วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 27 ฉบับที่ 4 ต.ค.–ธ.ค. 2560
The Journal of KMUTNB., Vol. 27, No. 4, Oct.–Dec. 2017

บทความวิจัย

การเพิ่มประสิทธิภาพการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิตซิง

จำลอง วงษ์ประเสริฐ*
สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08–1790–3737 อีเมล: jumlong.v@ubru.ac.th DOI: 10.14416/j.kmutnb.2017.10.004
รับเมื่อ 27 ตุลาคม 2559 ตอบรับเมื่อ 4 มกราคม 2560 เผยแพร่ออนไลน์ 30 ตุลาคม 2560
© 2017 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

บทคัดย่อ

บทความนี้มีจุดมุ่งหมายเพื่อเสนอแนวทางในการเพิ่มประสิทธิภาพในการประมาณค่าข้อมูลสูญหายโดยใช้รีจีมสวิตซิง โดยการนำเสนอการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิตซิงมีนและรีจีมสวิทซิงรีเกรสชัน โดยทำการทดลองเปรียบเทียบประสิทธิภาพระหว่างการประมาณค่าข้อมูลสูญหายด้วยค่าเฉลี่ยกับรีจีมสวิตซิงมีน และระหว่างวิธีการประมาณค่าข้อมูลสูญหายด้วยสมการถดถอยกับรีจีมสวิทซิงรีเกรสชัน โดยเปรียบจากค่า MSE พบว่า รีจีมสวิตซิงมีนมีประสิทธิภาพดีกว่าการประมาณค่าข้อมูลสูญหายด้วยค่าเฉลี่ย และรีจีมสวิทซิงรีเกรสชันมีประสิทธิภาพดีกว่า วิธีการประมาณค่าข้อมูลสูญหายด้วยสมการถดถอย

คำสำคัญ: ข้อมูลสูญหาย, รีจีมสวิทซิง

การอ้างอิงบทความ: จำลอง วงษ์ประเสริฐ, "การเพิ่มประสิทธิภาพการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิตซิง," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 27, ฉบับที่ 4, หน้า 821–827, ต.ค.–ธ.ค. 2560

Research Article

Increase Efficiency of Missing Data Imputation by Regime Switching

Jumlong Vongprasert*
Department of Applied Statistics, Faculty of Science, Ubon Ratchathani Rajabhat University, Ubon Ratchathani, Thailand

* Corresponding Author, Tel. 08–1790–3737, E–mail: jumlong.v@ubru.ac.th DOI: 10.14416/j.kmutnb.2017.10.004
Received 27 October 2016; Accepted 4 January 2017; Published online: 30 October 2017
© 2017 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

Abstract

The aim of this paper is to propose regime switching approach to missing data imputation for increasing efficiency by regime switching mean and regime switching regression. By using simulation data, the comparisons were made between mean imputation with regime switching mean and between regression imputation with regime switching regression. Under MSE regime switching mean outperformed the mean imputation and regime switching regression outperformed the regression imputation method.

Keywords: Missing Data, Regime Switching

Please cite this article as: J. Vongprasert, "Increase efficiency of missing data imputation by regime switching," *The Journal of KMUTNB.*, vol. 27, no. 4, pp. 821–827, Oct.–Dec. 2017 (in Thai).

