

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

สรุปผลการวิจัย

การวิจัยครั้งนี้ได้ศึกษาถึงผลกระทบของการวิเคราะห์ทางสถิติภายใต้เงื่อนไขการแจกแจงแบบปกติเมื่อข้อมูลที่มีการแจกแจงแบบโลจิสติก โดยส่วนที่ (1) ทำการเปรียบเทียบประสิทธิภาพของตัวสถิติทดสอบที่ใช้สำหรับคัดกรองการแจกแจงแบบโลจิสติกจากการแจกแจงแบบปกติ โดยใช้ตัวสถิติทดสอบ 6 ตัวสถิติ ได้แก่ ตัวสถิติ Kolmogorov-Smirnov, ตัวสถิติ Shapiro Wilk, ตัวสถิติ Anderson Darling, ตัวสถิติ Lilliefors, ตัวสถิติ Cramer Von Mises และตัวสถิติ Chi-square ในส่วนที่ (2) ทำการทดสอบตัวสถิติที่ใช้ในการทดสอบการแจกแจงแบบโคสแควร์ การแจกแจงแบบที และการแจกแจงแบบเอฟ เมื่อข้อมูลเริ่มต้นมีการแจกแจงแบบโลจิสติก ว่าสามารถที่จะพบนัยสำคัญหรือไม่ และส่วนที่ (3) ทำการศึกษาลักษณะของการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติก

การสรุปผลในส่วนที่ (1) ว่าตัวสถิติใดมีประสิทธิภาพในการคัดกรองการแจกแจงแบบโลจิสติกจากการแจกแจงแบบปกติ พิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของการเกิดคลาดเคลื่อนประเภทที่ 1 จากการทดลองเป็นอันดับแรก แล้วจึงพิจารณาค่ากำลังการทดสอบของตัวสถิติเป็นอันดับต่อไป ซึ่งผลสรุปทั้ง 2 ขั้นตอน เป็นดังนี้

1. พิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของการเกิดคลาดเคลื่อนประเภทที่ 1

ตัวสถิติ Kolmogorov-Smirnov, ตัวสถิติ Shapiro Wilk, ตัวสถิติ Anderson Darling, ตัวสถิติ Lilliefors และตัวสถิติ Cramer Von Mises สามารถควบคุมความน่าจะเป็นในการเกิดคลาดเคลื่อนประเภทที่ 1 ได้ในทุกขนาดตัวอย่าง 10, 25, 50 และ 100 ทุกระดับนัยสำคัญ 0.01, 0.05 และ 0.10

ตัวสถิติ Chi-square สามารถควบคุมความน่าจะเป็นในการเกิดคลาดเคลื่อนประเภทที่ 1 ได้ในกรณีต่างๆ ยกเว้นเมื่อขนาดตัวอย่าง 10 ระดับนัยสำคัญ 0.01 ไม่สามารถควบคุมความน่าจะเป็นในการเกิดคลาดเคลื่อนประเภทที่ 1 ได้ ทั้งนี้เนื่องจากขนาดตัวอย่างมีขนาดเล็กมาก ซึ่งส่งผลต่อข้อจำกัดในการแบ่งกลุ่มของตัวสถิติ Chi-square

2. พิจารณาค่ากำลังการทดสอบของตัวสถิติทดสอบ

สำหรับกรณีทั่วไปตัวสถิติ Shapiro Wilk ให้ค่ากำลังการทดสอบสูงที่สุด รองลงมาคือ ตัวสถิติ Anderson Darling ตัวสถิติ Cramer Von Mises ตัวสถิติ Lilliefors ตัวสถิติ Chi-square และ

ตัวสถิติ Kolmogorov – Smirnov ตามลำดับ ยกเว้นกรณีที่ขนาดตัวอย่างเท่ากับ 10 ระดับนัยสำคัญ 0.1 ที่ตัวสถิติ Anderson – Darling ตัวสถิติ Cramer Von Mises และตัวสถิติ Lilliefors ให้ค่าอำนาจการทดสอบสูงกว่าตัวสถิติ Shapiro – Wilk โดยเมื่อขนาดตัวอย่างเพิ่มมากขึ้น ค่ากำลังการทดสอบสูงขึ้น และที่ระดับนัยสำคัญ 0.1 จะให้ค่ากำลังการทดสอบสูงที่สุด รองลงมาคือที่ระดับนัยสำคัญ 0.05 และที่ระดับนัยสำคัญ 0.01 ตามลำดับ

