



บทที่ 5

สรุปผลการ วิจัยและข้อ เสนอแนะ

จากผลการวิจัยโดยใช้ข้อมูลและวิธีการวิเคราะห์ที่ได้กล่าวมาแล้วในบทที่ 2 พอดีสรุปที่สำคัญได้ดังนี้คือ

1. เมื่อศึกษาถึงผลของปัจจัยต่าง ๆ ต่อค่าความคลาดเคลื่อนเฉลี่ยของแต่ละวิธี พบว่าขนาดตัวอย่าง จำนวนตัวแปร ขนาดความสัมพันธ์ระหว่างตัวแปร และ INTERACTION มีผลต่อค่าความคลาดเคลื่อนเฉลี่ย จากนั้นทำการทดสอบต่อไป พบว่า ขนาดตัวอย่าง จำนวนตัวแปร ขนาดความสัมพันธ์ระหว่างตัวแปร INTERACTION ระหว่างจำนวนตัวแปรและขนาดความสัมพันธ์ระหว่างตัวแปร INTERACTION ระหว่างจำนวนตัวแปรและขนาดตัวอย่าง มีผลทำให้ค่าความคลาดเคลื่อนเฉลี่ยของวิธีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญ รวมทั้งวิธีวิเคราะห์ความถดถอย พหุเชิงเส้นตัดแปลง และวิธีวิเคราะห์ส่วนประกอบหลัก และ INTERACTION ระหว่างขนาดตัวอย่างและขนาดความสัมพันธ์ระหว่างตัวแปรมีผลทำให้ค่าความคลาดเคลื่อนเฉลี่ยของวิธีวิเคราะห์ความถดถอยพหุเชิงเส้นตัดแปลง แตกต่างกันอย่างมีนัยสำคัญ รวมทั้งวิธีวิเคราะห์ส่วนประกอบหลักด้วย

2. เมื่อศึกษาถึงความแตกต่างระหว่างค่าความคลาดเคลื่อนเฉลี่ยของแต่ละวิธี พบว่าที่ระดับนัยสำคัญ 0.05 ค่าความคลาดเคลื่อนเฉลี่ยของทั้ง 4 วิธี แตกต่างกันอย่างไม่มีนัยสำคัญ และยังพบว่าปัจจัยต่าง ๆ INTERACTION ต่าง ๆ ของปัจจัยต่าง ๆ ใน 3 ปัจจัยนี้ ทำให้ค่าความคลาดเคลื่อนเฉลี่ยแตกต่างกันอย่างไม่มีนัยสำคัญ

จากผลสรุปดังกล่าวอาจสรุปได้ว่า ถ้าหากมีปัญหาข้อมูลสูญหายและข้อมูลมีลักษณะตามสถานการณ์ต่าง ๆ ที่กำหนดไว้ จะสามารถประมาณค่าสูญหายโดยวิธีการใด ๆ ก็ได้ใน 4 วิธีการประมาณนี้ เนื่องจากมีผลทำให้ค่าความคลาดเคลื่อนเฉลี่ยไม่แตกต่างกันอย่างมีนัยสำคัญแต่อย่างไรก็ตาม ถ้าพิจารณาตามค่าความคลาดเคลื่อนเฉลี่ยที่ได้ในแต่ละสถานการณ์ แล้วจะเห็นว่าค่าไม่เท่ากัน ดังนั้น สิ่งที่เราควรคำนึงถึงเพิ่มเติมในการเลือกวิธีการประมาณค่าสูญหายในที่นี้ก็คือ การเรียงลำดับของค่าความคลาดเคลื่อนจากน้อยไปหามาก ถ้าค่าความคลาดเคลื่อน

เฉลี่ยน้อยก็ควรจะเป็นวิธีการประมาณที่ดีที่สุด และปัจจัยอีกอย่างหนึ่งที่ควรนำมาพิจารณาด้วยก็คือ เวลาในการประมาณค่าของแต่ละวิธี

2.1 ผลสรุปเมื่อพิจารณาการเรียงลำดับของค่าความคลาดเคลื่อนเฉลี่ยของแต่ละสถานการณ์พบว่า

ถ้า $p = 3$	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$\rho = 0.1$	วิธีที่ 1	ดีที่สุด	
	$n = 200$				$\rho = 0.1$	วิธีที่ 3	ดีที่สุด	
	$n = 30$	$n = 50$	$n = 70$	$n = 100$	$n = 200$	$\rho = 0.2-0.7$	วิธีที่ 3	ดีที่สุด
	$n = 30$	$n = 50$	$n = 70$			$\rho = 0.9$	วิธีที่ 4	ดีที่สุด
	$n = 100$	$n = 200$				$\rho = 0.9$	วิธีที่ 2	ดีที่สุด
ถ้า $p = 5$	$n = 30$	$n = 50$				$\rho = 0.1-0.2$	วิธีที่ 1	ดีที่สุด
	$n = 30$					$\rho = 0.4-0.9$	วิธีที่ 4	ดีที่สุด
	$n = 50$					$\rho = 0.3$	วิธีที่ 3	ดีที่สุด
	$n = 50$					$\rho = 0.35-0.9$	วิธีที่ 4	ดีที่สุด
	$n = 70$	$n = 100$	$n = 200$			$\rho = 0.2-0.4$	วิธีที่ 3	ดีที่สุด
	$n = 70$					$\rho = 0.5-0.8$	วิธีที่ 4	ดีที่สุด
	$n = 100$					$\rho = 0.7-0.8$	วิธีที่ 4	ดีที่สุด
	$n = 200$					$\rho = 0.8-0.9$	วิธีที่ 4	ดีที่สุด
ถ้า $p = 7$	$n = 30$	$n = 50$				$\rho = 0.1-0.2$	วิธีที่ 1	ดีที่สุด
	$n = 30$	$n = 50$				$\rho = 0.3-0.8$	วิธีที่ 4	ดีที่สุด
	$n = 70$	$n = 100$				$\rho = 0.3-0.8$	วิธีที่ 4	ดีที่สุด
	$n = 200$					$\rho = 0.2-0.4$	วิธีที่ 3	ดีที่สุด
	$n = 200$					$\rho = 0.5-0.8$	วิธีที่ 4	ดีที่สุด
ถ้า $p = 10$	$n = 70$	$n = 100$	$n = 200$			$\rho = 0.2-0.5$	วิธีที่ 4	ดีที่สุด

