The Journal of Applied Science **Vol. 17 No. 2: 98-117 [2018]**
วารสารวิทยาศาสตร์ประยุกต์ doi: 10.14416/j.appsci.2018.12.001
ISSN 1513-7805 (Printed in Thailand) ISSN 2586-9663 (Online)

Research Article

ตัวประมาณค่าเฉลี่ยประชากรที่ปรับปรุงโดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กันสำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม

An improved estimator of population mean using dual to ratio-cum-product for post-stratified sampling

จิตรานนท์ วรรณพงศ์ [1]* และ สุพรรณี อึ้งปัญสัตวงศ์[2]

Jittranon Wannapong [1]* and Supunnee Ungpansattawong[2]

[1]สาขาสถิติประยุกต์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น อำเภอเมือง จังหวัดขอนแก่น 40002

Division of Applied Statistics, Department of Statistics, Faculty of Science, Khon Kaen University, Muang, Khon Kaen 40002

[2]ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น อำเภอเมือง จังหวัดขอนแก่น 40002

Department of Statistics, Faculty of Science, Khon Kaen University, Muang, Khon Kaen 40002

***E-mail**: w.jittranon@gmail.com

Received: 05/09/2018; Accepted: 11/12/2018

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กันสำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่มซึ่งปรับปรุงมาจากตัวประมาณของ Singh et al. (2005) พร้อมทั้งหาค่าคลาดเคลื่อนกำลังสองเฉลี่ย (mean square error: MSE) ของตัวประมาณที่นำเสนอ และทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอกับตัวประมาณแบบอัตราส่วนร่วมกับผลคูณของ Tailor et al. (2016) โดยใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าประสิทธิภาพสัมพัทธ์ ผลการวิจัยพบว่า ตัวประมาณที่นำเสนอมีประสิทธิภาพสูงกว่าตัวประมาณแบบอัตราส่วนร่วมกับผลคูณของ Tailor et al. (2016) อย่างมีเงื่อนไขและผลการศึกษาเชิงทฤษฎีของค่าคลาดเลื่อนกำลังสองเฉลี่ยแสดงให้เห็นจริงโดยการคำนวณเชิงตัวเลข

คำสำคัญ: การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม, ค่าคลาดเคลื่อนกำลังสองเฉลี่ย, ค่าประสิทธิภาพสัมพัทธ์, ตัวประมาณอัตราส่วนร่วมกับผลคูณแบบคู่กัน

Abstract

The aim of research was to propose the estimator of population mean using dual to ratio cum product for post-stratified sampling. This proposed estimator was modified from the estimator of Singh et al. (2005). Mean squared error of the proposed estimator was obtained. The efficiency of this proposed estimator was compared with ratio cum product estimator by Tailor et al. (2016) on the basis of mean squared error and relative efficiency. The results revealed that the proposed estimator was more efficient than ratio cum product estimator by Tailor et al. (2016). In additions, the theoretical of mean square error of the estimator were illustrated with numerical computation.

Keywords: post-stratification, mean squared error, relative efficiency, dual to ratio cum product estimator

บทนำ

ในการศึกษาและการวิจัยเชิงสำรวจ ผู้วิจัยต้องทำการเก็บรวบรวมข้อมูลของประชากรจากประชากรเป้าหมาย ในกรณีที่ไม่อาจศึกษาประชากรได้ทั้งหมดจะต้องทำกรอบตัวอย่างในการสุ่มตัวอย่างเพื่อหาตัวแทนที่ดี ซึ่งการที่จะได้ตัวแทนที่ดีของประชากรนั้น ต้องอาศัยวิธีการทางสถิติที่เรียกว่า เทคนิคการเลือกตัวอย่าง (sampling techniques) เข้ามาช่วยเพื่อให้ได้ตัวอย่างที่เป็นตัวแทนที่ดีของประชากรมากที่สุด

เมื่อพิจารณาการสุ่มตัวอย่างขนาด n จากประชากรขนาด N ด้วยวิธีการเลือกตัวอย่างสุ่มแบบง่ายแบบไม่ใส่คืน เมื่อต้องการประมาณค่าเฉลี่ยประชากร $\left(\bar{Y}\right)$ ของตัวแปร y ที่สนใจศึกษา ตัวประมาณที่ใช้ประมาณค่าเฉลี่ยประชากรคือ $\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}$ ซึ่งเป็นตัวประมาณไม่เอนเอียง (unbiased estimator) ของ $\bar{Y}$ โดยตัวประมาณ ดังกล่าวมีความแปรปรวน คือ $V\left(\bar{y}\right)=\left(1-f\right)\frac{S^2}{n}$ โดยที่ $f=\frac{n}{N}$ คือ สัดส่วนตัวอย่าง (sample fraction)

ในบางสถานการณ์ เราอาจไม่มีจุดประสงค์ที่จะทำการศึกษาคุณลักษณะตามกลุ่มของตัวแปรที่สนใจศึกษา จึงไม่ได้แบ่งประชากรออกเป็นชั้นภูมิก่อนที่จะทำการเลือกตัวอย่างจากประชากร เช่น ในการสำรวจประชากรกลุ่มหนึ่ง เราต้องการศึกษาคุณสมบัติบางประการของประชากรทั้งหมด โดยการเลือกตัวอย่างสุ่มแบบง่ายแบบไม่ใส่คืนมา n หน่วย จากประชากรขนาด N หน่วย แล้วจึงเกิดประเด็นที่ต้องศึกษาแยกกลุ่มลงไปอีก จึงต้องนำหน่วยตัวอย่างขนาด n หน่วย ที่สุ่มมาได้นี้ มาแบ่งออกเป็นชั้นภูมิต่างๆ L ชั้นภูมิ ตามตัวแปรที่สนใจศึกษา เช่น เพศ การศึกษา เป็นต้น วิธีนี้ เรียกว่า การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม (post-stratified sampling)

เมื่อทำการเก็บรวบรวมข้อมูลจากตัวอย่างซึ่งเป็นส่วนหนึ่งของประชากร โดยวิธีการเลือกตัวอย่างแล้ว ตัวประมาณ (estimator) ถือเป็นสิ่งสำคัญสำหรับการวิเคราะห์และสรุปผลไปยังคุณลักษณะของประชากรที่สนใจ

ศึกษา ซึ่งคุณลักษณะต่าง ๆ ของประชากร เรียกว่า พารามิเตอร์ (parameter) ได้แก่ ยอดรวมประชากร ค่าเฉลี่ยประชากร สัดส่วนประชากร และความแปรปรวนของประชากร เป็นต้น

การประมาณค่าเฉลี่ยประชากรในรูปแบบหนึ่ง คือ การนำตัวแปรช่วย x เข้ามาช่วยเพิ่มประสิทธิภาพในการประมาณค่าเฉลี่ยประชากรให้ดียิ่งขึ้น โดยตัวแปรช่วย x ต้องมีความสัมพันธ์กับตัวแปร y ที่สนใจศึกษาในทิศทางเดียวกัน ตัวประมาณดังกล่าวเรียกว่า ตัวประมาณแบบอัตราส่วน (ratio estimator) แทนสัญลักษณ์ด้วย $\bar{y}_R$ โดยมีรูปแบบดังนี้ (Suwattee, 2011)

$$\bar{y}_R = \bar{y}\frac{\bar{X}}{\bar{x}}$$

แต่ถ้าตัวแปรช่วย x มีความสัมพันธ์กับตัวแปร y ที่สนใจศึกษา ในทิศทางตรงกันข้าม ตัวประมาณดังกล่าวเรียกว่า ตัวประมาณแบบผลคูณ (product estimator) แทนสัญลักษณ์ด้วย $\bar{y}_P$ โดยมีรูปแบบดังนี้

$$\bar{y}_P = \bar{y}\frac{\bar{x}}{\bar{X}}$$

นอกจากตัวประมาณแบบอัตราส่วนและตัวประมาณแบบผลคูณดังกล่าวข้างต้นแล้ว ยังมีการนำเสนอตัวประมาณแบบอื่นในการประมาณค่าเฉลี่ยประชากรอีกหลายตัว เช่น

Singh (1967) ได้นำเสนอตัวประมาณแบบอัตราส่วนร่วมกับผลคูณที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบง่าย ซึ่งเป็นการนำตัวประมาณแบบอัตราส่วนมาผสมกับตัวประมาณแบบผลคูณ เพื่อเพิ่มประสิทธิภาพของตัวประมาณโดยใช้ตัวแปรช่วย 2 ตัว คือตัวแปรช่วย x ซึ่งมีความสัมพันธ์ในทิศทางเดียวกันกับตัวแปร y ที่สนใจศึกษา และตัวแปรช่วย z ซึ่งมีความสัมพันธ์ในทิศทางตรงกันข้ามกับตัวแปร y ที่สนใจศึกษา มาใช้ในการสร้างตัวประมาณ แทนสัญลักษณ์ด้วย $\bar{y}_{RP}$ โดยมีรูปแบบดังนี้

$$\bar{y}_{RP} = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\frac{\bar{z}}{\bar{Z}}\right)$$

Srivenkataramana (1980) ได้นำเสนอตัวประมาณแบบอัตราส่วนคู่กันขึ้นเป็นครั้งแรก ซึ่งมีชื่อเรียกว่า ตัวประมาณแบบอัตราส่วนคู่กัน (dual to ratio estimator) ที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบง่าย แทนสัญลักษณ์ด้วย $\bar{y}_R^{(d)}$ โดยมีรูปแบบดังนี้

$$\bar{y}_R^{(d)} = \bar{y}\left(\frac{\bar{x}^*}{\bar{X}}\right)$$

เมื่อ $\bar{x}^* = \dfrac{N\bar{X} - n\bar{x}}{N-n}$

และในปีเดียวกัน Bandyopadyay (1980) ได้นำเสนอตัวประมาณแบบผลคูณคู่กันที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบง่าย แทนสัญลักษณ์ด้วย $\bar{y}_P^{(d)}$ โดยมีรูปแบบดังนี้

$$\bar{y}_P^{(d)} = \bar{y}\left(\frac{\bar{Z}}{\bar{z}^*}\right)$$

 doi: 10.14416/j.appsci.2018.12.001

เมื่อ $\bar{z}^* = \dfrac{N\bar{Z} - n\bar{z}}{N-n}$

Singh et al. (2005) ได้นำเสนอตัวประมาณแบบอัตราส่วนร่วมกับผลคูณคู่กันที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบง่าย โดยทำการปรับปรุงตัวประมาณค่าเฉลี่ยประชากรของ Singh (1967) ด้วยการแปลงค่า x_i และ z_i ให้อยู่ในรูปของ $x_i^* = \dfrac{\bar{X}N - x_i n}{N-n}$ และ $z_i^* = \dfrac{\bar{Z}N - z_i n}{N-n}$, $i = 1,2,3,\ldots,n$ แทนสัญลักษณ์ด้วย $\bar{y}_{RP}^*$ โดยมีรูปแบบดังนี้

$$\bar{y}_{RP}^* = \bar{y}\left(\frac{\bar{x}^*}{\bar{X}}\frac{\bar{Z}}{\bar{z}^*}\right)$$

Tailor et al. (2012) ได้นำเสนอตัวประมาณแบบอัตราส่วนร่วมกับผลคูณที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ โดยใช้ตัวแปรช่วย 2 ตัวแปร คือตัวแปรช่วย x และ z ซึ่งมีแนวคิดมาจากตัวประมาณแบบอัตราส่วนร่วมกับผลคูณที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบง่ายของ Singh (1967) แทนสัญลักษณ์ด้วย $\bar{y}_{RP}^{ST}$ โดยมีรูปแบบดังนี้

$$\bar{y}_{RP}^{ST} = \bar{y}_{st}\left(\frac{\bar{X}}{\bar{x}_{st}}\frac{\bar{z}_{st}}{\bar{Z}}\right)$$

Tailor et al. (2016) ได้นำเสนอตัวประมาณแบบอัตราส่วนร่วมกับผลคูณที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม ซึ่งมีแรงบันดาลใจมาจากตัวประมาณแบบอัตราส่วนร่วมกับผลคูณที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิ ของ Tailor et al. (2012) แทนสัญลักษณ์ด้วย $\bar{y}_{RP}^{PS}$ โดยมีรูปแบบดังนี้

$$\bar{y}_{RP}^{PS} = \bar{y}_{ps}\left(\frac{\bar{X}}{\bar{x}_{ps}}\frac{\bar{z}_{ps}}{\bar{Z}}\right)$$

เมื่อ $\bar{y}_{ps} = \sum_{h=1}^{L} W_h \bar{y}_h$, $\bar{x}_{ps} = \sum_{h=1}^{L} W_h \bar{x}_h$ และ $\bar{z}_{ps} = \sum_{h=1}^{L} W_h \bar{z}_h$

มีค่าความเอนเอียงและค่าคลาดเคลื่อนกำลังสองเฉลี่ย คือ

$$Bias\left(\bar{y}_{RP}^{PS}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h \left\{\frac{R_2}{\bar{Y}} S_{yzh} - \frac{R_1}{\bar{Z}} S_{xzh} + \frac{R_1}{\bar{X}} S_{xh}^2 - \frac{R_1}{\bar{Y}} S_{yxh}\right\}$$

$$MSE\left(\bar{y}_{RP}^{PS}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h \left[S_{yh}^2 - R_1^2 S_{xh}^2 + R_2^2 S_{zh}^2 - 2\left(R_1 S_{yxh} + R_1 R_2 S_{xzh} - R_2 S_{yzh}\right)\right]$$

จากการศึกษาดังกล่าวข้างต้น ผู้วิจัยสนใจที่จะนำเสนอตัวประมาณค่าเฉลี่ยประชากรที่ปรับปรุงใหม่ ของ Singh et al. (2005) ให้มีประสิทธิภาพมากยิ่งขึ้น ภายใต้การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม โดยการปรับปรุงตัวประมาณค่าเฉลี่ยประชากร ด้วยการแปลงค่า x_{hi} และ z_{hi} ให้อยู่ในรูปของ

$x_{hi}^* = \dfrac{\bar{X}_h N_h - x_{hi} n_h}{N_h - n_h}$ และ $h = 1,2,3,\ldots,L$

ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อนำเสนอตัวประมาณค่าเฉลี่ยประชากรโดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กันสำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่มและเปรียบเทียบประสิทธิภาพของตัวประมาณใหม่ที่นำเสนอในงานวิจัยนี้กับตัวประมาณของ Tailor et al. (2016)

วิธีการทดลอง

ในการวิจัยครั้งนี้ ผู้วิจัยได้เสนอตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กันสำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม และเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอ โดยใช้ค่าประสิทธิภาพสัมพัทธ์เป็นเกณฑ์ พร้อมทั้งแสดงตัวอย่างการคำนวณโดยใช้ข้อมูล 2 ชุด ได้แก่ ข้อมูลผลผลิตทางการเกษตรของประเทศอินเดีย ในปี 2014-2015 มีจำนวนผลผลิตทั้งหมด 72 ชนิด โดยอัตราความสามารถในการผลิตพืชผล (หน่วย : ตันต่อเฮกตาร์) เป็นตัวแปร y ที่สนใจศึกษา มีตัวแปรช่วย x และ z เป็นปริมาณการผลิตพืชผล (หน่วย : ตัน) และพื้นที่ที่ใช้ในการปลูกพืชผล (หน่วย : เฮกตาร์) ตามลำดับ และข้อมูลดัชนีมวลกาย (body mass index: BMI) ของคนไข้ที่เข้ามาใช้บริการที่โรงพยาบาลหัวใจและศูนย์วิจัย Rajaei เมืองเตหะราน ประเทศอิหร่าน มีจำนวนคนไข้ทั้งหมด 80 คน โดยค่าดัชนีมวลกาย (หน่วย : กิโลกรัม/เมตร^2) เป็นตัวแปร y ที่สนใจศึกษา มีตัวแปรช่วย x และ z เป็นน้ำหนัก (หน่วย : กิโลกรัม) และส่วนสูง (หน่วย : เมตร) ตามลำดับ มีขั้นตอนดังนี้

1. ตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม

ผู้วิจัยเสนอตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม โดยทำการแปลงค่า x_{hi} และ z_{hi} ให้อยู่ในรูปของ $x_{hi}^* = \frac{N_h\bar{X}_h - n_h x_{hi}}{N_h - n_h}$ และ $z_{hi}^* = \frac{N_h\bar{Z}_h - n_h z_{hi}}{N_h - n_h}$ แล้วจะได้ $\bar{x}_h^* = \frac{N_h\bar{X}_h - n_h\bar{x}_h}{N_h - n_h}$ และ $\bar{z}_h^* = \frac{N_h\bar{Z}_h - n_h\bar{z}_h}{N_h - n_h}$ เมื่อ $h = 1,2,3,\ldots,L$ ทำให้ได้ตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม แทนสัญลักษณ์ด้วย $\bar{y}_{RP}^{PS(d)}$ โดยมีรูปแบบดังนี้

$$\bar{y}_{RP}^{PS(d)} = \bar{y}_{ps}\left(\frac{\bar{x}_{ps}^*}{\bar{X}}\frac{\bar{Z}}{\bar{z}_{ps}^*}\right) \tag{1}$$

เมื่อ $\bar{y}_{ps} = \sum_{h=1}^{L} W_h\bar{y}_h$ คือ ค่าเฉลี่ยตัวอย่างในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่มของตัวแปร y ที่สนใจศึกษา ในชั้นภูมิที่ h

 doi: 10.14416/j.appsci.2018.12.001

$\bar{x}_{ps}^{*} = \sum_{h=1}^{L} W_h \bar{x}_h^{*}$ คือ ค่าเฉลี่ยตัวอย่างในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่มของตัวแปรช่วย x ในชั้นภูมิที่ h

$\bar{z}_{ps}^{*} = \sum_{h=1}^{L} W_h \bar{z}_h^{*}$ คือ ค่าเฉลี่ยตัวอย่างในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่มของตัวแปรช่วย z ในชั้นภูมิที่ h

$\bar{y}_h = \dfrac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปร y ที่สนใจศึกษา ในชั้นภูมิที่ h

$\bar{x}_h = \dfrac{\sum_{i=1}^{n_h} x_{hi}}{n_h}$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปรช่วย x ในชั้นภูมิที่ h

$\bar{z}_h = \dfrac{\sum_{i=1}^{n_h} z_{hi}}{n_h}$ คือ ค่าเฉลี่ยตัวอย่างของตัวแปรช่วย z ในชั้นภูมิที่ h

$\bar{X} = \sum_{h=1}^{L} W_h \bar{X}_h$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย x

$\bar{Z} = \sum_{h=1}^{L} W_h \bar{Z}_h$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย z

$\bar{X}_h = \dfrac{\sum_{h=1}^{N_h} x_{hi}}{N_h}$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย x ในชั้นภูมิที่ h

$\bar{Z}_h = \dfrac{\sum_{h=1}^{N_h} z_{hi}}{N_h}$ คือ ค่าเฉลี่ยประชากรของตัวแปรช่วย z ในชั้นภูมิที่ h

2. ค่าความเอนเอียงและค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ

2.1 การหาค่าความเอนเอียงของตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม

เมื่อนำเสนอตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม ให้อยู่ในรูปของ e_{ih} แล้วหาค่าความเอนเอียงของตัวประมาณได้จาก $Bias\left(\bar{y}_{RP}^{PS(d)}\right) = E\left(\bar{y}_{RP}^{PS(d)} - \bar{Y}\right)$

กำหนดให้ $\bar{y}_h, \bar{x}_h$ และ $\bar{z}_h$ เป็นตัวประมาณไม่เอนเอียงของ $\bar{Y}_h, \bar{X}_h$ และ $\bar{Z}_h$ ตามลำดับ ทำให้ได้ $\bar{y}_h = \bar{Y}_h\left(1+e_{0h}\right), \bar{x}_h = \bar{X}_h\left(1+e_{1h}\right)$ และ $\bar{z}_h = \bar{Z}_h\left(1+e_{2h}\right)$

จะได้ $e_{0h} = \dfrac{\bar{y}_h - \bar{Y}_h}{\bar{Y}_h}, e_{1h} = \dfrac{\bar{x}_h - \bar{X}_h}{\bar{X}_h}$ และ $e_{2h} = \dfrac{\bar{z}_h - \bar{Z}_h}{\bar{Z}_h}$

 doi: 10.14416/j.appsci.2018.12.001

เมื่อ $E(e_{0h}) = E(e_{1h}) = E(e_{2h}) = 0$

$$E(e_{0h}^2) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)C_{yh}^2$$

$$E(e_{1h}^2) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)C_{xh}^2$$

$$E(e_{2h}^2) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)C_{zh}^2$$

$$E(e_{0h}e_{1h}) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)\rho_{yxh}C_{yh}C_{xh}$$

$$E(e_{0h}e_{2h}) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)\rho_{yzh}C_{yh}C_{zh}$$

และ $E(e_{1h}e_{2h}) = \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)\rho_{xzh}C_{xh}C_{zh}$

จากนั้น แสดงตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม ให้อยู่ในรูปของ e_i

เมื่อ $e_0 = \frac{\sum_{h=1}^{L} W_h \bar{Y}_h e_{0h}}{\bar{Y}}, e_1 = \frac{\sum_{h=1}^{L} W_h \bar{X}_h e_{1h}}{\bar{X}}$ และ $e_2 = \frac{\sum_{h=1}^{L} W_h \bar{Z}_h e_{2h}}{\bar{Z}}$

เมื่อทำให้อยู่ในรูปของค่าคาดหวังจะทำให้ทราบความเอนเอียงของตัวประมาณที่นำเสนอ

เมื่อ $E(e_0) = E(e_1) = E(e_2) = 0$

$$E(e_0^2) = \frac{1}{\bar{Y}^2}\sum_{h=1}^{L} W_h^2 S_{yh}^2 \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

$$E(e_1^2) = \frac{1}{\bar{X}^2}\sum_{h=1}^{L} W_h^2 S_{xh}^2 \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

$$E(e_2^2) = \frac{1}{\bar{Z}^2}\sum_{h=1}^{L} W_h^2 S_{zh}^2 \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

$$E(e_0e_1) = \frac{1}{\bar{Y}\bar{X}}\sum_{h=1}^{L} W_h^2 S_{yxh} \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

$$E(e_0e_2) = \frac{1}{\bar{Y}\bar{Z}}\sum_{h=1}^{L} W_h^2 S_{yzh} \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

และ $E(e_1e_2) = \frac{1}{\bar{X}\bar{Z}}\sum_{h=1}^{L} W_h^2 S_{xzh} \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$

 doi: 10.14416/j.appsci.2018.12.001

2.2 การหาค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กันสำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม

หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม ได้จาก

$$MSE\left(\bar{y}_{RP}^{PS(d)}\right) = E\left(\bar{y}_{RP}^{PS(d)} - \bar{Y}\right)^2$$

เมื่อ $E(e_0) = E(e_1) = E(e_2) = 0$

$$E(e_0^2) = \frac{1}{\bar{Y}^2}\sum_{h=1}^{L} W_h^2 S_{yh}^2 \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

$$E(e_1^2) = \frac{1}{\bar{X}^2}\sum_{h=1}^{L} W_h^2 S_{xh}^2 \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

$$E(e_2^2) = \frac{1}{\bar{Z}^2}\sum_{h=1}^{L} W_h^2 S_{zh}^2 \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

$$E(e_0 e_1) = \frac{1}{\bar{Y}\bar{X}}\sum_{h=1}^{L} W_h^2 S_{yxh} \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

$$E(e_0 e_2) = \frac{1}{\bar{Y}\bar{Z}}\sum_{h=1}^{L} W_h^2 S_{yzh} \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$$

และ $E(e_1 e_2) = \frac{1}{\bar{X}\bar{Z}}\sum_{h=1}^{L} W_h^2 S_{xzh} \left(\frac{1}{nW_h} - \frac{1}{N_h}\right)$

3. การเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กันสำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม โดยใช้ค่าประสิทธิภาพสัมพัทธ์

ทำการเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม $\left(\bar{y}_{RP}^{PS(d)}\right)$ กับตัวประมาณแบบอัตราส่วนร่วมกับผลคูณที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม $\left(\bar{y}_{RP}^{PS}\right)$ ของ Tailor et al. (2016) ด้วยค่าประสิทธิภาพสัมพัทธ์ (relative efficiency: RE) (Ungpansattawong, 2011) ดังนี้

$$RE\left(\bar{y}_{RP}^{PS(d)}, \bar{y}_{RP}^{PS}\right) = \frac{MSE\left(\bar{y}_{RP}^{PS}\right)}{MSE\left(\bar{y}_{RP}^{PS(d)}\right)}$$

ถ้ามีค่าประสิทธิภาพสัมพัทธ์มากกว่า 1 แสดงว่า ตัวประมาณที่นำเสนอ $\left(\bar{y}_{RP}^{PS(d)}\right)$ มีประสิทธิภาพสูงกว่า ตัวประมาณ ของ Tailor et al. (2016) $\left(\bar{y}_{RP}^{PS}\right)$

Vol. 17 No. 2: 98-117 [2018]
doi: 10.14416/j.appsci.2018.12.001

4. ตัวอย่างเชิงตัวเลข

การแสดงตัวอย่างเชิงตัวเลขเพื่อทำการสนับสนุนผลของการศึกษาเชิงทฤษฎี โดยใช้ข้อมูล 2 ชุด ได้แก่

ข้อมูลชุดที่ 1 ข้อมูลผลผลิตทางการเกษตรของประเทศอินเดีย ในปี 2014-2015 มีจำนวนผลผลิตทั้งหมด 72 ชนิด โดยอัตราความสามารถในการผลิตพืชผล (หน่วย : ตันต่อเฮกตาร์) เป็นตัวแปร y ที่สนใจศึกษา มีตัวแปรช่วย x และ z เป็นปริมาณการผลิตพืชผล (หน่วย : ตัน) และพื้นที่ที่ใช้ในการปลูกพืชผล (หน่วย : เฮกตาร์) ตามลำดับ และข้อมูลชุดที่ 2 ข้อมูลดัชนีมวลกาย (body mass index: BMI) ของคนไข้ที่เข้ามาใช้บริการที่โรงพยาบาลหัวใจและศูนย์วิจัย Rajaei เมืองเตหะราน ประเทศอิหร่าน มีจำนวนคนไข้ทั้งหมด 80 คน โดยค่าดัชนีมวลกาย (หน่วย : กิโลกรัม/เมตร^2) เป็นตัวแปร y ที่สนใจศึกษา มีตัวแปรช่วย x และ z เป็นน้ำหนัก (หน่วย : กิโลกรัม) และส่วนสูง (หน่วย : เมตร) ตามลำดับ

ตารางที่ 1. ลักษณะของประชากรที่ศึกษาจากข้อมูลชุดที่ 1 และ 2

	ข้อมูลชุดที่ 1	**ข้อมูลชุดที่ 2**
N	72	80
$\bar{Y}$	13.787	26.86
$\bar{X}$	3518.632	72.39
$\bar{Z}$	218.555	164.26
S_y	8.340	126.709
S_x	5119.963	11.727
S_z	270.175	9.490
S_{yx}	9944.452	11.531
S_{yz}	470.507	-4.474
S_{xz}	746165.816	19.725

ข้อมูลชุดที่ 1 ข้อมูลผลผลิตทางการเกษตรของประเทศอินเดีย ในปี 2014-2015 มีจำนวนผลผลิตทั้งหมด 72 ชนิด โดยอัตราความสามารถในการผลิตพืชผล (หน่วย : ตันต่อเฮกตาร์) เป็นตัวแปร y ที่สนใจศึกษา มีตัวแปรช่วย x และ z เป็นปริมาณการผลิตพืชผล (หน่วย : ตัน) และพื้นที่ที่ใช้ในการปลูกพืชผล (หน่วย : เฮกตาร์) ตามลำดับ ผู้วิจัยทำการสุ่มตัวอย่างจากประชากร โดยขนาดตัวอย่างที่จะทำการศึกษาแบ่งออกเป็น 4 ระดับ คือ ตัวอย่างขนาด $n = 8, n = 10, n = 20$ และ $n = 30$ โดยใช้วิธีการเลือกตัวอย่างสุ่มแบบง่ายแบบไม่ใส่คืน ในแต่ละ

ขนาดตัวอย่างตามที่กำหนด เมื่อทราบรายละเอียดข้อมูลของตัวอย่างที่สุ่มมาแล้วนั้น จึงนำมาแบ่งออกเป็นชั้นภูมิตามที่กำหนดไว้ โดยแบ่งผลผลิตทางการเกษตรออกเป็น 2 ชั้นภูมิ ได้แก่ ชั้นภูมิที่ 1 คือ ผลผลิตของผลไม้ และชั้นภูมิที่ 2 คือ ผลผลิตของผัก แสดงดังตารางที่ 2-5 ค่าสถิติของประชากรที่ศึกษา

ตารางที่ 2. ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 1 ขนาดตัวอย่าง $n = 8$

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
N_h	36	36
n_h	4	4
$\bar{y}_h$	8.128	13.987
$\bar{x}_h$	1430.531	2839.799
$\bar{z}_h$	180.017	207.010
s_{yh}	2.907	2.000
s_{xh}	912.291	2577.583
s_{zh}	84.389	190.110
s_{yxh}	1292.232	-820.591
s_{yzh}	-43.531	-74.204
s_{xzh}	59512.611	489673.515

ตารางที่ 3. ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 1 ขนาดตัวอย่าง $n = 10$

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
N_h	36	36
n_h	6	4
$\bar{y}_h$	12.844	15.168
$\bar{x}_h$	1688.291	495.533
$\bar{z}_h$	103.250	34.167
s_{yh}	7.296	7.927
s_{xh}	2639.812	733.606
s_{zh}	150.271	38.025

Vol. 17 No. 2: 98-117 [2018]
doi: 10.14416/j.appsci.2018.12.001

ตารางที่ 3. (ต่อ) ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 1 ขนาดตัวอย่าง n = 10

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
s_{yxh}	9051.818	1190.619
s_{yzh}	434.590	-30.296
s_{xzh}	394842.7542	25920.887

ตารางที่ 4. ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 1 ขนาดตัวอย่าง n = 20

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
N_h	36	36
n_h	8	12
$\bar{y}_h$	500.050	15.984
$\bar{x}_h$	4221.977	9144.714
$\bar{z}_h$	268.654	490.747
s_{yh}	7.515	5.139
s_{xh}	3250.592	9104.228
s_{zh}	147.406	459.487
s_{yxh}	18156.970	30993.080
s_{yzh}	336.606	1418.795
s_{xzh}	407322.2163	4135123.318

ตารางที่ 5. ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 1 ขนาดตัวอย่าง n = 30

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
N_h	36	36
n_h	17	13
$\bar{y}_h$	10.246	19.031
$\bar{x}_h$	1987.557	2657.933
$\bar{z}_h$	184.729	145.005

doi: 10.14416/j.appsci.2018.12.001

ตารางที่ 5. (ต่อ) ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 1 ขนาดตัวอย่าง n = 30

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
s_{yh}	6.774	11.649
s_{xh}	2491.883	4138.898
s_{zh}	161.103	233.475
s_{yxh}	7322.860	186.250
s_{yzh}	100.693	-110.152
s_{xzh}	323422.388	957806.448

ข้อมูลชุดที่ 2 ข้อมูลดัชนีมวลกาย (body mass index: BMI) ของคนไข้ที่เข้ามาใช้บริการที่โรงพยาบาลหัวใจและศูนย์วิจัย Rajaei เมืองเตหะราน ประเทศอิหร่าน มีจำนวนคนไข้ทั้งหมด 80 คน โดยค่าดัชนีมวลกาย (หน่วย : กิโลกรัม/เมตร^2) เป็นตัวแปร y ที่สนใจศึกษา มีตัวแปรช่วย x และ z เป็นน้ำหนัก (หน่วย : กิโลกรัม) และส่วนสูง (หน่วย : เมตร) ตามลำดับ ผู้วิจัยทำการสุ่มตัวอย่างจากประชากร โดยขนาดตัวอย่างที่จะทำการศึกษาแบ่งออกเป็น 4 ระดับ คือ ตัวอย่างขนาด $n=8, n=10, n=20$ และ $n=30$ โดยใช้วิธีการเลือกตัวอย่างสุ่มแบบง่ายแบบไม่ใส่คืน ในแต่ละขนาดตัวอย่างตามที่กำหนด เมื่อทราบรายละเอียดข้อมูลของตัวอย่างที่สุ่มมาแล้วนั้น นำมาแบ่งออกเป็นชั้นภูมิ ตามที่กำหนดไว้ โดยแบ่งคนไข้ที่เข้ามาใช้บริการออกเป็น 2 ชั้นภูมิ ได้แก่ ชั้นภูมิที่ 1 คือ เพศหญิง และชั้นภูมิที่ 2 คือ เพศชาย แสดงดังตารางที่ 6-9 ค่าสถิติของประชากรที่ศึกษา

ตารางที่ 6. ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 2 ขนาดตัวอย่าง n = 8

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
N_h	35	45
n_h	2	6
$\bar{y}_h$	23.302	24.505
$\bar{x}_h$	63.000	71.667
$\bar{z}_h$	164.000	171.000
s_{yh}	3.123	1.637
s_{xh}	12.728	5.820
s_{zh}	5.657	5.177

Vol. 17 No. 2: 98-117 [2018]
doi: 10.14416/j.appsci.2018.12.001

ตารางที่ 6. (ต่อ) ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 2 ขนาดตัวอย่าง $n = 8$

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
s_{yxh}	39.748	6.666
s_{yzh}	17.666	-1.443
s_{xzh}	72.000	17.600

ตารางที่ 7. ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 2 ขนาดตัวอย่าง $n = 10$

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
N_h	35	45
n_h	6	4
$\bar{y}_h$	4.950	2.234
$\bar{x}_h$	11.714	6.778
$\bar{z}_h$	26.486	15.489
s_{yh}	1.572	0.452
s_{xh}	1.777	1.388
s_{zh}	2.867	1.286
s_{yxh}	2.434	0.439
s_{yzh}	-4.115	-0.230
s_{xzh}	-3.059	0.676

ตารางที่ 8. ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 2 ขนาดตัวอย่าง $n = 20$

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
N_h	35	45
n_h	9	11
$\bar{y}_h$	27.057	25.989
$\bar{x}_h$	66.000	76.636
$\bar{z}_h$	156.222	171.545

doi: 10.14416/j.appsci.2018.12.001

ตารางที่ 8. (ต่อ) ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 2 ขนาดตัวอย่าง n = 20

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
s_{yh}	4.668	2.641
s_{xh}	12.268	9.993
s_{zh}	8.318	6.699
s_{yxh}	46.827	21.531
s_{yzh}	-7.282	1.784
s_{xzh}	41.625	43.818

ตารางที่ 9. ค่าสถิติของประชากรที่ศึกษา จากข้อมูลชุดที่ 2 ขนาดตัวอย่าง n = 30

	ชั้นภูมิที่ 1	**ชั้นภูมิที่ 2**
N_h	35	45
n_h	16	14
$\bar{y}_h$	13.842	8.273
$\bar{x}_h$	33.486	24.133
$\bar{z}_h$	71.029	52.956
s_{yh}	2.816	1.845
s_{xh}	8.053	8.095
s_{zh}	4.269	3.569
s_{yxh}	19.990	14.169
s_{yzh}	0.633	3.902
s_{xzh}	17.691	23.416

ผลการทดลอง

ผลการวิจัยสำหรับการศึกษาครั้งนี้ คือ ตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กันสำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม และค่าความเอนเอียงและค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ จากนั้นทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอกับตัวประมาณ

 doi: 10.14416/j.appsci.2018.12.001

ของ Tailor et al. (2016) โดยใช้ค่าประสิทธิภาพสัมพัทธ์เป็นเกณฑ์ พร้อมทั้งแสดงตัวอย่างการคำนวณโดยใช้ข้อมูล 2 ชุด ซึ่งผลการวิจัยแสดงดังนี้

1. ค่าความเอนเอียงและค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ

ผู้วิจัยได้นำตัวประมาณที่นำเสนอ ที่อยู่ในรูปของ e ไปทำการคำนวณหาค่าความเอนเอียงและค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ แสดงได้ดังนี้

$$\bar{y}_{RP}^{PS(d)} - \bar{Y} = \bar{Y}\left(e_0 - g_h e_1 - g_h e_0 e_1 + g_h e_2 + g_h e_0 e_2 - g_h^2 e_1 e_2 + g_h e_2^2\right)$$

เมื่อ $e_0 = \dfrac{\sum_{h=1}^{L} W_h \bar{Y}_h e_{0h}}{\bar{Y}}, e_1 = \dfrac{\sum_{h=1}^{L} W_h \bar{X}_h e_{1h}}{\bar{X}}$ และ $e_2 = \dfrac{\sum_{h=1}^{L} W_h \bar{Z}_h e_{2h}}{\bar{Z}}$

ทำให้ได้ค่าความเอนเอียงและค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ ดังนี้

หาค่าความเอนเอียงของตัวประมาณที่นำเสนอ ซึ่งสามารถแสดงได้ดังนี้

จากสมการ $\bar{y}_{RP}^{PS(d)} - \bar{Y} = \bar{Y}\left(e_0 - g_h e_1 - g_h e_0 e_1 + g_h e_2 + g_h e_0 e_2 - g_h^2 e_1 e_2 + g_h e_2^2\right)$

หาค่าคาดหวังทั้ง 2 ข้างของสมการ จะได้

$$\begin{aligned}
E\left(\bar{y}_{RP}^{PS(d)} - \bar{Y}\right) &= \bar{Y}\left[E(e_0) - g_h E(e_1) - g_h E(e_0 e_1) + g_h E(e_2) + g_h E(e_0 e_2) - g_h^2 E(e_1 e_2) + g_h E(e_2^2)\right] \\
&= \bar{Y}\left[-g_h \frac{1}{\bar{Y}\bar{X}} \sum_{h=1}^{L} W_h^2 S_{yxh}\left(\frac{1}{nW_h} - \frac{1}{N_h}\right) + g_h \frac{1}{\bar{Y}\bar{Z}} \sum_{h=1}^{L} W_h^2 S_{yzh}\left(\frac{1}{nW_h} - \frac{1}{N_h}\right)\right. \\
&\quad \left. - g_h^2 \frac{1}{\bar{X}\bar{Z}} \sum_{h=1}^{L} W_h^2 S_{xzh}\left(\frac{1}{nW_h} - \frac{1}{N_h}\right) + g_h \frac{1}{\bar{Z}^2} \sum_{h=1}^{L} W_h^2 S_{zh}^2\left(\frac{1}{nW_h} - \frac{1}{N_h}\right)\right] \\
&= \bar{Y}\sum_{h=1}^{L} W_h^2\left(\frac{1}{nW_h} - \frac{1}{N_h}\right)\left[-g_h \frac{1}{\bar{Y}\bar{X}} S_{yxh} + g_h \frac{1}{\bar{Y}\bar{Z}} S_{yzh} - g_h^2 \frac{1}{\bar{X}\bar{Z}} S_{xzh} + g_h \frac{1}{\bar{Z}^2} S_{zh}^2\right]
\end{aligned}$$

จาก $W_h = \dfrac{N_h}{N}$

$$\begin{aligned}
E\left(\bar{y}_{RP}^{PS(d)} - \bar{Y}\right) &= \bar{Y}\sum_{h=1}^{L} W_h^2\left(\frac{1}{W_h}\right)\left(\frac{1}{n} - \frac{1}{N}\right)\left[-g_h \frac{1}{\bar{Y}\bar{X}} S_{yxh} + g_h \frac{1}{\bar{Y}\bar{Z}} S_{yzh} - g_h^2 \frac{1}{\bar{X}\bar{Z}} S_{xzh} + g_h \frac{1}{\bar{Z}^2} S_{zh}^2\right] \\
&= \bar{Y}\sum_{h=1}^{L} W_h\left(\frac{1}{n} - \frac{1}{N}\right)\left[-g_h \frac{1}{\bar{Y}\bar{X}} S_{yxh} + g_h \frac{1}{\bar{Y}\bar{Z}} S_{yzh} - g_h^2 \frac{1}{\bar{X}\bar{Z}} S_{xzh} + g_h \frac{1}{\bar{Z}^2} S_{zh}^2\right] \\
&= \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h\left[-g_h \frac{\bar{Y}}{\bar{Y}\bar{X}} S_{yxh} + g_h \frac{\bar{Y}}{\bar{Y}\bar{Z}} S_{yzh} - g_h^2 \frac{\bar{Y}}{\bar{X}\bar{Z}} S_{xzh} + g_h \frac{\bar{Y}}{\bar{Z}^2} S_{zh}^2\right]
\end{aligned}$$

จาก $R_1 = \dfrac{\bar{Y}}{\bar{X}}$ และ $R_2 = \dfrac{\bar{Y}}{\bar{Z}}$

$$E\left(\bar{y}_{RP}^{PS(d)} - \bar{Y}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L} W_h\left[-g_h \frac{1}{\bar{X}} S_{yxh} + g_h \frac{1}{\bar{Z}} S_{yzh} - g_h^2 \frac{R_1}{\bar{Z}} S_{xzh} + g_h \frac{R_2}{\bar{Z}} S_{zh}^2\right]$$

จาก $Bias\left(\bar{y}_{RP}^{PS(d)}\right) = E\left(\bar{y}_{RP}^{PS(d)} - \bar{Y}\right)$

ดังนั้น $Bias\left(\bar{y}_{RP}^{PS(d)}\right)$ คือ

$$Bias\left(\bar{y}_{RP}^{PS(d)}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L}W_h\left[-g_h\frac{1}{\bar{X}}S_{yxh}+g_h\frac{1}{\bar{Z}}S_{yzh}-g_h^2\frac{R_1}{\bar{Z}}S_{xzh}+g_h\frac{R_2}{\bar{Z}}S_{zh}^2\right] \quad (2)$$

เมื่อ $g_h=\dfrac{n_h}{N_h-n_h}, R_1=\dfrac{\bar{Y}}{\bar{X}}$ และ $R_2=\dfrac{\bar{Y}}{\bar{Z}}$

หาค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่นำเสนอ ซึ่งสามารถแสดงได้ดังนี้

จากสมการ $\bar{y}_{RP}^{PS(d)}-\bar{Y}\cong\bar{Y}\left(e_0-g_he_1-g_he_0e_1+g_he_2+g_he_0e_2-g_h^2e_1e_2+g_he_2^2\right)$

เมื่อยกกำลังสองทั้ง 2 ข้าง และไม่พิจารณาพจน์ของ e ที่มีเลขชี้กำลังมากกว่าสอง จะได้

$\left(\bar{y}_{RP}^{PS(d)}-\bar{Y}\right)^2=\left[\bar{Y}\left(e_0-g_he_1+g_he_2\right)\right]^2$ หาค่าคาดหวังทั้ง 2 ข้าง

$$\begin{aligned}
E\left(\bar{y}_{RP}^{PS(d)}-\bar{Y}\right)^2 &= E\left[\bar{Y}\left(e_0-g_he_1+g_he_2\right)\right]^2\\
&=\bar{Y}^2\left[E\left(e_0^2\right)+g_h^2E\left(e_1^2\right)+g_h^2E\left(e_2^2\right)-2g_hE\left(e_0e_1\right)+2g_hE\left(e_0e_2\right)-2g_h^2E\left(e_1e_2\right)\right]\\
&=\bar{Y}^2\left[\frac{1}{\bar{Y}^2}\sum_{h=1}^{L}W_h^2S_{yh}^2\left(\frac{1}{nW_h}-\frac{1}{N_h}\right)+g_h^2\frac{1}{\bar{X}^2}\sum_{h=1}^{L}W_h^2S_{xh}^2\left(\frac{1}{nW_h}-\frac{1}{N_h}\right)+g_h^2\frac{1}{\bar{Z}^2}\sum_{h=1}^{L}W_h^2S_{zh}^2\left(\frac{1}{nW_h}-\frac{1}{N_h}\right)\right.\\
&\quad-2g_h\frac{1}{\bar{Y}\bar{X}}\sum_{h=1}^{L}W_h^2S_{yxh}\left(\frac{1}{nW_h}-\frac{1}{N_h}\right)+2g_h\frac{1}{\bar{Y}\bar{Z}}\sum_{h=1}^{L}W_h^2S_{yzh}\left(\frac{1}{nW_h}-\frac{1}{N_h}\right)\\
&\quad\left.-2g_h^2\frac{1}{\bar{X}\bar{Z}}\sum_{h=1}^{L}W_h^2S_{xzh}\left(\frac{1}{nW_h}-\frac{1}{N_h}\right)\right]\\
&=\bar{Y}^2\sum_{h=1}^{L}W_h^2\left(\frac{1}{nW_h}-\frac{1}{N_h}\right)\left[\frac{1}{\bar{Y}^2}S_{yh}^2+g_h^2\frac{1}{\bar{X}^2}S_{xh}^2+g_h^2\frac{1}{\bar{Z}^2}S_{zh}^2-2g_h\frac{1}{\bar{Y}\bar{X}}S_{yxh}+2g_h\frac{1}{\bar{Y}\bar{Z}}S_{yzh}-2g_h^2\frac{1}{\bar{X}\bar{Z}}S_{xzh}\right]
\end{aligned}$$

จาก $W_h=\dfrac{N_h}{N}$

$$\begin{aligned}
E\left(\bar{y}_{RP}^{PS(d)}-\bar{Y}\right)^2 &=\bar{Y}^2\sum_{h=1}^{L}W_h^2\left(\frac{1}{W_h}\right)\left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{1}{\bar{Y}^2}S_{yh}^2+g_h^2\frac{1}{\bar{X}^2}S_{xh}^2+g_h^2\frac{1}{\bar{Z}^2}S_{zh}^2-2g_h\frac{1}{\bar{Y}\bar{X}}S_{yxh}+2g_h\frac{1}{\bar{Y}\bar{Z}}S_{yzh}-2g_h^2\frac{1}{\bar{X}\bar{Z}}S_{xzh}\right]\\
&=\bar{Y}^2\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L}W_h\left[\frac{1}{\bar{Y}^2}S_{yh}^2+g_h^2\frac{1}{\bar{X}^2}S_{xh}^2+g_h^2\frac{1}{\bar{Z}^2}S_{zh}^2-2g_h\frac{1}{\bar{Y}\bar{X}}S_{yxh}+2g_h\frac{1}{\bar{Y}\bar{Z}}S_{yzh}-2g_h^2\frac{1}{\bar{X}\bar{Z}}S_{xzh}\right]\\
&=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L}W_h\left[\frac{\bar{Y}^2}{\bar{Y}^2}S_{yh}^2+g_h^2\frac{\bar{Y}^2}{\bar{X}^2}S_{xh}^2+g_h^2\frac{\bar{Y}^2}{\bar{Z}^2}S_{zh}^2-2g_h\frac{\bar{Y}^2}{\bar{Y}\bar{X}}S_{yxh}+2g_h\frac{\bar{Y}^2}{\bar{Y}\bar{Z}}S_{yzh}-2g_h^2\frac{\bar{Y}^2}{\bar{X}\bar{Z}}S_{xzh}\right]
\end{aligned}$$

จาก $R_1=\dfrac{\bar{Y}}{\bar{X}}$ และ $R_2=\dfrac{\bar{Y}}{\bar{Z}}$

$$E\left(\bar{y}_{RP}^{PS(d)}-\bar{Y}\right)^2=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L}W_h\left[S_{yh}^2+g_h^2R_1^2S_{xh}^2+g_h^2R_2^2S_{zh}^2-2g_hR_1S_{yxh}+2g_hR_2S_{yzh}-2g_h^2R_1R_2S_{xzh}\right]$$

จาก $MSE\left(\bar{y}_{RP}^{PS(d)}\right)=E\left(\bar{y}_{RP}^{PS(d)}-\bar{Y}\right)^2$

ดังนั้น $MSE\left(\bar{y}_{RP}^{PS(d)}\right)$ คือ

$$MSE\left(\bar{y}_{RP}^{PS(d)}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L}W_h\left[S_{yh}^2+g_h^2R_1^2S_{xh}^2+g_h^2R_2^2S_{zh}^2-2g_hR_1S_{yxh}+2g_hR_2S_{yzh}-2g_h^2R_1R_2S_{xzh}\right] \quad (3)$$

เมื่อ $g_h = \dfrac{n_h}{N_h - n_h}, R_1 = \dfrac{\bar{Y}}{\bar{X}}$ และ $R_2 = \dfrac{\bar{Y}}{\bar{Z}}$

2. เงื่อนไขที่ทำให้ประสิทธิภาพของตัวประมาณที่นำเสนอสูงกว่าตัวประมาณของ Tailor et al. (2016) พิจารณาเงื่อนไข

$$MSE\left(\bar{y}_{RP}^{PS(d)}\right) < MSE\left(\bar{y}_{RP}^{PS}\right)$$

$$MSE\left(\bar{y}_{RP}^{PS(d)}\right) - MSE\left(\bar{y}_{RP}^{PS}\right) < 0$$

$$\begin{bmatrix} \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L} W_h\left[S_{yh}^2 + g_h^2R_1^2S_{xh}^2 + g_h^2R_2^2S_{zh}^2 - 2g_hR_1S_{yxh} + 2g_hR_2S_{yzh} - 2g_h^2R_1R_2S_{xzh}\right] - \\ \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{h=1}^{L} W_h\left[S_{yh}^2 + R_1^2S_{xh}^2 + R_2^2S_{zh}^2 - 2\left(R_1S_{yxh} - R_2S_{yzh} + R_1R_2S_{xzh}\right)\right] \end{bmatrix} < 0$$

$$\sum_{h=1}^{L} W_h \begin{bmatrix} g_h^2R_1^2S_{xh}^2 + g_h^2R_2^2S_{zh}^2 - 2g_hR_1S_{yxh} + 2g_hR_2S_{yzh} - 2g_h^2R_1R_2S_{xzh} \\ -R_1^2S_{xh}^2 - R_2^2S_{zh}^2 + 2R_1S_{yxh} - 2R_2S_{yzh} + 2R_1R_2S_{xzh} \end{bmatrix} < 0 \qquad (4)$$

จากเงื่อนไขข้างต้นจะเห็นว่าตัวประมาณ $\bar{y}_{RP}^{PS(d)}$ จะมีประสิทธิภาพสูงกว่าตัวประมาณ $\bar{y}_{RP}^{PS}$ ก็ต่อเมื่อ

$$\sum_{h=1}^{L} W_h \begin{bmatrix} g_h^2R_1^2S_{xh}^2 + g_h^2R_2^2S_{zh}^2 - 2g_hR_1S_{yxh} + 2g_hR_2S_{yzh} - 2g_h^2R_1R_2S_{xzh} \\ -R_1^2S_{xh}^2 - R_2^2S_{zh}^2 + 2R_1S_{yxh} - 2R_2S_{yzh} + 2R_1R_2S_{xzh} \end{bmatrix} < 0$$

3. ตัวอย่างการคำนวณ

ในการคำนวณเชิงตัวเลขเพื่อทำการสนับสนุนผลของการศึกษาเชิงทฤษฎี ผู้วิจัยใช้ข้อมูลประชากรที่ศึกษา 2 ชุด ได้แก่ ข้อมูลผลผลิตการเกษตรประเทศอินเดีย ในปี 2014-2015 และข้อมูลดัชนีมวลกาย (body mass index: BMI) ของคนไข้ที่เข้ามาใช้บริการที่โรงพยาบาลหัวใจและศูนย์วิจัย Rajaei ณ เมืองเตหะราน ประเทศอิหร่าน

คำนวณเชิงตัวเลขภายใต้ประชากรที่ศึกษา เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม $\left(\bar{y}_{RP}^{PS(d)}\right)$ กับตัวประมาณแบบอัตราส่วนร่วมกับผลคูณที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบบางชั้นภูมิภายหลังการสุ่ม $\left(\bar{y}_{RP}^{PS}\right)$ ของ Tailor et al. (2016) โดยพิจารณาค่าคลาดเคลื่อนกำลังสองเฉลี่ย ดังแสดงในตารางที่ 10 และ 11 พร้อมทั้งเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอกับตัวประมาณของ Tailor et al. (2016) จากค่าประสิทธิภาพสัมพัทธ์

ตารางที่ 10. ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าประสิทธิภาพสัมพัทธ์ (RE) ของตัวประมาณภายใต้ประชากรที่ศึกษา เมื่อคำนวณจาก ข้อมูลชุดที่ 1

ขนาดตัวอย่าง	ค่า Bias		ค่า MSE		ค่า
	$\bar{y}_{RP}^{PS}$	$\bar{y}_{RP}^{PS(d)}$	$\bar{y}_{RP}^{PS}$	$\bar{y}_{RP}^{PS(d)}$	$RE\left(\bar{y}_{RP}^{PS(d)}, \bar{y}_{RP}^{PS}\right)$
n = 8	-0.069	-2.224	0.603	0.579	1.041
n = 10	0.264	-5.501	7.677	4.336	1.771
n = 20	1.359	-6.578	129.106	17.285	7.470
n = 30	0.108	-4.347	891.758	1.811	492.358

ตารางที่ 11. ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าประสิทธิภาพสัมพัทธ์ (RE) ของตัวประมาณภายใต้ประชากรที่ศึกษา เมื่อคำนวณจาก ข้อมูลชุดที่ 2

ขนาดตัวอย่าง	ค่า Bias		ค่า MSE		ค่า
	$\bar{y}_{RP}^{PS}$	$\bar{y}_{RP}^{PS(d)}$	$\bar{y}_{RP}^{PS}$	$\bar{y}_{RP}^{PS(d)}$	$RE\left(\bar{y}_{RP}^{PS(d)}, \bar{y}_{RP}^{PS}\right)$
n = 8	0.012	-0.007	0.546	0.538	1.015
n = 10	-0.010	-0.007	0.106	0.079	1.341
n = 20	0.002	-0.016	0.494	0.247	2.002
n = 30	0.004	-0.021	0.093	0.013	7.413

จากตารางที่ 10 และ 11 พบว่า ค่าความเอนเอียงของตัวประมาณที่นำเสนอทั้งหมดมีค่าเป็นลบ และค่าความเอนเอียงของตัวประมาณของ Tailor et al. (2016) ส่วนมากมีค่าเป็นบวก และ ตัวประมาณที่นำเสนอมีประสิทธิภาพสูงกว่าตัวประมาณของ Tailor et al. (2016) เนื่องจากให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) น้อยกว่าในทุกระดับขนาดตัวอย่างที่กำหนด และค่าประสิทธิภาพสัมพัทธ์ (RE) มีค่ามากกว่า 1 ในทุกระดับขนาดตัวอย่างที่กำหนด

วิจารณ์และสรุปผลการทดลอง

การจากศึกษา ตัวประมาณของ Singh et al. (2005) ได้เสนอตัวประมาณแบบอัตราส่วนร่วมกับผลคูณแบบคู่กันที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบง่าย และตัวประมาณของ Tailor et al. (2016)

ได้เสนอตัวประมาณแบบอัตราส่วนร่วมกับผลคูณที่ใช้ในการประมาณค่าเฉลี่ยประชากรในการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม ผู้วิจัยได้ทำการปรับปรุงตัวประมาณค่าเฉลี่ยประชากร ให้มีประสิทธิภาพมากยิ่งขึ้น โดยการแปลงค่า x_{hi} และ z_{hi} ให้อยู่ในรูปของ $x_{hi}^{*}=\frac{\bar{X}_h N_h - x_{hi} n_h}{N_h - n_h}$ และ $z_{hi}^{*}=\frac{\bar{Z}_h N_h - z_{hi} n_h}{N_h - n_h}$, $h=1,2,3,...,L$ ภายใต้การเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม ดังนั้นผู้วิจัยจึงได้เสนอตัวประมาณค่าเฉลี่ยประชากร โดยใช้อัตราส่วนร่วมกับผลคูณแบบคู่กัน สำหรับการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม

เมื่อพิจารณาค่าความเอนเอียง จากข้อมูลทั้ง 2 ชุด พบว่า ค่าความเอนเอียงของตัวประมาณที่นำเสนอทั้งหมดมีค่าเป็นลบ และค่าความเอนเอียงของตัวประมาณของ Tailor et al. (2016) ส่วนมากมีค่าเป็นบวก จึงส่งผลให้ค่าประมาณที่คำนวณจากประมาณที่นำเสนอมีค่าต่ำกว่าค่าจริง ดังนั้น ผู้วิจัยจึงทำการเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอกับตัวประมาณของ Tailor et al. (2016) โดยพิจารณาจากค่าคลาดเคลื่อนกำลังสองเฉลี่ยและค่าประสิทธิภาพสัมพัทธ์ จากผลการคำนวณค่าเชิงตัวจากข้อมูลทั้ง 2 ชุด ดังตารางที่ 1 และ 2 ตามลำดับ พบว่าตัวประมาณที่นำเสนอมีประสิทธิภาพสูงกว่าตัวประมาณของ Tailor et al. (2016)

เมื่อพิจารณาจากการแบ่งขนาดตัวอย่างออกเป็น 4 ระดับ คือ ตัวอย่างขนาด $n=8, n=10, n=20$ และ $n=30$ จากผลการคำนวณค่าเชิงตัวเลขจากข้อมูลทั้ง 2 ชุด พบว่า ตัวประมาณที่นำเสนอมีประสิทธิภาพสูงขึ้น เมื่อขนาดตัวอย่างเพิ่มขึ้น จะสังเกตได้ว่า เมื่อตัวอย่างมีขนาดเล็กค่าคลาดเคลื่อนกำลังสองเฉลี่ยจะมีค่ามาก และค่าคลาดเคลื่อนกำลังสองเฉลี่ยลดลงเมื่อตัวอย่างมีขนาดใหญ่ขึ้น

แต่เมื่อพิจารณาที่ขนาดตัวอย่าง $n=30$ จากข้อมูลชุดที่ 1 จะเห็นได้ว่า ค่าประสิทธิภาพสัมพัทธ์เพิ่มขึ้นเป็น 492.358 เท่า ในขณะที่ข้อมูลชุดที่ 2 ค่าประสิทธิภาพสัมพัทธ์เพิ่มขึ้น 7.413 เท่า ซึ่งจะเห็นได้ว่าลักษณะของประชากรที่ศึกษาในข้อมูลชุดที่ 1 และ 2 (จากตารางที่ 1 ในหน้า 9) มีความแตกต่างกัน เมื่อพิจารณาจากค่าส่วนเบี่ยงเบนมาตรฐานหรือลักษณะของประชากรที่ศึกษาจากข้อมูลทั้ง 2 ชุด โดยข้อมูลชุดที่ 1 จะมีค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปรที่สนใจศึกษาสูง โดยเฉพาะตัวแปรช่วย x และ z ส่วนในข้อมูล ชุดที่ 2 ค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปรที่สนใจศึกษา มีค่าไม่มากนัก ดังนั้น เมื่อพิจารณาจากลักษณะของประชากรที่ศึกษา จากที่กล่าวมาข้างต้นทำให้เห็นถึงความแตกต่างของค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณที่นำเสนอ จากการเลือกตัวอย่างสุ่มแบบแบ่งชั้นภูมิภายหลังการสุ่ม ทำให้ตัวประมาณที่นำเสนอมีประสิทธิภาพเพิ่มมากขึ้น

เอกสารอ้างอิง

Bandyopadhyay, S. (1980). Improved ratio and product estimators. *Sankhyā Series C, 42*(2), 45-49.

Singh, H. P., Singh, R., Espejo, M. R., Pineda, M. D. & Nadarajah, S. (2005). On the efficiency of a dual to ratio-cum-product estimator in sample surveys. *Mathematical Proceedings of the Royal Irish Academy, 105A*(2), 51-56.

Vol. 17 No. 2: 98-117 [2018]
doi: 10.14416/j.appsci.2018.12.001

Singh, M. P. (1967). Ratio-cum-product method of estimation. *Metrika*, *12(*1), 34-42. https://doi.org/10.1007/BF02613481

Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. *Biometrika, 67*(1), 199-204. https://doi.org/10.1093/biomet/67.1.199

Suwattee, P. (2011). *Sample survey: sampling designs and analysis* (1st ed.). Bangkok, Thailand: NIDACIP.

Tailor, R., Chouhan, S., Tailor, R. & Garg, N. (2012). A ratio-cum product estimator of population mean in stratified random sampling using two auxiliary variables. *Statistica, 72*(3), 287-297. doi: 10.6092/issn.1973-2201/3648

Tailor, R., Lone, H. A., Pandey, R. & Kumar, M. (2016). Ratio-cum-product type estimator of finite population mean in case of post stratification. *Mathematical Sciences Letters*, *5*(1), 103-106. http://dx.doi.org/10.18576/msl/050115

Ungpansattawong, S. (2011). *Sample survey designs* (1st ed.). Khon Kaen, Thailand: Khon Kaen University Printing House.