



2.1 วิธีการหาการแจกแจงของรายได้อันต่อเนื่อง

รายได้อันต่อเนื่องของกรุงเทพมหานคร ในเขตเทศบาล และนอกเขตเทศบาล มีการแจกแจงด้วยฟังก์ชันการแจกแจงซึ่งทดสอบความเป็นไปได้ด้วย การทดสอบภาวะสำรूपสันติ หรือการทดสอบความเหมาะสมของการแจกแจง (Test of Goodness of Fit) ซึ่งเป็นการทดสอบความแตกต่างระหว่างความถี่ที่สังเกตได้ และความถี่ที่ได้จากการแจกแจงที่คาดหวังไว้ ค่าสถิติที่ใช้ในการทดสอบคือ ไคสแควร์ (χ^2)

$$\chi^2 (v) = \frac{\sum_{i=1}^n (O_i - E_i)^2}{E_i}$$

- เมื่อ n คือ จำนวนชั้นของข้อมูล
- O_i คือ ความถี่ของข้อมูลในชั้นที่ i รวมขอบเขตบนของข้อมูล แต่ไม่รวมขอบเขตล่าง
- E_i คือ ความถี่ที่ได้จากฟังก์ชันการแจกแจงที่คาดหวังในชั้นที่ i มีค่าเท่ากับ $N \times (F(L_{i+1}) - (F(L_i)))$
- N คือ ความถี่ทั้งหมดของค่าสังเกต
- L_{i+1} คือ ขอบเขตบนในชั้นที่ i
- L_i คือ ขอบเขตล่างในชั้นที่ i
- v คือ องศาแห่งความอิสระมีค่าเท่ากับ $n-k-1$
- k คือ จำนวนค่าสถิติที่ใช้ประมาณค่าพารามิเตอร์

ถ้า $f(x)$ เป็นฟังก์ชันความหนาแน่นของรายได้อันต่อเนื่องที่คาดหวังและมีการแจกแจงแบบต่อเนื่อง ฟังก์ชันการแจกแจงคือ

$$F(L_i) = \int_0^{L_i} f(x)dx ; x > 0$$

ลุ่มมติฐานที่ตั้งในการทดสอบ

H_0 : รายได้ครัวเรือนมีการแจกแจงด้วยฟังก์ชันการแจกแจงที่คาดหวัง
หวังไว้

H_a : รายได้ครัวเรือนไม่ได้มีการแจกแจงด้วยฟังก์ชันการแจกแจง
ตามที่คาดหวังไว้

ถ้าความถี่ที่สังเกตได้จากข้อมูลมีค่าใกล้เคียงกับความถี่ที่คาดหวังตามทฤษฎี χ^2

จากการคำนวณจะมีค่าน้อย แต่ถ้าความถี่ต่างกันมากค่าของ χ^2 จากการคำนวณจะมีค่ามากและ
เกณฑ์ที่จะยอมรับสมมติฐาน H_0 คือ χ^2 จากการคำนวณมีค่าน้อยกว่า χ^2_{α} ที่ได้จากตาราง
ที่ระดับนัยสำคัญ α

2.2 วิธีจัดกลุ่มรายได้ครัวเรือน

การจัดกลุ่มรายได้มีหลายวิธีขึ้นอยู่กับลักษณะการนำเอากลุ่มรายได้ที่จัดได้ไปใช้ประโยชน์
ดังนั้นในที่นี้จะพิจารณาตามลักษณะการใช้ประโยชน์ กลุ่มรายได้ที่จัดดังนี้

2.2.1 ถ้านำเอากลุ่มรายได้ที่จัดได้ไปใช้ในการสำรวจตัวอย่าง โดยมุ่งให้หน่วยตัวอย่าง
ในกลุ่มเดียวกันมีความคล้ายคลึงกันมาก ๆ และพิจารณาค่าความแปรปรวนวิธีที่ศึกษามี 2 วิธี คือ

