

ภาวะ Multicollinearity กับ การวิเคราะห์การถดถอยพหุคูณ

นัจริกรณ สิมมกรณ *

คำว่า “Multicollinearity” นั้น ในตำราที่เป็นภาษาไทยบางตำราที่ใช้คำว่า “ภาวะร่วมเส้นตรงหลายตัวแปร” “ภาวะร่วมเส้นตรงเชิงพหุ” คำเหล่านี้มีความหมายในทำนองเดียวกันคือ เป็นกรณีที่ตัวแปร ซึ่งอาจจะได้ทั้งในกรณีของตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์กันหรือไม่เป็นอิสระจากกัน ถ้าสัมพันธ์เพียง 2 ตัว จะเรียก Collinearity และถ้าสัมพันธ์กันหลายตัวจะเรียก Multicollinearity (อารี วิบูลย์พงศ์. มปป. : 159) ซึ่งปัญหาดังกล่าวมักเกิดขึ้นเสมอเมื่อมีการ นำเทคนิคการวิเคราะห์หลายตัวแปร (Multivariate Analysis) มาใช้ ในกรณีการวิเคราะห์ที่มีตัวแปรอิสระหลายตัว เหตุผลของนำตัวแปรหลายตัว มาเป็นตัวแปรอิสระในอธิบายการเปลี่ยนแปลงของตัวแปรตามตัวหนึ่งหรืออาจมากกว่า นั้นคือ นักวิจัยมีแนวความคิดว่า ตัวแปรอิสระเพียงตัวเดียวไม่สามารถที่จะอธิบายตัวแปรตามได้อย่างเพียงพอ และการที่จะศึกษาว่าตัวแปรอิสระแต่ละตัวมีอิทธิพลต่อตัวแปรตามเท่าใด ตัวแปรอิสระแต่ละตัวจะต้องไม่ขึ้นอยู่กับตัวแปรอิสระตัวใดตัวหนึ่งมากเกินไปหรือกล่าวอีกนัยหนึ่งได้ว่าตัวแปรอิสระไม่ควรมีความสัมพันธ์กันมากเกินไป มิฉะนั้นจะแยกผลของตัวแปรอิสระตัวนั้นว่ามีผลต่อตัวแปรตามมากน้อยเพียงใดไม่ได้ แต่ในความเป็นจริงโอกาสที่คุณสมบัติของหน่วยวิเคราะห์จะมีความสัมพันธ์กันมากก็มักเกิดขึ้นอยู่เสมอ (สุชาติ ประสิทธิ์รัฐสินธุ์. 2545 : 83) ดังจะเห็นได้จากในงานวิจัยหรือการศึกษาค้นคว้าที่มีการศึกษาความสัมพันธ์ระหว่างหลายตัวแปร อาทิเช่น การวิเคราะห์การถดถอยพหุคูณ การวิเคราะห์ความสัมพันธ์สาเหตุ การวิเคราะห์องค์ประกอบเชิงสำรวจหรือเชิงยืนยัน แบบจำลองสมการโครงสร้าง เป็นต้น มักจะพบปัญหาดังกล่าวเกิดขึ้นอยู่บ่อยครั้ง ซึ่งมักไม่เป็นผลดีต่อคำตอบของการวิจัยหรือการศึกษา เพราะถ้าตัวแปรอิสระมีความสัมพันธ์กันมากเกินไป ผลที่จะเกิดตามมาคือ ความคลาดเคลื่อนหรือการขาดความแม่นยำในการคาดเดาถึงคำตอบที่จะเกิดขึ้นให้ได้ใกล้เคียงกับความเป็นจริง ปัญหาดังกล่าวนักวิจัยไม่ควรจะเพิกเฉย แต่เป็นหน้าที่ของนักวิจัยที่จะต้องหาวิธีการหรือแนวทางที่จะแยกความสัมพันธ์ระหว่างตัวแปรอิสระเหล่านั้นออกจากกัน เพื่อให้ได้คำตอบของการวิจัยที่ดีที่สุดและเกิดความคลาดเคลื่อนน้อยที่สุด จากข้างต้นผู้อ่านคงจะพอมองภาพออกแล้วว่าภาวะ Multicollinearity จะส่งผล ในด้านลบต่อการศึกษาในลักษณะใดได้บ้าง ในที่นี้ผู้เขียนจะขอกล่าวถึงผลของปัญหาดังกล่าวต่อการวิเคราะห์การถดถอยพหุคูณเท่านั้น (Multiple Regression Analysis)

ผลของ Multicollinearity ที่มีต่อการวิเคราะห์การถดถอยพหุคูณ

ในการวิเคราะห์การถดถอยพหุคูณนั้นเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตาม(Dependent or Outcome Variable) ซึ่งบางครั้งเรียกว่า “ตัวแปรเกณฑ์ (Criterion variable) 1 ตัว กับตัวแปรอิสระ (Independent Variable) บางครั้งเรียกว่า “ตัวแปรพยากรณ์” หรือ “ตัวแปรทำนาย” (predictor) ตั้งแต่ 2 ตัวขึ้นไป จากนั้นไปเพื่อไม่ให้เกิดความสับสนผู้เขียนขอใช้คำว่า “ตัวแปรตาม” และ “ตัวแปรอิสระ” ซึ่งถือเป็นพื้นฐานที่ถูกนำไปประยุกต์ใช้ในวิธีการทางสถิติขั้นที่สูงกว่าเพื่อนำมาซึ่งคำตอบของงานวิจัยที่เป็นประโยชน์ทั้งทางด้านการศึกษา เศรษฐกิจ และสังคม มากมาย เมื่อมีการใช้ Multiple Regression Analysis ประกอบการศึกษาใด แน่แน่นอนว่าสิ่งที่ละเลยไม่ได้คือการพิจารณาว่า ตัวแปรอิสระเหล่านั้นมีความสัมพันธ์กันเกิดขึ้นในระดับใด เช่น ถ้าต้องการสร้างสมการความสัมพันธ์ระหว่างยอดขายเครื่องใช้เด็กอ่อน (Y) กับตัวแปรอิสระ 3 ตัว คือ X_1 = จำนวนเครื่องใช้เด็กอ่อนที่ขายได้ X_2 = จำนวนลูกค้าที่เข้ามาในร้าน X_3 = ราคาต่อหน่วยของเครื่องใช้เด็กอ่อน

*นิสิตปริญญาเอก สาขาวิจัยและประเมินผลการศึกษา มหาวิทยาลัยมหาสารคาม รุ่นที่ 4

ในกรณีนี้ จำนวนเครื่องใช้เด็กอ่อนที่ขายได้ (X_1) อาจมีความสัมพันธ์กับลูกค้าที่เข้ามาในร้าน (X_2) หรือราคาต่อหน่วยของเครื่องใช้เด็กอ่อน (X_3) อาจมีความสัมพันธ์กับจำนวนที่ขายได้ (X_1) เพราะถ้าความสัมพันธ์ระหว่างตัวแปรอิสระ ที่กล่าวมาข้างต้นนี้ มีความสัมพันธ์กันสูงก็จะส่งผลให้การประมาณค่าสถิติต่างๆ คลาดเคลื่อนไป และทำให้ได้สมการการถดถอยที่ไม่ถูกต้อง ซึ่งจะส่งผลต่อการประมาณค่าเฉลี่ยของ Y หรือการพยากรณ์ Y คลาดเคลื่อนจากความเป็นจริง เพื่อสะท้อนให้เห็นถึงความชัดเจนยิ่งขึ้น ผู้เขียนจึงขอกล่าวถึงผลของภาวะ Multicollinearity ที่มีต่อการวิเคราะห์การถดถอยพหุคูณ ในประเด็นต่อไปนี้

ประการที่ 1 ส่งผลให้ความสามารถในสังเกตอิทธิพลหรือการทำนายของตัวแปรอิสระแต่ละตัวลดลง ดังตัวอย่างการวิเคราะห์ปัจจัยที่ส่งผลต่อแรงจูงใจในการทำงาน (MOT) โดยมีตัวแปรพยากรณ์ 6 ตัว ได้แก่ TPW, HUM, TAT, PER, AGE และ SAL

Correlations								
		MOT	TPW	HUM	TAT	PER	AGE	SAL
Pearson Correlation	MOT	1.000	-.023	.743	.592	.403	.112	.115
	TPW	-.023	1.000	-.004	-.006	-.011	-.035	-.040
	HUM	.743	-.004	1.000	.519	.319	.096	.103
	TAT	.592	-.006	.519	1.000	.199	.084	.094
	PER	.403	-.011	.319	.199	1.000	.025	.055
	AGE	.112	-.035	.096	.084	.025	1.000	.839
	SAL	.115	-.040	.103	.094	.055	.839	1.000
Sig. (1-tailed)	MOT		.192	.000	.000	.000	.000	.000
	TPW	.192		.433	.409	.336	.093	.065
	HUM	.000	.433		.000	.000	.000	.000
	TAT	.000	.409	.000		.000	.001	.000
	PER	.000	.336	.000	.000		.175	.019
	AGE	.000	.093	.000	.001	.175		.000
	SAL	.000	.065	.000	.000	.019	.000	
N	MOT	1442	1442	1442	1442	1442	1442	1442
	TPW	1442	1442	1442	1442	1442	1442	1442
	HUM	1442	1442	1442	1442	1442	1442	1442
	TAT	1442	1442	1442	1442	1442	1442	1442
	PER	1442	1442	1442	1442	1442	1442	1442
	AGE	1442	1442	1442	1442	1442	1442	1442
	SAL	1442	1442	1442	1442	1442	1442	1442

ภาพที่ 1 แสดงเมตริกซ์ความสัมพันธ์ระหว่างตัวแปร

จากภาพที่ 1 พบว่า ตัวแปรพยากรณ์ AGE และ SAL มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .839 ซึ่งถือว่าเข้าข่ายภาวะ Multicollinearity เมื่อนำตัวแปรพยากรณ์ทั้งหมดเข้าสู่การวิเคราะห์การถดถอยโดยวิธี Enter ผลปรากฏดังนี้ กรณีที่นำ AGE และ SAL เข้าไปวิเคราะห์ร่วมกัน ได้ผลดังภาพที่ 2

Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	39.485	4.929		8.010	.000					
	TPW	-.778	.777	-.016	-1.002	.317	-.023	-.026	-.016	.998	1.002
	HUM	2.044	.073	.543	28.191	.000	.743	.597	.447	.679	1.472
	TAT	.312	.021	.273	14.866	.000	.592	.361	.233	.728	1.374
	PER	.764	.073	.175	10.447	.000	.403	.266	.166	.894	1.119
	AGE	.210	.145	.042	1.445	.149	.112	.038	.023	.296	3.379
	SAL	.000	.001	-.012	-.411	.681	.115	-.011	-.007	.295	3.388

a. Dependent Variable: MOT

a. Dependent Variable: MOT

ภาพที่ 2 แสดงผลการวิเคราะห์การถดถอยในกรณีที่นำ AGE และ SAL เข้าไปวิเคราะห์ร่วมกัน

กรณีที่นำเฉพาะ SAL เข้าไปวิเคราะห์ได้ผลดังภาพที่ 3

Coefficients ^a											
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics		
	B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF	
1	(Constant)	42.370	4.509		9.397	.000					
	TPW	-.781	.777	-.016	-1.006	.315	-.023	-.027	-.016	.998	1.002
	HUM	2.047	.073	.543	28.230	.000	.743	.597	.448	.680	1.471
	TAT	.312	.021	.273	14.664	.000	.592	.361	.233	.728	1.374
	PER	.759	.073	.174	10.386	.000	.403	.264	.165	.898	1.116
	SAL	.001	.000	.023	1.459	.145	.115	.038	.023	.985	1.015

a. Dependent Variable: MOT

ภาพที่ 3 แสดงผลการวิเคราะห์การถดถอย

ในกรณีที่นำเฉพาะ SAL เข้าไปวิเคราะห์จะพบว่าถ้าไม่มีตัวแปร AGE ค่าสัมประสิทธิ์การพยากรณ์ในรูปคะแนนมาตรฐานของ SAL จะเพิ่มขึ้นจาก -.012 เป็น 0.23

กรณีที่นำเฉพาะ AGE เข้าไปวิเคราะห์ ได้ผลดังภาพที่ 4

Coefficients ^a											
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	39.868	4.839		8.239	.000					
	TPW	-.772	.776	-.016	-.994	.320	-.023	-.026	-.016	.999	1.001
	HUM	2.044	.072	.542	28.197	.000	.743	.597	.447	.679	1.472
	TAT	.311	.021	.273	14.665	.000	.592	.361	.233	.728	1.374
	PER	.762	.073	.175	10.443	.000	.403	.266	.166	.896	1.116
	AGE	.180	.080	.032	2.013	.044	.112	.053	.032	.988	1.012

a. Dependent Variable: MOT

ภาพที่ 4 แสดงผลการวิเคราะห์การถดถอยในกรณีที่นำเฉพาะ AGE เข้าไปวิเคราะห์

จะพบว่าถ้าไม่มีตัวแปร SAL ค่าสัมประสิทธิ์การพยากรณ์ในรูปคะแนนมาตรฐานของ AGE แม้ว่าจะลดลงจาก .043 เป็น 0.32 แต่ตัวแปรดังกล่าวสามารถทำนายได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ประการที่ 2 สัมประสิทธิ์การถดถอยไม่คงที่ทั้งขนาดและเครื่องหมาย (อาจเป็นได้ทั้งบวกและลบ) ทั้งนี้เพราะความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์การถดถอยมีมาก ซึ่งจะนำไปสู่การได้สมการถดถอยที่ไม่แม่นยำ และยากต่อการมีนัยสำคัญทางสถิติ ดังกรณีของ SAL และ AGE ในตัวอย่างการวิเคราะห์ที่ผ่านมา

ประการที่ 3 ผลของการสร้างสมการถดถอยอาจมีนัยสำคัญของ R^2 รวม แต่ตัวแปรอิสระอาจไม่มีนัยสำคัญในการทำนายหรือเป็นการบอกได้ว่าตัวแปรอิสระดังกล่าวไม่มีอิทธิพลต่อตัวแปรตาม ทั้งนี้ไม่ใช่เพราะตัวแปรอิสระนั้นไม่มีอิทธิพลต่อตัวแปรตามแต่เป็นเพราะตัวอย่างหรือสาระของข้อมูลที่ได้มาไม่สามารถจำแนกอิทธิพลของตัวแปรอิสระนั้นได้อย่างแม่นยำ ซึ่งในกรณีนี้อาจทำให้เกิดข้อสงสัยผิดพลาดในส่วนของการตัดตัวแปรอิสระนั้นออกไป โดยอาจทำให้ผิดไปจากทฤษฎี ซึ่งก่อให้เกิดความคลาดเคลื่อนตามมา ผลในข้อนี้บางครั้งมาจากความผิดพลาดของผู้วิจัยในส่วนของการสร้างเครื่องมือในการเก็บรวบรวมข้อมูล การสุ่มกลุ่มตัวอย่างก็เป็นได้ประการที่ 4 คือ ตัวประมาณค่ามีความไวกับการเพิ่มขึ้นหรือลดลงของขนาดกลุ่มตัวอย่างหรือข้อมูล กล่าวคือ การตัดหรือเพิ่มจำนวนของกลุ่มตัวอย่างเพียงเล็กน้อยก็สามารถทำให้ค่าประมาณเปลี่ยนแปลงได้

การพิจารณาภาวะ Multicollinearity

ในการตรวจสอบปัญหาดังกล่าวสามารถทำได้หลายวิธี แต่ละวิธีนั้นเป็นวิธีที่ไม่มีสถิติรองรับในการระบุระดับของปัญหา แต่เป็นเพียงวิธีที่มีการกำหนดระดับในทางใดทางหนึ่งขึ้นมาเป็นค่าเกณฑ์ไว้เพื่อเปรียบเทียบ ดังตัวอย่างที่จะกล่าวต่อไปนี้

➤ การตรวจสอบเมตริกซ์ความสัมพันธ์ระหว่างตัวแปรอิสระ ถ้าตัวแปรอิสระตัวใด มีความสัมพันธ์กันสูง (ค่า r_{xy} มีค่าตั้งแต่ 0.80 หรือ 0.90 ขึ้นไป (อารี วิบูลย์พงศ์ : 164, ทรงศักดิ์ ภูสีอ่อน. 2551 : 280, Joseph F. Hair et al. 2010 : 200, Barbara G.Tabachnick and Linda S. Fidell.2007 : 89)) ก็แสดงว่าตัวแปรดังกล่าวมีความสัมพันธ์กันสูงเกินไป ซึ่งวิธีดังกล่าวเป็นวิธีที่ง่ายที่สุด ตัวอย่างการพิจารณาแสดงดังภาพที่ 5

Correlations								
		Y	X1	X2	X3	X4	X5	X6
Pearson Correlation	Y	1.000	.124	.108	-.050	.723	.579	.403
	X1	.124	1.000	.835	-.025	.131	.076	.044
	X2	.108	.835	1.000	-.041	.123	.069	.074
	X3	-.050	-.025	-.041	1.000	-.032	-.010	-.027
	X4	.723	.131	.123	-.032	1.000	.501	.321
	X5	.579	.076	.069	-.010	.501	1.000	.167
	X6	.403	.044	.074	-.027	.321	.167	1.000
Sig. (1-tailed)	Y		.000	.000	.060	.000	.000	.000
	X1	.000		.000	.223	.000	.009	.089
	X2	.000	.000		.101	.000	.016	.011
	X3	.060	.223	.101		.163	.379	.206
	X4	.000	.000	.000	.163		.000	.000
	X5	.000	.009	.016	.379	.000		.000
	X6	.000	.089	.011	.206	.000	.000	
N	Y	958	958	958	958	958	958	958
	X1	958	958	958	958	958	958	958
	X2	958	958	958	958	958	958	958
	X3	958	958	958	958	958	958	958
	X4	958	958	958	958	958	958	958
	X5	958	958	958	958	958	958	958
	X6	958	958	958	958	958	958	958

ภาพที่ 5 แสดงเมตริกซ์ความสัมพันธ์ระหว่างตัวแปร

จากภาพที่ 1 ผู้เขียนได้จำลองข้อมูลขึ้นมา โดยมี ตัวแปรตาม 1 ตัว คือ Y และตัวแปรอิสระจำนวน 6 ตัว คือ X1, X2, X3, X4, X5 และ X6 จากเมตริกซ์ความสัมพันธ์นี้พบว่าค่า สัมประสิทธิ์สหสัมพันธ์อย่างง่ายของ X1 และ X2 มีค่าเท่ากับ 0.835 ซึ่งอาจมีความหมายโดยนัยได้ว่าตัวแปรอิสระทั้งสองตัวนั้นมีความสัมพันธ์กันสูง

➤ ในการตรวจสอบภาวะ Multicollinearity อาจทำได้โดยวัดสัดส่วนความแปรปรวนของตัวแปรอิสระแต่ละตัวที่ถูกอธิบายโดยชุดของตัวแปรอิสระตัวอื่นๆ ในการศึกษาอื่นๆ โดยกำหนดให้ตัวแปรอิสระดังกล่าวกลายเป็นตัวแปรตาม และถูกอธิบายโดยชุดของตัวแปรอิสระตัวอื่นๆ โดยใช้รูปแบบของ การถดถอยเข้ามาช่วยในการวิเคราะห์ สามารถตรวจสอบได้ 2 วิธี คือ

- การพิจารณาค่า Tolerance ซึ่งเป็นค่าที่บอกถึงสัดส่วนของความแปรปรวนของ ตัวแปรอิสระ นั้นๆ ที่ไม่สามารถอธิบายได้ด้วยชุดของตัวแปรอิสระตัวอื่นๆ โดยมีขั้นตอนการตรวจสอบ คือ กำหนดให้ตัวแปรอิสระหนึ่งในจำนวนตัวแปรอิสระทั้งหมดเป็นตัวแปรตามในการสร้างสมการถดถอย คำนวณค่า R^2 และ คำนวณค่า Tolerance จากสูตร $1 - R^2$ เช่น ถ้าชุดตัวแปรอิสระอื่นๆ สามารถอธิบายความแปรปรวนในตัวแปรนั้นได้ 36 เปอร์เซ็นต์ ($R^2 = 0.36$) ค่า Tolerance จะเท่ากับ $1.0 - 0.36 = 0.64$ หมายความว่า ชุดของตัวแปรอิสระไม่สามารถอธิบายความแปรปรวนในตัวแปรนั้นได้ 64 เปอร์เซ็นต์ จะเห็นได้ว่า ถ้าค่า Tolerance มีค่าสูงจะสื่อให้เห็นว่า มีความสัมพันธ์พหุคูณกับตัวแปรอิสระด้วยกันในระดับต่ำ ในทางกลับกันถ้าค่านี้มีค่าต่ำแสดงว่า มีความสัมพันธ์พหุคูณกับตัวแปรอิสระด้วยกันในระดับสูง ซึ่งทั้งนี้ไม่ควรจะต่ำกว่า 0.1 (Joseph F. Hair et al. 2010 : 204, ทรงศักดิ์ ภูสีอ่อน. 2551 : 289)

- การพิจารณาค่า Variance Inflation Factor (VIF) เป็นค่าที่เป็นส่วนกลับของค่า Tolerance หาได้จากสูตร $1/\text{Tolerance}$ ตัวอย่างเช่น ถ้าค่า Tolerance มีค่าเป็น 0.64 ค่า VIF จะมีค่าเท่ากับ $1/0.64 \approx 1.56$ ในการพิจารณาค่า VIF นั้นไม่ควรมีค่าสูง (ไม่ควรเกิน 10 (Joseph F. Hair et al. 2010 : 204, ทรงศักดิ์ ภูสีอ่อน. 2551 : 289) เนื่องจากเป็นส่วนกลับของ Tolerance ถ้า VIF สูงก็จะบ่งบอกว่าค่า Tolerance ต่ำ ซึ่งหมายความว่า เกิดปัญหา Multicollinearity นั้นเอง โดยค่าทั้งสองนี้จะปรากฏอยู่ในส่วนของผลการวิเคราะห์ Collinearity Statistics ดังภาพที่ 6

Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	46.168	6.030		7.657	.000					
	X1	.270	.171	.057	1.576	.115	.124	.051	.032	.301	3.323
	X2	.000	.001	-.038	-1.042	.298	.108	-.034	-.021	.301	3.324
	X3	-1.270	.972	-.026	-1.307	.192	-.050	-.042	-.026	.997	1.003
	X4	1.912	.090	.515	21.276	.000	.723	.568	.426	.683	1.463
	X5	.314	.025	.280	12.447	.000	.579	.374	.249	.749	1.335
	X6	.807	.090	.189	8.954	.000	.403	.279	.179	.893	1.120

a. Dependent Variable: Y

ภาพที่ 6 แสดงตัวอย่างค่า Tolerance และ VIF ของตัวแปรอิสระแต่ละตัว

นอกจากวิธีที่กล่าวมาข้างต้นนี้แล้วการตรวจสอบภาวะ Multicollinearity ยังสามารถทำได้อีกหลายวิธี เช่น การใช้สมการถดถอยช่วย (Auxiliary regression) การใช้ Ridge Regression เป็นต้น ซึ่งสามารถศึกษาได้จากตำราสถิติทั่วไป

แนวทางในการแก้ไขเมื่อเกิดภาวะ Multicollinearity

ในการแก้ปัญหาดังกล่าวนั้นสามารถทำได้หลายวิธี ดังเช่น

ประการที่ 1 ไม่แก้ไขปัญหาที่เกิดขึ้น การปฏิบัติเช่นนี้จะเห็นเฉพาะในกรณีของสร้างสมการถดถอยเพื่อการพยากรณ์โดยรวมเท่านั้น ไม่ได้คำนึงถึงสัมประสิทธิ์การถดถอย (Joseph F. Hair et al. 2010 : 205) และการคัดเลือกตัวแปรที่ดีที่สุดในการทำนาย

ประการที่ 2 ตัดตัวแปรอิสระที่มีความสัมพันธ์กันสูงออกตัวใดตัวหนึ่ง โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่ายระหว่างตัวแปรอิสระกับตัวแปรตาม โดยเลือกตัวที่มีค่าสูงกว่าไว้ ตัวอย่างเช่น ถ้าย้อนกลับไปยังภาพที่ 1 ซึ่งพบว่า X_1 และ X_2 มีความสัมพันธ์กันสูง และเมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่ายระหว่างตัวแปรอิสระทั้งสองตัวกับตัวแปรตาม Y พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปร X_1 ($r_{xy} = 0.124$) มีค่าสูงกว่าตัวแปร X_2 ($r_{xy} = 0.108$) ดังนั้น ควรเลือกตัด ตัวแปร X_2 ออก เป็นต้น หรืออาจจะพิจารณาจากค่า Partial Correlation (ค่าสัมประสิทธิ์สหสัมพันธ์แยกส่วน) จากการวิเคราะห์การถดถอยก็ได้ โดยเลือกตัวที่มีค่าสูงกว่าไว้เช่นเดียวกัน ตัวอย่างเช่น ดังภาพที่ 2 ก็ยังคงเลือกตัดตัวแปร X_2 ออกเช่นเดิม เป็นต้น แต่วิธีการตัดตัวแปรนี้ถ้าไม่จำเป็นผู้วิจัยควรหลีกเลี่ยง เพราะอาจส่งผลต่อการเพิ่มขึ้นของค่าความคลาดเคลื่อนได้ ดังตัวอย่างภาพที่ 2 ถ้าพิจารณาให้ถี่ถ้วนแล้วจะพบว่า ค่า Tolerance และ VIF ของตัวแปรอิสระทุกตัวยังมีค่าอยู่เกณฑ์ที่กำหนดดังนั้น อาจไม่จำเป็นที่ต้องเลือกตัดตัวแปรอิสระตัวใดทิ้งก็ได้ แต่เลือกแก้ไขปัญหาโดยวิธีการอื่นแทน เช่น การแปลง ตัวแปร เป็นต้น

ประการที่ 3 ใช้วิธีการ Stepwise Regression เนื่องจากวิธีดังกล่าวเป็นวิธีการเลือก ตัวแปรอิสระเข้าไปในสมการถดถอยพหุคูณที่ดีที่สุด เพราะจะพยายามขจัดปัญหาการเกิด Multicollinearity ในตัวเองอยู่แล้ว

ประการที่ 4 การแปลงตัวแปร เมื่อเกิดปัญหา Multicollinearity ในขั้นรุนแรงและในทางทฤษฎีไม่สามารถตัดค่าตัวแปรที่มีปัญหาออกจากสมการได้ ตัวแปรอาจจะถูกแปลงให้อยู่ในรูปอื่นๆ แทน เช่น สร้างรูปแบบความสัมพันธ์เชิงเส้น (linear combination) ของตัวแปรที่มีปัญหา คือ สร้างตัวแปรใหม่ที่เป็นสมการกับตัวแปรที่มีปัญหา และใช้ตัวแปรใหม่แทนในการสร้างสมการถดถอย ตัวอย่างเช่น ถ้า X_1 และ X_2 มีความสัมพันธ์กันสูงมาก ก็สร้างตัวแปรใหม่ คือ $X_3 = X_1 + X_2$ แทนค่าลงในแบบจำลองเดิม

$$y_i = \beta_1 + \beta_3 x_3 + u_i$$

$$\text{นั่นคือ} \quad = \beta_1 + \beta_3 (x_1 + x_2) + u_i$$

นอกจากนี้ยังมีวิธีการแก้ไขอื่นๆ อีก เช่น การเลือกใช้วิธีการของ Bayesian regression (Ridge Regression) เป็นต้น

จากที่กล่าวข้างต้นจะเห็นว่าภาวะ Multicollinearity เป็นภาวะที่ตัวแปรมีความสัมพันธ์กันสูง จนเกือบจะเป็นตัวเดียวกัน ซึ่งในกรณีของการถดถอยพหุคูณนั้นเมื่อเกิดภาวะ Multicollinearity ขึ้น จะส่งผลให้เกิดความคลาดเคลื่อนในเรื่องของ

- 1) ความสามารถในการทำนายของตัวแปรทำนายที่เกิดภาวะดังกล่าวลดลง
- 2) สัมประสิทธิ์การถดถอยของตัวแปรนั้นเกิดความไม่คงที่ทั้งขนาดและทิศทาง
- 3) ตัวแปรนั้นทำนายอาจไม่มีนัยสำคัญในการทำนาย
- 4) การประมาณค่ามีความไวต่อขนาดของกลุ่มตัวอย่าง

สุดท้ายส่งผลให้การสร้างสมการพยากรณ์ตัวแปรเกณฑ์ขาดความแม่นยำ ดังนั้น ภาวะนี้จึงเป็นภาวะที่ผู้วิจัยไม่ควรละเลยต่อผลที่จะเกิดขึ้น ควรมีการตรวจสอบอย่างถี่ถ้วนในขั้นตอนของการวิเคราะห์การถดถอยพหุคูณว่า มีปัญหานี้เกิดขึ้นหรือไม่ ซึ่งสามารถกระทำได้หลายวิธี อาทิเช่น การตรวจสอบจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร ค่า Tolerance และค่า VIF เป็นต้น เมื่อพบว่ามีปัญหาในเรื่องของ Multicollinearity จริง ผู้วิจัยควรหาวิธีการในการแก้ไขปัญหาดังกล่าว เช่น การตัดตัวแปรตัวใดตัวหนึ่งออก การเลือกใช้การวิเคราะห์การถดถอยที่มีการคัดเลือกตัวแปรที่ดีที่สุดเข้าสู่สมการ การแปลงตัวแปรในรูปของสมการเชิงเส้นก่อนนำไปวิเคราะห์ใหม่ เป็นต้น ซึ่งในการแก้ไขปัญหานี้ผู้วิจัยควรคำนึงกรอบทฤษฎีที่ใช้ในการศึกษาด้วยเพื่อให้ได้แนวทางที่ดีที่สุด และส่งผลต่อการได้สมการพยากรณ์ที่มีประสิทธิภาพในการทำนายสูงสุด

เอกสารอ้างอิง

- ทรงศักดิ์ ภูสีอ่อน. (2551). การประยุกต์ใช้ SPSS วิเคราะห์ข้อมูลทางการวิจัย. มหาสารคาม : สำนักพิมพ์มหาวิทยาลัยมหาสารคาม.
- สุชาติ ประสิทธิ์รัฐสินธุ์. (2545). การใช้สถิติในงานวิจัยอย่างถูกต้องและได้มาตรฐาน. กรุงเทพฯ : เพื่องฟ้า พรินติ้ง.
- อารี วิบูลย์พงศ์. (มปป). เศรษฐกิจประยุกต์สำหรับการตลาดเกษตร. สืบค้นเมื่อวันที่ 15 สิงหาคม 2553, จาก http://web.agri.cmu.ac.th/aec/AEC_Home/web_econometric/link_econ.htm
- Tabachnick B.G. and Fidell L.S. (2007). Using Multivariate Statistics. Peason Prentice Hall.
- Hair, Joseph F. and others. (2010). Multivariate Data Analysis A Global Perspective. Peason Prentice Hall.