THE ANALYSIS OF BINARY DATA WITH LARGE, UNBALANCED, AND INCOMPLETE CLUSTERS USING RATIO MEAN AND WEIGHTED REGRESSION METHODS

by

ELLEN SIM SNYDER

Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2118T

September 1993
THE ANALYSIS OF BINARY DATA WITH LARGE, UNBALANCED, AND INCOMPLETE CLUSTERS USING RATIO MEAN AND WEIGHTED REGRESSION METHODS

by

Ellen Sim Snyder

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

1993

Approved by:

[Signatures]

Advisor

[Signature]

Reader

[Signature]

Reader

[Signature]

Reader

[Signature]

Reader

[Signature]
ABSTRACT

ELLEN SIM SNYDER. The Analysis of Binary Data with Large, Unbalanced, and Incomplete Clusters Using Ratio Mean and Weighted Regression Methods. (Under the direction of Gary G. Koch.)

A goal of many health services research studies is to quantify the level of utilization of a particular treatment and identify factors associated with its use. Often, medical practices are randomly selected from several geographic regions and utilization of the treatment of interest is noted for all or a random sample of patients in each practice. Outcomes for patients of the same practice are not independent since receipt of a medical service is under the control of the practitioner. In addition, multiple sites or opportunities to receive the treatment which vary in number for each patient may occur. Outcomes for the same patient are also not independent. In this type of study, it is the medical practice which defines clusters of correlated observations; and, thus, large and unequally sized clusters of (possibly incomplete, multivariate) correlated binary outcomes occur.

A weighted regression method for the analysis of large clusters of (possibly incomplete, multivariate), correlated binary data from one- and two-stage cluster samples based on ratio means is presented. Categorical predictor variables at the cluster-, patient-, and subpatient-levels may be accommodated. The method is demonstrated using data from a health services research study in dentistry on the use of crowns versus less costly, three or more surface restoration procedures, and
compared with other regression methods for (clustered) binary data. The method is also extended to accommodate more general multivariate, ordinal, and continuous outcomes from one-stage cluster samples with large and unequally sized clusters. Joint estimation of the population fraction with the treatment of interest and a new measure of intracluster correlation for binary data from one-stage cluster samples is also proposed. A weighted regression procedure for joint modeling of the population fraction and measure of intracluster correlation for multiple strata is also demonstrated with data from the crown utilization study.
ACKNOWLEDGMENTS

I feel very fortunate to have had the opportunity to work with my advisor, Dr. Gary G. Koch. I would like to express my sincere appreciation for his guidance and friendship during the course of this dissertation, and admiration for his extensive expertise. I would also like to thank Gary for always fitting me into his schedule and for going out of his way to do so on many occasions despite numerous responsibilities and other commitments.

I also want to thank my committee members: Drs. Lawrence Kupper, Lisa LaVange, Gregory Samsa, Jane Weintraub, and Daniel Shugars, for their review of these chapters, and for their insights and comments. Special thanks to Dan Shugars for allowing me to use his crown utilization study data which motivated the topic for this research, and to Lisa LaVange and Babubhai Shah of the Research Triangle Institute for providing me with a demonstration copy of SUDAAN™ software and other assistance. I am sincerely grateful to the Veterans Administration of Durham, NC, for the financial support I received through their predoctoral fellowship program.

My most sincere appreciation goes to my husband, Jeff, for enduring a year of separation while I lived in Chapel Hill to work on this dissertation, and for his patience, encouragement, and love throughout. Many thanks to my classmate, Sandra Stinnett, for her friendship and for lifting my spirits on so many occasions. Last, but certainly not least, I wish to thank my parents, Michael and Helen Sim, for their love, encouragement, support, and generosity throughout my life and especially during the years I spent at UNC.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>1.2</td>
<td>Literature Overview</td>
<td>7</td>
</tr>
<tr>
<td>I</td>
<td>1.3</td>
<td>Likelihood Based Methods</td>
<td>7</td>
</tr>
<tr>
<td>I</td>
<td>1.4</td>
<td>The Generalized Estimating Equation Approach</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1.4.1</td>
<td>First Order Generalized Estimating Equations</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>1.4.2</td>
<td>Second Order Generalized Estimating Equations</td>
<td>11</td>
</tr>
<tr>
<td>I</td>
<td>1.5</td>
<td>Adjustment to Chi-Squared Tests</td>
<td>14</td>
</tr>
<tr>
<td>I</td>
<td>1.6</td>
<td>Pseudo-Maximum Likelihood Methods (Survey Data Regression)</td>
<td>15</td>
</tr>
<tr>
<td>I</td>
<td>1.7</td>
<td>Weighted Least Squares (WLS) Methods for Categorical Data</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.7.1</td>
<td>Contingency Table Formulation</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1.7.2</td>
<td>Case-Record Formulation and Linear Models for Means</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>1.7.3</td>
<td>The Linear Model for Functions of Means</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>1.7.4</td>
<td>WLS Methods for Longitudinal Clustered Categorical Data</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>1.7.5</td>
<td>WLS Methods for Univariate Binary Data from Nested Designs with Balanced Clusters</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>1.7.6</td>
<td>WLS Methods for Bivariate and R-Variate Binary Data from Nested Designs with Balanced Clusters</td>
<td>27</td>
</tr>
<tr>
<td>I</td>
<td>1.8</td>
<td>Direct Estimation Methods</td>
<td>27</td>
</tr>
<tr>
<td>I</td>
<td>1.9</td>
<td>Summary</td>
<td>30</td>
</tr>
<tr>
<td>I</td>
<td>1.10</td>
<td>Overview of Research</td>
<td>31</td>
</tr>
</tbody>
</table>
Chapter II: Ratio Mean Regression for the Analysis of Binary Data from a One-Stage Cluster Sample Using Weighted Regression Methods

2.1 Introduction ........................................................................................................ 33
2.2 Ratio Mean Definition ........................................................................................ 34
  2.2.1 Stratification of Ratio Means by Cluster-Level Covariates .................. 37
  2.2.2 Comparing Ratio Means Between Cluster-Level Subgroups ............... 40
2.3 Subgroups of Ratio Means Defined by Patient-Level Characteristics ......... 41
  2.3.1 Comparing Ratio Means Between Patient-Level Subgroups ............... 45
2.4 Subgroups of Ratio Means Defined by Subpatient-Level Characteristics .... 46
  2.4.1 Comparing Ratio Means Between Subpatient-Level Subgroups ......... 47
2.5 Subgroups of Ratio Means Defined by Cluster-Level, Patient-Level, and Subpatient-Level Characteristics ................................................................. 48
  2.5.1 Matrix Notation ......................................................................................... 50
2.6 Ratio Mean Regression ...................................................................................... 52
  2.6.1 A Linear Model for R ................................................................................ 52
  2.6.2 A Linear Model for Logit(R) ................................................................. 54
2.7 Example ............................................................................................................. 55
2.8 Summary ............................................................................................................ 58

Chapter III: Ratio Mean Regression Extended to Multicategory, Ordinal, and Continuous Outcomes from a One-Stage Cluster Sample .................................................................. 60

3.1 Introduction ........................................................................................................ 60
3.2 Multicategory Outcomes .................................................................................... 60
  3.2.1 Ratio Mean Definition ............................................................................. 61
3.2.2 Ratio Means Defined by Cluster-, Patient, and Subpatient-Level Characteristics............................................. 65

3.2.3 Ratio Mean Regression Model for Multicategory Outcomes......70

3.3 Ordinal Outcomes.................................................................. 76

3.3.1 Ratio Mean Definition............................................................. 76

3.3.2 Ratio Means Defined by Cluster-, Patient-, and Subpatient-Level Characteristics.................................................. 79

3.4 Continuous Outcomes................................................................. 84

3.4.1 Ratio Mean Definition............................................................. 84

3.4.2 Alternative Ratio Mean Definitions.......................................... 85

Chapter IV: Ratio Mean Regression for the Analysis of Binary Data from a Two-Stage Cluster Sample.................................................. 86

4.1 Introduction.............................................................................. 86

4.2 Review of Two-Stage Cluster Sampling: Sampling With Replacement at Stage One and Without Replacement at Stage Two............... 87

4.2.1 Notation.................................................................................. 87

4.2.2 Estimators and their Variances for the Overall Population Mean and Cluster-Specific Means................................................. 89

4.3 Ratio Mean Definition for a Two-Stage Cluster Sample: Sampling With Replacement at Stage One and Without Replacement at Stage Two.....95

4.3.1 General Case........................................................................... 95

4.3.2 Special Sampling Schemes..................................................... 99

4.4 Subgroups of Ratio Means Defined by Cluster-, Patient-, and Subpatient-Level Characteristics.................................................. 101

4.5 Ratio Mean Regression.................................................................. 104

4.5.1 A Linear Model for Logit(B)..................................................... 104
4.5.2 Example........................................................................105

4.6 Summary........................................................................108

Chapter V: A Ratio Mean Approach for Joint Estimation of the Overall Mean
And Intracluster Correlation for Binary Data from a One-Stage Cluster
Sample...........................................................................110

5.1 Introduction....................................................................110
5.2 Notation for Population Quantities.................................111
5.3 Partitioning Variability..................................................114
5.4 Measures of Intracluster Correlation...............................122
5.5 Estimation for One-Stage Cluster Samples.......................131
  5.5.1 General Case........................................................131
  5.5.2 Special Cases.......................................................145
  5.5.3 Weighted Regression Applications............................147
5.6 Examples.......................................................................148
  5.6.1 Example 1............................................................148
  5.6.2 Example 2............................................................150
  5.6.3 Example 3............................................................151

Chapter VI: Summary and Directions for Further Research...........154

6.1 Summary.......................................................................154
6.2 Directions for Further Research......................................157

References..........................................................................159
Tables...............................................................................164
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Data Layout for an ((s \times r)) Contingency Table</td>
</tr>
<tr>
<td>2.1</td>
<td>Observed Ratio Means and Standard Errors Calculated Separately for Each Subpopulation Defined by the Crossclassification of Region, Age, and Tooth Arch</td>
</tr>
<tr>
<td>2.2</td>
<td>(10^5 \times) Estimated Covariance Matrix for (R_\tau)</td>
</tr>
<tr>
<td>2.3</td>
<td>Parameter Estimates and Standard Errors from Ratio Mean, Survey, and Ordinary Logistic Regression Methods</td>
</tr>
<tr>
<td>2.4</td>
<td>Analysis of Variance for Ratio Mean Regression Model</td>
</tr>
<tr>
<td>2.5</td>
<td>Predicted Ratio Means and Standard Errors for Each Subpopulation Defined by the Crossclassification of Region, Age, and Tooth Arch, Generated from the Ratio Mean Regression Model</td>
</tr>
<tr>
<td>2.6</td>
<td>Adjusted Odds Ratios and 95% Confidence Intervals for Selected Factor-Level Comparisons, Generated from the Ratio Mean Regression Model</td>
</tr>
<tr>
<td>4.1</td>
<td>Observed Ratio Means and Standard Errors Calculated Separately for Each Subpopulation Defined by the Crossclassification of Region, Age, and Tooth Arch</td>
</tr>
<tr>
<td>4.2</td>
<td>(10^5 \times) Estimated Covariance Matrix for (R_\tau)</td>
</tr>
<tr>
<td>4.3</td>
<td>Parameter Estimates and Standard Errors from Ratio Mean, Survey, and Ordinary Logistic Regression Models</td>
</tr>
<tr>
<td>4.4</td>
<td>Analysis of Variance for Ratio Mean Regression Model</td>
</tr>
<tr>
<td>4.5</td>
<td>Predicted Ratio Means and Standard Errors for Each Subpopulation Defined by the Crossclassification of Region, Age, and Tooth Arch, Generated from the Ratio Mean Regression Model</td>
</tr>
<tr>
<td>4.6</td>
<td>Adjusted Odds Ratios and 95% Confidence Intervals for Selected Factor-Level Comparisons, Generated from the Ratio Mean Regression Model</td>
</tr>
<tr>
<td>5.1</td>
<td>Estimates of (\psi) and (\beta_{y/x}) for each Region</td>
</tr>
<tr>
<td>5.2</td>
<td>(10^5 \times) Covariance Matrix for Estimate of ((\psi_1, \beta_{y/x,1}, \ldots, \psi_4, \beta_{y/x,4}))</td>
</tr>
</tbody>
</table>
Parameter Estimates and Standard Errors from Weighted Regression Model for Fisher Z-Transformation of Intracluster Correlation

Analysis of Variance

Predicted Estimates and Standard Errors for \( \hat{\beta}_{y/x} \), by Region

Parameter Estimates and Standard Errors from Weighted Regression Model for Logit(\( \theta \)) and Fisher Z-Transformation of \( \hat{\beta}_{y/x} \)

Analysis of Variance

Predicted Estimates and Standard Errors for \( \theta \) and \( \hat{\beta}_{y/x} \), By Region
CHAPTER I
INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

The primary outcome of interest in many medical, dental, and other types of health science studies is often a univariate or multivariate binary characteristic. Furthermore, the study designs or data collection strategies employed frequently give rise to dependencies among the outcomes for subsets of observations with some common feature for their source (i.e., sharing a common environment). A subset of data is said to form a "cluster" when the observations in the subset have intraclass correlation and so are not independent of one another.

Clustering may arise in a variety of ways. For example, natural groupings of the study subjects may give rise to clustered data. In teratology studies, dams may be exposed to a potential carcinogen and their offspring subsequently examined for malformations. The offspring of one dam cannot be assumed to be independent relative to the totality of offspring and so form a cluster. Clustering may also arise in longitudinal studies. The repeated measurements of an outcome for one study subject over time cannot be assumed as independent and form a cluster. Clustering may also arise from a (multi-stage) cluster sampling process in which subsets of population members are selected together into the sample, but the outcome of interest is observed for each member of the subset. For example, the subsets may be households, and the
outcome of interest observed for each household member may be the occurrence of a cold in the previous week. The cold status responses of members of the same household have intraclass correlation in the sense of being more internally similar relative to cold status of other persons so that they cannot be assumed as independent and form a cluster. Likewise, medical practitioners may be randomly selected from a target population in the first stage of a cluster sample, and patients of those practitioners may be selected into the sample in stage two. The outcome of interest for patients of the same practitioner have intraclass correlation and so cannot be assumed as independent, particularly if the outcome is receipt of a medical service which is under the control of the practitioner, as is often the case in many health services research studies. These types of health services research studies are known as utilization or outcomes studies and are of interest for this work.

A health services research study in dentistry will be used to illustrate the issues and methodologic developments for this research. The dental study is concerned with the identification of factors which may be associated with using a crown instead of a less costly crown substitute for a tooth requiring a greater than or equal to three surface restoration. The extent to which providers use the same type of restoration procedure for their patients is also of interest. Using a database of national computerized insurance claims from May 1, 1990 to May 1, 1991 as the source population, the included dental practices were considered conceptually equivalent to a sample randomly selected from the geographic regions of interest. For those practices, patients having at least one restoration with greater than or equal to three surfaces were identified and included in the sample. The factors (or covariates) which may be associated with the use of a
crown versus a less costly crown substitute for a restoration with greater than or equal to three surfaces may occur at the level of the dental practice, patient, or tooth. Examples of practice-level characteristics of interest include: region of the country, urban versus rural practice, and single versus group dental practice. Examples of patient-level characteristics of interest include: age, gender, and level of insurance coverage. Examples of tooth-specific characteristics include: tooth type (i.e., 1st premolar, 2nd premolar, 1st molar, 2nd molar), tooth arch (i.e., maxillary versus mandibular), and date of service. The binary outcome of interest is restoration category (i.e., crown versus crown substitute) and is observed for each tooth with a greater than or equal to three surface restoration of each patient selected into the study.

For the dental study, some practitioners may tend to use one treatment alternative (e.g. a crown) more frequently than another (e.g. a crown substitute) because of practice preference so that restorations for individuals who are patients of the same dental practice may be highly correlated. Hence, restorations for patients from the same dental practice cannot be assumed as independent and form a cluster. Likewise, practitioners (and/or their patients) may tend to prefer the same procedure for multiple sites in a patient and, hence, the type of restorations for multiple sites of the same patient may also be highly correlated. Hence, there is the potential for multiple levels of clustering in the dental data; there may be clustering at both the practice- and patient-levels.

A central issue for the analysis of clustered data is accounting for the dependence among cluster subunits in the analysis. If the dependence among observations in the clusters is not negligible and the usual statistical methods are applied, then the significance level of tests may be inflated or diminished.
depending upon whether there is positive or negative correlation among observations in the clusters, and whether the variables to be tested are constant for the cluster as a whole or vary according to cluster subunits.

Point and interval estimation of the overall mean, probability, or rate of response, and the exploration of the effects of covariates on the overall mean response are among the primary goals for most analyses of clustered data. Some studies which involve clustered data may, in addition, focus on analyzing the dependence structure among (classes of) observations within clusters, and obtaining estimates of the degree of dependence among observations within (and between these classes in) the clusters. There may be dependence among the elements of the outcome vector for the same subject and also between the outcomes of different subjects in the same cluster as in the dental study. Statistical methods for the analysis of clustered data may, therefore, fall into one of two categories: 1) those which provide estimates of the overall mean response and, perhaps the effects of covariates on the mean response, while accounting for the dependence among observations in the clusters but manage dependence as a nuisance, or, 2) those which jointly estimate the overall mean response and the degree of dependence among (classes of) cluster subunits. Joint estimation is more efficient because it uses information about the dependence structure when producing estimates. Measures of the degree of dependence for binary outcomes may include intraclass correlation coefficients or the odds ratio for pairwise dependencies. There are advantages and disadvantages associated with using each of these measures.

The response of interest does not have to be a univariate random variable, but may be bivariate or multivariate. An example of a clustered
bivariate binary outcome can be constructed using the cold status of household members example by increasing the number of questions asked of each household member from one to two: 1) did the household member have a cold in the previous week? and, 2) did the household member also have a headache in the previous week? If more than two binary questions are asked, then there may be clustering among the multivariate binary outcomes. The outcomes observed for each study subject in the dental study may be viewed as an (incomplete) sixteen-variate vector of binary responses relative to the two premolars and two molars in each of the four quadrants of the mouth being eligible or not for a restoration with at least three surfaces.

Clusters may also be "unbalanced," that is, the number of observations in a cluster (i.e., size of the cluster), differs among the clusters. Unbalanced clusters may result from subsampling unequal numbers of observations from each cluster. Unbalanced clusters may also occur when there are randomly missing vector elements for a clustered multivariate outcome or if subjects differ in the number of relevant vector elements for the analysis. For the dental data, of the 16 posterior teeth (excluding wisdom teeth) which are crown eligible, only the premolars and molars with a greater than or equal to three surface restoration of each patient are relevant for the analysis and, hence, the cluster sizes will necessarily vary. Similarly, some diseases may present either unilaterally or bilaterally (e.g., eye disease) and, thus, patients may contribute either a univariate (binary) observation or a bivariate (binary) outcome vector if the outcome of interest is, say, whether or not regression of disease has occurred. In this case, the analytic method must be able to differentiate between observations which are not relevant because they are not diseased and observations for which
the outcome is no disease regression.

While statistical methods for analyzing clustered continuous outcome data with multivariate normal distributions have been widely available (e.g., random effects and mixed model ANOVA), only recently have methods for clustered binary and other discrete outcomes been proposed in the literature. Most of the proposed methods for clustered binary data are, however, not useful for the analysis of large clusters like those for the dental study. If ten patients were selected from each dental practice in the sample and each patient can contribute between 1 and 16 teeth, then the cluster sizes could, theoretically, range between 10 and 160. Clusters of observations of size greater than or equal to 10 are considered to be large relative to computational feasibility considerations for many statistical analysis methods for discrete clustered data (e.g., generalized estimating equation methods (Zeger and Liang, 1986) require the number of clusters to be large but the cluster sizes to be relatively small). A considerable amount of variation among cluster sizes also affects efficiency and computational feasibility and is another difficulty to address.

This research will focus on statistical methods for analyzing multivariate binary outcomes from a cluster sample with large, unbalanced, and incomplete clusters. Some related consideration is additionally given to multicategory, ordinal, and continuous outcomes. It will be motivated largely by the data structure for the dental study. The methods developed in this research will address estimation of the overall mean response and the effects of covariates on the overall mean response while accounting for the clustering, and also methods which focus on the joint estimation of the overall mean response, effects of covariates, and measures of the dependence among observations in a cluster. The
remainder of this chapter will review existing statistical methods for the analysis of clustered binary data and will assess the extent to which these methods can be used for analysis of data which have the complex structure described for the dental study. Some extensions of existing methodology and new methods will then be proposed.

1.2 Literature Overview

The principal computational strategies for the analysis of univariate and multivariate binary data from a cluster sample are: 1) maximum likelihood relative to an assumed parametric distribution, 2) the generalized estimating equation (GEE) methods of Zeger and Liang (1986) which relax the assumptions about the underlying probability distribution, 3) adjustments to $\chi^2$ tests, 4) pseudo-maximum likelihood methods, 5) weighted least squares (WLS) relative to an assumed sampling framework, and 6) direct computation of estimators for the overall mean, standard error, and their covariance matrix relative to an assumed sampling framework.

1.3 Likelihood Based Methods

Many models have been proposed for the analysis of clustered categorical data which use maximum likelihood for estimation. An annotated bibliography of methods for analyzing correlated categorical data by Ashby, Neuhaus, Hauck et al. (1992) reviews many of these. The principal limitations of maximum likelihood based approaches for clustered categorical data include the required assumption of an underlying parametric distribution for the data which may not, in fact, be correct for the data (e.g., beta-binomial distribution) and, for some
models, computational tractability. The principal advantages of the maximum likelihood approach are that smaller samples may be analyzed with this approach relative to other approaches (e.g., weighted least squares), and parameter estimates and test statistics have desirable properties.

Methods which assume specific models for the intracluster correlations have been proposed. These include the beta-binomial model (Skellam, 1948; Williams, 1975; Crowder, 1978) and the two-parameter correlated binomial model (Kupper and Haseman, 1978). These methods can only accommodate cluster-level covariates.

Unconditional and conditional logistic regression with fixed effects for the clusters can also be used for the analysis of clustered binary data. Depending upon the sampling procedure for the data, the cluster effects may also be considered as random (Conaway, 1990). Unlike regression parameters from linear regression models based on normally distributed data, however, regression parameters for nonlinear categorical data models with fixed or random cluster effects have a cluster-specific rather than a population-averaged interpretation (Zeger, Liang, and Albert, 1988; Neuhaus, Kalbfleisch, and Hauck, 1991). Models with parameters which have population-averaged and not cluster-specific interpretations will be those of interest for the dental data example.

Log-linear models have also been used for the analysis of multivariate clustered categorical data (Rosner, 1984, 1989). Parameters in these models have a conditional rather than marginal interpretation; the parameters characterize probabilities which are conditional on particular values for all other elements of a multivariate response vector. This may be especially problematic when the multivariate outcomes are incomplete; the parameters from a log-linear model
have a different interpretation if one of the elements of a multivariate vector is missing than if the vector is complete. This type of model would be completely impractical for analyzing the dental data and other types of health services research studies which may give rise to large and incomplete clusters. In addition, conditioning on the outcomes for other elements of a multivariate vector may mask the effects of certain covariates which may be confounded with these other outcomes. The transitional (conditional) regressive logistic model proposed by Bonney (1987) also suffers from similar parameter interpretation problems. Conditional and transitional models may be most appropriate for genetic studies of familial aggregation of disease. Models for the marginal expectation of a multivariate outcome will be more useful for investigations such as the dental study.

Zhao and Prentice (1989) identified a class of quadratic exponential models useful for analyzing multivariate binary data when the number of observations for each subject is small.

1.4 The Generalized Estimating Equation Approach

The generalized estimating equation approach of Zeger and Liang (1986) and Liang and Zeger (1986) is becoming a popular approach for regression analysis of longitudinal and multivariate discrete outcomes. This method is a multivariate extension of quasi-likelihood analysis which was originally proposed by Wedderburn (1974). Quasi-likelihood for independent univariate outcomes, is an extension of the generalized linear models approach of McCullagh and Nelder (1989) for exponential family outcomes. The quasi-likelihood approach relaxes the full parametric distributional assumption of the response variable; only the
relationship between a function of the expected value of the response and the linear predictor (i.e. link function) and, the relationship between the variance and the mean up to a scale parameter, \( \phi \), needs to be specified. A system of score-like equations based on the (log) quasi-likelihood is then solved by an iteratively reweighted least squares algorithm. The resulting estimates of beta coefficients and standard errors from quasi-likelihood models are consistent and have asymptotically normal distributions (McCullagh, 1983).

1.4.1 **First Order Generalized Estimating Equations**

The generalized estimating equation method proposed by Zeger and Liang (1986) is a multivariate extension of quasi-likelihood analysis; the \( i \)-th subject or observational unit is allowed to contribute a \((d_i \times 1)\) vector of responses, \( \mathbf{y}_i = (y_{i1}, \ldots, y_{id_i})' \), where \( i = 1, \ldots, N \) indexes \( N \) clusters and \( j = 1, \ldots, d_i \) indexes the \( d_i \) responses for the \( i \)-th cluster. A link function which relates the marginal expectation of \( y_{ij} \), \( E(y_{ij}) = \mu_{ij} \), to the linear predictor, \( \mathbf{X}_{ij} \beta \), is assumed, where \( \mathbf{X}_{ij} \) is a \((1 \times p)\) vector of covariates for response \( j \) of cluster \( i \), and \( \beta \) is a \((p \times 1)\) vector of parameters to be estimated. A \((d_i \times d_i)\) working correlation matrix, \( \mathbf{R}_i(\phi) \), is specified for each \( y_i \), since the elements of \( y_i \) cannot be assumed as independent. \( \mathbf{R}_i(\phi) \) is assumed to be a function of a vector of unknown parameters, \( \phi \). The working covariance matrix for \( y_i \) is given by

\[
\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i(\phi) \mathbf{A}_i^{1/2} / \phi , \tag{1.1}
\]

where \( \mathbf{A}_i \) is a \((d_i \times d_i)\) diagonal matrix with \( v(\mu_{ij}) \) as the \( j \)-th diagonal element, where \( v(\cdot) \) is the variance function, and \( \mu_{ij} \) is the expected value of the \( j \)-th
element of $y_i$. The estimating equations for $\beta$ are given by

$$
\sum_{i=1}^{N} D_i' \mathcal{Y}_i^{-1} (y_i - \mu_i) = 0,
$$

(1.2)

where $D_i = \frac{\partial \mu_i}{\partial \beta}$ is a $(d_i \times p)$ matrix of first derivatives, and $\mu_i = (\mu_{i1}, ..., \mu_{id_i})'$. These equations are solved using an iteratively reweighted least squares algorithm consisting of two steps: 1) method of moments estimation for the elements of $R_i(\varphi)$ and the scale parameter $\varphi$ and, 2) a modified Fisher scoring procedure to update successive estimates $\hat{\beta}$ of $\beta$. Liang and Zeger (1986) showed that when the link function is correctly specified, the estimates for $\beta$ and $\mathcal{Y}_i \beta$ are consistent, even if the working correlation is misspecified and that $N^{1/2}(\hat{\beta} - \beta)$ is asymptotically multivariate normal with expected value $0$ and covariance matrix estimated by

$$
\mathcal{Y}_i \beta = \hat{\mathcal{Y}}_1^{-1} \hat{\mathcal{Y}}_0 \hat{\mathcal{Y}}_1^{-1},
$$

(1.3)

where

$$
\hat{\mathcal{Y}}_0 = \sum_{i=1}^{N} D_i' \hat{\mathcal{Y}}_i^{-1} (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)' \hat{\mathcal{Y}}_i^{-1} D_i
$$

(1.4)

and

$$
\hat{\mathcal{Y}}_1 = \sum_{i=1}^{N} D_i' \hat{\mathcal{Y}}_i^{-1} D_i.
$$

(1.5)

While choosing the correct working correlation matrix is not necessary for the consistency and multivariate normality of $\hat{\beta}$, it does, however, increase the efficiency of $\hat{\beta}$. Liang, Zeger, and Qaqish (1992) showed that the estimates for $\varphi$ using this approach can be quite inefficient.

1.4.2 Second Order Generalized Estimating Equations
Prentice (1988) extended the generalized estimating equation approach of Liang and Zeger (1986) to permit joint inference on the regression parameters, $\beta$, and pairwise correlations, $\varrho$. Lipsitz (1989) proposed using the odds ratio as the association parameter rather than the correlation for multivariate binary outcomes. Liang, Zeger, and Qaqish (1992) described a class of models for the marginal expectations and pairwise odds ratios for a multivariate binary vector of outcomes as a function of covariates, and used the estimation procedure proposed by Prentice (1988), which they refer to as second order generalized estimating equations, or, GEE2. Liang, Zeger, and Qaqish (1992) also compared the efficiency and robustness of the parameter estimates for $\hat{\beta}$ and $\varrho$ using GEE2, to those obtained using the procedure of Zeger and Liang (1986), which they refer to as GEE1, and relative to the maximum likelihood estimates for clusters of size 4 for six different models. Using simulated data, they demonstrated that estimation for $\hat{\beta}$ using either GEE1 or GEE2 was highly efficient relative to the maximum likelihood estimates. In addition, they showed that estimation for $\varrho$ using GEE2 was also highly efficient, however, estimation for $\varrho$ was very inefficient using GEE1. They concluded that GEE2 should be used for estimation when the estimation of $\varrho$ is of primary interest, otherwise, GEE1 can be used. However, the consistency of $\hat{\beta}$ when using GEE2 depends on the correct specification of the model for the pairwise odds ratios in addition to the GEE1 requirement of correctly specifying the model for the marginal mean.

Qaqish and Liang (1992) showed that proper specification of design matrices for the model proposed by Liang, Zeger, and Qaqish (1992) allows one to model data with multiple levels of nesting or multiple classes of observations. Their model, in theory, applies to the dental data, where teeth are nested within
patients, and patients are nested within dental practices.

The model proposed by Liang, Zeger, and Qaqish (1992) for multivariate binary data is described using notation similar to that used by Qaqish and Liang (1992) as follows. Let \( y_i = (y_{i1}, \ldots, y_{id_i})' \) be a \((d_i \times 1)\) vector of binary responses and \( w_i = (y_{i1}y_{i2}, y_{i1}y_{i3}, \ldots, y_{i(d_i-1)}y_{id_i})' \) be a \((d_i-1) \times 1\) vector of products of the elements from the i-th cluster. Let \( \psi_i = (y_i', w_i)' \) be an extended response vector.

Let \( \mu_i^{(1)} = (\mu_{i1}, \ldots, \mu_{id_i})' = E(y_i) \) be the marginal expectation for \( y_i \) and \( \mu_i^{(2)} = (\mu_{i12}, \mu_{i13}, \ldots, \mu_{i(d_i-1)d_i})' = E(w_i) \) be the marginal expectation for \( w_i \), and \( \mu_i = (\mu_i^{(1)}, \mu_i^{(2)})' \). Let \( \eta_{ij} = \mathcal{X}_{ij}\beta \) and \( \eta_{ijj} = \mathcal{Z}_{ijj}\varphi \) be linear predictors, where \( \varphi \) is a \((p \times 1)\) vector of unknown parameters, \( \mathcal{X}_{ij} \) is a \((1 \times p)\) vector of covariates for the j-th outcome in the i-th cluster, \( \varphi \) is a \((q \times 1)\) vector of unknown parameters, and \( \mathcal{Z}_{ijj} \) is a \((1 \times q)\) vector of covariates for the product of the j-th and \( j' \)-th outcomes in the i-th cluster. Let \( \varphi = (\varphi', \varphi')' \). Let \( \eta_i^{(1)} = (\eta_{i1}, \ldots, \eta_{id_i})' \) and \( \eta_i^{(2)} = (\eta_{i12}, \eta_{i13}, \ldots, \eta_{i(d_i-1)d_i})' \). Assume \( h_1(\cdot) \), the link function which relates \( \mu_i^{(1)} \) to \( \eta_i^{(1)} \) and, \( h_2(\cdot) \), the link function which relates \( \mu_i^{(2)} \) to \( \eta_i^{(2)} \), are known, and that they are the logit and log-odds-ratio link functions, respectively. Let \( \psi_{ijj} \) represent the pairwise odds ratio for \( y_{ij} \) and \( y_{ij'}, j = 1, \ldots, d_i, 1 \leq j < j' \leq d_i, \)

\[
\ln \psi_{jj'} = \frac{\mu_{ijj}(1 - \mu_{ij} - \mu_{ij'} + \mu_{ijj})}{(\mu_{ij} - 1)(\mu_{ij'} - 1)},
\]

and

\[
E[y_{ij}y_{ij'}] = \mu_{ij} \mu_{ij'} \text{ if } \psi_{ijj} = 1, \text{ and }
\]

\[
= 1 - (\mu_{ij} + \mu_{ij'})(1 - \psi_{ijj}) - \left\{ [1 - (\mu_{ij} + \mu_{ij'})(1 - \psi_{ijj}) - 4(\psi_{ijj} - 1)\mu_{ijj}\mu_{ij}{\mu_{ij'}}{1/2}] / 2(\psi_{ijj} - 1) \right\} \text{ if } \psi_{ijj} \neq 1.
\]

(1.6)
Then the full model is written as

$$
\eta = \begin{bmatrix}
\eta_1^{(1)} \\
\eta_2^{(2)}
\end{bmatrix} = h(\mu) = 
\begin{bmatrix}
h_1(\mu^{(1)}) \\
h_2(\mu^{(1)}, \mu^{(2)})
\end{bmatrix} = 
\begin{bmatrix}
X_\delta & 0 \\
0 & Z_\delta
\end{bmatrix}
\begin{bmatrix}
\delta \\
\theta
\end{bmatrix}
= Q \xi, \quad (1.8)
$$

where $\eta_1^{(1)} = (\eta_1^{(1)}_1, \ldots, \eta_N^{(1)}_1)'$, $\eta_2^{(2)} = (\eta_1^{(2)}_1, \ldots, \eta_N^{(2)}_1)'$, $\mu^{(1)} = (\mu_1^{(1)}, \ldots, \mu_N^{(1)})'$, and $\mu^{(2)} = (\mu_1^{(2)}, \ldots, \mu_N^{(2)})'$. Estimation of $\delta$ is carried out using GEE1 for the extended response vector, $y_*^{(1)}$. The estimating equation for $\delta$ is given by

$$
U(\delta) = \sum_{i=1}^N \mathcal{D}_i' \mathcal{V}_{i}^{-1} (y_{i}^* - \mu_i) = 0, \quad (1.9)
$$

where $\mathcal{D}_i = \frac{\partial(h(\mu^{(1)}, \mu^{(2)}))}{\partial \delta}$, and $\mathcal{V}_{i}$ is the working covariance matrix of $y_{i}^*$.

1.5 Adjustment to Chi-Squared Tests

Rao and Scott (1992) proposed a simple method for comparing groups of clustered binary data with group-specific covariates in which no model for the intraclass correlation was assumed. It is based on the concept of a design effect which is frequently used in traditional survey methodology. The design effect for an estimator is defined as the variance which takes into account the sample
design divided by the variance calculated as if the data were from a simple random sample. Simple adjustments to the data based on an estimate of the design effect are made prior to performing standard $\chi^2$ tests. Donald and Donner (1988), Donald and Banting (1988), and Donner (1989) discussed a method of adjusting the standard $\chi^2$ test for comparing groups of subjects with respect to overall prevalence of a certain characteristic when multiple correlated binary observations are made on each subject.

1.6 Pseudo-Maximum Likelihood Methods (Survey Data Regression)

Binder (1983) proposed a method for estimating the asymptotic covariance matrix of regression parameters within the class of generalized linear models (e.g., logistic regression) from data sampled from finite populations according to a complex survey design. Using a Taylor series linearization of the sample score equations, the asymptotic variance matrix of the estimators is obtained. This variance matrix is equivalent to the robust variance estimator proposed by Zeger and Liang (1986) with an identity working correlation matrix.

1.7 Weighted Least Squares (WLS) Methods for Categorical Data

The weighted least squares methodology of Grizzle, Starmer, and Koch (1969) allows the variation in a vector of estimates derived from categorical data to be modeled as a linear function of covariates. The estimates may be a vector of proportions, means, or other functions of the data, for which an essentially known covariance matrix is available. The main advantage of the weighted least squares method is that a homogeneous covariance matrix assumption across subpopulations based on the covariates is not required. Koch, Gillings, and
Stokes (1980) review the framework for inference for weighted least squares methods as compared with maximum likelihood and randomization based methods. Large sample sizes are required so that the functions being modeled have a multivariate normal distribution and so that consistency of the associated covariance matrix is ensured. The method is more generally described and extended in Koch and Imrey et al (1985). Weighted least squares methods have been proposed for the analysis of longitudinal studies (Koch, Landis, Freeman, et al, 1977), and for longitudinal studies with missing data (Stanish, Gillings, and Koch, 1978). Weighted least squares methodology for categorical data will be described in this section for the strictly linear model and the functional linear model for product multinomial data, and also for case record data for linear models for means and functional linear models for means, since these methods will be referred to frequently in later chapters.

1.7.1 Contingency Table Formulation

Let the subpopulations formed by the crossclassification of categorical explanatory variables for the population be indexed by \( h = 1, 2, ..., s \), and the elements of a categorical response vector be indexed by \( i = 1, 2, ..., r \). The samples of size \( n_{hi} = \sum_{i=1}^{r} n_{hl} \) are independently chosen by a process which is conceptually equivalent to simple random sampling from each of the \( s \) subpopulations. The data layout is summarized in an \( s \times r \) contingency table format in Table 1.1.

Let \( \mathbf{n}_h = (n_{h1}, n_{h2}, ..., n_{hr})' \) denote the \( r \times 1 \) vector of sample frequencies \( n_{hl} \) of the \( i \)-th response category among subjects from the \( h \)-th subpopulation. The vector \( \mathbf{n}_h \) follows a multinomial distribution with parameters \( n \) and
\[ z_h = (\pi_{h1}, \pi_{h2}, ..., \pi_{hr})', \text{ where } \pi_{hl} \text{ represents the probability that a randomly} \\
\text{selected subject from the } h\text{-th subpopulation is classified into the } l\text{-th response} \\
category and } n_h. \text{ is the size of the sample for subpopulation } h. \text{ The likelihood} \\
function for the data is} \\

\[ L = \prod_{h=1}^{s} \left\{ n_h ! \prod_{l=1}^{r} \frac{\pi_{hl}^{n_{hl}}}{n_{hl}!} \right\}, \quad (1.10) \]

\[ \text{subject to the constraint: } \sum_{l=1}^{r} \pi_{hl} = 1, \text{ for } h = 1, 2, ..., s. \text{ The sample} \\
\text{proportions, } p_{hl} = \frac{n_{hl}}{n_h}, \text{ } h = 1, ..., s; l = 1, ..., r, \text{ are unbiased estimators for the} \\
\pi_{hl} \text{ and the elements of the covariance matrix of } p_h = (p_{h1}, ..., p_{hr}) \text{ are given by} \\

\[ \text{Var}(p_{hl}) = \pi_{hl} \left(1 - \pi_{hl}\right) / n_h. \quad (1.11) \]

\[ \text{Cov}(p_{hl}, p_{hl'}) = -\pi_{hl} \pi_{hl'} / n_h, \quad l = l' \quad (1.12) \]

\[ \text{and } \ \text{Cov}(p_{hl}, p_{hl'}) = 0, \quad h \neq h'. \quad (1.13) \]

\[ \text{In matrix notation, let } \nabla_h(z_h) = \{(D_{z_h} - z_h \pi_h) / n_h \} \text{ denote the covariance} \\
\text{matrix for the } h\text{-th subpopulation where } D_{z_h} \text{ is a diagonal matrix with diagonal} \\
elements } z_h, \text{ and let } z = (z'_1, ..., z'_s)' \text{ denote the compound vector of parameters for} \\
\text{all subpopulations, and } p = (p'_1, ..., p'_s)' \text{ is the compound vector of sample} \\
\text{proportions.} \]

\[ \nabla(p) = z \quad (1.14) \]

\[ \text{and } \nabla \nabla(p) = D \text{ag} \{ \nabla_h(z_h) \} \quad (1.15) \]
where $\mathcal{D}_{\text{diag}}\{ \cdot \}$ denotes a block diagonal matrix with blocks $\mathcal{V}_h(\pi_h)$, $h = 1, \ldots, H$. If $\mathbf{p}$ is substituted for $\pi$ in $\mathcal{V}(\pi)$, then the resulting matrix $\mathcal{V}(\mathbf{p})$ is a consistent estimator of $\mathcal{V}(\pi)$.

A vector of functions $\mathbf{E} = \mathbf{E}(\mathbf{p}) = (F_1, \ldots, F_u)'$ obtained from linear, logarithmic, or exponential transformations of the sample proportions may be constructed, where $u \leq s(r - 1)$. A linearized Taylor series estimate of $\mathcal{V}_\mathbf{E} = \mathcal{V}_\mathbf{E}(\mathbf{p})$ may then be obtained as $\mathcal{H}(\mathbf{p})^T \mathcal{V}_\mathbf{E}(\mathbf{p}) \mathcal{H}(\mathbf{p})'$, where

$$\mathcal{H}(\mathbf{p}) = \left[ \frac{\partial \mathbf{E}(\mathbf{z})}{\partial \mathbf{z}} \mid \mathbf{z} = \mathbf{p} \right].$$

For linear functions, $\mathcal{A} \mathbf{z}$, $\mathcal{H} = \mathcal{A}$, for logarithmic functions, $\log(\mathbf{z})$, $\mathcal{H} = \mathcal{D}_{\mathbf{z}}^{-1}$, and for exponential functions $\exp(\mathbf{z})$, $\mathcal{H} = \mathcal{D}_{\exp(\mathbf{z})}$. A general class of linear models for describing the variation in $\mathbf{E}$ may be written as

$$\mathbf{E}_A(\mathbf{E}) = \mathbb{E}_A[\mathbf{E}(\mathbf{p})] = \mathbf{E}(\pi) = \mathbf{X}\hat{\beta}$$

(1.16)

where $\mathbb{E}_A(\cdot)$ denotes asymptotic expected value, and $\mathbf{X}$ is a $u \times t$ model specification matrix and $\hat{\beta}$ is a $t \times 1$ vector of unknown parameters to be estimated. The weighted least squares (WLS) estimates $\mathbf{b}$ of $\hat{\beta}$ and its estimated covariance matrix $\mathcal{V}_b$ of $\text{Var}(\mathbf{b})$ are given by

$$\mathbf{b} = (\mathbf{X}^T \mathcal{V}_\mathbf{E}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathcal{V}_\mathbf{E}^{-1} \mathbf{E}$$

(1.17)

and

$$\mathcal{V}_b = (\mathbf{X}^T \mathcal{V}_\mathbf{E}^{-1} \mathbf{X})^{-1},$$

(1.18)
As shown in Koch and Imrey et al. (1985), \( \hat{b} \) asymptotically has the multivariate normal distribution with mean vector

\[
E_A(\hat{b}) = \beta
\]

and variance-covariance matrix \( Var_A(\hat{b}) \) for which \( V_{\hat{b}} \) is a consistent estimator. A goodness of fit statistic for assessing the appropriates of the model is

\[
Q_w = (\hat{E} - \hat{X}\hat{b})' V_{\hat{E}}^{-1}(\hat{E} - \hat{X}\hat{b}).
\]

This statistic has an approximate \( \chi^2 \) distribution with d.f. = (u - t) when the subpopulations are moderately large (say, \( n_h \geq 20 \)). When model (1.16) provides a satisfactory description of the variation among the elements of \( E(\beta) \), tests of linear hypotheses \( H_0: \beta \sim 0 \), where \( \beta \) is a known \( c \times t \) matrix of full rank, are applicable through the Wald statistic

\[
Q_C = (\hat{C}\hat{b})'[\hat{C}V_{\hat{b}}\hat{C}]^{-1}(\hat{C}\hat{b}).
\]

Since \( \hat{C}\hat{b} \) approximately has a multivariate normal distribution with \( E_A(\hat{C}\hat{b}) = \hat{C} \beta \) and \( \hat{C}V_{\hat{b}}\hat{C}' \) is a consistent estimator of \( Var_A(\hat{C}\hat{b}) \), then \( Q_C \) approximately has a \( \chi^2 \) distribution with d.f. = c under \( H_0 \). Predicted values \( \hat{E} = \hat{X}\hat{b} \) and their asymptotic covariance matrix \( V_\hat{E} = \hat{X}V_{\hat{b}}\hat{X}' \) can also be obtained. The predicted values have smaller variance than their original counterparts in \( E \) since their estimation makes use of all of the data.
1.7.2 Case-Record Formulation and Linear Models for Means

While the previous section presented the weighted least squares method in terms of an underlying product-multinomial distribution for sets of counts, the weighted least squares method may also be formulated in terms of functions of means $F_k$ and their estimated covariance matrix. Let $h = 1, \ldots, s$ index a set of distinct subpopulations from which samples with sizes $\{n_h\}$ have been independently obtained by a process conceptually equivalent to simple random sampling. Let $i = 1, \ldots, n_h$ index subjects in the sample from the $h$-th subpopulation. Also let $y_{hik} = (y_{hi1}, y_{hi2}, \ldots, y_{hid})'$ denote the case-record vector of responses according to $d$ categorical variables for which the $k$-th has possible outcomes $l = 1, \ldots, L_k$. Let $a_k = (a_{k1}, \ldots, a_{KL_k})'$ denote a set of finite scores for the outcomes of the $k$-th response. Then, the sample mean for the $k$-th response among subjects from the $h$-th sample is given by

$$
\bar{y}_{hk} = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hki},
$$

(1.22)

and it follows that $\bar{y}_h = (\bar{y}_{h1}, \ldots, \bar{y}_{hd})'$ has

$$
E\{\bar{y}_h\} = \mu_h, \quad \text{and} \quad \text{Var}\{\bar{y}_h\} = \frac{1}{n_h} \sum_h
$$

(1.23)

where $\mu_h = (\mu_{h1}, \mu_{h2}, \ldots, \mu_{hd})'$ is the vector of expected values for the $d$ response variables for the $h$-th subpopulation. The covariance matrix $\sum_h$ for the case records $\{y_{hi}\}$ can be consistently estimated by

$$
\sum_h = \frac{1}{n_h} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)(y_{hi} - \bar{y}_h)'.
$$

(1.24)
The conceptual similarity between the sample means \( \{ \bar{y}_h \} \) and the function of the multinomially distributed counts \( \{ n_{hl} \} \) in the previous section point to the analogous use of weighted least squares methods to fit linear models to describe the variation among the \( \{ \mu_{hk} \} \).

Let \( \bar{y} = (\bar{y}_1', \bar{y}_2', ..., \bar{y}_s')' \) denote the compound vector of sample means for all \( s \) subpopulations and let \( \mu = (\mu'_1, \mu'_2, ..., \mu'_s)' \) denote the corresponding vector of expected values. A linear model for \( \mu \) can be expressed as

\[
E\{ \bar{y} \} = \mu = X\beta,
\]

where \( X \) is the known \((ds \times t)\) model specification matrix with full rank and \( \beta \) is the \((t \times 1)\) vector of unknown parameters. The covariance matrix for \( \bar{y} \) can be consistently estimated by

\[
\Sigma_{\bar{y}} = D_{\text{diag}} \{ \Sigma_{\bar{y}_h} \}
\]

(1.26)

where \( \Sigma_{\bar{y}_h} = \frac{1}{n_{hl}} \Sigma_{\bar{y}_h} \). The \( \{ n_{hl} \} \) are assumed to be sufficiently large that \( \Sigma_{\bar{y}} \) is nonsingular. The weighted least squares estimator for \( \beta \) is given by

\[
b = (X'\Sigma_{\bar{y}}^{-1}X)^{-1}X'\Sigma_{\bar{y}}^{-1}\bar{y}
\]

(1.27)

and a consistent estimator for its covariance matrix is

\[
\Sigma_b = (X'\Sigma_{\bar{y}}^{-1}X).
\]

(1.28)
The goodness of fit of the model can be assessed by the Wald statistic

\[ Q_w = (\bar{Y} - \bar{X} \beta)' \mathbf{V}_Y^{-1}(\bar{Y} - \bar{X} \beta) \]  

which approximately has the \( \chi^2 \) distribution with d.f. = (sd - t). Tests of full rank linear hypotheses \( H_0: \mathbf{C} \beta = 0 \) can be performed using

\[ Q_C = (\mathbf{C} \beta)'(\mathbf{C} \mathbf{V}_Y \mathbf{C})^{-1}(\mathbf{C} \beta), \]  

this statistic approximately has the \( \chi^2 \) distribution with d.f. = Rank(C).

1.7.3 **The Linear Model for Functions of Means**

This methodology can be generalized to the functional linear model

\[ \mathbf{F}(\mu) = \bar{X} \bar{\beta} \]  

where \( \mathbf{F}(\mu) = [F_1(\mu), F_2(\mu), ..., F_u(\mu)]' \) is a set of \( u < ds \) functions of interest.

The weighted least squares estimators for \( \bar{\beta} \) are given by

\[ \bar{\beta} = (\bar{X}' \mathbf{V}_F \mathbf{X})^{-1} \bar{X}' \mathbf{V}_F \mathbf{F} \]  

where \( \mathbf{F} = \mathbf{F}(\bar{Y}) \),

and

\[ \mathbf{V}_{\mathbf{F}} = \left[ \left. \frac{\partial \mathbf{F}}{\partial \bar{z}} \right| \bar{z} = \bar{Y} \right] \mathbf{V}_Y \left[ \left. \frac{\partial \mathbf{F}}{\partial \bar{z}} \right| \bar{z} = \bar{Y} \right]' \]  

(1.33)
is a consistent estimator for $\text{Var}_A(\mathbf{F})$. For large samples $\{n_h\}$, $\mathbf{b}$ approximately has a multivariate normal distribution with $\mathbb{E}_A(\mathbf{b}) = \mathbf{\theta}$ and a covariance matrix for which

$$
\Sigma_b = (X' \Sigma X)^{-1}
$$

is a consistent estimator. A goodness of fit statistic for assessing the fit of the model,

$$
Q_w = (\mathbf{F} - \mathbf{Xb})' \Sigma X^{-1} (\mathbf{F} - \mathbf{Xb}),
$$

has approximately a $\chi^2$ distribution with d.f. = $(u - t)$. Full rank linear hypotheses $\mathbb{C}_b = \mathbb{Q}$ can be tested with

$$
Q_C = (\mathbb{C}_b)' [\mathbb{C}_b \Sigma_b \mathbb{C}]^{-1} (\mathbb{C}_b),
$$

which has a $\chi^2$ distribution with d.f. = $\text{Rank}(\mathbb{C})$ in large samples. Predicted values $\hat{\mathbf{F}} = \hat{X} \hat{\mathbf{b}}$ can be used to generate more efficient estimates $\hat{\mathbf{F}}$ for $\mathbb{E}(\mu)$ than their observed counterparts. The properties of weighted least squares methods for functional linear models pertaining to mean vectors $\{\mathbf{y}_h\}$ are essentially the same as those for multinomially distributed counts $\{n_{hi}\}$.

1.7.4 **WLS Methods for Longitudinal Clustered Categorical Data**

Koch, Landis, Freeman, *et al.* (1977) proposed an approach for marginal
modeling of multivariate categorical data from repeated measures experiments. Their paper outlines appropriate hypotheses, estimation procedures (weighted least squares), and Wald statistics for testing linear hypotheses concerning the parameters. Special attention is paid to the problems of conceptually large tables and empty cells.

Stanish, Gillings, and Koch (1978), proposed using multivariate ratios and weighted least squares methods for the analysis of a longitudinal clinical trial with ordinal categorical data with missing data on one or more dependent variables. To deal with the missing data, ratio estimators for the means of the dependent variables (i.e., response at each of three visits) were modeled using weighted least squares regression methods. An estimate of the covariance matrix of these ratio estimators was obtained by the delta-method. Large sample sizes ensure consistency of the estimated covariance matrix, the bias of the ratio estimators to be negligible, and the distributions of the ratio estimates to be approximately normal. A fundamental assumption of this approach is that the missing data are independent of the outcome, and the method is not recommended in situations where more than ten percent of the observations are missing for any of the dependent variables under study.

LaVange, Keyes, Koch, et al. (1992) proposed using multivariate ratio methods for the analysis of incidence densities for an observational study of lower respiratory illness in children. The features of the data included 1) repeat occurrences of the outcomes for a given subject, inherent variability in at-risk periods, individual risks that changed over the course of the study. The method requires minimal distributional assumptions and accounts for variability in at-risk periods. Linear models were fit to the logarithm of incidence density ratios
using weighted least squares methods. Ratio means were used to take into account the variation in both the outcome and the exposure-time denominators.

1.7.5 **WLS Methods for Univariate Binary Data from Nested Designs with Balanced Clusters**

Kempthorne & Koch (1983) proposed a model for the analysis of clustered attribute data from a two-stage nested design, allowing the joint estimation of the overall mean response and intraclass correlation coefficient among elements within each cluster. They considered a random sample of clusters and observations within each cluster from populations which are (conceptually) infinite in size. Intermediate quantities were calculated within each cluster such as the cluster means and pairwise joint probability that two distinct observations in a cluster have the attribute, and were used to estimate the overall mean response and overall intraclass correlation. The weighted least squares estimation procedure of Grizzle, Starmer, and Koch (1969) with direct input of these quantities and their covariance matrix was used for the estimation (Koch and Imrey *et al.*, 1985). Two cases were considered: the case of balanced clusters and also the partially balanced case (i.e., when clusters are grouped so that clusters in a particular group have the same number of elements). For each of these cases they allowed either one or two subpopulations per cluster (e.g., males and females). In the case of two subpopulations, two overall means and three intraclass correlations were estimated: the overall mean response for each subpopulation, the intraclass correlations for observations within each subpopulation and the intraclass correlation between observations in different subpopulations within a cluster. Categorical covariates at the cluster- or patient-
levels (which define the subpopulations) may be accommodated by this method. This approach will accommodate large clusters (e.g., clusters with more than 10 observations each).

Prior to the work of Kempthorne and Koch (1983), Landis and Koch (1977) and Kleinman (1973 and 1975) also used estimates of cluster-level means to estimate the overall mean response. The intraclass correlation coefficient was also estimated, but no standard error was produced for this parameter by their procedure. Beitler & Landis (1985) proposed a two-way ANOVA model for multinomial data which is directly analogous to the mixed model for continuous data. It is a direct extension of the one-way model developed in Landis and Koch (1977). Model-based variance components are first directly estimated. Error variance components are allowed to vary across treatment levels. No standard error estimates are given for these quantities. The estimated variance components are then incorporated into the covariance structure of a weighted least squares general linear models framework which can be used for parameter estimation and hypothesis testing. A Wald statistic for testing (fixed-effects) treatment differences is proposed as long as the random effects assumptions of the mixed model are reasonable as assessed by a goodness of fit statistic. A similar model was proposed by Miller and Landis (1991) for a clustered categorical, and not necessarily binary, response variable. General correlation patterns were accounted for by a mixed effects modeling approach. Estimation of the cluster variance components was done through the method of moments, but no standard errors for these quantities were produced. Modeling of functions of the observed proportions was carried out using GEE1 rather than WLS methods.
1.7.6 WLS Methods for Bivariate and R-Variate Binary Data from Nested Designs with Balanced Clusters

Marques & Koch (1992) proposed a method for the analysis of bivariate dichotomous data from a (stratified) cluster sample. A model similar to the one used by Kempthorne and Koch (1983) was used, but, the response considered for Marques' and Koch's model was a bivariate vector of binary outcomes rather than a univariate binary outcome. The bivariate case requires the joint estimation of each marginal mean, the intraclass correlation among observations for different individuals in the same cluster, and an additional intraclass correlation parameter for association among observations for the same individual within a cluster. Weighted least squares was used for estimation. The cases of two or more subpopulations per cluster were also addressed. One disadvantage of this methodology is that observations which are incomplete must be deleted from the analysis (a problem which is characteristic of traditional multivariate analysis techniques). Also, the clusters must either be balanced or partially balanced, so that it may not be used for the most general and frequently occurring case of unbalanced clusters. Covariates at the cluster-, patient-, or subpatient-level which define the subpopulations may be accommodated, but they must be categorical. An advantage of this method is that it may be used to analyze large clusters.

1.8 Direct Estimation Methods

There are several approaches to directly estimating the overall mean response for unbalanced clustered binary data. One estimator is simply the overall mean which ignores the structure of the cluster. It is advantageous when
the intraclass correlation is near zero. This estimator will be used for comparative purposes since it is the estimate and usual standard error appropriate for a simple random sample. Another estimator is the mean of the cluster means. It has the limitation of not weighting each cluster by the amount of information in each cluster. A third estimator is one which was proposed by Koch (1967) which uses an alternate weighting scheme.

Kempthorne (1982) proposed several heuristic estimators for the overall mean response and intraclass correlation coefficient(s) for the unbalanced case. The underlying model is the same as the one used by Kempthorne and Koch (1983). The mean response of the sampled observations in each cluster, $\bar{y}_i$, is calculated. Each are unbiased estimators of the overall mean response, $\pi$. The sample variance of the elements in the i-th cluster, $s_i^2$, is unbiased for $\pi - \lambda$, where $\lambda$ is the probability that two distinct observations in a cluster have the attribute.

$$\text{Var}[\bar{y}_i, s_i^2] = \begin{bmatrix}
(\lambda - \pi^2) & (\lambda - \pi^2) - (\xi - \pi\lambda) \\
(\lambda - \pi^2) - (\xi - \pi\lambda) & (\lambda - \pi^2) - 2(\xi - \pi\lambda) + (\eta - \lambda^2)
\end{bmatrix}
$$

\begin{align*}
+ \frac{1}{d_i} \begin{bmatrix}
(\pi - \lambda) & (\pi - \lambda) - 2(\lambda - \xi) \\
(\pi - \lambda) - 2(\lambda - \xi) & (\pi - \lambda) - 4(\lambda - \xi) + 4(\xi - \eta)
\end{bmatrix}
\end{align*}

\begin{align*}
+ \frac{2}{d_i(d_i - 1)} \begin{bmatrix}
0 & 0 \\
0 & \lambda - 2\xi + \eta
\end{bmatrix},
\end{align*}

(1.37)
where $\xi$ is the probability that three distinct observations in a cluster have the attribute, $\eta$ is the probability that four distinct observations in a cluster have the attribute, and $d_i$, the number of observations sampled in cluster $i$, is greater than or equal to four for all clusters. Conditionally unbiased estimators of the linear parameter functions, $(\pi - \lambda)$, $(\lambda - \xi)$, $(\xi - \eta)$, and $(\lambda - 2\xi + \eta)$, which, in part, characterize $\text{Var}[\bar{Y}_i, s_i^2]$, exist within each cluster. Unbiased estimators of the nonlinear parameter functions, $(\lambda - \pi^2)$, $(\xi - \pi \lambda)$, and $(\eta - \lambda^2)$, do not exist within each cluster and must be estimated across clusters. A method of obtaining unbiased estimators of these parameter functions involving the evaluation of the expectations of the squares of certain differences among the observations (Koch, 1968), was used. These estimates of the linear and nonlinear parameter functions were used to estimate $\text{Var}[\bar{Y}_i, s_i^2]$ for the $i$-th cluster. Finally, the cluster-specific estimates of $\pi$ and $\pi - \lambda$ were combined over all clusters using either a mean of means, overall mean, or Koch's mean principle to obtain more efficient estimates of $\pi$ and $\pi - \lambda$. The respective variances and covariance between the overall estimates of $\pi$ and $\pi - \lambda$ are also obtained through linear combinations of the cluster-specific variances and covariances. This method does not necessarily produce optimal estimates. Weighted least squares was also mechanically applied to combine the cluster-specific estimates, $\bar{Y}_i$ and $s_i^2$, as yet another approach, but since each function was based on only one estimate, this procedure does not have the asymptotic properties weighted least squares estimators ordinarily have. It was performed for comparative purposes only.
1.9 Summary

Maximum likelihood models for clustered categorical data may only be used when it is reasonable to assume that the clustered data follow a particular parametric distribution (e.g., beta-binomial). Their usefulness is also limited in that many of these models can only accommodate cluster-level covariates. GEE methods relax the distributional assumptions. GEE1 gives consistent estimates which relate covariates to the outcome even when the working correlation matrix is misspecified. While an estimate of the correlation matrix is produced by this method, it has been shown to be highly inefficient (Liang, Zeger, and Qaqish, 1992). GEE2 is a modification of GEE1 in which both the outcome and pairwise joint probabilities are modeled simultaneously and provides more efficient estimates of second order model parameters. Neither GEE1 nor GEE2 currently accommodate complex survey sampling weights. The model discussed in Qaqish and Liang (1992) for clustered binary data with multiple levels of nesting, allows modeling of the mean and covariance structure simultaneously and uses GEE2 for estimation; however, it may be impractical for data with large clusters and cannot handle complex survey sample weights. Pseudo-maximum likelihood methods such as survey data logistic regression (Shah, Barnwell, Hunt et al., 1992) can handle sampling weights from complicated designs but cannot be used for analysis of the covariance structure among nested levels of observations within clusters, although design effects may be obtained for parameter estimates. The methods proposed by Kempthorne and Koch (1983) provide joint estimates of the overall response mean, intraclass correlation coefficients, and their covariance matrix; however, they are only applicable in the case of univariate outcomes. Marques and Koch (1992) developed extensions to the methods of
Kempthorne and Koch (1983) to handle clustered bivariate and multivariate binary data, but only for the balanced and partially balanced cases, and not for the most commonly occurring unbalanced case. Their method does not provide a way to handle incomplete or missing data. Multivariate ratio methods provide estimates of the overall mean response and standard errors which take into account variation in sample sizes due to cluster sampling; but no information about the dependence structure, with the exception of approximate design effects, is provided.

1.10 **Overview of Research**

Chapter II will extend simple ratio mean analysis for (multivariate) binary data from a one-stage cluster sample to regression analysis of ratio means using WLS methods. Both strictly linear and functional linear models for ratio means will be proposed. The estimates and standard errors from a ratio mean regression (RMR) model for the logit of the ratio mean will be compared with those estimates and standard errors from survey data logistic regression, GEE1, and ordinary logistic regression using the dental data. It will be demonstrated that the RMR approach will be useful in situations where GEE1 is not, that is, for situations with binary data from large clusters containing unequal numbers of elements.

Chapter III will extend the methods of Chapter II to accommodate multycategory, ordinal, and continuous outcomes from a one-stage cluster sample with large and unequally sized clusters. Thus, the flexibility of the ratio mean approach will be demonstrated.

Chapter IV will modify and extend the RMR methods of Chapter II to
two-stage cluster samples. The sampled clusters may still be large and contain unequal numbers of elements.

Chapter V will propose a new measure of intracluster correlation which is appropriate for binary data from a one-stage cluster sample with clusters that are unequal in size and vary with respect to the number of relevant observations they contain. The measure will simultaneously take into account the random sampling variability of the outcome of interest along with the number of relevant observations. Joint estimates of the overall (ratio) mean and this new measure of intracluster correlation will be proposed. Joint modeling of these two estimators for multiple strata defined by cluster-level covariates will be accomplished through the use of weighted regression methods. This analysis approach is an alternative to GEE2 for situations in which cluster sizes are large and when there is interest in modeling only cluster-level characteristics.
CHAPTER II
RATIO MEAN REGRESSION FOR THE ANALYSIS OF BINARY DATA
FROM A ONE-STAGE CLUSTER SAMPLE USING WEIGHTED
REGRESSION METHODS

2.1 Introduction

An analysis of data from a cluster sample must take into account the
correlation among observations in the same cluster. In addition, when the
number of observations per cluster is unbalanced, the contribution to the random
variation due to variable cluster sizes should also be taken into account. One
way to accomplish both of these goals is through the use of ratio means.

This chapter will focus on estimating the overall mean response for
observations of a binary outcome from subjects in a one-stage cluster sample with
large, unbalanced clusters, using ratio mean methods. The ratio mean estimator,
its standard error, and a large sample test statistic for contrasting two or more
(correlated) ratio means will be reviewed. These ideas will then be extended to
allow a vector of ratio means (or, a vector of logits of ratio means, or natural
logarithms of ratio means), corresponding to a crossclassification of categorical
covariates to be modeled using weighted least squares methods. The covariates
may apply to the cluster as a whole or vary according to cluster subunit
characteristics. The method will be illustrated using data from the dental study
described in Chapter I. The results will then be compared to similar models fit
using survey data logistic regression, GEE1, and ordinary logistic regression.
2.2 **Ratio Mean Definition**

A ratio mean estimator for the overall population mean per element of an attribute of interest will be defined in this section for observations of a binary outcome from subjects in a one-stage cluster sample with clusters that are unequal in size. The assumed sampling method for clusters is simple random sampling with replacement (or, equivalently, without replacement from a very large population). The binary outcome may be observed for one or more sites for each study subject. Although the notation used in this section may seem overly complex, its usefulness will become apparent later in the chapter.

Let $i = 1, \ldots, N$ index sample clusters, $j = 1, \ldots, \nu_i$ index patients in the $i$-th cluster, and $t = 1, \ldots, m_{ij}$ index multiple observations for patient $j$ in the $i$-th cluster. Here, $N$ represents the number of sampled clusters, $\nu_i$ represents the total number of patients in the $i$-th cluster, and $m_{ij}$ represents the total number of potential observations for patient $j$ in the $i$-th cluster. (For the dental data application, $m_{ij}$ will equal a constant, $m$). Let $x_{ijt}$ be a binary outcome which takes the value 1 if the $t$-th observation for the $j$-th patient in the $i$-th cluster is relevant (e.g. observed), and 0 otherwise, and let $y_{ijt}$ be a binary outcome which takes the value 1 if the $t$-th observation for patient $j$ in cluster $i$ is relevant and has the attribute of interest, and 0 otherwise. (For some applications, all observations for each patient will be relevant, so that $x_{ijt} = 1$). Also define

\[
y_{i..} = \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} y_{ijt} \quad \text{and} \quad x_{i..} = \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} x_{ijt}.
\]

Here, $y_{i..}$ is the total number of relevant observations with the attribute for the $i$-th cluster and $x_{i..}$ is the total number of relevant observations in the $i$-th cluster. Since the assumed sampling method for clusters is simple random sampling with replacement, the $(y_{i..}, x_{i..})$ are independent and identically
distributed. The ratio mean estimator for the proportion of relevant observations with the attribute over all clusters is defined as:

$$ R = \left( \frac{\sum_{i=1}^{N} y_i}{N} \right) / \left( \frac{\sum_{i=1}^{N} x_i}{N} \right) $$

$$ = \bar{y} / \bar{x} , $$

where $\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$ and $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$.

$R$ can be viewed as the estimated average number of occurrences of relevant observations with the attribute per cluster divided by the average number of relevant observations per cluster. Since the $N$'s cancel, $R$ is, more simply, the estimated total number of occurrences of relevant observations with the attribute divided by the estimated total number of relevant observations, or, the estimated proportion of occurrences of the attribute of interest among relevant observations. The numerator and denominator of $R$ are each means of $N$ independent and identically distributed random variates, but the numerator and denominator are correlated with each other.

For the dental study example, the relevant observations are those posterior teeth with restorations having greater than or equal to three surfaces indicated by $x_{ij} = 1$, with $m_{ij} = m = 16$ (the number of first and second molars and premolars which each person potentially has). The ratio mean estimator for the proportion of three or more surface restorations which are crowns, is the average number of crowns per cluster divided by the average number of teeth with restorations having greater than or equal to three surfaces per cluster, or, the total number of crowns divided by the total number of teeth with restorations having three or more surfaces. Another example is the choice of treatment option
by a medical practitioner for patients with a particular eye disease. The disease may be unilateral (i.e., occurring in one eye, only), or bilateral (i.e., occurring in both eyes). The diseased eyes would be the relevant eyes indicated by $x_{ijt}$, with $m_{ij}=2$.

A first-order Taylor series representation for $R$ at the population mean $(\mu_y, \mu_x)$ of the $(y_{i\ldots}, x_{i\ldots})$ is given by

$$R = \frac{\mu_y}{\mu_x} + \frac{1}{\mu_x} (\overline{y} - \mu_y) - \frac{\mu_y}{\mu_x^2} (\overline{x} - \mu_x) + O(\frac{1}{N})$$

(2.2)

$$= \frac{\mu_y}{\mu_x} + \frac{1}{\mu_x} \left\{ (\overline{y} - \theta \overline{x}) - (\mu_y - \theta \mu_x) \right\} + O(\frac{1}{N}).$$

The asymptotic expected value of a first-order Taylor series for $R$, $E_A(R)$, is $\frac{\mu_y}{\mu_x} = \theta$, the ratio mean in the population.

The variance of $R$, based on a Taylor series linearization, is given by:

$$\text{Var}(R) = \left( \frac{\mu_y}{\mu_x} \right)^2 \left\{ \frac{\text{Var}(\overline{y})}{\mu_y^2} - \frac{2 \text{Cov}(\overline{y}, \overline{x})}{\mu_y \mu_x} + \frac{\text{Var}(\overline{x})}{\mu_x^2} \right\}$$

(2.3)

$$= \frac{1}{\mu_x^2} \left\{ \text{Var}(\overline{y} - \theta \overline{x}) \right\}.$$

A consistent estimator for $\text{Var}(R)$ is given by

$$\text{var}(R) = \frac{R^2}{N} \left\{ \frac{\overline{y}^2}{N} - \frac{2 \overline{xy}}{\overline{y} \overline{x}} + \frac{\overline{x}^2}{N} \right\}$$

(2.4)

$$= \frac{1}{N \overline{x}^2} \sum_{i=1}^{N} (y_{i\ldots} - R x_{i\ldots})^2/(N-1),$$

where
\[ s_y^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{(N - 1)}, \]

\[ s_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{(N - 1)}, \]

\[ s_{xy} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{(N - 1)}, \]

\[ \bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}, \text{ and} \]

\[ \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}. \]

Note that expression (2.4) for \( \text{var}(R) \) is always positive. A large sample variance estimator for \( \text{Var}(\ln R) \), also based on a Taylor series linearization, is given by

\[ \text{var}(\ln R) = \frac{\text{var}(R)}{R^2}. \quad (2.5) \]

A large sample \((1 - \alpha)\)-level confidence interval for \( \theta \) is given by:

\[ \exp \left\{ \ln R \pm Z_{(1-\alpha/2)} \cdot \left( \text{var}(R)^{1/2} / R \right) \right\}. \quad (2.6) \]

2.2.1 Stratification of Ratio Means by Cluster-Level Covariates

Ratio means may be calculated separately for subgroups of observations corresponding to crossclassifications of cluster-level covariates. In this situation, the assumed sampling method for clusters is stratified simple random sampling.
with replacement. For the dental study example, ratio means may be computed using only urban practices and, likewise, only rural practices.

Let \( h=1,\ldots,H \) index strata formed by the crossclassification of cluster characteristics, \( i=1,\ldots,N_h \) index sample clusters in the \( h \)-th stratum, \( j=1,\ldots,\nu_{hi} \) index patients from the \( i \)-th cluster in the \( h \)-th stratum, and \( t=1,\ldots,m_{hij} \) index multiple observations for patient \( j \) in cluster \( i \) and stratum \( h \). Here, \( H \) represents the number of strata formed by the crossclassification of cluster-level characteristics, \( N_h \) represents the number of sampled clusters in stratum \( h \), \( \nu_{hi} \) represents the total number of patients in cluster \( i \) of stratum \( h \), and \( m_{hij} \) represents the total number of potential observations for patient \( j \) in cluster \( i \) of stratum \( h \). Define a binary outcome \( x_{hijt} \) which takes the value 1 if the \( t \)-th observation for patient \( j \) of cluster \( i \) in stratum \( h \) is relevant, and 0 otherwise, and let \( y_{hijt} \) be a binary outcome which takes the value 1 if the \( t \)-th observation for patient \( j \) of cluster \( i \) of stratum \( h \) is relevant and has the attribute, and 0 otherwise. Also define

\[
y_{hi.} = \frac{\sum_{j=1}^{\nu_{hi}} m_{hij} y_{hijt}}{\nu_{hi}} \quad \text{and} \quad x_{hi.} = \frac{\sum_{j=1}^{\nu_{hi}} m_{hij} x_{hijt}}{\nu_{hi}}.
\]

The ratio mean estimator for the \( h \)-th stratum is given by:

\[
R_h = \left( \frac{\sum_{i=1}^{N_h} y_{hi.}}{N_h} \right) / \left( \frac{\sum_{i=1}^{N_h} x_{hi.}}{N_h} \right)
\]

\[= \frac{\bar{y}_h}{\bar{x}_h},\]

where \( \bar{y}_h = \frac{\sum_{i=1}^{N_h} y_{hi.}}{N_h} \) and \( \bar{x}_h = \frac{\sum_{i=1}^{N_h} x_{hi.}}{N_h} \).

The quantity \( R_h \) is, more simply, the estimated average number of
relevant occurrences with the attribute per cluster in the h-th stratum divided by the average number of relevant observations per cluster in the h-th stratum. Since the $N_h$ cancel, it may also be interpreted as the estimated total number of relevant occurrences with the attribute in the h-th stratum divided by the estimated total number of relevant observations in the h-th stratum.

For the dental study example, the ratio mean for urban practices is the estimated total number of crowns for urban practices divided by the estimated total number restorations having at least three surfaces for urban practices, or, equivalently, the average number of crowns per urban practice divided by the average number of three or more surface restorations per urban practice.

A first-order Taylor series representation for $R_h$ at the population mean $(\mu_{hy}, \mu_{hx})$ of the $(y_{hi}, \ldots, x_{hi}, \ldots)$ is given by

$$R_h = \frac{\mu_{hy}}{\mu_{hx}} + \frac{1}{\mu_{hx}} (\bar{y}_h - \mu_{hy}) - \frac{\mu_{hy}}{2 \mu_{hx}^2} (\bar{x}_h - \mu_{hx}) + O\left(\frac{1}{N_h}\right) \quad (2.8)$$

The asymptotic expected value of a first-order Taylor series for $R_h$, $E_A(R_h)$, is $\frac{\mu_{hy}}{\mu_{hx}} = \theta_h$, the ratio mean in the population for the h-th stratum.

The variance of $R_h$, based on a Taylor series linearization, is given by:

$$\text{Var}(R_h) = \left(\frac{\mu_{hy}}{\mu_{hx}}\right)^2 \left\{ \frac{\text{Var}(\bar{y}_h)}{\mu_{hy}^2} - \frac{2 \text{Cov}(\bar{y}_h, \bar{x}_h)}{\mu_{hy} \mu_{hx}} + \frac{\text{Var}(\bar{x}_h)}{\mu_{hx}^2} \right\} \quad (2.9)$$

A large sample variance estimator for $R_h$ is given by

$$\text{var}(R_h) = \frac{R_h^2}{N_h} \left\{ \frac{s_{hy}^2}{\bar{y}_h^2} - \frac{2 s_{hxy}}{\bar{y}_h \bar{x}_h} + \frac{s_{hx}^2}{\bar{x}_h^2} \right\} \quad (2.10)$$

where
\[ s_{hy}^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2}{(N_h - 1)}, \]
\[ s_{hx}^2 = \frac{\sum_{i=1}^{N_h} (x_{hi} - \bar{x}_h)^2}{(N_h - 1)}, \]
\[ s_{hxy} = \frac{\sum_{i=1}^{N_h} (x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{(N_h - 1)}, \]
\[ \bar{y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}, \quad \text{and} \]
\[ \bar{x}_h = \frac{\sum_{i=1}^{N_h} x_{hi}}{N_h}. \]

It is also noted that a large sample variance estimator for \( \text{Var}(\ln R_h) \) based on a Taylor series linearization is given by

\[ \text{var}(\ln R_h) = \frac{\text{var}(R_h)}{R_h^2}. \tag{2.11} \]

A large sample \((1 - \alpha)\)-level confidence interval for \( \theta_h \) is given by:

\[ \exp \left\{ \ln R_h \pm Z_{(1-\alpha/2)} \cdot \left( \text{var}(R_h)^{1/2} / R_h \right) \right\}. \tag{2.12} \]

2.2.2 Comparing Ratio Means Between Cluster-Level Subgroups

In many situations, it is of interest to compare one or more ratio means for subgroups defined by cluster-level characteristics to evaluate the extent to which differences are not due to sampling variation. For the dental study
example, it may be of interest to compare the ratio means which estimate the proportion of restorations with three or more surfaces which are crowns, between urban and rural practices. A large sample test statistic which may be used to compare two ratio means, \( R_h \) and \( R_{h'} \), is given by:

\[
Q = \frac{\left\{ \ln \left( \frac{R_h}{R_{h'}} \right) \right\}^2}{\text{var} \left[ \ln \left( \frac{R_h}{R_{h'}} \right) \right]},
\]

where \( \text{var} \left[ \ln \left( \frac{R_h}{R_{h'}} \right) \right] = \left\{ \text{var}(R_h)/R_h^2 + \text{var}(R_{h'})/R_{h'}^2 \right\} \), is based on a Taylor series linearization. The test statistic, \( Q \), has approximately a chi-squared distribution with one degree of freedom in large samples.

A large sample \((1 - \alpha)\)-level confidence interval for the rate ratio, \( \theta_h/\theta_{h'} \), is given by

\[
\exp \left\{ \ln \left( \frac{R_h}{R_{h'}} \right) \pm Z_{(1-\alpha/2)} \cdot \sqrt{\text{var} \left[ \ln \left( \frac{R_h}{R_{h'}} \right) \right]} \right\}.
\]

(2.14)

2.3 Subgroups of Ratio Means Defined by Patient-Level Characteristics

Ratio means may, alternatively, be calculated for subgroups defined by patient-level characteristics. For the dental study example, ratio means may be computed separately for males and females and, likewise, by age category. Unlike ratio means stratified according to cluster-level characteristics, however, ratio means stratified on patient characteristics are not necessarily independent. They may be correlated since patients having different levels of a given patient-level characteristic may come from the same cluster.

For the sampling framework in Section 2.2, let \( i=1,\ldots,N \) index sample clusters, \( j=1,\ldots,\nu_i \) index patients sampled in cluster \( i \), \( t=1,\ldots,m_{ij} \) index multiple
observations for patient \( j \) in cluster \( i \), and \( k = 1, \ldots, K \) index patient types (i.e., crossclassifications of patient-level characteristics). Here, \( N \) represents the number of sampled clusters, \( \nu_i \) represents the number of patients in cluster \( i \), \( m_{ij} \) represents the total number of potential observations for patient \( j \) in cluster \( i \), and \( K \) represents the number of patient types. Define a binary outcome \( x_{ikjt} \) which takes the value 1 if the \( t \)-th observation for patient \( j \) of cluster \( i \) is relevant and is for a patient of type \( k \), and 0 otherwise, and let \( y_{ikjt} \) be a binary outcome which takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) is for a patient of type \( k \), is relevant, and has the attribute, and 0 otherwise. Thus, \( y_{ikjt} = y_{ijt} \cdot x_{ikjt} \). Also define

\[
y_{ik} = \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} y_{ikjt} \quad \text{and} \quad x_{ik} = \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} x_{ikjt}.
\]

Note that the summations over \( j \) encompass all \( \nu_{ik} \) patients of type \( k \) in the \( i \)-th cluster. Also, the vectors \((y_{i1}, \ldots, x_{i1}, \ldots, y_{iK}, \ldots, x_{iK}, \ldots)'\) are independent and identically distributed as a consequence of the assumed sampling method for clusters being simple random sampling with replacement. A ratio mean estimator for the subset of patients of type \( k \) is defined as:

\[
R_k = \left( \frac{\sum_{i=1}^{N} y_{ik}}{N} \right) / \left( \frac{\sum_{i=1}^{N} x_{ik}}{N} \right) \quad (2.15)
\]

The quantity \( R_k \) is the estimated average number of relevant occurrences of the attribute for patients of type \( k \) per cluster divided by the estimated average number of relevant observations for patients of type \( k \) per cluster. For the dental data example, the ratio mean for female patients is the estimated total number of crowns for female patients divided by the estimated total number of restorations with three or more surfaces for female patients, or, the average number of crowns
for female patients divided by the average number of greater than or equal to
three surface restorations for female patients.

A first-order Taylor series representation for $R_k$ at the population mean
$(\mu_{ky}, \mu_{kx})$ of the $(y_{ik}, \ldots, x_{ik}, \ldots)$ is given by

$$
R_k = \frac{\mu_{ky}}{\mu_{kx}} + \frac{1}{\mu_{kx}} (\overline{y}_k - \mu_{ky}) - \frac{\mu_{ky}}{\mu_{kx}^2} (\overline{x}_k - \mu_{kx}) + O(\frac{1}{N})
$$

$$
= \theta_k + \theta_k \frac{(\overline{y}_k - \mu_{ky})}{\mu_{ky}} - \theta_k \frac{\overline{x}_k - \mu_{kx}}{\mu_{kx}} + O(\frac{1}{N}) \quad (2.16)
$$

The asymptotic expected value of a first-order Taylor series for $R_k$,
$E_A(R_k)$, is $\mu_{ky} / \mu_{kx} = \theta_k$, the ratio mean for the k-th subpopulation.

The variance of $R_k$, based on a Taylor series linearization, is given by:

$$
\text{Var}(R_k) = \left(\frac{\mu_{ky}}{\mu_{kx}}\right)^2 \left\{ \frac{\text{Var}(\overline{y}_k)}{\mu_{ky}^2} - \frac{2 \text{Cov}(\overline{y}_k, \overline{x}_k)}{\mu_{ky} \mu_{kx}} + \frac{\text{Var}(\overline{x}_k)}{\mu_{kx}^2} \right\} \quad (2.17)
$$

since the $(y_{ik}, \ldots, x_{ik}, \ldots)$ are independent and identically distributed. A variance estimator for $R_k$ is given by

$$
\text{var}(R_k) = \frac{R_k^2}{N} \left\{ \frac{s_{ky}^2}{\overline{y}_k^2} - \frac{2 s_{kxy}}{\overline{y}_k \overline{x}_k} + \frac{s_{kx}^2}{\overline{x}_k^2} \right\} \quad (2.18)
$$

where

$$
s_{ky}^2 = \frac{N}{N-1} \sum_{i=1}^N (y_{ik} - \overline{y}_k)^2
$$

$$
s_{kx}^2 = \frac{N}{N-1} \sum_{i=1}^N (x_{ik} - \overline{x}_k)^2
$$

$$
s_{kxy} = \frac{N}{N-1} \sum_{i=1}^N (x_{ik} - \overline{x}_k)(y_{ik} - \overline{y}_k)
$$
\[ \bar{y}_k = \frac{\sum_{i=1}^{N} y_{ik}}{N}, \text{ and} \]

\[ \bar{x}_k = \frac{\sum_{i=1}^{N} x_{ik}}{N}. \]

The covariance of \( R_k \) and \( R_{k'} \), based on a Taylor series linearization, is given by:

\[
\text{Cov}(R_k, R_{k'}) = \theta_k \theta_k' \left\{ \frac{\text{Cov}(\bar{y}_k, \bar{y}_{k'})}{\mu_k \mu_{k'}'}, \frac{\text{Cov}(\bar{y}_k, \bar{x}_{k'})}{\mu_k \mu_{k'}'x_{k'}} - \frac{\text{Cov}(\bar{x}_k, \bar{x}_{k'})}{\mu_k \mu_{k'}'x_{k'}} \right\} + \frac{\text{Cov}(\bar{x}_k, \bar{x}_{k'})}{\mu_k \mu_{k'}'x_{k'}} \]

since the \((y_{ik}, x_{ik}, y_{ik'}, x_{ik'})\) are independent and identically distributed. An estimator of \( \text{Cov}(R_k, R_{k'}) \) is given by:

\[
\text{cov}(R_k, R_{k'}) = \frac{R_k R_{k'}'}{N} \left\{ \frac{s_{kk'y}}{y_k y_{k'}'} - \frac{s_{kk'y}x_{k'}}{y_k x_{k'}'} - \frac{s_{k'kyx}}{y_k x_{k'}}' + \frac{s_{kk'x}}{x_k x_{k'}'} \right\} \]

where

\[
s_{kk'y} = \sum_{i=1}^{N} (y_{ik} - \bar{y}_k)(y_{ik'} - \bar{y}_{k'}') / (N - 1),
\]

\[
s_{kk'y}x_{k'} = \sum_{i=1}^{N} (y_{ik} - \bar{y}_k)(x_{ik} - \bar{x}_k) / (N - 1),
\]

\[
s_{k'kyx} = \sum_{i=1}^{N} (y_{ik'} - \bar{y}_{k'})(x_{ik} - \bar{x}_k) / (N - 1),
\]

\[
s_{kk'x} = \sum_{i=1}^{N} (x_{ik} - \bar{x}_k)(x_{ik'} - \bar{x}_{k'}) / (N - 1),
\]

\[
\bar{y}_k = \frac{\sum_{i=1}^{N} y_{ik}}{N},
\]

\[
\bar{x}_k = \frac{\sum_{i=1}^{N} x_{ik}}{N}.
\]
and \[ \bar{x}_k = \frac{\sum_{i=1}^{N} x_{ik}}{N} \].

2.3.1 Comparing Ratio Means Between Patient-Level Subgroups

It may be of interest to compare two or more ratio means for subgroups of patients defined by patient-level characteristics to evaluate the extent to which differences are not due to sampling variation. For the dental study example, it may be of interest to compare ratio means which estimate the proportion of restorations with three or more surfaces which are crowns, between male and female patients. A large sample test statistic which may be used to compare two correlated ratio means, \( R_k \) and \( R_{k'} \), is given by:

\[ Q = \frac{\left\{ \ln (R_k/R_{k'}) \right\}^2}{\text{var[\ln (R_k/R_{k'})]}} \]  \hspace{1cm} (2.21)

where \( \text{var[\ln (R_k/R_{k'})]} = \left\{ \text{var}(R_k)/R_k^2 + \text{var}(R_{k'})/R_{k'}^2 - 2\text{cov}(R_k,R_{k'})/R_k R_{k'} \right\} \), is based on a Taylor series linearization. The test statistic, \( Q \), has approximately a chi-squared distribution with one degree of freedom in large samples.

A large sample \((1 - \alpha)\)-level confidence interval estimate for the rate ratio, \( \theta_k/\theta_{k'} \), is given by

\[ \exp \left\{ \ln (R_k/R_{k'}) \pm Z_{(1-\alpha/2)} \cdot \sqrt{\text{var[ln(R_k/R_{k'})]}} \right\} \] \hspace{1cm} (2.22)
2.4 Subgroups of Ratio Means Defined by Subpatient-Level Characteristics

Ratio means may also be calculated for subgroups of observations defined by subpatient-level characteristics. For the dental study example, ratio means may be computed separately for pre-molars versus molars and, likewise, for maxillary versus mandibular teeth. Like ratio means computed separately according to patient-level characteristics, ratio means computed separately according to subpatient-level characteristics are also not independent. They are correlated since subpatient-level observations with different levels of a characteristic may come from the same cluster (and patient).

Let \( i=1,...,N \) index clusters, \( j=1,...,\nu_i \) index patients in cluster \( i \), and \( t=1,...,m_{ij} \) index multiple observations for patient \( j \) in cluster \( i \), and \( l=1,...,L \) index subpatient-level observation types defined by cross-classifications of subpatient-level characteristics. Here, \( N \) represents the number of sampled clusters, \( \nu_i \) represents the total number of patients sampled in cluster \( i \), \( m_{ij} \) represents the total number of potential subpatient-level observations for patient \( j \) in cluster \( i \), and \( L \) represents the number of levels resulting from the cross-classification of subpatient-level characteristics. Define a binary outcome \( x_{ijlt} \) which takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) is relevant and of type \( l \), and 0 otherwise, and let \( y_{ijlt} \) be a binary outcome which takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) is relevant, of type \( l \), and has the attribute, and 0 otherwise.

Also define \( y_{i\cdot l} = \sum_{j=1}^{\nu_i} m_{ij} y_{ijlt} \) and \( x_{i\cdot l} = \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} x_{ijlt} \). Then the ratio mean for the \( l \)-th subpatient-level observation type is given by:

\[
R_l = \frac{\sum_{i=1}^{N} y_{i\cdot l} / N}{\sum_{i=1}^{N} x_{i\cdot l} / N}
\]  
(2.23)
The quantity $R_1$ is, more simply, the average number of occurrences of the attribute for relevant observations of type 1 per cluster divided by the average number of relevant observations of type 1 per cluster. For the dental data example, the ratio mean for the proportion of crowns in first molars, is the total number of crowns in first molars divided by the total number of first molars with a restoration with greater than or equal to three surfaces, or, the average number of crowns in first molars per cluster divided by the average number of first molars with restorations having three or more surfaces per cluster.

The asymptotic expectations and variances for $R_1$ and covariance between $R_1$ and $R_{1'}$ may be obtained by replacing the subscripts, $k$ and $k'$, in the expressions given in Section 2.3 by 1 and $1'$.

2.4.1 Comparing Ratio Means Between Subpatient-Level Subgroups

It may be of interest to compare two or more ratio means for subgroups based on subpatient-level characteristics to evaluate the extent to which differences are not due to sampling variation. For the dental study example, it may be of interest to compare mandibular versus maxillary teeth for the ratio means which estimate the proportion of crowns for greater than or equal to three surface restorations. A large sample test statistic which may be used to compare two correlated ratio means, $R_1$ and $R_{1'}$, is given by:

$$Q = \frac{\left\{ \ln \left( \frac{R_1}{R_{1'}} \right) \right\}^2}{\text{var}[\ln \left( \frac{R_1}{R_{1'}} \right)]},$$ (2.24)
where \( \text{var} [\ln(R_l/R_{l'})] = \left\{ \text{var}(R_l)/R_l^2 + \text{var}(R_{l'})/R_{l'}^2 - 2\text{cov}(R_l, R_{l'})/R_l R_{l'} \right\} \),

is based on a Taylor series linearization. The test statistic, \( Q \), has approximately a chi-squared distribution with one degree of freedom in large samples.

A large sample \((1 - \alpha)\)-level confidence interval estimate for the rate ratio \( \theta_l/\theta_{l'} \) is given by

\[
\exp\left\{ \ln (R_l/R_{l'}) \pm Z_{(1-\alpha/2)} \cdot \text{var} [\ln(R_l/R_{l'})]^{1/2} \right\}.
\]  

(2.25)

2.5 Subgroups of Ratio Means Defined by Cluster-Level, Patient-Level, and Subpatient-Level Characteristics

Ratio means do not have to be calculated separately for subgroups defined by cluster-, patient-, or subpatient-level characteristics; they may be calculated for subgroups of observations defined by the crossclassification of cluster-, patient-, and subpatient-level characteristics, simultaneously, provided there are adequate numbers of observations (at the cluster level) in each subpopulation of interest to assure asymptotic (log) normality of the ratio mean estimators.

Let \( h=1,\ldots,H \) index levels for the crossclassification of cluster-level characteristics, \( k=1,\ldots,K \) index levels for the crossclassification of patient-level characteristics, \( l=1,\ldots,L \) index levels for the crossclassification of subpatient-level characteristics, \( i=1,\ldots,N_h \) index sample clusters in stratum \( h \), \( j=1,\ldots,\nu_{hi} \) index patients in cluster \( i \) of stratum \( h \), \( t=1,\ldots,m_{hij} \) index multiple observations for patient \( j \) in cluster \( i \) of stratum \( h \). Here, \( N_h \) is the number of sample clusters in the \( h \)-th stratum, \( \nu_{hi} \) is the total number of patients in cluster \( i \) of stratum \( h \), \( m_{hij} \) is the total number of potential observations for patient \( j \) in cluster \( i \) of stratum \( h \), \( H \) is the number of levels for the crossclassification of cluster-level characteristics.
characteristics, \( K \) is the number of levels for the crossclassification of patient-level characteristics, and \( L \) is the number of levels for the crossclassification of subpatient-level characteristics. Define a binary outcome \( x_{hikjl} \) which takes the value 1 if the \( t \)-th observation for patient \( j \) of cluster \( i \) of stratum \( h \) is relevant and of type \( l \) and is from a patient of type \( k \), and 0 otherwise; and let \( y_{hikjl} \) be a binary outcome which takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) of stratum \( h \) is relevant, of type \( l \), for a patient of type \( k \), and has the attribute, and 0 otherwise. Also define

\[
y_{hik \cdot l} = \frac{\nu_{hi}}{\sum_{j=1}^{\nu_{hi}} \sum_{t=1}^{m_{hij}} y_{hikjl}} \quad \text{and} \quad x_{hik \cdot l} = \frac{\nu_{hi}}{\sum_{j=1}^{\nu_{hi}} \sum_{t=1}^{m_{hij}} x_{hikjl}}.
\]

Note that the summations over \( j \) and \( t \) encompass all \( \nu_{hikl} \) observations of type \( l \) for the patients in subgroup \( k \) of cluster \( i \). Also, for each stratum \( h \), the compound vectors \((y_{hi1 \cdot l}, x_{hi1 \cdot l}, \ldots, y_{hiK \cdot L}, x_{hiK \cdot L})'\) are independent and identically distributed as a consequence of the assumed sampling method for clusters within strata being simple random sampling with replacement. Then, the ratio mean for the subgroup of patients with the \( l \)-th subpatient-level observation type, \( k \)-th patient type, and \( h \)-th cluster type is given by

\[
R_{hkl} = \left( \frac{\sum_{i=1}^{N} y_{hik \cdot l}}{N} \right) / \left( \frac{\sum_{i=1}^{N} x_{hik \cdot l}}{N} \right)
\]

(2.26)

The quantity \( R_{hkl} \) is the estimated average number of relevant occurrences of the attribute for patients of type \( k \) with teeth of type \( l \) per cluster of type \( h \), divided by the average number of relevant observations for patients of type \( k \) with teeth of type \( l \) per cluster of type \( h \). It should be noted that some patients may only contribute a subset \( L^* \) of \( L \) subpatient types and that some clusters may only contribute a subset \( K^* \) of \( K \) patient types. The \( x_{hik \cdot l} \) can be
0 for some settings as long as their sums over i are clearly greater than 0.

For the dental data example, the ratio mean estimator for mandibular first molars for females aged 35 to 50 years old in urban clusters is the total number of crowns for mandibular first molars for females age 35 to 50 years old per urban cluster divided by the total number of mandibular first molars with restorations having greater than or equal to three surfaces for females aged 30 to 50 per urban cluster.

2.5.1 Matrix Notation

A matrix formulation for calculating ratio means for subgroups of observations defined by cluster-, patient-, and subpatient-level characteristics, simultaneously, will be presented in this section. The assumed sampling method for clusters is stratified simple random sampling with replacement. Also, the same subscripts as those used in the previous section will be used here. Let \( f_{hijkl} = (y_{hijkl}', x_{hijkl}')' \) be a bivariate binary observation where, \( y_{hijkl} \) is a binary outcome which takes the value 1 if the t-th observation of patient j of cluster i of stratum h is of type l, from a patient of type k, is relevant and has the attribute, and 0 otherwise; and \( x_{hijkl} \) is a binary outcome which takes the value 1 if the t-th observation for patient j of cluster i of type h is relevant and is of type l, and from a patient of type k, and 0 otherwise.

Let

\[
 f_{hijkl} = \sum_j \sum_i \left( y_{hijkl}', x_{hijkl}' \right)',
\]

\[
 f_{hi} = \left( f_{hi1}', f_{hi2}', ..., f_{hiKL}' \right), \text{ and}
\]
\[
\overline{I}_h = \frac{N_h}{\sum_{i=1}^{N_h} f_{hi}} / N_h.
\]

For each \( h \), the \( f_{hi} \) are independent and identically distributed as a consequence of the assumed sampling method for clusters being simple random sampling with replacement. An unbiased and, hence, consistent estimator of the variance of \( \overline{I}_h \) is given by

\[
V_{\overline{I}_h} = \frac{1}{N_h(N_h - 1)} \sum_{i=1}^{N_h} (f_{hi} - \overline{I}_h)(f_{hi} - \overline{I}_h)'.
\]  

(2.27)

Let

\[
\overline{I} = (\overline{I}_1', ..., \overline{I}_H'),
\]

so that

\[
V_{\overline{I}} = \text{Block} ( V_{\overline{I}_h} ),
\]  

(2.28)

where \( \text{Block}(\cdot) \) means a block diagonal matrix with the \( V_{\overline{I}_h} \) as diagonal blocks. The vector of ratio means resulting from the crossclassification of cluster-, patient-, and subpatient-level characteristics may be written as

\[
R = \exp \mathcal{A} \ln \overline{I},
\]  

(2.29)

where \( \mathcal{A} = [1, -1] \otimes I_{\text{HKL}} \),

\[ I_{\text{HKL}} \] is the HKL x HKL identity matrix, and "\( \otimes \)" is the symbol for the Kronecker product whereby the matrix \([1, -1]\) on the left multiplies each
element of the matrix $I_{HKL}$ on the right. As shown in Koch and Imrey et al (1985), the estimated variance of $R_J$ based on a first-order Taylor series approximation for $R_J$ is given by

$$V_R = H V_I H',$$  (2.30)

where

$$H = D_R A D_I^{-1},$$

and $D_R$ and $D_I$ are diagonal matrices with the elements of $R_J$ and $I_J$ as their respective diagonal elements.

2.6 **Ratio Mean Regression**

The number of pairwise comparisons of ratio means corresponding to a crossclassification of cluster-, patient-, and subpatient-level characteristics may be considerable. Furthermore, some of the cluster-, patient-, and subpatient-level characteristics may be correlated. Hence, the simultaneous examination of the effects of cluster-, patient-, and subpatient-level characteristics on the ratio mean through a regression procedure would be advantageous, and the ability to evaluate interactions between two or more characteristics would also be desirable. Such a regression procedure for ratio means (or functions of ratio means) is possible using the method of Grizzle, Starmer, and Koch (1969) (GSK) and extensions of this approach as discussed in Koch and Imrey et al (1985). This regression procedure will be presented in the remaining sections of this Chapter.

2.6.1 **A Linear Model for $R$**

Let $R = (R_{111}, \ldots, R_{HKL})'$ be a vector of ratio means for the subpopulations defined by the crossclassification of $H$ cluster-, $K$ patient-, and $L$
subpatient-level characteristics. The function vector $R$ may be modeled using the GSK method by directly forming $R$ and its covariance matrix (Koch, Freeman, and Freeman, 1975; Koch and Imrey et al., 1985). A linear model for $R$ is given by

$$E_A(R) = \chi \beta$$

where $E_A(\cdot)$ denotes asymptotic expected value, $\chi$ is the HKL×u design or model specification matrix and $\beta$ is a $u \times 1$ vector of unknown parameters to be estimated. The covariance matrix for $R$ may be consistently estimated by $V_R$ in (2.30). The weighted least squares estimator $b$ of $\beta$ is given by

$$b = (\chi'VR^{-1}R)^{-1}\chi'VR^{-1}R.$$  

(2.32)

In large samples, the estimator $b$ has approximately a normal distribution with

$$E_A(b) = \beta$$

(2.33)

and, for which a consistent estimator of the corresponding covariance matrix is

$$V_b = (\chi'VR^{-1}R)^{-1}. $$

(2.34)

The goodness of fit of the model in (2.31) can be evaluated with the residual weighted least squares statistic

$$Q = (R - \chi b)'VR^{-1}(R - \chi b).$$
When (2.31) applies, Q approximately has the chi-squared distribution with (HKL-u) degrees of freedom when all HKL subgroups have at least moderately large sample sizes.

2.6.2 A Linear Model for Logit(R)

Let

\[ \mathbf{\bar{F}} = (F_{111}, \ldots, F_{HKL})' \]

\[ = \logit \mathbf{R} \]

\[ = \ln (\mathbf{R}) - \ln (1 - \mathbf{R}), \]

be a vector of logits of the ratio means corresponding to the subpopulations defined by the crossclassification of the H cluster-, K patient-, and L subpatient-level characteristics. An estimate of the covariance matrix for \( \mathbf{\bar{F}} \), based on a first-order Taylor series approximation for \( \mathbf{\bar{F}} \), is given by

\[ \nabla_{\mathbf{\bar{F}}} = D_{\mathbf{\bar{F}}}^{-1} \left( \nabla (\mathbf{1} - \mathbf{R}) \right)^{-1} \nabla_{\mathbf{R}} \left( D_{\mathbf{R}} \mathbf{R} \right)^{-1} D_{\mathbf{\bar{F}}}^{-1} \]  

(2.35)

where \( D_{\mathbf{\bar{F}}} \) is a HKL×HKL matrix with the elements of the vector \( \mathbf{R} \) on the diagonal, \( D_{\mathbf{R}} \mathbf{R} \) is a HKL×HKL matrix with the elements of the vector \( (\mathbf{1} - \mathbf{R}) \) on the diagonal, and \( \nabla_{\mathbf{\bar{F}}} \) is the estimate of the covariance matrix for \( \mathbf{\bar{F}} \) given in (2.30). The function vector \( \mathbf{\bar{F}} \) may be modeled using the GSK method by directly forming \( \mathbf{\bar{F}} \) and its covariance matrix. A linear model for \( \mathbf{\bar{F}} \) is given by

\[ E_A(\mathbf{\bar{F}}) = X_{\mathbf{\bar{F}}} \]  

(2.36)
where $\mathbf{X}$ is the HKL x u design or model specification matrix and $\beta$ is a u x 1 vector of unknown parameters to be estimated. The covariance matrix of $\mathbf{F}$ may be consistently estimated by $\mathbf{V}_F$. The weighted least squares estimator $\hat{\beta}$ of $\beta$ is given by

$$
\hat{\beta} = (\mathbf{X}' \mathbf{V}_F^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}_F^{-1} \mathbf{F}.
$$

(2.37)

In large samples, the estimator $\hat{\beta}$ has approximately a normal distribution with

$$
E_A(\hat{\beta}) = \beta
$$

(2.38)

and for which

$$
\mathbf{V}_{\hat{\beta}} = (\mathbf{X}' \mathbf{V}_F^{-1} \mathbf{X})^{-1}
$$

(2.39)

is a consistent estimator of the covariance matrix.

2.7 Example

Using data from the dental study described in Section 1.1, the effects of the following factors: region of the country (Northeast, South, Central, and West), age of the patient (categorized as $\leq 34$, 35-50, and $\geq 51$), and tooth arch (mandibular and maxillary), on the proportion of restorations with greater than or equal to three surfaces which are crowns, were explored using the ratio mean regression (RMR) method. The sample consisted of 358 dental practices (clusters) selected from the population with replacement. The number of patients per sample cluster ranged from 5 to 757 with a median of 32. The number of relevant teeth per sample cluster ranged from 10 to 1010 with a median of 46. The median number of relevant teeth per person in the sample was 1, and ranged
from 1 to 13. The total number of patients in the sample was 18,335 and the total number of relevant teeth in the sample was 26,525.

Observed ratio means and their standard errors computed for each subpopulation formed by the crossclassification of region of the country, age of the patient, and tooth arch, are given in Table 2.1. The standard errors for each ratio mean were computed in two ways: by the Taylor series method and, for comparison purposes only, as if the data were from a simple random sample (SRS) with no correction for the clustering. The estimated covariance matrix for $R$, $\bar{R}$, was computed by the Taylor series method and is given in Table 2.2.

The following model for the $\logit(\bar{R})$ was fit to the dental data using the RMR method:

$$E(\bar{R}) = \beta_0 + \beta_1 \text{Region1} + \beta_2 \text{Region2} + \beta_3 \text{Region3} + \beta_4 \text{Age1} + \beta_5 \text{Age2} + \beta_6 \text{Mandib} + \beta_7 \text{Age1} \times \text{Mandib} + \beta_8 \text{Age2} \times \text{Mandib} \quad (2.40)$$

where: $\bar{R} = \logit(\bar{R})$; Region1, Region2, and Region3, are indicator variables for region 1, region 2, and region 3, respectively; Age1 and Age2 are indicator variables for age $\leq 34$ and age 35 to 50, respectively; Mandib is an indicator variable indicating the mandibular versus maxillary arch; and Age1 $\times$ Mandib and Age2 $\times$ Mandib are interaction terms.

Parameter estimates computed by the RMR method and their standard errors are given in Table 2.3. The analysis of variance (ANOVA) for the model using RMR is presented in Table 2.4. The goodness of fit (GOF) statistic given by the RMR method in Table 2.4 ($\chi^2 = 14.62$, df = 15, and p-value = 0.479) indicates that model (2.40) adequately represents the data. The ANOVA for the
RMR model shows that there is a significant interaction between age and tooth arch \( (\chi^2 = 20.36, \text{df} = 2, \text{p-value} < 0.001) \). It is noted that a model which excluded the interaction terms did not adequately represent the data \( (\chi^2 = 34.98, \text{df} = 17, \text{and p-value} < 0.006) \). The inadequacy of the model excluding the interaction terms may not have been detected using other approaches because of a lack of such a GOF statistic for these approaches.

Parameter estimates and standard errors for a model containing the same predictor variables fit to the data using survey data logistic regression, and also fit by ordinary logistic regression not accounting for the clustering, are also given in Table 2.3 for comparison. The model could not be fit to these data using the GEE1 method (with a mean-variance relationship \( \text{mean}(1-\text{mean}) \), a logit link, and an exchangable working correlation matrix) since the maximum cluster size allowed by the SAS macro (Karim and Zeger, 1988) to implement GEE1 was exceeded by this sample. Note that the survey data logistic regression results are equivalent to those which would be obtained from GEE1 with: a logit link, an independence working correlation matrix, and with \( \text{variance} = \text{mean} \times (1-\text{mean}) \) for the mean-variance relationship (LaVange, Keyes, Koch et al., 1993). The parameter estimates from the RMR model are similar to those from the survey logistic regression and ordinary logistic regression models. The RMR method may, in fact, yield more efficient parameter estimates because an estimate of the true covariance matrix of \( \beta \) is used for the estimation of \( \hat{\beta} \); the standard error for each parameter estimate is smaller for the RMR method than for the survey logistic regression (and, presumably GEE1) methods. They are not, however, smaller than the standard errors for cluster-level covariate parameters from the ordinary logistic regression model, as expected, since the RMR method takes into account the clustering of the data whereas the ordinary logistic regression does not. The standard errors for the RMR method and the other methods are similar
for the patient- and tooth-level covariates.

The predicted ratio means generated from the RMR model and their standard errors are given in Table 2.5. The standard errors for the model-predicted ratio means are smaller than those for the observed ratio means (which account for the clustering) in Table 2.1 because the RMR method uses all observations when calculating each standard error and not just the observations in a given subpopulation.

Adjusted odds ratios and 95% confidence intervals for selected factor-level comparisons are given in Table 2.6. It may be concluded from the model that crown use for restorations with greater than or equal to three surfaces is lower in region 1 than it is in regions 2, 3, and 4 which have similar levels of crown utilization. The odds of a crown are also greater for patients ≥51 years of age versus patients ≤34 years of age, for patients 35-50 years of age versus patients ≤34 years of age, and for patients ≥51 years of age versus patients 30-50 years of age. The odds ratios for these age comparisons are slightly larger for teeth on the maxillary arch; that is, there is an interaction between age and tooth arch.

2.8 Summary

The RMR method proposed in this chapter is a very general regression technique which may be used to analyze (multivariate) binary data from a one-stage cluster sample in which clusters have been selected by simple random sampling with replacement. It may be thought of as an alternative to survey data logistic regression and especially GEE1 when the cluster sizes are large. It may also be used more generally for outcomes with multiple categories which may or may not be ordered (the use of the RMR method for multiple category
outcomes is discussed in the next chapter). The methodology proposed in this chapter is also applicable in situations where the clusters are selected with unequal probability but still with replacement sampling. Further clarification of this issue is provided in Chapter IV. While the RMR method is particularly suited for the analysis of large clusters, it cannot be used when the number of clusters is small; each subpopulation should have at least 20 clusters. The RMR method also cannot accommodate continuous predictor values. Since continuous predictors often fail to be linear in the log odds for \( R \) and, hence, must be categorized, this may not, in fact, be much of a limitation. The RMR method yields more efficient estimates of model parameters than survey data logistic regression and GEE1 because it utilizes an estimate of the true covariance matrix for \( R \) for estimation. All three approaches: RMR, survey data logistic regression, and GEE1, give inconsistent results when the model for the (ratio) mean is misspecified. While statistical software packages may be used to generate the quantities needed to fit the RMR models, a dedicated software procedure that is easy to implement may be needed for the adoption of this approach in practice.
CHAPTER III
RATIO MEAN REGRESSION EXTENDED TO MULTICATEGORY, ORDINAL, AND CONTINUOUS OUTCOMES FROM A ONE-STAGE CLUSTER SAMPLE

3.1 Introduction

In this chapter, the ratio mean regression (RMR) method proposed in Chapter II for a cluster sample of multivariate binary outcomes will be extended to accommodate multivariate, ordinal, and continuous outcomes with large and unequally sized clusters. The ratio mean estimator, its standard error, and a large-sample test statistic for contrasting subsets of (correlated) ratio means defined by a cross-classification of explanatory factors, will be proposed for each type of outcome. Multiple regression models with categorical predictors will also be proposed for each type of outcome. The data from the dental study will, again, be used to illustrate the methods proposed in this chapter.

3.2 Multicategory Outcomes

This section will illustrate the ratio mean technique for the analysis of multicategory outcomes from a one-stage cluster sample with large and unequally sized clusters. Such multicategory outcomes are common in dental research. The binary outcome of crown versus no crown in Chapter II may be more precisely defined as a multi-level outcome with the following levels: crown, amalgam, plastic, no restoration, and no tooth. The topic for this section will be analysis
methods for this type of outcome from a one-stage cluster sample with large clusters of unequal sizes. The assumed sampling method for clusters is simple random sampling with replacement (or, equivalently, without replacement from a very large population). Within the selected clusters, all available units (e.g. sites for teeth) are observed for all subjects.

3.2.1 **Ratio Mean Definition**

Let \( i = 1, \ldots, N \) index sample clusters, \( j = 1, \ldots, \nu_i \) index patients in the \( i \)-th cluster, and \( t = 1, \ldots, m_{ij} \), index multiple observations for patient \( j \) in the \( i \)-th cluster. Here, \( N \) represents the number of clusters, \( \nu_i \) represents the total number of patients in the \( i \)-th cluster, and \( m_{ij} \) represents the total number of potential observations for patient \( j \) in the \( i \)-th cluster. Let \( x_{ijt} \) be a binary outcome which takes the value 1 if the \( t \)-th observation for the \( j \)-th patient in the \( i \)-th cluster is relevant, and 0 otherwise. A multicategory outcome with \( G \) levels will be represented by the set of binary indicators \( y_{gijt}, g = 1, \ldots, G \), where \( y_{gijt} \) takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) is relevant and has level \( g \), and 0 otherwise. Also define

\[
y_{gi..} = \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} y_{gijt} \quad \text{and} \quad x_{i..} = \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} x_{ijt}.
\]

Here, \( y_{gi..} \) is the total number of relevant observations with level \( g \) of the attribute in the \( i \)-th cluster and \( x_{i..} \) is the total number of relevant observations in the \( i \)-th cluster. Since the assumed sampling method for clusters is simple random sampling with replacement, the vectors \((y_{1i..}, y_{2i..}, \ldots, y_{Gi..}, x_{i..})\) are independent and identically distributed.
The ratio mean estimator for the proportion of relevant observations with level \( g \) over all clusters is defined as:

\[
R_g = \frac{\left( \sum_{i=1}^{N} y_{gi\ldots} / N \right)}{\left( \sum_{i=1}^{N} x_{i\ldots} / N \right)}
\]

\[
= \frac{\bar{y}_g}{\bar{x}}.
\]

where \( \bar{y}_g = \frac{\sum_{i=1}^{N} y_{gi\ldots}}{N} \) and \( \bar{x} = \frac{\sum_{i=1}^{N} x_{i\ldots}}{N} \).

\( R_g \) can be viewed as the average number of occurrences of level \( g \) among relevant observations per cluster divided by the average number of relevant observations per cluster.

The vector \( \bar{R} = (R_1, R_2, ..., R_G)' \) of the \( \{R_g\} \) can be obtained from the vector,

\[
\bar{f} = \frac{1}{N} \sum_{i=1}^{N} (y_{1i\ldots}, y_{2i\ldots}, ..., y_{Gi\ldots}, x_{i\ldots})'
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} f_i
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} (y_i', x_i')'
\]

\[
= (\bar{y}', \bar{x})'
\]
by the matrix operations

\[ R = \exp \left\{ [I_G, -I_G] \ln \bar{y} \right\} \]

where \( I_G \) is the \( G \times G \) identity matrix and \( I_G \) is a \( (G \times 1) \) vector of 1's.

A first-order Taylor series representation for \( R \) at the population mean \( \bar{\mu} = (\mu'_y, \mu'_x)' \) of the \( (\bar{y}', \bar{x}') \), where \( \mu'_y = (\mu_{y1}', \mu_{y2}', \ldots, \mu_{yG}') \), is given by

\[ R = \frac{\mu'_y}{\mu'_x} + \frac{1}{\mu'_x} (\bar{y} - \mu'_y) - \frac{\mu'_y}{\mu'_x^2} (\bar{x} - \mu'_x) + O(\frac{1}{N}) \] (3.2)

The asymptotic expected value of the first-order Taylor series for \( R \), \( E_A(R) \), is \( \frac{\mu'_y}{\mu'_x} = \theta \), the ratio mean for the \( G \) levels of the outcome in the population.

The covariance matrix for \( R \), based on a Taylor series linearization, the independent and identically distributed nature of the \( (\bar{y}_i', \ldots, \bar{x}_i') \), and the methods of Koch and Imrey et al. (1985), is given by

\[ \text{Var}_A(R) = D_\theta A D_\mu^{-1} [\text{Var}(\bar{y})] D_\mu^{-1} A' D_\theta \] (3.3)

where \( D_\theta \) and \( D_\mu \) are diagonal matrices with \( \theta \) and \( \mu \) as diagonal elements and \( A = [I_G, -I_G] \). A variance estimator for \( \text{Var}(R) \) is given by

\[ \text{var}(R) = D_R A D_\bar{y}^{-1} \bar{y} \bar{y}^{-1} A' D_R^{-1} \] (3.4)

where
\[ Y_{\bar{I}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (I_i - \bar{I})(I_i - \bar{I})' . \tag{3.5} \]

Since

\[ Y_{\bar{I}} = \begin{bmatrix} \bar{y} & \bar{y}, \bar{x} \\ \bar{y}', \bar{x} & \bar{v}_x \end{bmatrix} \]

where

\[ \bar{v}_y = \frac{1}{N(N-1)} \sum_{i=1}^{N} (y_i - \bar{y})(y_i - \bar{y})' \]

\[ \bar{v}_y, x = \frac{1}{N(N-1)} \sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x}) \]

\[ v_x = \frac{1}{N(N-1)} \sum_{i=1}^{N} (x_i - \bar{x})^2 , \]

one can alternatively express \( \text{var}(R_\lambda) \) as

\[
\text{var}(R_\lambda) = \mathcal{D}_{\bar{R}} \{ (\mathcal{D}_{\bar{Y}}^{-1} v_{\bar{y}} \mathcal{D}_{\bar{Y}}^{-1}) - (\mathcal{D}_{\bar{Y}}^{-1} v_{\bar{y}}, x \lambda G')/\bar{x} - (\lambda G v'_{\bar{y}}, x \mathcal{D}_{\bar{Y}}^{-1})/\bar{x} \\
+ \lambda G \lambda' G v_x/\bar{x}^2 \} \mathcal{D}_{\bar{R}}
\tag{3.6}
\]

This form for \( \text{var}(R_\lambda) \) is a matrix counterpart to expression (2.4) for a single ratio.

A large sample variance estimator for \( \text{Var}(\ln R_\lambda) \) based on a Taylor series linearization is given by
\[
\text{var}(\ln R_g) = D^{-1}_R \text{var}(R_g) D^{-1}_R.
\] (3.7)

A large sample \((1 - \alpha)-\text{level confidence interval for } \theta_g \text{ in } \theta \) is given by:

\[
\exp \left\{ \ln R_g \pm Z_{(1 - \alpha/2)} \cdot (\text{var}(R_g)^{1/2}/R_g) \right\},
\] (3.8)

where the \(\text{var}(R_g)\) are diagonal elements of \(\text{var}(R)\). Similarly, a large sample \((1 - \alpha)-\text{level confidence interval for } (\theta_g/\theta_{g'}) \text{ is given by}

\[
\exp \{ \ln(R_g/R_{g'}) \pm Z_{(1 - \alpha/2)} \text{var} [\ln(R_g/R_{g'})]^{1/2} \}
\] (3.9)

where

\[
\text{var}[\ln(R_g/R_{g'})] = \left\{ \frac{\text{var}(R_g)}{R^2_g} - \frac{2 \text{Cov}(R_g,R_{g'})}{R_g R_{g'}} + \frac{\text{var}(R_{g'})}{R^2_{g'}} \right\}
\] (3.10)

with \(\text{cov}(R_g, R_{g'})\) being the estimated covariance of \(R_g\) and \(R_{g'}\) from (3.).

3.2.2 **Ratio Means Defined by Cluster-, Patient-, and Subpatient-Level Characteristics**

Ratio means for multiclass outcomes may be defined for subsets of cluster-, patient-, and subpatient-level characteristics in a similar fashion to those defined in Chapter II for binary outcomes. Let \(h=1,\ldots,H\) index levels for the
crossclassification of cluster-level characteristics, \( k = 1, \ldots, K \) index levels for the crossclassification of patient-level characteristics, \( l = 1, \ldots, L \) index levels for the crossclassification of subpatient-level characteristics. Also let \( i = 1, \ldots, N_h \) index sample clusters in stratum \( h \), \( j = 1, \ldots, \nu_{hi} \) index patients in cluster \( i \) of stratum \( h \), \( t = 1, \ldots, m_{hij} \) index multiple observations for patient \( j \) in cluster \( i \) of stratum \( h \). Here, \( N_h \) is the number of sample clusters in the \( h \)-th stratum, \( \nu_{hi} \) is the total number of patients in cluster \( i \) of stratum \( h \), \( m_{hij} \) is the total number of potential observations for patient \( j \) in cluster \( i \) of stratum \( h \), \( H \) is the number of levels for the crossclassification of cluster-level characteristics, \( K \) is the number of levels for the crossclassification of patient-level characteristics, and \( L \) is the number of levels for the crossclassification of subpatient-level characteristics. The assumed sampling method for clusters is stratified simple random sampling (with replacement); i.e., simple random sampling for each stratum and independence of sampling in different strata.

Define a binary outcome \( x_{hikjlt} \) which takes the value 1 if the \( t \)-th observation for patient \( j \) of cluster \( i \) of stratum \( h \) is relevant and of type \( l \) and is from a patient of type \( k \), and 0 otherwise. Also define \( y_{ghikjlt} \) as a binary outcome which takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) of stratum \( h \) is relevant, and is of type \( l \), for a patient of type \( k \), and has level \( g \) of the multiclass outcome, and 0 otherwise. Also define

\[
y_{ghik.l} = \sum_{j=1}^{\nu_{hi}} \sum_{t=1}^{m_{hij}} y_{ghikjlt} \quad \text{and} \quad x_{hik.l} = \sum_{j=1}^{\nu_{hi}} \sum_{t=1}^{m_{hij}} x_{hikjlt}.
\]

Note that the summations over \( j \) and \( t \) encompass all \( \nu_{hikl} \) observations of type \( l \) for patients in subgroup \( k \) of cluster \( i \) in stratum \( h \). The \( x_{hik.l} \) are the
corresponding total numbers of observations which are relevant and the \( y_{ghikl} \) are the corresponding total numbers which are relevant and have level \( g \) of the multicategory outcome. Also, for each stratum, the assumption of stratified simple random sampling of entire clusters implies that the compound vectors 
\[
(y_{1hi1l}, \ldots, y_{Ghi1l}, x_{hi1l}, \ldots, y_{1hiKl}, \ldots, y_{GhiKl}, x_{hiKl})' \]
are independent and identically distributed. Then the ratio mean for level \( g \) for the subgroup of patients with the \( l \)-th subpatient-level observation type, \( k \)-th patient type, and \( h \)-th cluster type is given by

\[
R_{ghkl} = \left( \frac{\sum_{i=1}^{N} y_{ghikl}}{N} \right) / \left( \frac{\sum_{i=1}^{N} x_{hikl}}{N} \right)
\]  
(3.11)

The quantity \( R_{ghkl} \) is the estimated average number of relevant occurrences of level \( g \) of the multicategory outcome for patients of type \( k \) with teeth of type \( l \) per cluster of type \( h \), divided by the average number of relevant observations for patients of type \( k \) with teeth of type \( l \) per cluster of type \( h \). It should be noted that some patients may only contribute a subset of \( L \) subpatient types and that some clusters may only contribute a subset of \( K \) patient types. The \( x_{hikl} \) can be 0 for some settings as long as their sums over \( i \) are clearly greater than 0.

Matrix notation will be useful for describing subsets of ratio means defined by cluster-, patient-, and subpatient-level characteristics, and will be used to describe the regression model discussed in the next section.

Let
\[
\mathbf{f}_{hijkl} = (y_{1hijkl}, y_{2hijkl}, \ldots, y_{Ghijkl}, x_{hijkl})'
\]
be a vector of binary random variables where, \( y_{ghikjl} \) is a binary outcome which takes the value 1 if the \( t \)-th observation of patient \( j \) of cluster \( i \) of stratum \( h \) is of type \( l \) and from a patient of type \( k \) is relevant and has level \( g \) of the multivariate outcome, and 0 otherwise, and let \( x_{hijkl} \) be a binary outcome which takes the value 1 if the \( t \)-th observation for patient \( j \) of cluster \( i \) of type \( h \) is relevant and is of type \( l \) and from a patient of type \( k \), and 0 otherwise.

Let

\[
f_{hi} = \sum_{j} \sum_{l} (y_{1hikjl}, y_{2hikjl}, \ldots, y_{Ghikjl}, x_{hikjl}'.
\]

For the compound vector,

\[
f_{hi} = (f'_{hi11}, f'_{hi12}, \ldots, f'_{hiKL}'),
\]

express the sample mean for the \( h \)-th stratum as

\[
\bar{f}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} f_{hi} / N_h.
\]

An unbiased and consistent estimator of the covariance matrix of \( \bar{f}_h \) is given by

\[
\Sigma_{\bar{f}_h} = \frac{1}{N_h(N_h - 1)} \sum_{i=1}^{N_h} (f_{hi} - \bar{f}_h)(f_{hi} - \bar{f}_h)' \tag{3.12}
\]

since the \( f_{hi} \) are independent and identically distributed (by virtue of the simple
random sampling of entire clusters from the respective strata).

Let

\[ \tilde{I} = (\tilde{I}_1', ..., \tilde{I}_H')', \]

so that

\[ \nabla_{\tilde{I}} = \text{Block} (\nabla_{\tilde{I}_h}). \]

(3.13)

The vector of ratio means for all levels \( g = 1, ..., G \) resulting from the crossclassification of cluster-, patient-, and subpatient-level characteristics may be written as

\[ R = \exp A \ln \tilde{I}, \]

(3.14)

where \( A = [I_{G \times G}, -I_{G \times 1}] \otimes I_{HKL}, \)

\( I_{G \times G} \) is the \( G \times G \) identity matrix, \(-I_{G \times 1}\) is a \( G \times 1 \) vector containing the element \(-1\) for every entry, and \( I_{HKL} \) is the \( HKL \times HKL \) identity matrix. The symbol "\( \otimes \)" represents the Kronecker product whereby the matrix, \( [I_{G \times G} \parallel -I_{G \times 1}] \), on the left multiplies each element of the matrix, \( I_{HKL} \), on the right. As shown in Koch and Imrey et al (1985), the variance of \( R \) based on a first-order Taylor series approximation for \( R \) is given by

\[ \nabla_{R} = H \nabla_{\tilde{I}} H', \]

(3.15)

where \( H = D_R A D_{\tilde{I}}^{-1}. \)
with $\mathbb{D}_R$ and $\mathbb{D}_I$ being diagonal matrices with the elements of $R_0$ and $I$ as their respective diagonal elements.

A regression analysis for the simultaneous examination of the effects of cluster-, patient-, and subpatient-level characteristics and their interactions on the ratio mean (or functions of ratio means) is possible using the method of Grizzle, Starmer, and Koch (1969) (GSK) and extensions of this approach as discussed in Koch and Imrey et al (1985). Such a regression procedure will be presented in the following section.

3.2.3 Ratio Mean Regression Model For Clustered Multicategory Outcomes

The ratio mean regression model will initially be written for the comparison of the ratio mean for level $g$ versus the ratio mean for level $G$ of the multicategory outcome, $y_{ghikjl}$.

Let

$$E_{gG} = (F_{g111}, ..., F_{gHKL})'$$

be a vector of the natural logarithms of the ratio of $R_0$ divided by $R_G$ for each subpopulation defined by the crossclassification of the $H$ cluster-, $K$ patient-, and $L$ subpatient-level characteristics; i.e.,

$$E_{gG} = A_{gG} \ln(R_g)$$

where

$$A_{gG} = [0, ..., 1, ..., -1] \otimes I_{HKL},$$
with the 1 in \([0, \ldots, 1, \ldots, -1]\) being in the \(g\)-th column. An estimate of the covariance matrix for \(\mathbf{F}_{gG}\), based on a first-order Taylor series approximation for \(\mathbf{F}_{gG}\), is given by

\[
\mathcal{V}_{\mathbf{F}_{gG}} = \mathbf{A}_{gG} \mathbf{D}_R^{-1} \mathcal{Y}_{\mathbf{R}} \mathbf{D}_R^{-1} \mathbf{A}_{gG}',
\]  

(3.16)

where \(\mathbf{D}_R\) is a HKL\(\times\)HKL matrix with the elements of the vector \(\mathbf{R}\) on the diagonal, \(\mathbf{A}_{gG}\) is the matrix described above, and \(\mathcal{Y}_{\mathbf{R}}\) is the estimate of the covariance matrix for \(\mathbf{R}\) given in (3.15). The function vector \(\mathbf{F}_{gG}\) is a vector of log-odds for outcome \(g\) relative to outcome \(G\) and so may be modeled using the GSK method. For this purpose, one directly forms \(\mathbf{F}_{gG}\) and its covariance matrix. A linear model for \(\mathcal{F}_{gG}\) is given by

\[
\mathbf{E}_{A}(\mathbf{F}_{gG}) = \mathbf{X} \beta_{gG}
\]  

(3.17)

where \(\mathbf{X}\) is the HKL\(\times\)u design or model specification matrix and \(\beta_{gG}\) is a \(u \times 1\) vector of unknown parameters to be estimated. The weighted least squares estimator \(\hat{\beta}_{gG}\) of \(\beta_{gG}\), relative to the consistent estimator \(\mathcal{Y}_{\mathbf{F}_{gG}}\) for \(\mathbf{F}_{gG}\), is given by

\[
\hat{\beta}_{gG} = (\mathbf{X}' \mathcal{Y}_{\mathbf{F}_{gG}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathcal{Y}_{\mathbf{F}_{gG}}^{-1} \mathbf{F}_{gG}.
\]  

(3.18)

In large samples, the estimator \(\hat{\beta}_{gG}\) has approximately a multivariate normal distribution with
\[ E_A(b_g G) = \beta_{gG} \]  

(3.19)

and a covariance matrix which is consistently estimated with

\[ V_A(b_g G) = (X' V_{FG}^{-1} X)^{-1}. \]  

(3.20)

More generally, this type of model can be applied to \( F_{gG} \) for all \( g = 1, 2, \ldots, (G-1) \) simultaneously. For this log-linear model, the analysis is focused on the functions

\[ \mathcal{E} = \{ [\mathbf{G}_{-1}, -\mathbf{G}_{-1}] \otimes \mathbf{I}_{HKL} \} \ln \mathbf{R} \]

\[ = \mathbf{A}_G \ln \mathbf{R}. \]  

(3.21)

A consistent estimate for the covariance matrix of \( \mathcal{E} \) is

\[ V_{\mathcal{E}} = \mathbf{A}_G D_R^{-1} V_R D_R^{-1} A'_G. \]  

(3.22)

For \( \mathcal{E} \), a linear model can be specified as

\[ E_A(\mathcal{E}) = X_F \beta_F \]  

(3.23)

where \( X_F \) is the HKL(\( G-1 \)) \( \times \) \( u_F \) design or model specification matrix and \( \beta_F \) is the \( (u_F \times 1) \) vector of unknown parameters to be estimated. The weighted least squares estimates \( \hat{b}_F \) of \( \beta_F \) are given by
\[ b_{\mathcal{F}} = (X_{\mathcal{F}}' \gamma_{\mathcal{F}}^{-1} X_{\mathcal{F}})^{-1} X_{\mathcal{F}}' \gamma_{\mathcal{F}}^{-1} X_{\mathcal{F}} . \]  \tag{3.24}

A family of models which are often of interest for \( \mathcal{F} \) have

\[ X_{\mathcal{F}} = \mathcal{I}_{(G-1)} \otimes X, \]

where \( X \) is an \( (HKL \times u) \) matrix of explanatory variables which are common to the respective \( \mathcal{F}_{g \mathcal{G}} \) for \( g = 1, 2, ..., (G-1) \) which comprise \( \mathcal{F} \). In particular,

\[ \mathcal{F}_{g \mathcal{G}} = \{(0, ..., 1, ..., 0) \otimes I_{HKL}\} \mathcal{F} \]  \tag{3.25}

where the 1 in \( [0, ..., 1, ..., 0] \) is in the \( g \)-th column for \( g = 1, 2, ..., (G-1) \). The models of this type correspond to the simultaneous specification of the model (3.17) to all \( \mathcal{F}_{g \mathcal{G}} \) for \( g = 1, 2, ..., (G-1) \). For this reason, their corresponding \( \beta_{\mathcal{F}} \) is comprised of the \( \beta_{g \mathcal{G}} \) in (3.13) through

\[ \beta_{g \mathcal{G}} = \{(0, ..., 1, ..., 0) \otimes I_{u}\} \beta ; \]  \tag{3.26}

i.e., the \( (u_{\mathcal{F}} \times 1) \) vector \( \beta \) is comprised of \( (G-1) \) vectors \( \beta_{g \mathcal{G}} \) which are each \( (u \times 1) \). Also, the respective elements of \( \beta_{g \mathcal{G}} \) are the \( g \)-th, \( (G-1+g) \)-th, ..., \( (u-1)(G-1)+g \)-th elements of \( \beta \).

The parameters in this type of model can be alternatively expressed as

\[ E_{\mathcal{A}} \{ F_{ghkl} \} = E_{\mathcal{A}} \{ (R_{ghkl}/R_{Ghkl}) \} \]
\[ = \exp(\mathbf{Z}^\prime_{hkl} \mathbf{g}^G), \]  

where \( \mathbf{Z}^\prime_{hkl} \) is the (hkl)-th row of \( \chi^i \); i.e., it is the row of \( \chi^i \) which indicates explanatory variables for the combinations of the h-th stratum, k-th patient group, and l-th sub-patient level. This alternative specification is equivalent to the multicategory log-linear model

\[ \theta_{ghkl} = E_A \{ R_{ghkl} \} \]

\[ = \frac{\exp(\mathbf{Z}^\prime_{hkl} \mathbf{g}^G)}{\sum_{g=1}^{G} \exp(\mathbf{Z}^\prime_{hkl} \mathbf{g}^G)} \]  

for \( g = 1, 2, \ldots, G \) for which \( \mathbf{g}^G \) is used as a convention so that \( \exp(\mathbf{Z}^\prime_{hkl} \mathbf{g}^G) = 1 \); here \( \theta_{ghkl} \) is the population proportion with the g-th level of the outcome for relevant observations in the hkl-th subgroup. Thus, odds ratios of the form

\[ \frac{\theta_{ghkl} \theta_{gh'k'l'}}{\theta_{gh'k'l'} \theta_{ghkl}} = \psi_{ghkl, gh'k'l'} \]

can be expressed as

\[ \exp\left\{ (\mathbf{Z}_{hkl} - \mathbf{Z}_{h'k'l'}) (\mathbf{g}^G - \mathbf{g'}^G) \right\} \]
Also, when \( g' = G \), this expression simplifies to

\[
\exp\{((Z_{hkl} - Z_{h'k'l'})/\beta_{gG})\}
\]

since \( \beta_{GG} = 0 \) by convention. These odds ratios represent the extent to which the odds of the \( g \)-th versus \( g' \)-th outcome is greater for the \( hkl \)-th subgroup than for the \( h'k'l' \)-th subgroup. Estimates and confidence intervals for these odds ratios are obtained through their estimates from \( b_F \) and corresponding variances from \( \mathcal{V}(b_F) = (X_F' X_F)^{-1} \). In particular,

\[
(Z_{hkl} - Z_{h'k'l'})(b_{gG} - b_{g'G}) = (Z_{khl} - Z_{h'k'l'}) \{[0, ..., 1, ..., -1, ..., 0] \odot \mathcal{L} u \} b_F
\]

\[
= \xi'_{ghk'l'} b_F
\]

\[
= \xi b_F
\]

(3.29)

with subscripts for \( \xi \) dropped for convenience; also, in \([0, ..., 1, ..., -1, ..., 0]\), the 1 is in the \( g \)-th position and the -1 is in the \( g' \)-th position of this \((1 \times (G - 1))\) vector. Note that if \( g' = G \), then the -1 element is omitted. It then follows that a large sample \((1 - \alpha)\)-level confidence interval for the odds ratio \( \psi_{ghk'l'} \) is

\[
\exp\{\xi b_F \pm Z_{(1 - \alpha/2)}(\xi b_F \mathcal{V}(b_F) \xi)^{1/2}\}
\]

(3.30)
3.3 **Ordinal Outcomes**

Ordinal outcomes selected from a cluster sample may also be analyzed by a modification of the ratio mean approach proposed in Chapter II for clustered binary outcomes. The multicategory outcome with levels: crown, crown substitute, and no crown, defined in the previous section, may, alternatively, be considered as an example of an ordinal outcome and will be treated as such in this section.

3.3.1 **Ratio Mean Definition**

Let $i=1,\ldots,N$ index sample clusters, $j=1,\ldots,\nu_i$ index patients in the $i$-th cluster, and $t=1,\ldots,m_{ij}$, index multiple observations for patient $j$ in the $i$-th cluster, where $N$ represents the number of sampled clusters, $\nu_i$ represents the number of patients in the $i$-th cluster, and $m_{ij}$ represents the number of observations for patient $j$ in the $i$-th cluster. Let $x_{ijt}$ be a binary outcome which takes the value 1 if the t-th observation for the j-th patient in the i-th cluster is relevant, and 0 otherwise. An ordinal outcome with $G$ ordered categories will be represented by the set of binary indicators $y_{gijt}$, $g=1,\ldots,G$, where $y_{gijt}$ takes the value 1 if the t-th observation for patient $j$ in cluster $i$ is relevant and has level $\leq g$, and 0 otherwise. Also define

$$y_{gi} = \frac{\nu_i}{\nu_i} \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} y_{gijt}$$

and

$$x_i = \frac{\nu_i}{\nu_i} \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} x_{ijt}.$$
Here, $\tilde{y}_{gi\ldots}$ is the total number of relevant observations in the i-th cluster with levels $\leq g$ for the attribute and $x_{i\ldots}$ is the total number of relevant observations in the i-th cluster. Let

$$\tilde{R}_{g} = \frac{(\sum_{i=1}^{N}\tilde{y}_{gi\ldots}/N)}{(\sum_{i=1}^{N}x_{i\ldots}/N)}$$

$$= \frac{\tilde{y}_{g}}{\bar{x}},$$

where

$$\tilde{y}_{g} = \frac{\sum_{i=1}^{N}\tilde{y}_{gi\ldots}}{N} \quad \text{and} \quad \bar{x} = \frac{\sum_{i=1}^{N}x_{i\ldots}}{N}.$$

The ratio mean, $\tilde{R}_{g}$, estimates the proportion of relevant observations with level $\leq g$ of the ordinal outcome over all clusters in the population; also, it can be expressed as:

$$\tilde{R}_{g} = \sum_{g'=1}^{g} R_{g'}$$

(3.31)

where $R_{g'}$ is the ratio mean in (3.1) for the $g'$-th outcome. From its structure, $\tilde{R}_{g}$ can be viewed as the average number of occurrences of a level $\leq g$ for the ordinal outcome among relevant observations per cluster divided by the average number of relevant observations per cluster.

Since $\tilde{R}_{g}$ is the sum of $R_{g'}$ for $g' \leq g$, it follows that

$$E\{\tilde{R}_{g}\} = \sum_{g'=1}^{g} \theta_{g'} = \theta_{g}.$$

Also, the variance for $\tilde{R}_{g}$ from Taylor series linearization and the methods in
Koch and Imrey et al. (1985) is

\[ \text{Var}_A(\tilde{R}_g) = \xi' \text{Var}_A(\tilde{R}) \xi \]  

(3.32)

where \( \xi' = [1, 1, ..., 1, 0, ..., 0] \) is a row vector for which the first \( g \) elements are 1's and the remaining elements are 0's and \( \text{Var}_A(\tilde{R}) \) is the Taylor series linearization covariance matrix for \( \tilde{R} \) in Section 3.2.1. A consistent estimator for \( \text{Var}_A(\tilde{R}_g) \) is similarly given by

\[ \text{var}(\tilde{R}_g) = \xi' [ \text{var}(\tilde{R}) ] \xi \]  

(3.33)

where \( \text{var}(\tilde{R}) \) is the estimated covariance matrix for \( \tilde{R} \) in (3.4). More comprehensively, for \( g = 1, 2, ..., (G - 1) \), the vector \( \tilde{\xi} = (\tilde{R}_1, \tilde{R}_2, ..., \tilde{R}_{G-1})' \) can be obtained from \( \tilde{R} \) via the linear operations \( \tilde{\xi} = A_c \tilde{R} \) where

\[
A_c = \begin{bmatrix}
1 & 0 & 0 & 0 & . & . & . & 0 & 0 \\
1 & 1 & 0 & 0 & . & . & . & 0 & 0 \\
1 & 1 & 1 & 0 & . & . & . & 0 & 0 \\
. & . & . & . & . & . & . & . & . \\
1 & 1 & 1 & 1 & . & . & . & 1 & 0
\end{bmatrix}
\]

Thus, \( E_A(\tilde{\xi}) = A_c \xi = \hat{\xi} \) and

\[ \text{Var}_A(\tilde{R}) = A_c [ \text{Var}_A(\tilde{R}) ] A_c' \]  

(3.34)
Also, a consistent estimator for \( \text{Var}_A(\hat{\tau}) \) is

\[
\text{var}(\hat{\tau}) = A_c \left[ \text{var}(\hat{\tau}) \right] A_c^t = \mathcal{V}_{\hat{\tau}}. 
\]

(3.35)

As in Section 2.6.2, \( \hat{\tau} \) can be transformed to logits via

\[
\text{logit} \hat{\tau} = \ln \left( \frac{\hat{\tau}}{1 - \hat{\tau}} \right).
\]

(3.36)

By the methods leading to (2.34), a consistent estimator for the covariance matrix of logit \( \hat{\tau} \) is

\[
\text{var} (\text{logit}(\hat{\tau})) = \mathcal{D}_{\hat{\tau}}^{-1} \left( \mathcal{I} - \mathcal{D}_{\hat{\tau}} \right)^{-1} \mathcal{V}_{\hat{\tau}} \left( \mathcal{I} - \mathcal{D}_{\hat{\tau}} \right)^{-1} \mathcal{D}_{\hat{\tau}}^{-1}.
\]

(3.37)

3.3.2 Ratio Means Defined by Cluster-, Patient-, and Subpatient-Level Characteristics

Ratio means for ordinal outcomes may be defined for subsets of cluster-, patient-, and subpatient-level characteristics in a fashion similar to those defined for the binary outcome discussed in Chapter II. Let \( h = 1, \ldots, H \) index levels for the crossclassification of cluster-level characteristics, \( k = 1, \ldots, K \) index levels for the crossclassification of patient-level characteristics, \( l = 1, \ldots, L \) index levels for the crossclassification of subpatient-level characteristics, \( i = 1, \ldots, N_h \) index sample clusters in stratum \( h \), \( j = 1, \ldots, n_{hi} \) index patients in cluster \( i \) of stratum \( h \), \( t = 1, \ldots, m_{hij} \) index multiple observations for patient \( j \) in cluster \( i \) of stratum \( h \).
Here, \( N_h \) is the number of sample clusters in the \( h \)-th stratum, \( \nu_{hi} \) is the total number of patients in cluster \( i \) of stratum \( h \), \( m_{hij} \) is the total number of potential observations for patient \( j \) in cluster \( i \) of stratum \( h \), \( H \) is the number of levels for the crossclassification of cluster-level characteristics, \( K \) is the number of levels for the crossclassification of patient-level characteristics, and \( L \) is the number of levels for the crossclassification of subpatient-level characteristics. Define a binary outcome \( x_{hijkl} \) which takes the value 1 if the \( t \)-th observation for patient \( j \) of cluster \( i \) of stratum \( h \) is relevant and of type \( l \) and is from a patient of type \( k \), and 0 otherwise. Also define \( \bar{Y}_{ghikjl} \) as a binary outcome which takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) of stratum \( h \) is relevant, of type \( l \), for a patient of type \( k \), and has level \( \leq g \), and 0 otherwise. Also define

\[
\bar{Y}_{ghik \cdot l} = \frac{\nu_{hi}}{N_h} \sum_{j=1}^{m_{hij}} \sum_{t=1}^{\nu_{hi}} \bar{Y}_{ghikjt} ,
\]

\[
x_{hik \cdot l} = \frac{\nu_{hi}}{N_h} \sum_{j=1}^{m_{hij}} \sum_{t=1}^{\nu_{hi}} x_{hikjt} , \text{ and}
\]

\[
\hat{R}_{ghkl} = \left( \sum_{i=1}^{N_h} \bar{Y}_{ghik \cdot l} / N_h \right) / \left( \sum_{i=1}^{N_h} x_{hik \cdot l} / N_h \right).
\]

It should be noted that some patients may only contribute a subset of \( L \) subpatient types and that some clusters may only contribute a subset of \( K \) patient types. The \( x_{hik \cdot l} \) can be 0 for some settings as long as their sums over \( i \) are clearly greater than 0.

The ratio mean estimator \( \hat{R}_{ghkl} \) is the proportion of relevant
observations with level \( \leq g \) of the ordinal outcome for the subgroup of patients with the \( l \)-th subpatient-level observation type, \( k \)-th patient type, and \( h \)-th cluster type. Also

\[
\tilde{R}_{ghkl} = \sum_{g' = 1}^{g} R_{g'hlk}
\]  

(3.38)

where the \( R_{g'hlk} \) are the ratio means for the respective outcomes \( g' = 1, 2, \ldots, G \) individually.

Matrix notation will be useful for analyses with models for describing subsets of ratio means defined by cluster-, patient-, and subpatient-level characteristics. For this purpose, let

\[
\mathbf{R}_{hkl} = (R_{1hkl}, R_{2hkl}, \ldots, R_{Ghkl})^T
\]

be the vector of ratio means for outcomes 1, 2, \ldots, \( G \) for the \((hkl)\)-th subgroups. Also, let

\[
\mathbf{R} = (\mathbf{R}_{111}, \ldots, \mathbf{R}_{HKL})^T
\]

be the entire vector of ratio means for all outcomes in all subgroups. The vector \( \mathbf{R} \) is determined from the vector of means of indicator variables, \( \mathbf{I} \), in Section 3.2.2 by 3.14. From \( \mathbf{R} \), the vector \( \tilde{\mathbf{R}} \) of ratios \( \tilde{R}_{ghkl} \) is determined by the matrix operations

\[
\tilde{\mathbf{R}} = (A \odot I_{HKL}) \mathbf{R}
\]
\[ = \tilde{A} c \tilde{C} \]  
(3.39)

where \( \tilde{A} c \) is the matrix defined in Section 3.3.1. Thus, by linear Taylor series methods, a consistent estimator for the covariance matrix of \( \tilde{C} \) is

\[ V_{\tilde{C}} = \tilde{A} c V_{\tilde{R}} \tilde{A}^\prime c , \]  
(3.40)

where \( V_{\tilde{R}} \) is defined in (3.15). As noted previously, models for

\[ \text{logit} \tilde{C} = \ln \left( \frac{\tilde{C}}{1 - \tilde{C}} \right) \]  
(3.41)

are of interest. For \( \text{logit}(\tilde{R}) \), a consistent estimator for the covariance matrix is

\[ \text{var}(\text{logit}(\tilde{R})) = P \tilde{R}^{-1} (\tilde{U} - P \tilde{R})^{-1} V_{\tilde{R}} (\tilde{U} - P \tilde{R})^{-1} P \tilde{R}^{-1} . \]  
(3.42)

Thus, in reference to logits for the proportions of outcomes \( g \) among relevant observations for the respective subgroups, hkl, statistical models can be fit by weighted least squares. For this type of analysis, one works with functions \( \tilde{F} = \text{logit} \tilde{C} \) and a consistent estimate \( V_{\tilde{F}} = \text{var}(\text{logit} \tilde{C}) \) for their covariance matrix.

A linear model for \( \tilde{F} \) is given by

\[ E_A(\tilde{F}) = X_F \tilde{C} \]  
(3.43)

where \( X_F \) is the HKL\((G - 1) \times u_F \) design or model specification matrix and \( \tilde{C} \) is
a \( u_F \times 1 \) vector of unknown parameters to be estimated. The weighted least squares estimator \( \hat{b}_F \) of \( \beta_F \) is given by

\[
\hat{b}_F = (X_F'X_F)^{-1}X_F'Y_F. \tag{3.44}
\]

In large samples, the estimator \( \hat{b}_F \) has approximately a normal distribution with

\[
E_A(\hat{b}_F) = \beta_F \tag{3.45}
\]

and a covariance matrix which is consistently estimated by

\[
\Sigma_{\hat{b}_F} = (X_F'X_F)^{-1}. \tag{3.46}
\]

For ordinal data, one model of usual interest is the proportional odds model. This model has

\[
X_F = [\lambda(G-1)^{\otimes 1} \lambda_{HKL} \lambda(G-1)^{\otimes X}],
\]

where \( X \) is an \((HKL \times (u-1))\) matrix of explanatory variables in reference to the HKL subgroups and which are commonly applicable to all \((G-1)\) cumulative logits. For this model, the first \((G-1)\) parameters in \( \beta_F \) are intercepts and the last \((u-1)\) are slopes for the \((u-1)\) explanatory variables. The exponentials of these parameters are interpretable as odds ratios per unit change in the explanatory variables for all odds of outcomes \( \leq g \) versus those \( > g \) for all \( g \).

The statistic
\[ Q = (\mathbf{F} - \mathbf{X}_F \mathbf{b}_F)^\prime \mathbf{X}_F^{-1} (\mathbf{F} - \mathbf{X}_F \mathbf{b}_F) \]  

(3.47)

can be used to evaluate goodness of fit of the model. For an appropriate model, it approximately has the chi-squared distribution with \((HKL(G-1)-u_F)\) degrees of freedom. Through this goodness of fit test, one can assess the structure of proportional odds as well as the role of interactions.

3.4 Continuous Outcomes

The ratio mean technique proposed in Chapter II for clustered multivariate binary data may also be modified for the analysis of clustered continuous multivariate outcomes. A variable of interest for the dental study which is continuous is the cost of a restoration. The analysis of this type of variable will be explored in the remaining sections of this chapter.

3.4.1 Ratio Mean Definition

The analysis of clustered multivariate continuous outcomes may be carried out using the methodology proposed in Chapter II if the continuous outcome of interest is defined as follows. Let \( s_{ijt} \) be the continuous outcome of interest for the \( t \)-th observation for the \( j \)-th patient of the \( i \)-th cluster. Let \( y_{ijt} \) equal \( s_{ijt} \) if \( x_{ijt} = 1 \), and 0 otherwise, where \( x_{ijt} \) takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) is relevant, and 0 otherwise. The ratio mean methods proposed in Chapter II can be used to analyze \( y_{ijt} \). For the dental study example, if the continuous variable is the cost of a restoration, then the ratio mean estimator, 

\[ R = \frac{\sum_{i=1}^{N} y_{i} \ldots}{\sum_{i=1}^{N} x_{i} \ldots}, \]

will represent the average cost of a
relevant restoration. This quantity can then be examined by cluster-, patient-, and subpatient-level characteristics using the weighted regression techniques described in Chapter II.

3.4.2 Alternative Ratio Mean Definitions

Other quantities may also be estimated by the ratio mean. For example, if the proportion of the cost of restorations due to crowns were of interest, one could define the ratio mean estimator in the following way. Let \( y_{ijt} \) equal the cost of the restoration if the \( t \)-th observation for patient \( j \) of cluster \( i \) is a crown, and 0 otherwise. Also, let \( y_{2ijt} \) equal the cost of the restoration if the \( t \)-th observation for patient \( j \) of cluster \( i \) is a restoration of any type. Define

\[
y_{1i..} = \frac{\nu_i}{m_{ij}} \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} y_{1ijt} \quad \text{and} \quad y_{2i..} = \frac{\nu_i}{m_{ij}} \sum_{j=1}^{\nu_i} \sum_{t=1}^{m_{ij}} y_{2ijt}
\]

Then the ratio mean estimator is defined as

\[
R = \frac{\sum_{i=1}^{N} y_{1i..}}{\sum_{i=1}^{N} y_{2i..}} \quad (3.48)
\]

This ratio mean may be interpreted as estimating the proportion of the cost of restorations due to crowns. One may then proceed to use the analysis methods proposed in Chapter II by substituting \( y_{ijt} \) with \( y_{1ijt} \) and \( x_{ijt} \) with \( y_{2ijt} \). Other definitions of the ratio mean estimator may be defined to estimate similar quantities of potential interest involving continuous variables.
CHAPTER IV

RATIO MEAN REGRESSION FOR THE ANALYSIS OF BINARY DATA FROM A TWO-STAGE CLUSTER SAMPLE

4.1 Introduction

Statistical methods for the analysis of multivariate binary outcomes selected from a one-stage cluster sample with large and unequally sized clusters were presented in Chapter II. In that chapter, all elements from each sample cluster were used in the analysis; no subsampling of clusters was done. The inclusion of all elements from each cluster in the analysis will often be preferred when there is little or no extra cost involved in collecting and maintaining information on all elements versus only a sample of these elements. This may be the case when administrative records are the data source. Even if there is an increase in cost associated with using all cluster elements, the savings may be offset by the time and money saved due to avoidance of the more complex analytic techniques required for the analysis of subsampled clusters. However, the costs associated with primary data collection for all cluster elements may be prohibitive so that subsampling of clusters may be necessary.

In this chapter, methods for the analysis of multivariate binary outcomes that have been selected from a two-stage cluster sample with large and unequally sized clusters will be proposed. Each cluster selected into the sample is subsampled and, hence, the variance formulas of Chapter II no longer apply. It will be shown that the methods presented in Chapter II may, in fact, be used to
analyze binary data from a two-stage cluster sample under a particular sampling strategy. Modifications of the methods of Chapter II will be proposed for more general two-stage cluster samples of multivariate binary observations. The modifications will involve strategies which weight sample totals up to estimate cluster totals.

This chapter will begin with a review of two-stage cluster sampling when with replacement sampling at stage one and without replacement sampling at stage two are employed. Notation for population quantities will be introduced and estimators for quantities such as the overall mean and its variance will be given. The ratio mean will be defined within the context of this two-stage sampling strategy. Self-weighting samples and their properties with respect to ratio mean analysis will be discussed. Ratio means will also be defined for subgroups defined by cluster-, patient-, and subpatient-level characteristics. Finally, a weighted regression procedure for analyzing the effects of explanatory factors on the ratio mean will be presented and illustrated using data from the dental study.

4.2 **Review of Two-Stage Cluster Sampling: Sampling With Replacement at Stage One and Without Replacement at Stage Two**

4.2.1 **Notation**

Let $\gamma = 1, \ldots, \Gamma$ index clusters in a population, where $\Gamma$ represents the number of clusters. Let $\xi = 1, \ldots, \nu_\gamma$ index elements in cluster $\gamma$, where $\nu_\gamma$ represents the number of elements in cluster $\gamma$. Let $\mu_{\gamma \xi}$ be the (true) value for element $\xi$ of cluster $\gamma$ in the population. The total for cluster $\gamma$ is given by
\[
\mu_{\gamma}^+ = \sum_{\xi=1}^{\nu_{\gamma}} \mu_{\gamma\xi},
\] (4.1)

the mean for cluster \( \gamma \) is given by

\[
\mu_\gamma = \mu_{\gamma}^+ / \nu_\gamma,
\] (4.2)

the mean over all clusters is given by

\[
\mu = \frac{\sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_{\gamma}} \mu_{\gamma\xi}}{\sum_{\gamma=1}^{\gamma} \nu_{\gamma}},
\] (4.3)

the average cluster total is given by

\[
\bar{\mu} = \frac{\sum_{\gamma=1}^{\Gamma} \mu_{\gamma}^+}{\Gamma},
\] (4.4)

and the average cluster size is given by

\[
\bar{\nu} = \frac{\sum_{\gamma=1}^{\Gamma} \nu_{\gamma}}{\Gamma},
\] (4.5)

Note that \( \mu = \bar{\mu} / \bar{\nu} \).
4.2.2 Estimators and their Variances for the Overall Population Mean and Cluster-Specific Means

Let N clusters be selected from the population at random with replacement. (This is equivalent to sampling without replacement from a very large population). Let \( \varphi_\gamma \) equal the probability of selecting cluster \( \gamma \). Let \( n_\gamma \) elements be selected from cluster \( \gamma \) by simple random sampling without replacement. Let the \( j \)-th element in the sample from the \( i \)-th sample cluster be represented by \( y_{ij} \), where \( i = 1, \ldots, N \) and \( j = 1, \ldots, n_i \). Note that \( n_i = n_\gamma \) when cluster \( \gamma \) is selected. The sample mean for cluster \( i \) is given by

\[
\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}.
\]  

(4.6)

The conditional expected value of \( \bar{y}_i \) given the selection of cluster \( \gamma \) is

\[
E\{ \bar{y}_i \mid i = \gamma \} = E\left\{ \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \mid i = \gamma \right\}
\]

\[= \frac{1}{n_\gamma} \sum_{j=1}^{n_\gamma} E\{ y_{ij} \mid i = \gamma \}\]

\[= \frac{\nu_\gamma}{\xi} \sum_{\xi=1}^{\nu_\gamma} \frac{\mu_{\gamma \xi}}{\nu_\gamma}
\]

\[= \bar{\mu}_\gamma,
\]

(4.7)
the mean for cluster $\gamma$. The unconditional expected value of $\overline{y}_1$ is

$$E\{ \overline{y}_1 \} = E\{ E\{ \overline{y}_1 | i = \gamma \} \}$$
$$= E\{ \mu_\gamma \}$$
$$= \sum_{\gamma=1}^{\Gamma} \varphi_\gamma \mu_\gamma$$

(4.8)

Note that $\overline{y}_1$ is not generally an unbiased estimator of $\mu$.

Another estimator for the overall mean is given by

$$\overline{x}_i = \frac{\overline{y}_i \nu_i}{\overline{v}_1 \nu_+}$$

(4.9)

where $\nu_i$ is the number of elements in the population for the $i$th sample cluster, $\varphi_i$ is the selection probability for the $i$-th sample cluster, and $\nu_+$ is the total number of elements in the population. The conditional expected value of $\overline{x}_i$, given the selection of cluster $\gamma$, is

$$E\{ \overline{x}_i | i = \gamma \} = E\{ \frac{\overline{y}_i \nu_i}{\overline{v}_1 \nu_+} | i = \gamma \}$$
$$= \frac{\nu_\gamma}{\varphi_\gamma \nu_+} E\{ \overline{y}_i | i = \gamma \}$$
$$= \frac{\nu_\gamma}{\varphi_\gamma \nu_+} \mu_\gamma$$
$$= \frac{\mu_\gamma + \frac{\nu_\gamma}{\varphi_\gamma \nu_+}}{\varphi_\gamma \nu_+}$$

(4.10)

The unconditional expectation of $\overline{x}_i$ is
$$E\{ \bar{z}_i \} = E\{ E[\bar{z}_i | i = \gamma] \}$$

$$= E\{ \frac{\mu_{\gamma} + }{\varphi_{\gamma} \nu_{+}} \}$$

$$= \sum_{\gamma=1}^{\Gamma} \varphi_{\gamma} \left\{ \frac{\mu_{\gamma} + }{\varphi_{\gamma} \nu_{+}} \right\}$$

$$= \frac{\sum_{1}^{\nu_{\gamma}} \frac{\mu_{\gamma} \xi}{\sum_{1}^{\nu_{\gamma}} \nu_{\gamma}}}{\sum_{\gamma=1}^{\nu_{\gamma}}}$$

$$= \mu \quad (4.11)$$

Thus, although $\bar{z}_i = \frac{\bar{y}_i \nu_i}{\varphi_i \nu_{+}}$ is a biased estimator of the cluster mean, $\bar{\mu}_\gamma$, $\bar{z}_i$ is an unbiased estimator for the mean over all clusters, $\mu$. When $(\nu_i / \varphi_i \nu_{+}) = 1$, then $\bar{y}_i = \bar{z}_i$ are both unbiased estimates for the cluster mean and the overall mean.

The statistic, $\bar{z}_i = \frac{\bar{y}_i \nu_i}{\varphi_i \nu_{+}}$, may be more generally written as an average of weighted sample elements, $z_{ij} = w_i y_{ij}$, with $w_i = \frac{\nu_i}{\varphi_i \nu_{+}}$, as shown below:

$$\bar{z}_i = \frac{\bar{y}_i \nu_i}{\varphi_i \nu_{+}}$$

$$= w_i \bar{y}_i$$

$$= \frac{\sum_{j=1}^{n_i} w_i y_{ij}}{n_i}$$
\[
\begin{align*}
= \sum_{j=1}^{n_i} z_{ij} \frac{n_i}{\nu_i}.
\end{align*}
\]

Furthermore,
\[
\bar{y}_i = \frac{\bar{y}_i}{\bar{z}_i} \frac{n_i}{\nu_i} = \frac{1}{\nu_i} \sum_{j=1}^{n_i} \left( \frac{1}{\nu_i \bar{z}_i} \right) y_{ij}
\]
\[
= \sum_{j=1}^{n_i} \frac{w_{ij}}{\nu_i} y_{ij},
\]
where \(w_{ij} = \frac{1}{\nu_i \bar{z}_i} \).

The \(z_i\) are independent and identically distributed since clusters are selected with replacement, (i.e. independently by the same process), and samplings within clusters are independent of one another. The conditional variance for \(z_i\), given \(i = \gamma\), is
\[
\Var{\bar{z}_i | i = \gamma} = \Var{w_i \bar{y}_i | i = \gamma}
\]
\[
= w_{\gamma}^2 \Var{\bar{y}_i | i = \gamma}
\]
\[
= w_{\gamma}^2 (1 - f_{\gamma}) \frac{1}{n_{\gamma}} \tau_{\gamma}
\]
where \(\tau_{\gamma} = \frac{1}{(\nu_{\gamma} - 1)} \sum_{\xi=1}^{\nu_{\gamma}} (\mu_{\gamma} \xi - \bar{\mu}_{\gamma})^2, f_{\gamma} = \frac{n_{\gamma}}{n_{\gamma}}, \) and \((1 - f_{\gamma})\) is the finite population correction (Kish, 1965). The unconditional variance of \(z_i\) is given by
\begin{align*}
\text{Var}(\mathcal{Z}_i) &= \text{Var}\{ \mathbb{E}[\mathcal{Z}_i \mid i = \gamma]\} + \\
&\quad \mathbb{E}\{ \text{Var}[\mathcal{Z}_i \mid i = \gamma]\} \\
&= \text{Var}\{ \frac{\mu_{\gamma} + \mu}{\varphi_{\gamma} \nu +} \} \\
&= \sum_{\gamma = 1}^{\Gamma} \varphi_{\gamma} \left\{ \frac{\mu_{\gamma} + \mu}{\varphi_{\gamma} \nu +} - \mathbb{E}\left( \frac{\mu_{\gamma} + \mu}{\varphi_{\gamma} \nu +} \right) \right\}^2 \\
&= \sum_{\gamma = 1}^{\Gamma} \varphi_{\gamma} \left\{ \frac{\mu_{\gamma} + \mu}{\varphi_{\gamma} \nu +} - \mu \right\}^2 \\
&= \sum_{\gamma = 1}^{\Gamma} \varphi_{\gamma} \left\{ \omega_{\gamma} \mu - \mu \right\}^2 \\
&= \tau_b, \\
\end{align*}

where

and where

\begin{align*}
\mathbb{E}\{ \text{Var}[\mathcal{Z}_i \mid i = \gamma]\} &= \mathbb{E}\{ \omega_{\gamma}^2 (1 - f_{\gamma}) \frac{1}{n_{\gamma}} r_{\gamma}\} \\
&= \sum_{\gamma = 1}^{\Gamma} \varphi_{\gamma} \omega_{\gamma}^2 \frac{1 - f_{\gamma}}{n_{\gamma}} r_{\gamma} \\
&= \tau_w, \\
\end{align*}

so that
\[ \text{Var}(z_i) = \sum_{\gamma=1}^{\Gamma} \left\{ (w_{\gamma} \mu_{\gamma} - \mu)^2 \varphi_{\gamma} + \right. \\
\left. w_{\gamma}^2 \frac{1}{n_{\gamma} \nu_{\gamma}} \left( 1 - \frac{n_{\gamma}}{\nu_{\gamma}} \right) \varphi_{\gamma} \tau_{\gamma} \right\} \\
= \tau_b + \tau_w. \quad (4.17) \]

Let

\[ \bar{z} = \frac{1}{N} \sum_{i=1}^{N} z_i \]

\[ = \frac{1}{N} \sum_{i=1}^{N} w_i \bar{v}_i, \quad (4.18) \]

be the simple average of the \( z_i \). The expected value of \( \bar{z} \) is given by

\[ E(\bar{z}) = E\left( \frac{1}{N} \sum_{i=1}^{N} z_i \right) \]

\[ = \mu, \quad (4.19) \]

and the variance of \( \bar{z} \) is given by

\[ \text{Var}(\bar{z}) = \text{Var}\left( \frac{1}{N} \sum_{i=1}^{N} z_i \right) \]

\[ = \frac{1}{N^2} \sum_{i=1}^{N} \text{Var}(z_i) \]

\[ = \frac{1}{N} (\tau_b + \tau_w). \quad (4.20) \]
An unbiased estimator for \( \text{Var}(\overline{Z}) \) is given by

\[
V_{\overline{Z}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (Z_i - \overline{Z})^2 ,
\]

(4.21)

since the \( Z_i \) are independent and identically distributed random variables.

4.3 **Ratio Mean Definition for a Two-Stage Cluster Sample: Sampling With Replacement at Stage One and Without Replacement at Stage Two**

Let the two-stage cluster sampling procedure just described apply to the selection of practices and patients for the dental study example of Chapter II. That is, let dental practices be selected by simple random sampling with replacement at stage one, and let patients be selected by simple random sampling without replacement at stage two. Patients may, as in Chapter II, have multiple sites for which the outcome of interest may occur.

4.3.1 **General Case**

Let \( i=1,\ldots,N \) index sample clusters, \( j=1,\ldots,n_i \) index patients sampled in the \( i \)-th cluster, and \( t=1,\ldots,m_{ij} \) index multiple observations for patient \( j \) in the \( i \)-th cluster. Here \( N \) represents the number of sampled clusters, \( n_i \) represents the number of patients sampled in the \( i \)-th cluster, and \( m_{ij} \) represents the number of observations for patient \( j \) in the \( i \)-th cluster. Let \( x_{ijt2} \) be a binary outcome which takes the value 1 if the \( t \)-th observation for the \( j \)-th patient in the \( i \)-th cluster is relevant (e.g. observed), and 0 otherwise, and let \( y_{ijt2} \) be a binary outcome which takes the value 1 if the \( t \)-th observation for patient \( j \) in cluster \( i \) is relevant and has the attribute, and 0 otherwise. The subscript, 2, in \( y_{ijt2} \) and \( x_{ijt2} \) emphasizes
that these random variables resulted from the two-stage cluster sampling process described above. Let

\[ \bar{y}_{i \cdots 2} = \frac{\sum_{j=1}^{n_i} \left( \sum_{t=1}^{m_{ij}} y_{ijt2} \right)}{n_i} \]

and

\[ \bar{x}_{i \cdots 2} = \frac{\sum_{j=1}^{n_i} \left( \sum_{t=1}^{m_{ij}} x_{ijt2} \right)}{n_i} \]

be the sample mean for the number of relevant observations with the attribute per person in the i-th cluster and the sample mean for the number of relevant observations per person in the i-th cluster, respectively. Also define

\[ \bar{y}_{i \cdots 2w} = w_i \bar{y}_{i \cdots 2} \]

and

\[ \bar{x}_{i \cdots 2w} = w_i \bar{x}_{i \cdots 2} , \]

where \( w_i = \frac{\nu_i}{\nu_i + \nu_+} \) is a sampling weight. Since the sampling method for clusters is simple random sampling with replacement, the \( (\bar{y}_{i \cdots 2w}, \bar{x}_{i \cdots 2w}) \) are independent and identically distributed. The ratio mean estimator for the proportion of relevant observations with the attribute over all clusters for the two-stage sampling framework just described is defined as:

\[ R = \frac{\left( \sum_{i=1}^{N} \bar{y}_{i \cdots 2w} / N \right)}{\left( \sum_{i=1}^{N} \bar{x}_{i \cdots 2w} / N \right)} \]

\[ = \frac{\bar{y}_{2w}}{\bar{x}_{2w}} , \]
where $\bar{y}_{2w} = \frac{\sum_{i=1}^{N} y_{i \cdot 2w}}{N}$ and $\bar{x}_{2w} = \frac{\sum_{i=1}^{N} x_{i \cdot 2w}}{N}$.

$R$ is an estimate of the number of relevant observations with the attribute over all persons and clusters divided by the number of relevant observations over all persons and clusters. The numerator and denominator of $R$ are each sums of $N$ independent random variates, and the numerator and denominator are correlated.

$R$ may also be formulated using matrix notation. Let

$$f_{ijt2} = (y_{ijt2}, x_{ijt2})'$$

$$\bar{f}_{i2} = \frac{1}{m_{ij}} \sum_{t=1}^{m_{ij}} f_{ijt2}$$

$$\bar{f}_{i2w} = w_i \bar{f}_{i2}$$

and

$$\bar{f}_{2w} = \frac{\sum_{i=1}^{N} \bar{f}_{i2w}}{N}$$

Then, $R$ is given by

$$R = \exp A \ln \bar{f}_{2w}, \hspace{1cm} (4.23)$$

where $A = [1, -1]$.

A first-order Taylor series representation for $R$ at the population mean $\mu = (\mu_y, \mu_x)'$ of the $(\bar{y}_{2w}, \bar{x}_{2w})$ is given by
\[ R = \frac{\mu_y}{\mu_x} + \frac{1}{\mu_x} (\bar{y}_{2w} - \mu_y) - \frac{\mu_y}{\mu_x^2} (\bar{x}_{2w} - \mu_x) + O\left(\frac{1}{N}\right) \] (4.24)

The asymptotic expected value of a first-order Taylor series for \( R \), \( E_A(R) \), is \( \frac{\mu_y}{\mu_x} = \theta \), the ratio mean in the population. The variance for \( R \), based on a Taylor series linearization, the independent and identically distributed nature of the \( (\bar{y}_{i\ldots2w}, \bar{x}_{i\ldots2w}) \), and the methods of Koch and Imrey et al. (1985), is given by

\[ \text{Var}(R) = \mathcal{D}_\theta A \mathcal{D}_\mu^{-1} [\text{Var}(\bar{I}_{2w})] \mathcal{D}_\mu^{-1} A' \mathcal{D}_\theta \] (4.25)

where \( \mathcal{D}_\theta \) and \( \mathcal{D}_\mu \) are diagonal matrices with \( \mu \) and \( \theta \) as diagonal elements and \( A = [1, -1] \). Also, \( \text{Var}(\bar{I}_{2w}) \) has a structure which is a bivariate counterpart to \( \text{Var}(\bar{Z}) \) as discussed earlier. A variance estimator for \( \text{Var}(R) \) is given by

\[ \text{var}(R) = \mathcal{D}_R A \mathcal{D}_{\bar{I}_{2w}}^{-1} \mathcal{V}_{\bar{I}_{2w}} \mathcal{D}_{\bar{I}_{2w}}^{-1} A' \mathcal{D}_R \] (4.26)

where

\[ \mathcal{V}_{\bar{I}_{2w}} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\bar{I}_{i2w} - \bar{I}_{2w})(\bar{I}_{i2w} - \bar{I}_{2w})' \] (4.27)

A large sample variance estimator for \( \text{Var}(\ln R) \), based on a Taylor series linearization, is given by

\[ \text{var}(\ln R) = \frac{\text{var}(R)}{R^2} \] (4.28)
A large sample \((1 - \alpha)\)-level confidence interval for \(\theta\) is given by:

\[
\exp \left\{ \ln R \pm Z_{(1-\alpha/2)} \cdot \left( \text{var}(R)^{1/2} / R \right) \right\}.
\]  

(4.29)

### 4.3.2 Special Sampling Schemes

Note that expression (4.22) depends on \(w_i\). More specifically, expression (4.22) depends upon knowing: the total number of study subjects in cluster \(i\), \(\nu_i\), \(i = 1, ..., N\), the probability of selecting cluster \(i\), \(\varphi_i\), \(i = 1, ..., N\), and the total number study subjects in the population, \(\nu_+\). In practice, some of these quantities may not be known. Certain sampling schemes obviate the need to know some or all of these quantities for the analysis. Two options are discussed in this section.

Let \(\varphi_{\gamma} = \frac{1}{k}\) be the equal probability of selecting clusters into the sample using simple random sampling with replacement at stage one, where \(\gamma\) indexes clusters in the population. Let \(\pi\) be the constant probability of selecting elements within each cluster at stage two. Hence, the common selection probability for each population element is given by

\[
\varphi = \frac{1}{k} \cdot \pi = \frac{1}{k} \cdot \frac{\pi \nu_{\gamma}}{\nu_+} = \frac{1}{k} \cdot \frac{n_{\gamma}}{\nu_+},
\]

where \(n_{\gamma} = \pi \nu_{\gamma}\). This implies that \(w_i = \frac{\nu_i}{\varphi_i \nu_+} = \frac{n_i \Gamma}{\pi \nu_+}\). Thus,

\[
R = \frac{v}{x}
\]
\[
N \sum_{i=1}^{N} \left\{ \frac{\frac{\sum_{j=1}^{n_i} \sum_{t=1}^{m_{ij}} y_{ijt2}}{n_i}}{\frac{\sum_{j=1}^{n_i} \sum_{t=1}^{m_{ij}} x_{ijt2}}{n_i}} \right\}
= \frac{\sum_{i=1}^{N} n_i \sum_{j=1}^{n_i} \sum_{t=1}^{m_{ij}} y_{ijt2}}{\sum_{i=1}^{N} n_i \sum_{j=1}^{n_i} \sum_{t=1}^{m_{ij}} x_{ijt2}}.
\]

Expression (4.30) is similar to expression (2.1) in Chapter II, with the exception that the summation over j only includes subjects randomly selected into the sample in stage II, whereas in Chapter 2, the summation over j included all subjects in each selected cluster. Therefore, estimation of R is the same as it was in Chapter II. Although \( \text{Var}(R) \) still depends on \( \nu_\gamma \), the number of elements in the population for each cluster, its estimation via (4.26) does not require knowing \( \nu_\gamma \) nor \( \Gamma \) and \( \nu_+ \) since these quantities cancel out in (4.21) and in expression (4.26) for the Taylor series approximation for the \( \text{Var}(R) \). Nevertheless, the \( \nu_1 \) must be known for determination of \( n_1 = \pi \nu_1 \).

A sampling scheme which avoids specification of population quantities in the analysis altogether is the self-weighting design. Let \( \varphi_\gamma = \frac{\nu_\gamma}{\nu_+} \) be the probability of selecting cluster \( \gamma \) in stage one. That is, let clusters be selected proportional to the number of subjects in the cluster (i.e. 'proportional to size').
\[ R = \frac{\bar{y}}{\bar{x}} \]

\[
\frac{N}{\sum_{i=1}^{n_i} \sum_{j=1}^{m_{ij}} \sum_{t=1}^{n_{ij}} y_{ijt2}}
\]

\[
= \frac{N}{\sum_{i=1}^{n_i} \sum_{j=1}^{m_{ij}} \sum_{t=1}^{n_{ij}} x_{ijt2}}
\]

Again, (4.31) is the same as (2.1) in Chapter II for R with the exception that the summation over \( j \) is only over subjects selected into the sample in stage II. In this case, the methods of Chapter II for whole clusters may be used for subsampled clusters without modification. However, one should note that the \( \nu_\gamma \) need to be known for all clusters in order to determine the \( \varphi_\gamma \) as \( \nu_\gamma / \nu_+ \).

### 4.4 Subgroups of Ratio Means Defined by Cluster-, Patient-, and Subpatient-Level Characteristics

Let \( h=1,\ldots,H \) index levels for the crossclassification of cluster-level characteristics, \( k=1,\ldots,K \) index levels for the crossclassification of patient-level characteristics, \( l=1,\ldots,L \) index levels for the crossclassification of subpatient-level characteristics, \( i=1,\ldots,N_h \) index sample clusters in stratum \( h \), \( j=1,\ldots,n_{hi} \) index patients sampled in cluster \( i \) of stratum \( h \), and \( t=1,\ldots,m_{hij} \) index multiple observations for patient \( j \) in cluster \( i \) of stratum \( h \). Here, \( N_h \) is the number of
sample clusters in the h-th stratum, \( n_{hi} \) is the number of sample patients in
cluster i of stratum h, \( m_{hij} \) is the number of observations for patient j in cluster i
of stratum h, H is the number of levels for the crossclassification of cluster-level
characteristics, K is the number of levels for the crossclassification of patient-
level characteristics, and L is the number of levels for the crossclassification of
subpatient-level characteristics. Let \( x_{hijkl}^{t} \) and \( y_{hijkl}^{t} \) be binary random
variables observed from a two-stage cluster sample with replacement at stage one
and without replacement at stage two. Let \( x_{hijkl}^{t} \) take the value 1 if the t-th
observation for patient j of cluster i of stratum h is relevant and of type l and is
from a patient of type k, and 0 otherwise, and let \( y_{hijkl}^{t} \) take the value 1 if the
t-th observation for patient j in cluster i of stratum h is relevant and of type l
and is for a patient of type k, and has the attribute, and 0 otherwise. Also, let

\[
\bar{f}_{hikl} = \frac{1}{\sum_{j} n_{hi}} \sum_{i} (y_{hijkl}^{t} x_{hijkl}^{t})'
\]

\[
\bar{f}_{hi2} = (\bar{f}_{hi11}^{t}, \bar{f}_{hi12}^{t}, \ldots, \bar{f}_{hiKL}^{t})'
\]

\[
\bar{f}_{hi2w} = w_{hi} \bar{f}_{hi2}
\]

and

\[
\bar{f}_{h2w} = \frac{1}{N_{h}} \sum_{i=1}^{N_{h}} \bar{f}_{hi2w}
\]

where \( w_{hi} = \frac{\nu_{hi}}{\varphi_{hi} + \nu_{h}} \). A consistent estimator for the variance of \( \bar{f}_{h2w} \) is given
by
\[
\begin{align*}
V_{\bar{f}_{h2w}} &= \frac{1}{N_{h}(N_{h} - 1)} \sum_{i=1}^{N_{h}} (\bar{f}_{hi2w} - \bar{f}_{h2w}) (\bar{f}_{hi2w} - \bar{f}_{h2w})' .
\end{align*}
\] (4.32)

Let
\[
\bar{f}_{2w} = (\bar{f}_{12w}', ..., \bar{f}_{H2w}')',
\]
so that
\[
V_{\bar{f}_{2w}} = \text{Block} \left( V_{\bar{f}_{h2w}} \right).
\] (4.33)

The vector of ratio means resulting from the crossclassification of cluster-, patient-, and subpatient-level characteristics may be written as
\[
R = \exp A \ln \bar{f}_{2w},
\] (4.34)

where
\[
A = [1, -1] \otimes I_{HKL},
\]

where \(I_{HKL}\) is the HKL \(\times\) HKL identity matrix and "\(\otimes\)" is the symbol for the Kronecker product whereby the matrix, \([1, -1]\), on the left multiplies each element of the matrix, \(I_{HKL}\), on the right. As shown in Koch and Imrey et al., (1985), a consistent estimator for the variance of \(R\) based on a first-order Taylor series approximation for \(R\) is given by
\[
V_{R} = \mathcal{H} V_{\bar{f}_{2w}} \mathcal{H}',
\] (4.35)

where
\[
\mathcal{H} = D_{R} A D_{f_{2w}}^{-1}.
\]
4.5 **Ratio Mean Regression**

The simultaneous examination of the effects of cluster-, patient-, and subpatient-level characteristics on the ratio mean and the ability to evaluate interactions between two or more variables is possible using the method of Grizzle, Starmer, and Koch (1969) (GSK) and extensions of this approach as discussed in Koch and Imrey *et al* (1985). This regression procedure will be presented and applied in the remaining sections of this Chapter.

4.5.1 **A Linear Model for Logit(R)**

Let

\[
\bar{\mathbf{F}} = (F_{111}, \ldots, F_{HKL})
\]

\[
= \logit R
\]

\[
= \ln (R) - \ln (1 - R),
\]

be a vector of logits of the ratio means corresponding to the subpopulations defined by the crossclassification of the H cluster-, K patient-, and L subpatient-level characteristics. An estimate of the covariance matrix for \( \bar{\mathbf{F}} \), based on a first-order Taylor series approximation for \( \bar{\mathbf{F}} \), is given by

\[
\mathbf{V}_{\bar{\mathbf{F}}} = \mathbf{D}_{R}^{-1} (\mathbf{D}_{(1 - R)})^{-1} \mathbf{V}_{R} \mathbf{D}_{R}^{-1} (\mathbf{D}_{(1 - R)})^{-1}
\]

(4.36)

where \( \mathbf{D}_{R} \) is a HKL \times HKL matrix with the elements of the vector \( R \) on the diagonal, \( \mathbf{D}_{(1 - R)} \) is a HKL \times HKL matrix with the elements of the vector \( (1 - R) \) on the diagonal, and \( \mathbf{V}_{R} \) is the estimate of the covariance matrix for \( R \).
given in (4.). The function vector $\mathbf{F}$ may be modeled using the GSK method by directly forming $\mathbf{F}$ and its covariance matrix. A linear model for $\mathbf{F}$ is given by

$$E_A(\mathbf{F}) = \mathbf{X}\beta$$  \hspace{1cm} (4.37)

where $\mathbf{X}$ is the $\text{HKL} \times p$ design or model specification matrix and $\beta$ is a $p \times 1$ vector of unknown parameters to be estimated. The covariance matrix of $\mathbf{F}$ may be consistently estimated by $\mathbf{V}_{\mathbf{F}}$. The weighted least squares estimator $\hat{\beta}$ of $\beta$ is given by

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{F}. \hspace{1cm} (4.38)$$

In large samples, the estimator $\hat{\beta}$ has approximately a normal distribution with

$$E_A(\hat{\beta}) = \beta \hspace{1cm} (4.39)$$

and a covariance matrix which is consistently estimated by

$$V_A(\hat{\beta})=(\mathbf{X}'\mathbf{V}_F^{-1}\mathbf{X})^{-1}. \hspace{1cm} (4.40)$$

4.5.2 Example

The dental data used for the example in Section 2.7 were selected from the population by a one-stage cluster sample with replacement. Using that sample as the first stage of a two-stage cluster sample for this example, patients were then subsampled within each cluster with selection probability $p = .2$ by simple random sampling without replacement. The model fit to the data in
Chapter II was fit to the sample data for this example using weighted regression for the logit(\( R \)) as described in the previous section. The effects of region of the country (Northeast, South, Central, and West), age of the patient (≤ 34, 35-50, and ≥ 51), and tooth arch (mandibular versus maxillary) on the proportion of restorations with greater than or equal to three surfaces which are crowns were explored. The number of patients sampled per cluster ranged from 1 to 140 with a median of 6. The number of relevant teeth per sample cluster ranged from 1 to 151 with a median of 6. The median number of relevant teeth per person in the sample was 1, and ranged from 1 to 4. The total number of clusters in the sample was 358, the total number of patients in the sample was 3383, and the total number of relevant teeth in the sample was 3665.

Observed ratio means and their standard errors for each of the twenty-four subpopulations formed by the crossclassification of region, age, and tooth arch, appear in Table 4.1. Note that the standard errors computed by the Taylor series method are nearly as small as they were in Chapter II. (The standard errors which were computed as if the data were obtained from a simple random sample are given for comparison only; they are not the correct standard errors for this two-stage cluster sample). The estimated covariance matrix for \( R \), \( V R \), was computed by the Taylor series method and is given in Table 4.2.

The following model for the logit(\( R \)) was fit to the sample data:

\[
E(R) = \beta_0 + \beta_1 \text{Region1} + \beta_2 \text{Region2} + \beta_3 \text{Region3} + \\
\beta_4 \text{Age1} + \beta_5 \text{Age2} + \beta_6 \text{Mandib} + \\
\beta_7 \text{Age1} \times \text{Mandib} + \beta_8 \text{Age2} \times \text{Mandib}
\]  
(4.41)
where: $\bar{x} = \logit(R)$; Region1, Region2, and Region3, are indicator variables for region 1, region 2, and region 3, respectively; Age1 and Age2 are indicator variables for age \leq 34 and age 35 to 50, respectively; Mandib is an indicator variable indicating the mandibular versus maxillary arch; and Age1 \times Mandib and Age2 \times Mandib are interaction terms.

Parameter estimates computed by the ratio mean regression (RMR) method and their standard errors are given in Table 4.3. The analysis of variance (ANOVA) for the model is presented in Table 4.4. The goodness of fit (GOF) statistic given in Table 4.4 ($\chi^2 = 8.04$, df = 15, p-value = 0.922) indicates that model (4.41) adequately represents the data. The ANOVA for the RMR model shows that there is a significant interaction between age and tooth arch ($\chi^2 = 6.98$, df = 2, p-value = 0.030). It is noted that a model without any interaction terms also fit the data ($\chi^2 = 15.03$, df = 17, p-value = 0.594) since the interaction effect between age and tooth arch, while significant, was small in magnitude. Parameter estimates and standard errors for models containing the same predictor variables but fit to the data using survey data logistic regression and ordinary logistic regression are also given in Table 4.3 for comparison. The standard errors for the RMR model are slightly smaller than those for the survey logistic regression.

The predicted ratio means generated from the RMR model for each subpopulation and their standard errors are given in Table 4.5. The standard errors for the model-predicted ratio means are smaller than those for the observed ratio means since the model-predicted ratio means are computed using all of the data whereas the observed ratio means are computed using only the data in each respective subpopulation.
Adjusted odds ratio's (OR's) and 95% confidence intervals (CI's) for selected factor-level comparisons are given in Table 4.6. It may be concluded from the model that crown use for restorations with greater than or equal to three surfaces is lower in region 1 than it is in regions 2, 3, and 4, which have similar levels of crown utilization, although this result just borders on statistical significance. (Note that the model without the Age × Tooth Arch interaction has a significant region effect). The odds of a crown are also greater for patients ≥ 51 years of age versus patients ≤ 34 years of age, for patients 35-50 years of age versus patients ≤ 34 years of age, and for patients ≥ 51 years of age versus patients 30-50 years of age. The odds ratios for these age comparisons are slightly larger for maxillary vs mandibular teeth; that is, there is a significant age by tooth arch interaction.

4.6 Summary

A ratio mean approach to the analysis of multivariate binary outcomes selected from a two-stage cluster sample with replacement sampling of clusters at stage one, and without replacement sampling of patients at stage two, was proposed in this chapter. Ratio mean estimators of population quantities were modified to accommodate sampling weights appropriate to the design. Weighted regression as discussed in Koch and Imrey et al. (1985) was used to analyze the modified ratio mean estimates for subgroups formed by the crossclassification of cluster-, patient-, and subpatient-level characteristics. An example using the dental data illustrated that the results of a RMR model using only 3665 teeth were similar to those obtained for the same model using whole clusters in Chapter II which included 26,525 teeth. The model-predicted ratio means and
their standard errors were nearly the same as those presented in Chapter II.

It is noted that the cluster sizes for the example data were highly variable. Even better concordance of estimates between the one-stage sample of Chapter II and the two-stage sample of this chapter for cluster-level parameters may have been obtained with the RMR method if the cluster sizes had been less variable for the example data, or the two-stage cluster sample was selected with probability proportional to size, or the clusters in the population were stratified according to their size and then selected from within each of these stratum.

The methodology proposed in this chapter is useful for the analysis of two-stage cluster samples with large and unequally sized clusters. While the RMR method gives results similar to those obtained using survey data logistic regression and (presumably GEE1), the estimation techniques are more straightforward. The disadvantages of this method are the same as those noted in Chapter II: that continuous predictor variables must be categorized for inclusion in RMR models and that there is currently no software package which will easily implement this methodology.
CHAPTER V

A RATIO MEAN APPROACH FOR JOINT ESTIMATION OF THE
OVERALL MEAN AND INTRACLUSTER CORRELATION FOR BINARY
DATA FROM A ONE-STAGE CLUSTER SAMPLE

5.1 Introduction

Several of the preceding chapters have focused on the use of a ratio mean approach for estimating the overall population mean and the effects of covariates on subpopulation means for a (multivariate) binary attribute among relevant observations from one- and two-stage cluster samples. While the ratio mean approach of prior chapters accounts for potential dependencies among observations within clusters, it provides no information regarding the degree of dependence among observations within the clusters or whether it varies according to cluster characteristics. Joint estimation of both the overall mean and some measure of intracluster correlation will provide information about the degree of dependence among observations within clusters and also opportunities for more efficient estimation of the overall mean.

The objective of this chapter will be to extend the ratio mean approach of the previous chapters to include joint estimation of both the overall mean and a measure of intracluster correlation. The methodology will provide extensions of the methods discussed by Kempthorne (1982) to situations with a binary attribute among relevant observations and samples of clusters with varying size and selection probabilities. The methodology will be developed beginning with
one-stage sampling of whole clusters from one population. A weighted regression procedure for estimation of the effects of cluster-level covariates on the overall mean and measure of intracluster correlation will also be demonstrated. This methodology is well suited for the analysis of large clusters.

5.2 Notation for Population Quantities

Let $\gamma = 1, 2, \ldots, \Gamma$ index cluster selected in a population and let $\xi = 1, 2, \ldots,$ $\nu_\gamma$ index elements within the clusters, where $\Gamma$ represents the number of clusters in the population and $\nu_\gamma$ represents the number of elements in cluster $\gamma$. Let $y_{\gamma \xi} = 1$ if $\xi$ in $\gamma$ is relevant and has the attribute, and $y_{\gamma \xi} = 0$ if otherwise. Also, let $x_{\gamma \xi} = 1$ if $\xi$ in $\gamma$ is relevant, and $x_{\gamma \xi} = 0$ if otherwise. Let

$$\pi_y = \frac{\sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} y_{\gamma \xi}}{\sum_{\gamma=1}^{\Gamma} \nu_\gamma}$$

(5.1)

$$= \frac{\nu_y +}{\nu_+}$$

be the population proportion of observations which are relevant and have the attribute, where $\nu_y = \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} y_{\gamma \xi}$ and $\nu_+ = \sum_{\gamma=1}^{\Gamma} \nu_\gamma$. The $\pi_y$ also represents the probability that a randomly selected element from the population is relevant and has the attribute when all elements have equal selection probabilities.
Let

\[ \pi_x = \frac{\sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} x_{\gamma\xi}}{\sum_{\gamma=1}^{\Gamma} \nu_\gamma} \]

\[ = \frac{\nu_x +}{\nu_+} \] (5.2)

be the population proportion of relevant observations, where \( \nu_{x+} = \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} x_{\gamma\xi} \) and \( \nu_+ = \sum_{\gamma=1}^{\Gamma} \nu_\gamma \). The \( \pi_x \) also represents the probability that a randomly selected element from the population is relevant when all elements have equal selection probabilities.

Define

\[ \theta = \frac{\pi_y}{\pi_x} = \frac{\nu_+}{\nu_x +} \]

as the population proportion of elements with the attribute among relevant elements. Also, let

\[ \pi_{y\gamma} = \frac{\sum_{\xi=1}^{\nu_\gamma} y_{\gamma\xi}}{\nu_\gamma} = \frac{\nu_{y\gamma}}{\nu_\gamma} \] (5.3)

be the proportion of elements in the \( \gamma \)-th cluster in the population with relevance and the attribute, and let

\[ \pi_{x\gamma} = \frac{\sum_{\xi=1}^{\nu_\gamma} x_{\gamma\xi}}{\nu_\gamma} = \frac{\nu_{x\gamma}}{\nu_\gamma} \] (5.4)
be the proportion of elements in the $\gamma$-th cluster in the population with relevance. Therefore, alternative expressions for $\pi_y$ and $\pi_x$ are given by

$$\pi_y = \frac{\sum_{\gamma=1}^{\Gamma} \nu_y \pi_{y\gamma}}{\nu_+} \tag{5.5}$$

and

$$\pi_x = \frac{\sum_{\gamma=1}^{\Gamma} \nu_y \pi_{x\gamma}}{\nu_+} \tag{5.6}$$

respectively. Let $\theta_{\gamma}$ be the population proportion of elements in the $\gamma$-th cluster with the attribute among relevant elements in the $\gamma$-th cluster. It is defined as

$$\theta_{\gamma} = \frac{\pi_{y\gamma}}{\pi_{x\gamma}} = \frac{\nu_{y\gamma}}{\nu_{x\gamma}}, \text{ when } \nu_{x\gamma} > 0 \text{ and }$$

$$= 0, \text{ when } \nu_{x\gamma} = 0.$$

The quantity $\theta$ can alternatively be written as a function of weighted sums of the $\theta_{\gamma}$, weighted by the number of relevant observations $\nu_{x\gamma}$ in cluster $\gamma$, $\gamma = 1, 2, ..., \Gamma$, as follows:

$$\theta = \frac{\sum_{\gamma=1}^{\Gamma} \nu_{x\gamma} \theta_{\gamma}}{\sum_{\gamma=1}^{\Gamma} \nu_{x\gamma} \pi_{x\gamma}} \tag{5.7}$$

$$= \frac{\sum_{\gamma=1}^{\Gamma} \nu_{x\gamma} \theta_{\gamma}}{\sum_{\gamma=1}^{\Gamma} \nu_{x\gamma}}.$$
Thus, $\theta$ is a weighted average across clusters in the population in a sense like $\pi_y$ and $\pi_x$ are except the weights for it are $(\nu'_{x,\gamma}/\nu'_{x,+})$ whereas those for $\pi_y$ and $\pi_x$ are $(\nu'_{y}/\nu'_{+})$.

### 5.3 Partitioning Variability

Let

$$\eta_{yT} = \frac{1}{\nu^+} \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu'_{\gamma}} (y_{\gamma\xi} - \pi_y)^2$$  \hspace{1cm} (5.8)

represent the population variance for $y_{\gamma\xi}$. It is also the sampling variance which applies to a randomly selected element from the population when all elements have equal selection probabilities. Note that $\eta_{yT}$ simplifies as follows:

$$\eta_{yT} = \frac{1}{\nu^+} \sum_{\gamma=1}^{\Gamma} \nu'_{\gamma} \sum_{\xi=1}^{y_{\gamma\xi}} (y_{\gamma\xi} - 2y_{\gamma\xi}\pi_y + \pi^2_y)$$

$$= \frac{1}{\nu^+} \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{y_{\gamma\xi}} (y_{\gamma\xi} - 2y_{\gamma\xi}\pi_y + \pi^2_y)$$

$$= \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{y_{\gamma\xi}} \frac{y_{\gamma\xi}}{\nu^+} - 2\pi_y \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu'_{\gamma}} \frac{y_{\gamma\xi}}{\nu^+} + \pi^2_y$$

$$= \pi_y - 2\pi_y^2 + \pi^2_y$$

$$= \pi_y (1 - \pi_y)$$  \hspace{1cm} (5.9)

The quantity, $\eta_{yT}$, may also be partitioned into between- and within-cluster variance components as shown below:
\[ \eta_{yT} = \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_{\gamma}} (y_{\gamma \xi} - \bar{y}_y)^2 \]

\[ = \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_{\gamma}} (y_{\gamma \xi} - \bar{y}_y + \bar{y}_y - \bar{y}_y)^2 \]

\[ = \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_{\gamma}} \{ (y_{\gamma \xi} - \bar{y}_y)^2 \} + 2(y_{\gamma \xi} - \bar{y}_y)(\bar{y}_y - \bar{y}_y) + (\bar{y}_y - \bar{y}_y)^2 \}

\[ = \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} (\bar{y}_y - \bar{y}_y)^2 \]

\[ = \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_{\gamma}} \{ y_{\gamma \xi} - 2y_{\gamma \xi} \bar{y}_y + \bar{y}_y \}

\[ + \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} (\bar{y}_y - \bar{y}_y)^2 \]

\[ = \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} (\bar{y}_y - 2\bar{y}_y + \bar{y}_y) + \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} (\bar{y}_y - \bar{y}_y)^2 \]

\[ = \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} \pi_{y\gamma}(1 - \pi_{y\gamma}) + \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} (\bar{y}_y - \bar{y}_y)^2 \]

\[ = \eta_{yW} + \eta_{yA} \] \hspace{1cm} (5.10)

Here,

\[ \eta_{yW} = \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} \pi_{y\gamma}(1 - \pi_{y\gamma}) \]

represents within cluster variance for \( y_{\gamma \xi} \), and
\[ \eta_{yA} = \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{\gamma}}{\nu_{+}} (\pi_{y\gamma} - \pi_y)^2 \]

represents between cluster variance for \( y_{\gamma\xi} \). Likewise, let the population variance for \( x_{\gamma\xi} \) be represented by

\[ \eta_{xT} = \frac{1}{\nu_{+}} \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{\gamma}}{\nu_{+}} \sum_{\xi = 1}^{\nu_{\gamma}} (x_{\gamma\xi} - \pi_x)^2 \]

\[ = \pi_x (1 - \pi_x) \]

\[ = \frac{1}{\nu_{+}} \sum_{\gamma = 1}^{\Gamma} \nu_{\gamma} \pi_{x\gamma} (1 - \pi_{x\gamma}) + \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{\gamma}}{\nu_{+}} (\pi_{x\gamma} - \pi_x)^2 \]

\[ = \eta_{xW} + \eta_{xA} , \]

(5.12)

where

\[ \eta_{xW} = \frac{1}{\nu_{+}} \sum_{\gamma = 1}^{\Gamma} \nu_{\gamma} \pi_{x\gamma} (1 - \pi_{x\gamma}) \]

represents within cluster variance for \( x_{\gamma\xi} \) and

\[ \eta_{xA} = \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{\gamma}}{\nu_{+}} (\pi_{x\gamma} - \pi_x)^2 \]

represents between cluster variance for \( x_{\gamma\xi} \). The population covariance between \( y_{\gamma\xi} \) and \( x_{\gamma\xi} \) is given by

\[ \eta_{yxT} = \frac{1}{\nu_{+}} \sum_{\gamma = 1}^{\Gamma} \sum_{\xi = 1}^{\nu_{\gamma}} (y_{\gamma\xi} - \pi_y)(x_{\gamma\xi} - \pi_x) \]
\[ = \frac{1}{p_+} \sum_{\gamma=1}^{\nu} \sum_{\xi=1}^{\nu} (y_{\gamma\xi} x_{\gamma\xi} - y_{\gamma\xi} \sigma_x - x_{\gamma\xi} \sigma_y + \sigma_y \sigma_x) \]

\[ = \frac{1}{p_+} \sum_{\gamma=1}^{\nu} \sum_{\xi=1}^{\nu} (y_{\gamma\xi} - y_{\gamma\xi} \sigma_x - x_{\gamma\xi} \sigma_y + \sigma_y \sigma_x) \]

\[ = \frac{1}{p_+} (\nu y_+ - \nu y_+ \sigma_x - \nu x_+ \sigma_y + \nu \sigma_y \sigma_x) \]

\[ = \sigma_y - \sigma_y \sigma_x - \sigma_y \sigma_x + \sigma_y \sigma_x \]

\[ = \sigma_y (1 - \sigma_x) \quad (5.13) \]

The expression for \(\eta_{yXT}\) is also equivalent to:

\[ \eta_{yXT} = \frac{1}{p_+} \sum_{\gamma=1}^{\nu} \sum_{\xi=1}^{\nu} (y_{\gamma\xi} - \sigma_y)(x_{\gamma\xi} - \sigma_x) \]

\[ = \frac{1}{p_+} \sum_{\gamma=1}^{\nu} \sum_{\xi=1}^{\nu} [(y_{\gamma\xi} - \sigma_y \gamma) + (\sigma_y \gamma - \sigma_y)][(x_{\gamma\xi} - \sigma_x \gamma)(\gamma x_\gamma - \sigma_x)] \]

\[ = \frac{1}{p_+} \sum_{\gamma=1}^{\nu} \sum_{\xi=1}^{\nu} [(y_{\gamma\xi} - \sigma_y \gamma)(x_{\gamma\xi} - \sigma_x \gamma) + (y_{\gamma\xi} - \sigma_y \gamma)(\sigma x_\gamma - \sigma_x)] \]

\[ + (\sigma_y \gamma - \sigma_y)(x_{\gamma\xi} - \sigma_x \gamma) + (\sigma_y \gamma - \sigma_y)(\sigma x_\gamma - \sigma_x)] \]

\[ = \frac{1}{p_+} \sum_{\gamma=1}^{\nu} \sum_{\xi=1}^{\nu} (y_{\gamma\xi} - \sigma_y \gamma)(x_{\gamma\xi} - \sigma_x \gamma) + \sum_{\gamma=1}^{\nu} \sigma y_+ (\sigma y \gamma - \sigma y)(\sigma x_\gamma - \sigma_x) \]

\[ = \frac{1}{p_+} \sum_{\gamma=1}^{\nu} \sum_{\xi=1}^{\nu} (y_{\gamma\xi} - \sigma_y \gamma x_{\gamma\xi} - y_{\gamma\xi} \sigma x \gamma + \sigma y \sigma x_\gamma) \]

\[ + \sum_{\gamma=1}^{\nu} \frac{\nu y_+ (\sigma y \gamma - \sigma y)(\sigma x_\gamma - \sigma_x)}{p_+} \]
\[ \frac{1}{\nu_{+}} \sum_{\gamma=1}^{\Gamma} \left( \nu_{\gamma} \pi_{y\gamma} - \nu_{\gamma} \pi_{y\gamma} \pi_{x\gamma} - \nu_{\gamma} \pi_{y\gamma} \pi_{x\gamma} + \nu_{\gamma} \pi_{y\gamma} \pi_{x\gamma} \right) \]

\[ + \sum_{\gamma=1}^{\Gamma} \frac{\nu_{\gamma}}{\nu_{+}} (\pi_{y\gamma} - \pi_{y})(\pi_{x\gamma} - \pi_{x}) \]

\[ = \frac{1}{\nu_{+}} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} \pi_{y\gamma} (1 - \pi_{x\gamma}) + \sum_{\gamma=1}^{\Gamma} \frac{\nu_{\gamma}}{\nu_{+}} (\pi_{y\gamma} - \pi_{y})(\pi_{x\gamma} - \pi_{x}) \]

\[ = \eta_{yxW} + \eta_{yxA} \]

(5.14)

where

\[ \eta_{yxW} = \frac{1}{\nu_{+}} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} \pi_{y\gamma} (1 - \pi_{x\gamma}) \]

and

\[ \eta_{yxA} = \sum_{\gamma=1}^{\Gamma} \frac{\nu_{\gamma}}{\nu_{+}} (\pi_{y\gamma} - \pi_{y})(\pi_{x\gamma} - \pi_{x}) \]

Let

\[ \Omega_{T} = \frac{1}{\nu_{+}} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} \sum_{\xi=1}^{\sum} (y_{\gamma\xi} - \theta_{x\gamma\xi})^2 \]

(5.15)

be a measure of the population variability for \( y_{\gamma\xi} \) taking into account \( x_{\gamma\xi} \). While \( \Omega_{T} \) is not strictly a variance by definition, the usefulness of this quantity for estimating a measure of intracluster correlation will be seen later in the chapter.

One useful property of \( \Omega_{T} \) is that it may be expressed as

\[ \Omega_{T} = \frac{1}{\nu_{+}} \sum_{\gamma \text{ with } \nu_{x\gamma} > 0} \sum_{\xi \text{ with } x_{\gamma\xi} = 1} (y_{\gamma\xi} - \theta)^2; \]
and so \((\frac{\nu^+}{\nu^+_{X+}})\Omega_T\) is the population variance for \(y_{\gamma\xi}\) given that \(x_{\gamma\xi} = 1\). It is also the sampling variance which applies to a randomly selected, relevant element from the population when all such elements have equal selection probabilities.

Another alternative expression for the variability measure \(\Omega_T\) is as follows:

\[
\Omega_T = \frac{1}{\nu^+} \sum_{\gamma=1}^{\nu^+} \sum_{\xi=1}^{\nu^+\gamma} (y_{\gamma\xi} - \bar{y}_{\gamma\xi})^2
\]

\[
= \frac{1}{\nu^+} \sum_{\gamma=1}^{\nu^+} \sum_{\xi=1}^{\nu^+\gamma} \left( y_{\gamma\xi}^2 - 2y_{\gamma\xi}\bar{y}_{\gamma\xi} + \bar{y}_{\gamma\xi}^2 \right)
\]

\[
= \frac{1}{\nu^+} \sum_{\gamma=1}^{\nu^+} \sum_{\xi=1}^{\nu^+\gamma} \left( y_{\gamma\xi}^2 - 2y_{\gamma\xi}\bar{y}_{\gamma\xi} + \bar{y}_{\gamma\xi}^2 \right)
\]

\[
= \frac{1}{\nu^+} \left( \nu_y^+ - 2\bar{y}_{\gamma\xi} + \bar{y}_{\gamma\xi}^2 \right)
\]

\[
= \frac{1}{\nu^+} \left\{ (1-\theta)\nu_y^+ + \theta (\nu_{X+} - \nu_y^+) \right\}
\]

\[
= \frac{1}{\nu^+} \left\{ (1-\theta)\nu_y^+ + \theta (\nu_{X+} - \nu_{X+}) \right\}
\]

\[
= \frac{\nu_y^+}{\nu^+} (1-\theta)
\]

\[
= \pi_X \theta (1 - \theta) \tag{5.16}
\]

Thus, \((\frac{\nu^+}{\nu^+_{X+}})\Omega_T = (\frac{\Omega_T}{\Omega_X}) = \theta(1-\theta)\) is analogous in form relative to \(\theta\) as are \(\eta_{YXW}\) and \(\eta_{YXA}\) relative to \(\pi_y\) and \(\pi_X\). It will also be useful to partition \(\Omega_T\) into components which represent variability of \(y_{\gamma\xi}\) taking into account \(x_{\gamma\xi}\) within and between clusters. The derivation is given below:
$$\Omega_T = \nu_+ \sum_{\gamma=1}^{\Gamma} \nu_\gamma \sum_{\xi=1}^{\nu_\gamma} (y_{\gamma \xi} - \theta x_{\gamma \xi})^2$$

$$= \nu_+ \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} (y_{\gamma \xi} - \theta x_{\gamma \xi})^2$$

$$= \nu_+ \sum_{\gamma=1}^{\Gamma} \nu_\gamma \sum_{\xi=1}^{\nu_\gamma} ((y_{\gamma \xi} - \theta x_{\gamma \xi})^2$$

$$+ 2(y_{\gamma \xi} - \theta x_{\gamma \xi})(\theta - \theta) x_{\gamma \xi} + (\theta - \theta)^2 x_{\gamma \xi}^2)$$

$$= \nu_+ \sum_{\gamma=1}^{\Gamma} \nu_\gamma \sum_{\xi=1}^{\nu_\gamma} ((y_{\gamma \xi} - 2y_{\gamma \xi} \theta x_{\gamma \xi} + \theta^2 x_{\gamma \xi}^2)$$

$$+ 2(y_{\gamma \xi} - \theta x_{\gamma \xi})(\theta - \theta) + (\theta - \theta)^2 x_{\gamma \xi}^2)$$

$$= \nu_+ \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} (y_{\gamma \xi} - 2y_{\gamma \xi} \theta x_{\gamma \xi} + \theta^2 x_{\gamma \xi})$$

$$+ 2(y_{\gamma \xi} - \theta x_{\gamma \xi})(\theta - \theta) + (\theta - \theta)^2 x_{\gamma \xi}^2)$$

$$= \nu_+ \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} (y_{\gamma \xi}(1 - 2\theta) + x_{\gamma \xi} \theta^2)$$

$$+ \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} \frac{y_{\gamma \xi} \theta}{\nu_+} (x_{\gamma} - \theta)^2$$

$$= \nu_+ \sum_{\gamma=1}^{\Gamma} (y_{\gamma \xi} - 2\theta y_{\gamma \xi} + \theta^2 x_{\gamma \xi})$$

$$+ \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} \frac{y_{\gamma \xi} \theta}{\nu_+} (x_{\gamma} - \theta)^2$$

$$= \nu_+ \sum_{\gamma=1}^{\Gamma} (y_{\gamma \xi} - \theta y_{\gamma \xi} + \theta^2 x_{\gamma \xi})$$

$$+ \sum_{\gamma=1}^{\Gamma} \sum_{\xi=1}^{\nu_\gamma} \frac{y_{\gamma \xi} \theta}{\nu_+} (x_{\gamma} - \theta)^2$$
\[
= \frac{1}{V^+} \sum_{\gamma = 1}^{\Gamma} \nu_{X \gamma} \theta_{\gamma}(1 - \theta_{\gamma}) + \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{X \gamma}}{V^+} (\theta_{\gamma} - \theta)^2,
\]

since \(\theta_{\gamma} = \nu_{X \gamma}/(\nu_{X \gamma}^+ + \nu_{X \gamma}^-)\). Moreover, this partition of \(\Omega_T\) can be further refined to

\[
= \sum_{\gamma = 1}^{\Gamma} \left\{ \nu_{X \gamma} \theta_{\gamma}(1 - \theta_{\gamma}) + \nu_{X \gamma}(\theta_{\gamma} - \theta)^2 \right\} \frac{\nu_{X \gamma}^+}{\nu_{X \gamma}^+ + \nu_{X \gamma}^-}
\]

\[
= \pi_X \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{X \gamma}}{V^+} \theta_{\gamma}(1 - \theta_{\gamma}) + \pi_X \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{X \gamma}}{V^+} (\theta_{\gamma} - \theta)^2
\]

\[
= \Omega_W + \Omega_A, \tag{5.17}
\]

where

\[
\Omega_W = \pi_X \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{X \gamma}}{V^+} \theta_{\gamma}(1 - \theta_{\gamma})
\]

and

\[
\Omega_A = \pi_X \sum_{\gamma = 1}^{\Gamma} \frac{\nu_{X \gamma}}{V^+} (\theta_{\gamma} - \theta)^2.
\]

Note that

\[
\frac{\Omega_W}{\pi_X} = \sum_{\gamma \text{ with } \nu_{X \gamma} > 0} \frac{\nu_{X \gamma} \theta_{\gamma}(1 - \theta_{\gamma})}{\nu_{X \gamma}^+}
\]

and

\[
\frac{\Omega_A}{\pi_X} = \sum_{\gamma \text{ with } \nu_{X \gamma} > 0} \frac{\nu_{X \gamma} (\theta_{\gamma} - \theta)^2}{\nu_{X \gamma}^+}
\]

are directly analogous in definition to \(\eta_W\), \(\eta_X\), and \(\eta_A\), \(\eta_A\) for all relevant observations in the population.
5.4 Measures of Intracluster Correlation

Let the population fraction of pairs of elements in cluster $\gamma$ being both relevant and having the attribute be given by

$$\lambda_{\gamma y} = \sum_{\xi \neq \xi'} \frac{y_{\gamma \xi} y_{\gamma \xi'}}{\nu_{\gamma} (\nu_{\gamma} - 1)}.$$  \hfill (5.18)

When all pairs $\xi \xi'$ of elements within cluster $\gamma$ have equal probabilities of selection, then $\lambda_{\gamma y}$ represents the probability that both members of a randomly selected pair of elements from cluster $\gamma$ have the attribute and are relevant. Since

$$\sum_{\xi \neq \xi'} y_{\gamma \xi} y_{\gamma \xi'} = \left( \sum_{\xi = 1}^{\nu_{\gamma}} y_{\gamma \xi} \right)^2 - \sum_{\xi = 1}^{\nu_{\gamma}} y_{\gamma \xi}^2$$

$$= \left( \sum_{\xi = 1}^{\nu_{\gamma}} y_{\gamma \xi} \right)^2 - \sum_{\xi = 1}^{\nu_{\gamma}} y_{\gamma \xi}^2$$

$$= \nu_{y_{\gamma}}^2 - \nu_{y_{\gamma}}$$

$$= \nu_{y_{\gamma}} (\nu_{y_{\gamma}} - 1)$$

$$= \pi_{y_{\gamma}} \nu_{y_{\gamma}} (\pi_{y_{\gamma}} \nu_{y_{\gamma}} - 1),$$

then

$$\lambda_{\gamma y} = \sum_{\xi \neq \xi'} \frac{y_{\gamma \xi} y_{\gamma \xi'}}{\nu_{\gamma} (\nu_{\gamma} - 1)}$$

$$= \frac{\pi_{y_{\gamma}} \nu_{y_{\gamma}} (\pi_{y_{\gamma}} \nu_{y_{\gamma}} - 1)}{\nu_{\gamma} (\nu_{\gamma} - 1)}.$$  \hfill (5.19)
Similarly, let the population fraction of pairs of elements in which both elements of the pair are relevant be given by

$$\lambda_{\gamma\chi} = \sum_{\xi \neq \xi'} \frac{x_{\gamma\xi} x_{\gamma\xi'}}{\nu_{\gamma}(\nu_{\gamma} - 1)}$$

(5.20)

$$= \frac{\pi_{\gamma\nu_{\gamma}}(\pi_{\gamma\nu_{\gamma}} - 1)}{\nu_{\gamma}(\nu_{\gamma} - 1)}.$$  

(5.21)

A measure of intracluster correlation for y may be defined as follows:

$$\rho_y = 1 - \left( \frac{\eta_{yW}}{\eta_{yT}} \right)$$

(5.22)

$$= 1 - \frac{\sum_{\gamma=1}^{s} \nu_{\gamma} \pi_{y\gamma}(1 - \pi_{y\gamma})}{\pi_{y}(1 - \pi_{y})}.$$ 

Since

$$\nu_{\gamma} \pi_{y\gamma}(1 - \pi_{y\gamma}) = \nu_{\gamma} \pi_{y\gamma} - \nu_{\gamma} \pi_{y\gamma}^2$$

$$= \nu_{\gamma} \pi_{y\gamma} - \nu_{\gamma} \pi_{y\gamma}^2 + \pi_{y\gamma} - \pi_{y\gamma}$$

$$= (\nu_{\gamma} - 1) \pi_{y\gamma} - \pi_{y\gamma}(\pi_{y\nu_{\gamma}} - 1)$$
\[\begin{align*}
&= (\nu_\gamma - 1) \pi_{y\gamma} - (\nu_\gamma - 1) \lambda_{y\gamma} \\
&= (\nu_\gamma - 1) (\pi_{y\gamma} - \lambda_{y\gamma}),
\end{align*}\]

the quantity, \(\tilde{\rho}_y\), may be written as a function of the \(\pi_{y\gamma}\) and \(\lambda_{y\gamma}\) as follows:

\[
\begin{align*}
\tilde{\rho}_y &= 1 - \frac{\frac{1}{\nu_+} \sum_{\gamma=1}^{\nu} \nu_\gamma \pi_{y\gamma}(1 - \pi_{y\gamma})}{\pi_y(1 - \pi_y)} \\
&= 1 - \frac{\frac{1}{\nu_+} \sum_{\gamma=1}^{\nu} (\nu_\gamma - 1)(\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)} \\
&= 1 - \frac{\frac{1}{\nu_+} \sum_{\gamma=1}^{\nu} \nu_\gamma(\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)} + \frac{\frac{1}{\nu_+} \sum_{\gamma=1}^{\nu} (\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)} \\
&= 1 - \left\{ \frac{\nu_+ \sum_{\gamma=1}^{\nu} \nu_\gamma \pi_{y\gamma} - \sum_{\gamma=1}^{\nu} \nu_\gamma \lambda_{y\gamma}}{\pi_y(1 - \pi_y)} \right\} \\
&= 1 - \frac{\pi_y - \lambda_y}{\pi_y(1 - \pi_y)} + \frac{\frac{1}{\nu_+} \sum_{\gamma=1}^{\nu} (\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)},
\end{align*}\]

where \(\lambda_y = \frac{1}{\nu_+} \sum_{\gamma=1}^{\nu} \nu_\gamma \lambda_{y\gamma}\) represents the probability that two randomly selected
elements from the same cluster are both relevant and have the attribute with the following sampling process: selection of one cluster from the population with probability proportional to size \( \nu_\gamma \) and selection of one pair of elements within the selected cluster with equal probability. With additional simplification,

\[
\tilde{\rho}_y = 1 - \frac{(\pi_y - \lambda_y)}{\pi_y(1 - \pi_y)} + \frac{1}{\nu_+} \sum_{\gamma = 1}^{\Gamma} \frac{(\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)}
\]

\[
= \frac{\pi_y(1 - \pi_y)}{\pi_y(1 - \pi_y)} (\pi_y - \lambda_y) + \frac{1}{\nu_+} \sum_{\gamma = 1}^{\Gamma} \frac{(\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)}
\]

\[
= \frac{\pi_y - \pi^2_y - \pi_y + \lambda_y}{\pi_y(1 - \pi_y)} + \frac{1}{\nu_+} \sum_{\gamma = 1}^{\Gamma} \frac{(\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)}
\]

\[
= \frac{\lambda_y - \pi^2_y}{\pi_y(1 - \pi_y)} + \frac{1}{\nu_+} \sum_{\gamma = 1}^{\Gamma} \frac{(\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)}
\]

\[
= \rho_y + \frac{1}{\nu_+} \sum_{\gamma = 1}^{\Gamma} \frac{(\pi_{y\gamma} - \lambda_{y\gamma})}{\pi_y(1 - \pi_y)},
\]

where \( \rho_y = (\lambda_y - \pi^2_y)/\pi_y(1 - \pi_y) \) is a usual type of intraclass correlation in the sense of having a numerator which is a covariance for pairs of observations from the same cluster sampled according to the process described for \( \lambda_y \) and a denominator which is a variance for a single observation from that sampling process. One can note that when \( y_{y\xi} \) is the same for all elements within the same cluster (i.e., either \( y_{y\xi} = 1 \) for all \( \xi \) in \( \gamma \) or \( y_{y\xi} = 0 \) for all \( \xi \) in \( \gamma \)), then
\[ \pi_y (1 - \pi_y \gamma) = 0, \quad \pi_y \gamma = \lambda_y \gamma, \text{ and } \pi_y = \lambda_y, \text{ and it follows that } \bar{\rho}_y = \rho_y = 1 \text{ for this situation of complete within cluster homogeneity. Alternatively, when } \pi_y \gamma = \pi_y \text{ for all } \gamma (\text{i.e., each cluster is like the population)}, \bar{\rho}_y = 0 \text{ and } \rho_y < 0 \text{ since} \]

\[
\lambda_y = \left\{ \pi_y^2 - \frac{1}{\nu} \sum_{\gamma=1}^{\nu} \frac{\nu \pi_y (1 - \pi_y)}{\nu - 1} \right\}.
\]

When all \( \nu \gamma \) are very large, \( \bar{\rho}_y = \rho_y \), and so evaluation of \( \bar{\rho}_y \) is sufficient. If clusters have the same number of elements, \( \nu \), so that \( \nu_+ = \nu \gamma \) or \( \nu_\gamma = \nu \), then

\[
\bar{\rho}_y = \rho_y + \frac{\pi_y - \lambda_y}{\nu \pi_y (1 - \pi_y)}
\]

\[= \rho_y + \frac{(1 - \rho_y)}{\nu}\]

\[= \frac{1 + (\nu - 1) \rho_y}{\nu}\]

Here, it is easy to see that when \( \nu \) is large, \( \bar{\rho}_y = \rho_y \). A measure of intracluster correlation for \( x \) is given by

\[
\bar{\rho}_x = 1 - \left( \frac{\eta_{xW}}{\eta_{XT}} \right) \quad (5.24)
\]

\[
= 1 - \frac{\frac{1}{\nu_+} \sum_{\gamma=1}^{\nu} \nu \pi_y \gamma(1 - \pi_y \gamma)}{\pi_x (1 - \pi_x)} \quad (5.25)
\]
\[ = \rho_X + \frac{1}{\nu^+} \sum_{\gamma=1}^{\nu^-} \left( \pi_{X\gamma} - \lambda_{X\gamma} \right) \rho_{X\gamma} \left( 1 - \pi_{X\gamma} \right) \]

Its properties are the same as those noted above for \( \tilde{\rho}_{\gamma} \).

Define

\[ \lambda_{y/X,\gamma} = \frac{\lambda_{y\gamma}}{\lambda_{x\gamma}} \tag{5.26} \]

as the population fraction of pairs of distinct elements in the \( \gamma \)-th cluster to which relevance and having the attribute jointly apply, divided by the population fraction of such pairs to which relevance jointly applies. This quantity may be simplified as follows:

\[ \lambda_{y/X,\gamma} = \frac{\lambda_{y\gamma}}{\lambda_{x\gamma}} \]

\[ = \frac{\pi_{y\gamma} \nu_{\gamma} (\pi_{y\gamma} \nu_{\gamma} - 1) / \nu_{\gamma} (\nu_{\gamma} - 1)}{\pi_{x\gamma} \nu_{\gamma} (\pi_{x\gamma} \nu_{\gamma} - 1) / \nu_{\gamma} (\nu_{\gamma} - 1)} \]

\[ = \frac{\pi_{y\gamma} (\pi_{y\gamma} \nu_{\gamma} - 1)}{\pi_{x\gamma} (\pi_{x\gamma} \nu_{\gamma} - 1)} \]

\[ = \theta_{\gamma} (\pi_{x\gamma} \nu_{\gamma} - 1). \tag{5.27} \]

Note that when \( \nu_{\gamma} \) is large,

\[ \frac{(\pi_{y\gamma} \nu_{\gamma} - 1)}{(\pi_{x\gamma} \nu_{\gamma} - 1)} \approx \frac{\pi_{y\gamma}}{\pi_{x\gamma}} = \theta_{\gamma} \]

and

\[ \lambda_{y/X,\gamma} \approx \theta_{\gamma}^2. \tag{5.28} \]
Define

\[
\hat{\rho}_{y/x} = 1 - \frac{\Omega_W}{\Omega_T} = 1 - \frac{\sum_{\gamma=1}^{\Gamma} \frac{\pi_{x,\gamma}}{\nu_x + \theta(1-\theta)} \theta_{y}(1 - \theta_{y})}{\pi_x(1 - \theta)} \tag{5.29}
\]

as a measure of intracluster correlation for \( y \) taking into account \( x \). Since

\[
\theta_{y} - \lambda_{y/x,\gamma} = \theta_{y} - \frac{(\pi_{y,\gamma} \nu_{y,\gamma} - 1)}{\pi_{x,\gamma}(\nu_{x,\gamma} - 1)}
\]

\[
= \theta_{y} \left[ 1 - \frac{(\pi_{y,\gamma} \nu_{y,\gamma} - 1)}{(\pi_{x,\gamma}(\nu_{x,\gamma} - 1))} \right]
\]

\[
= \theta_{y} \left[ \frac{(\pi_{y,\gamma} \nu_{y,\gamma} - 1) - (\pi_{y,\gamma} \nu_{y,\gamma} - 1)}{(\pi_{x,\gamma}(\nu_{x,\gamma} - 1))} \right]
\]

\[
= \frac{\theta_{y} \pi_{x,\gamma} \nu_{y}(1 - \theta_{y})}{(\nu_{x,\gamma} - 1)}
\]

\[
= \frac{\nu_{x,\gamma} \theta_{y}(1 - \theta_{y})}{(\nu_{x,\gamma} - 1)}
\]

then \( \hat{\rho}_{y/x} \) may be written as a function of \( \theta_{y} \) and \( \lambda_{y/x,\gamma} \) as follows:

\[
\hat{\rho}_{y/x} = 1 - \frac{\Omega_W}{\Omega_T}
\]

\[
= 1 - \sum_{\gamma=1}^{\Gamma} \frac{\nu_{x,\gamma} \theta_{y}(1 - \theta_{y})}{\nu_x + \theta(1 - \theta)}
\]

\[
= 1 - \sum_{\gamma=1}^{\Gamma} \frac{(\nu_{x,\gamma} - 1)(\theta_{y} - \lambda_{y/x,\gamma})}{\nu_x + \theta(1 - \theta)}
\]
where \( \lambda_{y/x} = \frac{\sum_{\gamma=1}^{\Gamma} \nu_{x_{\gamma}y/x_{\gamma}}}{\nu_{x+}} \) represents the probability that two randomly selected elements from the same cluster have the attribute when both are relevant, in the idealized situation where one cluster is selected from the population with probability proportional to \( \nu_{x_{\gamma}} \) (relative to \( \nu_{x+} \)) and one pair of relevant elements within that cluster is selected with equal probability (relative to all such pairs). With additional simplification,

\[
\tilde{\rho}_{y/x} = \frac{\lambda_{y/x} - \varrho^2}{\theta (1 - \theta) + \frac{1}{\nu_{x+} \theta (1 - \theta)}} \sum_{\gamma=1}^{\Gamma} (\theta_{\gamma} - \lambda_{y/x, \gamma})
= \rho_{y/x} + \frac{1}{\nu_{x+} \theta (1 - \theta)} \sum_{\gamma=1}^{\Gamma} (\theta_{\gamma} - \lambda_{y/x, \gamma}),
\]

(5.30)

where \( \rho_{y/x} = \frac{\lambda_{y/x} - \varrho^2}{\theta (1 - \theta)} \) is a usual type of intraclass correlation in the sense of having a numerator which is a covariance for paired relevant observations from the same cluster under the idealized sampling process described for \( \lambda_{y/x} \) and a
denominator which is a variance for that sampling process. Note that \( \tilde{\beta}_{y/x} \geq \rho_{y/x} \) since \( \theta_{y} \geq \lambda_{y/x, \gamma} \). When \( \nu_{x/+} \) is large relative to \( \Gamma \) (that is, the average number of relevant observations per cluster, \( \frac{\nu_{x/+}}{\Gamma} \), is large), \( \tilde{\beta}_{y/x} \approx \rho_{y/x} \). If clusters have the same number of relevant observations so that \( \nu_{x/+} = \nu_{x} \Gamma \) or \( \nu_{x/\gamma} = \nu_{x} \), then

\[
\tilde{\beta}_{y/x} = \frac{\theta - \lambda_{y/x}}{\theta (1 - \theta) \nu_{x}}
\]

\[
= \rho_{y/x} + \frac{(1 - \rho_{y/x})}{\nu_{x}}
\]

\[
= \frac{1}{\nu_{x}} + \frac{(\nu_{x} - 1)}{\nu_{x}} \rho_{y/x}.
\]

Again, when \( \nu_{x} \) is large, \( \tilde{\beta}_{y/x} \approx \rho_{y/x} \). Also note that when \( \theta_{\gamma} = \theta \), for all \( \gamma = 1, 2, ..., \Gamma \), then

\[
\lambda_{y/x, \gamma} = \frac{(\nu_{\gamma} \pi_{x, \gamma} \theta - 1)}{(\nu_{\gamma} \pi_{x, \gamma} - 1)},
\]

and \( \lambda_{y/x} \) simplifies to

\[
\lambda_{y/x} = \frac{\sum_{\gamma=1}^{\Gamma} \nu_{x\gamma} \lambda_{y/x, \gamma}}{\nu_{x/+}}
\]

\[
= \sum_{\gamma=1}^{\Gamma} \frac{\nu_{x\gamma} \theta (\nu_{\gamma} \pi_{x, \gamma} \theta - 1)}{(\nu_{\gamma} \pi_{x, \gamma} - 1)}
\]

\[
= \theta \sum_{\gamma=1}^{\Gamma} \frac{\nu_{x\gamma} (\nu_{\gamma} \theta - 1)}{(\nu_{\gamma} - 1)}
\]
\[
\theta = \sum_{\gamma=1}^{\Gamma} \left( \frac{\nu_{x\gamma}^2}{\nu_{x}^2} + \frac{\nu_{x\gamma}^2 (\theta - 1)}{(\nu_{x\gamma} - 1) \nu_{x}} \right)
\]

\[
= \theta^2 \frac{\sum_{\gamma=1}^{\Gamma} \nu_{x\gamma}}{\nu_{x}} + \sum_{\gamma=1}^{\Gamma} \frac{\nu_{x\gamma}^2 (\theta - 1)}{(\nu_{x\gamma} - 1) \nu_{x}}
\]

\[
= \theta^2 - \sum_{\gamma=1}^{\Gamma} \frac{\nu_{x\gamma}^2 (\theta - 1)}{(\nu_{x\gamma} - 1) \nu_{x}}
\]

Thus, when \( \theta_{\gamma} = \theta \) for all \( \gamma \), \( \rho_{y/x} = -\sum_{\gamma=1}^{\Gamma} \frac{\nu_{x\gamma}}{\nu_{x} (\nu_{x\gamma} - 1)} < 0 \), and this simplifies to \( \rho_{y/x} = 0 \) for situations where all \( \nu_{x\gamma} \) are large. Also, in this case of \( \theta_{\gamma} = \theta \) for all \( \gamma \), \( \hat{\rho}_{y/x} = 0 \). Alternatively, \( \hat{\rho}_{y/x} = \rho_{y/x} = 1 \) when all clusters \( \gamma \) have \( \theta_{\gamma} = 0 \) or \( \theta_{\gamma} = 1 \) (i.e., when they are all internally homogeneous for attribute status given relevance). Since \( \hat{\rho}_{y/x} \) has the essential properties of an intraclass correlation coefficient, its evaluation is sufficient for evaluation of \( \rho_{y/x} \).

5.5 Estimation for One-Stage Cluster Samples

5.5.1 General Case

Let \( N \) clusters be selected from the population at random with replacement and let \( \varphi_{\gamma} \) be the probability of selection for cluster \( \gamma \). Let \( i = 1, \ldots, N \) index selected clusters and \( \xi = 1, 2, \ldots, \nu_{1} \) index elements of sample cluster \( i \), where \( \nu_{1} \) is the total number of elements in sample cluster \( i \). While the focus of this chapter is on estimating \( \theta \) and \( \hat{\rho}_{y/x} \) and their variance-covariance matrix with respect to the sampling scenario just described, individual estimates of
intermediate quantities of interest will be presented first.

An estimate of $\pi_y$ is given by

$$
\hat{\pi}_y = \frac{1}{N} \sum_{i=1}^{N} \frac{\nu_{yi}}{\varphi_i \nu_+ ^+} 
$$

$$
= \frac{1}{N} \sum_{i=1}^{N} \hat{\pi}_i y_i , 
$$

where $\hat{\pi}_i y_i = \frac{\nu_{yi}}{\varphi_i \nu_+ ^+}$. An estimate of $\pi_x$ is given by

$$
\hat{\pi}_x = \frac{1}{N} \sum_{i=1}^{N} \frac{\nu_{xi}}{\varphi_i \nu_+ ^+} 
$$

$$
= \frac{1}{N} \sum_{i=1}^{N} \hat{\pi}_i x_i , 
$$

where $\hat{\pi}_i x_i = \frac{\nu_{xi}}{\varphi_i \nu_+ ^+}$. Therefore, an estimate of $\theta$ is given by

$$
\hat{\theta} = \frac{\hat{\pi}_y}{\hat{\pi}_x} . 
$$

The expected value of $\hat{\pi}_y$ is

$$
E\{\hat{\pi}_y\} = E\{ \frac{1}{N} \sum_{i=1}^{N} \frac{\nu_{yi}}{\varphi_i \nu_+ ^+} \}
$$

$$
= \sum_{\gamma=1}^{\Gamma} \varphi_\gamma \frac{\nu_{y\gamma}}{\varphi_{y\gamma} \nu_+ ^+} 
$$

$$
= \pi_y
$$
and, similarly, the expected value of $\hat{r}_X$ is

$$E\{\hat{r}_X\} = \tau_X.$$  

The variance of $\hat{r}_Y$ is

$$\text{Var}\{\hat{r}_Y\} = \text{Var}\left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{\nu y_i}{\phi Y_i} \right\}$$

$$= \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \phi_{Y \gamma} \left( \frac{\nu y_{\gamma}}{\phi_{Y \gamma}} - \tau_Y \right)^2$$

$$= \frac{\sigma^2_{\nu, Y}}{N}, \quad (5.34)$$

and, similarly, the variance of $\hat{r}_X$ is

$$\text{Var}\{\hat{r}_X\} = \text{Var}\left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{\nu x_i}{\phi X_i} \right\}$$

$$= \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \phi_{X \gamma} \left( \frac{\nu x_{\gamma}}{\phi_{X \gamma}} - \tau_X \right)^2$$

$$= \frac{\sigma^2_{\nu, X}}{N}. \quad (5.35)$$

The covariance between $\hat{r}_Y$ and $\hat{r}_X$ is given by

$$\text{Cov}\{\hat{r}_Y, \hat{r}_X\} = \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \phi_{Y \gamma} \left\{ \frac{\nu y_{\gamma}}{\phi_{Y \gamma}} - \tau_Y \right\} \left( \frac{\nu x_{\gamma}}{\phi_{X \gamma}} - \tau_X \right)$$

$$= \frac{\sigma_{\nu, YX}}{N}. \quad (5.36)$$
The asymptotic variance of \( \hat{\beta} \), based on a Taylor series representation for \( \hat{\beta} \) at \((\pi_y, \pi_x)\) is

\[
\text{Var}_A(\hat{\beta}) = \sigma^2 \left\{ \frac{\sigma_{\phi,y}^2}{N \pi_y^2} - 2 \frac{\sigma_{\phi,yx}}{N \pi_y \pi_x} + \frac{\sigma_{\phi,x}^2}{N \pi_x^2} \right\}.
\] (5.37)

Since the \((\hat{\pi}_y, \hat{\pi}_x)\) are independent and identically distributed by virtue of the with replacement simple random sampling of clusters, unbiased estimates for \(\sigma_{\phi,y}^2\), \(\sigma_{\phi,x}^2\), and \(\sigma_{\phi,yx}^2\) are given by:

\[
\hat{\sigma}_{\phi,y}^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left( \frac{\nu_{yi}}{\nu_{i+y}^+ - \hat{\pi}_y} \right)^2,
\] (5.38)

\[
\hat{\sigma}_{\phi,x}^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left( \frac{\nu_{xi}}{\nu_{i+x}^+ - \hat{\pi}_x} \right)^2,
\] (5.39)

and

\[
\hat{\sigma}_{\phi,yx}^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left( \frac{\nu_{yi}}{\nu_{i+y}^+ - \hat{\pi}_y} \right) \left( \frac{\nu_{xi}}{\nu_{i+x}^+ - \hat{\pi}_x} \right),
\] (5.40)

respectively. Hence, an asymptotic, unbiased estimator for the variance of \( \hat{\beta} \), based on the Taylor series approximation is

\[
\text{Var}_A(\hat{\beta}) = \sigma^2 \left\{ \frac{\hat{\sigma}_{\phi,y}^2}{N \hat{\pi}_y^2} - 2 \frac{\hat{\sigma}_{\phi,yx}}{N \hat{\pi}_y \hat{\pi}_x} + \frac{\hat{\sigma}_{\phi,x}^2}{N \hat{\pi}_x^2} \right\}.
\] (5.41)

An alternative form for \( \text{Var}_A(\hat{\beta}) \) may be derived as follows:

\[
\text{Var}_A(\hat{\beta}) = \sigma^2 \left\{ \frac{\hat{\sigma}_{\phi,y}^2}{N \hat{\pi}_y^2} - 2 \frac{\hat{\sigma}_{\phi,yx}}{N \hat{\pi}_y \hat{\pi}_x} + \frac{\hat{\sigma}_{\phi,x}^2}{N \hat{\pi}_x^2} \right\}.
\] (5.42)

\[
= \frac{1}{N \hat{\pi}_x^2} \left\{ \hat{\sigma}_{\phi,y}^2 - 2 \hat{\sigma}_{\phi,yx} \hat{\pi}_y \hat{\pi}_x + \hat{\sigma}_{\phi,x}^2 \hat{\pi}_y^2 \right\}
\]
\[
= \frac{1}{N\hat{\sigma}_x^2} \left\{ \sum_{i=1}^{N} \frac{\nu_{yi} - \hat{\nu}_y}{(N-1)} \right. \\
\left. - 2\hat{\theta} \sum_{i=1}^{N} \frac{\nu_{yi} - \hat{\nu}_y}{(N-1)} \left\{ \frac{\nu_{xi} - \hat{\nu}_x}{(N-1)} \right. \\
+ \hat{\sigma}^2 \sum_{i=1}^{N} \frac{\nu_{xi} - \hat{\nu}_x}{(N-1)} \right\} \right.
\]

\[
= \frac{1}{N(N-1)\hat{\sigma}_x^2} \sum_{i=1}^{N} \left\{ \nu_{yi}^2 - 2\hat{\nu}_y (\nu_{yi}) + \hat{\nu}_y^2 \right\} \\
- 2\hat{\theta} \frac{\nu_{yi}}{\nu_{yi}} \frac{\nu_{xi}}{\nu_{yi}} - \frac{\nu_{yi}}{\nu_{yi}} \hat{\nu}_x - \frac{\nu_{xi}}{\nu_{yi}} \hat{\nu}_x + \hat{\nu}_x + \hat{\nu}_y \hat{\nu}_x \\
+ \hat{\sigma}^2 \left\{ (\nu_{xi})^2 - 2 (\nu_{yi} \nu_{xi}) + \hat{\nu}_x^2 \right\} \\
= \frac{1}{N(N-1)\hat{\sigma}_x^2} \sum_{i=1}^{N} \frac{(\nu_{yi} - \hat{\nu}_x)^2}{(N-1)\hat{\sigma}_x^2} \\
= \frac{N}{(N-1)} \sum_{i=1}^{N} \frac{(\nu_{yi} - \hat{\nu}_x)^2}{\nu_{yi}^2} \left\{ \frac{\nu_{xi}^2}{\nu_{yi}^2} \right\}^{-2}.
\]

(5.43)
Let an unbiased estimator for $\hat{\eta}_{yw}$ be given by

$$
\hat{\eta}_{yw} = \frac{1}{N} \sum_{i=1}^{N} \frac{\nu_i y_i (1 - y_i)}{\varphi_i \nu_i} ,
$$

(5.44)

and since $\nu_i y_i (1 - y_i) = (\nu_i - 1)(y_i - y_i)$, then an alternate form for $\hat{\eta}_{yw}$ is

$$
\hat{\eta}_{yw} = \frac{1}{N} \sum_{i=1}^{N} \frac{(\nu_i - 1)(y_i - y_i)}{\varphi_i \nu_i} .
$$

(5.45)

The expected value of $\hat{\eta}_{yw}$ is

$$
E\{\hat{\eta}_{yw}\} = E \left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{(\nu_i - 1)(y_i - y_i)}{\varphi_i \nu_i} \right\} \\
= \sum_{\gamma=1}^{\Gamma} \varphi_{\gamma} \frac{\nu_{\gamma} - 1)(y_{\gamma} - y_{\gamma})}{\varphi_{\gamma} \nu_{\gamma}} \\
= \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} (\nu_{\gamma} - 1) \left( y_{\gamma} - \frac{\nu_{\gamma} y_{\gamma} (y_{\gamma} - \nu_{\gamma} - 1)}{\nu_{\gamma} (\nu_{\gamma} - 1)} \right) \\
= \frac{1}{\nu_+} \sum_{\gamma=1}^{\Gamma} \pi_{\gamma} (1 - y_{\gamma}) \nu_{\gamma} \\
= \eta_{yw}. 
$$

The variance of $\hat{\eta}_{yw}$ is

$$
\text{Var}\{\hat{\eta}_{yw}\} = \text{Var} \left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{(\nu_i - 1)(y_i - y_i)}{\varphi_i \nu_i} \right\} \\
= \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \varphi_{\gamma} \left\{ \frac{(\nu_{\gamma} - 1)(y_{\gamma} - y_{\gamma})}{\varphi_{\gamma} \nu_+} - \eta_{yw} \right\}^2
$$
\[ = \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \varphi_{\gamma} \left( \frac{\nu_{\gamma} \pi_{\gamma} \gamma (1 - \pi_{\gamma} \gamma)}{\varphi_{\gamma} \nu_{\gamma} +} - \eta_{yW} \right)^2 \]  
(5.46)

An unbiased variance estimator for \( \text{Var}\{\hat{\eta}_{yW}\} \) is given by

\[ \text{Var}\{\hat{\eta}_{yW}\} = \frac{1}{N} \left( \frac{1}{(N-1)} \right) \sum_{i=1}^{N} \left( \frac{\nu_{y} \pi_{y} (1 - \pi_{y} I)}{\varphi_{y} \nu_{y} +} - \hat{\eta}_{yW} \right)^2 \]  
(5.47)

\[ = \frac{\sigma_{\varphi_{y} \eta_{y}}^{2}}{N}. \]

Similarly, an unbiased estimator for \( \eta_{xW} \) is given by

\[ \hat{\eta}_{xW} = \frac{1}{N} \sum_{i=1}^{N} \frac{\nu_{i} \pi_{i} x_{i} (1 - \pi_{i} x_{i})}{\varphi_{i} \nu_{i} +} \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \frac{(\nu_{i} - 1) (\pi_{i} x_{i} - \lambda_{i} x_{i})}{\varphi_{i} \nu_{i} +}. \]  
(5.48)

The expected value of \( \hat{\eta}_{xW} \) is

\[ E\{\hat{\eta}_{xW}\} = \nu_{+} \sum_{\gamma=1}^{\Gamma} \pi_{x \gamma} (\pi_{x \gamma} - 1) \nu_{\gamma} \]

\[ = \eta_{xW} \]

and the variance of \( \hat{\eta}_{xW} \) is

\[ \text{Var}\{\hat{\eta}_{xW}\} = \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \varphi_{\gamma} \left( \frac{(\nu_{\gamma} - 1) (\pi_{x \gamma} - \lambda_{x \gamma})}{\varphi_{\gamma} \nu_{\gamma} +} - \eta_{xW} \right)^2 \]

\[ = \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \varphi_{\gamma} \left( \frac{\nu_{\gamma} \pi_{x \gamma} (1 - \pi_{x \gamma})}{\varphi_{\gamma} \nu_{\gamma} +} - \eta_{xW} \right)^2. \]  
(5.49)
An unbiased variance estimator for $\text{Var}\{\hat{\eta}_{xW}\}$ is given by

$$\text{Var}\{\hat{\eta}_{xW}\} = \frac{1}{N} \frac{1}{(N-1)} \sum_{i=1}^{N} \left\{ \nu_{i} x_{i} \left( 1 - \frac{\nu_{i}}{\nu_{i}^{+}} \right) - \hat{\eta}_{xW} \right\}^2$$

$$= \frac{\sigma_\phi^2 \eta_{x} \eta_{xW}}{N}.$$

The covariance between $\hat{\eta}_{y}$ and $\hat{\eta}_{yW}$ is

$$\text{Cov}\{\hat{\eta}_{y}, \hat{\eta}_{yW}\} = \text{Cov}\left\{ \frac{1}{N} \sum_{i=1}^{N} \nu_{i} y_{i}, \frac{1}{N} \sum_{i=1}^{N} \nu_{i} x_{i} \right\}$$

$$= \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} \left( \frac{\nu_{\gamma} y_{\gamma}}{\nu_{\gamma}^{+}} - \hat{\eta}_{y} \right) \left( \frac{\nu_{\gamma} x_{\gamma} \nu_{\gamma} \left( 1 - \frac{\nu_{\gamma}}{\nu_{\gamma}^{+}} \right)}{\nu_{\gamma}^{+} - \eta_{yW}} \right).$$

An unbiased estimator for $\text{Cov}\{\hat{\eta}_{y}, \hat{\eta}_{yW}\}$ is given by

$$\text{Cov}\{\hat{\eta}_{y}, \hat{\eta}_{yW}\} = \frac{1}{N} \frac{1}{(N-1)} \sum_{i=1}^{N} \left( \nu_{i} y_{i} - \hat{\eta}_{y} \right) \left( \frac{\nu_{i} x_{i} \left( 1 - \frac{\nu_{i}}{\nu_{i}^{+}} \right)}{\nu_{i}^{+} - \eta_{yW}} \right)$$

$$= \frac{\sigma_\phi \eta_{y} \eta_{yW}}{N}.$$

The covariance between $\hat{\eta}_{xW}$ and $\hat{\eta}_{xW}$ is

$$\text{Cov}\{\hat{\eta}_{xW}, \hat{\eta}_{xW}\} = \text{Cov}\left\{ \frac{1}{N} \sum_{i=1}^{N} \nu_{i} x_{i}, \frac{1}{N} \sum_{i=1}^{N} \nu_{i} x_{i} \right\}$$

$$= \frac{1}{N} \sum_{\gamma=1}^{\Gamma} \nu_{\gamma} \left( \frac{\nu_{\gamma} x_{\gamma}}{\nu_{\gamma}^{+}} - \hat{\eta}_{x} \right) \left( \frac{\nu_{\gamma} x_{\gamma} \nu_{\gamma} \left( 1 - \frac{\nu_{\gamma}}{\nu_{\gamma}^{+}} \right)}{\nu_{\gamma}^{+} - \eta_{xW}} \right).$$

An unbiased estimator for $\text{Cov}\{\hat{\eta}_{x}, \hat{\eta}_{xW}\}$ is given by
\[ \text{Cov}\{\hat{x}_x, \hat{\eta}_{xW}\} = \frac{1}{N} \frac{1}{(N-1)} \sum_{i=1}^{N} \left( \frac{\nu \pi_1}{\varphi_1 \nu_+} - \hat{x}_x \right) \left( \frac{\nu \pi_1 (1 - \pi_1)}{\varphi_1 \nu_+} - \hat{\eta}_{xW} \right) \] (5.54)

\[ = \frac{\delta \varphi \pi \eta_x}{N}. \]

An unbiased estimator for \( \Omega \) is obtained through the quantity \( Q_i \) defined below:

\[ Q_i = \sum_{\xi=1}^{\nu_i} (y_{i\xi} - \theta_1 x_{i\xi})^2, \] and

\[ = 0 \text{ when } y_{i\xi} = x_{i\xi} = 0 \text{ for all } \xi \in i. \] (5.55)

The quantity \( Q_i \) may be rewritten in a more convenient form as follows:

\[ Q_i = \sum_{\xi=1}^{\nu_i} (y_{i\xi} - \theta_1 x_{i\xi})^2 \]

\[ = \sum_{\xi=1}^{\nu_i} (y_{i\xi} - \pi_y + \pi_y - \theta_1 x_{i\xi})^2 \]

\[ = \sum_{\xi=1}^{\nu_i} (y_{i\xi} - \pi_y + \pi_x - \theta_1 x_{i\xi})^2 \]

\[ = \sum_{\xi=1}^{\nu_i} [\pi_y - \pi_y + \theta_1 (\pi_x - x_{i\xi})]^2 \]

\[ = \sum_{\xi=1}^{\nu_i} \left( y_{i\xi} - \pi_y \right)^2 + 2 \theta_1 \sum_{\xi=1}^{\nu_i} \left( y_{i\xi} - \pi_y \right) (\pi_x - x_{i\xi}) \]

\[ + \theta_1^2 \sum_{\xi=1}^{\nu_i} (\pi_x - x_{i\xi})^2 \]

\[ = \nu_i \pi_y (1 - \pi_y) - 2 \theta_1 \nu_i \pi_y (1 - \pi_x) + \theta_1^2 \nu_i \pi_x (1 - \pi_x) \]
\[
\nu \pi y_i (1 - \pi y_i) - \theta^2 \nu \pi x_i (1 - \pi x_i)
\]

\[
\nu_i (\pi y_i - \theta^2 \pi x_i)
\]

\[
\nu \pi x_i (\theta_i - \theta^2_i)
\]

\[
\nu x_i \theta_i (1 - \theta_i).
\] (5.56)

Let an estimator for \( \Omega_W \) be

\[
\hat{\Omega}_W = Q = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{Q_i}{\varphi_{i^-} \varphi_{i^+}} \right).
\] (5.57)

The expected value of \( Q \) is

\[
E\{Q\} = \sum_{\gamma} \varphi_{\gamma} \left( \frac{Q_{\gamma}}{\varphi_{\gamma} \varphi_{\gamma^+}} \right)
\]

\[
= \sum_{\gamma} \nu_{\gamma} \varphi_{\gamma} (1 - \theta_{\gamma})
\]

\[
= \tau_x \sum_{\gamma} \nu_{\gamma} \varphi_{\gamma} (1 - \theta_{\gamma})
\]

\[
= \Omega_W.
\]

The variance of \( Q \) is

\[
\text{Var}(Q) = \frac{1}{N} \sum_{\gamma} \varphi_{\gamma} \left( \frac{Q_{\gamma}}{\varphi_{\gamma} \varphi_{\gamma^+}} - \Omega_W \right)^2
\]
\[ \begin{align*}
\frac{\sigma^2_Q}{N} & = \sigma^2_Q \tag{5.58} \\
\text{An unbiased variance estimator for } \text{Var}(Q) \text{ is given by} \\
\text{Var}(Q) & = \frac{1}{N} \left( \frac{1}{N-1} \right) \sum_{i=1}^{N} \left( \frac{Q_i}{x_i} - Q \right)^2. \tag{5.59}
\end{align*} \]

An estimator for \( \hat{\rho}_y \) is given by

\[ \hat{\rho}_y = 1 - \frac{\hat{\eta}_W}{\hat{\eta}_T} \]

\[ = 1 - \frac{\hat{\eta}_W}{\hat{\xi}_y(1 - \hat{\xi}_y)}, \tag{5.60} \]

and, likewise, an estimator for \( \hat{\rho}_x \) is given by

\[ \hat{\rho}_x = 1 - \frac{\hat{\eta}_W}{\hat{\eta}_T} \]

\[ = 1 - \frac{\hat{\eta}_W}{\hat{\xi}_x(1 - \hat{\xi}_x)}. \tag{5.61} \]

An estimator of \( \hat{\rho}_{y/x} \) is given by

\[ \hat{\rho}_{y/x} = 1 - \frac{\hat{\Omega}_W}{\hat{\Omega}_T} \]

\[ = 1 - \frac{\hat{\Omega}_W}{\hat{\xi}_x \hat{\delta}(1 - \hat{\delta})}. \]
\[ = 1 - \frac{\hat{\Omega}_W}{\hat{x}_y(\hat{x}_x - \hat{x}_y)} \]
\[ = 1 - \frac{\hat{\Omega}_W \hat{x}_x}{\hat{x}_y(\hat{x}_x - \hat{x}_y)}. \]  \hspace{1cm} (5.62)

The variances of \( \hat{\rho}_y, \hat{\rho}_x, \) and \( \hat{\rho}_{y/x} \) may be estimated by Taylor series approximations since \( \hat{\rho}_y, \hat{\rho}_x, \) and \( \hat{\rho}_{y/x} \) are functions of \( \hat{\eta}_W, \hat{s}_y, \hat{s}_x, \) and \( \hat{\Omega}_W, \) and variance estimates for these quantities are given above.

The main objective of this section is to produce joint estimates of \( \theta \) and \( \hat{\rho}_{y/x} \) and their variance-covariance matrix. Since \( \pi_x \) and \( \rho_x \) and their variance-covariance matrix may be of secondary interest, the remainder of this section will be directed at estimating the vector \( (\pi_x, \rho_x, \theta, \rho_{y/x})' \) and its variance-covariance matrix. The vector \( (\pi_x, \pi_y, \eta_{xW}, \Omega_W)' \) and its variance-covariance matrix is estimated first, and then transformed by a series of linear, logarithmic, and exponential operations to obtain an estimate of \( (\pi_x, \rho_x, \theta, \rho_{y/x})' \). The variance-covariance matrix for this vector may then be approximated by the Taylor series method.

Let
\[ F_i = \frac{1}{\varphi_1 \varphi_+} \{ \nu_{xi}, \nu_{yi}, \nu_{Xi}(1 - \pi_{xi}), \nu_{xi}(1 - \pi_{yi}) \}'. \]

Then an unbiased estimate of \( (\pi_x, \pi_y, \eta_{xW}, \Omega_W)' \) is given by
\[ \bar{F} = \frac{1}{N} \sum_{i=1}^{N} F_i \]
\[ = (\hat{s}_x, \hat{s}_y, \hat{\eta}_{xW}, \hat{\Omega}_W)'. \]  \hspace{1cm} (5.63)
An asymptotically unbiased estimator for the variance-covariance matrix of 
\( \bar{\Sigma}_\epsilon = (\hat{\pi}_y, \hat{\pi}_x, \hat{\eta}_{xW}, \hat{\Omega}_W)' \) is given by

\[
V_{\bar{\Sigma}_\epsilon} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\bar{\epsilon}_i - \bar{\bar{\epsilon}})(\bar{\bar{\epsilon}}_i - \bar{\bar{\epsilon}})'.
\] (5.64)

To produce an estimate of \((\pi_x, \hat{\rho}_x, \theta, \hat{\rho}_{y/x})'\), the vector \(\bar{\bar{\epsilon}}\) is transformed by the following series of linear, logarithmic, and exponential operations:

\[
f_3 = A_3 \{ \exp [A_2 (\ln \{A_1 \bar{\epsilon} + C_1\} ) ] \} + C_2,
\] (5.65)

\[
= (\hat{\pi}_x, \hat{\rho}_x, \hat{\theta}, \hat{\rho}_{y/x})'.
\]

where

\[
A_1 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
1 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix},
\]

\[
A_2 = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
-1 & -1 & 0 & 0 & 1 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & -1 & -1 & 0 & 1
\end{bmatrix}.
\]
\[
A_3 = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1 \\
\end{bmatrix},
\]

\[
C_1 = \begin{bmatrix}
0 \\
1 \\
0 \\
0 \\
0 \\
\end{bmatrix},
\]
and

\[
C_2 = \begin{bmatrix}
0 \\
1 \\
0 \\
1 \\
\end{bmatrix}
\]

An asymptotically unbiased estimator for the variance-covariance matrix of \((\hat{\sigma}_x, \hat{\sigma}_x, \hat{\hat{\sigma}}, \hat{\hat{\hat{\sigma}}}_y / x)'\) based on a Taylor series approximation is given by

\[
\text{Vár}(f_3) = A_3 D_{f_2} A_2 D_{f_1}^{-1} A_1 X \bar{X}^{-1} A_1' D_{f_1}^{-1} A_2' D_{f_2} A_3',
\]

where

\[
f_1 = A_1 \bar{X} + C_1,
\]

(5.66)

(5.67)
\[ f_2 = \exp \left\{ A_2 \left( \ln f_1 \right) \right\}, \tag{5.68} \]

\( \mathcal{D}_{f_1} \) is a diagonal matrix with the elements of \( f_1 \) on the diagonal, and \( \mathcal{D}_{f_2} \) is a diagonal matrix with the elements of \( f_2 \) on the diagonal.

5.5.2 **Special Cases**

In some situations information about \( \nu_+ \) may not be available. It is noted that estimation of the quantities \( \theta \) and \( \bar{\rho}_{y/x} \) do not require knowledge of \( \nu_+ \) because the constant \( \nu_+ \) cancels from both the numerator and denominator of the estimators for \( \theta \), \( \bar{\rho}_{y/x} \), and each entry of the estimate of their variance-covariance matrix. Furthermore, if clusters are selected with equal probability, then \( \varphi_i = \frac{1}{N} \), \( i = 1, \ldots, N \), and \( \theta \) and \( \bar{\rho}_{y/x} \) may be estimated without specifying either \( \Gamma \) or \( \nu_+ \) as follows:

Let

\[ \mathbf{F}_1^* = \{ \nu_{xi}, \nu_{yi}, \nu_{xi1}(1 - \theta_1) \}^t. \]

Then an unbiased estimate of \( (\pi_x^*, \pi_y^*, \Omega_{W}^*)^t \) is given by

\[ \mathbf{F}^* = \frac{1}{N} \sum_{i=1}^{N} \mathbf{F}_1^* \tag{5.69} \]

\[ = (\hat{\pi}_x^*, \hat{\pi}_y^*, \hat{\Omega}_W^*)^t. \]

An asymptotically unbiased estimator for the variance-covariance matrix of \( \mathbf{F}^* = (\pi_x^*, \pi_y^*, \Omega_{W}^*)^t \) is given by
\[ \mathcal{X}_{\vec{E}^*} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\mathcal{E}_{i}^* - \bar{\mathcal{E}}^*)(\mathcal{E}_{i}^* - \bar{\mathcal{E}}^*)' \]  

(5.70)

To produce an estimate of \((\theta, \hat{\rho}_{y/x}')\), the vector \(\vec{E}^*\) is transformed by a series of linear, log, and exponential operations to produce the vector \(\vec{F}^* = (\hat{\theta}, \hat{\rho}_{y/x}')\) as follows:

\[ \vec{F}^* = (\hat{\theta}^*, \hat{\rho}_{y/x}'^*) \]

\[ = A_3^* \{ \exp \{ A_2^* \ln \{ A_1^* \vec{E}^* \} \} \} + \zeta_2^* \]  

(5.71)

where

\[ A_1^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \]

\[ A_2^* = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \]

\[ A_3^* = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \]
\[ \mathcal{C}_2^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \]

An asymptotically unbiased estimator for the variance-covariance matrix of 
\((\hat{\theta}^*, \hat{\rho}^*_{y/x})^\prime\) based on a Taylor series approximation is given by

\[ \text{Var}\{ \mathbf{f}_3 \} = A_3^* D_{\mathbf{f}_2}^* A_2^* D_{\mathbf{f}_1}^{-1} A_1^* \mathbf{V}_{\mathbf{f}_2}^* A_1^\prime D_{\mathbf{f}_1}^{-1} A_2^\prime D_{\mathbf{f}_2}^* A_3^\prime, \]  

\[ (5.72) \]

where

\[ \mathbf{f}_1^* = A_1^* \mathbf{f}_1^*, \]

and

\[ \mathbf{f}_2^* = \exp \left\{ A_2^* \left( \ln \mathbf{f}_1^* \right) \right\}, \]

and where \( D_{\mathbf{f}_1} \) is a diagonal matrix with the elements of \( \mathbf{f}_1^* \) on the diagonal and \( D_{\mathbf{f}_2} \) is a diagonal matrix with the elements of \( \mathbf{f}_2^* \) on the diagonal.

**5.5.3 Weighted Regression Applications**

The vectors \( \mathbf{f}_3 \) (or \( \mathbf{f}_3^* \)) can be determined for each of several strata such as those based on different regions. By use of the corresponding estimated variance-covariance matrices, models to describe variation across strata can be fit by weighted least squares methods in ways similar to those described in Section 2.6. Applications of such analysis are described in Sections 5.6.2 and 5.6.3.
5.6 **Examples**

5.6.1 **Example 1**

In this example, $\theta$ and $\hat{\beta}_{y/x}$ were jointly estimated for the one-stage cluster sample of dental practices described in Section 2.7. Here, $\theta$ represents the overall population fraction of persons who have at least one crown in a molar among persons with at least one restoration with greater than or equal to three surfaces in a molar. The parameter $\hat{\beta}_{y/x}$ represents the intracluster correlation for having a crown in a molar, taking into account the variation among practices with respect to the number of persons with a restoration on at least three surfaces in a molar. The total number of persons in the sample with at least one crown in a molar was 6642 and the total number of persons in the sample with at least one restoration with greater than or equal to three surfaces in a molar was 13,674. (Refer to Section 2.7 for further description of characteristics of the sample).

An estimate of the vector $\hat{\mathbf{F}}^* = (\hat{\sigma}_X^*, \hat{\sigma}_W^*)'$ and its covariance matrix, $\hat{\mathbf{V}}_{\hat{\mathbf{F}}^*}$, are

\[
\begin{bmatrix}
38.1955 \\
18.5531 \\
7.51113
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
7.31855 & 3.77574 & 1.67164 \\
3.77574 & 2.32129 & 0.87691 \\
1.67164 & 0.87691 & 0.40455
\end{bmatrix}
\]

These quantities were transformed by the series of exponential, logarithmic, and
linear operations outlined in Section 5.4.2 to produce the following estimate of 
\((\hat{\theta}, \hat{\rho}_{y/x})'\) and its covariance matrix:

\[
\begin{bmatrix}
0.48574 \\
0.21276
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
0.00026 & -0.00004 \\
-0.00004 & 0.00032
\end{bmatrix}
\]

Thus, the fraction of persons with at least one crown out of persons with at least 
one restoration with greater than or equal to three surfaces in a molar, is 
approximately 0.49. The standard error for this quantity is 0.016. Note that the 
variance of \(\theta\) under the assumption of simple random sampling would be 0.00018 
and so the design effect for variance estimation of \(\theta\) is \(0.00026/0.000018 = 14.4\). 
If the 358 practices were of equal size, then each would have had about 
13674/358 = 38.2 relevant patients, and so the design effect for variance 
estimation of \(\theta\) would be expected to be about \(1 + (38.2 - 1)(0.21276)\) = 8.9. 
The larger design effect that actually applies is due to the heterogeneity of cluster 
sizes. The large value for \(\hat{\rho}_{y/x}'\) 0.21, and its small standard error of 0.018 
demonstrate that there is a considerable degree of intracluster correlation for this 
outcome.
5.6.2 Example 2

For this example, \( \bar{\rho}_{y/x} \) was estimated for each region of the country (Northeast, South, Central, West) separately, and the variation of the respective estimates among the four regions was explored using weighted regression. (It is noted that estimates of \( \bar{\rho}_{y/x} \) for each region do not necessarily have to be modeled separately from those of \( \theta \). A model which considers \( \theta \) and \( \bar{\rho}_{y/x} \) jointly for each region will be presented in Section 5.6.3.)

Table 5.1 contains estimates and standard errors for the intracluster correlation parameter \( \bar{\rho}_{y/x} \) for the attribute of at least one crowned molar among persons with at least one \( \geq 3 \) surface restoration in a molar, by region of the country. These estimates were produced by applying the methods of Section 5.52 to the data for each region, separately. The following model for Fisher's Z-transformation of \( \bar{\rho}_{y/x} \)

\[
FZ(\bar{\rho}_{y/x}) = \frac{1}{2} \ln[(1 + \bar{\rho}_{y/x})/(1 + \bar{\rho}_{y/x})]
\]

was fit to the data:

\[
FZ(\bar{\rho}_{y/x}) = \beta_1 \text{Int} + \beta_2 \text{Region1} + \beta_3 \text{Region3}
\]

(5.73)

where Int is the intercept, Region1 is an indicator variable for region 1, and Region3 is an indicator variable for region 3. Since \( \bar{\rho}_{y/x} \) varies between 0 and 1, the Fisher Z-transformation was used to improve approximate normality and approximate variance estimation. The parameter estimates and their standard errors for model (5.73) are given in Table 5.3. The Analysis of Variance (ANOVA) for model (5.73) is given in Table 5.4. The goodness of fit statistic \( \chi^2 = 0.22, \text{df} = 1, \text{p-value} = 0.638 \) shows that model (5.73) adequately represents the data. Significant differences were found between the four regions with respect
to \( FZ(\tilde{\rho}_{y/x}) \) \((\chi^2 = 10.39, \text{df} = 3, \text{p-value} = 0.016)\). The intracluster correlation is lower for regions 1 and 3 than it is for regions 2 and 4 \((\chi^2 = 10.17, \text{df} = 2, \text{p-value} = 0.006)\). The predicted values for \(\tilde{\rho}_{y/x} \) generated from model (5.73) and 95% confidence intervals are given in Table 5.5, by region. The model-predicted values for \(\tilde{\rho}_{y/x} \) are similar to the observed values for \(\tilde{\rho}_{y/x} \) given in Table 5.1, and their standard errors are slightly smaller than the standard errors given in Table 5.1. The standard errors for the model-predicted estimates are lower since they are based on all of the data while the standard errors for the observed values in Table 5.1 are based only on the data for each respective region.

5.6.3 Example 3

In this example, a model for the vector \((\theta', \tilde{\rho}'_{y/x})' = (\theta_1, \tilde{\rho}_{y/x}, \ldots, \theta_4, \tilde{\rho}_{y/x})' \) was fit to the data so that the variation in the elements of \(\theta \) and \(\tilde{\rho}_{y/x} \) among the four regions could be explored simultaneously. Table 5.1 contains estimates of the population fraction for the attribute of at least one crowned molar among persons with at least one \(\geq 3 \) surface restoration in a molar, by region of the country, in addition to those for the intracluster correlation parameter, \(\tilde{\rho}_{y/x} \) for each region. The Taylor series estimate for \(10^5 \) times the estimated variance-covariance matrix for these estimates are given in Table 5.2.
The model fit to these estimates was:

\[
\begin{bmatrix}
\text{logit}(\theta_1) \\
\text{FZ}(\hat{\gamma}_{y/x,1}) \\
\text{logit}(\theta_2) \\
\text{FZ}(\hat{\gamma}_{y/x,2}) \\
\text{logit}(\theta_3) \\
\text{FZ}(\hat{\gamma}_{y/x,3}) \\
\text{logit}(\theta_4) \\
\text{FZ}(\hat{\gamma}_{y/x,4})
\end{bmatrix}
= 
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5
\end{bmatrix}
\] (5.74)

where

\[
\text{logit}(\theta) = \ln \left( \frac{\theta}{1 - \theta} \right),
\]

\[
\text{FZ}(\hat{\gamma}_{y/x}) = \frac{1}{2} \ln \left( \frac{1 + \hat{\gamma}_{y/x}}{1 - \hat{\gamma}_{y/x}} \right),
\]

and where FZ(·) is Fisher's Z-transformation. The parameters \(\beta_1\) and \(\beta_2\) represent a reference value for logit(\(\theta\)) in regions 2, 3, and 4, and a differential region effect for region 1, respectively; and the parameters \(\beta_3\), \(\beta_4\), and \(\beta_5\) represent a reference value for \(\text{FZ}(\hat{\gamma}_{y/x})\) in regions 3 and 4, and differential region effects for regions 1 and 2, respectively. Fisher's Z-transformation was applied to the \(\hat{\gamma}_{y/x}\)'s in order to improve applicability of approximate normality and approximate variance estimation with linear Taylor series methods.

The parameter estimates and their standard errors for model (5.74) are
given in Table 5.6. The ANOVA for this model is given in Table 5.7. The goodness of fit statistic \( \chi^2 = 0.98, \text{ df } = 3, \text{ p-value } = 0.805 \) indicates that model (5.74) adequately fits the data. Model-predicted values for \( \gamma \) and \( \gamma_{y/x} \) are given in Table 5.8. These values are virtually the same as the predicted values from separate models for \( \gamma \) and \( \gamma_{y/x} \). Only slight variance reductions for the parameter estimates were achieved by modeling \( \gamma \) and \( \gamma_{y/x} \) jointly for the parameter estimates since the covariance between estimates for \( \gamma \) and \( \gamma_{y/x} \) was small.
CHAPTER VI
SUMMARY AND DIRECTIONS FOR FURTHER RESEARCH

6.1 Summary

This research focused on the development and application of statistical methods for the analysis of binary data from one- and two-stage cluster samples containing clusters which are large and unequal in size. For cases in which there were multiple opportunities to observe a binary outcome for each study subject, the multivariate binary outcome vector was allowed to be incomplete as well. A dental study of crown utilization motivated the statistical developments contained in this research.

Chapter I presented a review of the relevant literature for this research. It also presented an overview of the contents of Chapters II through V.

Chapter II presented ratio mean methodology for the analysis of (multivariate) binary data from a one-stage cluster sample with clusters selected by simple random sampling with replacement. The ratio mean was reviewed as an estimator for the overall population mean in this setting. Ratio means were then defined for various crossclassifications of cluster-, patient-, and subpatient-level characteristics. Weighted regression methods were used to analyze variation among the (correlated) elements of such vectors. A weighted regression model for the logit of the ratio mean was applied to a sample of data from the dental study
of crown utilization. The results were compared with those for survey data logistic regression, GEE1, and ordinary logistic regression models containing the same predictor variables.

Chapter III extended the ratio mean methods proposed in Chapter II for (multivariate) binary outcomes to multicategory, ordinal and continuous outcomes from a one-stage cluster sample with clusters selected by simple random sampling with replacement. Ratio mean definitions were modified to accommodate these types of outcomes. The ratio means were also defined for crossclassifications of cluster-, patient-, and subpatient-level characteristics. Weighted regression methods were proposed for the analysis of variation among such vector elements.

Chapter IV presented ratio mean methodology for the analysis of (multivariate) binary data from a two-stage cluster sample with clusters selected by simple random sampling with replacement, and elements within clusters selected by simple random sampling without replacement. Modifications of the methods in Chapter II were proposed to allow for the subsampling of clusters. Several self-weighting sampling schemes and their advantages were discussed. Weighted regression methods were used to explore the variation among ratio means for subpopulations defined by the crossclassification of cluster-, patient-, and subpatient-level characteristics. The sample data for the ratio mean regression example in Chapter IV were obtained by subsampling each cluster of the one-stage cluster sample of crown utilization data used in Chapter II. The results of a weighted regression model for the logit of the ratio mean fit to the subsampled data were similar to those obtained from a model containing the same predictor variables fit to the sample of whole clusters in Chapter II.
Chapter V presented methodology for the simultaneous estimation of the overall population mean and a measure of intracluster correlation for univariate binary data from a one-stage cluster sample. Unlike the usual intraclass correlation coefficient, the proposed measure of intracluster correlation takes into account the random sample size resulting from differing numbers of relevant observations among the clusters. Joint estimates of the population fraction for the attribute of interest and the proposed measure intracluster correlation were computed for multiple strata defined by cluster-level characteristics. Weighted regression methods were used to analyze the variation among elements of this vector of joint estimates.

For the dental study used to illustrate the methods, the proportion of crowns among posterior teeth with a restoration on three or more surfaces (excluding wisdom teeth) was nearly fifty percent. Since the source population was comprised of patients with dental insurance, the prevalence of crowns is likely to be higher in this population than in non-insured populations. The proportion of crowns was also estimated for subgroups defined by the crossclassification of the factors: region of the country, age of the patient, and tooth arch. The proportion of crowns for these subpopulations ranged from approximately twenty to sixty percent, and half were between forty and sixty percent. The odds of a crown versus crown substitute was lower in the northeast as compared with the southern, central, and western regions of the country. The odds of a crown were also higher in older versus younger patients and for teeth on the mandibular versus maxillary arch. The age differences with respect to the odds of a crown were greater for maxillary teeth. For molars in particular, an intracluster correlation of 0.21 indicated a moderate tendency for dental
practitioners to use one restoration method or the other for their patients.

While the statistical methods discussed in this work were motivated by the dental study, they were applied using only selected factors of interest to demonstrate the methods. Hence, these results should be viewed cautiously since additional factors may need to be taken into account. Issues concerning the use of claims data for inference should also be addressed. A thorough analysis of these data is, however, beyond the scope of this research.

6.2 Directions for Further Research

The statistical developments proposed in this research for the analysis of binary data from one- and two-stage cluster samples containing clusters which are large and unequal in size, lend themselves to further study and extensions. It would be of interest to identify situations in which the methods proposed in this research work best, and when their use should be avoided. Since the methodology may be inefficient for samples which contain grossly unbalanced clusters, sampling schemes which stratify on cluster size and then combine estimates over the strata could be evaluated. It would be of interest to determine the number of primary sampling units (i.e. clusters) needed to support the asymptotic approximations. While the methods proposed in this research focused on the analysis of one- and two-stage cluster samples with large and unbalanced clusters, strategies to improve smaller sample properties might also be developed. The extensions proposed in Chapter III for multicategory, ordinal, and continuous outcomes could be modified for two-stage cluster samples.

The methods of Chapter V for joint estimation of the overall mean and a measure of intracluster correlation for a univariate binary outcome from a one-
stage cluster sample could be extended in several directions. They may be extended to apply to binary outcomes with multiple opportunities for observation (i.e., to a multivariate outcome), and allow for covariate analysis of subpatient-level characteristics in ways like those demonstrated in Chapter V for practice-level characteristics. They may be extended so that covariate analysis of patient-level characteristics is also possible. These types of extensions would require the estimation of additional means and intracluster correlation parameters (i.e., those which account for correlation between different elements of the outcome vector for the same individual; between the same element for different individuals; and between different elements of the outcome vector for different individuals), and their variance-covariance matrix. The ability to jointly estimate means and intracluster correlation coefficients for subgroups regardless of whether the subgroups are defined by cluster-, patient-, or subpatient-level characteristics is the ultimate goal.
REFERENCES


Table 1.1 Layout for an \((s \times r)\) Contingency Table.

<table>
<thead>
<tr>
<th>Subpopulation</th>
<th>1</th>
<th>2</th>
<th>\ldots</th>
<th>r</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(n_{11})</td>
<td>(n_{12})</td>
<td>\ldots</td>
<td>(n_{1r})</td>
<td>(n_1).</td>
</tr>
<tr>
<td>2</td>
<td>(n_{21})</td>
<td>(n_{22})</td>
<td>\ldots</td>
<td>(n_{2r})</td>
<td>(n_2).</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>s</td>
<td>(n_{s1})</td>
<td>(n_{s2})</td>
<td>\ldots</td>
<td>(n_{sr})</td>
<td>(n_s).</td>
</tr>
</tbody>
</table>
Table 2.1 Observed Ratio Means and Standard Errors Calculated Separately For Each Subpopulation Defined by the Crossclassification of Region, Age, and Tooth Arch.

<table>
<thead>
<tr>
<th>Region</th>
<th>Age</th>
<th>Tooth Arch</th>
<th>R</th>
<th>SRS SE(R)</th>
<th>Taylor Series SE(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.26</td>
<td>0.019</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.19</td>
<td>0.015</td>
<td>0.035</td>
</tr>
<tr>
<td>35-50</td>
<td></td>
<td>mandibular</td>
<td>0.37</td>
<td>0.017</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.31</td>
<td>0.015</td>
<td>0.047</td>
</tr>
<tr>
<td>≥ 51</td>
<td></td>
<td>mandibular</td>
<td>0.43</td>
<td>0.026</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.37</td>
<td>0.024</td>
<td>0.052</td>
</tr>
<tr>
<td>2</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.40</td>
<td>0.012</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.26</td>
<td>0.010</td>
<td>0.023</td>
</tr>
<tr>
<td>35-50</td>
<td></td>
<td>mandibular</td>
<td>0.52</td>
<td>0.010</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.42</td>
<td>0.009</td>
<td>0.025</td>
</tr>
<tr>
<td>≥ 51</td>
<td></td>
<td>mandibular</td>
<td>0.60</td>
<td>0.015</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.52</td>
<td>0.014</td>
<td>0.030</td>
</tr>
<tr>
<td>3</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.38</td>
<td>0.014</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.27</td>
<td>0.013</td>
<td>0.027</td>
</tr>
<tr>
<td>30-50</td>
<td></td>
<td>mandibular</td>
<td>0.55</td>
<td>0.011</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.42</td>
<td>0.010</td>
<td>0.028</td>
</tr>
<tr>
<td>≥ 51</td>
<td></td>
<td>mandibular</td>
<td>0.63</td>
<td>0.016</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.59</td>
<td>0.015</td>
<td>0.034</td>
</tr>
<tr>
<td>4</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.42</td>
<td>0.025</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.28</td>
<td>0.021</td>
<td>0.052</td>
</tr>
<tr>
<td>35-50</td>
<td></td>
<td>mandibular</td>
<td>0.60</td>
<td>0.020</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.45</td>
<td>0.018</td>
<td>0.063</td>
</tr>
<tr>
<td>≥ 51</td>
<td></td>
<td>mandibular</td>
<td>0.62</td>
<td>0.026</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.58</td>
<td>0.024</td>
<td>0.050</td>
</tr>
</tbody>
</table>
Table 2.2 $10^5 \times$ Estimated Covariance Matrix for $\mathbf{R}$

<table>
<thead>
<tr>
<th></th>
<th>131</th>
<th>104</th>
<th>141</th>
<th>127</th>
<th>77</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>131</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>104</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>141</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>127</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>77</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>111</th>
<th>94</th>
<th>176</th>
<th>200</th>
<th>166</th>
<th>274</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>176</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>166</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>274</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>98</td>
<td>56</td>
<td>49</td>
<td>42</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>56</td>
<td>53</td>
<td>41</td>
<td>40</td>
<td>35</td>
<td>43</td>
</tr>
<tr>
<td>0</td>
<td>49</td>
<td>40</td>
<td>49</td>
<td>60</td>
<td>39</td>
<td>49</td>
</tr>
<tr>
<td>0</td>
<td>42</td>
<td>35</td>
<td>43</td>
<td>39</td>
<td>71</td>
<td>55</td>
</tr>
<tr>
<td>0</td>
<td>53</td>
<td>43</td>
<td>47</td>
<td>49</td>
<td>55</td>
<td>90</td>
</tr>
<tr>
<td>0</td>
<td>113</td>
<td>73</td>
<td>38</td>
<td>46</td>
<td>29</td>
<td>17</td>
</tr>
<tr>
<td>0</td>
<td>73</td>
<td>71</td>
<td>33</td>
<td>36</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>0</td>
<td>38</td>
<td>33</td>
<td>56</td>
<td>55</td>
<td>38</td>
<td>45</td>
</tr>
<tr>
<td>0</td>
<td>46</td>
<td>36</td>
<td>55</td>
<td>80</td>
<td>37</td>
<td>61</td>
</tr>
<tr>
<td>0</td>
<td>29</td>
<td>28</td>
<td>36</td>
<td>37</td>
<td>55</td>
<td>27</td>
</tr>
<tr>
<td>0</td>
<td>17</td>
<td>12</td>
<td>45</td>
<td>61</td>
<td>27</td>
<td>115</td>
</tr>
</tbody>
</table>

| 0  | 0   | 0   | 0   | 0   | 0   | 0   |
| 0  | 0   | 0   | 0   | 0   | 0   | 0   |
| 0  | 0   | 0   | 0   | 0   | 0   | 0   |
| 0  | 0   | 0   | 0   | 0   | 0   | 0   |
| 0  | 0   | 0   | 0   | 0   | 0   | 0   |
| 0  | 0   | 0   | 0   | 0   | 0   | 0   |

| 0  | 758 | 394 | 376 | 318 | 415 |
| 0  | 394 | 275 | 242 | 211 | 240 |
| 0  | 376 | 242 | 381 | 345 | 357 |
| 0  | 318 | 211 | 345 | 392 | 292 |
| 0  | 415 | 240 | 357 | 292 | 490 |
| 0  | 275 | 188 | 259 | 247 | 282 | 26 |
Table 2.3  Parameter Estimates and Standard Errors from Ratio Mean, Survey, and Ordinary Logistic Regression Methods¹.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Statistic</th>
<th>Ratio Mean</th>
<th>Survey</th>
<th>Ordinary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Estimate</td>
<td>0.293</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.196</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>Region1</td>
<td>Estimate</td>
<td>-0.783</td>
<td>-0.72</td>
<td>-0.72</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.257</td>
<td>0.30</td>
<td>0.05</td>
</tr>
<tr>
<td>Region2</td>
<td>Estimate</td>
<td>-0.154</td>
<td>-0.16</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.222</td>
<td>0.25</td>
<td>0.04</td>
</tr>
<tr>
<td>Region3</td>
<td>Estimate</td>
<td>-0.042</td>
<td>-0.10</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.220</td>
<td>0.25</td>
<td>0.04</td>
</tr>
<tr>
<td>Age ≤ 34</td>
<td>Estimate</td>
<td>-1.183</td>
<td>-1.19</td>
<td>-1.19</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.070</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>Age 35-50</td>
<td>Estimate</td>
<td>-0.490</td>
<td>-0.49</td>
<td>-0.49</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.053</td>
<td>0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Mandibular</td>
<td>Estimate</td>
<td>0.254</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.061</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Age ≤ 34 × Man-</td>
<td>Estimat</td>
<td>0.338</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>dibular</td>
<td>SE</td>
<td>0.075</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>Age 35-50 × Man-</td>
<td>Estimat</td>
<td>0.201</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>dibular</td>
<td>SE</td>
<td>0.063</td>
<td>0.07</td>
<td>0.06</td>
</tr>
</tbody>
</table>

¹ The model was not fit using GEE1 since the maximum cluster size exceeded that allowed by the SAS macro (Karim and Zeger, 1988) used to implement GEE1.
Table 2.4 Analysis of Variance for Ratio Mean Regression Model

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>$\chi^2$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>2.23</td>
<td>0.135</td>
</tr>
<tr>
<td>Region</td>
<td>3</td>
<td>18.17</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>2</td>
<td>284</td>
<td>0.000</td>
</tr>
<tr>
<td>Tooth Arch</td>
<td>1</td>
<td>17.56</td>
<td>0.000</td>
</tr>
<tr>
<td>Age $\times$ Tooth Arch</td>
<td>2</td>
<td>20.36</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>15</td>
<td>14.62</td>
<td>0.479</td>
</tr>
</tbody>
</table>
Table 2.5 Predicted Ratio Means and Standard Errors for Each Subpopulation Defined by the Crossclassification of Region, Age, and Tooth Arch, Generated from the Ratio Mean Regression Model.

<table>
<thead>
<tr>
<th>Region</th>
<th>Age</th>
<th>Tooth Arch</th>
<th>R</th>
<th>SE(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.25</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.16</td>
<td>0.021</td>
</tr>
<tr>
<td>35-50</td>
<td></td>
<td>mandibular</td>
<td>0.37</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.27</td>
<td>0.033</td>
</tr>
<tr>
<td>≥ 51</td>
<td></td>
<td>mandibular</td>
<td>0.44</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.38</td>
<td>0.039</td>
</tr>
<tr>
<td>2</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.39</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.26</td>
<td>0.020</td>
</tr>
<tr>
<td>35-50</td>
<td></td>
<td>mandibular</td>
<td>0.53</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.41</td>
<td>0.023</td>
</tr>
<tr>
<td>≥ 51</td>
<td></td>
<td>mandibular</td>
<td>0.60</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.53</td>
<td>0.026</td>
</tr>
<tr>
<td>3</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.42</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.28</td>
<td>0.020</td>
</tr>
<tr>
<td>30-50</td>
<td></td>
<td>mandibular</td>
<td>0.55</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.44</td>
<td>0.024</td>
</tr>
<tr>
<td>≥ 51</td>
<td></td>
<td>mandibular</td>
<td>0.62</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.56</td>
<td>0.024</td>
</tr>
<tr>
<td>4</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.43</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.29</td>
<td>0.042</td>
</tr>
<tr>
<td>35-50</td>
<td></td>
<td>mandibular</td>
<td>0.56</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.45</td>
<td>0.050</td>
</tr>
<tr>
<td>≥ 51</td>
<td></td>
<td>mandibular</td>
<td>0.63</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.57</td>
<td>0.048</td>
</tr>
</tbody>
</table>
Table 2.6 Adjusted Odds Ratios and 95% Confidence Intervals for Selected Factor-Level Comparisons, Generated from the Ratio Mean Regression Model.

<table>
<thead>
<tr>
<th>Subpopulation</th>
<th>Comparison</th>
<th>Adjusted Odds Ratio</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandibular</td>
<td>35-50 vs ≤ 34</td>
<td>1.7</td>
<td>(1.6 - 2.0)</td>
</tr>
<tr>
<td></td>
<td>≥ 51 vs ≤ 34</td>
<td>2.3</td>
<td>(2.0 - 2.7)</td>
</tr>
<tr>
<td>Maxillary</td>
<td>35-50 vs ≤ 34</td>
<td>2.0</td>
<td>(1.8 - 2.2)</td>
</tr>
<tr>
<td></td>
<td>≥ 51 vs ≤ 34</td>
<td>3.3</td>
<td>(2.8 - 3.7)</td>
</tr>
<tr>
<td>overall</td>
<td>Region 2 vs 1</td>
<td>1.9</td>
<td>(1.3 - 2.7)</td>
</tr>
<tr>
<td></td>
<td>Region 3 vs 1</td>
<td>2.1</td>
<td>(1.5 - 3.0)</td>
</tr>
<tr>
<td></td>
<td>Region 4 vs 1</td>
<td>2.2</td>
<td>(1.3 - 3.6)</td>
</tr>
</tbody>
</table>
Table 4.1 Observed Ratio Means and Standard Errors Calculated Separately For Each Subpopulation Defined by the Crossclassification of Region, Age, and Tooth Arch.

<table>
<thead>
<tr>
<th>Region</th>
<th>Age</th>
<th>Tooth Arch</th>
<th>R</th>
<th>SRS</th>
<th>SE(R)</th>
<th>Taylor Series SE(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.28</td>
<td>0.050</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.12</td>
<td>0.035</td>
<td>0.037</td>
<td></td>
</tr>
<tr>
<td>35-50</td>
<td>mandibular</td>
<td>0.47</td>
<td>0.051</td>
<td>0.064</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.29</td>
<td>0.039</td>
<td>0.049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 51</td>
<td>mandibular</td>
<td>0.47</td>
<td>0.078</td>
<td>0.071</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.45</td>
<td>0.070</td>
<td>0.072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.43</td>
<td>0.035</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.26</td>
<td>0.028</td>
<td>0.036</td>
<td></td>
</tr>
<tr>
<td>35-50</td>
<td>mandibular</td>
<td>0.54</td>
<td>0.027</td>
<td>0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.40</td>
<td>0.024</td>
<td>0.033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 51</td>
<td>mandibular</td>
<td>0.63</td>
<td>0.038</td>
<td>0.039</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.54</td>
<td>0.038</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.37</td>
<td>0.039</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.27</td>
<td>0.034</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>30-50</td>
<td>mandibular</td>
<td>0.55</td>
<td>0.028</td>
<td>0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.40</td>
<td>0.028</td>
<td>0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 51</td>
<td>mandibular</td>
<td>0.58</td>
<td>0.046</td>
<td>0.052</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.58</td>
<td>0.042</td>
<td>0.039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.48</td>
<td>0.074</td>
<td>0.129</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.37</td>
<td>0.057</td>
<td>0.103</td>
<td></td>
</tr>
<tr>
<td>35-50</td>
<td>mandibular</td>
<td>0.55</td>
<td>0.055</td>
<td>0.081</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.45</td>
<td>0.049</td>
<td>0.069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 51</td>
<td>mandibular</td>
<td>0.62</td>
<td>0.067</td>
<td>0.080</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.55</td>
<td>0.061</td>
<td>0.065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>412</td>
<td>70</td>
<td>202</td>
<td>128</td>
<td>123</td>
<td>62</td>
<td>0</td>
</tr>
<tr>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>---</td>
</tr>
<tr>
<td>70</td>
<td>137</td>
<td>75</td>
<td>80</td>
<td>16</td>
<td>62</td>
<td>0</td>
</tr>
<tr>
<td>202</td>
<td>75</td>
<td>410</td>
<td>182</td>
<td>23</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>128</td>
<td>80</td>
<td>182</td>
<td>235</td>
<td>39</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>123</td>
<td>16</td>
<td>23</td>
<td>39</td>
<td>501</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>62</td>
<td>62</td>
<td>109</td>
<td>175</td>
<td>24</td>
<td>521</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>228</td>
<td>74</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>74</td>
<td>129</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>67</td>
<td>51</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>66</td>
<td>50</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td>31</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>188</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>78</td>
<td>178</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>28</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>63</td>
<td>61</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>-18</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.2  $10^6 \times$ Estimated Covariance Matrix for $R_1$
Table 4.3 Parameter Estimates and Standard Errors from Ratio Mean, Survey, and Ordinary Logistic Regression Models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Statistic</th>
<th>Ratio Mean</th>
<th>Survey</th>
<th>Ordinary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Estimate</td>
<td>0.302</td>
<td>0.38</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.215</td>
<td>0.26</td>
<td>0.13</td>
</tr>
<tr>
<td>Region1</td>
<td>Estimate</td>
<td>-0.555</td>
<td>-0.67</td>
<td>-0.67</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.271</td>
<td>0.32</td>
<td>0.14</td>
</tr>
<tr>
<td>Region2</td>
<td>Estimate</td>
<td>-0.042</td>
<td>-0.15</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.241</td>
<td>0.29</td>
<td>0.11</td>
</tr>
<tr>
<td>Region3</td>
<td>Estimate</td>
<td>-0.087</td>
<td>-0.18</td>
<td>-0.18</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.236</td>
<td>0.29</td>
<td>0.12</td>
</tr>
<tr>
<td>Age ≤ 34</td>
<td>Estimate</td>
<td>-1.281</td>
<td>-1.25</td>
<td>-1.24</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.144</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td>Age 35-50</td>
<td>Estimate</td>
<td>-0.642</td>
<td>-0.61</td>
<td>-0.62</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.112</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>Mandibular</td>
<td>Estimate</td>
<td>0.202</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.132</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Age ≤ 34 × Man-</td>
<td>Estimate</td>
<td>0.413</td>
<td>0.45</td>
<td>0.44</td>
</tr>
<tr>
<td>dibular</td>
<td>SE</td>
<td>0.175</td>
<td>0.18</td>
<td>0.20</td>
</tr>
<tr>
<td>Age 35-50 × Man-</td>
<td>Estimate</td>
<td>0.376</td>
<td>0.39</td>
<td>0.38</td>
</tr>
<tr>
<td>dibular</td>
<td>SE</td>
<td>0.161</td>
<td>0.17</td>
<td>0.17</td>
</tr>
</tbody>
</table>

† The model was not fit using GEE1 since the maximum cluster size exceeded that allowed by the SAS macro (Karim and Zeger, 1988) used to implement GEE1.
Table 4.4 Analysis of Variance for Ratio Mean Regression Model

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>$\chi^2$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.97</td>
<td>0.161</td>
</tr>
<tr>
<td>Region</td>
<td>3</td>
<td>7.36</td>
<td>0.061</td>
</tr>
<tr>
<td>Age</td>
<td>2</td>
<td>79.86</td>
<td>0.000</td>
</tr>
<tr>
<td>Tooth Arch</td>
<td>1</td>
<td>2.33</td>
<td>0.127</td>
</tr>
<tr>
<td>Age x Tooth Arch</td>
<td>2</td>
<td>6.98</td>
<td>0.030</td>
</tr>
<tr>
<td>Residual</td>
<td>15</td>
<td>8.04</td>
<td>0.922</td>
</tr>
<tr>
<td>Region</td>
<td>Age</td>
<td>Tooth Arch</td>
<td>R</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maxillary</td>
<td>0.18</td>
</tr>
<tr>
<td>35-50</td>
<td>mandibular</td>
<td>0.42</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.29</td>
<td>0.038</td>
</tr>
<tr>
<td>≥ 51</td>
<td>mandibular</td>
<td>0.49</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.44</td>
<td>0.046</td>
</tr>
<tr>
<td>2</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.26</td>
<td>0.029</td>
</tr>
<tr>
<td>35-50</td>
<td>mandibular</td>
<td>0.55</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.41</td>
<td>0.030</td>
</tr>
<tr>
<td>≥ 51</td>
<td>mandibular</td>
<td>0.61</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.56</td>
<td>0.036</td>
</tr>
<tr>
<td>3</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.26</td>
<td>0.027</td>
</tr>
<tr>
<td>30-50</td>
<td>mandibular</td>
<td>0.54</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.39</td>
<td>0.030</td>
</tr>
<tr>
<td>≥ 51</td>
<td>mandibular</td>
<td>0.60</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.55</td>
<td>0.032</td>
</tr>
<tr>
<td>4</td>
<td>≤ 34</td>
<td>mandibular</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.27</td>
<td>0.047</td>
</tr>
<tr>
<td>35-50</td>
<td>mandibular</td>
<td>0.56</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.42</td>
<td>0.053</td>
</tr>
<tr>
<td>≥ 51</td>
<td>mandibular</td>
<td>0.62</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>maxillary</td>
<td>0.57</td>
<td>0.053</td>
</tr>
<tr>
<td>Subpopulation</td>
<td>Comparison</td>
<td>Adjusted Odds Ratio</td>
<td>95% Confidence Interval</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Mandibular</td>
<td>35-50 vs ≤34</td>
<td>1.8</td>
<td>(1.4 - 2.3)</td>
</tr>
<tr>
<td></td>
<td>≥ 51 vs ≤34</td>
<td>2.4</td>
<td>(1.8 - 3.1)</td>
</tr>
<tr>
<td>Maxillary</td>
<td>35-50 vs ≤34</td>
<td>1.9</td>
<td>(1.5 - 2.4)</td>
</tr>
<tr>
<td></td>
<td>≥ 51 vs ≤34</td>
<td>3.6</td>
<td>(2.7 - 4.8)</td>
</tr>
<tr>
<td>overall</td>
<td>Region 2 vs 1</td>
<td>1.7</td>
<td>(1.1 - 2.5)</td>
</tr>
<tr>
<td></td>
<td>Region 3 vs 1</td>
<td>1.6</td>
<td>(1.1 - 2.4)</td>
</tr>
<tr>
<td></td>
<td>Region 4 vs 1</td>
<td>1.7</td>
<td>(1.0 - 3.0)</td>
</tr>
</tbody>
</table>
Table 5.1 Estimates of $\theta$ and $\hat{\rho}_{y/x}$ for each Region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\theta$</td>
<td>0.35027</td>
<td>0.038341</td>
</tr>
<tr>
<td></td>
<td>$\hat{\rho}_{y/x}$</td>
<td>0.19044</td>
<td>0.037283</td>
</tr>
<tr>
<td>2</td>
<td>$\theta$</td>
<td>0.49228</td>
<td>0.023664</td>
</tr>
<tr>
<td></td>
<td>$\hat{\rho}_{y/x}$</td>
<td>0.24447</td>
<td>0.023022</td>
</tr>
<tr>
<td>3</td>
<td>$\theta$</td>
<td>0.51971</td>
<td>0.024083</td>
</tr>
<tr>
<td></td>
<td>$\hat{\rho}_{y/x}$</td>
<td>0.13290</td>
<td>0.030496</td>
</tr>
<tr>
<td>4</td>
<td>$\theta$</td>
<td>0.52314</td>
<td>0.057009</td>
</tr>
<tr>
<td></td>
<td>$\hat{\rho}_{y/x}$</td>
<td>0.27059</td>
<td>0.050299</td>
</tr>
<tr>
<td></td>
<td>147</td>
<td>82</td>
<td>0</td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>----</td>
<td>---</td>
</tr>
<tr>
<td>82</td>
<td>139</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>56</td>
<td>-17</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>-17</td>
<td>53</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5.3 Parameter Estimates and Standard Errors from Weighted Regression Model for Fisher Z-Transformation of Intracluster Correlation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.2543</td>
<td>0.0223</td>
</tr>
<tr>
<td>Region1</td>
<td>-0.0615</td>
<td>0.0447</td>
</tr>
<tr>
<td>Region3</td>
<td>-0.1206</td>
<td>0.0382</td>
</tr>
</tbody>
</table>
Table 5.4 Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>$\chi^2$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>129.77</td>
<td>0.000</td>
</tr>
<tr>
<td>Region</td>
<td>2</td>
<td>10.17</td>
<td>0.006</td>
</tr>
<tr>
<td>Residual</td>
<td>1</td>
<td>0.22</td>
<td>0.638</td>
</tr>
</tbody>
</table>
Table 5.5 Predicted Estimates and Standard Errors for $\hat{\rho}_{y/x}$, by Region.

<table>
<thead>
<tr>
<th>Region</th>
<th>$\hat{\rho}_{y/x}$</th>
<th>SE($\hat{\rho}_{y/x}$)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.19</td>
<td>0.037</td>
<td>(0.12 - 0.26)</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.021</td>
<td>(0.21 - 0.29)</td>
</tr>
<tr>
<td>3</td>
<td>0.13</td>
<td>0.030</td>
<td>(0.07 - 0.19)</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>0.021</td>
<td>(0.21 - 0.29)</td>
</tr>
</tbody>
</table>

† Computed by applying the inverse of Fisher's Z-transformation to the limits of a 95% confidence interval for the predicted values for each region from the weighted regression model.
Table 5.6 Parameter Estimates and Standard Errors from the Weighted Regression Model for Logit(θ) and Fisher Z-Transformation of $\beta_{y/x}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>-0.6178</td>
<td>0.1685</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.6410</td>
<td>0.1802</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.2498</td>
<td>0.0215</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>-0.0570</td>
<td>0.0443</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>-0.1183</td>
<td>0.0379</td>
</tr>
<tr>
<td>Source</td>
<td>DF</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>-----------------</td>
<td>----</td>
<td>----------</td>
</tr>
<tr>
<td>Intercept1</td>
<td>1</td>
<td>13.45</td>
</tr>
<tr>
<td>Region ($\theta$)</td>
<td>1</td>
<td>12.66</td>
</tr>
<tr>
<td>Intercept2</td>
<td>1</td>
<td>134.69</td>
</tr>
<tr>
<td>Region ($\bar{y}/x$)</td>
<td>2</td>
<td>9.94</td>
</tr>
<tr>
<td>Residual</td>
<td>3</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Table 5.8  Predicted Estimates and Standard Errors for $\theta$ and $\tilde{\theta}_{y/x}$ by Region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\theta$</td>
<td>0.35027</td>
<td>0.038339</td>
</tr>
<tr>
<td></td>
<td>$\tilde{\theta}_{y/x}$</td>
<td>0.19044</td>
<td>0.037325</td>
</tr>
<tr>
<td>2</td>
<td>$\theta$</td>
<td>0.50579</td>
<td>0.015961</td>
</tr>
<tr>
<td></td>
<td>$\tilde{\theta}_{y/x}$</td>
<td>0.24469</td>
<td>0.020232</td>
</tr>
<tr>
<td>3</td>
<td>$\theta$</td>
<td>0.50579</td>
<td>0.015961</td>
</tr>
<tr>
<td></td>
<td>$\tilde{\theta}_{y/x}$</td>
<td>0.13074</td>
<td>0.030325</td>
</tr>
<tr>
<td>4</td>
<td>$\theta$</td>
<td>0.50579</td>
<td>0.015961</td>
</tr>
<tr>
<td></td>
<td>$\tilde{\theta}_{y/x}$</td>
<td>0.24469</td>
<td>0.020232</td>
</tr>
</tbody>
</table>