1. บทนำ

ข้อมูลสูญหาย (Missing Data) คือค่าสังเกตที่ต้องการทราบค่าแต่ไม่สามารถทราบค่าได้ โดยที่ค่านั้นควรจะสามารถทราบค่าได้หากวิธีการที่ใช้ในการรวบรวมข้อมูลหรือในการวัดค่ามีประสิทธิภาพดีขึ้น หรือมีความเหมาะสมมากขึ้น ข้อมูลสูญหายเป็นปัญหาที่พบโดยทั่วไปในการวิจัยเชิงปริมาณ เป็นเรื่องที่ต้องเกิดขึ้นที่นักวิจัยจะต้องเผชิญ แม้ว่าจะได้ควบคุมการสำรวจหรือการทดลองไว้แล้วเป็นอย่างดีก็ตาม [1] นักวิจัยได้ให้ความสำคัญกับปัญหาข้อมูลสูญหายและให้ความสำคัญเพิ่มขึ้นเรื่อยๆ [2] นักวิจัยพยายามจะขจัดปัญหาข้อมูลสูญหาย โดยพยายามที่จะเก็บข้อมูลให้ได้ข้อมูลที่สมบูรณ์ แต่ปัญหาข้อมูลสูญหายนั้นเกิดจากหลากหลายสาเหตุ นักวิจัยจำเป็นต้องพิจารณาถึงแนวทางที่เหมาะสม สำหรับใช้จัดการกับข้อมูลสูญหายในทุกๆ ครั้งที่พบกับปัญหานี้ ซึ่งวิธีการที่ใช้สำหรับจัดการกับข้อมูลสูญหายมีทางเลือกให้พิจารณาค่อนข้างหลากหลาย หากเลือกใช้วิธีจัดการกับข้อมูลสูญหายที่ไม่เหมาะสมย่อมส่งผลทำให้เกิดการบิดเบือนผลการวิเคราะห์ วิธีการทางสถิติได้ถูกพัฒนาเพื่อการวิเคราะห์ข้อมูลสำหรับชุดข้อมูลที่สมบูรณ์ แต่เมื่อมีข้อมูลสูญหายย่อมส่งผลกระทบต่อประสิทธิภาพการวิเคราะห์ข้อมูล ในอดีตข้อมูลสูญหายได้ถูกจัดการโดยวิธีการต่างๆ วิธีการที่พยายามแก้ไขข้อมูลก่อนการวิเคราะห์ เช่น การลบข้อมูลที่ไม่สมบูรณ์ทิ้ง (Listwise Deletion) เป็นหนึ่งในกลยุทธ์ดังกล่าว วิธีอื่นๆ เช่น การแทนค่า ตัวแปรที่หายไปโดยใช้ค่าเฉลี่ยของตัวแปรที่เหลืออยู่ (Mean Substitution) แต่วิธีการเหล่านี้ทำให้ตัวประมาณ (Estimator) เกิดความเอนเอียง (Bias) จากการศึกษาของ Wood *et al.* [3] ที่ได้ทำการศึกษาจากผลงานวิจัยที่ได้รับการตีพิมพ์ในวารสารจำนวน 71 ชิ้น พบว่ามีงานวิจัยถึงร้อยละ 89 ที่มีปัญหาเรื่องข้อมูลสูญหาย และมีเพียงร้อยละ 21 เท่านั้นที่มีการจัดการกับปัญหาข้อมูลที่ไม่สมบูรณ์ จากผลการศึกษานี้แสดงให้เห็นว่า การจัดการกับปัญหาข้อมูลสูญหายยังคงถูกละเลยกันอย่างเป็นปกติ

การสูญหายของข้อมูลจะมีผลกระทบที่รุนแรง ในการวิเคราะห์ข้อมูลตัวแปรพหุ (Multivariate Data) เพราะถ้าพบว่าหน่วยวิเคราะห์ใดมีตัวแปรใดขาดข้อมูลไปแม้เพียงตัวแปรเดียวก็จะตัดหน่วยวิเคราะห์นั้นทิ้งทั้งหน่วย โดยไม่สนใจว่าจะยังมีตัวแปรอื่น อีกมากที่มีข้อมูลครบถ้วนหรือไม่ [4]

