

สรุปผลการวิจัยและข้อเสนอแนะ

จากผลการวิจัยโดยใช้ข้อมูลและวิธีการวิเคราะห์การถดถอยที่ได้กล่าวมาแล้วในบทที่ 2 พบว่าขนาดตัวแปรอิสระ (M) ขนาดตัวอย่าง (N) การกระจายของข้อมูล (C.V.) ค่าเบี่ยงเบนมาตรฐาน (d) และเปอร์เซ็นต์ข้อมูลสูญหาย (MM) มีผลต่อวิธีที่ใช้ในการประมาณค่าสูญหายของตัวแปรอิสระในการวิเคราะห์การถดถอย ซึ่งวิธีที่ใช้ในการประมาณค่าสูญหายของตัวแปรอิสระในการวิเคราะห์การถดถอยที่ใช้ในการศึกษาคั้งนี้คือ

- วิธีที่ 1. วิธีวิเคราะห์หสัมบูรณ์
- วิธีที่ 2. วิธีสูญหาย
- วิธีที่ 3. วิธีวิเคราะห์ความถดถอย
- วิธีที่ 4. วิธี MAXIMUM LIKELIHOOD
- วิธีที่ 5. วิธีค่าเฉลี่ย
- วิธีที่ 6. วิธีค่ามัธยฐาน

ในการตรวจสอบความเหมาะสมของวิธีที่ใช้ในการประมาณค่าสูญหายของตัวแปรอิสระดังกล่าวนั้นนอกจากจะใช้ MSE ของค่าประมาณเมื่อไม่มีข้อมูลสูญหายเป็นเกณฑ์แล้วยังได้นำเอาเวลาที่ใช้ในการประมวลผลแต่ละวิธีด้วยเครื่องคอมพิวเตอร์มาพิจารณาประกอบด้วย

ผลการตรวจสอบความเหมาะสมของวิธีที่ใช้ในการประมาณค่าสูญหายของตัวแปรอิสระสรุปได้ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย

1. ผลสรุปเมื่อพิจารณาการเรียงลำดับค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละสถานการณ์พบว่า

1.1 ถ้า M = 2	C.V. = 0.05	MM = 5%	N = 30, 100	วิธีที่ 5 ดีที่สุด
			N = 70	วิธีที่ 4 ดีที่สุด
		MM = 10%	N = 30, 70	วิธีที่ 5 ดีที่สุด
			N = 100	วิธีที่ 4 ดีที่สุด

	MM = 15%	N = 30, 70, 100	วิธีที่ 5 คัดสุด	
	C.V. = 0.2	MM = 5%	N = 30, 100	วิธีที่ 5 คัดสุด
		N = 70	วิธีที่ 4 คัดสุด	
	MM = 10%, 15%			
		N = 30, 70, 100	วิธีที่ 5 คัดสุด	
	C.V. = 1	MM = 5%	N = 30, 100	วิธีที่ 5 คัดสุด
		N = 70	วิธีที่ 4 คัดสุด	
	MM = 10%, 15%			
		N = 30, 70	วิธีที่ 5 คัดสุด	
		N = 100	วิธีที่ 3 คัดสุด	
1.2 ถ้า $M = 3$	C.V. = 0.05	MM = 5%	N = 30	วิธีที่ 4 คัดสุด
			N = 70, 100	วิธีที่ 5 คัดสุด
		MM = 10%	N = 30, 70	วิธีที่ 5 คัดสุด
			N = 100	วิธีที่ 4 คัดสุด
		MM = 15%	N = 30, 70, 100	วิธีที่ 5 คัดสุด
	C.V. = 0.2	MM = 5%	N = 30	วิธีที่ 4 คัดสุด
			N = 70, 100	วิธีที่ 5 คัดสุด
		MM = 10%	N = 30, 70, 100	วิธีที่ 4 คัดสุด
		MM = 15%	N = 30, 100	วิธีที่ 5 คัดสุด
			N = 70	วิธีที่ 4 คัดสุด
	C.V. = 1	MM = 5%	N = 30	วิธีที่ 4 คัดสุด
			N = 70, 100	วิธีที่ 5 คัดสุด
		MM = 10%	N = 30, 70, 100	วิธีที่ 4 คัดสุด
		MM = 15%	N = 30, 100	วิธีที่ 5 คัดสุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

	N = 70	วิธีที่ 4 คัดสุด
1.3 ถ้า $M = 5$ C.V. = 0.05	MM = 5% N = 30, 100	วิธีที่ 4 คัดสุด
	N = 70	วิธีที่ 5 คัดสุด
	MM = 10% N = 30, 70	วิธีที่ 5 คัดสุด
	N = 100	วิธีที่ 4 คัดสุด
	MM = 15% N = 30, 100	วิธีที่ 4 คัดสุด
	N = 70	วิธีที่ 3 คัดสุด
	C.V. = 0.2 MM = 5% N = 30, 100	วิธีที่ 4 คัดสุด
	N = 70	วิธีที่ 5 คัดสุด
	MM = 10% N = 30	วิธีที่ 5 คัดสุด
	N = 70	วิธีที่ 3 คัดสุด
	N = 100	วิธีที่ 4 คัดสุด
	MM = 15% N = 30	วิธีที่ 4 คัดสุด
	N = 70	วิธีที่ 3 คัดสุด
	N = 100	วิธีที่ 5 คัดสุด
	C.V. = 1 MM = 5% N = 30	วิธีที่ 5 คัดสุด
	N = 70, 100	วิธีที่ 4 คัดสุด
	MM = 10% N = 30	วิธีที่ 5 คัดสุด
	N = 70	วิธีที่ 3 คัดสุด
	N = 100	วิธีที่ 4 คัดสุด
	MM = 15% N = 30	วิธีที่ 4 คัดสุด
	N = 70	วิธีที่ 3 คัดสุด
	N = 100	วิธีที่ 5 คัดสุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.4 ถ้า $M = 7$ $C.V. = 0.05$ $MM = 5\%$ $N = 30, 70, 100$ วิธีที่ 5 คัดสุด

$MM = 10\%$ $N = 30$ วิธีที่ 3 คัดสุด

$N = 70, 100$ วิธีที่ 5 คัดสุด

$MM = 15\%$ $N = 30, 100$ วิธีที่ 5 คัดสุด

$N = 70$ วิธีที่ 4 คัดสุด

$C.V. = 0.2$ $MM = 5\%, 15\%$

$N = 30, 70, 100$ วิธีที่ 5 คัดสุด

$MM = 10\%$ $N = 30$ วิธีที่ 3 คัดสุด

$N = 70, 100$ วิธีที่ 5 คัดสุด

$C.V. = 1$ $MM = 5\%$ $N = 30, 70, 100$ วิธีที่ 5 คัดสุด

$MM = 10\%, 15\%$

$N = 30$ วิธีที่ 3 คัดสุด

$N = 70, 100$ วิธีที่ 5 คัดสุด

2. ผลสรุปเมื่อพิจารณาจากเวลาที่คอมพิวเตอร์ใช้ในการประมวลผลในกรณีที่มี $M = 7$, $N = 100$, $MM = 15\%$, $C.V. = 1$, $G = 25$ โดยกระทำซ้ำ 700 ครั้ง พบว่าเวลาที่ใช้ในการประมวลผลของแต่ละวิธีเป็นดังนี้

วิธีที่ใช้ในการประมวล	เวลาที่คอมพิวเตอร์ใช้ในการประมวลผล
1. วิธีวิเคราะห์สมบูรณ์	1.19.01
2. วิธีสุ่มหาย	1.35.11
3. วิธีวิเคราะห์ความถดถอย	2.30.07
4. วิธี MAXIMUM LIKELIHOOD	1.46.18
5. วิธีค่าเฉลี่ย	1.30.49
6. วิธีค่ามัธยฐาน	1.43.00

จะเห็นว่าวิธีค่าเฉลี่ยใช้เวลาการประมวลผลน้อยที่สุด ในขณะที่วิธีวิเคราะห์ความถดถอย ใช้เวลาการประมวลผลมากที่สุด

เมื่อเปรียบเทียบ MSE ของสมการถดถอยเชิงซ้อน โดยประมวลค่าสุทธัยทั้ง 4 วิธี กับ MSE ของสมการถดถอยเมื่อไม่มีข้อมูลสุทธัยมีผลต่างกันตามสถานการณ์ที่จำลองขึ้น สิ่งที่น่านำมาพิจารณาเพิ่มเติมในการวิเคราะห์ความถดถอยคือ อัตราส่วนความคลาดเคลื่อนรวมของค่าประมาณที่ได้จากสมการถดถอยเมื่อแทนค่าสุทธัยของตัวแปรอิสระด้วยวิธีการต่าง ๆ เมื่อเทียบกับวิธีที่ 1 ซึ่งพบว่าบางสถานการณ์ให้ผลสอดคล้องกับเมื่อพิจารณาจากค่า MSE และบางสถานการณ์ให้ผลไม่ สอดคล้องกับเมื่อพิจารณาจากค่า MSE ซึ่งสถานการณ์ที่ให้ผลไม่สอดคล้องคือ

			พิจารณาจาก อัตราส่วนความ คลาดเคลื่อนรวม	พิจารณาจาก MSE
เมื่อ $M = 2$	$C.V. = 0.05$	$MM = 10\%$ $N = 100$	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด
		$MM = 15\%$ $N = 70$	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
		$C.V. = 1$ $MM = 10\%, 15\%$ $N = 100$	วิธีที่ 5 ดีที่สุด	วิธีที่ 3 ดีที่สุด
เมื่อ $M = 3$	$C.V. = 0.05$	$MM = 5\%, 15\%$ $N = 70$	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
		$C.V. = 0.2$ $MM = 5\%$ $N = 70$	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
	$C.V. = 1$	$MM = 10\%$ $N = 30, 70$	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด
		$MM = 10\%$ $N = 30, 70$	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด
เมื่อ $M = 5$	$C.V. = 0.05$	$MM = 15\%$ $N = 70$	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด
		$MM = 5\%$ $N = 100$	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด
		$MM = 10\%$ $N = 30$	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
	$C.V. = 0.2$	$MM = 15\%$ $N = 30$	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด
		$N = 70$	วิธีที่ 4 ดีที่สุด	วิธีที่ 3 ดีที่สุด
		$MM = 5\%$ $N = 70$	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
		$N = 100$	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด

พิจารณาจาก
อัตราส่วนความ
คลาดเคลื่อนรวม

พิจารณาจาก
MSE

MM = 10%	N = 30	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
	N = 70	วิธีที่ 5 ดีที่สุด	วิธีที่ 3 ดีที่สุด
MM = 15%	N = 30	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด
	N = 70	วิธีที่ 5 ดีที่สุด	วิธีที่ 3 ดีที่สุด
C.V. = 1	N = 100	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
	MM = 5%, N = 30	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
	MM = 10%, 15%, N = 70	วิธีที่ 5 ดีที่สุด	วิธีที่ 3 ดีที่สุด
MM = 15%	N = 30	วิธีที่ 5 ดีที่สุด	วิธีที่ 4 ดีที่สุด
	N = 100	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
เมื่อ M = 7 C.V. = 0.05			
	N = 30	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
MM = 10%	N = 30	วิธีที่ 4 ดีที่สุด	วิธีที่ 3 ดีที่สุด
MM = 10%	N = 70	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
MM = 15%	N = 100	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
C.V. = 0.2			
	N = 30	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
MM = 10%	N = 30	วิธีที่ 4 ดีที่สุด	วิธีที่ 3 ดีที่สุด
MM = 10%	N = 70	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
C.V. = 1	MM = 5%, N = 30	วิธีที่ 4 ดีที่สุด	วิธีที่ 5 ดีที่สุด
	MM = 10%, 15%, N = 30	วิธีที่ 4 ดีที่สุด	วิธีที่ 3 ดีที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

นั่นคือ 272 สถานการณ์ (62.96%) จากการจำลองการทดลอง 432 สถานการณ์
ที่ทำให้ค่าอัตราส่วนดังกล่าวสอดคล้องกับเมื่อพิจารณาจากค่าคลาดเคลื่อนกำลังสองเฉลี่ย

เมื่อคำนึงถึงปัจจัยอย่างใดอย่างหนึ่งเพียงอย่างเดียวโดยไม่คำนึงถึงปัจจัยอย่างอื่น
ซึ่งคิดเป็นเปอร์เซ็นต์สามารถสรุปได้ดังนี้

1. พิจารณาจากจำนวนตัวแปรอิสระ

		% ของจำนวนสถานการณ์ที่ให้ผลใกล้เคียงกับ เมื่อมีข้อมูลครบถ้วน	
		เมื่อพิจารณาจาก ค่า MSE	เมื่อพิจารณาจาก อัตราส่วนความคลาดเคลื่อนรวม
M = 2	วิธีที่ 5 คัดสุ่ม	77.78 %	85.19 %
	วิธีที่ 4 คัดสุ่ม	14.81 %	14.81 %
	วิธีที่ 3 คัดสุ่ม	7.41 %	
M = 3	วิธีที่ 5 คัดสุ่ม	55.56 %	62.96 %
	วิธีที่ 4 คัดสุ่ม	44.44 %	37.04 %
M = 5	วิธีที่ 5 คัดสุ่ม	33.33 %	44.44 %
	วิธีที่ 4 คัดสุ่ม	48.15 %	55.56 %
	วิธีที่ 3 คัดสุ่ม	18.52 %	
M = 7	วิธีที่ 5 คัดสุ่ม	81.49 %	51.85 %
	วิธีที่ 4 คัดสุ่ม	3.70 %	48.15 %
	วิธีที่ 3 คัดสุ่ม	14.81 %	

2. พิจารณาจากเปอร์เซ็นต์ข้อมูลสูญหาย

		% ของจำนวนสถานการณ์ที่ให้ผลใกล้เคียงกับ เมื่อมีข้อมูลครบถ้วน	
		เมื่อพิจารณาจาก ค่า MSE	เมื่อพิจารณาจาก อัตราส่วนความคลาดเคลื่อนรวม
MM = 5%	วิธีที่ 5 คัดสุ่ม	66.67 %	52.78 %
	วิธีที่ 4 คัดสุ่ม	33.33 %	47.22 %

MM = 10%	วิธีที่ 5 คัดสุด	52.78 %	61.11 %
	วิธีที่ 4 คัดสุด	30.55 %	38.89 %
	วิธีที่ 3 คัดสุด	16.66 %	
MM = 15%	วิธีที่ 5 คัดสุด	69.45 %	66.67 %
	วิธีที่ 4 คัดสุด	19.44 %	33.33 %
	วิธีที่ 3 คัดสุด	11.11 %	

3. พิจารณาจากขนาดตัวอย่าง

๔ ของจำนวนสถานการณ์ที่ให้ผลใกล้เคียงกับ
เมื่อมีข้อมูลครบถ้วน

		เมื่อพิจารณาจาก ค่า MSE	เมื่อพิจารณาจาก อัตราส่วนความคลาดเคลื่อนรวม
N = 30	วิธีที่ 5 คัดสุด	61.11 %	55.56 %
	วิธีที่ 4 คัดสุด	27.78 %	44.44 %
	วิธีที่ 3 คัดสุด	11.11 %	
N = 70	วิธีที่ 5 คัดสุด	61.11 %	61.11 %
	วิธีที่ 4 คัดสุด	25.00 %	38.89 %
	วิธีที่ 3 คัดสุด	13.89 %	
N = 100	วิธีที่ 5 คัดสุด	63.89 %	72.22 %
	วิธีที่ 4 คัดสุด	30.56 %	27.78 %
	วิธีที่ 3 คัดสุด	5.55 %	

4. พิจารณาจากลักษณะการกระจายข้อมูล (C.V.)

๔ ของจำนวนสถานการณ์ที่ให้ผลใกล้เคียงกับ
เมื่อมีข้อมูลครบถ้วน

		เมื่อพิจารณาจาก ค่า MSE	เมื่อพิจารณาจาก อัตราส่วนความคลาดเคลื่อนรวม
C.V. = 0.05	วิธีที่ 5 คัดสุด	66.67 %	52.78 %
	วิธีที่ 4 คัดสุด	27.78 %	47.22 %

	วิธีที่ 3 คัดที่สุด	5.55 %	
C.V. = 0.2	วิธีที่ 5 คัดที่สุด	63.89 %	62.11 %
	วิธีที่ 4 คัดที่สุด	27.78 %	38.89 %
	วิธีที่ 3 คัดที่สุด	8.33 %	
C.V. = 1.0	วิธีที่ 5 คัดที่สุด	55.56 %	69.44 %
	วิธีที่ 4 คัดที่สุด	27.78 %	30.56 %
	วิธีที่ 3 คัดที่สุด	16.66 %	

จากผลสรุปดังกล่าวข้างต้นเมื่อพิจารณาจากค่า MSE และอัตราส่วนความคลาดเคลื่อนรวมแล้วในสถานการณ์ส่วนใหญ่วิธีที่ 5 (วิธีค่าเฉลี่ย) จะเป็นวิธีที่เหมาะสมที่สุดสำหรับการประมาณค่าสูญหายของตัวแปรอิสระในการวิเคราะห์การถดถอย ถึงแม้ว่าในบางสถานการณ์วิธีที่ 4 (วิธี MLE) จะให้ค่า MSE ใกล้เคียงกับ MSE ของวิธีที่ 1 (วิธีวิเคราะห์สมบูรณ์) ก็ตามแต่เมื่อพิจารณาถึงความแตกต่างระหว่าง MSE ของวิธีที่ 4 และวิธีที่ 5 พบว่ามีความแตกต่างกันไม่มากนักกล่าวคือแตกต่างกันในช่วง (0.00007, 2.22427) เมื่อ $d = 5$, (0.00042, 10.75102) เมื่อ $d = 10$ (0.00016, 33.08619) เมื่อ $d = 20$, (0.00781, 49.71848) เมื่อ $d = 25$ นอกจากนี้ในการประมวลผลของวิธีที่ 4 ใช้เวลาคอมพิวเตอร์มากกว่าวิธีที่ 5 ด้วย

ในสถานการณ์ที่ตัวแปรอิสระและเปอร์เซ็นต์ของข้อมูลสูญหายมีจำนวนมากในขณะที่ขนาดตัวอย่างเล็ก วิธีที่ 3 (วิธีวิเคราะห์ความถดถอย) จะเหมาะสมที่สุด เช่นเดียวกับเมื่อตัวแปรอิสระมีจำนวนน้อยและขนาดตัวอย่างใหญ่ MSE ของวิธีที่ 3 จะมีค่าใกล้เคียงกับ MSE ของวิธีที่ 1 ที่สุด อย่างไรก็ตามเมื่อพิจารณา MSE ของวิธีที่ 3 เปรียบเทียบกับวิธีที่ 2 (วิธีสูญหาย) ในสถานการณ์เช่นนี้พบว่าค่า MSE ใกล้เคียงกัน นั่นคือถ้าขนาดตัวอย่างใหญ่และจำนวนตัวแปรอิสระน้อย การตัดชุดข้อมูลสูญหายทั้งจะไม่มีความกระทบกระเทือนกับผลการวิเคราะห์การถดถอยโดยวิธีกำลังสองน้อยที่สุด

ดังนั้นถ้ามีข้อมูลอยู่ชุดหนึ่งและต้องการดูว่าวิธีใดที่จะประมาณค่าสูญหายในการวิเคราะห์สมการถดถอยดีที่สุดควรพิจารณาดังนี้

1. ขนาดตัวอย่าง

2. คุณลักษณะการกระจายของข้อมูลอย่างคร่าว ๆ โดยพิจารณาจากค่า C.V.
3. คำนวณค่าสหุหยาของตัวแปรว่ามีกี่เปอร์เซ็นต์

เมื่อตรวจสอบคุณลักษณะของข้อมูลอย่างคร่าว ๆ ข้างต้นแล้ว จึงคิดว่าวิธีใดที่ได้กล่าวมาแล้วเหมาะสมกับข้อมูลที่มีอยู่ ทั้งนี้อาจจะคำนึงถึงเวลาในการประมวลผลด้วย เพื่อที่จะลดค่าใช้จ่ายในการวิจัย

ข้อเสนอแนะ

สำหรับการวิจัยครั้งนี้เป็นการจำลองการทดลองขึ้นตามสถานการณ์ต่าง ๆ กัน 432 สถานการณ์ ในแต่ละสถานการณ์จะมีความแตกต่างกันในด้านการกระจายของข้อมูล ขนาดตัวอย่าง ขนาดตัวแปรอิสระ การสหุหยาของข้อมูล ความเบี่ยงเบนมาตรฐาน โดยที่สถานการณ์ดังกล่าวมีการแจกแจงของข้อมูลเป็นแบบปกติและตัวแปรอิสระไม่มีความสัมพันธ์กัน, สัดส่วนการสหุหยาของข้อมูลเท่ากันในทุกตัวแปร ดังนั้นในการวิจัยต่อ ๆ ไปควรศึกษาในกรณีที่ข้อมูลไม่ได้มีการแจกแจงแบบปกติและควรคำนึงถึงความสัมพันธ์ระหว่างตัวแปรอิสระและสัดส่วนการสหุหยาของข้อมูลไม่เท่ากันทุกตัวแปร และควรจะนำค่าสัมประสิทธิ์ความถดถอยที่ได้จากการสมการถดถอยที่ประมาณด้วยวิธีต่าง ๆ เปรียบเทียบกับวิธีที่ไม่มีข้อมูลสหุหยา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย