การคาดการณ์ประชากร (1):

แบบจำลองเชิงเส้นตรง และแบบจำลองเชิงทวีกำลัง

(Linear and Exponential Models)

โดย รศ. ดร. วรรณศิลป์ พีรพันธุ์ © 2001-2003

สาระสำคัญของบทความนี้ เป็นการอธิบายถึงวิธีการคาดการณ์จำนวนประชากรในอนาคตของพื้นที่โดยอาศัย "Extrapolation Techniques" ซึ่งเป็นการใช้แนวโน้มการเติบโตของประชากรในอดีตมาคาดการณ์จำนวนประชากรในอนาคต แบบจำลองที่จะกล่าวถึงต่อไปนี้เป็นแบบจำลองที่นิยมใช้ในการคาดการณ์ประชากรโดยทั่วไป ได้แก่ แบบจำลองเชิงเส้นตรง (Linear Model) และ แบบจำลองเชิงทวีกำลัง (Exponential Model)

1. แบบจำลองเชิงเส้นตรง (Linear Model)

แบบจำลองเชิงเส้นตรงเป็นแบบจำลองที่มีรูปแบบง่าย ๆ และมีการนำไปประยุกต์ใช้อย่างแพร่หลายในการคาดการณ์ประชากร แบบจำลองนี้จะใช้ได้เมื่อประชากรในอดีตของพื้นที่มีการเพิ่มขึ้นในแต่ละช่วงเวลาเป็นจำนวนค่อนข้างคงที่ และมีแนวโน้มว่ารูปแบบดังกล่าวจะยังคงดำเนินต่อไปในอนาคต ซึ่งในทางคณิตศาสตร์แล้วเราสามารถจะคาดการณ์ประชากรในอนาคตของพื้นที่ศึกษาได้โดยใช้สมการถดถอยเชิงเส้นตรงอย่างง่าย (Simple Linear Regression) ซึ่งมีรูปแบบทั่วไปคือ:

$$Y_c = a + bX \quad \text{..................... [1]}$$

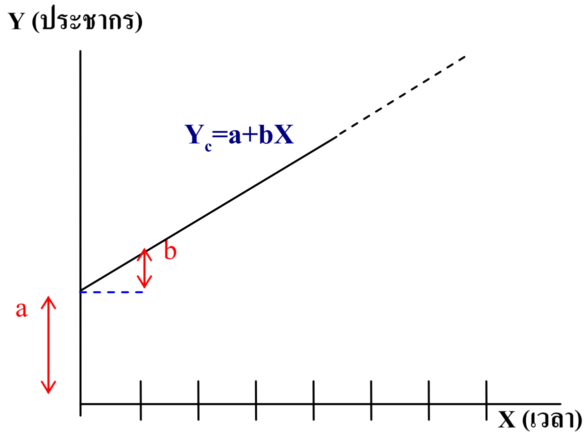
โดยที่: Y_c = ตัวแปรตาม (Dependent Variable) ซึ่งในที่นี้ได้แก่จำนวนประชากรที่คาดการณ์

X = ตัวแปรอิสระ (Independent Variable) ซึ่งในที่นี้ได้แก่ช่วงเวลา (Time Index)

A = ค่าตัวคั่น (Y-intercept) หรือค่าของ Y_c เมื่อ $X = 0$

b = ค่าความชันของเส้นสมการ (Slope) หรือค่า Y_c ที่เปลี่ยนแปลงไปเมื่อ X เปลี่ยนไป 1 หน่วย ซึ่งในที่นี้ก็คือจำนวนประชากรที่เพิ่มขึ้นหรือลดลงต่อหน่วยเวลา (เช่นต่อปี)

การที่เส้น Y_c ปรากฏเป็นเส้นตรงก็เนื่องจากการเพิ่มขึ้นต่อหน่วยของตัวแปรอิสระ X (ช่วงเวลา) ทำให้ตัวแปรตาม Y (ประชากร) เพิ่มขึ้นเป็นจำนวนที่คงที่ กล่าวอีกนัยหนึ่งก็คือค่าความชัน b ของเส้นสมการ (หรือจำนวนประชากรที่เปลี่ยนแปลงต่อหน่วยเวลา) มีค่าคงที่นั่นเอง



แผนภูมิที่ 1 ความสัมพันธ์ของจำนวนประชากรกับช่วงเวลาในสมการเชิงเส้นตรง

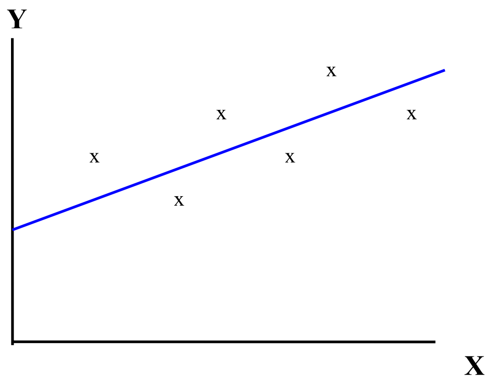
Least Squares Straight Line

โดยทั่วไปแล้ว เกือบจะเป็นไปไม่ได้เลยที่การเติบโตของประชากรในอดีตของพื้นที่จะมีลักษณะเป็นเส้นตรงโดยสมบูรณ์ ดังนั้นก่อนที่จะหาค่า a และ b ในสมการ [1] จึงจำเป็นต้องหาเส้นตรงซึ่งเป็นตัวแทนที่ดีที่สุดของข้อมูลในอดีตให้ได้เสียก่อน วิธีการที่นิยมกันมากก็คือการหาเส้นตรงกำลังสองน้อยที่สุด (Least Square Straight Line) ซึ่งเป็นเส้นตรงที่ก่อให้เกิดผลรวมของความแตกต่างยกกำลังสองน้อยที่สุด ระหว่างจำนวนประชากรจริงจากข้อมูลในอดีต (Observed Value, Y_o) และจำนวนประชากรที่คาดการณ์ได้ (Calculated Value, Y_c) หรืออีกนัยหนึ่งก็คือการหาเส้นตรงที่ทำให้ $\sum(Y_o-Y_c)^2$ มีค่าน้อยที่สุดนั่นเอง

คุณสมบัติของเส้นตรงกำลังสองน้อยที่สุด (Least Square Straight Line) ได้แก่:

1) ผลรวมของความแตกต่างระหว่าง Observed Value (Y_o) และ Calculated Value (Y_c) เท่ากับศูนย์ (เนื่องจากค่าที่เป็นบวกและค่าที่เป็นลบจะหักลบกันจนหมด):

$$\sum(Y_o-Y_c) = 0$$



แผนภูมิที่ 2 เส้นตรงกำลังสองน้อยที่สุด

2) ผลรวมของความแตกต่างข้างต้นยกกำลังสอง มีค่าน้อยที่สุด เมื่อเทียบกับเส้นตรงอื่น:

$$\sum(Y_o - Y_c)^2 = \text{Minimum}$$

ในการคำนวณหาค่า a และ b ของเส้นตรงกำลังสองน้อยที่สุดสำหรับข้อมูลชุดหนึ่ง สามารถอาศัยสมการที่เรียกว่า "Normal Equations" ได้ดังนี้:

$$\sum Y = na + b\sum X \quad \text{.........................} \ [2]$$

และ $$\sum XY = a\sum X + b\sum X^2 \quad \text{.........................} \ [3]$$

โดยที่ n เป็นจำนวนข้อมูล (Observations)

จากข้อมูลประชากรในอดีตดังตัวอย่างในตารางที่ 1 เราสามารถหาค่า $\sum Y$, $\sum XY$, n, $\sum X$ และ $\sum X^2$ ได้ จากนั้นจึงแทนค่าเหล่านี้ลงในสมการ [2] และ [3] เพื่อหาค่า a และ b ดังนี้:
แทนค่า $\sum Y$, n และ $\sum X$ ในสมการ [1] จะได้:

$$96.53 = 11a + 55b \quad \text{.........................} \ (1)$$

แทนค่า $\sum XY$, $\sum X$ และ $\sum X^2$ ในสมการ [2] จะได้:

