การวัดความสม่ำเสมอหรือความสอดคล้องของการกระจายตัวในพื้นที่

โดย รศ. ดร. วรรณศิลป์ พีรพันธุ์ © 2003

Lorenz Curve และ Concentration Index

Lorenz Curve เป็นวิธีการพล็อตข้อมูล 2 ชุดหรือตัวแปร 2 ตัว ลงบนกราฟเพื่อแสดงความสม่ำเสมอ หรือความสอดคล้องของการกระจายตัวในพื้นที่ (Areal Distribution) ตัวอย่างเช่น ตารางที่ 1 แสดงให้เห็นถึงตัวแปร 2 ตัวคือ X และ Y ที่กระจายตัวอยู่บนพื้นที่ A, B, C, D, E และ F เราสามารถใช้ Lorenz Curve มาช่วยพิจารณาว่าตัวแปรทั้งสองมีการกระจายตัวในพื้นที่ย่อยทั้งหกสอดคล้องกันมากน้อยเพียงใด เพื่อให้ง่ายต่อการอธิบายให้เราสมมุติว่า ตัวแปร X เป็นผู้ลงคะแนนเสียงเลือกตั้งที่เป็นชนชั้นกลาง และตัวแปร Y เป็นผู้ลงคะแนนเสียงที่เลือกพรรคไทยรักไทย (หน่วย : พันคน) และเราต้องการหาความสอดคล้องของการกระจายตัวของตัวแปรทั้งสองในพื้นที่เหล่านี้

ตารางที่ 1 การกระจายตัวของตัวแปร X และ Y ในพื้นที่ย่อย A,B,C,D,E และ F

พื้นที่	X_i	Y_i
A	30	30
B	20	15
C	10	8
D	10	5
E	20	19
F	30	23

ขั้นตอนในการสร้าง Lorenz Curve มีดังนี้

1) คำนวณค่าอัตราส่วน Y_i / X_i สำหรับแต่ละหน่วยพื้นที่ (i = 1,2,3,...,N)
2) จัดลำดับหน่วยพื้นที่เรียงตามค่า Y/X โดยเรียงจากค่าต่ำสุดไปหาสูงสุด
3) แปลงค่า X_i และ Y_i เป็นร้อยละ

4) หาค่าร้อยละสะสม (Cumulative Percentage) ของ X_i และ Y_i ตั้งแต่ลำดับแรกจนถึงลำดับสุดท้าย

5) พล็อตตำแหน่งของร้อยละสะสม C_{xi} และร้อยละสะสม Y (C_{Yi}) บน X-Y Graph แล้วลากเส้นเชื่อมต่อจุดที่พล็อตเพื่อสร้าง Lorenz Curve

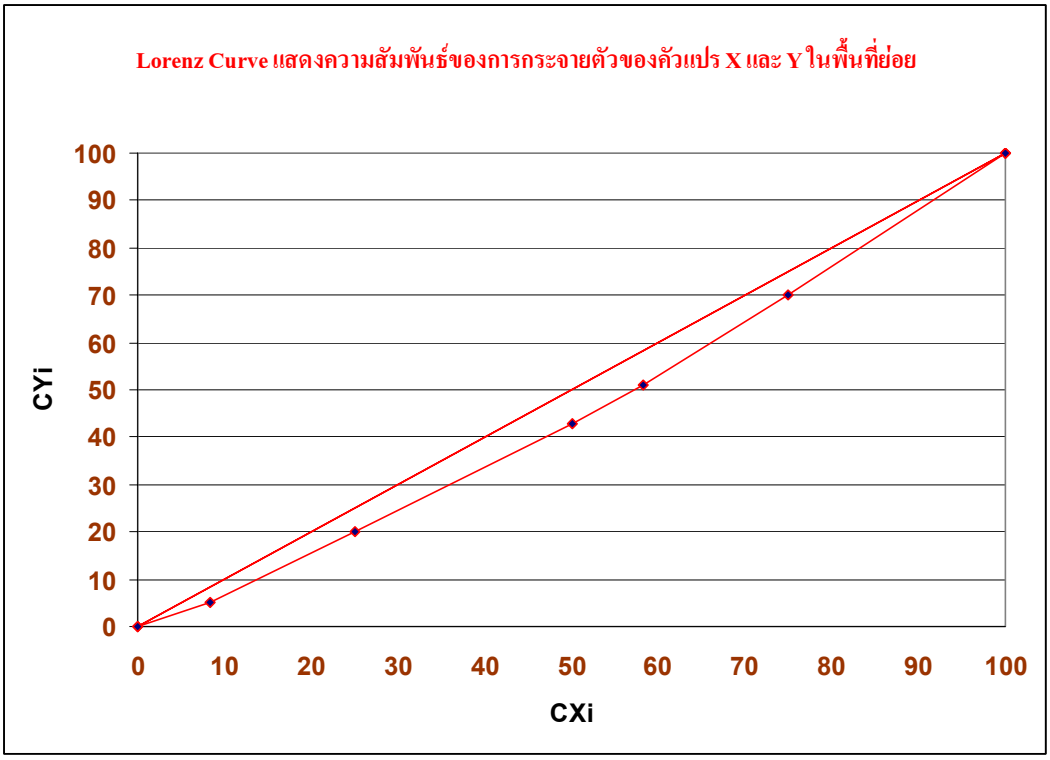
ถ้าการกระจายตัวของตัวแปรทั้งสองบนพื้นที่ย่อยมีความสอดคล้องกันมาก เส้น Lorenz Curve จะเข้าใกล้เส้นทแยงมุม (Diagonal) ถ้าเส้น Lorenz Curve อยู่ห่างจากเส้นทแยงมุมออกไป แสดงว่าการกระจายตัวของตัวแปรทั้งสองบนพื้นที่ย่อยไม่สอดคล้องกัน

เมื่อการกระจายตัวของตัวแปรทั้งสองในพื้นที่มีความสอดคล้องกันมากที่สุด (Maximum Areal Association) เส้น Lorenz Curve จะทับเส้นทแยงมุม (Diagonal) เมื่อการกระจายตัวของตัวแปรทั้งสองในพื้นที่มีความสอดคล้องกันน้อยที่สุด (Minimum Areal Association) เส้น Lorenz Curve จะทับแกน X หรือแกน Y (สัมผัสจุด Co-ordinate X = 100% หรือ Y = 100 %)

ตารางที่ 2 การแปลงคะแนนดิบเป็นร้อยละสะสมเพื่อสร้าง Lorenz Curve และคำนวณค่า Concentration Index

เรียงลำดับพื้นที่		แปลงเป็นร้อยละ		หาค่าร้อยละสะสม	
Y_i/X_i	พื้นที่	P_{Xi}	P_{Yi}	C_{Xi}	C_{Yi}
0.50	D	8.3	5.0	8.3	5.0
0.75	B	16.7	15.0	25.0	20.0
0.77	F	25.0	23.0	50.0	43.0
0.80	C	8.3	8.0	58.3	51.0
0.95	E	16.7	19.0	75.0	70.0
1.00	A	25.0	30.0	100.0	100.0

เอาไปพล็อตกราฟ



วิธีการวัดความสอดคล้องของการกระจายตัวในพื้นที่ (Areal Association) จาก Lorenz Curve ก็คือ การหาค่า Concentration Index (CI) ซึ่งในทางกราฟฟิกหาได้จากการวัดระยะห่างสุดในแนวตั้งระหว่างเส้นทแยงมุม (Diagonal) กับเส้น Lorenz Curve ซึ่งจะมีค่าอยู่ระหว่าง 0 % ถึง 100 % นอกจากนี้ ยังมีวิธีคำนวณค่า CI โดยไม่ต้องสร้าง Lorenz Curve อีก 2 วิธีคือ

1) โดยการคำนวณจากสูตร: $CI = Max (C_{Xi} - C_{Yi})$ .....(1)

โดยใช้ค่าร้อยละสะสมซึ่งใช้ในการสร้าง Lorenz Curve ซึ่งจากตัวอย่างในตารางที่ 2 สามารถคำนวณค่า CI ได้

$$CI = 58.3\ \% - 51.0\ \% \text{ หรือ } 7.3\ \%$$

2) คำนวณจากสูตร:

$$CI = \frac{\sum_{i=1}^{N} \left| P_{Xi} - P_{Yi} \right|}{2} \quad \text{.....(2)}$$

เมื่อ P_{Xi} และ P_{Yi} เป็นค่าร้อยละ (Uncumulative Percentage) ของแต่ละตัวแปร การใช้สูตรนี้ช่วยให้ไม่ต้องคำนวณค่าร้อยละสะสม และสามารถหาค่า CI ได้โดยไม่ต้องสร้าง Lorenz Curve ในบางครั้งการคำนวณหาค่า CI จะใช้สัดส่วน (Proportion) แทนค่าร้อยละ ในกรณีดังกล่าว ค่า CI จะอยู่ระหว่าง 0 และ 1

การประยุกต์ใช้ค่า CI

Concentration Index เป็นดัชนีวัดความสอดคล้องของการกระจายตัวของตัวแปรสองตัวในพื้นที่ ยิ่งการกระจายตัวของตัวแปรทั้งสองในพื้นที่ย่อยมีลักษณะใกล้เคียงกันเท่าใด (หรืออีกนัยหนึ่งก็คือ มีความสอดคล้องกันมากเท่าใด) ค่า CI ก็จะน้อยลงเท่านั้น

คำว่า Concentration Index บางครั้งก็เรียกว่า Index of Dissimilarity (D_A) ซึ่งทั้งสองคำเป็นการเรียกชื่อกลาง ๆ ทั่วไปเมื่อนำไปใช้ประยุกต์ใช้ในแต่ละด้านก็อาจเรียกชื่อต่างกันออกไปอีก เช่น

1) ในกรณีของข้อมูลการจ้างงาน ถ้า Y เป็นจำนวนคนงานในโรงงานที่ใช้เครื่องจักร และ X เป็นจำนวนคนงานในโรงงานผลิตรถยนต์ CI จะถูกเรียกว่า Coefficient of Geographical Association

2) ถ้าเปลี่ยนค่า X ในข้อ 1) เป็นจำนวนประชากร CI จะถูกเรียกว่า Degree of Market Orientation ของอุตสาหกรรม

3) ถ้าเปลี่ยนเป็นจำนวนประชากรกับขนาดพื้นที่ CI ถูกเรียกว่า Population Concentration Index

4) ถ้าให้ X เป็นการใช้ที่ดินรวม และ Y เป็นการใช้ที่ดินประเภทใดประเภทหนึ่ง CI จะถูกเรียกว่า Coefficient of Areal Localization

5) ถ้า X เป็นคนผิวดำ และ Y เป็นคนผิวขาว CI จะถูกเรียกว่า Index of Residential Segregation

ข้อจำกัดของ Lorenz Curve และ Concentration Index

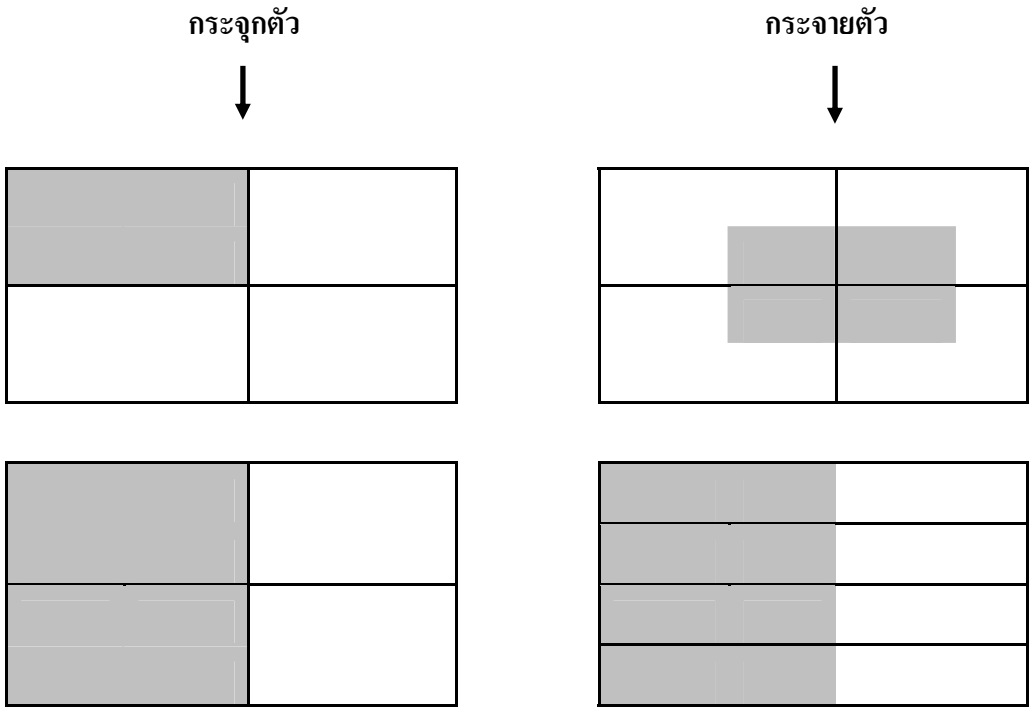
Lorenz Curve และ CI เป็นเครื่องมือที่มีประโยชน์และสามารถประยุกต์ใช้ได้อย่างกว้างขวางในการวิเคราะห์ด้านพื้นที่ อย่างไรก็ตาม ในการประยุกต์ใช้และการแปลความหมาย จะต้องคำนึงถึงข้อจำกัดดังต่อไปนี้:

1) ใช้กับตัวแปรที่อยู่ในรูปของสัดส่วนหรืออัตราส่วนอยู่แล้วไม่ได้ เช่น ใช้หาความสอดคล้องระหว่างความหนาแน่นประชากร (หน่วย: คน/พื้นที่) กับขนาดประชากรไม่ได้

2) ใช้กับค่าติดลบไม่ได้ เช่น การเปลี่ยนแปลงประชากรที่ลดลง ฯลฯ

3) ต้องมีตัวเลขสถิติในระดับหน่วยพื้นที่ย่อย ดังนั้น จึงใช้กับสถิติน้ำฝน อุณหภูมิ ฯลฯ ซึ่งเก็บข้อมูลเฉพาะจุด และเป็นค่ารวมของพื้นที่ทั้งหมดเพียงค่าเดียวไม่ได้

4) วิธีการแบ่งพื้นที่และขนาดของพื้นที่ย่อยที่ต่างกันจะส่งผลให้รูปร่างของ Lorenz Curve และ CI แตกต่างกันออกไป แม้ว่าจะเป็นการวัดเรื่องเดียวกัน ในพื้นที่เดียวกัน (ดังแสดงให้เห็นด้วยกราฟฟิคในแผนภูมิที่ 1) เพราะฉะนั้น การวัดเปรียบเทียบข้ามอนุภาคจึงไม่สามารถทำได้ เช่น จะเปรียบเทียบการกระจายตัวของประชากรในภาคใต้กับภาคกลาง โดยหา CI ของภาคกลางเปรียบเทียบกับ CI ของภาคใต้ไม่ได้



แผนภูมิที่ 1 การแบ่งพื้นที่ย่อยด้วยวิธีการและ/หรือขนาดที่ต่างกัน

อาจทำให้ผลการวิเคราะห์ที่ได้แตกต่างกัน

Gini Coefficient

Gini Coefficient ที่จริงแล้วก็คือค่า Concentration Index นั่นเอง จากสมการ (2) ซึ่งเป็นการคำนวณค่า CI โดยไม่ใช้ร้อยละสะสม มีรูปแบบคือ:

$$CI = \frac{\sum_{i=1}^{N} \left| P_{Xi} - P_{Yi} \right|}{2}$$

เมื่อ P_{Xi} และ P_{Yi} เป็นค่าร้อยละของแต่ละตัวแปร (i = 1, 2, 3,..., N) สูตรที่ใช้ในการคำนวณหาค่า Gini Coefficient ก็มีการคำนวณเช่นเดียวกันคือ:

$$G = 1/2 * \sum_{i=1}^{N} \left| P_{Xi} - P_{Yi} \right| \quad \text{..... (3)}$$

ในกรณีที่ใช้ข้อมูลดิบโดยไม่คำนวณค่าร้อยละ จะสามารถหาค่า Gini Coefficient ได้จากสูตร:

$$G = 1/2 * \sum_{i=1}^{N} \left| 100X_i/X_t - 100Y_i/Y_t \right| \quad \text{.....(4)}$$

เมื่อค่า X_i และ Y_i เป็นค่าคะแนนดิบของตัวแปร X และ Y และ X_t และ Y_t เป็นค่ารวม (Total) ของตัวแปร X และ Y ตามลำดับ ค่า G จะอยู่ระหว่าง 0 ถึง 100

ค่า G = 0 หมายถึง มีความสอดคล้องของการกระจายตัวของตัวแปรทั้งสองในพื้นที่สูงที่สุด (Exact Correspondence or No Concentration or Localization) ขณะที่ค่า G = 100 แสดงถึง มีความแตกต่างของการกระจายตัวของตัวแปรทั้งสองมากที่สุด โดยมีการกระจุกตัวของ ตัวแปรตัวหนึ่งอยู่ในพื้นที่ย่อยที่ไม่มีตัวแปรอีกตัวหนึ่งอยู่เลย (Absolutely Concentration)

Gini Coefficient มีที่ใช้และข้อจำกัดเช่นเดียวกับ Concentration Index (เพราะเป็นตัวเดียวกัน) โดยทั่วไปแล้วมักจะถูกนำไปใช้ในการวัดความสอดคล้องของการกระจายตัวของ

ตัวแปรที่เกี่ยวข้องกับประชาชนกลุ่มต่าง ๆ กัน โดยเฉพาะอย่างยิ่งการแบ่งกลุ่มประชาชนออกตามลักษณะทางเศรษฐกิจและสังคม

ตารางที่ 3 การคำนวณหาค่า Gini Coefficient โดยใช้สมการที่ (4)

พื้นที่	Xi	Yi	100Xi/Xt	100Yi/Yt	|100Xi/Xt - 100Yi /Yt|
A	30	30	25.00	30.00	5.00
B	20	15	16.67	15.00	1.67
C	10	8	8.33	8.00	0.33
D	10	5	8.33	5.00	3.33
E	20	19	16.67	19.00	2.33
F	30	23	25.00	23.00	2.00
รวม	**120**	**100**	**100.00**	**100.00**	**14.67**

G = 1/2 * 14.67 =7.34