

การกำหนดขนาดตัวอย่างเมื่อเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากัน Sample Size Estimation in Unequal Probability Sampling

ประยัด แสงงาม

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์ อ.เมือง

จ.นครปฐม 73000 e-mail: sangngam_p@su.ac.th

บทคัดย่อ

งานวิจัยนี้เสนอการกำหนดขนาดตัวอย่างเมื่อเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันแบบแทนที่ พารามิเตอร์ที่ต้องการประมาณ คือ ค่ารวมประชากร ผลการศึกษาพบว่าขนาดตัวอย่างเป็นฟังก์ชันเพิ่มบน ขนาดประชากร ค่าเฉลี่ยประชากร สัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรที่สนใจศึกษา สัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรช่วย และระดับความเชื่อมั่น โดยขนาดตัวอย่างมีความสัมพันธ์กับสัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรที่สนใจศึกษาและของตัวแปรช่วยในลักษณะกราฟเส้นโค้งหงาย อย่างไรก็ตามขนาดตัวอย่างเป็นฟังก์ชันลดบน ความผิดพลาดที่ยอมรับได้ และสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักระหว่างตัวแปรที่สนใจศึกษากับตัวแปรช่วย โดยขนาดตัวอย่างมีความสัมพันธ์กับสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักในรูปแบบฟังก์ชันเชิงเส้นตรงที่มีความชันเป็นลบ

คำสำคัญ: ขนาดตัวอย่าง, ตัวประมาณที่ไม่เอนเอียง, ความแปรปรวน, การเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากัน

Abstract

This paper derived a sample size in unequal probability sampling with replacement. The parameter to be estimated is the population total. The result showed that the sample size is an increasing function on population size, population mean, weight coefficients of variation of study variable and that of auxiliary variable and confidence level. The quadratic relationship between the sample size and weight coefficient of variation gives a graph to be concave up. However, the sample size is a decreasing function on the acceptance error and the weight coefficient of correlation. The relationship between the sample size and weight coefficient of correlation is the negative linear relationship.

Keywords: sample size, unbiased estimator, variance, unequal probability sampling

1. บทนำ

การกำหนดขนาดตัวอย่างเป็นปัญหาที่สำคัญอย่างหนึ่งในงานวิจัย การกำหนดขนาดตัวอย่างน้อยไปอาจทำให้ความน่าจะเป็นที่จะเกิดความผิดพลาดมีค่าสูง แต่การกำหนดขนาดตัวอย่างมากเกินไปก็อาจทำให้ต้องเสียเวลาและค่าใช้จ่ายในการวิจัยสูง ปัญหาที่เกิดขึ้นในการกำหนดขนาดตัวอย่างสำหรับผู้วิจัยซึ่งใช้สถิติในการวิเคราะห์ข้อมูล คือ ไม่ทราบว่าจะเลือกใช้สูตรในการกำหนดขนาดตัวอย่างใดให้เหมาะสมกับงานวิจัย ผู้วิจัยส่วนใหญ่มักจะกำหนดขนาดตัวอย่างโดยใช้การเปิดตาราง ใช้สูตรที่หาง่าย หรือสูตรที่คำนวณง่าย เป็นต้น ในบางครั้งผู้วิจัยก็อาจกำหนดขนาดตัวอย่างตามความสะดวกในการเก็บรวบรวมข้อมูล โดยที่ผู้วิจัยไม่ได้คำนึงถึงวัตถุประสงค์ของงานวิจัย สถิติที่ใช้วิเคราะห์ข้อมูล และความผิดพลาดที่จะเกิดขึ้นจากการใช้สถิตินั้น การกำหนดขนาดตัวอย่างลักษณะนี้ทำให้ผู้วิจัยอาจไม่สามารถบอกความผิดพลาดที่เกิดขึ้นจากการใช้สถิติได้

ในปี 1960 Harter ได้สร้างตารางการแจกแจงความน่าจะเป็นของตัวสถิติที่มีการแจกแจงปกติมาตรฐานหลายตัวแปรเพื่อใช้คำนวณขนาดตัวอย่างในการวิเคราะห์ความแปรปรวน ในปี 1969 Cohen ได้รวบรวมวิธีการกำหนดขนาดตัวอย่างกรณีที่ผู้วิจัยต้องการควบคุมความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 และกำลังการทดสอบของสถิติ เช่น การทดสอบค่าเฉลี่ยประชากร การทดสอบสัมประสิทธิ์การถดถอย และการวิเคราะห์โลจิสติก เป็นต้น สำหรับการประมาณค่าในการวิเคราะห์การถดถอยมีนักวิจัยหลายคนเสนอวิธีการกำหนดขนาดตัวอย่างเพื่อสร้างสมการถดถอยให้มีความถูกต้องมากที่สุด Sawyer(1982) ได้เสนอสูตรการกำหนดขนาดตัวอย่างโดยทำให้ความผิดพลาดสัมบูรณ์เฉลี่ยของค่าพยากรณ์มีค่าน้อยที่สุด ในปี 2012 Brooks และ Barcikowski ได้เสนอวิธีการกำหนดขนาดตัวอย่างในการวิเคราะห์การถดถอยเพื่อให้เปอร์เซ็นต์ความถูกต้องของการพยากรณ์ค่าของตัวแปรตามมีค่าสูง

สำหรับงานวิจัยเชิงสำรวจ การเลือกตัวอย่างโดยใช้ความน่าจะเป็นมีความสำคัญมาก เพราะถ้าผู้วิจัยเลือกตัวอย่างโดยไม่ใช้ความน่าจะเป็นจะไม่สามารถบอกความผิดพลาดที่เกิดจากการเลือกตัวอย่างได้ Cochran ปี 1977 ได้ให้สูตรในการกำหนดขนาดตัวอย่างภายใต้ทฤษฎีการเลือกตัวอย่าง กรณีเลือกตัวอย่างสุ่มแบบง่าย เลือกตัวอย่างแบบแบ่งชั้นภูมิ และเลือกตัวอย่างแบบกลุ่มด้วยความน่าจะเป็นเท่ากัน

Desu และ Daghavarao (1990) ได้รวบรวมวิธีการกำหนดขนาดตัวอย่างทั้งกรณีที่ต้องการประมาณค่าพารามิเตอร์และต้องการทดสอบสมมติฐาน กล่าวโดยสรุปการกำหนดขนาดตัวอย่างอาจแยกตามวัตถุประสงค์ของงานวิจัย คือ เพื่อการประมาณค่าพารามิเตอร์ การพยากรณ์ และการทดสอบสมมติฐาน ในการประมาณค่าพารามิเตอร์อาจจะแยกเป็น 2 แนวคิด คือ แนวคิดการเลือกตัวอย่างภายใต้ประชากรจำกัด และแนวคิดการเลือกตัวอย่างจากประชากรที่มีการแจกแจงความน่าจะเป็น งานวิจัยนี้สนใจศึกษาการกำหนดขนาดตัวอย่างภายใต้แนวคิดการเลือกตัวอย่างจากประชากรจำกัดและวัตถุประสงค์เพื่อประมาณค่าพารามิเตอร์

วิธีการเลือกตัวอย่างโดยใช้ความน่าจะเป็นที่เป็นพื้นฐานสำหรับวิธีการเลือกตัวอย่างแบบอื่น คือ การเลือกตัวอย่างสุ่มแบบง่าย การเลือกตัวอย่างแบบนี้จะมีประสิทธิภาพสูงเมื่อความแปรปรวนของประชากรมีค่าต่ำ แต่ในทางปฏิบัติประชากรมักมีความแปรปรวนสูง ผู้วิจัยอาจเพิ่มประสิทธิภาพของแผนการเลือกตัวอย่างได้โดยใช้การเลือกตัวอย่างที่ใช้ความน่าจะเป็นไม่เท่ากัน การเลือกตัวอย่างที่ใช้ความน่าจะเป็นไม่เท่ากันนี้มักใช้ในการสำรวจด้านประชากรศาสตร์ และด้านสาธารณสุขศาสตร์ โดยหน่วยตัวอย่างจะถูกรวมกันเป็นกลุ่มที่มีขนาดของกลุ่มแตกต่างกัน การสร้างตัวประมาณและการหาความผิดพลาดของค่าประมาณเมื่อใช้แผนการเลือกตัวอย่างโดยใช้ความน่าจะเป็นสามารถศึกษาได้ ในทฤษฎีการเลือกตัวอย่างของ Cochran (1977) การเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันถึงแม้จะมีการใช้อย่างแพร่หลาย แต่ความแปรปรวนของตัวประมาณที่สร้างขึ้นจะมีรูปแบบซับซ้อนและวิธีการสุ่มหน่วยตัวอย่างยังมีหลากหลายวิธี การสร้างสูตรในการกำหนดขนาดตัวอย่างเมื่อเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันจึงยังไม่มีที่แน่นอน นักวิจัยส่วนใหญ่จึงมักใช้สูตรการกำหนดขนาดตัวอย่างจากการเลือกตัวอย่างสุ่มแบบง่าย ทั้งที่

วิธีการวิจัยใช้แผนการเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากัน ซึ่งจะส่งผลให้ความถูกต้องของการประมาณไม่สอดคล้องกับค่าที่นักวิจัยกำหนด

งานวิจัยนี้จะสร้างสูตรในการกำหนดขนาดตัวอย่างเมื่อเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันแบบแทนที่ พารามิเตอร์ที่สนใจประมาณ คือ ค่ารวมประชากร อย่างไรก็ตามพารามิเตอร์ที่ต้องการประมาณในทฤษฎีการเลือกตัวอย่างอาจมีได้หลายตัว แม้ว่าการวิจัยนี้สนใจพารามิเตอร์เพียงตัวเดียวเท่านั้น แต่ก็สามารถนำหลักการไปประยุกต์ใช้เป็นแนวทางในการกำหนดขนาดตัวอย่าง เมื่อสนใจประมาณค่าพารามิเตอร์ตัวอื่นได้

2. การเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันแบบแทนที่

ให้ $U = \{1, 2, \dots, N\}$ แทน ประชากรจำกัดที่ประกอบขึ้นจากหน่วยตัวอย่างที่ $i = 1, 2, \dots, N$ และ กำหนดให้ y แทน ตัวแปรที่สนใจศึกษา ซึ่งค่าของตัวแปรที่สนใจศึกษาที่สอดคล้องกับประชากรแทนด้วย $y_1, y_2, \dots, y_N$ พารามิเตอร์ที่ต้องการประมาณในที่นี้ คือ ค่ารวมประชากร แทนด้วย $Y = \sum_{i=1}^N y_i$ การเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันต้องกำหนดความน่าจะเป็นเริ่มต้นให้กับหน่วยตัวอย่างทุกหน่วย กำหนดให้ p_i แทน ความน่าจะเป็นเริ่มต้นที่หน่วยที่ i จะถูกเลือกครั้งแรก เมื่อ $0 < p_i < 1$, $p_i \neq p_j$ บางค่า $i \neq j = 1, 2, \dots, N$ และ $\sum_{i=1}^N p_i = 1$ เรียก p_i ว่า ความน่าจะเป็นเริ่มต้น (Initial probability) การกำหนดค่า p_i ปกติจะใช้ตัวแปรช่วย (x) ที่มีความสัมพันธ์ทางบวกกับตัวแปรที่สนใจศึกษา โดยต้องทราบค่าตัวแปรช่วย x_i ของหน่วยตัวอย่าง i ทุกหน่วยในประชากร บางครั้งเรียกตัวแปรช่วยนี้ว่า ตัววัดขนาด (Size) กำหนดความน่าจะเป็นเริ่มต้นได้ดังนี้ $p_i = x_i / \sum_{i=1}^N x_i$ การเลือกตัวอย่างเมื่อใช้ตัววัดขนาดเป็นตัวแปรสร้างความน่าจะเป็นเริ่มต้นบางครั้งเรียกว่า การเลือกตัวอย่างแบบใช้ความน่าจะเป็นได้สัดส่วนกับขนาด (Sampling with Probability Proportional to Size) ตัววัดขนาดอาจเป็นฟังก์ชันของตัวแปรหลายตัวก็ได้ เช่น $x = f(x_1, x_2, \dots, x_k)$ เป็นฟังก์ชันของตัวแปรช่วยจำนวน k ตัว และค่าของฟังก์ชันต้องมีค่าเป็นจำนวนบวก การเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันสามารถทำได้ทั้งแบบแทนที่และไม่แทนที่ แต่ในงานวิจัยนี้จะศึกษาเฉพาะกรณีการเลือกตัวอย่างแบบแทนที่

กระบวนการเลือกตัวอย่างจะทำการสุ่มหน่วยตัวอย่างทีละหน่วยโดยกำหนดความน่าจะเป็นที่แต่ละหน่วยจะถูกสุ่มเท่ากับความน่าจะเป็นเริ่มต้น แล้วแทนที่หน่วยที่สุ่มได้คืนก่อนที่จะสุ่มตัวอย่างครั้งต่อไป ทำการสุ่มตัวอย่างจนกระทั่งได้ตัวอย่างขนาด n เรียกวิธีการสุ่มนี้ว่า การเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันแบบแทนที่ (Sampling with Unequal Probability and with Replacement) การดำเนินการสุ่มตัวอย่างด้วยความน่าจะเป็นไม่เท่ากันในแต่ละครั้ง อาจทำได้ 2 วิธี คือ วิธีสะสมขนาด และวิธีของลาฮารี สามารถศึกษาได้ใน ประชุม (2552)

เมื่อเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันแบบแทนที่ กำหนดให้ r_i แทน จำนวนครั้งที่หน่วย i ถูกเลือกเป็นตัวอย่าง และ $s = \{(1, r_1), (2, r_2), \dots, (N, r_N)\}$ แทน ตัวอย่างที่ไม่พิจารณาลำดับของหน่วยที่ถูกเลือก โดยที่ (i, r_i) แทน คู่อันดับของหน่วยตัวอย่าง i กับ จำนวนครั้งที่หน่วย i ถูกเลือกเป็นตัวอย่าง วิธีการสุ่มหน่วยตัวอย่างข้างต้นสอดคล้องกับปัญหาการสุ่มลูกบอลจำนวน n ลูก จากกล่องที่มีลูกบอลทั้งหมดจำนวน N ลูก ความน่าจะเป็นที่ลูกบอล i จะถูกสุ่มเท่ากับ p_i และการสุ่มแต่ละครั้งเป็นอิสระกัน ดังนั้น $r = (r_1, r_2, \dots, r_N)$ มีการแจกแจงพหุนามความน่าจะเป็นที่จะได้ r และได้ตัวอย่าง s มีค่าเท่ากัน คือ

$$P(s) = \frac{n!}{r_1! r_2! \dots r_N!} \prod_{i=1}^N p_i^{r_i}$$

เมื่อ $\sum_{i=1}^N r_i = n$

การนิยามตัวอย่างที่ไม่พิจารณาลำดับข้างต้นทำให้ง่ายต่อการสร้างแผนการเลือกตัวอย่าง แต่ในการเขียนตัวประมาณในรูปสัญลักษณ์จะกำหนดให้ $s^* = (1, 2, \dots, n)$ แทน ตัวอย่างที่พิจารณาลำดับ ภายใต้การเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันแบบแทนที่ ให้ y_i เป็นค่าสังเกตของตัวแปรที่สนใจศึกษาในตัวอย่าง s^* สอดคล้องกับความน่าจะเป็นเริ่มต้น p_i ที่หน่วยจะถูกเลือก เมื่อ $i = 1, 2, \dots, n$ ในปี ค.ศ. 1943 แชนเซนและเฮอริวิทซ์ ได้เสนอ ตัวประมาณที่ไม่เอนเอียงของค่ารวมประชากร ดังนี้

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{p_i} \quad (2.1)$$

ความแปรปรวนของ $\hat{Y}$ เท่ากับ

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i}{p_i} - \hat{Y} \right)^2 p_i \quad (2.2)$$

ตัวประมาณที่ไม่เอนเอียงของ $V(\hat{Y})$ คือ

$$\hat{V}(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{y_i}{p_i} - \hat{Y} \right)^2 \quad (2.3)$$

ถ้า p_i แปรผันตรงกับค่า y_i จะได้ว่าความน่าจะเป็นเริ่มต้นในการเลือกมีค่าเท่ากับ $p_i = y_i/Y$ ในกรณีที่ y_i มีค่ามากกว่า 0 จะทำให้สามารถสร้างความน่าจะเป็นเริ่มต้นนี้ได้ ดังนั้นในกรณีนี้ตัวประมาณค่ารวมประชากรจะให้ค่าความแปรปรวนเท่ากับ 0 กล่าวคือ ค่าประมาณมีค่าเท่ากับค่าพารามิเตอร์สำหรับทุกตัวอย่างที่เป็นไปได้ ดังนั้นการสร้างความน่าจะเป็นเริ่มต้นที่เหมาะสมที่สุด คือ ใช้ค่าสังเกตของตัวแปรที่สนใจศึกษา อย่างไรก็ตามในทางปฏิบัติถ้าทราบค่าสังเกตของตัวแปรที่สนใจศึกษาจากทุกหน่วยในประชากรก็ไม่จำเป็นต้องสุ่มตัวอย่าง จึงไม่เหมาะสมที่จะใช้ค่าสังเกตนี้ในการสร้างความน่าจะเป็นเริ่มต้น โดยทั่วไปการกำหนดความน่าจะเป็นเริ่มต้นจะใช้ตัวแปรช่วยที่ได้สัดส่วนใกล้เคียงกับตัวแปรที่สนใจศึกษา

3. การกำหนดขนาดตัวอย่าง

การกำหนดขนาดตัวอย่างรูปแบบหนึ่งอาจใช้ค่าใช้จ่ายเป็นเกณฑ์ โดยพิจารณาค่าใช้จ่ายในการวิจัยประกอบด้วย ส่วนใดบ้าง แล้วนำมาเขียนเป็นฟังก์ชันค่าใช้จ่ายและขนาดตัวอย่าง เช่น ฟังก์ชันค่าใช้จ่ายอย่างง่ายในการสำรวจตัวอย่าง

$$C = c_0 + cn$$

เมื่อ C แทน ค่าใช้จ่ายทั้งหมดในการสำรวจ

c_0 แทน ค่าใช้จ่ายในการจัดการ เช่น ค่าเอกสาร ค่าฝึกอบรมพนักงานภาคสนาม

c แทน ค่าใช้จ่ายต่อการเก็บหน่วยตัวอย่าง 1 หน่วย

n แทน ขนาดตัวอย่าง

ดังนั้นจะได้ขนาดตัวอย่างที่ต้องเก็บรวบรวมข้อมูล คือ

$$n = \frac{C - c_0}{c}$$

การกำหนดขนาดตัวอย่างในกรณีนี้จะไม่คำนึงถึงคุณภาพของตัวประมาณ ขนาดตัวอย่างที่ใช้จะขึ้นอยู่กับงบประมาณที่มีเท่านั้น ในทางปฏิบัติจึงไม่นิยมใช้

งานวิจัยนี้จะกำหนดขนาดตัวอย่างโดยพิจารณาความผิดพลาดที่ยอมรับได้ (d) ระดับความเชื่อมั่น $(1-\alpha)$ พารามิเตอร์ที่ต้องการประมาณ คือ ค่ารวมประชากร (Y) ตัวประมาณที่ใช้ คือ ตัวประมาณที่ไม่เอนเอียง $\hat{Y}$ แสดงดัง (2.1) และ กำหนดขนาดตัวอย่างภายใต้เงื่อนไข $P(|\hat{Y} - Y| \geq d) = \alpha$ ภายใต้ข้อสมมติว่าตัวประมาณ $\hat{Y}$ เข้าสู่การ

แจกแจงปกติที่ค่าเฉลี่ยเท่ากับ $E(\hat{Y}) = Y$ และความแปรปรวนเท่ากับ (2.2) แต่ความแปรปรวนที่สร้างในรูปแบบนี้จะยากในการอธิบาย การตีความหมายและการหาค่าจากงานวิจัยที่เกี่ยวข้องเพื่อใช้ในการประมาณขนาดตัวอย่าง งานวิจัยนี้จะพิจารณาความแปรปรวนในรูปอื่น ดังนี้

$$\begin{aligned} V(\hat{Y}) &= \frac{1}{n} \sum_{i=1}^N \left(\frac{y_i}{p_i} - Y \right)^2 p_i \\ &= \frac{1}{n} \sum_{i=1}^N \frac{1}{p_i} \left[(y_i - \bar{Y}) - (p_i Y - \bar{Y}) \right]^2 \\ &= \frac{1}{n} \sum_{i=1}^N \frac{1}{p_i} \left[(y_i - \bar{Y})^2 - 2N\bar{Y}(y_i - \bar{Y})(p_i - \bar{p}) + N^2\bar{Y}^2(p_i - \bar{p})^2 \right] \text{ เมื่อ } \bar{p} = \frac{1}{N} \sum_{i=1}^N p_i = \frac{1}{N} \\ &= \frac{N^2}{n} \left[\frac{1}{N} \sum_{i=1}^N \frac{\bar{X}}{x_i} (y_i - \bar{Y})^2 + \frac{R^2}{N} \sum_{i=1}^N \frac{\bar{X}}{x_i} (x_i - \bar{X})^2 - 2 \frac{R}{N} \sum_{i=1}^N \frac{\bar{X}}{x_i} (y_i - \bar{Y})(x_i - \bar{X}) \right] \text{ เมื่อ } R = \frac{\bar{Y}}{\bar{X}} \\ &= \frac{N^2}{n} (\sigma_y^{*2} + R^2 \sigma_x^{*2} - 2R\rho^* \sigma_y^* \sigma_x^*) \\ &= \frac{N^2 \bar{Y}^2}{n} (C_y^{*2} + C_x^{*2} - 2\rho^* C_y^* C_x^*) \end{aligned} \quad (3.1)$$

$$\begin{aligned} \text{เมื่อ } \sigma_y^{*2} &= \frac{1}{N} \sum_{i=1}^N \frac{\bar{X}}{x_i} (y_i - \bar{Y})^2 & \sigma_x^{*2} &= \frac{1}{N} \sum_{i=1}^N \frac{\bar{X}}{x_i} (x_i - \bar{X})^2 & \sigma_{xy}^* &= \frac{1}{N} \sum_{i=1}^N \frac{\bar{X}}{x_i} (y_i - \bar{Y})(x_i - \bar{X}) & C_x^* &= \frac{\sigma_x^*}{\bar{X}} \\ C_y^* &= \frac{\sigma_y^*}{\bar{Y}} \text{ และ } \rho^* &= \frac{\sigma_{xy}^*}{\sigma_x^* \sigma_y^*} \end{aligned}$$

เมื่อเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันแบบแทนที่ ตัวประมาณ $\hat{Y}$ เป็นตัวประมาณที่ไม่เอนเอียงของ Y และความแปรปรวนของ $\hat{Y}$ มีค่าเท่ากับ (3.1) โดยที่ $-\infty < Y < \infty$ และ $0 < V(\hat{Y}) < \infty$ เนื่องจาก $\frac{y_1}{p_1}, \frac{y_2}{p_2}, \dots, \frac{y_n}{p_n}$ เป็นตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระกัน ดังนั้น จากทฤษฎีบทขีดจำกัดส่วนกลาง

(Central limit theorem) ตัวแปรสุ่ม $\frac{\hat{Y} - Y}{\sqrt{V(\hat{Y})}}$ จะเข้าสู่เชิงการแจกแจงสู่ตัวแปรสุ่ม Z ซึ่งมีการแจกแจงปกติมาตรฐาน

จากเงื่อนไขการกำหนดขนาดตัวอย่าง จะได้ว่า $P\left(|Z| \geq \frac{d}{\sqrt{V(\hat{Y})}}\right) = \alpha$ เมื่อ $Z \sim N(0,1)$ ดังนั้น $\frac{d}{\sqrt{V(\hat{Y})}} = z_{\frac{\alpha}{2}}$ เมื่อ

$z_{\frac{\alpha}{2}}$ คือ ค่าของตัวสถิติ Z ที่ความน่าจะเป็นรวมด้านขวาเท่ากับ $\frac{\alpha}{2}$

เมื่อแก้สมการหาขนาดตัวอย่าง จะได้ขนาดตัวอย่างเมื่อเลือกตัวอย่างโดยใช้ความน่าจะเป็นไม่เท่ากันแบบแทนที่ เท่ากับ

$$n = \frac{z_{\frac{\alpha}{2}}^2 N^2 \bar{Y}^2 (C_y^{*2} + C_x^{*2} - 2\rho^* C_y^* C_x^*)}{d^2} \quad (3.2)$$

จากขนาดตัวอย่างข้างต้น พบว่า ขนาดตัวอย่างเป็นฟังก์ชันเพิ่มบน ขนาดประชากร (N) ค่าเฉลี่ยประชากร ($\bar{Y}$) สัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรที่สนใจศึกษา (C_y^*) สัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรช่วย (C_x^*) และระดับความเชื่อมั่น อย่างไรก็ตามขนาดตัวอย่างเป็นฟังก์ชันลดบน ความผิดพลาดที่ยอมรับได้ และสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก (ρ^*)

การกำหนดขนาดตัวอย่างข้างต้นอาจมีปัญหาในการกำหนดความผิดพลาดที่ยอมรับได้ d เพราะค่านี้จะขึ้นอยู่กับค่าพารามิเตอร์ที่ต้องการประมาณ การกำหนดขนาดตัวอย่างอาจพิจารณาความผิดพลาดที่ยอมรับได้สัมพัทธ์ (r) และกำหนดเงื่อนไข $P\left(\left|\frac{\hat{Y}-Y}{Y}\right| \geq r\right) = \alpha$ ภายใต้ข้อสมมติเดียวกันกับการกำหนดขนาดตัวอย่างที่พิจารณาความผิดพลาดที่ยอมรับได้ก่อนหน้า จะได้ขนาดตัวอย่างเท่ากับ

$$n = \frac{z_{\frac{\alpha}{2}}^2 (C_y^{*2} + C_x^{*2} - 2\rho^* C_y^* C_x^*)}{r^2} \quad (3.3)$$

ขนาดตัวอย่าง (3.3) จะไม่ขึ้นกับค่าพารามิเตอร์ที่ต้องการประมาณ และขนาดตัวอย่างเป็นฟังก์ชันลดบนความผิดพลาดที่ยอมรับได้สัมพัทธ์ กล่าวคือ เมื่อความผิดพลาดที่ยอมรับได้สัมพัทธ์มีค่าเพิ่มขึ้นขนาดตัวอย่างจะมีแนวโน้มลดลง

4. การจำลอง

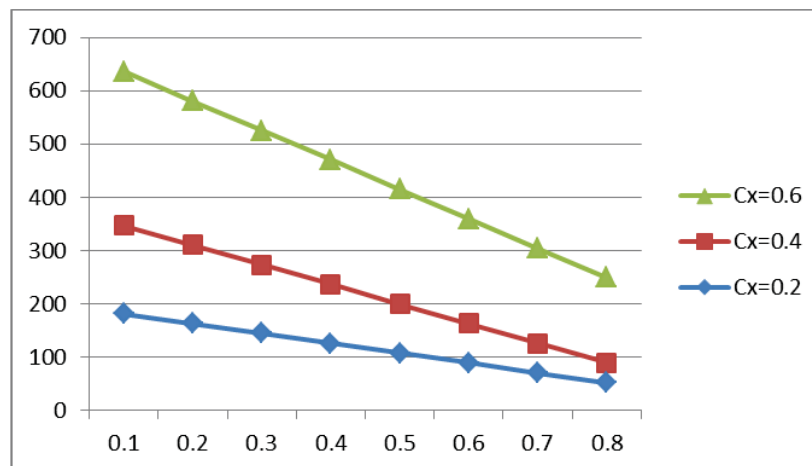
จากสูตรการกำหนดขนาดตัวอย่างข้างต้น หัวข้อนี้จะทำการศึกษาขนาดตัวอย่างเมื่อพารามิเตอร์บางตัวมีค่าเปลี่ยนไป สมมติว่าประชากรขนาด $N=50,000$ ที่มีค่าเฉลี่ยประชากรเท่ากับ $\bar{Y}=5,000$ กำหนดสัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรที่สนใจศึกษาและของตัวแปรช่วยมีค่าเท่ากับ 0.2 0.3 0.4 0.5 และ 0.6 กำหนดสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักระหว่างตัวแปรที่สนใจศึกษาและตัวแปรช่วยมีค่าเท่ากับ 0.1 ถึง 0.8 โดยให้ค่าเพิ่มขึ้นทีละ 0.1 ในที่นี้จะกำหนดความผิดพลาดที่ยอมรับได้สัมพัทธ์เท่ากับ 5% และกำหนดระดับความเชื่อมั่นเท่ากับ 95% ผลการศึกษาแสดงดังตาราง 4.1 และรูป 4.1 - 4.2

ตาราง 4.1 ขนาดตัวอย่างเมื่อกำหนดความผิดพลาดที่ยอมรับได้สัมพัทธ์ 5% และระดับความเชื่อมั่น 95%

C_y^*	C_x^*	ρ^*							
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.2	0.2	111	98	86	74	61	49	37	25
	0.3	181	163	144	126	108	89	71	52
	0.4	283	258	234	209	184	160	135	111
	0.5	415	384	353	323	292	261	230	200
	0.6	578	541	504	467	430	393	357	320
0.3	0.2	181	163	144	126	108	89	71	52
	0.3	249	221	194	166	138	111	83	55
	0.4	347	310	274	237	200	163	126	89
	0.5	476	430	384	338	292	246	200	154
	0.6	636	581	526	470	415	360	304	249
0.4	0.2	283	258	234	209	184	160	135	111
	0.3	347	310	274	237	200	163	126	89
	0.4	443	393	344	295	246	197	148	98
	0.5	569	507	446	384	323	261	200	138
	0.6	725	652	578	504	430	357	283	209
0.5	0.2	415	384	353	323	292	261	230	200
	0.3	476	430	384	338	292	246	200	154
	0.4	569	507	446	384	323	261	200	138

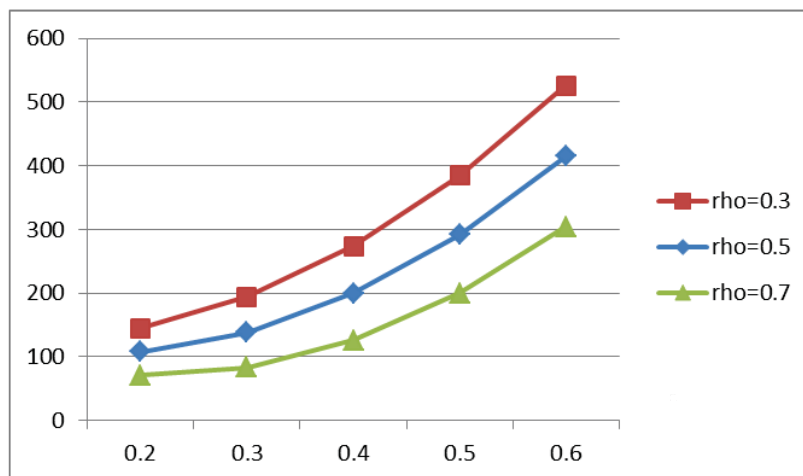
	0.5	691	615	538	461	384	307	230	154
	0.6	845	753	661	569	476	384	292	200
0.6	0.2	578	541	504	467	430	393	357	320
	0.3	636	581	526	470	415	360	304	249
	0.4	725	652	578	504	430	357	283	209
	0.5	845	753	661	569	476	384	292	200
	0.6	996	885	774	664	553	443	332	221

จากตาราง 4.1 พบว่า ขนาดตัวอย่างจะมีค่าอยู่ระหว่าง 25 ถึง 996 ถ้ากำหนดให้สัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรที่สนใจศึกษาและสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักคงตัว ขนาดตัวอย่างจะมีแนวโน้มเพิ่มขึ้นเมื่อสัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรช่วยเพิ่มขึ้น ถ้ากำหนดให้สัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรที่สนใจศึกษาและของตัวแปรช่วยมีค่าคงตัว ขนาดตัวอย่างจะมีแนวโน้มลดลงเมื่อสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักเพิ่มขึ้น



รูป 1 ขนาดตัวอย่างเมื่อ $C_x^* = 0.2$ 0.4 และ 0.6 และกำหนด $C_y^* = 0.3$

จากรูป 1 พบว่า เมื่อสัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรช่วยมีค่าสูงขนาดตัวอย่างจะมีค่าสูง เมื่อสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักเพิ่มขึ้นขนาดตัวอย่างมีแนวโน้มลดลงเชิงเส้นตรง เมื่อสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักเพิ่มขึ้น ความแตกต่างของขนาดตัวอย่างที่สัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรช่วยต่างกันมีแนวโน้มลดลง



รูป 2 ขนาดตัวอย่างเมื่อ $\rho^* = 0.3$ 0.5 และ 0.7 และกำหนด $C_y^* = 0.3$

จากรูป 2 พบว่า เมื่อสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักมีค่าน้อยขนาดตัวอย่างจะมีค่าสูง เมื่อสัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรช่วยเพิ่มขึ้นขนาดตัวอย่างมีแนวโน้มเพิ่มขึ้นในลักษณะเส้นโค้งหงาย เมื่อสัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรช่วยเพิ่มขึ้น ความแตกต่างของขนาดตัวอย่างระหว่างสัมประสิทธิ์สหสัมพันธ์ที่ต่างกันมีแนวโน้มเพิ่มขึ้น

5. สรุปและอภิปรายผล

เมื่อเลือกตัวอย่างด้วยความน่าจะเป็นไม่เท่ากันแบบแทนที่ ความแปรปรวนของตัวประมาณที่ไม่เอนเอียงของค่ารวมประชากรมีรูปแบบที่ซับซ้อนต่อการอธิบายและตีความหมาย งานวิจัยนี้สร้างรูปแบบความแปรปรวนในรูปที่ง่ายขึ้น และนำรูปแบบนี้ไปใช้ในการสร้างขนาดตัวอย่าง ผลจากการกำหนดขนาดตัวอย่างที่ได้พบว่า ขนาดตัวอย่างมีความสัมพันธ์กับสัมประสิทธิ์ความผันแปรแบบถ่วงน้ำหนักของตัวแปรที่สนใจศึกษาและของตัวแบบช่วยในรูปกราฟลักษณะเส้นโค้งหงาย ขนาดตัวอย่างมีความสัมพันธ์กับสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักระหว่างตัวแปรช่วยและตัวแปรที่สนใจศึกษาในรูปแบบเส้นตรงทิศทางตรงข้าม การกำหนดขนาดตัวอย่างโดยกำหนดความผิดพลาดที่ยอมรับได้มักเกิดปัญหาที่ไม่ทราบว่าควรกำหนดค่าความผิดพลาดเท่ากับค่าใด เพื่อให้ง่ายต่อการประยุกต์ใช้ งานวิจัยนี้ได้สร้างขนาดตัวอย่างเมื่อกำหนดความผิดพลาดที่ยอมรับได้สัมพัทธ์

แม้ว่าพารามิเตอร์ที่ต้องการประมาณในงานวิจัยนี้ คือ ค่ารวมประชากร การประยุกต์ใช้กับพารามิเตอร์ตัวอื่นก็สามารถสร้างได้ในทำนองเดียวกัน อย่างไรก็ตามการอธิบายขนาดตัวอย่างจากสัมประสิทธิ์แบบถ่วงน้ำหนักอาจจะนำผลลัพธ์ที่ได้ไปเปรียบเทียบกับกำหนัดขนาดตัวอย่างเมื่อเลือกตัวอย่างด้วยวิธีอื่นได้ยาก งานวิจัยต่อไปควรมีการศึกษาเปรียบเทียบสัมประสิทธิ์แบบถ่วงน้ำหนักและไม่ถ่วงน้ำหนักเมื่อลักษณะเส้นโค้งการแจกแจงของประชากรมีรูปแบบสมมาตรและไม่สมมาตร

6. กิตติกรรมประกาศ

ขอขอบพระคุณกรรมการประเมินบทความทั้ง 2 ท่าน ที่ช่วยให้ข้อเสนอแนะในการปรับปรุง แก้ไขบทความให้มีความถูกต้องมากยิ่งขึ้น และขอบคุณคณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร ที่สนับสนุนการทำวิจัยนี้

7. เอกสารอ้างอิง

- Brooks, P. B. and Barcikowski, R. S. (2012). The PEAR Method for Sample Sizes in Multiple Linear Regression. **Multiple Linear Regression Viewpoints**. 38(2):1-16.
- Cochran, W.G. (1977). **Sampling Techniques**. Newyork: Jonh Wiley and Sons.
- Cohen, J. (1969). **Statistical power analysis for the behavioural sciences**. New York: Academic Press.
- Desu, M. M. and Raghavarao, D. (1990). **Sample size methodology**. New York. Academic Press Inc.
- Hansen MM and WN Hurwitz. 1943. On the theory of sampling from finite populations. **Annals of Mathematical Statistics**. 14:333-362.
- Harter, H.L. (1960). Tables of range and studentized range. **Annals of Mathematical Statistics**. 31: 1122-1147.
- Sawyer, R. (1982). Sample size and the accuracy of prediction made from multiple regression equations. **Journal of Educational Statistics**. 7(2) : 91-104.
- ประชุม สุวดี (2552). การสำรวจด้วยตัวอย่าง: การชักตัวอย่างและการวิเคราะห์. กรุงเทพมหานคร: โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์.