METHODS FOR MULTIDIMENSIONAL CONTINGENCY TABLE ANALYSIS

by

GERALDINE HILKER

B.A., Kearney State College, 1973

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1975

Approved by:

K. E. Kemp
Major Professor

TABLE OF CONTENTS

1. Introduction

The purpose of this paper is to synthesize the diverse information in the literature of multidimensional contingency tables and to compare the techniques of analysis, computational complexity and types of hypotheses.

Multidimensional contingency table analysis was largely ignored until Bartlett's paper appeared in 1935. Since then numerous papers have been written attempting to analyze contingency tables in many fashions. Most of the methods are asymptotically equivalent (Berkson (1972)). The methods of maximum likelihood, minimum discrimination, and the general linear model are outlined in this paper. References are given for other methods as well. Estimation procedures are discussed and examples are given to test the hypothesis of no three-way interaction.

Contingency tables are characterized by two types of data: 1) factor - classification of the unit to which the experimental unit belongs and, 2) response - classification of what happened to the experimental unit. Hence, the data are the frequencies with which experimental units fall into the various combinations of factor-response categories.

The term 'interaction' has been used in different contexts in the statistical literature. Bhapkar and Koch (1968) list the four situations that arise in multidimensional contingency tables and the interpretation of the 'no interaction' hypothesis for each situation.

Model 1 - multi-response and no factor

a) concerned with the relationship between different responses

b) "No interaction" hypotheses pertain to whether some measure of association among the members of a certain set of responses depends upon the categories of the members of a subset of the remaining responses.

Model 2 - multi-response and one factor

a) concerned with the effect of the factor in the joint and marginal distribution of the responses and on measures of association among the responses.

b) "No interaction" hypotheses pertain to whether some measure of association among the members of a certain set of responses depends upon the categories of the members of a subset of the remaining responses and/or the categories of the factor.

Model 3 - multi-response and multi-factor

a) concerned with both the relationship among the responses and the way in which the factors combine.

b) "No interaction" hypothesis can take many forms. These forms question the pattern of association among responses and whether factor combinations affect the response distribution.

Model 4 - one response and multi-factor

a) concerned with the way that the factors influence the response (analogous to univatiate analysis of variance).

b) "No interaction" hypothesis pertains to the manner in which factors combine to determine the response distribution.

In all of the above models the marginals determined by adding factor frequencies are assumed to be fixed. Marginals determined

THIS BOOK CONTAINS NUMEROUS PAGES WITH THE PAGE NUMBERS CUT OFF

THIS IS AS RECEIVED FROM THE CUSTOMER

from response frequencies are variable. Also, the total sample size is fixed prior to conducting the experiment.

The procedure to be followed in contingency table analysis is: 1) specify the analysis method to be used, 2) formulate hypotheses of interest, 3) compute expected frequencies (if the method requires it) and 4) interpret the results.

2. Methods

Many methods have been proposed to analyse multidimensional contingency tables. The traditional Pearson's chi-square statistic

$$\chi^2_p = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

has been criticized because it doesn't produce any estimates of the effects that the variables have on each other. Three proposed methods of multidimension contingency table analysis are: 1) maximum likelihood; 2) minimum discrimination information; and 3) the general linear model. The background for these methods will now be presented. Later, examples will be given and other methods of analysis will be mentioned.

2.1 Maximum Likelihood

The maximum likelihood approach discussed here will be based on logits as given by Goodman (1970). Refer to Fryer (1966) for a discussion of logits. This type of analysis is appropriate for data of the Model 4 type and should be used primarily for stratified samples. For notational convenience a three-way contingency table (I x J x K) will be discussed.

Notation:

i=1,2,,I j=1,2,,J k=1,2,,K