$$459.44 = 55a + 385b \quad \text{.........................} \ (2)$$

(1) x 5: $$482.65 = 55a + 275b \quad \text{.........................} \ (3)$$

(2) - (3): $$-23.21 = 110b$$

$$b = 110/(-23.21) = -0.211$$

ตารางที่ 1 การคำนวณหาค่าต่าง ๆ สำหรับ Normal Equations

ปี ค.ศ.	ประชากร (*1,000 คน) Y_o	X	X^2	XY
1962	10.15	0	0	0
1963	9.49	1	1	9.49
1964	9.41	2	4	18.82
1965	9.54	3	9	28.62
1966	8.74	4	16	34.96
1967	8.00	5	25	40.00
1968	8.44	6	36	50.64
1969	8.81	7	49	61.67
1970	8.29	8	64	66.32
1971	7.68	9	81	69.12
1972	7.98	10	100	79.80
N=11	$\sum Y = 96.53$	$\sum X = 55$	$\sum X^2 = 385$	$\sum XY = 459.44$

แทนค่า b ใน (1) จะได้:

$$
\begin{aligned}
96.53 &= 11a + 55\,(-0.211) \\
96.53 &= 11a - 11.605 \\
108.135 &= 11a \\
a &= 108.135/11 \\
&= 9.8304
\end{aligned}
$$

ดังนั้น สมการทั่วไปที่ใช้ในการคาดการณ์ประชากรจากข้อมูลในตารางที่ 1 จึงสามารถเขียนได้เป็น:

$$Y_c = 9.8304 - 0.211X$$

เพื่อให้การคำนวณง่ายขึ้น เราอาจกำหนดให้ค่า X=0 อยู่ตรงกึ่งกลางของช่วงเวลาดังในตารางที่ 2 เพื่อให้ค่า $\sum X$ ในสมการ [2] และ [3] เป็น 0 ซึ่งเมื่อเราแทนค่า $\sum X = 0$ ในสมการ [2] จะได้:

$$\sum Y = na + b(0)$$

$$\sum Y = na$$

$$a = \sum Y/n \qquad \text{.............................[4]}$$

และเมื่อแทนค่า $\sum X = 0$ ในสมการ [3] จะได้:

$$\sum XY = a(0) + b\sum X^2$$

$$\sum XY = b\sum X^2$$

$$b = \sum XY/\sum X^2 \qquad \text{............................... [5]}$$

ตารางที่ 2 การปรับค่า $\sum X$ เพื่อช่วยให้การคำนวณง่ายขึ้น

กรณีข้อมูลเป็นจำนวนเลขคี่

ปี ค.ศ.	X
1962	-5
1963	-4
1964	-3
1965	-2
1966	-1
1967	0
1968	1
1969	2
1970	3
1971	4
1972	5
N = 11	**$\sum X = 0$**

กรณีข้อมูลเป็นจำนวนเลขคู่

ปี ค.ศ.	X
1962	-9
1963	-7
1964	-5
1965	-3
1966	-1
→	0
1967	1
1968	3
1969	5
1970	7
1971	9
N = 11	**$\sum X = 0$**

จากการคำนวณหาค่า a และ b ของข้อมูลประชากรชุดเดิมโดยใช้สมการ [4] และ [5] พบว่า:

$$\mathbf{a = 8.7754}$$

$$\mathbf{b = -0.211}$$

และสมการทั่วไปที่ใช้คาดการณ์ประชากรได้แก่:

$$\mathbf{Y_c = 8.7754 - 0.211X}$$

ซึ่งจะเห็นว่าค่า b หรือการเปลี่ยนแปลงประชากรต่อปียังคงเดิม ขณะที่ค่า a เปลี่ยนไป เนื่องจากการเรียงลำดับค่า X ใหม่ โดยไม่ได้เริ่มจาก 0 แต่เริ่มจาก -5 อย่างไรก็ตาม เมื่อนำเอาสมการทั้งสองไปคาดการณ์ประชากรในอนาคตจะได้ผลลัพธ์เช่นเดียวกันดังตัวอย่างต่อไปนี้:

$\mathbf{Y_c = 9.8304 - 0.211X}$ **ค่า X สำหรับปี 1962 = 0** $\mathbf{\rightarrow Y_{1962} = 9.8304 - 0.211(0)}$ $\mathbf{Y_{1962} = 9.8304}$	$\mathbf{Y_c = 8.7754 - 0.211X}$ **ค่า X สำหรับปี 1962 = -5** $\mathbf{\rightarrow Y_{1962} = 8.7754 - (0.211)(-5)}$ $\mathbf{Y_{1962} = 8.7754 + 1.055}$ $\mathbf{Y_{1962} = 9.8304}$

ในกรณีที่ข้อมูลในอดีตเป็นจำนวนเลขคู่แทนที่จะเป็นจำนวนเลขคี่อย่างในตัวอย่างข้างต้น จะพบว่าค่า X = 0 ตรงกึ่งกลางของช่วงเวลาอยู่ไม่ตรงกับข้อมูลในปีใดเลย ดังนั้นเราจึงจำเป็นจะต้องปรับค่า X เสียใหม่ ดังตัวอย่างในตารางที่ด้าน 2 ขวา โดยเมื่อไม่อาจให้ค่า X ของปีใดปีหนึ่งเป็น 0 ได้ จึงต้องให้ค่า X = 0 อยู่ระหว่างปี ค.ศ. 1966 กับ 1967 แล้วให้ค่า X ของ ค.ศ. 1966 เป็น -1 และค่า X ของ ค.ศ. 1967 เป็น +1 ทำให้มีความแตกต่างของค่า X อยู่ = +1 - (-1) = 2 ดังนั้นค่าดัชนี X จึงจะต้องเพิ่มทีละ 2 ทุกช่วงปี แทนที่จะเพิ่มทีละ 1 เหมือนในกรณีของข้อมูลที่มีจำนวนเป็นเลขคี่

2. แบบจำลองเชิงทวีกำลัง (Exponential Model)

โทมัส มัลทัส (Thomas Malthus) นักปราชญ์ชาวอังกฤษ เป็นผู้ตั้งข้อสังเกตว่าจำนวนประชากรโดยทั่วไปมีแนวโน้มที่จะเพิ่มขึ้นแบบอนุกรมเรขาคณิต (Geometric Growth) กล่าวคือ มีลักษณะเหมือนกับดอกเบี้ยเงินฝากหรือเงินกู้ที่เพิ่มขึ้นในอัตราส่วนหรือร้อยละที่คงที่ แทนที่จะเพิ่มเป็นจำนวนที่คงที่เหมือนกรณีแบบจำลองเชิงเส้นตรง รูปแบบสมการทั่วไปของแบบจำลองเชิงทวีกำลัง ได้แก่:

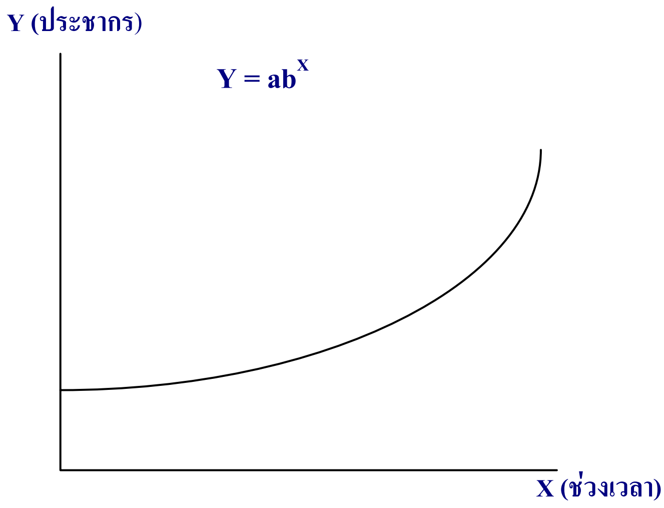
$$Y_c = ab^X \quad ....[6]$$

โดยที่: Y_c = จำนวนประชากรที่คาดการณ์ได้ เมื่อตัวแปรอิสระ = X

X = ช่วงเวลา (Time Index)

a = Y-intercept หรือค่าของ Y_c เมื่อ X = 0

b = 1.0 + อัตราการเติบโต (Growth Rate, r) เมื่อ อัตราการเติบโต (r) = จำนวนประชากรที่เปลี่ยนแปลงไปในช่วงเวลาใดเวลาหนึ่ง หารด้วย จำนวนประชากรในช่วงต้นของเวลา



