การเปรียบเทียบค่าประมาณที่ปรับน้ำหนักจากการไม่ได้รับคำตอบด้วยวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรและวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง

A Comparison of Population Estimates Using Nonresponse Unit Weighting Adjustment Methods: Raking and Poststratification

ดวงพร ทัชชะวณิช
รองศาสตราจารย์ประจำสาขาวิชาสถิติประยุกต์
คณะวิทยาศาสตร์ มหาวิทยาลัยหอการค้าไทย
E-mail: doungporn__hat@utcc.ac.th

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบค่าประมาณที่ปรับค่าน้ำหนักจากการไม่ได้รับคำตอบด้วยวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร และวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง จากการเปรียบเทียบค่าความเอนเอียงในการประมาณค่าเฉลี่ยของประชากร พบว่า วิธีทั้งสองให้ค่าประมาณที่มีความเอนเอียงไม่แตกต่างกันและจากการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเฉลี่ยของประชากร พบว่า ในกรณีที่ประชากรมีการกระจายน้อยวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง และวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร ให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเฉลี่ยของประชากรน้อยกว่ากรณีที่ไม่ได้มีการปรับค่าน้ำหนัก วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร จะให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานน้อยกว่าวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง เมื่อประชากรมีการกระจายมากขึ้นสองวิธีนี้จะให้ค่า

ประมาณของค่าเฉลี่ยของประชากรที่มีความคลาดเคลื่อนมาตรฐานแตกต่างกันน้อยลง นอกจากนี้ยังพบว่าเมื่อใช้ขนาดตัวอย่างมากขึ้นสองวิธีนี้จะให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าที่ใกล้เคียงกับกรณีที่ไม่ได้ปรับค่าน้ำหนักมากขึ้น

คำสำคัญ: วิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร การไม่ได้รับคำตอบ

Abstract

The objective of this study is to compare the precision of the parameter estimates of mean using two weighting adjustment methods for unit nonresponse, namely Raking and Postratification methods. The study shows that both methods produced similar biased estimates of the population mean. The standard errors of the estimates indicate that where there was less variation among population data, Poststratification and Raking methods provided better results than using only the base weighting adjustment method, and Raking gave smaller standard errors than Poststratification. In addition, when there was a greater variation in population data, both methods gave less difference in standard errors. Moreover, when a larger sample size was used, the estimates obtained from the two methods resembled standard errors compared to the results gained from using baseline weighting.

Keywords: Poststratification, Raking, Nonresponse

บทนำ

ในงานวิจัยส่วนมากมักประสบกับปัญหาในเรื่องการไม่ได้รับการตอบจากหน่วยตัวอย่างบางหน่วยซึ่งอาจจะเนื่องมาจากสาเหตุหลายประการจึงทำให้หน่วยตัวอย่างที่ตอบมีจำนวนน้อยกว่าที่ได้วางแผนเอาไว้ในตอนแรก หากข้อมูลจากผู้ที่ตอบและผู้ที่ไม่ตอบแตกต่างกันจะทำให้ค่าประมาณของค่าเฉลี่ยที่ได้มีความเอนเอียง จากการศึกษาของ Bethlehem and Schouten (2004) ได้ใช้ตัวแปรช่วย (auxiliary variable) ในการหาค่าน้ำหนักและได้เสนอวิธีเลือกตัวแบบถ่วงน้ำหนักที่ดีที่สุด (best weighting model) โดยไม่ได้ให้ความสนใจในเรื่องความแปรปรวนของค่าประมาณเพราะเชื่อว่ามีความสำคัญน้อยกว่าความเอนเอียง โดยเฉพาะเมื่อตัวอย่างขนาดใหญ่ แต่อย่างไรก็ตาม ความแปรปรวนเป็นตัวบ่งชี้ถึงคุณภาพของตัวประมาณ ดังนั้น เพื่อลดความเอนเอียงของค่าประมาณผู้วิจัยจึงอาจจะใช้วิธีปรับค่าน้ำหนักซึ่งเป็นการชดเชยให้กับหน่วยตัวอย่างบางหน่วยที่ไม่ตอบ ในการปรับค่าน้ำหนักสามารถทำได้ 2 วิธี คือ การใช้ข้อมูลจากตัวอย่างและการใช้ข้อมูลจากประชากร สำหรับในการศึกษาครั้งนี้จะใช้วิธีปรับค่าน้ำหนักโดยใช้ข้อมูลจากประชากร ซึ่งเป็นวิธีที่สามารถลดผลกระทบเนื่องจาก

ข้อมูลตัวอย่างไม่ครอบคลุมประชากรได้ โดยดำเนินการ 2 ขั้นตอน คือ ขั้นที่ 1 เป็นการปรับค่าน้ำหนักของตัวประมาณเนื่องจากการไม่ได้รับคำตอบจากหน่วยตัวอย่างและขั้นที่ 2 เป็นการปรับค่าน้ำหนักของตัวประมาณเพื่อให้ได้ค่าผลรวมเข้าใกล้ค่าควบคุม (control totals) ที่ถูกกำหนดจากค่าของประชากร ซึ่งสามารถทำได้ 2 วิธี คือ วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร (Raking) และวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง (Poststratification) ซึ่งเป็นที่คาดหวังว่าวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร อาจจะให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานน้อยกว่าวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง เพื่อหาข้อสรุปในการเลือกใช้วิธีที่มีประสิทธิภาพมากที่สุดจึงกำหนดวัตถุประสงค์ในการวิจัย ดังนี้

1) เพื่อเปรียบเทียบค่าความเอนเอียงของการประมาณค่าเฉลี่ยของประชากรที่ปรับค่าน้ำหนักจากการไม่ได้รับคำตอบโดยใช้วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร และวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง กับวิธีที่ไม่ได้ปรับค่าน้ำหนัก

2) เพื่อเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานของค่าประมาณของค่าเฉลี่ยของประชากรที่ปรับค่าน้ำหนักจากการไม่ได้รับคำตอบโดยใช้วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรและวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างกับวิธีที่ไม่ได้ปรับค่าน้ำหนัก

ทฤษฎีที่เกี่ยวข้อง

วิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง ในทางทฤษฎีถ้าตัวอย่างได้จากแผนการเลือกตัวอย่างโดยใช้ความน่าจะเป็นและไม่มีหน่วยตัวอย่างที่ไม่ตอบค่าผลรวมของน้ำหนักตัวอย่างจะเท่ากับค่าผลรวมของประชากร แต่ในทางปฏิบัติมักจะไม่เป็นเช่นนั้น เนื่องจากมีหลายปัจจัยที่เกี่ยวข้องโดยปัจจัยสำคัญที่มีส่วนทำให้ค่าน้ำหนักมีการเบี่ยงเบนไปมี 2 ปัจจัย คือ มีบางหน่วยตัวอย่างไม่ตอบและหน่วยตัวอย่างไม่ครอบคลุมลักษณะของประชากรซึ่งอาจจะเป็นเพราะกรอบตัวอย่างไม่สมบูรณ์การนำค่าน้ำหนักมาใช้ในการประมาณค่าจะสามารถลดความเอนเอียงของการประมาณค่าเฉลี่ยของประชากร เนื่องจากมีบางหน่วยตัวอย่างที่ไม่ตอบและเนื่องจากหน่วยตัวอย่างไม่ครอบคลุมประชากร ซึ่งผลรวมของค่าน้ำหนักหลังจากที่ได้ปรับแล้วมีค่าเท่ากับผลรวมของประชากร หากประชากร N หน่วยถูกแบ่งออกเป็น H ชั้นภูมิซึ่งในแต่ละชั้นภูมิมีขนาด $N_1, N_2, \ldots, N_H$ และให้ Y_{hi} เป็นค่าตัวแปรที่ต้องการศึกษาของหน่วยที่ i จากชั้นภูมิ h โดย h = 1, 2, ...,H และ i = 1, 2, ...,N_h จากตัวอย่างสุ่มที่ได้มีหน่วยตัวอย่างที่ตอบ n หน่วยโดยมีหน่วยตัวอย่างที่ตอบในแต่ละชั้นภูมิคือ $n_1, n_2, \ldots, n_H$ ค่าเฉลี่ยและค่าความแปรปรวนของตัวอย่างในชั้นภูมิ h คือ $\bar{y}_h = \frac{\sum y_h}{n_h}$ และ $S^2_h = \frac{\sum (y_{hi} - \bar{y}_h)^2}{(n_h - 1)}$ ตามลำดับ จะได้ตัวประมาณค่าเฉลี่ยของประชากร $(\overline{Y}_{PS})_{est} = \sum_h \frac{N_h \bar{y}_h}{N}$ ซึ่งในการหาค่าเฉลี่ยของแต่ละชั้นภูมิได้มีการถ่วงน้ำหนักด้วยขนาดสัมพัทธ์ของแต่ละชั้นภูมิ หากตัวอย่างที่ได้ในแต่ละชั้นภูมิไม่ได้เป็นสัดส่วนกับขนาดของประชากร ค่าประมาณที่ได้จากวิธีนี้ก็จะถูกปรับค่าโดยอัตโนมัติ ถึงกระนั้นวิธีนี้ก็ยังมีข้อเสียเนื่องจากไม่สามารถกำหนดขนาดตัวอย่างได้ล่วงหน้าจึงทำให้ในบางชั้นภูมิอาจจะไม่มีข้อมูล ในการประมาณค่าความแปรปรวนของตัวประมาณค่าเฉลี่ยประชากรมี 2 วิธี ดังนี้

1) ค่าประมาณความแปรปรวนของค่าเฉลี่ยประชากรแบบมีเงื่อนไข

$$V((\bar{Y}_{PS})est / n) = \sum_h \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right)\left(\frac{s_h^2}{n_h}\right)$$

2) ค่าประมาณความแปรปรวนของค่าเฉลี่ยประชากรแบบไม่มีเงื่อนไข

$$V((\overline{Y}_{PS})_{est}) = \sum_h \left(\frac{N_h}{N}\right)^2 s_h^2 \left(E\left(n_h^{-1}\right) - N_h^{-1}\right)$$

$$\approx (n^{-1} - N^{-1}) \sum_h \left(\frac{N_h}{N}\right) s_h^2 + n^{-2} \sum_h \left(1 - \frac{n_h}{N_h}\right) s_h^2$$

เมื่อต้องการประมาณค่าความแปรปรวนก่อนเลือกตัวอย่างจะใช้ค่าประมาณความแปรปรวนแบบมีเงื่อนไข แต่ในกรณีที่ต้องการประมาณค่าความแปรปรวนหลังจากที่ได้เลือกตัวอย่างแล้วก็ควรใช้ค่าประมาณความน่าจะเป็นแบบไม่มีเงื่อนไขซึ่งการใช้วิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างจำเป็นต้องทราบค่า $\frac{N_h}{N}$ และตัวแปรที่ใช้ในการจำแนกหน่วยตัวอย่างให้อยู่ในชั้นภูมิเดียวกัน โดยสิ่งสำคัญที่ผู้วิจัยจะต้องระมัดระวัง ก็คือ ต้องใช้หลักเกณฑ์เดียวกันในการจัดหน่วยตัวอย่างให้อยู่ในแต่ละชั้นภูมิ นอกจากนี้ ผู้วิจัยจำเป็นต้องทราบการแจกแจงของประชากร การหาค่าน้ำหนักโดยใช้ข้อมูลจากประชากรมีการดำเนินการ แบ่งเป็น 2 ขั้นตอน คือ ขั้นที่ 1 กำหนดตัวแปรที่ใช้ควบคุมของประชากรที่ต้องการศึกษา และขั้นที่ 2 คำนวณค่าน้ำหนักซึ่งจะเป็นการปรับค่าจากผลรวมของตัวอย่างให้มีค่าเท่ากับค่าควบคุมซึ่งจะทำให้ค่าความคลาดเคลื่อนมาตรฐานของค่าประมาณลดลงได้ ยกเว้นในกรณีที่ตัวแปรที่ใช้ควบคุมไม่มีความสัมพันธ์กับค่าตัวแปรที่ได้จากการสำรวจ หรือในกรณีที่บางกลุ่มมีค่าน้ำหนักมากเกินไปแต่มีจำนวนผู้ตอบน้อย ถ้าให้ m_h แทนจำนวนผู้ตอบในแต่ละกลุ่ม ตัวประมาณจากวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างหาจาก $\tilde{Y}_{ps} = \sum_1^H \frac{N_h}{m_h} \tilde{Y}_h$ โดยมีข้อสมมติที่สำคัญ 2 ข้อ คือ 1) การตอบของแต่ละหน่วยตัวอย่างในกลุ่มย่อยของประชากรเป็นอิสระกันและมีการแจกแจงเบอร์นูลลี ซึ่งมีความน่าจะเป็นที่แต่ละหน่วยจะตอบคือ $\Phi_h > 0$ 2) การตอบในแต่ละกลุ่มย่อยเป็นอิสระกัน (Choudhry and Valliant, 2002)

ตารางที่ 1 สัญลักษณ์ที่ใช้แทนค่าของข้อมูลที่ได้จากประชากรและตัวอย่าง

	ประชากร						ตัวอย่าง				
	1	**2**	**...**	**K**	**ผลรวม**		**1**	**2**	**...**	**K**	**ผลรวม**
1	W_{11}	W_{12}	...	W_{1K}	$W_{1.}$	1	q_{11}	q_{12}	...	q_{1K}	$q_{1.}$
2	W_{21}	W_{22}	...	W_{2K}	$W_{2.}$	2	q_{21}	q_{22}	...	q_{2K}	$q_{2.}$
.	.	.		.	.	.	.	.		.	.
.	.	.		.	.	.	.	.		.	.
.	.	.		.	.	.	.	.		.	.
H	W_{H1}	W_{H2}	...	W_{HK}	$W_{H.}$	H	q_{H1}	q_{H2}	...	q_{HK}	$q_{H.}$
ผลรวม	$W_{.1}$	$W_{.2}$	...	$W_{.K}$	$W_{..}$	ผลรวม	$q_{.1}$	$q_{.2}$	...	$q_{.K}$	$q_{..}$

วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร เป็นวิธีที่พัฒนาขึ้นโดย Deming and Stephan (1940) ได้ถูกนำมาใช้ในการหาตัวปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรแต่ไม่ทราบจำนวนข้อมูลในบางกลุ่มของตารางแจกแจงของ 2 ตัวแปรหรือจำนวนตัวอย่างในบางกลุ่มอาจจะมีจำนวนน้อยมากซึ่งจะส่งผลต่อประสิทธิภาพของค่าประมาณ การปรับค่าน้ำหนักด้วยวิธีนี้จะให้ค่าสถิติที่มีความต้องกัน (consistency) หากประชากรถูกแบ่งออกเป็นกลุ่มย่อย L กลุ่ม โดย L = NK ให้ m_{hk} แทนจำนวนหน่วยตัวอย่างที่ตอบ โดย h = 1, 2, …,H และ k = 1, 2, …,K ในการใช้วิธีนี้มีข้อสมมติที่จำเป็นต้องพิจารณา 3 ข้อ ดังนี้

1) ภายในกลุ่มย่อยแต่ละกลุ่มของประชากร การที่หน่วยตัวอย่างจะตอบหรือไม่ตอบเป็นอิสระต่อกันและมีการแจกแจงแบบเบอร์นูลลี โดยมีความน่าจะเป็นที่จะตอบ $\Phi_{hk} > 0$

2) การตอบหรือไม่ตอบในแต่ละกลุ่มย่อยของประชากรเป็นอิสระกัน

3) ถ้า $\{\Phi_{hk}\}$ เป็นเซตของความน่าจะเป็นที่หน่วยตัวอย่างในประชากรย่อยแต่ละกลุ่มย่อยจะตอบ ซึ่ง $\text{In}\{\Phi_{hk}/(1-\Phi_{hk})\} = \alpha_h + \alpha_k$ นั่นแสดงว่าค่าลอการิทึมของค่าความน่าจะเป็นที่จะตอบขึ้นกับค่าคงที่ 2 ค่า คือ $\{\alpha_h\}$ และ $\{\alpha_k\}$ หรืออาจจะกล่าวได้ว่าความน่าจะเป็นที่แต่ละหน่วยตัวอย่างจะตอบขึ้นกับผลรวมของแต่ละแถวหรือผลรวมของแต่ละคอลัมน์เท่านั้นแต่ไม่ขึ้นกับจำนวนหน่วยในกลุ่มย่อย

ดังนั้น ข้อสมมติสองข้อแรกจึงเหมือนกับการปรับค่าน้ำหนักด้วยวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง Kalton (1981) ได้อธิบายการใช้วิธีแทนค่าซ้ำโดยเริ่มต้นจากการให้ค่าน้ำหนัก $\frac{W_h}{q_h}$ แต่ละกลุ่มย่อยในแถว h (h = 1, …, H) ซึ่งทำให้ผลรวมของน้ำหนักในกลุ่มย่อยแถวที่ h $\frac{W_h}{q_h}\sum_h q_{hi}$ มีค่าเท่ากับ W_h เนื่องจากการปรับค่าน้ำหนักและผลรวมของคอลัมน์ของตัวอย่างจึงทำให้ $\sum_h q_{hk}\frac{W_h}{q_h} = \sum_h q'_{hk}$ ในรอบที่สองของการแทนค่าซ้ำ หน่วยตัวอย่างที่อยู่ในแต่ละกลุ่มย่อยในคอลัมน์ที่ k (k=1,…,K) ถูกปรับค่าน้ำหนักโดยใช้ $W_k q'_{k.}$ ผลรวมของน้ำหนักในคอลัมน์ k มีค่าเท่ากับ $W_{.k}$ ทำให้ q'_k มีค่าเปลี่ยนเป็น $\sum_h q'_{hk}\frac{W_h}{q'_h} = \sum_h q''_{hk} = q''_{k.}$ จากนั้นได้มีการแทนค่าซ้ำเช่นเดียวกับขั้นที่ 1 และจะหยุดเมื่อค่าน้ำหนักที่ปรับได้ลู่เข้าสู่ค่าที่ค่อนข้างจะคงที่ เมื่อนำค่าน้ำหนักที่ได้จากวิธีนี้ไปประมาณค่าจะทำให้ค่าประมาณที่ได้เข้าใกล้การแจกแจงปกติเมื่อใช้แผนการเลือกตัวอย่างแบบสุ่มอย่างง่าย (simple random sampling) ตัวประมาณค่าจากวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร $\tilde{Y}_r = \sum\sum \tilde{N}_{hk}\bar{y}^*_{hk}$ ซึ่ง $\tilde{N}_{hk} = \tilde{a}_h\tilde{b}_k m_{hk}$ โดย $\{\tilde{a}_h\}$ และ $\{\tilde{b}_k\}$ จะลู่เข้าสู่ค่าที่ใช้ปรับซึ่งได้จากการแทนค่าซ้ำ เนื่องจากวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร มีการควบคุมการแจกแจงของน้ำหนักของตัวอย่าง

จากการศึกษาของ Cervantes and Brick (2008) พบว่า การประมาณค่าด้วยการปรับค่าน้ำหนักโดยวิธี ปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร ได้ถูกใช้ในการเทียบเคียงกับการให้ค่าน้ำหนักเมื่อทราบค่าควบคุมซึ่งทำให้ค่าความคลาดเคลื่อนมาตรฐานของค่าประมาณลดลง เนื่องจากการใช้ตัวแปรช่วยซึ่งมีความสัมพันธ์กับการตอบในปัจจุบันวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรจึงถูกนำไปใช้เพื่อลดค่าความเอนเอียงที่เกิดจากการไม่ตอบ

Battaglia (2008) ได้ใช้วิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง ในงานสำรวจเรื่อง Behavioral Risk Factor Survillance System (BRFSS) โดยใช้ตัวแปรอายุ เพศ และสีผิว จากแหล่งข้อมูลที่เหมาะสมมีความเชื่อถือได้ เช่น ใช้ผลการสำมะโนประชากรปี 2000 การสำมะโนประชากรในปีปัจจุบัน เป็นต้น เป็นค่าควบคุมจากการใช้วิธี Logistic Regression และซอฟต์แวร์ Interaction Detection ทำให้สามารถกำหนดปัจจัยหลักซึ่งเป็นตัวแปรทางด้านเศรษฐกิจและสังคม และผลกระทบร่วมของตัวแปรเกี่ยวกับปัจจัยเสี่ยงทางสุขภาพสองตัว เนื่องจากข้อมูลที่ใช้ในการศึกษาเป็นงานสำรวจขนาดใหญ่งานหนึ่งในประเทศสหรัฐอเมริกาในทุก 10 ปี จึงทำการสำรวจเพียง 1 ครั้ง ในการศึกษาได้หาค่าควบคุมที่ต้องกันกับค่าประมาณของอายุ เพศ และสีผิว ซึ่งได้จากแหล่งข้อมูลทางธุรกิจและการแจกแจงของตัวแปรทางประชากรซึ่งได้จากงานสำรวจในปี 2008 ได้ใช้วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร หาน้ำหนักโดยใช้ปัจจัยหลักและผลรวมของผลกระทบร่วมของสองปัจจัย นอกจากนี้ ยังได้แสดงการให้ค่าน้ำหนักซึ่งจะทำให้ค่าความเอนเอียงเนื่องจากการไม่ตอบมีค่าลดลง

วิธีดำเนินการวิจัย

ในการศึกษาได้ใช้ข้อมูลที่ได้จากโครงการสำรวจภาวการณ์ครองชีพของข้าราชการพลเรือนสามัญ (ในสังกัด ก.พ.) พ.ศ. 2549 โดยจำลองสถานการณ์ของการไม่ตอบโดยใช้วิธีสุ่มและเจาะจงเฉพาะข้อมูลที่ผู้ตอบไม่ต้องการเปิดเผย โดยมีขั้นตอนในการศึกษา ดังนี้

1) กำหนดกรอบตัวอย่าง

2) คำนวณค่าสัมประสิทธิ์ความแปรผันของตัวแปรที่เลือกมาศึกษา (ตารางที่ 2)

3) ในข้อมูลทุกชุด สุ่มให้มีข้อมูลบางส่วนสูญหาย ซึ่งเป็นการจำลองสถานการณ์ของการไม่ตอบโดยใช้วิธีสุ่มและเจาะจงเฉพาะข้อมูลที่ผู้ตอบไม่ต้องการเปิดเผย ในแต่ละตัวแปรที่ศึกษาสุ่มให้มีผู้ไม่ตอบตั้งแต่ร้อยละ 10 จนถึงร้อยละ 90

4) ประมาณค่าของตัวแปรที่ต้องการโดยการปรับค่าน้ำหนักด้วยวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างและวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร ใช้ขนาดตัวอย่างเท่ากับ 300 500 700 และ 1000

5) คำนวณค่าความเอนเอียงสัมพัทธ์ (Relative Bias) จากสูตร $\frac{(\bar{y}_{predict} - \bar{Y}_{actual})}{\bar{Y}_{actual}} \times 100$ โดย $\bar{y}_{predict}$ เป็นค่าเฉลี่ยที่ได้จากวิธีที่ได้ปรับค่าน้ำหนักและวิธีที่ไม่ได้ปรับค่าน้ำหนัก $\bar{Y}_{actual}$ เป็นค่าเฉลี่ยของประชากรและคำนวณค่าความคลาดเคลื่อนมาตรฐานสัมพัทธ์ซึ่งใช้เปรียบเทียบประสิทธิภาพของค่าประมาณที่ได้จากวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างและวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรกับวิธีที่ไม่ได้ปรับน้ำหนักจากสูตร $RSE_{Post\ vs\ Base} = \frac{se_{post} - se_{Base}}{se_{Base}}$

$RSE_{Raking\ vs\ Base} = \frac{se_{Raking} - se_{Base}}{se_{Base}}$ ตามลำดับ ค่าความคลาดเคลื่อนมาตรฐานสัมพัทธ์ซึ่งใช้เปรียบเทียบประสิทธิภาพของค่าประมาณที่ได้จากวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรกับวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างจากสูตร

$$RSE_{Raking\ vs\ Post} = \frac{se_{Raking} - se_{Post}}{se_{Post}}$$

โดย se_{Post}, se_{Raking} เป็นค่าความคลาดเคลื่อนมาตรฐานที่ได้จากวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร ตามลำดับ

ตารางที่ 2 ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐานและค่าสัมประสิทธิ์ความแปรผันของตัวแปรที่ใช้ศึกษา

เพศ	ตัวแปร	ค่าเฉลี่ย (Mean)	ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)	ค่าสัมประสิทธิ์ความแปรผัน (C.V.)
ชาย	รายได้ข้าราชการ	28259.37	18583.61	0.66
	รายได้คู่สมรส	22225.62	22249.63	1.00
	หนี้สินทั้งหมด	665966.3	780816.05	1.17
	ค่าใช้จ่ายของครัวเรือน	33803.61	27367.04	0.81
	ค่าใช้จ่ายเพื่อการสะสมทุน	8839.009	20943.23	2.37
หญิง	รายได้ข้าราชการ	21478.99	13044.13	0.61
	รายได้คู่สมรส	26501.61	23436.88	0.88
	หนี้สินทั้งหมด	576601.2	659996.65	1.14
	ค่าใช้จ่ายของครัวเรือน	27594.12	21088.30	0.76
	ค่าใช้จ่ายเพื่อการสะสมทุน	6587.442	12856.31	1.95
	รายได้ข้าราชการ	24370.75	16002.91	0.66
	รายได้คู่สมรส	24450.56	22972.66	0.94
ผลรวม	หนี้สินทั้งหมด	614966.4	715700.28	1.16
	ค่าใช้จ่ายของครัวเรือน	30242.4	24162.96	0.80
	ค่าใช้จ่ายเพื่อการสะสมทุน	7547.709	16824.71	2.23

6) ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าสัมบูรณ์ของค่าความเอนเอียงสัมพัทธ์ของค่าประมาณที่ได้จากสามวิธีนี้โดยใช้วิธีวิเคราะห์ความแปรปรวน เมื่อจำแนกตามค่าสัมประสิทธิ์ความแปรผัน (แบ่งเป็น 3 กลุ่ม คือ (1) น้อยกว่า 1.00 (2) 1.00-1.17 (3) มากกว่า 1.17) ขนาดตัวอย่าง (300 500 700 และ 1000) และค่าร้อยละของหน่วยตัวอย่างที่ไม่ตอบ (แบ่งเป็น 3 กลุ่ม คือ (1) 0-30 (2) 31-60 (3) มากกว่า 60) โดยใช้ระดับนัยสำคัญ 0.05

7) ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าความคลาดเคลื่อนมาตรฐานสัมพัทธ์ของค่าประมาณที่ได้จากวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง เมื่อเทียบกับวิธีที่ไม่ได้ปรับค่าน้ำหนักกับค่าความคลาดเคลื่อนมาตรฐานสัมพัทธ์ของค่าประมาณที่ได้จากวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรเมื่อเทียบกับวิธีที่ไม่ได้ปรับค่าน้ำหนักโดยใช้ค่าสถิติทีด้วยระดับนัยสำคัญ 0.05

ในการศึกษาได้ประมวลผลโดยใช้โปรแกรม WesVar ซึ่งเป็นโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลที่ได้จากการสำรวจด้วยตัวอย่างโดยเฉพาะและทำงานภายใต้ระบบปฏิบัติการวินโดวส์ซึ่งมีความยืดหยุ่นสูง ใช้ได้กับข้อมูลที่ได้จากการวางแผนการสุ่มตัวอย่างหลายแบบ ทั้งแผนการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) และแผนการสุ่มตัวอย่างแบบซับซ้อน (Complex Sample Design) เช่น แผนการสุ่มตัวอย่างแบบหลายขั้น (Multistage Sampling) แผนการสุ่มตัวอย่างแบบชั้นภูมิ (Stratified Sampling) แผนการสุ่มตัวอย่างแบบใช้ความน่าจะเป็นไม่เท่ากัน

(Unequal Probability Sampling) เป็นต้น ในการหาค่าความแปรปรวนของค่าประมาณใช้เทคนิค Replication ซึ่งจะแบ่งข้อมูลตัวอย่างออกเป็นกลุ่มย่อยจำนวน G กลุ่ม คำนวณค่าประมาณที่สนใจจากกลุ่มตัวอย่างย่อยเหล่านั้นแล้วจึงนำค่าประมาณจากกลุ่มตัวอย่างย่อยไปหาค่าความแปรปรวนของค่าประมาณ (Westat, 2002)

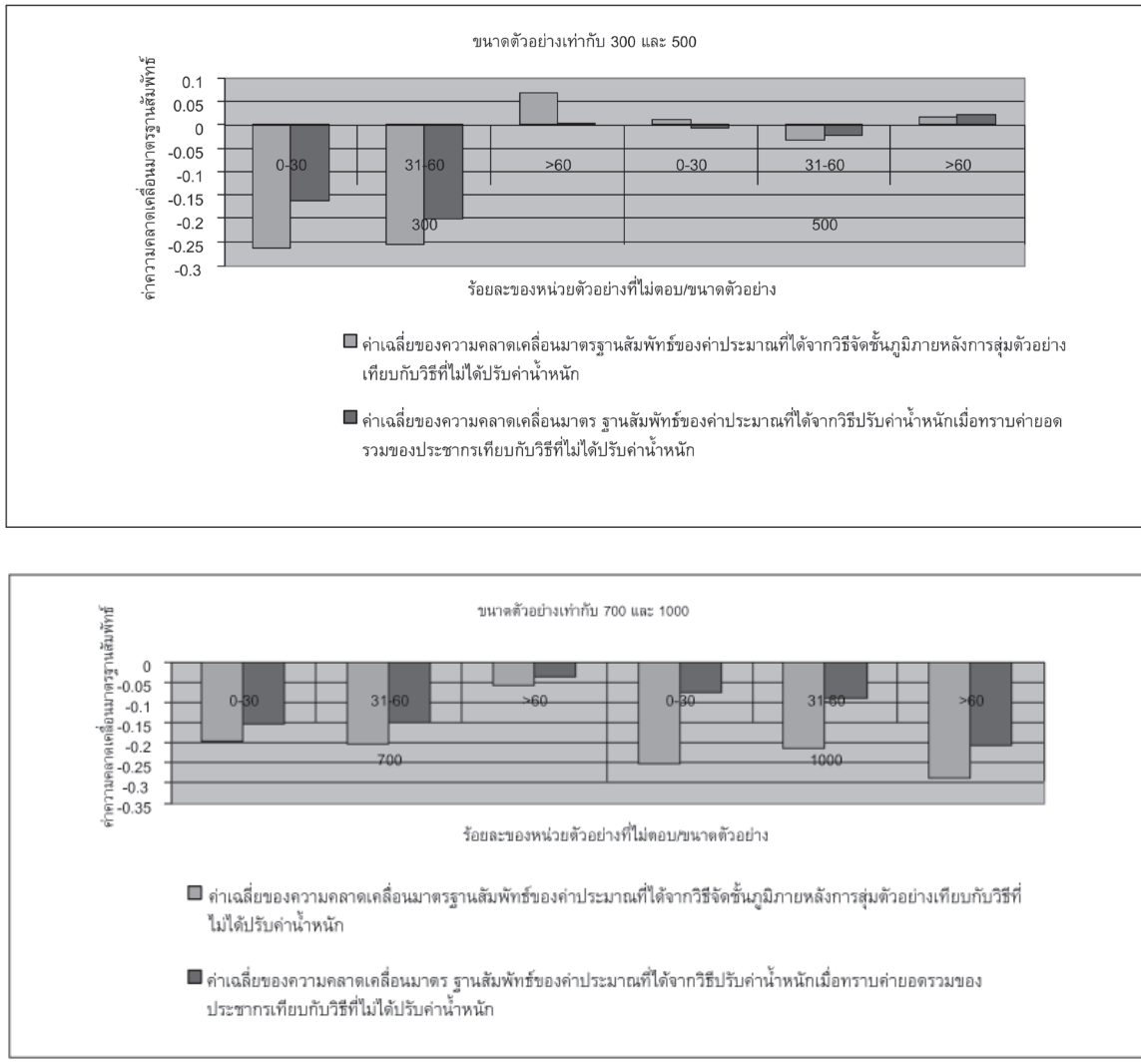
ผลการศึกษา

จากการเปรียบเทียบค่าความเอนเอียงในการประมาณค่าเฉลี่ยของประชากรที่ได้จากสองวิธีนี้ และค่าความเอนเอียงที่ได้เมื่อประมาณค่าโดยไม่ได้มีการปรับค่าน้ำหนัก (base weight) พบว่ามีค่าไม่ค่อยแตกต่างกันยกเว้นในกรณีที่ข้อมูลมีการกระจายน้อย (ค่าสัมประสิทธิ์ความแปรผัน = 0.66) และใช้ตัวอย่างขนาดเล็ก (ขนาด 300) หากมีผู้ที่ไม่ตอบมากกว่าร้อยละ 30 ค่าประมาณที่ไม่ได้มีการปรับค่าน้ำหนักจะมีความเอนเอียงมากกว่าค่าประมาณที่มีการปรับค่าน้ำหนักด้วย 2 วิธีนี้ และเมื่อเพิ่มขนาดตัวอย่างให้มากขึ้นค่าความเอนเอียงจะมีค่าน้อยลงในทุกๆ ระดับของค่าสัมประสิทธิ์ความแปรผัน

จากการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของค่าสัมบูรณ์ของค่าความเอนเอียงสัมพัทธ์ของค่าประมาณที่ได้จากสามวิธีนี้ เมื่อจำแนกตามค่าสัมประสิทธิ์ความแปรผัน (แบ่งเป็น 3 กลุ่ม คือ (1) น้อยกว่า 1.00 (2) 1.00-1.17 (3) มากกว่า 1.17) ขนาดตัวอย่าง (300 500 700 และ 1000) และค่าร้อยละของหน่วยตัวอย่างที่ไม่ตอบ (แบ่งเป็น 3 กลุ่ม คือ (1) 0-30 (2) 31-60 (3) มากกว่า 60) โดยใช้ระดับนัยสำคัญ 0.05 พบว่า มีผลการทดสอบไม่แตกต่างกัน นอกจากนี้ ยังพบว่าเมื่อมีจำนวนหน่วยตัวอย่างที่ไม่ตอบเพิ่มขึ้นค่าประมาณที่ได้จากสามวิธีนี้จะให้ค่าเฉลี่ยของความเอนเอียงสัมพัทธ์เพิ่มขึ้น และเมื่อพิจารณาในแต่ละระดับของค่าร้อยละของหน่วยตัวอย่างที่ไม่ตอบและขนาดตัวอย่างพบว่าค่าประมาณที่ได้จากข้อมูลที่มีค่าสัมประสิทธิ์ความแปรผันมากจะมีค่าเฉลี่ยของค่าความเอนเอียงสัมพัทธ์น้อยกว่าข้อมูลที่มีค่าสัมประสิทธิ์ความแปรผันน้อย

จากการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานสัมพัทธ์ พบว่า ในกรณีที่ประชากรมีการกระจายน้อย วิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างและวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร ให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าน้อยกว่าวิธีที่ไม่ได้ปรับค่าน้ำหนัก วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานน้อยกว่าวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง ในกรณีที่ประชากรมีการกระจายมากขึ้น สองวิธีนี้จะให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานแตกต่างกันน้อยลง นอกจากนี้ ยังพบว่าเมื่อใช้ขนาดตัวอย่างมากขึ้นวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างและวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานใกล้เคียงกับวิธีที่ไม่ได้ปรับน้ำหนักมากขึ้น

จากการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานสัมพัทธ์ของค่าประมาณที่ได้จากวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างเมื่อเทียบกับวิธีที่ไม่ได้ปรับค่าน้ำหนักกับค่าความคลาดเคลื่อนมาตรฐานสัมพัทธ์ของค่าประมาณที่ได้จากวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร เมื่อเทียบกับวิธีที่ไม่ได้ปรับค่าน้ำหนักโดยใช้ค่าสถิติทีด้วยระดับนัยสำคัญ 0.05 พบว่า ในกรณีที่ข้อมูลมีค่าสัมประสิทธิ์ความแปรผันน้อยกว่า



รูปที่ 1 ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานสัมพัทธ์ของค่าประมาณที่ได้จากวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรเทียบกับวิธีที่ไม่ได้ปรับค่าน้ำหนัก

1.00 ในทุกระดับของค่าร้อยละของหน่วยตัวอย่างที่ไม่ตอบและทุกระดับของขนาดตัวอย่างให้ผลการทดสอบที่แตกต่างกัน (จากรูปที่ 1 ในกรณีที่ใช้ตัวอย่างขนาด 700 และ 1000 พบว่า วิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่างให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าน้อยกว่าวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร) ส่วนในกรณีที่ข้อมูลมีค่าสัมประสิทธิ์ความแปรผันมากกว่า 1.00 พบว่า ในทุกกรณีให้ผลการทดสอบที่ไม่แตกต่างกัน

สรุปผลการวิจัย

ในการเปรียบเทียบค่าประมาณที่ปรับค่าจากการไม่ได้รับคำตอบด้วยวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากรและวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง พบว่า ถึงแม้สองวิธีนี้จะมีวิธีปรับค่าน้ำหนักที่แตกต่างกัน กล่าวคือ ในการใช้วิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง จำเป็นต้องทราบการแจกแจงร่วมของตัวแปรทุกตัวที่นำมาใช้ในการปรับค่าน้ำหนัก ส่วนวิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร จะปรับค่าหนักโดยการแทนค่าซ้ำ ดังนั้น จึงไม่จำเป็นต้องทราบการแจกแจงร่วมแต่สองวิธีนี้จะให้ค่าประมาณที่มีความเอนเอียงไม่แตกต่างกันและมีค่าไม่แตกต่างกับวิธีที่ไม่ได้ค่าปรับน้ำหนัก เมื่อเพิ่มขนาดตัวอย่างให้มากขึ้นก็พบว่าในทุกระดับของค่าสัมประสิทธิ์ความผันแปรจะให้ค่า ประมาณที่มีความเอนเอียงน้อยลง แต่ไม่สามารถที่จะกำจัดให้ค่าความเอนเอียงหมดไปได้ เนื่องจากในการสำรวจด้วยตัวอย่างมีหลายขั้นตอนและหน่วยตัวอย่างแต่ละหน่วยก็อาจมีสาเหตุของการไม่ตอบที่แตกต่างกัน จึงเป็นการยากที่จะประมาณค่าความเอนเอียง สำหรับการเปรียบเทียบประสิทธิภาพของค่าประมาณโดยพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานพบว่าในกรณีที่ประชากรมีการกระจายน้อยสองวิธีนี้ให้ค่าประมาณที่มีค่าความคลาดเคลื่อนมาตรฐานน้อยกว่าวิธีที่ไม่ได้ปรับน้ำหนัก วิธีปรับค่าน้ำหนักเมื่อทราบค่ายอดรวมของประชากร ให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานน้อยกว่าวิธีจัดชั้นภูมิภายหลังการสุ่มตัวอย่าง ในกรณีที่ประชากรมีการกระจายมากขึ้นสองวิธีนี้จะให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานแตกต่างกันน้อยลง นอกจากนี้ ยังพบว่าเมื่อใช้ขนาดตัวอย่างมากขึ้นการปรับค่าน้ำหนักด้วยสองวิธีนี้จะให้ค่าประมาณที่มีความคลาดเคลื่อนมาตรฐานใกล้เคียงกับวิธีที่ไม่ได้ปรับน้ำหนักมากขึ้น

ในการศึกษาหากมีการแบ่งช่วงของร้อยละของหน่วยตัวอย่างที่ไม่ตอบในช่วงร้อยละ 0 - ร้อยละ 30 ให้มีช่วงแคบกว่านี้และศึกษาเพิ่มเติมในกรณีที่ใช้ขนาดตัวอย่างเท่ากับ 100 และ 200 รวมทั้งการเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error) ของค่าประมาณจะทำให้ได้ข้อสรุปที่ชัดเจนมากขึ้น

กิตติกรรมประกาศ

ขอขอบคุณมหาวิทยาลัยหอการค้าไทยที่สนับสนุนทุนในการดำเนินงานวิจัย

บรรณานุกรม

Battaglia, Michael P. 2008. **Improving Standard Poststratification Techniques For Random-Digit-Dialing Telephone Surveys: Survey Research Methods** [Online]. Available: http://www.surveymethods.org

Bethlehem, J.G., and Schouten, B. 2004. **Nonresponse Adjustment in Household Surveys.** Voorburg/Heerlen: Statistics Netherlands.

Cervantes, Ismael Flores, and Brick, J. Michael. 2008. **Empirical Evaluation of Raking Ratio Adjustments for Nonresponse** [Online]. Available: http://www.amstat.org/meetings/JSM/2008/

Choudhry, G. Hussain, and Valliant, Richard. 2002. **Wesvar: Software for Complex Survey Data Analysis** [Online]. Available: http://www.statcan.gc.ca/pub/11-522-x/2002001/session6/6728-eng.pdf

Deming, W.E., and Stephan, F.F. 1940. **On a Least Square Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are known** [Online]. Available: http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aoms/1177731829

Kalton, Graham. 1981. **Compensating for Missing Survey Data.** Ann Arbor, MI: Institute for Social Research.

Westat. 2002. **WesVar 4.2 User's Guide.** Rockville, MD: Westat.

Assoc. Prof. Doungporn Hatchavanich received her Master of Science Degree in Statistics from Chulalongkorn University, Thailand. She is currently a lecturer in the School of Science, University of the Thai Chamber of Commerce. Her main interests are in Sampling Techniques and Regression Analysis.