Comparison of Different Measures of Correlation for Nominal Data

M. PHIL THESIS

(ECONOMETRICS)

ANSAR ABBAS

Reg. 01/MPhil- EST/ PIDE / 2014

SUPERVISOR

Dr. Atiq -ur- Rehman

Department of Econometrics and Statistics

Pakistan Institute of Development Economics (PIDE) Islamabad, Pakistan

بسم الله الرحمن الرحيم

DEDICATION

This humble effort is dedicated to
"My Family for Their Loving Wishes, Support Patience, Understanding and Guidance and All Those Who Seek Knowledge to Reach At Truth"

Pakistan Institute of Development Economics

CERTIFICATE

This is to certify that this thesis entitled: **"Comparison of Different Measures of Correlation for Nominal Data"** submitted by Mr. Anser Abbas is accepted in its present form by the Department of Econometrics and Statistics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree in **Master of Philosophy in Econometrics**.

Supervisor:

Dr. Atiq-ur-Rehman
Assistant Professor
PIDE, Islamabad

External Examiner:

Dr. Ateeq-ur Rahman Irshad
Assistant Professor
International Business & Marketing
NUST, Islamabad

Head,
Department of Econometrics and Statistics

10-4-2017.

Dr. Amena Urooj

Certificate

This is certified that Mr. Ansar Abbas Registration No: 01/MPhil- EST/ PIDE / 2014 thesis entitled **"Comparison of Different Measures of Correlation for Nominal Data"** submitted to the department of Econometrics and statistics and accepted in its original form. We have read this thesis and our opinion is that, this thesis satisfied all the requirements. This thesis is fully adequate in scope and quality as a thesis for the degree of Master of Philosophy in Econometrics.

Checked by ……………………………………….…………………………………..

Dr. Atiq – ur- Rehman,

Assistant Professor

Department of Econometrics and Statistics

Thesis Supervisor

Examining Committee:

Approval by …………………………………………………………………………..

External Examiner

Approval by …………………………………………………………………………..

Department of Econometrics and Statistics

Head of Department

Abstract

There are a number of tests available for testing association between categorical data arranged in contingency table. However, there is no clear guidance on the relative merits of these tests and about the appropriate choice of test, except for 2×2 contingency table. We evaluate various tests of independence including Pearson chi-square test of independence, Likelihood ratio chi-square test of independence, Goodman and Kruskal's lambda test, Uncertainty coefficient, Generalized McNemar's test (Stuart- Maxwell test), and Generalized fisher exact test (Fisher freeman-Halton test) on the basis of their size distortion and power for various alternatives by extensive Monte Carlo Simulation. We observe that there is no significant size distortion for all of these tests, therefore these tests are equivalent with respect to their size. However, the power of tests for various alternatives changes dramatically. We found that Generalized McNemar's test (Stuart- Maxwell test) outperforms other test in terms of power. Therefor we recommends the use of Generalized McNemar's test (Stuart- Maxwell test) for testing association in contingency table. So, this is the most powerful and robust test of independence/measures of correlation for nominal data.

Acknowledgment

All the Praise are to Allah, Almighty, Peace and Blessing of Allah on his Prophet Muhammad (SAW) who taught the Humanity to Human Beings.

A very special note of thanks goes to my parents whose heartfelt prayers, appreciation and support have always been a valuables asset and a great source of inspiration for me. They always encouraged me whenever I was demoralized during studies. They really deserve special thanks for enduring all my problems with great patience and love.

The acknowledgment will remain incomplete without special thanks to my supervisor Dr. Atiq-ur-Rehman Assistant Professor Department of Econometrics and Statistics Pakistan Institute of Development Economics (PIDE), Islamabad for his excellent cooperation with me during research.

I am also greatly indebted to Dr. Hafsa Hina Assistant Professor Department of Econometrics and Statistics Pakistan Institute of Development Economics, Islamabad for her valuable comments on my research proposal presentation and extend valuable cooperation throughout my research work. I am also thankful to all my teachers who taught me at any stage of my studies. I would like to acknowledge the teaching of Dr. Eatzaz Ahmad, Dr. Wasim Shahid Malik and Dr. Ahsan Ul Haq Satti for their kind style of teaching.

I would like to acknowledge many friends and colleague who help me to accomplish this document. Especial thanks to Mr. Shah Abbas Nagari who helped me at the early stage of M. Phil course work and provide me excellent company to enjoy.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background of Study

Correlation is one of the most important statistical measure that is often used by economists and social scientists. There are many ways to calculate the correlation for different types of data sets and measurement scales. The economic data to which correlation is applied can be categorized into various types. The data may be continuous and discrete (qualitative) in nature. The categorical data can be further divided into two scales of measurement that are mentioned below.

1. Nominal Scale
2. Ordinal or Rank Order Scale

The continuous data can be divided into ratio and interval scale. Qualitative data can be further divided into nominal, and ordinal or rank order scale. It is often said that researcher must be careful in the application of different tests of independence for different types of data. But however, no clear guidance is available that what kind of measure of correlation for a data. Many people argue that the use of Pearson product moment correlation co-efficient for continuous data is appropriate when our data possess the important assumption of normality. People apply Pearson correlation to categorical data as well. Whenever it is well known that this kind of data does not have normality. There is a huge amount of literature that exists about the interpretation, factors affecting the size of Person's r, its assumptions, sampling distribution, properties of sampling distribution, special cases of Person's r ,and applications of Person's for continuous data.

There are various tests of independence that can be applied to qualitative data. For example: Spearman rank correlation coefficient, Kendall tau co-efficient of rank correlation, Kendall's partial rank-order correlation, and many others. However, which one of is more appropriate, that yet to be explored.

The comparison of different measures of correlation for continuous and rank-order type of data exists in the literature. For example Cornbleet and Shea (1978) has compared Pearson product-moment correlation coefficient and rank correlation coefficient. (Barnhart, Lokhnygina, Kosinski, & Haber, 2007), they analyzed the Concordance Correlation Co-efficient and Co-efficient of Individual Agreement. The two studies had been conducted about the comparison of different measures of association.

There are many comparisons of correlation for continuous and rank order data but they are not sufficient to provide a complete guidance to the researcher about the uses and significance of the tests of independence because most of the tests of independence are being affected by table dimension, sample size, critical points of asymptotic, standard errors, and probability values. According to my knowledge, some researchers compared the different measures of association for nominal data on the basis of table dimension, sample size, critical points of asymptotic, and probability value but not on the size and power analysis. To fill this gap, we are conducting the study which compares six measures of correlation/independence for nominal data in terms of their size and power properties. Moreover, six tests of independence/measures of correlation are compared for two way $(R \times C)$ contingency table.

1.2 Objectives of the Study

The objective of this study is to explore size and power properties of various kinds of tests for correlation for nominal data and to find out optimal test for correlation for such data sets. Our first concern of this study is to set up a criterion to be used in comparing the performance of tests of independence.

1.3 Significance of the Study

The study enables the practitioner to distinguish between the best and the worst performance of a tests of independence for categorical data (nominal variables). It will also facilitate the researcher for the usage of a test of independence. This type of study also provides to some extent but not a complete guidance to a researcher that which of these tests of independence performs well when sample size, and categories or multiples of two nominal variables changes in different studies under different circumstances.

1.4 Plan of the Study

This study has following chapters:

First chapter describes about the introduction, 2nd chapter is about empirically/hierarchal literature review, 3rd chapter deals with theoretical framework, 4th chapter attempts to address data generating process, and comparing tests of independence for nominal data based on size and power analysis of the study. Chapter 5th and 6th deals with the computation of empirical size and power comparison and the last chapter of this study is devoted to summary, conclusion, and recommendations.

Chapter 2

Literature Review

2.1 Literature Review

The correct choice of statistical test for an experiment largely depends upon the nature of dependent and independent variables analyzed. Which statistical test is most appropriate? Should a parametric and non-parametric test be used? Parametric tests are appropriate when continuous variables follow a normal distribution and non-parametric tests are appropriate when they do not. An extensive literature is available for the comparison of the tests of independence when the variables are categorized into ratio, ordinal and rank-order levels. In many studies Pearson chi-square and likelihood ratio chi-square tests are comparatively analyzed based on observed and expected frequencies in 2x2 contingency tables.

In most of the studies it was observed that when non-normality of the distribution in any data set exists we cannot use Pearson product moment correlation co-efficient for continuous data. But in the study of Chok (2010), he concluded that, the permutation test, based on Pearson product moment correlation coefficient could offer a valuable advantage over Spearman's and Kendall's correlation co-efficient in non-normal distribution for continuous data. Hence, it is not a sole fact to disregard the use of Pearson co-efficient when non-normality of the distribution exists in a continuous data set but if when sample size guidelines are not met, these tests does not have sufficient power to provide meaningful results. We move toward other tests that are commonly used for categorical data.

Agresti and Kateri (2011) Analyzed that when expected frequency is less than five in contingency table, we must use likelihood ratio chi-square test instead of

Pearson Chi-square test as well as smallest expected frequency was at least one, the problem of type one error rate and misleading p- values arises under different sample size and different sampling designs. Some tests are comparatively analyzed based on type one error rate and p-values.

Cangur and Ankarali (2013) had compared Pearson Chi-square and Log-Likelihood Ratio test statistics in variety of conditions with respect to type I error rate which plays an important role in the selection of the tests. As a result of simulation, the Type I error rates of both tests are similar to each other at 5% when sample size is more than 100 and regardless of balance and unbalance of marginal row and column probabilities. When we regard balanced and unbalanced marginal row and column probabilities and sample size is less and 100 Type I error rate of Pearson Chi-square is at 5% and Type I error rate of Log-Likelihood Ratio is more 5% as a result Log-Likelihood Ratio test is more affected by the structure of the table, sample size and balanced and unbalanced marginal row or column probabilities than Pearson Chi-square test.

For example Lydersen and Laake (2003) had compared Exact Pearson's Chi-square, Likelihood ratio and Fisher's tests and their three versions that are standard, mid p, and randomized tests on the basis of the power and significance level for 2×2 contingency table using binomial and multinomial sampling. They concluded that for mid p type 1 error probabilities often exceed the nominal significance level. The mid p and randomized test versions have approximately the same power and their power is higher than the standard test version. When power of the Exact Pearson's Chi-square, Likelihood ratio and Fisher's tests differ then this difference of power occurs approximately in the same way for standard, mid p and randomized test versions. In many cases, Pearson's Chi-square and Fisher's tests have almost equal power and their power is higher than LR but when designs are poorly balanced, LR performs best. Fisher's test seems to be slightly more robust if the design is poor.

Özdemir and Eyduran (2005) Suggested that the values of chi-square and likelihood ratio chi-square statistic would be quite similar when observed frequencies in each cell would be more than five. However, it could be suggested that likelihood ratio (LR) chi-square statistics would be more effective than chi-square statistics when observed and expected frequencies would be less than five. Considering findings many authors suggested that to be made a very appropriate decision in selecting favorable statistics, their power values as well as probability values of two statistics should be examined. Consequently, to be making an exact decision on selecting favorable one of two statistics "power of test" for both statistics is more important concept than common concept on whether both observed and expected frequencies were less than five.

Lydersen, Fagerland, and Laake (2009) analyzed the critical points of asymptotic Pearson chi-square test for large samples and Fisher Irwin's called fisher's exact test for smaller samples. The asymptotic test may not preserve the test size that is, the actual significance level may be higher than the nominal significance level. Fisher's exact test is conservative that is other tests generally have larger power yet still preserve test size. To overcome the two important problems that is power and test size they extended their analysis on other significance tests depending on variety of choices including

- Level of conditioning in the sample space: should the p-value be computed conditionally or unconditionally on one or marginal sums in the observed table.
- Choice of test statistic, such as Pearson's of Fisher's.

- Exact or asymptotic calculation of p-value.
- Further adjustment such as the mid p-value.

The significance tests depending on above choices are listed below.

1. Pearson asymptotic chi-square test.
2. Fisher –Boschloo unconditional test.
3. Fisher exact conditional mid p-value test.
4. Fisher exact conditional.

Above four test are compared based on given choices and conclusion about the power and test size may be stated as follows.

When significance level is obtained based on fixed row sums then fisher's exact test is more conservative than Fisher's-Boschloo's unconditional test which also by definition preserves test size. Fisher conditional mid p-test and Fisher's-Boschloo's unconditional test performs equally and in this case mid p-test preserve test size. Pearson asymptotic chi-square test is neither conservative nor preserves test size. When power is a function of sample size then conditional tests has lowest power and mid p-test and unconditional tests has the same power. Hence unconditional tests and mid-p test are used more and more because they preserve the significance levels and have more power than traditional fisher exact test. The traditional Fisher's exact test practically can never be used. Behavior of nominal and ordinal variables is strongly influenced by sampling schemes and some other constraints.

Olszak and Ritschard (1995) Analyzed that important finding of simulations has relatively high discrepancy of the estimates that are based on nominal λ and ordinal γ of Goodman and Kruskal. So, to test the significance of association other measures should be used. So, direct and indirect linkages between variables should equal their partial association. Ordinal measures adopt such a constraint of partial association but nominal measures do not. So, the behavior of nominal and ordinal partial association measures should be interpreted with special care.

Göktas and Içi (2011) Has compared the tests of independence like that Spearman and Pearson correlation coefficient, gamma coefficient, Kendall's tau b, Kendall's tau c and summer's d for ordered contingency tables based on table dimension and sample size and determines which measure of association is more efficient. They found that as the number of table dimensions increases what the sample size is Pearson's and spearman's correlation coefficients increases on average but slightly they underestimate actual degree of ordinal measure of association. But Kendall's tau b, Kendall's tau c and summer's d increases as the table dimension increases but they always underestimate the actual degree of association. Although Pearson correlation coefficient is slightly larger than spearman' correlation coefficient in larger dimension table. But, at last they concluded that, Gamma co-efficient is good when table dimension and sample sizes are relatively small. It increases and over estimates as the sample size increases for any certain table dimension. In overall, Gamma co-efficient is best for square tables and preserves the actual degree of association in average.

Ferguson, Genest, and Hallin (2000) Showed that Kendall's tau can be used to test for a serial dependence in a univariate time series. They gave formula for both mean and variance of circular and non-circular versions of the statistic and proved its asymptotic normality under the hypothesis of independence. They further presented a Monte Carlo study comparing size and power based on Kendall's tau and related procedures that are based on alternative parametric and non-parametric measures of serial dependence. Their simulation results revealed that Kendall's tau outperforms

Spearman's rho in detecting first order auto regressive dependence, even though these two statistics are asymptotically equivalent under the null hypothesis.

Galla (1987) Non-parametric rank-order correlation tests such as Spearman's rho and Kendall's tau are often used as alternative to Pearson's r and their counterpart when the assumptions underlying these tests cannot be met. Generalized Kendall's tau is particularly used as alternative to a partial correlation coefficient.

Maturi and Elsayigh (2010) Has compared ten measures of correlation coefficients for rank-order data by using a three-step bootstrap procedure. These ten measures of correlation coefficient are listed below

(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 "ν*".

They used the standard error criterion for their comparison. 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 co-efficient has lowest standard error amongst other weighted correlation coefficients".

Newson (2002) Has reviewed the uses of Somers'D, Kendall's Tau and the Hodges-Lehmann median differences. The author found that the confidence limits for these parameters and their differences were more informative than the traditional practice of reporting only p-values.

For example (Agbedeyi and Igweze) has compared Tau-b, Tau-c and Gamma statistics and leads to the conclusion that gamma statistic is consistently better than tau-b and tau-c.

Similarly Hauke and Kossowski (2011) has compared the coefficients and statistical significance of Pearson's product- moment correlation and Spearman's rank correlation over the same sets of data.

Yang, Sun, and Hardin (2011) Argued that when data is matched pair with multiple categories, we commonly use Stuart -Maxwell and Bhapkar test for evaluating marginal homogeneity. But when data is collected in clustered matched-pair, two extended Obuchowski tests are proposed for this evaluating marginal homogeneity. A Monte Carlo simulation study illustrates that the extensions of Stuart –Maxwell and extended Obuchowski tests perform well with respect to power and nominal size. Although extended Bhapkar is asymptotically equivalent to the other three tests but it is not recommended due to its being liberal in the nominal size.

Islam (2012) Has compared four measures of correlation coefficients in 2×2 contingency table for categorical data (ordinal variables) based on size, power and stringency criterion. He 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. The clear distinction and contribution between these two studies is that I have compared different measures of correlation for nominal data and 3×3 contingency table.

It has been suggested that Likelihood ratio, Wald and Lagrange Multiplier tests are asymptotically equivalent and choice among them depends up computational convenience. But these tests (Zaman, 1996) have been compared graphically on the basis of stringencies and maximum power analysis. Plots of stringency show that LM test is extremely poor, Wald test comes in second, and LR test has small stringencies and maximum power of all alternatives. However, as a conclusion LR test is clearly winner.

4.2 Gap in Literature

The above literature revealed that there are no single criteria for the comparison of different measures of independence/association for nominal, and ordinal or rank-ordered data. The most important points of literature review of my study are summarized below.

Comparison of Pearson chi-square and likelihood ratio chi-square tests exists in literature based on observed and expected frequencies for only 2×2 contingency table. Likelihood ratio chi-square test is superior to Pearson chi-square test when expected frequency is less than five and limiting form of Pearson chi-square test is likelihood ratio chi-square test. Most of the authors suggested that the appropriate decision in selection of most favorable statistics is that their power and probability values should be examined. Pearson chi-square test for large samples and Fisher's exact test for small samples has been compared based on critical points of asymptotic. Most of the tests of independence for ordinal data have been compared based on table dimension, sample size and determines which measure of association is more efficient. Ten measures of correlation of rank-ordered data have been compared based on three step bootstrap procedure and standard error criterion. However, based on the literature review, it is concluded that there is lack of comparison of tests of independence for nominal data on the basis of size and power analysis.

