หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

วิธีการดำเนินการวิจัย

ในการวิจัยครั้งนี้ ต้องการศึกษาวิธีการประมาณค่ารวมประชากร เมื่อค่าสังเกตบางค่าในประชากรเป็นค่าสูงมาก และเมื่อเลือกตัวอย่างแบบสุ่มอย่างง่าย ชนิดไม่ใส่คืนแล้วได้หน่วยตัวอย่างบางหน่วยมีค่าสูงมาก โดยอาศัยคุณสมบัติของตัวประมาณความถดถอยและคุณสมบัติการแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้ว นอกจากนี้ยังต้องการศึกษาเปรียบเทียบตัวประมาณค่ารวมประชากรที่ผู้วิจัยเสนอแนะ กับตัวประมาณที่เสนอโดยไมเคิลและคาดาบา และตัวประมาณ $\hat{Y}_o$ โดยใช้ค่าประสิทธิภาพสัมพัทธ์เป็นหลักในการพิจารณา ทั้งนี้เทคนิคหรือวิธีที่ใช้ในการจำลองข้อมูลเพื่อหาค่าประสิทธิภาพสัมพัทธ์ก็คือวิธีมอนติคาร์โล (Monte Carlo Method)

ขั้นตอนการดำเนินงานวิจัย ที่จะกล่าวถึงในตอนแรกของบทนี้คือ วิธีมอนติคาร์โล แล้วจึงแสดงรายละเอียดแผนการทดลอง ขั้นตอนของการทดลอง และโปรแกรมที่ใช้ในการทดลอง ปรากฏรายละเอียดตามลำดับดังต่อไปนี้

4.1 วิธีมอนติคาร์โล

เทคนิคที่ใช้สำหรับแก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีมอนติคาร์โลก็เป็นเทคนิคหนึ่งที่จะใช้แก้ปัญหานี้ได้ ชื่อและการพัฒนาระบบการจำลองข้อมูลของวิธีมอนติคาร์โลถูกกล่าวถึงประมาณปี ค.ศ. 1944 และในปี ค.ศ. 1964 ฮัมเมอร์สเล่ย์ และ แฮนด์สคัม (Hammersley and Handscomb) กล่าวไว้ว่า วิธีมอนติคาร์โลเป็นสาขาหนึ่งของคณิตศาสตร์เชิงทดลอง มีหลักการคือจะใช้ตัวเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการ ในปัจจุบันนี้วิธีมอนติคาร์โลมีบทบาทสำคัญในทางปฏิบัติต่อการพัฒนาทางด้านวิชาการในสาขาต่าง ๆ มากยิ่งขึ้น เช่น สาขาทางด้านคณิตศาสตร์ประยุกต์ สาขาการวิจัยดำเนินงาน เป็นต้น ซึ่งขั้นตอนของวิธีมอนติคาร์โลแบ่งเป็น 3 ขั้นตอนดังนี้

4.1.1 การสร้างตัวเลขสุ่ม (random number) การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในวิธีมอนติคาร์โล ทั้งนี้เพราะว่าหลักการของวิธีนี้จะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา ลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบยูนิฟอร์ม (Uniform) ในช่วง (0,1) สำหรับวิธีการสร้างตัวเลขสุ่มมีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของตัวเลขสุ่มที่เกิดขึ้นจะต้องมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) และเป็นอิสระต่อกัน

4.1.2 การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรงแต่อาจจะมีขั้นตอนอื่น ๆ หลายขั้นตอนที่ต้องใช้ตัวเลขสุ่ม

4.1.3 การทดลองกระทำเมื่อประยุกต์ให้ใช้กับตัวเลขสุ่มได้แล้วขั้นตอนต่อไปก็คือการทดลองโดยใช้กระบวนการของการสุ่ม (random process) มากระทำในลักษณะที่ซ้ำ ๆ กัน (replication) เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

4.2 แผนการทดลอง

ในการวิจัยครั้งนี้ ต้องการศึกษาคุณสมบัติของตัวประมาณ $\hat{Y}_k$; k =1,2,3 และศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณดังกล่าวกับตัวประมาณ $\hat{Y}_o$ และ $\hat{Y}_{mkt}$; t = 1,2,3,4 โดยศึกษาจากค่าประสิทธิภาพสัมพัทธ์ เมื่อสมมติให้ตัวแปร Y ที่สนใจจะศึกษามีการแจกแจงแบบล็อกนอร์มอลและแบบแกมม่า เนื่องจากว่าการแจกแจงทั้ง 2 รูปแบบนี้เป็นการแจกแจงลักษณะแบบเบ้ขวา (positively skewed distribution) ที่นิยมใช้ศึกษาในเชิงทฤษฎีในสาขาต่าง ๆ เช่น เศรษฐศาสตร์ วิศวกรรมศาสตร์ เป็นต้น ตัวอย่างของตัวแปรที่มีการแจกแจงดังกล่าวเช่นรายได้ จำนวนสมาชิกในครัวเรือน น้ำหนักของคน เป็นต้น พร้อมทั้งสมมติให้การแจกแจงของความคลาดเคลื่อน (ε) มีลักษณะการแจกแจงเช่นเดียวกันกับตัวแปร Y สำหรับค่าพารามิเตอร์ที่เลือกใช้เพื่อที่จะทำให้มีโอกาสมากที่จะได้หน่วยที่มีค่าสูงมากในประชากรคือ กรณีที่ $\mu = 2$, $\sigma^2 = 1$ สำหรับการแจกแจงแบบล็อกนอร์มอล และ $\alpha = 2$ $\beta = 1$ สำหรับการแจกแจงแบบแกมม่า ส่วนค่าพารามิเตอร์ของความคลาดเคลื่อน กำหนดให้ $\mu = 0$, $\sigma^2 = 1$ ในการแจกแจงแบบล็อกนอร์มอล และมีค่า $\alpha = 0.02$, $\beta = 10$ ในการแจกแจงแบบแกมม่า เพื่อพยายามทำให้ความคลาดเคลื่อนมีค่าเฉลี่ยใกล้ศูนย์ และมีความ

แปรปรวนเท่ากับความแปรปรวนของตัวแปร Y มากที่สุด สำหรับค่าเฉลี่ยประชากรของตัวแปร X กำหนดให้มีค่าเท่ากันทั้ง 2 การแจกแจงคือกำหนดให้ $\mu_X = 2$ เหตุที่ศึกษาค่าเฉลี่ยประชากรของตัวแปร X ค่าเดียว เพราะค่าเฉลี่ยของประชากรขนาดจำกัดของตัวแปร X ไม่มีผลต่อการเปลี่ยนแปลงประสิทธิภาพของตัวประมาณความถดถอย สิ่งที่มีผลต่อการเปลี่ยนแปลงประสิทธิภาพของตัวประมาณความถดถอยก็คือความสัมพันธ์เชิงเส้นระหว่างตัวแปร Y และตัวแปร X ซึ่งในการวิจัยครั้งนี้จะแยกศึกษาเป็น กรณีตัวแปร Y และตัวแปร X มีความสัมพันธ์เชิงเส้นต่อกันทางบวก โดยกำหนดให้มีค่าสัมประสิทธิ์สหสัมพันธ์ (ρ_{XY}) 4 ค่าคือ 0.1 0.3 0.5 และ 0.7 และกรณีที่มีความสัมพันธ์เชิงเส้นต่อกันทางลบ จะกำหนดให้มีค่า ρ_{XY} 4 ค่าเช่นกันคือ -0.1 -0.3 -0.5 และ-0.7

เนื่องจากในการอนุมานอย่างไม่มีเงื่อนไขจำเป็นต้องทราบการแจกแจงของจำนวนหน่วยที่มีค่าสูงมากที่พบในตัวอย่าง ซึ่งในกรณีที่เลือกตัวอย่างสุ่มแบบง่ายชนิดไม่ใส่คืน การแจกแจงของจำนวนหน่วยที่มีค่าสูงมาก ในตัวอย่างจะเป็นแบบ positive Hypergeometric เพราะค่าของจำนวนหน่วยที่มีค่าสูงมากในตัวอย่างมีค่าเป็นไปได้เริ่มตั้งแต่ 1 ขึ้นไป ซึ่งมีปัญหาในการหาความน่าจะเป็นเมื่อขนาดประชากรขนาดใหญ่ (มากกว่า 1000) ดังนั้นขนาดประชากรที่ใช้ศึกษาจึงกำหนดให้มีขนาดเท่ากับ 500 และ 1000 ส่วนจำนวนค่าลักษณะที่เป็นค่าสูงมากในประชากรจะศึกษา 3 จำนวนในขนาดประชากร 500 คิดเป็นร้อยละได้เท่ากับ 1.8% 2.8% และ 3.2% และขนาดประชากร 1000 ศึกษา 3 จำนวนเช่นกัน คิดเป็นร้อยละได้เท่ากับ 1.8% 2.8% และ 3.3% สำหรับขนาดตัวอย่างที่สนใจศึกษามีขนาดเท่ากับ 50 100 และ 200 ส่วนจำนวนหน่วยตัวอย่างที่มีค่าสูงมากในตัวอย่างจะมีค่าเป็นไปได้ตั้งแต่ 2 จนถึงจำนวนค่าลักษณะในประชากรที่เป็นค่าสูงมาก โดยในแต่ละสถานการณ์ที่ศึกษาจะกระทำซ้ำ ๆ กัน 100 ครั้ง ในขนาดประชากร 500 และกระทำซ้ำ ๆ กัน 50 ครั้ง ในขนาดประชากรเท่ากับ 1000

4.3 ขั้นตอนในการทดลอง

ขั้นตอนในการทดลองสำหรับการวิจัยครั้งนี้ แบ่งออกเป็น 3 ขั้นตอนคือ

4.3.1 การสร้างโปรแกรมสำหรับสร้างลักษณะประชากรตามที่ศึกษาพร้อมทั้งตรวจสอบค่าเฉลี่ย ค่าความแปรปรวน จำนวนค่าลักษณะที่เป็นค่าสูงมากที่พบในตัวแปร Y และ X ค่าสัมประสิทธิ์สหสัมพันธ์ (ρ_{XY}) ที่คำนวณได้ และความเป็นเชิงเส้น (linearity)

4.3.2 การหาโมเมนต์ที่ k ของจำนวนหน่วยตัวอย่างที่มีค่าสูงมากที่พบในตัวอย่างที่สุ่มได้คือ $E(n_1^k)$; k = 1,2,3,4,...

4.3.3 การหาค่าประสิทธิภาพสัมพัทธ์ ของตัวประมาณ $\hat{Y}_k$; k = 1,2,3 หรือตัวประมาณ $\hat{Y}_{mkt}$; t = 1,2,3,4 เทียบกับตัวประมาณ $\hat{Y}_o$

ปรากฏรายละเอียดสำหรับแต่ละขั้นตอนดังนี้

4.3.1 การสร้างโปรแกรมสำหรับสร้างลักษณะประชากรตามที่ศึกษา

ก่อนอื่นจะกล่าวถึง การสร้างรูปแบบการแจกแจงของตัวแปร Y และ ε ทุกรูปแบบ ตามที่กล่าวไว้ในแผนการทดลองนั้นจะใช้โปรแกรมภาษาฟอร์แทรนโฟ (Fortran IV) โดยใช้กับเครื่อง IBM 370/3031 ซึ่งการสร้างลักษณะการแจกแจงแบบต่าง ๆ นั้น จำเป็นจะต้องใช้ตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) เป็นพื้นฐานในการสร้าง สำหรับโปรแกรมที่ใช้สร้างตัวเลขสุ่มในการวิจัยครั้งนี้ใช้วิธีที่ Shanon (1975: 352-256) เสนอไว้ คือใช้ลับรูทีนแรนดู (Randu) โดยคำสั่ง call RANDU (IX, IY, YEL) เมื่อ IX คือค่าเริ่มต้นที่ใช้ และ YEL คือตัวแปรสุ่มที่ได้ สำหรับรายละเอียดในการสร้างการแจกแจงแบบต่าง ๆ เป็นดังนี้

1. การแจกแจงแบบล็อกนอร์มอล โปรแกรมย่อยที่ใช้สร้างข้อมูลให้มีการแจกแจงแบบล็อกนอร์มอล นั้นสร้างขึ้นได้จากความสัมพันธ์ระหว่างการแจกแจงแบบปกติกับการแจกแจงแบบล็อกนอร์มอลที่ว่า

"ถ้าให้ N เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย = μ แปรปรวน = σ^2 จะได้ L = exp(N) เป็นตัวแปรสุ่มที่มีการแจกแจงแบบล็อกนอร์มอลที่มีค่าเฉลี่ย = $\exp(\mu + \frac{\sigma^2}{2})$ ความแปรปรวน = $\exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$"[1]

ดังนั้น ก่อนอื่นจึงต้องสร้างข้อมูลให้มีการแจกแจงแบบปกติก่อน โดยใช้ลับรูทีนเกาส์ (Guass) ซึ่งใช้คำสั่ง call GUASS (IX, EX, SD, X_1) เมื่อ IX คือค่าเริ่มต้น SD และ EX คือค่าเบี่ยงเบนมาตรฐานและค่าเฉลี่ยของประชากร เป็นค่าที่ส่งจากโปรแกรมหลัก ส่วนผลลัพธ์คือ X_1 ที่เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย EX และความแปรปรวน

---

[1] Sung, C. choi, Introduction Applied Statisties in Science.

คือ SD^2 แล้วใส่เอกซ์โปเนนเชียลของผลลัพธ์ X_1 ที่ได้ โดยใช้สมการ $Y = \exp(X_1)$ ก็จะได้ตัวแปรสุ่ม Y มีการแจกแจงแบบล็อกนอร์มอลที่มีค่าเฉลี่ยเท่ากับ $\exp(EX + SD^2/2)$ และความแปรปรวนเท่ากับ $\exp(2EX + 2SD^2) - \exp(2EX + SD^2)$

2. การแจกแจงแบบแกมม่า โปรแกรมย่อยแกม (gam) ที่ใช้สร้างข้อมูลให้มีการแจกแจงแบบแกมม่าที่มีค่าเฉลี่ย $\alpha\beta$ และความแปรปรวน $\alpha\beta^2$ นั้นใช้วิธี Inverse Transformation สร้างตัวแปร erlang (z_i) ที่เป็นผลบวกของ R_i ตัวแปร ; i = 1,2 ที่มีการแจกแจงแบบเอกซ์โปเนนเชียล[1] ที่มีค่าเฉลี่ยเท่ากับ β ค่าความแปรปรวนเท่ากับ β^2 เมื่อ R_1 เป็นจำนวนเต็มใหญ่ที่สุดที่เล็กกว่า α ($R_1 = [\alpha]$) และ R_2 เป็นเลขจำนวนเต็มเล็กที่สุดที่ใหญ่กว่า α ($R_2 = R_1 + 1$) แล้วผลิตตัวแปรสุ่ม R ขึ้นมาโดยถ้า R มีค่าน้อยกว่าหรือน้อยกว่า q (q มีค่าเท่ากับ $\alpha - R_1$) จะได้ z_2 เป็นตัวแปรสุ่มที่มีการแจกแจงแบบแกมม่าที่มีพารามิเตอร์ α และ β ในทางตรงกันข้าม ถ้า R มีค่ามากกว่า q จะได้ z_1 เป็นตัวแปรสุ่มที่มีการแจกแจงแบบแกมม่าที่มีพารามิเตอร์ α และ β แทน

สำหรับขั้นตอนการสร้างโปรแกรมสำหรับสร้างลักษณะประชากรตามที่กำหนดมีดังนี้คือ

1. สร้างตัวแปรสุ่ม Y และ ε ให้เป็นตัวแปรสุ่มที่มีการแจกแจงตามที่กำหนด
2. สร้างตัวแปร X ขึ้นมาให้มีความสัมพันธ์เชิงเส้นต่อกันกับตัวแปร Y โดยอาศัยสมการเชิงเส้น $Y_i = \beta_o + \beta_1 X_i + \varepsilon_i$; i = 1,2,...,N เมื่อ $\beta_o = \mu_Y - \beta_1\mu_X$ และ β_1 เป็นจำนวนจริงใด ๆ ที่ทำให้ตัวแปร Y และตัวแปร X มีความสัมพันธ์เชิงเส้นต่อกัน[2] โดยในตอนแรกนี้จะกำหนดให้มีค่าเท่ากับ 1000 ก่อน เพื่อเป็นการค้นหาค่าเริ่มต้นที่จะทำให้มีลักษณะประชากรเหมือนข้อ 3

---

[1] ฟังก์ชั่นความน่าจะเป็นของการแจกแจงแบบเอกซ์โปเนนเชียลมีลักษณะดังนี้

$$f(z) = 1/\beta \; e^{-z/\beta} \quad , z > 0$$

[2] อาศัยจากความสัมพันธ์ของ β_1 กับ ρ_{XY} ที่ว่า $\rho_{XY} = \beta_1 \sqrt{\sum_{i=1}^{N}(X_i-\bar{X})^2 / \sum_{i=1}^{N}(Y_i-\bar{Y})^2}$

3. สร้างตัวแปรสุ่ม Y และ X ให้มีจำนวนเท่ากับขนาดประชากรที่ศึกษาแล้วตรวจสอบค่าเฉลี่ย และค่าความแปรปรวนของตัวแปร Y ε และ X พร้อมกับ ตรวจสอบจำนวนหน่วยที่มีค่าสูงมากที่พบในตัวแปร Y และ X ว่ามีจำนวนเท่ากับที่ต้องการหรือไม่ ถ้าเท่าก็ไปทำข้อ 4 แต่ถ้าไม่เท่าก็เปลี่ยนค่าเริ่มต้นใหม่ แล้วย้อนกลับไปเริ่มต้นทำข้อ 1 ใหม่

4. เปลี่ยนค่า β_1 ไปเรื่อย ๆ เพื่อให้ได้ประชากรที่มีค่า ρ_{XY} ตามที่ศึกษา พร้อมทั้งตรวจสอบค่า ρ_{XY} ที่คำนวณได้ด้วยวิธีการทดสอบแบบ Fisher's Transformation

5. หลังจากที่สร้างประชากรให้มีขนาด (N) จำนวนค่าลังเกตที่เป็นค่าสูงมากที่พบในประชากร (N_1) และมีค่า ρ_{XY} ตรงตามที่ต้องการโดยอาศัยค่าเริ่มต้น 1 ค่า และไปทำในขั้นตอนที่ 4.3.3 (ซึ่งจะกล่าวต่อไป) แล้วขั้นต่อไปเปลี่ยนค่าเริ่มต้นตัวใหม่ และย้อนกลับไปเริ่มต้นทำข้อ 1 ถึงข้อ 4 ใหม่ เพื่อให้ได้ประชากรที่มีขนาด จำนวนค่าลังเกตที่เป็นค่าสูงมากที่พบในประชากรและค่า ρ_{XY} เท่าเดิม กระทำในทำนองเดียวกัน ซ้ำ ๆ กัน 100 ครั้ง ในขนาดประชากรเท่ากับ 500 หรือกระทำซ้ำ ๆ กัน 500 ครั้งในขนาดประชากรเท่ากับ 1000 จนกระทั่งครบทุกสถานการณ์ที่ศึกษา

4.3.2 การหาค่าโมเมนต์ที่ k ของจำนวนหน่วยตัวอย่างที่มีค่าสูงมากที่พบในตัวอย่าง (n_1)

เนื่องจากในการอนุมานอย่างไม่มีเงื่อนไขจำเป็นต้องอาศัยค่าโมเมนต์ที่ k ของจำนวนหน่วยตัวอย่างที่มีค่าสูงมากที่พบในตัวอย่าง (n_1) ดังนั้นจึงจำเป็นต้องหาค่าโมเมนต์ที่ k ของ n_1 ไว้ก่อน เพื่อเป็นการประหยัดเวลาที่ใช้ในการทำงานของโปรแกรม (run program) การหาค่าโมเมนต์ที่ k ของ n_1 จะหาจากสูตรที่ว่า $E(n_1^k) = \sum_{n_1=1}^{N_1} n_1^k \cdot p(n_1 \mid n,N,N_1)$

และเนื่องจาก n_1 มีการแจกแจงแบบ positive hypergeometric distribution

ดังนั้น $$p(n_1 \mid n,N,N_1) = \frac{1}{1-d} \frac{\binom{N_1}{n_1}\binom{N-N_1}{n-n_1}}{\binom{N}{n}} \quad ; n_1 = 1,2,.....,N_1 \text{--(1)}$$

$N-N_1 > n \geq N_1 \geq 1$

เนื่องจาก N มีขนาดใหญ่ การหาความน่าจะเป็นของการแจกแจงแบบ positive hypergeometric ที่อยู่ในรูปแฟคเตอร์เรียลหรือคอมบิเนชั่น จะมีปัญหาไม่สามารถที่จะทำค่าออกมาได้แม้ว่าในการวิจัยครั้งนี้จะใช้เครื่องคอมพิวเตอร์ IBM 370/3031 เพราะเนื่องจากเครื่องคอมพิวเตอร์มีเนื้อที่ในการเก็บหน่วยความจำเพียง $2^{\pm 32}$ ไบท์ (byte) เท่านั้น ดังนั้นจึงอาศัยลอการิทึมแฟคเตอร์เรียล หาค่าความน่าจะเป็นสำหรับแต่ละค่าของ n_1 เมื่อใช้ขนาดตัวอย่างและขนาดประชากรค่า ๆ หนึ่ง ซึ่งมีวิธีการหาค่าความน่าจะเป็นดังนี้คือ

1. จากสูตรสำหรับคำนวณหาค่าความน่าจะเป็นในสมการที่ (1) กระจายค่าคอมบิเนชั่นให้อยู่ในรูปแฟคเตอร์เรียลจะได้เป็น

$$p(n_1 \mid n,N,N_1) = \frac{1}{1-d} \frac{N_1!\ (N-N_1)!\ n!\ (N-n)!}{n_1!\ (N_1-n_1)!\ (n-n_1)!\ (N-N_1-n+n_1)!\ N!} \quad ---(2)$$

2. หลังจากนั้นใส่ลอกฐาน 10 ในสมการที่ 2 จะได้

$$\log p(n_1 \mid n,N,N_1) = \log(N_1!) + \log((N-N_1)!) + \log(n!) + \log((N-n)!)$$
$$-\log(1-d) - \log(n_1!) - \log((N_1-n_1)!) - \log((n-n_1)!)$$
$$-\log((N-N_1-n+n_1)!) - \log(N!)$$
$$= \left[\log(N_1!) + \log((N-N_1)!) + \log(n!) + \log((N-n)!)\right.$$
$$\left.-\log(N!) - \log(1-d)\right] - \left[\log(n_1!) + \log((N_1-n_1)!)\right.$$
$$\left.+\log((n-n_1)!) + \log((N-N_1-n+n_1)!)\right]$$

กำหนดให้ $T_1 = \log(N_1!) + \log((N-N_1)!) + \log(n!) + \log((N-n)!) - \log(N!) - \log(1-d)$

$T_2 = \log(n_1!) + \log((N_1-N_1)!) + \log((n-n_1)!) + \log((N-N_1-n+n_1)!)$

$$\therefore \quad \log p\,(n_1 \mid n,N,N_1) = T_1 - T_2$$

3. เปลี่ยนค่า $\log p\,(n_1 \mid n,N,N_1)$ ให้อยู่ในรูปแอนตี้ลอก (anti-log) แล้วจะได้

$$p\,(n_1 \mid n,N,N_1) = 10^{(T_1-T_2)}$$

แต่ทั้งนี้ถึงแม้ว่าจะสามารถหาค่าความน่าจะเป็นสำหรับแต่ละค่าของ n_1 ได้ก็ยังมีข้อจำกัดอีกคือ ขนาดประชากร N ไม่ควรมีค่ามากกว่า 1000 เพราะในตารางล็อกการิทึมแฟคตอเรียลมีค่าสูงสุดเพียง N = 1000 เท่านั้น

4.3.3 การหาค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณ $\hat{Y}_k$; k = 1,2,3 $\hat{Y}_{mkt}$; t = 1,2,3,4 เทียบกับตัวประมาณ $\hat{\hat{Y}}_o$

เมื่อสร้างลักษณะการแจกแจงของประชากรได้ตามที่ต้องการและคำนวณหาค่าโมเมนต์ที่ k ของ n_1 ได้แล้ว ขั้นต่อไปคือ การทดลอง เพื่อศึกษาคุณสมบัติของตัวประมาณ และเปรียบเทียบประสิทธิภาพของตัวประมาณ $\hat{Y}_k$ กับ $\hat{Y}_{mkt}$ และ $\hat{Y}_o$ ว่าตัวประมาณตัวใดจะมีประสิทธิภาพสูงสุด เพื่อใช้เป็นเกณฑ์อย่างหนึ่งในการตัดสินใจว่าควรจะเลือกใช้ตัวประมาณตัวใดในแต่ละสภาวะการณ์ตามวัตถุประสงค์ของการวิจัยนี้คือ

1. เมื่อความสัมพันธ์เชิงเส้นระหว่างตัวแปร Y และตัวแปร X มีค่ามากขึ้น
2. เมื่อในประชากรมีจำนวนค่าลังเกตที่เป็นค่าสูงมากเพิ่มขึ้น
3. เมื่อในตัวอย่างมีจำนวนหน่วยตัวอย่างที่มีค่าสูงมากเพิ่มขึ้น

ดังนั้น จึงจำเป็นต้องหาค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณดังที่กล่าวมาแล้วเทียบกับตัวประมาณ $\hat{Y}_o$ เพื่อหาข้อสรุปในการศึกษาคุณสมบัติของตัวประมาณ $\hat{Y}_k$ และเปรียบเทียบประสิทธิภาพของตัวประมาณ $\hat{Y}_k$ กับ $\hat{Y}_{mkt}$ และ $\hat{Y}_o$ ซึ่งกำหนดให้ค่าประสิทธิภาพสัมพัทธ์มีค่าดังนี้

$$\text{R.E.}\ (\hat{Y}_k \text{ เทียบกับ } \hat{Y}_o) = \frac{\text{MSE ของ } \hat{Y}_o}{\text{MSE ของ } \hat{Y}_k} \quad ; \quad k = 1,2,3$$

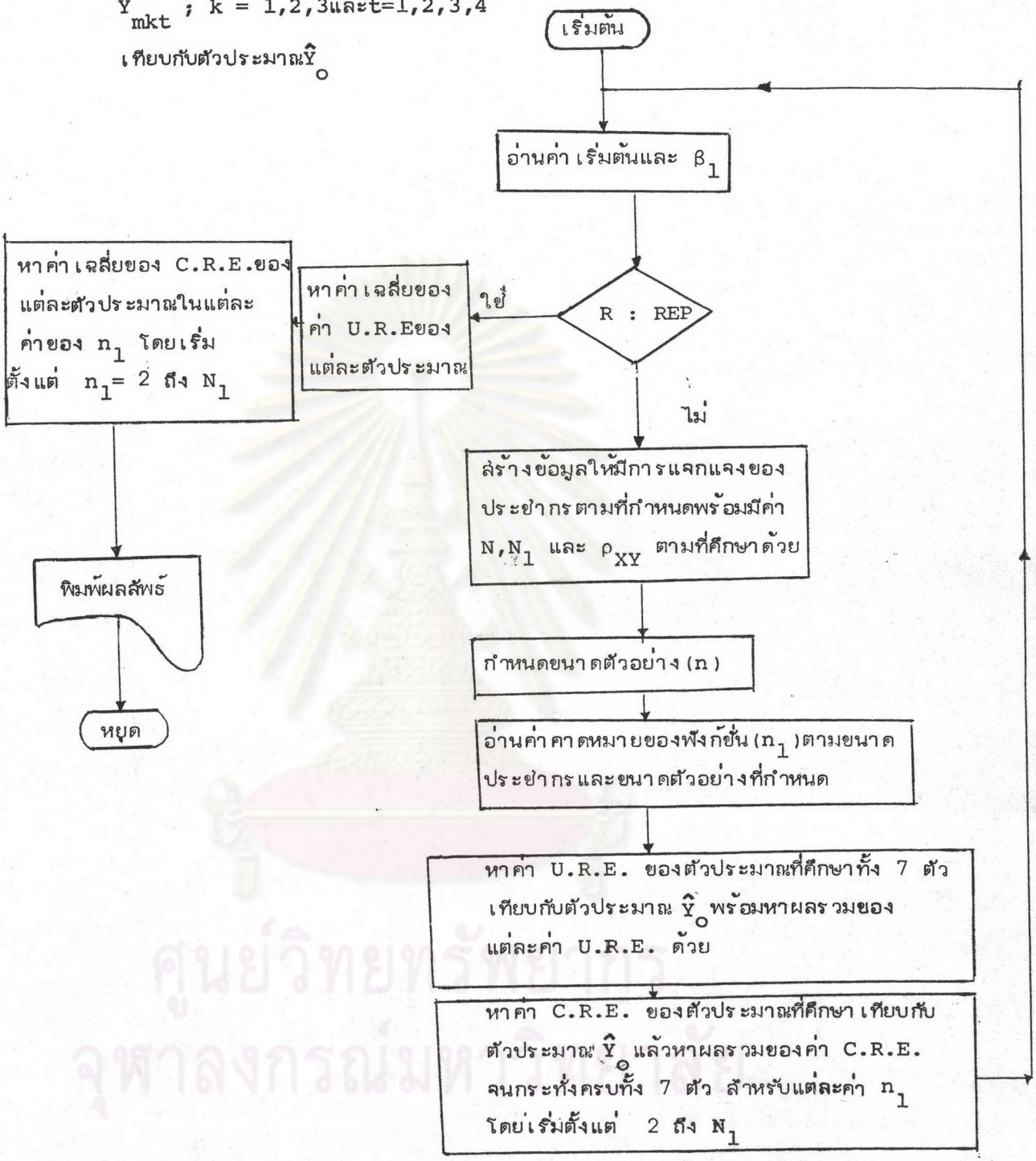
และ

$$R.E.(\hat{Y}_{mkt} \text{ เทียบกับ } \hat{Y}_o) = \frac{\text{MSE ของ } \hat{Y}_o}{\text{MSE ของ } \hat{Y}_{mkt}} \quad ; \quad t = 1,2,3,4$$

และเมื่อหาค่าประสิทธิภาพสัมพัทธ์ โดยใช้ค่าเริ่มต้นหนึ่งเสร็จแล้วจะย้อนกลับไปเปลี่ยนค่าเริ่มต้นตัวใหม่เพื่อสร้างประชากรขนาดเท่าเดิม มีจำนวนค่าสังเกตที่เป็นค่าสูงมากในประชากรค่า ρ_{XY} และขนาดตัวอย่างที่ใช้เท่าเดิมจนกระทั่งครบตามจำนวนซ้ำที่กำหนดใช้ในแผนการทดลอง จากนั้นก็เปลี่ยนค่าขนาดตัวอย่างไปจนกระทั่งครบทุกค่าตามที่ต้องการ ขั้นต่อไปจะเปลี่ยนค่า β_1 เพื่อทำให้มีค่า ρ_{XY} ใหม่ตามที่ศึกษา และเปลี่ยนค่าขนาดตัวอย่างจนกระทั่งครบทุกจำนวนเช่นเดิม หลังจากนั้นจะเปลี่ยนค่าเริ่มต้นตัวใหม่เพื่อที่จะทำให้มีจำนวนค่าสังเกตที่เป็นค่าสูงมากในประชากรเปลี่ยนค่าไปเท่ากับจำนวนที่ต้องการศึกษา แต่ขนาดประชากร ค่า ρ_{XY} และขนาดตัวอย่างที่ใช้เท่าเดิม แล้วกระทำซ้ำในทำนองเดียวกัน จนครบสถานการณ์ที่จะศึกษาในขนาดประชากรหนึ่ง ๆ ซึ่งต่อไปก็จะเปลี่ยนขนาดประชากรใหม่จนกระทั่งครบตามสถานการณ์ต่าง ๆ ที่จะศึกษาในหนึ่งลักษณะของการแจกแจง โดยที่ตอนสุดท้ายของแต่ละสถานการณ์ไม่มีที่คำนวณหาค่าเฉลี่ยของค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณแต่ละตัว เมื่อเสร็จสิ้นไปหนึ่งลักษณะการแจกแจงของตัวแปร Y แล้วขั้นต่อไปก็จะเปลี่ยนลักษณะการแจกแจงของตัวแปร Y ใหม่จนครบทุกการแจกแจง และกระทำในทำนองเดิม โดยในแต่ละสถานการณ์ที่ศึกษาจะกระทำซ้ำตามจำนวนที่กำหนดไว้ในแผนการทดลอง สรุปเป็นผังงานดังรูปที่ 4.1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 4.1 แสดงผังงานสำหรับการหาค่าประสิทธิภาพสัมพัทธ์ของการประมาณ $\hat{Y}_k$ และ $\hat{Y}_{mkt}$; k = 1,2,3และt=1,2,3,4 เทียบกับตัวประมาณ$\hat{Y}_o$



ความหมายของสัญลักษณ์ที่ใช้ในผังงานนี้คือ

R แทนจำนวนครั้งที่ทำซ้ำแล้ว

REP แทนจำนวนซ้ำที่ต้องการ

N แทนขนาดประชากรที่ใช้ศึกษา

n_1 แทนจำนวนหน่วยตัวอย่างที่มีค่าสูงมากที่พบในตัวอย่าง

N_1 แทนจำนวนค่าสังเกตในประชากรที่เป็นค่าสูงมาก

U.R.E. แทนประสิทธิภาพสัมพัทธ์อย่างไม่มีเงื่อนไข (Uncondition relative efficiency)

C.R.E. แทนประสิทธิภาพสัมพัทธ์อย่างมีเงื่อนไข (condition relative efficiency)

4.4 โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัย ทั้งหมดเขียนด้วยภาษาฟอร์แทน IV โดยใช้เครื่อง IBM 370/3031 ซึ่งลักษณะของโปรแกรมแบ่งได้เป็นดังตารางที่ 4.1 สำหรับโปรแกรมหนึ่ง ๆ ที่ใช้คำนวณหาประสิทธิภาพสัมพัทธ์ จะใช้ขนาดประชากร จำนวนของค่าสังเกตในประชากรที่เป็นค่าสูงมาก ค่าสัมประสิทธิ์สหสัมพันธ์ และขนาดตัวอย่าง แต่ละอย่างที่กล่าวมามีค่าเท่ากับ 1 ค่าเท่านั้น สำหรับในแต่ละลักษณะของการแจกแจงของประชากรหนึ่ง ๆ

ตารางที่ 4.1 แสดงลักษณะของโปรแกรมทั้งหมดที่ใช้ในการทำวิจัย

ลำดับที่ของโปรแกรม	ลักษณะการทำงาน	โปรแกรมย่อยที่เรียกใช้
1.	หาค่าคาดหมายของฟังก์ชั่น (n_1) ค่าเริ่มต้นที่ใช้และค่า β_1	โปรแกรมตัวเลขสุ่ม โปรแกรม ลักษณะการแจกแจงของประชากร
2.	สร้างข้อมูลให้มีลักษณะการแจกแจงของประชากรขนาด N ที่มี N_1 และ ρ_{XY} ตามที่ศึกษา	โปรแกรมสร้างตัวเลขสุ่ม โปรแกรมสร้าง ลักษณะการแจกแจงของประชากร
3.	คำนวณหาค่าประสิทธิภาพสัมพัทธ์ในการอนุมานแบบไม่มีเงื่อนไขและการอนุมานแบบมีเงื่อนไข โดยที่ในแบบที่มีเงื่อนไข จะเปลี่ยนค่า n_1 ไปเรื่อย ๆ ตั้งแต่ 2 จนถึง N_1	