Journal of Research Methodology, Volume 12, Number 2 (July - December 1999)

วารสารวิธีวิทยาการวิจัย ปีที่ 12 ฉบับที่ 2 (กรกฎาคม - ธันวาคม 2542)

Nonlinear Regression

Siridej Sujiva

ABSTRACT

The article aims to explain the concept and principle of nonlinear regression with its anlaysis method. It covers the characteristics of nonlinear model, nonlinear model transforming, the estimation of starting values and parameter and the approximation of confidence intervals for parameters along with explicit examples.

Journal of Research Methodology, Volume 12, Number 2 (July - December 1999)
วารสารวิธีวิทยาการวิจัย ปีที่ 12 ฉบับที่ 2 (กรกฎาคม - ธันวาคม 2542)

Nonlinear Regression

ศิริเดช สุชีวะ

บทคัดย่อ

บทความนี้มีวัตถุประสงค์เพื่ออธิบายแนวคิด หลักการ และวิธีการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น (nonlinear regression) โดยนำเสนอเนื้อหาเกี่ยวกับคุณลักษณะของโมเดลที่ไม่ใช่เชิงเส้น (nonlinear model) การแปลงเป็นโมเดลเชิงเส้น การกำหนดค่าเริ่มต้น การประมาณค่าพารามิเตอร์ในโมเดล และการประมาณช่วงความเชื่อมั่นของค่าพารามิเตอร์ โดยการยกตัวอย่างประกอบ

ความนำ

สถิติวิเคราะห์ที่ใช้ในการศึกษารูปแบบความสัมพันธ์ส่วนใหญ่จะอยู่บนพื้นฐานของโมเดลเชิงเส้น (linear model) โดยเฉพาะในกรณีที่นักวิจัยยังไม่มีโมเดลเชิงทฤษฎีเกี่ยวกับสิ่งที่จะวิเคราะห์ แต่ในหลาย ๆ ครั้งที่เห็นได้ชัดเจนว่าไม่สามารถจะใช้โมเดลเชิงเส้นมาวิเคราะห์ได้ อาทิ ความสัมพันธ์ระหว่างความเร็วรถกับระยะเวลาที่ใช้หยุดรถ ซึ่งกฎทางฟิสิกส์ก็บ่งชี้แล้วว่าไม่ใช่ความสัมพันธ์เชิงเส้น เราก็จะไม่วิเคราะห์ด้วยโมเดลเชิงเส้น ในการวิจัยทางพฤติกรรมศาสตร์ นักวิจัยนิยมใช้โมเดลเชิงเส้นเป็นส่วนใหญ่ เช่น กรณีของการวิเคราะห์เงินเดือนในฐานะที่เป็นฟังก์ชั่นเชิงเส้นกับตัวแปรอื่น ๆ เช่น อายุ ระดับการศึกษา ประสบการณ์ เป็นต้น เหตุผลหนึ่งที่ทำให้นักวิจัยเลือกใช้โมเดลเชิงเส้นก็คือ ความง่ายในการประมาณค่าทางสถิติและการทดสอบสมมติฐาน โดยกระบวนการในการประมาณค่าพารามิเตอร์ของโมเดลเชิงเส้นจะตรงไปตรงมา ได้คำตอบทันทีโดยไม่จำเป็นต้องคำนวณทวนซ้ำ (iteration) อย่างไรก็ดีในบางสถานการณ์ที่ข้อมูลนั้นไม่ได้มีความสัมพันธ์เชิงเส้น โดยเฉพาะกรณีของการวิเคราะห์ถดถอย โมเดลการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น (Nonlinear Regression) จะให้คำตอบกับนักวิจัยได้

คุณลักษณะของโมเดลที่ไม่ใช่เชิงเส้น (Nonlinear Model)

