BIBLIOGRAPHY

DAS-ILEN, ELVIRA D. and PADILAN, DIXY E. April 2008. Efficiency of CATANOVA in Measuring Association between Socio-Economic Variables and Motivations and Aspirations of Working Students. Benguet State University, La Trinidad, Benguet.

ADVISER: SALVACION Z. BELIGAN, Ph.D.

ABSTRACT

The study was conducted to determine the association between socio-economic variables and motivations and aspirations of working students at Benguet State University using the Analysis of Variance for Categorical data (CATANOVA) as an alternative tool for Pearson's Chi-square test. Specifically, this study aimed to determine the relationship of working student's motivation and aspiration with their socio-economic variables and to determine the proportion of variation of the response variable (motivations/aspirations) attributed to the independent variables (socio-economic variables).

A survey questionnaire was floated and distributed to the respondents through random sampling in order to obtain the needed data for this study.

Both chi-square and C statistics were computed to determine the significance of the association between the response and the independent variables.

Based on the results of the analysis employing both CATANOVA and Pearson's Chi-square students motivations was associated to the father's occupation, household size and sibling's position. Aspiration of working students was associated to gender and family income. With the use of CATANOVA, the percent contribution of the socio-economic variables on the variability of the working student's motivations and aspirations were determined.

CATANOVA statistics is at par with the Pearson's Chi-square statistics in determining associations but more efficient than Pearson's Chi-square in terms of their variability.

TABLE OF CONTENTS

Page

Bibliography ……………………………………………………………….………i

Abstract ……………………………………………………………………………i

Table of Contents ………………………………………………………………...iii

INTRODUCTION

Background of the Study………………………………….…………1

Objectives of the Study …………………………………………………...3

Importance of the Study ………………………………………………. 3

Scope and Delimitation of the Study ………………….…………………4

REVIEW OF RELATED LITERATURE

Studies Using CATANOVA ……………………………….…….…... 5

Studies on the Association of Educational
Aspirations and Socio-economic Variables ……………..….……..….. 8

THEORETICAL FRAMEWORK

Analysis of Variance for Categorical Data ………………………….…...11

Pearson's Chi-square Test ……….……………………………………….23

Distribution of Pearson's Chi-Square …….………………………………24

Efficiency of Pearson's Chi-square
and CATANOVA ………………………………….…….…………….24

Definition of Terms ………………………………..…………………...25

METHODOLOGY

Locale of the Study …………………………….…………………………27

Respondents of the Study …………………….……….………….…..27

Data Collection Instrument …………………….……………..……...…27

Analysis of Data ……………………………….……….…………28

RESULTS AND DISCUSSION

Associations Between Educational Motivations and Socio-Demographic Profile of the Working Students ………..…………29

Associations Between Educational Aspirations and Socio-Demographic Profile of the Working Students ……….…………31

Efficiency of CATANOVA and the Pearson's Chi-square Test in Measuring Associations ...………………...……..32

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Summary ……………………………………….….………………34

Conclusions ……………………………………….…….……………34

Recommendations ………………………………….…….…………..35

LITERATURE CITED ………………………………………….….…………..36

APPENDICES ………………………………………………………..…………37

INTRODUCTION

Background of the Study

Researchers usually grope on what statistical tool to be used in determining relationships of categorical variables. Several techniques for analyzing and testing contingency data are available in many literatures. The Chi-Square test is one of the most commonly used for determining the association or dependence of two variables. The Chi-Square test is most appropriate when the sample has been selected from an infinite or large bivariate multinomial population and the sample size n is reasonable large (Neter, et al., 1988). However, according to Light and Margolin (1974), Chi-square test is no longer appropriate when sample size n is small, that is, when more than 20% of the observed cell frequencies are lower than 5. Thus, a statistical tool known as Categorical Analysis of Variance (CATANOVA) derived earlier by Light and Margolin (1971) is suggested for use for small sample comparisons.

CATANOVA is a variant of the standard One-Way ANOVA since it also partition the variation of nominal data into additive components. The procedure deals with the frequency counts and row and column marginal totals in a two dimensional contingency table. Similar to the parametric One-Way ANOVA, the general approach to categorical data is to compute the total variation in the data and then partition this into specific components. Aside from the measurement of association statistic, CATANOVA can also determine the proportion of variation attributable to the independent variable which other test can not do, like the chi-square test.

For the application of CATANOVA, the researchers of this study sought to utilize the empirical data on the educational aspirations of working students.

One of the most important concerns of any College or University is the full development of students' potential through meaningful and relevant programs that respond to their varied backgrounds, orientations, personal and professional needs. To meet their needs, the provision of a well-defined student assistance program requires much attention, as it contributes to the total development of prospective professionals through varied learning experience in a work setting within the school. As provided for in the National Compensation Circular No. 64 (1990), students assistance maybe hired to render emergency or temporary services for the following reasons: 1) to provide practicum training for students, 2) to provide extra income for students and 3) to emphasize dignity in labor. Though these objectives of the NCC are for the benefits of the students, many students found working and studying at the same time a difficult task because it does not only consume much the time of students but also prevent them from joining extra curricular activities that will promote their personal growth.

Hence, the purpose of this study is two-fold, first, to find the relationship between students' motivation to work and their demographic profile and the relationship between educational aspirations and socio-economic background of the respondents, and second, to assess the applicability of CATANOVA in contingency tables with different dimensions.

Objectives of the Study

The study aimed to determine the efficiency of CATANOVA in measuring the degree of association between two categorical variables observed from working students at Benguet State University. Specifically, the study aimed:

1. To determine the association between educational motivation and socio-economic background of working students.

2. To determine the association between educational aspirations and socio-economic profile of the working students; and

3. To determine the efficiency of CATANOVA and Pearson Chi-square test in measuring the degree of association between two categorical variables.

Importance of the Study

Results of this study maybe utilized as an input to school administrators in policy formulation regarding student assistance programs.

The study is also important to researchers who use statistics in determining the degree of association between two variables measured in nominal scale and to any researcher who is dealing with quantitative variables.

Moreover, the result of the study will serve as a guide for researchers in determining the efficiency of CATANOVA over the Pearson's Chi-square test using the data on motivations and aspirations of working students.

Scope and Delimitation of the Study

This study was conducted at Benguet State University, La Trinidad, Benguet and it focused on the aspirations and motivations of working students in relation to their socio-economic profile using the CATANOVA.

The respondents were students from Benguet State University who are working inside the school as students' assistants, and those who are working outside BSU as service crews in McDonalds and Jollibee for a minimum of five (5) months.

The researchers gathered the needed information from the different colleges of Benguet State University through the administration of survey questionnaires.

REVIEW OF RELATED LITERATURE

Studies Using CATANOVA

Anestesiol (1995) evaluated the incidence of colonization and infection by methicillin-resistant in PICU. They studied two-hundred patients with duration of hospitalization longer than 24 hours out of the 255 patients who were hospitalized during the same period. The patients were divided in two groups according to the presence or the absences of MRS. The difference of the two populations were compared using the t-test and the CATANOVA. Wilcoxon's test was used to analyze the relation between the two values. The results were significant when p = 0.05 and Ct = 3.81. They concluded that the clinical impact of MRS in terms of morbidity and mortality in this PICU is modest. The prevention and limitation of the spread of MRS could be obtained by simple but essential measures of control.

Anderson et. al. (1980) presented an extension of the analysis of variance for categorical data (CATANOVA) procedure to multidimensional contingency tables involving several factors and a response variable measured on a nominal scale. Using an appropriate measure of total variation for multinomial data, partial and multiple association measures are developed as R^2 quantities which parallel the analogous statistics in multiple linear regression for quantitative data. In addition, test statistics are derived in terms of these R^2 criteria. Finally, this CATANOVA approach is illustrated within the context of 2 three-way contingency table from a multicenter clinical trial.

Singh B. (2004) examined the CATANOVA method for its suitability for analysis of nominal data from repeated measures design. The modified tests are developed for single and multi-group repeated measures designs, separately. The computed results reveal that in single group repeated measures designs the actual size of existing CATANOVA test is more for negative correlation and less for positive correlation than the stated size. This implies that the existing test may yield the non-existing effects as significant for negative correlation and may not even detect the real effects for positive correlation among repeated observations. Similar results are obtained for repeated factor and interaction effects and just reverse result for group effect in multi-group repeated measures design. These results show that the existing CATANOVA tests are not valid for repeated measures designs and hence the modified tests should be used for analysis of nominal data from such designs.

Souza et.al. (2001) studied the genetic divergence among organisms; generally the analysis is done directly from the DNA molecule. Therefore, a possible outcome is categorical being one out of four categories (looking at the nucleotide level). Light and Margolin (1971) developed an analysis of variance for categorical data (CATANOVA) and Pinheiro et al. (2000) employed a similar measure of variation and extended the CATANOVA procedure taking into account several positions in the sequence for balanced designs. Here we consider variable number of sequences in each group, where, the samples are unbalanced.

In order to test the null hypothesis of homogeneity among groups, the asymptotic distribution of the test statistic was found and its power is evaluated. An application of the test to real data is illustrated using resampling methods such as the bootstrap to generate the empirical distribution of the test statistics.

Light's et.al. (1979) revealed that the CATANOVA method for analysis of two ways classified nominal data was developed analogous to the least squares method of fitting constants for two way cross-classified quantitative data with disproportionate cell frequencies. This method has been compared with simple chi-square method through a numerical example. Exact small sample behavior in two-way contingency tables is investigated for Pearson's chi-square statistic χ^2, Light and Margolin's C statistic and its related R^2 measure of association, Kullback's minimum discrimination information statistic $f(\hat{2})$, and Goodman and Kruskal's Lambda. R^2 was shown to be identical to Goodman and Kruskal's λ, leading to a test for independence based on λ . In small samples from a product of multinomial model, the null distribution of C is better approximated by a chi-square distribution than is the null distribution of χ^2; both are considerably better approximated by a χ^2 distribution than is the null distribution of $f(\hat{2})$. It is proved for tables with two columns and any number of rows that if the column totals are equal, then χ^2 ; thus, χ^2 is more conservative than $f(\hat{2})$. Hence, use of $f(\hat{2})$ should be avoided in testing independence in tables with small samples.

Light's et. al. (1971) discussed a measure of variation for categorical data. He developed an analysis of variance for a one-way table, where the response variable is categorical. The data can be viewed alternatively as falling in a two-dimensional contingency table with one margin fixed. Components of variation are derived, and their properties are investigated under a common multinomial model. Using these components, we propose a measure of the variation in the response variable explained by the grouping variable. A test statistic is constructed on the basis of these properties, and its asymptotic behavior under the null hypothesis of independence is studied. Empirical sampling results confirming the asymptotic behavior and investigating power are included.

Lesser's et. al. (1980) studied the educational aspirations of 617 adolescent students in Denmark and was classified into five categories. CATANOVA and Chi-square test was used to determine the relationship of gender and educational aspirations of the students. They found out that gender is statistically associated to educational aspirations of the students although, only 2.1 percent of the variation was explained by knowledge of a respondents' gender.

Studies about Aspirations and Socio-
Economic Characteristics

Aspirations are strong desires to reach something high or great. Young people's aspirations guide what students learn in school, how they prepare for adult life, and what they eventually do (Walberg, 1989). This Digest reports on educational aspirations of rural youth compared with students living elsewhere, and suggests ways communities can work together to raise the sights of their young people.

Kayser (1973) tested the predictability of the aspirational changes using the simple, two-state, and discrete-time Markov model and to test for differences in selected background characteristics for students with different aspirational patterns. Major findings that there was a differential turnover between college and non-college plans, that students with a given level but different histories of aspirations were similar in selected background characteristics such as family status, significant others' support, and income aspirations, and that the two-state, discrete-time Markov model predicted the changes in educational aspirations for the students sampled. A major conclusion was that the assumption of aspirational stability was supported and the process appears to be Markovian in nature with one general process in operation over all the high school years.

Chiale et.al. (1965) investigated associations of overweight status and changes in overweight status over time with life satisfaction and future aspirations among a community sample of young women. Analyses were conducted using the

SPSS version 11.0 statistical software package (SPSS, Inc., Chicago, IL). χ^2 analysis was used to investigate differences in aspirations and life satisfaction among women in the four BMI categories (cross-sectional analysis) and across women in the four overweight status change categories (longitudinal analysis). Young women's aspirations were cross-sectional related to BMI category, such that obese women were less likely to aspire to further education, although this relationship maybe explained largely by current occupation. Even after adjusting for current occupation, young women who were obese were more dissatisfied with work/career/study, family relationships, partner relationships, and social activities. Weight status was also longitudinally associated with aspirations and life satisfaction. Women who were overweight or obese at both surveys were more likely than other women to aspire to "other" types of employment (including self-employed and unpaid work in the home) as opposed to full-time employment. They were also less likely to be satisfied with study or partner relationships. Women who resolved their overweight/obesity status were more likely to aspire to being childless than other women.

THEORETICAL FRAMEWORK

Analysis of Variance for Categorical Data (CATANOVA)

In this study, a technique derived by Light and Margolin (1974) was utilized as measure of association between categorical variables. The analysis begins with the partition of the total variation in the data into an explainable component and an unexplained component or noise. The general approach of this technique is to compute the total variation in the data and then partition this variation into specific components. The distributions of the various components are derived under an assumed model.

CATANOVA is a variant of the standard one-way analysis of variance. It is an analysis of variance of a one-way table with replication, where the variable being observed is nominal.

Since the data is nominal, this is one obstacle in defining measure of variation for categorical data because there is a tendency to think of variation as a measure of departure of a set of individual observations from their mean. Moreover, with categorical data, the mean is an undefined concept. However, there is an alternative way of looking at variation. Gini noted that the sum of squares of deviations from their mean for n quantitative measurements can be expressed solely as a function of the squares of the pair wise differences for all $\binom{n}{2}$ pairs. Specifically, if $X_1,...,X_n$ denote the measurements, then

$$SS = \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{2n}\sum_{j=1}^{n}\left(X_i - X_j\right)$$
$$= \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^{2} \qquad (1)$$

where,

$$\overline{X} = \sum_{i=1}^{n}\frac{X_i}{n}$$

$$d_{ij} = X_i - X_j.$$

Assuming there are G unordered experimental groups and unordered I response categories. Each response is in one of the I categories. Denote the number of responses in one category i for group $j, i = 1,..., I$ and $j - 1,..., G$ by n_{ij}. The number of responses, or group size, for group j is $n_{+j} = \sum_{j=1}^{G} n_{ij}$. The number of responses in the ith category is $n = \sum_{j+1}^{G} n_{ij}$.. Thus, the total number of responses in the study is;

$$n = \sum_{j=1}^{G} n_{+j} = \sum_{i=1}^{I} n_{i+} = \sum_{i=1}^{I}\sum_{j=1}^{G} n_{ij} \qquad (2)$$

An alternative way of viewing this data is via an I and G contingency table where n_{ij} is the count in the (i, j) cell. For this two-dimensional contingency table, or the equivalent one-way analysis of variance with categorical responses, the total variation in the response variable or “total sum of squares” is equal to;

$$TSS = \frac{n}{2} - \frac{1}{2n}\sum_{i=1}^{I} n_{i+}^2 \qquad (3)$$

In one-way analysis of variance with quantitative data one would compute the within-group component of variation and the between-group component of variation. We do this by applying the formula of variation for categorical responses $X_1,...,X_n$ which is;

$$\frac{1}{2n}\sum_{j=1}^{n}\sum_{i=1}^{n} d_{ij}^2 = \frac{1}{2n}\sum_{j=1}^{n}\sum_{i=1}^{n} d_{ij}, \qquad (4)$$

where,

$d_{ij} = 1$ if x_i and x_j name different categories

$= 0$ if x_i and x_j name the same category.

To the $\binom{n_{ij}}{2}$ pairs of responses within group $j, j = 1,...,G.$

The variation in the response variable within the j^{th} group is then

$$\frac{n_{ij}}{2} - \frac{1}{2n_{+j}}\sum_{i=1}^{I} n_{ij}^2 \qquad (5)$$

Hence the total within-group variation or within –group sum of squares (WSS) is found by assuming (5) over j ;

$$WSS = \sum_{j=1}^{G}\left(\frac{n_{+j}}{2} - \frac{1}{2n_{+j}}\sum_{i=1}^{I} n_{ij}^2\right)$$
$$= \frac{n}{2} - \frac{1}{2}\sum_{j=1}^{G}\frac{1}{n_{+j}}\sum_{i=1}^{I} n_{ij}^2 \qquad (6)$$

For balance design situations where $n_{+j} = n_+$, then

$$WSS = \frac{1}{2n}\left(Gn_{+}^{2} - \sum_{j=1}^{G}\sum_{i=1}^{I} n_{ij}^{2} \right).$$

The between-group variation or sum of squares (BSS) is defined as the differences between TSS and WSS. Hence:

$$BSS = \frac{1}{2}\left(\sum_{j=1}^{G} \frac{1}{n_{+j}} \sum_{i=1}^{I} n_{ij}^{2} \right) = \frac{1}{2n} \sum^{I} i = 1 n_{i+}^{2} \ . \qquad (7)$$

For balance design situations where $n_{+j} = n_{+}$, this reduces to

$$BSS = \frac{1}{2n_{+}G}\left[G\left(\sum_{j=1}^{G}\sum_{i=1}^{I} n_{ij}^{2} \right) - \sum_{i=1}^{I} n_{i+}^{2} \right].$$

We now turn to the problem posed in the introduction on measures of association for categorical. The three components of variation we have just derived enable us to define a measure of association between the grouping and response variables which maybe given a “proportion of variation” explained interpretation. We define this measure as:

$$R^{2} = \frac{\left(\sum_{j=1}^{G} \frac{1}{n_{+j}} \sum_{i=1}^{I} n_{ij}^{2} \right) - \frac{1}{n} \sum_{i=1}^{I} n_{i+}^{2}}{n - \frac{1}{n} \sum_{i=1}^{I} n_{i+}^{2}} = \frac{BSS}{TSS} \qquad (8)$$

It has the property that $R^{2} = 0$ if $n_{ij} / n_{+j} = f_{i}$; i = 1,…,I; j = 1,….,G, i.e., if for each j, j = 1 there exists a I such that $n_{ij} = n_{+j}$, i.e., if there is perfect predictability. Otherwise, $0 < R^{2} < 1$.

BSS, TSS and a Proposed Test Statistics

The previously proposed method of partitioning categorical variation will be referred to as a categorical analysis of variation, or CATANOVA. We developed the following C statistics to test the null hypothesis $p_{ij} = p_i$.

$$C = \frac{(n-1)(I-1)BSS}{TSS} \tag{9}$$

This test statistic, referred to as the CATANOVA C statistic, is asymptotically approximated as $\chi^2_{(I-1)(G-1)}$ under H_0.

Recalling the measure of association defined in (8), we note that $C = (n-1)(I-1)R^2$. Thus, to test the significance of the measure of association, we need to multiply it by $(n-1)(I-1)$ and approximate its distribution under H_0 by $\chi^2_{(I-1)(G-1)}$.

The derivation of the test statistic proceeds along asymptotic lines. This enables us to invoke further results for quadratic forms in normal variates, because V is asymptotically multivariate normal,

$$V \approx N(\mu, \mathrm{Z}) \tag{10}$$

In investigating BSS, we first prove that under H_0, the asymptotic distribution of BSS does not depend on $\underline{\mu}$. BSS can be written as;

$$BSS = V'BV = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{G}\left[\frac{n_{ij}}{\sqrt{n_{+j}}} - \frac{n_{i+}}{n}\right]^2 . \tag{11}$$

Let $\theta_{ij} = n_{ij} - n_{+j} p_i = n_{ij} - E_o(n_{ij})$. Then $\theta_{ij} = \sum_{j=1}^{G} \theta_{ij} = n_{i+} - np_i$. Moreover, under H_0,

$$\theta' = (\theta_{11},...,\theta_{I1},...,\theta_{1G},...,\theta_{IG})$$

is asymptotically

$$\theta' \approx N(0,Z) \qquad (12)$$

where,

Z is as in $Z_j = n_{+j} Z_o$ and $Z = M \otimes Z_o$ where,

$$. \; Z_o = \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & \cdots & -p_1 p_I \\ & p_2(1-p_2) & \cdots & -p_2 p_I \\ & & & \vdots \\ & & & p_I(1-p_I) \end{bmatrix}$$

Now,

$$BSS = V'BV$$
$$= \frac{1}{2}\sum_i \sum_j \left[\frac{\theta_{ij} + n_{ij} p_i}{\sqrt{n_{+j}}} - \frac{\theta_{i+} = np_i}{n}\sqrt{n_{+j}} \right]^2$$
$$= \frac{1}{2}\sum_{i=1}^{I} \sum_{j=1}^{G} \left[\frac{\theta_{ij}}{\sqrt{n_{+j}}} - \frac{\theta_{i+}}{n}\sqrt{n_{+j}} \right]^2$$
$$= \theta' B \theta$$

Hence, under H_0 in studying BSS, we may treat $\underline{\mu}$ as 0. From this, it follows that under H_0, BSS is asymptotically distributed as,

$$BSS \approx \sum_{i=1}^{IG} \lambda_i \chi_1^2 \qquad (13)$$

where,

$\{\lambda_i, i = 1,...,IG\}$ is the set of characteristic roots of BZ :

$$BZ = \frac{1}{2}\left[\left(M^{-1} - \frac{1}{n}\mu_G\right)\otimes I_1\right][M \otimes Z_o] \\ = \frac{1}{2}\left[I_G - \frac{1}{n}\mu_G M\right]\otimes Z_o \tag{14}$$

where,

$$B = T - W = \frac{1}{2}\left[\left(M^{-1} - \frac{1}{n}\mu_G\right)\otimes I\right] \text{ and}$$

$$Z = M \otimes Z_o$$

where,

Z_o is defined in (12).

The characteristic roots of BZ under H_0 are then one half the products of the characteristic roots of

$$I_G - \tfrac{1}{n}\cup_G M \tag{15}$$

With the characteristic roots of Z_o, the characteristic roots of (15) are the solutions to the determinantal equation:

$$\left| I_G(I - \lambda) - \frac{1}{n}\cup_G M \right| = 0 \tag{16}$$

The left-hand side of (2.7) is:

$$L.H.S. = \lambda(I - \lambda)^{G-1}$$

Hence, $\lambda = 0$ is a root withy multiplicity one and $\lambda = 1$ is a root with multiplicity $G - 1$.

Thus, (2.4) is really

$$BSS \approx \frac{1}{2}\sum_{i=1}^{I}\lambda_i \chi^2_{G-1} \qquad (17)$$

where,

$\{\lambda_i, i = 1,\dots, I\}$ is the set of characteristic roots of Z_o.

Recall that:

$$Z_o = \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & -p_1 p_I \\ & & \vdots \\ & & p_1(1-p_I) \end{bmatrix}. \qquad (18)$$

i.e., Z_o, is a multinomial covariance matrix. Roy et. Al., studied the determination of the characteristic roots of the general multinomial covariance matrix without presenting the roots. The characteristic equation they obtain is:

$$\left\{1 - \sum_{i=1}^{I}\left(\frac{p_i^2}{p_i - \lambda}\right)\right\}\prod_{i=1}^{I}(p_i - \lambda) = 0 \qquad (19)$$

Certainly $\lambda = 0$ is a root. This also follows because a multinomial covariance matrix is singular. The determination of the other roots in general appears to be an unsolved problem. We can make some further observations, however, by noting that (19) is equivalent to:

$$\left|\prod_{i=1}^{I}(p_i = \lambda)\right| - \sum_{i=1}^{I} p_i^2 \left|\prod_{j \neq i}(p_j - \lambda)\right| = 0 \qquad (20)$$

Then;

a. Unless there is a $j \neq i$ such that $p_i = p_j, \lambda = p_i$ is not a root.

b. If no two probabilities $p_i, p_j i \neq j$, are equal, then between each two successively ordered probabilities lies a characteristics roots, i.e., if we order the probabilities and the characteristic roots, we have:

$$p_{i1} < \lambda_{i1} < p_{i2} < \lambda_{i2} < \ldots < p_{i(I-1)} < \lambda_{i(I-1)} < p_{iI}$$

To see this, note that the left-hand side of (20) as a function of λ changes continuously from positive to negative or vise versa as λ changes from p_{ij} to $p_{i(j+1)}$.

c. If k of the probabilities $p_i = i = 1,..., I,$ are equal, i.e.,

$p_{i_r} = p_{i_{(r+1)}} \ldots = p_{i_{(r+k-1)}} = p$, then in (20), $(p - \lambda)^{k-1}$ is a factor and $\lambda = p$ is then a characteristic root with multiplicity $k = 1$.

Unfortunately the set of $\{\lambda_i, i = 1,..., I-1\}$. Suppose, however, that under H_o we approximate the distribution of $BSS \sim \frac{1}{2}\sum_{i=1}^{I-1} \lambda_{i\chi^2 G-1}$,

By (suggested also by Thompson):

$$S \sim \frac{1}{2}\sum_{i=1}^{I-1} \overline{\lambda}_{x^2(I-1)(G-1)}, \qquad (21)$$

where,

$$\bar{\lambda} = \frac{\sum_{i=1}^{I-1} \lambda_i}{I-1} = \frac{traceZ_o}{I-1} = \frac{1 - \sum_{i=1}^{I} p_i^2}{I-1}.$$

From (c) above, if $p_i = ... = p_i = 1/I$ then $\lambda = 1/I$ is a root multiplicity $I-1$ and

$$BSS \sim \frac{1}{2I} X^2_{(I-1)(G-1)}.$$

Thus, under H_o, at the center of the simplex defined by

$$\{(p_i,..., p_i): p_i \geq 0, i = 1,..., I\} \text{ and } \sum_{i=1}^{I} p_i = 1 \qquad (22)$$

The approximation of (17) by (21) is exact. Moreover, for $I = 2$ and all G, the approximation of (17) by (21) is exact.

In general, over the simplex of (22),

$$E_0 BSS = \frac{1}{2}\left(\sum_{i=1}^{I-1} \lambda_i\right)(G-1) = \frac{1}{2}\bar{\lambda}(I-1)(G-1)$$

$$= \quad E\left(\frac{1}{2}\bar{\lambda} x^2{}_{(I-1)(G-1)}\right) = E(S).$$

Thus, the approximation in (17) has the correct first moment under H_0 over the entire simplex. Further, under H_0:

Var (BSS)

$$= \frac{(G-1)}{2} \sum_{i=1}^{I-1} \lambda_i^2 = \frac{1}{2}(G-1) trace(Z_0^2)$$

$$= \frac{1}{2}(G-1)\left[\sum_{i=1}^{I}\left[p_i(1-p_i)^2\right] + 2\sum_{i=1}^{I}\sum_{j=i+1}^{I} p_i^2 p_j^2\right] \qquad (23)$$

$$=\frac{1}{2}(G-1)\left[\sum_{i=1}^{I} p_i^2 - 2\sum_{i=1}^{I} p_i^3 + \left(\sum_{i=1}^{I} p_i^2\right)^2\right].$$

The variance of the approximation is:

$$\text{Var}(S)=\frac{1}{2}(I-1)(G-1)\overline{\lambda}^2$$

$$=\frac{1}{2}\left(\frac{G-1}{I-1}\right)\left(1-\sum_{i=1}^{I} p_i^2\right)^2. \qquad (24)$$

For I =2 and all G, (23) and (24) are exactly equal over the entire simplex defined by (22) because the approximation is exact. Moreover, for all I and G, (23) and (24) are exactly equal at the center of the simplex defined (22), for the same reason. Numerical comparisons between (23) and (24) for various values of I and $(p_i,\ldots,p_n)$ indicate moderate to good agreement "near" the center of the simplex.

For example, with $I=4$ and $p_1=0.2, p_2=0.2,$ and $p_4=0.4,$ then (23) $=\frac{1}{2}(G-1)(.1824)$. Hence, we will approximate the asymptotic distribution of BSS as:

$$BSS \sim \left(\frac{1-\sum_{i=1}^{I} p_i^2}{2(I-1)}\right)\chi^2_{(I-1)(G-1)}. \qquad (25)$$

This still leaves the problem that $\frac{1}{2}\left(1-\sum_{i=1}^{I} p_i^2\right)$ is unknown. The minimum variance unbiased estimator under H_o of $\frac{1}{2}\left(1-\sum_{i=1}^{I} p_i^2\right)= \frac{1}{2}\sum_{i=1}^{I} p_i(1-p_1)$ can be shown to be:

$$\frac{1}{2n}\left(\frac{n}{n-1}\right)\sum_{i=1}^{I}\left(\frac{n_{i+}}{n}\right)\left(I-\frac{n_{i+}}{n}\right)$$
$$=\frac{1}{2(n-1)}\left[\sum_{i} n_{i+}-\frac{\sum_{i} n_{i+}^{2}}{n}\right] \tag{26}$$
$$=\frac{1}{2(n-1)}\left[n-\frac{1}{n}\sum_{i=1}^{I} n_{i+}^{2}\right]$$
$$=TMS$$

We already knew that $I_o(TMS)=\frac{1}{2}\left(1-\sum_{i=1}^{I} p_i^2\right)$

Since $n\frac{n_{i+}}{n}$ coverage's in probability to p_i under H_o, the asymptotic variance of TMS is zero.

Thus, under H_o, we view TMS as a constant equal to $\frac{1}{2}\left(1-\sum_{i=1}^{I} p_i^2\right)$ then,

$$C=\frac{(n-1)(I-1)BSS}{TSS}$$
$$=(n-1)\,(I-1)\left[\frac{\left(\sum_{j=1}^{G}\frac{1}{n+j}\sum_{i=1}^{I} n_{ij}^{2}\right)-\frac{1}{n}\sum_{i=1}^{I} n_{i+}^{2}}{n-\frac{1}{n}\sum_{i=1}^{I} n_{i+}^{2}}\right] \tag{27}$$

Maybe approximated as $\chi^2_{(I-1)(G-1)}$.

Pearson's Chi-square Test

Pearson's chi-square test (χ2) is one of a variety of chi-square tests – statistical procedures whose results are evaluated by reference to the chi-square distribution. It tests a null hypothesis that the relative frequencies of occurrence of observed events follow a specified frequency distribution. The events are assumed to be independent and have the same distribution, and the outcomes of each event must be mutually exclusive. A simple example is the hypothesis that an ordinary six-sided die is "fair", i.e., all six outcomes occur equally often. Pearson's chi-square is the original and most widely-used chi-square test.

Chi-square is calculated by finding the difference between each observed and theoretical frequency for each possible outcome, squaring them, dividing each by the theoretical frequency, and taking the sum of the results. The number of degrees of freedom is equal to the number of possible outcomes, minus 1:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad (28)$$

where,

O_i = an observed frequency;

E_i = an expected (theoretical) frequency, asserted by the null hypothesis;

n = the number of possible outcomes of each event.

A chi-square probability of 0.05 or less is commonly interpreted by applied workers as justification for rejecting the null hypothesis that the row variable is unrelated (that is, only randomly related) to the column variable. The alternate hypothesis is not rejected when the variables have an associated relationship.

Distribution of Pearson's Chi-Square

The null distribution of the Pearson statistic with j rows and k columns is approximated by the chi-square distribution with (k − 1) (j − 1) degrees of freedom. [28] This approximation arises as the true distribution, under the null hypothesis, if the expected value is given by a multinomial distribution. For large sample sizes, the central limit theorem says this distribution tends toward a certain multivariate normal distribution.

Efficiency of Pearson's Chi-square and C

The efficiency of both the χ^2 and the C statistics maybe obtained using the formula given below:

$$E = \frac{2k}{V(C)} \qquad (29)$$

$$E = \frac{2k}{V(\chi^2)} \qquad (30)$$

where k is the degrees of freedom of the r x c contingency table computed as

(r – 1)(c – 1).

Definition of Terms

Aspiration. A strong desire to achieve a goal.

CATANOVA. It is a technique designed to identify the variation between treatments of interest to the researcher. A measure of variation for categorical data.

Categorical. An unordered and discrete variable. It is data that can only be put into unordered groups.

Chi-square Test. A statistics used in the analysis of enumeration data. It reflects discrepancies between the observed and expected or theoretical frequencies of individuals, objects, or events falling in various categories.

Economic Profile. Refers to the respondent's age, sex, civil status, parent's occupation, parent's annual income and etc.

Household size. Refers to the number of family members of the respondents.

Motivation. A factor that encourages a person to pursue or achieve something.

Nominal data. These are categorical data where the order of the categories is arbitrary.

One-way analysis of variance. A procedure for comparing the mean scores of two or more groups based on one categorical independent variable.

Pearson's' Chi-square. It is used to test the hypothesis of no association of columns and rows in tabular data. Also known as the chi-square goodness-of-fit or chi-square test of independence.

Predictor Variables. These are variables from which projections are made in a prediction study. These include the socio-economic profile of the respondents.

Proportion of variation (R^2). Used to measure association.

Response variable. A variable on which information is collected and which there is an interest because of its direct policy relevance.

Test of goodness of fit. It is used to test if an observed distribution conforms to any other distribution. Establishes whether or not an observed frequency distribution differs from a theoretical distribution.

Test of independence. Assess whether paired observations on two variables expressed in a contingency table are independent of each other.

Working students. These are youths working their way through college.

METHODOLOGY

Locale of the study

The study was conducted at Benguet State University, La Trinidad, Benguet during the school year 2007-2008.

Respondents of the Study

A sample of 100 students was chosen at random from the College of Agriculture, College Nursing, College of Arts and Sciences, College of Engineering, College of Home Technology and Economics, College of Teacher Education, College of Forestry, and College of Veterinary Medicine.

This study included students working as student assistants within Benguet State University as the sampling units.

Data Collection Instrument

A questionnaire consisting of several questions on the respondents' socio-economic background and questions on their motivations and aspirations for working was developed and pre-tested.

The dependent variables included student's motivations and aspirations and the predictor variables were the socio-economic background of the respondents which included gender, father and mother's occupation, mother and fathers' educational attainment, household size, number of siblings, sibling position and parents' annual income.

Analysis of Data

The responses of the respondents on the different questions were tabulated in an r x c contingency table. The Chi-square test and the CATANOVA were both computed for each contingency table.

Significance of results of the χ^2 and C values were compared to the tabular value at 0.01 and 0.05 level of significance to determine the association of the respondents' socio-economic variables with their motivations and aspirations.

The R^2 which is a measure of the variation in the dependent variable due to the independent variable was likewise determined only in CATANOVA. The efficiency of both the χ^2 and CATANOVA were computed using equation (29) and (30).

RESULTS AND DISCUSSION

Associations Between Educational Motivations and Socio-Demographic Profile of the Working Students

For purposes of illustrating the computation of CATANOVA, the responses of 100 working students were classified into educational motivations by father's educational attainment, by household size, and by sibling's position. As revealed in Table 1, both CATANOVA C and the χ^2 test did not show significant association between educational motivations and father's educational attainment. This result indicates that student's motivation to work while studying

Table 1**.** Association between student's motivations to work as student's assistants and father's occupation and father's education.

DEMOGRAPHIC PROFILE	EDUCATIONAL MOTIVATION		
	Earn Extra Money for Allowance	Develop Interpersonal Skills	Spend Free Time Wisely
Father's Occupation			
Government Employee	3	8	0
Self-Employed	5	3	4
Farming	19	15	13
Laboring	16	8	6
	df=6 χ^2 =10.78,	P=.09, C=10.39	P=.10; R^2(%)=5.2
Father's Education			
Elementary	11	11	8
High School	20	9	13
College	12	14	2
	df=4 χ^2 =9.00,	P=.06, C=8.03	P=.09; R^2(%)=4.05

is weakly motivated by their father's educational attainment. However, the variation in the student's motivation to work attributed to their father's educational attainment was about 4.05 percent. The father's occupation was likewise weakly associated to their children's motivation to work while studying. As revealed by both tests, no significant association between father's educational attainment and father's occupation and student's motivations to work.

As shown in Table 2, chi-square test did not show significant association between educational motivations by household size. However, in the C test, there is a significant association between the 2 variables. This result indicates that student's motivation to work was motivated by household size. Although statistically associated to student's educational motivation, household size contributed only 2.30% of the variation in the student's educational motivations. For the association between sibling's position and students' educational motivation, both χ^2 and C tests revealed significant results. These findings indicate that student's motivation to work is strongly motivated by sibling's position. The variation in the students' motivation to work is attributed to their sibling's position by about 5.50 percent.

Table 2, shows significant association between students' motivation and household size and between student's motivation and sibling's position using the C test.

Table 2. Association between student's motivations to work as student's assistants and household size and sibling's position.

DEMOGRAPHIC PROFILE	EDUCATIONAL MOTIVATION		
	Earn Extra Money for Allowance	Develop Interpersonal Skills	Spend Free Time Wisely
Household Size			
1-5	9	6	5
6-9	18	21	13
10 or more	16	7	5
	df=4 χ^2 =4.00^{ns},	P=.40, C=9.23*	P=.05; R^2(%)=2.30
Sibling's Position			
Eldest	19	10	3
Middle	11	17	14
Youngest	13	7	6
	df=4 χ^2 =10.56*,	P=.03, C=10.86*	P=.02; R^2(%)=5.50

* = significant ns = not significant

These results suggest that both household size and sibling position had significant contribution on the student's to work as student assistants 2.30 % of the variation in the student's educational motivations.

<u>Association Between Educational Aspirations and Socio-Demographic Profile of the Working Students</u>

Table 3 shows both C and the χ^2 test showed strong evidence to declare significant association between student's educational aspirations and gender and between student's educational aspiration and parent's annual income.

The results indicate that student's aspiration to work while studying are strongly attributed to their gender and parent's annual income. Gender explained

Table 3. Association between student's aspirations to work as student's assistants and gender and family income.

DEMOGRAPHIC PROFILE	EDUCATIONAL ASPIRATION	
Gender	Finish a Degree	Preparation for Future Job
Male	14	21
Female	41	24
df=1 χ^2=4.90*, p=.03 C=4.85*, p=.03 R^2(%)=4.9		
Family Income		
30,000 and below	6	12
31,000 to 50,000	28	10
51,000 to 80,000	10	10
81,000 to 100,000	7	6
101,000 or more	4	7
df=4 χ^2=10.53*, p=.03 C=10.42*, p=.03 R^2(%)=10.53		

* - significant

4.9 % of the variation in student's aspiration while parent's income explained 10.53 % variation in the student's aspiration to work.

<u>Efficiency of CATANOVA and the Pearson's Chi-square Test in Measuring Associations</u>

Table 4 presents the efficiency of the Pearson's chi-square and CATANOVA C association tests for two variables cross-tabulated in contingency tables with different dimensions. For a 4 x 3 contingency table, the variability of the Pearson's chi-square statistic and the C statistics considered higher than the expected variance of 2k for the theoretical χ^2 distribution.

Table 4. Efficiency of CATANOVA and Chi-square Test with 100 Sample Size

Motivation Associated to	Dimension	Efficiency of Pearson's Chi-square $\frac{2k}{V(\chi^2)}$	Efficiency of CATANOVA $\frac{2k}{V(C)}$
Father's Occupation	4 x 3	0.99	0.98
Father's Educ'l Attainment	3 x 3	1.01	1.00
Household Size	3 x 3	1.01	1.00
Sibling's Position	3 x 3	1.01	1.00
Aspiration Associated to			
Gender	2 x 2	1.06	1.06
Parents Annual Income	5 x 2	1.04	1.04

The efficiency of Pearson's Chi-square in a 3 x 3 contingency table was found higher than the efficiency of the C statistic. This means that the Pearson's chi-square statistics is more variable than the theoretical χ^2 distribution.

SUMMARY, CONCLUSION AND RECOMMENDATION

Summary

Both chi-square and C statistics were computed to determine the significance of association between the response and independent variables.

Based on the results of the analysis employing both CATANOVA and Pearson's Chi-square, students' motivations was associated to the father's occupation, house hold size and sibling's position. Aspiration of working students was associated to gender and family income. With the use of CATANOVA, the percent contribution of the socio-economic variables on the variability of the working students' motivations and aspirations were determined.

CATANOVA statistics is at par with the Pearson's Chi-square statistics in determining associations but more efficient than Pearson's Chi-square in terms of their variability.

Conclusion

Based on the above results, the following conclusions were drawn:

The father's occupation, educational attainment, household size and sibling position had significant bearing on the respondent's motivation to work.

The respondent's aspirations to earn a degree were explained by their gender and their family income.

CATANOVA as a tool for determining association is at par with the Chi-square test but considered more efficient than the Pearson's Chi-square.

Recommendations

In dealing with qualitative data where the response variable is categorized with no order CATANOVA C is recommended for used because it does not only measure the relationship between the response and predictor variables it also measures the percent contribution explained by the predictor variables.

LITERATURE CITED

BALATERO, R.E. 2003. Factors Affecting Academic Performance of Working Students of Benguet State University, La Trinidad Benguet.

COBB, R. A., MCINTIRE, W.G., and PRATT, P. A.1989. Vocational and educational aspirations of high school students: A problem for rural America. In R. Quaglia (Ed.), Research in Rural Education

COCHRAN, W.G. (1950). "The Comparison of Percentages in Matched Samples," Biometrika.

D'AMBRA, ANTONELLO and PASQUALE SARNACCHIARO, 2007. Explorative Data Analysis and CATANOVA for Ordinal Variables: An Integrated Approach

HOEFFDING, W.1965. "Asymptotically Optimal tests for Multinomial Distributions," Annals of Mathematical Statistics.

LIGHT, R.J. MARGOLIN, H. B. 1969. "Analysis of Variance for Categorical Data, with Applications to Agreement and Association," Unpublished Ph.D. dissertation, Department of Statistics, Harvard University.

MACLI-ING and GUIMPAYAN. 2001. Factors Affecting Academic Performance of Working Students in McDonalds Center mall. Undergraduate Thesis. Benguet State University, La Trinidad Benguet.

PEARSON, K.1900. "On the Criterion that a Given System of Deviation from the probable in the Case of the Correlated System of Variables is Such that it can Be Reasonably Supposed to have Arisen from Random Sampling," Philosophical Magazine.

RJ ANDERSON, JR LANDIS, 1980. Communications in Statistics-Theory and Methods, informaworld.com

WALBERG, H. J. 1989. Student aspirations: National and international perspectives. In R. Quaglia (Ed.), Research in Rural Education.

APPENDICES

Appendix A: Letter of Communication

Benguet State University
College of Arts and Sciences
La Trinidad, Benguet

MR. ROMEO BABARAN
Manager
Jollibee La Trinidad
La Trinidad, Benguet

We the undersigned are students from Benguet State University taking up Bachelor of Science in Applied Statistics. We are conducting our research entitled "Efficiency of CATANOVA in Measuring Association Between Socio-economic Variables and Motivations and Aspirations of Working students." This is to comply with the requirements of the course.

In this connection, may we request for permission that we be allowed to administer our questionnaire to the working students of Benguet State University who are having part time job in your fast food establishment.

Your favorable action for this request is highly appreciated.

Thank you.

Researchers,

ELVIRA D. DAS-ILEN

DIXY E. PADILAN

Noted by:

SALVACION Z. BELIGAN
Adviser

Recommending Approval:

MARIA AZUCENA B. LUBRICA
Chairman-MPS Department

AUREA MARIE M. SANDOVAL
CAS-Dean

Benguet State University
College of Arts and Sciences
La Trinidad, Benguet

The Manager
McDonald La Trinidad
La Trinidad, Benguet

Greetings!

We the undersigned are students from Benguet State University taking up Bachelor of Science in Applied Statistics. We are conducting our research entitled "Efficiency of CATANOVA in Measuring Association Between Socio-economic Variables and Motivations and Aspirations of Working students." This is to comply with the requirements of the course.

In this connection, may we request for permission that we be allowed to administer our questionnaire to the working students of Benguet State University who are having part time job in your fast food establishment.

Your favorable action for this request is highly appreciated.

Thank you.

Researchers,

ELVIRA D. DAS-ILEN

DIXY E. PADILAN

Noted by:

SALVACION Z. BELIGAN
Adviser

Recommending Approval:

MARIA AZUCENA B. LUBRICA
Chairman-MPS Department

AUREA MARIE M. SANDOVAL
CAS-Dean

Appendix B: Sample Survey Questionnaire

Benguet State University
College of Arts and Sciences
La Trinidad, Benguet

Dear Respondents:

Greetings!

We the undersigned are students from Benguet State University taking up Bachelor of Science in Applied Statistics. We are conducting our research entitled "Efficiency of CATANOVA in Measuring Association Between Socio-economic Variables and Motivations and Aspirations of Working students." This is to comply with the requirements of the course.

In this connection, may we solicit your valued cooperation in answering the following questions honestly. Rest assured, your answers will be kept confidential.

Thank you and more power.

Researchers,

ELVIRA D. DAS-ILEN

DIXY E. PADILAN

Noted by:

SALVACION Z. BELIGAN
Adviser

DIRECTION: Please check or write on the blank that corresponds to your answer

Personal Data:

Name (optional): _______________ Gender: _____

Age: _____ Civil Status: ___

Degree/Course: ____________________ Year Level: _____

Number of Units: ________

A. Socio-economic Profile:

Direction. Please write your answer on the blank provided for the needed information.

1. Parents occupation. __________ Mother __________ Father

2. Parents highest educational attainment. __________ Mother ________ Father

3. Household size (# of family members). _____

4. Number of siblings in the family (including you): _____

5. Your sibling position. _____

6. Annual income of parents (just encircle your answer on the choices given below).

a. below 30,000 b. 31,000 to 50,000

c. 51,000 to 80,000 d. 81,000 to 100,000

e. 101,000 and above

B. On Motivation and Aspiration:

Direction. Please encircle only your main reason why you work while studying.

Motivations:

a. Augment the limited allowance from parents.

b. To develop interpersonal skills.

c. To spend free time wisely.

Aspirations:

a. To finish a degree.

b. For job preparation.

Appendix C: Variables Raw Data

Respondents	Gender	Father's Occupation	Mother's Occupation	Father's Educational Attainment	Mother's Educational Attainment	Household Size	Number of Siblings	Sibling Position	Annual Income Of Parents	Motivation	Aspiration
1	1	3	5	2	2	3	3	3	2	1	2
2	1	3	3	2	3	3	3	3	5	1	2
3	2	4	1	1	3	1	1	3	1	3	2
4	1	4	2	2	3	3	2	3	5	1	1
5	2	4	5	1	3	2	3	3	5	2	2
6	2	3	3	2	3	2	2	3	2	3	1
7	1	3	5	1	2	2	2	1	1	1	1
8	1	3	4	2	2	2	2	2	3	3	2
9	2	3	5	1	3	1	1	1	1	2	1
10	2	3	2	1	2	3	3	2	2	1	1
11	2	3	3	3	3	2	2	2	5	3	2
12	2	3	5	1	2	2	1	1	2	1	1
13	2	1	5	3	3	2	2	3	4	2	1
14	1	4	5	2	2	3	3	1	1	2	2
15	2	3	4	1	3	2	1	1	3	2	1
16	1	3	3	2	3	2	2	2	1	1	1
17	2	3	5	3	3	2	2	2	3	3	2
18	2	4	5	3	3	1	1	1	2	1	1
19	2	4	2	2	3	2	1	2	2	2	2
20	2	1	5	3	2	2	1	2	4	2	2
21	2	2	4	3	3	2	2	2	4	2	2
22	2	4	5	2	3	1	1	1	2	2	1
23	2	3	3	2	3	1	1	3	1	1	1
24	1	3	3	3	3	2	1	2	1	2	2
25	1	4	5	1	3	3	3	3	2	1	2
26	2	1	5	3	3	1	1	1	2	1	1
27	2	3	5	2	3	3	3	1	1	2	2
28	2	3	4	2	3	2	1	2	3	3	1
29	2	4	5	3	3	2	2	2	2	1	1
30	2	3	5	3	2	2	2	2	4	2	1
31	2	4	5	2	3	3	2	1	2	1	1

32	2	3	5	1	2	2	2	2	2	1	1
33	2	4	5	3	2	3	3	2	3	1	1
34	2	4	2	1	1	1	2	1	1	3	2
35	1	4	5	2	3	3	2	1	1	1	2
36	1	2	4	3	3	3	3	2	5	1	2
37	2	1	1	3	3	1	1	2	3	2	1
38	1	2	5	2	3	2	2	2	1	3	2
39	1	3	3	1	2	2	1	3	1	2	1
40	1	1	1	3	3	2	1	1	2	2	1
41	2	2	2	2	2	3	1	1	2	2	1
42	2	4	5	2	2	3	3	2	2	1	1
43	1	3	4	1	2	3	3	2	2	3	2
44	2	4	5	3	3	2	2	2	3	2	2
45	1	3	1	3	3	2	2	2	5	2	2
46	2	3	3	1	3	2	1	2	3	3	2
47	2	2	3	1	1	2	2	3	4	3	1
48	2	2	1	3	3	3	3	1	5	1	1
49	1	3	5	1	2	2	3	3	4	2	2
50	1	1	5	3	3	3	2	1	5	2	2
51	1	3	4	2	2	2	2	1	4	1	1
52	2	3	5	2	3	1	1	2	3	3	2
53	2	2	4	2	3	2	2	2	2	1	1
54	2	3	3	1	1	3	3	3	4	1	1
55	2	3	2	2	3	1	1	2	3	1	2
56	1	4	5	1	2	2	1	3	2	2	2
57	1	3	3	1	2	1	1	2	3	2	2
58	2	2	1	3	3	3	2	1	4	1	2
59	1	1	2	3	3	3	3	2	5	2	1
60	1	3	5	2	3	2	1	2	5	1	2
61	1	2	5	3	3	2	2	1	3	1	1
62	1	3	4	2	2	2	2	1	2	3	1
63	2	3	3	2	3	3	3	2	1	2	2
64	2	4	3	1	3	3	2	3	2	3	2
65	2	3	5	1	3	1	1	3	2	1	2
66	2	4	5	3	1	2	1	2	3	2	1
67	2	2	2	2	2	3	3	2	4	3	1
68	2	3	2	2	2	2	2	1	4	1	2
69	1	4	4	1	3	3	2	2	2	3	1
70	2	4	5	3	3	2	1	1	2	1	1
71	2	4	5	2	2	2	2	2	2	2	1
72	1	3	3	1	3	2	3	2	1	2	2
73	2	3	5	2	3	2	2	2	2	3	1
74	2	3	3	1	2	2	2	1	1	1	2
75	2	4	4	2	3	2	2	1	2	1	1
76	2	3	5	2	3	1	1	1	2	1	1

77	1	4	5	2	3	2	1	1	2	1	1
78	2	4	5	2	2	1	2	3	2	1	1
79	2	4	1	2	3	1	1	2	3	1	1
80	2	3	4	1	3	3	3	3	4	1	1
81	2	3	5	1	2	1	1	2	2	2	1
82	1	3	5	2	3	2	1	2	2	3	1
83	2	2	5	2	3	2	1	1	2	3	2
84	2	2	2	3	3	3	3	2	2	2	2
85	2	1	1	3	3	2	2	3	3	1	1
86	2	3	5	2	3	1	1	3	5	1	1
87	2	3	3	1	2	1	1	2	2	3	1
88	2	3	3	1	2	1	1	3	2	2	1
89	2	1	5	3	3	2	1	1	4	2	2
90	1	3	5	2	3	1	1	3	2	3	2
91	1	5	3	1	2	3	2	3	3	1	2
92	1	1	2	2	2	2	2	1	3	1	1
93	1	1	2	2	2	2	2	1	3	2	2
94	2	3	5	2	3	2	3	2	1	2	1
95	2	4	5	3	3	2	1	1	2	1	1
96	1	4	4	2	3	2	1	3	1	3	2
97	1	3	3	1	2	2	1	3	2	2	1
98	2	4	5	2	3	3	2	2	3	3	1
99	2	3	1	3	3	2	2	3	3	1	2
100	2	4	4	1	3	3	1	1	1	1	2

Legend:

Gender: 1 = male

2 = female

Parent's occupation:

1 = government employee

2 = self-employed

3 = farming

4 = laboring

5 = housekeeping

Parent's educational attainment:

1 = elementary

2 = high school

3 = college

Household size:

1 to 5 members = 1

6 to 9 members = 2

10 and above members = 3

Number of Siblings**:**

1 to 4 siblings = 1

5 to 8 siblings = 2

9 and above siblings = 3

Siblings Position:

Eldest = 1

Middle = 2

Youngest = 3

Parent's Annual income:

30,000 and below = 1

31,000 to 50,000 = 2

51,000 to 80,000 = 3

81,000 to 100,000 = 4

101,000 and above = 5

Motivations:

1= to augment the limited from parents

2= to develop interpersonal skills

3=to spend time productively

Aspirations:

1= to finish a degree

2= preparation for future job

Appendix D: Sample Computations

Table 1. Gender-Aspirations Cross Tabulation

	Gender		
Aspirations	Male	Female	Total
1	14	41	55
2	21	24	45
Total	35	65	100

TSS= $\frac{n}{2}-\frac{1}{2n}\sum_{i=1}^{I} n_{i+}^{2}$

WSS= $\frac{n}{2}-\frac{1}{2}\sum_{j=1}^{G}\frac{1}{n_{+j}} n_{ij}^{2}$

BSS= TSS – WSS

R^2= BSS/TSS

C= (n-1)(I-1) R^2

where,

n= total number of observation.

n_{ij}=the count in the ij^{th} cell

n_{i+}= the total count in the i^{th} cell

n_{+j}= the total in the j^{th} cell.

TSS= $\frac{100}{2}-\frac{1}{2(100)}\left[\left(55^2+45^2\right)\right]$

=50-25.25

=24.75

$$\text{WSS}= \frac{100}{2} - \frac{1}{2}\left[\left(\frac{14^2 + 21^2}{35}\right) + \left(\frac{41^2 + 24^2}{65}\right)\right]$$

= 50 -0.05(18.2 + 34.723077)

= 50 – 26.4615385

= 23.5384615

BSS= 24.75 – 23.5384615

= 1.211538

$$R^2 = \frac{1.211538}{24.75}$$

= 0. 048951

C= (100-1)(2-1)(0.048951)

= 4.846154

The efficiency is between gender and motivation.

$$\text{Eff}(\chi^2) = \frac{Approximated}{Observed(\chi^2)}$$

$$= \frac{4.000}{3.758}$$

=1.06

$$\text{Eff(C)} = \frac{Approximated}{Observed(C)}$$

$$= \frac{4.000}{3.766} \qquad = 1.06$$

Appendix E: EMPIRICAL SAMPLING UNDER Ho

ANOVA for Categorical Data

[a] = Approximated [b] = Observed CATANOVA [c] = Observed CHI-SQUARE

Simulation number	yi	G	Group sample size	Mean	Variance	PERCENTILE		
						Reference percentile		Empirical critical value
1	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	2	100 all	2.000[a]	4.000a	.99	.890	4.440
				1.973[b]	3.766b	.90	.944	5.763
				1.982[c]	3.758c	.95	.940	9.238
2	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	2	50 all	2.000[a]	4.000[a]	.90	.890	4.459
				1.994[b]	3.889[b]	.99	.951	6.034
				2.027[c]	3.918[c]	.99	.989	9.092
3	(.70,.15,.13)	2	30 all	2.000[a]	4.000[a]	.90	.903	4.720
				2.027[b]	4.160[b]	.95	.955	6.211
				2.077[c]	3.826 [c]	.99	.991	9.463
4	(.70,.15,.15)	2	60 all	2.000 [a]	4.000[a]	.90	.890	4.433
				1.951 [b]	3.940[b]	.95	.947	3.870
				1.963 [c]	3.380[c]	.99	.992	9.854
5	(.70,.15,.15)	2	100all	2.000[a]	4.000[a]	.90	.900	4.606
				1.951[b]	3.940[b]	.95	.955	6.149
				1.963[c]	3.380[c]	.99	.990	9.310
6	$\left(\frac{1}{2},\frac{1}{4},\frac{1}{4}\right)$	2	15all	2.000[a]	4.000[a]	.90	.895	4.528
				2.024[b]	4.224[b]	.95	.956	6.265
				2.078[c]	3.911[c]	.99	.992	9.770
7	(.90,.03,.03)	2	80all	2.000[a]	4.000[a]	.90	.904	4.710
				1.986[b]	3.989[b]	.95	.955	6.166

				2.013[c]	3.033[c]	.99	.990	9.126
8	(.48,.48,.04)	2	80all	2.000[a]	4.000[a]	.90	.916	4.973
				2.020[b]	5.498[b]	.95	.968	6.970
				2.101[c]	3.399[c]	.99	.995	10.619
9	(.30,.30,.40)	2	50all	2.000[a]	4.000[a]	.90	.895	4.496
				1.889[b]	3.562[b]	.95	.942	3.671
				1.929[c]	3.752[c]	.99	.982	8.085
10	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	2	20,40	2.000[a]	4.000[a]	.90	.896	4.569
				1.954[b]	3.822[b]	.95	.952	6.117
				1.989[c]	3.973[c]	.99	.988	9.019
11	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	2	30,100	2.000[a]	4.000[a]	.90	.890	4.422
				1.940[b]	3.602[b]	.95	.938	3.447
				1.957[c]	3.587[c]	.99	.986	8.688
12	(.70,.15,.15)	2	20,40	2.000[a]	4.000[a]	.90	.900	4.618
				2.029[b]	4.061[b]	.95	.945	5.805
				2.051[c]	3.446[c]	.99	.990	9.306
13	(.90,.05,.05)	2	20,40	2.000[a]	4.000[a]	.90	.902	4.642
				2.014[b]	4.024[b]	.95	.954	6.182
				2.092[c]	3.867[c]	.99	.970	9.344
14	$\left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{2}{5}\right)$	2	50all	3.000[a]	6.000[a]	.90	.901	6.298
				2.919[b]	6.095[b]	.95	.947	7.684
				2.963[c]	5.725[c]	.99	.989	11.180
15	$\left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right)$	2	50all	4.000[a]	8.000[a]	.90	.895	7.619
				3.999[b]	7.959[b]	.95	.947	9.260
				4.062[c]	7.926[c]	.99	.987	12.813
16	(.30,.30,.40)	3	50all	4.000[a]	8.000[a]	.90	.897	7.678
				4.012[b]	8.005[b]	.95	.950	9.460

				4.038[c]	7.954[c]	.99	.988	13.095
17	$\left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{2}{5}\right)$	3	80all	6.000[a]	12.000[a]	.90	.903	10.868
				6.116[b]	12.190[b]	.95	.949	12.544
				6.155[c]	12.113[c]	.99	.984	15.949
18	(.16,.16,.16,.16,. 16,.20)	2	100all	3.000[a]	10.000[a]	.90	.900	9.266
				3.035[b]	9.967[b]	.95	.954	11.391
				5.060[c]	9.854[c]	.99	.988	14.688
19	(.20,.20,.20,.40)	4	80all	9.000[a]	18.000[a]	.90	.895	14.283
				8.908[b]	17.054[b]	.95	.945	16.621
				8.971[c]	16.336[c]	.99	.986	21.064
20	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	4	30all	6.000[a]	12.000[a]	.90	.926	11.770
				6.280[b]	12.860[b]	.95	.961	13.357
				6.347[c]	13.018[c]	.99	.986	16.128
21	$\left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right)$	3	100all	8.000[a]	16.000[a]	.90	.890	12.986
				7.967[b]	15.887[b]	.95	.940	15.047
				7.992[c]	16.028[c]	.99	.987	19.580
22	(.30,.30,.40)	3	30all	8.000[a]	16.000[a]	.90	.890	12.993
				7.794[b]	15.389[b]	.95	.940	15.027
				7.815[c]	15.438[c]	.99	.981	18.554
23	$\left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right)$	4	100all	12.000[a]	24.000[a]	.90	.908	18.873
				12.200[b]	25.296[b]	.95	.956	21.569
				12.198[c]	24.856[c]	.99	.993	27.498
24	(.10,.10,.10,.10,. 10,.10,.10,.30)	6	100all	35.000[a]	70.000[a]	.90	.918	47.617
				35.035[b]	80.488[b]	.95	.958	51.247
				34.940[c]	69.204[c]	.99	.991	59.326

Source: Journal of the American Statistical Association, September 1971