

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เมื่อต้องการศึกษาคุณลักษณะของประชากร วิธีการที่จะทำให้ทราบลักษณะของประชากรที่สนใจศึกษาได้ถูกต้องคือ การเก็บข้อมูลจากทุกหน่วยตัวอย่างในประชากร หรือเรียกว่า การสำมะโน แต่หากประชากรที่ต้องการศึกษามีขนาดใหญ่มากและเป็นไปได้อย่างที่จะสามารถเก็บข้อมูลได้ครบทั้งประชากรเนื่องจากอาจมีข้อจำกัดในด้านเวลาค่าใช้จ่าย หรืออื่นๆ วิธีการที่เหมาะสมที่จะทำให้ได้ข้อมูลที่ใกล้เคียงกับประชากรมากที่สุดก็คือ การเลือกตัวอย่างที่เป็นตัวแทนที่ดีของประชากร ซึ่งการที่จะได้ตัวแทนที่ดีของประชากรนั้นจะต้องอาศัยวิธีการทางสถิติที่เรียกว่า เทคนิคการสุ่มตัวอย่าง (Sampling Techniques) เข้ามาช่วย เมื่อได้ตัวอย่างที่เป็นตัวแทนที่ดีแล้วนั้น การประมาณค่าของคุณลักษณะที่สนใจศึกษาของประชากรจากตัวอย่างก็จะได้ค่าใกล้เคียงความเป็นจริงเมื่อทำการเก็บรวบรวมข้อมูลจากตัวอย่างซึ่งเป็นส่วนหนึ่งของประชากร โดยการสุ่มตัวอย่างแล้ว ตัวประมาณ (estimator) ถือเป็นสิ่งสำคัญที่นักสถิติพยายามที่จะพัฒนาเพื่อจะนำไปใช้ในการวิเคราะห์และสรุปผลไปยังคุณลักษณะของประชากรที่ศึกษา

ในบางกรณี ค่าประชากรที่เราสนใจอาจเป็นค่าอัตราส่วนของตัวแปรสองตัวได้ โดยอาจเกิดขึ้นเนื่องจากลักษณะการจับคู่ข้อมูล หรือลักษณะข้อมูลที่มีอยู่ เช่น หากสนใจค่าใช้จ่ายด้านการบริโภคอาหารเฉลี่ยต่อคนในเขตกรุงเทพมหานคร ถ้าสามารถเก็บข้อมูลได้ว่าคนแต่ละคนมีรายจ่ายการบริโภคอาหารเท่าใด ก็จะหาค่าเฉลี่ยได้ แต่เนื่องจากการบริโภคอาหารในครัวเรือนมักเป็นการบริโภคร่วมกันที่ไม่สามารถแยกแยะออกมาได้ว่าแต่ละคนบริโภคเท่าไร ดังนั้นการเก็บรวบรวมข้อมูล จึงมักพิจารณาครัวเรือนเป็นหน่วยสำหรับเก็บข้อมูล โดยทำการเก็บข้อมูลค่าใช้จ่ายการบริโภคอาหารของครัวเรือนพร้อมกับข้อมูลจำนวนคนในครัวเรือนแต่ละครัวเรือน จากตัวอย่างนี้จะเห็นได้ว่าตัวแปรที่เป็นตัวแปรเป้าหมาย (target variable) คือ ค่าใช้จ่ายด้านการบริโภคอาหารในครัวเรือน และ ตัวแปรที่เป็นตัวแปรช่วย (auxiliary variable) คือ จำนวนคนในครัวเรือนแต่ละครัวเรือนเป้าหมาย

เมื่อข้อมูลที่ได้สามารถเขียนอยู่ในรูปอัตราส่วนประชากร ก็สามารถหาตัวประมาณอัตราส่วนประชากรที่จะนำไปใช้ในการประมาณค่าเฉลี่ยประชากรและค่าผลรวมประชากรต่อไป ซึ่งถ้าตัวประมาณของอัตราส่วนประชากรมีค่าใกล้เคียงกันกับค่าอัตราส่วน

ประชากรย่อมมีผลทำให้การประมาณค่าเฉลี่ยประชากรและค่าผลรวมประชากรใกล้เคียงกับค่าประชากรที่ได้ โดยการประมาณค่าของค่าพารามิเตอร์มี 2 แบบ คือ การประมาณค่าแบบจุด และการประมาณค่าแบบช่วง แต่การประมาณแบบช่วง (interval estimation) เป็น การประมาณโดยใช้ช่วงสุ่มที่คาดว่าจะครอบคลุมค่าพารามิเตอร์ที่เราต้องการประมาณ โดยช่วงสุ่มนั้นมีโอกาสครอบคลุมค่าพารามิเตอร์ด้วยความน่าจะเป็นที่กำหนดไว้ล่วงหน้า ซึ่งดีกว่าการประมาณแบบจุดที่ให้ค่าเพียงค่าเดียว โดยค่าที่ได้อาจเท่ากับค่าพารามิเตอร์หรือไม่เท่ากับค่าพารามิเตอร์ก็ได้

โดยในปี 2548 สุกัญญา อินทรภักดิ์ และ พิชณญ์ เจียวกุณ ได้ศึกษาและเปรียบเทียบวิธีการประมาณค่าช่วงความเชื่อมั่นสำหรับพารามิเตอร์ของข้อมูลชั่วคราว ซึ่งมีการแจกแจงล็อกนอร์มัล และการแจกแจงไวบูลล์ ต่อมาในปี 2551 วราภรณ์ กรงทองได้เปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นสำหรับอัตราส่วนออกดสีในการถดถอยโลจิสติกพหุคูณ และในปี 2550 และปี 2552 วราภรณ์ พานิชกิจโกศลกุล ได้เปรียบเทียบวิธีการประมาณค่าช่วงความเชื่อมั่นของพารามิเตอร์ของการแจกแจงแบบเลขชี้กำลัง เมื่อข้อมูลมีค่าผิดปกติ และได้เปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นของพารามิเตอร์ สำหรับตัวแบบตัวแบบอัตตสหสัมพันธ์อันดับ ที่หนึ่ง (AR1) เมื่อข้อมูลมีค่าผิดปกติและค่าสูญหาย ตามลำดับ การประมาณช่วงความเชื่อมั่นของอัตราส่วนประชากร โดยใช้ตัวประมาณของอัตราส่วนประชากรที่ต่างกัน คือ การใช้อัตราส่วนของค่าเฉลี่ย และการใช้ค่าเฉลี่ยของอัตราส่วน โดยที่ทราบค่าเฉลี่ยประชากรของตัวแปรช่วย แต่ในบางครั้งไม่ทราบค่าเฉลี่ยประชากรของตัวแปรช่วย จึงใช้ค่าเฉลี่ยตัวอย่างของตัวแปรช่วยมาใช้ในการคำนวณช่วงความเชื่อมั่น ดังนั้นการวิจัยครั้งนี้จึงสนใจ ศึกษาการใช้ค่าเฉลี่ยตัวอย่างแทนค่าเฉลี่ยประชากรในการประมาณช่วงความเชื่อมั่นของอัตราส่วนประชากร จากการสุ่มตัวอย่างอย่างง่าย

1.2 วัตถุประสงค์ของการศึกษา

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีสำหรับการประมาณอัตราส่วนประชากร โดยการใช้ (1) อัตราส่วนระหว่างค่าเฉลี่ยตัวอย่างของตัวแปรเป้าหมาย (target variable) และค่าเฉลี่ยตัวอย่างของตัวแปรช่วย (auxiliary variable) และ (2) ค่าเฉลี่ยของอัตราส่วนระหว่าง ค่าของตัวแปรเป้าหมาย (target variable) และ ค่าของตัวแปรช่วย (auxiliary variable) ที่ได้จากตัวอย่าง เมื่อใช้ค่าเฉลี่ยตัวอย่างของตัวแปรช่วย จากการสุ่มตัวอย่างอย่างง่าย คำนวณช่วงความเชื่อมั่นของอัตราส่วนประชากร

1.3 สมมุติฐานของการศึกษา

วิธีสำหรับการประมาณอัตราส่วนประชากร โดยการใช้ (1) อัตราส่วนระหว่างค่าเฉลี่ยตัวอย่างของตัวแปรเป้าหมาย (target variable) และค่าเฉลี่ยตัวอย่างของตัวแปรช่วย (auxiliary variable) และ (2) ค่าเฉลี่ยของอัตราส่วนระหว่าง ค่าของตัวแปรเป้าหมาย (target variable) และ ค่าของตัวแปรช่วย (auxiliary variable) ที่ได้จากตัวอย่าง ด้วยวิธีการสุ่มตัวอย่างอย่างง่ายมา คำนวณช่วงความเชื่อมั่นของอัตราส่วนประชากรมีความแตกต่างกัน

1.4 ขอบเขตการศึกษา

งานวิจัยนี้เป็นการศึกษาโดยมีขอบเขตการศึกษาดังนี้

(1) ประชากร ที่ศึกษา มีการแจกแจง 3 แบบ คือ การแจกแจงปกติสองตัวแปร (bivariate normal distribution), การแจกแจงปัวซองลิทกอนอร์มัลสองตัวแปร (bivariate poisson lognormal distribution) และการแจกแจงโคชีสองตัวแปร (bivariate cauchy distribution)

(2) สัมประสิทธิ์การแปรผันของประชากร (c.v.) ของทั้งตัวแปร X และตัวแปร Y โดยมีค่าเท่ากัน คือ 20% ,50% และ 80%

(3) สัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ) ระหว่างตัวแปร X และตัวแปร Y คือ $\pm 0.5, \pm 0.6, \pm 0.7$ และ ± 0.8

(4) ช่วงความเชื่อมั่น $(1-\alpha)100\%$ คือ 95%

(5) ขนาดตัวอย่างที่ศึกษา (n) คือ 10, 100 และ 200

(6) ในการจำลองจะทำซ้ำ 1,000 ครั้ง ในแต่ละเงื่อนไขต่างๆ ที่กำหนดในการ วิจัยครั้งนี้ โดยใช้โปรแกรม R เวอร์ชัน 2.10.1

1.5 เกณฑ์ที่ใช้ในการเปรียบเทียบ

การเปรียบเทียบวิธีสำหรับการประมาณอัตราส่วนประชากร จะพิจารณาจากค่าความน่าจะเป็นครอบคลุม (coverage probability)

1.6 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับในการวิจัยครั้งนี้ สามารถเลือกตัวประมาณอัตราส่วนที่เหมาะสมกับเงื่อนไขของประชากร เพื่อนำไปใช้ในการประมาณค่าเฉลี่ยประชากรและค่าผลรวมประชากรต่อไป ด้วยการประมาณแบบช่วงเมื่อไม่ทราบค่าเฉลี่ยประชากรของตัวแปรช่วย

1.7 สัญลักษณ์และคำจำกัดความที่ใช้ในงานวิจัย

X แทน ตัวแปรช่วย (auxiliary variable)

Y แทน ตัวแปรที่สนใจศึกษา (target variable)

N แทน ขนาดของประชากร

n แทน ขนาดของตัวอย่าง

μ แทน ค่าเฉลี่ยของประชากร

$$R = \frac{\frac{1}{N} \sum_{i=1}^N Y_i}{\frac{1}{N} \sum_{i=1}^N X_i} = \frac{\mu_y}{\mu_x} \quad \text{แทน อัตราส่วนประชากร}$$

$$r_1 = \hat{R} = \frac{\bar{Y}}{\bar{X}} \quad \text{แทน ตัวประมาณของอัตราส่วนประชากรตัวที่ 1 ในที่นี้}$$

เรียกว่า อัตราส่วนของค่าเฉลี่ย (ratio of mean)

$$r_2 = \hat{R} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{X_i} \quad \text{แทน ตัวประมาณของอัตราส่วนประชากรตัวที่ 2 ในที่นี้}$$

เรียกว่า ค่าเฉลี่ยของอัตราส่วน (mean of ratio)

คำจำกัดความ

ความน่าจะเป็นครอบคลุม (coverage probability) ของช่วงความเชื่อมั่น คือ สัดส่วนของจำนวนครั้งที่ช่วงครอบคลุมค่าจริงที่สนใจ

อัตราส่วนของค่าเฉลี่ย คือ อัตราส่วนระหว่างค่าเฉลี่ยของตัวแปรที่สนใจต่อค่าเฉลี่ยของตัวแปรช่วย ตัวอย่างเช่น $r_1 = \hat{R} = \frac{\bar{Y}}{\bar{X}}$

ค่าเฉลี่ยของอัตราส่วน คือ ค่าเฉลี่ยของอัตราส่วนระหว่างค่าของตัวแปรที่สนใจต่อ

ค่าของตัวแปรช่วย ตัวอย่างเช่น $r_2 = \hat{R} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{X_i}$