_i^A + u_j^B + \ldots + \sum_r \sum_s \sum_t \phi_{ir}\mu_{ir}\nu_{js}\eta_{tk}$$

where $F_{ijk}$ is a cell mean, the $u$-terms are analogous to the parameters of loglinear models, which represent main effects and possibly some 2–way interaction effects, and $\sum_r \sum_s \sum_t \phi_{ir}\mu_{ir}\nu_{js}\eta_{tk}$ represents the 3–way or a combination of 2– and 3–way interactions. Since log 3-mode models are multiplicative, many of the properties discussed in Chapter 3 that are characteristic of such models are also properties of log 3-mode models. For example, they will always yield non-negative fitted values and they are sub-table invariant.

Like $RC(M)$ association models, log 3–mode models provide representations of odds ratios and conditional odds ratios in terms of the intrinsic association parameters and the scale values. For example, the logarithm of the 3–way odds ratio in terms of the parameters of equation 6.1 is

$$\ln(\Theta_{ii',jj',kk'}) = \sum_{r=1}^{R} \sum_{s=1}^{S} \sum_{t=1}^{T} \phi_{rst}(\mu_{ir} - \mu_{i'r})(\nu_{js} - \nu_{j's})(\eta_{kt} - \eta_{k't}) \tag{6.3}$$

The parameter $\phi_{rst}$ is a measure of the strength of the 3–way association among variables $A$, $B$ and $C$ (effects included in $u^{(1,2)}$) on components $r$, $s$ and $t$, respectively. The

logarithm of the conditional odds ratio for category $k$ of variable $C$ is

$$\begin{aligned} \ln(\Theta_{ii',jj'(k)}) &= \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T}\phi_{rst}(\mu_{ir}-\mu_{i'r})(\nu_{js}-\nu_{j's})\eta_{kt} \\ &= \sum_{r=1}^{R}\sum_{s=1}^{S}\phi^{*}_{rsk}(\mu_{ir}-\mu_{i'r})(\nu_{js}-\nu_{j's}) \end{aligned} \tag{6.4}$$

where $\phi^{*}_{rsk} = \sum_{t=1}^{T}\phi_{rst}\eta_{kt}$. For level $k$ of variable $C$, $\phi^{*}_{rsk}$ is a measure of the strength of the relationship between variables $A$ and $B$ on components $r$ and $s$. For a unit change in scale values of variables $A$ and $B$ on components $r$ and $s$, respectively, $\phi^{*}_{rsk}$ is the contribution of components $r$ and $s$ to the logarithm of the conditional odds ratio.

In 3–mode principal components analyses of continuous data (Kroonenberg, 1983, 1984; Kroonenberg, Lammers and Stoop, 1985) and in 3-mode analyses of categorical data as presented in the previous chapter, the relationship among the levels of the observed modes and the relationship among the modes can be represented graphically in terms of the components. Various plots of scale values were discussed and illustrated in Chapter 5 where *TUCKALS3* analyses were performed on standardized residuals of loglinear models. Rather than representing structure in the data that is *not* captured by a loglinear model, plots of log 3-mode model scale values represent associations in the data that are captured by the model. Furthermore, for the log 3-mode models, plots analogous to those in Chapter 5 provide pictures of the associations in a 3-way tables where associations are defined in terms of odds ratios.

### 6.1.2 Estimation of Log 3-Mode Models

Identification constraints on the parameters of log 3-mode models are necessary to estimate the parameters. The constraints are arbitrary with respect to the fit of the model. Either zero-sum constraints (i.e., $\sum_i u_i^A = \sum_j u^B = \sum_k u_k^C = 0$, and for any first order interactions, $\sum_j u_{jk}^{BC} = \sum_k u_{jk}^{BC} = 0$) or fixing certain values to a constant (e.g., $u_1^A = u_1^B = u_1^C = u_{1k}^{BC} = u_{j1}^{BC} = 0$) can be used to identify the $u$-terms. With Tucker-3 decompositions of $u^{(1,2)}$, the scale values are constrained to be orthonormal,

$$\sum_{i=1}^{I} \mu_{ir}\mu_{ir'}h_i^A = \delta_{rr'} \tag{6.5}$$

$$\sum_{j=1}^{J} \nu_{js}\nu_{js'}h_j^B = \delta_{ss'} \tag{6.6}$$

$$\sum_{k=1}^{K} \eta_{kt}\eta_{kt'}h_k^C = \delta_{tt'} \tag{6.7}$$

where $h_i^A$,$h_j^B$ and $h_k^C$ are fixed and known weights, and $\delta_{rr'}$, $\delta_{ss'}$ and $\delta_{tt'}$ are Kronecker deltas (e.g., $\delta_{rr'} = 1$ for $r = r'$, and 0 for $r \neq r'$). Marginal probabilities, uniform weights and unit weights are among the possible choices for $h_i^A$, $h_j^B$ and $h_k^C$.

In the case of the $RC(M)$ association model for 2–way tables, centering constraints are necessary on the scale values of both the row and column variables; however, with log 3-mode models, the necessity of centering constraints to identify the scale values for a variable depends on which effects are included in $u^{(1)}$ and $u^{(1,2)}$. For the models listed

in Table 6.1, the equation numbers of the required centering constraints refer to

$$\sum_{i=1}^{I} \mu_{ir} h_i^A = 0 \tag{6.8}$$

$$\sum_{j=1}^{J} \nu_{js} h_j^B = 0 \tag{6.9}$$

$$\sum_{k=1}^{K} \eta_{kt} h_k^C = 0 \tag{6.10}$$

The scale values of variables *not* involved in the first order interaction terms that are included in the model need to be centered.

For example, in the $(A, BC)$ log 3-mode model, the terms $u_{jk}^{BC}$ are estimated and only the scale values for variable $A$ (i.e., the $\mu_{ir}$'s) need to be centered. Constraint 6.8 is necessary, because any set of scale values $\mu_{ir}^* = (\mu_{ir} + x)$ where $x$ is an arbitrary constant will lead to the same fitted values as $\mu_{ir}$. As another example, consider the $(AC, BC)$ model, which contains the first order interaction parameters $u_{ik}^{AC}$ and $u_{jk}^{BC}$. Since variable $A$ is not involved in $u_{jk}^{BC}$, and variable $B$ is not involved in $u_{ik}^{AC}$, the scale values corresponding to variables $A$ and $B$ need to be centered (i.e., equations 6.8 and 6.9).

For independent Poisson random variables, the maximum likelihood equations for the parameter estimates of the models listed in Table 6.1 are given in Appendix B. An iterative algorithm based on the unidimensional Newton-Raphson method is also given

in Appendix B. The algorithm for the first four models of Table 6.1 was implemented in a *FORTRAN* program, which is given in Appendix C.

Maximum likelihood algorithms require starting values for all of the model parameters. Good starting values for the log 3-mode parameters can be obtained either from the saturated loglinear model or the appropriate loglinear model. For example, good starting values for the $u$-terms of the $(AB, C)$ log 3-mode model are the $u$-terms (i.e., $u$, $u_i^A$, $u_j^B$, $u_k^C$ and $u_{ij}^{AB}$) from the saturated loglinear model. Alternatively, the estimated parameters of the $(AB, C)$ loglinear model can also be used as starting values for the $u$-terms of the $(AB, C)$ log 3-mode model. Good starting values for the intrinsic association parameters and the scale values are the least squares estimates of the core matrix and component scores from the *TUCKALS3* approximation of $u^{(1,2)}$, where $u^{(1,2)}$ is the sum of the appropriate $u$-terms from the saturated loglinear model. This method works well unless there are some cells with very small frequencies, which leads to estimates of particular $u$-terms that are very large or small relative to the other estimated parameters (i.e., for $f_{ijk} = 0$, $|u_{ijk}^{(1,2)}| \to \infty$). When some of cells have large $u^{(1,2)}$'s (i.e., small frequencies), a *TUCKALS3* decomposition of the $u^{(1,2)}$'s will be dominated by the cells with large values of $u^{(1,2)}$. When this occurs, reasonable starting values for the scale values and the intrinsic association parameters can be obtained from the appropriate *TUCKALS3* analysis of the standardized residuals from a loglinear model. For example, the *TUCKALS3* decomposition of the standardized residuals from the $(AB, C)$ loglinear model for particular values of $R$, $S$ and $T$ can be used as starting values for the scale

values and intrinsic association parameters of the $(AB, C)$ log 3-mode model with $R$, $S$ and $T$ components for modes $A$, $B$ and $C$, respectively.

### 6.1.3 Model Selection

The process of modeling data with log 3–mode models involves a number of steps. The first step is the selection of an appropriate basic model. The choice of a basic model should be based on substantive theory and how the data were collected. The next step is to fit the model to the data for different numbers of components and evaluate the fit of the models. The criteria for evaluating models include goodness-of-fit statistics, parsimony, interpretability of results and residual analyses. Residual analyses provide a way to check model assumptions and to ensure that the model is not missing any systematic structure that is present in the data.

For the first step, a preliminary analysis of the data using loglinear models is useful for determining which associations are present or "important" with respect to modeling the data. Substantive theory will dictate which associations are "interesting". The appropriate basic model is the one in which $u^{(1)}$ includes the important but uninteresting interactions and $u^{(1,2)}$ includes the substantively interesting interactions. When selecting a basic model, the method of data collection and the structure of the study should also be taken into consideration. For example, if the BC margin is fixed by design, then $u^{(1)}$ should include $u_{jk}^{BC}$.

Including interactions in $u^{(1,2)}$ that are not important will not overly influence the fit of the model or the estimated scale values and association parameters. For example, if

the 2–way partial association between variables $B$ and $C$ is negligible (i.e., $u_{jk}^{BC} \approx 0$ for all $j$ and $k$), then including $u_{jk}^{BC}$ in $u^{(1,2)}$ will have a minimal effect on the estimated scale values and fitted values of the log 3-mode model. The scale values estimated from the model where $u_{jk}^{BC}$ is included in $u^{(1,2)}$ will be approximately the same as those from the model where $u_{jk}^{BC}$ is included in $u^{(1)}$. When a 2–way association is negligible, the major advantage of including it in $u^{(1,2)}$ is that the model will have more degrees of freedom than the model in which it is included in $u^{(1)}$.

Once a basic model has been chosen, the next step is to fit the model for different numbers of components. A useful preliminary step is to first do *TUCKALS3* decompositions of either the estimates of $u^{(1,2)}$ from the saturated loglinear model or the standardized residuals from the loglinear model that corresponds to the basic log 3-mode model. Such analyses often approximate the log 3-mode model, which makes *TUCKALS3* analyses useful exploratory tools for determining a set of potentially "good" log 3-mode models. These analyses give an indication of the structure and complexity of the associations in the data and help determine the number of components needed. For example, if the $(AB, C)$ model is selected from Table 6.1, then the standardized residuals from the $(AB, C)$ loglinear model should be decomposed for various numbers of components. The residual $X^2$ from a *TUCKALS3* analysis often approximates Pearson's $X^2$ statistic for the corresponding log 3-mode model. Furthermore, the scale values and core elements from *TUCKALS3* are good starting values for the scale values and intrinsic association parameters for the algorithm used to fit the log 3-mode models to data.

After fitting log 3-mode models for various numbers of components, the goodness-of-fit of the models needs to evaluated. The natural and standard tool for this purpose is $G^2$, the likelihood ratio statistic. When the underlying sampling assumptions are valid, $G^2$ can be compared to a chi-squared distribution with the appropriate degrees of freedom to test whether the lack-of-fit of a model is statistically "large". Differences in fit between nested models can be tested by comparing conditional likelihood ratio statistics to the appropriate chi-squared distribution; that is, compare $G^2_{(1|2)} = (G^2_{(1)} - G^2_{(2)})$, where model (1) is the more restrictive model and (2) is the more general model, to the chi-squared distribution with $df_{(1|2)} = (df_{(1)} - df_{(2)})$.

When the sampling assumptions are violated, as is the case with longitudinal and repeated measures data, $G^2$ can still be used as an overall index of fit. The ratio $G^2/df$ is useful, because it takes into account degrees of freedom. The statistic $G^2/df$ can be used to identify a sub-set of models that provide acceptable fits. Regardless of whether the sampling assumptions are violated, $G^2/df$ can be used to compare non-nested models. The model with the smaller $G^2/df$ is the more parsimonious model.

The substantive interpretation of the data based on the models is another basis for discriminating among models. The interpretation of the data should make sense in light of knowledge of the subject matter, substantive theory, and previous analyses of the data. For example, the overall interpretation of the associations in the peer play data based on a log 3-mode model should be similar to the results found from the loglinear models, $RC(M)$ association models fit to 2–way formulations of the data, association model generalizations, and *TUCKALS3* analyses of the residuals of loglinear models.

Since the different analyses are based on slightly different models, some differences in substantive interpretations of the data are expected; however, in general, the results should be similar.

With log 3-mode models, there are two aspects of parsimony that need to be considered: degrees of freedom ($df$) and number of components ($R, SandT$). The degrees of freedom depend on how many parameters are estimated relative to the number of cells in the table. The parameters consist of the $u$-terms, the scale values and the intrinsic association parameters. The number of components indicates the complexity of the representation of the associations in the data. Models with fewer components yield simpler graphical representations than models with more components. With nested models, the more restrictive model always has fewer components than the more general model; therefore, the most parsimonious model in terms of degrees of freedom is also the most parsimonious in terms of number of components. This is not necessarily the case for non-nested models. For example, suppose that the set of possible models has been narrowed down to two models, the model where $R = T = 3$ and $S = 1$ (or "313") and the one where $R = S = T = 2$ (or "222"). The 313 model has more degrees of freedom than 222, but 222 has fewer components than 313.

An evaluation of a model is not complete without examining the residuals. The residuals should be studied to look for evidence that the model does not fit in some systematic way. For example, if there are outliers or the largest residuals are associated with particular categories or combinations of categories, then the model is not accounting for some aspect of the data. Theoretically, if a model is correct, then the standardized residuals

are asymptotically distributed as standard normal random variables. Empirically, the standardized residuals should be approximately normally distributed with mean 0 and variance slightly less than 1 (Agresti, 1991).

The residuals should also be studied to look for evidence that sampling assumptions have been violated. If observations are independent, random samples from a Poisson distribution, then certain structures and relationships are expected. For example, if the observations are from a Poisson distribution, then the variance of the fitted values should equal the mean. With the peer play data, the independence assumption is suspect and observations may be positively correlated. If this is the case, then the variance should exceed the mean ("overdispersion"). If the independent assumption is valid, then the correlations of the residuals between occasions and between play qualities should equal 0. These and other residual analyses are explained in detail and illustrated in Section 6.2 where the residual analyses of log 3-mode models fit to the peer play data are presented.

### 6.1.4 Log 3-Mode Modeling of the Peer Play Data

Since the $GT$ association is needed to fit the peer play data, is affected by the design of the study, but is not substantively interesting, appropriate basic models are those that include $u_{ik}^{GT}$ terms. Since the $GP$, $PT$ and $GPT$ associations are all substantively interesting, their effects are all included in $u^{(1,2)}$. While the $PT$ association is interesting, previous analyses indicate that it is relatively weak; thus including it in $u^{(1,2)}$ is expected to have a negligible effect on the estimated scale values of the log 3-mode model. The

basic model chosen for the peer play data is "$(GT, P)$":

$$\ln(P_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T}\phi_{rst}\mu_{ir}\nu_{js}\eta_{kt} \tag{6.11}$$

for various values of $R$, $S$ and $T$. As in the previous analyses of the peer play data, unit weights are used in the identification constraints (i.e., $h_i^G = h_j^P = h_k^T = 1$). Since $u_{ik}^{GT}$ is in the model, the $GT$ margin is fit perfectly and only the scale values for the play qualities are centered (i.e., $\sum_j \nu_{js} = 0$).

Both $(G, P, T)$ and $(GT, P)$ log 3-mode models were fit to the peer play data, but only the results of the $(GT, P)$ models are presented and discussed here. The fit statistics for the $(GT, P)$ models are given in Table 6.2. Particular models will be referred to by the number of components estimated for groups, play qualities and occasions, respectively. These are listed in the first three columns of the table. For example, the third model in the table is denoted by "$(GT, P)$–212" or "212", for short. The models are arranged from most restrictive (top) to least restrictive (bottom), and they are separated into blocks consisting of models with the same total number of components. For comparison to the 3-mode analyses of the standardized residuals from the $(GT, P)$ loglinear model, the last column contains the residual $X^2$ from Tuckals3.

The first model in the table, $(GT, P)$–000, is equivalent to the $(GT, P)$ loglinear model. The largest single improvement in fit between two nested models occurs between 000 and 111. The $X^2$ for the $(GT, P)$–111 log 3-mode model is approximately equal to the residual $X^2$ from the corresponding *TUCKALS3* decomposition. On the basis of fit

**Table 6.2**: Fit statistics of $(GT, P)$ log 3–mode models $(N = 25)$.

| Components | | | Log 3–mode model statistics | | | | *TUCKALS3* |
|---|---|---|---|---|---|---|---|
| Group | Play | Time | *df* | $G^2$ | $G^2/df$ | $X^2$ | $X^2$(Residual) |
| 0 | 0 | 0 | 340 | 446.24 | 1.31 | 422.96 | 422.96 |
| 1 | 1 | 1 | 320 | 339.26 | 1.06 | 320.24 | 319.30 |
| 2 | 1 | 2 | 311 | 307.08 | .99 | 282.26 | 292.74 |
| 1 | 2 | 2 | 308 | 326.02 | 1.06 | 305.51 | 310.61 |
| 2 | 2 | 1 | 306 | 309.87 | 1.01 | 293.41 | 294.82 |
| 3 | 1 | 3 | 304 | 288.96 | .95 | 264.21 | 273.17 |
| 1 | 3 | 3 | 298 | 318.36 | 1.07 | 296.50 | 306.11 |
| 3 | 3 | 1 | 294 | 290.70 | .99 | 273.48 | 275.34 |
| 2 | 2 | 2 | 297 | 293.57 | .99 | 268.83 | 264.17 |
| 2 | 2 | 3 | 291 | 287.01 | .99 | 265.03 | 257.16 |
| 3 | 2 | 2 | 289 | 264.63 | .92 | 244.46 | 262.35 |
| 2 | 3 | 2 | 286 | 280.17 | .98 | 257.62 | 256.61 |
| 3 | 2 | 3 | 281 | 251.88 | .90 | 235.79 | 238.08 |
| 2 | 3 | 3 | 278 | 264.84 | .95 | 239.24 | 242.89 |
| 3 | 3 | 2 | 276 | 243.88 | .88 | 225.33 | 232.73 |
| 3 | 3 | 3 | 265 | 219.93 | .83 | 201.81 | 208.67 |

statistics, the 111 model appears to fits the data rather well, especially considering that the model contains relatively few parameters.

The next block of models (i.e., 212, 122, 221) consists of models that are slightly more complex than 111, but they are still relatively simple. Model 212, which has 2 components for groups and occasions and one component for play qualities, is the "best" model among those in the third block. It has the most degrees of freedom, the smallest $G^2$, the smallest $X^2$, and the smallest $G^2/df$. This model also yields the largest improvement in fit relative to 111 (i.e., $(G^2_{(111)} - G^2_{(212)})/(df_{(111)} - df_{(212)}) = 3.58$), as well as, the smallest decrement in fit relative to more complex models.

**Table 6.3**: Analysis of association by $(GT, P)$ log 3–mode models.

| Source | Models | *df* | $G^2$ | $G^2/df$ | Percent | Cumulative |
|---|---|---|---|---|---|---|
| $\{\mu_{i1}\}, \{\nu_{j1}\}, \{\eta_{k1}\}$ | $000 - 111$ | 20 | 106.98 | 5.35 | 24.0% | 24.0% |
| $\{\mu_{i2}\}, \{\eta_{k2}\}$ | $111 - 212$ | 9 | 32.19 | 3.58 | 7.2% | 31.1% |
| $\{\nu_{j2}\}$ | $212 - 222$ | 14 | 13.51 | .97 | 3.0% | 34.2% |
| unexplained | 222 | 297 | 293.57 | .99 | 65.8% | 100.0% |
| total | 000 | 340 | 446.24 | 1.31 | | |

Model comparisons between nested models in the set consisting of 000, 111, 212 and 222 are summarized in an analysis of association table, Table 6.3. This table can be used to examine the relative contributions of various components of the 222 model to explaining the interactions in the data. The source or component effects are listed in the first column. The models used to assess the contribution of each of the sources are listed in the second column, where the first model is nested within the second. The $G^2$'s for the first 3 sources are differences between likelihood ratio statistics. The last two columns contain the percent and the cumulative percent of the "total association" that is accounted for by the components. The "total association" equals the lack-of-fit of the $(GT, P)$ loglinear model, which is measured by the $G^2$ for the $(GT, P)$ loglinear, and consists of the $GP$, $PT$ and $GPT$ partial associations.

The data that are "unexplained" by 222 (i.e., the lack-of-fit of the 222 model) is "small" relative to the degrees of freedom for this model (i.e., $G^2/df = .99$). The second component for the play qualities, $\{\nu_{j2}\}$, accounts for only 3% of the association, which is not a statistically large amount (i.e., $G^2/df = .97$ is "small"). The second components for groups and occasions, $\{\mu_{i2}\}$ and $\{\eta_{k2}\}$, account for 7.2%, and are important since

$G^2/df = 3.58$ is relatively large. The three components $\{\mu_{i1}\}$, $\{\nu_{j1}\}$ and $\{\eta_{k1}\}$ account for 24.0% of the association, and are also important since $G^2/df = 5.35$ is quite large. Based on this analysis, 212 is the "best" model.

In the *TUCKALS3* analyses of the residuals from the $(GT, P)$ loglinear model, neither the 111 nor the 212 decomposition appear to be adequate based on the percent of $X^2_{(GT,P)}$ accounted for by the decomposition. The 111 and 212 decompositions account for only 24.51% and 30.08%, respectively, of $X^2_{(GT,P)}$. However, with the log 3-mode modeling approach where degrees of freedom are taken into consideration, the $(GT, P)$–111 and 212 fit the data fairly well (i.e., $G^2_{(111)}/df_{(111)} = 1.06$, and $G^2_{(212)}/df_{(212)} = .99$ ). Even though the $(GT, P)$–212 log 3-mode model accounts for approximately a third of the $GP$, $PT$ and $GPT$ partial associations, the remaining or unexplained association in the data is "small" in the sense it can be attributed to chance or random variation.

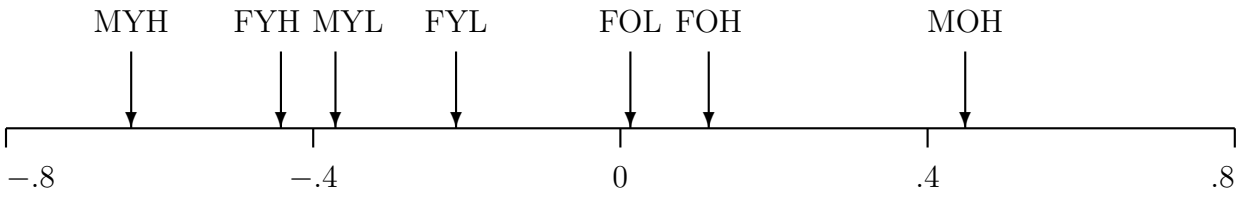
While the simple models, 111 and 212, are adequate, the more complex models 313 and 222, which include 111 and 212 as special cases, were also considered. Among the models in the fourth block, which consists of 313, 133 and 331, the "best" one is 313. It has the smallest $G^2$, $G^2/df$ and $X^2$, as well as the most degrees of freedom. The fifth block consists of just one model, 222. On the basis of fit statistics, 313 would be chosen over 222, because 313 has more degrees of freedom and fits better than 222 (i.e., $df_{(313} = 304 > df_{(222)}$, and $G^2_{(313)}/df_{(313)} < G^2_{(222)}/df_{(222)}$). On the basis of number of components, 222 is simpler than 313 (i.e., 6 components versus 7). For 313, either a three dimensional plot or a set of two dimensional of projections is required to graphically represent the

groups, occasions, and the relationship between them. For 222, two dimensional plots are sufficient.

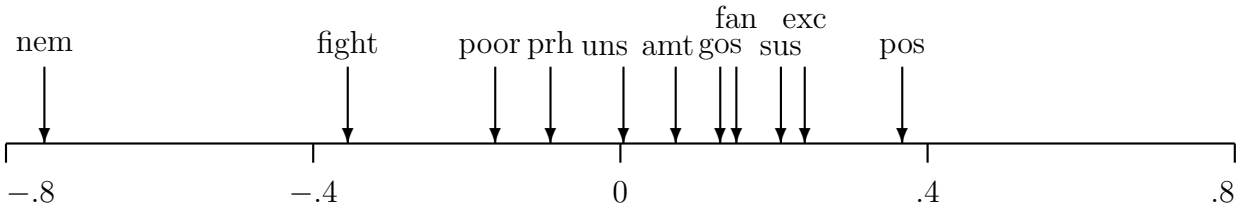
The interpretability of the data in terms of the estimated parameters of the models and the results of residual analyses were also considered. The scale values from the models 111, 212, 122, 221, 313, and 222 were plotted and examined, and the residuals from these models were analyzed. The 111 and 212 were selected as the two "best" models. On the basis of parsimony and interpretability, 111 is better than 212; however, 212 is only slightly more complex than 111, and there is some evidence that the second components for groups and occasions contribute a significant amount to explaining the data. The estimated scale values from both 111 and 212 are presented and discussed below, while results of residual analyses are presented in Section 6.2.

The estimated intrinsic association parameter of the $(GT, P)$–111 log 3-mode model is $\hat{\phi}_{111} = 4.307$. The estimated scale values of groups, play qualities and occasions are plotted in Figure 6.1. The relative order of the scale values is similar to the order from previous analyses, in particular the conditional $RC(1)$ association model with homogeneity constraints on the scale values for groups and play qualities (see Figure 4.4), and the 111 *TUCKALS3* decomposition of the standardized residuals from the $(GT, P)$ loglinear model (see Figure 5.5).

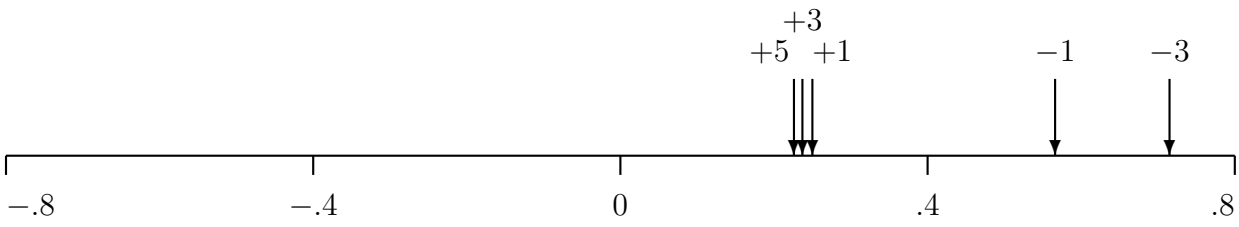
The older groups have positive scale values, while the younger ones have negative values. The group MOH has the largest positive value and MYH has the largest negative value. The scale values for MYH, FYH and MYL are spaced relatively further apart than the corresponding scale values from the *TUCKALS3* decomposition, and the scale value




Groups: $\hat{\mu}_{i1}$




Play Qualities: $\hat{\nu}_{j1}$




Occasions: $\hat{\eta}_{k1}$

**Figure 6.1**: Estimated scale values of the $(GT, P)$–111 log 3-mode model fit to the 25 child sub-sample.

for FYL is much closer to MYL than it is in the *TUCKALS3* decomposition. Overall, the relative order and spacing of the group scale values of the $(GT, P)$–111 log 3-mode model are more similar to those from the conditional $RC(1)$ association model than they are to those from the *TUCKALS3* decomposition.

The "good" play qualities have positive scale values and the "bad" qualities have negative scale values. Negative emotion has the largest negative value, followed by fight and poor play. Positive play has the largest positive value, and excite has the next largest value. Except for the three smallest scale values, which correspond to sustain, prohibit and amity, the relative order of the play quality scale values is the same as that from the 111 *TUCKALS3* decomposition (Figure 5.5), and the conditional $RC(1)$ association model with homogeneity constraints on groups and play qualities (Figure 4.4). With respect to spacing, the scale values for play qualities from the log 3-mode model are more spread apart relative to those from the conditional model and Tuckals3.

The order of the scale values for occasions is the same as the temporal order of the occasions; however, the values for the last three sessions are nearly equal. The order of the scale values for +1, +3 and +5 should not be over-interpreted, especially without information on the standard errors of these parameter estimates. Standard errors were not calculated because of dependencies among observations. Similar to the 111 *TUCKALS3* decomposition, the scale values for the occasions are all positive. This leads to the conclusion that the strength of the association between groups and play qualities decreases over time, especially after the birth of the siblings. The older children tend to show more positive or mature qualities, and the younger children tend to show

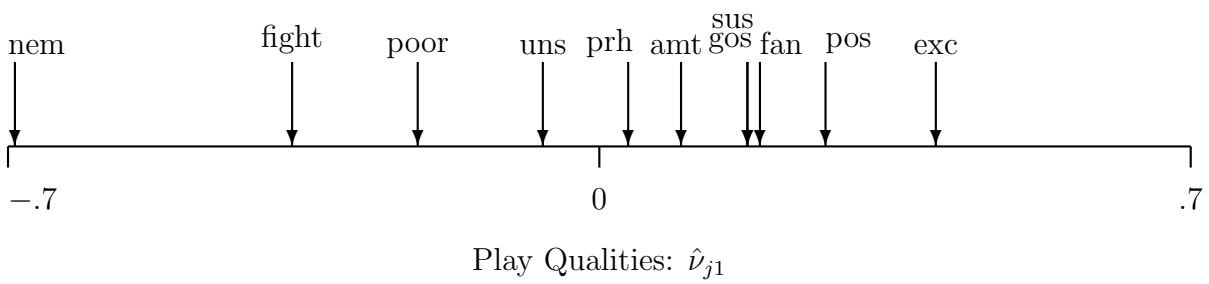



**Figure 6.2**: Estimated play quality scale values of the $(GT, P)$–212 log 3-mode model.

more negative or immature qualities. The children are more similar to each other in terms of the qualities observed in their play after the birth of their siblings than before the birth. This is interpretation is basically the same as that of the conditional $RC(1)$ association model with homogeneity constraints on the scale values for groups and play qualities and the 111 *TUCKALS3* of the standardized residuals.

Estimating a second component for groups and occasions leads to a slightly more complex interpretation of the associations in the peer play data than the 111 model. The interpretation of the associations in the data based on the 212 model will now be presented. The $(GT, P)$-212 model is similar to the joint $RC(1)$ association model fit to ${}_{TG}\mathbf{F}_{P}$, the conditional $RC(1)$ association model fit with homogeneity constraints on the scale values for play qualities, and the 212 *TUCKALS3* analysis of the standardized residuals from the $(GT, P)$ loglinear model. Since there is only one component for play qualities in the 212 model, the intrinsic association parameters form a diagonal matrix. The estimated values are $\hat{\phi}_{111} = 4.345$, which is approximately equal to the $\hat{\phi}$ from the 111 model, and $\hat{\phi}_{212} = 2.497$. Since $\hat{\phi}_{111}$ is larger than $\hat{\phi}_{212}$, the first components, $\{\mu_{i1}\}$

and $\{\eta_{k1}\}$ are more "important" than the second components, $\{\mu_{i2}\}$ and $\{\eta_{k2}\}$, with respect to accounting for the $GT$, $PT$ and $GPT$ interactions .

The scale values for the play qualities are plotted in Figure 6.2. Except for two minor differences, the order of the scale values in Figure 6.2 is the same as the order of the play quality scale values from the joint $RC(1)$ association model (Figure 4.2), the conditional $RC(1)$ association model with homogeneity constraints on the scale values for the play qualities (Figure 4.5), and the 212 *TUCKALS3* decomposition of the standardized residuals from the $(GT, P)$ loglinear model (Figure 5.6). The order of the scale values for gossip, fantasy and sustain in Figure 6.2 differs from the orders in Figures 4.2, 4.5, and 5.6; however, in all of these figures, the scale values for these qualities are nearly equal. The only other exception is that in Figure 5.6, the scale value for sustain is slightly larger than the scale value for positive play. The actual numerical values of the estimated scale values for the play qualities from the $(GT, P)$–212 model are nearly identical to the estimated values from the joint and conditional models.

The estimated scale values for groups and occasions from the $(GT, P)$-212 log 3-mode model are plotted in the same space in Figure 6.3. To take into account the differential contribution of each of the components with respect to explaining the interactions in the data, the scale values are multiplied by $\sqrt{\hat{\phi}_{r1r}}$ for $r = 1, 2$. Given this scaling, the inner product between vectors connecting the origin to the points corresponding to groups and occasions equal the "scores" of the groups on the occasions ($\phi^{**}_{i1k} = \sum_r \phi_{r1r}\mu_{ir}\eta_{kr}$).

The scale values for groups and occasions on the first component are similar to those from $(GT, P)$–111. The most noticeable difference is that the sign of the scale value for

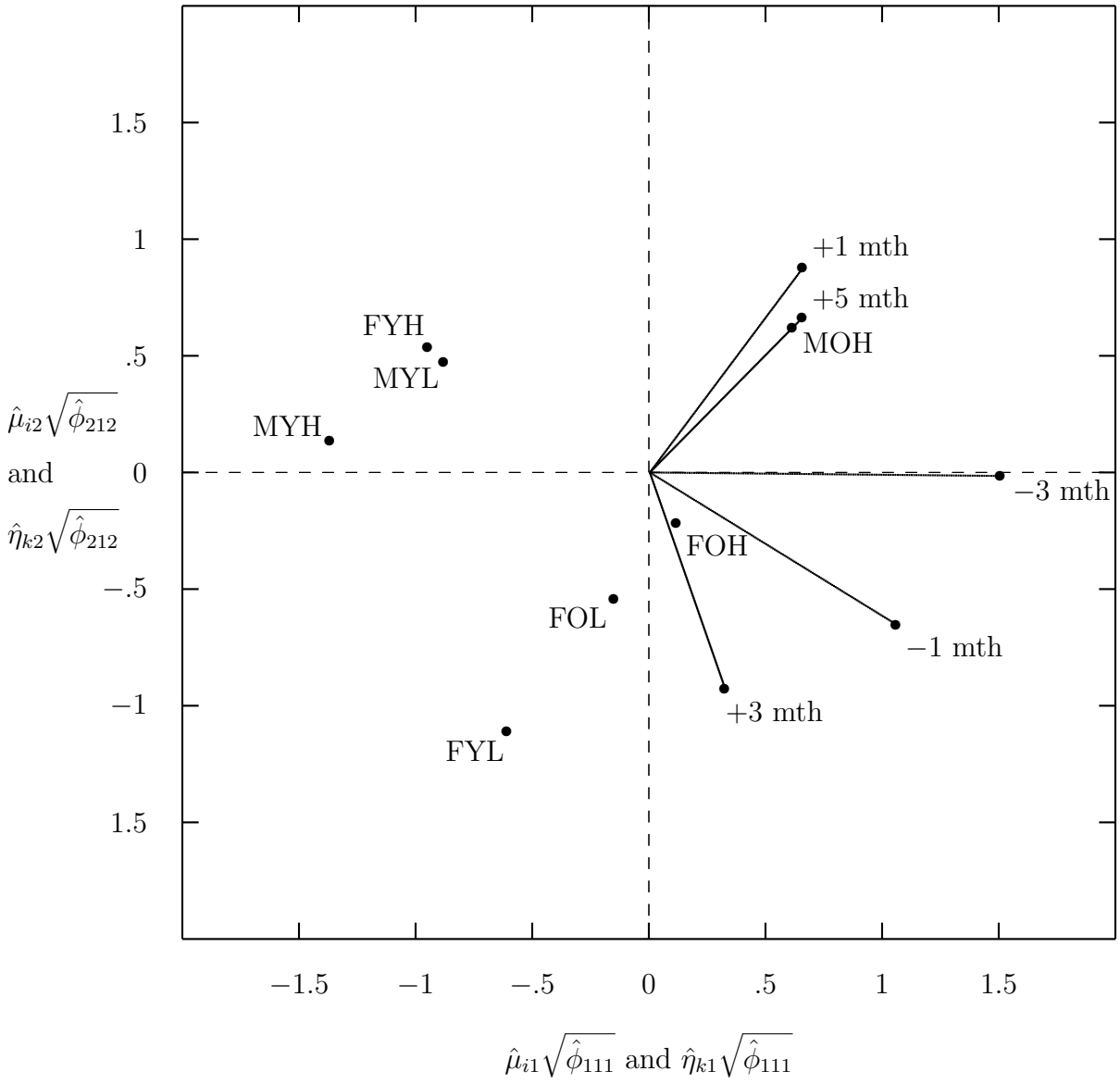



**Figure 6.3**: Joint plot of the estimated group and occasion scale values of the $(GT, P)$–212 log 3-mode model.

FOL differs; however, the estimated scale values for FOL in the 111 and 212 models are both small and close to zero. With respect to occasions, the overall pattern of the scale values on the first component is the same as that from 111. All of the scale values are positive, and they are ordered such that the scale value for first session is the largest, and the scale values for the last three sessions, which are nearly equal, are the smallest.

With respect to groups, the second component primarily contrasts the group FYL, which has a large negative scale value, from the groups FYH, MYL and MOH, which have nearly equivalent, positive values on the second component. Taking into account the scale values on both components simultaneously, the children in FYH and MYL are more similar to each other than they are to the children in any of the other groups (i.e., the points corresponding to FYH and MYL are very close). The play qualities exhibited on each occasion by the children in MYH are similar to that of the children in groups FYH and MYL (i.e., the point corresponding to MYH is relatively close to the points for FYH and MYL). The vectors connecting the origin to the points corresponding to FYL, FOL and FOH point in the same general direction (i.e., "down"), which indicates that the children in FYL, FOL and FOH exhibit a similar pattern of play qualities on each occasion. Since the point corresponding to FOH is the closest to the origin, the point for FOL is the next closest, and the point for FYL is the furthest, the pattern of interaction is the least pronounced or the weakest for the children in FOH, is the next weakest for FOL, and is the strongest for FYL.

With respect to occasions, the second component primarily contrasts the behavior of the children after the first session. The vector corresponding to the first occasion

is the longest and points in essentially the same direction as the first component (i.e., horizontal axis), which implies that the strongest interactions in the data occur at the first session. To the extent that the vector corresponding to the second session points in the same direction as the horizontal axis (i.e., the first session), the pattern of interactions observed at the first session carry over to the second session. Alternatively, to the extent that the vector for the second occasion deviates from horizontal and points in the same direction as the vector for +3 months, the pattern of interactions at the second session differs from that at the first session and is similar to those at the third session. The points corresponding to +1 and +5 months are relatively close to each other, which implies that the pattern and strength of associations in the data are similar on these two occasions.

There are similarities and differences in the configurations of points for groups and occasions in Figure 6.3 and the joint plot of the scale values for groups and occasions from the 212 *TUCKALS3* decomposition of the standardized residuals from the $(GT, P)$ loglinear model (Figure 5.8). For example, in both figures, the points corresponding to the groups MOH, FOH, FOL and FYL lie roughly on a straight line; however, in Figure 6.3, this line of points has a smaller slope and lies relatively closer to the origin than it does in Figure 5.8. The difference in slopes is trivial, because the axes can be arbitrarily rotated such that the slopes are identical. Relative distances between points and between points and the origin are not arbitrary, and they cannot be changed by rotation. In both figures, the vectors for occasions form a "fan" with +1 at the top, +3 at the bottom, and −3 approximately mid-way between +1 and +3. In Figure 5.8, the pattern for occasions differs in that the angle between the vectors corresponding to −3 and −1 is smaller than

it is in Figure 6.3, but in Figure 5.8, the angle between the vectors corresponding to +1 and +3 is larger than it is in Figure 6.3. While there are some differences, the overall configuration of points for group and occasions from the $(GT, P)$–212 log 3-mode model and the 212 *TUCKALS3* analysis are more similar than they are dissimilar. Further empirical comparisons of the results from these and other analyses of the peer play data are made in Chapter 7, where the theoretical basis for the empirical similarities among the models are discussed.

To aid the interpretation of the relationship between play qualities, groups and occasions, the scores for groups on each occasion, $\phi^{**}_{i1k} = \sum_r \phi_{r1r}\mu_{ir}\eta_{kr}$, are plotted in Figure 6.4. These scores are analogous and empirically similar to the estimated scale values for the groups at each occasion from the joint $RC(1)$ association model fit to ${}_{TG}\mathbf{F}_P$ table (Figure 4.3), the conditional $RC(1)$ association model fit to $\{\mathbf{F}_{GP|1}, \ldots, \mathbf{F}_{GP|5}\}$ with homogeneity constraints on the scale values for the play qualities (Figure 4.6), and the 212 *TUCKALS3* decomposition (Figure 5.9). Similarities among Figures 4.3, 4.6, and 5.9 were noted in Section 5.2.3 of the previous chapter. The interpretation of interactions in the data based on Figure 6.4 is the basically the same as the interpretation of Figures 4.3, 4.6 and 5.9.

In summary, on the basis of fit statistics alone, the $(GT, P)$–212 model appears to be the "best" model for the peer play data; however, the $(GT, P)$-111 model also yields a reasonable fit. These two models are relatively simple and both have reasonable substantive interpretations. The interpretation of the data based on these models leads to the conclusion that the younger children tend to exhibit more of the "bad" qualities than the

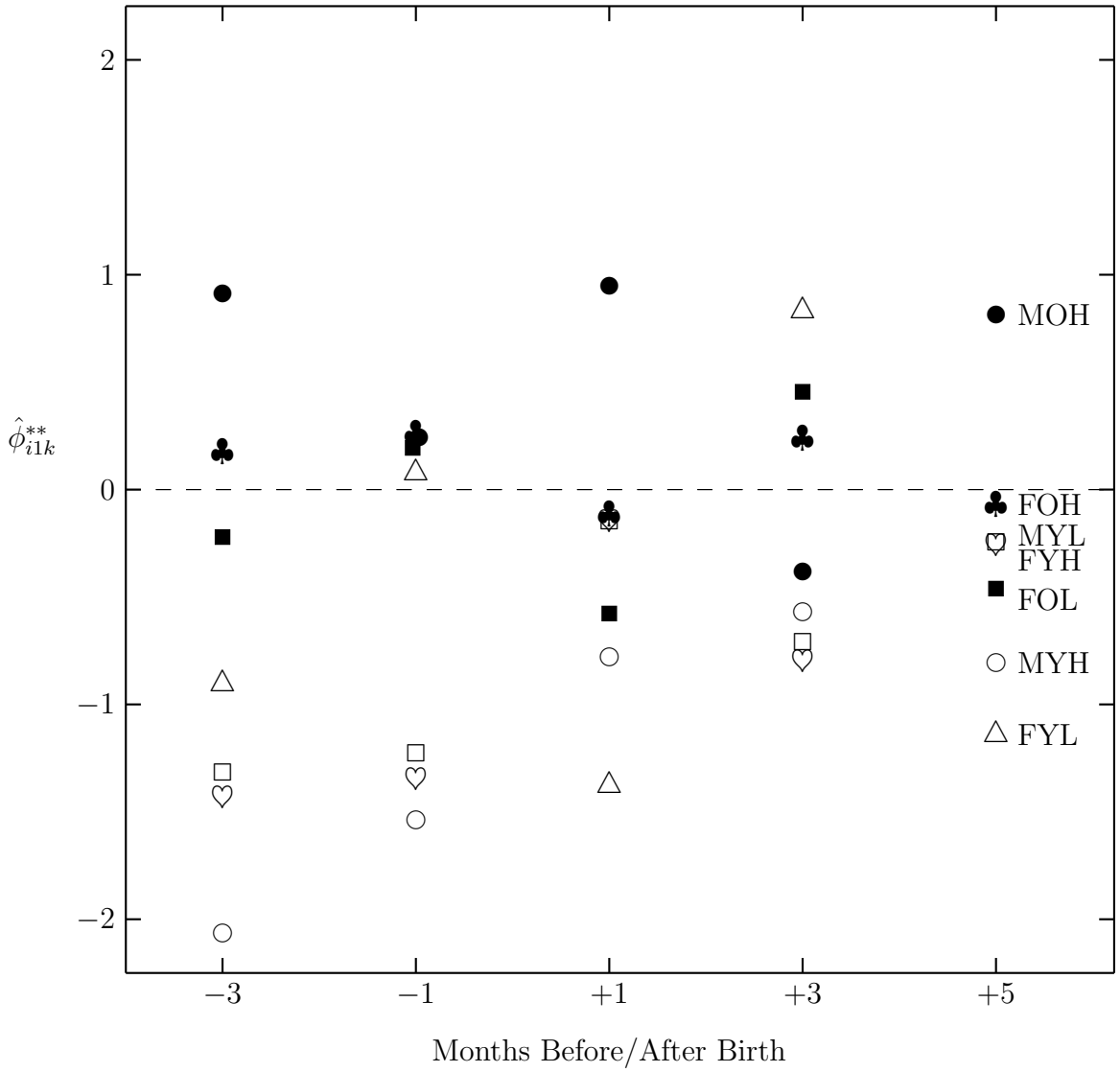



**Figure 6.4**: Estimated group scores at each occasion of the $(GT, P)$–212 log 3-mode model.

older children, the children in MOH exhibit more of the "good" qualities than children in any of the other groups, and the strength of this relationship decreases over time (i.e., the children are more similar to each other at the end of the study than they were at the beginning). The second components for groups and occasions from the $(GT, P)$-212 model adds to this interpretation primarily a description of differences between the groups after the first session. Residual analyses are presented in the next section for the 111 and 212 models to help decide whether they are adequate.

## 6.2 Residual Analyses

The residual analyses presented in this section are based on the hypothesis that a model, which includes the sampling assumptions and a particular mathematical structure, is "correct" or adequate for a particular data set. This premise leads to a number of predictions regarding the distribution of the residuals and particular relationships between residuals and fitted values. If the expected structures and relationships are present, then further support is gained for the conjecture that the model is correct.

Basic analyses of the residuals from all of the models fit to the peer play data were made. Since the models $(GT, P)$–111 and $(GT, P)$–212 were selected as the "best" models, more extensive analyses of the residuals from these model were performed. The results of the residual analyses for 111 and 212 are very similar. In the analyses where the results are essentially the same, only the results for the simpler model (i.e., 111) are

presented. When there are differences, the results for both the 111 and 212 models are presented.

In Section 6.2.1, predictions regarding the distribution of standardized residuals are discussed, and the observed distributions are compared with the predicted distributions. In Section 6.2.2, the residuals are examined and tested for systematic patterns where the models fail to fit the data. In Section 6.2.3, expectations based on the Poisson assumption are discussed, including implications of violations of the independence assumption.

### 6.2.1 Distribution of Standardized Residuals

If a model is correct, then the standardized residuals should be asymptotically distributed as standard normal variables; that is,

$$\hat{e}_{ijk} = \frac{f_{ijk} - \hat{f}_{ijk}}{\sqrt{\hat{f}_{ijk}}} \approx N(0,1)$$

where the $f_{ijk}$'s are the observed frequencies and the $\hat{f}_{ijk}$'s are the estimated fitted values from the model. In practice, standardized residuals tend to exhibit less variability than standard normal variables (Agresti, 1990).

In Figure 6.5, the standardized residuals from the $(GT, P)$-111 log 3-mode model are plotted against the expected values from the standard normal distribution. The line drawn in the figure is an identity line and represents the expected (asymptotic) relationship. Overall, the standardized residuals are slightly smaller than expected (i.e., for $\hat{e}_{ijk} < 0$, the points tend to be above the identity line, and for $\hat{e}_{ijk} > 0$, they tend to

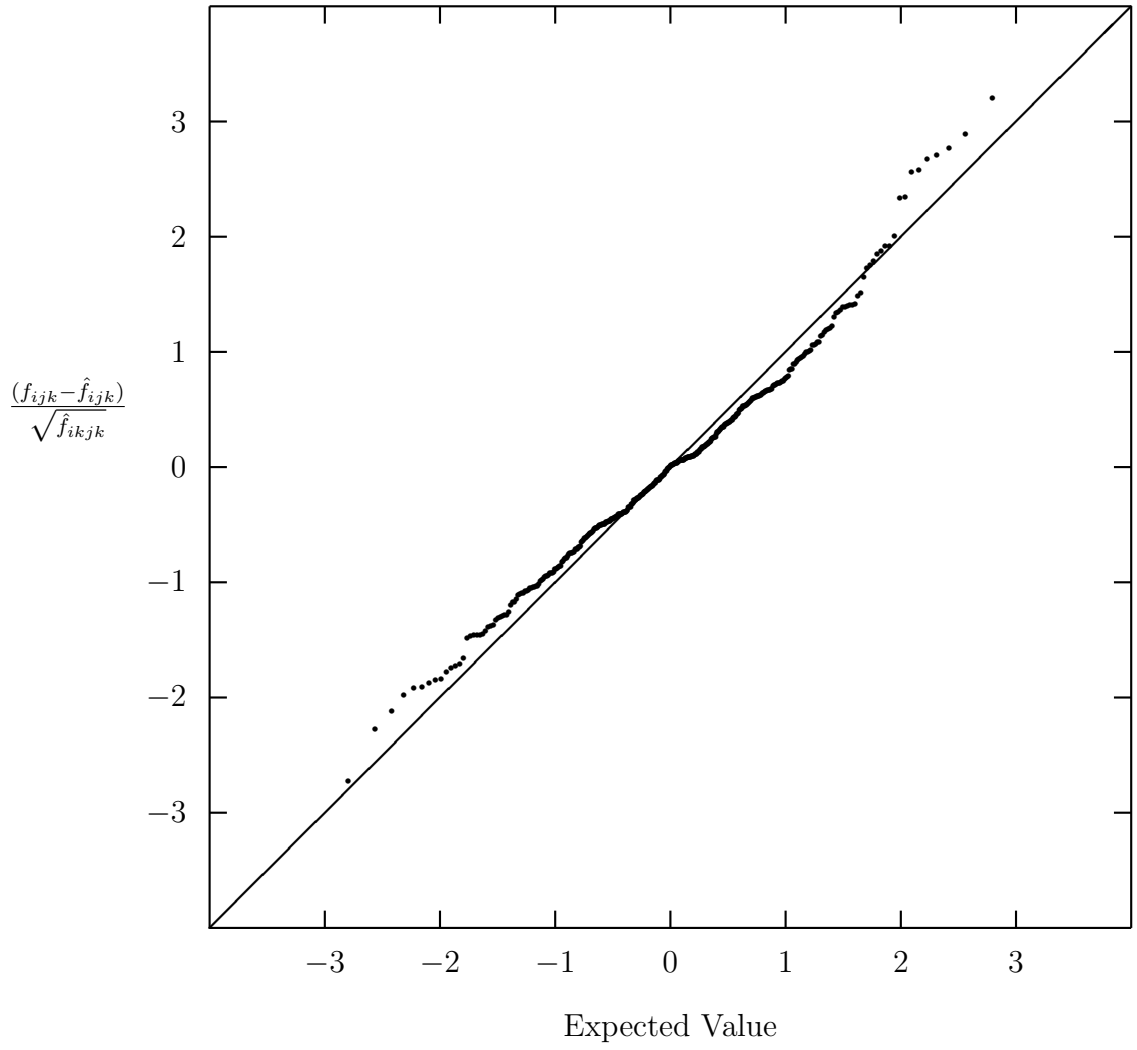



**Figure 6.5**: Normal probability plot of the standardized residuals from the $(GT, P)$–111 log 3-mode model.

be below the line). The major exceptions are 9 residuals, which are larger than expected and lie above the identity line in the upper right hand corner in Figure 6.5.

The variance of the $\hat{e}_{ijk}$'s equals .834, which is slightly smaller than the asymptotic expected value of 1. With $t_{384} = -0.052$ and $p$-value$> 0.9$, the mean, which equals $-0.0024$, is not statistically different from 0. The hypothesis that the $\hat{e}_{ijk}$'s come from a normal distribution was tested using the Shapiro-Wilk statistic $W$, which equals an estimated variance based on order statistics divided by the usual variance estimator (SAS Institute, 1988). The null hypothesis of normality is rejected for small values of $W$. The test statistic for the distribution of the $\hat{e}_{ijk}$'s equals $W = .9793$ and has a $p$-value$= .10$, which supports the conjecture that the $\hat{e}_{ijk}$'s are normally distributed.

Although there is not enough evidence to reject the hypothesis that the standardized residuals are distributed as expected under the assumption that the $(GT, P)$-111 model is correct, it may be possible that the distribution of the residuals for particular categories of the variables/modes deviate significantly from expectation. The distributions of the standardized residuals for each occasion, group and play quality were examined. These distributions should be asymptotically normally distributed with mean and variance equal to 0 and 1, respectively. The only outstanding deviations from this expectation are the variances for excite and fight, $\sigma^2_{\text{exc}} = 1.59$ and $\sigma^2_{\text{fight}} = 1.43$, respectively, which are relatively large.

Shapiro-Wilk statistics and $p$-values for testing the normality hypothesis for each group, play quality and occasion are reported in Table 6.4. The $p$-values for all the distributions except FOH, negative emotion, fight, and the third occasion (+1 mth) are

**Table 6.4**: Shapiro-Wilk test statistics for each mode from $(GT, P)$-111 log 3-mode model.

| Groups | W | $p$-value | Play | W | $p$-value | Occasions | W | $p$-value |
|---|---|---|---|---|---|---|---|---|
| FYL | .9639 | .190 | sus | .9710 | .553 | $-3$ mth | .9693 | .197 |
| FYH | .9740 | .471 | gos | .9714 | .565 | $-1$ mth | .9677 | .160 |
| FOL | .9764 | .561 | pos | .9694 | .508 | $+1$ mth | .9548 | .024 |
| FOH | .9475 | .032 | exc | .9783 | .768 | $+3$ mth | .9772 | .473 |
| MYL | .9582 | .106 | amt | .9693 | .505 | $+5$ mth | .9891 | .948 |
| MYH | .9857 | .896 | fan | .9588 | .267 | | | |
| MOH | .9957 | .999 | uns | .9800 | .816 | | | |
| | | | poor | .9697 | .515 | | | |
| | | | nem | .9177 | .014 | | | |
| | | | fght | .8956 | .003 | | | |
| | | | prh | .9792 | .794 | | | |

larger than .05. To control the family-wise error rate for each variable/mode, critical $p$-values for groups, play qualities and occasions were set equal to $\alpha_{\text{G}} = (.05/7) = .0071$, $\alpha_{\text{P}} = (.05/11) = .0045$ and $\alpha_{\text{T}} = (.05/5) = .01$, respectively. Using the more conservative values, the only distribution that significantly differs from normality is the distribution for fight. The distribution of the standardized residuals for fight is "well behaved", except for 5 of the residuals. The 5 large residuals are outliers and range from 2.009 to 2.705, while the next largest residual is 1.192.

The distribution of the standardized residuals from the $(GT, P)$-212 log 3-mode model is slightly better than the distribution from the 111 model. The $\hat{e}_{ijk}$'s from 212 are plotted in Figure 6.6. While the largest residuals from the 111 model were larger than expected, the largest $\hat{e}_{ijk}$'s from the 212 model are slightly smaller than expected. The variance of the $\hat{e}_{ijk}$'s equals .735, and the mean is not statistically different from zero ($t_{384} = -.026$, $p$-value= .979). The hypothesis of normality is not rejected ($W = .9822$, $p$-value= .284).

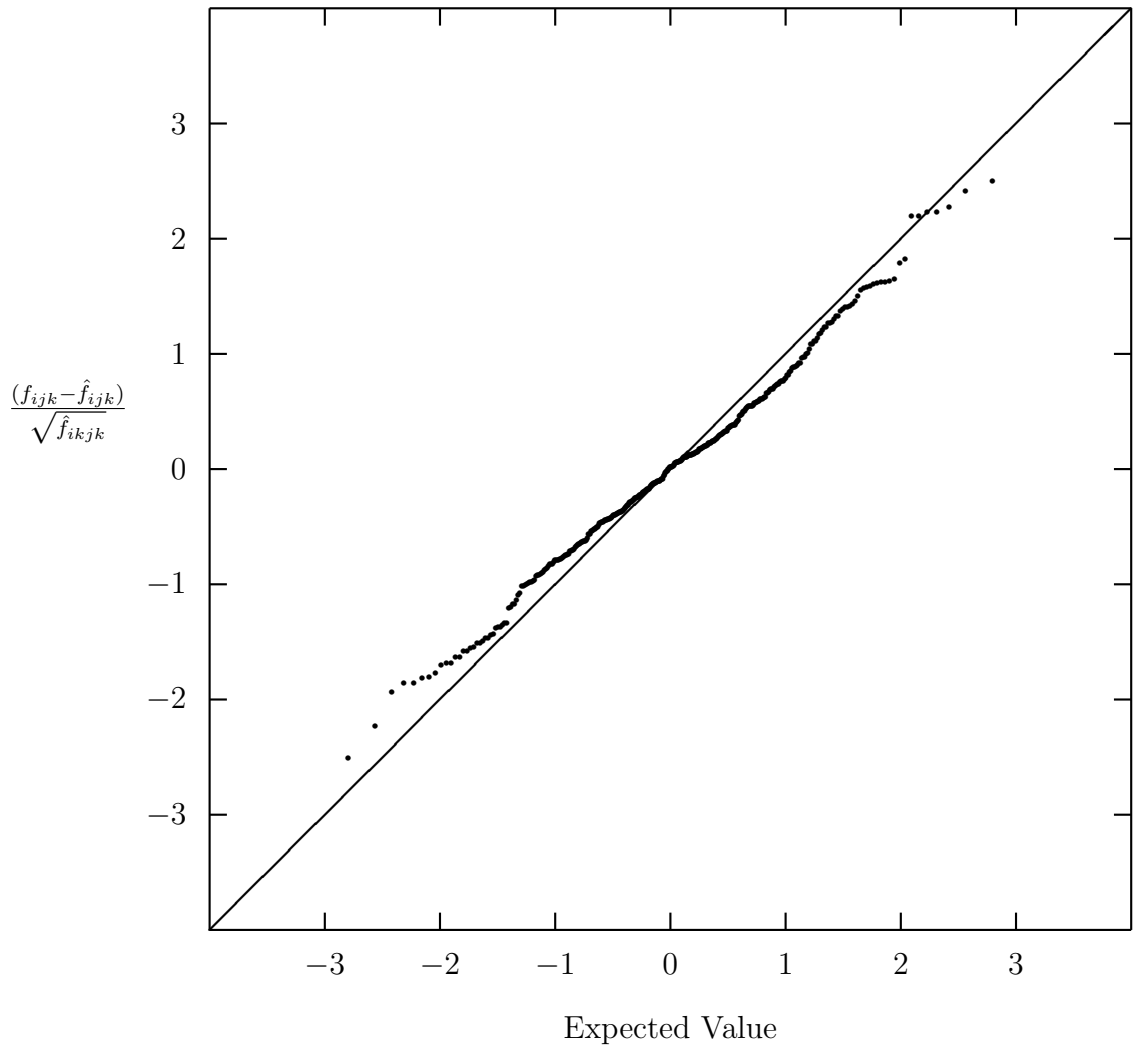



**Figure 6.6**: Normal probability plot of the standardized residuals from the $(GT, P)$–212 log 3-mode model.

The distribution of $\hat{e}_{ijk}$'s from the 212 model were also examined by categories of each mode, but no outstanding discrepancies with respect to expectations were found. None of the means are statistically different from 0, and none of the variances are large relative to 1. Except for FOH, which has a $W = .9390$ and $p$-value$= .012$, all of the tests for normality yield $p$-values greater than .05. Using the more conservative critical $p$-value for groups (i.e., $\alpha_{\mathrm{G}} = .0071$), the null hypothesis of normality for FOH is not rejected.

### 6.2.2 Systematic Discrepancies Between Fitted Values and Data

If a model is correct (or acceptable), then the residuals corresponding to certain categories or combinations of categories of the variables should not be consistently larger or smaller than 0. Since the standardized residuals are fairly well behaved (normal with approximately equal variances), one way to search for systematic structure in the residuals is to perform an ANOVA. Such an analysis can reveal categories and/or combinations of categories that have mean residuals that are statistically larger or smaller than zero. ANOVA's of the standardized residuals from both the 111 and 212 model were performed; however, none of the sources (i.e., 1– and 2–way marginal means of $\hat{e}_{ijk}$'s) were statistically different from 0.

Systematic patterns in the residuals were also searched for by identifying the largest residuals and determining whether too many of the large residuals are associated with a particular category or categories of the 3 variables/modes. Defining "large" as $|\hat{e}_{ijk}| > 2.00$, "large" standardized residuals were identified from both the 111 and 212 models. There are only 3 large negative residuals from the 111 model and 2 from the 212 model,

**Table 6.5**: The largest positive standardized residuals from the $(GT, P)$–111 log 3-mode model.

| Rank | Group | Play | Time | $f_{ijk}$ | $\hat{f}_{ijk}$ | $\hat{e}_{ijk}$ |
|---|---|---|---|---|---|---|
| 1 | FYL | nem | +1 | 9 | 3.233 | 3.208 |
| 2 | MYL | exc | +1 | 34 | 20.821 | 2.888 |
| 3 | FYL | exc | +3 | 14 | 6.779 | 2.774 |
| 4 | MOH | fght | +3 | 11 | 4.970 | 2.705 |
| 5 | MYL | fan | −3 | 77 | 56.837 | 2.675 |
| 6 | FYL | fght | +1 | 13 | 6.453 | 2.577 |
| 7 | FOH | poor | +1 | 18 | 9.929 | 2.562 |
| 8 | FOH | fght | −3 | 19 | 11.174 | 2.341 |
| 9 | MYH | fght | +5 | 10 | 4.859 | 2.332 |
| 10 | MYL | fght | +3 | 24 | 15.969 | 2.010 |

which can be seen in the normal probability plots (Figures 6.5 and 6.6). There does not appear to be any pattern in terms of the cells that are being overfit (i.e., negative residuals).

Both of the models underfit more cells of the table than they overfit (i.e., there are more positive than negative residuals). There are 10 "large" standardized residuals from the 111 model, 9 of which are noticeably larger than expected relative to the normal distribution. These residuals correspond to the positive $\hat{e}_{ijk}$'s in Figure 6.5 that are above the identity line. The 10 cells that are overfit are given in Table 6.5, where the observed frequencies ($f_{ijk}$), fitted values ($\hat{f}_{ijk}$) and standardized residuals ($\hat{e}_{ijk}$) are also reported.

Five of the residuals listed in Table 6.5 are from the play category fight and four of them are from the third occasion (+1). The groups with the most number of "large" residuals are MYL and FYL, each of which has 3 large residuals. Assuming that the

residuals from each category are equally probable, the probability of observing a certain number of large residuals from a particular category can be computed using the Binomial distribution. The computed probabilities can be used to decide whether the number of residuals observed from a category is statistically large and is an "unlikely event" under the hypothesis that the log 3-mode model is correct. For groups, the probability of observing 3 or more large residuals from the same group (i.e., MYL or FYL) is $\Pr[(X \geq 3|N = 10, p = (1/7)] = .161$, and for occasions, the probability of observing 4 or more large residuals from a single occasion (i.e., +1) is $\Pr[(X \geq 4|N = 10, p = (1/5)] = .121$. Neither of these probabilities provide evidence against the 111 model. For the play categories, the probability of observing 5 or more large residuals from a single category is $\Pr[X \geq 5|N = 10, p = (1/11)] = .001$, which implies that observing 5 of the largest 10 residuals from cells with the play category fight is an unlikely event. There are too many large residuals from the category fight, which was also noted when the distributions of the residuals were examined by category of the three modes.

For the 212 model, there are only 7 "large" positive standardized residuals, but only 2 of them are larger than expected relative to the normal distribution. Furthermore, none of the categories of any of the variables is over represented in the largest 7 $\hat{e}_{ijk}$'s from the 212 model.

### 6.2.3 Sampling Assumptions

Implications of the sampling assumption that observations are independent random samples from a Poisson distribution are outlined and examined in this section. Given the

design of the peer play study in which the same children are observed repeatedly at each occasion and across occasions, the independence assumption is highly suspect. It is quite likely that the observations are (positively) correlated within occasions, which will be referred to here as "re-sampling" dependence, and across occasions, which will be referred to as "serial" dependence. Another implicit assumption is that the children within the groups are homogeneous and have the same Poisson parameter (i.e., $\lambda_{g_{ijk}} = \lambda_{ijk}$, where $g_i$ indexes children within group $i$). Both heterogeneity and positively correlated observations are causes of "overdispersion" in which the variance of a count (i.e., observed frequency) is greater than the variance predicted by the model (Agresti, 1990; McCullagh & Nelder, 1989). To determine whether the sampling assumptions are violated and have an effect on the modeling, correlations between residuals, and the variance of the data and fitted values are examined in Sections 6.2.3.1 and 6.2.3.2, respectively.

#### 6.2.3.1 Correlations Between Standardized Residuals

Correlations between standardized residuals are examined primarily to assess whether observations are independent. Under the Poisson assumption, if observations are independent, for example, across index $i$, then the correlations between the standardized residuals $e_{ijk}$ and $e_{i'jk}$ should equal 0 for $i \neq i'$. Even when observations are independent, the computed correlations between standardized residual may not be zero. Small positive correlations could exist, because the residuals are a function of fitted values, which are a function of a common set of parameters. Since the number of cells in the tables is large (385), there is a fairly large number of parameters, and the function is nonlinear, this

**Table 6.6**: Inter-group correlations of the standardized residuals from $(GT, P)$–111 (lower triangle) and Bonferroni probabilities less than .999 that $\rho = 0$ (upper triangle).

| | FYL | FYH | FOL | FOH | MYL | MYH | MOH |
|---|---|---|---|---|---|---|---|
| FYL | — | | | | | | |
| FYH | −.004 | — | | .196 | | | |
| FOL | .075 | −.170 | — | | | | |
| FOH | .065 | .347 | .003 | — | .085 | | |
| MYL | −.041 | −.092 | −.258 | −.381 | — | .135 | |
| MYH | −.145 | −.204 | .113 | .179 | −.363 | — | |
| MOH | −.091 | .242 | −.165 | −.241 | .266 | −.137 | — |

influence is expected to be negligible. If observations are positively correlated, then the residuals should be positively correlated. If observations are negatively correlated, then the residuals should be negatively correlated.

Since groups consist of different children, it is unlikely that observations are dependent across groups; however, to ensure that no dependencies between groups were caused by any inadvertent systematic differences in the treatment of the children or the coding of data, correlations between groups were computed. The correlations for the $(GT, P)$-111 model are reported in the lower triangle of Table 6.6 and Bonferroni probabilities for the hypothesis that $\rho = 0$ are reported in the upper triangle. Since most of the Bonferroni probabilities are close to 1, only those than are less than .999 are reported. As expected, all of the correlations are relatively small and none of them are significantly different from 0. The same result was obtained for the 212 model.

Since the same children are observed repeatedly on the each occasion, the observations may be correlated across play categories. To examine whether dependencies due to repeated measures or re-sampling within occasions is a problem, the correlations between

**Table 6.7**: Inter-play quality correlations of the standardized residuals from $(GT, P)$–111 (lower triangle) and Bonferroni probabilities less than .999 $\rho = 0$ (upper triangle).

| | sus | gos | pos | exc | amt | fan | uns | poor | nem | fght | prh |
|---|---|---|---|---|---|---|---|---|---|---|---|
| sus | — | | | | | | | | .846 | | |
| gos | -.238 | — | | | | | | | | .690 | |
| pos | -.060 | -.094 | — | | | | | | | | |
| exc | -.063 | .211 | .036 | — | | | .675 | | | | |
| amt | -.099 | .291 | -.034 | .067 | — | | .016 | | | | |
| fan | .320 | .039 | -.345 | .083 | -.150 | — | | .231 | | | |
| uns | -.199 | -.237 | -.376 | -.419 | -.577 | .046 | — | | | | |
| poor | -.153 | .009 | .055 | -.373 | .183 | -.472 | -.005 | — | | | |
| nem | -.406 | -.157 | .065 | -.113 | .077 | -.352 | -.008 | -.053 | — | | |
| fght | -.133 | -.418 | .162 | -.321 | -.129 | -.176 | -.067 | -.045 | .225 | — | |
| prh | -.085 | -.158 | .043 | -.295 | -.264 | .025 | .278 | .124 | -.084 | -.106 | — |

the play categories of the standardized residuals from the 111 model were computed and are reported in Table 6.7. The largest correlation and the only one that is significantly different from 0 is the correlation between unsustain and amity, which equals $-.577$. While only one correlation is significantly different from 0, there does appear to be a slight pattern in Table 6.7. The largest correlations tend to be between the "good" qualities and the "bad" ones, and among these correlations, there are more negative correlations than positive (i.e., 21 negative versus 9 positive).

The correlations from the 212 model are similar to those in Table 6.7. The largest and only statistically large correlation is the correlation between unsustain and amity (i.e., $r = -.588$). The pattern observed in Table 6.7 where the correlations tend to be larger between the "good" and "bad" qualities rather than between two "good" qualities or two "bad" qualities is not found in the correlations from the 212 model. The pattern in Table 6.7 may be due just to chance.

Since the same children are observed on different occasions, the observations could be dependent across time. Correlations of the standardized residuals were computed between occasions to examine whether serial dependence is a problem. The correlations for the 111 model are reported in Table 6.8, along with the Bonferroni probabilities that are less than .999. The only correlation that is significantly different from 0 is the correlation between $-1$ and $+1$ month, which equals $-.335$ and has a $p$-value $= .029$. The correlations for the 212 model are similar to those in Table 6.8, except that the correlation between $-1$ and $+1$ is smaller and not significant (i.e., $r(-1, +1) = -.214$, $p$-value $= .621$). If serial dependence is present, then it is likely that the largest correlations would be between adjacent time points (e.g., the first and second occasion) and the smallest correlation would be between the occasions most separated in time (i.e., the first and last occasion). This pattern is not found in the correlations of the standardized residuals from the 111 or 212 model. The inter-occasion correlations do not provide evidence for serial dependence. If there was a serious problem with serial dependence, then it should have been apparent in the correlations. For example, Hausman, Hall and Griliches (1984) modeled count data from the same companies across time and found correlations between time points equal to approximately .8, which is clearly evidence for serial dependence.

Based on the inter-group and inter-occasion correlations of standardized residuals, which are relatively small and do not exhibit any consistent patterns, the assumption of independence across groups and across occasions is a close approximation to the truth. The inter-play quality correlations of the standardized residuals for the 111 and 212 model are too large to dismiss the possibility of re-sampling dependency; however, they

**Table 6.8**: Inter-occasion correlations of the standardized residuals from $(GT, P)$–111 (lower triangle) and Bonferroni probabilities less than .999 that correlations equal 0 (upper triangle).

| | −3 mth | −1 mth | +1 mth | +3 mth | +5 mth |
|---|---|---|---|---|---|
| −3 mth | — | | .199 | | .641 |
| −1 mth | .041 | — | .029 | | |
| +1 mth | −.265 | −.335 | — | | |
| +3 mth | −.187 | .127 | −.004 | — | |
| +5 mth | .212 | −.003 | .036 | −.111 | —- |

are small enough to retain the assumption of independence across play categories as an approximation to reality. In other words, there are some signs of dependency due to re-sampling, but the dependence does not appear to be very strong.

#### 6.2.3.2 The Variance Property of Poisson Variables

While the correlations between residuals do not provide enough evidence to rule out the independence assumption, dependency may be present, as well as other violations of the sampling assumptions (e.g., heterogeneity). Following Hausman, Hall and Griliches (1984), the variance property of Poisson distributed random variables is examined as an additional check for dependence and for heterogeneity.

Assume that $F_{ijk}$ is a random variable from a Poisson distribution where the Poisson parameter $\lambda_{ijk}$ equals

$$\lambda_{ijk} = \exp(u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \sum_r \sum_s \sum_t \phi_{rst} \mu_{ir} \nu_{js} \eta_{kt})$$

or

$$\ln(\lambda_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \sum_r \sum_s \sum_t \phi_{rst} \mu_{ir} \nu_{js} \eta_{kt}$$

Both the expected value, $E(F_{ijk})$, and the variance, $V(F_{ijk})$, equal $\lambda_{ijk}$. An additional useful property of Poisson random variables is that if observations are independent, for example, across index $i$, then $F_{+jk} = \sum_{i=1}^{I} F_{ijk}$ is a Poisson random variable with $\lambda_{jk} = \sum_{i=1}^{I} \lambda_{ijk}$, and the expected value and variance of $F_{+jk}$ equal $\lambda_{jk}$.

The children within groups may be heterogeneous, which would cause overdispersion (i.e., $E(F_{ijk}) < V(F_{ijk})$). For example, suppose that $\lambda_{g_i jk}$ equals the Poisson parameter of child $g_i$ in group $i$. If the $\lambda_{g_i jk}$'s are randomly distributed with mean $\lambda_{ijk}$, then there is additional variation in the data due to the variation of the $\lambda_{g_i jk}$'s around $\lambda_{ijk}$. Assuming that children are homogeneous, $F_{.jk} = (1/I) \sum_i F_{ijk}$ should be distributed as a Poisson random variable with mean and variance equal to $(1/I) \sum_i \lambda_{ijk}$. The mean of the fitted values $\hat{f}_{.jk}$ is an estimate of $F_{.jk}$, and the variance of the unstandardized residuals $e^*_{ijk} = (f_{ijk} - \hat{f}_{ijk})$ is an estimate of the variance of $F_{.jk}$ (i.e. $\hat{\sigma}^2_{jk} = \sum_i (e^*_{ijk} - \bar{e}^*_{.jk})^2/(I-1)$).

Since the distributions of the mean fitted values and the variance of residuals are quite skewed, the logarithms of the variances of the residuals are plotted against the logarithms of the mean fitted values in Figure 6.7. If the sampling assumptions are correct, then $\hat{f}_{.jk} = \hat{\sigma}^2_{jk}$ (or equivalently, $\ln(\hat{f}_{.jk}) = \ln(\hat{\sigma}^2_{jk})$), which corresponds to the identity line in Figure 6.7. The least squares regression line, which is labeled "Best fit" in Figure 6.7, has an intercept and slope equal to $-.530$ and 1.013, respectively. The 95% confidence intervals for the intercept and the slope are $[-1.496, .436]$ and $[.625, 1.401]$, respectively.

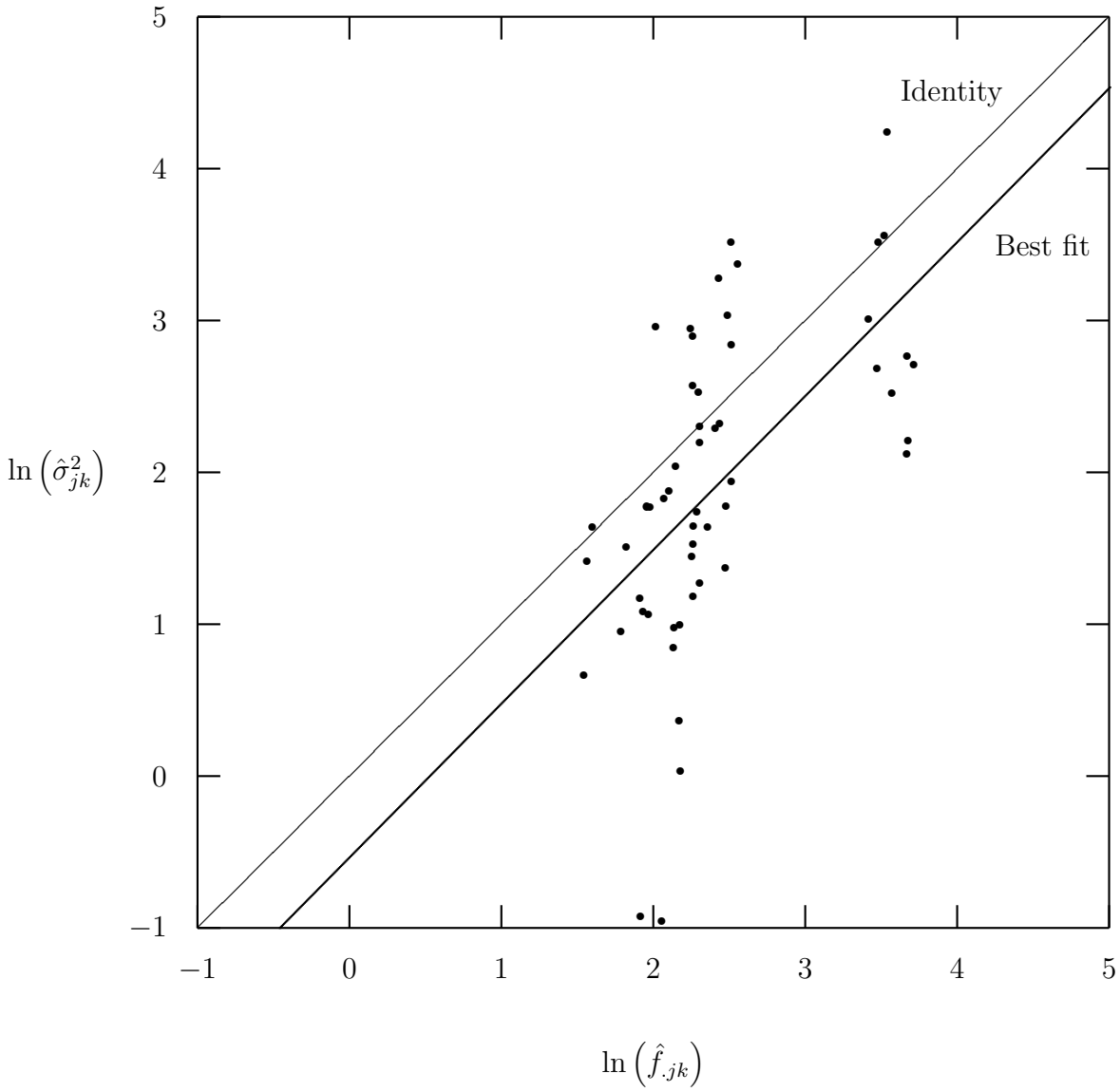



**Figure 6.7**: Within (Play × Occasion) residual variance versus average fitted values from $(GT, P)$–111.

Since the intervals contain 0 and 1, respectively, there is no evidence of overdispersion due to heterogeneity.

Overdispersion due to violation of the independence assumption by re-sampling within occasions can also be examined. If observations are independent within occasion, then $F_{i.k} = (1/J)\sum_j F_{ijk}$ should be randomly distributed as a Poisson variable with mean and variance equal to $\lambda_{i.k} = (1/J)\sum_j \lambda_{ijk}$. The mean of the fitted values $\hat{f}_{i.k} = (1/J)\sum_j \hat{f}_{ijk}$ was computed as an estimate of mean $F_{i.k}$ and the variance of the unstandardized residuals $\hat{\sigma}_{ik} = \sum_j (e^*_{ijk} - \bar{e}^*_{i.k})^2/(J-1)$, where $\bar{e}^*_{i.k} = \sum_j e^*_{ijk}/J$, was computed as an estimate of the variance. In Figure 6.8, the logarithms of the variance estimates, $\ln(\hat{\sigma}^2_{ik})$, are plotted against the logarithms of the mean estimates, $\hat{f}_{i.k}$. The least squares regression line, which has an intercept and slope equal to .086 and .844, respectively, is quite similar to the identity line. The 95% confidence intervals for intercept and slope (i.e., $[-.843, 1.015]$ and $[.481, 1.201]$, respectively) contain 0 and 1, respectively. Since the variance does not increase more rapidly than the mean, Figure 6.8 provides further support for the conclusion that dependence due to re-sampling within sessions is not a problem.

As a final check for serial dependence, the logarithms of the estimated variances of the unstandardized residuals $\ln(\hat{\sigma}^2_{ij})$ were plotted against the logarithms of the mean fitted values $\ln(\hat{f}_{ij.})$ in Figure 6.9. If there is serial dependence, then the variance should increase faster than the mean. The least squares regression line has a slope equal to .915 and intercept equal to $-.344$, and is not significantly different from the identity line (i.e., 95% confidence interval for the intercept, $[-1.014, .326]$, contains 0, and the interval for the slope, $[.689, 1.191]$, contains 1).

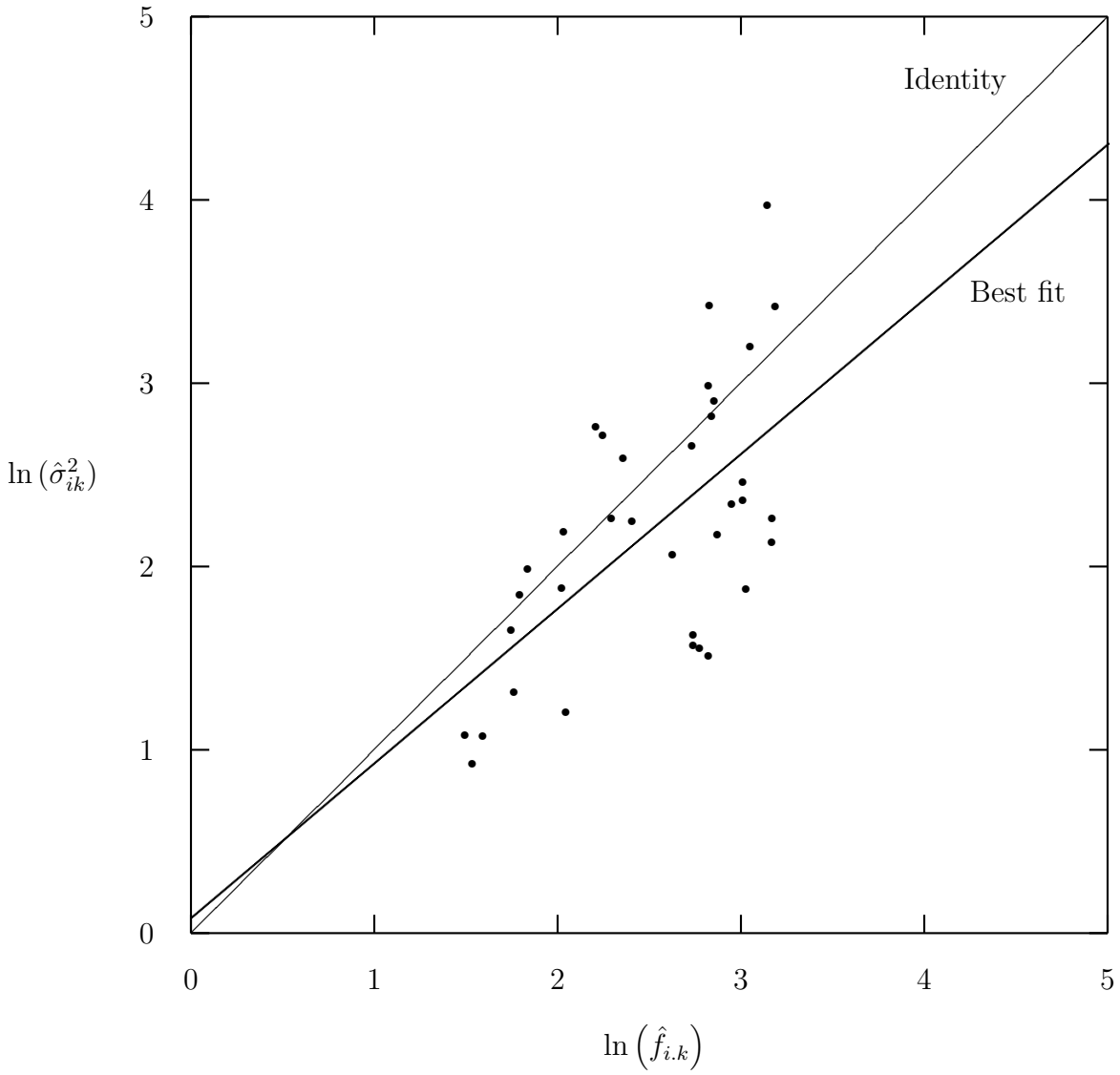



**Figure 6.8**: Within (Group × Occasion) residual variance versus average fitted values from $(GT, P)$–111.

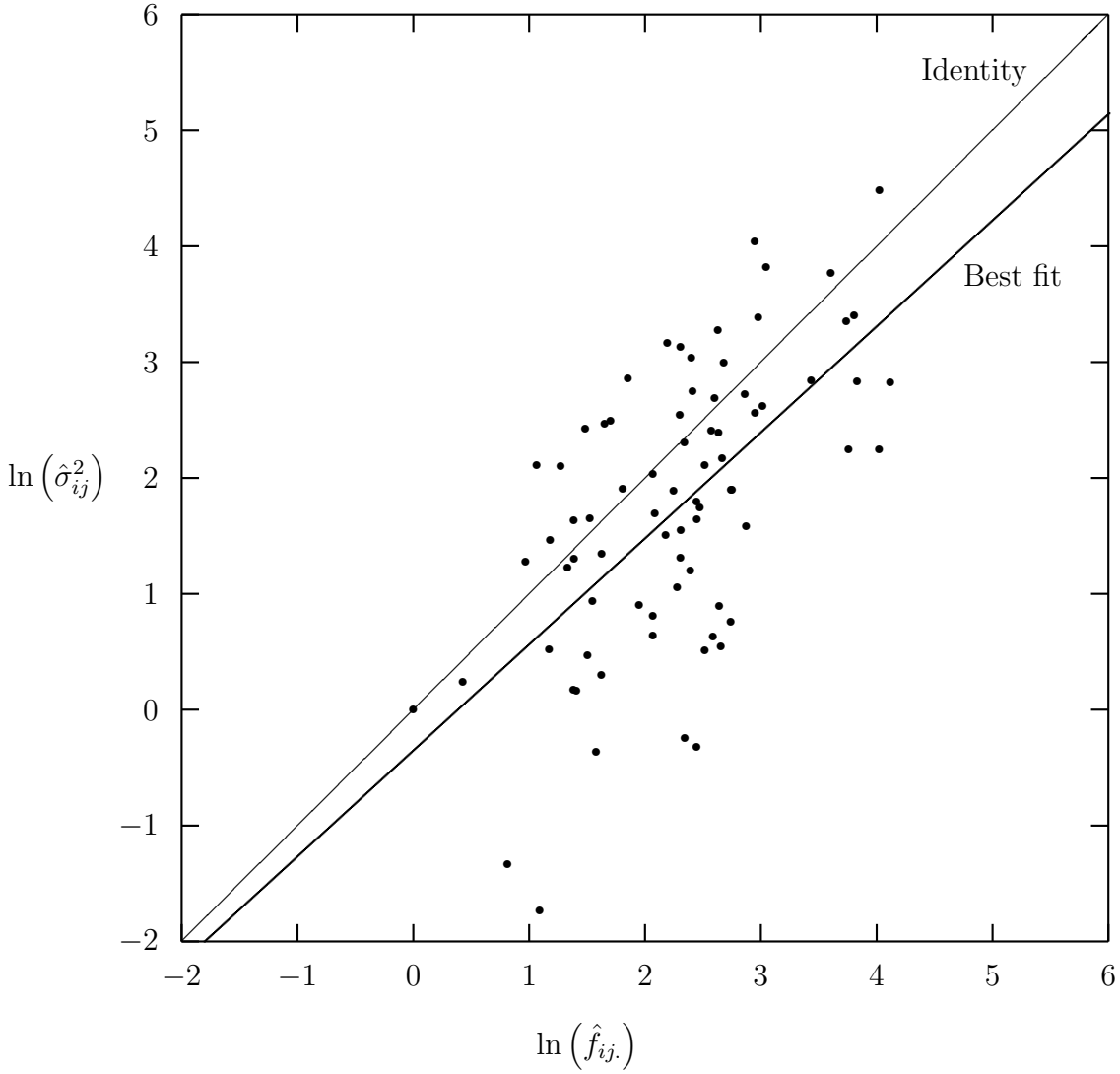



**Figure 6.9**: Within (Group × Play) residual variance versus average fitted values from $(GT, P)$–111.

The variance property was also examined for the $(GT, P)$–212 model to determine whether heterogeneity, re-sampling dependence and serial dependence are problems. The results are very similar to those from the 111 model, which are displayed in Figures 6.7, 6.8, and 6.9.

Since there is no evidence of overdispersion and the correlations between standardized residuals are small, the sampling assumption that observations are independent Poisson variables is reasonable. Given the design of the study, this result is somewhat surprising. Caution should still be exercised with respect to relying on $p$-values of likelihood ratio test statistics and using estimates of standard errors of parameters. (Even though standard errors of parameters would have been useful, they were not estimated to avoid the temptation of interpreting them.)

In summary, the 212 model yields a slightly better fit than the 111 model. The distributions of the residuals agree with expectations and no large systematic patterns of lack-of-fit were found for either the 111 or 212 model. The only exception is that for the 111 model, there are too many large residuals from the play category "fight". If the simplest possible model is desired, then the 111 model is preferable, because it is the most parsimonious model that yields an reasonable fit. If a more detailed picture of the associations and a closer approximation of the data is desired, then 212 is preferable.

## 6.3 Missing Data

There were 30 children in the peer play study, but since 5 of them were absent on the third occasion, the subset of the peer play data consisting only of the data from the 25 children present on all occasions (Table 2.2) was used in examples of the models and analyses presented in this and the previous 3 chapters. Of the 5 children with missing data, 4 of them were classified into one of the 7 groups studied in the 25 child sub-sample (one child for each of the groups FYL, FOL, FOH and MYH). The other child with missing data was an older male with low sibling acceptance (MOL). Since there was only one child in the MOL group, there are structural zeros in the table of data from all 30 children (Table 2.1).

Using only the data from the 25 children present on all occasions ignored the data from 5 children on 4 occasions. Including the data from the 4 children who were in one of the 7 groups studied in the 25 child sub-sample yields a $(7 \times 11 \times 5)$ table of frequencies based on a 29 child sub-sample. The frequencies from the 29 child sub-sample and the full 30 child sample are analyzed by log 3-mode models in Sections 6.3.1 and 6.3.2, respectively. The effect of the missing data on the analyses are summarized in Section 6.3.3.

### 6.3.1 The 29 Child Sub-sample

The 29 child sub-sample consists of the frequencies in Table 2.1 for the 7 groups, FYL, FYH, FOL, FOH, MYL, MYH and MOH. The data from 29 child sub-sample was modeled to determine the effect of slightly different group compositions on the results and

indicate the stability of the results based on the 25 child sub-sample. Since the missing data are from 4 of the 7 groups on the third occasion, the $GT$ margin is systematically affected by the missing data. The frequencies at +1 month for the groups FYL, FOL, FOH and MYH are smaller, not because of any change in the behavior of the children in these groups, but because of the missing data. Since the $GT$ margin is fit perfectly in the $(GT, P)$ log 3-mode model, there should only be a minimal effect on the estimated scale values that represent the $GP$, $PT$, and $GPT$ interactions. The basic interpretation of the associations in the data should be the same for the 25 and 29 child sub-samples.

The fit statistics for the $(GT, P)$ log 3-mode model fit to the 29 child sub-sample are reported in Table 6.9. The degrees of freedom for all of the models are the same as those for the 25 child sub-sample (see Table 6.2), but the fit statistics are smaller for the 29 child sub-sample than they are for the 25 child sub-sample. When sample sizes are increased, models usually fit worse; however, the data from the additional 4 children appear to have smoothed out some of the idiosyncrasies in the data.

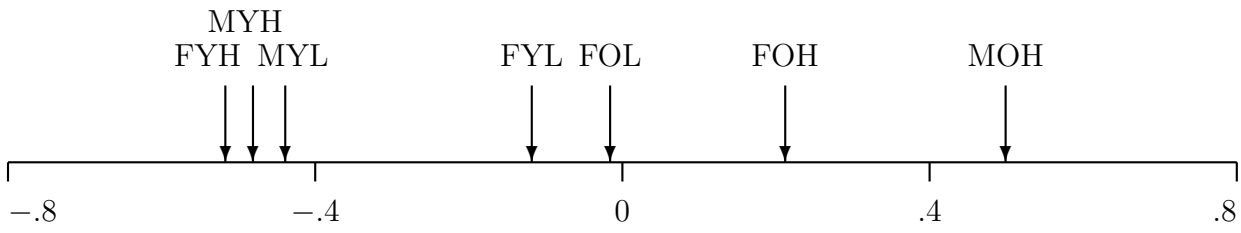
The $(GT, P)$–000 (loglinear) model clearly does not fit the data; however, adding just 1 component for each mode yields the 111 model, which fits the data quite well. Among the models in the next block, both the 212 and 221 models fit significantly better than the 111 model. In terms of fit statistics and parsimony, the best model in this block of model is the 212 model. The 212 and 221 models have nearly equal $G^2$'s, but 212 has fewer degrees of freedom than 221. The models with more components fit even better, but the additional complexity is not necessary to adequately model the data.

**Table 6.9**: Fit statistics of $(GT, P)$ log 3-mode models ($N = 29$).

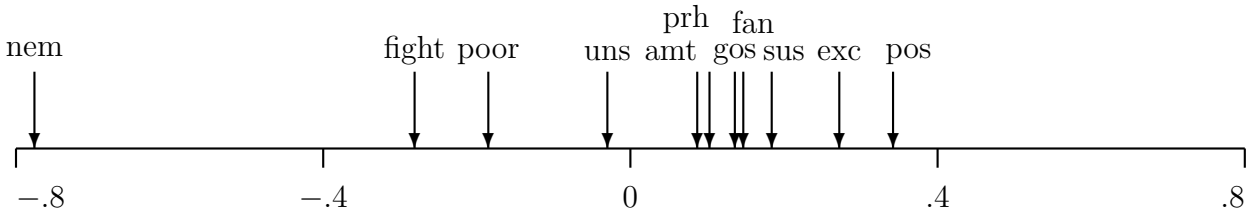
| Components | | | Log 3–mode model statistics | | | | | *TUCKALS3* |
|---|---|---|---|---|---|---|---|---|
| Group | Play | Time | *df* | $G^2$ | *p*-value | $G^2/df$ | $X^2$ | $X^2$(Residual) |
| 0 | 0 | 0 | 340 | 430.07 | .001 | 1.26 | 407.68 | 407.68 |
| 1 | 1 | 1 | 320 | 327.16 | .379 | 1.02 | 312.15 | 307.97 |
| 2 | 1 | 2 | 311 | 293.91 | .749 | .94 | 270.60 | 279.97 |
| 1 | 2 | 2 | 308 | 312.52 | .418 | 1.01 | 294.50 | 300.35 |
| 2 | 2 | 1 | 306 | 291.02 | .722 | .95 | 280.15 | 270.81 |
| 3 | 1 | 3 | 304 | 279.44 | .841 | .92 | 256.98 | 260.91 |
| 1 | 3 | 3 | 298 | 305.55 | .369 | 1.03 | 287.51 | 295.45 |
| 3 | 3 | 1 | 294 | 273.82 | .795 | .93 | 263.08 | 264.43 |
| 2 | 2 | 2 | 297 | 270.10 | .867 | .91 | 254.28 | 256.98 |
| 2 | 2 | 3 | 291 | 272.20 | .779 | .94 | 253.50 | 248.04 |
| 3 | 2 | 2 | 289 | 251.42 | .946 | .87 | 234.95 | 242.21 |
| 2 | 3 | 2 | 286 | 264.34 | .816 | .92 | 247.91 | 251.17 |
| 3 | 2 | 3 | 281 | 242.83 | .952 | .86 | 225.87 | 233.71 |
| 2 | 3 | 3 | 278 | 250.44 | .881 | .90 | 239.32 | 237.78 |
| 3 | 3 | 2 | 276 | 229.98 | .980 | .83 | 219.35 | 222.37 |
| 3 | 3 | 3 | 265 | 207.05 | .997 | .78 | 197.84 | 206.55 |

The estimated scale values and the residuals from the 111 and 212 models were studied and were found to be quite similar to those from the 25 child sub-sample. The residuals from the 212 model were remarkably close to being normally distributed. The 212 model is the better model, but for the sake of simplicity, the estimated scale values from the 111 model are presented here.

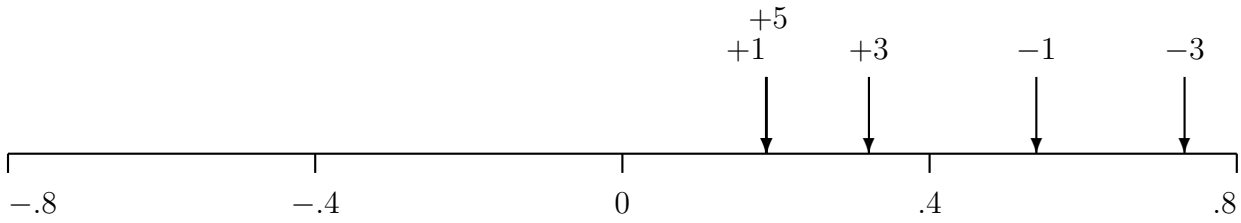
The estimated scale values of the 111 model fit to the 29 child sub-sample are plotted in Figure 6.10. The estimated intrinsic association parameter is $\hat{\phi}_{rst} = 3.761$. The order and relative spacing of the scale values for groups, play qualities and occasions are quite similar to the order and spacing of the estimated scale values for the 25 child sub-sample (see Figure 6.1).




Groups: $\hat{\mu}_{i1}$




Play Qualities: $\hat{\nu}_{j1}$




Occasions: $\hat{\eta}_{k1}$

**Figure 6.10**: Estimated scale values of the $(GT, P)$–111 log 3-mode model fit to the 29 child sub-sample.

Except for one scale value, the order of the scale values for groups is the same as the order from the 25 child sub-sample. The only exception is that the order of the scale values for FYH and MYH is reversed. The changes in the location of the scale values in Figure 6.10 relative to Figure 6.1 are minor, and they primarily involve groups that had missing values. For example, in Figure 6.10, MYH is closer to FYL and MYL, FYL is closer to zero, and FOH is closer to MOH than they are in Figure 6.1.

Except for one minor difference, the order of the scale values for play qualities is the same for the 25 and 29 child sub-samples. The only difference involves prohibit, which has a relatively small positive value in the 29 child sub-sample but has a small negative value in the 25 child sub-sample. The relative locations of the play qualities are quite similar for the two samples.

The most noticeable different in the relative position of the scale values for occasions is that instead of the last three occasions being clustered closely together as they are for the 25 child sub-sample, the fourth occasion (+3 mth) has a larger scale value than the third and fifth occasions (+1 and +5 mth), which have approximately equal scale values. The scale value for the fourth occasion is still closer to the scale values for the third and fifth occasions than it is to the scale value for the second occasion (−1 mth).

Overall, the interpretation of the associations in the data is the same for the 25 and 29 child sub-samples. The older children, in particular those classified as FOH or MOH, exhibit more "good" play qualities and fewer "bad" ones, while the younger children, in particular those classified as FYH, MYL or MYH, exhibit more immature qualities than mature ones. This relationship is the strongest at the first session (−3 mth), slightly

weaker on the second occasion ($-1$ mth), and is weaker for the three sessions after the birth of the children's siblings.

### 6.3.2 The Full Data Set

Even though Table 2.1, which consists of all of the available data, has 11 structural zeros, it can be analyzed by log 3-mode models. Only two minor adjustments are required to the algorithm that fits the model to the data. First, the fitted values for the cells containing the structural zeros need to be fixed to 0, which reduces the model degrees of freedom by 11, one for each structural zero. Since the observed frequencies and fitted values both equal 0 for these 11 cells, these cells have no effect on the up-dating equations for any of the model parameters. Second, since all of the structural zeros are for the MOL group classification on the third session, the interaction parameter $u^{GT}_{(\mathrm{MOL},3)}$ cannot be estimated, which increases the degrees of freedom by one. The details on how to fit models where the fitted values for particular cells are constrained to equal the observed frequencies are given in Appendix B.

The 29 child sub-sample falls into a $(7 \times 11 \times 5)$ sub-table of the $(8 \times 11 \times 5)$ table (consisting of the data from all 30 children). Since log 3-mode models are sub-table invariant, a log 3-mode model that fits a table *perfectly* will also fit any of its sub-tables perfectly. The parameters of the model fit to the sub-table equal either the parameters of the model fit to the full table or a simple re-scaling of these parameters. Some of the parameters need to be re-scaled to take into account the different number of frequencies and categories in the sub-table. For example, assume that a log 3-mode model perfectly

fits the $(8 \times 11 \times 5)$ table (i.e., the data from all 30 children) and that it is also fit to the $(7 \times 11 \times 5)$ sub-table (i.e., the 29 child sub-sample). The parameters of the log 3-mode model fit to the $(7 \times 11 \times 5)$ sub-table would equal the corresponding parameters of the model for the full table, except the parameters $u_i^G$, $u_{ik}^{GT}$, $\mu_{ir}$, $\phi_{rst}$, and $u$. These parameters equal the corresponding parameters of the model for the full table re-scaled to take into account the different number of categories for groups (e.g., $\mu_{ir}^* = \mu_{ir}/\sqrt{\sum_{i \neq \text{MOL}} \mu_{ir}^2}$, where $\mu_{ir}$ and $\mu_{ir}^*$ are parameters for the full table and sub-table, respectively), and the different sample sizes.

The sub-table invariance property holds *exactly* only when a log 3-mode model fits a table perfectly. However, when a model fits well but not perfectly, there will be just small differences in the estimated parameters of the models fit to the table and to any of its sub-tables. The estimated parameters from fitting the model to the table and a sub-table become more similar if the fit of the model to the table improves. With the peer play data, the estimated parameters of the log 3-model models fit to the frequencies from the 30 children will not be exactly equal to the parameters from the models for the 29 child sub-table, because the models do not fit the data perfectly. However, if a model fits the $(8 \times 11 \times 5)$ table well, the parameters should be very similar to those from the corresponding model fit to the $(7 \times 11 \times 5)$ sub-table.

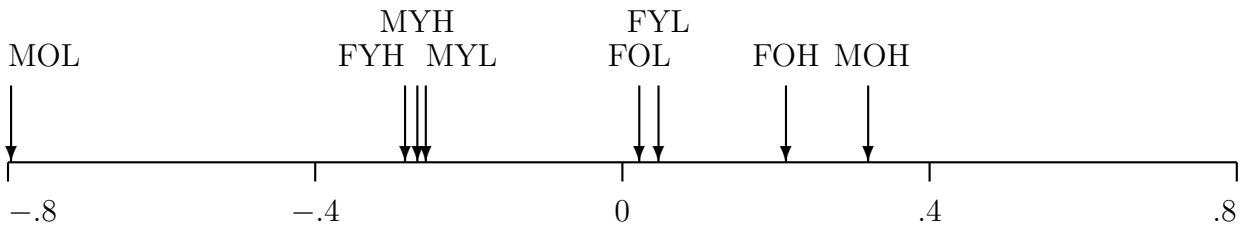
The fit statistics for the $(GT, P)$ log 3-mode models fit to the $(8 \times 11 \times 5)$ table are reported in Table 6.11. The pattern of the fit statistics for the models fit to the 30 child sample is very similar to that of the models that were fit to the 25 and 29 child sub-samples. In particular, the $(GT, P)$ loglinear model clearly does not fit, but

**Table 6.10**: Fit statistics of $(GT, P)$ log 3-mode models ($N = 30$).

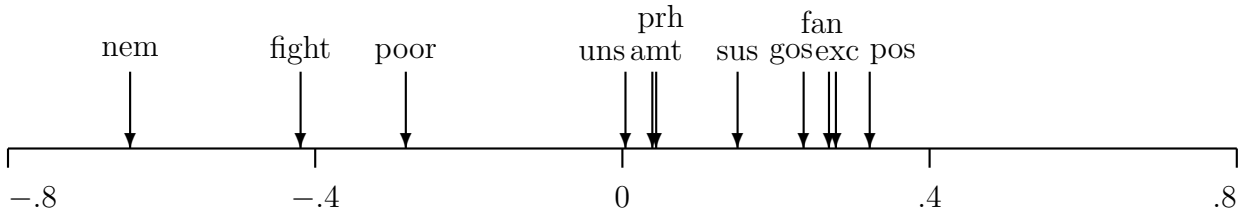
| Components | | | Log 3–mode model statistics | | | | | *TUCKALS3* |
|---|---|---|---|---|---|---|---|---|
| Group | Play | Time | *df* | $G^2$ | *p*-value | $G^2/df$ | $X^2$ | $X^2$(Residual) |
| 0 | 0 | 0 | 380 | 534.14 | .000 | 1.41 | 511.87 | 511.87 |
| 1 | 1 | 1 | 359 | 377.81 | .249 | 1.05 | 356.81 | 344.04 |
| 2 | 1 | 2 | 349 | 344.04 | .565 | .98 | 318.33 | 318.12 |
| 1 | 2 | 2 | 347 | 367.12 | .219 | 1.06 | 345.52 | 339.70 |
| 2 | 2 | 1 | 344 | 339.54 | .558 | .99 | 321.38 | 314.26 |
| 2 | 2 | 2 | 335 | 319.06 | .809 | .96 | 298.74 | 304.49 |
| 3 | 3 | 3 | 302 | 268.43 | .918 | .89 | 248.98 | 243.80 |

the 111 model fits reasonably well. Among the models in the next block, 212, 122 and 221, the 122 model does not fit much better than 111, and the 212 model is the best fitting model in the block. The 222 model fits better than the 212 model, but a second component for play qualities does not appear to be necessary. The reduction in model fit from deleting the second component for play in the 222 model is not statistically "large" (i.e., $G^2 = (344.04 - 319.06) = 24.09$, $df = 15$, and $p$-value= .064). Using a model more complex than 212 is not necessary.

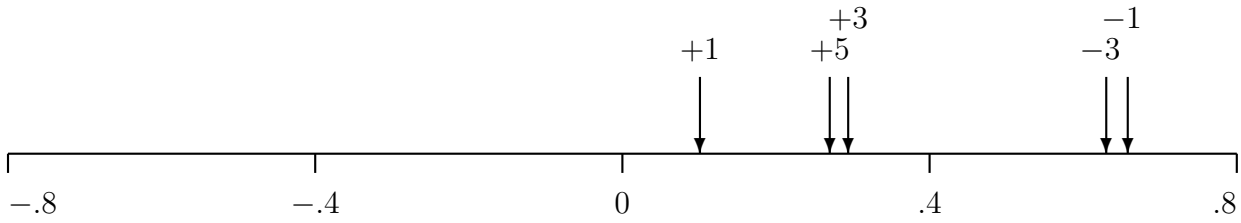
The scale values and residuals for the 111 and 212 models were examined and studied. The residuals are slightly better (smaller and more "well behaved") for the 212 model. The residuals for the 111 and 212 models are similar in value and distribution to the residuals from the 111 and 212 models, respectively, fit to the 29 child sub-table. The estimated scale values were found to be very similar to those from the 111 and 212 models fit to the 29 child sub-sample. For the sake of simplicity, the estimated scale values from the 111 model are presented and discussed here.




Groups: $\hat{\mu}_{i1}$




Play Qualities: $\hat{\nu}_{j1}$




Occasions: $\hat{\eta}_{k1}$

**Figure 6.11**: Estimated scale values of the $(GT, P)$–111 log 3-mode model fit to the entire 30 child sample.

The estimated scale values for the $(GT, P)$–111 log 3-mode model are plotted in Figure 6.11. The estimated intrinsic association parameter $\hat{\phi}_{111}$ is 5.485. The order and relative location of the scale values for the play qualities in Figure 6.11 is basically the same as the order and location in Figure 6.10. The only exception is that fantasy and gossip are closer to excite in Figure 6.11 than they are in Figure 6.10.

In Figure 6.11, the basic split between the scale values for the occasions before the birth and after the birth of the siblings is still present, but within the before and after sessions, there are differences in the order of the scale values. For the two sessions before the birth of the siblings, the scale values for the first 2 occasions ($-3$ and $-1$) are approximately equal in Figure 6.11, but in Figure 6.10, the scale value for $-3$ is larger than the scale value for $-1$. For the three sessions after the births, in Figure 6.11, the scale value for $+1$ is smaller than the scale values for $+3$ and $+5$, which are approximately equal, but in the Figure 6.10, the scale value for $+1$ is approximately equal to the value for $+5$, which is smaller than the scale value for $+3$.

The differences between the scale values for occasions in Figures 6.11 and 6.10 can be explained by the presence/absence of the MOL group. The child in the MOL group classification has the largest negative scale value, which is fairly extreme relative to the scale values for the other groups. The MOL child exhibited a lot more negative emotion, fighting and poor play and exhibited fewer of the "good" qualities than any of the other children. Since there are structural zeros for this group on the third occasion, the extreme behavior of this child did not affect the scale value for the $+1$ occasion, but it did affect the scale values for all of the other sessions. The extreme behavior of the child in the

MOL group contributes substantially to the association present in the table and "pushes" up the scale values assigned to $-1$ and $+5$ relative to the other scale values.

To facilitate a comparison of the group scale values of the 111 model fit to the 25, 29 and 30 child samples, the scale values from the 30 child sample were re-scaled such that the $\sum_{i \neq \text{MYL}} \hat{\mu}_i^2 = 1$. These re-scaled values and the estimated scale values from the 25 and 29 child sub-samples are all plotted in Figure 6.12. The re-scaled value for MOL from the 30 child sample, which equals $-1.313$, is off the scale in Figure 6.12. The scale values for the 7 groups from the 29 and 30 child samples are quite similar. They are not exactly the same, because the models do not fit the data perfectly. The only difference in the order of the scale values is that the scale value for FYL is slightly larger than FOL for the 30 child sample, while FYL is slightly smaller than FOL for the 29 child sub-sample. This is a minor difference, since the scale values for FYL and FOL are close to 0 for both the $N = 29$ and 30.

Given the overall similarities between the scale values fit to the $(8 \times 11 \times 5)$ table and the $(7 \times 11 \times 5)$ sub-table, the $(GT, P)$–111 log 3-mode model is approximately sub-table invariant. Although not presented here, the scale values of the $(GT, P)$–212 model fit to the full data set and the 29 child sub-sample are also quite similar.

### 6.3.3 The Effect of Missing Data

By including the data from children who had missing observations, the effect on the modeling and substantive interpretation of the associations in the data can be examined. From the log 3-mode modeling of the 25, 29 and 30 child samples, the $(GT, P)$–212 model

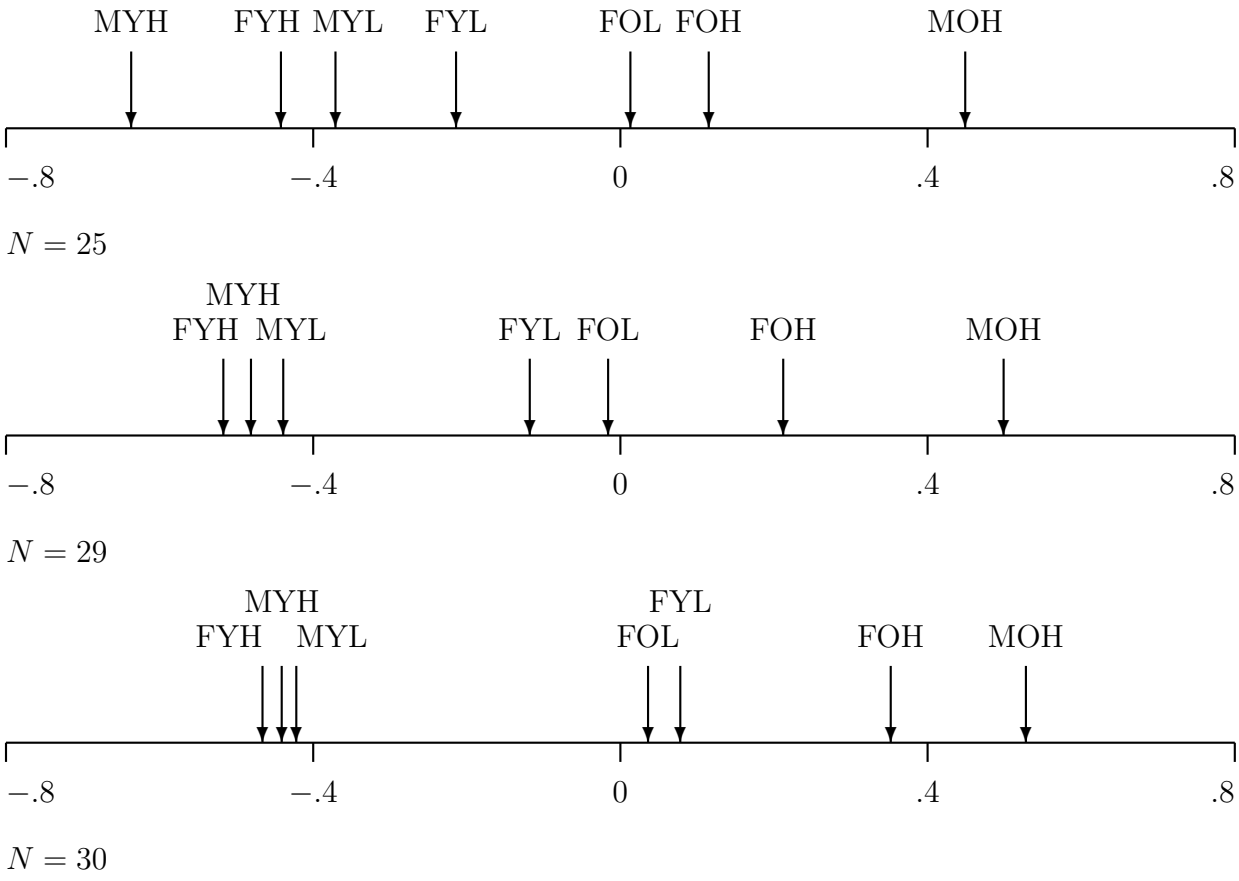



**Figure 6.12**: Estimated group scale values of the $(GT, P)$–111 log 3-mode models fit to the 25, 29 and 30 child samples.

was chosen as the "best" model in all cases, but since the 111 model provides a good fit, the scale values for this model were presented and used to describe the associations in the data.

Including the data from the children who had missing observations yields scale values for the group classifications that are more clustered relative to those obtained when only the data from the 25 children present on all occasions are analyzed. Age appears to play a primarily role in the group scale values, while sibling adjustment plays a secondary role. There do not appear to be any systematic differences between boys and girls. The groups with younger children tend to have negative scale values, while the groups with older children tend to have positive values. The major exception is the MOL classification, which contains the data from only one child. It is likely that the data for MOL are not very representative of older male children with low sibling adjustment. The groups with the largest positive scale values are the older children who show high sibling acceptance (i.e., FOH and MOH). In all of the analyses FOL is assigned a scale value close to 0.

The play qualities are fairly consistently ordered from "bad" to "good". Among the "bad" qualities, negative emotion is consistently assigned the largest negative scale value, followed by fight and prohibit, which are also assigned relatively large negative values. While there is less distinction among the "good" qualities, positive play is consistently assigned the largest positive scale value, followed by the cluster of qualities consisting of excite, fantasy, sustain, and gossip.

The occasions can be split into two groups: those before the birth of the children's siblings and those after. The occasions before the birth are consistently assigned scale

values that are larger than those assigned to the occasions after the births. Furthermore, the scale values assigned to occasions within these two classes tend to be more similar than those between classes. For example, the scale values for $-3$ and $-1$ are closer than the scale values for $-1$ and $+1$.

The interpretation of the associations in the peer play data is basically the same regardless of whether a sub-sample or all of the data are analyzed. The older children, in particular, those classified as FOH or MOH, exhibit more "good" play qualities and fewer "bad" ones, while the younger children, in particular, those classified as FYH, MYL or MYH, exhibit more immature qualities than mature ones. This relationship is the strongest at the first two sessions ($-3$ and $-1$) and is weaker at the three sessions after the birth of the children's siblings ($+1$, $+3$ and $+5$). This description of the relationship between groups, play qualities and time is based on the scale values from the $(GT, P)$-111 model. The description based on the 212 model for the 29 and 30 child samples is basically the same as that for the 25 child sub-sample. The description based on the 212 model includes a description of the differences between the groups across time with respect to the play qualities observed, especially for the 3 occasions after the birth and for the groups MOH and FYL.

## 6.4 Conclusion

Incorporating Tucker's 3–mode model into loglinear models yields models with many desirable properties and characteristics. The models have many of the properties associated

with multiplicative (loglinear) models. For example, they define and measure interactions in terms of odds and odds ratios, they always yield non-negative fitted values, and they are sub-table invariant. One of the criticisms of loglinear models for multiway tables is the proliferation of parameters and the difficulty of interpreting them (e.g., van der Heijden, 1985; van der Heijden & de Leeuw, 1985, 1987). This criticism is not necessarily applicable to log 3-mode models. Since combinations of interaction effects can be modeled by the 3–way decomposition, log 3-mode models tend to have fewer parameters than loglinear models, and the parameters have a clear and meaningful interpretation.

Log 3-mode models are designed to represent both the interesting and uninteresting systematic structure in data such that the residuals consist only of random, unsystematic deviations from the model. This modeling approach leads to natural ways of assessing the fit of models, including analyses of residuals to ensure that models are indeed adequately fitting the data.

With log 3-mode models, 3–way associations among categorical variables are analyzed by 3–factor decompositions, rather than by 2–factor decompositions as in the reduction approach. The addition of Tucker's 3-mode model is especially useful for analyzing longitudinal data, as was demonstrated on the peer play data.

The log 3-mode modeling of the peer play data yielded descriptions of the relationship between the group classifications, play qualities and occasions that are consistent with those from the analyses of the data in the previous chapters. The basic interpretation is the same whether or not the data from children with missing observations are included. The reasons for the similarities among the various analyses performed in this thesis is

discussed in the next chapter where further theoretical and empirical comparisons are presented.

The structural aspects (mathematical form) of the log 3-mode models is suitable for longitudinal categorical data, but the statistical model will not, in general, be adequate or even approximately adequate. While the residual analyses of the $(GT, P)$-212 log 3-mode model fit to the peer play data indicated that the assumption of independent Poisson random variables is a tenable approximation to the truth, additional model development is needed to explicitly incorporate serial and/or re-sampling dependence and the possibility of heterogeneity among subjects classified into the same groups. Explicitly incorporating dependence and heterogeneity would allow statistical tests for the presence of these effects. Besides statistical tests for dependence, heterogeneity and lack-of-fit of models, valid estimates of standard errors of model parameters could be computed.

# Chapter 7

# DISCUSSION AND CONCLUSION

The major purpose of this study was to develop models for multivariate, longitudinal categorical data that yield graphical representations of interactions in the data. To this end, many existing and new ways of analyzing categorical data were presented and discussed. The existing techniques considered were loglinear models, correspondence analysis, $RC(M)$ correlation models, $RC(M)$ association models, and various extensions and generalizations of them. The primary new development in this thesis is the log 3-mode models proposed in Chapter 6; however, other new ways of analyzing multivariate, longitudinal categorical data are also contained in this thesis. These additional ways consist of the following novel applications of existing methods and models:

1. The analysis longitudinal categorical data as described by van der Heijden (1987) and van der Heijden and de Leeuw (1989), except that the $RC(M)$ association model is used instead of correspondence analysis to analyze marginal and joint tables;

2. The use of conditional $RC(M)$ association models with homogeneity constraints on model parameters as proposed and developed by Becker and Clogg (1989a, 1989b), except that occasions rather than groups is the conditioning variable;

3. In the absence of a 3–way association, the use of bivariate interaction, log-bilinear models (including models where $u$ terms are included to ensure that particular 2-way margins are fit perfectly) to analyze longitudinal data; and

4. The use of *TUCKALS3* to decompose sums of particular interaction terms from saturated loglinear models fit to 3–way tables (as described in Chapter 6), rather than standardized residuals from unsaturated loglinear models.

Most of the existing models and the new proposals were illustrated by analyzing the peer play data by these methods. In Section 7.1, the various models fit to the data are summarized and their relative merits are evaluated. There is a remarkable similarity among the results from various models and the *TUCKALS3* analyses. In Section 7.2, the theoretical and empirical relationships among the models and analyses are examined. In Section 7.3, a general strategy for analyzing multivariate, categorical data is outlined. In the final section, Section 7.4, areas for future work are noted.

## 7.1 Summary of the Models Fit to the Data

Many models were fit to the peer play data in this thesis. In this section, these models are summarized, and the different type of models are evaluated against each other in terms of how well they fit the data. Since most of the analyses reported in the preceding

chapters used the data from the 25 child sub-sample and the pattern of results for the 25 child sub-sample is very similar to the pattern for the 29 and 30 child samples, the results discussed here are from the analyses of the data from the 25 child sub-sample.

In Figure 7.1, the likelihood ratio statistics from the models fit to the data from the 25 child sub-sample are plotted against degrees of freedom. Each point in the plot corresponds to a different model, most of which are labeled. Different symbols are used for different types of models: □ for loglinear models; ■ for bivariate interaction, log-bilinear models where the $GT$ margin is fit perfectly (i.e., equation 5.5); ○ for conditional and joint $RC(M)$ association models; and ● for log 3–mode models. The dotted line, labeled $\chi^2_{.95}$, corresponds to chi-squared values where the probability of observing a value as large as or larger than the chi-squared values is .05. The solid line, labeled $G^2 = df$, is the identity line and corresponds to chi-squared values where the probability of observing values as large as or larger than it is .5 (i.e., $\chi^2_{.5}$). The dotted and solid lines are used as (somewhat arbitrary) weak and strong criteria for evaluating the goodness-of-fit of the models.

Regardless of whether the weak or strong criterion is used, the loglinear models do not fit well; all of the □'s in Figure 7.1 lie above the dotted line. The models that come closest to fitting the data are the $(GT, GP)$ and $(GT, GP, PT)$ loglinear models, but these models are not adequate. The parameters of these models are hard to interpret, and they do not readily yield graphical representations of the interactions in the data. What the loglinear models do provide is an indication of the complexity of the structure of the interactions in the data. Specifically, there is no $PT$ partial (or marginal) association,

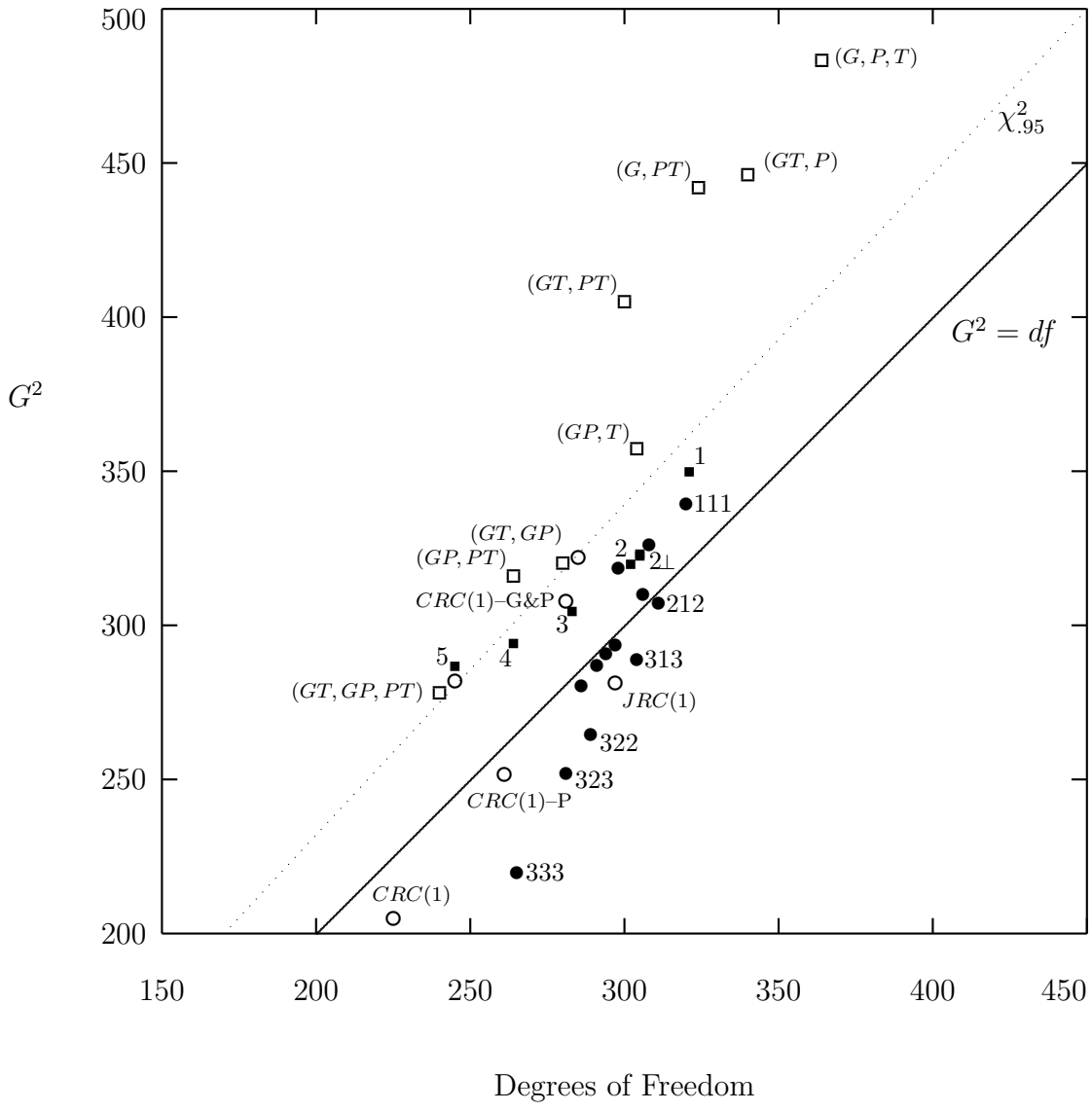



**Figure 7.1**: $G^2$ versus $df$ from models fit to the data from the 25 child sub-sample.

the most important (or strongest) interaction is the $GP$ partial association, and there is a slight 3–way association.

A large improvement in fit relative to the loglinear models is obtained when specific structures are assumed for the interactions. Even though none of the bivariate interaction, log-bilinear models represents the 3–way interaction, all of them, except the 5 dimensional model, fit better than the loglinear models. In the figure, the numbers next to the ■'s indicate the number of components per mode for that bivariate, log bilinear model. For $M = 2$, models with both orthogonal and non-orthogonal dimensions were fit, and these are labeled "2" and "2⊥", respectively. The components of the models with $M > 3$ were not restricted to be orthogonal. Recall that the $(GT, P)$ loglinear model is the bivariate, log bilinear model with $M = 0$ dimensions. When just one dimension is added to the $(GT, P)$ loglinear model, there is a large drop in $G^2$ relative to the number of additional parameters added to the model.

Even though all of the bivariate interaction, log-bilinear models, except the 5 dimensional model, have $G^2$'s that are less than the weaker criterion, $\chi^2_{.95}$, these models are of limited usefulness. Regardless of the number of dimensions estimated, a bivariate, log-bilinear model will never fit better than the $(GT, GP, PT)$ loglinear model. The $G^2$ of bivariate, log-bilinear models will always be larger than the $G^2$ of the $(GT, GP, PT)$ loglinear model. More importantly, the log-bilinear models do not represent the 3–way interaction.

The joint $RC(1)$ association model and four of the five conditional $RC(1)$ association models represent the 3–way interaction. The only conditional model that does not

represent the 3–way interaction is the one with homogeneity constraints on the intrinsic association parameters and the scale values for both groups and play qualities (in Figure 7.1, the unlabeled ○ to the right of the □ labeled "$(GT, GP)$"). This model is an extension of the $(GT, GP, PT)$ loglinear model where the $GP$ partial association is represented by a bilinear term.

The joint and three of the conditional models that do represent the 3–way interaction fit better than the models that do not represent the 3–way associations (i.e., the loglinear models, the bivariate, log-bilinear models, and the conditional $RC(1)$ association model with homogeneity constraints on all parameters of the bilinear term). The fit of the conditional $RC(1)$ association model with homogeneity constraints on the scale values for both groups and play qualities (in Figure 7.1, the ○ labeled "$CRC(1)$–G&P") meets the weaker criterion and is approximately equal the fit of the bivariate, log-bilinear model with $M = 3$. The joint $RC(1)$ association model ("$JRC(1)$"), the conditional $RC(1)$ association model with no homogeneity constraints ("$CRC(1)$"), and the conditional model with homogeneity constraints on the scale values for play qualities ("$CRC(1)$–P") are three of the better fitting models out of all of the models fit to the peer play data. Each of these three models meets the stronger criterion.

The $CRC(1)$ model with no homogeneity constraints imposed on the parameters has the smallest $G^2$ of all of the models fit to the peer play data; however, it is also the least parsimonious model. This model is essentially the same as fitting a separate $RC(1)$ association model to each of the five $\mathbf{F}_{GP|k}$ subtables. Of the three conditional models where the scale values for groups are forced to be the same across occasions, none of them

fit the data very well. This implies that there is a qualitative difference in the behavior of children within groups across time and that more than one dimension is needed to represent the groups. The conditional model with homogeneity constraints on the scale values for play qualities (i.e., the $CRC(1)$–P model) is a much more parsimonious model than the $CRC(1)$ model without any constraints. Since the $CRC(1)$–P model also fits the data well, only one set of scales values for play qualities is necessary.

The $JRC(1)$ model, which also fits the data well, is an even more parsimonious model than the $CRC(1)$–P model. This adds additional support to the speculation that only one dimension is needed for play qualities, and it further implies that the 2–way interaction terms $\{u_{jk}^{PT}\}$ are not essential.

The conditional and joint models provide good "benchmarks" against which to evaluate the $(GT, P)$ log 3–mode models. They can be used to assess whether there is an advantage of using a 3–way decomposition instead of a 2–way decomposition. More specifically, the differences between the log 3–mode models and the joint $RC(1)$ association model indicate the advantage of explicitly treating the 3 modes as separate entities and assigning scale values to the categories of each of them (versus assigning scale values to combinations of groups and occasions). The differences between the log 3–mode models and the conditional models indicate the advantage of only estimating one set of 2–way interaction parameters (i.e., $\{u_{ik}^{GT}\}$) and components for each the 3 modes versus estimating two sets of 2–way interaction parameters (i.e., $\{u_{ik}^{GT}\}$ and $\{u_{jk}^{PT}\}$), one component play qualities, and one component for groups (or combinations of groups and occasions).

When both $G^2$ and degrees of freedom are taken into consideration, some of the log 3-mode models are better than the joint and conditional $RC(1)$ association models. The most interesting log 3-mode models are the ones that are more parsimonious than the joint and conditional models. These log 3-mode models are also the ones that are empirically and theoretically very similar to the joint and conditional $RC(1)$ association models. For example, the $CRC(1)$–P model, the $JRC(1)$ model and the $(GT, P)$–212 log 3-mode model (in Figure 7.1, the ● labeled "212") all have similar substantive interpretations of the data. The $CRC(1)$–P and $JRC(1)$ models have smaller $G^2$'s than the $(GT, P)$–212 model, but the 212 model has fewer parameters and more degrees of freedom. Since all of these models have acceptable and comparable $G^2$'s relative to their degrees of freedom, the "best" model is the $(GT, P)$–212 log 3-mode model. This model is the simplest one. Thus, using the stronger criterion (the dark line), the "best" model is the $(GT, P)$–212 log 3-mode model.

As another example, the $CRC(1)$-G&P model and the $(GT, P)$–111 log 3-mode model have similar substantive interpretations. Of these two models, the conditional model has a smaller $G^2$, but the log 3-mode model has fewer parameters and more degrees of freedom. Both models have acceptable fits with respect to the weaker criterion, but since the 111 model is the simpler, it is the better model. Thus, even using the weaker criterion (the dotted line), a log 3-mode model is still the "best" model. The simplest, most parsimonious models among all of the models that meet the strong and weak criterion are the $(GT, P)$–212 and 111 models, respectively.

## 7.2 Theoretical and Empirical Comparisons

Regardless of the specific sample (i.e., $N = 25$, 29 or 30) or the specific model, the same basic description of the data was obtained from the models that fit the data well. In addition to the models, the corresponding *TUCKALS3* decompositions of the standardized residuals from the $(GT, P)$ loglinear model also gave this same basic interpretation. The similarity of the representations from the different models and analyses results from theoretical similarities between the models in conjunction with the relatively simple the structure of the associations in the peer play data.

In this section, theoretical and empirical comparisons are made between the conditional $RC(1)$ association models, the joint $RC(1)$ association model, *TUCKALS3* analyses, and the $(GT, P)$–111 and 212 log 3-mode models. In Section 7.2.1, the models and analyses analogous to the $(GT, P)$–111 log 3-mode model are examined, and in Section 7.2.2, those analogous to the $(GT, P)$–212 model are examined. While the models discussed in Sections 7.2.1 and 7.2.2 are theoretically and empirically similar, there is a fundamental difference between them. The joint and conditional models involve 2–way decompositions, while the log 3-mode models involve 3–way decompositions. This inherent distinction is discussed in Section 7.2.3.

### 7.2.1 One Dimensional Models

The $(GT, P)$–111 log 3-mode model is empirically very similar to the $CRC(1)$–G&P model and the 111 *TUCKALS3* analysis of the standardized residuals from the $(GT, P)$

loglinear model. In each of these analyses, one set of numbers (i.e., one dimension or component) is estimated for each of the three modes. Given the known structure in the data, comparing the mathematical form of the models underlying these three analyses reveals why the results are so similar.

The $(GT, P)$–111 log 3-mode model is

$$F_{ijk} = \exp(u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \phi\mu_i\nu_j\eta_k) \tag{7.1}$$

or

$$\ln(F_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \phi\mu_i\nu_j\eta_k \tag{7.2}$$

where $\sum_i \mu_i^2 = \sum_j \nu_j^2 = \sum_k \eta_k^2 = 1$, and only the scale values for play qualities are centered (i.e., $\sum_j \nu_j = 0$). The $CRC(1)$–G&P can be expressed in a form comparable to equation 7.2 as follows:

$$\begin{aligned} \ln(F_{ijk}) &= u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + u_{jk}^{PT} + \phi_{(k)}\mu_i\nu_j \\ &= u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + u_{jk}^{PT} + \phi^*\mu_i\nu_j\eta_k^* \end{aligned} \tag{7.3}$$

where $\sum_i \mu_i^2 = \sum_j \nu_j^2 = 1$, $\sum_i \mu_i = \sum_j \nu_j = 0$, $\phi^* = \sqrt{\sum_k \phi_{(k)}^2}$, and $\eta_k^* = \phi_{(k)}/\phi^*$, such that $\sum_k (\eta_k^*)^2 = 1$. The major difference between equations 7.2 and 7.3 is that equation 7.3 includes $u_{jk}^{PT}$, but equation 7.2 does not. Equation 7.2 is not a special case of equation 7.3, because slightly different constraints on the scale values are imposed to

identify the parameters. Only the scale values for play qualities are centered in the log 3-mode model, but the scale values for both the groups and play qualities are centered in the conditional $RC(1)$ association model. When homogeneity constraints are imposed on the parameters of the conditional $RC(M)$ association model, the "identification constraints" on the scale values are actually not completely arbitrary, but they are actually restrictive and affect the fit of the model.

Empirically, the $(GT, P)$–111 and the $CRC(1)$–G&P models yield similar results, because the two differences between the models have a negligible effect due to the structure of the associations in the data. First, the $PT$ partial association is not important, so $\hat{u}_{jk}^{PT} \approx 0$. Second, the estimated scale values for groups of the $(GT, P)$–111 model are both negative and positive and their sum is approximately equal to zero (i.e., $\sum_i \hat{\mu}_i \approx 0$). Thus, the difference in the constraints imposed on the scale values has a minimal effect.

The 111 *TUCKALS3* decomposition of the standardized residuals from the $(GT, P)$ loglinear model is

$$\frac{F_{ijk} - \exp(u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT})}{\left(\exp(u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT})\right)^{1/2}} = \tilde{g}\tilde{x}_i\tilde{y}_j\tilde{z}_k$$

For comparison to the log 3-mode mode (i.e, equation 7.1), the above equation can be re-written as

$$\begin{aligned} F_{ijk} \;=\; & \exp(u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT}) \left[1 + \left(\frac{\tilde{g}}{(\exp(u))^{1/2}}\right)\left(\frac{\tilde{x}_i}{(\exp(u_i^G))^{1/2}}\right)\right. \\ & \left.\left(\frac{\tilde{y}_j}{(\exp(u_j^P))^{1/2}}\right)\left(\frac{\tilde{z}_k}{(\exp(u_k^T))^{1/2}}\right)\left(\frac{1}{(\exp(u_{ik}^{GT}))^{1/2}}\right)\right] \end{aligned}$$

$$
\begin{aligned}
F_{ijk} &= \exp(u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT}) \left[1 + g^* x_i^* y_j^* z_k^* / w_{ik}\right] \\
F_{ijk} &= \exp(u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT}) \left[1 + g x_i y_j z_k / w_{ik}\right] \qquad (7.4)
\end{aligned}
$$

where $g^* = \tilde{g}/\sqrt{\exp(u)}$, $x_i^* = \tilde{x}_i/\sqrt{\exp(u_i^G)}$, $y_j^* = \tilde{y}_j/\sqrt{\exp(u_j^P)}$, $z_k^* = \tilde{z}_k/\sqrt{\exp(u_k^T)}$, $w_{ik} = \sqrt{\exp(u_{ik}^{GT})}$, $x_i$, $y_j$ and $z_k$ equal $x_i^*$, $y_j^*$ and $z_k^*$, respectively, re-scaled such that $\sum_i x_i^2 = \sum_j y_j^2 = \sum_k z_k^2 = 1$, (i.e., $x_i = x_i^*/(\sum_i (x_i^*)^2)^{1/2}$, $y_j = y_j^*/(\sum_j (y_j^*)^2)^{1/2}$ and $z_k = z_k^*/(\sum_k (z_k^*)^2)^{1/2}$), and $g = g^* \sqrt{\sum_i (x_i^*)^2} \sqrt{\sum_j (y_j^*)^2} \sqrt{\sum_k (z_k^*)^2}$. In the *TUCKALS3* analyses, the $u$-terms of the loglinear model were estimated first, and then the scale values and $g$ of the 3-way decomposition were computed.

When $g = \phi = \phi^* = 0$, equation 7.4 is equivalent to the $(GT, P)$–111 log 3-mode model (i.e., equation 7.1 ) and is approximately equivalent to the $CRC(1)$-G&P model (equation 7.3). If $u_{ik}^{GT}$ is "small" such that $w_{ik} \approx 1$ (i.e., $u_{ik}^{GT} \approx 0$) and $g x_i y_j z_k / w_{ik}$ is small relative to 1, then equation 7.4 will approximate equation 7.1. Thus, the smaller the strength of the association in the data that is represented by the scale values of the models, the greater the similarity of results.

In the previous chapters, empirical comparisons of the scale values from the $CRC(1)$–G&P model (Figure 4.4), the 111 *TUCKALS3* analysis of the standardized residuals from the $(GT, P)$ loglinear model (Figure 5.5), and the $(GT, P)$–111 log 3-mode model (Figure 6.1) were made by comparing the plots of scale values for groups and play qualities. From these comparisons it was noted that these analyses are all very similar, but that the scale values from the $CRC(1)$–G&P model appears to be more similar to those from

**Table 7.1**: Correlations of scale values for play qualities (upper triangle) and groups (lower triangle).

| | $CRC(1)$–G&P | $(GT, P)$–111 | 111 *TUCKALS3* |
|---|---|---|---|
| $CRC(1)$–G&P | — | .996 | .976 |
| $(GT, P)$–111 | .995 | — | .974 |
| 111 *TUCKALS3* | .967 | .952 | — |

the $(GT, P)$–111 log 3-mode model than they are to the parameters from the 111 *TUCKALS3* decomposition. This subjective judgment is also supported by the correlations of the scale values from the three different analyses.

The Pearson product moment correlations of the scale values for groups and play qualities from the $CRC(1)$–G&P model, $(GT, P)$–111 log 3-mode model, and the 111 *TUCKALS3* decomposition are reported in the lower and upper triangles, respectively, of Table 7.1. All of the correlations in Table 7.1 are quite large, but the largest ones (i.e., .996 and .995) are between the scale values from the $CRC(1)$-G&P model and $(GT, P)$-111 log 3-mode model. Thus, the relative order and spacing of the scale values for groups and play qualities from the conditional $RC(1)$ association model and the log 3-mode model are indeed more similar to each of other than they are to those from the *TUCKALS3* analysis.

While the group and play quality scale values from the conditional and log 3-mode models are nearly identical, the relative order and spacing of the scale values for occasions from the log 3-mode model are more similar to those from the *TUCKALS3* decomposition than they are to the estimated intrinsic association parameters for each occasion in the $CRC(1)$–G&P model, (i.e., $\phi_{(k)}$, for $k = 1, \ldots, 5$, or the $\eta_k^*$ in equation 7.3). The

**Table 7.2**: Correlations of association parameters and scale values for occasions

| | $CRC(1)$-G&P | $(GT, P)$–111 | 111 *TUCKALS3* |
|---|---|---|---|
| $CRC(1)$-G&P | — | | |
| $(GT, P)$–111 | .974 | — | |
| 111 *TUCKALS3* | .994 | .992 | — |

intrinsic association parameters of the $CRC(1)$–G&P model are analogous to the scale values assigned to occasions in the 111 *TUCKALS3* analysis and the $(GT, P)$-111 log 3-mode model. The correlations between the intrinsic association parameters and the scale values for occasions are reported in Table 7.2. All of the correlations are quite large. The smallest one (i.e., .974) is between the log 3-mode and the $CRC(1)$–G&P models.

The large correlations between the scale values for groups, play qualities and occasions from the $CRC(1)$–G&P model, the 111 *TUCKALS3* analysis, and the $(GT, P)$–111 log 3-mode model confirm the similarity observed in the plots of the scale values from these analyses. These three analyses are representing the same aspects of the data and yield nearly identical representations and substantive interpretations of the interactions in the data. The large correlations suggest that the models underlying each of these analyses are nearly equivalent, which in turn suggests that $u_{jk}^{PT} \approx 0$ and that the overall strength of the association represented is relatively small.

### 7.2.2 Multiple Dimensions for Groups and Occasions

The $JRC(1)$ model, $CRC(1)$–P model, the 212 *TUCKALS3* analysis, and the $(GT, P)$-212 log 3-mode model all have the same basic substantive interpretation of the data. After the theoretical similarities among these models are noted, the empirical similarities are examined.

In these four analyses, one dimension or component is estimated for the play qualities, but multiple components are estimated for the groups and occasions. The $(GT, P)$–212 log 3-mode model is

$$\ln(F_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \sum_{r=1}^{2}\sum_{t=1}^{2} \phi_{r1t}\mu_{ir}\nu_{j1}\eta_{kt}$$

where $\sum_i \mu_{ir}\mu_{ir'} = \delta_{rr'}$, $\sum_k \eta_{kt}\eta_{kt'} = \delta_{tt'}$, and $\sum_j \nu_{j1} = 0$. Since the scale values can be arbitrarily transformed such that $\phi_{r1t} = 0$ for $r \neq t$, the $(GT, P)$–212 model reduces to

$$\begin{aligned} \ln(F_{ijk}) &= u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \sum_{r=1}^{2} \phi_{r1r}\mu_{ir}\nu_{j1}\eta_{kr} \qquad (7.5) \\ &= u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \phi_{ik}^{**}\nu_{j1} \\ &= u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \phi^*\mu_{ik}^*\nu_{j1} \qquad (7.6) \end{aligned}$$

where $\phi_{ik}^{**} = \sum_r \phi_{r1r}\mu_{ir}\eta_{kr}$, $\phi^* = \sqrt{\sum_{i,k}(\phi_{ik}^{**})^2}$, and $\mu_{ik}^* = \phi_{ik}^{**}/\phi^*$, such that $\sum_{i,k}(\mu_{ik}^*)^2 = 1$.

The reduced form of the log 3-mode model, equation 7.6, is nearly identical to the $JRC(1)$ model, which is

$$\ln(F_{ijk}) = u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + \phi\mu_{(ik)}^{GT}\nu_j \tag{7.7}$$

where $\sum_{i,k}\mu_{(ik)}^{GT} = \sum_j \nu_j = 0$, and $\sum_{i,k}(\mu_{(ik)}^{GT})^2 = \sum_j \nu_j^2 = 1$. The major difference between the log 3-mode model and the joint model is that the group-occasion scores of the log 3-mode model, $\{\mu_{ik}^*\}$ (or $\{\phi_{ik}^{**}\}$), are a function of two group and two occasion components, while no specific structure is assumed by the joint model for the group-occasion scale values, $\{\mu_{(ik)}^{GT}\}$. An additional distinction, which stems from this difference, is that different constraints are needed to identify the scale values; namely, in equation 7.7 the $\{\mu_{ik}^{GT}\}$ are centered, but neither the analogous values in equation 7.6, $\{\mu_{ik}^*\}$, nor the component values, $\{\mu_{ir}\}$ and $\{\eta_{kr}\}$, are centered.

The conditional $RC(1)$ association model with homogeneity constraints on the scale values for play qualities is also quite similar to the equations 7.6 and 7.7. The $CRC(1)$–P model can be written as

$$\begin{aligned} \ln(F_{ijk}) &= u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + u_{jk}^{PT} + \phi_{(k)}\mu_{i(k)}\nu_j & (7.8)\\ &= u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT} + u_{jk}^{PT} + \phi^\star\mu_{ik}^\star\nu_j & (7.9) \end{aligned}$$

where $\sum_i \mu_{i(k)} = \sum_j \nu_j = 0$, $\sum_i \mu_{i(k)}^2 = \sum_j \nu_j^2 = 1$, $\phi^\star = \sqrt{\sum_{i,k}(\phi_{(k)}\mu_{i(k)})^2}$, and $\mu_{ik}^\star = \phi_{(k)}\mu_{i(k)}/\phi^\star$, such that $\sum_{i,k}(\mu_{ik}^\star)^2 = 1$ and $\sum_i \mu_{ik}^\star = 0$. The $u_{jk}^{PT}$ term in equation 7.9 is the primary difference between the conditional model and other two models. The conditional and joint models have slightly different identification constraints on the group-occasion scale values, both of which are different from those of the log 3-mode model. For the $CRC(1)$–P model, the $\{\mu_{i(k)}\}$ (and $\{\mu_{ik}^\star\}$) are centered over $i$ for each $k$, while for the $JRC(1)$ model, the $\{\mu_{ik}^{GT}\}$ are centered over both $i$ and $k$.

The mathematical model underlying the 212 *TUCKALS3* decomposition of the standardized residuals from the $(GT, P)$ loglinear model is

$$F_{ijk} = \exp(u + u_i^G + u_j^P + u_k^T + u_{ik}^{GT})(1 + \sum_{r=1}^{2} g_{r1r} x_{ir} y_{j1} z_{kr}/w_{ik}) \tag{7.10}$$

where $w_{ik} = \sqrt{\exp(u_{ik}^{GT})}$. Equation 7.10 is more similar to the log 3-mode model (equation 7.5) than it is to either the joint (equation 7.7) or the conditional (equation 7.9) $RC(1)$ association models. The 3-way decomposition in equation 7.10 has the same form as the one in the log 3-mode models (see equation 7.5). Equation 7.10 is different from the joint, conditional and log 3-mode models in that $F_{ijk}$ is an additive function of the fitted values from the $(GT, P)$ loglinear model and the combined effects the $GP$, $PT$ and $GPT$ interactions, while the other models are a log-additive function of these effects.

The $JRC(1)$ and $CRC(1)$–P models assign scale values to groups for each occasion, (i.e., $\mu_{(ik)}^{GT}$ and $\mu_{ik}^\star$, respectively) which are analogous to the group-occasion scores from the *TUCKALS3* analysis (i.e., $g_{i1k}^{**} = \sum_r \sum_t g_{r1t} x_{ir} z_{kt}$) and the log 3-mode model (i.e.,

**Table 7.3**: Correlations of group-occasion scores/scale values at each occasion (lower triangle) and play scale values (upper triangle).

| | $JRC(1)$ | $CRC(1)$–P | $(GT, P)$–212 | 212 *TUCKALS3* |
|---|---|---|---|---|
| $JRC(1)$ | — | .997 | .996 | .973 |
| $CRC(1)$–P | .972 | — | .992 | .972 |
| $(GT, P)$–212 | .898 | .858 | — | .973 |
| 212 *TUCKALS3* | .855 | .854 | .935 | — |

$\phi^{**}_{i1k} = \sum_r \sum_t \phi_{r1t}\mu_{ir}\eta_{kt}$, or equivalently, $\mu^*_{ik}$). The correlations of the group-occasion scores and scale values from these four models are reported in the lower triangle of Table 7.3, and the correlations of the scale values for play qualities are reported in the upper triangle.

The largest correlations in Table 7.3 are between the scale values assigned to the play qualities. Among these correlations, the three largest correlations are between the joint and conditional models (i.e., .997), the joint $RC(1)$ association and the $(GT, P)$-212 models (i.e., .996), and the conditional $RC(1)$ association and the $(GT, P)$–212 models (i.e., .992).

Among the correlations of the group-occasion scores and scale values, the largest one is between the scale values from the joint and conditional $RC(1)$ association models (i.e., .972), and the second largest is between the group-occasion scores from the $(GT, P)$-212 model and the *TUCKALS3* decomposition (i.e., .935). The correlations between the group-occasion scores from the $(GT, P)$-212 model and the scale values from the joint and conditional model (i.e., .898 and .858, respectively) are slightly larger than those from the *TUCKALS3* decomposition (i.e., .855 and .854, respectively).

While the scale values from all four models are quite similar (i.e., all of the correlations are relatively large), the pattern of correlations in Table 7.3 indicates that the joint and conditional $RC(1)$ association models are nearly identical and both are more similar to the $(GT, P)$-212 log 3-mode model than they are to the 212 *TUCKALS3* decomposition. These empirical observations confirm the expectations based on the theoretical similarities between the models.

Overall, the 212 *TUCKALS3* decomposition is empirically and theoretically more similar to the log 3-mode model than it is to either the joint or conditional model. The play quality scale values from the 212 *TUCKALS3* decomposition are equally similar to those from the other three models (i.e., correlations of .972–.973), but the group-occasion scores are noticeably more similar to those from the $(GT, P)$–212 log 3-mode model than they are to the scale values from the joint and conditional models (i.e., .935 versus .855 and .854). Unlike the conditional and joint models, the *TUCKALS3* decomposition also approximates the structure of the group-occasion scores of the log 3-mode model; that is, it approximates the two components for the groups and the two components for the occasions (i.e., $x_{ir} \approx \hat{\mu}_{ir}$ for $r = 1, 2$, and $z_{kt} \approx \hat{\eta}_{kt}$ for $t = 1, 2$).

### 7.2.3 2-Way Versus 3-Way Decompositions

Given the theoretical and empirical similarities among the various models fit to the peer play data, the $(GT, P)$–111 log 3-mode model can be thought of as the $CRC(1)$–G&P model where a simpler structure is imposed on the group-occasion scale values, and the $(GT, P)$–212 model can be thought of as the $JRC(1)$ or $CRC(1)$–P model where a simpler

**Table 7.4**: Correlations of group-occasion scores/scale values.

| | $JRC(1)$ | $CRC(1)$-P | $(GT, P)$-212 | *TUCKALS3* | $JRC(1)$-SVD2 |
|---|---|---|---|---|---|
| $JRC(1)$ | — | | | | |
| $CRC(1)$–P | .972 | — | | | |
| $(GT, P)$–212 | .898 | .858 | — | | |
| *TUCKALS3* | .855 | .854 | .935 | — | |
| $JRC(1)$–SVD2 | .901 | .906 | .956 | .917 | — |
| $CRC(1)$–SVD2 | .883 | .912 | .929 | .907 | .992 |

structure is imposed on the group-occasion scale values (Kroonenberg & de Leeuw, 1980; Kiers, 1991). From this viewpoint, it might appear that the rank two singular value decomposition (SVD) of the group-occasion scale values from either the joint and/or the conditional analyses is a reasonable way of approximating the scale values the log 3-mode model. For the peer play data, this procedure works quite well.

The first two right and left singular vectors of the $(7 \times 5)$ matrix with elements $\hat{\mu}^{GT}_{(ik)}$, which are the estimated group-occasion scale values from the joint $RC(1)$ association model, and the first two singular vectors of the matrix with elements $\hat{\mu}_{i(k)}\hat{\phi}_{(k)}$, which are from the $CRC(1)$–P model, are good approximations of the group and occasion scale values from the $(GT, P)$-212 log 3-mode model. The singular vectors are also quite similar to the scale values from the from 212 *TUCKALS3* analysis.

The rank 2 SVD approximations of the group-occasion scale values are also quite similar to the group-occasion scores from the log 3-mode model and the *TUCKALS3* analysis (i.e., $\hat{\phi}^{**}_{i1k}$ (or $\hat{\mu}^{*}_{ik}$) and $g^{**}_{i1k}$). The correlations between the rank 2 SVD approximations of the group-occasion scale values from the $JRC(1)$ and $CRC(1)$–P models, and the group-occasion scale values and scores from the various models and analyses are

given in Table 7.4. The rank 2 approximations of the group-occasion scale values from the joint model ("$JRC(1)$–SVD2") and the conditional model ("$CRC(1)$–SVD2") are more similar to the group-occasion scores from the $(GT, P)$–212 log 3-mode model than they are to the group-occasion scale values from the joint and conditional models, which were the values decomposed (i.e., for the joint model, $.956 > .901$, and for the conditional model, $.929 > .912$). This supports the idea that the log 3-mode model is basically just a joint (or conditional) $RC(1)$ association model where a more restrictive structure has been imposed on the scale values for combinations of groups and occasions.

The rank 2 approximations from the joint and conditional models are also similar to the group-occasion scores from the 212 *TUCKALS3* decomposition (i.e., .917 and .907), but the rank 2 approximations are more similar to the group-occasions scores from the log 3-mode model than they are to the scores from the *TUCKALS3* decomposition. This result is expected, because the joint and conditional models are theoretically more similar to the log 3-mode model than they are to the model underlying the *TUCKALS3* analysis.

While computing two successive, 2-way decompositions yields an approximation of the 3-way decomposition for the peer play data, such a procedure will not in general work. This procedure worked here mainly because the structure in the peer play data is relatively simple, in particular, only one dimension is required for the play qualities. However, even with a relatively simple structure, there is no guarantee that two successive, 2-way decompositions will lead to the same or even approximate decomposition of a 3–way array as when the array is decomposed into three ways simultaneously. Simply

changing the order that the 2–way decompositions are performed can lead to different results (Franc, 1989).

Even in cases where the SVD of scale values of two interactively coded variables from a joint and/or conditional $RC(M)$ association models approximates the scale values of a log 3-mode model, the 3-mode model is the better model. The statistics $G^2$ and $X^2$ are natural indices to assess the fit of log 3-mode models to the data, but no such natural measures exist for an $RC(M)$ association model followed by a SVD. The fit of the $RC(M)$ association model can be measured by $G^2$ and/or $X^2$, but this does not take into consideration that the scale values are then approximated by a lower rank SVD. Usually with lower rank SVD approximations, some sort of "percent of the fit" measure is used to assess how well the values decomposed are approximated; however, such measures do not take into consideration the number of parameters estimated and the degrees of freedom.

Another problem with $RC(M)$ association models followed by a secondary SVD is that there are unnecessary constraints imposed on the scale values. In cases where a joint or conditional model approximates a log 3-mode model, not all of the scale values of the log 3-mode model need to be centered, but the scale values of $RC(M)$ association models have to be centered to identify them.

The most serious problem with using $RC(M)$ association models followed by a secondary SVD is that when more than 1 component is necessary for each mode of a table, successive 2-way decompositions of 3–way arrays will not even lead to decompositions that have the same form as Tucker's 3-mode model. For example, suppose that the

$(GT, P)$-222 model is the "best" log 3-mode model, which is

$$\ln(F_{ijk}) = u + u_i^G + u_k^P + u_k^T + u_{ik}^{GT} + \sum_{r=1}^{2}\sum_{s=1}^{2}\sum_{t=1}^{2}\phi_{rst}\mu_{ir}\nu_{js}\eta_{kt} \tag{7.11}$$

and consider the joint $RC(M)$ association model fit to ${}_{TG}\mathbf{F}_P$, which has the same $u$-terms as equation 7.11. If two components are needed for play qualities, as in equation 7.11, then the joint $RC(1)$ association model would be too simple to fit the data and the 2 dimensional model,

$$\ln(F_{ijk}) = u + u_i^G + u_k^P + u_k^T + u_{ik}^{GT} + \sum_{m=1}^{2}\phi_m\mu_{(ik)m}^{GT}\nu_{jm} \tag{7.12}$$

would be needed. With the $RC(2)$ association model, there are two sets of group-occasion scale values (i.e., $\{\mu_{(ik)1}^{GT}\}$ and $\{\mu_{(ik)2}^{GT}\}$). Taking the rank 1, SVD approximation of each of them yields two components for groups and two for occasions; that is, $\mu_{(ik)m}^{GT[1]} \approx \lambda_{1(m)}v_{i1(m)}^G v_{k1(m)}^T$ where $\lambda_{1(m)}$ is the first (largest) singular value and $\{v_{i1(m)}^G\}$ and $\{v_{k1(m)}^T\}$ are the corresponding singular vectors from the (generalized) SVD of $\{\mu_{(ik)m}^{GT}\}$, for $m = 1, 2$. Substituting the SVD approximations of $\mu_{(ik)m}^{GT}$ into equation 7.12 the yields a diagonal 3-way decomposition,

$$\ln(F_{ijk}) = u + u_i^G + u_k^P + u_k^T + u_{ik}^{GT} + \sum_{m=1}^{2}\phi_m^{*[1]}v_{i1(m)}^G v_{k1(m)}^T\nu_{jm}$$

where $\phi_m^{*[1]} = \phi_m\lambda_{1(m)}$. This decomposition is more similar to CANDECOMP than it is to Tucker3.

Tucker's 3-mode model was chosen because of its potential usefulness in analyzing longitudinal data. The mathematical structure of Tucker's 3-mode decomposition permits the representation of change and stability in terms of underlying components for time or occasions. The other 2 modes of the data, groups (or individuals) and variables, are also represented by components, and the relationship among the 3-modes are represented by the elements of the core matrix. Complex interactions among 3-modes can be represented in a relatively simple form with Tucker's 3-mode model. When the number of underlying components is relatively small, the model provides a description and representation of the data that is simpler than the data itself. Furthermore, if the associations in the data are more complex than the ones in the peer play data, then attempting to analyze it with 2-way methods may lead to results that are more complex and harder to interpret than the results from a log 3-mode analysis. Three–way decompositions require 3–way methods.

## 7.3 Complementary Strategy for Analyzing Multivariate Data

The analyses performed in this study complemented each other. With the peer play data, the loglinear models, $RC(M)$ association models, and the *TUCKALS3* decompositions all contributed to our understanding of the structure in the data and gave indications of which log 3-mode models might best fit the data. In the end, the $(GT, P)$–111 and 212

log 3-mode models were chosen as the "best" models for this data set because they are the simplest models that adequately fit the data.

A general strategy for analyzing data sets other than the peer play data that are more or less complex than the play data is outlined here. This strategy uses the different methods discussed in this thesis and a hierarchy among the models. The hierarchical relations among the models is (in part) analogous to the hierarchy of 3–way methods described by Kiers (1989, 1991), except that the methods considered by Kiers do not pertain explicitly to categorical data. Unlike the 3–way methods considered by Kiers, the most complex models among those considered here are the loglinear models. In the hierarchy given by Kiers, the most complex model is a principal components analysis of a "supermatrix", which is analogous to a joint analysis where $RC(M)$ models are fit to the table in which the rows correspond to group-variable combinations and the columns correspond to occasions. The same basic eclectic strategy outlined by Kiers is suggested here; namely, the most complex model, which is also the least parsimonious one, is fit first, followed by successively more restrictive models until the most parsimonious model that provides an acceptable level of fit is identified.

Substantive considerations regarding how the data was collected and specific hypotheses and research questions are assumed to have already been taken into account. Considerations about how the data arose determine whether particular margins should be fit perfectly, and whether the standard sampling assumptions of statistical models are met. If the sampling assumptions are reasonable, then statistical hypothesis tests, such as goodness-of-fit tests, and estimates of standard errors of parameter estimates are

valid. Research questions and hypotheses determine which interactions are interesting and should be represented graphically (i.e., the ones to include in $u^{(1,2)}$).

Loglinear models are good for obtaining an initial assessment of the complexity of the associations in the data. Loglinear modeling will indicate whether there is a 3–way association. If there is no 3–way association, the simplest loglinear model that fits the data is the starting point for searching for even simpler models. The simpler models, which include $RC(M)$ association models and bivariate interaction, log-bilinear models, hypothesize particular structures for the associations. If there is a 3–way association, then extensions of saturated loglinear models are needed.

In the simplest case, complete independence, there is nothing in the data that warrants further study. If a joint or conditional independence loglinear model fits the data, then the partial associations in the data can be studied by fitting $RC(M)$ association models to the margins corresponding to the partial associations. If the "no 3–way interaction" loglinear model fits the data, then bivariate interaction, log-bilinear models can be fit to the data. In general, when there is no 3–way association, some sort of bivariate interaction, log-bilinear model can be found that fits the data. In the most parsimonious cases, the same scale values can be used for each of the 2–way associations, while in more complex cases, different scale values may be needed for each of them.

When there is a 3–way association, additional exploratory analyses of the data are recommended. The joint $RC(M)$ association model, the conditional $RC(M)$ association model (with and without homogeneity constraints), and/or *TUCKALS3* decompositions can be used to get a better understanding of the complexity of the associations. When

there are only one or two important but not interesting interactions (i.e., $u^{(1)}$ includes one or two terms) and when the structure of the higher-way association is relatively simple, the joint and conditional models are particularly useful. If $u^{(1)}$ includes all three 2–way effects terms such that $u^{(1,2)}$ only includes the 3–way association, then the scale values from the joint and conditional models will represent not only the 3–way association, which is of primary interest, but also the 2–way effects in $u^{(1)}$. When either the structure of the association in $u^{(1,2)}$ is relatively complex or $u^{(1,2)}$ only includes the 3–way association, *TUCKALS3* decompositions will be more useful than joint or conditional models.

*TUCKALS3* decompositions of estimates of $u_{ijk}^{(1,2)}$, which are computed using the estimated parameters from the saturated loglinear model, will yield fairly good approximations of log 3-mode models. With the peer play data, standardized residuals from the $(GT, P)$ loglinear model were used instead of the estimates of $u_{ijk}^{(1,2)}$ from the saturated loglinear model. For the peer play data, *TUCKALS3* decompositions of the estimated $u_{ijk}^{(1,2)}$'s were not useful, because some of the frequencies were very small (i.e., equal 0). For the small frequencies, the absolute value of the estimated $u_{ijk}^{(1,2)}$'s are very large relative to the estimates for the other cells (i.e., for $F_{ijk} = 0$, $u_{ijk}^{(1,2)} \to \infty$). The least squares estimates of the Tucker3 decomposition of $\{\hat{u}_{ijk}^{(1,2)}\}$ were dominated by the large $\hat{u}_{ijk}^{(1,2)}$'s. For small numbers of components, the sum of squared discrepancies is minimized by ensuring that the large $\hat{u}_{ijk}^{(1,2)}$'s are well approximated by the decomposition. To fit the large $\hat{u}_{ijk}^{(1,2)}$'s, non-zero scale values are assigned to the categories of the small frequencies, while relatively small scale values (close to zero) are assigned to the categories of the other frequencies.

When there are small frequencies (e.g., sampling zeros), better approximations of log 3-mode models can be obtained by *TUCKALS3* decompositions of the standardized residuals from an appropriate unsaturated loglinear model. How well a log 3-mode model will be approximated by a *TUCKALS3* decomposition depends both on the strength of the interactions that are represented by the decomposition and the effects included in $u^{(1)}$ and $u^{(1,2)}$. When the interactions in the data are relatively weak, *TUCKALS3* decompositions will approximate log 3-mode models better than when the associations are stronger. The best approximations of log 3-mode model are obtained when $u^{(1)}$ does not include any 2–way interaction effects, because the mathematical model underlying the *TUCKALS3* decomposition, which is

$$F_{ijk} = \exp(u + u_i^A + u_j^B + u_k^C)(1 + \sum_r \sum_s \sum_t \phi_{rst}\mu_{ir}\nu_{js}\eta_{kt}) \tag{7.13}$$

most resembles the corresponding log 3-mode model. The approximation will be the worst when $u^{(1)}$ includes all three of the 2–way interactions effects and $u^{(1,2)}$ includes only the 3–way interaction. In this case, the mathematical model underlying the *TUCKALS3* decomposition least resembles the corresponding log 3-mode model,

$$F_{ijk} = \exp(u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC})(1 + \sum_r \sum_s \sum_t \phi_{rst}\mu_{ir}\nu_{js}\eta_{kt}/w_{ij}w_{ik}w_{jk}) \tag{7.14}$$

where $w_{ij} = \sqrt{\exp(u_{ij}^{AB})}$, $w_{ik} = \sqrt{\exp(u_{ik}^{AC})}$, and $w_{jk} = \sqrt{\exp(u_{jk}^{BC})}$.

The results of exploratory analyses using *TUCKALS3* will narrow down the set of potential log 3-mode models, as well as provide starting values for the algorithm used to compute maximum likelihood estimates of the model parameters of the log 3-mode models. The results of the log 3-mode modeling, the *TUCKALS3* decompositions and any other analyses will be similar to the extent that the models are similar. If the results deviate substantially from each other, then either the models underlying the analyses are considerably different from each other, especially given the structure of the data, and/or mistakes were made in performing the analyses.

Once a few "good" log 3-mode models are identified, residuals analyses, such as those performed in Section 6.2 of Chapter 6, should be performed to determine whether any of the models are failing to model systematic structure in the data and to check the appropriateness of the sampling assumptions.

## 7.4 Future Work

By incorporating Tucker's 3-mode components model into loglinear models for 3–way tables, the resulting models represent change and stability over time in terms of a few number of components for the mode corresponding to occasions. The modes for groups and variables are also reduced to a smaller number of components than there are categories or levels of these modes. These structural aspects of log 3-mode models makes them ideal for analyzing longitudinal data. The various plots of component scores provide graphical representations of the relationship among the categories of a mode and the

relationship between different modes. With the peer play data, the log 3-mode models yielded very nice "pictures" of how the qualities of play exhibited by the children, who were classified into groups, changed over time and differed between groups.

While the log 3-mode modeling of the peer play data yields nice interpretations and representations of the data, there are two general areas that warrant future work and study: additional applications of the log 3-mode models as developed and presented in this thesis, and additional development of the models. The former is discussed in Section 7.4.1 and the latter is discussed in Section 7.4.2.

### 7.4.1 Additional Applications

To further assess the usefulness of the log 3-mode models, the methods discussed in this thesis should be applied to other multivariate 3-mode categorical data sets, especially those where the associations are more complex than those in the peer play data. Data sets that have been previously analyzed by methods such as correspondence analysis or other 2–way methods could be re-analyzing using log 3-mode models, and the results of the 2–way and 3–way analyses can be compared. Besides longitudinal data, log 3-mode models could also be used to analyze cross-sectional categorical data where one of the modes corresponds to time. With cross-sectional data, repeated observations are not made on the same individuals, and the independence assumption is likely to be valid. In such cases, no additional model development is need to estimate standard errors of parameter estimates, put confidence intervals on them, and compare $G^2$'s to chi-squared distributions to perform statistical tests of fit.

The models fit to the peer play data only fit one or none of the 2–way margins perfectly; however, log 3-mode models are flexible enough so that either two or all three of the 2–way margins can be fit perfectly. It would be interesting to compare the results of log 3-mode models where two or three margins are fit perfectly to the results obtained from 2–way reductions methods (i.e., marginal, joint and conditional $RC(M)$ association models). When two of the margins are fit perfectly, the results of conditional $RC(M)$ association models may be similar to log 3-mode models, but not to the results of marginal and joint analyses. When three 2–way margins are fit perfectly by a log 3-mode model, the representation of the interactions given by even a conditional $RC(M)$ association model could be quite different from that of the log 3-model model.

### 7.4.2 Additional Model Development

The least satisfactory aspect of the log 3-mode models for analyzing longitudinal data is the sampling assumption of independent observations. Violations of independence, which is to be expected for repeated measures data, primarily effects variance estimates and invalidates statistical test, but has minimal effects on the estimates of the model parameters. Based on residual analyses, the independence assumption appeared to be adequate for the peer play data; however, the violation of this assumption will likely be a problem for other longitudinal data sets, which limits the usefulness of log 3-mode models for analyzing such data. Additional model development is needed to deal with the non-independence of the observations. Besides the non-independence of observations, the possible heterogeneity among units (e.g., subjects, objects, children, etc.) that are

classified into the same group could also be a problem when log 3-mode models are fit to other data sets. This is another area for additional model development.

In the context of loglinear and Rasch models, methods have been proposed to deal with non-independence and heterogeneity, and these could also be used with log 3-mode models. One class of approaches, which are based on MLE, include marginal MLE (Cressie & Holland, 1983), conditional MLE (Conaway, 1992, 1988), "quasi" MLE (McCullagh & Nelder, 1989), and MLE with the additional assumption of a particular distribution for model parameters (Hausman, Hall & Griliches, 1984; Jansen & van Duijn, 1992). Marginal and conditional MLE are of limited usefulness for log 3-mode models, but quasi MLE and MLE with the additional assumption of a particular distribution for certain model parameters are more promising directions for further development. With quasi MLE, only the first two moments of the distribution are assumed to follow a Poisson distribution.

Generalized or weighted least squares estimation, GLS, (Grizzle, Starmer & Koch, 1969; Koch, Landis, Freeman, Freeman & Lehnen, 1977) can be used instead of MLE to deal with the situation where the assumption of independent Poisson variables is not valid. Inferences about parameters and the fit of the model to the data can be made with generalized least squares. Pannekoek (1992) uses GLS to fit the correlation model (i.e., correspondence analysis) to tables where observations are not independent (e.g., panel data, and composite or "stacked" tables) and describes how to adapt the method for $RC(M)$ association models. For the $RC(M)$ models, the trick is to obtain an estimate of the variance/covariance matrix of the relative frequencies. For log 3-mode modeling of

longitudinal data, a possible approach is to simultaneously model the cell frequencies and the pair-wise frequencies (e.g., the frequencies of group $i$ giving response $j$ on occasion $k$ and occasion $k+1$). This approach has been used for factor analysis of binary variables (Christoffersson, 1975; Muthèn & Christoffersson, 1981; Muthèn, 1978).

A potentially useful modification of the log 3-mode models given in Chapter 6 would be the addition of the option to impose linear restrictions on the scale values. This could be used to constrain some scale values to be equivalent. For example, with the peer play data, imposing equality restrictions on the scale values for some of the groups (e.g., MYL and FYL), play qualities (e.g., sustain and fantasy), and occasions (e.g., $-3$ and $-1$ months, and/or $+1$, $+3$ and $+5$ months) would indicated whether the difference between particular scale values are important or are due to chance.

Linear restrictions could also be used to incorporate additional or external information into the model. For example, with the peer play data, the group scale values could be restricted such that they are a linear function of "main" and "interaction" effects of the classification variables, age, gender and sibling acceptance. With respect to the components for occasions or time, orthogonal polynomials could be quite useful. With orthogonal polynomials, the components for occasions could be interpreted as change due to linear, quadratic, and so on trends.

A final area for future study is the use of alternative 3-way decompositions. One of the less desirable characteristics of Tucker's 3-mode model is the rotation indeterminacy of the components. Both Tucker's extended 3–mode model (Tucker2) and the canonical decomposition model (CANDECOMP/PARAFAC) were mentioned in this thesis,

but neither was explicitly incorporated into an extension of a loglinear model and neither has a problem with rotational indeterminacy. Another alternative model for the decomposition that could be used is a 3–way decomposition into directional components ("DEDICOM"), which combines the orthogonal components property of Tucker3 and the rotational determinacy (or "intrinsic axis") property of CANDECOMP (Harshman & Lundy, 1992).

# APPENDICES

## A Review of Correspondence Analysis

The different problems and approaches that lead to the technique of correspondence analysis are reviewed here (Greenacre, 1984; Tennenhaus & Young, 1985). The least squares estimation of the model is given by the *generalized* singular value decomposition of the matrix of probabilities or the centered matrix; that is,

$$(\mathbf{P} - \mathbf{E}) = \mathbf{X}\mathbf{D}_\lambda\mathbf{Y}'$$

where $\mathbf{P}$ and $\mathbf{E}$ are $(I \times J)$ matrices with elements $P_{ij}$ and $P_{i.}P_{.j}$, respectively, $\mathbf{D}_\lambda$ is a diagonal matrix with the singular values $\lambda_m$ on the diagonal, and the matrices $\mathbf{X}$ and $\mathbf{Y}$ contain the row and columm scale values, respectively, which are scaled such that $\mathbf{X}'\mathbf{D_r}\mathbf{X} = \mathbf{Y}'\mathbf{D_c}\mathbf{Y} = \mathbf{I}$, and $\mathbf{1}'\mathbf{D_r}\mathbf{X} = \mathbf{1}'\mathbf{D_c}\mathbf{Y} = \mathbf{0}$ (Greenacre, 1984, 1988). The matrices $\mathbf{D_r}$ and $\mathbf{D_c}$ are diagonal matrices with the row and column marginal probabilities, $P_{i.}$ and $P_{.j}$, respectively, on the diagonals. From the canonical correlation approach, the least squares estimation of the model is equivalently obtained from the ordinary singular

value decomposition of a matrix with elements $\{(P_{ij} - P_{i.}P_{.j})/(P_{i.}^{1/2}P_{.j}^{1/2})\}$; that is,

$$\mathbf{D_r}^{-1/2}(\mathbf{P} - \mathbf{E})\mathbf{D_c}^{-1/2} = \mathbf{U}\mathbf{D}_\lambda\mathbf{V}'$$

where $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}$ (Goodman, 1986; van der Heijden & de Leeuw, 1985; van der Heijden, de Falguerolles & de Leeuw, 1989; Escoufier, 1988, 1989). The row and column scale values in equation 3.8 are given by

$$\mathbf{X} = \mathbf{D_r}^{-1/2}\mathbf{U}$$

$$\mathbf{Y} = \mathbf{D_c}^{-1/2}\mathbf{V}$$

With respect to the canonical correlation analysis of continuous variables, the matrix $(\mathbf{P}-\mathbf{E})$ corresponds to a matrix of covariances between variables from two sets of variables (i.e., $\mathbf{\Sigma_{12}}$ or $\mathbf{S_{12}}$) and the matrices $\mathbf{D_r}$ and $\mathbf{D_c}$ correspond to the variance/covariance matrices of variables within each of the two sets, (i.e., $\mathbf{\Sigma_{11}}$ and $\mathbf{\Sigma_{22}}$ or $\mathbf{S_{11}}$ and $\mathbf{S_{22}}$).

From the dual or optimal scaling approach, CA is defined as a quantification technique that seeks to assign scale values to the rows and columns of a table such that the variance of the scale values assigned to the categories of each variable is maximized relative to the total variance, which is constrained to equal one (Greenacre, 1984). This maximization problem can be thought of as quantifying the categories of a variable by maximizing the "between group" variance relative to the total variance, where the "groups" are the category classifications of individuals or objects. Since the total variance is constrained to

equal one, the ratio of the "between group" and total variance is a squared "correlation ratio." Equivalently, the criterion can be stated as minimizing the "within group" variance relative to the total variance. Framing the problem in this way, CA can be interpreted as maximizing "internal consistency" within the category classifications (Greenacre, 1984).

The technique is called "dual" because the optimal scale values play two (dual) roles in the symmetric problems of finding optimal row scale values and optimal column scale values. The duality of the solution is easily seen in the reciprocal averaging approach. The reciprocal averaging approach seeks to assign scale values to the rows and columns such that the row scores are linear combinations of the column scores and vice versa (Greenacre, 1984). This approach is defined by the following "transition formulae" that relate row and column scale values:

$$\begin{aligned} x_{im} &= (1/\lambda_m) \sum_j (P_{ij}/P_{i.}) y_{jm} \\ y_{jm} &= (1/\lambda_m) \sum_i (P_{ij}/P_{.j}) x_{im} \end{aligned}$$

(Greenacre, 1984). As a scaling technique, usually only the scale values on the first dimension (i.e., $x_{i1}$ and $y_{j1}$), which have the largest possible correlation, are of interest.

As a descriptive technique, CA yields a lower dimensional matrix approximation of the original table. Normalized scale values are often plotted to provide a "picture" of the table. The geometric development and rational behind CA provides further insight into the interpretation of the plots. A few terms that are used in this development will now be defined. A "row profile" is a set of the conditional probabilities (or frequencies)

$P_{j|i} = P_{ij}/P_{i.}$, and an "average row profile" consists of the column marginals $P_{.j}$. Column profiles and average column profiles are similarly defined. A "chi–squared distance" between (for example) two row profiles is defined as

$$d^2(i, i') = \sum_j \frac{(P_{ij}/P_{i.} - P_{i'j}/P_{i'.})^2}{P_{.j}}$$

If two profiles are similar, then the chi–squared distance between them will be small. In CA, scale values are assigned to categories to reflect the chi–squared distances between row (column) profiles. In other words, similar profiles are assigned similar scale values. Therefore, in a plot of row scale values, proximity represents similarity of row profiles, and in plots of column scale values, proximity represents similarity of columns profiles. Alternatively, distances between row (column) points represent dissimilarity between the corresponding profiles.

In CA, the association between variables is reflected in the differences between row profiles (or between column profiles). If the row and column variables are independent, then all of the row profiles will equal the average row profile (i.e., the column marginals). Therefore, differences between the row profiles and the average row profile indicate the presence of an association between the row and column variables. To measure the association, the chi–squared distances between each row profile and the average row profile are computed using

$$d^2(i, .) = \sum_j \frac{(P_{ij}/P_{i.} - P_{.j})^2}{P_{.j}}$$

and the weighted sum of these distances is proportional to $X^2$,

$$\begin{aligned} X^2/N &= \sum_i d^2(i,.)P_{i.} \\ &= \sum_i \sum_j \frac{(P_{ij} - P_{i.}P_{.j})^2}{P_{i.}P_{.j}} \end{aligned}$$

which in CA, is called the total amount of "inertia" in a table. As previously discussed, $X^2$ is a measure of the total association in a table, and the *RC* correlation model provides a decomposition of $X^2$. The exact same result can be derived using the column profiles instead of row profiles. Another result that is apparent from the geometric development of CA is that the origin in the plots represents the average row and the average column profile. The closer a point is to the origin, the more similar it is to its average. In other words, points that are far from the origin contribute more to the $X^2$ than those that are close to the origin.

Normalized row and column scale values are frequently plotted in the same plot to provide a picture of the relationship between the variables; however, the distances between row and column points in these joint plots do *not* represent the relationship between the row and column variables (Greenacre, 1984; Goodman, 1986; Carroll, Green & Schaffer, 1986). Only the distances between rows or between columns are defined. The distances between rows and columns are *not* defined (meaningful). Since the row scores are weighted averages of the column scores and vice versa, the relationship or interaction between the rows and columns is represented in joint plots by the *relative* locations of individual row points with respect to all of the column points and vice versa.

Furthermore, the absolute distances between row points and column points in joint plots depend on how the scale values were normalized. Israëls (1987) thoroughly discusses different normalizations and their respective uses and interpretations.

# B Maximum Likelihood Estimation

The maximum likelihood (ML) equations for bivariate interaction, log bilinear models with consistency constraints on the scale values (e.g., equations 5.4 and 5.5) and log 3–mode models are given in Section B.1. A simple algorithm for computing the ML estimates of the model parameters using unidimensional Newton-Raphson is described in Section B.2. In Section B.3, alterations to the equations given in the previous sections are briefly described such that models can be fit to data that contain structural zeros, missing values, or cells that (for some other reason) should not be fit by the model.

## B.1 MLE Equations

Let $\mathbf{\Theta}$ equal the vector of parameters to be estimated (e.g., $\mathbf{\Theta} = (u, u_1^A, \ldots, \eta_{km})$). If the $F_{ijk}$'s are independent, Poisson random variables cross-classified into an $(I \times J \times K)$ table of variables $A$, $B$ and $C$, then the likelihood equation for $\{F_{ijk} | i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K\}$ is

$$L(\mathbf{\Theta}) = \prod_i \prod_j \prod_k \left( \frac{e^{-F_{ijk}} F_{ijk}^{f_{ijk}}}{f_{ijk}!} \right) \tag{B.1}$$

where $f_{ijk}$ is an observed frequency and $F_{ijk}$ is from a special case of the following general model

$$F_{ijk} = \exp(u_{ijk}^{(0)} + u_{ijk}^{(1)} + u_{ijk}^{(1,2)}) \tag{B.2}$$

or

$$\ln(F_{ijk}) = u^{(0)}_{ijk} + u^{(1)}_{ijk} + u^{(1,2)}_{ijk} \tag{B.3}$$

where $u^{(0)} = (u + u^A_i + u^B_j + u^C_k)$, $u^{(1)}$ equals any first order (2–way) interaction terms that are in the model (e.g., $u^{AB}_{ij}$, $u^{AC}_{ik}$, and/or $u^{BC}_{jk}$), and $u^{(1,2)}$ equals the sum of either bilinear or trilinear terms. This general model includes both bivariate interaction, log-bilinear models (referred to here as "log bilinear" models, for short), which were described in Chapter 5, and log 3–mode models, which were described in Chatper 6.

Rather than using equation B.1, it is easier to use the logarithm of it, which equals

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Theta}) &= \sum_i \sum_j \sum_k (-\ln(f_{ijk}!) + f_{ijk} \ln(F_{ijk}) - F_{ijk}) && \text{(B.4)} \\ &= \mathcal{C} + \sum_i \sum_j \sum_k f_{ijk} \ln(F_{ijk}) - \sum_i \sum_j \sum_k F_{ijk} && \text{(B.5)} \end{aligned}$$

where $\mathcal{C} = -\sum_i \sum_j \sum_k \ln(f_{ijk}!)$, which is a constant that depends only on the data (i.e., $\mathcal{C}$ does not depend on the model parameters).

To find the model parameters that maximize the likelihood, the specific model for $F_{ijk}$ is subsituted into the equation B.5 and the partial derivatives of $\mathcal{L}(\boldsymbol{\Theta})$ with respect to each of the model parameters are taken (i.e., the gradient). Setting these derivatives equal to zero and solving the set of equations for the model parameters yields the parameters that maximize $\mathcal{L}(\boldsymbol{\Theta})$.

Both the log bilinear and the log 3-mode models contain an overall effect and main effect terms. The ML equations for these parameters and any 2–way interactions terms that are in a model are the same for both types of models. The partial derivative of the

log likelihood function for the overall effect is

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial u} &= \sum_i \sum_j \sum_k f_{ijk} - F_{ijk} \\ &= f_{+++} - F_{+++} \end{aligned} \tag{B.6}$$

The partial derivatives for the main effect terms $u_i^A$, $u_j^B$ and $u_k^C$ are

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial u_i^A} = f_{i++} - F_{i++} \tag{B.7}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial u_j^B} = f_{+j+} - F_{+j+} \tag{B.8}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial u_k^C} = f_{++k} - F_{++k} \tag{B.9}$$

The second derivates of equations B.6 – B.9 are always negative. For example, the second derivate of equation B.7 is $\partial^2 \mathcal{L}(\boldsymbol{\Theta}) / \partial^2 u_i^A = -F_{i++} < 0$; therefore, setting equations B.6, B.7, B.8 and B.9 equal to zero yields a maximum. Thus, the ML equations for $u$, $u_i^A$, $u_j^B$ and $u_k^C$ are

$$f_{+++} - F_{+++} = 0 \tag{B.10}$$

$$f_{i++} - F_{i++} = 0 \tag{B.11}$$

$$f_{+j+} - F_{+j+} = 0 \tag{B.12}$$

$$f_{++k} - F_{++k} = 0 \tag{B.13}$$

For any 2–way interaction terms that are included in either a log bilinear model or log 3-mode model, the partial derivatives of the log likelihood for these terms are needed. The partial derivative for $u_{ij}^{AB}$ is

$$\begin{aligned} \frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial u_{ij}^{AB}} &= \sum_k (f_{ijk} - F_{ijk}) \\ &= f_{ij+} - F_{ij+} \end{aligned} \tag{B.14}$$

and the partial derivatives for $u_{ik}^{AC}$ and $u_{jk}^{BC}$ are

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial u_{ik}^{AC}} = f_{i+k} - F_{i+k} \tag{B.15}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial u_{jk}^{BC}} = f_{+jk} - F_{+jk} \tag{B.16}$$

Since the second derivatives of equations B.14, B.15 and B.16 are all negative, setting them equals to zero yields a maximum. Thus, the ML equations for $u_{ij}^{AB}$, $u_{ik}^{AC}$ and $u_{jk}^{BC}$ are

$$f_{ij+} - F_{ij+} = 0 \tag{B.17}$$

$$f_{i+k} - F_{i+k} = 0 \tag{B.18}$$

$$f_{+jk} - F_{+jk} = 0 \tag{B.19}$$

While the ML equations for the $u$-terms of log bilinear and log 3-mode models are the same (and are the same as those for loglinear models), the ML equations for the scale values and intrinsic association parameters of the two types of models are different. For bilinear models, the partial derivatives of the log likelihood with respect to the scale values $\mu_{im}$, $\nu_{jm}$ and $\eta_{km}$, and the intrinsic association parameters $\phi_m^{AB}$, $\phi_m^{AC}$ and $\phi_m^{BC}$ are

$$\frac{\partial \mathcal{L}(\mathbf{\Theta})}{\partial \mu_{im}} = \sum_j \sum_k (\phi_m^{AB} \nu_{jm} + \phi_m^{AC} \eta_{km})(f_{ijk} - F_{ijk}) \tag{B.20}$$

$$\frac{\partial \mathcal{L}(\mathbf{\Theta})}{\partial \nu_{jm}} = \sum_i \sum_k (\phi_m^{AB} \mu_{im} + \phi_m^{BC} \eta_{km})(f_{ijk} - F_{ijk}) \tag{B.21}$$

$$\frac{\partial \mathcal{L}(\mathbf{\Theta})}{\partial \eta_{km}} = \sum_i \sum_j (\phi_m^{AB} \mu_{im} + \phi_m^{BC} \eta_{km})(f_{ijk} - F_{ijk}) \tag{B.22}$$

$$\frac{\partial \mathcal{L}(\mathbf{\Theta})}{\partial \phi_m^{AB}} = \sum_i \sum_j (\mu_{im} \nu_{jm})(f_{ij+} - F_{ij+})$$

$$\frac{\partial \mathcal{L}(\mathbf{\Theta})}{\partial \phi_m^{AC}} = \sum_i \sum_k (\mu_{im} \eta_{km})(f_{i+k} - F_{i+k}) \tag{B.23}$$

$$\frac{\partial \mathcal{L}(\mathbf{\Theta})}{\partial \phi_m^{BC}} = \sum_j \sum_k (\nu_{jm} \eta_{km})(f_{+jk} - F_{+jk}) \tag{B.24}$$

The second derivatives of equations B.20 — B.24 are negative (e.g., $\partial^2 \mathcal{L}(\Theta)/\partial^2 \mu_{im} = -\sum_j \sum_k (\phi_m^{AB} \nu_{jm} + \phi_m^{AC} \eta_{km})^2 F_{ijk} < 0$). Thus, the ML equations for the scale values $\mu_{im}$,

$\nu_{jm}$ and $\eta_{km}$, and the intrinsic association parameters $\phi_m^{AB}$, $\phi_m^{AC}$ and $\phi_m^{BC}$ are

$$\sum_j \sum_k (\phi_m^{AB} \nu_{jm} + \phi_m^{AC} \eta_{km})(f_{ijk} - F_{ijk}) = 0 \quad \text{(B.25)}$$

$$\sum_i \sum_k (\phi_m^{AB} \mu_{im} + \phi_m^{BC} \eta_{km})(f_{ijk} - F_{ijk}) = 0 \quad \text{(B.26)}$$

$$\sum_i \sum_j (\phi_m^{AB} \mu_{im} + \phi_m^{BC} \eta_{km})(f_{ijk} - F_{ijk}) = 0 \quad \text{(B.27)}$$

$$\sum_i \sum_j (\mu_{im} \nu_{jm})(f_{ij+} - F_{ij+}) = 0 \quad \text{(B.28)}$$

$$\sum_i \sum_k (\mu_{im} \eta_{km})(f_{i+k} - F_{i+k}) = 0 \quad \text{(B.29)}$$

$$\sum_j \sum_k (\nu_{jm} \eta_{km})(f_{+jk} - F_{+jk}) = 0 \quad \text{(B.30)}$$

For log 3-mode models, the partial derivatives of the log likelihood with respect to the scale values $\mu_{ir}$, $\nu_{js}$ and $\eta_{kt}$ are

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial \mu_{ir}} = \sum_j \sum_k (\sum_s \sum_t \phi_{rst} \nu_{js} \eta_{kt})(f_{ijk} - F_{ijk}) \quad \text{(B.31)}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial \nu_{js}} = \sum_i \sum_k (\sum_r \sum_t \phi_{rst} \mu_{ir} \eta_{kt})(f_{ijk} - F_{ijk}) \quad \text{(B.32)}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial \eta_{kt}} = \sum_i \sum_j (\sum_r \sum_s \phi_{rst} \mu_{ir} \nu_{js})(f_{ijk} - F_{ijk}) \quad \text{(B.33)}$$

The partial derivative of the log likelihood with respect to $\phi_{rst}$ is

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta})}{\partial \phi_{rst}} = \sum_i \sum_j \sum_k (\mu_{ir} \nu_{js} \eta_{kt})(f_{ijk} - F_{ijk}) \quad \text{(B.34)}$$

The second partial derivatives of equations B.31 — B.34 are negative, thus setting these equations equal to zero yields a maximum. The ML equations for $\mu_{ir}$, $\nu_{js}$, $\eta_{kt}$ and $\phi_{rst}$ are

$$\sum_j \sum_k (\sum_s \sum_t \phi_{rst}\nu_{js}\eta_{kt})(f_{ijk} - F_{ijk}) = 0 \quad \text{(B.35)}$$

$$\sum_i \sum_k (\sum_r \sum_t \phi_{rst}\mu_{ir}\eta_{kt})(f_{ijk} - F_{ijk}) = 0 \quad \text{(B.36)}$$

$$\sum_i \sum_j (\sum_r \sum_s \phi_{rst}\mu_{ir}\nu_{js})(f_{ijk} - F_{ijk}) = 0 \quad \text{(B.37)}$$

$$\sum_i \sum_j \sum_k (\mu_{ir}\nu_{js}\eta_{kt})(f_{ijk} - F_{ijk}) = 0 \quad \text{(B.38)}$$

## B.2 Algorithm

The equations and steps necessary to create an iterative algorithm for estimating the parameters of any of the log bilinear (Chapters 5) or log 3–mode models (Chapter 6) are presented here. On each cycle of an algorithm, the following steps are executed:

1. Up-date estimates of all of the model parameters
2. Center main effect terms, and (depending on the specific model) center 2–way interaction effect terms and scale values.
3. Check for convergence

Steps $1 - 3$ are repeated until the estimated parameters have converged. Each of these steps are described in more detail in Sections B.2.1, B.2.2, and B.2.3. After convergence,

the last step is to re-scale the scale values such that the identification constraints are met (e.g, $\sum_i \mu_{ir}\mu_{ir'} = \delta_{rr'}$).

### B.2.1 Step 1: Up-date Parameter Estimates

The up-dating equations given here are based on the unidimensional version of the Netwon-Raphson method (Agresti, 1984). This method has been successfully used by others for estimating the parameters of $RC(M)$ association models and related models (Goodman, 1979, 1985; Becker, 1990a; Choulakian, 1988a, 1988b). Like other methods, the Netwon-Raphson method leads to an iterative algorithm that requires starting values for all of the parameters. The unidimensional version of Newton-Raphson is relatively simple and does not require inverting a potentially large matrix of second derivatives on each iteration. Disadvantages of this method are that it does not automatically yield estimates of standard errors and it can be slow to converge. If estimates of standard errors are desired, then they can be computed using the ML estimates of the model parameters.

The unidimensional method is essentially is the same as the Netwon-Raphson method, except that rather than using the entire matrix second derivatives, only the diagonals are used. The ML equation for a parameter is treated as a function of that parameter and the remaining parameters are treated as fixed. On each cycle of the algorithm, a sequence of up-dating equations are executed, one for each of the parameters.

The equations for up-dating the estimates of the model parameters are obtained as follows. Let $\theta^{[q]}$ be the value of a model parameter on the qth iteration. The value of $\theta$

on the next iteration, $q+1$, is

$$\theta^{[q+1]} = \theta^{[q]} - \frac{h(\theta^{[q]})}{h'(\theta^{[q]})} \tag{B.39}$$

where $h(\theta^{[q]})$ is the value of the maximum likelihood equation for parameter $\theta^{[q]}$ (i.e., the first derivative of the likelihood equation with respect to $\theta$) evaluated using the current estimates of the other parameters, and $h'(\theta^{[q]})$ is the value of the derivative of $h(\theta^{[q]})$, which is the second derivative of the likelihood equation for the parameter $\theta$.

For the log bilinear and 3–mode models (and loglinear models as well), the up-dating equations for main effect terms are

$$u_i^{A[q+1]} = u_i^{A[q]} + \frac{f_{i++} - F_{i++}^{[q]}}{F_{i++}^{[q]}} \tag{B.40}$$

$$u_j^{B[q+1]} = u_j^{B[q]} + \frac{f_{+j+} - F_{+j+}^{[q]}}{F_{+j+}^{[q]}} \tag{B.41}$$

$$u_k^{C[q+1]} = u_k^{C[q]} + \frac{f_{++k} - F_{++k}^{[q]}}{F_{++k}^{[q]}} \tag{B.42}$$

where $F_{i++}^{[q]} = \sum_j \sum_k F_{ijk}^{[q]}$, $F_{+j+}^{[q]} = \sum_i \sum_k F_{ijk}^{[q]}$, $F_{++k}^{[q]} = \sum_i \sum_j F_{ijk}^{[q]}$, and $F_{ijk}^{[q]}$ is the fitted value from a specific version of the general model (equation B.2) computed using the current parameters estimates, which are from the $q$th iteration.

The up-dating equations for any 2 way-interaction effect terms that are included in a model are

$$u_{ij}^{AB[q+1]} = u_{ij}^{AB[q]} + \frac{f_{ij+} - F_{ij+}^{[q]}}{F_{ij+}^{[q]}} \tag{B.43}$$

$$u_{ik}^{AC[q+1]} = u_{ik}^{AC[q]} + \frac{f_{i+k} - F_{i+k}^{[q]}}{F_{i+k}^{[q]}} \tag{B.44}$$

$$u_{jk}^{BC[q+1]} = u_{jk}^{BC[q]} + \frac{f_{+jk} - F_{+jk}^{[q]}}{F_{+jk}^{[q]}} \tag{B.45}$$

For a log bilinear model, the up-dating equations for scale values that are to be estimated are

$$\mu_{im}^{[q+1]} = \mu_{im}^{[q]} + \frac{\sum_j \sum_k (\phi_m^{AB[q]} \nu_{jm}^{[q]} + \phi_m^{AC[q]} \eta_{km}^{[q]})(f_{ijk} - F_{ijk}^{[q]})}{\sum_j \sum_k (\phi_m^{AB[q]} \nu_{jm}^{[q]} + \phi_m^{AC[q]} \eta_{km}^{[q]})^2 F_{ijk}^{[q]}} \tag{B.46}$$

$$\nu_{jm}^{[q+1]} = \nu_{jm}^{[q]} + \frac{\sum_i \sum_k (\phi_m^{AB[q]} \mu_{im}^{[q]} + \phi_m^{BC[q]} \eta_{km}^{[q]})(f_{ijk} - F_{ijk}^{[q]})}{\sum_i \sum_k (\phi_m^{AB[q]} \mu_{im}^{[q]} + \phi_m^{BC[q]} \eta_{km}^{[q]})^2 F_{ijk}^{[q]}} \tag{B.47}$$

$$\eta_{km}^{[q+1]} = \eta_{km}^{[q]} + \frac{\sum_i \sum_j (\phi_m^{AC[q]} \mu_{im}^{[q]} + \phi_m^{BC[q]} \nu_{jm}^{[q]})(f_{ijk} - F_{ijk}^{[q]})}{\sum_i \sum_j (\phi_m^{AC[q]} \mu_{im}^{[q]} + \phi_m^{BC[q]} \nu_{jm}^{[q]})^2 F_{ijk}^{[q]}} \tag{B.48}$$

If a 2–way interaction term, for example, $u_{ij}^{AB}$, is included in a log bilinear model, then $\phi_m^{AB}$ is set equal to zero in equations B.46 and B.47. Likewise, if $u_{ik}^{AC}$ and/or $u_{jk}^{BC}$ are included in a model, then $\phi_m^{AC}$ and/or $\phi_m^{BC}$, respectively, are set equal to zero.

For the log bilinear models, the up-dating equations for the intrinsic association parameters to be estimated (i.e., those that are not set equal to zero) are

$$\phi_m^{AB[q+1]} = \phi_m^{AB[q]} + \frac{\sum_i \sum_j (\mu_{im}^{[q]} \nu_{jm}^{[q]})(f_{ij+} - F_{ij+}^{[q]})}{\sum_i \sum_j (\mu_{im}^{[q]} \nu_{jm}^{[q]})^2 F_{ij+}^{[q]}} \tag{B.49}$$

$$\phi_m^{AC[q+1]} = \phi_m^{AC[q]} + \frac{\sum_i \sum_k (\mu_{im}^{[q]} \eta_{km}^{[q]})(f_{i+k} - F_{i+k}^{[q]})}{\sum_i \sum_k (\mu_{im}^{[q]} \eta_{km}^{[q]})^2 F_{i+k}^{[q]}} \tag{B.50}$$

$$\phi_m^{BC[q+1]} = \phi_m^{BC[q]} + \frac{\sum_j \sum_k (\nu_{jm}^{[q]} \eta_{km}^{[q]})(f_{+jk} - F_{+jk}^{[q]})}{\sum_j \sum_k (\nu_{jm}^{[q]} \eta_{km}^{[q]})^2 F_{+jk}^{[q]}} \tag{B.51}$$

For all of the log 3-mode models, regardless of which (if any) of the 2–way interaction effect terms included in the model, all of the following up-dating equations for the scale values and intrinsic association parameters are needed:

$$\mu_{ir}^{[q+1]} = \mu_{ir}^{[q]} + \frac{\sum_j \sum_k (\sum_s \sum_t \phi_{rst}^{[q]} \nu_{js}^{[q]} \eta_{kt}^{[q]})(f_{ijk} - F_{ijk}^{[q]})}{\sum_j \sum_k (\sum_s \sum_t \phi_{rst}^{[q]} \nu_{js}^{[q]} \eta_{kt}^{[q]})^2 F_{ijk}^{[q]}} \tag{B.52}$$

$$\nu_{js}^{[q+1]} = \nu_{js}^{[q]} + \frac{\sum_i \sum_k (\sum_r \sum_t \phi_{rst}^{[q]} \mu_{ir}^{[q]} \eta_{kt}^{[q]})(f_{ijk} - F_{ijk}^{[q]})}{\sum_i \sum_k (\sum_r \sum_t \phi_{rst}^{[q]} \mu_{ir}^{[q]} \eta_{kt}^{[q]})^2 F_{ijk}^{[q]}} \tag{B.53}$$

$$\eta_{kt}^{[q+1]} = \eta_{kt}^{[q]} + \frac{\sum_i \sum_j (\sum_r \sum_s \phi_{rst}^{[q]} \mu_{ir}^{[q]} \nu_{js}^{[q]})(f_{ijk} - F_{ijk}^{[q]})}{\sum_i \sum_j (\sum_r \sum_s \phi_{rst}^{[q]} \mu_{ir}^{[q]} \nu_{js}^{[q]})^2 F_{ijk}^{[q]}} \tag{B.54}$$

$$\phi_{rst}^{[q+1]} = \phi_{rst}^{[q]} + \frac{\sum_i \sum_j \sum_k (\mu_{ir}^{[q]} \nu_{js}^{[q]} \eta_{kt}^{[q]})(f_{ijk} - F_{ijk}^{[q]})}{\sum_i \sum_j \sum_k (\mu_{ir}^{[q]} \nu_{js}^{[q]} \eta_{kt}^{[q]})^2 F_{ijk}^{[q]}} \tag{B.55}$$

The set of up-dating equations for estimating the ML estimates of the parameters of a specific model should contain one equation for each parameter. The order that they

are executed is not critical; however, after the equations for one set of parameters (e.g., for $\{u_i^A | i = 1, \ldots, I\}$) are executed, the new estimates of these parameters should be used in the equations for the next set of parameters (e.g., for $\{u_j^B | j = 1, \ldots, J\}$). Once all the equations for all of the parameters have been executed, step 1 of the alogrithm is complete.

### B.2.2 Step 2: Re-center Parameter Estimates

In this step, the new estimates of the parameters are re-centered such that they meet the identification constraints needed for a specific model. For all of the log bilinear and 3-mode models, the typical identification constraints are placed on the main effect terms and any 2–way interaction effect terms (i.e., the $u$-terms); namely, either $\sum_i u_i^A = \sum_j u_j^B = \sum_k u_k^C = \sum_i u_{ij}^{AB} = \ldots = 0$. If the alternative identification contraints for the $u$-terms (i.e., $u_1^A = y_1^B = u_1^C = u_{1j}^{AB} = 0$) are used, then the $u$-terms do not need to be centered.

The constraints required to identify the scale values for the log bilinear and 3-mode models are different. For log bilinear models, all of the scale values are centered around 0,

$$\sum_i \mu_{im} h_{A(i)} = \sum_j \nu_{jm} h_{B(j)} = \sum_k \eta_{km} h_{C(k)} = 0 \tag{B.56}$$

where $h_{A(i)}$, $h_{B(j)}$, and $h_{C(k)}$ are fixed and known weights. The centering constraint, equation B.56, is imposed at step 2 of the alogrithm, but the scaling constraint is not

needed until after the alogrithm has converged. Recall that this scaling constraint is

$$\sum_i \mu_{im}^2 h_{A(i)} = \sum_j \nu_{jm}^2 h_{B(j)} = \sum_k \eta_{km}^2 h_{C(k)} = 1 \tag{B.57}$$

While it is not necessary to impose constraint B.57 at step 2 of the alogrithm, if the scale values are going to be restricted to be orthonormal (e.g., $\sum_i \mu_{im}\mu_{im'} = \delta_{mm'}$), then constraint B.57 along with the orthogonality restriction should be imposed at this point.

For the log 3-mode models, whether the scale values estimated for particular modes need to be centered depends on which 2–way interaction effects are in the model. See Table 6.1 for a list of the required centering constraints for each of the possible log 3-mode models. Regardless of the 2–way interaction effects, all scale values are constrained to be orthogonal:

$$\sum_i \mu_{ir}\mu_{ir'} = \delta_{rr'} \tag{B.58}$$

$$\sum_j \nu_{js}\nu_{js'} = \delta_{ss'} \tag{B.59}$$

$$\sum_k \nu_{kt}\nu_{kt'} = \delta_{tt'} \tag{B.60}$$

These orthogonality and scaling constraints, equations B.58, B.59, and B.60, do not have to be imposed at step 2 on each iteration, but can be imposed after the alogrithm has converged.

For both log bilinear and log 3-mode models, the scale values should be centered first, followed by centering any 2-way interaction effects, and lastly the main effect terms should be centered. If any scaling constraints or restrictions are going to be imposed on the scale values at step 2, they should be done after the parameters have all been re-centered.

### B.2.3 Step 3: Check for Convergence

Unlike loglinear models, not all of the ML equations (in general) will be completely satisfied. All of the 1-way margins will be fit perfectly (i.e., $\hat{F}_{i++} = f_{i++}$, $\hat{F}_{+j+} = f_{+j+}$, and $\hat{F}_{++k}$); therefore, equations B.11, B.12, and B.13 will be satisfied. The 2-way margins corresponding to any 2–way interaction effects that are included in the model will also be fit perfectly, and the ML equations for these parameters will also be met. The ML equations for scale values and intrinsic association parameters will not necessarly equal 0, but they will be minimized.

Rather than using the ML equations to determine whether the estimated parameters $\boldsymbol{\Theta}^{[\mathbf{q}]}$ have converged to the maximum likelihood estimates $\hat{\boldsymbol{\Theta}}^{[\mathbf{q}]}$, the fitted values based on the estimated parameters from the last iteration are compared with those from the next to last iteration. When the sum of the absolute values of the differences between the fitted values from the last and next to last iteration is less than epsilon (a small number set by the researcher), the fitted values have not changed or have changed very little, and the estimates have converged satisfactorily. If the sum is greater than epsilon, then another iteration of the algorithm should be performed. After the estimated parameters have

converged satisfactorily, the scale values should be re-scaled to meet the identification constraints.

## B.3 Missing or Fixed Values

Only minor modifications to the equations and alogrithm given in the two pervious section are required to estimated the parameters of a model where the fitted values of particular cells of the data/table are set equal to the observed value, zero, or some other value. Let $\omega_{ijk} = 0$ for those cells that are not fit by the model and are set equal to the observed or some other value. The general model, equation B.2, is replaced in the equations previous given by

$$F_{ijk} = \exp(u_{ijk}^{(0)} + u_{ijk}^{(1)} + u_{ijk}^{(1,2)})\omega_{ijk} + f_{ijk}(1 - \omega_{ijk}) \tag{B.61}$$

Working throught the derivation of the up-dating equations using equation B.61 in the log likelihood yields the same equations as previously given are obtained, except that equation B.61 should be used to compute $F_{ijk}^{[q]}$ rather than equation B.2. Additionaly, the degrees of freedom for a specific model are then reduced by the number of cells where $\omega_{ijk} = 0$.

# C FORTRAN Programs

The *FORTRAN* program used to fit the bivariate interaction, log-bilinear models to the peer play data in Section 5.1 is given in Section C.1, and the program used to fit the log 3-mode models in Chapter 6 is given in Section C.2. Both of these programs use the algorithm presented in Appendix B.

## C.1 Bivariate Interaction, Log-bilinear Models

```
C MDno3way.f
C
C  written by Carolyn J. Anderson
C  last revised:  October 30, 1991
C
C  Fits and estimates the parameters of multi-dimensional, no 3-way
C   association models with consistency constraints on scale values.
C
C  Either (i)  all 2-way interactions are decomposed or
C         (ii) the AC and BC margins are decomposed while the AB margin
C              is fit
C
C   IuAB  = 0  option (i) --->  fix uAB(i,j) = 0 for all i & j, and
C                               estimate phiAB(m)
C         = 1  option (ii) --->  fix phiAB(m) = 0 for all m, and
C                               estimate uAB(i,j)
C
C   Iorth = 0  sum of squared scale values = 1
C         = 1  scale values are orthonormalized
C
C
C  Uni-dimensional Newton-Raphson is used to find MLE of model parameters.
C
C   Mn = Max(NA, NB, NC)       = Max Number of levels/variable
C   Mdim = Min(NA, NB, NC) - 1  = Max Number of dimensions
C
C
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FAB(Mn,Mn),FAC(Mn,Mn),FBC(Mn,Mn),FA(Mn),
     +FB(Mn),FC(Mn),WA(Mn),WB(Mn),WC(Mn),FHAT(Mn,Mn,Mn),FHA(Mn),FHB(Mn)
     +,FHC(Mn),FHAB(Mn,Mn),FHAC(Mn,Mn),FHBC(Mn,Mn),uA(Mn),uB(Mn),uC(Mn)
     +,uAB(Mn,Mn),ETA(Mn,Mdim),phiAB(Mdim),phiAC(Mdim),phiBC(Mdim),
     +Ffit(Mn,Mn,Mn),ChiSq(Mn,Mn,Mn),Chi1(Mn),Chi2(Mn,Mn)
      double precision MU(Mn,Mdim),NU(Mn,Mdim)
      character*30 iname, oname
      character*79 title
      common / iter / CONV,DIFF,IMAX,ICALLS,ICONV,IHIST
      common / nunit / nuniti, nunito
      ZERO = 0.0
      nuniti = 3
      nunito = 4
```

```
    1 write(*,2)
    2 format('Enter name of input file:  ')
      read(*,600,end=1) iname
      open(unit=nuniti,file=iname,status='OLD')
    3 write(*,4)
    4 format('Enter name of output file:  ')
      read(*,600,end=3) oname
      open(unit=nunito,file=oname,status='NEW')
c                                                  Problem specifications
c                                                                  Input
      read(nuniti,610) title
      read(nuniti,*)  NA, NB, NC, NDim, delta, IuAB, Iorth,
     +              inWA, inWB, inWC, IMAX, CONV, IHIST
c                                                                 Output
      write(nunito,620) title
      write(nunito,300) NA, NB, NC, NDim, delta
      if (IuAB.eq.1) write(nunito,301) IuAB
      if (Iorth.eq.0) then
        write(nunito,302) Iorth
       else
        write(nunito,303) Iorth
      endif
      write(nunito,310) inWA,inWB,inWC
c                                                  Read in data  and Ffit
c                                                   fastest = variable C
c                                                   slowest = variable A
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
         read(nuniti,*) F(i,j,k), Ffit(i,j,k)
         F(i,j,k) = F(i,j,k) + delta
   10  continue
c                                                Compute observed margins
      call AMARG(F,FA,NA,NB,NC)
      call BMARG(F,FB,NA,NB,NC)
      call CMARG(F,FC,NA,NB,NC)
      call ABMARG(F,FAB,NA,NB,NC)
      call ACMARG(F,FAC,NA,NB,NC)
      call BCMARG(F,FBC,NA,NB,NC)
      call SUM1W(FA,NA,FTOT)
      if (Ndim.eq.0) goto 101
c                             Weights:  input, unit or marginal probabilties
c                                inWx =  0    1      2
      if (inWA.eq.0) then
        read(nuniti,*) (WA(i),i=1,NA)
      else
        if (inWA.eq.1) then
          call ONE1W(WA,NA)
        else
          call DivByX(FTOT,FA,WA,NA)
        endif
      endif
      if (inWB.eq.0) then
        read(nuniti,*) (WB(j),j=1,NB)
      else
        if (inWB.eq.1) then
          call ONE1W(WB,NB)
        else
          call DivByX(FTOT,FB,WB,NB)
        endif
      endif
```

```
      if (inWC.eq.0) then
        read(nuniti,*) (WC(k),k=1,NC)
      else
        if (inWC.eq.1) then
          call ONE1W(WC,NC)
        else
          call DivByX(FTOT,FC,WC,NC)
        endif
      endif
c                                                          Starting values
c
      read(nuniti,*) u,(uA(i),i=1,NA),(uB(j),j=1,NB),(uC(k),k=1,NC)
      if (IuAB.eq.1) then
        do i = 1, NA
          read(nuniti,*) (uAB(i,j),j=1,NB)
        end do
        call ZERO1W(phiAB,Ndim)
      endif
      if (Ndim.gt.0) then
        do m = 1, NDim
          read(nuniti,*)  (MU(i,m),i=1,NA)
        end do
        do m = 1, NDim
          read(Nuniti,*)  (NU(j,m),j=1,NB)
        end do
        do m = 1, NDim
          read(Nuniti,*)  (ETA(k,m),k=1,NC)
        end do
        if (IuAB.eq.0) then
          read(Nuniti,*)   (phiAB(m),m=1,Ndim)
          call ZERO2W(uAB,NA,NB)
        endif
        read(nuniti,*)  (phiAC(m),m=1,Ndim)
        read(nuniti,*)  (phiBC(m),m=1,Ndim)
      else
        call ZERO2W(MU,NA,Ndim)
        call ZERO2W(NU,NB,Ndim)
        call ZERO2W(ETA,NC,Ndim)
        call ZERO1W(phiAC,Ndim)
        call ZERO1W(phiBC,Ndim)
      endif
c                                                  Output starting values
      write(nunito,620) title
      write(nunito,320)
      call OUmain(u,uA,uB,uC,NA,NB,NC)
      call OUTint(MU,NU,ETA,phiAB,phiAC,phiBC,uAB,NA,NB,NC,NDim,
     +           IuAB,WA,WB,WC)
c                                       ......Heart of the program.......
c
      ICONV = 0
      call MODEL(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,NA,NB,NC,
     + Ndim,IuAB,Ffit,F,FAB,FAC,FBC,FA,FB,FC,FHAT,FHAB,FHAC,FHBC,FHA,
     + FHB,FHC,WA,WB,WC,Iorth)
c
      call ABMARG(FHAT,FHAB,NA,NB,NC)
      call ACMARG(FHAT,FHAC,NA,NB,NC)
      call BCMARG(FHAT,FHBC,NA,NB,NC)
      call AMARG(FHAT,FHA,NA,NB,NC)
      call BMARG(FHAT,FHB,NA,NB,NC)
      call CMARG(FHAT,FHC,NA,NB,NC)
c                                                          Output results
```

```
c                                        Convergence message & fit statistics
      if (ICONV.EQ.1) then
        write(nunito,350)
        write(*,350)
      else
        write(nunito,355)
        write(*,355)
      endif
      write(nunito,630)
      write(nunito,621) title
      write(nunito,410) ICALLS, IMAX, DIFF, CONV
c
      do 210 i = 1, NA
      do 210 j = 1, NB
      do 210 k = 1, NC
        if (FHAT(i,j,k).ne.ZERO)   ChiSq(i,j,k) =
     +    (F(i,j,k)-FHAT(i,j,k))*(F(i,j,k)-FHAT(i,j,k))/FHAT(i,j,k)
  210 continue
c                                                 Compute degrees of freedom
      Nless = 0
      do 423 i = 1, NA
       do 423 j = 1, NB
        do 423 k = 1, NC
  423     Nless = Nless + 1 - INT(Ffit(i,j,k))
      ndf = NA*NB*NC - 1 - (1+Ndim)*(NA+NB+NC-3) - Nless
     +      - IuAB*( (NA-1)*(NB-1) - Ndim) + Iorth*Ndim*(Ndim-1)*3/2
      write(nunito,420)
      write(nunito,430) ndf, Nless
      call GSQUAR(F,FHAT,GSQ,NA,NB,NC)
      call CHI(GSQ,ndf,pvalG)
      write(nunito,440) GSQ, pvalG
      call PEARSO(F,FHAT,XSQ,NA,NB,NC)
      call CHI(XSQ,ndf,pvalX)
      write(nunito,450) XSQ, pvalX
c                                    Observed data, fitted values, & residuals
      write(nunito,620) title
      write(nunito,500) FTOT
c                                                              1-way margins
      write(nunito,510)
      call AMARG(ChiSq,Chi1,NA,NB,NC)
      call OUT1W(FTOT,FA,FHA,Chi1,NA)
      write(nunito,520)
      call BMARG(ChiSq,Chi1,NA,NB,NC)
      call OUT1W(FTOT,FB,FHB,Chi1,NB)
      write(nunito,530)
      call CMARG(ChiSq,Chi1,NA,NB,NC)
      call OUT1W(FTOT,FC,FHC,Chi1,NC)
c                                                              2-way margins
      write(nunito,620) title
      write(nunito,540)
      call ABMARG(ChiSq,Chi2,NA,NB,NC)
      call OUT2W(FTOT,FAB,FHAB,Chi2,NA,NB)
      write(nunito,550)
      call ACMARG(ChiSq,Chi2,NA,NB,NC)
      call OUT2W(FTOT,FAC,FHAC,Chi2,NA,NC)
      write(nunito,560)
      call BCMARG(ChiSq,Chi2,NA,NB,NC)
      call OUT2W(FTOT,FBC,FHBC,Chi2,NB,NC)
c                                                               3-way table
      write(nunito,620) title
      call OUT3W(FTOT,F,FHAT,Ffit,NA,NB,NC)
```

```
c                                                        Parameter estimates
      write(nunito,620) title
      write(nunito,570)
      call OUmain(u,uA,uB,uC,NA,NB,NC)
      call OUTint(MU,NU,ETA,phiAB,phiAC,phiBC,uAB,NA,NB,NC,Ndim,
     +           IuAB,WA,WB,WC)
      if ((Ndim.gt.0).and.((inWA.eq.2).and.(inWB.eq.2)).or.
     +     ((inWA.eq.2).and.(inWC.ne.1)).or.
     +     ((inWB.eq.2).and.(inWC.eq.2))  )         write(nunito,580)
      if ((inWA.eq.2).and.(inWB.eq.2).and.(IuAB.eq.0)) then
         do m = 1, Ndim
           call CORR(FAB,FTOT,MU,NU,rhoAB,NA,NB,m)
           write(nunito,581) m, rhoAB
         end do
      endif
      if ((inWA.eq.2).and.(inWC.eq.2)) then
        do m = 1, Ndim
          call CORR(FAC,FTOT,MU,ETA,rhoAC,NA,NC,m)
          write(nunito,582) m, rhoAC
        end do
      endif
      if ((inWB.eq.2).and.(inWC.eq.2)) then
        do m= 1, Ndim
          call CORR(FBC,FTOT,NU,ETA,rhoBC,NB,NC,m)
          write(nunito,583) m, rhoBC
        end do
      endif
c
      close(3,status='keep')
      close(4,status='keep')
      stop
c                                                          Format Statements
  300 format(1x,'Problem Specifications for Multi-dimensional No 3-way',
     +' association model',/,
     +/, ' Number of rows    = ',i2,/,' Number of columns = ',i2,
     +/,' Number of layers  = ',i2,//,' Number of dimensions = ',i2,
     +//,' Delta = ',E15.5,' was added to each cell')
  301 format(/,1x,'IuAB = ',i1,' ----> uAB(i,j) estimated ')
  302 format(/,1x,'Iorth = ',i1,' ---> sum of squared scale values = 1')
  303 format(/,1x,'Iorth = ',i1,' ---> sum of squared scale values = 1 '
     +,'and ',/,16x,'scales are orthogonal')
  310 format(/,1x,'Weights for scale values are ',
     +'inWA = ',i1,', inWB = ',i1,', and inWC = ',i1,/,
     +1x,'where inWx = 0  --> input',/,7x,'inWx = 1  --> unit',/,
     +7x,'inWx = 2  --> marginal probabilties',/ )
  320 format(1x,'Starting Values:  ')
  350 format(/,1x,'Solution converged ',/)
  355 format(/,1x,'Warning:  solution did not converge for specified',
     +' criterion ',/)
  410 format(1x,'Summary of Convergence Information:',//,
     +1x,'Number of cycles:',10x,'ICALLS = ',i6,/,
     +1x,'Maximum number of cycles:',4x,'IMAX = ',i6,/,
     +1x,'Change in fitted values:',2x,'DIFF = ',f15.10,/,
     +1x,'Convergence critertion:',3x,'CONV = ',f15.10,/)
      write(nunito,420)
  420 format(/,' Fit Statistics:  ',/)
  430 format(1x,'Degrees of freedom = ',i5,5x,'Note:  ',i2,
     +' cells where fitted = observed')
  440 format(/,1x,'Likelihood ratio statistic,  G**2 = ',f15.4,
     +5x,'p-value = ',f8.5)
  450 format(1x,'Pearson chi-square statitic, X**2 = ',f15.4,
```

```
     +5x,'p-value = ',f8.5,/)
  500 format(1x,'Total number of observations in table = ',f10.2,//)
  510 format(1x,'Row Margin:  ',/,1x,' i',6x,'P(i++)',6x,'F(i++)',
     +6x,'FHAT(i++)',6x,'Residual',4x,'Sum of Chi-terms')
  520 format(//,1x,'Column Margin:  ',/,1x,' j',6x,'P(+j+)',6x,'F(+j+)'
     +,6x,'FHAT(+j+)',6x,'Residual',4x,'Sum of Chi-terms')
  530 format(//,1x,'Layer Margin:  ',/,1x,' k   ',3x,'P(++k)',6x,'F(++k)'
     +,6x,'FHAT(++k)',6x,'Residual',4x,'Sum of Chi-terms')
  540 format(1x,'2-Way margins:',//,1x,'Row x Column:  ', /,1x,
     +' i   j   ',3x,'P(i,j,+)',6x,'F(ij+)',6x,'FHAT(ij+)',5x,
     +'Residual',4x,'Sum of Chi-terms')
  550 format(//,1x,'Row x Layer:  ', /,1x,
     +' i   k   ',3x,'P(i+k)',6x,'F(i+k)',6x,'FHAT(i+k)',5x,
     +'Residual',4x,'Sum of Chi-terms')
  560 format(//,1x,'Column x Layer:  ', /,1x,
     +' j   k   ',3x,'P(+jk)',6x,'F(+jk)',6x,'FHAT(+jk)',5x,
     +'Residual',4x,'Sum of Chi-terms')
  570 format(1x,'Estimated Model Parameters:',/)
  580 format(/,1x,'Correlations between scale values:')
  581 format(' rhoAB(',i2,') = ',f8.5)
  582 format(' rhoAC(',i2,') = ',f8.5)
  583 format(' rhoBC(',i2,') = ',f8.5)
  600 format(1a30)
  610 format(1a79)
  620 format('1',1a79,/)
  621 format(1x,1a79,/)
  630 format(/,1x,79('='),/)
      end
C
C  Set elements of a One dimensional array = ZERO
C
      subroutine ZERO1W(X,N)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn)
      data ZERO / 0.0 /
      do 10 I = 1, N
   10   X(I) = ZERO
      return
      end
C
C Set elements of a Two dimensional array = ZERO
C
      subroutine ZERO2W(X,NR,NC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn)
      data ZERO / 0.0 /
      do 10 I = 1, NR
        do 10 J = 1, NC
   10     X(I,J) = ZERO
      return
      end
C
C Set elements of 1-dimensional array equal to one
C
      subroutine ONE1W(W,N)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension W(Mn)
      do 10 i = 1, N
```

```
  10    W(i) = 1.0
        return
        end
C
C Sum the elements of a 1-dimensional array
C
        subroutine SUM1W(X,N,SUM)
        parameter (Mn=20,Mdim=19)
        implicit double precision (A-H,O-Z), integer*4 (i-n)
        dimension X(Mn)
        SUM = 0.0
        do 10 i = 1, N
   10   SUM = SUM + X(i)
        return
        end
C
C Divide elements of an array/vector by a constant/scaler
C  -- Use to compute 1-way marginal probabilities or means
C
        subroutine DivByX(X,V,Vx,N)
        parameter (Mn=20,Mdim=19)
        implicit double precision (A-H,O-Z), integer*4 (i-n)
        dimension V(Mn), Vx(Mn)
        do 10 i = 1, N
   10   Vx(i) = V(i)/X
        return
        end
C                                                          A MARGin
        subroutine AMARG(X,XA,NA,NB,NC)
        parameter (Mn=20,Mdim=19)
        implicit double precision (A-H,O-Z), integer*4 (i-n)
        dimension X(Mn,Mn,Mn), XA(Mn)
        call ZERO1W(XA,NA)
        do 10 I = 1, NA
         do 10 J = 1, NB
          do 10 K = 1, NC
   10     XA(I) = XA(I) + X(I,J,K)
        return
        end
C                                                          B MARGin
        subroutine BMARG(X,XB,NA,NB,NC)
        parameter (Mn=20,Mdim=19)
        implicit double precision (A-H,O-Z), integer*4 (i-n)
        dimension X(Mn,Mn,Mn), XB(Mn)
        call ZERO1W(XB,NB)
        do 10 J = 1, NB
         do 10 I = 1, NA
          do 10 K = 1, NC
   10    XB(J) = XB(J) + X(I,J,K)
        return
        end
C                                                          C MARGin
        subroutine CMARG(X,XC,NA,NB,NC)
        parameter (Mn=20,Mdim=19)
        implicit double precision (A-H,O-Z), integer*4 (i-n)
        dimension X(Mn,Mn,Mn), XC(Mn)
        call ZERO1W(XC,NC)
        do 10 K = 1, NC
         do 10 I = 1, NA
          do 10 J = 1, NB
   10    XC(K) = XC(K) + X(I,J,K)
```

```
      return
      end
C                                                              AB MARGin
      subroutine ABMARG(X,XAB,NA,NB,NC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XAB(Mn,Mn)
      call ZERO2W(XAB,NA,NB)
      do 10 I = 1, NA
       do 10 J = 1, NB
        do 10 K = 1, NC
   10    XAB(I,J) = XAB(I,J) + X(I,J,K)
      return
      end
C                                                              AC MARGin
      subroutine ACMARG(X,XAC,NA,NB,NC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XAC(Mn,Mn)
      call ZERO2W(XAC,NA,NC)
      do 10 I = 1, NA
       do 10 J = 1, NB
        do 10 K = 1, NC
   10    XAC(I,K) = XAC(I,K) + X(I,J,K)
      return
      end
C                                                              BC MARGin
      subroutine BCMARG(X,XBC,NA,NB,NC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XBC(Mn,Mn)
      call ZERO2W(XBC,NB,NC)
      do 10 I = 1, NA
       do 10 J = 1, NB
        do 10 K = 1, NC
   10    XBC(J,K) = XBC(J,K) + X(I,J,K)
      return
      end
C
C  Compute predictions of consistent, multi-dimensional no 3-way association
C   model
C
      subroutine PRED(F,X,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     + NA,NB,NC,Ndim,Ffit)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),X(Mn,Mn,Mn),uA(Mn),uB(Mn),uC(Mn),uAB(Mn,Mn)
     +,ETA(Mn,Mdim),phiAB(Mdim),phiAC(Mdim),phiBC(Mdim),Ffit(Mn,Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      do 30 i = 1, NA
        A  = u + uA(i)
        do 20 j = 1, NB
          PAB = 0
          do 5 m = 1, Ndim
    5       PAB =  PAB + phiAB(m)*MU(i,m)*NU(j,m)
          AB = A + uB(j) + PAB + uAB(i,j)
          do 10 k = 1, NC
           PACBC = 0
           do 6 m = 1, Ndim
    6        PACBC = PACBC  +
     +              (phiAC(m)*MU(i,m) + phiBC(m)*NU(j,m))*ETA(k,m)
```

```
          X(i,j,k) = (dexp(AB + uC(k) + PACBC))*Ffit(i,j,k)
     +               + F(i,j,k)*(1-Ffit(i,j,k))
   10    continue
   20   continue
   30 continue
      return
      end
C
C Update U-terms:  main effect and/or interaction terms
C
      subroutine UFIT(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +NA,NB,NC,Ndim,IuAB,Ffit,F,FA,FB,FC,FAB,FHAT,FHA,FHB,FHC,FHAB)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension uA(Mn),uB(Mn),uC(Mn),ETA(Mn,Mdim),FHAT(Mn,Mn,Mn),
     +FHA(Mn),FHB(Mn),FHC(Mn),F(Mn,Mn,Mn),FA(Mn),FB(Mn),FC(Mn),
     +FAB(Mn,Mn),FHAB(Mn,Mn),phiAB(Mdim),phiAC(Mdim),phiBC(Mdim),
     +uAB(Mn,Mn),Ffit(Mn,Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      if (IuAB.eq.0) then
c                                                              up-date uA(i)
       call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +          NA,NB,NC,Ndim,Ffit)
       call AMARG(FHAT,FHA,NA,NB,NC)
       do 10 i = 1, NA
   10    uA(i) = uA(i) + (FA(i) - FHA(i))/FHA(i)
c                                                              up-date uB(j)
       call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +          NA,NB,NC,Ndim,Ffit)
       call BMARG(FHAT,FHB,NA,NB,NC)
       do 20 j = 1, NB
   20    uB(j) = uB(j) + (FB(j) - FHB(j))/FHB(j)
c                                                         up-date uAB(i,j)
      else
       call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +          NA,NB,NC,Ndim,Ffit)
       call ABMARG(FHAT,FHAB,NA,NB,NC)
       do 25 i = 1, NA
        do 25 j= 1, NB
   25    if (FHAB(i,j).ne.0)
     +    uAB(i,j) = uAB(i,j) + (FAB(i,j)-FHAB(i,j))/FHAB(i,j)
      endif
c                                                              up-date uC(k)
      call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +         NA,NB,NC,Ndim,Ffit)
      call CMARG(FHAT,FHC,NA,NB,NC)
      do 30 k = 1, NC
   30   uC(k) = uC(k) + (FC(k) - FHC(k))/FHC(k)
      return
      end
C
C Up-date category scale values
C
      subroutine SCVFIT(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +NA,NB,NC,Ndim,Ffit,F,FHAT)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension uA(Mn),uB(Mn),uC(Mn),ETA(Mn,Mdim),F(Mn,Mn,Mn),
     +FHAT(Mn,Mn,Mn),phiAB(Mdim),phiAC(Mdim),phiBC(Mdim),uAB(Mn,Mn),
     +Ffit(Mn,Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
```

```
      data ZERO / 0.0 /
c                                                     Loop through dimensions m
      do 100 m = 1, Ndim
c                                                                  up-date mu(i,m)
      call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +         NA,NB,NC,Ndim,Ffit)
      do 30 i = 1, NA
        X = ZERO
        XX = ZERO
        do 10 j = 1, NB
         do 10 k = 1, NC
            X = X + (phiAB(m)*NU(j,m) + phiAC(m)*ETA(k,m))
     +          *(F(i,j,k) - FHAT(i,j,k))
           XX = XX + (phiAB(m)*NU(j,m) + phiAC(m)*ETA(k,m))
     +         * (phiAB(m)*NU(j,m) + phiAC(m)*ETA(k,m))*FHAT(i,j,k)
   10  continue
       MU(i,m) = MU(i,m) + X/XX
   30 continue
c                                                                   up-date nu(j,m)
      call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +         NA,NB,NC,Ndim,Ffit)
      do 60 j = 1, NB
       X = ZERO
       XX = ZERO
       do 40 i = 1, NA
        do 40 k = 1, NC
         X = X + (phiAB(m)*MU(i,m) + phiBC(m)*ETA(k,m))
     +         *(F(i,j,k)-FHAT(i,j,k))
        XX = XX + (phiAB(m)*MU(i,m) + phiBC(m)*ETA(k,m))
     +          *(phiAB(m)*MU(i,m) + phiBC(m)*ETA(k,m))*FHAT(i,j,k)
   40  continue
       NU(j,m) = NU(j,m) + X/XX
   60 continue
c                                                                 up-date eta(k,m)
      call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +         NA,NB,NC,Ndim,Ffit)
      do 90 k = 1, NC
        X = ZERO
        XX = ZERO
        do 70 i = 1, NA
         do 70 j = 1, NB
          X = X + (phiAC(m)*MU(i,m) + phiBC(m)*NU(j,m))
     +          *(F(i,j,k) - FHAT(i,j,k))
         XX = XX + (phiAC(m)*MU(i,m) + phiBC(m)*NU(j,m))
     +           *(phiAC(m)*MU(i,m) + phiBC(m)*NU(j,m))*FHAT(i,j,k)
   70   continue
       ETA(k,m) = ETA(k,m) + X/XX
   90 continue
  100 continue
      return
      end
C
C Up-date phi parameters
C
      subroutine PHIFIT(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +NA,NB,NC,Ndim,IuAB,Ffit,F,FAB,FAC,FBC,FHAT,FHAB,FHAC,FHBC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension uA(Mn),uB(Mn),uC(Mn),ETA(Mn,Mdim),F(Mn,Mn,Mn),
     +FAB(Mn,Mn),FAC(Mn,Mn),FBC(Mn,Mn),FHAT(Mn,Mn,Mn),FHAB(Mn,Mn),
     +FHAC(Mn,Mn),FHBC(Mn,Mn),phiAB(Mdim),phiAC(Mdim),phiBC(Mdim),
```

```
     +uAB(Mn,Mn),Ffit(Mn,Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      data ZERO / 0.0 /
c                                                              update phiAB(m)
      if (IuAB.eq.0) then
      do 10 m=1, Ndim
       call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +          NA,NB,NC,Ndim,Ffit)
       call ABMARG(FHAT,FHAB,NA,NB,NC)
       X = ZERO
       XX = ZERO
       do 5 i = 1, NA
         do 5 j = 1, NB
           X = X + MU(i,m)*NU(j,m)*(FAB(i,j)-FHAB(i,j))
           XX = XX + MU(i,m)*NU(j,m)*MU(i,m)*NU(j,m)*FHAB(i,j)
    5  continue
       phiAB(m) = phiAB(m) + X/XX
   10 continue
      endif
c
      do 100 m = 1, Ndim
c                                                              update phiAC(m)
      call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +         NA,NB,NC,Ndim,Ffit)
      call ACMARG(FHAT,FHAC,NA,NB,NC)
      X = ZERO
      XX = ZERO
      do 30 i = 1, NA
        do 30 k = 1, NC
          X  = X + MU(i,m)*ETA(k,m)*(FAC(i,k)-FHAC(i,k))
          XX = XX + MU(i,m)*ETA(k,m)*MU(i,m)*ETA(k,m)*FHAC(i,k)
   30 continue
      phiAC(m) = phiAC(m) + X/XX
c                                                              update phiBC(m)
      call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +         NA,NB,NC,Ndim,Ffit)
      call BCMARG(FHAT,FHBC,NA,NB,NC)
      X = ZERO
      XX = ZERO
      do 50 j = 1, NB
        do 50 k = 1, NC
          X = X + NU(j,m)*ETA(k,m)*(FBC(j,k)-FHBC(j,k))
          XX = XX + NU(j,m)*ETA(k,m)*NU(j,m)*ETA(k,m)*FHBC(j,k)
   50 continue
      phiBC(m) = phiBC(m) + X/XX
  100 continue
      return
      end
C
C  ......The Main Event......
C  Multi-dimensional, no 3-way association models with consistency
C   constraints on the scale values.
C
      subroutine MODEL(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +NA,NB,NC,Ndim,IuAB,Ffit,F,FAB,FAC,FBC,FA,FB,FC,FHAT,FHAB,FHAC,
     +FHBC,FHA,FHB,FHC,WA,WB,WC,Iorth)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FAB(Mn,Mn),FAC(Mn,Mn),FBC(Mn,Mn),
     +FA(Mn),FB(Mn),FC(Mn),WA(Mn),WB(Mn),WC(Mn),FHAT(Mn,Mn,Mn),FHA(Mn),
     +FHB(Mn),FHC(Mn),FHAB(Mn,Mn),FHAC(Mn,Mn),FHBC(Mn,Mn),
```

```
     +uA(Mn),uB(Mn),uC(Mn),ETA(Mn,Mdim),phiAB(Mdim),phiAC(Mdim),
     +phiBC(Mdim),uAB(Mn,Mn),Ffit(Mn,Mn,Mn),FHold(Mn,Mn,Mn)
      double precision MU(Mn,Mdim),NU(Mn,Mdim)
      common / iter / CONV,DIFF,IMAX,ICALLS,ICONV,IHIST
      common / nunit / nuniti, nunito
      ICALLS = 0
      Ncycle = 1
c                                                          Output History?
      if (IHIST.eq.1) write(nunito,400)
c                                                     Initial fitted values
      call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +         NA,NB,NC,NDim,Ffit)
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
          FHold(i,j,k) = FHAT(i,j,k)
   10 continue
c                                                            The main loop
      do 1000 loop = 1, IMAX
c                                                          up-date u-terms
c
      call UFIT(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,NA,NB,NC,
     +         NDim,IuAB,Ffit,F,FA,FB,FC,FAB,FHAT,FHA,FHB,FHC,FHAB)
      if (Ndim.gt.0) then
c                                           up-date category scale values
c
       call SCVFIT(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,NA,NB,NC,
     +           NDim,Ffit,F,FHAT)
c                                                            up-date phi's
c
       call PHIFIT(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,NA,NB,NC,
     +          NDim,IuAB,Ffit,F,FAB,FAC,FBC,FHAT,FHAB,FHAC,FHBC)
      endif
c                                                       center new estimates
c
      call CENTER(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,NA,NB,NC,
     +          NDim,IuAB,WA,WB,WC)
c                                                        scale new estimates
      if (Ndim.gt.0) then
        call SCALE(Iorth,MU,phiAB,phiAC,WA,NA,NDim)
        call SCALE(Iorth,NU,phiAB,phiBC,WB,NB,NDim)
        call SCALE(Iorth,Eta,phiAC,phiBC,WC,NC,NDim)
      endif
      ICALLS = ICALLS + 1
c                                                          Check Convergence
c
      call PRED(F,FHAT,u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +         NA,NB,NC,NDim,Ffit)
      call CHECK(F,FHAT,FHold,NA,NB,NC)
c                                               ICONV = 1  ---> converged
      if (ICONV.EQ.1)  return
c                                                                  History
      if (IHIST.eq.1) then
        if (Ncycle.lt.10) then
          Ncycle = Ncycle + 1
        else
          call GSQUAR(F,FHAT,GSQ,NA,NB,NC)
          call PEARSO(F,FHAT,XSQ,NA,NB,NC)
          write(nunito,500) loop, DIFF, GSQ, XSQ
          Ncycle = 1
        endif
```

```
      endif
1000 continue
      return
  400 format('1History of Maximum Likelihood Estimation:',/,
     +1x,'Cycle:',8x,'DIFF',15x,'G**2',15x,'X**2')
  500 format(1x,i6,3(3x,f15.8))
      end
C
C  Check convergence
C
      subroutine CHECK(F,FHAT,FHold,NA,NB,NC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn),FHold(Mn,Mn,Mn)
      common / iter / CONV,DIFF,IMAX,ICALLS,ICONV,IHIST
      data ZERO / 0.0 /
      DIFF = ZERO
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
   10     DIFF = DIFF + DABS(FHAT(i,j,k) - FHold(i,j,k))
c
      if (DIFF .lt.  CONV) then
        ICONV = 1
      else
        do 20 i = 1, NA
         do 20 j = 1, NB
          do 20 k = 1, NC
   20       FHold(i,j,k) = FHAT(i,j,k)
      endif
      return
      end
C
C  Center category scale values and adjust main effect parameters
C   then center main effects parameters
C
      subroutine CENTER(u,uA,uB,uC,MU,NU,ETA,phiAB,phiAC,phiBC,uAB,
     +NA,NB,NC,Ndim,IuAB,WA,WB,WC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension uA(Mn),uB(Mn),uC(Mn),ETA(Mn,Mdim),WA(Mn),WB(Mn),WC(Mn),
     +phiAB(Mdim),phiAC(Mdim),phiBC(Mdim),uAB(Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      data ZERO / 0.0 /
      if (Ndim.gt.0)  then
      call SUM1W(WA,NA,A)
      call SUM1W(WB,NB,B)
      call SUM1W(WC,NC,C)
      do 35 m = 1, Ndim
C                                          Center scale values
       X = ZERO
       Y = ZERO
       Z = ZERO
       do 10 i = 1, NA
   10    X = X + MU(i,m)*WA(i)
       do 20 j = 1, NB
   20    Y = Y + NU(j,m)*WB(j)
       do 30 k = 1, NC
   30    Z = Z + ETA(k,m)*WC(k)
       XA = X/A
       YB = Y/B
```

```
       ZC = Z/C
       do 31 i = 1, NA
         uA(i) = uA(i) + (phiAB(m)*YB + phiAC(m)*ZC)*MU(i,m)
         MU(i,m) = MU(i,m) - XA
   31  continue
       do 32 j = 1, NB
        uB(j) = uB(j) + (phiAB(m)*XA + phiBC(m)*ZC)*NU(j,m)
        NU(j,m) = NU(j,m) - YB
   32  continue
       do 33 k = 1, NC
        uC(k) = uC(k) + (phiAC(m)*XA + phiBC(m)*YB)*ETA(k,m)
        ETA(k,m) = ETA(k,m) - ZC
   33  continue
       u = u - phiAB(m)*XA*YB - phiAC(m)*XA*ZC - phiBC(m)*YB*ZC
   35 continue
      endif
C                                                          Center u-terms
      if (IuAB.eq.1) call DBLCEN(uAB,uA,uB,u,NA,NB)
      call CEMAIN(uA,u,NA)
      call CEMAIN(uB,u,NB)
      call CEMAIN(uC,u,NC)
      return
      end
C
C CEnter MAIN effect around 0 and adjust constant parameter of model
C
      subroutine CEMAIN(X1,X,Nrow)
      parameter (Mn=20,Mdim=19)
      implicit double precision (a-h,o-z), integer*4 (i-n)
      dimension X1(Mn)
      Xbar = 0
      call SUM1W(X1,Nrow,Xbar)
      Xbar = Xbar/DBLE(Nrow)
      do 10 i = 1, Nrow
   10   X1(i) = X1(i) - Xbar
      X = X + Xbar
      return
      end
C
C Double Center matrix of interaction terms and adjust main & constant terms
C
      subroutine DBLCEN(u12,u1,u2,u,N1,N2)
      parameter (Mn=20,Mdim=19)
      implicit double precision (a-h,o-z), integer*4 (i-n)
      dimension u12(Mn,Mn), u1(Mn), u2(Mn), xbar1(Mn), xbar2(Mn)
      x1 = DBLE(N1)
      x2 = DBLE(N2)
      grand = 0.0
      call ZERO1W(xbar1,N1)
      call ZERO1W(xbar2,N2)
      do 10 i = 1, N1
      do 10 j = 1, N2
         grand = grand + u12(i,j)
         xbar1(i) = xbar1(i) + u12(i,j)
         xbar2(j) = xbar2(j) + u12(i,j)
   10 continue
      grand = grand/(x1*x2)
      call DivByX(x2,xbar1,xbar1,N1)
      call DivByX(x1,xbar2,xbar2,N2)
      do 20 i = 1, N1
       do 20 j = 1, N2
```

```
         u12(i,j) = u12(i,j) - xbar1(i) - xbar2(j) + grand
   20 continue
      do 30 i = 1, N1
         u1(i) = u1(i) + xbar1(i)
   30 continue
      do 40 j = 1, N2
         u2(j) = u2(j) + xbar2(j)
   40 continue
      u = u - grand
      return
      end
C
C Orthonormalize scale values via Grahm-Schmidt process such that
C  weighted sum of squares = 1 and weighted sum of crossproducts = 0.
C
      subroutine SCALE(Iorth,S,phi1,phi2,W,N,NDim)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4(I-N)
      dimension S(Mn,Mdim), phi1(Mdim), phi2(Mdim), W(Mn), XX(Mdim)
      data ZERO / 0.0 /
c                                                              1st dimension
      X = ZERO
      do 10 i = 1, N
        X = X + S(i,1)*S(i,1)*W(i)
   10 continue
      X = dsqrt(X)
      do 20 i = 1, N
        S(i,1) = S(i,1)/X
   20 continue
      phi1(1) = phi1(1)*X
      phi2(1) = phi2(1)*X
c                          if only want sum of squares = 1 (not orthogonal)
      if (NDim.lt.2) return
      if (Iorth.eq.0) then
         do 26 m = 2, NDim
           X = ZERO
          do 24 i = 1, N
   24       X = X + S(i,m)*S(i,m)*W(i)
          X = dsqrt(X)
          do 25 i = 1, N
   25       S(i,m) = S(i,m)/X
          phi1(m) = phi1(m)*X
          phi2(m) = phi2(m)*X
   26    continue
         return
c                                 Scale & orthogonalize remaining dimensions
      else
       do 100 m = 2, Ndim
        Last = m - 1
        do 40 L = 1, Last
          XX(L) = ZERO
          do 30 i = 1, N
   30       XX(L) = XX(L) + S(i,L)*S(i,m)*W(i)        !  weighted inner products
   40   continue
        do 60 L = 1, Last
          do 50 i = 1, N
   50       S(i,m) = S(i,m) - XX(L)*S(i,L)               ! vector - projection
   60   continue
        X = ZERO                                  ! adjust vector length & phi's
        do 70 i = 1, N
   70     X = X + S(i,m)*S(i,m)*W(i)
```

```
      X = dsqrt(X)
      do 80 i = 1, N
   80   S(i,m) = S(i,m)/X
      phi1(m) = phi1(m)*X
      phi2(m) = phi2(m)*X
  100 continue
    endif
    return
    end
C
C Compute likelihood ratio statistic:  G**2
C
    subroutine GSQUAR(F,FHAT,GSQ,NA,NB,NC)
    parameter (Mn=20,Mdim=19)
    implicit double precision (A-H,O-Z)
    dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn)
    integer NA,NB,NC
    data ZERO / 0.0 /
    GSQ = ZERO
    do 10 i = 1, NA
     do 10 j = 1, NB
      do 10 k = 1, NC
       if ( (F(i,j,k).eq.ZERO).or.(FHAT(i,j,k).eq.Zero) ) goto 10
       GSQ = GSQ + F(i,j,k) * dlog( F(i,j,k)/FHAT(i,j,k) )
   10 continue
    GSQ = 2*GSQ
    return
    end
C
C  Compute Pearson chi-square:   X**2
C
    subroutine PEARSO(F,FHAT,XSQ,NA,NB,NC)
    parameter (Mn=20,Mdim=19)
    implicit double precision (A-H,O-Z)
    dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn)
    integer NA,NB,NC
    data ZERO / 0.0 /
    XSQ = ZERO
    do 10 i = 1, NA
     do 10 j = 1, NB
      do 10 k = 1, NC
       if (FHAT(i,j,k).ne.ZERO)  XSQ = XSQ +
     +      (F(i,j,k)-FHAT(i,j,k))*(F(i,j,k)-FHAT(i,j,k))/FHAT(i,j,k)
   10 continue
    return
    end
C
C Compute the p-value of a chi-square statistic
C
    subroutine CHI(XX,N,pval)
    implicit double precision (a-h,o-z)
    z = dexp(-XX/2)
    m = n/2
    m = m+m
    if(m.eq.n) go to 10
c Odd degrees of freedom.
    prod = 0.79788456*z
    t = 1/(1+0.2316419*dsqrt(XX))
    pval = prod*t*(0.31938153+t*(-0.356563782+t*(1.781477937
   *   +t*(-1.821255978+1.330274429*t))))
    m = n-2
```

```
      prod = prod/dsqrt(XX)
      do 2 j=1,m,2
      prod = prod*XX/j
    2 pval = pval+prod
      go to 20
c Even degrees of freedom
   10 pval = z
      m = n-2
      if(n.le.2) go to 20
      prod = z
      do 12 j=2,m,2
      prod = prod*XX/j
   12 pval = pval+prod
   20 return
      end
C
C Computes the Pearson correlation between 2 variables where data is grouped
C
      subroutine CORR(FRC,Ftot,X1,X2,rho,NR,NC,m)
      parameter (Mn=20,Mdim=19)
      implicit double precision (a-h,o-z), integer*4 (i-n)
      dimension FRC(Mn,Mn), X1(Mn,Mdim), X2(Mn,Mdim)
      rho = 0.0
      do 10 i = 1, NR
       do 10 j = 1, NC
         rho = rho + X1(i,m)*X2(j,m)*FRC(i,j)
   10 continue
      rho = rho/Ftot
      return
      end
C
C  Output 1-way Table in vector format
C
      subroutine OUT1W(XTOT,X,HX,Chi1,NX)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z)
      dimension X(Mn),HX(Mn),Chi1(Mn)
      integer NX
      common / nunit / nuniti, nunito
      do 10 i = 1, NX
   10  write(nunito,100)
     +     i, X(i)/XTOT, X(i), HX(i), X(i)-HX(i), Chi1(i)
      return
  100 format(2x,i2,5(3x,f10.5))
      end
C
C  Output 2-way Table in vector format
C
      subroutine OUT2W(XTOT,XY,HXY,Chi2,NX,NY)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension XY(Mn,Mn),HXY(Mn,Mn),Chi2(Mn,Mn)
      common / nunit / nuniti, nunito
      do 10 i = 1, NX
       do 10 j = 1, NY
   10    write(nunito,100) i,j,XY(i,j)/XTOT,XY(i,j),HXY(i,j),
     +    XY(i,j)-HXY(i,j), Chi2(i,j)
      return
  100 format(2(2x,i2),5(3x,f10.5))
      end
C
```

```
C  Output 3-way Table
C
      subroutine OUT3W(FTOT,F,FHAT,Ffit,NA,NB,NC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn),SRes(Mn,Mn,Mn),
     + Ffit(Mn,Mn,Mn)
      common / nunit / nuniti, nunito
      write(nunito,100)
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
        resid =  F(i,j,k)-FHAT(i,j,k)
        write(nunito,110) i, j, k, F(i,j,k)/FTOT,F(i,j,k), FHAT(i,j,k),
     +   resid, resid/dsqrt(FHAT(i,j,k)), Ffit(i,j,k)
   10 continue
      return
  100 format(/,1x,'Observed, Fitted & Residual Frequencies:',/,
     +'   I   J   K   P(ijk)     F(ijk)     FHAT(ijk)   Residual',
     +'    Std Resid  F fit?')
  110 format(3(2x,i2),5(2x,f10.5),2x,f2.0)
      end
C
C  Output Model Parameters:   Constant and Main Effects
C
      subroutine OUmain(u,uA,uB,uC,NA,NB,NC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension uA(Mn),uB(Mn),uC(Mn)
      common / nunit / nuniti, nunito
      write(nunito,100) u
      write(nunito,110)
      do 10 i = 1, NA
   10   write(nunito,111) i, uA(i)
      write(nunito,120)
      do 20 j = 1, NB
   20   write(nunito,121) j, uB(j)
      write(nunito,130)
      do 30 k = 1, NC
   30   write(nunito,131) k, uC(k)
      return
  100 format(/,1x,'Constant:  ',f10.5)
  110 format(//,1x,'Main Effects for Rows:',/)
  111 format(1x,'uA(',i2,') = ',f15.5)
  120 format(//,1x,'Main Effects for Columns:',/)
  121 format(1x,'uB(',i2,') = ',f15.5)
  130 format(//,1x,'Main Effects for Layers:',/)
  131 format(1x,'uC(',i2,') = ',f15.5)
      end
C
C Output Model Parameters:  Association Model Interactions effects
C
      subroutine OUTint(MU,NU,ETA,phiAB,phiAC,phiBC,uAB,NA,NB,NC,
     +Ndim,IuAB,WA,WB,WC)
      parameter (Mn=20,Mdim=19)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension ETA(Mn,Mdim),WA(Mn),WB(Mn),WC(Mn),phiAB(Mdim),
     + phiAC(Mdim),phiBC(Mdim),uAB(Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      common / nunit / nuniti, nunito
      if (IuAB.eq.1) then
```

```
     write(nunito,105)
     write(nunito,106) (j,j=1,NB)
     do 5 i = 1, NA
 5     write(nunito,107) i, (uAB(i,j),j=1,NB)
   endif
   if (Ndim.gt.0) then
   write(nunito,110)
   write(nunito,101) (m,m=1,Ndim)
   do 10 i = 1, NA
10  write(nunito,111) i, (MU(i,m),m=1,Ndim), WA(i)
   write(nunito,120)
   write(nunito,101) (m,m=1,Ndim)
   do 20 j = 1, NB
20  write(nunito,121) j, (NU(j,m),m=1,Ndim), WB(j)
   write(nunito,130)
   write(nunito,101) (m,m=1,Ndim)
   do 30 k = 1, NC
30  write(nunito,131) k, (ETA(k,m),m=1,Ndim) , WC(k)
   write(nunito,140)
   write(nunito,101) (m,m=1,Ndim)
   if (IuAB.eq.0)  write(nunito,141) (phiAB(m),m=1,Ndim)
   write(nunito,151) (phiAC(m),m=1,Ndim)
   write(nunito,161) (phiBC(m),m=1,Ndim)
   endif
   return
101 format(9x,'m = ',12(4x,i2,4x))
105 format(/,1x,'AB interaction:  uAB(i,j)')
106 format(6x,'j = ',10(5x,i2,5x),/,10(5x,i2,5x))
107 format(1x,'i = ',i2,3x,10(2x,f10.5),/,10(2x,f10.5))
110 format(//,1x,'Row Scale Values and Weights:  ',/)
111 format(1x,'Mu(',i2,',m) = ',3x,12(2x,f8.5))
120 format(//,1x,'Column Scale Values and Weights:',/)
121 format(1x,'Nu(',i2,',m) = ',3x,12(2x,f8.5))
130 format(//,1x,'Layer Scale Values and Weights:',/)
131 format(1x,'Eta(',i2,',m) = ',1x,12(2x,f8.5))
140 format(//,1x,'Measures of association:')
141 format(5x,'phiAB = ',12(2x,f8.5))
151 format(5x,'phiAC = ',12(2x,f8.5))
161 format(5x,'phiBC = ',12(2x,f8.5))
   end
```

## C.2 Log 3-Mode Models

```
C log3mode.f
C
C  written by Carolyn J. Anderson
C  last revised:  July 15, 1992
C
C  Fits and estimates the parameters of log 3--mode association models
C   where some combination of the 2- and 3-way interactions/associations are
C   decomposed using Tucker's 3-mode principle components.
C
C  Either (i)  all 2-way and 3-way interactions are decomposed or
C         (ii) the AC, BC, and ABC associations are decomposed while the AB
C              margin is fit.
C
C   IuAB  = 0  option (i) --->  fix uAB(i,j) = 0
C         = 1  option (ii) -->  estimate uAB(i,j)
C
C
C
C  Uni-dimensional Newton-Raphson is used to find MLE of model parameters.
C
C   Mn = Max(NA, NB, NC)       = Max Number of levels/variable
C   Mdim = Min(NA, NB, NC) - 1  = Max Number of dimensions
C
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FAB(Mn,Mn),FAC(Mn,Mn),FBC(Mn,Mn),FA(Mn),
     +FB(Mn),FC(Mn),WA(Mn),WB(Mn),WC(Mn),FHAT(Mn,Mn,Mn),FHA(Mn),
     +FHB(Mn),FHC(Mn),FHAB(Mn,Mn),FHAC(Mn,Mn),FHBC(Mn,Mn),uA(Mn),
     +uB(Mn),uC(Mn),uAB(Mn,Mn),ETA(Mn,Mdim),phi(Mdim,Mdim,Mdim),
     +Ffit(Mn,Mn,Mn),ChiSq(Mn,Mn,Mn),Chi1(Mn),Chi2(Mn,Mn),DP(Mn,Mn)
      double precision MU(Mn,Mdim),NU(Mn,Mdim)
      character*30 iname, oname
      character*131 title
      common / iter / CONV,DIFF,IMAX,ICALLS,ICONV,IHIST
      common / nunit / nuniti, nunito
      none = 1
      nfour = 4
      zero = 0.0
      nuniti = 3
      nunito = 4
    1 write(*,2)
    2 format('Enter name of input file:  ')
      read(*,600,end=1) iname
      open(unit=nuniti,file=iname,status='OLD')
    3 write(*,4)
    4 format('Enter name of output file:  ')
      read(*,600,end=3) oname
      open(unit=nunito,file=oname,status='NEW')
c                                                  Problem specifications
c                                                                  Input
      read(nuniti,610) title
      read(nuniti,*)  NA, NB, NC, MA, MB, MC, delta, IuAB,
     +             inWA, inWB, inWC, IMAX, CONV, IHIST
c                                                                 Output
      write(nunito,620) title
      write(nunito,300) NA, MA, NB, MB, NC, MC, delta
      if (IuAB.eq.1) write(nunito,301) IuAB
      write(nunito,310) inWA,inWB,inWC
c                                                  Read in data  and Ffit
```

```
c                                                       fastest = variable C
c                                                       slowest = variable A
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
         read(nuniti,*) F(i,j,k), Ffit(i,j,k)
         F(i,j,k) = F(i,j,k) + delta
   10  continue
c                                                  Compute observed margins
      call AMARG(F,FA,NA,NB,NC)
      call BMARG(F,FB,NA,NB,NC)
      call CMARG(F,FC,NA,NB,NC)
      call ABMARG(F,FAB,NA,NB,NC)
      call ACMARG(F,FAC,NA,NB,NC)
      call BCMARG(F,FBC,NA,NB,NC)
      call SUM1W(FA,NA,FTOT)
      if (MA.ne.0) then
c                              Weights:  input, unit or marginal probabilties
c                                inW =  0    1      2
       if (inWA.eq.0) then
         read(nuniti,*) (WA(i),i=1,NA)
       else
         if (inWA.eq.1) then
           call ONE1W(WA,NA)
         else
           call DivByX(FTOT,FA,WA,NA)
         endif
       endif
       if (inWB.eq.0) then
         read(nuniti,*) (WB(j),j=1,NB)
       else
          if (inWB.eq.1) then
           call ONE1W(WB,NB)
         else
           call DivByX(FTOT,FB,WB,NB)
         endif
       endif
       if (inWC.eq.0) then
         read(nuniti,*) (WC(k),k=1,NC)
       else
         if (inWC.eq.1) then
           call ONE1W(WC,NC)
         else
           call DivByX(FTOT,FC,WC,NC)
         endif
       endif
      endif
c                                                          Starting values
c
      read(nuniti,*) u,(uA(i),i=1,NA),(uB(j),j=1,NB),(uC(k),k=1,NC)
      if (IuAB.eq.1) then
        do 20 i = 1, NA
   20     read(nuniti,*) (uAB(i,j),j=1,NB)
      else
        call ZERO2W(uAB,NA,NB)
      endif
      if (MA.gt.0) then
        do m1 = 1, MA
          read(nuniti,*)  (MU(i,m1),i=1,NA)
        end do
        do m2 = 1, MB
```

```
        read(Nuniti,*)  (NU(j,m2),j=1,NB)
       end do
       do m3 = 1, MC
        read(Nuniti,*)  (ETA(k,m3),k=1,NC)
       end do
       do 110 m1 = 1, MA
        do 110 m2 = 1, MB
 110      read(nuniti,*)  (phi(m1,m2,m3),m3=1,MC)
     else
       call ZERO2W(MU,NA,MA)
       call ZERO2W(NU,NB,MB)
       call ZERO2W(ETA,NC,MC)
       do 120 m1 = 1, MA
       do 120 m2 = 1, MB
       do 120 m3 = 1, MC
 120    phi(m1,m2,m3) = 0.0
     endif
c                                                    Output starting values
     write(nunito,620) title
     write(nunito,320)
     call OUmain(u,uA,uB,uC,NA,NB,NC)
     call OUTint(uAB,MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,IuAB,WA,WB,WC)
c                                       ......Heart of the program.......
c
     ICONV = 0
     call MODEL(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,
    +          F,FAB,FAC,FBC,FA,FB,FC,FHAT,FHAB,FHAC,FHBC,FHA,
    +          FHB,FHC,WA,WB,WC,Ffit,IuAB)
c
     call ABMARG(FHAT,FHAB,NA,NB,NC)
     call ACMARG(FHAT,FHAC,NA,NB,NC)
     call BCMARG(FHAT,FHBC,NA,NB,NC)
     call AMARG(FHAT,FHA,NA,NB,NC)
     call BMARG(FHAT,FHB,NA,NB,NC)
     call CMARG(FHAT,FHC,NA,NB,NC)
c                                                        Output results
c                                     Convergence message & fit statistics
     if (ICONV.EQ.1) then
       write(nunito,350)
       write(*,350)
     else
       write(nunito,355)
       write(*,355)
     endif
     write(nunito,630)
     write(nunito,621) title
     write(nunito,410) ICALLS, IMAX, DIFF, CONV
c                                                    Compute ChiSq-terms
     do 210 i = 1, NA
     do 210 j = 1, NB
     do 210 k = 1, NC
       if (FHAT(i,j,k).ne.ZERO)   ChiSq(i,j,k) =
    +    (F(i,j,k)-FHAT(i,j,k))*(F(i,j,k)-FHAT(i,j,k))/FHAT(i,j,k)
 210 continue
c                                                 Compute degrees of freedom
c                        Number where fit=observed (e.g., structural zeros)
     Nless = 0
     do 220 i = 1, NA
      do 220 j = 1, NB
       do 220 k = 1, NC
 220     Nless = Nless + 1 - INT(Ffit(i,j,k))
```

```
c                        Number of unique phi parameters estimated
     if (MA.eq.1) then
Nphi = min0(MB,MC)
     else
if (MB.eq.1) then
  Nphi = min0(MA,MC)
       else
  if (MC.eq.1) then
    Nphi = min0(MA,MB)
  else
           Nphi = MA*MB*MC
         endif
       endif
     endif
     ndf = NA*NB*NC +2 -NA*(MA+1) -NB*(MB+1) -NC*(MC+1) -Nphi -Nless
    *   - IuAB*((NA-1)*(NB-1)-MC) +(MA*(MA+1)+MB*(MB+1)+MC*(MC+1))/2
     write(nunito,420)
     write(nunito,430) ndf, Nless
     call GSQUAR(F,FHAT,GSQ,NA,NB,NC)
     call CHI(GSQ,ndf,pvalG)
     write(nunito,440) GSQ, pvalG
     call PEARSO(F,FHAT,XSQ,NA,NB,NC)
     call CHI(XSQ,ndf,pvalX)
     write(nunito,450) XSQ, pvalX
c                                    Observed data, fitted values, & residuals
     write(nunito,620) title
     write(nunito,500) FTOT
c                                                          1-way margins
     write(nunito,510)
     call AMARG(ChiSq,Chi1,NA,NB,NC)
     call OUT1W(FTOT,FA,FHA,Chi1,NA)
     write(nunito,520)
     call BMARG(ChiSq,Chi1,NA,NB,NC)
     call OUT1W(FTOT,FB,FHB,Chi1,NB)
     write(nunito,530)
     call CMARG(ChiSq,Chi1,NA,NB,NC)
     call OUT1W(FTOT,FC,FHC,Chi1,NC)
c                                                          2-way margins
     write(nunito,620) title
     write(nunito,540)
     call ABMARG(ChiSq,Chi2,NA,NB,NC)
     call OUT2W(FTOT,FAB,FHAB,Chi2,NA,NB)
     write(nunito,550)
     call ACMARG(ChiSq,Chi2,NA,NB,NC)
     call OUT2W(FTOT,FAC,FHAC,Chi2,NA,NC)
     write(nunito,560)
     call BCMARG(ChiSq,Chi2,NA,NB,NC)
     call OUT2W(FTOT,FBC,FHBC,Chi2,NB,NC)
c                                                            3-way table
     write(nunito,620) title
     call OUT3W(FTOT,F,FHAT,Ffit,NA,NB,NC)
c                                                     Parameter estimates
     write(nunito,620) title
     write(nunito,570)
     call OUmain(u,uA,uB,uC,NA,NB,NC)
     call OUTint(uAB,MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,IuAB,WA,WB,WC)
c                                                    Inner (dot) Products
     write(nunito,700)
c                                                    A x B | C
     do 250 m3 = 1, mC
      do 230 i = 1, NA
```

```
      do 230 j = 1, NB
DP(i,j) = zero
do 225 m1 = 1, MA
do 225 m2 = 1, MB
  225     DP(i,j) = DP(i,j) + Mu(i,m1)*phi(m1,m2,m3)*Nu(j,m2)
  230  continue
       write(nunito,701) m3
       write(nunito,710) (j,j=1,NB)
       do 240 i = 1, NA
  240    write(nunito,720) i,(DP(i,j),j=1,NB)
  250 continue
c                                                    A x C | B
     do 270 m2 = 1, MB
do 265 i = 1, NA
do 265 k = 1, NC
  DP(i,k) = zero
         do 260 m1 = 1, MA
  do 260 m3 = 1, MC
  260      DP(i,k) = DP(i,k) + Mu(i,m1)*phi(m1,m2,m3)*Eta(k,m3)
  265   continue
write(nunito,702) m2
write(nunito,710) (k,k=1,NC)
do 267 i = 1, NA
  267    write(nunito,720) i,(DP(i,k),k=1,NC)
  270 continue
c                                                    B x C | A
     do 290 m1 = 1, MA
do 280 j = 1, NB
do 280 k = 1, NC
  DP(j,k) = zero
  do 275 m2 = 1, MB
  do 275 m3 = 1, MC
  275      DP(j,k) = DP(j,k) + Nu(j,m2)*phi(m1,m2,m3)*Eta(k,m3)
  280   continue
write(nunito,703) m1
write(nunito,710) (k,k=1,NC)
do 285 j = 1, NB
  285    write(nunito,720) j,(DP(j,k),k=1,NC)
  290 continue
c
     close(3,status='keep')
     close(4,status='keep')
     stop
c                                                      Format Statements
  300 format(1x,'Problem Specifications for Log 3-mode multi-linear',
    +' association model',/,
    +/,' Number of rows    = ',i2,5x,'Number of dimensions = ',i2,
    +/,' Number of columns = ',i2,5x,'Number of dimensions = ',i2,
    +/,' Number of layers  = ',i2,5x,'Number of dimensions = ',i2,//,
    +//,' Delta = ',E15.5,' was added to each cell')
  301 format(/,1x,'IuAB = ',i1,' ----> uAB(i,j) estimated ')
  310 format(/,1x,'Weights for scale values are ',
    +'inWA = ',i1,', inWB = ',i1,', and inWC = ',i1,/,
    +1x,'where inW = 0  --> input',/,7x,'inW = 1  ---> unit',/,
    +7x,'inW = 2  --> marginal probabilties',/ )
  320 format(1x,'Starting Values:  ')
  350 format(/,1x,'Solution converged ',/)
  355 format(/,1x,'Warning:  solution did not converge for specified',
    +' criterion ',/)
  410 format(1x,'Summary of Convergence Information:',//,
    +1x,'Number of cycles:',10x,'ICALLS = ',i6,/,
```

```
     +1x,'Maximum number of cycles:',4x,'IMAX = ',i6,/,
     +1x,'Change in fitted values:',2x,'DIFF = ',f15.10,/,
     +1x,'Convergence critertion:',3x,'CONV = ',f15.10,/)
      write(nunito,420)
  420 format(/,' Fit Statistics:  ',/)
  430 format(1x,'Degrees of freedom = ',i5,5x,'Note:  ',i2,
     +' cells where fitted = observed')
  440 format(/,1x,'Likelihood ratio statistic,  G**2 = ',f15.4,
     +5x,'p-value = ',f8.5)
  450 format(1x,'Pearson chi-square statitic, X**2 = ',f15.4,
     +5x,'p-value = ',f8.5,/)
  500 format(1x,'Total number of observations in table = ',f10.2,//)
  510 format(1x,'Row Margin:  ',/,1x,' i',6x,'P(i++)',6x,'F(i++)',
     +6x,'FHAT(i++)',6x,'Residual',4x,'Sum of Chi-terms')
  520 format(//,1x,'Column Margin:  ',/,1x,' j',6x,'P(+j+)',6x,'F(+j+)'
     +,6x,'FHAT(+j+)',6x,'Residual',4x,'Sum of Chi-terms')
  530 format(//,1x,'Layer Margin:  ',/,1x,' k   ',3x,'P(++k)',6x,'F(++k)'
     +,6x,'FHAT(++k)',6x,'Residual',4x,'Sum of Chi-terms')
  540 format(1x,'2-Way margins:',//,1x,'Row x Column:  ', /,1x,
     +' i   j   ',3x,'P(i,j,+)',6x,'F(ij+)',6x,'FHAT(ij+)',5x,
     +'Residual',4x,'Sum of Chi-terms')
  550 format(//,1x,'Row x Layer:  ', /,1x,
     +' i   k   ',3x,'P(i+k)',6x,'F(i+k)',6x,'FHAT(i+k)',5x,
     +'Residual',4x,'Sum of Chi-terms')
  560 format(//,1x,'Column x Layer:  ', /,1x,
     +' j   k   ',3x,'P(+jk)',6x,'F(+jk)',6x,'FHAT(+jk)',5x,
     +'Residual',4x,'Sum of Chi-terms')
  570 format(1x,'Estimated Model Parameters:',/)
  600 format(1a30)
  610 format(1a131)
  620 format('1',1a131,/)
  621 format(1x,1a131,/)
  630 format(/,1x,131('='),/)
  700 format(//,'1--------------- Inner Products --------------',/)
  701 format(/,1x,'Inner Products between Mu and Nu for Eta(',i2,')',/)
  702 format(/,1x,'Inner Products between Mu and Eta for Nu(',i2,')',/)
  703 format(/,1x,'Inner Products between Nu and Eta for Mu(',i2,')',/)
  710 format(1x,5x,12(4x,i2,4x),/,6x,12(4x,i2,4x))
  720 format(1x,i2,3x,12(f8.5,2x),/,6x,12(f8.5,2x))
      end
C
C  Set elements of a One dimensional array = ZERO
C
      subroutine ZERO1W(X,N)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn)
      data ZERO / 0.0 /
      do 10 I = 1, N
   10   X(I) = ZERO
      return
      end
C
C Set elements of a Two dimensional array = ZERO
C
      subroutine ZERO2W(X,NR,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn)
      data ZERO / 0.0 /
      do 10 I = 1, NR
```

```
       do 10 J = 1, NC
   10    X(I,J) = ZERO
      return
      end
C
C Set elements of 1-dimensional array equal to one
C
      subroutine ONE1W(W,N)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension W(Mn)
      do 10 i = 1, N
  10    W(i) = 1.0
      return
      end
C
C Sum the elements of a 1-dimensional array
C
      subroutine SUM1W(X,N,SUM)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn)
      SUM = 0.0
      do 10 i = 1, N
   10   SUM = SUM + X(i)
      return
      end
C
C Divide elements of an array/vector by a constant/scaler
C  -- Use to compute 1-way marginal probabilities or means
C
      subroutine DivByX(X,V,Vx,N)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension V(Mn), Vx(Mn)
      do 10 i = 1, N
   10   Vx(i) = V(i)/X
      return
      end
C                                                          A MARGin
      subroutine AMARG(X,XA,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XA(Mn)
      call ZERO1W(XA,NA)
      do 10 I = 1, NA
       do 10 J = 1, NB
        do 10 K = 1, NC
   10     XA(I) = XA(I) + X(I,J,K)
      return
      end
C                                                          B MARGin
      subroutine BMARG(X,XB,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XB(Mn)
      call ZERO1W(XB,NB)
      do 10 J = 1, NB
       do 10 I = 1, NA
        do 10 K = 1, NC
   10    XB(J) = XB(J) + X(I,J,K)
```

```
      return
      end
C                                                          C MARGin
      subroutine CMARG(X,XC,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XC(Mn)
      call ZERO1W(XC,NC)
      do 10 K = 1, NC
       do 10 I = 1, NA
        do 10 J = 1, NB
   10    XC(K) = XC(K) + X(I,J,K)
      return
      end
C                                                           AB MARGin
      subroutine ABMARG(X,XAB,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XAB(Mn,Mn)
      call ZERO2W(XAB,NA,NB)
      do 10 I = 1, NA
       do 10 J = 1, NB
        do 10 K = 1, NC
   10    XAB(I,J) = XAB(I,J) + X(I,J,K)
      return
      end
C                                                          AC MARGin
      subroutine ACMARG(X,XAC,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XAC(Mn,Mn)
      call ZERO2W(XAC,NA,NC)
      do 10 I = 1, NA
       do 10 J = 1, NB
        do 10 K = 1, NC
   10    XAC(I,K) = XAC(I,K) + X(I,J,K)
      return
      end
C                                                          BC MARGin
      subroutine BCMARG(X,XBC,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension X(Mn,Mn,Mn), XBC(Mn,Mn)
      call ZERO2W(XBC,NB,NC)
      do 10 I = 1, NA
       do 10 J = 1, NB
        do 10 K = 1, NC
   10    XBC(J,K) = XBC(J,K) + X(I,J,K)
      return
      end
C
C  Compute predictions of log 3-mode association model
C
      subroutine PRED(F,X,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     +MA,MB,MC,Ffit)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),X(Mn,Mn,Mn),uA(Mn),uB(Mn),uC(Mn),uAB(Mn,Mn)
     +,ETA(Mn,Mdim),phi(Mdim,Mdim,Mdim),Ffit(Mn,Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      do 40 i = 1, NA
```

```
      A = u + uA(i)
      do 30 j = 1, NB
        AB = A + uB(j) + uAB(i,j)
        do 20 k = 1, NC
          ABC = 0.0
         do 10 m1 = 1, MA
         do 10 m2 = 1, MB
         do 10 m3 = 1, MC
   10      ABC = ABC + phi(m1,m2,m3)*Mu(i,m1)*Nu(j,m2)*Eta(k,m3)
        X(i,j,k) = dexp( AB + uC(k) + ABC)*Ffit(i,j,k)
     +             + F(i,j,k)*(1-Ffit(i,j,k))
   20   continue
   30  continue
   40 continue
      return
      end
C
C Update U-terms:  main effect and/or interaction terms
C
      subroutine UFIT(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     +MA,MB,MC,IuAB,F,FA,FB,FC,FAB,FHAT,FHA,FHB,FHC,FHAB,Ffit)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FA(Mn),FB(Mn),FC(Mn),FAB(Mn,Mn),
     +FHAT(Mn,Mn,Mn),FHA(Mn),FHB(Mn),FHC(Mn),FHAB(Mn,Mn),uA(Mn),
     +uB(Mn),uC(Mn),uAB(Mn,Mn),ETA(Mn,Mdim),phi(Mdim,Mdim,Mdim),
     +Ffit(Mn,Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      if (IuAB.eq.0) then
c                                                          up-date uA(i)
       call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
       call AMARG(FHAT,FHA,NA,NB,NC)
       do 10 i = 1, NA
   10    uA(i) = uA(i) + (FA(i) - FHA(i))/FHA(i)
c                                                          up-date uB(j)
       call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
       call BMARG(FHAT,FHB,NA,NB,NC)
       do 20 j = 1, NB
   20    uB(j) = uB(j) + (FB(j) - FHB(j))/FHB(j)
c                                                       up-date uAB(i,j)
      else
       call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
       call ABMARG(FHAT,FHAB,NA,NB,NC)
       do 25 i = 1, NA
        do 25 j= 1, NB
   25    if (FHAB(i,j).ne.0)
     +     uAB(i,j) = uAB(i,j) + (FAB(i,j)-FHAB(i,j))/FHAB(i,j)
       endif
c                                                          up-date uC(k)
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
      call CMARG(FHAT,FHC,NA,NB,NC)
      do 30 k = 1, NC
   30   uC(k) = uC(k) + (FC(k) - FHC(k))/FHC(k)
      return
      end
C
C Up-date category scale values
```

```
C
      subroutine SCVFIT(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,F,FHAT,Ffit)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn),Ffit(Mn,Mn,Mn),uA(Mn),
     +uB(Mn),uC(Mn),uAB(Mn,Mn),ETA(Mn,Mdim),phi(Mdim,Mdim,Mdim)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      data ZERO / 0.0 /
c                                                          up-date mu(i,m1)
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
      do 30 m1 = 1, MA
      do 30 i = 1, NA
        X = ZERO
        XX = ZERO
        do 20 j = 1, NB
        do 20 k = 1, NC
          W = ZERO
          do 10 m2 = 1, MB
          do 10 m3 = 1, MC
   10       W = W + phi(m1,m2,m3)*Nu(j,m2)*Eta(k,m3)
          X = X + W*(F(i,j,k)-FHAT(i,j,k))
          XX = XX + W*W*FHAT(i,j,k)
   20   continue
        Mu(i,m1) = Mu(i,m1) + X/XX
   30 continue
c                                                          up-date nu(j,m2)
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
      do 70 m2 = 1, MB
      do 70 j = 1, NB
        X = ZERO
        XX = ZERO
        do 60 i = 1, NA
        do 60 k = 1, NC
          W = ZERO
          do 50 m1 = 1, MA
          do 50 m3 = 1, MC
   50       W = W + phi(m1,m2,m3)*Mu(i,m1)*Eta(k,m3)
          X = X + W*(F(i,j,k)-FHAT(i,j,k))
          XX = XX + W*W*FHAT(i,j,k)
   60   continue
        Nu(j,m2) = Nu(j,m2) + X/XX
   70 continue
c                                                        up-date eta(k,m3)
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
      do 110 m3 = 1, MC
      do 110 k = 1, NC
        X = ZERO
        XX = ZERO
        do 100 i = 1, NA
        do 100 j = 1, NB
          W = ZERO
          do 90 m1 = 1, MA
          do 90 m2 = 1, MB
   90       W = W + phi(m1,m2,m3)*Mu(i,m1)*Nu(j,m2)
          X = X + W*(F(i,j,k)-FHAT(i,j,k))
          XX = XX + W*W*FHAT(i,j,k)
  100   continue
```

```
        Eta(k,m3) = Eta(k,m3) + X/XX
  110 continue
      return
      end
C
C Up-date phi parameters
C
      subroutine PHIFIT(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,F,FHAT,Ffit)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn),Ffit(Mn,Mn,Mn),uA(Mn),
     +uB(Mn),uC(Mn),uAB(Mn,Mn),ETA(Mn,Mdim),phi(Mdim,Mdim,Mdim)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      ZERO = 0.0
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
      do 20 m1 = 1, MA
      do 20 m2 = 1, MB
      do 20 m3 = 1, MC
        X = ZERO
        XX = ZERO
        do 10 i = 1, NA
        do 10 j = 1, NB
        do 10 k = 1, NC
          W = Mu(i,m1)*Nu(j,m2)*Eta(k,m3)
          X = X + W*(F(i,j,k)-FHAT(i,j,k))
          XX = XX + W*W*FHAT(i,j,k)
   10   continue
        phi(m1,m2,m3) = phi(m1,m2,m3) + X/XX
   20 continue
      return
      end
C
C  ......The Main Event......  Log-Three-mode association model
C
      subroutine MODEL(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,F,FAB,FAC,FBC,FA,FB,FC,FHAT,FHAB,FHAC,FHBC,FHA,
     + FHB,FHC,WA,WB,WC,Ffit,IuAB)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FAB(Mn,Mn),FAC(Mn,Mn),FBC(Mn,Mn),
     +FA(Mn),FB(Mn),FC(Mn),FHAT(Mn,Mn,Mn),FHA(Mn),FHB(Mn),FHC(Mn),
     +FHAB(Mn,Mn),FHAC(Mn,Mn),FHBC(Mn,Mn),WA(Mn),WB(Mn),WC(Mn),
     +Ffit(Mn,Mn,Mn),FHold(Mn,Mn,Mn),uA(Mn),uB(Mn),uC(Mn),
     +uAB(Mn,Mn),ETA(Mn,Mdim),phi(Mdim,Mdim,Mdim)
      double precision MU(Mn,Mdim),NU(Mn,Mdim)
      common / iter / CONV,DIFF,IMAX,ICALLS,ICONV,IHIST
      common / nunit / nuniti, nunito
      ICALLS = 0
      Ncycle = 1
      call SUM1W(WC,NC,WCsum)
c                                                       Output History?
      if (IHIST.eq.1) write(nunito,400)
c                                                    Initial fitted values
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     + MA,MB,MC,Ffit)
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
          FHold(i,j,k) = FHAT(i,j,k)
```

```
   10 continue
c                                                                The main loop
      do 1000 loop = 1, IMAX
c                                                              up-date u-terms
c
      call UFIT(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,IuAB,
     +         F,FA,FB,FC,FAB,FHAT,FHA,FHB,FHC,FHAB,Ffit)
      if (MA.gt.0) then
c                                             up-date category scale values
c
      call SCVFIT(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,
     +           F,FHAT,Ffit)
c                                                               up-date phi's
c
      call PHIFIT(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,
     +      F,Fhat,Ffit)
      endif
c                                                       center new estimates
c
      call CENTER(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,
     +           IuAB,WC,WCsum)
      ICALLS = ICALLS + 1
c                                                         Check Convergence
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     +         MA,MB,MC,Ffit)
      call CHECK(F,FHAT,FHold,NA,NB,NC)
c                                               ICONV = 1  ---> converged
      if (ICONV.eq.1) then
c                                                        scale new estimates
       if(MA.gt.0) call SCALE(MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,WA,WB,WC)
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     +         MA,MB,MC,Ffit)
       return
      endif
c                                                                  History
      if (IHIST.eq.1) then
        if (Ncycle.lt.10) then
          Ncycle = Ncycle + 1
        else
          call GSQUAR(F,FHAT,GSQ,NA,NB,NC)
          call PEARSO(F,FHAT,XSQ,NA,NB,NC)
          write(nunito,500) loop, DIFF, GSQ, XSQ
          Ncycle = 1
        endif
      endif
1000 continue
c                                                        scale new estimates
      if (MA.gt.0) call SCALE(MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,WA,WB,WC)
      call PRED(F,FHAT,u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     +         MA,MB,MC,Ffit)
      return
  400 format('1History of Maximum Likelihood Estimation:',/,
     +1x,'Cycle:',8x,'DIFF',15x,'G**2',15x,'X**2')
  500 format(1x,i6,3(3x,f15.8))
      end
C
C  Check convergence
C
      subroutine CHECK(F,FHAT,FHold,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
```

```
      dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn),FHold(Mn,Mn,Mn)
      common / iter / CONV,DIFF,IMAX,ICALLS,ICONV,IHIST
      data ZERO / 0.0 /
      DIFF = ZERO
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
   10    DIFF = DIFF + DABS(FHAT(i,j,k) - FHold(i,j,k))
c
      if (DIFF .lt.  CONV) then
        ICONV = 1
      else
        do 20 i = 1, NA
         do 20 j = 1, NB
          do 20 k = 1, NC
   20      FHold(i,j,k) = FHAT(i,j,k)
      endif
      return
      end
C
C  Center category scale values and adjust main effect parameters
C   then center main effects parameters
C
      subroutine CENTER(u,uA,uB,uC,uAB,MU,NU,ETA,phi,NA,NB,NC,
     +MA,MB,MC,IuAB,WC,C)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension uA(Mn),uB(Mn),uC(Mn),uAB(Mn,Mn),ETA(Mn,Mdim),
     +phi(Mdim,Mdim,Mdim),WC(Mn),zC(Mn),zAB(Mn,Mn)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      ZERO = 0.0
      if (IuAB.eq.1) then
       if (MA.gt.0) then
        do 30 m3 = 1, MC
          zC(m3) = ZERO
          do 10 k = 1, NC
   10       zC(m3) = zC(m3) + Eta(k,m3)*WC(k)
          zC(m3) = zC(m3)/C
          do 20 k = 1, NC
   20       Eta(k,m3) = Eta(k,m3) - zC(m3)
   30   continue
        do 50 i = 1, NA
        do 50 j = 1, NB
          zAB(i,j) = ZERO
          do 40 m1 = 1, MA
          do 40 m2 = 1, MB
          do 40 m3 = 1, MC
   40       zAB(i,j) = zAB(i,j) + phi(m1,m2,m3)*Mu(i,m1)*Nu(j,m2)*zC(m3)
          uAB(i,j) = uAB(i,j) + zAB(i,j)
   50   continue
       endif
       call DBLCEN(uAB,uA,uB,u,NA,NB)
      endif
      call CEMAIN(uA,u,NA)
      call CEMAIN(uB,u,NB)
      call CEMAIN(uC,u,NC)
      return
      end
C
C CEnter MAIN effect around 0 and adjust constant parameter of model
C
```

```
      subroutine CEMAIN(X1,X,Nrow)
      parameter (Mn=11,Mdim=11)
      implicit double precision (a-h,o-z), integer*4 (i-n)
      dimension X1(Mn)
      Xbar = 0
      call SUM1W(X1,Nrow,Xbar)
      Xbar = Xbar/DBLE(Nrow)
      do 10 i = 1, Nrow
   10   X1(i) = X1(i) - Xbar
      X = X + Xbar
      return
      end
C
C Double Center matrix of interaction terms and adjust main & constant terms
C
      subroutine DBLCEN(u12,u1,u2,u,N1,N2)
      parameter (Mn=11,Mdim=11)
      implicit double precision (a-h,o-z), integer*4 (i-n)
      dimension u12(Mn,Mn), u1(Mn), u2(Mn), xbar1(Mn), xbar2(Mn)
      x1 = DBLE(N1)
      x2 = DBLE(N2)
      grand = 0.0
      call ZERO1W(xbar1,N1)
      call ZERO1W(xbar2,N2)
      do 10 i = 1, N1
      do 10 j = 1, N2
         grand = grand + u12(i,j)
         xbar1(i) = xbar1(i) + u12(i,j)
         xbar2(j) = xbar2(j) + u12(i,j)
   10 continue
      grand = grand/(x1*x2)
      do 12 i = 1, N1
   12   xbar1(i) = xbar1(i)/x2
      do 14 j = 1, N2
   14   xbar2(j) = xbar2(j)/x1
      do 20 i = 1, N1
       do 20 j = 1, N2
         u12(i,j) = u12(i,j) - xbar1(i) - xbar2(j) + grand
   20 continue
      do 30 i = 1, N1
         u1(i) = u1(i) + xbar1(i)
   30 continue
      do 40 j = 1, N2
         u2(j) = u2(j) + xbar2(j)
   40 continue
      u = u - grand
      return
      end
C
C Orthonormalize scale values via Grahm-Schmidt process such that
C weighted sum of squares = 1 and weighted sum of crossproducts = 0.
C Then find transformation matrix by solving system of equations and
C use C the transformations matrix to adjust phi.
C
      subroutine SCALE(MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,WA,WB,WC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4(I-N)
      dimension Eta(Mn,Mdim), phi(Mdim,Mdim,Mdim),WA(Mn),WB(Mn),
     + WC(Mn),XX(Mn),EtaO(Mn,Mdim),A(Mdim,Mdim)
      double precision Mu(Mn,Mdim),Nu(Mn,Mdim),MuO(Mn,Mdim),
     + NuO(Mn,Mdim)
```

```
      ZERO = 0.0
c                                                              1st dimension
c                                                                 Mu(i,1)
      Xa = ZERO
      do 10 i = 1, NA
   10   Xa = Xa + Mu(i,1)*Mu(i,1)*WA(i)
      Xa = dsqrt(Xa)
      do 20 i = 1, NA
   20   MuO(i,1) = Mu(i,1)/Xa
c                                                                 Nu(j,1)
      Xb = ZERO
      do 40 j = 1, NB
   40   Xb = Xb + Nu(j,1)*Nu(j,1)*WB(j)
      Xb = dsqrt(Xb)
      do 50 j = 1, NB
   50   NuO(j,1) = Nu(j,1)/Xb
c                                                                 Eta(k,1)
      Xc = ZERO
      do 70 k = 1, NC
   70   Xc = Xc + Eta(k,1)*Eta(k,1)*WC(k)
      Xc = dsqrt(Xc)
      do 80 k = 1, NC
   80   EtaO(k,1) = Eta(k,1)/Xc
c                                                          Mu(i,m1), m1 > 1
      if (MA.gt.1) then
        do 150 m1 = 2, MA
          Last = m1 - 1
          do 105 L = 1, Last
            XX(L) = ZERO
            do 100 i = 1, NA
  100        XX(L) = XX(L) + MuO(i,L)*Mu(i,m1)*WA(i)
  105       continue
  do 115 i = 1, NA
   proj = ZERO
          do 110 L = 1, Last
  110       proj = proj + XX(L)*MuO(i,L)
   MuO(i,m1) = Mu(i,m1) - proj
  115       continue
          X = ZERO
          do 120 i = 1, NA
  120       X = X + MuO(i,m1)*MuO(i,m1)*WA(i)
          X = dsqrt(X)
          do 130 i = 1, NA
  130       MuO(i,m1) = MuO(i,m1)/X
  150     continue
       endif
c                                                          Nu(j,m2), m2 > 1
      if (MB.gt.1) then
        do 200 m2 = 2, MB
          Last = m2 -1
          do 165 L = 1, Last
            XX(L) = ZERO
            do 160 j = 1, NB
  160        XX(L) = XX(L) + NuO(j,L)*Nu(j,m2)*WB(j)
  165       continue
          do 175 j = 1, NB
           proj = ZERO
           do 170 L = 1, Last
  170        proj = proj + XX(L)*NuO(j,L)
           NuO(j,m2) = Nu(j,m2) - proj
  175     continue
```

```
          X = ZERO
          do 180 j = 1, NB
  180       X = X + NuO(j,m2)*NuO(j,m2)*WB(j)
          X = dsqrt(X)
          do 190 j = 1, NB
  190       NuO(j,m2) = NuO(j,m2)/X
  200    continue
      endif
c                                                    Eta(k,m3), m3 > 1
      if (MC.gt.1) then
        do 250 m3 = 2, MC
          Last = m3 - 1
          do 210 L = 1, Last
            XX(L) = ZERO
            do 205 k = 1, NC
  205       XX(L) = XX(L) + EtaO(k,L)*Eta(k,m3)*WC(k)
  210     continue
          do 225 k = 1, NC
          proj = ZERO
          do 220 L = 1, Last
  220       proj = proj + XX(L)*EtaO(k,L)
          EtaO(k,m3) = Eta(k,m3) - proj
  225     continue
          X = ZERO
          do 230 k = 1, NC
  230       X = X + EtaO(k,m3)*EtaO(k,m3)*WC(k)
          X = dsqrt(X)
          do 240 k = 1, NC
  240       EtaO(k,m3) = EtaO(k,m3)/X
  250    continue
      endif
c                      Adjust the phi's for transformation of scale values
      call solve(Mu,MuO,A,NA,MA)
      do 310 m1 = 1, MA
      do 310 m2 = 1, MB
      do 310 m3 = 1, MC
        zzz = zero
        do 300 k = 1, MA
  300     zzz = zzz + A(m1,k)*phi(k,m2,m3)
        phi(m1,m2,m3) = zzz
  310 continue
      call solve(Nu,NuO,A,NB,MB)
      do 320 m1 = 1, MA
      do 320 m2 = 1, MB
      do 320 m3 = 1, MC
        zzz = zero
        do 318 k = 1, MB
  318     zzz = zzz + A(m2,k)*phi(m1,k,m3)
        phi(m1,m2,m3) = zzz
  320 continue
      call solve(Eta,EtaO,A,NC,MC)
      do 340 m1 = 1, MA
      do 340 m2 = 1, MB
      do 340 m3 = 1, MC
        zzz = zero
        do 330 k = 1, MC
  330     zzz = zzz + A(m3,k)*phi(m1,m2,k)
        phi(m1,m2,m3) = zzz
  340 continue
c                      at last.....almost finished......
      do 500 m1= 1, MA
```

```
      do 500 i = 1, NA
  500  Mu(i,m1)=MuO(i,m1)
      do 501 m2= 1, MB
      do 501 j = 1, NB
  501  Nu(j,m2)=NuO(j,m2)
      do 502 m3 = 1, MC
      do 502 k = 1, NC
  502  Eta(k,m3)=EtaO(k,m3)
      return
      end
C
C Solve equation  Y = XB for B
C  where Y (nr x nc),  X (nr x nc), X'X=(X'X) = I (nc x nc) so...
C B = (X'X) X'Y = X'Y
C
      subroutine SOLVE(Y,X,B,nr,nc)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z)
      dimension Y(Mn,Mdim), X(Mn,Mdim), B(Mdim,Mdim)
      integer nr, nc
      zero = 0.0
      do 10 i = 1, nc
      do 10 j = 1, nc
        B(i,j) = zero
        do 5 k = 1, nr
          B(i,j) = B(i,j) + X(k,i)*Y(k,j)
    5   continue
   10 continue
      return
      end
C
C
C Compute likelihood ratio statistic:  G**2
C
      subroutine GSQUAR(F,FHAT,GSQ,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z)
      dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn)
      integer NA,NB,NC
      data ZERO / 0.0 /
      GSQ = ZERO
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
         if ( (F(i,j,k).eq.ZERO).or.(FHAT(i,j,k).eq.Zero) ) goto 10
         GSQ = GSQ + F(i,j,k) * dlog( F(i,j,k)/FHAT(i,j,k) )
   10 continue
      GSQ = 2*GSQ
      return
      end
C
C  Compute Pearson chi-square:   X**2
C
      subroutine PEARSO(F,FHAT,XSQ,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z)
      dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn)
      integer NA,NB,NC
      data ZERO / 0.0 /
      XSQ = ZERO
      do 10 i = 1, NA
```

```
       do 10 j = 1, NB
        do 10 k = 1, NC
         if (FHAT(i,j,k).ne.ZERO)    XSQ = XSQ +
     +       (F(i,j,k)-FHAT(i,j,k))*(F(i,j,k)-FHAT(i,j,k))/FHAT(i,j,k)
    10 continue
       return
       end
C
C Compute the p-value of a chi-square statistic
C
       subroutine CHI(XX,N,pval)
       implicit double precision (a-h,o-z)
       z = dexp(-XX/2)
       m = n/2
       m = m+m
       if(m.eq.n) go to 10
c        Odd degrees of freedom.
       prod = 0.79788456*z
       t = 1/(1+0.2316419*dsqrt(XX))
       pval = prod*t*(0.31938153+t*(-0.356563782+t*(1.781477937
     *   +t*(-1.821255978+1.330274429*t))))
       m = n-2
       prod = prod/dsqrt(XX)
       do 2 j=1,m,2
       prod = prod*XX/j
     2 pval = pval+prod
       go to 20
c        Even degrees of freedom
    10 pval = z
       m = n-2
       if(n.le.2) go to 20
       prod = z
       do 12 j=2,m,2
       prod = prod*XX/j
    12 pval = pval+prod
    20 return
       end
C
C  Output Model Parameters:  Constant and Main Effects
C
       subroutine OUmain(u,uA,uB,uC,NA,NB,NC)
       parameter (Mn=11,Mdim=11)
       implicit double precision (A-H,O-Z), integer*4 (i-n)
       dimension uA(Mn),uB(Mn),uC(Mn)
       common / nunit / nuniti, nunito
       write(nunito,100) u
       write(nunito,110)
       do 10 i = 1, NA
    10   write(nunito,111) i, uA(i)
       write(nunito,120)
       do 20 j = 1, NB
    20   write(nunito,121) j, uB(j)
       write(nunito,130)
       do 30 k = 1, NC
    30   write(nunito,131) k, uC(k)
       return
   100 format(/,1x,'Constant:  ',f10.5)
   110 format(//,1x,'Main Effects for Rows:',/)
   111 format(1x,'uA(',i2,') = ',f15.5)
   120 format(//,1x,'Main Effects for Columns:',/)
   121 format(1x,'uB(',i2,') = ',f15.5)
```

```
  130 format(//,1x,'Main Effects for Layers:',/)
  131 format(1x,'uC(',i2,') = ',f15.5)
      end
C
C Output Model Parameters:  Association Model Interactions effects
C
      subroutine OUTint(uAB,MU,NU,ETA,phi,NA,NB,NC,MA,MB,MC,IuAB,
     +WA,WB,WC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension uAB(Mn,Mn), ETA(Mn,Mdim),WA(Mn),WB(Mn),WC(Mn),
     +phi(Mdim,Mdim,Mdim)
      double precision MU(Mn,Mdim), NU(Mn,Mdim)
      common / nunit / nuniti, nunito
      if (IuAB.eq.1) then
        write(nunito,105)
        write(nunito,106) (j,j=1,NB)
        do 5 i = 1, NA
    5     write(nunito,107) i, (uAB(i,j),j=1,NB)
      endif
      if (MA.eq.0) return
      write(nunito,110)
      write(nunito,101) (m1,m1=1,MA)
      do 10 i = 1, NA
   10  write(nunito,111) i, (MU(i,m1),m1=1,MA), WA(i)
      write(nunito,120)
      write(nunito,102) (m2,m2=1,MB)
      do 20 j = 1, NB
   20  write(nunito,121) j, (NU(j,m2),m2=1,MB), WB(j)
      write(nunito,130)
      write(nunito,103) (m3,m3=1,MC)
      do 30 k = 1, NC
   30  write(nunito,131) k, (ETA(k,m3),m3=1,MC), WC(k)
      write(nunito,140)
      do 50 m1 = 1, MA
       write(nunito,150) m1
       write(nunito,103) (m3,m3=1,MC)
       do 40 m2 = 1, MB
   40   write(nunito,151) m2, (phi(m1,m2,m3),m3=1,MC)
   50 continue
      return
  101 format(9x,'mA = ',11(5x,i2,5x))
  102 format(9x,'mB = ',11(5x,i2,5x))
  103 format(9x,'mC = ',11(5x,i2,5x))
  105 format(/,1x,'AB interaction:  uAB(i,j)')
  106 format(6x,'j = ',10(5x,i2,5x),/,10(5x,i2,5x))
  107 format(1x,'i = ',i2,3x,10(2x,f10.5),/,10(2x,f10.5))
  110 format(//,1x,'Row Scale Values and Weights:  ',/)
  111 format(1x,'Mu(',i2,',m) = ',3x,12(2x,f10.5))
  120 format(//,1x,'Column Scale Values and Weights:',/)
  121 format(1x,'Nu(',i2,',m) = ',3x,12(2x,f10.5))
  130 format(//,1x,'Layer Scale Values and Weights:',/)
  131 format(1x,'Eta(',i2,',m) = ',1x,12(2x,f10.5))
  140 format(//,1x,'Measures of association --- "Core Matrix" of ',
     +' phi(mA,mB,mC):')
  150 format(/,1x,'  mA = ',i2)
  151 format(1x,'mB = ',i2,2x,12(2x,f10.5))
      end
C
C  Output 1-way Table in vector format
C
```

```
      subroutine OUT1W(XTOT,X,HX,Chi1,NX)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z)
      dimension X(Mn),HX(Mn),Chi1(Mn)
      integer NX
      common / nunit / nuniti, nunito
      do 10 i = 1, NX
   10  write(nunito,100)
     +     i, X(i)/XTOT, X(i), HX(i), X(i)-HX(i), Chi1(i)
      return
  100 format(2x,i2,5(3x,f10.5))
      end
C
C  Output 2-way Table in vector format
C
      subroutine OUT2W(XTOT,XY,HXY,Chi2,NX,NY)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension XY(Mn,Mn),HXY(Mn,Mn),Chi2(Mn,Mn)
      common / nunit / nuniti, nunito
      do 10 i = 1, NX
       do 10 j = 1, NY
   10    write(nunito,100) i,j,XY(i,j)/XTOT,XY(i,j),HXY(i,j),
     +    XY(i,j)-HXY(i,j), Chi2(i,j)
      return
  100 format(2(2x,i2),5(3x,f10.5))
      end
C
C  Output 3-way Table
C
      subroutine OUT3W(FTOT,F,FHAT,Ffit,NA,NB,NC)
      parameter (Mn=11,Mdim=11)
      implicit double precision (A-H,O-Z), integer*4 (i-n)
      dimension F(Mn,Mn,Mn),FHAT(Mn,Mn,Mn), Ffit(Mn,Mn,Mn)
      common / nunit / nuniti, nunito
      write(nunito,100)
      do 10 i = 1, NA
       do 10 j = 1, NB
        do 10 k = 1, NC
        resid =  F(i,j,k)-FHAT(i,j,k)
        write(nunito,110) i, j, k, F(i,j,k)/FTOT,F(i,j,k), FHAT(i,j,k),
     +    resid, resid/dsqrt(FHAT(i,j,k)), Ffit(i,j,k)
   10 continue
      return
  100 format(/,1x,'Observed, Fitted & Residual Frequencies:',/,
     +'   I   J   K   P(ijk)      F(ijk)      FHAT(ijk)   Residual',
     +'    Std Resid  F fit?')
  110 format(3(2x,i2),5(2x,f10.5),2x,f2.0)
      end
```

# REFERENCES


Aldrich, J.H., & Nelson, F.D. (1984). *Linear, Probability, Logit, and Probit Models*. Beverly Hills, CA: Sage Publications.

Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. New York: Wiley.

Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.

Bartholomew, D.J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society, B, 42*, 293–321.

Bartholomew, D.J. (1987). *Latent Variable Models and Factor Analysis*. New York, NY: Oxford University Press.

Becker, M.P. (1989a). Models for the analysis of association in multivariate contingency tables. *Journal of the American Statistical Association*, *84*, 1014–1019.

Becker, M.P. (1989b). On the bivariate normal distribution and association models for ordinal categorical data. *Statistics and Probability Letters*, *8*, 435–440.

Becker, M.P. (1990a). Algorithm AS 253: Maximum likelihood estimation of the RC(M) association model. *Applied Statistics*, *39*, 152–167.

Becker, M.P. (1990b). *Exploratory analysis of association models using loglinear models and singular value decompositions*. Paper presented at the $59^{\text{th}}$ International Workshop on Statistical Modelling; Touleure, FR.

Becker, M.P., & Clogg, C.G. (1989a). Analysis of sets of two-way contingency tables using association models. *Journal of the American Statistical Association*, *84*, 142–151.

Becker, M.P., & Clogg, C.G. (1989b). *Maximum likelihood estimation of association models for sets of two–way contingency tables* (Tech. Rep. No. 89-1). Ann Arbor: University of Michigan, Department of Biostatistics.

Bentler, P.M. (1973). Assessment of developmental factor change at the individual and group level. In J.R. Nesselroade and H.W. Resse (Eds.) *Life-Span Developmental Psychology* (pp. 145–174). New York: Academic Press.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete Multivariate Analysis*. Cambridge, MA: The MIT Press.

Bloxom, B. (1984). Tucker's three–mode factor analysis model. In H.G. Law, C.W. Synder, J. Hattie, and R.P. McDonald (Eds.) *Research Methods for Multimode Data Analysis* (pp. 104–121). New York, NY: Praeger.

Böckenholt, U. & Böckenholt, I. (1990). Correspondence analysis with linear constraints. *Psychometrika*, *55*, 633–639.

Carroll, J.D., & Chang, J. (1970). Analsyis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, *35*, 283–319.

Carroll, J.D., Green, P.E., & Schaffer, C.M. (1986). Interpoint distance comparisons in correspondence analysis. *Journal of Marketing Research*, *23*, 271–280.

Carroll, J.D. & Green, P.E. (1988). An INSCAL–base approach to multiple correspondence analysis. *Journal of Marketing Research*, *25*, 193–203.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5–32.

Choulakian, V. (1988a). Exploratory analysis of contingency tables by loglinear formulations and generalizations of correspondence analysis. *Psychometrika*, 53, 235–250.

Choulakian, V. (1988b). Errata. *Psychometrika*, 53, 593.

Clogg, C.C. (1986). Statistical modelling versus singular value decomposition. *International Statistical Review*, *54*, 284-288.

Clogg, C.C. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics*, *38*, 281.

Clogg, C.C., Eliason, S.R., & Whal, R.J. (1990). Labor–market experiences and labor–force outcomes. *American Journal of Sociology*, *95*, 1537–1576.

Conaway, M.R. (1989). Conditional likelihood methods for repeated categorical measurement. *Journal of the American Statistical Association*, *84*, 53–62.

Conaway, M.R. (1992). The analysis of repeated categorical measurements subject to nonignorable nonresponse. *Journal of the American Statistical Association*, *87*, 817–824.

Coppi, R., & Bolasco, S. (Eds.) (1989). *Multiway Data Analysis*. Amsterdam, The Netherlands: Elsevier Science Publishers.

Cox, D.R. (1983). Some remarks on overdispersion. *Biometrika, 70*, 269–274.

Cressie, N., & Holland, P.W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometric, 48*, 129–141.

D'Ambra, L., & Lauro, N. (1989). Non symmetrical analysis of three-way contingency tables. In R. Coppi & S. Bolasco (Eds.), *Multiway Data Analysis* (pp. 341–352). Amsterdam, The Netherlands: Elsevier Science Publishers.

D'Aubigny, G., & Polit, E. (1989). Some optimality properties of the generalization of the Tucker method to the analysis of n-way tables with specified metrics. In R. Coppi & S. Bolasco (Eds.), *Multiway Data Analysis* (pp. 39–52). Amsterdam, The Netherlands: Elsevier Science Publishers.

Darroch, J.N. (1974). Multiplicative and additive interaction in contingency tables. *Biometrika*, *61*, 207–214.

de Leeuw, J., & van der Heijden, P.G.M. (1988). The analysis of time–budgets with a latent time–budget model. In E. Diday (Ed.), *Data Analysis and Informatics, V*, (pp.159–166). Amsterdam: North-Holland.

de Leeuw, J., van der Heijden, P.G.M., & Verboon, P. (1989). *A latent time-budget model*, (PRM 01-89). Leiden: Department of Psychology.

de Leeuw, J., & van der Heijden, P.G.M. (1991). Reduced rank models for contingency tables. *Biometrika*, *78*, 229–232.

Denis, J.B., & Dhorne, T. (1989). Orthogonal tensor decomposition of 3–way tables. In R. Coppi & S. Bolasco (Eds.) *Multiway Data Analysis* (pp. 31–37). Amsterdam, The Netherlands: Elsevier Science Publishers.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*, 211–218.

Eliason, S.R. (1989). *The Categorical Data Analysis System*, version 3.00A. Department of Psychology, University of Iowa.

Escoufier, Y., & Junca, S. (1986). Least–squares approximation of frequencies or their logarithms. *International Statistical Review*, *54*, 279–283.

Escoufier, Y. (1988). Beyond correspondence analysis. In H. H. Bock (Ed.), *Classification and Related Methods of Data Analysis*, (pp. 505–514). New York: North-Holland.

Escoufier, Y. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics*, *38*, 281–282.

Fienberg, S.E. (1985). *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA: The MIT Press.

Franc, A. (1989). Multiway matrices: some algebraic remarks. In R. Coppi & S. Bolasco (Eds.) *Multiway Data Analysis* (pp. 19–29). Amsterdam, The Netherlands: Elsevier Science Publishers.

Giffi, A. (1990). *Nonlinear Multivariate Analysis*. Chichester, England: John Wiley & Sons.

Gilula, Z. (1982). A note on the analysis of association in cross-classifications having ordered categories. *Communications in Statistics: Theory and Methodology*, *11*, 1233–1240.

Gilula, Z. (1984). On some similarities between canonical correlation models and latent class models for two–way contingency tables. *Biometrika*, *71*, 523–529.

Gilula, Z., & Haberman, S.J. (1986). Canonical analysis of contingency tables by maximum likelihood. *Journal of the American Statistical Association, Theory and Methods*, *81*, 780–788.

Gilula, Z., & Haberman, S.J. (1988). The analysis of multiway contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association, Theory and Methods*, *83*, 760–771.

Gilula, Z., Krieger, M., & Ritov, Y. (1988). Ordinal association in contingency tables: Some interpretive aspects. *Journal of the American Statistical Association*, *83*, 540–545.

Goldstein, H. (1987). The choice of constraints in correspondence analysis. *Psychometrika*, *52*, 207–216 .

Goodman, L.A. (1978). *Analyzing Qualitative/Categorical Data*. Cambridge, MA: Abt Books.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, *74*, 537–552.

Goodman, L.A. (1981a). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, *68*, 347–355.

Goodman, L.A. (1981b). Association models and canonical correlation models in the analysis of cross-classifications having ordered categories. *Journal of the American Statistical Association*, *76*, 320–334.

Goodman, L.A. (1985). The analysis of cross-classified data having ordered and/or unorderd categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics*, *13*, 10–69.

Goodman, L.A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review*, *54*, 243–309.

Goodman, L.A. (1987a). The analysis of a set of multidimensional contingency tables using log-linear models, latent-class models, and correlation models: The Solomon data revisited. In A. E. Gelfand (Ed.), *Contributions to the Theory and Applications of Statistics* (pp. 429–484). New York: Academic Press.

Goodman, L.A. (1987b). New methods for analyzing the intrinsic character of qualitative variables using cross-classified data. *American Journal of Sociology*, *93*, 529–583.

Goodman, L.A. (1991). Measures, models, and graphical displays in the analysis of cross-classified data. *Journal of the American Statistical Association*, *86*, 1085–1111.

Goodman, L.A., & Kruskal, W.H. (1979). *Measures of Association for Cross Classifications*. New York: Springer-Verlag.

Gower, J.C. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics*, *38*, 273–274.

Greenacre, M.J. (1984). *Theory and Applications of Correspondence Analysis*. Orlando, Florida: Academic Press.

Greenacre, M.J. (1988). Correspondence analysis of multivariate categorical data by weighted least–squares. *Biometrika*, *75*, 457–467.

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

Harshman, R.A., & Lundy, M.E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H.G. Law, C.W. Synder, J. Hattie, and R.P. McDonald (Eds.) *Research Methods for Multimode Data Analysis* (pp 122–215). New York, NY: Praeger.

Harshman, R.A., & Lundy, M.E. (1992, July). Three-way DEDICOM: analyzing multiple matrices of asymmetric relationships. In S. Nishisato (chair), *Multidimensional Methods*. 57thAnnual Meeting of the Psychometric Society.

Hausman, J., Hall, B.H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents–R&D relationship. *Econometrica*, *52*, 909–938.

Healy, J.R., & Goldstein, H. (1976). An approach to the scaling of categorized attributes. *Biometrika*, *63*, 219–229.

Hill, M.O. (1974). Correspondence analysis: A neglected multivariate technique. *Applied Statistics*, *23*, 340–354.

Iacobucci, D.M. (1989). Modeling multivariate sequential dyadic interactions. *Social Networks, 11*, 315–362.

Israëls, A.Z. (1987). *Eigenvalue Techinques For Qualitative Data*. Leiden, The Netherlands: DWSO Press.

Israëls, A.Z. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics*, *38*, 283.

Jansen, M.G.H., & van Duijn, M.A.J. (1992). Extensions of Rasch's multiplicative poisson model. *Psychometrika*, *57*, 405–414.

Kiers, H.A.L. (1989). *Three–Way Methods for the Analysis of Qualitative and Quantitative Two–Way Data*. Leiden, The Netherlands: DWSO Press.

Kiers, H.A.L. (1991). Hierarchical relations among three-way methods. *Psychometrika, 56*, 449–470.

Kiers, H.A.L., Kroonenberg, P.M., & ten Berge, J.M.F. (1992). An efficient algorithm for TUCKALS3 on data with large numbers of observations. *Psychometrika*, *57*, 415–422.

Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., & Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics, 33*, 133–158.

Kramer, L., & Gottman, J.M (1992). Becoming a sibling: "With a little help from my friends". *Developmental Psychology, 28*, 685–699.

Krippendorff, K. (1986). *Information Theory*. Newbury Park, CA: Sage Publications.

Kroonenberg, P.M. (1983). *Three–Mode Principal Component Analysis: Theory and Applications*. Leiden, The Netherlands: DWSO Press.

Kroonenberg, P.M. (1984). Three–mode principal component analysis: Illustrated with an example from attachment theory. In H.G. Law, C.W. Synder, J. Hattie, and R.P. McDonald (Eds.) *Research Methods for Multimode Data Analysis* (pp 64–103). New York, NY: Praeger.

Kroonenberg, P.M. (1989). Singular value decompositions of interactions in three–way contingency tables. In R. Coppi & S. Bolasco (Eds.) *Multiway Data Analysis* (pp. 169–184). Amsterdam, The Netherlands: Elsevier Science Publishers.

Kroonenberg, P.M., & Brouwer, P. (1985a). *User's Guide to TUCKALS2, version 4.0*, (WR 85-12-RP). Leiden, The Netherlands: Department of Education.

Kroonenberg, P.M., & Brouwer, P. (1985b). *User's Guide to TUCKALS3, version 4.0*, (WR 85-09-RP). Leiden, The Netherlands: Department of Education.

Kroonenberg, P.M., & de Leeuw, J. (1980). Principal component analysis of three–mode data by means of alternating least squares algorithms. *Psychometrika*, *45*, 69–97.

Kroonenberg, P.M., Lammers, C.J., & Stoop, I. (1985). Three-mode principal comonent analysis of multivariate longitudinal organizational data. *Sociological Methods & Research*, *14*, 99–136.

Kruskal, J.B. (1984). Multilinear models. In H.G. Law, C.W. Synder, J. Hattie, and R.P. McDonald (Eds.) *Research Methods for Multimode Data Analysis* (pp 36–62). New York, NY: Praeger.

Kruskal, J.B. (1989). Rank, decomposition, and uniqueness for 3–way and $N$–way arrays. In R. Coppi & S. Bolasco (Eds.) *Multiway Data Analysis* (pp. 7–18). Amsterdam, The Netherlands: Elsevier Science Publishers.

Kruskal, J.B., Harshman, R.A., & Lundy, M.E. (1989). How 3–MFA can cause degenerate PARAFAC solutions, among other relationships. In R. Coppi & S. Bolasco (Eds.) *Multiway Data Analysis* (pp. 115–122). Amsterdam, The Netherlands: Elsevier Science Publishers.

Lancaster, H.O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, *44*, 289–292.

Lancaster, H.O. (1969). *The Chi-squared Distribution*. New York: Wiley.

Lauritzen, S.L., & Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, *17*, 31–57.

Lauro, N. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics*, *38*, 283–284.

Lauro, N., & D'Ambra, L. (1984). L'Analyse non symetrique des correspondances. In E. Diday, et al. (Eds.) *Data Analysis and Informatics, III* (pp. 433–446). Amsterdam: Elsevier Science Publishers.

Law, H.G., Synder, C.W., Hattie, J., & McDonald, R.P. (Eds.) (1984). *Research Methods for Multimode Data Analysis*. New York, NY: Praeger.

Lundy, M.E., Harshman, R.E., & Kruskal, J.B. (1989). A two stage procedure incorporating good features of both trilinear and quadrilinear models. In R. Coppi & S. Bolasco (Eds.) *Multiway Data Analysis* (pp. 123–130). Amsterdam, The Netherlands: Elsevier Science Publishers.

McCullagh, P., & Nelder, J.A. (1989). *Generalized Linear Models* (2nd ed.). London: Chapman and Hall.

McCuthcheon, A.L. (1987). *Latent Class Analysis*. Newbury Park, CA: Sage Publications.

Muthèn, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 407–419.

Muthèn, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

Nelder, J.A. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics*, *38*, 280.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and Its Applications*, Toronto, Canada: University of Toronto Press.

Pannekoek, J. (1985). Log–multiplicative models for multiway tables. *Sociological Methods & Reasearch*, *14*, 137–153.

Pannekoek, J. (1992). Generalized least squares estimation of correspondence analysis. Unpublished manuscript, Netherlands Centeral Bureau of Statistics, Department of Statistical Methods.

Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyâ*, *26*, 329–358.

Read, T.R.C., & Cressie, N.A.C (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. New York, NY: Springer-Verlag.

Rom, D., & Sarkar, S.K. (1990). Approximating probability intergrals of multivariate normal using association models. *Journal of Statistical Computation and Simulation*, *35*, 109–119.

Sabatier, R., Lebreton, J., & Chessel, D. (1989). Principal components analysis with instrumental variables as a tool for modelling composition data. In R. Coppi & S. Bolasco (Eds.) *Multiway Data Analysis* (pp. 341–352). Amsterdam, The Netherlands: Elsevier Science Publishers.

SAS Institute, INC. (1988). *SAS Procedures Guide* (Release 6.03 Edition). Cary, NC: author.

Takane, Y., Yanai, H., & Mayekawa, S. (1991). Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, *56*, 667–684.

Teeuwen, J.H.M. (1988). Applications of methods for ordinal data in contingency tables. In M.G.H. Jansen & W.H. van Schuur (Eds.), *The Many Faces of Multivariate Analysis: Vol. 1* (pp. 153–165). Groningen, West Germany: RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen.

Tenenhaus, M., & Young, F.W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, *50*, 91–119.

Tucker, L.R (1963). Implications of factor analysis of three-way matrices for measurement of change. In C.W. Harris (Ed.) *Problems in Measuring Change*, (pp. 122–137). Madison, WI: University of Wisconsin Press.

Tucker, L.R (1964). The extension of factor analysis to three-dimensional matrices. In N. Fredieriksen and H. Gulliksen (Eds.) *Contributions to Mathematrical Psychology*, (pp. 110–127). NY: Holt, Rinehart, and Winston, INC.

Tucker, L.R (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279-311.

van der Heijden, P.G.M. (1987). *Correspondence Analysis of Longitudinal Data*. Leiden, The Netherlands: DWSO Press.

van der Heijden, P.G.M., & de Leeuw, J. (1985). Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, *50*, 429–447.

van der Heijden, P.G.M., & de Leeuw, J. (1989). Correspondence analysis, with special attention to the analysis of panel data and event history data. In C.C. Clogg (Ed.) *Sociological Methodology* (pp 43–87). Cambridge, MA: Basil Blackwell.

van der Heijden, P.G.M., de Falugerolles, A., & de Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis. *Applied Statistics*, *2*, 249–292.

van der Heijden, P.G.M., & Meijerink, F. (1988). *Generalized Correpsondence Analysis of Multi-way Contingency Tables and Multi-way (Super-) Indicator Matrices*, (PRM 01-88). Leiden: Department of Psychology.

van der Heijden, P.G.M., Mooijaart, & de Leeuw, J. (1989). *Latent Budget Analysis*, (PRM 03-89). Leiden: Department of Psychology.

van der Heijden, P.G.M., & Worsley, K.J. (1988). Comment on 'Correspondence analysis used complementary to loglinear analysis'. *Psychometrika, 53*, 287–291.

Visser, R.A. (1982). *Analysis of Longitudinal Data in Behavioral and Social Research*. Leiden, The Netherlands: DSWO Press.

Wang, Y.J. (1987). The probability integrals of bivariate normal distributions: A contingency table approach. *Biometrika*, *74*, 185–190.

Wasserman, S., Faust, K., & Galaskiewicz, J. (1990) Correspondence and canonical analysis of relational data. *Journal of Mathematical Sociology*, *15*, 11–64.

Wickens, T.D. (1989). *Multiway Contingency Tables Analysis for the Social Sciences*. Hillsdale, NJ: Lawerence Erlbaum.

Whittaker, J. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics*, *38*, 278–279.

Yoshizawa, T. (1988). Singular value decomposition of multiarray data and its applications. In *Recent Developments in Clustering, and Data Analysis* (pp 241–257). San Diego, CA: Academic Press.

# VITA

Carolyn Jane Anderson
Department of Psychology
Quantitative Division
University of Illinois at Urbana-Champaign

## Education

University of Illinois at Urbana-Champaign
Ph.D in Quantitative Psychology, expected January 1993
Masters of Science in Statistics, 1986

University of California, Berkeley
Bachelor of Arts in Psychology and Economics, 1982

## Publications

Wasserman, S., & Anderson, C.J. (1987). Stochastic a posteriori blockmodels: Construction and assessment, *Social Networks, 9*, 1–36.

Birnbaum, M.H., Anderson, C.J., & Hynan, L.G. (1989). Two opertations for "Ratios" and "Differences" of distances of the mental map. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 785–796.

Anderson, C.J. (1989). Review of *Eigenvalue Techniques for Qualitative Data* by Abby Israëls. *Journal of Classification, 6*, 147–152.

Birnbaum, M.H., Anderson, C.J., & Hynan, L.G. (1990). Theories of bias in probability judgment. In J.P. Caverni, J.M. Fabre, & M. Gonzalez (Eds.) *Cognitive Biases* (pp. 477–498). Amsterdam: Elsevier Science.

Anderson, C.J., Wasserman, S., & Faust, K. (1992). Building stochastic blockmodels. *Social Networks, 14*, 137–161.

Weber, E.U., Anderson, C.J., & Birnbaum, M.H.(1992). A theory of percieved risk and attractiveness. *Organizational Behavior and Human Decision Processes, 52*, 492–523.

Anderson, C.J., & Wasserman, S. (1993, in press). Recent developments in categorical data analysis: A review of *Categorical Data Analysis* by Alan Agresti and *Multiway Contingency Tables Analysis for the Social Sciences* by Thomas Wickens. *Journal of Mathematical Psychology, 37*.