2.2.1.1 วิธีที่ใช้รากที่สองของความถี่สะสม (Cumulative \sqrt{f}) ของ Dalenius
และ Hodges¹ เป็นวิธีแบ่งกลุ่มที่ความแปรปรวนของตัวประมาณค่าเฉลี่ยรายได้ประชากรมีค่าต่ำ
สุด โดยใช้การบวกสะสมของรากที่สองของความถี่ของรายได้ครัวเรือน ($f(y)$) เมื่อ y
คือรายได้ครัวเรือนที่ใช้เป็นตัวแปรในการจัดกลุ่ม ทั้งนี้โดยมีขั้นตอนในการจัดกลุ่มดังนี้

1. หาค่าความถี่ของรายได้ครัวเรือนแต่ละค่า และหารากที่สองของความถี่
($\sqrt{f(y)}$)

2. บวกสะสมค่า $\sqrt{f(y)}$ เป็น cum $\sqrt{f(y)}$

3. หารผลรวมสะสมของค่า $\sqrt{f(y)}$ ด้วย จำนวนกลุ่มรายได้ที่ต้องการ
ซึ่งในที่นี้ต้องการจัดกลุ่มรายได้ เป็น 10 กลุ่ม 5 กลุ่ม และ 3 กลุ่ม ตามลำดับ

¹William G. Cochran, Sampling Techniques, 3d ed. (New York :
John Wiley & Sons, Inc., 1977), pp. 128-130.

4. ผลหารที่ได้จากข้อ 3 เป็นตัวกำหนดช่วงรายได้

2.2.1.2 วิธีที่ใช้ระยะทางกำลังสอง (Squared Distance) ของ Walter

D. Fisher¹ เป็นวิธีจัดกลุ่มที่ให้ผลรวมของความแปรปรวนแต่ละกลุ่มมีค่าน้อยที่สุด การจัดกลุ่มจากสมาชิกทั้งหมด N ตัว ถ้าต้องการจัดกลุ่ม G กลุ่มเมื่อ $G < N$ จะมีวิธีจัดได้ทั้งหมด $\binom{N-1}{G-1}$ วิธีเมื่อจัดกลุ่มได้แล้วจะเลือกวิธีที่ให้ค่าผลรวมของความแปรปรวนแต่ละกลุ่มต่ำสุด โดยสูตรในการหาผลรวมของความแปรปรวนแต่ละกลุ่มเป็นดังนี้

$$\text{ให้ } D = \sum_{i=1}^N w_i (a_i - \bar{a}_i)^2$$

เมื่อ a_i คือ รายได้ครัวเรือนที่ i

\bar{a}_i คือ รายได้เฉลี่ยของกลุ่มที่รายได้ครัวเรือนที่ i อยู่

w_i คือ ค่าถ่วงน้ำหนักของรายได้ครัวเรือนที่ i โดย $w_i = \frac{n_i}{N}$

n_i คือ ความถี่ของระดับรายได้ครัวเรือนที่ i

N คือ ความถี่ทั้งหมด

เมื่อเลือกวิธีให้ค่า D ต่ำสุดแล้วจะได้กลุ่มรายได้ซึ่งกำหนดโดยขอบเขตบนและล่าง ดังนี้ $(a_0 - a_1), (a_1 - a_2), \dots, (a_{g-1} - a_g)$ เมื่อ $a_0 < a_1 < a_2 < \dots < a_{g-1} < a_g$ แล้วปรับขอบเขตกลุ่มที่ได้ใหม่ จากการหาอนุพันธ์ของ D เทียบกับ $a_i, i = 1, 2, \dots, g-1$ เมื่อให้อนุพันธ์ $\frac{\partial D}{\partial a_i}$ มีค่าเท่ากับศูนย์แล้ว จะได้ค่า a_i ซึ่งเป็นขอบเขตที่ปรับใหม่ คือ $a_i = \frac{1}{2} (\bar{a}_i + \bar{a}_{i+1})$ (โดยที่ \bar{a}_i และ \bar{a}_{i+1} เป็นรายได้เฉลี่ยในกลุ่มที่ i และ $i + 1$ ตามลำดับ) และจะให้ D มีค่าน้อยที่สุด เรียก D ว่า Least Squared Partition

