1

CANADIAN THESES ON MICROFICHE

I.S.B.N.

THESES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

(7)

0-315-17833-7

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Division

Division des thèses canadiennes

Ottawa, Canada
K1A 0N4

65742

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

• Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

Georgia Ruth Roberts

Date of Birth — Date de naissance

15 / 5 / 48

Country of Birth — Lieu de naissance

Canada

Permanent Address — Résidence fixe

6 Lynhaven Cr.
Nepean Ont. K2E 5K2

Title of Thesis — Titre de la thèse

Contributions to Chi-Squared Tests
with Survey Data

University — Université

Carleton

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

Ph. D.

Year this degree conferred — Année d'obtention de ce grade

1985

Name of Supervisor — Nom du directeur de thèse

JNK Rao

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

Date

Jan 30 / 85

Signature

Georgia Roberts

CONTRIBUTIONS TO CHI-SQUARED TESTS

WITH

SURVEY DATA

by

Georgia Ruth Roberts, B.A. (Hons.), M.Sc.

A thesis submitted to the Faculty of
Graduate Studies in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University

Ottawa, Ontario

January, 1985

© copyright

1985, Georgia Roberts

The undersigned hereby recommend to the Faculty of Graduate
Studies and Research acceptance of the thesis, submitted by

Georgia Roberts, B.A.(Hons.), M.Sc.

in partial fulfilment of the requirements for the degree of
Doctor of Philosophy.

B.007. Puttanawana

Chairman, Department of
Mathematics and Statistics

○

JNK Rao

Thesis Supervisor

David Bellhouse

External Examiner

Carleton University

January, 1985

ABSTRACT

The impact of survey design on chi-squared test statistics applied to multiway contingency tables is studied. In Chapter 1, a review of the relevant literature is given.

Chapter 2 contains two extensions of the work by Rao and Scott (Annals of Statistics, 1984) on loglinear models. First, the form of a correction to the Pearson chi-squared statistic is developed for the case of one set of fixed marginal totals of the observed frequencies. The correction is expressed in terms of within-population cell proportions and associated design effects. Second, several results derived for the loglinear model are extended to a more general model. Two examples not covered by Rao and Scott's model are presented.

In Chapter 3, the asymptotic properties of chi-squared test statistics for binary response models are developed. First-order corrections to the test statistics requiring only a small amount of information are suggested. As well, for the particular binary response model, the logit model, conditions for an equivalent specification as a loglinear model with direct estimates are obtained.

Diagnostics for assessing a binary response model, taking the survey design into account, are suggested in Chapter 4. A logit model is then fitted to a set of binary response data from the Canadian Labour Force Survey and several of the diagnostics for the fit are examined.

In Chapter 5 the asymptotic properties of chi-squared test statistics for polytomous response models are developed. First-order corrections to the test statistics are suggested. As well, two particular models are examined.

Chapter 6 contains some suggestions for further research.

ACKNOWLEDGEMENTS

My supervisor, Professor J.N.K. Rao, is chiefly responsible for my interest in categorical data. I am deeply grateful for his direction, criticism and insight into the topics included in this thesis. I have thoroughly enjoyed working under his supervision.

As well, thanks go to Dr. Alistair Scott for his helpful correspondence on certain aspects of the thesis topic.

I would also like to acknowledge Gill Murray for her outstanding work in the preparation of this manuscript.

Finally, I give special thanks to my husband David, who has supported me through this endeavor and given me unlimited encouragement during the lulls in my progress.

TABLE OF CONTENTS

	Page
ACCEPTANCE SHEET	ii
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1: Introduction	1
1.1 Review of the Literature	1
1.2 Outline of Thesis	17
CHAPTER 2: Extensions of Rao-Scott theory for loglinear models	20
2.1 Introduction	20
2.2 Notation and Background Results for the Loglinear Model	21
2.3 δ . for Models with Direct Estimates and for Sampling with One Set of Fixed Marginal Totals	26
2.4 Extensions to General Models	39
CHAPTER 3: Binary Response Models	59
3.1 Introduction	59
3.2 Asymptotic Distribution of χ^2 for Binary Response Models	60
3.3 Examples of Binary Response Models	79
3.4 Logit Models and Direct Estimates	91
CHAPTER 4: A: Diagnostic Procedures for Binary Response Models	
B: Logit Model Analysis of Labour Force Survey Data	100

Table of Contents cont'd...

4.1	Introduction	100
4.2	Diagnostic Methods for Binary Response Models	100
4.3	Application of Logit Model to Labour Force Survey Data	108
CHAPTER 5: Polytomous Response Models		120
5.1	Introduction	120
5.2	Asymptotic Distribution of X^2 for Polytomous Response Models	121
5.3	Examples of Polytomous Response Models	134
CHAPTER 6: Further Research		141
APPENDIX		143
BIBLIOGRAPHY		151

LIST OF TABLES

TABLE	DESCRIPTION	PAGE
3.1	Three-dimensional Hierarchical Logit Models with Corresponding Loglinear Models	98
3.2	Four-dimensional Hierarchical Logit Models with Corresponding Loglinear Models	99
4.1	Values of $\hat{\theta}_i$, s.e. $(\hat{\theta}_i)$, $\chi^2_W(2/1) = \hat{\theta}_i^2 / \text{var}(\hat{\theta}_i)$ and $G^2(2/1) / \delta \cdot (H_1)$ for the LFS data under model (4.24)	115

LIST OF FIGURES

FIGURE	DESCRIPTION	PAGE
4.1	Normal Probability Plot of $\tilde{G}_1$	116
4.2	Index Plot of $\tilde{G}_1$	116
4.3	Index Plot of $\{\hat{\theta}_0 - \hat{\theta}_0(-l)\}/s.e.(\hat{\theta}_0)$	117
4.4	Index Plot of $\{\hat{\theta}_1 - \hat{\theta}_1(-l)\}/s.e.(\hat{\theta}_1)$	117
4.5	Index Plot of $\{\hat{\theta}_2 - \hat{\theta}_2(-l)\}/s.e.(\hat{\theta}_2)$	118
4.6	Index Plot of $\{\hat{\theta}_3 - \hat{\theta}_3(-l)\}/s.e.(\hat{\theta}_3)$	118
4.7	Index Plot of $\{G^2 - \tilde{G}^2(-l)\}/\hat{\delta}$	119
4.8	Index Plot of $\{G^2 - G^2(-l)\}/\hat{\delta}$	119

CHAPTER 1

INTRODUCTION

1.1 Review of the Literature

In recent years major advances have been made in the development of techniques for the analysis of multidimensional cross-classified categorical data. Under the assumption of loglinear models and multinomial sampling, maximum likelihood estimation has generally been employed along with the Pearson and likelihood ratio test statistics (Bishop, Fienberg and Holland (1975)). Alternatively, weighted least squares with the Wald statistic has been used (Grizzle, Starmer and Koch (1969)). The methodology has gained in popularity because of its similarity in approach to analysis of variance for continuous data and its applicability to a wide variety of situations.

1.1.2 Loglinear models

The *loglinear model*, as described in detail by Bishop, Fienberg and Holland (abbreviated BFH) (1975), has allowed for the estimation of parameters and the testing of hypotheses. Briefly, the loglinear model consists of defining the logarithm of the population proportion in a cell of the table as a linear function of the variables defining the dimensions of the table. BFH, by analogy with analysis of variance, mainly restrict themselves to loglinear models that are *hierarchical* in the sense that higher order terms are defined as deviations from lower order terms and higher order terms cannot be included in the

model unless all the related lower order terms are present.

As an illustration of such hierarchical models, the saturated hierarchical loglinear model in 3 dimensions is

$$\begin{aligned}
 (1.1) \quad \ln \pi_{ijk} = & u && \text{(constant term)} \\
 & + u_1(i) + u_2(j) + u_3(k) && \text{(one-factor terms)} \\
 & + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) && \text{(two-factor interactions)} \\
 & + u_{123}(ijk) && \text{(three-factor interaction)}
 \end{aligned}$$

where π_{ijk} is the proportion in the (i, j, k) th cell, $i=1, 2, \dots, I$; $j=1, 2, \dots, J$; $k=1, 2, \dots, K$. In order to have a one-to-one correspondence between the u terms and the π_{ijk} , the following symmetric constraints are applied:

$$(1.2) \quad 0 = \sum_i u_1(i) = \sum_j u_2(j) = \sum_k u_3(k),$$

$$0 = \sum_i u_{12}(ij) = \sum_j u_{12}(ij) = \sum_i u_{13}(ik) = \dots = \sum_k u_{23}(jk),$$

and

$$0 = \sum_i u_{123}(ijk) = \sum_j u_{123}(ijk) = \sum_k u_{123}(ijk).$$

Many hypotheses about the structure of the cell probabilities are formulated by omitting appropriate terms from the saturated model. Referring again to (1.1), examples of four different independence hypotheses are now described.

(a) Three-factor effect absent.

The model is

$$\ln \pi_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk)$$

This is equivalent to stating that each two-factor effect is the same for all values of the third variable.

(b) Three-factor and one two-factor effect absent

Selecting u_{12} as the absent two-factor effect yields

$$\ln \pi_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{23}(jk) + u_{13}(ik)$$

This model is equivalent to stating that variables 1 and 2 are independent for every level of variable 3, but each is associated with variable 3. Stated in other words, variables 1 and 2 are *conditionally independent*, given the level of variable 3.

(c) Three-factor and two two-factor effects absent.

Selecting u_{12} and u_{13} as the absent two-factor effects yields

$$\ln \pi_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{23}(jk)$$

This model is equivalent to stating that variable 1 is *completely independent* of the other two variables; variables 2 and 3 are associated.

(d) Three-factor and all two-factor effects absent.

The model is

$$\ln \pi_{ijk} = u + u_1(i) + u_2(j) + u_3(k)$$

It represents *complete independence* of all three variables.

The loglinear model of Rao and Scott (1984), which is more general than the hierarchical loglinear model described above, is of the form

$$(1.3) \quad \underline{\mu} = \bar{u}(\underline{\theta})\underline{1} + \underline{X}\underline{\theta},$$

where $\underline{\mu}$ is the T-vector of log cell probabilities $\mu_i = \ln \pi_i$, $\underline{X} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_T)'$ is a known $T \times r$ matrix of rank $r (\leq T-1)$, $\underline{X}'\underline{1} = \underline{0}$, $\underline{\theta}$ is an r-vector of parameters, $\underline{1}$ is a T-vector of 1's, and $\bar{u}(\underline{\theta}) = \ln\{1/(\sum \exp \underline{x}_i'\underline{\theta})\}$. (In the special case of the BFH model, $\underline{X}$ consists of 0's, 1's and -1's and $\underline{\theta}$ consists of u-terms of the type appearing in (1.1).) Under multinomial sampling, it is well-known that the *likelihood equations* appropriate to model (1.3) are

$$(1.4) \quad \underline{X}'\hat{\underline{\pi}} = \underline{X}'(\underline{n}/n),$$

where $\underline{n} = (n_1, n_2, \dots, n_T)'$ is the T-vector of observed frequencies in the sample of size n ($\sum n_t = n$) and $\hat{\underline{\pi}} = \underline{\pi}(\hat{\underline{\theta}})$ is the maximum likelihood estimator of $\underline{\pi} = (\pi_1, \pi_2, \dots, \pi_T)'$ under the model ($\sum \hat{\pi}_t = 1$). (Note: In the remainder of the thesis, a $\hat{\cdot}$ on a parameter denotes an estimate of the parameter.) Two methods in common use for solving (1.4) for $\hat{\underline{\pi}}$ are iterative proportional fitting (IPF) and Newton-Raphson. Newton-Raphson is more often used when the maximum likelihood estimate (abbreviated mle) of $\underline{\theta}$, as well as that of $\underline{\pi}$, is required. In the case of the hierarchical loglinear models of BFH, IPF consists of successive proportional adjustment of cell estimates so that they fit a set of sufficient marginal totals determined by the particular model used; if the model is such that the cell estimates

can be written in closed form, IPF yields the exact estimates $\hat{\pi}$ in one cycle. Of the four hierarchical models in 3 dimensions described previously, iteration is required only for model (a) to obtain $\hat{\pi}_{ijk}$; for the other three models the closed form estimates are

$$(b) \quad \hat{\pi}_{ijk} = (x_{i+k} x_{+jk}) / x_{++k}$$

$$(c) \quad \hat{\pi}_{ijk} = (x_{+jk} x_{i++}) / n$$

and

$$(d) \quad \hat{\pi}_{ijk} = x_{i++} x_{+j+} x_{++k} / n^2$$

where n is the total sample size, x_{ijk} is the number of sample units in the (ijk) th cells, and "+" replacing one or more subscripts indicates sums of elementary cell counts (for example, $x_{i+k} = \sum_j x_{ijk}$). For testing hypotheses, no more than $\hat{\pi}$ is needed, but if the mle's of the individual u-terms are required, they are obtained directly as functions of $\hat{\pi}$.

Since $\underline{n}/n$ is assumed to be multinomial with parameter $\underline{\pi}$, then $\sqrt{n}(\underline{n}/n - \underline{\pi})$ converges in distribution to a random vector with distribution $N_T(0, P)$ where $P = D_{\underline{\pi}} - \underline{\pi} \underline{\pi}'$ and $D_{\underline{\pi}}$ is a diagonal matrix with i th diagonal element being π_i (in short, $D_{\underline{\pi}} = \text{diag}(\underline{\pi})$). It follows from a result of Birch (1964) that the asymptotic covariance matrix of the mle $\hat{\theta}$ is

$$(1.5) \quad C(\hat{\theta}) = n^{-1} (X' P X)^{-1}$$

and the asymptotic covariance matrix of $\hat{\pi} = \underline{\pi}(\hat{\theta})$ is

$$(1.6) \quad C(\hat{\pi}) = PXC(\hat{\theta})X'P.$$

When testing hypotheses about the loglinear model, three statistics have been in common use: the *Pearson chi-squared test statistic* x^2 , the *likelihood ratio test statistic* G^2 and the *Wald statistic* x_W^2 . For testing the hypothesis $H: \underline{\mu} = \underline{u}(\underline{\theta})\underline{1} + X\underline{\theta}$, the forms of x^2 and G^2 are

$$(1.7) \quad x^2 = n \sum_{i=1}^T (n_i/n - \hat{\pi}_i)^2 / \hat{\pi}_i = (\underline{n}/n - \hat{\pi})' D_{\hat{\pi}}^{-1} (\underline{n}/n - \hat{\pi})$$

and

$$(1.8) \quad G^2 = 2 \sum_{i=1}^T n_i \ln(n_i/n\hat{\pi}_i),$$

where $D_{\hat{\pi}} = \text{diag}(\hat{\pi})$. In order to construct the parallel Wald statistic, let C be any $T \times (T-1-r)$ matrix of rank $(T-1-r)$ with $C'X = \underline{0}$ and $C'\underline{1} = \underline{0}$. Then the hypothesis H is equivalent to $H': \underline{\phi} = C'\underline{\mu} = \underline{0}$ and the Wald statistic is

$$x_W^2 = \hat{\underline{\phi}}' (\hat{C}(\hat{\underline{\phi}}))^{-1} \hat{\underline{\phi}}$$

where $\hat{\underline{\phi}} = C'\hat{\underline{\mu}}$, $\hat{\underline{\mu}} = (\ln \hat{\pi}_1, \ln \hat{\pi}_2, \dots, \ln \hat{\pi}_T)'$, and $\hat{C}(\hat{\underline{\phi}})$ is the estimated asymptotic covariance matrix of $\hat{\underline{\phi}}$, determined as follows. Since, by the delta method, $\hat{\underline{\mu}} - \underline{\mu} \sim D_{\hat{\pi}}^{-1}(\hat{\pi} - \pi)$ (where " $\sim$ " denotes asymptotic equivalence), it follows that

$$(1.9) \quad C(\hat{\underline{\phi}}) = C'D_{\hat{\pi}}^{-1}C(\hat{\pi})D_{\hat{\pi}}^{-1}C = n^{-1}CX(X'PX)^{-1}X'C.$$

The estimator $\hat{C}(\hat{\underline{\phi}})$ is obtained from (1.9) by replacing P with $\hat{P} = D_{\hat{\pi}} - \hat{\pi}\hat{\pi}'$. All three statistics, x^2 , G^2 and x_W^2 , are

asymptotically equivalent under H with asymptotic distribution being χ^2_{T-r-1} (i.e. a central chi-squared with $T-r-1$ degrees of freedom).

Denoting the model (1.3) by M_1 , let $X = (X_1, X_2)$ and $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2)'$ where X_1 is $T \times s$, X_2 is $T \times u$, $\underline{\theta}_1$ is $s \times 1$ and $\underline{\theta}_2$ is $u \times 1$ ($s+u = r$, $X_1' \underline{1} = \underline{0}$, $X_2' \underline{1} = \underline{0}$). It is often of interest to test the hypothesis $H : \underline{\theta}_2 = \underline{0}$, thus leading to the reduced model

$$M_2: \quad \underline{\mu} = \bar{u} \begin{pmatrix} \underline{\theta}_1 \\ \underline{0} \end{pmatrix} \underline{1} + X_1 \underline{\theta}_1. \quad (1.10)$$

M_2 is said to be nested within M_1 . Under M_2 , mle's of $\underline{\theta}_1$ and $\underline{\pi}$, denoted by $\hat{\underline{\theta}}_1$ and $\hat{\underline{\pi}}$, are the solutions of the maximum likelihood equations

$$X_1' \hat{\underline{\pi}} = X_1' (\underline{n}/n). \quad (1.11)$$

To test the fit of M_2 , given M_1 , again Pearson chi-squared, likelihood ratio and Wald test statistics have been used. These statistics take the forms

$$\chi^2 = n \sum_{i=1}^T (\hat{\pi}_i - \hat{\pi}_i)^2 / \hat{\pi}_i = n(\hat{\underline{\pi}} - \hat{\underline{\pi}})' D_{\hat{\underline{\pi}}}^{-1} (\hat{\underline{\pi}} - \hat{\underline{\pi}}),$$

$$G^2 = 2n \sum_{i=1}^T \hat{\pi}_i \ln(\hat{\pi}_i / \hat{\pi}_i)$$

$$\text{and } \chi^2_W = \hat{\underline{\phi}}' (\hat{C}(\hat{\underline{\phi}}))^{-1} \hat{\underline{\phi}}$$

where $\hat{\underline{\phi}} = C_1' \hat{\underline{\mu}}$, $\hat{C}(\hat{\underline{\phi}}) = n^{-1} C_1' X (X' P X)^{-1} X' C_1$ and C_1 is any $T \times u$ matrix of rank u with $C_1' X_1 = \underline{0}$, $C_1' \underline{1} = \underline{0}$ and $C_1' X_2$ nonsingular.

Again, X^2 , G^2 , and X^2 have the same asymptotic distribution under H , which is χ^2_u . Note that " X^2 " represents a test statistic while " X " denotes a matrix.

Until this point, it has been assumed that the data was collected by multinomial sampling. However, a slightly more complex method of sampling is the product multinomial, which would fix a set of marginal totals in the table of counts. BFH (1975) showed that, under product multinomial sampling, the mle's of the cell proportions, π_i , and the u -terms in a hierarchical loglinear model are unchanged (from the values obtained under multinomial sampling) as long as the parameters corresponding to the fixed margin are in the model. This condition ensures that the estimated expected frequencies sum to the fixed marginal totals.

In recent years there has been a growing trend to use survey data for more than the estimation of descriptive parameters. Because of the availability of survey data in tabular form and computer statistical packages, the methodology developed for multinomial samples is easy to apply. However, due to cost and operational constraints, most surveys consist of more complex sampling schemes than the multinomial. As an example, the Canada Health Survey (see Hidiroglou and Rao (1981)) involved 12,000 households (40,000 persons) from 100 sample geographical clusters. These 100 clusters were allocated initially to the provinces proportional to the square root of their 1971 census populations. Three major strata - major cities, other urban areas, and rural areas -

were formed within each province. Quebec and Ontario were further subdivided into three health regions. The allocation of the provincial clusters to the major strata was done proportional to their respective 1971 Census populations with the requirement that the minimum allocation to a stratum be two clusters. Further guidelines regulated the choice of the clusters within the strata and the households within the clusters. Obviously, the clustering, combined with the stratification and multistage sampling in such a complex survey design leads to correlations between the units.

Many subject-matter researchers choose to ignore the sampling plan and proceed as if the statistical tests will behave the same as under multinomial sampling. They calculate their point estimates and test statistics in the usual manner, except that $\underline{n}/n$, the sample estimate of $\underline{\pi}$ under multinomial sampling, is replaced by $\hat{\underline{p}}$, an estimate of $\underline{\pi}$ derived from the actual survey design used. Generally, $\hat{\underline{p}}$ is a ratio estimate with the property that $\sqrt{n}(\hat{\underline{p}} - \underline{\pi})$ converges in distribution to a $N_{\underline{T}}(\underline{0}, V)$ random vector as $n \rightarrow \infty$. As an example, the "pseudo mle" $\hat{\underline{\pi}}$ is obtained from the "pseudo likelihood equations"

$$(1.12) \quad X' \hat{\underline{\pi}} = X' \hat{\underline{p}},$$

(derived from (1.4)), the Pearson chi-squared statistic for testing the goodness of fit of model (1.3) is

$$(1.13) \quad X^2 = n(\hat{\underline{p}} - \hat{\underline{\pi}})' D_{\hat{\underline{\pi}}}^{-1} (\hat{\underline{p}} - \hat{\underline{\pi}}),$$

(derived from (1.7)), and the corresponding likelihood ratio test

statistic is

$$G^2 = 2n \sum \hat{p}_i \ln(\hat{p}_i / \hat{\pi}_i) \quad (\text{derived from (1.8)}).$$

Rao and Scott (1981, 1984), Holt, Scott and Ewings (1980), Fellegi (1980), Brier (1978), Hidiroglou and Rao (1981, 1983) and others have shown that the nominal significance levels of traditional chi-square tests can be severely distorted by clustering, thus yielding misleading results. It therefore seems essential that the sampling plan be taken into account when assessing survey data. Two approaches have been utilized by various authors.

First, the Wald statistic approach generalizes easily to complex sampling designs, as illustrated by Koch, Freeman and Freeman (1975), provided that a consistent estimate $\hat{V}$ of the covariance matrix of the cell proportions is available. As an example, it was noted earlier that testing the fit of model (1.3) was equivalent to testing $H' : \underline{\phi} = C'\underline{\mu} = 0$, where C is any $T \times (T-1-r)$ matrix of rank $(T-1-r)$, with $C'X = 0$ and $C'1 = 0$. The Wald statistic for testing H' is

$$x_W^2 = \hat{\underline{\phi}}' (\hat{C}(\hat{\underline{\phi}}))^{-1} \hat{\underline{\phi}}$$

where $\hat{\underline{\phi}} = C'\hat{\underline{\mu}}$, $\hat{\underline{\mu}} = (\ln \hat{\pi}_1, \ln \hat{\pi}_2, \dots, \ln \hat{\pi}_T)$, $\hat{\underline{\pi}} = (\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_T)$ is the solution of (1.12) and $\hat{C}(\hat{\underline{\phi}})$ is the estimated asymptotic covariance matrix of $\hat{\underline{\phi}}$ determined as follows. Since

$$(1.14) \quad C(\hat{\underline{\phi}}) = C'D_{\hat{\underline{\pi}}}^{-1} C(\hat{\underline{\pi}}) D_{\hat{\underline{\pi}}}^{-1} C,$$

where

$$C(\hat{\underline{\pi}}) = n^{-1} P X (X' P X)^{-1} (X' V X) (X' P X)^{-1} X' P,$$

then $\hat{C}(\hat{\phi})$ is obtained by replacing π by $\hat{\pi}$ and V by $\hat{V}$ in (1.14).

The property of the Wald statistic that is so useful in hypothesis testing is that its asymptotic distribution is central chi-squared.

However, when working with published tables, the researcher generally does not have access to the appropriate information for a covariance matrix estimate. Furthermore, even when the information is available, the Wald statistic may not behave well. Hidioglou and Rao (1983) found that the Wald statistic, when applied to a data set from the Canada Health Survey, behaved erratically, perhaps due to the instability of the estimated covariance matrix. As well, Thomas and Rao (1984), in a Monte Carlo study on the exact significance levels of several goodness-of-fit statistics under cluster sampling, found that the Wald statistic leads to unacceptably high levels, particularly as the number of clusters decreases and the number of categories increases.

A second approach to take account of the survey design is to modify the customary Pearson chi-squared statistic X^2 (or the likelihood ratio statistic G^2). Rao and Scott (1981, 1984) and Scott and Rao (1981) have examined X^2 for such hypotheses as goodness of fit, independence in a two-dimensional table, homogeneity of proportions across populations and nested loglinear models. In each case, they showed that the asymptotic distribution of the appropriate X^2 is a weighted sum, $\sum_{i=1}^{\ell} \delta_i Z_i^2$, of independent χ_1^2 variables. The weights δ_i are the eigenvalues of an $\ell \times \ell$ matrix whose size ℓ and form depends on the particular hypothesis being tested but which involves

the asymptotic covariance matrix V/n of the cell proportions

$\hat{p}$. For testing the nested loglinear hypothesis $H: \theta_2 = 0$, given that $\mu = \bar{u}(\theta)1 + x_{1-1}\theta_1 + x_{2-2}\theta_2$, for example, the δ_i are the u eigenvalues of $(\bar{X}_2' P \bar{X}_2)^{-1} (\bar{X}_2' V \bar{X}_2)$, where $\bar{X}_2 = (I - X_1(X_1' P X_1)^{-1} X_1' P) X_2$. Observing that $X^2/\delta_\cdot$ has the same asymptotic expectation as a X_θ^2 variable (where $\delta_\cdot = \lambda^{-1} \sum \delta_i$), they suggested a first-order correction to X^2 of the form $X^2/\hat{\delta}_\cdot$, where $\hat{\delta}_\cdot$ is a survey estimate of $\delta_\cdot$.

For several of the hypotheses that they examined, they found that estimation of $\delta_\cdot$ does not require knowledge of the complete covariance matrix. In particular, for the goodness of fit hypothesis, $\pi = \pi_0$, they need only the design effects (see Appendix A6) for individual cells; in the population homogeneity problem they need the design effects (abbreviated "deffs") of proportions in each population; for testing the fit of hierarchical loglinear model M_2 nested with M_1 , they require only the cell deffs and the deffs of collapsed tables, provided that the likelihood equations for M_1 and M_2 appropriate to multinomial sampling admit explicit solutions.

Rao and Scott (1984) made further recommendations for approximating $\delta_\cdot$ in the case of testing M_2 nested within M_1 , when at least one of M_2 and M_1 does not admit an explicit solution and when the full covariance matrix is unknown. They suggest taking the $\delta_\cdot$ corresponding to a hypothesis "closest" to H (in the sense of having approximately the same u -terms in the model) and admitting explicit solution. However, when nesting is within the saturated model, they suggest an

approximation to $\delta_{\cdot}$ given by $\hat{\delta}_{\cdot} = \lambda_{\cdot}$, where $(T-1)\lambda_{\cdot} = \sum (1-\pi_i)d_i$ and d_i is the deff of the i^{th} cell. Limited empirical studies by Holt, Scott and Ewings (1980) and Hidiroglou and Rao (1981, 1983) have indicated that $X^2/\hat{\lambda}_{\cdot}$ and a similar correction $X^2/\hat{d}_{\cdot}$ suggested by Fellegi (1980), where $\hat{\lambda}_{\cdot}$ is a survey estimate of $\lambda_{\cdot}$ and $\hat{d}_{\cdot}$ is a survey estimate of $d_{\cdot} = T^{-1} \sum_{i=1}^T d_i$ may not perform well. It may be noted that $\hat{d}_{\cdot}$ and $\hat{\lambda}_{\cdot}$ are independent of the hypothesis being tested.

Better corrections to X^2 than $\delta_{\cdot}$ are possible if more information is available on the δ_i 's. If $\sum \hat{\delta}_i^2$ is also known, which generally requires an estimate of the full covariance matrix, then, following Satterthwaite (1946), $X^2/(1+a^2)\delta_{\cdot}$ may be treated as a χ_v^2 variable, where $v = \ell/(1+a^2)$ and a^2 is the coefficient of variation of the δ_i 's. In the Monte Carlo study by Thomas and Rao (1984), this statistic performed well.

1.1.3 Response Models

Loglinear models are primarily designed for investigating relationships among several categorical variables. There are, however, many sets of cross-classified data in which one of the classifying variables would obviously be considered a response variable and the others would be considered explanatory. Frequently, the response variable is binary, with such classifications as "success and failure" or "employed and unemployed". In fact, for this type of data, the full contingency table may not be published; instead, the presentation may be in the

form of a table of rates (such as success rates), so that the dimensionality of the original table of frequencies is reduced by one.

Because the loglinear methodology is so readily available, many researchers attempt to use a loglinear model with the table of frequencies, regardless of their questions of primary interest. However, another possible approach is to work directly with models for the logits (or other transformations) of the success proportions in the table of reduced dimension. It is a well-known fact (see, for example, Bishop^o (1969)) that for any hierarchical logit model there is a hierarchical loglinear model that gives the same estimated expected cell frequencies. Yet, Kalbfleisch (1984) succinctly argues for an analysis in terms of logit models:

"The loglinear analysis treats all of the classification variables symmetrically and attempts to model simultaneously both the relationships among the explanatory variables and their effects on survival. The result of trying to do everything at once is to greatly complicate the analysis and its interpretation. Furthermore, inferences about the effects of the explanatory variables on survival are now dependent upon what is assumed about the relationships among the explanatory variables. On the other hand, the logistic analysis is conditional, and hence inferences from it do not depend upon assumptions as to how the explanatory variables are interrelated. Indeed, the sample frequencies of these variables need not even be representative of their population frequencies. The conditional logistic analysis requires less stringent assumptions and in that sense is more general than the unconditional loglinear analysis."

A further argument against a loglinear analysis, as pointed out by Kalbfleisch (1984), is that a model for a transformation of the success rates other than the logit transformation (say, the probit transformation or the complementary log-log) may be appropriate and such models do not have an equivalent loglinear form.

Traditionally, direct analysis of success rates has been based on the assumption of product binomial sampling. Under this assumption, a limited number of well-known results are now given. Suppose that P_i represents the true success rate in the i^{th} domain, $i = 1, 2, \dots, R$ and that the model $P_i = m_i(\underline{\theta})$ is under consideration, where $\underline{\theta}$ is an r -vector of unknown parameters and m_i is a function of known form. Also, assume that each domain is sampled independently, with a random sample of size n_i , of which y_i are successes, being taken from the i^{th} domain. The maximum likelihood estimates of $\underline{\theta}$ and P_i , denoted by $\hat{\underline{\theta}}$ and $\hat{m}_i = m_i(\hat{\underline{\theta}})$ respectively, are obtained from the likelihood equations

$$(1.14) \quad \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\hat{\underline{\theta}}} D(\underline{m}(\hat{\underline{\theta}}))^{-1} D(\underline{1}-\underline{m}(\hat{\underline{\theta}}))^{-1} D(\underline{n}/n) \underline{q} = \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\hat{\underline{\theta}}} D(\underline{1}-\underline{m}(\hat{\underline{\theta}}))^{-1} (\underline{n}/n)$$

where $\frac{\partial \underline{m}}{\partial \underline{\theta}}$ is the $R \times r$ matrix having $(i, j)^{\text{th}}$ element equal to $\partial m_i(\underline{\theta}) / \partial \theta_j$, $n = \sum n_i$, $\underline{n} = (n_1, n_2, \dots, n_R)'$, $\underline{q} = (y_1/n_1, y_2/n_2, \dots, y_R/n_R)'$, $\underline{m}(\underline{\theta}) = (m_1(\underline{\theta}), \dots, m_R(\underline{\theta}))'$ and $D(\underline{x})$ denotes $\text{diag}(\underline{x})$ for any vector $\underline{x}$. In the case of the logit model, $\ln(P_i/(1-P_i)) = \underline{x}_i' \underline{\theta}$, these likelihood equations have the form

$$(1.15) \quad \underline{x}' D(\underline{n}/n) \underline{m}(\hat{\underline{\theta}}) = \underline{x}' D(\underline{n}/n) \underline{q},$$

where $\underline{x} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_R)'$.

The Pearson chi-squared statistic χ^2 used to test the hypothesis

$H: \underline{P} = \underline{m}(\underline{\theta})$, has the form

$$(1.16) \quad \chi^2 = \sum_{i=1}^R (y_i - n_i m_i(\hat{\underline{\theta}}))^2 / [n_i m_i(\hat{\underline{\theta}}) (1 - m_i(\hat{\underline{\theta}}))]$$

$$= n(\underline{q} - \underline{m}(\hat{\theta}))' D(n/n) D(\underline{m}(\hat{\theta}))^{-1} D(\underline{1} - \underline{m}(\hat{\theta}))^{-1} (\underline{q} - \underline{m}(\hat{\theta}))$$

where χ^2 is asymptotically distributed as a χ^2_{R-r} variable when H is true (see Appendix A7).

Finally, the asymptotic covariance matrices of the mle's $\hat{\theta}$ and $\hat{\underline{m}} = \underline{m}(\hat{\theta})$ are given by

$$C(\hat{\theta}) = n^{-1} \left(\frac{\partial \underline{m}'}{\partial \theta} E \frac{\partial \underline{m}}{\partial \theta} \right)^{-1}$$

and

$$C(\underline{m}(\hat{\theta})) = \frac{\partial \underline{m}}{\partial \theta} C(\hat{\theta}) \frac{\partial \underline{m}}{\partial \theta}$$

where $E = D(\underline{W}) D(\underline{m}(\theta))^{-1} D(\underline{1} - \underline{m}(\theta))^{-1}$ and $\underline{W}$ is the vector of the proportions of the population in each of the R domains.

Instead of the response variable being binary, it could have more than two categories. Fienberg (1977) and Haberman (1982) suggest extensions of the logit transformation for this type of data. McCullagh and Nelder (1983) make similar suggestions, but also present models to take account of an ordering in the scale of measurement of the response variable. In all cases, multinomial (product multinomial) sampling is assumed when the properties of the estimated models are developed.

After a model has been fitted to a table of cross-classified data, it is often desirable to make a critical assessment of the fit by identifying any outlying cell proportions and influential points in the factor space. Pregibon (1981) and Cook and Weisberg (1982) make several suggestions for such an assessment for a logit model. However,

they assume binomial sampling in obtaining their results.

1.2 Outline of Thesis

1. Rao and Scott (1984) developed the asymptotic null distribution of the Pearson chi-squared test statistic X^2 (and the likelihood ratio test statistic G^2) for nested loglinear models of form (1.3), under a general sampling plan. They also obtained a simple correction, $\hat{\delta}$, to X^2 (or G^2) which requires only the knowledge of cell deffs and the deffs for marginal tables whenever the models admit direct solution of the likelihood equations under multinomial sampling. In Section 1 of Chapter 2, the form of this correction $\hat{\delta}$ will be derived, in the case of one set of fixed marginal totals of the observed frequencies in the sample. This case arises frequently in practice, since a large survey design often consists of independent samples from several populations. The correction will be expressed in terms of within-population cell proportions and associated deffs. The published tables usually report the latter quantities.
2. The form of Rao and Scott's loglinear model does not cover all hypotheses of interest about the relationships among the cell proportions. As an example, marginal homogeneity in a two-dimensional table is not covered. Thus, a model more general than the loglinear (but having the loglinear as a special case) will be examined in Section 2 of Chapter 2. Following the pattern of Rao and Scott (1984), the asymptotic theory of the Pearson chi-squared statistic for testing nested hypotheses with the model will be developed. Corrections to the statistic

to take account of the true sampling plan will be obtained. Finally, two examples not covered by Rao and Scott's model will be presented.

3. When multidimensional survey data consists of a binary response variable and several explanatory variables, little work has been done on the properties of chi-squared test statistics for models of functions of the binary response probabilities. In Chapter 3, the asymptotic properties of the Pearson chi-squared statistic χ^2 for nested hypotheses are developed under the general model form $P_i = m_i(\theta)$, where P_i is the rate of response in the i^{th} domain, θ is an unknown parameter and m_i is a function satisfying specified regularity conditions. Upper bounds on the first order correction δ to χ^2 that involve only deffs of the response proportions are developed.

For those binary response models $P_i = m_i(\theta)$ that can be expressed in the form $g(P_i) = x_i'\theta$, Wald statistics are also developed.

4. As noted in Section 1.1.2, loglinear models with explicit (or direct) cell estimates yield a simple correction to the Pearson chi-squared statistic. Chapter 3 concludes with a specification of the logit models having an equivalent specification as a loglinear model with direct estimates. For these logit models, the first order correction δ to the Pearson chi-squared statistic may be expressed in terms of deffs of variables in the logit specification.

5. In Chapter 4, diagnostics for assessing a binary response model are developed, with the survey design being taken into account. As

well, a logit model is fitted to a set of binary response data from the Canadian Labour Force Survey, and several diagnostics for the fit are examined.

6. In Chapter 5, models for multidimensional data consisting of a polytomous response variable and several explanatory variables are introduced. The asymptotic properties of the Pearson chi-squared statistic X^2 for nested hypotheses are developed. Two particular polytomous response models, one suggested by Haberman (1982) and the other by McCullagh and Nelder (1983), are examined.

7. Chapter 6 contains suggestions for further research.

CHAPTER 2

EXTENSIONS OF RAO-SCOTT THEORY FOR LOGLINEAR MODELS

2.1 Introduction

The recent development of methodology for loglinear models has greatly aided in the analysis of multidimensional tables of data. Many analytical tools are now available (for example, BPH (1975), Fienberg (1977), and Haberman (1978)) and can be readily implemented by various computer software packages. However, in much of this literature, multinomial or product multinomial sampling is assumed, thus ignoring the clustering present in most complex sample surveys. To date, researchers wishing to take account of the effect of the survey design have either been using a Wald statistic (eg. Koch et al. (1975)) which incorporates the sample design into the estimation of a covariance matrix, or have been making corrections to the usual Pearson chi-squared statistic (eg. Rao and Scott (1981, 1984)). This latter method is often more easily implemented since, in many situations, estimates of the full covariance matrix are not required.

This chapter will cover two extensions of the work of Rao and Scott (1984). First, continuing with the loglinear model, a correction, δ , to the Pearson chi-squared statistic will be developed in the case of a general sampling plan with one set of fixed marginal totals of the observed frequencies, where the loglinear model being used admits direct estimates. The u-terms corresponding to the fixed margins must all be included in the model, as outlined in Chapter 1. Section 2.2 will

describe required notation and background results for the loglinear model. Section 2.3 will develop the general form of δ in terms of estimated cell proportions and their marginal totals within cell populations corresponding to fixed margins, and of the associated design effects (abbreviated "deffs"), as usually reported in published tables.

Second, in Section 2.4, the asymptotic theory of the Pearson chi-squared statistic will be developed under a specified sampling scheme when models more general than the loglinear are used. As in the case of loglinear models, the parameters of these models will be estimated by maximum likelihood methods appropriate to multinomial sampling. Corrections to the Pearson chi-squared statistic to take account of the true sampling plan will be obtained.

2.2 Notation and Background Results for the Loglinear Model

Let $\underline{\pi} = (\pi_1, \pi_2, \dots, \pi_t, \dots, \pi_T)'$ denote the T-vector of population cell proportions in a multiway table, with the obvious restriction $\sum \pi_t = 1$. (In general practice, individual terms of $\underline{\pi}$ will have multiple subscripts denoting the levels of each dimension, rather than a single subscript between 1 and T). Suppose a sample $\tilde{s}$ of n ultimate units is drawn according to a specified survey design $p(\tilde{s})$. Let $\hat{\underline{p}} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_T)'$ denote a consistent estimator of $\underline{\pi}$ under the design, with $\sum \hat{p}_t = 1$. Also assume that a central limit theorem exists for $p(\tilde{s})$ so that $\sqrt{n}(\hat{\underline{p}} - \underline{\pi})$ is asymptotically distributed as a $N(0, V)$ vector as $n \rightarrow \infty$. (See Rao and Scott (1981) for references pertaining to central limit theorems in sampling).

A loglinear model M_1 for π may be written as

$$(2.1) \quad \underline{\mu} = \bar{u}(\underline{\theta}) \underline{1} + X \underline{\theta}$$

where $\underline{\mu}$ is a T -vector with $\mu_t = \ln \pi_t$, $\underline{1}$ is a T -vector of 1's, $X = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_r)'$ is a $T \times r$ matrix of full rank $r (\leq T-1)$ and satisfying $X' \underline{1} = \underline{0}$, $\underline{\theta}$ is an r -vector of unknown parameters and $\bar{u}(\underline{\theta}) = \ln[1/\sum \exp(\underline{x}_t' \underline{\theta})]$. The form of the normalizing factor $\bar{u}(\underline{\theta})$ is determined by the restriction $\sum \pi_t = 1$. The u -notation for loglinear models employed by Bishop, Fienberg and Holland (1975) satisfies the form of (2.1). As an example, the independence model for a 2×3 table may be written as

$$\mu_{ij} = \bar{u} + u_{1(i)} + u_{2(j)} \quad i=1,2 \text{ and } j=1,2,3$$

where $u_{1(2)} = -u_{1(1)}$ and $u_{2(3)} = -u_{2(1)} - u_{2(2)}$ or as

$$\underline{\mu} = \bar{u} \underline{1} + \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} u_{1(1)} \\ u_{2(1)} \\ u_{2(2)} \end{bmatrix}$$

where $\underline{\theta} = (u_{1(1)}, u_{2(1)}, u_{2(2)})'$ and X is 6×3 of rank 3.

If the sample $\tilde{s}$ were drawn by multinomial sampling, it is well-known that the likelihood equations for π under model (2.1) would be

$$(2.2) \quad X' \tilde{\pi} = X' (n/n)$$

where $\tilde{\pi} = \pi(\tilde{\theta})$ are the maximum likelihood estimates and $\underline{n}$ is the

T-vector of cell counts n_t , $\sum n_t = n$. Since the likelihood equations under model (2.1) for a general survey design $p(\tilde{s})$ are often very complex or impossible to derive, "pseudo maximum likelihood estimates" of $\underline{\pi}$ (or of $\underline{\theta}$) are obtained by replacing (n/n) in (2.2) by $\hat{p}$ and solving equations

$$(2.3) \quad X' \underline{\hat{\pi}} = X' \hat{p}.$$

The resulting estimates $\underline{\hat{\pi}}$ are consistent under the model.

To test the goodness of fit of model M_1 , the Pearson chi-squared statistic given by

$$X^2 = n \sum (\hat{p}_i - \hat{\pi}_i)^2 / \hat{\pi}_i = n (\hat{p} - \underline{\hat{\pi}})' D_{\hat{\pi}}^{-1} (\hat{p} - \underline{\hat{\pi}})$$

or the loglikelihood ratio statistic given by

$$G^2 = 2n \sum \hat{p}_t \ln(\hat{p}_t / \hat{\pi}_t)$$

are often employed, where $D_{\underline{\pi}} = \text{diag}(\underline{\pi})$. The equivalent asymptotic distributions of X^2 and G^2 under M_1 are a special case of asymptotic distributions of the test statistics for a nested hypothesis described below.

Suppose $\underline{\theta} = (\underline{\theta}_1', \underline{\theta}_2')'$ where $\underline{\theta}_1$ is $s \times 1$ and $\underline{\theta}_2$ is $u \times 1$ ($s+u = r$). Also, let $X = (X_1, X_2)'$ where X_1 is $T \times s$ and X_2 is $T \times u$. It is often of interest to test the hypothesis $H: \underline{\theta}_2 = \underline{0}$, thus leading to the reduced model M_2 :

$$(2.4) \quad \underline{\mu} = \bar{u} \begin{pmatrix} \underline{\theta}_1 \\ \underline{0} \end{pmatrix} \underline{1} + X_1 \underline{\theta}_1.$$

M_2 is said to be nested within M_1 . Under M_2 , "pseudo maximum likelihood estimates" - or pseudo mle's, for short - of $\underline{\theta}_1$ and $\underline{\pi}$

are denoted by $\hat{\theta}_1$ and $\hat{\pi} = \pi(\hat{\theta}_1)$ and are obtained by solving equations

$$X_1' \hat{\pi} = X_1' \hat{p}.$$

For testing H , the Pearson chi-squared statistic given by

$$(2.5) \quad X^2 = n \sum (\hat{\pi}_t - \hat{\pi})^2 / \hat{\pi}_t = n (\hat{\pi} - \hat{\pi})' D_{\hat{\pi}}^{-1} (\hat{\pi} - \hat{\pi}),$$

or the likelihood ratio statistic

$$\begin{aligned} (2.6) \quad G^2 &= 2n \sum \hat{\pi}_t \ln(\hat{\pi}_t / \hat{\pi}_t) \\ &= 2n \sum \hat{p}_t \ln(\hat{p}_t / \hat{\pi}_t) - 2n \sum \hat{p}_t \ln(\hat{p}_t / \hat{\pi}_t) \quad (\text{BFH (1975), p.126}) \\ &= G_2^2 - G_1^2 \end{aligned}$$

are often employed. Under H , the statistics X^2 and G^2 may be shown to be asymptotically equivalent, using a proof of a form similar to lemma 14.9.1, BFH (1975). Rao and Scott (1984) showed that, under H , X^2 (or G^2) is asymptotically equivalent to $\sum_{i=1}^u \delta_i Z_i^2$ where the Z_i^2 's are independent χ_1^2 random variables and the δ_i 's are the eigenvalues (all greater than 0) of the matrix $(\bar{X}_2' P \bar{X}_2)^{-1} (\bar{X}_2' V \bar{X}_2)$, where $\bar{X}_2 = (I - X_1 (X_1' P X_1)^{-1} X_1' P) X_2$ and $P = D_{\pi} - \pi \pi'$.

In the same paper, the authors suggested that, for practical purposes, it is often adequate to treat $X^2 / \hat{\delta}_.$ or $G^2 / \hat{\delta}_.$ as χ_u^2 under H , where $u \hat{\delta}_. = \text{tr}(\bar{X}_2' P \bar{X}_2)^{-1} (\bar{X}_2' V \bar{X}_2)$ and $\hat{\delta}_.$ is an estimate of $\delta_.$. They showed that, in those cases where models M_1 and M_2 admitted explicit solutions for $\hat{\pi}$ and $\hat{\pi}$ respectively, $\delta_.$ could be expressed simply in terms of the variance estimates or deffs of single cells and marginal tables. Details of this result will now be provided.

Using the notation of BFH(1975), a model yielding direct estimates may be expressed in the form

$$(2.7) \quad \pi_{\theta} = [\prod_i \pi_{\theta_i}] / [\prod_j \pi_{\phi_j}]$$

where θ denotes an arbitrary cell, π_{θ_i} are marginal totals of the π_{θ} corresponding to the minimal sufficient statistics (under multinomial sampling) for the model, and π_{ϕ_j} are also marginal totals of the π_{θ} with $\phi_j = \theta_i \cap \theta_k$ for some i and k . As an example, the conditional independence of variables 1 and 2 in a three-way table may be expressed as $\pi_{ijk} = (\pi_{i+k} \pi_{+jk}) / \pi_{++k}$. Here, $\theta = \{i, j, k\}$, $\theta_1 = \{i, k\}$, $\theta_2 = \{j, k\}$ and $\phi_1 = \{k\}$. For model (2.7) the pseudo mle of π_{θ} is

$$(2.8) \quad \hat{\pi}_{\theta} = [\prod_i \hat{p}_{\theta_i}] / [\prod_j \hat{p}_{\phi_j}]$$

where $\hat{p}_{\theta_i}$ and $\hat{p}_{\phi_j}$ are the marginal totals of $\hat{p}_{\theta}$ corresponding to θ_i and ϕ_j respectively. Suppose that both models M_1 and M_2 can be expressed in the form of (2.7). Then Rao and Scott showed that

$$(2.9) \quad u\delta = \overline{EG}_2^2 - \overline{EG}_1^2$$

(with $\overline{E}$ denoting asymptotic expectation) and that

$$(2.10) \quad \overline{EG}_1^2 = \sum_{\theta} (1 - \pi_{\theta}) d_{\theta} - \sum_i [\sum_{\theta_i} (1 - \pi_{\theta_i}) d_{\theta_i}] + \sum_j [\sum_{\phi_j} (1 - \pi_{\phi_j}) d_{\phi_j}],$$

where d_{θ} is the deff of $\hat{p}_{\theta}$, d_{θ_i} is the deff of $\hat{p}_{\theta_i}$ and d_{ϕ_j} is the deff of $\hat{p}_{\phi_j}$ under M_1 , i.e.

$$d_{\theta} = \frac{nV(\hat{p}_{\theta})}{\pi_{\theta}(1-\pi_{\theta})}, \quad d_{\theta_i} = \frac{nV(\hat{p}_{\theta_i})}{\pi_{\theta_i}(1-\pi_{\theta_i})}, \quad \text{and} \quad d_{\phi_j} = \frac{nV(\hat{p}_{\phi_j})}{\pi_{\phi_j}(1-\pi_{\phi_j})},$$

where π_{θ} is given by the expression of form (2.7) for M_1 . A similar result holds for $\overline{EG}_2^2$.

As noted by the authors, the result of the previous paragraph also holds true for the case of one set of fixed marginal totals of the observed frequencies in the sample, provided that the u-terms corresponding to the fixed margins are included in the model. However, a researcher will often have estimated cell proportions and marginal totals, plus the corresponding deffs, just within each population corresponding to the fixed margins. Therefore, it would be useful to express δ in terms of these given quantities. This is done in Section 2.3.

2.3 δ for Models with Direct Estimates and for Sampling with One Set of Fixed Marginal Totals

Suppose that $\theta = \gamma + \mu$, where θ stands for the set of all indexing variables and γ stands for the subset of those variables indexing the set of fixed marginal frequencies. For example, in a three-way table where n_{i++} are fixed, $\theta = \{i, j, k\}$, $\gamma = \{i\}$ and $\mu = \{j, k\}$. Let $P_{\mu}(\gamma)$ denote the proportion of the γ^{th} sample in the μ^{th} cell, and let $\hat{p}_{\mu}(\gamma)$ be a survey estimate of $P_{\mu}(\gamma)$, where $\sum_{\mu} P_{\mu}(\gamma) = \sum_{\mu} \hat{p}_{\mu}(\gamma) = 1$. Published tables, such as those in The Health of Canadians (1981), report the $\hat{p}_{\mu}(\gamma)$. It will be assumed that samples are independent and that deffs of $\hat{p}_{\mu}(\gamma)$ and of marginal

totals of $\hat{p}_\mu(\gamma)$ are also available. Following in the notation of BFH(1975), let n_θ be the observed frequency in a single cell and let n_γ be the sum of frequencies over all variables other than γ . The n_γ are fixed by the survey design and $\sum_\theta n_\theta = \sum_\gamma n_\gamma = n$. Finally,

$$\pi_\theta = (n_\gamma/n) P_\mu(\gamma) \quad \text{and} \quad \hat{p}_\theta = (n_\gamma/n) \hat{p}_\mu(\gamma).$$

As stated in Section 2.2, when testing $H: \theta_2 = 0$ for a log-linear model, use of the statistic x^2/δ . (or G^2/δ .) is often practical, where $u\delta = \overline{EG}_2^2 - \overline{EG}_1^2$. If model M_1 is saturated, $\hat{\pi}_\theta = \hat{p}_\theta$, so that $\overline{EG}_1^2 = G_1^2 = 0$ and $u\delta = \overline{EG}_2^2 = \overline{EG}^2$. For ease of presentation we examine δ in this case since the extension to the unsaturated situation is immediate due to relationship (2.9).

Suppose that model M_2 yields direct estimates, so that it can be expressed in the form

$$(2.11) \quad \pi_\theta = \frac{n_\gamma}{n} P_\mu(\gamma) = [\prod_i \pi_{\theta_i}] / [\prod_j \pi_{\phi_j}].$$

Then, by (2.10),

$$(2.12) \quad u\delta = \sum_\theta (1-\pi_\theta) d_\theta - \sum_i [\sum_{\theta_i} (1-\pi_{\theta_i}) d_{\theta_i}] + \sum_j [\sum_{\phi_j} (1-\pi_{\phi_j}) d_{\phi_j}]$$

where d_θ , d_{θ_i} and d_{ϕ_j} are the deffs of $\hat{p}_\theta$, $\hat{p}_{\theta_i}$ and $\hat{p}_{\phi_j}$ respectively under H . We will now express (2.12) in terms of $\hat{p}_\mu(\gamma)$ and its marginal totals, and their corresponding deffs.

$$(a) \quad (1-\pi_\theta) d_\theta :$$

$$\begin{aligned} (1-\pi_\theta) d_\theta &= nV(\hat{p}_\theta)/\pi_\theta = (n_\gamma/n) n_\gamma V(\hat{p}_\mu(\gamma))/\pi_\theta = n_\gamma V(\hat{p}_\mu(\gamma))/P_\mu(\gamma) \\ &= (1-P_\mu(\gamma)) d(\hat{p}_\mu(\gamma)), \end{aligned}$$

where $P_\mu(\gamma)$ is given by (2.11) and $d(\hat{p}_\mu(\gamma))$ denotes the deff of $\hat{p}_\mu(\gamma)$ under H , i.e.

$$(2.13) \quad d(\hat{p}_\mu(\gamma)) = n_Y V(\hat{p}_\mu(\gamma)) / (P_\mu(\gamma) (1 - P_\mu(\gamma)))$$

$$(b) \quad (1 - \pi_{\theta_i}) d_{\theta_i} \quad \text{or} \quad (1 - \pi_{\phi_j}) d_{\phi_j} :$$

There are several possible configurations for θ_i or ϕ_j , when compared to the variables γ determining the fixed marginals. The

θ_i or ϕ_j can

- (1) equal γ ,
- (2) contain γ plus some other variables,
- (3) contain a portion of γ and no other variables. (A θ_i cannot be of this form since some θ_k must contain all of γ , because the model includes the u terms determined by γ , so that another θ_k consisting of just a portion of γ would be redundant),
- (4) contain a portion of γ plus other variables,
- (5) be independent of γ and contain all other variables.

or

- (6) be independent of γ and contain only a portion of the other variables.

The form of $(1 - \pi_{\theta_i}) d_{\theta_i}$ (or, similarly, $(1 - \pi_{\phi_j}) d_{\phi_j}$) will now be developed under each of these configurations.

- (1) θ_i (or ϕ_j) equals γ .

$$\hat{p}_{\theta_i} = \hat{p}_\gamma = (n_Y/n) \sum_\mu \hat{p}_\mu(\gamma) = n_Y/n = \text{constant.}$$

Therefore $V(\hat{p}_{\theta_i}) = 0$ and hence

$$(1 - \pi_{\theta_i}) d_{\theta_i} = 0.$$

(2) θ_i (or ϕ_j) contains γ plus some other variables.

Recall that $\theta = \gamma + \mu$. Let $\mu = a + b$ and suppose $\theta_i = \gamma + a$.

Then

$$\hat{p}_{\theta_i} = \hat{p}_{\gamma, a} = \sum_b (n_{\gamma}/n) \hat{p}_{a, b}(\gamma) = (n_{\gamma}/n) \hat{p}_{a, +}(\gamma),$$

where $\hat{p}_{a, +}(\gamma)$ denotes $\sum_b \hat{p}_{a, b}(\gamma) = \sum_b \hat{p}_{\mu}(\gamma)$ when $\theta = \gamma + a + b$.

Thus,

$$\begin{aligned} (1 - \pi_{\theta_i}) d_{\theta_i} &= nV((n_{\gamma}/n) \hat{p}_{a, +}(\gamma)) / \pi_{\theta_i} = (n_{\gamma}/n) n_{\gamma} V(\hat{p}_{a, +}(\gamma)) / \pi_{\theta_i} \\ &= n_{\gamma} V(\hat{p}_{a, +}(\gamma)) / p_{a, +}(\gamma) = (1 - p_{a, +}(\gamma)) d(\hat{p}_{a, +}(\gamma)), \end{aligned}$$

where $d(\hat{p}_{a, +}(\gamma))$ denotes the deff of $\hat{p}_{a, +}(\gamma)$.

(3) ϕ_j contains a portion of γ and no other variables.

Suppose that $\theta = \gamma + \mu$, $\gamma = c + d$ and $\phi_j = c$. Then

$$\begin{aligned} \hat{p}_{\phi_j} &= \hat{p}_c = \sum_d \sum_{\mu} (n_{\gamma}/n) \hat{p}_{\mu}(\gamma) = \sum_d (n_{c, d}/n) \sum_{\mu} \hat{p}_{\mu}(\gamma) \\ &= \sum_d (n_{c, d}/n) = n_c/n = \text{constant}. \end{aligned}$$

Thus,

$$(1 - \pi_{\phi_j}) d_{\phi_j} = 0.$$

(4) θ_i (or ϕ_j) contains a portion of γ plus other variables.

Suppose $\theta = \gamma + \mu$, $\gamma = c + d$, $\mu = a + b$ and $\theta_i = c + a$.

Then

$$\begin{aligned} \hat{p}_{\theta_i} &= \hat{p}_{c, a} = \sum_d \sum_b (n_{\gamma}/n) \hat{p}_{\mu}(\gamma) = \sum_d (n_{c, d}/n) \sum_b \hat{p}_{a, b}(c, d) \\ &= \sum_d (n_{\gamma}/n) \hat{p}_{a, +}(c, d). \end{aligned}$$

Hence, taking account of the independent sampling for each value of γ ,

$$\begin{aligned}(1-\pi_{\theta_i})d_{\theta_i} &= nv(\sum_d (n_Y/n) \hat{p}_{a,+}(c,d)) / \pi_{\theta_i} \\ &= [\sum_d (n_Y/n) n_Y v(\hat{p}_{a,+}(c,d))] / \pi_{\theta_i} \\ &= \frac{[\sum_d (n_Y/n) p_{a,+}(c,d) (1-p_{a,+}(c,d)) d(\hat{p}_{a,+}(c,d))]}{[\sum_d (n_Y/n) p_{a,+}(c,d)]}\end{aligned}$$

(5) θ_i (or ϕ_j) is independent of γ and contains all other variables.

In this case, $\theta = \gamma + \mu$ and $\theta_i = \mu$ so that

$$\hat{p}_{\theta_i} = \hat{p}_{\mu} = \sum_Y (n_Y/n) \hat{p}_{\mu}(\gamma)$$

Then

$$\begin{aligned}(1-\pi_{\theta_i})d_{\theta_i} &= nv[\sum_Y (n_Y/n) \hat{p}_{\mu}(\gamma)] / \pi_{\theta_i} \\ &= \sum_Y (n_Y/n) n_Y v(\hat{p}_{\mu}(\gamma)) / \pi_{\theta_i} \text{ because of independent sampling} \\ &= [\sum_Y (n_Y/n) p_{\mu}(\gamma) (1-p_{\mu}(\gamma)) d(\hat{p}_{\mu}(\gamma))] / [\sum_Y (n_Y/n) p_{\mu}(\gamma)]\end{aligned}$$

(6) θ_i (or ϕ_j) is independent of γ and contains only a portion of the other variables.

Let $\theta = \mu + \gamma$, $\mu = a + b$ and $\theta_i = a$. Then

$$\hat{p}_{\theta_i} = \hat{p}_a = \sum_b \sum_Y (n_Y/n) \hat{p}_{\mu}(\gamma) = \sum_Y (n_Y/n) \hat{p}_{a,+}(\gamma)$$

Therefore,

$$\begin{aligned}(1-\pi_{\theta_i})d_{\theta_i} &= nv(\sum_Y (n_Y/n) \hat{p}_{a,+}(\gamma)) / \pi_{\theta_i} \\ &= \sum_Y (n_Y/n) n_Y v(\hat{p}_{a,+}(\gamma)) / \pi_{\theta_i} \text{ because of independent sampling} \\ &= \frac{[\sum_Y (n_Y/n) p_{a,+}(\gamma) (1-p_{a,+}(\gamma)) d(\hat{p}_{a,+}(\gamma))]}{[\sum_Y (n_Y/n) p_{a,+}(\gamma)]}\end{aligned}$$

For any particular model with direct estimates and for a sampling scheme with a set of fixed marginal totals, an expression for δ can be obtained readily using the above general results, provided that the model contains the u -terms corresponding to the fixed margins. Below, the formulae are given for all permissible cases in two and three-dimensional tables and for one particular example in a four-dimensional table.

In the tabular results that follow, let "Hypothesis (within)" denote the hypothesis expressed in terms of within-population values, and let "con- l " indicate that a specified θ_i or ϕ_j has the configuration l ($l = 1, 2, \dots, 6$) described above.

Two-Dimensional Table

Hypothesis: $\pi_{ij} = \pi_{i+} \pi_{+j}$ ($\Sigma u_{12} = 0$)

(a) Fixed margins: n_{i+}

$$\pi_{ij} : (n_{i+}/n) P_j(i)$$

$$\text{Hypothesis (within)} : P_j(i) = \sum_i (n_{i+}/n) P_j(i) = Q_j$$

This is testing of homogeneity of proportions across populations.

$$\theta_1 = \{i\} \quad \text{con-1}$$

$$\theta_2 = \{j\} \quad \text{con-5}$$

$$(1-\pi_{\theta_1})d_{\theta_1} = (1-P_{j(i)})d(\hat{p}_{j(i)}) = (1-Q_j)d(\hat{p}_{j(i)})$$

$$(1-\pi_{\theta_1})d_{\theta_1} = 0$$

$$(1-\pi_{\theta_2})d_{\theta_2} = [\sum_i (n_{i+}/n) P_j(i) (1-P_{j(i)})d(\hat{p}_{j(i)})] / [\sum_i (n_{i+}/n) P_j(i)]$$

$$= (1-Q_j) \sum_i (n_{i+}/n) d(\hat{p}_j(i)) .$$

Therefore

$$\begin{aligned} (I-1)(J-1)\delta &= \sum_{i,j} (1-Q_j) d(\hat{p}_{jk(i)}) - \sum_j (1-Q_j) \sum_i (n_{i+}/n) d(\hat{p}_{jk(i)}) \\ &= \sum_{i,j} (1-n_{i+}/n) (1-Q_j) d(\hat{p}_{jk(i)}) . \end{aligned}$$

This is the result obtained by Rao and Scott (1984).

(b) Fixed margins : n_{+j}

$$\pi_{ij} : (n_{+j}/n) p_{i(j)} .$$

Results are parallel to (a), with i and j interchanged.

Three-Dimensional Table

Hypothesis A: $\pi_{ijk} = (\pi_{i+k} \pi_{+jk}) / \pi_{++k} \quad (\equiv u_{12} = u_{123} = 0)$

(Aa) Fixed margins : n_{i++}

$$\pi_{ijk} : (n_{i++}/n) p_{jk(i)}$$

$$\text{Hypothesis (within): } p_{jk(i)} = \frac{p_{+k(i)} [\sum_i (n_{i++}/n) p_{jk(i)}]}{[\sum_i (n_{i++}/n) p_{+k(i)}]}$$

This hypothesis is hard to interpret.

$$\theta_1 = \{i, k\} \quad \text{con-2}$$

$$\theta_2 = \{j, k\} \quad \text{con-5}$$

$$\phi_1 = \{k\} \quad \text{con-6}$$

Therefore

$$\begin{aligned} (I-1)(J-1)K\delta &= \sum_{i,j,k} (1-p_{jk(i)}) d(\hat{p}_{jk(i)}) - \sum_{i,k} (1-p_{+k(i)}) d(\hat{p}_{+k(i)}) \\ &\quad - \sum_{j,k} ([\sum_i n_{i++} p_{jk(i)}] (1-p_{jk(i)}) d(\hat{p}_{jk(i)}) / [\sum_i n_{i++} p_{+k(i)}]) \end{aligned}$$

$$+ \sum_k \{ [\sum_i n_{i++} P_{+k(i)} (1 - P_{+k(i)})^d (\hat{p}_{+k(i)})] / [\sum_i n_{i++} P_{+k(i)}] \}$$

(Ab) Fixed margins: n_{+j+}

$$\pi_{ijk} : (n_{+j+}/n) P_{ik(j)}$$

Results are parallel to (Aa) above, with i and j interchanged.

(Ac) Fixed margins : n_{i+k}

$$\pi_{ijk} : (n_{i+k}/n) P_{j(i,k)}$$

Hypothesis (within): $P_{j(i,k)} = [\sum_i (n_{i+k}/n) P_{j(i,k)}] / (n_{++k}/n) = Q_j(k)$

$$\theta_1 = \{i, k\} \quad \text{con-1}$$

$$\theta_2 = \{j, k\} \quad \text{con-4}$$

$$\phi_1 = \{k\} \quad \text{con-3}$$

Therefore

$$(I-1)(J-1)K\delta = \sum_{ijk} (1 - Q_j(k)) (1 - n_{i+k}/n_{++k})^c d(\hat{p}_{j(i,k)})^c.$$

(Ad) Fixed margins: n_{+jk}

$$\pi_{ijk} : (n_{+jk}/n) P_{i(j,k)}$$

Results are parallel to (Ac) above, with i and j interchanged.

(Ae) Fixed margins : n_{++k}

$$\pi_{ijk} : (n_{++k}/n) P_{ij(k)}$$

Hypothesis (within): $P_{ij(k)} = P_{i+(k)} P_{+j(k)}$

This is testing of independence of variables

1 and 2 within each population.

$$\theta_1 = \{i, k\} \quad \text{con-2}$$

$$\theta_2 = \{j, k\} \quad \text{con-2}$$

$$\phi_1 = \{k\} \quad \text{con-1}$$

Therefore

$$\begin{aligned} (I-1)(J-1)K\delta &= \sum_{ijk} (1-p_{i+(k)}) p_{+j(k)} d(\hat{p}_{ij(k)}) \\ &\quad - \sum_{ik} (1-p_{i+(k)}) d(\hat{p}_{i+(k)}) \\ &\quad - \sum_{jk} (1-p_{+j(k)}) d(\hat{p}_{+j(k)}) \end{aligned}$$

(Af) It is not permissible to have $\{n_{ij+}\}$ as fixed margins since $u_{12(ij)}$ is not in the model.

Hypothesis B: $\pi_{ijk} = \pi_{+jk} \pi_{i++}$ ($\equiv u_{12} = u_{13} = u_{123} = 0$)

(Ba) Fixed margins : n_{i++}

$$\pi_{ijk} : (n_{i++}/n) p_{jk(i)}$$

Hypothesis(within): $p_{jk(i)} = \sum_i (n_{i++}/n) p_{jk(i)} = q_{jk}$

This is testing of homogeneity of proportions
across populations.

$$\theta_1 = \{j, k\} \quad \text{con-5}$$

$$\theta_2 = \{i\} \quad \text{con-1}$$

Therefore

$$(I-1)(JK-1)\delta = \sum_{ijk} (1-q_{jk}) (1-n_{i++}/n) d(\hat{p}_{jk(i)})$$

(Bb) Fixed margins : n_{+j+}

$$\pi_{ijk} : (n_{+j+}/n) p_{ik(j)}$$

Hypothesis(within): $p_{ik(j)} = \sum_i p_{ik(j)} [\sum_j (n_{+j+}/n) p_{i+(j)}]$

This hypothesis is hard to interpret.

$$\theta_1 = \{j, k\} \quad \text{con-2}$$

$$\theta_2 = \{i\} \quad \text{con-6}$$

Therefore

$$\begin{aligned} (I-1)(JK-1)\delta &= \sum_{ijk} (1-p_{ik(j)}) d(\hat{p}_{ik(j)}) \\ &\quad - \sum_{jk} (1-p_{+k(j)}) d(\hat{p}_{+k(j)}) \\ &\quad - \sum_i ([\sum_{j+} n_{ij+} p_{i+(j)}] (1-p_{i+(j)}) d(\hat{p}_{i+(j)}) / [\sum_{j+} n_{ij+} p_{i+(j)}]) \end{aligned}$$

(Bc) Fixed margins: n_{++k}

$$\pi_{ijk} : (n_{++k}/n) p_{ij(k)}$$

Results are parallel to (Bb) above, with j and k interchanged.

(Bd) Fixed margins : n_{+jk}

$$\pi_{ijk} : (n_{+jk}/n) p_{i(jk)}$$

Hypothesis(within): $p_{i(jk)} = \sum_{jk} (n_{+jk}/n) p_{i(jk)} = Q_i$

This is testing of homogeneity of proportions
across populations.

$$\theta_1 = \{j, k\} \quad \text{con-1}$$

$$\theta_2 = \{i\} \quad \text{con-5}$$

Therefore

$$(I-1)(JK-1)\delta = \sum_{ijk} (1-Q_i)(1-n_{+jk}/n) d(\hat{p}_{i(jk)})$$

(Be) It is not permissible to have $\{n_{ij+}\}$ as fixed margins since $u_{12}(ij)$ is not in the model.

(Bf) It is not permissible to have $\{n_{i+k}\}$ as fixed margins since $u_{13}(ik)$ is not in the model.

Hypothesis C: $\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k} \quad (\equiv u_{12} = u_{13} = u_{23} = u_{123} = 0)$

(Ca) Fixed margins : n_{i++}

$$\pi_{ijk} : (n_{i++}/n)P_{jk(i)}$$

$$\begin{aligned} \text{Hypothesis (within)} : P_{jk(i)} &= [\sum_i (n_{i++}/n)P_{j+(i)}] [\sum_i (n_{i++}/n)P_{+k(i)}] \\ &= Q_j R_k \end{aligned}$$

$$\theta_1 = \{i\} \quad \text{con-1}$$

$$\theta_2 = \{j\} \quad \text{con-6}$$

$$\theta_3 = \{k\} \quad \text{con-6}$$

Therefore

$$\begin{aligned} [IJK - (I+J+K) + 2]\delta &= \sum_{ijk} (1 - Q_j R_k) d(\hat{p}_{jk(i)}) \\ &\quad - \sum_{ij} (n_{i++}/n) (1 - Q_j) d(\hat{p}_{j+(i)}) \\ &= \sum_{ik} (n_{i++}/n) (1 - R_k) d(\hat{p}_{+k(i)}) \end{aligned}$$

(Cb) Fixed margins: n_{+j+}

$$\pi_{ijk} : (n_{+j+}/n)P_{ik(j)}$$

Results are parallel to (Ca) above, with i and j interchanged.

(Cc) Fixed margins: n_{++k}

$$\pi_{ijk} : (n_{++k}/n)P_{ij(k)}$$

Results are parallel to (Ca) above, with i and k interchanged.

(Cd) It is not permissible to have $\{n_{ij+}\}$, $\{n_{i+k}\}$ or $\{n_{+jk}\}$ as the set of fixed margins since $u_{12}(ij)$, $u_{13}(ik)$ and $u_{23}(jk)$ respectively are not in the model.

Four-Dimensional Table

Hypothesis: $\pi_{ijkl} = \pi_{ijk+} \pi_{i++l} / \pi_{i+++}$ ($\equiv u_{1234} = u_{124} = u_{134} = u_{234} = u_{24} = u_{34} = 0$)

Fixed margins : n_{i+++}

π_{ijkl} : $(n_{i+++}/n) = P_{jkl(i)}$

Hypothesis(within): $P_{jkl(i)} = P_{jk+(i)} P_{++l(i)}$

This is testing the independence of the pair of variables 2 and 3 with variable 4, within each population.

$\theta_1 = \{i, j, k\}$ con-2

$\theta_2 = \{i, l\}$ con-2

$\phi_1 = \{i\}$ con-1

$(1-\pi_{\theta}) d_{\theta} = (1-P_{jk+(i)} P_{++l(i)}) d(\hat{p}_{jkl(i)})$

$(1-\pi_{\theta_1}) d_{\theta_1} = (1-P_{jk+(i)}) d(\hat{p}_{jk+(i)})$

$(1-\pi_{\theta_2}) d_{\theta_2} = (1-P_{++l(i)}) d(\hat{p}_{++l(i)})$

$(1-\pi_{\phi_1}) d_{\phi_1} = 0$

Therefore

$$\begin{aligned} I(L-1)(KJ-1)\delta &= \sum_{ijkl} (1-P_{jk+(i)} P_{++l(i)}) d(\hat{p}_{jkl(i)}) \\ &\quad - \sum_{ijk} (1-P_{jk+(i)}) d(\hat{p}_{jk+(i)}) \\ &\quad - \sum_{il} (1-P_{++l(i)}) d(\hat{p}_{++l(i)}) \end{aligned}$$

Any other specified hypothesis admitting direct estimates and specified fixed margins in four-dimensional tables could be handled routinely, by using the general algorithm developed above.

2.4 Extensions to General Models

Although loglinear models are very versatile in covering many problems encountered by the researcher with multiway tables of data, situations arise which cannot be handled by loglinear models. As an example, marginal homogeneity in a 2-dimensional square table cannot be expressed as a loglinear model (BFH (1975), p. 282). Or, as another example, Grizzle, Starmer and Koch (1969) present the model $G\pi = H\theta$, where G is a $u \times T$ matrix of rank $u \leq T-1$ and H is a $u \times v$ matrix of rank $v \leq u$, so not following the form of (2.1). However, in such instances, it would be desirable to have a chi-squared or likelihood ratio statistic for testing feasible hypotheses, and to know the asymptotic properties of such statistics. As well, when these asymptotic distributions are complex, it would be important to have corrections to the statistics that could be practically implemented with survey data.

