

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยนี้มีวัตถุประสงค์ 2 ประการ ประการแรกคือเปรียบเทียบประสิทธิภาพของสัมประสิทธิ์การตัดสินใจ (R^2 , R^2_i) และสัมประสิทธิ์การตัดสินใจที่ปรับค่า ($R^2_{O,adj,MS}$, $R^2_{I,adj,MS}$, $R^2_{O,adj,LM}$, $R^2_{I,adj,LM}$, $R^2_{I,adj,SAS_{DEV}}$, $R^2_{I,adj,SAS_{AIC}}$) สำหรับตัวแบบการถดถอยโลจิสติก ภายใต้ขนาดตัวอย่าง (n) เท่ากับ 50, 100, 1000 และ 2000 และจำนวนตัวแปรอธิบาย (p) ในตัวแบบการถดถอยเท่ากับ 1, 5 และ 10 ตัว วัตถุประสงค์ประการที่สองคือศึกษาความเอนเอียงสัมพัทธ์ และรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของสัมประสิทธิ์การตัดสินใจ และสัมประสิทธิ์การตัดสินใจที่ปรับค่าที่กล่าวข้างต้น เมื่อกำหนดตัวแบบไม่ถูกต้องใน 3 ลักษณะ คือ (1) ลักษณะที่รูปแบบของตัวแปรอธิบายไม่ถูกต้อง (wrong functional form of the covariate) (2) ลักษณะที่กำหนดฟังก์ชันเชื่อมโยงของตัวแบบไม่ถูกต้อง (misspecified link function) และ (3) ลักษณะที่กำหนดให้ตัวแบบมีตัวแปรอธิบายที่มีความสัมพันธ์กับตัวแปรตามขาดหายไป (missing covariate) ดังนั้นเพื่อให้บรรลุวัตถุประสงค์ดังกล่าว จึงสรุปผลการวิจัยที่ได้จากการดำเนินการตามขอบเขตการศึกษาที่กล่าวไว้ในบทที่ 1 หน้า 8 โดยพิจารณาจากค่าต่างๆ ดังต่อไปนี้

1. ความเอนเอียง (bias) และความเอนเอียงสัมพัทธ์ (relative bias) ของค่าประมาณมัธยฐาน ($\hat{R}^2$) ที่ได้จากการทำซ้ำจำนวน 1,000 ครั้ง
2. ค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (root mean square error, RMSE) ซึ่งคำนวณแยกตามวิธีการที่ใช้ในการคำนวณค่า R^2 และ R^2_{adj} เมื่อ $RMSE_O$ และ $RMSE_I$ คือค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญ และวิธีความควรจะเป็นสูงสุด ตามลำดับ
3. ร้อยละของค่าประมาณสัมประสิทธิ์การตัดสินใจ (R^2 , R^2_i) และสัมประสิทธิ์การตัดสินใจที่ปรับค่า ($R^2_{O,adj,MS}$, $R^2_{I,adj,MS}$, $R^2_{O,adj,LM}$, $R^2_{I,adj,LM}$, $R^2_{I,adj,SAS_{DEV}}$, $R^2_{I,adj,SAS_{AIC}}$) ที่มีค่าอยู่นอกช่วง $[0,1]$ จากการทำซ้ำจำนวน 1,000 ครั้ง

สรุปผลการวิจัย

ผลการวิจัยสามารถสรุปได้ดังนี้

1. ประสิทธิภาพของสัมประสิทธิ์การตัดสินใจ (R_O^2, R_I^2) และสัมประสิทธิ์การตัดสินใจที่ปรับค่า ($R_{O,adj,MS}^2, R_{I,adj,MS}^2, R_{O,adj,LM}^2, R_{I,adj,LM}^2, R_{I,adj,SAS_{DEV}}^2, R_{I,adj,SAS_{AIC}}^2$)

เมื่อกำหนดฟังก์ชันเชื่อมโยงที่แท้จริงเป็นฟังก์ชันโลจิสท พบว่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วยสถิติดีไวเนียนซ์จากวิธีความควรจะเป็นสูงสุดที่ใช้ใน PROC LOGISTIC ของโปรแกรม SAS[®] ($R_{I,adj,SAS_{DEV}}^2$) มีประสิทธิภาพน้อยที่สุดในทุกสถานการณ์ที่ทำการศึกษาลำหรับประสิทธิภาพของค่า R^2 และ R_{adj}^2 อื่นๆ นั้นขึ้นอยู่กับขนาดตัวอย่าง และจำนวนตัวแปรอธิบาย โดยที่เมื่อมีตัวแปรอธิบายเพียงตัวเดียว R^2 และ R_{adj}^2 ทุกตัวมีประสิทธิภาพดีไม่แตกต่างกันในทุกขนาดตัวอย่างที่ทำการศึกษา แต่เมื่อมีจำนวนตัวแปรอธิบายเพิ่มมากขึ้นและมีขนาดตัวอย่างจำนวน 50 และ 100 ค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วย AIC จากวิธีความควรจะเป็นสูงสุดที่ใช้ใน PROC GENMOD ของโปรแกรม SAS[®] ($R_{I,adj,SAS_{AIC}}^2$) มีประสิทธิภาพไม่แตกต่างจากสัมประสิทธิ์การตัดสินใจที่ไม่ปรับค่าที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญ (R_O^2) และวิธีความควรจะเป็นสูงสุด (R_I^2) ซึ่งมีประสิทธิภาพต่ำกว่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าของมิตเติลแบ็คและสเชมเปอร์ (Mittlbuck and Schemper) ($R_{I,adj,MS}^2$) และของเลียโอและแมคกี (Liao and McGee) ($R_{I,adj,LM}^2$) ในกรณีที่ตัวแปรอธิบายจำนวน 10 ตัว และขนาดตัวอย่างเท่ากับ 50 ประสิทธิภาพของสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วยวิธีความควรจะเป็นสูงสุดดีกว่าวิธีกำลังสองน้อยสุดแบบสามัญ และวิธีการของเลียโอและแมคกีมีประสิทธิภาพสูงกว่าวิธีของมิตเติลแบ็คและสเชมเปอร์ แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็น 1,000 และ 2,000 ค่า R^2 และ R_{adj}^2 ทุกตัว (ยกเว้น $R_{I,adj,SAS_{DEV}}^2$) มีประสิทธิภาพไม่แตกต่างทั้งในกรณีที่ตัวแปรอธิบายจำนวน 5 และ 10 ตัว

ดังนั้นจากผลการวิจัยที่ได้ สามารถสรุปได้ว่าค่า $R_{I,adj,SAS_{DEV}}^2$ มีประสิทธิภาพน้อยที่สุดในทุกสถานการณ์ที่ทำการศึกษา และค่า $R_{I,adj,LM}^2$ มีความเหมาะสมที่จะใช้อธิบายสัดส่วนของความผันแปรที่เกิดขึ้นในความน่าจะเป็นที่จะเกิดความสำเร็จของตัวแปรตามที่เกิดเนื่องจากอิทธิพลของตัวแปรอธิบาย หรือใช้ในการคัดเลือกตัวแบบการถดถอยโลจิสติกที่เหมาะสมสำหรับข้อมูลที่ต้องการวิเคราะห์ โดยเฉพาะในกรณีที่ตัวแปรอธิบายจำนวนมาก และขนาดตัวอย่างเล็ก

2. ความเอนเอียงสัมพัทธ์ (relative bias) และรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของสัมประสิทธิ์การตัดสินใจและสัมประสิทธิ์การตัดสินใจที่ปรับค่าเมื่อกำหนดตัวแบบการถดถอยไม่ถูกต้อง

2.1 กำหนดรูปแบบของตัวแปรอธิบายไม่ถูกต้อง โดยกำหนดให้ตัวแปรอธิบาย X_1 มีรูปแบบเป็นกำลังสอง (X_1^2)

เมื่อกำหนดรูปแบบของตัวแปรอธิบายไม่ถูกต้อง พบว่ารูปแบบของตัวแปรอธิบายมีอิทธิพลต่อค่าความเอนเอียงสัมพัทธ์ และรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของ R^2 และ R_{adj}^2 โดยจะขึ้นอยู่กับขนาดตัวอย่าง และประเภทของการคำนวณค่า R^2 และ R_{adj}^2 โดยพบว่าเมื่อกำหนดรูปแบบของตัวแปรอธิบายไม่ถูกต้องค่าความเอนเอียงสัมพัทธ์ของ R^2 และ R_{adj}^2 ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดมีค่ามากกว่าวิธีกำลังสองน้อยสุดแบบสามัญ แต่ค่า RMSE ของ R^2 และ R_{adj}^2 ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดมีค่าน้อยกว่าวิธีกำลังสองน้อยสุดแบบสามัญ และพบว่าเมื่อมีขนาดตัวอย่าง 50 และ 100 ค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วย AIC จากวิธีความควรจะเป็นสูงสุดที่ใช้ใน PROC GENMOD ของโปรแกรม SAS[®] ($R_{l,adj,SAS_{AIC}}^2$) มีความเอนเอียงสัมพัทธ์สูงที่สุด แต่เมื่อมีขนาดตัวอย่าง 1,000 และ 2,000 พบว่าค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วยวิธีความควรจะเป็นสูงสุดของมิตเทิลแบ็คและสเชมเปอร์ (Mittlbach and Schemper) ($R_{l,adj,MS}^2$) และของเลียวและแมคกี (Liao and McGee) ($R_{l,adj,LM}^2$) ให้ค่าความเอนเอียงสัมพัทธ์ที่สูงและใกล้เคียงกับ $R_{l,adj,SAS_{AIC}}^2$

ดังนั้นเมื่อกำหนดให้ตัวแบบการถดถอยมีรูปแบบของตัวแปรอธิบายไม่ถูกต้องสามารถสรุปได้ว่าค่า $R_{l,adj,SAS_{AIC}}^2$ มีความเอนเอียงสัมพัทธ์มากที่สุดเมื่อตัวอย่างมีขนาดเล็ก ส่วนค่า $R_{l,adj,MS}^2$ และ $R_{l,adj,LM}^2$ มีความเอนเอียงสัมพัทธ์มากไม่แตกต่างกับ $R_{l,adj,SAS_{AIC}}^2$ เมื่อตัวอย่างมีขนาดใหญ่ และค่า RMSE ของ R^2 และ R_{adj}^2 ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดมีค่าน้อยกว่าวิธีกำลังสองน้อยสุดแบบสามัญ

2.2 กำหนดฟังก์ชันเชื่อมโยงของตัวแบบไม่ถูกต้อง โดยกำหนดให้ฟังก์ชันเชื่อมโยงไม่ถูกต้อง 2 รูปแบบคือฟังก์ชันคอมพลีเมนทาล็อก-ล็อก (complementary log-log) และฟังก์ชันโพรบิท (probit)

เมื่อกำหนดฟังก์ชันเชื่อมโยงของตัวแบบการถดถอยไม่ถูกต้อง โดยที่ฟังก์ชันโลจิทเป็นฟังก์ชันเชื่อมโยงที่แท้จริง พบว่าฟังก์ชันโพรบิทมีอิทธิพลต่อค่าความเอนเอียงสัมพัทธ์ และรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของ R^2 และ R_{adj}^2 ต่างๆ ไม่แตกต่างจากฟังก์ชันโลจิท ซึ่งเป็นไปตามที่คาดไว้ ทั้งนี้เพราะฟังก์ชันโลจิทและฟังก์ชันโพรบิทมีลักษณะที่คล้ายกันเพียงแต่ฟังก์ชันโพรบิทจะมีค่าที่เข้าใกล้ 0 หรือ 1 ค่อนข้างเร็วกว่าฟังก์ชันโลจิท และฟังก์ชันทั้งสองจะมีลักษณะสมมาตร ในขณะที่ฟังก์ชันคอมพลีเมนทาล็อก-ล็อก นั้นมีลักษณะไม่สมมาตร จึงทำให้ฟังก์ชันคอมพลีเมนทาล็อก-ล็อกมีอิทธิพลต่อค่าความเอนเอียงสัมพัทธ์ และค่า RMSE ของ R^2 และ R_{adj}^2 แตกต่างจากฟังก์ชันโลจิท เนื่องจากความเบ้ (skewness) ของฟังก์ชันเชื่อมโยงจะมีผลกระทบต่อค่าประมาณของตัวแบบการถดถอยโลจิสติก (Czado and Santner, 1992)

สำหรับการกำหนดฟังก์ชันเชื่อมโยงไม่ถูกต้องแบบคอมพลีเมนทาล็อก-ล็อกจะมีอิทธิพลต่อค่าความเอนเอียงสัมพัทธ์และค่า RMSE ของ R^2 และ R_{adj}^2 โดยจะขึ้นอยู่กับจำนวนตัวแปรอธิบาย และประเภทของการคำนวณค่า R^2 และ R_{adj}^2 โดยมีผลทำให้ค่าความเอนเอียงสัมพัทธ์ และ RMSE ของ R^2 และ R_{adj}^2 ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดมีค่ามากกว่าวิธีกำลังสองน้อยสุดแบบสามัญ และพบว่าเมื่อมีตัวแปรอธิบาย 1 ตัว ค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วยวิธีความควรจะเป็นสูงสุดของเลียโอและแมคกี (Liao and McGee) ($R_{l,adj,LM}^2$) มีความเอนเอียงสัมพัทธ์มากที่สุด แต่เมื่อมีตัวแปรอธิบาย 5 และ 10 พบว่าค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วย AIC ที่ใช้ใน PROC GENMOD ของโปรแกรม SAS[®] ($R_{l,adj,SAS_{AIC}}^2$) และค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วยวิธีความควรจะเป็นสูงสุดของมิตเติลแบ็คและสเชมเปอร์ (Mittlbach and Schemper) ($R_{l,adj,MS}^2$) มีความเอนเอียงสัมพัทธ์มากที่สุด

ดังนั้นเมื่อกำหนดให้ตัวแบบการถดถอยมีลักษณะฟังก์ชันเชื่อมโยงไม่ถูกต้องสามารถสรุปได้ว่าค่า $R_{l,adj,SAS_{AIC}}^2$ และ $R_{l,adj,MS}^2$ มีความเอนเอียงสัมพัทธ์มากที่สุด โดยเฉพาะอย่างยิ่งในกรณีที่จำนวนตัวแปรอธิบายมากกว่า 1 ตัว และค่า RMSE ของ R^2 และ R_{adj}^2 ที่คำนวณด้วยวิธีกำลังสองน้อยสุดแบบสามัญมีค่าน้อยกว่าวิธีความควรจะเป็นสูงสุด

2.3 กำหนดให้ตัวแบบการถดถอยมีตัวแปรอธิบายที่มีความสัมพันธ์กับตัวแปรตาม ขาดหายไป

เมื่อกำหนดให้ตัวแปรอธิบายที่มีความสัมพันธ์กับตัวแปรตามขาดหายไปจากตัวแบบ
ที่ศึกษา พบว่าการที่ตัวแบบมีตัวแปรอธิบายขาดหายไปจะมีอิทธิพลต่อค่าความเอนเอียงสัมพัทธ์
และรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของ R^2 และ R_{adj}^2 โดยจะขึ้นอยู่กับ
กับจำนวนตัวแปรอธิบายที่ขาดหายไป ขนาดตัวอย่าง และประเภทของการคำนวณค่า R^2 และ
 R_{adj}^2 คือถ้าตัวแปรอธิบายขาดหายไปเพิ่มจำนวนมากขึ้น ค่าความเอนเอียงสัมพัทธ์ และ
RMSE ของ R^2 และ R_{adj}^2 จะมีค่าเพิ่มสูงขึ้น และพบว่าค่า RMSE ของ R^2 และ R_{adj}^2
ที่คำนวณด้วยวิธีความควรจะเป็นสูงสุดมีค่าน้อยกว่าวิธีกำลังสองน้อยสุดแบบสามัญ แต่ความ
เอนเอียงสัมพัทธ์ของสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วยวิธีความควรจะเป็นสูงสุดมีค่ามากกว่า
วิธีกำลังสองน้อยสุดแบบสามัญ โดยที่สัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วย AIC ที่ใช้ใน
PROC GENMOD ของโปรแกรม SAS[®] ($R_{l,adj,SAS_{AIC}}^2$) จะมีความเอนเอียงสัมพัทธ์มากที่สุดใน
ทุกสถานการณ์ และสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วยวิธีความควรจะเป็นสูงสุดของมิตเทิลแบ็ค
และสเชมเปอร์ (Mittlbach and Schemper) ($R_{l,adj,MS}^2$) และของเลียโอและแมคกี (Liao
and McGee) ($R_{l,adj,LM}^2$) มีความเอนเอียงสัมพัทธ์ที่สูงและใกล้เคียงกับ $R_{l,adj,SAS_{AIC}}^2$ เมื่อมี
ขนาดตัวอย่าง 1,000 และ 2,000

ดังนั้นเมื่อกำหนดให้ตัวแปรอธิบายที่มีความสัมพันธ์กับตัวแปรตามขาดหายไปจาก
ตัวแบบการถดถอย สามารถสรุปได้ว่าค่า $R_{l,adj,SAS_{AIC}}^2$ มีความเอนเอียงสัมพัทธ์สูงสุด เมื่อ
ตัวอย่างมีขนาดเล็ก ส่วนค่า $R_{l,adj,MS}^2$ และ $R_{l,adj,LM}^2$ มีความเอนเอียงสัมพัทธ์ไม่แตกต่างกับ
 $R_{l,adj,SAS_{AIC}}^2$ เมื่อตัวอย่างมีขนาดใหญ่ และค่า RMSE ของ R^2 และ R_{adj}^2 ที่คำนวณด้วย
วิธีความควรจะเป็นสูงสุดมีค่าน้อยกว่าวิธีกำลังสองน้อยสุดแบบสามัญ

ข้อเสนอแนะ

การวิจัยครั้งนี้ ผู้วิจัยมีข้อเสนอแนะดังต่อไปนี้

1. ในการวิเคราะห์การถดถอยโลจิสติก กรณีที่ตัวแปรตามมีลักษณะข้อมูลแบบทวิภาค พบว่าสำหรับตัวแบบที่มีตัวแปรอธิบายจำนวนมาก และขนาดตัวอย่างเล็ก ควรเลือกใช้สัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วยวิธีความควรจะเป็นสูงสุดของเลียโอและแมคกี (Liao and McGee) ($R^2_{l,adj,LM}$) เพื่อใช้อธิบายสัดส่วนของความผันแปรที่เกิดขึ้นในความน่าจะเป็นที่จะเกิดความสำเร็จของตัวแปรตามที่เกิดเนื่องจากอิทธิพลของตัวแปรอธิบาย หรือใช้ในการคัดเลือกตัวแบบการถดถอยโลจิสติกที่เหมาะสมสำหรับข้อมูลที่ต้องการวิเคราะห์
2. ในการวิเคราะห์การถดถอยโลจิสติก กรณีที่ตัวแปรตามมีลักษณะข้อมูลแบบทวิภาค พบว่าตัวอย่างที่เหมาะสมมีขนาดตั้งแต่ 1,000 ขึ้นไป เพราะค่าประมาณของ R^2 และ R^2_{adj} ส่วนใหญ่จะมีค่าเข้าใกล้สัมประสิทธิ์การตัดสินใจที่แท้จริง (R^2_{true}) เมื่อขนาดตัวอย่างเท่ากับ 1,000 และ 2,000
3. จากการวิจัยครั้งนี้ สำหรับสัมประสิทธิ์การตัดสินใจ (R^2) ที่มีค่าน้อยเมื่อทำการปรับค่าสัมประสิทธิ์การตัดสินใจ (R^2_{adj}) พบว่าค่าประมาณของ R^2_{adj} ที่ได้ในบางครั้งให้ค่าที่ติดลบ นั้นแสดงว่าค่าที่นำมาใช้ในการปรับแก้สัมประสิทธิ์การตัดสินใจมีค่ามากเกินไป เพราะฉะนั้นในทางปฏิบัติควรกำหนดค่าที่ติดลบให้มีค่าเป็นศูนย์ ซึ่งหมายความว่าสัดส่วนของความแปรผันของตัวแปรตามไม่สามารถอธิบายได้ด้วยตัวแปรอธิบายที่มีในตัวแบบการถดถอย
4. เมื่อกำหนดตัวแบบการถดถอยโลจิสติกไม่ถูกต้อง พบว่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าด้วย AIC ใน PROC GENMOD ของโปรแกรม SAS[®] ($R^2_{l,adj,SAS_{AIC}}$) จะให้ค่าความเอนเอียงสัมพัทธ์สูงสุด
5. การวิจัยครั้งนี้ศึกษากรณีที่ค่าของตัวแปรตามมีโอกาสเกิดลักษณะที่สนใจเท่ากับ 0.5 ผู้สนใจอาจศึกษากรณีที่ค่าของตัวแปรตามมีโอกาสเกิดลักษณะที่สนใจในอัตราที่แตกต่างกัน
6. การวิจัยครั้งนี้ศึกษากรณีที่ตัวแปรอธิบายในตัวแบบการถดถอยเป็นอิสระต่อกัน (independent) ผู้สนใจอาจศึกษากรณีที่ตัวแปรอธิบายมีภาวะร่วมหลายตัวเชิงเส้นต่อกัน (multicollinearity)