บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การศึกษาวิธีการประมาณการแจกแจงเป็นหัวข้อหนึ่งของงานวิจัยทางสถิติที่ยังเป็นที่น่าสนใจอยู่ โดยเฉพาะการศึกษาเกี่ยวกับเวลาการอยู่รอด (Survival time) ซึ่งพบได้ในงานวิจัยด้านต่างๆ เช่น งานด้านประกันภัย สนใจศึกษาเวลาการคงอยู่ของกรมธรรม์ งานด้านการแพทย์ สนใจศึกษาเวลาที่คนไข้เริ่มได้รับการรักษาจนกระทั่งเสียชีวิต และงานด้านอุตสาหกรรม สนใจศึกษาเวลาที่เครื่องจักรทำงานเป็นปกติจนกระทั่งเครื่องจักรเกิดการขัดข้อง ซึ่งข้อมูลลักษณะนี้สามารถจัดให้อยู่ในรูปของการแจกแจงได้หลายแบบด้วยกัน เช่น การแจกแจงแบบชี้กำลัง การแจกแจงแบบแกมมา และการแจกแจงแบบไวบูลล์ เป็นต้น ในที่นี้จะขอกล่าวถึงการแจกแจงแบบล็อกนอร์มอล, การแจกแจงแบบล็อกโลจิสติก และการแจกแจงแบบไวบูลล์ โดยที่การแจกแจงทั้ง 3 การแจกแจง สามารถใช้อธิบายในสถานการณ์เกี่ยวกับอายุการใช้งานหรือความทนทานของวัตถุ โดยวัดที่เวลาเริ่มต้นจนกระทั่งวัตถุนั้นเสีย ซึ่งในทางปฏิบัติการเก็บรวบรวมข้อมูลบางครั้งขนาดตัวอย่างหรือจำนวนค่าสังเกตมีขนาดใหญ่แต่จำนวนการเกิดเหตุการณ์ที่สนใจหรือความล้มเหลว (Failure) มีจำนวนน้อย หรือต้องเสียเวลาในการรอคอยจนกระทั่งเกิดความล้มเหลว ซึ่งเป็นระยะเวลาที่ยาวนานจนไม่สามารถรอคอยให้เกิดความล้มเหลวครบทุกหน่วยตัวอย่างหรือทุกค่าสังเกตได้ นอกจากตัวอย่างขนาดใหญ่และการเสียเวลาแล้วยังจะต้องเสียค่าใช้จ่ายในการเก็บข้อมูลมาก ดังนั้นข้อมูลที่นำมาวิเคราะห์จะมีลักษณะเป็นข้อมูลที่ถูกเซ็นเซอร์ ซึ่งข้อมูลที่ถูกเซ็นเซอร์นั้นมีความสำคัญที่จะต้องนำมาใช้วิเคราะห์เพื่อวางแผนหรือปรับปรุง ดังนั้นเพื่อการตัดสินใจจึงควรทำการประมาณค่าที่ถูกตัดทิ้ง โดยใช้วิธีการที่มีความเหมาะสม จะได้ค่าประมาณที่มีค่าใกล้เคียงกับค่าจริง และมีจำนวนข้อมูลเพียงพอที่จะนำไปศึกษาตามวัตถุประสงค์ เป็นต้นว่าเพื่อสร้างตัวแบบที่เหมาะสม หรือเพื่อศึกษาถึงการแจกแจงของข้อมูล

เนื่องจากในงานวิจัยส่วนใหญ่ มักพบว่าการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดเป็นวิธีที่มีประสิทธิภาพดีกว่าวิธีการประมาณอื่นๆ (Genschel & Meeker, 2010; Olteanu & Freeman, 2010) แต่ในทางปฏิบัติการวิเคราะห์ข้อมูลด้วยวิธีภาวะน่าจะเป็นสูงสุดนั้นทำได้ยาก เพราะวิธีภาวะน่าจะเป็นสูงสุดไม่มีรูปแบบสมการที่แน่นอน ทำให้คำนวณได้ยากและใช้เวลานาน ดังนั้นเราจึงสนใจที่จะศึกษาวิธีการประมาณค่าพารามิเตอร์แบบใช้กราฟที่มีความสะดวกและง่ายต่อการคำนวณมากกว่าวิธีภาวะน่าจะเป็นสูงสุด และทำการเปรียบเทียบว่าวิธีการประมาณแบบใดมีประสิทธิภาพดีกว่า และควรจะใช้การประมาณแบบใดในแต่ละสถานการณ์

นอกจากนี้ยังศึกษาวิธีการประมาณแบบใช้กราฟที่ปรับค่าเอนเอียงว่าจะมีประสิทธิภาพดีกว่าวิธีแบบใช้กราฟหรือไม่ เพื่อนำมาเป็นแนวทางในการปฏิบัติจริง

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและเปรียบเทียบการประมาณค่าพารามิเตอร์ของการแจกแจงจากข้อมูลที่ถูกเซ็นเซอร์แบบช่วง โดยใช้วิธีภาวะน่าเป็นสูงสุด (Maximum Likelihood Estimation : MLE) และการประมาณด้วยวิธีแบบใช้กราฟ (Graphical Estimation : GE)

2. เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณแบบใช้กราฟที่ปรับค่าเอนเอียง (Bias Correcting) กับวิธีการประมาณแบบใช้กราฟ

ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้กระทำภายใต้ขอบเขต ดังนี้

1. การแจกแจงที่นำมาใช้ในการศึกษา มี 3 การแจกแจง คือ
   1.1 การแจกแจงแบบล็อกนอร์มอล
   1.2 การแจกแจงแบบล็อกโลจิสติก
   1.3 การแจกแจงแบบไวบูลล์
2. ขนาดตัวอย่าง (n) ที่ใช้ในการศึกษาเท่ากับ 100, 250, 500 และ 1000
3. ข้อมูลที่ศึกษามีลักษณะเป็นข้อมูลที่ถูกเซ็นเซอร์แบบช่วง
4. วิธีการประมาณค่าพารามิเตอร์ที่นำมาใช้เปรียบเทียบ มี 3 วิธี ดังนี้
   4.1 วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation : MLE)
   4.2 วิธีแบบใช้กราฟ (Graphical Estimation : GE)
   4.3 วิธีแบบใช้กราฟที่ปรับค่าเอนเอียง (Bias Corrected Graphical Estimation : BCGE)
5. การศึกษาครั้งนี้ประมวลผลโดยโปรแกรม R เวอร์ชัน 2.9.2 (R Development Core Team, 2009) โดยการจำลองข้อมูลในแต่ละสถานการณ์จะกระทำซ้ำ 5000 รอบ

เกณฑ์ที่ใช้ในการตัดสินใจ

เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพของตัวประมาณ

1. เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ เป็นการประมาณค่าแบบจุด โดยจะพิจารณาเปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) แต่เนื่องจากพารามิเตอร์ที่ต้องการประมาณค่ามี 2 พารามิเตอร์ ดังนั้นจะต้องหาค่าเฉลี่ยของค่า MSE ของตัวประมาณทั้ง 2

กำหนดให้ θ_p เป็นค่าจริงของพารามิเตอร์ เมื่อ $p=1,2$

$\hat{\theta}_{pi}$ เป็นค่าประมาณของพารามิเตอร์

θ_1 เป็น location parameter

θ_2 เป็น scale parameter

$$MSE(\theta_p)=\frac{\sum_{i=1}^{M}(\hat{\theta}_{pi}-\theta_p)^2}{M}$$

เมื่อ M คือ จำนวนครั้งของการทดลองในแต่ละสถานการณ์ และค่าเฉลี่ยของค่า MSE ของตัวประมาณทั้ง 2 เรียกว่า ค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย โดยพิจารณาให้พารามิเตอร์ที่ต้องการประมาณค่าทั้ง 2 มีน้ำหนักเท่าๆกัน คือพารามิเตอร์ที่ต้องการประมาณค่าทั้ง 2 มีความสำคัญไม่ต่างกัน

$$\overline{MSE}=\frac{MSE(\theta_1)+MSE(\theta_2)}{2}$$

ถ้าวิธีการประมาณค่าที่มีค่า $\overline{MSE}$ ต่ำกว่าจะถือว่าการประมาณค่าวิธีนั้นมีประสิทธิภาพดีกว่า สำหรับการประมาณค่าแบบจุดในสถานการณ์นั้นๆ

2. การปรับค่าเอนเอียงของวิธีการประมาณแบบใช้กราฟ จะพิจารณาจากการประมาณค่าแบบช่วง โดยเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่น (Confidence Coefficient)

$$\text{ค่าสัมประสิทธิ์ความเชื่อมั่น}=\frac{\text{จำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นครอบคลุมค่าพามิเตอร์ }\theta}{M}$$

ถ้าวิธีการประมาณค่าที่มีค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าจะถือว่าการประมาณค่าวิธีนั้นมีประสิทธิภาพดีกว่า สำหรับการประมาณค่าแบบช่วงในสถานการณ์นั้นๆ

คำจำกัดความที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้มีคำจำกัดความที่ใช้ในการวิจัย ดังนี้

1. ข้อมูลที่ถูกเซ็นเซอร์ (Censored data) คือ ข้อมูลที่ไม่ทราบค่าจริงของข้อมูล แต่รู้ช่วงของค่าที่เป็นไปได้เท่านั้น

