บทความวิจัย

การประมาณค่าเฉลี่ยประชากรในการสำรวจตัวอย่างเมื่อมีข้อมูลสูญหาย

พัชรีพร สนรักษา พัชรี วงษ์เกษม และ คณินทร์ ธีรภาพโอพาร*

บทคัดย่อ

วัตถุประสงค์ของการศึกษาในครั้งนี้ คือ นำเสนอตัวประมาณค่าเฉลี่ยประชากรในการสำรวจตัวอย่างเมื่อมีข้อมูลสูญหาย โดยการประยุกต์ใช้ตัวประมาณค่าเฉลี่ย T_{d1} ของ Singh และคณะ [2] เมื่อเปรียบเทียบค่า the root mean square error (RMSE) ของตัวประมาณค่าที่นำเสนอและตัวประมาณค่า T_{d1}, T_{d2}, ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยวิธีค่าเฉลี่ย (mean imputation), ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยค่าอัตราส่วน (ratio imputation), ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยค่าคอมโพรไมซ (compromised imputation) และตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก (weighted nearest neighbor and regression imputation) (ภายใต้เงื่อนไขที่กำหนด) เราพบว่าตัวประมาณค่าที่นำเสนอมีประสิทธิภาพดีกว่าตัวประมาณค่า T_{d1}, T_{d2}, ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยวิธีค่าเฉลี่ย (mean imputation), ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยค่าอัตราส่วน (ratio imputation), ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยค่าคอมโพรไมซ (compromised imputation) และตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก (weighted nearest neighbor and regression imputation)

คำสำคัญ: ตัวประมาณค่าเฉลี่ยประชากร ข้อมูลสูญหาย

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ชลบุรี 20131

*ผู้นิพนธ์ประสานงาน, email : kanint@buu.ac.th

Estimation of Population Mean in Sample Survey with Missing Data

Phatchareephorn Sonraksa, Patcharee Wongkasem and Kanint Teerapabolarn*

ABSTRACT

The objective of this study is to propose an estimation of population mean in sample survey with missing data, which is obtained by applying the method of T_{d1} estimator of Singh, et al. [2]. By comparing the root mean square error (RMSE) of the proposed estimator and T_{d1} estimator, T_{d2} estimator, estimators based on mean imputation, ratio imputation, compromised imputation and weighted nearest neighbor and regression imputation, under the specific condition, it indicates that the proposed estimator is more efficient than T_{d1} estimator, T_{d2} estimator, mean imputation, ratio imputation, compromised imputation and weighted nearest neighbor and regression imputation.

Keywords: estimation of population mean, missing data

Department of Mathematics, Faculty of Science, Burapha University

*Corresponding author, email: kanint@buu.ac.th

บทนำ

ปัญหาหลักปัญหาหนึ่งซึ่งพบโดยบังเอิญจากการเก็บรวบรวมข้อมูลจากแบบสอบถามในการสำรวจตัวอย่าง คือ การไม่ตอบ หรือเรียกว่า ข้อมูลสูญหาย (missing data) ข้อมูลสูญหายจากการสำรวจตัวอย่างโดยทั่วไปมี 2 ประเภท คือ unit nonresponse และ item nonresponse โดย Kalton และ Kasprzyk [1] ได้ให้นิยามของ unit nonresponse และ item nonresponse ไว้ดังนี้ unit nonresponse คือ การไม่ตอบสำหรับหน่วยตัวอย่างบางหน่วย ซึ่งอาจเป็นผลสืบเนื่องมาจากการไม่เข้าใจความหมายของคำต่างๆ ที่ใช้ในแบบสอบถาม สำหรับ item nonresponse คือ การสูญหายของข้อมูลที่เกิดจากการไม่ตอบเฉพาะบางคำถามหรือบางตัวแปร ซึ่งอาจเกิดจากการที่ออกแบบคำถามในแบบสอบถามไม่ครอบคลุม ทำให้ผู้ตอบแบบสอบถามไม่สามารถตอบคำถามบางคำถามได้ หรืออาจเกิดความผิดพลาดในการบันทึกข้อมูลทำให้ต้องตัดข้อมูลบางตัวออกไป ส่วนใหญ่การแก้ไขปัญหากรณีข้อมูลสูญหายที่เกิดจากการไม่ตอบเฉพาะบางคำถามหรือบางตัวแปร ทางสถิติมักใช้วิธีการประมาณค่าข้อมูลสูญหาย (imputation) แล้วนำค่าประมาณที่ได้ไปแทนข้อมูลสูญหายเพื่อให้ได้ชุดข้อมูลสมบูรณ์ ซึ่งในงานวิจัยนี้สนใจศึกษาข้อมูลสูญหายที่เกิดจากการไม่ตอบเฉพาะบางคำถาม หรือบางตัวแปร (item nonresponse)

ข้อมูลสูญหายเป็นสถานการณ์ที่เกิดขึ้นจริงและไม่สามารถหลีกเลี่ยงได้ ในงานวิจัยการตัดหน่วยตัวอย่างที่ไม่ตอบออกจากการวิเคราะห์ข้อมูลเป็นสิ่งที่อาจกระทำได้ถ้าข้อมูลนั้นหามาได้โดยง่าย มีค่าใช้จ่ายน้อย และมีการสำรองข้อมูลจำนวนมากๆ แต่ในความเป็นจริงแล้วอาจเสียเวลา ค่าใช้จ่าย กำลังแรงงานและทรัพยากรอื่นๆ เป็นจำนวนมาก การจัดการกับข้อมูลสูญหายนั้นมีหลายวิธีให้เลือกใช้ การพิจารณาเลือกใช้วิธีการใดนั้นขึ้นอยู่กับลักษณะของข้อมูลสูญหายที่เกิดขึ้น ดังนั้นการแก้ไขปัญหาที่เกิดจากข้อมูลสูญหายเป็นสิ่งสำคัญ และต้องเลือกใช้วิธีการแก้ไขปัญหาที่เหมาะสม หรือค้นหาวิธีใหม่ที่เหมาะสมกว่าการประมาณค่าข้อมูลสูญหาย มีผู้ที่เสนอวิธีการประมาณค่าข้อมูลสูญหายหลายๆ วิธี ได้แก่

1. วิธีการประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย Mean imputation (MI)
2. วิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอย Regression imputation (RI)
3. วิธีการประมาณข้อมูลสูญหายด้วยค่าใกล้สุด Nearest neighbor imputation (NNI)
4. วิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก Weighted nearest neighbor and regression imputation (WNR)
5. วิธีการประมาณข้อมูลสูญหายด้วยค่าอัตราส่วน Ratio imputation (RAT)
6. วิธีการประมาณข้อมูลสูญหายด้วยค่าฮอทเดค Hot deck imputation (HDI)
7. วิธีการประมาณข้อมูลสูญหายด้วยค่าโคลเดค Cold deck imputation (CDI)
8. วิธีการประมาณข้อมูลสูญหายด้วยค่าคอมโพรไมซ Compromised imputation (COMP)
9. วิธีการประมาณค่าข้อมูลสูญหายของ Singh และคณะ [2]

Singh และคณะ [2] ได้ศึกษาวิจัยเรื่อง Estimation of population mean using imputation techniques in sample surveys โดยเสนอตัวประมาณค่าข้อมูลสูญหาย 2 วิธี วิธีแรก คือ ข้อมูลตัวแปรช่วย x เป็นข้อมูลที่สมบูรณ์ทั้งหมด n หน่วย การแทนค่าข้อมูลสูญหายโดยใช้เฉพาะข้อมูลของตัวแปรช่วย x เป็นข้อมูลที่สมบูรณ์ทั้งหมด r หน่วย ของหน่วยตัวอย่างที่ข้อมูลของตัวแปรที่สนใจ y ที่เก็บได้ ซึ่งมีทั้งหมด r หน่วย ส่วนที่เหลือ n-r หน่วย เป็นข้อมูลสูญหาย ซึ่ง ($r < n$) โดยกำหนดให้

$$y_{\cdot i} = \begin{cases} y_i & if\ i \in R \\ \dfrac{1}{(n-r)}\bar{y}_r\left[\dfrac{n\left\{\left(A+C\right)\bar{X}+fB\bar{x}_r\right\}}{\left\{\left(A+fB\right)\bar{X}+C\bar{x}_r\right\}}-r\right] & if\ i \in R^c \end{cases} \tag{1}$$

เมื่อ $A = (d-1)(d-2)$, $B = (d-1)(d-4)$, $C = (d-2)(d-3)(d-4)$ และ $f = \dfrac{n}{N}$
ซึ่ง d เป็นจำนวนจริงบวก

ตัวประมาณค่าเฉลี่ยประชากรที่ใช้ข้อมูลที่ประมาณค่าข้อมูลสูญหายด้วยสมการ (1) อยู่ในรูป

$$T_{d1} = \bar{y}_r\left[\frac{(A+C)\bar{X}+fB\bar{x}_r}{(A+fB)\bar{X}+C\bar{x}_r}\right] \tag{2}$$

เป็นตัวประมาณค่าที่เอนเอียง โดยมีค่าเอนเอียง (Bias) ดังนี้

$$b(T_{d1}) = \left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}\varphi\left[\rho_{YX}C_YC_X-\varphi_2C_X^2\right] \tag{3}$$

เมื่อ $\varphi = \varphi_1 - \varphi_2$, $\varphi_1 = \dfrac{fB}{A+fB+C}$, และ $\varphi_2 = \dfrac{C}{A+fB+C}$

โดยที่ C_Y คือ สัมประสิทธิ์การแปรผันของข้อมูลของตัวแปรที่สนใจ Y
C_X คือ สัมประสิทธิ์การแปรผันของข้อมูลของตัวแปรช่วย X
และ ρ_{YX} คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรที่สนใจ Y และตัวแปรช่วย X

และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE(T_{d1}) = \left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}^2\left[C_Y^2+\varphi^2C_X^2+2\varphi\rho_{YX}C_YC_X\right] \tag{4}$$

ซึ่งพบว่า $MSE(T_{d1})$ มีค่าต่ำสุดก็ต่อเมื่อ

$$\varphi = -\rho_{YX}\frac{C_Y}{C_X} \tag{5}$$

จาก (5) เมื่อทราบค่า ρ_{YX}, C_Y และ C_X สามารถหาค่า d ที่เหมาะสมได้ และจะได้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยที่เหมาะสม ดังนี้

$$MSE(T_{d1})_{opt} = \left(\frac{1}{r} - \frac{1}{N}\right)(1 - \rho_{YX}^2)S_Y^2 \quad (6)$$

เมื่อ S_Y^2 คือ ความแปรปรวนของข้อมูลของตัวแปรที่สนใจ Y

สำหรับวิธีที่สองที่ Singh และคณะ [2] เสนอการแทนค่าข้อมูลสูญหายโดยใช้ข้อมูลของตัวแปรช่วย x จากทั้ง n หน่วย โดยกำหนดให้

$$y_{\cdot i} = \begin{cases} y_i & if\ i \in R \\ \frac{1}{(n-r)}\bar{y}_r\left[\frac{n\left\{(A+C)\bar{X} + fB\bar{x}_n\right\}}{\left\{(A+fB)\bar{X} + C\bar{x}_n\right\}} - r\right] & if\ i \in R^c \end{cases} \quad (7)$$

เมื่อ $A = (d-1)(d-2)$, $B = (d-1)(d-4)$, $C = (d-2)(d-3)(d-4)$ และ $f = \frac{n}{N}$
ซึ่ง d เป็นจำนวนจริงบวก

ตัวประมาณค่าเฉลี่ยประชากรที่ใช้ข้อมูลที่ประมาณค่าข้อมูลสูญหายด้วยสมการ (7) อยู่ในรูป

$$T_{d2} = \bar{y}_r\left[\frac{(A+C)\bar{X} + fB\bar{x}_n}{(A+fB)\bar{X} + C\bar{x}_n}\right] \quad (8)$$

เป็นตัวประมาณค่าที่เอนเอียง โดยมีค่าเอนเอียง (Bias) ดังนี้

$$b(T_{d2}) = \left(\frac{1}{n} - \frac{1}{N}\right)\bar{Y}\varphi\left[\rho_{YX}C_Y C_X - \varphi_2 C_X^2\right] \quad (9)$$

มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE(T_{d2}) = \bar{Y}^2\left[\left(\frac{1}{r} - \frac{1}{N}\right)C_Y^2 + \left(\frac{1}{n} - \frac{1}{N}\right)\left[\varphi^2 C_X^2 + 2\varphi\rho_{YX}C_Y C_X\right]\right] \quad (10)$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยที่เหมาะสม ดังนี้

$$MSE(T_{d2})_{opt} = \left[\left(\frac{1}{r} - \frac{1}{N}\right) - \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{YX}^2\right]S_Y^2 \quad (11)$$

ดังนั้น ในการศึกษาครั้งนี้เราจึงสนใจศึกษาตัวประมาณค่าเฉลี่ยประชากร โดยประยุกต์ใช้ตัวประมาณค่า T_{d1} ของ Singh และคณะ [2]

วัตถุประสงค์ของการวิจัย

1. เพื่อเสนอตัวประมาณค่าเฉลี่ยประชากรในการสำรวจตัวอย่างเมื่อมีข้อมูลสูญหายโดยการปรับตัวประมาณค่า T_{d1}

2. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าที่นำเสนอกับตัวประมาณค่า T_{d1}, T_{d2}, ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยวิธีการประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย การประมาณข้อมูลสูญหายด้วยค่าอัตราส่วน การประมาณข้อมูลสูญหายด้วยค่าคอมโพรไมซ และวิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก

วิธีการดำเนินการวิจัย

วิธีการดำเนินการวิจัยมีขั้นตอนดังนี้

1. ศึกษางานวิจัยดังที่ได้กล่าวมาข้างต้น โดยเรามีความสนใจที่จะปรับตัวประมาณค่า T_{d1} ของ Singh และคณะ [2] ให้เป็นตัวประมาณค่าเฉลี่ยประชากรใหม่ ให้ T_{new} แทนตัวประมาณค่าเฉลี่ยประชากรใหม่ ซึ่งกำหนดดังนี้

$$T_{new} = \lambda T_{d1} \tag{12}$$

เมื่อ T_{d1} คือ ตัวประมาณค่าเฉลี่ยประชากรของ Singh และคณะ [2] และ λ คือ ค่าคงตัวที่ทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ T_{new} มีค่าต่ำสุด

2. หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่าที่นำเสนอ

3. เปรียบเทียบประสิทธิภาพของตัวประมาณค่าที่นำเสนอและตัวประมาณค่า T_{d1}, T_{d2}, ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยวิธีการประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย การประมาณข้อมูลสูญหายด้วยค่าอัตราส่วน การประมาณข้อมูลสูญหายด้วยค่าคอมโพรไมซและวิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก โดยใช้ค่า the root mean square error (RMSE) ซึ่งเรากล่าวว่า ตัวประมาณค่า $\hat{\theta}_1$ เป็นตัวประมาณค่าที่มีประสิทธิภาพสูงกว่า $\hat{\theta}_2$ ก็ต่อเมื่อ $RMSE(\hat{\theta}_1) < RMSE(\hat{\theta}_2)$

4. แสดงตัวอย่างเชิงตัวเลขเพื่อสนับสนุนผลการศึกษาในเชิงทฤษฎีโดยใช้เครื่องคอมพิวเตอร์เขียนด้วยโปรแกรม R ซึ่งดำเนินการ ดังนี้

4.1 ข้อมูลจากราคาทองคำย้อนหลังจากสมาคมค้าทองคำ (Gold Traders Association) ตั้งแต่วันที่ 28 มีนาคม 2553 จนถึงวันที่ 1 ตุลาคม 2554 ขนาดประชากร N เท่ากับ 1,219 ซึ่งประกอบด้วยตัวแปรช่วย X คือ ราคาทองต่างประเทศ ตัวแปรที่สนใจ Y คือ ราคาทองคำแท่ง 96.5% ณ ระดับค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) เท่ากับ 0.97 และข้อมูลจากราคาข้าวหอมมะลิย้อนหลังจากกรมการค้าภายใน กระทรวงพาณิชย์ ตั้งแต่วันที่ 1 มกราคม 2550 จนถึงวันที่ 1 พฤศจิกายน 2554 ขนาดประชากร N เท่ากับ 1,147 ประกอบด้วยตัวแปรช่วย X คือ ราคาข้าวเปลือกหอมมะลิ (ชนิดสีได้ต้นข้าว 32-46 กรัม) น้ำหนัก 1000 กิโลกรัม ตัวแปรที่สนใจ Y คือ ราคาข้าวหอมมะลิ 100% ชั้น 1 น้ำหนัก 100 กิโลกรัม ณ ระดับค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) เท่ากับ 0.91

4.2 ทดสอบระดับค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรที่สนใจ Y และตัวแปรช่วย X ว่ามีระดับความสัมพันธ์ตามที่ระบุไว้หรือไม่

4.3 ชักตัวอย่าง (n) ขนาด = 1% และ 2% ของขนาดประชากร (N) ด้วย วิธีการชักตัวอย่างแบบง่ายชนิดแทนที่ (simple random sampling with replacement หรือ SRS) Chaimongkol และ Suwattee [3]

