# A SOFTWARE PACKAGE FOR DETERMINATION OF THE SAMPLE SIZE

PATHOMPORN THITATAN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR

THE DEGREE OF MASTER OF SCIENCE (BIOSTATISTICS)

FACULTY OF GRADUATE STUDIES

MAHIDOL UNIVERSITY

2003

ISBN 974-04-3478-9

COPYRIGHT OF MAHIDOL UNIVERSITY

Copyright by Mahidol University

Thesis

entitled

**A SOFTWARE PACKAGE FOR DETERMINATION OF THE SAMPLE SIZE**

Pathomporn Thitatan
.....
Miss. Pathomporn Thitatan
Candidate

Wilai Kusolvisitkul
.....
Associate Prof. Wilai Kusolvisikul,
M.Sc. (Biostat)
Major-Advisor

Junya Pat
.....
Assistant Prof. Junya Pattaraarchachai,
Sc.D. (Biostat)
Co-Advisor

Pannawish Wong
.....
Mr. Pannawish Wongwiwattananon,
M.D.,Cert. in Rehabilitation Medicine
Co-Advisor

Rassm Hoon
.....
Associate Prof. Rassmidara Hoonsawat,
Ph.D.
Dean
Faculty of Graduate Studies

S. Singhaniyom
.....
Assistant Prof. Sumalee Singhaniyom,
M.Sc.(Biostat)
Acting Chair
Master of Science Programme
In Biostatistics
Faculty of Public Health

Copyright by Mahidol University

Thesis

entitled

**A SOFTWARE PACKAGE FOR DETERMINATION OF THE SAMPLE SIZE**

was submitted to the Faculty of Graduate Studies, Mahidol University

for the degree of Master of Science(Biostatistics)

on

May 21 , 2003

Pathomporn Thitatan

Miss. Pathomporn Thitatan
Candidate

Wilai Kusolvisitkul

Associate Prof. Wilai Kusolvisikul,
M.Sc. (Biostat)
Chairman

Junya P

Assistant Prof. Junya Pattaraarchachai,
Sc.D. (Biostat)
Member

Chulaluk Komoltri

Miss.Chulaluk Komoltri,
Dr.PH.(Biostatistics)
Member

Pannawish Wong.

Mr. Pannawish Wongwiwattananon,
M.D.,Cert. in Rehabilitation Medicine
Member

Rass Hoon

Associate Prof. Rassmidara Hoonsawat,
Ph.D.
Dean
Faculty of Graduate Studies

C. Chaikittiporn

Associate Prof. Chalermchai Chaikittiporn,
Dr.PH.(Epidemiology)
Dean
Faculty of Public Health
Mahidol University

Copyright by Mahidol University

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude and deep appreciation to Associate Prof. Wilai Kusolvisikul my principle advisor for her guidance, invaluable advisor, supervision and encouragement throughout. She was never lacking in kindness and support. I am equally grateful to Assistant Prof. Junya Pattaraarchachai my co-advisor for her constructive comments, supervision and encouragement. She was always nice and friendly support.

Special thanks to Dr. Chulaluk Komoltri for his constructive comments. I wish to thanks Doctor Pannawish Wongwiwattananon for encouragement. He was always nice and friendly support. I would like to thank Lecturer Supachai Sangrattanagul for their great assistance in computer applications. And also special unforgettable thanks to everybody in the Department of Biostatistics, Faculty of Public Health.

Finally, I am extreme gratefully to my family for their infinite love, kindness, understanding and cheering myself throughout my study.

Pathomporn Thitatan

Copyright by Mahidol University

A SOFTWARE PACKAGE FOR DETERMINATION OF THE SAMPLE SIZE.

PATHOMPORN THITATAN 4137776 PHBS/M

M.Sc. (BIOSTATISTICS)

THESIS ADVISORS: WILAI KUSOLVISIKUL, M.Sc., JANYA PATTARAARCHACHAI, Sc.D., PANNAWISH WONGWIWATTANANON, M.D.

ABSTRACT

In this study, the program developer aimed to improve the sample size determination using the Microsoft Visual Basic Professional Edition 6.0 to write the program and calculate the sample size. In addition, Microsoft Excel was also used. This sample size determination program is composed of four parts: 1.Sample size for sample survey, 2. Sample size for observational study, 3. Sample size for logistic regression and 4. Sample size for correlation. Concerning the sample size for a sample survey, it is composed of an estimation of the parameters and hypothesis testing, in which the estimation of the parameters have three designs: Simple Random Sampling, Stratified Sampling and Cluster Sampling. Each design uses allowable absolute error, and relative error. Sample size for observational study is divided into two categories: Case-Control Study and Cohort Study. Sample size for logistic regression is composed of Simple Logistic Regression and Multiple Logistic Regression. Sample size for correlation is divided into a Test of a Single Correlation and Test of Two Independent Correlations.

This program was evaluated by 28 evaluators, using different sample size backgrounds, with 71.4 % female and 28.6 % male. There were 42.9 % of the evaluators who used to work on other sample size programs, while other 57.1 % have never used any program at all. The evaluators gave a very good score of evaluation regarding data input with 57.1% positive in the area of speed with a mean of 3.57; and 46.4% in the area of convenience with a mean of 3.36; and last of all, 42.9% in the area of coverage with a mean of 3.36 sample size solution coverage. In the area of features and format beauty, it needed improvement (39.3%, mean 2.71). The advantage of the program is that it is easy and convenient to use and appropriate for sample size coverage. It has many type I error and type II error values; moreover, it has a “Help” menu assisting the user in case they need guidance. This program can give many sample size results without going back to the square one. However, this program is unable to find the sample size value of Survival and Clinical trials, this being the only disadvantage of this program.

KEY WORDS : PACKAGE /SAMPLE SIZE/SURVEY/OBSERVATIONAL/ LOGISTIC REGRESSION/CORRELATION

93 P. ISBN 974-04-3478-9

Copyright by Mahidol University

โปรแกรมสำเร็จรูปสำหรับการคำนวณหาขนาดตัวอย่าง (A SOFTWARE PACKAGE FOR DETERMINATION OF THE SAMPLE SIZE)

ปฐมพร ฐิตะฐาน 4137776 PHBS/M

วท.ม. (ชีวสถิติ)

คณะกรรมการควบคุมวิทยานิพนธ์ : วิไล กุศลวิศิษฏ์กุล, M.Sc., จรรยา ภัทรอาชาชัย, Sc.D., ปัณณวิชญ์ วงศ์วิวัฒนานนท์, M.D.

## บทคัดย่อ

จากการพัฒนาโปรแกรมการคำนวณหาขนาดตัวอย่างครั้งนี้ได้ใช้โปรแกรม Microsoft Visual Basic Professional Edition 6.0 มาเขียนและคำนวณ แต่ในการคำนวณบางเรื่องอาจใช้โปรแกรม Microsoft Excel มาคำนวณร่วมด้วย ในโปรแกรมการคำนวณหาขนาดตัวอย่างนี้ได้ทำการคำนวณหาขนาดตัวอย่าง 4 เรื่อง ดังนี้ 1.การหาขนาดตัวอย่างในการสำรวจ 2. การหาขนาดตัวอย่างในการสังเกต 3. การหาขนาดตัวอย่างสำหรับlogistic regression 4. การหาขนาดตัวอย่างใน correlation ในเรื่องของการหาขนาดตัวอย่างในการสำรวจ แบ่งได้ 2 เรื่อง คือ การประมาณค่าพารามิเตอร์และการทดสอบสมมุติฐาน และในการประมาณค่าพารามิเตอร์ ประกอบด้วยแบบแผนการสุ่มตัวอย่าง 3 แบบ คือ การสุ่มตัวอย่างแบบง่าย,การสุ่มตัวอย่างแบบชั้นภูมิ,การสุ่มตัวอย่างแบบกลุ่มและในแต่ละแบบแผนการสุ่มตัวอย่างยังใช้การกำหนดความคลาดเคลื่อนสัมบูรณ์ และกำหนดความคลาดเคลื่อนสัมพัทธ์ มาใช้ในการคำนวณ การหาขนาดตัวอย่างในการสังเกตยังสามารถแบ่งเป็น 2 วิธีคือ Case-Control study และ Cohort study การหาขนาดตัวอย่างสำหรับ logistic regression แบ่งเป็น 2 เรื่อง คือ Simple logistic regression and Multiple logistic regression การหาขนาดตัวอย่างใน correlation แบ่งเป็น Test of a single correlation and Test of two independent correlations

จากการนำโปรแกรมการคำนวณหาขนาดตัวอย่างไปให้ผู้ประเมินโปรแกรมได้ใช้ทั้งหมด 28 คนโดยทั้ง 28 คนมีความรู้เกี่ยวกับการหาขนาดตัวอย่างทั้งหมด โดยแบ่งเป็นเพศหญิง 71.4% เพศชาย 28.6% เคยใช้โปรแกรมการคำนวณหาขนาดตัวอย่างมาก่อน 42.9%และไม่เคยใช้ 57.1% จากการแสดงความคิดเห็นต่างๆของผู้ประเมินความรวดเร็วในการป้อนข้อมูลและความสะดวกในการป้อนข้อมูลและความครอบคลุมในเนื้อหาดีที่สุด 57.1%,46.4%,42.9%ตามลำดับ ค่าเฉลี่ย 3.57,3.36,3.36 ตามลำดับ และความสวยงามยังต้องปรับปรุง 39.3% ค่าเฉลี่ย 2.71 และจุดเด่นของโปรแกรมคือ ใช้ง่าย,สะดวก,ครอบคลุมเนื้อหาการคำนวณหาขนาดตัวอย่าง มีค่า type I error ($\alpha$), type II error ($\beta$) หลายค่า มีเมนู Help แสดงสูตรการคำนวณอีกทั้งแสดงผลลัพธ์หลายค่าโดยไม่ต้องทำการคำนวณใหม่ ข้อเสีย ยังไม่สามารถคำนวณหาขนาดตัวอย่างในเรื่อง Survival และ Clinical trial ได้

93 หน้า. ISBN 974-04-3478-9

Copyright by Mahidol University

# CONTENTS


| | | Page |
|---|---|---|
| **ACKNOWLEDGEMENT** | | **iii** |
| **ABSTRACT** | | **iv** |
| **LIST OF TABLES** | | **viii** |
| **LIST OF FIGURES** | | **ix** |
| **CHAPTER** | | |
| **I** | **INTRODUCTION** | |
| | Significance of the problem | **1** |
| | Objective | **2** |
| | Scope of the study | **2** |
| | Benefits anticipated from this program | **2** |
| | Scopes in Writing This Program | **3** |
| **II** | **LITERATURE REVIEW** | |
| | Sample size calculation | **4** |
| | Package for determination of the sample size | **34** |
| **III** | **METHODOLOGY** | |
| | Research Instruments | **38** |
| | Step for writing A Software package for determination of the sample size | **40** |
| | The sample size determination solution comparing with other programs | **44** |
| **IV** | **RESULTS** | |
| | The program character. | **46** |
| | Parts of A Software package for determination of the sample size | **46** |
| | Steps inputting data and Error data show in data input | **51** |
| | The program evaluation by the users. | **51** |


Copyright by Mahidol University

# CONTENTS (Continues)


| | | Page |
|---|---|---|
| **V** | **DISCUSSION** | |
| | The developer gave discussion toward the A-SIZE Program | **56** |
| | Comparing A-SIZE program to other Sample size programs which is appropriately able to work well with sample size calculation | **56** |
| **VI** | **CONCLUSION AND RECOMMENDATION** | |
| | Conclusion | **59** |
| | Recommendation | **60** |
| **REFERENCES** | | **61** |
| **APPENDIX A** | | |
| | A Software package for determination of the sample size hand book | **63** |
| **APPENDIX B** | | |
| | Evaluation Form | **90** |
| **BIOGRAPHY** | | **93** |


Copyright by Mahidol University

# LIST OF TABLES


| Tables | | Page |
|---|---|---|
| 1 | Formula total estimator within “d” percentage points | 16 |
| 2 | Formula total estimator within “r” percentage points in stratified sampling | 17 |
| 3 | Formula mean estimator within “d” percentage points in stratified sampling | 18 |
| 4 | Formula mean estimator within “r” percentage points in stratified sampling | 19 |
| 5 | Formula proportion estimator within “d” percentage points in stratified sampling | 20 |
| 6 | Formula proportion estimator within “r” percentage points in stratified sampling | 21 |
| 7 | Unit normal deviates $Z_{\alpha}$ and $Z_{\beta}$ for Selected Values of $\alpha$ and $\beta$. | 26 |
| 8 | Comparison of Seven Software for Sample Size Calculation | 45 |
| 9 | Number and Percentage of Evaluators’ General Data | 52 |
| 10 | Number and Percentage Concerning Sample Size by Evaluators Experiences | 52 |
| 11 | Number and Percentage of Software Using by Evaluators’ Opinion | 53 |


Copyright by Mahidol University

# LIST OF FIGURES


| Figures | | Page |
|---|---|---|
| 1 | Research flowchart | 41 |
| 2 | Designing a software package flowchart | 43 |
| 3 | Show menu bar in estimation | 47 |
| 4 | Show menu bar in hypothesis testing | 48 |
| 5 | Show menu bar in sample size for observational study | 49 |
| 6 | Show menu bar in sample size for logistic regression | 49 |
| 7 | Show menu bar in sample size for correlation | 50 |


Copyright by Mahidol University

# CHAPTER I
# INTRODUCTION

## Significance of the Problem

It is an important question a statistician frequently asks in planning a survey how to determine the appropriate sample size. In general research, it is unable to use the number of the entire population to determine statistical research, but to use sampling method instead. The sample size can refer to the population. Good sample represents and covers the entire population without prejudice. Having proper size, good enough in gaining confidence statistically, its data must be referable to its own population correctly. Therefore, it is not easy to determine the sample size, on the other hand, if the sample size is too small, it will be unable to be referred to the entire population. In vise versa, if the sample size is too large, it will lead to the waste of time and cost spending. Moreover, as the sample size increased, the statistical error can be commonly created, too. Thus, what need to done is to determine good sample size. There are various programs used in determining sample size on nowadays, but none of these programs will cover the entire required patterns. Generally, these are good in specific uses only, for instance, some may emphasize in Epidemiology; some are in Clinical trial; some are invented to use in few patterns; while some don't have complete parameter making the users have got no choice and finally, confused. Some users who did not have enough background in statistics might choose to use the program inappropriately or even calculate manually which simply leads to complexity, wasting of time and even come up with wrong conclusion.

As it is mentioned above, the researcher can simply see the significant of developing a software package for determination of the sample size to cover as many methods as it can be to eliminate the mistake of calculation, time spending, and cost. Furthermore, the user may comfortably use the program under the instruction guided within the program itself in case the user is not statistician or having expertise in

Copyright by Mahidol University

computer. Since on nowadays computer is more widely used on daily basis, this program will help the user cut down cost and time as well.

**Objective**

To develop " A Software package for determination of the sample size " as a tool to be used in assisting researchers to calculate for the results of sample size.

**Scopes of the Study**

1. Sample size for sample survey
   - 1.1 Estimating the population
     - 1.1.1 Simple random sampling
     - 1.1.2 Systematic sampling
     - 1.1.3 Stratified sampling
     - 1.1.4 Cluster sampling
   - 1.2 Hypothesis testing
     - 1.2.1 Mean
     - 1.2.2 Proportion
2. Sample size for observational study
   - 2.1 Case-Control study
   - 2.2 Cohort study
3. Sample size for logistic regression
   - 3.1 Simple logistic regression
   - 3.2 Multiple logistic regression
4. Sample size for correlation
   - 4.1 Test of a single correlation
   - 4.2 Test of two independent correlations

**Benifits Anticipated from this Program**

1. A software program for convenient-ship to compute sample size, cutting down time consuming.

Copyright by Mahidol University

2. Students will save money and, they do not need to purchase the other expensive programs.
3. To develop ready to use program to calculate sample size.
4. To modify computer technology in statistical use.

**Scopes in Writing This Program.**

1. Software package is to be used on microcomputer.
2. Minimum requirement of PC is not lesser than Pentium and runs on Microsoft Windows's 95 or higher.
3. Microsoft Visual Basic Professional Edition 6.0 makes the software development for determination of the sample size calculation.

Copyright by Mahidol University

# CHAPTER II
# LITERATURE REVIEW

This chapter is to mention

1. Sample size for calculation
   1.1 Defining term
   1.2 Sample size for sample survey
   1.3 Sample size for observational study
   1.4 Sample size for logistic regression
   1.5 Sample size for correlation
2. Package for determination of the sample size.

## 1. Sample size for calculation

### 1.1 Defining Term

**The Population**

The population (or universe or target population) is the entire set of individuals to which findings of the survey are to be extrapolated.(1)

**Elementary Units**

Elementary units or elements of the population is the individual members of the population whose characteristics are to be measured

**Sample**

A part of a population or subset of population. The primary objective of a sample survey are to take a sample from the population and to estimate population parameters from that sample.

**Population Parameters**

Population Parameter is the values of the distribution of a specified variable or characteristic in a population, and for given population a parameter is constant.

Copyright by Mahidol University

These are four values of population parameters, that is, population total, population mean, population proportion, population ratio.

**Population Total**

The population total of a characteristic X is generally denoted by T(x) and is the sum of the values of the characteristic over all elements in the population. The population total is given by (1)

$$T(x) = \sum_{i=1}^{N} x_i$$

**Population Mean**

The population mean with respect to a characteristic X is given by

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

**Population Proportion**

When the characteristic being measured represents the presence or absence of some dichotomous attribute. It is often desired to estimate the proportion of elementary units in the population having the attribute. The P denoted the population proportion of elements having the attribute and is given by

$$P = \frac{\sum_{i=1}^{N} x_i}{N}, x = 0,1$$

**Population Ratio**

The population ratio of a characteristic Y are X is generally denoted by R. The population ratio is given by

$$R = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i} \quad \text{or} \quad \frac{\mu_y}{\mu_x}$$

**Estimator**

A sample statistic that is used to determine a probable value of a population parameter as one might use a known value of a sample to estimate the value of the population. Good estimators should be unbiased, consistent and efficient.(2)

Copyright by Mahidol University

**Sample Total**

An estimate of the population total can be obtained from the sample total as given by

$$\hat{T}(x) = \frac{N}{n}\sum_{i=1}^{n} x_i$$

**Sample Mean**

An estimate of the population mean can be obtained from the sample mean as given by

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

**Sample Proportion**

An estimate of the population proportion can be obtained from the sample proportion as given by

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}, x_i = 0,1$$

**Sample Ratio**

An estimate of the population ratio can be obtained from the sample ratio as given by

$$\hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\hat{T}(y)}{\hat{T}(x)} = \frac{\overline{y}}{\overline{x}}$$

**Probability Sampling**

A probability samples has the characteristic that every element in the population has a known. Because every element has a known chance of being selected, estimate of population parameters are linear function of the observations (population mean, totals, proportions, ratio) can be constructed from the sample data. Also, the standard errors of these estimates can be estimated under the condition that the second-order inclusion probabilities are known. This gives the users of the survey estimates insight into how much value can be placed on the estimates.

Copyright by Mahidol University

**Coefficient of Variation**

Coefficient of variation or CV denotes the relative variation of the distribution. The $CV(\hat{\theta})$ of an estimate $(\hat{\theta})$ of a population parameter, $\theta$, and its standard error $SE(\hat{\theta})$ divided by the true value $\theta$ of the parameter being estimated, the formula is

$$CV(\hat{\theta}) = \frac{SE(\hat{\theta})}{\theta}$$

**1.2 Sample Size for Sample Survey**

Determining sample size for sample survey its have sampling plan such as Simple random sampling, Systematic sampling, Stratified sampling and Cluster sampling.

**1.2.1 Simple Random Sampling**

Simple random sampling, or random sampling without replacement, is the sampling design in which n distinct units are selected from the N units in the population in such a way that every possible combination of n units is equally likely to be the sample selected. The sample may be obtained through n selections in which at each step every unit of the population not already selected has equal chance of selection. Equivalently, one may make a sequence of independent selections from the whole population, each unit having equal probability of selections at each step, discarding repeat selections and continuing until n distinct unit are obtained.

With simple random sampling, the probability that the i th unit of the population is included in the sample is $p = n/N$, so that the inclusion probability is the same for each unit. Designs other than simple random sampling may give each unit equal probability of being included in the sample, but only with simple random sampling does each possible sample of n units have the same probability. (3)

**1.2.1.1 Sample Size for Estimating Parameter**

To determine sample size for estimating parameter can be done by putting allowable absolute error of estimation, or putting relative error of estimation

Copyright by Mahidol University

which are giving the same result or answer but different aspect. The symbol “d” tells allowable absolute error while “r” tells how much percent that “d” is away from the real value.There are four values of the estimating parameter such as total, mean, proportion and ratio (4),(5).

**a. Sample size for estimating a population total**

**To determine allowable absolute error**

Find the allowable absolute error or margin error by putting estimating parameter as the target, then, setting the acceptable maximum error which is a given number having the symbol of “d”. (4)

Estimating Total (T) to within “d” absolute percentage points

| **One sample** | **Two sample $T_1 - T_2$** |
|---|---|
| $n = \frac{n_0}{1+\frac{n_0}{N}}$, $n_0 = \frac{Z_{\frac{\alpha}{2}}^2 N^2 s^2}{d^2}$ | $n = \frac{n_0}{1+\frac{n_0}{N}}$, $n_0 = \frac{Z_{\frac{\alpha}{2}}^2 (N_1^2 s_1^2 + N_2^2 s_2^2)}{d^2}$ |

The sample variance is defined as

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

| | | |
|---|---|---|
| N | = | Population size |
| n | = | Sample size for finite population |
| s | = | Sample standard deviation |
| $\alpha$ | = | Type I error |
| d | = | Allowable absolute error |
| $n_0$ | = | Sample size for infinite population |

Copyright by Mahidol University

**Finding the relative error.**

This is similar to finding the value of "d" but given in term of how many times fold of error or in term of percentage of parameter.(4)

Estimating 'T to within "r" of T

$$n = \frac{n_0}{1 + \frac{n_0}{N}}, \qquad n_0 = \frac{Z^2_{\frac{\alpha}{2}} CV^2}{r^2}$$

$$CV = \frac{s}{\bar{x}}$$

| | | |
|---|---|---|
| N | = | Population size |
| n | = | Sample size for finite population |
| s | = | Sample standard deviation |
| CV | = | Coefficient of variation |
| r | = | Relative error |
| $\alpha$ | = | Type I error |

**b. Sample size for estimating a population mean**

**To determine allowable absolute error**

Estimating $\mu$ to within "d" absolute percentage points

| **One sample** | **Two sample $\mu_1 - \mu_2$** |
|---|---|
| $n = \frac{n_0}{1 + \frac{n_0}{N}}$, $n_0 = \frac{Z^2_{\frac{\alpha}{2}} s^2}{d^2}$, $d = \lvert\hat{\mu} - \mu\rvert$ | $n = \frac{n_0}{1 + \frac{n_0}{N}}$, $n_0 = \frac{Z^2_{\frac{\alpha}{2}} s_p^2}{d^2}$ |

Copyright by Mahidol University

$$n_1 = n_2 = n \text{ and } s_1^2 = s_2^2 = s^2$$

$$s_p^2 = \frac{s_1^2 + s_2^2}{2}, n_1 = n_2$$

**Finding the relative error**

Estimating μ to within "r" of μ

$$n = \frac{n_0}{1 + \frac{n_0}{N}}, \quad n_0 = \frac{Z_{\frac{\alpha}{2}}^2 CV^2}{r^2}$$

$$CV = \frac{s}{\bar{x}} \qquad r = \frac{|\hat{\mu} - \mu|}{\mu}$$

**c. Sample size for estimating a population proportion**

**To determine allowable absolute error**

Estimating P to within "d" absolute percentage points

| One sample | Two sample $P_1 - P_2$ |
|---|---|
| $n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$, $n_0 = \frac{Z_{\frac{\alpha}{2}}^2 pq}{d^2}$, $d = \lvert \hat{p} - p \rvert$ | $n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$, $n_1 = n_2 = n$, $n_0 = \frac{Z_{\frac{\alpha}{2}}^2 (p_1q_1 + p_2q_2)}{d^2}$, $n_2 = kn_1$, $n_0 = \frac{Z_{\frac{\alpha}{2}}^2 (kp_1q_1 + p_2q_2)}{kd^2}$ |

N = Population size

n = Sample size for finite population

Copyright by Mahidol University

$p$ = Sample proportion

$p_1$ = Sample proportion group 1

$p_2$ = Sample proportion group 2

$q_1$ = $1- p_1$

$k$ = Constant

When it is desired that $n_2$ be proportionate to $n_1$ ($n_2 = kn_1$)

$\alpha$ = Type I error

$d$ = Allowable absolute error

**Finding the relative error.**

Estimating P to within "r" of P

$$n = \frac{n_0}{1+\frac{n_0 - 1}{N}} \qquad n_0 = \frac{Z^2_{\frac{\alpha}{2}} CV^2}{r^2}$$

$$r = \frac{|\hat{p} - p|}{p} \qquad CV = \sqrt{\frac{q}{p}}$$

**d. Sample size for estimating a population ratio**

**To determine allowable absolute error**

Estimating R to within "d" absolute percentage points

| **One sample** | **Two sample $R_1 - R_2$** |
|---|---|
| $n = \frac{n_0}{1+\frac{n_0}{N}}$, <br> $n_0 = \frac{Z^2_{\frac{\alpha}{2}} s_r^2}{\bar{x}^2 d^2}$ | $n = \frac{n_0}{1+\frac{n_0}{N}}$, <br> $n_0 = \frac{Z^2_{\frac{\alpha}{2}} (\bar{X}_2^2 s_1^2 + \bar{X}_1^2 s_2^2)}{d^2 (\bar{X}_1^2 \bar{X}_2^2)}$ |

$$s_r^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{R}x_i)^2$$

Copyright by Mahidol University

**Notation**

N = Population size
n = Sample size for finite population
d = Allowable absolute error
r = Relative error
$s^2$ = Sample variance
CV = Coefficient of variation
$\hat{p}$ = Proportion estimation
$\hat{q}$ = $1-\hat{p}$
$\bar{x}$ = Mean

**1.2.1.2 Sample Size for Hypothesis Testing**

As mentioned earlier, it is just an estimation of sample size determination. Practically, a researcher can calculate sample size on Hypothesis testing.

**a. Hypothesis testing for mean**

**Hypothesis testing for a single population mean**

| One – sided test | Two – sided test |
|---|---|
| $H_o:\mu=\mu_o; H_a:\mu>\mu_o, H_a:\mu<\mu_o$ | $H_o:\mu=\mu_o; H_a:\mu\neq\mu_o$ |
| $n=\frac{s^2(Z_\alpha+Z_\beta)^2}{(\mu_a-\mu_0)^2}$ | $n=\frac{s^2(Z_{\frac{\alpha}{2}}+Z_\beta)^2}{(\mu_a-\mu_0)^2}$ |

n = Sample size
$\mu_0$ = Population mean in null hypothesis
$\mu_a$ = Population mean in alternative hypothesis
$s^2$ = Sample variance

Copyright by Mahidol University

**Hypothesis testing for two population mean**

| Independent sample | | Related sample or Matched design |
|---|---|---|
| **One–sided test** | **Two–sided test** | |
| $H_0 : \mu_1 = \mu_2; H_a : \mu_1 > \mu_2, H_a : \mu_1 < \mu_2$ | $H_0 : \mu_1 = \mu_2; H_a : \mu_1 \neq \mu_2$ | $H_0 : \mu_d = 0$ |
| $n = \frac{2s^2(Z_\alpha + Z_\beta)^2}{(\mu_1 - \mu_2)^2}$ | $n = \frac{2s^2(Z_{\frac{\alpha}{2}} + Z_\beta)^2}{(\mu_1 - \mu_2)^2}$ | $n = \frac{s_d^2(Z_{\frac{\alpha}{2}} + Z_\beta)^2}{\mu_d^2}$ <br> $\mu_d = \mu_2 - \mu_1$ |

**b. Hypothesis testing for proportion**

**Hypothesis testing for a single population proportion**

| One – sided test | Two – sided test |
|---|---|
| $H_o : P = P_o; H_a : P > P_o; H_a : P < P_o$ | $H_o : P = P_o; H_a : P \neq P_o$ |
| $n = \frac{\left(Z_\alpha \sqrt{P_o Q_o} + Z_\beta \sqrt{P_a Q_a}\right)^2}{(P_a - P_o)^2}$ | $n = \frac{\left(Z_{\frac{\alpha}{2}} \sqrt{P_o Q_o} + Z_\beta \sqrt{P_a Q_a}\right)^2}{(P_a - P_o)^2}$ |

n = Sample size

$p_o$ = Population proportion in null hypothesis

$p_a$ = Population proportion in alternative hypothesis

$s^2$ = Sample variance

Copyright by Mahidol University

**Hypothesis testing for two population proportion**

| One – sided test | Two – sided test |
|---|---|
| $H_0 : P_1 = P_2, H_a : P_1 > P_2; H_a : P_1 < P_2$ | $H_0 : P_1 = P_2; H_a : P_1 \neq P_2$ |
| $n = \frac{(Z_\alpha \sqrt{2\bar{P}\bar{Q}} + Z_\beta \sqrt{P_1Q_1 + P_2Q_2})^2}{(P_1 - P_2)^2}$ <br> $\bar{P} = \frac{P_1 + P_2}{2}, n_1 = n_2 = n$ | $n = \frac{(Z_{\frac{\alpha}{2}} \sqrt{2\bar{P}\bar{Q}} + Z_\beta \sqrt{P_1Q_1 + P_2Q_2})^2}{(P_1 - P_2)^2}$ |

### 1.2.2 Systematic Sampling

This method of sampling is at first sight quite different from simple random sampling. Suppose that the N units in the population are numbered 1 to N in some order. To select a sample of n units, we take a unit at random from the first k units and every th unit thereafter. For instance, if k is 15 and if the first unit drawn is number 13,the subsequent units are number 28, 43, 58, and so on. The selection of the first unit determines the whole sample. This type is called an every k th systematic sample.

The apparent advantages of this method over simple random sampling are as follows:

1. It is easier to draw a sample and often easier to execute without mistakes. This is particular advantage when the drawing is done in the field. Even when drawing is done in an office there may be a substantial saving in time. For instance, if the units are described on cards that are all of the same size and lie in a file drawer, a card can be drawn out every inch along the file as measured by a ruler. This operation is speedy, whereas simple random sampling would be slow. Of course, this method departs slightly from the strict “every k th” rule.

2. Intuitively, systematic sampling seems likely to be more precise than simple random sampling. (6 )

Copyright by Mahidol University

### 1.2.3 Stratified Sampling

Stratified sampling is the sampling frame can be partitioned into groups or strata, and the sampling can be performed separately within each stratum. Stratified random sample is a sampling plan in which a population is divided in L mutually exclusive and exhaustive strata, and a simple random sample of $n_h$ elements is taken within each stratum h. The sampling is performed independently within each stratum. In essence, we can think of a stratified random sampling scheme as being L separate simple random samples.(1)

Stratification is a common technique. There are many reasons for this; the principal ones are the following:

1. If data of know precision are wanted for certain subdivisions of the population, it is advisable to treat each subdivision as a "population" in its own right

2. Administrative convenience may dictate the use of stratification; for example, the agency conducting the survey may have field offices, each of which can supervise the survey for a part of the population.

3. Sampling problems may differ markedly in different parts of the population. With human populations, people living in institutions (e.g., hotels, hospitals, prisons) are often placed in a different stratum from people living in ordinary homes because a different approach to the sampling is appropriate for the two situations. In sampling businesses we may possess a list of the large firms, which are placed in a separate stratum. Some type of area sampling may have to be used for the smaller firms.

4. Stratification may produce a gain in precision in the estimates of characteristics of the whole population. It may be possible to divide a heterogeneous population into subpopulation, each of which is internally homogeneous. This is suggested by the name strata, with its implication of a division into layers. If each stratum is homogeneous, in that the measurements vary little from one unit to another, a precise estimate of any stratum mean can be obtained from a small sample in that stratum. These estimates can then be combined into a precise estimate for the whole population. The theory of stratified sampling deals with the properties of the estimates from a stratified sample and with the best choice of the sample sizes $n_h$ to obtain maximum precision. In this development it is taken for granted that the strata have

Copyright by Mahidol University

already been constructed. The problems of how to construct strata and of how many strata there should be are postponed to a later stage. (6)

The Formula in stratified sampling was shown in the following table.

**Table 1.** Formula for total estimator within "d" percentage points

| **Allocation** | **n** | **$n_h$** |
|---|---|---|
| Equal sample | $n = \frac{Z^2_{\frac{\alpha}{2}} L \sum_{h=1}^{L} N_h s_h^2}{d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{n}{L}$ |
| Proportional allocation | $n = \frac{Z^2_{\frac{\alpha}{2}} N \sum_{h=1}^{L} N_h s_h^2}{d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{N_h}{\sum_{h=1}^{L} N_h} n$ |
| Neyman's allocation | $n = \frac{Z^2_{\frac{\alpha}{2}} \left( \sum_{h=1}^{L} N_h s_h \right)^2}{d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{N_h s_h}{\sum_{h=1}^{L} N_h s_h} n$ |
| Optimum allocation for fixed cost | $n = \frac{Z^2_{\frac{\alpha}{2}} \left[ \sum_{h=1}^{L} N_h s_h \sqrt{c_h} \right] \frac{\left[ \sum_{h=1}^{L} N_h s_h \right]}{\sqrt{c_h}}}{d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{\frac{N_h s_h}{\sqrt{c_h}}}{\frac{\sum_{h=1}^{L} N_h s_h}{\sqrt{c_h}}} n$ |

$N = \sum_{n=1}^{L} N_h$ = The total number of units in the population

$n = \sum_{n=1}^{L} n_h$ = Total sample size

$N_h$ = Total number of units in stratum h

$n_h$ = The number of units in the sample from stratum

$s_h$ = Sample standard deviation in stratum h

$c_h$ = The cost per unit observed in stratum h

L = Number of stratum

Copyright by Mahidol University

**Table 2.** Formula for total estimator within "r" percentage points

| **Allocation** | **n** | **$n_h$** |
|---|---|---|
| Equal sample | $n = \dfrac{Z^2_{\frac{\alpha}{2}} L \sum_{h=1}^{L} N_h^2 s_h^2}{r^2 \sum_{h=1}^{L} N_h \bar{x}_h^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \dfrac{n}{L}$ |
| Proportional allocation | $n = \dfrac{Z^2_{\frac{\alpha}{2}} N \sum_{h=1}^{L} N_h s_h^2}{r^2 \sum_{h=1}^{L} N_h \bar{x}_h^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \dfrac{N_h}{\sum_{h=1}^{L} N_h} n$ |
| Neyman's allocation | $n = \dfrac{Z^2_{\frac{\alpha}{2}} \left( \sum_{h=1}^{L} N_h s_h \right)^2}{r^2 \sum_{h=1}^{L} N_h \bar{x}_h^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h}$ | $n_h = \dfrac{N_h s_h}{\sum_{h=1}^{L} N_h s_h} n$ |
| Optimum allocation for fixed cost | $n = \dfrac{Z^2_{\frac{\alpha}{2}} \left[ \sum_{h=1}^{L} N_h s_h \sqrt{c_h} \right] \dfrac{\left[ \sum_{h=1}^{L} N_h s_h \right]}{\sqrt{c_h}}}{r^2 \sum_{h=1}^{L} N_h \bar{x}_h^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \dfrac{\dfrac{N_h s_h}{\sqrt{c_h}}}{\dfrac{\sum_{h=1}^{L} N_h s_h}{\sqrt{c_h}}} n$ |

$N_h$ = Total number of units in stratum h

$n_h$ = The number of units in the sample from stratum

$s_h$ = Sample standard deviation in stratum h

$c_h$ = The cost per unit observed in stratum h

L = Number of stratum

r = Relative error

$\bar{x}_h$ = Mean of stratum h

Copyright by Mahidol University

**Table 3.** Formula for mean estimator within "d" percentage points (5)

| Allocation | n | $n_h$ |
|---|---|---|
| Equal sample | $n = \frac{Z^2_{\frac{\alpha}{2}} L \sum_{h=1}^{L} N_h^2 s_h^2}{N^2 d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{n}{L}$ |
| Proportional allocation | $n = \frac{Z^2_{\frac{\alpha}{2}} N \sum_{h=1}^{L} N_h s_h^2}{N^2 d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{N_h}{\sum_{h=1}^{L} N_h} n$ |
| Neyman's allocation | $n = \frac{Z^2_{\alpha/2} \left( \sum_{h=1}^{L} N_h s_h \right)^2}{N^2 d^2 + Z^2_{\alpha/2} \sum_{h=1}^{L} N_h s_h}$ | $n_h = \frac{N_h s_h}{\sum_{h=1}^{L} N_h s_h} n$ |
| Optimum allocation for fixed cost | $n = \frac{Z^2_{\alpha/2} \left[ \sum_{h=1}^{L} N_h s_h \sqrt{c_h} \right] \frac{\left[ \sum_{h=1}^{L} N_h s_h \right]}{\sqrt{c_h}}}{N^2 d^2 + Z^2_{\alpha/2} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{\frac{N_h s_h}{\sqrt{c_h}}}{\frac{\sum_{h=1}^{L} N_h s_h}{\sqrt{c_h}}} n$ |

$N_h$ = Total number of units in stratum h

$n_h$ = The number of units in the sample from stratum

$s_h$ = Sample standard deviation in stratum h

$c_h$ = The cost per unit observed in stratum h

L = Number of stratum

d = Allowable absolute error

Copyright by Mahidol University

**Table 4.** Formula for mean estimator within "r" percentage points

| **Allocation** | **n** | **$n_h$** |
|---|---|---|
| Equal sample | $n = \frac{Z^2_{\frac{\alpha}{2}} L \sum_{h=1}^{L} N_h^2 s_h^2}{r^2\left(\sum_{h=1}^{L} N_h \mu_h\right)^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{n}{L}$ |
| Proportional allocation | $n = \frac{Z^2_{\frac{\alpha}{2}} N \sum_{h=1}^{L} N_h s_h^2}{r^2\left(\sum_{h=1}^{L} N_h \mu_h\right)^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^{\,2}}$ | $n_h = \frac{N_h}{\sum_{h=1}^{L} N_h} n$ |
| Neyman's allocation | $n = \frac{Z^2_{\frac{\alpha}{2}} \left(\sum_{h=1}^{L} N_h s_h\right)^2}{r^2\left(\sum_{h=1}^{L} N_h \mu_h\right)^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^{\,2}}$ | $n_h = \frac{N_h s_h}{\sum_{h=1}^{L} N_h s_h} n$ |
| Optimum allocation for fixed cost | $n = \frac{Z^2_{\frac{\alpha}{2}} \left[\sum_{h=1}^{L} N_h s_h \sqrt{c_h}\right] \frac{\left[\sum_{h=1}^{L} N_h s_h\right]}{\sqrt{c_h}}}{r^2\left(\sum_{h=1}^{L} N_h \mu_h\right)^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h s_h^2}$ | $n_h = \frac{\frac{N_h s_h}{\sqrt{c_h}}}{\frac{\sum_{h=1}^{L} N_h s_h}{\sqrt{c_h}}} n$ |

$N_h$ = Total number of units in stratum h

$n_h$ = The number of units in the sample from stratum

$s_h$ = Sample standard deviation in stratum h

$c_h$ = The cost per unit observed in stratum h

L = Number of stratum

r = Relative error

$\mu_h$ = Mean of stratum h

Copyright by Mahidol University

**Table 5.** Formula for proportion estimator within "d" percentage points

| Allocation | n | $n_h$ |
|---|---|---|
| Equal sample | $n = \dfrac{Z^2_{\frac{\alpha}{2}} L \sum_{h=1}^{L} N_h^2 \frac{N_h}{N_{h-1}} \hat{p}_h(1-\hat{p}_h)}{N^2 d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h \frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}$ | $n_h = \dfrac{n}{L}$ |
| Proportional allocation | $n = \dfrac{Z^2_{\frac{\alpha}{2}} N \sum_{h=1}^{L} N_h \frac{N_h}{N_{h-1}} \hat{p}_h(1-\hat{p}_h)}{N^2 d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h \frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}$ | $n_h = \dfrac{N_h}{\sum_{h=1}^{L} N_h} n$ |
| Neyman's allocation | $n = \dfrac{Z^2_{\frac{\alpha}{2}} \left[ \sum_{h=1}^{L} N_h \sqrt{\frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)} \right]^2}{N^2 d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h \frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}$ | $n_h = \dfrac{N_h \sqrt{p_h q_h}}{\sum_{h=1}^{L} N_h p_h q_h} n$ |
| Optimum allocation for fixed cost | $n = \dfrac{Z^2_{\frac{\alpha}{2}} \left( \sum_{h=1}^{L} N_h \sqrt{\frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h) c_h} \right) \left( \frac{\sum_{h=1}^{L} N_h \sqrt{\frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}}{\sqrt{c_h}} \right)}{N^2 d^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h \frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}$ | $n_h = \dfrac{\frac{N_h \sqrt{p_h q_h}}{\sqrt{c_h}}}{\sum_{h=1}^{L} \frac{N_h p_h q_h}{\sqrt{c_h}}} n$ |

$N_h$ = Total number of units in stratum h

$n_h$ = The number of units in the sample from stratum

$c_h$ = The cost per unit observed in stratum h

L = Number of stratum

$\hat{p}_h$ = Proportion of sample in stratum h

Copyright by Mahidol University

**Table 6.** Formula for proportion estimator within "r" percentage points

| **Allocation** | **n** | **$n_h$** |
|---|---|---|
| Equal sample | $n = \dfrac{Z^2_{\frac{\alpha}{2}} L \sum_{h=1}^{L} N_h^2 \frac{N_h}{N_{h-1}} \hat{p}_h(1-\hat{p}_h)}{r^2(\sum_{h=1}^{L} N_h P_h)^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h \frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}$ | $n_h = \dfrac{n}{L}$ |
| Proportional allocation | $n = \dfrac{Z^2_{\frac{\alpha}{2}} N \sum_{h=1}^{L} N_h \frac{N_h}{N_{h-1}} \hat{p}_h(1-\hat{p}_h)}{r^2(\sum_{h=1}^{L} N_h P_h)^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h \frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}$ | $n_h = \dfrac{N_h}{\sum_{h=1}^{L} N_h} n$ |
| Neyman's allocation | $n = \dfrac{Z^2_{\frac{\alpha}{2}} \left[ \sum_{h=1}^{l} N_h \sqrt{\frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)} \right]^2}{r^2(\sum_{h=1}^{L} N_h P_h)^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h \frac{N_h}{N_{h-1}} \hat{p}_h(1-\hat{p}_h)}$ | $n_h = \dfrac{N_h \sqrt{p_h q_h}}{\sum_{h=1}^{L} N_h p_h q_h} n$ |
| Optimum allocation for fixed cost | $n = \dfrac{Z^2_{\frac{\alpha}{2}} \left( \sum_{h=1}^{L} N_h \sqrt{\frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h) c_h} \right) \left( \frac{\sum_{h=1}^{L} N_h \sqrt{\frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}}{\sqrt{c_h}} \right)}{r^2(\sum_{h=1}^{L} N_h P_h)^2 + Z^2_{\frac{\alpha}{2}} \sum_{h=1}^{L} N_h \frac{N_h}{N_h - 1} \hat{p}_h(1-\hat{p}_h)}$ | $n_h = \dfrac{\frac{N_h \sqrt{p_h q_h}}{\sqrt{c_h}}}{\sum_{h=1}^{L} \frac{N_h p_h q_h}{\sqrt{c_h}}} n$ |

$N_h$ = Total number of units in stratum h

$n_h$ = The number of units in the sample from stratum

$c_h$ = The cost per unit observed in stratum h

L = Number of stratum

$\hat{p}_h$ = Proportion of sample in stratum h

$P_h$ = Proportion of population in stratum h

Copyright by Mahidol University

### 1.2.4. Cluster Sampling

A method for drawing a sample from a population in two or more stages. It is typically used when researchers cannot get a complete list of the members of a population they wish to study but can get a complete list of groups or "clusters" of the population. It is also used when a random sample would produce a list of subject so widely scattered that surveying them would be prohibitively expensive. Generally, the researcher wishes to get cluster containing subjects as diverse as possible, whereas in stratified sampling the goal is often to find strata with subjects as similar to one another as possible.

The disadvantage of cluster sampling is that each stage of the process increases sampling error. The margin of error is therefore larger in cluster sampling than in simple or stratified random sampling; but, because cluster sampling is usually much easier (cheaper), this error can be compensated for by increasing the sample size. (2)

**Mean Estimator**

$$\hat{V}ar(\bar{x})_{srs} = \frac{s^2}{n}$$

$$\hat{V}ar(\bar{x})_{cluster} = \frac{\sum_{i=1}^{n} (\bar{x}_i - \bar{G})^2}{n(n-1)}$$

$$\text{Design effect} = \frac{\hat{V}ar(\bar{x})_{cluster}}{\hat{V}ar(\bar{x})_{srs}}$$

$$n_{cluster} = \text{design effect} * n_{srs}$$

$x_i$ = Mean in cluster i

$\bar{G}$ = Grand mean

n = Number of clusters

$\hat{V}ar(\bar{x})_{cluster}$ = Variance of $\bar{x}$ in cluster sampling

Copyright by Mahidol University

$\hat{V}ar(\bar{x})_{srs}$ = Variance of $\bar{x}$ in simple random sampling

$n_{srs}$ = Sample size for simple random sampling

$n_{cluster}$ = Sample size for cluster

**Proportion Estimator**

$$\hat{V}ar(\hat{p})_{srs} = \frac{\hat{p}(1-\hat{p})}{n}$$

$$\hat{V}ar(\hat{p})_{cluster} = \frac{\sum_{i=1}^{n}(p_i - p)^2}{k(k-1)}$$

$$\text{Design effect} = \frac{\hat{V}ar(\hat{p})_{cluster}}{\hat{V}ar(\hat{p})_{srs}}$$

$n_{cluster}$ = design effect * $n_{srs}$

$p_i$ = Proportion in cluster i

p = Global proportion

k = Number of clusters

$\hat{V}ar(\hat{p})_{cluster}$ = Variance of $\hat{p}$ in cluster sampling

$\hat{V}ar(\hat{p})_{srs}$ = Variance of $\hat{p}$ in simple random sampling

$n_{srs}$ = Sample size for simple random sampling

$n_{cluster}$ = Sample size for cluster

### 1.3 Sample Size for Observational study

#### 1.3.1 Case-Control Study

**Definition**

"Case- control", "Case – history", "Retrospective", "Backward- looking" refers to an important form of epidemiological study under – taken to test hypotheses regarding the causation of disease. This type of study aims to look backward from the EFFECT to CAUSE.

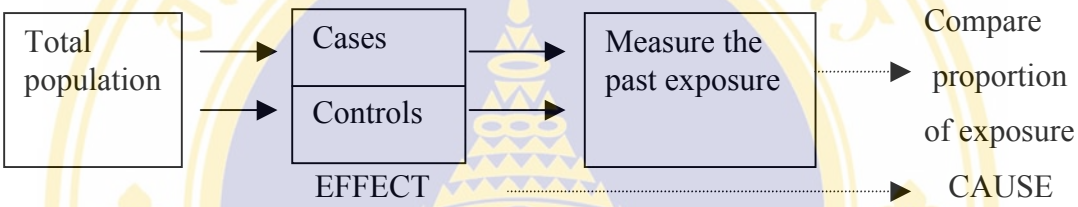
The hypothesis is tested through the comparison of the frequency of suspected causal factors in two groups of people, namely, Case and Control, which are as

Copyright by Mahidol University

similar as possible in all respects except for the presence or absence of the disease under study. (8)

**The Features**

1. All relevant events under study have already occurred by the time the study begins
2. The study involves the selection of a group having the disease (the cases )and appropriate group without the disease (the control )
3. The study is an indirect approach to the question of exposure versus increased risk, and the measurement of risk is called Relative Odds Ratio. The estimation of frequency of exposure (exposure rate) among cases and controls can be computed.




**Analysis and Interpretation**

Odds Ratio

| | Disease (D+) | non disease (D-) | |
|---|---|---|---|
| Exposed (E+) | a | b | a+b |
| Non – exposed (E- ) | c | d | c+d |
| | a+c | b+d | N |

Relative risk can be estimated, using Odds Ratio as:

$$OR = ad / bc$$

The number of subjects to be selected for study of a specific disease exposure relationship is a fundamental consideration in planning a case-control investigation. A study should be sufficiently large to avoid two sources of error: 1 claiming that exposure is associated with disease when, in fact it is not, and 2 claiming that exposure is not associated with disease, when in fact it is. When the frequency of

Copyright by Mahidol University

exposure between cases and controls is compared by means of a statistical test, the probability of making the first error is called the level of significance, and is commonly denoted by "α." The probability of making the second error is represented by "β" and the quantity 1-β is called the power of the study. Assuming that the relative risk in the target population differs from unity, the power (1-β) is the probability of finding that the sample estimate of relative risk (odd ratio) differs significantly from unity

Basically, an answer to the question of how many subjects should be selected for a case-control study depends on the specification of four values:1. the relative frequency of exposure among controls in the target population, $p_0$ 2. a hypothesized relative risk associated with exposure that would have sufficient biologic or public health importance to warrant its detection, R 3. the desired level of significance, α; 4. the desired study power, 1-β.

**1.3.1.1 Sample Size and Power for Unmatched Studies**

**Sample Size**

Suppose that one want to assess whether a particular exposure, regarded simply as having been either present or absent. Is associated with an alteration in the risk of a particular disease. For unmatched samples of cases and controls, the odd ratio in a simple 2 x 2 table can be tested by means of the standard chi-square test. Let $p_0$ denote the estimated exposure rate (proportion exposed) among controls, and let R denote the relative risk corresponding to the smallest increase or decrease in risk of interest. Specifying desired values for α and β, such as α=0.05 and β = 0.01, and assuming that equal numbers of cases and controls will be selected, the required sample size for each group (n per group) is calculated as follows.

$$n = \frac{Z_\alpha \sqrt{2\overline{pq}} + Z_\beta \sqrt{p_1 q_1 + p_0 q_0})^2}{(p_1 - p_0)^2} \longrightarrow (1)$$

Copyright by Mahidol University

$$p_1 = \frac{p_0 RR}{[1 + p_0(R-1)]} \longrightarrow (2)$$

$$\bar{p} = \frac{p_1 + p_0}{2} \qquad \bar{q} = 1 - \bar{p}$$

$$q_1 = 1 - p_1 \qquad q_0 = 1 - p_0$$

A formula that is simple than 1 and for practical purposes equivalent, is given by

$$n = \frac{2\bar{p}\bar{q}(Z_\alpha + Z_\beta)^2}{(p_1 - p_0)^2} \longrightarrow (3)$$

Corresponding to $\alpha = 0.05$ (two-sided) and $\beta = 0.10$, one has $Z_\alpha = 1.96$ and $Z_\beta = 1.28$, so that equation 3 reduces to a particularly simple formula:

$$n = \frac{21\bar{p}\bar{q}}{(p_1 - p_0)^2}$$

**Table7.** Unit normal deviates $Z_\alpha$ and $Z_\beta$ for Selected Values of $\alpha$ and $\beta$.

| **Values** $\alpha$ and $\beta$ | **One-sided test** $Z_\alpha$ or $Z_\beta$ | **Two-sided test** $Z_\alpha$ |
|---|---|---|
| 0.005 | 2.576 | 2.813 |
| 0.01 | 2.326 | 2.576 |
| 0.025 | 1.960 | 2.242 |
| 0.05 | 1.645 | 1.960 |
| 0.10 | 1.282 | 1.645 |
| 0.20 | 0.842 | 1.282 |
| 0.30 | 0.524 | 1.036 |

**Study Power**

In practice, the size of a study is often restricted by financial resources, the number of available cases, or a time limitation. For example, an important policy decision concerning an industrial, environmental, or drug exposure may require that a study be done quickly. In this situation the rate of accrual of cases will limit the size

Copyright by Mahidol University

of the case series. Similarly, the rate of accrual of cases will limit the size of the case series. Similarly, the number of cancers of specific type in a tumor registry will restrict the size of a case series based on prevalent cases. If a study were proposed within a fixed administrative setting, cases arising within a given hospital, or from the facilities of a prepaid medical plan, the case series would be limited by the size and type of population covered. Although in principle one can always extend the duration of study or include more catchment units, such an option may not be considered practical. Finally, there are always financial constraints and limitations of staffing

As a consequence of various constraints on the size of an investigation, a sample size analysis often begins with a fixed value for n, and then determines the resulting power (1-β) of detecting various values of R. Inversion of formulas 1 and 3 gives

$$\hat{Z}_\beta = \frac{[\sqrt{n(p_1 - p_0)^2} - Z_\alpha \sqrt{2\overline{pq}}]}{\sqrt{p_1 q_1 + p_0 q_0}} \longrightarrow (4)$$

and

$$\hat{Z}_\beta = [\frac{n(p_1 - p_0)^2}{2\overline{pq}}]^{1/2} - Z_\alpha \longrightarrow (5)$$

respectively. The power is determined from tables of the normal distribution by finding the probability with which the calculated value of $\hat{Z}_\beta$ is not exceeded:

$$\text{Power} = 1-\beta = P(Z \le \hat{Z}_\beta)$$

**1.3.1.2 Sample Size and Power with Multiple Controls per Case**

**Sample Size Determination with k controls per case unmatched studies**

A simple extension of equation 1 and 3 applies to the determination of sample size with multiple controls per case. With k controls per case, the sample size formula corresponding to 1 is

$$n = \frac{[Z_\alpha \sqrt{(1 + 1/k)\overline{p}'\overline{q}'} + Z_\beta \sqrt{p_1 q_1 + p_0 q_0 / k}]^2}{(p_1 - p_0)^2} \longrightarrow (6)$$

where

$$p_1 = \frac{p_0 RR}{[1 + p_0(RR - 1)]}$$

$$\overline{p}' = (p_1 + k p_0)/(1 + k) \longrightarrow (7)$$

Copyright by Mahidol University

$\bar{q}' = 1 - \bar{p}'$, and $p_1$ is calculated using equation 2

The term n represents the number of cases to be taken, so that kn gives the number of controls. Note that when k =1, indicating a 1:1 case: control ratio, formula 6 reduces to 1. The term k need not be an integer. Thus taking k = 0.8, implying more cases than controls, or k = 1.5 is perfectly acceptable. The sample size formula which corresponds to equation 3 is

$$n = \frac{(1+\frac{1}{k})\bar{p}'\bar{q}'(Z_\alpha + Z_\beta)^2}{(p_1 - p_0)^2} \quad \rightarrow \quad (8)$$

**Study Power with k controls per case unmatched studies**

To evaluate the power of a study having unequal numbers of cases and controls, one inverts equation 6 or 8, solving for $Z_\beta$:

$$\hat{Z}_\beta = \frac{[\sqrt{n(p_1 - p_0)^2} - Z_\alpha\sqrt{(1+1/k)\bar{p}'\bar{q}'}]}{\sqrt{p_1q_1 + p_0q_0/k}} \quad \rightarrow \quad (9)$$

and

$$\hat{Z}_\beta = [\frac{n(p_1 - p_0)^2}{(1+\frac{1}{k})\bar{p}'\bar{q}'}]^{1/2} - Z_\alpha \quad \rightarrow \quad (10)$$

In the above formula, the term n represents the number of cases, and the term k represents the number of controls per case. The power is determined from tables of the standard normal distribution by find the probability with which the calculated value of

$\hat{Z}_\beta$ is not exceed : Power = $P(\left(Z \le \hat{Z}_\beta\right))$

Equation 9 and 10 may be used either in planning an investigation or in evaluating the power of a completed study

Copyright by Mahidol University

**1.3.1.3 Sample Size and Power for Matched Studies**

**Sample Size**

For each pair in a matched case – control study, there are four possible exposures out comes. Denoting an exposed individual by (+) and an unexposed individual by (-), one may write these as (case, control): (++), (+-), (-+), and (--). Thus, for example, (+-) indicates that the case was exposed and the control was not exposed. For specified $\alpha$and $\beta$, the number of discordant pairs required to detect a relative risk R is given by

$$m = \frac{Z_{\frac{\alpha}{2}} + Z_{\beta}\sqrt{P(1-P)}]^2}{(P-\frac{1}{2})^2} \quad \longrightarrow \quad (11)$$

where

$$P=RR/(1+RR) \quad \longrightarrow \quad (12)$$

The term RR denotes the disease-exposure odds ratio and RR denotes the relative risk. Letting $p_e$ denote the probability of an exposure-discordant pair, the total number of pair M required on average to yield m discordant pairs is $M = m / p_e$. Although strictly speaking $p_e$ depends on the matching criteria, one may use as a crude approximation

$$p_e = (p_0 q_1 + p_1 q_0) \quad \longrightarrow \quad (13)$$

where $p_o$ is the estimated proportion of exposed controls in the target population and $p_1$ is the estimated proportion of exposed cases determined from equation 2. The quantities $q_o$ and $q_1$ are defined by $q_o=1-p_o$ and $q_1=1-p_1$, respectively. Thus

$$M = \frac{m}{(p_0 q_1 + p_1 q_0)} \quad \longrightarrow \quad (14)$$

As an application of the preceding sample size analysis, consider doing a matched case-control study of oral contraceptives Assume again that the proportion of exposed controls is $p_o=0.3$ and that $\alpha=0.05$(two-sided) and $\beta=0.1$. In order to detect a twofold increased risk (RR=2) with these specifications, one needs 90 exposure discordant pairs:

Copyright by Mahidol University

$$m=[1.96/2+1.28\sqrt{(2/3)(1/3)}]^2/(2/3-1/2)^2=90$$

As a crude approximation to the total number of pairs that are required, one calculates $p_e = (0.3)(0.54)+(0.46)(0.7) = .484$ from 13, so that M = 186 from equation 14. For an unmatched study with the above specifications, one estimates a sample size of n = 188 per group.

**Power**

$Z_\beta$ allows one to calculate the power of a matched study based on m discordant pairs:

$$\hat{Z}_\beta = [-Z_{\frac{\alpha}{2}} + \sqrt{m(P-1/2)^2}]/\sqrt{P(1-P)} \longrightarrow (15)$$

For specified values of m, R, and $\alpha$ can be used to calculate $\hat{Z}_\beta$ . The power is then determined from tables of the normal distribution by finding the probability that $\hat{Z}_\beta$ is not exceeded, Power = P( $Z \leq \hat{Z}_\beta$ ).(8)

**1.3.2. Cohort Study**

**Definition**

The word "cohort", "Incidence", "longitudinal", "prospective" and "forward – looking" study can be used interchangeably and refers to study which is one important form of epidemiological study undertaken to test hypotheses regarding the causation of disease. It is the type of observational study that looks from CAUSE TO EFFECT.

**The Features**

a. The study starts with a group of individuals with and without an etiological characteristic or exposure to a suspected etiologic factor (x) , and all of them are free from the specified disease to be studied ( disease – free cohort )

b. Both exposed group and non – exposed group (control group) are observed over a period of time (weeks, month, or years) and the development of the disease in all subjects is periodically searched for.

c. The observed incidence rates of disease (D) in the two groups are compared.

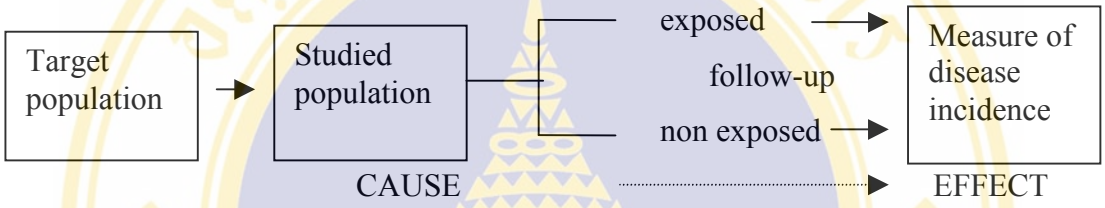
Copyright by Mahidol University

d. This study provides direct estimates of the risk of developing disease in both groups which are supposed to be similar in all other characteristics. (9)

**Analysis and Interpretation**

At the end of the study, the data are analyzed in terms of incidence rates and the relative risk in order to determine the association between the suspected cause and effect

BASIC STUDY FORMAT




**Relative Risk (RR)**

| | Disease (D+) | non disease (D-) | |
|---|---|---|---|
| Exposed (E +) | a | b | a+b |
| Non – exposed (E - ) | c | d | c+d |
| | a+c | b+d | N |

$$RR = \frac{a / (a+b)}{c / (c+d)}$$

**Sample Size Determination**

Formula 1

$$n_{group} = \frac{\left\{Z_{\alpha}\sqrt{2\bar{p}\bar{q}} + Z_{\beta}\sqrt{p_1[1+RR-p_1(1+RR^2)]}\right\}^2}{(p_1(1-RR))^2}$$

n = Number of sample for each group

$p_1$ = Probability of disease among non - exposed

Copyright by Mahidol University

RR = Relative risk (of public health importance)

$$\bar{p} = \frac{p_1(1-RR)}{2} \qquad \bar{q} = 1-\bar{p}$$

Formula 2

$$n = \frac{\left\{ Z_{\alpha/2}\sqrt{2\bar{p}\bar{q}} + Z_{\beta}\sqrt{p_1(1-p_1)+p_2(1-p_2)} \right\}^2}{(p_1-p_2)^2}$$

$$\bar{p} = \frac{(p_1-p_2)}{2} \qquad p_1 = RR \times p_2$$

$$p = (p_1+p_2)/2 = p_2(RR+1)/2 \qquad \bar{q} = 1-\bar{p}$$

### 1.4 Sample Size for Logistic Regression

#### 1.4.1 Simple Logistic Regression

In a simple logistic regression model, relate a covariate $X_1$ to the binary response variable Y in a model log (P/(1-P)) = $\beta_0 + \beta_1 X_1$ where P = prob(Y =1). We are interested in testing the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1: \beta_1 = \beta^*$, where $\beta^* \neq 0$, that the covariate is related to the binary response variable. When the covariate is a binary variable, say X = 0 or 1, the log odds value $\beta_1 = 0$ if and only if the two event rates are equal. The sample size formula for the total sample size required for comparing two independent event rates has the following form formula

$$n = \frac{\left( Z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{B}} + Z_{1-\beta}\sqrt{P_1(1-P_1)+\frac{P_2(1-P_2)(1-B)}{B}} \right)^2}{[(P_1-P_2)^2(1-B)}$$

where : $P = (1-B)P_1 + BP_2$ is the overall event rate ;B is the proportion of the sample with X=1 ; $P_1$ and $P_2$ are the event rates at X = 0 and X = 1, respectively.

Copyright by Mahidol University

### 1.4.2 Multiple Logistic Regression

When there is more than one covariate in the model, a hypothesis of interest is the effect of a specific covariate in the presence of other covariates. In terms of log odds parameters, the null hypothesis for multiple logistic regression is $H_0 : [\beta_1,\beta_2,\ldots,\beta_p] = [0,\beta_{2,\ldots,}\beta_p]$ against the alternative $[\beta^*,\beta_2,\ldots,\beta_p]$. Let $b_1$ be the maximum likelihood estimate of $\beta_1$. Whittemore has shown that, for continuous, normal covariates X, the variance of $b_1$ in the multivariable setting with p covariates, var $_p(b_1)$,can be approximated by inflating the variance of $b_1$ obtained from the one parameter model,$var_1(b_1)$,by multiplying by $1/(1-\rho^2_{1.2\ 3\ldots.p)}$ where $\rho_{1.2\ 3\ \ldots p}$ is the multiple correlation coefficient relating $X_1$ with $X_2,\ldots..X_p$. That is, approximately

$$Var_p(b_1) = var_1(b_1)/(1-\rho^2_{1.23\ldots p})$$

The squared multiple correlation coefficient $\rho^2_{1.23\ldots p,}$ also known as $R^2$, is equal to the proportion of the variance of $X_1$ explained by the regression relationship with $X_2,\ldots..X_p$. The term $1/(1-\rho^2_{1.23\ldots p})$ will be referred to as a variance inflation factor (VIF). The required sample size for the multivariable case can also be approximated from the univariable case by inflating it with the same factor $1/(1-\rho^2_{1.23\ldots p})$. Following the relationship of the variances, we have

$$n_p = n_1/ (1 -\rho^2_{1.23\ldots p})$$

where $n_p$ and $n_1$ are the sample size required for a logistic regression model with p and 1 covariates, respectively. (10),(11)

## 1.5 Sample Size for Correlation

### 1.5.1 Test of a Single Correlation

$$n = (Z_\alpha + Z_\beta)^2 + 3Z_F$$

$Z_F =$ Fisher's Z transformation

$$Z = \frac{1}{2}\log_e\frac{1+r}{1-r} = 1.1513\ \log\left(\frac{1+r}{1-r}\right)$$

Copyright by Mahidol University

### 1.5.2 Test of Two Independent Correlations

$$n = \frac{(Z_{\alpha} + Z_{\beta})^2}{(Z_{F1} - Z_{F2})^2}$$

$Z_{F2}$ = Fisher's Z transformation 1(Group1)

$Z_{F2}$ = Fisher's Z transformation 2(Group2) (12)

## 2 Package for Determination of the Sample Size

### 2.1 SAM Program

SAM performs sample size calculations for four of the most common types of study involving the comparison of two groups (though sometimes we are interested in three or more groups). These study types are categorized by the intended analysis of the primary endpoint: a comparison of proportions, survival, means or, for case-control studies, examing the risk-ratio.

**Advantages of Using SAM Program**

1. It is an easy program can be used without handbooks.
2. Shows sample size varied to the change outcome.
3. When we know the sample size, we are able to inversely calculate to find the power value.

**Disadvantages of Using SAM Program**

1. Unable to calculate sample size in one or two independent samples.
2. Unable to calculate too high or too low sample size.

The size of program is 141 kb and it is able to run on DOS. (13)

### 2.2 Power Program

The Power program does not calculate the sample size required to estimate a single mean or a single proportion to with some specified level of precision

The Power program can provide sample size or power calculations for the following situations:

Copyright by Mahidol University

1. Comparison of mean for independent samples
2. Comparison of mean for dependent samples
3. Comparison of proportion for independent samples
4. Comparison of proportion for dependent samples

Study of survival times to be analysis using the log rank test

This Module only discusses situations 1 and 3

Given any two of the following study characteristics POWER can calculate the third:

1.power

2.sample size

3.detectable difference (clinically or biologically or socially important difference that is worth detecting)

**2.3 Minitab Program**

Minitab is a leader is a leader is delivering statistical software and services for qualityimprovement, education and research. Minitab statistical software provides tools to analyze data across a variety of disciplines and is targeted for users at every level: scientists, business and industrial users, faculty and student. Minitab is available on the most widely-used computer platforms, including windows, dos, macintosh, open VMS and Unix. Minitab is quite easy to learn and use, right out of the box. Pull-down menus and dialog boxes give you easy prompts every step of the way. Data are entered in Minitab's data window and may be imported directly from a variety of the formats, including Lotus, Excel, Syphony, Quattro Pro, dbase and text (ASCII) files. The price is $995 per 1 copies for the Minitab 13 for windows95, 98, 2000,windows me or windows NT4 or if you own a previous release of Minitab, upgrade to a new release for only $325.(These are the prices for customers in the US and Cannada)

Use Minitab's power and sample size capabilities to evaluate power and sample size before you design and run an experiment (prospective) or after you perform an experiment (retrospective). Minitab provide power, sample size, and difference (effect) calculation (also the number of center points for factorial and Plackett Burman designs) for the following procedures one sample Z, one sample t, two

Copyright by Mahidol University

sample t, one sample proportion, two sample proportion, one way analysis of variance, two level factorial designs, Plackett Burman designs (14)

### 2.4 Stata Program

Stata is an environment for manipulating and exploring data using statistical and graphical methods. Stata is not a collection of modules: Data management, statistical and graphical commands can be interspersed. Results from analyses can be saved and used in subsequent commands. Furthermore, Stata technical bulletin(STB), a bi-monthly publication, provides a forum for Stata user of all disciplines and levels of sophistication. STB plays an important role in making Stata a state-of-the-art statistical package. Stata is particularly useful to professionals working in the area of medical research, economic research and panel data. The prices is $995,New copy of stata7 for single-user license and $249 stat/transfer 6.0 for single-user license. Samsi estimates reguired sample size or power of tests for studies comparing two group sampsi can be used when comparing means or proportions for simple studies where only one measurement of the outcome is planned, and for comparing mean summary statisties for more complex studies where repeated measurement of the outcome on each experimental unit are planned.

For simple studies, where only one measurement of the outcome is plane, sampsi computes sample size or power for four types of tests

1. Two – Sample comparison of mean
2. One – Sample comparison of mean to hypothesized value
3. Two – Sample comparison of proportion
4. One – Sample comparison of proportion to hypothesized value (15)

### 2.5 Epi Info Program

Epi Info is a series of programs for Microsoft Windows 95, 98, NT, and 2000 for use by public health professionals in conducting outbreak investigations, managing databases for public health surveillance and other tasks, and general database and statistics applications. With Epi Info and a personal computer, physicians, epidemiologists, and other public health and medical workers can rapidly

Copyright by Mahidol University

develop a questionnaire or form, customize the data entry process, and enter and analyze data. Epi Info is written in Visual Basic Edition 6.0, from Microsoft, Inc.

The majority of statistical and epidemiological analyses performed by Epi Info are based on the assumption that the data were collected from an infinitely large population using simple random sampling. Point estimates, such as proportions, risk ratios, odds ratios, means, etc., are sample estimates. Epi Info program be able to calculated sample size 4 such as single proportion, two proportion, cohort study, case control study and be able to find the value of power (case control and cohort). (16), (17)

**2.6 East 2000 Program**

East 2000, a Microsoft Windows software package for the design and interim monitoring of group sequential clinical trials. East 2000 software adds considerable functionality to its Ms-Dos predecessor through a superior user interface and through the addition of new statistical methods. New user interface East 2000 is developed on the Microsoft Windows platform. It supports a highly interactive user interface with read access to stopping boundary charts, error spending function charts, power charts and the ability to present the results as reports in Microsoft office. New statistical methods east 2000 has greatly expanded the design and interim monitoring capabilities previously available in East-Dos. In addition east 2000 provides a simulation module for investigating how the power of a sequential design is affected by different assumptions about the magnitude of the treatment difference.(18)

Copyright by Mahidol University

# CHAPTER III
# METHODOLOGY

This chapter is to mention

1. Research instruments
2. Step for writing A Software package for determination of the sample size
3. The sample size determination solution comparing with other programs

## 1. Research Instruments

The instruments used in writing this program are Hardware, Software and Microsoft Visual Basic Edition 6.0 program.
as follows:

### 1.1 Hardware:

| | | |
|---|---|---|
| CPU | : | Intel Pentium Compatible |
| RAM | : | At least 64 MB |
| Hard disk | : | At least 1 GB |
| VGA Card Ram | : | At least 2 MB |

### 1.2 Software:

| | | |
|---|---|---|
| Operating System | : | Microsoft Windows 95 or Windows98 |
| Object - Oriented Application | : | Microsoft Visual Basic Edition 6.0 |
| HTML program | : | Microsoft Front Page 2000 |
| Help program | : | HTML Help Workshop |
| Picture program | : | Adobe Photoshop 5.5 |
| Document tools | : | Microsoft word 97 |

Copyright by Mahidol University

**1.3. Microsoft Visual Basic Program**.

Today, there are many programs used in writing software package such as : "C, Pascal, Fortran, Delphi, and Visual Basic program. In writing the Software Package for sample size, the developer chooses to use the Microsoft Visual Basic as it is not complicated and convenient to use. Microsoft Visual Basic is the program famously used to develop and write the Application Programs as Epi Info 2000 program.

Microsoft Visual Basic program is written under the Windows program designed by using visualize, putting objects in the monitor to communicate with the user. These objects will change accordingly with events such as to move the mouse or type the keyboard to command. To develop the program on Windows by using Microsoft Visual Basic which has less steps to do, the writer just chooses suitable form and control, then, draw it in the screen. And then, put the code into the Microsoft Visual Basic, which it saves a lot of time in study. Microsoft Visual Basic program is appropriate to users in different levels, from beginner to professional.(19)

**2. Steps for Writing Software Package for Determination of the Sample Size**

Software package for determination of the sample size consists of 5 steps. (Figure1)

2.1. Review of the literature

2.2. Design a software package for determination of the sample size

2.3. Develop a software package for determination of the sample size

2.4. Test a software package for determination of the sample size

2.5. Writing a research report

Copyright by Mahidol University

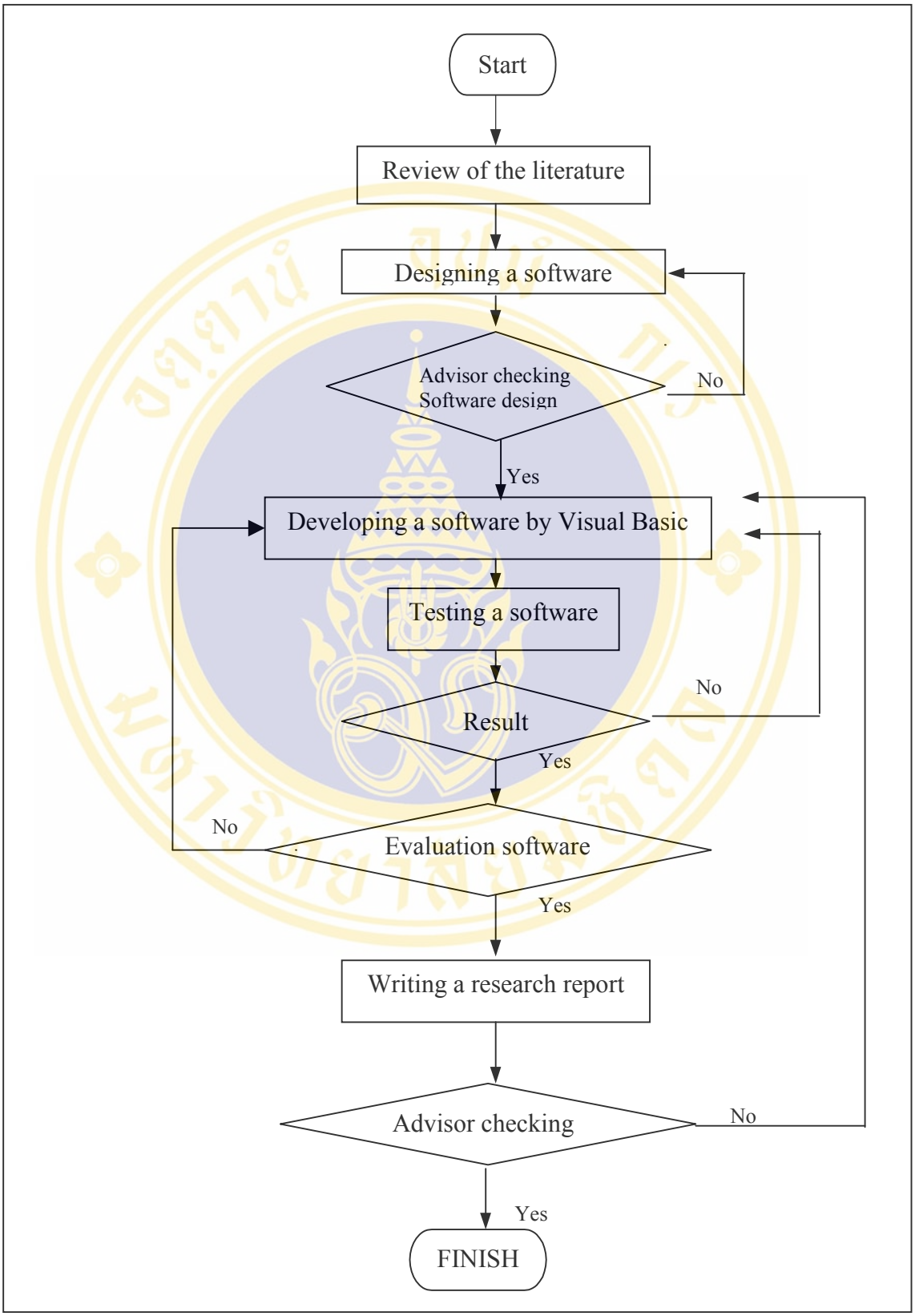



**Figure 1** Research flowchart

Copyright by Mahidol University

**2.1 Review of the Literature**

Review theories and related information work from printed materials, for examples, books, journals, research papers, technical report and thesis. Furthermore, collect information in electronic material that can find from many virtual libraries and organization pages in World Wide Web.

Study and learn to understand the methods of Object-Oriented Analysis and Design (OOA and OOD) before real system analysis and design next. Study and learn to understand the Object-Oriented application program that are expected to use to develop this a software package for determination of the sample size.

Microsoft Visual Basic Professional Edition 6.0: an Object-Oriented, visual programming applications for Microsoft Windows 98.

**2.2 Design a Software Package for Determination of the Sample Size**

First, to design the interface plotting how it is coming out. The developer intends to use menu bars, which each menu bar will show how to find sample size. When the user clicks each menu bar, the forms will be displayed leading to calculation of sample size. The user inputs data in text boxes, and then press Calculate button, which the result will be issued right the way. However, if there is such mistake in calculation as putting data out of range and incomplete, the symbol “Error” will show warning the user. This program is able to be saved and printed out. (Figure 2)

Copyright by Mahidol University

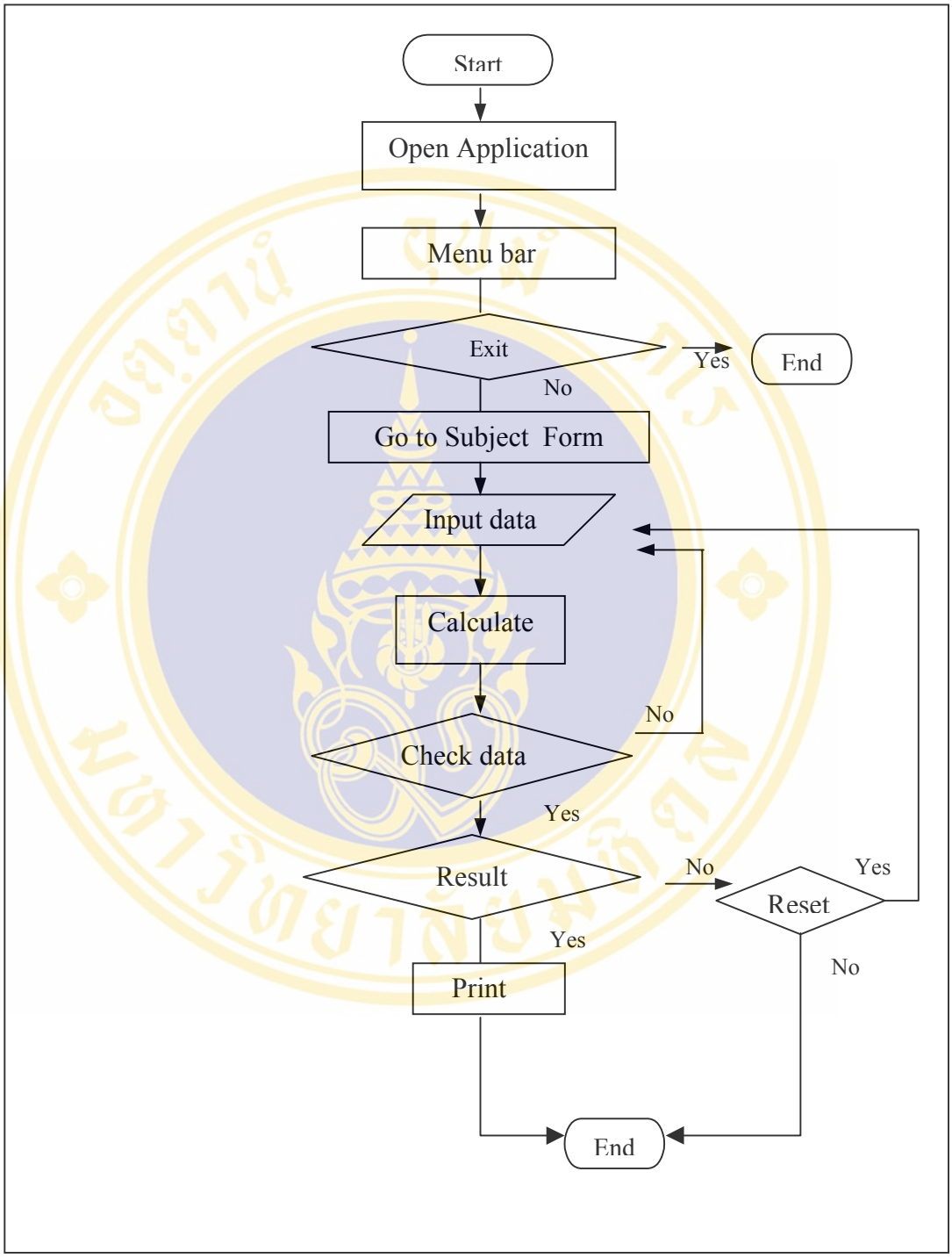



**Figure 2.** Designing a software package flowchart

Copyright by Mahidol University

**2.3 Develop a Software Package for Determination of the Sample Size**

A Software package for determination of the sample size will be written in Basic language. Microsoft Visual Basic version 6 can be used on Microsoft Window 98 to create program interface. The Microsoft Visual Basic can be used conveniently in writing the program since it is fully equipped with pre-programmed instruments. Next is to write the code to making interface ready to work. There is complexity in writing codes as well as it is easy to make error leading to functional mistake. Good developer must be very careful in writing codes as much as the test must be done all the way through whether or not the work is done appropriately. Some forms must be done through the Microsoft Excel program, commanded linking to the Microsoft Visual Basic to the Microsoft Excel.

**2.4 Test a Software Package for Determination of the Sample Size**

The testing is aimed at the program performs correctly function and satisfies the requirement of user. Testing phrases are divided into 4 steps:

2.4.1 Object Testing

2.4.2 Form Testing

2.4.3 System Testing

2.4.4 Software Evaluation

**2.4.1 Object Testing**

Objects mean commands, text boxes, list boxes, Grid table, key buttons and other controls that are combined into a Form. Object testing is a testing individual object in each form to ensure that they operate correctly. For example, “Print” command was tested that it retrieves correct data from database and prints them to printer correctly and perfectly. All of objects must be tested. Result of testing is that object perform the correct functions.

**2.4.2 Form Testing**

It is a testing the overall functions of objects in individual interface forms. One form consists of several objects. Each object has a unique function. When they are

Copyright by Mahidol University

integrated into a form, they may operate unsuitable functions with one another or some errors arise. Therefore, form testing is needed for checking these unpredictable errors.

### 2.4.3 System Testing

The forms are integrated to make up system. When individual form is tested until satisfaction, another importance testing is an integrated testing or called system testing. The testing is concerned with finding error with result from unanticipated interactions between forms and form objects.

### 2.4.4 Software Evaluation

When the program is complete, it needs to be evaluated by users. The evaluators come from different social status and education level such as statisticians, researchers, lecturers and students who have knowledge concerning sample size determination. The A-SIZE program was given to the evaluators with the instruction of the program had been given prior to the using of the program, and the users can work it out with no time frame limit. After the program had been observed satisfactorily, the users answered the evaluation form giving the result to the program developer. Any suggestions are valuable to the developer concerning program development in the future if needed.

### 2.5 Writing a Research Report

After all the above mentioned steps have been done, user manual and research documentation are made up for giving information about research and program.

## 3. The sample size determination solution comparing with other programs

There are many programs which is capable to be used in sample size determination: SAM program, Power program, Epi Info program, Stata program, Minitab program and East 2,000, which each program is distinctively different in sample size determination. This is the way to compare the A-SIZE program in sample size determination with others. (Table 8)

Copyright by Mahidol University

**Table 8. Comparison of Seven Software for Sample Size Calculation**

| | SAM* | Power* | Epi Info** | Stata** | Minitab** | East 2000** | A-SIZE** |
|---|---|---|---|---|---|---|---|
| **Sample size for sample surveys** | | | | | | | |
| **Estimation** | | | | | | | |
| **Simple random Sampling / Systematic sampling** | | | | | | | |
| **Estimation of parameter within allowable absolute error"d"** | | | | | | | |
| **One sample** | | | | | | | |
| - Total | × | × | × | × | × | × | ✓ |
| - Mean | × | × | × | × | ✓ | × | ✓ |
| - Proportion | × | × | ✓ | × | × | × | ✓ |
| - Ratio | × | × | × | × | × | × | ✓ |
| **Two sample** | | | | | | | |
| - Total | × | × | × | × | × | × | ✓ |
| - Mean | × | × | × | × | ✓ | × | ✓ |
| - Proportion | × | × | ✓ | × | × | × | ✓ |
| - Ratio | × | × | × | × | × | × | ✓ |
| **Estimation of parameter within relative error "r"** | | | | | | | |
| - Total | × | × | × | × | × | × | ✓ |
| - Mean | × | × | × | × | × | × | ✓ |
| - Proportion | × | × | × | × | × | × | ✓ |
| **Stratified sampling** | | | | | | | |
| - Total | × | × | × | × | × | × | ✓ |
| - Mean | × | × | × | × | × | × | ✓ |
| - Proportion | × | × | × | × | × | × | ✓ |
| **Cluster sampling** | | | | | | | |
| - Mean | × | × | × | × | × | × | ✓ |
| - Proportion | × | × | × | × | × | × | ✓ |
| **Hypothesis testing** | | | | | | | |
| **One – sample** | | | | | | | |
| - Mean | × | × | × | ✓ | × | × | ✓ |
| - Proportion | × | × | × | ✓ | ✓ | × | ✓ |
| **Two – sample** | | | | | | | |
| - Mean | ✓ | ✓ | × | ✓ | × | × | ✓ |
| - Proportion | ✓ | ✓ | × | ✓ | ✓ | × | ✓ |
| **Sample size for Observational study** | | | | | | | |
| **Case – control study** | ✓ | × | ✓ | × | × | × | ✓ |
| **Cohort study** | × | × | ✓ | × | × | × | ✓ |
| **Sample size for logistic regression** | | | | | | | |
| **Logistic regression** | × | × | × | × | × | × | ✓ |
| **Sample size for correlation** | | | | | | | |
| **Correlation** | × | × | × | × | × | × | ✓ |
| **Power of test** | × | ✓ | ✓ | × | ✓ | × | ✓ |
| **One-way ANOVA** | × | × | × | × | ✓ | × | × |
| **Clinical trial** | × | × | × | × | × | ✓ | × |
| **Survival** | ✓ | ✓ | × | × | × | × | × |

* = run on DOS ** = run on Windows

Copyright by Mahidol University

# CHAPTER IV
# RESULTS

There are 4 parts of the software package:

1. The program character.
2. Parts of A Software package for determination of the sample size.
3. Steps inputting data and Error data shown in data input.
4. The program evaluation by the users.

## 1. The Program Character

The program developer gives the name A-SIZE program, which is simple and plain to call. This program, A-SIZE, is 15 MB suitable to the computer with CPU starting from Pentium and up, having Ram of at least 32 MB and VGA of 800 X 600, working on the Microsoft Windows 95 or higher. If the capacity of the computer has not reached this requirement, the program will not work as well as it should be; or it won't work at all.

## 2. Parts of A Software Package for Determination of a Sample Size.

This program is composed of 6 menu bars as follows:

**2.1 Survey.** Determining sample size for sample survey to estimate parameter and hypothesis testing which has sub-menu as is shown below.(Figure3)

**Estimation** is the way to determine sample size for estimating parameter, which is divided into 3 designs: Simple random sampling/Systematic sampling, Stratified sampling, Cluster sampling.

These sampling designs, Simple random sampling/Systematic sampling, can be used to determine sample size in total estimation, mean, proportion and ratio of one

Copyright by Mahidol University

sample population and two sample population. In determination of sample size, there are two choices to make: allowable absolute error (d) and relative error(r).

Stratified sampling can be used to determine sample size in the total, mean and proportion estimation. Moreover, there are 3 ways to determine: Proportional allocation, Neyman's allocation, Optimum allocation for fixed cost. In order to determine the sample size as mentioned, allowable absolute error (d) or relative error (r) can also be used for the determination.

Cluster sampling can be used in determination of sample size, mean and proportion.

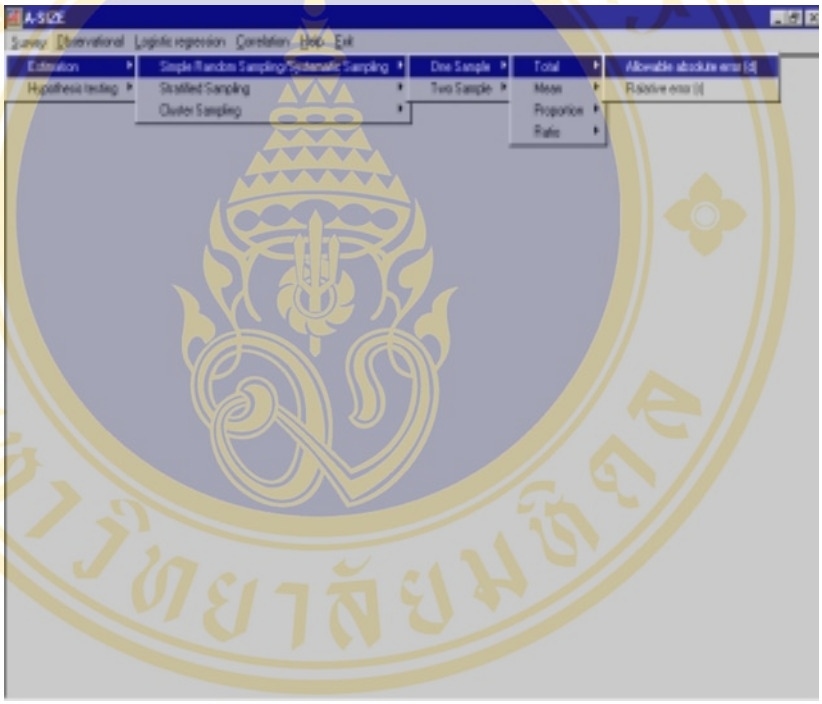



**Figure 3.** Show menu bar in estimation

**Hypothesis Testing** is sample size determination concerning mean and proportion which is used for one or two sampling group; further more, it can be used in one-sided test and two-sided test as well.(Figure 4)

Copyright by Mahidol University

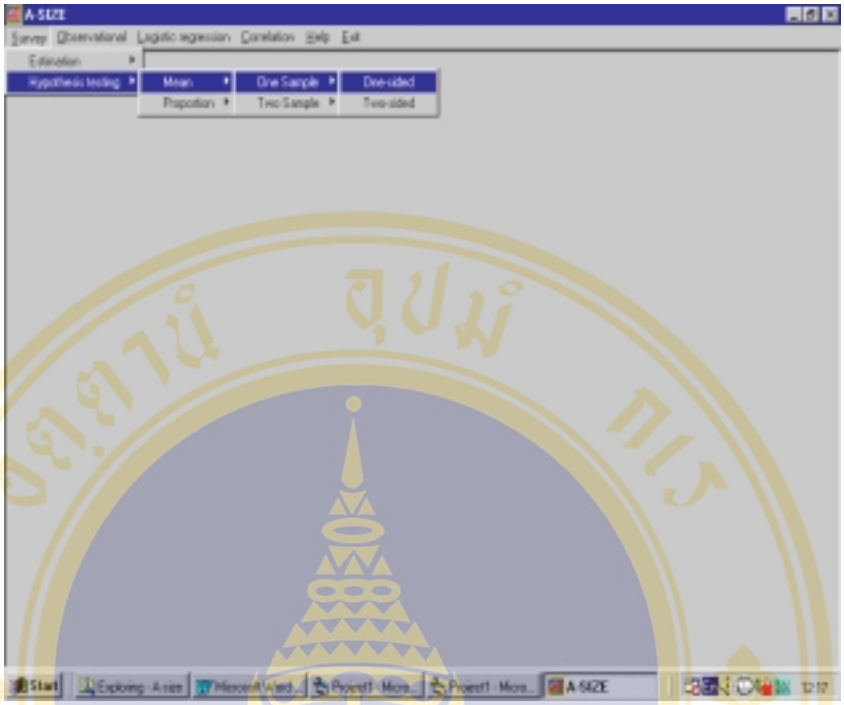



**Figure 4.** Show menu bar in hypothesis testing

**2.2 Observational study** is divided into two categories Case-Control study and Cohort study.(Figure5)

**Case – Control Study** is “Retrospective”, ”Backward-looking” refers to an important form of epidemiological study under- taken to test hypotheses regarding the causation of disease. This type of study aims to look backward from the effect to cause divided in to unmatched studies, k controls per case unmatched studies, matched studies. Each study is used to calculate sample size and power.

**Cohort** is “Incidence”, “longitudinal”, “prospective” and “forward – looking” study can be used interchangeably and refers to study which is one important form of epidemiological study undertaken to test hypotheses regarding the causation of disease. It is the type of observational study that looks from cause to effect.

Copyright by Mahidol University

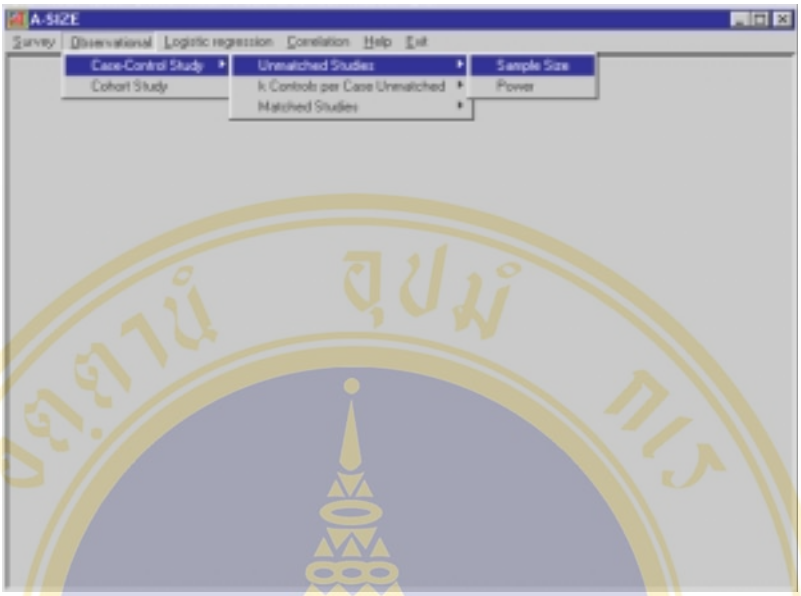



**Figure 5.** Show menu bar in sample size for observational study

**2.3 Logistic regression** is divided into two : Simple logistic regression and Multiple logistic regression.(Figure 6)

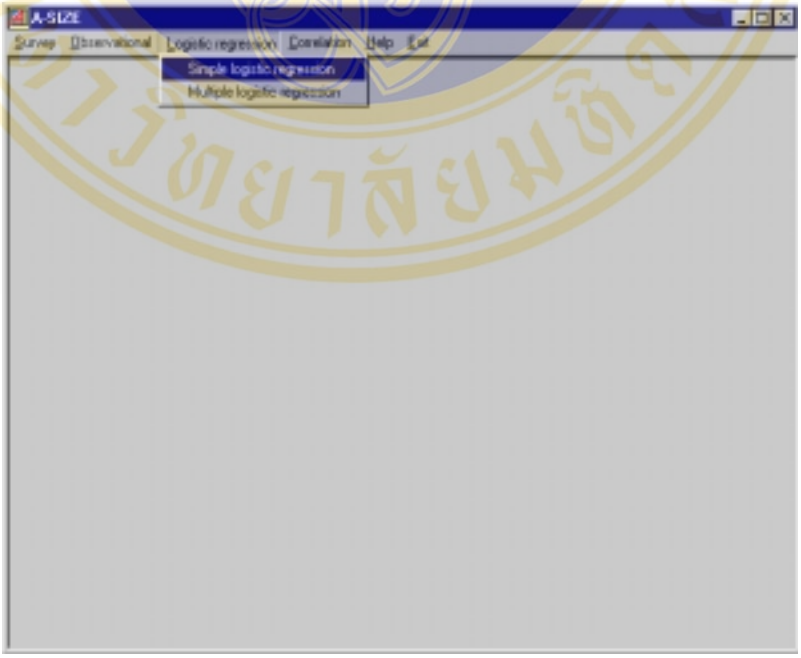



**Figure 6.** Show menu bar in sample size for logistic regression

Copyright by Mahidol University

**2.4 Correlation** is divided into two test of a single correlation and test of two independent correlation. (Figure 7)

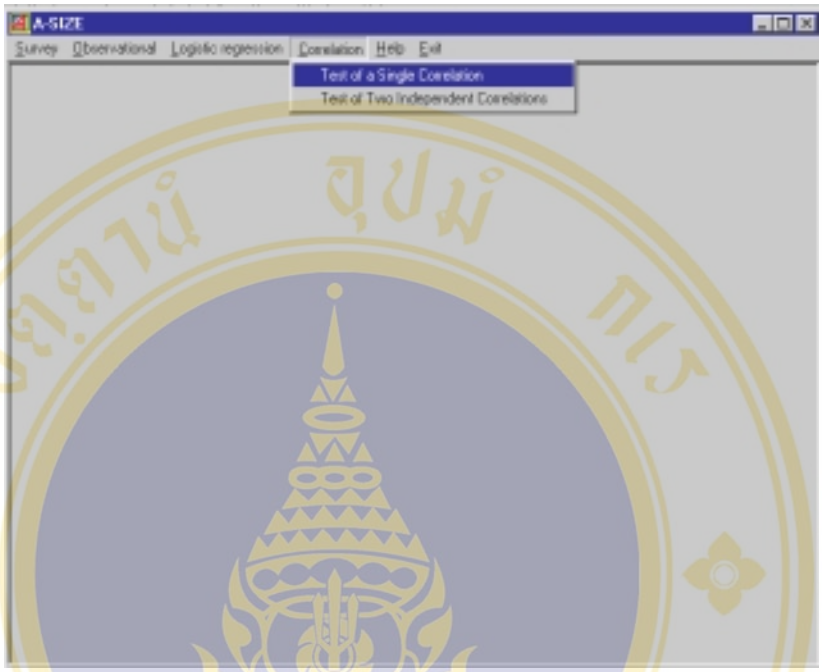



**Figure 7.** Show menu bar in sample size for correlation

**2.5 Help** or menu help is the menu used in giving explanation in each form, telling what the program has, including example and explanation in using the program.

