บทที่ 5

บทสรุป

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี ตรวจสอบความแม่นยำและอำนาจการทดสอบที่ได้จากวิธีการจัดการข้อมูลสูญหาย วิธีการสุ่มตัวอย่าง ความสัมพันธ์ระหว่างตัวแปร และจำนวนข้อมูลสูญหายที่แตกต่างกัน และศึกษาปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร ที่มีต่อความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ โดยมีตัวแปรและเงื่อนไขที่ใช้ในการวิจัยดังต่อไปนี้

1. ตัวแปรที่ศึกษามีดังต่อไปนี้

1.1 ตัวแปรอิสระ

1.1.1 วิธีการสุ่ม จำแนกเป็น 3 วิธี คือ

1.1.1.1 การสุ่มแบบแบ่งชั้น

1.1.1.2 การสุ่มแบบกลุ่ม

1.1.1.3 การสุ่มแบบหลายขั้นตอน

1.1.2 วิธีการจัดการข้อมูลสูญหาย จำแนกเป็น 3 วิธี คือ

1.1.2.1 การตัดข้อมูลออกแบบลิสท์ไวส์ (Listwise deletion)

1.1.2.2 การแทนค่าข้อมูลด้วยวิธีอีเอ็ม (Em algorithm or Expectation maximization)

1.1.2.3 การแทนค่าข้อมูลด้วยวิธีอีพีเอสเอสอี (Estimated parameter and Smallest standard error)

1.1.3 ความสัมพันธ์ระหว่างตัวแปร 3 ลักษณะ คือ ความสัมพันธ์ระดับต่ำ (r=.30) ความสัมพันธ์ระดับปานกลาง (r=.50) และความสัมพันธ์ระดับสูง (r=.70)

1.1.4 จำนวนข้อมูลสูญหาย 5% 10% 20% 30%

1.2 ตัวแปรตาม คือ ความแม่นยำ และอำนาจการทดสอบ

2. ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลที่ได้จากการจำลองสถานการณ์โดยใช้เทคนิคมอนติ คาร์โล ซิมูเลชั่น (Monte Carlo Simulation Technique))

3. ความแม่นยำ หมายถึง ความใกล้เคียงกันระหว่างค่าสถิติกับค่าพารามิเตอร์ พิจารณาเฉพาะค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์

4. อำนาจการทดสอบคำนวณจากการใช้สถิติทดสอบที (t-test) การวิเคราะห์ความแปรปรวนและการทดสอบสหสัมพันธ์

5. การวิจัยครั้งนี้ใช้ข้อมูลที่มีลักษณะการแจกแจงแบบปกติสองตัวแปร (Bivariate normal distribution) โดยกำหนดให้ตัวแปรเกณฑ์ (Y) คือ เกรดเฉลี่ยของนักเรียน และตัวแปรทำนาย (X) คือ คะแนนสอบของนักเรียน และกำหนดให้มีความสัมพันธ์ต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70)

6. ข้อมูลสูญหายที่จะศึกษาในครั้งนี้จะศึกษาเฉพาะตัวแปรเกณฑ์ (Y) คือ เกรดเฉลี่ยของนักเรียน

7. การประมาณขนาดตัวอย่างเป็นไปตามวิธีของนีย์แมน

8. เกณฑ์ในการเสนอผลการเปรียบเทียบว่าวิธีการจัดการข้อมูลสูญหาย วิธีการสุ่มตัวอย่าง ความสัมพันธ์ระหว่างตัวแปร และจำนวนข้อมูลสูญหายที่ต่างกัน วิธีใดดีที่สุดใช้ค่าเฉลี่ยกำลังสองของความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และค่าสัมประสิทธิ์สหสัมพันธ์ถ้าดัชนีดังกล่าวมีค่าเท่ากัน วิธีที่ให้ค่าอำนาจการทดสอบสูงกว่าจะถือว่าเป็นวิธีที่ดีที่สุด

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง ผู้วิจัยจำลองสถานการณ์โดยใช้เทคนิค มอนติคาร์โล ซิมูเลชั่น (Monte Carlo Simulation Technique) ด้วยเครื่องคอมพิวเตอร์จุลภาค เขียนโปรแกรมด้วยภาษาฟอกโปร (FOXPRO) ได้ข้อมูลผลสัมฤทธิ์ทางการเรียน ประกอบด้วย เกรดเฉลี่ยเป็นตัวแปรเกณฑ์ (Y) และตัวทำนายคือคะแนนสอบของนักเรียน (X) และผู้วิจัยได้สร้างประชากรขึ้นมาอีก 2 กลุ่ม เพื่อใช้ในการศึกษาค่าอำนาจการทดสอบโดยให้มีค่าเฉลี่ยของเกรดเฉลี่ยมากกว่าค่าเฉลี่ยของเกรดเฉลี่ยจากประชากรในการจำลองสถานการณ์ครั้งแรกเท่ากับ 1σ และน้อยกว่าค่าเฉลี่ยของเกรดเฉลี่ยจากประชากรในการจำลองครั้งแรกเท่ากับ -1σ แล้วจึงนำข้อมูลจากประชากรทั้งหมดมากำหนดว่าแต่ละคนเป็นกลุ่มใดและระดับชั้นใด

ดำเนินการวิเคราะห์ข้อมูลจากประชากรกลุ่มแรกที่มีความสัมพันธ์ต่ำ (r=.30) โดยหาค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ แล้วจึงสุ่มตัวอย่างโดยใช้คอมพิวเตอร์สุ่มจากประชากรกลุ่มนี้ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวน 350 คน สร้างข้อมูลสูญหายแบบสุ่ม (randomly missing data) โดยการเขียนคำสั่งคอมพิวเตอร์ให้ลบข้อมูลเกรดเฉลี่ยออกไปเท่ากับ 5% ดำเนินการจัดกระทำข้อมูลสูญหายโดยการตัดข้อมูลสูญหายออก

แบบลิสท์ไวส์ คำนวณหาค่าเฉลี่ย ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ และคำนวณความแม่นยำโดยพิจารณาจากความแตกต่างของค่าเฉลี่ยเลขคณิต ความแปรปรวนและสัมประสิทธิ์สหสัมพันธ์ ที่ได้จากการสุ่มตัวอย่างแบบแบ่งชั้นและตัดข้อมูลสูญหายออกแบบลิสท์ไวส์กับค่าที่คำนวณจากประชากร หลังจากนั้นจึงคำนวณอำนาจการทดสอบจากการทดสอบความสัมพันธ์

เปลี่ยนกลุ่มประชากรเป็นกลุ่มที่สองซึ่งมีค่าเฉลี่ยของเกรดเฉลี่ยมากกว่าประชากรกลุ่มแรกเท่ากับ 1σ สุ่มตัวอย่างโดยใช้คอมพิวเตอร์สุ่มจากประชากรกลุ่มนี้ด้วยวิธีการสุ่มแบบแบ่งชั้น จำนวน 350 คน สร้างข้อมูลสูญหายแบบสุ่ม (randomly missing data) จำนวน 5% จัดกระทำข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ แล้วคำนวณอำนาจการทดสอบจากการทดสอบที (t-test) โดยใช้ข้อมูลจากการสุ่มประชากรกลุ่มแรกและประชากรกลุ่มที่สอง เปลี่ยนกลุ่มประชากรเป็นกลุ่มที่สามซึ่งมีค่าเฉลี่ยของเกรดเฉลี่ยน้อยกว่าประชากรกลุ่มแรกเท่ากับ -1σ สุ่มตัวอย่างโดยใช้การสุ่มแบบแบ่งชั้น จำนวน 350 คน สร้างข้อมูลสูญหายแบบสุ่ม จำนวน 5% แล้วจึงตัดข้อมูลสูญหายออกแบบลิสท์ไวส์ คำนวณอำนาจการทดสอบจากการทดสอบที (t-test) โดยใช้ประชากรกลุ่มที่สองกับกลุ่มที่สามมาเปรียบเทียบกัน และคำนวณอำนาจการทดสอบจากการทดสอบเอฟ (F-test) หรือการวิเคราะห์ความแปรปรวน โดยใช้ข้อมูลจากการสุ่มประชากรกลุ่มแรก กลุ่มที่สอง และกลุ่มที่สาม มาเปรียบเทียบกัน ดำเนินการทำซ้ำแบบเดิมดังที่กล่าวมาจำนวน 1,000 ครั้ง แล้วจึงคำนวณหาค่าเฉลี่ยกำลังสองความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ คำนวณอำนาจการทดสอบโดยพิจารณาสัดส่วนจากจำนวนครั้งของการปฏิเสธสมมุติฐานสูญของการทดสอบความสัมพันธ์ การทดสอบที (t-test) และการทดสอบเอฟ (F-test) กับจำนวนครั้งของการทดสอบทั้งหมด

ดำเนินการวิเคราะห์ข้อมูลในลักษณะเดียวกัน โดยกำหนดความสัมพันธ์ระหว่างตัวแปรเท่ากับ .50 และ .70 ใช้วิธีการสุ่มตัวอย่างแบบกลุ่มและแบบหลายขั้นตอน จำนวนข้อมูลสูญหายเท่ากับ 10% 20% และ 30% วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี ได้สถานการณ์จำลอง 108 เงื่อนไขการทดลอง ในแต่ละสถานการณ์จะมีข้อมูล 1,000 กรณี นำข้อมูลมาเปรียบเทียบกันโดยทำการทดสอบความแตกต่างของค่าเฉลี่ยความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ จำแนกตามวิธีการสุ่มตัวอย่าง จำนวนข้อมูลสูญหาย ความสัมพันธ์ระหว่างตัวแปร และวิธีการจัดการข้อมูลสูญหาย โดยการวิเคราะห์ความแปรปรวนแบบทางเดียว (One Way ANOVA) เมื่อทดสอบแล้วพบว่ามีนัยสำคัญทางสถิติได้วิเคราะห์เปรียบเทียบความแตกต่างของค่าเฉลี่ยเป็นรายคู่โดยใช้วิธี

ของทูกี้ (Tukey) หลังจากนั้นผู้วิจัยใช้คอมพิวเตอร์สุ่มตัวอย่างจำนวน 100 กรณี ในแต่ละเงื่อนไขการทดลองนำข้อมูลทั้งหมดมารวมกันได้ข้อมูล 10,800 กรณี แล้วจึงดำเนินการวิเคราะห์ปฏิสัมพันธ์ระหว่างวิธีการวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร ที่มีต่อความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ โดยการวิเคราะห์ความแปรปรวน 4 ทาง (Four-Way ANOVA) เมื่อเกิดปฏิสัมพันธ์ระหว่างตัวแปร ผู้วิจัยได้ค้นหาคำตอบเกี่ยวกับการเกิดปฏิสัมพันธ์โดยการวิเคราะห์ Simple Interaction Effect , Simple Main Effect และ Simple Simple Effect

ผลการศึกษา

ผลการศึกษาที่สำคัญสรุปได้ดังนี้

1. ผลการพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี

การพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี มีขั้นตอนดังนี้

1.1 ศึกษาเอกสาร แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับวิธีการจัดการข้อมูลสูญหาย

1.2 วิเคราะห์แนวคิดของวิธีการจัดการข้อมูลสูญหายแบบต่าง ๆ โดยพิจารณาจากจุดเด่น จุดด้อยของวิธีการเหล่านั้นเพื่อหาแนวคิดที่จะนำเสนอวิธีการจัดการข้อมูลสูญหายแบบใหม่

1.3 เสนอวิธีการจัดการข้อมูลสูญหายแบบใหม่

จากการศึกษาวิธีการจัดการข้อมูลสูญหายประกอบกับการศึกษาจุดเด่นและจุดด้อยของวิธีการต่าง ๆ พบว่าเทคนิควิธีการที่ค่อนข้างมีเหตุผลในการประมาณค่าพารามิเตอร์ได้ถูกต้องแม่นยำก็คือการทำซ้ำ (Iteration) โดยเฉพาะวิธี EM algorithm จะได้รับการตรวจสอบและยืนยันจากรายงานการวิจัยมากที่สุดว่าเป็นวิธีการที่ดี แต่วิธีการดังกล่าวก็มีจุดบกพร่องอยู่หลายประการ เช่น

1. การแทนค่าสูญหายครั้งแรกในสมการ $\hat{Y} = a + bX$ ซึ่ง $a = \overline{Y} - b\overline{X}$ และ $b = r_{xy}\frac{S_y}{S_x}$ ประกอบไปด้วยค่าสถิติ คือ $\overline{X}, \overline{Y}, r_{xy}, S_x$ และ S_y ได้มาจากกลุ่มตัวอย่างที่มีข้อมูลสมบูรณ์ได้ตัดหน่วยตัวอย่างที่มีข้อมูลสูญหายออกไป ดังนั้นการประมาณค่าพารามิเตอร์ด้วยค่าเหล่านี้จึงทำให้มีอคติ (bias) หมายความว่าค่าที่ได้อาจจะไม่ใช่ค่าของประชากรที่แท้จริง

2. วิธีการจัดการข้อมูลสูญหายแบบ EM algorithm ใช้วิธีการประมาณค่าพารามิเตอร์ด้วยวิธี Maximum likelihood มีข้อจำกัดในการประมาณค่าพารามิเตอร์คือใช้เทคนิคการทำซ้ำ (Iteration algorithms) ซึ่งใช้เวลาและเสียค่าใช้จ่ายค่อนข้างสูง และไม่สามารถนำไปรวมไว้กับโปรแกรมคอมพิวเตอร์มาตรฐานทั่ว ๆ ไป และประเด็นที่ถูกวิพากษ์วิจารณ์ก็คือความคลาดเคลื่อนมาตรฐานไม่ตรง (invalid) คือไม่ได้คำนวณอย่างตรงไปตรงมาในการวิเคราะห์ขั้นสุดท้ายดังนั้นจะมีวิธีการอื่น ๆ ที่มีความตรงในการประมาณค่าความคลาดเคลื่อนมาตรฐาน แต่วิธีนี้นักสถิติแนะนำในการใช้แก้ปัญหาข้อมูลสูญหายเพราะใช้วิธีการถดถอยซึ่งน่าจะแทนค่าข้อมูลสูญหายได้ถูกต้องมากยิ่งขึ้น

ดังนั้นผู้วิจัยจึงใช้วิธีการประมาณค่าข้อมูลสูญหายโดยมีขั้นตอนดังต่อไปนี้

1. ประมาณค่าข้อมูลสูญหายโดยวิธีการถดถอยอย่างง่าย (Simple regression) เป็นเทคนิคพื้นฐานของวิธีการอื่น ๆ จากสมการ $\hat{Y} = a + bX$ ใช้ค่าสถิติคือ $a = \overline{Y} - b\overline{X}$ และ $b = r_{xy}\frac{S_y}{S_x}$ เป็นค่าโดยประมาณแต่ถ้าเราสามารถประมาณค่าพารามิเตอร์ได้ก็จะสามารถทำให้สมการ $\hat{Y} = a + bX$ เป็นสมการที่แท้จริงของประชากร การประมาณค่าข้อมูลสูญหายก็จะได้ค่าที่ถูกต้อง ดังนั้นค่าสถิติ $\overline{X}, \overline{Y}, r_{xy}, S_x$ และ S_y จะได้มาโดยการประมาณค่าแบบทำซ้ำตามทฤษฎีกล่าวว่า ค่าคาดหวังของค่าเฉลี่ยจะเท่ากับค่าเฉลี่ยของประชากร $\left(E(\overline{X}) = \mu\right)$ หมายความว่า ถ้ามีค่าเฉลี่ยจำนวนมากแล้วนำมาหาค่าเฉลี่ยจะได้ค่าเฉลี่ยของประชากร ดังนั้นในขั้นตอนแรกนี้จะทำการสุ่มตัวอย่างซ้ำจากข้อมูลที่เหลือเมื่อตัดข้อมูลสูญหายออกไปแล้วเป็นจำนวน 1,000 ครั้ง แล้วนำค่าเฉลี่ยทั้ง 1,000 ครั้งมาหาค่าเฉลี่ยก็จะได้ค่าเฉลี่ยของประชากร

การประมาณค่า S_x และ S_y ก็ทำในลักษณะเดียวกันเพราะ $E(S^2) = \sigma^2$ การประมาณค่าความแปรปรวนดังกล่าวสอดคล้องกับการประมาณค่าความแปรปรวนที่เสนอโดยลิทเทิลและรูบิน ที่กล่าวว่าถ้าให้ $\hat{\theta}_1, ..., \hat{\theta}_k$ เป็นตัวแปรสุ่มซึ่งไม่มีความสัมพันธ์ต่อกันและมีค่าเฉลี่ยคือ μ จะได้ว่า $\overline{\theta} = \sum_{j=1}^{k}\frac{\hat{\theta}_j}{k}$ ดังนั้นในการประมาณค่าความแปรปรวนกรณีที่มีข้อมูลสูญหายก็สามารถทำได้โดยการสุ่มข้อมูลที่สมบูรณ์จำนวน 1,000 ครั้ง ในแต่ละครั้งหาความแปรปรวนแล้วนำความแปรปรวนทั้งหมดนั้นมาหาค่าเฉลี่ยก็จะได้ตัวประมาณค่าความแปรปรวนของประชากร

สำหรับการประมาณค่าความสัมพันธ์ของประชากรก็จะทำในลักษณะเดียวกันด้วยเหตุผลต่อไปนี้ สมการแสดงความแปรปรวนร่วมของตัวแปรต้นและตัวแปรตามของกลุ่มตัวอย่าง

$$\begin{aligned}COV(X,Y) &= E\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)\\ &= E\left(XY-\overline{X}Y-\overline{Y}X+\overline{X}\,\overline{Y}\right)\\ &= E(XY)-\overline{X}E(Y)-\overline{Y}E(X)+\overline{X}\,\overline{Y}\\ &= E(XY)-\overline{X}\,\overline{Y}-\overline{Y}\overline{X}+\overline{X}\,\overline{Y}\\ &= E(XY)-\overline{X}\,\overline{Y}\end{aligned}$$

ดังนั้นสามารถหาค่าคาดหวังความแปรปรวนร่วมของกลุ่มตัวอย่างได้ดังนี้

$$E(COV(X,Y)) = E(E(XY)-\overline{X}\,\overline{Y})$$

$$E(COV(X,Y)) = E(E(XY)) - E(\overline{X}\,\overline{Y})$$

ถ้า a และ b เป็นค่าคงที่ จะได้ E(aX+b)=aE(X)+b ถ้า a=0 แล้ว จะได้ E(b)=b เนื่องจาก E(XY) เป็นค่าคงที่ตัวหนึ่ง ดังนั้น E(E(XY)) = E(XY) และเนื่องจาก $\overline{X}$ และ $\overline{Y}$ เป็นค่าคงที่จึงน่าจะเป็นอิสระต่อกัน และถ้า X และ Y เป็นอิสระต่อกันแล้ว E(XY)=E(X)E(Y) ดังนั้น $E(\overline{X}\,\overline{Y}) = E(\overline{X})E(\overline{Y})$ แต่ $E(\overline{X})=\mu_X$ และ $E(\overline{Y})=\mu_X$ จะทำให้ได้ว่า

$$E(COV(X,Y)) = E(XY)-\mu_X\mu_y = E(X-\mu_X)(Y-\mu_y)$$

ซึ่งสมการดังกล่าวก็คือความแปรปรวนร่วมของประชากร (σ_{xy}) นั่นเอง

แสดงว่าค่าคาดหวังความแปรปรวนร่วมของกลุ่มตัวอย่างก็คือความแปรปรวนร่วมของประชากร ดังสมการ

$$\sigma_{xy} = E\left(X-\mu_x\right)\left(Y-\mu_y\right)$$

$$\sigma_{xy} = E\left(COV(X,Y)\right)$$

แต่เนื่องจาก $\rho_{XY}=\sigma_{XY}/\sigma_X\sigma_Y$

ดังนั้นเมื่อสามารถประมาณค่าความแปรปรวนร่วมของประชากรได้ก็สามารถประมาณค่าความสัมพันธ์ของประชากรได้เช่นเดียวกัน

ในการประมาณค่าความแปรปรวนร่วมของประชากรกรณีที่มีข้อมูลสูญหาย ผู้วิจัยจะสุ่มข้อมูลที่สมบูรณ์มาจำนวน 1,000 ครั้ง แต่ละครั้งหาความแปรปรวนร่วม แล้วนำความแปรปรวนร่วมทั้งหมดมาหาค่าเฉลี่ยก็จะได้ความแปรปรวนร่วมของประชากร แล้วนำมาหาความสัมพันธ์ของประชากรโดยใช้สูตร $\rho_{XY}=\sigma_{XY}/\sigma_X\sigma_Y$ ต่อไป

2. จากการประมาณค่าข้อมูลสูญหายโดยการสร้างสมการทำนาย $\hat{Y}=a+bX$ และใช้การประมาณค่าพารามิเตอร์ในขั้นตอนที่ 1 ก็ทำให้มั่นใจได้ว่า การทำนายข้อมูลสูญหายน่าจะถูกต้องแม่นยำมากยิ่งขึ้น ลักษณะการสร้างสมการทำนายจะต้องทำให้มีความคลาดเคลื่อนน้อยที่สุด ทำให้การทำนายค่า Y (ข้อมูลสูญหาย) ได้อย่างไม่มีอคติค่าที่ได้ใกล้เคียงกับค่าของ

ประชากร ซึ่งความแตกต่างระหว่าง $Y - Y_p$ (Y_p เป็นค่าที่ได้จากการทำนาย) เป็นความคลาดเคลื่อน ($e = Y - Y_p$) ค่าความคลาดเคลื่อนมีทั้งค่าที่เป็นบวกและลบ เส้นถดถอยที่ไม่มีอคติคือเส้นที่ให้ค่าผลรวมของความคลาดเคลื่อนที่เป็นบวกและลบเท่าๆ กัน แต่เนื่องจาก $\sum(Y - Y_p) = 0$ เพื่อตัดค่าเครื่องหมายลบออกไปจึงยกกำลังสองค่าความคลาดเคลื่อนนั่นคือ $\sum(Y - Y_p)^2 = \sum e^2$ มีค่าน้อยที่สุดจึงจะได้สมการทำนายที่ดีที่สุด แต่ลักษณะการประมาณค่าในขั้นตอนที่ 1 สร้างจากข้อมูลที่สมบูรณ์และข้อมูลสูญหายที่ได้จากการทำนายจึงทำให้ไม่สามารถมั่นใจได้ว่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard error of the estimate) จะมีค่าน้อยที่สุด ดังนั้นจึงจำเป็นต้องเลือกสมการที่มีความคลาดเคลื่อนน้อยที่สุด (ถ้าสร้างสมการถดถอยจากข้อมูลสมบูรณ์จะมีความคลาดเคลื่อนน้อยที่สุดตามวิธีกำลังสองน้อยที่สุด) เพื่อให้สมการทำนายข้อมูลสูญหายได้อย่างถูกต้องแม่นยำ

ดังนั้นเมื่อแทนค่าข้อมูลสูญหายในขั้นตอนที่ 1 แล้ว ผู้วิจัยจึงสร้างสมการการทำนายข้อมูลสูญหายจากข้อมูลทั้งหมดทั้งข้อมูลสมบูรณ์และข้อมูลสูญหายที่ได้จากการทำนายโดยใช้วิธีการถดถอยอย่างง่าย (Simple regression) จำนวน 1,000 ครั้ง แล้วเลือกสมการที่มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าน้อยที่สุดเป็นสมการทำนายข้อมูลสูญหายในการศึกษาวิจัยครั้งนี้

จากที่กล่าวมาจะเห็นได้ว่าในขั้นตอนที่ 1 เป็นการประมาณค่าพารามิเตอร์ (Estimated Parameter) และในขั้นตอนที่ 2 เป็นการคัดเลือกสมการที่มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าน้อยที่สุด (Smallest Standard Error) ดังนั้นผู้วิจัยจึงตั้งชื่อวิธีการจัดการข้อมูลสูญหายแบบนี้ว่า วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี (EPSSE)

2. ผลการเปรียบเทียบความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์

การเปรียบเทียบความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ จำแนกตามวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร สรุปได้ดังนี้

2.1 ค่าความแม่นยำของค่าเฉลี่ยเลขคณิต เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหายค่อนข้างน้อยคือเท่ากับ 5%-10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ แบบอีเอ็ม และแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด แต่เมื่อใช้วิธีการสุ่มตัวอย่างแบบเดิม จำนวนข้อมูลสูญหายสูงขึ้นเป็น 20%-30%

ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) เช่นเดียวกัน วิธีการจัดการข้อมูลสูญหายแบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

เมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายเท่ากับ 5%-20% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ แบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่เมื่อใช้วิธีการสุ่มตัวอย่างแบบเดิม จำนวนข้อมูลสูญหายอยู่ในระดับสูงสุดเท่ากับ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

เมื่อใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จำนวนข้อมูลสูญหายค่อนข้างน้อย คือเท่ากับ 5%-10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ แบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด แต่เมื่อใช้วิธีการสุ่มตัวอย่างแบบเดิม จำนวนข้อมูลสูญหายสูงขึ้นเป็น 20%-30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

2.2 ค่าความแม่นยำของความแปรปรวน เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหาย 5% 10% 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของความแปรปรวนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

เมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายอยู่ในระดับน้อยที่สุดคือเท่ากับ 5% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ แบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของความแปรปรวนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีเพียงกรณีเดียว คือ เมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายเท่ากับ 5% ความสัมพันธ์ระหว่าง

ตัวแปรอยู่ในระดับปานกลาง (r=.50) ได้ค่าความแม่นยำของความแปรปรวนแตกต่างจากกรณีอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่ต่ำที่สุด แต่เมื่อใช้วิธีการสุ่มตัวอย่างแบบเดิม จำนวนข้อมูลสูญหายเท่ากับ 10% 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของความแปรปรวน ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

เมื่อใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จำนวนข้อมูลสูญหาย 5% 10% 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของความแปรปรวนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

2.3 ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหายเท่ากับ 5% 10% และ 20% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ แบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด แต่เมื่อใช้วิธีการสุ่มตัวอย่างแบบเดิม จำนวนข้อมูลสูญหายสูงที่สุดคือเท่ากับ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์แตกต่างจากกรณีอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

เมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายค่อนข้างน้อยคือเท่ากับ 5% -10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์แบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด แต่เมื่อใช้วิธีการสุ่มตัวอย่างแบบเดิม จำนวนข้อมูลสูญหายสูงขึ้นเป็น 20%-30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์แตกต่างจากกรณีอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

เมื่อใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จำนวนข้อมูลสูญหายเท่ากับ 5% และ 20% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ แบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด ในกรณีที่จำนวนข้อมูล

สูญหายเท่ากับ 10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบอีเอ็มและแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด แต่เมื่อจำนวนข้อมูลสูญหายสูงที่สุดคือเท่ากับ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์แตกต่างจากกรณีอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด

3. ผลการเปรียบเทียบอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์ การทดสอบทีและการทดสอบเอฟ

การเปรียบเทียบอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์ การทดสอบทีและการทดสอบเอฟ จำแนกตามวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร สรุปได้ดังนี้

3.1 อำนาจการทดสอบที่ได้จากการทดสอบความสัมพันธ์ เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหายมีค่าเพิ่มขึ้น (5% , 10% , 20% และ 30%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ อำนาจการทดสอบมีค่าลดลง และเมื่อใช้วิธีการสุ่มแบบเดิม จำนวนข้อมูลสูญหายมีค่าเพิ่มขึ้น (5% , 10% , 20% และ 30%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) เช่นเดียวกัน จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็ม อำนาจการทดสอบก็มีค่าลดลงเช่นเดียวกัน เมื่อพิจารณาเปรียบเทียบกับวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์ สูงกว่าวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ และการแทนค่าแบบอีเอ็ม แต่เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหาย 5% , 10% , 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับปานกลาง (r=.50) และสูง (r=.70) วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ การแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี ได้ค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์สูงเป็น 1 ทุกกรณี

เมื่อพิจารณาอำนาจการทดสอบที่ได้จากการทดสอบความสัมพันธ์ เมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหาย 5% , 10% , 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ และการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี ค่าอำนาจการทดสอบมีค่าสูงเป็น 1 เกือบทุกเงื่อนไขการทดลอง ยกเว้นในกรณีที่ใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม

จำนวนข้อมูลสูญหายสูงสุด (30%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ ได้ค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์ต่ำสุด

เมื่อพิจารณาอำนาจการทดสอบที่ได้จากการทดสอบความสัมพันธ์ เมื่อใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน มีแนวโน้มเหมือนกันกับวิธีการสุ่มตัวอย่างแบบกลุ่ม คือ มีค่าสูงเป็น 1 เกือบทุกเงื่อนไขการทดลอง ยกเว้นในกรณีที่ใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จำนวนข้อมูลสูญหายสูงสุด (30%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์และการแทนค่าข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์เท่ากับ .9990 แต่ก็เป็นค่าที่ใกล้เคียงกันมากกับวิธีการจัดการข้อมูลสูญหายแบบอีเอ็ม

3.2 อำนาจการทดสอบ (1σ และ 2σ) จากการทดสอบทีและการทดสอบเอฟ โดยใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จำนวนข้อมูลสูญหายเท่ากับ 5% , 10% , 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ การแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี ได้ค่าอำนาจการทดสอบเท่ากับ 1 ทุกกรณี

โดยสรุปการเปรียบเทียบอำนาจการทดสอบถ้าใช้การทดสอบความสัมพันธ์ การทดสอบทีและการทดสอบเอฟ ได้ผลสอดคล้องกันทุกกรณี ไม่ว่าจะใช้วิธีการสุ่มตัวอย่างแบบใด จำนวนข้อมูลสูญหายระดับใด ความสัมพันธ์ระหว่างตัวแปรระดับใด และวิธีการจัดการข้อมูลสูญหายแบบใด

4. ผลการศึกษาปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร

การศึกษาปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย **จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร ที่มีต่อความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ สรุปได้ดังนี้**

4.1 ปฏิสัมพันธ์สี่ทางและสามทางระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปรที่มีต่อความแม่นยำของค่าเฉลี่ยเลขคณิต ไม่มีนัยสำคัญทางสถิติที่ระดับ .01

4.2 ปฏิสัมพันธ์สี่ทางระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร ที่มีต่อความแม่นยำของความแปรปรวน ไม่มีนัยสำคัญทางสถิติที่ระดับ .01 แต่พบว่า ปฏิสัมพันธ์สามทางมีนัยสำคัญทางสถิติที่ระดับ .01 ดังต่อไปนี้

4.2.1 ปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายและจำนวนข้อมูลสูญหาย มีนัยสำคัญทางสถิติที่ระดับ .01 ผลการวิเคราะห์ Simple Interaction Effect , Simple Main Effect และ Simple Simple Effect พบว่า ค่าความแม่นยำของความแปรปรวนจะสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายทุกระดับ จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ และเมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายน้อยที่สุด (5%) จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี ค่าความแม่นยำของความแปรปรวนจะต่ำลงเรื่อย ๆ เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่ม และแบบหลายขั้นตอน จำนวนข้อมูลสูญหายค่อนข้างสูง (20%-30%) จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี

4.2.2 ปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปรมีนัยสำคัญทางสถิติที่ระดับ .01 ผลการวิเคราะห์ Simple Interaction Effect, Simple Main Effect และ Simple Simple Effect พบว่า ค่าความแม่นยำของความแปรปรวนจะสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม ความสัมพันธ์ระหว่างตัวแปรทุกระดับ จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ และเมื่อใช้วิธีการสุ่มแบบหลายขั้นตอน ความสัมพันธ์ระหว่างตัวแปรทุกระดับ จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ แต่ถ้าใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับปานกลาง (r=.50) และสูง (r=.70) จัดการข้อมูลสูญหายโดยการ ตัดออกแบบลิสท์ไวส์ ค่าความแม่นยำของความแปรปรวนจึงจะสูงที่สุด ค่าความแม่นยำของความแปรปรวนจะต่ำลงเรื่อย ๆ เมื่อระดับความสัมพันธ์ระหว่างตัวแปรต่ำลง โดยเฉพาะเมื่อใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี ความสัมพันธ์ระหว่างตัวแปรต่ำสุด (r=.30) ได้ค่าความแม่นยำของความแปรปรวนต่ำที่สุด

4.2.3 ปฏิสัมพันธ์ระหว่างวิธีการจัดการข้อมูลสูญหายกับจำนวนข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปร มีนัยสำคัญทางสถิติที่ระดับ .01 ผลการวิเคราะห์ Simple Interaction Effect , Simple Main Effect และ Simple Simple Effect พบว่า ค่าความแม่นยำของความแปรปรวนจะสูงที่สุดเมื่อใช้วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์

จำนวนข้อมูลสูญหายทุกระดับและความสัมพันธ์ระหว่างตัวแปรทุกระดับ แต่ถ้าใช้วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี จำนวนข้อมูลสูญหายจะต้องอยู่ในระดับน้อยที่สุด คือ เท่ากับ 5% เท่านั้น และความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับปานกลาง (r=.50) และสูง (r=.70) ค่าความแม่นยำของความแปรปรวนจะต่ำลงเรื่อย ๆ เมื่อจำนวนข้อมูลสูญหายเพิ่มขึ้นและความสัมพันธ์ระหว่างตัวแปรลดลง โดยเฉพาะเมื่อใช้วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี ทุกระดับความสัมพันธ์ระหว่างตัวแปร จำนวนข้อมูลสูญหายอยู่ในระดับสูง คือ เท่ากับ 20% และ 30% ได้ค่าความแม่นยำของความแปรปรวนต่ำที่สุด

4.3 ปฏิสัมพันธ์สี่ทางระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวน ข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปร ที่มีต่อความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ ไม่มีนัยสำคัญทางสถิติที่ระดับ .01 แต่พบว่าปฏิสัมพันธ์สามทางมีนัยสำคัญทางสถิติที่ระดับ .01 ระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายและจำนวนข้อมูลสูญหาย ผลการวิเคราะห์ Simple Interaction Effect , Simple Main Effect และ Simple Simple Effect พบว่า ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์จะสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นแบบกลุ่มและแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ จำนวนข้อมูลสูญหายอยู่ระหว่าง 5%-20% และเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี จะต้องใช้ข้อมูลที่มีจำนวนข้อมูลสูญหายค่อนข้างน้อยคืออยู่ระหว่าง 5%-10% เท่านั้น จึงจะได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์สูงที่สุด ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์จะต่ำลงเรื่อย ๆ เมื่อจำนวนข้อมูลสูญหายเพิ่มมากขึ้น โดยเฉพาะเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นแบบกลุ่มและแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี จำนวนข้อมูลสูญหายสูงสุดคือเท่ากับ 30% ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ต่ำที่สุด

อภิปรายผลการวิจัย

การวิจัยครั้งนี้ได้ข้อค้นพบเกี่ยวกับวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีที่ผู้วิจัยพัฒนาขึ้น ตรวจสอบความแม่นยำและอำนาจการทดสอบที่ได้จากวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีกับวิธีอื่น ๆ และศึกษาปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการ

ข้อมูลสูญหายกับความสัมพันธ์ระหว่างตัวแปรและจำนวนข้อมูลสูญหาย ซึ่งข้อค้นพบมีทั้งเป็นไปตามสมมุติฐานและไม่เป็นไปตามสมมุติฐาน ดังที่ผู้วิจัยจะได้นำมาอภิปรายดังต่อไปนี้

1. จากสมมุติฐานการวิจัยที่กล่าวว่า วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีน่าจะมีความแม่นยำดีกว่าวิธีอื่น เมื่อใช้วิธีการสุ่มตัวอย่าง ความสัมพันธ์ระหว่างตัวแปร และจำนวนข้อมูลสูญหายที่แตกต่างกัน

1.1 ผลการวิจัย พบว่า วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอีได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิต ไม่แตกต่างจากวิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์และแบบอีเอ็ม อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จำนวนข้อมูลสูญหายอยู่ระหว่าง 5%-10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) และวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอีได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิต ไม่แตกต่างจากวิธีอีเอ็ม อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นแบบกลุ่มและแบบหลายขั้นตอน จำนวนข้อมูลสูญหายอยู่ในระดับสูงที่สุดคือเท่ากับ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) ซึ่งไม่สอดคล้องกับสมมุติฐานการวิจัยที่ตั้งไว้ ทั้งนี้สามารถอธิบายได้ว่า ในการประมาณค่าโดยใช้วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีมีขั้นตอนในการประมาณค่าข้อมูล 2 ขั้นตอน 1). ประมาณค่าข้อมูลสูญหายโดยวิธีการถดถอยอย่างง่าย (Simple regression) จากสมการ $\hat{Y} = a + bX$ ใช้ค่าสถิติคือ $a = \overline{Y} - b\overline{X}$ และ $b = r_{xy}\frac{S_y}{S_x}$ เป็นค่าโดยประมาณ แต่วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี ประมาณค่า $\overline{X}, \overline{Y}, r_{xy}, S_x$ และ S_y โดยการทำซ้ำ จำนวน 1,000 ครั้ง จึงทำให้มั่นใจได้ว่า สมการ $\hat{Y} = a + bX$ เป็นสมการที่แท้จริงของประชากร วิธีการแทนค่าข้อมูลสูญหายโดยใช้การถดถอย (Regression imputation) จะรักษาข้อมูลไว้มากกว่าวิธีการตัดข้อมูลสูญหายออกแบบลิสท์ไวส์ เนื่องจากข้อมูลของแต่ละคนไม่ได้ถูกลบออกไปจากการวิเคราะห์ซึ่งเป็นผลมาจากการสูญหายข้อมูลที่ได้จากการประมาณค่ารักษาการแจกแจงของค่าเฉลี่ย (Roth, 1994. p. 541) 2). สร้างสมการทำนายข้อมูลสูญหายจากข้อมูลทั้งหมด ทั้งข้อมูลสมบูรณ์และข้อมูลสูญหายที่ได้จากการทำนายโดยใช้วิธีการถดถอยอย่างง่าย (Simple regression) จำนวน 1,000 ครั้ง แล้วเลือกสมการที่มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าน้อยที่สุด ประกอบกับค่าสถิติค่าเฉลี่ยได้มาจากสูตร $\bar{X} = \frac{\sum X}{N}$ เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางตัวหนึ่งที่ใช้ข้อมูลทั้งหมดมาคำนวณ ถึงแม้ว่าการแทนค่าข้อมูลสูญหายด้วยวิธีอีพีเอสเอสอีจะทำให้ข้อมูลที่นำไปแทนข้อมูลสูญหาย

แตกต่างจากค่าของข้อมูลจริงไปบ้าง แต่ด้วยเหตุผลที่กล่าวมาข้างต้นทำให้ค่าเฉลี่ยที่คำนวณได้ไม่แตกต่างจากค่าของประชากรมากนัก จึงเป็นเหตุผลที่น่าจะทำให้ค่าความแม่นยำที่ได้จากวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีสูงที่สุด และข้อค้นพบจากการวิจัยพบว่าวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตไม่แตกต่างจากวิธีการจัดการข้อมูลสูญหายแบบอีเอ็ม ทั้งนี้สามารถอธิบายได้ว่า วิธีการจัดการข้อมูลสูญหายแบบอีเอ็มแทนค่าข้อมูลสูญหายโดยใช้วิธีการถดถอยแล้วประมาณค่าเมตริกซ์ความแปรปรวนร่วม ดำเนินการทำซ้ำ (iterative method) จนค่าของเมตริกซ์ความแปรปรวนร่วมไม่แตกต่างกัน ซึ่งเป็นการประมาณค่าพารามิเตอร์แบบหนึ่งที่เรียกว่า แมกซิมัมไลค์ลิฮูด โดยทำการประมาณค่าพารามิเตอร์ที่มีโอกาสเกิดขึ้นสูงที่สุด (Enders, 2001. p. 715) วิชุดา บัวคง (2533) ได้ศึกษาเปรียบเทียบประสิทธิผลของวิธีประมาณค่าพารามิเตอร์ของแบบจำลองโลจีสติก 3 พารามิเตอร์ระหว่างวิธีแมกซิมัมไลค์ลิฮูด วิธีฮิวลิสติก และวิธีของเบย์ ในแบบทดสอบวัดผลสัมฤทธิ์และแบบทดสอบวัดความถนัด ผลการวิจัย พบว่า วิธีแมกซิมัมไลค์ลิฮูดมีประสิทธิภาพสูงที่สุดในการประมาณค่าพารามิเตอร์ ประกอบการประมาณค่าของวิธีอีเอ็มทำซ้ำจนเมตริกซ์ความแปรปรวนร่วมไม่แตกต่างกันซึ่งแสดงถึงค่าสถิติต่าง ๆ จะคงที่ไม่เปลี่ยนแปลง ค่าความแปรปรวนร่วมจะประกอบไปด้วยค่าสถิติ r_{xy}, S_x, S_y ซึ่งเมื่อค่าสถิติเหล่านี้มีค่าคงที่ จึงทำให้ได้ผลการประมาณค่าใกล้เคียงกับวิธีอีพีเอสเอสอีที่ผู้วิจัยพัฒนาขึ้น

เมื่อพิจารณาเงื่อนไขอื่น ๆ ที่ทำให้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตที่ได้จากวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีสูงที่สุด คือ จำนวนข้อมูลสูญหายค่อนข้างสูง (30%) และความสัมพันธ์สูง (r=.70) ทั้งนี้อาจเป็นเพราะว่า เมื่อข้อมูลสูญหายน้อย เช่น 5% ข้อมูลที่เหลืออยู่ยังมีความเป็นตัวแทนของข้อมูลทั้งหมด ดังนั้นการตัดข้อมูลออกไปหรือการแทนค่าข้อมูลสูญหายก็ไม่ทำให้ความแม่นยำของค่าเฉลี่ยเลขคณิตแตกต่างกัน แต่เมื่อจำนวนข้อมูลสูญหายเพิ่มมากขึ้นความเป็นตัวแทนของข้อมูลทั้งหมดก็จะน้อยลงจึงทำให้วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตแตกต่างจากวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีและแบบอีเอ็ม สอดคล้องกับคำกล่าวของรูท (Roth, 1994. p. 551) ที่กล่าวว่าเราจะไม่เลือกใช้วิธีการจัดการข้อมูลสูญหายเมื่อจำนวนข้อมูลสูญหายน้อยกว่า 10% แม้ว่าจะสูญหายแบบสุ่ม (missing at random) หรือสูญหายแบบเป็นระบบ (systematic pattern) การเลือกใช้ข้อมูลสูญหายจะมีความสำคัญมากขึ้นเมื่อจำนวนข้อมูลสูญหายอยู่ระหว่าง 15%-20% และจะมีความสำคัญมากที่สุดเมื่อมีจำนวนข้อมูลสูญหาย 30%-40% ที่จำนวนข้อมูลสูญหายระดับนี้วิธีการจัดการข้อมูลสูญหายจะให้ผลแตกต่างกัน

และคำกล่าวของ เชฟเฟอร์ และลิทเทิลและรูบิน (Schafer, 1997. p. 1 ; Little and Rubin, 1987. p. 5) ที่กล่าวว่า ถ้ามีข้อมูลสูญหายจำนวนน้อยประมาณ 5% การตัดหน่วยตัวอย่างออกไปดูเหมือนว่าจะมีเหตุผลในการแก้ปัญหาข้อมูลสูญหาย แต่ถ้ามีการสูญหายมากการตัดข้อมูลออกจะไม่มีประสิทธิภาพข้อมูลที่เหลืออยู่จะไม่เป็นตัวแทนของประชากรซึ่งมีเป้าหมายในการอ้างอิง

ตัวแปรอีกตัวหนึ่งคือความสัมพันธ์ระหว่างตัวแปร จากผลการวิจัย พบว่า วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีและแบบอีเอ็มจะมีความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุดเมื่อความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) ทั้งนี้อาจเป็นเพราะว่า วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีใช้วิธีการถดถอยทำนายข้อมูลสูญหาย การทำนายข้อมูลสูญหายโดยใช้วิธีการถดถอยจะมีความถูกต้องมากยิ่งขึ้นเมื่อระดับความสัมพันธ์ระหว่างตัวแปรสูง ยิ่งความสัมพันธ์สูงมากขึ้นเท่าไร การทำนายก็จะมีความถูกต้องมากยิ่งขึ้น ซึ่งสอดคล้องกับงานวิจัยของรูท (Roth, 1994. pp. 537-560) ที่พบว่า ถ้าความสัมพันธ์ระหว่างตัวแปรมีค่ามากวิธีการจัดการข้อมูลสูญหายโดยใช้การถดถอยจะดีกว่าเมื่อข้อมูลมีความสัมพันธ์ระหว่างตัวแปรน้อย แต่ทั้งนี้ก็ขึ้นอยู่กับตัวแปรอื่น ๆ ด้วย

1.2 ผลการวิจัย พบว่า วิธีการจัดการข้อมูลสูญหายโดยการตัดข้อมูลสูญหายออกแบบลิสท์ไวส์ได้ค่าความแม่นยำของความแปรปรวนแตกต่างจากวิธีอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด ตามเงื่อนไขต่อไปนี้ 1). เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหาย 5% 10% 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) 2). เมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายเท่ากับ 10% 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) 3). เมื่อใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จำนวนข้อมูลสูญหาย 5% 10% 20% และ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ปานกลาง (r=.50) และสูง (r=.70) ซึ่งไม่สอดคล้องกับสมมุติฐานการวิจัยที่ตั้งไว้ ทั้งนี้สามารถอธิบายได้ว่า ในการวิจัยครั้งนี้ผู้วิจัยสร้างข้อมูลสูญหายแบบสุ่ม (randomly missing data) เมื่อตัดข้อมูลสูญหายออกจึงยังทำให้ข้อมูลที่เหลืออยู่เป็นตัวแทนของข้อมูลทั้งหมด จากข้อมูลที่สร้างขึ้นมามีคะแนนสอบของนักเรียน (X) และเกรดเฉลี่ยของนักเรียน (Y) แล้วสร้างข้อมูลสูญหายแบบสุ่มในตัวแปรเกรดเฉลี่ย ผู้วิจัยได้นำข้อมูลมาสร้างตัวแปรใหม่อีก 1 ตัวแปร คือ ตัวแปรสูญหาย (MISS) ค่าของตัวแปรสูญหายมีค่าเท่ากับ 1 เมื่อเกรดเฉลี่ยของนักเรียนมีค่าสูญหาย และมีค่าเท่ากับ 0 เมื่อตัวแปรเกรดเฉลี่ยของนักเรียนไม่มีค่าสูญหาย แล้วนำค่าของ

ตัวแปรใหม่นี้ไปหาความสัมพันธ์กับคะแนนสอบของนักเรียน ผลปรากฏว่า ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าสาเหตุของการสูญหายไม่มีความเกี่ยวข้องกับคะแนนสอบ ลักษณะของการสูญหายแบบนี้เรียกว่าเป็น MCAR (missing completely at random) ถ้าข้อมูลมีลักษณะเป็น MCAR ก็จะประมาณค่าได้อย่างดี (Little & Rubin, 1987; Rubin, 1976) และถ้าใช้วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ก็จะไม่มีอคติ (Cerrilo et al., 2000. http://www.lrdc.pitt.edu/hple/Publications/PERSIST.PDF) ประกอบกับความแปรปรวนได้มาจากสูตร $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$ เป็นการพิจารณาข้อมูลแต่ละตัวเทียบกับค่าเฉลี่ย จากที่กล่าวมาเมื่อข้อมูลมีการสูญหายแบบสุ่ม ข้อมูลที่เหลืออยู่จะมีความเป็นตัวแทนของข้อมูลทั้งหมดค่อนข้างสูง เมื่อนำมาหาความแปรปรวนจึงไม่ทำให้โครงสร้างของข้อมูลเปลี่ยนแปลงไป ค่าความแปรปรวนที่ได้จึงน่าจะใกล้เคียงกับค่าของประชากร ทำให้ความแม่นยำมีค่าสูง แต่เมื่อพิจารณาเปรียบเทียบกับวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี ข้อมูลสูญหายจะถูกแทนค่าด้วยข้อมูลที่สร้างขึ้นมาโดยใช้การแทนค่าด้วยวิธีการถดถอยจากสมการ $\hat{Y} = a + bX$ ซึ่งไม่ได้นำค่าความคลาดเคลื่อนรวมเข้าไปในสมการดังกล่าว ถึงแม้ว่าวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอีจะใช้สมการถดถอยที่มีความคลาดเคลื่อนน้อยที่สุดแต่ก็ยังมีความคลาดเคลื่อนอยู่ค่าที่แทนข้อมูลสูญหายจึงอาจจะแตกต่างจากค่าจริง ทั้งนี้เมื่อนำมาหาค่าเฉลี่ยอาจจะมีความแตกต่างจากค่าของประชากรน้อยเพราะเป็นภาพรวมของข้อมูลทั้งหมด ดังที่ได้กล่าวมาแล้ว แต่เมื่อนำมาหาค่าความแปรปรวนอาจจะทำให้โครงสร้างของข้อมูลเปลี่ยนไป ด้วยเหตุผลที่กล่าวมา จึงทำให้ค่าความแม่นยำของความแปรปรวนที่ได้จากวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์สูงที่สุด ซึ่งผลการวิจัยดังกล่าวสอดคล้องกับงานวิจัยของเฟรดและลี (Fred & Lii, 1998. http://www.findarticles.com) ที่พบว่าวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ มีความถูกต้องเพราะว่า เมื่อลบข้อมูลออกไปข้อมูลที่เหลืออยู่ยังเป็นโครงสร้างของข้อมูลจริง และรูท (Roth, 1999) ได้วิพากษ์วิจารณ์ว่าวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าด้วยวิธีการถดถอย (regression imputation) ดีกว่าวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ แต่จากงานวิจัยของเขา พบว่าวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์มีอคติน้อยหรือไม่มีเลยและการกระจายรอบ ๆ คะแนนจริงน้อย นอกจากนี้ ซู (Zhou, 2002. p. 11) ได้ศึกษาวิจัยแล้วพบว่า วิธีการประมาณค่าข้อมูลสูญหายโดยใช้วิธีการถดถอยจะประมาณค่าความแปรปรวนของมาตรวัด (scale variance) ได้น้อยกว่าปกติ อาจเป็นไปได้ว่าวิธีนี้ไม่ได้รวมความคลาดเคลื่อนไว้ด้วย และ

วิธีการจัดการข้อมูลสูญหายโดยไม่รวมความคลาดเคลื่อนมีความถูกต้องน้อยในการประมาณค่าส่วนเบี่ยงเบนมาตรฐาน (Zhou, 2002. p. 52)

1.3 ผลการวิจัย พบว่า วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอส เอสอี ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ ไม่แตกต่างจาก วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์และแบบอีเอ็ม อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่ม และแบบหลายขั้นตอน จำนวนข้อมูลสูญหายค่อนข้างต่ำ คือเท่ากับ 5%-10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) และวิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ แตกต่างจากวิธีการจัดการข้อมูลสูญหายแบบอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเป็นค่าที่สูงที่สุด เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่ม และแบบหลายขั้นตอน จำนวนข้อมูลสูญหายอยู่ในระดับสูงที่สุดคือเท่ากับ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) ซึ่งไม่สอดคล้องกับสมมุติฐานการวิจัยที่ตั้งไว้ ทั้งนี้สามารถอธิบายได้ว่า ในการวิจัยครั้งนี้ผู้วิจัยสร้างข้อมูลสูญหายแบบสุ่ม (randomly missing data) เมื่อตัดข้อมูลสูญหายออกจึงยังทำให้ข้อมูลที่เหลืออยู่เป็นตัวแทนของข้อมูลทั้งหมด จากข้อมูลที่สร้างขึ้นมามีคะแนนสอบของนักเรียน (X) และเกรดเฉลี่ยของนักเรียน (Y) แล้วสร้างข้อมูลสูญหายแบบสุ่มในตัวแปรเกรดเฉลี่ย ผู้วิจัยได้นำข้อมูลมาสร้างตัวแปรใหม่อีก 1 ตัวแปร คือ ตัวแปรสูญหาย (MISS) ค่าของตัวแปรสูญหายมีค่าเท่ากับ 1 เมื่อเกรดเฉลี่ยของนักเรียนมีค่าสูญหาย และมีค่าเท่ากับ 0 เมื่อตัวแปรเกรดเฉลี่ยของนักเรียนไม่มีค่าสูญหาย แล้วนำค่าของตัวแปรใหม่นี้ไปหาความสัมพันธ์กับคะแนนสอบของนักเรียน ผลปรากฏว่า ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าสาเหตุของการสูญหายไม่มีความเกี่ยวข้องกับคะแนนสอบ ลักษณะของการสูญหายแบบนี้เรียกว่าเป็น MCAR (missing completely at random) ถ้าข้อมูลมีลักษณะเป็น MCAR ก็จะประมาณค่าได้อย่างดี (Little & Rubin, 1987; Rubin, 1976) และถ้าใช้วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ก็จะไม่มีอคติ (Cerrilo et al., 2000. http://www.lrdc.pitt.edu/hple/Publications/PERSIST.PDF) ประกอบกับค่าสัมประสิทธิ์สหสัมพันธ์ได้มาจากสูตร $r_{xy} = \frac{cov(x,y)}{S_x S_y} = \frac{\sum(x-\bar{x})(y-\bar{y})}{S_x S_y(n-1)}$ ค่าที่เปลี่ยนแปลงไปคือ $(y-\bar{y})$ และ s_y เพราะตัวแปร Y คือเกรดเฉลี่ยของนักเรียนจะเกิดการสูญหายจากการกำหนดของผู้วิจัย จากการศึกษาวิจัยพบว่าเมื่อจำนวนข้อมูลสูญหายสูงขึ้นถึง 30% วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของความแปรปรวนสูงที่สุด ดังนั้นจึงน่าจะเป็นเหตุผลหนึ่งที่ทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้เข้าใกล้ค่าของประชากร ส่วนเงื่อนไขอื่น ๆ ที่ทำให้ค่าความแม่นยำของ

สัมประสิทธิ์สหสัมพันธ์ มีค่าสูงที่สุด คือ จำนวนข้อมูลสูญหายค่อนข้างสูง (10%-30%) และความสัมพันธ์สูง (r=.70) ได้อภิปรายมาแล้วในหัวข้อ 2.1 ซึ่งผลการวิจัยดังกล่าวสอดคล้องกับงานวิจัยของเฟรดและลี (Fred & Lii, 1998. http://www.findarticles.com ; Kromrey & Hines, 1994. p. 573-593) ที่พบว่า วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์มีความถูกต้องเพราะว่าเมื่อลบข้อมูลออกไปข้อมูลที่เหลืออยู่ยังเป็นโครงสร้างของข้อมูลจริง และสามารถประมาณค่าพารามิเตอร์ได้ถูกต้องเมื่อจำนวนข้อมูลสูญหายเท่ากับ 30% และรูท (Roth, 1999) ได้วิพากษ์วิจารณ์ว่าวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าด้วยวิธีการถดถอย (regression imputation) ดีกว่าวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ แต่จากงานวิจัยของเขา พบว่า วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์มีอคตินอ้ยหรือไม่มีเลย และการกระจายรอบ ๆ คะแนนจริงน้อย ยังเกอร์และฟอร์ไซท์ (Younger & Forsyth, 1998. p. 203) ได้กล่าวว่า เมื่อใช้วิธีการจัดการข้อมูลสูญหาย การประมาณค่าสัมประสิทธิ์การถดถอย ค่าเฉลี่ย ความแปรปรวน สัมประสิทธิ์สหสัมพันธ์ และอื่น ๆ มีอคติ ถึงแม้ว่าหลายสถานการณ์วิธีการแทนค่าข้อมูลสูญหายอาจจะให้อคติในการประมาณค่าพารามิเตอร์น้อยกว่าการตัดข้อมูลสูญหายออก แต่ก็ยังไม่มีความชัดเจนในวิธีการแทนค่าหลาย ๆ วิธี

2. จากสมมุติฐานการวิจัยที่กล่าวว่า วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอีน่าจะมีอำนาจการทดสอบดีกว่าวิธีอื่น เมื่อใช้วิธีการสุ่มตัวอย่าง ความสัมพันธ์ระหว่างตัวแปร และจำนวนข้อมูลสูญหายที่แตกต่างกัน

2.1 ผลการวิจัย พบว่า วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าอำนาจการทดสอบ เมื่อใช้การทดสอบความสัมพันธ์สูงกว่าวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์และการแทนค่าแบบอีเอ็ม เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหายทุกระดับ (5%,10%,20% และ 30%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ซึ่งสอดคล้องกับสมมุติฐานการวิจัยที่ตั้งไว้ ทั้งนี้สามารถอธิบายได้ว่าในการประมาณค่าโดยใช้วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี มีขั้นตอนในการประมาณค่าข้อมูล 2 ขั้นตอน 1). ประมาณค่าข้อมูลสูญหายโดยวิธีการถดถอยอย่างง่าย (Simple regression) จากสมการ $\hat{Y} = a + bX$ ใช้ค่าสถิติคือ $a = \overline{Y} - b\overline{X}$ และ $b = r_{xy}\frac{S_y}{S_x}$ เป็นค่าโดยประมาณ แต่วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี ประมาณค่า $\overline{X}, \overline{Y}, r_{xy}, S_x$ และ S_y โดยการทำซ้ำ จำนวน 1,000 ครั้ง จึงทำให้มั่นใจได้ว่า สมการ $\hat{Y} = a + bX$ เป็นสมการที่แท้จริงของประชากร 2). สร้างสมการทำนายข้อมูลสูญหายจากข้อมูลทั้งหมด ทั้งข้อมูลสมบูรณ์

และข้อมูลสูญหายที่ได้จากการทำนายโดยใช้วิธีการถดถอยอย่างง่าย (Simple regression) จำนวน 1,000 ครั้ง แล้วเลือกสมการที่มีความคลาดเคลื่อนมาตรฐานในการประมาณค่าน้อยที่สุด เป็นสมการทำนายข้อมูลสูญหายข้อมูลที่แทนเข้าไปจึงน่าจะใกล้เคียงกับข้อมูลเดิมของประชากร เมื่อนำไปคำนวณสัมประสิทธิ์สหสัมพันธ์แล้วทดสอบสมมุติฐานเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร การปฏิเสธสมมุติฐานสูญน่าจะมีมากขึ้น เพราะค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้มีความใกล้เคียงกับค่าของประชากรดังกล่าว แต่เนื่องจากสัมประสิทธิ์สหสัมพันธ์ของประชากรมีค่าไม่เท่ากับศูนย์ การปฏิเสธสมมุติฐานสูญในการทดสอบนี้ ก็คือ อำนาจการทดสอบ ดังนั้นเมื่อใช้วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี จึงน่าจะทำให้อำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์มีค่าสูงสุด สอดคล้องกับคำกล่าวของราจเมเคอร์ ที่กล่าวว่า ประเด็นที่สำคัญของวิธีการแทนค่า คือ รักษาขนาดของกลุ่มตัวอย่างและผลที่ตามมาก็คือ รักษาอำนาจการทดสอบ (Raaijmakers, 1999. p. 729)

เมื่อเปรียบเทียบวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอีกับวิธีการจัดการข้อมูลสูญหายแบบอีเอ็ม (EM algorithm) พบว่า วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์สูงกว่าวิธีการจัดการข้อมูลสูญหายแบบอีเอ็ม ทั้งนี้สามารถอธิบายได้ว่า วิธีการจัดการข้อมูลสูญหายแบบอีเอ็ม (EM algorithm) จะแทนค่าข้อมูลสูญหายจากวิธีการถดถอย โดยการสร้างค่าสถิติต่าง ๆ จากข้อมูลที่เหลืออยู่และประมาณค่าความแปรปรวนร่วม ดำเนินการทำซ้ำจนค่าความแปรปรวนร่วมไม่แตกต่างกัน แต่เนื่องจากการแทนค่าข้อมูลสูญหายครั้งแรกในสมการ $\hat{Y} = a + bX$ ใช้ค่าสถิติที่ได้มาจากกลุ่มตัวอย่างที่มีข้อมูลสมบูรณ์ตัดหน่วยตัวอย่างที่มีข้อมูลสูญหายออกไปการประมาณค่าพารามิเตอร์ด้วยค่าเหล่านี้จึงมีอคติ (bias) หมายความว่า ค่าที่ได้อาจจะไม่ใช่ค่าที่แท้จริงของประชากร เมื่อนำไปสร้างสมการทำนายข้อมูลสูญหายแล้วคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ และทดสอบสมมุติฐานว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันหรือไม่ จึงน่าจะทำให้การปฏิเสธสมมุติฐานสูญได้น้อยกว่าเมื่อใช้วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี การปฏิเสธสมมุติฐานสูญที่ไม่จริงคืออำนาจการทดสอบ ดังนั้นค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์จึงน้อยกว่าวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี และเมื่อใช้วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ ได้ค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์ต่ำกว่าวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ทั้งนี้สามารถอธิบายได้ว่า วิธีการจัดการข้อมูลสูญหายโดยการตัดข้อมูลสูญหายออกแบบลิสท์ไวส์ตัดกลุ่มตัวอย่างที่มีข้อมูลสูญหายออกไปทำให้ขนาดกลุ่มตัวอย่างน้อยลง ลดอำนาจการทดสอบทางสถิติ ขนาดของกลุ่ม

ตัวอย่างมีผลต่ออำนาจการทดสอบ ถ้าขนาดของกลุ่มตัวอย่างใหญ่ขึ้น อำนาจการทดสอบจะสูงขึ้น (สุชาดา บวรกิติวงศ์, 2541. หน้า 16-19) ในการทดสอบความสัมพันธ์อำนาจการทดสอบจะเพิ่มขึ้นเมื่อค่าความสัมพันธ์แตกต่างจากศูนย์มากขึ้น ในการวิจัยโดยทั่วไปไม่สามารถควบคุมค่าความสัมพันธ์ได้ นักวิจัยสามารถควบคุมขนาดของกลุ่มตัวอย่างและระดับนัยสำคัญได้ กล่าวได้ว่า เมื่อ $\rho \# 0$ อำนาจการทดสอบจะเพิ่มขึ้นเมื่อขนาดของกลุ่มตัวอย่างและค่านัยสำคัญเพิ่มมากขึ้น

เมื่อพิจารณาเงื่อนไขอื่น ๆ มีประเด็นที่น่าสนใจ ผู้วิจัยได้นำมาอภิปรายดังต่อไปนี้ คือ เมื่อใช้วิธีการสุ่มแบบแบ่งชั้น ได้ค่าอำนาจการทดสอบความสัมพันธ์ต่ำกว่าวิธีการสุ่มตัวอย่างแบบกลุ่มและแบบหลายขั้นตอน ทั้งนี้สามารถอธิบายได้ว่า วิธีการสุ่มตัวอย่างทั้งสามวิธีจะแบ่งข้อมูลออกเป็นชั้นภูมิย่อย ๆ วิธีการสุ่มตัวอย่างแบบแบ่งชั้นแบ่งข้อมูลออกเป็น 3 ชั้นภูมิ วิธีการสุ่มตัวอย่างแบบกลุ่มแบ่งข้อมูลออกเป็นชั้นภูมิย่อย 14 ชั้นภูมิ และวิธีการสุ่มตัวอย่างแบบหลายขั้นตอนแบ่งข้อมูลออกเป็นชั้นภูมิย่อย 21 ชั้นภูมิ การแบ่งข้อมูลออกเป็นชั้นภูมิย่อยมากกว่าจะทำให้การประมาณค่ามีความแม่นยำมากกว่า (Cochran, 1977) จึงน่าจะเป็นเหตุผลหนึ่งที่ทำให้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นได้ค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์ต่ำกว่าวิธีการสุ่มตัวอย่างแบบกลุ่มและวิธีการสุ่มตัวอย่างแบบหลายขั้นตอน และเมื่อใช้ข้อมูลที่มีความสัมพันธ์ต่ำ (r=.30) ได้ค่าอำนาจการทดสอบความสัมพันธ์ต่ำกว่าข้อมูลที่มีความสัมพันธ์ปานกลาง (r=.50) และสูง (r=.70) ทั้งนี้สามารถอธิบายได้ว่า ในการทดสอบสมมุติฐานเกี่ยวกับความสัมพันธ์จะนำค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้แปลงเป็นค่าที แล้วจึงทดสอบสมมุติฐานตามสูตร $t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$ จะเห็นว่าถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าน้อยค่าทีก็จะมีค่าน้อยกว่าเมื่อสัมประสิทธิ์สหสัมพันธ์มีค่าสูง เมื่อนำไปทดสอบสมมุติฐานโอกาสที่จะปฏิเสธสมมุติฐานสูญเมื่อสมมุติฐานสูญไม่จริง (อำนาจการทดสอบ) ก็จะมีน้อย ซึ่งสอดคล้องกับหลักการในการทดสอบสมมุติฐานเกี่ยวกับความสัมพันธ์ เมื่อค่าความสัมพันธ์แตกต่างจากศูนย์มากขึ้น อำนาจการทดสอบจะมากขึ้น ดังนั้นจึงน่าจะเป็นเหตุผลหนึ่งที่ทำให้ เมื่อใช้ข้อมูลที่มีความสัมพันธ์ต่ำอำนาจการทดสอบความสัมพันธ์ก็จะต่ำด้วย

2.2 ผลการเปรียบเทียบอำนาจการทดสอบ ถ้าใช้การทดสอบความสัมพันธ์ การทดสอบที และการทดสอบเอฟ ได้ผลสอดคล้องกันทุกกรณี ไม่ว่าจะใช้วิธีการสุ่มตัวอย่างแบบใด จำนวนข้อมูลสูญหายระดับใด ความสัมพันธ์ระหว่างตัวแปรระดับใดและวิธีการจัดการข้อมูลสูญหายแบบใด ทั้งนี้สามารถอธิบายได้ว่า การทดสอบสมมุติฐานโดยใช้การทดสอบที (t-test)

ในการวิจัยครั้งนี้ใช้สูตร $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$ กรณีความแปรปรวนเท่ากัน $(\sigma_1^2 = \sigma_2^2)$ เพราะกลุ่มประชากรที่ผู้วิจัยสร้างขึ้นมาทั้ง 2 กลุ่มมีความแปรปรวนเท่ากัน ค่าสถิติที่จะขึ้นอยู่กับความแตกต่างระหว่างค่าเฉลี่ยและจำนวนกลุ่มตัวอย่าง จึงจะทำให้โอกาสในการปฏิเสธสมมุติฐานสูญสูง เมื่อพิจารณาค่าเฉลี่ยในการวิจัยครั้งนี้ ผู้วิจัยกำหนดให้ค่าเฉลี่ยมีความแตกต่างกัน 1 เท่า ของส่วนเบี่ยงเบนมาตรฐาน (1σ) ซึ่งเป็นความแตกต่างที่ค่อนข้างสูง ประกอบกับจำนวนกลุ่มตัวอย่างที่ใช้ในการทดสอบมีจำนวนเท่ากับ 350 คน ซึ่งเป็นขนาดกลุ่มตัวอย่างจำนวนมาก จึงทำให้ค่าความคลาดเคลื่อนที่ได้จากสูตร $\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}$ มีค่าน้อย ค่าทีที่คำนวณได้จึงน่าจะมีค่าสูงโอกาสในการปฏิเสธสมมุติฐานสูญจะมีมากขึ้น (ค่าเฉลี่ยของประชากรสองกลุ่มไม่เท่ากัน ดังนั้นในการปฏิเสธสมมุติฐานสูญ $H_0 : \mu_1 = \mu_2$ เมื่อสมมุติฐานสูญไม่จริง คือ อำนาจการทดสอบ) ทำให้ค่าอำนาจการทดสอบ (1σ) เมื่อใช้การทดสอบทีมีค่าสูง ซึ่งข้อมูลจากผลการวิจัย พบว่า ค่าอำนาจการทดสอบเท่ากับ 1 ทุกกรณี ดังนั้นเมื่อใช้การทดสอบทีในกรณีที่ความแตกต่างของค่าเฉลี่ยเท่ากับ 2 เท่าของส่วนเบี่ยงเบนมาตรฐาน (-1σ กับ $+1\sigma$) จึงทำให้ค่าอำนาจการทดสอบสูงเท่ากับ 1 ทุกกรณีด้วยเหตุผลดังที่ได้กล่าวมาแล้ว เมื่อพิจารณาอำนาจการทดสอบจากการทดสอบเอฟ ด้วยเหตุที่ว่า ค่าสถิติทีและค่าสถิติเอฟมีความสัมพันธ์กัน พิจารณาได้จากสูตร

1). $t = \frac{z}{\sqrt{\frac{\chi^2_{(n-1)}}{(n-1)}}}$

2). $F = \frac{\chi^2_{(\upsilon_1)}/\upsilon_1}{\chi^2_{(\upsilon_2)}/\upsilon_2}$

เมื่อนำค่าสถิติทีมายกกำลังสองจะได้เท่ากับ

$$t^2 = \frac{z^2}{\chi^2_{(n-1)}/(n-1)} = \frac{\chi^2_{(1)}/1}{\chi^2_{(n-1)}/(n-1)}$$

ค่าสถิติทียกกำลังสองก็คือค่าสถิติเอฟที่มีชั้นแห่งความเป็นอิสระเท่ากับ 1 และ υ_2

$$F = \frac{\chi^2_{(1)}/1}{\chi^2_{(\upsilon_2)}/\upsilon_2}$$

ดังนั้น $t^2(\upsilon) = F(1,\upsilon_2)$ เมื่อ $\upsilon = \upsilon_2$ จากผลการศึกษาที่พบว่าอำนาจการทดสอบเมื่อใช้การทดสอบทีทดสอบความแตกต่างของค่าเฉลี่ยที่มีความแตกต่างกัน 1 เท่า และ 2 เท่า ของส่วนเบี่ยงเบนมาตรฐาน มีค่าสูงเป็น 1 ทุกกรณี และจากความสัมพันธ์ของค่าสถิติทีและค่าสถิติเอฟดังกล่าว จึงทำให้อำนาจการทดสอบเมื่อใช้การทดสอบเอฟสูงเป็น 1 ทุกกรณีด้วย และเมื่อพิจารณาอำนาจการทดสอบจากการทดสอบความสัมพันธ์ก็พบว่ามีค่าอำนาจการทดสอบค่อนข้างสูง เช่นเดียวกัน โดยเฉพาะเมื่อความสัมพันธ์ระหว่างข้อมูลอยู่ในระดับปานกลาง (r=.50) และสูง (r=.70) อำนาจการทดสอบมีค่าสูงเป็น 1 ทุกกรณี ทั้งนี้การทดสอบสมมุติฐานเกี่ยวกับความสัมพันธ์ใช้สูตร $t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$ ซึ่งมีความสัมพันธ์กับการทดสอบที ดังที่ได้อภิปรายเกี่ยวกับค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์มาแล้ว ประกอบกับการทดสอบสมมุติฐานเกี่ยวกับตัวแปรสองตัวว่ามีความสัมพันธ์กันหรือไม่ เมื่อทดสอบแล้วพบว่ามีความสัมพันธ์กัน ถ้านำไปทดสอบความแตกต่างก็จะมีความแตกต่างกันด้วย ซึ่งจะเห็นได้ว่า การทดสอบความสัมพันธ์ การทดสอบที และการทดสอบเอฟ มีความสัมพันธ์กันจากเหตุผลที่กล่าวมาจึงทำให้ผลการเปรียบเทียบอำนาจการทดสอบ ถ้าใช้การทดสอบความสัมพันธ์ การทดสอบที และการทดสอบเอฟ ได้ผลสอดคล้องกันทุกกรณี

3. จากสมมุติฐานการวิจัยที่กล่าวว่า ความแม่นยำน่าจะขึ้นอยู่กับปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายกับจำนวนข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปรที่แตกต่างกัน

3.1 ผลการวิจัย พบว่า ปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปร ที่มีต่อความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ ไม่มีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งไม่สอดคล้องกับสมมุติฐานการวิจัย ทั้งนี้สามารถอธิบายได้ว่า การมีปฏิสัมพันธ์ของตัวแปร 3 ตัว เมื่อเปรียบเทียบไปยังระดับต่าง ๆ ของตัวแปรที่ 4 กระสวนการเพิ่มขึ้นของตัวแปรตามความแม่นยำของค่าเฉลี่ย ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ มีลักษณะสม่ำเสมอ ลักษณะดังกล่าวบ่งบอกถึง การเพิ่มหรือสดลงของค่าตัวแปรตามเมื่อเปรียบเทียบไปยังระดับต่าง ๆ ของตัวแปรที่ 4 มีความคงที่ จึงทำให้ไม่เกิดปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปร เมื่อพิจารณาข้อค้นพบจากการวิจัย พบว่า ค่าความแม่นยำของค่าเฉลี่ยเลขคณิต เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น จำนวนข้อมูลสูญหายอยู่ในระดับน้อยที่สุด (5%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็ม ได้ค่าความแม่นยำของค่าเฉลี่ย

เลขคณิตสูงที่สุด แต่เมื่อจำนวนข้อมูลสูญหายสูงขึ้นเป็น 10%-30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) เช่นเดียวกัน วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุด

เมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายน้อยที่สุด (5%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็ม ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุด ในกรณีจำนวนข้อมูลสูญหายเท่ากับ 10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็ม ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุด แต่เมื่อจำนวนข้อมูลสูญหายอยู่ในระดับสูง (20% และ 30%) ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุด

เมื่อใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จำนวนข้อมูลสูญหายอยู่ระหว่าง 5%-20% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็ม ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุด ถึงแม้ว่าในกรณีที่จำนวนข้อมูลสูญหายเท่ากับ 10% วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตดีกว่าวิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็ม แต่ค่าความแม่นยำของทั้งสองวิธีเกือบเท่ากัน แต่เมื่อจำนวนข้อมูลสูญหายสูงที่สุดคือเท่ากับ 30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอี ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุด

จากข้อค้นพบความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) จะทำให้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุด ไม่ว่าจะใช้วิธีการสุ่มตัวอย่างแบบใด วิธีการจัดการข้อมูลสูญหายแบบใด และจำนวนข้อมูลสูญหายระดับใด มีเพียงในกรณีที่ใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายเท่ากับ 10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็ม ได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิตสูงที่สุด และเมื่อพิจารณาเปอร์เซ็นต์ความแม่นยำของค่าเฉลี่ยเลขคณิต พบว่า มีค่าใกล้เคียงกันมาก หมายความว่า ค่าเฉลี่ยเลขคณิตที่คำนวณได้จากเงื่อนไขต่าง ๆ ของการทดลองใกล้เคียงกันมากกับค่าของประชากร เมื่อพิจารณาค่าความแม่นยำของความแปรปรวน วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์จะทำให้ค่าความแม่นยำของความแปรปรวนสูงที่สุด ไม่ว่าจะใช้วิธีการสุ่มตัวอย่างแบบใด ความสัมพันธ์ระหว่างตัวแปรระดับใด และจำนวน

ข้อมูลสูญหายระดับใด และเมื่อพิจารณาค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) จะทำให้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์สูงที่สุด ไม่ว่าจะใช้วิธีการสุ่มตัวอย่างแบบใด วิธีการจัดการข้อมูลสูญหายแบบใด และจำนวนข้อมูลสูญหายระดับใด ลักษณะแบบนี้น่าจะทำให้ผลต่างของตัวแปรตาม (ความแม่นยำของค่าเฉลี่ยเลขคณิต ความแม่นยำของความแปรปรวน และความแม่นยำของสัมประสิทธิ์สหสัมพันธ์) ภายใต้ระดับหนึ่งของแฟคเตอร์หนึ่ง เมื่อเปรียบเทียบไปยังอีกระดับหนึ่งของแฟคเตอร์หนึ่งไม่แตกต่างกัน หรือความแตกต่างของตัวแปรตามเนื่องมาจากความแตกต่างของตัวแปรอิสระตัวใดตัวหนึ่งมีค่าเท่า ๆ กันในทุกระดับของตัวแปรอีกตัวหนึ่งซึ่งแสดงว่าไม่มีผลเนื่องมาจากปฏิสัมพันธ์ระหว่างตัวแปร (บุญเรียง ขจรศิลป์, 2543. หน้า 87) ลักษณะดังกล่าวจึงมีผลทำให้ไม่เกิดปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการจัดการข้อมูลสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธ์ระหว่างตัวแปร

3.2 ผลการวิจัย พบว่า ปฏิสัมพันธ์สามทางที่มีต่อความแม่นยำของความแปรปรวนมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีข้อค้นพบดังต่อไปนี้

3.2.1 ปฏิสัมพันธ์ระหว่างระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายและจำนวนข้อมูลสูญหาย มีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่า ค่าความแม่นยำของความแปรปรวนจะสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายทุกระดับ จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ และเมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายน้อยที่สุด (5%) จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี ค่าความแม่นยำของความแปรปรวนจะต่ำลงเรื่อย ๆ เมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นแบบกลุ่ม และแบบหลายขั้นตอน จำนวนข้อมูลสูญหายค่อนข้างสูง (20%-30%) จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี

3.2.2 ปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปร มีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่า ค่าความแม่นยำของความแปรปรวนจะสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม ความสัมพันธ์ระหว่างตัวแปรทุกระดับ จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ และเมื่อใช้วิธีการสุ่มแบบหลายขั้นตอน ความสัมพันธ์ระหว่างตัวแปรทุกระดับ จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ แต่ถ้าใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับปานกลาง (r=.50) และสูง (r=.70) จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ ค่าความแม่นยำของความแปรปรวนจึงจะสูงที่สุด ค่าความแม่นยำของความแปรปรวนจะต่ำลงเรื่อย ๆ เมื่อระดับความ

สัมพันธ์ระหว่างตัวแปรต่ำลง โดยเฉพาะเมื่อใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี ความสัมพันธ์ระหว่างตัวแปรต่ำสุด (r=.30) ได้ค่าความแม่นยำของความแปรปรวนต่ำที่สุด

3.2.3 ปฏิสัมพันธ์ระหว่าง วิธีการจัดการข้อมูลสูญหายกับจำนวนข้อมูลสูญหายและความสัมพันธ์ระหว่างตัวแปร มีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่า ค่าความแม่นยำของความแปรปรวนจะสูงที่สุดเมื่อใช้วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ จำนวนข้อมูลสูญหายทุกระดับและความสัมพันธ์ระหว่างตัวแปรทุกระดับ แต่ถ้าใช้วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี จำนวนข้อมูลสูญหายจะต้องอยู่ในระดับน้อยที่สุด คือ เท่ากับ 5% เท่านั้น และความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับปานกลาง (r=.50) และสูง (r=.70) ค่าความแม่นยำของความแปรปรวนจะต่ำลงเรื่อย ๆ เมื่อจำนวนข้อมูลสูญหายเพิ่มขึ้นและความสัมพันธ์ระหว่างตัวแปรลดลง โดยเฉพาะเมื่อใช้วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี ทุกระดับความสัมพันธ์ระหว่างตัวแปร จำนวนข้อมูลสูญหายอยู่ในระดับสูง คือ เท่ากับ 20% และ 30% ได้ค่าความแม่นยำของความแปรปรวนต่ำที่สุด

จากข้อค้นพบดังกล่าว วิธีการสุ่มตัวอย่างแบบกลุ่มและวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ ค่อนข้างมีผลอย่างมากที่ทำให้ค่าความแม่นยำของความแปรปรวนดีที่สุด ทั้งนี้สามารถอธิบายได้ว่า วิธีการสุ่มตัวอย่างแบบกลุ่มในการวิจัยครั้งนี้มีขั้นตอนการสุ่มดังนี้ 1). แบ่งประชากรออกเป็นกลุ่มทั้งหมด 14 กลุ่ม 2). สุ่มกลุ่มออกมาครึ่งหนึ่งโดยการสุ่มอย่างง่าย 3). กำหนดขนาดกลุ่มตัวอย่างในแต่ละกลุ่มแบบนีย์แมน 4). สุ่มตัวอย่างตามกลุ่มที่สุ่มได้โดยการสุ่มอย่างง่าย วิธีการสุ่มตัวอย่างแบบกลุ่ม แบ่งประชากรออกเป็นกลุ่มย่อยถึง 14 กลุ่ม แล้วจึงสุ่มออกมา 7 กลุ่ม การแบ่งประชากรออกเป็นกลุ่มย่อย ๆ ที่มากขึ้นน่าจะประมาณค่าพารามิเตอร์ได้ถูกต้องมากขึ้น ซึ่งสอดคล้องกับที่ คอคแครน ได้กล่าวว่า จำนวนชั้นภูมิที่มากกว่าจะทำให้การประมาณค่าพารามิเตอร์มีความแม่นยำมากกว่า (Cochran, 1977) เมื่อพิจารณาจากข้อมูล ค่าความแปรปรวนในแต่ละกลุ่มมีค่าใกล้เคียงกับค่าความแปรปรวนของประชากรมาก ($\sigma^2 = .15$) เมื่อสุ่มตัวอย่างในแต่ละกลุ่มขึ้นมาจึงน่าจะทำให้การสุ่มแบบนี้รักษาความแปรปรวนของประชากรไว้ได้มากกว่าการสุ่มแบบอื่น ๆ ประกอบกับเมื่อใช้วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ วิธีนี้จะตัดข้อมูลสูญหายออกไปแล้วนำข้อมูลที่เหลืออยู่มาวิเคราะห์ ข้อมูลสูญหายในการวิจัยครั้งนี้เป็นการ สูญหายแบบสุ่ม (ramdomly

missing data) เมื่อตัดข้อมูลสูญหายออกไปจึงยังทำให้ข้อมูลที่เหลืออยู่เป็นตัวแทนของประชากรทั้งหมดได้ และผู้วิจัยได้ตรวจสอบลักษณะของข้อมูลเพิ่มเติม โดยการสร้างตัวแปรใหม่ขึ้นมาอีก 1 ตัวแปร คือ ตัวแปรสูญหาย (MISS) ค่าของตัวแปรสูญหายมีค่าเท่ากับ 1 เมื่อเกรดเฉลี่ยของนักเรียนมีค่าสูญหาย และมีค่าเท่ากับ 0 เมื่อตัวแปรเกรดเฉลี่ยของนักเรียนไม่มีค่าสูญหาย แล้วนำค่าของตัวแปรใหม่นี้ไปหาความสัมพันธ์กับคะแนนสอบของนักเรียน ผลปรากฏว่า ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าสาเหตุของการสูญหายไม่มีความเกี่ยวข้องกับคะแนนสอบ ลักษณะของการสูญหายแบบนี้เรียกว่าเป็น MCAR (missing completely at random) ถ้าข้อมูลมีลักษณะเป็น MCAR ก็จะประมาณค่าได้อย่างดี (Little & Rubin, 1987; Rubin, 1976) และถ้าใช้วิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ก็จะไม่มีอคติ (Cerrilo et al., 2000. p. 7) ประกอบกับความแปรปรวนได้มาจากสูตร $S^2 = \frac{\sum(X - \bar{X})^2}{n-1}$ เป็นการพิจารณาข้อมูลแต่ละตัวเทียบกับค่าเฉลี่ย จากที่กล่าวมาเมื่อข้อมูลมีการสูญหายแบบสุ่ม ข้อมูลที่เหลืออยู่จะมีความเป็นตัวแทนของข้อมูลทั้งหมดค่อนข้างสูง เมื่อนำมาหาความแปรปรวนจึงไม่ทำให้โครงสร้างของข้อมูลเปลี่ยนแปลงไป จากเหตุผลทั้งสองประการที่กล่าวมาจึงทำให้วิธีการสุ่มตัวอย่างแบบกลุ่มและวิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์มีผลอย่างมากทำให้ค่าความแม่นยำของความแปรปรวนสูงที่สุด

