The Philippine Statistician, 1998
Vol. 47, Nos. 1-4, pp.19-33

Identifying Correlates Of Poverty In Complex Survey Data

Zita V.J. Albacea and Ann Inez N. Gironella.[1]

ABSTRACT

This study identifies some correlates of poverty using the 1994 Family Income and Expenditure Survey data (FIES) of the Philippines. The association of some selected variables was tested against the poverty status of a household using chi-square test of independence corrected according to the sampling design of FIES. The educational attainment of the household head was found to be associated with poverty status. Likewise, at the household level, whether agriculture is the main source of income of the household; total number of household members and presence of members aged between 1 to 6 years old in the household were also declared to be associated with poverty status. The computed values of the chi-square test statistic were found to change depending on the design effects. On the average, the corrected chi-square value based on first order correction factor $\bar{d}.$, was reduced by as much as 1.72% of the uncorrected value.

Keywords: Poverty; Complex Survey; Chi-square Test of Independence

1. Introduction

Poverty is commonly defined as a form of deprivation. Specifically, the United States Panel on Poverty and Family Assistance (1995) defines poverty as an *economic* deprivation. It pertains to people's lack of economic resources like income for consumption of economic goods and services. There are other forms of personal deprivation like psychological deprivation or lack of self-esteem. These forms of deprivation which a household member may suffer may be related to his household's economic status. For example, one who loses his job or who has never been successful in finding one may suffer a deprivation of both income and psychic esteem.

Measures of poverty commonly used are income-based; that is, household income is used as an indicator of a household's economic resources. However, such measures are considered inadequate in fully describing the conditions of households who are 'in poverty'. Studies of Orbeta and Hilario (1995) and Reyes, et al. (1996) on Minimum Basic Needs (MBN) indicators as poverty measures, among others, aimed to identify characteristics of the household that are related to its poverty status which are not income-based.

In the Philippines, official poverty measures are mainly based on data from the Family Income and Expenditure Survey (FIES) conducted by the National Statistics Office every three years since 1985. Up until the 1994 round (a new sample design was in effect in the 1997 round) every time the survey was conducted about 25,000 sample households were interviewed. This sample of households is deemed sufficient to provide reliable estimates of income and expenditure levels for each province, each key city, and each key municipality of the country.

The 1994 FIES used a stratified two-stage cluster sampling design. The urban and rural areas of each province are the principal domains for the survey. Barangays, which are

[1] Assistant Professor and Professor of Statistics, respectively, Institute of Statistics, College of Arts and Sciences, University of the Philippines Los Baños.

classified as either urban or rural, are the primary sampling units (PSUs) and the households within each sample barangay comprise the secondary sampling units (SSUs). Survey data items include characteristics of the household head and household characteristics that can be and have been correlated to the household's poverty status.

The National Statistical Coordination Board (NSCB) in its publication 'Philippine Poverty Statistics' (December, 1996) defined *poor households* as those whose income are not adequate to meet the poverty threshold. Poverty threshold refers to the basic food and non-food requirement (valued in peso) of a household. The methodology used to measure poverty and other official poverty statistics are well described in several papers, like Virola (1994), Marquez and Virola (1995), and Virola and David (1995).

Marquez and Virola (1995) present an updated profile of the Philippine poor based on 1985, 1988, 1991, and 1994 FIES. In their study, poor households are characterized with respect to age, sex and occupation of the household head, source of income, expenditure and consumption pattern, payment of taxes, type and tenure of housing, source of water supply, kind of toilet facilities, availability of electricity, ownership of household appliances, and involvement in entrepreneurial activities. The study concludes that the following characteristics of the poor have not changed significantly since 1985:

- Poor households have an average of 6 members, one more than the average size of non-poor households.
- Heads of poor households are employed and are found in the agriculture, fishing and forestry sector.

Bautista (1991) used discriminant analysis to differentiate families below and above the reported national poverty threshold. She concludes that families below the poverty threshold or poor families have large family sizes and the household heads are agricultural workers with low educational attainment. Cosme (1998) also applied discriminant analysis to the 1994 FIES data to identify non-income-based discriminating variables in classifying poor and non-poor families. The discriminating variables identified are age and highest educational attainment of the household head, employment status of the household head and his wife, total number of household members who were employed, whether agriculture is the household's main source of income or not, and clothing, education, household and medical expenditures. Noriega (1995) used the results of the 1994 Poverty Mapping Survey conducted by the Department of Social Welfare and Development to identify variables describing the well-being of households for classifying poor from the non-poor households. A poverty classification scheme based on the social, economic and nutrition needs of the households is developed.

Orbeta and Hilario (1995), as cited in Reyes, et al. (1996) reviewed the relationship between the Minimum Basic Needs (MBN) indicators and income groupings using the results of the 1992 Socio-Economic Survey of Special Group of Families. The study concludes that income is a good indicator of deprivation except that of enabling needs of the bottom 30% of the population. Income adequately captures the incidence of deprivation in terms of survival and security needs. Reyes, et al. (1996) extended this study to look at the relationship of income and MBN indicators on the population at large utilizing data from the 1991 FIES.

