# CONFIDENCE INTERVALS FOR POPULATION SIZE BASED ON A CAPTURE-RECAPTURE DESIGN

by

JIANJUN HUA

---

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
2011

Approved by:

Major Professor
Paul Nelson

# Copyright



**This document is presented using LaTeX**

# Abstract

Capture-Recaputre (CR) experiments stemmed from the study of wildlife and are widely used in areas such as ecology, epidemiology, evaluation of census undercounts, and software testing, to estimate population size, survival rate, and other population parameters. The basic idea of the design is to use "overlapping" information contained in multiple samples from the population. In this report, we focus on the simplest form of Capture-Recapture experiments, namely, a two-sample Capture-Recapture design, which is conventionally called the "Petersen Method."

We study and compare the performance of three methods of constructing confidence intervals for the population size based on a Capture-Recapture design, asymptotic normality estimation, Chapman estimation, and "inverting a $\chi^2$ test" estimation, in terms of coverage rate and mean interval width. Simulation studies are carried out and analyzed using R and SAS. It turns out that the "inverting a $\chi^2$ test" estimation is better than the other two methods. A possible solution to the "zero recapture" problem is put forward. We find that if population size is at least a few thousand, two-sample CR estimation provides reasonable estimates of the population size.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgments

This research report would not have been possible without the support of many people. The author owes his deepest gratitude to his supervisor, Dr. Paul Nelson whose abundant help, invaluable assistance, support and guidance from the initial to the final level enabled the author to develop an understanding of the subject and tide the author over the tough time in his life. The author is very grateful to the members of the supervisory committee, Dr. Leigh Murray, and Dr. Haiyan Wang, without whose knowledge and assistance this study would not have been successful. The author would also like to convey wholehearted thanks to the Department Head, Dr. James Neill for his understanding and providing the financial aids in the past two and a half years. Special thanks also to all other faculty, especially Dr. Juan Du for her invaluable advice and suggestion, and the author's peer graduate students. Not forgetting to his best friends who always been there. The author wishes to express his love and gratitude to his beloved families; for their understanding and endless love, through the duration of his studies.

Lastly, the author offers his regards and blessings to all of those who supported him in any respect during the completion of the research.

# Chapter 1

# Introduction

## 1.1 The Capture-Recapture Method of Estimating the Size of a Population

### 1.1.1 A General Procedure for a CR Design

How many endangered or threatened animals, such as Giant Pandas, Bald Eagles, whales, polar bears, etc., are still living in their natural habitats? How many fish are there in Tuttle Creek Reservoir? How many people have type II diabetes? How many faults are there in a newly-developed software? These are but few instances of the important problem of estimating a population size $N$ when a census cannot be taken. In its simplest form, a Capture-Recapture(CR) experiment randomly selects $n_1$ units from the population and *marks* them. A captured fish, for example, could be marked by placing a tag on its tail. The tagged units are then returned to the population and assumed to be in their original conditions. A second random sample of size $n_2$ is then selected, resulting in $x$ marked units. An estimate of the population size $N$ can be obtained from $(n_1, n_2, x)$. The detailed estimation procedure and underlying assumptions are described in Chapter 2.

### 1.1.2 Giant Pandas and Bald Eagles:Endangered and Threatened Species

The Giant Panda is among the world's most threatened animals. There are probably only about 1,000 Giant Pandas [*] left in the wild and they are classified as an endangered species. Giant Pandas used to range throughout southern and eastern China, Myanmar, and the northern part of Vietnam [†]. Habitat loss caused by logging is the largest threat to Giant Pandas today. Now, they are found only in a small part of China, in mountainous bamboo forests of southwestern China. Giant Pandas also live in zoos in China, the United States, Mexico, Japan, Germany, and North Korea.

Another well known "threatened species" is the Bald Eagle, the national emblem of the United States. The Bald Eagle was threatened with extinction in the US because of DDT (a type of pesticide) poisoning [‡]. Protection under the Endangered Species Act, together with reintroduction programs, increased eagle populations, and the species was reclassified as "Threatened" in 1995. They were finally removed from the list in June 2007, a true conservation success story [§].

According to the definitions given by Wikipedia [¶], an endangered species is a population which is at risk of becoming extinct because it is either few in numbers, or threatened by changing environmental or predation parameters. A threatened species [‖] is one which is vulnerable to endangerment in the near future. Giant Pandas and Bald Eagles are just but two pronounced examples of endangered and threatened species which are under protection.

In the United States alone, the "known species threatened with extinction is ten times higher than the number protected under the Endangered Species Act" [14]. The US Fish and Wildlife Service as well as the National Marine Fisheries Service are held responsible for classifying and protecting endangered species. Yet, adding a particular species to the list is

[*] http://animals.nationalgeographic.com/animals/mammals/giant-panda/
[†] http://www.bearsoftheworld.net/giant_panda_bears.asp
[‡] http://animals.nationalgeographic.com/birds/bald-eagle/
[§] http://www.worldwildlife.org/who/media/press/2007/WWFPresitem974.html
[¶] http://en.wikipedia.org/wiki/Endangered_species
[‖] http://en.wikipedia.org/wiki/Threatened_species

a long, controversial process and in reality it represents only a fraction of imperiled plant and animal life[14].

Taking timely actions to protect endangered species and remove them from the list is imperative for humans because the loss in the ecological environment may eventually adversely impact mankind's life on the earth. Therefore, accurately estimating the population sizes of endangered species is crucial before effective and practical protection measures can be taken.

Now, a question arises: what data can be used to support the claim that a species is endangered or threatened?

A direct measure of a population size, i.e, a census, is frequently not doable due to geographical and economical constraints. In such cases, we can apply some statistical methodologies to estimate the population size. There are several such statistical methods. One of them is the Capture-recapture method, which is the focus of this report.

### 1.1.3 The Definition of Capture-Recapture Method

The basic idea of the Capture-Recapture (CR) Method is to use the "overlap information" contained in different samples from a population to estimate some characteristics of a population, such as the population size, survival rate, etc. A detailed explanation of the CR method will be given in Chapter 2. Here, an outline for the history and relevant literature are briefly reviewed, as follows.

### 1.1.4 Historical Background and Literature Review

CR experiments and methods of analysis (also called Petersen Methods when only 2 samples are used) have a long history. A rudimentary CR analysis was carried out by Laplace in 1783, who tried to estimate the French population. A register of births for the whole country (the "marked" individuals) was his first sample. His second "sample" consisted of a number of parishes of known total population size, and the "overlap" was the number of births recorded for these parishes[10,13].

The original application of a CR method was in the study of wildlife management[13], just like the ones used to estimate the population sizes of Giant Pandas and Bald Eagles.

Later, in 1940, Green and Evans[9] studied the fluctuations in the abundance of Snowshoe Hares in the area around Lake Alexander, Minnesota, from 1932 to 1939, using a series of Petersen estimates. In 1961, Odum and Pontin[11]employed radioactive $P^{32}$ to tag ants in a colony and used the Petersen method to estimate the number of ants in the colony.

Nowadays, applications of CR experiments can be found in a wide spectrum of fields, such as the evaluation of census undercounts[6], epidemiology[1], software testing[3,7,16], and many applications in ecology[2]. Different researchers often use different names for the same method, such as mark-recapture, capture-mark-recapture, sight-resight, mark-release-recapture, multiple systems estimation and band recovery, etc.

1n 1993, Darroch[6] and his colleagues used a capture-recapture method to estimate the census undercount, which involved a second independent sample (besides the census data) or alternative data source to be matched with census and Post-Enumeration Survey (PES) data. This kind of CR is also called the dual-system methodology, which is largely based on three assumptions[6]:

(1) Perfect matching: Individuals in the second list (the PES) can be matched with those in the first list (the census) without error.

(2) Independence of lists: The probability of an individual being included in the first list does not depend on whether he or she was included in the second list.

(3) Homogeneity: The probability of inclusion on a list does not vary from individual to individual.

Actually, the second and third assumptions are related. Suppose that the lists are independent within strata, namely some interesting subsets of the lists, but the probability of capture or inclusion varies across strata. When the strata are combined, independence will generally no longer exist in the resulting data.

Abeni and his colleagues[1] used CR methods to complement estimates of the size of the population having HIV-1 infection in Lazio, Italy, during 1990 obtained from surveillance, surveys, and dynamic mathematical models. They used reports from four large testing sites to generate incomplete, partially overlapping lists of HIV-positive subjects. They adopted log-linear models to estimate prevalence of this disease. It turned out that this method provides a simple and inexpensive means of obtaining accurate estimates of the total number of HIV seropositives. It could be applied easily in all situations in which data from multiple sources is available.

CR methods have also evolved as promising methods in software testing. In 1995, Wohlin[16]used capture-recapture to estimate the residual faults. Combined with a filtering technique, he found that CR estimations provide an important method to perform fault content estimations throughout the software life cycle, which helps decision makers accept documents in software development.

In 2000, Ebrahimi[7] and coworkers noticed that sometimes the failure of complex software systems is due to the faults introduced in the requirements and design stages of the development process. Using different reviews cannot guarantee the complete removal of the faults until the software is developed. In their paper, they propose a procedure based on a CR design to estimate the number of faults which are not discovered. The main advantage of their procedure is that the independence among reviewers is not necessary. Briand and colleagues[3] also applied CR methods to software testing.

## 1.2 Our Goal

In this report, simulations will be used to study and compare the performance of interval estimates of the size of a population based on a CR experiment. Although there is an extensive literature on the CR method, we have not found any references to studies investigating the performance of interval estimators of the size of a population.

## 1.3 Outline of This Report

In the following chapters, I present the framework and assumptions underlying the analysis of CR experiments (Chapter 2). Then, I explain the simulation methods adopted to construct confidence intervals and compare those methods(Chapter 3). Finally, I sum up my results and propose some potential extensions and expansions of this project (Chapter 4). In the Appendix, I attach the codes written in R and SAS which were used to carry out the simulations.

# Chapter 2

# Methods for Constructing Confidence Intervals for the Size of a Population Based on a Capture-Recapture Experiment

## 2.1 Underlying Assumptions

The simplest CR method is the so-called two-sample model, which is the model adopted in this project, used solely to estimate the unknown population size $N$[10,13]. For the simplest CR model, $x$, the number of marked items in the second sample, has the hypergeometric distribution given in Eq. 2.1. Assumptions used in this report are as follows.

(i) Simple random sampling without replacement (SRSWOR) is used at each stage (but the entire sample is replaced after marking).

(ii) There is no change to the population during the investigation. In other words, the population is closed.

(iii) For each sample, each individual has the same chance of being sampled.

(iv) The two samples are independent.

(v) There is no loss of marks (individuals can be matched from capture to recapture).

(vi) Marking does not affect the catchability of any individual.

Assumption (iv) really follows from (iii) since the latter implies that marked and unmarked items have the same probability of being caught in the second sample so that capture in the first sample does not affect capture in the second sample. However, most of the time, researchers list them separately. When assumptions (i), (ii), (iii), and (v) are satisfied, the conditional distribution of $x$ given $n_1$ and $n_2$, is the hypergeometric distribution and $N$ is the size of the population, $n_1$ is the size of the first sample, $n_2$ is the size of the second sample.

$$f(x|n_1, n_2) = \frac{\binom{n_1}{x}\binom{N-n_1}{n_2-x}}{\binom{N}{n_2}}, \tag{2.1}$$

where $max\{0, n_1 + n_2 - N\} \leq x \leq min(n_1, n_2)$.

Hence, $E(\frac{x}{n_2}) = \frac{n_1}{N}$, the proportion of marked items in the population. A method of moments of point estimate of $N$ can then be obtained by setting

$$\frac{x}{n_2} = \frac{n_1}{N},$$

and solving for $N$, yielding

$$\hat{N} = \frac{n_1 n_2}{x}, \tag{2.2}$$

and its corresponding variance[10]

$$\hat{V}(\hat{N}) = \left(\frac{n_1 n_2}{x}\right)^2 \frac{n_2 - x}{x(n_2 - 1)} \approx \frac{n_1^2 n_2 (n_2 - x)}{x^3}, \tag{2.3}$$

for $x > 0$.

In this simple form, capture-recapture is a special case of ratio estimation of a population total. To verify this, let

$$y_i = 1 \text{ for every individual in the population}$$

and

$$x_i = \begin{cases} 1 & \text{if individual } i \text{ is marked} \\ 0 & \text{if individual } i \text{ is not marked.} \end{cases}$$

The ratio estimate of $N = t_y = \sum_{i=1}^{N} y_i$ is $\hat{t}_{yr} = t_x \hat{B} = \hat{N}$, where $t_x = \sum_{i=1}^{N} x_i = n_1$ and $\hat{B} = \frac{\bar{y}}{\bar{x}} = \frac{n_2}{x}$. Note that for $x > 0$,

$$\hat{N} = \hat{t}_{yr} = \frac{n_1 n_2}{x},$$

is also approximately the maximum likelihood estimator[10].

The properties of $\hat{N}$ have been fully discussed by Chapman[4]. He showed that although $\hat{N}$ is a best asymptotically normal estimate of $N$ as $N \to \infty$, it is biased, and the bias can be large for small samples. However, when $n_1 + n_2 \geq N$, his modified estimate

$$\tilde{N} = \frac{(n_1 + 1)(n_2 + 1)}{(x + 1)} - 1 \tag{2.4}$$

is exactly unbiased. The variance for the unbiased estimate $\tilde{N}$ was given by Seber[12] and Wittes[15], namely

$$\hat{V}(\tilde{N}) = \frac{(n_1 + 1)(n_2 + 1)(n_1 - x)(n_2 - x)}{(x + 1)^2 (x + 2)}. \tag{2.5}$$

Although $n_1 + n_2 \geq N$ is very restrictive condition for unbiasedness, in reality, the use of Eq. 2.4 and Eq. 2.5 has some flexibility: Researchers apply them to their CR studies even if $n_1 + n_2 \leq N$[10]. Specifically, in our simulation study, the Chapman estimation method is utilized as a substitute for Eq. 2.2 and Eq. 2.3 whenever the recapture $x = 0$.

## 2.2 Some Methods for Constructing Confidence Sets for the Size of a Population

We consider several ways to construct confidence sets for the unknown population size $N$.

(1) Asymptotic normality[10,13]: $\hat{N} \pm z_{\alpha/2}\sqrt{\hat{V}(\hat{N})}$, where $\hat{N}$ (Eq.2.2) is the *mle* or $\tilde{N} \pm z_{\alpha/2}\sqrt{\hat{V}(\tilde{N})}$, where $\tilde{N}$ (Eq.2.4) is the Chapman's estimator. Their corresponding variances are given Eq.2.3 and Eq.2.5.

(2) Inverting a test for $N$[10]. Consider testing $H_0 : N = N_0$ vs $H_0 : N \neq N_0$. Suppose that $X = x$ tagged fish have been observed in the second example so that failing to reject

$H_0$ leads to an exact type I error rate $1-\alpha$ test for all $N_0$ if $T(x, N_0)$ lies in the region $A(N_0)$ . Then, having observed $X = x$, $C(x) = \{N\colon T(x, N) \in A(N)\}$ is a $1-\alpha$ confidence set for $N$.

We use an approximate version of this approach. Let $T = \chi^2(N) \equiv \chi^2(x_{22}^*)$ be the $\chi^2$ test statistic for testing for independence in the two-way contingency Table 2.1, where $N = x_{11} + x_{12} + x_{21} + x_{22}$, $x_{11} \equiv x$ denotes the number of marked units caught in the second sample and $x_{22}$ is the number of units observed in neither sample and is unknown. By guessing $x_{22}$ using a particular value $x_{22}^*$, $x_{2+}$ can be obtained using $x_{21} + x_{22}^*$ and $x_{+2}$ can be obtained using $x_{12} + x_{22}^*$, in turn, $x_{++}$ can be obtained using either $x_{1+} + x_{2+}$ or $x_{+1} + x_{+2}$. Then, a $p$-value can be obtained from a $\chi^2$ test using $x_{22} = x_{22}^*$. If the p-value were greater than a predetermined significance level, say, 0.05, then the corresponding value of $x_{++}$ would be placed inside the 95% CI for $N$[10]. Specifically, the confidence set contains all values of $x_{++}(= N)$ for which p-value $> 0.05$.

**Table 2.1**: Capture-Recapture Data In a Two-way Contingency Table

| | | In Sample 2? | | |
|---|---|---|---|---|
| | | Yes | No | |
| In Sample 1? | Yes | $x_{11}(= x)$ | $x_{12}$ | $x_{1+}(= n_1)$ |
| | No | $x_{21}$ | $x_{22}$ | $x_{2+}$ |
| | | $x_{+1}(= n_2)$ | $x_{+2}$ | $x_{++}(= N)$ |

## 2.3 Two Scenarios

Actually, there are at least two probability models used by CR researchers:

1. $n_1$ and $n_2$ are fixed, but $x$ is the observed value of random variable $X$.

2. $n_1$, $n_2$ and $x$ are all observed values of random variables.

To simplify the analysis, the first scenario was adopted here. In reality, either by design or acting conditionally, $n_1$ and $n_2$ are often considered as fixed resulting in $X$ having the hypergeometric distribution (cf.Eq. 2.1).

In the next chapter, I will explain how simulations were carried out to implement the three methods for constructing confidence sets for $N$ and compare their performances, and how to overcome some potential problems of each method.

# Chapter 3

# Simulation Studies

In this chapter, my simulation experiment is described. Tables and figures are used to illustrate comparisons for the approaches adopted to construct confidence intervals for $N$ based on a capture-recapture design. Cochran's test[5,8] is employed to compare the coverage rates of nominal .95 confidence sets constructed using the methods. Solutions to the potential problems in the simulations are provided at the end of this chapter.

## 3.1 Simulations

### 3.1.1 Basic Procedure

Simulation studies were carried out and visualized using R and SAS to evaluate the performance in terms of actual coverage rates and mean width of the three CR methods of constructing confidence intervals for the population size $N$: asymptotic normality, Chapman, "inverting $\chi^2$ test", based on the data $(n_1, n_2, x)$ . The three methods are denoted "Nh", "Nt", and "Nchisq" respectively, in the following tables and figures. The simulations are based on the underlying assumptions for a CR design which were stated in Chapter 2.

Two simple random samples ($n_1$ and $n_2$) without replacement (SRSWOR) are selected from a predetermined population size $N$. The unique values among the two samples, namely "overlap", are used as the "recapture $x$". The samples are selected in terms of the proportions ($p_1$ and $p_2$) of the population size $N$, namely $n_1 = p_1 \times N$ and $n_2 = p_2 \times N$, where $p_1$

and $p_2$ are looped from 0.05 to 0.45 in steps of 0.10. For each pair of ($p_1$, $p_2$), 1000 independent sets of $n_1$ and $n_2$ are chosen and then corresponding recapture counts $x$ are tallied for each pair of samples. Eq. 2.2 and Eq. 2.4 are used to calculate the point estimators $\hat{N}$ and $\tilde{N}$. The corresponding estimates for variances for $\hat{N}$ and $\tilde{N}$ are calculated using Eq. 2.3 and Eq. 2.5. Confidence sets (CIs) are obtained as described in Chapter 2. $\widehat{Nchisq}$ is obtained by a "searching" method, i.e. "inverting $\chi^2$ test." For given ($n_1$, $n_2$, $x$) in Table 2.1, $x_{22}$ is looped from $0.1N$ to $2.0N$; then a $\chi^2$ test is performed for the "two-way contingency table." If the p-value were greater than a predetermined significance level, say, 0.05, then the corresponding value of $x^*_{++}$ would be saved to form the 95% confidence set for $N$. Finally, estimated coverage rate, mean square error (MSE), mean half width, and tolerance for each estimator are calculated in order to compare the three methods' performance. Note that all three methods are used on each data set.

### 3.1.2 Notation

Estimated coverage rate (denoted C in the following tables) is defined as the ratio of the number of CIs which cover true population size $N$ (denoted $\mathrm{M}_0$) to the total number of CIs constructed (denoted nci, in the simulation nci=1000), i.e.

$$\mathrm{C} = \frac{\mathrm{M}_0}{\mathrm{nci}}, \tag{3.1}$$

is unbiased for the true coverage rate ($\underline{C}$), i.e. $\underline{\mathrm{C}} = \mathrm{E}(\mathrm{C})$.