3.3 ผลการวิจัย พบว่า ปฏิสัมพันธ์สามทางระหว่างวิธีการสุ่มตัวอย่างกับวิธีการจัดการข้อมูลสูญหายและจำนวนข้อมูลสูญหาย ที่มีต่อความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ มีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่า ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์จะสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ จำนวนข้อมูลสูญหายอยู่ระหว่าง 5%-20% และเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี จะต้องใช้ข้อมูลที่มีจำนวนข้อมูลสูญหายค่อนข้างน้อยคืออยู่ระหว่าง 5%-10% เท่านั้น จึงจะได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์สูงที่สุด ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์จะต่ำลงเรื่อย ๆ เมื่อจำนวนข้อมูลสูญหายเพิ่มมากขึ้น โดยเฉพาะเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี จำนวนข้อมูลสูญหายสูงสุดคือเท่ากับ 30% ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์ต่ำที่สุด จากข้อค้นพบดังกล่าว จำนวนข้อมูลสูญหายและวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ ค่อนข้างมีผลอย่างมากที่ทำให้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์สูงที่สุด ทั้งนี้สามารถอธิบายได้ว่า เมื่อข้อมูลสูญหายมี

เป็นจำนวนน้อย 5%-10% ความเป็นตัวแทนของข้อมูลก็น่าจะมีมากกว่าเมื่อข้อมูลสูญหายเป็นจำนวนมาก และเมื่อใช้วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและแบบอีพีเอสเอสอี ก็น่าจะช่วยรักษาความเป็นตัวแทนของข้อมูลไว้ โครงสร้างของข้อมูลไม่น่าจะเปลี่ยนแปลงไปมากนัก ดังจะเห็นได้ว่าเมื่อจำนวนข้อมูลสูญหายตั้งแต่ 20% ขึ้นไป วิธีการแทนค่าข้อมูลสูญหายแบบอีเอ็มและแบบอีพีเอสเอสอี ก็ไม่ทำให้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์สูงที่สุด ประกอบผู้วิจัยได้อภิปรายมาแล้วว่า ข้อมูลสูญหายที่เกิดขึ้นเป็นการสูญหายแบบสุ่ม (randomly missing data) จึงทำให้ข้อมูลที่เหลืออยู่มีความเป็นตัวแทนของข้อมูลทั้งหมดค่อนข้างสูง ถึงแม้ว่าจะตัดข้อมูลสูญหายออกไปถึง 20% โดยใช้วิธีการตัดออกแบบลิสท์ไวส์ ก็ยังทำให้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์สูงที่สุด ดังนั้นจึงเป็นเหตุผลที่ทำให้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์จะสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ จำนวนข้อมูลสูญหายอยู่ระหว่าง 5%-20% และเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่มและแบบหลายขั้นตอน จัดการข้อมูลสูญหายโดยการแทนค่าแบบอีเอ็มและการแทนค่าแบบอีพีเอสเอสอี จะต้องใช้ข้อมูลที่มีจำนวนข้อมูลสูญหายค่อนข้างน้อยคืออยู่ระหว่าง 5%-10% เท่านั้น

ข้อค้นพบของการเกิดปฏิสัมพันธ์ที่ผู้วิจัยกล่าวมาสอดคล้องกับคำกล่าวของ เฟรดและลี (Fred & Lii, 1998) ที่กล่าวว่า การประเมินความถูกต้องของวิธีการจัดการข้อมูลสูญหายควรจะศึกษาในรูปของปฏิสัมพันธ์ด้วย วิธีการจัดการข้อมูลสูญหายที่แตกต่างกันย่อมมีปฏิสัมพันธ์กับตัวแปรอื่น ๆ ซึ่งมีผลต่อข้อค้นพบของการศึกษา เช่นเดียวกับที่ คอร์ทินา ให้ข้อสังเกตว่าผลของการจัดการข้อมูลสูญหายจะปะปนกันในสภาพการณ์ที่มีปฏิสัมพันธ์กัน (Fred & Lii , 1998. p. unpaged citting Cortina, 1993. p.915-922) คำกล่าวของนักวิจัยทั้งสองท่านสอดคล้องกับงานวิจัยของ เอ็นเดอร์ ที่ศึกษาพบว่า วิธีการจัดการข้อมูลสูญหาย ลักษณะความสัมพันธ์ระหว่างตัวแปร และจำนวนข้อมูลสูญหายมีปฏิสัมพันธ์กัน (Enders, 2001. p. 713-740)

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีที่ผู้วิจัยพัฒนาขึ้นจะดีที่สุดในบางสถานการณ์ ดังนั้นการเลือกใช้จะขึ้นอยู่กับเงื่อนไขต่าง ๆ ที่เหมาะสม โดยเฉพาะจำนวนกลุ่มตัวอย่างที่ใช้ในการวิจัยจะต้องใกล้เคียงกับการวิจัยครั้งนี้ จากผลการวิจัยมีข้อเสนอแนะสำหรับการนำไปใช้ดังต่อไปนี้

1. นักวิจัยควรดำเนินการทุกอย่างเพื่อป้องกันไม่ให้เกิดข้อมูลสูญหายเพราะยังไม่มีวิธีประมาณค่าข้อมูลสูญหายที่สามารถประมาณค่าได้เท่ากับข้อมูลจริง

2. นักวิจัยควรกำหนดวิธีการจัดการข้อมูลสูญหายไว้ในรายงานการวิจัย และเหตุผลของการใช้วิธีนั้น ข้อมูลเหล่านี้จะช่วยให้ผู้อ่านงานวิจัยเข้าใจผลการวิเคราะห์และการสรุปผลการวิจัยได้อย่างดี

3. เมื่อเกิดข้อมูลสูญหายขึ้นควรทำการวิเคราะห์ถึงสาเหตุของการสูญหายว่าเป็นแบบใด เพราะการเลือกใช้วิธีการจัดการข้อมูลสูญหายมีข้อตกลงเบื้องต้นเกี่ยวกับสาเหตุของการสูญหาย และสาเหตุของการสูญหายทำให้วิธีการจัดการข้อมูลสูญหายประมาณค่าพารามิเตอร์ได้แตกต่างกัน นักวิจัยสามารถทำได้โดยสร้างตัวแปรสูญหายขึ้นมาจากข้อมูลสูญหายแล้วนำไปหาความสัมพันธ์กับคะแนนอีกตัวแปรหนึ่งจะทำให้ทราบถึงสาเหตุของการสูญหายว่าเป็นแบบใด MCAR (Missing completely at random) หรือ MAR (Missing at random) การเลือกใช้วิธีการจัดการข้อมูลสูญหายก็จะถูกต้องมากยิ่งขึ้น

4. เมื่อนักวิจัยเลือกใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น ถ้าจำนวนข้อมูลสูญหายมีค่าน้อย คือเท่ากับ 5%-10% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) สามารถเลือกใช้วิธีการจัดการข้อมูลสูญหายแบบใดก็ได้ เพราะสามารถประมาณค่าพารามิเตอร์ คือ ค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ ไม่แตกต่างกันและอำนาจการทดสอบสูง ในกรณีที่ทำการทดสอบความสัมพันธ์ เมื่อความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ควรใช้วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีเพราะได้ค่าอำนาจการทดสอบสูงสุด แต่ถ้าจำนวนข้อมูลสูญหายอยู่ระหว่าง 20%-30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) ควรใช้วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ในกรณีที่ทำการทดสอบความสัมพันธ์ เมื่อความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ควรใช้วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีเพราะได้ค่าอำนาจการทดสอบสูงสุด

5. เมื่อนักวิจัยใช้วิธีการสุ่มตัวอย่างแบบกลุ่ม จำนวนข้อมูลสูญหายน้อยที่สุดคือเท่ากับ 5% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) สามารถเลือกใช้วิธีการจัดการข้อมูลสูญหายแบบใดก็ได้ เพราะสามารถประมาณค่าพารามิเตอร์ คือ ค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ ไม่แตกต่างกันและอำนาจการทดสอบสูง ในกรณีที่ทำการทดสอบความสัมพันธ์ เมื่อความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) ควรใช้วิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีเพราะได้ค่าอำนาจการทดสอบสูงสุด แต่ถ้าจำนวนข้อมูลสูญหายอยู่ระหว่าง 20%-30% ความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับสูง (r=.70) ควรใช้วิธี

การจัดการข้อมูลสูญหายแบบลิสท์ไวส์ ในกรณีที่ทำการทดสอบความสัมพันธ์ เมื่อความสัมพันธ์ระหว่างตัวแปรอยู่ในระดับต่ำ (r=.30) จำนวนข้อมูลสูญหายสูงที่สุดคือเท่ากับ 30% ไม่ควรใช้วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์เพราะได้ค่าอำนาจการทดสอบต่ำสุด

6. เมื่อนักวิจัยใช้วิธีการสุ่มตัวอย่างแบบหลายขั้นตอน จำนวนข้อมูลสูญหายทุกระดับ (5%-30%) ควรเลือกใช้วิธีการจัดการข้อมูลสูญหายแบบลิสท์ไวส์ เพราะสามารถประมาณค่าพารามิเตอร์ คือ ค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ไม่แตกต่างกันมากนัก และค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์ การทดสอบที และการทดสอบเอฟค่อนข้างสูงเกือบ 1 ทุกกรณี

ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. วิธีการจัดการข้อมูลสูญหายโดยการแทนค่าแบบอีพีเอสเอสอีผู้วิจัยได้พัฒนาขึ้นจากสมการ $\hat{Y} = a + bX$ โดยคัดเลือกสมการที่มีความคลาดเคลื่อนน้อยที่สุด แต่ก็ยังมีความคลาดเคลื่อนอยู่ ดังนั้นในการทำวิจัยต่อไปควรศึกษาถึงความคลาดเคลื่อนในการสร้างสมการถดถอยแล้วหาวิธีนำมารวมในสมการเพื่อทำให้การแทนค่าได้ถูกต้องมากยิ่งขึ้น

2. ควรศึกษาเปรียบเทียบวิธีการจัดการข้อมูลสูญหายโดยการตัดออกแบบลิสท์ไวส์ การแทนค่าแบบอีเอ็ม และการแทนค่าแบบอีพีเอสเอสอีที่รวมความคลาดเคลื่อน โดยใช้ข้อมูลที่มีการสูญหายแบบสุ่มที่ใช้ในการวิจัยครั้งนี้ และการสูญหายแบบเป็นระบบ (systematic missing data) คือ ลักษณะการสูญหายอาจจะเกิดขึ้นที่คะแนนสูงหรือคะแนนต่ำไม่ได้เกิดแบบสุ่ม

3. ควรนำตัวแปรจำนวนกลุ่มตัวอย่างที่นักวิจัยเชิงสำรวจค่อนข้างนิยมใช้ เช่น 200 300 400 500 600 จนถึง 1,000 เข้ามาเป็นตัวแปรในการศึกษาเปรียบเทียบวิธีการจัดการข้อมูลสูญหาย โดยกำหนดความสัมพันธ์ระหว่างตัวแปรให้อยู่ในระดับสูง (r=.70) ว่าจะได้ผลเป็นอย่างไร

4. ควรมีการศึกษาวิธีการจัดการข้อมูลสูญหายกับค่าสถิติอื่น ๆ เช่น ค่าสัมประสิทธิ์การถดถอยทั้งในรูปคะแนนดิบและคะแนนมาตรฐานโดยเฉพาะค่าสัมประสิทธิ์อิทธิพลในโมเดลลิสเรลและค่าสถิติที่เกี่ยวของกับการวัดคุณลักษณะทางจิตวิทยา เช่น ความตรง และความเที่ยง ฯลฯ

5. ควรมีการศึกษาวิธีการจัดการข้อมูลสูญหายที่นักสถิติได้พัฒนาไว้แล้วและมีโปรแกรมสำเร็จรูปช่วยในการวิเคราะห์ เช่น Hot deck imputation , FIML (Full information maximization likelihood) และ Multiple imputation เปรียบเทียบกับวิธีอื่น ๆ ในสถานการณ์ต่าง ๆ ว่าจะได้ผลเป็นอย่างไร

6. ควรมีการศึกษาวิจัยในลักษณะนี้แต่ใช้ตัวแปรทำนายมากกว่า 1 ตัวแปร อาจจะเป็น 2 ตัวแปร หรือ 3 ตัวแปร เพราะการใช้ตัวแปรทำนายมากกว่า 1 ตัวแปรน่าจะให้ผลการทำนายถูกต้องมากยิ่งขึ้น