

การพัฒนาวิธีการประมาณค่าข้อมูลสูญหายสำหรับการวัดซ้ำ ด้วยวิธีการแทนค่า แบบสองทางและตัวแบบออโตรีเกรสซีฟอันดับที่ 1

THE DEVELOPMENT ESTIMATING MISSING DATA METHOD FOR REPEATED MEASUREMENTS BY TWO-WAY IMPUTATION AND FIRST-ORDER AUTOREGRESSIVE MODEL

พรณิภา ภูกองพลอย* และ ทรงศักดิ์ ภูสีอ่อน
Punnipa Poogungploy* and Songsak Phusee – orn

สาขาวิชาวิจัยและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม
Educational Measurement, Faculty of Education, Mahasarakham University

บทคัดย่อ

การวิจัยในครั้งนี้ มีความมุ่งหมายเพื่อพัฒนาวิธีการประมาณค่าข้อมูลสูญหาย โดยการถ่วงน้ำหนักด้วยวิธีการแทนค่าแบบสองทางและตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง (IWTWAR (1)) และศึกษาประสิทธิภาพของการประมาณค่าข้อมูลสูญหายโดยเปรียบเทียบกับ การประมาณค่าข้อมูลสูญหายแบบเดิมโดยใช้จำลองข้อมูล ขอบเขตของการวิจัยใช้จำนวน การวัดซ้ำ 3 ครั้ง และ 6 ครั้ง ขนาดของตัวอย่าง แบ่งเป็น 10, 20 และ 50 และร้อยละของ ข้อมูลสูญหายแบ่งเป็น ร้อยละ 5, 10, 15 และร้อยละ 20 พิจารณาประสิทธิภาพของวิธีการ ประมาณค่าข้อมูลสูญหาย โดยใช้เกณฑ์ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) ผลการศึกษาพบว่า การประมาณค่าข้อมูลสูญหายโดยการถ่วงน้ำหนักด้วยวิธีการแทนค่า แบบสองทางและตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง มีประสิทธิภาพสูงสุด ในกรณีที่มี จำนวนการวัดซ้ำ 6 ครั้ง คะแนนพัฒนาการมีอัตราพัฒนาการเท่ากับ 0.5 และ 2.5 ขนาด ตัวอย่างเท่ากับ 10 หน่วย และ 20 หน่วย มีข้อมูลสูญหายร้อยละ 5, 10, 15 และร้อยละ 20 ทั้งในกรณีที่คะแนนพัฒนาการเป็นเส้นตรงและไม่เป็นเส้นตรง โดยวิธีการที่ผู้วิจัยพัฒนา ขึ้นมีประสิทธิภาพแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับวิธีการประมาณ

* ผู้ประสานงาน: พรณิภา ภูกองพลอย

อีเมล: jeab.stat@hotmail.com

ค่าข้อมูลสูญหายด้วยตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง (AR (1)) และวิธีการประมาณค่าข้อมูลสูญหายโดยใช้ค่าเฉลี่ย (MI) และในกรณีที่มีจำนวนการวัดซ้ำ 3 ครั้ง การประมาณค่าข้อมูลสูญหายด้วยตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง จะมีประสิทธิภาพสูงสุดทุกกรณี

คำสำคัญ: การประมาณค่าข้อมูลสูญหาย, การวัดซ้ำ, ออโตรีเกรสซีฟ

Abstract

This research aimed to develop a method estimate a missing data for repetition by two-way imputation and First-Order autoregressive model (IWTWAR (1)) and to compare the efficiency of estimation of missing data by comparing the estimation of missing data by using the traditional method. The scope of the study was a number of repeat measurements divided into 3 and 6 times. The size of the sample was divided into 10, 20 and 50 and the number of percent of missing data was divided into 5, 10, 15 and 20 percent. The researcher considered the efficiency of the missing data estimating method by using mean squared error (MSE). The result showed that the estimating of missing data by two-way imputation and the first-order autoregressive model for the most effective if repetition 6 times. Considering the developmental scores in a Linear Growth Model and the developmental scores were equal to 0.5 and 2.5. When the sample sizes were equal 10 and 20 units, missing data were 5, 10, 15 and 20 percent, in case a linear and non-linear score. The researcher developed the efficient estimating methods varying with different methods in the case of developmental point's equal 0.5 and were significantly different from the missing data estimating method with the First-Order of autoregressive model (AR (1)) and using missing data estimating methods by mean substitution (MI). The case repetitions 3 times estimating missing data with the First-Order of the autoregressive model showed the highest efficiency in all cases.

Keywords: Missing data, Repeated Measurements, Autoregressive

บทนำ

ในการวิจัยระยะยาวปัญหาที่นักวิจัยพบคือ ปัญหาการข้อมูลสูญหาย(Dropout) ของตัวอย่าง ข้อมูลสูญหาย (Missing data) หมายถึง การขาดหายไปของตัวแปรที่สนใจ หรือค่าสังเกตที่ต้องการทราบค่าแต่ไม่สามารถทราบค่าได้ โดยทั่วไปข้อมูลสูญหายจะขัดขวางความสามารถในการอธิบายและทำความเข้าใจกับปรากฏการณ์ที่ศึกษา ซึ่งอาจเกิดจากผู้ตอบแบบสอบถามปฏิเสธที่จะตอบคำถามบางข้อหรือผู้ตอบแบบสอบถามไม่มีข้อมูลเกี่ยวกับคำถามนั้น (Allison, 2012: 1145) สำหรับข้อมูลสูญหายที่เกิดขึ้นกับการวัดซ้ำในทางการศึกษา อาจเกิดจากนักเรียนย้ายโรงเรียนในระหว่างที่ดำเนินการทดลอง ผลกระทบที่เกิดจากการสูญหายของข้อมูลจะส่งผลอย่างมีนัยสำคัญต่อผลการวิเคราะห์ข้อมูล โดยส่งผลกระทบต่อค่าพารามิเตอร์ของประชากร ซึ่ง DeSilvio (2009: 5698-A) กล่าวว่า เมื่อมีข้อมูลสูญหายเกิดขึ้นค่าสถิติที่ใช้วัดแนวโน้มเข้าสู่ส่วนกลางจะมีความเอนเอียงที่อาจสูงขึ้นหรือต่ำลง การแทนที่ของข้อมูลสูญหายด้วยค่าทดแทนที่เหมาะสมจะลดความเอนเอียงได้ ถ้าการวิเคราะห์ข้อมูลนั้นจำเป็นต้องใช้วิธีวิเคราะห์แบบหลายตัวแปร (Multivariate Data) อาทิ การวิเคราะห์ข้อมูลจากการวิจัยที่มีการวัดซ้ำ การสูญหายของข้อมูลจะยังมีผลกระทบที่รุนแรงมากขึ้น (Heeringa. 2000; Roth. 1994: 537-560 อ้างถึงใน จำลอง วงษ์ประเสริฐ, 2554 : 2) ดังที่ Lynette and Jorgensen (2003: 429) ซึ่งกล่าวว่าข้อมูลสูญหายจะพบบ่อยในการศึกษา ข้อมูลหลายตัวแปร โดยเฉพาะการวัดซ้ำที่มีการวัดตัวแปรในหลายช่วงเวลาทำให้ตัวอย่างที่ใช้ในการเก็บรวบรวมข้อมูลได้รับความเสียหายและไม่สามารถอธิบายปรากฏการณ์ที่ต้องการศึกษาได้ เมื่อเกิดข้อมูลสูญหายเมตริกซ์ข้อมูลจะไม่สมบูรณ์ และส่งผลกระทบต่อ การวิเคราะห์ข้อมูล ซึ่งถ้าแก้ไขปัญหามูลสูญหายโดยการตัดตัวอย่างที่มีข้อมูลสูญหายออกไปและ Cismondi and Other (2013: 63) กล่าวว่า การสูญเสียจำนวนตัวอย่างและสูญเสียสารสนเทศของข้อมูลที่สูญหายอย่างมีนัยสำคัญอาจมีผลต่อประสิทธิภาพการทำงานของ การสร้างแบบจำลอง ทั้งนี้การวิเคราะห์ หลายตัวแปรนั้นหากตัวแปรบางตัวหรือหลายตัวมีปัญหา จะส่งผลไปยังตัวแปรอื่นและทำให้ ผลการวิเคราะห์มีปัญหาด้วย

นักวิชาการได้พัฒนาวิธีการในการจัดการกับข้อมูลสูญหายโดยการหาค่าที่เหมาะสม ในการหาค่าทดแทนข้อมูลสูญหาย ซึ่งในงานวิจัยนี้ผู้วิจัยจะเรียกวิธีการจัดการข้อมูลสูญหาย ว่า “การประมาณค่าข้อมูลสูญหาย”เช่น การแทนค่าข้อมูลสูญหาย ด้วยค่าเฉลี่ย (Mean substitution) ซึ่งเป็นวิธีที่ง่ายที่สุดแต่การแก้ไขปัญหาวด้วยแนวทางนี้จะทำให้ค่าประมาณ

ของข้อมูลสูญหายนั้นไม่เพียงตรง (Little and Rubin, 2002: 205) การประมาณค่าข้อมูลสูญหายโดยวิธีการถดถอย (Regression approach) ซึ่งเป็นวิธีการที่อาศัยข้อมูลในเมทริกซ์ข้อมูลซึ่งจะมีความยุ่งยากซับซ้อนขึ้นหรือการประมาณค่าข้อมูลสูญหายด้วยวิธีการวนซ้ำ (Expectation maximization approach: EM) แต่ข้อเสียของวิธีการประมาณค่าข้อมูลสูญหายด้วยสองวิธีหลังสุดนี้มีข้อเสียคือค่อนข้างยุ่งยากในการทำความเข้าใจ (Rencher and Christensen, 2012: 83) ผู้วิจัยจึงพยายามศึกษาวิธีการประมาณค่าข้อมูลสูญหายด้วยวิธีอื่นซึ่งวิธีการที่น่าสนใจคือ การประมาณค่าข้อมูล สูญหายด้วยวิธีการแทนค่าแบบสองทางด้วยความคลาดเคลื่อนแบบการแจกแจงแบบปกติ (Two-way imputation with normally distributed errors: TW-E) (Bernaards and Sijtsma, 2000: 321-364) โดยที่แนวคิดพื้นฐานของการประมาณค่าข้อมูลสูญหายด้วยวิธีนี้พิจารณาค่าเฉลี่ยแบบสองทางคล้ายกับการวิเคราะห์ความแปรปรวน มีข้อดีคืออาศัยความเรียบง่ายในการวิเคราะห์ข้อมูลและอาศัยแนวคิดของการหาค่าเฉลี่ยของแต่ละคนและการวัดซ้ำ (Ginkel, 2007: 4013-4027) แนวคิดนี้เป็นการประมาณค่าข้อมูลสูญหายที่อาศัยความเรียบง่ายและมีความคลาดเคลื่อนน้อย (Schafer and Schenker, 2000: 144-154) หากพิจารณาแนวคิดและวัตถุประสงค์ของการศึกษาที่มีการวัดซ้ำ พบว่ามีแนวคิดคล้ายกับการวิเคราะห์อนุกรมเวลาซึ่งพิจารณาพฤติกรรมในอดีตของตัวแปรที่ศึกษาเพื่ออธิบายพฤติกรรมในอนาคตของตัวแปร ทั้งนี้การวิเคราะห์อนุกรมเวลาโดยใช้ตัวแบบ Box-jenkis พิจารณาสหสัมพันธ์ในตนเองของชุดข้อมูลรูปแบบของสมการพยากรณ์จะถูกกำหนดขึ้นมาเอง ตามลักษณะของสหสัมพันธ์ในตนเองและสหสัมพันธ์ในตนเองส่วนย่อยของข้อมูลชุดนั้น (สุพรรณิ อึ้งปัญญาวงศ์, 2555: 125) ตัวแบบที่ง่ายและพิจารณาข้อมูลย้อนหลังที่ใกล้กับเวลาที่พยากรณ์มากที่สุด นั่นคือ ตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง (AR (1)) ซึ่งเป็นตัวแบบที่ค่าปัจจุบันขึ้นอยู่กับค่าในอดีต 1 หน่วยเวลาย้อนหลังและความคลาดเคลื่อนสุ่มในปัจจุบัน (วิจิต หล่อจิระชุมกุล และคณะ, 2539: 163) ดังนั้นการพยากรณ์โดยใช้ตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่งจึงได้รับความนิยม (วรารุทธิ์ พานิชกิจโกศลกุล, 2554: 2)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องทั้งในประเทศและต่างประเทศพบว่าเทคนิคในการประมาณค่าข้อมูลสูญหายแต่ละวิธีนั้น ไม่สามารถใช้ประมาณค่าข้อมูลสูญหายได้ทุกรูปแบบ วิธีในการประมาณค่าข้อมูลสูญหายจะมีลักษณะเฉพาะในการวิเคราะห์ และใช้กับลักษณะของข้อมูลที่แตกต่างกัน จุดประสงค์หลักของการประมาณค่าข้อมูลสูญหายทุกวิธี

คือการได้มาซึ่งค่าประมาณที่ดีที่สุดส่วนจะเลือกใช้วิธีใดในการประมาณค่านั้นขึ้นอยู่กับความแม่นยำ ความถูกต้อง ความสมบูรณ์และความคลาดเคลื่อนของการประมาณค่า หากนำค่าประมาณเดี่ยวจากแต่ละวิธีมาพิจารณาร่วมกันโดยหาตัวถ่วงน้ำหนักที่เหมาะสมสำหรับแต่ละวิธีซึ่งค่าประมาณใหม่ที่ได้นั้นจะให้ความคลาดเคลื่อนต่ำสุด เรียกค่าประมาณนั้นว่าค่าประมาณร่วม (Combined estimates) (เสาวณิต สุขภารังษ, 2546: 61) ซึ่งการประมาณค่าสิ่งที่สำคัญที่ต้องคำนึงถึงคือ ค่าประมาณที่ได้ต้องเป็นค่าประมาณที่ไม่เอนเอียงและมีความแปรปรวนต่ำสุด ซึ่งอาจพิจารณาจากการประมาณค่าร่วม โดยการนำวิธีประมาณค่าที่แตกต่างกันที่มีความเหมาะสมมาทำการประมาณค่าร่วมกันโดยอาศัยวิธีการทางสถิติ (จำลอง วงษ์ประเสริฐ, 2554: 38)

จากแนวคิดข้างต้น ผู้วิจัยจึงมีความสนใจที่จะพัฒนาวิธีการประมาณค่าข้อมูลสูญหายสำหรับการวัดซ้ำ โดยอาศัยการประมาณค่าร่วมแบบถ่วงน้ำหนักระหว่างการประมาณค่าข้อมูลสูญหายด้วยวิธีการแทนค่าแบบสองทางด้วยความคลาดเคลื่อนแบบการแจกแจงแบบปกติ (TW-E) และการประมาณค่าข้อมูลสูญหายโดยการวิเคราะห์อนุกรมเวลาโดยใช้ตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง (AR(1)) ซึ่งต่อไปนี้จะเรียกวิธีการประมาณค่าข้อมูลสูญหายสำหรับการวัดซ้ำที่พัฒนาขึ้นนี้ว่าการประมาณค่าข้อมูลสูญหายโดยการถ่วงน้ำหนักด้วยการประมาณค่าข้อมูลสูญหายด้วยวิธีการแทนค่าแบบสองทางและตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง (Weighted Two-way imputation Method and First-Order Autoregressive Model: IWTWAR(1))

วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาวิธีการประมาณค่าข้อมูลสูญหายสำหรับการวัดซ้ำด้วยวิธีการแทนค่าแบบสองทางและตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง
2. เพื่อศึกษาประสิทธิภาพของการประมาณค่าข้อมูลสูญหายโดยเปรียบเทียบกับ การประมาณค่าข้อมูลสูญหายแบบเดิมโดยใช้จำลองข้อมูล

วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้เป็นการพัฒนาวิธีการประมาณค่าข้อมูลสูญหายสำหรับการวัดซ้ำจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง มีรายละเอียดดังนี้

1. ประชากรและกลุ่มตัวอย่าง

การจำลองข้อมูลใช้โมเดลของการวัดคะแนนพัฒนาการทางการศึกษาโดยอาศัยแนวคิดโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง (Latent growth curve model) ทั้งแบบเชิงเส้นและไม่เชิงเส้น สำหรับวิธีการประมาณค่าข้อมูลสูญหายที่นำมาเปรียบเทียบกับประกอบด้วยการประมาณค่าข้อมูลสูญหายโดยใช้ค่าเฉลี่ย (MI) การประมาณค่าข้อมูลสูญหายด้วยวิธีการแทนค่าแบบสองทาง (TW-E) และการประมาณค่าข้อมูลสูญหายโดยใช้ตัวแบบอตรีเกรสซีฟอันดับที่หนึ่ง (AR (1))

2. เครื่องมือที่ใช้ในการวิจัย

การศึกษาประสิทธิภาพของวิธีการประมาณค่าข้อมูลสูญหายใช้ข้อมูลจากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ซิมูเลชันโดยใช้โปรแกรมสำเร็จรูป (โปรแกรม Mplus) ในการจำลองข้อมูล

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยทำการวิเคราะห์รูปแบบของการวิเคราะห์ข้อมูลระยะยาว และโมเดลของการวัดคะแนนพัฒนาการทางการศึกษา และแนวคิดเกี่ยวกับการประมาณค่าข้อมูลสูญหายจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่าวิธีการประมาณค่าข้อมูลสูญหายมีข้อจำกัดในการเลือกใช้วิธีการประมาณค่าข้อมูลสูญหาย การเลือกใช้วิธีการประมาณค่าข้อมูลสูญหายด้วยวิธีการใดวิธีการหนึ่งอาจจะให้ค่าประมาณที่ไม่เหมาะสม และหากสามารถใช้วิธีการประมาณค่าข้อมูลสูญหายร่วมกันสองวิธีที่เหมาะสมกับข้อมูลในลักษณะที่ต้องการวิเคราะห์จะทำให้ค่าประมาณที่ได้มีความแม่นยำสูงและมีคลาดเคลื่อนต่ำ (เสาวณิต สุขภารังสี, 2546: 61) โดยที่เรียกค่าประมาณที่ได้จากการประมาณค่า แต่ละวิธีว่า ค่าประมาณแบบเดี่ยว (Individual estimates) และเมื่อนำค่าประมาณเดี่ยวจากแต่ละวิธีมาพิจารณาร่วมกันโดยหาตัวถ่วงน้ำหนักที่เหมาะสมสำหรับแต่ละวิธีซึ่งค่าประมาณใหม่ที่ได้นั้นจะให้ความคลาดเคลื่อนต่ำสุด เรียกค่าประมาณนั้นว่า ค่าประมาณร่วม (Combined estimates) โดยข้อดีของการประมาณค่าข้อมูลสูญหายด้วยวิธีการถ่วงน้ำหนักคือ ไม่ต้องใช้สมมติฐานใด ๆ เกี่ยวกับ $f(Y_i^M | Y_i^o, X_i)$ โดยทำการปรับการวิเคราะห์โดยใช้น้ำหนักของข้อมูลที่สังเกตได้ (Garrett and James, 2011: 526) แนวทางในการพัฒนาวิธีการประมาณค่าข้อมูลสูญหายนั้นผู้วิจัยจะพิจารณาจากการประมาณค่าข้อมูลสูญหายโดยการถ่วงน้ำหนักของวิธีการประมาณค่าข้อมูลสูญหาย 2 วิธี ดังนี้