The studies reviewed above had the common primary objective of determining indicators of poverty. However, in some of the studies mentioned particularly those that used statistical inference techniques, the sampling design of the survey was ignored proceeding with the analysis of the data as though they were obtained by simple random sampling. Nathan (1975) studied the test of independence in contingency tables from stratified proportional samples. Although the results of the study showed that the usual Pearson chi-square test can be used in data from stratified proportional sampling, there is still evidence of loss in power of the test when the sampling design is ignored in the analysis. Holt, et al. (1980) studied the effect of the sampling design on the behavior of usual Pearson chi-square tests for goodness of fit, homogeneity and independence. Specifically, the results of the study suggest that the corrections work well for test of goodness of fit and homogeneity and tend to be conservative in the independence case, at least for the type of variables measured in the survey data used by the study (Britain General Household Survey, 1971 and Britain Election Survey, 1974).

This study aimed to identify some indicators of poverty based on the 1994 FIES. In addition to the identification process which would update observations from similar studies, this study assessed alternative test statistics for the chi-square test of independence. These statistics "correct" the usual Pearson chi-square statistic to account for possible effects of the sampling design used in the survey.

2. Methodology

2.1 Overview

Each household interviewed in the 1994 FIES is classified as either 'poor' or 'non-poor' using the official regional poverty threshold (see Appendix Table A for the official regional poverty thresholds) and this classification is referred to in this study as the household's poverty status. The association of the household's poverty status with some of the variables observed in the survey was evaluated using the chi-square test of independence. The usual Pearson test statistic $\chi^2(P)$, and three design-based corrections were computed and compared with respect to magnitude of values and estimated probabilities of a type I error. Two versions, $\chi^2(\hat{\delta})$ and $\chi^2(\bar{d})$, of the first order correction and a second order correction $\chi^2(S,\alpha)$ were defined following Holt, et al. (1980), Fellegi (1980), and Hidiriglou and Rao (1987). Furthermore, residual analysis was done to assess and interpret deviations from the null hypothesis. The Statistical Analysis System (SAS) software was used to implement the computations.

2.2 Correcting for design effect

Statistical independence of two variables in a nominal scale of measurement can be evaluated using chi-square test of independence. The test procedure assumes that a simple random sample is taken from the population and each element is cross-classified according to the levels of the two variables producing an R × C contingency table where R refers to the number of rows and C to the number of columns. The general form of a chi-square test statistic is given as follows:

$$\chi^2 = \sum_{all\ cells} \frac{(observed - expected)^2}{expected}.$$

Under the null hypothesis that the two variables used in the row and column classification of the table are independent, the test statistic follows a chi-square distribution with (R-1)(C-1) degrees of freedom. Also, under the same null hypothesis the probability that a unit is in the ijth cell is equal to the product of the marginal probabilities of the ith row and jth column, in symbols, $p_{ij} = p_{i+} \cdot p_{+j}$ where p_{ij} is the probability that a unit is in the ijth cell, p_{i+} is the marginal probability that the unit is in the ith row and p_{+j} is the marginal probability that the unit is in the jth column of the contingency table. Using the empirical estimates of the marginal probabilities, $\hat{p}_{i+} = \frac{\sum_j^C n_{ij}}{n}$ and $\hat{p}_{+j} = \frac{\sum_i^R n_{ij}}{n}$, an estimate of p_{ij} is given by:

$$\hat{p}_{ij} = \hat{p}_{i+} \cdot \hat{p}_{+j} \quad (1)$$

where n_{ij} is the observed number of units in (i,j)th cell and n is the total number of units cross-classified in the contingency table. With an estimate of p_{ij}, the expected frequency in the ijth cell is equal to $n\,\hat{p}_{ij}$.

Rao and Scott (1979) as cited by Holt, et al. (1980) show that Pearson's χ^2 statistic computed from survey data is not asymptotically distributed as a chi-square random variable. Rather, this statistic is asymptotically distributed as the weighted sum of independent chi-square random variables, W_s with 1 degree of freedom, that is, $\sum_{s=1}^{(R-1)(C-1)} \delta_s W_s$. Rao and Scott (1981) showed that the weights, δ_s, estimated by $\hat{\delta}_s$, are the eigenvalues of design effect (*deff*) matrix **D**, estimated by $\hat{\boldsymbol{D}}$ where

$$\hat{\boldsymbol{D}} = n\left(\hat{\boldsymbol{P}}^{-1}_{(R-1)+} \otimes \hat{\boldsymbol{P}}^{-1}_{+(C-1)}\right)\hat{\Gamma} \quad (2)$$

where $\otimes$ denotes direct product operator. Also, $\hat{\boldsymbol{P}}_{(R-1)+} = diag\left(\hat{\boldsymbol{p}}_{i+}\right) - \hat{\boldsymbol{p}}_{i+}\hat{\boldsymbol{p}}'_{i+}$, $\hat{\boldsymbol{P}}_{+(C-1)} = diag\left(\hat{\boldsymbol{p}}_{+j}\right) - \hat{\boldsymbol{p}}_{+j}\hat{\boldsymbol{p}}'_{+j}$ and $\hat{\Gamma}$ is the estimated covariance matrix of $\hat{h}_{ij} = \hat{p}_{ij} - \hat{p}_{i+}\hat{p}_{+j}$ for i = 1, 2, ..., (R-1) and j = 1, 2, ..., (C-1). The vectors, $\hat{\boldsymbol{p}}_{i+}$ and $\hat{\boldsymbol{p}}_{+j}$ are the column vectors of the estimated marginal probabilities, that is, $\hat{\boldsymbol{p}}_{i+} = \left(\hat{p}_{1+}\hat{p}_{2+}\cdot\cdot\hat{p}_{(R-1)+}\right)'$ and $\hat{\boldsymbol{p}}_{+j} = \left(\hat{p}_{+1}\hat{p}_{+2}\cdot\cdot\hat{p}_{+(C-1)}\right)'$.

Hidiroglou and Rao (1987) obtained estimates of p_{ij}'s in a stratified two-stage cluster sample as $\left(\hat{p}_{ij} = \frac{\hat{N}_{ij}}{\hat{N}}\right)$ which is the ratio of two post-stratified estimates where $\hat{N}_{ij} = \sum_{h=1}^{L}\sum_{t=1}^{r_h}\sum_{k=1}^{m_{ht}} w_{htk}\, y_{ij(htk)}$ and $\hat{N} = \sum_{i=1}^{R}\sum_{j=1}^{C}\hat{N}_{ij}$. Here, $Y_{ij(htk)}$ takes the value one if the kth sampled secondary sampling unit in the tth primary sampling unit of the hth stratum belongs to the ijth cell of the contingency table and zero, otherwise. The basic sampling weight attached to the (htk)th element denoted by w_{htk} is taken as the inverse of the probability of inclusion of the element in the sample. Defining the estimates of marginal probabilities in terms of $\hat{p}_{ij}$'s, $\hat{p}_{i+}$ is $\sum_{j=1}^{C}\hat{p}_{ij}$ while $\hat{p}_{+j}$ is $\sum_{i=1}^{R}\hat{p}_{ij}$. The estimated variance of $\hat{p}_{ij}$ is given by

$$\hat{\text{var}}\left(\hat{p}_{ij}\right) = \hat{N}^{-2}\sum_{h=1}^{L}\left(\frac{r_h}{r_h - 1}\right)\sum_{t=1}^{r_h}(z_{ijht} - \bar{z}_{ijh})^2 \tag{3}$$

where $z_{ijht} = \sum_k w_{htk}Y_{ij(htk)} - \hat{p}_{ij}\sum_i\sum_j\sum_k w_{htk}Y_{ij(htk)}$ and $\bar{z}_{ijh} = \sum_t z_{ijht} \Big/ r_h$ and r_h is the number of primary sampling units in the hth stratum. The estimated variance of the estimate of a marginal probability is a linear combination of $\hat{\text{var}}\left(\hat{p}_{ij}\right)$ and $\hat{\text{cov}}\left(\hat{p}_{ij}, \hat{p}_{i'j'}\right)$, a measure of linear relationship between $\hat{p}_{ij}$ and $\hat{p}_{i'j'}$ where $i' \neq i$ and $j' \neq j$.

In terms of cell and marginal probabilities, the null hypothesis can be stated as Ho: $p_{ij} = p_{i+} \cdot p_{+j}$, where i = 1,2, ... R and j =1, 2, ..., C. The test statistic which is commonly known as Pearson's χ^2 statistic can therefore be expressed as:

$$\chi^2(P) = n\sum_{i=1}^{R}\sum_{j=1}^{C}\frac{\left(\hat{p}_{ij} - \hat{p}_{i+}\hat{p}_{+j}\right)^2}{\hat{p}_{i+}\hat{p}_{+j}} \tag{4}$$

where n is the total sample size used in the survey. Also, Hidiroglou and Rao (1987) gave the following expression for $\hat{\gamma}_{ij\ i'j'}$, elements of $\hat{\Gamma}$:

$$\hat{\gamma}_{ij\ i'j'} = \hat{N}^{-2}\sum_{h=1}^{L}\left(\frac{r_h}{r_h - 1}\right)\sum_{t=1}^{r_h}(z_{ijht} - \bar{z}_{ijh})(z_{i'j'ht} - \bar{z}_{i'j'h}) \tag{5}$$

where $z_{ijht} = \sum_k w_{htk}Y_{ij(htk)} - \hat{p}_{i+}\sum_i\sum_k w_{htk}Y_{ij(htk)} - \hat{p}_{+j}\sum_j\sum_k w_{htk}Y_{ij(htk)} + \hat{p}_{i+}\hat{p}_{+j}\sum_i\sum_j\sum_k w_{htk}Y_{ij(htk)}$.

It should be noted that the elements of the estimated covariance matrices of $\hat{p}_{ij}$ and $\hat{h}_{ij}$ depend on the primary sampling unit totals $\left(\sum_k w_{htk}Y_{ij(htk)}\right)$ and their marginal totals.

Using $\hat{\delta}_s$, Holt, et al. (1980) gave a correction factor for the chi-square test of independence from survey data. However, he left a challenge to others to obtain an expression of the correction factor using the cell and marginal probabilities. This challenge was answered by Gross in 1984 when he obtained an expression in terms of the cell and marginal probabilities and the asymptotic variances of their estimates without evaluating the individual eigenvalues, $\hat{\delta}_s$. Hidiroglou and Rao (1987) refer to this correction as a first-order correction of the chi-square statistic.

The first of this type of correction depends on the estimated cell *deffs*, $\hat{d}_{ij} = \frac{\hat{\text{var}}(\hat{p}_{ij})}{\hat{p}_{ij}(1 - \hat{p}_{ij})n^{-1}}$, and the estimated *deffs* of margins, $\hat{d}_{A(i)} = \frac{\hat{\text{var}}(\hat{p}_{i+})}{\hat{p}_{i+}(1 - \hat{p}_{i+})n^{-1}}$ and $\hat{d}_{B(j)} = \frac{\hat{\text{var}}(\hat{p}_{+j})}{\hat{p}_{+j}(1 - \hat{p}_{+j})n^{-1}}$, and is given by:

$$\chi^2(\hat{\delta}) = \frac{\chi^2(P)}{\hat{\delta}_\bullet} \tag{6}$$

where $(R-1)(C-1)\hat{\delta}_\bullet = \sum_{i=1}^{R}\sum_{j=1}^{C}\frac{\hat{p}_{ij}(1-\hat{p}_{ij})}{\hat{p}_{i+}\hat{p}_{+j}}\hat{d}_{ij} - \sum_{i=1}^{R}(1-\hat{p}_{i+})\hat{d}_{A(i)} - \sum_{j=1}^{C}(1-\hat{p}_{+j})\hat{d}_{B(j)}$. The statistic $\chi^2(\hat{\delta})$ is said to be distributed as central chi-square with (R-1)(C-1) degrees of freedom under the null hypothesis.

A second type of first order correction makes use of Fellegi's suggestion (1980) to use the average cell design effect $\bar{d}$ as a divisor, that is,

$$\chi^2(\bar{d}) = \frac{\chi^2(P)}{\bar{d}} \tag{7}$$

where $\bar{d} = \frac{\sum_{i=1}^{R}\sum_{j=1}^{C}\hat{d}_{ij}}{RC}$. The null distribution of $\chi^2(\bar{d})$ is also central chi-square with (R-1)(C-1) degrees of freedom.

Hidiroglou and Rao (1987) also reported a second order correction to the Pearson statistic based on the Satterthwaite approximation to $\sum_{s=1}^{(R-1)(C-1)}\delta_s W_s$. This correction is obtained by treating $\chi^2(S) = \frac{\chi^2(P)}{\hat{\delta}_\bullet(1+\hat{C}_\delta^2)}$ as χ^2_ν where $\nu = \frac{(R-1)(C-1)}{1+\hat{C}_\delta^2}$ and $\hat{C}_\delta^2$ is the estimated coefficient of variation of the $\hat{\delta}_s$. This second order correction takes account of the variability in the $\hat{\delta}_s$ unlike the first order correction. Also, like $\hat{\delta}_s$, $\hat{C}_\delta^2$ can be calculated without evaluating the individual eigenvalues. The mathematical expression for $\hat{C}_\delta^2$ is given by:

$$\hat{C}_\delta^2 = \frac{\sum\hat{\delta}_s^2}{(R-1)(C-1)\hat{\delta}_\bullet^2} - 1 \tag{8}$$

where $\sum_{s=1}^{(R-1)(C-1)}\hat{\delta}_s^2 = n^2\sum_{j'}^{C}\sum_{j}^{C}\sum_{i'}^{R}\sum_{i}^{R}\frac{\gamma_{ij\,i'j'}^2}{\hat{p}_{i+}\hat{p}_{+j}\hat{p}_{i'+}\hat{p}_{+j'}}$. The computed $\chi^2(S)$ is usually modified to $\chi^2(S,\alpha) = \chi^2(S)\left(\frac{\chi^2(\alpha)}{\chi^2_\nu(\alpha)}\right)$ where $\chi^2(\alpha)$ is the customary upper α-point of chi-square distribution with (R-1)(C-1) degrees of freedom. The null hypothesis is rejected at the α level if $\chi^2(S,\alpha)$ exceeds $\chi^2(\alpha)$.

Assuming that the Satterthwaite approximation is accurate, $\Pr[\chi^2(S,\alpha) > \chi^2(\alpha)|Ho] \doteq \alpha$, the type I error rates of $\chi^2(P)$ and $\chi^2(\hat{\delta})$ can be estimated as

$\Pr\left[\chi_{\nu}^{2} \geq \frac{\chi^{2}(\alpha)}{\hat{\delta}_{\bullet}(1+\hat{C}_{\delta}^{2})}\right]$ and $\Pr\left[\chi_{\nu}^{2} \geq \frac{\chi^{2}(\alpha)}{(1+\hat{C}_{\delta}^{2})}\right]$, respectively, for nominal level α. Similarly, an estimate of the type I error rate using Felligi's correction is $\Pr\left[\chi_{\nu}^{2} \geq \bar{d}\,\frac{\chi^{2}(\alpha)}{\hat{\delta}_{\bullet}(1+\hat{C}_{\delta}^{2})}\right]$.

Hidiroglou and Rao (1987) discuss the usefulness of an analysis of residual in detecting deviations from the null hypothesis. The standardized residuals $\hat{e}_{ij} = \frac{e_{ij}}{\sqrt{\hat{d}_{ij}(h)}}$ are approximately N(0,1) under the null hypothesis where $e_{ij} = \hat{h}_{ij} \Big/ \sqrt{\hat{p}_{i+}\hat{p}_{+j}(1-\hat{p}_{i+})(1-\hat{p}_{+j})n^{-1}}$ are the standardized residuals under the assumption of simple random sampling (Haberman, 1973). The estimated *deff* of $\hat{h}_{ij}$ under the null hypothesis is defined as $\hat{d}_{ij}(h) = \frac{\hat{\gamma}_{ijij}}{\sqrt{\hat{p}_{i+}\hat{p}_{+j}(1-\hat{p}_{i+})(1-\hat{p}_{+j})n^{-1}}}$.

3. Results and Discussion

For clarity, the results of the study are presented in two parts. The first part discusses the results of procedures for identifying correlates of poverty status of a household from information in the 1994 FIES while the second part presents the effects of the survey design on the computed chi-square statistics together with their estimated size of type I error.

3.1 Identifying Correlates of Poverty Status

Six non-income based characteristics of the household head and eleven characteristics of the household were evaluated for their association with the poverty status of a household. The six characteristics of the household head are sex, marital status, age, educational attainment, employment status and class of worker of an employed household head. The household characteristics are household type, employment status of the spouse of the household head, major source of income, indicator as to whether the main source of income of the household is agriculture or not (or, agriculture indicator), total number of household members, and composition of the household in terms of the total number of household members and presence of non-relatives, household members aged less than 1 year old, between 1 to 6 years old, between 7 to 14 years old, and between 15 to 24 years old.

The chi-square test statistics from the test of independence discussed in section 2 of these 17 variables against poverty status are shown in Table 1. All the uncorrected chi-square values $\chi^{2}(P)$ are very large with the highest value of 3,683 corresponding to the agriculture indicator. Among the characteristics of the household head, educational attainment has the highest computed uncorrected chi-square test statistic of 2,529.

$\chi^{2}(\hat{\delta})$ and $\chi^{2}(\bar{d})$ values shown in Table 1 are the first order corrections while $\chi^{2}(S,\ 0.05)$ are the chi-square values with second order correction. It can be observed that the values of the corrected statistics still lead to the rejection of the null hypothesis.

Table 1. Uncorrected and corrected computed values of the chi-square test statistics of selected characteristics form the 1994 FIES cross-classified with poverty status.

	$R \times C$	$\hat{\delta}$	$\bar{d}$	$\chi^2(P)$	$\chi^2(\hat{\delta})$	$\chi^2(\bar{d})$	$\chi^2(S,$ 0.05
Characteristic of the household head							
Sex	2×2	0.395	1.113	260	659	234	707
Marital status	3×2	0.235	0.748	219	931	292	728
Age	4×2	0.579	0.744	308	532	414	543
Educational attainment	4×2	0.718	0.850	2529	3520	2974	3758
Employment status	2×2	0.418	1.150	427	1023	372	1029
Class of worker	3×2	0.623	0.964	826	1326	857	1438
Characteristic of the household (hhld)							
Household type	2×2	0.590	1.254	128	218	102	235
Total number of hhld members	4×2	0.422	0.644	1482	3509	2299	3176
Total number of employed hhld members	2×2	0.475	1.179	38	80	32	72
Presence of hhld mem. aged <1 yr old	2×2	0.288	1.050	242	842	231	796
Presence of hhld mem. aged bet. 1 to 6 yrs	2×2	0.763	1.176	1283	1683	1091	1536
Presence of hhldmem. aged bet. 7 to 14 yrs	2×2	0.756	1.165	230	304	197	282
Presence of hhldmem aged bet. 15 to24 yrs	2×2	0.625	1.157	519	830	448	596
Presence of non-relative hhld members	2×2	0.182	1.074	418	2297	389	2863
Employment status of the hhld head spouse	2×2	0.645	1.172	260	404	222	298
Agriculture indicator	2×2	1.250	1.512	3683	2945	2435	2576
Major source of income	3×2	0.832	1.139	967	1162	849	1301

A measure of the degree of association between the variables is the coefficient of contingency. Table 2 shows the computed coefficients based on the Pearson statistic for the selected variables and poverty status. It can be observed that there are contingency coefficients very near zero implying very weak association of the characteristic with poverty status. Using the criterion of declaring only those characteristic with contingency coefficient of at least 0.20 as having association with poverty status, four characteristics "qualify" as correlates of poverty.

Table 2. Contingency coefficients associated with the uncorrected chi-square test statistic.	*Contingency Coefficient*
Characteristic of the household head	
Sex	0.102
Marital status	0.094
Age	0.111
Educational attainment	0.304
Employment status	0.130
Class of worker	0.195
Characteristic of the household	
Household Type	0.072
Total number of household members	0.237
Total number of employed household members	0.039
Presence of household members aged less than 1 year old	0.098
Presence of household members aged between 1 to 6 years old	0.222
Presence of household members aged between 7 to 14 years old	0.096
Presence of household members aged between 15 to 24 years old	0.143
Presence of non-relative household member	0.129
Employment status of the household head spouse	0.115
Agriculture indicator	0.360
Major source of income	0.194

One of the four characteristics associated with poverty status is a characteristic of the household head and the rest are characteristics of the household itself. Educational attainment has a coefficient equal to 0.304. Deviations from the null hypothesis of independence as measured by the computed standardized residuals of the different categories of educational attainment variable with poverty status are shown in Table 3. All the $\hat{e}_{ij}$'s are very large in absolute values. The largest residual is -4,679 of poor households in household headed by a person with college education. From the large positive values $\hat{e}_{31} = 438$ and $\hat{e}_{41} = 1{,}613$, it can be concluded that a significantly greater proportion of heads of non-poor households have higher educational attainment (at least high school) compared to heads of poor households.

The characteristics of the household associated with poverty status include agriculture indicator, total number of household members, and presence of household members aged between 1 to 6 years old. The agriculture indicator has the highest contingency coefficient equal to 0.360 (see Table 2). The standardized residuals are given in Table 4 with -1,479 as the largest residual corresponding to the category of poor households in the household whose main source of income is not agriculture. The negative value means that poor households are part of households whose main source of income is agriculture.

Table 3. Design-based standardized residuals of the educational attainment of the household head.

Educational Attainment of the Household Head	Poverty Status	
	Non-poor	Poor
No schooling	-790	595
Reached elementary or elementary graduate	-970	968
Reached high school or high school graduate	438	-601
Reached College or college graduate	1613	-4679

Table 4. Design-based standardized residuals of the agriculture indicator of the household.

Agriculture Indicator	Poverty Status	
	Non-poor	Poor
Agriculture as main source of income	-1455	1290
Non-agriculture as main source of income	1346	-1479

The other two characteristics of the household found to be associated with poverty status are concerned with the composition of the household membership. The total number of household members has a coefficient of 0.237 while the presence of household members aged between 1 to 6 years old has 0.222 as coefficient. The standardized residuals for these two characteristics are shown in Tables 5 and 6, respectively.

Table 5. Design-based standardized residuals of the total number of household members.

Total Number of Household Members	Poverty Status	
	Non-poor	Poor
Less than 5 members	946	-1182
Between 5 to 7 members	-380	370
Between 8 to 12 members	-1308	1036
Greater than 12 members	-4087	6681

Table 6. Design-based standardized residuals of presence of household members aged between 1 to 6 years old.

Presence of Household Members Aged Between 1 to 6 years old	Poverty Status	
	Non-poor	Poor
Present	-1001	934
Absent	876	-968

In Table 5, the largest residual is 6,681 and from the large positive residuals, $\hat{e}_{22} = 370$, $\hat{e}_{32} = 1{,}036$ and $\hat{e}_{42} = 6{,}681$, it can be concluded that greater proportion of poor households are part of households with large membership. In Table 6, the largest residual is -1,001 which implies that non-poor households are part of households without member aged between 1 to 6 years old.

3.2 Effects of the Sampling Design

Referring back to Table 1 which shows the computed values of the uncorrected and corrected chi-square test statistics, it can be observed that there are substantial changes in the computed values upon the application of the correction factors. Table 7 shows the changes in the computed values expressed as percentages of the uncorrected values. Except for agriculture indicator, the corrected chi-square values are larger when the first order correction based on $\hat{\delta}$ is applied. The second order correction likewise yields large values. The values of $\hat{\delta}_s$ are mostly less than 1, hence its mean when applied to the uncorrected chi-square values leads to higher values. Since values of individual $\hat{\delta}_s$'s are close to its mean value it also has a small coefficient of variation and thus results in large corrected chi-square values when the second order correction is applied. On the other hand, most of the corrected chi-square values based on d yield smaller values since the computed values of $\bar{d}$'s are greater than 1.

Table 7. Percentage increase (decrease) in the values of the corrected from the uncorrected chi-square test statistic.

	Percentage increase (decrease) obtained in		
	$\chi^2(\hat{\delta})$	$\chi^2(\bar{d})$	$\chi^2(S,\ 0.05)$
Characteristic of the household head			
Sex	153.419	(10.168)	171.995
Marital status	325.478	33.622	232.986
Age	72.722	34.473	76.056
Educational attainment	39.199	17.608	48.582
Employment status	139.292	(13.013)	140.709
Class of worker	60.484	3.714	73.985
Characteristic of the household			
Household Type	69.529	(20.236)	82.713
Total number of household members	136.834	55.183	114.377
Total number of employed household members	110.628	(15.212)	89.268
Presence of household members aged less than 1 year old	247.669	(4.801)	228.544
Presence of household members aged between 1 to 6 years old	31.141	(14.988)	19.685
Presence of household members aged between 7 to 14 years old	32.217	(14.200)	22.966
Presence of household members aged between 15 to 24 yrs. old	60.020	(13.586)	15.023
Presence of non-relative household members	449.481	(6.923)	584.784
Employment status of the household head spouse	55.116	(14.670)	14.368
Agriculture indicator	(20.022)	(33.871)	(30.056)
Major source of income	20.187	(12.190)	34.594

On the average, the first order correction based on $\hat{\delta}$ results in an increase of 116.67% of the uncorrected chi-square value while the correction based on $\bar{d}$ leads to an average decrease of 1.72% of the uncorrected value. The average increase when the second order correction is applied is 112.98%, smaller than the average increase obtained under the first order correction. It is only for the agriculture indicator where all of the corrections applied result in smaller corrected chi-square values. The uncorrected chi-square value is 3,683 which led to a corrected chi-square value equal to 2,435 when first order correction based on $\bar{d}$ was applied. The percentage decrease in this correction is 33.87% which is the largest percentage decrease obtained.

The estimated probabilities of committing type I error based on the Satterthwaite approximation are shown in Table 8. The largest estimated size of the Type I error is 0.1230 for agriculture indicator. This estimate is associated with the uncorrected chi-square value and it means that without corrections one is declaring dependence of agriculture indicator and poverty status when in fact the variables are independent with probability 0.1230 and not 0.05, the assumed level of significance.

Table 8. Estimated sizes of type I error associated with the uncorrected and corrected chi-square values of selected characteristics form the 1994 FIES cross-classified with poverty status.

	Estimated size of type I error associated with		
	$\chi^2(P)$	$\chi^2(\hat{\delta})$	$\chi^2(\bar{d})$
Characteristic of the household head			
Sex	0.0003	0.040	0.0001
Marital status	< 0.0001	0.106	0.0002
Age	0.0016	0.046	0.0128
Educational attainment	0.0047	0.036	0.0143
Employment status	0.0003	0.049	< 0.0001
Class of worker	0.0035	0.037	0.0044
Characteristic of the household			
Household Type	0.0042	0.040	0.0010
Total number of household members	0.0008	0.069	0.0129
Total number of employed household members	0.0034	0.068	0.0012
Presence of household members aged less than 1 year old	0.0004	0.057	0.0003
Presence of household members aged between 1 to 6 years old	0.0320	0.061	0.0200
Presence of household members aged between 7 to 14 years old	0.0297	0.059	0.0190
Presence of household members aged between 15 to 24 years old	0.0319	0.116	0.0185
Presence of non-relative household members	< 0.0001	0.021	< 0.0001
Employment status of the household head spouse	0.0325	0.110	0.0180
Agriculture indicator	0.1230	0.073	0.0421
Major source of income	0.0146	0.033	0.0075

However, substantial decreases may be observed in the estimated sizes of type I error when corrections are made as shown in Table 8. From a value of 0.1230, the estimated size of type I error associated with the uncorrected chi-square value of agriculture indicator, is much lower at 0.0421 when the first order correction based on $\bar{d}$ was applied. On the average, the first order correction based on $\bar{d}$ gave larger reductions in estimated sizes of type I error compared to reductions of the correction based on $\hat{\delta}$. However, estimated sizes of type I error associated with $\chi^2(\hat{\delta})$ are also closer to the nominal level of significance compared to those associated with $\chi^2(\bar{d})$. This is so since the Satterthwaite's approximation is based on the individual values of $\hat{\delta}_s$ and not on $\hat{d}_{ij}$ values.

4. Conclusions

When applied to the 1994 FIES data to determine the association of poverty status with some characteristics of the household head and of the household itself, the Pearson's chi-square statistic for the test of independence yields large computed test statistics. Upon application of the first and second order corrections for survey design effects to the chi-square test statistic, the computed values are reduced by as much as 33.87% of the uncorrected computed chi-square value. However, the resulting corrected values remain very much greater than the critical value at 5% level of significance. Hence, the same decision (rejection of the null hypothesis of independence) is arrived at based on either the uncorrected or corrected chi-square values.

Four characteristics with contingency coefficients computed from the uncorrected Pearson statistic of at least 0.20 are considered by the study as significantly associated with poverty status. One of the four characteristics refers to the household head while the rest describe the household. Educational attainment of the household head is a correlate of poverty status of a household. Poor households tend to belong to households headed by a person with low educational attainment. In terms of the characteristics of the household, agriculture indicator and household membership defined as the total number of members and presence of children aged between 1 to 6 years old are also correlates of poverty status of the household. Poor households tend to belong to households with agriculture as main source of income. Also, poor households are found in households characterized by large number of household members and where, in addition, there are children aged between 1 to 7 years old who are members of the household.

The substantial effect of the survey design is the reduction of computed values of Pearson chi-square test statistic when the design-based corrections are applied. Corresponding differences in the estimated sizes of type I error associated with the uncorrected and corrected chi-square test statistics are also established in this study. An estimated probability of type I error associated with the uncorrected chi-square test statistic value as high as 0.1230 is reduced to a value near the nominal level of α upon application of the corrections to the chi-square test statistic. Hence, if one does not adjust the chi-square test statistic the probability of committing type I error will be higher than the set value of α. This result indicates that although the usual Pearson's chi-square test when applied to survey data leads to same decision as when the corrected chi-square test statistic at least for the 1994 FIES data, it is the probability of committing type I error which is very much affected by the sampling design. Such effect has an implication on conclusions made based on the results of the tests. It is therefore recommended that chi-square test statistics corrected for inclusion probabilities be used when testing for independence. Such correction can be easily made if *deffs* can be published as well. Further studies need to be done, however, to support the empirical results presented, not only on type I error but on the power of the test, as well.

5. Acknowledgment

The authors would like to thank Dr. Dalisay S. Maligalig, Dr. Isidoro P. David and the anonymous referee for their critical comments and suggestions which improved the paper considerably.

6. Appendix

Appendix Table A. 1994 Official Regional Poverty Thresholds

REGION														
NCR	CAR	1	2	3	4	5	6	7	8	9	10	11	12	ARMM
11,230	10,853	10,022	8,316	9,757	9,537	8,319	8,197	6,425	6,444	7,074	7,938	8,201	8,971	8,889

Source: Philippine Poverty Statistics released by National Statistical Coordination Board, December 1996.

7. References

BAUTISTA, R.P. 1991 *A study of poor and non-poor families using discriminant analysis*, Proceedings of the Fourth National Convention on Statistics.

COSME, A. 1998 *Non-income based determinants of poor and non-poor families*, Undergraduate Special Problem, University of the Philippines Los Baños.

FELLEGI, I.P. 1980 *On adjusting the Pearson chi-square statistic for cluster sampling*, Journal of the American Statistical Association, Vol. 71, pp. 665-670.

GROSS, W.F. 1984 *A note on chi-square test with survey data*, Journal of the Royal Statistical Society, Series B., Vol. 46, pp. 270-272.

HABERMAN, S.J. 1973 *The analysis of residuals in cross-classified tables*, Biometrics, Vol.29., pp. 205-220.

HIDIROGLOU, M.A. and J.N.K. RAO 1987 *Chi-squared tests with categorical data from complex surveys: Part I- Simple goodness of fit, homogeneity and independence in a two-way table with application to the Canada Health Survey (1978-1979)*, Journal of Official Statistics, Vol. 3, No. 2, pp.117-132.

HOLT, D. , A.J. SCOTT AND P.D. EWINGS 1980 *Chi-squared tests with survey data*, Journal of the Royal Statistical Society, Series A., Vol. 143, pp. 303-320.

MARQUEZ, N.R. AND R.A. VIROLA 1995 *Monitoring changes in the characteristics of the Philippine poor: 1985-1994*, Proceedings of the Sixth National Convention on Statistics, pp. 434-460.

NATHAN, G 1975 *Test of independence in contingency tables from stratified proportional samples*, Sankhya: The Indian Journal of Statistics, Vol. 87, Series C, pp. 77-87.

NORIEGA, R.M. 1995 *Alternative methodology for determining the poor from the non-poor: Philippine case studies on poverty mapping*, Proceedings of the Sixth National Convention on Statistics, pp. 419-433.

National Statistics Coordination Board 1996 *Philippine Poverty Statistics, Metro Manila*, Philippines

ORBETA, A. AND A.R. HILARIO 1995 *A note on the adequacy of income as poverty indicator*, Paper presented at the MIMAP Advisory Council meeting.

RAO, J.N.K. AND SCOTT, A.J. 1981 *The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables*, Journal of the American Statistical Association, Vol. 76, pp. 221-230.

REYES, M., R.R. CANCO AND K.C. ILARDE 1981 *Using MBN indicators for poverty mapping*, Paper presented at the MIMAP Advisory Council meeting.

VIROLA, R.A. 1994 *Counting the poor in the Philippines*, Paper presented during the second quarterly meeting of the Philippine Statistical Association.

VIROLA, R.A. AND DAVID, I.P. 1995 *A comparative review of poverty assessment: Philippines and Indonesia Proceedings of the Sixth National Convention on Statistics*, pp. 593-609.

Us Panel On Poverty And Family Assistance *1995 Measuring Poverty: A New Approach*, National Academy of Science, USA.