LOGISTIC REGRESSION ANALYSIS FOR COMPLEX SAMPLE DATA

by

Lloyd E. Chambless

and

Kerrie E. Boyle

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series Number 1440

June 1983

LOGISTIC REGRESSION ANALYSIS FOR COMPLEX SAMPLE DATA

Lloyd E. Chambless
Kerrie E. Boyle

Lipid Research Clinics Program
Department of Biostatistics
University of North Carolina
Chapel Hill, NC 27514

Key Words: Asymptotic normality, Estimation for subdomains, Horvitz-Thompson estimators, Maximum likelihood estimation, Modeling for finite populations, Stratified sampling, Taylor series expansion, Variance estimation

This research was supported by NHLBI Contract No. M01-HV-1-223-L.

SUMMARY

The estimation of parameters in linear and logistic regression based on a stratified random sample from a finite population is investigated. The notion of an infinite superpopulation facilitates the development of asymptotic normal distribution theory. Procedures for parameter estimation and statistical inference on a subdomain of the finite population are also developed. We illustrate these modeling techniques, showing the effects on inference of ignoring the sampling design in estimation of model coefficients. Finally, we extend our methods to logistic regression on data gathered in a two-stage stratified sample design.

1. INTRODUCTION

In the situation where data arise from a sampling scheme more complex than simple random sampling, we would like to account for the sampling design in our analysis of the data. While this is often done when the analysis entails only estimates of population means or totals, the situation is otherwise for more complex analyses. However, interest in accounting for the sampling design in analysis has increased in recent years; see Fuller (1975), Sárndal (1978), O'Brien (1981), Binder (1981), and DuMouchel and Duncan (1981). In this study we consider primarily logistic regression for stratified random sample data, though we also include an extension to a two-stage stratified sampling situation.

That it is desirable to account for the sampling design in estimating population means seems generally accepted. In a linear model setting, we would be "modeling" $Y = X\underset{\sim}{\beta}$, where Y is the variable of interest, and X is identically one. The controversy begins when more variables are included in X. Suppose X contains only an intercept term and an indicator variable, say for smoking versus non-smoking. If the sampling of the U.S. population were done by race, with equal numbers of whites and blacks, little argument would be raised in accounting for the sampling scheme, via weighted averages, when estimating the population means of Y for each level of the smoking variable. Suppose we want to adjust for age differences, by adding age as an independent variable in our budding model. At this point there are those who would argue for a return to the ordinary least squares (OLS) fit of the homoscedastic model to "test" the difference between smokers and non-smokers. Clearly if the mean of Y differs between the races, this method will lead to "adjusted means" by smoking class that are "biased" toward the mean for blacks, who are oversampled. Note that such a "bias" may or may not be present in the estimate of the difference in expected Y between smokers and non-smokers.

DuMouchel and Duncan (1981) present several models for consideration. One is the OLS model just discussed. Another, the "mixture model", is essentially OLS within each sampling stratum, then a weighted average thereof; i.e., a $\underset{\sim}{\beta}_j$ is estimated for each stratum, and the parameter of interest is the $\bar{\underset{\sim}{\beta}} = \Sigma\, w_j \underset{\sim}{\beta}_j$, where the weights w_j are proportional to the inverse sampling proportion. But $\bar{\underset{\sim}{\beta}}$ is problematic when there are large numbers of small strata, since stratum-specific variable estimates become unstable. In addition, this "mean of parameters" parameter may not be the desired one, i.e., there are instances where an "overall population" parameter is the object of interest. Another model, of the type considered in this paper, defines $\underset{\sim}{\beta}$ as giving the best linear predictor $X\underset{\sim}{\beta}$ for Y in the sense of minimizing the expected squared error of prediction. Or in logistic regression, our main focus, $\underset{\sim}{\beta}$ will be defined to give the best linear predictor $X\underset{\sim}{\beta}$ of logit {Pr(Y=1)} in the sense of maximizing the expected "likelihood" for the finite population. Of course if the relationship of interest varies too much by sampling strata, an "overall" $\underset{\sim}{\beta}$ may be inappropriate. How much variation is too much for an overall measure to be useful? How feasible is it to show stratum-specific results? The answers to such questions are not fixed, but depend on the particular problem at hand. In a broader context, whether to include some of the stratum-defining variables in the model is really just part of a more general model building procedure.

In addition to concern about the validity of estimation which ignores the sampling design, one must also be concerned with variance estimates. Even if assuming away the sampling design has little effect on the estimates of the parameters of interest, there may be a substantial effect on the estimated variances of the estimators (O'Brien, 1981).

The analytic model of interest in this paper is logistic regression; however, the basic techniques are applicable to other transformations of binary data. After introducing some notation, we define the parameter of interest and develop large sample distribution theory, including hypothesis testing procedures. In applications we have applied an iteratively reweighted least squares algorithm to maximum likelihood estimation of the model's coefficients, and briefly discuss this. We also investigate logistic regression modeling on a subdomain of the population. We then utilize our modeling techniques in an example. Parameter estimates and statistical inferences obtained from a logistic regression analysis that accounts for the stratified random sampling scheme are compared to those from an analogous model that assumes a simple random sample. Finally, we present an extension of our techniques for two-stage stratified sampling.

2. NOTATION

We consider a population

$$\mathcal{Q} = \bigcup_{i=1}^{K} \mathcal{Q}_i,$$

a disjoint union of K strata. We also consider subsets $\mathcal{S}_i \subseteq \mathcal{R}_i \subseteq \mathcal{Q}_i$ for each i, where

$$n_i = \text{size of } \mathcal{S}_i, \qquad n = \sum_{i=1}^{K} n_i, \qquad \mathcal{S} = \bigcup_{i=1}^{K} \mathcal{S}_i,$$

$$N_i = \text{size of } \mathcal{R}_i, \qquad N = \sum_{i=1}^{K} N_i, \qquad \mathcal{R} = \bigcup_{i=1}^{K} \mathcal{R}_i.$$

If Y is a variable of interest, we will use Y_{ij} to designate the j^{th} observation on Y in $\mathcal{R}_i$ or $\mathcal{S}_i$. We use "tilded" capital script characters to designate the whole vector of observations on Y in $\mathcal{R}$: $\underset{\sim}{\mathcal{Y}}' = (Y_{11} \ldots Y_{1N_1} \ldots Y_{KN_K})$. For the vector of observations on Y in $\mathcal{S}$ we used "tilded" caps: $\underset{\sim}{Y}' = (Y_{11} \ldots Y_{1n_1} \ldots Y_{Kn_K})$.

We will interpret $\mathcal{R}$ as the finite population of interest, considered as a random sample of the infinite superpopulation $\mathcal{Q}$. $\mathcal{S}$ will be taken as a stratified random (non-replacement) sample from $\mathcal{R}$. Our interest is in statistical inference from $\mathcal{S}$ to $\mathcal{R}$.

3. DEFINING THE PARAMETER OF INTEREST

Suppose we had variables Y and X on $\mathcal{Q}$, and some hypothesized "relation" $F(Y) = G(X;\underset{\sim}{\beta})$ with G depending on parameter $\underset{\sim}{\beta}$. The usual problem is to estimate $\underset{\sim}{\beta}$, an rx1 vector, from sample data. Examples of such relations are linear regression models $Y = X\underset{\sim}{\beta}$, or, more generally, nonlinear regression models $Y = G(X;\underset{\sim}{\beta})$. Another example is the case where F is a density or probability function we want to estimate. This paper deals primarily with this latter example, though we will briefly mention the "regression" example for motivational purposes.

3.1 Linear Regression

Consider the model $E(Y) = X\underset{\sim}{\beta}$ (or $Y = X\underset{\sim}{\beta} + \varepsilon$) on R. Using R we would estimate $\underset{\sim}{\beta}$, using the least squares criterion, by

$$\underset{\sim}{B} = (X'X)^{-1} X'\underset{\sim}{Y} \qquad (1)$$

where X_{ij} is an r dimensional row vector (X_{ijt}) of "independent" variables. Actually our main interest in this study does not lie with $\underset{\sim}{\beta}$ and the "superpopulation" Q. Rather, we are concerned with $\underset{\sim}{B}$ and the finite population R, and so take (1) as the definition of a parameter on the finite population R. Not actually having R, we estimate $\underset{\sim}{B}$ by $\hat{\underset{\sim}{B}} = (X'WX')^{-1} X'W\underset{\sim}{Y}$ where $\underset{\sim}{Y}$ and X are defined on the sample S similarly to $\underset{\sim}{Y}$ and X and

$$W = \begin{pmatrix} w_{11} & & 0 \\ & \ddots & \\ 0 & & w_{Km_K} \end{pmatrix},$$

for $w_{ij} = f_i^{-1}$, where f_i is the sampling proportion in the i^{th} stratum. Then $X'W\underset{\sim}{Y}$ and $X'WX$ are the (unbiased) Horvitz-Thompson estimators (Cochran, 1977) of $X'\underset{\sim}{Y}$ and $X'X$, respectively.

3.2 Nonlinear Regression

For a least squares fit of a nonlinear regression model $E(Y) = G(X_i\underset{\sim}{\beta})$, the normal equations on R are

$$Z'\underset{\sim}{G} = Z'\underset{\sim}{Y}$$

where by Z we mean the Nxr matrix with entries

$$\frac{\partial G_{ij}}{\partial \beta_t} X_{ij}.$$

We can then "estimate" $Z'\underset{\sim}{G}$ and $Z'\underset{\sim}{Y}$ by Horvitz-Thompson estimators $Z'W\underset{\sim}{G}$ and $Z'W\underset{\sim}{Y}$, respectively, with W as before and Z, $\underset{\sim}{G}$, and $\underset{\sim}{Y}$ the stratified random sample relatives of Z, $\underset{\sim}{G}$, and $\underset{\sim}{Y}$. To estimate $\underset{\sim}{B}$, we then equate $Z'W\underset{\sim}{G}$ and $Z'W\underset{\sim}{Y}$ and solve for $\hat{\underset{\sim}{B}}$, generally by iterative methods.

3.3 Density

Suppose now that we hypothesize $E\{F(Y)\} = X\underset{\sim}{\beta}$. If F is invertible, we could rewrite this, heuristically, $E(Y) = F^{-1}(X\underset{\sim}{\beta})$ and apply the methods of the previous paragraph. However, if Y is a zero-one random variable, so that $E(Y) = pr(Y=1) = F^{-1}(X\underset{\sim}{\beta})$, instead of applying nonlinear least squares, we take a maximum likelihood approach.

The likelihood of our sample R is:

$$\begin{aligned} L(\underset{\sim}{\beta}) &= \prod_{i=1}^{K} \prod_{j=1}^{N_i} F^{-1}(X_{ij}\underset{\sim}{\beta})^{Y_{ij}} \{1-F^{-1}(X_{ij}\underset{\sim}{\beta})\}^{1-Y_{ij}} \\ &= \prod_{i=1}^{K} \prod_{j=1}^{N_i} P_{ij}(\underset{\sim}{\beta})^{Y_{ij}} \{1-P_{ij}(\underset{\sim}{\beta})\}^{1-Y_{ij}}, \end{aligned}$$

where

$$P_{ij} = P_{ij}(\underset{\sim}{\beta}) = pr(Y_{ij}=1|X_{ij}) = F^{-1}(X_{ij}\underset{\sim}{\beta}). \qquad (2)$$

The resulting likelihood equations are

$$0 = \underset{\sim}{U} = \frac{\partial \log L}{\partial \underset{\sim}{\beta}} = Z'(\underset{\sim}{Y}-\underset{\sim}{P}), \qquad (3)$$

where by Z we mean the Nxr matrix with entries

$$\frac{\partial P_{ij}}{\partial \beta_t} \frac{1}{P_{ij}(1-P_{ij})},$$

X, $\underset{\sim}{Y}$, $\underset{\sim}{P}$ are on the finite population and we assume there is a solution $\underset{\sim}{\beta}$. This we take as defining the parameter of interest on the finite population R. Not really having R but only the stratified random sample $S \subseteq R$, we utilize S to compute the "Horvitz-Thompson" estimators $\widehat{Z'\underset{\sim}{Y}}$ and $\widehat{Z'\underset{\sim}{P}}$ where

$$\widehat{Z'\underset{\sim}{Y}} = Z'W\underset{\sim}{Y} = \sum_{i=1}^{K} f_i^{-1} \sum_{j=1}^{n_i} Z'_{ij} Y_{ij},$$

and similarly for $\widehat{Z'\underset{\sim}{P}}$. Then the equations

$$0 = \hat{\underset{\sim}{U}} = Z'W\underset{\sim}{Y} - Z'W\underset{\sim}{P} \tag{4}$$

are solved, iteratively, for $\hat{\underset{\sim}{\beta}}$. Note that these "likelihood" equations can be written $\hat{\underset{\sim}{U}} = \frac{\partial \log \hat{L}}{\partial \underset{\sim}{\beta}} = 0$, where

$$\hat{L} = \prod_{i=1}^{K} \left[\prod_{j=1}^{n_i} F^{-1}(X_{ij}\underset{\sim}{\beta})^{Y_{ij}} \{1-F^{-1}(X_{ij}\underset{\sim}{\beta})\}^{1-Y_{ij}} \right]^{f_i^{-1}}.$$

For the rest of the paper we will restrict our attention to logistic regression, where

$$F^{-1}(X_{ij}\underset{\sim}{\beta}) = \exp(X_{ij}\underset{\sim}{\beta})/\{1 + \exp(X_{ij}\underset{\sim}{\beta})\}.$$

In this case the likelihood equations are

$$X'(\underset{\sim}{Y}-\underset{\sim}{P}) = 0, \tag{3'}$$

or using the stratified random sample,

$$X'W(\underset{\sim}{Y}-\underset{\sim}{P}) = \underset{\sim}{0}. \tag{4'}$$

Before proceeding to derive the large sample distribution of the estimators of $\underset{\sim}{\beta}$, we point out that (4') reduces to the "right" estimates in the simple case where X is a zero-one variable. Specifically, for the i^{th} stratum consider the 2x2 table

		Y: 1	Y: 0
X	1	a_i	b_i
	0	c_i	d_i

where $a_i + b_i + c_i + d_i = n_i$. The model is

$$\mathrm{pr}(Y=1|X) = \frac{\exp(\alpha + \beta X)}{1 + \exp(\alpha + \beta X)} ,$$

and the maximum likelihood estimates for α and β are

$$\exp(\hat{\alpha}) = \frac{\sum_{i=1}^{K} f_i^{-1} c_i}{\sum_{i=1}^{K} f_i^{-1} d_i} ,$$

and

$$\exp(\hat{\beta}) = \frac{\left(\sum_{i=1}^{K} f_i^{-1} a_i\right)\left(\sum_{i=1}^{K} f_i^{-1} d_i\right)}{\left(\sum_{i=1}^{K} f_i^{-1} b_i\right)\left(\sum_{i=1}^{K} f_i^{-1} c_i\right)} .$$

If we then estimate $\mathrm{pr}(Y=1|X=0) = P_o$ by

$$\hat{P}_o = \frac{e^{\hat{\alpha}}}{1 + e^{\hat{\alpha}}}$$

we can show that

$$\hat{P}_0 = \frac{\sum_{i=1}^{K} \sum_{\substack{j=1,\\ X=0}}^{n_i} f_i^{-1} Y_{ij}}{\sum_{i=1}^{K} \sum_{\substack{j=1,\\ X=0}}^{n_i} f_i^{-1}} .$$

This is just the estimate we would get from a weighted mean approach in which the sampling proportions on the $X = 0$ subpopulation are assumed to be the same as those on the full population.

4. LARGE-SAMPLE THEORY

As the next step in our investigation of the parameter of interest $\underset{\sim}{B}$, we will find the asymptotic distribution of $\hat{\underset{\sim}{B}} - \underset{\sim}{B}$. To obtain our asymptotic results we use an approach similar to Fuller (1975): Take a sequence $\{R_m\}$ of finite populations, with strictly increasing size N_m. Each R_m is considered as a random sample from the infinite superpopulation Q. Let $R_{mi} = R_m \cap Q_i$. Write N_{mi} for the size of R_{mi}. For each m let a stratified random non-replacement sample S_m of size n_m be selected from R_m, where S_{mi} is the sample from the i^{th} stratum. Write n_{mi} for the size of S_{mi} so that $n_m = \Sigma_i n_{mi}$. Further, write $n_{mi}/N_{mi} = f_{mi}$ and let $\lim f_{mi} = f_i$, for $0 < f_i < 1$. Also, let $\lim N_{mi}/N_m = \pi_i$, where $0 < \pi_i < 1$. Finally, write $g_i = 1-f_i$.

Theorem 1. Assuming the existence on each Q_i of the second moment of the random variable $X'(Y-P)$ and the first moment equal to zero, we have

$$n_m^{-\frac{1}{2}} (\hat{\underset{\sim}{U}} - \underset{\sim}{U}) \rightarrow N(0,\Sigma)$$

where

$$\Sigma = \sum_{i=1}^{K} \pi_i \; f^{-1} \frac{g_i}{f_i} \Sigma_i ,$$

$$f = \sum_i f_i \pi_i = \sum_i \lim_{m\to\infty} f_{mi} \frac{N_{mi}}{N_m} = \lim_{m\to\infty} \sum_i \frac{n_{mi}}{N_m} = \lim_{m\to\infty} \frac{n_m}{N_m},$$

and Σ_i is the variance of $X'\cdot(Y-P)$ on Q_i.

Proof: Note that we will generally drop the subscript m from the $\hat{\underset{\sim}{U}}_m$, n_{mi}, N_{mi}, f_{mi}, etc.

$$n^{-\frac{1}{2}}(\hat{\underset{\sim}{U}} - U) = n^{-\frac{1}{2}} \left\{ \sum_{i=1}^{K} f_i^{-1} \sum_{j=1}^{n_i} X'_{ij}(Y_{ij} - P_{ij}) - \sum_{i=1}^{K} \sum_{j=1}^{N_i} X'_{ij}(Y_{ij} - P_{ij}) \right\}$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{K} (f_i^{-1} - 1) \sum_{j=1}^{n_i} X'_{ij}(Y_{ij} - P_{ij}) - n^{-\frac{1}{2}} \sum_{i=1}^{K} \sum_{j=n_i+1}^{N_i} X'_{ij}(Y_{ij} - P_{ij}).$$

We have

$$n_i^{-\frac{1}{2}} (f_i^{-1} - 1) \sum_{j=1}^{n_i} X'_{ij}(Y_{ij} - P_{ij}) \to (f_i^{-1} - 1)\, N(0,\Sigma_i),$$

and

$$n_i^{-\frac{1}{2}} \sum_{j=n_i+1}^{N_i} X'_{ij}(Y_{ij} - P_{ij}) = (1 - f_i)^{\frac{1}{2}} f_i^{-\frac{1}{2}} (N_i - n_i)^{-\frac{1}{2}} \sum_{j=n_i+1}^{N_i} X'_{ij}(Y_{ij} - P_{ij})$$

$$\to (1 - f_i)^{\frac{1}{2}} f_i^{-\frac{1}{2}} N(0,\Sigma_i)$$

by the Lindberg-Levy Central Limit Theorem. Thus,

$$A_i = n_i^{-\frac{1}{2}} \left\{ f_i^{-1} \sum_{j=1}^{n_i} X'_{ij}(Y_{ij} - P_{ij}) - \sum_{j=1}^{N_i} X'_{ij}(Y_{ij} - P_{ij}) \right\}$$

$$\to N(0,\{(f_i^{-1} - 1)^2 + (1 - f_i) f_i^{-1}\} \Sigma_i) = N(0, n_i^{-1} N_i \frac{g_i}{f_i} \Sigma_i).$$

So:

$$n^{-\frac{1}{2}}(\hat{\underset{\sim}{U}} - \underset{\sim}{U}) = \sum_{i=1}^{K} \left(\frac{n_i}{n}\right)^{\frac{1}{2}} A_i \to N(0, \sum_{i=1}^{K} \pi_i f^{-1} \frac{g_i}{f_i} \Sigma_i).$$

Theorem 2 (see Binder (1981)). Under the assumptions of Theorem 1, plus an additional regularity condition as specified in Rao (1973), p.293, on the third-order partial derivatives (with respect to $\underset{\sim}{\beta}$) of $P_{ij}(\underset{\sim}{\beta})$, i.e., on the $X_r X_s X_t\ P(1-P)(1-2P)$, $n^{\frac{1}{2}}(\hat{\underset{\sim}{B}} - \underset{\sim}{B})$ is asymptotically $N(0, \mathcal{I}^{-1}\Sigma\mathcal{I}^{-1})$, where

$$\mathcal{I} = f^{-1} \sum_{i=1}^{K} \pi_i \mathcal{I}_i ,$$

and

$$\mathcal{I}_i = \text{PLIM } N_i^{-1} \sum_{j=1}^{N_i} X'_{ij} X_{ij}\ P_{ij}(\underset{\sim}{\beta})\{1-P_{ij}(\underset{\sim}{\beta})\}.$$

Proof: By a Taylor series expansion, for any j=i,...,r,

$$n^{-\frac{1}{2}}\, U_j(\underset{\sim}{B}) = n^{-\frac{1}{2}}\, U_j(\underset{\sim}{\beta}) + n^{-\frac{1}{2}} \left(\frac{\partial U_j}{\partial \underset{\sim}{\beta}}\right) (\underset{\sim}{B} - \underset{\sim}{\beta}) + 0_p(n^{-\frac{1}{2}}),$$

using the additional regularity condition and the fact that $\underset{\sim}{B}$ is a consistent estimator of $\underset{\sim}{\beta}$. But since $\underset{\sim}{U}(\underset{\sim}{B}) = 0$, we have

$$\text{PLIM} \left\{ n^{-\frac{1}{2}}\, \underset{\sim}{U}(\underset{\sim}{\beta}) + n^{-\frac{1}{2}} \left(\frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}}\right) (\underset{\sim}{B} - \underset{\sim}{\beta}) \right\} = 0.$$

$$\frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}} = - \sum_{i=1}^{K} \sum_{j=1}^{N_i} X'_{ij} X_{ij}\ P_{ij}(\underset{\sim}{\beta})\{1-P_{ij}(\underset{\sim}{\beta})\},$$

so,

$$\text{PLIM } \left(N^{-1} \frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}}\right) = - \text{PLIM} \sum_{i=1}^{K} \frac{N_i}{N} N_i^{-1} \sum_{j=1}^{N_i} X'_{ij} X_{ij}\ P_{ij}(\underset{\sim}{\beta})\{1-P_{ij}(\underset{\sim}{\beta})\}$$

$$= - \sum_{i=1}^{K} \pi_i \mathcal{I}_i = - f\mathcal{I}.$$

Thus,

$$\text{PLIM} \left\{ n^{-\frac{1}{2}}\, \underset{\sim}{U}(\underset{\sim}{\beta}) + \frac{N}{n} \left(N^{-1} \frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}}\right) n^{\frac{1}{2}} (\underset{\sim}{B} - \underset{\sim}{\beta}) \right\}$$

$$= \text{PLIM} \left\{ n^{-\frac{1}{2}}\, \underset{\sim}{U}(\underset{\sim}{\beta}) + f^{-1}(-f\mathcal{I}) n^{\frac{1}{2}} (\underset{\sim}{B} - \underset{\sim}{\beta}) \right\}$$

$$= \text{PLIM} \left\{ n^{-\frac{1}{2}}\, \underset{\sim}{U}(\underset{\sim}{\beta}) - \mathcal{I} n^{\frac{1}{2}} (\underset{\sim}{B} - \underset{\sim}{\beta}) \right\} = 0. \qquad \text{(a)}$$

Also,

$$\text{PLIM}\left\{n^{-\frac{1}{2}}\,\hat{U}(\beta) + n^{-\frac{1}{2}}\left(\frac{\partial\hat{U}}{\partial\beta}\right)(\hat{B}-\beta)\right\} = 0,$$

and

$$\text{PLIM}\; n^{-1}\frac{\partial\hat{U}}{\partial\beta} = \text{PLIM}\left\{-n^{-1}\sum_{i=1}^{K} f_i^{-1}\sum_{j=1}^{n_i} X_{ij}'X_{ij}\,P_{ij}(1-P_{ij})\right\}$$

$$= -\sum_{i=1}^{K}\text{PLIM}\left\{\frac{n_i}{N_i}\frac{N_i}{N}\frac{N}{n} f_i^{-1}\, n_i^{-1}\sum_{j=1}^{n_i} X_{ij}'X_{ij}\,P_{ij}(1-P_{ij})\right\}$$

$$= -f^{-1}\sum_{i=1}^{K}\pi_i \mathcal{I}_i = -\mathcal{I}\,.$$

So,

$$\text{PLIM}\left\{n^{-\frac{1}{2}}\,\hat{U}(\beta) + n^{-1}\left(\frac{\partial\hat{U}}{\partial\beta}\right) n^{\frac{1}{2}}(\hat{B}-\beta)\right\} = \text{PLIM}\left\{n^{-\frac{1}{2}}\,\hat{U}(\beta) - \mathcal{I}n^{\frac{1}{2}}(\hat{B}-\beta)\right\} = 0. \quad \text{(b)}$$

From (a) and (b) we obtain:

$$\text{PLIM}\left[n^{\frac{1}{2}}(\hat{B}-B) - \mathcal{I}^{-1}n^{-\frac{1}{2}}\{\hat{U}(\beta) - U(\beta)\}\right]$$

$$= \text{PLIM}\left\{n^{\frac{1}{2}}(\hat{B}-\beta) - \mathcal{I}^{-1}n^{-\frac{1}{2}}\,\hat{U}(\beta)\right\}$$

$$- \text{PLIM}\left\{n^{\frac{1}{2}}(B-\beta) - \mathcal{I}^{-1}n^{-\frac{1}{2}}\,U(\beta)\right\}.$$

Thus, $n^{\frac{1}{2}}(\hat{B}-B) \to N(0, \mathcal{I}^{-1}\Sigma\,\mathcal{I}^{-1})$.

<u>Theorem 3</u>. Under the hypothesis of Theorems 1 and 2,

(i) a consistent estimator for Σ is:

$$\sum_{i=1}^{K}\frac{N_i}{n}\frac{g_i}{f_i}\hat{\Sigma}_i = \sum_{i=1}^{K}\frac{n_i g_i}{n f_i^2}\hat{\Sigma}_i\,,$$

where

$$\hat{\Sigma}_i = \frac{1}{n_i-1}\left(\sum_{j=1}^{n_i} X_{ij}'X_{ij}\{Y_{ij} - P_{ij}(\hat{B})\}^2\right.$$

$$\left. - \frac{1}{n_i}\left[\sum_{j=1}^{n_i} X_{ij}'\{Y_{ij} - P_{ij}(\hat{B})\}\cdot\sum_{j=1}^{n_i} X_{ij}\{Y_{ij} - P_{ij}(\hat{B})\}\right]\right);$$

and

(ii) a consistent estimator for $I^{-1}\Sigma I^{-1}$ is $\hat{I}^{-1}\hat{\Sigma}\hat{I}^{-1}$, where

$$\hat{I} = \sum_{i=1}^{K} \frac{N_i}{n} \hat{I}_i = \sum_{i=1}^{K} \frac{N_i}{n} \frac{\sum_{j=1}^{n_i} X'_{ij}X_{ij} \; P_{ij}(\hat{\beta})\{1-P_{ij}(\hat{\beta})\}}{n_i}$$

$$= n^{-1} \sum_{i=1}^{K} f_i^{-1} \sum_{j=1}^{n_i} X'_{ij}X_{ij} \; P_{ij}(\hat{\beta})\{1-P_{ij}(\hat{\beta})\}.$$

Proof: This follows from the weak law of large numbers and the fact that $\hat{B}$ converges to β in probability.

Note that the proofs of Theorems 1 through 3 had nothing to do with the particular form of $P = F^{-1}(X\beta)$. The conclusions of the theorems hold whatever the form of these functions, as long as the related regularity conditions on the functions hold and we replace X by the Nxr matrix with entries

$$\frac{\partial P_{ij}}{\partial \beta_t} \frac{1}{P_{ij}(1-P_{ij})}$$

in U and change

$$\frac{\partial U}{\partial \beta}$$

accordingly.

5. HYPOTHESIS TESTING

We wish to show that the usual type of test statistics for testing hypotheses of the form H_0: $AB = C$ are valid in our sampling framework.

1) Wald Statistic: $W = n(A\hat{B}-C)'(AI^{-1}\Sigma I^{-1}A')(A\hat{B}-C)$

Under H_0, $n^{\frac{1}{2}}(A\hat{B}-C) \sim N(0, AI^{-1}\Sigma I^{-1}A')$, so the distribution of W is asymptotically χ^2 with rank $(A\mathcal{I}^{-1}\Sigma\mathcal{I}^{-1}A') =$ rank (A) degrees of freedom. Of course, we use the estimate $A\hat{\mathcal{I}}^{-1}\hat{\Sigma}\hat{\mathcal{I}}^{-1}A'$ for the variance.

2) Likelihood Ratio Statistic: $LR = -2\log\{\hat{L}(\hat{\underset{\sim}{B}}^*)/\hat{L}(\hat{\underset{\sim}{B}})\}$, where $\hat{\underset{\sim}{B}}^*$ maximizes $\hat{L}(\underset{\sim}{\beta})$ under the restriction $A\underset{\sim}{\beta} = C$. One can show that under H_0, PLIM $[-2\{\hat{\ell}(\hat{\underset{\sim}{B}}^*) - \hat{\ell}(\hat{\underset{\sim}{B}})\}] =$ PLIM $[-2\{\ell(\underset{\sim}{B}^*) - \ell(B)\}]$, where ℓ is the log-likelihood and $\hat{\ell}$ its sample estimate. By the usual MLE theory the latter is χ^2 with rank (A) d.f.

3) Score Statistic: $S = n^{-1}\ \hat{\underset{\sim}{U}}(\hat{\underset{\sim}{B}}^*)'\Sigma(\hat{\underset{\sim}{B}}^*)^{-1}\ \hat{\underset{\sim}{U}}(\hat{\underset{\sim}{B}}^*)$

Now $n^{-\frac{1}{2}}\ (\hat{\underset{\sim}{U}}(\underset{\sim}{\beta}) - \underset{\sim}{U}(\underset{\sim}{\beta}))$ is asymptotically $N(0,\Sigma(\underset{\sim}{\beta}))$. But $\underset{\sim}{U}(\underset{\sim}{B}) = 0$, so $n^{-\frac{1}{2}}\ \hat{\underset{\sim}{U}}(B)$ is asymptotically $N(0,\Sigma(\underset{\sim}{B}))$.

We can write $A\underset{\sim}{B} = C$ equivalently, with a possible re-ordering of the components of $\underset{\sim}{B}$, as $\underset{\sim}{B} = (\underset{\sim}{B}_1'\ f(\underset{\sim}{B}_1)')'$, where f is an affine function, and $\underset{\sim}{B}_1$ has dimension r-rank (A). Then $\hat{\underset{\sim}{B}}^* = (\hat{\underset{\sim}{B}}_1', f(\hat{\underset{\sim}{B}}_1)')'$, where $\hat{\underset{\sim}{B}}_1$ is the MLE for $\hat{L}$ considered as a function of r-rank (A) parameters through the relation $\underset{\sim}{B} = (\underset{\sim}{B}_1', f(\underset{\sim}{B}_1)')'$. Write $\hat{\underset{\sim}{U}} = (\hat{\underset{\sim}{U}}_1', \hat{\underset{\sim}{U}}_2')'$, where $\hat{\underset{\sim}{U}}_2$ is (rank Ax1). By definition $\hat{\underset{\sim}{U}}_1(\hat{\underset{\sim}{B}}_1) = 0$. So $n^{-\frac{1}{2}}\ \hat{\underset{\sim}{U}}(\hat{\underset{\sim}{B}}^*) = n^{-\frac{1}{2}}\ (0', \hat{\underset{\sim}{U}}_2(\hat{\underset{\sim}{B}}^*)')'$ is asymptotically $N(0,\Sigma(\hat{\underset{\sim}{B}}^*))$, under H_0, since $(\hat{\underset{\sim}{B}}^* - \underset{\sim}{B}) \to 0$ in probability under H_0. Thus, $n^{-1}\ \hat{\underset{\sim}{U}}(\hat{\underset{\sim}{B}}^*)\ \Sigma(\hat{\underset{\sim}{B}}^*)^{-1}\ \hat{\underset{\sim}{U}}(\hat{\underset{\sim}{B}}^*) \sim \chi^2_{\text{rank (A)}}$.

6. FITTING A MODEL ON A SUBDOMAIN

Suppose we are interested in fitting a model on a subpopulation $\mathcal{R}'$ of our finite population $\mathcal{R}$. Designate with primes the parameters associated with $\mathcal{R}'$, e.g., N_i' is the size of the i^{th} stratum of subpopulation $\mathcal{R}'$. Let $\mathcal{S}' = \mathcal{S} \cap \mathcal{R}'$. If the N_i' are known, then we could assume that $\mathcal{S}'$ is a stratified random sample of $\mathcal{R}'$, with sampling proportion $f_i' = n_i'/N_i'$, and proceed as before. However, if the N_i' are not known, we must proceed otherwise.

We can estimate N_i' as $\hat{N}_i' = (n_i'/n_i)N_i$. Suppose we wanted to fit a linear model $E(Y) = X\underset{\sim}{B}$ on R'. Write

$$\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{Y}_2 \end{pmatrix}, \qquad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \qquad \text{on} \qquad \begin{pmatrix} R' \\ R - R' \end{pmatrix},$$

with similar notation on S. We want the least squares estimate of $\underset{\sim}{Y}_1 = X_1\underset{\sim}{B}$, that is, we want to solve the equation $(X_1' X_1)\underset{\sim}{B} = X_1'\underset{\sim}{Y}_1$. If we knew the sampling proportions f_i' for the subpopulation, we would estimate $X_1'X_1$ and $X_1'\underset{\sim}{Y}_1$ by the Horvitz-Thompson estimators as we did on the entire finite population. Not knowing the f_i', we immediately think of estimating them by:

$$\hat{f}_i' = \frac{n_i'}{\hat{N}_i'} = \frac{n_i'}{\frac{n_i'}{n_i} N_i} = \frac{n_i}{N_i} = f_i,$$

and then employing these in estimating $X_1'X_1$ and $X_1'\underset{\sim}{Y}_1$. Writing the weight matrix W as

$$\begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix},$$

where W_1 corresponds to the subpopulation of interest, the estimates for $X_1'X_1$ and $X_1'\underset{\sim}{Y}_1$ are then $X_1'W_1X_1$ and $X_1'W_1\underset{\sim}{Y}_1$, so that $\hat{\underset{\sim}{B}} = (X_1'W_1X_1)^{-1}(X_1'W_1\underset{\sim}{Y}_1)$.

To get the distribution of our estimator, however, we need to return to the sampling framework of the entire finite population R. Define

$$\underset{\sim}{Y}^* = \begin{cases} \underset{\sim}{Y} & \text{in } R' \\ 0 & \text{off } R' \end{cases}$$

with $\underset{\sim}{X}^*$ defined similarly. Then

$$\underset{\sim}{Y}^* = \begin{bmatrix} \underset{\sim}{Y}_1 \\ 0 \end{bmatrix} \quad \text{and} \quad X^* = \begin{bmatrix} X_1 \\ 0 \end{bmatrix} \quad \text{on } S.$$

If we fit the model $Y^* = X^*\underset{\sim}{B}^*$ on R, we get $\hat{\underset{\sim}{B}}^* = (X^{*\prime}WX^*)^{-1}X^{*\prime}W\underset{\sim}{Y}^* =$

$$\left[\begin{pmatrix} X_1' & 0 \end{pmatrix}\begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix}\begin{pmatrix} X_1 \\ 0 \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} X_1' & 0 \end{pmatrix}\begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix}\begin{pmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{0} \end{pmatrix}\right] = (X_1'W_1X_1)^{-1}\, X_1'W_1\underset{\sim}{Y}_1 = \hat{\underset{\sim}{B}},$$

which is the estimator derived above. Now we consider the estimated variance of $\hat{\underset{\sim}{B}} - \underset{\sim}{B}$:

$$\begin{aligned} \text{var}(\hat{\underset{\sim}{B}} - \underset{\sim}{B}) = \text{var}(\hat{\underset{\sim}{B}}^* - \underset{\sim}{B}^*) &= (X^{*\prime}WX^*)^{-1}\, \hat{G}(X^{*\prime}WX^*)^{-1} \\ &= (X_1'W_1X_1)^{-1}\, \hat{G}(X_1'W_1X_1), \qquad (5) \end{aligned}$$

where

$$\hat{G} = \sum_{i=1}^{K} \frac{n-1}{n-r}\, \frac{n_i}{n_i-1}\, \frac{1-f_i}{f_i^2}\, X_i^{*\prime}\, (V_i^{*2} - \frac{1}{n_i}\, V_i^*\, J_i^*\, V_i^*)\, X_i^*,$$

J_i^* is an $n_i x n_i$ matrix of units, and $V_i^* = \text{diag}(\underset{\sim}{Y}_i^* - X_i^*\, \hat{\underset{\sim}{B}}^*)$ is the diagonal matrix formed from $\underset{\sim}{Y}_i^* - X_i^*\, \hat{\underset{\sim}{B}}^*$ (see Fuller (1975)).

If the stratum-specific subpopulation sampling proportions f_i' were presumed to equal f_i, then the corresponding estimated variance, denoted V, would "underestimate" var $\hat{B}^*$. For large n_i' we have approximately

$$\text{var}\, \hat{\underset{\sim}{B}}^* - V = \sum_{i=1}^{K} \left(\frac{1-f_i}{f_i^2}\right)\left(\frac{1}{n_i'} - \frac{1}{n_i}\right) X_i^{*\prime}\, V_i^*\, J_i^*\, V_i^*\, X_i^* \,.$$

It is clear that the inaccuracy in using V as an estimate of var $\hat{\underset{\sim}{B}}^*$ worsens as the disparity between the subpopulation sample size and the full sample size increases within each stratum.

To bring the reader to more familiar ground, we can take $\underset{\sim}{B}$ as a scalar parameter B, and $\underset{\sim}{X} = 1$, and it can then be shown that (5) simplifies to the estimate in Cochran (1977) of the variance of the ratio estimator of domain means.

Now consider logistic regression on the finite subpopulation R'. With X* being 0 off R', we get

$$L(\underset{\sim}{\beta};X^*) = \prod_{i=1}^{K} \prod_{j\varepsilon R_i'} \left\{\frac{\exp(X_{ij}\underset{\sim}{\beta})}{1+\exp(X_{ij}\underset{\sim}{\beta})}\right\}^{Y_{ij}} \left\{\frac{1}{1+\exp(X_{ij}\underset{\sim}{\beta})}\right\}^{1-Y_{ij}}$$

$$\cdot \prod_{i=1}^{K} \prod_{j\varepsilon R_i/R_i'} \left(\frac{1}{2}\right)^{Y_{ij}} \left(\frac{1}{2}\right)^{1-Y_{ij}}$$

$$= \sum_{i=1}^{K} 2^{n_i'-n_i} \cdot \prod_{i=1}^{K} \prod_{j\varepsilon R_i'} \left\{\frac{\exp(X_{ij}\underset{\sim}{\beta})}{1+\exp(X_{ij}\underset{\sim}{\beta})}\right\}^{Y_{ij}} \left\{\frac{1}{1+\exp(X_{ij}\underset{\sim}{\beta})}\right\}^{1-Y_{ij}}$$

$$= \text{constant} \cdot L_1(\underset{\sim}{\beta}),$$

where $L_1(\underset{\sim}{\beta})$ is the likelihood on R'. We see that the device of equating X to 0 off R' gives the "right" likelihood, up to a constant, and the "right" likelihood equation,

$$\underset{\sim}{U}_1 = \sum_{i=1}^{K} \sum_{j=1}^{N_i'} X_{ij}'(Y_{ij} - P_{ij}) = \sum_{i=1}^{K} \sum_{j=1}^{N_i} X_{ij}^{*\prime}(Y_{ij} - P_{ij}).$$

The sample estimate of $\underset{\sim}{U}_1$ is the same, using X* (on R) or X (on R') with the estimated weights f_i^{-1}:

$$\hat{\underset{\sim}{U}}_1 = \sum_{i=1}^{K} \sum_{j=1}^{n_i'} f_i^{-1} X_{ij}'(Y_{ij} - P_{ij}) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} f_i^{-1} X_{ij}^{*\prime}(Y_{ij} - P_{ij}).$$

Then, from the right-hand expression and Theorem 1, $n^{-\frac{1}{2}}(\hat{\underset{\sim}{U}}_1 - \underset{\sim}{U}_1) \to N(0,\Sigma_1)$, where

$$\Sigma_1 = \sum_{i=1}^{K} \pi_i f_i^{-1} \frac{g_i}{f_i} \Sigma_{1i},$$

and

$$\Sigma_{1i} = \text{var}\ \{X^{*\prime}(Y-P)\}$$

on Q_i. Moreover, a consistent estimator for Σ_1 is:

$$\hat{\Sigma}_1 = \sum_{i=1}^{K} \frac{N_i}{n} \frac{g_i}{f_i} \hat{\Sigma}_{1i} = \sum_{i=1}^{K} \frac{n_i g_i}{n\ f_i^2} \hat{\Sigma}_{1i},$$

where $\hat{\Sigma}_{1i}$ is the usual sample covariance of $X_i^{*\prime}(Y_i-P_i)$ on the i^{th} stratum:

$$\hat{\Sigma}_{1i} = \frac{1}{n_i-1}\left[\sum_{j=1}^{n_i'} X_{ij}'X_{ij}(Y_{ij}-P_{ij})^2 - \frac{1}{n_i}\left\{\sum_{j=1}^{n_i'} X_{ij}'(Y_{ij}-P_{ij})\right\}\right.$$

$$\left.\cdot\left\{\sum_{j=1}^{n_i'} X_{ij}(Y_{ij}-P_{ij})\right\}\right].$$

If we assumed the subpopulation sampling proportions to be known, with $f_i'=f_i$, then the corresponding estimated variance Σ^* would be like $\hat{\Sigma}_1$, with every n_i in the formula replaced by n_i'. For n_i' large, we get $\hat{\Sigma}_1 - \Sigma^*$ approximately as

$$\sum_{i=1}^{K} \frac{g_i}{nf_i^2}\left(\frac{1}{n_i'} - \frac{1}{n_i}\right)\left\{\sum_{j=1}^{n_i'} X_{ij}'(Y_{ij}-P_{ij})\right\}\cdot\left\{\sum_{j=1}^{n_i'} X_{ij}(Y_{ij}-P_{ij})\right\}.$$

This difference then becomes smaller as $n_i' \to n_i$.

Writing $\underset{\sim}{B}^*$ for the parameter to be estimated on the finite subpopulation, we have from Theorem 2

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{B}}^* - \underset{\sim}{B}^*) \to N(0, I_1^{-1}\ \Sigma_1\ I_1^{-1}),$$

where $I_1^{-1}\ \Sigma_1\ I_1^{-1}$ is consistently estimated by $\hat{I}_1^{-1}\ \hat{\Sigma}_1\ \hat{I}_1^{-1}$ with

$$\hat{I}_1 = n^{-1}\sum_{i=1}^{K} f_i^{-1} \sum_{j=1}^{n_i} X_{ij}^{*\prime}\ X_{ij}^{*}P_{ij}(1-P_{ij}) = n^{-1}\sum_{i=1}^{K} f_i^{-1}\sum_{j=1}^{n_i'} X_{ij}'X_{ij}P_{ij}(1-P_{ij}).$$

Hence, the total error in estimating $\text{var}(\hat{\underset{\sim}{B}}^* - \underset{\sim}{B}^*)$ by taking the subpopulation weights as f_i^{-1} and known comes in the estimate of Σ_1.

The procedure discussed above to estimate $\underset{\sim}{B}$ on a subpopulation R' made no use of the value of Y off the subpopulation. Since $P_{ij} = ½$ off R', it may be of use to set $Y = ½$ on $R - R'$ so that the residual $Y - P$ is identically zero off R'. This would have particular application for goodness-of-fit tests which involved the residuals.

One may also want to fit a separate model on each of R' and $R - R'$. Defining an indicator variable I to identically zero on R' and unity off, we could then take

$$P_{ij} = \frac{\exp\{X_{ij}\underset{\sim}{B}_1 + (X_{ij}I_{ij})\underset{\sim}{B}_2\}}{1 + \exp\{X_{ij}\underset{\sim}{B}_1 + (X_{ij}I_{ij})\underset{\sim}{B}_2\}} .$$

We assume here that X contains a column of ones, i.e., that the model has an intercept term, so that X•I includes I as a variable in the model. Then $\underset{\sim}{B}_1$ corresponds to the subpopulation of interest, and $\underset{\sim}{B}_3 = \underset{\sim}{B}_1 + \underset{\sim}{B}_2$ to its complement. The likelihood of R then breaks up into two factors, one a function of $\underset{\sim}{B}_1$ and the other of $\underset{\sim}{B}_3$, and maximization of the product is equivalent to maximization of each factor separately, so that we get the same estimates of $\underset{\sim}{B}$ on Q as before. This approach will often be impractical when it comes to computations, either because of the increase in the number of variables or because of a lack of data in $R - R'$.

7. ALGORITHMS FOR ESTIMATION

In order to estimate the coefficients of the logistic regression model, an iteratively reweighted least squares program was used. As in Jennrich and Moore (1975), one can show that the "normal" equations from maximum likelihood estimation of our model are the same as those for a least squares estimation for a related nonlinear model.

Write $f(\hat{\underset{\sim}{\beta}}) = (\underset{\sim}{Y} - \underset{\sim}{P}(\underset{\sim}{\beta}))'WT(\underset{\sim}{Y} - \underset{\sim}{P}(\underset{\sim}{\beta}))$ for the weighted sums of squares of residuals to be minimized. Taking partial derivation of f with respect to $\underset{\sim}{\beta}$ we get the normal equations

$$0 = (\frac{\partial \underset{\sim}{P}}{\partial \underset{\sim}{\beta}})' \cdot WT \cdot (\underset{\sim}{Y} - \underset{\sim}{P}).$$

Since $\frac{\partial \underset{\sim}{P}}{\partial \underset{\sim}{\beta}_t} = X_{ijt} P_{ij}(1 - P_{ij})$, if we take WT as a diagonal matrix with diagonal elements $f_i^{-1} P_{ij}^{-1}(1 - P_{ij})^{-1}$ we see that the normal equations (4') result. Of course the weights depend on $\underset{\sim}{\beta}$, so that at each stage of the iteratively weighted least squares routine, the weights are re-calculated using the most recent estimates of the $\underset{\sim}{\beta}$. We have adapted PROC NLIN in SAS (1979) to get our estimates of $\underset{\sim}{\beta}$, and have written our own software to utilize these estimates in calculating variances and test statistics.

8. ILLUSTRATION

Using data from the Lipid Research Clinics Program Prevalence Study, parameter estimates from a weighted logistic regression model are compared to those from an analogous model that ignores the stratified random sampling scheme. The Prevalence Study consisted of two screening visits. The first, Visit 1, was designed as a complete screen of ten well-defined target populations: consider these Visit 1 subjects as the finite population R. The details of the actual complex sampling strategy are not presented; but, with some simplification, we take the Visit 2 sample as a stratified random sample of the finite Visit 1 population (S), where the strata are defined by race and lipid zone. This latter variable, lipid zone, has three categories: elevated levels of either serum cholesterol or triglyceride at Visit 1; borderline elevated levels of cholesterol or triglyceride; and normal serum cholesterol and triglyceride levels. The sampling proportions for the first two zones were 100% and 15% for each race and for the third zone, 25% for whites and 32% for blacks.

The logistic regression framework is used to estimate age-adjusted race-specific Visit 1 prevalences of Type IIa dyslipoproteinemia (essentially, serum low density lipoprotein cholesterol (LDL), for males aged 6-19 years from one urban clinic. Define the dependent variable Y to be one for Type IIa subjects and zero, otherwise. Variable R is a race variable with value zero for whites and one for blacks, whereas A is an age variable, equal to age minus 12. The model of interest is

$$\text{logit } \Pr\{\underset{\sim}{Y}=1\} = B_0 + B_1 A + B_2 R$$

Parameter estimates ($\hat{\underset{\sim}{B}}$) when this model is considered in a finite random sampling scheme as well as estimates obtained within a finite stratified random sampling scheme are displayed in Table 1. The parameter estimates and the estimated standard errors for the simple random sampling scheme have been obtained from PROC LOGIST in SAS (1979). However, the standard errors that appear in Table 1 have been modified by a finite population correction factor ($\sqrt{1-f_i}$) to improve comparability.

Inspection of Table 1 reveals a notable difference between the two estimates of the intercept: the estimate calculated within the stratified random sampling scheme is much smaller than that calculated in a simple random sampling framework (-2.95 vs -1.93). The estimated prevalences of phenotype IIa are shown in Table 2. There are substantial differences between these two sets of probabilities. Since LDL is defined as "elevated" if above the age-specific 95th percentile for the white population, the "weighted" estimates must be the correct ones. Ignoring the stratified random sampling design in this nonlinear regression model results in an overestimation of these two prevalences. Such an error would have serious implications, for example, in health planning if this proportion were used to calculate an expected number of individuals in a population requiring specific health resources. The number of blacks would be overestimated by 90% and the number of whites by 130%, resulting in overestimation of required health services.

TABLE 1

Coefficient Estimates for Logistic Regression Models
Ignoring or Accounting for Sampling Design

	Ignoring Design		Considering Design	
Variable	$\hat{\beta}$	Std.Err.($\hat{\beta}$)	$\hat{\beta}$	Std.Err.($\hat{\beta}$)
Intercept	-1.93	0.11	-2.95	0.13
Age	-0.08	0.03	-0.05	0.03
Race	0.43	0.19	0.70	0.24

TABLE 2

Estimated Age-Adjusted Race-Specific Prevalences of Type IIa Dyslipoproteinemia Ignoring or Accounting for Sampling Design

	Ignoring Design	Considering Design
P(Type IIa|White)*	0.13	0.05
P(Type IIa|Black)*	0.18	0.10

*Calculated at Age = 12.

Suppose the primary hypothesis of interest is a black-white comparison in the prevalence of phenotype IIa. The two approaches yield roughly the same conclusion that blacks have an elevated prevalence of phenotype IIa, but the magnitude of this excess differs. Accounting for the sampling scheme, we find that juvenile black males have 2.0 (= exp(0.70)) times the odds of being Type IIa as whites, while if we fit the model ignoring the sampling scheme we get an odds ratio of 1.53 (= exp(0.43)).

9. EXTENSION OF LARGE SAMPLE THEORY TO A TWO-STAGE STRATIFIED SAMPLE

In this section the proposed methods are extended to a two-stage stratified sampling design. Firstly, a choice must be made on what types of "asymptotics" to consider. We have followed Krewski and Rao (1981) in focusing "on surveys with large numbers of strata with relatively few primary sampling units selected within each stratum," so that the limiting processes consider the number of strata as increasing. We begin by presenting the framework for the asymptotic theory.

Consider the sequence $\{R_K\}$ of finite populations such that R_K

has K strata with N_i primary sampling units in the i^{th} stratum, $i=1,\ldots,K$. Suppose n_i primary sampling units (PSU's) are selected from the i^{th} stratum, with single-draw sampling probabilities $\pi_{ij} > 0$ $(j=1,\ldots,N_i)$, where $\Sigma_j \pi_{ij} = 1$ for all i. The j^{th} primary sampling unit of the i^{th} stratum contains M_{ij} subjects, of which a simple random non-replacement sample of m_{ij} units is drawn. Assume the primary sampling units are selected with replacement and that independent samples are selected within those PSU's selected more than once. Note that in the above and subsequent notation for parameters for the k^{th} population R_K, K has consistently been suppressed for simplicity.

When the finite population in the j^{th} PSU of the i^{th} stratum of the finite population R_K is considered as a random sample from an infinite superpopulation, the likelihood on R_K can be written as

$$L(\underset{\sim}{\beta}) = \prod_{i=1}^{K} \prod_{j=1}^{N_i} \prod_{k=1}^{M_{ij}} \left\{\frac{\exp(X_{ijk}\underset{\sim}{\beta})}{1+\exp(X_{ijk}\underset{\sim}{\beta})}\right\}^{Y_{ijk}} \left\{\frac{1}{1+\exp(X_{ijk}\underset{\sim}{\beta})}\right\}^{1-Y_{ijk}}$$

$$= \prod_{i=1}^{K} \prod_{j=1}^{N_i} \prod_{k=1}^{M_{ij}} P_{ijk}^{\;Y_{ijk}} (1 - P_{ijk})^{1-Y_{ijk}},$$

where

$$P_{ijk} = pr(Y_{ijk}=1 | X_{ijk}) = \frac{\exp(X_{ijk}\underset{\sim}{\beta})}{1+\exp(X_{ijk}\underset{\sim}{\beta})} .$$

Of particular interest are the likelihood equations:

$$U_S = \frac{\partial \log L(\underset{\sim}{\beta})}{\partial \beta_S}$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{N_i} \sum_{k=1}^{M_{ij}} X_{ijk}(Y_{ijk} - P_{ijk}) = 0.$$

Let

$$w_{ij} = \left(\frac{\pi_{ij}\, m_{ij}\, n_i}{M_{ij}}\right)^{-1},$$

$$\hat{L}(\underset{\sim}{\beta}) = \prod_{i=1}^{K} \prod_{j=1}^{n_i} \left\{ \prod_{k=1}^{m_{ij}} P_{ijk}^{\;Y_{ijk}} (1 - P_{ijk})^{1-Y_{ijk}} \right\}^{w_{ij}},$$

and

$$\hat{U}_S = \frac{\partial \log \hat{L}(\beta)}{\partial \beta_S} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} w_{ij} X_{ijks} (Y_{ijk} - P_{ijk}).$$

Then $\hat{U}_S$ is an unbiased estimator of U_S.

The solution to the equations $\underset{\sim}{U} = 0$ defines the parameter of interest $\underset{\sim}{B}$, and the solution to $\hat{\underset{\sim}{U}} = 0$ provides our estimate $\hat{\underset{\sim}{B}}$. Asymptotic results are expressed in the following two theorems.

Theorem 4. Under suitable regularity conditions, $K^{-\frac{1}{2}} (\hat{\underset{\sim}{U}} - \underset{\sim}{U}) \to N(0,\Gamma)$ as $K \to \infty$, where Γ is defined below.

Proof:

$$\hat{\underset{\sim}{U}} - \underset{\sim}{U} = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} w_{ij} X'_{ijk} (Y_{ijk} - P_{ijk}) - \sum_{i=1}^{K} \sum_{j=1}^{N_i} \sum_{k=1}^{M_{ij}} X'_{ijk} (Y_{ijk} - P_{ijk})$$

$$= \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} X'_{ijk} (Y_{ijk} - P_{ijk})(w_{ij} - 1)$$

$$- \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{k=m_{ij}+1}^{m_{ij}} X'_{ijk} (Y_{ijk} - P_{ijk})$$

$$- \sum_{i=1}^{K} \sum_{j=n_i+1}^{N_i} \sum_{k=1}^{M_{ij}} X'_{ijk} (Y_{ijk} - P_{ijk})$$

$$= \sum_{i=1}^{K} \underset{\sim}{A}_i - \sum_{i=1}^{K} \underset{\sim}{B}_i - \sum_{i=1}^{K} \underset{\sim}{C}_i,$$

where the definitions of $\underset{\sim}{A}_i$, $\underset{\sim}{B}_i$, $\underset{\sim}{C}_i$ are the obvious ones.

Consider $\underset{\sim}{A}_i$ first. Let

$$\underset{\sim}{Z}_{ijk} = X'_{ijk}(Y_{ijk} - P_{ijk}).$$

Then

$$\underset{\sim}{A}_i = \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} \underset{\sim}{Z}_{ijk}(w_{ij} - 1) = \sum_{j=1}^{n_i} \underset{\sim}{A}_{ij},$$

and the expected value of $\underset{\sim}{A}_i$ given the n_i PSU is

$$E(\underset{\sim}{A}_i|\text{PSU}) = \sum_{j=1}^{n_i} (w_{ij} - 1)m_{ij}\, E(\underset{\sim}{Z}_{ijk}) = 0,$$

assuming $E(\underset{\sim}{Z}_{ijk}) = 0$ (regularity condition number one). Clearly $E(\underset{\sim}{A}_i) = 0$.

Now examine the variance of $\underset{\sim}{A}_i$:

$$\text{var}(\underset{\sim}{A}_i) = E(\underset{\sim}{A}_i\underset{\sim}{A}_i')$$

$$= E(\sum_{j=1}^{n_i} \underset{\sim}{A}_{ij}\underset{\sim}{A}_{ij}' + 2\sum_j \sum_{j'>j} \underset{\sim}{A}_{ij}\underset{\sim}{A}_{ij'}').$$

We know $E(\underset{\sim}{A}_{ij}\underset{\sim}{A}_{ij'}'|j\neq j') = 0$, and

$$E(\underset{\sim}{A}_{ij}\underset{\sim}{A}_{ij}'|\text{PSU}) = (w_{ij} - 1)^2\, E\left\{(\sum_{k=1}^{m_{ij}} \underset{\sim}{Z}_{ijk})(\sum_{k=1}^{m_{ij}} \underset{\sim}{Z}_{ijk}')\right\}$$

$$= (w_{ij} - 1)^2 \sum_{k=1}^{m_{ij}} E(\underset{\sim}{Z}_{ijk}\underset{\sim}{Z}_{ijk}').$$

$$= (w_{ij} - 1)^2\, m_{ij}\, \Sigma_{ij}.$$

Averaging over all PSU's produces the desired result:

$$\text{var}(\underset{\sim}{A}_i) = \sum_{j=1}^{N_i} \pi_{ij}n_i(w_{ij} - 1)^2\, m_{ij}\, \Sigma_{ij}$$

$$= \sum_{j=1}^{N_i} \frac{(M_{ij} - \pi_{ij}m_{ij}\, n_i)^2}{\pi_{ij}m_{ij}\, n_i}\, \Sigma_{ij}.$$

With suitable conditions on moments of the $\underset{\sim}{A}_i$,

$$K^{-\frac{1}{2}} \sum_{i=1}^{K} \underset{\sim}{A}_i \rightarrow N(0,\Gamma_1),$$

where

$$\Gamma_1 = \underset{K\rightarrow\infty}{\text{PLIM}} \frac{1}{K} \sum_{i=1}^{K} \text{var}(\underset{\sim}{A}_i)$$

is assumed to be positive definite. Now we repeat this process for the $\underset{\sim}{B}_i$ and $\underset{\sim}{C}_i$.

$$\underset{\sim}{B}_i = \sum_{j=1}^{n_i} \sum_{k=m_{ij}+1}^{M_{ij}} \underset{\sim}{Z}_{ijk} = \sum_{j=1}^{n_{ij}} \underset{\sim}{B}_{ij}.$$

As before, $E(\underset{\sim}{B}_i) = 0$, and

$$\text{var}(\underset{\sim}{B}_i) = \sum_{j=1}^{N_i} \pi_{ij} n_i \; E(\underset{\sim}{B}_{ij} \underset{\sim}{B}_{ij}' | j)$$

$$= \sum_{j=1}^{N_i} \pi_{ij} n_i (M_{ij} - m_{ij}) \; \Sigma_{ij}.$$

Again, with the right regularity conditions,

$$K^{-\frac{1}{2}} \sum_{i=1}^{K} \underset{\sim}{B}_i \rightarrow N(0,\Gamma_2),$$

where

$$\Gamma_2 = \underset{K\rightarrow\infty}{\text{PLIM}} \frac{1}{K} \sum_{i=1}^{K} \text{var}(\underset{\sim}{B}_i)$$

is assumed positive definite. Further,

$$K^{-\frac{1}{2}} \sum_{i=1}^{K} \underset{\sim}{C}_i \rightarrow N(0,\Gamma_3),$$

where

$$\Gamma_3 = \underset{K\rightarrow\infty}{\text{PLIM}} \frac{1}{K} \sum_{i=1}^{K} \text{var}(\underset{\sim}{C}_i),$$

$$\underset{\sim}{C}_i = \sum_{j=n_i+1}^{N_i} \sum_{k=1}^{M_{ij}} \underset{\sim}{Z}_{ijk},$$

and

$$\mathrm{var}(\underset{\sim}{C}_i) = \sum_{j=1}^{N_i} (N_i - n_i)\pi_{ij}M_{ij}\,\Sigma_{ij}.$$

Thus $K^{-\frac{1}{2}}(\hat{\underset{\sim}{U}} - \underset{\sim}{U}) \to N(0,\Gamma)$, for

$$\Gamma = \sum_{t=1}^{3}\Gamma_t = \underset{K\to\infty}{\mathrm{PLIM}}\frac{1}{K}\sum_{i=1}^{K}\{\mathrm{var}(\underset{\sim}{A}_i) + \mathrm{var}(\underset{\sim}{B}_i) + \mathrm{var}(\underset{\sim}{C}_i)\}$$

$$= \underset{K\to\infty}{\mathrm{PLIM}}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{N_i}\left\{\frac{(M_{ij} - \pi_{ij}m_{ij}\,n_i)^2}{n_i\,\pi_{ij}m_{ij}} + n_i\pi_{ij}(M_{ij} - m_{ij})\right.$$

$$\left. + \pi_{ij}(N_i - n_i)M_{ij}\right\}\Sigma_{ij}$$

$$= \underset{K\to\infty}{\mathrm{PLIM}}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{N_i} M_{ij}(M_{ij} - 2\pi_{ij}m_{ij}\,n_i + n_i m_{ij}\,\pi_{ij}^2 N_i)\,\Sigma_{ij}.$$

<u>Theorem 5</u>. Under suitable regularity conditions

$$K^{\frac{1}{2}}(\hat{\underset{\sim}{B}} - \underset{\sim}{B}) \to N(0,\, \Delta^{-1}\,\Gamma\,\Delta^{-1}),$$

where Δ is defined below.

<u>Proof</u>: As in Theorem 2,

$$\mathrm{PLIM}\,\{K^{-\frac{1}{2}}\,\underset{\sim}{U}(\underset{\sim}{\beta}) + K^{-\frac{1}{2}}\left(\frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}}\right)(\underset{\sim}{B} - \underset{\sim}{\beta})\} = 0.$$

Now

$$\frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}} = -\sum_{i=1}^{K}\sum_{j=1}^{N_i}\sum_{k=1}^{M_{ij}} X'_{ijk}X_{ijk}\,P_{ijk}(1 - P_{ijk}).$$

Write

$$D_{ijk} = X'_{ijk}X_{ijk}\,P_{ijk}(1 - P_{ijk}), \quad \text{and} \quad E(\underset{\sim}{D}_{ijk}|i,j) = \nu_{ij}.$$

Then

$$E\Big(\sum_{j=1}^{N_i}\sum_{k=1}^{M_{ij}} D_{ijk}\Big) = \sum_{j=1}^{N_i} M_{ij}\nu_{ij}\,.$$

So,

$$\underset{K\to\infty}{\text{PLIM}} \left(\frac{1}{K}\frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}} + \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{N_i} M_{ij}\nu_{ij}\right) = 0$$

under suitable regularity conditions. We assume that

$$\lim_{K\to\infty} \frac{\sum_{i=1}^{K}\sum_{j=1}^{N_i} M_{ij}\nu_{ij}}{K} = \Delta$$

is positive definite. Thus,

$$0 = \text{PLIM}\ \{K^{-\frac{1}{2}}\ \underset{\sim}{U}(\underset{\sim}{\beta}) + K^{-\frac{1}{2}}\ (\frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}})(\underset{\sim}{B} - \underset{\sim}{\beta})\} \qquad \text{(c)}$$

$$= \text{PLIM}\ \{K^{-\frac{1}{2}}\ \underset{\sim}{U}(\underset{\sim}{\beta}) + (K^{-1}\ \frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{\beta}})\ K^{\frac{1}{2}}\ (\underset{\sim}{B} - \underset{\sim}{\beta})\}$$

$$= \text{PLIM}\ \{K^{-\frac{1}{2}}\ \underset{\sim}{U}(\underset{\sim}{\beta}) - \Delta\ K^{\frac{1}{2}}\ (\underset{\sim}{B} - \underset{\sim}{\beta})\}.$$

Also,

$$\text{PLIM}\ \{K^{-\frac{1}{2}}\ \hat{\underset{\sim}{U}}(\underset{\sim}{\beta}) + K^{-\frac{1}{2}}\ (\frac{\partial \hat{\underset{\sim}{U}}}{\partial \underset{\sim}{\beta}})(\hat{\underset{\sim}{B}} - \underset{\sim}{\beta})\} = 0.$$

Now

$$\frac{1}{K}\frac{\partial \hat{\underset{\sim}{U}}}{\partial \underset{\sim}{\beta}} = -\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}} w_{ij}\underset{\sim}{D}_{ijk},$$

and

$$E\left(\sum_{j=1}^{n_i}\sum_{k=1}^{m_{ij}} w_{ij}\underset{\sim}{D}_{ijk}\right) = \sum_{j=1}^{N_i} (n_i\pi_{ij}\ m_{ij}w_{ij}\ \nu_{ij}) = \sum_{j=1}^{N_i} M_{ij}\nu_{ij}.$$

So,

$$\underset{K\to\infty}{\text{PLIM}}\ \left(\frac{1}{K}\frac{\partial \hat{\underset{\sim}{U}}}{\partial \underset{\sim}{\beta}}\right) = -\ \Delta,$$

and

$$0 = \text{PLIM}\ \{K^{-\frac{1}{2}}\ \hat{\underset{\sim}{U}}(\underset{\sim}{\beta}) - \Delta K^{\frac{1}{2}}\ (\hat{\underset{\sim}{B}} - \underset{\sim}{\beta})\}. \qquad \text{(d)}$$

As in Theorem (3), from (c) and (d) we get

$$\text{PLIM}\ [K^{\frac{1}{2}}\ (\hat{\underset{\sim}{B}} - \underset{\sim}{B}) - \Delta^{-1}\ K^{-\frac{1}{2}}\ \{\hat{\underset{\sim}{U}}(\underset{\sim}{\beta}) - \underset{\sim}{U}(\underset{\sim}{\beta})\}] = 0,$$

hence

$$K^{\frac{1}{2}}\ (\hat{\underset{\sim}{B}} - \underset{\sim}{B}) \to N(0,\ \Delta^{-1}\Gamma\ \Delta^{-1}).$$

For a consistent estimator of Δ, we take

$$\hat{\Delta} = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} \sum_{k=1}^{m_{ij}} w_{ij} \hat{D}_{ijk},$$

where $\hat{D}_{ijk}$ is D_{ijk} evaluated at the MLE. For Γ, take

$$\hat{\Gamma} = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} \frac{M_{ij}(M_{ij} - 2\pi_{ij} m_{ij} n_i + n_i m_{ij} \pi_{ij}^2 N_i)}{\pi_{ij} n_i} \hat{\Sigma}_{ij}$$

for $\hat{\Sigma}_{ij}$, the sample covariance of D_{ijk} on the j^{th} PSU of the i^{th} stratum.

The hypothesis tests which follow from these theorems can then be derived just as in Section 5.

ACKNOWLEDGEMENT

The authors are grateful to Dawn Stewart for her programming assistance. We also wish to thank Dave DeLong for helpful discussions regarding PROC NLIN and Paul Stewart for his contribution to the subdomain methodology. The skillful typing of this manuscript by Ernestine Bland is acknowledged.

REFERENCES

Barr, A.J., Goodnight, J.H., and Sall, J.P. (1979). SAS Users Guide, SAS Institute Inc.: North Carolina.

Binder, D.A. (1981). "On the variance of asymptotically normal estimators from complex surveys", Statistics Canada, Industrial and Agricultural Survey Methods Division, Ottowa.

Cochran, W.G. (1977). Sampling Techniques, 3rd edition, John Wiley and Sons: New York.

DuMouchel, W.H. and Duncan, G.J. (1981). "Using sample survey weights in multiple regression analysis of stratified samples", American Statistical Association, 1981 Proceedings of the Section on Survey Research Methods, 629-637.

Fuller, W.A. (1975). "Regression analysis for sample survey", Sankhyā 37, 117-132.

Jennrich, R.I. and Moore, R.H. (1975). "Maximum likelihood estimation by means of nonlinear least squares", American Statistical Association, 1975 Proceedings of the Statistical Computing Section, 57-65.

Krewski, D. and Rao, J.N.K. (1981). "Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods", The Annals of Statistics 9, 1010-1019.

O'Brien, K.F. (1981). "Life table analysis for complex survey data", Ph.D. Thesis, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina.

Rao, C.R. (1973). Linear Statistical Inference and Its Application, 2nd edition, John Wiley and Sons: New York.

Sárndal, C.E. (1978). "Design-based and model-based inference in survey sampling", Scand. J. Statist. 5, 25-52.