3.1 การประมาณค่าข้อมูลสูญหายด้วยวิธีการแทนค่าแบบสองทางด้วยความคลาดเคลื่อนแบบการแจกแจงแบบปกติ (Two-way imputation with normally distributed errors: TW-E) ซึ่งค่าประมาณข้อมูลสูญหายที่ได้จากวิธี TW-E จะพิจารณาข้อมูลทั้งอิทธิพลของหน่วยตัวอย่างอื่นของช่วงเวลาที่ทำกรประมาณค่าและอิทธิพลจำนวนการวัดซ้ำ ใช้ตัวแปรที่มีอยู่ในการอธิบายลักษณะของข้อมูลที่สูญหาย และค่าประมาณข้อมูลสูญหายได้จากตัวสถิติที่เป็นตัวสถิติที่มีความแปรปรวนและเป็นตัวประมาณไม่เอนเอียงค่าเฉลี่ยประชากร (Bernaards and Sijtsma, 2000: 321-364)

3.2 การประมาณค่าข้อมูลสูญหายโดยใช้ตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง (First-Order Autoregressive Model: AR (1)) ทั้งนี้เนื่องจากการวัดซ้ำตัวแปรที่สนใจศึกษาเพื่อพิจารณาการเปลี่ยนแปลงของปรากฏการณ์ต่างๆ ตามช่วงเวลาเปลี่ยนแปลงไป มีแนวคิดคล้ายกับการวิเคราะห์อนุกรมเวลาซึ่งมีแนวคิดพื้นฐานในการอธิบายปรากฏการณ์ของตัวแปร โดยอาศัยพฤติกรรมในอดีตของตัวแปรที่ศึกษาว่าควรจะเป็นเพียงพที่จะอธิบายพฤติกรรมในอนาคตของตัวแปรได้ (วิชิต หล่อจิระชุมกุล และ จิราวัลย์ จิตรถเวช, 2548: 1) ตัวแบบออโตรีเกรสซีฟแทนค่าปัจจุบันของอนุกรมเวลาในเทอมของความแปรปรวนปัจจุบันและในอดีต โดยตัวแบบที่ง่ายและพิจารณาข้อมูลย้อนหลังที่ใกล้กับเวลาที่พยากรณ์มากที่สุด นั่นคือตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่งเป็นปฏิบัติ (วิชิต หล่อจิระชุมกุล และคณะ, 2539: 163) กระบวนการ Autoregressive ที่ง่ายที่สุดและมีความสำคัญในทาง เป็นตัวแบบที่ค่าปัจจุบันขึ้นอยู่กับค่าในอดีต 1 หน่วยเวลาย้อนหลังและความคลาดเคลื่อนสุ่มในปัจจุบัน ทั้งนี้ผู้วิจัยเลือกตัวแบบ ตัวแบบออโตรีเกรสซีฟอันดับที่หนึ่ง (AR (1)) เนื่องจากการวัดซ้ำทางการศึกษานั้น จำนวนของการวัดซ้ำจะไม่มากและมีโอกาสที่จะเกิดข้อมูลสูญหายในช่วงต้นของการศึกษาได้ เช่น ช่วงเวลาที่ 2 หรือ 3 หากใช้ตัวแบบอัตสหสัมพันธ์ในอันดับที่ 2 หรือมากกว่าจะทำให้ไม่สอดคล้องกับบริบทของการเก็บรวบรวมข้อมูลทางการศึกษา ในการวิจัยครั้งนี้ใช้วิธีการประมาณค่าพารามิเตอร์คือวิธีการกำลังสองน้อยที่สุดซึ่งเป็น การประมาณค่าพารามิเตอร์ที่ทำให้ผลรวมค่ากำลังสองของความคลาดเคลื่อนภายใต้พารามิเตอร์ของตัวแบบมีค่าต่ำที่สุด และทำให้ค่าพยากรณ์ที่ได้จากการวิเคราะห์ มีความคลาดเคลื่อนต่ำ (So and Shin, 1999: 65-73)

4. การวิเคราะห์ข้อมูล

การศึกษาประสิทธิภาพของวิธีการประมาณค่าข้อมูลสูญหายสำหรับการวัดซ้ำมีรายละเอียดการดำเนินการดังนี้

4.1 การกำหนดเงื่อนไขสำหรับการจำลองข้อมูลโดยอาศัยเทคนิคมอนติคาร์โล ซิมูเลชัน ซึ่งผู้วิจัยใช้โปรแกรมสำเร็จรูปในการจำลองข้อมูล โดยอาศัยแนวคิดโมเดลโค้ง พัฒนาการที่มีตัวแปรแฝง โดยมีเงื่อนไขในการจำลองสถานการณ์ ดังนี้

4.1.1 ลักษณะของข้อมูลสูญหาย กำหนดให้มีลักษณะการสูญหายของข้อมูล แบบ Missing completely at random (MCAR)

4.1.2 จำนวนการวัดซ้ำ แบ่งเป็น 3 ครั้ง และ 6 ครั้ง

4.1.3 ร้อยละของข้อมูลสูญหาย คือ ร้อยละ 5, 10, 15 และ 20

4.1.4 รูปแบบของการเปลี่ยนแปลงคะแนนพัฒนาการ 2 แบบ คือ คะแนนพัฒนาการทางการศึกษาที่เป็นเส้นตรง และการเปลี่ยนแปลงคะแนนพัฒนาการทางการศึกษาที่ไม่ใช่เส้นตรง แบ่งออกเป็น 2 กรณี คือ

4.1.4.1 กรณีที่การเปลี่ยนแปลงคะแนนพัฒนาการของผู้เรียนเป็นเส้นตรง (Linear Growth Model: LIN Model) จะแบ่งออกเป็น 2 กรณี คือ