หมายเหตุ ค่ากำลังการทดสอบที่ได้จากการศึกษาครั้งนี้ให้ค่าไม่สูงนัก ทั้งนี้เนื่องจากข้อมูลที่มีการแจกแจงแบบโลจิสติกที่นำมาใช้ไม่ได้ทำการคัดกรองจากการแจกแจงแบบปกติก่อนนำไปศึกษา หากทำการคัดกรองข้อมูลก่อนนำไปศึกษา จะให้ค่ากำลังการทดสอบที่สูงขึ้น

การสรุปผลส่วนที่ (2) การทดสอบตัวสถิติที่ใช้ในการทดสอบการแจกแจงแบบโคสแควร์ การแจกแจงแบบที และการแจกแจงแบบเอฟ เมื่อข้อมูลเริ่มต้นมีการแจกแจงแบบโลจิสติก ว่าสามารถที่จะพบนัยสำคัญหรือไม่ ซึ่งผลสรุปทั้ง 3 การแจกแจง เป็นดังนี้

1. การแจกแจงแบบโคสแควร์

สำหรับการทดสอบการแจกแจงแบบโคสแควร์ เมื่อข้อมูลเริ่มต้นมีการแจกแจงแบบโลจิสติก พบว่า ตัวสถิติที่นำมาใช้ในการทดสอบพบความแตกต่างอย่างมีนัยสำคัญในทุกขนาดตัวอย่าง และทุกระดับนัยสำคัญ โดยค่ากำลังการทดสอบจะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่ม และเมื่อระดับนัยสำคัญเพิ่มขึ้น

2. การแจกแจงแบบที

สำหรับการทดสอบการแจกแจงแบบทีเมื่อข้อมูลเริ่มต้นมีการแจกแจงแบบโลจิสติกนั้น พบว่า ตัวสถิติที่นำมาใช้ในการทดสอบ ไม่พบนัยสำคัญของความแตกต่างในทุกขนาดตัวอย่าง และทุกระดับนัยสำคัญ ค่ากำลังการทดสอบที่ได้มีค่าใกล้เคียงกับระดับนัยสำคัญที่ตั้งไว้ในทุกกรณี

3. การแจกแจงแบบเอฟ

สำหรับการทดสอบการแจกแจงแบบเอฟเมื่อข้อมูลเริ่มต้นมีการแจกแจงแบบโลจิสติก 2 ชุด พบว่า ตัวสถิติที่นำมาใช้ในการทดสอบพบความแตกต่างอย่างมีนัยสำคัญในทุกขนาดตัวอย่าง ทุกรูปแบบที่ทดสอบ และทุกระดับนัยสำคัญ โดยค่ากำลังการทดสอบจะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่ม และเมื่อระดับนัยสำคัญเพิ่มขึ้น

สำหรับการทดสอบการแจกแจงแบบเอฟเมื่อข้อมูลเริ่มต้นมีการแจกแจงแบบโลจิสติกชุดเดียว พบว่า ตัวสถิติที่นำมาใช้ในการทดสอบพบความแตกต่างอย่างมีนัยสำคัญในทุกขนาดตัวอย่าง และทุกระดับนัยสำคัญ โดยค่ากำลังการทดสอบจะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่ม และเมื่อระดับนัยสำคัญเพิ่มขึ้น

การสรุปผลส่วนที่ (3) การศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติก โดยพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่ 1 ในการประมาณค่าสัมประสิทธิ์การถดถอย และทดสอบการแจกแจงของค่าสัมประสิทธิ์การถดถอย ซึ่งผลสรุป เป็นดังนี้

สำหรับกรณีที่ความคลาดเคลื่อนได้จากการสร้างข้อมูลโดยตรง พบว่า ค่าสัมประสิทธิ์การถดถอยเมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติก (ภายใต้เงื่อนไข ε มีการแจกแจงแบบปกติ) ให้ผลสอดคล้องกับค่าจริงในทุกกรณี

การทดสอบการแจกแจงของค่าสัมประสิทธิ์การถดถอย พบว่า ผลรวมเชิงเส้นของสัมประสิทธิ์การถดถอยที่ได้ไม่ได้มีการแจกแจงแบบปกติ ($H_0: \sum \beta_i$ มีการแจกแจงแบบปกติ) ในหลายกรณี ซึ่งทำให้สรุปได้ว่าไม่ได้มีคุณสมบัติที่ว่ามีการแจกแจงแบบปกติหลายตัวแปร โดยเฉพาะกรณีข้อมูลมีขนาดเล็ก แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นปัญหาดังกล่าวจะค่อย ๆ ลดลง ค่าสัมประสิทธิ์การถดถอยที่ได้จากตัวแปรอิสระที่มาจากการแจกแจงแบบ $\text{Exp}(0.023)$ พบปัญหาดังกล่าวมากกว่าค่าสัมประสิทธิ์การถดถอยที่ได้จากตัวแปรอิสระที่มาจากการแจกแจงแบบ $U(-100,100)$ และการแจกแจงแบบ $N(0,1111.56)$

สำหรับกรณีที่ทำการคัดกรองความคลาดเคลื่อนจากการแจกแจงแบบปกติก่อนนำไปทดสอบ พบว่า ค่าสัมประสิทธิ์การถดถอยเมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติก (ภายใต้เงื่อนไข ε มีการแจกแจงแบบปกติ) ให้ผลสอดคล้องกับค่าจริงภายใต้ระดับนัยสำคัญ 0.01 (เมื่อขนาดตัวอย่างมากกว่า 50), 0.05 และ 0.10 กรณีที่ควรระวังในการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติก คือ เมื่อตัวแปรอิสระมีการแจกแจงแบบ $\text{Exp}(0.023)$ ที่ขนาดตัวอย่างน้อยกว่า 50 ซึ่งค่าสัมประสิทธิ์การถดถอยที่ประมาณได้แตกต่างจากค่าสัมประสิทธิ์การถดถอยเริ่มต้นอย่างมีนัยสำคัญ

การทดสอบการแจกแจงของค่าสัมประสิทธิ์การถดถอย พบว่า ผลรวมเชิงเส้นของสัมประสิทธิ์การถดถอยที่ได้ไม่ได้มีการแจกแจงแบบปกติ ($H_0: \sum \beta_i$ มีการแจกแจงแบบปกติ) ในหลายกรณี ซึ่งทำให้สรุปได้ว่าไม่ได้มีคุณสมบัติที่ว่ามีการแจกแจงแบบปกติหลายตัวแปร โดยเฉพาะกรณีข้อมูลมีขนาดเล็ก แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นปัญหาดังกล่าวจะค่อย ๆ ลดลง ค่าสัมประสิทธิ์การถดถอยที่ได้จากตัวแปรอิสระที่มาจากการแจกแจงแบบ $\text{Exp}(0.023)$ พบปัญหาดังกล่าวมากกว่าค่าสัมประสิทธิ์การถดถอยที่ได้จากตัวแปรอิสระที่มาจากการแจกแจงแบบ $U(-100,100)$ และการแจกแจงแบบ $N(0,1111.56)$

ข้อเสนอแนะ

1. การวิจัยครั้งนี้ได้ศึกษาเปรียบเทียบประสิทธิภาพการคัดกรองการແຈແຈແຈแบบโลจิสติกจากการແຈແຈแบบปกติของตัวสถิติที่ใช้กันโดยทั่วไป ซึ่งความสามารถในการคัดกรองไม่สูงมากนัก จึงควรมีการศึกษา และพัฒนาวิธีการคัดกรองการແຈແຈแบบโลจิสติกจากการແຈແຈแบบปกติให้มีประสิทธิภาพสูงยิ่งขึ้นต่อไป
2. การวิจัยครั้งนี้ศึกษาผลกระทบของการประมาณค่าสัมประสิทธิ์การถดถอย เมื่อความคลาดเคลื่อนมีการແຈແຈแบบโลจิสติก ในการวิเคราะห์การถดถอย เท่านั้น ในการวิจัยครั้งต่อไป ควรมีการศึกษาผลกระทบในการวิเคราะห์ทางสถิติด้านอื่นๆ เช่น การวิเคราะห์ความแปรปรวน (ANOVA) เป็นต้น