แต่เนื่องจากข้อมูลรายได้ครัวเรือนมีมากจนเกินกว่าที่จะหาจำนวนวิธีในการจัดกลุ่มได้ครบทั้งหมด จึงต้องจัดกลุ่มรายได้เป็นกลุ่มย่อย ๆ K กลุ่ม ($K > G$) ก่อนที่จะจัดเข้าเป็นกลุ่มใหญ่ G กลุ่ม ดังนั้นการหาผลรวมของความแปรปรวนแต่ละกลุ่มจำเป็นต้องหาผลรวมของความแปรปรวน

¹Walter D. Fisher, "On Grouping for Maximum Homogeneity,"

ของค่าเฉลี่ยแต่ละกลุ่มแทน และต้องปรับค่าจำกัดความของสัญลักษณ์ต่าง ๆ ใหม่ดังนี้

- a_i คือ รายได้เฉลี่ยของครัวเรือนในกลุ่มย่อยที่ i
 \bar{a}_i คือ รายได้เฉลี่ยของครัวเรือนในกลุ่มใหญ่ที่กลุ่มย่อยที่ i อยู่
 w_i คือ ค่าถ่วงน้ำหนักในกลุ่มย่อยที่ i โดย $w_i = \frac{n_i}{N}$
 n_i คือ ความถี่ของรายได้ในกลุ่มย่อยที่ i
 N คือ ความถี่ทั้งหมด

สำหรับขั้นตอนในการแบ่งกลุ่มมีดังนี้

1. แบ่งกลุ่มรายได้ออกเป็นกลุ่มย่อย ๆ จำนวนพอประมาณในที่นี้ใช้กลุ่มย่อยจำนวน 30 กลุ่มย่อย ซึ่งหาได้ด้วยวิธีที่ใช้รากที่สองของความถี่สะสม เพื่อให้หน่วยตัวอย่างในกลุ่มเดียวกันมีความคล้ายคลึงกัน และสะดวกในการคำนวณหาค่า D

2. เขียนความเป็นไปได้ในการจัดกลุ่มว่า จะมีกลุ่มย่อยจำนวนเท่าใดและกลุ่มใดบ้างอยู่ในกลุ่มที่ต้องการจัด ดูรายละเอียดเพิ่มเติมในภาคผนวก ก.

3. ใช้คอมพิวเตอร์ช่วยคำนวณค่าผลรวมของ $w_i (a_i - \bar{a}_i)^2$ ของกลุ่มย่อยที่จะอยู่ในกลุ่มเดียวกัน ทุกกลุ่มที่เป็นไปได้ โปรแกรมที่ใช้แสดงไว้ในภาคผนวก ข.

4. หาผลรวมของ $w_i (a_i - \bar{a}_i)^2$ ของแต่ละวิธีที่จัดกลุ่มตามจำนวนกลุ่มที่ต้องการ คือ 10 กลุ่ม 5 กลุ่ม และ 3 กลุ่ม แต่ละชุด แล้วเลือกวิธีที่ให้ค่า D ต่ำสุดของแต่ละชุด

2.2.2 ถ้านำเอากรุปรายได้ที่จัดได้นั้นไปใช้ในการประมาณค่าดัชนีชี้วัดความไม่เท่าเทียมกันของรายได้ วิธีที่ใช้ศึกษาคือ วิธีที่ใช้เงื่อนไขค่าเฉลี่ย (The Average Condition) ของ B.B. Aghevli และ F. Mehran¹

2.2.1.7) วิธีที่ใช้เงื่อนไขค่าเฉลี่ย (The Average Condition) ของ B.B. Aghevli และ F. Mehran เป็นวิธีจัดกลุ่มที่ทำให้ผลรวมของความแตกต่างของรายได้แต่ละกลุ่ม (θ) มีค่าน้อยที่สุด และยังให้ขอบเขตกลุ่มที่เหมาะสมด้วย ความแตกต่างของรายได้วัดได้ด้วย Gini's

¹B.B. Aghevli and F. Mehran, "Optimal Grouping of Income Distribution Data," Journal of the American Statistical Association 76 (March, 1981) : 22-26.

absolute pairwise difference ถ้า x, y, \dots แทนรายได้ครัวเรือนของแต่ละครัวเรือนที่ไม่เป็นลบ มีการแจกแจงด้วยฟังก์ชัน F ที่ต่อเนื่อง ซึ่งสามารถหาค่าอนุพันธ์และค่าเฉลี่ยได้ ถ้าต้องการจัดกลุ่มรายได้ g กลุ่ม ซึ่งกำหนดโดยขอบเขตล่างและบนดังนี้ $(a_0 - a_1), (a_1 - a_2), \dots, (a_{g-1} - a_g)$ เมื่อ $a_0 < a_1 < \dots < a_{g-1} < a_g$

ผลรวมของความแตกต่างของรายได้แต่ละกลุ่มคือ

$$\phi = \sum_{i=1}^g \int_{a_{i-1}}^{a_i} \int_{a_{i-1}}^{a_i} |x-y| dF(x) dF(y)$$

ขอบเขตกลุ่มที่จะแบ่งรายได้ออกเป็นกลุ่ม ๆ หนึ่งหาได้ด้วยการหาอนุพันธ์ของ ϕ เทียบกับ $a_i, i = 1, 2, \dots, g-1$ ให้อนุพันธ์ $\frac{\partial \phi}{\partial a_i}$ มีค่าเท่ากับศูนย์จะได้ค่า a_i ที่ทำให้ค่า ϕ น้อยที่สุด คือ

$$a_i = \frac{\int_{a_{i-1}}^{a_{i+1}} x dF(x)}{\int_{a_{i-1}}^{a_{i+1}} dF(x)} \dots \dots \dots (1)$$

หรือ $a_i = E(X | a_{i-1} \leq X < a_{i+1})$

แต่เนื่องจากฟังก์ชันการแจกแจงรายได้อาจหาได้จากผลการวิเคราะห์นั้น ไม่สามารถอินทิเกรตเพื่อหาค่า a_i ที่แม่นยำ (exact) ได้ จึงได้ทำการประมาณค่า a_i ด้วยการหาค่า a_i ซ้ำ ๆ กัน (Iterative procedure) เมื่อเกิดกรณีที่ไม่สามารถหาค่า a_i ที่แท้จริงได้ ผลของการจัดกลุ่มด้วยวิธีนี้อาจจะไม่ให้ขอบเขตกลุ่มที่เหมาะสมที่สุด แต่จะให้ค่าต่ำสุดของ ϕ โดยประมาณ

สำหรับขั้นตอนในการแบ่งกลุ่มมีดังนี้

1. แบ่งกลุ่มรายได้ออกเป็น 10 กลุ่ม 5 กลุ่ม และ 3 กลุ่ม ตามลำดับ โดยแบ่งช่วงรายได้ทั้งหมดออกเป็นช่วงสั้น ๆ ที่มีความกว้างพอสมควร ในที่นี้แบ่งเท่า ๆ กัน ยกเว้นช่วงสุดท้าย
2. ให้ $a_{i(0)}$ เป็นขอบเขตเริ่มต้นของกลุ่มเมื่อ $i=1, 2, \dots, 9; i=1, 2, \dots, 4; i=1, 2$ สำหรับการสัดเป็น 10 กลุ่ม 5 กลุ่ม และ 3 กลุ่มตามลำดับ

3. คำนวณหาค่า $a_i(1)$ จาก $a_i(0)$ ด้วยสูตร (1) โดยใช้เครื่องคอมพิวเตอร์
ประมาณค่าในกรณีอื่นๆ

4. คำนวณ $a_i(2)$ จาก $a_i(1)$ ต่อไปเรื่อยๆ จนกระทั่ง $a_i(s+1)$ ที่คำนวณ
จาก $a_i(s)$ ให้ค่าที่ไม่เปลี่ยนแปลง ในที่ใช้เกณฑ์การเปลี่ยนแปลงในหลักหน่วย

โปรแกรมที่ใช้ประมาณค่าอินทิกรัล และคำนวณค่า a_i ต่อเนื่องกันไปเรื่อยๆ ได้แสดง
ไว้ในภาคผนวก ข.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย