FITTING GENERAL LINEAR MODEL FOR LONGITUDINAL SURVEY DATA UNDER INFORMATIVE SAMPLING

ABDULHAKEEM A.H. EIDEH[1]
AL-QUDS UNIVERSITY, PALESTINE

Statistics in Transition-New Series, Vol 11, No 3, 2010

ABSTRACT

The purpose of this article is to account for informative sampling in fitting superpopulation model for multivariate observations, and in particular multivariate normal distribution, for longitudinal survey data. The idea behind the proposed approach is to extract the model holding for the sample data as a function of the model in the population and the first order inclusion probabilities, and then fit the sample model using maximum likelihood, pseudo maximum likelihood and estimating equations methods. As an application of the results, we fit the general linear model for longitudinal survey data under informative sampling using different covariance structures: the exponential correlation model, the uniform correlation model, and the random effect model, and using different conditional expectations of first order inclusion probabilities given the study variable. The main feature of the present estimators is their behaviours in terms of the informativeness parameters.

Key words: General Linear Model, Informative sampling, Longitudinal Survey Data, Maximum Likelihood, and Sample distribution.

1. Introduction

Sampling designs for surveys are often complex and informative, in the sense that the selection probabilities are correlated with the variables of interest, even when conditioned on explanatory variables. In this case conventional analysis that disregards the informativeness can be seriously biased, since the sample distribution differs from that of the population. Most of the studies in social surveys are based on data collected from complex sampling designs. Standard analysis of survey data often fails to account for the complex nature of the sampling design such as the use of unequal selection probabilities, clustering, and post-stratification. The effect of the sample design on the analysis is due to the fact that the models in use typically do not incorporate all the design variables determining the sample selection, either because there may be too many of them or because they are not of substantive interest.

[1] *Address for correspondence*: Dr. ABDULHAKEEM A.H. EIDEH, Department of Mathematics, College of Science and Technology, Al-Quds University, Abu-Dies campus, Palestine, P.O. Box 20002, Jerusalem. E-mail: msabdul@science.alquds.edu.

However, if the sampling design is informative in the sense that the outcome variable (variable of interest) is correlated with the design variables not included in the model, even after conditioning on the model covariates, standard estimates of the model parameters can be severely biased, leading possibly to false inference. Pfeffermann (1993, 1996) reviews many examples reported in the literature that illustrate the effects of ignoring the sampling process when fitting models to survey data and discusses methods that have been proposed to deal with this problem, see also Skinner, Holt, and Smith (1989), Kasprzyk, Duncan, Kalton and Singh (1989), Hoem, (1989), and Chambers and Skinner (2003). It should be emphasized that standard inference may be biased even when the original sample is a simple random sample, due to non-response, attrition and imperfect frames that results in de facto a posterior differential inclusion probabilities.

To overcome the difficulties associated with the use of classical inference procedures for cross sectional survey data, Pfeffermann, Krieger and Rinott (1998) proposed the use of the sample distribution induced by the assumed population models, under informative sampling, and developed expressions for its calculation. Similarly, Eideh and Nathan (2006) fitted time series models for longitudinal survey data under informative sampling. Furthermore, Eideh (2008) fitted random effects or subject-specific effects models for analyzing normal data, which are assumed to be correlated, under the concept of informative sampling.

The plan of this paper is as follows. In Section 2 we define sample distribution and sample likelihood. In Section 3 we extract the sampled distribution of the multivariate normal distribution under informative sampling. In Section 4 we fit the general linear model for longitudinal survey data. Section 5 provides a discussion of the results.

2. Sample distribution and sample likelihood

Let $U = \{1,...,N\}$ denote a finite population consisting of N units. Let y be the target or study variable of interest and let y_i be the value of y for the ith population unit. Let x_i, $i \in U$ be the value of an auxiliary variable(s), x, and $\mathbf{z} = \{z_1,...,z_N\}$ be the values of a known design variable, used for the sample selection process but not included in the working model under consideration. In what follows we consider a sampling design with selection probabilities $\pi_i = \Pr(i \in s) > 0$, and sampling weight $w_i = 1/\pi_i$; $i = 1,...,N$. In practice, the π_i's may depend on the population values (x,y,z). We express this by writing: $\pi_i = \Pr(i \in s \mid x,y,z)$. The sample s consists of the subset of U selected at random by the sampling scheme with inclusion probabilities $\pi_1,...,\pi_N$. Denote by $\mathbf{I} = (I_1,...,I_N)'$ the N by 1 sample indicator (vector) variable such that $I_i = 1$ if unit $i \in U$ is selected to the sample and $I_i = 0$ if otherwise. The sample s is defined accordingly as $s = \{i \mid i \in U, I_i = 1\}$ and its complement by $c = \bar{s} = \{i \mid i \in U, I_i = 0\}$. We assume probability sampling, so that $\pi_i = \Pr(i \in s) > 0$ for all units $i \in U$.

We now consider the population values $y_1,...,y_N$ as random variables, which are independent realizations from a distribution with probability density function $f_p(y_i \mid x_i,\theta)$, indexed by a vector parameter θ.
According to Krieger and Pfeffermann (1997), the (marginal) sample probability density function of y_i is defined as:

$$\begin{aligned} f_s(y_i \mid x_i,\theta,\gamma) &= f_p(y_i \mid x_i,\theta,\gamma, i \in \mathrm{s}) \\ &= \frac{\Pr(i \in \mathrm{s} \mid x_i, y_i,\gamma) f_p(y_i \mid \mathbf{x}_i,\theta)}{\Pr(i \in \mathrm{s} \mid x_i,\theta,\gamma)} \\ &= \frac{E_p(\pi_i \mid x_i, y_i,\gamma) f_p(y_i \mid x_i,\theta)}{E_p(\pi_i \mid x_i,\theta,\gamma)} \end{aligned} \tag{1}$$

where θ is the parameter of the population distribution, γ is the parameter indexing $\Pr(i \in s \mid x_i, y_i,\gamma)$ and

$$E_p(\pi_i \mid x_i,\theta,\gamma) = \int E_p(\pi_i \mid x_i, y_i,\gamma) f_p(y_i \mid x_i,\theta) dy_i$$

Having derived the sample distribution, Pfeffermann, Krieger and Rinott (1998) proved that if the population measurements y_i are independent, then as $N \to \infty$ (with n fixed, where n is the sample size), the sample measurements are asymptotically independent, so we can apply standard inference procedures to complex survey data by using the marginal sample distribution for each unit. Based on the sample data $\{y_i, x_i, w_i;\ i \in s\}$, we can estimate the parameters of the population model in two steps:

Step-one: According to Pfeffermann and Sverchkov (1999), estimate the informativeness parameters γ using the following relationship:

$$E_s(w_i \mid x_i, y_i,\gamma) = 1 / E_p(\pi_i \mid x_i, y_i,\gamma) \tag{2}$$

Thus the informativeness parameters can be estimated using regression analysis. Denoting the resulting estimate of γ by $\tilde{\gamma}$.

Step-two: Substitute $\tilde{\gamma}$ in the sample log-likelihood function, and then maximize the resulting sample log-likelihood function with respect to the population parameters, θ:

$$\begin{aligned} l_{rs}(\theta,\tilde{\gamma}) &= l_{srs}(\theta) - \sum_{i=1}^{n} \log E_p(\pi_i \mid \mathbf{x}_i,\theta,\tilde{\gamma}) \\ &= l_{srs}(\theta) + \sum_{i=1}^{n} \log E_s(w_i \mid \mathbf{x}_i,\theta,\tilde{\gamma}) \end{aligned} \tag{3}$$

where $l_{rs}(\theta,\tilde{\gamma})$ is the sample log-likelihood after substituting $\tilde{\gamma}$ in the sample log-likelihood function and where

$$l_{srs}(\theta) = \sum_{i \in s} \log\{f_p(y_i \mid x_i, \theta)\}$$

is the classical log-likelihood obtained by ignoring the sample design.

3. Multivariate normal distribution under informative sampling

Most classical methods of multivariate analysis of continuous data are based on an examination of the structure of population mean vectors and covariance matrices. We consider the problem of estimating the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\mathbf{V}$ of a vector of study variables $\mathbf{y}$, from survey data obtained under informative sampling. The original theoretical basis for this is the multivariate normal distribution. The existing work in this area deals with the estimation problem when the sampling scheme is noninformative; see for example Smith and Holmes (1989).

The following theorem focuses on multivariate normal distribution under different modeling of the population conditional expectation of first order inclusion probabilities.

Theorem 1. Assume that the population distribution is q-dimensional multivariate normal with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$, that is:

$$\mathbf{y}_i = (y_{i1},..., y_{iq})' \underset{p}{\sim} N_q(\boldsymbol{\mu}, \mathbf{V}), i = 1,..., N$$

Let $E_p(y_{ij}) = \mu_j$ and $Cov_p(y_{ij}, y_{ik}) = v_{jk}$, $i = 1,..., N$; $j, k = 1,2,..., q$.

1. Under the exponential inclusion probability model – exponential sampling:

$$\begin{aligned} E_p(\pi_i \mid \mathbf{y}_i) &= \exp(a_0 + \mathbf{a}'\mathbf{y}_i) \\ &= \exp(a_0)\prod_{j=1}^{q} \exp(a_j y_{ij}) \end{aligned} \tag{4}$$

The sample probability density function of $\mathbf{y}_i$ is:

$$f_s(\mathbf{y}_i) = (2\pi)^{-0.5q} |\mathbf{V}|^{-0.5} \exp\left[-0.5\{\mathbf{y}_i - (\boldsymbol{\mu} + \mathbf{V}\mathbf{a})\}' \mathbf{V}^{-1} \{\mathbf{y}_i - (\boldsymbol{\mu} + \mathbf{V}\mathbf{a})\}\right] \tag{5}$$

That is, $\mathbf{y}_i = (y_{i1},..., y_{iq})' \underset{s}{\sim} N_q(\boldsymbol{\mu} + \mathbf{V}\mathbf{a}, \mathbf{V}), i = 1,..., n$. So that,

$$\begin{aligned} E_s(\mathbf{y}_i) &= E_p(\mathbf{y}_i \mid i \in s) \\ &= \boldsymbol{\mu} + \mathbf{V}\mathbf{a} = E_p(\mathbf{y}_i) + \mathbf{V}\mathbf{a} \end{aligned}$$

and

$$\begin{aligned} Var_s(\mathbf{y}_i) &= Var_p(\mathbf{y}_i \mid i \in s) \\ &= \mathbf{V} = Var_p(\mathbf{y}_i) \end{aligned}$$

2. Under the linear inclusion probability model – linear sampling:

$$\begin{aligned} E_p(\pi_i \mid \mathbf{y}_i) &= b_0 + \mathbf{b}'\mathbf{y}_i \\ &= b_0 + \sum_{t=1}^{q} b_j y_{ij} \end{aligned} \tag{6}$$

The sample probability density function of $\mathbf{y}_i$ is:

$$f_s(\mathbf{y}_i) = \frac{(b_0 + \mathbf{b}'\mathbf{y}_i)(2\pi)^{-0.5q}|\mathbf{V}|^{-0.5}\exp\left\{-0.5(\mathbf{y}_i - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}}{b_0 + \mathbf{b}'(\boldsymbol{\mu})} \tag{7}$$

Furthermore,

$$\begin{aligned} E_s(\mathbf{y}_i) &= \boldsymbol{\mu} + \frac{\mathbf{Vb}}{b_0 + \mathbf{b}'\boldsymbol{\mu}} \\ &= E_p(\mathbf{y}_i) + \frac{\mathbf{Vb}}{b_0 + \mathbf{b}'\boldsymbol{\mu}} \end{aligned} \tag{8a}$$

and

$$\begin{aligned} Var_s(\mathbf{y}_i) = Cov_s(\mathbf{y}_i) &= \mathbf{V} - \frac{(\mathbf{Vb})(\mathbf{Vb})'}{(b_0 + \mathbf{b}'\boldsymbol{\mu})^2} \\ &= Var_p(\mathbf{y}_i) - \frac{(\mathbf{Vb})(\mathbf{Vb})'}{(b_0 + \mathbf{b}'\boldsymbol{\mu})^2} \end{aligned} \tag{8b}$$

Proofs:

1. As an extension of equation (1), the sample probability density function of $\mathbf{y}_i$ is given by:

$$f_s(\mathbf{y}_i) = f_p(\mathbf{y}_i \mid i \in \mathrm{s}) = \frac{E_p(\pi_i \mid \mathbf{y}_i) f_p(\mathbf{y}_i)}{E_p(\pi_i)} \tag{9}$$

So that,

$$\begin{aligned} f_s(\mathbf{y}_i) &= \frac{\exp(\mathbf{a}'\mathbf{y}_i)(2\pi)^{-0.5q}|\mathbf{V}|^{-0.5}\exp\left\{-0.5(\mathbf{y}_i - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}}{\exp(\mathbf{a}'\boldsymbol{\mu} + 0.5\mathbf{a}'\mathbf{V}\mathbf{a})} \\ &= (2\pi)^{-\frac{q}{2}}|\mathbf{V}|^{-0.5}\exp(-0.5\mathbf{a}'\mathbf{V}\mathbf{a})\exp(-0.5Q) \end{aligned} \tag{10}$$

where $Q = (\mathbf{y}_i - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) - 2\mathbf{a}'(\mathbf{y}_i - \boldsymbol{\mu})$.

Setting $\mathbf{C}_i = \mathbf{y}_i - \boldsymbol{\mu}$, then Q can be written as: $\mathrm{Q} = \mathbf{C}_i'\mathbf{V}^{-1}\mathbf{C}_i - 2\mathbf{a}'\mathbf{C}_i$ .Using theorems in multivariate statistical analysis, see Johnson and Wichern (1998), page 68, we have:

$$Q = \mathbf{C}_i' \mathbf{V}^{-0.5}\mathbf{V}^{-0.5}\mathbf{C}_i - 2\mathbf{a}'\mathbf{V}^{0.5}\mathbf{V}^{-0.5}\mathbf{C}_i$$

Now let $\mathbf{D}_i = \mathbf{V}^{-0.5}\mathbf{C}_i$, then :

$$\begin{aligned} Q &= \mathbf{D}_i' \mathbf{D}_i - 2\mathbf{a}'\mathbf{V}^{0..5}\mathbf{D}_i \\ &= \mathbf{D}_i' \mathbf{D}_i - 2\mathbf{a}'\mathbf{V}^{0.5}\mathbf{D}_i + \mathbf{a}'\mathbf{V}^{0.5}\mathbf{V}^{0.5}\mathbf{a} - \mathbf{a}'\mathbf{V}^{0.5}\mathbf{V}^{0.5}\mathbf{a} \\ &= \left(\mathbf{D}_i - \mathbf{V}^{0.5}\mathbf{a}\right)'\left(\mathbf{D}_i - \mathbf{V}^{0.5}\mathbf{a}_1\right) - \mathbf{a}'\mathbf{V}\mathbf{a} \end{aligned}$$

But $\mathbf{C}_i = \mathbf{y}_i - \boldsymbol{\mu}$ and $\mathbf{D}_i = \mathbf{V}^{-0.5}\mathbf{C}_i$, so that Q can be expressed equivalently as:

$$\begin{aligned} Q &= \left(\mathbf{V}^{-0.5}\left(\left(\mathbf{y}_i - \boldsymbol{\mu}\right) - \mathbf{V}^{0.5}\mathbf{V}^{0.5}\mathbf{a}\right)\right)'\left(\mathbf{V}^{-0.5}\left(\left(\mathbf{y}_i - \boldsymbol{\mu}\right) - \mathbf{V}^{0.5}\mathbf{V}^{0.5}\mathbf{a}\right)\right) - \mathbf{a}'\mathbf{V}\mathbf{a} \\ &= \left(\mathbf{y}_i - \left(\boldsymbol{\mu} + \mathbf{V}\mathbf{a}\right)\right)'\mathbf{V}^{-1}\left(\mathbf{y}_i - \left(\boldsymbol{\mu} + \mathbf{V}\mathbf{a}\right)\right) - \mathbf{a}'\mathbf{V}\mathbf{a} \end{aligned}$$

Thus after substituting this expression of Q in (10), we get (5).
Hence the multivariate normal distribution in the sample is the same as in the population, except that the mean is shifted by the constant $\mathbf{Va}$. Notice that the sample probability density function does not depend on a_0 . Note also:

$$E_s\left(y_{ij}\right) = \mu_j + a_1 v_{j1} + \ldots + a_j v_{jj} + \ldots + a_q v_{jq} \tag{11a}$$

and

$$Var_s\left(y_{ij}\right) = v_{jj}, Cov_p\left(y_{ij}, y_{ik}\right) = v_{jk},\ j \neq k = 1,\ldots,q \tag{11b}$$

2. Substitute (6) in (9) we get (7).

Let us now compute the first two moments of this sample pdf. In order to do this we will use the moment generating function technique. The moment generating function of the sample probability density function is given by:

$$\begin{aligned} M_s\left(\mathbf{u}_i\right) &= E\left(\exp\left(\mathbf{u}_i'\mathbf{y}_i\right)\right) = \int \left(\exp\left(\mathbf{u}_i'\mathbf{y}_i\right)\right)\frac{b_0 + \mathbf{b}'\mathbf{y}_i}{b_0 + \mathbf{b}'\boldsymbol{\mu}} f_p\left(\mathbf{y}_i\right)d\mathbf{y}_i \\ &= \frac{b_0}{b_0 + \mathbf{b}'\boldsymbol{\mu}} M_p\left(\mathbf{u}_i\right) + \frac{\mathbf{b}'\dfrac{dM_p\left(\mathbf{u}_i\right)}{d\mathbf{u}_i}}{b_0 + \mathbf{b}'\boldsymbol{\mu}} \end{aligned}$$

where

$$M_p\left(\mathbf{u}_i\right) = \exp\left(\mathbf{u}_i'\boldsymbol{\mu} + .5\mathbf{u}_i'\mathbf{V}\mathbf{u}_i\right)$$

and

$$\frac{dM_p(\mathbf{u}_i)}{d\mathbf{u}_i} = M_p(\mathbf{u}_i)(\boldsymbol{\mu} + \mathbf{V}\mathbf{u}_i)$$

Thus we have:

$$M_s(\mathbf{u}_i) = \frac{M_p(\mathbf{u}_i)}{b_0 + \mathbf{b}'\boldsymbol{\mu}}(b_0 + \mathbf{b}'(\boldsymbol{\mu} + \mathbf{V}_0\mathbf{u}_i)) \tag{12}$$

This equation gives explicitly the relationship between the population and the sample moment generating functions. Notice that the sample and population moment generating functions are different unless $\mathbf{b} = \mathbf{0}$, in which case the sampling mechanism is noninformative.

Let $R_s(\mathbf{u}_i) = \log M_s(\mathbf{u}_i)$. Differentiating this expression twice and setting $\mathbf{u}_i = \mathbf{0}$, we get (8a) and (8b).
From (8a) and (8b), we can see that:

$$E_s(y_{ij}) = \mu_j + \frac{b_1 v_{j1} + ... + b_j v_{jj} + ... + b_q v_{jq}}{b_0 + b_1\mu_1 + ... + b_j\mu_j + ... + b_q\mu_{iq}}, \tag{13a}$$

$$Var_s(y_{ij}) = v_{jj} - \frac{(b_1 v_{j1} + ... + b_j v_{jj} + ... + b_q v_{jq})^2}{(b_0 + b_1\mu_1 + ...b_j\mu_j + ... + b_q\mu_q)^2} \tag{13b}$$

$$Cov_p(y_{ij}, y_{ik}) = v_{jk} - \frac{(b_1 v_{j1} + ... + b_j v_{jj} + ... + b_q v_{jq})(b_1 v_{k1} + ... + b_j v_{kj} + ... + b_q v_{kq})}{(b_0 + b_1\mu_1 + ...b_j\mu_j + ... + b_q\mu_q)^2}, \tag{13c}$$

for $i = 1,...,n,\ j \neq k = 1,2,...,q.$

Also $Var_s(y_{ij}) \leq Var_p(y_{ij})$ and the equality holds if and only if $\mathbf{b} = \mathbf{0}$, that is when the sampling mechanism is noninformative. On the other hand, if **b≠0**, then the means the variances and the covariances change, in contradiction to what happens for the sample probability density function (5), where only the means change.

To illustrate the results, from now on, we only consider the particular cases of the exponential inclusion probability model – exponential sampling, see equation (4).
The following are special cases of Theorem 1.

Corollary 1. (Univariate normal distribution, $q = 1$)
Let $E_p(y_{i1}) = \mu_1$ and $Var_p(y_{i1}) = v_{11}$. Under the exponential sampling:

$$E_p(\pi_i \mid y_{i1}) = \exp(a_0 + a_1 y_{i1}) \tag{14}$$

We get the following result:

$$y_{i1} \underset{s}{\sim} N(\mu_1 + a_1 v_{11}, v_{11}) \tag{15}$$

Corollary 2. (Bivariate normal distribution, $q = 2$).
Let $E_p(y_{i1}) = \mu_1, E_p(y_{i2}) = \mu_2,\ Var_p(y_{i1}) = v_{11},\ Var_p(y_{i2}) = v_{22},\ Cov_p(y_{i1}, y_{i2}) = v_{12}$ and $Cor_p(y_{i1}, y_{i2}) = \rho_{12}$. Under the exponential sampling:

$$E_p(\pi_i \mid y_{i1}, y_{i2}) = \exp(a_0 + a_1 y_{i1} + a_2 y_{i2}) \tag{16}$$

and using the properties of multivariate normal distribution, see Johnson and Wichern (1998), page 171, we obtain the following results:

1. The joint sample probability density function of (y_{i1}, y_{i2}) is:

$$(y_{i1}, y_{i2})' \underset{s}{\sim} N_2\left\{\begin{pmatrix} \mu_1 + a_1 v_{11} + a_2 v_{12} \\ \mu_2 + a_1 v_{21} + a_2 v_{22} \end{pmatrix}, \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}\right\} \tag{17a}$$

2. The sample marginal probability density functions of y_{i1} and y_{i2} are respectively given by:

$$y_{i1} \underset{s}{\sim} N(\mu_1 + a_1 v_{11} + a_2 v_{12}, v_{11}) \tag{17b}$$

and

$$y_{i2} \underset{s}{\sim} N(\mu_2 + a_1 v_{21} + a_2 v_{22}, v_{22}) \tag{17c}$$

3. The conditional sample probability density function of y_{i1} given y_{i2} is univariate normal distribution with conditional mean:

$$\begin{aligned} E_s(y_{i1} \mid y_{i2}) &= E_s(y_{i1}) + \rho_{12}\sqrt{\frac{v_{22}}{v_{11}}}(y_{i2} - E_s(y_{i2})) \\ &= \mu_1 + a_1 v_{11} + a_2 v_{12} + \rho_{12}\sqrt{\frac{v_{11}}{v_{22}}}(y_{i2} - (\mu_2 + a_1 v_{21} + a_2 v_{22})) \\ &= \mu_1 + \rho_{12}\sqrt{\frac{v_{11}}{v_{22}}}(y_{i2} - \mu_2) + a_1\left(v_{11} - \frac{v_{12}^2}{v_{22}}\right) \\ &= \mu_1 + \rho_{12}\sqrt{\frac{v_{11}}{v_{22}}}(y_{i2} - \mu_2) + a_1 v_{11}(1 - \rho_{12}^2) \end{aligned} \tag{17d}$$

and conditional variance:

$$V_s(y_{i1} \mid y_{i2}) = v_{11}(1 - \rho_{12}^2) = V_p(y_{i1} \mid y_{i2}) \tag{17e}$$

That is,

$$y_{i1} \mid y_{i2} \underset{s}{\sim} N\left\{\mu_1 + \rho_{12}(v_{11}/v_{22})^{0.5}(y_{i2} - \mu_2) + a_1 v_{11}(1 - \rho_{12}^2), v_{11}(1 - \rho_{12}^2)\right\} \tag{17f}$$

Notice that:

$$E_s(y_{i1} \mid y_{i2}) = E_p(y_{i1} \mid y_{i2}) + a_1 V_p(y_{i1} \mid y_{i2}) \tag{17g}$$

4. Similarly, the conditional sample probability density function of y_{i2} given y_{i1} is:

$$y_{i2} \mid y_{i1} \underset{s}{\sim} N\left\{\mu_2 + \rho_{12}(v_{22}/v_{11})^{0.5}(y_{i1} - \mu_1) + a_2 v_{22}(1-\rho_{12}^2), v_{22}(1-\rho_{12}^2)\right\} \tag{17h}$$

Thus the sample and population probability density functions are different, but belong to the same family of distribution, which is normal. Also the change occurs only for the means and conditional means, whereas the variances, conditional variances, and covariance do not change.

In particular if $a_2 = 0$, that is the inclusion probabilities depend only on y_{i1} (which is the case in panel surveys), then we have:

1. The joint sample probability density function of (y_{i1}, y_{i2}) is:

$$(y_{i1}, y_{i2})' \underset{s}{\sim} N_2\left\{\begin{pmatrix} \mu_1 + a_1 v_{11} \\ \mu_2 + a_1 v_{12} \end{pmatrix}, \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}\right\} \tag{18a}$$

2. The marginal sample probability density functions of y_{i1} and y_{i2} are respectively given by:

$$y_{i1} \underset{s}{\sim} N(\mu_1 + a_1 v_{11}, v_{11}) \tag{18b}$$

and

$$y_{i2} \underset{s}{\sim} N(\mu_2 + a_1 v_{12}, v_{22}) \tag{18c}$$

3. The conditional sample probability density function of y_{i1} given y_{i2} is:

$$y_{i1} \mid y_{i2} \underset{s}{\sim} N\left\{\mu_1 + a_1 v_{11}(1-\rho_{12}^2) + \rho_{12}(v_{11}/v_{22})^{0.5}(y_{i2} - \mu_2), v_{11}(1-\rho_{12}^2)\right\} \tag{18d}$$

4. The conditional sample probability density function of y_{i2} given y_{i1} is:

$$y_{i2} \mid y_{i1} \underset{s}{\sim} N\left\{\mu_2 + \rho_{12}(v_{22}/v_{11})^{0.5}(y_{i1} - \mu_1), v_{22}(1-\rho_{12}^2)\right\} \tag{18e}$$

Notice that the sample and population probability density functions of $y_{i2} \mid y_{i1}$ are same, while the other sample and population distributions are different.

Birnbaum et al. (1950) studied the effect of selection performed on some coordinates of a multi-dimensional population, but in different point of view.

Notice that the sample and population pdf's of $y_{i2} \mid y_{i1}$ are same, while the other sample and population distributions are different.

Corollary 3. (Bivariate normal distribution, $q = 2$).
Let $E_p(y_{i1}) = \mu_1, E_p(y_{i2}) = \mu_2$, $Var_p(y_{i1}) = v_{11}$, $Var_p(y_{i2}) = v_{22}$, $Cov_p(y_{i1}, y_{i2}) = v_{12}$ and $Cor_p(y_{i1}, y_{i2}) = \rho_{12}$. Under the linear model:

$$E_p(\pi_i \mid y_{i1}, y_{i2}) = b_0 + b_1 y_{i1} + b_2 y_{i2}$$

and using the properties of multivariate normal distribution, we have the following results:

1. The joint sample probability density function of (y_{i1}, y_{i2}) is:

$$f_s(y_{i1}, y_{i2}) = \frac{b_0 + b_1 y_{i1} + b_2 y_{i2}}{b_0 + b_1 \mu_1 + b_2 \mu_2} f_p(y_{i1}, y_{i2}) \tag{19a}$$

2. Integrating (19a), with respect to y_{i2}, we have:

$$\begin{aligned}
f_s(y_{i1}) &= \int f_s(y_{i1}, y_{i2}) dy_{i2} = \int \frac{b_0 + b_1 y_{i1} + b_2 y_{i2}}{b_0 + b_1 \mu_1 + b_2 \mu_2} f_p(y_{i1}, y_{i2}) dy_{i2} \\
&= \frac{1}{b_0 + b_1 \mu_1 + b_2 \mu_2} \left(\int b_0 f_p(y_{i1}, y_{i2}) dy_{i2} + \int b_1 y_{i1} f_p(y_{i1}, y_{i2}) dy_{i2} \right) + \\
&\quad \frac{1}{b_0 + b_1 \mu_1 + b_2 \mu_2} \int b_2 y_{i2} f_p(y_{i1}, y_{i2}) dy_{i2} \\
&= \frac{1}{b_0 + b_1 \mu_1 + b_2 \mu_2} \left(\int b_0 f_p(y_{i1}, y_{i2}) dy_{i2} + \int b_1 y_{i1} f_p(y_{i1}, y_{i2}) dy_{i2} \right) + \\
&\quad \frac{1}{b_0 + b_1 \mu_1 + b_2 \mu_2} \int b_2 y_{i2} f_p(y_{i1}) f_p(y_{i2} | y_{i1}) dy_{i2} \\
&= \frac{1}{b_0 + b_1 \mu_1 + b_2 \mu_2} \left(b_0 f_p(y_{i1}) + b_1 y_{i1} f_p(y_{i1}) + b_2 f_p(y_{i1}) E_p(y_{i2} | y_{i1}) \right) + \\
&\quad \frac{1}{b_0 + b_1 \mu_1 + b_2 \mu_2} \int b_2 y_{i2} f_p(y_{i1}) f_p(y_{i2} | y_{i1}) dy_{i2} \\
&= \left\{ \frac{b_0 + b_1 y_{i1} + \left(\mu_2 + \rho_{12} \left(\frac{v_{22}}{v_{11}} \right)^{0.5} (y_{i1} - \mu_1) \right)}{b_0 + b_1 \mu_1 + b_2 \mu_2} \right\} f_p(y_{i1})
\end{aligned} \tag{19b}$$

3. Similarly, integrating (19a), with respect to y_{i1}, we obtain the following marginal sample probability density function of y_{i2}:

$$f_s(y_{i2}) = \left\{ \frac{b_0 + b_2 y_{i2} + b_1 \left(\mu_1 + \rho_{12} \left(\frac{v_{11}}{v_{22}} \right)^{0.5} (y_{i2} - \mu_2) \right)}{b_0 + b_1 \mu_1 + b_2 \mu_2} \right\} f_p(y_{i2}) \qquad (19c)$$

4. Using (19a), (19b), and (19c), we get the following conditional sample probability density function of y_{i1} given y_{i2}:

$$\begin{aligned} f_s(y_{i1} \mid y_{i2}) &= \frac{f_s(y_{i1}, y_{i2})}{f_s(y_{i2})} \\ &= \frac{b_0 + b_1 y_{i1} + b_2 y_{i2}}{b_0 + b_1 \mu_1 + b_2 \mu_2} f_p(y_{i1}, y_{i2}) \div \frac{b_0 + b_2 y_{i2} + b_1 E_p(y_{i2} | y_{i1})}{b_0 + b_1 \mu_1 + b_2 \mu_2} f_p(y_{i2}) \\ &= \frac{b_0 + b_1 y_{i1} + b_2 y_{i2}}{b_0 + b_1 E_p(y_{i1} \mid y_{i2}) + b_2 y_{i2}} \frac{f_p(y_{i1}, y_{i2})}{f_p(y_{i2})} \\ &= \left\{ \frac{b_0 + b_1 y_{i1} + b_2 y_{i2}}{b_0 + b_1 E_p(y_{i1} \mid y_{i2}) + b_2 y_{i2}} \right\} f_p(y_{i1} \mid y_{i2}) \end{aligned} \qquad (19d)$$

where

$$E_p(y_{i1} \mid y_{i2}) = \mu_1 + \rho_{12} \left(\frac{v_{11}}{v_{22}} \right)^{0.5} (y_{i2} - \mu_2)$$

5. Similarly, the conditional sample probability density function of y_{i2} given y_{i1} is:

$$f_s(y_{i2} \mid y_{i1}) = \left\{ \frac{b_1 y_{i1} + b_2 y_{i2} + b_0}{b_2 E_p(y_{i2} \mid y_{i1}) + b_1 y_{i1} + b_0} \right\} f_p(y_{i2} \mid y_{i1}) \qquad (19e)$$

where

$$E_p(y_{i2} \mid y_{i1}) = \mu_2 + \rho_{12} \left(\frac{v_{22}}{v_{11}} \right)^{0.5} (y_{i1} - \mu_1)$$

Thus in this case we see that the sample probability density functions and population probability density functions are very different. Also notice that formulas (19b, c) are true, not only for the bivariate normal distribution, but for any joint probability density function of (y_{i1}, y_{i2}), provided that the marginal and conditional distributions and the corresponding moments are exist.

The following theorem provides the maximum likelihood estimators of the mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$.

Theorem 2. Assume $\mathbf{y}_i = (y_{i1},...,y_{iq})' \underset{p}{\sim} N_q(\boldsymbol{\mu}, \mathbf{V}), i = 1,...,N$ are independent. Let $\mathbf{y}_1,...,\mathbf{y}_n$ be a sample of size n selected by informative sampling.
1. Under the exponential sampling – equation (4): the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\mathbf{V}$ are given by:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{y}} - \hat{\mathbf{V}}\tilde{\mathbf{a}} \tag{20a}$$

and

$$\hat{\mathbf{V}} = n^{-1}\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' = \{\hat{\mathbf{V}}_{ij}\} = \{s_{ij}\} \tag{20b}$$

where $\tilde{\mathbf{a}}$ is the least square estimator under the model: $E_s(w_i \mid \mathbf{y}_i) = \exp(-a_0 - \mathbf{a}'\mathbf{y}_i)$, $\bar{\mathbf{y}} = (\bar{y}_1,...,\bar{y}_q)'$, $\bar{y}_i = n^{-1}\sum_{j=1}^{n} y_{ij}$ and $\hat{V}_{ij} = s_{ij} = n^{-1}\sum_{k=}^{n}(y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j)$.

2. Under the linear sampling – equation (6): the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\mathbf{V}$ are defined by the equations:

$$(\bar{\mathbf{y}} - \hat{\boldsymbol{\mu}})(\tilde{b}_0 + \tilde{\mathbf{b}}'\hat{\boldsymbol{\mu}}) = \hat{\mathbf{V}}\tilde{\mathbf{b}} \tag{21a}$$

and

$$n\hat{\mathbf{V}} = \sum_{i=1}^{n}(\bar{\mathbf{y}} - \hat{\boldsymbol{\mu}})(\bar{\mathbf{y}} - \hat{\boldsymbol{\mu}})' \tag{21b}$$

where $\tilde{b}_0$ and $\tilde{\mathbf{b}}$ are the least squares estimators under the model: $E_s(w_i \mid \mathbf{y}_i) = 1/(-b_0 - \mathbf{b}'\mathbf{y}_i)$. Solve (21a) and (21b) iteratively. Start with classical maximum likelihood estimators.

Proofs:
1. Exponential sampling. Using the two-step method of estimation.

Step-one. Estimation of informativeness parameters a_0 and $\mathbf{a}$ via the relationship (2).

Step-two. After substituting $\tilde{\mathbf{a}}$ in (5), the resulting sample log-likelihood is given by:

$$l_{rs}(\boldsymbol{\mu}, \mathbf{V}) = -0.5nq\log 2\pi - 0.5n\log|\mathbf{V}| - 0.5\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu}^*)'\mathbf{V}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}^*) \tag{22}$$

where $\boldsymbol{\mu}^* = \boldsymbol{\mu} + \mathbf{V}\tilde{\mathbf{a}}$.

According to Johnson and Wichern (1998), page 182, the maximum likelihood estimators of $\boldsymbol{\mu}^*$ and $\mathbf{V}$ are given by:

$$\hat{\boldsymbol{\mu}}^{*} = \bar{\mathbf{y}}$$

and

$$\hat{\mathbf{V}} = n^{-1}\sum_{i=1}^{n}\left(\mathbf{y}_i - \bar{\mathbf{y}}\right)\left(\mathbf{y}_i - \bar{\mathbf{y}}\right)'.$$

Hence the result – equation (20).

2. Linear model. Using the two-step method of estimation.

Step-one. Estimation of informativeness parameters, b_0 and $\mathbf{b}$ via the relationship (2).

Step-two. After substituting $\tilde{b}_0$ and $\tilde{\mathbf{b}}$ in (7), the resulting sample log-likelihood is given by:

$$l_{rs}(\boldsymbol{\mu},\mathbf{V}) = -\frac{1}{2}nq\log 2\pi - \frac{1}{2}n\log|\mathbf{V}| - \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{y}_i - \boldsymbol{\mu}\right)'\mathbf{V}^{-1}\left(\mathbf{y}_i - \boldsymbol{\mu}\right) - n\log\left(\tilde{b}_0 + \tilde{\mathbf{b}}'\boldsymbol{\mu}\right) \quad (23)$$

Now differentiating $l_{rs}(\boldsymbol{\mu},\mathbf{V})$ with respect to $\boldsymbol{\mu}$ and $\mathbf{V}$, we can show that the maximum likelihood estimators of $\boldsymbol{\mu}$ and $\mathbf{V}$ are defined by equation (21).

The following corollary provides the maximum likelihood estimators of the mean μ and the variance σ^2 when the population model is univariate normal and the sampling process is exponential and linear, which is a particular case of Theorem 2 with $q = 1$.

Corollary 4. Assume $y_i \underset{p}{\sim} N(\mu,\sigma^2), i = 1,...,N$ are independent. Let $y_1,...,y_n$ be a sample of size n selected under the following sampling schemes.
1.Exponential sampling. The maximum likelihood estimators of μ and σ^2 are given by:

$$\hat{\mu} = \bar{y} - \tilde{a}_1\hat{\sigma}^2 \quad (24a)$$

and

$$\hat{\sigma}^2 = s^2 = n^{-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \quad (24b)$$

where $\tilde{a}_1$ is the least square estimator of the informativeness parameter a_1.

2. Linear sampling. The maximum likelihood estimators of μ and σ^2 are given by:

$$\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\left(y_i - \hat{\mu}\right) - n\tilde{b}_1\left(\tilde{b}_0 + \tilde{b}_1\hat{\mu}\right)^{-1} = 0 \quad (25a)$$

and

$$\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{\mu}\right)^2 \tag{25b}$$

where $\tilde{b}_0$ and $\tilde{b}_1$ are the estimators of the informativeness parameters b_0 and b_1.

Corollary 5. Assume $\mathbf{y}_i = \left(y_{i1}, y_{i2}\right)' \underset{p}{\sim} N_2\left(\boldsymbol{\mu}, \mathbf{V}\right), i = 1,..., N$. Under the exponential sampling – equation (4): the maximum likelihood estimators of μ_1, μ_2, v_{11}, v_{22}, v_{11} and are given by:

$$\hat{\mu}_1 = \bar{y}_1 - \tilde{a}_1 s_{11} - \tilde{a}_2 s_{12} \tag{26a}$$

$$\hat{\mu}_2 = \bar{y}_2 - \tilde{a}_2 s_{22} - \tilde{a}_1 s_{12} \tag{26b}$$

$$\hat{\mathbf{V}} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \left\{s_{ij}\right\}, i, j = 1,2 \tag{27}$$

4. Application – fitting general linear model for longitudinal survey data under informative sampling

4.1 Population model:

Let $y_{it}, i = 1,..., N; t = 1,..., T$ be the measurement on the *i-th* subject at time $t = 1,..., T$. Associated with each y_{it} are the (known) values, $x_{itk}, k = 1,...p$, of p explanatory variables. We assume that the y_{it} follow the regression model:

$$y_{it} = \beta_1 x_{it1} + ... + \beta_p x_{itp} + \varepsilon_{it} \tag{28}$$

where ε_{it} are random sequence of length T associated with each of the N subjects. In our context, the longitudinal structure of the data means that we expect the ε_{it} to be correlated within subjects.

Let $\mathbf{y}_i = \left(y_{i1},..., y_{iT}\right)'$, $\mathbf{x}_{it} = \left(x_{it1},..., x_{itp}\right)'$ and let $\boldsymbol{\beta} = \left(\beta_1,..., \beta_p\right)'$ be the vector of unknown regression coefficients. The general linear model for longitudinal survey data treats the random vectors $\mathbf{y}_i, i = 1,..., N$ as independent multivariate normal variables, that is

$$\mathbf{y}_i \underset{p}{\sim} N_q\left(\mathbf{x}_i\boldsymbol{\beta}, \mathbf{V}\right) \tag{29}$$

where $\mathbf{x}_i$ is the matrix of size T by p of explanatory variables for subject i, and $\mathbf{V}$ has $\left(jk\right) - th$ element, $v_{jk} = \mathrm{cov}_p\left(y_{ij}, y_{ik}\right), j, k = 1,..., T$; see Diggle, Liang and Zeger (1994).

4.2 Covariance structure of $\mathbf{y}_i$:

It is useful at this stage to consider what form the matrix **V** might take. We consider three cases: the exponential correlation model, the uniform correlation model; see Diggle, Liang and Zeger (1994), and the random effect model; see Skinner and Holmes (2003).

Case1: The exponential correlation model:

In this model, the $(t,s)-$ th element of **V** has the form:

$$v_{ts} = \text{cov}_p(y_{it}, y_{is}) = \sigma^2 \rho^{|t-s|}, t, s = 1,...,T \tag{30}$$

Note that the correlation, v_{ts}/σ^2, between a pair of measurements on the same unit decays toward zero as the time separation between the measurements increases.

Case2: The uniform correlation model:

In this model, we assume that there is a positive correlation, ρ, between any two measurements on the same subject. So that the $(t,s)-th$ element of **V** has the form:

$$v_{ts} = \text{cov}_p(y_{it}, y_{is}) = \sigma^2 \rho, t \neq s = 1,...,T \;\; ; \;\; v_{tt} = \sigma^2, t = 1,...,T \tag{31}$$

Case3: Random effects models:

Under this model the multivariate outcomes $\mathbf{y}_i = (y_{i1},..., y_{iT})', i = 1,..., N$ are independent with mean vector and covariance matrix given respectively by:

$$E_p(\mathbf{y}_i) = (\beta_1,..., \beta_T) = \boldsymbol{\mu} \tag{32a}$$

$$\text{cov}_p(\mathbf{y}_i) = \sigma_u^2 \mathbf{J}_T + \sigma^2 \mathbf{V}_T = \sum \tag{32b}$$

where $\mathbf{J}_T$ denotes the T by T matrix all of whose elements are one, and the $(t,t')-th$ element of $\mathbf{V}_T$ is $\rho^{|t-t'|}; t, t' = .,...,T$.

4.3 Sampling design:

We assume a single-stage informative sampling design, where the sample is a panel sample selected at time $t = 1$ and all units remain in the sample till time $t = T$. Examples of longitudinal surveys, some of which are based on complex sample designs, and of the issues involved in their design and analysis can be found in Herriot and Kasprzyk (1984), and Nathan (1999). In many of the cases described in these papers, a sample is selected for the first round and continues to serve for several rounds. Then it is intuitively reasonable to assume that the first order inclusion probabilities, π_i, depend on the population values of the response variable at the first

occasion only, the values y_{i1}, and on $\mathbf{x}_{i1}=\left(x_{i11},\ldots,x_{i1p}\right)'$, and the values of known design variable, $\mathbf{z}=\{z_1,...,z_N\}$, used for the sample selection, but not included in the working model under consideration.

4. 4 Sample distribution:

Under exponential inclusion probability model:

$$E_p\left(\pi_i \mid y_{i1},\mathbf{x}_{i1}\right)=\exp\left(a_0^*+a_0y_{i1}+a_1x_{i11}+a_2x_{i12}+...+a_px_{i1p}\right) \tag{33}$$

Using (29), (33) and Theorem 1, we have:

$$\mathbf{y}_i\left|\mathbf{x}_i,\boldsymbol{\theta},a_0\underset{s}{\sim}N_q\left(\boldsymbol{\mu}^*,\mathbf{V}\right)\right. \tag{34}$$

where,

$$\boldsymbol{\mu}^*=\left[\mathbf{x}_{i1}'\boldsymbol{\beta}+a_0v_{11},\mathbf{x}_{i2}'\boldsymbol{\beta}+a_0v_{12},...,\mathbf{x}_{iT}'\boldsymbol{\beta}+a_0v_{1T}\right]'$$

Note that, let $\mathbf{y}_{i,T-1}=\left(y_{i2},y_{i3},\ldots,y_{iT}\right)'$. Alternatively, the sample probability density function of $\mathbf{y}_i$ is can be written as:

$$f_s\left(\mathbf{y}_i\left|\mathbf{x}_i\right.\right)=f_s\left(y_{i1}\left|\mathbf{x}_{i1}\right.\right)f_p\left(\mathbf{y}_{i,T-1}\mid y_{i1},\mathbf{x}_i\right) \tag{35}$$

where

$$f_s\left(y_{i1}\left|\mathbf{x}_{i1},\boldsymbol{\theta},\boldsymbol{\gamma}\right.\right)=\frac{1}{\sqrt{2\pi v_{11}}}\exp\left[-\frac{1}{2v_{11}}\left(y_{i1}-\mathbf{x}_{i1}'\boldsymbol{\beta}-a_0v_{11}\right)^2\right] \tag{36}$$

$$f_p\left(\mathbf{y}_{i,T-1}\mid y_{i1},\mathbf{x}_i\right)=\frac{1}{\left(2\pi\right)^{T-1}\left|V_{0,T-1}^*\right|^{1/2}}\exp\left[-\frac{1}{2}\left(\mathbf{y}_{i,T-1}-\boldsymbol{\mu}_{T-1}\right)'\left(V_{0,T-1}^*\right)^{-1}\left(\mathbf{y}_{i,T-1}-\boldsymbol{\mu}_{T-1}\right)\right] \tag{37}$$

$$\begin{aligned}\boldsymbol{\mu}_{T-1}&=E_p\left[\mathbf{y}_{i,T-1}\mid y_{i1},\mathbf{x}_i\right]\\&=\left[\mathbf{x}_{i2}'\boldsymbol{\beta}+\frac{v_{21}^*}{v_{11}^*}\left(y_{i1}-\mathbf{x}_{i1}'\boldsymbol{\beta}\right),...,\mathbf{x}_{iT}'\boldsymbol{\beta}+\frac{v_{T1}^*}{v_{11}^*}\left(y_{i1}-\mathbf{x}_{i1}'\boldsymbol{\beta}\right)\right]'\end{aligned}$$

with general term:

$$v_{tt'}=v_{t+1,t'+1}-\left(v_{11}\right)^{-1}v_{t+1,1}v_{1,t'+1};\ t,t'=1,...,T-1.$$

So that we have the following sample model:

$$\begin{aligned}y_{it}&=\beta_0+\beta_1x_{it1}+...+\beta_px_{itp}+\varepsilon_{it}\\&=\mathbf{x}_i^*\boldsymbol{\beta}^*+\varepsilon_{it};i=1,2,\ldots,n.\end{aligned}, \tag{38}$$

where $\beta_0 = a_0 v_{11}$, $\boldsymbol{\beta}^* = (\beta_0, \beta_1, ..., \beta_p)'$, $\mathbf{x}_i^* = (\mathbf{1}, \mathbf{x}_i)$ and the ε_{it} are a random sequence correlated within subjects .
Note that if $a_0 = 0$, that is the sampling design is noninformative, then the population and sample models are the same.

4.5 Estimation:

We consider three method of estimation, namely: unweighted maximum likelihood, pseudo maximum likelihood, and two-step estimation based on the sample distribution.

4.5.1 Unweighted maximum likelihood:

Maximum likelihood for the case where the sampling design is ignorable: the value of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{V})$ that satisfy:

$$\frac{\partial}{\partial \boldsymbol{\theta}} l_{srs}(\boldsymbol{\mu}, \mathbf{V}) = -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\theta}}\left\{ nq \log 2\pi + n \log|\mathbf{V}| + \sum_{i=1}^{n} (\mathbf{y}_i - \boldsymbol{\mu})' \mathbf{V}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) \right\} \qquad (39)$$

4.5.2 Pseudo maximum likelihood:

The pseudo maximum likelihood estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{V})$ is defined as the solution of:

$$\hat{U}_w(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i \in s} w_i \frac{\partial}{\partial \boldsymbol{\theta}}\left[q \ln 2\pi + \ln|\mathbf{V}| + \{\mathbf{y}_i - (\boldsymbol{\mu} + \mathbf{V}\mathbf{a})\}' \mathbf{V}^{-1} \{\mathbf{y}_i - (\boldsymbol{\mu} + \mathbf{V}\mathbf{a})\} \right] = 0 \quad (40)$$

$$\hat{U}_{ws}(\boldsymbol{\theta}) = \sum_{i \in s} w_i \frac{\partial}{\partial \boldsymbol{\theta}} \ln\{f_s(y_{i1} | \mathbf{x}_{i1})\} + \frac{N}{n}\sum_{i \in s} \frac{\partial}{\partial \boldsymbol{\theta}} \ln\{f_p(\mathbf{y}_{i,T-1} \mid y_{i1}, \mathbf{x}_i)\} = 0 \qquad (41)$$

For more discussion, see Eideh and Nathan (2006).

4.5.3 Two-step method:

Step one. Estimate a_0 via the model: $E_s(w_i \mid y_{i1}, \mathbf{x}_{i1}) = \exp(-a_0^* - a_0 y_{i1} - \mathbf{a}'\mathbf{x}_{i1})$**.**

Step two. Using Theorem 2, the maximum likelihood estimators of $\boldsymbol{\beta}^*$ and $\mathbf{V}$ are defined as the solution of:

$$\frac{\partial}{\partial \boldsymbol{\theta}} l_{rs}(\boldsymbol{\beta}^*, \mathbf{V}) = -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\theta}}\left\{ nq \log 2\pi + n \log|\mathbf{V}| + \sum_{i=1}^{n} (\mathbf{y}_i - \boldsymbol{\mu}^*)' \mathbf{V}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}^*) \right\} = 0 \quad (42)$$

where $\boldsymbol{\beta}^* = (\beta_0, \beta_1, ..., \beta_p)'$ and $\boldsymbol{\mu}^* = [\mathbf{x}_{i1}'\boldsymbol{\beta} + a_0 v_{11}, \mathbf{x}_{i2}'\boldsymbol{\beta} + a_0 v_{12}, ..., \mathbf{x}_{iT}'\boldsymbol{\beta} + a_0 v_{1T}]'$.

4.6 Variance estimation:

For variance estimation we use the inverse of Fisher information matrix and the bootstrap approach for variance, see Pfeffermann and Sverchkov (1999, 2003) and Eideh and Nathan (2006).

(a) Fisher information matrix approach:

The inverse of the observed Fisher information matrix evaluated at $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}\right)$ is given by:

$$\begin{aligned} \hat{V}_s\left(\hat{\boldsymbol{\theta}}\right) &= \left[I_s\left(\hat{\boldsymbol{\theta}}\right)\right]^{-1} \\ &= \left\{ -\frac{1}{n}\left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right\}^{-1} \end{aligned} \tag{43}$$

(b) Bootstrap approach:

Let $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}\right)$ be the sample maximum likelihood estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{V})$ based on any of the equations (39-42) and $\hat{\boldsymbol{\theta}}_b = \left(\hat{\boldsymbol{\beta}}_b, \hat{\mathbf{V}}_b\right)$ be the ML estimator computed from the bootstrap sample $b = 1,..., B$, with the same sample size, drawn by simple random sampling with replacement from the original sample – the sample drawn under informative sampling design. The bootstrap variance estimator of $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}\right)$ is defined as:

$$\hat{V}_{boot}\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\boldsymbol{\theta}}_b - \bar{\hat{\boldsymbol{\theta}}}_{boot}\right)\left(\hat{\boldsymbol{\theta}}_b - \bar{\hat{\boldsymbol{\theta}}}_{boot}\right)' \tag{44}$$

where

$$\bar{\hat{\boldsymbol{\theta}}}_{boot} = \frac{1}{B}\sum_{b}^{B}\hat{\boldsymbol{\theta}}_b \, .$$

5. Conclusions

In this paper, we extend the definition of univariate sample distribution into multivariate random variables. Also, we consider a new method of estimating the parameters of the superpopulation model for analyzing multivariate normal observations from finite population when the sampling design is informative. Furthermore, the general linear model for longitudinal survey data under informative sampling using different covariance structures: the exponential correlation model, the uniform correlation model, and the random effect model, was fitted under informative sampling.

The main feature of the present estimators is their behaviours in terms of the informativeness parameters.

The paper is purely mathematical. The role of informativeness of sampling mechanism in adjusting various estimators for bias reduction, based on simulation

study, under different population models and different modeling of conditional expectations of first order inclusion probabilities given response variable and covariates, can be found in Pfeffermann and Sverchkov (1999, 2003), Nathan and Eideh (2004), Eideh (2008), and Eideh and Nathan (2006, 2009).
I hope that the new mathematical results obtained will encourage further theoretical, empirical and practical research in these directions.

ACKNOWLEDGEMENTS

The author is grateful to the referees for their valuable comments.

REFERENCES

Birnbaum Z.W., Paulson E., and Andrews F.C. (1950). On the Effect of Selection Performed on Some Coordinates of a Multi-Dimensional Population. Psychometrika, 15, pp 191-204.
Chambers, R. and Skinner, C. (2003). *Analysis of Survey Data*. New York: John Wiley.
Diggle, P. J., Liang, K. Y, and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Science Publication.
Eideh A.H. (2008). Estimation and Prediction of Random Effects Models for Longitudinal Survey Data under Informative Sampling. *Statistics in Transition – New Series .Volume 9, Number 3, December 2008,* pp 485 – 502.
Eideh, A. H. and Nathan, G. (2009). Two-Stage Informative Cluster Sampling with application in Small Area Estimation. *Journal of Statistical Planning and Inference.139*, pp 3088-3101.
Eideh, A. H. and Nathan, G. (2006) Fitting Time Series Models for Longitudinal Survey Data under Informative Sampling. *Journal of Statistical Planning and Inference*, 136, 3052-3069.
Fahrmeir, L. and Tutz, G. (2001). *Multivariate statistical modeling based on generalized linear models*, 2nd edn. New York: Springer.
Herriot, R.A., and Kasprzyk, D. (1984). The survey of income and program participation. *American Statistical Association, Proceedings of the Social Statistics Section*, pp.107-116.
Hoem, J.M. (1989). The issue of weights in panel surveys of individual behaviors. In Panel Surveys, (Eds.), Kasprzyk, D., Duncan, G.J., Kalton, G., and Singh, M.P., New York: Wiley, pp. 539-565.
Lohnson, R.A., and Wichern, D.W. (1998). *Applied Multivariate Statistical Analysis, 4th edn*. New Jersey: Prentice Hall.
Kasprzyk, D., Duncan, G.J, Kalton, G., and Singh, M.P. (Eds.) (1989). Panel Surveys. New York: Wiley.
Krieger, A.M, and Pfeffermann, D. (1997) Testing of distribution functions from complex sample surveys. *Journal of Official Statistics*. 13: 123-142.
Mardia, K.V., Kent, T.J. and Bibby, J.M. (1979). *Multivariate analysis*, New York: Academic Press.
Nathan, G. (1999). A review of sample attrition and representativeness in three longitudinal surveys. GSS methodology Series No. 13. London: Office of National Statistics.

Nathan, G. and Eideh, A. H. (2004). L'analyse des données issues des enquêtes longitudinales sous un plan de sondage informatif. in**:** *Échantillonage et Méthodes d'Enquêtes*. (P. Ardilly – Ed.) Paris: Dunod, pp 227-240.

Pfeffermann, D. (1993). The role of sampling weight when modeling survey data. International Statistical Review 61: 317-337.

Pfeffermann, D. (1996). The use of sampling weights for survey data analysis. Statistical Methods in Medical Research, V.5: 239-261.

Pfeffermann, D., Krieger, A. M, and Rinott, Y. (1998). Parametric Distributions of Complex Survey Data under Informative Probability Sampling. *Statistica Sinica*, 8, 1087 1114.

Pfeffermann, D. and Sverchkov, M. (1999). Parametric and Semi-Parametric Estimation of Regression Models Fitted to Survey Data. *Sankhya***,** 61, Ser. B, 66-186.

Pfeffermann, D. and Sverchkov, M. (2003). Fitting Generalized Linear Models under Informative Probability Sampling. *Analysis of Survey Data*. (eds. R. Chambers and C. J. Skinner), pp. 175-195. New York: Wiley**.**

Skinner, C.J., Holt, D., and Smith, T.M.F (Eds.) (1989). Analysis of Complex Surveys, New York: Wiley.

Skinner, C.J., and Holmes, D. (2003). Random Effects Models for Longitudinal Data. *Analysis of Survey Data***.** (eds. R. Chambers and C. Skinner), pp. 175-195. New York: Wiley**.**

Smith, T.M.F. and Holmes, D.J. (1989). Multivariate analysis. In *Analysis of Complex Surve*ys, eds. C.J. Skinner, D. Holt and T.M.F. Smith, New York: Wiley, pp. 165-190.