1) กรณีจำนวนวัดซ้ำ 3 ครั้ง กำหนดสัมประสิทธิ์พื้นฐานในโมเดลการวัด (B) เท่ากับ 0, 1 และ 2

2) กรณีจำนวนวัดซ้ำ 6 ครั้ง กำหนดสัมประสิทธิ์พื้นฐานในโมเดลการวัด (B) เท่ากับ 0, 0.5, 1.0, 1.5, 2.0 และ 2.5

4.1.4.2 กรณีที่การเปลี่ยนแปลงคะแนนพัฒนาการของผู้เรียนไม่ใช่เส้นตรง (Fixed Curve Parameter Growth Model: FIC Model) แบ่งออกเป็น 2 กรณี คือ

1) กรณีจำนวนวัดซ้ำ 3 ครั้ง กำหนดสัมประสิทธิ์พื้นฐานในโมเดลการวัด (B) เท่ากับ 0, 1 และ 2.5

2) กรณีจำนวนวัดซ้ำ 6 ครั้ง กำหนดสัมประสิทธิ์พื้นฐานในโมเดลการวัด (B) เท่ากับ 0, 0.15, 0.60, 1.50, 1.80 และ 2.5

4.1.5 ตัวแปรแฝงระดับคะแนนในการวัดครั้งแรก แบ่งเป็น 3 ระดับ คือ 20 คะแนน 40 คะแนน และ 80 คะแนน

4.1.6 ตัวแปรแฝงความชื้นหรืออัตราพัฒนาการ แบ่งเป็น 2 ระดับ คือ 0.5 และ 2.5 (Hertzog, 2008; John & Kristen, 2010; David & Shannon, 2011)

4.1.7 จำนวนหน่วยตัวอย่างในแต่ละสถานการณ์ การจำลองข้อมูลในแต่ละสถานการณ์ใช้จำนวนหน่วยตัวอย่าง 1,000 หน่วย

4.2 สถิติที่ใช้ในการเปรียบเทียบประสิทธิภาพของข้อมูลสูญหายจะ คำนวณค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Mean squared error: MSE)

4.3 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าข้อมูลสูญหายทั้ง 4 วิธี ด้วยการวิเคราะห์ความแปรปรวนทางเดียว (One Way ANOVA) และเปรียบเทียบค่าเฉลี่ยรายคู่ด้วยวิธีของ Tukey's Method) เพื่อตรวจสอบว่าวิธีการประมาณค่าข้อมูลสูญหายวิธีการใดบ้างที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติกับวิธีการประมาณค่าข้อมูลสูญหายที่มีประสิทธิภาพสูงสุดในแต่ละเงื่อนไขของการวิจัย

สรุปผลการวิจัย

1. ผลการพัฒนาวิธีการประมาณค่าข้อมูลสูญหาย

จากการพิจารณาแนวคิดในการประมาณค่าข้อมูลสูญหายด้วยวิธี TW-E และ AR(1) ผู้วิจัยจะนำมาพัฒนาเพื่อนำเสนอวิธีการประมาณค่าข้อมูลสูญหายโดยการถ่วงน้ำหนักด้วยการประมาณค่าข้อมูลสูญหายด้วย IWTWAR (1) โดยพิจารณาข้อมูลที่มีการวัดซ้ำ ให้ Y_{ij} แทนผลที่ได้จากการวัดตัวแปรสำหรับคนที่ i ที่เวลา j โดยที่ $i = 1, 2, \dots, n$ และ $j = 1, 2, \dots, r$ โดยกำหนดให้ $y_{n1}, y_{n2}, \dots, y_{nr}$ เป็นค่าที่สังเกตได้ของคนที่ n ส่วนค่า y_{nt+1} เป็นค่าข้อมูลสูญหายที่ต้องประมาณค่า มีขั้นตอนดังนี้

ขั้นตอนที่ 1 การประมาณค่าข้อมูลสูญหายด้วย TW-E ดังนี้

1) กรณีที่มีข้อมูลสูญหาย ตำแหน่ง y_{nt+1} จากข้อมูลที่สมบูรณ์ของคนที่ n นั่นคือ $y_{n1}, y_{n2}, \dots, y_{nr}$ นำมาหาค่าเฉลี่ยของคะแนนของคนที่ n (อิทธิพลของจำนวนการวัดซ้ำ) ดังนี้

$$\bar{y}_{i\cdot} = \frac{\sum_{j=1}^r Y_{ij}}{r-1}$$

2) จากข้อมูลที่สมบูรณ์ในซ้ำที่ t นั่นคือ $y_{1t}, y_{2t}, \dots, y_{nt}$ นำมาหาค่า $\bar{y}_{\cdot j}$ เป็นค่าเฉลี่ย ณ เวลาที่ t หรือซ้ำที่ t นั้นเอง (อิทธิพลของคนในการวัดซ้ำ ณ เวลา t) ดังนี้

$$\bar{y}_{\cdot j} = \frac{\sum_{i=1}^n Y_{ij}}{n-1}$$

3) จากข้อมูลที่สมบูรณ์ $y_{11}, y_{12}, \dots, y_{nt-1}$ นำมาหาค่า $\bar{y}_{..}$ ซึ่งเป็นค่าเฉลี่ยรวมดังนี้

$$\bar{y}_{..} = \frac{\sum_{i=1}^{n-1} \sum_{j=1}^{r-1} Y_{ij}}{(n-1)(r-1)}$$

4) ประมาณค่าข้อมูลสูญหายสำหรับคะแนนที่ขาดหายไปในเซลล์ $(n, t+1)$ ด้วยวิธี TW-E ดังนี้

$$\hat{Y}_{nt+1} = \bar{y}_{i\cdot} + \bar{y}_{\cdot j} - \bar{y}_{..}$$

กำหนดให้ $\hat{Y}_{nt+1}$ เป็นค่าประมาณข้อมูลสูญหายที่ได้จากการประมาณค่าด้วยวิธี TW-E ต่อไปนี้จะเขียนแทนด้วย $\hat{Y}_{TW-E}$ ค่าประมาณข้อมูลสูญหายได้จากตัวสถิติที่เป็นตัวสถิติที่มีความแกร่ง (Robust) และเป็นตัวประมาณไม่เอนเอียง (Unbias) ของค่าเฉลี่ยของประชากร (Schafer and Schenker, 2000: 105)

ขั้นตอนที่ 2 ประมาณค่าข้อมูลสูญหายด้วยวิธี AR (1) มีขั้นตอนดังนี้

1) ประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด (So and Shin, 1999: 65-73) ดังนี้

$$\hat{\phi} = \frac{\sum_{j=2}^t (Y_{nj} - \bar{Y})(Y_{nj-1} - \bar{Y})}{\sum_{j=2}^t (Y_{nj-1} - \bar{Y})^2}$$

2) นำค่าประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแทนค่าในสมการประมาณค่าข้อมูลสูญหายในหน่วยตัวอย่างคนที่ n ในเวลาที่ $t+1$ จะได้สมการประมาณค่าดังนี้

$$\hat{Y}_{nt+1} = \bar{Y} + \hat{\phi}(Y_{nt} - \bar{Y})$$

โดยที่ $\hat{Y}_{nt+1}$ แทนค่าพยากรณ์ของข้อมูล ณ เวลาที่ $t+1$

$\hat{\phi}$ แทนค่าพารามิเตอร์ของตัวแบบ

Y_{nt} แทนค่าที่สังเกตได้ ณ เวลา t

$\bar{Y}$ แทนค่าเฉลี่ยของข้อมูลที่สมบูรณ์ของคนที่ n

จะได้ค่าประมาณของข้อมูลสูญหายด้วยวิธีการวิเคราะห์หอนุกรมต่อไปนี้ผู้วิจัย

จะเขียนแทนด้วย $\hat{Y}_{AR(1)}$

ขั้นตอนที่ 3 นำค่าประมาณของข้อมูลสูญหายในขั้นตอนที่ 1 และขั้นตอนที่ 2 มาถ่วงน้ำหนักเพื่อให้ค่าประมาณที่ได้มีความแปรปรวนต่ำสุด

$$\hat{Y}_{IWTWAR(1)} = w_k \hat{Y}_{TW-E} + (1 - w_k) \hat{Y}_{AR(1)}$$

เมื่อ w_k เป็นค่าถ่วงน้ำหนักระหว่าง $\hat{Y}_{TW-E}$ และ $\hat{Y}_{AR(1)}$ โดยที่ $0 \leq w_k \leq 1$

นำค่า $\hat{Y}_{IWTWAR(1)}$ แทนข้อมูลสูญหาย จากนั้นนำค่าแทนข้อมูลสูญหายที่ได้จากขั้นตอนที่ 3 รวมเข้ากับข้อมูลที่สมบูรณ์ ทำซ้ำขั้นตอนที่ 1 ถึง 3 จนครบทุกค่าของข้อมูล

2. ผลการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าข้อมูลสูญหาย

2.1 กรณีรูปแบบของคะแนนพัฒนาการเป็นเส้นตรงผลการศึกษาพบว่า

2.1.2 กรณีที่มีจำนวนการวัดซ้ำเท่ากับ 3 ครั้ง การประมาณค่าข้อมูลสูญหายด้วยวิธี AR (1) จะมีประสิทธิภาพสูงสุด รองลงมาคือ วิธี IWTWAR (1) และประสิทธิภาพของวิธี AR (1) แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับวิธีการประมาณค่าข้อมูลสูญหายด้วยวิธีการประมาณค่าแบบอื่น

2.1.2 กรณีที่มีจำนวนการวัดซ้ำเท่ากับ 6 ครั้ง ผลการศึกษาพบว่า เมื่อคะแนนพัฒนาการมีอัตราพัฒนาการเท่ากับ 0.5 การประมาณค่าข้อมูลสูญหายด้วยวิธี IWTWAR (1) มีประสิทธิภาพสูงสุด และเมื่อมีหน่วยตัวอย่างขนาด 10 หน่วย จะมีประสิทธิภาพแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับวิธี AR (1) และวิธี MI และเมื่อมีขนาดตัวอย่างเท่ากับ 20 หน่วยและ 50 หน่วย วิธี IWTWAR(1) จะมีประสิทธิภาพแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับวิธี TW-E และวิธี MI และเมื่อคะแนนพัฒนาการ

มีอัตราพัฒนาการเท่ากับ 2.5 ขนาดตัวอย่างจำนวน 10 หน่วย วิธี IWTWAR(1) มีประสิทธิภาพสูงสุดและมีประสิทธิภาพแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับวิธีอื่นเมื่อขนาดตัวอย่างจำนวน 20 หน่วย และ 50 หน่วย วิธี AR(1) มีประสิทธิภาพสูงสุดและมีประสิทธิภาพแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับ IWTWAR (1)

2.2 กรณีรูปแบบของคะแนนพัฒนาการไม่เป็นเส้นตรง ผลการศึกษาพบว่า

2.2.1 กรณีที่มีจำนวนการวัดซ้ำเท่ากับ 3 ครั้ง การประมาณค่าด้วยวิธี TW-E และวิธี AR (1) จะมีประสิทธิภาพสูงสุด และมีประสิทธิภาพแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับวิธีอื่น

2.2.2 กรณีที่มีจำนวนการวัดซ้ำเท่ากับ 6 ครั้ง ผลการศึกษาพบว่า คะแนนพัฒนาการมีอัตราพัฒนาการเท่ากับ 0.5 และ 2.5 เมื่อมีขนาดตัวอย่างเท่ากับ 10 หน่วย และ 20 หน่วย วิธี IWTWAR (1) มีประสิทธิภาพสูงสุด และมีประสิทธิภาพแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับวิธีอื่นเมื่อมีขนาดตัวอย่าง 20 หน่วย และ 50 หน่วย วิธี IWTWAR (1) มีประสิทธิภาพสูงสุด แต่มีประสิทธิภาพสูงสุดแต่มีประสิทธิภาพไม่แตกต่างกับวิธีอื่น เมื่อมีคะแนนพัฒนาการมีอัตราพัฒนาการเท่ากับ 2.5 และมีขนาดตัวอย่าง 10 หน่วย วิธี IWTWAR (1) มีประสิทธิภาพสูงสุด และมีประสิทธิภาพแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 กับวิธีอื่นแต่เมื่อมีขนาดตัวอย่างเท่ากับ 50 หน่วย วิธี TW-E และวิธี AR (1) จะมีประสิทธิภาพสูงสุดแต่มีประสิทธิภาพไม่แตกต่างกับวิธีอื่น

หมายเหตุ: รายละเอียดของผลการวิเคราะห์เชิงสถิติให้ดูในเล่มดัชนีพนธ์มหาวิทยาลัยมหาสารคาม ชื่อเรื่องการพัฒนาวิธีการประมาณค่าข้อมูลสูญหายสำหรับการวัดซ้ำ

อภิปรายผล

1. วิธีการประมาณค่าข้อมูลสูญหายที่ผู้วิจัยพัฒนาขึ้นมีประสิทธิภาพสูงสุดเมื่อจำนวนการวัดซ้ำเท่ากับ 6 ครั้ง ขนาดตัวอย่างเท่ากับ 10 หน่วย และ 20 หน่วย มีประสิทธิภาพแตกต่างกับวิธีการประมาณค่าวิธีอื่น เนื่องจากการพัฒนาวิธีการประมาณค่าที่ผู้วิจัยพัฒนาขึ้นอาศัยแนวคิดของตัวแบบของการวิเคราะห์ข้อมูลระยะยาว และโมเดลของการวัดคะแนนพัฒนาการทางการศึกษา โดยนักวิจัยและนักวัดผลมีความพยายามที่จะประมาณค่าตัวแปรตาม (Y_{ij}) โดยใช้สารสนเทศที่ได้จากการวัดตัวแปรอิสระ (t_{ij}) โดยการประมาณค่าพารามิเตอร์ซึ่งประกอบด้วยจุดตัดของตัวแบบเชิงเส้น และแนวโน้มของตัวแบบเชิงเส้น ซึ่งจุดตัดของตัว

แบบเชิงเส้นสามารถประมาณค่าได้จากการวิเคราะห์เส้นถดถอยของการวัดซ้ำของแต่ละบุคคล และแนวโน้มของตัวแบบเชิงเส้นสามารถประมาณค่าได้จากการวิเคราะห์ตัวแปรร่วม (Garrett and James, 2011: 84) ทั้งนี้ในการพัฒนาวิธีการประมาณค่าข้อมูลสูญหายที่ผู้วิจัยพัฒนาขึ้น อาศัยการพิจารณาข้อมูลทั้งอิทธิพลของตัวอย่างอื่นของช่วงเวลาที่ทำการประมาณค่าและอิทธิพลจำนวนการวัดซ้ำ ใช้ตัวแปรที่มีอยู่ในการอธิบายลักษณะของข้อมูลที่สูญหายและค่าประมาณข้อมูลสูญหายได้จากตัวสถิติที่เป็นตัวสถิติที่มีความแข็งแกร่งและเป็นตัวประมาณไม่เอนเอียงค่าเฉลี่ยประชากร (Bernaards and Sijtsma, 2000: 321-364) ร่วมกับการพิจารณาการเปลี่ยนแปลงของปรากฏการณ์ต่างๆ ตามช่วงเวลาที่เปลี่ยนแปลงไปโดยอาศัยพฤติกรรมในอดีตของตัวแปรที่ศึกษาว่าควรจะมีเพียงพที่จะอธิบายพฤติกรรมในอนาคตของตัวแปรได้ (วิชิต หล่อจิระชุนกุล และจิราวัลย์ จิตรถเวช, 2548: 1) เมื่อมีการถ่วงน้ำหนักระหว่างวิธีการประมาณค่าสองค่าด้วยกัน และวิธีการประมาณค่านั้นเหมาะสมกับข้อมูลในลักษณะที่ต้องการวิเคราะห์ จะทำให้ค่าประมาณที่ได้มีความแม่นยำสูง และมีคลาดเคลื่อนต่ำ (เสาวณิต สุขภารังสี, 2546: 61) ด้วยสาเหตุข้างต้น ทำให้วิธีการประมาณค่าข้อมูลสูญหายที่ผู้วิจัยพัฒนาขึ้น มีประสิทธิภาพสูงในกรณีที่มีจำนวนการวัดซ้ำ 6 ครั้ง และจำนวนตัวอย่างน้อย เนื่องจาก ทั้งนี้เมื่อขนาดของตัวอย่างมีจำนวนเพิ่มขึ้น วิธี IWTSAR (1) ที่ผู้วิจัยพัฒนาขึ้นมีประสิทธิภาพลดลง ทั้งนี้เนื่องจากแนวคิดในการพัฒนาวิธีการประมาณค่าผู้วิจัยได้นำเอาวิธีการประมาณค่าแบบ TW-E มาถ่วงน้ำหนักร่วมกับการประมาณค่าข้อมูลสูญหายด้วยวิธี AR(1) ซึ่งแนวคิดของวิธี TW-E ให้ผลลัพธ์ที่เที่ยงตรงในการวิเคราะห์สัมประสิทธิ์อัลฟาของครอนบาค ค่าเฉลี่ยกำลังสองในการวิเคราะห์ความแปรปรวนแต่มีความเอนเอียง (Bias) ในการทดสอบคะแนนเฉลี่ย การวิเคราะห์ค่าน้ำหนักปัจจัย (Factor loadings) การวิเคราะห์องค์ประกอบหลัก (Components analysis) แต่วิธี TW-E จะส่งผลให้มีส่วนเบี่ยงเบนมาตรฐานมากกว่าข้อมูลที่สมบูรณ์เล็กน้อยและก่อให้เกิดผลลัพธ์ที่มีความลำเอียงมากขึ้นโดยเฉพาะในกรณีที่ร้อยละของข้อมูลสูญหายมีจำนวนมาก และขนาดตัวอย่างจำนวนเพิ่มขึ้นเนื่องจากสัดส่วนการเพิ่มของส่วนเบี่ยงเบนมาตรฐาน (Ginkel, 2007: 4013-4027) ดังนั้นจึงเป็นเหตุให้วิธีที่ผู้วิจัยพัฒนาขึ้น (IWTSAR (1)) มีประสิทธิภาพสูงสุดในบางกรณี โดยเฉพาะในกรณีที่ร้อยละของข้อมูลสูญหายน้อย เช่น ร้อยละ 5 และร้อยละ 10 ซึ่งสอดคล้องกับผลการศึกษาของ จำลอง วงษ์ประเสริฐ (2554: 304-320) ได้ศึกษาการพัฒนาวิธีการประมาณข้อมูลสูญหายโดยการถ่วงน้ำหนักแบบวนซ้ำด้วยวิธีของแจ๊คไนฟ์และ

การวิเคราะห์การถดถอย (ไอบีบลิวเจอร์) ซึ่งพบว่าการประมาณค่ามีประสิทธิภาพสูงที่สุดในบางเงื่อนไขของการศึกษา

2. วิธีการประมาณค่า AR (1) มีประสิทธิภาพสูงที่สุดเมื่อมีจำนวนการวัดซ้ำ 3 ครั้ง และกรณีที่ขนาดตัวอย่าง 50 หน่วย เนื่องจากมีแนวคิดพื้นฐานในการอธิบายปรากฏการณ์ของตัวแปร โดยอาศัยพฤติกรรมในอดีตของตัวแปรที่ศึกษาว่าควรจะเพียงพอที่จะอธิบายพฤติกรรมในอนาคตของตัวแปรได้ (วิจิต หล่อจิระขุนทด และจิราวัลย์ จิตรถเวช, 2548: 1) ซึ่งวิธี AR (1) จะประมาณค่าข้อมูลสูญหายโดยแทนค่าปัจจุบันของการวัดในเทอมของความแปรปรวนปัจจุบันและในอดีต โดยตัวแบบที่ง่ายและพิจารณาข้อมูลย้อนหลังที่ใกล้กับเวลาที่ประมาณค่าข้อมูลสูญหายมากที่สุดคือตัวแบบ AR (1) เป็นกระบวนการ Autoregressive ที่ง่าย (วิจิต หล่อจิระขุนทด และคณะ, 2539: 163) โดยที่ผู้วิจัยใช้วิธีการประมาณค่าพารามิเตอร์คือ วิธีกำลังสองน้อยที่สุด ซึ่งเป็นการประมาณค่าพารามิเตอร์ที่ทำให้ผลรวมค่ากำลังสองของความคลาดเคลื่อนภายใต้พารามิเตอร์ของตัวแบบมีค่าต่ำที่สุดและทำให้ค่าพยากรณ์ที่ได้จากการวิเคราะห์มีความคลาดเคลื่อนต่ำ (So and Shin, 1999: 65-73) วิธี MI และวิธี TW-E ไม่พบว่ามีประสิทธิภาพสูงที่สุดในทุกเงื่อนไข ทั้งนี้เนื่องจากเมื่อเทียบกับวิธีการประมาณค่าทั้ง 4 วิธี MI และ TW-E ใช้สารสนเทศน้อยที่สุดในการประมาณค่าข้อมูลสูญหายเมื่อเทียบกับ IWTWAR (1) และ AR (1)

เอกสารอ้างอิง

จำลอง วงษ์ประเสริฐ. (2554). การพัฒนาวิธีการประมาณข้อมูลสูญหายโดยการถ่วงน้ำหนักแบบวนซ้ำด้วยวิธีของแจ๊คไนฟ์และการวิเคราะห์การถดถอย. ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, มหาวิทยาลัยมหาสารคาม.

Wongprasert, S. (2011). The development of the method of estimating the lost data by loop weighted using the method of jackknife and regression analysis. Doctorate of Philosophy, Maha Sarakham University. (in Thai).

วราฤทธิ์ พานิชกิจโกศลกุล. (2554). การเปรียบเทียบวิธีการพยากรณ์ข้อมูลถ่วงน้ำหนักแบบเวลาสำหรับตัวแบบอัตตสหสัมพันธ์อันดับที่หนึ่ง. วารสารวิทยาศาสตร์ ม. อุบลฯ ฉบับพิเศษ, 1(มกราคม-ธันวาคม), 1-9.

- Phanitchakitkosolkun, W. (2011). A comparison of the one time period forward forecasting method for first order correlation models. **Special edition Ubon Journal of Science Journal**, 1 (January-December), 1–9. (in Thai).
- วิจิต หล่อจิระชุมกุล. (2539). **เทคนิคการพยากรณ์**. กรุงเทพมหานคร: สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- Lorjirunhoonkul, W. (1996). **Forecasting techniques**. Bangkok: National Institute of Development Administration. (in Thai).
- วิจิต หล่อจิระชุมกุล และจิราวัลย์ จิตรถเวช. (2548). **เทคนิคการพยากรณ์**. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- Lorjirunhoonkul, W., & Jitaravej, J. (2005). **Forecasting Techniques**. 3rd edition, Bangkok: The National Institute of Development Administration. (in Thai).
- สุพรรณิ อึ้งปัญสัตวงศ์. (2555). **เทคนิคการพยากรณ์เชิงสถิติ**. ขอนแก่น: คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.
- Ungpunsatwong, W. (2012). **Statistical forecasting techniques**. Khon Kaen: Faculty of Science, Khon Kaen University. (in Thai).
- Allison. (2012). Handling Missing Data by Maximum Likelihood. **Statistics and Data Analysis**, 312(2012), 1145–1152.
- Bernaards, C.A., & Sijtsma, K. (2000). Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. **Multivariate Behav**, 3(5), 321–364.
- Cismondi & other. (2013). Weighted Nearest Neighbor and Regression Imputation. **In Paper Presented at the Meeting of 9th Asia-Pacific Decision Sciences Institute**. Korea: APDSI-KOPOMS Seoul.
- DeSilvio, M. (2009). A Variance Ratio Statistics for Assessing the Missing Data Mechanism: An Empirical Study. **Dissertation Abstracts International**, 63(1), 5698–A.
- Garrett, M, F., Nan M.L., & James. H. W. (2011). **Applied Longitudinal Analysis**. New Jersey: Wiley& sons.

- Ginkel, J. R. V. (2007). Two-way imputation: A Bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation. **Computational Statistics & Data Analysis**, 51(2007), 4013 – 4027.
- Heeringa, S. G. (2000). Multivariate Imputation of Coarsened Survey on Household Wealth. **Diddertation Abstracts International**, 61(1), 1112–1120.
- Little, R. J. and Rubin, D. (2002). Statistical Analysis with Missing Data (2nd ed.). New York: John Wiley and Sons.
- Lynette H. & Murray J. (2003). Mixture model clustering for mixed data with missing information. **Computational Statistics & Data Analysis**, 41(2003), 429 – 440.
- Rencher & Christensen. (2012). **Method of Multivariate Analysis** (3rd ed.). New York: Wiley Series.
- Schafer, J. L., & Schenker, N. (2000). Inference with Imputed Conditional Means. **Journal of the American Statistical Association**, 95(March), 144–154.
- So, B. S., & Shin, D. (1999). Recursive Mean Adjustment in Time Series Inferences. **Statistics and Probability Letters**, 4(3), 65–73.