**2.6 Exit** is the closing program

There is calculation form in the program. Each form is capable to calculate 1 thing. It is devised for save and print. The sample size will be given in term of text box and grid table. Scroll bar is movable up and down to see the sample size. The calculation can issue 50 sample sizes, but be able to print only 18 answers depending on a monitor size.

Copyright by Mahidol University

**3. Steps Inputting Data and Error Data Shown in Data Input**

When the first screen of the monitor is on, click a mouse to menu bar of given sampling design. When the sampling design form is appeared, data is needed to be put into each text box. There are two ways to move cursor from one dialog box to another: using mouse click and putting enter. In case there is a downward arrow, using upperward arrow↑ and downward arrow ↓ instead of clicking mouse. Press the Calculate button when data is completely input in dialog box and the answer will be given from the sample size dialog box and grid table. Press the print button when printing is needed. Press the print button then the form print will be appeared. The save button is an alternative if printing is not immediately needed, and it can be done by clicking a save button in term of .txt.

If data is not completely in put, the program will show error, "Input data not complete", then it needs to be done appropriately again. If the data in put is not in the information range, the signal, "Value out of range $(0 \le p \le 1)$" will be shown, then the data in put is needed again. The monitor illustration is given in Appendix B.

**4. The Program Evaluation by the Users**

Evaluated by users. This evaluation form is shown Appendix B is composed of 3 parts, 17 questions. Part one is evaluators' general data. Part two shows opinions concerning software using. Part three shows general opinions of evaluators divided into 3 parts: Advantages, Disadvantages and Recommendation for improvement.

**4.1 Evaluators' General Data.**

The results were shown that the samples were 71.4% female, 28.6% male. The majority of age was 20-29 years old (60.7%), minimum 22 years old, maximum 46 years old, mean 29.11 years old and standard deviation 5.44 years old. For the present occupation, 53.5% were researcher or statistician, 28.6% were student, 17.9 % were lecturer. The data were shown in the table 9

Copyright by Mahidol University

**Table9. Number and percentage of evaluators' general data**

| General data | Number | Percentage |
|---|---|---|
| **Sex** | | |
| Female | 20 | 71.4 |
| Male | 8 | 28.6 |
| **Age(years)** | | |
| 20-29 | 17 | 60.7 |
| 30-39 | 9 | 32.1 |
| ≥40 | 2 | 7.2 |
| Mean ± SD | 29.11 ± 5.44 | |
| Min- Max | 22-46 | |
| **The present occupation** | | |
| Researcher/Statistician | 15 | 53.5 |
| Student | 8 | 28.6 |
| Lecturer | 5 | 17.9 |

### 4.2 Sample Size Experiences of Evaluators

The majority of sample (85.7%) have experiences in area of the determination of sample size and all person (100.0%) have knowledge background of the determination of sample size. Approximately equal proportion used any software related to the determination of sample size. The data were shown in the table10.

**Table10. Number and Percentage Concerning Sample Size by Evaluators Experiences**

| Sample Size Experience | Number | Percentage |
|---|---|---|
| **Experiences in area of the determination of sample size** | | |
| Yes | 24 | 85.7 |
| No | 4 | 14.3 |

Copyright by Mahidol University

**Table10.** Continued

| Sample Size Experience | Number | Percentage |
|---|---|---|
| **Knowledge background of the determination of sample size** | | |
| Yes | 28 | 100.0 |
| No | 0 | 0.0 |
| **Used any software related to the determination of sample size** | | |
| Yes | 12 | 42.9 |
| No | 16 | 57.1 |

### 4.3 General opinions of evaluators

From 7 areas: 1.convenience of supplying data 2. Fastness of supplying data 3. Difficulty of amending data 4. Feature of software 5. Accuracy of sample size result 6. Sample size coverage and 7. Satisfaction of this program in whole, the evaluators gave different opinion score as followed: The highest score of 57.1%, very good, and mean of 3.25, in the area of fastness of supplying data. The second highest score of 46.4%, very good, and mean of 3.36, is in the area of convenience of supplying data. The lowest score concerning feature of software is 3.6%, needing improvement, with the mean of 2.71. The evaluators gave 100% opinion score in the area whether or not they may use his program in the future. The data were shown in the table11.

**Table11. Number and Percentage of Software Using by Evaluators' Opinion**

| Opinion | Number | Percentage | Mean(SD) |
|---|---|---|---|
| **Conveniences of supplying data** | | | |
| Improved | 0 | 0.0 | 3.36(0.68) |
| Fairly good | 3 | 10.7 | |
| Good | 12 | 42.9 | |
| Very good | 13 | 46.4 | |
| **Fastness of supplying data** | | | |
| Improved | 0 | 0.0 | 3.57(0.50) |
| Fairly good | 0 | 0.0 | |
| Good | 12 | 42.9 | |
| Very good | 16 | 57.1 | |

Copyright by Mahidol University

**Table 11** Continued

| Opinion | Number | Percentage | Mean(SD) |
|---|---|---|---|
| **Difficulty of amending data** | | | |
| Improved | 1 | 3.6 | 3.25(0.65) |
| Fairly good | 0 | 0.0 | |
| Good | 18 | 64.3 | |
| Very good | 9 | 32.1 | |
| **Features of software** | | | |
| Improved | 1 | 3.5 | 2.71(0.81) |
| Fairly good | 11 | 39.3 | |
| Good | 11 | 39.3 | |
| Very good | 5 | 17.9 | |
| **Accuracy of sample size results** | | | |
| Improved | 0 | 0.0 | 3.21(0.69) |
| Fairly good | 4 | 14.3 | |
| Good | 14 | 50.0 | |
| Very good | 10 | 35.7 | |
| **Sample size coverage** | | | |
| Improved | 0 | 0.0 | 3.36(0.62) |
| Fairly good | 2 | 7.1 | |
| Good | 14 | 50.0 | |
| Very good | 12 | 42.9 | |
| **Total satisfaction of software** | | | |
| Improved | 0 | 0.0 | 3.11(0.69) |
| Fairly good | 5 | 17.8 | |
| Good | 15 | 53.6 | |
| Very good | 8 | 28.6 | |
| **If you need to estimate sample size, will you use this program** | | | |
| Yes | 28 | 100.0 | |
| No | 0 | 0.0 | |

### 4.4 The Evaluation of This Program Gives Both Advantage and Disadvantage.

The users have given evaluation from both sides as follows

**Advantages of This Program**

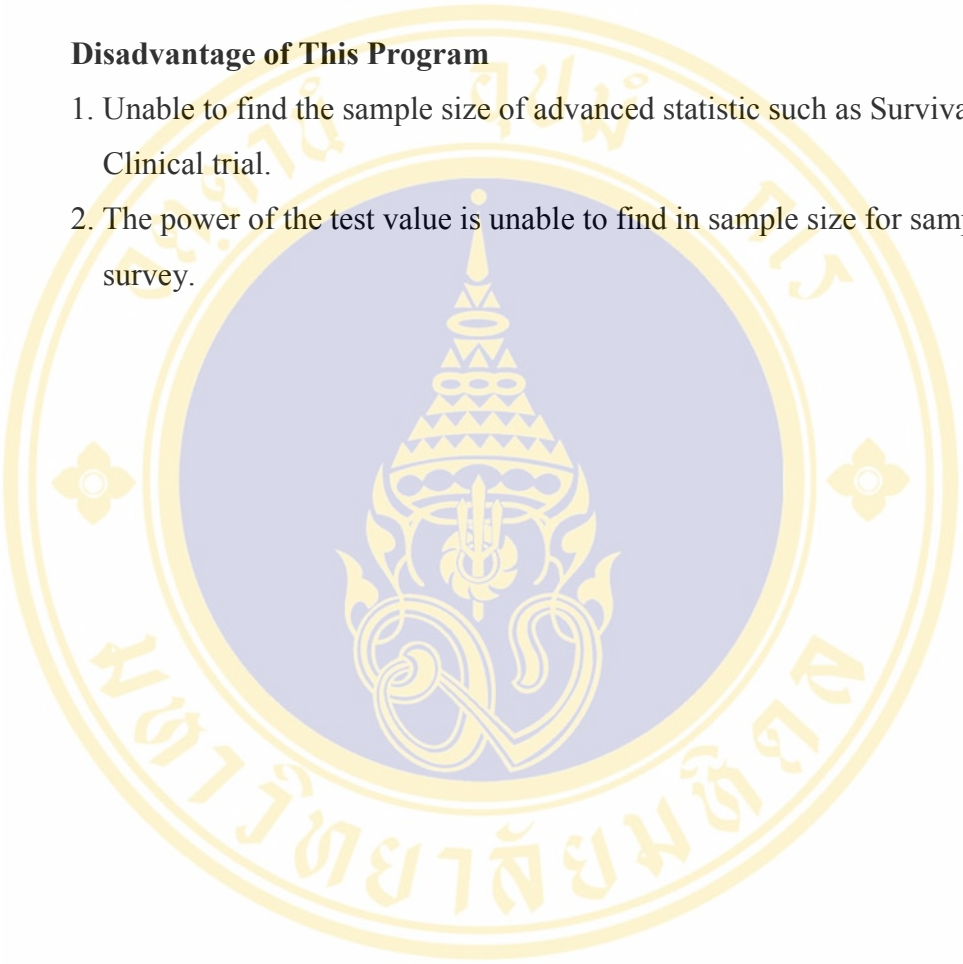
1. Easy to use and calculate
2. It covers sampling designs and parameter.
3. Be able to fix $\alpha$, $\beta$ values in sample size calculation.

Copyright by Mahidol University

4. There is a Menu Help to show steps and formula in using including examples.
5. Be able to show sample sizes without being back to the square one.
6. Be able to save it.

**Disadvantage of This Program**

1. Unable to find the sample size of advanced statistic such as Survival and Clinical trial.
2. The power of the test value is unable to find in sample size for sample survey.



Copyright by Mahidol University

# CHAPTER V
# DISCUSSION

There are two parts in this chapter:

1. The developer gave discussion toward the A-SIZE Program .
2. Comparing A-SIZE program to other Sample size programs which is appropriately able to work well with sample size calculation.

## 1. The Developer Gave Discussion Toward the A-SIZE Program .

This program gives varied statistical values except One-way ANOVA, survival and Clinical Trials. Its only limitation is to find the power of the test on Case-Control study. However, in case of sample size for sample survey, it is able to find estimation of any such parameters as Total, Mean, Proportion and Ratio, dividing into one sample and two sample, using allowable absolute error (d) and relative error (r) in calculation. Other programs are unable to calculate sample size through relative error (r). Moreover, the A-SIZE program can also show the result in Grid Table, which many recent results are still showing and available to be saved and printed out.

## 2. Comparing A-SIZE Program to other Sample Size Programs which is Appropriately able to Work well with Sample Size Calculation

In the part of sample size for sample survey, it is divided into estimation of parameter and hypothesis testing. Estimation of parameter can be divided into one sample and two samples, using allowable absolute error (d) and relative error(r). The A-SIZE program is able to calculate the 4 parameters Total, Mean, Proportion and Ratio to determine the allowable absolute error of sample size, while the relative error is able to find only 3 parameters Total, Mean and Proportion. The SAM program, Power program, Stata program and East 2000 are unable to find any of these parameters. The Epi Info program is available to find estimation of parameter within allowable absolute error. It is able to find one-sample proportion, two-sample

Copyright by Mahidol University

proportions. In the comparison between the A-SIZE program and Epi Info program concerning one-sample proportion, the result of sample size value is the same.

**Stratified Sampling**. None of these programs are able to calculate sample size of Stratified Sampling except the A-SIZE program.

**Hypothesis Testing** is composed of one sample and two sample, having two parameter values: mean and proportion. The Stata and the A-SIZE are perfectly able to calculate sample size while SAM program and Power program can be used to calculate sample size concerning two-sample mean and two-sample proportion; Minitab program can also be used in finding sample size one-sample proportion and two-sample proportion; and last of all, the Epi Info and East 2000 are unable to calculate the hypothesis testing.

**Cluster Sampling.** The A-SIZE program is able to calculate both Mean Estimator and proportion estimator, but the Epi Info can calculate only proportion estimator while others are unable to do at all.

**Case-Control study.** Case-Control study can be calculated by using the SAM program, Epi Info program and the A-SIZE program. But the SAM program and Epi Info grogram can only calculate for the unequal study while the A-SIZE program is able to calculate all three: unmatched study, k controls per case unmatched study and matched study.

**Cohort study**. Cohort study sample size can be determined by using the Epi Info program and A-SIZE program. The Epi Info program can calculate power of the test, but the A-SIZE program can't do it.

**Power of the test.** Power of the test can be calculated by using the Power program, the Epi Info program, Minitab program and A-SIZE program. However, the calculation may not be done completely in every statistic value. The A-SIZE program can calculate only case control study. Epi Info program can be used calculating sample size in Case-Control Study and Cohort study. Minitab program can be used in calculating Power of the test in many ways. There are one sample Z, one sample T, two sample T, one sample Proportion, two sample Proportion, one-way ANOVA, two level factorial designs and Plackett Burman designs. The power program can be used in determine power of the test in two sample mean and two sample proportion.

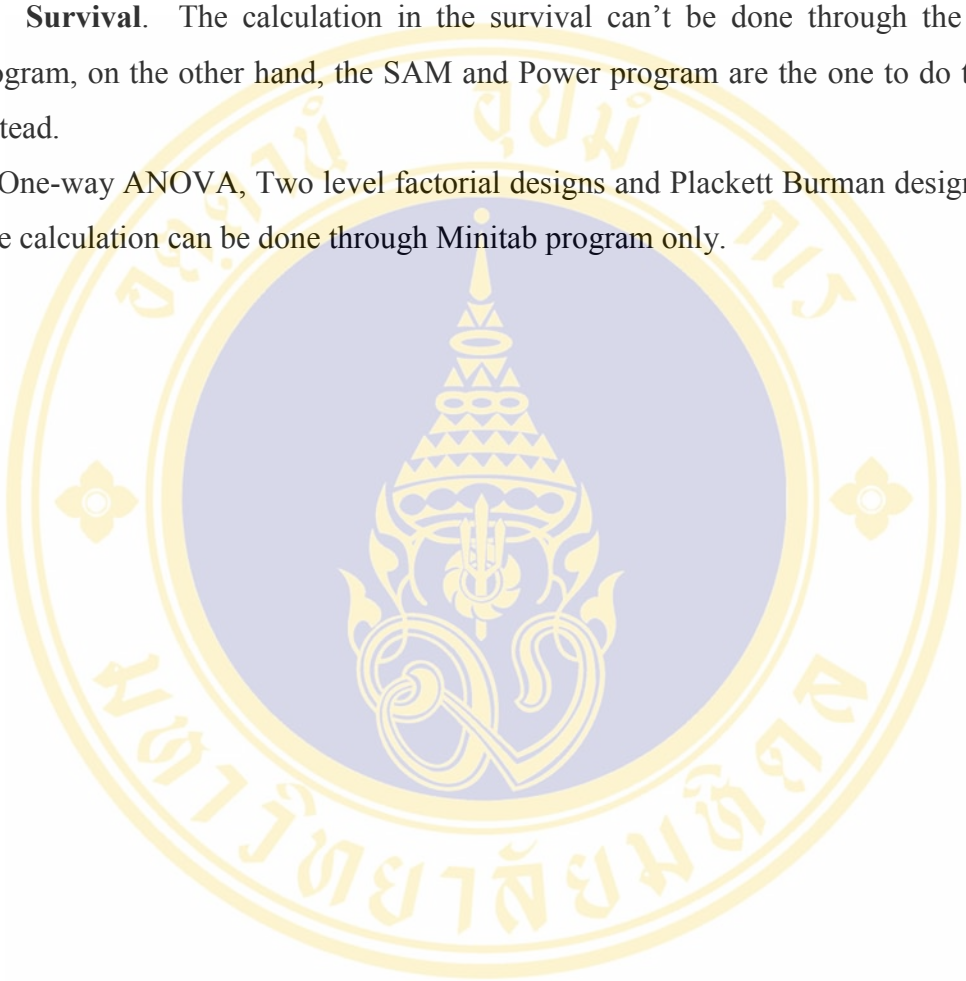
Copyright by Mahidol University

**Clinical trial.** The sample size calculation on Clinical trial can't be done by the A-SIZE program, but the East 2000 is the only program that can handle this job.

**The logistic regression and correlation** can be done through the A-SIZE program while the others can't.

**Survival**. The calculation in the survival can't be done through the A-SIZE program, on the other hand, the SAM and Power program are the one to do the work instead.

In One-way ANOVA, Two level factorial designs and Plackett Burman design sample size calculation can be done through Minitab program only.



Copyright by Mahidol University

# CHAPTER VI
# CONCLUSION AND RECOMMENDATION

## 1. Conclusion

This study emphasizes in the development of sample size determination using the Microsoft Visual Basic program. This Microsoft program is pimple to be used. However, it has some disadvantages concerning arithmetical symbol such as Greek or Latin symbol which can't be displayed. The detail of this program is composed of 4 parts: Sample size for sample survey, Sample size for observational study, Sample size for logistic regression and Sample size for correlation. The sample size for sample survey is composed of sample size for estimating parameter and sample size for hypothesis testing. The sample size for observational study is composed of case-control study and cohort study. The sample size for logistic regression is composed of single logistic regression and multiple logistic regression. The sample size for correlation is composed of test of a single correlation and test of two independent correlations. The using of this program is similar to other programs, having Menu bar displayed prior to moving for any further process. A user can find any sample size by just going to the menu which it will show forms of input data. Next, the user need to input data into all text boxes respectively. Input data needs to be done correctly and completely otherwise the program will warn with the "Error" sign, for instance, proportion value which normally must be in between 0-1 will show Error if inputting data is out of range. After inputting data is done, the user clicks calculate key bottom. The result will be displayed in two spots Text box and Grid table. The Grid table shows varied sample size results, which it is useful as the user can choose a suitable result for the task. However, if the Grid table is not used, it will give only one result or answer. Next, the user can print it out by clicking "Print" or clicking "Save" if a user wants to save data for further uses.

When this program was tested by 28 users dealing with sample size such as researchers/statisticians, lecturers and students, these evaluators gave good and very good score opinion. The score opinion concerns statistic sample size coverage, convenience-ship and speed of using, but in term of feature, it still needs improvement.

Copyright by Mahidol University

To the question whether they will love to use this program or not, all of evaluators give the same answer, “yes, they will”.

**2 Recommendation**

There are 2 recommendations: using this program and suggestions for further program development.

**2.1 Recommendation concerning using this program**

1. The Microcomputer’s ability is to be considered prior to the use of this program, otherwise, it will not give the best result or unable to work at all.
2. The program is made for personal computer use under the Microsoft Windows system.
3. Prior to the use of this program, the user needs to study the program thoroughly as the wrong use can create a wrong outcome.
4. The user needs to be acknowledgeable of how to determine sample size.

**2.2 Recommendation concerning Suggestions for Further Program development**

1. Add more advanced statistical sample size determination such as Survival and Clinical Trial to the perfection of the work.
2. A suggestion for whoever will develop this program in the future concerning good feather is making screen picture more interesting and beautiful.

Copyright by Mahidol University

# REFERENCES

1. Paul S. Levy, Stanley Lemeshow. Sampling of population. 3rd ed. New York: John Wiley & Sons; 1999:13,15,58.
2. W Paul Vogt. Dictionary of statistics and methodology: a nontechnical guide for the social sciences: Sage publications; The United States of America; 1993:83.
3. Steven K Thompson. Sampling. New York : John Wiley & Sons;1992:11.
4. Wilai Kusolvisikul. Sampling theory and applications in public health. Mahidol University; 1998.
5. Lemeshow S, and Lwanga S.K. Adequacy of Sample Size in Health Studies. Chichester: John Wiley & Son; 1990.
6. Cochran,W.G. Sampling Techniques. 3rded. New York: John Wiley & Sons ; 1963:89-90,205.
7. Anusorn Bunsong. Patterns of personal health practice: high school students in Ubonratchathani province [M Sc. Thesis in Biostatistics]. Bangkok: Faculty of Graduate Studies, Mahidol University; 2001.
8. James J. Schlesselman. Case-Control Studies Design, Conduct, Analysis. NewYork: Oxford University Press ; 1982:145-162.
9. Somchai S, Amornrath P. Principles of epidemiology. 1th ed. n.p.; n.d.
10. F.Y. Hsieh, Daniel A. Bloch and Michael D. Larsen. A Simple method of sample size calculation for linear and logistic regression. Statist Med1998; 17:624-5.
11. Chaiyuth Jetiyanuwatr. Factors associated with smoking cessation of public health officials in kanchanaburi province [M.Sc. Thesis in Biostatistics]. Bangkok: Faculty of Graduate Studies, Mahidol University; 2001; 41-4.
12. ธวัชชัย วรพงศธร. หลักการวิจัยทางสาธารณสุขศาสตร์. กรุงเทพมหานคร:สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย; 2540:438-39.

Copyright by Mahidol University

13. Glasziou P. SAM : A Sample size calculation. NHMRC Clinical Trial Centre; 1991.
14. Minitab Inc. Minitab user's guide2. United States of America; 9-2,3.
15. Rodney Hayward. Stata Program. University of Michigan's schools of Medicine &Public Health.
16. Andrew G. Dean, Juan Carlos Zubieta, and Kevin M. Manual Epi info 2000.
17. Lakkana T. Program Epi Info version 6. 1st ed. Nontaburi: Ministry of Public Health; 1995.
18. Christopher Jennison and Bruce Turnbull. East 2000. Group Sequential Methods with Applications to Clinical Trials; 2000
19. Tharin Sittitumchari. Microsoft Visual Basic Version 6. Bangkok. Successmedia; 1989.

Copyright by Mahidol University

# APPENDIX A

## Part of a Software package for determination of the sample size hand book

## ส่วนหนึ่งของคู่มือการใช้โปรแกรมสำเร็จรูปสำหรับการคำนวณหาขนาดตัวอย่าง

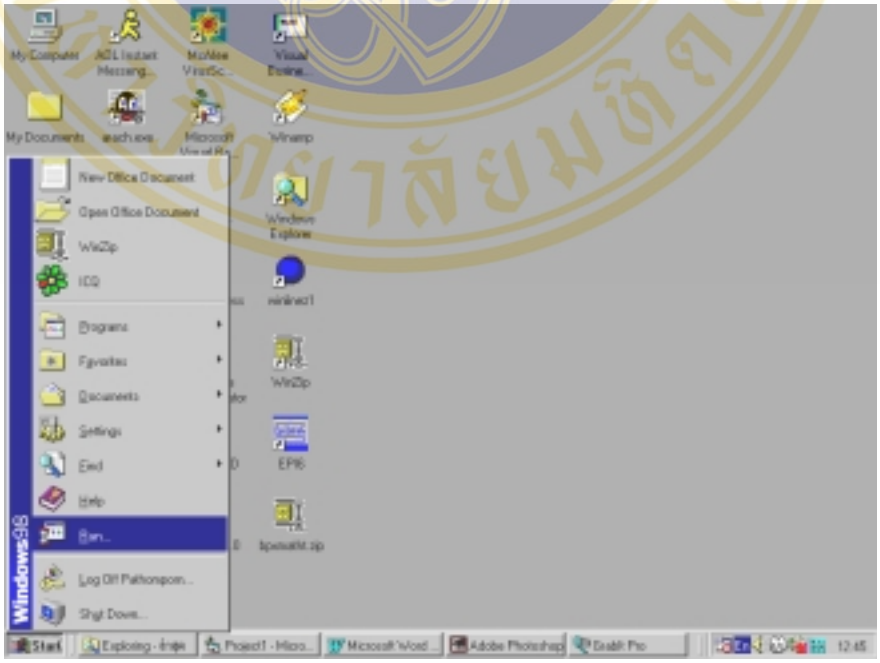
Copyright by Mahidol University

# วิธีการติดตั้งโปรแกรม

โปรแกรมสำเร็จรูปสำหรับการคำนวณหาขนาดตัวอย่างสามารถเรียกใช้งานบนเครื่องไมโครคอมพิวเตอร์ ตั้งแต่รุ่น Pentium ขึ้นไป จอภาพแสดงผลชนิด Super VGA ความละเอียด 800X600จุด ภายใต้การทำงานของโปรแกรมไมโครซอฟ วินโดวส์ เวอร์ชัน 95 ขึ้นไป แต่จะทำงานช้ากว่าไมโครซอฟ วินโดวส์ เวอร์ชัน 98 เนื้อที่ว่างบนฮาร์ดดิสต์อย่างน้อย 16 เมกะไบต์ และ CD ROM

**วิธีการติดตั้งโปรแกรม** มีขั้นตอนดังนี้

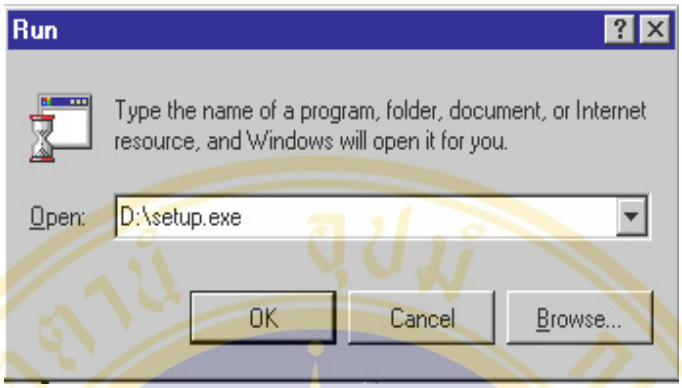
1. เปิดเครื่องคอมพิวเตอร์ จนหน้าจอแสดงการทำงานภายใต้โปรแกรมไมโครซอฟ วินโดวส์ เวอร์ชัน 95 ขึ้นไป ถ้าจะให้ทำงานได้เต็มประสิทธิภาพควรตั้งแต่ ไมโครซอฟ วินโดวส์ เวอร์ชัน 98 ขึ้นไป
2. ใส่แผ่น CD ในไดร์ฟ CD ROM
3. คลิกที่คำสั่ง<RUN> ในเมนู <Start> จอภาพจะปรากฏดังรูปนี้




รูปที่1 แสดงการเลือกเมนู<Start>

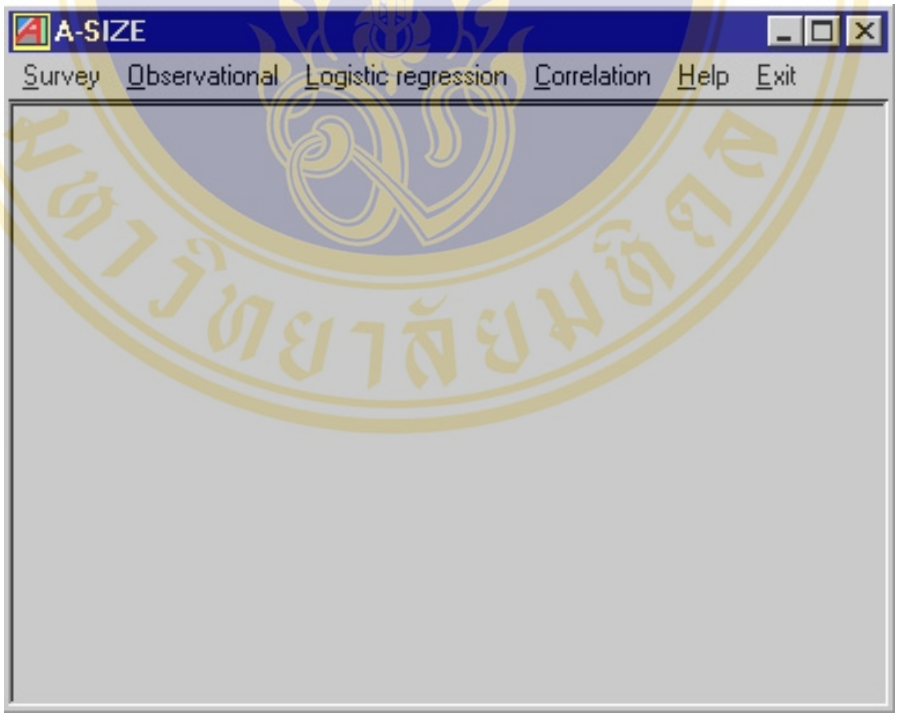
Copyright by Mahidol University

4. ทำการ Browse หา Setup.exe แล้วคลิกที่ปุ่ม<OK> ดังรูป




รูปที่ 2 แสดงข้อความเมื่อเลือกคำสั่ง <RUN>จากเมนู<Start>

5. เมื่อเข้าสู่โปรแกรมแล้วจอภาพจะปรากฏดังนี้




รูปที่ 3 แสดงหน้าจอเริ่มต้นของโปรแกรม

Copyright by Mahidol University

**การใช้เมนูบาร์**

| เมนูบาร์ | ความหมาย |
|---|---|
| **Survey + Estimation** | |
| Simple Random Sampling + One Sample + Total + Allowable absolute error(d) | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Total ในประชากร1กลุ่ม โดยใช้การแจกแจงแบบ Allowable absolute error(d) |
| Simple Random Sampling + One Sample + Total + Relative error(r) | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Total ในประชากร1กลุ่ม โดยใช้การแจกแจงแบบ Relative error(r) |
| Simple Random Sampling + One Sample + Mean + Allowable absolute error(d) | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Mean ในประชากร1กลุ่ม โดยใช้การแจกแจงแบบ Allowable absolute error(d) |
| Simple Random Sampling + One Sample + Mean + Relative error(r) | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Mean ในประชากร1กลุ่ม โดยใช้การแจกแจงแบบ Relative error(r) |
| Simple Random Sampling + One Sample + Proportion + Allowable absolute error(d) | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Proportion ในประชากร1กลุ่ม โดยใช้การแจกแจงแบบ Allowable absolute error(d) |
| Simple Random Sampling + One Sample + Proportion + Relative error(r) | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Proportion ในประชากร1กลุ่ม โดยใช้การแจกแจงแบบ Relative error(r) |
| Simple Random Sampling + One Sample + Ratio + Allowable absolute error(d) | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Ratio ในประชากร1กลุ่ม โดยใช้การแจกแจงแบบ Allowable absolute error(d) |
| Simple Random Sampling + Two Sample + Total | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Total ในประชากร2กลุ่ม |
| Simple Random Sampling + Two Sample + Mean | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Mean ในประชากร2กลุ่ม |
| Simple Random Sampling + Two Sample + Proportion | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Proportionในประชากร2กลุ่ม |
| Simple Random Sampling + Two Sample + Ratio | เข้าสู่การคำนวณหาขนาดตัวอย่างในการประมาณค่าพารามิเตอร์ Ratio ในประชากร2กลุ่ม |

Copyright by Mahidol University

| เมนูบาร์ | ความหมาย |
|---|---|
| **Survey + Estimation** | |
| Stratified Sampling + Total + Proportional allocation + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าTotalโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Proportional allocation โดยใช้ Allowable absolute error(d) |
| Stratified Sampling + Total + Proportional allocation + Relative error(r) | เข้าสู่การคำนวณการประมาณค่าTotalโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Proportional allocation โดยใช้ Relative error(r) |
| Stratified Sampling + Total + Neyman's allocation + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าTotalโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Neyman's allocation โดยใช้ Allowable absolute error(d) |
| Stratified Sampling + Total + Neyman's allocation + Relative error(r) | เข้าสู่การคำนวณการประมาณค่าTotalโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Neyman's allocation โดยใช้ Relative error(r) |
| Stratified Sampling + Total + Optimum allocation for fixed cost + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าTotalโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Optimum allocation for fixed costโดยใช้ Allowable absolute error(d) |
| Stratified Sampling + Total + Optimum allocation for fixed cost + Relative error(r) | เข้าสู่การคำนวณการประมาณค่าTotalโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Optimum allocation for fixed costโดยใช้ Relative error(r) |
| Stratified Sampling + Mean + Proportional allocation + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าMeanโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Proportional allocation โดยใช้ Allowable absolute error(d) |
| Stratified Sampling + Mean + Proportional allocation + Relative error(r) | เข้าสู่การคำนวณการประมาณค่าMeanโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบProportional allocation โดยใช้ Relative error(r) |
| Stratified Sampling + Mean + Neyman's allocation + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าMeanโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Neyman's allocation โดยใช้ Allowable absolute error(d) |

Copyright by Mahidol University

| เมนูบาร์ | ความหมาย |
|---|---|
| **Survey + Estimation** | |
| Stratified Sampling + Mean + Neyman's allocation + Relative error(r) | เข้าสู่การคำนวณการประมาณค่าMeanโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Neyman's allocation โดยใช้ Relative error(r) |
| Stratified Sampling + Mean + Optimum allocation for fixed cost + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าMeanโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Optimum allocation for fixed costโดยใช้ Allowable absolute error(d) |
| Stratified Sampling + Mean + Optimum allocation for fixed cost +Relative error(r) | เข้าสู่การคำนวณการประมาณค่าMeanโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Optimum allocation for fixed costโดยใช้ Relative error(r) |
| Stratified Sampling + Proportion + Proportional allocation + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าProportionโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Proportional allocation โดยใช้ Allowable absolute error(d) |
| Stratified Sampling + Proportion + Proportional allocation + Relative error(r) | เข้าสู่การคำนวณการประมาณค่าProportionโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Proportional allocation โดยใช้ Relative error(r) |
| Stratified Sampling + Proportion + Neyman's allocation + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าProportionโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Neyman's allocation โดยใช้ Allowable absolute error(d) |
| Stratified Sampling + Proportion + Neyman's allocation +Relative error(r) | เข้าสู่การคำนวณการประมาณค่าProportionโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Neyman's allocation โดยใช้ Relative error(r) |
| Stratified Sampling + Proportion + Optimum allocation for fixed cost + Allowable absolute error(d) | เข้าสู่การคำนวณการประมาณค่าProportionโดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Optimum allocation for fixed costโดยใช้ Allowable absolute error(d) |

Copyright by Mahidol University

| เมนูบาร์ | ความหมาย |
|---|---|
| **Survey + Estimation** | |
| Stratified Sampling + Proportion + Optimum allocation for fixed cost + Relative error(r) | เข้าสู่การคำนวณการประมาณค่า Proportion โดยใช้วิธีจัดสรรแจกจ่ายตัวอย่างแบบ Optimum allocation for fixed cost โดยใช้ Relative error(r) |
| Cluster Sampling + Mean Estimator | การคำนวณหาขนาดตัวอย่างของ Cluster sampling ในเรื่อง Mean Estimator |
| Cluster Sampling + Proportion Estimator | การคำนวณหาขนาดตัวอย่างของ Cluster sampling ในเรื่อง Proportion Estimator |
| **Survey + Hypothesis testing** | |
| Mean + One Sample + One-sided | การคำนวณหาขนาดตัวอย่างในการทดสอบสมมติฐานของ Mean ในประชากร 1 กลุ่ม 1 ทาง |
| Mean + One Sample + Two-sided | การคำนวณหาขนาดตัวอย่างในการทดสอบสมมติฐานของ Mean ในประชากร 1 กลุ่ม 2 ทาง |
| Mean + Two Sample + One-sided | การคำนวณหาขนาดตัวอย่างในการทดสอบสมมติฐานของ Mean ในประชากร 2 กลุ่ม 1 ทาง |
| Mean + Two Sample + Two-sided | การคำนวณหาขนาดตัวอย่างในการทดสอบสมมติฐานของ Mean ในประชากร 2 กลุ่ม 2 ทาง |
| Proportion + One Sample + One-sided | การคำนวณหาขนาดตัวอย่างในการทดสอบสมมติฐานของ Proportion ในประชากร1 กลุ่ม 1 ทาง |
| Proportion + One Sample + Two-sided | การคำนวณหาขนาดตัวอย่างในการทดสอบสมมติฐานของ Proportion ในประชากร1 กลุ่ม 2 ทาง |
| Proportion + Two Sample + One-sided | การคำนวณหาขนาดตัวอย่างในการทดสอบสมมติฐานของ Proportion ในประชากร 2 กลุ่ม 1 ทาง |

Copyright by Mahidol University

| เมนูบาร์ | ความหมาย |
|---|---|
| **Survey + Hypothesis testing** | |
| Proportion + Two Sample +Two-sided | การคำนวณหาขนาดตัวอย่างในการทดสอบสมมติฐานของ Proportion ในประชากร 2 กลุ่ม 2 ทาง |
| **Observational study** | |
| Case-Control Study + Unmatched Studies + Sample size | การคำนวณหาขนาดตัวอย่างใน Unmatched Studies |
| Case-Control Study + Unmatched Studies + Power | การคำนวณหา Powerใน Unmatched Studies |
| Case-Control Study + k Controls per Case Unmatched + Sample size | การคำนวณหาขนาดตัวอย่างใน k Controls per Case Unmatched |
| Case-Control Study + k Controls per Case Unmatched + Power | การคำนวณหา Powerใน k Controls per Case Unmatched |
| Case-Control Study + Matched Studies+Sample size | การคำนวณหาขนาดตัวอย่างใน Matched Studies |
| Case-Control Study + Matched Studies+Power | การคำนวณหา Powerใน Matched Studies |
| Cohort Study | การคำนวณหาขนาดตัวอย่างใน Cohort Study |
| **Multivariate** | |
| Simple logistic regression | การคำนวณหาขนาดตัวอย่างใน Simple logistic regression |
| Multiple logistic regression | การคำนวณหาขนาดตัวอย่างใน Multiple logistic regression |
| **Correlation** | |
| Test of a Single Correlation | การคำนวณหาขนาดตัวอย่างใน Test of a Single Correlation |
| Test of Two Independent Correlations | การคำนวณหาขนาดตัวอย่างใน Test of Two Independent Correlation |
| Help | ช่วยสอนวิธีการใช้โปรแกรม |
| Exit | ออกจากโปรแกรม |

Copyright by Mahidol University

**ตารางอธิบายปุ่มต่างๆที่ใช้ในโปรแกรม**

| ปุ่ม | ความหมาย |
|---|---|
| Calculate | คำนวณหาขนาดตัวอย่าง |
| Exit | ออกจากโปรแกรมการคำนวณ |
| Reset | ลบข้อมูลหมดทั้งหน้าจอ |
| Remove | ลบข้อมูลทีละชุด |
| Add | เพิ่มข้อมูลทีละชุด |
| Print | พิมพ์ค่าที่เราคำนวณได้ |

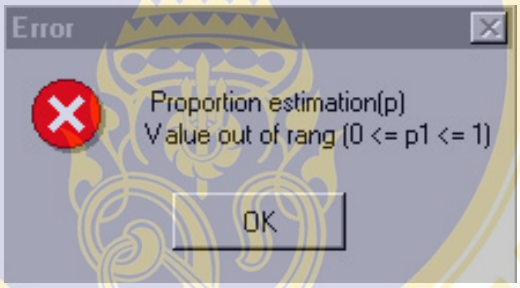
Copyright by Mahidol University

**การแสดงผลของความผิดพลาดในการป้อนข้อมูล**

1. การบอกความผิดพลาดในการป้อนข้อมูล แสดงว่าป้อนข้อมูลไม่ครบถ้วน ผู้ใช้โปรแกรมจะต้องป้อนข้อมูลให้ครบถ้วนก่อนถึงจะคำนวณค่าได้

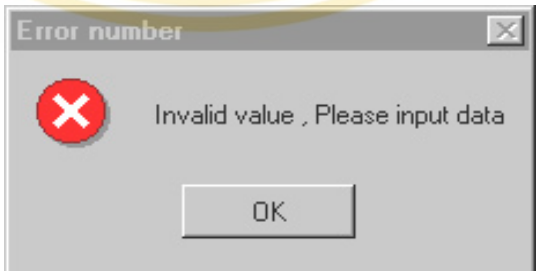
รูป 4 แสดงการป้อนข้อมูลไม่ครบถ้วน

2. การบอกความผิดพลาดในการป้อนค่า p1 คือค่า proportion ต้องอยู่ระหว่าง $0 \leq p1 \leq 1$




รูป 5 แสดงการป้อนข้อมูลผิดพลาด

3.การบอกความผิดในการป้อนค่าเข้าไปในตาราง Grid ถ้าไม่ได้ป้อนค่าให้สมบูรณ์จะแสดงขึ้นมา




รูป 6 แสดงการป้อนข้อมูลผิดพลาดในตาราง

Copyright by Mahidol University

**โจทย์ตัวอย่างในแต่ละเรื่อง**

**1. การคำนวณหาขนาดตัวอย่างในเรื่องการประมาณค่าเฉลี่ย(Sample size for estimating a population mean)**

**ตัวอย่าง.** ขนาดประชากรนักเรียน 1,000 คน ค่าเฉลี่ยคะแนนความรู้เท่ากับ 5 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ควรสุ่มตัวอย่างมาเท่าใด เมื่อกำหนดความคลาดเคลื่อนสัมบูรณ์เท่ากับ 0.1 สัมประสิทธิ์ความแปรปรวนเท่ากับ 20% ความคลาดเคลื่อนสัมพัทธ์เท่ากับ 2% ความเชื่อมั่นเท่ากับ 95%

**วิธีทำ** โจทย์กำหนด $N = 1{,}000,\ \alpha = 0.05,\ s^2 = 1,\ d = 0.1$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \qquad n_0 = \frac{Z^2_{\alpha/2} s^2}{d^2} \qquad n_0 = \frac{1.96^2 1^2}{0.1^2} = 384.16$$

$$n = \frac{384.16}{1 + \frac{384.16}{1{,}000}} = 277.5 = 278$$

**แหล่งที่มา** : วิไล กุศลวิศิษฏ์กุล. ทฤษฎีการสุ่มตัวอย่างและการประยุกต์ในงานสาธารณสุข. มหาวิทยาลัยมหิดล; 2541:203

ในโจทย์ตัวอย่างนี้ทำการคำนวณหาขนาดตัวอย่างในเรื่อง Survey ต้องการประมาณค่าเฉลี่ยโดยการสุ่มตัวอย่างแบบง่ายของประชากร 1 กลุ่ม โดยกำหนดความคลาดเคลื่อนสัมบูรณ์ (Allowable absolute error (d)) และความคลาดเคลื่อนสัมพัทธ์

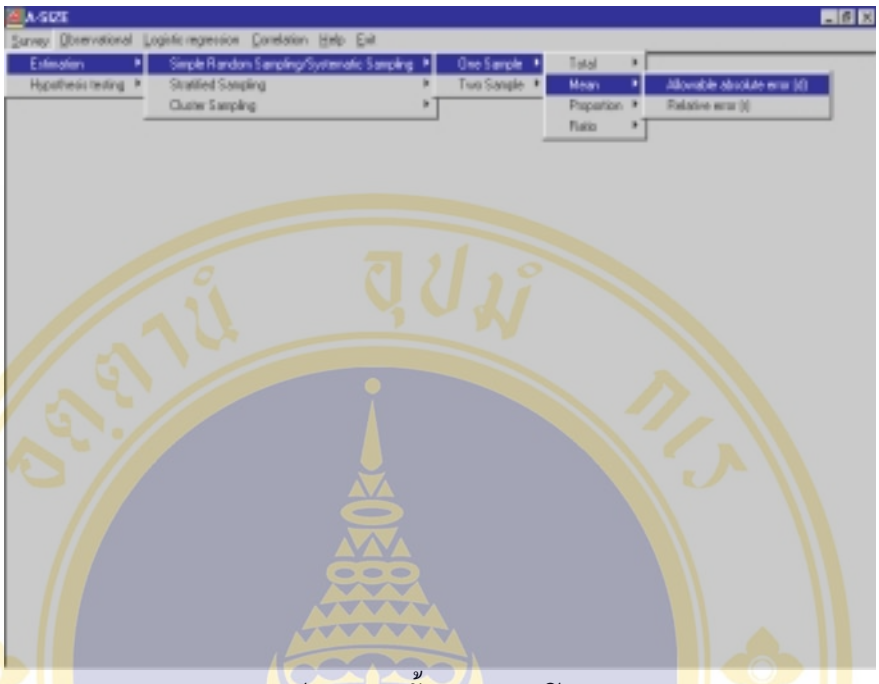
**จากโจทย์แสดงขั้นตอนการทำงานดังนี้**

1. ให้ทำการเลือก Survey ที่ปรากฏอยู่บน Menu bar

2. จาก Menu Survey ทำการเลือกเมนูต่อได้ โดยการใช้เมาส์คลิกต่อไปเรื่อยๆ จะปรากฏ Form สำหรับการคำนวณขึ้นมา ในขั้นตอนนี้ผู้ใช้จะเป็นคนเลือกเองว่าจะใช้การคำนวณโดยวิธีไหน ใช้เมาส์คลิกได้ดังนี้ Estimation →Simple random sampling/Systematic sampling→

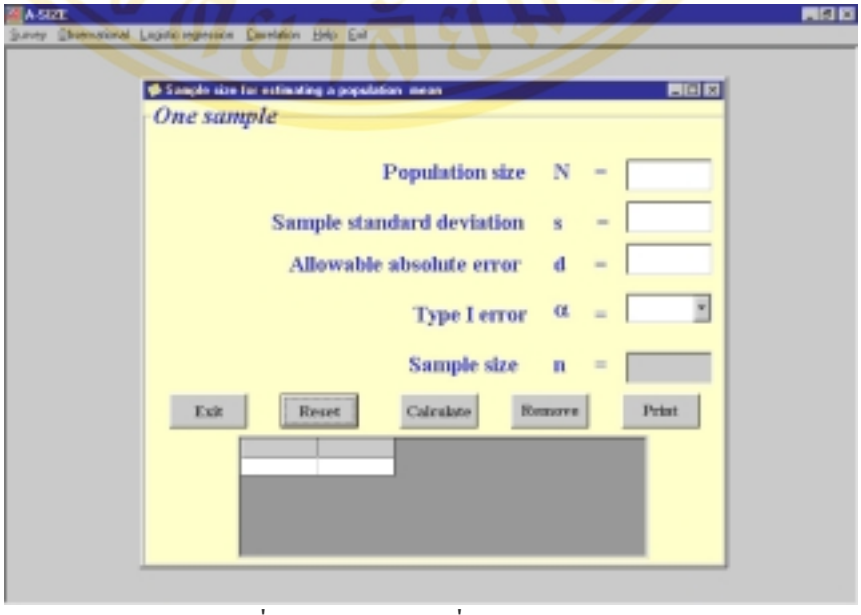
Copyright by Mahidol University

Mean→ One Sample → Allowable absolute error (d)




รูป 7 แสดงขั้นตอนการเปิดเมนู

3. ป้อนค่าที่จะคำนวณไปยัง Text box แต่ละ Text box ถ้าทำการป้อนเสร็จแล้วให้กดปุ่ม Enter หรือ ใช้ Mouse Click เลือกในแต่ละช่องก็ได้ ในขั้นตอนนี้จะต้องทำการป้อนข้อมูลให้ถูกต้อง มิฉะนั้น โปรแกรมจะแสดงค่า Error ออกมา



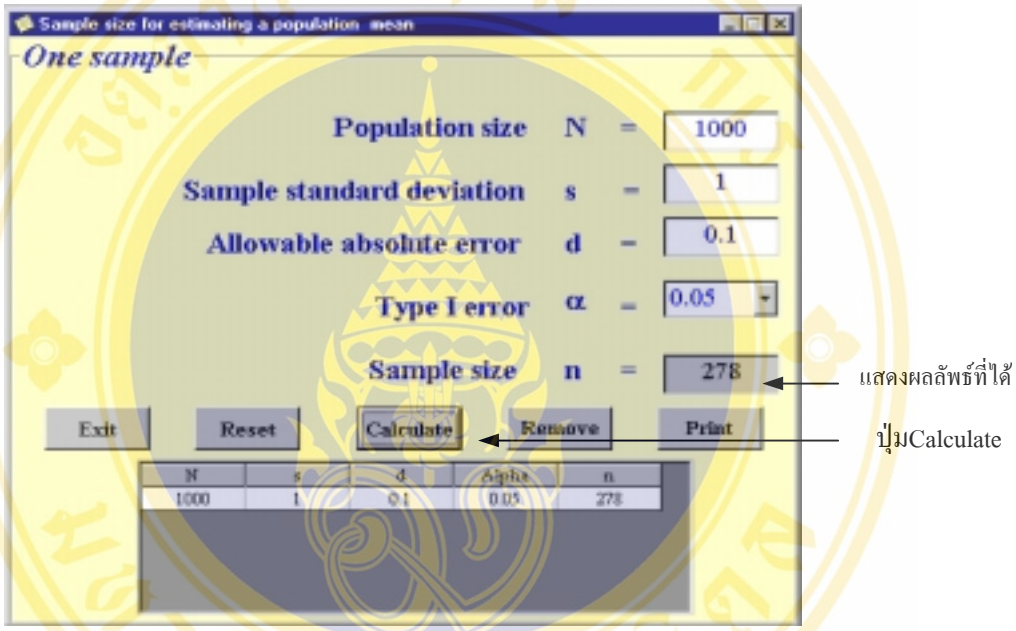

รูปที่ 8 แสดงหน้าจอที่จะทำการป้อนข้อมูล

Copyright by Mahidol University

4. ในช่อง Drop down list สามารถทำได้ 2 วิธีคือ ใช้ Mouse คลิกเลือกหรือใช้ Keyboard แทนได้ โดยการเลือกปุ่มลูกศร ↑ , ↓

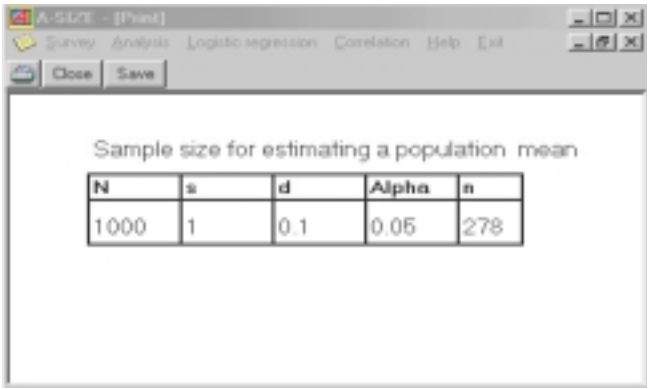
5. เมื่อทำการป้อนข้อมูลครบถ้วนแล้วให้กดปุ่ม **Calculate** ได้เลย

6. ผลลัพธ์จะแสดงออกมาในช่อง **Sample size (n)**




รูปที่ 9 แสดงผลลัพธ์ที่คำนวณได้

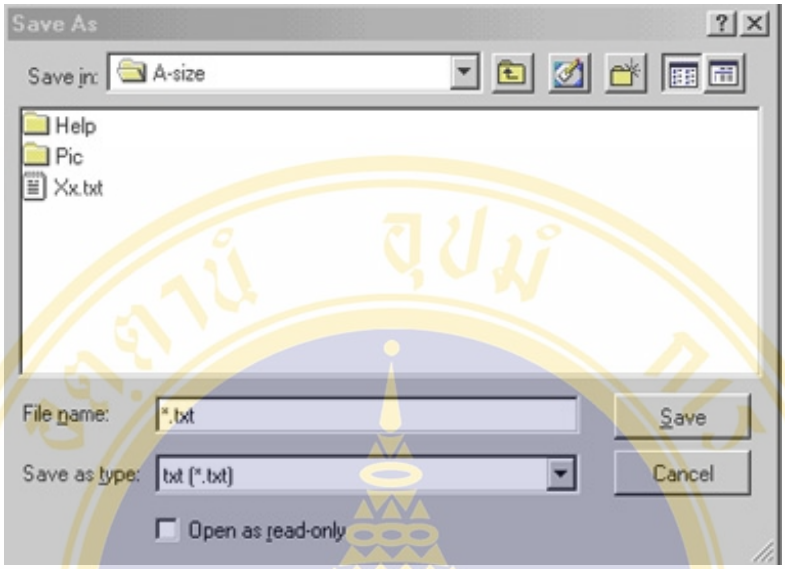
7. ถ้าต้องการ Print ข้อมูลออกทางเครื่องพิมพ์ให้คลิกไปที่ปุ่ม Print และคลิกที่รูปPrint อีกโปรแกรมจะทำการ Print ข้อมูลออกมา




รูปที่ 10 แสดงหน้าจอ Print

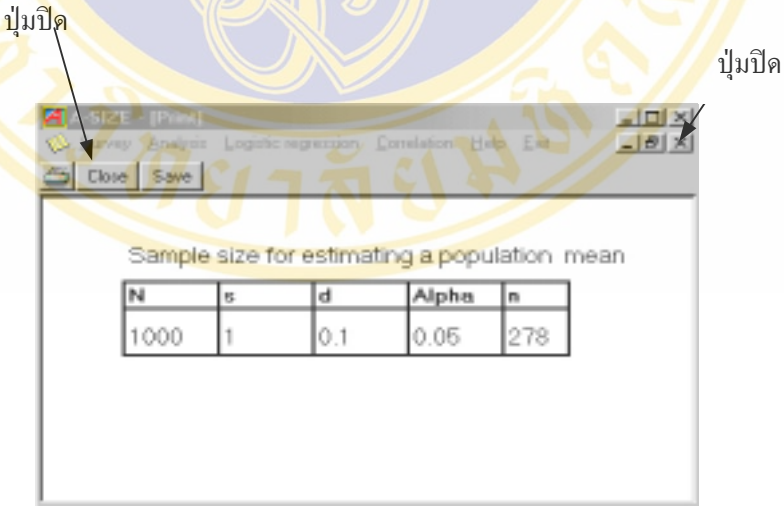
Copyright by Mahidol University

8. ถ้าจะทำการSave ให้คลิกMouse ไปที่ปุ่ม Saveได้เลย จะทำการ Save fileเป็น .txt




รูปที่ 11 แสดงหน้าจอ Save

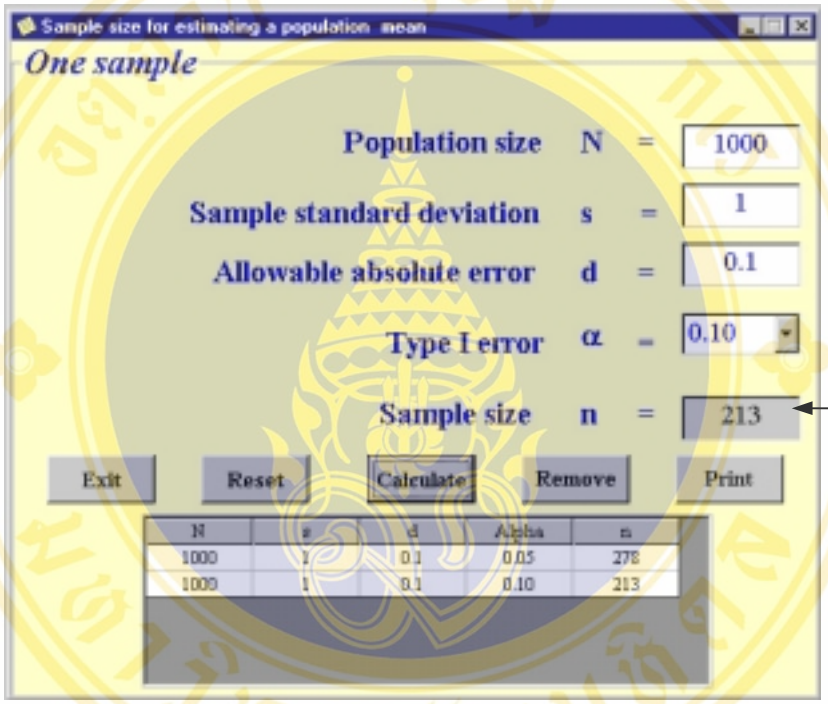
9. ถ้าจะทำการปิดหน้าจอ Print ทำได้ 2 วิธีคือ 1.กดที่ปุ่ม Close 2. คลิกMouse ไปที่เครื่องหมาย X ของ Windows แสดงตามลูกศรได้ดังนี้




รูปที่ 12 แสดงการปิดหน้าจอ Print

Copyright by Mahidol University

10. ถ้าจะทำการคำนวณหาขนาดตัวอย่างต่อโดยการเปลี่ยนค่าเพียงบางค่าให้คลิกเปลี่ยนค่าได้เลย โดยที่ไม่ต้องเข้าสู่ Form การคำนวณใหม่ ในที่นี้ลองเปลี่ยนค่า $\alpha$ จาก 0.05 เป็น 0.1 และกดปุ่ม Calculate ได้เลยขนาดตัวอย่างที่ได้จะแสดงในรูปตาราง Grid ในที่นี้ได้ขนาดตัวอย่าง เท่ากับ 213 ดังแสดงในรูป




รูปที่ 13 แสดงการคำนวณซ้ำโดยการเปลี่ยนค่า $\alpha$

ผู้ใช้สามารถคำนวณได้หลายๆครั้งโดยที่ไม่ต้องเริ่ม Form ใหม่ มีข้อดีตรงที่ผู้ใช้สามารถเลือกขนาดตัวอย่างที่ต้องการได้

ในโจทย์เดิมทำการหาขนาดตัวอย่างโดยการกำหนดความคลาดเคลื่อนสัมพัทธ์ (Relative error) วิธีการทำเหมือนขั้นตอนแรกแต่เปลี่ยนจากการกำหนดความคลาดเคลื่อนสัมบูรณ์Allowable absolute error (d) มาเป็น การกำหนดความคลาดเคลื่อนสัมพัทธ์ Relative error (r)

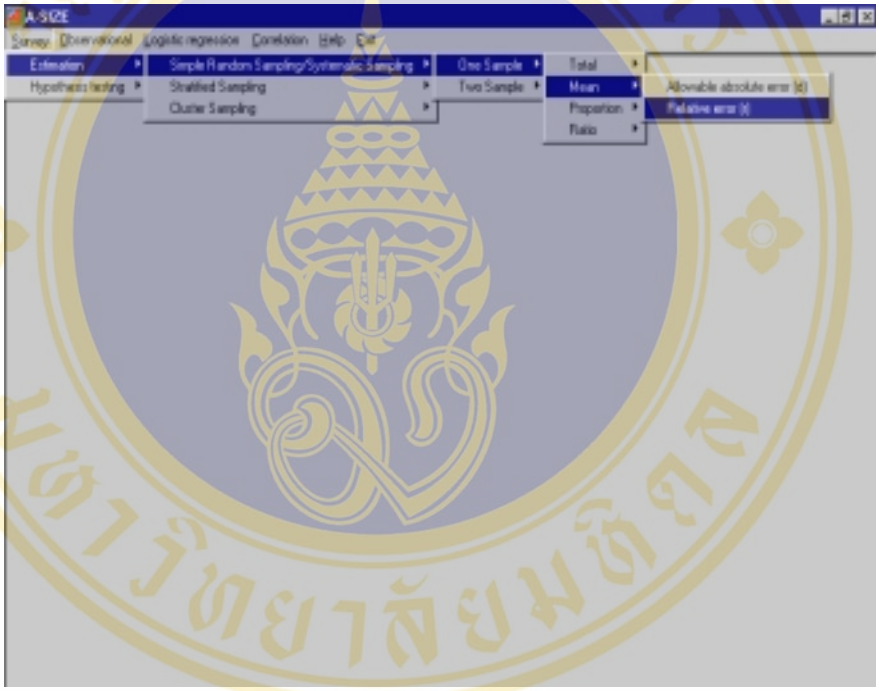
Copyright by Mahidol University

แสดงขั้นตอนการทำได้ดังนี้

1. ให้ทำการเลือก Menu Survey ที่ปรากฏอยู่บน Menu bar

2. จาก Menu Survey ทำการเลือกเมนูต่อได้ โดยการใช้เมาส์คลิกต่อไปเรื่อยๆ จะปรากฏForm สำหรับการคำนวณ ใช้เมาส์คลิกได้ดังนี้

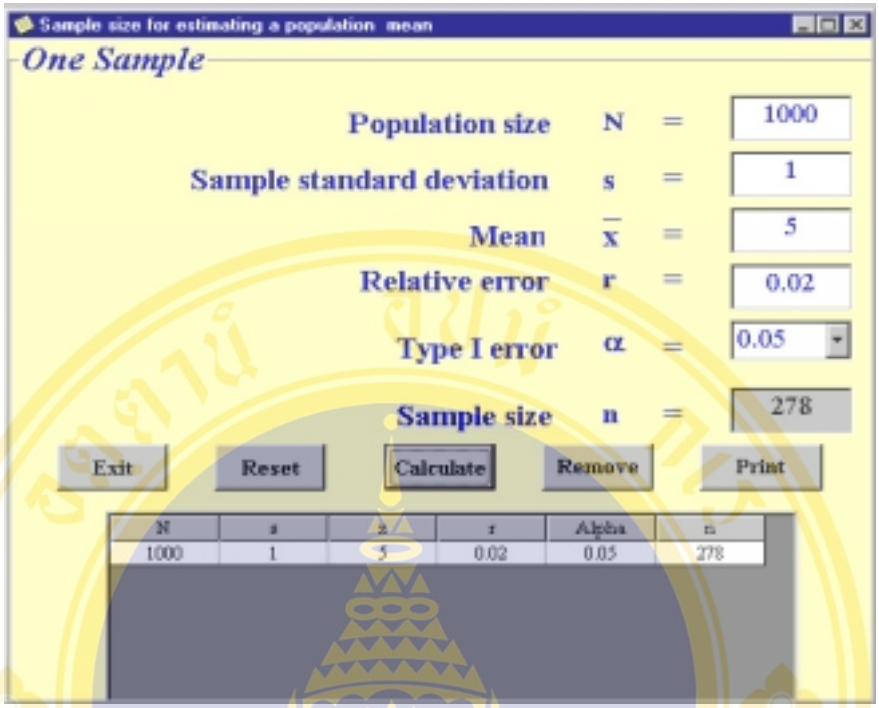
Estimation →Simple random sampling/Systematic sampling→One Sample → Mean→ Relative error (r)




รูปที่ 14 แสดงขั้นตอนการเปิดเมนู Surveyเมื่อกำหนดค่า Relative error

พอทำการคลิกจะปรากฏ Form สำหรับการคำนวณขึ้นมา ให้เราป้อนค่าที่จะคำนวณลงไปได้เลย พอป้อนค่าเสร็จเรียบร้อยแล้วให้ทำการกดปุ่ม Calculate ผลลัพธ์ที่ได้จะแสดงออกมาในช่องของ Sample size ดังแสดงในรูป

Copyright by Mahidol University




รูปที่ 15 แสดงผลลัพธ์ของการคำนวณโดยกำหนดค่า Relative error

**แสดงการคำนวณหาขนาดตัวอย่างในแผนการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified sampling)**

**ตัวอย่าง** จากการสำรวจเบื้องต้นเกี่ยวกับค่าใช้จ่ายด้านการรักษาพยาบาลในครัวเรือนขนาดต่างๆ พบว่า

| ขนาดครัวเรือน | จำนวนครัวเรือน | ส่วนเบี่ยงเบนมาตรฐาน($S_h$) |
|---|---|---|
| ขนาดเล็ก | 600 | 200 |
| ขนาดกลาง | 300 | 300 |
| ขนาดใหญ่ | 100 | 500 |
| **รวม** | **1000** | |

จงประมาณค่าขนาดตัวอย่าง เมื่อกำหนดให้ความคลาดเคลื่อนสมบูรณ์ (Allowable absolute error)ไม่เกิน 30 บาท ที่ระดับนัยสำคัญ 0.01 ในที่นี้จะแสดงวิธีทำสำหรับ Neyman's allocation. เท่านั้น

Copyright by Mahidol University

$$n = \frac{Z^2_{\frac{\alpha}{2}}\left(\sum_{h=1}^{L} N_h S_h\right)^2}{N^2 d^2 + Z^2_{\frac{\alpha}{2}}\sum_{h=1}^{L} N_h S_h{}^2}$$

$$= \frac{2.576^2[(600*200)+(300*300)+(100*500)]^2}{1000^2*30^2+2.576^2[(600*200^2)+(300*300^2)+(100*500^2)]}$$

$$= 320.43$$

ดังนั้น ต้องใช้ขนาดตัวอย่าง 321 คน

สามารถจัดสรรตัวอย่างของแต่ละชั้นภูมิได้ ดังนี้

$$n_1 = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} n \qquad = (120000/260000) * 321 = 148$$

$$n_2 = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} n \qquad = (90000/260000) * 321 = 111$$

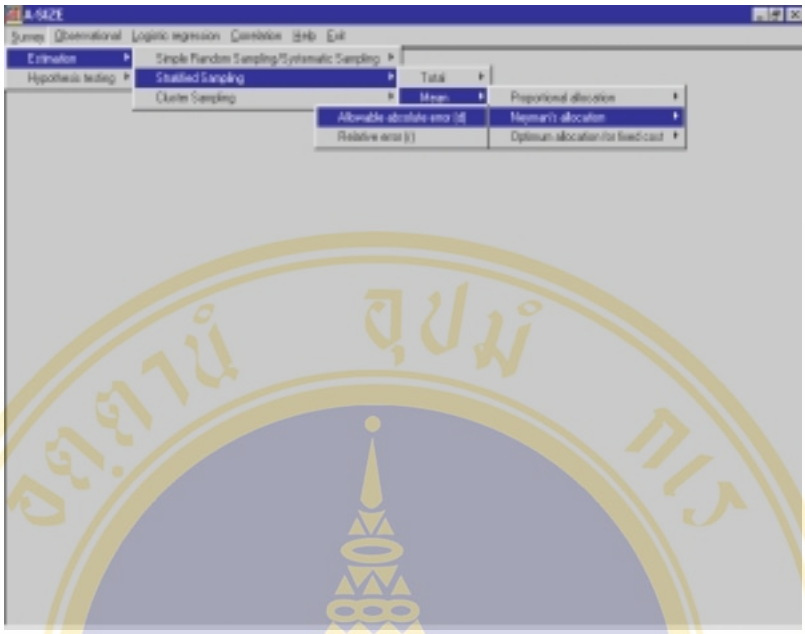
$$n_3 = \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h} n \qquad = (50000/260000) * 321 = 62$$

แสดงขั้นตอนการคำนวณหาขนาดตัวอย่าง ดังนี้

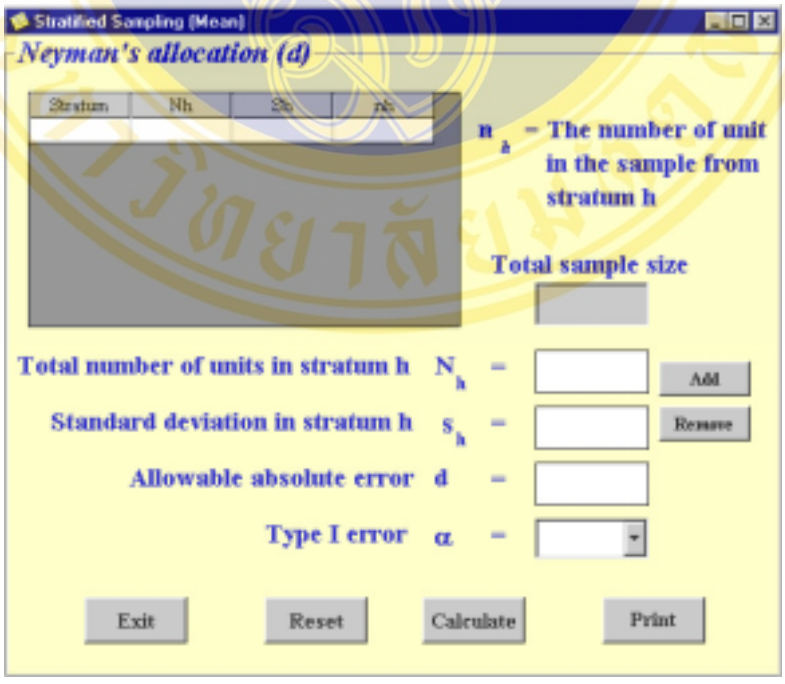
1. ให้ทำการเลือกMenu Survey ที่จะปรากฏอยู่บน Menu bar

2. จาก Menu Survey ทำการเลือกเมนูต่อได้ โดยการใช้เมาส์คลิกต่อไปเรื่อยๆ จะปรากฏ Form ที่ทำการคำนวณ ในขั้นตอนนี้ผู้ใช้จะเป็นคนเลือกเองว่าจะใช้การคำนวณโดยวิธีไหน ใช้เมาส์คลิกได้ ดังนี้ Estimation→ Stratified Sampling → Mean → Neyman's allocation → Allowable absolute error(d)

Copyright by Mahidol University



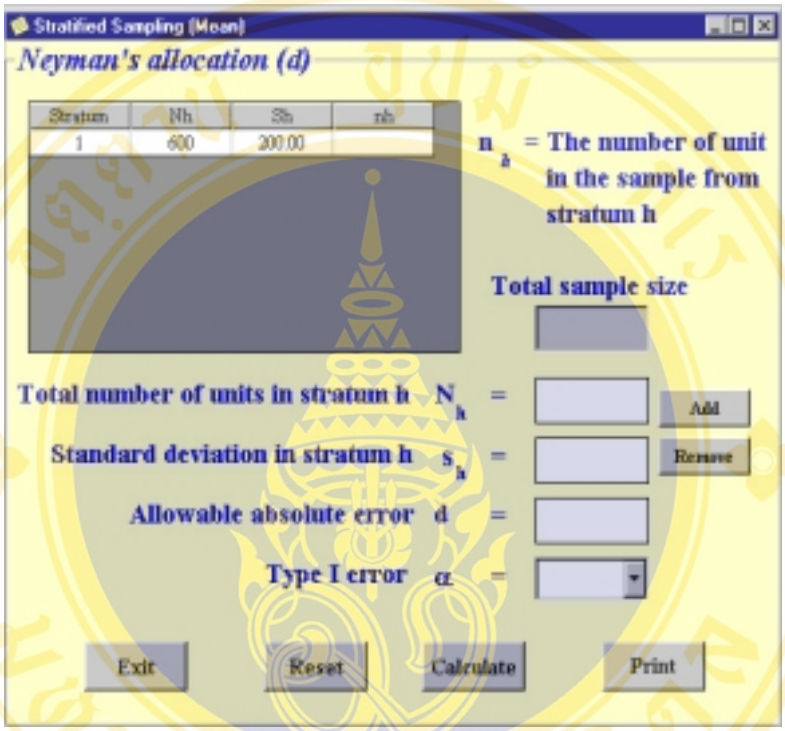

รูปที่ 16 แสดงเมนูการคำนวณในเรื่อง Stratified Sampling




รูปที่ 17 แสดง Form การป้อนข้อมูลในการคำนวณ Stratified Sampling

Copyright by Mahidol University

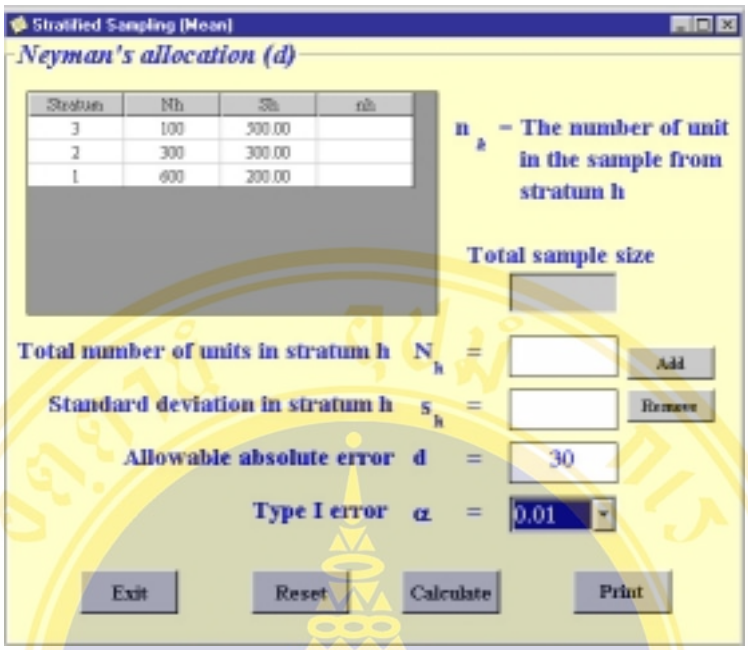
3. เมื่อปรากฏ Formที่จะทำการป้อนข้อมูลมาแล้วให้ทำการป้อนข้อมูลเข้าไปทีละ 1 ชุด เสร็จแล้ว กดปุ่ม Add เพื่อจะให้ค่าแสดงในตาราง Grid ในที่นี้ให้ใส่ค่า $N_h$ และ $S_h$ ก่อนและกดปุ่มAdd




