สำนักงานคณะกรรมการวิจัยแห่งชาติ
ห้องสมุดงานวิจัย
วันที่ 3 มิ.ย. 2555
เลขทะเบียน 245643
เลขเรียกหนังสือ

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ทำการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์สำหรับข้อมูลที่ถูกเซ็นเซอร์แบบช่วง และศึกษาการปรับค่าเอนเอียงของวิธีการประมาณแบบใช้กราฟ โดยใช้โปรแกรม R เวอร์ชัน 2.9.2 (R Development Core Team, 2009) ในการจำลองข้อมูล ในบทนี้จะกล่าวถึงแผนการดำเนินการวิจัย ขั้นตอนในการดำเนินการวิจัย และขั้นตอนการทำงานของโปรแกรม ซึ่งมีรายละเอียด ดังนี้

3.1 ลักษณะข้อมูล

ข้อมูลที่ศึกษามีลักษณะข้อมูลที่ถูกเซ็นเซอร์แบบช่วง การแบ่งช่วงข้อมูลจะใช้การเรียงค่าจากน้อยไปมาก และแบ่งช่วงข้อมูลเป็น 2 กรณีศึกษา

กรณีศึกษาที่ 1 : กำหนดจำนวนจุดเวลาในการสังเกต (observed time) เป็น 5, 10 และ 20 จุด พบว่าระยะห่างระหว่างเวลาจะเท่ากัน เพื่อดูว่ามีข้อมูลถูกเซ็นเซอร์ในแต่ละช่วงเวลาเท่าใด โดยระยะห่างระหว่างเวลามีขนาดเท่ากับ 1000/จำนวนจุดเวลาในการสังเกต ซึ่งจำนวนข้อมูลที่ศึกษามีค่าไม่เท่ากับ n เนื่องจากศึกษาเฉพาะข้อมูลที่มีค่าไม่เกิน 1000

เช่น ถ้าข้อมูลขนาด n = 100 และจำนวนจุดเวลาในการสังเกตเท่ากับ 5 จุด ระยะห่างระหว่างเวลาแต่ละจุดจะเท่ากับ 1000/5 = 200 หน่วยเวลา

ตัวอย่างการแบ่งช่วงข้อมูลโดยกำหนดจำนวนจุดเวลาในการสังเกต

ข้อมูลที่มีการแจกแจงแบบล็อกนอร์มอลขนาด n = 100 คือ 16 ,17 ,35 ,36 ,46 ,47 ,56 ,57 ,59 ,62 ,63 ,64 ,65 ,66 ,72 ,76 ,77 ,78 ,79 ,81 ,86 ,93 ,100 ,103 ,107 ,108 ,109 ,117 ,119 ,121 ,124 ,126 ,134 ,139 ,140 ,141 ,143 ,145 ,148 ,149 ,153 ,154 ,162 ,170 ,184 ,185 ,194 ,210 ,217 ,218 ,223 ,224 ,233 ,237 ,240 ,242 ,245 ,247 ,248 ,252 ,255 ,258 ,267 ,271 ,296 ,303 ,304 ,312 ,323 ,329 ,331 ,333 ,337 ,338 ,353 ,362 ,366 ,367 ,368 ,379 ,380 ,398 ,414 ,445 ,509 ,527 ,561 ,656 ,664 ,721 ,764 ,801 ,869 ,887 ,915 ,919 ,922 ,1302 ,1404 ,1440

และจำนวนจุดเวลาในการสังเกตเท่ากับ 5 จุด ระยะห่างระหว่างเวลาแต่ละจุดจะเท่ากับ 200 หน่วยเวลา

ซึ่งสามารถแสดงการแบ่งช่วงข้อมูลโดยกำหนดจำนวนจุดเวลาในการสังเกตได้ในตารางที่ 3.1

ตารางที่ 3.1 แสดงการแบ่งข้อมูลเป็นช่วงเมื่อทำการแบ่งข้อมูลขนาด n = 100 ด้วยวิธีกำหนดจำนวนจุดเวลาในการสังเกตเท่ากับ 5 จุด

I	ระยะห่างระหว่างเวลา	จำนวนข้อมูลที่ถูกเซ็นเซอร์
1	0 - 200	47
2	201 - 400	35
3	401 - 600	5
4	601 - 800	4
5	801 - 1000	6

กรณีศึกษาที่ 2 : กำหนดเปอร์เซ็นต์การเซ็นเซอร์ของข้อมูล (censoring rate : CR) เป็น 5%, 10% และ 20% พบว่าระยะห่างระหว่างเวลาจะไม่เท่ากันขึ้นอยู่กับเปอร์เซ็นต์การเซ็นเซอร์ของข้อมูล เพื่อดูว่าเมื่อข้อมูลถูกเซ็นเซอร์จะอยู่ที่เวลาเท่าใด โดยระยะห่างระหว่างข้อมูลมีขนาดเท่ากับ n*เปอร์เซ็นต์การเซ็นเซอร์ของข้อมูล จำนวนข้อมูลที่ศึกษามีค่าเท่ากับ n และข้อมูลถูกเซ็นเซอร์แบบช่วงทั้งหมดเท่ากับ n/ระยะห่างระหว่างข้อมูล

เช่น ถ้าข้อมูลมีขนาดเท่ากับ n = 100 และเปอร์เซ็นต์การเซ็นเซอร์ของข้อมูลเท่ากับ 20% ระยะห่างระหว่างเวลาแต่ละจุดจะเท่ากับ100*(20/100) = 20 ข้อมูล และมีข้อมูลถูกเซ็นเซอร์แบบช่วงทั้งหมดเท่ากับ100/20 = 5 ช่วง

ตัวอย่างการแบ่งช่วงข้อมูลโดยกำหนดเปอร์เซ็นต์การเซ็นเซอร์ของข้อมูล

ข้อมูลที่มีการแจกแจงแบบล็อกนอร์มอลขนาด n = 100 คือ 16 ,17 ,35 ,36 ,46 ,47 ,56 ,57 ,59 ,62 ,63 ,64 ,65 ,66 ,72 ,76 ,77 ,78 ,79 ,81 ,86 ,93 ,100 ,103 ,107 ,108 ,109 ,117 ,119 ,121 ,124 ,126 ,134 ,139 ,140 ,141 ,143 ,145 ,148 ,149 ,153 ,154 ,162 ,170 ,184 ,185 ,194 ,210 ,217 ,218 ,223 ,224 ,233 ,237 ,240 ,242 ,245 ,247 ,248 ,252 ,255 ,258 ,267 ,271 ,296 ,303 ,304 ,312 ,323 ,329 ,331 ,333 ,337 ,338 ,353 ,362 ,366 ,367 ,368 ,379 ,380 ,398 ,414 ,445 ,509 ,527 ,561 ,656 ,664 ,721 ,764 ,801 ,869 ,887 ,915 ,919 ,922 ,1302 ,1404 ,1440

และเปอร์เซ็นต์การเซ็นเซอร์ของข้อมูลเท่ากับ 20% ระยะห่างระหว่างเวลาแต่ละจุดจะเท่ากับ 20 ข้อมูล และมีข้อมูลถูกเซ็นเซอร์แบบช่วงทั้งหมดเท่ากับ 5 ช่วง

ซึ่งสามารถแสดงการแบ่งช่วงข้อมูลโดยกำหนดเปอร์เซ็นต์การเซ็นเซอร์ของข้อมูลได้ในตารางที่ 3.2

ตารางที่ 3.2 แสดงการแบ่งข้อมูลเป็นช่วงเมื่อทำการแบ่งข้อมูลขนาด n = 100 ด้วยวิธีกำหนดเปอร์เซ็นต์การเซ็นเซอร์ของข้อมูลเท่ากับ 20%

I	ระยะห่างระหว่างข้อมูล	จำนวนข้อมูลที่ถูกเซ็นเซอร์
1	0 – 20	20
2	21 – 40	20
3	41 - 60	20
4	61 - 80	20
5	81 - 100	20

3.2 แผนการดำเนินการวิจัย

การวิจัยครั้งนี้ ได้กำหนดสถานการณ์ต่างๆ ในการศึกษา ดังนี้

1. เมื่อข้อมูลมีการแจกแจงแบบล็อกนอร์มอล กำหนดให้ข้อมูลมีค่าพารามิเตอร์ $\mu = 5.263$ และ $\sigma = 1$

2. เมื่อข้อมูลมีการแจกแจงแบบล็อกโลจิสติก กำหนดให้ข้อมูลมีค่าพารามิเตอร์ $\mu = 3.963$ และ $\sigma = 1$

3. เมื่อข้อมูลมีการแจกแจงแบบไวบูลล์ กำหนดให้ข้อมูลมีค่าพารามิเตอร์ $\eta = 111.428$, $\beta = 0.5$, $\eta = 333.808$, $\beta = 1$ และ $\eta = 577.761$, $\beta = 2$

4. วิธีการประมาณค่าพารามิเตอร์ที่นำมาใช้เปรียบเทียบ มี 3 วิธี ดังนี้

- วิธีภาวะน่าจะเป็นสูงสุด (MLE)
- วิธีแบบใช้กราฟ (GE)
- วิธีแบบใช้กราฟที่ปรับค่าเอนเอียง (BCGE)

5. ขนาดตัวอย่าง (n) ที่ใช้ในการศึกษาเท่ากับ 100, 250, 500 และ 1000

6. ระดับความเชื่อมั่นที่ใช้ในการประมาณค่าแบบช่วง คือ 0.90, 0.95 และ 0.99

7. การสุ่มตัวอย่างในวิธีบูตสแตรปเพื่อหาค่าประมาณแบบช่วงจะกระทำซ้ำ 500 รอบ

8. การจำลองข้อมูลในแต่ละสถานการณ์จะกระทำซ้ำ 5000 รอบ

หมายเหตุ ค่า μ และ η ที่เลือกศึกษาจะให้ค่าเปอร์เซ็นต์ไทล์ที่ 95 มีค่าประมาณ 1000 สำหรับในแต่ละการแจกแจง

3.3 ขั้นตอนในการดำเนินการวิจัย

การดำเนินการวิจัย แบ่งออกเป็น 2 ส่วน คือ การเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์ และศึกษาการปรับค่าเอนเอียงของวิธีการประมาณแบบใช้กราฟ ซึ่งในแต่ละส่วนมีขั้นตอนในการดำเนินการวิจัย ดังนี้

1. การเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์

1.1 สร้างข้อมูลที่มีการแจกแจงล็อกนอร์มอล การแจกแจงล็อกโลจิสติก และการแจกแจงไวบูลล์ ให้มีจำนวนตามขนาดตัวอย่าง (n) ที่กำหนด

1.2 แบ่งข้อมูลออกเป็นช่วงตามลักษณะข้อมูลที่กำหนด

เมื่อทำการแบ่งข้อมูลออกเป็นช่วงตามกรณีศึกษาที่ 1 และ 2 จะสามารถแบ่งข้อมูลออกเป็นช่วงที่ขึ้นกับเวลา ซึ่งสามารถแบ่งเวลาที่อยู่ในรูปของ $(0,\infty)$ ให้เป็น $m+1$ ช่วง ดังนี้

$$(t_0,t_1],(t_1,t_2],\ldots,(t_{m-1},t_m],(t_m,t_{m+1}), \qquad t_0=0 \text{ และ } t_{m+1}=\infty$$

โดยที่ t_m คือ เวลาสุดท้ายที่เก็บค่า

1.3 คำนวณค่าพารามิเตอร์โดยวิธีการประมาณแบบภาวะน่าจะเป็นสูงสุด (MLE) กับวิธีการประมาณแบบใช้กราฟ (GE)

1.3.1 วิธีการประมาณแบบภาวะน่าจะเป็นสูงสุด (MLE)

ฟังก์ชันภาวะน่าจะเป็น (Likelihood Function : L) ที่มีข้อมูลถูกเซ็นเซอร์สามารถเขียนได้ดังนี้ (Meeker & Escobar, 1998)

$$L(\theta;x_1,\ldots,x_n)=\prod_{i=1}^{n}L_i(\theta;x_1,\ldots,x_n)$$

$$=\prod_{i=1}^{m+1}[F(t_i)]^{l_i}[F(t_i)-F(t_{i-1})]^{d_i}[1-F(t_i)]^{r_i}$$

ที่ซึ่ง $n=\sum_{j=1}^{m+1}(d_j+r_j+l_j)$

โดยที่ d คือ จำนวนข้อมูลเซ็นเซอร์แบบช่วง

r คือ จำนวนข้อมูลเซ็นเซอร์ทางขวา

l คือ จำนวนข้อมูลเซ็นเซอร์ทางซ้าย

1.3.2 วิธีการประมาณแบบใช้กราฟ (GE)

การพล็อตกราฟระหว่าง $x_{(i)}$ กับ $F_{0,1}^{-1}(p_{(i)})$ โดยที่ $x_{(i)}$ เป็น sample quantile และ $p_{(i)}$ เป็นลำดับควอนไทล์ของ $x_{(i)}$ ค่าประมาณของ intercept และ slope จากกราฟสามารถนำมาใช้ในการประมาณค่า μ และ σ นั้น สามารถแยกตามการแจกแจงแต่ละการแจกแจงได้ดังนี้ (Lawless, 2003; Meeker & Escobar, 1998)

- การแจกแจงแบบล็อกนอร์มอล

จะพล็อตกราฟระหว่าง $x_{(i)}$ กับ $F_{0,1}^{-1}(p_{(i)})$ ได้ดังนี้

$x_{(i)}$ คือ $\log(t_i)$

$F_{0,1}^{-1}(p_{(i)})$ คือ ฟังก์ชันควอนไทล์ (quantile function) ของ $F_{0,1}(t_i)$

- การแจกแจงแบบล็อกโลจิสติก

จะพล็อตกราฟระหว่าง $x_{(i)}$ กับ $F_{0,1}^{-1}(p_{(i)})$ ได้ดังนี้

$x_{(i)}$ คือ $\log(t_i)$

$F_{0,1}^{-1}(p_{(i)})$ คือ $\log\left[\frac{F(t_i)}{S(t_i)}\right]$

- การแจกแจงแบบไวบูลล์

จะพล็อตกราฟระหว่าง $x_{(i)}$ กับ $F_{0,1}^{-1}(p_{(i)})$ ได้ดังนี้

$x_{(i)}$ คือ $\log(t_i)$

$F_{0,1}^{-1}(p_{(i)})$ คือ $\log[-\log[S(t_i)]]$

โดยที่ $S(t)$ เป็น ฟังก์ชันการรอดชีพ (Survival Function) จะอยู่ในรูปของ $S(t)=1-F(t)$

หมายเหตุ วิธีภาวะน่าจะเป็นสูงสุด และวิธีแบบใช้กราฟ สามารถใช้คำนวณในการแบ่งข้อมูลออกเป็นช่วงตามลักษณะข้อมูลที่ศึกษาได้ทั้ง 2 กรณี

2. การปรับค่าเอนเอียงของวิธีการประมาณแบบใช้กราฟ (GE)

2.1 ใช้วิธีบูตสแตรป สร้างตัวอย่างชุดใหม่จากตัวอย่างสุ่มชุดเดิม เพื่อหาค่าประมาณแบบช่วงของวิธีการประมาณแบบใช้กราฟ (GE) ที่ระดับความเชื่อมั่นที่กำหนด และตรวจสอบว่าช่วงครอบคลุมค่าพารามิเตอร์หรือไม่ โดยจะทำซ้ำจำนวน 500 รอบ

2.2 หาค่าเอนเอียง โดยนำตัวอย่างชุดใหม่ที่ได้จากวิธีบูตสแตรปมาหาค่าเฉลี่ย แล้วไปลบกับค่าประมาณพารามิเตอร์ด้วยวิธีแบบใช้กราฟ (GE)

2.3 นำค่าประมาณพารามิเตอร์ด้วยวิธีแบบใช้กราฟ (GE) มาลบกับค่าเอนเอียง ได้เป็นค่าประมาณพารามิเตอร์ด้วยวิธีแบบใช้กราฟที่ปรับค่าเอนเอียง (BCGE)

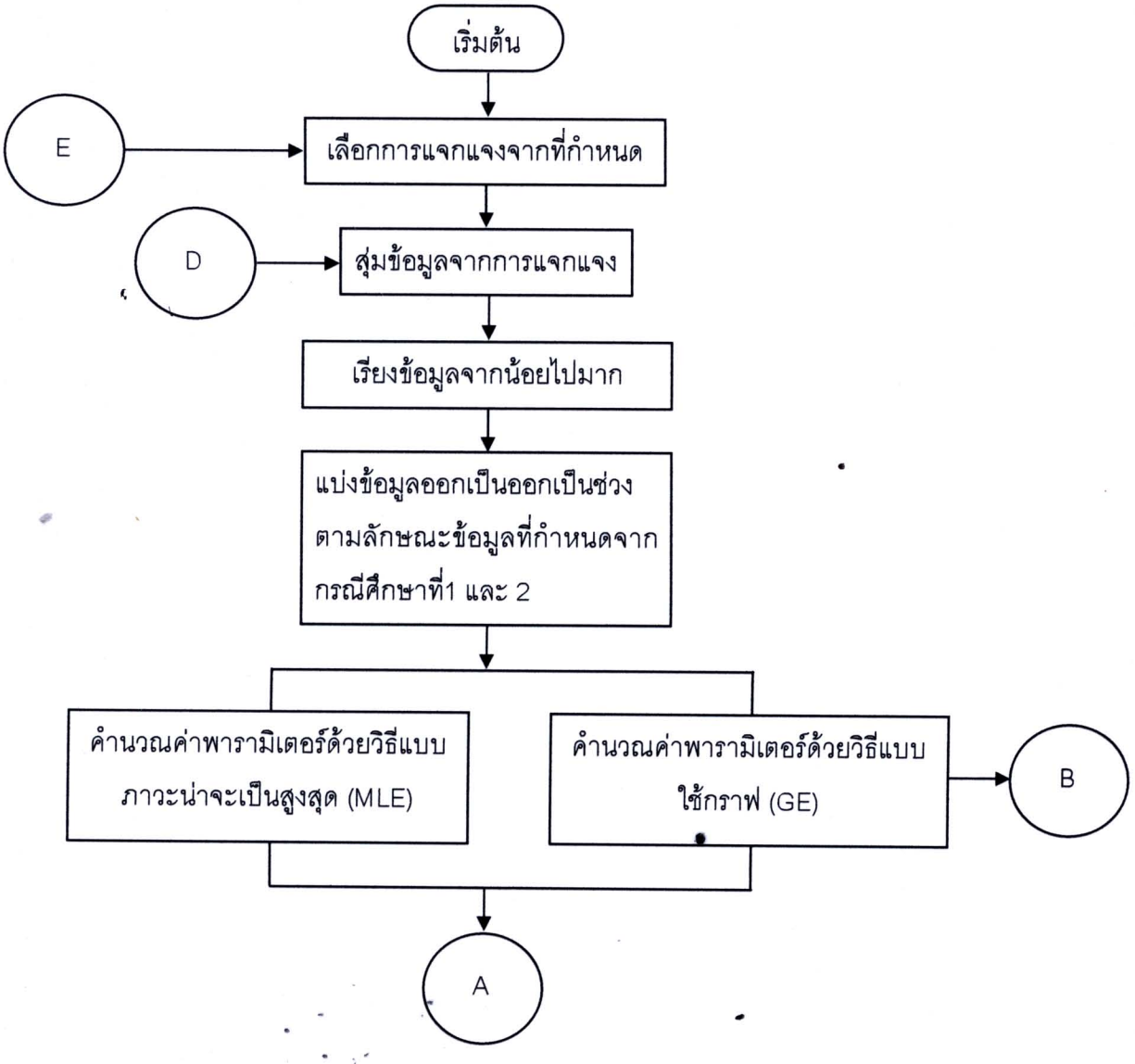
2.4 หาค่าประมาณแบบช่วงของวิธีการประมาณแบบใช้กราฟที่ปรับค่าเอนเอียง (BCGE) โดยนำมาค่าประมาณแบบช่วงของวิธีการประมาณแบบใช้กราฟ (GE) มาลบกับค่าเอนเอียง และตรวจสอบว่าช่วงครอบคลุมค่าพารามิเตอร์หรือไม่

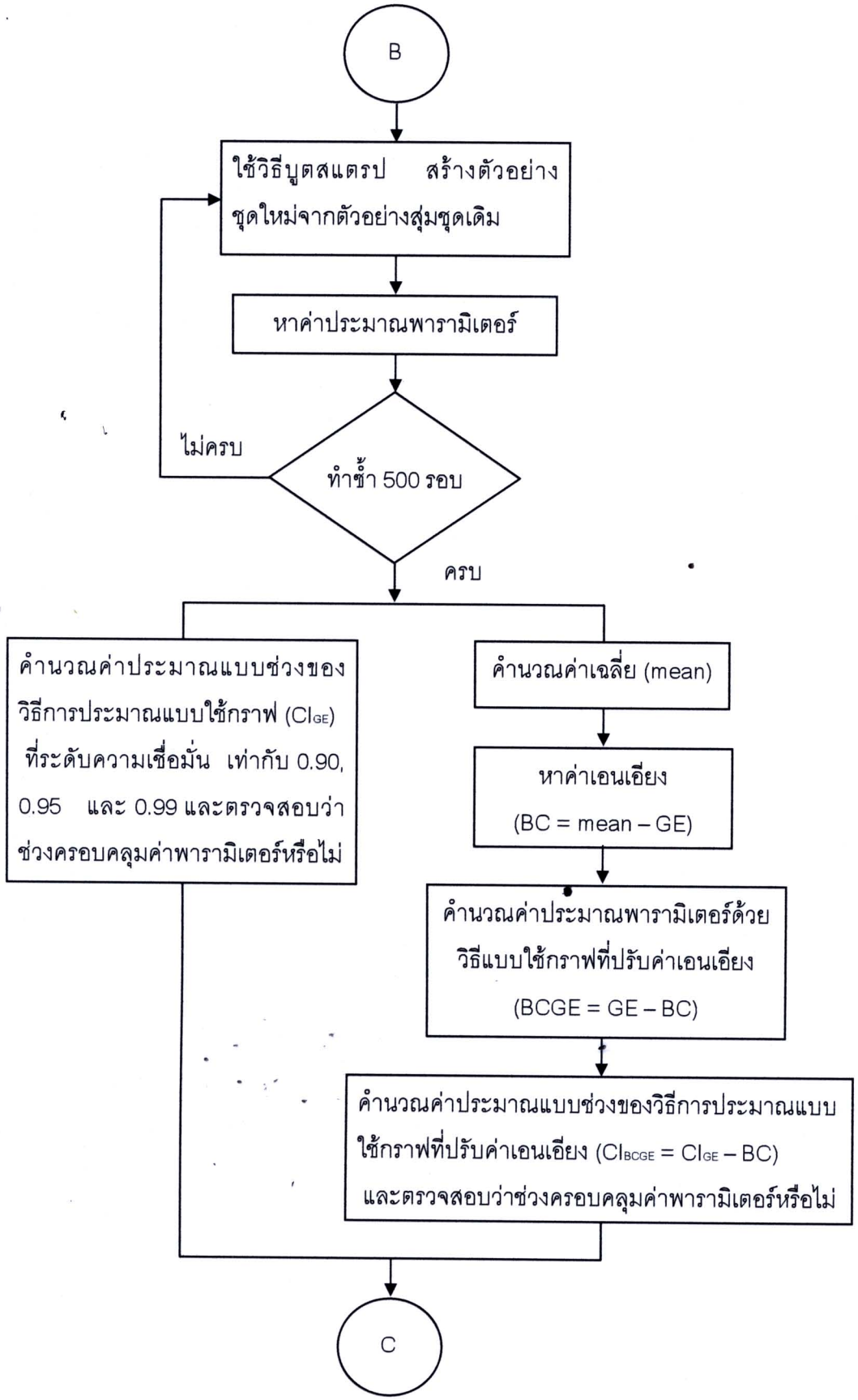
3. ทำซ้ำข้อ 1. - 2. จำนวน 5000 รอบ

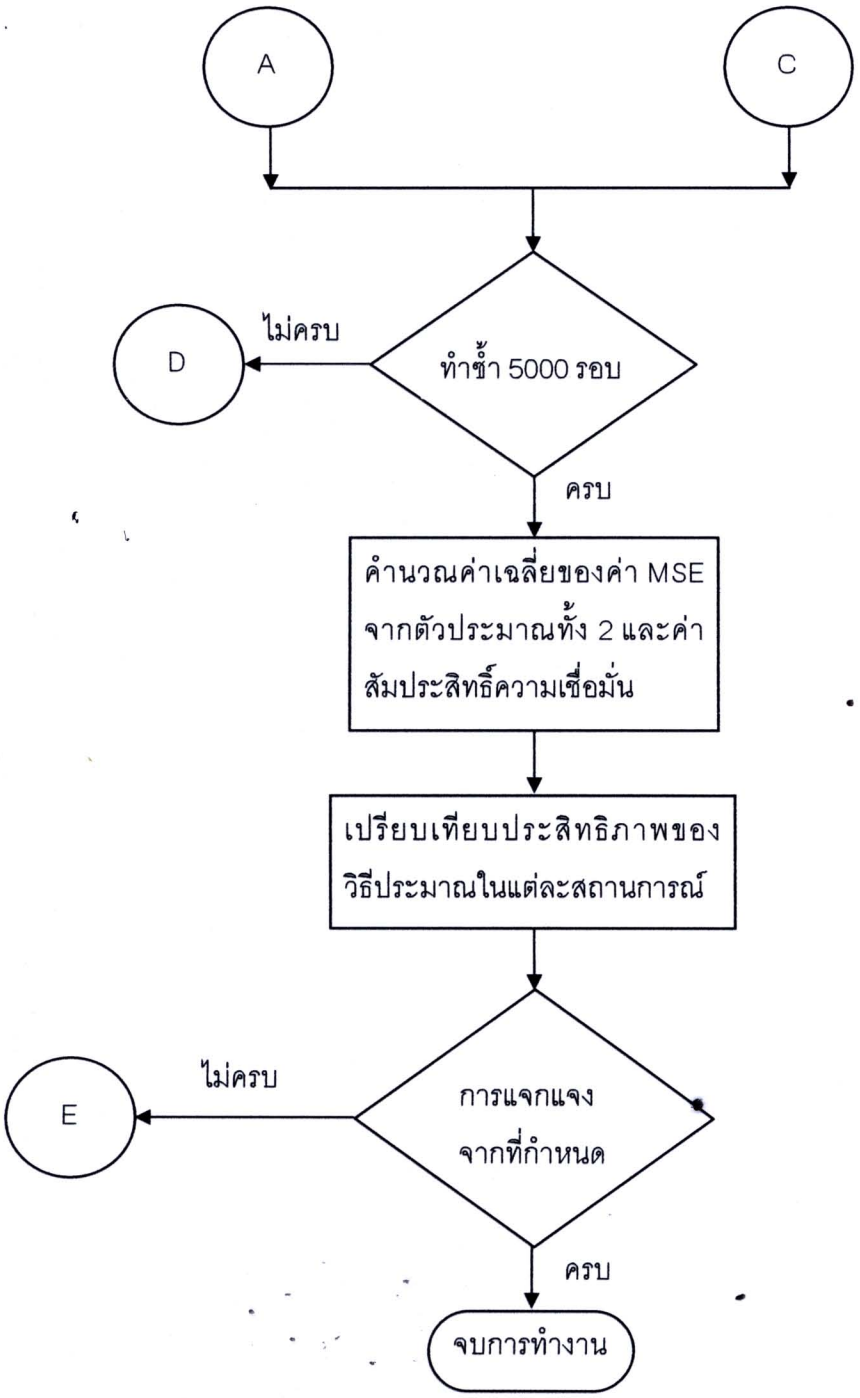
4. คำนวณค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าสัมประสิทธิ์ความเชื่อมั่น

3.4 ขั้นตอนการทำงานของโปรแกรม

ภาพที่ 3.1 แสดงแผนผังขั้นตอนการทำงานของโปรแกรม




B
ใช้วิธีบูตสแตรป สร้างตัวอย่างชุดใหม่จากตัวอย่างสุ่มชุดเดิม
หาค่าประมาณพารามิเตอร์
ทำซ้ำ 500 รอบ
ไม่ครบ
ครบ
คำนวณค่าประมาณแบบช่วงของวิธีการประมาณแบบใช้กราฟ (CI_{GE}) ที่ระดับความเชื่อมั่น เท่ากับ 0.90, 0.95 และ 0.99 และตรวจสอบว่าช่วงครอบคลุมค่าพารามิเตอร์หรือไม่
คำนวณค่าเฉลี่ย (mean)
หาค่าเอนเอียง (BC = mean – GE)
คำนวณค่าประมาณพารามิเตอร์ด้วยวิธีแบบใช้กราฟที่ปรับค่าเอนเอียง (BCGE = GE – BC)
คำนวณค่าประมาณแบบช่วงของวิธีการประมาณแบบใช้กราฟที่ปรับค่าเอนเอียง ($CI_{BCGE} = CI_{GE} – BC$) และตรวจสอบว่าช่วงครอบคลุมค่าพารามิเตอร์หรือไม่
C


A
C
ทำซ้ำ 5000 รอบ
ไม่ครบ
D
ครบ
คำนวณค่าเฉลี่ยของค่า MSE จากตัวประมาณทั้ง 2 และค่าสัมประสิทธิ์ความเชื่อมั่น
เปรียบเทียบประสิทธิภาพของวิธีประมาณในแต่ละสถานการณ์
การแจกแจงจากที่กำหนด
ไม่ครบ
E
ครบ
จบการทำงาน