2.2 ผลสรุปเมื่อพิจารณาจากเวลาที่คอมพิวเตอร์ใช้ในการประมวลผลในกรณี

$P = 30$ $n = 30$ $\rho = 0.2$ โดยทำการทดลอง 1000 ครั้ง พบว่า เวลาที่ใช้ในการประมวลผลของแต่ละวิธีเป็นดังนี้

วิธีที่ใช้ในการประมาณ	เวลาที่คอมพิวเตอร์ใช้ในการประมวลผล
วิธีที่ใช้ค่าเฉลี่ย	1 นาที 31.03 วินาที
วิธีวิเคราะห์ความถดถอยพหุเชิงเส้น	3 นาที 10.93 วินาที
วิธีวิเคราะห์ความถดถอยพหุเชิงเส้น ดัดแปลง	3 นาที 09.02 วินาที
วิธีวิเคราะห์ส่วนประกอบหลัก	5 นาที 45.89 วินาที

จะเห็นว่าวิธีค่าเฉลี่ยใช้เวลาในการประมวลผลน้อยที่สุด และวิธีวิเคราะห์ส่วนประกอบหลัก ใช้เวลาในการประมวลผลมากที่สุด

3. ในการวิจัยครั้งนี้นอกจากจะพิจารณาค่าเฉลี่ยของค่าความคลาดเคลื่อนเฉลี่ยจาก 1000 ครั้งแล้ว ยังสนใจเปรียบเทียบผลของการถ่วงน้ำหนักของการเรียงลำดับค่าความคลาดเคลื่อนเฉลี่ยในแต่ละ SIMULATION พบว่าในการจำลองการทดลอง 106 สถานการณ์ ผลการวิจัยสอดคล้องกับเมื่อพิจารณาการเรียงลำดับของค่าความคลาดเคลื่อนเฉลี่ย 98 สถานการณ์ คิดเป็น 92.45 เปอร์เซ็นต์ และมีผลการวิจัยแตกต่างเมื่อพิจารณาการเรียงลำดับของค่าความคลาดเคลื่อนเฉลี่ย 8 ครั้ง คิดเป็น 7.55 เปอร์เซ็นต์ ดังนั้นอาจสรุปได้ว่าการพิจารณาตามค่าความคลาดเคลื่อนเฉลี่ย ซึ่งถือว่าเป็นข้อมูลปริมาณ และการพิจารณาค่าคะแนนรวมถ่วงน้ำหนัก ซึ่งถือว่าเป็นข้อมูลปริมาณที่แปลงมาจากข้อมูลคุณภาพให้ผลสอดคล้องกัน 92.45 เปอร์เซ็นต์

4. ข้อเสนอแนะ

4.1 สำหรับการวิจัยครั้งนี้เป็นการจำลองการทดลองขึ้นตามสถานการณ์ต่าง ๆ กัน ทั้งหมด 106 สถานการณ์ ในแต่ละสถานการณ์จะมีความแตกต่างกันในด้านขนาดตัวอย่าง จำนวนตัวแปร และขนาดความสัมพันธ์ระหว่างตัวแปร สถานการณ์หนึ่ง ๆ ใช้เวลาในการประมวลผลมาก ทั้งนี้อาจเป็นเพราะว่าเป็นการวิเคราะห์ของตัวแปรพหุ ดังนั้น ในการวิจัยต่อ ๆ ไป ที่เกี่ยวข้องควรจะลดเวลาในการประมวลผลลง โดยลดจำนวนรอบของการประมวลผลลงจาก 1000 รอบเหลือประมาณ 100-500 รอบ แต่ก่อนอื่นควรที่จะทำการเปรียบเทียบผลเมื่อประมวลผล 1000 รอบ และ 100-500 รอบ ว่าแตกต่างกันหรือไม่

4.2 ในการวิจัยครั้งนี้เสนอวิธีการประมาณค่าสู่ศูนย์หายทั้งหมด 4 วิธี จำนวนตัวแปร $P = 3\ 5\ 7\ 10$ ขนาดตัวอย่าง $n = 30\ 50\ 70\ 100\ 200$ และสัดส่วนของข้อมูลสูญหายของแต่ละตัวแปรเป็น 10% ดังนั้นการวิจัยครั้งนี้สามารถสรุปผลได้เฉพาะสถานการณ์เหล่านี้เท่านั้น สิ่งที่น่าสนใจเพิ่มเติมก็คือ ถ้าหากสัดส่วนของข้อมูลสูญหายแตกต่างกันไปคือ 15% 20% 30% และ 35% จำนวนตัวแปรเพิ่มขึ้นเป็น 30 และขนาดตัวอย่างเพิ่มขึ้นเป็น 500 ตัวอย่างแล้ว ผลการวิจัยจะสอดคล้องหรือแตกต่างกันอย่างไร



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย