AN ABSTRACT OF THE THESIS OF

Mozhu Liu for the Master of Science

in Mathematics presented on April 17, 2012

Title: Power and Sample Size for some Chi-Square Goodness of Fit Tests

Abstract approved: ______________________________

The first decision a researcher must make is to decide what sample size will be used in the experiment. Many researchers are familiar with sample size issues for the simple t-test, approximate binomial tests, two-sample t-test, and the analysis of variance. However, it is very difficult to find anyone who is familiar with the power and sample size issues for the Chi-Square goodness of fit test.

For example, the first hypothesis examined is testing to see if two binomial proportions are equal. An approximate test of this hypothesis can be conducted using either a z-test or a Chi-Square goodness of fit test. These tests are equivalent tests since $z^2 = \chi^2$. The power of these tests can be approximated by using the standard normal distribution or a non-central Chi-Square distribution. I derived the non-centrality parameter for this test and the other related goodness of fit tests. It is somewhat surprising that the power of the test computed using the standard normal is not identical to the power based on the Chi-Square. Even though the values are not equal they are quite close.

A simulation is also conducted to estimate alpha and power. The results show that the empirical level of significance for the Chi-Square goodness test is close to alpha. It is also seen that simulated powers tend to be quite close to powers computed using the non-central Chi-Square. Some simple iterative programs are included that can be used to

compute the sample size needed to detect a given departure from the null hypothesis with a desired power.

Approved by the Department Chair

Approved by Committee Member

Approved by Committee Member

Approved by Committee Chair

Approved by Dean of the Graduate School and Distance Education

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to my advisor, Dr. Larry Scott, for his patience, encouragement, massive help and expert advice. He spent lots of his time to help me revise and improve my thesis.

I also would like to thank my committee members: Dr. Brian Hollenbeck, and Dr. Daniel J. Miller, for developing my thesis and giving helpful advice.

My thanks also go to my parents and my aunt who have supported me to finish the thesis. I hope they receive my gratitude for their love, their education and encouragement.

Finally, I would like to thank my boyfriend, Zhe Peng, for his support. When I spent a lot of time for the sleepless nights, he always accompanied with me to do the thesis.

PREFACE

I have four chapters in my thesis. The first chapter gives some statistical background and discusses some Chi-Square goodness of fit test. In Chapter 2, the non-centrality parameter is derived for five Chi-Square goodness of fit tests. Chapter 3 gives Calculated and simulated sample sizes for these Chi-Square goodness of fit tests. The last chapter summarizes my conclusions regarding power and the sample size for Chi-Square goodness of fit tests.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

POWER AND SAMPLE SIZE FOR SOME CHI-SQUARE GOODNESS OF FIT TESTS

A Thesis

Presented to

The Department of Mathematics

EMPORIA STATE UNIVERSITY

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

By

Mozhu Liu

April 19, 2012

CHAPTER 1

INTRODUCTION

Many statistical methods assume a probability distribution to derive the methods. One common approach for checking to see if the data follows the specified distributions is a Chi-Square goodness of fit test. My thesis involves the effect of sample size on the power and validity of the Chi-Square goodness of fit test. If sample size is too small, important differences may not be detected or the test may not be valid. However, if sample size is too large, differences may be detected that are too small to be meaningful. In addition, I am investigating how different categorizations of a distribution affect the Chi-Square goodness of fit test.

It can be argued that the most important decision a researcher makes is the decision of what sample size will be used in the experiment. Many statistical methods books such as Zar (1999) discuss this issue for the simple t-test, approximate binomial tests, two-sample t-test, and the analysis of variance. Some analytical work by Patnaik (1949) has been done on the power of the Chi-Square goodness of fit test. O'Brien and Shieh (2000) presented power and sample size calculations for F-tests. However, it is very difficult to find any detailed discussion of the power and sample size issues for the Chi-square goodness of fit test.

The purpose of this thesis is to investigate how large the sample size should be when using the Chi-Square goodness of fit test. I start by looking at computing the power of some Chi-Square goodness of fit tests. Slakter (1968) investigated the power of the Chi-Square goodness of fit test with small expected frequencies. There are a number of other situations where the Chi-Square goodness of fit test can be used. In this study five different hypotheses are examined where a Chi-square goodness of fit test might be applied. For example, the first hypothesis examined to determine if two binomial proportions are equal. This hypothesis can be tested using either a z-test or a Chi-Square goodness of fit test. The power of these tests is computed for several different situations using the standard normal as well as the non-central Chi-Square. A method for computing the non-centrality parameter for the Chi-Square test is provided. It is somewhat surprising that the computed powers for these equivalent tests are approximately equal though not identical. A simulation is also conducted to estimate alpha and power. The results show that the empirical level of significance for the Chi-Square goodness of fit test is close to alpha. In most situations, when using the non-central Chi-Square to compute the power of the test, the simulated powers and the calculated powers tend to be very close to each other.

Simple iterative programs are used to compute the sample size needed to detect departure from the null hypothesis. It should be noted that sample size may be described as total number of experimental units or the number of experimental units per group depending on exactly which null hypothesis is being tested. In general it is best to use an equal number of observations per group. However, in situations where the number of observations in a given group is limited, the sample size needed to obtain the desired power can be found and provided that there is departure from the null hypothesis in the other groups.

Chapter 2 will discuss some statistical background as well as some hypotheses that can be tested by using the Chi-Square goodness of fit test. Methods for calculating power will be investigated. The accuracy of the results will be confirmed using simulations. Chapter 3 will address various questions about sample size.

The goodness of fit test is a statistical procedure to determine whether an assumed distribution is consistent with the data collected. The Chi-Square goodness of fit test is the major topic of this thesis. In addition, case V in Chapter 2 can be looked as an extension of the goodness of fit test. It is frequently referred to as a test of homogeneity. The standard Chi-Square test statistic for the test of homogeneity is

$$\chi^2 = \sum \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2\big((r-1)(k-1)\big),$$

where i, j index the rows and columns of table and $(r - 1)(k - 1)$ are the degrees-of-freedom.

CHAPTER 2

Chi-Square goodness of fit tests are used to test a wide variety of different situations. Five null hypotheses are examined in this thesis:

Case I, $H_0: p_1 = p_2 = \cdots = p_k$, and $\sum p_i \neq 1$. Under this hypothesis there are k binomial populations with each population having the same p. With n_i observations are taken from each population.

Case II, $H_0: p_1 = p_2 = \cdots = p_k = p_0$; Under this hypothesis there are k binomial populations with each population having the same specified p = p_0. With n_i observations are taken from each population.

Case III, $H_0: p_1 = p_2 = \cdots = p_k$, and $\sum p_i = 1$. Under this hypothesis the distribution is multinomial. The n observations are taken from one multinomial population. Cochran (1952) discussed this hypothesis in detail.

