

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์ที่ได้จากวิธีการคัดเลือกตัวแบบการถดถอย ในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ ซึ่งวิธีการคัดเลือกสมการถดถอย ที่นำมาใช้ในการสร้างตัวแบบการถดถอยเชิงเส้นพหุคูณ มี 5 วิธี ดังนี้

1. การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของอาโคเคะที่ปรับค่าความเสี่ยง โดยใช้ตัวประมาณข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information)
2. การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_C ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence)
3. การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_T ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence)
4. การคัดเลือกตัวแบบ Cross-Validation (C_p) ที่ใช้สำหรับตัวอย่างขนาดเล็ก หรือ G_n^{CV} ที่ใช้ $C_n^{(R)}$
5. การคัดเลือกตัวแบบโดยใช้สถิติทดสอบเอฟบางส่วน (The Partial F-test Statistic) ด้วยวิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method)

โดย 3 วิธีแรกเป็นวิธีที่พัฒนาต่อเนื่องมาจากข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) และวิธีที่ 4 เป็นวิธีที่พัฒนาต่อเนื่องมาจากวิธี Cross Validation ซึ่งแต่ละวิธีเป็นเครื่องมือในการหาตัวแบบที่ให้ค่าพยากรณ์ที่เหมาะสมที่สุด ในที่นี้ผู้วิจัยได้นำวิธีการทั้ง 4 ที่กล่าวข้างต้น มาเปรียบเทียบกับวิธีการถดถอยแบบขั้นบันได ซึ่งเป็นวิธีการที่พบในโปรแกรมสำเร็จรูปทางสถิติทั่วไป เพื่อศึกษาว่าวิธีการภายใต้แนวทางต่างๆ มีประสิทธิภาพแตกต่างกันหรือไม่ เมื่อขนาดตัวอย่างเล็กถึงปานกลาง เนื่องจากเมื่อขนาดตัวอย่างใหญ่ การวิเคราะห์การถดถอยเชิงเส้นพหุคูณใน 4 วิธีแรก จะเข้าใกล้การวิเคราะห์ด้วยวิธีดั้งเดิมของแนวคิดนั้นๆ ซึ่งมีการคำนวณที่ง่ายกว่า ดังนั้น ในการวิจัยครั้งนี้ผู้วิจัยจึงได้ทำการเปรียบเทียบวิธีการคัดเลือกสมการถดถอยทั้ง 5 วิธีข้างต้น เพื่อศึกษาวิธีการใดที่ให้ค่าพยากรณ์ที่เหมาะสมที่สุด ภายใต้ข้อมูลที่เป็นไปตามข้อตกลงเบื้องต้นในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ ซึ่งมีการกำหนดสถานการณ์ต่างๆ ในการวิจัย ดังนี้