รูปที่ 18 แสดงการป้อนข้อมูล 1 ชุด

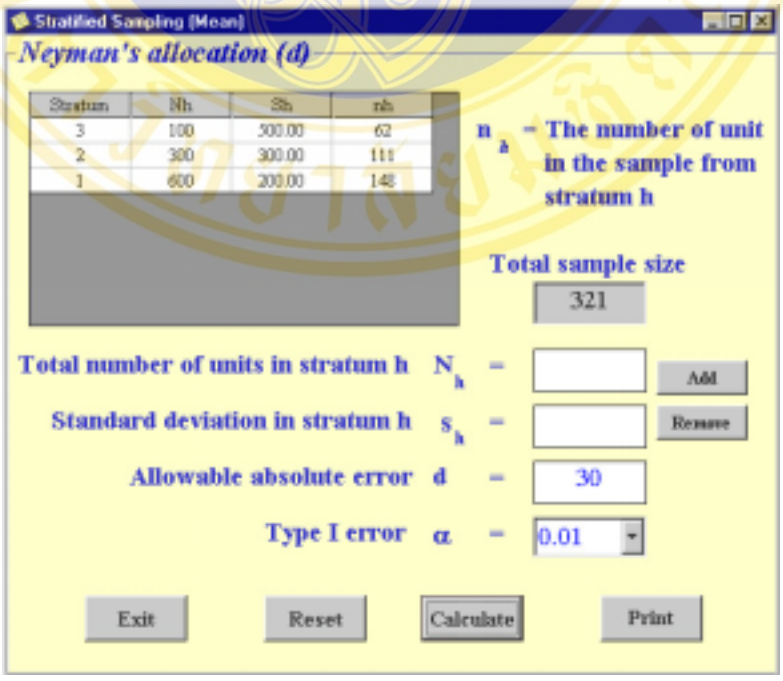
ต่อจากนั้นทำการกรอกข้อมูลจนครบ

Copyright by Mahidol University




รูปที่ 19 แสดงการป้อนข้อมูลจนครบ

4. เมื่อกรอกข้อมูลครบถ้วนแล้วนั้น กดปุ่ม Calculate จะได้ค่าตามรูป




รูปที่ 20 แสดงผลลัพธ์ของการคำนวณหาขนาดตัวอย่างใน Stratified Sampling

Copyright by Mahidol University

Total sample size คือ ค่าของขนาดตัวอย่างทั้งหมด มีค่าเท่ากับ 321

$n_h$ คือ ค่าของขนาดตัวอย่างแต่ละชั้นภูมิ ชั้นภูมิที่ 1 มีค่าเท่ากับ 148

ชั้นภูมิที่ 2 มีค่าเท่ากับ 111

ชั้นภูมิที่ 3 มีค่าเท่ากับ 62

**แสดงการคำนวณหาขนาดตัวอย่างในแผนการสุ่มตัวอย่างแบบกลุ่ม (Cluster sampling)**

**ตัวอย่าง** ประชากรกลุ่มเป้าหมายของนักเรียนมัธยมศึกษาตอนปลาย ภาคการศึกษาที่2 ปี2543 มีนักเรียนระดับมัธยมศึกษาตอนปลายในประชากรเป้าหมายทั้งหมด 29,664 คน โดยแบ่งได้เป็น 11 สหวิทยาเขต จากการสำรวจนำร่องเก็บข้อมูลจากแบบสอบถามที่สร้างไว้โดยให้นักเรียนตอบแบบสอบถามเองมีนักเรียนที่ตอบข้อมูลที่สามารถคำนวณเป็นคะแนนพฤติกรรมได้ จำนวน 272 คน ค่าเฉลี่ยของคะแนนพฤติกรรมสุขภาพได้เท่ากับ 100.83 และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 8.91 คำนวณหาขนาดตัวอย่างที่ได้จากวิธีการสุ่มตัวอย่างแบบง่ายได้ 707 คำนวณหาขนาดของ Design effect

$$\text{Design effect} = \frac{\hat{V}ar(\bar{x})_{cluster}}{\hat{V}ar(\bar{x})_{srs}}$$

คำนวณหาค่าความแปรปรวนของตัวแปรที่ได้จากการสุ่มตัวอย่างแบบกลุ่ม ได้ดังนี้

$$\hat{V}ar(\bar{x})_{cluster} = \frac{\sum_{i=1}^{n} \bar{x}_i - \bar{G})^2}{n(n-1)}$$

n = จำนวนสหวิทยเขต (cluster) 11 แห่ง

$\bar{x}$ = ค่าเฉลี่ยของคะแนนพฤติกรรมสุขภาพในสหวิทยาเขตกลุ่มที่ i (i=1,2,3..,11)

$\bar{G}$ = ค่าเฉลี่ยของคะแนนพฤติกรรมสุขภาพของนักเรียน 272 คน

Copyright by Mahidol University

| สหวิทยาเขต | $\bar{x}$ |
|---|---|
| 1. | 95.58 |
| 2. | 99.26 |
| 3. | 101.54 |
| 4. | 101.26 |
| 5. | 103.55 |
| 6. | 101.9 |
| 7. | 99.6 |
| 8. | 102.84 |
| 9. | 101.33 |
| 10. | 101.05 |
| 11. | 97.77 |

(Grand Mean 100.83)

$$\hat{V}ar(\bar{x})_{cluster} = \frac{\sum_{i=1}^{n} \bar{x}_i - \bar{G})^2}{n(n-1)}$$

$$\hat{V}ar(\bar{x})_{cluster} = \frac{(95.58-100.83)^2 + (99.26-100.83)^2 + .... + (97.77-100.83)^2}{11(11-1)}$$

$$= 0.413$$

การคำนวณหาค่าความแปรปรวนที่ได้จากการสุ่มตัวอย่างแบบง่าย

$$\hat{V}ar(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

n = จำนวนนักเรียนตัวอย่าง 272 คน

$x_i$ = คะแนนพฤติกรรมสุขภาพนักเรียนในนักเรียนคนที่ i(i=1,2,…,272)

$\bar{x}$ = ค่าเฉลี่ยของคะแนนพฤติกรรมสุขภาพในนักเรียนตัวอย่าง

แทนค่าได้

$$\hat{V}ar(x) = \frac{\sum_{i=1}^{n} (x_i - 100.83)^2}{272-1} = 79.38$$

Copyright by Mahidol University

$$\hat{V}ar(\bar{x})_{srs} = \frac{s^2}{n}\left[\frac{N-n}{N}\right]$$

โดยที่ N = ขนาดประชากร 29,664 คน

n = จำนวนนักเรียนตัวอย่าง 272 คน

$$\hat{V}ar(\bar{x})_{srs} = \frac{79.38}{272}\left[\frac{29,664-272}{29,664}\right] = 0.289$$

คำนวณหาขนาดตัวอย่าง Design effect ได้ดังนี้

$$\text{Design effect} = \frac{\hat{V}ar(\bar{x})_{cluster}}{\hat{V}ar(\bar{x})_{srs}}$$

$$= \frac{0.413}{0.289} = 1.427$$

$$n_{cluster} = \text{design effect} * n_{srs}$$

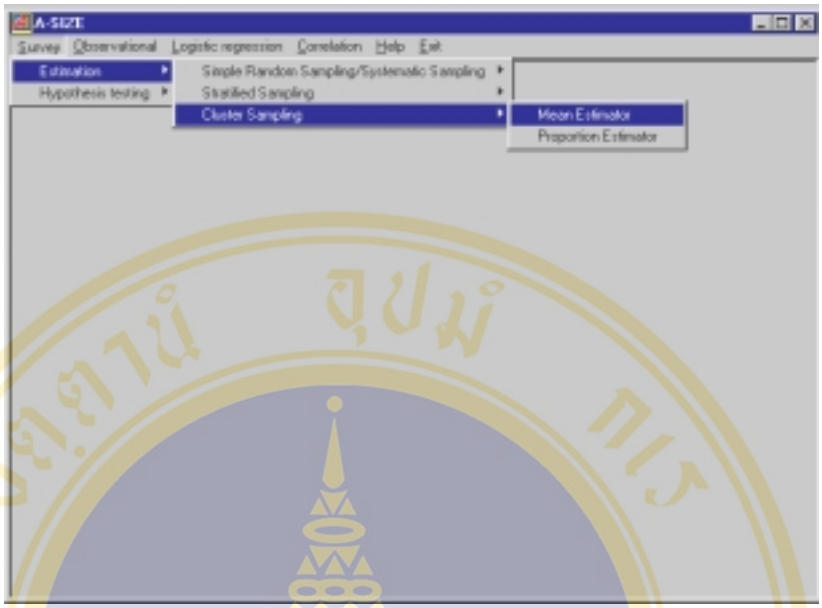
$$= 1.427*707$$

$$= 1,009$$

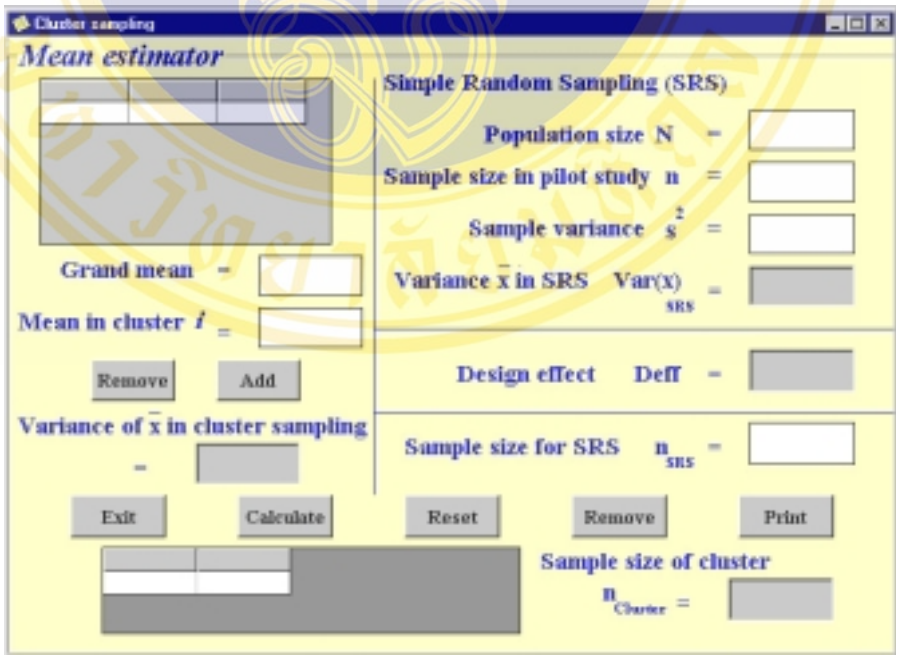
แสดงขั้นตอนการคำนวณหาขนาดตัวอย่าง ดังนี้

1. ให้ทำการเลือก Menu Survey ที่จะปรากฏอยู่บน Menu bar
2. จาก Menu Survey ทำการเลือกเมนูต่อได้ โดยการใช้เมาส์คลิกต่อไปเรื่อยๆ จะปรากฏFormที่ทำการคำนวณหาขนาดตัวอย่าง ใช้เมาส์คลิกได้ดังนี้ Estimation →Cluster Sampling→Mean Estimator

Copyright by Mahidol University



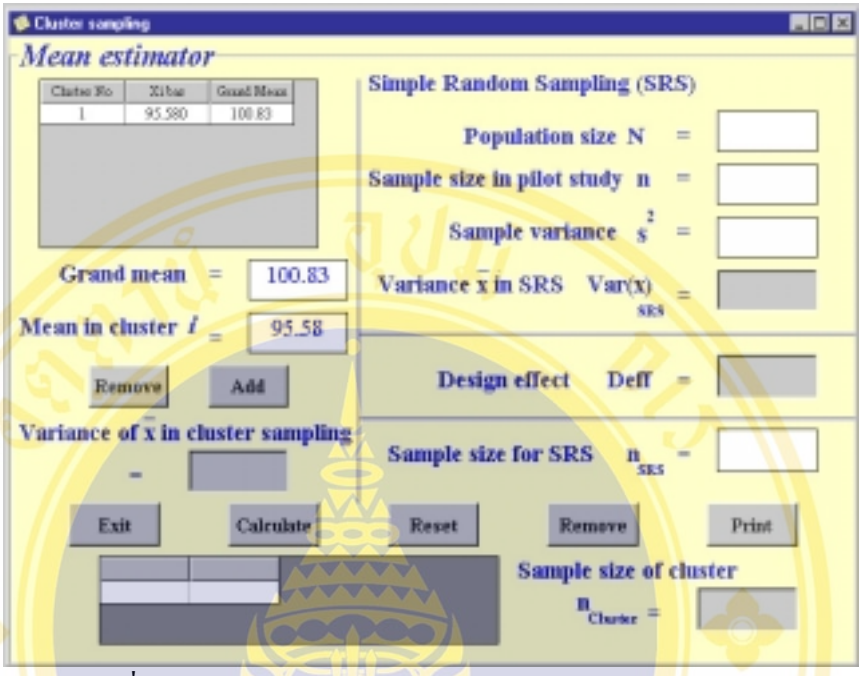

รูปที่ 21 แสดงเมนูการคำนวณในเรื่อง Cluster sampling




รูปที่ 22 แสดง Form การป้อนข้อมูลใน Cluster sampling

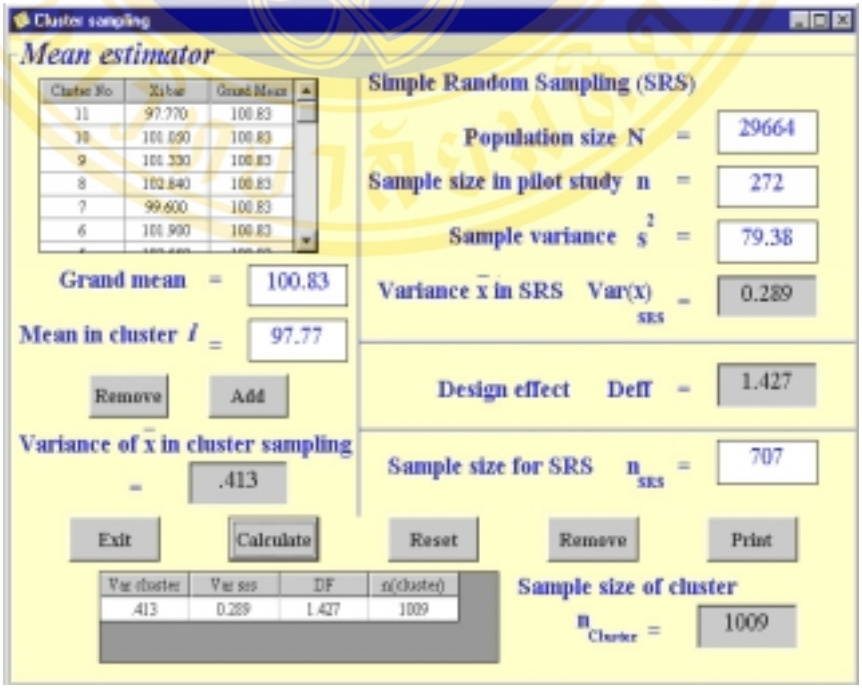
Copyright by Mahidol University

3. ต่อจากนั้นทำการป้อนค่าที่จะนำไปคำนวณ โดยป้อนข้อมูลทีละชุด เสร็จแล้วกดปุ่มAdd




รูปที่ 23 แสดงการป้อนข้อมูลทีละชุดใน Cluster sampling

4. เมื่อกรอกข้อมูลครบถ้วนแล้วนั้น กดปุ่มCalculate จะได้ค่าตามรูปตามรูปนี้



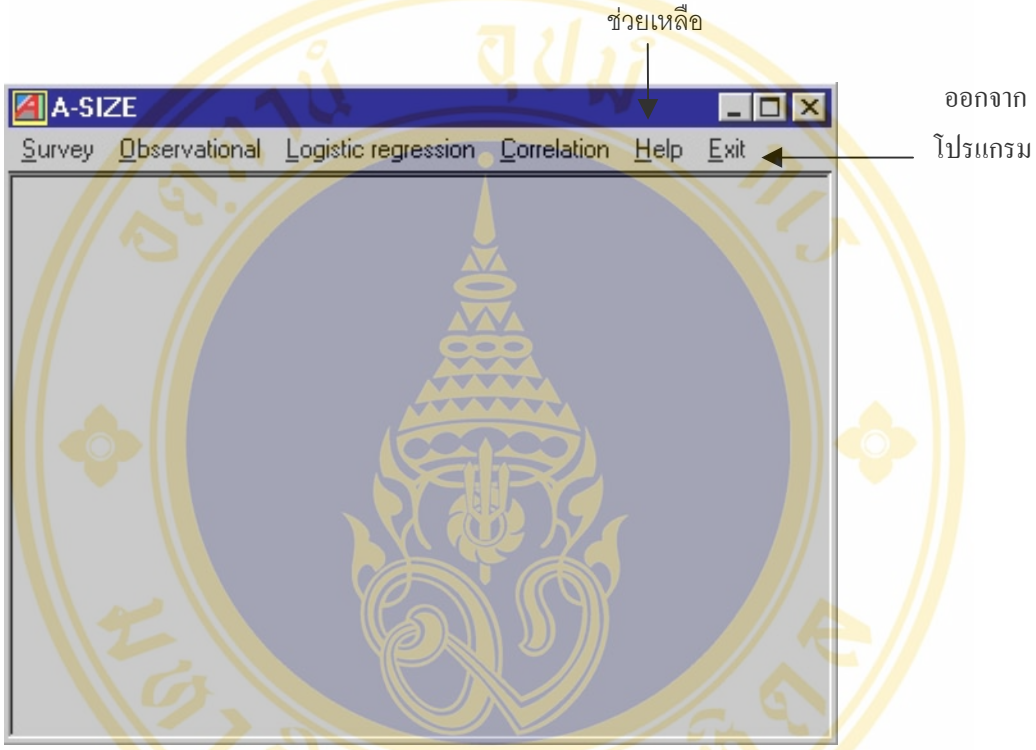

รูปที่ 24 แสดงการคำนวณในเรื่อง Cluster sampling

Copyright by Mahidol University

ขนาดตัวอย่างที่คำนวณได้มีค่า 1009

**การขอความช่วยเหลือ** ทำได้ดังนี้ คือกดปุ่ม Help ที่ Menu bar

**การออกจากโปรแกรม** ทำได้ ดังนี้ คือกดปุ่ม Exit ที่ Menu bar




รูปที่ 25 แสดงMenu Help และ Exit

Copyright by Mahidol University

# APPENDIX B
# EVALUATION FORM
# แบบประเมินผลซอฟต์แวร์

Copyright by Mahidol University

**แบบประเมินผลซอฟต์แวร์"การคำนวณหาขนาดตัวอย่าง"**
**โดยผู้ที่เกี่ยวข้องกับการคำนวณหาขนาดตัวอย่าง**

---

**ส่วนที่ 1 ข้อมูลทั่วไปเกี่ยวกับผู้ตอบแบบสอบถาม (ทำเครื่องหมาย ✓ หน้าข้อความที่ต้องการเลือก)**

1. เพศ

❑ หญิง ❑ ชาย

2.ปัจจุบันท่าน อายุ……ปี

3.ประสบการณ์ในการทำงานที่เกี่ยวกับการคำนวณหาขนาดตัวอย่าง

❑ มี ❑ ไม่มี

4. ท่านมีความรู้เกี่ยวกับการหาขนาดตัวอย่างมาก่อนหรือไม่

❑ มี ❑ ไม่มี

5. ท่านเคยใช้ซอฟต์แวร์ที่เกี่ยวกับการหาขนาดตัวอย่างหรือไม่

❑ เคยระบุชื่อซอฟต์แวร์…… ❑ ไม่เคย

6.ลักษณะงานที่ทำอยู่ในปัจจุบัน

❑ นักวิจัย/นักสถิติ ❑ อาจารย์

❑ นักศึกษา ❑ อื่นๆ(ระบุ……)

---

**ส่วนที่ 2 ความคิดเห็นเกี่ยวกับการใช้ซอฟต์แวร์ "การคำนวณหาขนาดตัวอย่าง"**
**(ทำเครื่องหมาย ✓ ตรงช่องที่ต้องการเลือก) หลังจากที่ได้ทดลองใช้ซอฟต์แวร์"การคำนวณหาขนาดตัวอย่าง"แล้ว**

| ข้อคิดเห็น | ควรปรับปรุง (1) | พอใช้ (2) | ดี (3) | ดีมาก (4) |
|---|---|---|---|---|
| 1. ความสะดวกในการป้อนข้อมูล | | | | |
| 2. ความรวดเร็วในการป้อนข้อมูล | | | | |
| 3. ความยากง่ายในการแก้ไขข้อมูล | | | | |
| 4.ความสวยงามและความน่าใช้ของซอฟต์แวร์นี้ | | | | |
| 5.ความถูกต้องของขนาดตัวอย่างที่ได้ | | | | |
| 6.ความครอบคลุมในเนื้อหาการหาขนาดตัวอย่าง | | | | |
| 7.ความพึงพอใจในภาพรวมของซอฟต์แวร์ | | | | |

Copyright by Mahidol University

**ส่วนที่ 3 ความคิดเห็นอื่นๆเกี่ยวกับซอฟต์แวร์"การคำนวณหาขนาดตัวอย่าง"**

ด้านจุดเด่น 1) . ……

2). ……

3). ……..……..…….

4). ……..……..

5). ……..……..

ด้านจุดด้อย 1) . ……

2). ……

3). ……..……

4). ……..……..

5). ……..……..

ข้อควรพัฒนาและข้อควรปรับปรุง

1) . ……

2). ……

3). ……..……

4). ……..……...

5). ……..……..

ถ้าท่านต้องหาขนาดตัวอย่างท่านจะใช้โปรแกรมนี้หรือไม่

❑ ใช้ ❑ ไม่ใช้เพราะ……..

♣♣♣♣♣♣♣♣♣♣♣

**ขอขอบคุณทุกท่านที่ช่วยกรุณาตอบแบบสอบถาม**

Copyright by Mahidol University

# BIOGRAPHY

| | |
|---|---|
| **NAME** | Miss Pathomporn Thitatan |
| **DATE OF BIRTH** | May 7, 1976 |
| **PLACE OF BIRTH** | Bangkok, Thailand |
| **INSTITUTIONS ATTENDED** | King Mongkut's Institute of Technology North Bangkok, 1994-1997, Degree Bechelor of Science(Applied Statistics)<br>Mahidol University, 1998-2003, Degree Master of Science(Biostatistics) |
| **POSITION & OFFICE** | 15 June – 30 September 2000: Data Preparation Branch Data Processing Operations Division<br>1 October 2000 – Present : Working Quality Control and Evaluation Branch Statistical Techniques Division at the National Statistical Office. |

Copyright by Mahidol University