Case IV, $H_0: p_1 = p_{1,0},\ p_2 = p_{2,0}, \ldots\ p_k = p_{k,0}$; Under this hypothesis there are k binomial populations with the specified p's or there is a multinomial population with $\sum p_{i,0} = 1$. Broffitt and Randles (1977) approximated the power of this test for small samples.

Case V, H_0: The r samples are homogeneous. Under this hypothesis, the case of interest assumes that the r samples come from the same multinomial distribution.

The purpose of this chapter is to examine the power of the Chi-Square tests used to test each of the above hypotheses. These tests are closely related but important differences are noted. I first calculate the approximate power using analytical methods and then compare these probabilities to simulated powers. The first case examined is a very simple case where the hypothesis can be tested by using either a z-test or a Chi-Square goodness of fit test.

Let $H_0: p_1 = p_2$, $H_A: p_1 \neq p_2$. It is common to use the following z-test as an approximate test statistic for testing this hypothesis. The test statistic can be written in several different algebraic forms as shown below.

$$z = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\frac{\bar{p}\bar{q}}{n}}} = \frac{\frac{x_1}{n} - \frac{x_2}{n}}{\sqrt{\frac{\bar{p}\bar{q}}{n}}} = \frac{x_1 - x_2}{\sqrt{n\bar{p}\bar{q}}} = \frac{x_1 - n\bar{p}}{\sqrt{n\bar{p}\bar{q}}} - \frac{x_2 - n\bar{p}}{\sqrt{n\bar{p}\bar{q}}},$$

Many statistical methods textbooks such as Zar (1999) discuss how to compute power and sample size using this z-test. It is easy to show that the power of this approximate test can be computed as follows:

$$\text{power} = P\left[Z \leq \frac{Z_{\alpha/2}\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}} - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}\right] + P\left[Z \geq \frac{Z_{(1-\frac{\alpha}{2})}\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}} - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}\right]$$

The hypothesis could also be tested using a Chi-Square goodness of fit test. If the above z statistic is squared then Zar (1999) gives the following test statistic.

$$z^2 = \left(\frac{x_1 - n\bar{p}}{\sqrt{n\bar{p}\bar{q}}} - \frac{x_2 - n\bar{p}}{\sqrt{n\bar{p}\bar{q}}}\right)^2 = \frac{(x_1 - n\bar{p})^2 - 2(x_1 - n\bar{p})(x_2 - n\bar{p}) + (x_2 - n\bar{p})^2}{n\bar{p}\bar{q}} = \frac{{x_1}^2 - 2x_1x_2 + {x_2}^2}{n\bar{p}\bar{q}} =$$

$$\frac{(x_1 - x_2)^2}{n\bar{p}\bar{q}} = \sum_{j=1}^{2} \frac{(o_j - e_j)^2}{e_j} = \chi^2(1).$$

It is known that the power of the Chi-Square goodness of fit test can be approximated based on using a non-central χ^2 distribution. In order to do this the non-centrality parameter needs to be found. Patnaik (1949) discussed computing the non-centrality parameter. A convenient way of thinking about this is that when the test statistic is computed under the alternative hypothesis, quantity is being added to the Chi-Square test statistic. That quantity is called the non-centrality parameter ϕ. A modern notation ϕ is $\phi = \sum \frac{(o_j - e_j)^2}{e_j}$, where $e_j = n\bar{p}, \bar{p} = \frac{p_1 + p_2}{2}$, $\bar{q} = 1 - \bar{p}$ and $o_j = np_j$ with j=1, 2. The test statistic is $\sum_{i=1}^{k} \frac{(X_i - n_i\bar{p})^2}{n_i\bar{p}\bar{q}}$. Assume the rejection rule is to reject H_0 if the value of the test statistic is greater than C. To compute the power, the probability is found that this non-central Chi-Square with 1 degree of freedom is greater than C.

I was interested in determining how well these tests work. To examine this question, I started by testing to see if two binomial populations have the same probability of a success. Let $H_0: p_1 = p_2$, $H_A: p_1 \neq p_2$, with $n_1 = 120, n_2 = 120, p_1 = 0.45, p_2 = 0.25, \alpha = 0.05$. For this hypothesis, the calculated power using the approximation for a z-test is 0.90613.

The power using the non-central Chi-Square approximation is 0.90113. These values are very close to each other but it may be somewhat surprising that they differ. I was interested in determining how accurately they reflect the real power of this test. To help answer this question, I wrote a Statistical Analysis System (SAS) program to simulate the power of this test (see Appendix A). I ran the simulation twice, and the results were 0.90636 and 0.90674. Each run was based on 50,000 replications of the test. These values are very close to the calculated power for the z-test and not far from the value calculated for the Chi-Square test.

Next, I wanted to try some more examples to investigate how the calculated power of the z-test approximation and the non-central Chi-Square approximation relate to simulated power. The results are shown in Table 1.

Table 1: Powers with Two Samples based on Non-Central Chi-Square Test and Z-Test

n_1	n_2	p_1	p_2	Power-Chi	Power-z	Power-Simu
50	50	0.25	0.35	0.19360	0.19211	0.19139
75	75	0.25	0.35	0.26692	0.26565	0.26141
100	100	0.25	0.35	0.33859	0.33766	0.33989
150	150	0.25	0.35	0.47210	0.47193	0.47284
200	200	0.25	0.35	0.58794	0.58846	0.58973
300	300	0.25	0.35	0.76197	0.76329	0.76433
400	400	0.25	0.35	0.86994	0.87137	0.87219

After trying a few more examples using the same p_1, p_2, and increasing the sample size n, I saw that the calculated and simulated power also increases, but the calculated powers between the z-test and the Chi-Square test do not show any consistent pattern. To simulate the power, the simulation program was run twice to see the stability of the simulations. Then I checked alpha to see if the simulated alpha is close to the chosen value of 0.05. Letting p_1 and p_2 be the same value, say 0.25, and running the program a few times for various sample sizes (n > 50), the simulated alpha values were either a little larger than 0.05 or a little less than 0.05. For these situations, the test seemed to be very accurate.

Table 2 shows some additional examples where the calculated power is based on the non-central Chi-Square approximation. The test statistic is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - n_i\bar{p})^2}{n_i\bar{p}\bar{q}}$$

Table 2: Powers with Two Samples Binomial Test based on Non-Central Chi-Square Test

$$H_0: p_1 = p_2$$

n_1	n_2	p_1	p_2	$\bar{p}$	Cal. Power	Sim. Power
200	200	0.3	0.4	0.35	0.55435	0.55746
500	500	0.3	0.4	0.35	0.91229	0.91298
200	200	0.35	0.45	0.4	0.53242	0.52796
300	300	0.35	0.45	0.4	0.70542	0.7071
50	45	0.75	0.50	0.625	0.71112	0.71395

This table shows that the simulated powers are very close to the calculated powers. Note that the sample size does not have to be equal, but equal sample size will lead to more powerful tests.

Now, consider an experiment where there are k binomial populations. The null hypothesis is, $H_0: p_1 = p_2 = \cdots = p_k$, and the alternative hypothesis is that at least one of the p_i's is different from the others. The z-test can no longer be used to test this hypothesis, but the Chi-Square test can still be used. The non-centrality parameter is now given as:

$\phi = \sum_{j=1}^{2k} \frac{(o_j - e_j)^2}{e_j}$ where $\bar{p} = \sum_{j=1}^{k} \frac{p_j}{k}$, $\bar{q} = 1 - \bar{p}$, $e_j = \begin{cases} n\bar{p} & j = 1, 2, \ldots, k \\ n\bar{q} & j = k+1, \ldots, 2k \end{cases}$ and $o_j = \begin{cases} np_j & j = 1, 2, \ldots, k \\ nq_j & j = k+1, \ldots, 2k \end{cases}$. The test statistic is $\chi^2 = \sum_{j=1}^{k} \frac{(X_j - n_j\bar{p})^2}{n_j\bar{p}\bar{q}} = \sum_{j=1}^{2k} \frac{(X_j - e_j)^2}{e_j}$ with k-1 degrees of freedom. Assume the rejection rule is to reject H_0 if the value of the test statistic is greater than C. To compute the power, it is necessary to find the probability that this non-central Chi-Square with k-1 degree of freedom is greater than C.

Suppose the sample size is 200, and $p_1 = 0.3, p_2 = 0.4, p_3 = 0.35, \bar{p} = 0.35$. So the non-centrality parameter is $\Phi = \frac{(60-70)^2}{70} + \frac{(80-70)^2}{70} + \frac{(70-70)^2}{70} + \frac{(140-130)^2}{130} + \frac{(120-130)^2}{130} + \frac{(130-130)^2}{130} = 4.39560$. Then the power is 0.45116. If $\alpha = 0.05$, the critical value for this test is 7.815, and the power can be approximated by the probability that the non-central Chi-Square is greater than 7.875. Using SAS to compute this value gives 0.45116.

Table 3 shows some additional examples where the calculated power is based on the non-central Chi-Square approximation. Zar (1999) gave the following test statistic as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - n_i\bar{p})^2}{n_i\bar{p}\bar{q}}.$$

Table 3: Powers with Three Samples based on Non-Central Chi-Square Test

$$H_0: p_1 = p_2 = p_3$$

n_1	n_2	n_3	p_1	p_2	p_3	$\bar{p}$	Calculated Power	Simulated Power
200	200	200	0.3	0.4	0.35	0.35	0.45116	0.45259
500	500	500	0.3	0.4	0.35	0.35	0.85224	0.85658
200	200	200	0.25	0.35	0.5	0.4	0.99821	0.99856
300	300	300	0.25	0.35	0.5	0.4	0.99997	0.99999

This table shows that the simulated powers are very close to the calculated powers. Note that the sample sizes for the k binomial populations do not have to be equal, but care must be taken when drawing conclusions about power when at least one $p_i = \bar{p}$.

The second hypothesis suggested in the beginning of Chapter 2 is $H_0: p_1 = p_2 = \cdots = p_k = p_0$ with the alternative hypothesis stating that at least one $p_i \neq p_0$. This hypothesis can be tested using a Chi-Square goodness of fit type test. It should be noted that the big difference between this test and the previous test is that this test has k degrees of freedom. The non-centrality parameter is $= \sum_{j=1}^{2k} \frac{(o_j - e_j)^2}{e_j}$,

where $e_j = \begin{cases} np_0 & j = 1, 2, \ldots, k \\ nq_0 & j = k+1, \ldots, 2k \end{cases}$ and $o_j = \begin{cases} np_j & j = 1, 2, \ldots, k \\ nq_j & j = k+1, \ldots, 2k \end{cases}$. The test statistic is

$\chi^2 = \sum_{j=1}^{k} \frac{(X_j - n_j p_0)^2}{n_j p_0 q_0} = \sum_{j=1}^{2k} \frac{(X_j - e_j)^2}{e_j}$ with k degrees of freedom. Assume the rejection rule is to reject H_0 if the value of the test statistic is greater than C. To compute the power, it is necessary to find the probability that this non-central Chi-Square with k degree of freedom is greater than C. For example, if $H_0: p_1 = p_2 = p_0$, and $n_1 = 200, n_2 = 200, p_0 = 0.4, p_1 = 0.3, p_2 = 0.4$, then the non-centrality parameter is $\phi = \frac{(60-80)^2}{80} + \frac{(140-120)^2}{120} + \frac{(80-80)^2}{80} + \frac{(120-120)^2}{120} = 5 + 3.33 = 8.33$. Using $\alpha = 0.05$, the rejection rule for this test is to reject the null hypotheses if the test statistic is greater than 7.815. The power can be approximated by the probability that the non-central Chi-Square is greater than 7.875. Using SAS to compute this value we get 0.73623. Using the SAS program provided in the Appendix C, the simulated power is 0.7594. The calculated and simulated powers are close. Some additional examples are given in Table 4. Zar (1999) also gave the following test statistic as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - n_i p_0)^2}{n_i p_0 (1 - p_0)}.$$

Table 4: Powers with Two and Three Samples

$$H_0: p_1 = p_2 = p_0$$

n_1	n_2	p_1	p_2	p_0	Cal. Power	Sim. Power
200	200	0.3	0.4	0.4	0.73623	0.7594
500	500	0.3	0.4	0.4	0.98827	0.99166
200	200	0.35	0.45	0.4	0.43060	0.44129
300	300	0.35	0.45	0.4	0.60275	0.60246

$$H_0: p_1 = p_2 = p_3 = p_0$$

n_1	n_2	n_3	p_1	p_2	p_3	p_0	Cal. Power	Sim. Power
200	200	200	0.3	0.4	0.4	0.4	0.67398	0.68262
500	500	500	0.3	0.4	0.4	0.4	0.97981	0.98448
200	200	200	0.35	0.35	0.5	0.4	0.85617	0.85222
300	300	300	0.35	0.35	0.5	0.4	0.96594	0.94179

By running the simulation when the null hypothesis is true, and getting a relative frequency of rejecting the null hypothesis approximately equal to $\alpha = 0.05$, tends to verify the validity test as well as the validity of the simulation program. From the examples, I put all the p's equal to 0.4 and ran the simulations. I saw that the relative frequency of rejecting the null hypothesis is very close to 0.05; sometimes it is bigger and sometimes it is smaller. So this approximate test performs well in these situations.

There were three other hypotheses mentioned in this chapter. For these three hypotheses, I assumed that the sample or samples came from a multinomial distribution. For case III, the null hypothesis is $H_0: p_1 = p_2 = \cdots = p_k$, where $\sum p_i = 1$. Under this hypothesis the distribution is multinomial. This hypothesis is another example where a Chi-Square goodness of fit type test is commonly used. Slakter (1968) looked at approximating the power of this test. The non-centrality parameter is $\phi = \sum \frac{(o_j - e_j)^2}{e_j}$, where $e_j = np$, $p = \frac{1}{k}$, and $o_j = np_j$ with $j = 1, 2, \cdots, k$. The test statistic is $\chi^2 = \sum_{i=1}^{k} \frac{(X_i - n_i p)^2}{n_i p(1-p)}$. Assume the rejection rule is to reject H_0 if the value of the test statistic is greater than C. To compute the power, it is necessary to find the probability that this non-central Chi-Square with k-1 degrees of freedom is greater than C. First, I checked to see if the relative frequency of rejecting a true null hypothesis is close to 0.05. I ran two simulations, and I got 0.04684 and 0.04728. Each run was based on 100,000 replications of the test. Then I changed the values of p to reflect a false null hypothesis. The true values of p were denoted by p_{1j}, where j denotes the group. Some simulated and calculated powers are given in Table 5:

Table 5: Calculated and Simulated Powers

p_{11}	p_{12}	p_{13}	p_{14}	n	Cal. Power	Sim. Power
0.20	0.25	0.25	0.30	100	0.19224	0.18871
0.20	0.25	0.25	0.30	150	0.27464	0.26886

0.20	0.25	0.25	0.30	200	0.35853	0.35870
0.15	0.25	0.25	0.35	200	0.93409	0.94352
0.10	0.20	0.30	0.40	100	0.97507	0.98540

When the values of p are constant, increasing the sample size should increase the power of the test, and it does for the examples given here. It is hard to tell if very small increases in n would increase real power. It certainly increased calculated power; when the differences between values of p were increased, the power of the test was also increased.

The fourth hypothesis studied in this Chapter is $H_0: p_1 = p_{1,0},\ p_2 = p_{2,0}, \ldots\ p_k = p_{k,0}$. Under this hypothesis there are k binomial populations with the specified p's, or there is a multinomial population with $\sum p_{i,0} = 1$. This hypothesis can be simulated by binomial or multinomial distributions. The multinomial case is the only case studied in this thesis. Assume $H_0: p_1 = 0.1, p_2 = 0.4, p_3 = 0.4, p_4 = 0.1,\ p_{1,0} = 0.13,\ p_{2,0} = 0.37,\ p_{3,0} = 0.37, p_{4,0} = 0.13$ where $p_{i,0}$ is the true value of p_i. Table 6 gives the calculated power and simulated power for several different sample sizes with $\alpha = 0.05$.

Table 6: Calculated and Simulated Powers

n	Cal Power	Simulated Power		
150	0.25847	0.22742	0.22906	0.22970
200	0.33697	0.31259	0.31425	0.31258
300	0.48873	0.48115	0.48218	0.48230
500	0.72979	0.74164	0.73974	0.74108

When I checked for alpha for a few times based on 100,000 replications, the results showed 0.0495, 0.05022, 0.04962, 0.05124, 0.04864 and so on. The simulated alpha's were close to the chosen alpha of 0.05.

The fifth hypothesis studied is H_0: The r samples are homogeneous. Under this hypothesis the r samples all come from the same multinomial distribution. This test can be looked at as an extension of the Chi-Squared goodness of fit test. It is usually called a Chi-Square test of homogeneity.

Suppose a researcher is interested in testing to see if the age structure of a deer population is changing over time. Deer are aged, put into one of four classes. And data were collected for three years. I tried to compare the calculated and simulated powers of the test of homogeneity. First, using an SAS simulation for alpha, I gave all three years the same age distribution with the sample size n=200. Then I ran the program and got a simulated alpha of 0.04862. It was not exactly 0.05, but it is close enough to assume that the level of significance is close to the chosen value of alpha.

Table 7: the Values of p

	1	2	3	4
1	0.30	0.25	0.25	0.20
2	0.32	0.27	0.23	0.18
3	0.35	0.30	0.20	0.15

I wanted to compare the calculated power to the simulated power. A program is given in the Appendix A to calculate the power of this test based on the non-central Chi-

Square distribution. Letting alpha equal 0.05 and using the probabilities given in Table 7, I got the calculated power equal to 0.29323. Another SAS program is provided to simulate the power of this test. The simulated power turned out to be 0.28502. The simulated power was a little smaller than the calculated power.

Since I needed more examples to draw conclusions, I chose a new set of values for p. These values are given in Table 8.

Table 8: Another Example for the Values of p

	1	2	3	4
1	0.30	0.25	0.25	0.20
2	0.35	0.30	0.20	0.15
3	0.40	0.35	0.15	0.10

Using the SAS program, I got a calculated power of 0.92937. Then I ran the other program a few times to see the simulated power based on 50,000 replications. I saw the simulated powers are 0.91272, 0.91232, 0.9129 and 0.91448; they were all around 0.91, which is smaller than the calculated power. When I simulated alpha for this test of homogeneity, I got 0.04954 and 0.0512. The simulated alpha values are about what I would expect for a valid test. As expected, when the differences in the values of p are increased, both the calculated and simulated powers also increase.

CHAPTER 3

This chapter focuses on finding the sample size needed to design an experiment with a given power. The investigator needs to be able to specify the differences that should be detected and the desired power. Again, suppose there are two binomial populations. Let $H_0: p_1 = p_2$, $H_A: p_1 \neq p_2$. Zar (2010) gives the approximate solution for sample size as $n = \frac{(z_\alpha + z_{\beta(1)})^2}{2(\arcsin\sqrt{p_1} - \arcsin\sqrt{p_2})^2}$. Another example is examined. Let $H_0: p_1 = p_2$, $H_A: p_1 \neq p_2$ with $\alpha = 0.05$. Suppose the researcher would like to have a 90% chance of detecting a difference if $p_1 = 0.45$ and $p_2 = 0.25$. This gives, $\bar{p} = \frac{0.45+0.25}{2} = 0.35$, $\bar{q} = 1 - 0.35 = 0.65$, $z_{1-\frac{\alpha}{2}} = 1.96$, $z_{1-\beta} = 1.2816$. Using the above equation shows n = 117.44. Thus, the sample size needs to be at least 118. This can be checked using the simulation program that is attached to this thesis (see Appendix C). Three runs of this simulation program showed an average power of approximately 0.902. That means the power is a little higher than the desired power when n=118. It should be noted that the sample size was rounded up, which should increase the power slightly.

An iterative SAS program is attached to this thesis that can be used to determine the sample size based on the power approximated using a non-central χ^2. This program gives n = 120. Simulations using n = 120 yield a power of approximately 0.9055. If a researcher needs the sample sizes to test $H_0: p_1 = p_2$, $H_A: p_1 \neq p_2$ with $\alpha = 0.05$, $p_1 = 0.75$, and $p_2 = 0.50$ and desired power of at least 0.80; the SAS programs provided in the Appendix B that can be used to find the minimum sample size needed based on approximating the power by using a non-central Chi-Square and for the approximation based on the z-test. The non-central Chi-Square approach yields n = 59, while the z-test approximation yields n = 58. These two results are not the same, but they are very close. Table 9 gives the sample size results for some other values of p_1, p_2 and desired power.

Table 9: Hypothesis I for the Sample Sizes

p_1	p_2	power	n-chi	n-z test
0.75	0.5	0.8	59	58
0.5	0.6	0.7	306	305
0.5	0.6	0.8	389	388
0.5	0.6	0.9	521	519
0.2	0.3	0.6	184	183
0.2	0.3	0.7	232	230
0.45	0.55	0.7	309	308

These results showed computing the power based on the non-central Chi-Square test will yield a sample size that is 1 or 2 units bigger than using the approximation based on the z-test. It is not surprising that it takes a larger sample size to detect a specific difference when p is close to 0.5 than it does when p is far from 0.5.

The following work gives some results when there are three binomial populations involved in the experiment. In this case, the approximation based on the z-test can no longer be used. An SAS program for the case involving three binomial populations is provided (see Appendix B). Simulations were run and indicated that the non-central Chi-Square approach works well for this case. When I took the calculated power of 0.75 with the values of p being 0.28, 0.38 and 0.35, the sample size is 365. Then I ran the SAS program to simulate the power and the sample size for a few times, and I saw that the simulated power is around 0.756 with the sample size 365.

Another SAS program is given in the Appendix B to calculate the sample size for Case II. With the desired power of 0.65 and $p_1 = 0.31, p_2 = 0.39, p_3 = 0.42, p_0 = 0.4$, I got the sample size of n = 222. When I simulated to investigate the power with the computed sample size, I found that the simulated power was around of 0.67, which is slightly bigger than the desired power.

For case III, I tried to get the sample size with the given power and values of p. Table 10 gives n for some related values of p and desired power.

Table 10: Hypothesis III for the Sample Sizes

p_{11}	p_{12}	p_{13}	p_{14}	Power	n
0.20	0.25	0.25	0.30	0.7	440
0.20	0.25	0.25	0.30	0.8	546
0.20	0.25	0.25	0.30	0.9	709
0.15	0.25	0.25	0.35	0.6	90
0.15	0.25	0.25	0.35	0.7	110
0.15	0.25	0.25	0.35	0.8	137
0.15	0.25	0.25	0.35	0.9	178
0.10	0.20	0.30	0.40	0.9	71

These values of n are supported by simulations. The values vary as might be expected for the given values of p and power. As the values of p get farther apart, the sample size decreases. As the desired power is increased, the sample size must be increased.

In Case IV, I used the same values of p as the ones used in Chapter 2. Let $p_1 = 0.1$, $p_2 = 0.4$, $p_3 = 0.4$, $p_4 = 0.1$, $p_{1,0} = 0.13$, $p_{2,0} = 0.37$, $p_{3,0} = 0.37, p_{4,0} = 0.13$. Then I used the desired powers of 0.5, 0.6, 0.7 and 0.9. I got the sample sizes of 308, 383, 470 and 758. In Chapter 2, a sample size of 300 was specified, and a calculated power of 0.48873 was obtained. In Chapter 3, a desired power of 0.5 was specified, and a sample size of 308 was obtained. These results make a lot of sense.

For Case V, I used another SAS program to get the sample size needed for this hypothesis (see Appendix B). The results look good. I tried to look at the case where one of the samples had a fixed size and wanted to find the other sample sizes. This situation may be very tricky and I need to do more in the future study for this part.

First, I made the values of p table, since I used the same group of p values with different powers to find sample size (see Table 11).

Table 11: the Values of p

p_{11}	p_{12}	p_{13}	p_{14}
0.30	0.25	0.25	0.20
p_{21}	p_{22}	p_{23}	p_{24}
0.32	0.27	0.23	0.18
p_{31}	p_{32}	p_{33}	p_{34}
0.35	0.30	0.20	0.15

Then, I ran the program with different powers to get the sample size with

$n_1 = n_2 = n_3$.

Table 12: Hypothesis V for the Sample Sizes

Power	n_1	n_2	n_3	Sim-Power
0.3	205	205	205	0.30045
0.5	345	345	345	0.50061
0.6	422	422	422	0.60014
0.7	512	512	512	0.70042
0.8	626	626	626	0.80028
0.9	800	800	800	0.90005

For this hypothesis, the sample size increased as the power increased. The simulated powers are also close to the real power.

CHAPTER 4

CONCLUSION

Since I did this research about sample size and power based on the Chi-Square goodness of fit tests, I realize there is a close connection among various tests. I also realize that there are important differences among these tests. I have investigated five null hypotheses and used simulations to confirm my results. Some of these five hypotheses assume the data come from a binomial distribution, some assume the data come from a multinomial distribution, and some can be applied to data that come from either a binomial or multinomial distribution. I found a simple way to calculate the non-centrality parameter for each of these tests and used it to compute the power of the tests. Simulations seem to confirm that the computed power is very close to the actual power.

When the sample size is increased the power is expected to increase. This will always occur with larger increases in sample size, but it may not always occur for small increases in sample size when using approximate test like the Chi-Square goodness of fit test. I have not attempted to investigate this issue in this thesis. It would be a good topic for future research.

In Chapter 2, I examined the power of the Chi-Square tests and used it to test each of the above hypotheses with a given sample size. In Chapter 3, I did it the opposite way, using the given power to find the sample size. When testing H_0: $p_1 = p_2$, the approximate power of the test can be calculated using either the non-central Chi-Square or the standard normal. Intuitively, I expected these two approximations to be the same. It turns out that they are close but not equal. The question of why they are not equal might be an interesting topic for future research. An SAS program was written to simulate the power of this test (see Appendix C). The result showed the simulated power was either bigger or smaller than the calculated power, but they were always close. I also did the simulation for checking alpha to see if the actual level of significance of the test was 0.05. The result showed it was a little bit bigger or smaller than 0.05. The test seems to be valid but it should still be viewed as an approximate test.

These results were generalized and applied to the other tests of interest. Simulations were used to confirm the results. All of these tests seem to be valid but they should still be viewed as approximate tests. In conclusion, the non-central Chi-Square distribution gives us a good method for determining the sample size needed when using a Chi-Square goodness of fit test.

BIBLIOGRAPHY

Broffitt, James D. and Ronald H, Randles. "A Power Approximation for the Chi-Square goodness-of-Fit Test: Simple Hypothesis Case." *Journal of the American Statistical Association* 72 (1977): 604-4.

Cochran, William G. "The χ^2 Test of Goodness of Fit." *Annals of Mathematical Statistics* 23 (1952): 315-45.

O'Brien, R.G. and G.Shieh. *Pragmatic, Unifying Algorithm Gives Power Probabilities for Common F Tests of the Multivariate General Linear Hypothesis.* 2000. www.bio.ri.ccf.org/UnifyPow.

Patnaik, P. B. "The noncentral χ^2- and F-distributions and their applications." *Biometrika* 36 (1949): 202-32.

Slakter, Malcolm J. "Accuracy of an Approximation to the Power of the Chi-Square Goodness-of-Fit Test with Small but Equal Expected Frequencies." *Journal of the American Statistical Association* 63 (1968): 912-8.

Zar, Jerrold H. *Biostatistical Analysis*. Upper Saddle River, NJ: Prentice Hall, 1999. Print.

APPENDICES

Appendix A

This program can be used to compute the power for Case I.

```
Data H1power;
Input n p;
q = 1-p;
Datalines;
200 0.3
200 0.4
200 0.35
Proc Means; Var p q;
Output Out=Stat Sum=Sump Sumq n=k;
Data Stat; Set Stat;
df1 = k-1;
pbar = Sump/k;
qbar = Sumq/k;
Do I = 1 to k;
Output;
End;
Proc Print;
Data Stat1; Merge H1power Stat;
o = n*p; e = n*pbar; phi = (o-e)**2/e; Output;
o = n*q; e = n*qbar; phi = (o-e)**2/e; Output;
Proc Print;
Proc Means; Var phi; Id df1;
Output Out=Stat2 Sum=phi;
Data Final; Set Stat2;
ChiSq = cinv(0.95, df1);
Power = 1-ProbChi(ChiSq, df1,phi);
Proc Print; Var df1 phi power;
Run;
```

This program can be used to compute the power for Case II.

```
Data H2CalP;
Input n p;
p0 = 0.4;
q0 = 1-p0;
q = 1-p;
Datalines;
300 0.35
300 0.35
300 0.5
Proc Print;
Data Try;
Data Stat1; Set H2CalP;
o = n*p; e = n*p0; phi = (o-e)**2/e; Output;
o = n*q; e = n*q0; phi = (o-e)**2/e; Output;
Proc Print;
Proc Means; Var phi; Id p0;
Output Out=Stat2 Sum=phi n=k;
Data Final; Set Stat2;
df1=k/2;
ChiSq = cinv(0.95, df1);
Power = 1-ProbChi(ChiSq, df1,phi);
Proc Print; Var p0 df1 phi power;
Run;
```

This program can be used to compute the power for Case III.

```
Data H3CalP;
Input p @@;
n=100;
k=4;
pk = 1/k;;
Datalines;
.1 .2 .3 .4
*Proc Print;
Data Try;
Data Stat1; Set H3CalP;
o = n*p; e = n*pk; phi = (o-e)**2/e; Output;
*Proc Print;
Proc Means NoPrint;     Var phi; Id n k pk;
Output Out=Stat2 Sum=phi;
Data Final; Set Stat2;
df1=k-1;
ChiSq = cinv(0.95, df1);
Power = 1-ProbChi(ChiSq, df1,phi);
Proc Print; Var n k pk df1 phi power;
Run;
```

This program can be used to compute the power for Case IV.

```
Data H4CalP;
n = 200;
p1 = .10; p2 = .4; p3 = .4; p4 = .10;
p10 = .13; p20 = .37; p30 = .37; p40 = .13;
e1 = n*p10;
e2 = n*p20;
e3 = n*p30;
e4 = n*p40;
o1 = n*p1;
o2 = n*p2;
o3 = n*p3;
o4 = n*p4;
df1=3;
phi = (o1-e1)**2/e1 + (o2-e2)**2/e2 + (o3-e3)**2/e3 + (o4-e4)**2/e4;
ChiSq = cinv(0.95, df1);
Power = 1-ProbChi(ChiSq, df1,phi);
Proc Print; Var p1 p2 p3 p4 p10 p20 p30 p40 df1 phi power;
Run;
```

This program can be used to compute the power for Case V.

```
Data H5CalP;
n=200;
df = 6;
Input Year Age p @@;
e = n*p;
Datalines;
1 1 .30 1 2 .25 1 3 .25 1 4 .20
2 1 .35 2 2 .30 2 3 .20 2 4 .15
3 1 .40 3 2 .35 3 3 .15 3 4 .10
Proc Sort; By Age;
Proc Print;
Proc Means; By Age; Var p;
Output Out=Stat mean=pbar; Id n df;
Data Stat1; Set Stat;
Do I = 1 to 3;
   o = n*pbar;
Output;
End;
Proc Print;
Data Stat2; Merge Homog Stat1;
C = (o - e)**2/e;
Proc Print;
Proc means; Var C;
Output Out=Stat3 Sum=phi; Id df;
Data Stat4; Set Stat3;
ChiSq = cinv(0.95, df);
Power = 1-ProbChi(ChiSq, df,phi);
Proc Print; Var df phi power;
Run;
```

Appendix B

This program can be used to calculate the sample size for Case I with two binomial populations.

```
Data H1CalcN;
n = 20;
Power = 0.9;
p1 = 0.25;
q1 = 1-p1;
p2 = 0.45;
q2 = 1-p2;
pbar = (p1+p2)/2;
qbar = (q1+q2)/2;
o1 = n*p1;
o2 = n*p2;
o3 = n*q1;
o4 = n*q2;
e1 = n*pbar;
e2 = n*pbar;
e3 = n*qbar;
e4 = n*qbar;
df1=1;
phi = (o1-e1)**2/e1 + (o2-e2)**2/e2 + (o3-e3)**2/e3 + (o4-e4)**2/e4;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
Do while (P < Power);
n = n+1;
o1 = n*p1;
o2 = n*p2;
o3 = n*q1;
o4 = n*q2;
e1 = n*pbar;
e2 = n*pbar;
e3 = n*qbar;
e4 = n*qbar;
df1=1;
phi = (o1-e1)**2/e1 + (o2-e2)**2/e2 + (o3-e3)**2/e3 + (o4-e4)**2/e4;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
End;
Proc Print; Var n p1 p2 df1 phi power P;
Run;
```

This program can be used to calculate the sample size for Case II with k=3.

```
Data H2CalN;
n1 = 20; n2 = n1; n3 = n2;
Power = 0.65;
p1 = 0.31;
q1 = 1-p1;
p2 = 0.39;
q2 = 1-p2;
p3 = 0.42;
q3 = 1-p3;
p0 = 0.4;
q0 = 1-p0;
o1 = n1*p1;
o2 = n2*p2;
o3 = n3*p3;
o4 = n1*q1;
o5 = n2*q2;
o6 = n3*q3;
e1 = n1*p0;
e2 = n2*p0;
e3 = n3*p0;
e4 = n1*q0;
e5 = n2*q0;
e6 = n3*q0;
df1=3;
phi = (o1-e1)**2/e1 + (o2-e2)**2/e2 + (o3-e3)**2/e3 + (o4-e4)**2/e4
      + (o5-e5)**2/e5 + (o6-e6)**2/e6;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
Do while (P < Power and n2 < 5000);
n1 = n1+1;n2 = n1;n3 = n2;
o1 = n1*p1;
o2 = n2*p2;
o3 = n3*p3;
o4 = n1*q1;
o5 = n2*q2;
o6 = n3*q3;
e1 = n1*p0;
e2 = n2*p0;
e3 = n3*p0;
e4 = n1*q0;
e5 = n2*q0;
e6 = n3*q0;
ph1 = (o1-e1)**2/e1;
ph2 = (o2-e2)**2/e2;
ph3 = (o3-e3)**2/e3;
ph4 = (o4-e4)**2/e4;
ph5 = (o5-e5)**2/e5;
ph6 = (o6-e6)**2/e6;
phi = ph1+ph2+ph3+ph4+ph5+ph6;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
End;
Proc Print; Var n1 n2 n3 p0 p1 p2 p3 df1 phi power P;
Run;
```

This program can be used to calculate the sample size for Case III.

```
Data H3CalN;
n = 20;
k = 4;
Power = 0.934;
pk = 1/k;
p1 = .15; p2 = .25; p3 = .25; p4 = .35;
df1=k-1;
e1 = n*pk;
e2 = n*pk;
e3 = n*pk;
e4 = n*pk;
o1 = n*p1;
o2 = n*p2;
o3 = n*p3;
o4 = n*p4;
phi = (o1-e1)**2/e1 + (o2-e2)**2/e2 + (o3-e3)**2/e3 + (o4-e4)**2/e4;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
Do while (P < Power);
n = n+1;
e1 = n*pk;
e2 = n*pk;
e3 = n*pk;
e4 = n*pk;
o1 = n*p1;
o2 = n*p2;
o3 = n*p3;
o4 = n*p4;
phi = (o1-e1)**2/e1 + (o2-e2)**2/e2 + (o3-e3)**2/e3 + (o4-e4)**2/e4;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
End;
Proc Print; Var p1 p2 p3 p4 pk df1 phi n power p;
Run;
```

This program can be used to calculate the sample size for Case IV.

```
Data H4CalcN;
n = 20;
Power = 0.9;
p1 = .10; p2 = .4; p3 = .4; p4 = .10;
p10 = .13; p20 = .37; p30 = .37; p40 = .13;
df1=3;
e1 = n*p10;
e2 = n*p20;
e3 = n*p30;
e4 = n*p40;
o1 = n*p1;
o2 = n*p2;
o3 = n*p3;
o4 = n*p4;
phi = (o1-e1)**2/e1 + (o2-e2)**2/e2 + (o3-e3)**2/e3 + (o4-e4)**2/e4;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
Do while (P < Power);
n = n+1;
e1 = n*p10;
e2 = n*p20;
e3 = n*p30;
e4 = n*p40;
o1 = n*p1;
o2 = n*p2;
o3 = n*p3;
o4 = n*p4;
phi = (o1-e1)**2/e1 + (o2-e2)**2/e2 + (o3-e3)**2/e3 + (o4-e4)**2/e4;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
End;
Proc Print; Var p1 p2 p3 p4 p10 p20 p30 p40 df1 phi n power p;
Run;
```

This program can be used to calculate the sample size for Case V.

```
Data H5CalcN;
n1 = 20; n2 = 20; n3 = 20;
df1=6; Power = 0.9;
p11 = .30; p12 = .25; p13 = .25; p14 = .20;
p21 = .35; p22 = .30; p23 = .20; p24 = .15;
p31 = .40; p32 = .35; p33 = .15; p34 = .10;
p1b = (p11+p21+p31)/3; p2b = (p12+p22+p32)/3;
p3b = (p131+p23+p33)/3;p4b = (p14+p24+p34)/3;
e11 = n1*p11; e12 = n1*p12; e13 = n1*p13; e14 = n1*p14;
e21 = n2*p21; e22 = n2*p22; e23 = n2*p23; e24 = n2*p24;
e31 = n3*p31; e32 = n3*p32; e33 = n3*p33; e34 = n3*p34;
o11 = n1*p1b; o12 = n1*p2b; o13 = n1*p3b; o14 = n1*p4b;
o21 = n2*p1b; o22 = n2*p2b; o23 = n2*p3b; o24 = n2*p4b;
o31 = n3*p1b; o32 = n3*p2b; o33 = n3*p3b; o34 = n3*p4b;
phi = (o11-e11)**2/e11 + (o12-e12)**2/e12 + (o13-e13)**2/e13 + (o14-
e14)**2/e14 +
      (o21-e21)**2/e21 + (o22-e22)**2/e22 + (o23-e23)**2/e23 + (o24-
e24)**2/e24 +
      (o31-e31)**2/e31 + (o32-e32)**2/e32 + (o33-e33)**2/e33 + (o34-
e34)**2/e34;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
Do while (P < Power);
n1 = n1+1; n2 = n2+1; n3 = n3+1;
p1b = (p11+p21+p31)/3; p2b = (p12+p22+p32)/3;
p3b = (p13+p23+p33)/3;p4b = (p14+p24+p34)/3;
e11 = n1*p11; e12 = n1*p12; e13 = n1*p13; e14 = n1*p14;
e21 = n2*p21; e22 = n2*p22; e23 = n2*p23; e24 = n2*p24;
e31 = n3*p31; e32 = n3*p32; e33 = n3*p33; e34 = n3*p34;
o11 = n1*p1b; o12 = n1*p2b; o13 = n1*p3b; o14 = n1*p4b;
o21 = n2*p1b; o22 = n2*p2b; o23 = n2*p3b; o24 = n2*p4b;
o31 = n3*p1b; o32 = n3*p2b; o33 = n3*p3b; o34 = n3*p4b;
phi = (o11-e11)**2/e11 + (o12-e12)**2/e12 + (o13-e13)**2/e13 + (o14-
e14)**2/e14 +
      (o21-e21)**2/e21 + (o22-e22)**2/e22 + (o23-e23)**2/e23 + (o24-
e24)**2/e24 +
      (o31-e31)**2/e31 + (o32-e32)**2/e32 + (o33-e33)**2/e33 + (o34-
e34)**2/e34;
ChiSq = cinv(0.95, df1);
P = 1-ProbChi(ChiSq, df1,phi);
End;
Proc Print; Var p11--p14;
Proc Print; Var p21--p24;
Proc Print; Var p31--p34;
Proc Print; Var df1 phi n1 n2 n3 power p;
Run;
```

Appendix C

Simulates power for Case I testing p1 = p2 = p3.

```
Data H1SimPow;
n1 = 365;
n2 = n1;
n3 = n1;
p1 = 0.28;
p2 = 0.38;
p3 = 0.35;
Reps = 100000;
alpha = 0.05;
Do K =  1 to Reps;
  x1 = RanBin(0, n1, p1);
  p1hat = x1/n1;
  y1 = n1-x1;
  x2 = RanBin(0, n2, p2);
  p2hat = x2/n2;
  y2 = n2-x2;
  x3 = RanBin(0, n1, p3);
  p3hat = x3/n3;
  y3 = n3-x3;
pbar = (x1 + x2 + x3)/(n1 + n2 + n3);
qbar = 1-pbar;
e1 = n1*pbar;
e2 = n2*pbar;
e3 = n3*pbar;
e4 = n1*qbar;
e5 = n2*qbar;
e6 = n3*qbar;
Output;
End;
*Proc Print;
Data Stat; Set H1SimPow;
ChiSq = ((x1-e1)**2)/e1 + ((x2-e2)**2)/e2 +((x3-e3)**2)/e3 +
        ((y1-e4)**2)/e4 + ((y2-e5)**2)/e5 + ((y3-e6)**2)/e6;
C1 = (x1-n1*pbar)**2/(n1*pbar*(1-pbar)) + (x2-n2*pbar)**2/(n2*pbar*(1-
pbar))
     +(x3-n3*pbar)**2/(n3*pbar*(1-pbar));
pvalue = 1-ProbChi(ChiSq, 2);
prob1 =  1-ProbChi(C1, 2);
If pvalue < alpha then Reject+1;
If prob1 < alpha then Ch+1;
Data Final; Set Stat;
If _n_ = Reps;
Power = Reject/Reps;
Power1 = Ch/Reps;
Proc Print;Var alpha n1 n2 n3 p1 p2 p3 ChiSq C1 Power Power1;
Run;
```

Case II simulation of power for Testing p1 = p2 = p3 + p0.

```
Data H2SimP;
n1 = 222;
n2 = n1;
n3 = n1;
p1 = 0.31;
p2 = 0.38;
p3 = 0.42;
p0 = 0.40;
q0 = 1-p0;
Reps = 100000;
alpha = 0.05;
Do K =  1 to Reps;
  x1 = RanBin(0, n1, p1);
  p1hat = x1/n1;
  y1 = n1-x1;
  x2 = RanBin(0, n2, p2);
  p2hat = x2/n2;
  y2 = n2-x2;
  x3 = RanBin(0, n1, p3);
  p3hat = x3/n3;
  y3 = n3-x3;
e1 = n1*p0;
e2 = n2*p0;
e3 = n3*p0;
e4 = n1*q0;
e5 = n2*q0;
e6 = n3*q0;
Output;
End;
*Proc Print;
Data Stat; Set H2SimP;
ChiSq = ((x1-e1)**2)/e1 + ((x2-e2)**2)/e2 +((x3-e3)**2)/e3 +
        ((y1-e4)**2)/e4 + ((y2-e5)**2)/e5 + ((y3-e6)**2)/e6;
C1 = (x1-n1*p0)**2/(n1*p0*(1-p0)) + (x2-n2*p0)**2/(n2*p0*(1-p0)) +
     (x3-n3*p0)**2/(n3*p0*(1-p0));
pvalue = 1-ProbChi(ChiSq, 3);
prob1 =  1-ProbChi(C1, 3);
If pvalue < alpha then Reject+1;
If prob1 < alpha then Ch+1;
Data Final; Set Stat;
If _n_ = Reps;
Power = Reject/Reps;
Power1 = Ch/Reps;
Proc Print;Var alpha n1 n2 n3 p1 p2 p3 p0 ChiSq C1 Power Power1;
Run;
```

Simulates the power for Case III.

```
Data H3SimPow;
Reps = 10000;
n=100;
alpha = 0.05;
p11 = .1; p12 = .2; p13 = .3; p14 = .4;
Do K =  1 to Reps;
Do Sample = 1 to n;
  Age = RanTBL(0, p11, p12, p13, p14);
Output;
End;
End;
*Proc Print;
Proc Freq NoPrint; Tables Age/ChiSq Testp=(25 25 25 25);
Output Out=Stat Pchi;
By K;
*Proc Print;
Data Sim; Set Stat;
Reps = 10000;
n=100;
alpha = 0.05;
p11 = .1; p12 = .2; p13 = .3; p14 = .4;
If P_PChi < Alpha Then Chi+1;
PowerChi=Chi/K;
Proc GPlot; Plot PowerChi*K;
Data Final;Set Sim;
If K=Reps;
Proc Print;Var Alpha p11 p12 p13 p14 Reps n PowerChi;
Run;
```

Simulates the power for Case IV.

```
Data H4SimPow;
Title Simulated Power # 4;
n = 100;
p1 = .10; p2 = .4; p3 = .4; p4 = .10;
p10 = .13; p20 = .37; p30 = .37; p40 = .13;
e1 = n*p10;
e2 = n*p20;
e3 = n*p30;
e4 = n*p40;
Reps = 100000;
alpha = 0.05;
Do K =  1 to Reps;
Do Sample = 1 to n;
  x = RanTBL(0, p1, p2, p3, p4);
  If x=1 then x1+1;
else if x=2 then x2+1;
else if x=3 then x3+1;
else  x4+1;
End;
Output;
x1=0; x2=0; x3=0; x4=0;
End;
*Proc Print;
Data Stat1; Set H4SimPow;
*Proc Print;
Data Stat; Set Stat1;
ChiSq = ((x1-e1)**2)/e1 + ((x2-e2)**2)/e2 +((x3-e3)**2)/e3 + ((x4-
e4)**2)/e4;
pvalue = 1-ProbChi(ChiSq, 3);
If pvalue < alpha then Reject+1;
Data Final; Set Stat;
If _n_ = Reps;
Power = Reject/Reps;
Proc Print;Var n Reps p1 p2 p3 p4 p10 p20 p30 p40 Power;
Run;
```

Simulates the power for Case V.

```
Data H5SimP;
Reps = 5000;
n=200;
alpha = 0.05;
p11 = .30; p12 = .25; p13 = .25; p14 = .20;
p21 = .30; p22 = .25; p23 = .25; p24 = .20;
p31 = .30; p32 = .25; p33 = .25; p34 = .20;
Do K =  1 to Reps;
Do Sample = 1 to n;
  Age = RanTBL(0, p11, p12, p13, p14);Year=1;
Output;
  Age = RanTBL(0, p21, p22, p23, p24);Year=2;
Output;
  Age = RanTBL(0, p31, p32, p33, p34);Year=3;
Output;
End;
End;
*Proc Print;
Proc Freq NoPrint; Tables Year*Age/ChiSq;
Output Out=Stat Pchi LRChi;
By K;
*Proc Print;
Data Sim; Set Stat;
Alpha = 0.05; Reps=5000;
If P_PChi < Alpha Then Chi+1;
If P_LrChi < Alpha Then LR+1;
PowerChi=Chi/K;
PowerLR = LR/K;
Proc GPlot; Plot PowerChi*K;
Data Final;Set Sim;
If K=Reps;
Proc Print;Var Alpha Reps PowerChi PowerLR;
Run;
```

I, Mozhu Liu , hereby submit this thesis/report to Emporia State University as partial fulfillment of the requirements for an advanced degree. I agree that the Library of the University may make it available to use in accordance with its regulations governing materials of this type. I further agree that quoting, photocopying, digitizing or other reproduction of this document is allowed for private study, scholarship (including teaching) and research purposes of a nonprofit nature. No copying which involves potential financial gain will be allowed without written permission of the author.

Signature of Author

Date

Power and Sample Size for
some Chi-Square Goodness of Fit Tests

Title of Thesis

Signature of Graduate School Staff

Date Received