2. ความเอนเอียง (Bias) คือ ค่าที่ใช้วัดค่าเฉลี่ยของตัวสถิติที่ได้ห่างจากฟังก์ชันพารามิเตอร์ที่สนใจมากน้อยเพียงใด นอกจากนั้นยังบอกทิศทางได้ด้วยว่า ตัวสถิติที่ได้มีค่าสูงหรือต่ำกว่าพารามิเตอร์ โดย $\mathbf{Bias} = \mathrm{E}(\hat{\theta}) - \theta$ เมื่อ $\hat{\theta}$ เป็นค่าประมาณของ θ

3. ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณกับค่าจริง โดย $\mathbf{MSE} = \mathrm{E}[\hat{\theta} - \theta]^2$ เมื่อ $\hat{\theta}$ เป็นค่าประมาณของ θ

4. ช่วงความเชื่อมั่น (Confidence interval) คือ ช่วงของการประมาณค่า เป็นช่วงที่อยู่รอบจุดของค่าประมาณ นอกจากนั้นยังบอกความเชื่อมั่นว่ามีค่าพารามิเตอร์อยู่ในช่วงนี้

ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับในงานวิจัยครั้งนี้ คือ

1. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณโดยวิธีภาวะน่าจะเป็นสูงสุดและตัวประมาณโดยใช้กราฟสำหรับข้อมูลที่ถูกเซ็นเซอร์แบบช่วง

2. เพื่อเป็นแนวทางในการเลือกใช้วิธีการประมาณค่าพารามิเตอร์สำหรับข้อมูลที่ถูกเซ็นเซอร์แบบช่วง ซึ่งมีการแจกแจงแบบล็อกนอร์มอล การแจกแจงแบบล็อกโลจิสติก และการแจกแจงแบบไวบูลล์

3. เพื่อเป็นแนวทางในการศึกษาทางด้าน Reliability ต่อไป

วิธีดำเนินการวิจัย

1. การเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์

1.1 ศึกษาค้นคว้าเอกสารและข้อมูลเกี่ยวกับการประมาณค่าพารามิเตอร์จากข้อมูลที่ถูกเซ็นเซอร์แบบช่วง

1.2 สร้างข้อมูลที่มีการแจกแจง ลักษณะข้อมูล และขนาดตัวอย่าง (n) ตามที่กำหนด

1.3 คำนวณค่าพารามิเตอร์โดยวิธีการประมาณแบบภาวะน่าจะเป็นสูงสุดกับวิธีการประมาณแบบใช้กราฟ

2. การปรับค่าเอนเอียงของวิธีการประมาณแบบใช้กราฟ

2.1 หาค่าประมาณแบบช่วงสำหรับวิธีการประมาณแบบใช้กราฟ ที่ระดับความเชื่อมั่นเท่ากับ 0.90, 0.95 และ 0.99 และตรวจสอบว่าช่วงครอบคลุมค่าพารามิเตอร์หรือไม่

2.2 หาค่าเอนเอียง

2.3 คำนวณค่าประมาณพารามิเตอร์ด้วยวิธีแบบใช้กราฟที่ปรับค่าเอนเอียง

2.4 หาค่าประมาณแบบช่วงสำหรับวิธีการประมาณแบบใช้กราฟที่ปรับค่าเอนเอียง ที่ระดับความเชื่อมั่นเท่ากับ 0.90, 0.95 และ 0.99 และตรวจสอบว่าช่วงครอบคลุมค่าพารามิเตอร์หรือไม่

3. ทำซ้ำข้อ 1. - 2. จำนวน 5000 รอบ

4. คำนวณค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าสัมประสิทธิ์ความเชื่อมั่น

ลำดับขั้นตอนในการเสนอผลการวิจัย

การวิจัยครั้งนี้ นำเสนอผลการวิจัยโดยเสนอค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าสัมประสิทธิ์ความเชื่อมั่น ในแต่ละสถานการณ์ นำเสนอในรูปแบบกราฟและตาราง โดยการนำเสนอผลการวิจัยแบ่งออกเป็น 2 ส่วน

ส่วนที่ 1 นำเสนอผลการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์ 3 วิธี คือ วิธีการประมาณแบบวิธีภาวะน่าจะเป็นสูงสุด วิธีการประมาณแบบใช้กราฟ และวิธีการประมาณแบบใช้กราฟที่ปรับค่าเอนเอียง โดยจะนำเสนอค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ส่วนที่ 2 นำเสนอผลการเปรียบเทียบประสิทธิภาพวิธีการประมาณแบบใช้กราฟที่ปรับค่าเอนเอียงกับวิธีการประมาณแบบใช้กราฟ เพื่อที่จะทดสอบว่า เมื่อปรับค่าเอนเอียง (Bias Correction) แล้ว จะสามารถประมาณค่าพารามิเตอร์ได้ค่าใกล้เคียงค่าจริงมากกว่าหรือไม่ โดยจะนำเสนอค่าสัมประสิทธิ์ความเชื่อมั่น