COMPARISON OF DIFFERENT MEASURES OF CORRELATION FOR CATEGORICAL DATA

Researcher:

ASAD UL ISLAM KHAN

Reg. No. 25-SE/MS.Et/F09

Supervisor:

Prof. Dr. Asad Zaman

Co Supervisor:

Dr. Atiq-ur-Rehman

School of Economics

International Institute of Islamic Economics

INTERNATIONAL ISLAMIC UNIVERSITY

ISLAMABAD.

...orical data - Economics

COMPARISON OF DIFFERENT MEASURES OF CORRELATION FOR CATEGORICAL DATA

Asad ul Islam Khan

Reg. No. 25-SE/MS.Et/F09

Submitted in partial fulfillment of the requirements for the

Master of Philosophy in Econometrics

at International Institute of Islamic Economics

International Islamic University, Islamabad

Supervisor: Dr. Asad Zaman

Co Supervisor: Dr. Ateeq-ur-Rehman

January 2012

APPROVAL SHEET

Comparison of Different Measures of Correlation for Categorical Data

by

Asad ul Islam Khan

Reg. No: 25-SE/MS.Et/F09

Accepted by the International Institute of Islamic Economics, International Islamic University, Islamabad, as partial fulfillment of the requirements for the award of degree of MASTER OF PHILOSOPHY in ECONOMETRICS.

Supervisor: ____________________ Co-Supervisor: ____________________

Dr. Asad Zaman Dr. Atiq-ur-Rehman

____

Dr. Asad Zaman

Director General (DG)

International Institute of Islamic Economics,

International Islamic University,

Islamabad.

DECLARATION

I hereby declare that this thesis, neither as a whole nor as a part thereof, has been copied out from any source. It is further declared that I have carried out this research by myself and have completed this thesis on the basis of my personal efforts under the guidance and help of my supervisor. If any part of this thesis is proven to be copied out or earlier submitted, I shall stand by the consequences. No portion of work presented in this thesis has been submitted in support of any application for any other degree or qualification in International Islamic University or any other university or institute of learning.

Asad Ul Islam Khan


بسم الله الرحمن الرحيم

ABSTRACT

"Comparison of Different Measures of Correlation for Categorical Data"

Testing of independence is widely and frequently used by the practitioners/researchers in every field of science. Some tests of independence are used for continuous and categorical data while there are some tests that can be used only for categorical data. For the continuous data there is a lot of literature on comparison of different measures of association, but for the categorical data we were unable to find any comparison of various measures of independence. This study compares four measures of correlation/tests of independence for categorical data, categorized into 2×2 contingency table on the basis of size of test and stringency criterion. We found that Fisher's exact test of independence (1934) is the robust and best test of all the four tests of independence/measures of correlation for categorical data.

ACKNOWLEDGEMENTS

All praises are to Allah, Almighty, Peace and Blessing of Allah on his last Prophet Muhammad who taught the humanity to human beings.

All achievements by me are nothing but a blessing of Allah. I am most obliged to my parents, sisters, brothers, bhabi, Dr Atiq-ur-Rehman and Dr Asad Zaman who made it possible for me to earn this degree.

At this occasion, I am thankful to all my teachers who taught me at any stage of my studies. I would acknowledge the teaching of Mr. Mumtaz Ahmed and Mr. Faiz-ur-Raheem for their kind style of teaching. I must acknowledge cooperation and support of my teachers and seniors including Iftikhar Hussain Adil and Saud Ahmed Khan. I would like to acknowledge many friends and colleagues who helped me during my M.Phil and in my early career and provided me excellent company to enjoy. I would thank all my class fellows during M.Phil including Khan Bahadar, Mehmood Hussan, Irfan Malik, Ali Raza, Yasir Riaz, Shahid Akbar, Rana Muhammad Imran, Gulfam Haider, Waqar Muhammad Khan, Imad Ullah, Zeshan Anwar, Asim Latif, Abdul Ghafar and Abdul Waheed whose company I always enjoyed. I also acknowledge the cooperation of the staff of IIIE including Syed Niaz Ali Shah, Zaheer Sahib, Tauqeer, Farooque & Rashid Rauf. Many thanks are also due to my class fellows Ali Raza and Shahid Akbar for cooperation in preparing initial draft of this thesis. Thanks are also due to my brother Murad Ul Islam Khan, sisters and wife for their continuous cooperation. Last but not the least I am thankful to my parents for their prayers, sacrifices, physical contribution, patience and immortal love that encouraged me at every flash of study period.

Asad ul Islam Khan

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Brief Introduction

Tests of Independence are part of tools of every social scientist. These tools are frequently used as a routine practice in Economics, Econometrics, Statistics, Social science and Biological sciences. The data on which these tools are to be applied is of various types and nature. The data may be of continuous or qualitative type. The continuous data can be further divided into ratio and interval. Similarly the qualitative data can be further divided into Nominal, Ordinal and Rank-Order. Keeping view all the types of data the researcher must be careful in conducting the test of independence because there are different tests available for different types of data. But usually the researchers do not take care of the data type and use Pearson Product Moment Coefficient of Correlation. If the data is continuous and it satisfies normality assumptions then there is no problem with it. But if the data is qualitative then the use of Pearson Product Moment Coefficient of Correlation is not appropriate and may produce wrong results. There is huge amount of literature that can lead researcher to appropriate choice of test if the data is continuous.

There are several tests of independence which can be applied to qualitative data. For example Spearman rank correlation coefficient and Kendall tau coefficient of rank correlation can be applied to rank-order data, Goodman and Kruskal's gamma measure of association and chi-square test of independence can be applied to ordinal or categorical data and chi-square test of independence and Fisher's exact test can be applied to nominal data.

For continuous type of data or the rank-order type of data the comparison of correlation coefficients exist in literature. For example Tahani A. Mature and Elsayigh (2010) compared ten correlation coefficients. Similarly Jann Hauke and Tomasz Kossowski compared Pearson and Spearman correlation coefficient for the same sets of data. Cornbleet and Shea (1978) compared product moment coefficient of correlation and rank correlation coefficients. Barnhart et.al compared the concordance correlation coefficient and coefficient of individual agreement.

So there is a guideline available for comparison and usage of a test of independence/measure of correlation for continuous data in literature. But there is no guideline for the comparison of tests of independence/measures of correlation for categorical data.

Nobody (to the best of my knowledge) has compared the coefficients of correlation/ tests of independence for categorical data. Especially for Ordinal or Nominal type of data categorized in a contingency table I was unable to find any comparison. So in order to fill this gap in literature this study has been conducted. In this study four tests of independence/measures of correlation are compared for the simplest type of contingency table i.e 2×2 contingency table. The four tests of independence/measures of correlation that are compared in this study are

(i) Goodman & Kruskal's Gamma measure of association

(ii) Chi Square test of independence

(iii) Pearson product moment coefficient of correlation

(iv) Fisher's exact test of independence

1.2 Motivation

Once I was being asked that which test of independence may be used for categorical data from the available tests of independence in literature. Then I searched for the comparison of tests of independence/measures of correlation for categorical data using the best search facilitators like Google and I was astonished that no such comparison is there. This led me to think of investigating this problem myself.

As there are various tests of independence for categorical data so one can apply any one of them and it is commonly accepted that the conclusions drawn from different tests of independence/measures of correlation will be same. But in reality different tests of independence lead to different conclusion. For example Young and Winn (2003) noted the sighting of Gymnothorax Moringa and G. vicinus (species of eel) in an area of 150 by 250 meters. The two variables of classification were the species of eel and the type of habitat. Young and Winn used Chi-square test of Independence (Chi-square = 6.26 and p-value = 0.044) and concluded that the two variables of classification are dependent as the p-value of chi-square was less than nominal value of 0.05. When I used the same data to test the null hypothesis of independence using Goodman and Kruskal's gamma measure of association, I cannot reject the null hypothesis of independence (Gamma = -0.142, Z = -1.61 and p-value = 0.1074) as the p-value of gamma is greater than the nominal value of 0.05.

In such a situation one cannot decide that which test of independence /measure of correlation may be used or which test of independence/measure of correlation will performs best, unless one has an idea about the size and power of these tests. This motivated and inspired me to conduct this study so that there may be a crystal clear most appropriate test of independence/measure of correlation for categorical data.

1.3 Objectives of the study

There are various tests of Independence/measures of Correlation for data categorized in a contingency table available in the literature. Different tests can lead to different conclusion and nobody knows which of the tests should be used for the data in hand. Therefore objective of this study is to find out a test of independence which gives the best size and power properties for this type of data.

1.4 Significance of the study

As the usage of the test of independence/measures of correlation for categorical data is common practice in every field of science, so this study will help the practitioners/researchers of these fields enabling them to apply the most appropriate test of independence/measure of correlation. This study will also help them in updating their knowledge about the empirical performance of the four tests of independence/measures of correlation for categorical data.

Chapter 2

LITERATURE REVIEW

As a result of extensive search on the internet it is found that the tests of Independence/measures of Correlation for categorical data are not compared in the literature (to the best on my knowledge). However the comparisons of tests of Independence/measures of Correlation are found in literature for continuous data and for rank order data. One of the recent comparisons in which two measures of correlation for rank order data are also compared is of Maturi and Elsayigh (2010). They compared ten correlation coefficients using a three step bootstrap procedure. The ten correlation coefficients compared by them were

(i) Pearson product moment "r"

(ii) Spearman's rho "ρ"

(iii) Kendall's tau "τ"

(iv) Spearman's Foot rule "Ft"

(v) Symmetric Foot rule "C"

(vi) The Greatest deviation "Rg"

(vii) The Top-Down "rT"

(viii) Weighted Kendall's tau "τw"

(ix) Blest "ν"

(x) Symmetric Blest's coefficient "v^*".

They used bootstrap method for their comparison. In order to decide that how many replications should be used to get the most accurate results, they used a three step bootstrap approach which was introduced by Andrews and Buchinsky (2002). They used this approach to determine the optimal number of replications for a specified degree of accuracy. They also used this approach to estimate the standard errors of the correlation coefficients without imposing the null hypothesis. They compared the correlation coefficients on basis of standard error criterion that is a correlation coefficient having the lowest standard error was considered as the best by them.

Three-step bootstrap approach can be used to determine the number of replications with pre-fixed degrees of accuracy which is applied for many different bootstrap problems, such as estimating the standard error, confidence interval and p-value for classical statistical techniques. Accuracy is measured by the percentage deviation of the bootstrap standard error estimate based on N bootstrap simulations from the corresponding ideal bootstrap standard error estimate for which $N = \infty$. A bound of the relevant percentage deviation, *pdb,* is specified such as the actual percentage deviation is less than this bound with a specified probability, $(1-\theta)$ tends to one. That is, for given (pdb, θ), the optimal number of repetitions, N^* satisfies

$$P^*\left(100\frac{\left|\hat{SE}_{N^*}-\hat{SE}_{\infty}\right|}{\hat{SE}_{\infty}}\leq pdb\right)=1-\theta$$

Where $\hat{SE}_{\infty}$ an ideal bootstrap estimate of standard error (SE) is, $\hat{SE}_N$ is the bootstrap approximation of $\hat{SE}_{\infty}$ based on N bootstrap replications. Also P^* represents the probability with respect to randomness in bootstrap samples.

They concluded that "one should use the PEARSON correlation coefficient if the data meets the normality assumption; otherwise, the GREATEST DEVIATION performs well especially when the data has outliers. However, when we want emphasis on the initial (top) data, the SYMMETRIC BLEST'S coefficient has lowest standard error amongst other weighted correlation coefficients".

Chapter 3

INDEPENDENCE IN CATEGORICAL DATA

In this chapter Independence for a $r \times c$ contingency table is defined, the significance of Independence is discussed, independence for 2×2 contingency table is also explained and a brief overview of tests of independence/measures of correlation that are compared in this study is given. At the end of the chapter the procedure with which the data is generated in 2×2 contingency table is described.

3.1 Independence when the data are categorized in $r \times c$ Contingency Table

Let the data on two variables/factors say X and Y is categorized in a $r \times c$ two way contingency table, where r is the number of rows and c is the number of columns. X variable/factor is categorized along r rows and Y variable/factor is categorized along c columns as depicted in Table 3.1.

Table: - 3.1

$r \times c$ Contingency Table

	Y variable/factor				Marginal Probability
X variable/factor	x_{11} (π_{11})	x_{12} (π_{12})	....	x_{1c} (π_{1c})	$\pi_{1\bullet}$
	x_{21} (π_{21})	x_{22} (π_{22})	....	x_{2c} (π_{2c})	$\pi_{2\bullet}$
	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
	x_{r1} (π_{r1})	x_{r2} (π_{r2})	....	x_{rc} (π_{rc})	$\pi_{r\bullet}$
Marginal Probability	$\pi_{\bullet 1}$	$\pi_{\bullet 2}$	....	$\pi_{\bullet r}$	1

where x_{ij} is the number of observations lying in the cell which is the intersection of ith row and jth column and π_{ij} is the corresponding cell probability.

The null hypothesis for any test of independence is that the two factors/variables are independent i.e.

$$H_0: \pi_{ij} = \pi_{i\bullet}\pi_{\bullet j} \quad \text{or} \quad H_0: \pi_{ij} - \pi_{i\bullet}\pi_{\bullet j} = 0 \quad \forall\ i = 1, 2, \ldots, r \text{ and } j = 1, 2, \ldots, c$$

The general alternative hypothesis for any test of independence is that the two factors/variables are not independent i.e.

$$H_A: \pi_{ij} \neq \pi_{i\bullet}\pi_{\bullet j} \quad \text{or} \quad H_A: \pi_{ij} - \pi_{i\bullet}\pi_{\bullet j} \neq 0 \quad \forall\ i = 1, 2, \ldots, r \text{ and } j = 1, 2, \ldots, c$$

where $\pi_{i\bullet}$ is the total of probabilities of ith row i.e

$$\pi_{i\bullet} = \sum_{j=1}^{c} \pi_{ij}$$

and $\pi_{\bullet j}$ is the total of probabilities for jth column i-e

$$\pi_{\bullet j} = \sum_{i=1}^{r} \pi_{ij}$$

3.2 Significance of Tests of Independence

The tests of independence are widely and frequently used in every field of science. But their applications are much more in Medical sciences, Social sciences and Biological sciences. These tests are used to draw the conclusion about the dependence or independence of two variables/factors.

In Medical sciences when a pharmacist wants to know the effect of medicine. He takes the data of subjects before and after the medicine or the groups the subjects into two. One is the group of subjects which has been given the medicine called as treatment group and the other group which has not been given the medicine called as control group. In either of the two situations he uses any of the tests of independence to tests whether the medicine has any effect or not.

In Education if a researcher wants to know the effect of a new technique of teaching, that whether it is effective or not. He takes the grades of students before and after the technique and tests them for independence using any of the tests of independence to know whether the technique is effective or not.

Similarly in Biological sciences if a researcher want to know the effect of a fertilizer. He takes the data before and after the use of fertilizer or groups the plants into two groups. One is the treatment group and the other is control group. Treatment group is that group of plants which are fertilized and the control group is that group of plants which are not fertilized. In either of the situations the researcher uses a test of independence to know whether that fertilizer has a significant effect or not.

All of the above examples show that the tests of independence are frequently used and their result/conclusion is very important. If a test of independence leads to a wrong conclusion then it has very adverse effects on a human life and society.

3.3 Overview of Tests of Independence/Measures of Correlation

A brief overview of each of test of Independence/measure of Correlation is given as under.

3.3.1 Pearson Product Moment Co-efficient of Correlation

Developed by Pearson (1896), the Pearson Product Moment correlation coefficient is employed with interval/ratio data to determine the degree to which two variables co vary (i.e., vary in relationship to one another. The statistic computed for the Pearson product-moment correlation coefficient is represented by the letter r. r is an estimate of ρ (the Greek letter rho), which is the true correlation between the two variables in the underlying population. r can assume any value within the range of -1 to $+1$. The absolute value of r indicates the strength of the relationship between the two variables. As the absolute value of r approaches 1, the degree of linear relationship between the variables becomes stronger.

The sign of r indicates the nature or direction of the linear relationship that exists between the two variables. A positive sign indicates a direct linear relationship, whereas a negative sign indicates an indirect (or inverse) linear relationship. A direct linear relationship is one in which a change on one variable is associated with a change on the other variable in the same direction (i.e., an increase on one variable is associated with an increase on the other variable, and a decrease on one variable is associated with a decrease on the other variable). An indirect/inverse relationship is one in which a change in one variable is associated with a change in the other variable in the opposite direction (i.e., an increase on one variable is associated with a decrease in the other variable, and a decrease in one variable is associated with an increase in the other variable).

r is computed using the relation

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{n\sum X^2 - (\sum X)^2 \quad n\sum Y^2 - (\sum Y)^2}}$$

Where n is the total number of observations and X and Y are two series.

r can be used as a test of independence under the null hypothesis of independence i-e

$H_0 : \rho = 0$ against the alternative of non independence using the relation

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where t follows Student's t-distribution with n-2 degree of freedom.

3.3.2 Goodman & Kruskal's Gamma Measure of Association

It is a measure of strength of association for the contingency table data when both variables are in ordinal scale of measurement. It was developed by Leo Goodman and William Kruskal's in as series of papers from 1954 to 1972 and named after its two developers. The range of its value is from -1 (100% negative association) to +1 (100% positive association). A zero value of Goodman & Kruskal's Gamma indicates that there is no association in the variables. A pair is said to be concordant if it is ranked in the same order for both variables and a pair is said to be disconcordant if it is ranked in different order for both variables i.e any pair of observations $(x_i\ ,y_i)$ and $(x_j\ ,\ y_j)$ are said to be concordant if the ranks for both elements agree: that is, if both $x_i > x_j$ and $y_i > y_j$ or if both $x_i < x_j$ and $y_i < y_j$, they are said to be discordant, if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$ and If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant, nor disconcordant. G is a symmetrical measure of association and depends upon number of concordant pairs (N_s) and number of disconcordant pairs (N_d). The relation for Gamma is given as under

$$G = \frac{N_s - N_d}{N_s + N_d}$$

If $N_s > N_d$ i.e. there are more number of concordant pairs then number of disconcordant pairs then Gamma is positive showing that there is a positive association between the variables. If $N_s < N_d$ i.e. there are more number of disconcordant pairs then number of concordant pairs then Gamma is negative showing that there is a negative association between the variables. If $N_s = N_d$ i.e. the number of concordant pairs is equal to the number of disconcordant pairs then the gamma will neither be positive nor negative, it will be zero showing that there is no association between variables.

Gamma can be used as a test of independence using a Z score where the null hypothesis is $\gamma = 0$ (No Association) against the alternative hypothesis of association

$$Z = G.\sqrt{\frac{N_s + N_d}{n(1 - G^2)}}$$

where *"n"* is the total number of observations and rest of the terms are defined above.

3.3.3 Chi-square test of independence

The Chi-square test of Independence checks the interdependence between two variables. This test criterion was first developed by Karl Pearson in 1900. The null hypothesis of chi-square test of independence states that the two variables of classification are independent and the alternative hypothesis states that the two variables of classification are not independent, instead they are associated. Let the two variables are categorized in a $r \times c$ contingency table then the test statistic of the Chi-square test of independence is

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

where o_{ij} is the observed frequency *ith* row and *jth* column cell and e_{ij} is the corresponding expected frequency.

The critical value of the test depends upon the significance level α and the number of degree of freedom which is $(r-1)(c-1)$.

The expected frequencies can be calculated by using the formula

$$e_{ij} = \frac{A_i B_j}{n}$$

where A_i is the total of i^{th} row, B_j is the total of j^{th} column and "n" is the total number of observations.

3.3.4 Fisher's exact test of independence

According to Daniel (1990) the Fisher exact test, which is also referred to as the Fisher–Irwin test, was simultaneously proposed by Fisher (1934, 1935), Irwin (1935), and Yates (1934).The test is used with 2 × 2 contingency tables involving small as well as large sample sizes when one or neither of the marginal sums is predetermined by the researcher. It is one of a class of exact tests, so called because the significance of the deviation from a null hypothesis can be calculated exactly, rather than relying on an approximation that becomes exact in the limit as the sample size grows to infinity, as with many statistical tests. This test is useful for categorical data that result from classifying objects in two different ways. It is used to examine the significance of the association between the two kinds of classification.

Following equation is used for the computation of the exact probability P of obtaining a specific set of observed frequencies in a 2 × 2 contingency table.

$$P = \frac{(a+c)!(b+d)!(a+b)!(c+d)!}{n!a!b!c!d!}$$

Where a is the number of observations lying in 1st row and 1st column cell, b is the number of observations lying in 1st row and 2nd column cell, c is the number of observations lying in 2nd row and 1st column cell and d is the number of observations lying in 2nd row and 2nd column cell and "!" represents factorial notation.

Then these probabilities *P's* are calculated for the observed data and for all extreme cases. By summing all *P's*, P_T (Total Probability and p-value) is obtained. If this P_T is less than or equal to size of test then null hypothesis of independence is rejected otherwise null hypothesis cannot be rejected.

3.4 Generation of Data in 2×2 contingency table with a specific measure of association

Null hypothesis in our study is H_0: The two variables/factors are independent i-e

$$H_0 : \pi_{ij} - \pi_{i\bullet}\pi_{\bullet j} = \Delta = 0 \qquad \forall \ i=1,\ 2 \ and\ j=1,\ 2.$$

Alternative hypothesis in our study is H_A: The two variables/factors are dependent i-e

$$H_A : \pi_{ij} - \pi_{i\bullet}\pi_{\bullet j} = \Delta \neq 0 \qquad \forall \ i=1,\ 2 \quad j=1,\ 2 \ and\ 0 < \Delta < 1.$$

In order to get the four cell probabilities $\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$ following four equations are used.

$$\pi_{11} = \pi_{1\bullet}\pi_{\bullet 1} + \Delta$$

$$\pi_{12} = \pi_{1\bullet}\pi_{\bullet 2} - \Delta$$

$$\pi_{21} = \pi_{2\bullet}\pi_{\bullet 1} - \Delta$$

$$\pi_{22} = \pi_{2\bullet}\pi_{\bullet 2} + \Delta$$

If we put $\Delta = 0$ then the four cell probabilities $\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$ will be obtained for null hypothesis of independence and if we put a value of Δ other than zero then the four cell probabilities $\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$ will be obtained under alternative hypothesis. For data to be generated under alternative hypothesis the values of Δ that are used in our study are $\Delta = [0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18]$.

The two marginal probabilities $\pi_{1\bullet}$ and $\pi_{\bullet 1}$ play a vital role in generation of data in 2×2 contingency table. As the sum of total probabilities is 1, so $\pi_{\bullet 2} = 1-\pi_{\bullet 1}$ and $\pi_{2\bullet} = 1-\pi_{1\bullet}$. In our study we have taken $\pi_{\bullet 1} = [0.1, 0.2, 0.3, 0.4, 0.5]$ and $\pi_{1\bullet} = [0.1, 0.2, 0.3, 0.4, 0.5]$. There were total 25 combinations of $\pi_{\bullet 1}$ and $\pi_{1\bullet}$ where the value of Δ is taken as zero and these combinations are named as "Combinations of Independence".

There are (25×9) 225 combinations of $\pi_{1\bullet}$ and $\pi_{\bullet 1}$ where the value of Δ is taken from the set $\Delta = [0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18]$. Out of these 225 combinations there were 90 combinations where the cell probabilities were negative or greater than 1. So these 90 combinations were excluded from further analysis and we were left with only 135 combinations and these combinations were named as "Combinations of Alternative".

Uniform random numbers are generated in the interval (0, 1) of a specific sample size '*n*'. Then these "*n*" uniform random numbers are converted to 2×2 contingency using all of the above 160 combinations of $\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$ (25 combinations of independence and 135 combinations of alternative) using the following procedure.

If the random number is less than or equal to π_{11} then it is counted for 1st row and 1st column cell or if the random number is greater than π_{11} but less than or equal to $\pi_{11}+\pi_{12}$ then it is counted for 1st row and 2nd column cell or if the random number is greater than $\pi_{11}+\pi_{12}$ but less than or equal to $\pi_{11}+\pi_{12}+\pi_{21}$ then it is counted for 2nd row and 1st column cell or if the random number is greater than $\pi_{11+}\ \pi_{12+}\ \pi_{21}$ then it is counted for 2nd row and 2nd column cell.

Similarly in order to use Pearson Product Moment coefficient of correlation, following procedure is used for generation of data in two series *X* and *Y*.

If the random number is less than or equal to π_{11} then X and Y are assigned zero or if the random number is greater than π_{11} but less than or equal to $\pi_{11+}\ \pi_{12}$ then X is assigned zero and Y is assigned 1 or if the random number is greater than $\pi_{11+}\ \pi_{12}$ but less than or equal to $\pi_{11+}\ \pi_{12+}\ \pi_{21}$ then X is assigned 1 and Y is assigned zero or if the random number is greater than $\pi_{11+}\ \pi_{12+}\ \pi_{21}$ then X and Y are assigned 1.

Chapter 4

COMPARING TESTS OF INDEPENDENCE

The four tests of Independence/measures of Correlation (Chi-square test of Independence, Goodman & Kruskal's Gamma measure of association, Pearson Product Moment coefficient of correlation and Fisher Exact Test) are compared in this study on the basis of Size and Power of each test and Stringency criterion.

In this chapter the stepwise procedure for Monte Carlo simulations is stated. The Stringency criterion that is used as the basis of comparison in this study is also explained in this chapter. At the end of chapter the Limitation of our study is discussed.

4.1 Monte-Carlo Simulation Design

The step wise procedure that is used to estimate the Simulated Critical values, Size of test and Power of test for each test of Independence/measure of Correlation is given as under.

(i) Generation of data in 2×2 contingency table using those combinations of π_{11}, π_{12} and π_{21} which satisfy the Null Hypothesis of independence and calculation of Simulated Critical Values for the three tests of Independence/measures of Correlation i.e. Chi-square test of Independence, Goodman & Kruskal's Gamma measure of Association, Pearson Product Moment Coefficient of Correlation.

(ii) Generation of data in 2×2 contingency table using those combinations of π_{11}, π_{12} and π_{21} which satisfy the Null Hypothesis of independence and calculation of size of test for the four tests of Independence/measures of Correlation i.e. Chi-square test of Independence, Goodman & Kruskal's Gamma measure of Association, Pearson Product Moment Coefficient of Correlation and Fisher exact test.

(iii) Generation of data in 2×2 contingency table using those combinations of π_{11}, π_{12} and π_{21} which satisfy the Alternatives Hypothesis of Non-Independence and calculation of Power "P" of each of the tests of Independence/measures of Correlation at different alternatives.

4.2 Stringency Criterion

The two characteristics Size and Power of any test are very important for the comparison of different tests. The tests can be compared on the basis of Size and Power. Let there are two tests $Test_1$ and $Test_2$ that are needed to be compared. Then these two tests can be compared on the basis of their Size and Power. But this approach does not give a satisfactory conclusion as at some alternatives $Test_1$ may be more powerful as compared to $Test_2$ and at some other alternatives $Test_2$ may be more powerful as compared to $Test_1$. So in order to address the problem of a single standard to compare the tests a technique known as stringency is used.

A brief layout of Stringency Criterion for the comparison of four tests of Independence/measures of Correlation is

(i) Calculation of critical values for each test of independence.

(ii) Drawing of Power curve for each test of Independence/measure of correlation by taking different alternatives along x-axis and Power of that test along y-axis i-e $P\ T_{\theta_i}^{k}$ Versus θ_i, where θ_i are different alternatives and $P\ T_{\theta_i}^{k}$ is the Power of test 'T^k' at different alternatives 'θ_i'.

(iii) Drawing of approximated Power Envelope (APE) by taking different alternatives along x-axis and the approximation of maximum Power (AMP) along y-axis i-e the plot of $P^{*}\ T_{\theta_j}^{k}$ versus θ_j. Where $P^{*}\ T_{\theta_j}^{k}$ is the approximated Maximum Power at θ_j for a specific 'j'.

(iv) Calculation of Short Comings of each Test of Independence i-e the Maximum difference between the Power Curve of a Test of Independence and approximated Power Envelope.

$$S\ T_{\theta_{\alpha k}} = \max\left[P^{*}\ T_{\theta_j}^{k} - P\ T_{\theta_i}^{k}\right]$$

(iv) Identification of the most stringent test i-e that test which has minimum of the shortcomings of the four tests of independence i-e Goodman & Kruskal's Gamma Measure of Association, Chi-Square Test of Independence, Pearson Product Moment Co-efficient of Correlation and Fisher Exact Test.

All of the above steps in 2.1 and 2.2 are repeated for different sample sizes of 25, 50 and100.

4.3 Approximating the Power Envelope

In this study in order to obtain power Envelope a test of independence based on Neyman Pearson Lemma is constructed. According to Neyman Pearson (NP) Lemma when one is testing the point null hypothesis

$$H_0: \theta = \theta_0$$

against the point Alternative hypothesis

$$H_A: \theta = \theta_1$$

then test based on Likelihood ratio which rejects H_o in favor of H_A

$$LR(x) = \frac{LH(\theta_0/x)}{LH(\theta_1/x)} \leq \upsilon \text{.............}(4.2.1)$$

Where $LH(\theta_0 / x)$ is the likelihood function under H_0 , $LH(\theta_1 / x)$ is the likelihood function under H_A and

$$P\ (\ LR(x) \leq \upsilon / H_0) = \alpha$$

is the most powerful test of size "α" for a threshold "υ".

In order to find the Likelihood function for a 2×2 contingency table the Multinomial model proposed by G. Rodriguez (2010) is used. According to G. Rodriguez the log-Likelihood function is given as

$$Log\text{-}LH = \sum_{j=1}^{2}\sum_{i=1}^{2} x_{ij}\ \pi_{ij} \quad \text{.........................}(4.2.2)$$

where x_{ij} is the number of observations lying in the ith row and jth column cell and π_{ij} is the cell probability of ith row and jth column.

In order to calculate NP test, data are generated in *2x2* contingency table either using combinations of independence if we are estimating simulated critical values of NP test or using combinations of alternatives if we are estimating powers of NP test. Then log-likelihood function under null hypothesis is calculated using relation 4.2.2 where $\pi's$ are taken from respective combination of independence. Similarly same relation 4.2.2 is used to calculate the log-likelihood function under alternative by using the $\pi's$ from respective combinations of alternative. Finally relation 4.2.1 is used to calculate the test statistic of NP test.

For each fixed alternative there are 25 powers of NP test corresponding to each point in the null hypothesis of independence. So in order to choose the LEAST FAVORABLE null hypothesis we take the smallest power as an approximation to maximum power (AMP).

4.4 Limitation of Study

The scope of this study is limited to 2×2 contingency table. Four tests of independence/measures of correlation (Goodman & Kruskal's Gamma Measure of Association, Chi-Square Test of Independence, Pearson Product Moment Co-efficient of Correlation and Fisher Exact Test) are compared for 2×2 contingency table in this study.

Chapter 5

COMPUTATION OF SIMULATED CRITICAL VALUES AND EMPIRICAL SIZE

In this chapter it is explained that why the simulated critical values are needed? Simulated critical values for each test of independence/measure of correlation are also estimated in this chapter. Size and distortion in size of each test of independence/measure of correlations is also calculated in this chapter and at the end of chapter the conclusion is stated.

5.1 Why the Simulated Critical Values are needed?

Most of the tests are based on asymptotic critical values which do not provide reliable estimates when sample size is small, so there is a need to get critical values which work well even in small samples. That's why we used simulated critical values for our analysis The critical value of each test of independence/measure of correlation is already given in the literature e.g. the Critical Value for Chi-Square test of independence is the value from which the area of chi-squares distribution with $(r-1)(c-1)$ degree of freedom is greater than 0.95 keeping the size of test constant at 5%. When I used these critical values here after named as asymptotic critical values to estimate the size of each test of independence/measure of correlation, I found substantial distortion in the size of each test of independence/measure of correlation (documented in Section 5.2). So in order to keep the size of test constant at nominal size of 5% the simulated critical values for each test of independence are estimated.

5.2 Size distortion when asymptotic critical values are used

The data are generated in 2×2 contingency table using all of the 25 combinations of independence and the size of each test of Independence/measure of Correlation is calculated by taking the asymptotic critical values for different sample sizes of 25, 50, and 100 using a Monte Carlo sample size of 10,000.

The Size of test for Chi-square test of Independence using asymptotic critical values at three different sample sizes of 25, 50 and 100 are given in Table 5.2.1.

The Size of test for Goodman and Kruskal's Gamma measure of Association using asymptotic critical values at three different sample sizes of 25, 50 and 100 are given in Table 5.2.2.

The Size of test for Pearson Product Moment coefficient of Correlation using asymptotic critical values at three different sample sizes of 25, 50 and 100 are given in the Table 5.2.3.

Fisher's exact test is not asymptotic and its exact critical values can be computed. The size of Fisher's exact test at three different sample sizes of 25, 50 and 100 are given in Table 5.4.4.

Table: - 5.2.1 Size of test for Chi-square (using asymptotic critical values)

Combinations of INDEPENDENCE			Size for CHISQUARE		
PI(1,1)	PI(1,2)	PI(2,1)	n=25	n=50	n=100
0.01	0.09	0.09	10.0	7.1	4.8
0.02	0.08	0.18	9.9	6.9	5.1
0.03	0.07	0.27	10.0	6.8	5.2
0.04	0.06	0.36	9.9	6.9	5.0
0.05	0.05	0.45	10.0	7.0	5.0
0.02	0.18	0.08	10.0	7.0	5.1
0.04	0.16	0.16	10.0	7.0	5.0
0.06	0.14	0.24	10.1	6.9	5.0
0.08	0.12	0.32	10.1	7.0	5.0
0.10	0.10	0.40	10.0	7.0	4.8
0.03	0.27	0.07	9.9	7.0	4.8
0.06	0.24	0.14	9.9	6.9	5.1
0.09	0.21	0.21	9.9	7.0	4.8
0.12	0.18	0.28	9.9	7.2	5.4
0.15	0.15	0.35	10.0	7.0	4.9
0.04	0.36	0.06	9.9	6.9	4.9
0.08	0.32	0.12	10.0	7.1	4.9
0.12	0.28	0.18	10.1	7.0	5.0
0.16	0.24	0.24	10.0	6.9	5.0
0.20	0.20	0.30	9.9	7.0	5.1
0.05	0.45	0.05	10.0	6.9	4.9
0.10	0.40	0.10	10.0	7.0	4.9
0.15	0.35	0.15	10.2	6.9	5.0
0.20	0.30	0.20	10.0	7.1	5.0
0.25	0.25	0.25	10.0	7.2	4.8

The above table shows substantial distortion in size of Chi-square test of independence when asymptotic critical values are used. However when sample size increases the distortion in size decreases.

Table: - 5.2.2 Size of test for Gamma (using asymptotic critical values)

Combinations of INDEPENDENCE			Size for GAMMA		
PI(1,1)	PI(1,2)	PI(2,1)	n=25	n=50	n=100
0.01	0.09	0.09	12.0	9.0	5.8
0.02	0.08	0.18	12.1	9.0	6.1
0.03	0.07	0.27	11.9	9.0	6.0
0.04	0.06	0.36	11.8	8.9	5.8
0.05	0.05	0.45	12.1	9.1	6.1
0.02	0.18	0.08	12.0	9.1	5.9
0.04	0.16	0.16	12.1	9.1	6.0
0.06	0.14	0.24	12.0	9.0	5.6
0.08	0.12	0.32	12.0	9.0	5.9
0.10	0.10	0.40	12.0	8.9	6.2
0.03	0.27	0.07	12.0	9.0	5.9
0.06	0.24	0.14	12.1	9.0	5.8
0.09	0.21	0.21	12.0	9.1	6.1
0.12	0.18	0.28	12.0	9.0	6.1
0.15	0.15	0.35	12.0	9.0	6.0
0.04	0.36	0.06	12.0	8.9	6.0
0.08	0.32	0.12	12.0	9.1	5.8
0.12	0.28	0.18	11.9	9.0	6.0
0.16	0.24	0.24	12.0	8.9	6.2
0.20	0.20	0.30	12.1	8.9	6.0
0.05	0.45	0.05	12.0	8.9	6.1
0.10	0.40	0.10	12.0	8.9	5.9
0.15	0.35	0.15	11.9	9.0	6.0
0.20	0.30	0.20	12.0	9.0	6.1
0.25	0.25	0.25	12.0	9.0	6.2

The above table shows substantial distortion in size of Goodman and Kruskal's Gamma measure of Association when asymptotic critical values are used. However when sample size increases the distortion in size decreases.

Table: - 5.2.3 Size of test for Pearson (using asymptotic critical values)

Combinations of INDEPENDENCE			Size for PEARSON		
PI(1,1)	PI(1,2)	PI(2,1)	n=25	n=50	n=100
0.01	0.09	0.09	9.1	7.0	5.2
0.02	0.08	0.18	9.1	7.1	5.0
0.03	0.07	0.27	9.0	7.1	5.2
0.04	0.06	0.36	8.9	6.9	4.8
0.05	0.05	0.45	9.0	7.0	5.1
0.02	0.18	0.08	9.0	6.8	4.9
0.04	0.16	0.16	8.9	6.9	5.0
0.06	0.14	0.24	8.8	6.9	5.1
0.08	0.12	0.32	8.9	7.0	5.2
0.10	0.10	0.40	8.9	7.1	5.1
0.03	0.27	0.07	9.1	7.0	5.0
0.06	0.24	0.14	8.9	7.1	5.1
0.09	0.21	0.21	9.1	6.9	5.0
0.12	0.18	0.28	9.1	7.0	5.0
0.15	0.15	0.35	9.1	7.2	5.1
0.04	0.36	0.06	8.9	7.1	5.0
0.08	0.32	0.12	8.9	7.1	5.0
0.12	0.28	0.18	8.8	6.8	5.1
0.16	0.24	0.24	8.9	7.0	5.0
0.20	0.20	0.30	9.0	6.9	5.0
0.05	0.45	0.05	9.2	7.1	5.0
0.10	0.40	0.10	9.1	7.0	5.0
0.15	0.35	0.15	9.0	7.1	5.0
0.20	0.30	0.20	8.9	7.0	5.0
0.25	0.25	0.25	9.1	6.9	5.1

The above table shows substantial distortion in size of Pearson product moment coefficient of correlation when asymptotic critical values are used. However when sample size increases the distortion in size decreases.

5.3 Computation of Simulated critical values

In order to find the simulated critical values "n" uniform random numbers are generated in interval *(0,1)* and then these random numbers are converted in 2×2 contingency table using all of the 25 combinations of independence (see Section 3.4). For a Monte Carlo sample size of 10,000 the simulated critical values for Chi-square Test of Independence, Goodman and Kruskal's Gamma measure of Association and Pearson Coefficient of Correlation are obtained at three different samples sizes of 25, 50 and 100.

The critical values are different for different combinations of independence at a specific sample size for particular test of Independence/measure of Correlation, so we chose the supremum of these critical values as our Simulated Critical value in order to keep the size of test less than or equal to nominal size of 5%. If there are lower critical values (LCVs) just as for Pearson and Gamma the minimum of these LCVs is chosen as our simulated critical value in order to keep the size of test less than or equal to nominal size of 5%.

For example the simulated critical values for Goodman and Kruskal's Gamma measure of Association at sample size of 25 are given in the Table 5.3.1. The least LCV is -0.72 and the supremum UCV is 0.91 so we take these two critical values as simulated critical values for Goodman and Kruskal's Gamma at sample size of 25.

In the same manner the critical values for three tests of independence are chosen. The simulated critical values that are used in further analysis for each test of independence at three sample sizes of 25, 50 and 100 are given in Table 5.3.2.

Table:- 5.3.1 Simulated critical Values for Goodman and Kruskal's Gamma

Combinations of INDEPENDENCE			LCV	UCV
PI(1,1)	PI(1,2)	PI(2,1)		
0.01	0.09	0.09	-0.04	0.91
0.02	0.08	0.18	-0.28	0.90
0.03	0.07	0.27	-0.45	0.81
0.04	0.06	0.36	-0.57	0.74
0.05	0.05	0.45	-0.66	0.66
0.02	0.18	0.08	-0.28	0.90
0.04	0.16	0.16	-0.49	0.85
0.06	0.14	0.24	-0.58	0.81
0.08	0.12	0.32	-0.66	0.76
0.10	0.10	0.40	-0.71	0.71
0.03	0.27	0.07	-0.45	0.80
0.06	0.24	0.14	-0.60	0.80
0.09	0.21	0.21	-0.66	0.79
0.12	0.18	0.28	-0.71	0.76
0.15	0.15	0.35	-0.72	0.72
0.04	0.36	0.06	-0.56	0.74
0.08	0.32	0.12	-0.66	0.76
0.12	0.28	0.18	-0.71	0.75
0.16	0.24	0.24	-0.71	0.73
0.20	0.20	0.30	-0.71	0.71
0.05	0.45	0.05	-0.66	0.66
0.10	0.40	0.10	-0.71	0.71
0.15	0.35	0.15	-0.72	0.72
0.20	0.30	0.20	-0.72	0.73
0.25	0.25	0.25	-0.71	0.73

Where 'LCV' means lower critical value and 'UCV' means upper critical value.

Table:- 5.3.1 Simulated critical Values

Sample Size	Critical Values for Chi-square	Critical Values for Goodman & Kruskal's GAMMA		Critical Values for Pearson Coefficient of Correlation	
		LCV	UCV	LCV	UCV
25	5.21	-0.72	0.91	-0.41	0.55
50	5.95	-0.68	0.89	-0.29	0.35
100	4.61	-0.65	0.75	-0.20	0.23

As the Test of Independence based on Neyman Pearson Lemma (NP test) is the ratio of the Likelihood Function under Null and Alternative hypothesis. So the critical values will be different for different alternative hypothesis. Also for a particular alternative hypothesis there will be 25 different critical values corresponding to 25 different combination of Independence for a specified sample size.

5.4 Computation of Size of Test based on simulated critical values

The data are generated in 2×2 contingency table using all of the 25 combinations of independence and by taking the critical values obtained in previous section 5.3, the size of each test of Independence/measure of Correlation is calculated for different sample sizes of 25, 50, and 100 using a Monte Carlo sample size of 10,000.

The Size of test for Chi-square test of Independence at three different sample sizes of 25, 50 and 100 are given in Table 5.4.1.

The Size of test for Goodman and Kruskal's Gamma measure of Association at three different sample sizes of 25, 50 and 100 are given in Table 5.4.2.

The Size of test for Pearson Product Moment coefficient of Correlation at three different sample sizes of 25, 50 and 100 are given in the Table 5.4.3.

The Size of test for Fisher exact test of independence at three different sample sizes of 25, 50 and 100 are given in the Table 5.4.4.

Table:- 5.4.1 Size of Chi-square (using simulated critical values)

Combinations of INDEPENDENCE			Size for CHISQUARE		
PI(1,1)	PI(1,2)	PI(2,1)	n=25	n=50	n=100
0.01	0.09	0.09	5.1	4.6	5.0
0.02	0.08	0.18	3.0	2.2	2.8
0.03	0.07	0.27	1.2	0.8	2.2
0.04	0.06	0.36	0.4	0.4	2.4
0.05	0.05	0.45	0.1	0.2	2.3
0.02	0.18	0.08	3.3	2.0	2.8
0.04	0.16	0.16	2.4	1.5	2.4
0.06	0.14	0.24	1.8	1.2	3.1
0.08	0.12	0.32	1.0	0.9	3.0
0.10	0.10	0.40	0.8	1.1	2.9
0.03	0.27	0.07	1.1	0.7	2.7
0.06	0.24	0.14	2.0	1.3	3.0
0.09	0.21	0.21	1.6	1.1	2.8
0.12	0.18	0.28	1.6	1.3	2.9
0.15	0.15	0.35	1.6	1.3	3.4
0.04	0.36	0.06	0.5	0.3	2.3
0.08	0.32	0.12	1.0	1.0	2.9
0.12	0.28	0.18	1.6	1.2	3.5
0.16	0.24	0.24	1.9	1.8	3.4
0.20	0.20	0.30	1.9	1.4	3.5
0.05	0.45	0.05	0.2	0.3	2.4
0.10	0.40	0.10	0.7	0.9	3.1
0.15	0.35	0.15	1.6	1.4	3.0
0.20	0.30	0.20	1.9	1.7	3.5
0.25	0.25	0.25	2.1	1.6	3.3

The above table shows that size of Chi-square test of independence is less or equal to nominal size of 5%, as we have used simulated critical values.

Table:-5.4.2 Size of Gamma (using simulated critical values)

Combinations of INDEPENDENCE			Size for GAMMA		
PI(1,1)	PI(1,2)	PI(2,1)	n=25	n=50	n=100
0.01	0.09	0.09	4.8	2.7	2.5
0.02	0.08	0.18	1.3	0.3	0.8
0.03	0.07	0.27	0.3	0.3	2.3
0.04	0.06	0.36	0.3	1.1	3.4
0.05	0.05	0.45	1.0	2.2	3.2
0.02	0.18	0.08	1.5	0.5	0.8
0.04	0.16	0.16	0.7	0.6	2.4
0.06	0.14	0.24	0.7	1.5	1.5
0.08	0.12	0.32	1.4	2.3	0.6
0.10	0.10	0.40	2.0	2.2	0.3
0.03	0.27	0.07	0.3	0.5	2.3
0.06	0.24	0.14	0.8	1.6	1.6
0.09	0.21	0.21	1.7	2.2	0.3
0.12	0.18	0.28	2.1	1.6	0.2
0.15	0.15	0.35	2.7	1.0	0.1
0.04	0.36	0.06	0.2	1.1	3.3
0.08	0.32	0.12	1.4	2.2	0.7
0.12	0.28	0.18	2.3	1.5	0.2
0.16	0.24	0.24	2.7	0.5	0.0
0.20	0.20	0.30	2.7	0.6	0.0
0.05	0.45	0.05	0.9	2.3	3.2
0.10	0.40	0.10	1.9	2.1	0.4
0.15	0.35	0.15	2.8	1.0	0.0
0.20	0.30	0.20	2.4	0.4	0.0
0.25	0.25	0.25	2.5	0.3	0.0

The above table shows that size of Goodman and Kruskal's Gamma measure of association is less or equal to nominal size of 5%, as we have used simulated critical value.

Table:- 5.4.3 Size of Pearson (using simulated critical values)

Combinations of INDEPENDENCE			Size for PEARSON		
PI(1,1)	PI(1,2)	PI(2,1)	n=25	n=50	n=100
0.01	0.09	0.09	2.9	2.2	2.3
0.02	0.08	0.18	1.4	1.7	2.1
0.03	0.07	0.27	0.5	1.5	3.1
0.04	0.06	0.36	0.6	1.4	2.8
0.05	0.05	0.45	1.1	1.8	3.4
0.02	0.18	0.08	1.3	1.6	2.2
0.04	0.16	0.16	0.9	1.5	3.2
0.06	0.14	0.24	1.2	2.0	3.3
0.08	0.12	0.32	1.5	2.5	3.1
0.10	0.10	0.40	2.1	2.7	3.4
0.03	0.27	0.07	0.6	1.4	2.5
0.06	0.24	0.14	1.0	2.1	3.3
0.09	0.21	0.21	1.6	2.6	3.7
0.12	0.18	0.28	1.9	2.5	3.1
0.15	0.15	0.35	2.2	3.2	3.0
0.04	0.36	0.06	0.5	1.7	3.2
0.08	0.32	0.12	1.7	2.3	3.3
0.12	0.28	0.18	2.3	2.8	3.0
0.16	0.24	0.24	2.4	2.5	3.4
0.20	0.20	0.30	2.3	2.9	3.3
0.05	0.45	0.05	1.0	2.2	3.5
0.10	0.40	0.10	2.0	2.9	3.5
0.15	0.35	0.15	2.5	2.6	3.3
0.20	0.30	0.20	2.7	2.6	3.5
0.25	0.25	0.25	2.2	2.7	3.7

The above table shows that size of Pearson product moment coefficient of correlation is less or equal to nominal size of 5%, as we have used simulated critical values.

Table:- 5.4.4 Size of Fisher

Combinations of INDEPENDENCE			Size for FISHER		
PI(1,1)	PI(1,2)	PI(2,1)	n=25	n=50	n=100
0.01	0.09	0.09	1.1	4.0	3.7
0.02	0.08	0.18	1.1	3.0	2.7
0.03	0.07	0.27	0.6	2.0	2.5
0.04	0.06	0.36	0.3	1.1	2.5
0.05	0.05	0.45	0.2	0.8	2.6
0.02	0.18	0.08	1.1	2.6	2.4
0.04	0.16	0.16	1.6	2.6	2.7
0.06	0.14	0.24	1.3	2.1	3.1
0.08	0.12	0.32	0.9	2.2	3.3
0.10	0.10	0.40	0.8	2.5	3.2
0.03	0.27	0.07	0.5	1.8	2.1
0.06	0.24	0.14	1.2	2.2	3.1
0.09	0.21	0.21	1.8	2.8	3.7
0.12	0.18	0.28	1.9	3.1	3.9
0.15	0.15	0.35	2.3	3.2	4.0
0.04	0.36	0.06	0.2	0.9	2.5
0.08	0.32	0.12	1.1	2.4	3.7
0.12	0.28	0.18	1.8	3.1	3.9
0.16	0.24	0.24	3.1	3.4	3.8
0.20	0.20	0.30	3.0	3.5	4.4
0.05	0.45	0.05	0.1	0.7	2.6
0.10	0.40	0.10	0.7	2.7	3.7
0.15	0.35	0.15	2.3	3.1	3.8
0.20	0.30	0.20	3.0	3.2	3.7
0.25	0.25	0.25	4.6	4.8	5.1

The above table shows that size of Fisher exact test of independence is less or equal to nominal size of 5%.

Size of Goodman and Kruskal's Gamma and Pearson product moment coefficient of correlation always remain below 5% because the distribution of these two measures in non symmetric. The non symmetry of Goodman and Kruskal's Gamma and Pearson are shown in Figure 5.4.1 and 5.4.2.

Figure: - 5.4.1 Probability distribution of Goodman and Kruskal's Gamma

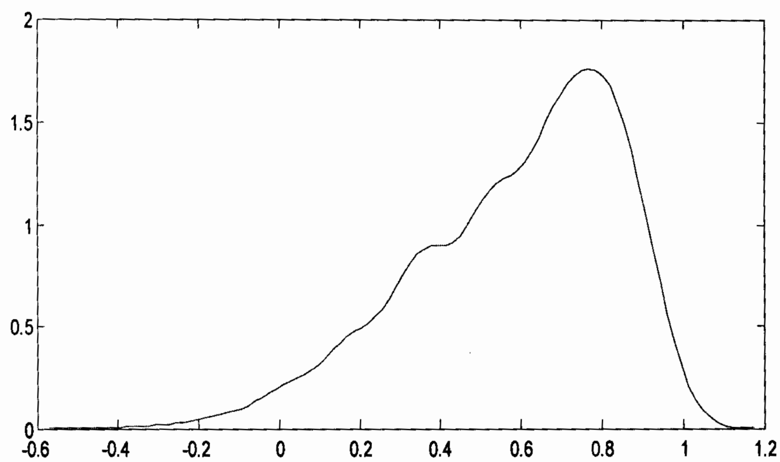
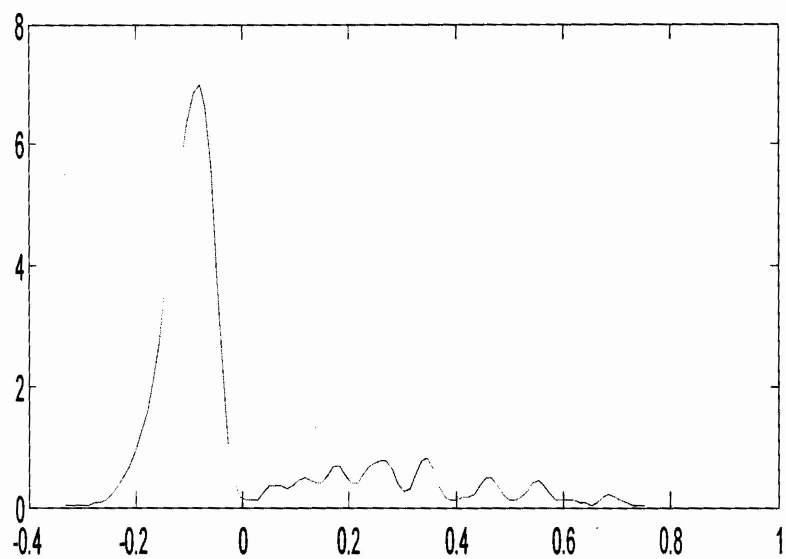


Figure: - 5.4.2 Probability distribution of Pearson Coefficient of Correlation



5.5 Comparison on the basis of size of test

In order to observe that which test of independence has the minimum distortion in size, the average of deviations of size (using asymptotic critical values) of three tests excluding Fisher's exact test around 5 are calculated for three sample sizes of 25, 50 and 100. Fisher's exact test was excluded because it was observed that Fisher exact test has the zero distortion in size.

The average of positive deviations of size of each test around 5 excluding Fisher's exact test when asymptotic critical values are used for three sample sizes of 25, 50 and 100 are given in Tables 5.5.1. Fisher's exact test has been excluded from the table because it has zero distortion.

Table: - 5.5.1 Distortion in size (using asymptotic critical values)

Sample Size	Gamma	Chi-square	Pearson
25	7.0	5.0	4.0
50	4.0	2.0	2.0
100	1.0	0.1	0.1

There is zero distortion in size of four tests of independence when simulated critical values are used.

5.5 Conclusion

As we have discussed, the space of null hypothesis is not a singleton, rather it contains many points. Fisher's exact test shows zero distortion in size. Therefore on the basis of size, Fisher exact test of Independence is the best of all the four tests.

Chapter 6

POWER COMPARISION

In this chapter the Powers of each test of Independence/measure of Correlations are estimated. Difference in power of each test of Independence/measure of Correlation from approximate Power Envelope is also shown using a Pictorial representation in this chapter. The most stringent test of all the four tests of independence/measures of correlations is also identified and at the last of the chapter the Conclusion is stated.

6.1 Computation of Power of Each Test

In order to calculate the Power of each test of Independence/measure of Correlation the data are generated using all of the 135 combinations of Alternative. The powers of Goodman & Kruskal's Gamma Measure of Association, Chi-Square Test of Independence, Pearson Product Moment co-efficient of correlation and Fisher Exact test are obtained for a Monte Carlo sample size of 10,000.

Powers of Goodman & Kruskal's Gamma Measure of Association, Chi-Square Test of Independence, Pearson Product Moment co-efficient of correlation and Fisher Exact test for different combinations of marginal probabilities $(\pi_{1\bullet} and \pi_{1\bullet})$ are given in tables 6.1.1

6.2 Computation of Power Envelope

The Powers of NP test at least favorable null hypothesis is used as an approximation of maximum power (AMP).

Table: - 6.1.1 Powers of tests at sample size of 25 when $\pi_{1\bullet} = 0.4 and \pi_{\bullet 1} = 0.4$.

Alternative Δ	Powers				
	Chi	***Gamma***	***Fisher***	***Pearson***	***AMP***
0.02	3.58	1.33	4.72	1.61	11.77
0.04	6.93	1.54	9.71	2.80	20.05
0.06	14.48	2.58	17.33	6.84	35.56
0.08	25.73	5.98	29.33	13.25	51.28
0.10	40.12	12.85	46.13	25.18	67.41
0.12	57.67	22.91	62.47	41.83	81.12
0.14	73.68	38.81	77.87	60.72	92.52
0.16	87.45	57.28	89.51	78.54	97.47
0.18	95.28	74.93	96.34	91.96	99.17

Table: - 6.1.2 Powers of tests at sample size of 50 when $\pi_{1\bullet} = 0.4 and \pi_{\bullet 1} = 0.4$.

Alternative Δ	Powers				
	Chi	***Gamma***	***Fisher***	***Pearson***	***AMP***
0.02	2.75	1.1	5.02	1.55	11.14
0.04	6.02	1.25	8.82	2.81	20.29
0.06	12.93	2.17	17.02	6.06	31.76
0.08	22.45	4.81	28.45	12.39	48.61
0.1	36.73	10.35	43.45	23.29	65.29
0.12	54.09	19.34	60.56	39.47	82.16
0.14	70.41	33.46	75.89	58.83	92.04
0.16	85.48	50.12	88.43	77.98	97.53
0.18	93.98	70.53	95.83	92.61	99.49

Table: - 6.1.3 Powers of tests at sample size of 100 when $\pi_{1\bullet} = 0.4 and \pi_{\bullet 1} = 0.4$.

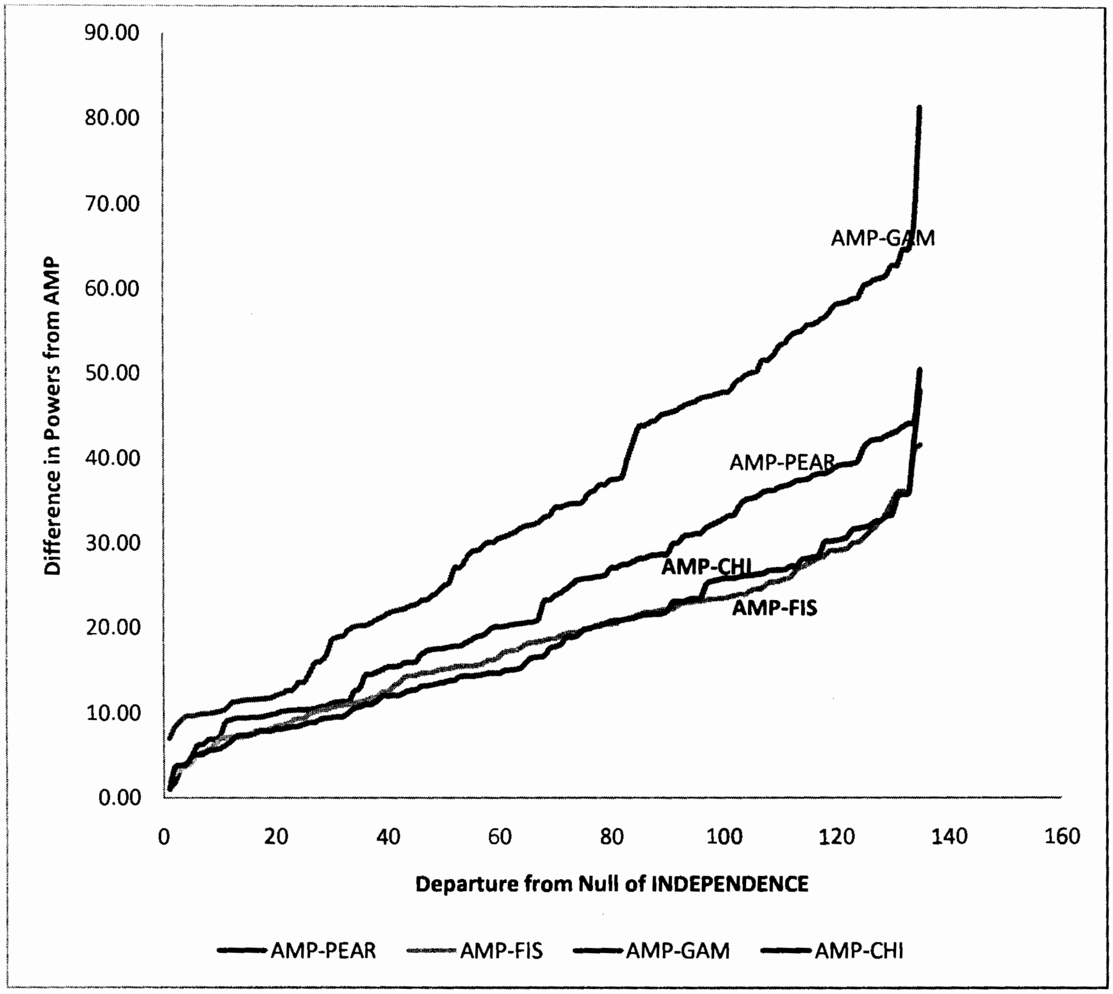
Alternative Δ	Powers				
	Chi	***Gamma***	***Fisher***	***Pearson***	***AMP***
0.02	3.25	0.12	6.78	3.48	15.58
0.04	9.69	0.19	16.92	9.29	32.17
0.06	24.07	0.76	35.47	22.87	52.35
0.08	45.6	2.62	59.5	44.63	75.78
0.1	69.16	9.24	81.18	68.47	91.59
0.12	88.28	23.22	93.86	87.85	98.1
0.14	97.24	48.14	98.96	96.84	99.72
0.16	99.69	76.64	99.88	99.66	99.97
0.18	100	95.22	99.99	100	100

6.3 Comparison on the basis of Power curves

In order to compare the tests on the basis of powers, power of each test at each alternative is subtracted from AMP and then they are sorted in ascending order of that difference. These sorted differences are then plotted using a scatter plot by taking the number of combination along x-axis and the difference of power from AMP along y-axis.

Figure: - 6.3.1

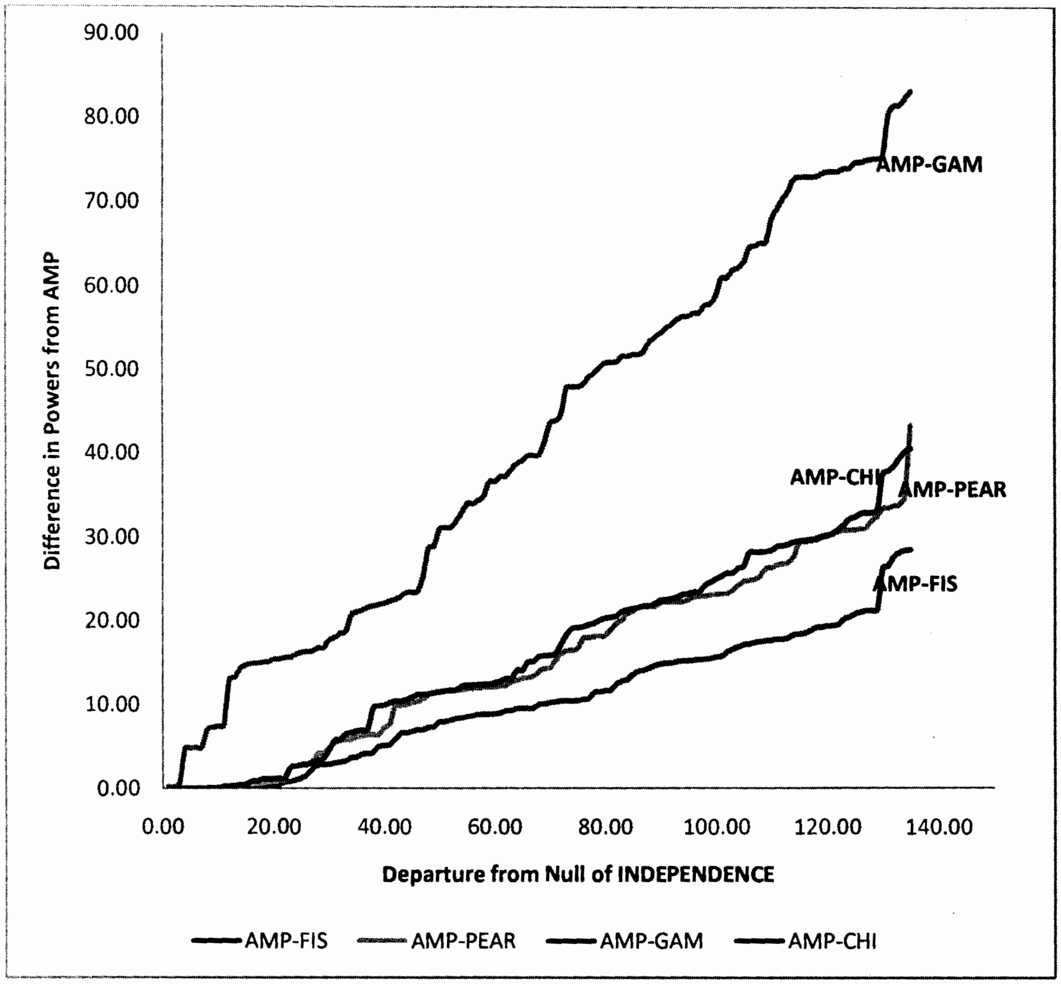
Difference in powers of each test from AMP at sample size of 25



The above figure depicts that Fisher's exact test and Chi-square test has minimum deviations from AMP. Whereas the Gamma has the maximum deviations from AMP. So Fisher's exact test and Chisquare test are more powerful tests as compared to the rest of two i-e Gamma and Pearson.

Figure: - 6.3.2

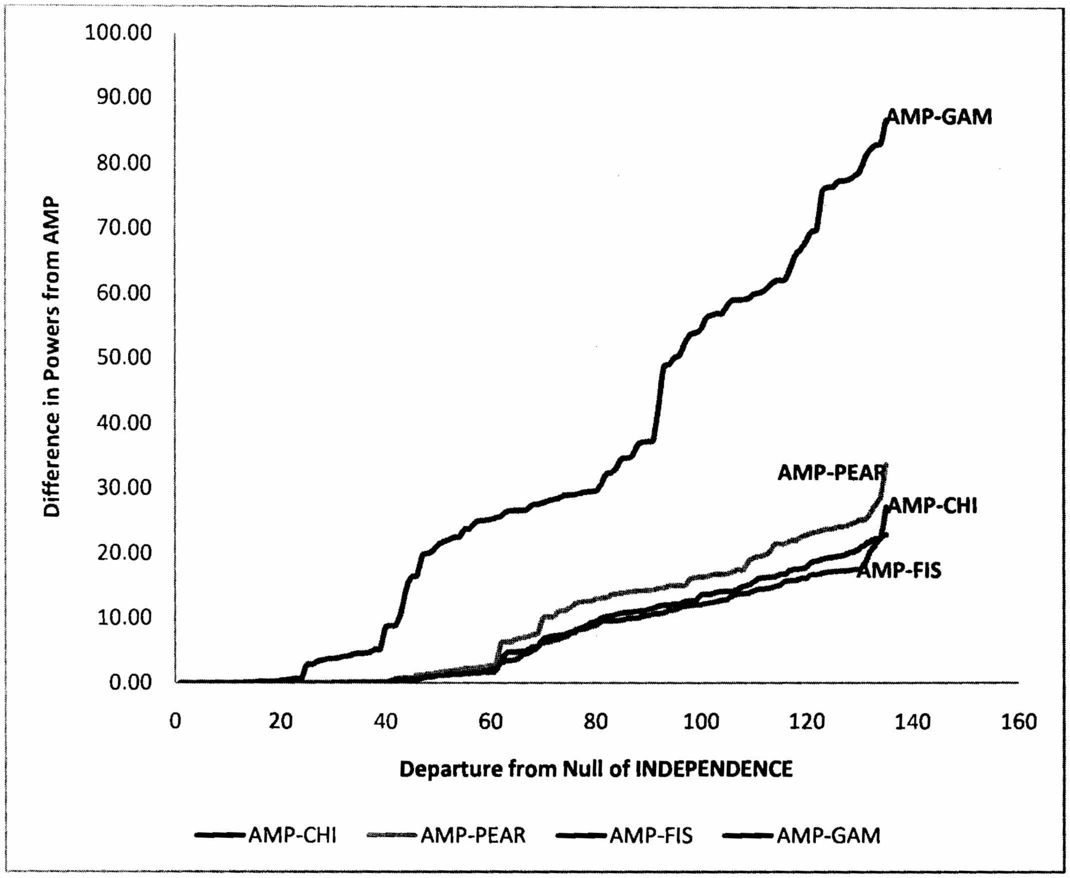
Difference in powers of each test from AMP at sample size of 50



The above figure depicts that Fisher's exact test has minimum deviations from AMP. Whereas the Gamma has the maximum deviations from AMP. So Fisher's exact test is more powerful tests as compared to the rest of three i-e Gamma, Chisquare and Pearson.

Figure: - 6.3.3

Difference in powers of each test from AMP at sample size of 100



The above figure depicts that Fisher's exact test, Chisquare test and Pearson Product Moment Coefficient has minimum deviations from AMP. Whereas the Gamma has the maximum deviations from AMP. So Fisher's exact test, Chisquare test and Pearson Product Moment Coefficient is more powerful tests as compared to Gamma measure of associaton.

6.3.1 Conclusion

All of the above figures show that Fisher's exact test has the minimum deviation from AMP among all the four tests of independence/measures of correlation. So Fisher's exact test is the best test of independence of all the four tests/measures on the basis of power. Also it is clear from the figures that the Goodman and Kruskal's Gamma measure of association is the worst test of all the four test of independence.

6.4 Identification of Most Stringent Test

The Power of each test of independence/measure of correlation is subtracted from the Power of NP Test and then to calculate the Short Coming of each test of independence/measure of correlation the maximum of these differences is taken for a specific combination of marginal probabilities ($\pi_{1\bullet} and \pi_{\bullet 1}$) at three sample sizes of 25,50 and 100. All of these results are shown in tables 6.4.1, 6.4.2 and 6.4.3.

Table: - 6.4.1 Most stringent test

Sample Size	Shortcoming				Most stringent test
	Chi	Gamma	Fisher	Pearson	
25	43.03	88.41	41.58	45.57	Fisher
50	41.31	83.42	35.37	41.23	Fisher
100	35.93	68.41	31.86	38.78	Fisher

6.4.1 Conclusion

The Fisher's exact test of independence is the best test of all the four tests of independence/measures of correlation for categorical data on the basis of stringency criterion.

6.5 Conclusion

As discussed in 6.3.1, Fisher's exact test of independence is the best test of all the four tests of independence/measures of correlation for categorical data on the basis of power and as discussed in 6.4.1, Fisher's exact test of independence is the best test of all the four tests of independence/measures of correlation for categorical data on the basis of stringency criterion, so Fisher's exact test is the best test of all the tests of independence.

Chapter 7

THE BEST AND WORST CASES OF EACH TEST OF INDEPENDENCE

As already it is stated in section 6.3.1 that Goodman and Kruskal's Gamma measure of association performs worst, so it is excluded from the further analysis. In this chapter combination of alternative where each of the three tests of independence (Fisher's exact test, Chi-square test and Pearson product moment coefficient of correlation) performs best and worst are given by a pictorial representation and in form of a table. Those combinations at which the powers of three tests of independence (Fisher's exact test, Chi-square test and Pearson product moment coefficient of correlation) are greater than 90% are eliminated from further analysis.

7.1 Combinations of Alternative where tests perform BEST

For all three sample sizes of 25, 50 and 100, those combinations of alternative are considered best where the difference between power of a particular test and AMP is less than 5%.

7.1.1 When sample size is 25

Combinations of alternative where Chi-square performs best at sample size of 25 are shown in figure 7.1.1.1 and Table 7.1.1.1.

Combinations of alternative where Fisher's exact test performs best at sample size of 25 are shown in figure 7.1.1.2 and Table 7.1.1.2.

Combinations of alternative where Pearson product moment coefficient of correlation performs best at sample size of 25 are shown in figure 7.1.1.3 and Table 7.1.1.3.

Figure: - 7.1.1.1 Combinations of alternative where Chi-square performs best

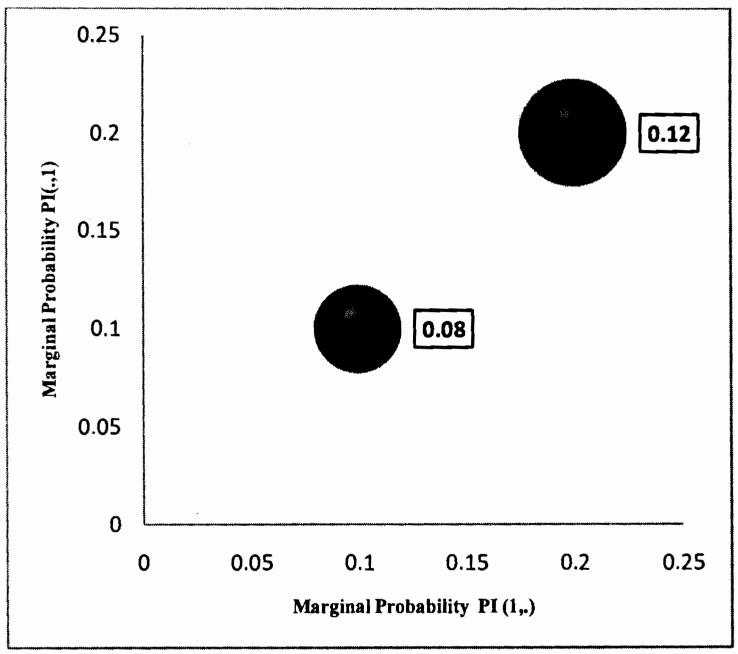


Table: - 7.1.1.1 Combinations of alternative where Chi-square performs best

PI(1,.)	PI(.,1)	delta	AMP	CHI
0.1	0.1	0.08	86.21	82.63
0.2	0.2	0.12	91.36	86.53

Figure: - 7.1.1.2 Combinations of alternative where Fisher performs best

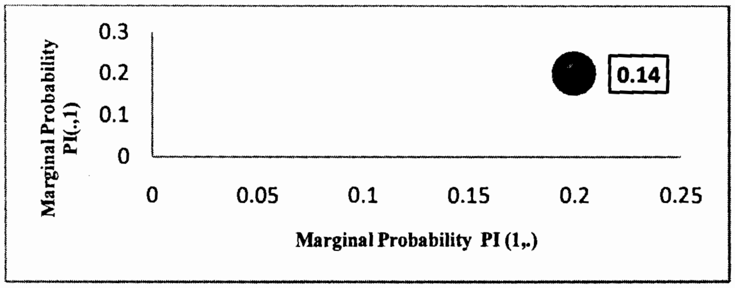


Table: - 7.1.1.2 Combinations of alternative where Fisher performs best

PI(1,.)	PI(.,1)	delta	AMP	FIS
0.2	0.2	0.14	94.71	89.69

Figure: - 7.1.1.3 Combinations of alternative where Pearson performs best

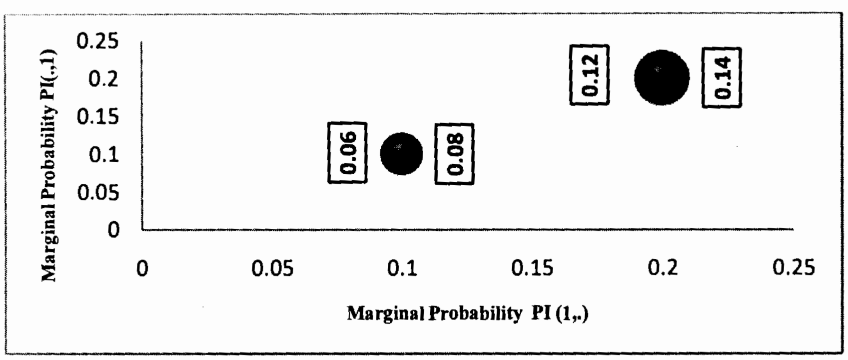


Table: - 7.1.1.3 Combinations of alternative where Pearson performs best

PI(1,.)	PI(.,1)	delta	AMP	PEAR
0.1	0.1	0.08	86.21	84.41
0.2	0.2	0.14	94.71	91.30
0.1	0.1	0.06	72.21	68.15
0.2	0.2	0.12	91.36	86.51

7.1.2 When sample size is 50

Combinations of alternative where Chi-square performs best at sample size of 50 are shown in figure 7.1.2.1 and Table 7.1.2.1.

Combinations of alternative where Fisher's exact test performs best at sample size of 50 are shown in figure 7.1.2.2 and Table 7.1.2.2.

Combinations of alternative where Pearson product moment coefficient of correlation performs best at sample size of 50 are shown in figure 7.1.2.3 and Table 7.1.2.3.

Figure: - 7.1.2.1 Combinations of alternative where Chi-square performs best

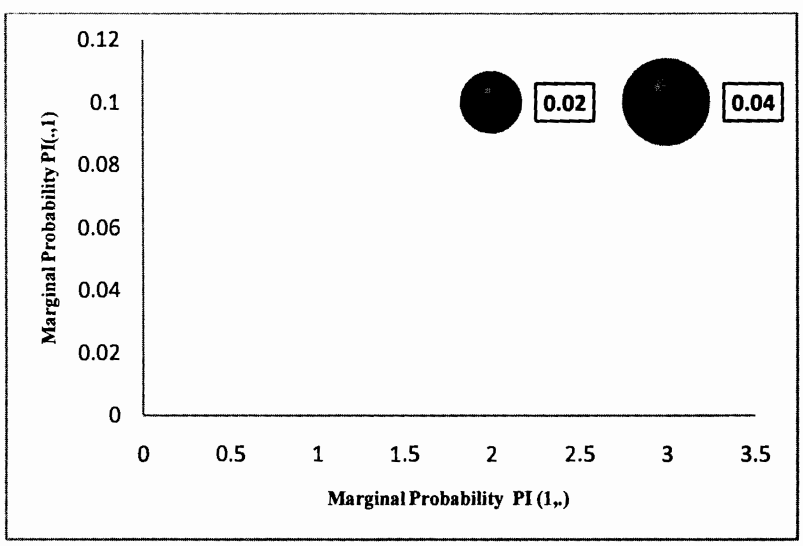


Table: - 7.1.2.1 Combinations of alternative where Chi-square performs best

PI(1,.)	PI(.,1)	delta	AMP	CHI
0.10	0.10	0.02	32.25	29.62
0.10	0.10	0.04	69.46	65.90