4.4 สุ่มตำแหน่งการสูญหายของข้อมูลให้กับตัวแปรที่สนใจ Y โดยสุ่มขนาด 5%, 10%, 15% และ 20% ของขนาดตัวอย่าง (n) Little และ Rubin [4]

4.5 หาค่า d ที่เหมาะสมโดยใช้เครื่องคอมพิวเตอร์เขียนด้วยโปรแกรม maple จาก

$$-\rho_{yx}\frac{C_y}{C_x}=\frac{\frac{n}{N}\left[(d-1)(d-4)\right]-\left[(d-2)(d-3)(d-4)\right]}{(d-1)(d-4)+\frac{n}{N}\left[(d-1)(d-4)\right]+\left[(d-2)(d-3)(d-4)\right]} \tag{13}$$

แล้วนำค่า d ที่ได้ทั้ง 3 ค่า แทนลงในค่า $A = (d-1)(d-2)$.

$B = (d-1)(d-4)$ และ $C = (d-2)(d-3)(d-4)$

4.6 นำค่า A, B และ C ที่ได้แทนค่าลงในตัวประมาณค่า T_{d1}, T_{d2} และตัวประมาณค่า T_{new} ประมาณค่าข้อมูลสูญหายด้วยตัวประมาณค่า T_{d1}, T_{d2} และตัวประมาณค่า T_{new}

4.7 เปรียบเทียบประสิทธิภาพของค่า (the root mean square error หรือ RMSE) ของตัวประมาณค่าที่นำเสนอ และตัวประมาณค่า T_{d1}, T_{d2}, ตัวประมาณค่า เมื่อประมาณค่าข้อมูลสูญหายด้วยวิธีการประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย, การประมาณข้อมูลสูญหายด้วยค่าอัตราส่วน, การประมาณข้อมูลสูญหายด้วยค่าคอมโพรไมซ และวิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก

ผลการวิจัย

ผลการวิจัยที่เราต้องการสำหรับการศึกษาครั้งนี้ คือ ตัวประมาณค่าที่ประยุกต์มาจากแนวคิดของ Singh และคณะ [2] เงื่อนไขที่ทำให้ประสิทธิภาพของตัวประมาณค่าที่นำเสนอดีกว่าตัวประมาณค่า T_{d1}, T_{d2}, ตัวประมาณค่า เมื่อประมาณค่าข้อมูลสูญหายด้วยวิธีการประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย, การประมาณข้อมูลสูญหายด้วยค่าอัตราส่วน, การประมาณข้อมูลสูญหายด้วยค่าคอมโพรไมซและวิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก ซึ่งเราจะนำเสนอเป็นลำดับ ดังนี้

1. ตัวประมาณค่าที่ประยุกต์มาจากแนวคิดของ Singh และคณะ [2] พิจารณาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่า T_{new} นั่นคือ

$$\begin{aligned} MSE(T_{new}) &= E[T_{new} - \bar{Y}]^2 \\ &= E[(T_{new})^2 - 2T_{new}\bar{Y} + \bar{Y}^2] \\ &= E[\lambda^2 T_{d1}{}^2 - 2\lambda T_{d1}\bar{Y} + \bar{Y}^2] \\ &= \lambda^2 E(T_{d1}{}^2) - 2\lambda\bar{Y}E(T_{d1}) + \bar{Y}^2 \\ &= \lambda^2 E(T_{d1}{}^2) - 2\lambda\bar{Y}E(T_{d1}) + \bar{Y}^2 + \lambda^2[E(T_{d1})]^2 - \lambda^2[E(T_{d1})]^2 \\ &= \lambda^2[E(T_{d1}{}^2) - [E(T_{d1})]^2] + [\lambda E(T_{d1}) - \bar{Y}]^2 \end{aligned}$$

จากความเอนเอียงของตัวประมาณค่า T_{d1} คือ $b(T_{d1}) = E(T_{d1}) - \bar{Y}$ หรือ $E(T_{d1}) = \bar{Y} + b(T_{d1})$ ดังนั้นเราจะได้

$$\begin{aligned} MSE(T_{new}) &= \lambda^2 Var(T_{d1}) + [\lambda(\bar{Y} + b(T_{d1})) - \bar{Y}]^2 \\ &= \lambda^2 Var(T_{d1}) + \bar{Y}^2(\lambda-1)^2 + 2\lambda\bar{Y}b(T_{d1})(\lambda-1) + \lambda^2 b^2(T_{d1}) \\ &= \lambda^2 Var(T_{d1}) + \lambda^2 b^2(T_{d1}) + \bar{Y}^2(\lambda-1)^2 + 2\lambda\bar{Y}b(T_{d1})(\lambda-1) \\ &= \lambda^2 MSE(T_{d1}) + \bar{Y}^2(\lambda-1)^2 + 2\lambda\bar{Y}b(T_{d1})(\lambda-1) \end{aligned}$$

โดยที่

$$MSE(T_{new}) = \lambda^2 MSE(T_{d1}) + \bar{Y}(\lambda-1)[\bar{Y}(\lambda-1) + 2\lambda b(T_{d1})] \qquad (14)$$

ซึ่งได้จากสมการ (6)

$$MSE(T_{d1}) = \left(\frac{1}{r} - \frac{1}{N}\right)(1 - \rho_{YX}^2)S_Y^2$$

ต่อไปหาค่าความเอนเอียงของตัวประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจาก

$$\begin{aligned} b(T_{new}) &= E(\lambda T_{d1} - \bar{Y}) \\ &= E(\lambda T_{d1}) - \bar{Y} \\ &= \lambda E(T_{d1}) - \bar{Y} \\ &= \lambda\left[\bar{Y} + \bar{Y}\varphi\left(\frac{1}{r} - \frac{1}{N}\right)\left[\rho_{YX}C_Y C_X - \varphi_2 C_X^2\right]\right] - \bar{Y} \\ &= \bar{Y}\left\{\lambda + \lambda\varphi\left(\frac{1}{r} - \frac{1}{N}\right)\left[\rho_{YX}C_Y C_X - \varphi_2 C_X^2\right] - 1\right\} \end{aligned}$$

ดังนั้นเราจะได้

$$b(T_{new}) = \bar{Y}\left\{\lambda + \lambda\varphi\left(\frac{1}{r} - \frac{1}{N}\right)\left[\rho_{YX}C_Y C_X - \varphi_2 C_X^2\right] - 1\right\} \tag{15}$$

ต่อไปเราจะหาค่า λ ที่ทำให้ความคลาดเคลื่อนกำลังสองเฉลี่ยของ T_{new} มีค่าต่ำสุด ดังนี้

$$\begin{aligned}
\frac{\partial MSE(T_{new})}{\partial\lambda} &= \frac{\partial}{\partial\lambda}\left\{\lambda^2 MSE(T_{d1}) + \bar{Y}(\lambda-1)[\bar{Y}(\lambda-1) + 2\lambda b(T_{d1})]\right\} \\
&= \frac{\partial}{\partial\lambda}\left\{\lambda^2 MSE(T_{d1}) + \bar{Y}^2(\lambda-1)^2 + 2\lambda\bar{Y}b(T_{d1})(\lambda-1)\right\} \\
&= \frac{\partial}{\partial\lambda}\left\{\lambda^2 MSE(T_{d1}) + \bar{Y}^2(\lambda-1)^2 + 2\lambda^2\bar{Y}b(T_{d1}) - 2\lambda\bar{Y}b(T_{d1})\right\} \\
&= 2\lambda MSE(T_{d1}) + 2\bar{Y}^2(\lambda-1) + 4\lambda\bar{Y}b(T_{d1}) - 2\bar{Y}b(T_{d1})
\end{aligned}$$

ซึ่งทำให้ได้ $MSE(T_{new})$ มีค่าต่ำสุดเมื่อ

ให้ $\dfrac{\partial MSE(T_{new})}{\partial\lambda} = 0$

$$\begin{aligned}
\lambda &= \frac{\bar{Y}[\bar{Y} + b(T_{d1})]}{MSE(T_{d1}) + \bar{Y}[\bar{Y} + 2b(T_{d1})]} \\
&= \frac{\bar{Y}[\bar{Y} + b(T_{d1})]}{Var(T_{d1}) + \left(b(T_{d1})\right)^2 + \bar{Y}^2 + 2\bar{Y}b(T_{d1})} \\
&= \frac{\bar{Y}[\bar{Y} + b(T_{d1})]}{Var(T_{d1}) + [\bar{Y} + b(T_{d1})]^2}
\end{aligned}$$

$$\lambda = \frac{\bar{Y}[\bar{Y} + b(T_{d1})]}{Var(T_{d1}) + [\bar{Y} + b(T_{d1})]^2} \tag{16}$$

ดังนั้นตัวประมาณค่าที่ประยุกต์มาจากแนวคิดของ Singh และคณะ [2] หรือ

$$T_{new} = \lambda T_{d1}$$

มีค่า $MSE(T_{new})$ ต่ำสุดก็ต่อเมื่อค่า λ ได้มาจากสมการ (16)

2. เงื่อนไขที่ทำให้ประสิทธิภาพของ T_{new} ดีกว่า T_{d1} สามารถพิจารณาได้ ดังนี้

$$
\begin{aligned}
MSE(T_{new}) - MSE(T_{d1}) &= (\lambda^2 - 1)MSE(T_{d1}) + \bar{Y}(\lambda - 1)[\bar{Y}(\lambda - 1) + 2\lambda b(T_{d1})] \\
&= (\lambda - 1)\left\{(\lambda + 1)MSE(T_{d1}) + \bar{Y}[\bar{Y}(\lambda - 1) + 2\lambda b(T_{d1})]\right\} \\
&= (\lambda - 1)\left\{\lambda MSE(T_{d1}) + MSE(T_{d1}) + \bar{Y}^2\lambda - \bar{Y}^2 + 2\lambda\bar{Y}b(T_{d1})\right\} \\
&= (\lambda - 1)\left\{MSE(T_{d1}) - \bar{Y}^2 + \lambda MSE(T_{d1}) + \bar{Y}^2\lambda + 2\lambda\bar{Y}b(T_{d1})\right\} \\
&= (\lambda - 1)\left\{MSE(T_{d1}) - \bar{Y}^2 + \lambda[MSE(T_{d1}) + \bar{Y}^2 + 2\bar{Y}b(T_{d1})]\right\} \\
&= (\lambda - 1)\left\{MSE(T_{d1}) - \bar{Y}^2 + \lambda[Var(T_{d1}) + (b(T_{d1}))^2 + \bar{Y}^2 + 2\bar{Y}b(T_{d1})\right\} \\
&= (\lambda - 1)\left\{MSE(T_{d1}) - \bar{Y}^2 + \lambda[Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2]\right\}
\end{aligned}
$$

ซึ่ง $MSE(T_{new})-MSE(T_{d1}) < 0$ ก็ต่อเมื่อ

$$(\lambda - 1)\left\{MSE(T_{d1}) - \bar{Y}^2 + \lambda[Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2]\right\} < 0 \qquad (17)$$

ต่อไปเราจะหาเงื่อนไขที่เหมาะสมของค่า λ ที่ทำให้ประสิทธิภาพของ T_{new} ดีกว่า T_{d1} โดยแบ่งการพิจารณาออกเป็นสองกรณี ดังนี้

กรณีที่ 1 ถ้า $\lambda < 1$

จากอสมการ (17) หารตลอดด้วย $(\lambda-1)$ เราจะได้ว่า

$$MSE(T_{d1}) - \bar{Y}^2 + \lambda[Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2] > 0 \quad \Longleftrightarrow \quad 1 > \lambda > \frac{\bar{Y}^2 - MSE(T_{d1})}{Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2}$$

โดยกำหนดให้ $\dfrac{\bar{Y}^2 - MSE(T_{d1})}{Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2} = k^*$

ในกรณีนี้เงื่อนไขที่เหมาะสมของค่า λ คือ

$$k^* < \lambda < 1 \qquad (18)$$

กรณีที่ 2 ถ้า $\lambda > 1$

จากอสมการ (17) หารตลอดด้วย $(\lambda-1)$ เราจะได้ว่า

$$MSE(T_{d1}) - \bar{Y}^2 + \lambda[Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2] < 0 \quad \Longleftrightarrow 1 < \lambda < \frac{\bar{Y}^2 - MSE(T_{d1})}{Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2}$$

ในกรณีนี้เงื่อนไขที่เหมาะสมของค่า λ คือ

$$1 < \lambda < k^* \tag{19}$$

ดังนั้น จากทั้งสองกรณี $MSE(T_{new})$ มีค่าน้อยกว่า $MSE(T_{d1})$ หรือ $MSE(T_{new})$ มีประสิทธิภาพดีกว่า $MSE(T_{d1})$ ก็ต่อเมื่อค่า λ สอดคล้องกับอสมการเงื่อนไข (18) หรือ (19)

ข้อสังเกต

1. เราจะเห็นได้ว่าค่า λ ที่ทำให้ประสิทธิภาพของ T_{new} ดีกว่า T_{d1} มีได้มากมายหลายค่า แต่ค่า λ ที่ทำให้ T_{new} มีประสิทธิภาพสูงสุด คือ ค่า λ ที่ได้มาจากสมการ (16)

2. พิจารณาค่า λ ในสมการ (16) และค่า k^* ในอสมการ (18) และ (19) เราจะเห็นว่า λ และ k^* จะมีค่าน้อยกว่าหนึ่งก็ต่อเมื่อ $b(T_{d1}) > \dfrac{-MSE(T_{d1})}{\bar{Y}}$ และในทางกลับกัน λ และ k^* จะมีค่ามากกว่าหนึ่งก็ต่อเมื่อ $b(T_{d1}) < \dfrac{-MSE(T_{d1})}{\bar{Y}}$

ตัวอย่างเชิงตัวเลข

ในหัวข้อนี้เป็นการศึกษาหรือทดสอบว่าผลลัพธ์เชิงตัวเลขมีความสอดคล้องกับผลลัพธ์เชิงทฤษฎีหรือไม่ ดังนั้น เราจึงนำเสนอตัวอย่างเชิงตัวเลขเพื่อแสดงให้เห็นว่าผลลัพธ์เชิงตัวเลขสนับสนุนผลการศึกษาในเชิงทฤษฎีที่ได้ในหัวข้อที่ผ่านมา ซึ่งตัวอย่างที่ผู้วิจัยนำมาเสนอในงานวิจัยนี้นำข้อมูลมาจากสมาคมค้าทองคำ (Gold Traders Association)

ตัวอย่างที่ 1 ประชากรที่เป็นตัวอย่างในการคำนวณใช้ข้อมูลจากราคาทองคำย้อนหลังจากสมาคมค้าทองคำ (Gold Traders Association) ตั้งแต่วันที่ 28 มีนาคม 2553 จนถึงวันที่ 1 ตุลาคม 2554 คำนวณด้วยเครื่องคอมพิวเตอร์โดยใช้โปรแกรม R 2.13.1 และโปรแกรม Maple7 ซึ่งข้อมูลประชากร N มีขนาดเท่ากับ 1219 ประกอบด้วย ตัวแปรช่วย X คือ ราคาทองต่างประเทศ ตัวแปรที่สนใจ Y คือ ราคาทองคำแท่ง 96.5% ณ ระดับค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) เท่ากับ 0.97 ลักษณะของประชากรของตัวแปรช่วย $\bar{X}$ = 1479.72 ตัวแปรที่สนใจ $\bar{Y}$ = 21441.3 และ ρ = 0.97748

ตารางที่ 1 แสดงการเปรียบเทียบวิธีการประมาณค่าข้อมูลสูญหายโดยค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) เท่ากับ 0.97

n	m	d	k*	λ	root mean square error						
					MI	RAT	COMP	WNR	T_{d2}	T_{d1}	T_{new}
13	5	1.510149	0.9999930	0.9999980	636.69169	612.38411	611.76421	757.36279	193.79686	73.37749	62.46432
		2.060511	0.9999930	0.9999980	636.69169	612.38411	611.76421	757.36279	193.79686	73.37749	62.46430
		12.090167	0.9999930	0.9999980	636.69169	612.38411	611.76421	757.36279	193.79686	73.37749	62.46433
	10	1.541730	1.0051560	1.0025780	754.32537	725.75905	724.84092	1719.46140	233.44700	91.35741	72.84549
		2.063123	1.0051560	1.0025780	754.32537	725.75905	724.84092	1719.46140	233.44700	91.35741	72.84536
		11.417991	1.0051560	1.0025780	754.32537	725.75905	724.84092	1719.46140	233.44700	91.35741	72.84550
	15	1.542186	1.0052166	1.0026395	754.46726	725.91520	724.99481	1848.87480	234.17731	93.25965	75.24578
		2.063164	1.0051539	1.0025770	754.46726	725.91520	724.99481	1848.87480	234.17731	93.25965	75.23389
		11.408835	1.0051539	1.0025770	754.46726	725.91520	724.99481	1848.87480	234.17731	93.25965	75.23390
	20	1.541401	1.0062017	1.0031018	826.25457	762.36571	760.27051	4084.34390	346.37943	95.91018	69.34235
		2.063094	1.0061998	1.0030999	826.25457	762.36571	760.27051	4084.34390	346.37943	95.91018	69.34218
		11.424602	1.0061998	1.0030999	826.25457	762.36571	760.27051	4084.34390	346.37943	95.91018	69.34237
25	5	1.397881	1.0030220	1.0015110	605.87223	593.73868	593.58472	658.55214	140.40754	74.01101	66.56700
		2.099960	1.0030220	1.0015110	605.87223	593.73868	593.58472	658.55214	140.40754	74.01101	66.56697
		14.709459	1.0030220	1.0015110	605.87223	593.73868	593.58472	658.55214	140.40754	74.01101	66.56700
	10	1.396770	1.0033203	1.0016686	633.04556	607.43082	607.10502	638.65743	191.97563	76.02422	67.30464
		2.099859	1.0033032	1.0016516	633.04556	607.43082	607.10502	638.65743	191.97563	76.02422	67.30359
		14.749100	1.0033032	1.0016516	633.04556	607.43082	607.10502	638.65743	191.97563	76.02422	67.30366
	15	1.393099	1.0036310	1.0018134	661.09628	620.28705	619.77642	5286.90410	233.92496	59.98271	45.65767
		2.099530	1.0036352	1.0018176	661.09628	620.28705	619.77642	5286.90410	233.92496	59.98271	45.65756
		14.881649	1.0036352	1.0018176	661.09628	620.28705	619.77642	5286.90410	233.92496	59.98271	45.65760
	20	1.392961	1.0040014	1.0019998	692.80209	635.03895	634.30802	12908.90100	278.89470	52.89651	31.04113
		2.099517	1.0040032	1.0020016	692.80209	635.03895	634.30802	12908.90100	278.89470	52.89651	31.04089
		14.886680	1.0040032	1.0020016	692.80209	635.03895	634.30802	12908.90100	278.89470	52.89651	31.04113

จากตารางที่ 1 เราจะเห็นว่า $MSE(T_{new})$ มีค่าน้อยกว่า $MSE(T_{d1})$ และเมื่อพิจารณาค่า λ ที่คำนวณได้จากสมการ (16) เราพบว่า $\dfrac{\bar{Y}^2 - MSE(T_{d1})}{Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2} = k^*$ มีค่าน้อยกว่า λ ซึ่งมีค่าน้อยกว่า 1 แสดงว่าค่า λ ที่คำนวณได้สอดคล้องกับอสมการ (18) และ $\dfrac{\bar{Y}^2 - MSE(T_{d1})}{Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2} = k^*$ มีค่ามากกว่า λ ซึ่งมีค่ามากกว่า 1 แสดงว่าค่า λ ที่คำนวณได้สอดคล้องกับอสมการ (19) จึงทำให้ประสิทธิภาพของ T_{new} ดีกว่า T_{d1}

ตัวอย่างที่ 2 ประชากรที่เป็นตัวอย่างในการคำนวณใช้ข้อมูลจากราคาข้าวหอมมะลิย้อนหลังจากกรมการค้าภายใน กระทรวงพาณิชย์ ตั้งแต่วันที่ 1 มกราคม 2550 จนถึงวันที่ 1 พฤศจิกายน 2554 คำนวณด้วยเครื่องคอมพิวเตอร์โดยใช้โปรแกรม R 2.13.1 และโปรแกรม Maple7 ซึ่งข้อมูลประชากร N มีขนาดเท่ากับ 1147 ประกอบด้วย ตัวแปรช่วย X คือ ราคาข้าวเปลือกหอมมะลิ (ชนิดสีได้ต้นข้าว 32-46 กรัม) น้ำหนัก 1000 กิโลกรัม ตัวแปรที่สนใจ คือ Y ราคาข้าวหอมมะลิ 100% ชั้น 1 น้ำหนัก 100 กิโลกรัม ณ ระดับค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) เท่ากับ 0.91 ลักษณะของประชากรของตัวแปรช่วย $\bar{X}$ = 13269.49 ตัวแปรที่สนใจ $\bar{Y}$ = 2713.94 และ ρ = 0.916703

ตารางที่ 2 แสดงการเปรียบเทียบวิธีการประมาณค่าข้อมูลสูญหายโดยค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) เท่ากับ 0.91

n	m	d	k*	λ	root mean square error						
					MI	RAT	COMP	WNR	T_{d2}	T_{d1}	T_{new}
12	5	1.966920	1.0001122	1.0000561	48.58824	49.67618	48.10681	1790.01678	35.85464	34.88393	34.88359
		2.451553	1.0001090	1.0000545	48.58824	49.67618	48.10681	1790.01678	35.85519	34.88454	34.88423
		5.274912	1.0001122	1.0000561	48.58824	49.67618	48.10681	1790.01678	35.85464	34.88393	34.88359
	10	1.742112	1.0056660	1.0028978	125.48274	122.02115	121.59833	797.55077	72.46244	64.99265	64.56069
		2.092745	1.0055357	1.0027678	125.48274	122.02115	121.59833	797.55077	72.46244	64.99265	64.55974
		8.398822	1.0055357	1.0027678	125.48274	122.02115	121.59833	797.55077	72.46244	64.99265	64.55974
	15	1.513531	1.0134302	1.0068189	177.07469	164.53838	163.67565	829.56058	87.44966	57.73937	54.92440
		2.110635	1.0119090	1.0059686	177.07469	164.53838	163.67565	829.56058	88.42489	59.49806	57.30014
		11.519923	1.0119090	1.0059686	177.07469	164.53838	163.67565	829.56058	88.15752	59.01918	55.79973
	20	1.033993	1.0234480	1.0118619	197.91859	189.76037	189.76001	4704.33402	73.03193	42.11720	28.44016
		2.041570	1.0231659	1.0115829	197.91859	189.76037	189.76001	4704.33402	73.03193	42.11720	28.43014
		170.460293	1.0231658	1.0115829	197.91859	189.76037	189.76001	4704.33402	73.03193	42.11720	28.43031
23	5	1.538243	1.0043931	1.0022128	108.50140	106.64775	106.58313	277.80846	51.85900	47.79757	47.43174
		2.114195	1.0043603	1.0021801	108.50140	106.64775	106.58313	277.80846	51.85900	47.79757	47.43165
		11.004862	1.0043603	1.0021801	108.50140	106.64775	106.58313	277.80846	51.85900	47.79757	47.43165
	10	1.567199	1.0055358	1.0028396	113.22061	111.20858	111.11813	237.63493	51.86731	46.93469	46.36585
		2.118875	1.0053916	1.0026958	113.22061	111.20858	111.11813	237.63493	51.86731	46.93469	46.36421
		10.454388	1.0053916	1.0026958	113.22061	111.20858	111.11813	237.63493	51.86731	46.93469	46.36421
	15	1.458283	1.0070480	1.0035417	121.08739	114.08033	113.92242	657.49496	61.53041	44.29611	43.26966
		2.103847	1.0070123	1.0035061	121.08739	114.08033	113.92242	657.49496	61.53041	44.29611	43.26954
		12.860635	1.0070123	1.0035061	121.08739	114.08033	113.92242	657.49496	61.53041	44.29611	43.26955
	20	1.513531	1.0080288	1.0040648	129.03307	118.76485	118.43549	6612.15355	67.33408	42.72617	41.36213
		2.110635	1.0079272	1.0039636	129.03307	118.76485	118.43549	6612.15355	67.33408	42.72617	41.36122
		11.519923	1.0079272	1.0039636	129.03307	118.76485	118.43549	6612.15355	67.33408	42.72617	41.36123

จากตารางที่ 1 เราจะเห็นว่า $MSE(T_{new})$ มีค่าน้อยกว่า $MSE(T_{d1})$ และเมื่อพิจารณาค่า λ ที่คำนวณได้จากสมการ (16) เราพบว่า $\frac{\bar{Y}^2 - MSE(T_{d1})}{Var(T_{d1}) + (\bar{Y} + b(T_{d1}))^2} = k^*$ มีค่ามากกว่า λ ซึ่งมีค่ามากกว่า 1 แสดงว่าค่า λ ที่คำนวณได้สอดคล้องกับอสมการ (19) จึงทำให้ประสิทธิภาพของ T_{new} ดีกว่า T_{d1}

สรุปผลการวิจัย

ตัวประมาณค่าเฉลี่ยประชากรในการสำรวจตัวอย่างเมื่อมีข้อมูลสูญหายที่นำเสนอในงานวิจัยนี้ เราได้ประยุกต์ใช้วิธีการหาตัวประมาณค่าของ Singh และคณะ [2] ในการปรับตัวประมาณค่า (T_{d1}) ให้เป็นตัวประมาณค่าใหม่ (T_{new}) และเมื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าตัวใหม่ที่นำเสนอกับตัวประมาณค่า T_{d1}, T_{d2},ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยวิธีประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย, การประมาณข้อมูลสูญหายด้วยค่าอัตราส่วน, การประมาณข้อมูลสูญหายด้วยค่าคอมโพรไมซและวิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก (เปรียบเทียบประสิทธิภาพโดยใช้ค่า the root mean square error หรือ RMSE พบว่าตัวประมาณค่า T_{new} มีประสิทธิภาพดีกว่าตัวประมาณค่า T_{d1}, T_{d2}, ตัวประมาณค่าเมื่อประมาณค่าข้อมูลสูญหายด้วยวิธีประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย, การประมาณข้อมูลสูญหายด้วยค่าอัตราส่วน, การประมาณข้อมูลสูญหายด้วยค่าคอมโพรไมซและวิธีการประมาณข้อมูลสูญหายด้วยค่าถดถอยของค่าเข้าใกล้สุดแบบถ่วงน้ำหนัก ภายใต้เงื่อนไขที่กำหนด (ค่า λ ในสมการ (16) ต้องสอดคล้องกับอสมการเงื่อนไข (18) หรือ (19)) นอกจากนี้ผลลัพธ์เชิงตัวเลขที่ได้นำเสนอทั้งสองตัวอย่างยังสนับสนุนผลการศึกษาในเชิงทฤษฎีข้างต้นได้เป็นอย่างดี

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณผู้ทรงคุณวุฒิที่ได้อ่านผลงานวิจัยและได้กรุณาให้คำแนะนำและแก้ไขข้อบกพร่องต่างๆ จนทำให้ผลงานวิจัยนี้มีความถูกต้องและสมบูรณ์มากยิ่งขึ้น

เอกสารอ้างอิง

1. Kalton, G., and Kasprzyk, D. 1982. Imputing for Missing Survey Responses. Available from URL: http://www.amstat.org/seetions/srms/Proceedings/allyearsf.html. 26 November 2011.
2. Singh, G. N., Priyanka, K., Kim, J. M., and Singh, S. 2010. Estimation of Population Mean Using Imputation Techniques in Sample Surveys. *Journal of the Korean Statistical Society* 39: 67-74.
3. Chaimongkol, W., and Suwattee, P. 2005. Three Composite Imputation Methods for Item Nonresponse Estimation in Sample Surveys. Doctor of Philosophy (Statistics). School of Applied Statistics, National Instituted of Development Administration. Bangkok. National Institute of Development Administration.
4. Little R. J. A., and Rubin, D. B. 1987. Statistical Analysis with Missing Data. New York. John Wiley and Sons, Inc.

ได้รับบทความวันที่ 19 ธันวาคม 2554
ยอมรับตีพิมพ์วันที่ 7 กุมภาพันธ์ 2555