In often tests of econometrics and statistics we know the form of sampling distribution of our test statistic when the sample size is large. When we are applying these tests with a sample that is relatively small, we must be careful because the actual sampling distribution of our statistic can be very different from its asymptotic distribution. If that case happens then the correct critical values for small samples can be quite different those ones if we choose them on the assumption that the asymptotic distribution is appropriate. We may conclude that we shall be using the significance level that is different from what we want that is called the size distortion of tests of independence for nominal data. This important point of size distortion is missing from the previous literature of the comparison of different measures of association for nominal, ordinal and rank- order data. So, in this study, I will analyze important point of size and power properties and few other basis of comparison for different measure of association for nominal data.

Variable Y	**Variable** X x_1 x_2,....................................,x_c	Row total and marginal probabilities

Chapter 3

Theoretical Framework

This chapter explains that what is the contingency table, how to interpret it, how to define association in this table and tests for association in contingency table?

3.1 What is Contingency Table?

The term contingency table was first introduced by Karl Pearson in 1900 and used in 1904 when he studied the mathematical statistics. A contingency table is a table of counts. A $k-$dimensional contingency table is formed by classifying subjects by two variables. One variable determines the row categories, the other variable defines the column categories. The combination of row and column categorize are called cells. For the mathematician, a $k-$dimensional contingency table with R rows and C column is the set $\{x_{i,j} = i = 1,2,.....,r, j = 1,2,.....,c\}$. Let the data is obtained for two variables X and Y that have fix number of individuals. Moreover, the data is displayed as $n_{i,j}$ in a table with R rows and C columns. Contingency table with cell frequencies, marginal totals with rows and column, joint probabilities with cell counts and marginal probabilities are depicted below.

y_1 y_2 $\vdots$ $\vdots$ y_r	$n_{11}(\pi_{11})$ $n_{12}(\pi_{12})\cdots\cdots n_{1c}(\pi_{1c})$ $n_{21}(\pi_{21})$ $n_{22}(\pi_{22})\cdots\cdots n_{2c}(\pi_{2c})$ $\vdots$ $\vdots$ $\vdots$ $\vdots$ $\vdots$ $\vdots$ $n_{r1}(\pi_{r1})$ $n_{r2}(\pi_{22})\cdots\cdots n_{rc}(\pi_{rc})$	$n_{1.}(\pi_{1.})$ $n_{2.}(\pi_{2.})$ $\vdots$ $\vdots$ $n_{r.}(\pi_{r.})$
Column total and marginal probabilities	$n_{.1}(\pi_{.1})$ $n_{.2}(\pi_{.2})\cdots\cdots n_{.c}(\pi_{.c})$	$n_{..}(1)$

3.2 How to Interpret Contingency Table?

Consider a $R \times C$ contingency table with both nominal categories with the explanatory variable X and response variable Y. Let $n_{i,j}$ denotes the cell frequencies and corresponding $\pi_{i,j}$ joint probabilities that an observation will fall in the *ith* category of Y and *jth* category of X that is $I = 1,2,......R, J = 1,2,....C$. We denote the sum of marginal totals of cell frequencies

And sum of corresponding marginal probabilities of *ith* category of Y and *jth* category of X as

$$n_{i.} = \sum_{i=1}^{R} n_{i,j}, n_{.j} = \sum_{i=1}^{C} n_{i,j}, \pi_{i.} = \sum_{i=1}^{R} \pi_{i,j} \text{ and } \pi_{.j} = \sum_{i=1}^{C} \pi_{i,j}$$

Row or column sum of marginal totals is equal to grand total and row or column sum of marginal probabilities is equal to unity i.e.

$$n_{.j} = \sum_{j=1}^{C} n_{i,j} = n_{i.} = \sum_{i=1}^{R} n_{i,j} = n_{..}$$

$$\text{And } \pi_{i.} = \sum_{i=1}^{R} \pi_{i,j} = \pi_{.j} = \sum_{j=1}^{C} \pi_{i,j} = 1$$

3.3 How to Define Association in Contingency Table?

The categorical (nominal) variables under two-way contingency table are statistically independent when their joint probabilities are equal to the marginal probabilities otherwise they are dependent i.e.

$$\pi_{i,j} = \pi_{.i}\pi_{j.} \text{ And } \pi_{i,j} \neq \pi_{.i}\pi_{j.}.$$

We formulate our null and alternative hypothesis for any test of independence in such a way that the two variables are independent and dependent. i.e.

$$H_o : \pi_{i,j} = \pi_{.i}\pi_{j.} \, or \, H_o : \pi_{i,j} - \pi_{.i}\pi_{j.} = 0 \;\forall\; i = 1,2,......,r \text{ and } j = 1,2, ...,c$$

$$H_1 : \pi_{i,j} \neq \pi_{.i}\pi_{j.} \, or \, H_1 : \pi_{i,j} - \pi_{.i}\pi_{j.} \neq 0 \;\forall\; i = 1,2,......,r \text{ and } j = 1,2, ...,c$$

3.4 Overview of Tests of Independence/Measures of Correlation

A brief overview of each test of independence/measures of correlation is given as

3.4.1 Pearson Chi- square Test of Independence

Pearson chi-square test χ^2 is a statistical test that was designed by Karl Pearson in 1900. It tests the null hypothesis that the joint probability $\pi_{i,j}$ that the outcome is in row i and column j is equal to the product of two marginal probabilities $\pi_{.i}\pi_{j.}$. i.e.

$$H_o : \pi_{i,j} = \pi_{.i}\pi_{j.} \, or \, H_o : \pi_{i,j} - \pi_{.i}\pi_{j.} = 0 \ \forall \ i = 1, 2, ..., r \ and \ j = 1, 2, ..., c$$

The alternative hypothesis states that the two variables of classification are not independent. Which is the test for independence of the attributes of X and Y that are categorized in $R \times C$ contingency table. Then the value of test statistic is

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{i,j} - E_{i,j})^2}{E_{i,j}}$$

The expected frequencies can be calculated by using the formula

$$E_{i,j} = \frac{n_{.i}\, n_{j.}}{n_{..}}$$

Or value of test statistic can also be calculated in terms of joint and marginal probabilities as

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{i,j} - \frac{n_{i.}n_{j.}}{n}\right)^2}{\frac{n_{i.}n_{j.}}{n}}$$

If H_0 is true this distribution follows approximately a chi-square distribution with $df = (r-1)(c-1)$ where r and c are number of rows and columns in contingency table. We reject H_0 at $\alpha = \alpha_0$ if $\chi_c^{\,2} \geq \chi^2_{\alpha 0}(r-1)(c-1)$.

The degree of freedom can be explained as follows: since the contingency table has kh categories the $df = (rc-1)$. But we need to estimate only $\{p_{i.}\}$ and $\{p_{.j}\}$ that of which satisfies constraint in $(0,1)$, thus we need to estimate $(r-1+c-1)$ parameters. Thus $df = rc-1-(r-1+c-1) = (r-1)(c-1)$.

3.4.2 Likelihood ratio chi-square test of independence

The likelihood ratio chi-square also known as G-test was first investigated by Robert R. Sokal and F. James Rohlf in1981. The asymptotic distribution of G-test that is chi-square distribution with same degree of freedom as the Pearson χ^2 test. Suppose that we have $R \times C$ contingency table in which $n_{i,j}$ individuals with row i and column j are classified independently. Let $\pi_{i,j}$ be the corresponding probability that an individual is classified in row i and column j so that $\pi_{i,j} > 0$ and $\sum_{i=1}^{r}\sum_{j=1}^{c}\pi_{i,j} = 1$.

Then the variables $n_{i,j}$ have a multinomial distribution with parameters N and $\pi_{i,j}$ for $i = 1,2,......, r$ and $j = 1,2,......, c$.

Now we consider the generalized likelihood ratio test. The likelihood function is

$$L(\pi / n) = \frac{n_{..}!}{\prod_{i=1}^{r}\prod_{j=1}^{c} n_{i,j}!}\prod_{i=1}^{r}\prod_{j=1}^{c}\pi_{i,j}{}^{n\,i,j} = A\prod_{i=1}^{r}\prod_{j=1}^{c}\pi_{i,j}{}^{n\,i,j}$$

Now the log likelihood function is

$$l(\pi/n) = \ln(A) + \sum_{i=1}^{r}\sum_{j=1}^{c} n_{i,j}\ \ln(\pi_{i,,j})$$

We must maximize this, subject to the constraint $\sum_{i=1}^{r}\sum_{j=1}^{c}\pi_{i,j} = 1$. We use the lagrangian multiplier theorem and differentiate the log likelihood function with respect to $\pi_{i,j}$, so that the maximum likelihood estimate of $\pi_{i,j}$ is $\pi_{i,j} = \frac{n_{i,j}}{n_{..}}$ for $i = 1,2,......, r$ and $j = 1,2,......, c$.

Hence it follows that

$$l(\hat{\pi}/n) = \ln(A) + \sum_{i=1}^{r}\sum_{j=1}^{c} n_{i,j} \ \ln\left(\frac{n_{i,j}}{n_{..}}\right).$$

Maximum likelihood estimate under restricted model that is

$$H_o : \pi_{i,j} = \pi_{.i}\pi_{j.}$$

With two constraints $\pi_{i.} = \sum_{i=1}^{R}\pi_{i,j} = \pi_{.j} = \sum_{i=1}^{C}\pi_{i,j} = 1$

$$l(\hat{\pi}_0/n) = \ln(A) + \sum_{i=1}^{r}\sum_{j=1}^{c} n_{i,j} \ \ln\left(\frac{e_{i,j}}{n..}\right).$$

Hence final form of likelihood ratio chi-square test statistic can be written as

$\lambda(y) = 2(URLLF - RLLF) = 2\sum_{i=1}^{r}\sum_{j=1}^{c} n_{i,j} \ \ln\left(\frac{e_{i,j}}{n..}\right)$ which follows chi-square distribution with degree of freedom $(r-1)(c-1)$.

Where

$URLLF$ = Un-restricted log likelihood function and

$RLLF$ = Restricted log likelihood function.

The Pearson chi-square test statistic and likelihood ratio chi-square test statistic are approximately equivalent as $n \to \infty$ with same degree of freedom $v = (r-1)(c-1)$. If H_0 is true this distribution follows approximately a chi-square distribution with $df = (r-1)(c-1)$. We reject H_0 at $\alpha = \alpha_0$ if $\chi_c^2 \geq \chi^2_{\alpha 0}(r-1)(c-1)$.

3.4.3 Goodman and Kruskal's Lambda Test

Goodman and kruskal's lambda (λ) is a measure of proportional reduction in error in contingency table analysis. This test was proposed by Goodman and Kruskal (1954). In probability theory and contingency table this measure is explained as follows. This measure describes the relative decrease in probability of making an error in predicting the values of Y when the values of X is known. This test measures the strength of association between dependent and independent variable in two-way contingency table analysis when two variables are categorized with nominal levels.

Consider X is explanatory variable and Y is response variable that are categorized in $R \times C$ contingency table with probability $p_{i,j}$ that will fall in the i^{th} category of X and j^{th} category of $Y\,(i=(1,2,......,r\,;j=1,2,.........,c)$. Goodman and kruskal defined the asymmetric and symmetric lambda (proportional reduction of error) measure as

Asymmetric Lambda (directional lambda)

i. Asymmetric lambda is calculated for columns C_j in contingency table that have X is explanatory variable and Y is response variable. It is denoted by $\lambda(R/C)$.

ii. Asymmetric lambda is calculated for rows R_i in contingency table that have Y is explanatory variable and X is response variable. It is denoted by $\lambda(C/R)$.

These two asymmetric lambdas symbolically expressed as:

$$\lambda(R/C) = \frac{\sum_{i=1}^{R} \max_j(\pi_{i,j}) - \max_j(\pi_{.c})}{1-\max_j(\pi_{.c})} \quad \text{And}$$

Its general standard error may be given as follows:

$$ASE_1 = \frac{\sqrt{\sum_{i=1}^{R}\sum_{j=1}^{C}\pi_{i,j}(\delta_{i,j}-\delta_j+\lambda\delta_j)^2-\lambda(R/C)}}{1-\max_j(\pi_{.c})}$$

Where

ASE = Asymptotic standard error

$$\delta_{i,j} = \begin{cases} 1 & \text{if } j \text{ is column index for } \max_j \pi_{i,j} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_j = \begin{cases} 1 & \text{if } j \text{ is index for } \max_j (\pi_{.c}) \\ 0 & \text{otherwise} \end{cases}$$

$$\lambda(C/R) = \frac{\sum_{j=1}^{C} \max_i (\pi_{i,j}) - \max_i (\pi_{r.})}{1 - \max_i (\pi_{r.})}$$

Its general standard error may be written as follows:

$$ASE_1 = \frac{\sqrt{\sum_{i=1}^{R}\sum_{j=1}^{C} \pi_{i,j} (\delta_{i,j} - \delta_j + \lambda \delta_j)^2 - \lambda (R/C)}}{1 - \max_j (\pi_{.c})}$$

Where

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i \text{ is row index for } \max_i \pi_{i,j} \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_i = \begin{cases} 1 & \text{if } i \text{ is index for } \max_i \pi_{r.} \\ 0 & \text{otherwise} \end{cases}$$

Symmetric Lambda (non-directional Lambda)

Now we consider the situation in which explanatory and response variables are not defined. PRE-measure is denoted by λ and symbolically may be defined as

$$\lambda = \frac{\sum_{i=1}^{R} \max_j (\pi_{i,j}) + \sum_{j=1}^{C} \max_i (\pi_{i,j}) - \max_j (\pi_{.c}) - \max_i (\pi_{r.})}{2 - \max_i (\pi_{r.}) - \max_j (\pi_{.c})}$$

Its general standard error may be obtained as follows:

$$ASE_1 = \frac{\sqrt{\sum_{i=1}^{R}\sum_{j=1}^{C} \pi_{i,j} \left[\delta_{i,j}^{r} + \delta_{i,j}^{c} - \delta_i^{r} - \delta_j^{c} + \lambda (\delta_i^{r} + \delta_j^{c})\right]^2 - 4\lambda^2}}{2 - \max_i (\pi_{r.}) - \max_j (\pi_{.c})}$$

Where

$$\delta_{i,j}^{r}=\begin{cases}1 & if\ i\ is\ the\ row\ index\ for \max_{j}\pi_{i,j}\\ 0 & otherwise\end{cases}$$

$$\delta_{i}^{r}=\begin{cases}1\ if\ i\ is\ index\ for\ \max_{i}\pi_{r.}\\ 0 \qquad otherwise\end{cases}$$

And where

$$\delta_{i,j}^{c}=\begin{cases}1\ if\ j\ is\ column\ index\ for\ \max_{i}\pi_{i,j}\\ 0 \qquad otherwise\end{cases}$$

$$\delta_{j}^{c}=\begin{cases}1\ if\ j\ is\ index\ for\ \max_{j}\pi_{.c}\\ 0 \qquad otherwise\end{cases}$$

Under the null hypothesis of independence, the symmetric version of Lambda test statistic asymptotically has standard normal distribution

$$Z_{calculated}=\frac{\lambda_{symmetric}}{ASE_1}.$$

3.4.4 Uncertainty Coefficient

The uncertainty coefficient is a measure of association for nominal variables. It was first introduced by Henri Theil in 1970 and is based on the concept of information entropy. We define the information function (uncertainty function) that quantify the information of any event. i.e.

$$I(\pi_i)=\ln\left(\frac{1}{\pi_i}\right)=-\ln(\pi_i)$$

Uncertainty of an event is low when π is close to one and high if π is close to zero.

Entropy: Entropy of a random variable X is defined as the expected value of its information function is called entropy and symbolically expressed as

$$H(x)=E[I(\pi)]=-\sum_{i=1}^{I}\pi_i\ln(\pi)$$

Now we define the entropies for rows and columns respectively

$$H(x) = -\sum_j \pi_{.j} \ln(\pi_{.j}) \text{ And } H(y) = -\sum_i \pi_{i.} \ln(\pi_{i.}) .$$

Entropy of two rows and columns together

$$H(x, y) = -\sum_{i,j} \pi_{i,j} \ln(\pi_{i,j}) .$$

Entropy of y given x and x given y

$$H(y/x) = -\sum_{i,j} \pi_{i,j} \frac{\pi_{i,j}}{\pi_{j.}} \text{ And } H(x/y) = -\sum_{i,j} \pi_{i,j} \frac{\pi_{i,j}}{\pi_{i.}}$$

When x and y are independent then

$$H(y/x) = H(y) \text{ And } H(xy) = H(x) + H(y)$$

The uncertainty coefficient for $U(y/x)$, $U(x/y)$ and overall for both rows and columns is the proportion of entropy in the variable $y(x)$ explained by $x(y)$

$$U(y/x) = \frac{H(y) - H(y/x)}{H(y)} = \frac{H(x) + H(y) - H(xy)}{H(y)}$$

The standard error under the null hypothesis that $U(y/x)$ equals zero is computed as:

$$ASE_0 = \frac{\sqrt{p - [H(x) + H(y) - H(xy)]^2}}{H(y)}$$

Where

$$p = \sum_{i,j} \pi_{i,j} \ln\left(\frac{\pi_{.c}\pi_{r.}}{\pi_{i,j}}\right)^2$$

$$U(x/y) = \frac{H(x) + H(y) - H(xy)}{H(x)}$$

The variance under the null hypothesis that $U(x/y)$ equals zero is computed as:

$$ASE_0 = \frac{\sqrt{p - [H(x) + H(y) - H(xy)]^2}}{H(x)}$$

The symmetric version of uncertainty coefficient is defined as follows:

$$U = \frac{2\left[H(x) + H(y) - H(xy)\right]}{H(x) + H(y)}$$

The asymptotic variance under the null hypothesis that U equals zero is computed as:

$$ASE_0 = \frac{2}{H(x) + H(y)}\sqrt{p - [H(x) + H(y) - H(xy)]^2}$$

This measure lies between 0 and 1. Value 0 indicates that x and y have no association and 1 indicates that x and y are completely associated.