The mean square error ($\underline{\mathrm{MSE}}$) of an estimator $\hat{\theta}$ is one of many ways to quantify the difference between it and the target value $\theta$ being estimated, and is expressed as follows [*]:

$$\underline{\mathrm{MSE}}(\hat{\theta}) = \mathrm{E}[(\hat{\theta} - \theta)^2]$$

Here $\theta$ is $N$ and $\hat{\theta}$ are the estimators of $N$ which is given in Eq. 2.2 and Eq. 2.4, respectively. Estimated MSE (denoted M in the following tables) in the simulation is calculated

[*] http://en.wikipedia.org/wiki/Mean_squared_error

as follows:

$$\mathrm{M}(\hat{\theta}) = \frac{1}{\mathrm{nci}} \sum_{\mathrm{i}=1}^{\mathrm{nci}} [(\hat{\theta} - \theta)^2] \tag{3.2}$$

Mean half width (denoted MHW in the following tables) is defined as the average of half lengths of simulated CIs. Specifically,

$$\mathrm{MHW} = \frac{\text{upper bound of a CI} - \text{lower bound of a CI}}{2}. \tag{3.3}$$

Since there is no formula to construct confidence interval using "inverting a test" method, to keep the comparisons among the methods consistent, we define mean half width instead of using margin of errors.

Tolerance (denoted T in the following tables) is defined in the simulations as follows:

$$\mathrm{T} = \frac{\sqrt{\mathrm{M}(\hat{\theta})}}{\mathrm{N}}, \tag{3.4}$$

which is used to measure relatively how far the estimate is dispersed around the true value of $N$.

## 3.2 Coverage Rate

The results of the simulations are summarized and presented in tables and figures, where the predetermined population size $N$ is set to 1000 and 5000, respectively and 1000 sets of confidence intervals (CIs) are constructed for the 25 settings, namely 25 pairs of $p_1 = \frac{n_1}{N}$ and $p_2 = \frac{n_2}{N}$, specified in Section 3.1 using the three methods at a nominal 95% confidence level.

We also performed hypothesis tests to check if each coverage rate is statistically significantly different than the nominal confidence level, 0.95, i.e.

$H_0 : \underline{C} = 0.95,$

$H_a : \underline{C} \neq 0.95,$

for all 3 methods for all 25 settings, $H_0$ is rejected if

$$|C - 0.95| \geq 1.96\sqrt{\frac{0.95(0.05)}{\text{nci}}} \tag{3.5}$$

All statistically significant results from the test are colored in red in Table 3.1 and Table 3.2. Some of the entries in Table 3.1 and Table 3.2 will be explained later.

From the tables, it can be seen that most of "red boxes" appear in the first (for both $N = 1000$ and $N = 5000$) and the second (for $N = 1000$) columns, which means that asymptotic normality estimation and Chapman estimation are not as good as "inverting a $\chi^2$" estimation.

From Table 3.1, it can be seen that asymptotic normality estimation ($\hat{N}$) provides the largest coverage rates compared to the other two methods, except for $p_1 = 0.05$ and $p_2 = 0.05$ where "inverting $\chi^2$ test" estimation gives the best coverage rate. The second highest is "inverting a $\chi^2$ test" estimation, whereas Chapman estimation is the lowest. However, in terms of mean half width, Chapman estimation gives the narrowest mean half width, while "inverting a $\chi^2$ test" estimation is the second and asymptotic normality the largest. Therefore, on the whole, "inverting a $\chi^2$ test" estimation is the best estimation method in terms of both coverage rate and mean half width. In other words, the "inverting a $\chi^2$ test" estimation can be practically used to estimate the size of a population. It might take longer to construct since it is actually a "searching method", but the better accuracy offset the extra time for the estimation. In the tables, pk is "consistency" which will be explained in Section 3.4 and pi_est is "missing rate" which will be interpreted in Section 3.5.1.

Comparing Table 3.2 with Table 3.1, it is evident that increasing the population size results in better performance, no matter which method is used.

Figures 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 are more detailed displays of estimated coverage rates. For given $N$, "inverting $\chi^2$ test" method shows quite flat, smoother top surface closer to 0.95 than the other two methods, where coverage rates change dramatically from small ($p_1$, $p_2$) to large ones. When $N$ becomes larger (from 1000 to 5000), comparing Figures 3.1, 3.2, 3.3 to Figures 3.4, 3.5, 3.6, it can be seen that all the estimation methods give more consistent

**Comparisons: Nh vs Nt vs Nchisq**
**(N = 1000)**

| ID | p1 | p2 | CNh | CNt | CNc | pk | Cochran | MHWNh | MHWNt | MHWNc | MNh | MNt | TNh | TNt | pi_est |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.05 | 0.05 | 0.896 | 0.774 | 0.969 | 0.805 | 0 | 2022 | 959 | 940 | 845 | 554 | 0.029 | 0.024 | 0.075 |
| 2 | 0.05 | 0.15 | 0.925 | 0.885 | 0.959 | 0.908 | 0 | 941 | 621 | 623 | 626 | 371 | 0.025 | 0.019 | 0 |
| 3 | 0.05 | 0.25 | 0.968 | 0.903 | 0.953 | 0.920 | 0 | 602 | 448 | 501 | 294 | 244 | 0.017 | 0.016 | 0 |
| 4 | 0.05 | 0.35 | 0.982 | 0.933 | 0.947 | 0.929 | 0 | 485 | 355 | 406 | 205 | 183 | 0.014 | 0.014 | 0 |
| 5 | 0.05 | 0.45 | 0.993 | 0.932 | 0.966 | 0.924 | 0 | 420 | 291 | 327 | 165 | 152 | 0.013 | 0.012 | 0 |
| 6 | 0.15 | 0.05 | 0.937 | 0.886 | 0.970 | 0.904 | 0 | 871 | 619 | 629 | 565 | 402 | 0.024 | 0.020 | 0.001 |
| 7 | 0.15 | 0.15 | 0.951 | 0.925 | 0.948 | 0.945 | 0 | 405 | 344 | 381 | 203 | 184 | 0.014 | 0.014 | 0 |
| 8 | 0.15 | 0.25 | 0.974 | 0.940 | 0.958 | 0.955 | 0 | 310 | 256 | 274 | 147 | 139 | 0.012 | 0.012 | 0 |
| 9 | 0.15 | 0.35 | 0.986 | 0.955 | 0.968 | 0.960 | 0 | 257 | 201 | 211 | 103 | 99 | 0.010 | 0.010 | 0 |
| 10 | 0.15 | 0.45 | 0.991 | 0.942 | 0.960 | 0.934 | 0 | 222 | 161 | 167 | 83 | 81 | 0.009 | 0.009 | 0 |
| 11 | 0.25 | 0.05 | 0.948 | 0.902 | 0.963 | 0.915 | 0 | 532 | 441 | 499 | 287 | 238 | 0.017 | 0.015 | 0 |
| 12 | 0.25 | 0.15 | 0.964 | 0.936 | 0.957 | 0.949 | 0 | 289 | 254 | 273 | 140 | 132 | 0.012 | 0.011 | 0 |
| 13 | 0.25 | 0.25 | 0.976 | 0.946 | 0.955 | 0.954 | 0 | 218 | 184 | 191 | 99 | 96 | 0.010 | 0.010 | 0 |
| 14 | 0.25 | 0.35 | 0.982 | 0.933 | 0.943 | 0.932 | 0 | 184 | 146 | 150 | 81 | 79 | 0.009 | 0.009 | 0 |
| 15 | 0.25 | 0.45 | 0.992 | 0.933 | 0.941 | 0.928 | 0 | 161 | 118 | 121 | 63 | 62 | 0.008 | 0.008 | 0 |
| 16 | 0.35 | 0.05 | 0.939 | 0.939 | 0.935 | 0.946 | 1 | 413 | 361 | 413 | 217 | 192 | 0.015 | 0.014 | 0 |
| 17 | 0.35 | 0.15 | 0.960 | 0.947 | 0.957 | 0.960 | 0 | 225 | 201 | 211 | 110 | 106 | 0.010 | 0.010 | 0 |
| 18 | 0.35 | 0.25 | 0.975 | 0.949 | 0.958 | 0.962 | 0 | 172 | 147 | 151 | 76 | 75 | 0.009 | 0.009 | 0 |
| 19 | 0.35 | 0.35 | 0.985 | 0.939 | 0.952 | 0.938 | 0 | 143 | 114 | 116 | 59 | 58 | 0.008 | 0.008 | 0 |
| 20 | 0.35 | 0.45 | 0.992 | 0.949 | 0.953 | 0.943 | 0 | 127 | 93 | 94 | 48 | 48 | 0.007 | 0.007 | 0 |
| 21 | 0.45 | 0.05 | 0.957 | 0.926 | 0.962 | 0.931 | 0 | 325 | 292 | 328 | 170 | 156 | 0.013 | 0.012 | 0 |
| 22 | 0.45 | 0.15 | 0.954 | 0.928 | 0.945 | 0.943 | 0 | 178 | 160 | 166 | 85 | 84 | 0.009 | 0.009 | 0 |
| 23 | 0.45 | 0.25 | 0.983 | 0.950 | 0.954 | 0.947 | 0 | 139 | 119 | 122 | 61 | 60 | 0.008 | 0.008 | 0 |
| 24 | 0.45 | 0.35 | 0.984 | 0.928 | 0.946 | 0.936 | 0 | 116 | 93 | 94 | 49 | 49 | 0.007 | 0.007 | 0 |
| 25 | 0.45 | 0.45 | 0.993 | 0.960 | 0.959 | 0.957 | 0 | 103 | 76 | 76 | 37 | 36 | 0.006 | 0.006 | 0 |

**Table 3.1**: Performance Comparisons 1: ID, identification number; $p_1$ and $p_2$, proportions for samples $n_1$ and $n_2$; C, estimated coverage rate; pk, consistency (Ref: Section 3.4); Cochran, result of Cochran's test where "1" is "equal" and "0" is "not equal"(Ref: Section 3.4); MHW, estimated mean half width; M, estimated mean square error; T, tolerance; pi_est, simulated missing rate (Ref: Section 3.5.1); Nx, x can be "h", "t", and "chisq", referring to the three methods: asymptotic normality, Chapman, and "inverting a $\chi^2$ test" estimation, respectively. Predetermined population size $N = 1000$; significance level $\alpha = 0.05$ The cells in red contain the results which are significantly different from 0.95.

results: the top surfaces become flatter, smoother and closer to 0.95.

**Comparisons: Nh vs Nt vs Nchisq**
**(N = 5000)**

| ID | p1 | p2 | CNh | CNt | CNc | pk | Cochran | MHWNh | MHWNt | MHWNc | MNh | MNt | TNh | TNt | pi_est |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.05 | 0.05 | 0.966 | 0.889 | 0.977 | 0.904 | 0 | 3135 | 2567 | 2679 | 1621 | 1307 | 0.008 | 0.007 | 0 |
| 2 | 0.05 | 0.15 | 0.961 | 0.947 | 0.951 | 0.956 | 0 | 1631 | 1430 | 1526 | 791 | 742 | 0.006 | 0.005 | 0 |
| 3 | 0.05 | 0.25 | 0.979 | 0.942 | 0.949 | 0.942 | 0 | 1245 | 1050 | 1094 | 574 | 552 | 0.005 | 0.005 | 0 |
| 4 | 0.05 | 0.35 | 0.989 | 0.942 | 0.945 | 0.934 | 0 | 1034 | 819 | 843 | 429 | 420 | 0.004 | 0.004 | 0 |
| 5 | 0.05 | 0.45 | 0.996 | 0.954 | 0.963 | 0.947 | 0 | 911 | 667 | 683 | 339 | 333 | 0.004 | 0.004 | 0 |
| 6 | 0.15 | 0.05 | 0.956 | 0.945 | 0.950 | 0.951 | 1 | 1553 | 1437 | 1537 | 798 | 747 | 0.006 | 0.005 | 0 |
| 7 | 0.15 | 0.15 | 0.972 | 0.935 | 0.955 | 0.951 | 0 | 862 | 782 | 799 | 407 | 400 | 0.004 | 0.004 | 0 |
| 8 | 0.15 | 0.25 | 0.978 | 0.951 | 0.960 | 0.966 | 0 | 663 | 569 | 577 | 291 | 288 | 0.003 | 0.003 | 0 |
| 9 | 0.15 | 0.35 | 0.982 | 0.929 | 0.951 | 0.940 | 0 | 558 | 447 | 451 | 233 | 231 | 0.003 | 0.003 | 0 |
| 10 | 0.15 | 0.45 | 0.984 | 0.937 | 0.937 | 0.946 | 0 | 493 | 364 | 367 | 194 | 193 | 0.003 | 0.003 | 0 |
| 11 | 0.25 | 0.05 | 0.968 | 0.958 | 0.960 | 0.968 | 1 | 1106 | 1046 | 1090 | 535 | 515 | 0.005 | 0.005 | 0 |
| 12 | 0.25 | 0.15 | 0.969 | 0.947 | 0.935 | 0.958 | 0 | 624 | 570 | 577 | 294 | 291 | 0.003 | 0.003 | 0 |
| 13 | 0.25 | 0.25 | 0.985 | 0.964 | 0.964 | 0.969 | 0 | 483 | 416 | 419 | 202 | 200 | 0.003 | 0.003 | 0 |
| 14 | 0.25 | 0.35 | 0.981 | 0.944 | 0.949 | 0.955 | 0 | 406 | 326 | 328 | 172 | 171 | 0.003 | 0.003 | 0 |
| 15 | 0.25 | 0.45 | 0.991 | 0.952 | 0.957 | 0.957 | 0 | 359 | 265 | 266 | 137 | 137 | 0.002 | 0.002 | 0 |
| 16 | 0.35 | 0.05 | 0.961 | 0.948 | 0.949 | 0.959 | 0 | 861 | 822 | 846 | 429 | 419 | 0.004 | 0.004 | 0 |
| 17 | 0.35 | 0.15 | 0.970 | 0.949 | 0.956 | 0.967 | 0 | 488 | 447 | 451 | 228 | 227 | 0.003 | 0.003 | 0 |
| 18 | 0.35 | 0.25 | 0.976 | 0.958 | 0.959 | 0.975 | 0 | 379 | 327 | 328 | 166 | 166 | 0.003 | 0.003 | 0 |
| 19 | 0.35 | 0.35 | 0.989 | 0.953 | 0.954 | 0.958 | 0 | 319 | 257 | 258 | 129 | 129 | 0.002 | 0.002 | 0 |
| 20 | 0.35 | 0.45 | 0.993 | 0.935 | 0.934 | 0.933 | 0 | 282 | 209 | 209 | 112 | 112 | 0.002 | 0.002 | 0 |
| 21 | 0.45 | 0.05 | 0.965 | 0.953 | 0.954 | 0.971 | 0 | 694 | 666 | 681 | 345 | 340 | 0.004 | 0.004 | 0 |
| 22 | 0.45 | 0.15 | 0.964 | 0.952 | 0.955 | 0.985 | 0 | 397 | 364 | 367 | 185 | 184 | 0.003 | 0.003 | 0 |
| 23 | 0.45 | 0.25 | 0.977 | 0.945 | 0.947 | 0.959 | 0 | 306 | 265 | 265 | 136 | 136 | 0.002 | 0.002 | 0 |
| 24 | 0.45 | 0.35 | 0.980 | 0.948 | 0.954 | 0.967 | 0 | 259 | 208 | 208 | 108 | 107 | 0.002 | 0.002 | 0 |
| 25 | 0.45 | 0.45 | 0.994 | 0.953 | 0.943 | 0.948 | 0 | 229 | 170 | 170 | 85 | 85 | 0.002 | 0.002 | 0 |

**Table 3.2**: Performance Comparisons 2: ID, identification number; $p_1$ and $p_2$, proportions for samples $n_1$ and $n_2$; C, estimated coverage rate; pk, consistency (Ref: Section 3.4); Cochran, result of Cochran's test where "1" is "equal" and "0" is "not equal"(Ref: Section 3.4); MHW, estimated mean half width; M, estimated mean square error; T, tolerance; pi_est, simulated missing rate (Ref: Section 3.5.1); Nx, x can be "h", "t", and "chisq", referring to the three methods: asymptotic normality, Chapman, and "inverting a $\chi^2$ test" estimation, respectively. Predetermined population size $N = 5000$; significance level $\alpha = 0.05$. The cells in red contain the results which are significantly different from 0.95.

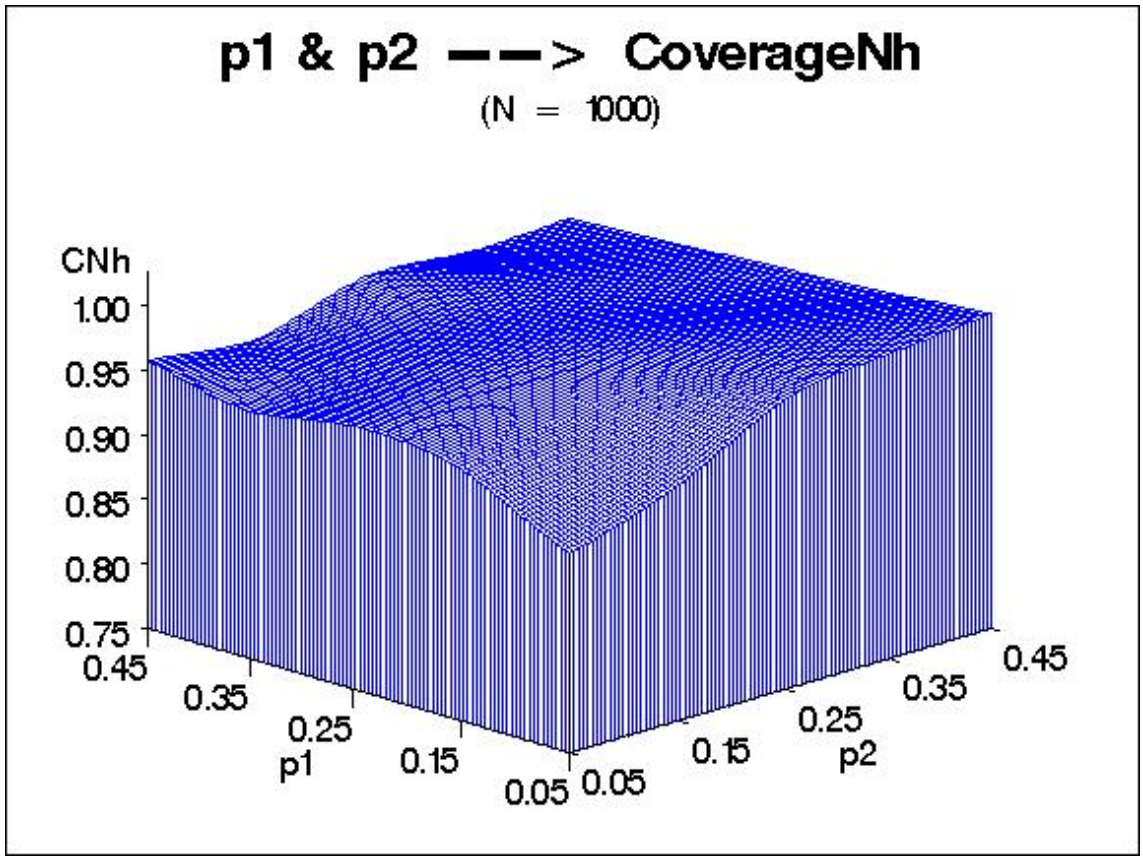


**Figure 3.1**: Coverage Rate Based On $\hat{N}$ For $N = 1000$

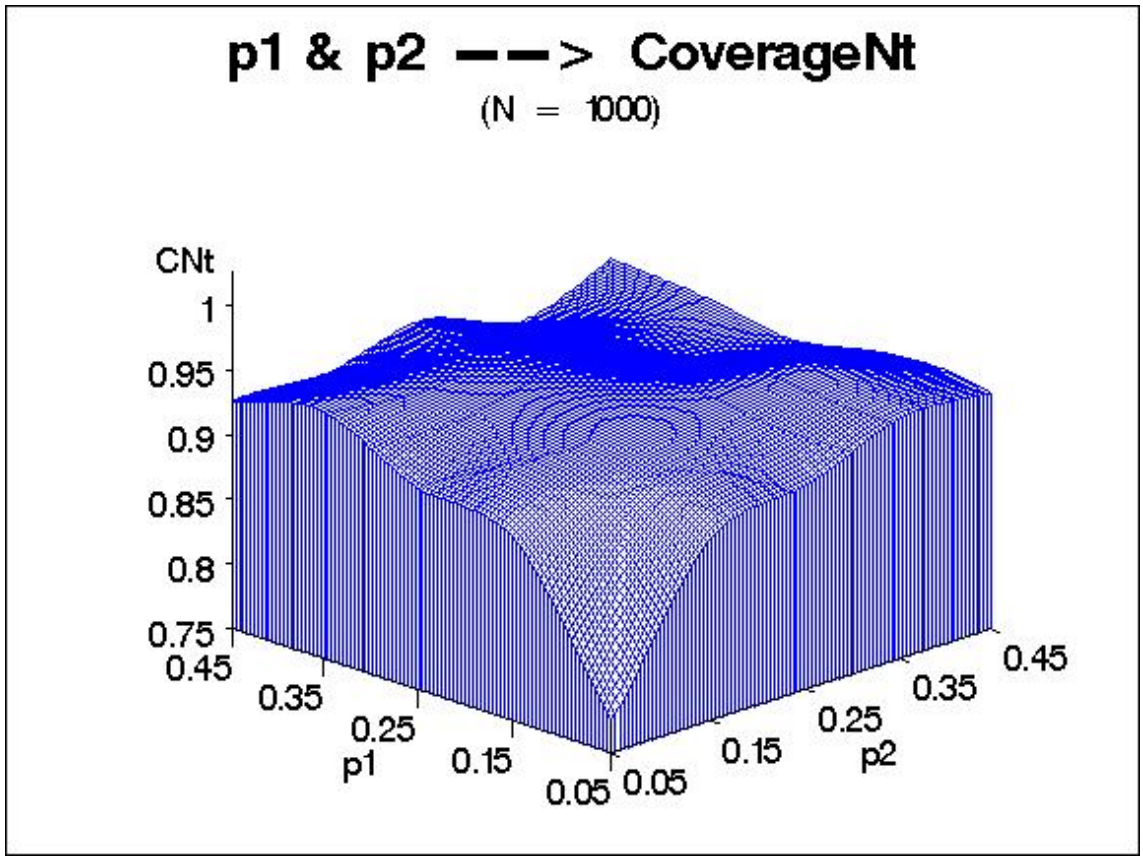


**Figure 3.2**: Coverage Rate Based On $\tilde{N}$ For $N = 1000$

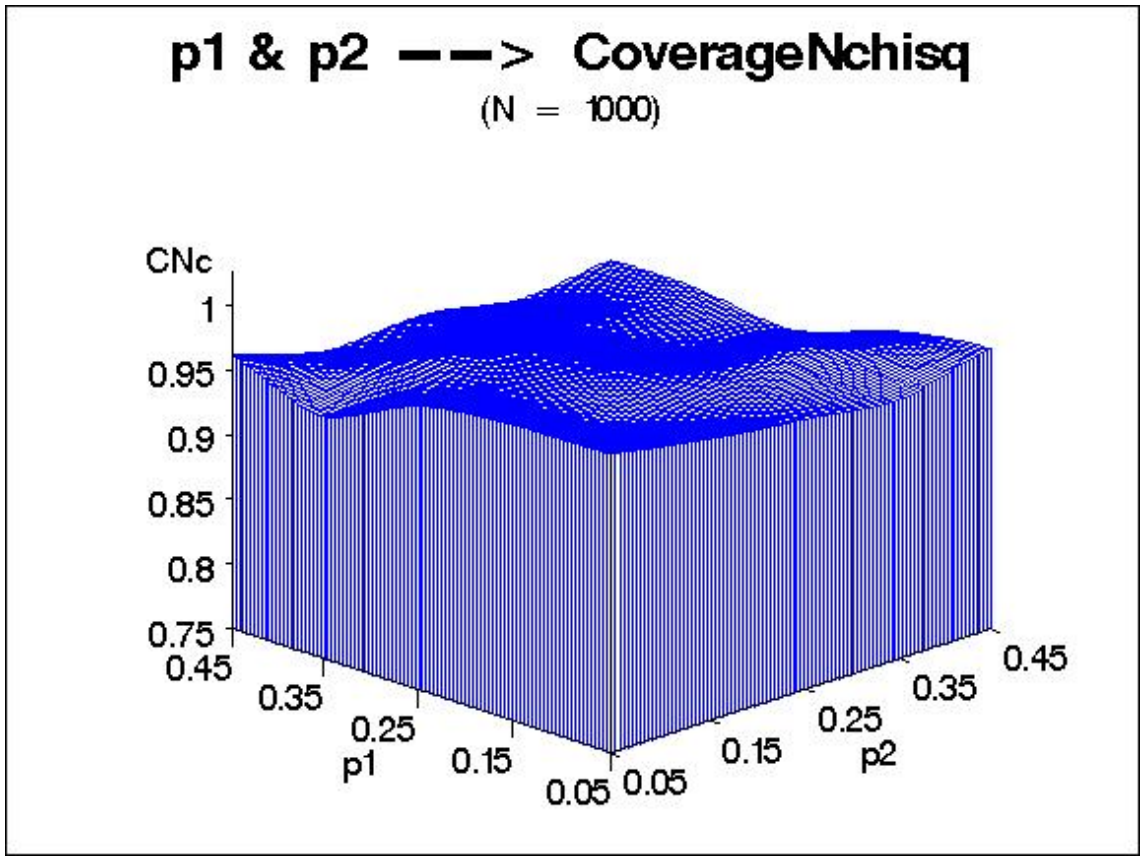


**Figure 3.3**: Coverage Rate Based On "Inverting a $\chi^2$ Test" For $N = 1000$

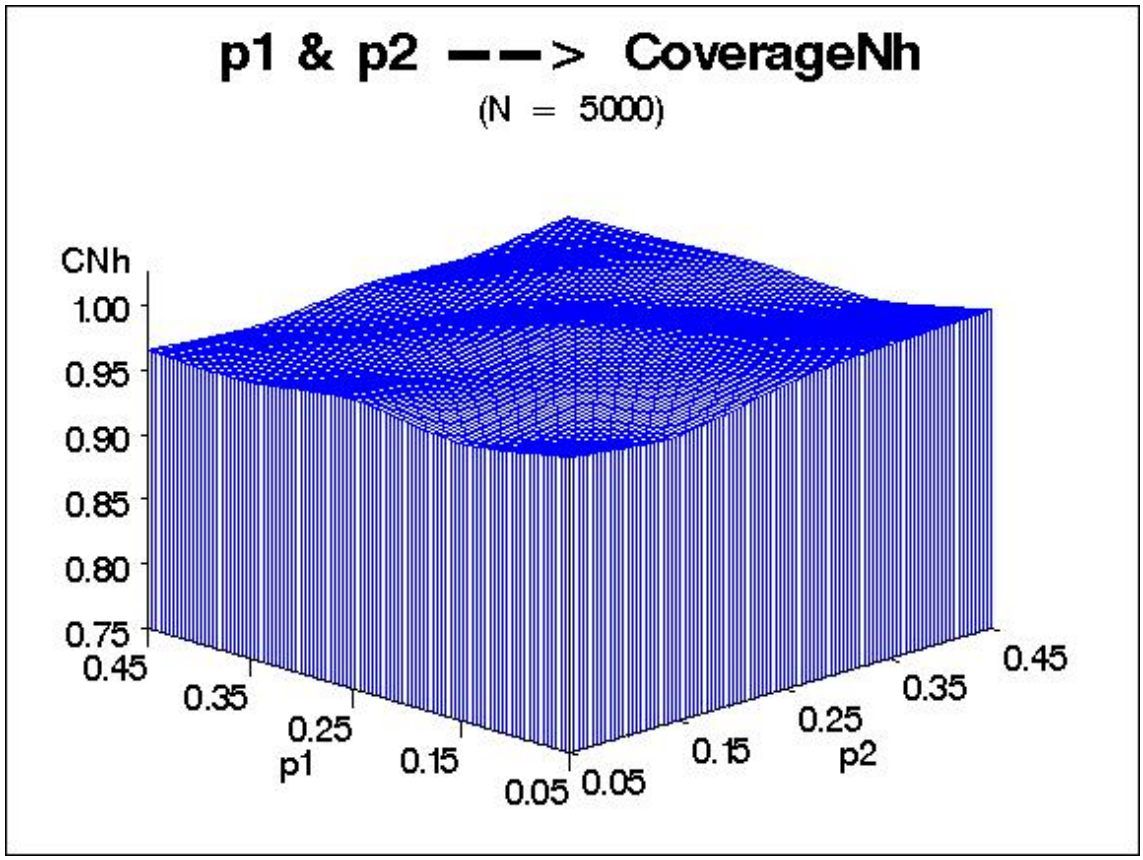


**Figure 3.4**: Coverage Rate Based On $\hat{N}$ For $N = 5000$

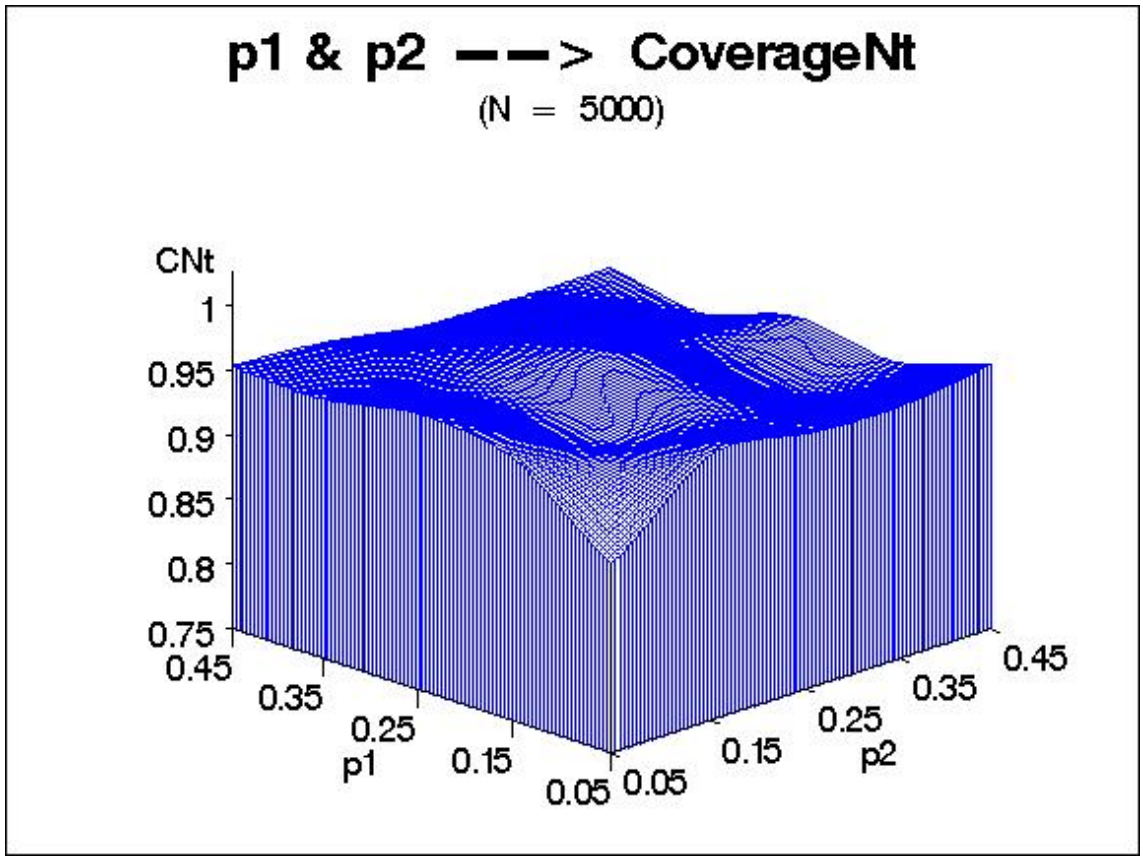


**Figure 3.5**: Coverage Rate Based On $\tilde{N}$ For $N = 5000$

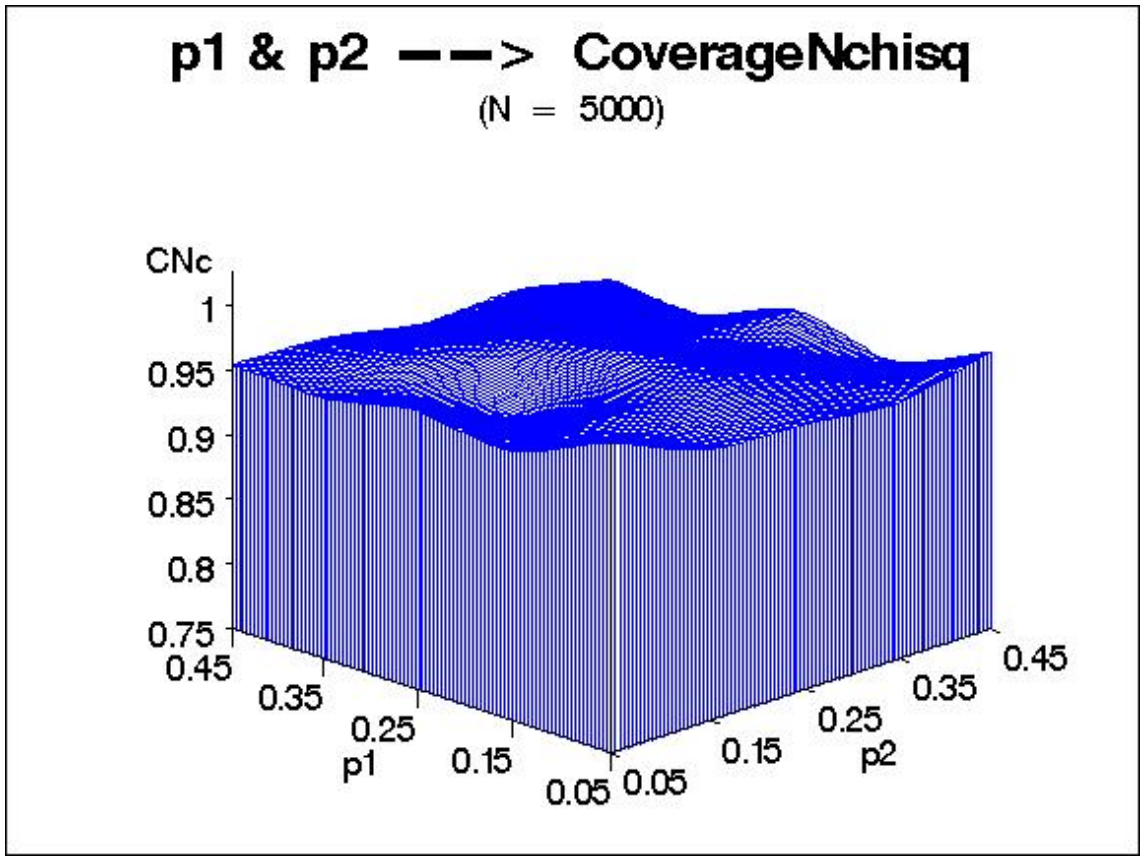


**Figure 3.6**: Coverage Rate Based On "Inverting a $\chi^2$ Test" For $N = 5000$

## 3.3 Analyses on Mean Half Width

As the above analyses showed, coverage rate is a very useful indicator for the performance of each of the three methods used to construct confidence sets. Here, we discuss another useful indicator—mean half width—to make the comparisons among the three methods and present a formal analysis of the inherent relationships between MHW and the sizes of the two samples and population size in a CR design.

### 3.3.1 Comparisons of Mean Half Widths Using Boxplots and Contour Plots

We plotted boxplots of Mean Half Widths (MHW) for the three methods studied in this report and put them side by side to make comparisons. In Figure 3.7, the MHWs are normalized by dividing $N$, which are called "relative mean half widths", denoted RMHW, to make the comparisons on the same scale, in the following sections, and the data for $N = 1000$ and $N = 5000$ are combined.

From the boxplots (Figure 3.7), values of RMHWs from the three methods are right-skewed with the means (denoted "+") being greater than the medians (denoted a horizontal line inside the box). The asymptotic normality and the "inverting a $\chi^2$ test" methods display very similar inter-quartile ranges, the sizes of the boxes, whereas the Chapman method shows a smaller "box size". There are a few points (extreme points) outside the 1.5 inter-quartile ranges (indicated by two horizontal lines outside the boxes). Those points are obtained when the two sample sizes ($p_1$ and $p_2$) are small ($<= 0.15$). "Invert a $\chi^2$ test" and the Chapman methods share more similarities. They have almost the same means, medians, and even similar extreme points. However, since the Chapman method has a slightly smaller box size, its inter-quartile range is smaller than the former. Thus, if one compares the three methods in terms of the ranges defined by the 1.5 inter-quartile ranges of RMHW's, the Chapman method leads to the smallest range among all the three, while "inverting a $\chi^2$ test" and the asymptotic normality methods provide similar ranges. On the other hand, the

asymptotic normality method generates a larger mean than the other two, which makes it more right-skewed. What is more, the asymptotic normality method produces more extreme points compared to the other two, which is a sign that this method is not that stable as the other two.

Contour plots for relative mean half widths are another nice tool to demonstrate the differences in RMHW's among the three methods. From Figures 3.10, 3.8, 3.9 and Figures 3.13, 3.11, 3.12, we see that the method based on asymptotic normality provides the largest relative mean half widths at the same set of $p_1$ and $p_2$, while "inverting a $\chi^2$ test" and the Chapman methods furnish very similar relative mean half widths. The increasing trends of relative mean half widths are evident for each method with the decrease of $p_1$ and $p_2$, while decreasing trends with the increase of $N$.

### 3.3.2 Regression of Mean Half Widths on $p_1$, $p_2$ and $N$

We regressed estimated MHW on $p_1$, $p_2$ and $N$ using SAS. From the SAS output(Page 34), it can be seen that MHW is negatively related to $p_1$ and $p_2$ for fixed $N$. Specifically, MHW decreases as $p_1$ or $p_2$ increase, which supports our intuition that, with larger samples, the estimate of $N$ becomes more precise, which is indicated by a narrower MHW.

The estimated regression coefficients of $N$ in the three methods are similar, which suggests that, for fixed $p_1$ and $p_2$, the population size is not an important factor in distinguishing the effectiveness of performances of the three methods. The sampling fractions $p_1$ and $p_2$ have very similar regression coefficients in the "inverting a $\chi^2$ test" and Chapman estimation methods and almost the same magnitudes for each of the two models. For the asymptotic normality estimation method, the estimated coefficients of $p_1$ and $p_2$ have quite different magnitudes and both are greater than their counterparts in the other two methods. Also, it seems that sample sizes in the asymptotic normality estimation method have larger impact on MHW than in the other two methods.

From the regression analysis, the R-Square is 0.668 for the asymptotic normality method,

0.697 for the Chapman method and 0.692 for the "inverting a $\chi^2$ test" method, which indicates that, on the one hand, the three independent variables, i.e. $p_1$, $p_2$, and $N$, explain a fair amount of variation in the dependent variable RMHW. On the other hand, it seems that more independent variables might need to be added to the model to better interpret the variability of RMHW. In terms of R-Square, the regression is better explained by the Chapman and the "inverting a $\chi^2$ test" methods.

Studentized residual plots demonstrate the inadequacy of our simple model (Figure 3.14, 3.15, and 3.16) and the possible need of more independent variables, interactions or higher order terms, or maybe transformation. However, any such changes to the models would make interpretation for the results very complicated. Here, we only want to know the approximate relationships between MHW and $p_1$, $p_2$ and $N$ and the linear model catch the most part of the variation in MHW, therefore, it serves our goal.

### 3.3.3 Relative Mean Half Width vs. Coverage Rate

We also investigated the relationship between relative mean half width and coverage rate. From Figures 3.17 and 3.18, it can be seen that, after removing the extreme point due to the smallest sample sizes at the coverage rate of around 0.77, there is not much of a relationship between relative mean half width and coverage rate. Intuitively, we expected to see that increase in coverage rate would lead to increase in RMHW, in other words, sacrifice in accuracy in the interval estimation. However, from the outcomes, there are no evident increasing trends appearing between coverage rate and RMHW for all the three methods. For the asymptotic normality and the Chapman methods, roughly speaking, there are very weak decreasing trends between coverage rate and RMHW, while for "inverting a $\chi^2$ test" method, there is not any obvious pattern between coverage and RMHW. On the other hand, it is clear that population size $N$ has a great impact on the relationships between coverage rate and RMHW: increasing $N$ makes coverage rate bigger, which conform to our intuitive guess. All these "surprises" need to be further explored in future investigation.

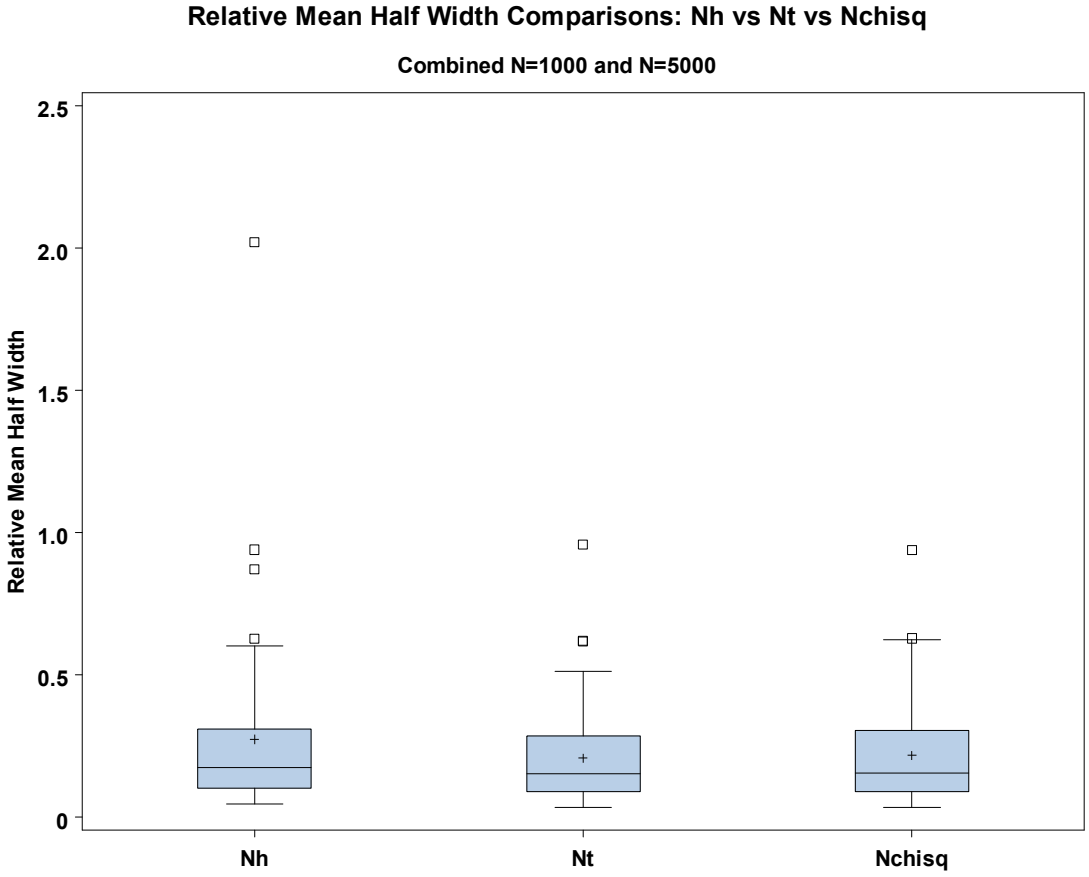


**Figure 3.7**: Boxplot Comparisons of Mean Half Widths Among the Three Methods over Combined Data for $N = 1000$ and $N = 5000$: Nx, the three methods, where x can be "h", "t", and "chisq", referring to asymptotic normality, Chapman, and "inverting a $\chi^2$ test" estimation, respectively; Significance level $\alpha = 0.05$; Relative Mean Half Width, normalized mean half width values by $N$.

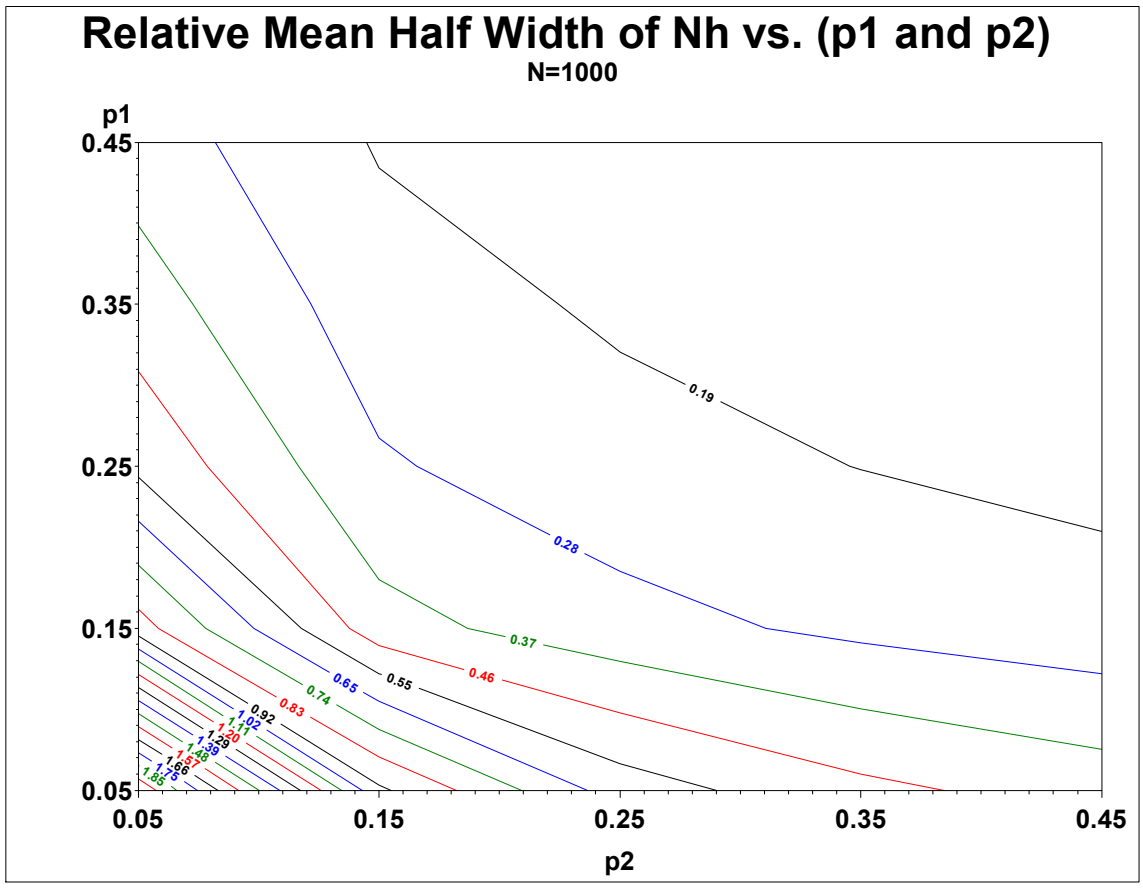


**Figure 3.8**: Contour Plot for Relative Mean Half Widths for Asymptotic Normality Method (Nh) ($N = 1000$).

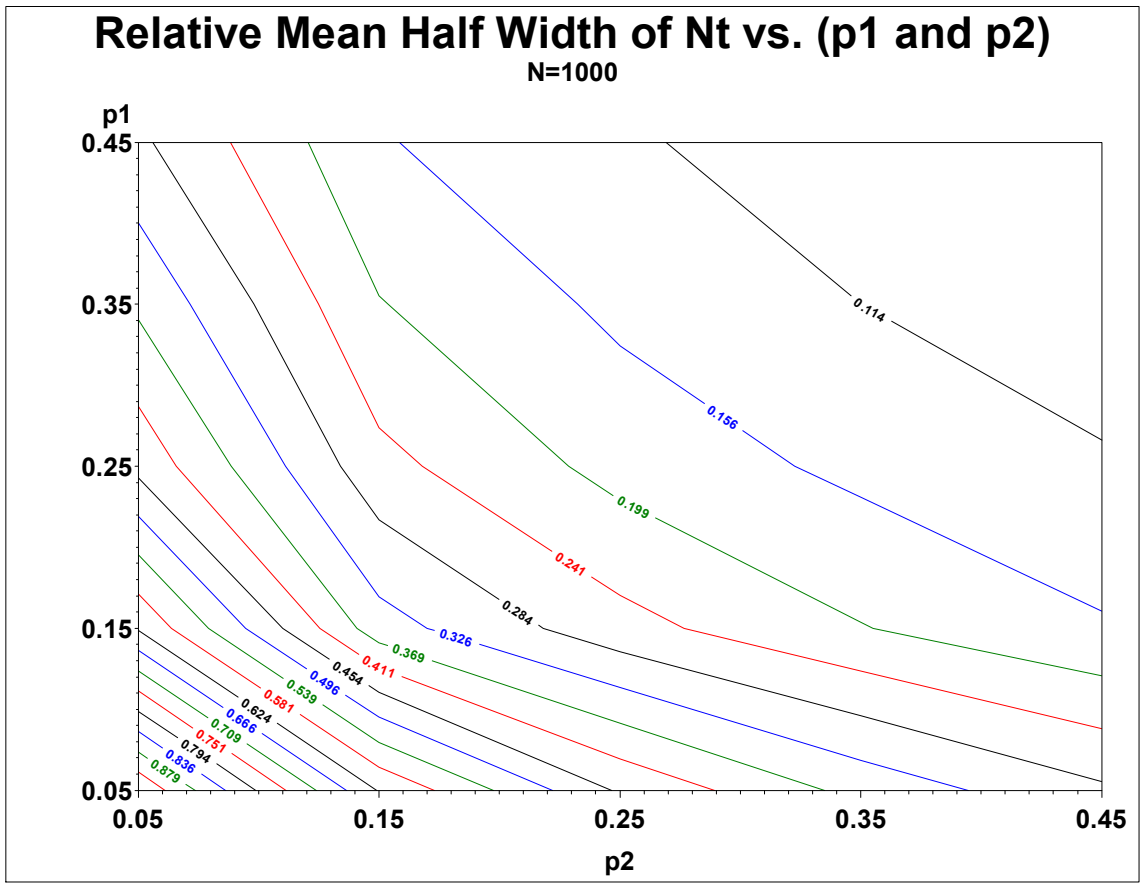


**Figure 3.9**: Contour Plot for Relative Mean Half Widths for Chapman Method (Nt) ($N = 1000$).

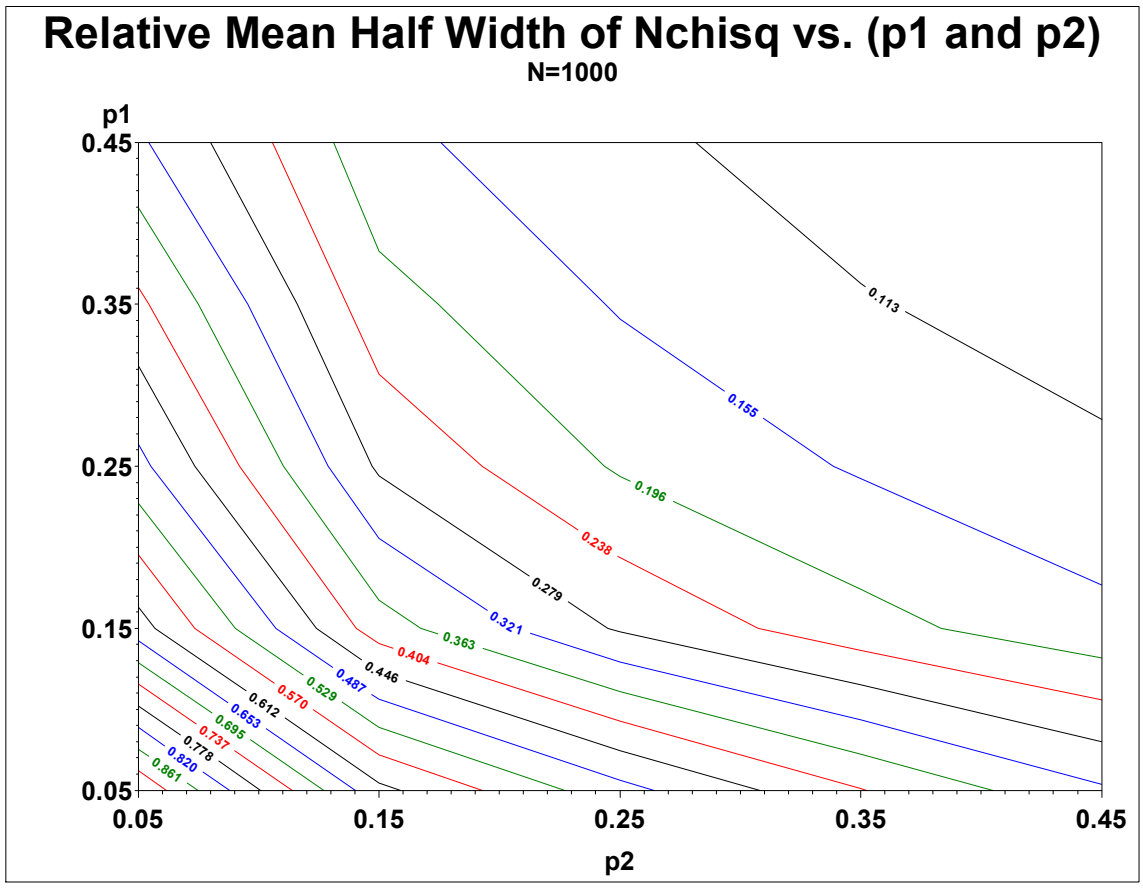


**Figure 3.10**: Contour Plot for Relative Mean Half Widths for "Inverting a $\chi^2$ Test" Method (Nchisq) ($N = 1000$).

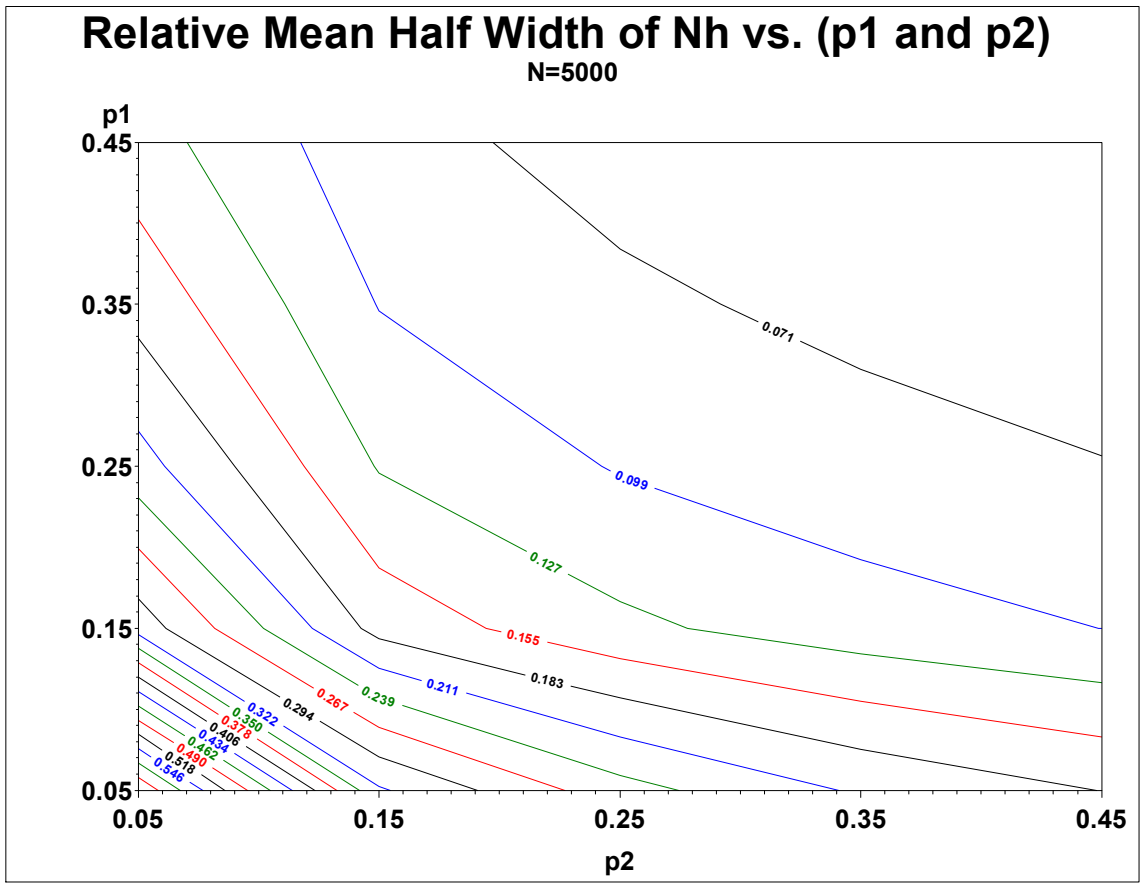


**Figure 3.11**: Contour Plot for Relative Mean Half Widths for Asymptotic Normality Method (Nh) ($N = 5000$).

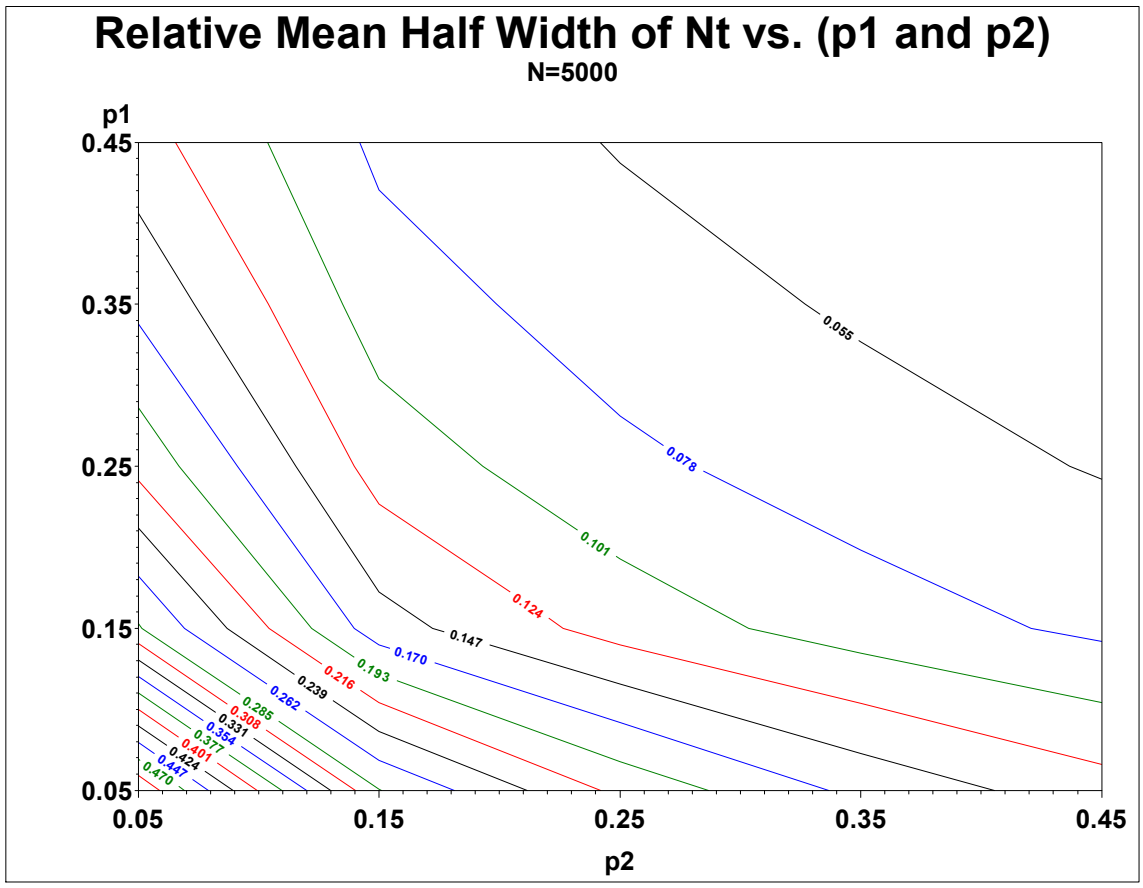


**Figure 3.12**: Contour Plot for Relative Mean Half Widths for Chapman Method (Nt) ($N = 5000$).

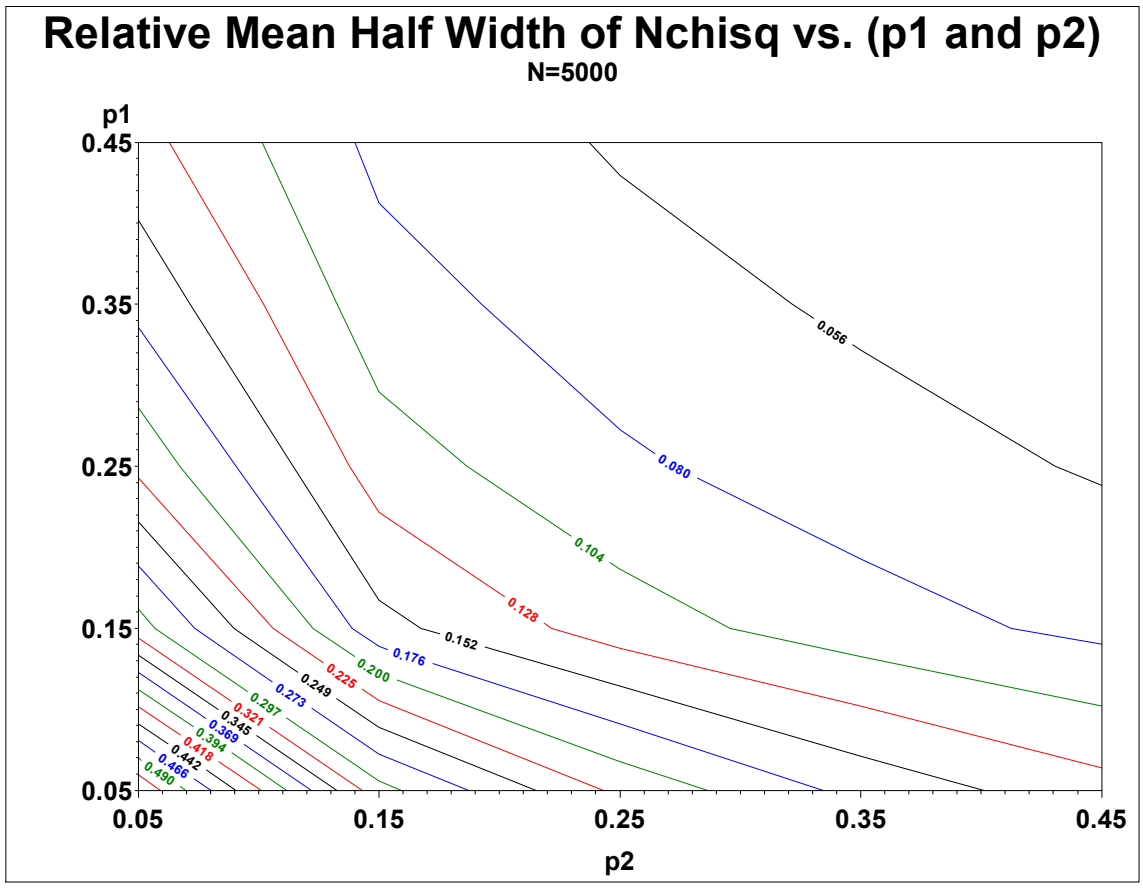


**Figure 3.13**: Contour Plot for Relative Mean Half Widths for "Inverting a $\chi^2$ Test" Method (Nchisq) ($N = 5000$).

***The REG Procedure***
***Model: MODEL1***
***Dependent Variable: MHWNh***

| | |
|---|---|
| **Number of Observations Read** | 50 |
| **Number of Observations Used** | 50 |

| **Analysis of Variance** | | | | | |
|---|---|---|---|---|---|
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 10170395 | 3390132 | 30.84 | <.0001 |
| **Error** | 46 | 5057034 | 109936 | | |
| **Corrected Total** | 49 | 15227430 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | 331.56527 | **R-Square** | 0.6679 |
| **Dependent Mean** | 582.74000 | **Adj R-Sq** | 0.6462 |
| **Coeff Var** | 56.89763 | | |

| **Parameter Estimates** | | | | | |
|---|---|---|---|---|---|
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
| **Intercept** | 1 | 1321.08000 | 144.52595 | 9.14 | <.0001 |
| **p1** | 1 | -2214.50000 | 331.56527 | -6.68 | <.0001 |
| **p2** | 1 | -1868.90000 | 331.56527 | -5.64 | <.0001 |
| **N** | 1 | 0.09417 | 0.02345 | 4.02 | 0.0002 |

***The REG Procedure***
***Model: MODEL1***
***Dependent Variable: MHWNt***

| | |
|---|---|
| Number of Observations Read | 50 |
| Number of Observations Used | 50 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 3 | 6770087 | 2256696 | 35.32 | <.0001 |
| Error | 46 | 2939384 | 63900 | | |
| Corrected Total | 49 | 9709471 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 252.78382 | R-Square | 0.6973 |
| Dependent Mean | 470.88000 | Adj R-Sq | 0.6775 |
| Coeff Var | 53.68328 | | |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | 1 | 977.52500 | 110.18591 | 8.87 | <.0001 |
| p1 | 1 | -1579.00000 | 252.78382 | -6.25 | <.0001 |
| p2 | 1 | -1579.90000 | 252.78382 | -6.25 | <.0001 |
| N | 1 | 0.09436 | 0.01787 | 5.28 | <.0001 |

***The REG Procedure***
***Model: MODEL1***
***Dependent Variable: MHWNc***

| | |
|---|---|
| **Number of Observations Read** | 50 |
| **Number of Observations Used** | 50 |

| **Analysis of Variance** | | | | | |
|---|---|---|---|---|---|
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 7350120 | 2450040 | 34.46 | <.0001 |
| **Error** | 46 | 3270483 | 71097 | | |
| **Corrected Total** | 49 | 10620603 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | 266.64108 | **R-Square** | 0.6921 |
| **Dependent Mean** | 489.66000 | **Adj R-Sq** | 0.6720 |
| **Coeff Var** | 54.45433 | | |

| **Parameter Estimates** | | | | | |
|---|---|---|---|---|---|
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
| **Intercept** | 1 | 1033.98500 | 116.22615 | 8.90 | <.0001 |
| **p1** | 1 | -1660.60000 | 266.64108 | -6.23 | <.0001 |
| **p2** | 1 | -1663.30000 | 266.64108 | -6.24 | <.0001 |
| **N** | 1 | 0.09555 | 0.01885 | 5.07 | <.0001 |

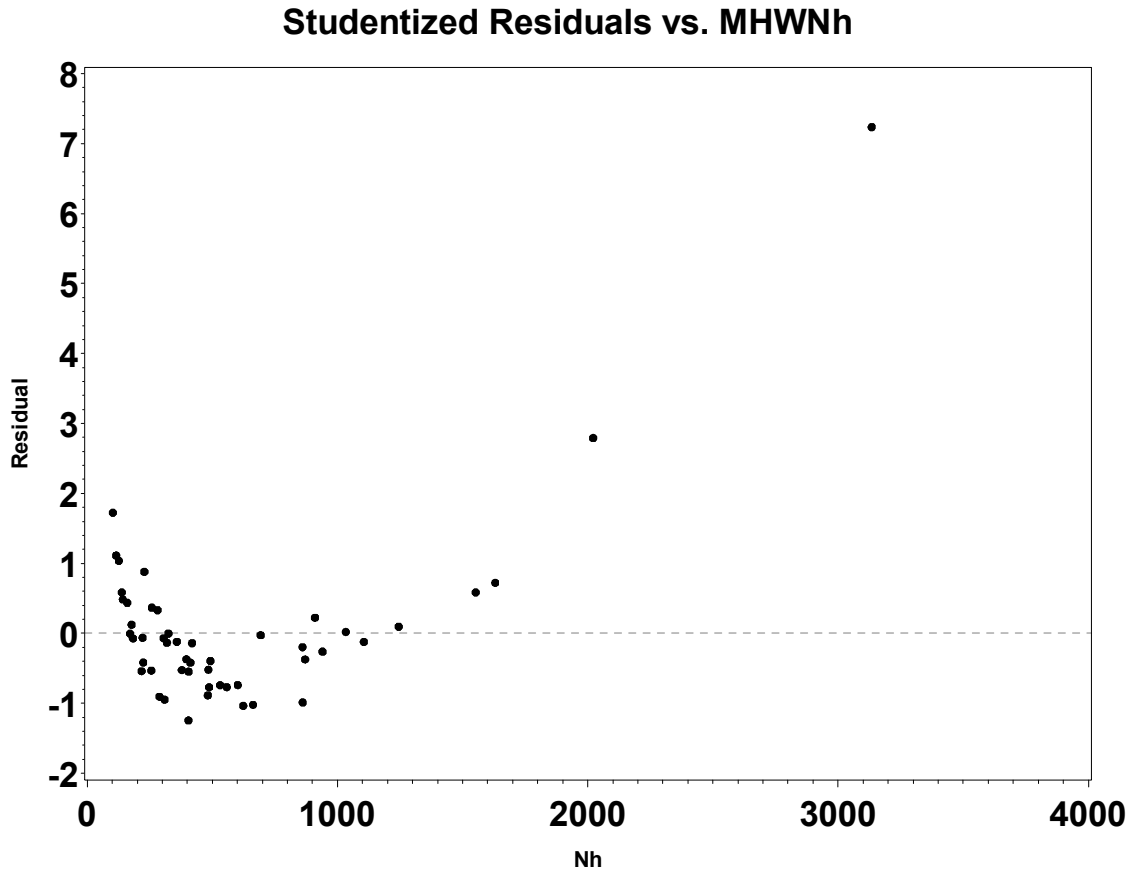


**Figure 3.14**: Residual Plots for Mean Half Widths for Asymptotic Normality Method (Nh). Combined data for $N = 1000$ and $N = 5000$.

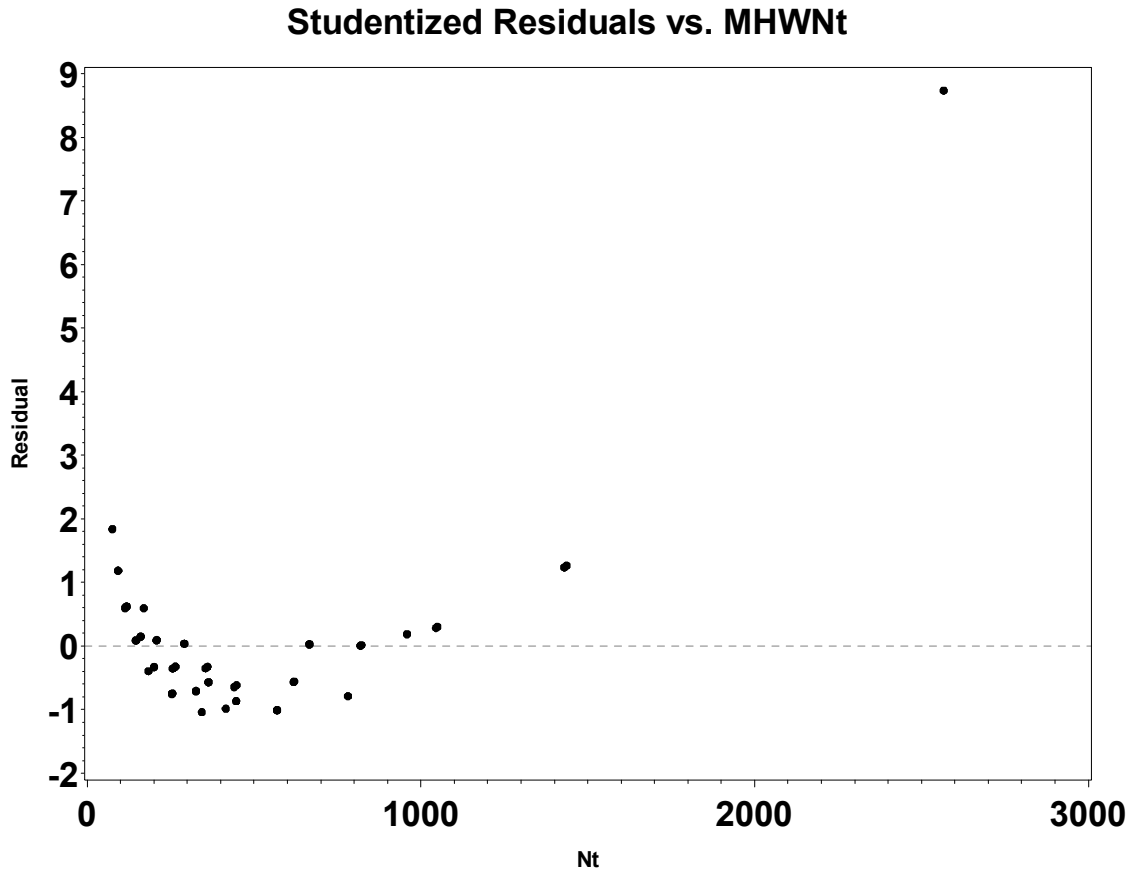


**Figure 3.15**: Residual Plots for Mean Half Widths for Chapman Method (Nt). Combined data for $N = 1000$ and $N = 5000$.

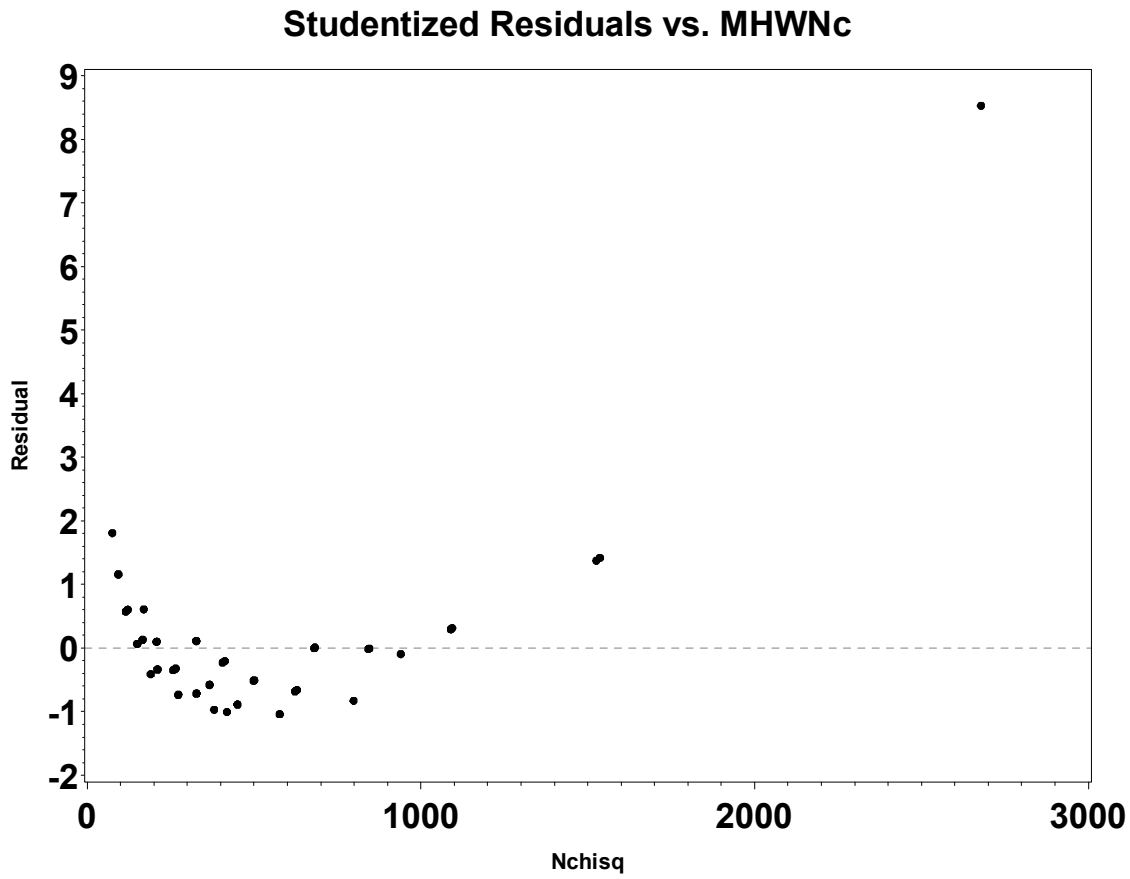


**Figure 3.16**: Residual Plots for Mean Half Widths for "Inverting a $\chi^2$ Test" Method (Nchisq). Combined data for $N = 1000$ and $N = 5000$.

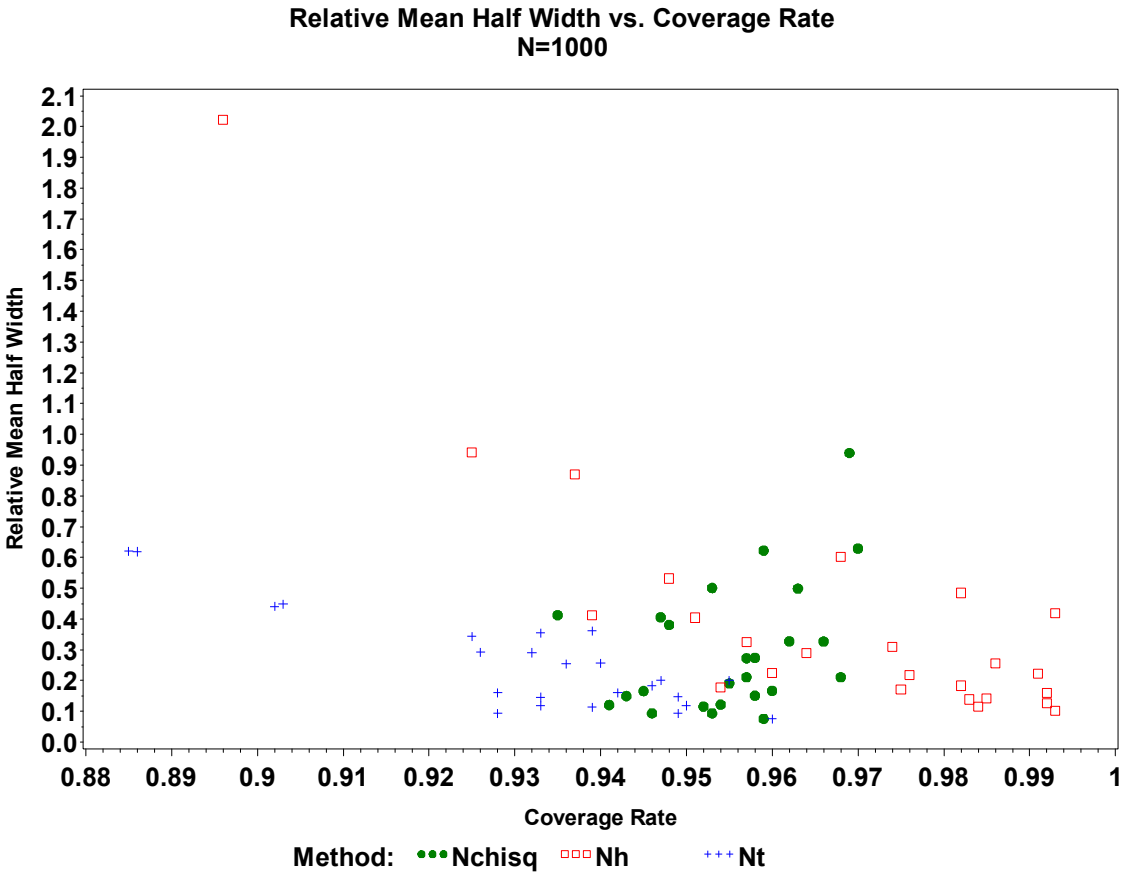


**Figure 3.17**: Relative Mean Half Width vs. Coverage Rate at $N = 1000$.

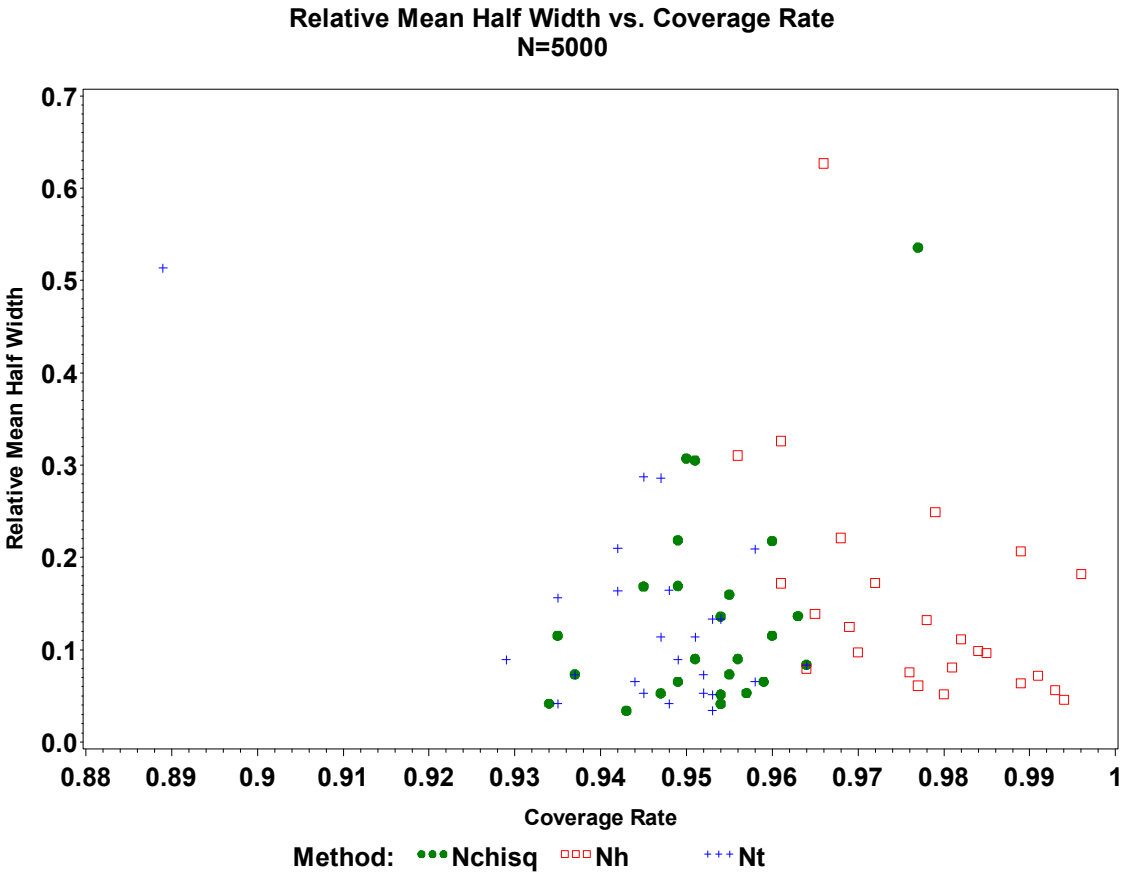


**Figure 3.18**: Relative Mean Half Width vs. Coverage Rate at $N = 1000$.

## 3.4 Cochran's Test And Consistency

To test if the three methods of constructing confidence sets have the same coverage rates, Cochran's test was conducted (denoted "Cochran" in the tables). Recall that all three methods were applied to each data set. Cochran's test is a non-parametric statistical test[5,8] of the hypothesis that $k$ treatments have identical effects based on data from a randomized block design where the response variable can take only two possible outcomes (coded as 0 and 1). Here, the blocks are the b = nci (nci=1000 in the simulation) data sets and the $k = 3$ methods are the treatments. The response is a "1" if the confidence set contains $N$ and "0" otherwise. The data structure is presented in Table 3.3.

**Table 3.3**: Cochran's Test Table

| | **Treatment 1** | **Treatment 2** | **...** | **Treatment k** |
|---|---|---|---|---|
| Block 1 | $X_{11}$ | $X_{12}$ | ... | $X_{1k}$ |
| Block 2 | $X_{21}$ | $X_{22}$ | ... | $X_{2k}$ |
| Block 3 | $X_{31}$ | $X_{32}$ | ... | $X_{3k}$ |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ |
| Block $b$ | $X_{b1}$ | $X_{b2}$ | ... | $X_{bk}$ |

The hypotheses tested given by

$H_0 : \underline{C}_1 = \underline{C}_2 = \ldots = \underline{C}_k$

$H_a : \underline{C}_i \neq \underline{C}_j$ for some treatments $i$ and $j$

where, $\underline{C}_i = P$ (scoring a "1" in column $i$).

Therefore, here, Cochran's test is testing if the three methods have the same coverage rates (under $H_0$) or not (under $H_a$).

Cochran's test statistic in the general case is given by

$$T = k(k-1)\sum_{j=1}^{k}\left(X_{\bullet j} - \frac{N_0}{k}\right)^2 \Big/ \sum_{i=1}^{b} X_{i\bullet}(k - X_{i\bullet}), \tag{3.6}$$

where

$k$ is the number of treatments

$X_{\bullet j}$ is the column total for the $j^{\text{th}}$ treatment

$b$ is the number of blocks

$X_{i\bullet}$ is the row total for the $i^{\text{th}}$ block

$N_0$ is the grand total

In the simulations, cell value ($X_{ij}$) is coded as "1" if the CI using method $j$ obtained from the $i^{\text{th}}$ simulated data set contains the true population size $N$ and "0" otherwise. In this setting, the symbols in Eq. 3.6 are defined as follows:

$k$ is the number of methods for constructing CIs

$X_{\bullet j}$ is the column total for the $j^{\text{th}}$ approach

$b$ is the number of pairs of samples (1000 in the simulations, i.e. nci in Eq. 3.1) taken for each set of $p_1$ and $p_2$ to construct CIs

$X_{i\bullet}$ is the row total for the $i^{\text{th}}$ pair-samples

$N_0$ is the grand total, i.e. the sum of all "1"s in the two-way table.

At nominal significance level $\alpha$, the critical region is

$$T > \chi^2_{1-\alpha,k-1} \tag{3.7}$$

where $\chi^2_{1-\alpha,k-1}$ is the $(1-\alpha)$-quantile of the $\chi^2$ distribution with $k-1$ degrees of freedom. The null hypothesis is rejected if the test statistic is in the critical region.

Cochran's test is based on the following assumptions:

1. The sample size is large, particularly, it assumes that $b$ is "large".

2. $\{X_{ij}, j = 1, 2, \cdots, k\}$, $i = 1, 2, \cdots, b$ are identically distributed.

3. The blocks were randomly selected from the population of all possible blocks.

For $i = 1, 2, \cdots, b$, $Cov(X_{ij}, X_{ij'})$ are equal where $j = 1, 2, \cdots, k$ and $j \neq k$.

For a fixed pair of ($p_1$,$p_2$), the two way table generated by the simulation looks like Table 3.4.

**Table 3.4**: Table For Fixed $p_1$ And $p_2$ Generated By the Simulation—Here $\hat{N}$,$\tilde{N}$,$\widehat{Nchisq}$ are used to denote asymptotic normality, Chapman, and "inverting a $\chi^2$ test" methods respectively; "1" means C.I. covers the true population size $N$, whereas "0" means the opposite; nci=1000 denotes the total number of data sets used in the simulation.

| i (Data Set) | $\hat{N}$ | $\tilde{N}$ | $\widehat{Nchisq}$ |
|---|---|---|---|
| 1 | 1 | 0 | 1 |
| 2 | 0 | 1 | 1 |
| 3 | 1 | 1 | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| nci | 1 | 1 | 1 |
| Coverage Rate | $\underline{C}_1 = \frac{\sum_{i=1}^{\text{nci}} X_{i1}}{\text{nci}}$ | $\underline{C}_2 = \frac{\sum_{i=1}^{\text{nci}} X_{i2}}{\text{nci}}$ | $\underline{C}_3 = \frac{\sum_{i=1}^{\text{nci}} X_{i3}}{\text{nci}}$ |

A summary of the resulting p-value is presented in Table 3.1 and Table 3.2. As shown in Table 3.1 and Table 3.2, we infer that at least two of the three coverage rates differ in each of the 25 conditions.

However, "consistency"-pk's in Table 3.1 and Table 3.2, defined as proportion of times that the three methods give the same answers (either all "1" or all "0"), are mostly greater than 0.90. This apparent contradiction is probably due to the high power of Cochran's test here since $b$ is nci=1000, a large number. But the coverage rates are similar in many cases. To understand this, one has to pay attention to the number of CIs constructed in the simulation for each pair of $p_1$ and $p_2$: 1000. Here the large number of CIs make the small differences detectable. A much smaller number of CIs, such as 50, were also tried to run the simulation. It was easy to get "equal" effect results. In a nutshell, smaller differences can be detected with larger number of CIs.

Cochran's test supports the conclusion that the coverage rates of the confidence intervals constructed by these three methods are not identical in each of the 25 cases considered here.

However, the 25 tests are powerful, each being based on 1000 blocks, and in our judgment the differences among coverage rates are not big enough to be of much practical significance.

## 3.5 Potential Problem and Possible Solutions

As noted above, the method of moments estimator $\hat{N}$ is undefined, if $x$, the number of recaptures, is equal to zero. One possible solution to this problem is as follows. As usual, one takes the first sample of size $n_1$ and marks the $n_1$ individuals in the first sample. Then, one takes the second sample of size $n_2$. If the recapture $x = 0$ in the second sample, return the second sample to the population without marking them. Later on, take a third sample of size $n_2$. If $x$ in the third sample is now positive, use ($n_1$, $n_2$, $x$) to estimate $N$, otherwise return the third sample without marking them and select the fourth sample of size $n_2$ and continue until $x > 0$. This procedure is usually referred to as "multi-sampling". Another solution to this problem is to use the Chapman estimator of $N$, which is defined even if $x = 0$. In the following two sections, missing rate, which is related to the stated problem, and the results from the simulations based on "multi-sampling" process are given.

### 3.5.1 Missing Rate

*Missing rate* is defined as the probability of zero "recapture" in the samples. In other words, missing rate is the probability $P(x = 0)$ for a hyper-geometric distribution (Eq. 2.1).

$$\begin{aligned}
\pi\ (n_1, n_2, N) &= P(x = 0|n_1, n_2, N) \\
&= \frac{\binom{n_1}{0}\binom{N-n_1}{n_2}}{\binom{N}{n_2}} \\
&= \frac{\binom{N-n_1}{n_2}}{\binom{N}{n_2}} \\
&= \frac{(N-n_1)!}{n_2!(N-n_1-n_2)!} \cdot \frac{n_2!(N-n_2)!}{N!} \\
&= \frac{(N-n_1)!}{(N-n_1-n_2)!} \cdot \frac{(N-n_2)!}{N!}
\end{aligned} \tag{3.8}$$

**Missing rate**
**(N=100)**

| ID | p1 | p2 | pi_est | pi_real |
|---|---|---|---|---|
| 1 | 0.05 | 0.05 | 0.762 | 0.770 |
| 2 | 0.05 | 0.15 | 0.432 | 0.436 |
| 3 | 0.05 | 0.25 | 0.218 | 0.229 |
| 4 | 0.05 | 0.35 | 0.109 | 0.110 |
| 5 | 0.05 | 0.45 | 0.052 | 0.046 |
| 6 | 0.15 | 0.05 | 0.444 | 0.436 |
| 7 | 0.15 | 0.15 | 0.068 | 0.071 |
| 8 | 0.15 | 0.25 | 0.014 | 0.009 |
| 9 | 0.15 | 0.35 | 0.000 | 0.001 |
| 10 | 0.15 | 0.45 | 0.000 | 0.000 |
| 11 | 0.25 | 0.05 | 0.247 | 0.229 |
| 12 | 0.25 | 0.15 | 0.012 | 0.009 |
| 13 | 0.25 | 0.25 | 0.000 | 0.000 |
| 14 | 0.25 | 0.35 | 0.000 | 0.000 |
| 15 | 0.25 | 0.45 | 0.000 | 0.000 |
| 16 | 0.35 | 0.05 | 0.116 | 0.110 |
| 17 | 0.35 | 0.15 | 0.002 | 0.001 |
| 18 | 0.35 | 0.25 | 0.000 | 0.000 |
| 19 | 0.35 | 0.35 | 0.000 | 0.000 |
| 20 | 0.35 | 0.45 | 0.000 | 0.000 |
| 21 | 0.45 | 0.05 | 0.039 | 0.046 |
| 22 | 0.45 | 0.15 | 0.000 | 0.000 |
| 23 | 0.45 | 0.25 | 0.000 | 0.000 |
| 24 | 0.45 | 0.35 | 0.000 | 0.000 |
| 25 | 0.45 | 0.45 | 0.000 | 0.000 |

**Table 3.5**: Missing rate comparison: ID, identification number; $p_1$ and $p_2$, percentage sizes for samples $n_1$ and $n_2$; pi_est, simulated missing rate; pi_real, real missing rate; Predetermined population size $N = 100$; significance level $\alpha = 0.05$.

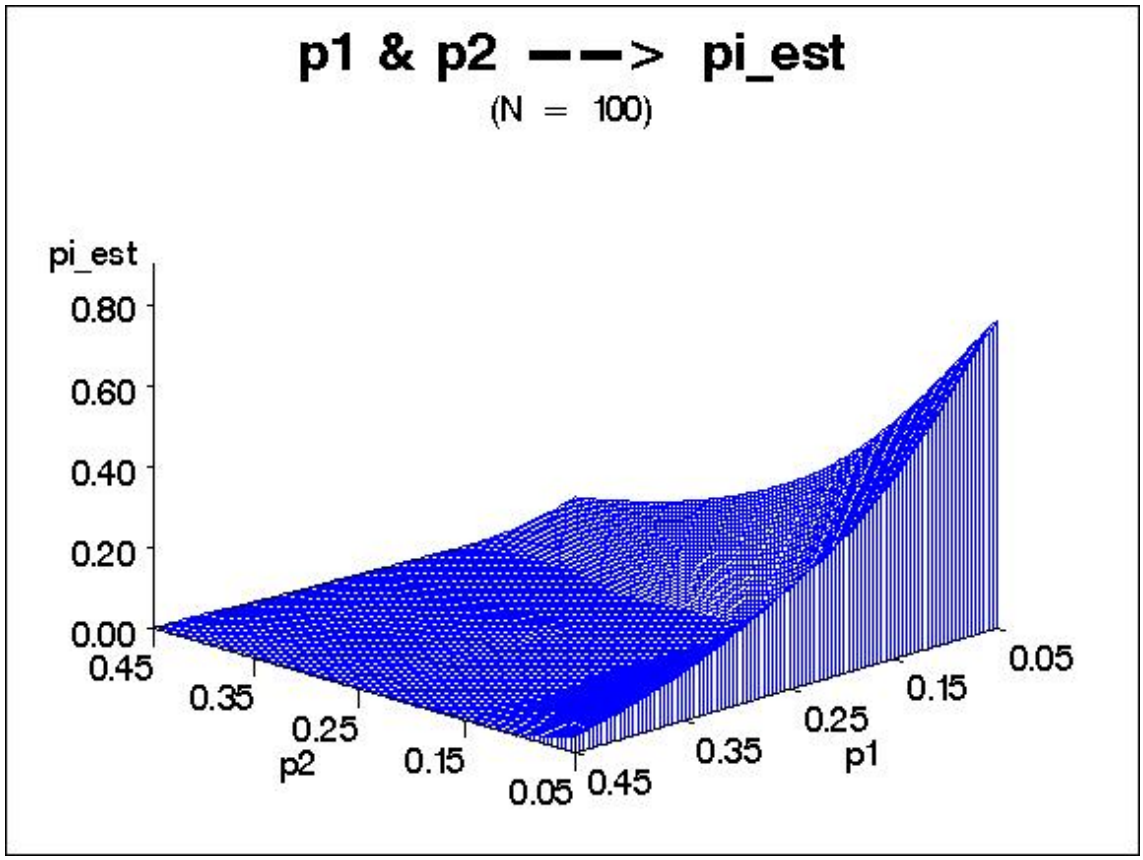


**Figure 3.19**: Estimated Missing Rate Based On $\hat{N}$ For $N = 100$

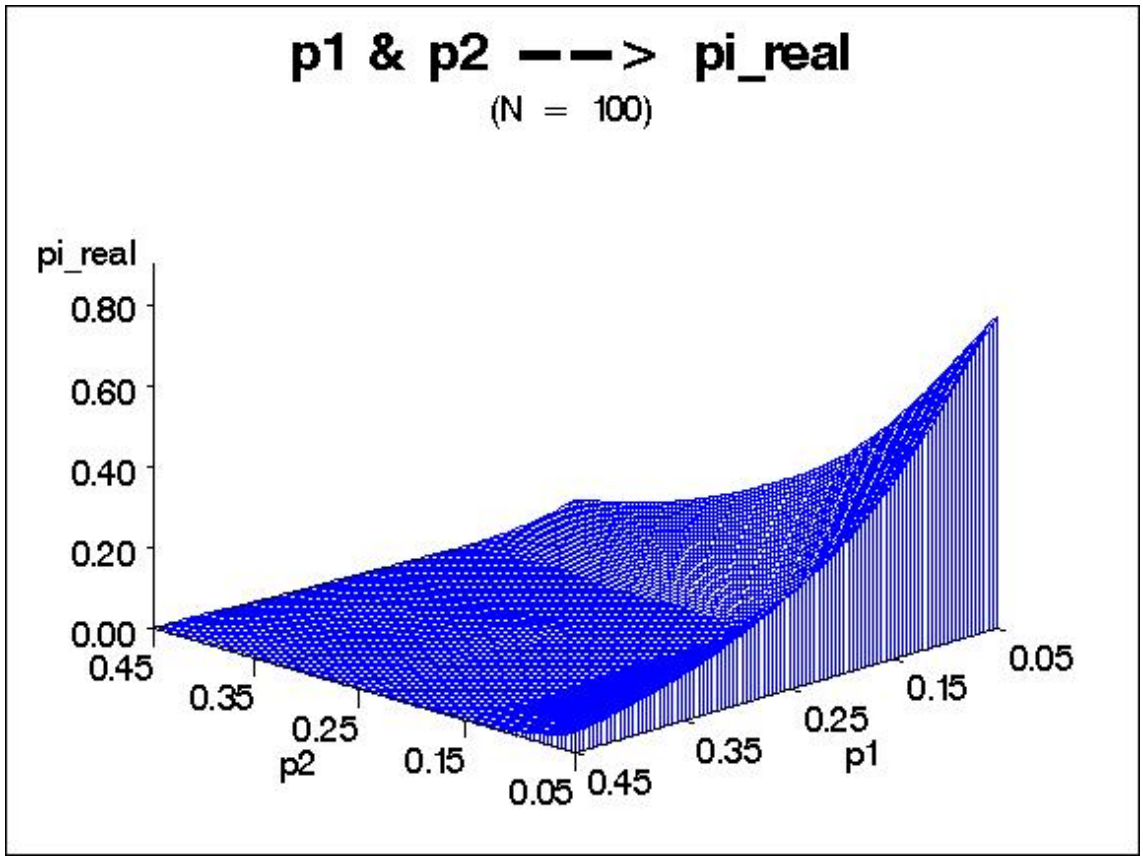


**Figure 3.20**: Real Missing Rate Based On $\hat{N}$ For $N = 100$

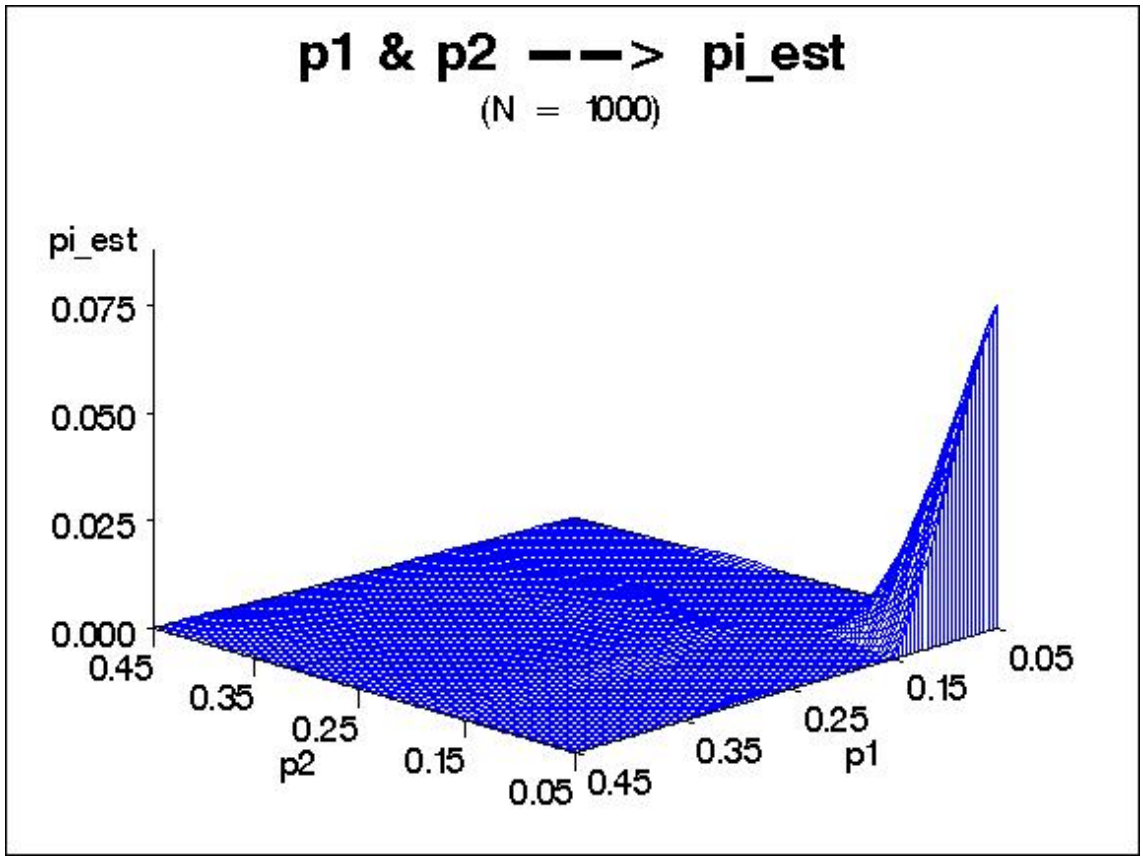


**Figure 3.21**: Estimated Missing Rate Based On $\hat{N}$ For $N = 1000$

Several values of $\pi(n_1, n_2, N)$ are presented in Table 3.5, where they are called "pi_real." On the other hand, a simulated estimate of the missing rate ($\hat{\pi}$) is as follows.

$$\hat{\pi} = \frac{\#\{i, x_i = 0\}}{\text{nci}} = \hat{\pi}(n_1, n_2, N), \tag{3.9}$$

where nci=1000 is the number of CIs constructed for each set of $p_1$ and $p_2$ in Eq. 3.1. The simulated missing rate is denoted "pi_est" in Table 3.5. Therefore, a comparison can be made between the real "missing rate" and its simulated counterpart.

From Table 3.5, Figure 3.19 and Figure 3.20, it can be seen that for $N = 100$ under the same conditions, the real missing rate and its simulated counterpart have a high agreement over almost any choices of $p_1$ and $p_2$.

From Fig. 3.21, it is quite clear that the "missing" only happens in a significant manner when the two samples are small (less than 15% of $N$).

On the other hand, when the population size is much larger than the sizes of the two samples, each sampling process turns to be independent to each other, which approximately leads to a binomial distribution. Under the binomial approximation, the missing rate can be expressed as follows.

$$\pi\ (n_1, n_2, N) \cong\ (1 - \frac{n_1}{N})^{n_2} \tag{3.10}$$

From Eq. 3.10, the missing rate $\pi$ gets larger as the population size $N$ increases, which means that if we know beforehand that the population is very large, we have to select big samples to avoid the "missing".

### 3.5.2 Multi-sampling

As described in the previous section, "multi-sampling" is a solution to "zero recapture" problem. Simulation was conducted to study the effect of "multi-sampling" and the outcomes are illustrated in the following two tables. Since "zero recapture" can only cause a problem for asymptotic normality estimation (Ref: Eq. 2.2), in the following tables we only

report the coverage rates for asymptotic normality, i.e. CNh. From the tables, ns_mean, i.e. the number of subsequent samples after the first sample, is only greater than one when the two sample sizes are small. In other words, more than two samples are only needed if the population size is small, such as 100 in Table 3.6, and the first two samples are small, otherwise it would not be necessary if the population size is large, such as 1000 in Table 3.7. Therefore, if it is possible to get a rough estimate from a literature, an educated guess or simply by common sense, that the size of population of interest is more that a few thousands, then, two-sample CR, i.e. Petersen Method, would be good enough to estimate the population size.

**Multi-sampling CR**
**(N = 100)**

| ID | p1 | p2 | CNh | MHWNh | MNh | TNh | ns_mean |
|---|---|---|---|---|---|---|---|
| 1 | 0.05 | 0.05 | 0.000 | 41 | 76 | 0.087 | 4.513 |
| 2 | 0.05 | 0.15 | 0.720 | 115 | 40 | 0.063 | 1.769 |
| 3 | 0.05 | 0.25 | 0.850 | 153 | 37 | 0.061 | 1.302 |
| 4 | 0.05 | 0.35 | 0.940 | 184 | 52 | 0.072 | 1.119 |
| 5 | 0.05 | 0.45 | 0.980 | 180 | 63 | 0.079 | 1.048 |
| 6 | 0.15 | 0.05 | 0.720 | 105 | 40 | 0.064 | 1.820 |
| 7 | 0.15 | 0.15 | 0.950 | 178 | 66 | 0.081 | 1.087 |
| 8 | 0.15 | 0.25 | 0.950 | 144 | 75 | 0.086 | 1.010 |
| 9 | 0.15 | 0.35 | 0.980 | 107 | 59 | 0.077 | 1.000 |
| 10 | 0.15 | 0.45 | 0.980 | 83 | 38 | 0.062 | 1.001 |
| 11 | 0.25 | 0.05 | 0.880 | 140 | 36 | 0.060 | 1.274 |
| 12 | 0.25 | 0.15 | 0.960 | 143 | 77 | 0.088 | 1.011 |
| 13 | 0.25 | 0.25 | 0.950 | 85 | 45 | 0.067 | 1.000 |
| 14 | 0.25 | 0.35 | 0.960 | 68 | 40 | 0.064 | 1.000 |
| 15 | 0.25 | 0.45 | 0.980 | 55 | 23 | 0.048 | 1.000 |
| 16 | 0.35 | 0.05 | 0.940 | 153 | 50 | 0.071 | 1.124 |
| 17 | 0.35 | 0.15 | 0.920 | 95 | 57 | 0.075 | 1.000 |
| 18 | 0.35 | 0.25 | 0.970 | 62 | 30 | 0.055 | 1.000 |
| 19 | 0.35 | 0.35 | 0.970 | 49 | 23 | 0.048 | 1.000 |
| 20 | 0.35 | 0.45 | 0.980 | 43 | 17 | 0.042 | 1.000 |
| 21 | 0.45 | 0.05 | 0.870 | 147 | 63 | 0.079 | 1.048 |
| 22 | 0.45 | 0.15 | 0.950 | 71 | 41 | 0.064 | 1.000 |
| 23 | 0.45 | 0.25 | 0.930 | 47 | 22 | 0.047 | 1.000 |
| 24 | 0.45 | 0.35 | 0.970 | 39 | 17 | 0.041 | 1.000 |
| 25 | 0.45 | 0.45 | 0.980 | 33 | 13 | 0.036 | 1.000 |

**Table 3.6**: Multiple Sampling 1: ID, identification number; $p_1$ and $p_2$, percentage sizes for samples $n_1$ and $n_2$; C, coverage rate; Se, standard error; MHW, mean half width; M, mean square error; T, tolerance; ns_mean, the number of subsequent samples after the first sample; Nh, asymptotic normality estimation. Predetermined population size $N = 100$; significance level $\alpha = 0.05$ The cells in red contain the results which are significantly different from 0.95.

**Multi-sampling CR**
**(N = 1000)**

| ID | p1 | p2 | CNh | MHWNh | MNh | TNh | ns_mean |
|---|---|---|---|---|---|---|---|
| 1 | 0.05 | 0.05 | 0.880 | 1942 | 778 | 0.028 | 1.084 |
| 2 | 0.05 | 0.15 | 0.930 | 946 | 632 | 0.025 | 1.000 |
| 3 | 0.05 | 0.25 | 0.970 | 597 | 286 | 0.017 | 1.000 |
| 4 | 0.05 | 0.35 | 0.980 | 486 | 209 | 0.014 | 1.000 |
| 5 | 0.05 | 0.45 | 0.990 | 420 | 165 | 0.013 | 1.000 |
| 6 | 0.15 | 0.05 | 0.940 | 854 | 493 | 0.022 | 1.001 |
| 7 | 0.15 | 0.15 | 0.960 | 407 | 194 | 0.014 | 1.000 |
| 8 | 0.15 | 0.25 | 0.980 | 303 | 135 | 0.012 | 1.000 |
| 9 | 0.15 | 0.35 | 0.990 | 254 | 107 | 0.010 | 1.000 |
| 10 | 0.15 | 0.45 | 0.990 | 223 | 83 | 0.009 | 1.000 |
| 11 | 0.25 | 0.05 | 0.950 | 561 | 326 | 0.018 | 1.000 |
| 12 | 0.25 | 0.15 | 0.970 | 287 | 135 | 0.012 | 1.000 |
| 13 | 0.25 | 0.25 | 0.980 | 220 | 98 | 0.010 | 1.000 |
| 14 | 0.25 | 0.35 | 0.980 | 184 | 76 | 0.009 | 1.000 |
| 15 | 0.25 | 0.45 | 0.990 | 161 | 58 | 0.008 | 1.000 |
| 16 | 0.35 | 0.05 | 0.930 | 424 | 245 | 0.016 | 1.000 |
| 17 | 0.35 | 0.15 | 0.960 | 221 | 106 | 0.010 | 1.000 |
| 18 | 0.35 | 0.25 | 0.970 | 171 | 77 | 0.009 | 1.000 |
| 19 | 0.35 | 0.35 | 0.980 | 144 | 59 | 0.008 | 1.000 |
| 20 | 0.35 | 0.45 | 1.000 | 127 | 47 | 0.007 | 1.000 |
| 21 | 0.45 | 0.05 | 0.950 | 326 | 166 | 0.013 | 1.000 |
| 22 | 0.45 | 0.15 | 0.960 | 180 | 87 | 0.009 | 1.000 |
| 23 | 0.45 | 0.25 | 0.980 | 139 | 64 | 0.008 | 1.000 |
| 24 | 0.45 | 0.35 | 0.980 | 117 | 50 | 0.007 | 1.000 |
| 25 | 0.45 | 0.45 | 0.990 | 102 | 39 | 0.006 | 1.000 |

**Table 3.7**: Multiple Sampling 2: ID, identification number; $p_1$ and $p_2$, percentage sizes for samples $n_1$ and $n_2$; C, coverage rate; Se, standard error; MHW, mean half width; M, mean square error; T, tolerance; ns_mean, the number of subsequent samples after the first sample; Nh, asymptotic normality estimation. Predetermined population size $N = 1000$; significance level $\alpha = 0.05$ The cells in red contain the results which are significantly different from 0.95.

# Chapter 4

# Conclusions and Future Work

## 4.1 Conclusions

In this report, we compared the performances of three methods of constructing confidence intervals based on a Capture-Recapture design, i.e., asymptotic normality estimation, Chapman estimation, and "inverting a $\chi^2$ test" estimation, in terms of coverage rate and mean width. Simulation studies were carried out using R and the visualization of results was realized by SAS. It turns out that the "inverting a $\chi^2$ test" estimation is the best in the settings investigated here among the three methods. A multi-sampling search method was simulated to solve the "zero recapture" problem. If, based on past experience or knowledge, the population size can be estimated, to be at least several thousand, the outcomes from multi-sampling simulation show that two-samples will generally be good enough to provide a reasonable estimate for population size.

## 4.2 Future Work

In our future studies, we might try other "inverting a test" methods, such as likelihood ratio test, to find out if this kind of methods is generally better than the ones based on ratio estimates. We might also extend our research to compare the effectiveness of CR methods in terms of other population parameters, such as survival rates; and expand our investigation by using multi-sampling method with after-first samples being marked and

returned to population and a few ordered recaptures being used for estimating population parameters.

# Bibliography

[1] Abeni, D. A., G. Brancato, and C. A. Perucci (1994). Capture-recapture to estimate the size of the population with human immunodeficiency virus type 1 infection. *Epidemiology 5(4)*, 410–414.

[2] Boswell, M. T., K. P. Burnham, and G. P. Patil (1988). Role and use of composite sampling and capture-recapture sampling in ecological studies. *Handbook of Statistics 6*, 469–488.

[3] Briand, L. C., K. E. Emam, B. G. Freimut, and O. Laitenberger (2000). A comprehensive evaluation of capture-recapture models for estimating software defect content. *IEEE Transactions on Software Engineering 26(6)*, 518–540.

[4] Chapman, D. G. *Some properties of the hypergeometricc distribution with applications to zoological censuese*, Volume 1. Univ. Calif. Public. Stat.

[5] Conover, W. J. (1999). *Practical Nonparametric Statistics* (3 ed.). John Wiley and Sons, Inc.

[6] Darroch, J. N., S. E. Fienberg, G. F. V. Glonek, and B. W. Junker (1993). A three-sample multiple- recapture approach to census population estimation with heterogeneous catchability. *J.Am.Statist. Ass. 88*, 1137–1148.

[7] Ebrahimi, N. B. (1997). On the statistical analysis of the number of errors remaining in software design document after inspection. *IEEE Transactions on Software Engineering 23*, 529–532.

[8] Gayle, S. S. (2010). *Masters' Report: 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*. Kansas State University.

[9] Green, R. G. and C. A. Evans (1940). Studies on a population cycle of snowshoe hares on the Lake Alexander area. *I: Gross annual census; J. Wildl. Manag.*, 1932–1939; 4:220–238.

[10] Lohr, S. L. (2010). *Sampling: Design and Analysis* (2nd ed.). Richard Stratton.

[11] Odum, E. P. and A. J. Pontin (1961). Population density of the underground ant and lasius flavus, as determined by tagging with $p^{32}$. *Ecology 42*, 186–188.

[12] Seber, G. A. F. (1970). The effects of trap response on tag-recapture estimates. *Biometrika 26*, 13–22.

[13] Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters* (2nd ed.). New York: MacMillan Publishing.

[14] Wilcove, D. S. and L. L. Master (2008). How many endangered species are there in the united states? *Frontiers in Ecology and the Environment 3(8)*, 414–420.

[15] Wittes, J. T. (1972). On the bias and estimated variance of chapman's two-sample capture-recapture population estimate. *Biometrics 28*, 592–597.

[16] Wohlin, C., P. Runeson, and B. Brantestam (1995). An experimental evaluation of capture-recapture in software inspection. *Software Testing, Verification and Reliability 5*, 213–232.

# Appendix A

# *R* Codes

```
{
cat("\n\n")
cat("   ======================================== ","\n")
cat("            Two-sample CR Simulation        ","\n")
cat("   ----------------------------------------  ","\n")
cat("              Author: Jianjun Hua           ","\n")
cat("                Time: July 2011             ","\n")
cat("   ======================================== ","\n")

#"Invert a Chi-square test" function
#List of arguments
#n1, n2: the two samples
#m: recapture in the second sample
#N: predetermined population size

chisquare<-function(n1,n2,m,N) {

Nchisq=0;n=0
x11=m;x12=n1-m;x1r=n1;
x21=n2-m;xc1=n2

for ( x22 in round(0.1*N):round(2.0*N) ) {
x2r=x21+x22
xc2=x12+x22
xcr=x1r+x2r

m11=x1r*xc1/xcr
m12=x1r*xc2/xcr
m21=x2r*xc1/xcr
m22=x2r*xc2/xcr

chisq=((x11-m11)^2)/m11+((x12-m12)^2)/m12+((x21-m21)^2)/m21+((x22-m22)^2)/m22
if ( chisq <= qchisq(0.95,1) ){
n=n+1
Nchisq[n]=xcr
}

}

out=c(min(Nchisq), max(Nchisq))
return(out)

}

#Function for computing pi_real
pifun<-function(n1,n2,N){(factorial(N-n1)*factorial(N-n2))/(factorial(N)*factorial(N-n1-n2))}

#Calculate the square root of MSE
rmse <- function(obs, pred) sqrt(mean((obs-pred)^2))

#Two-sample CR function for comparing the three methods starts here.
#List of arguments:
#N    : predetermined population size
#k    : number of pairs of p1 and p2
#pt   : base proportion (0.05)
#nci  : number of confidence sets for each pair of p1 and p2
#alpha: significance level
#seed : seed for "sample"

cr.compare<-function(N,k,pt,nci,alpha,seed){

#Matrix for the two-way table for Cochran's test
pres<-array(0, dim=c(nci,3))

#Initialization of variables
#n    : counter used to track the number of inner loops
#m    : "recapture" matrix
#Nh   : asymptotic normality estimate for N
#Nt   : Chapman estimate for N
#llNx : lower bound for Nx method
#ulNx : upper bound for Nx method
#SeNx : standard error for Nx method
#MOENx: margin of error for Nx method
#VNx  : variance for Nh method
```

```
n=0;m=0
Nh=0;Nt=0;llNh=0;ulNh=0;llNt=0;ulNt=0;SeNh=0;SeNt=0;MOENh=0;MOENt=0
CoverageNh=0
CoverageNt=0
MeanWidthNh=0
MeanWidthNt=0
MSENh=0
MSENt=0
TolNh=0
TolNt=0
MeanSeNh=0
MeanSeNt=0

llchisq=0;ulchisq=0
CoverageNchisq=0;CIchisq=0;MeanWidthNchisq=0

#Initialize "missing rate"
pi.est=0

#Initialize "consistency"
pk=0

#Initialize the statistic for Cochran's test
T=0
Cochran=as.character()

set.seed(seed)

for(i in 1:k){
p1=(2*i-1)*pt
n1=round(p1*N)
for(j in 1:k){
n=n+1
p2=(2*j-1)*pt
n2=round(p2*N)

npi=0

npk=0

for(ici in 1:nci){

s1=sample(1:N,n1)
s2=sample(1:N,n2)

m[ici]=(n1+n2)-length(unique(c(s1,s2)))
if (m[ici] < max(0,n1+n2-N) | m[ici] > min(n1,n2)){print("error");break}
if (m[ici] != 0)
{
Nh[ici]=(n1*n2)/m[ici]
VNh=Nh[ici]^2*(Nh[ici]-n1)/(n1*n2)
SeNh[ici]=sqrt(VNh)
MOENh[ici]=1.96*SeNh[ici]
llNh[ici]=Nh[ici]-MOENh[ici]
if (llNh[ici] <= m[ici]) {llNh[ici]=m[ici]}
ulNh[ici]=Nh[ici]+MOENh[ici]

out=chisquare(n1,n2,m[ici],N)
llchisq[ici]=out[1]
ulchisq[ici]=out[2]

}
else
{
npi=npi+1
}

Nt[ici]=(n1+1)*(n2+1)/(m[ici]+1)-1
VNt=((n1+1)*(n2+1)*(n1-m[ici])*(n2-m[ici]))/((m[ici]+1)^2*(m[ici]+2))
SeNt[ici]=sqrt(VNt)
MOENt[ici]=1.96*SeNt[ici]
llNt[ici]=Nt[ici]-MOENt[ici]
if (llNt[ici] <= m[ici]) {llNt[ici]=m[ici]}
```

```
ulNt[ici]=Nt[ici]+MOENt[ici]

if (m[ici]==0) {

Nh[ici]=Nt[ici]
SeNh[ici]=SeNt[ici]
MOENh[ici]=MOENt[ici]
llNh[ici]=llNt[ici]
ulNh[ici]=ulNt[ici]

llchisq[ici]=llNt[ici]
ulchisq[ici]=ulNt[ici]

}

pres[ici,1]=ifelse(((llNh[ici]<=N) & (ulNh[ici]>=N)), 1, 0)
pres[ici,2]=ifelse(((llNt[ici]<=N) & (ulNt[ici]>=N)), 1, 0)
pres[ici,3]=ifelse(((llchisq[ici]<=N) & (ulchisq[ici]>=N)), 1, 0)

if (pres[ici,1]==pres[ici,2] && pres[ici,2]==pres[ici,3]) {npk=npk+1}
}

#Fraction for recapture m=0
pi.est[n]=round(npi/nci,digits=3)

#Cochran's test to test if the coverage rates from the three methods are the same at alpha level

pk[n]=round(npk/nci,digits=3)

T[n]=6*sum((apply(pres,2,sum)-sum(pres)/3)^2)/sum(apply(pres,1,sum)*(3-apply(pres,1,sum)))

Cochran[n]=ifelse((T[n]>qchisq(1-alpha,df=2)),0,1)

#C.I.

CINh=cbind(llNh,ulNh)
CINt=cbind(llNt,ulNt)

#Coverage Rate

CoverageNh[n]=round(length(CINh[N>=llNh & N<=ulNh])/(2*nci),digits=3)
CoverageNt[n]=round(length(CINt[N>=llNt & N<=ulNt])/(2*nci),digits=3)
CIchisq=cbind(llchisq,ulchisq)
CoverageNchisq[n]=round(length(CIchisq[N>=llchisq & N<=ulchisq])/(2*nci),digits=3)

#MSE

MSENh[n]=rmse(Nh,N)
MSENt[n]=rmse(Nt,N)

#Tolerance

TolNh[n]=sqrt(MSENh[n])/N
TolNt[n]=sqrt(MSENt[n])/N

#The average of STD
MeanSeNh[n]=round(mean(SeNh))
MeanSeNt[n]=round(mean(SeNt))

#Mean Width

MeanWidthNh[n]=round(mean(MOENh))
MeanWidthNt[n]=round(mean(MOENt))
MeanWidthNchisq[n]=round(mean(ulchisq-llchisq)/2)

}
}

p1=rep((2*(1:k)-1)*pt,each=k)
p2=rep((2*(1:k)-1)*pt,k)
n1=round(p1*N)
n2=round(p2*N)

pi.real=0
```

```
if (N<=100 & k<=5){
for (i in 1:n) {
pi.real[i]=round(pifun(n1[i],n2[i],N),digits=3)
}
}

if (N<=100 & k<=5){
data.frame(p1,p2,CoverageNh, CoverageNt, CoverageNchisq, pk, Cochran, MeanSeNh, MeanSeNt, MeanWid
}
else {
data.frame(p1,p2,CoverageNh, CoverageNt, CoverageNchisq, pk, Cochran, MeanSeNh, MeanSeNt, MeanWid
}

}

#Two-sample CR function for comparing the three methods ends here.

}

#Substantialize the function and output
N=1000;k=5;pt=0.05;nci=1000;alpha=0.05;seed=888;
z1=Sys.time()
CR.output=cr.compare(N,k,pt,nci,alpha,seed)
CR.output
z2=Sys.time()
difftime(z2,z1)#track CPU time for the simulation

write.csv(CR.output,"CR3N1000k5nci1000alpha5.csv")
```

```
{
cat("\n\n")
cat("   ======================================== ","\n")
cat("         Multi-sampling CR Simulation       ","\n")
cat("   ----------------------------------------  ","\n")
cat("              Author: Jianjun Hua           ","\n")
cat("                Time: July 2011             ","\n")
cat("   ======================================== ","\n")

#Calculate the square root of MSE

rmse <- function(obs, pred) sqrt(mean((obs-pred)^2))

#Multi-sampling function starts here.
#List of arguments:
#N    : predetermined population size
#k    : number of pairs of p1 and p2
#pt   : base proportion (0.05)
#nci  : number of confidence sets for each pair of p1 and p2
#seed : seed for "sample"

multiple.sampling<-function(N,k,pt,nci,seed){

#Initialization of variables
#n    : counter used to track the number of inner loops
#m    : "recapture" matrix
#Nh   : asymptotic normality estimate for N
#llNh : lower bound for Nh method
#ulNh : upper bound for Nh method
#SeNh : standard error for Nh method
#MOENh: margin of error for Nh method
#VNh  : variance for Nh method
#ns   : number of subsequent samples after the first sample

n=0;m=as.numeric(0)
Nh=0;llNh=0;ulNh=0;SeNh=0;MOENh=0
CoverageNh=0
MeanWidthNh=0
MSENh=0
TolNh=0
MeanSeNh=0
ns=0;ns.mean=0

set.seed(seed)

for(i in 1:k){
p1=(2*i-1)*pt
n1=round(p1*N)
for(j in 1:k){
n=n+1
p2=(2*j-1)*pt
n2=round(p2*N)

for(ici in 1:nci){
s1=sample(1:N,n1)

#Subsequent sample(s): if m=0 for nn=1, keep sampling until m!=0
nn=0;m[ici]=0
while (m[ici]==0) {
nn=nn+1
s2=sample(1:N,n2)
m[ici]=(n1+n2)-length(unique(c(s1,s2)))
}

ns[ici]=nn

if (m[ici] < max(0,n1+n2-N) | m[ici] > min(n1,n2)){print("error");break}

Nh[ici]=(n1*n2)/m[ici]
VNh=Nh[ici]^2*(Nh[ici]-n1)/(n1*n2)
SeNh[ici]=sqrt(VNh)
MOENh[ici]=1.96*SeNh[ici]
llNh[ici]=Nh[ici]-MOENh[ici]
if (llNh[ici] <= m[ici]) {llNh[ici]=m[ici]}
```

```
ulNh[ici]=Nh[ici]+MOENh[ici]

}

#average number of subsequent samples

ns.mean[n]=mean(ns)

#C.I.

CINh=cbind(llNh,ulNh)

#Coverage Rate

CoverageNh[n]=round(length(CINh[N>=llNh & N<=ulNh])/(2*nci),digits=2)

#MSE

MSENh[n]=rmse(Nh,N)

#Tolerance

TolNh[n]=sqrt(MSENh[n])/N

#The average of STD
MeanSeNh[n]=round(mean(SeNh))

#Mean Width

MeanWidthNh[n]=round(mean(MOENh))

}
}

p1=rep((2*(1:k)-1)*pt,each=k)
p2=rep((2*(1:k)-1)*pt,k)
n1=round(p1*N)
n2=round(p2*N)

data.frame(p1,p2,CoverageNh, MeanSeNh, MeanWidthNh, MSENh, TolNh, ns.mean);

}

#Multi-sampling function ends here.

}

#Substantialize the function and output
N=1000;k=5;pt=0.05;nci=1000;seed=888;
z1=Sys.time()
CR.output=multiple.sampling(N,k,pt,nci,seed)
CR.output
z2=Sys.time()
difftime(z2,z1) #track CPU time for the simulation

write.csv(CR.output,"CRMSN1000k5nci1000.csv")
```

# Appendix B

# *SAS* Codes

```
/*******************************************************************************
 * Purpose: This code is used to analyze and visualize the results from CR simulation *
 * Author : Jianjun Hua                                                        *
 * Time   : August 2011                                                        *
 *******************************************************************************/

options nodate pageno=1;

%let path=%str(d:\jjhua\MSReport\);
%let libref=CROut;
libname &libref "&path";

%let dtn=CR; * dtn = data set name;

%let alpha=5;
%let ext=csv;

* For comparisons between the three methods;

%macro CRInput(N);

%let filename=&path.CR3N&N.k5nci1OOOalpha&alpha..&ext;
%put &filename; * Check the filenames;

proc import datafile="&filename"
            out=work.N&N
            dbms= dlm replace;
    guessingrows = 5OO;
    delimiter = ',';
run;

* Prepare the table for comparisons between the three methods;

data CR&N;
  set work.N&N;
  CCRN=1*Cochran;
  TolNh=round(TolNh,0.OO1);
  TolNt=round(TolNt,0.OO1);
  MSENh=round(MSENh,1);
  MSENt=round(MSENt,1);
  label id='ID'
        CoverageNh='CNh' CoverageNt='CNt' CoverageNchisq='CNc'
        MeanSeNh='SeNh' MeanSeNt='SeHt'
        MeanWidthNh='MHWNh' MeanWidthNt='MHWNt' MeanWidthNchisq='MHWNc'
        MSENh='MNh' MSENt='MNt'
        TolNh='TNh' TolNt='TNt'
        CCRN='Cochran';
  drop Cochran;
run;

proc print data=CR&N label noobs;
  var ID--pk CCRN MeanWidthNh MeanWidthNt MeanWidthNchisq MSENh MSENt TolNh TolNt pi_est;
  title 'Comparisons: Nh vs Nt vs Nchisq';
  title2 "(N = &N)";
run;

ods rtf file="&path.temp.doc"; /* Temporary file for the traffic-lighting tables */

proc report data=CR&N nowd;
column ID--pk CCRN MeanWidthNh MeanWidthNt MeanWidthNchisq MSENh MSENt TolNh TolNt pi_est;
define ID / order;
define p1 / display;
```

```
define p2 / display;
define pk / display format=5.3;
define CoverageNh / display format=5.3;
define CoverageNt / display format=5.3;
define CoverageNchisq /display format=5.3;
define CCRN / display;
define MeanWidthNh / display;
define MeanWidthNt / display;
define MeanWidthNchisq / display;
define MSENh / display;
define MSENt / display;
define TolNh / display format=5.3;
define TolNt / display format=5.3;
define pi_est / display;

compute CoverageNh;
if abs(CoverageNh-0.95) >= 1.96*sqrt(0.95*.05/1000) then call define(_col_,"style","style={background=red}");
endcomp;

compute CoverageNt;
if abs(CoverageNt-0.95) >= 1.96*sqrt(0.95*.05/1000) then call define(_col_,"style","style={background=red}");
endcomp;

compute CoverageNchisq;
if abs(CoverageNchisq-0.95) >= 1.96*sqrt(0.95*.05/1000) then call define(_col_,"style","style={background=red}");
endcomp;

run;

ods rtf close;

* Make 3D plots to demonstrate the difference between the three methods;

data plot3d;
   set CR&N;
   keep p1 p2 CoverageNh CoverageNt CoverageNchisq pi_est;
run;

proc g3grid data=plot3d out=default;
   grid p1*p2=CoverageNh CoverageNt CoverageNchisq pi_est / spline naxis1=100 naxis2=100;
run;

filename Fig "&path\Report\figures";

ods listing gpath="&path\Report\figures" image_dpi=300;
ods graphics on / reset=index imagefmt=jpeg;

goptions reset=all border
         gunit=pct htext=4 ftext=swiss ftitle=swissb htitle=6 ctext=black
         device=JPEG xmax=6IN ymax=4.5IN
         interpol=none cback=white gsfname=Fig gsfmode=replace
         colors=(black blue green red);

title 'p1 & p2 --> CoverageNh';
title2 "(N = &N)";
proc g3d data=default;
  plot p1*p2=CoverageNh /
       side
       rotate=45 ctop=blue cbottom=black
       xticknum=5 yticknum=5 zmin=0.75 zmax=1.00 zticknum=6;
  format p1 p2 CoverageNh 4.2;
run;
```

```
title 'p1 & p2 --> CoverageNt';
title2 "(N = &N)";
proc g3d data=default;
  plot p1*p2=CoverageNt /
       side
       rotate=45 ctop=blue cbottom=black
       xticknum=5 yticknum=5 zmin=0.75 zmax=1.00 zticknum=6;
  format p1 p2 CoverageNh 4.2;
run;

title1 'p1 & p2 --> CoverageNchisq';
title2 "(N = &N)";
proc g3d data=default;
  plot p1*p2=CoverageNchisq /
       side
       rotate=45 ctop=blue cbottom=black
       xticknum=5 yticknum=5 zmin=0.75 zmax=1.00 zticknum=6;
  format p1 p2 CoverageNh 4.2;
run;

title1 'p1 & p2 --> pi_est';
title2 "(N = &N)";
proc g3d data=default;
  plot p1*p2=pi_est /
       side
       rotate=135 ctop=blue cbottom=black
       xticknum=5 yticknum=5 zmin=0 zmax=0.075;
  format p1 p2 4.2 pi_est 5.3;
run;

ods graphics off;

*For MHW Comparisons (prepare Boxplots);
proc transpose data=CR&N (keep=p1 p2 MeanWidthNh MeanWidthNt MeanWidthNchisq)
  out=MHWNAnalysis
  name=MHWN;
  by p1 p2;
run;

proc print data=MHWNAnalysis label;
  title "MHWNAnalysis";
run;

data MHWNOut;
  set MHWNAnalysis;
  RMHWNV=COL1/&N;
  label RMHWNV='Relative Mean Half Width'
        name='MHWN';
  drop COL1;
run;

proc sort data=MHWNOut out=MHWNOut&N;
  by MHWN;
run;

proc print data=MHWNout&N label;
run;

%mend;

options nomprint nomlogic nosymbolgen; /* For checking macros */
```

```
%let N1=1000;
%CRInput(&N1)
%let N2=5000;
%CRInput(&N2)

* For MHW comparisons using Boxplots;
data MHWNOutComb;
   set MHWNOut&N1 MHWNOut&N2;
run;

proc sort data=MHWNOutComb out=MHWNOutComb1;
    by MHWN;
run;

data MHWNOutComb2;
   set MHWNOutComb1 (drop=_label_);
   MHWN=tranwrd(MHWN,'MeanWidthNh','Nh');
   MHWN=tranwrd(MHWN,'MeanWidthNt','Nt');
   MHWN=tranwrd(MHWN,'MeanWidthNchisq','Nchisq');
run;

data Box1 Box2;
   set MHWNOutComb2;
   if _N_<= 50 then output Box1;
   else
   output Box2;
run;

data Box;
   set Box2 Box1;
run;

*For checking dataset before Boxplots;
proc print data=Box;
run;

*For drawing Boxplots;
options nonumber papersize=(11in 8.5in);;
goptions reset=all;

ods listing close;
ods pdf file="&path.Report\figures\MHWBoxplotComb.pdf" notoc;

goptions ftext='Arial/bo' htext=2.5 gunit=pct;
symbol color = black h = 2.0;
proc boxplot data=Box;
  plot RMHWNV*MHWN /nohlabel boxstyle=schematic boxwidth=10;
  title1 h=3 'Relative Mean Half Width Comparisons: Nh vs Nt vs Nchisq';
  title3 h=2.5 "Combined N=1000 and N=5000";
run;
ods pdf close;
goptions reset=all;
ods listing;

*For plot RMHW(=MHW/N) vs. CR, p1, p2;

%macro RMHWCR(N);
data MHWNOut&N (drop=MeanWidthNh MeanWidthNt MeanWidthNchisq);
    set CR&N (keep=p1 p2 MeanWidthNh MeanWidthNt MeanWidthNchisq);
    Nh=MeanWidthNh/&N;
    Nt=MeanWidthNt/&N;
    Nchisq=MeanWidthNchisq/&N;
```

```
run;

proc transpose data=MHWNout&N
   out=tm
   prefix=RMHW
   name=N;
   by p1 p2;
run;

proc sort data=tm out=tsm(rename=(RMHW1=RMHW));
   by N;
run;

proc print data=tsm;
run;

data COut(rename=(CoverageNh=Nh CoverageNt=Nt CoverageNchisq=Nchisq));
    set CR&N (keep=p1 p2 CoverageNh CoverageNt CoverageNchisq);
run;

proc transpose data=COut
   out=tc
   prefix=CV
   name=N;
   by p1 p2;
run;

proc sort data=tc out=tcm (drop=_label_ rename=(CV1=CR));
   by N;
run;

proc print data=tcm;
run;

data MCMerge;
   merge tsm tcm;
   by N;
run;

proc sort data=MCMerge out=MCMergeOut;
   by N CR;
run;

proc print data=MCMergeOut;
run;

goptions reset=all gunit=pct htext=3 ftext='Arial/bo' htitle=3;
symbol1 v=dot height=2 c=green width=2 i=none;
symbol2 v=square height=2 c=red width=2 i=none;
symbol3 v=plus height=2 c=blue width=2 i=none;

ods listing close;
ods pdf file="&path.Report\figures\RMHWvsCRN&N..pdf" notoc;

axis1 label=(h=2.5 angle=90 rotate=0 "Relative Mean Half Width");
axis2 label=(h=2.5 "Coverage Rate");
proc gplot data=MCMergeOut(where=(CR>0.8));
   title1 "Relative Mean Half Width vs. Coverage Rate";
   title2 "N=&N";
   plot RMHW*CR=N /vaxis=axis1 haxis=axis2;
   label N='Method:';
run;
```

```
ods pdf close;
goptions reset=all;
ods listing;

goptions reset=all gunit=pct border cback=white
colors=(black blue green red)
ftext='Arial/bo' htext=3.0 htitle=5;

/* Contour plot for RMHW vs. (p1 and p2) */
ods listing close;
ods pdf file="&path.Report\figures\RMHWvsP1P2N&N..pdf" notoc;

/* Process the original data with PROC G3GRID */
proc g3grid data=MCMergeOut;
   grid p1*p2=RMHW;
run;

/* Create the contour graph with a spline interpolation */
proc gcontour data=MCMergeOut (firstobs=1 obs=25);
   title1 "Relative Mean Half Width of Nchisq vs. (p1 and p2)";
   title2 "N=&N";
   plot p1*p2=RMHW / nolegend autolabel=(check=none) nlevels=20;
run;

proc gcontour data=MCMergeOut (firstobs=26 obs=50);
   title1 "Relative Mean Half Width of Nh vs. (p1 and p2) ";
   title2 "N=&N";
   plot p1*p2=RMHW / nolegend autolabel=(check=none) nlevels=20;
run;

proc gcontour data=MCMergeOut (firstobs=51 obs=75);
   title1 "Relative Mean Half Width of Nt vs. (p1 and p2)";
   title2 "N=&N";
   plot p1*p2=RMHW / nolegend autolabel=(check=none) nlevels=20;
run;

quit;

ods pdf close;
ods listing;

%mend;

%RMHWCR(1000)
%RMHWCR(5000)


* For regress MHW on p1, p2 and N for each method;

%macro MHWReg(N,k);
%let filename=&path.CR3N&N.k5nci1000alpha&alpha..&ext;
%put &filename; * Check the filename;

proc import datafile="&filename"
            out=temp&k
            dbms= dlm replace;
    guessingrows = 500;
    delimiter = ',';
run;

data temp&k.out (rename=(MeanWidthNh=MHWNh MeanWidthNt=MHWNt MeanWidthNchisq=MHWNc));
```

```
   set temp&k (keep=p1 p2 MeanWidthNh MeanWidthNt MeanWidthNchisq);
   N=&N;
run;
%mend;

%MHWReg(1000,1)
%MHWReg(5000,2)

data comb12;
  set temp1out temp2out;
run;

proc print data=comb12;
run;

options nonumber papersize=(11in 8.5in);;
goptions reset=all;

ods listing close;
ods pdf file="&path.Report\figures\MHWResidualComb.pdf" notoc;

goptions ftext='Arial/bo' htext=4.0 htitle=4.0 gunit=pct;
symbol value=dot color = black h = 1.5;
axis1 label=(h=2.5 angle=90 rotate=0 "Residual");
axis2 label=(h=2.5 "Nh");
axis3 label=(h=2.5 "Nt");
axis4 label=(h=2.5 "Nchisq");

/* Using macros might simplify the following codes */
proc reg data=comb12;
  model MHWNh = p1 p2 N;
  output out=res_nh (keep=MHWNh rhat_nh yhat_nh) rstudent=rhat_nh predicted=yhat_nh cookd=cook_distance_nh;
run;

proc reg data=comb12;
  model MHWNt = p1 p2 N;
  output out=res_nt (keep=MHWNt rhat_nt yhat_nt) rstudent=rhat_nt predicted=yhat_nt cookd=cook_distance_nt;
run;

proc reg data=comb12;
  model MHWNc = p1 p2 N;
  output out=res_nc (keep=MHWNc rhat_nc yhat_nc) rstudent=rhat_nc predicted=yhat_nc cookd=cook_distance_nc;
run;

title "Studentized Residuals vs. MHWNh";
proc gplot data=res_nh;
   plot rhat_nh*MHWNh / vaxis=axis1 haxis=axis2 vref=0 lvref=2 lh=15;
run;

title "Studentized Residuals vs. MHWNt";
proc gplot data=res_nt;
   plot rhat_nt*MHWNt / vaxis=axis1 haxis=axis3 vref=0 lvref=2 lh=15;
run;

title "Studentized Residuals vs. MHWNc";
proc gplot data=res_nc;
   plot rhat_nc*MHWNc / vaxis=axis1 haxis=axis4 vref=0 lvref=2 lh=15;
run;

ods pdf close;
goptions reset=all;
ods listing;
```

```
* For multisampling;

%let N=1000;  /* Change N if necessary. */
%let filename=&path.CRMSN&N.k5nci1000alpha&alpha..&ext;
%put &filename; * Check the filename;

proc import datafile="&filename"
            out=work.N&N
            dbms= dlm replace;
    guessingrows = 500;
    delimiter = ',';
run;

data &libref..&dtn;
  set work.N&N;
  MeanWidthNh=round(MeanWidthNh,1);
  MSENh=round(MSENh,1);
  TolNh=round(TolNh,0.001);
  ns_mean=round(ns_mean,0.001);
  label
        CoverageNh='CNh'
        MeanSeNh='SeNh'
        MeanWidthNh='MHWNh'
        MSENh='MNh'
        TolNh='TNh';
run;

proc print data=&libref..&dtn label noobs;
  title 'Multi-sampling CR';
  title2 "(N = &N)";
run;

ods rtf file="&path.temp.doc";

proc report data=&libref..&dtn nowd;
column ID p1 p2 CoverageNh MeanWidthNh MSENh TolNh ns_mean;
define ID / order;
define p1 / display;
define p2 / display;
define CoverageNh / display format=5.3;
define MeanWidthNh / display;
define MSENh / display;
define TolNh / display format=5.3;
define ns_mean / display format=5.3;

compute CoverageNh;
if abs(CoverageNh-0.95) >= 1.96*sqrt(0.95*.05/1000) then call define(_col_,"style","style={background=red}");
endcomp;

run;

ods rtf close;

* For "Missing Rate";

%let N=100; /* Change N if necessary. */
filename CRNIn "&path.CRN&N.k5nci1000.txt";

data CRN&N;
   infile CRNIn missover;
   input p1 p2 CNh CNt SeNh SeNt MWN MWNt MNh MNt TNh TNt pi_est pi_real;
```

```
   ID=_N_;
run;

proc print data=CRN&N;
   title 'Missing rate';
   title2 "(N=&N)";
   var p1 p2 pi_est pi_real;
   ID ID;
run;

proc g3grid data=CRN&N(drop=ID) out=default;
   grid p1*p2=pi_est pi_real / spline naxis1=100 naxis2=100;
run;

ods listing gpath="&path\Report\figures" image_dpi=300;
ods graphics on / reset=index imagefmt=jpeg;

goptions reset=all border
         gunit=pct htext=4 ftext=swiss ftitle=swissb htitle=6 ctext=black
         device=JPEG xmax=6IN ymax=4.5IN
         interpol=none cback=white gsfname=Fig gsfmode=replace
         colors=(black blue green red);

title1 'p1 & p2 --> pi_est';
title2 "(N = &N)";
proc g3d data=default;
  plot p1*p2=pi_est /
       side
       rotate=135 ctop=blue cbottom=black
       xticknum=5 yticknum=5 zmin=0 zmax=0.80 zticknum=5;
  format p1 p2 4.2 pi_est 5.2;
run;

title1 'p1 & p2 --> pi_real';
title2 "(N = &N)";
proc g3d data=default;
  plot p1*p2=pi_real /
       side
       rotate=135 ctop=blue cbottom=black
       xticknum=5 yticknum=5 zmin=0 zmax=0.80 zticknum=5;
  format p1 p2 4.2 pi_real 5.2;
run;

ods graphics off;

/*********************************End of the Code**********************************/
```