แผนภูมิที่ 3 ความสัมพันธ์ระหว่างประชากรกับช่วงเวลาในแบบจำลองเชิงทวีกำลัง

การแปลงสมการเชิงทวีกำลังให้อยู่ในรูปแบบของสมการเชิงเส้นตรง

วิธีที่ง่ายที่สุดในการหาเส้นสมการที่ใช้คาดการณ์ประชากรของแบบจำลองเชิงทวีกำลัง ก็คือ การแปลงสมการ [6] ให้อยู่ในรูปของสมการเชิงเส้นตรง ($Y_c = a + bX$) โดยอาศัย Logarithms

หลักเกณฑ์ทั่วไปของ Logarithms คือ:

ถ้า $N = a^X$ ดังนั้น $\text{Log}_a N = X$

หรืออธิบายได้ว่า เมื่อ N เป็นค่าผลลัพธ์ของเลขจำนวนหนึ่ง, a, ยกกำลัง X (โดยที่ a ไม่เท่ากับ 1) เราอาจแสดงค่า Log ของ N ที่มีฐาน a ได้เท่ากับ X

ตัวอย่างเช่น	ถ้า $10 = 10^1$	ดังนั้น	$\text{Log}_{10}10 = 1$
และ	ถ้า $100 = 10^2$	ดังนั้น	$\text{Log}_{10}100 = 2$
และ	ถ้า $1{,}000 = 10^3$	ดังนั้น	$\text{Log}_{10}1{,}000 = 3$
และ	ถ้า $10{,}000 = 10^4$	ดังนั้น	$\text{Log}_{10}10{,}000 = 4$
และ	ถ้า $100{,}000 = 10^5$	ดังนั้น	$\text{Log}_{10}100{,}000 = 5$ ฯลฯ

จะเห็นได้ว่าการแปลงค่าในลักษณะดังกล่าวส่งผลถึงค่าในช่วงที่อยู่สูงขึ้นไปอย่างมีนัยสำคัญกับค่าที่อยู่ในช่วงที่ต่ำกว่า การเปลี่ยนแปลงประชากรจริงจาก 10 มาเป็น 100 หรือเพิ่มขึ้นสุทธิ 90 คน เมื่อแสดงเป็นค่า Log จะเปลี่ยนแปลงไปเพียง 1 คือจาก 1 มาเป็น 2 เท่านั้น ขณะที่การเปลี่ยนแปลงจริงจาก 100 มาเป็น 1,000 หรือเพิ่มขึ้นสุทธิ 900 คน (มากกว่าเดิม 10 เท่า) ก็ยังคงแสดงเป็นค่า Log ได้ต่างกันเท่ากับ 1 คือจาก 2 มาเป็น 3 ในทำนองเดียวกัน การเปลี่ยนแปลงจริงจาก 1,000 มาเป็น 10,000 หรือเพิ่มขึ้น 9,000 คน ก็แสดงเป็นค่า Log ได้ต่างกันเท่ากับ 1 เช่นเดิม คือจาก 3 มาเป็น 4

ตารางที่ 3 เป็นตัวอย่างของการแปลงตัวเลขประชากรจากค่าปกติ (P และ ΔP) มาเป็นค่า Log ($\text{Log}_{10}P$ และ $\Delta\text{Log}_{10}P$) จะเห็นว่าตัวเลขการเปลี่ยนแปลงสุทธิทางซ้ายมือ (ΔP) มีขนาดใหญ่ขึ้นทุกทีตามช่วงเวลา (t) ที่เพิ่มขึ้น โดยมีอัตราส่วนการเติบโตที่คงที่คือ 10 เท่าของจำนวนเดิมทุกปี อันเป็นลักษณะของการเติบโตแบบทวีกำลัง ตัวเลขทางขวามือ ($\Delta\text{Log}_{10}P$) แสดงการเปลี่ยนแปลงค่า $\text{Log}_{10}P$ ซึ่งเพิ่มขึ้นเป็นจำนวนที่คงที่ (คือ 1 ต่อปี) อันเป็นลักษณะของการเติบโตแบบเส้นตรง ดังนั้น การแปลงค่าจริงมาเป็นค่า Log จึงสามารถแปลงรูปแบบของสมการทั่วไปเชิงทวีกำลัง ให้อยู่ในรูปแบบของสมการทั่วไปเชิงเส้นตรงได้

ตารางที่ 3 เปรียบเทียบการเปลี่ยนแปลงประชากรระหว่างค่าปกติและค่า Log

Time Index (t)	Population (P)	ΔP	$Log_{10}P$	$\Delta Log_{10}P$
0	10		1	1
1	100	90	2	1
2	1,000	900	3	1
3	10,000	9,000	4	1
4	100,000	90,000	5	1

ในการแปลงสมการทั่วไปของแบบจำลองเชิงทวีกำลังให้อยู่ในรูปของแบบจำลองเชิงเส้นตรง เราอาศัยกฎของ Logarithms สองข้อคือ:

1) Logarithm ของตัวเลข 2 ตัวคูณกัน (หรือหารกัน) เท่ากับผลบวก (หรือผลต่าง) ของ Logarithm ของตัวเลขแต่ละตัว:

$$\log(ab) = \log a + \log b$$

$$\log(a/b) = \log a - \log b$$

2) Logarithm ของตัวเลขใดยกกำลัง (หรือถอดราก) เท่ากับ Logarithm ของตัวเลขนั้นคูณกับ (หรือหารด้วย) ค่ากำลัง (หรือราก):

$$\log(a^b) = b \log a$$

$$\log(a^{1/b}) = (\log a)/b$$

จากกฎทั้งสองข้อ เราสามารถนำมาใช้ในการแปลงสมการ [6] ให้อยู่ในรูปของสมการเชิงเส้นตรงได้ดังนี้:

$$Y_c = ab^X$$

$$\log Y_c = \log(ab^X)$$

$$\log Y_c = \log a + \log b^X$$

$$\log Y_c = \log a + (\log b)X \quad \text{.....................} \ [7]$$

เราสามารถนำ Least Squares Criterion มาใช้ในการหาเส้นสมการที่เป็นเส้นตรงกำลังสองน้อยที่สุดได้เช่นเดียวกับแบบจำลองเชิงเส้นตรง โดยใช้ค่า log Y แทนที่ Y ในการคำนวณหาค่า log a และ log b ดังนี้:

Arithmatic Least-Squares Straight Line	**Logarithmic Least-Squares Straight Line**
$Y_c = a + bX$ .... [1]	$\log Y_c = \log a + (\log b)X$ .... [7]
โดยที่:	**โดยที่ :**
$a = \sum Y/n$ .... [4]	$\log a = \sum(\log Y)/n$ .... [8]
$b = \sum XY/\sum X^2$ .... [5]	$\log b = \sum(X\log Y)/\sum X^2$ ...... [9]

เมื่อแทนค่า log a และ log b ในสมการ [7] แล้ว เราก็สามารถคำนวณค่า $\log Y_c$ ของแต่ละช่วงเวลาได้โดยการ Antilog ทั้งสองข้างของสมการ ตัวอย่างเช่น เราคำนวณค่า log a ได้เท่ากับ 0.9449 และ log b ได้เท่ากับ -0.0047 เมื่อเราเอาค่าทั้งสองไปแทนในสมการ [7] จะได้:

$$\log Y_c = 0.9449 - 0.0047X$$

ในกรณีที่ X = 5 เราสามารถคำนวณค่า $\log Y_c$ ได้ดังนี้:

$$\log Y_c = 0.9449 - 0.0047(5)$$

$$\log Y_c = 0.9214$$

เมื่อเรา Antilog ทั้งสองข้าง จะได้:

$$\text{Antilog}(\log Y_c) = \text{Antilog}(0.9214)$$

$$Y_c = 10^{0.9214}$$

$$Y_c = 8.35$$