Figure: - 7.1.2.2 Combinations of alternative where Fisher performs best

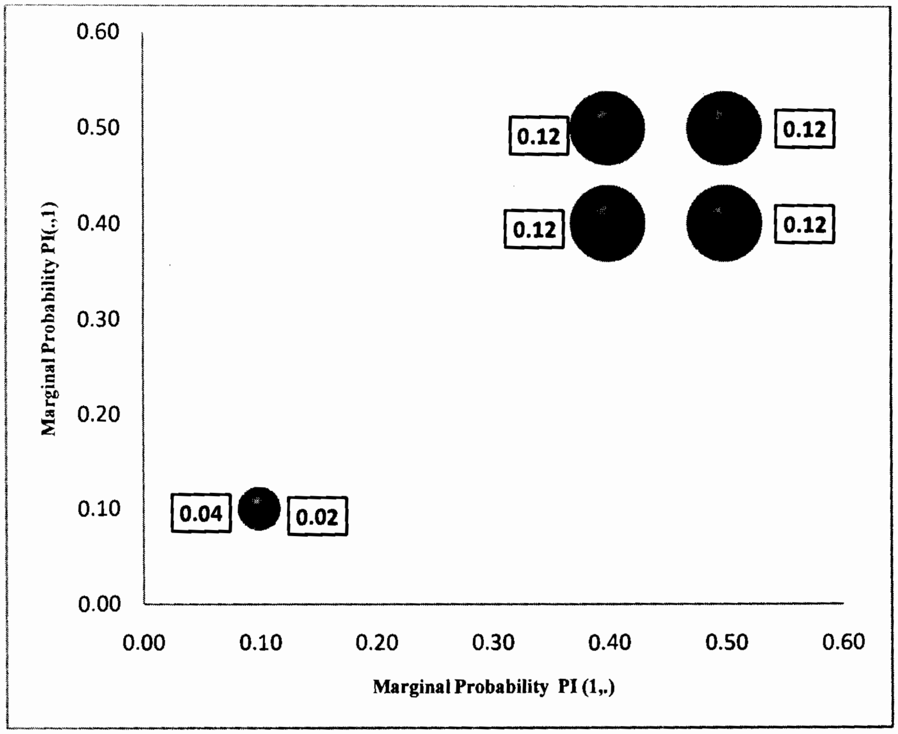


Table: - 7.1.2.2 Combinations of alternative where Fisher performs best

PI(1,.)	PI(.,1)	delta	AMP	FIS
0.10	0.10	0.02	32.25	29.22
0.40	0.40	0.12	98.02	94.32
0.10	0.10	0.04	69.46	65.38
0.50	0.40	0.12	97.86	93.74
0.40	0.50	0.12	98.10	93.86
0.50	0.50	0.12	97.43	92.47

Figure: - 7.1.2.3 Combinations of alternative where Pearson performs best

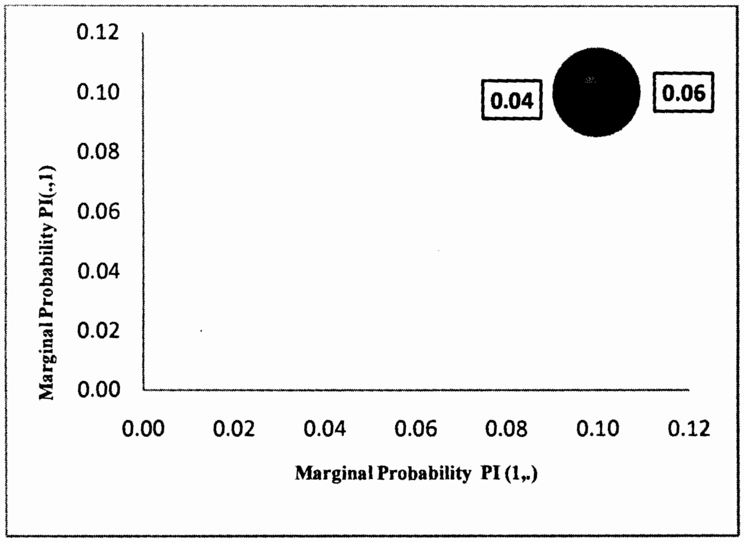


Table: - 7.1.2.3 Combinations of alternative where Pearson performs best

PI(1,.)	PI(.,1)	delta	AMP	PEAR
0.10	0.10	0.06	93.18	89.05
0.10	0.10	0.04	69.46	65.31

7.1.3 When sample size is 100

Combinations of alternative where Chi-square performs best at sample size of 100 are shown in figure 7.1.3.1 and Table 7.1.3.1.

Combinations of alternative where Fisher's exact test performs best at sample size of 100 are shown in figure 7.1.3.2 and Table 7.1.3.2.

Figure: - 7.1.3.1 Combinations of alternative where Chi-square performs best

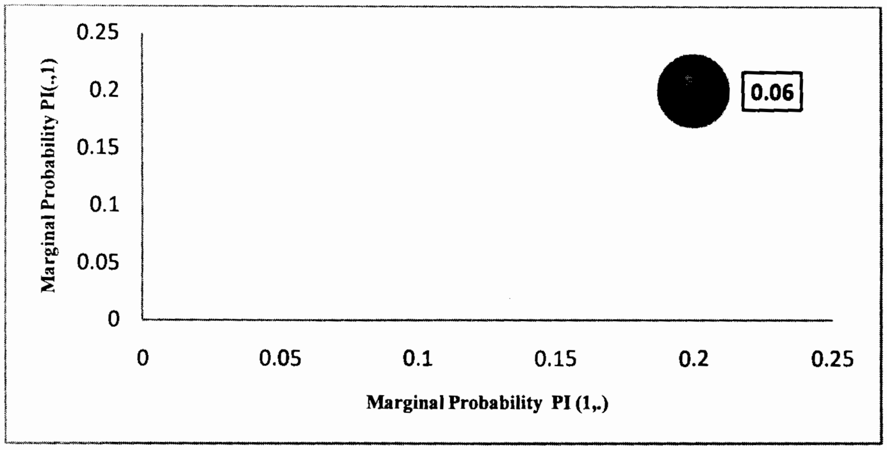


Table: - 7.1.3.1 Combinations of alternative where Chi-square performs best

PI(1,.)	PI(.,1)	delta	AMP	CHI
0.2	0.2	0.06	96.05	91.50

Figure: - 7.1.3.2 Combinations of alternative where Fisher performs best

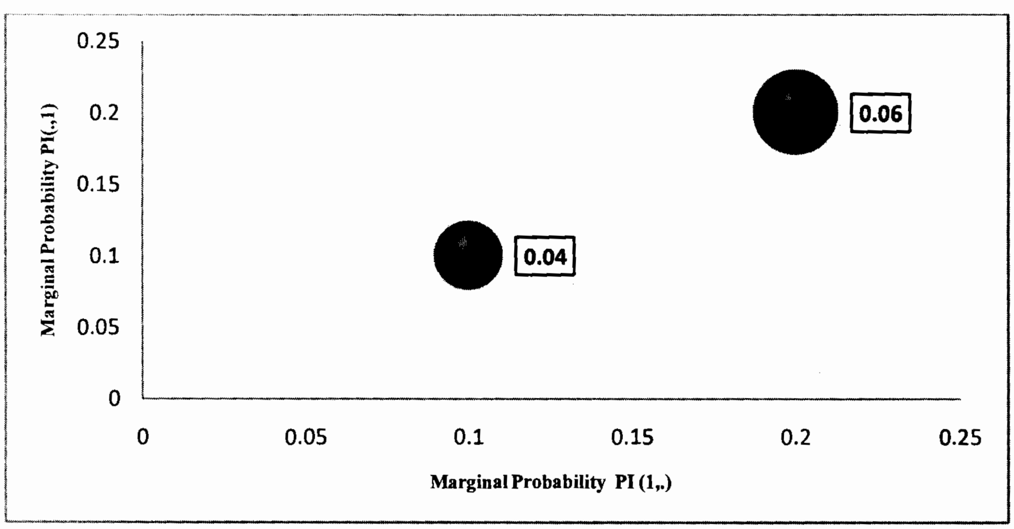


Table: - 7.1.3.2 Combinations of alternative where Fisher performs best

PI(1,.)	PI(.,1)	delta	AMP	FIS
0.1	0.1	0.04	93.01	89.64
0.2	0.2	0.06	96.05	91.11

7.2 Combinations of Alternative where tests perform WORST

For sample size of 25, those combinations of alternative are considered worst where the difference between power of a particular test and AMP is greater than 40%. Similarly for sample size of 50, those combinations of alternative are considered worst where the difference between power of a particular test and AMP is greater than 35% and for sample size of 100, those combinations of alternative are considered worst where the difference between power of a particular test and AMP is greater than 20%.

7.2.1 When sample size is 25

Combinations of alternative where Chi-square performs worst at sample size of 25 are shown in figure 7.2.1.1 and Table 7.2.1.1.

Combinations of alternative where Fisher's exact test performs worst at sample size of 25 are shown in figure 7.2.1.2 and Table 7.2.1.2.

Combinations of alternative where Pearson coefficient of correlation performs worst at sample size of 25 are shown in figure 7.2.1.3 and Table 7.2.1.3.

Figure:-7.2.1.1Combinations of alternative where Chi-square performs worst

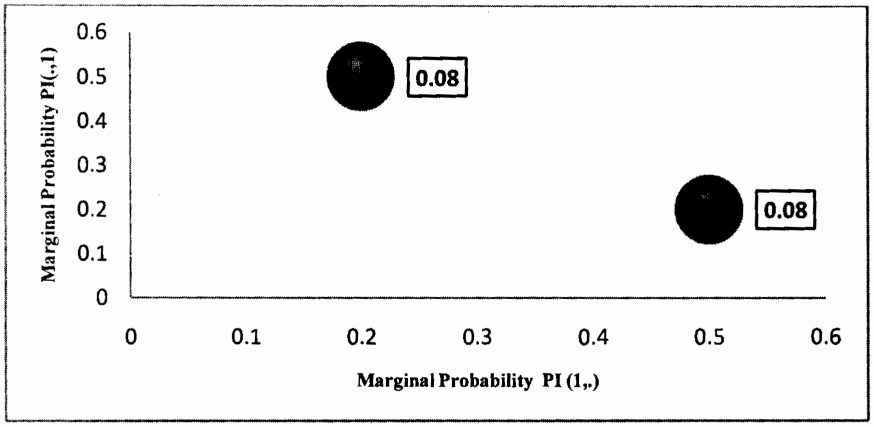


Table:-7.2.1.1Combinations of alternative where Chi-square performs worst

PI(1,.)	PI(.,1)	delta	AMP	CHI
0.2	0.5	0.08	58.03	15.69
0.5	0.2	0.08	58.51	10.48

Figure: - 7.2.1.2 Combinations of alternative where Fisher performs worst

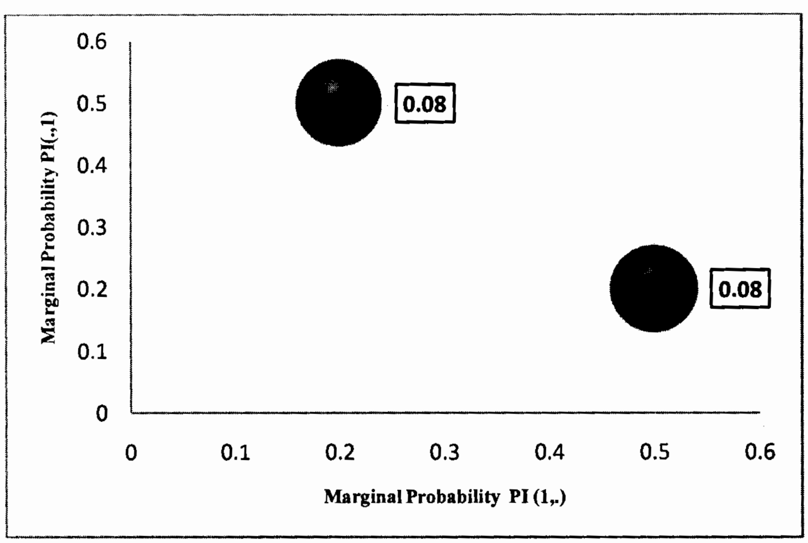


Table: - 7.2.1.2 Combinations of alternative where Fisher performs worst

PI(1,.)	PI(.,1)	delta	AMP	FIS
0.2	0.5	0.08	58.03	17.08
0.5	0.2	0.08	58.51	16.93

Figure: - 7.2.1.3 Combinations of alternative where Pearson performs worst

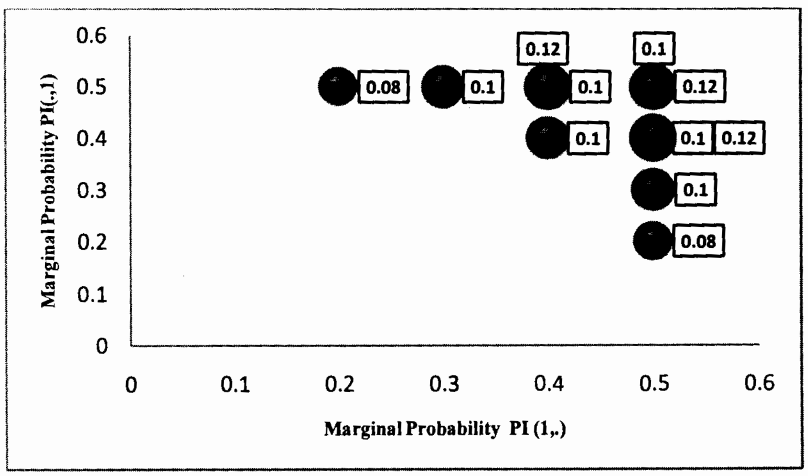


Table: - 7.2.1.3 Combinations of alternative where Pearson performs worst

PI(1,.)	PI(.,1)	delta	AMP	PEAR
0.5	0.4	0.12	80.39	39.06
0.4	0.5	0.1	65.29	23.29
0.4	0.4	0.1	67.41	25.18
0.3	0.5	0.1	69.95	27.65
0.4	0.5	0.12	82.16	39.47
0.2	0.5	0.08	58.03	15.05
0.5	0.3	0.1	70.39	27.14
0.5	0.4	0.1	67.13	23.35
0.5	0.5	0.12	80.07	35.93
0.5	0.2	0.08	58.51	14.24
0.5	0.5	0.1	66.53	15.96

7.2.2 When sample size is 50

Combinations of alternative where Chi-square performs worst at sample size of 50 are shown in figure 7.2.2.1 and Table 7.2.2.1.

Combinations of alternative where Pearson coefficient of correlation performs worst at sample size of 50 are shown in figure 7.2.2.2 and Table 7.2.2.2.

Figure: 7.2.2.1Combinations of alternative where Chi-square performs worst

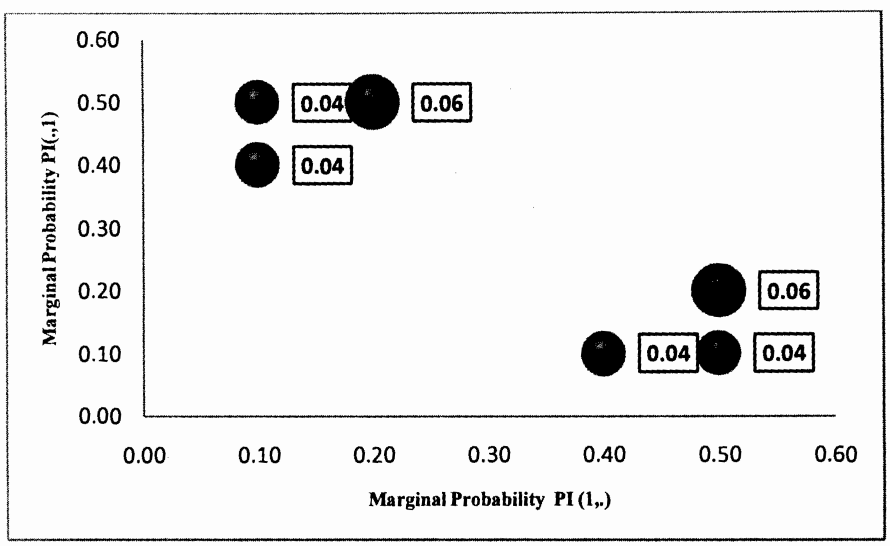


Table: 7.2.2.1Combinations of alternative where Chi-square performs worst

PI(1,.)	PI(.,1)	delta	AMP	CHI
0.40	0.10	0.04	51.86	14.36
0.10	0.50	0.04	52.89	15.01
0.10	0.40	0.04	53.03	14.55
0.20	0.50	0.06	69.14	29.73
0.50	0.10	0.04	54.36	14.13
0.50	0.20	0.06	70.30	29.73

Figure: - 7.2.2.2 Combinations of alternative where Pearson performs worst

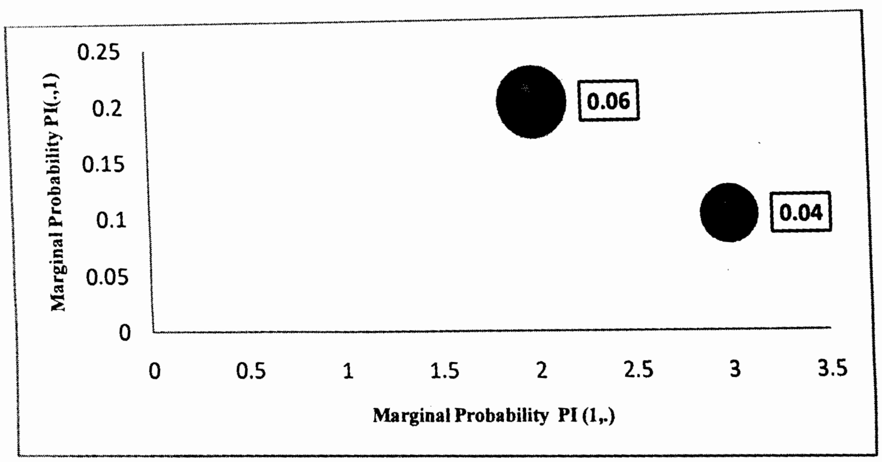


Table: - 7.2.2.2 Combinations of alternative where Pearson performs worst

PI(1,.)	PI(.,1)	delta	AMP;	PEAR
0.50	0.20	0.06	70.30	35.37
0.50	0.10	0.04	54.36	10.99

7.2.3 When sample size is 100

Combinations of alternative where Chi-square performs worst at sample size of 100 are shown in figure 7.2.3.1 and Table 7.2.3.1.

Combinations of alternative where Fisher's exact test performs worst at sample size of 100 are shown in figure 7.2.3.2 and Table 7.2.3.2.

Figure:-7.2.3.1Combinations of alternative where Chi-square performs worst

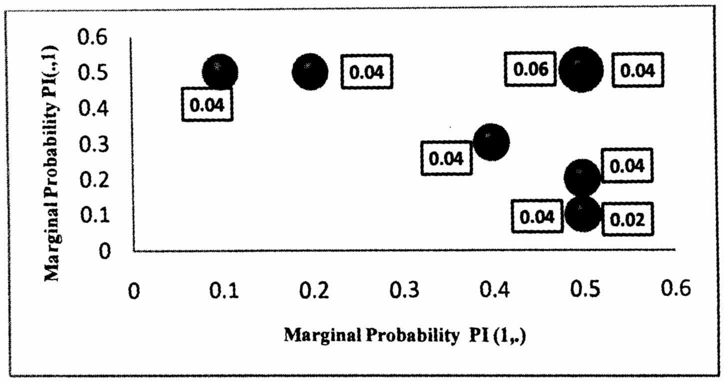


Table:-7.2.3.1Combinations of alternative where Chi-square performs worst

PI(1,.)	PI(.,1)	delta	AMP	CHI
0.5	0.2	0.04	64.82	44.72
0.5	0.5	0.04	50.82	30.55
0.4	0.3	0.04	55.76	34.97
0.5	0.1	0.02	37.74	16.38
0.5	0.5	0.06	82.64	60.74
0.5	0.1	0.04	87.23	64.97
0.2	0.5	0.04	67.15	44.74
0.1	0.5	0.04	87.12	59.96

Figure: - 7.2.3.2 Combinations of alternative where Fisher performs worst

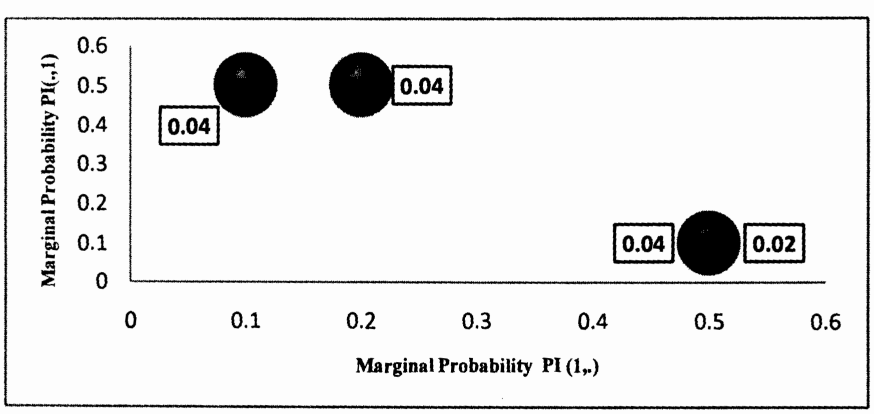


Table: - 7.2.3.2 Combinations of alternative where Fisher performs worst

PI(1,.)	PI(.,1)	delta	AMP	FIS
0.5	0.1	0.02	37.74	17.42
0.2	0.5	0.04	67.15	45.85
0.1	0.5	0.04	87.12	64.62
0.5	0.1	0.04	87.23	64.33

Figure: - 7.2.3.3 Combinations of alternative where Pearson performs worst

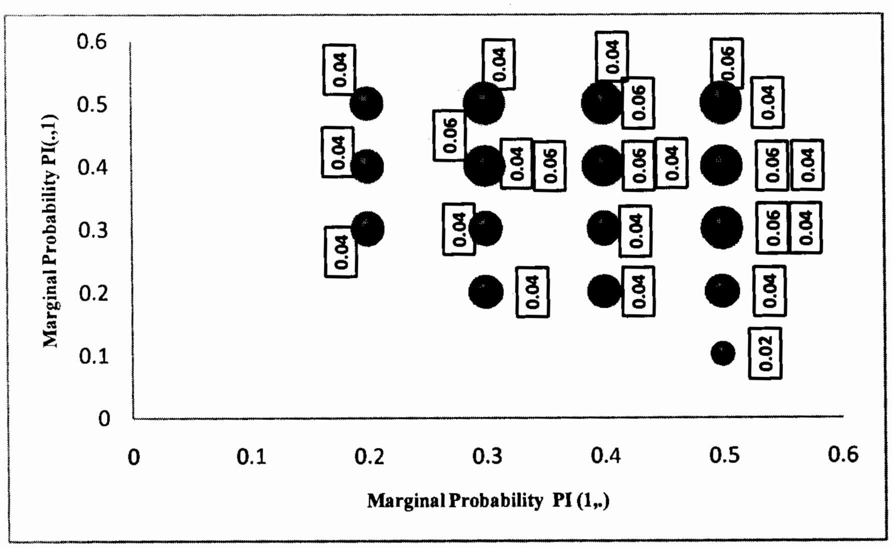


Table: - 7.2.3.3 Combinations of alternative where Pearson performs worst

PI(1,.)	PI(.,1)	delta	AMP	PEAR
0.3	0.4	0.06	85.37	65.08
0.5	0.3	0.06	84.44	62.98
0.4	0.5	0.06	78.43	56.96
0.3	0.2	0.04	66.83	45.33
0.3	0.5	0.06	85.43	63.43
0.5	0.1	0.02	37.74	15.73
0.3	0.3	0.04	57.94	35.35
0.5	0.4	0.06	79.87	57.08
0.4	0.4	0.06	80.86	57.67
0.4	0.4	0.04	50.49	27.18
0.4	0.2	0.04	64.17	40.59
0.4	0.5	0.04	49.03	25.33
0.2	0.3	0.04	68.73	44.95
0.5	0.4	0.04	49.42	25.35
0.2	0.4	0.04	64.96	40.84
0.3	0.4	0.04	54.75	30.37
0.4	0.3	0.04	55.76	31.04
0.5	0.2	0.04	64.82	39.65
0.5	0.3	0.04	53.77	28.56
0.3	0.5	0.04	55.09	28.93
0.5	0.5	0.04	50.82	23.26
0.5	0.5	0.06	82.64	53.92
0.2	0.5	0.04	67.15	33.37

Chapter 8

CONCLUSION, RECOMMENDATIONS AND DIRECTION FOR FUTURE RESEARCH

In this chapter the conclusion of this study and in the light of this conclusion a recommendation is given. The directions for future research are also discussed at the end of chapter.

8.1 Conclusion

In light of the conclusion of chapter number 5 (Section 5.5), it is observed that the Fisher's exact test of independence is the robust test in terms of Size. Any combination of independence can be taken for the computation of size of Fisher's exact test of Independence and distortion in size is the minimum of all the four tests of independence.

In chapter 6, it was concluded (Section 6.5) that the Fisher's exact test of independence is the robust test in terms of power, also the Fisher's exact test of independence is the Most Stringent test of all the four tests of independence.

So it is concluded that the Fisher's exact test of independence is the robust test in terms of both Size and Power and is also the Most Stringent test of all the four tests of independence.

8.2 Recommendation

As the Fisher's exact test of independence performs best as compared to all of the four tests of Independence/measure of Correlation for Categorical data on the basis of both Size of test and Stringency criterion. So the practitioners/researchers should use Fisher's exact test of independence when they are testing of independence for categorical data.

8.3 Direction for Future Research

This study can be extended to the data categorized in a two 3×3 contingency table or 4×4 contingency table or so on. Also more co-efficient of correlation/Tests of independence for categorical data can be included in future research.

REFERENCES

1. Abdi H (2007) "Kendall coefficient of rank correlation." Encyclopedia of Mathematics.
2. Barnhart H. X et.al "Comparison of Concordance Correlation Coefficient and Coefficient of Individual Agreement in Assessing Agreement." Duke University.
3. Cornbleet P. J and Shea M. C (1978) "Comparison of Product Moment and Rank Correlation Coefficient in the Assessment of Laboratory Method-Comparison Data." Clin. Chem, 24/6, 857-861.
4. Hauke J and Kossowski T "Comparison of values of Pearson and Spearman correlation coefficient on the same sets of data." Adam Mickiewicz University Poland.
5. Maturi T. A. (May 2010), "A Comparison of correlation coefficients via a Three-Step Bootstrap Approach." Journal of Mathematics Research, Vol.2, No.2.
6. Sheskin D. J. "Hand book of Parametric and Non Parametric Statistical Procedures." Second Edition.
7. Rodriguez G. (August 2010) "Log Linear Models for contingency tables."
8. Young R.F and Winn H.E (2003) "Activity patterns, diet and shelter site use for two species of moray eels, Gymnothorax Moringa and Gymnothorax Vicinus, in Belize." Copeia 2003, 44-55.
9. Zaman A. "Statistical Foundations for Econometrics Techniques."