Under the null hypothesis of independence or no association, the asymptotic test statistics has standard normal distribution which is given as

$$Z_{calculated} = \frac{U_{symmetric}}{ASE_0}.$$

3.4.5 Generalized McNemar's test (Stuart- Maxwell test)

Generalized McNemar's test was first proposed by Stuart (1955) and Maxwell (1970) and this test is used to test the marginal homogeneity of the outcomes when two nominal variables are classified into multiple categories. To test the marginal homogeneity of the outcomes of two nominal variables in a square contingency table, we assess the following hypothesis. i.e.

$$H_0 : \pi_{i.} = \pi_{.i} \qquad \forall i = 1, 2, ..., I$$

$$H_A : \pi_{i.} \neq \pi_{.i}$$

The Stuart-Maxwell statistic is calculated as:

$$X^2 = d' S^{-1} d \text{, With degree of freedom } k - 1$$

Where d is a column vector that contains any $k-1$ values of $k \times k$ contingency table i.e. $d_1, d_2, \ldots\ldots., d_k$

Where

$$d_i = \pi_{.i} - \pi_{i.} \qquad i = 1,2,\ldots\ldots.., k$$

Let S denotes the variance covariance matrix of the elements of d. The elements of S are equal to:

$$s_{ii} = \pi_{i.} + \pi_{.i} - 2\pi_{ii}$$

$$s_{ij} = -(\pi_{ij} + \pi_{ji})$$

Where d' is the transpose of d and S^{-1} is the inverse of S matrix?

Stuart Maxwell test statistic X^2 is an asymptotically chi-square test statistic χ^2 with degree of freedom $k-1$ as $N \to \propto$. For 2×2 contingency table McNemar's and generalized McNemar's both are equal.

3.4.6 Generalized Fisher Exact Test (Fisher Freeman-Halton Test)

The generalization of the fisher exact test is known as the Fisher-Freeman-Halton (F-F-H) test (Freeman and Halton 1951). When sample sizes and asymptotic approximation do not work properly, then to perform the exact inference the marginal totals are fixed by design. In general, for a $R \times C$ contingency table, we test following hypothesis for F-F-H test

$$H_o : \pi_{i,j} = \pi_{.i}\pi_{j.} \quad \forall\; i = 1, 2, \cdots\cdots, r \;\; and \;\; j = 1, 2, \cdots\cdots, c$$

$$H_A : \pi_{i,j} \neq \pi_{.i}\pi_{j.}$$

The exact tests such as F-F-H tests are based on the multiple Hypergeometric distributions. The main objective of this test is that to remove the effect of all nuisance parameters from the distribution of y to make exact inference for categorical data. Thus, for F-F-H test we define a reference set. i.e.

$$\Re = \left\{ y : y, is.r \times c \sum_{i-1}^{r} n_{i,j} = n_{j.} ; \sum_{i-1}^{r} n_{i,j} = n_{.i} \right\}$$

We will use X to denote $R \times C$ table constructed from the sample data and Y to denote $R \times C$ table constructed from $\Re$. The probability of Y that belongs of $\Re$.i.e. $Y \in \Re$ under the null hypothesis that rows and columns are independent is given by

$$p(y) = \frac{\prod_{i=1}^{r} n_{.i}! \prod_{j=1}^{c} n_{j.}!}{N! \prod_{i=1}^{r} \prod_{j=1}^{c} n_{i,j}!}$$

The probability value $p\ (y)$ is compared with the value of the discrepancy measure (test statistic) $D(y)$ to perform the exact test of independence for null hypothesis. Large value of $D(y)$ provides evidence against null hypothesis and small value of $D(y)$ is consistent with null hypothesis. The test statistic for F-F-H test is denoted by $D_0(x)$ and is calculated as

$$D_0(x) = -2\log(\lambda(x)),$$

Where

$$\gamma = (2\pi)^{(r-1)(c-1)/2} (N)^{-(rc-1)/2} \prod_{i=1}^{r} (n_{i.})^{(c-1)/2} \prod_{j=1}^{c} (n_{.j})^{(r-1)/2}$$

The exact p value for testing the null hypothesis under F-F-H test is denoted by p_2 and calculated by accumulating the null probabilities of all the $R \times C$ tables in $\Re$ that are at least as extreme as the observed table x by using F-F-H test statistic.

$$p_2 = pr\left[D(y) \geq D_0\,(x) \right] = \sum_{y:D(y) \geq D_0(x)} p\ (y).$$

Chapter 4

Generation of Data and Comparing Test of Independence

This chapter explains that how data is generated. The six tests of independence/measures of correlation (Pearson chi-square test of independence, Likelihood ratio chi-square test of independence, Goodman and Kruskal's lambda test, Uncertainty coefficient, Generalized McNemar's test (Stuart- Maxwell test), Generalized fisher exact test (Fisher freeman-Halton test)) are compared in this study based on size and power analysis. In this chapter, the stepwise procedure for Monte-Carlo Simulation is explained and end of the chapter a brief conclusion is also stated.

4.1 Data Generating Process (DGP) in 3×3 Contingency Table

To generate the data of contingency table for nominal variables, there are lots of techniques in the literature of statistical simulation. The following method of data generating process is used for null hypothesis of independence

$$H_o : \pi_{i,j} = \pi_{.i}\pi_{j.} \; or \; H_o : \pi_{i,j} - \pi_{.i}\pi_{j.} = 0$$

We are dealing with categorical data following the above $R \times C$ contingency table, the data generating process is as follows.

Row selection

For each draw row will be selected in such a way that the probability of selection of each row follows the marginal probability of that row according to the contingency table.

Column Selection

Once the selection of row is done, the column will be selected in such a way that conditional probability of each column as per contingency table is observed.

Program which computes the data generating process and simulation design is given in appendix of this thesis.

4.2 Comparing Test of Independence

4.2.1 Basis of comparison

The comparison of various measures of association is based on the various assumptions. For example, Pearson chi-square test of independence is more sensitive to the sample size and dimension of the contingency table. So, when the sample sizes and table dimension differ the value of chi-square test statistic and its significance may lead to a researcher on wrong conclusion. To detect such type of effects a researcher must compute the size and power of a test. The power of a hypothesis test is the probability when a test correctly rejects the null hypothesis when alternative hypothesis is true and size of a test is vice versa. So, size and power properties are used to analyzed the performance of tests of independence.

The comparison of the tests based on size and power do not provide a satisfactory conclusion at different alternatives. As one test is more powerful at some alternatives as compare to other tests and at some alternatives may be any other test is more powerful. To overcome this problem and to calculate the best point optimal test stringency criteria is used. Point optimal test can be perform only when a test having one dimensional parameter of interest with a one-sided alternative hypothesis. The class of tests that is included in our study have higher dimensions, or two sided alternatives, it is easily seen that point optimal test cannot perform reasonably. However, for practical it is more important that tests which maximizes power over suitably distant alternatives are also preferred.

The test should have no size distortion when applied to data fulfilling the null hypothesis. If there are multiple designs of contingency table each fulfilling the null of independence, the test should have same power for all such designs. This can be assumed if the simulated critical values for different designs are same. If we are having comparable size, the power would help in choice of optimal test. So, two characteristics, size and power, have essential importance in assessing a test.

4.3 Simulation Design

We use the Monte-Carlo Simulation design to obtain the most powerful test the stepwise procedure is adopted. The stepwise procedure that is used to estimate the size and power of statistical tests of independence/measure of correlation for nominal data is stated as follows.

(i) Generation of data in 3×3 contingency table using those different designs of contingency table which satisfy the null hypothesis of independence and calculation of size of test for six tests of independence/measure of correlation i.e. Likelihood ratio chi-square test of independence, Goodman and Kruskal's lambda test, Uncertainty coefficient, Generalized McNemar's test (Stuart-Maxwell test), Generalized fisher exact test (Fisher freeman-Halton test).

(ii) Generation of data in 3×3 contingency table using those different designs of contingency table which satisfy the alternative hypothesis of non-independence and calculation of power of these six tests of independence/measure of correlation at different alternatives.

(iii) The above two steps will be repeated for each test of independence 20,000 times with sample sizes of 25, 50 and 100 and levels of significance 1% and 5%.

(iv) Calculate the p value for different alternatives of no association. This value will be called the proportion of rejection and will serve as the power of each test of independence.

(v) The powers of each test of independence is plotted against the alternative hypothesis to get the power curves for these tests.

4.4 Conclusion

To compute the power of test statistic T which depends upon critical value (CV) and alternative hypothesis (ρ) needs two counters n and m. A new value of T is generated and update the total counter $n+1$. The success counter m is updated $m = m+1$ only if $T > CV$. At the end, we calculate the probability of rejection to be m/n. The scatter plot of probability of rejection and alternative hypothesis will generate the power curve.

Chapter 5

Computation of Simulated Critical Values

And Empirical Size

5.1 Introduction of Chapter

This chapter explains that why the simulated critical values (finite sample critical values) are required for different test statistics? Simulated critical values are also calculated in this chapter for each test of independence/measure of correlation. We use a Monte-Carlo simulation approach which could be helpful in calculating the size and distortion of size of each test of independence and at the end of this chapter conclusion is stated.

5.2 Why the simulated critical values are needed?

In often cases we know the form of sampling distribution of our test statistic when the sample size is relatively large. So, we know the asymptotic sampling distribution. But if we are applying these tests with a sample that is relatively small, we have to be careful because the actual sampling distribution of our test statistic can be very different from its asymptotic distribution. If that case happens, then correct critical values for small samples can be quite different if we choose them on the assumption that the asymptotic distribution is appropriate. Thus, in turn, we are using the significance level that is different what we want. This creates the distortion in size of a test. Actual sampling distribution of our test statistic could be quite different what we want. It usually depends upon the testing problem, unknown values of the parameters and nature of the data that we are using. So, in order to keep the size of a test constant at nominal size of 5% and 1% the simulated critical values for each test of independence are generated at sample sizes 25, 50 and 100.

5.3 Computation of size of test based on simulated critical values

After the generation of data under the null hypothesis, we generate the critical values for six tests of independence/measures of correlation. These critical values for each test of independence is calculated under sample size of 25, 50 and 100, levels of significance α = (5%, 1%) and 20,000 Monte-Carlo replications.

Table 5. 1: Size of Tests of Independence with Sample Size 25

(Simulated Actual – Finite Sample Critical Values)

List of Tests	Levels of Significance	Sample size 25 and $\alpha = 0.05, 0.01$									
		D1	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
$Chi(\chi^2)$ $Test$	$\alpha = 0.05$	9.25	9.25	9.27	9.17	9.17	9.20	9.39	9.14	9.22	9.23
	$\alpha = 0.01$	12.27	12.62	12.80	12.87	12.87	12.51	12.69	12.62	12.58	12.60
$LR.Chi$ (χ^2) $Test$	$\alpha = 0.05$	6.14	5.98	5.87	5.65	5.65	6.05	5.88	5.81	6.09	6.06
	$\alpha = 0.01$	8.26	7.89	8.05	7.80	7.80	8.12	7.92	7.79	8.16	8.14
$Lembda$ (λ) $Test$	$\alpha = 0.05$	2.34	2.34	2.52	2.18	2.18	2.34	2.34	2.34	2.34	2.34
	$\alpha = 0.01$	3.07	3.07	3.41	3.11	3.11	3.07	3.15	3.15	3.07	3.07
UC $Test$	$\alpha = 0.05$	1.12	1.09	1.10	1.09	1.09	1.10	1.09	1.08	1.11	1.11
	$\alpha = 0.01$	1.50	1.46	1.51	1.53	1.53	1.50	1.47	1.50	1.50	1.49
$S-M$ $Test$	$\alpha = 0.05$	9.42	7.41	6.09	6.30	6.30	8.37	7.23	6.63	8.58	10.39
	$\alpha = 0.01$	12.16	10.12	8.47	8.82	8.82	11.13	10.00	9.36	11.17	13.14
$G.F.E$ $Test$	$\alpha = 0.05$	8.68	8.63	9.78	12.73	12.73	8.58	8.79	8.83	8.63	8.61

As we can see, the results depend to some degree on the form of contingency table. The above table shows that the size of each test of independence is less or equal to the nominal size of 5% and 1%, as we have used simulated critical values. For all α levels considered, the empirical significance levels for each test of independence are consistent with the nominal significance levels, as the nominal significance levels are with in the corresponding 95% and 99% confidence intervals.

Table 5. 2: Size of Tests of Independence with Sample Size 50

(Simulated Actual – Finite Sample Critical Values)

		Sample size 50 and $\alpha = 0.05, 0.01$									
List of Tests	**Levels of Significance**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
$Chi(\chi^2)$ $Test$	$\alpha = 0.05$	9.39	9.30	9.28	9.64	9.22	9.38	9.80	9.39	9.29	9.35
	$\alpha = 0.01$	12.63	12.89	12.72	13.07	12.95	12.85	13.48	13.15	12.95	12.59
$LR.Chi$ (χ^2) Tes	$\alpha = 0.05$	8.62	8.46	8.34	8.64	7.69	8.52	8.60	8.21	8.45	8.59
	$\alpha = 0.01$	11.82	11.68	11.29	11.95	10.80	11.88	11.77	11.39	12.05	11.92
$Lembda$ (λ) $Test$	$\alpha = 0.05$	1.58	1.53	1.41	1.64	1.21	1.58	1.53	1.58	1.58	1.57
	$\alpha = 0.01$	2.05	1.98	1.98	2.26	1.89	2.05	2.11	2.19	2.05	2.05
UC $Test$	$\alpha = 0.05$	0.78	0.77	0.79	0.84	0.78	0.77	0.80	0.78	0.76	0.78
	$\alpha = 0.01$	1.06	1.07	1.07	1.16	1.10	1.07	1.09	1.09	1.08	1.07
$S-M$ $Test$	$\alpha = 0.05$	13.47	9.47	6.40	6.54	7.17	11.22	9.06	8.01	11.68	15.49
	$\alpha = 0.01$	17.36	13.05	9.52	9.68	10.33	14.98	12.60	11.39	15.40	19.52
	$\alpha = 0.05$	8.95	8.91	8.77	9.17	8.51	8.97	9.31	8.79	8.89	8.94
$G.F.E$ $Test$	$\alpha = 0.01$	12.18	12.26	11.94	12.32	11.86	12.47	12.83	12.18	12.33	12.19

As we can see, the results depend to some degree on the form of contingency table. The above table shows that the size of each test of independence is less or equal to the nominal size of 5% and 1%, as we have used simulated critical values. For all α levels considered, the empirical significance levels for each test of independence are consistent with the nominal significance levels, as the nominal significance levels are with in the corresponding 95% and 99% confidence intervals.

Table 5. 3: Size of Tests of Independence with Sample Size 100

(Simulated Actual – Finite Sample Critical Values)

	Sample size 100 and $\alpha = 0.05, 0.01$										
List of Tests	**Levels of Significance**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
$Chi(\chi^2)$ $Test$	$\alpha = 0.05$	9.45	9.38	9.46	10.43	9.37	9.41	10.49	9.48	9.37	9.46
	$\alpha = 0.01$	13.00	12.80	13.17	14.21	12.99	13.07	14.26	13.08	13.14	13.02
$LR.Chi$ (χ^2) $Test$	$\alpha = 0.05$	9.67	9.55	9.41	10.39	9.08	9.62	10.71	9.27	9.58	9.63
	$\alpha = 0.01$	13.46	13.22	13.24	14.32	12.57	13.38	14.50	13.09	13.43	13.41
$Lembda$ (λ) $Test$	$\alpha = 0.05$	1.12	1.01	0.86	1.02	0.52	1.06	1.06	1.12	1.06	1.04
	$\alpha = 0.01$	1.41	1.34	1.25	1.57	0.98	1.41	1.43	1.48	1.39	1.39
UC $Test$	$\alpha = 0.05$	0.43	0.43	0.44	0.51	0.47	0.43	0.49	0.44	0.43	0.43
	$\alpha = 0.01$	0.60	0.60	0.63	0.71	0.65	0.60	0.67	0.62	0.60	0.60
$S - M$ $Test$	$\alpha = 0.05$	20.01	12.68	6.91	7.36	8.46	15.89	11.81	9.94	16.79	24.02
	$\alpha = 0.01$	25.44	17.13	10.45	10.77	12.03	20.88	16.19	13.97	21.70	29.30
$G.F.E$ $Test$	$\alpha = 0.05$	9.30	9.22	9.26	10.21	9.05	9.27	10.36	9.26	9.24	9.32
	$\alpha = 0.01$	12.94	12.75	12.87	14.02	12.47	12.85	14.18	12.77	12.96	12.90

As we can see, the results depend to some degree on the form of contingency table. The above table shows that the size of each test of independence is less or equal to the nominal size of 5% and 1%, as we have used simulated critical values. For all α levels considered, the empirical significance levels for each test of independence are consistent with the nominal significance levels, as the nominal significance levels are with in the corresponding 95% and 99% confidence intervals.

5.3.1 Comparison on the basis of size of a test

Table 5. 4: Distortion in size (using simulated critical values).

	Mean Deviation					
	Chi-Square Test	LR Test	Lambda test	UC Test	Stuart-Maxwell Test	Generalized Fisher Exact Test
Sample Size 25	0.05	0.15	0.06	0.01	1.21	1.29
	0.13	0.15	0.07	0.02	1.26	0.15
Sample Size 50	0.13	0.20	0.08	0.01	2.49	0.19
	0.19	0.30	0.08	0.02	2.75	0.33
Sample Size 100	0.33	0.34	0.12	0.02	4.63	0.41
	0.38	0.38	0.10	0.03	5.24	

We will be interested in whether our results are sensitive to the form of the data. As we can see, these results depend to some degree on the form of the contingency tables. If we were to use the asymptotic critical values as an approximation, we shall be using critical values that are too small. As a result, asymptotic critical values are too small, so significance level that we are using on simulated critical values would be larger from those one that we are using on asymptotic critical values. Significance level that we are using on simulated critical values is called nominal significance level of nominal size of the test. The difference between two quantities is the degree of size distortion. Hence asymptotic critical values would lead to very misleading results. This misleading result would be in terms of size distortion.

5.4 Conclusion

To test whether two nominal variables are independent or not, it is necessary to know the critical values at different sample sizes and significance levels. Initially we have critical values in the literature but these critical are not available in a wide range for us. So, we have generated extent tables of critical values for different sample sizes and significance levels using Monte-Carlo simulation for testing independence for two nominal variables. Based on those simulated critical values, we conclude that the space of null hypothesis is not a single criterion, rather it contains many points. Stuart-Maxwell test is more sensitive to the form of the degree of contingency table. This test creates huge distortion in size if we use asymptotic critical values.

Chapter 6

Results of Power Comparisons

In this section, we present the results of empirical power comparisons for six tests of independence for nominal data. Asymptotic critical values do not work for small sample size, because these critical values only work when sample size is large. Therefore, we do not use asymptotic critical values. Since we are using simulated critical values, this ensures that there is no size distortion.

6.1 Power of Tests

We determine the empirical power of each test of independence by simulating the contingency tables under the dependence structure, and computing the proportion of times the independence hypothesis is rejected at a given significance levels α. We generate 20,000 (3×3) contingency tables with sample sizes 25, 50 and 100 based on a specified dependence structure. For each sample, we compute the test statistic for Pearson chi-square test of independence, Likelihood ratio chi-square test of independence, Goodman and Kruskal's lambda test, Uncertainty coefficient, Generalized McNemar's test (Stuart- Maxwell test), Generalized fisher exact test (Fisher freeman-Halton test). We reject the independence hypothesis at significance level α if the test statistic of each test exceeds the critical value at level α. For several (3×3) contingency tables under dependence structure, the critical values and proportion of times the independence hypothesis is rejected are given in appendix for 5% and 1% significance levels.

6.2 Comparison based ON Power Curves

In this empirical study, we show the power curve for all tests for testing $H_o : \pi_{i,j} = \pi_{.i}\pi_{j.}$, against $H_O : \pi_{i,j} \neq \pi_{.i}\pi_{j.}$. To draw the power plots at different alternatives or possible values of distance are taken along X-axis and power of each test of independence/measure of correlation are taken along Y-axis. Scatter plots are drawn at three different sample sizes of 25, 50 and 100 for different cut-off points of distance or alternative hypothesis. Significance levels have been chosen to be 5% and 1%.

Fig.6.1.1

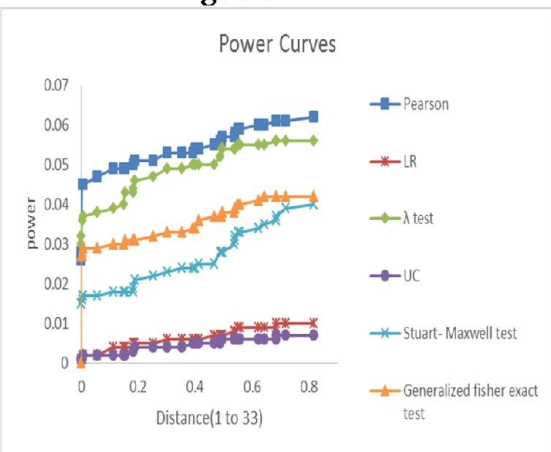


Fig.6.1.2

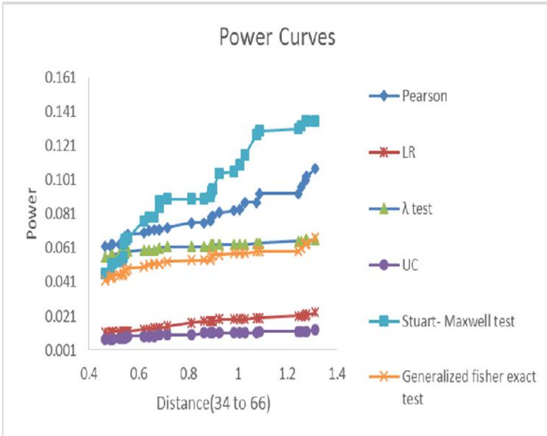


Fig.6.1.3

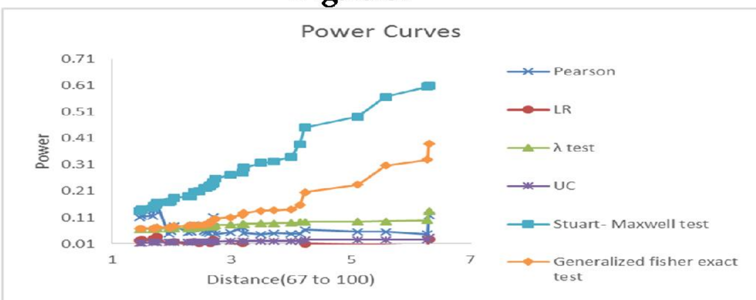


Figure 6. 1: Power curves at sample size 25 and α =5%:

Empirical study show the power plots of all six tests included in our study. In calculating the power of every possible value of distance or alternative hypothesis, we observe that when the distance is close to zero, the power of test is close to nominal size or type I error. As the distance increases, the power of test also increases. It is also visible from the graph that the power curve of Stuart-Maxwell test seems to outperform all other tests as its power about null hypothesis is close to the nominal size and it attains highest power for larger alternatives. Moreover, we observed that LR test and UC were completely insensitive to this type of dependence, and power of these tests was as low as the size of the tests. Note that when the sample size is 25 the power of all tests is less than significance level (α) or size of test. The tests are biased for the chosen values of the parameters.

Similar examples, such as one of the two figures (6.1.2 & 6.1.3) depicted above exhibit that increasing the power of all tests as well as increase over the part of parameter space. The empirical study shows that the Stuart-Maxwell test is more powerful up to highest parameter space. This test is unbiased for the chosen values of the parameters.

Figure 6. 2: Power curves at sample size 25 and α=1%:

Fig.6.2.1

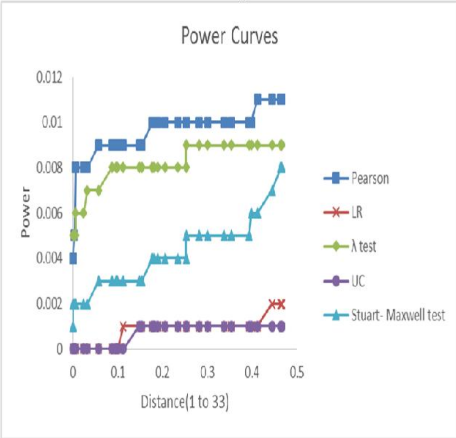


Fig.6.2.2

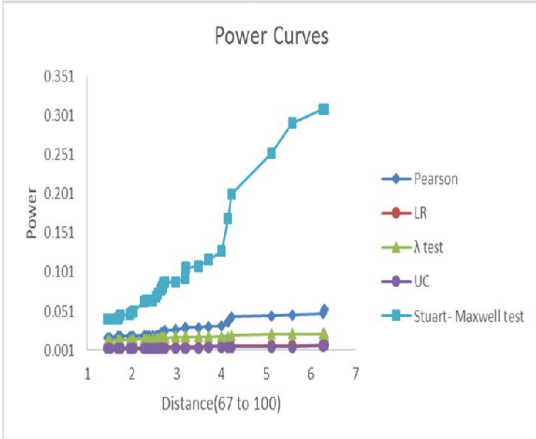
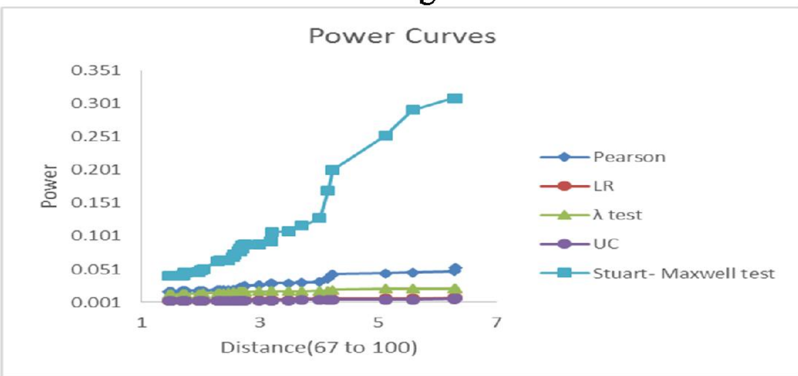


Fig.6.2.3



Above graph shows the power curves of all six tests included in our study. In calculating the power of every possible value of distance or alternative hypothesis, we observe that when the distance is close to zero, the power of test is close to nominal size or type I error. As the distance increases, the power of test also increases. It is also visible from the graph that the power curve of Stuart-Maxwell test seems to outperform all other tests as its power about null hypothesis is close to the nominal size and it attains highest power for larger alternatives. Two tests LR and UC perform very poorly in terms of power and their power is not very different from nominal size. Note that when the sample size is 25 the power of all tests is less than significance level (α).The tests are biased for the chosen values of the parameters.

Similar examples, such as one of the two figures (6. 2.2 & 6.2.3) depicted above exhibit that increasing the power of all tests as well as increase over the part of parameter space. The most powerful test is the Stuart-Maxwell test over increase in the part of parameter space. This test is unbiased for the chosen values of the parameters. But, in resulting as well as level of significance decreases, also decrease in power of all tests.

Figure 6. 3: Power curves at sample size 50 and α = 5%:

Fig.6.3.1

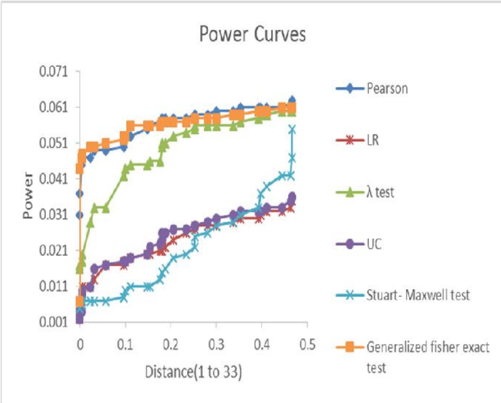


Fig.6.3.2

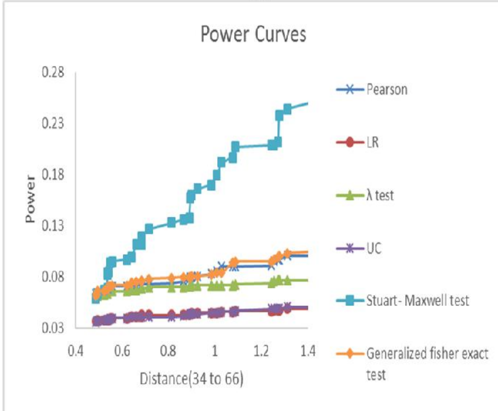


Fig.6.3.3

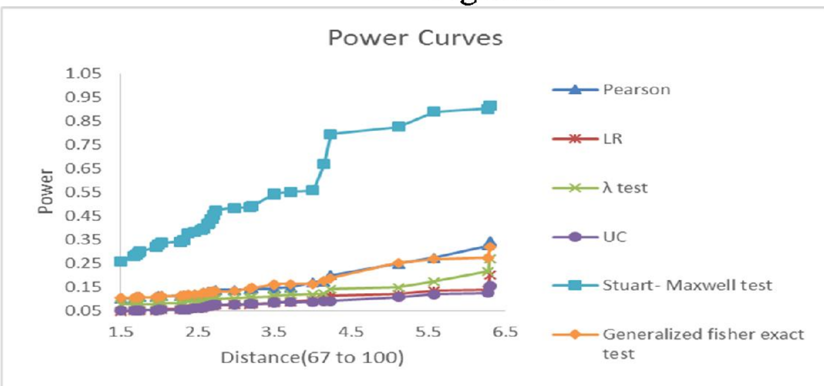


Fig. (6.3.1) represents that when sample size is 50, the power of all tests is less than significance level (α). The tests are biased for the chosen values of the parameter space or distance. Fig. (6.3.2) has been prepared from the same sample size except that some tests like Pearson Chi-square, generalized Fisher exact test and Stuart-Maxwell test have more power than significance level (α). These tests are unbiased for chosen values of the parameter space or distance. As a matter of fact, λ test, LR and UC still have less power than significance level (α). These measures of association always underestimate and biased the actual degree of association of nominal variables for the chosen values of the parameter space or distance. In overall, Stuart-Maxwell test has fairly large power than all tests. Fig. (6.3.3) represents that all tests have larger power than significance level (α) and are unbiased for the chosen values of the parameter space. In practice, Stuart-Maxwell test outperforms than all remaining tests for the chosen values of the parameter space.

Figure 6. 4: Power curves at sample size 50 and α = 1%:

Fig.6.4.1

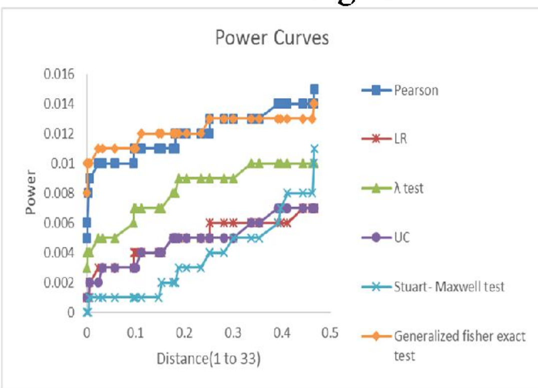


Fig.6.4.2

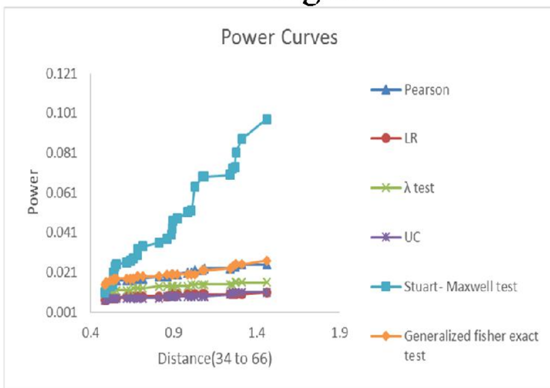


Fig.6.4.3

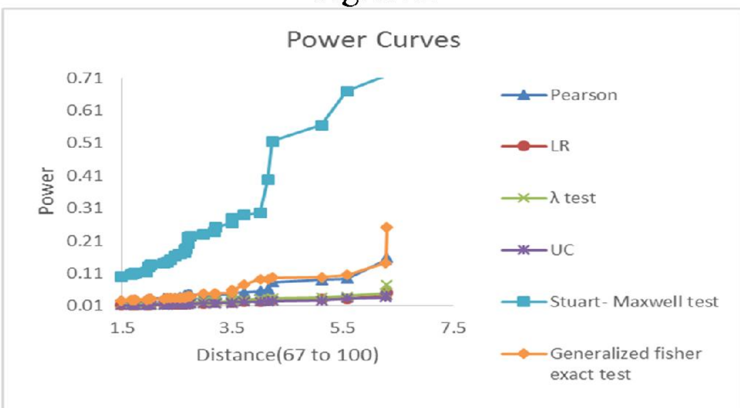


Fig. (6.4.1) shows that when sample size is 50, the power of all tests is less than significance level(α). The tests are biased for the chosen values of the parameter space or distance. Figs. (6.4.2 and (6.4.3) have been prepared from the same sample size although all tests have more power than significance level(α). All tests are unbiased for the chosen values of the parameter space or distance. We observe that the true power of all tests tends to increase as well as increase in parameter space or distance and Stuart-Maxwell test has higher power than all other tests of nominal association.

Figure 6. 5: Power curves at sample size 100 and $\alpha = 5\%$:

Fig.6. 5.1

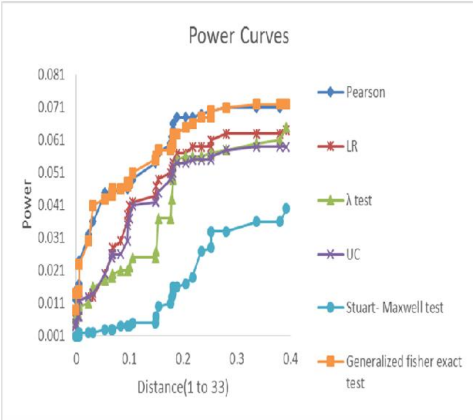


Fig.6.5.2

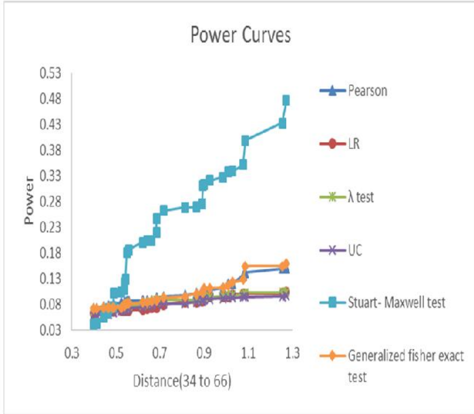


Fig.6.5.3

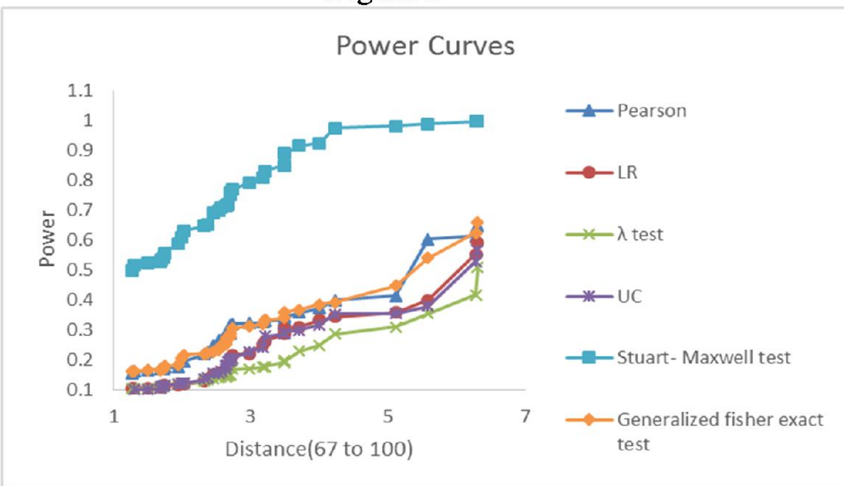


Fig. (6.5.1) shows that when sample size is 100, the power of all tests is less than significance level (α). The tests are biased for the chosen values of the parameter space or distance. Figs. (6.5.2) and (6.5.3) have been prepared from the same sample size although all tests have more power than significance level (α). All tests are unbiased for the chosen values of the parameter space or distance. We observe that the true power of all tests tends to increase as well as increase in parameter space or distance and Stuart-Maxwell test has higher power than all other tests. We have confirmed that at higher sample size, level of significance and parameter space or distance levels power of all tests have been increased.

Figure 6. 6 : Power curves at sample size 100 and $\alpha = 1\%$:

Fig.6.6.1

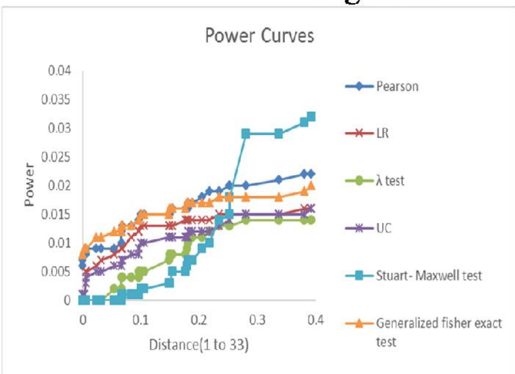


Fig.6.6.2

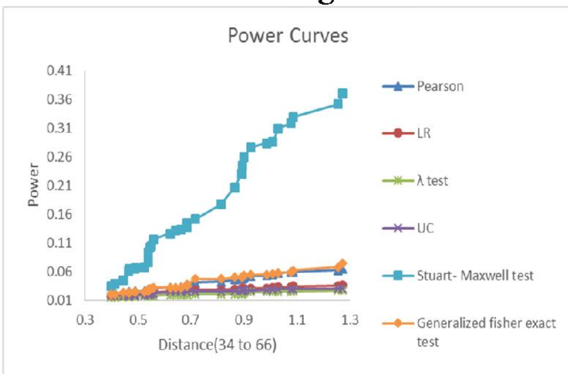


Fig.6.6.3

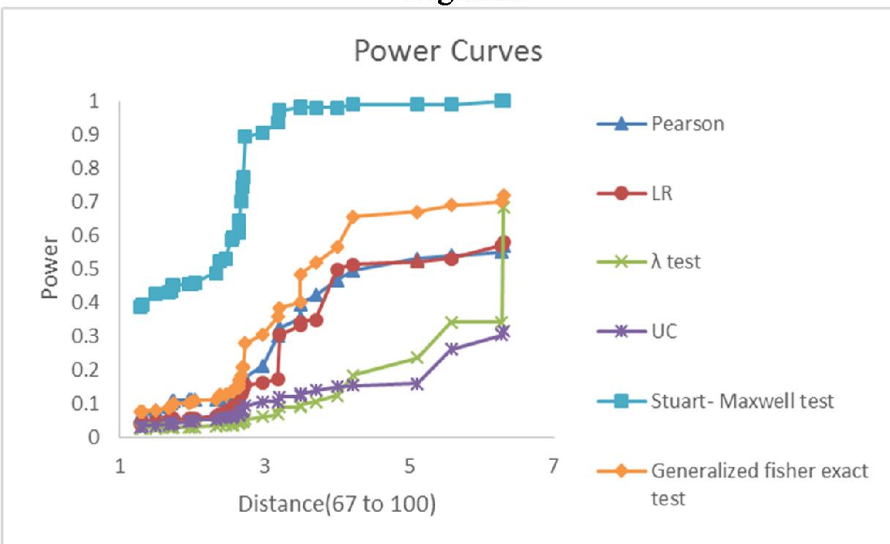


Fig. (6.6.1) shows that when sample size is 100, the power of all tests is less than significance level(α). These tests are biased for the chosen values of the parameter space or distance. Figs. (6.6.2) and (6.6.3) have been prepared from the same sample size although all tests have more power than significance level(α). All these tests are unbiased for the chosen values of the parameter space or distance. We observe that the true power of all tests tends to increase as well as increase in parameter space or distance and Stuart-Maxwell test has higher power than all other tests. We have confirmed that at smaller level of significance, power of all tests has been decreased.

6.3 Conclusion

All the above figures depict that when sample size, significance level(α), parameter space or level of distance increases power of all tests of independence also increases. Stuart-Maxwell test at a specific significance level α is most powerful against a specific alternative than all other tests of independence at same significance level. But this test is most powerful for some alternatives but less powerful for others. When a test at a specific significance level is more powerful against all alternatives than all other tests, we called this test uniformly most powerful. Comparing power curves for test of independence, we observe that for near alternatives with parameter space close to zero Pearson Chi-square test is the more powerful test. For far alternative with parameter space large Stuart-Maxwell test is the more powerful test. However, we may conclude that comparing tests of independence based on their power, Stuart-Maxwell test has substantially superior performance for larger alternatives than all other tests.

Chapter 7

Summary, Conclusions, Recommendations and Direction for Future Research

In this chapter, we stated the conclusion of our study and in the light of this conclusion recommendations are also given.

7.1 Summary

Pearson chi-square test of independence, Likelihood ratio chi-square test of independence, Goodman and Kruskal's lambda test, Uncertainty coefficient, Generalized McNemar's test (Stuart- Maxwell test), Generalized fisher exact test (Fisher freeman-Halton test) are used to test the association of unordered 3×3 contingency table. Null and alternative hypothesis are used to compute the test power and significance level for these test statistics in 3×3 contingency table. After extensive simulations, in general we recommend that for small values of parameter space which near to zero Pearson Chi-square test is more powerful and for large values of parameter space far zero Stuart- Maxwell test is more powerful.

7.2 Conclusion

Most of the test of hypothesis for nominal association become invalid and give misleading conclusions about the statistical significant of those ones. So, to overcome this problem we introduce some concept of simulated critical values in this study. After the extensive simulation in chapter 5, it was observed that the size of all six tests of independence is less or equal to the nominal size of 5% and 1%, as we have used simulated critical values. Moreover, to overcome the problem of size distortion for a test of independence of nominal association, we recommend that a researcher must use simulated critical values instead of asymptotic ones when sample size is small.

Based on those simulated critical values, we conclude that the space of null hypothesis is not a single criterion, rather it contains many points. Stuart-Maxwell test is more sensitive to the form of the degree of contingency table. This test creates huge distortion in size if we use asymptotic critical values. In the previous studies, we found that asymptotic tests are less reliable and their significance may fluctuate substantially. Hence, asymptotic tests should never be used for small samples $r \times c$ contingency tables.

From simulation results of chapter 6, we confirmed that for near alternatives with parameter space close to zero Pearson Chi-square test is the more powerful test. For far alternative with parameter space large Stuart-Maxwell test is the more powerful test. Power plots for all tests of independence represent that when sample size, significance level (α), parameter space or level of distance increases power of all tests of independence also increases.

7.3 Recommendations

The empirical study of Comparison of different measures of nominal correlation, a researcher must use the simulated critical values instead of asymptotic ones to overcome the problem of size distortion of test. From simulation results we confirmed that for near alternatives with parameter space close to zero Pearson Chi-square test is the more powerful test. For large alternative with parameter space far to zero, Stuart-Maxwell test is the more powerful test. So, a researcher/practioner should use simulated critical values for statistical significance because Stuart-Maxwell test is more sensitive to the form of degree of contingency table and this test is the more powerful for large alternatives.

7.4 Direction for Future Research

This study can be extended to the data categorized in multi-way contingency table. These six tests of independence for nominal variables can be compared based on asymptotic, exact conditional, or exact conditional with mid-p adjustment that are commonly used for computation the p-values. Comparison can be carried out based on marginal distributions, confidence intervals and standard errors of these tests.

References

1. Agbedeyi, O. D., & Igweze, A. H. Robust Method for Testing the Significance of Bivariate Correlation of Ordinal Data.

2. Agresti, A., & Kateri, M. (2011). *Categorical data analysis*: Springer.

3. Barnhart, H. X., Lokhnygina, Y., Kosinski, A. S., & Haber, M. (2007). Comparison of concordance correlation coefficient and coefficient of individual agreement in assessing agreement. *Journal of Biopharmaceutical Statistics, 17*(4), 721-738.

4. Cangur, S., & Ankarali, H. Comparison of Pearson Chi-Square and Log–Likelihood Ratio Statistics in RxC Tables with Regard to Type I Error.

5. Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data.* University of Pittsburgh.

6. Cornbleet, P. J., & Shea, M. C. (1978). Comparison of product moment and rank correlation coefficients in the assessment of laboratory method-comparison data. *Clinical chemistry, 24*(6), 857-861.

7. Ferguson, T. S., Genest, C., & Hallin, M. (2000). Kendall's tau for serial dependence. *Canadian Journal of Statistics, 28*(3), 587-604.

8. Galla, J. P. (1987). Kendall's tau and kendall's partial correlation: two BASIC programs for microcomputers. *Behavior Research Methods, 19*(1), 55-56.

9. Göktas, A., & Içi, Ö. (2011). A comparison of the most commonly used measures of association for doubly ordered square contingency tables via simulation. *Metodološki zvezki, 8*(1), 17-37.

10. Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae, 30*(2), 87-93.

11. Islam, A. U. (2012). Comparison of different measures of correlation for catagorical data. *Islamic International University Islamabad*, 60.

12. Lydersen, S., Fagerland, M. W., & Laake, P. (2009). Recommended tests for association in 2× 2 tables. *Statistics in medicine, 28*(7), 1159-1175.

13. Lydersen, S., & Laake, P. (2003). Power comparison of two□sided exact tests for association in 2× 2 contingency tables using standard, mid p and randomized test versions. *Statistics in medicine, 22*(24), 3859-3871.

14. Maturi, T., & Elsayigh, A. (2010). A comparison of correlation coefficients via a three-step bootstrap approach. *Journal of Mathematics Research, 2*(2), 3.

15. Newson, R. (2002). Parameters behind" nonparametric" statistics: Kendall's tau, Somers' D and median differences.

16. Olszak, M., & Ritschard, G. (1995). The behaviour of nominal and ordinal partial association measures. *The statistician*, 195-212.

17. Özdemir, T., & Eyduran, E. (2005). Comparison of Chi-Square and Likelihood Ratio Chi-Square Tests: Power of Test. *Journal of Applied Sciences Research, 1*(2), 242-244.

18. Yang, Z., Sun, X., & Hardin, J. W. (2011). Testing marginal homogeneity in clustered matched-pair data. *Journal of Statistical Planning and Inference, 141*(3), 1313-1318.

19. Zaman, A. (1996). *Statistical foundations for econometric techniques*: Academic Press.

Appendix –A

Power of Tests

		Power of tests at 5% critical values, Sample size 25					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart- Maxwell test	Generalized fisher exact test
D1.1	0	2.6%	0.1%	3.0%	0.1%	1.5%	0.0%
D1.2	0	2.8%	0.1%	3.2%	0.1%	1.5%	2.7%
D1.3	0.0034	4.5%	0.1%	3.6%	0.1%	1.6%	2.7%
D1.4	0.0055	4.5%	0.1%	3.6%	0.1%	1.7%	2.8%
D1.5	0.006	4.5%	0.2%	3.7%	0.2%	1.7%	2.9%
D1.6	0.0569	4.7%	0.2%	3.8%	0.2%	1.7%	2.9%
D1.7	0.1117	4.9%	0.4%	3.9%	0.2%	1.8%	3.0%
D1.8	0.1479	4.9%	0.4%	4.0%	0.2%	1.8%	3.0%
D1.9	0.1536	4.9%	0.4%	4.3%	0.2%	1.8%	3.1%
D1.10	0.1813	5.0%	0.4%	4.3%	0.3%	1.8%	3.1%
D2.1	0.1814	5.0%	0.5%	4.4%	0.3%	1.9%	3.1%
D2.2	0.1883	5.1%	0.5%	4.6%	0.4%	2.1%	3.1%
D2.3	0.2524	5.1%	0.5%	4.7%	0.4%	2.2%	3.2%
D2.4	0.3009	5.3%	0.6%	4.9%	0.4%	2.3%	3.3%
D2.5	0.3537	5.3%	0.6%	4.9%	0.4%	2.4%	3.3%
D2.6	0.3925	5.3%	0.6%	5.0%	0.5%	2.4%	3.4%
D2.7	0.3982	5.4%	0.6%	5.0%	0.5%	2.4%	3.4%
D2.8	0.4114	5.4%	0.6%	5.0%	0.5%	2.5%	3.6%
D2.9	0.4663	5.5%	0.7%	5.0%	0.5%	2.5%	3.7%
D2.10	0.4861	5.6%	0.7%	5.2%	0.5%	2.8%	3.7%
D3.1	0.4911	5.6%	0.7%	5.4%	0.6%	2.8%	3.7%
D3.2	0.4952	5.7%	0.7%	5.4%	0.6%	2.8%	3.8%
D3.3	0.5372	5.7%	0.8%	5.4%	0.6%	3.0%	3.8%
D3.4	0.5381	5.7%	0.8%	5.4%	0.6%	3.1%	3.9%
D3.5	0.5382	5.8%	0.8%	5.4%	0.6%	3.2%	3.9%
D3.6	0.5495	5.9%	0.9%	5.5%	0.6%	3.3%	4.0%
D3.7	0.5573	5.9%	0.9%	5.5%	0.6%	3.3%	4.0%
D3.8	0.6205	6.0%	0.9%	5.5%	0.6%	3.4%	4.1%
D3.9	0.6416	6.0%	0.9%	5.5%	0.6%	3.5%	4.2%
D3.10	0.6845	6.1%	0.9%	5.6%	0.6%	3.6%	4.2%
D4.1	0.6845	6.1%	1.0%	5.6%	0.7%	3.7%	4.2%
D4.2	0.716	6.1%	1.0%	5.6%	0.7%	3.9%	4.2%
D4.3	0.8137	6.2%	1.0%	5.6%	0.7%	4.0%	4.2%

		Power of tests at 5% critical values, Sample size 25					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D4.4	0.4667	6.2%	1.1%	5.6%	0.7%	4.6%	4.2%
D4.5	0.4861	6.2%	1.1%	5.6%	0.7%	4.6%	4.4%
D4.6	0.4911	6.3%	1.1%	5.7%	0.7%	5.2%	4.4%
D4.7	0.4952	6.3%	1.2%	5.7%	0.7%	5.2%	4.5%
D4.8	0.523	6.3%	1.2%	5.8%	0.8%	5.3%	4.5%
D4.9	0.5372	6.4%	1.2%	5.8%	0.8%	5.4%	4.5%
D4.10	0.5381	6.6%	1.2%	5.8%	0.8%	5.6%	4.6%
D5.1	0.5382	6.6%	1.2%	5.8%	0.8%	6.4%	4.6%
D5.2	0.5442	6.8%	1.2%	5.8%	0.8%	6.6%	4.8%
D5.3	0.5495	6.8%	1.2%	5.9%	0.9%	6.6%	4.8%
D5.4	0.5573	6.9%	1.2%	5.9%	0.9%	6.7%	4.9%
D5.5	0.6205	7.0%	1.3%	6.0%	0.9%	7.7%	5.0%
D5.6	0.6416	7.1%	1.3%	6.0%	0.9%	7.9%	5.1%
D5.7	0.6615	7.2%	1.4%	6.0%	0.9%	7.9%	5.2%
D5.8	0.6845	7.2%	1.4%	6.1%	1.0%	8.5%	5.2%
D5.9	0.6845	7.2%	1.4%	6.1%	1.0%	8.9%	5.2%
D5.10	0.716	7.3%	1.5%	6.2%	1.0%	9.0%	5.3%
D6.1	0.8137	7.6%	1.7%	6.2%	1.0%	9.0%	5.4%
D6.2	0.8648	7.6%	1.8%	6.2%	1.1%	9.0%	5.4%
D6.3	0.8908	7.7%	1.8%	6.2%	1.1%	9.1%	5.4%
D6.4	0.8944	7.9%	1.8%	6.3%	1.1%	9.4%	5.6%
D6.5	0.8996	8.0%	1.8%	6.3%	1.1%	9.6%	5.7%
D6.6	0.9267	8.2%	1.9%	6.3%	1.1%	10.5%	5.7%
D6.7	0.9847	8.3%	1.9%	6.3%	1.1%	10.6%	5.8%
D6.8	1.0093	8.4%	1.9%	6.3%	1.1%	11.0%	5.8%
D6.9	1.03	8.8%	1.9%	6.3%	1.1%	11.6%	5.8%
D6.10	1.0769	8.8%	2.0%	6.4%	1.1%	12.8%	5.9%
D7.1	1.0879	9.3%	2.0%	6.4%	1.2%	13.0%	5.9%
D7.2	1.2444	9.3%	2.1%	6.5%	1.2%	13.1%	5.9%
D7.3	1.2545	9.7%	2.1%	6.5%	1.2%	13.3%	6.1%
D7.4	1.2732	10.1%	2.1%	6.6%	1.2%	13.6%	6.3%
D7.5	1.2775	10.3%	2.2%	6.6%	1.2%	13.6%	6.4%
D7.6	1.3125	10.8%	2.3%	6.6%	1.3%	13.6%	6.7%

		Power of tests at 5% critical values, Sample size 25					
Design	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D7.7	1.4631	11.1%	2.3%	6.6%	1.3%	13.6%	6.8%
D7.8	1.5019	11.4%	2.4%	6.6%	1.3%	13.9%	6.8%
D7.9	1.68	11.5%	2.7%	6.6%	1.4%	14.5%	6.8%
D7.10	1.706	13.8%	2.9%	6.7%	1.4%	15.5%	7.0%
D8.1	1.7308	16.0%	2.9%	6.7%	1.4%	15.8%	7.1%
D8.2	1.7487	16.7%	3.8%	6.7%	1.5%	16.5%	7.2%
D8.3	1.9545	4.8%	0.3%	6.8%	1.5%	16.8%	7.2%
D8.4	1.9965	5.7%	0.8%	6.8%	1.6%	17.0%	7.3%
D8.5	2.0367	7.8%	1.8%	6.8%	1.6%	18.5%	7.7%
D8.6	2.2765	5.3%	0.6%	6.8%	1.6%	18.9%	7.8%
D8.7	2.3256	5.4%	0.6%	6.8%	1.6%	19.1%	7.9%
D8.8	2.3798	8.2%	1.8%	6.9%	1.7%	20.7%	7.9%
D8.9	2.4609	6.6%	1.2%	7.3%	1.7%	20.9%	8.1%
D8.10	2.5427	5.8%	0.9%	7.5%	1.7%	22.3%	8.3%
D9.1	2.5714	5.2%	0.6%	7.5%	1.7%	22.3%	8.5%
D9.2	2.6374	5.1%	0.5%	7.5%	1.7%	22.7%	8.7%
D9.3	2.6413	6.0%	0.9%	7.5%	1.7%	23.0%	8.7%
D9.4	2.6532	6.9%	1.2%	7.5%	1.8%	23.3%	8.9%
D9.5	2.6729	7.4%	1.7%	7.6%	1.8%	23.7%	10.1%
D9.6	2.6963	7.5%	1.7%	7.8%	1.8%	24.1%	10.1%
D9.7	2.703	11.3%	2.4%	7.9%	1.9%	24.4%	10.1%
D9.8	2.735	4.7%	0.2%	8.0%	1.9%	25.8%	10.6%
D9.9	2.9786	5.3%	0.6%	8.2%	2.0%	27.2%	11.1%
D9.10	3.1815	7.2%	1.3%	8.4%	2.0%	27.9%	11.9%
D10.1	3.2076	5.1%	0.5%	8.5%	2.0%	29.9%	12.7%
D10.2	3.4921	4.6%	0.2%	8.7%	2.1%	31.7%	13.6%
D10.3	3.7118	5.0%	0.4%	8.8%	2.1%	32.3%	13.7%
D10.4	4.0049	4.8%	0.3%	8.9%	2.1%	34.0%	14.1%
D10.5	4.1481	4.8%	0.2%	9.2%	2.1%	38.8%	15.7%
D10.6	4.2308	6.3%	1.1%	9.3%	2.5%	45.1%	20.5%
D10.7	5.1166	5.5%	0.7%	9.4%	2.6%	49.3%	23.4%
D10.8	5.5797	5.5%	0.7%	9.6%	2.6%	56.8%	30.6%
D10.9	6.2799	4.7%	0.2%	9.9%	2.6%	60.6%	32.9%
D10.10	6.303	11.7%	2.8%	13.5%	3.5%	61.0%	38.9%

		Power of tests at 1% critical values, Sample size 25				
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart- Maxwell test
D1.1	0	0.4%	0.0%	0.5%	0.0%	0.1%
D1.2	0	0.4%	0.0%	0.5%	0.0%	0.2%
D1.3	0.0034	0.5%	0.0%	0.5%	0.0%	0.2%
D1.4	0.0055	0.8%	0.0%	0.5%	0.0%	0.2%
D1.5	0.006	0.8%	0.0%	0.6%	0.0%	0.2%
D1.6	0.0236	0.8%	0.0%	0.6%	0.0%	0.2%
D1.7	0.0316	0.8%	0.0%	0.7%	0.0%	0.2%
D1.8	0.0569	0.9%	0.0%	0.7%	0.0%	0.3%
D1.9	0.0868	0.9%	0.0%	0.8%	0.0%	0.3%
D1.10	0.0963	0.9%	0.0%	0.8%	0.0%	0.3%
D2.1	0.0968	0.9%	0.0%	0.8%	0.0%	0.3%
D2.2	0.0998	0.9%	0.0%	0.8%	0.0%	0.3%
D2.3	0.1117	0.9%	0.1%	0.8%	0.0%	0.3%
D2.4	0.1479	0.9%	0.1%	0.8%	0.1%	0.3%
D2.5	0.1536	0.9%	0.1%	0.8%	0.1%	0.3%
D2.6	0.1768	1.0%	0.1%	0.8%	0.1%	0.4%
D2.7	0.1813	1.0%	0.1%	0.8%	0.1%	0.4%
D2.8	0.1814	1.0%	0.1%	0.8%	0.1%	0.4%
D2.9	0.1883	1.0%	0.1%	0.8%	0.1%	0.4%
D2.10	0.2052	1.0%	0.1%	0.8%	0.1%	0.4%
D3.1	0.234	1.0%	0.1%	0.8%	0.1%	0.4%
D3.2	0.2524	1.0%	0.1%	0.8%	0.1%	0.4%
D3.3	0.2524	1.0%	0.1%	0.9%	0.1%	0.5%
D3.4	0.2808	1.0%	0.1%	0.9%	0.1%	0.5%
D3.5	0.3009	1.0%	0.1%	0.9%	0.1%	0.5%
D3.6	0.3373	1.0%	0.1%	0.9%	0.1%	0.5%
D3.7	0.3537	1.0%	0.1%	0.9%	0.1%	0.5%
D3.8	0.3925	1.0%	0.1%	0.9%	0.1%	0.5%
D3.9	0.3982	1.0%	0.1%	0.9%	0.1%	0.6%
D3.10	0.4114	1.1%	0.1%	0.9%	0.1%	0.6%
D4.1	0.4444	1.1%	0.2%	0.9%	0.1%	0.7%
D4.2	0.4635	1.1%	0.2%	0.9%	0.1%	0.8%
D4.3	0.4663	1.1%	0.2%	0.9%	0.1%	0.8%

		Power of tests at 1% critical values, Sample size 25				
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test
D4.4	0.4667	1.1%	0.2%	0.9%	0.1%	0.9%
D4.5	0.4861	1.1%	0.2%	0.9%	0.1%	0.9%
D4.6	0.4911	1.1%	0.2%	1.0%	0.1%	0.9%
D4.7	0.4952	1.2%	0.2%	1.0%	0.1%	0.9%
D4.8	0.523	1.2%	0.2%	1.0%	0.1%	1.0%
D4.9	0.5372	1.2%	0.2%	1.0%	0.2%	1.1%
D4.10	0.5381	1.2%	0.2%	1.0%	0.2%	1.2%
D5.1	0.5382	1.2%	0.2%	1.0%	0.2%	1.2%
D5.2	0.5442	1.2%	0.2%	1.0%	0.2%	1.3%
D5.3	0.5495	1.2%	0.2%	1.0%	0.2%	1.4%
D5.4	0.5573	1.2%	0.2%	1.0%	0.2%	1.4%
D5.5	0.6205	1.2%	0.2%	1.0%	0.2%	1.7%
D5.6	0.6416	1.2%	0.2%	1.0%	0.2%	1.9%
D5.7	0.6615	1.3%	0.2%	1.1%	0.2%	2.0%
D5.8	0.6845	1.3%	0.2%	1.1%	0.2%	2.0%
D5.9	0.6845	1.3%	0.2%	1.1%	0.2%	2.1%
D5.10	0.716	1.3%	0.2%	1.1%	0.2%	2.1%
D6.1	0.8137	1.4%	0.2%	1.1%	0.2%	2.1%
D6.2	0.8648	1.4%	0.2%	1.1%	0.2%	2.2%
D6.3	0.8908	1.4%	0.2%	1.1%	0.2%	2.2%
D6.4	0.8944	1.4%	0.2%	1.1%	0.2%	2.3%
D6.5	0.8996	1.4%	0.2%	1.1%	0.2%	2.3%
D6.6	0.9267	1.4%	0.2%	1.1%	0.2%	2.6%
D6.7	0.9847	1.4%	0.3%	1.2%	0.2%	2.7%
D6.8	1.0093	1.5%	0.3%	1.2%	0.2%	2.9%
D6.9	1.03	1.5%	0.3%	1.2%	0.2%	3.1%
D6.10	1.0769	1.5%	0.3%	1.2%	0.2%	3.6%
D7.1	1.0879	1.5%	0.3%	1.2%	0.2%	3.8%
D7.2	1.2444	1.5%	0.3%	1.2%	0.2%	3.9%
D7.3	1.2545	1.6%	0.3%	1.2%	0.2%	3.9%
D7.4	1.2732	1.6%	0.3%	1.3%	0.2%	4.0%
D7.5	1.2775	1.6%	0.3%	1.3%	0.2%	4.0%
D7.6	1.3125	1.7%	0.3%	1.3%	0.3%	4.1%

		Power of tests at 1% critical values, Sample size 25				
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test
D7.7	1.4631	1.7%	0.3%	1.3%	0.3%	4.1%
D7.8	1.5019	1.7%	0.3%	1.3%	0.3%	4.1%
D7.9	1.68	1.8%	0.3%	1.4%	0.3%	4.1%
D7.10	1.706	1.9%	0.3%	1.4%	0.3%	4.2%
D8.1	1.7308	1.9%	0.4%	1.4%	0.3%	4.6%
D8.2	1.7487	1.9%	0.4%	1.4%	0.3%	4.6%
D8.3	1.9545	1.9%	0.4%	1.4%	0.3%	4.7%
D8.4	1.9965	1.9%	0.4%	1.4%	0.3%	5.0%
D8.5	2.0367	1.9%	0.4%	1.5%	0.3%	5.1%
D8.6	2.2765	2.0%	0.4%	1.5%	0.3%	6.3%
D8.7	2.3256	2.0%	0.4%	1.5%	0.3%	6.5%
D8.8	2.3798	2.0%	0.4%	1.5%	0.3%	6.5%
D8.9	2.4609	2.0%	0.4%	1.6%	0.3%	6.5%
D8.10	2.5427	2.0%	0.4%	1.6%	0.3%	7.0%
D9.1	2.5714	2.0%	0.4%	1.6%	0.4%	7.4%
D9.2	2.6374	2.1%	0.4%	1.6%	0.4%	7.8%
D9.3	2.6413	2.1%	0.5%	1.7%	0.4%	7.9%
D9.4	2.6532	2.3%	0.5%	1.7%	0.4%	7.9%
D9.5	2.6729	2.3%	0.5%	1.7%	0.4%	8.2%
D9.6	2.6963	2.3%	0.5%	1.7%	0.4%	8.2%
D9.7	2.703	2.5%	0.5%	1.7%	0.4%	8.7%
D9.8	2.735	2.6%	0.5%	1.7%	0.4%	8.8%
D9.9	2.9786	2.7%	0.5%	1.8%	0.4%	8.9%
D9.10	3.1815	3.0%	0.5%	1.8%	0.4%	9.3%
D10.1	3.2076	3.0%	0.5%	1.8%	0.4%	10.7%
D10.2	3.4921	3.0%	0.5%	1.8%	0.4%	10.9%
D10.3	3.7118	3.1%	0.6%	1.8%	0.5%	11.7%
D10.4	4.0049	3.2%	0.6%	1.9%	0.5%	12.8%
D10.5	4.1481	3.7%	0.6%	1.9%	0.5%	17.0%
D10.6	4.2308	4.4%	0.6%	2.0%	0.5%	20.1%
D10.7	5.1166	4.5%	0.7%	2.2%	0.5%	25.3%
D10.8	5.5797	4.6%	0.7%	2.2%	0.5%	29.2%
D10.9	6.2799	4.8%	0.8%	2.2%	0.6%	31.0%
D10.10	6.303	5.4%	60.1%	2.2%	0.7%	31.0%

		Power of tests at 5% critical values, Sample size 50					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D1.1	0	3.1%	0.2%	1.6%	0.2%	0.5%	0.7%
D1.2	0	3.7%	0.4%	1.7%	0.4%	0.5%	4.4%
D1.3	0.0034	4.5%	0.5%	1.8%	0.4%	0.6%	4.7%
D1.4	0.0055	4.7%	0.9%	1.8%	0.9%	0.6%	4.7%
D1.5	0.006	4.7%	1.1%	2.0%	1.0%	0.7%	4.8%
D1.6	0.0236	4.7%	1.1%	2.9%	1.1%	0.7%	5.0%
D1.7	0.0316	4.9%	1.3%	3.3%	1.6%	0.7%	5.0%
D1.8	0.0569	4.9%	1.7%	3.3%	1.7%	0.7%	5.1%
D1.9	0.0963	5.0%	1.7%	4.2%	1.8%	0.8%	5.2%
D1.10	0.0968	5.1%	1.8%	4.2%	1.8%	0.8%	5.3%
D2.1	0.0998	5.2%	1.8%	4.4%	1.8%	1.0%	5.3%
D2.2	0.1117	5.3%	1.9%	4.5%	1.9%	1.1%	5.6%
D2.3	0.1479	5.5%	2.0%	4.5%	2.0%	1.1%	5.6%
D2.4	0.1536	5.6%	2.0%	4.6%	2.2%	1.1%	5.6%
D2.5	0.1768	5.7%	2.1%	4.6%	2.3%	1.3%	5.6%
D2.6	0.1813	5.7%	2.1%	5.0%	2.4%	1.5%	5.7%
D2.7	0.1814	5.8%	2.1%	5.1%	2.6%	1.5%	5.7%
D2.8	0.1883	5.8%	2.2%	5.1%	2.6%	1.6%	5.7%
D2.9	0.2052	5.8%	2.4%	5.3%	2.7%	1.9%	5.7%
D2.10	0.234	5.8%	2.6%	5.4%	2.7%	2.0%	5.7%
D3.1	0.2524	5.9%	2.7%	5.5%	2.8%	2.2%	5.8%
D3.2	0.2524	5.9%	2.8%	5.6%	2.8%	2.5%	5.8%
D3.3	0.2808	5.9%	2.8%	5.6%	2.9%	2.6%	5.8%
D3.4	0.3009	6.0%	2.8%	5.6%	3.0%	2.8%	5.8%
D3.5	0.3373	6.0%	2.9%	5.6%	3.1%	2.9%	5.9%
D3.6	0.3537	6.1%	3.0%	5.7%	3.2%	3.1%	5.9%
D3.7	0.3925	6.1%	3.0%	5.8%	3.2%	3.3%	6.0%
D3.8	0.3982	6.1%	3.1%	5.9%	3.2%	3.7%	6.0%
D3.9	0.4114	6.1%	3.2%	5.9%	3.3%	3.9%	6.0%
D3.10	0.4444	6.1%	3.2%	6.0%	3.3%	4.2%	6.1%
D4.1	0.4635	6.2%	3.3%	6.0%	3.5%	4.2%	6.1%
D4.2	0.4663	6.2%	3.5%	6.0%	3.6%	4.7%	6.1%
D4.3	0.4667	6.3%	3.5%	6.1%	3.6%	5.5%	6.1%

		Power of tests at 5% critical values, Sample size 50					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D4.4	0.4861	6.4%	3.7%	6.1%	3.6%	5.9%	6.2%
D4.5	0.4911	6.5%	3.7%	6.2%	3.7%	6.3%	6.3%
D4.6	0.4952	6.5%	3.8%	6.2%	3.7%	6.4%	6.6%
D4.7	0.523	6.7%	3.8%	6.3%	3.8%	6.7%	6.6%
D4.8	0.5372	6.8%	3.8%	6.4%	3.8%	6.9%	6.8%
D4.9	0.5381	6.8%	3.8%	6.6%	3.8%	8.2%	6.8%
D4.10	0.5382	6.9%	3.9%	6.6%	3.9%	8.4%	7.0%
D5.1	0.5442	6.9%	3.9%	6.6%	3.9%	8.5%	7.1%
D5.2	0.5495	7.1%	4.0%	6.6%	4.0%	9.4%	7.1%
D5.3	0.5573	7.1%	4.0%	6.6%	4.0%	9.5%	7.2%
D5.4	0.6205	7.1%	4.0%	6.6%	4.0%	9.7%	7.2%
D5.5	0.6416	7.1%	4.1%	6.7%	4.1%	9.9%	7.4%
D5.6	0.6615	7.2%	4.1%	6.7%	4.1%	11.2%	7.5%
D5.7	0.6845	7.2%	4.1%	6.9%	4.1%	11.2%	7.6%
D5.8	0.6845	7.3%	4.3%	6.9%	4.1%	11.9%	7.6%
D5.9	0.716	7.3%	4.3%	7.0%	4.1%	12.7%	7.8%
D5.10	0.8137	7.4%	4.3%	7.0%	4.1%	13.3%	7.9%
D6.1	0.8648	7.6%	4.3%	7.0%	4.2%	13.6%	8.0%
D6.2	0.8908	7.8%	4.3%	7.1%	4.3%	13.8%	8.0%
D6.3	0.8944	8.0%	4.4%	7.1%	4.4%	15.7%	8.0%
D6.4	0.8996	8.0%	4.5%	7.2%	4.4%	16.0%	8.1%
D6.5	0.9267	8.1%	4.5%	7.2%	4.4%	16.6%	8.1%
D6.6	0.9847	8.3%	4.5%	7.2%	4.5%	17.0%	8.2%
D6.7	1.0093	8.6%	4.5%	7.2%	4.5%	18.0%	8.4%
D6.8	1.03	9.0%	4.6%	7.2%	4.6%	19.2%	8.4%
D6.9	1.0769	9.0%	4.6%	7.2%	4.6%	19.7%	9.4%
D6.10	1.0879	9.0%	4.7%	7.3%	4.7%	20.7%	9.5%
D7.1	1.2444	9.1%	4.7%	7.4%	4.9%	20.9%	9.5%
D7.2	1.2545	9.6%	4.8%	7.6%	4.9%	20.9%	9.7%
D7.3	1.2732	9.7%	4.8%	7.7%	4.9%	21.2%	9.9%
D7.4	1.2775	9.7%	4.8%	7.7%	4.9%	23.8%	10.0%
D7.5	1.3125	10.1%	4.9%	7.7%	5.1%	24.4%	10.3%
D7.6	1.4631	10.1%	4.9%	7.7%	5.1%	25.3%	10.5%

		Power of tests at 5% critical values, Sample size 50					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D7.7	1.5019	10.3%	5.0%	7.8%	5.1%	26.1%	10.6%
D7.8	1.68	10.4%	5.1%	7.8%	5.1%	28.1%	10.7%
D7.9	1.706	10.8%	5.1%	7.8%	5.3%	28.7%	10.8%
D7.10	1.7308	10.8%	5.2%	8.0%	5.4%	29.7%	10.8%
D8.1	1.7487	10.9%	5.3%	8.1%	5.4%	30.2%	10.8%
D8.2	1.9545	10.9%	5.4%	8.1%	5.4%	32.0%	10.8%
D8.3	1.9965	11.2%	5.5%	8.2%	5.5%	33.0%	11.1%
D8.4	2.0367	11.3%	5.8%	8.4%	5.5%	33.8%	11.2%
D8.5	2.2765	11.4%	5.8%	8.4%	5.5%	34.2%	11.4%
D8.6	2.3256	11.6%	5.9%	8.6%	5.5%	35.0%	11.7%
D8.7	2.3798	11.8%	6.1%	9.2%	5.6%	37.9%	12.0%
D8.8	2.4609	12.1%	6.2%	9.2%	6.4%	38.6%	12.0%
D8.9	2.5427	12.1%	6.2%	9.4%	6.5%	39.0%	12.4%
D8.10	2.5714	12.7%	6.6%	9.4%	6.8%	39.9%	12.6%
D9.1	2.6374	13.0%	6.9%	9.4%	6.9%	41.6%	12.6%
D9.2	2.6413	13.0%	7.3%	9.6%	7.0%	41.8%	12.9%
D9.3	2.6532	13.1%	7.3%	9.8%	7.0%	42.1%	13.1%
D9.4	2.6729	13.2%	7.3%	9.8%	7.5%	43.7%	13.2%
D9.5	2.6963	13.2%	7.4%	9.9%	7.5%	43.8%	13.2%
D9.6	2.703	13.4%	7.5%	10.0%	7.6%	45.6%	13.3%
D9.7	2.735	13.9%	7.6%	10.1%	7.7%	47.4%	13.4%
D9.8	2.9786	14.0%	7.6%	10.3%	7.8%	48.5%	13.4%
D9.9	3.1815	14.1%	7.9%	10.7%	8.0%	48.7%	14.6%
D9.10	3.2076	14.4%	8.0%	10.8%	8.1%	49.1%	14.8%
D10.1	3.4921	14.5%	8.4%	11.0%	8.4%	54.2%	16.1%
D10.2	3.4921	14.6%	8.8%	11.6%	8.7%	54.5%	16.4%
D10.3	3.7118	15.0%	9.0%	12.0%	8.8%	55.4%	16.4%
D10.4	4.0049	17.2%	9.6%	12.1%	8.9%	55.8%	16.5%
D10.5	4.1481	17.5%	10.4%	12.4%	9.2%	66.9%	17.7%
D10.6	4.2308	20.1%	11.4%	14.2%	9.2%	79.6%	18.9%
D10.7	5.1166	24.9%	12.1%	15.0%	10.8%	82.7%	25.4%
D10.8	5.5797	27.5%	13.6%	17.5%	12.2%	89.1%	26.9%
D10.9	6.2799	32.7%	13.9%	21.7%	12.7%	90.3%	27.5%
D10.10	6.303	34.7%	20.1%	27.0%	15.6%	91.5%	32.0%

		Power of tests at 1% critical values, Sample size 50					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D1.1	0	0.5%	0.1%	0.3%	0.1%	0.0%	0.8%
D1.2	0	0.6%	0.1%	0.4%	0.1%	0.0%	1.0%
D1.3	0.0034	0.8%	0.1%	0.4%	0.1%	0.0%	1.0%
D1.4	0.0055	0.9%	0.1%	0.4%	0.2%	0.1%	1.0%
D1.5	0.006	0.9%	0.2%	0.4%	0.2%	0.1%	1.0%
D1.6	0.0236	1.0%	0.3%	0.5%	0.2%	0.1%	1.1%
D1.7	0.0316	1.0%	0.3%	0.5%	0.3%	0.1%	1.1%
D1.8	0.0569	1.0%	0.3%	0.5%	0.3%	0.1%	1.1%
D1.9	0.0963	1.0%	0.3%	0.6%	0.3%	0.1%	1.1%
D1.10	0.0968	1.1%	0.4%	0.7%	0.3%	0.1%	1.1%
D2.1	0.0998	1.1%	0.4%	0.7%	0.3%	0.1%	1.1%
D2.2	0.1117	1.1%	0.4%	0.7%	0.4%	0.1%	1.2%
D2.3	0.1479	1.1%	0.4%	0.7%	0.4%	0.1%	1.2%
D2.4	0.1536	1.1%	0.4%	0.7%	0.4%	0.2%	1.2%
D2.5	0.1768	1.1%	0.5%	0.8%	0.5%	0.2%	1.2%
D2.6	0.1813	1.1%	0.5%	0.8%	0.5%	0.2%	1.2%
D2.7	0.1814	1.2%	0.5%	0.8%	0.5%	0.2%	1.2%
D2.8	0.1883	1.2%	0.5%	0.9%	0.5%	0.3%	1.2%
D2.9	0.2052	1.2%	0.5%	0.9%	0.5%	0.3%	1.2%
D2.10	0.234	1.2%	0.5%	0.9%	0.5%	0.3%	1.2%
D3.1	0.2524	1.2%	0.5%	0.9%	0.5%	0.4%	1.3%
D3.2	0.2524	1.3%	0.6%	0.9%	0.5%	0.4%	1.3%
D3.3	0.2808	1.3%	0.6%	0.9%	0.5%	0.4%	1.3%
D3.4	0.3009	1.3%	0.6%	0.9%	0.5%	0.5%	1.3%
D3.5	0.3373	1.3%	0.6%	1.0%	0.6%	0.5%	1.3%
D3.6	0.3537	1.3%	0.6%	1.0%	0.6%	0.5%	1.3%
D3.7	0.3925	1.4%	0.6%	1.0%	0.7%	0.6%	1.3%
D3.8	0.3982	1.4%	0.6%	1.0%	0.7%	0.7%	1.3%
D3.9	0.4114	1.4%	0.6%	1.0%	0.7%	0.8%	1.3%
D3.10	0.4444	1.4%	0.7%	1.0%	0.7%	0.8%	1.3%
D4.1	0.4635	1.4%	0.7%	1.0%	0.7%	0.8%	1.3%
D4.2	0.4663	1.4%	0.7%	1.0%	0.7%	1.0%	1.4%
D4.3	0.4667	1.5%	0.7%	1.0%	0.7%	1.1%	1.4%

Designs	Distance	Power of tests at 1% critical values, Sample size 50 Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D4.4	0.4861	1.5%	0.7%	1.0%	0.7%	1.1%	1.5%
D4.5	0.4911	1.5%	0.7%	1.1%	0.7%	1.3%	1.5%
D4.6	0.4952	1.6%	0.7%	1.1%	0.8%	1.3%	1.6%
D4.7	0.523	1.6%	0.8%	1.1%	0.8%	1.4%	1.7%
D4.8	0.5372	1.6%	0.8%	1.1%	0.8%	1.5%	1.7%
D4.9	0.5381	1.7%	0.8%	1.1%	0.8%	2.1%	1.7%
D4.10	0.5382	1.7%	0.8%	1.1%	0.8%	2.1%	1.7%
D5.1	0.5442	1.7%	0.8%	1.1%	0.8%	2.1%	1.8%
D5.2	0.5495	1.7%	0.8%	1.1%	0.8%	2.4%	1.8%
D5.3	0.5573	1.7%	0.8%	1.2%	0.8%	2.5%	1.8%
D5.4	0.6205	1.7%	0.9%	1.2%	0.8%	2.6%	1.8%
D5.5	0.6416	1.7%	0.9%	1.2%	0.8%	2.7%	1.8%
D5.6	0.6615	1.8%	0.9%	1.3%	0.8%	2.8%	1.8%
D5.7	0.6845	1.8%	0.9%	1.3%	0.8%	3.0%	1.9%
D5.8	0.6845	1.8%	0.9%	1.3%	0.8%	3.3%	1.9%
D5.9	0.716	1.8%	0.9%	1.3%	0.8%	3.4%	1.9%
D5.10	0.8137	1.9%	0.9%	1.4%	0.8%	3.6%	1.9%
D6.1	0.8648	1.9%	0.9%	1.4%	0.9%	3.8%	2.0%
D6.2	0.8908	2.0%	1.0%	1.4%	0.9%	4.0%	2.0%
D6.3	0.8944	2.0%	1.0%	1.4%	0.9%	4.3%	2.0%
D6.4	0.8996	2.0%	1.0%	1.4%	0.9%	4.7%	2.0%
D6.5	0.9267	2.0%	1.0%	1.4%	0.9%	4.8%	2.0%
D6.6	0.9847	2.1%	1.0%	1.4%	0.9%	5.1%	2.0%
D6.7	1.0093	2.1%	1.0%	1.5%	0.9%	5.2%	2.0%
D6.8	1.03	2.2%	1.0%	1.5%	0.9%	6.4%	2.0%
D6.9	1.0769	2.2%	1.0%	1.5%	0.9%	6.9%	2.2%
D6.10	1.0879	2.3%	1.0%	1.5%	0.9%	6.9%	2.2%
D7.1	1.2444	2.3%	1.0%	1.5%	1.0%	7.0%	2.3%
D7.2	1.2545	2.4%	1.0%	1.5%	1.1%	7.3%	2.4%
D7.3	1.2732	2.4%	1.0%	1.6%	1.1%	7.4%	2.5%
D7.4	1.2775	2.5%	1.0%	1.6%	1.1%	8.1%	2.5%
D7.5	1.3125	2.5%	1.0%	1.6%	1.1%	8.8%	2.5%
D7.6	1.4631	2.5%	1.1%	1.6%	1.1%	9.8%	2.7%

		Power of tests at 1% critical values, Sample size 50					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D7.7	1.5019	2.6%	1.1%	1.6%	1.1%	10.0%	2.7%
D7.8	1.68	2.6%	1.1%	1.6%	1.1%	10.5%	2.8%
D7.9	1.706	2.7%	1.1%	1.6%	1.1%	10.8%	2.8%
D7.10	1.7308	2.7%	1.2%	1.6%	1.2%	11.0%	2.9%
D8.1	1.7487	2.9%	1.2%	1.6%	1.2%	11.2%	2.9%
D8.2	1.9545	2.9%	1.3%	1.6%	1.2%	11.5%	3.0%
D8.3	1.9965	3.1%	1.3%	1.6%	1.2%	13.0%	3.1%
D8.4	2.0367	3.2%	1.4%	1.6%	1.3%	13.6%	3.1%
D8.5	2.2765	3.3%	1.4%	1.6%	1.3%	14.2%	3.2%
D8.6	2.3256	3.5%	1.4%	1.7%	1.3%	14.4%	3.3%
D8.7	2.3798	3.5%	1.4%	1.7%	1.4%	15.2%	3.3%
D8.8	2.4609	3.5%	1.5%	1.8%	1.4%	16.3%	3.3%
D8.9	2.5427	3.6%	1.5%	1.8%	1.5%	16.8%	3.3%
D8.10	2.5714	3.8%	1.5%	1.8%	1.5%	17.0%	3.4%
D9.1	2.6374	4.0%	1.6%	1.9%	1.6%	17.3%	3.5%
D9.2	2.6413	4.2%	1.6%	1.9%	1.6%	17.6%	3.7%
D9.3	2.6532	4.3%	1.7%	2.0%	1.6%	18.3%	3.8%
D9.4	2.6729	4.3%	1.8%	2.1%	1.6%	18.6%	3.8%
D9.5	2.6963	4.4%	1.9%	2.1%	1.6%	20.3%	3.9%
D9.6	2.703	4.4%	1.9%	2.1%	1.7%	22.0%	4.0%
D9.7	2.735	4.5%	1.9%	2.2%	1.7%	22.5%	4.0%
D9.8	2.9786	4.5%	1.9%	2.2%	1.8%	22.9%	4.7%
D9.9	3.1815	4.5%	2.0%	2.3%	1.8%	23.8%	4.8%
D9.10	3.2076	4.5%	2.0%	2.3%	1.9%	25.0%	4.9%
D10.1	3.4921	4.6%	2.1%	2.5%	1.9%	26.5%	5.0%
D10.2	3.4921	4.8%	2.1%	2.9%	2.1%	27.8%	5.8%
D10.3	3.7118	5.1%	2.3%	3.0%	2.2%	28.9%	7.5%
D10.4	4.0049	5.6%	2.3%	3.1%	2.2%	29.6%	9.2%
D10.5	4.1481	6.1%	2.5%	3.2%	2.4%	39.6%	9.2%
D10.6	4.2308	8.3%	3.0%	3.3%	2.5%	51.6%	9.7%
D10.7	5.1166	9.0%	3.1%	3.6%	2.7%	56.5%	9.8%
D10.8	5.5797	9.4%	3.2%	4.1%	3.3%	66.9%	10.5%
D10.9	6.2799	15.1%	4.2%	5.0%	3.6%	71.6%	14.2%
D10.10	6.303	15.7%	5.2%	7.5%	4.3%	71.8%	25.0%

		Power of tests at 5% critical values, Sample size 100					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D1.1	0	1.2%	0.4%	0.6%	0.4%	0.1%	0.9%
D1.2	0	1.5%	0.6%	0.7%	0.5%	0.1%	1.4%
D1.3	0.0034	1.5%	0.9%	0.8%	0.7%	0.1%	1.4%
D1.4	0.0055	1.7%	0.9%	1.0%	0.7%	0.1%	1.5%
D1.5	0.006	2.4%	1.2%	1.0%	1.2%	0.2%	2.3%
D1.6	0.0236	3.2%	1.3%	1.1%	1.3%	0.2%	3.0%
D1.7	0.0316	3.6%	1.3%	1.6%	1.4%	0.2%	4.1%
D1.8	0.05435	4.5%	2.0%	1.8%	2.0%	0.3%	4.3%
D1.9	0.0668	4.5%	2.6%	1.9%	2.5%	0.3%	4.4%
D1.10	0.0668	4.5%	2.6%	1.9%	2.6%	0.3%	4.5%
D2.1	0.0684	4.5%	2.8%	2.0%	2.6%	0.3%	4.6%
D2.2	0.0831	4.6%	3.0%	2.1%	2.6%	0.4%	4.6%
D2.3	0.0963	4.6%	3.6%	2.1%	3.0%	0.4%	4.7%
D2.4	0.0968	4.7%	3.8%	2.1%	3.5%	0.4%	4.8%
D2.5	0.0998	4.7%	4.1%	2.2%	3.7%	0.4%	4.8%
D2.6	0.10565	4.9%	4.2%	2.5%	4.1%	0.5%	5.1%
D2.7	0.1479	5.4%	4.4%	2.5%	4.2%	0.5%	5.5%
D2.8	0.14995	5.6%	4.7%	2.7%	4.2%	0.7%	5.6%
D2.9	0.1536	5.7%	4.9%	3.7%	4.5%	1.0%	5.8%
D2.10	0.1768	6.0%	5.1%	3.7%	4.9%	1.1%	5.8%
D3.1	0.1787	6.2%	5.2%	4.3%	5.0%	1.3%	5.9%
D3.2	0.1813	6.4%	5.3%	4.9%	5.1%	1.4%	6.0%
D3.3	0.1814	6.6%	5.4%	4.9%	5.3%	1.6%	6.3%
D3.4	0.1883	6.8%	5.7%	5.5%	5.4%	1.6%	6.3%
D3.5	0.2052	6.8%	5.7%	5.6%	5.4%	1.7%	6.5%
D3.6	0.21765	6.8%	5.9%	5.6%	5.5%	1.9%	6.6%
D3.7	0.234	6.9%	5.9%	5.6%	5.5%	2.7%	6.8%
D3.8	0.2524	7.0%	5.9%	5.7%	5.5%	2.8%	6.8%
D3.9	0.2524	7.0%	6.1%	5.7%	5.6%	3.3%	7.0%
D3.10	0.2808	7.1%	6.3%	5.8%	5.8%	3.3%	7.1%
D4.1	0.3373	7.1%	6.3%	6.0%	5.9%	3.6%	7.2%
D4.2	0.38065	7.1%	6.3%	6.1%	5.9%	3.6%	7.2%
D4.3	0.3925	7.2%	6.4%	6.5%	5.9%	4.0%	7.2%

Designs	Distance	Power of tests at 5% critical values, Sample size 100					
		Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D4.4	0.3982	7.2%	6.4%	6.7%	5.9%	4.2%	7.2%
D4.5	0.4114	7.3%	6.5%	6.9%	6.0%	4.4%	7.3%
D4.6	0.4444	7.6%	6.6%	7.0%	6.0%	5.7%	7.4%
D4.7	0.4635	7.6%	6.6%	7.1%	6.1%	6.3%	7.4%
D4.8	0.4663	7.8%	6.6%	7.1%	6.1%	7.3%	7.4%
D4.9	0.4861	7.9%	6.7%	7.2%	6.5%	7.6%	7.4%
D4.10	0.4911	8.2%	6.8%	7.3%	6.6%	10.2%	7.5%
D5.1	0.523	8.4%	6.8%	7.3%	6.8%	10.5%	7.5%
D5.2	0.5372	8.4%	6.9%	7.4%	6.9%	11.0%	7.8%
D5.3	0.5381	8.4%	6.9%	7.5%	6.9%	11.3%	7.8%
D5.4	0.5382	8.5%	6.9%	7.5%	7.0%	12.3%	8.1%
D5.5	0.5442	8.5%	6.9%	7.5%	7.0%	13.0%	8.2%
D5.6	0.5495	8.6%	6.9%	7.7%	7.4%	18.1%	8.3%
D5.7	0.5573	8.7%	6.9%	7.8%	7.4%	18.7%	8.3%
D5.8	0.6205	8.8%	7.0%	7.9%	7.5%	20.1%	8.4%
D5.9	0.6416	8.9%	7.3%	8.2%	7.8%	20.4%	8.5%
D5.10	0.6615	8.9%	7.4%	8.4%	7.8%	20.5%	8.6%
D6.1	0.6845	8.9%	7.5%	8.7%	8.1%	22.1%	8.8%
D6.2	0.6845	9.3%	7.8%	8.9%	8.1%	24.8%	9.1%
D6.3	0.716	9.6%	8.0%	8.9%	8.2%	26.3%	9.3%
D6.4	0.8137	9.9%	8.5%	8.9%	8.3%	26.9%	9.6%
D6.5	0.8648	10.2%	8.5%	9.0%	8.5%	27.0%	10.1%
D6.6	0.8908	10.2%	8.6%	9.1%	8.6%	27.5%	10.5%
D6.7	0.8944	10.7%	9.1%	9.3%	8.8%	31.2%	10.5%
D6.8	0.8996	10.9%	9.2%	9.5%	9.1%	31.4%	11.1%
D6.9	0.9267	10.9%	9.5%	9.5%	9.1%	32.2%	11.2%
D6.10	0.9847	11.2%	9.5%	9.7%	9.2%	32.8%	11.3%
D7.1	1.0093	12.0%	9.5%	9.8%	9.3%	33.8%	11.7%
D7.2	1.03	12.1%	10.0%	9.9%	9.3%	34.0%	12.4%
D7.3	1.0769	13.7%	10.1%	10.0%	9.5%	35.2%	12.9%
D7.4	1.0879	14.2%	10.1%	10.3%	9.5%	39.9%	15.5%
D7.5	1.2545	15.0%	10.2%	10.4%	9.6%	43.3%	15.5%
D7.6	1.2732	15.1%	10.4%	10.5%	9.8%	47.7%	15.9%

Designs	Distance	Power of tests at 5% critical values, Sample size 100					
		Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D7.7	1.2775	15.6%	10.5%	10.6%	9.8%	49.8%	16.3%
D7.8	1.3125	16.3%	10.8%	10.6%	10.1%	51.8%	16.3%
D7.9	1.5019	16.6%	10.8%	10.7%	10.3%	52.2%	16.7%
D7.10	1.68	17.1%	11.0%	10.9%	10.5%	52.9%	16.7%
D8.1	1.706	17.2%	11.1%	11.0%	11.3%	53.6%	17.0%
D8.2	1.7308	17.2%	11.6%	11.3%	11.4%	54.1%	17.1%
D8.3	1.7487	17.6%	11.7%	11.7%	11.7%	55.7%	18.1%
D8.4	1.9545	17.7%	11.8%	12.2%	12.0%	58.8%	18.3%
D8.5	1.9965	19.7%	11.9%	12.3%	12.3%	61.0%	20.7%
D8.6	2.0367	19.7%	11.9%	12.4%	12.4%	63.2%	21.7%
D8.7	2.3256	21.9%	13.0%	13.0%	13.8%	64.7%	22.1%
D8.8	2.3798	22.6%	13.7%	13.5%	14.1%	65.1%	22.4%
D8.9	2.4609	24.9%	15.1%	14.0%	15.6%	69.1%	22.9%
D8.10	2.5427	26.7%	15.5%	14.3%	15.7%	69.9%	23.6%
D9.1	2.5714	26.8%	15.8%	14.5%	16.4%	70.8%	25.0%
D9.2	2.6374	27.1%	16.5%	14.6%	17.2%	71.2%	25.3%
D9.3	2.6413	28.7%	16.5%	14.6%	18.1%	71.2%	26.0%
D9.4	2.6532	28.7%	17.2%	14.7%	18.3%	71.8%	27.5%
D9.5	2.6729	28.8%	17.8%	14.8%	19.7%	71.9%	27.6%
D9.6	2.6963	28.8%	20.6%	15.0%	19.9%	75.2%	27.8%
D9.7	2.703	31.8%	20.7%	15.4%	20.2%	75.9%	28.8%
D9.8	2.735	32.2%	21.8%	17.0%	20.6%	76.9%	30.7%
D9.9	2.9786	32.3%	22.1%	17.2%	22.9%	79.2%	31.3%
D9.10	3.1815	32.6%	25.3%	17.7%	24.1%	80.8%	32.1%
D10.1	3.2076	33.1%	26.4%	17.9%	28.0%	83.2%	33.4%
D10.2	3.4921	33.6%	28.7%	19.3%	28.7%	85.0%	34.0%
D10.3	3.4921	34.4%	30.5%	19.9%	29.2%	89.0%	36.0%
D10.4	3.7118	35.9%	31.0%	23.1%	29.9%	91.5%	36.8%
D10.5	4.0049	37.3%	33.3%	24.9%	31.7%	92.5%	38.6%
D10.6	4.2308	40.0%	34.4%	28.7%	35.3%	97.5%	39.2%
D10.7	5.1166	41.4%	35.8%	31.0%	35.6%	98.1%	44.8%
D10.8	5.5797	60.4%	40.0%	35.5%	38.0%	98.8%	54.0%
D10.9	6.2799	61.4%	55.4%	41.7%	52.9%	99.7%	62.5%
D10.10	6.303	65.0%	59.1%	50.8%	56.8%	99.8%	65.8%

		Power of tests at 1% critical values, Sample size 100					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D1.1	0	0.6%	0.1%	0.0%	0.0%	0.0%	0.8%
D1.2	0	0.7%	0.1%	0.0%	0.1%	0.0%	0.8%
D1.3	0.0034	0.8%	0.1%	0.0%	0.1%	0.0%	0.9%
D1.4	0.0055	0.8%	0.5%	0.0%	0.3%	0.0%	0.9%
D1.5	0.006	0.9%	0.5%	0.0%	0.4%	0.0%	0.9%
D1.6	0.0236	0.9%	0.6%	0.0%	0.5%	0.0%	1.1%
D1.7	0.0316	0.9%	0.7%	0.0%	0.5%	0.0%	1.1%
D1.8	0.05435	0.9%	0.8%	0.2%	0.6%	0.0%	1.2%
D1.9	0.0668	1.0%	0.9%	0.2%	0.6%	0.0%	1.2%
D1.10	0.0668	1.2%	0.9%	0.4%	0.7%	0.1%	1.3%
D2.1	0.0684	1.3%	0.9%	0.4%	0.7%	0.1%	1.3%
D2.2	0.0831	1.3%	1.1%	0.4%	0.8%	0.1%	1.3%
D2.3	0.0963	1.4%	1.2%	0.4%	0.8%	0.1%	1.4%
D2.4	0.0968	1.4%	1.2%	0.5%	0.9%	0.1%	1.4%
D2.5	0.0998	1.5%	1.3%	0.5%	1.0%	0.2%	1.5%
D2.6	0.10565	1.5%	1.3%	0.5%	1.0%	0.2%	1.5%
D2.7	0.1479	1.5%	1.3%	0.7%	1.1%	0.3%	1.5%
D2.8	0.14995	1.5%	1.3%	0.8%	1.1%	0.3%	1.6%
D2.9	0.1536	1.6%	1.3%	0.8%	1.1%	0.5%	1.6%
D2.10	0.1768	1.6%	1.4%	0.8%	1.1%	0.5%	1.6%
D3.1	0.1787	1.6%	1.4%	0.9%	1.1%	0.6%	1.7%
D3.2	0.1813	1.6%	1.4%	1.0%	1.1%	0.7%	1.7%
D3.3	0.1814	1.7%	1.4%	1.1%	1.2%	0.7%	1.7%
D3.4	0.1883	1.7%	1.4%	1.1%	1.2%	0.7%	1.7%
D3.5	0.2052	1.8%	1.4%	1.1%	1.2%	0.9%	1.7%
D3.6	0.21765	1.9%	1.4%	1.2%	1.2%	1.0%	1.7%
D3.7	0.234	1.9%	1.5%	1.3%	1.3%	1.4%	1.8%
D3.8	0.2524	2.0%	1.5%	1.3%	1.4%	1.5%	1.8%
D3.9	0.2524	2.0%	1.5%	1.3%	1.5%	1.8%	1.8%
D3.10	0.2808	2.0%	1.5%	1.4%	1.5%	2.9%	1.8%
D4.1	0.3373	2.1%	1.5%	1.4%	1.5%	2.9%	1.8%
D4.2	0.38065	2.2%	1.6%	1.4%	1.5%	3.1%	1.9%
D4.3	0.3925	2.2%	1.6%	1.4%	1.6%	3.2%	2.0%

		Power of tests at 1% critical values, Sample size 100					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D4.4	0.3982	2.2%	1.7%	1.4%	1.7%	3.5%	2.1%
D4.5	0.4114	2.3%	1.7%	1.4%	1.8%	3.9%	2.1%
D4.6	0.4444	2.3%	1.7%	1.6%	1.8%	4.4%	2.3%
D4.7	0.4635	2.4%	1.7%	1.6%	1.8%	6.0%	2.3%
D4.8	0.4663	2.5%	1.8%	1.6%	1.8%	6.4%	2.3%
D4.9	0.4861	2.5%	1.9%	1.6%	1.8%	6.5%	2.4%
D4.10	0.4911	2.6%	2.0%	1.7%	1.9%	6.6%	2.4%
D5.1	0.523	2.6%	2.0%	1.7%	1.9%	6.7%	2.5%
D5.2	0.5372	2.6%	2.0%	1.7%	2.0%	7.6%	2.6%
D5.3	0.5381	2.7%	2.2%	1.8%	2.0%	8.9%	2.8%
D5.4	0.5382	2.9%	2.2%	1.8%	2.0%	9.3%	2.9%
D5.5	0.5442	2.9%	2.3%	1.8%	2.1%	10.1%	2.9%
D5.6	0.5495	2.9%	2.4%	1.8%	2.2%	10.4%	3.1%
D5.7	0.5573	3.1%	2.4%	1.8%	2.2%	11.6%	3.2%
D5.8	0.6205	3.2%	2.5%	1.9%	2.4%	12.6%	3.2%
D5.9	0.6416	3.2%	2.5%	1.9%	2.4%	13.2%	3.2%
D5.10	0.6615	3.3%	2.8%	1.9%	2.4%	13.3%	3.2%
D6.1	0.6845	3.3%	2.8%	1.9%	2.4%	13.7%	3.4%
D6.2	0.6845	3.4%	2.8%	2.0%	2.5%	14.4%	3.7%
D6.3	0.716	4.1%	2.8%	2.1%	2.5%	15.1%	4.7%
D6.4	0.8137	4.2%	2.8%	2.1%	2.5%	17.7%	4.7%
D6.5	0.8648	4.6%	2.9%	2.1%	2.6%	20.6%	5.0%
D6.6	0.8908	4.7%	3.0%	2.1%	2.6%	23.0%	5.0%
D6.7	0.8944	4.8%	3.0%	2.2%	2.7%	24.4%	5.1%
D6.8	0.8996	4.9%	3.0%	2.3%	2.7%	25.9%	5.3%
D6.9	0.9267	5.2%	3.1%	2.4%	2.7%	27.6%	5.4%
D6.10	0.9847	5.3%	3.1%	2.4%	2.8%	28.3%	5.5%
D7.1	1.0093	5.5%	3.2%	2.4%	2.9%	28.6%	5.6%
D7.2	1.03	5.7%	3.3%	2.5%	2.9%	30.9%	5.7%
D7.3	1.0769	5.9%	3.3%	2.5%	2.9%	31.8%	5.9%
D7.4	1.0879	5.9%	3.4%	2.6%	2.9%	32.9%	6.2%
D7.5	1.2545	6.2%	3.5%	2.7%	3.0%	35.1%	6.8%
D7.6	1.2732	6.4%	3.6%	2.8%	3.1%	37.0%	7.5%

		Power of tests at 1% critical values, Sample size 100					
Designs	Distance	Pearson χ^2	LR	λ test	UC	Stuart-Maxwell test	Generalized fisher exact test
D7.7	1.2775	6.4%	4.1%	2.8%	3.1%	38.6%	7.6%
D7.8	1.3125	6.5%	4.5%	2.8%	3.4%	39.4%	7.8%
D7.9	1.5019	8.0%	4.8%	2.8%	3.4%	42.8%	8.1%
D7.10	1.68	9.2%	5.0%	2.9%	3.8%	42.9%	8.8%
D8.1	1.706	10.0%	5.0%	2.9%	4.0%	43.2%	9.3%
D8.2	1.7308	11.0%	5.1%	2.9%	4.2%	45.3%	9.4%
D8.3	1.7487	11.0%	5.5%	3.0%	4.3%	45.3%	10.1%
D8.4	1.9545	11.1%	5.6%	3.2%	4.7%	45.4%	10.3%
D8.5	1.9965	11.1%	5.7%	3.2%	4.9%	45.5%	10.6%
D8.6	2.0367	11.1%	5.7%	3.2%	5.3%	46.0%	11.0%
D8.7	2.3256	11.2%	6.4%	3.4%	5.3%	48.8%	11.2%
D8.8	2.3798	11.5%	6.8%	3.6%	5.8%	52.2%	12.6%
D8.9	2.4609	12.5%	7.5%	3.7%	5.8%	53.0%	12.6%
D8.10	2.5427	12.9%	8.7%	3.7%	5.8%	58.6%	13.6%
D9.1	2.5714	13.0%	9.1%	3.7%	6.1%	59.5%	13.6%
D9.2	2.6374	13.2%	10.9%	4.1%	6.5%	60.7%	14.7%
D9.3	2.6413	14.4%	11.2%	4.3%	6.5%	63.5%	15.9%
D9.4	2.6532	15.2%	12.5%	4.4%	6.9%	64.3%	17.1%
D9.5	2.6729	15.2%	12.5%	4.7%	7.8%	70.2%	18.5%
D9.6	2.6963	15.7%	13.6%	4.7%	8.2%	74.4%	20.8%
D9.7	2.703	15.8%	14.8%	4.8%	8.2%	77.1%	20.8%
D9.8	2.735	17.7%	15.5%	5.4%	9.4%	89.3%	27.9%
D9.9	2.9786	21.4%	16.3%	6.1%	10.7%	90.5%	30.6%
D9.10	3.1815	30.1%	17.3%	6.9%	10.7%	93.6%	35.8%
D10.1	3.2076	32.6%	30.4%	9.0%	12.1%	97.2%	38.5%
D10.2	3.4921	35.4%	33.3%	9.0%	12.3%	98.3%	40.0%
D10.3	3.4921	39.5%	34.1%	9.6%	13.0%	113.8%	48.4%
D10.4	3.7118	42.2%	34.8%	10.7%	14.0%	119.9%	51.8%
D10.5	4.0049	46.7%	49.9%	12.4%	15.1%	210.0%	56.4%
D10.6	4.2308	49.5%	51.3%	18.4%	15.4%	214.6%	65.5%
D10.7	5.1166	124.4%	126.5%	23.7%	15.9%	214.6%	139.2%
D10.8	5.5797	146.3%	152.6%	34.1%	26.2%	245.3%	169.7%
D10.9	6.2799	227.7%	214.8%	34.1%	30.4%	450.6%	218.5%
D10.10	6.303	414.8%	481.1%	68.2%	31.6%	636.9%	453.9%

Appendix-B

Programming for Data Generating Process and Simulation Design

```
function [CTsimulated]=CT(X,n)
a=sum(sum(X));
Y=X/a;
b=sum(Y')';
c=sum(Y);
d=sum(sum(Y));
CTsimulated=zeros(3,3);
for i=1:n
    x=rand;
    if x<b(1,1);
        RN=1;
    else if x< (b(1,1)+b(2,1));
        RN=2;
        else
            RN=3;
        end
    end
    y=rand;
    if y< Y(RN,1)/b(RN,1);
        CN=1;
    else if y<Y(RN,1)/b(RN,1)+ Y(RN,2)/b(RN,1);
            CN=2;
        else
            CN=3;
        end
    end
     CTsimulated(RN,CN)=CTsimulated(RN,CN)+1;
end
```

Simulated critical values

```
X=[6 5 5;12 10 10;18 15 15];
tic
resAT=zeros(20000,7);
for i=1:20000
    X2=CT(X,100);
    a=chi1(X2);
    b=LR1(X2);
    c=PR(X2);
    ASE2=SPC(X2);
    Z1=c/ASE2;
    c1=PR1(X2);
    ASE1=SPR(X2);
     Z2=c1/ASE1;
    c2=PR2(X2);
    d=UC1(X2);
    ASE3=UR(X2);
    Z3=d/ASE3;
    d1=UC2(X2);
    ASE4=UC(X2);
    Z4=d1/ASE4;
    d2=UC3(X2);
    e=SM(X2);
    f=Exact(X2);

    resAT(i,1)=a;
    resAT(i,2)=b;
```

```
    resAT(i,3)=Z1;
    resAT(i,4)=Z3;
    resAT(i,5)=e;
    resAT(i,6)=f;
end
cv=prctile(resAT,95)
save D:\resATcvqfile290.txt resAT cv -ascii
toc
```

Computation of Power

```
X=[5 5 1;10 10 10;15 15 15];
tic
Pr=zeros(1,6);
for i=1:100
    X2=CT(X,25);
    a=chi1(X2);
    b=LR1(X2);
    c=PR(X2);
    ASE2=SPC(X2);
    Z1=c/ASE2;
    c2=PR2(X2);
    d=UC1(X2);
    ASE3=UR(X2);
    Z3=d/ASE3;
    e=SM(X2);
    f=Exact(X2);
  teststat=[a b Z1 Z3 e f];
  cv=[9.22783   5.91621 2.32551 1.09965 7.6716  9.59985];
   pr2=teststat>cv;
   Pr=Pr+pr2;
end
Pr
toc
```