การตัดสินว่าโมเดลใดเป็นโมเดลเชิงเส้นนั้น จะมีวิธีการอย่างไร ขอให้ท่านพิจารณาสมการต่อไปนี้

$$Y = B_0 + B_1 X_1^2 \qquad (1) \quad \text{(Norusis, 1994)}$$

ท่านคิดว่าสมการนี้เป็นสมการเชิงเส้นหรือไม่ แน่นอนว่าถ้าแทนค่าในสมการนี้แล้วจะไม่ได้กราฟเส้นตรง แต่จะเป็นรูปพาราโบลา แต่คำว่า ‘เชิงเส้น’ ในบริบทนี้ ไม่ได้อยู่ที่ว่าสมการนั้นจะให้ผลเป็นเส้นตรงหรือเส้นโค้ง แต่อยู่ที่รูปแบบฟังก์ชั่นของสมการ กล่าวคือ เราจะพิจารณาว่าตัวแปรตามนั้นสามารถแทนได้ด้วยผลรวมเชิงเส้น (linear combination) ของค่าพารามิเตอร์ของตัวแปรอิสระแต่ละตัวหรือไม่ ค่าพารามิเตอร์จะต้องเป็นเชิงเส้น ในขณะที่ตัวแปรอิสระจะถูกเปลี่ยนให้อยู่ในรูปใดก็ได้ ซึ่งจะไม่กระทบกับค่าพารามิเตอร์

ดังนั้นสมการ (1) เป็นโมเดลเชิงเส้น เพราะว่าโมเดลนี้เป็น nonlinear เฉพาะตัวแปรอิสระ X เท่านั้น แต่มีความเป็นเชิงเส้นในค่าพารามิเตอร์ B_0 และ B_1 ซึ่งเราสามารถเขียนสมการดังกล่าวในรูปนี้ก็ได้

$$Y = B_0 + B_1 X' \qquad (2) \quad \text{(Norusis, 1994)}$$

โดยที่ X' เป็นกำลังสองของ X_1 นักวิจัยสามารถประมาณค่าพารามิเตอร์ในโมเดลนี้ โดยใช้โมเดลเชิงเส้นได้เลย

การแปลงโมเดลให้เป็นเชิงเส้น

ขอให้ท่านผู้อ่านพิจารณาสมการ (3) และสมการ (4)

$$Y = e^{B_0 + B_1 X_1 + B_2 X_2 + E} \quad \text{(3)} \quad \text{(Norusis, 1994)}$$

$$Y = B_0 + B_1 Z_1 + B_2 Z_2 + \dots + B_P Z_P + E \quad \text{(4)} \quad \text{(Norusis, 1994)}$$

สมการ (4) มี b เป็นค่าพารามิเตอร์ และ Z เป็นฟังก์ชั่นของตัวแปรอิสระ ในขณะที่สมการ (3) เป็นโมเดลที่ไม่ใช่เชิงเส้น อย่างไรก็ดีถ้าทำสมการ (3) ให้เป็นล็อกธรรมชาติทั้งสองข้าง เราจะได้สมการ (5) ดังนี้

$$\ln(Y) = B_0 + B_1 X_1 + B_2 X_2 + E \quad \text{(5)} \quad \text{(Norusis, 1994)}$$

สมการ (5) ที่แปลงแล้วนี้เป็นเชิงเส้นในค่าพารามิเตอร์ ซึ่งเราสามารถใช้เทคนิคเชิงเส้นในการประมาณค่าได้ โมเดลที่ไม่ใช่เชิงเส้นซึ่งสามารถแปลงให้อยู่ในรูปโมเดลเชิงเส้นได้ เรียกว่า intrinsically linear model ดังนั้นนักวิจัยควรพิจารณาก่อนว่าโมเดลที่จะวิเคราะห์นั้นสามารถแปลงให้อยู่ในรูปโมเดลเชิงเส้นได้หรือไม่ อันจะทำให้การประมาณค่าพารามิเตอร์สะดวกกว่ากันมาก ดังตัวอย่างต่อไปนี้

$$Y = e^B X + E \quad \text{(6)} \quad \text{(Norusis, 1994)}$$

ถ้าเราแปลงค่า e^B ให้เป็น B'(จะได้เป็นสมการ (7)

$$Y = B' X + E \quad \text{(7)} \quad \text{(Norusis, 1994)}$$

ซึ่งเราสามารถใช้วิธีการปกติในการประมาณค่า B′ และใช้ล็อกธรรมชาติแก้สมการให้ได้ค่าของ b

ในโมเดลที่เป็นเชิงเส้นและที่ไม่เป็นเชิงเส้น เรามีข้อตกลงเบื้องต้นว่าเทอมความคลาดเคลื่อนเป็นโมเดลเชิงบวก (additive) ดังนั้นเมื่อนักวิจัยแปลงโมเดลให้เป็นเชิงเส้นแล้ว ต้องตรวจสอบให้แน่ใจว่าเทอมความคลาดเคลื่อนที่แปลงแล้วยังคงสอดคล้องกับข้อตกลงเบื้องต้นหรือไม่ ตัวอย่างเช่น ในกรณีที่โมเดลแรกเริ่มของเราเป็นดังนี้

$$Y = e^{BX} + E \quad \text{(8)} \quad \text{(Norusis, 1994)}$$

จากสมการ (8) การทำให้เป็นล็อกธรรมชาติ ไม่ได้มีผลในโมเดลซึ่งมีเทอมความคลาดเคลื่อนที่เป็นเชิงบวก แต่ก็มีบางโมเดลที่เราไม่สามารถแปลงให้เป็นเชิงเส้นได้ ที่เรียกว่า intrinsically nonlinear ดังโมเดลในสมการ (9)

$$Y = B_0 + e^{B_1 X_1} + e^{B_2 X_2} + e^{B_3 X_3} + E \quad (9)$$ (Norusis, 1994)

เราต้องประมาณค่าพารามิเตอร์โดยใช้การวิเคราะห์ถดถอยแบบไม่ใช่เชิงเส้น (nonlinear regression) การวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้นนี้คล้ายกับการวิเคราะห์ถดถอยเชิงเส้นปกติตรงที่เราจะเลือกค่าของพารามิเตอร์เพื่อให้ผลรวมกำลังสองของค่าเศษเหลือ (sum of squared residual) มีค่าน้อยที่สุด แต่ในการประมาณค่าพารามิเตอร์ในการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้นนี้ต้องใช้การคำนวณทวนซ้ำ (interation)

ตัวอย่างของโมเดลที่ไม่เป็นเชิงเส้น เช่น โมเดลการเพิ่มประชากร อัตราการเพิ่มประชากรมักจะแสดงด้วยโมเดลโลจิสติก ดังนี้

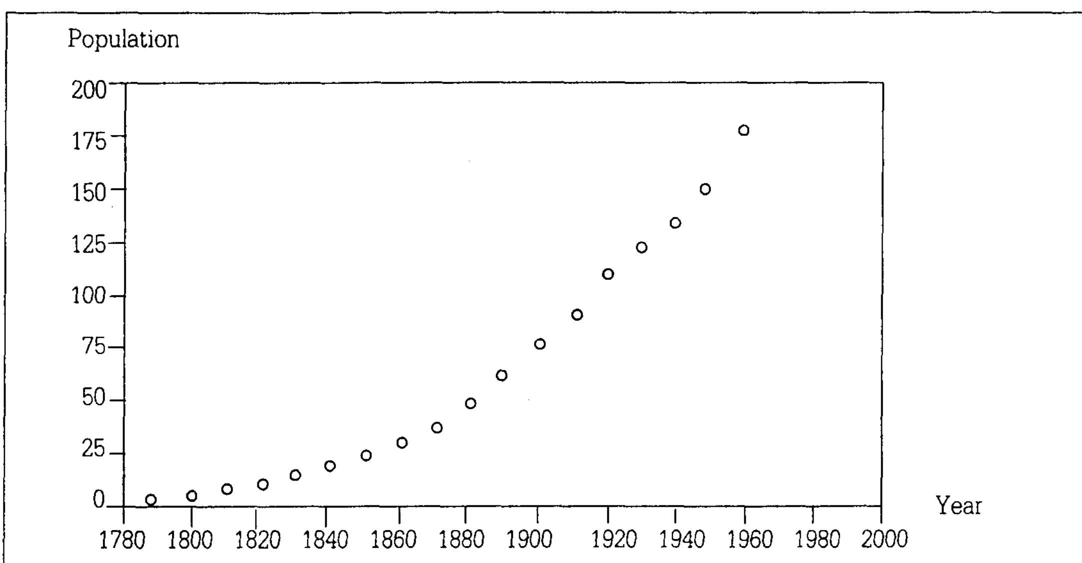
$$Y_i = \frac{C}{1 + e^{A+BT_i}} + E_i \quad (10)$$ (Norusis, 1994)

โดยที่ Y_i คือขนาดของประชากร ณ เวลา T_i แม้ว่าโมเดลนี้จะเหมาะกับข้อมูลที่สังเกตได้ แต่อาจจะฝ่าฝืนข้อตกลงเบื้องต้นเรื่องความคลาดเคลื่อนที่เป็นอิสระและความคงที่ของความแปรปรวน เนื่องจากความคลาดเคลื่อนของข้อมูลอนุกรมเวลาไม่เป็นอิสระ และขนาดของความคลาดเคลื่อนก็อาจขึ้นอยู่กับขนาดของประชากร อีกทั้งโมเดลโลจิสติกของการเพิ่มประชากรไม่สามารถแสดงให้เป็นโมเดลเชิงเส้นได้ ดังนั้น นักวิจัยจึงต้องใช้การวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้นมาประมาณค่าพารามิเตอร์

เพื่อให้เข้าใจได้ชัดเจน ขอให้ผู้อ่านศึกษาตัวอย่างการวิจัยของ Fox (1984) เกี่ยวกับโมเดลการเพิ่มประชากรที่แสดงในคู่มือโปรแกรม SPSS Advanced Statistics (Norusis, 1994) ตารางที่ 1 แสดงจำนวนประชากร (ล้านคน) ในสหรัฐอเมริกา ตั้งแต่ปี ค.ศ. 1790 ถึง 1960 กราฟรูปที่ 2 เป็นการพล็อตข้อมูลชุดเดียวกันนี้

POP	YEAR	DECADE
3.895	1790	0
5.267	1800	1
7.182	1810	2
9.566	1820	3
12.834	1830	4
16.985	1840	5
23.069	1850	6
31.278	1860	7
38.416	1870	8
49.924	1880	9
62.692	1890	10
75.734	1900	11
91.812	1910	12
109.806	1920	13
122.775	1930	14
131.669	1940	15
150.697	1950	16
178.464	1960	17

ตารางที่ 1 จำนวนประชากรของสหรัฐอเมริกา (Norusis, 1994)



กราฟที่ 2 แสดงความสัมพันธ์ระหว่างปี ค.ศ. กับจำนวนประชากร (Norusis, 1994)

ในการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น เราจะใช้ตัวแปร “ทศวรรษ” (DECADE) ซึ่งเป็นตัวเลขแสดงจำนวนทศวรรษ ตั้งแต่ปี 1790 เป็นตัวแปรอิสระ ซึ่งมีข้อดีคือสามารถหลีกเลี่ยงความยุ่งยากในการคำนวณตัวเลขมาก ๆ ได้

เมื่อเริ่มการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น เราจำเป็นต้องมีค่าพารามิเตอร์เริ่มต้นซึ่งสำคัญมาก เพราะผลของการประมาณค่าพารามิเตอร์ในโมเดลนี้ไม่ใช่เชิงเส้น มักจะขึ้นอยู่กับการมีค่าเริ่มต้นในการประมาณค่าพารามิเตอร์ที่ดี ในกรณีตัวอย่างการวิเคราะห์นี้ เราหาค่าเริ่มต้นได้จากข้อตกลงของโลจิสติกโมเดลของการเพิ่มประชากร ที่ว่าค่าพารา-มิเตอร์ C จะเป็นค่า asymptote เราจึงเลือกค่า asymptote ซึ่งใกล้เคียงกับค่าสังเกตได้ที่มากที่สุดเป็นค่าเริ่มต้น เราใช้ค่า asymptote เป็น 200 เนื่องจากค่าสังเกตได้ของประชากรมีค่าสูงสุดเป็น 178