In the work to follow, π will again denote the T -vector of population cell proportions in a multiway table, with the same assumptions about the survey design as given in the first paragraph of Section 2.2. However, instead of a loglinear model for π , we will assume the general model M_1 :

$$(2.14) \quad \pi_i = f_i(\theta) \quad i = 1, 2, \dots, T$$

where θ is an r -vector of unknown parameters and the f_i are known functions satisfying the regularity conditions specified by Birch (1964)

(see BFH (1975), p.510 or Appendix A8).

The likelihood under the assumption of multinomial sampling is

$$(2.15) \quad L = \frac{n!}{\prod n_i!} \prod (f_i(\underline{\theta}))^{n_i}.$$

When L is maximized, under the restriction $\sum f_i(\underline{\theta}) = 1$, the likelihood equations are

$$(2.16) \quad \left(\frac{\partial f}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\underline{\hat{\theta}}} D(\underline{f}(\underline{\hat{\theta}}))^{-1} (\underline{n}/n) = \underline{0}$$

where $\frac{\partial f}{\partial \underline{\theta}}$ is the $T \times r$ Jacobian matrix with (i,j) th element $\partial f_i(\underline{\theta}) / \partial \theta_j$, $D(\underline{f}(\underline{\hat{\theta}})) = \text{diag}(\underline{f}(\underline{\hat{\theta}}))$, and $\underline{n}/n = (n_1/n, n_2/n, \dots, n_T/n)'$. Because the true sampling plan is not multinomial, "pseudo likelihood equations" are obtained by replacing $(\underline{n}/n)$ in (2.16) with $\underline{\hat{p}}$, yielding

$$(2.17) \quad \left(\frac{\partial f}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\underline{\hat{\theta}}} D(\underline{f}(\underline{\hat{\theta}}))^{-1} \underline{\hat{p}} = \underline{0}.$$

The solution of (2.17) for $\underline{\hat{\theta}}$, and the resulting $\underline{\hat{\pi}} = \underline{f}(\underline{\hat{\theta}})$ are the "pseudo mle's" for $\underline{\theta}$ and $\underline{\pi}$ respectively. The same result (2.17) would obviously be obtained if n_i in (2.15) were replaced by $n \hat{p}_i$, to yield $\bar{L}$, and then $\bar{L}$ was maximized.

The Pearson chi-squared statistic used to test $\underline{\pi} = \underline{f}(\underline{\theta})$ in the classical multinomial situation is

$$x_c^2 = \sum_i (n_i - n f_i(\underline{\hat{\theta}}))^2 / n f_i(\underline{\hat{\theta}}) = n \sum_i (n_i/n - f_i(\underline{\hat{\theta}}))^2 / f_i(\underline{\hat{\theta}}),$$

where $\underline{\hat{\theta}}$ is the solution of (2.16). For a general sampling scheme, the parallel statistic, obtained by substituting $\hat{p}_i$ for n_i/n and having $\underline{\hat{\theta}}$ the solution of (2.17), is

$$(2.18) \quad x^2 = n \sum_i (\hat{p}_i - f_i(\hat{\theta}))^2 / f_i(\hat{\theta}) = n(\hat{p} - \underline{f}(\hat{\theta}))' D(\underline{f}(\hat{\theta}))^{-1} (\hat{p} - \underline{f}(\hat{\theta}))$$

where $D(\underline{f}(\hat{\theta})) = \text{diag}(\underline{f}(\hat{\theta})) = \text{diag}(\hat{\pi})$. The asymptotic distribution of x^2 will be developed as a special case of the asymptotic distribution of a statistic used for nested models.

2.4.1. Asymptotic Covariance Matrix of $\hat{\pi}$

Lemma. Under the regularity conditions of Birch (1964),

$$(2.19) \quad \hat{\theta} - \underline{\theta} = (F' D_{\pi}^{-1} F)^{-1} F' D_{\pi}^{-1} (\hat{p} - \underline{f}(\hat{\theta})) + o_p(n^{-1/2})$$

where F is the $T \times r$ matrix $\frac{\partial f}{\partial \theta}$ and $D_{\pi} = \text{diag}(\pi)$.

Proof. The function $\bar{L}$ which is maximized to obtain $\hat{\theta}$ is defined by

$$\begin{aligned} -2 \ln \bar{L} &= -2n \sum_i \hat{p}_i \ln f_i(\hat{\theta}) - 2n \ln(n! / \prod (n \hat{p}_i)!) \\ &= 2n \sum_i \hat{p}_i \ln \hat{p}_i - 2n \sum_i \hat{p}_i \ln f_i(\hat{\theta}) - 2n \sum_i \hat{p}_i \ln \hat{p}_i - 2n \ln(n! / \prod (n \hat{p}_i)!) \\ &= 2n \bar{G}^2(\hat{p}, \underline{f}(\hat{\theta})) + \text{terms not involving } \hat{\theta} \end{aligned}$$

where $\bar{G}^2(\underline{x}, \underline{y}) = \sum x_i \ln(x_i / y_i)$. To maximize $\bar{L}$, we must choose the value of $\hat{\theta}$ to minimize $\bar{G}^2(\hat{p}, \underline{f}(\hat{\theta}))$. Birch (1964) defines $\underline{\theta}(\underline{p})$ to be that value of $\underline{\theta}$ such that

$$\bar{G}^2(\underline{p}, \underline{f}(\underline{\theta}(\underline{p}))) = \inf_{\underline{\theta}} \bar{G}^2(\underline{p}, \underline{f}(\underline{\theta})).$$

The value of $\hat{\theta}$ to maximize $\bar{L}$ is thus $\hat{\theta} = \underline{\theta}(\hat{p})$. Birch (1964) proved that

$$\underline{\theta}(\underline{p}) = \underline{\theta} + (F'D_{\pi}^{-1}F)^{-1}F'D_{\pi}^{-1}(\underline{p} - \underline{f}(\underline{\theta})) + o(\|\underline{p} - \underline{f}(\underline{\theta})\|)$$

$$\text{as } \underline{p} \rightarrow \underline{\pi} = \underline{f}(\underline{\theta}).$$

Since, by assumption, $\sqrt{n}(\hat{\underline{p}} - \underline{f}(\underline{\theta}))$ is asymptotically distributed as a $N(0, V)$ vector as $n \rightarrow \infty$, then

$$\hat{\underline{p}} - \underline{f}(\underline{\theta}) = o_p(n^{-1/2}).$$

Therefore, it follows that

$$(2.20) \quad \hat{\underline{\theta}} = \underline{\theta}(\hat{\underline{p}}) = \underline{\theta} + (F'D_{\pi}^{-1}F)^{-1}F'D_{\pi}^{-1}(\hat{\underline{p}} - \underline{f}(\underline{\theta})) + o_p(n^{-1/2}).$$

Q.E.D.

Since we assumed that $\sqrt{n}(\hat{\underline{p}} - \underline{f}(\underline{\theta})) \rightsquigarrow N(0, V)$, where " $\rightsquigarrow$ " denotes convergence in distribution, the asymptotic covariance matrix of $\hat{\underline{p}}$ is $C(\hat{\underline{p}}) = V/n$. This fact together with (2.19), implies that the asymptotic covariance matrix of $\hat{\underline{\theta}}$ is

$$(2.21) \quad C(\hat{\underline{\theta}}) = n^{-1}(F'D_{\pi}^{-1}F)^{-1}(F'D_{\pi}^{-1}VD_{\pi}^{-1}F)(F'D_{\pi}^{-1}F)^{-1}.$$

Because of the assumed regularity conditions, (Birch (1964)), we have

$$(2.22) \quad \underline{f}(\hat{\underline{\theta}}) = \underline{f}(\underline{\theta}) + F(\hat{\underline{\theta}} - \underline{\theta}) + o_p(n^{-1/2}).$$

Substituting (2.20) into (2.22), we have

$$(2.23) \quad \underline{f}(\hat{\underline{\theta}}) - \underline{f}(\underline{\theta}) = F(F'D_{\pi}^{-1}F)^{-1}F'D_{\pi}^{-1}(\hat{\underline{p}} - \underline{f}(\underline{\theta})) + o_p(n^{-1/2}),$$

so that the asymptotic covariance matrix of $\underline{f}(\hat{\underline{\theta}})$ is

$$(2.24) \quad C(\underline{f}(\hat{\underline{\theta}})) = F \cdot C(\hat{\underline{\theta}}) \cdot F'.$$

The vector $\underline{\hat{p}} - \underline{f}(\underline{\hat{\theta}})$ is called the vector of residuals. Since

$$\begin{aligned} (2.25) \quad \underline{\hat{p}} - \underline{f}(\underline{\hat{\theta}}) &= (\underline{\hat{p}} - \underline{f}(\underline{\theta})) - (\underline{f}(\underline{\hat{\theta}}) - \underline{f}(\underline{\theta})) \\ &= (\underline{I} - \underline{F}(\underline{F}'\underline{D}_{\pi}\underline{F})^{-1}\underline{F}'\underline{D}_{\pi}^{-1})(\underline{\hat{p}} - \underline{f}(\underline{\theta})) + o_p(n^{-\frac{1}{2}}) \end{aligned}$$

then the asymptotic covariance matrix of the residual vector is

$$(2.26) \quad \text{Cov}(\underline{\hat{p}} - \underline{f}(\underline{\hat{\theta}})) = n^{-1} \underline{A} \underline{V} \underline{A}'$$

where $\underline{A} = (\underline{I} - \underline{F}(\underline{F}'\underline{D}_{\pi}\underline{F})^{-1}\underline{F}'\underline{D}_{\pi}^{-1})$. Consistent estimates of the above covariance matrices are obtained by replacing $\underline{\theta}$ by $\underline{\hat{\theta}}$ and $\underline{V}$ by its survey estimate.

The residual vector is often useful in detecting model deviations. As an example, let the vector of standardized residuals be denoted by $\underline{e} = (e_1, e_2, \dots, e_T)'$, where $e_i = r_i / [\hat{C}_{ii}(\underline{x})]^{\frac{1}{2}}$ with $r_i = \hat{p}_i - f_i(\underline{\hat{\theta}})$ and $\hat{C}_{ii}(\underline{x})$ being the estimated asymptotic variance of r_i obtained from (2.26) by replacing $\underline{\theta}$ and $\underline{V}$ by $\underline{\hat{\theta}}$ and $\hat{\underline{V}}$ respectively. Since the e_i 's are approximately $N(0,1)$ if the model fitted is true, the expected numbers of residuals e_i exceeding 1.96 and 2.58 in absolute value are $(.05)T$ and $(.01)T$ respectively. These expected numbers provide a rough guide for identifying any outlying cells.

2.4.2 Nested Models

In model $M_1: \underline{\pi} = \underline{f}(\underline{\theta})$, suppose that $\underline{\theta} = (\underline{\theta}_1', \underline{\theta}_2')'$ where $\underline{\theta}_1$ is $s \times 1$ and $\underline{\theta}_2$ is $u \times 1$ ($s+u = r$). If we wished to test the hypothesis $H: \underline{\theta}_2 = \underline{0}$; we would have the reduced model

$M_2 : \pi = \underline{f}(\underline{\theta}_1)$ where " $\underline{f}(\underline{\theta}_1)$ " denotes " $\underline{f}(\underline{\theta}_1, \underline{0})$ ". Let $\hat{\underline{\theta}}_1$ and $\hat{\underline{\pi}} = \underline{f}(\hat{\underline{\theta}}_1)$ be the pseudo mle's of $\underline{\theta}_1$ and $\underline{\pi}$ respectively under M_2 , obtained from the pseudo likelihood equations

$$\left(\frac{\partial \underline{f}}{\partial \underline{\theta}_1} \right)_{\underline{\theta} = (\hat{\underline{\theta}}_1, \underline{0})} D(\underline{f}(\hat{\underline{\theta}}_1))^{-1} \hat{\underline{p}} = \underline{0}.$$

The Pearson chi-squared statistic for testing H is given by

$$(2.27) \quad \chi^2 = n \sum_i \frac{(f_i(\hat{\underline{\theta}}) - f_i(\hat{\underline{\theta}}_1))^2}{f_i(\hat{\underline{\theta}}_1)} = n (\underline{f}(\hat{\underline{\theta}}) - \underline{f}(\hat{\underline{\theta}}_1))' D(\underline{f}(\hat{\underline{\theta}}_1))^{-1} (\underline{f}(\hat{\underline{\theta}}) - \underline{f}(\hat{\underline{\theta}}_1)).$$

By definition,

$$F = (F_1, F_2),$$

where F_1 and F_2 denote $\frac{\partial \underline{f}}{\partial \underline{\theta}_1}$ and $\frac{\partial \underline{f}}{\partial \underline{\theta}_2}$ respectively. Since F is of rank r , F_1 must be of rank s and F_2 of rank u . Because of Birch's assumptions,

$$(2.28) \quad \underline{f}(\hat{\underline{\theta}}) - \underline{f}(\underline{\theta}) \sim F(\hat{\underline{\theta}} - \underline{\theta}) = F_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1) + F_2(\hat{\underline{\theta}}_2 - \underline{\theta}_2),$$

and similarly, under H ,

$$(2.29) \quad \underline{f}(\hat{\underline{\theta}}_1) - \underline{f}(\underline{\theta}) \sim F_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1).$$

where " $A \sim B$ " means $\sqrt{n}(A-B) = o_p(1)$. Thus, under H ,

$$(2.30) \quad \underline{f}(\hat{\underline{\theta}}) - \underline{f}(\hat{\underline{\theta}}_1) \sim F_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1) + F_2\hat{\underline{\theta}}_2 - F_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1)$$

where, parallel to (2.19),

$$(2.31) \quad (\hat{\underline{\theta}}_1 - \underline{\theta}_1) = (F_1' D_{\pi}^{-1} F_1)^{-1} F_1' D_{\pi}^{-1} (\hat{\underline{p}} - \underline{f}(\underline{\theta})) + o_p(n^{-1/2}).$$

Examining (2.30) it can be seen that if $(\hat{\theta}_1 - \theta_1)$ could be expressed asymptotically as $(\hat{\theta}_1 - \theta_1) + Q\hat{\theta}_2$ for some matrix Q , we could express $\underline{f}(\hat{\theta}) - \underline{f}(\hat{\theta}_1)$ asymptotically as a function just of $\hat{\theta}_2$. To accomplish this, we recall that, under H ,

$$(2.32) \quad \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 \end{pmatrix} \sim \begin{pmatrix} F_1' D_{\pi}^{-1} F_1 & F_1' D_{\pi}^{-1} F_2 \\ F_2' D_{\pi}^{-1} F_1 & F_2' D_{\pi}^{-1} F_2 \end{pmatrix}^{-1} F_2' D_{\pi}^{-1} (\hat{\theta} - \underline{f}(\theta)).$$

Using the standard formula for the inverse of a partitioned matrix (eg. Seber (1977), page 390), and (2.31), we find (see Appendix) that $Q = (F_1' D_{\pi}^{-1} F_1)^{-1} (F_1' D_{\pi}^{-1} F_2)$. Substituting into (2.30), it follows that

$$(2.33) \quad \underline{f}(\hat{\theta}) - \underline{f}(\hat{\theta}_1) \sim (F_2 - F_1 Q) \hat{\theta}_2 = A_2 \hat{\theta}_2,$$

where

$$(2.34) \quad A_2 = (I - F_1' (F_1' D_{\pi}^{-1} F_1)^{-1} F_1' D_{\pi}^{-1}) F_2.$$

Because of (2.33) and the fact that $D(\underline{f}(\hat{\theta}_1))^{-1} \sim D_{\pi}^{-1}$ under H , we have now shown that

Theorem 2.1. Under $H : \theta_2 \neq 0$,

$$(2.35) \quad X^2 \sim n \hat{\theta}_2' A_2' D_{\pi}^{-1} A_2 \hat{\theta}_2.$$

We can also express the asymptotic distribution of X^2 as a weighted sum of chi-squared variables as follows.

In the expression for $C(\hat{\theta})$ given by (2.21), if $F' D_{\pi}^{-1} F$ is partitioned as in (2.32) and is inverted, the appropriate terms

pertaining to $C(\hat{\theta}_2)$ are:

$$(2.36) \quad C(\hat{\theta}_2) = n^{-1} (A_2' D_{\pi}^{-1} A_2)^{-1} (A_2' D_{\pi}^{-1} V D_{\pi}^{-1} A_2) (A_2' D_{\pi}^{-1} A_2)^{-1}.$$

Thus, $\hat{\theta}_2 \approx N_u(0, C(\hat{\theta}_2))$ and

$$X^2 \approx \sum \delta_i W_i,$$

where the W_i are independent χ_1^2 variables and the δ_i are eigenvalues of

$$(2.37) \quad nC(\hat{\theta}_2) (A_2' D_{\pi}^{-1} A_2) = (A_2' D_{\pi}^{-1} A_2)^{-1} (A_2' D_{\pi}^{-1} V D_{\pi}^{-1} A_2).$$

It should be noted that the asymptotic covariance matrix of $A_2' D_{\pi}^{-1} \hat{\theta}_2$ is $n^{-1} (A_2' D_{\pi}^{-1} V D_{\pi}^{-1} A_2)$ under the survey design and $n^{-1} (A_2' D_{\pi}^{-1} P D_{\pi}^{-1} A_2) = n^{-1} (A_2' D_{\pi}^{-1} A_2)$ under multinomial sampling, so that the δ_i can be regarded as the "generalized design effects" of the vector $A_2' D_{\pi}^{-1} \hat{\theta}_2$, as defined by Rao and Scott, (1981).

Under multinomial sampling, $V = P$ and $(A_2' D_{\pi}^{-1} A_2)^{-1} (A_2' D_{\pi}^{-1} P D_{\pi}^{-1} A_2) = I$, so that $\delta_i = 1$, $i = 1, 2, \dots, u$, and we get the standard result that $X^2 \approx \chi_u^2$ under H , as a special case.

2.4.3 Statistics of the Wald Type

We will now restrict our models $\pi_i = f_i(\theta)$ to those for which there exists a monotonic differentiable function g such that

$$(2.38) \quad g_i = g(\pi_i) = x_{oi} \bar{u}(\theta) + x_i' \theta \quad i = 1, 2, \dots, T,$$

where $\underline{\theta}$ is an $r \times 1$ vector, $X = [\underline{x}_1, \underline{x}_2, \dots, \underline{x}_T]'$ is a $T \times r$ matrix of rank r ($\leq T-1$), $\underline{x}_0 = [x_{01}, x_{02}, \dots, x_{0T}]'$ is a T -vector with $X' \underline{x}_0 = \underline{0}$ and $\bar{u}(\underline{\theta})$ is a constant that is a function of $\underline{\theta}$, whose form is determined by the restriction $\sum \pi_i = 1$. As an example, for the loglinear model, $g(\pi_i) = \ln \pi_i$, $x_{0i} = 1$ and $\bar{u}(\underline{\theta}) = \ln[1/\sum \exp(\underline{x}_i' \underline{\theta})]$. Another special case is that of the model for marginal homogeneity in a square table, as will be shown in Subsection 2.4.5.

If a consistent estimator $\hat{V}/n$ of the covariance matrix of $\hat{p}$ is available, then a Wald statistic x_W^2 that is asymptotically χ_u^2 under $H : \underline{\theta}_2 = \underline{0}$ may be constructed as follows.

Model (2.38) may be written as

$$(2.39) \quad \underline{g} = \bar{u}(\underline{\theta}) \underline{x}_0 + \underline{x}_1 \underline{\theta}_1 + \underline{x}_2 \underline{\theta}_2,$$

where $\underline{g} = (g_1, g_2, \dots, g_T)'$, $X = (\underline{x}_1, \underline{x}_2)$, $\underline{\theta} = (\underline{\theta}_1', \underline{\theta}_2')'$, $\underline{\theta}_1$ is $s \times 1$ and $\underline{\theta}_2$ is $u \times 1$ ($s+u = r$). Since X is of full rank r , $\underline{x}_1$ is of full rank s and $\underline{x}_2$ of full rank u . Let C be any $T \times u$ matrix of rank u such that $C' \underline{x}_1 = \underline{0}$, $C' \underline{x}_0 = \underline{0}$ and $C' \underline{x}_2$ is nonsingular. Then $H : \underline{\theta}_2 = \underline{0}$ is equivalent to $H' : \underline{\phi} = C' \underline{g} = \underline{0}$. A Wald statistic for testing H , based on $\hat{\pi}$, would be

$$(2.40) \quad x_W^2 = \hat{\underline{\phi}}' [\hat{C}(\hat{\underline{\phi}})]^{-1} \hat{\underline{\phi}}$$

where $\hat{\underline{\phi}} = C' \hat{\underline{g}}$, $\hat{\underline{g}}$ is the T -vector with $\hat{g}_i = g(\hat{\pi}_i)$ and $\hat{C}(\hat{\underline{\phi}})$ is the estimated covariance matrix of $\hat{\underline{\phi}}$ obtained in the following manner.

By the delta method (see, for example, BFH, (1975)), $\hat{\underline{g}} - \underline{g} \sim D_{\underline{g}}(\hat{\underline{\pi}} - \underline{\pi})$

where D_g is the diagonal matrix with i th diagonal element equal to $\partial g_i / \partial \pi_i$. We then get

$$(2.41) \quad C(\hat{\phi}) = C' D_g C(\hat{\pi}) D_g C$$

where $C(\hat{\pi})$ is given by (2.24). Finally, we obtain $\hat{C}(\hat{\phi})$ by replacing π by $\hat{\pi}$, θ by $\hat{\theta}$ and V by $\hat{V}$ in (2.41). Now, $\hat{\phi}$ has an asymptotic distribution of $N_u(0, C(\hat{\phi}))$ so that by standard results (see C.R. Rao (1965)), the asymptotic distribution of X_W^2 is χ_u^2 .

The Wald statistic X_W^2 requires the knowledge of $\hat{V}$. If $\hat{V}$ is not available, a Wald-type statistic may be formed by ignoring the survey design and replacing V/n in $C(\hat{\phi})$ by P/n , the covariance matrix of $\hat{p}$ under multinomial sampling, thus yielding

$$(2.42) \quad C_M(\hat{\phi}) = C' D_g (F C_M(\hat{\theta}) F') D_g C$$

where

$$(2.43) \quad C_M(\hat{\theta}) = n^{-1} (F' D_{\pi}^{-1} F)^{-1} (F' D_{\pi}^{-1} P D_{\pi}^{-1} F) (F' D_{\pi}^{-1} F)^{-1} \\ = n^{-1} (F' D_{\pi}^{-1} F)^{-1}$$

because $D_{\pi}^{-1} P D_{\pi}^{-1} = D_{\pi}^{-1} - \underline{1}\underline{1}'$ and $\underline{1}' F = 0$. This then leads to the test statistic

$$(2.44) \quad \bar{X}_W^2 = \hat{\phi}' (\hat{C}_M(\hat{\phi}))^{-1} \hat{\phi},$$

where $\hat{C}_M(\hat{\phi})$ is obtained by replacing θ by $\hat{\theta}$ and π by $\hat{\pi}$ in $C_M(\hat{\phi})$ of (2.42). Since $\hat{\phi} \approx N(0, C(\hat{\phi}))$ and $\hat{C}_M(\hat{\phi}) \sim C_M(\hat{\phi})$, then by standard results on quadratic forms, the asymptotic distribution

of $\bar{X}_W^2$ is a weighted sum $\sum v_i w_i$ of independent χ_1^2 variables, where v_i are the eigenvalues of $C_M(\hat{\phi})^{-1} C(\hat{\phi})$.

Theorem 2.2. Under $H: \theta_2 = 0$, $\bar{X}_W^2 \sim \chi^2$.

Proof. The (i,j) th element of the matrix $D_g F$ is $(\partial g_i / \partial \pi_i)(\partial \pi_i / \partial \theta_j)$ which equals $\partial g_i / \partial \theta_j$, so that

$$(2.45) \quad D_g F = \frac{\partial g}{\partial \theta},$$

where $\partial g / \partial \theta$ is the $T \times r$ matrix with (i,j) th element equal to $\partial g_i / \partial \theta_j$. However, because of (2.38),

$$(2.46) \quad \frac{\partial g}{\partial \theta} = X' + X_0' \frac{\partial \bar{u}}{\partial \theta},$$

where $\frac{\partial \bar{u}}{\partial \theta}$ is the $1 \times r$ vector with i th element defined by $\partial \bar{u}(\theta) / \partial \theta_i$.

Taking into account the properties of C , this leads to

$$(2.47) \quad C' \frac{\partial g}{\partial \theta} = (0, C' X_2) .$$

Under H , we have

$$\begin{aligned} (2.48) \quad \hat{\phi} &= \hat{\phi} - \phi = C'(\hat{\theta} - \underline{g}) \\ &\sim C' D_g(\hat{\pi} - \pi) \quad \text{by the delta method} \\ &\sim C' D_g F(\hat{\theta} - \underline{\theta}) \quad \text{by (2.22)} \\ &= C' X_2 \hat{\theta}_2 \quad \text{by (2.45) and (2.47).} \end{aligned}$$

Also under H , $\hat{C}_M(\hat{\phi}) \sim C_M(\hat{\phi})$ where

$$(2.49) \quad C_M(\hat{\phi}) = \frac{1}{n} C' \frac{\partial g}{\partial \theta} (F' D_{\pi}^{-1} F)^{-1} \frac{\partial g}{\partial \theta}' C$$

$$= C'X_2(A_2'D_\pi^{-1}A_2)^{-1}X_2'C$$

because of (2.47) and the standard formula for the inverse of $(F'D_\pi^{-1}F)$, partitioned as in (2.32). Under H , using the fact that $C'X_2$ is nonsingular, we have

$$\begin{aligned} (2.50) \quad \bar{X}_W^2 &\sim n\hat{\theta}'_2 X_2'C(A_2'D_\pi^{-1}A_2)^{-1}X_2'C^{-1}C'X_2\hat{\theta}_2 \\ &= n\hat{\theta}'_2(A_2'D_\pi^{-1}A_2)\hat{\theta}_2 \end{aligned}$$

which is the asymptotic distribution of X^2 , as proved in Theorem 2.1.

It then follows that $\bar{X}_W^2$ is also asymptotically distributed as

$\sum \delta_i w_i$, where the w_i are independent χ^2_1 variables and the δ_i are the eigenvalues of (2.37). Thus $\{\delta_1, \delta_2, \dots, \delta_u\}$ is identical to $\{v_1, v_2, \dots, v_u\}$.

2.4.4. Special Case of Saturated Model M_1

If $s+u(=r) = T-1$, model M_1 is said to be saturated since the number of independent parameters in the model equals the number of independent proportions in the population. A general hypothesis about the fit of a model $\pi = f(\theta)$ can actually be treated as a hypothesis about the fit of a nested model, where the nesting is within a saturated M_1 .

Restricting ourselves to the form of models described in (2.38) and (2.39), we will now show that the matrix with eigenvalues δ_i in the asymptotic distribution of X^2 (or $\bar{X}_W^2$) can be obtained, when M_1 is saturated, without the need to compute A_2 . Under saturated

M_1 , $\hat{\pi} = \hat{p}$, so that $C(\hat{\pi}) = V/n$ under the given survey design and $C(\hat{\pi}) = P/n$ under multinomial sampling. Substituting these values into (2.41), we then find that the δ_i 's are the eigenvalues of

$$(2.51) \quad (C_M(\hat{\phi}))^{-1} C(\hat{\phi}) = (C'D_g P D_g C)^{-1} (C'D_g V D_g C).$$

2.4.5. Examples

1. Loglinear models.

As noted previously, the loglinear model is one particular case of model (2.38), with $g(\pi_i) = \ln \pi_i$, $x_0 = \underline{1}$ and $\bar{u}(\theta) = \ln[1/\sum \exp(x_i'\theta)]$. Use of the general theory of Section 2.4.2 for testing a nested hypothesis leads to the same asymptotic distribution of X^2 as that obtained by Rao and Scott (1984). This fact will now be shown. Since $f_i(\theta) = \exp(\bar{u}(\theta) + x_i'\theta)$, then

$$F_1 = D_\pi(I - \underline{1}_T \pi') X_1$$

and

$$F_2 = D_\pi(I - \underline{1}_T \pi') X_2,$$

where $\underline{1}_T$ is the $T \times 1$ vector of ones. Then, since

$$(I - \pi \underline{1}_T') D_\pi (I - \underline{1}_T \pi') = P,$$

we obtain the result that

$$A_2 = D_\pi [I - \underline{1}_T \pi'] \bar{X}_2,$$

where $\bar{X}_2 = (I - X_1(X_1' P X_1)^{-1} X_1' P) X_2$.

Taking into account that $V \frac{1}{\pi} = \frac{0}{\delta}$, we then find, by substitution, that

$$(A_2' D_{\pi}^{-1} A_2)^{-1} (A_2' D_{\pi}^{-1} V D_{\pi}^{-1} A_2) = (\bar{X}_2' P \bar{X}_2)^{-1} (\bar{X}_2' V \bar{X}_2),$$

which is the same asymptotic matrix found by Rao and Scott (1984, Theorem 1) for the loglinear model.

2. Test of marginal homogeneity in a square table

There are, as well, models following the form of (2.38) which are not loglinear models. Consider, for example, the hypothesis of marginal homogeneity in a square table. For ease of presentation, we will consider a 3×3 table, where marginal homogeneity is defined by the conditions $\pi_{1+} = \pi_{+1}$, $\pi_{2+} = \pi_{+2}$ and $\pi_{3+} = \pi_{+3}$. Any two of these conditions imply the third, and in particular, if $\pi_{1+} = \pi_{+1}$ and $\pi_{2+} = \pi_{+2}$, then

$$\pi_{3+} = \pi_{++} - \pi_{1+} - \pi_{2+} = \pi_{++} - \pi_{+1} - \pi_{+2} = \pi_{+3}.$$

In matrix form, the marginal homogeneity conditions are

$$(2.52) \quad \begin{bmatrix} 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{13} \\ \pi_{21} \\ \pi_{22} \\ \pi_{23} \\ \pi_{31} \\ \pi_{32} \\ \pi_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

If we rearrange the π_{ij} 's and let $\underline{\pi} = (\pi_{12}\pi_{13}\pi_{11}\pi_{21}\pi_{22}\pi_{23}\pi_{31}\pi_{32}\pi_{33})$
 $= (\pi_{12}\pi_{13}\pi_{-2})'$, then (2.52) is equivalent to

$$(2.53) \quad \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \pi_{12} \\ \pi_{13} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 \end{bmatrix} \underline{\pi}_{-2}$$

Since the 2×2 matrix on the left is nonsingular, (2.53) is equivalent to

$$(2.54) \quad \begin{bmatrix} \pi_{12} \\ \pi_{13} \end{bmatrix} = B \underline{\pi}_{-2},$$

where

$$B = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 \end{bmatrix},$$

or to

$$(2.55) \quad \underline{\pi} = \begin{bmatrix} B \\ I_7 \end{bmatrix} \underline{a},$$

where $\underline{a}$ is a 7×1 vector of parameters. Obviously, the coefficient matrix in (2.55) is 9×7 of rank 7. In the form (2.55), the marginal homogeneity conditions hold, but the restriction $\underline{1}'\underline{\pi} = 1$ will be satisfied only if $\underline{1}'(B', I)' \underline{a} = 1$, i.e.

$$(2.56) \quad a_1 = 1 - 2a_2 - a_3 - a_4 - 2a_5 - a_6 - a_7.$$

Using the partitioning of $(B', I)'$,

$$(2.57) \quad \begin{bmatrix} B \\ I \end{bmatrix} = (\underline{x}_0, X),$$

where $\underline{x}_0$ is a 9×1 vector and X is a 9×6 matrix, model (2.55)

with additional restriction (2.56) may be written as

$$(2.58) \quad \underline{\pi} = \bar{u}(\underline{\theta}) \underline{x}_0 + \underline{x} \underline{\theta},$$

where $\underline{\theta}$ is a 6×1 vector of parameters, $\bar{u}(\underline{\theta})$ is a constant satisfying $\bar{u}(\underline{\theta}) = 1 - 2\theta_1 - \theta_2 - \theta_3 - 2\theta_4 - \theta_5 - \theta_6$, $\underline{x}$ is a 9×6 matrix of rank 6 and $\underline{x}_0' \underline{x} = \underline{0}$. The marginal homogeneity model is thus expressed in the form of (2.38), with $g(\pi_i) = \pi_i$ so that $D_g = I$. To test for the fit of this marginal homogeneity model, we would solve the "pseudo likelihood equations" (2.17) from the sample data, obtaining $\hat{\underline{\theta}}$ and $\hat{\underline{\pi}} = \underline{f}(\hat{\underline{\theta}})$, and then calculate statistic X^2 given by (2.18).

According to the results of Section 2.4.4, the asymptotic distribution of X^2 would be $\delta_1 W_1 + \delta_2 W_2$, where W_1 and W_2 are independent X_1^2 variables and δ_1 and δ_2 are the eigenvalues of $(C'PC)^{-1}(C'VC)$, C being any 9×2 matrix of rank 2 such that $C'X = \underline{0}$ and $C'\underline{x}_0 = \underline{0}$. One possible method to obtain a C matrix is now described. Let A represent the 9×7 matrix $(\underline{x}_0, \underline{x})$, where A is of rank 7, and let $\bar{A}$ be the projection matrix $\bar{A} = I - A(A'A)^{-1}A'$, where, using the fact that $\underline{x}_0' \underline{x} = \underline{0}$, we have $\bar{A}\underline{x}_0 = \underline{0}$ and $\bar{A}\underline{x} = \underline{0}$. Since $\bar{A}$ is 9×9 of rank 2, any 2 linearly independent columns of $\bar{A}$ would satisfy the requirements for the matrix C .

3. Model of Grizzle, Starmer and Koch

As noted in the first paragraph of Section 2.4, the model $G\underline{\pi} = H\underline{\theta}$, where G is $u \times T$ of rank u and H is $u \times v$ of rank v presented by Grizzle, Starmer, and Koch (1969) does not follow the form of a loglinear model. It can, however, be put into the form of model (2.38). We will now outline a method of doing this. First of all, we show in the Appendix that $G\underline{\pi} = H\underline{\theta}$ is equivalent to

$\pi = FV$, where F is $T \times T-u+v$ of rank $s = T-u+v$. Then, using Theorem 8.9.7 of Graybill (1969), F may be expressed as

$$F = Q \begin{pmatrix} D \\ 0 \end{pmatrix} P,$$

where Q is a $T \times T$ orthogonal matrix, P is an $s \times s$ orthogonal matrix, D is an $s \times s$ diagonal matrix with the (nonzero) eigenvalues of $F'F$ on the diagonal and 0 is a $T-s \times s$ matrix of zeroes. Now, the matrix $\tilde{Q} = Q \begin{pmatrix} D \\ 0 \end{pmatrix}$, which is $T \times s$, has orthogonal columns and is of rank s , may be expressed as $\tilde{Q} = (x_0, X)$, where $x_0'X = 0$. Also, let $PV = \underline{\beta} = (\beta_1, \beta_2)'$, where β_2 is $(s-1) \times 1$. Then $\pi = FV$ may be expressed as

$$\pi = \beta_1 x_0 + X\beta_2,$$

where β_1 is the function of β_2 defined by

$$\beta_1 = (1'x_0)^{-1}(1 - 1'X\beta_2)$$

so that the restriction $1'\pi = 1$ is satisfied. The model is now in the form of (2.38). It should be noted that the marginal homogeneity model above can be expressed as a special case of the Grizzle, Starmer and Koch (1969) model, with $H = 0$.

2.4.7. Adjustments to X^2

Since the asymptotic distribution of X^2 (or G^2) is complex, it is important to have a correction for X^2 that could be practically implemented with survey data. One approach likely to be adequate in

many situations is to treat $X^2 \hat{\delta}$ as χ_u^2 under H , where δ may be expressed as

$$(2.59) \quad u\delta = \sum \delta_i = \text{tr}(A_2' D_{\pi}^{-1} A_2)^{-1} (A_2' D_{\pi}^{-1} V D_{\pi}^{-1} A_2),$$

and $\hat{\delta}$ is the survey estimate of δ . However, $\hat{\delta}$ depends on the full estimated covariance matrix of $\hat{p}$ except in the case of loglinear models with direct estimates (see Section 2.2). Since the estimated covariance matrix is frequently unavailable, it is worthwhile, therefore, to obtain an upper bound on δ which depends only on the cell design effects or on generalized design effects independent of any null hypothesis. One method to obtain such a bound is described below.

Since $A_2' D_{\pi}^{-1} A_2$ is a symmetric matrix, there exists a nonsingular matrix B such that

$$B A_2' D_{\pi}^{-1} A_2 B = D,$$

where D is a diagonal matrix (C.R. Rao (1965), p.20). Let $Y = B A_2'$
 $= (Y_1, Y_2, \dots, Y_u)'$. We then have $Y D_{\pi}^{-1} Y' = D$ so that

$$Y_i' D_{\pi}^{-1} Y_j = 0 \text{ for } i \neq j.$$

Therefore,

$$\begin{aligned} \text{tr}(A_2' D_{\pi}^{-1} A_2)^{-1} (A_2' D_{\pi}^{-1} V D_{\pi}^{-1} A_2) &= \text{tr}(B')^{-1} (A_2' D_{\pi}^{-1} A_2)^{-1} B^{-1} B (A_2' D_{\pi}^{-1} V D_{\pi}^{-1} A_2) B' \\ &= \text{tr}(Y D_{\pi}^{-1} Y)^{-1} (Y D_{\pi}^{-1} V D_{\pi}^{-1} Y) \\ (2.60) \quad &= \sum_{i=1}^u \frac{Y_i' D_{\pi}^{-1} V D_{\pi}^{-1} Y_i}{Y_i' D_{\pi}^{-1} Y_i} \end{aligned}$$

$$= \sum_{i=1}^u \frac{\underline{w}_i' (D_{\pi}^{-1/2} V D_{\pi}^{-1/2}) \underline{w}_i}{\underline{w}_i' \underline{w}_i},$$

where $\underline{w}_i = D_{\pi}^{-1/2} \underline{y}_i$. Now, since $\underline{w}_i' \underline{w}_j = 0$ for $i \neq j$, we can conclude that

$$(2.61) \quad \text{tr}(A_2' D_{\pi}^{-1} A_2)^{-1} (A_2' D_{\pi}^{-1} V D_{\pi}^{-1} A_2) \leq \lambda_1 + \lambda_2 + \dots + \lambda_u,$$

where $\lambda_1 \geq \lambda_2 \dots \geq \lambda_{T-1}$ are the non-zero eigenvalues of $D_{\pi}^{-1/2} V D_{\pi}^{-1/2}$, or, equivalently, of $D_{\pi}^{-1} V$. The inequality in (2.61) follows from the standard matrix result of C.R. Rao (1965), p.63. From (2.59) and (2.61) it follows that

$$(2.62) \quad \delta \leq u^{-1} \sum_{i=1}^u \lambda_i.$$

It should be noted that the eigenvalues of $D_{\pi}^{-1} V$ are independent of any hypothesis, so that if all these eigenvalues were reported along with the published tables, an estimated upper bound on δ could be obtained for any hypothesis, using (2.62).

If the estimated eigenvalues of $D_{\pi}^{-1} V$ are not available, but deffs, d_i , of the $\hat{p}_i$, the survey estimates of the cell proportions, are published, an upper bound may still be estimated by noting that

$$(2.63) \quad u\delta \leq \sum_{i=1}^u \lambda_i \leq \sum_{i=1}^T \lambda_i = T\lambda.$$

and

$$(2.64) \quad T\lambda = \text{tr } D_{\pi}^{-1} V = \sum_{i=1}^T v_{ii}/\pi_i = \sum_{i=1}^T (1-\pi_i) d_i$$

where v_{ii} is the i^{th} diagonal element of V and d_i is the deff of

$\hat{p}_i$. However, this upper bound on δ is not satisfactory if u is small compared to T .

When the full estimated covariance matrix of $\hat{p}$ is available, a more accurate correction to X^2 (or G^2) is obtained by the Satterthwaite approximation in which

$$(2.65) \quad X_s^2 = X^2 / [(1+a^2)\delta]$$

is treated as χ_v^2 , where $v = u/(1+a^2)$ and $a^2 = \sum (\delta_i - \delta)^2 / u\delta^2$ is the square of the coefficient of variation of the δ_i 's. As noted by Rao & Scott (1984), "it is, however, not necessary to evaluate the individual δ_i 's in order to compute X_s^2 , since $\sum \delta_i = \bar{E}(X^2)$ and $2\sum \delta_i^2 = \bar{V}(X^2)$ where $\bar{E}$ and $\bar{V}$ respectively denote the asymptotic expectation and asymptotic variance operators." In fact, Rao and Scott's approach may be duplicated, except for the modification that $\sqrt{n}(\hat{\pi} - \pi) \approx Y$, where Y is a $N(0, B)$ vector with $B = A_2' C(\hat{\theta}_2) A_2$ and $C(\hat{\theta}_2)$ is given by (2.36).

CHAPTER 3

BINARY RESPONSE MODELS

3.1 Introduction

Loglinear models and their generalizations, such as those described in Chapter 2, are primarily designed for investigating relationships among several categorical variables. The distinction between a response variable and an explanatory variable is often not crucial. There are, however, many data sets to be analysed in which one variable is considered a response variable and all other variables are considered to be explanatory. If the response variable can take on only two values and if the explanatory variables can be categorical or continuous, models describing the relationship between the response variable and the explanatory variables are called binary response models. Two such models frequently encountered are the logit and probit models.

Possible labels applied to the categories of the response variable include "alive or dead", "present or absent" and "yes or no". In keeping with the convention set out by other authors (see, for example, Dobson (1983)), we will generally use the terminology "success or failure" for the response categories in the work to follow.

Traditionally, when a binary response model has been employed, the sampling scheme for collecting data has been the product binomial. With this sampling plan, the Pearson chi-squared statistics for testing hypotheses are asymptotically distributed as χ^2 random variables. (See, for example, Cox (1970) and McCullagh and Nelder (1983).)

However, as was the case for loglinear models, the standard asymptotic theory is not valid when binary response models are used for assessing survey data involving complex sampling schemes. In Section 3.2 the asymptotic theory for the Pearson chi-squared statistic will be developed under a specified sampling scheme when a general binary response model is used. The parameters of these models will be estimated by maximum likelihood methods appropriate to product binomial sampling. Corrections to the Pearson chi-squared statistic to take account of the true sampling plan will be obtained. Section 3.3 will apply the results of Section 3.2 to four particular binary response models - logit, probit, complementary log-log and Box-Cox transformation models.

A subset of logit models may be derived from hierarchical loglinear models. For this subset, the existence of direct estimates of success rates is of interest. For loglinear models existence of such estimates meant that δ , (used to modify the Pearson chi-squared statistic) could be expressed in terms of just the cell and marginal deffs. In Section 3.4, the conditions for the existence of such estimates for logit models are developed.

3.2 Asymptotic Distribution of X^2 for Binary Response Models

Suppose that the finite population of interest of size N is partitioned into R mutually exclusive exhaustive domains, the i^{th} of size N_i . The partitioning is often done according to the levels of one or more factors measured in the survey. In the i^{th} domain, suppose that there are N_{i1} successes, and thus a success rate of

$$p_i = N_{i1}/N_i$$

A general binary response model for the proportions P_i , $i=1,2,\dots,R$ may be written as

$$(3.1) \quad P_i = m_i(\underline{\theta})$$

where $\underline{\theta}$ is an r -vector of unknown parameters ($r \leq R$) and m_i is a function of known form (generally depending on the domain) satisfying the regularity conditions of Birch (1964), given by BFH (1975, p.510).

Suppose that a sample $\tilde{s}$ of n ultimate units is drawn from the population according to a specified survey design $p(\tilde{s})$. Let $\hat{N}$, $\hat{N}_i$ and $\hat{N}_{i1}$ be the survey estimates of the total population size, the size of the i^{th} domain and the number of successes in the i^{th} domain respectively, such that the following two assumptions hold:

$$(a) \quad \underline{w} = \underline{W} + o_p(n^{-1/2}) \quad \text{where} \quad \underline{w} = (w_1, w_2, \dots, w_R)',$$

$$w_i = \hat{N}_i / \hat{N}, \quad W = (W_1, W_2, \dots, W_R) \quad \text{and} \quad W_i = N_i / N.$$

(b) $\sqrt{n}(\hat{\underline{P}} - \underline{P})$ is asymptotically distributed as a $N(\underline{0}, \underline{\Sigma})$ vector as $n \rightarrow \infty$, where $\underline{P} = (P_1, P_2, \dots, P_R)$, $\hat{\underline{P}} = (\hat{P}_1, \hat{P}_2, \dots, \hat{P}_R)$ and $\hat{P}_i = \hat{N}_{i1} / \hat{N}_i$ is the survey estimate of the proportion of successes in the i^{th} domain. The estimates w_i and $\hat{P}_i$ are typically ratio estimates.

The likelihood under the assumption of independent binomial sampling in each domain is

$$(3.2) \quad L = \prod_{i=1}^R \binom{n_i}{x_i} m_i(\underline{\theta})^{x_i} (1 - m_i(\underline{\theta}))^{n_i - x_i}$$

where n_i and x_i are the size of sample and number of sampled successes respectively in the i^{th} domain, with $\sum n_i = n$. When L is maximized, the likelihood equations are

$$(3.3) \quad \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\underline{\hat{\theta}}} D(\underline{m}(\underline{\hat{\theta}}))^{-1} D(\underline{1}-\underline{m}(\underline{\hat{\theta}}))^{-1} D(\underline{n}/n) \underline{q} = \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\underline{\hat{\theta}}} D(\underline{1}-\underline{m}(\underline{\hat{\theta}}))^{-1} (\underline{n}/n)$$

where $\underline{q} = (q_1, q_2, \dots, q_R)$ with $q_t = x_t/n_t$, $\left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)$ is the $R \times r$ Jacobian matrix with (i, j) th element $\partial m_i(\underline{\theta})/\partial \theta_j$, $\underline{n}/n = (n_1/n, n_2/n, \dots, n_R/n)'$, $D(\underline{m}(\underline{\hat{\theta}})) = \text{diag}(m_1(\underline{\hat{\theta}}), m_2(\underline{\hat{\theta}}), \dots, m_R(\underline{\hat{\theta}}))$ and similarly for $D(\underline{1}-\underline{m}(\underline{\hat{\theta}}))$ and $D(\underline{n}/n)$. Because the true sampling plan is not product binomial, "pseudo likelihood equations" are obtained by replacing $(\underline{n}/n)$ with $\underline{w}$ and $\underline{q}$ with $\underline{\hat{p}}$ in (3.3), yielding

$$(3.4) \quad \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\underline{\hat{\theta}}} D(\underline{m}(\underline{\hat{\theta}}))^{-1} D(\underline{1}-\underline{m}(\underline{\hat{\theta}}))^{-1} D(\underline{w}) \underline{\hat{p}} = \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\underline{\hat{\theta}}} D(\underline{1}-\underline{m}(\underline{\hat{\theta}}))^{-1} \underline{w}.$$

The solution of (3.4) for $\underline{\hat{\theta}}$, and the resulting $\underline{\hat{m}} = \underline{m}(\underline{\hat{\theta}})$ are the pseudo mle's for $\underline{\theta}$ and $\underline{p}$ respectively. The same result (3.4) would obviously be obtained if n_i and x_i in (3.2) were replaced by $n\bar{w}_i$ and $n_i \hat{p}_i$ respectively, to yield $\bar{L}$, and then $\bar{L}$ was maximized. The consistency of $\underline{\hat{p}}$ and $\underline{w}$ ensures the consistency of $\underline{\hat{\theta}}$ and $\underline{\hat{m}}$, under standard regularity conditions.

The Pearson chi-squared statistic used to test $H: \underline{p} = \underline{m}(\underline{\theta})$ in the classical product binomial situation is

$$(3.5) \quad \begin{aligned} X_C^2 &= \sum (x_i - n_i m_i(\underline{\hat{\theta}}))^2 / [n_i m_i(\underline{\hat{\theta}}) (1 - m_i(\underline{\hat{\theta}}))] \\ &= n \sum (n_i/n) (x_i/n_i - m_i(\underline{\hat{\theta}}))^2 / [m_i(\underline{\hat{\theta}}) (1 - m_i(\underline{\hat{\theta}}))] \end{aligned}$$

where $\underline{\hat{\theta}}$ is the solution of (3.3). For a general sampling scheme, the parallel statistic obtained by substituting w_i and $\hat{p}_i$ for n_i/n and x_i/n_i respectively and having $\underline{\hat{\theta}}$ the solution of (3.4) is

$$(3.6) \quad \begin{aligned} X^2 &= n \sum w_i (\hat{p}_i - m_i(\underline{\hat{\theta}}))^2 / [m_i(\underline{\hat{\theta}}) (1 - m_i(\underline{\hat{\theta}}))] \\ &= n (\underline{\hat{p}} - \underline{m}(\underline{\hat{\theta}}))' D(\underline{w}) D(\underline{m}(\underline{\hat{\theta}}))^{-1} D(\underline{1}-\underline{m}(\underline{\hat{\theta}}))^{-1} (\underline{\hat{p}} - \underline{m}(\underline{\hat{\theta}})). \end{aligned}$$

The asymptotic distribution of X^2 will be developed as a special case of the asymptotic distribution of a statistic used for nested models.

3.2.1 Asymptotic Covariance Matrix of $\hat{\pi}$

Lemma: Under the regularity conditions of Birch (1964),

$$(3.7) \quad \hat{\theta} - \theta = \left(\frac{\partial m'}{\partial \theta} E \frac{\partial m}{\partial \theta} \right)^{-1} \frac{\partial m'}{\partial \theta} E (\hat{P} - \underline{m}(\theta)) + o_p(n^{-1/2})$$

where $E = D(W)D(\underline{m}(\theta))^{-1}D(\underline{1} - \underline{m}(\theta))^{-1}$ and " $\sim$ " denotes asymptotic equivalence.

Proof: The function $\bar{L}$ which is maximized to obtain $\hat{\theta}$ is defined by

$$\begin{aligned} -2 \ln \bar{L} &= -2n \sum_i w_i \hat{p}_i \ln m_i(\theta) - 2n \sum_i w_i (1 - \hat{p}_i) \ln (1 - m_i(\theta)) - 2 \sum_i \ln \left(\frac{nw_i}{nw_i \hat{p}_i} \right) \\ &= 2n \bar{G}^2(\hat{V}, \underline{Y}(\theta)) + 2n \sum_i w_i \ln w_i - 2n \sum_i w_i \hat{p}_i \ln w_i \hat{p}_i \\ &\quad - 2n \sum_i w_i (1 - \hat{p}_i) \ln w_i (1 - \hat{p}_i) - 2 \sum_i \ln \left(\frac{nw_i}{nw_i \hat{p}_i} \right) \\ &= 2n \bar{G}^2(\hat{V}, \underline{Y}(\theta)) + \text{terms not involving } \theta \end{aligned}$$

where

$$(3.8) \quad \hat{V} = (\hat{P}' D(W), (1 - \hat{P})' D(W))',$$

$$(3.9) \quad \underline{Y}(\theta) = (\underline{m}(\theta)' D(W), (1 - \underline{m}(\theta))' D(W))',$$

and $\bar{G}^2(\underline{x}, \underline{y}) = \sum x_i \ln(x_i/y_i)$. It is readily proved that $\underline{1}' \hat{V} = \underline{1}' \underline{Y}(\theta) = 1$. To maximize $\bar{L}$, we must choose the value of θ to minimize $\bar{G}^2(\hat{V}, \underline{Y}(\theta))$. Birch (1964) defines $\hat{\theta}(\hat{V})$ to be that value of θ such that

$$\bar{G}^2(\underline{v}, \underline{Y}(\underline{\theta}(\underline{v}))) = \inf_{\underline{\theta}} \bar{G}^2(\underline{v}, \underline{Y}(\underline{\theta})).$$

The value of $\underline{\theta}$ to maximize $\bar{L}$ is thus $\hat{\underline{\theta}} = \underline{\theta}(\hat{\underline{v}})$. Birch (1964) proved that

$$(3.10) \quad \underline{\theta}(\underline{v}) = \underline{\theta} + (B'B)^{-1} B'D(\underline{Y}(\underline{\theta}))^{-1} (\underline{v} - \underline{Y}(\underline{\theta})) + o(\|\underline{v} - \underline{Y}(\underline{\theta})\|)$$

as $\underline{v} \rightarrow \underline{Y}(\underline{\theta})$, where B is the $2R \times r$ matrix defined by $B = D(\underline{Y}(\underline{\theta}))^{-1} \frac{\partial \underline{Y}}{\partial \underline{\theta}}$, and $\underline{v}$ has the form $(\underline{P}'D(\underline{w}), (1-\underline{P})'D(\underline{w}))'$.

In order to express (3.10) in terms of the vector $\underline{m}(\underline{\theta})$, we note that

$$\underline{v} - \underline{Y}(\underline{\theta}) = (D(\underline{w}), -D(\underline{w}))'(\underline{P} - \underline{m}(\underline{\theta})),$$

$$D(\underline{Y}(\underline{\theta}))^{-1} = \begin{bmatrix} D(\underline{w})^{-1} D(\underline{m}(\underline{\theta}))^{-1} & \\ 0 & D(\underline{w})^{-1} D(1 - \underline{m}(\underline{\theta}))^{-1} \end{bmatrix},$$

and

$$\frac{\partial \underline{Y}}{\partial \underline{\theta}} = (D(\underline{w}), -D(\underline{w}))' \frac{\partial \underline{m}}{\partial \underline{\theta}},$$

so that

$$B'B = \frac{\partial \underline{m}'}{\partial \underline{\theta}} \bar{E} \frac{\partial \underline{m}}{\partial \underline{\theta}},$$

and

$$B'D(\underline{Y}(\underline{\theta}))^{-1} (\underline{v} - \underline{Y}(\underline{\theta})) = \frac{\partial \underline{m}'}{\partial \underline{\theta}} \bar{E} (\underline{P} - \underline{m}(\underline{\theta}))$$

where $\bar{E} = D(\underline{w})D(\underline{m}(\underline{\theta}))^{-1}D(1 - \underline{m}(\underline{\theta}))^{-1}$. Substituting the above expressions into (3.10), we have

$$\underline{\theta}(\underline{v}) = \underline{\theta} + \left(\frac{\partial \underline{m}'}{\partial \underline{\theta}} \bar{E} \frac{\partial \underline{m}}{\partial \underline{\theta}} \right)^{-1} \frac{\partial \underline{m}'}{\partial \underline{\theta}} \bar{E} (\underline{P} - \underline{m}(\underline{\theta})) + o(\|\underline{v} - \underline{Y}(\underline{\theta})\|) \text{ as } \underline{v} \rightarrow \underline{Y}(\underline{\theta}).$$

Since, by assumption, $\sqrt{n}(\underline{\hat{p}} - \underline{m}(\underline{\theta})) \approx N(\underline{0}, \Sigma)$, and $\underline{w} = \underline{W} + o_p(n^{-1/2})$, we then have $\underline{\hat{p}} - \underline{m}(\underline{\theta}) = O_p(n^{-1/2})$ and $\underline{\hat{y}} - \underline{y}(\underline{\theta}) = O_p(n^{-1/2})$.

Therefore it follows that

$$(3.11) \quad \underline{\hat{\theta}} = \underline{\theta}(\underline{\hat{p}}) = \underline{\theta} + \left(\frac{\partial \underline{m}'}{\partial \underline{\theta}} \right)^{-1} \frac{\partial \underline{m}'}{\partial \underline{\theta}} \underline{E}(\underline{\hat{p}} - \underline{m}(\underline{\theta})) + o_p(n^{-1/2}). \quad \text{Q.E.D.}$$

Since we assumed that $\sqrt{n}(\underline{\hat{p}} - \underline{m}(\underline{\theta})) \approx N(\underline{0}, \Sigma)$, the asymptotic covariance matrix of $\underline{\hat{p}}$ is

$$(3.12) \quad C(\underline{\hat{p}}) = \Sigma/n.$$

This fact, together with (3.7), implies that the asymptotic covariance matrix of $\underline{\hat{\theta}}$ is

$$(3.13) \quad C(\underline{\hat{\theta}}) = n^{-1} (\underline{M}' \underline{E} \underline{M})^{-1} (\underline{M}' \underline{E} \Sigma \underline{E} \underline{M}) (\underline{M}' \underline{E} \underline{M})^{-1},$$

where, for the remainder of the chapter, $\underline{M}$ denotes the matrix $\frac{\partial \underline{m}}{\partial \underline{\theta}}$.

In the case of simple random sampling, we have $\Sigma = \underline{E}^{-1}$ and $C(\underline{\hat{\theta}})$ reduces to

$$C_{\underline{M}}(\underline{\hat{\theta}}) = n^{-1} (\underline{M}' \underline{E} \underline{M})^{-1}.$$

Because of the regularity conditions, we have

$$(3.14) \quad \underline{m}(\underline{\hat{\theta}}) - \underline{m}(\underline{\theta}) = \underline{M}(\underline{\hat{\theta}} - \underline{\theta}) + o_p(n^{-1/2}).$$

Substituting (3.11) into (3.14), we have

$$(3.15) \quad \underline{m}(\underline{\hat{\theta}}) - \underline{m}(\underline{\theta}) = \underline{M}(\underline{M}' \underline{E} \underline{M})^{-1} \underline{M}' \underline{E}(\underline{\hat{p}} - \underline{m}(\underline{\theta})) + o_p(n^{-1/2})$$

so that the asymptotic covariance matrix of $\underline{m}(\underline{\hat{\theta}})$ is

$$(3.16) \quad C(\underline{m}(\underline{\hat{\theta}})) = \underline{M} C(\underline{\hat{\theta}}) \underline{M}'.$$

As noted in Chapter 2, the vector $\underline{\hat{p}} - \underline{m}(\underline{\hat{\theta}})$, called the vector of

residuals, is often useful in detecting model deviations. Since

$$(3.17) \quad \begin{aligned} \underline{\hat{p}} - \underline{m}(\underline{\hat{\theta}}) &= (\underline{\hat{p}} - \underline{m}(\underline{\theta})) - (\underline{m}(\underline{\hat{\theta}}) - \underline{m}(\underline{\theta})) \\ &= (I - M(M'EM)^{-1}M'E)(\underline{\hat{p}} - \underline{m}(\underline{\theta})) + o_p(n^{-1/2}), \end{aligned}$$

then the asymptotic covariance matrix of the residual vector is

$$(3.18) \quad C(\underline{\hat{p}} - \underline{m}(\underline{\hat{\theta}})) = n^{-1} A \Sigma A'$$

where

$$A = (I - M(M'EM)^{-1}M'E).$$

Consistent estimates of the above covariance matrices are obtained by replacing $\underline{\theta}$ by $\underline{\hat{\theta}}$ and Σ by the survey estimate $\hat{\Sigma}$ (when available).

3.2.2. Nested Models

Suppose that, in model $M_1 : \underline{p} = \underline{m}(\underline{\theta})$, the vector of parameters $\underline{\theta}$ could be partitioned as $\underline{\theta} = (\underline{\theta}_1', \underline{\theta}_2')'$, where $\underline{\theta}_1$ is $s \times 1$ and $\underline{\theta}_2$ is $u \times 1$ ($s+u = r$). If we wished to test the hypothesis $H : \underline{\theta}_2 = \underline{0}$ given M_1 , we would have the reduced model $M_2 : \underline{p} = \underline{m}(\underline{\theta}_1)$ where " $\underline{m}(\underline{\theta}_1)$ " denotes " $\underline{m}(\underline{\theta}_1, \underline{0})$ ". Let $\underline{\hat{\theta}}_1$ and $\underline{\hat{m}} = \underline{m}(\underline{\hat{\theta}}_1)$ be the pseudo mle's of $\underline{\theta}_1$ and $\underline{p}$ respectively under M_2 , obtained from the pseudo likelihood equations

$$\left(\frac{\partial \underline{m}'}{\partial \underline{\theta}_1} \right)_{\underline{\theta}=(\underline{\hat{\theta}}_1, \underline{0})} D(\underline{m}(\underline{\hat{\theta}}_1))^{-1} D(\underline{1}-\underline{m}(\underline{\hat{\theta}}_1))^{-1} D(\underline{w}) \underline{\hat{p}} = \left(\frac{\partial \underline{m}'}{\partial \underline{\theta}_1} \right)_{\underline{\theta}=(\underline{\hat{\theta}}_1, \underline{0})} D(\underline{1}-\underline{m}(\underline{\hat{\theta}}_1))^{-1} \underline{w}.$$

The Pearson chi-squared statistic for testing H is given by

$$(3.19) \quad \begin{aligned} \chi^2 &= n \sum_i w_i (\underline{m}_i(\underline{\hat{\theta}}) - \underline{m}_i(\underline{\hat{\theta}}_1))^2 / (\underline{m}_i(\underline{\hat{\theta}}_1) (1 - \underline{m}_i(\underline{\hat{\theta}}_1))) \\ &= n (\underline{m}(\underline{\hat{\theta}}) - \underline{m}(\underline{\hat{\theta}}_1))' D(\underline{w}) D(\underline{m}(\underline{\hat{\theta}}_1))^{-1} D(\underline{1}-\underline{m}(\underline{\hat{\theta}}_1))^{-1} (\underline{m}(\underline{\hat{\theta}}) - \underline{m}(\underline{\hat{\theta}}_1)). \end{aligned}$$

By definition,

$$\frac{\partial \underline{m}}{\partial \underline{\theta}} = \left(\frac{\partial \underline{m}}{\partial \underline{\theta}_1}, \frac{\partial \underline{m}}{\partial \underline{\theta}_2} \right)$$

and, since $\frac{\partial \underline{m}}{\partial \underline{\theta}}$ is $R \times r$ of rank r , $\frac{\partial \underline{m}}{\partial \underline{\theta}_1}$ must be of rank s and $\frac{\partial \underline{m}}{\partial \underline{\theta}_2}$ of rank u . Let M_1 denote $\frac{\partial \underline{m}}{\partial \underline{\theta}_1}$ and M_2 denote $\frac{\partial \underline{m}}{\partial \underline{\theta}_2}$.

Because of the regularity assumptions,

$$(3.20) \quad \underline{m}(\hat{\underline{\theta}}) - \underline{m}(\underline{\theta}) \sim M(\hat{\underline{\theta}} - \underline{\theta}) = M_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1) + M_2(\hat{\underline{\theta}}_2 - \underline{\theta}_2)$$

and, similarly, under H ,

$$(3.21) \quad \underline{m}(\hat{\underline{\theta}}_1) - \underline{m}(\underline{\theta}) \sim M_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1).$$

Thus, under H ,

$$(3.22) \quad \underline{m}(\hat{\underline{\theta}}) - \underline{m}(\hat{\underline{\theta}}_1) \sim M_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1) + M_2 \hat{\underline{\theta}}_2 - M_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1)$$

where, parallel to (3.7),

$$(3.23) \quad (\hat{\underline{\theta}}_1 - \underline{\theta}_1) \sim (M_1' E M_1)^{-1} M_1' E(\underline{P} - \underline{m}(\underline{\theta})).$$

Examining (3.22), it can be seen that, if $(\hat{\underline{\theta}}_1 - \underline{\theta}_1)$ could be expressed asymptotically as $(\hat{\underline{\theta}}_1 - \underline{\theta}_1) + Q \hat{\underline{\theta}}_2$ for some matrix Q , we could express $\underline{m}(\hat{\underline{\theta}}) - \underline{m}(\hat{\underline{\theta}}_1)$ asymptotically just as a function of: $\hat{\underline{\theta}}_2$. To accomplish this, we recall that, under H ,

$$(3.24) \quad \begin{pmatrix} \hat{\underline{\theta}}_1 - \underline{\theta}_1 \\ \hat{\underline{\theta}}_2 \end{pmatrix} \sim \begin{pmatrix} M_1' E M_1 & M_1' E M_2 \\ M_2' E M_1 & M_2' E M_2 \end{pmatrix}^{-1} M' E(\underline{P} - \underline{m}(\underline{\theta})).$$

Using the standard formula for the inverse of a partitioned matrix (eg. Seber (1977), p.390), and (3.22), we find (see Appendix) that

$$Q = (M_1' E M_1)^{-1} (M_1' E M_2) .$$

Substituting for $\hat{\underline{\theta}}_1 - \underline{\theta}_1$ in (3.22) we find

$$(3.25) \quad \underline{m}(\hat{\underline{\theta}}) - \underline{m}(\hat{\underline{\theta}}_1) \sim (M_2 - M_1 Q) \hat{\underline{\theta}}_2 = H_2 \hat{\underline{\theta}}_2 ,$$

where

$$(3.26) \quad H_2 = (I - M_1 (M_1' E M_1)^{-1} M_1' E) M_2 .$$

THEOREM 3.1. Under $H : \underline{\theta}_2 = \underline{0}$,

$$(3.27) \quad X^2 \sim n \hat{\underline{\theta}}_2' (H_2' E H_2) \hat{\underline{\theta}}_2 .$$

PROOF: Under H , $D(\underline{m}(\hat{\underline{\theta}}_1))^{-1} \sim D(\underline{m}(\underline{\theta}))^{-1}$ and

$$D(\underline{1} - \underline{m}(\hat{\underline{\theta}}_1))^{-1} \sim D(\underline{1} - \underline{m}(\underline{\theta}))^{-1} \text{ where } \underline{\theta} = (\underline{\theta}_1', \underline{0})' .$$

Thus, since by assumption, $D(\underline{w}) \sim D(\underline{W})$,

$$D(\underline{w}) D(\underline{m}(\hat{\underline{\theta}}_1))^{-1} D(\underline{1} - \underline{m}(\hat{\underline{\theta}}_1))^{-1} \sim E \text{ under } H .$$

Using result (3.25), we then conclude that

$$X^2 \sim n \hat{\underline{\theta}}_2' H_2' E H_2 \hat{\underline{\theta}}_2 .$$

We can also express the asymptotic distribution of X^2 as a weighted sum of chi-squared variables as follows. In the expression for $C(\hat{\underline{\theta}})$ given by (3.13), if $M' E M$ is partitioned as in (3.24) and is inverted, the terms pertaining to $C(\hat{\underline{\theta}}_2)$ (see Appendix) are

$$(3.28) \quad C(\hat{\underline{\theta}}_2) = n^{-1} (H_2' E H_2)^{-1} (H_2' E \Sigma E H_2) (H_2' E H_2)^{-1} .$$

Thus, noting that $\hat{\underline{\theta}}_2 \approx N_u(\underline{0}, C(\hat{\underline{\theta}}_2))$, where " $\approx$ " denotes "asymptotically distributed as", we have

$$X^2 \approx \sum_{i=1}^u \delta_i Z_i^2,$$

where the Z_i^2 are independent χ_1^2 variables and the δ_i are the eigenvalues of

$$(3.29) \quad nC(\hat{\theta}_2)(H_2'EH_2) = (H_2'EH_2)^{-1}(H_2'E\bar{\Sigma}E'H_2).$$

As a special case, under simple random sampling the $\hat{p}_i$ are independent with asymptotic variance $p_i(1-p_i)/(nW_i)$. Thus, $\bar{\Sigma} = E^{-1}$ under H , and $(H_2'EH_2)^{-1}(H_2'EE^{-1}EH_2) = I$ so that $\delta_i = 1$, $i = 1, 2, \dots, u$, and we get the standard result that $X^2 \approx \chi_u^2$ under H .

It should be noted that the asymptotic covariance matrix of $H_2'E\hat{p}$ is $n^{-1}(H_2'E\bar{\Sigma}EH_2)$ under the survey design and $n^{-1}(H_2'EE^{-1}EH_2) = n^{-1}(H_2'EH_2)$ under simple random sampling, so that the δ_i are the "generalized design effects" of the vector $H_2'E\hat{p}$, as defined by Rao and Scott (1981).

3.2.3 Statistics of the Wald Type

We will now restrict our models $P_i = m_i(\theta)$ to those for which there exists a monotone differentiable function g such that

$$(3.30) \quad g_i = g(P_i) = \underline{x}_i'\underline{\theta}, \quad i = 1, 2, \dots, R,$$

where $\underline{\theta}$ is an $r \times 1$ vector of parameters and $X = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_R)'$ is a $R \times r$ matrix of rank r ($\leq R$). As an example, for the logit model we have $P_i = m_i(\theta) = (1 + \exp(-\underline{x}_i'\underline{\theta}))^{-1}$ and $g(P_i) = \ln(P_i/(1-P_i)) = \underline{x}_i'\underline{\theta}$.

If a consistent estimator $\hat{\Sigma}/n$ of the covariance matrix of $\hat{p}$ is available, then a Wald statistic X_W^2 , that is asymptotically χ_u^2 under $H: \underline{\theta}_2 = \underline{0}$ may be constructed as follows.

Model (3.30) may be written as

$$(3.31) \quad \underline{q} = X_1 \underline{\theta}_1 + X_2 \underline{\theta}_2$$

where $\underline{q} = (q_1, q_2, \dots, q_R)'$, $X = (X_1, X_2)$, $\underline{\theta} = (\underline{\theta}_1', \underline{\theta}_2')'$, $\underline{\theta}_1$ is $s \times 1$ and $\underline{\theta}_2$ is $u \times 1$ ($s+u=r$). Since X is of full rank r , X_1 is of full rank s and X_2 is of full rank u . Let C be any $T \times u$ matrix of rank u such that $C'X_1 = 0$ and $C'X_2$ is nonsingular. (If $X_2'X_1 = 0$, then X_2 would be a choice for C .) Then, $H: \underline{\theta}_2 = \underline{0}$ is equivalent to

$$(3.32) \quad H': \underline{\phi} = C' \underline{q} = \underline{0}.$$

A Wald statistic for testing H' , based on $\hat{P}$, would be

$$(3.33) \quad \chi_W^2 = \hat{\phi}' [\hat{C}(\hat{\phi})]^{-1} \hat{\phi}$$

where $\hat{\phi} = C' \hat{g}$, $\hat{g}$ is the R -vector with $\hat{g}_i = g(\hat{\theta}_i) = \underline{x}_i' \hat{\theta}$ and $\hat{C}(\hat{\phi})$ is the estimated covariance matrix of $\hat{\phi}$ obtained in the following manner. By the delta method (see, for example, BPH (1975)),

$$\hat{g} - \underline{g} \sim \frac{\partial g}{\partial \underline{\theta}} (\hat{\underline{\theta}} - \underline{\theta}) = X(\hat{\underline{\theta}} - \underline{\theta}). \text{ We then get}$$

$$(3.34) \quad C(\hat{\phi}) = C'X C(\hat{\underline{\theta}})X'C$$

where $C(\hat{\underline{\theta}})$ is given by (3.13). Finally we obtain $\hat{C}(\hat{\phi})$ by replacing $\underline{\theta}$ by $\hat{\underline{\theta}}$, Σ by $\hat{\Sigma}$ and $\underline{w}$ by $\underline{w}$ in (3.34). Now, $\hat{\phi}$ has an asymptotic distribution of $N_u(\underline{0}, C(\hat{\phi}))$ and $\hat{C}(\hat{\phi})$ is a consistent estimator of $C(\hat{\phi})$, so that by standard results (see C.R. Rao (1965)) the asymptotic distribution of χ_W^2 is χ_u^2 .

The Wald statistic χ_W^2 requires the knowledge of the full covariance matrix of $\underline{P}$, $\hat{\Sigma}/n$. If $\hat{\Sigma}$ is not available, a Wald-type

statistic may be formed by ignoring the survey design and replacing E/n in $C(\hat{\phi})$ by E^{-1}/n , the asymptotic covariance matrix of $\hat{p}$ under simple random sampling, thus yielding

$$(3.35) \quad C_M(\hat{\phi}) = C' X C_M(\hat{\theta}) X' C,$$

where

$$(3.36) \quad C_M(\hat{\theta}) = n^{-1} \left(\frac{\partial m'}{\partial \theta} E \frac{\partial m}{\partial \theta} \right)^{-1}$$

This then leads to the test statistic

$$(3.37) \quad \bar{X}_W^2 = \hat{\phi}' (\hat{C}_M(\hat{\phi}))^{-1} \hat{\phi}$$

where $\hat{C}_M(\hat{\phi})$ is obtained by replacing θ by $\hat{\theta}$ and $\underline{W}$ by $\underline{w}$ in (3.35). Since $\hat{\phi} \approx N(\underline{0}, C(\hat{\phi}))$ and $\hat{C}_M(\hat{\phi}) \sim C_M(\hat{\phi})$, then by standard results on quadratic forms, the asymptotic distribution of $\bar{X}_W^2$ is a weighted sum $\sum \gamma_i Z_i^2$ of independent χ_1^2 variables where γ_i are the eigenvalues of $(C_M(\hat{\phi}))^{-1} C(\hat{\phi})$.

THEOREM 3.2. Under $H : \theta_2 = \underline{0}$, $\bar{X}_W^2 \sim \chi^2$.

Proof: Under H , we have

$$\begin{aligned} (3.38) \quad \hat{\phi} &= \hat{\phi} - \phi = C'(\hat{g} - g) \\ &\sim C'X(\hat{\theta} - \theta) \\ &= C'[X_1(\hat{\theta}_1 - \theta_1) + X_2\hat{\theta}_2] \\ &= C'X_2\hat{\theta}_2 \end{aligned}$$

because of the delta method and the properties of C . Also under H ,

$\hat{C}_M(\hat{\phi}) \sim C_M(\hat{\phi})$ where

$$\begin{aligned} (3.39) \quad C_M(\hat{\phi}) &= n^{-1} C'X(M'EM)^{-1}X'C \\ &= n^{-1} C'X_2(H_2'EH_2)^{-1}X_2'C \end{aligned}$$

because of the standard formula for the inverse of $(M' E M)$ partitioned as in (3.24) (see Appendix). Under H , using the fact that $C'X_2$ is nonsingular, we then have

$$(3.40) \quad \bar{X}_W^2 = \hat{\phi}' (\hat{C}_M(\hat{\phi}))^{-1} \hat{\phi} \sim n\hat{\theta}'_2 X_2' C [C'X_2 (H_2' E H_2)^{-1} X_2' C]^{-1} C'X_2 \hat{\theta}_2 \\ = n\hat{\theta}'_2 (H_2' E H_2) \hat{\theta}_2$$

which is the asymptotic distribution of X^2 , as proved in Theorem 3.1.

It then follows that $\bar{X}_W^2$ is also asymptotically distributed as

$\sum \delta_i Z_i^2$ where the Z_i^2 are independent χ_1^2 variables and the δ_i are the eigenvalues of (3.29). Thus $\{\delta_1, \delta_2, \dots, \delta_u\}$ is identical to $\{\gamma_1, \gamma_2, \dots, \gamma_u\}$.

3.2.4 Special Case of Saturated Model M_1

If $s+u=R$, model M_1 is said to be saturated, since the number of independent parameters in the model equals the number of independent proportions in the population. A general hypothesis about the fit of a model $\underline{p} = \underline{m}(\theta)$ can actually be treated as a hypothesis about the fit of a nested model, where the nesting is within a saturated M_1 .

Restricting ourselves to the subset of binary response models described in (3.30), we will now show that the matrix with eigenvalues δ_i in the asymptotic distribution of X^2 (or $\bar{X}_W^2$) can be obtained, when M_1 is saturated, without the need to compute H_2 . Under saturated M_1 , M is a $R \times R$ nonsingular matrix so that solution of the pseudo likelihood equations yields $\hat{M} = \hat{P}$. This result implies that $C(\hat{P}) = E/n$ under the survey design and $C_M(\hat{P}) = E^{-1}/n$ under simple

random sampling. However, since $\hat{g} - g \sim D_g(\hat{p} - p)$ by the delta method, where D_g is the diagonal matrix with i diagonal element equal to $\partial g_i / \partial p_i$, we know that $C(\hat{g}) = D_g C(\hat{p}) D_g$ and $C_M(\hat{g}) = D_g C_M(\hat{p}) D_g$. As well, because $\hat{\phi} = C' \hat{g}$, we deduce that $C(\hat{\phi}) = C' C(\hat{g}) C$ and $C_M(\hat{\phi}) = C' C_M(\hat{g}) C$ where C is $R \times R-s$ of rank $R-s$ with $C' X_1 = 0$. We thus find that the δ_i 's are the eigenvalues of

$$(3.41) \quad (C_M(\hat{\phi}))^{-1} C(\hat{\phi}) = (C' D_g E^{-1} D_g C)^{-1} (C' D_g \Sigma D_g C).$$

3.2.5. Adjustments to X^2

Since the asymptotic distribution of X^2 is complex, it is important to have a correction for X^2 that can be practically implemented with survey data. One approach likely to be adequate in many situations is to treat X^2/δ as χ_u^2 where

$$(3.42) \quad u\delta = \Sigma \delta_i = \text{tr}(H_2' E H_2)^{-1} (H_2' E \Sigma E H_2),$$

and $\hat{\delta}$ is the survey estimate of δ . However, $\hat{\delta}$ depends on the full estimated covariance matrix of $\hat{p}$, except in the case of some restricted logit models admitting direct estimates (see Section 3.4). Since the estimated covariance matrix is frequently unavailable, it is worthwhile, therefore, to obtain an upper bound on δ which depends only on the cell design effects or on generalized design effects independent of any null hypothesis. One method to obtain such a bound is described below.

Now, δ may be written as

$$(3.42) \quad u\delta = \text{tr}(H_2' E H_2)^{-1} (H_2' E \Sigma E H_2),$$

where H_2 is defined in (3.26). Since $H_2' E H_2$ is a symmetric matrix, there exists a nonsingular matrix B such that

$$B H_2' E H_2 B' = D,$$

where D is a diagonal matrix (C.R. Rao (1965), p.20). Let

$$Y = B H_2' = (Y_1, Y_2, \dots, Y_u)'.$$

We then have $Y E Y' = D$ so that

$$(3.43) \quad Y_i' E Y_j = 0 \quad \text{for } i \neq j.$$

Therefore,

$$\begin{aligned} (3.44) \quad \text{tr}(H_2' E H_2)^{-1} (H_2' E \Sigma E H_2) &= \text{tr}(B')^{-1} (H_2' E H_2)^{-1} B^{-1} B (H_2' E \Sigma E H_2) B' \\ &= \text{tr}(Y E Y')^{-1} (Y E \Sigma E Y') \\ &= \sum_{i=1}^u \frac{Y_i' (E \Sigma E) Y_i}{Y_i' E Y_i}. \end{aligned}$$

Lemma 3.2. Let $Y_1, Y_2, \dots, Y_u$ be $R \times 1$ vectors such that $Y_i' E Y_j = 0$, $i \neq j$. Then

$$(3.45) \quad \sum_{i=1}^u \frac{Y_i' (E \Sigma E) Y_i}{Y_i' E Y_i} \leq \lambda_1 + \lambda_2 + \dots + \lambda_u,$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_R > 0$ are the eigenvalues of $E \Sigma$.

Proof. Since E^{-1} is a nonsingular covariance matrix, E is positive definite, and can thus be expressed as

$$(3.46) \quad E = QQ' ,$$

where Q is a nonsingular matrix (C.R. Rao (1965), p.36). We then have

$$(3.47) \quad Y_i' E Y_j = z_i' z_j = 0 , \quad i \neq j ,$$

where $z_i = Q' Y_i$. Since $(Q^{-1})' E \Sigma E (Q^{-1})'$ is symmetric,

$$(3.48) \quad \sum_{i=1}^u \frac{Y_i' (E \Sigma E) Y_i}{Y_i' E Y_i} = \sum_{i=1}^u \frac{z_i' (Q^{-1})' E \Sigma E (Q^{-1})' z_i}{z_i' z_i} \leq \sum_{i=1}^u \lambda_i ,$$

where $\lambda_1 \geq \lambda_2 \dots \geq \lambda_R > 0$ are the eigenvalues of $(Q^{-1})' E \Sigma E (Q^{-1})'$.

(The inequality in (3.48) follows from the standard matrix result in C.R. Rao (1965), p.36). Also, by definition, an eigenvalue λ of $(Q^{-1})' E \Sigma E (Q^{-1})'$ must satisfy

$$(3.49) \quad |Q^{-1} E \Sigma E Q^{-1} - \lambda I| = 0 .$$

Since E is nonsingular, $|E| = |QQ'| = |Q| |Q'| \neq 0$, which then implies that (3.49) is equivalent to

$$|Q| |Q^{-1} E \Sigma E Q^{-1} - \lambda I| |Q'| |E^{-1}| = |E \Sigma + \lambda I| = 0 .$$

Thus, the eigenvalues of $E \Sigma$ are equivalent to the eigenvalues of $(Q^{-1})' E \Sigma E (Q^{-1})'$, and (3.45) holds. Q.E.D.

From the above lemma it follows that

$$(3.50) \quad \delta \leq u^{-1} \sum_{i=1}^u \lambda_i .$$

It should be noted that the eigenvalues of $E \Sigma$ are independent of any

hypothesis. Hence, it should be possible to report the eigenvalues with the published tables. Then (3.50) may be used to obtain an estimated upper bound on δ for any hypothesis. It may be noted that a particular hypothesis specifies u and hence the u largest eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_u$.

If the estimated eigenvalues of $E\Sigma$ are not available, but deffs of the estimated proportions $\hat{p}$ are published, an upper bound may still be estimated by noting that

$$(3.51) \quad u\delta \leq \sum_{i=1}^u \lambda_i \leq \sum_{i=1}^R \lambda_i = R\lambda.$$

and

$$(3.52) \quad R\lambda = \text{tr } E\Sigma = \sum_{i=1}^R v_{ii} / [P_i(1-P_i)/W_i] \\ = \sum_{i=1}^R D_i = RD.$$

where v_{ii} is the i th diagonal element of Σ and D_i is the deff of $\hat{p}_i$. However, this upper bound on δ is not satisfactory if u is small compared to R . Rao and Scott (1984a) suggest the use of an F statistic, $[G^2(2/1)/u]/[G^2(1)/(R-s)]$, in this case. This statistic performs well in controlling the size of the test when $R-s$ is large. The F statistic is incorporated into the GLIM program (McCullagh and Nelder (1983)).

When model M_1 is saturated, that is $s+u = R$, (3.51) and (3.52) lead to

$$(3.53) \quad (R-s)\delta \leq RD.$$

It follows from (3.53) that $D.$ is a satisfactory upper bound on $\delta.$ if $s \ll R$ as in the application of the results of this section to Canadian Labor Force Survey data in Chapter 4. The inequality (3.53) may also be obtained directly from (3.42) as follows:

$$\begin{aligned}
 (3.54) \quad u\delta. &= \text{tr}(H_2'EH_2)^{-1}(H_2'E \Sigma EH_2) \\
 &= \text{tr}(M'E M)^{-1}(M'E \Sigma E M) - \text{tr}(M_1'EM_1)^{-1}(M_1'E \Sigma EM_1) \quad (\text{see Appendix}) \\
 &= (s+u)\gamma. - s\gamma_1. \quad (\text{say}) \\
 &\leq (s+u)\gamma.,
 \end{aligned}$$

since all eigenvalues of $(M_1'E M_1)^{-1}(M_1'E \Sigma EM_1)$ are non-negative. When M_1 is saturated, (3.54) reduces to

$$(3.55) \quad (R-s)\delta. \leq R\gamma.$$

where, since M is a nonsingular $R \times R$ matrix in this case,

$$\begin{aligned}
 R\gamma. &= \text{tr}(M'E M)^{-1}(M'E \Sigma E M) \\
 &= \text{tr} M^{-1}E^{-1}(M')^{-1}(M'E \Sigma E M) \\
 &= \text{tr} \Sigma E \\
 &= R\lambda. = RD.
 \end{aligned}$$

When the full estimated covariance matrix of $\hat{p}$ is available, a more accurate correction to X^2 (or G^2) is obtained by a Satterthwaite approximation in which

$$(3.56) \quad X_s^2 = X^2 \gamma[(1+a^2)\delta.] \quad (G_s^2 = G^2 / [(1+a^2)\delta.])$$

is treated as χ_v^2 where $v = u/(1+a^2)$ and $a^2 = \sum (\delta_i - \bar{\delta})^2 / u \bar{\delta}^2$ is the square of the coefficient of variation of the δ_i 's. As noted by Rao and Scott (1984), "it is, however, not necessary to evaluate the individual δ_i 's in order to compute X_s^2 , since $\sum \delta_i = \bar{E}(X^2)$ and $2\sum \delta_i^2 = \bar{V}(X^2)$ where $\bar{E}$ and $\bar{V}$ respectively denote the asymptotic expectation and asymptotic variance operators." Now, $\bar{E}(X^2)$ and $\bar{V}(X^2)$ may be calculated from the covariance matrix of $\hat{\underline{p}}$ as follows. Because of the assumptions made,

$$D(\underline{w})D(\underline{m}(\hat{\underline{\theta}}_1))D(\underline{1} - \underline{m}(\hat{\underline{\theta}}_1)) \sim E \text{ under } H,$$

so that, from (3.19),

$$X^2 \sim n \sum (m_i(\hat{\underline{\theta}}) - m_i(\hat{\underline{\theta}}_1))^2 / e_i$$

where e_i is the i th diagonal element of E . However, from (3.25) we can deduce that $\sqrt{n}(\underline{m}(\hat{\underline{\theta}}) - \underline{m}(\hat{\underline{\theta}}_1))$ has the asymptotic distribution of a vector $\underline{Y}$ where $\underline{Y} \sim N(\underline{0}, B)$ with $B = (b_{t\ell}) = H_2 C(\hat{\underline{\theta}}_2) H_2'$ and $C(\hat{\underline{\theta}}_2)$ given by (3.28). Thus,

$$\bar{E}(X^2) = \sum_{t=1}^R E(Y_t^2) / e_t$$

and

$$\bar{V}(X^2) = \sum_{t=1}^R V(Y_t^2) / e_t^2 + \sum_{i \neq j}^R \text{cov}(Y_i^2, Y_j^2) / e_i e_j.$$

However, using the properties of the moment generating function of a multivariate normal variable, we know that $E(Y_t^2) = b_{tt}$, $V(Y_t^2) = 2b_{tt}^2$ and $\text{cov}(Y_i^2, Y_j^2) = 2b_{ij}^2$, $i \neq j = 1, 2, \dots, R$, so that

$$(3.57) \quad \Sigma \delta_i = \overline{E(x^2)} = \Sigma b_{tt}/e_t$$

and

$$(3.58) \quad 2\Sigma \delta_i^2 = \overline{V(x^2)} = 2 \sum_{i=1}^R \sum_{j=1}^R b_{ij}^2 / e_i e_j .$$

Thus, with $\hat{\Sigma}/n$ available, an estimate $\hat{B}$ of B may be obtained, from which estimates $\hat{a}$ and $\hat{\delta}$ of a and δ respectively, may be calculated using (3.57) and (3.58). Inserting these estimates into the expression for X_s^2 yields a test statistic to be compared to a χ_v^2 variable.

3.3 Examples of Binary Response Models

In Section 3.2, results were obtained for general binary response models of the form $P_i = m_i(\theta)$. For the development of the Wald and Wald-type statistics, the models were restricted to those for which a monotone differentiable function g of P_i exists such that $g(P_i) = \underline{x}_i' \theta$. In this section, these results are applied to the following four binary response models.

(a) Logit Model

$$P_i = m_i(\theta) = (1 + \exp(-\underline{x}_i' \theta))^{-1}$$

$$g(P_i) = \ln(P_i / (1 - P_i)) = \underline{x}_i' \theta$$

(b) Probit Model

$$P_i = m_i(\theta) = \Phi(\underline{x}_i' \theta)$$

$$g(P_i) = \Phi^{-1}(P_i) = \underline{x}_i' \theta$$

where $\Phi(t)$ denotes the value of the cumulative normal curve at the

point t .

(c) Complementary log-log model (or extreme value model)

$$P_i = m_i(\underline{\theta}) = 1 - \exp(-\exp(\underline{x}_i' \underline{\theta}))$$

$$g(P_i) = \ln(-\ln(1 - P_i)) = \underline{x}_i' \underline{\theta}$$

(d) Box-Cox model

$$\underline{\theta} = (\lambda, \underline{\beta})'$$

$$P_i = m_i(\underline{\theta}) = \begin{cases} (1 + \exp(-\underline{x}_i' \underline{\beta}))^{-1} & \text{if } \lambda = 0 \\ [1 + (1 - \lambda \underline{x}_i' \underline{\beta})^{-(1/\lambda)}]^{-1} & \text{if } \lambda \neq 0. \end{cases}$$

The logit and probit models became popular for describing a wide range of physicochemical and biological processes, and have been used extensively for analysing assays dependent upon a quantal response (see Ashton (1972) and Cox (1970)). These models are also extensively used with survey data, but ignoring the survey design and applying the standard methods for independent binomial sampling. The complementary log-log model arose to explain certain types of survival data (see McCullagh and Nelder (1983)). Finally, the Box-Cox model was proposed as a generalization of the logit model. Guerrero and Johnson (1982) applied it to Mexican household survey data but ignored the survey design. The choice among the four models is often determined by goodness of fit or ease of application.

3.3.1. Models of the form $g(P_i) = \underline{x}_i' \underline{\theta}$

As shown above, the probit, logit and complementary log-log models may be expressed in the form $g_i = g(P_i) = \underline{x}_i' \underline{\theta}$. This property implies

that

$$\partial m_i / \partial \theta_j = (\partial m_i / \partial g_i) (\partial g_i / \partial \theta_j) = (\partial g_i / \partial m_i)^{-1} x_{ij},$$

or, in matrix form,

$$(3.59) \quad M = \frac{\partial m}{\partial \theta} = D X,$$

where D is a diagonal matrix with i th diagonal element equal to $(\partial g_i / \partial p_i)^{-1}$ and X is the $R \times r$ matrix defined by $X = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_R)'$. It then follows that

$$M_1 = \frac{\partial m}{\partial \theta_1} = D X_1,$$

$$M_2 = \frac{\partial m}{\partial \theta_2} = D X_2,$$

and

$$\begin{aligned} (3.60) \quad H_2 &= (I - D X_1 (X_1' D E D X_1)^{-1} X_1' D E) D X_2 \\ &= D (I - X_1 (X_1' \Delta X_1)^{-1} X_1' \Delta) X_2 \\ &= \overline{D X_2}, \end{aligned}$$

where

$$(3.61) \quad \Delta = D E D,$$

$$(3.62) \quad \overline{X_2} = (I - X_1 (X_1' \Delta X_1)^{-1} X_1' \Delta) X_2$$

and $\underline{x}_i' = (\underline{x}_{1i}', \underline{x}_{2i}')$ so that $X_1 = (\underline{x}_{11}', \underline{x}_{12}', \dots, \underline{x}_{1R})'$ and $X_2 = (\underline{x}_{21}', \underline{x}_{22}', \dots, \underline{x}_{2R})'$.

For the logit model,

$$g_i = g(P_i) = \ln(P_i/(1-P_i)) = \underline{x}_i' \underline{\theta} ,$$

so that

$$\partial g_i / \partial P_i = P_i^{-1} (1-P_i)^{-1}$$

and thus

$$(3.63) \quad D = D(\underline{P}) D(1-\underline{P}) = D(\underline{m}(\underline{\theta})) D(1-\underline{m}(\underline{\theta})) . \quad (\text{LOGIT})$$

where $D(\underline{a}) = \text{diag}(a_1, a_2, \dots, a_R)$ for any vector $\underline{a} = (a_1, a_2, \dots, a_R)'$.

Let $\Phi(x)$ represent the cumulative normal curve at the point x and let $\phi(x) = (\sqrt{2\pi})^{-1} \exp(-x^2/2)$ denote the normal frequency at point x . Then the probit model states that

$$m_i(\underline{\theta}) = \Phi(\underline{x}_i' \underline{\theta}) = \int_{-\infty}^{\underline{x}_i' \underline{\theta}} \phi(y) dy .$$

From standard results on improper integrals (eg., Widder (1961)) we deduce that

$$(3.64) \quad \partial m_i(\underline{\theta}) / \partial (\underline{x}_i' \underline{\theta}) = \phi(\underline{x}_i' \underline{\theta}) ,$$

and thus

$$(3.65) \quad \partial m_i(\underline{\theta}) / \partial \theta_j = [\partial m_i(\underline{\theta}) / \partial (\underline{x}_i' \underline{\theta})] [\partial (\underline{x}_i' \underline{\theta}) / \partial \theta_j] = \phi(\underline{x}_i' \underline{\theta}) x_{ij} ,$$

which, in matrix form, is

$$(3.66) \quad \frac{\partial \underline{m}}{\partial \underline{\theta}} = \underline{M} = D(\underline{\phi}) \underline{X} ,$$

where $D(\underline{\phi})$ is the diagonal matrix with i th diagonal element equal to $\phi(\underline{x}_i' \underline{\theta})$. Thus, for the probit model,

$$(3.67) \quad \mathcal{D} = D(\underline{\phi}) \quad (\text{PROBIT})$$

For the complementary log-log model,

$$g_i = g(P_i) = \ln(-\ln(1-P_i)) = \underline{x}_i' \underline{\theta},$$

so that

$$\partial g_i / \partial P_i = -[\ln(1-P_i)]^{-1} (1-P_i)^{-1}$$

and thus

$$(3.68) \quad \mathcal{D} = -D(\ln(1 - \underline{m}(\underline{\theta}))) D(1 - \underline{m}(\underline{\theta})) \quad (\text{LOG-LOG})$$

where $D(\ln(1 - \underline{m}(\underline{\theta})))$ is the diagonal matrix with i th diagonal element equal to $\ln(1 - m_i(\underline{\theta}))$.

For the models for which $g(P_i) = \underline{x}_i' \underline{\theta}$, we obtain, by substitution into (3.13), (3.16) and (3.25) respectively,

$$C(\hat{\underline{\theta}}) = n^{-1} (\underline{X}' \Delta \underline{X})^{-1} (\underline{X}' \Delta \mathcal{D}^{-1} \Sigma \mathcal{D}^{-1} \Delta \underline{X}) (\underline{X}' \Delta \underline{X})^{-1},$$

$$C(\underline{m}(\hat{\underline{\theta}})) = \underline{X}' \mathcal{D} C(\hat{\underline{\theta}}) \mathcal{D} \underline{X},$$

and

$$(3.69) \quad C(\underline{m}(\hat{\underline{\theta}}) - \underline{m}(\hat{\underline{\theta}}_1)) = \mathcal{D} \bar{\underline{X}}_2' C(\hat{\underline{\theta}}_2) \bar{\underline{X}}_2 \mathcal{D}.$$

Theorem 3.3. The asymptotic distribution of the Pearson chi-squared statistic for testing $M_2 : g(P_i) = \underline{x}_{i1}' \underline{\theta}_1$, given that $M_1 : g(P_i) = \underline{x}_i' \underline{\theta}$ is true, is $\sum \delta_i z_i^2$ where the z_i^2 are independent χ_1^2 variables and the δ_i are the eigenvalues of

$$(3.70) \quad (\bar{\underline{X}}_2' \Delta \bar{\underline{X}}_2)^{-1} (\bar{\underline{X}}_2' \Delta \mathcal{D}^{-1} \Sigma \mathcal{D}^{-1} \Delta \bar{\underline{X}}_2).$$

Proof. Substitute (3.60) and (3.61) into (3.29).

Theorem 3.4. The asymptotic distribution of the Pearson chi-squared statistic for testing goodness of fit of the model $M_1 : g(\mathcal{P}_i) = \underline{x}_i' \underline{\theta}$ is $\sum_{i=1}^{R-s} \delta_i Z_i^2$ where the Z_i^2 are independent χ_1^2 variables and the δ_i are the eigenvalues of

$$(3.71) \quad (C' \Delta^{-1} C)^{-1} (C' D^{-1} \Sigma D^{-1} C)$$

with C being any $R \times R-s$ matrix of rank $R-s$ having the property $C'X = \underline{0}$.

Proof. Substitute (3.61) into (3.41).

It should be noted that the asymptotic distribution of the goodness of fit statistic may also be obtained as a particular case of Theorem 3.3, with model M_2 being the model tested and M_1 being a saturated model having M_2 nested within it.

The Wald statistic for testing $H : \underline{\theta}_2 = \underline{0}$, given that $M_1 : \underline{q} = \underline{X} \underline{\theta} = \underline{X}_1 \underline{\theta}_1 + \underline{X}_2 \underline{\theta}_2$ is true, is

$$\chi_W^2 = \hat{\underline{\phi}}' [\hat{C}(\hat{\underline{\phi}})]^{-1} \hat{\underline{\phi}},$$

where $\hat{\underline{\phi}} = C' \hat{\underline{q}} = C' X \hat{\underline{\theta}}$, C being any $R \times u$ matrix of rank u such that $C' X_1 = \underline{0}$ and $C' X_2$ is nonsingular. Now, $\hat{C}(\hat{\underline{\phi}}) = C' X \hat{C}(\hat{\underline{\theta}}) X' C$, where, by (3.13),

$$(3.72) \quad C(\hat{\underline{\theta}}) = n^{-1} (X' \Delta X)^{-1} (X' \Delta D^{-1} \Sigma D^{-1} \Delta X) (X' \Delta X)^{-1}$$

and $\hat{C}(\hat{\theta})$ is obtained by substituting $\hat{\theta}$, $\underline{w}$ and $\hat{\Sigma}$ into (3.72) for $\underline{\theta}$, $\underline{w}$, and Σ respectively. Asymptotically, χ^2_W is distributed as a χ^2_u variable.

The Wald-type statistic χ^2_W of (3.37), which has the same asymptotic distribution as the Pearson chi-squared statistic χ^2 for testing $H : \underline{\theta}_2 = \underline{0}$, is

$$\chi^2_W = \hat{\underline{\phi}}' (\hat{C}_M(\hat{\phi}))^{-1} \hat{\underline{\phi}},$$

where $\hat{\underline{\phi}}$ is the same as in the last paragraph,

$$\begin{aligned} (3.73) \quad C_M(\hat{\phi}) &= C'X C_M(\hat{\theta}) X'C \\ &= n^{-1} C'X(X' \Delta X)^{-1} X'C, \end{aligned}$$

and $\hat{C}_M(\hat{\phi})$ is obtained by replacing $\underline{\theta}$ by $\hat{\theta}$ and $\underline{w}$ by $\underline{w}$ in (3.73).

The adjustments to χ^2 for models of the form $g(P_i) = \underline{x}'_i \underline{\theta}$ follow directly from the results of subsection 3.2.5.

3.3.2. Logit model

Since the logit model is the binary response model most extensively used for survey work, the results of the last subsection will be detailed for it. Recalling that, for the logit model,

$$\mathcal{D} = D(\underline{m}(\underline{\theta})) D(\underline{1} - \underline{m}(\underline{\theta}))$$

and

$$\Delta = D(\underline{w}) D(\underline{m}(\underline{\theta})) D(\underline{1} - \underline{m}(\underline{\theta})),$$

we then have

$$C(\hat{\theta}) = n^{-1} (X' \Delta X)^{-1} (X' D(\underline{W}) \Sigma D(\underline{W}) X) (X' \Delta X)^{-1},$$

$$C(\underline{m}(\hat{\theta})) = X' D(\underline{m}(\hat{\theta})) D(1 - \underline{m}(\hat{\theta})) C(\hat{\theta}) D(\underline{m}(\hat{\theta})) D(1 - \underline{m}(\hat{\theta})) X,$$

and

$$C(\underline{m}(\hat{\theta}) - \underline{m}(\hat{\theta}_1)) = D(\underline{m}(\hat{\theta})) D(1 - \underline{m}(\hat{\theta})) \bar{X}_2' C(\hat{\theta}_2) \bar{X}_2 D(\underline{m}(\hat{\theta})) D(1 - \underline{m}(\hat{\theta})).$$

As well, the asymptotic distribution of the Pearson chi-squared statistic for the nested hypothesis $M_2 : g(P_i) = \underline{x}_{1i-1}' \theta$ given $M_1 : g(P_i) = \underline{x}_i' \theta$ is $\sum_{i=1}^u \delta_i Z_i^2$ where the Z_i^2 are independent χ_1^2 variables and the δ_i are eigenvalues of

$$(\bar{X}_2' \Delta \bar{X}_2)^{-1} (\bar{X}_2' D(\underline{W}) \Sigma D(\underline{W}) \bar{X}_2).$$

Finally, the asymptotic distribution of the Pearson chi-squared statistic for testing goodness of fit of the model $M_1 : g(P_i) = \underline{x}_i' \theta$ is $\sum_{i=1}^{R-s} \delta_i Z_i^2$ where the Z_i^2 are independent χ_1^2 variables and the δ_i are the eigenvalues of

$$(C' \Delta^{-1} C)^{-1} (C' D(\underline{m}(\hat{\theta}))^{-1} D(1 - \underline{m}(\hat{\theta}))^{-1} \Sigma D(\underline{m}(\hat{\theta}))^{-1} D(1 - \underline{m}(\hat{\theta}))^{-1} C)$$

with C being any $R \times R-s$ matrix of rank $R-s$ having the property $C'X = 0$.

3.3.3 Box-Cox model

As noted in the introduction to the section, the Box-Cox model is the only model of the four considered which cannot be put into the

form $g(P_1) = \underline{x}_1' \underline{\theta}$. Because of this, the Wald statistic X_W^2 and the Wald-type statistic $\bar{X}_W^2$ will not be discussed.

Let the vector of parameters of length $(s+1)$ be denoted by $\underline{\theta} = (\lambda, \underline{\beta}')'$. The Box-Cox model to be considered is

$$m_i(\underline{\theta}) = \begin{cases} [1 + (1 + \lambda \underline{x}_i' \underline{\beta})^{-1/\lambda}]^{-1} & \text{if } \lambda \neq 0 \\ (1 + \exp(-\underline{x}_i' \underline{\beta}))^{-1} & \text{if } \lambda = 0 \text{ (logit model).} \end{cases}$$

Now, $m_i(\underline{\theta})$ is a continuous function for all values of $\underline{\theta}$, since $m_i(\underline{\theta})$ is continuous for all points where $\lambda \neq 0$, and, for fixed $\underline{\beta}$,

$$(3.74) \quad \lim_{\lambda \rightarrow 0} m_i(\lambda, \underline{\beta}) = m_i(0, \underline{\beta})$$

due to the fact that

$$(3.75) \quad \lim_{\lambda \rightarrow 0} (1 + \lambda \underline{x}_i' \underline{\beta})^{-1/\lambda} = \exp(-\underline{x}_i' \underline{\beta}).$$

Lemma 3.1. If w and y are functions of x , then

$$(3.76) \quad \partial w / \partial x = -w^2 (\partial w^{-1} / \partial x)$$

and

$$(3.77) \quad \partial (w^y) / \partial x = y w^{y-1} (\partial w / \partial x) + w^y \ln w (\partial y / \partial x).$$

Proof. Both results are straightforward.

We need the derivatives $\frac{\partial m}{\partial \theta}$ to determine the eigenvalues δ_i or the correction δ . To derive $\frac{\partial m}{\partial \theta}$ we must examine the following four cases.

22
OF / DE

$$(i) \quad \underline{\partial m_i(\theta) / \partial \lambda} \quad (\lambda \neq 0)$$

$$\begin{aligned} \partial m_i^{-1}(\theta) / \partial \lambda &= \frac{\partial}{\partial \lambda} [1 + \lambda \underline{x_i}' \beta]^{-1/\lambda} \\ &= -(1/\lambda) (1 + \lambda \underline{x_i}' \beta)^{-1/\lambda - 1} \underline{x_i}' \beta \\ &\quad + (1 + \lambda \underline{x_i}' \beta)^{-1/\lambda} (\ln(1 + \lambda \underline{x_i}' \beta)) (1/\lambda^2) \quad \text{by (3.77)} \\ &= (1/\lambda) (1 + \lambda \underline{x_i}' \beta)^{-1/\lambda} [(1/\lambda) \ln(1 + \lambda \underline{x_i}' \beta) - (1 + \lambda \underline{x_i}' \beta)^{-1} \underline{x_i}' \beta] . \end{aligned}$$

Thus, by (3.76)

$$(3.78) \quad \partial m_i(\theta) / \partial \lambda = m_i(\theta) (1 - m_i(\theta)) (1/\lambda^2) (1 - z_i + \ln z_i) , \quad (\lambda \neq 0)$$

$$\text{where } z_i = (1 + \lambda \underline{x_i}' \beta)^{-1}$$

$$(ii) \quad \underline{\partial m_i(\theta) / \partial \beta_j} \quad (\lambda \neq 0)$$

$$\begin{aligned} \partial m_i^{-1}(\theta) / \partial \beta_j &= -(1/\lambda) (1 + \lambda \underline{x_i}' \beta)^{-1/\lambda - 1} \lambda x_{ij} \\ &= -(m_i^{-1}(\theta) - 1) (1 + \lambda \underline{x_i}' \beta)^{-1} x_{ij} . \end{aligned}$$

Therefore,

$$(3.79) \quad \partial m_i(\theta) / \partial \beta_j = m_i(\theta) (1 - m_i(\theta)) z_i x_{ij} \quad (\lambda \neq 0)$$

$$(iii) \quad \underline{m_i(\theta) / \partial \beta_j} \quad (\lambda = 0)$$

$$\begin{aligned} (3.80) \quad \partial m_i(0, \beta) / \partial \beta_j &= x_{ij} (1 + \exp(-\underline{x_i}' \beta))^{-2} \exp(-\underline{x_i}' \beta) \\ &= m_i(0, \beta) (1 - m_i(0, \beta)) x_{ij} . \end{aligned}$$

$$(iv) \quad \underline{\partial m_i(\theta) / \partial \lambda} \quad (\lambda = 0) .$$

By the definition of a derivative and the use of l'Hospital's rule,

$$(3.81) \quad \frac{\partial m_i(0, \underline{\beta})}{\partial \lambda} = \lim_{\lambda \rightarrow 0} [m_i(\lambda, \underline{\beta}) - m_i(0, \underline{\beta})] / \lambda \\ = -m_i(0, \underline{\beta}) (1 - m_i(0, \underline{\beta})) (\underline{x}_i' \underline{\beta})^2 / 2 .$$

Combining the results above,

$$(3.82) \quad \frac{\partial \underline{m}}{\partial \theta} = \begin{cases} D(\underline{m}(\theta)) D(1 - \underline{m}(\theta)) D(\underline{z}) [\underline{u}, \underline{x}] & \text{if } \lambda \neq 0 \\ D(\underline{m}(0, \underline{\beta})) D(1 - \underline{m}(0, \underline{\beta})) [\underline{w}, \underline{x}] & \text{if } \lambda = 0 , \end{cases}$$

where $\underline{w} = (w_1, w_2, \dots, w_R)'$, $w_i = -(\underline{x}_i' \underline{\theta})^2 / 2$,

$\underline{u} = (u_1, u_2, \dots, u_R)'$ and $u_i = (z_i \lambda^2)^{-1} (1 - z_i + \ln z_i)$.

Now, $\lim_{\lambda \rightarrow 0} \underline{u} = \underline{w}$ and $\lim_{\lambda \rightarrow 0} D(\underline{z}) = I$, so that

$$(3.83) \quad \left(\frac{\partial \underline{m}}{\partial \theta} \right)_{\lambda=0} = \lim_{\lambda \rightarrow 0} \frac{\partial \underline{m}}{\partial \theta} .$$

We now derive the asymptotic distribution of X^2 for the Box-Cox model, utilizing the derivatives (3.82).

Theorem 3.5. Let $\underline{\beta} = (\underline{\beta}_1', \underline{\beta}_2')'$ where $\underline{\beta}_1$ is $s \times 1$ and $\underline{\beta}_2$ is $u \times 1$ ($s+u = r$). The asymptotic distribution of the Pearson chi-squared statistic for testing $H: \underline{\beta}_2 = \underline{0}$, given that the Box-Cox model applies, is $\sum_{i=1}^u \delta_i z_i^2$, where the z_i^2 are independent χ_1^2 variables and the δ_i are the eigenvalues of

$$(3.84) \quad (\bar{B}_2' \nabla \bar{B}_2)^{-1} (\bar{B}_2' D(\underline{W}) \Sigma D(\underline{W}) \bar{B}_2) ,$$

with

$$B = D(\underline{z})[\underline{u}, X] , \quad B_1 = D(\underline{z})[\underline{u}, X_1] , \quad B_2 = D(\underline{z})X_2 ,$$

$$\nabla = D(\underline{m}(\underline{\theta}))D(\underline{1} - \underline{m}(\underline{\theta}))D(\underline{W}) ,$$

and

$$\bar{B}_2 = [I - B_1(B_1' \nabla B_1)^{-1} B_1' \nabla] B_2 .$$

Proof. By (3.82) we see that

$$(3.85) \quad \frac{\partial \underline{m}}{\partial \underline{\theta}} = D(\underline{m}(\underline{\theta}))D(\underline{1} - \underline{m}(\underline{\theta}))B ,$$

$$(3.86) \quad \frac{\partial \underline{m}}{\partial \theta_1} = D(\underline{m}(\underline{\theta}))D(\underline{1} - \underline{m}(\underline{\theta}))B_1 ,$$

and

$$(3.87) \quad \frac{\partial \underline{m}}{\partial \theta_2} = D(\underline{m}(\underline{\theta}))D(\underline{1} - \underline{m}(\underline{\theta}))B_2 .$$

When these expressions are substituted into (3.29), the required result follows.

Corollary. The asymptotic distribution of the Pearson chi-squared statistic for the goodness of fit of the Box-Cox model is $\sum_{i=1}^{R-s-1} \delta_i Z_i^2$ where Z_i^2 are independent χ_1^2 variables and the δ_i are the eigenvalues of

$$(3.88) \quad (\bar{B}_{e2}' \nabla \bar{B}_{e2})^{-1} (\bar{B}_{e2}' D(\underline{W}) \sum D(\underline{W}) \bar{B}_{e2})$$

with

$$B_{e2} = D(\underline{z})X_e ,$$

and

$$\bar{B}_{e2} = [I - B(B' \nabla B)^{-1} B' \nabla] B_{e2} ,$$

where X_e is any $R \times (R-s-1)$ matrix such that $[\underline{u}, X, X_e]$ is non-singular.

Proof. Testing of goodness of fit of the Box-Cox model may be treated as a special case of Theorem 3.5, where the Box-Cox model is the model to be tested and the given model is a saturated model of the Box-Cox form. To create a saturated model, the vector $\underline{\beta}$ must be extended to $\bar{\underline{\beta}} = (\underline{\beta}, \underline{\beta}_e)$ where $\underline{\beta}_e$ is of length $R-s-1$ and the matrix X must be extended to (X, X_e) where X_e is $R \times (R-s-1)$ so that $(\underline{u}, X, X_e)$ is nonsingular.

3.4. Logit Models and Direct Estimates

When a loglinear model has direct estimates (i.e. explicit mle's), Rao and Scott (1984) showed that the value of $\hat{\delta}^*$ used in the first order correction of the Pearson chi-squared statistic, $X^2/\hat{\delta}^*$, may be expressed just in terms of the deffs of individual cells and of marginal totals corresponding to "sufficient configurations". This result is of significance since such deffs could often be provided along with published tables.

In Subsection 3.2.5, it was stated that, for binary response models, direct estimates do not generally exist so that in calculating first order corrections of the Pearson chi-squared statistic of the form $X^2/\hat{\delta}^*$, the best that can be done, when an estimated covariance matrix is unavailable, is to calculate an upper bound for $\hat{\delta}^*$. However, for the logit model, the most popular binary response model in survey

work, there exist a limited number of situations, as described below, in which loglinear direct estimate results can improve on this correction.

It is a well-known fact that for every hierarchical logit model there is a corresponding loglinear model yielding the same estimated expected cell frequencies. If a hypothesized logit model should correspond to a loglinear model with direct estimates, then the hypothesis and the corresponding Pearson chi-squared statistic could be expressed in terms of the loglinear model for which the direct estimate results of Rao and Scott could then be utilized. The following theorem delineates the subset of hierarchical logit models corresponding to loglinear models with direct estimates.

Following the notation of BFH (1975), Chapter 3, if a model can be expressed by configurations $C_{\theta_1}, C_{\theta_2}, \dots, C_{\theta_t}$, where each θ_i represents a subset of variables, then the model contains terms $u_{\theta_1}, u_{\theta_2}, \dots, u_{\theta_t}$ and all their lower order relatives.

Theorem 3.6. A hierarchical logit model has corresponding loglinear model with direct estimates iff the logit model is expressible as a single configuration.

Proof. Consider a logit model arising from an ℓ -dimensional contingency table, where the ℓ th variable is considered to be the binary response variable and the remaining are considered factors. Let the variables be denoted by $1, 2, \dots, \ell$.

Sufficiency. Suppose that the logit model may be expressed by the configuration $C_{\underline{a}}$ where $\underline{a} \subset \{1, 2, \dots, l-1\}$. Then, it is a well-known fact (see, for example, Kalbfleisch (1982)) that the loglinear model with the same mle's of cell proportions has minimal sufficient configurations $C_{\underline{a}, l}$ and $C_{1, 2, \dots, l-1}$. However, according to Theorem 3.4-1 of BFH (1975), if there are only two minimal sufficient configurations, direct estimates exist.

Necessity. Suppose that the logit model needs more than one configuration to be expressed. For simplicity, assume that it has two configurations, say $C_{\underline{a}}$ and $C_{\underline{b}}$. Since these are minimal configurations, then $\underline{a} \subset \{1, 2, \dots, l-1\}$ and $\underline{b} \subset \{1, 2, \dots, l-1\}$ but $\underline{a} \not\subset \underline{b}$ and $\underline{b} \not\subset \underline{a}$. Let $\underline{a} = \underline{d} \cup \underline{e}$ and $\underline{b} = \underline{d} \cup \underline{f}$ where $\underline{d}$, $\underline{e}$ and $\underline{f}$ are mutually exclusive and $\underline{d}$ could be the empty set. Also, let $\underline{c} = \{1, 2, \dots, l-1\} \setminus (\underline{a} \cup \underline{b})$ so that $\underline{d}$, $\underline{e}$, $\underline{f}$ and $\underline{c}$ are disjoint, with their union being $\{1, 2, \dots, l-1\}$. Now, the loglinear model with the same mle's as the logit model may be defined by minimal sufficient configurations $C_{\underline{a}, l}$, $C_{\underline{b}, l}$ and $C_{1, 2, \dots, l-1}$. These configurations may be expressed equivalently as $C_{\underline{d} \cup \underline{e}, l}$, $C_{\underline{d} \cup \underline{f}, l}$ and $C_{\underline{d} \cup \underline{e} \cup \underline{f} \cup \underline{c}}$. Applying the classifying rules of BFH (1975), page 76, we do the following.

(a) Delete $\underline{d}$ which is in all configurations, leaving

$$C_{\underline{e}, l}, C_{\underline{f}, l} \text{ and } C_{\underline{e} \cup \underline{f} \cup \underline{c}}.$$

(b) Delete $\underline{c}$ since it appears in only 1 configuration, leaving

$$C_{\underline{e}, l}, C_{\underline{f}, l} \text{ and } C_{\underline{e} \cup \underline{f}}.$$

(c) There are no redundant configurations, so that no further steps can be taken: We are left with three configurations, and thus no direct estimates exist.

As a result, if the logit model is described by more than one configuration, the corresponding loglinear model does not have direct estimates. Q.E.D.

Tables 3.1 and 3.2 contain examples of all hierarchical logit models in three and four dimensions respectively, together with their corresponding loglinear models. The cases in which the loglinear models have direct estimates are indicated.

In view of Theorem 3.6, making use of results about loglinear models with direct estimates is unrealistic in the context of logit models since they do not even cover such realistic logit models as those involving two or more factors acting additively without interaction.

However, when the loglinear model corresponding to a logit model has direct estimates, the form of these direct estimates and an expression for δ may be obtained, as follows. Let θ denote the subscript for a general cell and Q denote a general subscript for the first $(l-1)$ variables, so that $\theta \equiv (Q, l)$. Suppose that the configuration describing the logit model is $C_{\underline{d}}$, where $\underline{d} \subset \{1, 2, \dots, l-1\}$. Then the minimal sufficient configurations for the corresponding loglinear model are $C_{\underline{d}, l}$ and $C_{1, 2, \dots, l-1}$, where $C_{1, 2, \dots, l-1}$ can also be denoted as $C_{\underline{d} \cup \underline{g}}$ with $\underline{d} \cup \underline{g} = \{1, 2, \dots, l-1\}$ and $\underline{d} \cap \underline{g} = \emptyset$. Since $\{\underline{d}, l\} \cap \{\underline{d} \cup \underline{g}\} = \underline{d}$, the loglinear model may be expressed as

$$(3.89) \quad \pi_{\theta} = \frac{\pi_{\underline{d},l} \pi_{1,2,\dots,l-1}}{\pi_{\underline{d}}}$$

or, equivalently, as

$$(3.90) \quad \pi_{Q,l} = \frac{\pi_{\underline{d},l} \pi_Q}{\pi_{\underline{d}}} \quad l = 1, 2$$

where, in both cases, any subscript not appearing is assumed to be summed.

The proportion P_Q of successes in the Q^{th} domain is defined as

$$(3.91) \quad P_Q = \frac{\pi_{Q,1}}{\pi_{Q,1} + \pi_{Q,2}} = \frac{\pi_{Q,1}}{w_Q},$$

Thus, $\pi_{Q,1} = w_Q P_Q$ and $\pi_{Q,2} = w_Q (1 - P_Q)$. It follows that

$$\pi_{\underline{d},l=1} = \sum_{Q \in \underline{d}} w_Q P_Q = \sum_{\underline{g}} w_{\underline{g}} P_{\underline{g}}$$

$$\pi_{\underline{d},l=2} = \sum_{\underline{g}} w_{\underline{g}} (1 - P_{\underline{g}}),$$

and

$$\pi_{\underline{d}} = \pi_{\underline{d},l=1} + \pi_{\underline{d},l=2} = \sum_{\underline{g}} w_{\underline{g}}.$$

The loglinear model (3.90), expressed in terms of the P_Q , is then

$$(3.92) \quad P_Q = (\sum_{\underline{g}} w_{\underline{g}} P_{\underline{g}}) / (\sum_{\underline{g}} w_{\underline{g}}) = f(\underline{d}).$$

Let $\hat{P}_Q$ and w_Q be survey estimates of P_Q and w_Q respectively. Then an estimate $\hat{m}_Q$ of P_Q under the hypothesized model is

$$(3.93) \quad \hat{m}_Q = (\sum_{\underline{g}} w_{\underline{g}} \hat{P}_{\underline{g}}) / (\sum_{\underline{g}} w_{\underline{g}}) = \hat{f}(\underline{d}).$$

If the survey and model estimates of the cell proportions are defined in terms of the logit proportions then

$$\hat{p}_{Q,1} = w_Q \hat{p}_Q, \hat{p}_{Q,2} = w_Q (1 - \hat{p}_Q), \hat{\pi}_{Q,1} = w_Q \hat{m}_Q \text{ and } \hat{\pi}_{Q,2} = w_Q (1 - \hat{m}_Q).$$

The Pearson chi-squared statistic for the loglinear model is

$$\begin{aligned} (3.94) \quad x^2 &= n \sum_{\theta} \frac{(\hat{p}_{\theta} - \hat{\pi}_{\theta})^2}{\hat{\pi}_{\theta}} = n \sum_Q \sum_{\ell} \frac{(\hat{p}_{Q,\ell} - \hat{\pi}_{Q,\ell})^2}{\hat{\pi}_{Q,\ell}} \\ &= n \sum_Q w_Q \frac{(\hat{p}_Q - \hat{m}_Q)^2}{\hat{m}_Q (1 - \hat{m}_Q)} \\ &= n \sum_Q w_Q \frac{(\hat{p}_Q - \hat{m}_Q)^2}{\hat{m}_Q (1 - \hat{m}_Q)} \\ &= n \sum_Q \frac{w_Q (\hat{p}_Q - P_Q)^2}{\hat{m}_Q (1 - \hat{m}_Q)} + n \sum_Q \frac{w_Q (\hat{m}_Q - P_Q)^2}{\hat{m}_Q (1 - \hat{m}_Q)} - 2n \sum_Q \frac{w_Q (\hat{p}_Q - P_Q)(\hat{m}_Q - P_Q)}{\hat{m}_Q (1 - \hat{m}_Q)} \end{aligned}$$

Since, under the hypothesized model, (3.92) and (3.93) are true, it follows that (3.94) may be written as

$$\begin{aligned} (3.95) \quad x^2 &\sim n \sum_Q \frac{w_Q (\hat{p}_Q - P_Q)^2}{\hat{P}_Q (1 - \hat{P}_Q)} + n \sum_{\underline{d}} \sum_{\underline{g}} w_Q \frac{(\hat{f}(\underline{d}) - f(\underline{d}))^2}{f(\underline{d}) (1 - f(\underline{d}))} - 2n \sum_{\underline{d}} \frac{(\hat{f}(\underline{d}) - f(\underline{d}))}{f(\underline{d}) (1 - f(\underline{d}))} \\ &\quad \cdot \sum_{\underline{g}} w_Q (\hat{P}_Q - P_Q) \\ &\sim n \sum_Q \frac{w_Q (\hat{P}_Q - P_Q)^2}{\hat{P}_Q (1 - \hat{P}_Q)} - n \sum_{\underline{d}} \sum_{\underline{g}} w_Q \frac{(\hat{f}(\underline{d}) - f(\underline{d}))^2}{f(\underline{d}) (1 - f(\underline{d}))}. \end{aligned}$$

Therefore,

$$(3.96) \quad u\delta = \overline{EX}^2 = n \sum_Q \frac{w_Q v(\hat{P}_Q)}{\hat{P}_Q (1 - \hat{P}_Q)} - n \sum_{\underline{d}} \frac{(\sum_{\underline{g}} w_Q) v(\hat{f}(\underline{d}))}{\hat{f}(\underline{d}) (1 - \hat{f}(\underline{d}))}$$

$$= \sum_Q d(\hat{p}_Q) - \sum_{\underline{d}} d(\hat{f}(\underline{d})) ,$$

where

$$d(\hat{p}_Q) = \frac{n w_Q v(\hat{p}_Q)}{\hat{p}_Q (1 - \hat{p}_Q)} \quad \text{and} \quad d(\hat{f}(\underline{d})) = \frac{(\sum w_Q) v(\hat{f}(\underline{d}))}{\hat{f}(\underline{d}) (1 - \hat{f}(\underline{d}))}$$

Example

Consider the hypothesized logit model, in four dimensions,

$$\ln(p_{ijk}/(1-p_{ijk})) = v + v_1(i) .$$

The corresponding loglinear model is

$$\pi_{ijkl} = (\pi_{ijk+} \pi_{i++l}) / \pi_{i+++} ,$$

which, in terms of the logit proportions, is

$$p_{ijk} = (\sum_{jk} w_{ijk} p_{ijk}) / \sum_{jk} w_{ijk} = f(i) .$$

We then have

$$\hat{m}_{ijk} = (\sum_{jk} w_{ijk} \hat{p}_{ijk}) / \sum_{jk} w_{ijk} = \hat{f}(i)$$

and

$$u\delta = \overline{Ex}^2 = \sum_{ijk} d(\hat{p}_{ijk}) - \sum_i d(\hat{f}(i)) ,$$

where

$$d(\hat{p}_{ijk}) = \frac{n w_{ijk} v(\hat{p}_{ijk})}{\hat{p}_{ijk} (1 - \hat{p}_{ijk})} \quad \text{and} \quad d(\hat{f}(i)) = \frac{(\sum w_Q) v(\hat{f}(i))}{\hat{f}(i) (1 - \hat{f}(i))} .$$

Table 3.1

Three-dimensional Hierarchical Logit Models with Corresponding Loglinear Models

Let π_{ijk} denote the proportion in the (i,j,k) th cell, $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; and $k = 1, 2$. The third variable is the binary response variable and $L_{ij} = \ln(\pi_{ij1}/\pi_{ij2})$.

Logit Model	# of Versions	Configuration form of Logit Model	Loglinear Model with same mle's	Existence of Direct Estimates
$L_{ij} = w$	1	constant	$C_3 C_{12}$	Yes
$L_{ij} = w + w_{1(i)}$	2	C_1	$C_{13} C_{12}$	Yes
$L_{ij} = w + w_{1(i)} + w_{2(j)}$	1	$C_1 C_2$	$C_{13} C_{23} C_{12}$	No
$L_{ij} = w + w_{1(i)} + w_{2(j)} + w_{12(ij)}$	1	C_{12}	C_{123}	$\hat{\pi}_{ijk} = \hat{p}_{ijk}$ (saturated model)

Table 3.2

Four-dimensional Hierarchical Logit Models with Corresponding Loglinear Models

Let $\pi_{ijk\ell}$ denote the proportion in the (i,j,k,ℓ) th cell, $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$; $k = 1, 2, \dots, K$; $\ell = 1, 2$. The fourth variable is the binary response variable and $L_{ijk} = \ln(\pi_{ijk1}/\pi_{ijk2})$.

LOGIT MODEL				LOGLINEAR MODEL WITH SAME MLE'S				Existence of direct estimates	
Example	# of effects removed 3f*	2f	1f	# of versions	Example	# of effects removed 4f	3f		2f
Constant	1	3	3	1	$C_4 C_{123}$	1	3	3	Yes
C_1	1	3	2	3	$C_{14} C_{123}$	1	3	2	Yes
$C_1 C_2$	1	3	1	3	$C_{14} C_{24} C_{123}$	1	3	1	No
$C_1 C_2 C_3$	1	3	0	1	$C_{14} C_{24} C_{34} C_{123}$	1	3	0	No
C_{12}	1	2	1	3	$C_{124} C_{123}$	1	2	1	Yes
$C_{12} C_{13}$	1	1	0	3	$C_{124} C_{134} C_{123}$	1	1	0	No
$C_{12} C_{13} C_{23}$	1	0	0	1	$C_{124} C_{134} C_{234} C_{123}$	1	0	0	No
C_{123}	0	0	0	1	C_{1234}	0	0	0	$\hat{\pi}_{ijk\ell} = \hat{p}_{ijk\ell}$

* f denotes factor

CHAPTER 4

A: DIAGNOSTIC PROCEDURES FOR BINARY RESPONSE MODELS

B: LOGIT MODEL ANALYSIS OF LABOUR FORCE SURVEY DATA

4.1. Introduction

As noted in Chapter 3, binary response models have been fitted to a wide variety of data. However, as is the case in ordinary least squares regression, the fitted model is likely to be sensitive to outlying responses and extreme values in the explanatory variables. Pregibon (1981) and Cook and Weisberg (1982), among others, have examined the logit model under product binomial sampling and have suggested methods of identifying outliers and influential points and of assessing their impact on various aspects of the fit of the logit model. In Section 4.2 a selection of these methods are extended to general sampling plans and to those binary response models expressible in the form $g(P_i) = \underline{x}_i' \underline{\theta}$. As noted in Chapter 3, this class includes the commonly used logit, probit, and complementary log-log models.

In Section 4.3, the results of the analysis of a set of data from the October 1980 Canadian Labour Force Survey (LFS) are presented. Logit models were fitted to the data and test statistics were calculated, using the techniques described in Chapter 3. As well, the diagnostic techniques outlined in Section 4.2 were utilized. This work was carried out jointly with S. Kumar and J.N.K. Rao.

4.2. Diagnostic Methods for Binary Response Models

4.2.1 Background and Notation

As in Chapter 3, let P_i denote the success rate in domain i , $i = 1, 2, \dots, R$. Suppose that the binary response model under consideration is $P_i = m_i(\underline{\theta})$, such that there exists a monotone differentiable function g with the property

$$(4.1) \quad g_i = g(P_i) = g(m_i(\underline{\theta})) = \underline{x}_i' \underline{\theta} \quad i=1, 2, \dots, R,$$

where $\underline{\theta}$ is an r -vector of unknown parameters and $\underline{x}_i'$ is a known vector of explanatory variables. Let $\hat{\underline{p}} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_R)'$ be the survey estimate of $\underline{p} = (p_1, p_2, \dots, p_R)'$ and let X denote the $R \times r$ matrix $X = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_R)'$.

By (3.4), the pseudo likelihood equations which are solved to obtain the pseudo mle's $\hat{\underline{\theta}}$ and $\hat{\underline{m}} = \underline{m}(\hat{\underline{\theta}})$ of $\underline{\theta}$ and $\underline{m}(\underline{\theta}) = (m_1(\underline{\theta}), m_2(\underline{\theta}), \dots, m_R(\underline{\theta}))'$ respectively, are

$$(4.2) \quad \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\hat{\underline{\theta}}} D(\hat{\underline{m}})^{-1} D(\underline{1}-\hat{\underline{m}})^{-1} D(\underline{w}) \hat{\underline{p}} = \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)'_{\underline{\theta}=\hat{\underline{\theta}}} D(\underline{1}-\hat{\underline{m}})^{-1} \underline{w},$$

where $\underline{w} = (w_1, \dots, w_R)'$ and w_i is a survey estimate of the proportion of the population in the i th domain. However, when the binary response models are restricted to those of form (4.1), $\frac{\partial \underline{m}}{\partial \underline{\theta}}$ has the special form:

$$(4.3) \quad \frac{\partial \underline{m}}{\partial \underline{\theta}} = D X,$$

where D is a diagonal matrix with i th diagonal element d_{ii} equal to $(\partial g_i / \partial m_i)^{-1}$, as shown in Section 3.3. Substituting (4.3) into (4.2) and rearranging terms, the pseudo mle's may be expressed as

$$(4.4) \quad X' \hat{\underline{Y}} = \underline{0},$$

where

$$\hat{\underline{y}} = \hat{\underline{D}}(\hat{\underline{m}})^{-1} \underline{D}(1-\hat{\underline{m}})^{-1} \underline{D}(\underline{w}) (\hat{\underline{P}} - \hat{\underline{m}})$$

and $\hat{\underline{D}}$ is obtained from $\underline{D}$ by substituting $\hat{\underline{m}}$ for $\underline{m}$. Now, the k th equation in (4.4) is

$$(4.5) \quad \sum_{i=1}^R x_{ik} \hat{y}_i = \sum_{i=1}^R x_{ik} \hat{d}_{ii} w_i (\hat{p}_i - \hat{m}_i) / [\hat{m}_i (1 - \hat{m}_i)] = 0$$

and, by straightforward computation,

$$(4.6) \quad \partial \left(\sum_{i=1}^R x_{ik} \hat{y}_i \right) / \partial \theta_j = - \sum_{i=1}^R x_{ik} \hat{b}_i x_{ij},$$

where

$$(4.7) \quad \hat{b}_i = \hat{d}_{ii}^2 w_i \hat{m}_i^{-1} (1 - \hat{m}_i)^{-1} \left\{ 1 + (\hat{p}_i - \hat{m}_i) \left[\hat{d}_{ii} \frac{\partial (\hat{d}_{ii}^{-1})}{\partial \hat{m}_i} + \frac{(1 - 2\hat{m}_i)}{\hat{m}_i (1 - \hat{m}_i)} \right] \right\}.$$

Thus

$$(4.8) \quad -\partial(X' \hat{\underline{y}}) / \partial \underline{\theta} = X' \hat{\underline{B}} X,$$

with $\hat{\underline{B}}$ being a diagonal matrix having $\hat{b}_i$ on the i th diagonal. It should be noted that $\hat{b}_i$ contains the factor $w_i / [\hat{m}_i (1 - \hat{m}_i)]$ which, apart from the multiplier n , is the inverse of the estimate of the asymptotic variance of $\hat{p}_i$ under product binomial sampling and model (4.1). Also to be noted is that, for the logit model defined by

$$(4.9) \quad g(m_i) = \ln(m_i / (1 - m_i)) = \underline{x}_i' \underline{\theta},$$

the value of $\hat{d}_{ii}$ is $\hat{m}_i (1 - \hat{m}_i)$ so that

$$(4.10) \quad \hat{b}_{i, \text{logit}} = w_i \hat{m}_i (1 - \hat{m}_i).$$

Since equations (4.4) are nonlinear in $\hat{\theta}$, iterative methods are required for their solution. A common iterative procedure, Newton-Raphson, leads to the iterative scheme

$$(4.11) \quad \underline{\theta}^{t+1} = \underline{\theta}^t - \left[\frac{\partial (X'Y^t)}{\partial \underline{\theta}^t} \right]^{-1} X'Y^t$$

$$= \underline{\theta}^t + (X' B^t X)^{-1} X' Y^t,$$

where $\underline{\theta}^t$ is the value of $\underline{\theta}$ after t iterations and where B^t and Y^t are obtained from $\hat{B}$ and $\hat{Y}$ by replacing $\hat{\theta}$ with $\underline{\theta}^t$. Instead of doing the iterations in the form of (4.11), the iterative process is often carried out by the method of iteratively reweighted least squares (IRLS) as follows. By defining the pseudo-observation vector

$$(4.12) \quad \underline{z}^t = X \underline{\theta}^t + (B^t)^{-1} Y^t,$$

then (4.11) may be written as

$$(4.13) \quad \underline{\theta}^{t+1} = (X' B^t X)^{-1} X' B^t \underline{z}^t,$$

i.e. in the form of a weighted least squares estimator. Starting with a trial value $\underline{\theta}^0$, the sequence $\{\underline{\theta}^t\}$ is obtained iteratively from (4.13), so that at convergence,

$$(4.14) \quad \underline{z} = X \hat{\theta} + (\hat{B})^{-1} \hat{Y},$$

and the solution of (4.4) is

$$(4.15) \quad \hat{\theta} = (X' \hat{B} X)^{-1} X' \hat{B} \underline{z}.$$

The vector of pseudo-residuals is

$$\begin{aligned}
 (4.16) \quad \underline{z} - \underline{X}\hat{\underline{\theta}} &= (\underline{I} - \underline{X}(\underline{X}'\hat{\underline{B}}\underline{X})^{-1}\underline{X}'\hat{\underline{B}})\underline{z} \\
 &= \hat{\underline{B}}^{-1/2}\bar{\underline{M}}\hat{\underline{B}}^{1/2}\underline{z},
 \end{aligned}$$

where

$$(4.17) \quad \bar{\underline{M}} = \underline{I} - \underline{H} = \underline{I} - \hat{\underline{B}}^{1/2}\underline{X}(\underline{X}'\hat{\underline{B}}\underline{X})^{-1}\underline{X}'\hat{\underline{B}}^{1/2}.$$

Substituting $\underline{X}\hat{\underline{\theta}} + \hat{\underline{B}}^{-1/2}\underline{\hat{Y}}$ for $\underline{z}$ on both sides of (4.16), produces

$$(4.18) \quad \hat{\underline{B}}^{-1/2}\underline{\hat{Y}} = \hat{\underline{B}}^{-1/2}\bar{\underline{M}}\hat{\underline{B}}^{1/2}\underline{\hat{Y}},$$

and premultiplication by $\hat{\underline{B}}^{1/2}$ yields

$$(4.19) \quad \underline{r} = \bar{\underline{M}}\underline{r},$$

where $\underline{r} = \hat{\underline{B}}^{1/2}\underline{\hat{Y}}$. In the case of the logit model under product binomial sampling, the i^{th} element of $\underline{r}$, say r_i , is

$$(4.20) \quad r_i = (\hat{p}_i - \hat{m}_i) / [w_i \hat{m}_i (1 - \hat{m}_i)]^{1/2},$$

which could be called a standardized residual, so that, by (4.19), $\bar{\underline{M}}$ spans the standardized residual space.

4.2.2 Diagnostic Methods

Diagnostic methods may be considered to fall into two categories, namely those which call attention to outlying and influential points and those which assess the impact of points on various aspects of fit of the model. A selection of methods falling into each category will now be described.

(a) Diagnostics for detecting outlying and influential points

The vector of residuals is often a useful tool for detecting

outlying cell proportions. Unlike the standard linear model, residuals for a binary response model can be defined on many scales. One choice that takes account of the survey design is the vector of standardized residuals

$$e_i = (\hat{p}_i - \hat{m}_i) / [\hat{v}_{ii}(\hat{p}_i - \hat{m}_i)]^{1/2},$$

where $\hat{v}_{ii}(\hat{p}_i - \hat{m}_i)$ is estimated from the i^{th} diagonal element of $C(\hat{p} - \hat{m})$ in (3.18), by replacing θ and Σ by $\hat{\theta}$ and $\hat{\Sigma}$. Since the e_i 's are approximately $N(0,1)$ under model (4.1), the expected numbers of residuals e_i exceeding 1.96, 2.33, and 2.58 in magnitude are 0.05R, 0.02R and 0.01R respectively (where R is the number of residuals). These expected numbers provide a rough guide for identifying any outlying cells.

In the binomial case, Pregibon (1981) proposed another set of residuals based on the components of the Pearson chi-squared statistic, X^2 , and the likelihood ratio statistic, G^2 . Similar residuals that take the sample design into account are the components of the corrected statistics $X^2/\hat{\delta}_i$ and $G^2/\hat{\delta}_i$. Recalling the forms of X^2 and G^2 from previous chapters, these components have the forms

$$(4.21) \quad \tilde{X}_i = X_i / (\hat{\delta}_i)^{1/2} = (nw_i)^{1/2} (\hat{p}_i - \hat{m}_i) / [\hat{\delta}_i \hat{m}_i (1 - \hat{m}_i)]^{1/2}$$

and

$$(4.22) \quad \tilde{G}_i = (2nw_i \{ \hat{p}_i \ln(\hat{p}_i / \hat{m}_i) + (1 - \hat{p}_i) \ln((1 - \hat{p}_i) / (1 - \hat{m}_i)) \})^{1/2}.$$

Large individual components should indicate cells poorly explained by

the model. Two useful ways of displaying these components are as index plots (i.e. plots of X_i and G_i versus i) and as normal probability plots (i.e. the ordered values plotted against standard normal quantiles).

Pregibon (1981) also suggested the use of the diagonal elements m_{ii} of the projection matrix $\bar{M}$ (defined in (4.17)) for detecting influential points. Small values of m_{ii} are useful in detecting extreme points in the factor space. It should be noted that $\bar{M}$ arises naturally if the pseudo-likelihood equations are solved by iterative reweighted least squares. An index plot of m_{ii} versus i often provides a useful display. As well, parallel to a suggestion made by Pregibon (1981), another useful graph is a scatter plot of $\tilde{X}_i^2/(X^2/\hat{\delta})$ versus h_{ii} , where h_{ii} is the i^{th} diagonal element of H defined in (4.17). It may be noted that these procedures for detecting influential points are independent of the survey design since the fitting is based on pseudo likelihood equations and since $\tilde{X}_i^2/(X^2/\hat{\delta}) = X_i^2/X^2$.

(b) Diagnostics for assessing the impact of points on aspects of fit

While the quantities introduced above should indicate if any of the observations are not well explained by the model or are dominating some aspect of the fit of the model, they do not adequately measure the impact of such observations on the components of the fitted model. Two approaches for quantifying the effect of individual observations on the fit are

- (i) assessment by deletion, in which the change in some aspect of the fit incurred by deleting one or more data points, is computed.
- (ii) assessment by infinitesimal perturbations which involves weighting the contribution of one of the observations to the overall loglikelihood by the value $0 \leq w \leq 1$ and observing the effect on various quantities, of changes in w .

The three measures described below are of type (i).

- (1) Coefficient sensitivity: The quantity

$$(4.23) \quad \Delta_j(l) = [\hat{\theta}_j - \hat{\theta}_j(-l)] / \text{s.e.}(\hat{\theta}_j),$$

where $\hat{\theta}_j(-l)$ is the pseudo mle of θ_j after deletion of the l th point, provides a measure of the j th coefficient sensitivity to the l th point. The index plots of $\Delta_j(l)$ versus l are useful displays, but could become unwieldy for picking out observations with elevated influence on the fit if there are a large number of explanatory variables.

- (2) Sensitivity of fitted values: Appreciable changes in the estimates of some of the θ_j 's when the l th point is deleted does not always imply that the estimated values $\hat{m}$ differ significantly from $\underline{m}(\hat{\theta}(-l))$, the vector obtained from $\hat{m}$ by replacing $\hat{\theta}$ with $\hat{\theta}(-l)$. If $\tilde{G}^2(-l)$ and $\tilde{X}^2(-l)$ denote the usual likelihood ratio and Pearson chi-squared statistics, with $\underline{m}(\hat{\theta}(-l))$ replacing $\hat{m}$, then $[G^2 - \tilde{G}^2(-l)]/\hat{\delta}$ or $[X^2 - \tilde{X}^2(-l)]/\hat{\delta}$ may be used to assess the impact of the l th point on the fitted $\hat{m}_i$ values.

(3) Goodness of Fit: Let $G^2(-l)$ and $X^2(-l)$ denote the likelihood ratio and chi-squared statistics calculated from the $(R-1)$ data points other than the l^{th} . Then a measure of goodness-of-fit sensitivity is given by $[G^2 - G^2(-l)]/\hat{\delta}$. or $[X^2 - X^2(-l)]/\hat{\delta}$.

4.3 Application of Logit Model to Labour Force Survey Data

The results of Chapter 3 and Section 4.2 will now be applied to a logit analysis of a set of data from the October 1980 Canadian Labour Force Survey (LFS). Described below are the survey design, the hypothesis testing done, and the results of the application of the diagnostics to the final fitted model.

4.3.1 Survey design

The purpose of the analysis was to use the two factors, age and education, to explain the variation in unemployment rates using a logit model. The sample consisted of males aged 15-64 who were in the labour force and not full-time students. Age-group levels were formed by dividing the interval (15,64) into ten groups with the j^{th} age group being the interval $[10 + 5j, 14 + 5j]$, $j = 1, 2, \dots, 10$, and then using the mid-point of each interval, $A_j = 12 + 5j$, as the age value for all persons in that age group. Similarly, six education group levels were formed, with the value E_k assigned to the k^{th} group being based on the median years of schooling. The resultant values of E_k were 7, 10, 12, 13, 14 and 16. The age by education cross-classification thus produced a two-way table of $R = 60$ cell proportions.

The LFS design consisted of stratified multistage cluster sampling with two stages in the self-representing (SR) urban areas and three, or four stages in non-self-representing (NSR) areas in each province. The survey estimates $\hat{p}_{jk}$ of the unemployment rate in the j^{th} (age group and k^{th} education level were adjusted for post-stratification, using the projected census age-sex distribution at the provincial level. The survey estimate $\hat{\Sigma}$ of the covariance matrix of the $\hat{p}_{jk}$'s was based on more than 450 first-stage units so that the degrees of freedom for $\hat{\Sigma}$ were large compared to $R = 60$.

4.3.2 Tests of Hypotheses

The logit model initially considered to explain the variation in the p_{jk} 's was

$$(4.24) \quad v_{jk} = \ln(p_{jk}/(1-p_{jk})) = \theta_0 + \theta_1 A_j + \theta_2 A_j^2 + \theta_3 E_k + \theta_4 E_k^2, \quad j=1,2,\dots,10 \\ k=1,2,\dots,6,$$

where the number of parameters is $s = 5$. The form of this model was arrived at after examination of scatter plots of the logits $\hat{v}_{jk}$ ($= \ln(\hat{p}_{jk}/(1-\hat{p}_{jk}))$) versus age levels A_j at each education level E_k . These scatter plots tended to increase with age to a maximum, and then decrease. As well, some previous work in the sociological literature (Block and Smith, 1977) supported a model of this form.

Applying the formulae developed in Chapter 3, the following values were obtained.

$$\begin{aligned}
 (4.25) \quad X^2 &= 98.9 & G^2 &= 101.2 \\
 X^2/\hat{\delta} &= 52.5 & G^2/\hat{\delta} &= 53.7 & \hat{\delta} &= 1.88
 \end{aligned}$$

Since $\hat{\Sigma}$ was available, the test statistic of choice is the Satterthwaite X_s^2 , which, when adjusted to refer to $X_{.05}^2(55) = 73.3$, (where $X_{.05}^2(55)$ denotes the upper 5% point of the χ^2 distribution with $R-s = 55$ degrees of freedom) had the value 47.7, which is not significant. Both X^2 and G^2 are larger than $X_{.05}^2(55)$, which implies that the model would have been rejected if the survey design had been ignored. On the other hand, the values of $X^2/\hat{\delta}$ or $G^2/\hat{\delta}$ indicate (as did X_s^2) that the model is adequate, with the significance level (i.e. P value) being approximately 0.52. It should be noted that, since $\hat{\Sigma}$ was available, $\hat{\delta}$ could be estimated directly, instead of having to be approximated by an upper bound. However, the upper bound $(R/R-s)\hat{D} = (60/55)1.905 = 2.078$ depending only on the cell deffs $\hat{D}_i$ is close to $\hat{\delta}$ since $R = 60$ is large relative to $s = 5$.

Since two of the cells had $\hat{p}_{jk} = 1$, the Wald statistic X_W^2 was not defined. However, when minor perturbations were made to the estimated counts to ensure that $\hat{p}_{jk} < 1$ for all cells, and X_W^2 then calculated, the values of X_W^2 were all large compared to $X^2/\hat{\delta}$ (in fact, at least 30 times larger), and highly variable.

Table 4.1 contains the pseudo mle's of the θ_j 's under model (4.24) together with their standard errors and the corresponding standard errors if the sampling plan had been product binomial. As

expected, the true standard errors are larger than the corresponding binomial standard errors.

A series of nested hypotheses were also tested. Individually, each of the non-intercept terms of model (4.24) was deleted, leading to hypotheses $H_i : \theta_i = 0$, $i = 1, 2, 3, 4$, given model (4.24). The values of the Wald statistic $X_W^2(2/1)$ and the corrected G^2 statistic $G^2(2/1)/\hat{\delta} \cdot (H_i)$ are given in Table 4.1, for each of the H_i , where $\hat{\delta} \cdot (H_i)$ denotes the estimate of δ appropriate to H_i . The values of $X_W^2(2/1)$ and $G^2(2/1)/\hat{\delta} \cdot (H_i)$ were compared to $\chi_{.05}^2(1) = 3.84$, for testing the significance of the individual hypotheses. The coefficient θ_2 of A_i^2 was found to be highly significant, while the hypothesis $H_4 : \theta_4 = 0$ (i.e. the coefficient of E_j^2 is zero) is not rejected at the 5% level by either test.

Two other nested hypotheses, given model (4.24), were also tested, namely $H_{34} : \theta_3 = \theta_4 = 0$ (i.e. no education effect) and $H_{24} : \theta_2 = \theta_4 = 0$ (i.e. no quadratic effects). For these two hypotheses, the values of the Wald statistic and the adjusted G^2 statistic, which were compared to $\chi_{.05}^2(2) = 5.99$, were

$$(4.26) \quad G^2(2/1)/\hat{\delta} \cdot (H_{34}) = 282.2/1.64 = 172.1 \quad X_W^2(2/1) = 165.6 \text{ for } H_{34},$$

$$G^2(2/1)/\hat{\delta} \cdot (H_{24}) = 242.2/2.28 = 106.3 \quad X_W^2(2/1) = 162.1 \text{ for } H_{24}.$$

Obviously, H_{34} and H_{24} were both highly significant. Unlike the goodness-of-fit case, the Wald statistics yielded values close to the corresponding adjusted G^2 statistic, for all the nested hypotheses.

Through the results of the tests of goodness-of-fit and the tests of nested hypotheses described above, we have arrived at the following 4-parameter model to describe the variation of the P_{jk} 's:

$$(4.27) \quad v_{jk} = \ln(P_{jk}/(1 - P_{jk})) = \theta_0 + \theta_1 A_j + \theta_2 A_j^2 + \theta_3 E_k.$$

When this model was fitted, the estimates of the parameters (followed by their standard errors) were

$$(4.28) \quad \begin{array}{ll} \hat{\theta}_0 = -3.10 & \text{s.e.} = 0.247 \\ \hat{\theta}_1 = 0.211 & \text{s.e.} = 0.0130 \\ \hat{\theta}_2 = -0.00218 & \text{s.e.} = 0.000172 \\ \hat{\theta}_3 = 0.1509 & \text{s.e.} = 0.0115 \end{array}$$

Model (4.27) will be used for calculating the diagnostic measures in the following section.

4.3.3. Diagnostics

(i) Residual Analysis

The 60 cells in the two-way table were numbered lexicographically, and the standardized residuals e_i were computed under model (4.27). Cells numbered 6 and 54, both with $\hat{P}_{jk} = 1$, had very large e_i values, namely 166.6 and 6.2 respectively due to unstable variance estimates. Among the remaining cells, those numbered 7, 27 and 59 had values 3.84, 2.73 and 2.52. Since the expected number of $|e_i|$ exceeding 2.33 under model (4.27) is roughly $0.02 \times 60 = 1.2$, and since only 3 cells (other than those with $\hat{P}_{jk} = 1$) have $|e_i| > 2.33$, then cells 7 and 27 could correspond to outlying cell proportions.

The normal probability plot of $\tilde{G}_i$ is displayed in Figure 4.1; since the plot of $\tilde{X}_i$ was similar, it is not provided. There are no large deviations from a straight line configuration in Figure 4.1, and the index plot of $\tilde{G}_i$, Figure 4.2, is consistent with this. Thus, from the components $\tilde{G}_i$ of the corrected G^2 statistic, there is no evidence of outlying cell proportions.

(ii) Detection of Influential Cells

Examination of an index plot of m_{ii} and a plot of x_i^2/x^2 versus h_{ii} indicated that cells 2, 3, and 55 could be influential points.

(iii) Coefficient Sensitivity

The index plots of $\Delta_j(\ell)$ versus ℓ , defined in (4.23), for θ_0 , θ_1 , θ_2 and θ_3 are presented in Figures 4.3, 4.4, 4.5 and 4.6 respectively. From inspection of these plots it is evident that cells 2 and 3 cause instability in $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\theta}_2$ while $\hat{\theta}_3$ is particularly affected by cell 7.

(iv) Sensitivity of Fitted Values

The plot of $[G^2 - G^2(-\ell)]/\hat{\delta} = c_\ell$ versus ℓ is displayed in Figure 4.7. This plot is used for assessing the impact of individual cells on fitted values. Cells 2 and 3, and to a lesser extent, cell 7, produce outstanding peaks in this plot. Following Pregibon (1981), the comparison of c_ℓ to the percentage points of χ_s^2 (where s is the number of coefficients in the model) gives a rough guide as to which contour of the confidence region the pseudo mle is displaced

because of deletion of the l^{th} cell. The value $c_2 = 2.1$ (for cell 2) roughly corresponds to the 78% contour of the confidence region.

(v) Goodness of Fit Sensitivity

The plot of $[G^2 - G^2(-l)]/\hat{\delta}$ is displayed in Figure 4.8. (The plot of $[X^2 - X^2(-l)]/\hat{\delta}$ is not given since it is similar.) Cells 2, 3, 7, 27, 39, and 54 produce outstanding peaks in this plot, the largest peak being at cell 7, with a value of 5.4. By deleting cell 7 and calculating the appropriate value of $\hat{\delta}$ denoted by $\hat{\delta}(-7)$, for the remaining 59 points, it was found that $G_C^2(-7) = G^2(-7)/\hat{\delta}(-7) = 48.43$ with 55 d.f. compared to $G_C^2 = G^2/\hat{\delta} = 55.3$ with 56 d.f..

From the above investigation, we conclude that cells 7, 2 and 3 are possible candidates for deletion. However, the impact of these 3 cells may not be significant enough to warrant their deletion.

A Box-Cox model, as described in Subsection 3.3.3, was also considered for the LFS data. However, since λ did not appear to be significantly different from 0, with $\underline{\beta}$ consisting of (A_i, A_i^2, E_k, E_k^2) , the logit model was used for more extensive testing.

Table 4.1

Values of $\hat{\theta}_i$, $s.e(\hat{\theta}_i)$, $x_w^2(2/1) = \hat{\theta}_i^2/\text{var}(\hat{\theta}_i)$ and $G^2(2/1)/\hat{\delta} \cdot (H_i)$ for the LFS data under model (4.2)

i	$\hat{\theta}_i$	<u>s.e. ($\hat{\theta}_i$)</u>		$x_w^2(2/1)$	$G^2(2/1)/\hat{\delta} \cdot (H_i)$
		True	Binomial		
0	-2.76	0.5557	*		
1	0.209	0.0132	0.012	250.6	168.4
2	-0.00217	0.000173	0.000136	157.3	102.1
3	0.0913	0.0891	0.068	1.04	1.01
4	0.00276	0.00411	0.0030	0.45	<u>0.46</u>

* value not available

Figure 4.1: Normal Probability Plot of $\tilde{G}_i$

Figure 4.2: Index Plot of $\tilde{G}_i$

Figure 4.3: Index Plot of $\{\hat{\theta}_0 - \hat{\theta}_0(-L)\}/s.e.(\hat{\theta}_0)$

Figure 4.4: Index Plot of $\{\hat{\theta}_1 - \hat{\theta}_1(-L)\}/s.e.(\hat{\theta}_1)$

Figure 4.5: Index Plot of $\{\hat{\theta}_2 - \hat{\theta}_2(-l)\} / \text{s.e.}(\hat{\theta}_2)$

Figure 4.6: Index Plot of $\{\hat{\theta}_3 - \hat{\theta}_3(-l)\} / \text{s.e.}(\hat{\theta}_3)$

Figure 4.7: Index Plot of $\{G^2 - G^2(-1)\}/\delta$

Figure 4.8: Index Plot of $\{G^2 - G^2(-1)\}/\delta$

CHAPTER 5

POLYTOMOUS RESPONSE MODELS

5.1 Introduction

As noted in Section 3.1, the categorical variables defining the dimensions of a multidimensional data set are often considered to be a response variable together with several explanatory variables. When the response variable takes on only 2 values, as in Chapters 3 and 4, models describing the relationship between the response variable and the explanatory variables are called binary response models. On the other hand, when the response variable takes on more than 2 values, the models are labelled *polytomous response models*.

It seems reasonable that the statistical model employed in a particular situation should reflect the nature of the scale of measurement of the polytomous response variable. Of major interest in the analysis of survey data are nominal responses and ordinal responses. If the integers 1 to k are the labels for the k possible response categories, then for nominal responses the integers are simply labels and could equally well be associated with any permutation of the response categories; on the other hand, for ordinal responses, the labels 1 to k are associated with the natural order of the response categories. For ordinal responses, many models have been suggested (see Fienberg (1977), or McCullagh and Nelder (1983)) which have been based on partial sums of probabilities, rather than on the individual

probabilities for the response categories themselves. Generally, for nominal responses, models are in terms of response category probabilities.

Traditionally, when a polytomous response model has been used, the sampling scheme for collecting the data has been the product-multinomial, so that the test statistics are asymptotically distributed as χ^2 random variables. However, the asymptotic theory is not valid with survey data involving complex sampling schemes. In Section 5.2, the asymptotic theory of the Pearson chi-squared statistic (or the likelihood ratio statistic) will be developed under a specified sampling scheme when a general polytomous response model is used. The parameters of the model will be estimated by maximum likelihood methods appropriate to product-multinomial sampling. Corrections to the Pearson chi-squared statistic that take account of the true sampling plan will be obtained.

In Section 5.3, these results will be applied to two specific polytomous response models, one of which, proposed by Haberman (1982), would be suitable for nominal response variables. The other, described by McCullagh and Nelder (1983), would be appropriate for ordinal response variables.

5.2 Asymptotic Distribution of χ^2 for Polytomous Response Models

Suppose that the finite population of interest of size N is partitioned into R mutually exclusive exhaustive domains, the i th domain of size N_i . In the i th domain, suppose that N_{ij} units fall

into the j^{th} response category, $j = 1, 2, \dots, (J+1)$. The proportion of units in the i^{th} domain with the j^{th} response is thus

$$P_{j(i)} = N_{ij}/N_i \quad (\text{where } \sum_{j=1}^{J+1} P_{j(i)} = 1).$$

A general polytomous response model for the proportions $P_{j(i)}$ is denoted as

$$(5.1) \quad P_{j(i)} = m_{ij}(\underline{\theta}) \quad i=1, 2, \dots, R; \quad j=1, 2, \dots, (J+1),$$

where $\underline{\theta}$ is an r -vector of unknown parameters ($r \leq JR$) and $m_{ij}(\underline{\theta})$ is a function of known form satisfying the regularity conditions of Birch (1964), given by BFH (1975, p.510). For ease of presentation of the results to be developed, the following notation will be used.

$$(5.2) \quad \underline{P}_i = (P_{1(i)}, P_{2(i)}, \dots, P_{J(i)})' \quad (J \times 1 \text{ vector})$$

$$(5.3) \quad \underline{P}_i = (P_{1(i)}, P_{2(i)}, \dots, P_{J+1(i)})' \quad ((J+1) \times 1 \text{ vector})$$

$$(5.4) \quad \underline{P} = (\underline{P}_1', \underline{P}_2', \dots, \underline{P}_R')' \quad (JR \times 1 \text{ vector})$$

and

$$(5.5) \quad \underline{P} = (\underline{P}_1', \underline{P}_2', \dots, \underline{P}_R')' \quad ((J+1)R \times 1 \text{ vector})$$

Suppose that a sample $\tilde{s}$ of n ultimate units is drawn from the population according to a specified survey design $p(\tilde{s})$. Let $\hat{N}$, $\hat{N}_i$ and $\hat{N}_{ij}$ be the survey estimates of the total population size, the size of the i^{th} domain, and the number of units with the j^{th} response in the i^{th} domain respectively, such that the following two assumptions hold:

(a) $\underline{w} = \underline{W} + o_p(n^{-1/2})$ where $\underline{w} = (w_1, w_2, \dots, w_R)'$,
 $w_i = \hat{N}_i / \hat{N}$, $\underline{W} = (W_1, W_2, \dots, W_R)'$ and $W_i = N_i / N$.

(b) $\sqrt{n}(\hat{\underline{P}} - \underline{P})$ is asymptotically distributed as a $N(0, \Sigma)$ vector as
 $n \rightarrow \infty$, where $\hat{P}_{j(i)} = \hat{N}_{ij} / \hat{N}_i$ is the survey estimate of $P_{j(i)}$,
and $\hat{\underline{P}}$ is obtained from (5.4) by replacing $P_{j(i)}$ with $\hat{P}_{j(i)}$.

The likelihood under the assumption of independent multinomial sampling in each domain is

$$(5.6) \quad L = \prod_{i=1}^R \left[\frac{n_i!}{J+1} \prod_{j=1}^{J+1} (m_{ij}(\underline{\theta}))^{n_{ij}} \right],$$

where n_i and n_{ij} are the size of sample and the number of sampled units with the j^{th} response respectively in the i^{th} domain, with

$$\sum_{j=1}^{J+1} n_{ij} = n_i \quad \text{and} \quad \sum_{i=1}^R n_i = n. \quad \text{When } L \text{ is maximized, subject to the}$$

constraints that $\sum_{j=1}^{J+1} m_{ij}(\underline{\theta}) = 1$, $i = 1, 2, \dots, R$, the likelihood equations are

$$(5.7) \quad \sum_{i=1}^R \frac{n_i}{n} \sum_{j=1}^{J+1} \frac{n_{ij}}{n_i} \frac{1}{m_{ij}(\underline{\theta})} \frac{\partial m_{ij}(\underline{\theta})}{\partial \theta_k} \bigg|_{\underline{\theta} = \hat{\underline{\theta}}} = 0 \quad k=1, 2, \dots, r.$$

Because the true sampling plan is not product-multinomial, "pseudo likelihood equations" are obtained by replacing n_i/n with w_i and n_{ij}/n_i with $\hat{P}_{j(i)}$ in (5.7), yielding

$$(5.8) \quad \sum_{i=1}^R w_i \sum_{j=1}^{J+1} \frac{\hat{P}_{j(i)}}{m_{ij}(\underline{\theta})} \frac{\partial m_{ij}(\underline{\theta})}{\partial \theta_k} \bigg|_{\underline{\theta} = \hat{\underline{\theta}}} = 0 \quad k=1, 2, \dots, r.$$

The solution of (5.8) for $\hat{\theta}$, and the resulting $\hat{m}_{ij} = m_{ij}(\hat{\theta})$ are the pseudo mle's for θ and $P_{j(i)}$ respectively. The same result (5.8) would obviously be obtained if n_i/n and n_{ij}/n_i in (5.6) were replaced by w_i and $\hat{P}_{j(i)}$ respectively, to yield $\bar{L}$, and then $\bar{L}$ was maximized. The consistency of $\hat{P}$ and $\underline{w}$ ensures the consistency of $\hat{\theta}$ and $\hat{\underline{m}}$, under standard regularity conditions, where $\hat{\underline{m}} = (\hat{\underline{m}}_1', \hat{\underline{m}}_2', \dots, \hat{\underline{m}}_R')'$ and $\hat{\underline{m}}_i = (\hat{m}_{i1}, \hat{m}_{i2}, \dots, \hat{m}_{iJ})'$.

The Pearson chi-squared statistic used to test $H : \underline{P} = \underline{m}(\theta)$ (where $\underline{m}(\theta) = (\underline{m}_1(\theta)', \underline{m}_2(\theta)', \dots, \underline{m}_R(\theta)')'$ and $\underline{m}_i(\theta) = (m_{i1}(\theta), m_{i2}(\theta), \dots, m_{iJ}(\theta))$), in the classical product multinomial situation is

$$(5.9) \quad \chi_c^2 = \sum_{i=1}^R \sum_{j=1}^{J+1} (n_{ij} - n_i m_{ij}(\hat{\theta}))^2 / n_i m_{ij}(\hat{\theta}),$$

where $\hat{\theta}$ is the solution of (5.7). For a general sampling scheme, the parallel statistic obtained by substituting w_i and $\hat{P}_{j(i)}$ for n_i/n and n_{ij}/n_i respectively and having $\hat{\theta}$, the solution of (5.8), is

$$(5.10) \quad \chi^2 = n \sum_{i=1}^R w_i \sum_{j=1}^{J+1} (\hat{P}_{j(i)} - m_{ij}(\hat{\theta}))^2 / m_{ij}(\hat{\theta}).$$

If Q_i denotes the $J \times J$ matrix

$$(5.11) \quad Q_i = \text{diag}(\underline{m}_i) - \underline{m}_i \underline{m}_i'$$

and Q denotes the $JR \times JR$ block-diagonal matrix with the Q_i 's on the diagonal, then (5.10) may be written as

$$(5.12) \quad \begin{aligned} \chi^2 &= n \sum_{i=1}^R w_i (\hat{\underline{P}}_i - \underline{m}_i(\hat{\theta}))' \hat{Q}_i^{-1} (\hat{\underline{P}}_i - \underline{m}_i(\hat{\theta})) \\ &= n (\hat{\underline{P}} - \hat{\underline{m}})' \hat{Q}^{-1} (D(\underline{w}) \otimes I_J) (\hat{\underline{P}} - \hat{\underline{m}}), \end{aligned}$$

where $D(\underline{w}) = \text{diag}(\underline{w})$, I_J is the $J \times J$ identity matrix, $\otimes$ denotes the direct product operator and $\hat{Q}$ is obtained from Q by replacing $\underline{m}_i = \underline{m}_i(\underline{\theta})$ with $\hat{\underline{m}}_i$.

With the notation introduced in the last paragraph, the pseudo likelihood equations of (5.8) may be written in matrix form as

$$(5.13) \quad \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)' \bigg|_{\underline{\theta} = \hat{\underline{\theta}}} (D(\underline{w}) \otimes I_J) \hat{Q}^{-1} \hat{\underline{P}} = \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)' \bigg|_{\underline{\theta} = \hat{\underline{\theta}}} (D(\underline{x}) \otimes I_J) \underline{1}_{JR}$$

where $D(\underline{x}) = \text{diag}(\underline{x})$, $x_i = w_i (1 - \underline{1}_J' \hat{\underline{m}}_i)^{-1}$, $\underline{1}_{JR}$ and $\underline{1}_J$ are unit vectors of lengths JR and J respectively, and $\frac{\partial \underline{m}}{\partial \underline{\theta}}$ is the $JR \times r$ matrix of the derivatives of the functions $m_{ij}(\underline{\theta})$ with respect to the elements of $\underline{\theta}$. It is assumed that $\left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)$ is of full rank r .

5.2.1 Asymptotic Covariance matrix of $\hat{\underline{m}}$:

Lemma 5.1. Under the regularity conditions of Birch (1964),

$$(5.14) \quad \hat{\underline{\theta}} - \underline{\theta} = \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} \right)' \nabla \frac{\partial \underline{m}}{\partial \underline{\theta}}^{-1} \frac{\partial \underline{m}}{\partial \underline{\theta}} \nabla (\hat{\underline{P}} - \underline{m}(\underline{\theta})) + o_p(n^{-1/2})$$

where ∇ is the $JR \times JR$ matrix defined by

$$(5.15) \quad \nabla = (D(\underline{w}) \otimes I_J) \hat{Q}^{-1}.$$

Proof: The function $\bar{L}$ which is maximized to obtain $\hat{\underline{\theta}}$ is defined by

$$-2 \ln \bar{L} = -2n \sum_{i=1}^R w_i \sum_{j=1}^{J+1} \hat{p}_{j(i)} \ln m_{ij}(\underline{\theta}) + a$$

where

$$a = \ln \left[\prod_{i=1}^R (nw_i)! / \prod_{j=1}^{J+1} (nw_i \hat{p}_{j(i)}) \right].$$

This quantity may also be expressed as

$$\begin{aligned}
 (5.16) \quad -2 \ln \bar{L} &= 2n \sum_i \sum_j w_i \hat{p}_{j(i)} \ln w_i \hat{p}_{j(i)} - 2n \sum_i \sum_j w_i \hat{p}_{j(i)} \ln w_i m_{ij}(\underline{\theta}) \\
 &\quad - 2n \sum_i \sum_j w_i \hat{p}_{j(i)} \ln w_i \hat{p}_{j(i)} + 2n \sum_i \sum_j w_i \hat{p}_{j(i)} \ln w_i + a \\
 &= 2n \bar{G}^2(\hat{\underline{v}}, \underline{\gamma}(\underline{\theta})) + \text{terms not involving } \underline{\theta},
 \end{aligned}$$

where

$$(5.17) \quad \hat{\underline{v}} = [w_{1=1} \hat{p}', w_{2=2} \hat{p}', \dots, w_{R=R} \hat{p}']' = [D(\underline{w}) \otimes I_{J+1}] \hat{\underline{p}},$$

$$(5.18) \quad \underline{\gamma}(\underline{\theta}) = [w_{1=1} m(\underline{\theta})', w_{2=2} m(\underline{\theta})', \dots, w_{R=R} m(\underline{\theta})']' = (D(\underline{w}) \otimes I_{J+1}) \underline{m}(\underline{\theta}),$$

$$\underline{m}_i(\underline{\theta}) = (m_{i1}(\underline{\theta}), m_{i2}(\underline{\theta}), \dots, m_{i,J+1}(\underline{\theta}))',$$

and

$$(5.19) \quad \bar{G}^2 = (\underline{x}, \underline{y}) = \sum_i x_i \ln(x_i/y_i), \text{ for any vectors } \underline{x} \text{ and } \underline{y}.$$

It is readily proved that $\underline{1}' \hat{\underline{v}} = \underline{1}' \underline{\gamma}(\underline{\theta}) = 1$. To maximize $\bar{L}$ we must choose the value of $\underline{\theta}$ to minimize $\bar{G}^2(\hat{\underline{v}}, \underline{\gamma}(\underline{\theta}))$. Birch (1964) defines $\underline{\theta}(\underline{v})$ to be that value of $\underline{\theta}$ such that

$$\bar{G}^2(\underline{v}, \underline{\gamma}(\underline{\theta}(\underline{v}))) = \inf_{\underline{\theta}} \bar{G}^2(\underline{v}, \underline{\gamma}(\underline{\theta})).$$

The value of $\underline{\theta}$ to maximize $\bar{L}$ is thus $\hat{\underline{\theta}} = \underline{\theta}(\hat{\underline{v}})$. Birch (1964) proved that

$$(5.20) \quad \underline{\theta}(\underline{v}) = \underline{\theta} + (B'B)^{-1} B'D(\underline{\gamma}(\underline{\theta}))^{-1} (\underline{v} - \underline{\gamma}(\underline{\theta})) + o(\|\underline{v} - \underline{\gamma}(\underline{\theta})\|)$$

as $\underline{v} \rightarrow \underline{\gamma}(\underline{\theta})$, where B is the $(J+1)R \times r$ matrix defined by

$B = D(\underline{Y}(\underline{\theta}))^{-1} \frac{\partial \underline{Y}}{\partial \underline{\theta}}$ and $\underline{v}$ has the form $[w_1 p'_1, \dots, w_{R-R} p'_{R-R}]'$.

In order to express (5.20) in terms of the vector $\underline{m}(\underline{\theta})$, we note that, due to the constraints, $\sum_{j=1}^{J+1} m_{ij}(\underline{\theta}) = 1$, $i=1,2,\dots,R$, then $\underline{m}(\underline{\theta})$ may be written as

$$(5.21) \quad \underline{m}(\underline{\theta}) = [\underline{m}_1(\underline{\theta})', 1 - \underline{1}'_J \underline{m}_1(\underline{\theta}), \underline{m}_2(\underline{\theta})', 1 - \underline{1}'_J \underline{m}_2(\underline{\theta}), \dots, \underline{m}_R(\underline{\theta})', 1 - \underline{1}'_J \underline{m}_R(\underline{\theta})]'$$

As a result of (5.21),

$$(5.22) \quad \frac{\partial \underline{Y}}{\partial \underline{\theta}} = (D(\underline{w}) \otimes I_{J+1}) \left[\frac{\partial \underline{m}_1}{\partial \underline{\theta}}, -\frac{\partial \underline{m}_1}{\partial \underline{\theta}} \underline{1}_J, \dots, \frac{\partial \underline{m}_R}{\partial \underline{\theta}}, -\frac{\partial \underline{m}_R}{\partial \underline{\theta}} \underline{1}_J \right]$$

and thus,

$$\begin{aligned} (5.23) \quad B'B &= \sum_{i=1}^R w_i \left[\frac{\partial \underline{m}_i}{\partial \underline{\theta}} D(\underline{m}_i)^{-1} \frac{\partial \underline{m}_i}{\partial \underline{\theta}} + \frac{\partial \underline{m}_i}{\partial \underline{\theta}} \underline{1}_J (1 - \underline{1}'_J \underline{m}_i)^{-1} \underline{1}'_J \frac{\partial \underline{m}_i}{\partial \underline{\theta}} \right] \\ &= \sum_{i=1}^R w_i \frac{\partial \underline{m}_i}{\partial \underline{\theta}} \left[D(\underline{m}_i)^{-1} + (1 - \underline{1}'_J \underline{m}_i)^{-1} \underline{1}_J \underline{1}'_J \right] \frac{\partial \underline{m}_i}{\partial \underline{\theta}} \\ &= \sum_{i=1}^R w_i \frac{\partial \underline{m}_i}{\partial \underline{\theta}} Q_i^{-1} \frac{\partial \underline{m}_i}{\partial \underline{\theta}} \\ &= \frac{\partial \underline{m}}{\partial \underline{\theta}} [D(\underline{w}) \otimes I_J] Q^{-1} \frac{\partial \underline{m}}{\partial \underline{\theta}} \end{aligned}$$

and

$$(5.24) \quad B'D(\underline{Y}(\underline{\theta}))^{-1} (\underline{v} - \underline{Y}(\underline{\theta})) = \frac{\partial \underline{m}}{\partial \underline{\theta}} [D(\underline{w}) \otimes I_J] Q^{-1} (\underline{p} - \underline{m}(\underline{\theta})).$$

Substituting the above expressions into (5.20), we have

$$\underline{\theta}(\underline{v}) = \underline{\theta} + \left(\frac{\partial \underline{m}}{\partial \underline{\theta}} [D(\underline{w}) \otimes I_J] Q^{-1} \frac{\partial \underline{m}}{\partial \underline{\theta}} \right)^{-1} \frac{\partial \underline{m}}{\partial \underline{\theta}} [D(\underline{w}) \otimes I_J] Q^{-1} (\underline{p} - \underline{m}(\underline{\theta})) + o(\|\underline{v} - \underline{Y}(\underline{\theta})\|)$$

as $\underline{v} \rightarrow \underline{Y}(\underline{\theta})$

Since, by assumption, $\sqrt{n}(\hat{\underline{P}} - \underline{m}(\underline{\theta})) \approx N(0, \Sigma)$ and $\underline{w} = \underline{W} + o_p(n^{-1/2})$, we then have $\hat{\underline{P}} - \underline{m}(\underline{\theta}) = o_p(n^{-1/2})$, $\hat{\underline{\gamma}} - \underline{\gamma}(\underline{\theta}) = o_p(n^{-1/2})$, and $[D(\underline{w}) \otimes I_J]Q^{-1} = \underline{V} + o_p(n^{-1/2})$. Therefore it follows that

$$(5.25) \quad \hat{\underline{\theta}} = \underline{\theta}(\hat{\underline{P}}) = \underline{\theta} + \left(\frac{\partial \underline{m}'}{\partial \underline{\theta}} \underline{V} \frac{\partial \underline{m}}{\partial \underline{\theta}}\right)^{-1} \frac{\partial \underline{m}'}{\partial \underline{\theta}} \underline{V}(\hat{\underline{P}} - \underline{m}(\underline{\theta})) + o_p(n^{-1/2}) \quad \text{Q.E.D.}$$

Since it was assumed that $\sqrt{n}(\hat{\underline{P}} - \underline{m}(\underline{\theta})) \approx N(0, \Sigma)$, the asymptotic covariance matrix of $\hat{\underline{P}}$ is

$$(5.26) \quad C(\hat{\underline{P}}) = \Sigma/n.$$

This fact, together with (5.25), implies that the asymptotic covariance matrix of $\hat{\underline{\theta}}$ is

$$(5.27) \quad C(\hat{\underline{\theta}}) = n^{-1}(\underline{M}' \underline{V} \underline{M})^{-1} (\underline{M}' \underline{V} \Sigma \underline{V} \underline{M}) (\underline{M}' \underline{V} \underline{M})^{-1},$$

where, for the remainder of the chapter, $\underline{M}$ denotes the $JR \times r$ matrix $\left(\frac{\partial \underline{m}}{\partial \underline{\theta}}\right)$.

In the case of product multinomial sampling,

$$\Sigma = \underline{V}^{-1} = (D(\underline{w})^{-1} \otimes I_J)Q \quad \text{and} \quad C(\hat{\underline{\theta}}) \text{ reduces to } C_M(\hat{\underline{\theta}}) = n^{-1}(\underline{M}' \underline{V} \underline{M})^{-1}.$$

Because of the regularity conditions,

$$(5.28) \quad \underline{m}(\hat{\underline{\theta}}) - \underline{m}(\underline{\theta}) = \underline{M}(\hat{\underline{\theta}} - \underline{\theta}) + o_p(n^{-1/2}).$$

Substituting (5.25) into (5.28), it follows that

$$(5.29) \quad \underline{m}(\hat{\underline{\theta}}) - \underline{m}(\underline{\theta}) = \underline{M}(\underline{M}' \underline{V} \underline{M})^{-1} \underline{M}' \underline{V}(\hat{\underline{P}} - \underline{m}(\underline{\theta})) + o_p(n^{-1/2})$$

so that the asymptotic covariance matrix of $\underline{m}(\hat{\theta})$ is

$$(5.30) \quad C(\underline{m}(\hat{\theta})) = MC(\hat{\theta})M'.$$

As noted in Chapter 2, the vector $\hat{\underline{p}} - \underline{m}(\hat{\theta})$, called the vector of residuals, is often useful in detecting model deviations. Since

$$(5.31) \quad \begin{aligned} \hat{\underline{p}} - \underline{m}(\hat{\theta}) &= (\hat{\underline{p}} - \underline{m}(\underline{\theta})) - (\underline{m}(\hat{\theta}) - \underline{m}(\underline{\theta})) \\ &= (I - M(M'VM)^{-1}M'V)(\hat{\underline{p}} - \underline{m}(\underline{\theta})) + o_p(n^{-1/2}), \end{aligned}$$

then the asymptotic covariance matrix of the residual vector is

$$(5.32) \quad C(\hat{\underline{p}} - \underline{m}(\hat{\theta})) = n^{-1}A\Sigma A',$$

where

$$A = (I - M(M'VM)^{-1}M'V).$$

Consistent estimators of the above covariance matrices are obtained by replacing $\underline{\theta}$ by $\hat{\underline{\theta}}$ and Σ by the survey estimate $\hat{\Sigma}$ (when available).

5.2.2. Nested Models

Let M_1 denote the model

$$p_{j(i)} = m_{ij}(\underline{\theta}) \quad i=1,2,\dots,R, \quad j=1,2,\dots,J+1,$$

which is equivalent to

$$(5.33) \quad \underline{p} = \underline{m}(\underline{\theta}),$$

because of the restrictions $\sum_{j=1}^{J+1} p_{j(i)} = \sum_{j=1}^{J+1} m_{ij}(\underline{\theta}) = 1, i=1,2,\dots,R.$

Suppose that the vector $\underline{\theta}$ can be partitioned as $\underline{\theta} = (\underline{\theta}_1', \underline{\theta}_2')'$, where $\underline{\theta}_1$ is $s \times 1$ and $\underline{\theta}_2$ is $u \times 1$ ($s+u = r$). If we wished to test the hypothesis $H : \underline{\theta}_2 = \underline{0}$, given M_1 , we would have the reduced model $M_2 : \underline{p} = \underline{m}(\underline{\theta}_1)$ where " $\underline{m}(\underline{\theta}_1)$ " denotes " $\underline{m}(\underline{\theta}_1, \underline{0})$ ". Let $\hat{\underline{\theta}}_1$ and $\hat{\underline{m}} = \underline{m}(\hat{\underline{\theta}}_1)$ be the pseudo mle's of $\underline{\theta}_1$ and $\underline{p}$ respectively under M_2 , obtained from the pseudo likelihood equations

$$(5.34) \quad \left. \frac{\partial \underline{m}'}{\partial \underline{\theta}_1} \right|_{\underline{\theta} = (\hat{\underline{\theta}}_1, \underline{0})} (D(\underline{w}) \otimes I_J) \hat{\underline{Q}}^{-1} \hat{\underline{p}} = \left. \frac{\partial \underline{m}'}{\partial \underline{\theta}_1} \right|_{\underline{\theta} = (\hat{\underline{\theta}}_1, \underline{0})} (D(\underline{x}_1) \otimes I_J) \underline{1}_J,$$

where $\hat{\underline{Q}} = Q(\hat{\underline{\theta}}_1, \underline{0})$ and $D(\underline{x}_1) = \text{diag}(\underline{x}_1)$ with $x_{1i} = w_i (1 - \underline{1}' \underline{m}_i(\hat{\underline{\theta}}_1))^{-1}$.

The Pearson chi-squared statistic for testing H is given by

$$\begin{aligned} (5.35) \quad X^2 &= n \sum_{i=1}^R w_i \sum_{j=1}^{J+1} (m_{ij}(\hat{\underline{\theta}}) - m_{ij}(\hat{\underline{\theta}}_1))^2 / m_{ij}(\hat{\underline{\theta}}_1) \\ &= n \sum_{i=1}^R w_i (\underline{m}_i(\hat{\underline{\theta}}) - \underline{m}_i(\hat{\underline{\theta}}_1))' \hat{\underline{Q}}_i^{-1} (\underline{m}_i(\hat{\underline{\theta}}) - \underline{m}_i(\hat{\underline{\theta}}_1)) \\ &= n (\underline{m}(\hat{\underline{\theta}}) - \underline{m}(\hat{\underline{\theta}}_1))' \hat{\underline{Q}}^{-1} (D(\underline{w}) \otimes I_J) (\underline{m}(\hat{\underline{\theta}}) - \underline{m}(\hat{\underline{\theta}}_1)). \end{aligned}$$

By definition,

$$M = \frac{\partial \underline{m}}{\partial \underline{\theta}} = \left(\frac{\partial \underline{m}}{\partial \underline{\theta}_1}, \frac{\partial \underline{m}}{\partial \underline{\theta}_2} \right),$$

and, since M is $JR \times r$ of rank r , $\frac{\partial \underline{m}}{\partial \underline{\theta}_1}$ must be of rank s and $\frac{\partial \underline{m}}{\partial \underline{\theta}_2}$ of rank u . Let M_1 denote $\frac{\partial \underline{m}}{\partial \underline{\theta}_1}$ and M_2 denote $\frac{\partial \underline{m}}{\partial \underline{\theta}_2}$.

Theorem 5.1. Under $H : \underline{\theta}_2 = \underline{0}$,

$$(5.36) \quad X^2 \sim n \hat{\underline{\theta}}_2' (H_2' \nabla H_2) \hat{\underline{\theta}}_2;$$

where $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ is the pseudo mle of θ under M_1 , and

$$(5.37) \quad H_2 = (I - M_1 (M_1' \nabla M_1)^{-1} M_1' \nabla) M_2 .$$

Furthermore,

$$X^2 \approx \sum_{i=1}^u \delta_i Z_i^2 ,$$

where the Z_i^2 are independent χ_1^2 variables and the δ_i are the eigenvalues of

$$(5.38) \quad n C(\hat{\theta}_2) (H_2' \nabla H_2) = (H_2' \nabla H_2)^{-1} (H_2' \nabla \Sigma \nabla H_2) .$$

Proof: The proof parallels that for the similar results for the binary response model in Chapter 3, with ∇ replacing E .

As a special case, under product multinomial sampling, $\Sigma = \nabla^{-1}$ under H , and $(H_2' \nabla H_2)^{-1} (H_2' \nabla \nabla^{-1} \nabla H_2) = I$ so that $\delta_i = 1, i=1,2,\dots,u$ and we get the standard result that $X^2 \approx \chi_u^2$ under H .

It should be noted that the asymptotic covariance matrix of $H_2' \nabla \hat{p}$ is $n^{-1} (H_2' \nabla \Sigma \nabla H_2)$ under the survey design and $n^{-1} (H_2' \nabla \nabla^{-1} \nabla H_2) = n^{-1} (H_2' \nabla H_2)$, under product multinomial sampling, so that the δ_i are the "generalized design effects" of the vector $H_2' \nabla \hat{p}$, as defined by Rao and Scott (1981).

It should also be noted that the asymptotic distribution of the Pearson chi-squared statistic (5.12) for testing the goodness of fit of model M_1 may be obtained as a special case of Theorem 5.1. For this case, M_1 must be considered to be nested within a saturated model, i.e. a model where $s+u = JR$.

5.2.3. Adjustments to χ^2

Since the asymptotic distribution of χ^2 is complex, it is important to have a correction for χ^2 that can be practically implemented with survey data. One approach likely to be adequate in many situations is to treat $\chi^2/\hat{\delta}$ as χ^2_u where

$$(5.39) \quad u\hat{\delta} = \sum \delta_i = \text{tr}(H_2' \nabla H_2)^{-1} (H_2' \nabla \Sigma \nabla H_2),$$

and $\hat{\delta}$ is the survey estimate of δ . However, $\hat{\delta}$ depends on the full estimated covariance matrix of $\hat{p}$, which is frequently unavailable. It is worthwhile, therefore, to obtain an upper bound on δ which depends only on the cell design effects or on generalized design effects. One method to obtain such a bound is described below.

Since $H_2' \nabla H_2$ is a symmetric matrix, there exists a nonsingular matrix B such that

$$BH_2' \nabla H_2 B' = D,$$

where D is a diagonal matrix (C.R. Rao (1965), p.20). Let

$$Y = BH_2' = (Y_1, Y_2, \dots, Y_u)'$$

It follows that $Y \nabla Y' = D$ and thus

$$(5.40) \quad Y_i \nabla Y_j = 0 \quad \text{for } i \neq j.$$

Therefore,

$$\begin{aligned}
 (5.41) \quad u\delta &= \text{tr}(\mathbf{H}_2' \nabla \mathbf{H}_2)^{-1} (\mathbf{H}_2' \nabla \Sigma \nabla \mathbf{H}_2) \\
 &= \text{tr}(\mathbf{B}')^{-1} (\mathbf{H}_2' \nabla \mathbf{H}_2)^{-1} \mathbf{B}^{-1} \mathbf{B} (\mathbf{H}_2' \nabla \Sigma \nabla \mathbf{H}_2) \mathbf{B}' \\
 &= \text{tr}(\mathbf{Y} \nabla \mathbf{Y}')^{-1} (\mathbf{Y} \nabla \Sigma \nabla \mathbf{Y}') \\
 &= \sum_{i=1}^u [\mathbf{y}_i' (\nabla \Sigma \nabla) \mathbf{y}_i] / \mathbf{y}_i' \nabla \mathbf{y}_i.
 \end{aligned}$$

Now, since $\mathbf{Q}_i$ is a positive definite covariance matrix for each i , it follows that $\nabla = (\mathbf{D}(\mathbf{W}) \times \mathbf{I}_J) \mathbf{Q}$ is also positive definite. Then, by lemma 3.2,

$$\sum_{i=1}^u [\mathbf{y}_i' (\nabla \Sigma \nabla) \mathbf{y}_i] / \mathbf{y}_i' \nabla \mathbf{y}_i \leq \lambda_1 + \lambda_2 + \dots + \lambda_u,$$

where $\lambda_1 \geq \lambda_2 \dots \geq \lambda_{JR} > 0$ are the eigenvalues of $\nabla \Sigma$. Thus,

$$(5.42) \quad \delta \leq u^{-1} \sum_{i=1}^u \lambda_i.$$

It should be noted that the eigenvalues of $\nabla \Sigma$ are independent of any hypothesis. Hence, it should be possible to report all the (estimated) eigenvalues with the published tables, so that the required number for bounding δ for any particular hypothesis could be chosen.

If the estimated eigenvalues of $\nabla \Sigma$ are not available but deffs of the estimated proportions $\hat{p}_{j(i)}$ are published, an upper bound may still be estimated by noting that

$$(5.43) \quad u\delta \leq \sum_{i=1}^u \lambda_i \leq \sum_{i=1}^{JR} \lambda_i = JR\lambda.$$

and

$$(5.44) \quad JR\lambda = \text{tr } \nabla \Sigma = \sum_{i=1}^R \sum_{j=1}^J v_{jj(i)} / [p_{j(i)} (1 - p_{j(i)}) / w_i]$$

$$= \sum_{i=1}^R \sum_{j=1}^J D_{jj(i)} = JRD.$$

where $v_{jj(i)}$ and $D_{jj(i)}$ are the variance and deff respectively of $\hat{p}_{j(i)}$. However, this upper bound on δ may not be satisfactory if u is small compared to JR .

When model M_1 is saturated, that is $s+u = JR$, (5.43) and (5.44) lead to

$$(5.45) \quad (JR - s)\delta \leq JRD.$$

It follows from (5.45) that D is a satisfactory upper bound on δ if $s \ll JR$.

When the full estimated covariance matrix of $\hat{p}$ is available, a more accurate correction to X^2 than X^2/δ is obtained by a Satterthwaite approximation in which

$$(5.46) \quad X_s^2 = X^2 / [(1+a^2)\delta]$$

is treated as χ_v^2 where $v = u/(1+a^2)$ and $a^2 = \sum (\delta_i - \delta)^2 / u\delta^2$ is the square of the coefficient of variation of the δ_i 's. It is not necessary, however, to evaluate the individual δ_i 's in order to compute X_s^2 . The approach used for the Satterthwaite correction for the binary response model in Chapter 3 may be followed for the polytomous response models as well.

5.3. Examples of Polytomous Response Models

In Section 5.2, results were obtained for general polytomous

response models of the form $P_{j(i)} = m_{ij}(\theta)$. In this section, a selection of these results are applied to two particular models. First to be examined is a model suggested by McCullagh and Nelder (1983), which is particularly applicable to data for which there is a natural ordering of the categories of the response variable, since the model is expressed in terms of the cumulative distribution probabilities. Next to be examined is a model studied by Haberman (1982) and Press (1972), frequently called the multinomial response model. Since, in this model, no ordering of the response categories is assumed, the model is more suited to a nominal response variable.

5.3.1. McCullagh and Nelder's model for ordinal scales of measurement

As in Section 5.2, let $P_{j(i)}$ denote the proportion with the j th response in the i th domain, $j=1,2,\dots,J+1$; $i=1,2,\dots,R$. Also, let $G_{j(i)}$ denote the j th cumulative probability in the i th domain, defined by

$$(5.47) \quad G_{j(i)} = \sum_{k=1}^j P_{k(i)}$$

The model proposed by McCullagh and Nelder (1983) has the form

$$(5.48) \quad \ln(G_{j(i)}/(1-G_{j(i)})) = v_j - \beta' y_i, \quad j=1,2,\dots,J$$

where v_j and β are unknown parameters and y_i is a known vector of length ℓ . Now, from the definition of the cumulative probabilities, it is obvious that

$$(5.49) \quad P_{j(i)} = \begin{cases} G_{1(i)} & \text{for } j=1 \\ G_{j(i)} - G_{j-1(i)} & \text{for } j=2,3,\dots,J \\ 1 - G_{J(i)} & \text{for } j=J+1 \end{cases}$$

or, equivalently,

$$(5.50) \quad \underline{G}_i = \underline{L} \underline{P}_i$$

where $\underline{G}_i = (G_{1(i)}, G_{2(i)}, \dots, G_{J(i)})$ and $\underline{L}$ is the $J \times J$ nonsingular matrix of the form

$$(5.51) \quad \underline{L} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}$$

and with an inverse of the form

$$\underline{L}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

In order to apply the results of Section 5.2, the matrix $\underline{M} = \left(\frac{\partial m}{\partial \theta} \right)$ must be obtained, where $\underline{\theta} = (v_1, v_2, \dots, v_J, \underline{\beta}')'$ is of length $J+1$, and where $\underline{M}$ is the matrix of partial derivatives derived from the model (5.48) expressed in the form $P_{j(i)} = m_{ij}(\underline{\theta})$. Using the facts that $\underline{P}_i = \underline{L}^{-1} \underline{G}_i$ and that, from (5.48),

$$(5.52) \quad G_{j(i)} = e^{v_j} / (e^{v_j} + e^{\underline{\beta}' \underline{y}_i}),$$

routine calculations yield

$$(5.53) \quad \partial G_{j(i)} / \partial v_k = 0 \quad \text{if } j \neq j, \quad \text{if } j \neq j,$$

$$(5.54) \quad \partial G_{j(i)} / \partial v_j = G_{j(i)} (1 - G_{j(i)}) ,$$

$$(5.55) \quad \partial G_{j(i)} / \partial \beta_k = -G_{j(i)} (1 - G_{j(i)}) y_{ik} ,$$

and thus

$$(5.56) \quad M = (I_R \otimes L^{-1}) G [I_R \otimes I_J, A] ,$$

where G is a $JR \times JR$ block-diagonal matrix with the $J \times J$ matrix $D(\underline{G}_i) D(\underline{1} - \underline{G}_i)$ forming the i th block on the diagonal, and

$$(5.57) \quad A = \begin{pmatrix} \underline{1}_J \otimes Y_1' \\ \underline{1}_J \otimes Y_2' \\ \vdots \\ \underline{1}_J \otimes Y_R' \end{pmatrix}$$

Suppose that $\underline{\beta} = (\underline{\beta}_1', \underline{\beta}_2')'$, where $\underline{\beta}_1$ is $s \times 1$ and $\underline{\beta}_2$ is $u \times 1$ ($s+u = l$), and that the nested hypothesis $H : \underline{\beta}_2 = \underline{0}$ is of interest. Under H , the reduced model $M_2 : P = \underline{m}(\underline{v}, \underline{\beta}_1, \underline{0})$ applies.

An alternative specification of M_2 is

$$\ln(G_{j(i)} / (1 - G_{j(i)})) = v_j - \beta_1' Y_{1i}$$

where

$$Y_1' = (Y_{11}', Y_{12}')$$

By definition,

$$M = \left[\frac{\partial m}{\partial v}, \frac{\partial m}{\partial \beta_1}, \frac{\partial m}{\partial \beta_2} \right].$$

If $X = [y_1, y_2, \dots, y_R]'$ is of rank l and has the property that $X'1 = 0$, then $[\underline{1}_R \otimes I_J, A]$ is of rank $J+l$, and thus $\frac{\partial m}{\partial \beta_1}$ and $\frac{\partial m}{\partial \beta_2}$ are of full rank s and u respectively.

From Theorem 5.1, the Pearson chi-squared statistic χ^2 for testing $H: \beta_2 = 0$, given that model (5.48) applies, has asymptotic distribution $\sum_{i=1}^u \delta_i Z_i^2$ where Z_i^2 are χ_1^2 variables and δ_i are the eigenvalues of

$$(H_2' \nabla H_2)^{-1} (H_2' \nabla \Sigma \nabla H_2).$$

For the McCullagh and Nelder model, tedious calculations yield

$$(5.58) \quad H_2 = (I_R \otimes L^{-1}) G \bar{Y}_2,$$

where

$$(5.59) \quad \bar{Y}_2 = [I - Y_1(Y_1' B Y_1)^{-1} Y_1' B] Y_2,$$

$$(5.60) \quad Y_1 = [\underline{1}_R \otimes I_J, \begin{bmatrix} \underline{1}_J \otimes Y_{11} \\ \vdots \\ \underline{1}_J \otimes Y_{R1} \end{bmatrix}]'$$

$$(5.61) \quad Y_2 = \begin{bmatrix} \underline{1}_J \otimes Y_{12} \\ \vdots \\ \underline{1}_J \otimes Y_{R2} \end{bmatrix},$$

and B is a block diagonal matrix, with the i th $J \times J$ block being

$$(5.62) \quad W_1 D(G_1) D(1 - G_1) [LQ_1 L']^{-1} D(G_1) D(1 - G_1).$$

It then follows that the δ_i are the eigenvalues of

$$(5.63) \quad (\bar{Y}_2' B \bar{Y}_2)^{-1} (\bar{Y}_2' G(D(W) \otimes L^{-1}) Q^{-1} \Sigma Q^{-1} (D(W) \otimes L^{-1}) G \bar{Y}_2).$$

5.3.2 Haberman's multinomial response model.

As in Section 5.2, let $P_{j(i)}$ denote the proportion with the j th response in the i th domain, $j=1,2,\dots,J+1$, $i=1,2,\dots,R$. A form of model for the $P_{j(i)}$'s studied by Haberman (1982) and others is

$$(5.64) \quad P_{j(i)} = m_{ij}(\theta) = \exp(x_{ij}' v_j) / \sum_{\ell=1}^{J+1} \exp(x_{ij}' v_\ell) \quad i=1,2,\dots,R; j=1,2,\dots,J+1.$$

with

$$(5.65) \quad \sum_{\ell=1}^{J+1} v_\ell = 0,$$

where the x_i are known vectors of length q , $\theta = (v_1', v_2', \dots, v_J')$ is an unknown parameter of length qJ and $v_j = (v_{1j}, v_{2j}, \dots, v_{qj})'$. In order to apply the results of Section 5.2, the matrix $M = (\frac{\partial m}{\partial \theta})$ must be derived.

Because of constraint (5.65), model (5.64), may be expressed as

$$(5.66) \quad m_{ij}(\theta) = \exp(x_{ij}' v_j) / [\sum_{\ell=1}^J \exp(x_{ij}' v_\ell) + \prod_{\ell=1}^J \exp(-x_{ij}' v_\ell)] \quad \begin{matrix} i=1,2,\dots,R \\ j=1,2,\dots,J \end{matrix}$$

Using (5.66), the following partial derivatives may be derived.

$$(5.67) \quad \partial m_{ij} / \partial v_{kj} = m_{ij}(\theta) x_{ik} (1 - m_{ij}(\theta) + m_{i,J+1}(\theta))$$

and

$$(5.68) \quad \partial m_{ij} / \partial v_{kl} = m_{ij}(\underline{\theta}) x_{ik} (-m_{il}(\underline{\theta}) + m_{i,J+1}(\underline{\theta})) , \quad l \neq j.$$

By definition,

$$M = \left[\frac{\partial m_1}{\partial \underline{\theta}}, \frac{\partial m_2}{\partial \underline{\theta}}, \dots, \frac{\partial m_R}{\partial \underline{\theta}} \right],$$

where

$$\frac{\partial m_i}{\partial \underline{\theta}} = \left[\frac{\partial m_i}{\partial v_{-1}}, \frac{\partial m_i}{\partial v_{-2}}, \dots, \frac{\partial m_i}{\partial v_{-J}} \right]$$

is a $J \times qJ$ matrix.

By (5.67) and (5.68), $\frac{\partial m_i}{\partial \underline{\theta}}$ has the form

$$(5.69) \quad \frac{\partial m_i}{\partial \underline{\theta}} = [Q_i + (1 - \underline{1}_J' \underline{m}_i) (\underline{1}_J' \otimes \underline{m}_i)] \otimes \underline{x}_i'.$$

Substituting (5.69) into (5.13), the pseudo likelihood equations for the Haberman model are

$$(5.70) \quad \begin{aligned} & \sum_{i=1}^R w_i \{ [I_J + (1 - \underline{1}_J' \underline{m}_i) (\underline{1}_J' \otimes \underline{m}_i) \hat{Q}_i^{-1}] \otimes \underline{x}_i \} \hat{p}_i \\ &= \sum_{i=1}^R w_i \{ [(1 - \underline{1}_J' \underline{m}_i)^{-1} \hat{Q}_i' + (\underline{1}_J \otimes \underline{m}_i')] \otimes \underline{x}_i' \} \underline{1}_J. \end{aligned}$$

A possible nested hypothesis would conceivably consist of setting a fixed set of parameters in each v_j equal to 0. In order to apply the nested results of Section 5.2 it would be advisable to reorder $\underline{\theta}$ (and also M) so that the non-zero parameters appeared first, followed by the ones hypothesized to be zero.

CHAPTER 6

FURTHER RESEARCH

While it is hoped that the results of this research will be of use in the application of loglinear, binary response and polytomous response models to survey data, there are many areas in which further research could be undertaken profitably. Three such areas are outlined below.

Many of the results in the previous chapters are asymptotic, thus only known to hold true for very large samples. However, surveys and analyses of survey data are carried out with finite sample sizes. It would therefore be profitable to assess the finite sample performance of the test statistics proposed, through a simulation study. One possible approach, say for nested hypotheses with a logit model, is to simulate two-stage cluster sampling through the use of Dirichlet-multinomial variables, as done by Thomas and Rao (1984). This allows considerable controlled deviation from simple product binomial sampling.

Another area requiring further work is the development of better bounds for the correction δ to χ^2 . As noted in Chapter 3, when little information is available on the covariance matrix of the estimated survey proportions, the bounds proposed there could be excessively large for some models.

Finally, all of the test statistics proposed in the previous chapters use χ^2 critical points. Thomas and Rao (1984) found that, for small and moderate numbers of clusters, better control of actual

significance levels, in the case of goodness-of-fit with loglinear models, was obtained by referring suitably scaled statistics to an F distribution. A similar approach should be considered for binary response and polytomous response-models.

APPENDIX

A1 To show equivalence of models $G\underline{\pi} = H\underline{\theta}$ and $\underline{\pi} = F\underline{v}$, where G is $u \times T$ of rank u , H is $u \times v$ of rank v , $v \leq u \leq T-1$, F is $T \times s$ of rank s , where $s = T - u + v$, and $\underline{\pi}$ has the property $\underline{1}'\underline{\pi} = 1$.

Proof: Since G is $u \times T$ of rank u , a matrix G_1 exists of size $(T-u) \times T$ and rank $T-u$ with the property $G_1 G' = \underline{0}$. The matrix $C = [G', G_1']'$ is, thus, nonsingular and provides a 1-1 transformation between $\underline{\pi}$ and $\underline{\pi}^* = C\underline{\pi}$. Obviously, since $\underline{1}'\underline{\pi} = 1$, then, if $\pi_1^*, \pi_2^*, \dots, \pi_{T-1}^*$ are known, π_T^* is determined by $\underline{1}'C^{-1}\underline{\pi}^* = 1$. Let $\underline{\pi}_0^*$ be partitioned as $\underline{\pi}^* = (\underline{\pi}_1^*, \underline{\pi}_2^*, \pi_T^*) = (\underline{\pi}^*, \pi_T^*)'$, where $\underline{\pi}_1^*$ is $u \times 1$, $\underline{\pi}_2^*$ is $(T-u-1) \times 1$ and $\underline{\pi}^* = (\underline{\pi}_1^*, \underline{\pi}_2^*)'$. The model $G\underline{\pi} = H\underline{\theta}$ states that $\underline{\pi}_1^* = H\underline{\theta}$, where H is $u \times v$ of rank v . Since $\underline{\pi}_2^*$ is unrestricted under the model, let $\underline{\pi}_2^*$ be denoted by the $(T-u-1) \times 1$ vector $\underline{\delta}$. Then

$$\underline{\pi}^* = \begin{pmatrix} \underline{\pi}_1^* \\ \underline{\pi}_2^* \end{pmatrix} = \begin{pmatrix} H\underline{\theta} \\ \underline{\delta} \end{pmatrix} = \begin{pmatrix} H & \underline{0} \\ \underline{0} & I \end{pmatrix} \begin{pmatrix} \underline{\theta} \\ \underline{\delta} \end{pmatrix} = J \begin{pmatrix} \underline{\theta} \\ \underline{\delta} \end{pmatrix}.$$

Since H is $u \times v$ of rank v , and $(H', \underline{0})'$ is orthogonal to $(\underline{0}, I)'$, then J is a $(T-1) \times (T-u-1+v)$ matrix of rank $(T-u-1+v)$.

If $\underline{\beta}$ denotes $(\underline{\theta}', \underline{\delta}')'$, then

$$\underline{\pi}^* = \begin{pmatrix} \underline{\pi}^* \\ \pi_T^* \end{pmatrix} = \begin{pmatrix} J\underline{\beta} \\ \pi_T^* \end{pmatrix} = \begin{pmatrix} J & \underline{0} \\ \underline{0} & 1 \end{pmatrix} \begin{pmatrix} \underline{\beta} \\ \pi_T^* \end{pmatrix} = K \begin{pmatrix} \underline{\beta} \\ \pi_T^* \end{pmatrix} = K\underline{v}$$

where K is $T \times (T-u+v)$ of rank $T-u+v$, and $\underline{v} = (\underline{\beta}', \pi_T^*)'$.

Finally,

$$\underline{\pi} = C^{-1} \underline{\pi}^* = C^{-1} K \underline{v} = F \underline{v}$$

where $F = C^{-1} K$ is $T \times T-u+v$ of rank $s = T-u+v$.

A2. Let $\underline{f}(\hat{\underline{\theta}}) - \underline{f}(\hat{\underline{\theta}}_1) \approx F_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1) + F_2 \hat{\underline{\theta}}_2 - F_1(\hat{\underline{\theta}}_1 - \underline{\theta}_1)$
 under M_2 , where $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2)$ is the pseudo mle of $\underline{\theta}$ under
 the model $M_1 : \underline{\pi} = \underline{f}(\underline{\theta})$; $\hat{\underline{\theta}}_1$ is the pseudo mle of $\underline{\theta}_1$ under
 the nested model $M_2 : \underline{\theta}_2 = \underline{0}$, and $F_1 = \partial \underline{f} / \partial \underline{\theta}_1$. Also,
 suppose that, under M_2 ,

$$(A.1) \quad (\hat{\underline{\theta}} - \underline{\theta}) = (\hat{\underline{\theta}}_1 - \underline{\theta}_1, \hat{\underline{\theta}}_2)' \sim (F' A F)^{-1} F' A (\underline{p} - \underline{f}(\underline{\theta}))$$

and

$$(A.2) \quad (\hat{\underline{\theta}}_1 - \underline{\theta}_1) \sim (F_1' A F_1)^{-1} F_1' A (\underline{p} - \underline{f}(\underline{\theta})),$$

for some nonsingular matrix A , where $F = (F_1, F_2)$. Then, if
 $(\hat{\underline{\theta}}_1 - \underline{\theta}_1) \sim (\hat{\underline{\theta}}_1 - \underline{\theta}_1) + Q \hat{\underline{\theta}}_2$, the matrix Q has the form
 $Q = (F_1' A F_1)^{-1} (F_1' A F_2)$.

Proof: By Seber (1977), the inverse of a partitioned matrix is

$$(A.3) \quad \begin{pmatrix} C & B \\ B' & D \end{pmatrix}^{-1} = \begin{pmatrix} C^{-1} + G L^{-1} G' & -G L^{-1} \\ -L^{-1} G' & L^{-1} \end{pmatrix},$$

where $G = C^{-1} B$ and $L = D - B' C^{-1} B = D - B' G$. Suppose that

$C = (F_1' A F_1)$, $B = (F_1' A F_2)$ and $D = (F_2' A F_2)$. Then

$$G = (F_1' A F_1)^{-1} (F_1' A F_2) \text{ and}$$

$$L = (F_2' A F_2) - (F_2' A F_1) (F_1' A F_1)^{-1} (F_1' A F_2) = \bar{F}_2' A \bar{F}_2,$$

where $\bar{F}_2 = (I - F_1(F_1'AF_1)^{-1}F_1'A)F_2 = F_2 - F_1G$. If $F'AF$ is partitioned and result (A.3) used, then (A.1) may be written as

$$(A.4) \quad \begin{aligned} (\hat{\underline{\theta}} - \underline{\theta}) &\sim \begin{bmatrix} C^{-1} + GL^{-1}G' & -GL^{-1} \\ -L^{-1}G' & L^{-1} \end{bmatrix} \begin{bmatrix} F_1' \\ F_2' \end{bmatrix} A(\hat{\underline{p}} - \underline{f}(\underline{\theta})) \\ &= \begin{bmatrix} C^{-1}F_1' - GL^{-1}\bar{F}_2' \\ L^{-1}\bar{F}_2' \end{bmatrix} A(\hat{\underline{p}} - \underline{f}(\underline{\theta})). \end{aligned}$$

Thus,

$$(A.5) \quad (\hat{\underline{\theta}}_1 - \underline{\theta}_1) + Q\hat{\underline{\theta}}_2 \sim (C^{-1}F_1' - GL^{-1}\bar{F}_2' + QL^{-1}\bar{F}_2')A(\hat{\underline{p}} - \underline{f}(\underline{\theta})).$$

On the other hand, by (A.2),

$$(A.6) \quad (\hat{\underline{\theta}}_1 - \underline{\theta}_1) \sim C^{-1}F_1'A(\hat{\underline{p}} - \underline{f}(\underline{\theta})).$$

From examination of (A.5) and (A.6), for the condition $(\hat{\underline{\theta}}_1 - \underline{\theta}_1) + Q\hat{\underline{\theta}}_2 \sim (\hat{\underline{\theta}}_1 - \underline{\theta}_1)$ to hold, Q could take on the value

$$(A.6) \quad Q = G = (F_1'AF_1)^{-1}(F_1'AF_2').$$

A3. Let $(\hat{\underline{\theta}} - \underline{\theta})' = (\hat{\underline{\theta}}_1' - \underline{\theta}_1', \hat{\underline{\theta}}_2' - \underline{\theta}_2') \sim (F'AF)^{-1}F'A(\hat{\underline{p}} - \underline{f}(\underline{\theta}))$, for a nonsingular matrix A , where the asymptotic covariance matrix of $\hat{\underline{p}}$ is

$$(A.7) \quad C(\hat{\underline{p}}) = V/n$$

and F is defined in A2.

$$\text{Then } C(\hat{\underline{\theta}}_2) = n^{-1}(\bar{F}_2'A\bar{F}_2)^{-1}(\bar{F}_2'AVA'\bar{F}_2)(\bar{F}_2'AF_2)^{-1}$$

where $\bar{F}_2$ is defined in A2.

Proof: If $(F'AF)$ is partitioned and inverted as in A2, then, extracting the appropriate terms from (A.4),

$$(A.8) \quad (\hat{\theta}_2 - \underline{\theta}) \sim L^{-1} \bar{F}_2' A (\hat{p} - \underline{f}(\underline{\theta})) = (\bar{F}_2' A \bar{F}_2)^{-1} \bar{F}_2' A (\hat{p} - \underline{f}(\underline{\theta})) .$$

Then, from (A.7) and (A.8) may be deduced

$$\begin{aligned} C(\hat{\theta}_2) &= (\bar{F}_2' A \bar{F}_2)^{-1} \bar{F}_2' A C(\hat{p}) A' \bar{F}_2 (\bar{F}_2' A \bar{F}_2)^{-1} \\ &= n^{-1} (\bar{F}_2' A \bar{F}_2)^{-1} (\bar{F}_2' A V A' \bar{F}_2) (\bar{F}_2' A \bar{F}_2)^{-1} . \end{aligned}$$

A4. Let $X = (X_1, X_2)$ where X_1 is $T \times s$ of rank s and X_2 is $T \times u$ of rank u . Suppose that $F = (F_1, F_2)$ is defined as in A2, with F_1 being $T \times s$ and F_2 being $T \times u$. Also, suppose that W is any $T \times u$ matrix of rank u such that $W'X_1 = \underline{0}$ and $W'X_2$ is nonsingular. Then for any nonsingular matrix A ,

$$W'X(F'AF)^{-1}X'W = W'X_2(\bar{F}_2' A \bar{F}_2)^{-1}X_2'W ,$$

where $\bar{F}_2$ is defined in A2.

Proof: Because of the conditions on W ,

$$W'X = (W'X_1, W'X_2) = (\underline{0}, W'X_2) .$$

If $F'AF$ is partitioned as

$$(A.9) \quad F'AF = \begin{pmatrix} F_1' A F_1 & F_1' A F_2 \\ F_2' A F_1 & F_2' A F_2 \end{pmatrix} ,$$

and result (A.3) is used, then

$$W'X(F'AF)^{-1}X'W = (0, W'X_2)(F'AF)^{-1}(0, W'X_2)' = W'X_2 L^{-1} X_2' W,$$

where $L^{-1} = (\bar{F}_2' A \bar{F}_2)$.

A5. Let A be a nonsingular matrix, and let $F = (F_1, F_2)$ and $\bar{F}_2$ be defined as in A2. Then

$$\text{tr}(\bar{F}_2' A \bar{F}_2)^{-1} (\bar{F}_2' A V A \bar{F}_2) = \text{tr}(F' A F)^{-1} (F' A V A F) - \text{tr}(F_1' A F_1)^{-1} (F_1' A V A F_1).$$

Proof: If $F' A F$ is partitioned as in (A.9) and is inverted using result (A.3), and if $F' A V A F$ is partitioned as

$$F' A V A F = \begin{pmatrix} F_1' A V A F_1 & F_1' A V A F_2 \\ F_2' A V A F_1 & F_2' A V A F_2 \end{pmatrix},$$

then

$$\begin{aligned} & \text{tr}(F' A F)^{-1} (F' A V A F) \\ &= \text{tr} \begin{pmatrix} (C^{-1} + GL^{-1}G') (F_1' A V A F_1) - GL^{-1} (F_2' A V A F_1) & (C^{-1} + GL^{-1}G') (F_1' A V A F_2) - GL^{-1} (F_2' A V A F_2) \\ -L^{-1}G' (F_1' A V A F_1) + L^{-1} (F_2' A V A F_1) & -L^{-1}G' (F_2' A V A F_1) + L^{-1} (F_2' A V A F_2) \end{pmatrix} \\ &= \text{tr}[(C^{-1} + GL^{-1}G') (F_1' A V A F_1) - GL^{-1} (F_2' A V A F_1)] + \text{tr}[-L^{-1}G' (F_2' A V A F_1) + L^{-1} (F_2' A V A F_2)] \end{aligned}$$

Also,

$$\text{tr}(F_1' A F_1)^{-1} (F_1' A V A F_1) = \text{tr} C^{-1} (F_1' A V A F_1).$$

Thus,

$$\begin{aligned}
& \text{tr}(F'AF)^{-1}(F'AVAF) - \text{tr}(F_1'AF_1)^{-1}(F_1'AVAF_1) \\
(A.10) \quad & = \text{tr}(GL^{-1}G')(F_1'AVAF_1) - \text{tr} GL^{-1}(F_2'AVAF_1) \\
& \quad + \text{tr}(-L^{-1}G')(F_2'AVAF_1) + \text{tr} L^{-1}(F_2'AVAF_2) \\
& = \text{tr} L^{-1}[G'(F_1'AVAF_1)G - (F_2'AVAF_1)G - G'(F_2'AVAF_1) + (F_2'AVAF_2)].
\end{aligned}$$

On the other hand, using (A.3),

$$\begin{aligned}
& \text{tr}(\bar{F}_2'AF_2)^{-1}(\bar{F}_2'AVAF_2) = \text{tr} L^{-1}(\bar{F}_2'AVAF_2) \\
(A.11) \quad & = \text{tr} L^{-1}(F_2 - F_1G)'(AVA)(F_2 - F_1G)
\end{aligned}$$

When the right hand side of (A.11) is expanded, the expansion equals the right hand side of (A.10), so that the required result follows.

A6. Design Effect

The design effect of an estimate $\hat{p}$ is the ratio of the variance of $\hat{p}$ under the sampling plan actually used, to the variance of $\hat{p}$ under simple random sampling.

A7. Form of Pearson Chi-squared Statistic in Binary Response Models

For ease of illustration, consider product binomial sampling in each of k domains, with a sample of n_i units, y_i being successes (and $(n_i - y_i)$ being failures). Consider that, under the hypothesis being considered, the estimate of the number of successes in the i th domain is $n_i m_i(\hat{\theta})$ (and thus the estimate of the number of failures is $n_i(1 - m_i(\hat{\theta}))$). The Pearson chi-squared statistic is defined to be

$$x^2 = \sum_{i=1}^R (y_i - n_i m_i(\hat{\theta}))^2 / n_i m_i(\hat{\theta}) + \sum_{i=1}^R [(n_i - y_i) - n_i (1 - m_i(\hat{\theta}))]^2 / n_i (1 - m_i(\hat{\theta})),$$

which reduces to

$$x^2 = \sum_{i=1}^R (y_i - n_i m_i(\hat{\theta}))^2 / [n_i m_i(\hat{\theta}) (1 - m_i(\hat{\theta}))],$$

because of the fact that

$$a/nm + a/(n(1-m)) = a/(nm(1-m))$$

for any values of a , n and m ($\neq 0$). A similar reduction of terms occurs in the Pearson chi-squared statistic used for a general sampling plan (see (3.6)).

A8. Regularity conditions of Birch (1964)

Let $\pi_i = f_i(\underline{\theta}_0)$, $i=1,2,\dots,T$, and let $\underline{\theta}_0$ be an r -vector, where r is less than the number of independent π_i 's (which, for a loglinear model, implies $r < T-1$). The assumptions made by Birch (1964) are:

1. $\underline{\theta}_0$ is an interior point of the parameter space Θ , so that $\underline{\theta}_0$ is not on the boundary of Θ and there is an r -dimensional neighbourhood of $\underline{\theta}_0$ completely contained in Θ .

2. $\pi_i = f_i(\underline{\theta}_0) > 0$ for $i=1,2,\dots,T$.

3. Each mapping f_i is totally differentiable at $\underline{\theta}_0$, so that the partial derivatives of f_i with respect to each θ_j exist at $\underline{\theta}_0$ and $f_i(\underline{\theta})$ has a linear approximation at $\underline{\theta}_0$ given by $f_i(\underline{\theta}) =$

$$f_i(\underline{\theta}_0) + \sum_{j=1}^r (\theta_j - \theta_{0j}) \frac{\partial f_i(\underline{\theta}_0)}{\partial \theta_j} + o(\|\underline{\theta} - \underline{\theta}_0\|) \quad \text{as } \underline{\theta} \rightarrow \underline{\theta}_0.$$

4. The Jacobian matrix $\frac{\partial f}{\partial \theta}$ whose (i,j) th element is $\partial f_i(\underline{\theta}_0)/\partial \theta_j$, is of full rank r .
5. Each inverse mapping f_i^{-1} is continuous at $\underline{f}(\underline{\theta}_0) = \underline{\pi}_i$.
6. Each mapping f_i is continuous at every point $\underline{\theta}$ in Θ .

It should be noted that this is a minimal set of regularity conditions.

BIBLIOGRAPHY

- Ashton, W.D. (1972). *The Logit Transformation*. Griffins Statistical Monographs and Courses, #32.
- Birch, M.W. (1964). A new proof of the Pearson-Fisher theorem. *Ann. Math. Statist.*, 35, 817-824.
- Bishop, Y.M.M. (1969). Full contingency tables, logits, and split contingency tables. *Biometrics*, 25, 383-400.
- Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Mass.
- Bloch, F.E. and Smith, S.P. (1977). Human capital and labour market employment. *J. Human Resources*, 12, 550-559.
- Brier, S.E. (1978). *Categorical Data Models for Complex Data Structures*. Unpublished Ph.D. Dissertation, School of Statistics, University of Minnesota.
- Cook, R.D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall, London.
- Cox, D.R. (1970). *Analysis of Binary Data*. Chapman and Hall, London.
- Dobson, A.J. (1983). *Introduction to Statistical Modelling*. Chapman and Hall, London.
- Fellegi, I.P. (1980). Approximate tests of independence and goodness of fit based on stratified multistage samples. *J. Amer. Statist. Ass.*, 71, 665-70.

Fienberg, S.E. (1977). *The Analysis of Cross-Classified Data*. MIT Press, Cambridge, Mass.

Graybill, F.A. (1969). *Introduction to Matrices with Applications in Statistics*. Wadsworth, California.

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

Guerrero, V.M. and Johnson, R.A. (1982). Use of the Box-Cox transformation with binary response models. *Biometrika*, 69, 309-14.

Haberman, S.J. (1978). *Analysis of Qualitative Data. Volume 1: Introductory Topics*. Academic Press, New York.

Haberman, S.J. (1982). Analysis of dispersion of multinomial responses. *J. Amer. Statist. Ass.*, 77, 568-580.

Hidiroglou, M.A. and Rao, J.N.K. (1981). Chisquare tests for the analysis of categorical data from the Canada Health Survey. Paper presented at the Intl. Statist. Inst. Meetings, Buenos Aires, 1981.

Hidiroglou, M.A. and Rao, J.N.K. (1983). Chi-square tests for the analysis of three-way contingency tables from the Canada Health Survey. Technical Report, Statistics Canada.

Holt, D., Scott, A.J., and Ewings, P.O. (1980). Chi-squared tests with survey data. *J. Roy. Statist. Soc., Ser. A*, 143, 302-320.

Kalbfleisch, J.G. (1984). Aspects of categorical data analysis.

In: *Topics in Applied Statistics* (Y.P. Chaubey and T.D. Dwivedi,

eds.), Dept. of Mathematics, Concordia University, Montreal,
139-150.

Koch, G.G., Freeman, D.H. Jr., and Freeman, J.L. (1975). Strategies
in the multivariate analysis of data from complex surveys.
Intl. Statist. Rev., 43, 59-78.

McCullagh, P. and Nelder, J.A. (1983). *Generalized Linear Models*.
Chapman and Hall, London.

Pregibon, D. (1981). Logistic regression diagnostics. *Ann. Statist.*,
9, 705-724.

Rao, C.R. (1965). *Linear Statistical Inference and its Applications*.
Second Edition. Wiley, New York.

Rao, J.N.K. and Scott, A.J. (1981). The analysis of categorical data
from complex sample surveys: chi-squared tests for goodness of
fit and independence in two-way tables. *J. Amer. Statist. Ass.*,
76, 221-30.

Rao, J.N.K. and Scott, A.J. (1984). On chi-squared tests for multiway
contingency tables with cell proportions estimated from survey
data. *Ann. Statist.*, 12, 46-60.

Rao, J.N.K. and Scott, A.J. (1984a). On simple adjustments to chisquared
tests with survey data: log-linear and logit models. Unpublished
manuscript.

Satterthwaite, F.E. (1946). An approximate distribution of estimates
of variance components. *Biometrics*, 2, 110-114.

S
Scott, A.J. and Rao, J.N.K. (1981). Chi-squared tests for contingency tables with proportions estimated from survey data. In *Current Topics in Survey Sampling* (D. Krewski, R. Platek and J.N.K. Rao, eds.) 247-266.

Seber, G.A.F. (1977). *Linear Regression Analysis*. Wiley, New York.

Thomas, D.R. and Rao, J.N.K. (1984). A Monte-Carlo study of exact levels for chi-squared goodness-of-fit statistics under cluster sampling. Technical Report Series of the Laboratory for Research in Statistics and Probability, 35, Carleton University.

The Health of Canadians (1981). Report of the Canada Health Survey, Health and Welfare and Statistics Canada. Statistics Canada Catalogue 82-538E.

Widder, D.V. (1961). *Advanced Calculus*, Second Edition. Prentice-Hall, N.J.

END

28H10H85

FIN