สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

โปรแกรมคำนวณค่าสถิติเบื้องต้นตามการสุ่มตัวอย่าง

นายดนุรุจ วินมูน

นายยงยุทธ รอดภัย

นางสาววรารัตน์ เอื้ออำนวยชัย

นายวิสิทธิ์ ธงธนานุรักษ์

วิทย.
ว 124 ว
2536
เลขหมู่.....
เลขทะเบียน.....
วัน,เดือน,ปี.....
6126525x

โครงงานพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตร์บัณฑิต

ภาควิชา สถิติประยุกต์

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2536

Computer Software for Estimation of Parameters in Sampling Methods

Mr. Dnuruth Winmoon

Mr. Yongyuth Rodpai

Miss Vararut Auamnuaychai

Mr. Visit Tongtananurug

A Special Project Submitted in Partial Fulfillment of the Requirement for the Degree of Bachelor of Science

Department of Applied Statistics

Faculty of Science

King Mongkut's Institute of Technology Ludkrabang

1993

หัวข้อปัญหาพิเศษ โปรแกรมคำนวณค่าสถิติเบื้องต้นตามการสุ่มตัวอย่าง

โดย นายคนุรุจ วินมูน

นายยงยุทธ รอดภัย

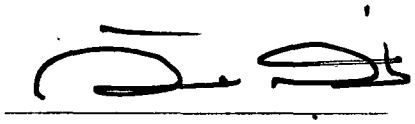
นางสาววราภรณ์ เอื้ออำนวยชัย

นายวิสิทธิ์ ธงธนานุรักษ์

ภาควิชา สถิติประยุกต์

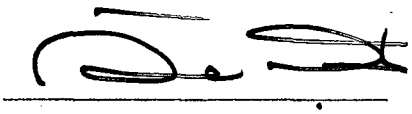
อาจารย์ที่ปรึกษา อาจารย์น้อมจิต กิตติวัชดิพาณิชย์

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์สถาบันพระจอมเกล้าเจ้าคุณทหารลาดกระบัง อนุมัติให้นับโครงงานพิเศษฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตร์บัณฑิต



(ผ.ศ.วีรศักดิ์ สุรพัฒน์) หัวหน้าภาค

(อ.น้อมจิต กิตติวัชดิพาณิชย์) ประธานกรรมการ



(ผ.ศ.วีรศักดิ์ สุรพัฒน์) กรรมการ

(อ.บุญสิทธิ์ วรจันทร์) กรรมการ

ลิขสิทธิ์ของภาควิชา สถิติประยุกต์ คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

หัวข้อปัญหาพิเศษ โปรแกรมคำนวณค่าสถิติเบื้องต้นตามการสุ่มตัวอย่าง

นักศึกษา นายคนุรุจ วินมูน
นายยงยุทธ รอดภัย
นางสาววรารัตน์ เอื้ออำนวยชัย
นายวิสิทธิ์ ธงธนานุรักษ์

อาจารย์ที่ปรึกษา อาจารย์น้อมจิต กิตติโชติพานิชย์

ภาควิชา สถิติประยุกต์

ปีการศึกษา 2536

บทคัดย่อ

การศึกษาครั้งนี้ได้จัดทำโปรแกรมคำนวณค่าสถิติตามวิธีการในแต่ละแผนแบบการสุ่มตัวอย่าง โดยใช้ภาษาซีและทำเป็นแบบเมนูเลือก ซึ่งสามารถหาแผนแบบการสุ่มตัวอย่างที่เหมาะสมกับลักษณะของประชากร หาขนาดตัวอย่างที่เหมาะสมเพื่อให้มีความคลาดเคลื่อนของค่าประมาณไม่เกินที่กำหนดไว้ พร้อมทั้งคำนวณและประมาณค่าสถิติเบื้องต้นของแต่ละแผนแบบการสุ่มตัวอย่างได้อย่างแม่นยำรวดเร็ว และมีการอธิบายข้อสงสัยต่างๆ ในส่วนของคำแนะนำ โดยการแสดงผลทั้งหมดเป็นภาษาไทย ซึ่งโปรแกรมนี้ใช้เนื้อที่ความจุ 900 KB และใช้กับเครื่องPC จอภาพ VGA มีหน่วยความจำไม่น้อยกว่า 640 KB

Special Project Title Computer Software for Estimation of Parameters in Sampling Methods

Name Mr. Dnuruth Winmoon
Mr. Yongyuth Rodpai
Miss Vararut Auamnuaychai
Mr. Visit Tongtananurug

Special Project Advisor Miss Nomchit Kittichotipanit

Department Applied Statistics

Acadamic Year 1993

Abstract

This computer software is developed to estimate the parameter of the population for various sampling methods using C programming language and can choose the menu for finding sampling methods suiting the population. This software can find sample size providing least error of estimated parameters, and also calculate and estimate parameters in each sampling methods precisely and quickly. There are description in help sector for any unknowns. All outputs of this software are Thai language. This program consumes 900 KB and specify to use with computer PC, console VGA, memory more than 640 KB.

กิตติกรรมประกาศ

รายงานปัญหาพิเศษฉบับนี้ที่สามารถสำเร็จลุล่วงไปได้ด้วยดีนั้น ทางคณะผู้จัดทำต้องขอขอบคุณบุคคลต่อไปนี้ ที่ช่วยเหลือในเรื่องต่าง ๆ ตลอดมา

ผ.ค.วีรศักดิ์ สุรพัฒน์ อาจารย์น้อมจิต กิตติโชติพาณิชย์ อาจารย์บุญญสิทธิ์ วรจันทร์ ซึ่งให้คำปรึกษา และแนะนำ

เจ้าหน้าที่บรรณารักษ์ห้องสมุดของมหาวิทยาลัยเกษตรศาสตร์ และของคณะวิทยาศาสตร์ลาดกระบังที่ช่วยแนะนำและค้นหาหนังสืออ้างอิงต่าง ๆ

นาย กีรติ พิฑูรพงศ์ ช่วยในเรื่องของเทคนิคการเก็บหน้าจอแต่ละหน้าจอของโปรแกรมไว้เพื่อพิมพ์ลงกระดาษ สำหรับใช้ในการจัดทำคู่มือการใช้งานของโปรแกรม

พี่ ๆ เจ้าหน้าที่ห้องภาควิชาสถิติประยุกต์ ทุกท่านที่ช่วยเหลือในการทำเรื่องขอห้องและติดต่ออาจารย์ในเรื่องต่าง ๆ

และสุดท้ายนี้ขอขอบคุณคณาจารย์ภาควิชาสถิติประยุกต์ทุกท่าน ที่ประสิทธิประสาทวิชาความรู้ ให้มีความสามารถในการที่จะทำปัญหาพิเศษฉบับนี้จนสำเร็จลุล่วงไปได้ด้วยดี

คณะผู้จัดทำ

สารบัญ

	หน้า
บทคัดย่อปัญหาพิเศษภาษาไทย	ก
บทคัดย่อปัญหาพิเศษภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
บทที่ 1 บทนำ	1
บทที่ 2 ทฤษฎีและหลักเกณฑ์ที่เกี่ยวข้อง	7
บทที่ 3 ขั้นตอนการดำเนินงาน	42
บทที่ 4 ผลการวิจัย และวิจารณ์	51
บทที่ 5 สรุปผลการศึกษาวิจัยและข้อเสนอแนะ	89
ภาคผนวก - คู่มือการใช้โปรแกรม EPSM.EXE	93
- คู่มือการใช้ภาษาไทย TSM	115
เอกสารอ้างอิง	

บทที่ 1
บทนำ

ความสำคัญ/ที่มาของปัญหา

ในปัจจุบันมีการทำวิจัยศึกษาค้นคว้ากันอย่างกว้างขวาง เพื่อให้ได้มาซึ่งความรู้และทฤษฎีใหม่ ๆ ซึ่งจะต้องมีการวางแผนการศึกษาให้ถูกต้องเหมาะสม เพื่อให้สิ้นเปลืองน้อยที่สุดโดยให้ผลตอบแทนมากที่สุดและมีความเชื่อถือได้มากที่สุด โดยเฉพาะอย่างยิ่งในโครงการใหญ่ ๆ ย่อมต้องมีการวางแผนให้ดีและรัดกุมที่สุด การตัดสินใจใด ๆ โดยมีเหตุผลย่อมต้องการข้อเท็จจริงมาประกอบการพิจารณา ดังนั้นในกิจการต่าง ๆ เช่น อุตสาหกรรม ธุรกิจ องค์การรัฐบาล ฯลฯ จึงต้องการข้อมูลสถิติเพื่อประโยชน์ในการวางแผนและการตัดสินใจ ดังนั้นจึงมีการรวบรวมข้อเท็จจริงทางสถิติและนำเอาข้อมูลเหล่านั้นมาวิเคราะห์และเสนอให้เหมาะสมเพื่อใช้เป็นรากฐานในการตัดสินใจวางนโยบายในกิจการต่างๆ

แต่ในการที่จะรวบรวมข้อเท็จจริงอาจจะมีข้อผิดพลาดได้ถึงแม้จะใช้วิธีสำมะโน ซึ่งเป็นวิธีที่สำรวจหน่วยทุกหน่วยที่อยู่ในประชากรที่สนใจก็ตาม และยังมีปัญหาว่าถ้าประชากรมีขนาดใหญ่มาก ๆ การเก็บรวบรวมข้อมูลได้ทั้งหมด จะทำให้เสียเวลาและค่าใช้จ่ายมาก หากประชากรประกอบด้วยหน่วยที่มีคุณสมบัติคล้ายคลึงกัน การเก็บรวบรวมข้อมูลทั้งหมดจะทำให้เสียเวลาและค่าใช้จ่ายโดยเปล่าประโยชน์ เพราะการที่เลือกใช้หน่วยใดหน่วยหนึ่งในประชากรเพื่อมาสรุปผลเกี่ยวกับประชากร ย่อมจะได้ผลดีเท่ากับการเลือกใช้หลายๆ หน่วยและจะเลือกใช้หน่วยใดก็ได้ผลเท่ากัน แต่โดยทั่วไปแล้วประชากรจะประกอบไปด้วยหน่วยที่มีความแตกต่างกัน การใช้หน่วยใดหน่วยหนึ่งหรือเพียงบางส่วนของประชากรมาสรุปผลเกี่ยวกับประชากรอาจจะได้ข้อเท็จจริงที่ไม่ตรงกับประชากร และข้อเท็จจริงนี้จะแตกต่างกันไป แล้วแต่ว่าจะเลือกได้หน่วยใด

เพื่อที่จะได้ตัวอย่างที่เป็นตัวแทนของประชากรได้ดี จึงมักใช้การสุ่ม(Randomization)ในการเลือกตัวอย่างจากประชากร กล่าวคือไม่มีการเลือกหน่วยประชากรหน่วยใดโดยจงใจ ไม่มีการใช้ความรู้สึก ไม่ใช้ความชำนาญ หรือความรู้เกี่ยวกับหน่วยนั้นๆ ซึ่งเป็นการเลือกตัวอย่างโดยลำเอียง ตัวอย่างที่เลือกได้โดยการสุ่มเรียกว่า ตัวอย่างสุ่ม (Random Sample) และในการสุ่มตัวอย่างจากประชากรจะมีข้อจำกัดเกี่ยวกับลักษณะของประชากรว่ามีความคล้ายคลึงกันหรือแตกต่างกันมากน้อยเพียงใด จึงต้องอาศัยทฤษฎีการสุ่มตัวอย่าง

สถิติทางด้านการสุ่มตัวอย่างได้เข้ามามีบทบาทสำคัญในงานวิจัยต่าง ๆ เช่น

- โครงการสำรวจแรงงาน
- การสำรวจรายได้-รายจ่ายของครัวเรือน
- การสำรวจผลผลิตข้าว และต้นทุนการผลิตข้าว
- การสำรวจปริมาณการขนส่งทางบกและทางน้ำ
- การสำรวจการเปลี่ยนแปลงของประชากร
- การสำรวจวิทยุโทรทัศน์
- การสำรวจรายจ่ายนักท่องเที่ยว

นอกจากนั้นการเก็บข้อมูลในงานวิจัยต่าง ๆ ส่วนมากประชากรจะมีขนาดใหญ่ จึงทำให้ไม่สามารถทำการเก็บรวบรวมข้อมูลทั้งหมดได้ในเวลาและงบประมาณที่จำกัด เป็นผลให้ต้องใช้ตัวอย่างเพื่อความสะดวกและลดค่าใช้จ่าย ซึ่งแผนแบบการสุ่มตัวอย่างก็มีอยู่หลายวิธีด้วยกันคือ

1. Simple Random Sampling
2. Systematic Sampling
3. Sampling with Unequal Probability
4. Stratified Sampling
5. Cluster Sampling
6. Multi-Stage Sampling

โดยที่ในแต่ละแผนแบบการสุ่มตัวอย่าง จะมีสูตรในการประมาณค่าสถิติที่แตกต่างกัน ผู้ใช้ต้องเลือกใช้แผนแบบการสุ่มตัวอย่างที่ เหมาะสมกับลักษณะของประชากรที่ต้องการเก็บรวบรวมข้อมูลมากที่สุด และสูตรในการคำนวณค่าของการสุ่มแต่ละวิธีค่อนข้างยุ่งยาก ซับซ้อน และใช้เวลาในการคำนวณมาก

ปัจจุบันนี้เทคโนโลยีในด้านต่าง ๆ ได้พัฒนามากขึ้นและคอมพิวเตอร์ก็เป็นเทคโนโลยีใหม่ที่ช่วยอำนวยความสะดวก ประหยัดเวลา และสามารถประยุกต์ใช้กับงานต่าง ๆ ได้มากมาย ซึ่งใช้กันอย่างแพร่หลายในปัจจุบัน จึงทำให้มีผู้สร้างโปรแกรมสำเร็จรูปขึ้นมามากมาย เพื่ออำนวยความสะดวกในการใช้งานด้านต่าง ๆ รวมทั้งโปรแกรมคำนวณทางด้านสถิติ เช่น SPSS/PC^{+} , SAS , MINITAB , MICROSTAT ฯลฯ ก็ได้ถูกสร้างขึ้นมากมายเพื่อช่วยให้คำนวณค่าได้โดยง่ายและประหยัดเวลา แต่โปรแกรมเหล่านี้ยังมีข้อจำกัด คือใช้กับแผนแบบการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) เท่านั้น แต่ถ้าการสุ่มตัวอย่างเป็นแบบวิธีอื่นก็อาจจะใช้โปรแกรมวิเคราะห์ดังกล่าวได้แต่ต้องมีการปรับปรุงข้อมูลดิบก่อน จึงควรจะมีโปรแกรมสำหรับคำนวณค่าต่าง ๆ ตามแผนแบบการสุ่มตัวอย่างที่เลือกใช้ เพื่อให้ได้ผลลัพธ์ที่ถูกต้องมากยิ่งขึ้นและจะพบว่าในปัจจุบันโปรแกรมต่าง ๆ ได้ถูกพัฒนาให้ใช้งานกับระบบภาษาไทย แต่โปรแกรมคำนวณทางด้านสถิติยังไม่ค่อยพัฒนาให้ใช้กับระบบภาษาไทยเท่าใดนัก จึงได้จัดทำโปรแกรมในการคำนวณค่าสถิติเบื้องต้นต่าง ๆ ตามแผนแบบการสุ่มตัวอย่างโดยใช้กับระบบภาษาไทยเพื่อช่วยให้ผู้ทำการวิจัยสามารถคำนวณค่าต่าง ๆ ได้สะดวก ประหยัดเวลา ใช้โปรแกรมได้โดยง่าย และเป็นจุดเริ่มหรือแนวทางในการพัฒนาโปรแกรมอื่น ๆ ต่อไป

วัตถุประสงค์ของปัญหา

1. เพื่อสร้างโปรแกรมสำหรับใช้ในการคำนวณค่าสถิติเบื้องต้นต่าง ๆ ตามแผนแบบการสุ่มตัวอย่าง และช่วยผู้ใช้งานเลือกแผนแบบสุ่มตัวอย่างและหาขนาดตัวอย่างพร้อมทั้งสร้างคู่มือการใช้งานของโปรแกรม เพื่อเป็นประโยชน์แก่นักวิจัยหรือผู้สนใจ ให้สามารถคำนวณข้อมูลได้ถูกต้องยิ่งขึ้น

2. เพื่อนำคอมพิวเตอร์และระบบภาษาไทยบนคอมพิวเตอร์เข้ามาช่วยในการคำนวณค่าสถิติทางด้านการสุ่มตัวอย่าง

ขอบเขตและข้อจำกัดของงานวิจัย

จัดทำโปรแกรมที่ใช้ระบบภาษาไทยเข้าช่วยในการติดต่อกับผู้ใช้ ซึ่งใช้กับเครื่อง IBM หรือ IBM Compatible ขนาด 16 บิตขึ้นไป จอภาพ VGA มีหน่วยความจำ (RAM) ไม่น้อยกว่า 640 KB โดยโปรแกรมนี้มีความสามารถดังต่อไปนี้

1. คำนวณและแสดงค่าสถิติเบื้องต้นตามวิธีการสุ่มตัวอย่าง มีรายละเอียดดังนี้

1.1 Simple Random Sampling (SRS)

1.2 Systematic Sampling

1.3 Sampling with Probability Proportional to size (PPS)

1.4 Stratified Sampling เฉพาะกรณีที่ในแต่ละ stratum สุ่มแบบ SRS เท่านั้น

1.5 Cluster Sampling

1.6 Multi-Stage Sampling เฉพาะ Two-Stage บางกรณีเท่านั้น คือ

1.6.1 ขั้นแรกสุ่มแบบ SRS ขั้นที่สองสุ่มแบบ SRS

1.6.2 ขั้นแรกสุ่มแบบ PPS ขั้นที่สองสุ่มแบบ SRS

โดยทุกการสุ่มจะมีการคำนวณค่าดังนี้

ค่าประมาณยอดรวม ($\hat{Y}$) , ค่าประมาณค่าเฉลี่ย ($\bar{y}$) , ค่าประมาณสัดส่วน (p) , ค่าประมาณอัตราส่วน (r) , พร้อมทั้งความแปรปรวนของค่าประมาณ , ค่าสัมประสิทธิ์ของความแปรผัน (CV.) และค่าช่วงความเชื่อมั่น (CI.) ของแต่ละค่าประมาณ

ซึ่งโปรแกรมสามารถคำนวณค่าประมาณต่าง ๆ ได้ละเอียดถึง 15 หลักและจะแสดงผลในรูปทศนิยม 4 ตำแหน่ง

2. คำนวณและแสดงค่าขนาดตัวอย่างที่เหมาะสม โดยแยกตามแผนแบบการสุ่มตัวอย่าง การกำหนดค่าความผิดพลาดที่ยอมให้มีได้ และค่าประมาณที่ต้องการประมาณ มีรายละเอียดดังนี้

2.1 Simple Random Sampling (SRS)

- เพื่อประมาณค่าเฉลี่ย ($\bar{y}$)
  - กำหนดค่าผิดพลาดเป็นแบบ absolute error (d)
  - กำหนดค่าผิดพลาดเป็นแบบ relative error (r)
- เพื่อประมาณสัดส่วน (p)
  - กำหนดค่าผิดพลาดเป็นแบบ absolute error (d)
  - กำหนดค่าผิดพลาดเป็นแบบ relative error (r)

2.2 Sampling with Probability Proportional to size (PPS)

- เพื่อประมาณค่าเฉลี่ย ($\bar{y}$)
  - กำหนดค่าผิดพลาดเป็นแบบ absolute error (d)
  - กำหนดค่าผิดพลาดเป็นแบบ relative error (r)

2.3 Stratified Sampling

- เพื่อประมาณค่าเฉลี่ย ($\bar{y}$)
  - กำหนดค่าผิดพลาดเป็นแบบ absolute error (d)
    - Optimum allocation
    - Neyman allocation
    - Proportional allocation
    - Equal allocation
  - กำหนดค่าผิดพลาดเป็นแบบ relative error (r)
    - Optimum allocation
    - Neyman allocation
    - Proportional allocation
    - Equal allocation

2.4 Cluster Sampling

- เพื่อประมาณค่าเฉลี่ย ($\bar{y}$)

- กำหนดค่าผิดพลาดเป็นแบบ absolute error (d)

- กำหนดค่าผิดพลาดเป็นแบบ relative error (r)

ซึ่ง absolute error (d) = $\left|\text{ค่าประมาณ} - \text{ค่าจริง}\right|$

relative error (e) = $\left|\frac{\text{ค่าประมาณ} - \text{ค่าจริง}}{\text{ค่าจริง}}\right|$

3. หาแผนแบบการสุ่มตัวอย่างที่เหมาะสม สำหรับกรณีที่ผู้ใช้รู้ลักษณะประชากร แต่ไม่ทราบว่าควรจะเลือกแผนแบบการสุ่มตัวอย่างแบบใด

ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้การคำนวณค่าสถิติเบื้องต้นตามแผนแบบการสุ่มตัวอย่างในแต่ละแบบ ทำได้สะดวกรวดเร็วและได้ค่าที่ถูกต้องยิ่งขึ้น

2. ช่วยในการหาแผนแบบการสุ่มตัวอย่าง สำหรับผู้วิจัยที่ไม่ทราบว่าจะสุ่มตัวอย่างแบบใดถึงจะเหมาะสม

3. ช่วยคำนวณหาขนาดตัวอย่างที่เหมาะสม

4. เพื่อเป็นแนวทางสำหรับผู้ที่สนใจ สามารถนำไปศึกษาและทำการพัฒนาโปรแกรมให้คำนวณค่าด้วยวิธีการสุ่มตัวอย่างอื่น ๆ อีกได้ หรือพัฒนาโปรแกรมให้สามารถนำค่าที่ได้มาประยุกต์ใช้กับการทดสอบสมมติฐานหรือการคำนวณค่าสถิติอื่น ๆ ได้กว้างขึ้น

บทที่2

ทฤษฎีและหลักเกณฑ์ที่เกี่ยวข้อง

เนื่องจากในการที่จะอธิบายทฤษฎีการสุ่มตัวอย่างนั้นจะมีคำศัพท์เฉพาะที่อาจทำให้ไม่เข้าใจได้จึงขออธิบายไว้ดังนี้

นิยามศัพท์ที่ใช้ในการสุ่มตัวอย่าง

1. พารามิเตอร์ (Parameter) หมายถึง ค่าคงที่ที่แสดงคุณลักษณะของประชากร ดังนั้นพารามิเตอร์จึงเป็นค่าที่คำนวณได้จากหน่วยทุกหน่วยในประชากร

2. ตัวประมาณค่า (Estimator) หมายถึง รูปทั่วไปของตัวสถิติที่ได้จากวิธีการใดวิธีการหนึ่งเพื่อใช้ประมาณค่าพารามิเตอร์

3. ค่าประมาณ หรือ ค่าสถิติ (Estimate) หมายถึง ค่าหนึ่ง ๆ ของตัวประมาณที่ได้มาโดยการแทนค่าของข้อมูลจากตัวอย่างหนึ่ง ๆ ที่เลือกได้ในรูปทั่วไปของตัวประมาณค่า

4. หน่วยตัวอย่าง (Sampling Unit) หมายถึง หน่วยที่ต้องการหรือวัดรายละเอียดข้อเท็จจริงบางประการ หน่วยตัวอย่างอาจจะเป็นหน่วยเล็ก ๆ เพียงหน่วยเดียว

5. ตัวอย่าง (Sample) หมายถึง หน่วยตัวอย่างที่เป็นส่วนหนึ่งของประชากรที่ถูกเลือกขึ้นมาเพื่อเป็นตัวแทนในการศึกษาหาข้อเท็จจริงที่ต้องการเกี่ยวกับประชากรนั้น

6. ขนาดของประชากร (Population size) หมายถึง จำนวนของหน่วยตัวอย่างทั้งหมดของประชากร มักใช้สัญลักษณ์ว่า N

7. ขนาดตัวอย่าง หมายถึง จำนวนของหน่วยตัวอย่างทั้งหมดในตัวอย่างมักใช้สัญลักษณ์ว่า n

8. กรอบตัวอย่าง (Sampling frame) หมายถึง รายชื่อหรือทำเนียบของหน่วยตัวอย่างทุกหน่วยในประชากร พร้อมทั้งตำบลที่อยู่ที่สามารถติดต่อได้ในปัจจุบัน หรืออาจจะหมายถึงแผนที่ที่แสดงอาณาเขตของหน่วยตัวอย่างต่าง ๆ อย่างชัดแจ้ง

9. ความแม่น (accuracy)หมายถึงการที่ค่าสถิติที่คำนวณได้มาจากการสุ่มตัวอย่าง มีค่าใกล้เคียงกับค่าที่แท้จริง ถือว่าค่าที่คำนวณได้จากตัวอย่างที่สุ่มนี้มีความแม่นสูง

10. ความเที่ยง (Precision) หมายถึงการดำเนินการสุ่มตัวอย่างในเงื่อนไขเดิม และได้ค่าประมาณเท่าเดิม หรือใกล้เคียงซ้ำกันหลาย ๆ ครั้งนั้นถือว่ามีความเที่ยงสูง เช่น คน หรืออาจจะเป็นกลุ่มของหน่วยเล็ก ๆ ก็ได้ เช่น ครัวเรือน

การสุ่มตัวอย่างมีวัตถุประสงค์ที่สำคัญ คือ ให้ได้มาซึ่งข้อมูลที่เป็นตัวแทนที่ดีของประชากร และนำข้อมูลที่ได้มาจากการสุ่มตัวอย่างนั้นไปประมาณค่าพารามิเตอร์ ซึ่งแผนแบบการสุ่มตัวอย่างที่นิยมใช้มีดังนี้

1. แผนแบบการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling, SRS)
2. แผนแบบการสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling)
3. แผนแบบการสุ่มตัวอย่างโดยกำหนดความน่าจะเป็น (Sampling with Unequal probability)
4. แผนแบบการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Sampling)
5. แผนแบบการสุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster Sampling)
6. แผนแบบการสุ่มตัวอย่างหลายขั้นตอน (Multi-stage Sampling)

โดยที่ในแต่ละกรณีนั้นมีการคำนวณที่น่าสนใจคือ

1. การประมาณค่าของ Parameter แบ่งเป็น
   - Population Mean ประมาณค่าเฉลี่ยของประชากร
   - Population Total ประมาณยอดรวมของประชากร
   - Population Proportion ประมาณสัดส่วนหรือเปอร์เซนต์ของหน่วยตัวอย่างในประชากรที่อยู่ในส่วนที่สนใจ
   - Population Ratio เป็นการสนใจอัตราส่วนของตัวแปร 2 ตัวจากประชากรขนาด N

2. การหาขนาดตัวอย่าง

ในแต่ละแผนแบบการสุ่มตัวอย่างนั้นจะมีข้อจำกัดในการใช้งานและสูตรการคำนวณค่าประมาณต่าง ๆ ที่แตกต่าง ๆ กัน ดังต่อไปนี้

1.แผนแบบการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling, SRS)

เป็นวิธีการสุ่มตัวอย่างขนาด n หน่วย จากประชากรขนาด N หน่วย โดยกำหนดให้หน่วยตัวอย่างแต่ละหน่วยในประชากรมีโอกาสจะถูกเลือกเป็นตัวแทนได้เท่า ๆ กัน คือ

P(หน่วยตัวอย่างแต่ละหน่วยจะถูกเลือกในการเลือกแต่ละครั้ง) = $^{1}/_{N}$

และตัวอย่างที่ได้โดยวิธีนี้เรียกว่า ตัวอย่างสุ่ม (Random Sample)

แผนแบบการสุ่มตัวอย่างแบบง่ายนี้เหมาะสำหรับประชากรที่มีลักษณะคล้ายคลึงกัน (Homogeneous) แต่ถ้าใช้การสุ่มตัวอย่างแบบ SRS กับประชากรที่มีลักษณะแตกต่างกัน ค่า n จะต้องมาก เพื่อให้ตัวอย่างที่ได้ครอบคลุมทุกลักษณะของประชากร

วิธีการเลือกหน่วยตัวอย่าง

แบ่งออกเป็น 2 วิธีคือ

1. การเลือกแบบไม่มีการแทนที่ (Sampling without replacement)เป็นการเลือกหน่วยตัวอย่าง โดยที่ถ้าหน่วยตัวอย่างใดถูกเลือกไปแล้ว จะไม่มีโอกาสถูกเลือกซ้ำอีก

2. การเลือกแบบมีการแทนที่ (Sampling with replacement) เป็นการเลือกหน่วยตัวอย่าง โดยที่ถ้าหน่วยตัวอย่าง ใดถูกเลือกไปแล้ว จะมีโอกาสถูกเลือกซ้ำอีก

ซึ่งทั้งการเลือกแบบแทนที่ และการเลือกแบบไม่แทนที่ สามารถทำได้ดังนี้คือ

- จับฉลาก (ใช้สำหรับประชากรที่มีขนาดเล็ก) โดยให้หมายเลขกำกับแต่ละสลากตั้งแต่เลข 1 จนถึง N สุ่มจับทีละสลากจนครบ n สลาก ถ้าสุ่มแบบมีการแทนที่สลากที่ถูกสุ่มขึ้นมาเป็นตัวอย่างแล้วก็จะถูกใส่กลับเข้าไปใหม่ ทำให้มีโอกาสที่จะถูกสุ่มขึ้นมาเป็นตัวแทนอีก ถ้าเป็นการสุ่มแบบไม่มีการแทนที่ สลากที่ตกเป็นตัวอย่างแล้วจะไม่ใส่กลับคืนเข้าไป ดังนั้น จึงไม่มีโอกาสที่จะถูกสุ่มขึ้นมาเป็นตัวแทนอีก

- ตารางเลขสุ่ม เป็นตารางที่ได้จัดเตรียมไว้สำหรับการสุ่มตัวอย่างโดยเฉพาะ เหมาะสำหรับงานขนาดใหญ่ ซึ่งประชากรมีจำนวนไม่จำกัด กล่าวคือ ประชากรประกอบด้วยหน่วยตัวอย่างจำนวนมาก วิธีใช้คือ

1. ให้หมายเลขแก่หน่วยตัวอย่างทุกหน่วยในกรอบตัวอย่าง(Sampling Frame)

2. สุ่มหน้าของตารางเลขสุ่ม (ในกรณีที่ตารางสุ่มมีหลายหน้า)

3. ในหน้าที่ตกเป็นตัวอย่าง สุ่มหาจุดเริ่มต้น ซึ่งเรียกว่า Random Start

4. การเลือกหน่วยถัดไป จะเลือกหน่วยที่อยู่ทางซ้าย, ขวา, บน หรือล่างก็ได้ และเลือกต่อไปจนครบตามที่ต้องการ

ข้อดี

1. เป็นแผนแบบการสุ่มตัวอย่างที่ง่าย และใช้สะดวก

2. การประมาณค่าต่าง ๆ ง่าย

3. ขนาดตัวอย่าง (n) ไม่ครบตามที่ต้องการก็ไม่มีปัญหาเนื่องจากสามารถใช้ขนาดตัวอย่าง (n) ในการคำนวณเท่ากับที่เก็บรวบรวมได้

ข้อเสีย

1. ถ้าประชากรมีลักษณะแตกต่างกันขนาดตัวอย่างที่ใช้จะใหญ่มากซึ่งจะทำให้ยุ่งยากเสียเวลา และเสียค่าใช้จ่ายสูง ในการเก็บรวบรวมข้อมูล

2. ตัวประมาณค่าทุกตัวจะเป็น Unbiased Estimator แต่ค่าความแปรปรวนไม่ได้มีค่าน้อยที่สุด เนื่องจากยังมีวิธีอื่นที่ให้ค่าความแปรปรวนน้อยกว่าวิธีนี้

3. ไม่สามารถคำนวณค่าของรายเขตย่อย ๆ ได้

สูตรที่ใช้

การประมาณค่าเฉลี่ย

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$\widehat{V(\bar{y})} = \frac{N-n}{n} \cdot \frac{s^2}{n}$$

โดยที่ $s^2 = \dfrac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$

<u>การประมาณค่ายอดรวม</u>

$$\hat{Y} = N\bar{y}$$

$$\widehat{V(\hat{Y})} = N^2 \widehat{V(\bar{y})}$$

<u>การประมาณค่าสัดส่วน</u>

$$p = \frac{a}{n} = \bar{y}$$

$$\widehat{V(p)} = \frac{N-n}{N} \cdot \frac{pq}{n-1} = \widehat{V(\bar{y})} \quad , \quad s^2 = \frac{npq}{n-1}$$

<u>การประมาณอัตราส่วน</u>

$$r = \frac{\bar{y}}{\bar{x}} = \frac{\hat{Y}}{\hat{X}}$$; r=ค่าประมาณอัตราส่วน

$$\widehat{V(r)} = \frac{1}{\bar{x}^2} \cdot \frac{N-n}{Nn} \cdot [s_y^2 + r^2 s_x^2 - 2rs_{xy}]$$;r=ค่าประมาณอัตราส่วน

$$s_y^2 = \frac{\sum_{i=1}^{n} (y_i-\bar{y})^2}{n-1}$$

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

$$s_{xy} = \frac{\sum_{i=1}^{n} (y_i-\bar{y})(x_i-\bar{x})}{n-1}$$

<u>การหาขนาดตัวอย่างเมื่อต้องการประมาณค่าเฉลี่ย</u>

- กรณีใช้ absolute error (d)

$$n = \frac{N Z^2_{\alpha/2} S^2}{N d^2+Z^2_{\alpha/2} \cdot S^2}$$

$$d = \left|\bar{y}-\bar{Y}\right|$$

- กรณีใช้ relative error (r)

$$n = \frac{n_o}{1+\frac{n_o}{N}}$$

$$n_o = \frac{Z^2_{\alpha/2} S^2}{r^2\bar{Y}^2}$$; r=ค่า relative error

<u>การหาขนาดตัวอย่างเมื่อต้องการประมาณสัดส่วน</u>

- กรณีใช้ absolute error (d)

$$n = \frac{N Z^2_{\alpha/2} PQ}{(N-1)d^2+Z^2_{\alpha/2}PQ}$$

$$d = \left|p-P\right|$$

- กรณีใช้ relative error (r)

$$n = \frac{N Z^2_{\alpha/2}PQ}{(N-1)(rP)^2+Z^2_{\alpha/2}PQ}$$

<u>การหาสัมประสิทธิ์ของความแปรผันของตัวประมาณค่า</u>

$$CV.(\hat{\theta}) = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}} \cdot 100$$

<u>การหาช่วงความเชื่อมั่นของตัวประมาณค่า</u>

$$CI.(\hat{\theta}) \text{ คือ } \hat{\theta}-Z_{\alpha/2}\sqrt{\hat{V}(\hat{\theta})} < \theta < \hat{\theta}+Z_{\alpha/2}\sqrt{\hat{V}(\hat{\theta})}$$

y_i ค่าสังเกตของหน่วยตัวอย่างที่ i ในตัวอย่าง

n ขนาดของตัวอย่าง

$\bar{y}$ ค่าประมาณค่าเฉลี่ย

$\hat{V}(\bar{y})$ ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ย

N ขนาดของประชากร

s^2 ความแปรปรวนของตัวอย่าง

p สัดส่วนของหน่วยตัวอย่างที่สนใจในตัวอย่าง

q สัดส่วนของหน่วยตัวอย่างที่ไม่สนใจในตัวอย่าง

$Z_{\alpha/2}$ ค่าจากตารางการแจงแจกปกติที่ $\alpha/2$

α ระดับนัยสำคัญ

S^2 ค่าความแปรปรวนของประชากร

d ค่า absolute error

$\bar{Y}$ ค่าเฉลี่ยของประชากร

$\hat{Y}$ ค่าประมาณยอดรวม

$\hat{V}(\hat{Y})$ ค่าประมาณความแปรปรวนของค่าประมาณยอดรวม

a จำนวนหน่วยตัวอย่างที่อยู่ในพวกที่สนใจในตัวอย่าง

$\hat{V}(p)$ ค่าประมาณความแปรปรวนของค่าประมาณสัดส่วน

P สัดส่วนของประชากรที่สนใจ

Q สัดส่วนของประชากรที่ไม่สนใจ

x_i ค่าสังเกต x ของหน่วยตัวอย่างที่ i ในตัวอย่าง

$\bar{x}$ ค่าประมาณค่าเฉลี่ยของ x

$\hat{V}(r)$ ค่าประมาณความแปรปรวนของค่าประมาณอัตราส่วน

s_y^2 ค่าความแปรปรวนของตัวแปร y ในตัวอย่าง

s_x^2 ค่าความแปรปรวนของตัวแปร x ในตัวอย่าง

s_{xy} ค่าความแปรปรวนร่วมของ x และ y ในตัวอย่าง

$CV.(\hat{\theta})$ ค่าสัมประสิทธิ์ของความแปรผันของค่าประมาณพารามิเตอร์ θ

$\hat{V}(\hat{\theta})$ ค่าประมาณความแปรปรวนของค่าประมาณพารามิเตอร์ θ

$\hat{\theta}$ ค่าประมาณพารามิเตอร์ θ

$CI.(\hat{\theta})$ ค่าช่วงความเชื่อมั่นของพารามิเตอร์ θ ที่ระดับความเชื่อมั่น $(1-\alpha)100$ %

2.แผนแบบการสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling)

เป็นการสุ่มตัวอย่าง โดยจะสุ่มตัวอย่างขึ้นมา 1 หน่วยจากทุกๆ k หน่วยตัวอย่าง โดยมีขั้นตอนในการสุ่มดังนี้

1. กำหนด k หรือความกว้างของช่วง k เท่ากับ N/n

2. สุ่มตัวอย่างหน่วยตัวอย่าง 1 หน่วยใน k โดยใช้การจับสลากหรือวิธีการใดก็ได้ เป็นหน่วยตัวอย่างหน่วยแรกที่ตกเป็นตัวอย่าง เรียกว่า random start

3. จากหน่วยตัวอย่างหน่วยแรกที่ตกเป็นตัวอย่าง(จากข้อ 2) สมมติว่าให้เท่ากับ r (โดยที่ค่า r มีค่าได้ตั้งแต่ 1 ถึง k) หน่วยถัดไปคือ r+k, r+2k, เรื่อยไปจนครบ n หน่วย

แผนแบบการสุ่มตัวอย่างแบบมีระบบนี้ เหมาะสำหรับประชากรที่มีการเรียงกันแบบสุ่ม หรือลำดับธรรมชาติ (Random) หรือแบบการเรียงลำดับ (Order) จะเรียงจากน้อยไปมากหรือมากไปน้อยก็ได้ แต่จะไม่ใช้กับข้อมูลที่มีฤดูกาลมาเกี่ยวข้อง และถ้าประชากรมีลักษณะการเรียงแบบ Random แล้วแผนการสุ่มตัวอย่างแบบมีระบบจะมีประสิทธิภาพเหมือนกับ SRS ดังนั้นในการประมาณค่าสถิติต่างๆ จึงใช้สูตรของแผนแบบการสุ่มตัวอย่างแบบง่าย

ข้อดี

1. การเลือกตัวอย่างทำได้ง่าย ประหยัดเวลา และมักไม่ค่อยผิด โดยเฉพาะอย่างยิ่งในทางปฏิบัติ ถ้าหน่วยตัวอย่างอยู่เรียงกันเป็นจำนวนมาก

2. การสุ่มตัวอย่างแบบมีระบบ มักจะมีความแม่นยำกว่าการสุ่มตัวอย่างแบบง่าย เพราะว่าตัวอย่าง ที่ได้จะกระจายไปทั่วประชากร

3. ถ้าไม่ทราบขนาดประชากร เราสามารถทำการสุ่มตัวอย่างได้โดยกำหนดค่า k ขึ้นมา เช่น สอบถามทัศนคติของผู้เข้าชมนิทรรศการ เป็นต้น และสามารถหาขนาดประชากรได้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ข้อเสีย

1. ขนาดตัวอย่างไม่แน่นอน ซึ่งอาจจะแตกต่างจากที่ตั้งไว้ ถ้า N/n ไม่ลงตัว หรือ N ไม่เท่ากับ nk

2. เมื่อ N/n ไม่ลงตัว หรือ N ไม่เท่ากับ nk ค่าเฉลี่ยของตัวอย่างที่ได้จะเป็นตัวประมาณค่าที่เป็น Biased Estimator ของค่าเฉลี่ยของประชากร

3. การประมาณค่าความแปรปรวนของค่าเฉลี่ยของตัวอย่าง จะประมาณโดยใช้ตัวอย่างเดียว และไม่สามารถหาค่าประมาณของความแปรปรวนของค่าเฉลี่ยของตัวอย่างที่ไม่มีความเอียงเอนได้

สูตรที่ใช้

นิยมใช้สูตรเดียวกับแผนแบบการสุ่มตัวอย่างแบบง่าย

3. แผนแบบการสุ่มตัวอย่างโดยกำหนดความน่าจะเป็นของการเลือกไม่เท่ากัน (Sampling with Unequal probability)

จากแผนแบบการสุ่มตัวอย่างโดยใช้ Probability Sampling จากประชากรที่มีอยู่ทั้งหมด N หน่วย จะเห็นได้ว่า Probability Sampling ที่ง่ายและสะดวกที่สุดคือ Equal Probability Sampling หรือ แผนแบบการสุ่มตัวอย่างแบบง่าย โดยเป็นการเลือกที่ให้โอกาสแก่ทุกหน่วยในประชากรที่จะถูกเลือกเท่า ๆ กัน

ในกรณีที่ประชากรมีลักษณะแตกต่างกัน การที่กำหนดให้หน่วยตัวอย่างแต่ละหน่วยในประชากรมีโอกาสถูกเลือกเท่าๆ กันจะทำให้มีความเที่ยง (precision) น้อย และถ้าต้องการจะเพิ่มความเที่ยงให้มากขึ้นจะมีวิธีการเลือกหน่วยตัวอย่าง โดยให้หน่วยตัวอย่างแต่ละหน่วยในประชากรมีโอกาสถูกเลือกไม่เท่ากัน ซึ่งก็คือวิธี Sampling with Unequal Probability จะเป็นการเลือกตัวอย่างที่มีการกำหนดค่าความน่าจะเป็นที่หน่วยตัวอย่างแต่ละหน่วยถูกเลือกไม่เท่ากัน โดยอาจกำหนดตามขนาดของหน่วยตัวอย่าง (size) ก็ได้โดยให้ขนาดของหน่วยตัวอย่างนั้นจะต้องมีความสัมพันธ์กับสิ่งที่ต้องการศึกษา คือ ถ้าหน่วยตัวอย่างใดมีขนาดใหญ่ก็จะมีโอกาสที่ถูกเลือกมากกว่าหน่วยตัวอย่างที่มีขนาดเล็ก และการเลือกตัวอย่างโดยกำหนดความน่าจะเป็นที่แต่ละหน่วยตัวอย่างจะถูกเลือกตามขนาดนี้จะเรียกว่า Sampling with Probability Proportional to Size (PPS)

แผนแบบการสุ่มตัวอย่างนี้ เหมาะสำหรับประชากรที่มีลักษณะแตกต่างกัน และไม่สามารถจัดเป็นกลุ่มได้

วิธีการเลือกหน่วยตัวอย่าง

1. PPS with replacement

จากกรอบตัวอย่าง (Sampling frame) ที่มีประชากรขนาด N ต้องการสุ่มตัวอย่าง ขนาด n ทำโดยสุ่มเลขขึ้นมา n ตัว จากตารางเลขสุ่มถ้าเลขสุ่มนั้นตกอยู่ในผลบวกสะสมของขนาดของหน่วยตัวอย่าง (Cummulative Size) ของหน่วยตัวอย่างใด

หน่วยตัวอย่างนั้นจะตกเป็นตัวอย่าง และหน่วยตัวอย่างแต่ละหน่วยมีโอกาสถูกเลือกซ้ำได้

2. PPS without replacement

วิธีการคล้ายกับ PPS with replacement แต่ตัดตัวเลือกที่ซ้ำออก แล้วเลือก Random number ขึ้นมาใหม่ จนครบ n นั่นคือวิธีนี้จะไม่มี sampling unit ใดถูกเลือกซ้ำเลย

3. PPS Sampling Systematic Sampling

เป็นวิธีการเลือกแบบ PPS ที่ใช้ Systematic เข้าช่วย

Sampling Frame จะต้องประกอบด้วยรายชื่อและที่อยู่ของทุกหน่วยตัวอย่าง และขนาดของหน่วยตัวอย่าง

<u>สูตรที่ใช้</u>

<u>การประมาณค่าเฉลี่ย</u>

$$\bar{y} = \frac{\hat{Y}}{N}$$

$$\widehat{V(\bar{y})} = \frac{1}{N^2} \cdot \frac{1}{n(n-1)} \cdot \sum_{i=1}^{n} \left[\frac{y_i - \hat{Y}}{p_i} \right]^2$$

$$p_i = \frac{M_i}{\sum_{i=1}^{N} M_i}$$

<u>การประมาณค่ายอดรวม</u>

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i}$$

$$\widehat{V(\hat{Y})} = \frac{1}{n(n-1)} \cdot \sum_{i=1}^{n} \left[\frac{y_i - \hat{Y}}{p_i} \right]^2$$

การประมาณสัดส่วน

$$p = \bar{y}$$

$$\widehat{V(p)} = \widehat{V(\bar{y})}$$

การประมาณอัตราส่วน

$$\hat{R} = \frac{\hat{Y}}{\hat{X}}$$

$$\widehat{V(\hat{R})} \simeq \left[\frac{\hat{Y}}{\hat{X}} \right]^2 \cdot \left[\frac{\sum_{i=1}^{n} \frac{(y_i - \hat{Y})^2}{p_i}}{n(n-1)\,\hat{Y}} + \frac{\sum_{i=1}^{n} \frac{(x_i - \hat{X})^2}{p_i}}{n(n-1)\,\hat{X}} - \frac{2\sum_{i=1}^{n} \frac{(y_i - \hat{Y})(x_i - \hat{X})}{p_i}}{n(n-1)\,\hat{X}\hat{Y}} \right]$$

การหาขนาดตัวอย่างเมื่อต้องการประมาณค่าเฉลี่ย

- กรณีใช้ absolute error (d)

$$n = \frac{Z^2_{\alpha/2}}{N^2 d^2} \cdot \left[\sum_{i=1}^{N} \frac{Y_i^2}{P_i} - Y^2 \right]$$

- กรณีใช้ relative error (r)

$$n = Z^2_{\alpha/2} \; \frac{1}{N^2} \cdot \frac{1}{r^2 \bar{Y}^2} \cdot \left[\sum_{i=1}^{N} \frac{Y_i^2}{p_i} - Y^2 \right]$$

การหาสัมประสิทธิ์ของความแปรผันของตัวประมาณค่า

$$CV.(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}} \cdot 100$$

การหาช่วงความเชื่อมั่นของตัวประมาณค่า

$$CI.(\hat{\theta}) \text{ คือ } \hat{\theta}-Z_{\alpha/2}\sqrt{V(\hat{\theta})} < \theta < \hat{\theta}+Z_{\alpha/2}\sqrt{V(\hat{\theta})}$$

y_i ค่าสังเกต y ของหน่วยตัวอย่างที่ i ในตัวอย่าง

n ขนาดของตัวอย่าง

$\bar{y}$ ค่าประมาณค่าเฉลี่ย

$\hat{V}(\bar{y})$ ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ย

p_i ค่าความน่าจะเป็นที่จะตกเป็นตัวอย่างของหน่วยตัวอย่างที่ i

M_i ขนาดของหน่วยตัวอย่างที่ i

N ขนาดประชากร

$Z_{\alpha/2}$ ค่าจากตารางการแจงแจกปกติที่ $\alpha/2$

d ค่า absolute error

$\bar{Y}$ ค่าเฉลี่ยของประชากร

$\hat{Y}$ ค่าประมาณยอดรวม

$\hat{V}(\hat{Y})$ ค่าประมาณความแปรปรวนของค่าประมาณยอดรวม

$\hat{V}(p)$ ค่าประมาณความแปรปรวนของค่าประมาณสัดส่วน

x_i ค่าสังเกต x ของหน่วยตัวอย่างที่ i ในตัวอย่าง

$\bar{x}$ ค่าประมาณค่าเฉลี่ยของ x

$\hat{R}$ ค่าประมาณอัตราส่วน

$\hat{V}(\hat{R})$ ค่าประมาณความแปรปรวนของค่าประมาณอัตราส่วน

CV.($\hat{\theta}$) ค่าสัมประสิทธิ์ของความแปรผันของค่าประมาณพารามิเตอร์ ๑

$\widehat{V(\hat{\theta})}$ ค่าประมาณความแปรปรวนของค่าประมาณพารามิเตอร์ ๑

$\hat{\theta}$ ค่าประมาณพารามิเตอร์ ๑

CI.($\hat{\theta}$) ค่าช่วงความเชื่อมั่นของพารามิเตอร์ ๑

4. แผนแบบการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Sampling)

แผนแบบการสุ่มตัวอย่างแบบแบ่งชั้นภูมิเป็น แผนแบบการสุ่มตัวอย่างจากประชากรที่มีการแบ่งหน่วยตัวอย่างในประชากรออกเป็นกลุ่ม ๆ ตามลักษณะบางอย่าง เช่น แบ่งท้องที่ออกเป็นในเขตเทศบาลกับนอกเขตเทศบาล แบ่งประชากรออกเป็นกลุ่ม ๆ ตามอาชีพ เป็นต้น แล้วจึงเลือกตัวแทนจากแต่ละกลุ่มขึ้นมาจำนวนหนึ่งเพื่อเป็นตัวอย่าง โดยที่การเลือกตัวอย่างจากแต่ละชั้นภูมิจะเป็นอิสระกันและใช้วิธีสุ่มตัวอย่างต่างกันได้ขึ้นอยู่กับลักษณะของประชากรในกลุ่มนั้น

วิธีการแบ่งประชากรออกเป็นกลุ่ม ๆ ตามลักษณะบางอย่างจะเรียกว่า Stratification และแต่ละกลุ่มของประชากรที่แบ่งออกไป เรียกว่า ชั้นภูมิ (Stratum หรือ Strata)

หลักสำคัญในการใช้ Stratification ก็คือ พยายามแบ่งประชากรออกเป็นกลุ่มๆ ตามลักษณะบางอย่างของประชากร โดยให้แต่ละกลุ่มประกอบด้วยหน่วยตัวอย่างที่มีความคล้ายคลึงกันมากที่สุด และมีความแตกต่างกันระหว่างกลุ่มมากที่สุด

ในที่นี้ใช้เฉพาะกรณีที่ในแต่ละ Statum สุ่มด้วยแผนแบบการสุ่มตัวอย่างแบบง่าย (Stratified Random Sampling)

ข้อดี

1. เป็นการลดความคลาดเคลื่อนที่เกิดจากการสุ่มตัวอย่าง หรือเป็นการเพิ่มความเที่ยง เพราะว่าในการสุ่มตัวอย่างได้แบ่งหน่วยตัวอย่างออกเป็นกลุ่ม ๆ ก่อน แล้วจึงสุ่มตัวอย่างจากทุกกลุ่ม ทำให้ได้ตัวอย่างที่เป็นตัวแทนประชากรทุกลักษณะในการสำรวจ

2. เนื่องจากแบ่งประชากรออกเป็นกลุ่ม ทำให้สามารถเสนอผลการสำรวจระดับย่อยได้ เช่น การสำรวจผลผลิตข้าวของประเทศอาจมีผู้สนใจต้องการทราบตัวเลขเป็นรายภาคพร้อม ๆ กับตัวเลขที่เป็นส่วนรวมทั้งประเทศ

3. ทำให้สามารถใช้วิธีการเลือกตัวอย่างต่าง ๆ กันได้สำหรับ Stratum ที่ต่างกัน เช่น ชั้นภูมิที่ 1 ใช้ SRS

ชั้นภูมิที่ 2 ใช้ PPS

ชั้นภูมิที่ 3 ใช้ Systematic

และนอกจากนี้การใช้ Stratified Sampling ยังทำให้สามารถกำหนดขนาดตัวอย่างสำหรับแต่ละ Stratum ได้อย่างเหมาะสม

4. เสียงบประมาณน้อยกว่า SRS เช่น ในการประมาณผลผลิตข้าว ถ้าสุ่มตัวอย่างแบบ SRS ข้อมูลจะกระจายไปทั่วทั้ง 72 จังหวัด โดยที่จังหวัดหนึ่งอาจจะมีจำนวนหน่วยตัวอย่างที่ตกเป็นตัวอย่างน้อย อีกจังหวัดหนึ่งมีจำนวนหน่วยตัวอย่างที่ตกเป็นตัวอย่างมาก ซึ่งทำให้ต้องใช้กำลังคนมาก แต่ถ้าสุ่มแบบ Stratified โดยแบ่ง unit ที่ตกเป็นตัวอย่าง ตามสัดส่วนของแต่ละจังหวัด จะทำให้เสียงบประมาณน้อยลง

<u>ข้อเสีย</u>

1. ถ้าจำนวน Stratum มากเกินไป อาจทำให้ไม่สามารถจำกัดขนาดตัวอย่างให้น้อยตามที่ต้องการได้ เช่น จำแนกนิสิตนักศึกษาออกตามมหาวิทยาลัยสาขาวิชาเอก เพศ และชั้นปีการศึกษา สมมติว่าทั้งหมดมี 8 มหาวิทยาลัย 6 สาขาวิชา 2 เพศและ 4 ชั้นปีการศึกษา ดังนั้นจะแบ่งประชากรออกเป็น 8 X 6 X 4 X 2 = 384 stratum ถ้าใช้ตัวอย่าง stratum ละ 10 คน จะต้องใช้ตัวอย่างทั้งหมด 3840 คน ซึ่ง อาจจะเป็นตัวอย่างที่มากเกินไปสำหรับการสำรวจที่มีงบประมาณ และ เวลาในการดำเนินการจำกัด

2. Sampling frame จะต้องมี information เพิ่ม เพื่อนำมาใช้ในการแบ่งหน่วยตัวอย่างทั้งหมดออกเป็นกลุ่ม เช่นแบ่งร้านค้าตามประเภทของสินค้าที่ขาย Sampling frame จะต้องประกอบด้วย ชื่อร้านค้า ที่อยู่และยังต้องมีรายการสินค้าที่ร้านค้านั้นขายด้วย

3. การประมาณผลมีความยุ่งยากกว่า SRS

สูตรที่ใช้

การประมาณค่าเฉลี่ย

$$\bar{y}_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h}$$

$$\hat{Y}_h = \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} = N_h \bar{y}_h$$

$$\bar{y}_{st} = \frac{\hat{Y}_{st}}{N}$$

$$\hat{V}(\bar{y}_{st}) = \frac{1}{N^2} \hat{V}(\hat{Y}_{st})$$

การประมาณค่ายอดรวม

$$\hat{Y}_{st} = \sum_{h=1}^{L} \hat{Y}_h$$

$$\hat{V}(\hat{Y}_{st}) = \sum_{h=1}^{L} N_h^2 \cdot \frac{N_h - n_h}{N_h} \cdot \frac{s_h^2}{n_h}$$

$$s_h^2 = \sum_{i=1}^{n_h} \frac{(y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

การประมาณสัดส่วน

$$p_{st} = \bar{y}_{st}$$

$$\hat{V}(p_{st}) = \hat{V}(\bar{y}_{st})$$

การประมาณอัตราส่วน

$$r_c = \frac{\hat{Y}_{st}}{\hat{X}_{st}} = \frac{\bar{y}_{st}}{\bar{x}_{st}}$$

$$\hat{Y}_{RC} = r_c \hat{X}_{st}$$

$$\hat{V}(\hat{Y}_{RC}) = \sum_{h=1}^{L} X_h^2 \cdot \frac{N_h - n_h}{N_h n_h} \cdot \frac{1}{\bar{X}_h^2} \cdot (s_{yh}^2 + r_c^2 s_{xh}^2 - 2 r_c s_{xyh})$$

$$s_{yh}^2 = \sum_{i=1}^{n_h} \frac{(y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

$$s_{xh}^2 = \sum_{i=1}^{n_h} \frac{(x_{hi} - \bar{x}_h)^2}{n_h - 1}$$

$$s_{xyh}^2 = \sum_{i=1}^{n_h} \frac{(y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h)}{n_h - 1}$$

$$\hat{V}(r) = \frac{1}{X^2} \hat{V}(\hat{Y}_{RC})$$

การหาขนาดตัวอย่างเมื่อต้องการประมาณค่าเฉลี่ย

กรณีใช้ absolute error(d)

- กรณี Optimum allocation

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

- กรณี Neyman allocation

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

- กรณี Proportional allocation

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

- กรณี Equal allocation

$$n = \frac{L\, Z^2_{\alpha/2} \sum_{h=1}^{L} N_h^2 S_h^2}{N^2 d^2 + Z^2_{\alpha/2} \sum_{h=1}^{L} N_h S_h^2}$$

กรณีใช้ relative error(r)

ใช้สูตรเหมือนกรณี absolute error แต่ให้เปลี่ยน d เป็น rY

การหาสัมประสิทธิ์ของความแปรผันของตัวประมาณค่า

$$CV.(\hat{\theta}) = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}} \cdot 100$$

การหาช่วงความเชื่อมั่นของตัวประมาณค่า

$$CI.(\hat{\theta}) \text{ คือ } \hat{\theta} - Z_{\alpha/2}\sqrt{\hat{V}(\hat{\theta})} < \theta < \hat{\theta} + Z_{\alpha/2}\sqrt{\hat{V}(\hat{\theta})}$$

$\bar{y}_{st}$ ค่าประมาณค่าเฉลี่ย

$\hat{Y}_{st}$ ค่าประมาณยอดรวมของตัวแปร Y

N ขนาดของประชากร

$\hat{V}(\bar{y}_{st})$ ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ย

$\hat{V}(\hat{Y}_{st})$ ค่าประมาณความแปรปรวนของค่าประมาณยอดรวม

$\hat{Y}_h$ ค่าประมาณยอดรวมของ Stratum ที่ h

n ขนาดของตัวอย่าง

N_h ขนาดของประชากรใน Stratum ที่ h

n_h ขนาดตัวอย่างใน Stratum ที่ h

s_h^2 ค่าความแปรปรวนของตัวอย่างใน Stratum ที่ h

y_{hi} ค่าสังเกตที่ i ใน Stratum ที่ h

$\bar{y}_h$ ค่าประมาณค่าเฉลี่ยของ Stratum ที่ h

p_{st} ค่าประมาณสัดส่วน

$\hat{V}(p_{st})$ ค่าประมาณความแปรปรวนของค่าประมาณสัดส่วน

r_c ค่าประมาณอัตราส่วนทั้งหมด

$\hat{X}_{st}$ ค่าประมาณยอดรวมของตัวแปร X

$\bar{x}_{st}$ ค่าเฉลี่ยของตัวแปร x

$\hat{Y}_{RC}$ ค่าประมาณยอดรวมจากค่าประมาณอัตราส่วน

$\hat{V}(\hat{Y}_{RC})$ ค่าประมาณความแปรปรวนของค่ายอดรวมจากค่าประมาณอัตราส่วน

L จำนวน Stratum ทั้งหมด

$Z_{\alpha/2}$ ค่าจากตารางการแจงแจกปกติที่ $\alpha/2$

S_h^2 ค่าความแปรปรวนของประชากรใน Stratum ที่ h

S_h ค่าส่วนเบี่ยงเบนมาตรฐานของประชากรใน Stratum ที่ h

C_h ค่าใช้จ่ายในการเก็บข้อมูลใน Stratum ที่ h

d ค่า absolute error

CV.($\hat{\theta}$) ค่าสัมประสิทธิ์ของความแปรผันของค่าประมาณพารามิเตอร์ θ

$\hat{V}(\hat{\theta})$ ค่าประมาณความแปรปรวนของค่าประมาณพารามิเตอร์ θ

$\hat{\theta}$ ค่าประมาณพารามิเตอร์ θ

CI.($\hat{\theta}$) ค่าช่วงความเชื่อมั่นของพารามิเตอร์ θ

5. แผนแบบการสุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster Sampling)

แผนแบบการสุ่มตัวอย่างแบบแบ่งกลุ่มเป็นแผนแบบการสุ่มตัวอย่างที่แบ่งหน่วยตัวอย่างในประชากรออกเป็นกลุ่มๆ ตามธรรมชาติ สภาพภูมิประเทศ หรือตามลักษณะของหน่วยงาน เช่น หมู่บ้าน กอง เป็นต้น โดยกำหนดให้หน่วยตัวอย่างในแต่ละกลุ่มมีลักษณะที่เป็นตัวแทนของประชากรมากที่สุดคือ ให้แต่ละกลุ่มประกอบด้วยหน่วยตัวอย่างที่มีความแตกต่างกัน และระหว่างกลุ่มมีความคล้ายคลึงกันมากที่สุด โดยแต่ละกลุ่มเรียกว่า Cluster จากนั้นจะสุ่ม Cluster ขึ้นมาจำนวนหนึ่งแล้วทำการสำรวจทุก ๆ หน่วยตัวอย่างใน Cluster ที่ตกเป็นตัวอย่างนั้น

จะเห็นว่าแผนแบบการสุ่มตัวอย่างแบบนี้ จะมีลักษณะเดียวกันกับแผนแบบการสุ่มตัวอย่างแบบแบ่งชั้นภูมิแต่จะแตกต่างจากการสุ่มตัวอย่างแบบแบ่งชั้นภูมิตรงที่สุ่ม Cluster ขึ้นมา แล้วจึงสำรวจหน่วยตัวอย่างทุกหน่วยใน Cluster แต่ในการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ จะสุ่มตัวอย่างจากในแต่ละ Stratum โดยต้องสุ่มทุก Stratum

ในที่นี้ใช้เฉพาะกรณีที่แต่ละ Cluster สุ่มด้วยแผนแบบการสุ่มตัวอย่างแบบง่าย

ประโยชน์

1. ประหยัดค่าใช้จ่าย เพราะการเก็บรวบรวมข้อมูลจากหน่วยตัวอย่างที่อยู่ใกล้เคียงกัน ย่อมทำได้ง่าย รวดเร็ว เสียค่าใช้จ่ายน้อย และสะดวกกว่าการเก็บรวบรวมข้อมูลจากหน่วยที่อยู่กระจัดกระจายกันไกลๆ เช่น จะทำการสำรวจครัวเรือนในหมู่บ้าน ซึ่งทำได้โดยการสุ่มตัวอย่างหมู่บ้านขึ้นมาแล้วสำรวจครัวเรือนทุกครัวเรือนในหมู่บ้านนั้น

2. ความสะดวก เพราะปัญหาที่ผู้ที่ทำการสำรวจหรือวิจัยประสบอยู่เสมอคือ การไม่มี Sampling frame หรือมีก็เชื่อถือไม่ค่อยได้ จะสร้างขึ้นใหม่ก็เสียค่าใช้จ่ายสูง เช่น ไม่สามารถหารายชื่อของครัวเรือนทุกครัวเรือนในประเทศไทยได้ แต่ถ้าใช้ Cluster Sampling คือทำการสุ่มหมู่บ้านแทน ซึ่งจะทำให้เราสามารถหากรอบตัวอย่าง ซึ่งเป็นรายชื่อหมู่บ้านทั้งหมดของประเทศได้แล้วจึงทำการสุ่มหาหมู่บ้านตัวอย่างขึ้นมา จากหมู่บ้านที่ตกเป็นตัวอย่าง ก็สำรวจครัวเรือนทุกครัวเรือนในหมู่บ้านที่ตกเป็นตัวอย่างนั้น

ข้อเสีย

แผนแบบการสุ่มตัวอย่างแบบแบ่งกลุ่มนี้ จะทำให้ได้ความแปรปรวนของตัวประมาณค่ามีค่าสูง แต่ในทางปฏิบัติส่วนใหญ่ก็ยังนิยมใช้ Cluster Sampling ถึงแม้ว่าจะมีประสิทธิภาพน้อยกว่าแผนแบบการสุ่มตัวอย่างแบบง่าย ทั้งนี้เป็นเพราะว่าเสียเวลาและค่าใช้จ่ายน้อย

สูตรที่ใช้

$$y_{t.} = \sum_{j=1}^{M_t} y_{tj}$$

$$\bar{y}_{t.} = \frac{y_{t.}}{M_t}$$

การประมาณค่ายอดรวม

$$\hat{Y}_{..} = \frac{N}{n} \sum_{t=1}^{n} y_{t.}$$

$$\hat{V}(\hat{Y}_{..}) = M_0^2 \hat{V}(\bar{y}_{..}) \quad , \quad M_0 = \sum_{t=1}^{N} M_t$$

การประมาณค่าเฉลี่ย ค่าสัดส่วน ค่าอัตราส่วน

ที่ M_t

$$\bar{y}_{..} = \frac{\sum_{t=1}^{n} y_{t.}}{M_0}$$

$$\widehat{V(\bar{y}_{..})} = \frac{N^2}{M_o^2} \cdot \frac{N-n}{Nn} \cdot \sum_{i=1}^{n} \frac{(y_{i.} - \bar{y}_{..})^2}{n-1}$$

$$p = \bar{y}_{..}$$

$$\widehat{V(p)} = \widehat{V(\bar{y}_{..})}$$

ไม่มี M_i

$$\bar{y}_{..} = \frac{\sum_{i=1}^{n} y_{i.}}{\sum_{i=1}^{n} M_i} = r$$

$$\bar{m}_o = \frac{\sum_{i=1}^{n} M_i}{n}$$

$$\widehat{V(\bar{y}_{..})} = \frac{N-n}{Nn} \cdot \frac{1}{\bar{m}_o^2} \cdot \sum_{i=1}^{n} \frac{(y_{i.} - \bar{y}_{..} M_i)^2}{n-1} = \widehat{V(r)}$$

การหาขนาดตัวอย่างเมื่อต้องการประมาณค่าเฉลี่ย

- กรณีใช้ absolute error (d)

$$n = \frac{N Z^2_{\alpha/2} S_c^2}{N\bar{M}_c^2 d^2 + Z^2_{\alpha/2} S_c^2}$$

$$S_c^2 = \frac{\sum_{i=1}^{N} (Y_{i.} - \bar{Y}_{..} M_i)^2}{N-1}$$

$$\bar{M}_c = \frac{\sum_{i=1}^{N} M_i}{N}$$

- กรณีใช้ relative error (r)

$$n = \frac{n_c}{1+\frac{n_c}{N}}$$

$$n_c = \frac{Z^2_{\alpha/2} S_c^2}{\bar{M}_c^2 r^2 \bar{Y}_{..}^2}$$

การหาสัมประสิทธิ์ของความแปรผันของตัวประมาณค่า

$$CV.(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}} \cdot 100$$

<u>การหาช่วงความเชื่อมั่นของตัวประมาณค่า</u>

$CI.(\hat{\theta})$ คือ $\hat{\theta}-Z_{\alpha/2}\sqrt{V(\hat{\theta})} < \theta < \hat{\theta}+Z_{\alpha/2}\sqrt{V(\hat{\theta})}$

y_{ij} ค่าสังเกตที่ j ใน Cluster ที่ i

$y_{i.}$ ค่ายอดรวมของ Cluster ที่ i ในตัวอย่าง

$\bar{y}_{i.}$ ค่าเฉลี่ยของ Cluster ที่ i ในตัวอย่าง

M_i จำนวนประชากรใน Cluster ที่ i

$\hat{Y}_{..}$ ค่าประมาณยอดรวมของประชากร

N จำนวน Cluster ทั้งหมด

n จำนวน Cluster ที่ตกเป็นตัวอย่าง

$V(\hat{Y}_{..})$ ค่าประมาณความแปรปรวนของค่าประมาณยอดรวมของประชากร

$V(\bar{y}_{..})$ ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ย

$\bar{y}_{..}$ ค่าประมาณของประชากรเฉลี่ย

p ค่าประมาณสัดส่วน

$V(p)$ ค่าประมาณความแปรปรวนของสัดส่วน

r ค่าประมาณอัตราส่วน

$V(r)$ ค่าประมาณความแปรปรวนของอัตราส่วน

$Z_{\alpha/2}$ ค่าจากตารางการแจงแจกปกติที่ $\alpha/2$

$Y_{i.}$ ค่ายอดรวมของหน่วยตัวอย่างทั้งหมดใน Cluster ที่ i ในประชากร

d ค่า absolute error

$CV.(\hat{\theta})$ ค่าสัมประสิทธิ์ของความแปรผันของค่าประมาณพารามิเตอร์ θ

$V(\hat{\theta})$ ค่าประมาณความแปรปรวนของค่าประมาณพารามิเตอร์ θ

$\hat{\theta}$ ค่าประมาณพารามิเตอร์ θ

$CI.(\hat{\theta})$ ค่าช่วงความเชื่อมั่นของพารามิเตอร์ θ

6. แผนแบบการสุ่มตัวอย่างหลายขั้นตอน (Multi-stage Sampling)

การสุ่มตัวอย่างแบบหลายขั้นตอน คือ วิธีการสุ่มตัวอย่างที่ทำเป็นขั้น ๆ หลายขั้นด้วยกันโดยที่แต่ละขั้นอาจใช้แผนแบบการสุ่มชนิดใดก็ได้ และการเก็บข้อมูลจะเก็บจากหน่วยตัวอย่างย่อยที่เลือกได้ในขั้นสุดท้าย

วิธีการสุ่มตัวอย่างแบบ 2 ขั้น

จะเหมือนกับ Cluster Sampling คือแบ่งหน่วยตัวอย่างออกเป็นกลุ่ม ซึ่งจะเรียกว่า PSU (Primary Sampling Unit) แล้วจึงสุ่ม PSU ขึ้นมาจำนวนหนึ่งเพื่อเป็นตัวอย่างและแทนที่จะทำการสำรวจทุกๆ หน่วยตัวอย่างใน PSU ที่ตกเป็นตัวอย่างเหมือนใน Cluster Sampling แต่จะสุ่มหน่วยตัวอย่างใน PSU ที่ตกเป็นตัวอย่างเพียงบางหน่วยขึ้นมาเป็นตัวอย่าง และทำการสำรวจเฉพาะหน่วยตัวอย่างที่ตกเป็นตัวอย่างเท่านั้น โดยที่การสุ่มตัวอย่างในแต่ละขั้นจะใช้วิธีการสุ่มตัวอย่างแบบเดียวกัน หรือต่างกันก็ได้

กรอบตัวอย่างที่ใช้

มี 2 อย่างคือ

1. กรอบของกลุ่มที่เรียกว่า Primary Sampling Unit (PSU)

2. กรอบของหน่วยตัวอย่างใน PSU ที่ตกเป็นตัวอย่าง เรียกว่า Secondary Sampling Unit (SSU) (SSU คือ หน่วยตัวอย่างแต่ละหน่วยที่อยู่ใน PSU)

ตัวอย่าง เช่น ต้องการสำรวจผลผลิตข้าวทั่วประเทศ

- SRS Sampling frame ประกอบด้วย รายชื่อครัวเรือนเกษตรทุกครัวเรือนในประเทศ

- Cluster Sampling ถ้าให้หมู่บ้านเป็น Cluster frame จะต้องประกอบด้วยรายชื่อหมู่บ้านทุกหมู่บ้านในประเทศ และจะสุ่มหมู่บ้านที่ตกเป็นตัวอย่าง จากนั้นจะสำรวจทุกครัวเรือนในหมู่บ้านที่ตกเป็นตัวอย่าง

- Two-Stage Sampling จะสุ่มหมู่บ้านจาก Sampling Frame แล้วจึงทำ

การสุ่มครัวเรือนเกษตร ในหมู่บ้านที่ตกเป็นตัวอย่างดังนั้น Sampling frame จะประกอบด้วย

1. รายชื่อหมู่บ้านทุกหมู่บ้าน (PSU)
2. รายชื่อครัวเรือนเกษตรในหมู่บ้านที่ตกเป็นตัวอย่าง (SSU)

จะเห็นว่า Two-Stage Sampling ใช้งบประมาณในการทำ Sampling frame น้อยกว่าแผนแบบการสุ่มตัวอย่างแบบง่ายเพราะว่าในแผนแบบการสุ่มตัวอย่างแบบง่าย จะต้องมีรายชื่อครัวเรือนเกษตรทุกครัวเรือน และระหว่าง Cluster Sampling และ Two-Stage Sampling Cluster Sampling จะใช้งบประมาณน้อยกว่า Two-Stage Sampling แต่ถ้าหน่วยตัวอย่างทุกหน่วยใน Cluster มีความคล้ายคลึงกันการสำรวจทุกหน่วยตัวอย่างใน Cluster ที่ตกเป็นตัวอย่างจะให้ผลเหมือนกับการสุ่มตัวอย่าง SSU ใน PSU ที่ตกเป็นตัวอย่างขึ้นมาสำรวจ ดังนั้นในทางปฏิบัติจึงนิยมใช้ Two-Stage Sampling มากกว่า Cluster Sampling เพราะว่าสามารถได้หน่วยตัวอย่างที่ตกเป็นตัวอย่างกระจายมากกว่า

<u>สูตรที่ใช้</u>

<u>การหาสัมประสิทธิ์ของความแปรผันของตัวประมาณค่า</u>

$$CV.(\hat{\theta}) = \frac{\sqrt{V(\hat{\theta})}}{\hat{\theta}} \cdot 100$$

<u>การหาช่วงความเชื่อมั่นของตัวประมาณค่า</u>

$CI.(\hat{\theta})$ คือ $\hat{\theta}-Z_{\alpha/2}\sqrt{V(\hat{\theta})} < \theta < \hat{\theta}+Z_{\alpha/2}\sqrt{V(\hat{\theta})}$

$CV.(\hat{\theta})$ ค่าสัมประสิทธิ์ของความแปรผันของค่าประมาณพารามิเตอร์ θ

$\widehat{V(\hat{\theta})}$ ค่าประมาณความแปรปรวนของค่าประมาณพารามิเตอร์ ๑

$\hat{\theta}$ ค่าประมาณพารามิเตอร์ ๑

CI.$(\hat{\theta})$ ค่าช่วงความเชื่อมั่นของพารามิเตอร์ ๑

SRS SRS

การประมาณค่ายอดรวม

$$\hat{Y}_{..} = \frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i} y_{ij}$$

$$\widehat{V(\hat{Y}_{..})} = \frac{N^2.N-n.s_{1b}^2}{Nn} + \frac{N}{n}\sum_{i=1}^{n} M_i^2.\frac{M_i-m_i}{M_i}.\frac{s_i^2}{m_i}$$

$$s_i^2 = \frac{\sum_{i=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2}{n-1}$$

$$s_{2}^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2$$

$$s_i^2 = \frac{\sum_{j=1}^{m}(y_{ij}-\bar{y}_{i.})^2}{m-1}$$

$$s_{tb}^2 = \frac{\sum_{i=1}^{n} (y_{i.} - \bar{y})^2}{n-1}$$

การประมาณค่าเฉลี่ย ค่าสัดส่วน ค่าอัตราส่วน

$$\bar{y} = \frac{\hat{Y}_{..}}{N}$$

รู้ M_o

$$M_o = \sum_{i=1}^{N} M_i$$

$$\bar{y}_{..} = \frac{\hat{Y}_{..}}{M_o}$$

$$\hat{V}(\bar{y}_{..}) = \frac{1}{M_o^2} \hat{V}(\hat{Y}_{..})$$

ไม่รู้ M_o

$$\hat{M}_o = \frac{N}{n} \sum_{i=1}^{n} M_i$$

$$\bar{y}_{..} = \frac{\hat{Y}_{..}}{\hat{M}_o} = r$$

$$\widehat{V(\bar{y}_{..})} = \frac{N-n}{Nn}\cdot\frac{1}{\bar{m}_o^2}\cdot\frac{n}{n-1}\sum_{i=1}(y_{i.}-M_i\bar{y}_{..})^2 + \frac{1}{Nn\bar{m}_o^2}\cdot\sum_{i=1}^{n}\frac{M_i^2(M_i-m_i)}{M_i}\cdot\frac{s_i^2}{m_i}$$

n จำนวน PSU ที่ตกเป็นตัวอย่าง

m_i จำนวน SSU ที่ตกเป็นตัวอย่างใน PSU ที่ i ในตัวอย่าง

$\hat{Y}_{..}$ ค่าประมาณยอดรวม

N จำนวน PSU ในประชากร

M_i จำนวน SSU ทั้งหมดใน PSU ที่ i

y_{ij} ค่าสังเกตใน SSU ที่ j PSU ที่ i ในตัวอย่าง

$\widehat{V(\hat{Y}_{..})}$ ค่าประมาณความแปรปรวนของค่าประมาณยอดรวม

$\bar{y}_{i.}$ ค่าประมาณค่าเฉลี่ยของ PSU ที่ i ในตัวอย่าง

$\bar{y}_{..}$ ค่าประมาณค่าเฉลี่ยต่อหน่วย

$\widehat{V(\bar{y}_{..})}$ ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ย

r ค่าประมาณอัตราส่วน

$\bar{y}$ ค่าประมาณค่าเฉลี่ยต่อ PSU

PPS SRS

การประมาณค่ายอดรวม

$$\hat{Y}_{..} = \frac{1}{n}\cdot\sum_{i=1}^{n}\frac{1}{p_i}\cdot\frac{M_i}{m_i}\cdot\sum_{i=1}^{m_i}y_{ij}$$

$$\widehat{V(\hat{Y}_{..})} = \frac{1}{n(n-1)} \cdot \sum_{i=1}^{n} \frac{(y_{i.} - \hat{Y}_{..})^2}{p_{i.}}$$

การประมาณค่าเฉลี่ย ค่าสัดส่วน ค่าอัตราส่วน

รู้ M_0

$$M_0 = \sum_{i=1}^{n} M_i$$

$$\bar{y}_{..} = \frac{\hat{Y}_{..}}{M_0}$$

$$\widehat{V(\bar{y}_{..})} = \frac{1}{M_0^2} \widehat{V(\hat{Y}_{..})}$$

ไม่รู้ M_0

$$\hat{M}_0 = \frac{N}{n} \sum_{i=1}^{n} M_i$$

$$\bar{y}_{..} = \frac{\hat{Y}_{..}}{\hat{M}_0}$$

$$\widehat{V(\bar{y}_{..})} = \frac{1}{X^2 n(n-1)} \sum_{i=1}^{n} \frac{(\hat{Y}_{i.} - r\hat{X}_{i.})^2}{p_i} = \widehat{V(r)}$$

$\bar{y}_{..}$ ค่าประมาณค่าเฉลี่ย

$\hat{Y}_{..}$ ค่าประมาณยอดรวม

n จำนวน PSU ที่ตกเป็นตัวอย่าง

M_i จำนวน SSU ทั้งหมดใน PSU ที่ i

$V(\hat{\bar{y}}_{..})$ ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ย

y_{ij} ค่าสังเกตใน SSU ที่ j PSU ที่ i ในตัวอย่าง

$V(\hat{\hat{Y}}_{..})$ ค่าประมาณความแปรปรวนของค่าประมาณยอดรวม

$y_{i.}$ ค่าประมาณยอดรวมของ PSU ที่ i ในตัวอย่าง

$p_{i.}$ ค่าความน่าจะเป็นที่จะถูกเลือกเป็นตัวอย่างของ PSU ที่ i

r ค่าประมาณอัตราส่วน

X ค่ายอดรวมทั้งหมดของตัวแปร X ของประชากร

$\hat{Y}_{i.}$ ค่าประมาณของยอดรวมของตัวแปร y ของ PSU ที่ iในตัวอย่าง

$\hat{X}_{i.}$ ค่าประมาณของยอดรวมของตัวแปร x ของ PSU ที่ iในตัวอย่าง

บทที่ 3
ขั้นตอนการดำเนินงาน

ขั้นตอนการดำเนินงานแบ่งออกเป็น

1. ศึกษาแผนแบบการสุ่มตัวอย่าง และสูตรการคำนวณค่าประมาณของแต่ละวิธีโดยละเอียดและทำการดัดแปลงสูตรให้สามารถคำนวณค่าได้เร็วขึ้นโดยสามารถคำนวณค่าจากการอ่านข้อมูลเพียงรอบเดียว และศึกษาความเป็นไปได้ในการนำคอมพิวเตอร์มาช่วยในการคำนวณ

2. ศึกษาโปรแกรมสำเร็จรูปทางด้านสถิติที่มีอยู่แล้ว ที่สามารถคำนวณค่าในเรื่องเดียวกัน ซึ่งโปรแกรมที่ศึกษาคือโปรแกรมสำเร็จรูป SPSS/PC$^+$ EPISTAT ซึ่งเป็นโปรแกรมสำเร็จรูปที่ใช้กันอย่างแพร่หลายในปัจจุบัน โดยสังเกตรูปแบบการอ่านข้อมูล สูตรที่ใช้ในการคำนวณ ผลลัพธ์ เปรียบเทียบลักษณะการใช้งาน เพื่อให้ผู้ใช้สามารถเข้าใจและใช้งานโปรแกรมที่จะสร้างได้ง่ายขึ้นเนื่องจากจะใช้พื้นฐานเดียวกัน

ผลจากการศึกษาโปรแกรมดังกล่าวได้ว่า

- โปรแกรมสำเร็จรูป SPSS/PC$^+$ เป็นโปรแกรมสำหรับคำนวณค่าสถิติและสามารถทดสอบสมมติฐานทางด้านสถิติได้ ซึ่งการคำนวณค่าสถิติเบื้องต้นใช้สูตรการคำนวณแบบเดียวกับแผนแบบการสุ่มตัวอย่างแบบง่าย (SRS) ถ้าจะคำนวณค่าจากข้อมูลที่ได้จากการสุ่มตัวอย่างอื่น ๆ ต้องทำการแปลงข้อมูลก่อนถึงจะคำนวณได้ โดยผู้ใช้ต้องรู้สูตรการคำนวณค่าของแผนแบบการสุ่มที่ตนเองใช้เพื่อจะได้ทำการแปลงข้อมูลได้ถูกต้อง และในการใช้โปรแกรมต้องศึกษาลักษณะคำสั่งต่าง ๆ ที่ใช้กับโปรแกรมก่อนจึงจะสามารถใช้โปรแกรมได้

- โปรแกรมสำเร็จรูป EPISTAT เป็นโปรแกรมคำนวณค่าสถิติต่าง ๆ ซึ่งมีเมนูให้เลือกมากมาย มีประสิทธิภาพมากสามารถทำงานได้หลายอย่าง แต่มีเมนูที่เกี่ยวข้องกับเรื่องที่กำลังศึกษาคือเรื่องหาขนาดตัวอย่างและหาตัวอย่าง โดยในโปรแกรมสำเร็จนี้หา

ขนาดตัวอย่างจากสูตรการคำนวณค่าแบบ SRS เพียงอย่างเดียว และค่าช่วงความช่วงมั่นที่ใช้ก็สามารถใช้ได้เพียง 4 ค่าที่กำหนดไว้ในโปรแกรมเท่านั้น ซึ่งในขั้นตอนการกรอกข้อมูลแต่ละขั้นไม่สามารถยกเลิกการทำงานได้ ต้องทำการกรอกให้ครบทุกขั้นตอนจนเสร็จจึงจะสามารถยกเลิกการทำงานได้ และเมื่อกรอกข้อมูลเสร็จโปรแกรมทำการคำนวณและแสดงผลลัพธ์บนจอภาพเรียบร้อยแล้ว ไม่มีข้อความบอกว่าควรจะทำอย่างไรในขั้นตอนต่อไป เพื่อออกหรือทำงานต่อผู้ใช้ต้องมีคู่มือหรือสุ่มกดปุ่มใด ๆ เอาเอง และในส่วนของการหาตัวอย่างจะทำการสุ่มหมายเลขของข้อมูลที่จะตกเป็นตัวอย่างให้ แต่ไม่ได้คำนวณค่าประมาณต่าง ๆ ให้ และในการออกจากโปรแกรมครั้งสุดท้ายไม่มีการคืนสีของตัวอักษรเดิมตอนก่อนเข้าโปรแกรมโดยจะค้างสีของเมนูไว้

3. ออกแบบการอ่านแฟ้มข้อมูลที่ไม่ยุ่งยาก โดยใช้หน่วยความจำอย่างประหยัดที่สุดและกำหนดผลลัพธ์ที่ต้องแสดง ซึ่งได้ออกแบบให้ข้อมูลนำเข้าเป็นข้อมูลดิบที่สำรวจได้ และในส่วนของการอ่านข้อมูลมาเก็บไว้ในหน่วยความจำนั้นจะไม่อ่านข้อมูลมาเก็บไว้ทั้งหมด เพราะถ้าข้อมูลมากจะทำให้เปลืองหน่วยความจำมาก จึงทำการอ่านข้อมูลและนำมาหาผลรวมที่จำเป็นต้องใช้ในการคำนวณค่าต่าง ๆ แล้วจึงเก็บในรูปของผลรวมนั้น ๆ ซึ่งมีไม่กี่จำนวน

4. เขียน Algolithm หรือ Flowchart แสดงขั้นตอนในการคำนวณค่าตามขอบเขตที่กำหนดไว้ได้ดังนี้

<u>เปิดแฟ้มข้อมูล</u>

1. รับชื่อแฟ้มข้อมูล
2. ตรวจสอบว่าเปิดแฟ้มข้อมูลได้หรือไม่
   - ได้ ไปทำข้อ 3
   - ไม่ได้ แสดงข้อผิดพลาดแล้วไปทำข้อ 4
3. อ่านแฟ้มข้อมูล จำนวนตัวแปรในแฟ้มข้อมูลคอลัมน์เริ่มต้น คอลัมน์สิ้นสุด
4. จบการทำงาน กลับสู่เมนูที่เรียก

อ่านแฟ้มข้อมูล

1. รับค่าจำนวนตัวแปรสูงสุด
2. เลือกตัวแปรที่ต้องการแก้ไขตำแหน่งคอลัมน์
   - เลือก ไปทำข้อ 3
   - ไม่เลือก ไปทำข้อ 5
3. รับค่าตำแหน่งคอลัมน์เริ่มต้น คอลัมน์สิ้นสุด
4. ตรวจสอบว่าค่าของตำแหน่งคอลัมน์ถูกต้องหรือไม่
   - ใช่ กลับไปทำข้อ 2
   - ไม่ใช่ แสดงข้อผิดพลาดแล้วกลับไปทำข้อ 2
5. จบการทำงาน กลับสู่เมนูที่เรียก

แสดงแฟ้มข้อมูล

1. เปิดแฟ้มข้อมูลได้หรือไม่
   - ได้ ไปทำข้อ 2
   - ไม่ได้ แสดงข้อผิดพลาดแล้วไปทำข้อ 4
2. แสดงค่าของข้อมูลใน 1 หน้าจอ
3. ตรวจสอบว่าต้องการแสดงค่าของข้อมูลต่อหรือไม่
   - ใช่ กลับไปทำข้อ 2
   - ไม่ใช่ ไปทำข้อ 4
4. จบการทำงาน กลับสู่เมนูที่เรียก

คำนวณค่าสถิติ

1. ตรวจสอบว่าเปิดแฟ้มข้อมูลได้หรือไม่ จำนวนตัวแปรถูกต้องหรือไม่
   - ใช่ ไปทำข้อ 2
   - ไม่ใช่ แสดงข้อผิดพลาดแล้วไปทำข้อ 10
2. รับค่าตัวแปรที่เกี่ยวข้อง ตัวแปรที่ต้องการจะหาค่า

3. ตัวสอบค่าของตัวแปรว่าถูกต้องหรือไม่

   - ใช่ ไปทำข้อ 4

4. อ่านแฟ้มข้อมูล เก็บค่าที่จะต้องคำนวณ

5. รับค่าตัวแปรต่าง ๆ ที่เกี่ยวข้องในการคำนวณหาค่าประมาณ

6. ตรวจสอบค่าของตัวแปรว่าอยู่ในขอบเขตที่ตั้งไว้หรือไม่

   - ใช่ ไปทำข้อ 7

   - ไม่ใช่ แสดงข้อผิดพลาดแล้วไปทำข้อ 10

7. คำนวณค่าสถิติค่าประมาณต่าง ๆ ที่ได้เลือกไว้

8. ตรวจสอบค่าที่คำนวณได้ว่าถูกต้องหรือไม่

   - ใช่ ไปทำข้อ 9

   - ไม่ใช่ แสดงข้อผิดพลาดแล้วไปทำข้อ 10

9. แสดงค่าที่คำนวณได้

10. จบการทำงาน กลับสู่เมนูที่เรียก

<u>หาขนาดตัวอย่าง</u>

1. รับค่าของตัวแปรต่าง ๆ ที่เกี่ยวข้องในการคำนวณ

2. ตรวจสอบว่าค่าอยู่ในขอบเขตที่กำหนดหรือไม่

   - ใช่ ไปทำข้อ 3

   - ไม่ใช่ แสดงข้อผิดพลาดแล้วไปทำข้อ 1

3. คำนวณหาค่าที่ต้องการ

4. ตรวจสอบค่าที่คำนวณได้ และแสดงผลออกหน้าจอ

5. จบการทำงาน กลับสู่เมนูที่เรียก

<u>เลือกแผนแบบการสุ่มตัวอย่าง</u>

1. รับจำนวนขั้นตอนในการสุ่มตัวอย่าง

   - ถ้ามากกว่า 1 ขั้นตอน แสดงว่าเป็น Multi-Stage ไปทำข้อ 2

- ถ้าเท่ากับ 1 ขั้นตอน ไปทำข้อ 2

2. เช็คว่าวนทำจนครบจำนวนขั้นตอนในการสุ่มที่รับมาในข้อ 1 หรือไม่

- ครบ ไปทำข้อ 6

- ไม่ครบ ไปทำข้อ 3

3. ขึ้นถามว่า "ประชากรมีลักษณะแตกต่างกันหรือไม่"

- ใช่ ไปทำข้อ 4

- ไม่ใช่ เป็น SRS ไปทำข้อ 2

4. ขึ้นถามว่า "มีการแบ่งกลุ่มประชากรหรือไม่"

- ใช่ ไปทำข้อ 5

- ไม่ใช่ เป็น PPS ไปทำข้อ 2

5. ขึ้นถามว่า "ในกลุ่มเดียวกันประชากรมีลักษณะเหมือนกันหรือไม่"

- ใช่ เป็น Stratified ไปทำข้อ 2

- ไม่ใช่ เป็น Cluster ไปทำข้อ 2

6. แสดงผลแล้วจบการทำงาน

<u>เปลี่ยนสี</u>

1. เลือกชนิดของรูปแบบที่ต้องการแก้ไข

- เลือก ไปทำข้อ 2

- ไม่เลือก ไปทำข้อ 3

2. แก้ไขสีใหม่หรือไม่

- แก้ ติดตั้งสีใหม่แทนสีเดิมแล้วกลับไปทำข้อ 1

- ไม่แก้ กลับไปทำข้อ 1

3. จบการทำงาน

<u>เลือกเสียง</u>

1. ต้องการแก้ไขเสียงหรือไม่

- ใช่ ไปทำข้อ 2
- ไม่ใช่ ไปทำข้อ 3

2. เลือกมีเสียงหรือไม่มีเสียง
3. ติดตั้งตามที่เลือกในข้อ 2
4. จบการทำงาน

<u>ออกจากโปรแกรม</u>

1. ขึ้นถามว่า "ต้องการเลิกใช้โปรแกรมหรือไม่"
   - ใช่ ไปทำข้อ 2
   - ไม่ใช่ กลับสู่เมนูที่เรียก
2. คืนค่าต่าง ๆ แล้วเลิกการทำงาน
3. ยกเลิกระบบภาษาไทย
4. จบการทำงาน

5. ศึกษาโปรแกรมระบบภาษาไทยต่าง ๆ เพื่อเลือกใช้กับโปรแกรมที่จะเขียน โดยได้ศึกษาระบบภาษาไทย 3 ระบบด้วยกันคือ VTHAI, THAISCT, TSM ซึ่งทำให้ได้รู้รายละเอียดดังนี้

- VTHAI เป็นภาษาไทยที่ถูกพัฒนาโดย สถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย

คุณสมบัติ : ใช้กับจอ VGA และ Monochorm สามารถใช้สีได้ 16 สี แต่ไม่สามารถทำให้ข้อความกระพริบได้ และเวลาออกจากระบบภาษาไทยแล้วจะมีปัญหาที่ว่า cursor จะมีลักษณะผิดปกติคือ กลายเป็นแถบทึบ ไม่เป็นขีดดังเดิม

- THAISCT เป็นภาษาไทยที่จัดทำโดยบริษัท SCT

คุณสมบัติ : ใช้กับจอ VGA เท่านั้น สามารถใช้สีได้เพียง 8 สี สามารถทำให้ข้อความกระพริบได้ แต่ลักษณะตัวอักษรมีขนาดเล็กและไม่สามารถติดตั้งไม่ให้ผู้ใช้ใช้ปุ่มสำหรับเปลี่ยนหน้าจอเป็น graphic mode กับ text mode

- TSM (Thai System Manager) เป็นภาษาไทยที่จัดทำโดยบริษัท ThaiSoft ซึ่งจัดทำขึ้นเพื่อใช้งานกับโปรแกรม Lotus

คุณสมบัติ : ใช้กับจอ VGA เท่านั้น สามารถใช้สีได้เพียง 8 สี สามารถทำให้ข้อความกระพริบได้ ลักษณะตัวอักษรมีขนาดใหญ่ แต่สามารถติดตั้งไม่ให้ผู้ใช้ใช้ปุ่มคีย์บอร์ดสำหรับเปลี่ยนหน้าจอเป็น graphic mode กับ text mode

จากที่ได้ศึกษาคุณสมบัติดังกล่าวแล้วจึงได้เลือกใช้ระบบภาษาไทย TSM เนื่องจากมีขนาดตัวอักษรค่อนข้างใหญ่สามารถอ่านได้สะดวก สามารถทำให้ข้อความกระพริบได้ซึ่งโปรแกรมที่จะสร้างมีการแจ้งข้อผิดพลาดต่าง ๆ ด้วยข้อความกระพริบเพื่อให้สะดุดตา และสามารถติดตั้งไม่ให้ผู้ใช้ใช้ปุ่มคีย์บอร์ดสำหรับเปลี่ยนหน้าจอเป็น graphic mode กับ text mode ซึ่งคู่มือการใช้ภาษาไทย TSM ได้ใส่ไว้ในภาคผนวกแล้ว

6. จาก Alogolithm หรือ Flowchart ที่ได้นำมาเขียนโปรแกรมภาษาซี โดยใช้ภาษาซี version 2.0 ของ Borland International ซึ่งเหตุผลที่เลือกใช้ภาษาซีในการเขียนโปรแกรมนี้เนื่องจากเป็นภาษาที่มีข้อดีในหลาย ๆ ด้านคือ

- ความเร็วในการคำนวณทางคณิตศาสตร์ค่อนข้างสูงเมื่อเทียบกับภาษาอื่นเช่น Foxpro

- สามารถติดต่อกับ file ได้สะดวก มีฟังก์ชั่นสำหรับใช้งานกับ file มากกว่าภาษาอื่นเช่น Pascal , Basic

- การทำงานของโปรแกรมภาษาซีจะมีความเร็วกว่าภาษา Basic หรือ Fortran เพราะภาษาซีเป็นภาษาที่ใกล้เคียงกับภาษาเครื่องมาก แต่ที่ไม่ใช้ภาษาเครื่องเช่น ภาษา Assembly ในการเขียนโปรแกรมนี้เนื่องจากการเขียนด้วยภาษา Assembly มีความยุ่งยากและซับซ้อนมากสำหรับโปรแกรมขนาดใหญ่ อีกทั้ง source code ยังมีความยาวมากกว่าภาษาซีอีกด้วย

- ขนาดของ Execute File ที่ได้จากภาษาซี ของโปรแกรมขนาดใหญ่จะมีขนาดเล็กกว่า Execute File ของภาษาอื่น เช่น Pascal , Basic , Foxpro

- มีความยืดหยุ่นในการใช้คำสั่งต่าง ๆ เช่น สามารถนำคำสั่งหลายคำสั่งมารวมเขียนเป็นประโยคเดียวได้ เช่น If((key=getkey())==F10) ซึ่งในภาษาอื่นเช่น Basic ทำไม่ได้

- เป็นภาษาที่สามารถใช้กับระบบปฏิบัติการอื่น ๆ ได้โดยการแก้ไขโปรแกรมเพียงบางส่วนเท่านั้น เช่น แก้ไขใช้กับ windows , unix เป็นต้น

- เป็นภาษาที่มี Editor ในตัวเองซึ่งในภาษาอื่นเช่น Fortran , Assembly , Cobol ไม่มี ทำให้สะดวกในการ complie และแก้ไขโปรแกรม

- เป็นภาษาที่นิยมใช้งานกันอย่างกว้างขวางทำให้ผู้อื่นสามารถดัดแปลงแก้ไข หรือพัฒนาโปรแกรมนี้ให้ดีและสมบูรณ์ยิ่งขึ้น หรือให้เหมาะกับงานที่ต้องการใช้ได้

7. ทดสอบความถูกต้องของการคำนวณค่า โดยใช้ข้อมูลจากตัวอย่างและแบบฝึกหัดที่เคยเรียน , ตัวอย่างจากหนังสือในเอกสารอ้างอิง

8. นำโปรแกรมที่เสร็จเรียบร้อยแล้วไปให้ผู้ใช้ทดลองใช้เพื่อหาสิ่งผิดพลาดและสิ่งที่ผู้ใช้ไม่เข้าใจ ซึ่งได้ข้อคิดเห็นบางประการจากผู้ทดลองใช้ดังนี้

- ค่าบางค่าควรมีขอบเขตของค่าที่กรอกเข้ามา แต่โปรแกรมไม่ได้ตรวจสอบในส่วนนี้เช่น ค่า relative error ควรให้กรอกเข้ามาได้แค่ค่า 0 ถึง 1 เท่านั้นและถ้ากรอกผิดควรมีข้อความแจ้งเตือนผู้ใช้

- คำถามบางประโยคไม่ชัดเจน ผู้ใช้อ่านแล้วไม่เข้าใจ

- คำศัพท์ทางสถิติบางคำผู้ใช้ไม่ทราบความหมาย ควรมีคำอธิบายไว้ด้วย

- เมนูการคำนวณค่าของแผนแบบการสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) ควรจะมีให้เห็นถึงแม้จะใช้สูตรการคำนวณแบบเดียวกับแผนแบบการสุ่มตัวอย่างแบบง่าย (SRS)

- การแสดงผลลัพธ์จากการคำนวณควรเด่นและน่าสนใจกว่าเดิม

9. ปรับปรุงโปรแกรมให้สมบูรณ์ยิ่งขึ้น และเข้าใจง่าย โดยแก้ข้อผิดพลาดต่าง ๆ,

จัดทำให้มีการกดปุ่ม F1 เพื่อให้คำแนะนำในการใช้โปรแกรมและแนะนำเกี่ยวกับแผนแบบการสุ่มตัวอย่างต่าง ๆ ,แก้ข้อความบนหน้าจอต่าง ๆ ให้ใช้ข้อความที่ผู้ใช้สามารถเข้าใจได้ดีขึ้น , จัดทำคำอธิบายค่าของข้อมูลที่ต้องกรอกบ้างค่าที่บรรทัดเกือบสุดท้ายของจอภาพ

10. จัดทำคู่มือในการใช้งาน เพื่อเป็นแนวทางในการใช้โปรแกรมนี้แก่ผู้ใช้ โดยละเอียด ซึ่งได้ใส่ไว้ในภาคผนวก

บทที่ 4

ผลการวิจัย และวิจารณ์

ผลการคำนวณค่าต่าง ๆ โดยใช้โปรแกรมคำนวณค่าสถิติเบื้องต้นตามการสุ่มตัวอย่าง (Computer Software for Estimation of Parameters in Sampling Methods) ซึ่งใช้ชื่อโปรแกรมว่า EPSM.EXE ให้ค่าที่รวดเร็วกว่าการคำนวณด้วยมือและค่าที่ได้มีความละเอียดถูกต้อง เนื่องจากไม่มีการปัดเศษในขั้นตอนคำนวณจะปัดเศษตอนแสดงผลครั้งสุดท้ายเท่านั้น โดยการแสดงผลครั้งสุดท้ายจะแสดงในรูปทศนิยม 4 ตำแหน่ง ซึ่งผลการคำนวณเขียนแยกเป็นหัวข้อได้ดังนี้

1. SRS

1.1 คำนวณค่าสถิติ

1.1.1 หาค่าเฉลี่ยและค่ายอดรวม (ตัวอย่างข้อมูลจากหนังสือของ สุบงกช จามีกร เทคนิคการสุ่มตัวอย่างกับงานวิจัย หน้า 55 ตัวอย่างที่ 5.1)

โรงพยาบาลแห่งหนึ่งต้องการจะประมาณหนี้สินเฉลี่ยของผู้ป่วยในซึ่งขณะนี้มีจำนวนผู้ป่วยในอยู่ทั้งสิ้น 484 คน ทำการสุ่มตัวอย่างผู้ป่วยมา 9 คนชนิดไม่ใส่คืน เพื่อศึกษาดูหนี้สินจากบัญชีที่ค้างชำระ รายละเอียดที่สุ่มตัวอย่างมีดังนี้

คนไข้	หนี้สินที่ค้างชำระ (บาท)	คนไข้	หนี้สินที่ค้างชำระ (บาท)
1	3350	6	4100
2	3200	7	4500
3	5200	8	4250
4	4300	9	3900
5	4000		

ให้ประมาณ

ก) หนี้ค้างชำระต่อคนของคนไข้ในโดยใช้ช่วงความเชื่อมั่น 95%

ข) ยอดหนี้สินทั้งหมดของคนไข้ในโดยใช้ช่วงความเชื่อมั่น 95%

<u>วิธีทำ</u> การประมาณค่าพารามิเตอร์ทั้ง 2 ทำดังนี้

ก) <u>การประมาณค่าเฉลี่ยของประชากร</u>

ค่าเฉลี่ยของหนี้สินค้างชำระ = 36800/9 = 4088.89

ความแปรปรวนของหนี้สินของคนไข้ที่ตกเป็นตัวอย่าง

$= (1/8) * [153325000-(36800)^2/9]$

$= (1/8) * [153325000-150471111.1]$

$= 356736.1125$

ค่าประมาณความแปรปรวนของค่าเฉลี่ยของหนี้สินค้างชำระ

$= [(484-9)/484 * (356736.1125/9)]$

$= (.9814)(39637.3458)$

$= 38900.0912$

ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยของหนี้สินที่ค้างชำระ

$= 197.2311$

การประมาณหนี้สินเฉลี่ยโดยใช้ช่วงความเชื่อมั่น 95%

$= 4088.89 \pm 1.96(197.2311)$

สรุปผลการประมาณหนี้สินเฉลี่ยได้ดังนี้ หนี้สินเฉลี่ยค้างชำระของคนไข้ในโรงพยาบาลแห่งนี้จะอยู่ในช่วง 4088.89 ± 386.5729 บาท (3702.3171,4475.4629) โดยใช้ช่วงความเชื่อมั่น 95%

ผลจากโปรแกรม

```
                        SRS Mean
แฟ้มข้อมูลปัจจุบัน = "A:\SRS4.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของค่าเฉลี่ย      =       4088.8889
  ความแปรปรวนของค่าเฉลี่ย    =      38900.2876
  C.V. ของค่าเฉลี่ย          =          4.8236%
   95.00% C.I.ของค่าเฉลี่ย   =       3702.3150  ถึง       4475.4627
```

ข) การประมาณยอดรวมของประชากร

ค่าประมาณยอดรวมของหนี้สินค้างชำระ = 484(4088.89)

=1979022.8 บาท

ค่าประมาณความแปรปรวนของค่าประมาณยอดรวมของหนี้สินที่ค้างชำระ

$= (484)^2 * 38900.0912$

$= 9,112,579,764$

95% ช่วงความเชื่อมั่นของหนี้สินที่ค้างชำระ

$= 1979022.8 \pm 1.96(95459.8333)$

สรุปผลการประมาณหนี้สินรวมได้ดังนี้ หนี้สินค้างชำระรวมของคนไข้ในโรงพยาบาลนี้จะอยู่ในช่วง $1979022.8 \pm 187101.2733$ (1791921.527,2166124.073) สำหรับการประมาณโดยใช้ช่วงความเชื่อมั่น 95%

ผลจากโปรแกรม

```
                    SRS Total
แฟ้มข้อมูลปัจจุบัน = "A:\SRS4.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของผลรวม      =    1979022.2222
  ความแปรปรวนของผลรวม    = 9112625771.6049
  C.V. ของผลรวม          =          4.8236%
   95.00% C.I.ของผลรวม   =   1791920.4767  ถึง   2166123.9677
```

1.1.2 หาค่ายอดรวมและค่าสัดส่วน(ตัวอย่างข้อมูลจากหนังสือของ สุบงกช จามีกร เทคนิคการสุ่มตัวอย่างกับงานวิจัย หน้า 62 ตัวอย่างที่ 5.2)

ทำการสุ่มตัวอย่างแบบธรรมดาจากนิสิตชั้นปีที่ 4 จำนวน 100 คน จากนิสิตทั้งหมด 300 คน ว่าจะมีแผนการณ์ที่จะศึกษาต่อในระดับปริญญาโทหรือไม่ และได้สอบถามว่าขณะนี้ทำงานนอกเวลาเพื่อที่จะหาเงินหรือไม่ ได้ข้อมูลดังนี้

นิสิตที่ตกเป็นตัวอย่าง	มีแผนการณ์ที่จะศึกษา	ทำงานขณะเรียน
1	1	1
2	0	0
3	0	1
4	1	0
5	0	0
.	.	.

นิสิตที่ตกเป็นตัวอย่าง	มีแผนการที่จะศึกษา	ทำงานขณะเรียน
.	.	.
.	.	.
98	1	0
99	0	1
100	1	1
	ผลรวม =15	ผลรวม = 65

ให้ประมาณ

ก. จำนวนของผู้ที่มีแผนการจะศึกษาต่อพร้อมทั้งประมาณค่าส่วนเบี่ยงเบนมาตรฐาน

ข. สัดส่วนของผู้ที่จะทำงานขณะเรียนต่อ พร้อมทั้งประมาณค่าสัดส่วนเบี่ยงเบนมาตรฐาน

<u>วิธีทำ</u>

ก. สัดส่วนของนิสิตที่มีแผนการศึกษาต่อ = (15/100) = 0.15

ค่าประมาณของจำนวนนิสิตที่จะศึกษาต่อ = 300*0.15 = 45 คน

ค่าประมาณความแปรปรวนของค่าประมาณของจำนวนนิสิตที่จะศึกษาต่อ

$= 300^2[((300-100)/300) * ((.15)(.85)/(100-1))]$

$= 77.2727$

ค่าประมาณส่วนเบี่ยงเบนมาตรฐาน = $\sqrt{77.2727} = 8.7905$

ผลจากโปรแกรม

```
                    SRS Total
แฟ้มข้อมูลปัจจุบัน = "A:\SRS3.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของผลรวม      =        45.0000
  ความแปรปรวนของผลรวม    =         77.2727
  C.V. ของผลรวม          =         19.5344%
   95.00% C.I.ของผลรวม   =         27.7706  ถึง        62.2294
```

ข.สัดส่วนของนิสิตขณะเรียน = 65/100 = 0.65

ความแปรปรวนของสัดส่วนของนิสิตที่ทำงานขณะเรียน

= (300-100)/300 * (.65)(.35)/(100-1)

= .00154

ค่าประมาณส่วนเบี่ยงเบนมาตรฐาน

= $\sqrt{.00154}$ = .0392

ผลจากโปรแกรม

```
                    SRS Propotion
แฟ้มข้อมูลปัจจุบัน = "A:\SRS3.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของสัดส่วน    =          0.6500
  ความแปรปรวนของสัดส่วน  =          0.0015
  C.V. ของสัดส่วน        =          6.0216%
   95.00% C.I.ของสัดส่วน =          0.5733  ถึง          0.7267
```

1.1.3 ค่าอัตราส่วน

การสุ่มตัวอย่างแบบธรรมดาจำนวน 15 ครัวเรือน จากประชากร 8,000 ครัวเรือน พบว่ามีจำนวนคนเป็นโรคฟันในรอบ 12 เดือนที่แล้ว ดังนี้

ครัวเรือน	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
จำนวนคน(x)	5	6	3	3	2	3	3	3	4	4	3	2	7	4	5
เป็นโรคฟัน(y)	1	0	1	2	0	0	1	1	1	0	1	0	4	3	2

ถ้าทราบว่ามีจำนวนคนทั้งหมด 30,000 คน จงประมาณอัตราส่วนของคนที่เป็นโรคฟัน พร้อมทั้งค่าความแปรปรวน

วิธีทำ ค่าประมาณอัตราส่วน = (17/15) / (57/15)

= 0.2982

ค่าเฉลี่ย y = (1+0+1+2+0+0+...+4+3+2)/15 = 1.1333

ค่าเฉลี่ย x = (5+6+3+3+2+3+...+7+4+5)/15 = 3.8

s_y^2 = [(1-1.1333)^2+(0-1.1333)^2+...+(2-1.1333)^2]/(15-1)

= (1.1872)^2

s_x^2 = [(5-3.8)^2+(6-3.8)^2+...+(5-3.8)^2]/(15-1)

= (1.4243)^2

s_{xy} = [(1-1.1333)(5-3.8)+(0-1.1333)(6-3.8)+...

+(2-1.1333)(5-3.8)]/(15-1)

= 0.8857

ค่าประมาณความแปรปรวน = [1/(3.8)^2]*[(8000-15)/(8000*15)]

*[(1.1872)^2+(0.2982)^2(1.4243)^2

-2(0.2982)(0.8857)]

= 0.0049

ผลจากโปรแกรม

```
                        SRS Ratio
แฟ้มข้อมูลปัจจุบัน = "A:\SRS2.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของอัตราส่วน      =          0.2982
  ความแปรปรวนของอัตราส่วน    =           0.0049
  C.V. ของอัตราส่วน           =          23.4519%
   95.00% C.I.ของอัตราส่วน    =           0.1612  ถึง          0.4353
```

1.2 หาขนาดตัวอย่าง

1.2.1 เพื่อประมาณค่าเฉลี่ยโดยกำหนดค่าผิดพลาดแบบ absolute error

(ตัวอย่างข้อมูลจากหนังสือของสุบงกช จามีกร เทคนิคการสุ่มตัวอย่างกับงานวิจัย หน้า 69 ตัวอย่างที่ 5.3)

ต้องการที่จะคำนวนหาขนาดตัวอย่างในการประมาณหนี้สินค้างชำระเฉลี่ยต่อคนไข้ 1 คน ของโรงพยาบาลเอกชนแห่งหนึ่ง ซึ่งมีคนไข้ในจำนวน 1000 คน และทางโรงพยาบาลอนุญาติให้คนไข้ติดค้างเงินได้ไม่เกินคนละ 2000 บาท การประมาณครั้งนี้กำหนดความคลาดเคลื่อนในการประมาณค่าเฉลี่ยเท่ากับ 60 บาท โดยใช้ช่วงความเชื่อมั่นในการประมาณ ประมาณ 95%

วิธีทำ ในการดำเนินงานต้องทำการกะประมาณค่า σ^2 ในที่นี้จะใช้นิสัยคือการยอมให้คนไข้ค้างชำระ

พิสัย = 2000 บาท

$\ell \approx 2000/4 = 500$

$\ell^2 \approx 500^2 = 250,000$

$n = [(1000)(1.96)^2(250000)] / [(1000*60^2)+(1.96)^2(250000)]$

$= 210.5956$

ดังนั้น ควรสุ่มตัวอย่าง 211 คน

ผลจากโปรแกรม

```
                    N SRS absolute
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
  absolute error (d) :          60.0000
  ส่วนเบี่ยงเบนเฉลี่ย (S) :     500.0000
  ขนาดประชากร (N) :   1000
  ช่วงความเชื่อมั่น(1-alpha)  0.9500      ค่า Z =       1.9600
  ------------------------------
     ขนาดตัวอย่าง (n) = 211
  ------------------------------
```

1.2.2 เพื่อประมาณค่าเฉลี่ยโดยกำหนดค่าผิดพลาดแบบ relative error

(ตัวอย่างข้อมูลจากเอกสารประกอบการเรียนการสอนวิชา Sampling ชุดการสุ่มตัวอย่างแบบ SRS หน้า 6 ตัวอย่าง 4)

ในเรือนเพาะชำแห่งหนึ่ง มีกะบะเพาะชำขนาด 1x430 $(ฟุต)^2$ ซึ่งได้ทดลองเพาะต้นเมเปิ้ลเพื่อดูอัตราการงอกของเมล็ด โดยเพาะต้นเมเปิ้ล 1 หลุม ต่อความยาว 1 ฟุต สมมุติว่าเมื่อสิ้นสุดการทดลอง ได้ค่าเฉลี่ยของการงอกเป็น 19 และ $s^2 = 85.6$

ถามว่า ถ้าต้องการทดลองในคราวต่อไปต้องการประมาณค่าเฉลี่ยไม่ให้มี relative error เกิน 10% ด้วยความน่าจะเป็น 0.95 จะต้องใช้ตัวอย่างเป็นจำนวนเท่าใด

<u>วิธีทำ</u> $n_0 = (1.96)^2 * 85.6/(.10*19)^2$

$= 91.09$

$n = 91.09/(1+(91.09/430))$

$= 75.17$

~ 76

<u>ผลจากโปรแกรม</u>

```
                    N SRS relative
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
  relative error (r) :            0.1000
  ค่าเฉลี่ยของประชากรทั้งหมด :      19.0000
  ส่วนเบี่ยงเบนเฉลี่ย (S) :         9.2520
  ขนาดประชากร (N) :   430
  ช่วงความเชื่อมั่น(1-alpha)  0.9500      ค่า Z =       1.9600
  ------------------------------
     ขนาดตัวอย่าง (n) = 76
  ------------------------------
```

<u>1.2.3 เพื่อประมาณค่าสัดส่วนโดยกำหนดค่าผิดพลาดแบบ absolute error</u>

(ตัวอย่างข้อมูลจากหนังสือของสุบงกช จามีกร เทคนิคการสุ่มตัวอย่างกับงานวิจัย หน้า 72 ตัวอย่างที่ 5.5)

กองกิจการนิสิตต้องการจะประมาณสัดส่วนของนิสิตที่ต้องการพักในหอพักของมหาวิทยาลัยจำนวนนิสิตทั้งหมดมี 2000 คน การกะประมาณนี้จะดำเนินการสุ่มตัวอย่างและใช้ช่วงความเชื่อมั่น 95% โดยประมาณ โดยกำหนดว่า จะยอมให้ช่วงความคลาดเคลื่อนเท่ากับ 5 % เท่านั้น ให้คำนวนหาขนาดตัวอย่างในการสำรวจครั้งนี้

<u>วิธีทำ</u> เนื่องจากไม่มีข้อมูลใด ๆ ที่จะมาช่วยในการประมาณ p และ q ได้ จึงใช้ p=.5 และ q = 1-p = 1-.5 = .5

$$n = [(2000)(1.96)^2(.5)(.5)] / [(2000-1)(0.05)^2+(1.96)^2(.5)(.5)]$$

$$= 322.3955$$

ดังนั้น ควรสุ่มตัวอย่างนิสิต 323 คน

<u>ผลจากโปรแกรม</u>

```
                    N SRS absolute
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
  absolute error (d) :          0.0500
  ค่าสัดส่วนของประชากรทั้งหมด (P) :          0.5000
  ขนาดประชากร (N) :   2000
  ช่วงความเชื่อมั่น(1-alpha)  0.9500      ค่า Z =       1.9600
  ------------------------------
     ขนาดตัวอย่าง (n) = 323
  ------------------------------
```

2. PPS

คำนวณค่าสถิติ (ค่ายอดรวมและค่าเฉลี่ย)

ในอำเภอหนึ่งมี 15 ตำบล และในแต่ละตำบลนั้นมี Supermarkets อยู่ ต้องการประมาณจำนวนสินค้าทั้งหมดใน Stock ของ Supermakets ทั้งหมดในตำบลและจำนวนสินค้าเฉลี่ย จึงทำการสุ่มตัวอย่างตำบลขึ้นมา 6 ตำบล โดยวิธี PPS without replacement มีข้อมูลดังนี้ (โดยจำนวน Supermarket ทั้งหมด = 150)

ตำบล	จำนวน Supermarkets	p_i	จำนวนสินค้า (y_i)	y_i/p_i
1	14	14/150	12	128.5714
2	12	12/150	16	200.0000
3	24	24/150	20	125.0000
4	6	6/150	28	700.0000
5	8	8/150	12	225.0000
6	13	13/150	16	184.6154

วิธีทำ ค่าประมาณจำนวนสินค้าทั้งหมด = (1/6)*(128.5714+200+...+184.6154)

= (1/6)*(1563.1868)

= 260.5145

ค่าประมาณความแปรปรวนของจำนวนสินค้าทั้งหมด

= [1/(6)(6-1)]*[(128.5714-260.5145)^2+(200-260.5145)^2+...

+(18406154-260.5145)^2]

= 239604.6239/[(6)(6-1)]

= 7986.8208

ค่าประมาณจำนวนสินค้าโดยเฉลี่ย = 260.5145/15 = 17.3676

ค่าประมาณความแปรปรวนของจำนวนสินค้าโดยเฉลี่ย

= [1/(15)^2]*[7986.8208] = 35.49698

ผลจากโปรแกรม

```
                          PPS Total
แฟ้มข้อมูลปัจจุบัน = "A:\PPS_R.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของผลรวม      =        260.5504
  ความแปรปรวนของผลรวม    =       7986.4432
  C.V. ของผลรวม          =         34.2993%
   95.00% C.I.ของผลรวม   =         85.3913  ถึง        435.7095
```

```
                          PPS Mean
แฟ้มข้อมูลปัจจุบัน = "A:\PPS_R.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของค่าเฉลี่ย      =        17.3700
  ความแปรปรวนของค่าเฉลี่ย    =         35.4953
  C.V. ของค่าเฉลี่ย          =         34.2993%
   95.00% C.I.ของค่าเฉลี่ย   =         -5.6928  ถึง        29.0473
```

3. Stratified

3.1 คำนวณค่าสถิติ

3.1.1 หาค่าเฉลี่ย (ตัวอย่างข้อมูลจากหนังสือของ สุบงกช จามีกร เทคนิคการสุ่มตัวอย่างกับงานวิจัย หน้า 80 ตัวอย่างที่ 6.1)

บริษัทธุรกิจโฆษณาแห่งหนึ่ง ต้องการจะประมาณจำนวนชั่วโมงเฉลี่ยที่ประชาชนในชุมชนแห่งหนึ่งชมโทรทัศน์ในหนึ่งสัปดาห์ โดยที่ชุมชนแห่งนั้นถูกแบ่งออกเป็น 3 ชั้นภูมิ ตามลักษณะการประกอบอาชีพ ชั้นภูมิ I คือ เมือง ก. ประชาชนส่วนใหญ่ทำงานโรงงานอุตสาหกรรม มีประชากร 155 คน ชั้นภูมิที่ II คือ เมือง ข. ประชากรส่วนใหญ่เป็นเด็ก และคนชรา ประชากรทั้งหมด 62 คน ชั้นภูมิที่ III เป็นเขตชนบทรอบๆ เมือง ก. และเมือง ข. ประชากร 93 คน จากชั้นภูมิ I,II,III ทำการสุ่มตัวอย่างแบบธรรมดา ขนาด $n = 40$, และ $n_1 = 20$, $n_2 = 8$, $n_3 = 12$ ตามลำดับ ได้ข้อมูลดังต่อไปนี้

ตารางที่ 6.1 จำนวนชั่วโมงที่ชมโทรทัศน์ต่อสัปดาห์

ชั้นภูมิ 1 เมือง ก	ชั้นภูมิ 2 เมือง ข	ชั้นภูมิ 3 เขตชนบท
35 28 26 41	27 4 49 10	8 15 21 7
43 29 32 37	15 41 25 30	14 30 20 11
36 25 29 31		12 32 34 24
39 38 40 45		
28 27 35 34		

ให้ประมาณ จำนวนชั่วโมงเฉลี่ยที่ชมโทรทัศน์ต่อสัปดาห์ของชุมชน

จากตารางที่ 6.1 คำนวณรายละเอียดได้ดังนี้

ตารางที่ 6.2 รายละเอียดที่คำนวณจากตารางที่ 6.1

ชั้นภูมิ 1 เมือง ก	ชั้นภูมิ 2 เมือง ข	ชั้นภูมิ 3 เขตชนบท
ขนาดตัวอย่าง = 20	ขนาดตัวอย่าง = 8	ขนาดตัวอย่าง = 12
ค่าเฉลี่ย = 33.9	ค่าเฉลี่ย = 25.125	ค่าเฉลี่ย = 19
ความแปรปรวน = 35.358	ความแปรปรวน = 232.411	ความแปรปรวน = 87.636
ประชากรทั้งหมด = 155	ประชากรทั้งหมด = 62	ประชากรทั้งหมด = 93

วิธีทำ

ค่าประมาณค่าเฉลี่ย

$= 1/310[(155)(33.9) + (62)(25.125) + (93)(19)$

$= 27.675$

ความแปรปรวนของค่าประมาณค่าเฉลี่ย

$= 1/(310)^2[(155)^2(.871)(35.358)/20+(62)^2(.871)(232.411)/8 + (93)^2(.871)(87.636)/12]$

$= 1.97$

ในการประมาณถ้าใช้ช่วงความเชื่อมั่นประมาณ 95%

ค่าประมาณค่าเฉลี่ย $= 27.675 \pm 1.96\sqrt{1.97} = 27.675 \pm (2.751)$

$= (24.924, 30.426)$

ผลจากโปรแกรม

```
                    Stratified Mean
แฟ้มข้อมูลปัจจุบัน = "A:\STRAT2.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของค่าเฉลี่ย      =        27.6750
  ความแปรปรวนของค่าเฉลี่ย    =         1.9695
  C.V. ของค่าเฉลี่ย          =         5.0710%
   95.00% C.I.ของค่าเฉลี่ย   =        24.9243  ถึง        30.4257
```

3.1.2 หาค่ายอดรวม(ตัวอย่างข้อมูลจากหนังสือของสุบงกช จามีกร เทคนิคการสุ่มตัวอย่างกับงานวิจัย หน้า 83 ตัวอย่างที่ 6.2)

จากตัวอย่างที่ 6.1 ให้ท่านประมาณยอดรวมของจำนวนชั่วโมงต่อสัปดาห์ที่ประชาชนในชุมชนชมโทรทัศน์ ให้ใช้ช่วงความเชื่อมั่นในการกะประมาณ ประมาณ 95%

วิธีทำ

ค่าประมาณยอดรวมของจำนวนชั่วโมงต่อสัปดาห์ = 8,579.250 ชั่วโมง

ค่าประมาณความแปรปรวนของค่าประมาณยอดรวมของจำนวนชั่งโมงต่อสัปดาห์

$= (310)^2(1.97)$

$= 189,278.560$ ชั่วโมง

ประมาณโดยใช้ช่วงความเชื่อมั่นประมาณ 95%

ค่าประมาณยอดรวมของจำนวนชั่วโมงต่อสัปดาห์อยู่ในช่วง

$= 8,579.25 \pm 1.96\sqrt{189,278.56}$

$= 8,579.25 \pm 1.96(435.06)$

$= 8,579.25 \pm 852.7207$

$= (7726.5293, 9431.9707)$

สรุปได้ว่า ประชาชนในชุมชนทั้งหมดใช้เวลาดูโทรทัศน์เป็นจำนวนทั้งสิ้น 8,579.25 ชั่วโมงต่อสัปดาห์ มีความคลาดเคลื่อนไม่เกิน 870.12 ชั่วโมง โดยมีความมั่นใจในการกะประมาณ ประมาณ 95%

ผลจากโปรแกรม

```
                    Stratified Total
แฟ้มข้อมูลปัจจุบัน = "A:\STRAT2.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของผลรวม      =       8579.2500
  ความแปรปรวนของผลรวม    =     189270.8086
  C.V. ของผลรวม          =          5.0710%
   95.00% C.I.ของผลรวม   =       7726.5468  ถึง       9431.9532
```

3.1.3 หาค่าสัดส่วน (ตัวอย่างข้อมูลจากหนังสือของสุบงกช จามีกร เทคนิคการสุ่มตัวอย่างกับงานวิจัย หน้า 98 ตัวอย่างที่ 6.7)

บริษัทวิจัยธุรกิจต้องการที่จะประมาณสัดส่วนของครัวเรือนในชุมชนแห่งหนึ่งว่าสนใจรสชาดของนมปรุงแต่งรสใหม่ที่บริษัทรับมาทำโฆษณาหรือไม่ ในชุมชนนั้นถูกแบ่งเป็น 3 ชั้นภูมิ ได้ทำการสุ่มตัวอย่างมาศึกษาจำนวน 40 ครัวเรือน โดยใช้วิธีการจัดสรรขนาดตัวอย่างให้เป็นสัดส่วนกับขนาดของชั้นภูมินั้นการสุ่มตัวอย่างมีรายระเอียดดังนี้

ชั้นภูมิ	ขนาดตัวอย่าง	จำนวนครัวเรือนที่เห็นด้วย	P_h
1	$n_1 = 20$	16	.8
2	$n_2 = 8$	2	.25
3	$n_3 = 12$	6	.50

ให้ประมาณสัดส่วนหรือเปอร์เซ็นต์ของครัวเรือนที่สนใจรสชาดของนมปรุงแต่งยี่ห้อดังกล่าว พร้อมทั้งหาค่าประมาณของความแปรปรวนของค่าประมาณที่ได้มาด้วย

<u>วิธีทำ</u> ค่าประมาณสัดส่วนของครัวเรือนที่สนใจ = (.5*.8)+(.2)(.25)+(.3)(.5)

= .60

ค่าประมาณความแปรปรวนของค่าประมาณสัดส่วน

= $1/(310)^2[115^2*.007+62^2*.024+93^2(.020)]$

= .0045

สัดส่วนของครัวเรือนที่สนใจในรสชาดของนมปรุงแต่งแบบใหม่ คือ .60 และมีช่วงความเชื่อมั่น 95 % ในการกะประมาณเท่ากับ .60 $\pm$ 1.96(.0671) = .60 $\pm$ .1315

<u>ผลจากโปรแกรม</u>

```
                    Stratified Propotion
แฟ้มข้อมูลปัจจุบัน = "A:\STRAT3.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของสัดส่วน      =          0.6000
  ความแปรปรวนของสัดส่วน    =          0.0045
  C.V. ของสัดส่วน           =         11.2402%
   95.00% C.I.ของสัดส่วน    =          0.4678  ถึง          0.7322
```

3.2 หาขนาดตัวอย่าง

3.2.1 เพื่อประมาณค่าเฉลี่ยโดยกำหนดค่าผิดพลาดแบบ absolute error แบบ optimum allocation

(ตัวอย่างข้อมูลจากเอกสารประกอบการเรียนการสอนวิชา Sampling แต่ได้ทำการแปลงโจทย์เพื่อให้เหมาะที่จะคำนวณตามโปรแกรม แบบฝึกหัดชุดที่ 5 หน้า 2 ข้อ 4)

สมาคมหอการค้าไทยต้องการทราบราคาข้าว และผลิตภัณฑ์ทางเกษตรในตลาดต่างประเทศ เพื่อนำมาใช้ในการวางแผนการจำหน่าย เราจึงใช้โทรศัพท์ทางไกลสัมภาษณ์ทูตพาณิชย์เกี่ยวกับราคาข้าวและราคาผลิตภัณฑ์ทางเกษตรโดยแบ่งออกเป็น 3เขต คือ เอเชีย ยุโรปและอเมริกาเหนือ ดังนั้นจึงใช้การสุ่มแบบ stretified random sampliing ปรากฏว่าค่าใช้จ่ายการสัมภาษณ์ต่อครั้ง ความแปรปรวนของราคาสินค้าที่ทูตพาณิชย์รายงานมาและจำนวนทูตพาณิชย์ N_i ในแต่ละเขตเป็นดังนี้

strarum 1 (เอเชีย)	stratum 2 (ยุโรป)	stratum 3 (อเมริกาเหนือ)
C_1 = 9 บาท	C_2 = 25 บาท	C_3 = 36 บาท
$(S_1)^2$ = 2.25	$(S_2)^2$ = 3.24	$(S_3)^2$ = 3.24
N_1 = 112	N_2 = 68	N_3 = 39

ทางสมาคมต้องการกะประมาณการแจ้งราคาโดยเฉลี่ยโดยให้มี d = 0.1 และช่วงความเชื่อมั่น 95 % จะต้องสุ่มตัวอย่างขนาดเท่าใด เพื่อให้เป็นไปตามเงื่อนไขที่กำหนด

วิธีทำ

$$\sum_{h=1}^{L} N_h S_h \sqrt{C_h} = 112*\sqrt{2.25}*\sqrt{9} + 68*\sqrt{3.24}*\sqrt{25} + 39*\sqrt{3.24}*\sqrt{36}$$

$$= 1537.2$$

$$\sum_{h=1}^{L} N_h S_h/\sqrt{C_h} = 112*\sqrt{2.25}/\sqrt{9} + 68*\sqrt{3.24}/\sqrt{25} + 39*\sqrt{3.24}/\sqrt{36}$$

$$= 92.18$$

$$\sum_{h=1}^{L} N_h S_h^2 = 112*2.25 + 68*3.24 + 39*3.24$$

$$= 598.68$$

$$n = (1.96)^2(1537.2)(92.18)/(479.61+1.96^2*598.68)$$

$$= (544351.2472)/(2779.4991)$$

$$= 195.8451$$

$$\approx 196$$

เพราะฉะนั้นจะต้องสุ่มตัวอย่างขนาด 196

ผลจากโปรแกรม

```
                N Stratify absolute
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
 absolute error (d) :            0.1000
 ช่วงความเชื่อมั่น(1-alpha)  0.9500      ค่า Z =       1.9600
 มีกี่ stratum : 3
 ------------------------------

    ขนาดตัวอย่าง (n) = 196

 ------------------------------
```

3.2.2 เพื่อประมาณค่าเฉลี่ยโดยกำหนดค่าผิดพลาดแบบ relative error แบบ Neyman allocation

(ตัวอย่างข้อมูลจากเอกสารประกอบการเรียนการสอนวิชา Sampling แต่ได้ทำการแปลงโจทย์เพื่อให้เหมาะที่จะคำนวณตามโปรแกรม แบบฝึกหัดชุดที่ 5 หน้า 3 ข้อ 7

ถ้าทำการแบ่ง Unit ในประชากรออกเป็น 2 Strata ตามตารางข้างล่างนี้ (ค่า S_h^2 คือ ค่าของ S^2 ใน Stratum ที่ประมาณจากการ Survey ที่ทำมาก่อน)

	N_h	S_h^2
Stratum 1	200,000	324
Stratum 2	600,000	81

ค่าเฉลี่ยประชากรจากการ Survey ที่ทำมาก่อน = 25

สมมติว่า ค่าใช้จ่ายในการทำการสำรวจในแต่ละ Stratum เท่ากัน และกำหนดให้ relative error = 0.002 ช่วงความเชื่อมั่น 95% ถ้าต้องการประมาณค่าเฉลี่ย ให้คำนวณขนาดตัวอย่างที่ต้องใช้

วิธีทำ

$$\sum_{h=1}^{L} N_h S_h = 200,000*\sqrt{324} + 600,000*\sqrt{81} = 9,000,000$$

$$\sum_{h=1}^{L} N_h S_h^2 = 200,000*324 + 600,000*81 = 113,400,000$$

$$n = (1.96)^2(9,000,000)^2/[(800,000)^2(.002*25)^2+(1.96)^2(113,400,000)]$$

$$= 152861.0124 \approx 152862$$

เพราะฉะนั้นขนาดตัวอย่างที่ต้องใช้เท่ากับ 1708

ผลจากโปรแกรม

```
                    N Stratify relative
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
  relative error (r) :              0.0020
  ค่าเฉลี่ยของประชากรทั้งหมด :       25.0000
  ช่วงความเชื่อมั่น(1-alpha)  0.9500      ค่า Z =        1.9600
  มีกี่ stratum : 2
  ------------------------------
     ขนาดตัวอย่าง (n) = 152862
  ------------------------------
```

3.2.3 เพื่อประมาณค่าเฉลี่ยโดยกำหนดค่าผิดพลาดแบบ relative error

(ตัวอย่างข้อมูลจากเอกสารประกอบการเรียนการสอนวิชา Sampling ชุดตัวอย่าง Stratified Sampling หน้า 6)

ในการสำรวจภาวะเศรษฐกิจของครอบครัวในชุมชนแห่งหนึ่ง ซึ่งประกอบด้วยครัวเรือนทั้งสิ้น 1,000 ครัวเรือน จำแนกออกเป็น 3 ชั้นภูมิตามจำนวนสมาชิกของครัวเรือน คือ ครัวเรือนที่มีสมาชิกไม่เกิน 3 คน ถือว่าเป็นขนาดเล็กมีอยู่ทั้งสิ้น 600 ครัวเรือน ครัวเรือนที่มีสมาชิก 4-8 คน ถือว่าเป็นขนาดกลาง มีอยู่ทั้งสิ้น 300 ครัวเรือน และครัวเรือนที่มีสมาชิกตั้งแต่ 9 คน ถือว่าเป็นขนาดใหญ่มีอยู่ทั้งสิ้น 100 ครัวเรือน

จากการสำรวจเบื้องต้นเฉพาะรายการข้อมูลเกี่ยวกับค่าใช้จ่ายด้านอาหาร พบว่าแต่ละชั้นภูมิมีความแปรปรวนภายในเป็น 400 900 และ 2500 บาท ตามลำดับ

จงกำหนดขนาดตัวอย่างเพื่อกะประมาณค่าใช้จ่ายเฉลี่ยด้านอาหารของแต่ละครัวเรือนให้สอดคล้องกับระดับความถูกต้องแม่นยำ d = $\pm$ 3 บาท และระดับความเชื่อถือ 99% ทั้งนี้ให้คำนวณขนาดตัวอย่างไว้สำหรับแผนจัดสรรตัวอย่างทุกแผน สำหรับกรณีของแผนการ

จัดสรรแบบอุตมช (Optimum allocation) ให้ถือว่า

C_1 = 25 บาท C_2 = 30 บาท C_3 = 44 บาท

<u>วิธีทำ</u> คำนวณค่าได้ดังนี้

ชั้นภูมิ	N_h	S_h^2	$N_hS_h^2$	$N_h^2S_h^2$	$\sqrt{C_h}$
1	600	400	240,000	144,000,000	5.00
2	300	900	270,000	81,000,000	5.48
3	100	2500	250,000	25,000,000	6.63
รวม	1000		760,000	250,000,000	

กรณี Proportional allocation

$$n = [(1000)(2.575)^2(760,000)] / [(1000)^2(3)^2+(2.575)^2(760,000)]$$

$$= 358.94 \sim 359$$

<u>ผลจากโปรแกรม</u>

```
                N Stratify absolute
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
 absolute error (d) :            3.0000
 ช่วงความเชื่อมั่น(1-alpha)  0.9900      ค่า Z =       2.5750
 มีกี่ stratum : 3
 ------------------------------
    ขนาดตัวอย่าง (n) = 359
 ------------------------------
```

กรณี Neyman allocation

$n = [(2.575)^2(26,000)^2] / [(1000)^2(3)^2+(2.575)^2(760,000)]$

$= 319.27$

~ 320

ผลจากโปรแกรม

```
                    N Stratify absolute
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
  absolute error (d) :              3.0000
  ช่วงความเชื่อมั่น(1-alpha)  0.9900      ค่า Z =       2.5750
  มีกี่ stratum : 3
  ------------------------------
     ขนาดตัวอย่าง (n) = 320
  ------------------------------
```

กรณี Equal allocation

$n = [(3)(2.575)^2(250,000,000)] / [(1000)^2(3)^2+(2.575)^2(760,000)]$

$= 354.22$

~ 355

ผลจากโปรแกรม

```
                    N Stratify absolute
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
  absolute error (d) :              3.0000
  ช่วงความเชื่อมั่น(1-alpha)  0.9900      ค่า Z =        2.5750
  มีกี่ stratum : 3
  ------------------------------
     ขนาดตัวอย่าง (n) = 355
  ------------------------------
```

4. Cluster

4.1 คำนวณค่าสถิติ

4.1.1 หาค่าเฉลี่ย (ตัวอย่างข้อมูลจากเอกสารประกอบการเรียนการสอนวิชา Sampling แบบฝึกหัดชุดที่ 6 หน้า 2 ข้อ 3)

นักวัดผลต้องการวัดความวิตกกังวลของนักเรียนทั้งหมดในโรงเรียนแห่งหนึ่ง แต่อาจารย์ใหญ่ของโรงเรียนนั้นไม่ยอมให้นักวัดผลทำการสุ่มนักเรียนเป็นรายบุคคลเพื่อแยกออกมาทำการทดสอบ เพราะจะทำให้เกิดความยุ่งยากในการจัดชั่วโมงสอบของวิชาอื่นๆ แต่ยอมให้ทำการทดสอบแก่นักเรียนทั้งหมดในห้องนั้น ถ้าหากว่าตกเป็นตัวอย่าง นักวัดผลจึงทำการสุ่มตัวอย่างมาจำนวน 25 ห้อง แบบ SRS without replacement จากจำนวนทั้งหมด 108 ห้อง แล้วจึงทำการวัดความวิตกกังวล ซึ่งปรากฏผลดังนี้

ห้องเรียน	จำนวนนักเรียน(คน)	คะแนนรวม	ห้องเรียน	จำนวนนักเรียน(คน)	คะแนนรวม
1	31	1,590	14	40	1,980
2	29	1,510	15	38	1,990
3	25	1,490	16	28	1,420
4	35	1,610	17	17	900
5	15	800	18	22	1,080
6	31	1,720	19	41	2,010
7	22	1,310	20	32	1,740
8	27	1,427	21	35	1,750
9	25	1,290	22	19	890
10	19	860	23	29	1,470

ห้องเรียน	จำนวนนักเรียน(คน)	คะแนนรวม	ห้องเรียน	จำนวนนักเรียน(คน)	คะแนนรวม
11	30	1,620	24	18	910
12	18	710	25	31	1,740
13	21	1,140			

จงประมาณคะแนนเฉลี่ยที่นักเรียนทำได้ในแบบทดสอบฉบับนี้ พร้อมทั้งความแปรปรวนของค่าประมาณ

<u>วิธีทำ</u> ค่าประมาณคะแนนเฉลี่ย = 34957/678

= 51.56

ค่าประมาณความแปรปรวนของค่าประมาณคะแนนเฉลี่ย

= [108-25/(108)(25)]*[(259410.712)/(27.12)^2(24))]

= 0.4518

<u>ผลจากโปรแกรม</u>

```
                    Cluster SRS Ratio
แฟ้มข้อมูลปัจจุบัน = "A:\CLUSTER6.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของอัตราส่วน      =        51.5590
  ความแปรปรวนของอัตราส่วน    =         0.4518
  C.V. ของอัตราส่วน          =         1.3036%
   95.00% C.I.ของอัตราส่วน   =        50.2416  ถึง        52.8764
```

4.1.2 หาค่ายอดรวมและค่าอัตราส่วน (ตัวอย่างข้อมูลจากเอกสารประกอบการเรียนการสอนวิชา Sampling แบบฝึกหัดชุดที่ 6 หน้า 1 ข้อ 1,2ก)

โรงงานผลิตเครื่องคิดเลขแห่งหนึ่งต้องการประมาณค่าใช้จ่าย ในการซ่อมเครื่องคิดเลขต่อเดือนแก่บริษัทต่าง ๆ ที่ขายเครื่องคิดเลขไปเขาจึงใช้เทคนิคการสุ่มแบบแบ่งกลุ่มโดยใช้บริษัทแต่ละบริษัทเป็นกลุ่ม (Cluster) แล้วทำการสุ่มตัวอย่างมาจำนวน 20 บริษัทจากจำนวนทั้งหมด 96 บริษัท แบบ SRS without replacement ปรากฏว่าบริษัทที่ตกเป็นตัวอย่างเสียค่าใช้จ่ายในการซ่อมเครื่องคิดเลขดังนี้

บริษัท	จำนวนเครื่องคิดเลข	ค่าใช้จ่ายในการซ่อมของเดือนที่แล้ว (บาท)	บริษัท	จำนวนเครื่องคิดเลข	ค่าใช้จ่ายในการซ่อมของเดือนที่แล้ว (บาท)
1	3	50	11	8	140
2	7	110	12	6	130
3	11	230	13	3	70
4	9	140	14	2	50
5	2	60	15	1	10
6	12	280	16	4	60
7	14	240	17	12	280
8	3	45	18	6	150
9	5	60	19	5	110
10	9	230	20	8	120

ก. จงประมาณค่าใช้จ่ายในการซ่อมทั้งหมดที่บริษัททั้ง 96 แห่งต้องจ่ายไปพร้อมทั้งความแปรปรวนของค่าประมาณ

ข. จงประมาณค่าซ่อมเฉลี่ยต่อเครื่องของเดือนที่แล้ว พร้อมทั้งความแปรปรวนของ

ค่าประมาณ

วิธีทำ

ก. ค่าประมาณค่าใช้จ่ายในการซ่อมทั้งหมด = (96/20)*2565 = 12,312

ค่าประมาณค่าเฉลี่ย = 2565/20 = 128.25

ค่าประมาณความแปรปรวนของค่าประมาณค่าใช้จ่ายในการซ่อมทั้งหมด

$= [(96)^2*(96-20)/((96)(20)(19))]*[(50-128.25)^2+(110-128.25)^2$

$+...+(120-128.25)^2]$

= 2520264

ผลจากโปรแกรม

```
                    Cluster SRS Total
แฟ้มข้อมูลปัจจุบัน = "A:\CLUSTER1.DAT"
   ค่าที่คำนวนได้
 ประมาณค่าของผลรวม      =      12312.0000
 ความแปรปรวนของผลรวม    =    2520264.0000
 C.V. ของผลรวม          =         12.8942%
  95.00% C.I.ของผลรวม   =       9200.4335  ถึง      15423.5665
```

ข. ค่าประมาณค่าซ่อมเฉลี่ยต่อเครื่อง = 2565/130 = 19.7308

ค่าประมาณความแปรปรวนของค่าประมาณค่าซ่อมเฉลี่ยต่อเครื่อง

$= [(96-20)/(96)(20)]*[1/(6.5)^2(19)]*[(50-19.7308(3))^2+$

$(110-19.7308(7))^2+...+(120-19.708(8))^2]$

= 0.7922

ผลจากโปรแกรม

```
                    Cluster SRS Ratio
แฟ้มข้อมูลปัจจุบัน = "A:\CLUST10.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของอัตราส่วน      =        19.7308
  ความแปรปรวนของอัตราส่วน    =         0.7922
  C.V. ของอัตราส่วน          =         4.5110%
   95.00% C.I.ของอัตราส่วน   =       17.9863  ถึง       21.4753
```

4.2 หาขนาดตัวอย่าง

จาก 4.1.2 ถ้าโรงงานต้องการทำการประมาณค่าใช้จ่ายในการซ่อมเฉลี่ยต่อเครื่องในเดือนต่อไป โดยให้ค่าประมาณแตกต่างจากค่าพารามิเตอร์น้อยกว่า 2 บาทจะต้องสุ่มตัวอย่างให้มีจำนวน Cluster เท่าใด ที่ระดับความเชื่อมั่น 95%

วิธีทำ

$$n = [(96)(1.96)^2(845.56)]/[(96)(6.5)^2(2)^2+(1.96)^2(845.56)]$$

$$= 16.0144 \sim 17$$

ผลจากโปรแกรม

```
                    N Cluster absolute
แฟ้มข้อมูลปัจจุบัน = "ยังไม่ได้เปิดแฟ้มข้อมูล"
  absolute error (d) :              2.0000
  ส่วนเบี่ยงเบนเฉลี่ย (Sc) :        29.0785
  ช่วงความเชื่อมั่น(1-alpha)  0.9500      ค่า Z =        1.9600
  ขนาดcluster (N) :  96
  ------------------------------
     ขนาดตัวอย่าง (n) = 17
  ------------------------------
```

5. Multi-Stage

5.1 SRS SRS (ค่าเฉลี่ยและค่ายอดรวม)

ตัวอย่างข้อมูลจากเอกสารประกอบการเรียนการสอนวิชา Sampling บทที่ 7 หน้า 7-16 ตัวอย่าง 2

ในพื้นที่แห่งหนึ่งมีโรงงานอุตสาหกรรมทอผ้าจำนวน 90 โรงงานต้องการจะประมาณเวลาเฉลี่ยที่เครื่องจักรต้องหยุดซ่อมเมื่อเดือนที่แล้ว เนื่องจากโรงงานอยู่อย่างกระจัดกระจายเขาจึงใช้ Cluster sampling โดยให้โรงงานแต่ละแห่งเป็น Cluster ในแต่ละโรงงานจะมีเครื่องจักรหลายเครื่อง และเป็นการเสียเวลามากที่จะตรวจสอบการใช้เวลาซ่อมเครื่องจักรแต่ละเครื่อง เขามีเวลาและงบประมาณพอที่จะสำรวจโดยสุ่มตัวอย่าง โรงงานขึ้นมา 10 โรงงาน และจากโรงงานที่ตกเป็นตัวอย่างจะสุ่มเครื่องจักรเพียง 20% ของแต่ละโรงงานดังข้อมูลดังนี้

โรงงานที่ตกเป็นตัวอย่าง	จำนวนจักรเย็บผ้า		เวลาที่ส่งซ่อม	y_i	s_i^2
	M_i ทั้งหมด	m_i ตัวอย่าง	(y_{ij})		
1	50	10	5,7,9,0,11,2,8,4,3,5	5.40	11.38
2	65	13	4,3,7,2,11,0,1,9,4,3,2,1,5	4.00	10.67
3	45	9	5,6,4,11,12,0,1,8,4	5.67	16.75
4	48	10	6,4,0,1,0,9,8,4,6,10	4.80	13.29
5	52	10	11,4,3,1,0,2,8,6,5,3	4.30	11.12
6	58	12	12,11,3,4,2,0,0,1,4,3,2,4	3.83	14.88
7	42	8	3,7,6,7,8,4,3,2	5.00	5.14
8	66	13	3,6,4,3,2,2,8,4,0,4,5,6,3	3.85	4.31

โรงงานที่ตก เป็นตัวอย่าง	จำนวนจักรเย็บผ้า		เวลาที่ส่งซ่อม	y_i	s_i^2
	M_i ทั้งหมด	m_i ตัวอย่าง	(y_{ij})		
9	40	8	6,4,7,3,9,1,4,5	4.88	6.13
10	56	11	6,7,5,10,11,2,1,4,0,5,4	5.00	11.80

ก. จงประมาณเวลาเฉลี่ยที่ใช้ซ่อมเครื่องจักรต่อเครื่อง ถ้าหากว่ามีเครื่องจักรทั้งหมด 4,500 เครื่อง

ข. จงประมาณจำนวนชั่วโมงทั้งหมดที่เครื่องจักรต้องหยุดซ่อมเมื่อเดือนที่แล้ว

ค. จงประมาณค่าเฉลี่ยของจำนวนชั่วโมงต่อเครื่องจักร 1 เครื่องที่ต้องหยุดซ่อมในเดือนที่แล้ว ถ้าไม่ทราบจำนวนเครื่องจักรทั้งหมด

ง. จงประมาณสัดส่วนของจำนวนเครื่องจักรที่จะต้องหยุดซ่อมใหญ่

พร้อมทั้งประมาณความแปรปรวนของค่าประมาณ

<u>วิธีทำ</u>

ก.ค่าประมาณเวลาที่ใช้ในการซ่อมเครื่องจักรต่อเครื่อง

= [(90/10)*(2400.59)]/4500

= 21,605.31/4500

= 4.8 ชั่วโมง

s_{1b}^2 = [583,198.67-(240059)^2/10]/9

= 768.3822

ค่าประมาณความแปรปรวน

$= [1/(4500)^2]*[90^2*((90-10)/90)*(768.3822/10)+(90/10)*21,987.508]$

$= 0.0371$

<u>ผลจากโปรแกรม</u>

```
                    Multi-stage SRS-SRS Mean
แฟ้มข้อมูลปัจจุบัน = "A:\MULTI3.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของค่าเฉลี่ย      =          4.8004
  ความแปรปรวนของค่าเฉลี่ย    =          0.0371
  C.V. ของค่าเฉลี่ย          =          4.0119%
   95.00% C.I.ของค่าเฉลี่ย   =          4.4229  ถึง          5.1778
```

ข.ค่าประมาณเวลาทั้งหมดที่เครื่องจักรต้องหยุดซ่อม

$= 90/100*(2400.59)$

$= 21,605.31$ ชั่วโมง

ค่าประมาณความแปรปรวน

$= (4500)^2(0.0371)$

$= 751,275$ ชั่วโมง^2

ผลจากโปรแกรม

```
                Multi-stage SRS-SRS Total
แฟ้มข้อมูลปัจจุบัน = "A:\MULTI3.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของผลรวม      =      21601.6154
  ความแปรปรวนของผลรวม    =     751063.2037
  C.V. ของผลรวม          =          4.0119%
   95.00% C.I.ของผลรวม   =     19903.0029  ถึง      23300.2279
```

ค.ค่าประมาณค่าเฉลี่ยของจำนวนชั่วโมงต่อเครื่องจักร 1 เครื่องที่ต้องหยุดซ่อมในเดือนที่แล้ว

$= [(90/10)*2400.59]/[(90/10)*522]$

$= 21605.31/4698$

$= 4.6$ ชั่วโมง/เครื่อง

$\sum_{t=1}^{n} M_t \bar{y}_{t.} = 126,530.87$

$\sum_{t=1}^{n} (y_{t.} - M_t \bar{y}_{..})^2 = [583,19867 + (460)^2(27,978)] - [2(4.6)(126,530.87)]$

$= 11,129.15$

ค่าประมาณความแปรปรวน

$= [(90-10)/(90*10)]*(1/52.2^2)*(11129.15/9) + [1/(10*90*52.2^2)]*21987.508$

$= 0.0493$

ผลจากโปรแกรม

```
                Multi-stage SRS-SRS Mean
แฟ้มข้อมูลปัจจุบัน = "A:\MULTI3.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของค่าเฉลี่ย      =          4.5980
  ความแปรปรวนของค่าเฉลี่ย    =          0.0496
  C.V. ของค่าเฉลี่ย          =          4.8442%
   95.00% C.I.ของค่าเฉลี่ย   =          4.1615  ถึง          5.0346
```

5.2 PPS SRS (ค่าเฉลี่ยและค่ายอดรวม)

ในอำเภอหนึ่งแบ่งออกเป็น 15 ตำบลและในแต่ละตำบลนั้นมี Supermarkets ซึ่งจำนวน Supermarket ในทุกตำบล = 150 ต้องการประมาณจำนวนสินค้าทั้งหมดใน Stock ของ Supermarkets จึงทำการสุ่มตัวอย่างตำบลขึ้นมา 6 ตำบล โดยวิธี PPS without replacement และจากตำบลที่ตกเป็นตัวอย่างสุ่ม Supermarket ตัวอย่างขึ้นมา 4 แห่ง จากแต่ละตำบล แบบ SRS without replacement มีข้อมูลดังนี้

ตำบล	จำนวน Supermarket (M_i)	p_i	จำนวนสินค้าใน Supermarket			
			1	2	3	4
1	14	14/150	2	3	2	5
2	12	12/150	2	5	5	4
3	24	24/150	1	4	8	7

ตำบล	จำนวน Supermarket (M_i)	p_i	จำนวนสินค้าใน Supermarket			
			1	2	3	4
4	6	6/150	4	8	9	7
5	8	8/150	1	1	6	4
6	13	13/150	4	3	3	6

<u>วิธีทำ</u> ค่าประมาณจำนวนสินค้าทั้งหมด = 650

ค่าประมาณความแปรปรวน = 8500

ค่าประมาณจำนวนสินค้าโดยเฉลี่ย = 650/150

= 4.333

ค่าประมาณความแปรปรวน = $[1/(150)^2]*[8500]$

= 0.3778

<u>ผลจากโปรแกรม</u>

```
                Multi-stage PPS-SRS Total
แฟ้มข้อมูลปัจจุบัน = "A:\MULTI20.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของผลรวม      =        650.0000
  ความแปรปรวนของผลรวม    =       8500.0000
  C.V. ของผลรวม           =         14.1839%
   95.00% C.I.ของผลรวม    =       469.2969  ถึง       830.7031
```

```
                    Multi-stage PPS-SRS Mean
แฟ้มข้อมูลปัจจุบัน = "A:\MULTI20.DAT"
    ค่าที่คำนวนได้
  ประมาณค่าของค่าเฉลี่ย      =          4.3333
  ความแปรปรวนของค่าเฉลี่ย    =           0.3778
  C.V. ของค่าเฉลี่ย          =          14.1839%
   95.00% C.I.ของค่าเฉลี่ย   =           3.1286  ถึง          5.5380
```

บทที่ 5

สรุปผลการศึกษาวิจัยและข้อเสนอแนะ

สรุปผล

จากการที่ได้จัดทำโปรแกรมสำหรับคำนวณค่าสถิติเบื้องต้นตามการสุ่มตัวอย่าง ดังขอบเขตที่ได้กำหนดไว้นั้น จะเห็นได้ว่าโปรแกรมได้ช่วยอำนวยความสะดวกในการคำนวณค่าสถิติได้อย่างมาก และการใช้งานก็เป็นไปโดยง่าย โดยที่ผู้ใช้ไม่ต้องมีความรู้พื้นฐานเกี่ยวกับสูตรในการคำนวณค่านั้น ๆ มาก่อน เพียงแต่ผู้ใช้มีข้อมูลดิบที่เก็บมาก็สามารถหาค่าต่าง ๆ ได้โดยง่าย และไม่ต้องเสียเวลาในการศึกษาโปรแกรมเท่าใดนัก เนื่องจากได้ออกแบบโปรแกรมเป็นแบบหัวข้อให้เลือกใช้โดยไม่มีการกรอกคำสั่งใด ๆ จึงทำให้สามารถใช้งานได้เลย โดยเฉพาะได้พัฒนาโปรแกรมให้เป็นระบบภาษาไทยจึงทำให้การใช้งานง่ายขึ้น

อนึ่งปัญหาพิเศษนี้ ได้พยายามสร้างขอบเขตในการคำนวณค่าต่าง ๆ ให้ครอบคลุมถึงเนื้อหา และค่าสถิติที่ต้องการคำนวนในการสุ่มตัวอย่างมากเพียงพอที่นักวิจัยส่วนใหญ่นิยมใช้

ข้อเสนอแนะ

เนื่องจากในการทำงานวิจัยปัญหานิเศษนี้มีช่วงเวลาในการทำ 1 เทอม ทำให้ไม่สามารถพัฒนาโปรแกรมให้มีขอบเขตการทำงานได้กว้างกว่านี้ จึงจะขอเสนอแนวทางในการพัฒนาโปรแกรมให้สามารถทำงานได้กว้างขวางกว่าเดิม เพื่อผู้ที่สนใจจะได้พัฒนาและนำโปรแกรมไปใช้ประโยชน์ได้มากขึ้น ซึ่งมีข้อเสนอแนะดังนี้

1. ควรจะคำนวณค่าประมาณของหลาย ๆ ตัวแปรได้ ในการคำนวณเพียงครั้งเดียว เพื่อความสะดวกของผู้ใช้ในกรณีที่ต้องการหาค่าประมาณต่าง ๆ ของหลาย ๆ ตัวแปรที่มีอยู่ในแฟ้มข้อมูลเดียวกัน

2. ในแผนแบบการสุ่มตัวอย่างแบบหลายขั้นตอน ควรจะสุ่มตัวอย่างได้มากกว่าสองขั้นตอนและในแต่ละขั้นตอนสามารถใช้แผนแบบการสุ่มได้กว้างกว่านี้คือ สามารถสุ่มเป็นแบบ Stratified หรือ Cluster ได้

3. ในแผนแบบการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ ในแต่ละ Stratum ควรจะสามารถใช้แผนแบบในการสุ่มได้กว้างกว่านี้

4. ในแผนแบบการสุ่มตัวอย่างแบบแบ่งกลุ่ม ในแต่ละ Cluster ควรจะสามารถใช้แผนแบบในการสุ่มได้กว้างกว่านี้

5. ควรจะพัฒนาให้มีส่วนของการทดสอบสมมติฐานต่าง ๆ ทางด้านสถิติเช่นt-test, z-test , F-test และการทดสอบความแปรปรวน ฯลฯ

อย่างไรก็ดี รายละเอียดของปัญหาพิเศษนี้ก็ยังใช้เป็นพื้นฐานและแนวทางให้ผู้สนใจที่จะพัฒนาโปรแกรมทางด้านสถิติต่าง ๆ พัฒนาโปรแกรมโดยคำนึงถึงว่าค่าสถิติเบื้องต้นที่จะนำมาวิเคราะห์หรือทดสอบสมมติฐานนั้น ถ้าข้อมูลใช้แผนแบบการสุ่มตัวอย่างที่แตกต่างกันต้องใช้สูตรในการคำนวณค่าต่าง ๆ นั้นแตกต่างกันด้วย

ภาคผนวก

คู่มือการใช้โปรแกรม

EPSM.EXE

คู่มือการใช้โปรแกรม EPSM.EXE

เป็นโปรแกรมสำหรับช่วยผู้ทำวิจัยหรือผู้ใช้งานที่มีข้อมูลจากการสุ่มตัวอย่าง แล้วต้องการคำนวณหาค่าสถิติเบื้องต้นต่าง ๆ โปรแกรมนี้จะสามารถช่วยให้การคำนวณหาค่าต่าง ๆ เป็นไปอย่างสะดวกและรวดเร็วโดยไม่ต้องเสียเวลาศึกษาสูตรการคำนวณค่าต่าง ๆ ของทฤษฎีการสุ่มตัวอย่างซึ่งมีอยู่มากมายและค่อนข้างยุ่งยาก และยังสามารถช่วยหาแผนแบบการสุ่มตัวอย่างและขนาดตัวอย่างที่เหมาะสมสำหรับผู้ที่ทราบลักษณะของประชากรแต่ไม่ทราบว่าควรจะสุ่มตัวอย่างโดยใช้แผนแบบการสุ่มตัวอย่างแบบใด ขนาดเท่าไรจึงจะเหมาะสมและให้ค่าความคลาดเคลื่อนของค่าประมาณที่ได้น้อยที่สุด

โปรแกรมที่สำคัญในการใช้งาน

1. โปรแกรมคำนวณ EPSM.EXE
2. โปรแกรมภาษาไทย TSM

โปรแกรมที่ควรจะมีประกอบการใช้งาน

1. โปรแกรมเกี่ยวกับคำแนะนำ - EPSM.HCL
   - EPSM.HCN
   - EPSM.HCY
   - EPSM.HGT
   - EPSM.HLF
   - EPSM.HLP
   - EPSM.HMU
   - EPSM.HSC
   - EPSM.HSR
2. โปรแกรมตัวอย่างการใช้งาน DEMO.EXE

ลักษณะของแฟ้มข้อมูล

กรอกข้อมูลโดยใช้ Editor ต่าง ๆ เช่น SIDEKICK QEDIT CW ฯลฯ โดยค่าของข้อมูลตัวแปรเดียวกันต้องอยู่ในช่วงของคอลัมน์ที่ผู้ใช้กำหนด ซึ่งอาจจะกรอกแต่ละค่าของข้อมูลแยกกันด้วยการเว้นวรรคหรือไม่เว้นก็ได้ แต่ถ้าเว้นวรรคจะสะดวกในขณะที่อ่านข้อมูลขึ้นมาคือไม่ต้องไปแก้คอลัมน์เริ่มต้นและคอลัมน์สิ้นสุดให้กับโปรแกรม และในการกรอกข้อมูลแต่ละชุดให้กรอกภายในหนึ่งบรรทัด(ใช้การขึ้นบรรทัดใหม่เป็นการแบ่งแต่ละชุดข้อมูล) ตัวอย่างดังรูป ก-1

```
 1  82 24  383  90 0
 2  50 18  638  45 0
 3 183  9  370 165 0
 4 182  -    -  120 0
 5  47 34 1326  47 0
 6  85 34  784  62 0
 7  38 32 1500  32 0
 8  23 22 1500  15 1
 9  38 28 1500  28 0
10  45 35 1500  40 1
```

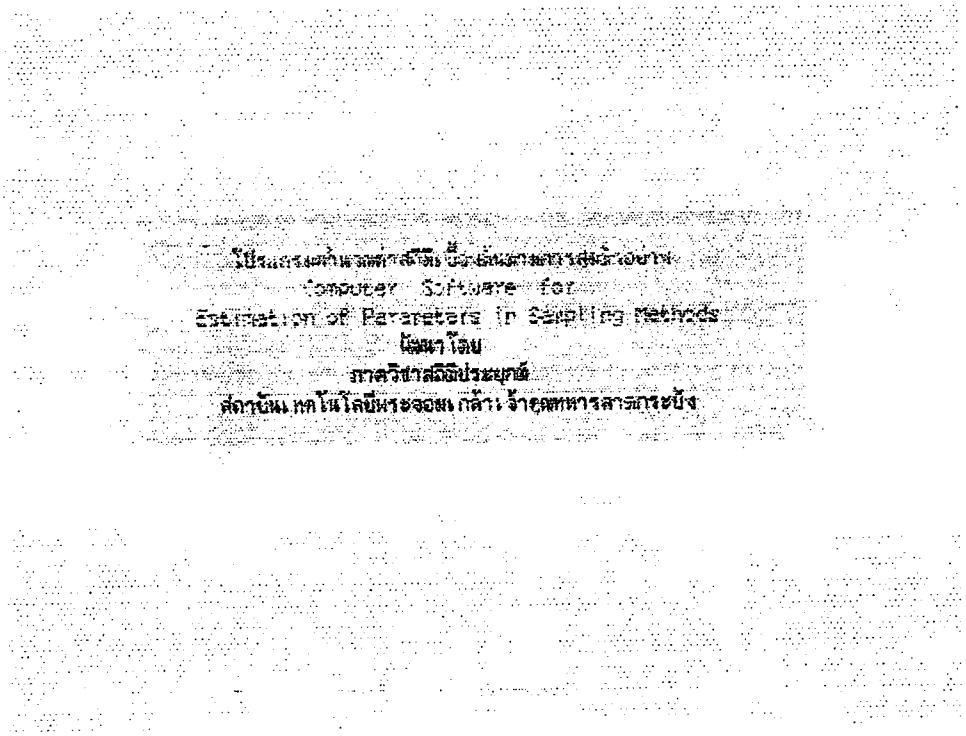
รูป ก-1 แสดงแฟ้มข้อมูลของแผนแบบการสุ่มตัวอย่างแบบง่าย

วิธีใช้งาน

เข้าสู่การทำงานของโปรแกรมโดยการคีย์คำว่า "EPSM" ดังตัวอย่างในรูป ก-2 (โดยสมมติว่าถ้าขณะนั้นเราอยู่ที่ไดร์ฟ A) และเมื่อเข้าสู่โปรแกรมแล้วจะเห็น Title ของโปรแกรมดังรูป ก-3 ให้กดปุ่มใด ๆ เพื่อทำงานต่อ

```
A:\>EPSM
```

รูป ก-2 แสดงคำสั่งในการเข้าสู่โปรแกรม EPSM



รูป ก-3 แสดง Title ของโปรแกรม EPSM

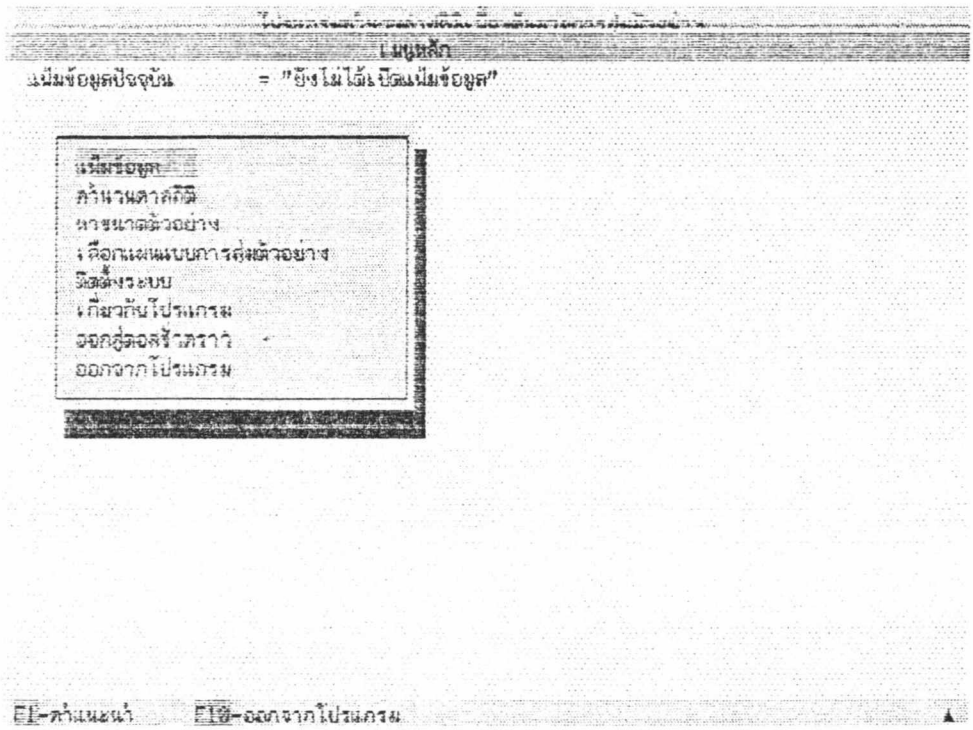
โปรแกรมจะแสดงเมนูหลักให้เลือก 8 เมนู (ดังรูป ก-4) คือ

1. แฟ้มข้อมูล
2. คำนวณค่าสถิติ
3. หาขนาดตัวอย่าง
4. เลือกแผนแบบการสุ่มตัวอย่าง
5. ติดตั้งระบบ

6. เกี่ยวกับโปรแกรม

7. ออกสู่ดอสชั่วคราว

8. ออกจากโปรแกรม

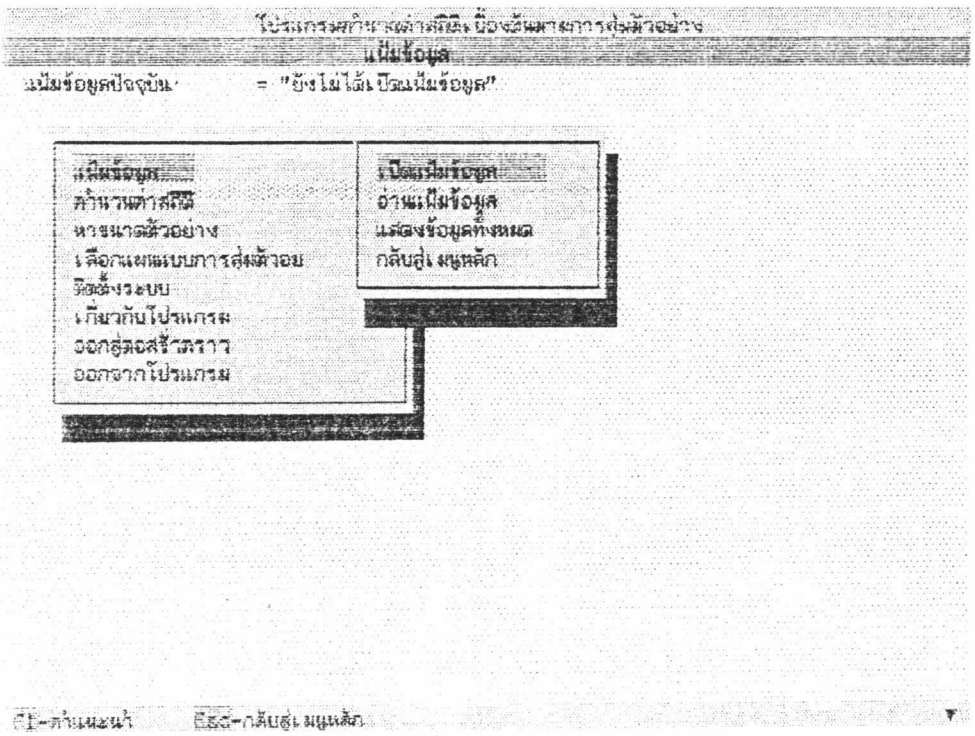


รูป ก-4 แสดงเมนูหลักของโปรแกรม

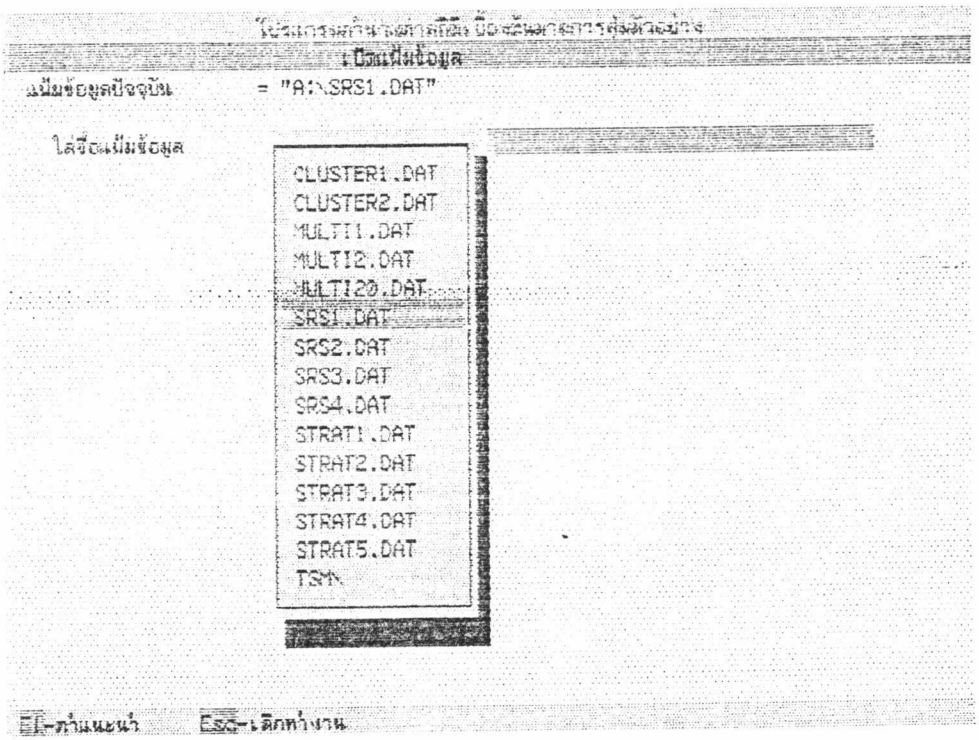
ซึ่งแต่ละเมนูมีรายละเอียดการทำงานดังนี้

1. แฟ้มข้อมูล (ดังในรูป ก-5) มีเมนูย่อยคือ

1.1 เปิดแฟ้มข้อมูล สำหรับบอกรายละเอียดเกี่ยวกับแฟ้มข้อมูลว่าอยู่ที่ Drive ใด Sub directory อะไร และชื่อแฟ้มข้อมูล ดังรูป ก-6

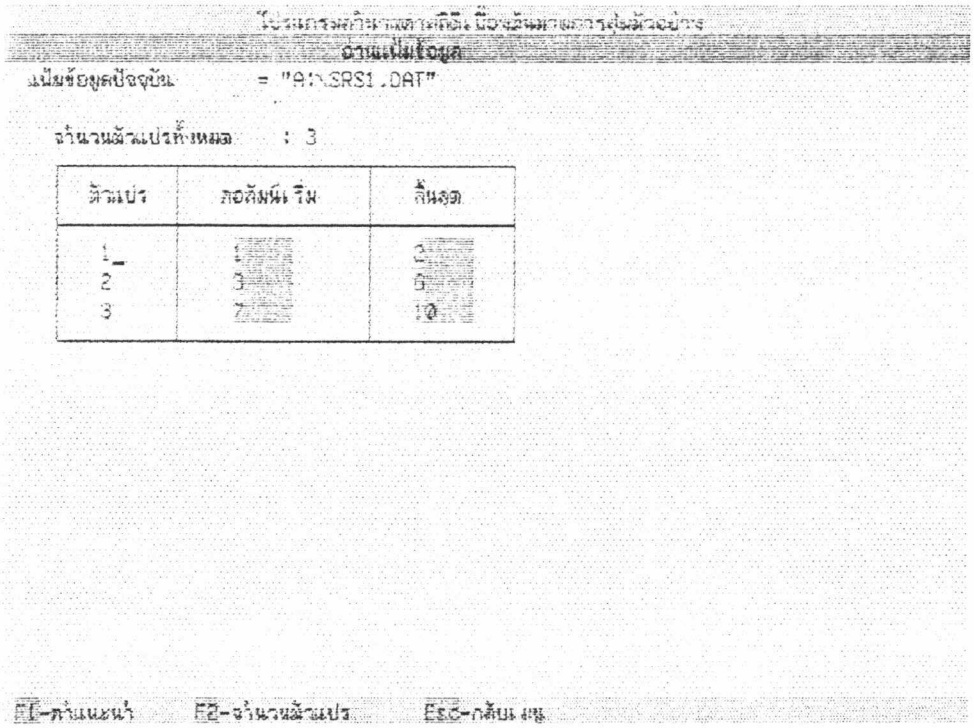


รูป ก-5 แสดงเมนูแฟ้มข้อมูล



รูป ก-6 แสดงตัวอย่างการเปิดแฟ้มข้อมูล

1.2 อ่านแฟ้มข้อมูล สำหรับอ่านแฟ้มข้อมูลตามชื่อไฟล์ที่ระบุในเมนูเปิดแฟ้มข้อมูล โดยโปรแกรมจะอ่านข้อมูล และใช้การเว้นวรรคของแฟ้มข้อมูลแบ่งข้อมูลออกเป็นแต่ละตัวแปร และใช้การขึ้นบรรทัดใหม่ของแฟ้มข้อมูลเป็นการบอกว่าจบข้อมูลหนึ่งชุด จากนั้นโปรแกรมก็จะแสดงรายละเอียดที่อ่านได้ เช่น มีกี่ตัวแปร แต่ละตัวแปรเริ่มจากคอลัมน์ที่เท่าไหร่ และจบด้วยคอลัมน์ที่เท่าไหร่ มีข้อมูลทั้งหมดกี่ชุด และผู้ใช้สามารถแก้ไขรายละเอียดต่าง ๆ ได้ใหม่ในกรณีที่โปรแกรมอ่านเข้ามาไม่ตรงลักษณะความต้องการของผู้ใช้ ดังรูป ก-7



รูป ก-7 แสดงตัวอย่างการอ่านแฟ้มข้อมูล

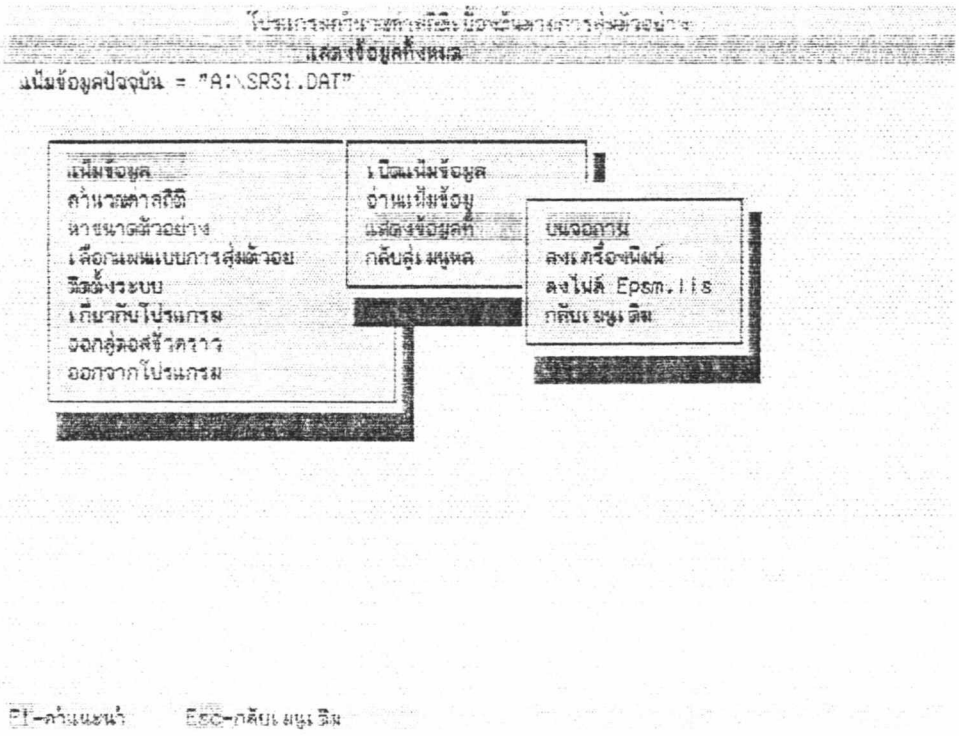
1.3 แสดงข้อมูลทั้งหมด สำหรับแสดงและตรวจสอบข้อมูลที่อ่านขึ้นมาว่าตรงกับที่ต้องการ หรือมีค่าผิดพลาดหรือไม่ ดังรูป ก-8 มีเมนูย่อยคือ

1.3.1 จอภาพ

1.3.2 ลงเครื่องพิมพ์

1.3.3 ลงไฟล์ Epsm.lis

1.4 กลับสู่เมนูหลัก สำหรับกลับไปอยู่ที่เมนูหลัก



รูป ก-8 แสดงเมนูการแสดงข้อมูลทั้งหมด

2. คำนวณค่าสถิติ (ดังในรูป ก-9) มีเมนูย่อยคือ

2.1 SRS คำนวณค่าสถิติต่าง ๆ ตามแผนแบบการสุ่มตัวอย่างแบบง่าย(Simple Random Sampling)

2.2 PPS คำนวณค่าสถิติต่าง ๆ ตามแผนแบบการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นของขนาดตัวอย่างเข้ามาเกี่ยวข้อง (Sampling with Probability Proportional to size)

2.3 Stratified คำนวณค่าสถิติต่าง ๆ ตามแผนแบบการสุ่มตัวอย่างแบบแบ่งประชากรออกเป็นชั้นภูมิ (Stratified Random Sampling)

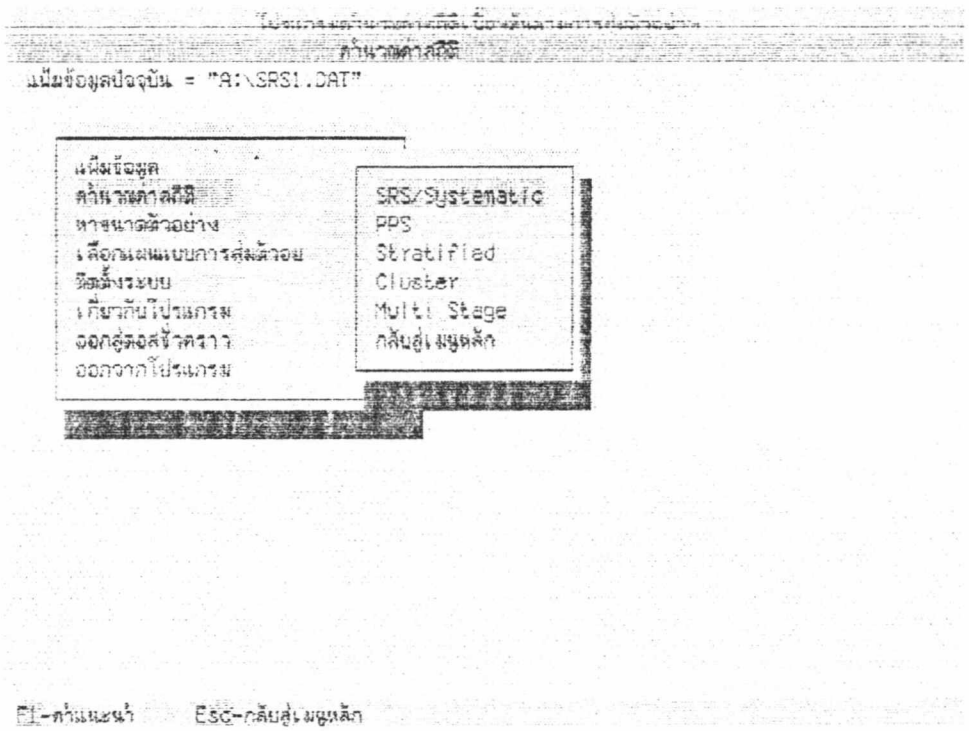
2.4 Cluster คำนวณค่าสถิติต่าง ๆ ตามแผนแบบการสุ่มตัวอย่างแบบแบ่งกลุ่ม (Cluster Sampling)

2.5 Multi-Stage คำนวณค่าสถิติต่าง ๆ ตามแผนแบบการสุ่มตัวอย่างแบบ

สองขั้นตอน (Two-Stage Sampling) โดยมีเมนูย่อยคือ

2.5.1 SRS-SRS แผนแบบการสุ่มตัวอย่างแบบสองขั้นตอนโดยทั้งสองขั้นตอนใช้วิธีการสุ่มตัวอย่างแบบ Sample Random Sampling

2.5.2 PPS-SRS แผนแบบการสุ่มตัวอย่างแบบสองขั้นตอนโดยขั้นตอนใช้วิธีการสุ่มตัวอย่างแบบ Sampling with Probability Proportional to size และขั้นตอนที่สองใช้วิธีการสุ่มตัวอย่างแบบ Sample Random Sampling



รูป ก-9 แสดงเมนูคำนวณค่าสถิติ

โดยทุกการสุ่มจะมีเมนูย่อย (ดังในรูป ก-10) คือ

- Estimate Mean สำหรับคำนวณค่าประมาณค่าเฉลี่ยของประชากร ค่าประมาณความแปรปรวนของค่าประมาณค่าเฉลี่ยของประชากร ค่าสัมประสิทธิ์ของความผันแปร(Coefficeincy of Variance) ค่าช่วงความเชื่อมั่น(Confidence Interval)

- Estimate Total สำหรับคำนวณค่าประมาณยอดรวมของประชากร ค่าประมาณความแปรปรวนของค่าประมาณยอดรวมของประชากร ค่าสัมประสิทธิ์ของความผัน

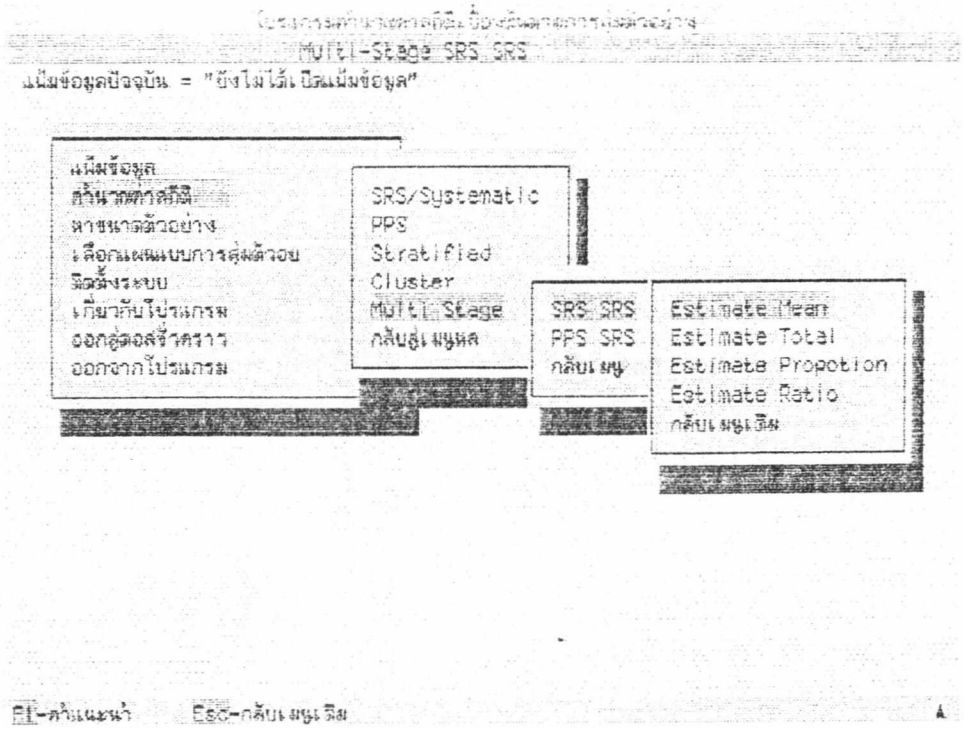
แปร(Coefficeincy of Variance)ค่าช่วงความเชื่อมั่น (Confidence Interval)

- Estimate Propotinal สำหรับคำนวณค่าประมาณสัดส่วนของประชากร ค่าประมาณความแปรปรวนของค่าประมาณสัดส่วนของประชากร ค่าสัมประสิทธิ์ของความผันแปร (Coefficeincy of Variance) ค่าช่วงความเชื่อมั่น (Confidence Interval)

- Estimate Ratio สำหรับคำนวณ ค่าประมาณอัตราส่วนของประชากร ค่าประมาณความแปรปรวนของค่าประมาณอัตราส่วนของประชากร ค่าสัมประสิทธิ์ของความผันแปร(Coefficeincy of Variance) ค่าช่วงความเชื่อมั่น (Confidence Interval)

- กลับสู่เมนูเดิม สำหรับกลับไปเลือกเมนูก่อนหน้านี้

2.6 กลับสู่เมนูหลัก สำหรับกลับไปอยู่ที่เมนูหลัก



รูป ก-10 แสดงเมนูย่อยในเมนูคำนวณค่าสถิติ

3. หาขนาดตัวอย่าง สำหรับหาขนาดตัวอย่างต่ำที่สุดที่ควรจะสุ่ม ตามแผนแบบการสุ่มตัวอย่าง ชนิดของค่าผิดพลาดและชนิดของค่าประมาณที่ต้องการจะประมาณ มีเมนูย่อยคือ (ดังในรูป ก-11)

3.1 SRS

- ค่าเฉลี่ยแบบ absolute error สำหรับขนาดตัวอย่างที่จะนำไปประมาณค่าเฉลี่ยของประชากร โดยหาจากการกำหนดค่า absolute error

- ค่าเฉลี่ยแบบ relative error สำหรับขนาดตัวอย่างที่จะนำไปประมาณค่าเฉลี่ยของประชากร โดยหาจากการกำหนดค่า relative error

- สัดส่วนแบบ absolute error สำหรับขนาดตัวอย่างที่จะนำไปประมาณสัดส่วนของประชากร โดยหาจากการกำหนดค่า absolute error

- สัดส่วนแบบ relative error สำหรับขนาดตัวอย่างที่จะนำไปประมาณสัดส่วนของประชากร โดยหาจากการกำหนดค่า relative error

3.2 PPS

- ค่าเฉลี่ยแบบ absolute error สำหรับขนาดตัวอย่างที่จะนำไปประมาณค่าเฉลี่ยของประชากร โดยหาจากการกำหนดค่า absolute error

- ค่าเฉลี่ยแบบ relative error สำหรับขนาดตัวอย่างที่จะนำไปประมาณค่าเฉลี่ยของประชากร โดยหาจากการกำหนดค่า relative error

3.3 Stratified

- ค่าเฉลี่ย ($\bar{y}$) สำหรับขนาดตัวอย่างที่จะนำไปประมาณค่าเฉลี่ยของประชากร มีเมนูย่อยคือ

  - absolute error
  - relative error

  และแต่ละเมนูมีเมนูย่อยคือ

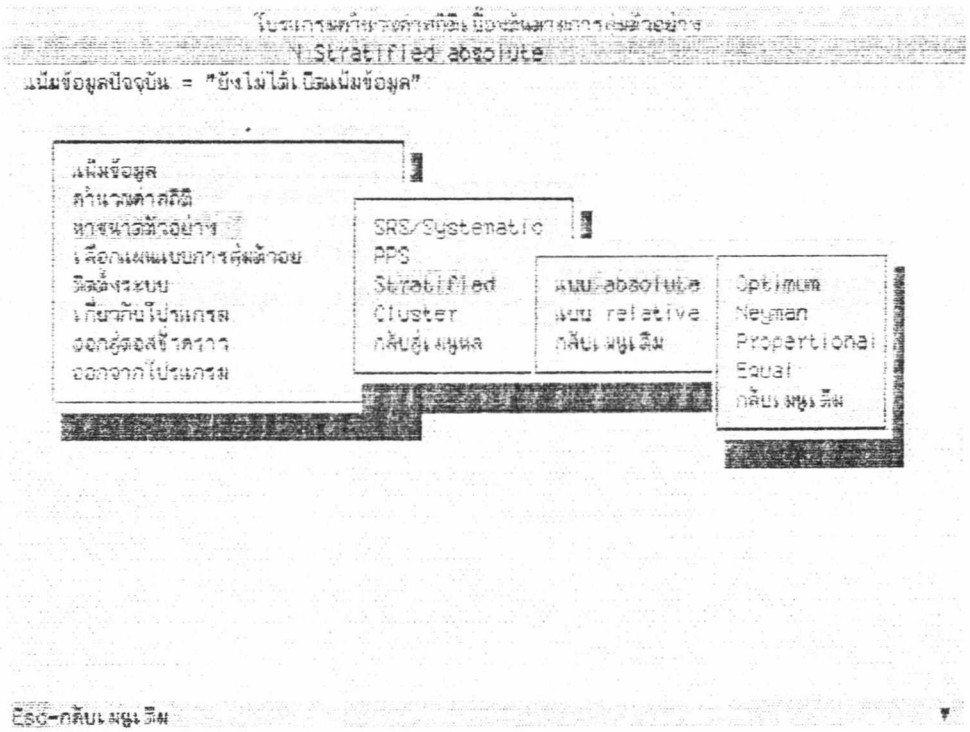
  - Optimum allocation
  - Neyman allocation
  - Proportional allocation
  - Equal allocation

3.4 Cluster

- ค่าเฉลี่ยแบบ absolute error สำหรับขนาดตัวอย่างที่จะนำไปประมาณค่าเฉลี่ยของประชากร โดยหาจากการกำหนดค่า absolute error

- ค่าเฉลี่ยแบบ relative error สำหรับขนาดตัวอย่างที่จะนำไปประมาณค่าเฉลี่ยของประชากร โดยหาจากการกำหนดค่า relative error

3.5 กลับสู่เมนูหลัก สำหรับกลับไปอยู่ที่เมนูหลัก



รูป ก-11 แสดงเมนูหาขนาดตัวอย่าง

ซึ่ง absolute error ชนิดของค่าผิดพลาดซึ่งมีค่าเท่ากับ $\left|\text{ค่าประมาณ} - \text{ค่าจริง}\right|$

relative error ชนิดของค่าผิดพลาดซึ่งมีค่าเท่ากับ $\left|\dfrac{\text{ค่าประมาณ} - \text{ค่าจริง}}{\text{ค่าจริง}}\right|$

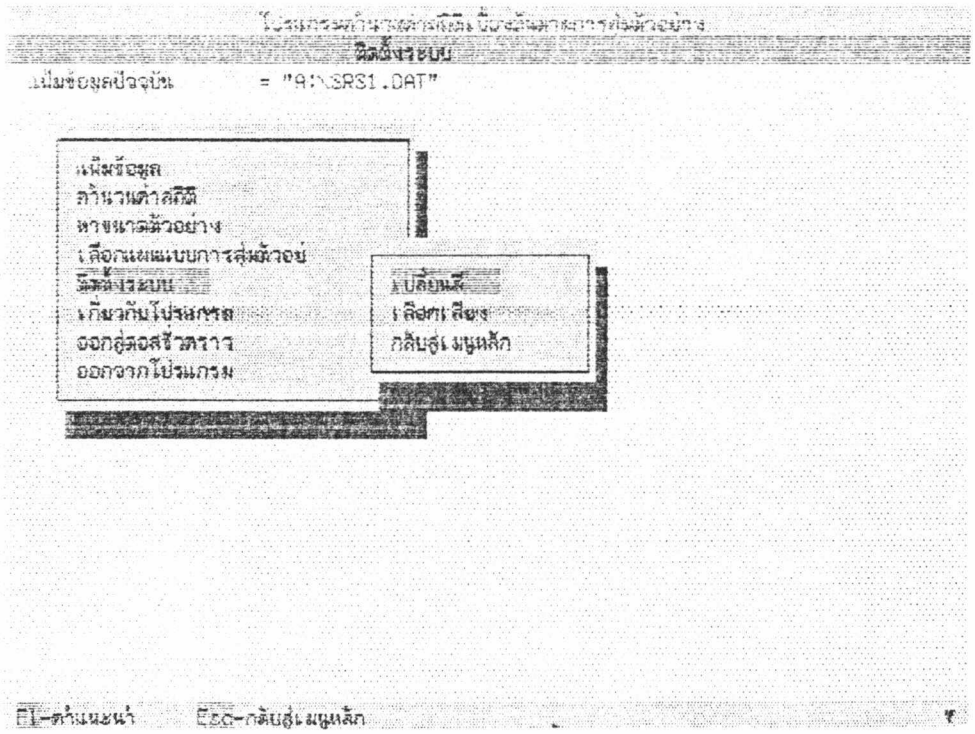
4. เลือกแผนแบบการสุ่มตัวอย่าง สำหรับช่วยผู้ที่รู้ลักษณะของประชากรที่จะทำการสุ่มตัวอย่าง แต่ไม่ทราบว่าควรจะเลือกแผนแบบการสุ่มตัวอย่างใดจึงจะเหมาะสม

5. ติดตั้งระบบ (ดังในรูป ก-12) มีเมนูย่อยคือ

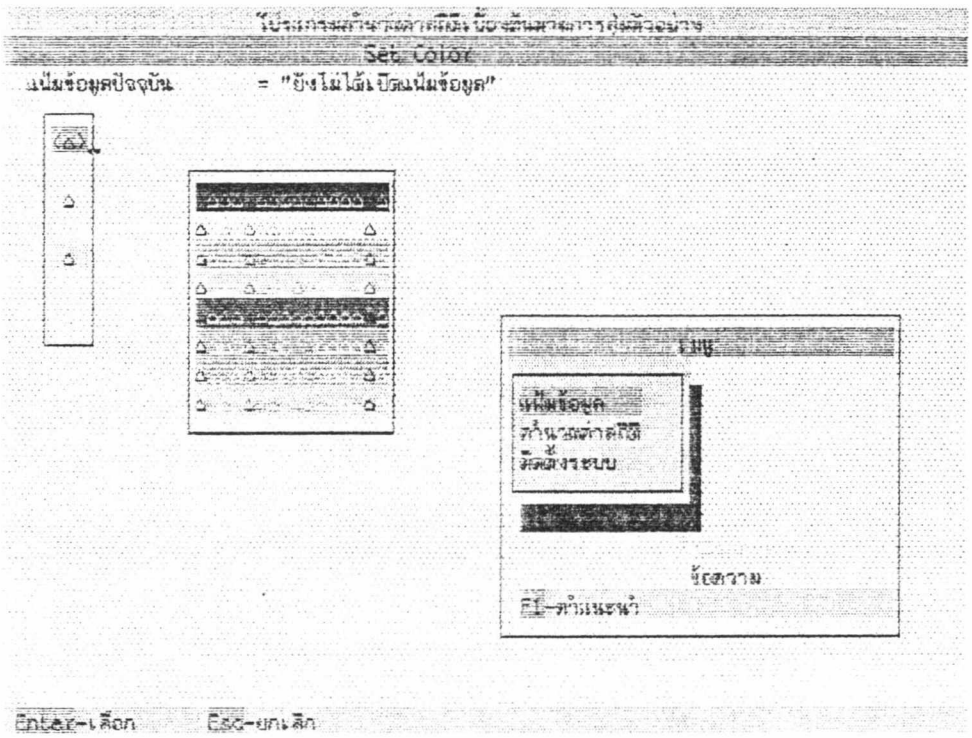
5.1 เปลี่ยนสี สำหรับเปลี่ยนสีบนหน้าจอที่ใช้งานตามแต่ผู้ใช้ต้องการ ดังในรูป ก-13

5.2 เลือกเสียง สำหรับเลือกว่าโปรแกรมจะมีเสียงเตือนหรือไม่มีเสียงเตือน ดังในรูป ก-14

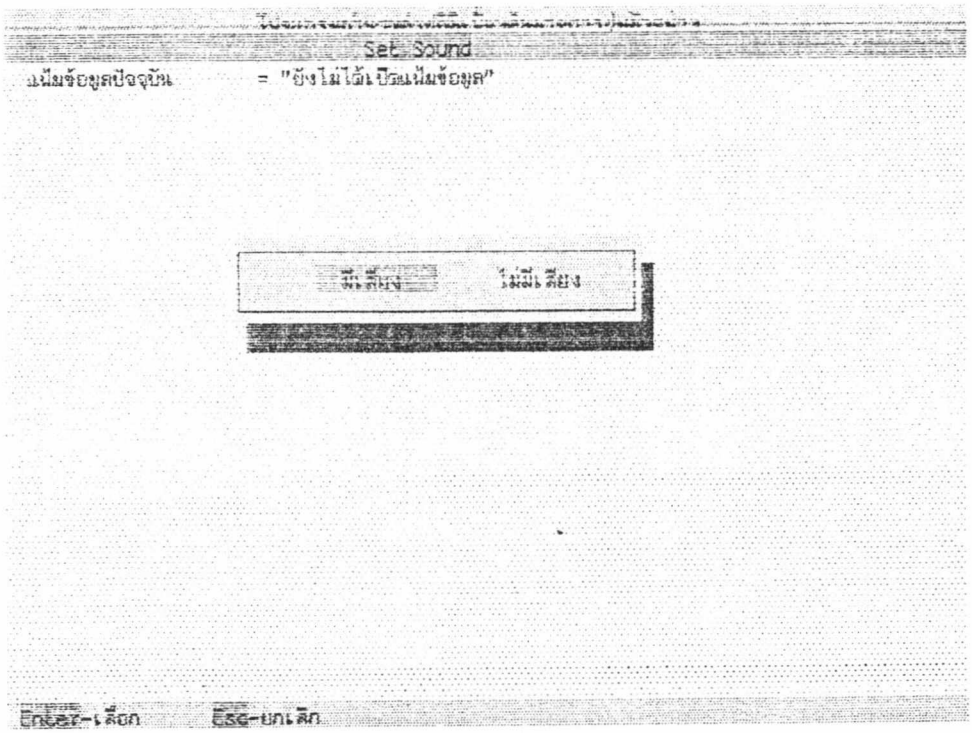
5.3 กลับสู่เมนูหลัก สำหรับกลับไปอยู่ที่เมนูหลัก



รูป ก-12 แสดงเมนูติดตั้งระบบ

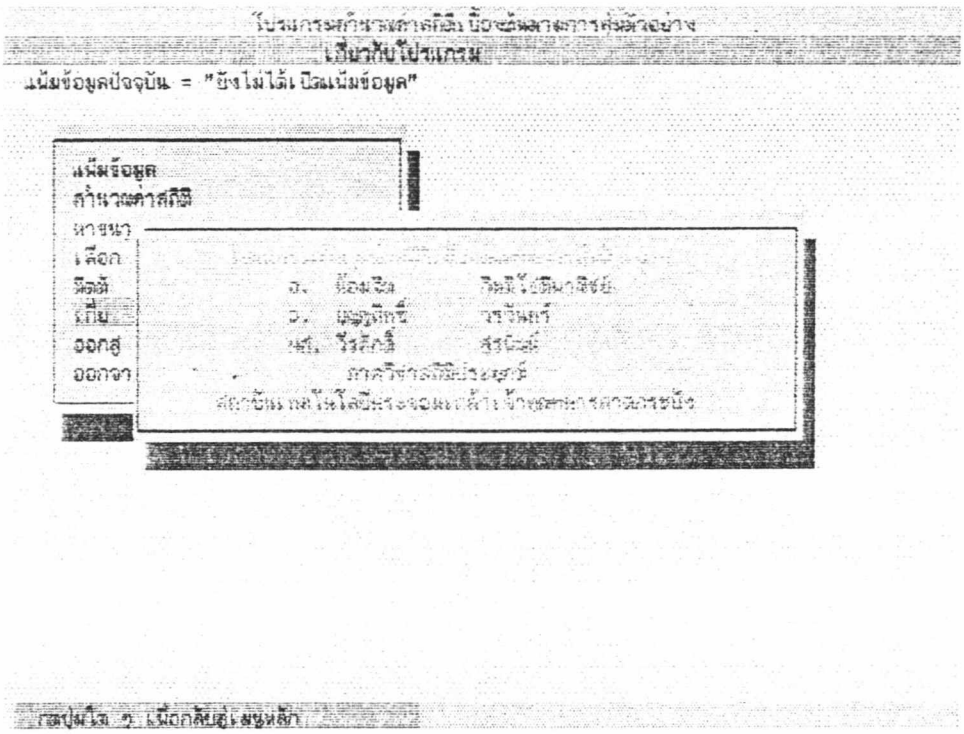


รูป ก-13 แสดงการเปลี่ยนสี



รูป ก-14 แสดงการเลือกเสียง

6. เกี่ยวกับโปรแกรม สำหรับบอกรายละเอียดของโปรแกรม เช่น บอกว่าเป็นโปรแกรม Version เท่าไหร่ และบุคคลที่สามารถติดต่อขอรายละเอียดเกี่ยวกับโปรแกรม ดังรูป ก-15



รูป ก-15 แสดงรายละเอียดเกี่ยวกับโปรแกรม

7. ออกสู่ดอสชั่วคราว สำหรับออกไปใช้คำสั่งดอส แต่ยังต้องการใช้งานโปรแกรมต่ออีกเวลาจะกลับมาทำงานต่อในโปรแกรมให้คีย์คำว่า "EXIT" ดังตัวอย่างในรูป ก-16 (โดยสมมติว่าถ้าขณะนั้นเราอยู่ที่ไดร์ฟ A)

```
A:\>EXIT
```

รูป ก-16 แสดงคำสั่งในการกลับสู่โปรแกรม EPSM

8. ออกจากโปรแกรม สำหรับเมื่อต้องการเลิกใช้งานแล้ว

ปุ่มคีย์บอร์ดที่ใช้ทั่วไป

↑	เลื่อนขึ้น	↓	เลื่อนลง
<-	เลื่อนไปทางซ้าย	->	เลื่อนไปทางขวา
Page Up	เลื่อนขึ้น 1 หน้าจอ	Page Down	เลื่อนลง 1 หน้าจอ
Enter	เลือกรายการนั้น ๆ	Esc	ยกเลิกการทำงานนั้น ๆ
Insert	พิมพ์แทรก/พิมพ์ทับ	Delete	ลบ
F1	คำแนะนำ	F10	ออกจากโปรแกรม

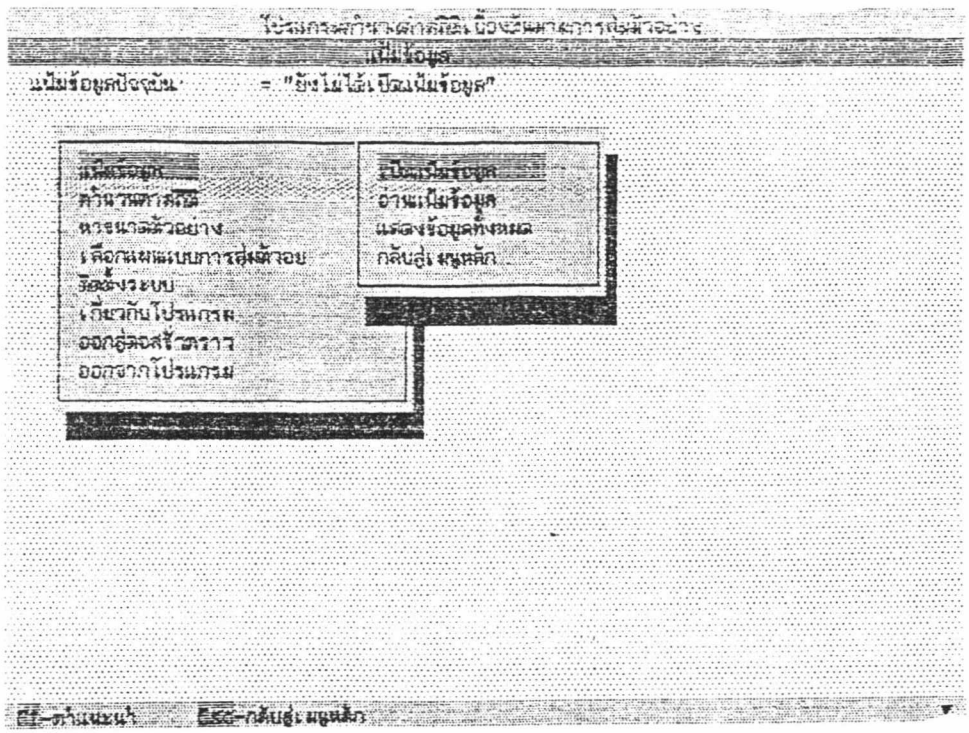
ตัวอย่างการใช้งาน

ในที่นี้จะแสดงขั้นตอนการใช้โปรแกรม โดยจะแสดงการคำนวณค่าเฉลี่ยด้วยแผนแบบการสุ่มแบบง่าย จากแฟ้มข้อมูล SRS1.DAT ซึ่งมีข้อมูลดังนี้

1	11	20
2	6	11
3	12	10
4	9	15
5	8	17
6	11	25
7	10	23
8	9	22
9	12	21
10	13	20

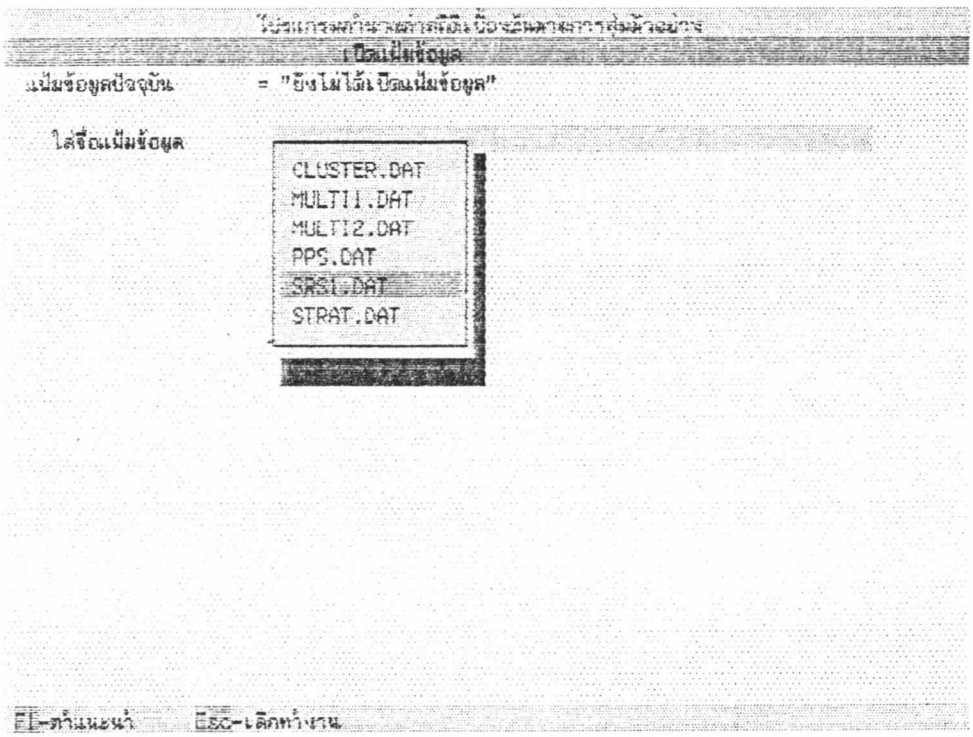
11	11	19
12	10	26
13	17	27
14	19	13
15	20	15
16	23	16
17	17	17
18	18	18
19	19	19
20	15	20

ขั้นตอนแรกให้เลื่อนแถบแสงที่เมนูหลักไปที่เมนูแฟ้มข้อมูลแล้วกด Enter จะปรากฏเมนูย่อยให้เลือกจัดการเกี่ยวกับแฟ้มข้อมูล ดังรูป ก-17



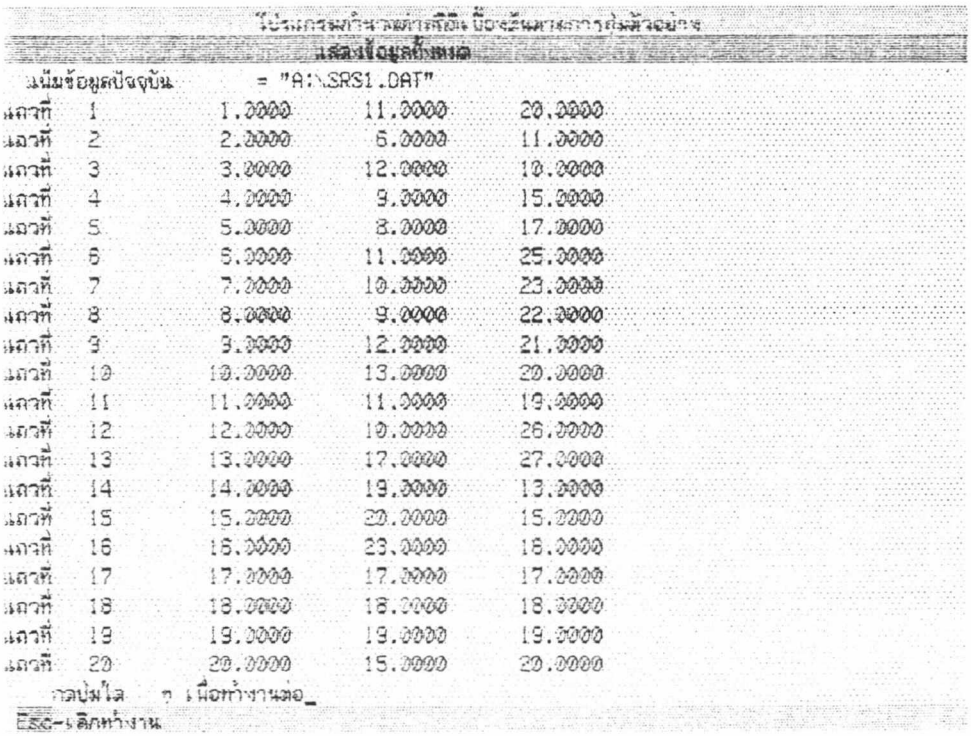
รูป ก-17 แสดงเมนูแฟ้มข้อมูล

หลังจากนั้นเลื่อนแถบแสงไปที่เมนูเปิดแฟ้มข้อมูล แล้วกด Enter จะปรากฏหน้าจอให้กรอกชื่อแฟ้มข้อมูล ในที่นี้เลือกใช้แฟ้มข้อมูล SRS1.DAT ดังรูป ก-18



รูป ก-18 แสดงตัวอย่างการเปิดแฟ้มข้อมูล SRS1.DAT

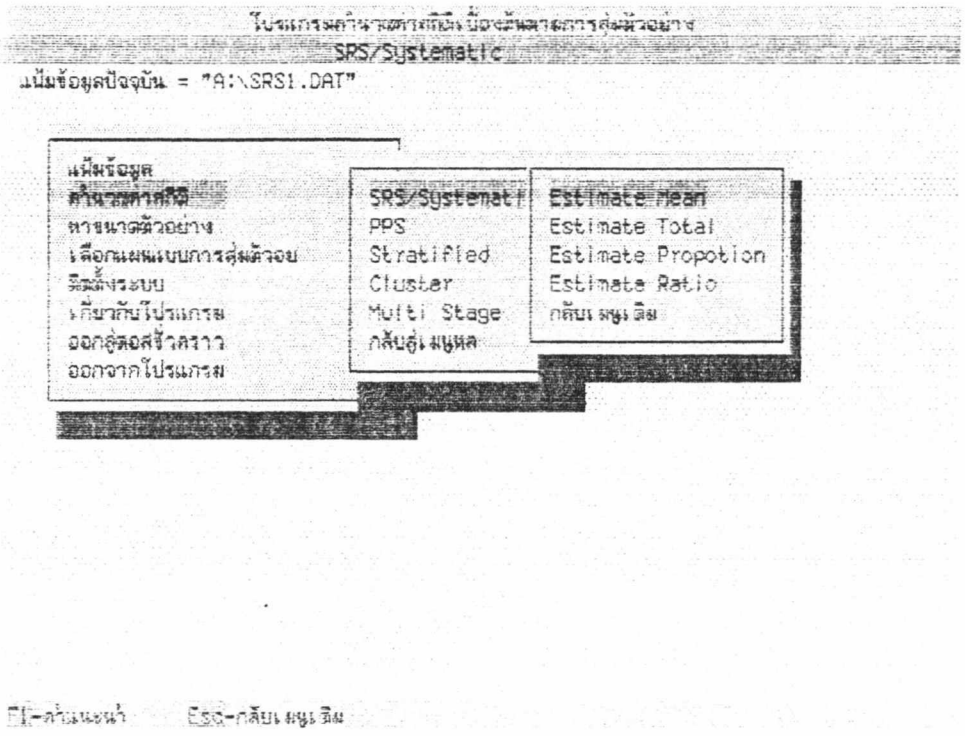
เพื่อทำการตรวจสอบข้อมูลว่าโปรแกรมอ่านค่าขึ้นมาตรงกับข้อมูลในแฟ้มหรือไม่ จึงเลื่อนแถบแสงมาที่เมนูแสดงข้อมูลทั้งหมด ดังรูป ก-19 ถ้าข้อมูลไม่ตรงอาจเกิดจากกำหนดตำแหน่งคอลัมน์เริ่มต้นและสิ้นสุดของแต่ละตัวแปรผิด ให้ไปแก้ไขที่เมนูอ่านแฟ้มข้อมูล



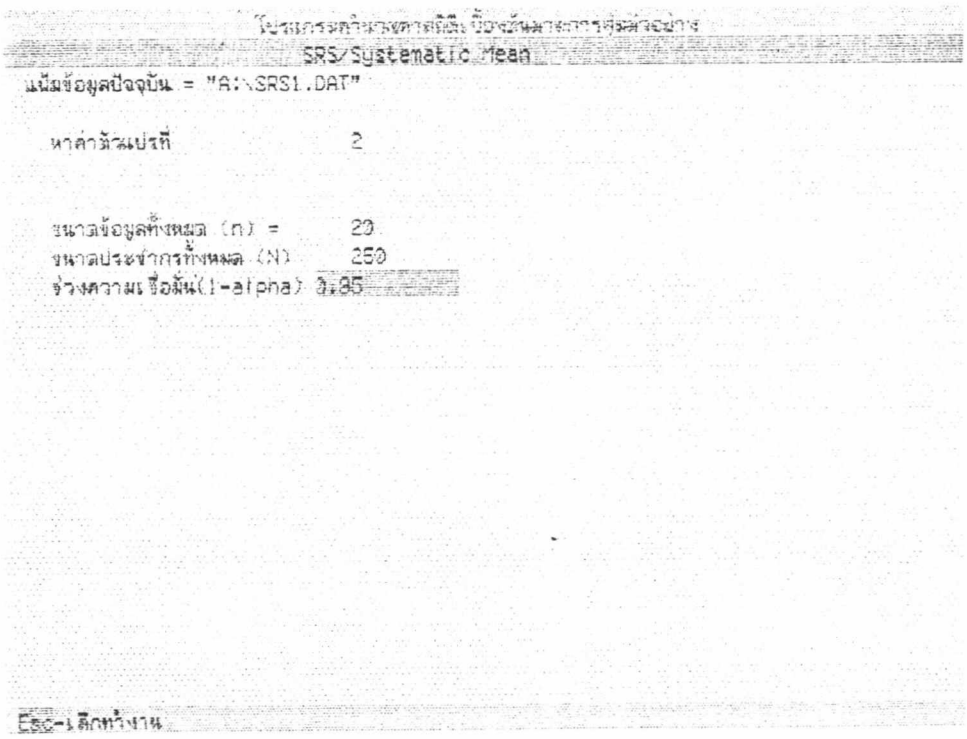
รูป ก-19 แสดงข้อมูลทั้งหมดของแฟ้มข้อมูล SRS1.DAT

เมื่อต้องการจะคำนวณค่าสถิติ ให้เลื่อนแถบแสงที่เมนูหลักไปที่เมนูคำนวณค่าสถิติ แล้วกด Enter จะปรากฏเมนูย่อยแผนแบบการสุ่มตัวอย่างให้เลือก ในที่นี้จะเลือก SRS แล้วจะปรากฏเมนูย่อยสำหรับคำนวณค่าสถิติที่ต้องการ ซึ่งจะเลือกหาค่าเฉลี่ยให้เลื่อนแถบแสงมาที่เมนู Estimae Mean ดังรูป ก-20

เมื่อกด Enter จะปรากฏหน้าจอให้กรอกตัวแปรที่ต้องการจะหาค่า เมื่อต้องการหาค่าตัวแปรตัวที่ 2 ให้กรอก 2 แล้วกด Enter. โปรแกรมจะทำการอ่านแฟ้มข้อมูลได้ขนาดข้อมูลทั้งหมด(n) = 20 หลังจากนั้นให้กรอกขนาดประชากรทั้งหมด(N)คือ 250 และกรอกช่วงความเชื่อมั่นคือ 0.95 ดังรูป ก-21



รูป ก-20 แสดงเมนูย่อยของ SRS ของเมนูคำนวณค่าสถิติ



รูป ก-21 แสดงหน้าจอในการกรอกข้อมูลที่ใช้ในการคำนวณค่าประมาณต่างๆ ของค่าเฉลี่ย

เมื่อเสร็จสิ้นการกรอกข้อมูลแล้วโปรแกรมจะทำการคำนวณค่าประมาณต่าง ๆ ได้ดังนี้

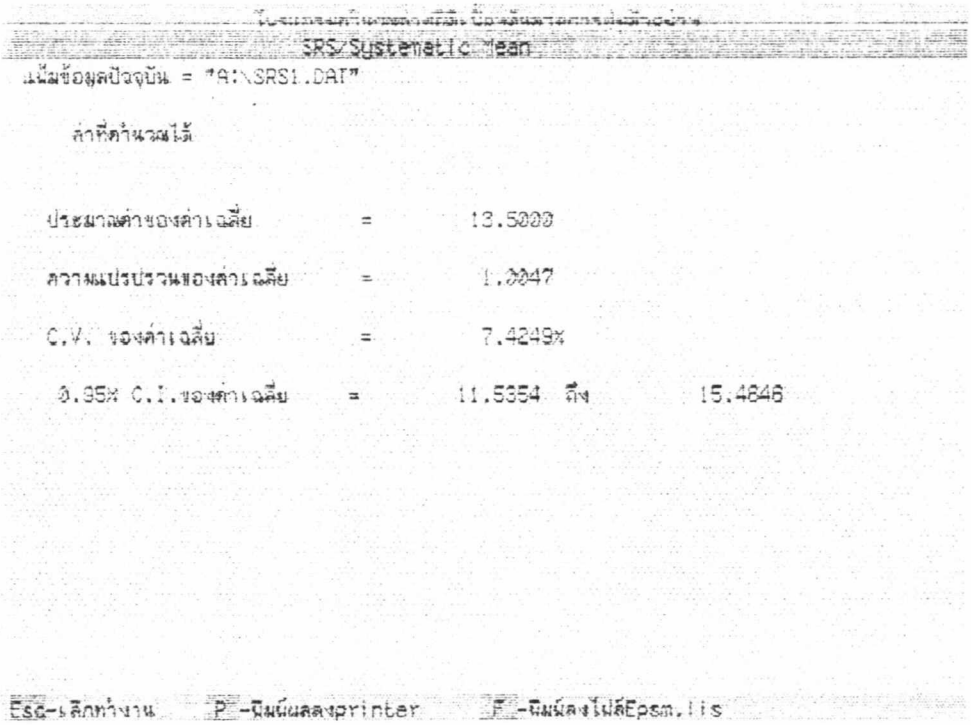
ประมาณของค่าเฉลี่ย = 13.5000

ความแปรปรวนของค่าเฉลี่ย = 1.0047

C.V. ของค่าเฉลี่ย = 7.4249%

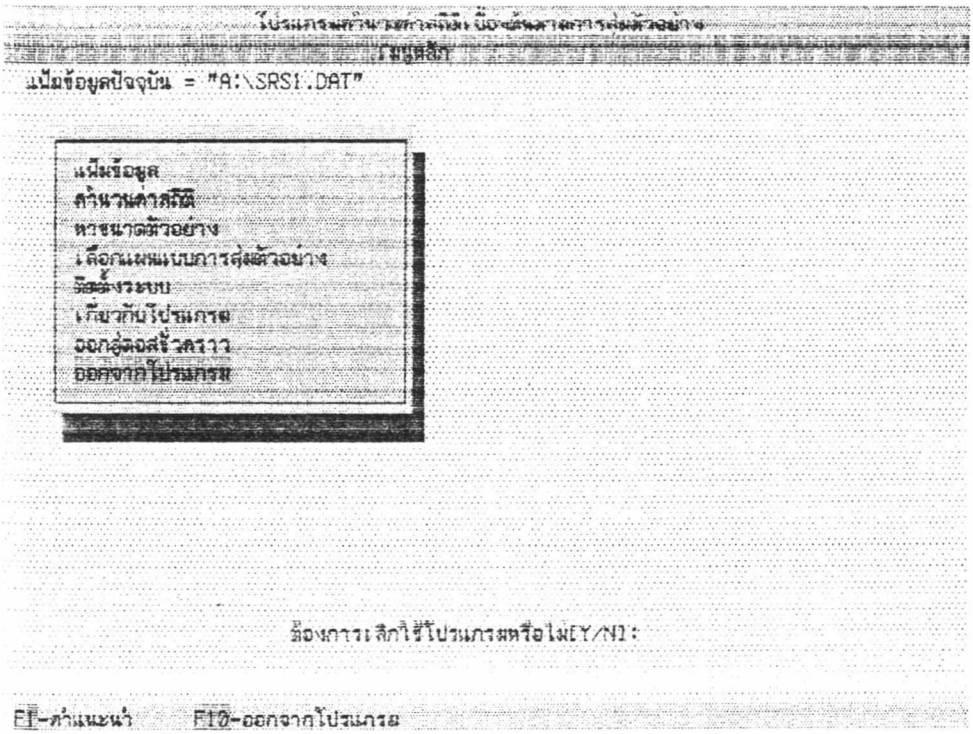
0.95% C.I.ของค่าเฉลี่ย = 11.5354 ถึง 15.4646

ดังรูป ก-22



รูป ก-22 แสดงผลการคำนวณค่าประมาณต่าง ๆ ของค่าเฉลี่ย

เมื่อต้องการจะออกจากโปรแกรมให้กด Esc เพื่อกลับสู่เมนูหลักแล้วเลื่อนแถบแสงไปที่เมนูออกจากโปรแกรม หรือกด F10 ดังรูป ก-23 แล้วกด Y เพื่อเลิกใช้โปรแกรม



รูป ก-23 แสดงการเลิกใช้โปรแกรม

คู่มือการใช้ภาษาไทย TSM

คู่มือการใช้ระบบภาษาไทย TSM

การเรียกใช้ TSM จากบรรทัดคำสั่งของ DOS

โปรแกรม TSM มีรูปแบบคำสั่งที่สามารถเรียกจากบรรทัดคำสั่งของ DOS เพื่อเลือกแบบการทำงาน คือ TSM [แบบการทำงาน]

ซึ่งแบบการทำงานจะเลือกได้ดังนี้

/D ใช้สำหรับถอนการติดตั้งของโปรแกรม TSM ออกจากหน่วยความจำ แต่บางครั้งอาจจะสั่งไม่ได้เพราะมีโปรแกรมประเภทฝังตัวในหน่วยความจำ (Terminated and Stay Resident) อยู่ในหน่วยความจำขณะนั้นด้วย

/E ใช้เพื่อเลือกแบบการแสดงผลทางจอภาพตอนตั้งต้นเป็นแบบใช้รูปร่างตัวอักษรตามปกติของเครื่องคอมพิวเตอร์ คือ ตัวอักษรไทยจะเห็นเป็นตัวอักษรเส้นกรอบหรือสัญลักษณ์

/T ใช้เพื่อเลือกแบบการแสดงผลทางจอภาพตอนตั้งต้นเป็นแบบใช้รูปร่างตัวอักษรไทย แต่ตัวอักษรไทยที่เป็นสระบน สระล่าง และวรรณยุกต์ จะไม่แสดงผลจัดระดับ

/R ใช้เพื่อเลือกแบบการแสดงผลทางจอภาพตอนตั้งต้นเป็นแบบใช้รูปร่างตัวอักษรไทย และแสดงผลจัดระดับตัวอักษรไทยที่เป็นสระบน สระล่าง และวรรณยุกต์

/I ใช้เพื่อสั่งให้โปรแกรม TSM ไม่ต้องอ่านค่าตั้งต้นการทำงานจากแฟ้ม TSM.CFG

/W ใช้เพื่อสั่งให้โปรแกรม TSM หยุดรอก่อนที่จะอ่านค่าตั้งต้นการทำงานจากแฟ้ม TSM.CFG จะมีประโยชน์ในกรณีที่แฟ้ม TSM.EXE และแฟ้ม TSM.CFG ไม่ได้รวมอยู่ในแผ่นเดียวกัน

/BW เลือกใช้สีขาว-ดำ (สำหรับจอภาพ LCD หรือ VGA mono)

/NA ไม่ตรวจสอบชนิดของ video adapter (สำหรับ VGA หรือ EGA compatibles บางชนิด)

/HELP ใช้เพื่อแสดงคำแนะนำในการใช้งาน

การเลือกแบบการทำงานของ TSM จากแป้นพิมพ์

โปรแกรม TSM สามารถที่จะเลือกแบบการทำงานที่ชนะใด ๆ ด้วยการกดแป้นพิมพ์ Alt พร้อมกับแป้นพิมพ์เลือกอื่น ๆ ได้ดังนี้

กดแป้นพิมพ์ Alt พร้อมกับ Shift-Left

ใช้เพื่อสลับเลือกชุดตัวอักษรที่จะแสดงบนจอภาพ ระหว่างตัวอักษรไทย (สมอ. ของสหวิริยา:AT-8708) กับตัวอักษรเส้นกรอบหรือสัญลักษณ์ (US Code Page 437)

กดแป้นพิมพ์ Alt พร้อมกับ Shift-Right

ใช้เพื่อสลับเลือกวิธีการแสดงตัวอักษรไทยบนจอภาพ ระหว่างแบบไม่จัดระดับ กับแบบจัดระดับอัตโนมัติ

กดแป้นพิมพ์ Alt พร้อมกับ Ins

ใช้ชนะที่อยู่ในการแสดงผลแบบจัดระดับอัตโนมัติ เพื่อสลับเลือกใช้หรือไม่ใช้วิธีการจัดวรรคตอนอัตโนมัติ (คือ การปรับตำแหน่งตัวอักษรไม่ให้เลื่อนไปข้างหน้า เนื่องจากสระบน สระล่าง หรือวรรณยุกต์) มีลำดับการกดแป้นพิมพ์ดังนี้

1. กดแป้นพิมพ์ Num Lock ให้แป้นพิมพ์ตัวเลขอยู่ในสภาวะของแป้นลูกศร
2. กดแป้นพิมพ์ Alt พร้อมกับ Ins แล้วปล่อย

กดแป้นพิมพ์ Alt พร้อมกับลูกศรซ้าย (Left-Arrow)

ใช้เพื่อสลับเลือกเปลี่ยนแป้นพิมพ์ตัวเลข(Numeric Keypad)ของแป้นพิมพ์แบบ 101 (Enchanced Keybord) ระหว่างชุดตัวเลขอารบิค กับตัวอักษรกรอบ มีลำดับการกดแป้นพิมพ์ดังนี้

1. กดแป้นพิมพ์ Num Lock ให้แป้นพิมพ์ตัวเลขอยู่ในสภาวะของแป้นลูกศร
2. กดแป้นพิมพ์ Alt พร้อมกับลูกศรซ้าย แล้วปล่อย

3. กดแป้นพิมพ์ Num Lock ให้แป้นพิมพ์ตัวเลขอยู่ในสภาวะของแป้นตัวเลข ซึ่งแป้นตัวเลข 1-9 จะกลายเป็นตัวอักษรกรอบ และยังคงสามารถใช้แป้นลูกศรได้อยู่ด้วยการเลือกจากแป้นพิมพ์ Num Lock

4. เมื่อต้องการยกเลิกเพื่อใช้แป้นพิมพ์ตัวเลขดังเดิม ให้กดแป้นพิมพ์ Num Lock ให้อยู่ในสภาวะของแป้นลูกศร แล้วกดแป้นพิมพ์ Alt พร้อมกับลูกศรซ้ายตาม

ลำดับการการค้นหาแฟ้ม TSM.CFG ของ TSM

ทุกครั้งที่สั่งเรียกโปรแกรม TSM มาใช้งาน นอกจากโปรแกรม TSM จะรับคำสั่งจากบรรทัดคำสั่งของ DOS แล้ว โปรแกรม TSM ยังอ่านค่าตั้งต้นอื่นๆ จากแฟ้ม TSM.CFG

ซึ่งเป็นแฟ้มเก็บค่าตั้งต้นการทำงานของโปรแกรม TSM ที่ถูกสร้างจากโปรแกรม TSMSETUP โดยตามปกติแล้วโปรแกรม TSM จะมีลำดับในการค้นหาแฟ้ม TSM.CFG ดังต่อไปนี้

1. ค้นหาแฟ้ม TSM.CFG จากสารบบที่อยู่ขณะนั้น (current directory)

2. ค้นหาแฟ้ม TSM.CFG จากสารบบที่ระบุโดยตัวแปรของระบบ (DOS Environment) ที่ชื่อว่า TSM เช่น ถ้าต้องการให้ค้นหาแฟ้ม TSM.CFG จากสารบบ C:\PROGRAM ให้นิมพ์ข้อความที่บรรทัดคำสั่งของ DOS ว่า

```
SET TSM=C:\PROGRAM
```

3. ค้นหาแฟ้ม TSM.CFG จากสารบบที่โปรแกรม TSM อยู่ขณะนั้น(เพราะโปรแกรม TSM อาจถูกเรียกใช้งานจากสารบบอื่นซึ่งมิใช่สารบบที่อยู่ขณะนั้น)

ในกรณีที่โปรแกรม TSM ไม่สามารถค้นหาแฟ้ม TSM.CFG จากสารบบใดเลย ก็จะใช้ค่าตั้งต้นในตัวเอง

เอกสารอ้างอิง

ดร. อนันต์ ศรีโสภา . เทคนิคการสุ่มตัวอย่าง , คณะศึกษาศาสตร์ ศรีนครินทรวิโรฒ ประสานมิตร , 2524

นิยม ปุราคำ. ทฤษฎีของการสำรวจสถิติจากตัวอย่างและการประยุกต์ ,ค.ส. การพิมพ์ กรุงเทพฯ

น้อมจิต กิตติโชติพาณิชย์. "เอกสารประกอบการเรียนการสอนวิชา Sampling",ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , 2535

พงษ์ระพี เตชพาหพงษ์ . แอดวานซ์เอ็มเอสดอส , เอช-เอน การพิมพ์ กรุงเทพฯ,2532

ศิวัฒน์ศิวะบวร , พรชัย จักรธำรงค์ , จิรศักดิ์ ชัยวิริยะกุล. การประยุกต์ใช้งานภาษาซี, เอช-เอน การพิมพ์ กรุงเทพฯ , 2532

วรชัย เยาวปาณี . โปรแกรมสำเร็จรูปทางสถิติสำหรับไมโครคอมพิวเตอร์ SPSS/PC+ ขั้นพื้นฐาน, กรุงเทพฯ:โอเดียนสโตร์ , 2532

สุบงกช จามีกร . เทคนิคการสุ่มตัวอย่างกับงานวิจัย , --[กรุงเทพ]:ภาควิชาสถิติคณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ , 2529

อภิชาติ พงษ์สรีหดุลชัย . การสุ่มสำรวจ , --[กรุงเทพ]:ศูนย์สถิติการเกษตรและสหกรณ์, 2530

อุทุมพร (ทองอุไทย) จามรมาน คณะครุศาสตร์ จุฬาฯ. การสุ่มตัวอย่างทางการศึกษา , พิมพ์ครั้งที่ 1 กรุงเทพฯ:[มปพ], 2530 (กรุงเทพฯ: โครงการตำราวิทยาศาสตร์อุตสาหกรรม)

Cochram,William Gemmell. Sampling techniques ,3nd ed New York , John Wiley , 1977.

E.K. Foreman . Survey sampling principle , -- New York : Marcel Dekker , 1991.

Tim Grady.TURBO C!PROGRAMMING PRINCPLES and PRACTICES,McGRAWHILL BOOK COPMPANY Singapore , 1989.

W.Edwards Deming. Sample design in business research , New York , : John Wiley , 1960.

William Mendenhall , Lyman Ott and Richard L. Scheaffer. Belmont. Elementary survey Sampling , Calif., Duxbury Press , 1971.