# บทที่ 5

## สรุปผลการวิจัยและข้อเสนอแนะ

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยในสมการเชิงเส้นพหุของทั้ง 5 วิธี เมื่อข้อมูลมีค่าผิดปกติ โดยการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีการ 5 วิธี คือ วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method : OLS) วิธีตัวประมาณ M (M-Estimator Method) และวิธีตัวประมาณ Bounded-Influence (Bounded-Influence Estimator Method : BI Estimator Method) เมื่อใช้เกณฑ์ความแกร่งของ Huber และของ Tukey และวิธีการใดที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) น้อยกว่าจะเป็นวิธีที่มีประสิทธิภาพในการประมาณค่าสัมประสิทธิ์การถดถอยมากกว่า

### ผลสรุปการวิจัย

ผลสรุปจากการเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของการประมาณค่าสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นด้วย วิธีกำลังสองน้อยที่สุด วิธีตัวประมาณ M และวิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Huber และของ Tukey ภายใต้สถานการณ์ที่ข้อมูลเกิดค่าผิดปกติในกรณีต่าง ๆ ซึ่งจำแนกตามชนิดของตัวแปรที่เกิดค่าผิดปกติและตามตำแหน่งของการเกิดค่าผิดปกติ มีดังนี้

1. **กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรตาม**

เมื่อความคลาดเคลื่อน ($\varepsilon$) มีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 3 สำหรับทุกขนาดตัวอย่าง ในสถานการณ์ที่อัตราส่วนปลอมปนของ $\varepsilon$ (PE) เท่ากับ 0.01 วิธีกำลังสองน้อยที่สุด วิธีตัวประมาณ M และวิธีตัวประมาณ BI มีค่า MSE ใกล้เคียงกัน และเมื่อเพิ่มอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เป็น 0.05 และ 0.10 โดยทั่วไปวิธีตัวประมาณ M และวิธีตัวประมาณ BI มีค่า MSE ใกล้เคียงกันและต่ำกว่าวิธีอื่น

เมื่อความคลาดเคลื่อน ($\varepsilon$) มีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 10 สำหรับทุกขนาดตัวอย่าง ในสถานการณ์ที่อัตราส่วนปลอมปนของ $\varepsilon$ (PE)

เท่ากับ 0.01 โดยทั่วไป วิธีตัวประมาณ M และวิธีตัวประมาณ BI มีค่า MSE ใกล้เคียงกันและต่ำที่สุด และเมื่อเพิ่มอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เป็น 0.05 และ 0.10 โดยทั่วไปวิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Tukey มีค่า MSE ต่ำที่สุด

**2. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม**

2.1 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ $x_1$ และตัวแปรตาม

เมื่อความคลาดเคลื่อน ($\varepsilon$) มีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 3 สำหรับทุกขนาดตัวอย่าง ในสถานการณ์ของทุกระดับค่าผิดปกติและทุกอัตราส่วนค่าผิดปกติของตัวแปรอิสระ $x_1$ (VX1 และ PX1) และมีอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เท่ากับ 0.01 โดยทั่วไป วิธีกำลังสองน้อยที่สุด วิธีตัวประมาณ M และวิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Huber มีค่า MSE ใกล้เคียงกันและต่ำที่สุด และเมื่อเพิ่มอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เป็น 0.05 และ 0.10 โดยทั่วไปวิธีตัวประมาณ M มีค่า MSE ต่ำที่สุด

เมื่อความคลาดเคลื่อน ($\varepsilon$) มีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 10 สำหรับทุกขนาดตัวอย่าง ในสถานการณ์ของทุกระดับค่าผิดปกติและทุกอัตราส่วนค่าผิดปกติของตัวแปรอิสระ $x_1$ (VX1 และ PX1) และมีอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เท่ากับ 0.01 โดยทั่วไปวิธีตัวประมาณ M มีค่า MSE ต่ำที่สุด และเมื่อเพิ่มอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เป็น 0.05 และ 0.10 โดยทั่วไปวิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Tukey มีค่า MSE ต่ำที่สุด

2.2 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ $x_2$ และตัวแปรตาม

เมื่อความคลาดเคลื่อน ($\varepsilon$) มีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 3 สำหรับทุกขนาดตัวอย่าง ในสถานการณ์ของทุกระดับค่าผิดปกติและทุกอัตราส่วนค่าผิดปกติของตัวแปรอิสระ $x_2$ (VX2 และ PX2) และมีอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เท่ากับ 0.01 โดยทั่วไป วิธีกำลังสองน้อยที่สุด วิธีตัวประมาณ M และวิธีตัวประมาณ BI มีค่า MSE ใกล้เคียงกัน และเมื่อเพิ่มอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เป็น 0.05 และ 0.10 โดยทั่วไป วิธีตัวประมาณ M มีค่า MSE ต่ำที่สุด

เมื่อความคลาดเคลื่อน ($\varepsilon$) มีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 10 ผลสรุปโดยทั่วไปคล้ายกับกรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ $x_1$ และตัวแปรตาม ยกเว้นที่อัตราส่วนปลอมปนของ $\varepsilon$ (PE) เท่ากับ 0.01 พบว่า วิธีตัวประมาณ M และวิธีตัวประมาณ BI มีค่า MSE ใกล้เคียงกันและต่ำกว่าวิธีอื่น

2.3 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ $x_1$ ตัวแปรอิสระ $x_2$ และตัวแปรตาม

เมื่อความคลาดเคลื่อน ($\varepsilon$) มีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 3 สำหรับทุกขนาดตัวอย่าง ในสถานการณ์ของทุกระดับค่าผิดปกติของตัวแปรอิสระ $x_1$ และตัวแปรอิสระ $x_2$ (VX1 และ VX2) และทุกอัตราส่วนค่าผิดปกติของตัวแปรอิสระ $x_1$ และตัวแปรอิสระ $x_2$ (PX1 และ PX2) และมีอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เท่ากับ 0.01 โดยทั่วไปวิธีกำลังสองน้อยที่สุด และวิธีตัวประมาณ M มีค่า MSE ใกล้เคียงกันและต่ำที่สุด และเมื่อเพิ่มอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เป็น 0.05 และ 0.10 โดยทั่วไปวิธีตัวประมาณ M มีค่า MSE ต่ำที่สุด

เมื่อความคลาดเคลื่อน ($\varepsilon$) มีการแจกแจงแบบปกติปลอมปน ที่สเกลแฟคเตอร์เท่ากับ 10 สำหรับทุกขนาดตัวอย่าง ในสถานการณ์ของทุกระดับค่าผิดปกติของตัวแปรอิสระ $x_1$ และตัวแปรอิสระ $x_2$ (VX1 และ VX2) และทุกอัตราส่วนค่าผิดปกติของตัวแปรอิสระ $x_1$ และตัวแปรอิสระ $x_2$ (PX1 และ PX2) และมีอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เท่ากับ 0.01 โดยทั่วไปวิธีตัวประมาณ M มีค่า MSE ต่ำที่สุด และเมื่อเพิ่มอัตราส่วนปลอมปนของ $\varepsilon$ (PE) เป็น 0.05 และ 0.10 โดยทั่วไป วิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Tukey มีค่า MSE ต่ำที่สุด

**3. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม ณ ตำแหน่งเดียวกัน**

3.1 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ $x_1$ และตัวแปรตาม ณ ตำแหน่งเดียวกัน

3.2 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ $x_2$ และตัวแปรตาม ณ ตำแหน่งเดียวกัน

3.3 กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ $x_1$ หรือตัวแปรอิสระ $x_2$ และตัวแปรตาม ณ ตำแหน่งเดียวกัน

ผลสรุปของทั้ง 3 กรณี คือ สำหรับทุกขนาดตัวอย่าง เมื่อความแปรปรวนของความคลาดเคลื่อน ($\varepsilon$) ที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระ ผิดปกติมีค่าเท่ากับ 9 โดยทั่วไป วิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Huber มีค่า MSE ต่ำที่สุด และเมื่อความแปรปรวนของความคลาดเคลื่อน ($\varepsilon$) ที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระ ผิดปกติ มีค่าเท่ากับ 100 โดยทั่วไป วิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Tukey มีค่า MSE ต่ำที่สุด

ค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยของทุกวิธีมีแนวโน้มลดลงด้วยอัตราที่ลดลง เมื่อขนาดตัวอย่างที่เพิ่มขึ้น และโดยทั่วไปมีแนวโน้มเพิ่มขึ้นด้วยอัตราที่คงที่ เมื่ออัตราส่วนปลอมปนของความคลาดเคลื่อนมีค่าเพิ่มขึ้น

โดยทั่วไปค่า MSE ของการประมาณค่าสัมประสิทธิ์การถดถอยของวิธีกำลังสองน้อยที่สุดมีแนวโน้มเพิ่มขึ้นด้วยอัตราที่คงที่ ส่วนค่า MSE ของวิธีตัวประมาณ M และวิธีตัวประมาณ BI มีแนวโน้มเพิ่มขึ้นด้วยอัตราที่ลดลง เมื่อสเกลแฟคเตอร์และความแปรปรวนของความคลาดเคลื่อนมีค่าเพิ่มขึ้น

จากการสรุปผลการวิจัยข้างต้น เพื่อความสะดวกในการพิจารณาเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยที่เหมาะสมสำหรับข้อมูลที่เกิดค่าผิดปกติในกรณีต่าง ๆ ดังนั้นผู้วิจัยจึงทำการสรุปไว้ในตารางที่ 5.1

**ตารางที่ ๖.1** (ต่อ)

3. ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระ และตัวแปรตาม ณ ตำแหน่งเดียวกัน

| ข้อมูลของตัวแปรอิสระ | ความแปรปรวนของ $\varepsilon$ ที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระผิดปกติ | |
|---|---|---|
| | 9 | 100 |
| 1. ตัวแปรอิสระ $x_1$ มีค่าผิดปกติ<br>VX1 = ปานกลาง, รุนแรง<br>PX1 = 0.05, 0.10 | BI-H | BI-T |
| 2. ตัวแปรอิสระ $x_2$ มีค่าผิดปกติ<br>VX2 = ปานกลาง, รุนแรง<br>PX2 = 0.05, 0.10 | BI-H | BI-T |
| 3. ตัวแปรอิสระ $x_1$ มีค่าผิดปกติ<br>VX1 = ปานกลาง, รุนแรง<br>PX1 = 0.05, 0.10 และ<br>ตัวแปรอิสระ $x_2$ มีค่าผิดปกติ<br>VX2 = ปานกลาง, รุนแรง<br>PX2 = 0.05, 0.10 | BI-H | BI-T |

**หมายเหตุ** OLS หมายถึง วิธีกำลังสองน้อยที่สุด

M-H หมายถึง วิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Huber

M-T หมายถึง วิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Tukey

BI-H หมายถึง วิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Huber

BI-T หมายถึง วิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Tukey

$\varepsilon$ หมายถึง ความคลาดเคลื่อน

CN(PE,9) หมายถึง การแจกแจงแบบปกติปลอมปน ที่มีอัตราส่วนการปลอมปนเท่ากับ PE และมีสเกลแฟคเตอร์เท่ากับ 3

CN(PE,100) หมายถึง การแจกแจงแบบปกติปลอมปน ที่มีอัตราส่วนการปลอมปนเท่ากับ PE และมีสเกลแฟคเตอร์เท่ากับ 10

VX1 และ PX1 หมายถึง ระดับค่าผิดปกติและอัตราส่วนการเกิดค่าผิดปกติของตัวแปรอิสระ $x_1$

VX2 และ PX2 หมายถึง ระดับค่าผิดปกติและอัตราส่วนการเกิดค่าผิดปกติของตัวแปรอิสระ $x_2$

และสามารถสรุปผลโดยรวมได้ดังนี้

1. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรตาม (ความคลาดเคลื่อน)

โดยทั่วไปวิธีตัวประมาณ M เป็นวิธีที่มีประสิทธิภาพมากกว่าวิธีอื่น

เนื่องจาก วิธีตัวประมาณ M เป็น วิธีการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรตาม ซึ่งการที่ข้อมูลมีค่าผิดปกติในตัวแปรตามอาจเป็นผลมาจากการที่ค่าความคลาดเคลื่อนบางค่ามีค่าสูงมากหรือต่ำมากกว่าปกติ การแก้ไขของวิธีตัวประมาณ M คือ จะใช้หลักการลดอิทธิพลของค่าสังเกตที่มีค่าความคลาดเคลื่อนผิดปกติ โดยเงื่อนไขที่จะนำมาลดอิทธิพลของค่าผิดปกติจะขึ้นอยู่กับเกณฑ์ความแกร่งหรือรูปแบบของฟังก์ชัน $\psi$ และในที่นี้ใช้เกณฑ์ความแกร่งของ Huber ซึ่งเหมาะสำหรับกรณีที่ข้อมูลมีค่าผิดปกติที่ไม่รุนแรงและเกณฑ์ความแกร่งของ Tukey ซึ่งเหมาะสำหรับกรณีที่ข้อมูลมีค่าผิดปกติที่รุนแรง

ดังนั้นเมื่อสเกลแฟคเตอร์มีค่าเพิ่มขึ้น จะมีผลให้วิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Tukey มีค่า MSE ต่ำกว่าวิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Huber เนื่องจากการเพิ่มขึ้นของสเกลแฟคเตอร์จะส่งผลให้ค่าความคลาดเคลื่อนบางค่ามีค่าผิดปกติที่รุนแรงมาก ส่วนกรณีที่ความคลาดเคลื่อนเกิดค่าผิดปกติที่ไม่รุนแรงและมีปริมาณน้อย (กรณีที่สเกลแฟคเตอร์เท่ากับ 3 และอัตราส่วนปลอมปนเท่ากับ 0.01) พบว่า โดยทั่วไปค่า MSE ของทั้ง 5 วิธี มีค่าใกล้เคียงกัน ในทางปฏิบัติควรเลือกใช้วิธีกำลังสองน้อยที่สุด เนื่องจากเป็นวิธีที่คำนวณง่าย และสะดวกเมื่อนำไปใช้อนุมานทางสถิติ และในกรณีที่วิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Huber และของ Tukey มีค่า MSE ใกล้เคียงกัน ในทางปฏิบัติควรเลือกใช้วิธีตัวประมาณ M เมื่อใช้เกณฑ์ความแกร่งของ Huber เนื่องจากใช้จำนวนรอบของการลู่เข้า (Convergence) น้อยกว่า

2. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม

โดยทั่วไปวิธีตัวประมาณ M เป็นวิธีที่มีประสิทธิภาพมากกว่าวิธีอื่นและผลสรุปโดยทั่วไปคล้ายกับกรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรตาม (ความคลาดเคลื่อน)

เนื่องจากภายใต้สถานการณ์ที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม ที่ผู้วิจัยจำลองขึ้นนั้น พบว่า ค่าผิดปกติของตัวแปรอิสระอาจจะเป็นค่าผิดปกติที่ดี และโอกาสที่ค่าผิดปกติของตัวแปรอิสระจะเป็นค่าผิดปกติที่ดีมีมาก ซึ่งจะช่วยให้สมการถดถอยเหมาะสมกับข้อมูลมากขึ้นและอาจทำให้ค่าประมาณมีความแม่นยำมากขึ้น

ดังนั้นการเพิ่มขึ้นของระดับค่าผิดปกติและอัตราส่วนค่าผิดปกติของตัวแปรอิสระ จะส่งผลให้โดยทั่วไปค่า MSE ของเกือบทุกวิธีมีค่าลดลง ส่วนโอกาสที่ค่าผิดปกติของตัวแปรตามจะเกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระผิดปกติ (Outliers at High Leverage Points) มีน้อย ดังนั้นโอกาสที่ค่าผิดปกติของตัวแปรตามจะมีอิทธิพลต่อการประมาณค่าสัมประสิทธิ์การถดถอยเช่นเดียวกับกรณีที่ข้อมูลเกิดค่าผิดปกติเฉพาะในตัวแปรตามจึงมีมาก และทำให้โดยทั่วไป วิธีตัวประมาณ BI ให้น้ำหนักแก่ค่าผิดปกติเหล่านี้ไม่เหมาะสม และให้ตัวประมาณที่มีประสิทธิภาพต่ำกว่าวิธีตัวประมาณ M

3. กรณีที่ข้อมูลเกิดค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม ณ ตำแหน่งเดียวกัน

โดยทั่วไป วิธีตัวประมาณ BI เป็นวิธีที่มีประสิทธิภาพมากที่สุด

เนื่องจากโอกาสที่ค่าผิดปกติของตัวแปรตามจะเกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระผิดปกติ (Outliers at High Leverage Points) มีมาก และวิธีตัวประมาณ BI เป็นวิธีที่มีความแกร่งสำหรับกรณีที่ค่าผิดปกติของตัวแปรตามเกิดขึ้น ณ ตำแหน่งที่มีค่าผิดปกติของตัวแปรอิสระ วิธีนี้จะทำการแก้ไขโดยใช้หลักการลดอิทธิพลของค่าผิดปกติ ซึ่งจะพิจารณาจากค่าผิดปกติของตัวแปรตามและของตัวแปรอิสระไปพร้อม ๆ กัน และเงื่อนไขของการลดอิทธิพลของค่าผิดปกติเหล่านี้จะใช้เกณฑ์ความแกร่งเช่นเดียวกับวิธีตัวประมาณ M ดังนั้นวิธีตัวประมาณ BI จะให้น้ำหนักแก่ค่าผิดปกติเหล่านี้อย่างเหมาะสมและให้ตัวประมาณค่าสัมประสิทธิการถดถอยที่มีประสิทธิภาพมากกว่าวิธีตัวประมาณ M และวิธีกำลังสองน้อยที่สุด และกรณีที่ค่าผิดปกติที่เกิดขึ้นมีความรุนแรงมากขึ้น (ความแปรปรวนของ $\varepsilon$ ที่เกิดขึ้น ณ ตำแหน่งที่มีค่าของตัวแปรอิสระผิดปกติ มีค่ามากขึ้น) วิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Tukey มีค่า MSE ต่ำกว่าวิธีตัวประมาณ BI เมื่อใช้เกณฑ์ความแกร่งของ Huber เนื่องจากเกณฑ์ความแกร่งของ Tukey มีความแกร่งต่อค่าผิดปกติที่รุนแรง

เมื่อพิจารณาค่า MSE ของตัวประมาณค่าสัมประสิทธิ์การถดถอยของแต่ละตัวในกรณีที่ข้อมูลมีค่าผิดปกติในกรณีต่าง ๆ พบว่า

กรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรตาม จะทำให้ค่า MSE ของตัวประมาณสัมประสิทธิ์การถดถอยของทุกตัวมีค่าเพิ่มขึ้นด้วยอัตราส่วนที่ใกล้เคียงกัน

กรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรอิสระ จะทำให้ค่า MSE ของตัวประมาณสัมประสิทธิ์การถดถอยของทุกตัวมีค่าเพิ่มขึ้น โดยเฉพาะค่า MSE ของตัวประมาณสัมประสิทธิ์

การถดถอยของตัวแปรอิสระที่มีค่าผิดปกติจะมีเปอร์เซ็นต์การเพิ่มขึ้นที่สูงกว่าค่า MSE ของตัวประมาณสัมประสิทธิ์การถดถอยตัวอื่น

กรณีที่ข้อมูลมีค่าผิดปกติในตัวแปรอิสระและตัวแปรตาม จะทำให้ค่า MSE ของตัวประมาณค่าสัมประสิทธิ์การถดถอยมีค่าเพิ่มขึ้น

**ข้อเสนอแนะ**

เพื่อเป็นแนวทางในการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นพหุ เมื่อข้อมูลมีค่าผิดปกติ และเพื่อเป็นแนวทางสำหรับการศึกษาวิจัยนอกเหนือจากการวิจัยครั้งนี้ ดังนั้นผู้วิจัยขอเสนอแนะเป็น 2 ด้าน คือ

1. **ด้านการนำไปใช้ประโยชน์**

ในทางปฏิบัติของการวิเคราะห์การถดถอย เมื่อผู้วิเคราะห์ได้รับข้อมูลของตัวแปรตามและตัวแปรอิสระ ขั้นแรกควรทำการตรวจสอบว่า ข้อมูลของตัวแปรแต่ละตัวมีค่าผิดปกติหรือไม่ ซึ่งอาจใช้วิธีการเขียนกราฟแบบ กล่อง (Box Plot) หรือพิจารณาค่า z-scores (ศึกษารายละเอียดได้จาก Statistics for Business and Economics. 5th ed., David R. Anderson, Dennis J. Sweeney and Thomas A. Williams., 1993) เป็นต้น

ถ้าผลของการตรวจสอบพบว่า มีค่าผิดปกติและผู้วิเคราะห์ทราบว่ามีสาเหตุมาจากความผิดพลาดของการวัด เช่น การบันทึกข้อมูลผิดพลาด ก็ควรจะทำการแก้ไขโดยอาจใช้วิธีตัดค่าผิดปกติออกไป เมื่อข้อมูลมีจำนวนมาก แต่ถ้าข้อมูลมีจำนวนไม่มาก อาจใช้วิธีการปรับแก้ค่านั้นตามความเหมาะสมก่อนที่จะนำข้อมูลไปใช้ในการวิเคราะห์ต่อไป แต่ถ้าไม่ทราบถึงสาเหตุของการเกิดค่าผิดปกติ ผู้วิเคราะห์ไม่ควรที่จะตัดค่าผิดปกติออกทันที เนื่องจากค่าผิดปกติอาจให้ข้อมูลที่เป็นประโยชน์ในการหาสมการถดถอยที่เหมาะสม

ถ้าผู้วิเคราะห์ต้องการนำข้อมูลเหล่านี้มาใช้ในการวิเคราะห์การถดถอยเชิงเส้น วิธีแก้ไข คือ อาจใช้วิธีการประมาณค่าสัมประสิทธิ์การถดถอยที่มีความแกร่ง ซึ่งวิธีการเหล่านี้จะใช้หลักการลดอิทธิพลของค่าสังเกตที่มีค่าผิดปกติลง โดยสร้างสมการถดถอยสำหรับข้อมูลส่วนใหญ่ พร้อมกับทำการตรวจสอบข้อมูลที่ผิดปกติโดยใช้สมการถดถอยที่สร้างเป็นเกณฑ์ และในที่นี้ขอแนะนำวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเมื่อข้อมูลมีค่าผิดปกติในกรณีต่าง ๆ ต่อไปนี้

| ข้อมูลของตัวแปรอิสระ | ข้อมูลของตัวแปรตาม | |
|---|---|---|
| | ปกติ | ผิดปกติ |
| ปกติ | วิธี OLS | วิธีตัวประมาณ M |
| ผิดปกติ | วิธีตัวประมาณ BI | วิธีตัวประมาณ M และ/ หรือ BI |

ถ้าค่าผิดปกติของตัวแปรตามมีความรุนแรง ก็ควรเลือกใช้เกณฑ์ความแกร่งที่มีรูปแบบของฟังก์ชัน $\psi$ เป็นแบบ redesending เช่น เกณฑ์ความแกร่งของ Tukey เป็นต้น นอกจากวิธีการประมาณค่าสัมประสิทธิ์การถดถอยที่เสนอ ผู้วิเคราะห์อาจใช้วิธีการแปลงข้อมูล (Data Transformation) เช่น การใส่ลอกการิทึม (Logarithmic) หรือใส่รากที่สอง (Square Root) เป็นต้น

**2. ด้านการศึกษาวิจัย**

2.1 ควรศึกษาการประมาณค่าพารามิเตอร์ (สัมประสิทธิ์การถดถอย) ในสมการถดถอยเชิงเส้นพหุเมื่อข้อมูลมีค่าผิดปกติ ด้วยวิธีการอื่น เช่น วิธีกำลังสองของมัธยฐานน้อยที่สุด (The Least Median of Squares Estimator) และวิธีตัวประมาณ Bounded-Influence ของ Koenker-Bassett และของ Welsh เป็นต้น

2.2 ควรศึกษาอำนาจการทดสอบของตัวสถิติที่มีพื้นฐานจากวิธีตัวประมาณ M และวิธีตัวประมาณ Bounded-Influence ในการทดสอบสมมุติฐานของสัมประสิทธิ์การถดถอย เมื่อข้อมูลมีค่าผิดปกติ