



บทที่ 5

สรุปผลการวิจัย

การวิจัยนี้เป็นการวิจัยเพื่อเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยโลจิสติก 3 วิธี คือ วิธีภาวะน่าจะเป็นสูงสุด (MLE) วิธีฟังก์ชันจำแนกประเภท (DF) และวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (WLS) วิจัยด้วยข้อมูลที่ไม่ได้จัดกลุ่ม (ungrouped data) ตัวแปรตาม (y) มีการแจกแจงทวินาม (Binomial Distribution) พารามิเตอร์ $n_i = 1$ และ $p_i \in (0,1)$

การวิจัยจะศึกษาในสถานการณ์ต่างๆ ที่กำหนด ดังนี้

1. การแจกแจงของตัวแปรอธิบาย และความแปรปรวนของตัวแปรอธิบาย
2. สัดส่วนของตัวแปรตาม $y=1$ (p)
3. ขนาดตัวอย่าง (N)
4. จำนวนตัวแปรอธิบาย

การสรุปผลว่าวิธีประมาณใดจะเป็นวิธีที่ดีที่สุดจะพิจารณาจาก ค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และใช้ค่า Deviance ประกอบการพิจารณา วิธีประมาณใดที่มีค่า RMSE และ ค่า Deviance ต่ำที่สุด จะเป็นวิธีประมาณที่ดีที่สุด ผลการวิจัยมีข้อสรุปดังนี้

การเปรียบเทียบค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของแต่ละวิธี

ในการวิจัยครั้งนี้ ผลวิจัยแสดงว่า วิธี WLS ไม่ดีเท่าวิธี MLE และวิธี DF ทั้งนี้เนื่องจาก วิธี WLS เป็นวิธีที่ใช้วิเคราะห์ข้อมูลจัดกลุ่ม ที่ต้องมีจำนวนค่าสังเกตในแต่ละระดับ x_i ก่อนข้างมาก ดังนั้นในการเปรียบเทียบค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของแต่ละวิธี จึงขอเปรียบเทียบเฉพาะวิธี MLE และวิธี DF โดยจะพิจารณาตามการแจกแจงของตัวแปรอธิบาย ตามขนาดตัวอย่าง และตามสัดส่วนของตัวแปรตาม $y=1$ ดังนี้

1. กรณีที่ตัวแปรอธิบาย 1 ตัว มีการแจกแจงปกติ
วิธี MLE และ วิธี DF เป็นวิธีที่ดีพอๆ กันสำหรับตัวอย่างทุกขนาด ทุกสัดส่วนของตัวแปรตาม $y=1$ ยกเว้น ตัวอย่างขนาดเล็ก ($N \leq 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ ค่อนข้างสูง ($p > 0.75$) วิธี DF ดีกว่า วิธี MLE
2. กรณีที่ตัวแปรอธิบาย 1 ตัว มีการแจกแจงแบบชี้กำลัง
วิธี MLE เป็นวิธีที่ดีกว่าวิธี DF สำหรับตัวอย่างทุกขนาด ทุกสัดส่วนของตัวแปรตาม $y=1$ ยกเว้น ตัวอย่างที่มีสัดส่วนของตัวแปรตาม $y=1$ สูง ($p > 0.80$) วิธี DF ดีกว่า วิธี MLE
3. กรณีที่ตัวแปรอธิบาย 1 ตัว มีการแจกแจงแบบไวบูลล์
วิธี MLE เป็นวิธีที่ดีกว่าวิธี DF สำหรับตัวอย่างทุกขนาด ทุกสัดส่วนของตัวแปรตาม $y=1$ ยกเว้น ตัวอย่างขนาดเล็ก ($N \leq 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ ค่อนข้างสูง ($p > 0.75$) และ ตัวอย่างขนาดใหญ่ ($N > 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ สูง ($p > 0.80$) วิธี DF ดีกว่า วิธี MLE
4. กรณีที่ตัวแปรอธิบาย 2 ตัว มีการแจกแจงปกติ และการแจกแจงแบบชี้กำลัง
วิธี MLE เป็นวิธีที่ดีกว่าวิธี DF สำหรับตัวอย่างทุกขนาด ทุกสัดส่วนของตัวแปรตาม $y=1$ ยกเว้น ตัวอย่างขนาดเล็ก ($N \leq 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ ค่อนข้างสูง ($p > 0.75$) และ ตัวอย่างขนาดใหญ่ ($N > 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ สูงมาก ($p > 0.85$) วิธี DF ดีกว่า วิธี MLE
5. กรณีที่ตัวแปรอธิบาย 2 ตัว มีการแจกแจงปกติ และการแจกแจงแบบไวบูลล์
วิธี MLE เป็นวิธีที่ดีกว่าวิธี DF สำหรับตัวอย่างทุกขนาด ทุกสัดส่วนของตัวแปรตาม $y=1$ ยกเว้น ตัวอย่างขนาดเล็ก ($N \leq 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ ค่อนข้างสูง ($p > 0.75$) และ ตัวอย่างขนาดใหญ่ ($N > 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ สูงมาก ($p > 0.85$) วิธี DF ดีกว่า วิธี MLE

6. กรณีที่ตัวแปรอธิบาย 2 ตัว มีการแจกแจงแบบชี้กำลัง และการแจกแจงแบบไวบูลล์

วิธี MLE เป็นวิธีที่ดีกว่าวิธี DF สำหรับตัวอย่างทุกขนาด ทุกสัดส่วนของตัวแปรตาม $y=1$ ยกเว้น ตัวอย่างขนาดเล็ก ($N \leq 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ ก่อนข้างสูง ($p > 0.75$) และ ตัวอย่างขนาดใหญ่ ($N > 40$) ที่มีสัดส่วนของตัวแปรตาม $y=1$ สูงมาก ($p > 0.85$) วิธี DF ดีกว่า วิธี MLE

สรุปโดยรวมว่า สำหรับข้อมูลที่ตัวแปรอธิบายมีการแจกแจงปกติ การแจกแจงแบบชี้กำลัง หรือ การแจกแจงแบบไวบูลล์ ทั้งกรณีตัวแปรอธิบาย 1 ตัว หรือ 2 ตัว วิธี MLE เป็นวิธีที่ดีกว่าวิธี DF สำหรับตัวอย่างทุกขนาด ยกเว้นตัวอย่างที่มีสัดส่วนของตัวแปรตาม $y=1$ ก่อนข้างสูง ($p > 0.75$) วิธี DF ดีกว่าวิธี MLE

ปัจจัยที่มีผลต่อค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของแต่ละวิธี

1. สัดส่วนของตัวแปรตาม $y=1$ เมื่อสัดส่วนเพิ่มขึ้น มีผลทำให้ ค่า RMSE ของทุกวิธีลดลง
2. ขนาดตัวอย่าง การเพิ่มขนาดตัวอย่าง มีผลทำให้ ค่า RMSE ของทุกวิธีลดลง แต่ละวิธีมีแนวโน้มแบบต่างๆ ตามสัดส่วนของตัวแปรตาม $y=1$
3. การแจกแจงของตัวแปรอธิบาย ค่า RMSE ของแต่ละวิธีมีแนวโน้มต่างกัน เมื่อตัวแปรอธิบายมีการแจกแจงต่างกัน แต่ค่า RMSE ของแต่ละวิธีมีแนวโน้มในลักษณะเดียวกันเมื่อตัวแปรอธิบายมีการแจกแจงเดียวกัน ที่ความแปรปรวนต่างกัน

ข้อเสนอแนะ

การเลือกใช้วิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยโลจิสติก พอสรุปได้ดังนี้

1. ควรเลือกใช้วิธี MLE ถึงแม้จะไม่ทราบการแจกแจงของข้อมูล ทั้งตัวอย่างขนาดเล็ก และตัวอย่างขนาดใหญ่ แต่ถ้าทราบว่าข้อมูลมีสัดส่วนของตัวแปรตาม $y=1$ สูง ($p > 0.75$) ทั้งตัวอย่างขนาดเล็ก และตัวอย่างขนาดใหญ่ ควรเลือกใช้ วิธี DF

2. สำหรับผู้ที่สนใจที่ต้องการประมาณค่าพารามิเตอร์ของสมการถดถอยโลจิสติกที่มีจำนวนตัวแปรอธิบาย 1 หรือ 2 ตัว โดยไม่คำนึงถึงความผิดพลาดมากนัก สามารถที่จะใช้วิธี DF เพราะเป็นวิธีที่สะดวก คิดได้ด้วยมือ โดยไม่จำเป็นต้องใช้เครื่องคำนวณอิเล็กทรอนิกส์
3. เสนอตัวสถิติ Deviance เป็นเกณฑ์สำคัญในการตัดสินใจว่า วิธีการประมาณใดดีที่สุด โดยเหตุผลที่ว่า ตัวสถิติ Deviance มีค่าขึ้นอยู่กับค่าความน่าจะเป็นที่ประมาณได้เท่านั้น
4. ถ้าผู้สนใจต้องการศึกษาวิธีการประมาณแบบใหม่เพื่อประมาณค่าพารามิเตอร์ในสมการถดถอยโลจิสติก น่าจะทำการวิจัยศึกษาในเรื่อง ตัวประมาณค่าแบบริดจ์ (Ridge Estimators. *Houwelingen* และ *Le Cessie* (1992))