A SIMULATION STUDY OF THE SIZE AND POWER OF COCHRAN'S Q VERSUS THE STANDARD CHI-SQUARE TEST FOR TESTING THE EQUALITY OF CORRELATED PROPORTIONS

by

SUELEN S GAYLE

B.S., University of the West Indies Mona, 1999

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2010

Approved by:

Major Professor:
Dr. Paul Nelson

Abstract

The standard Chi-square test for the equality of proportions of positive responses to *c* specified binary questions is valid when the observed responses arise from independent random samples of units. When the responses to all *c* questions are recorded on the same unit, a situation called *correlated proportions*, the assumptions under which this test is derived are no longer valid. Under the additional assumption of compound symmetry, the Cochran-Q test is a valid test for the equality of proportions of positive responses. The purpose of this report is to use simulation to examine and compare the performance of the Cochran-Q test and the standard Chi-square test when testing for the equality of correlated proportions. It is found that the Cochran-Q test is superior to the Chi-square test in terms of size and power, especially when the common correlation among the binary responses is large.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First I would like to thank my major professor, Dr. Paul Nelson, for his supervision, support, encouragement, and patience throughout this project. I would like to thank the members of my committee: Dr. John Boyer and Dr. Suzanne Dubnicka both for asking good questions in the process and for helping with my programming needs. I would also like to thank Dr. Andresen and Aaron Tygart in the CIS Department of the College of Engineering for allowing me to use Beocat, the K-State beowulf cluster. Thank you to my colleagues in the Statistics department who have provided assistance in many ways. Finally, I also thank my family for their support.

CHAPTER 1 - Introduction

Suppose that binary responses $\{X_{ij},\ j=1,2,...,c;\ i=1,2,...,n_j\}$ are recorded on units exposed to *c* environments $\{E_j\}$. Let X_{ij} =1 denote a *positive* response to E_j and let the expected values E(X_{ij}) = π_j, j=1,2,...,c . We are interested in using the observed data $\{x_{ij}\}$ to test the hypotheses

$$H_0 : \pi_1 = \pi_2 = \cdots \pi_c \quad \text{vs} \quad H_a : \pi_i \neq \pi_j \text{ for at least one pair } (i, j) . \tag{1.1}$$

For the purposes of this study, it is helpful to think of the *c* environments as questions and that we are interested in comparing the proportions of *positive* (x_{ij} =1) responses to these questions. An important consideration in choosing the appropriate method to carry out this test is specification of how the data were collected. Here, we consider two cases. In Case 1, X_{il} and X_{jk} are mutually independent if $(i,l) \neq (j,k)$. In Case 2, X_{il} and X_{ik} are recorded on the same unit and hence, are typically not independent. Case 2 is referred to as the problem of testing for the equality of *correlated proportions*. The data structure in Case 1 is in the form of a completely randomized, one-way design with n_j binary responses in environment E_j, *j* = *1,2…,c*, as laid out in Table 1.1 below.

Table 1.1: Case 1 { X_{ij} } Independent

Environments			
E_1	E_2	$\cdots$	E_c
X_{11}	X_{12}	$\cdots$	X_{1c}
X_{21}	X_{22}	$\cdots$	X_{2c}
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$X_{n_1 1}$	$X_{n_2 2}$	$\cdots$	$X_{n_c c}$

In Case 2, each of r units is exposed to all c environments. The exposures to the environments are independently randomized over each unit. This data structure amounts to a complete, randomized block design where the units are blocks and the responses are binary.

Table 1.2: Case 2 Correlated Proportions

	Environments				
Blocks	1	2	…	c	Row Totals
1	X_{11}	X_{12}	…	X_{1c}	R_1
2	X_{21}	X_{22}	…	X_{2c}	R_2
…	…	…	…	…	…
r	X_{r1}	X_{r2}	…	X_{rc}	R_r
Column Totals	C_1	C_2	…	C_c	M

Example 1.1: Suppose that an investigator wants to compare the proportions of positive responses to c fixed questions in a large target population of students. Using a Case 1 design, $n_1 + n_2 + \cdots + n_c$ students would be randomly selected from the population and, at random, n_j students would be asked to respond to question j, $j = 1, 2, ..., c$ and their responses scored as right or wrong. In Case 2, r students would be randomly selected and their responses, in random order, to all c questions recorded.

Example 1.2: Suppose the investigator wants to carry out a simulation study to compare the power of c hypothesis tests, the environments, at some fixed alternative, H_a. Using the design in Case 2, the investigators simulates r independent data sets under H_a, and carries out all c tests on each data set and records the proportions of times out of r that each test leads to the rejection of H_0. In Case 1, the investigator would have simulated rc independently, identically distributed samples under H_a. Each of the c tests would be carried out on r of these samples and a record made of whether or not the null hypothesis was rejected. In this setting, the blocking in Case 2 requires fewer units than Case 1 and has the advantage of controlling for heterogeneity among units. Specifically, Case 1 requires the generation of *rc* data sets while Case 1 requires only *r* datasets. However, Case 2 requires an additional assumption of compound symmetry, as described below, in order to carry out a valid test of (1.1). The proper analyses for these two designs in the general setting should be carried out as follows.

Case 1: The Standard Chi-Square test for Equality of Proportions

Sum the rows in Table 1.1 and display the results in Table 1.3, called a 2x*c* contingency table. The entries $\left\{O_{1j} = \sum_{i=1}^{n_j} X_{ij} = C_j, \, j = 1,2,...,c\right\}$ and $\{O_{2j} = n_j - O_{1j}\}$ and in Table 1.3 are obtained by tallying the total number of ones and zeros, now called 'successes' and 'failures', respectively, in each column of Table 1.1. The row totals { $R_i^{\#}$ } of Table 1.3 are respectively the total number of observed successes and failures. Note that $R_1^{\#} + R_2^{\#} = n_1 + n_2 ... + n_c = N$ and $O_{i1} \sim B(n_i, \pi_i)$, $i = 1, 2, ..., c$, are independent, binomial random variables.

Table 1.3: Data Structure for Standard Chi-Square

Outcomes	E_1	E_2	…	E_c	*Totals*
Success	O_{11}	O_{12}	…	O_{1c}	$R_1^{\#}$
Failure	O_{21}	O_{22}	…	O_{2c}	$R_2^{\#}$
Total	n_1	n_2	…	n_c	N

Increasing values of the **test statistic**

$$T = \sum_{i=1}^{r} \sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \text{ where } E_{ij} = \frac{n_i C_j}{N}, \qquad (1.2)$$

provide increasing support for H_a over H_0 given in (1.1). The test statistic measures the relative squared difference between the estimated expected counts $\{E_{ij}\}$ under H_0 and the observed counts $\{Oij\}$ in Table 1.3. One of the main assumptions of this test is that of mutual independence among the samples, as specified in Case1.

Decision Rule: For large samples, reject H_0 at approximate type 1 error rate α if $T \geq \chi^2_{(1-\alpha,c-1)}$, the 100(1-$\alpha$) percentage point of a chi-square distribution with $c-1$ degrees of freedom.

Case 2: Cochran's Q Test

Recall that $X_{ij} \sim B(1,\pi_j)$, X_{ij} is independent of X_{kl} for $i \neq k$ but X_{ij} may be correlated with X_{il}, $l \neq j$. Thus, the rows of Table 1.2 are independent random vectors. But, random variables within a row may be correlated with $E(X_{ij} \cdot X_{ik}) = \pi_{j,k}$ *,j≠k=1,2,…c; i = 1,2,…r.*

Cochran (1950) developed a randomization test in this setting which actually tests the more restricted null hypothesis

$$H_{00}: \pi_1 = \pi_2 = ... = \pi_c \text{ and } \pi_{\text{ij}} = \pi_{1,2} = \pi_{1,3} = ... = \pi_{(c-1),c}, \quad (1.3a)$$

where $\pi_{lm} = P(X_{il} = 1, X_{im} = 1)$, i = 1,2,…,r, $l \neq m$.

Clearly, H_{00} implies H_0 in (1.1). But, the converse is not true. Note that H_{00} is equivalent to

$$H^*_{00}: \pi_1 = \pi_2 = \cdots = \pi_c \text{ and } \rho_{ij} = \rho_{1,2} = \rho_{1,3} = ... = \rho_{(c-1),c}, \quad (1.3b)$$

where ρ_{lm} = corr(X_{il}, X_{im}), $l \neq m$.

The second condition in (1.3b) is commonly called *compound symmetry*. Mandansky (1963) called H_{00} the hypothesis of interchangeability. Berger and Gold (1973) and Bapkara and Somes (1977) showed that the Q- statistic only has a limiting chi-square distribution under H^*_{00} or H_{00} .

Cochran's Q

Cochran's (1950) test statistic is given by

$$Q = c(c-1)\sum_{j=1}^{c}(C_j - M/c)^2 / (cM - \sum_{i=1}^{r} R_i^2) \quad (1.4)$$

where $M = \sum C_j = \sum R_i$.

An asymptotically size α test is given by:

Decision Rule: Reject H_{00} at approximate type I error rate α if $Q > \chi_{1-\alpha,\nu}$, $\nu = (c-1)$.

Berger and Gold (1973) showed that under H_0, the asymptotic distribution of Q is given by a linear combination of independent, single degree of freedom chi-square variates, where the coefficients are difficult to estimate. Mandansky (1963) showed that as a test of exchangeability, Cochran's test is not consistent against all alternatives. Wallenstein and Berger (1981) developed an approximate test for H_0 that performed reasonably well in terms of size and power in a small scale simulation study. Vitalliano (1979) conducted a simulation study which indicates that Cochran's Q, used as a test statistic for H_0, tends to be conservative in small samples and performs reasonably well in terms of type I error rate unless the hypothesis of compound symmetry is grossly violated.

When c = 2, Cochran's Q is equivalent to the well known McNemar's (1947) test.

McNemar's Test

For c = 2, the rows of Table 1.2 may be viewed as independent realizations of random variables (Y_1, Y_2), where Y_j = 1 if the response to condition *j* is a success and 0 otherwise, j = 1,2. Table 1.4 summarizes the entries in Table 1.2 in a 2 x 2 table, where the four cells correspond to counts of the possible values of (Y_1, Y_2).

Table 1.4: Summary Counts for Comparing Two Proportions

	Y_2		
Y_1	**0**	**1**	Total
0	e	f	$e+f$
1	g	h	$g+h$
Total	$e+g$	$f+h$	N

McNemar's test statistic $\chi^2 = \dfrac{(f-g)^2}{f+g}$ is used to test, in the notation of (1.1) for c=2,

$$H_o : \pi_1 \equiv P\left(Y_1 = 1\right) = P\left(Y_2 = 1\right) \equiv \pi_2 ,$$

$$H_a : P\left(Y_1 = 1\right) \neq P\left(Y_2 = 1\right) .$$

Note that this test statistic only uses those blocks in which a unit responds differently to the two environments. This makes sense since these are the only blocks that contain information about the $\pi_1 - \pi_2$. Compound symmetry holds by default here since c=2. The decision rule for McNemar's test is given by:

Reject H_0 at approximate level of significance α if $\chi^2 > \chi^2_{1-\alpha,1}$.

Relationship Between the McNemar's and Cochran's Q test

McNemar's and Cochran's Q tests are equivalent when there are c=2 responses in each row of Table 1.2. To verify this statement, we start out by setting c = 2 in Q,

$$Q = 2\sum_{j=1}^{c}(C_j - M/c)^2 / (2\sum_{i=1}^{r} R_i - \sum\nolimits_{i=1}^{r} R_i^2)$$

Since $M = \sum C_j = C_1 + C_2$,

$$2\sum_{j=1}^{2}(C_j - M/c)^2 = 2[(C_1 - (C_1 + C_2)/2)^2 + (C_2 - (C_1 + C_2)/2)^2] \qquad (1.5)$$

$$= (C_1 - C_2)^2$$

$$= (g + h - f - h)^2$$

$$= (f - g)^2.$$

Similarly,

$$2\sum_{i=1}^{r} R_i - \sum_{i=1}^{r} R_i^2 = 2(f+g+2h)-(f+g+4h) \tag{1.6}$$

$$= (f+g)$$

Dividing (1.5) by (1.6) yields the desired result.

An Example

(Adapted from Conover ,1980)

Each of three basketball enthusiasts had devised his own system for predicting the outcomes of collegiate basketball games. Twelve games were selected at random, and each sportsman presented a prediction of the outcome of each game. After the games were played, the results were tabulated, using "1" for successful prediction and "0" for unsuccessful prediction. Table 1.4 summarizes the outcomes. This example falls under Case 2. We have c=3 environments which are each of the sportsmen. The r=12 games are the blocks. Here, Cochran's-Q is used to test the hypothesis:

$$H_0 : \pi_1 = \pi_2 = \pi_3, \qquad \text{vs}$$

$$H_a : \pi_i \neq \pi_j \text{ for at least two } i,j$$

Using equation 1.4, one finds a value of $Q = 2.8$, and after comparing Q to the tabled value of $\chi_{(0.05,2)} = 5.99$, we fail to reject H_0 at the nominal type I error rate 0.05 and conclude that there is not a statistically significant difference among the three prediction systems. Table 1.5 also gives pairwise correlations of the responses, values that can in an informal manner be used to assess the validity of the assumption of compound symmetry.

Chapter Two will provide the background on the algorithm used in generating the artificial binary data and the software used to generate the data. Chapter Three will present the findings of the study while Chapter Four reviews finding and gives recommendations.

Table 1.5: Example Data

	Sportsman			
Game	1	2	3	Totals
1	1	1	1	3
2	1	1	1	3
3	0	1	0	1
4	1	1	0	2
5	0	0	0	0
6	1	1	1	3
7	1	1	1	3
8	1	1	0	2
9	0	0	1	1
10	0	1	0	1
1	1	1	1	3
2	1	1	1	3
Totals	8	10	7	25
Sample proportions	0.667	0.833	0.583	1
Pairwise Correlations	1vs2	1vs3	2vs3	
	0.633	0.478	0.076	

CHAPTER 2 - Generating Artificial Binary Correlated Data and Simulation Study Design

Leisch et al. (1998) outlined an algorithm for generating correlated binary data from multivariate binary distributions. In addition, they created an R package called bindata, which allows the user to specify values of the correlation matrix, marginal success rates, and sample size. The result is a simulated dataset of 1's and 0's with correlated columns. By default, if the correlation structure is not specified, the data generated will be independent across columns. Appendix A provides more information on the bindata package.

The simulation experiment conducted here specified various levels of *r*, *c*, correlations $\rho_{jk} = \text{corr}(X_{ij}, X_{ik})$, $j \neq k$, and success rates $\{\pi_j\}$ and simulated data in the form of Table 1.2. In deciding on the specific values of ρ_{jk} and *r* to use in the study, a preliminary study was done using various combinations of each. After trying different combinations, correlations between 0 and 0.6 were found to be best, as other combinations were not always compatible with the bindata package. From the preliminary study the information on the number of cpu hours for different combinations was collected. This was very useful in deciding on final values of *r* and *c* since the code is a fairly slow one. Based on these results it was decided to look at values of *r* = 50, 100 and 200. Values of *r* greater than 300 were initially considered but in most cases produced powers equal to one therefore not providing much variability for analysis. The marginal proportions $\{\pi_i\}$ used were centered about 0.5. In all cases the data generated satisfied the condition of compound symmetry.

Each simulation was carried out 1000 times, thus generating 1000 binary correlated data sets on which both the Cochran-Q test and the standard Chi-square test were carried out. Estimated rejection rates of H_0 given in (1.3b) are summarized, and an assessment as to how well Cochran's Q and the standard Chi-square test performs in terms of size and power is made. From each type of test the p-values were stored and compared to $\alpha = 0.05$, scoring a 1 for a p-value less than α alpha and 0 otherwise. The proportion of ones in each type of test was then recorded. This data was used to assess the estimated powers and the estimated type one error rates of both tests. In addition the differences in the results of both tests were compared. Since the data being used to compare the two tests was based on the same original data set, the appropriate test to compare the power and type I error rates is Cochran-Q test.

CHAPTER 3 - Simulation Study Results

The estimated type 1 error rates as well as the power of both the Cochran-Q test and the standard Chi-square test are reported. In addition, the p-values for the comparison of the performance of the two tests are reported. All these tests were carried out at nominal type I error rate $\alpha = 0.05$. Tables 3.1– 3.8 present the results of the simulation study.

To facilitate comparisons based on the marginal distribution, the non-centrality type parameter δ was introduced, where $\delta = r\sum_{c}(\pi_c - \bar{\pi})^2/(c-1)$, { π_i }are the specified success rates for each of the *c* populations under the alternative, and $\bar{\pi}$ is the mean of the specified proportions. The parameter δ is zero under H_o and increases as the variation among the specified proportions increases. The values of δ for the different scenarios considered are shown in Table 3.1.

Table 3.1: Non-Centrality Parameters under the Alternative Hypothesis

	Marginal Probability		
c=2	**(0.4,0.6)**	**(0.45,0.55)**	**(0.48,0.52)**
r			
50	1.000	0.250	0.040
100	2.000	0.500	0.080
200	4.000	1.000	0.160
c=3	**(0.40,0.50,0.60)**	**(0.45,0.50,0.55)**	**(0.48,0.5,0.52)**
r			
50	0.500	0.125	0.020
100	1.000	0.250	0.400
200	2.000	0.500	0.080

In addition, the estimated standard errors are calculated using the binomial distribution for both the estimated type one error rates and the estimated powers. In each case $\hat{\pi} = y/N$ and

$se(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{N}}$, where *y* is the number of times the null hypothesis is rejected by a given test and *N=1000* and is the number of simulations carried out.

These standard errors are used to compute the estimated margins of errors (m.o.e) for a 95% confidence level and are presented in Table 3.2. This is calculated only for a few representative values of $\hat{\pi}$ that give the reader an idea of the estimated m.o.e for values of $\hat{\pi}$ close to those in Table 3.2.

Table 3.2: Standard Errors for Estimates for c=2 and c=3

$\hat{\pi}$	2s.e($\hat{\pi}$)
0.01	0.003
0.05	0.007
0.10	0.009
0.15	0.011
0.20	0.013
0.50	0.016

For valid comparisons of power of Cochran's Q and the standard Chi-square we want the estimated type I error rates to be equal or relatively close for the two tests. This difference needs to be considered throughout the analysis. The estimated type I error rates for the Cochran-Q are closer to the nominal type I error rate of 0.05 while the estimated type I error rates for the Chi-square test are much smaller than 0.05, the exception being when $\rho = 0$. The tradeoff between the lower type I error rates and the power is noticeable in the relatively lower power of the Chi-Square test for $\rho > 0$. This is true both in the case of c=2 and c=3.

Two Environments: c=2

In Table 3.3 the estimated type I error rates are shown for the Cochran-Q and the standard Chi-square tests for c=2. The estimated type I error rates average 0.05 for all values of ρ under the Cochran-Q test. For the Chi-square test, the averages were 0.06, 0.014 and 0.003

for ρ = 0, 0.4, 0.6 respectively. The Cochran-Q performed similarly for all values of ρ, while the Chi-square had estimated type I error rate much smaller than the nominal type I error rate of 0.05 for non-zero values of ρ.

Table 3.4 presents estimated powers for c=2, using three different pairs of marginal distributions. Figures 3.1-3.3 also makes comparisons of the estimated power as δ changes and for different values of ρ. The power of the Cochran-Q test increased as δ increased. In all cases of ρ (including zero), the power of the test increased as δ increased. This was also true when the standard Chi-square test was used instead of the recommended Cochran's Q test.

With respect to changes in power as ρ changed, all else held constant, for Cochran's test, the larger ρ the greater the power of the test. The opposite was true for the estimated powers under the Chi-square test, instead the powers were higher for lower values of ρ, all else held constant. This was expected based on the low type I error rates for the chi-square test when ρ is large. As mentioned before the two tests have different estimated type one error rates which show up here in the differences in power.

For fixed marginal probabilities and ρ, the power increased as r, the number of rows in Table 1.2, increased for both the Cochran-Q and the standard Chi-square.

Using the Cochran-Q test to compare the power of the two tests gave p-values less than 0.05 in all cases of ρ = 0.4 and 0.6 leading to the conclusion that there is a statistically significant difference between the result we get from the two tests. The only exception to this was r =200 and ρ = 0. For ρ = 0 however the results were mixed, the differences were not statistically significant when r =200. The p-values for the difference between the two test (see Table 3.5) indicates a significant difference between the results of the two test when ρ = 0.4 and ρ = 0.6.

Table 3.3: Estimated Type I Error Rates for c=2

Cochran's Q				Chi-Square			
r=50	ρ			***r=50***	ρ		
Marginal Probabilities	**0**	**0.4**	**0.6**	**Marginal Probabilities**	**0**	**0.4**	**0.6**
(0.4,0.4)	0.068	0.056	0.049	(0.4,0.4)	0.067	0.016	0.004
(0.5,0.5)	0.069	0.054	0.047	(0.5,0.5)	0.076	0.016	0.004
(0.6,0.6)	0.064	0.050	0.046	(0.6,0.6)	0.062	0.011	0.003
r=100	ρ			***r=100***	ρ		
Marginal Probabilities	**0**	**0.4**	**0.6**	**Marginal Probabilities**	**0**	**0.4**	**0.6**
(0.4,0.4)	0.061	0.054	0.051	(0.4,0.4)	0.063	0.013	0.000
(0.5,0.5)	0.073	0.069	0.060	(0.5,0.5)	0.081	0.016	0.003
(0.6,0.6)	0.070	0.051	0.054	(0.6,0.6)	0.066	0.013	0.006
r=200	ρ			***r=200***	ρ		
Marginal Probabilities	**0**	**0.4**	**0.6**	**Marginal Probabilities**	**0**	**0.4**	**0.6**
(0.4,0.4)	0.052	0.050	0.045	(0.4,0.4)	0.055	0.015	0.000
(0.5,0.5)	0.034	0.032	0.040	(0.5,0.5)	0.036	0.009	0.003
(0.6,0.6)	0.046	0.040	0.035	(0.6,0.6)	0.041	0.011	0.001

Table 3.4: Estimated Powers for c=2

Cochran's Q					**Chi-Square**				
	r=50	ρ				***r=50***	ρ		
Marginal Probabilities	δ	**0**	**0.4**	**0.6**	**Marginal Probabilities**	δ	**0**	**0.4**	**0.6**
(0.40,0.60)	**1**	0.526	0.727	0.909	(0.40,0.60)	**1**	0.539	0.547	0.557
(0.45,0.55)	**0.250**	0.197	0.244	0.310	(0.45,0.55)	**0.250**	0.208	0.138	0.082
(0.48,0.52)	**0.04**	0.089	0.089	0.096	(0.48,0.52)	**0.04**	0.102	0.030	0.011
	r=100	ρ				***r=100***	ρ		
Marginal Probabilities	δ	**0**	**0.4**	**0.6**	**Marginal Probabilities**	δ	**0**	**0.4**	**0.6**
(0.40,0.60)	**2.00**	0.793	0.956	0.999	(0.40,0.60)	**2.00**	0.804	0.881	0.928
(0.45,0.55)	**0.50**	0.299	0.448	0.613	(0.45,0.55)	**0.50**	0.326	0.269	0.215
(0.48,0.52)	**0.08**	0.109	0.123	0.147	(0.48,0.52)	**0.08**	0.124	0.045	0.018
	r=200	ρ				***r=200***	ρ		
Marginal Probabilities	δ	**0**	**0.4**	**0.6**	**Marginal Probabilities**	δ	**0**	**0.4**	**0.6**
(0.40,0.60)	**4.00**	0.977	1.000	1.000	(0.40,0.60)	**4.00**	0.980	0.998	1.000
(0.45,0.55)	**1.00**	0.496	0.723	0.897	(0.45,0.55)	**1.00**	0.500	0.516	0.511
(0.48,0.52)	**0.16**	0.121	0.152	0.230	(0.48,0.52)	**0.16**	0.118	0.062	0.032

Figure 3.1

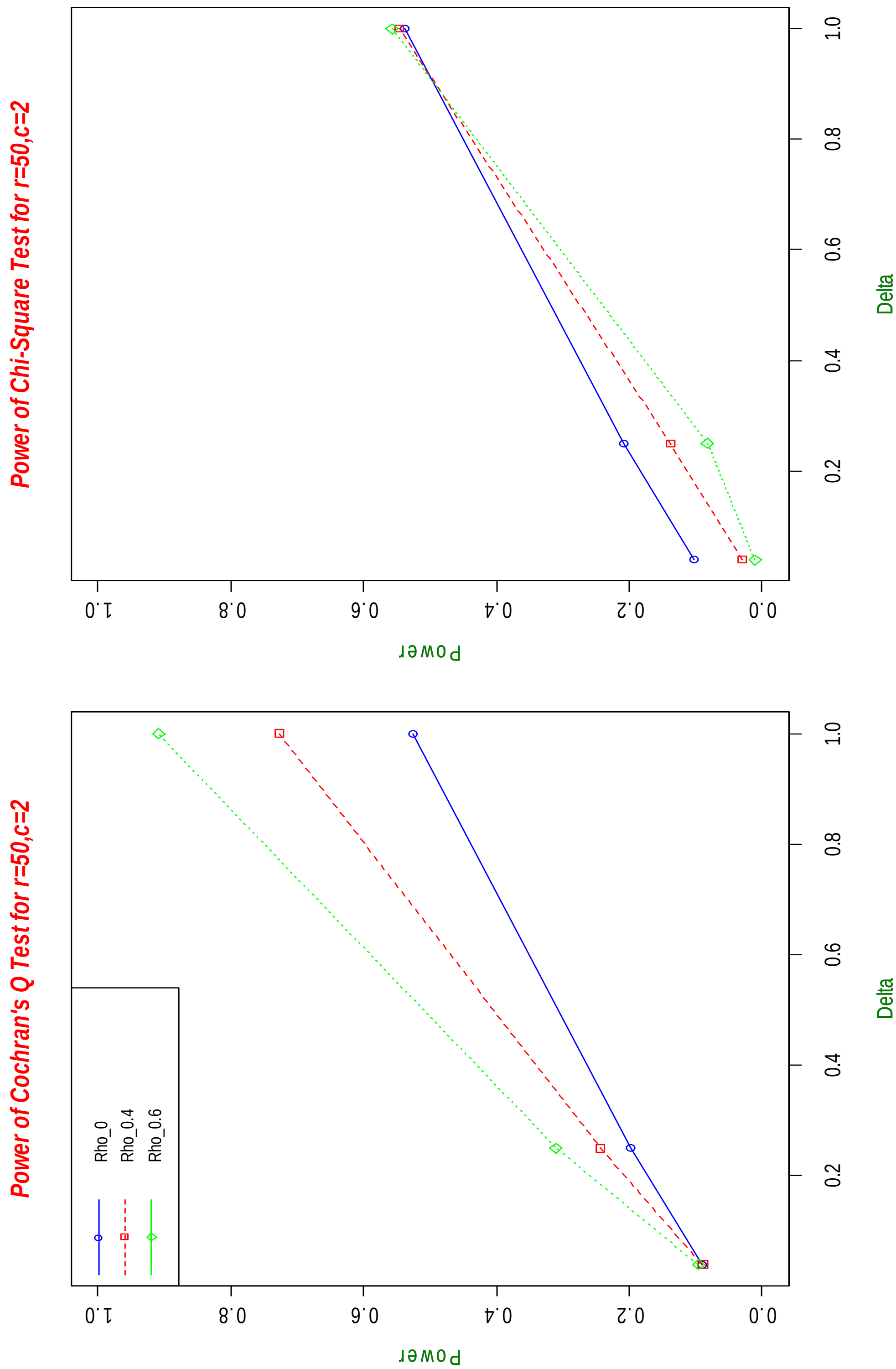
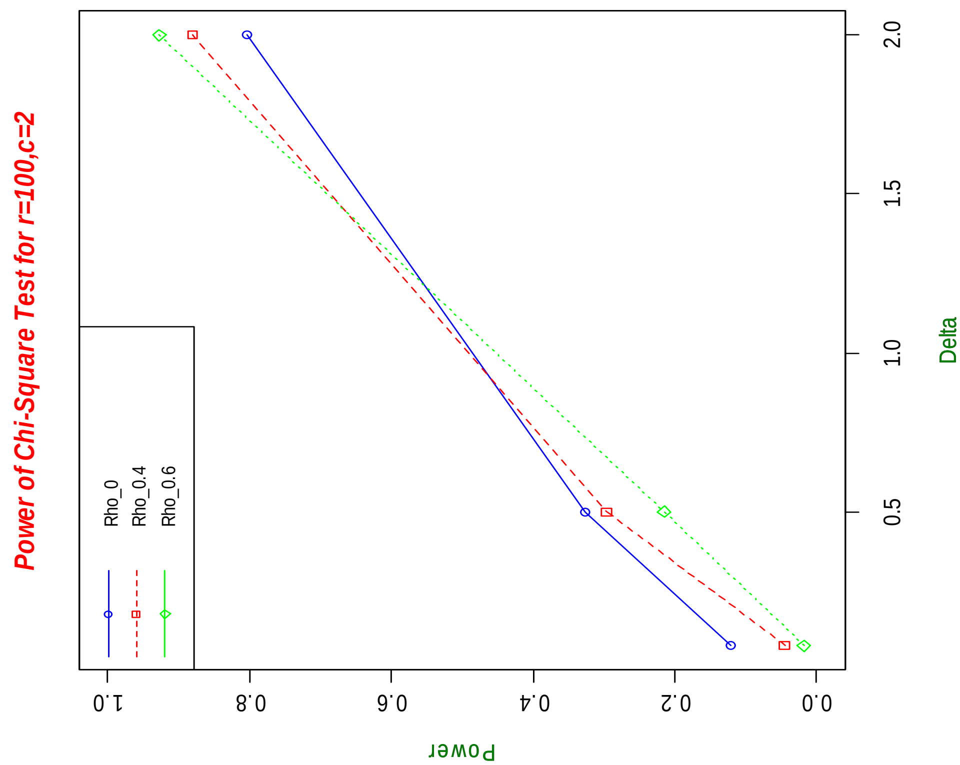

Power of Chi-Square Test for r=50,c=2
Power
Delta
0.0
0.2
0.4
0.6
0.8
1.0
Power of Cochran's Q Test for r=50,c=2
Power
Delta
Rho_0
Rho_0.4
Rho_0.6

Figure 3.2


Power of Chi-Square Test for r=100,c=2
Rho_0
Rho_0.4
Rho_0.6
Power
Delta
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0

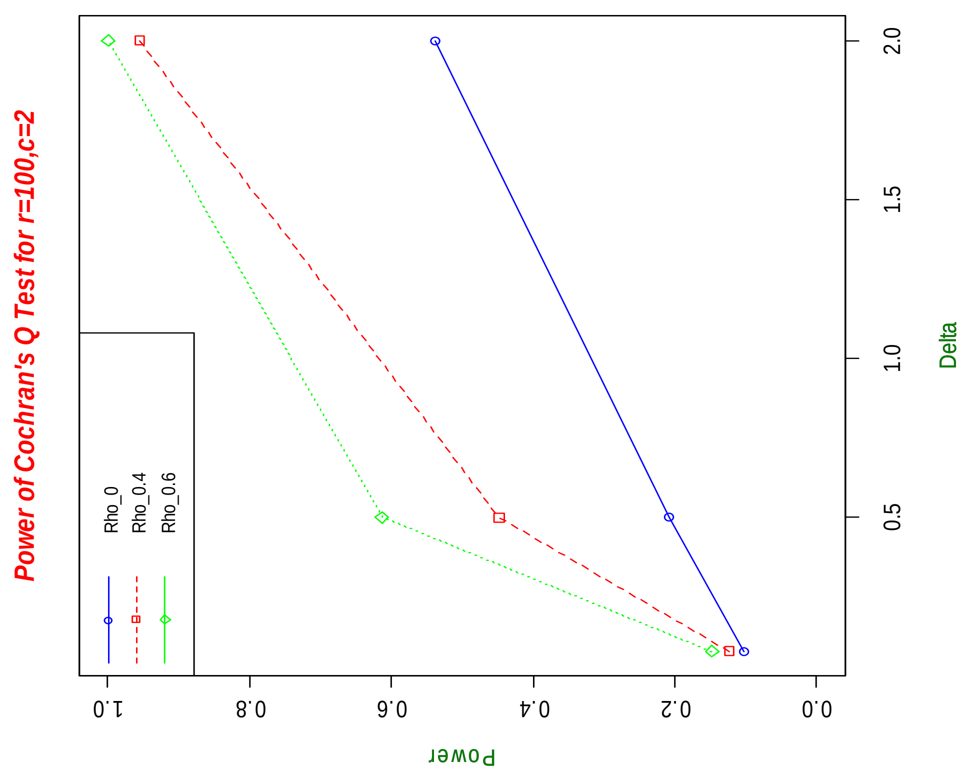
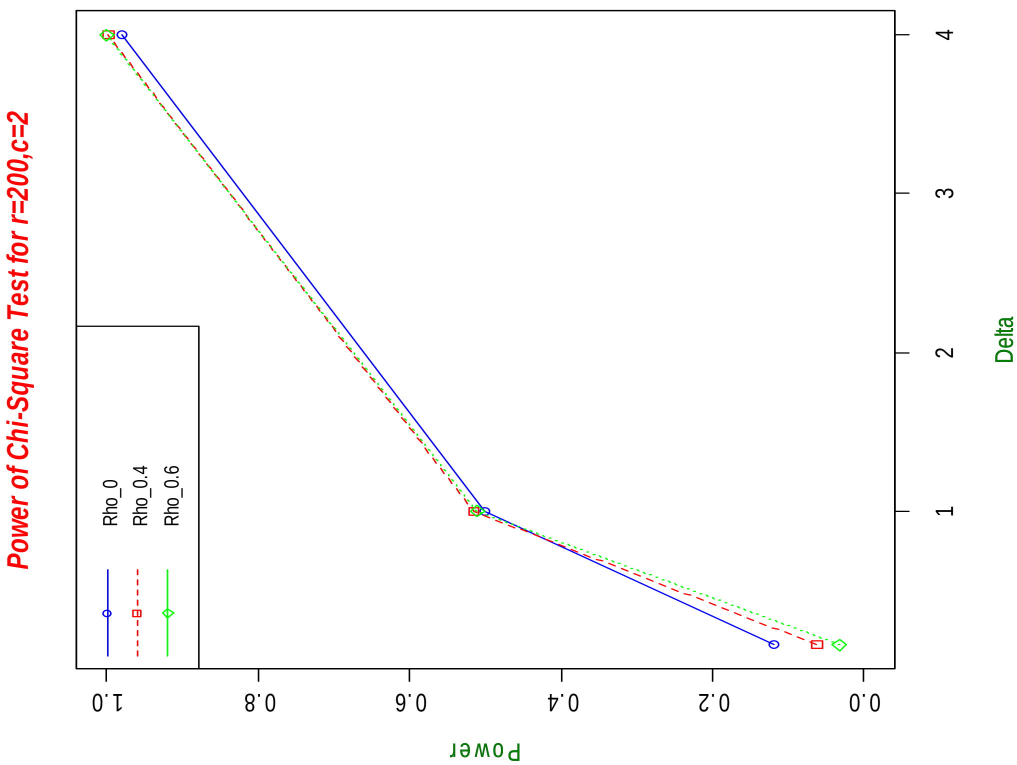
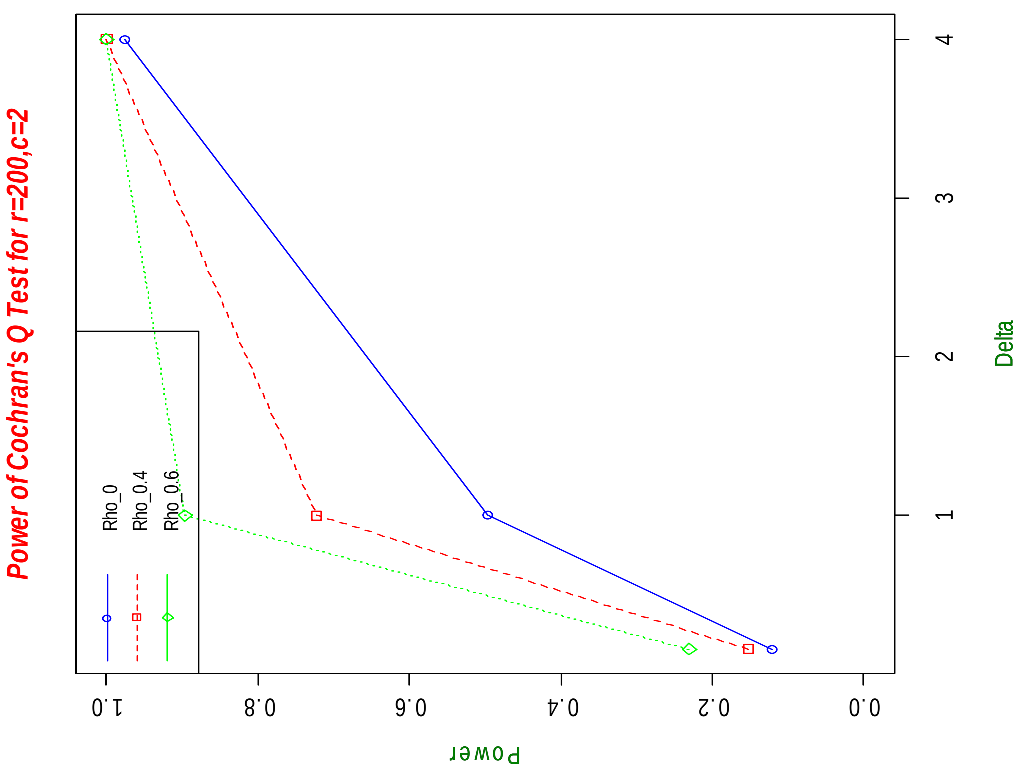

Power of Cochran's Q Test for r=100,c=2
Rho_0
Rho_0.4
Rho_0.6
Power
Delta
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0

Figure 3.3



Power of Chi-Square Test for r=200,c=2
Rho_0
Rho_0.4
Rho_0.6
Power
0.0
0.2
0.4
0.6
0.8
1.0
Delta
1
2
3
4



Power of Cochran's Q Test for r=200,c=2
Rho_0
Rho_0.4
Rho_0.6
Power
0.0
0.2
0.4
0.6
0.8
1.0
Delta
1
2
3
4

Table 3.5: P-values for Comparing Cochran's Q and The Standard Chi-square Tests

r=50				
Marginal Probabilities	δ	ρ		
		0	**0.4**	**0.6**
(0.40,0.60)	**1**	0.0279	<0.0001	<0.0001
(0.45,0.55)	**0.250**	0.0482	<0.0001	<0.0001
(0.48,0.52)	**0.04**	0.0286	<0.0001	<0.0001
r=100				
Marginal Probabilities	δ	ρ		
		0	**0.4**	**0.6**
(0.40,0.60)	**2.00**	0.0116	<0.0001	<0.0001
(0.45,0.55)	**0.50**	<0.0001	<0.0001	<0.0001
(0.48,0.52)	**0.08**	0.0023	<0.0001	<0.0001
r=200				
Marginal Probabilities	δ	ρ		
		0	**0.4**	**0.6**
(0.40,0.60)	**4.00**	0.0833	0.1573	NA[1]
(0.45,0.55)	**1.00**	0.4328	<0.0001	<0.0001
(0.48,0.52)	**0.16**	0.3657	<0.0001	<0.0001

[1] This NA indicates that the Q or McNemar's test statistics could not be calculated based on the data that was produced in Table 3.4 for r=200 and ρ =0.60.

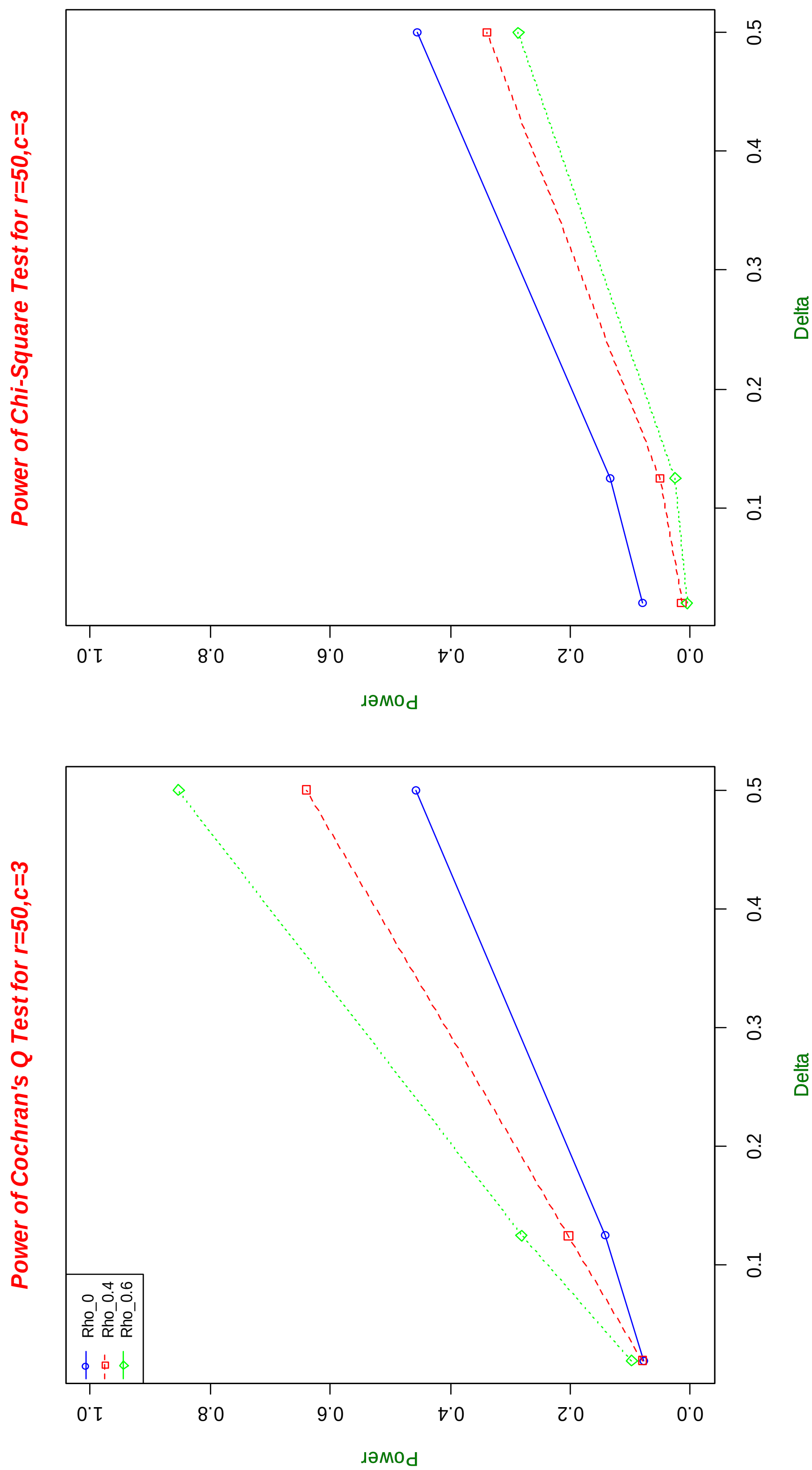
Three Environments: c=3

Tables 3.6 – 3.8 presents results for c=3. In Table 3.6 the estimated type I error rates are shown for the Cochran-Q and the standard Chi-square tests. The estimated type I error rates average 0.05 for all values of ρ under the Cochran-Q test. For the Chi-square test, the averages were 0.05, 0.008 and 0.0007 for $\rho = 0, 0.4, 0.6$ respectively. Again, the Cochran-Q performed similarly for all ρ while the Chi-square became increasingly conservative as ρ increased.

The estimated powers when c=3 are presented in Table 3.7. Figures 3.4-3.6 also compare the estimated powers for changes in δ and for different values of ρ. For the Cochran-Q test, the powers increased as δ increased. For all values of r, the higher values of ρ were associated with higher powers. The Chi-square tests also had higher powers for higher values of δ and as in the case of c=2, opposite to the behavior of the Cochran-Q test, higher values of ρ were associated with lower powers.

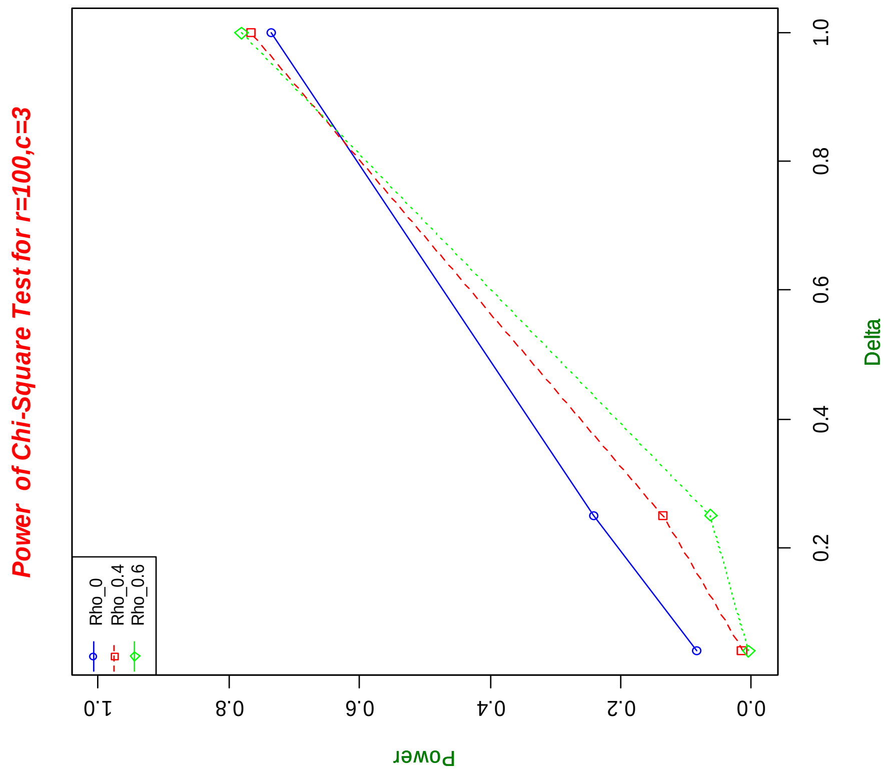
Interestingly, for both c=2 and c=3, with as expected, the exception of $\rho = 0$, the power of the Cochran-Q test is generally higher than that for the chi-square test. For r = 50 this was even more pronounced. In general the powers were below 0.50 for the chi-square. Again the p-values for difference between the two test indicates a statistically significant difference between the powers of the two tests when $\rho = 0.4$ and $\rho = 0.6$.

Figure 3.4



Power of Chi-Square Test for r=50,c=3
Power
Delta
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
Power of Cochran's Q Test for r=50,c=3
Rho_0
Rho_0.4
Rho_0.6
Power
Delta

Figure 3.5


Power of Chi-Square Test for r=100,c=3
Rho_0
Rho_0.4
Rho_0.6
Power
Delta
0.0
0.2
0.4
0.6
0.8
1.0

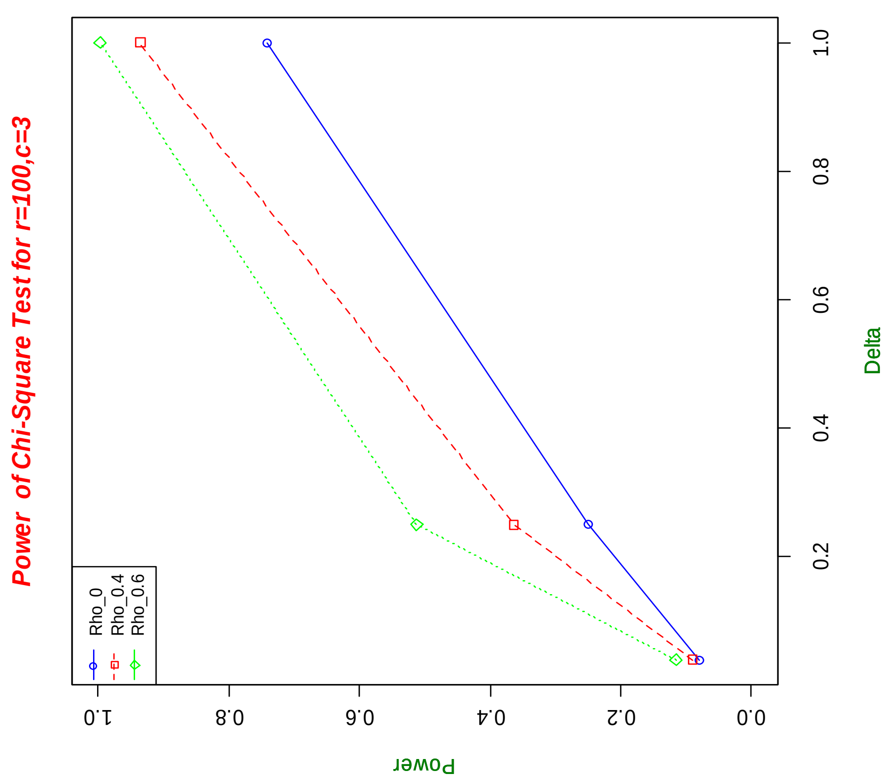
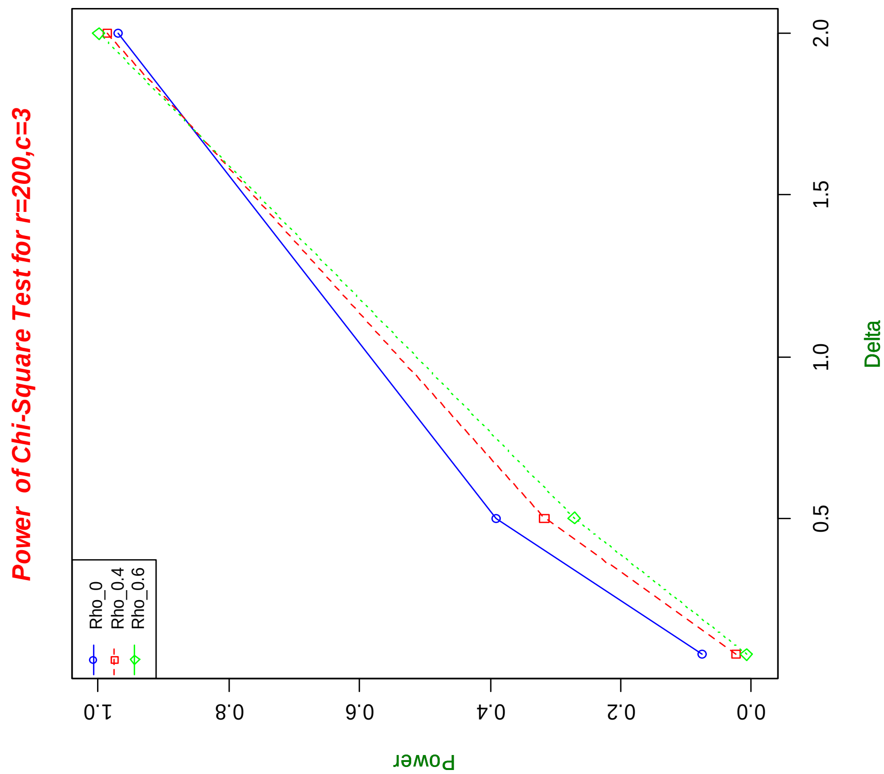
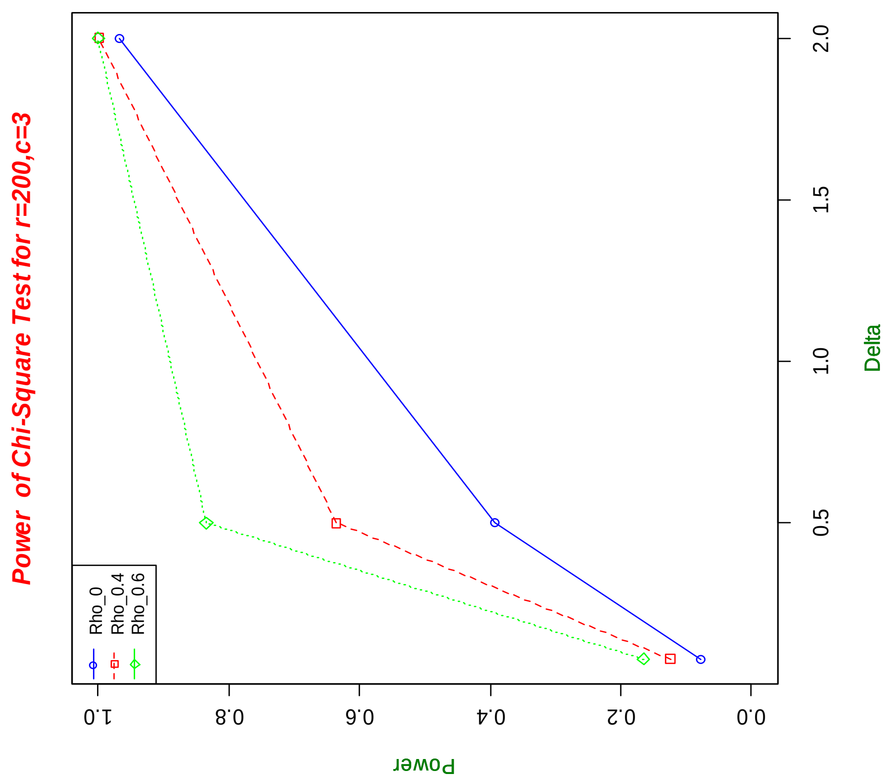

Power of Chi-Square Test for r=100,c=3
Rho_0
Rho_0.4
Rho_0.6
Power
Delta
0.0
0.2
0.4
0.6
0.8
1.0

Figure 3.6



Power of Chi-Square Test for r=200,c=3
Rho_0
Rho_0.4
Rho_0.6
Power
0.0
0.2
0.4
0.6
0.8
1.0
Delta
0.5
1.0
1.5
2.0



Power of Chi-Square Test for r=200,c=3
Rho_0
Rho_0.4
Rho_0.6
Power
0.0
0.2
0.4
0.6
0.8
1.0
Delta
0.5
1.0
1.5
2.0

Table 3.6: Estimated Type I Error Rates for c=3

Cochran's Q				**Chi-Square**			
r=50	ρ			***r=50***	ρ		
Marginal Probabilities	**0**	**0.4**	**0.6**	**Marginal Probabilities**	**0**	**0.4**	**0.6**
(0.4,0.4,0.4)	0.064	0.052	0.039	**(0.4,0.4,0.4)**	0.067	0.007	0.000
(0.5,0.5,0.5)	0.062	0.071	0.058	**(0.5,0.5,0.5)**	0.068	0.016	0.002
(0.6,0.6,0.6)	0.054	0.054	0.066	**(0.6,0.6,0.6)**	0.057	0.009	0.002
r=100	ρ			***r=100***	ρ		
Marginal Probabilities	**0**	**0.4**	**0.6**	**Marginal Probabilities**	**0**	**0.4**	**0.6**
(0.4,0.4,0.4)	0.065	0.055	0.045	**(0.4,0.4,0.4)**	0.069	0.009	0.000
(0.5,0.5,0.5)	0.059	0.060	0.056	**(0.5,0.5,0.5)**	0.062	0.010	0.000
(0.6,0.6,0.6)	0.057	0.057	0.060	**(0.6,0.6,0.6)**	0.056	0.007	0.002
r=200	ρ			***r=200***	ρ		
Marginal Probabilities	**0**	**0.4**	**0.6**	**Marginal Probabilities**	**0**	**0.4**	**0.6**
(0.4,0.4,0.4)	0.048	0.046	0.045	**(0.4,0.4,0.4)**	0.043	0.007	0.000
(0.5,0.5,0.5)	0.033	0.040	0.040	**(0.5,0.5,0.5)**	0.034	0.006	0.001
(0.6,0.6,0.6)	0.036	0.037	0.045	**(0.6,0.6,0.6)**	0.033	0.004	0.000

Table 3.7: Estimated Powers for c=3

Cochran's Q					Chi-Square				
	r=50	ρ				***r=50***	ρ		
Marginal Probabilities	δ	**0**	**0.4**	**0.6**	**Marginal Probabilities**	δ	**0**	**0.4**	**0.6**
(0.40,0.50,0.60)	**0.500**	0.458	0.640	0.853	**(0.40,0.50,0.60)**	**0.500**	0.456	0.339	0.287
(0.45,0.50,0.55)	**0.125**	0.141	0.203	0.281	**(0.45,0.50,0.55)**	**0.125**	0.134	0.051	0.025
(0.48,0.50,0.52)	**0.020**	0.078	0.079	0.098	**(0.48,0.50,0.52)**	**0.020**	0.081	0.014	0.005
	r=100	ρ				***r=100***			
Marginal Probabilities	δ	**0**	**0.4**	**0.6**	**Marginal Probabilities**	δ	**0**	**0.4**	**0.6**
(0.40,0.50,0.60)	**1.000**	0.741	0.936	0.997	**(0.40,0.50,0.60)**	**1.000**	0.736	0.766	0.781
(0.45,0.50,0.55)	**0.250**	0.249	0.364	0.513	**(0.45,0.50,0.55)**	**0.250**	0.241	0.136	0.062
(0.48,0.50,0.52)	**0.400**	0.079	0.090	0.115	**(0.48,0.50,0.52)**	**0.400**	0.083	0.014	0.004
	r=200	ρ				***r=200***			
Marginal Probabilities	δ	**0**	**0.4**	**0.6**	**Marginal Probabilities**	δ	**0**	**0.4**	**0.6**
(0.40,0.50,0.60)	**2.000**	0.967	0.999	1.000	**(0.40,0.50,0.60)**	**2.000**	0.969	0.987	0.999
(0.45,0.50,0.55)	**0.500**	0.393	0.636	0.835	**(0.45,0.50,0.55)**	**0.500**	0.391	0.317	0.271
(0.48,0.5,0.52)	**0.080**	0.077	0.124	0.165	**(0.48,0.5,0.52)**	**0.080**	0.076	0.023	0.007

Table 3.8: P-values for Comparing Cochran's Q and The Standard Chi-square Tests

r=50				
Marginal Probabilities	δ	ρ		
		0	**0.40**	**0.60**
(0.40,0.50,0.60)	**0.500**	0.7237	<0.0001	<0.0001
(0.45,0.50,0.55)	**0.125**	0.0897	<0.0001	<0.0001
(0.48,0.50,0.52)	**0.020**	0.6171	<0.0001	<0.0001
r=100				
Marginal Probabilities	δ	ρ		
		0	**0.40**	**0.60**
(0.40,0.50,0.60)	**1.000**	0.1317	<0.0001	<0.0001
(0.45,0.50,0.55)	**0.250**	0.1025	<0.0001	<0.0001
(0.48,0.50,0.52)	**0.400**	0.2852	<0.0001	<0.0001
r=200				
Marginal Probabilities	δ	ρ		
		0	**0.40**	**0.60**
(0.40,0.50,0.60)	**2.000**	0.3173	0.0005	0.3173
(0.45,0.50,0.55)	**0.500**	0.6171	<0.0001	<0.0001
(0.48,0.50,0.52)	**0.080**	0.6547	<0.0001	<0.0001

CHAPTER 4 - Conclusion

The aim of this study was to assess the effect of using the inappropriate standard Chi-square test instead of the Cochran-Q test when working with correlated binary data. The result of the study revealed that there is a statistically significant difference in the powers of the two tests, with the Cochran-Q test being the more powerful of the two.

This result also holds regardless of the sample size. The power of the tests was affected by the correlation structure of the data. Specifically, the higher the correlation, the higher the power under Cochran-Q, while the power was lower for higher correlations under the Chi-square test. Thus, Cochran's Q and not the standard Chi-square should be used to compare correlated proportions when compound symmetry holds.

The study focused on rates of success centered on 0.50 and so for further study it would be interesting to how the results might differ if values closer to say 0.20 or 0.80 were considered.

The issue of compound symmetry was raised in discussion of the assumptions of the Cochran-Q test. It would be interesting to consider data in which the pair wise correlations are not all equal.

References

Bhapkar, Vasant P., and Somes, Grant W. (1977), "Distribution of Q When Testing Equality of Matched Proportions," *Journal of the American Statistical Association*, 72, 658-661.

Berger, Agnes, and Gold, Ruth (1973), "Note on Cochran's Q Test for the Comparison of Correlated Proportions," *Journal of the American Statistical Association*, 68, 989-993.

Cochran,W.G. (1950). *"The comparison of percentages in matched samples". Biometrika*, 37, 256-266.

Conover, W.J. (1980). Practical Nonparametric Statistics, 2nd ed. John Wiley & Sons Inc

Madansky, A., (1963) "Tests of Homogeneity for Correlated Samples," *Journal of the American Statistical Association*, 58 ,97-119

McNemar, Quinn (1947). "Note on the sampling error of the difference between correlated proportions or percentages". *Psychometrika* 12 (2): 153-157.

Vitaliano, Peter P. (1979), "The Relative Merits of Two Statistics for Testing A Hypothesis about the Equality of Three Correlated Proportions," *Journal of the American Statistical Association*, 74, 232-237.

Wallenstein, Sylvan and Agnes Berger (1981), " On the Asymptotic Power of tests for comparing K Correlated Proportions," *Journal of the American Statistical Association*, 76, 114-118.

Appendix A - Bindata Package

Source: http://cran.r-project.org/web/packages/bindata/bindata.pdf

Package ‘bindata’ - November 22, 2009

Version 0.9-17

Date 2009-11-22

Title Generation of Artificial Binary Data

Author Friedrich Leisch and Andreas Weingessel and Kurt Hornik

Maintainer Friedrich Leisch Friedrich.Leisch@R-project.org

Description Generation of correlated artificial binary data.

License GPL-2

Depends e1071, mvtnorm (>= 0.7-0)

Repository CRAN

Date/Publication 2009-11-22 19:06:36

rmvbin Multivariate Binary Random Variates

Description

Creates correlated multivariate binary random variables by thresholding a normal distribution. The correlations of the components can be specified either as common probabilities, correlation matrix of the binary distribution, or covariance matrix of the normal distribution.

Usage

rmvbin(n, margprob, commonprob=diag(margprob),
bincorr=diag(length(margprob)),
sigma=diag(length(margprob)),
colnames=NULL, simulvals=NULL)

Arguments

n number of observations.

margprob	margin probabilities that the components are 1.
commonprob	matrix of probabilities that components i and j are simultaneously 1.
bincorr	matrix of binary correlations.
sigma	covariance matrix for the normal distribution.
colnames	vector of column names for the resulting observation matrix.
simulvals	result from simul.commonprob, a default data array is automatically loaded if this argument is omitted.

Details

Only one of the arguments commonprob, bincorr and sigma may be specified. Default are uncorrelated components.

n samples from a multivariate normal distribution with mean and variance chosen in order to get the desired margin and common probabilities are sampled. Negative values are converted to 0, positive values to 1.

Author(s)

Friedrich Leisch

References

Friedrich Leisch, Andreas Weingessel and Kurt Hornik (1998). On the generation of correlated artificial binary data. Working Paper Series, SFB “Adaptive Information Systems and Modeling in Economics and Management Science”, Vienna University of Economics, http://www.wu-wien.ac.at/am

Appendix B - R Code

```
#function for Cochran-Q test
cochranq.test <- function(mat)
{
  k <- ncol(mat)
  C <- sum(colSums(mat) ^ 2)
  R <- sum(rowSums(mat) ^ 2)
  T <- sum(rowSums(mat))
  num <- (k - 1) * ((k * C) - (T ^ 2))
  den <- (k * T) - R
  Q <- num / den
  df <- k - 1
  names(df) <- "df"
  names(Q) <- "Cochran's Q"
  p.val <- pchisq(Q, df, lower = FALSE)
  QVAL <- list(statistic = Q, parameter = df, p.value = p.val,
        method = "Cochran's Q Test for Dependent Samples",
        data.name = deparse(substitute(mat)))
  class(QVAL) <- "htest"
  return(QVAL)
}
########################################################################
library(bindata)          # code requires the 'Bindata' package
#next I specify the marginal probabilities and correlation structure
margprob<-c(0.40, 0.50, 0.60)
rho<-cbind(c(1,0,0),c(0,1,0),c(0,0,1))

M=1000        #number of simulations
r=50          #number of rows in binary dataset
########################################################################
```

```
#Initialization
pval.cochran=rep(NA,M)
pval.cochran.1=rep(NA,M)
pval.cochran.2=rep(NA,M)
pval.cochran.3=rep(NA,M)
ind.cochran=rep(NA,M)
ind.cochran.1=rep(NA,M)
ind.cochran.2=rep(NA,M)
ind.cochran.3=rep(NA,M)
pval.chisquare=rep(NA,M)
pval.chisquare.1=rep(NA,M)
pval.chisquare.2=rep(NA,M)
pval.chisquare.3=rep(NA,M)
ind.chisquare=rep(NA,M)
ind.chisquare.1=rep(NA,M)
ind.chisquare.2=rep(NA,M)
ind.chisquare.3=rep(NA,M)
##########################################################################
ptm <- proc.time()
for (i in 1:M)
{
##############################################
simdata<-matrix(NA,nrow=r,ncol=3)
# this next for loop generates r independent rows and 3 dependent columns and saves as
'simdata'

for (j in 1:r)
{
set.seed(i*r+j) #did not use set.seed(i) b/c I would end up with the same
        # binary dataset for each simulation
```

```
simdata[j,]<-rmvbin(1,margprob=margprob,bincorr=rho)
}

##############################################
#the next set of steps stores p-values from Cochran-Q test
pval.cochran[i]<-cochranq.test(simdata)$p.value
if(pval.cochran[i]<0.05)
{ind.cochran[i]=1}
else {ind.cochran[i]=0}

##############################################
#the next set of steps stores p-values from Standard Chisquare test
#chisquare test
a=sum(simdata[,1])
b=sum(simdata[,2])
c=sum(simdata[,3])
pval.chisquare[i]<- prop.test(x = c(a,b,c), n= c(r,r,r), correct = FALSE)$p.value
if(pval.chisquare[i]<0.05)
{ind.chisquare[i]=1}
else {ind.chisquare[i]=0}




##############################################
sum(ind.chisquare)/M
sum(ind.cochran)/M
###############################################
#Next compare the success rates in the two test to see if their
#Differences are statistically significant
y<-matrix(NA,M,2)
```

```
y[,1] <- ind.cochran
y[,2 ] <- ind.chisquare
data <- data.frame(x1=y[,1], x2=y[,2])
cochranq.test(data)
proc.time() - ptm
```