การตัดข้อมูลทิ้งเป็นวิธีปริยาย (Default) ของการวิเคราะห์ข้อมูลตัวแปรพหุในโปรแกรมสถิติทั่วไป Kim และ Curry [5] ได้ทำการศึกษาโดยวิธีทดลองพบว่า ถ้าตัวแปรแต่ละตัวมีข้อมูลหายไปโดยสุ่มเพียงร้อยละ 10 จะมีผลให้ต้องตัดหน่วยวิเคราะห์ทิ้งถึงร้อยละ 59 เห็นได้ว่าเป็นความสูญเสียในอัตราที่สูงมาก การวิเคราะห์ข้อมูลจากเฉพาะข้อมูลที่เหลืออยู่ภายหลังจากที่ได้ตัดทิ้งค่าสังเกตที่ไม่สมบูรณ์ไปแล้ว ผลการวิเคราะห์จะเอนเอียงและไม่ถูกต้อง และเป็นการสูญเสียสารสนเทศ เสียค่าใช้จ่าย ดังนั้นควรหาวิธีการประมาณค่าข้อมูลที่สูญหายไปนั้น ด้วยวิธีที่เหมาะสมก่อนทำการวิเคราะห์ข้อมูล เพราะหากได้มีการทดแทนรายการข้อมูลที่สูญหายด้วยวิธีการที่เหมาะสมก่อนการวิเคราะห์ข้อมูล จะพบว่าสถิติทดสอบ (Test Statistics) และสถิติอื่นๆ มีอำนาจทดสอบ (Power of Test) สูงขึ้น การประมาณค่าข้อมูลสูญหายเพื่อทดแทนข้อมูลที่สูญหายไปนั้น ทำให้ประสิทธิภาพของการประมาณค่าและการสรุปผล การวิจัยสูงขึ้นอย่างน่าประหลาดใจ [6]

วัตถุประสงค์การวิจัยครั้งนี้เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าข้อมูลสูญหายด้วยค่าเฉลี่ยกับวิธีการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิทชิงมีน และวิธีการประมาณค่าข้อมูลสูญหายด้วยการถดถอย กับวิธีการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิทชิงรีเกรสชัน

2. ข้อมูลและวิธีการวิจัย

1. ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลที่ได้จากการจำลองสถานการณ์ โดยใช้เทคนิคมอนติคาร์โลซิมูเลชัน โดยที่ตัวแปรสองตัวมีสัมประสิทธิ์สหสัมพันธ์กำหนดระดับความสัมพันธ์ระหว่างตัวแปร 3 ระดับ คือ ระดับความสัมพันธ์ระหว่างตัวแปรต่ำ ($\rho = 0.3$) ปานกลาง

(ρ = 0.5) และสูง (ρ = 0.7) ขนาดตัวอย่างในการศึกษา คือ 100, 200 และ 500 หน่วย ประเภทของข้อมูลสูญหาย เป็นการสูญหายแบบสุ่มอย่างสมบูรณ์ ร้อยละข้อมูลสูญหายร้อยละ 5, 10, 15 และ 20 [7]–[14]

แฮมิลตัน [15] แนะนำรีจีมสวิตชิง (Regime Switching) โดยใช้ตัวแบบออโตรีเกรสซีฟ (Autoregressive; AR) กระบวนการรีจีมสวิตชิง (Regime Switching Process) และตัวแบบมาร์คอฟสวิทชิง (Markov Switching Models) ในการอธิบายข้อมูลที่ไม่เป็นอิสระกัน เขาแสดงให้เห็นถึงการสลับไปมาระหว่างความผันผวนสูงและความผันผวนต่ำ (High Volatility and Low Volatility) ที่มีการเปลี่ยนแปลงเป็นระยะ ๆ ตัวแบบรีจีมสวิตชิง พฤติกรรมของกลุ่มตัวแปรจะขึ้นอยู่กับสภาวะของระบบทำให้การเปลี่ยนสภาวะของระบบที่มีการเปลี่ยนแปลงเป็นระยะ ๆ ระหว่างสภาวะผันผวนต่ำที่มีเสถียรภาพและความไม่แน่นอนมีมากขึ้นสำหรับสภาวะความผันผวนสูง

ตัวแบบรีจีมสวิตชิงอาศัยทฤษฎีความน่าจะเป็นในการอธิบายการเปลี่ยนแปลงของสภาวะที่เกิดขึ้น ตัวแบบคณิตศาสตร์ที่ใช้อธิบายการสลับไปมาระหว่างสภาวะผันผวนต่ำและสภาวะผันผวนสูงแสดงดังต่อไปนี้

$$P(y_t \,|\, s_t, s_{t-1}, ..., s_{(t-k_s)}, y_t, y_{t-1}, ..., y_{(t-k_y)}; \theta)$$

เมื่อ $y_1, y_2, \ldots, y_t \overset{iid}{\sim} N(0, \sigma^2)$

วิธีการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิทชิงมีน มีขั้นตอนดังนี้

1) จากข้อมูลที่สมบูรณ์ $(y_1,x_1),(y_2,x_2),...(y_r,x_r)$ นำมาแบ่งกลุ่มโดยใช้ตัวแปร y แบ่งกลุ่มของข้อมูล โดยใช้รีจีมสวิทชิง

2) จากข้อมูลที่สมบูรณ์ $y_{g1}, y_{g2}, \ldots y_{gr}$ ของแต่ละกลุ่ม นำมาหาค่าเฉลี่ยซึ่งการหาค่าประมาณจะถูกกำหนดโดย

$$\overline{y}_g = \frac{\sum_{i=1}^{gr} \overline{y}_{gi}}{gr}; g = 1, 2, \ldots, k$$

3) นำค่า $\overline{y}_g$ ที่ได้ทำนายค่าที่สูญหายในแต่ละกลุ่ม

วิธีการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิทชิงรีเกรสชัน มีขั้นตอนดังนี้

จากข้อมูลที่สมบูรณ์ จากข้อมูลที่สมบูรณ์ (y_{g1},x_{g1}), $(y_{g2},x_{g2}),...(y_{gr},x_{gr})$ ของแต่ละกลุ่ม นำมาสร้างสมการถดถอย กำหนดให้

$$\hat{y}_{gj} = \hat{\beta}_{g0} + \hat{\beta}_{g1} x_{gj}; r+1 \le j \le n$$

สำหรับ $r+1 \le j \le n$

โดยที่ $\hat{\beta}_{g0} = \overline{y}_{gr} - \hat{\beta}_{g1} \overline{x}_{gr}$

s_t; t = 1,2,3,...,k แทนสภาวะที่เกิดขึ้นในระบบ

θ แทนค่าสัมประสิทธิ์ของตัวแปร y

โดยตัวแบบรีจีมสวิตชิง จะได้

$$P[s_t \,|\, s_1, s_2, ..., s_{t-1}] = P[s_t \,|\, s_{t-1}]$$

ความน่าจะเป็นที่จะทำให้เกิดการเปลี่ยนสภาวะจากสภาวะ i ไปยังสภาวะ j แสดงได้ดังนี้

$$p_{ij} = P[s_t = j \,|\, s_{t-1} = i]$$

สภาวะ s_t ไม่สามารถสังเกตได้โดยตรงแต่สามารถสรุปได้จากตัวแปร y และความน่าจะเป็น p ดังนั้นจะได้

$$P(s_t = j \,|\, y_t, y_{t-1}, ..., y_1; p, \theta) = p_j \text{ และ}$$

$$P(s_{t-1} = i \,|\, y_t, y_{t-1}, ..., y_1; p, \theta) = p_i.$$

$$\hat{\beta}_{g1} = \frac{\sum_{i=1}^{gr} (x_{gi} - \overline{x}_{gr})(y_{gi} - \overline{y}_{gr})}{\sum_{i=1}^{gr} (x_{gi} - \overline{x}_{gr})^2}$$

$$\overline{y}_{gr} = \frac{1}{gr} \sum_{i=1}^{gr} y_{gi} \qquad \overline{x}_{gr} = \frac{1}{gr} \sum_{i=1}^{gr} x_{gi}$$

นำสมการถดถอยที่ได้ทำนายค่าที่สูญหาย ได้ค่า $\hat{y}_{gj}$

การดำเนินการทดลอง มีรายละเอียดดังนี้

1) สร้างประชากรจำนวน 10,000 หน่วย

2) สุ่มตัวอย่างจากประชากร ด้วยวิธีการสุ่มตัวอย่างอย่างง่าย โดยมีขนาดตัวอย่างตามที่กำหนด

3) สร้างข้อมูลสูญหายแบบสุ่มสมบูรณ์ โดยมีร้อยละข้อมูลสูญหายตามที่กำหนด

4) จัดการข้อมูลสูญหายด้วยวิธีการประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย วิธีการประมาณข้อมูลสูญหายด้วยการถดถอย วิธีการประมาณข้อมูลสูญหายโดยรีจีมสวิทชิงมีน และวิธีการประมาณข้อมูลสูญหายโดยรีจีมสวิทชิงรีเกรสชัน

5) คำนวณค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Mean Square Error; MSE) ที่ได้จากแต่ละวิธี โดยที่

$$MSE = \frac{\sum_{i=1}^{n}(\theta_i - \hat{\theta}_i)^2}{n}$$

6) ทำซ้ำขั้นตอนที่ 1 ถึง 4 จำนวน 10,000 ครั้ง

7) คำนวณค่าเฉลี่ยของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสอง ของวิธีการจัดการข้อมูลสูญหายแต่ละวิธี

ดำเนินการในลักษณะเดียวกันจากขั้นตอนที่ 1 ถึงขั้นตอนที่ 6 โดยกำหนดขนาดตัวอย่าง ระดับความสัมพันธ์ระหว่างตัวแปรและร้อยละข้อมูลสูญหาย ดังต่อไปนี้

2. ขนาดตัวอย่าง 100, 200 และ 500

3. ระดับความสัมพันธ์ระหว่างตัวแปรต่ำ ปานกลาง และสูง

4. ร้อยละข้อมูลสูญหายร้อยละ 5, 10, 15 และ 20

3. ผลการศึกษา

ผลการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าข้อมูลสูญหาย จำแนกตามขนาดตัวอย่าง ระดับความสัมพันธ์ระหว่างตัวแปร ร้อยละข้อมูลสูญหายและวิธีการจัดการข้อมูลสูญหาย แสดงดังตารางที่ 1 และตารางที่ 2

ตารางที่ 1 ค่า MSE ของวิธีการประมาณค่าข้อมูลสูญหาย จำแนกตามขนาดตัวอย่าง ระดับความสัมพันธ์ระหว่างตัวแปร ร้อยละข้อมูลสูญหายระหว่างวิธีการจัดการข้อมูลสูญหาย

ขนาดตัวอย่าง	ระดับความสัมพันธ์ระหว่างตัวแปร	ร้อยละข้อมูลสูญหาย	วิธีการประมาณค่าข้อมูลสูญหาย			
			ค่าเฉลี่ย	รีจีมสวิทชิงมีน	สมการถดถอย	รีจีมสวิทชิงรีเกรสชัน
100	ต่ำ	5	0.2868	0.0568	0.2679	0.1437
		10	0.5695	0.0563	0.5335	0.1337
		15	0.8565	0.0565	0.8045	0.1046
		20	1.1379	0.0563	1.0706	0.0810
	ปานกลาง	5	0.2944	0.0582	0.2270	0.0588
		10	0.5839	0.0577	0.4512	0.0446
		15	0.8781	0.0580	0.6787	0.0426
		20	1.1665	0.0577	0.9026	0.0422
	สูง	5	0.2944	0.0583	0.1567	0.0724
		10	0.5868	0.0580	0.3111	0.0614
		15	0.8824	0.0582	0.4710	0.0470
		20	1.1731	0.0580	0.6258	0.0392
200	ต่ำ	5	0.5755	0.0573	0.5311	0.1357
		10	1.1414	0.0568	1.0523	0.0830
		15	2.2773	0.0566	2.1062	0.0540
		20	2.2773	0.0566	2.1062	0.0540
	ปานกลาง	5	0.5853	0.0582	0.4470	0.0453
		10	1.1758	0.0584	0.8946	0.0434
		15	1.7649	0.0585	1.3413	0.0432
		20	2.3476	0.0584	1.7885	0.0432
	สูง	5	0.5906	0.0587	0.3081	0.0613
		10	1.1865	0.0590	0.6185	0.0377
		15	1.7854	0.0592	0.9266	0.0312
		20	2.3651	0.0588	1.2362	0.0299
500	ต่ำ	5	1.4385	0.0574	1.3125	0.0679
		10	2.8543	0.0570	2.6103	0.0517
		15	4.2962	0.0572	3.9292	0.0517
		20	5.7413	0.0573	5.2490	0.0518
	ปานกลาง	5	1.4771	0.0589	1.1080	0.0437
		10	2.9617	0.0591	2.2207	0.0438
		15	4.4125	0.0587	3.3256	0.0438
		20	5.8909	0.0588	4.4403	0.0438
	สูง	5	1.4827	0.0592	0.7661	0.0343
		10	2.9597	0.0590	1.5345	0.0304
		15	4.4577	0.0593	2.3056	0.0304
		20	5.9279	0.0591	3.0667	0.0303

ตารางที่ 2 เปรียบเทียบประสิทธิภาพวิธีการประมาณค่าข้อมูลสูญหาย จำแนกตามขนาดตัวอย่างและวิธีการจัดการข้อมูลสูญหาย

	วิธีการประมาณค่าข้อมูลสูญหาย			
	ค่าเฉลี่ย	**รีจีมสวิทชิงมีน**	**สมการถดถอย**	**รีจีมสวิทชิงรีเกรสชัน**
ขนาดตัวอย่าง				
100	0.7259	0.0575	0.5417	0.0726
200	1.5061	0.0580	1.1131	0.0552
500	3.6584	0.0584	2.6557	0.0436
ระดับความสัมพันธ์				
ต่ำ	1.9544	0.0568	1.7978	0.0844
ปานกลาง	1.9615	0.0584	1.4855	0.0449
สูง	1.9744	0.0587	1.0272	0.0421
ร้อยละข้อมูลสูญหาย				
5	0.7806	0.0581	0.5694	0.0737
10	1.5577	0.0579	1.1363	0.0589
15	2.4012	0.0580	1.7654	0.0498
20	3.1142	0.0579	2.2762	0.0462
ภาพรวม				
	1.9634	0.0580	1.4368	0.0571

จากตารางที่ 1 พบว่าค่า MSE ของวิธีการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิทชิงมีนมีค่าต่ำกว่าวิธีการประมาณค่าข้อมูลสูญหายด้วยค่าเฉลี่ยทุกกรณี และค่า MSE ของวิธีการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิทชิงรีเกรสชันมีค่าต่ำกว่าวิธีการประมาณค่าข้อมูลสูญหายด้วยสมการถดถอยทุกกรณี

จากตารางที่ 2 พบว่า เมื่อพิจารณาค่าจำแนกตามขนาดตัวอย่าง ระดับความสัมพันธ์ระหว่างตัวแปร ร้อยละข้อมูลสูญหายและภาพรวม ค่า MSE ของวิธีการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิทชิงมีนมีค่าต่ำกว่าวิธีการประมาณค่าข้อมูลสูญหายด้วยค่าเฉลี่ยทุกกรณี และค่า MSE ของวิธีการประมาณค่าข้อมูลสูญหายด้วยรีจีมสวิทชิงรีเกรสชันมีค่าต่ำกว่าวิธีการประมาณค่าข้อมูลสูญหายด้วยสมการถดถอยทุกกรณี

4. สรุป

จากการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าข้อมูลสูญหายระหว่างวิธีการประมาณค่าข้อมูลสูญหายด้วยค่าเฉลี่ยกับรีจีมสวิทชิงมีน พบว่ารีจีมสวิทชิงมีนมีประสิทธิภาพสูงกว่าทุกกรณี และการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าข้อมูลสูญหายระหว่างวิธีการประมาณค่าข้อมูลสูญหายด้วยสมการถดถอยกับรีจีมสวิทชิงรีเกรสชัน พบว่ารีจีมสวิทชิงรีเกรสชันมีประสิทธิภาพสูงกว่าทุกกรณี

5. กิตติกรรมประกาศ

ขอขอบคุณหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิจัยและประเมินผลการศึกษา มหาวิทยาลัยราชภัฏอุบลราชธานี สำหรับทุนที่สนับสนุนในการทำการวิจัยครั้งนี้

เอกสารอ้างอิง

[1] D. D. Leeuw, J.Hox, and M. Huisman, "Prevention and treatment of item nonresponse," *Journal of Official Statistics*, vol. 19, no. 2, pp. 153–176, June, 2003.

[2] A. Carlson, "Data mining: Finding nuggets of knowledge in mountains of data," *Northwest Science & Technology*, vol. 78, pp. 24–25, 2001.

[3] A. M. Wood, I. R. White, and S. G. Thompson, "Are missing outcome data adequately handled: A review of published randomized controlled trials in major medical journals," *Clinical Trial*. vol. 1, pp 368–376, 2004.

[4] N. Tsikriktsis, "A review of techniques for treating missing data in OM survey research," *Journal of Operations Management*, vol. 24, pp. 53–62, 2005.

[5] P. Ogoke Uchenna and E. C. Nduka, "Methods of analysing missing values in a regression model," *Indian Journal of Science and Technology*, vol. 5, no. 2, pp. 2013–2016, 2012.

[6] M. R. Raymond, "Missing data in evaluation research," *Evaluation & the Health Professions*, vol. 9, no. 4, pp. 395–420, 1986.

[7] J. W. France, "Some simple procedures for handling missing data in multivariate analysis," *Psychometrika*, vol. 41, pp. 409–415, 1976.

[8] L. L. Brockmeier, J. D. Kromrey, and C. V. Hines, "Systematically missing data and multiple regression analysis : An empirical comparison of election and imputation techniques," *Multiple Linear Regression Viewpoints*, vol. 25, pp. 20–39, 1998.

[9] N. H. Timm, "The estimation of variance-covariance and correlation matrices from incomplete data," *Psychometrika*, vol. 35, no. 4, pp. 417–437, 1970.

[10] E. M. L. Beale and R.J.A. Little, "Missing values in multivariate analysis," *Journal of the Royal Statistical Society*, vol. 37, pp. 129–145, 1975.

[11] T. C. Gleason and R. Staelin, "A proposal for handling missing data," *Psvchometrika*, vol. 40, pp. 229–252, 1975.

[12] L. S. Chan, J. A. Gilman, and O. J. Dunn, "Alternative approaches to missing values in discriminant analysis," *Journal of the American Statistical Association*, vol. 71, pp. 842–844, 1976.

[13] R. J. A. Little, "Inference about means from incomplete multivariate data," *Biometrika*, vol. 63, pp. 593–604, 1976.

[14] M. R. Raymond and D. M. Roberts, "A comparison of methods for treating incomplete data in selection research," *Educational and Psychological Measurement*, vol. 47, pp. 13–26, 1987.

[15] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, pp. 357–384, 1989.