The Philippine Statistician, 2005
Vol. 54, Nos. 1-4, pp. 47-62

Loglinear-Logit Models With Covariates For The Changes of Categorical Variables[1]

Emeterio S. Solivas[2]

Received July, 2005; Revised November, 2005

ABSTRACT

This paper considers some formulations of the logit models with covariates to analyze changes in categorical variables observed in repeated measurements. The objective is to illustrate how existing techniques for the analysis of loglinear-logit models can be extended in an analysis of covariance (ANACOVA) type of analysis to study processes of change of categorical response variables. The framework begins with the formulation of response functions and linear modeling using the weighted least squares (WLS) method of parameter estimation and inference . The general approach is then applied to loglinear-logit modeling and specifically illustrated on a 2x2 contingency table. Some loglinear-logit models are then formulated to include a fixed covariate to be analyzed by the WLS approach. Finally, the technique is considered for the study of changes in categorical response variable. The techniques are then applied with the analysis of an illustrative data.

Keywords: response function, design matrix, logit formulation, reparametrization, persistence and symmetry coefficients.

1. INTRODUCTION

While the classical log-linear modeling is used to explain association among discrete or categorical variables where the data are cross-tabulated in contingency table, the logit modeling is used when one of the variables is considered as response variable and the others as the causal variables. Oftentimes, some of the causal variables are continuous and others are discrete. When the researcher is interested on the effects these mixed causal variables on the responses, reformulation of the log-linear logit model to incorporate some continuous and discrete covariates needs to be specified.

The objectives of this paper are: to find how loglinear-logit models can be reformulated to include parameters which measure the main and interaction effects of factors and covariates; to apply the general linear models approach of estimation and inference to the reformulated models; to extend the procedure to study changes of categorical responses in repeated measurements, and to illustrate how the analyses are implemented.

[1] Earlier version of this paper was presented during the plenary session of the First Joint Student – Faculty Conference of the Statistical Sciences at the CAS, UPLB, 27 September 2003.

[2] Associate Professor of Statistics. Institute of Statistics, U.P. Los Baños, College, Laguna. Email: ess@instat.uplb.edu.ph and terrysolivas@yahoo.com .

II. THEORETICAL FRAMEWORK

2.1. Preliminaries: The standard loglinear modeling of contingency tables.

Given two categorical variables Y_1 and Y_2. Consider Y_1 to represent set of *explanatory variables* which give a set of r categories denoted as *subpopulations* and Y_2 to represent set of *response variables* which give a set of c categories denoted as *response levels* The bivariate probability distribution of Y_1 and Y_2 is summarized in the following two-way table:

Table 2.1. Density table of a two categorical variables Y_1 and Y_2.

Subpopulation	Y_2 *(response levels)*				
Y_1	1	2	...	c	Total
1	π_{11}	π_{12}		π_{1c}	$\pi_{1.}$
2	π_{21}	π_{22}		π_{2c}	$\pi_{2.}$
:					
r	π_{r1}	π_{r2}		π_{rc}	$\pi_{r.}$
Total	$\pi_{.1}$	$\pi_{.2}$		$\pi_{.c}$	*1.00*

where: π_{ij} = probability of the jth response of subpopulation i. Let the vectorized π_{ij}'s be an (sx1) vector as $\pi' = (\pi_1\ \pi_2 \dots \pi_s)$, where $s = r$ x c. The objective of the analysis is to model the relationships among functions of the response probabilities π_{ij}'s, say $\gamma(\pi)$. These functions may be given by a set of p linear equations:

$$\gamma_1 = a_{11}\pi_1 + a_{12}\pi_2 + \dots + a_{1s}\pi_s$$
$$\gamma_2 = a_{21}\pi_1 + a_{22}\pi_2 + \dots + a_{2s}\pi_s$$
$$\vdots$$
$$\gamma_p = a_{p1}\pi_1 + a_{p2}\pi_2 + \dots + a_{ps}\pi_s$$

In matrix notation , $\gamma = \gamma(\pi) = \mathbf{A}\pi$, where $\mathbf{A}$ is the pxs matrix of linear coefficients and the rank of A is p.

For illustration, consider a 3x2 table. The vector of response probabilities is

$$\pi' = [\pi_{11}\ \ \pi_{12}\ \ \pi_{21}\ \ \pi_{22}\ \ \pi_{31}\ \ \pi_{32}]$$

and if $\gamma(\pi)$ is formulated as:

$$\gamma_1 = \pi_{11} - \pi_{12};\quad \gamma_2 = \pi_{21} - \pi_{22};\quad \gamma_3 = \pi_{31} - \pi_{32}$$

then the matrix A is

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

2.1.1. The weighted least squares method.

The response probability functions $\gamma(\pi)$ are expressed in a linear model,

$$\gamma(\pi) = \mathbf{Z}\beta$$

where **Z** is the px q design matrix and β is a qx1 vector of parameters be estimated.

Considering $\gamma(\pi)$ as a function of the true unknown proportions, then $\gamma(\pi)$ is estimated from the sample contingency table. An exercise in statistical theory can show that the maximum likelihood estimate of π_{ij} is p_{ij},

$$p_{ij} = n_{ij}/n_{i.}$$

where the variances and covariances are given as

$$V(p_{ij}) = \pi_{ij}(1 - \pi_{ij})/n_{i.}, \qquad i = 1,\ldots, r \quad j = 1, \ldots, c$$

$$Cov(p_{ij},p_{ik}) = -\pi_{ij}\pi_{ik}/n_{i.}, \qquad j \neq k$$

$$Cov(p_{ij},p_{lk}) = 0 \qquad i \neq j \neq k \neq l.$$

Define the vectorized p_{ij} as $\mathbf{p}' = [p_1 \ p_2 \ \ldots \ p_s]$. The covariance matrix of **p** is (sxs) block diagonal denoted as $V(\mathbf{p}) = \Sigma$. For a 2x2 table,

$$\Sigma = \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix}$$

where the ith block diagonal submatrix, the covariance matrix for the ith row, is

$$\Sigma_i = \begin{bmatrix} \pi_{i1}(1 - \pi_{i1})/n_{i.} & -\pi_{i1}\pi_{i2}/n_{i.} \\ -\pi_{i2}\pi_{i1}/n_{i.} & \pi_{i2}(1 - \pi_{i2})/n_{i.} \end{bmatrix}$$

The estimate of Σ, denoted as **S**, is obtained by replacing π_{ij} by the corresponding p_{ij}. Also, replacing π_{ij} by the corresponding p_{ij} in the response function gives the sample response function denoted as $\mathbf{g} = \mathbf{g}(\mathbf{p})$.

Defining the (px1) error vector $\varepsilon = \mathbf{g} - \gamma$ gives the linear model

$$\mathbf{g} = \mathbf{Z}\beta + \varepsilon$$

It is assumed that $E(\varepsilon) = \mathbf{0}$ and that the covariance matrix of ε is $E(\varepsilon\varepsilon') = \mathbf{A}\Sigma\mathbf{A}'$ which is estimated by $\mathbf{V} = \mathbf{ASA}'$. This implies that the estimate of $V(\mathbf{g}) = \mathbf{V} = \mathbf{ASA}'$.

From the theory of linear models, the weighted least squares (WLS) estimator of β, denoted as **b**, is given by

$$\mathbf{b} = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{g}$$

and the estimate of the covariance matrix of **b** is

$$\hat{\mathbf{V}}(\mathbf{b}) = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}$$

A test of the goodness-of-fit of the linear model $\mathbf{g} = \mathbf{Z}\beta + \varepsilon$ is done by the test statistic

$$\chi^2 = \mathbf{g}'\mathbf{V}^{-1}\mathbf{g} - \mathbf{b}'(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{b}'$$

which, from linear model theory, has a large sample chi-square distribution with $(p\text{-}q)$ degrees of freedom under the null hypothesis.

A test about a linear combination of β, say Ho: $\mathbf{C}\beta = \mathbf{0}$, may be done by the test statistic

$$\chi^2 = (\mathbf{Cb})'[\mathbf{C}'(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})\mathbf{C}]^{-1}(\mathbf{Cb}), \text{ where } \mathbf{C} \text{ is } (m\text{x}q) \text{ of rank } m.$$

From linear model theory, this test under the null hypothesis, also has a large sample chi-square distribution with m degrees of freedom.

2.1.2. Loglinear modeling by the weighted least squares (WLS) method.

The WLS approach can be used for loglinear modeling where the logarithms of response proportions, denoted by the vector ln(**p**), are used in the model. For this, the estimate of the covariance matrix of ln(**p**) is (in Jobson, 1992, p 102),

$$\hat{V}(\ln(\mathbf{p})) = \mathbf{S}^* = \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1}$$

where **D** is a diagonal matrix with elements of **p** on the diagonal.

A logit response model is obtained by the linear transformation of ln(**p**). Denote again the matrix of linear transformation by **A**, then the logit responses as

$$\mathbf{g} = \mathbf{A}\ln(\mathbf{p})$$

When **A** = **I**, the logit model is the usual loglinear model. As before we express the model

$$\mathbf{g} = \mathbf{Z}\beta + \varepsilon.$$

The WLS estimation and hypothesis testing procedures follow as above where **S*** is used.

2.1.4. The loglinear model for a 2x2 contingency table.

For simple illustration, the 2x2 contingency table is used. Consider the following table:

Subpopulation	Y_2 *(response variable)*		
Y_1	1	2	Total
1	π_{11}	π_{12}	$\pi_{1.}$
2	π_{21}	π_{22}	$\pi_{2.}$
Total	$\pi_{.1}$	$\pi_{.2}$	*1.00*

Using the last-category-set-to zero restriction, the saturated linear model $\gamma = \mathbf{Z}\beta$ is

$$\gamma_1 = \ln(\pi_{11}) = \mu + \mu_{1(1)} + \mu_{2(1)} + \mu_{12(11)}$$

$$\gamma_2 = \ln(\pi_{12}) = \mu + \mu_{1(1)}$$

$$\gamma_3 = \ln(\pi_{21}) = \mu \qquad + \mu_{2(1)}$$

$$\gamma_4 = \ln(\pi_{22}) = \mu$$

In matrix notation, the components γ, $\mathbf{Z}$, and β are given as

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix} \qquad \mathbf{Z} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} \mu \\ \mu_{1(1)} \\ \mu_{2(1)} \\ \mu_{12(11)} \end{pmatrix}$$

In general: $\gamma_i = \mu + \mu_{1(1)}Z_{i1} + \mu_{2(1)}Z_{i2} + \mu_{12(11)}Z_{i1}Z_{2i2}$

where Z_{i1} and Z_{i2} are the *i*th element of the second and third columns of **Z**, respectively.

2.2. The 2x2 contingency table loglinear model with covariate.

Suppose a covariate X is considered to have an association with Y_1 and Y_2. In the contingency table, enter into the *ij*th cell the mean of X observations on that cell. For a 2x2 table, the sample frequencies and covariate means may be summarized as follows:

Table 2.2. Contingency table with covariate

Subpopulation	Y_2 *(response level)*		
Y_1	*1*	2	Total
1	n_{11} X_{11}	n_{12} X_{12}	$n_{1.}$
2	n_{21} X_{21}	n_{22} X_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	N

Let F_{ij} be the expected frequency of the *ij*th cell. For a 2x2 table, the saturated loglinear linear model $\gamma(\ln(\mathbf{F})) = \mathbf{W}\beta$ that incorporates the covariate X can be formulated in a variety of models. This paper illustrates only one of many possible alternative model formulations. Let us denote this as our model A.

Model with Covariate: Model A.

Using the last-category-set-to-zero constraint, a proposed model, referred as ***Model A*** is:

$$\gamma_1 = \ln(F_{11}) = \mu + \mu_{1(1)} + \beta_1 X_{11}$$

$$\gamma_2 = \ln(F_{12}) = \mu + \mu_{1(1)}$$

$$\gamma_3 = \ln(F_{21}) = \mu \qquad + \beta_2 X_{21}$$

$$\gamma_4 = \ln(F_{22}) = \mu$$

where β_1 is the interaction effect of covariate X and $Y_1 = 1$ on the response function γ_i and β_2 is the interaction effect of covariate X and $Y_1 = 2$. In matrix notation, the matrices **W** and β are

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & X_{11} & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & X_{21} \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu \\ \mu_{1(1)} \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

In general,

$$\gamma_i = \mu + \mu_{1(1)} Z_i + \beta_1 X_{1i} + \beta_2 X_{2i}$$

where Z_i, X_{1i} and X_{2i} are the *i*th element of the second, third and last columns of **W**, respectively.

2.3. The repeated measurement loglinear model with covariate to measure changes.

We now construct the loglinear model with a covariate X to measure changes in the response variable in repeated measurement data. Consider that a categorical response variable Y with c categories or classification are observed at two occasions from a sample of n individuals. Let Y_1 be the observation on variable Y on the first occasion and Y_2 at the second occasion. When the data on Y_1 and Y_2 of the n individuals are cross-tabulated we have a special square c x c contingency table data. Let n_{ij} be the observed frequency in the cell (i,j) where $i = 1, 2, ..., c$ and $j = 1, 2, ..., c$.

Table 2.3. Repeated measurement contingency table with covariate.

Occasion 1	Y_2 (Occasion 2)		
Y_1	1	2	Total
1	n_{11} X_{11}	n_{12} X_{12}	$n_{1.}$ $X_{1.}$
2	n_{21} X_{21}	n_{22} X_{22}	$n_{2.}$ $X_{2.}$
Total	$n_{.1}$ $X_{.1}$	$n_{.2}$ $X_{.2}$	n $X_{..}$

2.3.1. Some logit formulations to measure changes in the 2x2 contingency table.

Let F_{ij} be the expected frequency in the ijth cell of the table.

1. $\ln(F_{11}/F_{12})$ is the logit of remaining on level 1 instead of changing to level 2 of Y.
2. $\ln(F_{21}/F_{22})$ is the logit of changing to level 1 instead of remaining to level 2 of Y.
3. $\ln(F_{11}/F_{22})$ is the logit of those remaining on level 1 compared to those remaining on level 2 of Y.
4. $\ln(F_{21}/F_{12})$ is the logit of those changing from level 2 to level 1 compared to those changing from level 1 to level 2 of Y.

For the proposed model, these logits are also functions of the covariate X. The functions are summarized as follows:

Table 2.4. Some logits as function of covariate X.

Logit	Model A
$\ln(F_{11}/F_{12})$	$\beta_1 X_{11}$
$\ln(F_{21}/F_{22})$	$-\beta_2 X_{21}$
$\ln(F_{11}/F_{22})$	$\mu_{1(1)} + \beta_1 X_{11}$
$\ln(F_{21}/F_{12})$	$-\mu_{1(1)} + \beta_2 X_{21}$

2.3.2. The model for changes in the 2x2 contingency table with covariate.

Looking at the cells in the 2x2 contingency table, we see that cells (1,1) and (2,2) denote *persistence* – (1,1) represents persistence of level 1 and (2,2) represents persistence of level 2. We also see that cells (1,2) and (2,1) denote *change* – (1,2) represents change from level 1 to level 2, and (2,1) represents change from level 2 to level 1. Thus, the magnitude of n_{12} and n_{21} tell us the change that occurred. The changes over occasions are related to n_{12} and n_{21} and the persistence are related to n_{11} and n_{22}. The situation $n_{12} = n_{21}$ denotes *symmetry* (S) and the situation $n_{11} = n_{22}$ denotes *persistence* (P). Let F_{ij} be the expected frequency of the ijth cell. Then to study changes in a 2x2 table we have to investigate the relationship between F_{12} *and* F_{21} and between F_{11} and F_{22}.

We formulate the model

$$\ln(n_{ij}) = \mu + \mu_P Z^P_{ij} + \mu_S Z^S_{ij} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \varepsilon_{ij}$$

where Z^P_{ij}, and Z^S_{ij} are indicator variables (with values 0 or 1) and the superscripts and superscript S and P denote symmetry and persistence, respectively.

The model design matrix **W** is constructed as follows:

Cell(I,j)	Design matrix				
	1	Z^P_{ij}	Z^S_{ij}	X_{ij}	X_{ij}
(1,1)	1	1	0	X_{11}	0
(1,2)	1	0	-1	0	0
(2,1)	1	1	0	0	X_{21}
(2,2)	1	0	-1	0	0

The model for each cell log frequency is

$$\ln(F_{11}) = \mu + \mu_P \qquad + \beta_1 X_{11}$$
$$\ln(F_{12}) = \mu \qquad - \mu_S$$
$$\ln(F_{21}) = \mu \qquad + \mu_S \qquad + \beta_2 X_{21}$$
$$\ln(F_{22}) = \mu - \mu_P$$

Defining the logits $\ln(F_{21}/F_{12})$ and $\ln(F_{21}/F_{12})$, we have

$$\ln(F_{11}/F_{22}) = 2\mu_P + \beta_1 X_{11} \qquad \ln(F_{21}/F_{12}) = 2\mu_S + \beta_2 X_{21}$$

this gives, from Table 2.4

$$\mu_P = \mu_{1(1)}/2 \quad \text{and} \quad \mu_S = -\mu_{1(1)}/2.$$

III. DATA AND METHODS

3.1. Data used in the study

The data used to illustrate the techniques described earlier were part of the data set gathered and used by Pacia and Prosuelo (1998) in their Special Problem (Stat 190). The data set were obtained from a random sample of 200 UPLB employees. A questionnaire was administered to the sample employees on the second week of January 1998 and then the same questionnaire was administered again to the sample employees on the first week of March.

The data were about the preference of the sample employees on the leading presidential aspirants during the first week of January 1998. They were De Villa, Lim, Roco and Others (consisting of De Venecia, Estrada and Santiago). The preferences were summarized in the following table.

Table 3.1. Preference and transition rates of presidentiables. Jan - March, 1998.

January	March				Total
	1. De Villa	2. Lim	3. Roco	4. Others	
1. De Villa	19 (51.4)	7 (18.9)	8 (21.6)	3 (8.1)	37 (18.5)
2. Lim	6 (8.5)	51 (71.8)	12 (16.9)	2 (2.8)	71 (35.5)
3. Roco	2 (3.6)	6 (10.7)	47 (83.9)	1 (1.8)	56 (28.0)
4. Others	1 (2.8)	2 (5.6)	10 (27.8)	23 (63.9)	36 (18.0)
Total	28 (14.0)	66 (33.0)	77 (38.5)	29 (14.5)	200 (100)

L^2 = 177.27 with p-value = 0.000 Cramer's V = 0.60 with p-value = 0.000

To illustrate the analysis of 2 x 2 contingency table, the categories (presidentiables) were dichotomized. The resulting four sub-tables were given in Tables 3.2a – 3.2d.

Table 3.1a. Frequency of preferences for De Villa, January-March, 1998.

January	March		Total
	Yes	No	
Yes	19 (43.8)	18 (39.1)	37
No	9 (38.6)	154 (40.1)	163
Total	28	172	200

Figures in parentheses are the transition rates. $L^2 = 41.09$ with p-value = 0.000

Table 3.2b. Frequency of preferences for Lim, January-March, 1998.

January	March		Total
	Yes	No	
Yes	51 (40.3)	20 (42.1)	71
No	15 (39.3)	114 (40.1)	129
Total	66	134	200

Figures in parentheses are the transition rates. $L^2 = 76.51$ with p-value = 0.000

Table 3.2c. Frequency of preferences for Roco, January-March, 1998.

January	March		Total
	Yes	No	
Yes	47 (38.8)	9 (43.0)	56
No	30 (39.4)	114 (40.9)	144
Total	77	123	200

Figures in parentheses are the transition rates. $L^2 = 69.83$ with p-value p = 0.000

Table 3.2d. Frequency of preferences for *Others*, January-March, 1998.

January	March		Total
	Yes	No	
Yes	23 (41.7)	13 (35.2)	36
No	6 (44.8)	158 (40.3)	164
Total	29	171	200

Figures in parentheses are the transition rates. $L^2 = 67.01$ with p-value = 0.000

IV. RESULTS AND DISCUSSION

The test of independence on each of Tables 3.2a – 3.2d showed significant relationship between January and March preferences. Next step was to model these relationships. The saturated loglinear model without covariate was first fitted on each 2x2 sub-table. The estimates of the parameters were summarized as follows:

Table 4.0. Estimates of the parameters of the saturated model without covariate

2x2 sub-table	Estimate of parameter			
	μ	$\mu_{1(1)}$	$\mu_{2(1)}$	$\mu_{12(11)}$
1. De Villa	5.037	-2.147	-2.840	2.894
2. Lim	4.736	-1.741	-2.028	2.964
3. Roco	4.736	-2.539	-1.335	2.988
4. Others	5.063	-2.498	-3.271	3.841

All estimates are significant at higher than 1% level of significance.

4.1 On the analysis of 2 x 2 contingency table using Model A.

The loglinear model with covariate was then fitted on each sub-table to obtain estimates of the parameters. The results were summarized in the table that follows.

Table 4.1.1. Estimates of the parameters of change in the January-March preferences.

2x2 sub-table	Estimate of parameter			
	μ	$\mu_{1(1)}$	β_1	β_2
1. De Villa	5.037	-2.147	0.0012^{ns}	-0.0737
2. Lim	4.736	-1.741	0.0232	-0.0516
3. Roco	4.736	-2.539	0.0426	-0.0339
4. Others	5.063	-2.498	0.0137^{ns}	-0.0730

Estimates are significant at higher than 1% level of significance except those with ns.

From Table 4.1.1, estimates of the log frequencies $\ln(F_{ij})$ were obtained using equations (2.2.1). These estimates are for:

$\ln(F_{11})$ - the log frequency of voters choosing the particular candidate in January and to choose this candidate again in March,

$\ln(F_{12})$ - the log frequency of voters choosing the particular candidate in January but to choose another candidate in March.

The results for each candidate are summarized as follows:

Table 4.1.2. Estimates of the $\ln(F_{11})$ and $\ln(F_{12})$ as a function of age.

2x2 sub-table	Estimate of the	
	$\text{Ln}(F_{11})$	$\ln(F_{12})$
1. De Villa	$2.890 + 0.0012X_{11}$	2.890
2. Lim	$2.996 + 0.0232X_{11}$	2.996
3. Roco	$2.197 + 0.0426X_{11}$	2.197
4. Others	$2.565 + 0.0137X_{11}$	2.565

The estimate of $\ln(F_{11})$ is a direct function of age. Among those voters who preferred a particular candidate in January and again in March, the increased preferential response to increasing age is highest for Roco, followed by Lim and the lowest by De Villa. These comparisons are shown in the following graph.

Also from Table 4.1.1, estimates of other log frequencies $\ln(F_{ij})$ were obtained using equations (2.2.1). These estimates are for:

$\ln(F_{21})$ - the log frequency of voters not choosing a particular candidate in January to choose this candidate in March,

$\ln(F_{22})$ - the log frequency of voters choosing a particular candidate in January and still not choosing the same candidate in March.

The results for each candidate are summarized as follows:

Table 4.1.3. Estimates of the $\ln(F_{12})$ and $\ln(F_{21})$ as a function of age.

2x2 sub-table	Estimate of the	
	$\text{Ln}(F_{21})$	$\ln(F_{22})$
1. De Villa	$5.037 - 0.0737X_{21}$	5.037
2. Lim	$4.736 - 0.0516X_{21}$	4.736
3. Roco	$4.736 - 0.0339X_{21}$	4.736
4. Others	$5.063 - 0.0730X_{21}$	5.063

Among those voters who did not prefer a particular candidate in January but preferred this candidate in March, the preferential changes are negative functions of age. Decreased preferential response to increasing age is highest for De Villa, followed by Others and the lowest is for Roco. These comparisons are shown in the following graph:

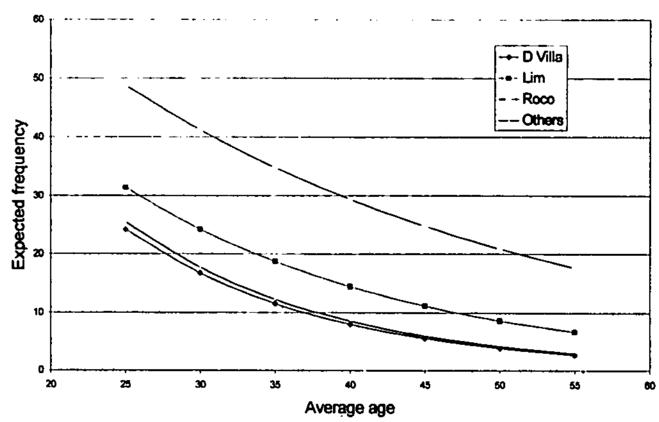

Fig. 2. Estimates of the expected frequencies F12 using model A.

Expressing the preferences in terms of logits, estimates of the logits $\ln(F_{12}/F_{21})$ and $\ln(F_{12}/F_{21})$ are computed and the results were summarized in the next table.

Table 4.1.4. Estimates of the logits $\ln(F_{12}/F_{21})$ and $\ln(F_{12}/F_{21})$.

2x2 sub-table	$\ln(F_{11}/F_{12})$	$\ln(F_{21}/F_{22})$
1. De Villa	$0.0012X_{11}$	$-0.0737X_{21}$
2. Lim	$0.0232X_{11}$	$-0.0516X_{21}$
3. Roco	$0.0426X_{11}$	$-0.0339X_{21}$
4. Others	$0.0137X_{11}$	$-0.0730X_{21}$

The logit $\ln(F_{11}/F_{12})$ indicates the log-odd of an individual who preferred a candidate to stay with that candidate instead of changing to another candidate.

- A positive logit indicates that there are more of those staying with the candidate than those changing to another, and the larger the positive value indicates more staying.
- A negative logit indicates that there are less of those staying with the candidate than those changing to another, and the larger the positive value indicates more staying.
- A unitary logit indicates that individuals staying with the candidate and those changing to another are equal.

The logit $\ln(F_{21}/F_{22})$ indicates the log-odd of an individual who did not preferred a candidate to change preference to that candidate instead of staying with other candidates.

- A positive logit indicates that there are more individuals changing to the candidate than those staying with another; the larger the absolute value indicates more changings.
- A negative logit indicates that there are less of those changing to the candidate than those staying with another; the larger the absolute value indicates more staying.
- A unitary logit indicates that the number of individuals changing to the candidate and those staying with another are equal.

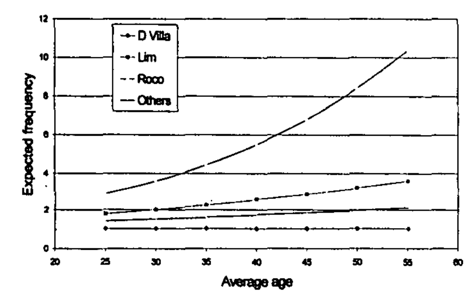

Fig. 3. Estimates of expected frequencies F11/F12 by model A.

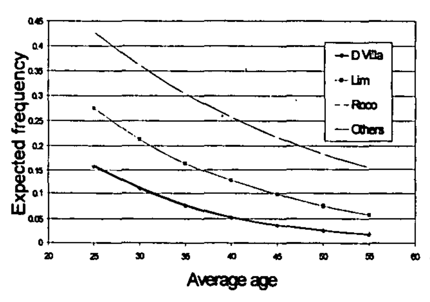

Fig. 4. Estimates of expected frequencies F21/F22: model A

Logits that compare persistence and changes are formulated. Again, from the estimates of the parameters in Table 4.1.1, estimates of the logits $\ln(F_{11}/F_{22})$ and $\ln(F_{21}/F_{12})$ were computed. The results were summarized in the next table.

Table 4.1.5. Estimates of the logits $\ln(F_{11}/F_{22})$ and $\ln(F_{21}/F_{12})$.

2x2 sub-table	$\ln(F_{11}/F_{22})$	$\text{Ln}(F_{21}/F_{12})$
1. De Villa	$-2.147 + 0.0012X_{11}$	$2.147 - 0.0737X_{21}$
2. Lim	$-1.741 + 0.0232X_{11}$	$1.741 - 0.0516X_{21}$
3. Roco	$-2.539 + 0.0426X_{11}$	$2.539 - 0.0339X_{21}$
4. Others	$-2.498 + 0.0137X_{11}$	$2.498 - 0.0730X_{21}$

The logit $\ln(F_{11}/F_{22})$ indicates the log-odd of the individuals who persisted on preferring a particular candidate compared to individuals who persisted not preferring this candidate.

- A positive logit indicates that there are more of those who persisted preferring than those who persisted not preferring the candidate.
- A negative logit indicates that there are more of those who persisted not preferring than those who persisted preferring the candidate.
- A unitary logit indicates equal persistence of those who persisted preferring and those who persisted not preferring the candidate.

The logit $\ln(F_{21}/F_{12})$ indicates the log-odd of the individuals who are at first against a candidate but later changed favor for the candidate compared to individuals who are at first in favor of the candidate but later changed favor against the candidate.

- A positive logit indicates that there are more of those who persisted preferring than those who persisted not preferring the candidate.
- A negative logit indicates that there are less of those who persisted preferring than those who persisted not preferring the candidate.
- A unitary logit indicates that those who persisted preferring and those who persisted not preferring the candidate are equal.

The comparisons of these logits are clearly illustrated in the graphs below:

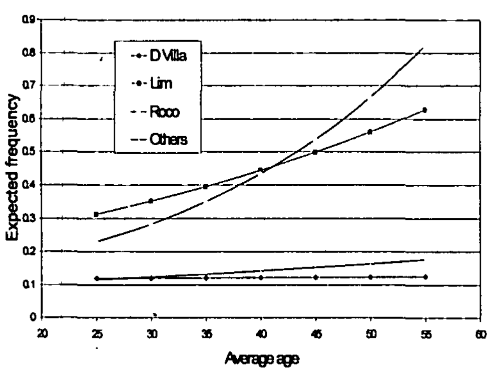

Fig. 5. Estimates of the frequencies F11/F22 using model A

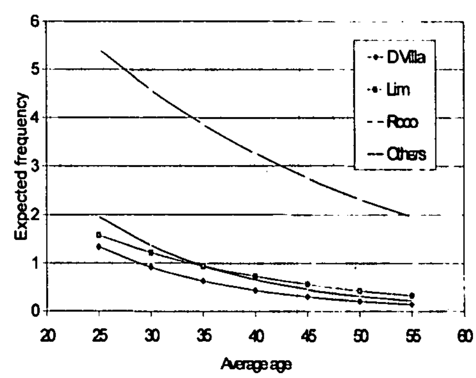

Fig. 6. Estimates of the frequencies F21/F12 using model A

Another measure of the changes that takes place in the table, coefficients of persistence and symmetry were likewise estimated from the parameter estimates in Table 4.1.1. The results were summarized in the next table.

Table 4.1.6. Estimates of the persistence (μ_P) and symmetry (μ_S) parameters of the January- March preferences.

2x2 sub-table	Estimate of	
	μ_P	μ_S
1. De Villa	-1.074	1.074
2. Lim	-0.871	0.871
3. Roco	-1.270	1.270
4. Others	-1.249	1.249

1. The $\mu_P = -\beta_1 X_{11}/2$ indicates equal persistence between the levels of preference.
   $\mu_P > -\beta_1 X_{11}/2$ indicates more voters who remained preferring a candidate than voters who remained not preferring the candidate.
   $\mu_P < -\beta_1 X_{11}/2$ indicates more voters who remained not preferring a candidate than voters who remained preferring the candidate.

2. The $\mu_S = -\beta_2 X_{21}/2$ indicates that changes from choosing to not choosing a candidate and vice versa cancel out.
   $\mu_S > -\beta_2 X_{21}/2$ indicates more voters changing from first not choosing a candidate then later choosing the candidate than vice versa.
   $\mu_S < -\beta_2 X_{21}/2$ indicates less voters changing from first not choosing a candidate then later choosing the candidate than vice versa.

V. SUMMARY AND RECOMMENDATION

The technique illustrated in this paper is very much applicable in many studies. It may be used in political research to measure changes of voting preferences. For market research, the technique may be used for measuring brand loyalty, TV program loyalty. For development research, it may be used to measure growth and progress. Many more.

The technique illustrated is only a starting point. It has numerous natural extensions. The immediate exercise that may be done is to work out the technique on a c x c contingency table using the same data in this paper. The procedure may be extended to the case of multi-wave panel contingency table. The technique may also be extended to incorporate a mixture of continuous and discrete covariates. If the categories of the response and explanatory variables are ordinal, the procedure may be extended to exploit the ordinality of the variables. The technique may also be extended to incorporate covariates that vary during the several periods of measurement. More complicated technique may be developed to handle combinations of the natural extensions just mentioned.

ACKNOWLEDGMENT

This author wishes to thank the editors of the Philippine Statistician and the anonymous referee for the helpful comments and corrections. All remaining errors are the author's responsibility.

References

AGRESTI, A. 1990. *Categorical Data Analysis.* Wiley and Sons, Inc.

JOBSON, J.D. 1992. *Applied Multivariate Data Analysis. II. Categorical and Multivariate Methods.* Springer-Verlag. New York.

NERI, Liza. 1992. *Log-linear Analysis of Categorical Data. Special Problem* . UPLB

PACIA, A. S. and B.B. PORSUELO. 1998. *Log-linear Modeling: An Approach to the Analysis of Changes in Nominal Variables.* Special Problem. UPLB.

SOLIVAS, E.S. 1998. *Statistical Models for the Changes of Preferential Variables: I. Dichotomous Classification.* The Philippine Statistician, 47:1-7, pp 77-88.