1. จำนวนตัวแปรอิสระ เท่ากับ 3 , 6 , 9 และ 12 ตัว

2. กำหนดให้ความคลาดเคลื่อนสุ่ม มีค่าเฉลี่ยเท่ากับ 0 และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1, 5 และ 10
 3. กำหนดขนาดตัวอย่าง (n) ที่ศึกษาขึ้นกับจำนวนตัวแปรอิสระ ดังนี้
 - 3.1 กรณีจำนวนตัวแปรอิสระ 3 ตัว ขนาดตัวอย่างเป็น 10, 20, 30 และ 40
 - 3.2 กรณีจำนวนตัวแปรอิสระ 6 ตัว ขนาดตัวอย่างเป็น 10, 20, 30 และ 40
 - 3.3 กรณีจำนวนตัวแปรอิสระ 9 ตัว ขนาดตัวอย่างเป็น 15, 25, 35 และ 45
 - 3.4 กรณีจำนวนตัวแปรอิสระ 12 ตัว ขนาดตัวอย่างเป็น 20, 30, 40 และ 50
 4. กำหนดค่าพารามิเตอร์เริ่มต้นในการจำลองข้อมูล
 - 4.1 กรณีตัวแปรอิสระ 3 ตัว ($\beta_0, \beta_1, \beta_2, \beta_3$) เป็น (1,1,1,1)
 - 4.2 กรณีตัวแปรอิสระ 6 ตัว ($\beta_0, \beta_1, \dots, \beta_6$) เป็น (1,1,1,1,1,1)
 - 4.3 กรณีตัวแปรอิสระ 9 ตัว ($\beta_0, \beta_1, \dots, \beta_9$) เป็น (1,1,1,1,1,1,1,1,1)
 - 4.4 กรณีตัวแปรอิสระ 12 ตัว ($\beta_0, \beta_1, \dots, \beta_{12}$) เป็น (1,1,1,1,1,1,1,1,1,1,1,1)
 5. เกณฑ์การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของอาโคเคะ ที่ปรับค่าความเสี่ยงโดยใช้ตัวประมาณของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) กำหนดให้ $\alpha = 3/4$ และ $\varepsilon = 2/5$
 6. เกณฑ์การคัดเลือกตัวแบบวิธีการคัดเลือกตัวแบบ โดยใช้สถิติทดสอบเอฟบางส่วน (The Partial F-test Statistic) ด้วยวิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method) กำหนดระดับนัยสำคัญของการทดสอบเพื่อนำตัวแปรอิสระเข้าสมการเป็น 0.05 และระดับนัยสำคัญของการทดสอบเพื่อนำตัวแปรอิสระออกจากสมการเป็น 0.10
- การวิจัยครั้งนี้ จำลองข้อมูลด้วยโปรแกรม S-plus 2000 เพื่อคำนวณค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย และ ค่าส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย เพื่อเปรียบเทียบประสิทธิภาพในการคัดเลือกตัวแบบของแต่ละวิธี โดยเกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการคัดเลือกตัวแบบการถดถอยวิธีใด มีความถูกต้องและแม่นยำมากที่สุด พิจารณาจากเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Square Error (AMSE)) และเกณฑ์ที่ใช้ประกอบการตัดสินใจ จะใช้เกณฑ์ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Square Error (RDAMSE)) ซึ่งวิธีใดให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด จะเป็นวิธีที่ดีที่สุด ผลการวิจัยได้ข้อสรุปดังนี้

เมื่อขนาดตัวอย่างเล็ก จากการเปรียบเทียบค่า AMSE ส่วนใหญ่พบว่าค่า AMSE เรียงลำดับจากน้อยไปมาก คือ วิธี SR ,วิธี AIC_{π} , วิธี KIC_I , วิธี G_n^{CV} ที่ใช้ $C_n^{(R)}$ และ วิธี KIC_C ตามลำดับ และ วิธี SR กับ วิธี AIC_{π} ให้ผลที่ใกล้เคียงกันมากแสดงว่าวิธี วิธี SR กับวิธี AIC_{π} ให้ค่าพยากรณ์ที่มีความถูกต้องแม่นยำที่สุด ในการคัดเลือกตัวแบบสมการถดถอยเชิงเส้นพหุคูณ แต่เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้นมีแนวโน้มว่าค่า AMSE ของ วิธี AIC_{π} จะน้อยกว่า วิธี SR แสดงว่า เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้นส่งผลให้วิธี AIC_{π} มีแนวโน้มว่าสามารถคัดเลือกตัวแบบที่ดีที่สุดได้ดีกว่า เนื่องจาก วิธี SR ใช้หลักการคัดเลือกตัวแปรผ่านการทดสอบสมมติฐาน แต่ วิธี AIC_{π} พิจารณาจากค่าความแปรปรวน กับ จำนวนพารามิเตอร์ที่ไม่ทราบค่า และสำหรับวิธี KIC_I ให้ค่า AMSE ต่ำกว่า วิธี KIC_C ทุกกรณี และวิธี G_n^{CV} ที่ใช้ $C_n^{(R)}$ ให้ค่าใกล้เคียงวิธี KIC_I กับ วิธี KIC_C

เมื่อขนาดตัวอย่างปานกลาง จากการเปรียบเทียบค่า AMSE พบว่าค่า AMSE เรียงลำดับจากน้อยไปมาก คือ วิธี SR ,วิธี AIC_{π} , วิธี KIC_C , วิธี KIC_I และ วิธี G_n^{CV} ที่ใช้ $C_n^{(R)}$ ตามลำดับ ซึ่ง วิธี SR ,วิธี AIC_{π} , วิธี KIC_C ให้ค่าเท่ากันหรือใกล้เคียงกันมาก นั่นคือ ผลการวิเคราะห์สมการถดถอยเชิงเส้นของ วิธี SR ,วิธี AIC_{π} , วิธี KIC_C ใกล้เคียงกันมาก

เมื่อขนาดตัวอย่างใหญ่ จากการเปรียบเทียบค่า AMSE พบว่าค่า AMSE เรียงลำดับจากน้อยไปมาก คือ วิธี SR ,วิธี AIC_{π} , วิธี KIC_C , วิธี KIC_I และ วิธี G_n^{CV} ที่ใช้ $C_n^{(R)}$ ตามลำดับ ซึ่ง วิธี SR ,วิธี AIC_{π} , วิธี KIC_C และ วิธี KIC_I ให้ค่า AMSE เท่ากัน หรือกล่าวได้ว่า ทั้ง 4 วิธีให้ผลเช่นเดียวกัน

ข้อเสนอแนะ

การวิจัยครั้งนี้ ผู้วิจัยมีข้อเสนอแนะ ดังต่อไปนี้

1. นำการวิจัยครั้งนี้ไปศึกษาในกรณีที่ตัวแปรอิสระมีตัวแปรตามมีประกอบ
2. การวิจัยครั้งนี้ศึกษาเฉพาะกรณีที่ข้อมูลที่นำมาศึกษาเป็นไปตามข้อตกลงเบื้องต้น ในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ ซึ่งในความเป็นจริงส่วนใหญ่จะพบว่าข้อมูลที่นำมาศึกษาไม่เป็นไปตามข้อตกลงเบื้องต้น ซึ่งน่าจะมีการนำเกณฑ์การคัดเลือกตัวแบบทั้ง 5 วิธี มาศึกษาในกรณีดังกล่าว
3. การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของอาโคเคะที่ปรับค่าความเสี่ยง โดยใช้ตัวประมาณข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_C ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_I ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) และเกณฑ์การคัดเลือกตัวแบบ Cross-Validation (C_p) ที่ใช้สำหรับตัวอย่างขนาดเล็ก หรือ G_n^{CV} ที่ใช้ $C_n^{(R)}$ สามารถนำไปประยุกต์ใช้กับตัวแบบอื่น โดยเฉพาะอย่างยิ่ง ตัวแบบอนุกรมเวลา เพื่อศึกษาว่าเกณฑ์การคัดเลือกตัวแบบทั้ง 4 วิธี สามารถคัดเลือกตัวแบบอนุกรมเวลาได้แม่นยำหรือไม่ เมื่อเปรียบเทียบกับวิธีพื้นฐาน
4. การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของอาโคเคะที่ปรับค่าความเสี่ยง โดยใช้ตัวประมาณข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_C ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_I ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) และเกณฑ์การคัดเลือกตัวแบบ Cross-Validation (C_p) ที่ใช้สำหรับตัวอย่างขนาดเล็ก หรือ G_n^{CV} ที่ใช้ $C_n^{(R)}$ นำไปศึกษาและพัฒนาเป็นเกณฑ์การคัดเลือกตัวแบบในการวิเคราะห์ในกรณีที่ตัวแปรตามเป็นตัวแปรเชิงกลุ่ม หรือ ตัวแบบการถดถอยโลจิสติก เป็นต้น