เมื่อใช้ค่า 200 แทนค่า C เราสามารถประมาณค่า A จากค่าสังเกตได้ของประชากร ณ เวลา 0 ดังสมการต่อไปนี้

$$3.895 = \frac{200}{1+e^{A}} \qquad (11) \quad \text{(Norusis, 1994)}$$

ดังนั้น

$$A = \ln\left(\frac{200}{3.895} - 1\right) = 3.9 \qquad (12) \quad \text{(Norusis, 1994)}$$

ในการประมาณค่า B เราใช้ค่าประชากร ณ เวลาเท่ากับ 1 กับค่าที่ประมาณได้ของ C และ A ดังนี้

$$5.267 = \frac{200}{1 + e^{B+3.9}} \qquad (13) \quad \text{(Norusis, 1994)}$$

ซึ่งเราสามารถเขียนได้ว่า

$$B = \ln\left(\frac{200}{5.27} - 1\right) - 3.9 = -0.29 \qquad (14) \quad \text{(Norusis, 1994)}$$

เราใช้ค่าที่ประมาณได้เหล่านี้เป็นค่าเริ่มต้นในการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น ดังจะได้กล่าวต่อไป

การประมาณค่าเริ่มต้นที่ดี นอกจากจะทำให้ได้คำตอบที่ถูกต้องด้วยการคำนวณทวนซ้ำที่น้อยกว่าแล้วยังจะช่วยหลีกเลี่ยงปัญหาในการคำนวณได้ด้วย การประมาณค่าเริ่มต้นในการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้นมีหลายวิธี ถ้าหากเราไม่มีค่าเริ่มต้นก็ไม่ควรกำหนดให้ค่าเริ่มต้นทั้งหมดเป็นศูนย์ ขอให้ใช้ค่าที่คาดหวังว่าจะเป็นมากที่สุดเป็นค่าเริ่มต้น

ถ้าหากเราไม่สนใจเทอมความคลาดเคลื่อน เราก็อาจจะใช้การวิเคราะห์ถดถอยเชิงเส้นในการคำนวณค่าเริ่มต้นได้ ดังเช่นโมเดลต่อไปนี้

$$Y = e^{A+BX} + E \quad \text{(15)} \quad \text{(Norusis, 1994)}$$

หากเราทำให้เป็นล็อกธรรมชาติทั้งสองข้างของสมการ ก็จะได้โมเดลดังนี้

$$ln\ (Y) = A + BX \quad \text{(16)} \quad \text{(Norusis, 1994)}$$

เราก็สามารถใช้การวิเคราะห์ถดถอยเชิงเส้นประมาณค่า A และ B แล้วใช้ค่าเหล่านี้เป็นค่าเริ่มต้นในการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้นได้

ในบางกรณีเราทราบค่าตัวแปรตามจากผลรวมของค่าพารามิเตอร์ ดังเช่นตัวอย่างต่อไปนี้

$$Y = e^{A+BX} \quad \text{(17)} \quad \text{(Norusis, 1994)}$$

เราทราบว่า เมื่อ X เท่ากับ 0 แล้ว Y จะเท่ากับ 2 เราจึงเลือกล็อกธรรมชาติของ 2 เป็นค่าเริ่มต้นสำหรับ A นักวิจัยอาจพิจารณาค่าสูงสุด ค่าต่ำสุดของสมการ และเมื่อตัวแปรอิสระทุกตัวมีค่าเข้าใกล้ 0 หรืออนันต์ มาช่วยในการกำหนดค่าเริ่มต้นก็ได้

การประมาณค่าพารามิเตอร์ในการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น

ตารางที่ 3 แสดงค่าผลรวมกำลังสองของค่าเศษเหลือ (residual sum of squares) และค่าพารามิเตอร์ที่ประมาณได้ในแต่ละรอบของการคำนวณทวนซ้ำ ในขั้นที่ 1 ค่าประมาณพารามิเตอร์จะเป็นค่าเริ่มต้นที่เรากำหนดให้ การคำนวณทวนซ้ำหลักซึ่งมักกำหนดค่าเริ่มต้นเป็นเลขจำนวนเต็มจะเป็นการประเมินทิศทาง ส่วนการคำนวณทวนซ้ำรองจะเป็นการกำหนดระยะทาง ดังที่กล่าวไว้ตอนท้ายของตารางที่ 2 การคำนวณทวนซ้ำจะหยุดเมื่อการเปลี่ยนแปลงสัมพัทธ์ของผลรวมกำลังสองของเศษเหลือระหว่างรอบของการคำนวณทวนซ้ำ น้อยกว่าหรือเท่ากับเกณฑ์ลู่เข้า (convergence criterion)

Iteration	Residual SS	A	B	C
1	969.6898219	3.90000000	-.30000000	200.000000
1.1	240.3756627	3.87148504	-.27852485	237.513990
2	240.3756627	3.87148504	-.27852485	237.513990
2.1	186.5020615	3.89003377	-.27910189	243.721558
3	186.5020615	3.89003377	-.27910189	243.721558
3.1	186.4972404	3.88880287	-.27886478	243.975460
4	186.4972404	3.88880287	-.27886478	243.975460
4.1	186.4972278	3.88885123	-.27886164	243.985980
5	186.4972278	3.88885123	-.27886164	243.985980
5.1	186.4972277	3.88884856	-.27886059	243.987296

Run stopped after 10 model evaluations and 5 derivative evaluations.
Iterations have been stopped because the relative reduction between successive residual sums of squares is at most SSCON = 1.00E-08

ตารางที่ 2 ผลการประมาณค่าพารามิเตอร์ (Norusis, 1994)

ผลการวิเคราะห์ค่าสถิติของการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น แสดงไว้ในตารางที่ 3 สถิติทดสอบที่ใช้ในโมเดลเชิงเส้น ไม่เหมาะที่จะใช้กับโมเดลที่ไม่ใช่เชิงเส้น ดังนั้นในสถานการณ์นี้ ค่า residual mean square (RMS) ไม่ใช่ตัวประมาณค่าที่ไม่ลำเอียงของความแปรปรวนของความคลาดเคลื่อน ในทางปฏิบัติ เรายังคงสามารถเปรียบเทียบความแปรปรวนของเศษเหลือกับค่าประมาณของความแปรปรวนทั้งหมด แต่สถิติทดสอบ F ไม่สามารถใช้ในการทดสอบสมมติฐานได้

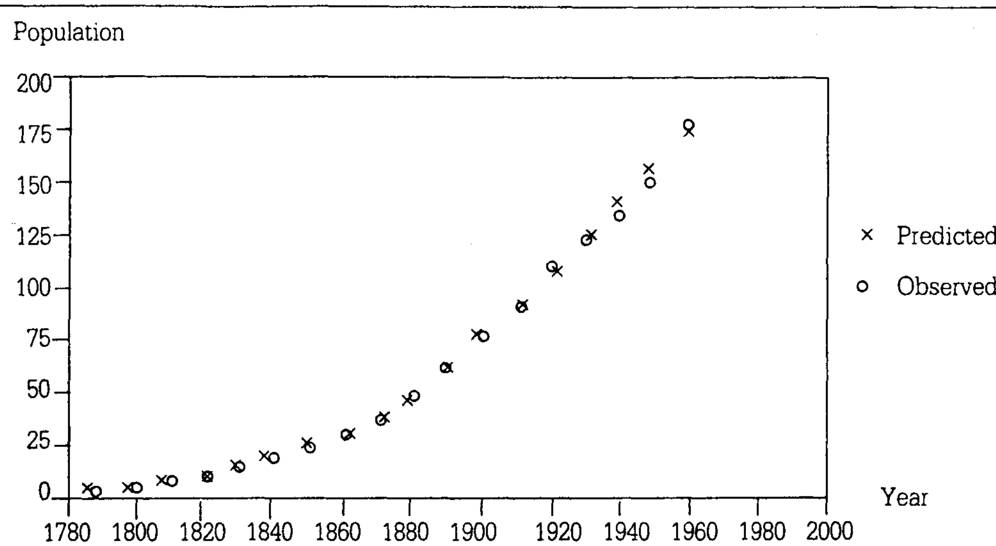
จากผลการวิเคราะห์ในตารางที่ 3 ในส่วน Uncorrected Total เป็นค่าผลรวมกำลังสองของตัวแปรตาม ส่วน Corrected Total เป็นค่าผลรวมของส่วนเบี่ยงเบนกำลังสองโดยรอบค่าเฉลี่ย ค่า Regression Sum of Squares เป็นค่าผลรวมของกำลังสองของค่าที่ทำนายได้ และ R Square เป็นค่าสัมประสิทธิ์การทำนาย (coefficient of determination) ซึ่งอาจแปลความหมายได้ว่าเป็นสัดส่วนของความแปรปรวนทั้งหมดของตัวแปรตามโดยรอบค่าเฉลี่ยที่สามารถอธิบายได้ด้วยโมเดลนี้ ในโมเดลที่ไม่ใช่เชิงเส้น ค่าจะเป็นลบเมื่อโมเดลที่ได้มีความเหมาะสมกับข้อมูลน้อยกว่าค่าเฉลี่ย จากค่า R^2 ที่เท่ากับ 0.9965 แสดงให้เห็นว่าโมเดลเหมาะสมกับข้อมูลเชิงประจักษ์ กราฟรูปที่ 2 แสดงความสัมพันธ์ระหว่างค่าที่สังเกตได้และค่าที่ทำนายได้จากโมเดล

Nonlinear Regression Summary Statistics — Dependent Variable POP

Source	DF	Sum of Squares	Mean Square
Regression	3	123053.53112	41017.84371
Residual	15	186.49723	12.43315
Uncorrected Total	18	123240.02834	
(Corrected Total)	17	53293.92477	

R Squared = 1 - Residual SS / Corrected SS = .99650

ตารางที่ 3 ผลการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น (Norusis, 1994)



กราฟที่ 2 ความสัมพันธ์ระหว่างค่าสังเกตและค่าที่ทำนายได้ (Norusis, 1994)

การประมาณช่วงความเชื่อมั่นของค่าพารามิเตอร์

ในการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น เราไม่สามารถคำนวณช่วงความเชื่อมั่นที่แน่นอนสำหรับพารามิเตอร์แต่ละตัวได้ เราจึงต้องอาศัยการประมาณค่าจากกลุ่มตัวอย่างขนาดใหญ่ในตารางที่ 4 แสดงค่าประมาณพารามิเตอร์ ค่าความคลาดเคลื่อนมาตรฐานและช่วงความเชื่อมั่น 95% ส่วนตารางที่ 5 แสดงเมตริกสหสัมพันธ์ของค่าประมาณพารามิเตอร์ ซึ่งถ้าหากว่าค่าสัมประสิทธิ์สหสัมพันธ์มีขนาดใหญ่ ไม่ว่าจะเป็นบวกหรือลบก็ตาม อาจจะเป็นไปได้ว่าโมเดลนั้นประมาณค่า

พารามิเตอร์มากเกินไป (over parameterized) ดังนั้นโมเดลที่มีค่าพารามิเตอร์น้อยกว่านี้อาจจะสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า แต่ก็ไม่ได้หมายความว่าโมเดลไม่เหมาะสม แต่หมายความว่ามีข้อมูลไม่เพียงพอที่จะประมาณค่าพารามิเตอร์ทั้งหมด

			Asymptotic 95%	
		Asymptotic	Confidence Interval	
Parameter	Estimate	Std. Error	Lower	Upper
A	3.888848562	.093704407	3.689122346	4.088574778
B	-.278860588	.015593951	-.312098308	-.245622868
C	243.98729636	17.967399750	205.69069033	282.28390239

ตารางที่ 4 ค่าประมาณพารามิเตอร์ (Norusis,1994)

Asymptotic Correlation Matrix of the Parameter Estimates

	A	B	C
A	1.0000	-.7244	-.3762
B	-.7244	1.0000	.9042
C	-.3762	.9042	1.0000

รูปที่ 7 เมตริกสหสัมพันธ์ของค่าประมาณพารามิเตอร์ (Norusis,1994)

จากผลการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้นเพื่อทำนายจำนวนประชากรของสหรัฐอเมริกาในรอบ 10 ปี ซึ่งจากข้อมูลเชิงประจักษ์พบว่าความสัมพันธ์ระหว่างจำนวนประชากร และปีค.ศ.ที่เพิ่มขึ้นไม่ได้เป็นเชิงเส้น ปรากฏว่าสมการถดถอยที่ไม่ใช่เชิงเส้นสามารถทำนายได้ถึงร้อยละ 99 ในขณะที่ข้อมูลชุดเดียวกันนี้เมื่อวิเคราะห์ด้วยการวิเคราะห์ถดถอยเชิงเส้น (linear Regression) จะสามารถทำนายได้เพียงร้อยละ 92 (adjusted R^2 = 0.921) ดังตารางที่ 6

Model		Sum of Squares	df	Mean Squares	F	Sig.
1	Regression	49337.679	1	49337.679	199.533	.000[a]
	Ressidual	3956.246	16	247.265		
	otal	53293.925	17			

a. Predictors: (Constant), DECADE

b. Dependent Variable : POP

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.962[a]	.962[a]	.961	15.72467

ตารางที่ 6 ผลการวิเคราะห์ถดถอยเชิงเส้น

สรุป

แม้ว่าการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์ส่วนใหญ่นิยมที่จะใช้โมเดลเชิงเส้นด้วยความสะดวกในการประมาณค่าพารามิเตอร์และการทดสอบสมมติฐาน แต่ในกรณีที่ข้อมูลที่จะทำการวิเคราะห์ มิได้เป็นความสัมพันธ์เชิงเส้น โดยเฉพาะในกรณีของการวิเคราะห์ถดถอย นักวิจัยก็ควรหันมาใช้โมเดลการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น (nonlinear regression) ซึ่งนักวิจัยควรจะตรวจสอบก่อนว่าโมเดลที่จะวิเคราะห์นั้นสามารถแปลงให้เป็นโมเดลเชิงเส้น (linear model) ได้หรือไม่ ถ้าโมเดลนั้นไม่สามารถแปลงให้เป็นเชิงเส้นได้ (intrincially nonlinear) จึงจะใช้การวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้น ขั้นตอนสำคัญของการวิเคราะห์ถดถอยที่ไม่ใช่เชิงเส้นขั้นตอนหนึ่ง คือ การกำหนดค่าเริ่มต้นในโมเดลซึ่งจะมีผลต่อคุณภาพของผลการวิเคราะห์ ส่วนการประมาณค่าพารามิเตอร์และการประมาณค่าช่วงความเชื่อมั่นก็มีส่วนที่แตกต่างไปจากผลการวิเคราะห์ถดถอยเชิงเส้นบ้าง ซึ่งนักวิจัยควรศึกษาทำความเข้าใจให้รอบคอบก่อน

รายการอ้างอิง

Draper, N.R., and Smith, H.(1981). ***Applied regression analysis.*** New York : John Wiley and Sons.

Fox, J. (1984) ***Linear statistical models and related methods : With applications to social research.*** New York : John Wiley and Sons.

Norusis, M.J. (1994)***SPSS advanced statistics user's guide.*** Chicago : SPSS Inc.