การจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบ ความสามารถของผู้สอบ และการทำหน้าที่ต่างกันของข้อสอบ

เขมิกา อารมณ์

ดุษฎีนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาวิจัย วัดผลและสถิติการศึกษา
คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา
สิงหาคม 2561
ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

คณะกรรมการควบคุมดุษฎีนิพนธ์และคณะกรรมการสอบดุษฎีนิพนธ์ ได้พิจารณาดุษฎีนิพนธ์ของ เขมิกา อารมณ์ ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาวิจัย วัดผลและสถิติการศึกษา ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมดุษฎีนิพนธ์

.....อาจารย์ที่ปรึกษาหลัก

(รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม)

..... อาจารย์ที่ปรึกษาร่วม

(ดร.สมพงษ์ ปั้นหุ่น)

คณะกรรมการสอบดุษฎีนิพนธ์

.....ประธาน

(ดร.อาวีพร ปานทอง)

.....กรรมการ

(รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม)

..... กรรมการ

(ดร.สมพงษ์ ปั้นหุ่น)

..... กรรมการ

(ดร.ณัฐพร ภักดี)

คณะศึกษาศาสตร์อนุมัติให้รับดุษฎีนิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาวิจัย วัดผลและสถิติการศึกษา ของมหาวิทยาลัยบูรพา

..... คณบดีคณะศึกษาศาสตร์

(รองศาสตราจารย์ ดร.วิชิต สุรัตน์เรืองชัย)

วันที่ 9 เดือน พฤษภาคม พ.ศ. 2561

กิตติกรรมประกาศ

ดุษฎีนิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความกรุณาจาก รองศาสตราจารย์ ดร.ไพรัตน์ วงษ์นาม อาจารย์ที่ปรึกษาหลัก ดร.สมพงษ์ ปั้นหุ่น อาจารย์ที่ปรึกษาร่วม ที่กรุณาให้คำปรึกษา แนะนำ แนวทางตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ด้วยความละเอียดถี่ถ้วนและเอาใจใส่เสมอมา ผู้วิจัยรู้สึก ซาบซึ้งเป็นอย่างยิ่ง จึงขอกราบขอบพระคุณเป็นอย่างสูง ไว้ ณ โอกาสนี้

ขอขอบพระคุณ ดร.อาวีพร ปานทอง ประธานสอบปากเปล่า และ ดร.ณัฐพร ภักดี กรรมการสอบปากเปล่า ที่ได้กรุณาให้ข้อเสนอแนะ ในการปรับปรุงแก้ไขทำให้ดุษฎีนิพนธ์ฉบับนี้ มีความสมบูรณ์ยิ่งขึ้น

ขอขอบคุณ คุณมาโนชญ์ ใจกว้าง นักวิชาการคอมพิวเตอร์ คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา และอาจารย์สกรณ์ บุษบง อาจารย์ประจำสาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ที่กรุณาให้ความรู้ ให้คำปรึกษาในด้านการเขียน โปรแกรมทางสถิติ ทำให้งานวิจัยมีความสมบูรณ์ยิ่งขึ้น นอกจากนี้ยังได้รับความช่วยเหลือจาก เพื่อน ๆ พี่ ๆ และน้อง ๆ จากสาขาวิจัย วัดผลและสถิติการศึกษา สาขาวิชาจิตวิทยาการปรึกษา ที่ให้ความช่วยเหลือในด้านต่าง ๆ ทำให้ดุษฎีนิพนธ์ฉบับนี้สำเร็จลงได้ด้วยดี

ขอขอบคุณมหาวิทยาลัยราชภัฏบุรีรัมย์ ที่ให้การสนับสนุนทุนทางการศึกษา ผู้วิจัยจึงขอขอบพระคุณมา ณ ที่นี้ด้วย

ขอขอบพระคุณ คุณพ่ออุทิศ คุณแม่พลอย อารมณ์ นางสาวพิชญา นายบุญช่วย อารมณ์ พี่ ๆ น้อง ๆ ทุกคน ที่ให้กำลังใจและสนับสนุนผู้วิจัยเสมอมา

คุณค่าและประโยชน์ของดุษฎีนิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นกตัญญูกตเวทิตา แด่บุพการี บูรพาจารย์ และผู้มีพระคุณทุกท่านทั้งในอดีตและปัจจุบัน ที่ทำให้ข้าพเจ้าเป็นผู้มี การศึกษา และประสบความสำเร็จมาจนตราบเท่าทุกวันนี้

เขมิกา อารมณ์

53810169: สาขาวิชา: วิจัย วัดผลและสถิติการศึกษา; ปร.ด. (วิจัย วัดผลและสถิติการศึกษา)
คำสำคัญ: ข้อมูลสูญหาย/ การประมาณค่าทดแทนพหุ/ การทำหน้าที่ต่างกันของข้อสอบ

เขมิกา อารมณ์: การจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบ ความสามารถของผู้สอบ และการทำหน้าที่ต่างกันของข้อสอบ (A MANAGEMENT OF MISSING DATA TREATMENTS FOR ESTIMATING ITEM PARAMETERS, PERSON ABILITY AND DIFFERENTIAL ITEM FUNCTIONING) กรรมการควบคุมดุษฎีนิพนธ์: ไพรัตน์ วงษ์นาม, ค.ด., สมพงษ์ ปั้นหุ่น, ค.ด. 213 หน้า. ปี พ.ศ. 2561.

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อ 1) วิเคราะห์ค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบจากวิธีจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) 2) ศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio Test: LRT) ดำเนินการศึกษาจากการจำลองข้อมูลภายใต้โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 3PL) โดยการตรวจให้คะแนน 2 ค่า และจำลองข้อมูลภายใต้เงื่อนไขที่แปรเปลี่ยน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดของการทำหน้าที่ต่างกันของข้อสอบ จำนวน 32 เงื่อนไข (2x2x2x2x2) ในแต่ละเงื่อนไขจำลองข้อมูลวนซ้ำ 100 ครั้ง ผลการวิจัยสรุปได้ ดังนี้

1. ผลการประมาณค่าพารามิเตอร์ด้วยวิธีประมาณค่าพหุ พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าใกล้เคียงกัน และค่าที่ได้มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ ค่าความยากของข้อสอบ (b) ของกลุ่มอ้างอิง ค่าที่ได้ส่วนใหญ่สูงกว่าค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ และค่าต่ำกว่าในกลุ่มเปรียบเทียบ ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่มีแนวโน้มใกล้เคียงค่าความสามารถของผู้สอบของข้อมูลที่สมบูรณ์มากกว่า ค่าพารามิเตอร์ความสามารถของผู้สอบภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน

2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ผ่านตามเกณฑ์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 ($p < 0.05$) แต่มีอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำ ไม่ผ่านเกณฑ์ตามที่กำหนดทุกเงื่อนไข

53810169: MAJOR: EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS;
Ph.D. (EDUCATIONAL RESEARCH, MEASUREMENT AND STATISTICS)
KEYWORDS: MISSING DATA/ MULTIPLE IMPUTATION/ DETECTING DIFFERENTIAL ITEM FUNCTIONING

KHEMEKA AR-ROM: A MANAGEMENT OF MISSING DATA TREATMENTS FOR ESTIMATING ITEM PARAMETERS, PERSON ABILITY AND DIFFERENTIAL ITEM FUNCTIONING. DISSERTATION ADVISORY COMMITTEE: PAIRATANA WONGNAM, Ph.D., SOMPONG PANHOON, Ph.D. 213 P. 2018.

The purposes of this research were to: 1) analyze the item parameters and an examinee's ability through the treatment of the missing data with multiple imputation (MI), and 2) study the differential item functioning of the examination with the Likelihood Ratio Test (LRT). The study was conducted by simulating a two-parameter logistic model (2PL) on 2 values and setting up the data model under 5 variables including the conditions of missing the data, the sample size, the missing rate, the test length of the examination, the magnitude of DIF in 32 conditions (2x2x2x2x2) on an examination. Each function consisted of 100 replicates. The results were summarized as follows:

1. The item parameter estimated by multiple imputation (MI) were performed under missing at random (MAR) and missing not at random (MNAR) at the same level. The discrimination powers of the examination (a) was similar. The obtained value was higher than the parameter in the complete data. Although the obtained value on the difficulty of the test (b) in the reference group was greater than the parameter in the complete data, it was lower than the parameter in the complete data in the focal group. At the same level of MAR and MNAR, more item parameter estimates on an examinee's ability (θ) under MAR approached to parameter estimates in the complete data than those under missing not at random (MNAR).

2. The results of DIF with LRT could be used to control the type I error rate with the statistical significance at 0.05 ($p < 0.05$), but all of the powers to find the DIF was lower than the criteria.

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ..... จ

สารบัญ..... ฉ

สารบัญตาราง..... ช

สารบัญภาพ..... ฌ

บทที่

1 บทนำ..... 1

ความเป็นมาและความสำคัญของปัญหา..... 1

คำถามการวิจัย..... 9

วัตถุประสงค์ของการวิจัย..... 10

ขอบเขตของการวิจัย..... 10

นิยามศัพท์เฉพาะ..... 12

กรอบแนวคิดในการวิจัย..... 18

ประโยชน์ที่ได้รับจากการวิจัย..... 20

2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 21

ตอนที่ 1 แนวคิด ทฤษฎีเกี่ยวกับข้อมูลสูญหาย และแนวทาง
การจัดการข้อมูลสูญหาย..... 22

ตอนที่ 2 แนวคิด เกี่ยวกับโมเดลทฤษฎีการตอบสนองข้อสอบ..... 34

ตอนที่ 3 แนวคิด ทฤษฎีเกี่ยวกับการประมาณค่าพารามิเตอร์..... 42

ตอนที่ 4 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล
(Monte cario simulation study)..... 54

ตอนที่ 5 คุณสมบัติของตัวประมาณค่าที่ดี 56

ตอนที่ 6 แนวคิด ทฤษฎีเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ..... 60

ตอนที่ 7 งานวิจัยที่เกี่ยวข้อง..... 78

สารบัญ (ต่อ)

บทที่		หน้า
3	วิธีดำเนินการวิจัย	95
	ตัวแปรที่ใช้ในการวิจัย	95
	เงื่อนไขที่ใช้ในการจำลอง	97
	การจำลองข้อมูล	101
	การวิเคราะห์ข้อมูล	103
4	ผลการวิเคราะห์ข้อมูล	112
	ตอนที่ 1 ผลการวิเคราะห์ค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก (a) และความยาก (b) ด้วยวิธีจัดการข้อมูลสูญหายโดยใช้วิธีประมาณค่าพหุ (Multiple imputation: MI)	113
	ตอนที่ 2 ผลการวิเคราะห์ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ด้วยวิธีจัดการข้อมูลสูญหายโดยวิธีประมาณค่าพหุ (Multiple imputation: MI)	150
	ตอนที่ 3 ผลการเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก และความยาก	153
	ตอนที่ 4 ผลการเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ)	174
	ตอนที่ 5 ผลการตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบ	176
5	สรุป อภิปรายผล และข้อเสนอแนะ	180
	สรุปผลการวิจัย	180
	อภิปรายผลการวิจัย	187
	ข้อเสนอแนะ	188
บรรณานุกรม		190
ภาคผนวก		202
ประวัติย่อของผู้วิจัย		213

สารบัญตาราง

ตารางที่		หน้า
2-1	อนุพันธ์อันดับที่ 1 และอนุพันธ์อันดับที่ 2 สำหรับฟังก์ชันไลค์ลูฮูดของค่าพารามิเตอร์ข้อสอบ (a, b, c) และความสามารถของผู้เข้าสอบ (θ) ในแบบจำลองโลจิสติค 3 พารามิเตอร์.....	47
2-2	ค่าฟังก์ชันสารสนเทศสำหรับค่าพารามิเตอร์ข้อสอบ และค่าความสามารถของผู้สอบในแบบจำลองโลจิสติค 3 พารามิเตอร์.....	49
2-3	วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ที่เป็นเอกมิติจำแนกตามลักษณะของข้อมูล	69
2-4	วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีการตรวจให้คะแนนแบบทวิภาค (Dichotomous DIF) และพหุวิภาค (Polytomous DIF)	71
2-5	สรุปประเด็นรายละเอียดจากการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับแนวทางการจัดการกับข้อมูลสูญหาย.....	86
3-1	ค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบที่ใช้ในการจำลองข้อมูล.....	100
4-1	ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) โดยภาพรวม จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของกลุ่มอ้างอิง	114
4-2	ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของกลุ่มอ้างอิง.....	116
4-3	ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) โดยรายข้อ จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหาย ของกลุ่มอ้างอิง	120

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-4 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนก ของข้อสอบ (a) โดยภาครวม จำแนกตามความยาวของแบบสอบ อัตราสูญหาย ของข้อมูลขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหาย ของกลุ่มเปรียบเทียบ..... 123

4-5 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนก ของข้อสอบ (a) รายข้อ จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการ สูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหาย ของกลุ่มเปรียบเทียบ..... 125

4-6 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_R) รายข้อ จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของกลุ่มอ้างอิง... 129

4-7 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_R) โดยภาพรวม จำแนกตามความยาวของแบบสอบอัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหาย ของกลุ่มอ้างอิง..... 132

4-8 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_R) รายข้อ จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ..... 134

4-9 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_R) รายข้อ จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ..... 138

4-10 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนก ของข้อสอบ (b_F) โดยภาพรวม จำแนกตามความยาวของแบบสอบ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของกลุ่มเปรียบเทียบ..... 141

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-11 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_F) รายข้อ จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ..... 143

4-12 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_F) รายข้อ จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ..... 147

4-13 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) โดยภาพรวม จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล..... 151

4-14 ค่าเฉลี่ย diff และ diff error ของค่าความยากของข้อสอบ (b_R) โดยภาพรวมจากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง... 153

4-15 ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อจากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง..... 155

4-16 ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อจากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง..... 157

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4-17	ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) โดยภาพรวม จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ.....	159
4-18	ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ.....	160
4-19	ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ.....	162
4-20	ค่าเฉลี่ย diff และ diff error ของค่าความยากของข้อสอบ (b_R) โดยภาพรวม จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มอ้างอิง..	164
4-21	ค่าเฉลี่ย diff และ diff error ของค่าความยากของข้อสอบ (b_R) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง.....	165

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-22 ค่าเฉลี่ย diff และ diff error ของค่าความยากของข้อสอบ (b_R) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มอ้างอิง…….. 167

4-23 ค่าเฉลี่ย diff และ diff error ค่าอำนาจจำแนกของข้อสอบ (b_F) โดยภาพรวม จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ.. 169

4-24 ค่าเฉลี่ย diff และ diff error ค่าอำนาจจำแนกของข้อสอบ (b_F) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ…… 170

4-25 ค่าเฉลี่ย diff และ diff error ค่าอำนาจจำแนกของข้อสอบ (b_F) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ…… 172

4-26 ค่าเฉลี่ย diff และ diff error ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) โดยภาพรวม จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล… 175

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4-27	ค่าเฉลี่ยของความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF และสภาวะการสูญหายของข้อมูล.....	176
4-28	อำนาจการทดสอบ ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF และสภาวะการสูญหายของข้อมูล.....	178

สารบัญภาพ

ภาพที่		หน้า
1-1	กรอบแนวคิดในการวิจัย……	19
2-1	โค้งลักษณะข้อสอบแบบโมเดลการตอบสนองข้อสอบแบบโลจิสติกชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) ของตัวอย่างข้อสอบ 5 ข้อ….....	38
2-2	โค้งลักษณะข้อสอบแบบโมเดลการตอบสนองข้อสอบแบบโลจิสติกชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) ของตัวอย่างข้อสอบ 3 ข้อ……..	40
2-3	โค้งลักษณะข้อสอบแบบโมเดลการตอบสนองข้อสอบแบบโลจิสติกชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) ของตัวอย่างข้อสอบ 1 ข้อ……..….	41
2-4	กราฟแสดงการแจกแจงความน่าจะเป็นสะสมของตัวแปรสุ่ม X……..	55
2-5	โค้งคุณลักษณะข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) …….....	62
2-6	โค้งคุณลักษณะข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) ……	63
3-1	แผนการจำลองข้อมูล……..…	100
3-2	ขั้นตอนการจำลองของ Multiple Imputation……	103
3-3	ขั้นตอนการประมาณค่าและการตรวจสอบการการทำหน้าที่ต่างกันของข้อสอบ ….....	106
3-4	แสดงขั้นตอนการจำลองและการวิเคราะห์ข้อมูล……..….....	111
4-1	กราฟแสดงค่าเฉลี่ยของความคลาดเคลื่อนประเภทที่ 1 จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง อัตราสูญหายของข้อมูล ขนาดของ DIF และสภาวะการสูญหายของข้อมูล……	177
4-2	กราฟแสดงอำนาจการทดสอบ จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง อัตราสูญหายของข้อมูล ขนาดของ DIF และสภาวะการสูญหายของข้อมูล ……..	179

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

จากสภาพความเป็นจริงในปัจจุบัน การบริหารงานด้านต่าง ๆ เช่น ด้านการแพทย์ การศึกษา สังคมศาสตร์ วิทยาศาสตร์ ข้อมูลสารสนเทศ ถือว่าเป็นข้อมูลที่สำคัญที่นำมาประกอบการตัดสินใจ เพราะในการนำข้อมูลไปวิเคราะห์ประมวลผลเพื่อให้ได้ผลลัพธ์ที่ถูกต้องนั้นเป็นประโยชน์ต่อการนำไปใช้ในการวางแผนตัดสินใจอย่างมาก ซึ่งปัจจุบันได้มีการนำสถิติมาช่วยในการตัดสินใจมากขึ้น แต่สิ่งที่เป็นปัญหาละพบบ่อย คือ ข้อมูลไม่สมบูรณ์ (Incomplete data) ซึ่งความไม่สมบูรณ์อาจเกิดจากการที่ผู้ให้ข้อมูล ตอบคำถามไม่ครบถ้วน บางครั้งผู้ตอบไม่ตอบคำถามเมื่อมีการสอบถามซ้ำ หรือมีการลงข้อมูลไม่ครบถ้วน เราเรียกข้อมูลดังกล่าวว่า “ข้อมูลสูญหาย (Missing data)” ซึ่งปัญหาที่เกิดขึ้นดังกล่าว แม้กระทั่งนักจิตวิทยา นักวิจัย และผู้เชี่ยวชาญด้านการวัดผล ประเมินผลส่วนใหญ่มักประสบปัญหาคล้ายกันเกี่ยวกับปรากฏการณ์การสูญหายของการตอบสนองข้อสอบหรือแบบสอบ โดยในส่วนของการวัดผลทั้งด้านความสามารถทางสมองและอารมณ์ความรู้สึก ผู้สอบอาจละเว้นการตอบสนองข้อสอบหนึ่งข้อหรือมากกว่าหนึ่งข้อขึ้นไปเนื่องจากความไม่ตั้งใจ ไม่ทราบคำตอบ หรือไม่กล้าเดาคำตอบ ส่วนการตอบแบบสอบถามนั้น ผู้ตอบบางคนอาจเกิดความรู้สึกต่อต้านไม่อยากตอบคำถามบางลักษณะ และยับยั้งการตอบข้อคำถาม ซึ่งเป็นประเด็นที่ไวต่อความรู้สึกของผู้ตอบ (Sensitive topic) ซึ่งสาเหตุเหล่านี้ ล้วนนำไปสู่ปัญหา ข้อมูลสูญหายทั้งสิ้น (Finch, 2008; Howell, 2009; Robitzsch & Rupp, 2009; Schafer & Graham, 2002 อ้างถึงใน กมลทิพย์ ศรีหาเศษ, 2555, หน้า 1)

ปัญหาข้อมูลสูญหายอาจถือว่าเป็นปัญหาที่ไม่รุนแรง ถ้าการวิเคราะห์ข้อมูลจะทำการวิเคราะห์ด้วยสถิติที่วิเคราะห์ข้อมูลตัวแปรเดียว (Univariate data) เช่น ค่าร้อยละ ค่าเฉลี่ย หรือสถิติพรรณนาตัวอื่น แต่ถ้าการวิเคราะห์ข้อมูลนั้นจำเป็นต้องใช้วิธีวิเคราะห์ข้อมูลหลายตัวแปร (Multivariate data) เช่น การวิเคราะห์การถดถอยพหุคูณ (Multiple regression analysis) การวิเคราะห์เส้นทาง (Path analysis) การวิเคราะห์ปัจจัย (Factor analysis) การวิเคราะห์จัดกลุ่ม (Cluster analysis) การวิเคราะห์ การจำแนกกลุ่ม (Discriminant analysis) และการถดถอยโลจีสติกส์ (Logistic regression) เป็นต้น ในกรณีนี้ การสูญหายของข้อมูลจะมีผลกระทบที่รุนแรง เพราะถ้าพบว่า หน่วยวิเคราะห์ใดมีตัวแปรใดที่ข้อมูลสูญหายไปแม้เพียงตัวแปรเดียวก็จะตัดหน่วย

วิเคราะห์นั้นทิ้งทั้งหน่วย โดยไม่สนใจว่าจะยังมีตัวแปรอื่นอีกมากที่มีข้อมูลครบถ้วนหรือไม่ (Heeringa, 2000, pp. 1-19; Roth, 1994, pp. 537-560) นักวิจัยจึงจำเป็นต้องพิจารณาแนวทางที่เหมาะสม สำหรับใช้จัดการข้อมูลสูญหายในทุก ๆ ครั้งที่พบปัญหานี้ ซึ่งวิธีการที่ใช้สำหรับจัดการกับข้อมูลสูญหายมีทางเลือกให้พิจารณา ค่อนข้างหลากหลาย หากเลือกใช้วิธีการจัดการกับข้อมูลสูญหายที่ไม่เหมาะสมย่อมส่งผลทำให้เกิดการบิดเบือนการวิเคราะห์ วิธีการทางสถิติโดยทั่วไปได้ถูกพัฒนาเพื่อการวิเคราะห์ข้อมูลที่สมบูรณ์ แต่เมื่อมีข้อมูลสูญหายย่อมส่งผล กระทบต่อประสิทธิภาพการวิเคราะห์ข้อมูล จากการศึกษาของ Wood et al. (2004, pp. 368-376) ได้ทำการศึกษาจากผลงานวิจัยที่ได้รับการตีพิมพ์ในวารสาร จำนวน 71 ชิ้น พบว่า มีงานวิจัยถึงร้อยละ 89 ที่มีปัญหาเรื่องข้อมูลสูญหาย และมีเพียงร้อยละ 21 เท่านั้นที่มีการจัดการกับข้อมูลสูญหาย นั่นแสดงให้เห็นว่า การจัดการกับปัญหาข้อมูลสูญหายยังคงถูกละเลยกันอย่างเป็นปกติ (ไพฑูรย์ มุลิวัลย์, 2556, หน้า 1)

ในทฤษฎีการตอบสนองข้อสอบ เมื่อร้อยละของข้อมูลสูญหายเพิ่มขึ้น การพยายามที่จะทำให้แบบจำลองสอดคล้องกับทฤษฎีเป็นไปได้ยากเพิ่มขึ้นและการวินิจฉัยรายบุคคลผิดพลาดเพิ่มขึ้น ผลกระทบของข้อมูลสูญหายที่เด่นชัดที่สุด คือ การแสดงสาเหตุที่ผิดพลาดในการวิเคราะห์เชิงสาเหตุ และยังส่งผลให้แบบจำลองจากการวัดไม่สอดคล้องกับแบบจำลองตามทฤษฎี (Zhang & Walker, 2008, pp. 466-479) ผลกระทบของข้อมูลสูญหายที่มีต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) ได้แก่ ความเอนเอียงของการอธิบายความผันแปรของข้อสอบ ความคลาดเคลื่อนประเภทที่หนึ่ง และความคลาดเคลื่อนประเภทที่สองได้รับผลกระทบโดยตรงและค่อนข้างจะมีผลกระทบที่รุนแรง (Robitszsch & Rupp, 2009, pp. 18-34) และการประมาณพารามิเตอร์ของข้อสอบส่งผลกระทบในทางลบและเพิ่มมากขึ้นเมื่อปริมาณข้อมูลสูญหายเพิ่มขึ้น (Furlow et al., 2007, pp. 388-403) นักวัดผลทางการศึกษา จึงคิดการพัฒนาวิธีการในการจัดการข้อมูลสูญหายหรือนำวิธีการทดแทนข้อมูลสูญหายที่มีประสิทธิภาพ มาประยุกต์ใช้สำหรับการประมาณค่าองค์ประกอบ การอนุมานทางสถิติ และการแปลผลการวิเคราะห์ข้อมูลเมื่อข้อมูลที่นำมาวิเคราะห์มีข้อมูลสูญหายเกิดขึ้น (Nitko, 2008, pp. 4-7 อ้างถึงใน จำลอง วงษ์ประเสริฐ, 2554, หน้า 3)

อย่างไรก็ตาม ยังมีประเด็นหรือปัญหาหนึ่งของการศึกษาในบริบทของทฤษฎีการตอบสนองข้อสอบที่ยังไม่ได้รับการศึกษาวิจัยมากนัก นั่นคือ การศึกษาวิธีจัดการข้อมูลสูญหาย และผลกระทบต่อค่าพารามิเตอร์ของข้อสอบจากการจัดการข้อมูลสูญหายและการวิเคราะห์ในระดับทุติยภูมิ ซึ่งปัญหาข้อมูลสูญหายนี้ นับเป็นปัญหาที่ส่งผลกระทบต่อผล

การวิเคราะห์ และความน่าเชื่อถือทางการวิจัย เนื่องจาก หากขนาดของข้อมูลสูญหายเพิ่มมากขึ้น นั่นหมายความว่าขนาดข้อมูลสมบูรณ์ที่ใช้ได้ย่อมน้อยลงตามไปด้วย และยิ่งรวมถึงการทำให้ผลการประมาณค่าพารามิเตอร์แตกต่างหรือคลาดเคลื่อนจากค่าที่ควรจะเป็นด้วย ดังนั้น การศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ จึงเป็นประเด็นที่ควรพิจารณาเป็นอย่างยิ่ง เนื่องจากทฤษฎีการตอบสนองข้อสอบซึ่งมีพื้นฐานความเชื่อว่าพฤติกรรมการตอบสนองข้อสอบของผู้สอบถูกกำหนด โดยคุณลักษณะภายในหรือความสามารถที่อยู่ภายในบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง วิธีหนึ่งในทางปฏิบัติคือการใช้แบบสอบที่วิเคราะห์ผลภายใต้ทฤษฎีการตอบสนองข้อสอบเพื่อประมาณค่าพารามิเตอร์ของข้อสอบหรือความสามารถของผู้สอบ ซึ่งการวัดแต่ละครั้งย่อมมีความคลาดเคลื่อนเกิดขึ้นเสมอ ทั้งนี้โมเดลการวัดของทฤษฎีการตอบสนองข้อสอบซึ่งเป็นระบบความสัมพันธ์ระหว่างตัวแปรอิสระที่ร่วมกันสำหรับทำนายตัวแปรตาม ในการอธิบายสิ่งที่ต้องการวัดนั้น โดยพัฒนาขึ้นภายใต้ข้อตกลงที่ว่าข้อมูลการตอบของผู้สอบสมบูรณ์ เมื่อเกิดข้อมูลสูญหายขึ้นเนื่องจากผู้สอบละเว้นการตอบสนองข้อสอบหรือตอบสนองข้อสอบไม่ครบทุกข้อ ย่อมทำให้ ความคลาดเคลื่อนในการวัดเพิ่มมากยิ่งขึ้น ซึ่งจะส่งผลกระทบโดยตรงต่อการประมาณค่าพารามิเตอร์ของข้อสอบหรือความสามารถของผู้สอบ ดังนั้น เมื่อการสอบมีค่าสูญหายเนื่องจากการตอบสนองข้อสอบของผู้สอบเกิดขึ้น จึงเกิดคำถามว่า จำนวนข้อมูลสมบูรณ์ที่เหลืออยู่นั้น ยังคงมีความเป็นตัวแทนของประชากรและเพียงพอที่จะวิเคราะห์ข้อมูลและให้ผลการประมาณค่าพารามิเตอร์ของข้อสอบหรือความสามารถของผู้สอบที่น่าเชื่อถือ เกิดความยุติธรรม และความถูกต้องแม่นยำได้หรือไม่ และวิธีการที่ใช้จัดการข้อมูลสูญหายวิธีการใด จึงจะทำให้เกิดความยุติธรรม และความถูกต้องแม่นยำของการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบในบริบทของทฤษฎีการตอบสนองข้อสอบได้ดีที่สุด (กมลทิพย์ ศรีหาเศษ, 2555, หน้า 7)

จากการศึกษาของ Mislevy and Wu (1988) เกี่ยวกับการประมาณค่าความสามารถของผู้สอบเมื่อผู้สอบไม่ได้ตอบคำถามทุกรายการและยังพบหลักฐานที่สนับสนุนเรื่องความขัดแย้งเพิ่มเติมที่ว่าเมื่อกลไกของข้อมูลที่หายไปนั้นไม่เกี่ยวข้องหรือมีความสัมพันธ์กับค่าพารามิเตอร์ที่ต้องการตรวจสอบ (ในกรณีนี้คือความสามารถในการทำข้อสอบ) ซึ่งทำให้คุณภาพของการประมาณค่าดังกล่าวได้รับผลกระทบอย่างใหญ่หลวง ในทางกลับกันถ้าผู้วัดผลไม่สามารถตรวจสอบเหตุผลของการตอบข้อสอบไม่ครบทุกข้อ (ซึ่งส่วนใหญ่มักจะเป็นกรณีนี้ Mislevy and Wu (1998) เสนอแนะว่าควรรวมกลไกของข้อมูลที่สูญหายเข้ากับการประมาณค่าพารามิเตอร์

และเพื่อหลีกเลี่ยงผลประเมินที่อาจเกิดความลำเอียง ถ้ากลไกนี้ไม่สามารถทำได้ก็เป็นไปได้มาก ที่การประมาณการพารามิเตอร์ของข้อสอบ และความสามารถของผู้สอบจะมีความลำเอียงอยู่ด้วย

และการศึกษาของ Andreis and Ferrari (2012) ได้ศึกษาเกี่ยวกับข้อมูลสูญหาย และการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ พบว่าข้อมูลที่ได้จากการสำรวจจำนวนมากมีข้อมูลสูญหายและส่งผลกระทบให้เกิดปัญหาในการวิเคราะห์ทางสถิติ ถ้าการดำเนินการทดแทนข้อมูลที่สูญหายไม่เหมาะสมกับค่าที่สูญหาย อาจนำไปสู่การอ้างอิงที่ไม่ถูกต้อง และวิธีการทางสถิติบางวิธีกำหนดให้มีความครบถ้วนสมบูรณ์ของเมทริกซ์ข้อมูล มีวิธีการดำเนินการหลายวิธีปรากฏและนำเสนอขึ้นมาเพื่อจัดการกับปัญหาเหล่านั้น เมื่อเกี่ยวข้องกับตัวแปรเชิงปริมาณ (Little & Rubin, 1987) โดยวิธีบางวิธีนำมาใช้กับข้อมูลจำแนกกลุ่มได้ ในตัวแบบทฤษฎีการตอบสนองแบบพหุมิตินักวิจัยสามารถเลือกจาก 3 ทางเลือก คือ 1) การตัดชุดข้อมูลที่ไม่สมบูรณ์ และพิจารณาเฉพาะค่าที่สังเกตที่ได้มาครบถ้วน (Listwise deletion) 2) ละเว้นค่าสูญหาย และใช้ค่าสังเกตที่มี (Passive approach) หรือ 3) พิจารณาทุกตัวแปรและทุกหน่วยนับ และทดแทนกับค่าที่สูญหาย การตัดข้อมูลทิ้งอาจนำไปสู่การหายไปของข้อมูลสารสนเทศ (Information) ที่เป็นสาระสำคัญ และนำไปสู่ความผิดพลาดของการประมาณค่าพารามิเตอร์ที่ วิธีการที่ 2 วิธี Passive approach เป็นวิธีที่แพร่หลาย ให้การอ้างอิงที่ถูกต้องเมื่อตัวแบบหุ่นเป็นแบบง่าย ๆ เช่น ทฤษฎีการตอบสนองข้อสอบมิติเดียว ในการประมาณค่าพารามิเตอร์ที่เสนอมีค่าสถิติที่มีความพอเพียงทางสถิติเกิดขึ้นเมื่อตัวแปรแบบหุ่นหรือโมเดลมีความซับซ้อนวิธีการที่ 2 นี้ไม่พบว่ามีความเหมาะสม และวิธีการดำเนินการกับข้อมูลตัวที่สูญหาย หรือการทดแทนข้อมูลที่สูญหาย น่าจะเป็นวิธีที่ดีกว่า นอกจากนั้นในบางกรณีหรือบางสถานการณ์กำหนดให้มีข้อมูลครบถ้วนในเมทริกซ์ข้อมูล ดังนั้นจึงจำเป็นต้องกระทำดังที่กล่าว

ซึ่ง Andreis and Ferrari (2012) ได้ศึกษาวิธีการในการเปรียบเทียบวิธีการประมาณค่าทดแทนข้อมูลสูญหาย 4 วิธีดังนี้ คือ 1) Complete data (CD) เป็นวิธีการที่ง่ายที่สุด และเป็นวิธีที่มีชื่อเสียงมากที่สุด เป็นวิธีการที่ตัดทิ้งทั้งแถวในเมทริกซ์ข้อมูล ซึ่งเป็นแถวที่มีข้อมูลสูญหาย ซึ่งข้อจำกัดของข้อมูลที่ตัดไป ยังทำให้มีความเพียงพอในข้อมูลที่จะวิเคราะห์ 2) Forward imputation (FD) เป็นลักษณะที่ตัวแปรหุ่นหรือโมเดลจะกระทำการทดแทน (Imputation) ต่อไปเรื่อย ๆ โดยใช้วิธีการวิเคราะห์ NLPCA (Non linear principle component analysis) และนำผลที่ได้ใส่ลงในตารางข้อมูลจำแนก (Categorical data) 3) Miss forest (MF) แบบค่าสูญหายไม่มีพารามิเตอร์เป็นวิธีการประมาณค่าทดแทนข้อมูลสูญหายที่ใช้ใน Random forest

4) Multivariate imputation by chained equation (MICE) เป็นการประมาณค่าทดแทนแบบพหุคูณ ซึ่งมีการกำหนดการแจกแจงแบบมีเงื่อนไข สำหรับตัวแปรทั้งหลายในชุดข้อมูล

Schmitt, Mandel and Guedj (2015) ได้ศึกษาการเปรียบเทียบวิธีจัดการข้อมูลสูญหาย โดยการศึกษาในครั้งนี้ได้เปรียบเทียบวิธีจัดการข้อมูลสูญหาย 6 วิธี คือ 1) วิธีค่าเฉลี่ย (Mean) 2) วิธีสมาชิกเพื่อนบ้านที่ใกล้ที่สุด (K-nearest neighbors: KNN) 3) วิธีฟุซซีเคมีน (Fuzzy K-means: FKM) 4) วิธีการวิเคราะห์องค์ประกอบพื้นฐานแบบเบย์ (Bayesian principal component analysis: BPCA) 5) การแตกค่าแบบเอกฐาน (Singular value decomposition: SVD) 6) วิธี multiple imputations by chained equations: MICE) โดยผลการศึกษาพบว่า วิธีการวิเคราะห์องค์ประกอบพื้นฐานแบบเบย์ และวิธีฟุซซีเคมีน เป็นวิธีที่น่าสนใจเพราะนำไปสู่การปฏิบัติได้ หรือเป็นไปได้ในทางปฏิบัติ

Finch (2011) ทำการศึกษาการเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่แตกต่างกันของข้อสอบจากการจัดการข้อมูลสูญหายด้วยวิธี Multiple Imputation ทำการจำลองข้อมูลภายใต้โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) และวิธีการจัดการข้อมูลสูญหายที่แตกต่างกัน 4 วิธี คือ 1) ข้อมูลสมบูรณ์ (complete data) 2) การลบตามรายการ (Liswise deletion) 3) การละเว้นการตอบแบบตอบผิด (Omitted as incorrect) 4) การประมาณค่าพหุ (Multiple Imputation) วิธีตรวจสอบการทำหน้าที่แตกต่างกันของข้อสอบ 3 วิธี คือ 1) Mantel-haenszel 2) Logistic regression 3) SIBTEST ความยาวแบบสอบ จำนวน 40 ข้อ ภายใต้เงื่อนไขที่แปรเปลี่ยน 5 ปัจจัย คือ สภาวะการสูญหาย ขนาดกลุ่มตัวอย่าง อัตราการสูญหาย ขนาดการทำหน้าที่ต่างกันของข้อสอบ และระดับความยากของข้อสอบ ผลการศึกษาพบว่า ปัจจัยขนาดกลุ่มตัวอย่าง ระดับผลกระทบ วิธีจัดการข้อมูลสูญหาย และสภาวะการข้อมูลสูญหาย พบว่าเมื่อระดับความยากและ อัตราการสูญหายที่สูงขึ้น วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธี สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ไม่แตกต่างกัน โดยเฉพาะภายใต้สภาวะการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) และภายใต้สภาวะการข้อมูลสูญหายอย่างสุ่ม (MAR)

การวิจัยที่เกี่ยวข้องกับการจัดการกับข้อมูลสูญหายในอดีตที่ผ่านมามีหลายวิธีการให้เลือกใช้ การพิจารณาเลือกใช้วิธีการใดขึ้นอยู่กับลักษณะของข้อมูลสูญหายที่เกิดขึ้น ซึ่งการพิจารณาประเภทของข้อมูลสูญหายเป็นขั้นตอนที่สำคัญ ทั้งนี้เพราะหากสามารถทราบถึงลักษณะของข้อมูลสูญหายจะช่วยในการพิจารณาแนวทางสำหรับจัดการกับปัญหาความไม่สมบูรณ์ของข้อมูลได้อย่างเหมาะสม โดยทั่วไปมักจำแนกข้อมูลสูญหายออกเป็น

3 ประเภทด้วยกัน (Little & Rubin, 2002) ดังนี้ คือ 1) การสูญหายโดยสมบูรณ์อย่างสุ่ม (Missing completely at random: MCAR) เป็นการสูญหายซึ่งอาจสูญหายจากความบกพร่องของเครื่องมือที่ใช้ เช่น ปัญหาเนื่องจากสภาพอากาศ หรือการเจ็บป่วยของกลุ่มตัวอย่าง หรือการนำเข้าข้อมูลไม่ถูกต้อง 2) การสูญหายอย่างสุ่ม (Missing at random: MAR) เป็นลักษณะของข้อมูลสูญหายซึ่งไม่ได้เกิดขึ้นอย่างสุ่มจากค่าสังเกตทั้งหมดภายหลังการควบคุมตัวแปรอื่นแล้ว แต่เกิดขึ้นอย่างสุ่มภายในบางส่วนหรือบางกลุ่มของค่าสังเกต 3) การสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR) เป็นลักษณะของข้อมูลสูญหายซึ่งไม่ได้เกิดขึ้นอย่างสุ่ม โดยค่าของข้อมูลสูญหายขึ้นอยู่กับค่าของข้อมูลสมบูรณ์ในตัวแปรเดียวกัน รวมถึงตัวแปรตัวอื่น

จากประเภทของข้อมูลสูญหายที่กล่าวมาแล้ว หากผู้วิจัยเลือกวิธีการที่ไม่เหมาะสมมาใช้อาจเป็นการเพิ่มค่าความคลาดเคลื่อนและทำลายผลลัพธ์ที่ควรจะได้ โดยวิธีการจัดการข้อมูลสูญหายที่มักถูกเลือกนำมาใช้ในการจัดการกับข้อมูลสูญหายที่ผ่านมามีหลายวิธี แต่ทั้งนี้สามารถแบ่งเป็น 2 กลุ่มคือ กลุ่มวิธีจัดการข้อมูลสูญหายแบบเก่าหรือแบบดั้งเดิมที่ ได้แก่ การตัดข้อมูลสูญหายตามรายการ (Listwise deletion) เป็นวิธีจัดการข้อมูลสูญหายที่ง่ายมากนั่นคือไม่สนใจข้อมูลสูญหายที่เกิดขึ้น โดยจะทำการวิเคราะห์ข้อมูลจากข้อมูลเฉพาะส่วนที่สมบูรณ์ และวิธีการนี้เหมาะกับข้อมูลที่สูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) และกลุ่มตัวอย่างมีขนาดใหญ่ การตัดข้อมูลสูญหายแบบเป็นรายคู่ (Pairwise deletion) เป็นวิธีการจัดการข้อมูลสูญหายสำหรับกรณีที่ทำการวิเคราะห์ระดับความสัมพันธ์ระหว่างตัวแปรคู่ โดยจะทำการวิเคราะห์ข้อมูลจากข้อมูลส่วนที่มีค่าสมบูรณ์ทั้งสองตัวแปร เหมาะกับข้อมูลสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) แต่จะทำให้เกิด Covariance matrix ซึ่งมีความสัมพันธ์ทางบวกไม่ชัดเจน แต่ทำให้เกิดความลำเอียงน้อยสำหรับกรณีข้อมูลสูญหายอย่างสุ่ม (MAR) แต่ความเหมาะสมขององศาอิสระสำหรับทดสอบนัยสำคัญนั้นยังไม่ชัดเจนนัก วิธีการประมาณข้อมูลสูญหายด้วยการแทนที่ด้วยค่าเฉลี่ย (Mean substitution) วิธีการแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยของข้อมูลที่ทราบค่า ซึ่งวิธีการแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยอาจเป็นทางเลือกที่ ไม่เหมาะสมที่สุดเนื่องจากทำให้ความแปรปวนลดลงและทำให้เกิดความไม่แน่นอนของความลำเอียงเมื่อจำนวนค่าสูญหายของตัวแปรที่แตกต่างกันมีความแตกต่างกันสูง วิธีการประมาณข้อมูลสูญหายแบบฮ็อทเดคก์ (Hot-deck imputation) เป็นวิธีการพิจารณาเลือกหน่วยตัวอย่างที่มีลักษณะคล้ายคลึงกันมากที่สุดกับหน่วยตัวอย่างที่เกิดค่าสูญหาย จากนั้นแทนค่าที่สูญหายด้วยค่าของหน่วยตัวอย่างที่คล้ายคลึงนั้น วิธีการประมาณข้อมูลสูญหายด้วยการถดถอย (Regression based method) ทำการสร้างสมการถดถอยระหว่างตัวแปรใด ๆ ที่ต้องการจากข้อมูลที่สมบูรณ์ โดยกำหนด

ให้ตัวแปรตามเป็นตัวแปรที่มีข้อมูลไม่สมบูรณ์ จากนั้นใช้สมการถดถอยที่ได้ทำการประมาณค่าของข้อมูลที่ไม่สมบูรณ์ แม้ว่าวิธีจัดการข้อมูลสูญหายแบบดั้งเดิมในอดีตจะนิยมใช้กันทั่วไป แต่ไม่มีวิธีจัดการข้อมูลสูญหายแบบดั้งเดิมวิธีใดที่ระบุว่าสามารถใช้แก้ปัญหาข้อมูลสูญหายได้เหมาะสม ยกเว้นภายใต้สถานการณ์ที่มีลักษณะเฉพาะ ซึ่งวิธีการเหล่านี้อาจให้ผลที่ลำเอียงทั้งในทางบวกและทางลบ การเพิ่ม Type II error และการประมาณค่าสหสัมพันธ์และน้ำหนักสัมประสิทธิ์การทำนายต่ำ (Acock, 2005; Baraldi & Enders, 2010; Peugh & Enders, 2004 อ้างถึงใน กมลทิพย์ ศรีหาเศษ, 2555; จำลอง วงษ์ประเสริฐ, 2554)

และกลุ่มวิธีการจัดการข้อมูลสูญหายกลุ่มที่สอง เป็นวิธีจัดการข้อมูลสูญหายแบบใหม่ เช่น วิธีค่าคาดหมายสูงสุด (Expectation maximization approach: EM) การประมาณค่าทดแทนด้วยการแทนที่ค่าเฉลี่ยข้อที่ถูก (Corrected item mean substitution imputation: CM) การประมาณค่าทดแทนด้วยฟังก์ชันการตอบสนอง (Response function imputation: RF) การประมาณค่าทดแทนพหุ (Multiple imputation: MI) วิธีการประมาณค่าที่เป็นไปได้สูงสุด (Maximum likelihood methods) และวิธีการที่ไม่อิงการประมาณค่าทดแทน (Nonimputation-based approaches) โดยในช่วงแรกของการศึกษาเกี่ยวกับข้อมูลสูญหาย วิธีจัดการอย่างง่ายส่วนใหญ่ที่ใช้ในการจัดการข้อมูลสูญหายคือ การคัดข้อมูลที่มีค่าสูญหายทิ้งแล้ววิเคราะห์เฉพาะกรณีที่ข้อมูลสมบูรณ์ ซึ่งก่อให้เกิดความผิดพลาดเกี่ยวกับความถูกต้องแม่นยำสำหรับการประมาณค่าพารามิเตอร์ในการวิเคราะห์ตัวแปรศึกษาอย่างมีนัยสำคัญ (Robitzsch & Rupp, 2009) นอกจากนี้ ผู้วิจัยมักนำเสนอเฉพาะการประมาณค่าพารามิเตอร์ที่ไม่มีความลำเอียงเท่านั้น ซึ่งสำหรับข้อมูลสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) นั้น แม้ว่าจะใช้การวิเคราะห์ความถี่ แต่ผู้วิจัยที่ต้องการใช้วิธีอย่างง่าย มักถูกกระตุ้นด้วยความไม่พอเพียงของความรู้ในเรื่องวิธีการหรือขาดการศึกษารายละเอียดหรือเงื่อนไขของวิธีการที่ใช้ จึงทำให้การจัดการข้อมูลสูญหายเป็นการจัดการที่ขาดการพิจารณาลักษณะหรือธรรมชาติของข้อมูลสูญหาย ต่อมาได้มีการพัฒนาวิธีจัดการกับข้อมูลสูญหาย โดยเฉพาะการมุ่งศึกษาในประเด็นผลของการใช้วิธีจัดการกับข้อมูลสูญหายที่ให้ผลความถูกต้องแม่นยำของการประมาณค่าพารามิเตอร์ของข้อสอบหรือความสามารถของผู้สอบสูงที่สุด

วิธีการแก้ค่าพารามิเตอร์ทางสถิติที่ดีที่สุดสำหรับการประมาณค่าทดแทนในหลายสถานการณ์นิยมใช้วิธีการประมาณค่าที่เป็นไปได้สูงสุด (Maximum likelihood estimation: ML) และการประมาณค่าทดแทนพหุ (Multiple imputation: MI) อย่างไรก็ตามในการวิเคราะห์

บางครั้งจะใช้กรอบการวิเคราะห์ Bayesian (Rupp, Dey, & Zumbo, 2004 cited in Robitzsch & Rupp, 2009) ในวิธีการประมาณค่าที่เป็นไปได้สูงสุดนั้น การประมาณค่าพารามิเตอร์จะถูกคำนวณโดยการรวมค่าข้อมูลสูญหายออกจาก likelihood ซึ่งมีคุณสมบัติบางอย่างเหมือนกันในทางทฤษฎีสำหรับค่าที่เป็นไปได้ทั้งหมดที่ข้อมูลสูญหายจะมีผล แนวคิดนี้ให้ผลที่เชื่อถือได้ว่าข้อตกลงเบื้องต้นของการสูญหายอย่างสุ่มเหมาะสม แต่ข้อบกพร่องของแนวคิดนี้คือเป็นโมเดลที่มีความเฉพาะเจาะจง เช่น Likelihood เป็นข้อจำกัดของโมเดล (Sinharay, Stern, & Russell, 2001) ในกรณีข้อมูลที่มีการตอบแบบ Dichotomous ในทฤษฎีการตอบสนองข้อสอบ เช่น Rasch model/ one-parameter logistic model (1PL) และ Two-parameter logistic model (2PL) หรือ Three-parameter logistic model (3PL) (Embretson & Reise, 2000) เป็นตัวอย่างของการใช้โดยอิงแนวคิดการประมาณค่าที่เป็นไปได้สูงสุด นอกจากการพัฒนาวิธีการหรือแนวทางการจัดการข้อมูลสูญหาย สำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบแล้ว ยังมีการพัฒนาโปรแกรมสำเร็จรูปที่อำนวยความสะดวกในการประมาณค่าแต่ละวิธีด้วย เช่น โปรแกรมสำหรับการประมาณค่าทดแทนพหุ คือ Amella, SAS/ IML Multiple imputation programs, Paul Allison's SAS macro, SIRNORM SAS macro และ SOLAS เป็นต้น (University of Texas, 2004) โดยตัวอย่างโปรแกรม NORM ใช้ Multivariate normal model (MICE) (Van Buuren, Boshuizen, & Knook, 1999) ซึ่งไม่มีข้อตกลงของการแจกแจงที่เฉพาะเจาะจงในการสร้าง แต่ทั้งนี้ หากวิธีการที่ใช้ในการจัดการข้อมูลสูญหายมีความซับซ้อนยุ่งยากมากขึ้นเท่าไร โปรแกรมสำเร็จรูปที่ใช้ ในการวิเคราะห์ย่อมมีความยุ่งยากในการใช้ตามไปด้วยเช่นกัน (กมลทิพย์ ศรีหาเศษ, 2555, หน้า 84)

จากการศึกษาเกี่ยวกับงานวิจัยที่ผ่านมาทั้งในอดีตและปัจจุบัน ทั้งทางการศึกษา ทางการแพทย์ สังคมศาสตร์ และวิทยาศาสตร์ ล้วนต่างพยายามค้นหาวิธีในการจัดการกับข้อมูลสูญหายทั้งสิ้น โดยวิธีที่นิยมใช้มีทั้งวิธีประมาณค่าทดแทนข้อมูลสูญหาย ทั้งแบบดั้งเดิมและแบบใหม่ โดยทั้ง 2 แบบ ล้วนเป็นวิธีการเพื่อจัดการกับข้อมูลที่มีการสูญหาย ให้เป็นข้อมูลที่มีความสมบูรณ์มากที่สุด เพื่อนำไปสู่การวิเคราะห์ข้อมูลให้ได้ผลการวิเคราะห์ที่ถูกต้องมากที่สุด ดังนั้นการศึกษาในครั้งนี้ ผู้วิจัยจึงมุ่งศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) เพื่อแก้ไขปัญหาความคลาดเคลื่อนในการ ประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ในกรณีที่มีข้อมูลสูญหายเนื่องจากผู้ทำการสอบละเว้นจากการตอบข้อสอบ หรือตอบข้อสอบไม่ครบทุกข้อ ซึ่งสาเหตุอาจมาจากผู้ทำการสอบไม่ตั้งใจ

ละเว้นข้อสอบ ผู้สอบอาจไม่ทราบคำตอบจึงทำให้เกิดละเว้นการตอบสนองข้อสอบ หรือผู้สอบอาจจะตั้งใจละเว้นการตอบข้อสอบ ถึงแม้ว่าผู้สอบจะสามารถตอบข้อสอบข้อนั้นได้ก็ตาม โดยในการวิเคราะห์ข้อมูลใช้กระบวนการจำลองสถานการณ์ ภายใต้เงื่อนไขที่แปรเปลี่ยนต่าง ๆ เพื่อให้ครอบคลุมเงื่อนไขทั้งหมดที่เป็นไปได้ และได้ข้อค้นพบที่ครอบคลุมในทุกกรณีตัวอย่างด้วยเทคนิคมอนติคาร์โล (Monte carlo simulation) ซึ่งเป็นวิธีการหนึ่งที่ใช้ในกระบวนการออกแบบจำลอง โดยในการจำลองข้อมูลเพื่อการศึกษาครั้งนี้ ผู้วิจัยจำลองข้อมูลครอบคลุมเงื่อนไขภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัยหลัก คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่างอัตราการสูญหายของข้อมูล ความยาวแบบสอบที่ต่างกัน และขนาดอิทธิพล DIF เพื่อให้เกิดความยุติธรรมและความถูกต้องแม่นยำของการประมาณค่าพารามิเตอร์ของข้อสอบ และความสามารถของผู้สอบในบริบทของทฤษฎีการตอบสนองข้อสอบมากที่สุด ซึ่งเป็นการขยายองค์ความรู้ในการจัดการข้อมูลสูญหาย และเป็นประโยชน์ต่อวงการศึกษาของประเทศไทยทั้งในปัจจุบันและอนาคต

คำถามการวิจัย

1. วิธีจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (MI) ภายใต้สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวแบบสอบ และขนาดอิทธิพล DIF จะให้ผลการประมาณค่าพารามิเตอร์ของข้อสอบ ความสามารถของผู้สอบ และความลำเอียงในการประมาณค่าแตกต่างกันอย่างไร

2. อัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนสองค่าด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) ภายใต้สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวแบบสอบ และขนาดอิทธิพล DIF มีความแตกต่างกันหรือไม่

3. อำนาจการทดสอบของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนสองค่าด้วยวิธีวิธีทดสอบอัตราส่วนความควรจะเป็น ภายใต้สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวแบบสอบ และขนาดอิทธิพล DIF มีความแตกต่างกันหรือไม่

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ด้วยวิธีจัดการข้อมูลสูญหายโดยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF

2. เพื่อเปรียบเทียบประสิทธิภาพของวิธีจัดการข้อมูลสูญหาย สำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ โดยวิธีจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF

3. เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนสองค่าด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT) จากการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF

4. เพื่อเปรียบเทียบอำนาจการทดสอบ (Power of test) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT) จากการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF

ขอบเขตของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบ และความสามารถของผู้สอบ ภายใต้ปัจจัยที่แตกต่างกัน รายละเอียดดังต่อไปนี้

1. การศึกษาครั้งนี้ใช้โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) นั่นคือทำการศึกษาพารามิเตอร์ข้อสอบ 2 พารามิเตอร์ ได้แก่ ความยาก และอำนาจจำแนก และพารามิเตอร์ผู้สอบ ได้แก่ ความสามารถผู้สอบ

2. ข้อมูลที่ใช้ในการศึกษาในครั้งนี้ ใช้ข้อมูลจำลองที่ได้จากโปรแกรม R โดยเป็นข้อมูลที่มีวิธีการให้คะแนนรายข้อแบบสองค่า (Dichotomous scoring) ซึ่งทำการจำลองภายใต้เงื่อนไข จำนวน 32 เงื่อนไข (2x2x2x2x2) ในแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 100 ครั้ง จำนวนการทำซ้ำภายใต้เงื่อนไขที่แปรเปลี่ยนทั้งหมด 3,200 ครั้ง โดยมีรายละเอียดของเงื่อนไข ดังนี้

2.1 สภาวะข้อมูลสูญหาย มี 2 ประเภท ดังนี้

2.1.1 การสูญหายอย่างสุ่ม (Missing random: MAR)

2.1.2 การสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not random: MNAR)

2.2 ขนาดของกลุ่มตัวอย่าง ประกอบด้วย กลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีสัดส่วนจำนวนผู้สอบในแต่ละกลุ่ม เป็นสัดส่วน 1:1 ดังนี้

2.2.1 จำนวน 500: 500 คน

2.2.2 จำนวน 1,500: 1,500 คน

2.3 อัตราการสูญหายของข้อมูล มี 2 ระดับ ดังนี้

2.3.1 อัตราการสูญหาย 10%

2.3.2 อัตราการสูญหาย 30%

2.4 ความยาวของแบบสอบ มี 2 ระดับ โดยที่ระดับความยาวแบบสอบต่างกันมีการกำหนดจำนวนข้อที่ทำหน้าที่ต่างกันของข้อสอบ ดังนี้

2.4.1 แบบสอบจำนวน 20 ข้อ กำหนดข้อที่ 1 เป็นข้อสอบที่ทำหน้าที่ต่างกัน

2.4.2 แบบสอบจำนวน 40 ข้อ กำหนดข้อที่ 1 และ 2 เป็นข้อสอบที่ทำหน้าที่ต่างกัน

2.5 ขนาดอิทธิพลของ DIF มี 2 ขนาด โดยค่าความยากเป็นค่าพารามิเตอร์ที่มีอิทธิพลของ DIF ดังนี้

2.5.1 ขนาดอิทธิพล = 0.5

2.5.2 ขนาดอิทธิพล = 1.0

3. ในการศึกษาในครั้งนี้ ทำการศึกษาภายใต้วิธีการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI)

4. ในการศึกษาในครั้งนี้ ทำการศึกษาภายใต้วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด (Maximum likelihood: ML)

5. ในการศึกษาในครั้งนี้ ทำการศึกษาภายใต้วิธีการตรวจสบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT)

6. ในการศึกษาครั้งนี้ พิจารณาการเปรียบเทียบประสิทธิภาพ 2 ส่วน ได้แก่

6.1 การวัดประสิทธิภาพการประมาณค่าพารามิเตอร์ พิจารณาจากดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย และค่าที่ประมาณได้ (diff error) และค่าดัชนีผลต่าง (diff)

6.2 การวัดประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพิจารณาจาก

6.2.1 อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

6.2.2 อำนาจการทดสอบ (Power of test) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

7. ทำการศึกษาภายใต้การจำลองข้อมูล เพื่อศึกษาการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ซึ่งในการจำลองข้อมูล แต่ละเงื่อนไขในการศึกษาที่กำหนดผู้วิจัยทำซ้ำ จำนวน 100 ครั้ง (Replication)

8. แบบสอบที่ใช้ในการศึกษาครั้งนี้ เป็นแบบสอบที่มีการตรวจให้คะแนนแบบ 2 ค่า (Dichotomous) คือ ผู้สอบได้ 1 คะแนนเมื่อตอบข้อสอบถูก และได้ 0 คะแนนเมื่อตอบข้อสอบผิด โดยไม่มีการลงโทษเมื่อผู้สอบตอบข้อสอบผิดหรือละเว้นการตอบสนองข้อสอบ นั่นคือ ไม่มีการให้คะแนนติดลบกรณีผู้สอบตอบข้อสอบผิดหรือละเว้นการตอบสนองข้อสอบ

นิยามศัพท์เฉพาะ

ข้อมูลสูญหาย (Missing data) หมายถึง การไม่ปรากฏค่าสังเกตได้ของการตอบข้อสอบ ข้อมูลสูญหายที่เกิดขึ้นเป็นตัวบ่งบอกถึงการไม่ตอบข้อสอบของผู้สอบ (Lack of response) อาจจะเกิดจากหลายสาเหตุ เช่น “ไม่ทราบ” “ไม่ตอบ/ ปฏิเสธที่จะตอบ” หรือ “ไม่มีความรู้เพียงพอ” เป็นต้น ในการศึกษาครั้งนี้ ข้อมูลสูญหายเกิดจากการที่ผู้สอบละเว้นการตอบสนองข้อสอบ (Omitted) ซึ่งอาจมีลักษณะการสูญหายอย่างสุ่ม (Missing at random: MAR) หรือการสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR)

การสูญหายอย่างสุ่ม (Missing at random: MAR) หมายถึง การขาดหายไปของข้อมูล โดยที่ความน่าจะเป็นของการสูญหายเป็นผลจากค่าตัวแปรสังเกตได้อื่น ๆ เท่านั้น ซึ่งหมายความว่า กระบวนการสูญหายอาจสัมพันธ์กับการสูญหายโดยสมบูรณ์อย่างสุ่มจากเงื่อนไขของตัวแปรร่วมที่สังเกตได้ (Observable covariates) เช่น การละเว้นหรือข้ามข้อสอบของผู้สอบไม่สัมพันธ์กับระดับความสามารถของผู้สอบ นั่นคือ ผู้สอบละเว้นหรือข้ามข้อสอบ

ของผู้สอบด้วยความตั้งใจแม้ว่าจะสามารถทำข้อสอบข้อนั้นได้ แต่เป็นผลมาจากการที่ผู้สอบมีระดับแรงจูงใจต่ำ ผู้สอบจึงละเว้นหรือข้ามข้อสอบ เป็นต้น ซึ่งสามารถเขียนสมการได้ ดังนี้

$$P(M|Y_{com}) = P(M|Y_{obs}, Y_{mis}) = P(M|Y_{obs})$$

ในการศึกษาครั้งนี้ กำหนดให้ข้อมูลที่มี**การ**สูญหายอย่างสุ่ม เป็นผลมาจากตัวแปรแรงจูงใจในการสอบของผู้สอบ เช่น ผู้สอบที่มีระดับแรงจูงใจในการสอบต่ำ มีอัตราการละเว้นการตอบสนองข้อสอบสูงกว่าผู้สอบที่มีระดับแรงจูงใจในการสอบสูง

การสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR) **หมายถึง** การขาดหายไปของข้อมูลโดยที่ผลของตัวแปรอิสระที่สัมพันธ์กับการสูญหาย ไม่สามารถอธิบายการสูญหายด้วยตัวแปรสังเกตได้ เพราะการสูญหายเป็นผลที่เกิดขึ้นแน่นอนจากค่าที่ไม่ได้สังเกต (Ymis) เช่น การที่ผู้สอบที่มีระดับความสามารถต่ำ ละเว้นหรือข้ามข้อสอบข้อที่ตนไม่ทราบคำตอบหรือมีความรู้ไม่เพียงพอที่จะตอบข้อสอบข้อนั้นได้ถูกต้อง เป็นต้น ซึ่งเขียนสมการได้ ดังนี้

$$P(M|Y_{com}) = P(M|Y_{obs}, Y_{mis}) \neq P(M|Y_{obs})$$

ในการศึกษาครั้งนี้ ศึกษาการสูญหายที่ไม่ใช่อย่างสุ่ม โดยเพิ่มเงื่อนไขสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) เข้าไปในเงื่อนไขสภาวะการสูญหายอย่างสุ่ม (MAR) โดยที่ความน่าจะเป็นของค่าสูญหายที่เกิดขึ้นยังคงมีความสัมพันธ์กับตัวแปรของมันเองภายใต้การควบคุมตัวแปรสังเกตได้อื่นแล้ว นั่นคือ ข้อมูลที่มีการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) เป็นผลมาจากการตอบถูกหรือตอบผิดในกลุ่มผู้สอบที่มีระดับแรงจูงใจในการสอบต่ำเท่านั้น โดยผู้ตอบผิดในกลุ่มนี้มีอัตราการละเว้นการตอบสนองข้อสอบสูงกว่าผู้สอบในกลุ่มนี้ที่ตอบถูก หรือกล่าวอีกนัยหนึ่งคือ ผู้สอบที่มีระดับแรงจูงใจในการสอบต่ำ และมีระดับความรู้ความสามารถต่ำมีอัตราการละเว้น การตอบสนองข้อสอบสูงกว่าผู้สอบที่มีระดับแรงจูงใจในการสอบต่ำแต่มีระดับความรู้ความสามารถสูง

วิธีจัดการข้อมูลสูญหาย หมายถึง แนวทาง หรือเทคนิควิธีในการจัดการข้อมูลที่มีค่าสูญหายเกิดขึ้นให้มีความถูกต้องในการสรุปอ้างอิงเกี่ยวกับประชากรที่ศึกษา ซึ่งในการศึกษาครั้งนี้ได้แก่ วิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI)

การประมาณค่าทดแทนพหุ (Multiple imputation: MI) หมายถึง วิธีจัดการข้อมูลสูญหายที่ใช้วิธี Markov chain monte carlo (MCMC) ในการเพิ่มข้อมูลไปยังการแจกแจงเบื้องต้นจากข้อมูลที่เพิ่มขึ้นจากการสร้างค่าประมาณทดแทนกระบวนการประมาณค่าทดแทนนี้

ประมาณค่าทดแทนซ้ำ M ครั้ง (ชุดข้อมูลย่อย) เพื่อสร้างชุดข้อมูลตัวแปรอิสระ (Schafer & Olsen, 1998) แต่ละชุดข้อมูลจะกลายเป็นกรณีในการวิเคราะห์สิ่งที่สนใจศึกษาต่อไป เช่น การประมาณค่าพารามิเตอร์ทฤษฎีการตอบสนองข้อสอบ (IRT) ผลของการวิเคราะห์แยก M ในภายหลังจะรวมเป็นค่าเดียวในการศึกษาครั้งนี้

วิธีการประมาณค่าพารามิเตอร์ หมายถึง วิธีที่ใช้ในการคำนวณหาค่าพารามิเตอร์ความยากของข้อสอบและความสามารถของผู้สอบ โดยใช้ค่าสถิติจากข้อมูลที่จำลอง โดยใช้วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮู้ด (Maximum likelihood: ML)

วิธีแมกซิมัมไลค์ลิฮู้ด (Maximum likelihood: ML) หมายถึง การประมาณค่าพารามิเตอร์ โดยอาศัยผลที่ได้จากตัวอย่างที่สุ่มเลือกมาจากการแจกแจงที่ทราบรูปแบบของฟังก์ชัน ความหนาแน่นแต่ไม่ทราบค่าพารามิเตอร์ ดังนั้น จึงใช้หลักของความน่าจะเป็นในการเลือกตัวอย่าง และวัดค่าได้จากกลุ่มตัวอย่างที่ถูก มาพิจารณาค่าประมาณของค่าพารามิเตอร์ที่ต้องการ

คุณภาพของวิธีจัดการข้อมูลสูญหาย หมายถึง ความถูกต้องแม่นยำของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ซึ่งดัชนีที่นำมาใช้ในการเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหาย ในครั้งนี้ ประกอบด้วย ดัชนีค่าความลำเอียง (BIAS) และดัชนีค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) โดยเปรียบเทียบค่าที่ได้ ถ้าเงื่อนไขที่ศึกษาใดมีค่าที่น้อยกว่าแสดงว่ามีคุณภาพในการจัดการข้อมูลสูญหายได้ดีกว่า

ดัชนีผลต่าง หมายถึง ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ของผู้สอบและความสามารถของผู้สอบ โดยใช้ค่าความแตกต่างระหว่างค่าที่กำหนดจากการจำลองกับค่าประมาณที่ได้จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI) ซึ่งเป็นการแสดงให้เห็นว่าค่าที่ได้จากการประมาณค่าทดแทนข้อมูลสูญหายมีผลต่างจากค่าที่กำหนดจากการจำลองโดยเฉลี่ยมากน้อยเพียงใด ซึ่งดัชนีผลต่าง แสดงถึงความถูกต้อง (Accuracy) ของการประมาณค่า โดยใช้เกณฑ์ในการพิจารณาคือ ถ้าดัชนีผลต่างมีค่าใกล้ศูนย์มากกว่าแสดงให้เห็นว่า การประมาณค่าใกล้เคียงกับค่าที่กำหนดจากการทำการจำลอง ดังสมการ

$$\text{diff} = \frac{\sum(A-\widehat{A})}{n}$$

เมื่อ

diff แทน ดัชนีผลต่าง

A แทน ค่าเฉลี่ยของค่าพารามิเตอร์ข้อสอบ และค่าความสามารถของผู้สอบที่ได้จากการจำลองข้อมูล

$\widehat{A}$ แทน ค่าประมาณที่ได้จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI)

n แทน จำนวนครั้งที่ทำซ้ำในการจำลองข้อมูล

ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย

ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย หมายถึง ความเบี่ยงเบนของค่าที่กำหนดจากการจำลองข้อมูล และค่าที่ประมาณได้ ซึ่งเป็นค่าที่สะท้องถึงความคงเส้นคงวา (Consistency) ของค่าพารามิเตอร์ข้อสอบและความสามารถผู้สอบที่ประมาณได้กับข้อมูลที่สมบูรณ์ โดยมีเกณฑ์ในการพิจารณาคือ ถ้าค่า diff error น้อยกว่า แสดงว่ามีความความคงเส้นคงวา ของการประมาณค่าสูงกว่าสะท้อนให้เห็นถึงความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์มีน้อย แต่ถ้าค่า diff error มากกว่า แสดงว่ามีความความคงเส้นคงวาของการประมาณค่าพารามิเตอร์ต่ำกว่าสะท้อนให้เห็นถึงความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์มีมาก ดังสมการ

$$\text{diff error} = \sqrt{\frac{\sum\left(A-\widehat{A}\right)^2}{n}}$$

เมื่อ

diff error แทน ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย

A แทน ค่าเฉลี่ยของค่าพารามิเตอร์ข้อสอบ และค่าความสามารถของผู้สอบที่ได้จากการจำลองข้อมูล

$\hat{A}$ แทน ค่าประมาณที่ได้จากการจัดการข้อมูลสูญหาย
ด้วยวิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI)

n แทน จำนวนครั้งที่ทำซ้ำในการจำลองข้อมูล

ความเอนเอียง หมายถึง คุณลักษณะของตัวประมาณค่าที่มีแนวโน้มในการให้ค่าประมาณเข้าใกล้ค่าพารามิเตอร์ โดยการพิจารณาจากค่าดัชนีผลต่าง (diff) และดัชนีค่ารากที่สองของผลต่างระหว่างค่าเฉลี่ยจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย (diff error)

ความคงเส้นคงวา หมายถึง คุณลักษณะของตัวประมาณค่าที่มีแนวโน้มในการให้ค่าประมาณเข้าใกล้ค่าพารามิเตอร์เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น (An estimator tends to get closer to the value of the parameter as the sample size becomes larger) หรือเมื่อขนาดกลุ่มตัวอย่างลู่เข้าสู่อนันต์ ($N \rightarrow \infty$)

ความพอเพียง หมายถึง คุณสมบัติของตัวประมาณค่าที่สามารถใช้ข้อมูลจากทุกหน่วยตัวอย่างที่มีในการประมาณค่าพารามิเตอร์ในประชากร แต่เนื่องจากการวิจัยในครั้งนี้ผู้วิจัยใช้ข้อมูลที่ได้จากการจำลองสถานการณ์ จึงสามารถประมาณค่าพารามิเตอร์ได้จากข้อมูลทั้งหมดที่มี คุณสมบัติความพอเพียงจึงไม่นำมาพิจารณา

โมเดลการตอบสนองข้อสอบ หมายถึง ระบบที่แสดงความสัมพันธ์ระหว่างโอกาสของการตอบข้อสอบถูก (P_i) กับความสามารถที่มีอยู่ภายในของผู้สอบ (θ) ในรูปโค้งลักษณะข้อสอบ (ICC) ซึ่งมีลักษณะเป็นฟังก์ชันโลจิสติก โดยความสัมพันธ์ดังกล่าวอธิบายด้วยลักษณะของข้อสอบอันประกอบด้วยค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดาข้อสอบถูก (C)

ค่าความสามารถที่แท้จริงของผู้สอบ (θ) หมายถึง ระดับความสามารถของผู้สอบแต่ละคนที่ประมาณค่าจากการตอบข้อสอบ ซึ่งมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0 และ 1 ตามลำดับ ค่าความสามารถที่แท้จริงมีค่าอยู่ระหว่าง $-\infty$ ถึง ∞

ค่าพารามิเตอร์ของข้อสอบ (β) หมายถึง ค่าที่บ่งบอกถึงลักษณะของข้อสอบซึ่งประกอบด้วยค่าความยากง่าย (b_i) ค่าอำนาจจำแนก (a_i) ของข้อสอบ

ค่าความยากของข้อสอบ (b_i) หมายถึง ค่าพารามิเตอร์ของข้อสอบที่อยู่ตำแหน่งบนโค้งลักษณะข้อสอบซึ่งอยู่บนมาตรความสามารถซึ่งทำให้โอกาสของการตอบข้อสอบถูกเท่ากับ $\frac{1+c_i}{2}$ โดยข้อสอบที่ยากจะมีค่าความยากของข้อสอบไปทางขวา จุดเปลี่ยนโค้งลักษณะข้อสอบเป็นจุดที่ค่า $\theta = b_i$ สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 และ 2 พารามิเตอร์ ค่าความยากของข้อสอบเป็นระดับความสามารถของผู้สอบที่มีความน่าจะเป็นในการตอบข้อสอบถูกเท่ากับ 0.5 แต่ถ้ามีการเดาแล้ว ค่าความยากของข้อสอบจะเป็นระดับความสามารถของผู้สอบที่มีค่า

ความน่าจะเป็นในการตอบข้อสอบถูกมากกว่าหรือเท่ากับค่า c แต่มีค่าน้อยกว่าหรือเท่ากับ 1.0 ค่าความยากง่ายที่นิยมใช้มีค่าอยู่ระหว่าง -2.50 ถึง +2.50 แต่ในทางทฤษฎีจะมีค่าอยู่ระหว่าง $(-\infty, +\infty)$

ค่าอำนาจจำแนกของข้อสอบ (a_i) หมายถึง ค่าพารามิเตอร์ของข้อสอบที่เป็นสัดส่วนของความชันของโค้งลักษณะข้อสอบที่ตำแหน่ง b_i ระหว่างผู้ที่มีความสามารถน้อยกว่าหรือเท่ากับค่า θ กับมีค่ามากกว่าค่า θ ณ จุดเปลี่ยนโค้งซึ่งจำแนกค่าความน่าจะเป็นในการตอบข้อสอบถูก ค่าอำนาจจำแนกของข้อสอบที่สูงแสดงถึงความสามารถในการจำแนกผู้สอบที่มีความสามารถสูงออกจากผู้สอบที่มีความสามารถต่ำได้ดี ค่าอำนาจจำแนกของข้อสอบนิยมใช้อยู่ในช่วง 0.50 ถึง +2.50 แต่ในทางทฤษฎีมีค่าอยู่ระหว่าง $(-\infty, +\infty)$

โค้งลักษณะข้อสอบ หมายถึง กราฟของฟังก์ชันการตอบสนองข้อสอบที่แสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับความน่าจะเป็นในการตอบข้อสอบข้อหนึ่ง ๆ ได้ถูกต้องของผู้สอบ

การทำหน้าที่ต่างกันของข้อสอบ (Differential item function: DIF) หมายถึง ผู้สอบจากกลุ่มที่แตกต่างกัน และมีการจับคู่ความสามารถตามที่ข้อสอบต้องการวัดเท่ากัน มีความน่าจะเป็นในการตอบข้อสอบที่ถูกต้องไม่เท่ากัน

ประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความถูกต้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จากการตรวจสอบด้วยวิธีแม็กซิมั่มไลค์ลิฮูด (Maximum likelihood: ML) ซึ่งพิจารณาได้จากอัตราความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ

กลุ่มอ้างอิง (Reference group: R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้เปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่ต่างกัน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่ากลุ่มสนใจ

กลุ่มเปรียบเทียบ (Focal group: F) หมายถึง กลุ่มผู้สอบที่เป็นเป้าหมายของการศึกษาซึ่งคาดว่าจะเสียเปรียบในการตอบข้อสอบเมื่อข้อสอบทำหน้าที่ต่างกัน โดยมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องน้อยกว่า

ข้อสอบที่ทำหน้าที่ต่างกันแบบรูปคงที่ (Uniform or inconsistent DIF) หมายถึง ข้อสอบที่ให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอกันในทุกระดับความสามารถ เมื่อพิจารณาโค้งลักษณะข้อสอบของผู้สอบทั้งสองกลุ่มจะพบว่าไม่มีปฏิสัมพันธ์กันระหว่างโค้งลักษณะข้อสอบในทุก ๆ ระดับความสามารถ

ข้อสอบที่ทำหน้าที่ต่างกันแบบรูปไม่คงที่ (Nonuniform or inconsistent DIF) หมายถึง ข้อสอบที่ให้โอกาสการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มไม่เท่ากันในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบพบว่าในบางระดับความสามารถจะมีปฏิสัมพันธ์ระหว่างโค้งลักษณะข้อสอบ

ความยาวของแบบสอบ หมายถึง จำนวนข้อสอบในแบบทดสอบ ในการวิจัยครั้งนี้ศึกษาความยาวของแบบทดสอบ 2 แบบ คือ 20 ข้อ และ 40 ข้อ

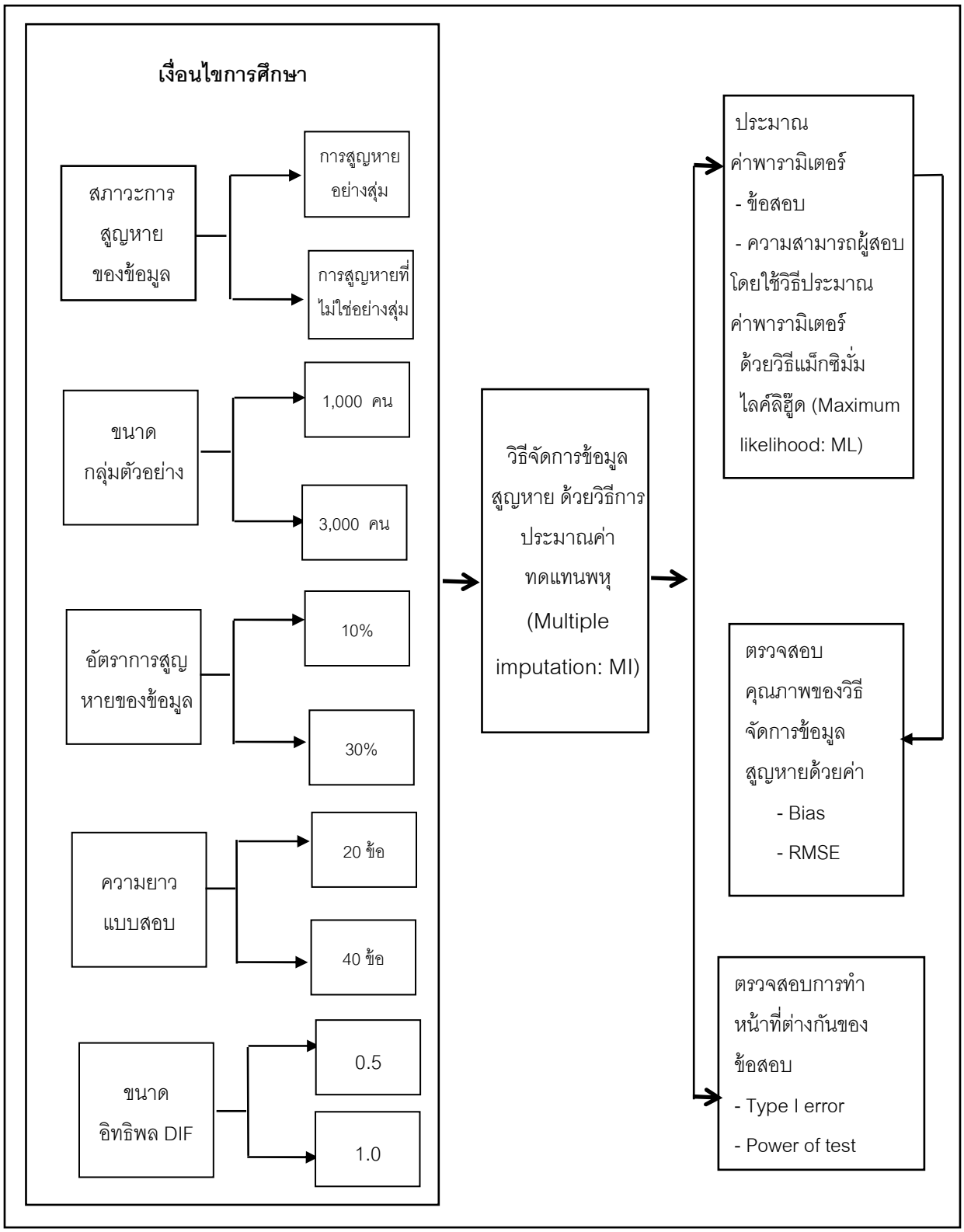
อำนาจการทดสอบ (Power of test) หมายถึง ความน่าจะเป็นที่จะปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นเป็นเท็จจะมีค่าเท่ากับ (1-β) กล่าวคือ ความน่าจะเป็นที่พบข้อสอบที่ตรวจสอบได้ถูกต้องว่าข้อสอบนั้นทำหน้าที่ต่างกันของข้อสอบจริง

อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) หมายถึง ความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐานว่าง โดยที่สมมติฐานว่างเป็นจริง (α) กล่าวคือ ความผิดพลาดที่พบข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบนั้นทำหน้าที่ไม่ต่างกันต่อจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ

กรอบแนวคิดในการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายเพื่อนำมาใช้ในการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบที่วัดความสามารถมิติเดียวและให้คะแนนสองค่า และการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI โดยมีเงื่อนไขในการศึกษาคือ สภาวะการสูญหายของข้อมูลที่ครอบคลุม การสูญหาย 2 ประเภท คือ การสูญหายอย่างสุ่ม (Missing at random: MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR) โดยรูปแบบการตอบสนองข้อสอบภายใต้สภาวะการสูญหายของข้อมูลแต่ละประเภท มี 3 แบบ คือ การตอบถูก (Correct: CR) ตอบผิด (Incorrect: IN) และละเว้นการตอบสนองข้อสอบ (Omitted: OM) ใช้ขนาดกลุ่มตัวอย่าง 2 ขนาด คือ 1,000 และ 3,000 คน อัตราการสูญหายของข้อมูลที่ต่างกัน 2 ระดับคือ 10% และ 30% ความยาวแบบสอบ 2 ระดับ คือ 20 และ 40 และขนาดอิทธิพลของ DIF 2 ขนาด คือ 0.5 และ 1.0 เกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ พิจารณาจากดัชนีค่าความลำเอียง (BIAS) และดัชนีค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยพิจารณาจาก Type I error และ Power of test เพื่อประโยชน์ในการนำไปเลือกใช้กับ

สถานการณ์ต่าง ๆ ให้เหมาะสมกับข้อมูลจริงต่อไปในอนาคต โดยสามารถแสดงกรอบแนวคิดในการวิจัย ดังภาพที่ 1-1



ภาพที่ 1-1 กรอบแนวคิดในการวิจัย

ประโยชน์ที่ได้รับจากการวิจัย

1. ผลการศึกษาทำให้ทราบถึงวิธีการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ภายใต้เงื่อนไขในการศึกษาคือ สภาวะการสูญหาย ขนาดกลุ่มตัวอย่าง อัตราการสูญหาย ความยาวแบบสอบ และขนาดอิทธิพลของ DIF ซึ่งผลการวิจัยที่ได้นำไปสู่การสรุปผลเกี่ยวกับ การประมาณค่าพารามิเตอร์ข้อสอบ ภายใต้สถานการณ์ที่มีความแตกต่างกัน

2. ผลการศึกษาทำให้ทราบเกี่ยวกับวิธีการประมาณค่าพารามิเตอร์ข้อสอบ และความสามารถผู้สอบว่าวิธีแมกซิมัมไลค์ลิฮู๊ด (Maximum likelihood: ML) มีลักษณะเป็นอย่างไร ซึ่งทำให้สามารถเลือกใช้วิธีการประมาณค่าด้วยวิธีดังกล่าวกับข้อมูลลักษณะอื่นได้อย่างเหมาะสม

3. เพื่อเป็นแนวทางในการจัดการข้อมูลสูญหาย และวิธีการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบ และวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อีกวิธีหนึ่งให้แก่นักวัดผลการศึกษา ในการตรวจสอบและพัฒนาคุณภาพข้อสอบ

4. ผลที่ได้จากการวิจัย ทำให้ผู้สนใจสามารถตัดสินใจเลือกใช้วิธีในการจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบ ได้ถูกต้องภายใต้เงื่อนไขที่สนใจศึกษาที่แตกต่างกัน

5. ได้แนวทางในการแก้ไขจุดอ่อน สำหรับวิธีที่ใช้ในการจัดการข้อมูลสูญหายในประเด็นที่แตกต่างกัน

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ผู้วิจัยมุ่งศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหาย เพื่อนำมาใช้ในการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบ และการทำหน้าที่ต่างกันของข้อสอบ ภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดของกลุ่มตัวอย่าง อัตราการสุญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพลของ DIF ในการศึกษาดังกล่าวผู้วิจัยได้ศึกษากรอบแนวคิด ทฤษฎี หลักการ รูปแบบ วิธีการจัดการข้อมูลสูญหาย การทำหน้าที่ต่างกันของข้อสอบ ตลอดจนงานวิจัยต่าง ๆ ที่เกี่ยวข้อง เพื่อนำองค์ความรู้ที่ได้มาประยุกต์ใช้กับการศึกษาในครั้งนี้ โดยผู้วิจัยแบ่งการนำเสนอเนื้อหาสาระออกเป็น 7 ตอน ดังนี้

ตอนที่ 1 แนวคิด ทฤษฎีเกี่ยวกับข้อมูลสูญหาย และแนวทางการจัดการข้อมูลสูญหาย

1.1 ความหมายของข้อมูลสูญหาย

1.2 สาเหตุการเกิดข้อมูลสูญหาย

1.3 ประเภทของข้อมูลสูญหาย

1.4 ความเสียหายที่เกิดขึ้นจากการมีข้อมูลสูญหาย

1.5 วิธีการจัดการข้อมูลสูญหาย

ตอนที่ 2 แนวคิด เกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

2.1 โมเดลพื้นฐานของทฤษฎีการตอบสนองข้อสอบ

2.2 โมเดลทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 3 แนวคิด ทฤษฎีเกี่ยวกับการประมาณค่าพารามิเตอร์

ตอนที่ 4 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte cario simulation study)

ตอนที่ 5 คุณสมบัติของตัวประมาณค่าที่ดี

ตอนที่ 6 แนวคิด ทฤษฎีเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ

ตอนที่ 7 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 แนวคิด ทฤษฎีเกี่ยวกับข้อมูลสูญหาย และแนวทางการจัดการข้อมูลสูญหาย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับแนวทางการจัดการข้อมูลสูญหายนั้น ผู้วิจัยได้นำเสนอ ดังนี้

ความหมายของข้อมูลสูญหาย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องได้มีนักวิจัยหลายท่านได้ให้ความหมายของข้อมูลสูญหายไว้ ดังนี้

ข้อมูลสูญหาย คือค่าสังเกตที่ต้องการทราบค่าแต่ไม่สามารถทราบค่าได้ โดยที่ค่านั้นควรจะสามารถทราบค่าได้หากวิธีการที่ใช้ในการรวบรวมข้อมูลหรือในการวัดค่ามีประสิทธิภาพดีขึ้นหรือมีความเหมาะสมมาก (Huisman, 1998 อ้างถึงใน จำลอง วงษ์ประเสริฐ, 2554)

ข้อมูลสูญหาย หมายถึง ค่าสังเกตได้ที่ต้องการทราบค่าแต่ ไม่สามารถทราบค่าได้ ซึ่งเป็นตัวบ่งบอกถึงความไม่พอเพียงของการตอบ (Lack of response) เช่น “ไม่ทราบ” “ไม่ตอบ/ ปฏิเสธที่จะตอบ” “ไม่มีความรู้เพียงพอ” หรือเหตุผลอื่น ๆ (Schafer & Graham, 2002; Little & Rubin,1987 อ้างถึงใน กมลทิพย์ ศรีหาเศษ, 2555)

สาเหตุหลักของการเกิดข้อมูลสูญหาย

สาเหตุการเกิดข้อมูลสูญหายมักเป็นผลอันเนื่องมาจากหลายกรณี โดยเหตุผลพื้นฐานมักเป็นผลจาก (ปิยะภรณ์ ประสิทธิ์วัฒนเสรี และสุคนธ์ ประสิทธิ์วัฒนเสรี, 2552, หน้า 55)

1. การไม่ยอมมาแสดงตัวของหน่วยตัวอย่างในช่วงเวลาเฝ้าติดตามหน่วยตัวอย่าง

2. หน่วยตัวอย่างปฏิเสธการตอบคำถามในบางคำถามของแบบฟอร์มที่ใช้รวบรวมข้อมูล ซึ่งส่วนใหญ่มักเป็นคำถามที่กระทบต่อความรู้สึกได้ง่าย

3. หน่วยตัวอย่างไม่ทราบคำตอบ ทั้งนี้อาจเป็นผลมาจากปัญหาในเรื่องของความจำ เช่น อาจจำไม่ได้ว่าเคยเข้ารับการตรวจสุขภาพมาแล้วกี่ครั้ง หรือเข้ารับการตรวจสุขภาพครั้งสุดท้ายเมื่อไร นอกจากนี้อาจเป็นผลเนื่องจากการไม่เข้าใจในความหมายของคำต่าง ๆ ที่ใช้ในการสร้างคำถาม

4. คำถามที่ใช้ไม่ครอบคลุมทุกกรณีจึงทำให้เกิดข้อมูลสูญหาย ยกตัวอย่างเช่น คำถามที่ต้องการทราบว่าหน่วยตัวอย่างได้รับการฉีดวัคซีนป้องกันเชื้อบาดทะยักครั้งสุดท้ายเมื่อใด ซึ่งในรายของผู้ที่ไม่เคยฉีดวัคซีนอาจไม่ตอบคำถามในข้อนี้เพราะเห็นว่าไม่เกี่ยวข้องกับตนเอง หรือในกรณีที่ต้องการทราบว่า การเข้าพบแพทย์ครั้งล่าสุดในรอบ 12 เดือนที่ผ่านมา สำหรับผู้ที่ไม่เคยเข้าพบแพทย์ใน 12 เดือนที่ผ่านมา ก็อาจไม่ตอบคำถามในข้อนี้

1.2.5 การนำข้อมูลเข้าสู่ระบบประมวลผลทางคอมพิวเตอร์ ซึ่งอาจเป็นผลจากความผิดพลาดของระบบฐานข้อมูล การเชื่อมโยงระหว่างข้อมูลจากฝ่ายต่าง ๆ

ประเภทของข้อมูลสูญหาย

ชนิดหรือประเภทของข้อมูลสูญหาย (Missing data) นั้น เป็นปรากฏการณ์ทั่วไปที่พบมาก ในข้อมูลที่ได้จากการสังเกตและข้อมูลจากการทดลอง ซึ่งอาจเกี่ยวกับโครงสร้างการออกแบบของการประเมินผลในกรณีที่เป็นการประเมินผล Large-scale ของผลสัมฤทธิ์ของผู้เรียน ซึ่งใช้การออกแบบ Complex rotation ในการทดสอบความแตกต่างของข้อสอบเพื่อนำไปสู่ความแตกต่างของกลุ่มผู้เรียน นอกจากนี้ข้อมูลสูญหายยังอาจเกี่ยวกับลักษณะการตอบสนอง ข้อสอบของผู้สอบ ซึ่งอาจเป็นตัวแปรอิสระของการออกแบบการศึกษา เช่น ความเหนื่อยล้าของผู้สอบ ความไม่พอเพียงของเวลา หรือการลังเลในการตอบสนองข้อสอบที่มาจากสาเหตุอื่น ซึ่งประเภทของข้อมูลสูญหายนั้น จำแนกลักษณะพิเศษระหว่างข้อมูลสูญหายได้เป็น 3 ประเภท คือ การสูญหายโดยสมบูรณ์อย่างสุ่ม (Missing completely at random: MCAR) การสูญหายอย่างสุ่ม (Missing at random: MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR) ทั้งนี้ ในการจำแนกประเภทของข้อมูลสูญหายตามแนวคิดของนักวิชาการนั้น แบ่งได้เป็น 2 กลุ่ม โดยกลุ่มแรกแบ่งประเภทของข้อมูลสูญหายเป็น 2 ประเภท คือ การสูญหายที่สามารถละเลยได้ (Lgnorable missing) กับการสูญหายที่ละเลยไม่ได้ (Nonignorable missing) (นงลักษณ์ วิรัชชัย, 2552; Garson, 2008; Statistics Solutions, 2009) ส่วนกลุ่มหลังแบ่งประเภท ของข้อมูลสูญหายเป็น 3 ประเภท คือ การสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) การสูญหายอย่างสุ่ม (MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) (Allison, 2002; Finch, 2008; Howell, 2009; Little & Rubin, 2002; Robitzsch & Rupp, 2009; Rubin, 1976; Schafer & Graham, 2002) ในที่นี้จะจำแนกประเภทข้อมูลสูญหายเป็น 3 ประเภท ซึ่งมีรายละเอียดดังนี้

1. **การสูญหายโดยสมบูรณ์อย่างสุ่ม** (Missing completely at random: MCAR)

ในการสูญหายของข้อมูลนั้นมีเหตุผลหลายประการว่าเพราะเหตุใดข้อมูลจึงเกิดการสูญหาย ซึ่งอาจสูญหายจากความบกพร่องของเครื่องมือที่ใช้ ปัญหาเนื่องจากสภาพอากาศ หรือการเจ็บป่วยของกลุ่มตัวอย่าง หรือการนำเข้าข้อมูลไม่ถูกต้อง เหล่านี้เรียกว่าการสูญหายโดยสมบูรณ์อย่างสุ่ม (Missing completely at random: MCAR) เมื่อกล่าวว่าข้อมูลเป็นการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) นั้น มีความหมายว่า ความน่าจะเป็นของค่าสังเกต (x_i) เป็นการสูญหายที่ไม่สัมพันธ์กับค่าสังเกต (x_i) หรือค่าของตัวแปรอิสระตัวอื่น

สมการความน่าจะเป็นของการสูญหาย สามารถอธิบายเพิ่มเติมได้ ดังนี้ (Allison, 2002; Little & Rubin, 2002; Schafer & Graham, 2002)

เมื่อให้ Y_{com} แทน ชุดข้อมูลสมบูรณ์ (Complete data set) ซึ่งเป็นชุดข้อมูลซึ่งประกอบด้วยข้อมูลจากการสังเกต (Observed data: Y_{obs}) และข้อมูลสูญหาย (Missing data: Y_{mis}) และ M แทนเมทริกซ์ ซึ่งมีตัวบ่งชี้ว่าข้อมูลเป็นการสูญหายหรือไม่

ข้อมูลซึ่งเป็นการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) จะมีรูปแบบการสูญหายเป็นผลที่เกิดขึ้นหรือเป็นตัวแปรตามที่ไม่ใช่ผลจากค่าของตัวแปรอิสระที่ได้จากการวัดและค่าที่ ไม่ได้สังเกตของตัวแปรนั้น การแจกแจงของรูปแบบการสูญหายเมื่อให้ข้อมูลสมบูรณ์นั้น เขียนสมการได้ดังนี้

$$P(M|Y_{com}) = P(M|Y_{obs}, Y_{mis}) = P(M)$$

นอกจากนี้ Rubin (1976) ได้ยกตัวอย่างว่า กรณีการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) นี้อาจเป็นผลมาจากการที่ผู้สอบไม่มีสมาธิ หรืออาจเป็นผลมาจากการออกแบบเงื่อนไขการวัด (Facet) เช่น การใช้หลักฐานคะแนนผู้สอบจากมาตรวัดประสิทธิภาพตนเองแบบสั้น (Short form of the efficacy scale) ซึ่งผู้วิจัยออกแบบการเก็บรวบรวมข้อมูลโดยการสุ่มชุดข้อมูลผู้สอบจากการสอบฉบับเต็ม

ข้อดีของข้อมูลประเภทการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) คือการไม่เกิดความลำเอียงในการวิเคราะห์ ซึ่งอาจเสียอำนาจการทดสอบจากการออกแบบการศึกษา แต่การประมาณค่าพารามิเตอร์จะไม่เกิดความลำเอียงเนื่องจากการสูญหายของข้อมูล

2. การสูญหายอย่างสุ่ม (Missing at random: MAR)

บ่อยครั้งที่การสูญหายของข้อมูลไม่ใช่การสูญหายโดยสมบูรณ์อย่างสุ่ม แต่จัดอยู่ในประเภทการสูญหายอย่างสุ่ม (Missing at random: MAR) สำหรับข้อมูลที่เป็นการ สูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) นั้น ความน่าจะเป็นของค่าสังเกต (x_i) ซึ่งเกิดการสูญหายนั้นไม่สัมพันธ์กับค่าสังเกต (x_i) หรือตัวแปรอื่นในการวิเคราะห์ แต่ข้อมูลจะพิจารณาว่าเป็นข้อมูลสูญหายอย่างสุ่ม (MAR) เมื่อพบเงื่อนไขว่าการสูญหายนั้นไม่ได้ขึ้นกับค่าสังเกต (x_i) ภายหลังการควบคุมตัวแปรอื่นแล้ว

สำหรับข้อมูลซึ่งเป็นการสูญหายอย่างสุ่ม (MAR) นั้น ความน่าจะเป็นของการสูญหายเป็นผลจากค่าตัวแปรสังเกตได้อื่น ๆ เท่านั้น ซึ่งหมายความว่า กระบวนการสูญหายอาจสัมพันธ์กับการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) จากเงื่อนไขของตัวแปรร่วมที่สังเกตได้ (Observable covariates) เขียนสมการได้ ดังนี้

$$P(M|Y_{com}) \quad = \quad P(M|Y_{obs}, Y_{mis}) \quad = \quad P(M|Y_{obs})$$

นอกจากนี้ การสูญหายอย่างสุ่ม (MAR) ยังสามารถเรียกได้อีกอย่างว่า การสูญหายที่ละเลยได้ (Ignorable missing) เนื่องจากในการดำเนินการในการกระบวนการสรุปอ้างอิง เช่น การประมาณค่าที่อิง Likelihood ข้อมูลที่ไม่ได้สังเกตอาจถูกละเลยใน Lkelihood

ในกรณีนี้ Rubin (1976) ได้ยกตัวอย่างว่า ผู้สอบที่มีระดับแรงจูงใจต่ำมักจะข้ามการตอบข้อคำถามในแบบสอบถามการประเมินประสิทธิภาพตนเอง นอกจากนี้ ค่าสูญหายที่เกิดขึ้นในกลุ่มผู้สอบที่มีระดับแรงจูงใจต่ำ เมื่อพิจารณาเทียบกับกลุ่มที่มีแรงจูงใจระดับปกติ พบว่า ไม่มีความสัมพันธ์กันระหว่างแนวโน้มของค่าสูญหายหรือประสิทธิภาพตนเองของผู้สอบ ซึ่งหมายความว่า ไม่มีความสัมพันธ์ส่วนที่เหลือระหว่างการสูญหายกับประสิทธิภาพตนเอง เมื่อระดับแรงจูงใจถูกควบคุมแล้ว จึงเป็นข้อสังเกตที่สำคัญของการสูญหายอย่างสุ่ม (MAR) ที่คะแนนแรงจูงใจถูกรวมเข้าไปในการวิเคราะห์คะแนนประสิทธิภาพตนเอง

อย่างไรก็ตาม หากคะแนนแรงจูงใจไม่ได้รวมผลการวัดเข้าไปในโมเดลการวิเคราะห์แล้วกลไกการสูญหายจะมีลักษณะเป็นการสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR)

3. การสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR)

ถ้าข้อมูลไม่ได้เกิดการสูญหายอย่างสุ่ม หรือสูญหายโดยสมบูรณ์อย่างสุ่มแล้ว การสูญหายของข้อมูลจะเป็นการสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR)

สำหรับข้อมูลที่เป็นการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) นั้น ไม่สามารถอธิบายการสูญหายด้วยตัวแปรสังเกตได้ เพราะการสูญหายเป็นผลที่เกิดขึ้นแน่นอนจากค่าที่ไม่ได้สังเกต Y_{mis} เขียนสมการได้ ดังนี้

$$P(M|Y_{com}) \quad = \quad P(M|Y_{obs}, Y_{mis}) \quad \neq \quad P(M|Y_{obs})$$

นอกจากนี้การสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) สามารถเรียกได้อีกอย่างว่า การสูญหายที่ละเลยไม่ได้ (Nonignorable missing) และวิธีการเฉพาะภายใต้ข้อมูลสูญหาย ซึ่งต้องใช้โมเดลในการกำจัดความลำเอียงในการอ้างอิงผลสรุป

ตัวอย่างเกี่ยวกับ การสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) นี้ Rubin (1976) ยกตัวอย่างว่า ผู้สอบที่มีระดับความสามารถทางคณิตศาสตร์ต่ำจะข้ามข้อสอบที่ผู้สอบไม่มีความเชื่อมั่นว่า

จะแก้ปัญหาเพื่อหาคำตอบได้ ซึ่งในกรณีนี้ ค่าสูญหายจะสัมพันธ์กับระดับของการประเมินความสามารถของผู้สอบ หรืออาจกล่าวได้อีกอย่างหนึ่งว่า การสูญหายที่ไม่ใช่อย่างสุ่มเป็นผลของตัวแปรอิสระที่สัมพันธ์กับการสูญหาย

กรณีที่ผู้วิจัยมีข้อมูลสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ปัญหาที่เกิดขึ้นนั้น แนวทางเดียวที่จะได้มาซึ่งการประมาณค่าที่ไม่ลำเอียงของค่าพารามิเตอร์ คือ การสร้างโมเดลการสูญหาย นอกจากนี้ อาจจำเป็นต้องเขียนโมเดลซึ่งอธิบายการสูญหายของข้อมูล โดยโมเดลนี้จะถูกรวมเข้าไปในโมเดลที่สมบูรณ์กว่าภายหลังสำหรับการประมาณค่าการสูญหายซึ่งเป็นเรื่องที่สามารถทำได้

4. ความเสียหายที่เกิดขึ้นจากการมีข้อมูลสูญหาย

ความเสียหายที่เกิดขึ้นจากการมีข้อมูลสูญหาย ก่อให้เกิดปัญหาดังต่อไปนี้ (จำลอง วงษ์ประเสริฐ, 2554)

4.1 การสรุปผลผิดพลาด (Misleading) ในกรณีที่ข้อมูลสูญหายมิได้มีค่าใกล้เคียงกับข้อมูลของตัวแปรเดียวกันที่เหลืออยู่ แต่กลับมีค่าผิดแผกแตกต่างไปเป็นอย่างมาก การสรุปผลจากการวิเคราะห์ข้อมูลเท่าที่มีอยู่จะได้คำตอบที่ผิดจากความเป็นจริง

4.2 ค่าประมาณที่ได้ไม่มีประสิทธิภาพ (Inefficient) ค่าประมาณที่ดีต้องอาศัยข้อมูลของตัวแปรที่สนใจที่มากพอ และต้องมีความหลากหลาย เมื่อมีข้อมูลสูญหายทำให้ค่าประมาณที่ได้ไม่เป็น ตัวแทนที่ดีสำหรับสิ่งที่ต้องการประมาณ

4.3 สถิติที่ใช้ในการวิเคราะห์ข้อมูลไม่เหมาะสม สถิติที่ใช้ในการวิเคราะห์ข้อมูลที่ได้วางแผนไว้กับข้อมูลที่ต้องได้จากข้อมูลครบถ้วนตามขนาดตัวอย่าง (Sample size) ซึ่งได้มีการกำหนดไว้แล้ว ถ้าข้อมูลเกิดการสูญหายสภาพการสุ่ม (Random) ย่อมเสียไป จึงกลายเป็นสภาพสุ่มบางส่วนบางที่

4.4 อำนาจการทดสอบลดลง (Power of the test decreased) ข้อมูลสูญหายทำให้ตัวสถิติที่ใช้ทดสอบสมมุติฐาน (Hypothesis testing) เกี่ยวกับตัวแปรหรือสมมุติฐานเกี่ยวกับระดับความสัมพันธ์ระหว่างตัวแปรผิดพลาดคลาดเคลื่อน (Witta, 2000, pp. 1-5)

4.5 ความสัมพันธ์ต่ำกว่าความเป็นจริง (Downward bias) ข้อมูลสูญหายทำให้ค่าสหสัมพันธ์ที่คำนวณได้จากกลุ่มตัวอย่างคาดคะเนความสัมพันธ์จริงผิดพลาด โดยความผิดพลาดจะปรากฏในลักษณะต่ำกว่าความเป็นจริง สาเหตุที่สำคัญคือ การสูญหายของข้อมูลทำให้ความแปรปรวนของตัวแปรเปลี่ยนไป ระดับความสัมพันธ์กับตัวแปรอื่นจึงเปลี่ยนแปลงตาม (Kim & Curry, 1977, pp. 215-240)

1.4.6 สถิติวัดคุณภาพแบบทดสอบได้รับผลกระทบ ในการวัดคุณภาพของมาตรวัด เช่น แบบสอบถาม แบบประเมินต่าง ๆ นั้น เราต้องการมาตรวัดที่สามารถวัดได้ถูกต้อง ผลของการสูญหายของข้อมูลที่มีต่อค่าเฉลี่ยและความแปรปรวน ทำให้คุณภาพแบบทดสอบได้รับผลกระทบซึ่งมีผลทำให้แบบทดสอบที่ผ่านการวัดขาดความถูกต้องและความเชื่อถือได้

1.4.7 การสรุปผลเกิดความเอนเอียง ผลการศึกษามิได้ชี้ไปที่ความจริงของปรากฏการณ์นั้นแต่ชี้ไปที่อื่นเนื่องจากเกิดข้อมูลสูญหาย (Kwang, 2000; DeSilvio, 1999)

5. วิธีการจัดการข้อมูลสูญหาย

วิธีการจัดการกับข้อมูลสูญหาย (Methods of handling missing data) จะมีหลายวิธีการให้เลือกใช้ โดยการพิจารณาเลือกใช้วิธีการใดขึ้นอยู่กับลักษณะของข้อมูลสูญหายที่เกิดขึ้น ซึ่งหากเลือกวิธีการที่ไม่เหมาะสมมาใช้ในการจัดการกับข้อมูลสูญหายดังกล่าวแล้ว อาจเป็นการเพิ่มค่าความคลาดเคลื่อนและทำลายผลลัพธ์ที่ควรจะได้ สำหรับวิธีการจัดการกับข้อมูลสูญหายที่มักถูกเลือกนำมาใช้สามารถจำแนกได้ 2 กลุ่ม คือ วิธีจัดการข้อมูลสูญหายแบบดั้งเดิม และวิธีการจัดการข้อมูลสูญหายแบบใหม่ รายละเอียดดังนี้ (ปิยะภรณ์ ประสิทธิ์วัฒนเสรี และสุคนธ์ ประสิทธิ์วัฒนเสรี, 2552, หน้า 57-59; อรวรรณ กีรติสิโรจน์, 2558, หน้า 27-30; กมลทิพย์ ศรีหาเศษ, 2555, หน้า 46-61)

5.1 วิธีจัดการข้อมูลสูญหายแบบดั้งเดิม

5.1.1 Listwise data deletion

Complete-case analysis หรือที่รู้จักในชื่ออื่น ๆ ได้แก่ Casewise deletion หรือ Listwise deletion วิธีการคือ ตัดข้อมูลของตัวอย่างรายที่มีข้อมูลสูญหายอย่างน้อย 1 ตำแหน่งออกจากการวิเคราะห์ นั่นคือไม่สนใจข้อมูลสูญหายที่เกิดขึ้น โดยจะทำการวิเคราะห์ข้อมูลจากข้อมูลเฉพาะส่วนที่สมบูรณ์ แนวทางนี้จะมีความเหมาะสมในกรณีที่ข้อมูลสูญหายมีจำนวนน้อยมาก และเป็นเทคนิคที่มีการใช้กันอย่างกว้างขวางและยังเป็นเทคนิคพื้นฐานในการจัดการข้อมูลสูญหาย เนื่องจากเป็นวิธีการที่ง่าย อย่างไรก็ตามเทคนิคนี้มีข้อจำกัด เนื่องจากการตัดข้อมูลออกจากการวิเคราะห์เป็นการลดขนาดตัวอย่าง จึงทำให้สูญเสียความแม่นยำ (Precision) ในการประมาณค่าและอาจจะทำให้เกิดอคติ (Bias) ในการประมาณค่าได้อีกด้วย เว้นแต่ว่าข้อมูลมีกลไกการสูญหายแบบ MCAR (Missing completely at random) นอกจากนี้การมีข้อมูลสูญเสียความแม่นยำและอคติให้เกิดน้อยที่สุดสำหรับการเลือกใช้ Complete-case analysis

5.1.2 Available-case analysis

Available-case analysis หรือ Pairwise deletion เป็นวิธีการที่คล้ายกับ Complete-case analysis แต่จะตัดข้อมูลที่มีการสูญหายเฉพาะตัวแปรนั้น ๆ ไม่ได้ตัดออกทั้งตัวอย่าง ใช้ในกรณีที่ต้องการวิเคราะห์หรือเปรียบเทียบข้อมูลเฉพาะตัวแปรที่สนใจ เช่น การวิเคราะห์สถิติพรรณา (Descriptive statistic) การวิเคราะห์ตัวแปรเดี่ยว (Anivariable analysis) แต่ไม่เหมาะสำหรับการวิเคราะห์ตัวแปรหลายตัว (Multivariable analysis) ข้อดีคือเทคนิคนี้มีประสิทธิภาพมากกว่า Complete-case analysis เนื่องจากตัวอย่างถูกตัดออกจากการวิเคราะห์น้อยกว่า แต่ก็มีข้อด้อยเนื่องจากความหลากหลายของจำนวนตัวอย่างในการวิเคราะห์แต่ละครั้ง จึงเป็นการลดความแม่นยำในการประมาณค่าและก่อให้เกิดอคติขึ้นได้ เช่นเดียวกับเทคนิค Complete-case analysis จึงควรใช้กับข้อมูลสูญหายที่มีกลไกแบบ MCAR

5.1.3 Weight complete-case analysis

เป็นเทคนิคที่ประยุกต์จาก Complete-case analysis โดยการถ่วงน้ำหนักระหว่างข้อมูลที่สมบูรณ์และข้อมูลที่สูญหายเพื่อลดการเกิดอคติ วิธีการนี้นิยมใช้ในการวิจัยเชิงสำรวจ แม้ว่าเทคนิคการถ่วงน้ำหนักนี้จะช่วยลดอคติ แต่ความแปรปรวนจะเพิ่มขึ้น ดังนั้นจึงเป็นการลดความแม่นยำในการประมาณค่า นอกจากนี้วิธีการคำนวณความแปรผันดังกล่าวยังมีความซับซ้อน โดยสรุปแล้วการวิเคราะห์ด้วยการถ่วงน้ำหนักนี้เหมาะสมกับกรณีที่มีข้อมูลขนาดใหญ่และมีแนวโน้มที่จะเกิดอคติมากกว่าการสูญเสียความแม่นยำ

5.1.4 Mean and median imputation

เป็นเทคนิคพื้นฐานของ Single imputation โดยการแทนค่าสูญหายด้วยค่ากลาง ได้แก่ ค่าเฉลี่ย (Mean) หรือค่ามัธยฐาน (Median) ของข้อมูลที่ทราบค่าภายในตัวแปรนั้น ๆ แม้ว่าเทคนิคนี้จะได้รับความนิยม อย่างไรก็ตามอาจเลิกอคติในการประมาณค่าพารามิเตอร์ได้เนื่องจากข้อมูลสูญหายจะถูกแทนที่ด้วยค่าของตัวแปรนั้น ๆ ไม่ขึ้นอยู่กับตัวแปรอื่น ๆ ยกเว้นกรณีทีข้อมูลมีกลไกสูญหายแบบ MCAR จึงไม่เกิดอคติ วิธีนี้อาจทำให้ค่าความแปรปรวนต่ำกว่าความเป็นจริงจนค่าความแปรปรวนอาจจะเล็กมากเมื่อข้อมูลสูญหายมีจำนวนมากขึ้น ดังนั้นค่าสูญหายที่ถูกแทนที่ด้วยวิธีนี้อาจจะไม่เหมาะสม เมื่อมีการปรากฏให้เห็นว่าความน่าจะเป็นของ Type I error มีขนาดมากขึ้น

5.1.5 Regression imputation

เป็นเทคนิคที่เริ่มมีตัวแปรอื่น ๆ เข้ามาเกี่ยวข้องโดยการสร้างสมการถดถอยจากข้อมูลของตัวแปรอื่น ๆ ที่สมบูรณ์เพื่อทำนายค่าของตัวแปรที่มีข้อมูลสูญหาย ข้อด้อยคือเทคนิคนี้

ไม่ได้มีค่าความคลาดเคลื่อน (Error term) รวมอยู่ในการประมาณค่าด้วย ดังนั้น เส้นสมการถดถอยที่สร้างขึ้นจึงไม่มีค่า Residual variance ซึ่งทำให้สมการถดถอยนี้มีค่าความสัมพันธ์เกินจริงเทคนิคนี้เหมาะสำหรับกลไกการสูญหายแบบ MAR (Missing at random)

5.1.6 Stochastic regression imputation

เป็นการแทนค่าข้อมูลสูญหายซึ่งพัฒนามาจากเทคนิค Regression imputation ด้วยการเพิ่มค่าความคลาดเคลื่อนเข้าไปในสมการถดถอย เทคนิคนี้ก่อให้เกิดอคติน้อยกว่าเทคนิคอื่น ๆ ข้างต้น อย่างไรก็ตามวิธีการนี้ได้รวมความไม่แน่นอนไว้ในการทำนาย ค่าสูญหายจึงทำให้มีค่าความแปรปรวนเพิ่มขึ้น ซึ่งถือเป็นข้อจำกัดสำคัญของเทคนิค Regression imputation ทุกรูปแบบ

5.1.7 Hot and cold desk imputation

Hot desk คือวิธีการแทนค่าข้อมูลสูญหายจากค่าของตัวอย่างรายที่มีลักษณะใกล้เคียงกันมากที่สุดภายในข้อมูลชุดเดียวกัน ข้อดี คือเทคนิคนี้ไม่ต้องการข้อสมมุติฐานของพารามิเตอร์และไม่ต้องระมัดระวังในการสร้างโมเดลเพื่อระบุค่า ข้อด้อยคือ อคติที่อาจจะเกิดขึ้นเนื่องจากการแทนค่าสูญหายจากข้อมูลที่สมบูรณ์เท่านั้น วิธีการนี้ทำให้ค่าความแปรปรวนต่ำกว่าความเป็นจริงเช่นเดียวกับการแทนค่า ด้วยค่ากลางเนื่องจากข้อมูลสูญหายจะถูกแทนค่าด้วยค่าของตัวอย่างที่สมบูรณ์เท่านั้น จึงทำให้ข้อมูลสูญหายที่ถูกแทนค่ามีค่าใกล้เคียงกัน นอกจากนี้ยังต้องการข้อมูลที่สมบูรณ์จำนวนมาก เพื่อที่จะได้เพิ่มโอกาสในการเลือกตัวอย่างที่มีลักษณะใกล้เคียงกัน และยังไม่มีเกณฑ์ที่ชัดเจนในการเลือกตัวอย่างที่มีข้อมูลสมบูรณ์มาแทนค่าข้อมูลสูญหาย ด้วยเหตุนี้ Hot desk จึงเป็นตัวเลือกสุดท้ายที่จะใช้จัดการข้อมูลสูญหายกรณีที่มีกลไกการเกิดแบบ MAR

Cold desk มีวิธีการเช่นเดียวกับ Hot desk แต่จะคัดข้อมูลจากตัวอย่างในชุดข้อมูลอื่นเพื่อมาแทนค่าข้อมูลสูญหาย จึงมีข้อด้อยเพิ่มเติมจากการที่ใช้ข้อมูลภายนอกคือเป็นการเพิ่มอคติ ดังนั้น เทคนิคนี้จึงไม่เป็นที่นิยม

5.1.8 Last observation arried forward

เป็น Single imputation รูปแบบหนึ่งที่ใช้กับงานวิจัยระยะยาว (Longitudinal studyl7d) เช่น การทดลองยา ซึ่งมีการเก็บข้อมูลซ้ำจากตัวอย่างเดิมเมื่อเวลาผ่านไป จึงมักจะมีตัวอย่างที่ออกจากการศึกษาก่อนกำหนด (Dropout) เทคนิคนี้มีวิธีการแทนค่าสูญหายโดยใช้ค่าสุดท้ายที่สามารถรวบรวมได้ของตัวอย่างรายนั้น ๆ โดยมีข้อสมมุติว่าข้อมูลไม่เปลี่ยนแปลงไปจากเดิม และมีกลไกการสูญหายแบบ MAR

5.1.9 Worst case analysis

เป็นเทคนิคที่นิยมใช้เมื่อข้อมูลสูญหายเกิดขึ้นกับตัวแปรตาม แม้ว่าจะสามารถใช้ได้เมื่อข้อมูลสูญหายเกิดขึ้นกับตัวแปรอิสระด้วยเช่นกัน วิธีการคือแทนค่าสูญหายด้วยค่าที่แย่ที่สุดหรือภายใต้สถานการณ์ที่เลวร้ายที่สุด (Worst case scenario) เช่น แทนค่าด้วยการเสียชีวิต สำหรับการศึกษาเกี่ยวกับระยะรอดชีพ (Survival outcome) เทคนิคนี้ไม่ได้มีวัตถุประสงค์ที่จะแทนค่าที่ถูกต้อง แต่แนวทางนี้ย่อมดีกว่าการปล่อยให้ข้อมูลมีค่าสูญหาย ซึ่งแสดงให้เห็นถึงความไม่มีคุณภาพของการศึกษา

5.1.10 Missing indicator method

การใช้ตัวบ่งชี้ตัวแปรที่มีค่าสูญหายเป็นอีกวิธีการหนึ่งที่ง่ายสำหรับการจัดการข้อมูลสูญหายของตัวแปรร่วม (Covariate) หรือตัวแปรกวน (Confounder) เมื่อมีการวิเคราะห์ถดถอย โดยการเพิ่มตัวแปรเข้าไปในโมเดลการวิเคราะห์บ่งชี้ว่ามีค่าของบางตัวแปรที่เป็นค่าสูญหาย แม้ว่าวิธีการนี้จะทำให้เกิดอคติ โดยขนาดของอคติจะขึ้นอยู่กับลักษณะของตัวแปรตาม อย่างไรก็ตาม เทคนิคนี้ยังมีการใช้บ่อย เนื่องจากเป็นวิธีที่ช่วยปรับค่าสูญหายที่ใช้งานง่าย

5.2 วิธีจัดการข้อมูลสูญหายแบบใหม่

5.2.1 Expectation maximization (EM) approach

วิธีการนี้เป็นการอาศัยหลักของกระบวนการวนซ้ำ (Iterative procedure) ระหว่าง 2 ขั้นตอน โดยขั้นตอนแรก เป็นขั้นตอนที่เรียกว่า Expectation (E) step ซึ่งจะทำการประมาณค่าคาดหวังจากฟังก์ชัน Likelihood ภายใต้ข้อมูลที่สมบูรณ์ สำหรับขั้นตอนที่สองเป็นขั้นตอนที่เรียกว่า Maximization (M) step เพื่อทำการแทนค่าคาดหวังของข้อมูลสูญหายด้วยค่าที่ได้จาก E step และทำการประมาณค่าคาดหวังจากฟังก์ชัน Likelihood ในกรณีถ้าไม่เกิดข้อมูลสูญหาย โดยจะทำการวนซ้ำระหว่าง 2 ขั้นตอนจนกว่าจะเกิดค่าที่ลู่เข้า (Convergence) หรือค่าที่มีการเปลี่ยนแปลงน้อยมาก ใช้ค่านั้นแทนค่าข้อมูลสูญหายที่เกิดขึ้น

5.2.2 Raw maximum likelihood methods:

เป็นวิธีการที่อาศัยข้อมูลสมบูรณ์ในการสร้างค่า Maximum likelihood ภายใต้ตัวแบบทางสถิติที่เหมาะสม ไม่ว่าจะเป็น Structural equation model, Regression model, ANOVA และ ANCOVA models

5.2.3 **การประมาณค่าทดแทนพหุ** (Multiple imputation: MI)

เป็นเทคนิคการแทนค่าข้อมูลสูญหายที่มีความหลากหลาย ค่อนข้างง่ายต่อการนำไปใช้ ซึ่งเป็นวิธีการที่ผสมผสานระหว่างวิธีการ EM และ Raw maximum likelihood methods ร่วมกับความสามารถของคุณสมบัติ Hot deck เพื่อทำการสร้างชุดจำลองของข้อมูล

ที่ได้ทำการแทนค่าข้อมูลสูญหายด้วย Imputed value แล้วขึ้นมาหลาย ๆ ชุด (ประมาณ 5 ถึง 10 ชุด) จากนั้นทำการวิเคราะห์ข้อมูลจากชุดต่าง ๆ บันทึกผลการวิเคราะห์ที่ได้

เป็นเทคนิคที่ค่อนข้างง่ายต่อการนำไปใช้ และมีการให้บริการจำนวนมากในการวิเคราะห์ด้วยโปรแกรมทางสถิติ เช่น SAS (SAS Institute, Cary, North Carolina), S-plus (Insightful Corporation, Seattle, WA), R (Foundation, Vienna, Austria), and Stata (StataCorp, College Station, TX)

การประมาณค่าทดแทนพหุ ได้รับการอธิบายรายละเอียดที่สมบูรณ์ไว้ในบทความหลายเรื่อง (Leite & Beretvas, 2004; Schafer, 1997; Schafer & Graham, 2002; Schafer & Olsen, 1998; Sinharay, Stern, & Russell, 2001) ได้รับการพัฒนาโดย Rubin (1987) โดยในเบื้องต้นมีจุดมุ่งหมายของการพัฒนา เพื่อให้เป็นทางเลือกของวิธีการที่ง่ายต่อการประมาณค่าทดแทน เช่น ตัวแทนของค่าเฉลี่ย (Mean substitution) การประมาณค่าทดแทนแบบ Hot deck การประมาณค่าทดแทนโดยใช้ Regression-based และการประมาณค่าทดแทนโดยการพิจารณาการแจกแจง (Huisman & Molenaar, 2001; Madow, Nisselson, & Olkin, 1983) วิธีการดังกล่าวที่ยกตัวอย่างมามีข้อแตกต่างกับเทคนิคการประมาณทดแทนเชิงเดี่ยวนี้ โดยการประมาณค่าทดแทนพหุใช้อธิบายสำหรับความคลุมเครือที่มีมาตั้งแต่เริ่มต้นในการสุ่มตัวอย่างจากประชากร โดยการนำเสนอ Degree ของการสุ่ม ในการประมาณค่าทดแทนและสร้างชุดข้อมูลประมาณทดแทน M แต่ละวิธีสามารถวิเคราะห์ได้ด้วยวิธีมาตรฐาน การประมาณค่าทดแทนพหุสามารถรวมสารสนเทศจากตัวแทนอื่นเข้าด้วยกันในกระบวนการประมาณค่าทดแทนในการนำเสนอค่าที่ถูกต้องแม่นยำกว่า การใช้การประมาณค่าทดแทนพหุต้องการข้อตกลงเกี่ยวกับรูปแบบความน่าจะเป็นภายใต้ชุดของข้อมูล เช่น การแจกแจงของพหุตัวแปรซึ่งในที่นี้คือความถี่ที่ใช้สำหรับตัวแปรต่อเนื่อง หรือการแจกแจงพหุเอกนามซึ่งปกติใช้กับตัวแปรไม่ต่อเนื่อง ที่ผ่านมา รูปแบบความน่าจะเป็นที่ถูกเลือกในการประมาณค่าพารามิเตอร์ทำไปถึงการใช้การแจกแจงแบบอิง Bayesian ภายหลัง ในการนำเสนอโมเดล Likelihood function ข้อมูลจากการสังเกตและการแจกแจงในเบื้องต้นจะใช้วิธี Markov chain monte carlo (MCMC) ในการเพิ่มข้อมูลไปยังการแจกแจงเบื้องต้นจากข้อมูลที่เพิ่มขึ้น จากการสร้างค่าประมาณทดแทน กระบวนการประมาณค่าทดแทนนี้ทำวนซ้ำ M ครั้ง เพื่อสร้างชุดข้อมูลตัวแปรอิสระ (Schafer & Olsen, 1998) แต่ละชุดข้อมูลจะกลายเป็นกรณี ในการวิเคราะห์สิ่งที่สนใจศึกษาต่อไป เช่น การประมาณค่าพารามิเตอร์ทฤษฎีการตอบสนองข้อสอบ (IRT) ผลของการวิเคราะห์แยก M ในภายหลังจะรวมเป็นค่าเดียวจากสมการ

$$\bar{Q} = \frac{\sum \hat{Q}_m}{M}$$

ความแปรปรวนของการประมาณค่านี้ประกอบด้วย 2 ส่วนคือ ความแปรปรวนระหว่างการประมาณค่าทดแทนและความแปรปรวนภายในการประมาณค่าทดแทน

ความแปรปรวนระหว่างการประมาณค่าทดแทนได้ดังสมการ

$$B = \frac{\sum (\hat{Q}_m - \bar{Q})^2}{M-1}$$

ความแปรปรวนภายในการประมาณค่าทดแทน ($\bar{U}$) เป็นค่าเฉลี่ยของการประมาณค่าความแปรปรวนข้ามการประมาณค่าทดแทน M ความแปรปรวนสำหรับการประมาณค่าทดแทนพหุ จะคำนวณภายหลังด้วยสมการ ดังนี้

$$T = \bar{U} + \left(1 + \frac{1}{M}\right)B$$

5.2.4 วิธีการประมาณค่าที่เป็นไปได้สูงสุดด้วยข้อมูลดิบ (Raw maximum likelihood methods) เป็นวิธีที่ใช้ข้อมูลสมบูรณ์ในการสร้างค่า Maximum likelihood ภายใต้รูปแบบทางสถิติที่เหมาะสม ได้แก่ Structural equation model, Regression model, ANOVA และ ANCOVA models

5.2.5 แนวคิด ที่ไม่อิงวิธีการของการประมาณค่าทดแทน

(Nonimputation-based approaches) แนวคิดที่ไม่อิงวิธีการของการประมาณค่าทดแทนวิธีหนึ่งสำหรับจัดการกับการตอบสนองข้อสอบที่สูญหายเป็นการจัดกระทำการตอบสนองข้อสอบที่สูญหายแบบการไม่แสดงคำตอบ (No presented: NP) ทางเลือกที่เหมาะสมจะใช้ใน BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 2003) ในกรณีนี้การตอบสนองข้อสอบที่ถูกทำให้สูญหายสำหรับผู้สอบซึ่งไม่รวมอยู่ในการประมาณค่าของพารามิเตอร์ข้อสอบ ประหนึ่งว่า ผู้ตอบไม่เคยให้โอกาสในการตอบสนองข้อสอบ วิธีข้อมูลสูญหาย

ลำดับที่สองที่เหมาะจะใช้สำหรับผู้ใช้ ของ BILOG-MG เป็นการจัดกระทำการตอบสนองข้อสอบสูญหายแบบตอบผิด (Incorrect: IN) ในทางตรงกันข้าม ไม่ว่าการตอบสนองข้อสอบสูญหายใดที่ถูกจัดกระทำเป็นแบบตอบผิดในการให้คะแนนของแบบสอบและการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ ลำดับสุดท้าย วิธีการที่ไม่อิงการประมาณค่าทดแทนสำหรับจัดการกับการตอบสนองข้อสอบสูญหายที่เป็นการจัดกระทำแบบตอบถูกบางส่วน (Fractionally correct: FR) เช่น เมื่อใช้โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) จะหมายความว่า ถ้ามี 5 ทางเลือกที่การละเว้นการตอบจะให้คะแนนแบบตอบถูกบางส่วนด้วยสัดส่วน $\frac{1}{5}$ (Zimowski et al., 2003)

สรุปการศึกษาในครั้งนี้ ผู้วิจัยเลือกวิธีในการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) (Schafer & Graham, 2002; Van Buuren & Groothuis-Oudshoorn, 2011) เพราะเป็นเทคนิคที่สามารถคำนวณโดยใช้ใช้โปรแกรม R เนื่องจากเป็นโปรแกรมคำนวณทางสถิติที่อนุญาตให้ผู้สนใจใช้ได้โดยไม่ละเมิดยังมี Package สำเร็จรูปสำหรับประมาณค่าทดแทน ซึ่งพัฒนาขึ้นโดยผู้เชี่ยวชาญที่มีความรู้และเชี่ยวชาญเกี่ยวกับสถิติและการจำลองข้อมูล (Honaker, King, & Blackwell, 2011; Rizopoulos, 2006; Van Buuren & Groothuis-Oudshoorn, 2011) คือ Package ‘MICE’ ซึ่งเป็น Package สำเร็จรูป ในการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าทดแทนพหุ มีขั้นตอนการคำนวณ 3 ขั้นตอนหลัก ดังนี้ การแทนค่าทดแทนแต่ละชุดข้อมูล (Imputation) การประมาณค่า (Estimation) และการรวมผลการประมาณค่าทดแทนข้อมูลสูญหายแต่ละชุดข้อมูลเป็นค่าเดียว (Pooling)

ขั้นตอน 1 การแทนค่าทดแทนแต่ละชุดข้อมูล (Imputation) คือขั้นตอนการแทนค่าของชุดข้อมูลที่มีค่าสูญหาย โดยการสร้างชุดข้อมูลที่ถูกแทนค่าขึ้นมาหลาย ๆ ชุด จำนวน M ชุด จากภาพที่ 1-1 การประมาณค่าทดแทนด้วยวิธี MI จากการศึกษาที่ผ่านมาพบว่า จำนวนครั้งในการประมาณค่าซ้ำสัมพันธ์กับประสิทธิภาพในการประมาณค่า ซึ่งกรณีข้อมูลสูญหายน้อยกว่า 30% จำนวนการประมาณค่าทดแทนซ้ำ 3-5 ครั้ง สามารถให้ผลการประมาณค่าทดแทนที่มีประสิทธิภาพเพียงพอ (Bodner, 2008; Graham, 2009) แต่ทั้งนี้ Van Buuren and Groothuis-Oudshoorn (2011 อ้างถึงใน กมลทิพย์ ศรีหาเศษ, 2555) เสนอแนะว่าจำนวนการประมาณค่าทดแทนซ้ำขั้นต่ำที่มีประสิทธิภาพในการประมาณค่าทดแทนพหุ คือ 10 ครั้ง ขึ้นไป ดังนั้นเพื่อให้การวิเคราะห์ข้อมูลที่ถูกต้องและเกิดความคลาดเลื่อนน้อยที่สุดในการศึกษาในครั้งนี้ ผู้วิจัยจึงประมาณค่าทดแทนซ้ำด้วยวิธี MI จำนวน 10 ครั้ง (M = 10)

ขั้นตอน 2 การประมาณค่า (Estimation) คือขั้นตอนการวิเคราะห์ข้อมูลตามปกติด้วยโมเดลที่เหมาะสมกับข้อมูล

ขั้นตอน 3 การรวมผลการประมาณค่าทดแทนข้อมูลสูญหายแต่ละชุดข้อมูลเป็นค่าเดียว (Pooling) คือ ขั้นตอนการสรุปผลการวิเคราะห์ โดยการรวมผลการวิเคราะห์ของแต่ละชุดข้อมูลเพื่อออกมาเป็นผลลัพธ์สุดท้าย (Final resul/ pooled results) ขั้นตอนการประมาณค่าและการรวม อาจเรียกรวมกันว่าขั้นตอนการวิเคราะห์ (Analysis)

ตอนที่ 2 แนวคิด เกี่ยวกับโมเดลทฤษฎีการตอบสนองข้อสอบ

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวกับทฤษฎีการตอบสนองข้อสอบและการประมาณค่าพารามิเตอร์ของข้อสอบนั้น ผู้วิจัยได้นำเสนอ ดังนี้

การวัดข้อสอบมีความเชื่อในความไม่แปรเปลี่ยนถ้ากลุ่มผู้สอบแตกต่างกัน ถ้ามีผู้ตอบมีระดับตัวแปรแฝงภายใน (θ) เท่ากันแล้ว ความน่าจะเป็นของการเห็นด้วยในข้อคำถามควรเท่ากัน (Andrich, 1988) ในทฤษฎีการตอบสนองข้อสอบ (Item response theory: IRT) ได้เสนอวิธีการสำหรับเปรียบเทียบการวัดข้อสอบข้ามกลุ่มผู้สอบ โดยโมเดลการตอบสนองข้อสอบเป็นฟังก์ชันของระดับความสามารถของผู้สอบและคุณลักษณะของข้อสอบในแบบสอบ (Andrich, 1988; Embretson & Reise, 2000) และยอมให้ข้อสอบที่ใช้ในการสอบมีความแตกต่างของฟังก์ชันข้อสอบและความไม่เท่าเทียมกันในการวัด (Embertson & Reise, 2000) ทฤษฎีการตอบสนองข้อสอบถูกนำไปประยุกต์ใช้อย่างกว้างขวางในงานวิจัยทางการศึกษาและจิตวิทยา และในปัจจุบันยังมีการนำไปประยุกต์ใช้ในสาขาสังคมวิทยาด้วย (Ghuman, Lee, & Smith, 2004; MacIntosh,1998; Smith & Furstenberg, 1994; Smith & Morgan,1994) โดยโมเดลทฤษฎีการตอบสนองข้อสอบ เป็นโมเดลทางคณิตศาสตร์ที่มีความสัมพันธ์กับความน่าจะเป็นของการตอบที่แสดงถึงความสามารถของผู้สอบและคุณลักษณะของข้อสอบ มาตรวัดทั่วไปมีจุดกึ่งกลางการตัดสินใจคือศูนย์ หน่วยของการวัดของค่าในหนึ่งหน่วยอยู่ในช่วง -4 ถึง +4 ซึ่งเป็นคะแนนที่เรียกว่า z score (Camilli & Shepard, 1994) แต่ละค่าของตัวแปรแฝงในมาตรวัดการวัดเป็นความน่าจะเป็นในการตอบข้อสอบของผู้สอบซึ่งเป็น การตรวจให้คะแนนแบบ 2 ค่า (Dichotomous) นั้น ความน่าจะเป็นจะต่ำในกลุ่มผู้สอบที่มีระดับความสามารถต่ำ และจะสูงสำหรับกลุ่มผู้สอบที่มีความสามารถสูง ซึ่งฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบ

จะเป็นโค้งความน่าจะเป็นที่มีรูปร่างแบบ S-shaped แต่มีข้อยกเว้นสำหรับข้อสอบที่มีค่าอำนาจจำแนกต่ำ ซึ่งโค้งนี้เป็นที่รู้จักกันดีในชื่อของโค้งลักษณะข้อสอบ (Item characteristic curve: ICC) และเป็นโค้งลักษณะเฉพาะของข้อสอบ (Baker, 2001; Embertson & Reise, 2000) โค้งลักษณะข้อสอบจะมี 2 คุณลักษณะคือ ความยากและอำนาจจำแนก ซึ่งความยากจะอธิบายถึงตำแหน่งการวัดของข้อสอบแต่ละข้อในมาตรวัดระดับความสามารถของผู้สอบ จุดเปลี่ยนโค้งจะตรงกับค่าความน่าจะเป็น 0.5 ซึ่งผู้สอบมีโอกาสตอบถูกหรือผิดเท่ากัน จุดโค้งของข้อสอบที่ง่ายจะเกิดขึ้นได้น้อยกว่าข้อสอบที่ยากในแต่ละระดับความสามารถของผู้สอบ นั้นคือโค้งลักษณะข้อสอบของข้อสอบที่ยากจะต่ำกว่าโค้งลักษณะของข้อสอบข้อที่ง่ายกว่า คุณลักษณะที่สองของอำนาจจำแนกของข้อสอบ คือ การวัดความชันของเส้นโค้ง และข้อสอบจะจำแนกผู้สอบที่มีความสามารถสูงกว่าจุดเปลี่ยนโค้งกับผู้สอบที่ระดับความสามารถต่ำกว่าจุดเปลี่ยนโค้ง อำนาจจำแนกยังเป็นหน่วยการวัดความเที่ยง ข้อสอบที่มีอำนาจจำแนกสูงจะมีเส้นโค้งชันกว่าและความเที่ยงสูงกว่าเมื่อเปรียบเทียบกับข้อสอบที่ง่ายกว่า นั่นคือความน่าจะเป็นในการตอบจะเพิ่มขึ้นตามการเพิ่มระดับความสามารถของผู้สอบ (Tfaily, 2006) ดัชนีอำนาจจำแนกของข้อสอบเป็นตัวแปรอิสระของระดับความยากของข้อสอบข้อนั้น ๆ (Baker, 2001; Embretson & Reise, 2000) มาตรวัดความชันเป็นส่วนกลับของมาตรวัดความยาก และด้วยเหตุนี้ ในการศึกษาโดยทั่วไปการประมาณค่าสองพารามิเตอร์ จึงไม่ควรเปรียบเทียบโดยตรง (Camilli & Shepard, 1994) นอกจากนี้ คุณลักษณะที่สำคัญของทฤษฎีการตอบสนองข้อสอบอีกประการคือ การประมาณค่าของพารามิเตอร์ข้อสอบ ซึ่งเป็นตัวแปรอิสระของการแจกแจงของความสามารถของกลุ่มตัวอย่าง (Bond & Fox, 2001)

ทฤษฎีการวัดของโมเดลทฤษฎีการตอบสนองข้อสอบนั้น Thissen and Orlando (2001) ได้อภิปรายถึงสองแนวคิดของการสร้างโมเดลในทฤษฎีการตอบสนองข้อสอบ ดังนี้

แนวคิดแรกเป็นการพัฒนาเป็นโมเดลความสอดคล้องพอดี (Well-fitting model) ในการสะท้อนข้อมูลของการตอบสนองข้อสอบด้วยพารามิเตอร์ความสามารถหรือปัจจัยที่สนใจ รวมทั้งคุณลักษณะของข้อสอบ เป้าหมายของแนวคิดนี้คือ การวิเคราะห์ข้อสอบ ซึ่งโมเดลควรสะท้อนถึงคุณลักษณะของข้อมูลการตอบสนองข้อสอบด้านความพอเพียงและความถูกต้องแม่นยำ ดังนั้น พฤติกรรมของข้อสอบจึงเป็นผลรวมของพารามิเตอร์ข้อสอบ ปรัชญาของข้อสอบถูกกำหนดโดยการวัดของข้อสอบ ไม่ใช่สิ่งที่วัดควรจะเป็น ซึ่งแนวคิดนี้การสร้างโมเดลมีความเชื่อว่าทฤษฎีของการวัดจะนำไปสู่การอธิบายข้อมูล

ส่วนอีกแนวคิดหนึ่งของการสร้างทฤษฎีการตอบสนองของข้อสอบ เป็นการได้มาซึ่งคุณลักษณะเฉพาะของการวัด ซึ่งนิยามโดยโมเดลของข้อมูลการตอบสนองข้อสอบต้องสอดคล้องกัน ถ้าข้อสอบหรือผู้สอบไม่มีความสอดคล้องภายในของคุณลักษณะในการวัดของโมเดลทฤษฎีการตอบสนองข้อสอบ การประเมินผลโดยการวิเคราะห์ส่วนที่เหลือ เช่น ความสอดคล้องทางสถิติของข้อสอบและผู้สอบ ข้อสอบหรือผู้สอบจะถูกปฏิเสธ แนวคิดนี้ดำเนินตามแนวคิดโมเดลของ Rasch (1960) และในกรณีที่ข้อมูลสอดคล้องกับโมเดล จะนำเสนอการอธิบายอย่างง่ายสำหรับการวิเคราะห์ข้อสอบและมาตรการให้คะแนน แนวคิดนี้ การสร้างโมเดลมีความเชื่อว่าการวัดที่ดีที่สุดเป็นการนิยามทางคณิตศาสตร์ ซึ่งนักวิจัยในกลุ่มนี้เชื่อว่าโมเดลในตระกูลของ Rasch เป็นตัวเลือกที่เหมาะสมที่สุดของโมเดลทฤษฎีการตอบสนองข้อสอบ เนื่องจากยังคงคุณลักษณะทางคณิตสตร์ไว้อย่างเหนียวแน่น โดยตัวอย่างการศึกษามีวัตถุประสงค์เฉพาะ เช่น การแยกการประมาณค่าพารามิเตอร์ผู้สอบและพารามิเตอร์ข้อสอบ และรวมคะแนนจำนวนที่พอเพียงอย่างง่ายที่ไม่ใช่สารสนเทศจากรูปแบบการตอบสนองที่ต้องการ

ทั้งนี้ Embretson and Rasch (2000) ได้เสนอแนะเกี่ยวกับการใช้โมเดล Rasch ว่าควรใช้โมเดลตระกูลของ Rasch เมื่อข้อสอบแต่ละข้อได้รับการถ่วงน้ำหนักเท่ากัน นั่นคือ ข้อสอบแต่ละข้อมีความสำคัญเท่ากัน ในการนิยามตัวแปรแฝงภายใน และเมื่อคุณลักษณะของโมเดลการวัดมีจุดแข็ง นั่นคือ มีวัตถุประสงค์เฉพาะหรือมีความพอเพียงอย่างง่าย ซึ่งเป็นสิ่งที่ควรจะมี ถ้าหนึ่งในความสอดคล้องที่ต้องการของโมเดลทฤษฎีการตอบสนองข้อสอบเป็นการคงอยู่ของข้อมูลหรือการประมาณค่าที่มีความถูกต้องแม่นยำสูงแล้ว โมเดลที่มีความสมบูรณ์มากกว่าจะถูกนำมาใช้

นอกจากนี้ โมเดลที่สัมพันธ์กับทฤษฎีการตอบสนองข้อสอบ โมเดลการตอบสนองข้อสอบแบบโลจิสติก (Logist model) ที่สำคัญมี 3 โมเดล คือ โมเดลการตอบสนองข้อสอบแบบโลจิสตริก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) โมเดลการตอบสนองข้อสอบแบบโลจิสตริส ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) และโมเดลการตอบสนองข้อสอบสนองข้อสอบแบบโลจริสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) (ศิริชัย กาญจนวาสี, 2550; Baker, 1992; Hambleton & Swaminathan, 1985; Reeve, 2003) ซึ่งมีรายละเอียด ดังนี้

1. โมเดลการตอบสนองข้อสอบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL)

การพัฒนาของ Rasch model เป็นตัวแปรเริ่มต้นของการพัฒนาโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) แต่ทั้ง Rasch model และโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) มีลักษณะที่คล้ายกันทางคณิตศาสตร์ โดยโค้งคุณลักษณะของโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) สำหรับข้อสอบข้อที่ i (ศิริชัย กาญจนวาสี, 2550) เขียนสมการได้ ดังนี้

$$P_i(\theta)=\frac{1}{1+e^{(\theta \quad b_i)}}$$

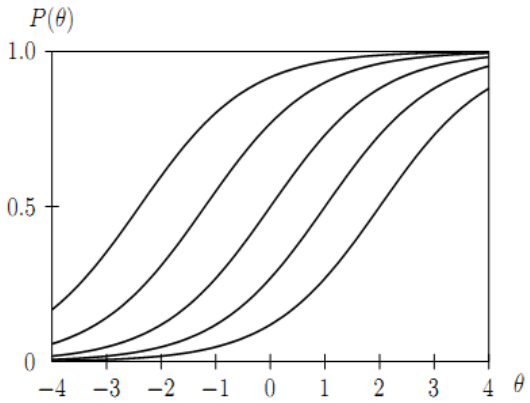
เมื่อ

$P_i(\theta)$	แทน	ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง
b_i	แทน	ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของของโค้งลักษณะข้อสอบ ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.05
e	แทน	ค่าคงที่ของลอการิทึมธรรมชาติ (Natural log) มีค่าเท่ากับ 2.718

ใน Rasch model กำหนดความชันที่แน่นอนสำหรับข้อสอบทั้งหมด และโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) ต้องการเพียงความชันที่มีค่าเท่ากันสำหรับข้อสอบทั้งหมด (Thissen & Orlando, 2001)

การแจกแจงของประชากรของประชากรของตัวแปรความสามารถ (θ) สำหรับโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) เป็นลักษณะเฉพาะทั่วไปที่มีค่าเฉลี่ยของประชากรเป็นศูนย์และความแปรปรวนเป็นหนึ่ง พารามิเตอร์ความยาก (b_i) มีความสัมพันธ์เป็นศูนย์ ซึ่งเป็นค่าเฉลี่ยระดับความสามารถในประชากรและความชัน พารามิเตอร์ a นำค่าความสัมพันธ์ไปยังหน่วยของส่วนเบี่ยงเบนมาตรฐานของ θ ดังนั้น โมเดลของตัวแปรความสามารถ (θ) จึงกำหนดให้มีการแจกแจงปกติ ไม่ใช่การตอบสนองข้อสอบที่ไม่ต่อเนื่อง (Thissen & Orlando, 2001; Thissen & Steinberg,

1988) ซึ่งโค้งลักษณะข้อสอบของโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) แสดงได้ดังภาพ



ภาพที่ 2-1 โค้งลักษณะข้อสอบแบบโมเดลการตอบสนองข้อสอบแบบโลจิสติกชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) ของตัวอย่างข้อสอบ 5 ข้อ
(Partchev, 2004, p. 14)

2. โมเดลการตอบสนองข้อสอบแบบโลจิสติกชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL)

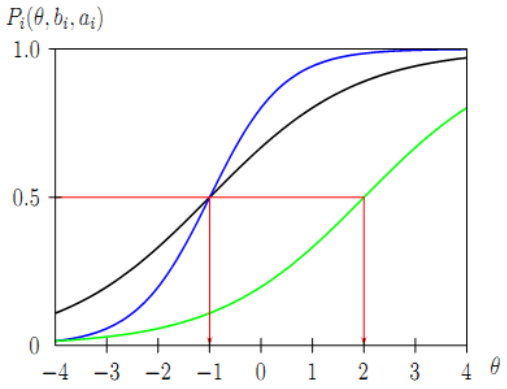
Birnbaum (1968) กล่าวว่า โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) จะพิจารณาความชันหรือพารามิเตอร์อำนาจจำแนก (a) ที่แปรเปลี่ยนข้ามข้อสอบ แทนที่จะถูกจำกัดให้มีค่าเท่ากันดังใน โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) หรือ Rasch model ความสัมพันธ์ที่สำคัญของความแตกต่างระหว่างระดับความสามารถของผู้สอบและความยากของข้อสอบถูกกำหนดด้วยขนาดของอำนาจจำแนกของข้อสอบ (Emberatson & Rasch, 2000) โค้งคุณลักษณะข้อสอบของ โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) สำหรับความน่าจะเป็นของโมเดลการตอบข้อสอบข้อที่ i ถูก สำหรับผู้สอบที่มีระดับความสามารถ θ (ศิริชัย กาญจนวาสี, 2550) เขียนสมการได้ ดังนี้

$$P_i(\theta) = \frac{1}{1+e^{-Da_i(\theta-b_i)}}$$

เมื่อ

$P_i(\theta)$	แทน	ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง
b_i	แทน	ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของโค้งลักษณะข้อสอบ ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50
a_i	แทน	ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าความชันของโค้งลักษณะของข้อสอบ ณ ตำแหน่ง $b_{i\,i}$
e	แทน	ค่าคงที่ของลอการิทึมธรรมชาติ (Natural log) มีค่าเท่ากับ 2.718
D	แทน	ค่าองค์ประกอบของการปรับสเกล (Scaling factor) เพื่อทำให้ Logistic fuction กับ Nature ogive function ใกล้เคียงกัน มีค่าเท่ากับ 1.70

ในโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) ค่าคงที่ D ซึ่งเท่ากับ 1.70 จะถูกเพิ่มเข้าไปในโมเดล โดยเป็นการจัด โมเดลการตอบสนองข้อสอบแบบโลจิสติก (Logistic model) ให้ใกล้เคียงกับโมเดลปกติสะสม (Normal ogive model) จากการศึกษาเอกสารที่เกี่ยวข้องโดย Thissen and Steinberg (1988) พบว่าครึ่งหนึ่งจากเอกสารที่เกี่ยวข้องทั้งหมดเป็นการศึกษาเกี่ยวกับวิธีการประมาณค่า ในขณะที่อีกครึ่งหนึ่งเป็นการศึกษาในวิธีการอื่น ๆ ซึ่งโค้งลักษณะข้อสอบของ โมเดลการตอบสนองข้อสอบแบบโลจิสติกชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) แสดงได้ดังภาพ



ภาพที่ 2-2 โค้งลักษณะข้อสอบแบบโมเดลการตอบสนองข้อสอบแบบโลจิสติกชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) ของตัวอย่างข้อสอบ 3 ข้อ (Partchev, 2004, p. 25)

3. โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL)

Lord (1980) กล่าวว่าโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) ได้รับการพัฒนาในการสอบทางการศึกษาที่ขยายผลการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบไปยังแบบสอบหลายตัวเลือก ซึ่งอาจนำมาซึ่งการเดา โดยสำหรับข้อสอบข้อที่ i โค้งคุณลักษณะข้อสอบของโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) (ศิริชัย กาญจนวาสี, 2550) เขียนสมการได้ ดังนี้

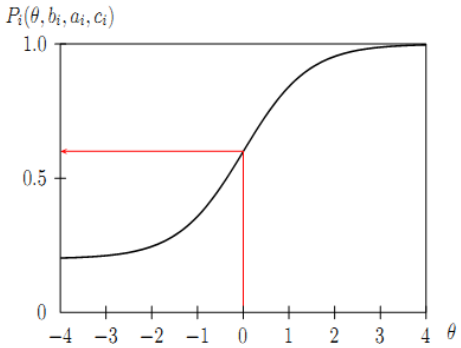
$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta - b_i)}}$$

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบ ข้อที่ i ได้ถูกต้อง

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของโค้งลักษณะข้อสอบ ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1 + c_i}{2}$

a_i	แทน	ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าความชันของโค้งลักษณะข้อสอบ ณ ตำแหน่ง b_i
c_i	แทน	ค่าพารามิเตอร์โอกาสในการเดา
e	แทน	ค่าคงที่ของลอการิทึมธรรมชาติ (Natural log) มีค่าเท่ากับ 2.718
D	แทน	ค่าองค์ประกอบของการปรับสเกล (Scaling factor) เพื่อทำให้ Logistic fuction กับ Nature ogive function ใกล้เคียงกัน มีค่าเท่ากับ 1.70

การอธิบายพารามิเตอร์โอกาสในการเดาสำหรับแบบสอบหลายตัวเลือกในการวัดผลทางการศึกษามีการศึกษาอย่างต่อเนื่อง อย่างไรก็ตาม พารามิเตอร์โอกาสในการเดาอาจให้สารสนเทศซึ่งเข้าใจอย่างถ่องแท้ในการเข้าใจพฤติกรรมของผู้สอบจากข้อสอบ สิ่งที่ควรพิจารณาของผู้สอบในการศึกษาว่า การที่ผู้สอบตอบสนองข้อสอบถูกนั้น อาจเป็นผลมาจากตัวแปรอื่นในการวัดการทดสอบความสามารถในบริบทของการศึกษา โดยทั่วไปกำหนดว่าผู้สอบทำให้ระดับความสามารถสูงขึ้นได้จากการเดาคำตอบได้ถูกต้อง (Reeve, 2003) ซึ่งโค้งลักษณะข้อสอบของ โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) แสดงได้ดังภาพ



ภาพที่ 2-3 โค้งลักษณะข้อสอบแบบโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) ของตัวอย่างข้อสอบ 1 ข้อ (Partchev, 2004, p. 34)

จากที่กล่าวมา จะพบว่า ข้อสอบในโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 1 พารามิเตอร์ (One-parameter logistic model: 1PL) ซึ่งบางครั้งเรียกว่า Rasch model กำหนดว่ามีอำนาจจำแนกเท่ากัน และมีเพียงพารามิเตอร์ความยากเท่านั้นที่เปลี่ยนแปลง ผู้สอบทุกคนที่มีคะแนนการตอบถูกเท่ากันจะถูกประมาณค่าว่ามีระดับความสามารถเดียวกัน แต่กรณีนี้จะไม่จริงในกรณีของโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) หรือ โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) ซึ่งในส่วนของโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) จะพิจารณาการเดาในการตอบข้อสอบด้วย

ตอนที่ 3 แนวคิด ทฤษฎีเกี่ยวกับการประมาณค่าพารามิเตอร์

การวิเคราะห์ข้อสอบมีจุดมุ่งหมายหลักเพื่อวิเคราะห์หาสารสนเทศของการทดสอบ ซึ่งได้แก่ สารสนเทศของข้อสอบและสารสนเทศของผู้สอบ การหาสารสนเทศดังกล่าวจะต้องมีการประมาณค่าพารามิเตอร์ ซึ่งได้แก่ ระดับความสามารถของผู้สอบ และการประมาณค่าพารามิเตอร์ของข้อสอบ ซึ่งได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าโอกาสการเดาคำตอบของข้อสอบนั้นถูก ดังนั้นการประมาณค่าพารามิเตอร์สำหรับการวิเคราะห์ข้อสอบจึงมีความสำคัญและมีความจำเป็นสำหรับการวิเคราะห์ข้อสอบ โดยในขั้นตอนนี้จะเป็นการนำเสนอวิธีการประมาณค่าพารามิเตอร์ วิธีแม็กซิมั่มไลค์ลิฮูด (Maximum likelihood) (Hambleton & Cook, 1977; Loard, 1980; รัตนา ศรีเหรัญ, 2539; ศิริชัย กาญจนวาศรี, 2550; สุนทร เทียนงาม, 2551 อ้างถึงใน ชนะศึก นิชานนท์, 2553, หน้า 89) ซึ่งมีรายละเอียด ดังนี้

วิธีแม็กซิมั่มไลค์ลิฮูด (Maximum likelihood estimation: ML)

การประมาณค่าพารามิเตอร์ด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุด เป็นการประมาณค่าพารามิเตอร์โดยอาศัยผลที่ได้จากตัวอย่างที่สุ่มเลือกมาจากการแจกแจงที่ทราบรูปแบบของฟังก์ชันความหนาแน่นแต่ไม่ทราบค่าพารามิเตอร์ ดังนั้น จึงใช้หลักของความน่าจะเป็นในการเลือกตัวอย่าง และวัดค่าได้จากกลุ่มตัวอย่างที่ถูกเลือก ($U_1 = u_1, U_2 = u_2, ..., U_n = u_n$) มาพิจารณาค่าประมาณของค่าพารามิเตอร์ที่ต้องการ

จากการนิยามความน่าจะเป็น ของผู้เข้าสอบในการตอบข้อสอบข้อที่ i เมื่อ $U_{ij} = 1$ สำหรับการตอบถูก และ $U_{ij} = 0$ สำหรับการตอบผิด ซึ่งสามารถแสดงสมการได้ ดังนี้

$$P(U_{ij}\left|\theta,b,a,c\right.) = P(U_{ij}=1\left|\theta,b,a,c\right.)\ P(U_{ij}=0\left|\theta,b,a,c\right.)$$

$$= P_{ij}^{u_{ij}}\left(1-P_{ij1}\right)^{1-u_{ij}}$$

$$= P_{ij}^{u_{ij}}Q^{1-u_{ij}}Q_{ij}$$

$$= 1-P_{ij}$$

ถ้าผู้เข้าสอบตอบข้อสอบ n ข้อ และแบบทดสอบมีลักษณะการวัดเพียงมิติเดียว (คือมีความเป็นอิสระเฉพาะที่) แล้วความน่าจะเป็นของการตอบแสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม (Joint density function) ซึ่งสามารถแสดงสมการได้ ดังนี้

$$P(U:\theta,a,b) = \prod_{i=1}^{N}\prod_{j=1}^{n}P_{ij}^{U_{ij}}Q_{ij}^{1-U_{ij}}$$

เมื่อ

$P(U:\theta,a,b)$ = ความน่าจะเป็นในการตอบข้อสอบ n ข้อ ของผู้เข้าสอบ N คน

U = เวคเตอร์ (Vector) ของตัวแปรที่แสดงผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบ N คน

θ = เวคเตอร์ (Vector) ความสามารถของผู้เข้าสอบ N คน

a = เวคเตอร์ (Vector) ค่าอำนาจจำแนกของข้อสอบ n ข้อ

b = เวคเตอร์ (Vector) ค่าความยากของข้อสอบ n ข้อ

สมการดังกล่าวเป็นความน่าจะเป็นของการตอบข้อสอบ n ข้อ ที่สามารถวัดหรือสังเกตได้โดยที่ $U_1,U_2,...,U_n$ เป็นตัวแปรสุ่มที่มีค่าเฉพาะเป็น $U_1,U_2,...,U_n$ เมื่อ U_1 มีค่าเท่ากับ 1 หรือ 0 และเนื่องจากสมการนี้เป็นฟังก์ชันทางคณิตศาสตร์ของค่า θ, b, a, c ที่จะบอกว่าตัวแปรสุ่มนี้มีโอกาสเกิดขึ้นเมื่อใด จึงเรียกฟังก์ชันนี้ว่าฟังก์ชันความน่าจะเป็นหรือ ฟังก์ชันไลค์ลิฮูด (Likelihood function) ซึ่งแสดงได้ดังสมการ

$$L(U_1,U_2,\ldots,U_n\mid\theta,b,a,c) \quad = \quad P_j^{u_j} Q_j^{1-u_j}$$

เมื่อ $U_j = 1$ ค่าของ Q_j ก็หมดไป เพราะ $Q_j^{1-u_j} = 1$ และเมื่อ $U_j = 0$ ค่าของ P_j จะหมดไป เพราะ $P_j^{u_j} = 1$ และฟังก์ชันไลค์ลิฮูดที่มีผู้สอบ N คน ตอบข้อสอบ n ข้อ มีสมการ ดังนี้

$$L(U\mid\theta,b,a,c) = L(U_1,U_2,\ldots,U_n\mid\theta_1,\theta_2,\ldots,\theta_n,$$

$$b_1,b_2,\ldots,b_n,a_1,a_2,\ldots,a_n,c_1,c_2,\ldots,c_n)$$

$$L(U\mid\theta,b,a,c) = \prod_{i=1}^{N}\prod_{j=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

เมื่อ U = เวคเตอร์ผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบ N คน

U_i = เวคเตอร์ผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบคนที่ i

P_{ij} = $P_{ij}(\theta_i,b_1,a_1,c_1)$

= ฟังก์ชันไลค์ลิฮูดของผลการสอบ

การประมาณค่าพารามิเตอร์ด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุด คือ การหาค่าประมาณของ ค่าพารามิเตอร์ที่จะทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด ซึ่งโดยปกติจะทำให้การหาค่าพารามิเตอร์ที่ทำให้ลอการิทึม (Logarithm) ของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด ทั้งนี้เพราะค่าประมาณของพารามิเตอร์ที่ทำให้ฟังก์ชันทั้งสองฟังก์ชันมีค่าสูงสุดเป็นค่าเดียวกัน แต่การหาค่าประมาณที่ทำให้ลอการิทึมของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดนั้น ทำได้ง่ายกว่า ซึ่งลอการิทึมของฟังก์ชันไลค์ลิฮูดของตัวแปร θ , b, a, c คือ

$$\ln(U\mid\theta,b,a,c) \quad = \quad \sum_{i=1}^{N}\sum_{j=1}^{n}(U_{ij}\ln P_{ij} + (1-U_{ij})\ln Q_{ij})$$

และการหาค่าพารามิเตอร์ θ,b,a,c ที่ทำให้ $\ln(U\mid\theta,b,a,c)$ มีค่าสูงสุดทำได้โดยกำหนดให้อนุพันธ์ของ $\ln(U\mid\theta,b,a,c)$ มีค่าเท่ากับศูนย์ แล้วแก้สมการหาค่ารากของอนุพันธ์ ดังนี้

$$d\ln(U|\theta,b,a,c)/d\theta_i = 0 \quad \text{เมื่อ } i = 1,2,...,N$$

$$d\ln(U|\theta,b,a,c)/db_i = 0 \quad \text{เมื่อ } i = 1,2,...,N$$

$$d\ln(U|\theta,b,a,c)/da_i = 0 \quad \text{เมื่อ } i = 1,2,...,N$$

$$d\ln(U|\theta,b,a,c)/dc_i = 0 \quad \text{เมื่อ } i = 1,2,...,N$$

โดยทั่วไปสมการไลค์ลิฮูดนี้ไม่เป็นสมการเส้นตรง ดังนั้น การหาค่า θ_i,a_j,b_j,c_j ที่ทำให้ $\ln(U|\theta,b,a,c)$ มีค่าสูงสุดจึงไม่สามารถหาได้ด้วยการใช้วิธีการอย่างง่าย แต่หาได้โดยใช้วิธีของนิวตัน ราฟสัน (Newton-raphson procedure) (Lord, 1980) ซึ่งเป็นการหาค่าประมาณโดยการประมาณค่าซ้ำ (Iterative) จนค่าที่ได้มีค่าคงที่ ซึ่งมีขั้นตอนดังต่อไปนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i และค่าพารามิเตอร์ a_j,b_j และ c_i

1. ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i คือ

$$\theta_i^{(0)} = \ln\left(\frac{x_i}{n-x_i}\right)$$

เมื่อ

ln = Natural Logarithm

X_i = คะแนนสอบของคนที่ i

n = จำนวนข้อสอบ

2. ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ a_j,b_j และ c_j อาจใช้ค่าประมาณจากวิธีฮิวริสติค หรือค่าอื่น ๆ ที่เหมาะสม

ขั้นที่ 2 ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคน θ_i โดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่ ดังนี้

$$\theta_i^{(m+1)} = \theta_i^{(m)} - \frac{g(\theta_i^{(m)})}{h(\theta_i^{(m)})}$$

เมื่อ

$\theta_i^{(m)}$, $\theta_i^{(m+1)}$ = ค่าประมาณความสามารถของผู้เข้าสอบคนที่ i ครั้งที่ m และครั้งที่ $m+1$

$g(\theta_i^{(m)})$ = $d \ln L(U|\theta,b,a,c)/d\theta_i$; ประมาณที่ $\theta_i^{(m)}$

$h(\theta_i^{(m)})$ = $d^2 \ln L(U|\theta,b,a,c)/d\theta_i$; ประมาณที่ $\theta_i^{(m)}$

การประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณความสามารถ (θ_i) จะมีค่าเข้าใกล้ค่าใดค่าหนึ่ง (Convergence) คือ ค่าประมาณครั้งที่ $m+1$ และครั้งที่ m มีค่าแตกต่างกันน้อยกว่าค่าที่กำหนดไว้ เช่น 0.001

ขั้นที่ 3 ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ a_j, b_j และ c_j โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่

$$x_j^{(m+1)} = x_j^{(m)} - \frac{g(x_j^{(m)})}{h(x_j^{(m)})}$$

เมื่อ

x_j = เวคเตอร์ของค่า a, b และ c ของข้อสอบที่ j

$x_j^{(m)}$, $x_i^{(m+1)}$ = ค่าประมาณพารามิเตอร์ของข้อสอบที่ j ครั้งที่ m และครั้งที่ $m+1$

$g(x_j^{(m)})$ = $d \ln L(U|\theta,b,a,c)/d\theta_i$ ของค่าที่ $x_j^{(m)}$

$h(x_j^{(m)})$ = $d^2 \ln L(U|\theta,b,a,c)/d\theta_i$ ของค่าที่ $x_j^{(m)}$

ในกระบวนการประมาณค่า จะมีการประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณจะลู่เข้าใกล้ค่าใดค่าหนึ่ง (Convergence)

ขั้นที่ 4 ประมาณค่าซ้ำในขั้นที่ 2 และขั้นที่ 3 จนกว่าค่าประมาณ θ_i, a_j, b_j และ c_j จะมีค่าที่คงที่ และมีความถูกต้องเพียงพอหรือทำให้ inL(θ, a, b, c) มีค่าสูงที่สุด

ค่าอนุพันธ์อันดับที่ 1 และค่าอนุพันธ์อันดับที่ 2 ของค่าพารามิเตอร์ θ, a, b, c แสดงค่าสุดท้ายที่ได้จากการประมาณจะเรียกว่า ค่าที่ได้จากการประมาณโดยวิธีแมกซิมั่มไลค์ลิฮู้ด และเพื่อให้ได้ค่าคงที่เร็วขึ้นในการประมาณค่าซ้ำแต่ละครั้ง ค่าพารามิเตอร์ความสามารถของข้อสอบจะถูกปรับให้มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ซึ่งจะเป็นผลให้ค่า a, b และ c ต้องถูกปรับตามค่าพารามิเตอร์ความสามารถของผู้สอบไปด้วย (Hambleton & Swaminathan,1985)

ตารางที่ 2-1 อนุพันธ์อันดับที่ 1 และอนุพันธ์อันดับที่ 2 สำหรับฟังก์ชันไลค์ลิฮู้ดของค่าพารามิเตอร์ข้อสอบ (a, b, c) และความสามารถของผู้เข้าสอบ (θ) ในแบบจำลองโลจิสติค 3 พารามิเตอร์

อนุพันธ์ (Derivative)	**สัญลักษณ์แสดง** (Expression)
$dlnL/d\theta_i$	$D\sum_{j=1}^{N}\frac{a_j(P_{ij}-c_j)(u_{ij}-P_{ij})}{(1-c_j)P_{ij}}$
$dlnL/da_j$	$\frac{D}{(1-c_j)}\sum_{j=1}^{N}\frac{(\theta_i-b_j)(P_{ij}-c_j)(u_{ij}-P_{ij})}{P_{ij}}$
$dlnL/db_j$	$\frac{-Da_j}{(1-c_j)}\sum_{i=1}^{N}\frac{(P_{ij}-c_j)(u_{ij}-P_{ij})}{P_{ij}}$
$dlnL/db_j$	$\frac{-Da_j}{(1-c_j)}\sum_{i=1}^{N}\frac{(P_{ij}-c_j)(u_{ij}-P_{ij})}{P_{ij}}$
$dlnL/dc_j$	$\frac{1}{(1-c_j)}\sum_{i=1}^{N}\frac{(u_{ij}-P_{ij})}{P_{ij}}$

ตารางที่ 2-1 (ต่อ)

อนุพันธ์ (Derivative)	**สัญลักษณ์แสดง** (Expression)
$d^2 \ln L / d\theta_i$	$D^2 \sum_{j=1}^{n} \frac{a_j (P_{ij} - c_j) Q_{ij}}{(1-c_j)^2 P_{ij}} \left[\frac{u_{ij} c_j}{P_{ij}} - P_{ij} \right]$
$d^2 \ln L / da_j$	$\frac{D^2}{(1-c_j)} \sum_{i=1}^{N} \left[\frac{(\theta_i - b_j)^2 (P_{ij} - c_j) Q_{ij}}{P_{ij}} \left\{ \frac{u_{ij}}{P_{ij}} c_j - P_{ij} \right\} \right]$
$d^2 \ln L / dd_j$	$\frac{D^2 a_j}{(1-c_j)^2} \sum_{i=1}^{N} (P_{ij} - c_j) \frac{Q_{ij}}{P_{ij}} \left[\frac{u_{ij} c_j}{P_{ij}} - P_{ij} \right]$
$d^2 \ln L / dc_j$	$\frac{1}{(1-c_j)^2} \sum_{i=1}^{N} \left[\frac{u_{ij} - 1}{P_{ij}} \right] - \left[\frac{u_{ij} Q_{ij}}{P_{ij}} \right]$

เมื่อ

$$P_{ij} = c_j + \frac{(2 - c_j)}{\left(1 + \exp^{-Da_j(\theta_j - b_j)}\right)}$$

U_{ij} = ผลการตอบข้อที่ j ของคนที่ i

จากวิธีการประมาณค่าพารามิเตอร์ข้างต้นเป็นการประมาณค่าตามโมเดลการตอบสนองข้อสอบโลจิสติค 3 พารามิเตอร์ (3PL) ถ้าผู้วิเคราะห์ข้อสอบต้องการประมาณค่าแบบโมเดลการตอบสนองข้อสอบโลจิสติค 2 พารามิเตอร์ (2PL) ให้กำหนดค่า c_j เป็น 0 และถ้าต้องการประมาณค่าแบบโมเดลการตอบสนองข้อสอบโลจิสติค 3 พารามิเตอร์ (3PL) ให้กำหนดค่า c_j เป็น 0 และ a_j เป็น 1

ตารางที่ 2-2 ค่าฟังก์ชันสารสนเทศสำหรับค่าพารามิเตอร์ข้อสอบ และค่าความสามารถของผู้เข้าสอบในแบบจำลองโลจิสติค 3 พารามิเตอร์

พารามิเตอร์ (Parameter)	**ฟังก์ชันสารสนเทศ** (Information function)
θ_i	$D^2 \sum_{j=1}^{n} \frac{a_j^2 (P_{ij} - c_j) Q_{ij} (c_j - P_{ij})}{P_{ij}(1-c_j)^2}$
a_j	$\frac{D^2}{(1-c_j)^2} \sum_{i=1}^{N} (\theta_i - b_j)^2 (P_{ij} - c_j)^2 Q_{ij}$
b_j	$\frac{D^2 a_j^2}{(1-c_j)^2} \sum_{i=1}^{N} \frac{(\theta_i - b_j)(P_{ij} - c_j)^2 Q_{ij}}{P_{ij}}$
c_j	$\frac{D}{(1-c_j)^2} \sum_{i=1}^{N} \frac{(\theta_i - b_j)(P_{ij} - c_j) Q_{ij}}{P_{ij}}$

วิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทาง (Marginal maximum likelihood estimation procedure: MMLE)

การประมาณค่าพารามิเตอร์ด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทางนี้ เป็นวิธีการประมาณค่าโดยสามารถประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบได้พร้อมกัน ซึ่งการประมาณค่าด้วยวิธีการนี้จะไม่ได้อิงกับค่าพารามิเตอร์ของผู้สอบ นอกจากนี้ยังมีการแจกแจงความสามารถของผู้สอบ เพื่อให้ได้ฟังก์ชันความน่าจะเป็นของพารามิเตอร์ของข้อสอบ หลังจากนั้นจึงนำค่าที่ได้ไปใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ การประมาณค่าด้วยวิธีนี้สามารถใช้ได้กับแบบทดสอบที่มีทั้งจำนวนข้อมากและน้อยได้ วิธีการประมาณค่าเริ่มจากรูปแบบการตอบ ข้อสอบของผู้สอบโดยสามารถแทนได้ด้วยสมการดังนี้

$$X = (x_1, x_2, x_3, \ldots, x_n)$$

และสำหรับผู้ที่มีความสามารถ θ ค่าความเป็นไปได้สามารถแสดงได้ดังนี้

$$P(X|\theta) = \prod_{j=1}^{n} P_{ij}(\theta)^{x_j} [1-P(\theta)]^{1-x_j}$$

ค่าความน่าจะเป็นที่ได้เป็นความน่าจะเป็นที่มีเงื่อนไขว่าทราบค่าความสามารถของผู้สอบ (θ) ซึ่งจากความน่าจะเป็นของรูปแบบ X จากผู้ไม่ทราบค่าความสามารถของผู้สอบ (θ) ซึ่งสุ่มมาจากประชากรที่มีการกระจายของค่าระดับความสามารถ (θ) อยู่ในรูปฟังก์ชันความหนาแน่นต่อเนื่อง (Continuous density g (θ) ที่อยู่ในรูปของความน่าจะเป็นที่ไม่มีเงื่อนไข และสามารถแสดงได้ ดังนี้

$$P(X) = \int_{-\infty}^{+\infty} P(X|\theta)g(\theta)\, de$$

ค่านี้เรียกว่าความน่าจะเป็นโดยปราศจากเงื่อนไขของ X และเนื่องจากค่าความสามารถ θ ถูกอินทริเกรต (Integrate) ออกไป ค่านี้จึงเป็นฟังก์ชันของพารามิเตอร์ของข้อสอบเท่านั้น ในวิธีมาร์จินัลแมกซิมัมไลค์ลิฮูด ค่าของพารามิเตอร์ของข้อสอบจะพิจารณาใช้ค่าที่ทำให้สมการข้างล่างมีค่าสูงสุด

$$\log L_M = \sum_{l=1}^{S} r_l \log_\theta \bar{P}(X_l)$$

เมื่อ

$P(X_l)$ ประมาณได้จากสูตร Gaussian Guadrature

r_l คือ ความถี่ของรูปแบบการตอบข้อสอบรูปแบบ X_l จากจำนวนผู้สอบทั้งหมด (N) และ

S คือ จำนวนรูปแบบที่เป็นไปได้ทั้งหมดจากข้อมูลการตอบข้อสอบ

ในแบบจำลองโลจิสติค 3 พารามิเตอร์ ค่าสูงสุดของสมการดังกล่าวสำหรับ ข้อ j หาได้จากสมการไลค์ลิฮูด ดังนี้

$$\sum_{k=1}^{a} \frac{\bar{r}_{jk} - \bar{N}_k P_j(X_k)}{P_j(X_k)\left[1-P_j(X_k)\right]} \frac{dP_j(X_k)}{d\begin{bmatrix} c_j \\ a_j \\ b_j \end{bmatrix}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

เมื่อ

$$\bar{r}_{jk} = \sum_{l}^{S} r_l X_{jl} P(X_l / X_k) A(X_k) / \bar{P}(X_l)$$

$$\bar{N}_k = \sum_{l}^{S} r_l P(X_l / X_k) A(X_k) / \hat{P}(X_l)$$

โดยสมการดังกล่าวจะสามารถคำนวณค่าความคาดหวังภายหลังของจำนวนข้อสอบที่ถูกและจำนวนข้อสอบที่ทำที่จุด X_k (X_{jl}) คือ คะแนน 0, 1 สำหรับข้อสอบ j ในรูป แบบการตอบ i นอกจากนี้การประมาณค่าด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทาง สามารถสรุปเป็นสมการทางคณิตศาสตร์ได้ ดังนี้

Marginal Maximum Likelihood of Item Paramrter (β) :

$$\text{Maximize} L(a,b,c) = \prod_{a=1}^{N} \int_{-\infty}^{\infty} g(a) L(\theta_a; a,b,c) d\theta_a$$

ปัจจุบันในการคำนวณเพื่อประมาณค่าพารามิเตอร์ด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทาง สามารถทำได้โดยสะดวก เนื่องจากมิสเลวี่ และบ็อค (Mislevy & Bock,1990) ได้พัฒนาโปรแกรมคำสั่งชื่อ BILOG แต่ในการใช้โปรแกรมดังกล่าวต้องใช้กลุ่มผู้สอบที่มีขนาดใหญ่ เพื่อทำให้สามารถประมาณค่าพารามิเตอร์ของข้อสอบได้ อย่างคงที่ นอกจากนี้จำนวนข้อสอบสำหรับการประมาณค่ายังต้องมีจำนวนมาก สำหรับการประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้โมเดลการตอบสนองข้อสอบแบบ 3 พาราเตอร์ (3 PL) ต้องมีผู้สอบที่มีระดับความสามารถต่ำที่เพียงพอด้วยจึงจะสามารถทำการประมาณค่าพารามิเตอร์ค่าโอกาสในการเดาข้อสอบถูก (c) ได้อย่างมีความน่าเชื่อถือ นอกจากมีผลกระทบต่อค่าพารามิเตอร์โอกาสในการเดาข้อสอบถูก (c) แล้ยังอาจจะส่งผลกระทบถึงค่าพารามิเตอร์ความยากของข้อสอบ (b) และค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) อีกด้วย แต่ปัญหาดังกล่าวสามารถแก้ไขได้โดยการประมาณค่าพารามิเตอร์ด้วยวิธีการของเบย์ (Bayesian) ซึ่งจะมีการกำหนดการแจกแจงเบื้องต้นของค่าพารามิเตอร์ค่าโอกาสในการเดาข้อสอบถูก (c) ก่อนในโปรแกรม BILOG (ศิริชัย กาญจนวาสี, 2550) นอกจากนี้ยังมีโปรแกรมคอมพิวเตอร์บางโปรแกรม เช่น โประแกรม LOGIST สามารถทำการค้นหาค่าความสามารถและค่าพารามิเตอร์ของข้อสอบ (a, b, c) ที่จะทำให้ฟังก์ชันประมาณค่า ที่เป็นไปได้สูงสุด

วิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าร่วมกัน (Joint maximum likelihood estimation procedure: JMLE)

สำหรับการประมาณค่าด้วยวิธีประมาณค่าที่เป็นไปได้สูงสุด วิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าร่วมกัน (Joint maximum likelihood estimation procedure: JMLE) ในการประมาณค่าจะใช้ N แทนจำนวนผู้สอบ $P_i(\theta_a)$ แทนความน่าจะเป็นในการตอบข้อสอบถูกในข้อที่ i ณ ผู้สอบระดับความสามารถ $\theta_a, Q_i(\theta_a) = 1 - P_i(\theta_a), u_{ia}$ แทนคำตอบข้อที่ i (0, 1) และ $g(\theta_a)$ แทนการแจกแจงพารามิเตอร์โดยมีสูตรดังนี้ (Lord, 1984)

$$\text{MaximizeL}(\theta;a,b,c) = \prod_{a=1}^{N}\prod_{i=1}^{n}\left[P_i(\theta_a)\right]^{U_{ia}}\left[Q_i(\theta_a)\right]^{1-U_{ia}}$$

หรือ

$$\log L(\theta;a,b,c) = \prod_{a=1}^{N}\prod_{i=1}^{n}\left[U_{ia}\log P_i(\theta_a)+(1+U_{ia})\log Q_i(\theta_a)\right]$$

สำหรับขั้นตอนแรกของการประมาณค่าด้วยวิธีประมาณแบบแม็กซิมั่มไลค์ลิฮูดร่วมกันมีการกำหนดค่าเริ่มต้นของ θ_a โดยใช้ค่า log ของอัตราส่วนจำนวนข้อที่ตอบถูกต้องจำนวนข้อที่ตอบผิดสำหรับผู้ตอบแต่ละคนแปลงเป็น คะแนนมาตรฐานเพื่อใช้เป็นค่าเริ่มต้นของ θ_a จากนั้น ทำการประมาณค่าพารามิเตอร์ของข้อสอบจากค่าเริ่มต้นที่กำหนดขึ้นของ θ_a จากขั้นตอนแรก สำหรับขั้นตอนที่สอง จากค่าพารามิเตอร์ของข้อสอบที่ได้จากขั้นตอนแรก ทำเสมือนทราบค่าพารามิเตอร์ของข้อสอบเพื่อประมาณค่าความสามารถของผู้สอบวิธีนี้จะกระทำซ้ำตามขั้นตอนทั้งสองจนกระทั่งได้ค่าประมาณ 2 ครั้งหลังที่ไม่เปลี่ยนแปลงโปรแกรมคอมพิวเตอร์สำหรับการประมาณค่าด้วยวิธีการนี้ ได้แก่ โปรแกรม LOGIST สำหรับการวิเคราะห์ข้อสอบตามโมเดลการตอบสนอง 1, 2 และ 3 พารามิเตอร์ และโปรแกรม BICAL และโปรแกรม BIGSCALE สำหรับการวิเคราะห์ข้อสอบตามโมเดลการตอบสนอง 1 พารามิเตอร์ (ศิริชัย กาญจนวาสี, 2550)

ข้อพึงระวังสำหรับการประมาณค่าด้วยวิธีประมาณแบบแม็กซิมั่มไลค์ลิฮู้ดร่วมกัน คือ การประมาณค่าด้วยวิธีการนี้ไม่สามารถประมาณค่าความสามารถได้ในกรณีมีผู้สอบได้คะแนนเต็มหรือได้ศูนย์คะแนน นอกจากนี้ยังไม่สามารถประมาณค่าพารามิเตอร์ข้อสอบได้ ในกรณีข้อสอบที่มีผู้ตอบถูกหรือผิดหมดทุกคน และในการวิเคราะห์โมเดล 2 และ 3 พารามิเตอร์ ค่าพารามิเตอร์ของโมเดลการตอบสนองข้อสอบจะมีความคงเส้นคงวาได้ ก็ต่อเมื่อมีจำนวน

ข้อสอบมากและกลุ่มผู้สอบมีขนาดใหญ่ (Hambleton, Swaminathan & Rogers,1991 อ้างถึงใน ศิริชัย กาญจนวาสี, 2545)

เมื่อเปรียบเทียบวิธีแม็กซิมั่มไลค์ลิฮู้ดทั้ง 2 วิธี พบว่าวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบวิธีมาร์จินัลแม็กซิมั่มไลค์ลิฮู้ด มีข้อดีกว่าวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบร่วมกัน ที่สำคัญคือ สามารถประมาณค่าผูกสอบโดยไม่ต้องประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ รวมทั้งความแม่นยำของการวิเคราะห์ เมื่อมีจำนวนผู้สอบ 1,000 ถึง 2,000 คน ทำข้อสอบ 40 ข้อ จะให้ค่าประมาณที่แตกต่างกันเล็กน้อย แต่ในกรณีที่ผู้สอบทำข้อสอบ 10 ถึง 15 ข้อ วิธีการประมาณค้าที่เป็นไปได้สูงสุดแบบร่วมกัน จะให้ค่าประมาณความสามารถที่ลำเอียง โดยเฉพาะในผู้สอบที่มีความสามารถต่ำ ดังนั้น จึงทำให้การประมาณค่าพารามิเตอร์มีความคลาดเคลื่อนแม้จะมีผู้สอบจำนวนมาก (Hambleton & Swaminathan, 1985)

ข้อดีของการประมาณค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด คือ การประมาณค่าพารามิเตอร์ด้วยวิธีการนี้จะให้สารสนเทศของค่าพารามิเตอร์ที่ต้องการได้ทั้งหมด ไม่ว่าจะเป็นพารามิเตอร์ของผู้สอบ ซึ่งได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบ หรือพารามิเตอร์ของข้อสอบ ซึ่งได้แก่ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และค่าโอกาสในการเดาข้อสอบได้ถูกต้อง การประมาณค่าพารามิเตอร์ของวิธีนี้ขึ้นอยู่กับจำนวนของกลุ่มผู้เข้าสอบและข้อสอบ ถ้ามีจำนวนเพิ่มขึ้น การประมาณค่าก็จะมีความคงที่ไปสู่ค่าพารามิเตอร์เพิ่มมากขึ้น

ส่วนข้อจำกัดของการประมาณค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮูด คือ **ประเด็นแรก** ในการประมาณค่าพารามิเตอร์ในขั้นที่ 2 และ 3 โดยใช้ค่าอนุพันธ์อันดับที่ 2 ในกระบวนการ นิวตัน ราฟสัน นั้นมีโอกาสที่ค่าประมาณที่ได้จะไม่ลู่เข้าสู่ค่าคงที่ **ประเด็นที่สอง** สำหรับการประมาณค่าในสมการไลค์ลิฮู้ดไม่ใช้สมการเชิงเส้นตรง จะทำให้การหารากของสมการที่ทำให้ฟังก์ชันไลค์ลิฮู้ดมีค่าสูงสุดได้หลายค่า แต่ค่าเหล่านี้ไม่สามารถนำไปใช้หรือประกันได้ว่าเป็นค่าพารามิเตอร์ที่แท้จริงได้ **ประเด็นที่สาม** ในบางครั้งค่าพารามิเตอร์หรือค่าที่ได้จากการประมาณค่าไม่ตกอยู่ในขอบเขตของค่าพารามิเตอร์ กล่าวคือ อาจมีค่าใดค่าหนึ่งอยู่ภายนอกขอบเขตที่ยอมรับได้ ในกรณีเช่นนี้ต้องมีการกำหนดขอบเขตจำกัดของค่าประมาณไว้ เพื่อให้ค่าประมาณที่ได้ไม่สูงหรือต่างเกินไปนัก แต่การกระทำเช่นนี้ เป็นจุดอ่อนของกระบวนการประมาณ ค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด โดยเฉพาะในแบบจำลอง 2 และ 3 พารามิเตอร์ จึงทำให้เกิดปัญหาตามมาเกี่ยวกับความตรง (Validity) ของค่าที่ประมาณได้ และ**ประเด็นสุดท้าย**เนื่องจากการประมาณค่าพารามิเตอร์ในแบบจำลอง 3 พารามิเตอร์มีสมการหลายสมการที่ต้องหารากที่สองทำให้ฟังก์ชันไลค์ลิฮู้ดมีค่าสูงสุดด้วยวิธีของ นิวตัน ราฟสัน จึงจำเป็นต้องใช้เครื่องคอมพิวเตอร์ขนาดใหญ่ในการทำงาน

ตอนที่ 4 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte cario simulation Study)

การจำลอง (Simulation) เป็นวิธีการที่ใช้ศึกษาระบบหรือเป็นกระบวนการออกแบบจำลองระบบจริง โดยทดลองสถานการณ์ที่จำลองขึ้นด้วยการสุ่มข้อมูลตามเงื่อนไขต่าง ๆ ซึ่งเทคนิคในการจำลองข้อมูลนั้น เป็นเทคนิคที่ดำเนินการทดลองกับแบบจำลองของระบบ หรือกล่าวอีกนัยหนึ่งคือ เป็นการสร้างเครื่องมือทดลอง ซึ่งจะทำงานเลียนแบบระบบของสิ่งที่สนใจ (Eppen, Gould, & Schmidt, 1985; Frenkel, 2004; Hillier & Lieberman, 1980; Muthen & Muthen, 2007) เนื่องจากในโลกความเป็นจริงของระบบงานจริงนั้นมีความสลับซับซ้อน จึงเป็นไป ได้ยากที่จะแก้ไขโดยคณิตศาสตร์ ดังนั้นการจำลองข้อมูลจึงเป็นการสร้างสถานการณ์ที่น่าจะ เป็นไปได้โดยใช้คอมพิวเตอร์ช่วยและใช้ข้อมูลที่เก็บรวบรวมมาเสมือนว่ามาจากระบบจริงแล้วทำการจำลองระบบขึ้นมา

การศึกษาแบบมอนติคาร์โล (Monte carlo simulation study) เป็นวิธีหนึ่งที่นิยมใช้ในการหาคำตอบทางสถิติ พิจารณาการจำลองสุ่มและกระทำซ้ำมาช่วยในการหาคำตอบที่ต้องการศึกษา ซึ่งการศึกษาแบบมอนติคาร์โลเป็นการศึกษาซึ่งโดยทั่วไปจะใช้สำหรับการตรวจสอบ ประสิทธิภาพของค่าพารามิเตอร์ทางสถิติภายใต้สถานการณ์ที่แตกต่างกัน นอกจากนี้ยังเป็น เทคนิคหนึ่งสำหรับการจำลองข้อมูลทางทฤษฎีด้วย

เทคนิคการจำลองข้อมูลแบบมอนติคาร์โล เป็นการจำลองแบบด้วยคอมพิวเตอร์ (Computer simulation) โดยอาศัยตัวเลขสุ่ม (Random number) มาสร้างตัวแปรให้เหมือนสถานการณ์จริง และมีการทดลองซ้ำหลาย ๆ ครั้ง เพื่อให้ได้ค่าที่แน่นอนที่จะเป็นข้อสรุป หรืออธิบายปรากฏการณ์ต่าง ๆ ในสถานการณ์จริง หรือช่วยหาคำตอบในเรื่องต่าง ๆ ที่ยังไม่แน่ใจในผลที่เกิดขึ้น กล่าวคือ เป็นการจำลองที่อยู่ในรูปของคอมพิวเตอร์โปรแกรมซึ่งเป็นที่นิยมใช้กันมากมีการนำข้อมูลเข้า ซึ่งต้องจัดเตรียมและวิเคราะห์ข้อมูล ซึ่งอาจใช้วิธีการทางสถิติช่วยได้ นอกจากนี้การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ยังเป็นเทคนิคหนึ่งที่ใช้ในกระบวนการออกแบบจำลองสถานการณ์สำหรับแก้ปัญหาทางด้านคณิตศาสตร์ วิศวกรรมศาสตร์ หรือทางด้านสถิติ ซึ่งการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลนั้นจะช่วยในการขยายความรู้ทางทฤษฎีทางด้านสถิติให้กว้างขวางมากยิ่งขึ้น

เทคนิคมอนติคาร์โลเป็นเทคนิคการสร้างข้อมูลด้วยการใช้ตัวเลขแบบสุ่มและความน่าจะเป็นสะสม โดยมีขั้นตอน ดังนี้

1. สร้างกราฟ หรือตารางความน่าจะเป็นสะสมของข้อมูลที่ต้องการ

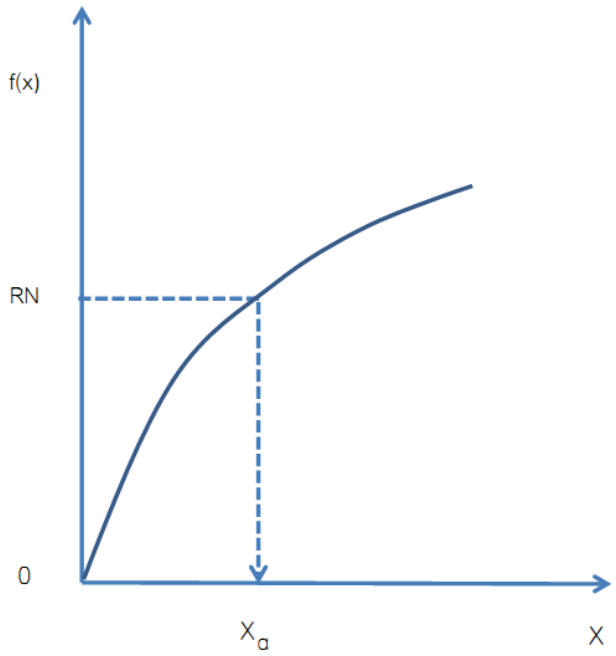
2. สร้างตัวเลขสุ่ม (Generate random number) นิยมสร้างตัวเลขสุ่มเทียมโดยอาศัยสูตรทางคณิตศาสตร์ ซึ่งการใช้เลขสุ่มเป็นสิ่งที่มีความสำคัญมากต่อระเบียบวิธีการมอนติคาร์โล เนื่องจากหลักการของเทคนิคมอนติคาร์โลนั้น อาศัยตัวเลขสุ่มที่มีลักษณะการกระจายแบบสม่ำเสมอ (Uniform) ในช่วงตัวเลข (0, 1) มาช่วยในการหาคำตอบของปัญหา

3. ใช้ค่าตัวเลขสุ่มที่ได้ในข้อ 2 แทนค่าความน่าจะเป็นสะสม หรือประยุกต์ใช้กับสถานการณ์หรือเงื่อนไขที่ต้องการศึกษา

4. ทดลองซ้ำหลาย ๆ ครั้ง เพื่อลดความคลาดเคลื่อน

5. อ่านค่าของข้อมูลจากกราฟหรือตาราง ซึ่งมีค่าความน่าจะเป็นสะสมเท่ากับตัวเลขในข้อ 3 และค่าที่ได้นี้คือค่าของข้อมูลที่ต้องการ

เช่น สมมติว่าตัวแปรแบบสุ่ม X มีลักษณะการแจกแจงของความน่าจะเป็นสะสม



ภาพที่ 2-4 กราฟแสดงการแจกแจงความน่าจะเป็นสะสมของตัวแปรสุ่ม X

วิธีการอ่านค่าของข้อมูลจากกราฟ มีวิธีการดังนี้

5.1 จากตารางตัวเลขสุ่ม (หรือจากการสร้างตัวเลขแบบสุ่มด้วยวิธีอื่น) ได้ตัวเลขสุ่ม (Random number: RN)

5.2 กำหนด RN ให้ได้ค่าอยู่ระหว่าง 0 ถึง 1 โดยใส่จุดทศนิยมข้างหน้า

5.3 หาค่า RN บนแกนตั้ง ลากเส้นนอนจากแกนตั้งที่ค่า RN ตัดกับกราฟที่ใด ลากเส้น ดิ่งลงมาหาแกนนอนซึ่งเป็นแกนของค่าตัวแปรสุ่ม X มีค่า Xa

6. ทดลองกระทำเมื่อประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่มแล้ว จะใช้กระบวนการของการสุ่มมากระทำในลักษณะซ้ำ ๆ กัน เพื่อหาคำตอบของปัญหาที่ต้องการ

ข้อดีของการใช้ข้อมูลที่ได้จากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

ข้อดีของการใช้ข้อมูลที่ได้จากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล มี 4 ประการ ดังนี้

1. การใช้ข้อมูลจากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล จะเป็นการได้มาซึ่งข้อมูลที่สามารถทำซ้ำได้

2. มีประสิทธิภาพมากในการเก็บข้อมูลที่ใช้ในทฤษฎีของการแจกแจงความน่าจะเป็น

3. ง่ายที่จะเปลี่ยนพารามิเตอร์ของทฤษฎีการแจกแจง

4. ทำให้ได้ข้อมูลตามสถานการณ์ที่กำหนดได้ในหลายสถานการณ์ซึ่งข้อมูลเชิงประจักษ์ ที่ได้จากการเก็บรวบรวมนั้นจะให้รูปแบบของข้อมูลเพียงแบบเดียวเท่านั้น

ตอนที่ 5 คุณสมบัติของตัวประมาณค่าที่ดี

เนื่องจากค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบ เป็นค่าที่สะท้อนคุณลักษณะ หรือปรากฏการณ์ในประชากร นักวิจัยไม่สามารถวัดค่าได้โดยตรง จึงจำเป็นต้องศึกษาจากค่าสถิติจากกลุ่มตัวอย่าง แล้วทำการสรุปอ้างอิงไปยังค่าของประชากรซึ่งเรียกว่า สถิติอ้างอิง (Inferential statistic) ซึ่งประกอบด้วย การประมาณค่า และการทดสอบสมมุติฐานเกี่ยวกับค่าพารามิเตอร์ (ชยุตม์ ภิรมย์สมบัติ, 2547, หน้า 25)

Mandenhall, Beaver, and Beaver (1994) and Glass and Hopkins (1996) กล่าวว่า ตัวประมาณค่าที่ดีควรเป็นตัวประมาณค่าที่ไม่มีความลำเอียง (Unbiased) กล่าวคือ ค่าประมาณ (Estimate) มีค่าเท่ากับค่าพารามิเตอร์ (Parameter) หรืออย่างน้อยที่สุดควรเป็นตัวประมาณค่าที่มีความลำอียงต่ำเมื่อเทียบกับตัวประมาณค่าอื่น ๆ นอกจากนี้จะต้องมีความคงเส้นคงวา (Consistency) มีความพอเพียง (Sufficiency) ในการใช้ข้อมูลจากทุกหน่วยตัวอย่างมาใช้ประมาณค่า และมีประสิทธิภาพสัมพัทธ์สูง (Relative efficiency) ในงานวิจัยนี้พิจารณาคุณสมบัติของตัวประมาณค่าในด้าน ความลำเอียง ความคงเส้นคงวา สำหรับความพอเพียงไม่ได้นำมาพิจารณาเนื่องจากตัวประมาณค่าทุกตัวใช้ข้อมูลจากผู้สอบทั้งหมดในการคำนวณอยู่แล้ว

ดัชนีค่าความลำเอียง (BIAS) หมายถึง ความคลาดเคลื่อนในการประมาณค่าค่าพารามิเตอร์ของผู้สอบและความสามารถของผู้สอบ โดยใช้ค่าความแตกต่างระหว่างค่าคาดหวังของตัวประมาณค่ากับค่าที่แท้จริง (Expected-parameter) ซึ่งเป็นการแสดงให้เห็นว่าค่าพารามิเตอร์ที่ได้จากการประมาณค่าทดแทนข้อมูลสูญหาย มีผลต่างจากค่าพารามิเตอร์ที่แท้จริงโดยเฉลี่ยมากน้อยเพียงใด ซึ่งดัชนีค่าความลำเอียง (BIAS) แสดงถึงความถูกต้อง (Accuracy) ของการประมาณค่าพารามิเตอร์ โดยใช้เกณฑ์ในการพิจารณาคือ ถ้าดัชนีค่าความลำเอียง (BIAS) มีค่าใกล้ศูนย์มากกว่าแสดงให้เห็นว่า การประมาณค่าใกล้เคียงกับค่าพารามิเตอร์ที่แท้จริงมากกาว่า ดังสมการ

$$= \frac{\sum(\theta - \hat{\theta})}{n}$$

เมื่อ

θ	แทน	ค่าพารามิเตอร์ที่แท้จริง
$\hat{\theta}$	แทน	ค่าพารามิเตอร์ที่ได้จากการประมาณค่า
n	แทน	จำนวนครั้งที่ทำซ้ำในการจำลองข้อมูล

ความคงเส้นคงวา หมายถึง คุณลักษณะของตัวประมาณค่าที่มีแนวโน้มในการให้ค่าประมาณเข้าใกล้ค่าพารามิเตอร์เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น (An estimator tends to get closer to the value of the parameter as the sample size becomes larger) หรือเมื่อขนาดกลุ่มตัวอย่างลู่เข้าสู่อนันต์ ($N \longrightarrow \infty$)

ในการพิจารณาคุณสมบัติความคงเส้นคงวา พิจารณาจาก ดัชนีค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย ซึ่งหมายถึง ความเบี่ยงเบนของค่าพารามิเตอร์ที่แท้จริง และค่าที่ประมาณได้ หรือ Root mean square error: RMSE) ซึ่งความเบี่ยงเบนของค่าพารามิเตอร์ที่แท้จริง และค่าที่ประมาณได้ สะท้องถึงความคงเส้นคงวา (Consistency) ของค่าพารามิเตอร์ข้อสอบและความสามารถผู้สอบที่ประมาณได้จากวิธีจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าทดแทนพหุ และข้อมูลที่สมบูรณ์ โดยมีเกณฑ์ในการพิจารณาคือ ถ้าค่า RMSE น้อยกว่า แสดงว่ามีความคงที่ของการประมาณค่าพารามิเตอร์สูงกว่าสะท้อนให้เห็นถึงความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์มีน้อย แต่ถ้าค่า RMSE มากกว่า แสดงว่า

มีความคงที่ของการประมาณค่าพารามิเตอร์ต่ำกว่าสะท้อนให้เห็นถึงความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์มีมาก ดังสมการ

$$\text{RMSE} = \sqrt{\frac{\sum(\hat{\theta}-\theta)^2}{n}}$$

เมื่อ

θ	แทน	ค่าพารามิเตอร์ที่แท้จริง
$\hat{\theta}$	แทน	ค่าพารามิเตอร์ที่ได้จากการประมาณค่า
n	แทน	จำนวนครั้งที่ทำซ้ำในการจำลองข้อมูล

ความพอเพียง หมายถึง คุณสมบัติของตัวประมาณค่าที่สามารถใช้ข้อมูลจากทุกหน่วยตัวอย่างที่มีในการประมาณค่าพารามิเตอร์ในประชากร แต่เนื่องจากการวิจัยในครั้งนี้ผู้วิจัยใช้ข้อมูลที่ได้จากการจำลองสถานการณ์ จึงสามารถประมาณค่าพารามิเตอร์ได้จากข้อมูลทั้งหมดที่มีคุณสมบัติความพอเพียงจึงไม่นำมาพิจารณา

เนื่องจากการทำวิจัยในครั้งนี้ ผู้วิจัยได้ทำการจำลองค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบด้วยการสุ่มค่าดังกล่าวจากการแจกแจงความน่าจะเป็นของค่าพารามิเตอร์ที่ได้ศึกษามาแล้ว ซึ่งค่าที่ได้ไม่ใช่ค่าพารามิเตอร์ที่แท้จริง ดังนั้นการศึกษาค่าความถูกต้องของการประมาณค่า วัดได้จากค่าต่อไปนี้

ดัชนีผลต่าง (diff) หมายถึง ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบ โดยใช้ค่าความแตกต่างระหว่างค่าเฉลี่ยค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบที่ได้จากการจำลองกับค่าประมาณที่ได้จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI) ซึ่งเป็นการแสดงให้เห็นว่าค่าที่ได้จากการประมาณค่าทดแทนข้อมูลสูญหาย มีผลต่างจากค่าที่กำหนดจากการจำลองโดยเฉลี่ยมากน้อยเพียงใด ซึ่งดัชนีผลต่าง แสดงถึงความถูกต้อง (Accuracy) ของการประมาณค่า โดยใช้เกณฑ์ในการพิจารณาคือ ถ้าดัชนีผลต่าง มีค่าใกล้ศูนย์มากกว่าแสดงให้เห็นว่า การประมาณค่าใกล้เคียงกับค่าที่กำหนดจากการทำการจำลอง ดังสมการ

$$\text{diff} = \frac{\sum(A-\widehat{A})}{n}$$

เมื่อ

diff แทน ดัชนีผลต่าง

A แทน ค่าเฉลี่ยของค่าพารามิเตอร์ข้อสอบ และค่าความสามารถของผู้สอบที่ได้จากการจำลองข้อมูล

$\widehat{A}$ แทน ค่าประมาณที่ได้จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI)

n แทน จำนวนครั้งที่ทำซ้ำในการจำลองข้อมูล

ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย

ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย หมายถึง ความเบี่ยงเบนของค่าเฉลี่ยค่าพารามิเตอร์ข้อสอบและความสามารถของผู้สอบที่ได้จากการจำลองข้อมูล และค่าที่ประมาณได้ ซึ่งเป็นค่าที่สะท้อนถึงความคงเส้นคงวา (Consistency) ของค่าพารามิเตอร์ข้อสอบและความสามารถผู้สอบที่ประมาณได้กับข้อมูลที่สมบูรณ์ โดยมีเกณฑ์ในการพิจารณาคือ ถ้าค่า diff error น้อยกว่า แสดงว่ามีความความคงเส้นคงวา ของการประมาณค่าสูงกว่าสะท้อนให้เห็นถึงความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์มีน้อย แต่ถ้าค่า diff error มากกว่า แสดงว่ามีความความคงเส้นคงวาของการประมาณค่าพารามิเตอร์ต่ำกว่าสะท้อนให้เห็นถึงความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์มีมาก ดังสมการ

$$\text{diff error} = \sqrt{\frac{\sum\left(A-\widehat{A}\right)^2}{n}}$$

เมื่อ

diff error แทน ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย

A แทน ค่าเฉลี่ยของค่าพารามิเตอร์ข้อสอบ และค่าความสามารถของผู้สอบที่ได้จากการจำลองข้อมูล

$\widehat{A}$	แทน	ค่าประมาณที่ได้จากการจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI)
n	แทน	จำนวนครั้งที่ทำซ้ำในการจำลองข้อมูล

ตอนที่ 6 แนวคิด ทฤษฎีเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ

กาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบจึงเป็นขั้นตอนที่มีความสำคัญยิ่ง ในการพัฒนาแบบสอบ ที่จะประเมินว่าข้อสอบที่สร้างขึ้นมีความยุติธรรม หรือความเท่าเทียมกัน ในทุกกลุ่มผู้สอบหรือไม่ในระยะเริ่มต้นของการศึกษามีการใช้คำว่า ความลำเอียงของข้อสอบ (Item bias) ซึ่งคำดังกล่าวอาจก่อให้เกิดความเข้าใจที่ไม่ตรงกัน เพราะสามารถแปลความหมาย ได้ทั้ง หลักการทางสังคมศาสตร์ และสถิติศาสตร์ จึงมีการเปลี่ยนมาใช้คำใหม่ที่มีความเหมาะสมกว่า คือ การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) ซึ่งเป็นความหมายที่ เน้นกระบวนการทางสถิติเข้ามาตรวจสอบ เพื่อให้ได้ข้อมูลสารสนเทศเกี่ยวกับการทำหน้าที่ต่างกัน ของข้อสอบสำหรับผู้สอบแต่ละกลุ่มย่อยที่มีความสามารถหลัก (θ) เหมือนกัน แต่อาจแตกต่างกัน ในลักษณะเฉพาะบางอย่าง (ศิริชัย กาญจนวาสี, 2545) ซึ่งจากการศึกษา พบว่ามีนักวัดผล หลายท่าน ได้ให้ความหมายไว้ ดังนี้

ศิริชัย กาญจนวาสี (2550, หน้า 117) ได้นิยามการทำหน้าที่ต่างกันของข้อสอบ หมายถึง การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัด เท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองข้อสอบ แตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบเกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับผู้สอบกลุ่มย่อย ที่ต่างกันและมีความสามารถหลัก (Primary ability) ระดับเดียวกันหรือมีคุณลักษณะแฝง (Latent trait) ที่ต้องการวัดเท่ากัน แต่มีความสามารถรอง (Secondary ability) แตกต่างกัน ทำให้ผู้สอบต่างกลุ่มที่นำมาจับคู่เปรียบเทียบกันมีโอกาสตอบข้อสอบถูกแตกต่างกัน

สุพัฒนา หอมบุปผา (2556) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสของการตอบข้อสอบได้ถูกต้องแตกต่างกัน สำหรับผู้สอบที่มีคุณลักษณะ หรือความสามารถในระดับเดียวกัน แต่มาจากกลุ่มประชากรย่อยที่แตกต่างกัน

อาวีพร ปานทอง (2558, หน้า 32) ได้ให้ความหมายแบบทดสอบที่ทำหน้าที่ต่างกัน ของข้อสอบ หมายถึง แบบทดสอบที่เมื่อนำไปทดสอบกับกลุ่มผู้สอบที่มีความสามารถเท่าเทียมกัน จากประชากรเดียวกัน แต่มีลักษณะของกลุ่มย่อยแตกต่างกัน เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม สภาพภูมิศาสตร์ สถานะทางเศรษฐกิจ เป็นต้น แล้วปรากฏว่า ความน่าจะเป็น ในการตอบข้อเดียวกันของกลุ่มผู้สอบมีความถูกต้องไม่เท่ากัน ทำให้มีการได้เปรียบเสียเปรียบ

จากการตอบข้อสอบข้อเดียวกันและใช้เกณฑ์จากการวิเคราะห์ผลการตอบ ข้อสอบตามกลุ่มย่อยของผู้สอบเพื่อบ่งบอกว่ากลุ่มใดเป็นฝ่ายได้เปรียบหรือเสียเปรียบ จากการตอบข้อสอบนั้น

อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551, หน้า 51) ได้ให้ความหมาย การทำหน้าที่ต่างกันของข้อสอบ หมายถึง การที่ข้อสอบทำให้ผู้สอบที่มีลักษณะหรือมาจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน

Angoff (1993) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบ หมายถึง การสังเกตได้ว่าข้อสอบข้อนั้นแสดงคุณสมบัติทางสถิติที่ต่างกัน เมื่อใช้ข้อมูลจากผู้ตอบข้อสอบที่มีความสามารถเท่ากัน แต่อยู่ในกลุ่มที่ต่างกัน

Camilli and Shepard (1994) ได้นิยามการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่มีความลำเอียงเป็นความคลาดเคลื่อนอย่างเป็นระบบหรือความไม่เที่ยงตรงของการวัดทำให้สมาชิกบางกลุ่ม ได้ประโยชน์หรือเสียประโยชน์ เช่น แบบทดสอบหรือข้อสอบที่เข้าข้างกลุ่มเพศชายมากกว่าเพศหญิง หรือคนผิวขาวกับคนผิวดำ หรือระหว่างคนที่อยู่ในที่ตั้งทางภูมิศาสตร์แตกต่างกัน

Dorans and Kulick (1986) การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่ง มีค่าต่ำกว่าหรือสูงกว่าผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน

Hambleton, Swaminathan, and Rogers (1991, p. 110) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ผู้สอบที่มีความสามารถระดับเดียวกัน แต่มาจากกลุ่มย่อยที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกไม่เท่ากัน

Narayanan and Swaminathan (1996) ได้นิยามการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ฟังก์ชันการตอบสนองข้อสอบซึ่งคำนวณจากกลุ่มผู้สอบกลุ่มย่อยที่ต่างกันมีค่าไม่เท่ากัน

จากความหมายของการทำหน้าที่ต่างกันของข้อสอบที่กล่าวมา สรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบซึ่งมีลักษณะต่างกัน แต่มีความสามารถระดับเดียวกัน มีโอกาสตอบข้อสอบถูกไม่เท่ากัน

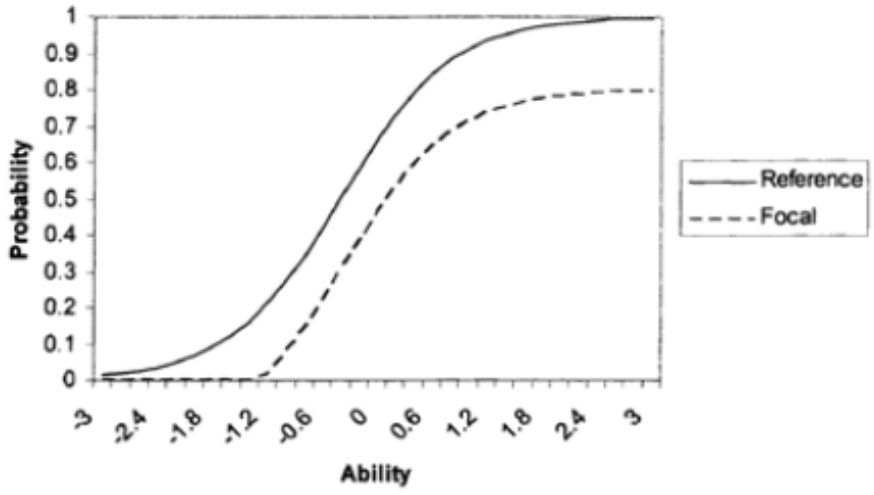
ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป ปกตินิยมเปรียบเทียบ 2 กลุ่ม คือ กลุ่มแรก เป็นกลุ่มที่คาดว่าจะเป็นกลุ่มที่เสียประโยชน์จากข้อสอบที่แสดงออกถึงการทำหน้าที่ต่างกันของข้อสอบและเป็นกลุ่มที่นักวิจัยสนใจศึกษา เรียกว่า กลุ่มเปรียบเทียบ (Focal group หรือกลุ่ม F) ส่วนกลุ่มที่สองเป็น

กลุ่มที่คาดว่าจะเป็นกลุ่มที่ได้ประโยชน์จากข้อสอบที่แสดงออกถึงการทำหน้าที่ต่างกันของข้อสอบ เรียกว่า กลุ่มอ้างอิง (Reference group หรือกลุ่ม R)

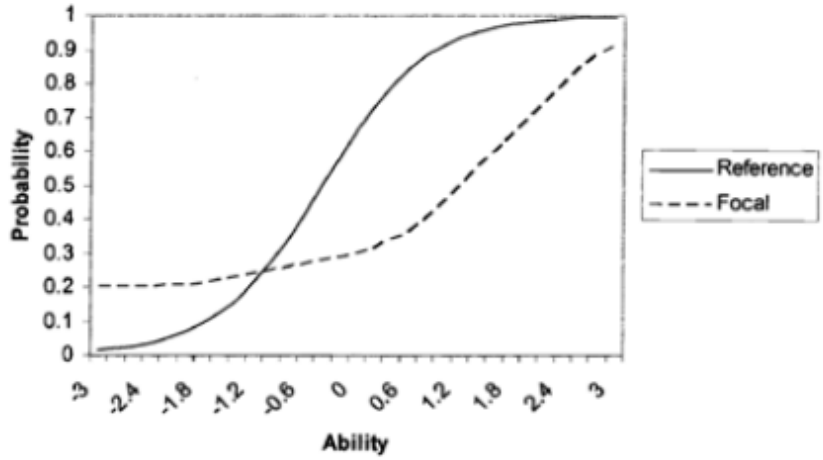
การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ได้แบ่งลักษณะของข้อสอบที่หน้าที่ต่างกัน แบ่งเป็น 2 ประเภท ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) และข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Non-uniform DIF) (ศิริชัย กาญจนวาสี, 2550; Sedivy, 2009)

1. ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึง ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งสม่ำเสมอกันทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบ (ICC) ของผู้สอบ 2 กลุ่ม พบว่าไม่มีปฏิสัมพันธ์ (Interaction) ระหว่างโค้งคุณลักษณะผู้สอบในทุกระดับความสามารถ ดังภาพที่ 2-5



ภาพที่ 2-5 โค้งคุณลักษณะข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF)

2. การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF) หมายถึง ข้อสอบที่ทำให้โอกาสการตอบข้อสอบถูกต้องของผู้สอบระหว่างกลุ่มไม่สม่ำเสมอกันทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบทั้ง 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างโค้งคุณลักษณะ เช่น ที่ระดับความสามารถหนึ่งผู้สอบกลุ่มอ้างอิงมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มเปรียบเทียบ แต่อีกที่ระดับความสามารถหนึ่งผู้สอบกลุ่มเปรียบเทียบมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มอ้างอิง ดังภาพที่ 2-6



ภาพที่ 2-6 โค้งคุณลักษณะข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF)

โดยทั่วไปในแบบสอบมาตรฐานมักมีข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปมากกว่าข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป แต่ในข้อมูลจริงจะมีข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้มากกว่า (สุพัฒนา หอมบุปผา, 2556) เนื่องจากในการทำหน้าที่ต่างกันของข้อสอบ คือ การที่ข้อสอบวัดความสามารถรอง (Secondary abilities หรือ Nuisance dimension หรือ η) หรือ คุณลักษณะแฝงอื่นนอกเหนือจากความสามารถหลัก (Primary abilities หรือ Primary dimension หรือ η) หรือคุณลักษณะแฝงที่ต้องการวัด ที่จะส่งผลให้ผู้สอบต่างกลุ่มที่นำเข้ามาจับคู่เปรียบเทียบกัน มีโอกาสในการตอบข้อสอบได้ถูกต้องต่างกัน ทั้ง ๆ ที่มีความสามารถหลักที่ต้องการวัดเท่ากัน นั่นคือ การทำหน้าที่ต่างกันจะเกิดเมื่อมีค่าเฉลี่ย ηไม่เท่ากัน ซึ่งการตัดสินการทำหน้าที่ต่างกันของข้อสอบมีได้ 4 สถานการณ์ ดังนี้ (Sedivy, 2009)

1. กลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเฉลี่ย ηและ ηเท่ากัน แสดงว่าไม่มีการทำหน้าที่ต่างกัน (No Bias/ DIF)

2. กลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเฉลี่ย ηเท่ากัน แต่มีค่าเฉลี่ย ηต่างกัน แสดงว่ามี Impact นั่นคือ ความน่าจะเป็นในการตอบข้อสอบถูกมาจากความแตกต่างของความสามารถหลัก

3. กลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเฉลี่ย ηเท่ากัน แต่มีค่าเฉลี่ย η ต่างกัน แสดงว่ามีการทำหน้าที่ต่างกัน

4. กลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเฉลี่ยแตกต่างกันทั้ง ηและ η แสดงว่ามีการทำหน้าที่ต่างกัน และมี Impact

หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีหลายวิธี เนื่องจากมีวิธีการศึกษาและการคิดค้นวิธีการต่าง ๆ เพื่อให้สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างมีประสิทธิภาพมากที่สุด ซึ่งสามารถแบ่งตามประเภทการวิเคราะห์ได้เป็น 2 กลุ่ม ดังนี้

1. กลุ่มที่ใช้คะแนนสังเกตได้ (Observe score) เป็นกลุ่มที่ใช้คะแนนรวมของแบบสอบเป็นเกณฑ์ในการจับคู่ผลสอบสองกลุ่มตามความรู้ หรือความสามารถที่แท้จริงของผู้สอบ วิธีการในกลุ่มนี้ ได้แก่ วิธีแมนเทล-แฮนส์เซล วิธีถดถอยโลจิสติก และวิธีทำให้เป็นมาตรฐาน จุดเด่นของวิธีการในกลุ่มนี้ คือ กลุ่มตัวอย่างขนาดเล็ก การวิเคราะห์ไม่ยุ่งยากซับซ้อน ส่วนจุดด้อยของวิธีการในกลุ่มนี้ คือ ค่าสถิติจะเปลี่ยนไปตามขนาดของกลุ่มตัวอย่าง เมื่อกลุ่มตัวอย่างที่ศึกษาเปลี่ยนไปผลการศึกษาก็อาจเปลี่ยนแปลงไปด้วย

2. กลุ่มที่ใช้คะแนนที่สังเกตไม่ได้ หรือเป็นตัวแปรแฝง (Latent variable) เป็นกลุ่มวิธีที่มีทฤษฎีการทดสอบเป็นพื้นฐาน ใช้การประมาณค่าคุณลักษณะแฝง (Latent trait) หรือใช้คะแนนจริงของผู้สอบเป็นเกณฑ์ในการจับคู่เปรียบเทียบผู้สอบ วิธีการในกลุ่มนี้ ได้แก่ วิธีการตอบสนองข้อสอบ (IRT) และวิธีซิบเทสท์ (SIBTEST) เป็นต้น

ศิริชัย กาญจนวาสี (2550) ได้เสนอหลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบที่ต้องใช้วิธีการจับคู่ตามเกณฑ์ความสามารถ เพราะเป็นเงื่อนไขสำคัญของการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ที่นิยมมี 2 วิธีที่สำคัญ ดังนี้

1. **เกณฑ์ภายนอก** (External criterion)

การทำหน้าที่ต่างกันของข้อสอบโดยใช้เกณฑ์ภายนอกนี้ สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อ และแบบสอบทั้งฉบับ โดยใช้คะแนนจากแบบสอบอื่นเป็นเกณฑ์ภายนอก แล้วใช้เทคนิคการวิเคราะห์ถดถอย (Regression analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ

หลักการนี้มีจุดมุ่งหมาย เพื่อสร้างสมการทำนายตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบทดสอบอื่นจากตัวแปรทำนายที่เป็นคะแนนรายข้อ หรือคะแนนแบบทดสอบระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ ในการวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบ จะใช้คะแนนรวมของแบบทดสอบทั้งฉบับเป็นตัวแปรทำนาย สำหรับตัวแปรเกณฑ์ที่ใช้เป็นเกณฑ์ภายนอกอาจใช้คะแนนรวมทั้งฉบับหรือเกรดเฉลี่ย หรือคะแนนจากงานที่เกี่ยวข้องของผู้สอบ (Cronbach, 1970) สมการทำนายสำหรับกลุ่มอ้างอิง และกลุ่มเปรียบเทียบแสดงได้ดังนี้ คือ

กลุ่มอ้างอิง (R) $Y_i = A_R + B_R X_i$

กลุ่มเปรียบเทียบ(F) $Y_i = A_F + B_F X_i$

โดยที่ Y_i = คะแนนของตัวแปรเกณฑ์ภายนอก

x_i = คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (Intercept)

B = ค่าความชัน (Slope)

จากฟังก์ชันการทำนายดังกล่าว สามารถเปรียบเทียบค่าตัดแกน (A) และค่าความชัน (B) ของเส้นกราฟระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบได้ ถ้าเส้นกราฟมีค่าความชันหรือค่าตัดแกนแตกต่างกัน สำหรับข้อสอบใด หรือแบบสอบใดแสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกัน โดยเข้าข้างกลุ่มผู้สอบที่มีค่าตัดแกนหรือค่าความชันที่สูงกว่า

การใช้เกณฑ์ภายนอกมีข้อดี คือ เกณฑ์ที่ใช้มีความเป็นอิสระจากข้อสอบ และแบบสอบที่ต้องการตรวจสอบ แต่มีจุดอ่อน คือ ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ ในทางปฏิบัติเป็นการยากที่จะหาเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความตรงเชิงทำนาย และมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าว จะทำให้ผลวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบขาดความแม่นยำและความสมบูรณ์

2. เกณฑ์ภายใน (Internal criterion)

การวิเคราะห์การทำหน้าที่ต่างกันโดยใช้เกณฑ์ภายในเป็นการนำวิธีการทางสถิติตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หรือแบบสอบ โดยเน้นการพิจารณาจากโครงสร้างภายในของแบบสอบเป็นหลัก ด้วยการวิเคราะห์ผลจากการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ที่มีความสามารถหรือคะแนนจริงเท่ากัน ว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบการวิเคราะห์ลักษณะนี้นิยมใช้ค่าสถิติต่าง ๆ เป็นตัวบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ ค่าสถิติที่นิยมนำมาใช้สรุปได้ ดังนี้

2.1 การทดสอบปฏิสัมพันธ์ (Interaction)

ในระยะเริ่มต้นของการศึกษาการทำหน้าที่ต่างกันของข้อสอบ มีการใช้สถิติทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับข้อสอบ หากผลการทดสอบมีนัยสำคัญ สรุปได้ว่ามีการทำหน้าที่ต่างกันของข้อสอบแล้ว

จากนั้นทำการวิเคราะห์ต่อด้วยการเปรียบเทียบเป็นรายคู่ (Post hoc) เพื่อระบุข้อสอบที่มีผลต่อการเกิดปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน

2.2 การวัดความเบี่ยงเบนสัมพัทธ์ (Relative deviation)

วิธีการนี้เป็นการคำนวณค่าพารามิเตอร์ข้อสอบจำแนกตามกลุ่ม แล้วแปลงให้เป็นค่าความยากมาตรฐาน (␣) สามารถนำมาเขียนเป็นกราฟเปรียบเทียบเป็นรายข้อ ถ้าข้อใดมีค่าเบี่ยงเบนไปจากแกนหลักที่คาดหมาย หรือเบี่ยงเบนเกินจากความคลาดเคลื่อนมาตรฐานของค่าความยากที่กำหนดก็ย่อมเป็นเครื่องบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ

วิธีการนี้มีข้อดีและข้อเสียคล้ายกับการทดสอบปฏิสัมพันธ์ และค่าความยากของข้อสอบไม่ใช่ตัวแทนของค่าความยากที่แท้จริงของข้อสอบ และอาจได้รับอิทธิพลจากตัวแปรอื่น เช่น ค่าอำนาจจำแนกและความสามารถของผู้เข้าสอบ

2.3 การเปรียบเทียบน้ำหนักองค์ประกอบ (Factor loading)

วิธีการนี้ใช้การวิเคราะห์องค์ประกอบ (Factor analysis) เป็นเทคนิคทางสถิติที่นิยมใช้ตรวจสอบความตรงเชิงโครงสร้าง (Construct validity) โดยนำการวิเคราะห์องค์ประกอบมาใช้ในการวิเคราะห์โครงสร้างของแบบทดสอบแยกตามกลุ่มผู้เข้าสอบ ความไม่สอดคล้องกันระหว่างน้ำหนักองค์ประกอบบนคุณลักษณะสำคัญในสิ่งที่มุ่งวัด หรือความแตกต่างของค่าเฉลี่ยคะแนนองค์ประกอบ (Factor score) ระหว่างกลุ่มผู้เข้าสอบ ย่อมสะท้อนให้เห็นถึงการทำหน้าที่ต่างกันของข้อสอบและแบบทดสอบ

ในการใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis: EFA) สำหรับศึกษาการทำหน้าที่ต่างกันของข้อสอบ มีจุดอ่อนเรื่องความไม่สอดคล้องระหว่างน้ำหนักองค์ประกอบ ซึ่งอาจเกิดจากความแตกต่างของความสามารถระหว่างกลุ่มก็ได้ แนวทางที่เหมาะสมจึงควรใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory factor analysis: CFA) นอกจากนี้ยังใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่มในลักษณะความสามารถหลักหรือความสามารถรองได้อีก

2.4 การเปรียบเทียบโอกาสตอบข้อสอบถูก

วิธีการนี้จะทำการวิเคราะห์เปรียบเทียบโอกาสของการตอบข้อสอบถูกของผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน ซึ่งเป็นแนวทางที่ได้รับความนิยมและเป็นที่ยอมรับในปัจจุบัน ซึ่งเป็นตัวบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีแนวทางในการคำนวณค่าสถิติใน 2 แนวทางหลัก ดังนี้

2.4.1 การเปรียบเทียบค่าสัดส่วนความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากัน

2.4.2 การเปรียบเทียบค่าฟังก์ชั่นการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากัน ซึ่งวิธีการนี้ตั้งอยู่บนหลักการของทฤษฎี IRT เช่น วิธีการวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีการวัดความแตกต่างของพื้นที่ เป็นต้น วิธีการนี้มีข้อดี คือ การคำนวณค่าสถิติของข้อสอบมีความน่าเชื่อถือ มีการควบคุมความสามารถของผู้เข้าสอบโดยการจับคู่กลุ่มความสามารถ เพื่อเปรียบเทียบ ณ ตำแหน่งต่าง ๆ ที่มีความสามารถระดับเท่ากัน จึงเป็นวิธีการที่ยอมรับกันโดยทั่วไป แต่ก็มีข้อจำกัด คือ ความซับซ้อนของแนวคิดพื้นฐาน และการใช้โปรแกรมสำเร็จรูปเพื่อวิเคราะห์ข้อมูลได้บางโปรแกรม

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การจำแนกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สามารถจำแนกตามลักษณะของเกณฑ์ประกอบด้วย ลักษณะการตรวจให้คะแนน (แบบทวิภาคและแบบพหุภาค) มิติของตัวแปร (กลุ่มวิธีที่ใช้คะแนนสังเกตได้และกลุ่มวิธีที่ใช้คะแนนของตัวแปรแฝง) มิติลักษณะของสถิติวิเคราะห์ (กลุ่มที่ใช้พาราเมตริก และกลุ่มที่ใช้สถิตินันพาราเมตริก) รายละเอียดแสดงดังนี้

1. จำนกตามลักษณะการตรวจให้คะแนน

1.1 กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบทวิวิภาค (Dichotomous DIF methods) หรือการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่า (Dichotomous DIF procedures) แบบสอบที่มีลักษณะของการตรวจให้คะแนนแบบนี้ ได้แก่ แบบสอบชนิดเลือกตอบที่มีการให้คะแนนในการตอบถูกเป็น 1 คะแนน ในขณะที่ตอบผิดได้ 0 คะแนนการนำเสนอการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบสอบถูกกำหนดอย่างจริงจังเพื่อให้เกิดความยุติธรรมในการใช้แบบสอบและเกิดความตรงต่อการแสดงความหมายที่แฝงอยู่ของคะแนนเหล่านั้น งานวิจัยที่เกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบมีความหลากหลายมากขึ้น เดิมเน้นไปที่ข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค ในช่วงระยะเวลาที่ผ่านมามีความพยายามพัฒนาแนวทางเลือกใหม่ของวิธีการวัดซึ่งช่วยจุดประกายให้เกิดประเด็นที่น่าสนใจ และยังมีการทำหน้าที่ต่างกันของข้อสอบชนิดอื่นที่นอกเหนือจากข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค (Kim, Chosen, & Kim, 2007

1.2 กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาค (Polytomously methods) เช่น ข้อสอบวัดภาคปฏิบัติ (Performance assessment) ข้อสอบความเรียง (Essay items) การตัดสินคุณภาพของแฟ้มสะสมงาน (Portfolio assessment) ข้อสอบที่วัดการอ่าน (Reading item) และข้อสอบที่วัดการเขียน (Writing item) รวมไปถึงข้อสอบปลายเปิด (Open-endeditem) เป็นต้น

2. จำแนกตามมิติของตัวแปรเกณฑ์

2.1 กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้ (Observed score) ค่าพารามิเตอร์แปรเปลี่ยนไปตามกลุ่มผู้สอบ วิธีในกลุ่มนี้มักวิเคราะห์ตามทฤษฎีทางการสอบแบบดั้งเดิม เรียกกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non-IRT approach) ใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ การวิเคราะห์ความแปรปรวน (ANOVA) การวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (Polytomous logistic regression) วิธีแมนเทล-แฮนส์เซลทั่วไป (General mantel-haenszel) และวิธีดัชนีมาตรฐานพหุวิภาค (Polytomous standardization)

2.2 กลุ่มวิธีที่ใช้คะแนนของคุณลักษณะหรือตัวแปรแฝง (Latent variable) วิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ ตัวแปรแฝงหรือคุณลักษณะดังกล่าวถูกใช้เป็นเกณฑ์การจับคู่กลุ่มผู้สอบ ค่าพารามิเตอร์คงที่ไม่ว่าจะใช้กลุ่มผู้สอบใด วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ วิธีการให้คะแนนบางส่วนทั่วไป (Generalized partial credit model: GPCM) วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (General IRT likelihood ratio) วิธีการให้คะแนนบางส่วน (Partial credit model: PCM) และวิธีซิปเทสท์พหุวิภาค (Polytomous SIBTEST)

3. จำแนกตามมิติลักษณะของสถิติวิเคราะห์

3.1 กลุ่มที่ใช้สถิติพาราเมตริก (Parametric approach) การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบมีข้อตกลงเบื้องต้นของโมเดลสำหรับอธิบายความสัมพันธ์ระหว่างคะแนนของข้อสอบและการจับคู่ตัวแปร

3.2 กลุ่มที่ใช้สถิตินันพาราเมตริก (Nonparametric approach) การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบไม่มีข้อตกลงเบื้องต้นของโมเดล ความหลากหลายของวิธีทางสถิติสามารถช่วยพัฒนาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิวิภาคและพหุวิภาค (Penfield, 2005)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่สำคัญ ๆ แสดงในตารางที่ 2-3, 2-4 (Feinstein,1995; Potenza & Dorans,1995 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550)

ตารางที่ 2-3 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับรูปแบบการตรวจให้คะแนนแบบทวิวิภาค ที่เป็นเอกมิติจำแนกตามลักษณะของข้อมูล (Marie, 2007)

วิธีการ (Method)	**สถิติพาราเมติก par)/ สถิตินันพาราเมติก (par)**	**การจับคู่ตัวแปรคะแนนที่สังเกตได้ (Obs)/ คะแนนของตัวแปรแฝง**	**ลักษณะการให้คะแนนทวิภาค(D)/ พหุภาค (P)**	**การทดสอบ(T)/ การวัด (M)**	**DIF เอกรูป (U)/ DIF อเนกรูป (N)**
Mentel-Haenszel	Non-p	Obs	D/P	T/M	U
Standardization	Non-p	Obs	D	M	U
Chi-square methods	Non-p	Obs	D	T	U/N
SIBTEST	Non-p	Obs	D/P	T/M	U/N
Logistic Regression	Par	Obs	D/P	T/M	U/N
Likelihood ratio test	Par	Obs/Lat	D/P	T/M	U/N
Prob. Diff. indices	Par	Lat	D	M	U/N
B parameter indices	Par	Lat	D	M	U/N
General IRT-LR	Par	Lat	D/P	T/M	U/N
IRT LRT	Par	Lat	D/P	T	U/N
IRT methods	Par	Lat	D/P	T/M	U/N
Lord's chi-squared test	Par	Lat	D	T	U/N
Log linear models	Par	Obs	D/P	T	U/N
Mixed effect models	Par	Lat	D/P	T	U/N

หมายเหตุ: (Par)ametric/ (non-p)arametric, (Obs)erved/ (Lat)ent, (D)ichotomously/ (P)olytomously, (T)est DIF/ (M)easure DIF, (U)niform/ (N)onuniform

จากตารางที่ 2-3 แสดงวิธีตรวจสอบ DIF ใน 14 วิธีการ สำหรับข้อมูลที่มีลักษณะเป็นเอกมิติ (Uni-dimensional) (Marie, 2007) จำแนกรายละเอียดที่ต้องการศึกษาออกเป็น 5 มิติ ได้แก่ มิติประเภทของสถิติ (พาราเมตริก (Par) ametric และนอนพาราเมตริก (non-p) arametric) มิติประเภทของคะแนน (คะแนนที่สังเกตได้ (Obs) erved และคะแนนแฝง (Lat)ent) มิติรูปแบบการตรวจให้คะแนน (ทวิวิภาค (D) ichotomously และพหุวิภาค (P) olytomously) มิติประเภทของการตรวจสอบ (การตรวจสอบ (T) estDIF และการวัดขนาด (M) easure DIF) มิติรูปแบบDIF (เอกรูป(U) niform และอเนกรูป (N) onuniform) ข้อมูลในตารางที่เสนอจะนำไปสู่การตัดสินใจเปรียบเทียบการเลือกใช้วิธีการตรวจสอบเพื่อให้สอดคล้องและเหมาะสมตามจุดมุ่งหมายของสารสนเทศที่ต้องการ

วิธีการตรวจสอบการทำหน้าที่ต่างกันที่มีการให้คะแนนแบบทวิวิภาค (Dichotomous DIF) และพหุวิภาค (Polytomous DIF) แสดงในตารางที่ 2-4

ตารางที่ 2-4 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบทวิภาค (Dichotomous DIF) และพหุวิภาค (Polytomous DIF) (ศิริชัย กาญจนวาศรี, 2550)

ประเภทและตัวแปรเกณฑ์	พาราเมตริก	นันพาราเมตริก
1. DIF แบบทวิภาค (Dichotomous DIF)		
1.1 คะแนนที่สังเกตได้	ANOVA	TID, MH
(Observed score)	Logistic Regression	STND
1.2 คุณลักษณะ/ ตัวแปรแฝง	IRT-D^2, *Lord's* χ^2	SIBTEST
(Latent variable)	General IRTLR	
	Loglinear IRTLR	
2. DIF แบบทวิภาค (Dichotomous DIF)		
2.1 คะแนนที่สังเกตได้	ANOVA	
(Observed score)	Polytomous Logistic Regression	Polytomous STND GMH
1.2 คุณลักษณะ/ ตัวแปรแฝง	General IRTLR	Polytomous SIBTEST
(Latent variable)	PCM	PCM

ในการศึกษาครั้งนี้ผู้วิจัยเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิภาค (Dichotomous) โดยใช้วิธีทดสอบอัตราส่วนความควรจะเป็น รายละเอียดดังนี้ (อาวีพร ปานทอง, 2558, หน้า 38)

วิธีทดสอบอัตราส่วนความควรจะเป็น (Likeliihood ratio test: LRT)

Thissen, Steinberg, and Wainer (1988) ได้เสนอวิธีการทดสอบอัตราส่วนความเป็นไปได้ (Likeliihood ratio test: LRT) สำหรับใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยจะทำการทดสอบความแตกต่างของผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม วิธีการทดสอบอัตราส่วนความเป็นไปได้ แบ่งออกเป็น 3 วิธีย่อย ๆ คือ 1) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบในรูปทั่วไป (General LR) เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยทำการประมาณค่าพารามิเตอร์ในโมเดลการตอบข้อสอบด้วยวิธีความเป็นไปได้สูงสุดแบบมาร์จินอล (Marginal maximum likelihood; MML) 2) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบในรูปลอกลิเนียร์ (Loglinear LR) เป็นการประมาณค่าพารามิเตอร์ด้วยวิธีความเป็นไปได้สูงสุด (Maximum likelihood; ML) 3) วิธีการทดสอบอัตราส่วนความเป็นไปได้ในทฤษฎีการตอบข้อสอบในรูปสารสนเทศที่มีขอบเขตจำกัด (Limited information LR) สำหรับวิธีนี้จะใช้การประมาณค่าพารามิเตอร์ในโมเดลการตอบแบบ Normal ogive ด้วยวิธีกำลังสองน้อยที่สุดในรูปทั่วไป (Generalized least squares: GLS) จะเห็นว่าทั้ง 3 วิธีดังกล่าวมาจะใช้การทดสอบอัตราส่วนความเป็นไปได้เพื่อทดสอบนัยสำคัญในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งผลความกลมกลืนของโมเดลเป็นที่รู้จักโดยทั่วไปในความกลมกลืนของฟังก์ชั่นในทฤษฎีการตอบข้อสอบ (IRT) ค่าความกลมกลืนของฟังก์ชั่นเป็นค่าดัชนีบอกว่าโมเดลดังกล่าวความเหมาะสมกับข้อมูลที่จะใช้ในกระบวนการประมาณค่าความเป็นไปได้สูงสุดสำหรับการประมาณค่าพารามิเตอร์ข้อสอบ (Camilli & Shepard, 1994)

หลักการตรวจสอบความเท่าเทียมกันของการวัดด้วยวิธีการทดสอบอัตราส่วนความเป็นไปได้ จะทำการเปรียบเทียบระหว่าง 2 โมเดล คือ โมเดลพื้นฐาน (Compact) และโมเดลเปรียบเทียบ (Augmented) โดยในโมเดลแรกสมมติให้มีกลุ่มผู้สอบแตกต่างกัน ดังนั้นจึงมีการบังคับให้พารามิเตอร์ของข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีค่าเท่ากัน (Group 1-Item 1 และ Group 2-Item 1) (Thissen, Steinberg, & Wainer, 1988, 1993) โมเดลพื้นฐานนี้เป็นการจัดเตรียมค่าความเป็นไปได้พื้นฐานสำหรับค่าความกลมกลืนของพารามิเตอร์ข้อสอบในโมเดล จากนั้นจะกระทำการทำข้อสอบแต่ละข้อที่ทำหน้าที่ต่างกันโดยแยกข้อมูลวิเคราะห์ในแต่ละข้อ ซึ่งคล้ายกับการประมาณค่าพารามิเตอร์ของข้อสอบที่บังคับ

ให้เท่ากันข้ามกลุ่ม ยกเว้นค่าพารามิเตอร์ของข้อสอบเริ่มต้น การทดสอบการทำหน้าที่ต่างกันของข้อสอบ โดยโมเดลเปรียบเทียบจัดเตรียมค่าความเป็นไปได้ที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ ข้อที่ i ในแต่ละประชากร ค่าความเป็นไปได้นี้ สามารถเปรียบเทียบทุกพารามิเตอร์ของข้อสอบทั้งหมดที่บังคับให้ข้ามกลุ่มของโมเดลพื้นฐาน

สมมติฐานในการทดสอบสำหรับโมเดลของพารามิเตอร์ของข้อสอบโดยใช้สูตร ดังนี้

$H_0 : a_{jR} = a_{jF}$ และ $b_{jR} = b_{jF}$ สำหรับทุก j

H_A : พารามิเตอร์ของข้อสอบข้อที่ j ของทั้งสองกลุ่มไม่เท่ากันอย่างน้อย 1 พารามิเตอร์

สูตรการทดสอบอัตราส่วนความเป็นไปได้ของโมเดล แสดงได้ดังนี้

$$LR = \frac{L^*(Model_C)}{L^*(Model_A)}$$

เมื่อ $L^*(Model_C)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลพื้นฐาน (จะเป็นค่าพารามิเตอร์ที่น้อยกว่า)

$L^*(Model_A)$ แทน แทนฟังก์ชันความเป็นไปได้ของโมเดลเปรียบเทียบที่ยอมให้พารามิเตอร์ข้อสอบของข้อสอบข้อที่ j ข้ามกลุ่มผู้สอบมีความหลากหลาย

นั่นคือในโมเดลพื้นฐาน (Compact) จะประกอบด้วยข้อสอบที่ทำหน้าที่ไม่ต่างกัน สำหรับในโมเดลเปรียบเทียบจะประกอบด้วยข้อสอบที่มีค่าพารามิเตอร์ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีค่าเปลี่ยนไปตามกลุ่มโมเดลเปรียบเทียบ (Augmented) อาจมีข้อสอบจำนวน 1 ข้อ หรือมากกว่าที่ทำหน้าที่ต่างกัน นอกจากนี้ระหว่าง 2 โมเดลจะต้องมีข้อสอบร่วมซึ่งเป็นข้อสอบที่สมมติว่าทำหน้าที่ไม่ต่างกัน จากนั้นจึงทำการเปรียบเทียบระหว่าง 2 โมเดลด้วยสถิติอัตราส่วนความเป็นไปได้ ดังนี้

$$G_j^2 = -2Ln(LR)$$

$$G_j^2 = -2Ln(L^*(Model_C) - \left| -2Ln(L^*(Model_A)) \right|$$

เมื่อ G_j^2 แทน สถิติอัตราส่วนความเป็นไปได้ของข้อสอบข้อที่ i

$L^*(Model_C)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลพื้นฐาน (Compact)

$L^*(Model_A)$ แทน ฟังก์ชันความเป็นไปได้ของโมเดลเปรียบเทียบ (Augmented)

โดยทั่วไปแล้ว $L^*(Model_C) < L^*(Model_A)$ และสถิติ $G_j^2 > 0$ มีการแจกแจงแบบไค-สแควร์ ซึ่งระดับของความเป็นอิสระเท่ากับจำนวนพารามิเตอร์ในโมเดล L ($Model_C$) และโมเดล L ($Model_A$) ดัชนีความไม่กลมกลืนในระดับนัยสำคัญทางสถิติของผลการทดสอบจะชี้บอกว่าโมเดลพื้นฐานกลมกลืนน้อยกว่าโมเดลเปรียบเทียบอย่างมีนัยสำคัญทางสถิติในการใช้การทดสอบแบบอัตราส่วนความเป็นไปได้ ค่าไค-สแควร์ของข้อสอบเหล่านั้นมีนัยสำคัญทางสถิติแสดงว่า มีการทำหน้าที่ต่างกันของข้อสอบ

ในการเปรียบเทียบค่าพารามิเตอร์ในโมเดลที่ตั้งสมมติฐานให้มีค่าเท่ากันระหว่างกลุ่มเปรียบเทียบและกลุ่มอ้างอิง สถิติที่ใช้ทดสอบอัตราส่วนความควรจะเป็น คือ -2 log *likelihood* (-2LL) และค่าที่ได้เปรียบเทียบการแจกแจง χ^2 กับ df เท่ากับจำนวนพารามิเตอร์ที่ทดสอบ ถ้าผลปรากฏว่าปฏิเสธสมมติฐานว่างแสดงว่าข้อสอบทำหน้าที่ต่างกัน นั้นคือค่าพารามิเตอร์อำนาจจำแนก (a_i) หรือ ค่าพารามิเตอร์ลำดับขั้น (b_{jk}) หรือทั้งสองค่า มีค่าไม่เท่ากันระหว่างกลุ่มอ้างอิงกับกลุ่มสนใจ (Woods, 2011) ค่า Chi-squared (χ^2) จะมี Degree or Freedom (df) เท่ากับจำนวนตัวแปรอิสระ หรือจำนวนพารามิเตอร์ของข้อสอบ ค่า χ^2 สามารถคำนวณได้ดังนี้

$$\chi^2 = -2\left[LL(C) - LL(A)\right]$$

โดยที่ LL(C) คือ ค่า Log-likelihood เมื่อ โมเดลไม่มีข้อจำกัด (Unrestricted model) และ LL(A) คือค่า Log-likelihood เมื่อโมเดลมีข้อจำกัด (Restricted model) เนื่องจากค่า LL(C) มากกว่า LL(A) ดังนั้น χ^2 จึงมีค่าเป็นบวกเสมอ การทดสอบอัตราความควรจะเป็นจะมีลักษณะ

คล้ายกับแบบ SIBTEST ในส่วนของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สถิติที่ใช้ของ Log-likelihood เขียนสูตรได้ ดังนี้ (Finch & French, 2007)

$$LL_{equal} = \sum_{G=1}^{2}\sum_{p=1}^{N} \ln\left[\sum_{I}^{q}\prod_{i=1}^{nitems}\left(T_{iG}\left(\mu_{ipG}\right)\varphi_{G}(\theta)d\theta\right)\right]$$

เมื่อ $T_{iG}\left(\mu_{ipG}\right)$ แทน ค่าพารามิเตอร์ของกลุ่ม G ที่ถูกกำหนดให้มีค่าพารามิเตอร์เท่ากันสำหรับทั้งสองกลุ่ม

$\varphi_{G}(\theta)$ แทน การแจกแจงของพารามิเตอร์คุณลักษณะของผู้ตอบ (θ) ของกลุ่ม G

ให้ U_{jki} แทน เมตริกซ์ของการตอบสนองข้อสอบ (Response pattern) คนที่ i U_{jki}=1 เมื่อตอบตัวเลือก k ในข้อที่ j และตัวเลือกอื่น ๆ ในข้อที่ j ให้ U_{jki} = 0 โดยให้ผู้ตอบเป็นอิสระกัน ความน่าจะเป็นแบบมีเงื่อนไขของการตอบสนองข้อสอบคนที่ i กำหนดให้เป็น θ การตอบตัวเลือก m และจำนวนข้อทั้งหมด n กำหนดให้หา ความน่าจะเป็นร่วมของเมตริกซ์ $U_i = \left(U_{jk}\right)_i$ (Joint probability) ดังนี้

$$P\left(U_i|\theta\right) = P\left[\left(U_{jk}\right)|\theta\right] = \prod_{j=1}^{n}\prod_{k=1}^{m}\left[P_{jk}(\theta)\right]^{U_{jki}}$$

เมื่อ U_{jki} = 1 ถ้าเลือกตอบตัวเลือกที่ k ในข้อที่ j และมีค่าเป็น 0 เมื่อเลือกตัวเลือกอื่น ๆ สำหรับผู้สอบที่ทำการสุ่มจากประชากรในการแจกแจงปกติของคุณลักษณะแฝง และ $\varphi(\theta)$ แทน ความน่าจะเป็นที่จะเกิดเหตุการณ์ในการตอบข้อที่ j (Marginal probability) คือ

$$P_i\left[(U_{jk})\right] = \int_{-\infty}^{\infty} P_i\left[(U_{jk})\middle|\theta\right]\varphi(\theta)d(\theta)$$

ถ้าผู้ตอบตอบข้อคำถามทั้งหมด n ข้อที่มีตัวเลือกทั้งสิ้น m ตัวเลือก รูปแบบการตอบสนองข้อสอบของคนที่ i สามารถระบุเป็นหนึ่งใน m^n ของรูปแบบทั้งหมด ดังนั้นกำหนดให้

r_i แทนจำนวนการตอบสนองข้อสอบ ในรูปแบบที่ i และให้ N แทน จำนวนตัวอย่างผู้ตอบทั้งหมดที่สุ่มมาจากประชากร ดังนั้น r_i แทน การแจกแจงแบบพหุนาม (Multinomial distribution) กับพารามิเตอร์ N และ $P_i\left[(U_{jk})\right]$ ดังสมการต่อไปนี้

$$L=\frac{N!}{\prod_{i=1}^{m^n} r!}\prod_{i=1}^{m^n}\left\{P_i\left[(U_{jk})\right]\right\}^n$$

$$\ln L=\ln N!-\sum_{i=1}^{m^n}\ln r_i!+\sum_{i=1}^{m^n}\ln r_i P_i[(U_{jk})]$$

สมการไลด์ลิฮูดสำหรับ $\hat{a}_j, \hat{b}_j$ และ $\hat{d}_k$ สามารถหาค่าได้จากอนุพันธ์ย่อยของสมการข้างต้นโดยการแทนค่าและกำหนดให้สมการมีค่าเป็นศูนย์

การประมาณค่าพารามิเตอร์ กำหนดให้ υ_j แทนพารามิเตอร์ $\hat{a}_j$ หรือ $\hat{b}_j$ ถ้าแทนค่า υ_h หมายความว่า ค่าพารามิเตอร์ υ_j สำหรับข้อที่ j มีค่าเท่ากับ h สามารถหาค่าอนุพันธ์ของความควรจะเป็นในสมการข้างต้นได้ ดังนี้

$$\frac{\partial \ln L}{\partial \upsilon_h}=\sum_{i=1}^{m^n}\frac{r_i}{P_i(U_{jk})}\int_{-\infty}^{\infty}P_i[(U_{jk})\big|\theta]\sum_{k=1}^{m}\frac{\partial[P_i(\theta)^{U_{hkl}}}{\partial \upsilon_h}\frac{\varphi(\theta)d(\theta)}{[P_{hk}(\theta)]^{U_{hki}}} \qquad (1)$$

ผลของการตอบสนองข้อสอบโดยกำหนด l = 1,2,…,S เมื่อ $S\le \min(N,m^n)$

ถ้าจำนวนการตอบของผู้สอบตัวเลือก l กำหนดให้เป็น r_l แล้ว $\sum_{l=1}^{S} r_l = N$

ในอนุพันธ์อันดับหนึ่งของความควรจะเป็นในสมการ (1) สามารถประมาณค่าโดยวิธีของเกาส์เฮอร์มิทควอดาร์เจอร์ (Gauss-hermite quadrature) ได้ดังนี้

$$\frac{\partial \ln L}{\partial \upsilon_k}\approx\sum_{l=1}^{S}\sum_{f=1}^{F}\frac{r_l L_l(X_f)A(X_f)}{\tilde{P}_l}\sum_{k=1}^{m}\frac{\partial[P_i(\theta)]^{U_{hkl}}}{\partial \upsilon_h}\frac{1}{[P_{hk}(X_f)]^{U_{hkl}}} \qquad (2)$$

เมื่อ

$$\tilde{P} = \sum_{f=1}^{F} L_l(X_f) A(X_f)$$

และ

$$L_l(X_f) = \prod_{j=1}^{n} \prod_{k=1}^{m} [P_{jk}(X_f)]^{U_{jkl}}$$

ในสมการ (2) $A(X_f)$ แทนค่าน้ำหนักในการหาพื้นที่โดยวิธีเกาส์เฮอร์มิท และ X_f แทนจุดของการหาพื้นที่ (Quadarture point) (Stroud & Secrest, 1996) น้ำหนักของการพื้นที่ $A(X_f)$ มีการแจกแจงเป็นปกติมาตรฐานที่จุด X_f จะได้ว่า $\sum_{f=1}^{F} A(X_f) = 1$

เมื่อ F แทนจำนวนจุดในการหาพื้นที่ทั้งหมด เพราะ U_{hkl} มีค่าที่เป็นไปได้ lสองค่า คือ 1 และ 0 จากสมการ (2) สามารถเขียนใหม่ได้ว่า

$$\sum_{f=1}^{F} \sum_{k=1}^{m} \frac{\bar{r}_{hkf}}{P_{hk}(X_f)} \frac{\partial P_{hk}(X_f)}{\partial \upsilon_h}$$

เมื่อ

$$\bar{r}_{hkf} = \sum_{t=1}^{s} \frac{r_i L_i(X_f) A(X_f) U_{hkl}}{\tilde{P}_i}$$

และ $\bar{r}_{hkf}$ แทน จำนวนของการตอบตัวเลือก k ของข้อที่ h ณ จุดของการหาพื้นที่ที่ f การหาพารามิเตอร์จุดตัดของการเลือกตอบ (Threshold parameter estimate) ค่าจุดตัดของการเลือกตอบตัวเลือกที่ h แทนสัญลักษณ์คือ d_g ซึ่งก็คือค่าพารามิเตอร์ d_k สำหรับแต่ละค่าของตัวเลือกที่ k มีค่าเป็น g ดังนั้น $P_{jk}(\theta)(k = 1,2,\cdots,m)$ แสดงดังสมการต่อไปนี้ในการหาอนุพันธ์อันดับที่หนึ่งของฟังก์ชันความควรจะเป็น ในสมการ (3) หาค่าเป็น d_g

$$\frac{\partial L_l(X_f)}{\partial d_g} = L_l(X_f) \sum_{j=1}^{n} \sum_{k=1}^{m} \frac{U_{jkl}}{P_{jk}(X_f)} \frac{\partial P_{jk}(X_f)}{\partial d_g} \qquad (3)$$

ตอนที่ 7 งานวิจัยที่เกี่ยวข้อง

จากการศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องกับแนวทางการจัดการข้อมูลสูญหายในทฤษฎีการตอบสนองข้อสอบ รวมทั้งการประมาณค่าพารามิเตอร์ รายละเอียด ดังนี้

งานวิจัยในประเทศ

กมลทิพย์ ศรีหาเศษ (2555) ได้ศึกษาเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหายเพื่อนำมาใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ด้วยวิธีจัดการข้อมูลสูญหาย 3 วิธี คือ วิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) วิธีค่าคาดหมายสูงสุด (Expectation-maximization algorithm: EM) และวิธีการประมาณค่าที่เป็นไปได้สูงสุด (Maximum likelihood estimation: ML) ผลการวิจัยได้ดังนี้

1. ผลการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบด้วยวิธีจัดการข้อมูลสูญหาย 3 วิธี พบว่า ภายใต้สภาวะการสูญหายทั้ง 2 ประเภท ค่าพารามิเตอร์ของข้อสอบ จากการประมาณค่าทั้ง 3 วิธี มีค่าใกล้เคียงกัน โดยวิธี EM ให้ค่าอำนาจจำแนก (a) ใกล้เคียงค่าพารามิเตอร์ที่แท้จริงของข้อสอบมากที่สุด ส่วนค่าความยากของข้อสอบ (b) และค่าโอกาสในการเดา (c) วิธี ML ให้ค่าพารามิเตอร์ใกล้เคียงค่าพารามิเตอร์ที่แท้จริงมากที่สุด ภายใต้สภาวะการสูญหายทั้ง 2 ประเภท ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) จากการประมาณค่าทั้ง 3 วิธี นั้น วิธี ML ให้ค่าพารามิเตอร์ใกล้เคียงค่าพารามิเตอร์ที่แท้จริงมากที่สุด

2. ผลการเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ ของข้อสอบและความสามารถของผู้สอบด้วยวิธีที่แตกต่างกัน 3 วิธี พบว่า ผลที่ได้จากการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบจากวิธีจัดการข้อมูลสูญหายทั้ง 3 วิธี ส่วนใหญ่ให้คุณภาพของการประมาณค่าพารามิเตอร์ ของข้อสอบและความสามารถของผู้สอบไม่ต่างกัน โดยภายใต้สภาวะการสูญหายทั้ง 2 แบบ วิธี EM ให้ค่า BIAS และ RMSE ในการประมาณค่าอำนาจจำแนกของข้อสอบ (a) น้อยที่สุด ส่วนวิธี ML ให้ค่า BIAS และ RMSE ในการประมาณค่าความยากของข้อสอบ (b) และค่าโอกาสในการเดา (c) น้อยที่สุด ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) วิธี ML ให้ค่าดัชนี BIAS และ RMSE ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) น้อยที่สุด ส่วนภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) วิธี MI ให้ค่าดัชนี BIAS ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) น้อยที่สุด แต่เมื่อพิจารณา ดัชนี RMSE ทั้ง 3 วิธี ให้ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ใกล้เคียงกัน

จำลอง วงษ์ประเสริฐ (2554) ได้ศึกษาพัฒนาวิธีการประมาณข้อมูลสูญหาย โดยการถ่วงน้ำหนักแบบวนซ้ำด้วยวิธีของแจ็คไนฟ์และการวิเคราะห์การถดถอย (IWJR) และเปรียบเทียบประสิทธิภาพในการประมาณค่าเฉลี่ย ความแปรปรวน สัมประสิทธิ์สหสัมพันธ์ และอำนาจการทดสอบ ภายใต้ข้อมูลสูญหายแบบสุ่มอย่างสมบูรณ์และการสุ่มตัวอย่างแบบง่าย กับการตัดข้อมูลสูญหายแบบลิสท์ไวส์ (LD) วิธีการประมาณข้อมูลสูญหายด้วยค่าเฉลี่ย (MI) วิธีการประมาณข้อมูลสูญหายด้วยการถดถอย (RI) และวิธีการจัดการข้อมูลสูญหายแบบอีพีอาร์ (EPR) โดยใช้ข้อมูลจากการจำลองและข้อมูลจริง การเปรียบเทียบกระทำภายใต้เงื่อนไขดังต่อไปนี้ 1) ขนาดตัวอย่าง 3 ขนาด (100 200 และ 500) 2) ระดับความสัมพันธ์ระหว่างตัวแปร 3 ระดับ (ต่ำ ปานกลางและสูง) และ 3) ร้อยละของข้อมูลสูญหาย 4 ระดับ (ร้อยละ 5 10 15 และ 20) และใช้ข้อมูลจากการจำลองศึกษาปฏิสัมพันธ์ระหว่างวิธีการประมาณข้อมูลสูญหายขนาดตัวอย่าง ระดับความสัมพันธ์ระหว่างตัวแปรและร้อยละข้อมูลสูญหาย ที่ระดับนัยสำคัญ .05 พบว่า IWJR มีความแกร่งต่อขนาดตัวอย่าง ความสัมพันธ์ระหว่างตัวแปรและร้อยละข้อมูลสูญหาย ในการประมาณค่าเฉลี่ย ความแปรปรวนและสัมประสิทธิ์สหสัมพันธ์ และมีประสิทธิภาพสูง เมื่อร้อยละข้อมูลสูญหายมีจำนวนน้อย

ณรงค์ โพธิ และสมชาย ปราการเจริญ (2553, หน้า 146-156) ได้ศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายโดยวิธีการวิเคราะห์จำแนกประเภท ต้นไม้การตัดสินใจและค่าเฉลี่ย โดยมีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสูญหายโดยวิธีการทางสถิติ ได้แก่ 1) การวิเคราะห์จำแนกประเภท 2) ต้นไม้การตัดสินใจ 3) ค่าเฉลี่ย ซึ่งใช้เกณฑ์เปรียบเทียบประสิทธิภาพของการประมาณค่าสูญหายโดยใช้ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (MMRE) ข้อมูลที่ใช้ในการศึกษาวิจัย คือฐานข้อมูลภาพถ่ายก้อนเนื้อเต้านม (Mammographic mass) ของ UCI Machine learning repository data set จำนวน 800 ชุด (Case) วิธีการวิจัยมีดังต่อไปนี้ 1) แบ่งฐานข้อมูลเป็น 2 กลุ่มที่ระดับความเชื่อมั่น 95% ด้วยวิธีของ Taro Yamane คือกลุ่มข้อมูลเรียนรู้ (Training data) 533 ชุด และกลุ่มข้อมูลทดสอบ (Testing data) 267 ชุด 2) นำข้อมูลเรียนรู้มาหาสมการทดแทนค่าสูญหายของแต่ละวิธี 3) นำข้อมูลทดสอบมาแทนค่าสมการที่ได้เพื่อหาค่าทดแทนค่าสูญหาย 4) คำนวณหาค่าความคลาดเคลื่อนสัมพัทธ์และค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ของแต่ละวิธีผลการวิจัยพบว่า วิธีที่ดีที่สุดเรียงตามลำดับจากมากไปน้อยเป็นดังนี้ 1) การวิเคราะห์จำแนกประเภท (MMRE = 26.56%) 2) ต้นไม้การตัดสินใจ (MMRE = 33.30%) และ 3) ค่าเฉลี่ย (MMRE = 63.26%) ดังนั้นการวิเคราะห์จำแนกประเภท

จึงมีความเหมาะสมกับข้อมูลที่มีความสัมพันธ์กันและมีการกระจายออกเป็นกลุ่มที่มีความแตกต่างกันอย่างชัดเจนการประมาณค่าสูญหาย จึงจะมีความใกล้เคียงกับค่าของข้อมูลจริงมากที่สุด

เชาว์ อินใย (2552, หน้า 111-118) ได้ศึกษาการพัฒนาวิธีการประมาณข้อมูลสูญหายแบบอีพีอาร์ กับวิธีการประมาณข้อมูลสูญหายด้วยวิธีอีเอ็ม และการตัดข้อมูลสูญหายแบบลิสท์ไวส์ตามวิธีการ โดยทำการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่ม และแบบหลายขั้นตอน ที่ระดับความสัมพันธ์ระหว่างตัวแปรระดับต่ำ (r = 0.30) ปานกลาง (r = 0.50) และสูง (r = 0.70) และจำนวนข้อมูลสูญหาย 5% 10% 20% และ 30% และศึกษาปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการประมาณข้อมูลสูญหาย สัดส่วนข้อมูลสูญหาย และระดับความสัมพันธ์ระหว่างตัวแปรที่มีต่อความแม่นยำของการประมาณค่าเฉลี่ยเลขคณิต ความแปรปรวนและสัมประสิทธิ์สหสัมพันธ์ ข้อมูลที่ใช้ศึกษามีลักษณะการแจกแจงแบบปกติ สองตัวแปร และใช้วิธีมอนติคาร์โล ผลการศึกษาพบว่าวิธีการประมาณข้อมูลสูญหายแบบอีพีอาร์ ได้ค่าความแม่นยำของการประมาณค่าเฉลี่ยเลขคณิตไม่แตกต่างจากวิธีการประมาณข้อมูลสูญหายด้วยวิธีอีเอ็ม และการตัดข้อมูลสูญหายแบบลิสท์ไวส์ แต่จะมีความแม่นยำสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่ม และแบบหลายขั้นตอน สัดส่วนข้อมูลสูญหายสูงสุด 5% และ 20% ทุกระดับความสัมพันธ์ระหว่างตัวแปร และเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นแบบกลุ่ม และแบบหลายขั้นตอน สัดส่วนข้อมูลสูญหายสูงสุด 30% ระดับความสัมพันธ์ระหว่างตัวแปรระดับสูง วิธีการประมาณข้อมูลสูญหายแบบอีพีอาร์ และการตัดข้อมูลสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของการประมาณค่าความแปรปรวนแตกต่างจากวิธีการประมาณข้อมูลสูญหายด้วยวิธีอีเอ็มอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และมีความแม่นยำสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบกลุ่ม และแบบหลายขั้นตอน สัดส่วนข้อมูลสูญหายร้อยละ 5% 10% 20% และ 30% ทุกระดับความสัมพันธ์ระหว่างตัวแปรวิธีการประมาณข้อมูลสูญหายแบบอีพีอาร์ กับวิธีประมาณข้อมูลสูญหายด้วยวิธีอีเอ็ม และการตัดข้อมูลสูญหายแบบลิสท์ไวส์ได้ค่าความแม่นยำของการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ไม่แตกต่างจากวิธีอื่น และมีความแม่นยำสูงที่สุดเมื่อใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นแบบกลุ่ม และแบบหลายขั้นตอน สัดส่วนข้อมูลสูญหายสูงสุดร้อยละ 5% 10% 20% และ 30% ระดับความสัมพันธ์ระหว่างตัวแปรสูง ค่าอำนาจการทดสอบเมื่อใช้การทดสอบความสัมพันธ์ การทดสอบที และการทดสอบเอฟ ไม่แตกต่างกันทั้ง 3 วิธี ไม่มีปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการประมาณข้อมูลสูญหาย และระดับความสัมพันธ์ระหว่างตัวแปรที่มีต่อความแม่นยำของการประมาณค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์

งานวิจัยต่างประเทศ

โดยในการค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องกับแนวทางการจัดการข้อมูลสูญหายในทฤษฎีการตอบสนองข้อสอบ ผู้วิจัยได้นำเสนอรายละเอียดแสดง ดังนี้

Schmitt et al. (2015) ได้ศึกษาเปรียบเทียบวิธีการจัดการข้อมูลสูญหาย 6 วิธี ดังนี้ 1) วิธีค่าเฉลี่ย (Mean) 2) วิธีสมาชิกที่ใกล้ที่สุด (K-nearest neighbors (KNN)) 3) วิธีฟุซซีเคมีน (Fuzzy K-means (FKM)) 4) วิธีการแตกค่า แบบเอกฐาน (singular value decomposition (SVD)) 5) วิธีการวิเคราะห์องค์ประกอบพื้นฐานแบบเบย์ (Bayesian principal component analysis (bPCA)) และ 6) วิธี multiple imputations by chained equations (MICE) เพื่อจัดการกับข้อมูลสูญหายให้มีความเหมาะสมเพื่อใช้ในการวิเคราะห์ข้อมูลได้ถูกต้อง และมีประสิทธิภาพ โดยทำการศึกษากับข้อมูลจริง 4 ชุด ซึ่งมีจำนวนตัวแปรแตกต่างกัน (4-65 ตัวแปร) ภายใต้เงื่อนไขการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) และมีอัตราการสูญหาย 5% 15% 25% 35% 45% และมีจำนวนตัวอย่าง 80, 89, 100, 129 ผลการศึกษาพบว่า วิธี fuzzy K-means (FKM) และ bayesian principal component analysis (bPCA) เป็นวิธีที่เหมาะสมดีมากในการนำไปใช้ โดยเฉพาะวิธี Fuzzy K-means (FKM) มีนัยสำคัญความได้เปรียบเมื่อข้อมูลมีขนาดเล็ก แต่วิธี bayesian principal component analysis (bPCA) ให้ผลดีมากเมื่อข้อมูลมีขนาดใหญ่

Andreis and Ferrari (2012) ศึกษาผลกระทบสำหรับการประมาณค่าพารามิเตอร์จากวิธีจัดการข้อมูลสูญหายสำหรับโมเดลโลจิสติกแบบพหุมิติ ชนิด 2 พารามิเตอร์ (Multidimensional two-parameter logistic model: M2PL) โดยใช้การจำลองข้อมูลในการศึกษาวิธีจัดการข้อมูลสูญหาย 4 วิธี คือ 1) Complete data (CD) 2) Forward imputation (FI) 3) Miss forest (MF) 4) Multivariate imputation by chained equations (MICE) ภายใต้เงื่อนไขการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) การสูญหายอย่างสุ่ม (MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) และมีอัตราการสูญหาย 5% 10% 30% และจำนวนข้อสอบ 10 ข้อ ผลการวิจัยพบว่า วิธี Complete data ให้ผลลัพธ์แสดงค่าพารามิเตอร์ซึ่งเป็นที่น่าพอใจมาก ส่วนวิธี Forward imputation (FI) และ Miss forest (MF) ควรเลือกใช้เมื่อพบว่าข้อมูลสูญหายมาก โดยพิจารณาจากค่า Bias แต่วิธี Miss forest (MF) และ Multivariate imputation by chained equations (MICE) จะดีกว่าถ้าสนใจจะประมาณค่าพารามิเตอร์ a_1, a_2

Finch (2011) ทำการศึกษาการเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่แตกต่างกันของข้อสอบจากการจัดการข้อมูลสูญหายด้วยวิธี Multiple imputation ทำการจำลองข้อมูลภายใต้โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic

model: 3PL) และวิธีการจัดการข้อมูลสูญหายที่แตกต่างกัน 4 วิธี คือ 1) ข้อมูลสมบูรณ์ (Complete data) 2) การลบตามรายการ (Liswise deletion) 3) การละเว้นการตอบแบบตอบผิด (Omitted as incorrect) 4) การประมาณค่าพหุ (Multiple Imputation) วิธีตรวจสอบการทำหน้าที่แตกต่างกันของข้อสอบ 3 วิธี คือ 1) Mantel-haenszel 2) Logistic regression 3) SIBTEST ความยาวแบบสอบ จำนวน 40 ข้อ ภายใต้เงื่อนไขที่แปรเปลี่ยน 6 ปัจจัยดังนี้ สภาวะการสูญหาย ขนาดกลุ่มตัวอย่าง อัตราการสูญหาย ขนาดการทำหน้าที่ต่างกันของข้อสอบ และระดับความยากของข้อสอบ ผลการศึกษาเมื่อพิจารณาจากค่าอัตราความคลาดเคลื่อนประเภทที่ 1(Type I error) พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้

ปัจจัยขนาดกลุ่มตัวอย่าง ระดับผลกระทบ วิธีจัดการข้อมูลสูญหาย และสภาวะการข้อมูลสูญหาย พบว่าเมื่อระดับความยากและ อัตราการสูญหายที่สูงขึ้น วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธีสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ไม่แตกต่างกัน โดยเฉพาะภายใต้สภาวะการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) และภายใต้สภาวะการข้อมูลสูญหายอย่างสุ่ม (MAR) และภายใต้ข้อมูลที่มีการละเว้นการตอบแบบตอบผิด (Omitted as incorrect) ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 มีค่าสูงขึ้น อยู่ในช่วง 0.14-0.30 และยังมีค่าเพิ่มขึ้นเมื่อขนาดตัวอย่าง และระดับขนาดอิทธิพลเพิ่มมากขึ้น

ปัจจัยระดับความยาก และสภาวะการสูญหา พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้ข้อมูลที่มีการละเว้นการตอบแบบตอบผิด (Omitted as incorrect) อัตราความคลาดเคลื่อนประเภทที่ 1 เป็นไปตามเกณฑ์ ($\leq$ 0.05) นั่นคือวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธี สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ระบุการทำหน้าที่ต่างกันของข้อสอบในข้อที่ไม่มีการทำหน้าที่ต่างกันของข้อสอบได้จริง

ปัจจัยอัตราการสูญหายที่ต่างกัน พบว่า ภายใต้ข้อมูลที่มีการสูญหายจากการละเว้นการตอบแบบตอบผิด (Omitted as incorrect) อัตราความคลาดเคลื่อนประเภทที่ 1 เป็นไปตามเกณฑ์ปกติ ($\leq$ 0.05) โดยเฉพาะที่อัตราความคลาดเคลื่อน 15%

Robitzsch and Rupp (2009) ศึกษาผลกระทบของข้อมูลสูญหายในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยศึกษากรณีการวิเคราะห์ด้วยวิธี Mantel-haenszel และ Logistic regression โดยใช้การจำลองข้อมูลในการศึกษาวิธีจัดการข้อมูลสูญหาย 4 วิธีคือ 1) การลบตามรายการ (Listwise deletion) 2) Zero imputation 3) Two-way imputation และ 4) Response function imputation ว่ามีปฏิสัมพันธ์อย่างไรกับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งสองวิธีคือ Mantel-haenszel statistic และ Logistic regression

analysis ภายใต้กลไกการสูญหายสามแบบคือ การสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) การสูญหายอย่างสุ่ม (MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ที่ทำให้เกิดการประมาณค่าสูงหรือต่ำ ของขนาดผลกระทบและอัตราการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และอัตราที่มีอิทธิพลมากที่สุดในการอธิบายการเปลี่ยนแปลงความลำเอียง รากที่สองของความคลาดเคลื่อน กำลังสองเฉลี่ย (Root mean square error) และอัตราการปฏิเสธ การจัดกระทำที่ไม่ถูกต้องของข้อมูลสูญหายอาจนำไปสู่การเพิ่มขึ้นของอัตราความคลาดเคลื่อนแบบ Type I และ Type II แต่อย่างไรก็ตาม Robitzsch and Rupp ให้ข้อคิดเห็นว่า ตัวเลือกระหว่างวิธีการตรวจสอบการทำหน้าที่ต่างกันสองวิธีที่ใช้ในการศึกษาครั้งนี้ ยังไม่ใช่วิธีที่ดีที่สุด

Finch (2008) ได้ศึกษาการประเมินผลกระทบของวิธีการที่ใช้ในการจัดการกับข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ของข้อสอบสำหรับ โดยใช้โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 3 พารามิเตอร์ (Three-parameter logistic model: 3PL) ทำการศึกษาโดยใช้การจำลองข้อมูลที่มีการศึกษามาก่อน โดยการทดสอบผลกระทบของการสูญหายของการตอบสนองข้อสอบในการประมาณค่าพารามิเตอร์ความยากและอำนาจจำแนก โดยใช้วิธีการจัดการข้อมูลสูญหายด้วยการ 4 วิธี คือ 1) Multiple imputation 2) Response function imputation 3) EM algorithm 4) Corrected item mean substitution imputation ผลการศึกษาพบว่าแต่ละวิธีมีความหลากหลายของระดับการอภิปรายประสิทธิผลในการประมาณค่าทดแทนข้อมูลความถูกต้องของการประมาณค่าตัวอย่างของพารามิเตอร์ความยาก และอำนาจจำแนกของข้อสอบตามลำดับ

Allison (2006) ศึกษาและขยายผลการศึกษางานของ Ake (2005) ต่อโดยทดสอบการใช้การประมาณค่าทดแทนพหุอิงการแจกแจงปกติ (Normal based multiple imputation) กับข้อมูลที่มีการตรวจให้คะแนนแบบ 2 ค่า (Dichotomous) โดยทำการศึกษาภายใต้การจำลองข้อมูล ซึ่งวิธีการที่ใช้ในการจัดการข้อมูลสูญหาย 2 วิธี คือ 1) วิธี listwise deletion 2) วิธี Linear imputation without rounding 3) วิธี Linear imputation rounding 4) วิธี Logistic regression imputation 5) วิธี Discriminant function imputation ผลการศึกษาของ Allison พบว่า เมื่อการประมาณค่าทดแทนพหุหมุนรอบใกล้จำนวนเต็ม 0 หรือ 1 มากที่สุดแล้ว ผลลัพธ์ที่ได้จากการประมาณค่าของค่าเฉลี่ยจะทำให้เกิดความลำเอียง และในทางตรงกันข้าม หากค่าที่ได้จากการประมาณค่าทดแทนพหุไม่ได้หมุนรอบใกล้จำนวนเต็ม 0 หรือ 1 แล้วผลลัพธ์ที่ได้จากการประมาณค่าของค่าเฉลี่ยที่จะทำให้เกิดความลำเอียงจะพบน้อยมาก

Ake (2005) ศึกษาผลกระทบของการใช้การประมาณค่าทดแทนพหุอิงโมเดลการแจกแจงปกติ (Normal based multiple imputation) ศึกษาภายใต้การจำลองข้อมูลกับข้อมูล Ordinal ขนาดตัวอย่าง 200 3,000 5,000 และภายใต้อัตราการสูญหาย 1% 5% 10% 20% และ 40% พบว่า เมื่อค่าประมาณทดแทนหมุนรอบการประมาณค่าจำนวนเต็มของค่าเฉลี่ยที่ได้จากการเพิ่มความลำเอียง เช่น ร้อยละของข้อมูลการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) ที่เพิ่มขึ้น ในทางตรงกันข้าม Ake พบว่า เมื่อการหมุนรอบไม่ได้ถูกจัดกระทำแล้วจะมีความลำเอียงน้อยมากในการประมาณค่าเฉลี่ย

Velicer and Colby (2005) ทำการศึกษาเปรียบเทียบวิธีจัดการข้อมูลสูญหายในการวิเคราะห์อนุกรมเวลาแบบ Arima (Arima time-series analysis) เนื่องจากข้อมูลสูญหายเป็นปัญหาทั่วไป ในทางปฏิบัติสำหรับการออกแบบการศึกษาระยะยาว การวิเคราะห์อนุกรมเวลาเป็นวิธีการศึกษาที่ใช้เวลานานในการสังเกตหน่วยตัวอย่าง ซึ่งในการศึกษาของ Velicer and Colby ทำการศึกษาวิธีการจัดการข้อมูลสูญหายด้วยวิธีการที่แตกต่างกัน 4 วิธี คือ 1) การลบทิ้ง (Deletion) 2) การแทนที่ค่าเฉลี่ย (Mean substitution) 3) mean of adjacent observations และ 4) การประมาณค่าที่เป็นไปได้สูงสุด (Maximum likelihood estimation) ทำการศึกษาภายใต้การจำลองข้อมูลอนุกรมเวลา จำนวน 100 หน่วยตัวอย่าง ภายใต้เงื่อนไขที่แตกต่างกันทั้งหมด 50 เงื่อนไข ดังนี้ 1) สหสัมพันธ์ภายในตัวเอง 5 ระดับ และ 2) ความชัน 2 ระดับและสัดส่วนข้อมูลสูญหาย 5 ระดับ คือ 10%, 20%, 30%, 40%, 50% และ 60% แล้วทำการเปรียบเทียบวิธีการโดยการพิจารณาความถูกต้องแม่นยำของการประมาณค่าพารามิเตอร์ 4 แบบ คือ 1) Level 2) Error variance 3) Degree of autocorrelation และ 4) Slope ผลการศึกษาพบว่า การประมาณค่าที่เป็นไปได้สูงสุดมีความถูกต้องแม่นยำในการประมาณค่าพารามิเตอร์มากที่สุดภายใต้เงื่อนไขการทดสอบทั้งหมด ส่วน Mean of adjacent observations เป็นวิธีที่มีความถูกต้องแม่นยำน้อยที่สุด ดังนั้น การประมาณค่าที่เป็นไปได้สูงสุดจึงเป็นวิธีการทางสถิติที่ดีสำหรับการประมาณค่าข้อมูลสูญหาย

Enders (2004) ศึกษาผลกระทบของวิธีในการจัดการกับข้อมูลสูญหายในการประมาณค่าความเที่ยงสำหรับข้อมูลแบบ Likert โดยการใช้ Cronbach's alpha ทำการศึกษาภายใต้การจำลองข้อมูล และวิธีการในการจัดการข้อมูลสูญหาย 4 วิธี คือ 1) วิธี EM algorithm 2) วิธี Listwise deletion 3) วิธี Pairwise deletion 4) วิธี Mean imputation ผลการศึกษา พบว่า วิธีค่าคาดหมายสูงสุด (EM) ให้ผลการศึกษาดีกว่าวิธีการลบตามรายการ (Listwise deletion) การลบเป็นรายคู่ (Pairwise deletion) การประมาณค่าทดแทนค่าเฉลี่ย (Mean imputation)

ที่มีความลำเอียง รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root mean square error) และขอบเขตช่วงความช่วงความเชื่อมั่นของสัมประสิทธิ์แอลฟ่า (Confidence interval coverage for coefficient alpha) โดย Enders ใช้การจำลองกลุ่มตัวอย่าง 200 คน ในการตอบสนองข้อสอบ 7 ข้อ และกำหนดช่วงการประมาณข้อมูลสูญหายที่ 11% ซึ่งจากผลการศึกษานั้น Ender เสนอแนะว่า ในการวิจัยผู้วิจัยควรพิจารณาผลกระทบของข้อมูลสูญหายในการประมาณค่าความเที่ยง และวิธีค่าคาดหมายสูงสุด (EM) เป็นวิธีที่ควรใช้ในการประมาณค่าทดแทนการตอบสนองข้อสอบที่สูญหาย

Gibson and Olejnik (2003) ศึกษาปัญหาของข้อมูลสูญหายในขั้นที่สองของโครงสร้างข้อมูล hierarchical 2 ระดับ ทำการศึกษาภายใต้การจำลองข้อมูลระดับมัธยมศึกษาตอนปลาย ค.ศ. 1982 และก่อนปี ค.ศ. 1982 โดยการจัดการข้อมูลสูญหายที่แตกต่างกัน 5 วิธีคือ 1) การลบตามรายการ (Listwise deletion) 2) การแทนที่ค่าเฉลี่ยรวมทุกกรณี (Overall mean substitution) 3) การประมาณค่าทดแทนค่าเฉลี่ยระดับกลุ่ม (Group mean imputation) 4) วิธีค่าคาดหมายสูงสุด (EM) และ 5) การประมาณค่าทดแทนพหุ (MI) ภายใต้เงื่อนไขที่แตกต่างกัน 4 เงื่อนไข คือ 1) จำนวนของตัวแปรขั้นที่สอง 2) ขนาดกลุ่มตัวอย่างขั้นที่สอง 3) สหสัมพันธ์จุดตัดความชัน และ 4) ร้อยละของข้อมูลสูญหาย ผลการศึกษาพบว่า วิธีการลบตามรายการ การประมาณค่าทดแทนค่าเฉลี่ยระดับกลุ่ม และวิธีค่าคาดหมายสูงสุด (EM) ให้ผลการดำเนินการเท่ากันในระดับดีสำหรับ ตัวแปรที่มีค่าสูญหาย ส่วนตัวแปรที่ไม่มีค่าสูญหายนั้น วิธีการลบตามรายการ และวิธีค่าคาดหมายสูงสุด (EM) ให้ผลการดำเนินการในระดับพอใจ มีเพียงวิธีการลบตามรายการเท่านั้นที่ให้ผลการดำเนินการในระดับดีในการประมาณค่าผลกระทบอย่างสุ่ม ยกเว้นกรณีขนาดกลุ่มตัวอย่างในขั้นที่สองสูญหายร้อยละ 30 หรือ 40

ตารางที่ 2-5 สรุปประเด็นรายละเอียดจากการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับแนวทางการจัดการกับข้อมูลสูญหาย

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
งานวิจัยในประเทศ				
กมลทิพย์ ศรีหาเศษ (2555)	ศึกษาเปรียบเทียบวิธีการจัดการข้อมูลสูญหาย	จำลองข้อมูล	1) วิธีการประมาณค่าทดแทนพหุ (Mmultiple imputation: MI) 2) วิธีค่าคาดหมายสูงสุด (Expectation-maximization algorithm: EM) 3) วิธีการประมาณค่าที่เป็นไปได้สูงสุด (Maximum likelihood estimation: ML)	**1) สภาวะการสูญหายของข้อมูล** MAR MNAR **2) ขนาดของกลุ่มตัวอย่าง** 1,500 3,000 4,500 **3) อัตราการสูญหายของข้อมูล** 5% 10% 15% และ 20% **4) ความยาวแบบสอบ** 20 ข้อ 40 ข้อ
จำลอง วงษ์ประเสริฐ (2554)	พัฒนาวิธีการประมาณข้อมูลสูญหาย	ข้อมูลจากการจำลองและข้อมูลจริง		**1) สภาวะการสูญหายของข้อมูล** MCAR **2) ขนาดของกลุ่มตัวอย่าง** 100 200 และ 500 **3) อัตราการสูญหายของข้อมูล** 5% 10% 15% และ 20% **3) ระดับความสัมพันธ์ระหว่างตัวแปร** ต่ำ ปานกลาง และ สูง

ตารางที่ 2-5 (ต่อ)

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
ณรงค์ โพธิและสมชาย ปราการเจริญ (2553)	ศึกษาเปรียบเทียบวิธีการจัดการข้อมูลสูญหาย	ข้อมูลจริง	1) การวิเคราะห์จำแนกประเภท 2) ต้นไม้การตัดสินใจ 3) ค่าเฉลี่ย	1) กลุ่มข้อมูลเรียนรู้ (Training data) 533 ชุด 2) กลุ่มข้อมูลทดสอบ (Testing data) 267 ชุด
เชาว์ อินใย (2552)	ศึกษาการพัฒนาวิธีการประมาณข้อมูลสูญหาย	ข้อมูลจากการจำลอง	1) การตัดข้อมูลออกแบบลิสท์ไวส์ (Listwise deletion) 2) การแทนค่าข้อมูลด้วยวิธีอีเอ็ม (Em algorithm) 3) การแทนค่าข้อมูลด้วยวิธีอีพีเอสเอสอี (Estimated parameter and smallest standard error)	**1) ความสัมพันธ์ระหว่างตัวแปร** - ความสัมพันธ์ระดับต่ำ (r=0.30) - ความสัมพันธ์ระดับปานกลาง(r=0.50) - ความสัมพันธ์ระดับสูง (r=1.0) **2) วิธีการสุ่ม** - การสุ่มแบบแบ่งชั้น - การสุ่มแบบกลุ่ม - การสุ่มแบบหลายขั้นตอน **3) อัตราการสูญหายของข้อมูล** 5% 10% 20% และ 30%

ตารางที่ 2-5 (ต่อ)

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
งานวิจัยต่างประเทศ				
Peter Schmitt, Jonas Mandel and Mickael Guedj (2015)	ศึกษาเปรียบเทียบวิธีการจัดการข้อมูลสูญหาย	ข้อมูลจริง	1) Mean 2) K-nearest neighbors (KNN) 3) Fuzzy K-means (FKM) 4) Singular value decomposition (SVD) 5) Bayesian principal component analysis (bPCA) 6) Multiple imputations by chained equations (MICE)	**1) สภาวะการสูญหายของข้อมูล** MCAR **2) ขนาดของกลุ่มตัวอย่าง** 80, 89, 100 และ 129 **3) อัตราการสูญหายของข้อมูล** 5% 15% 25% และ 45% **4) จำนวนตัวแปร** 4, 8, 60 และ 65
Federico Andreis and Pier Alda Ferrari (2012)	ผลกระทบในการประมาณค่าพารามิเตอร์	จำลองข้อมูล	1) Complete Data 2) Forward Imputation 3) Miss Forest 4) Multivariate Imputation by Chained Equations	**1) สภาวะการสูญหายของข้อมูล** MCAR, MAR และ MNAR **2) ขนาดของกลุ่มตัวอย่าง** 100 **3) อัตราการสูญหายของข้อมูล** 5% 10% และ 30% **4) ความยาวแบบสอบ** 10 **5) โมเดล** M2PL

ตารางที่ 2-5 (ต่อ)

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
Finch (2011)	การเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่แตกต่างกันของข้อสอบจากการจัดการข้อมูลสูญหายด้วยวิธี Multiple Imputation	การจำลองข้อมูลภายใต้ 3pL	1) Complete data 2) Liswise deletion 3) Omitted as incorrect 4) Multiple Imputation	**1) สภาวะการสูญหายของข้อมูล** MCAR MAR และ MNAR **2) ขนาดของกลุ่มตัวอย่าง** 500, 1,000 และ 2,000 **3) อัตราการสูญหายของข้อมูล** 5% และ 15% **4) ความยาวแบบสอบถาม** 40 **5) ขนาดอิทธิพล DIF** 0, และ 0.5 **6) วิธีตรวจสอบ DIF** - Mantel-Haenszel - logistic regression - SIBTEST **7) ระดับความยาก** - 1, 0, 1 **8) ขนาดการทำหน้าที่ต่างกันของข้อสอบ** 0, 0.3 และ 0.6

ตารางที่ 2-5 (ต่อ)

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
Robitzsch and Rupp (2009)	ผลกระทบของข้อมูลสูญหายในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	จำลองข้อมูล	1) Listwise deletion 2) Zero imputation 3) Two-way imputation 4) Two-way adjacent imputation 5) Response function imputation	**1) สภาวะการสูญหายของข้อมูล** MCAR MAR และ MNAR **2) ขนาดของกลุ่มตัวอย่าง** 500, 2,000 และ 8,000 **3) อัตราการสูญหายของข้อมูล** 10% และ 30% **4) ความยาวแบบสอบถาม** 20 และ 40 **5) ขนาดอิทธิพล DIF** 0, 0.2, 0.4 และ 0.6 **6) วิธีตรวจสอบ DIF** Mantel-Haenszel statistic และ logistic regression analysis **7) ผลกระทบ** Negative: focal N(0, 1), Reference: N(–.5, 1) Positive: focal N(0, 1), reference N(.5, 1) None: focal N(0, 1), reference N(0, 1)

ตารางที่ 2-5 (ต่อ)

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
Finch (2008)	ผลกระทบของวิธีการหลากหลายที่ใช้ในการจัดการกับข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ของข้อสอบ	จำลองข้อมูล	1) Multiple imputation 2) Response function imputation 3) EM algorithm 4) Corrected item mean substitution imputation	**1) สภาวะการสูญหายของข้อมูล** MAR และ MNAR **2) ขนาดกลุ่มตัวอย่าง** 500 และ 1,000 **3) อัตราการสูญหายของข้อมูล** 5%, 15% และ 30% **4) ความยาวแบบสอบ** 20
Allison (2006)	ทดสอบการใช้การประมาณค่าทดแทนพหุอิงการแจกแจงปกติ	จำลองข้อมูล	1) Listwise deletion 2) Linear imputation without rounding 3) Linear imputation rounding 4) Logistic regression imputation 5) Discriminant function imputation	**1) สภาวะการสูญหายของข้อมูล** MCAR และ MAR **2) ขนาดกลุ่มตัวอย่าง** 500 **3) อัตราการสูญหายของข้อมูล** 1%, 5%, 20% และ 50%
Ake (2005)	ผลกระทบของการใช้การประมาณค่าทดแทนพหุอิงโมเดลการแจกแจงปกติ	จำลองข้อมูล	1) Markov Chain Monte Carlo 2) Multiple imputation	**1) สภาวะการสูญหายของข้อมูล** MAR **2) ขนาดกลุ่มตัวอย่าง** 200 3,000 5,000 และ 35,000 **3) อัตราการสูญหายของข้อมูล** 1% 5% 10% 20% และ 40%

ตารางที่ 2-5 (ต่อ)

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
Velicer and Colby (2005)	วิธีจัดการข้อมูลสูญหายในการวิเคราะห์อนุกรมเวลา	จำลองข้อมูล	1) Deletion 2) Mean substitution 3) Mean of adjacent observations 4) Maximum likelihood estimation	**1) สภาวะการสูญหายของข้อมูล** MCAR **2) ขนาดกลุ่มตัวอย่าง** 100 **3) อัตราการสูญหายของข้อมูล** 10%, 20%, 30%, 40%, 50% และ 60% **4) Degrees of dependency** –0.80, –0.40, 0.00, 0.40 และ 0.80 **5) ลักษณะของความชัน** 0 และ a positive slope of 15
Enders (2004)	ผลกระทบของวิธีการในการจัดการกับข้อมูลสูญหายในการประมาณค่าความเที่ยงสำหรับข้อมูลแบบ Likert	จำลองข้อมูล	1) EM algorithm 2) Listwise deletion 3) Pairwise deletion 4) Mean imputation	**1) สภาวะการสูญหายของข้อมูล** MCAR และ MAR **2) ขนาดกลุ่มตัวอย่าง** 200 **3) อัตราการสูญหายของข้อมูล** 10% และ 20%

ตารางที่ 2-5 (ต่อ)

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
Leite and Beretvas (2004)	สมรรถนะของการประมาณค่าทดแทนพหุอิงการแจกแจงปกติโดยการจำลองสหสัมพันธ์	จำลองข้อมูล	1) Multiple imputation	**1) สภาวะการสูญหายของข้อมูล** MCAR, MAR และ MNAR **2) ขนาดกลุ่มตัวอย่าง** 1,000 **3) อัตราการสูญหายของข้อมูล** 2%, 15%, 20%, 30% และ 50% **4) ระดับความสัมพันธ์ภายในตัวแปร** 0.3 และ 0.8 **5) ระดับมาตรวัด Likert** 3, 5 และ 7
Gibson and Olejnik (2003)	ปัญหาของข้อมูลสูญหายในขั้นที่สองของโครงสร้างข้อมูล hierarchical 2 ระดับ	จำลองข้อมูล	1) Listwise deletion 2) Overall mean substitution 3) Group mean imputation 4) EM algorithm 5) Multiple imputation	**1) สภาวะการสูญหายของข้อมูล** MAR **2) ขนาดกลุ่มตัวอย่าง** 30 และ 160 **3) อัตราการสูญหายของข้อมูล** 10% และ 40% **4) สหสัมพันธ์ระหว่างค่าตัดแกน Y และความชัน** 0.2 และ 0.8 **5) จานวนของตัวแปรแต่ละระดับ** Level 1: 1 Level 2: 2 และ 4

ตารางที่ 2-5 (ต่อ)

ผู้วิจัย	ประเด็นวิจัย	ข้อมูลที่ใช้ศึกษา	วิธีจัดการข้อมูลสูญหาย	เงื่อนไขในการศึกษา
Enders (2001)	ทดสอบความเหมาะสมของผลการประมาณค่าสารสนเทศแบบเต็มรูปที่เป็นไปได้สูงสุด	จำลองข้อมูล	1) Full information maximum likelihood 2) Listwise deletion 3) Pairwise deletion 4) Mean imputation	**1) สภาวะการสูญหายของข้อมูล** MCAR, MAR และ MNAR **2) ขนาดกลุ่มตัวอย่าง** 100, 250, 500 และ 750 **3) อัตราการสูญหายของข้อมูล** 2%, 5%, 10%, 15% และ 25% **4) จำนวนของตัวแปรที่ใช้ศึกษา** 9 ตัว **5) ระดับน้ำหนักองค์ประกอบ** 0.40, 0.60 และ 0.80

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายสำหรับนำมาใช้ในการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก และความยาก) และพารามิเตอร์ความสามารถของผู้สอบ (θ) และศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด (ML) โดยข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลที่จำลองโดยใช้โปรแกรม R โดยทำการประมาณค่าพารามิเตอร์ และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้โปรแกรม IRTPRO มีรายละเอียดขั้นตอนการดำเนินการวิจัยตามลำดับ ดังนี้

1. ตัวแปรที่ใช้ในการวิจัย
2. เงื่อนไขที่ใช้ในการจำลอง
3. การจำลองข้อมูล
4. การวิเคราะห์ข้อมูล

จากขั้นตอนการดำเนินการวิจัย ที่ได้กล่าวข้างต้น ผู้วิจัยสามารถสรุปรายละเอียดโดยรวมจากการศึกษาค้นคว้า จากเอกสารและงานวิจัยที่เกี่ยวข้องดังรายละเอียดต่อไปนี้

ตัวแปรที่ใช้ในการวิจัย

การศึกษาครั้งนี้ใช้โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) นั่นคือทำการศึกษาพารามิเตอร์ข้อสอบ 2 พารามิเตอร์ ได้แก่ ความยาก และอำนาจจำแนก และพารามิเตอร์ผู้สอบ ได้แก่ ความสามารถผู้สอบ ดังรายละเอียด

1. ข้อมูลที่ใช้ในการศึกษาในครั้งนี้ ใช้ข้อมูลจำลองที่ได้จากโปรแกรม R โดยเป็นข้อมูลที่มีวิธีการให้คะแนนรายข้อแบบสองค่า (Dichotomous scoring) ซึ่งทำการจำลองภายใต้เงื่อนไขจำนวน 32 เงื่อนไข (2x2x2x2x2) ในแต่ละเงื่อนไขจำลองข้อมูลซ้ำ 100 ครั้ง จำนวนการทำซ้ำภายใต้เงื่อนไขที่แปรเปลี่ยนทั้งหมด 3,200 ครั้ง โดยมีรายละเอียดของเงื่อนไข ดังนี้

1.1 สภาวะข้อมูลสูญหาย มี 2 ประเภท ดังนี้

1.1.1 การสูญหายอย่างสุ่ม (Missing random: MAR)

1.1.2 การสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not random: MNAR)

1.2 ขนาดของกลุ่มตัวอย่าง มี 2 ขนาด ประกอบด้วยกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ที่มีสัดส่วนจำนวนผู้สอบในแต่ละกลุ่ม เป็นสัดส่วน 1: 1 ดังนี้

1.2.1 จำนวน 500: 500 คน

1.2.2 จำนวน 1,500: 1,500 คน

1.3 อัตราการสูญหายของข้อมูล มี 2 ระดับดังนี้

1.3.1 อัตราการสูญหาย 10%

1.3.2 อัตราการสูญหาย 30%

1.4 ความยาวของแบบสอบ มี 2 ระดับ โดยที่ระดับความยาวแบบสอบต่างกันมีการกำหนดจำนวนข้อที่ทำหน้าที่ต่างกันของข้อสอบ ดังนี้

1.4.1 แบบสอบจำนวน 20 ข้อ กำหนดข้อที่ 1 เป็นข้อสอบที่ทำหน้าที่ต่างกัน

1.4.2 แบบสอบจำนวน 40 ข้อ กำหนดข้อที่ 1 และ 2 เป็นข้อสอบที่ทำหน้าที่ต่างกัน

1.5 ขนาดอิทธิพลของ DIF มี 2 ขนาด โดยค่าความยากเป็นค่าพารามิเตอร์ที่มีอิทธิพลของ DIF ดังนี้

1.5.1 ขนาดอิทธิพล = 0.5

1.5.2 ขนาดอิทธิพล = 1.0

2. ในการศึกษาในครั้งนี้ ทำการศึกษาภายใต้วิธีการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI)

3. ในการศึกษาในครั้งนี้ ทำการศึกษาภายใต้วิธีการประมาณค่าพารามิเตอร์ ด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด (Maximum likelihood: ML)

4. ในการศึกษาในครั้งนี้ ทำการศึกษาภายใต้วิธีการตรวจสบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT)

5. ในการศึกษาครั้งนี้ พิจารณาการเปรียบเทียบประสิทธิภาพ 2 ส่วน ได้แก่

5.1 การวัดประสิทธิภาพการประมาณค่าพารามิเตอร์ พิจารณาจากดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย และค่าที่ประมาณได้ (diff error) และค่าดัชนีผลต่าง (diff)

5.2 การวัดประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพิจารณาจาก

5.2.1 อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

5.2.2 อำนาจการทดสอบ (Power) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

6. ทำการศึกษาภายใต้การจำลองข้อมูล เพื่อศึกษาการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ซึ่งในการจำลองข้อมูล แต่ละเงื่อนไขในการศึกษาที่กำหนด ผู้วิจัยทำซ้ำ จำนวน 100 ครั้ง (Replication)

7. แบบสอบที่ใช้ในการศึกษาครั้งนี้ เป็นแบบสอบที่มีการตรวจให้คะแนนแบบ 2 ค่า (Dichotomous) คือ ผู้สอบได้ 1 คะแนนเมื่อตอบข้อสอบถูก และได้ 0 คะแนนเมื่อตอบข้อสอบผิด โดยไม่มีการลงโทษเมื่อผู้สอบตอบข้อสอบผิดหรือละเว้นการตอบสนองข้อสอบ นั่นคือ ไม่มีการให้คะแนนติดลบกรณีผู้สอบตอบข้อสอบผิดหรือละเว้นการตอบสนองข้อสอบ

เงื่อนไขที่ใช้ในการจำลอง

การศึกษาครั้งนี้ ทำการศึกษาภายใต้เงื่อนไขที่สำคัญ 5 เงื่อนไข ดังนี้

1. วิธีจัดการข้อมูลสูญหาย

วิธีจัดการข้อมูลสูญหาย มี 1 วิธี ได้แก่ วิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI)

2. สภาวะการสูญหายของข้อมูล

ในการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการจัดการข้อมูลสูญหาย พบว่าสภาวะการสูญหายของข้อมูลในการศึกษาที่ผ่านมา มีการศึกษาวิธีการประมาณค่าทดแทนค่าสูญหายกับข้อมูลที่มีการสูญหายทั้งแบบการสูญหายโดยสมบูรณ์อย่างสุ่ม (Missing completely at random: MCAR) การสูญหายอย่างสุ่ม (Missing at random: MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR) แต่เนื่องจากข้อมูลที่มีรูปแบบการสูญหายโดยสมบูรณ์อย่างสุ่ม (Missing completely at random: MCAR) นั้นไม่ก่อให้เกิดปัญหาเรื่องลำเอียงในการวิเคราะห์ข้อมูล ซึ่งแม้ว่าอาจทำให้เสียอำนาจการทดสอบเนื่องจากการออกแบบการศึกษา แต่การประมาณค่าพารามิเตอร์ที่ได้จะไม่เกิดความลำเอียงเนื่องจากการสูญหายของข้อมูล ดังนั้นในการศึกษาครั้งนี้ จึงจำลองเงื่อนไขการศึกษาครอบคลุมสภาวะการสูญหายของข้อมูล 2 ประเภท คือ การสูญหายอย่างสุ่ม (Missing at random: MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR) โดยรูปแบบการตอบสนองข้อสอบภายใต้สภาวะการสูญหายของข้อมูลแต่ละประเภท มี 3 แบบ คือ การตอบถูก (Correct: CR) ตอบผิด (Incorrect: IN) และละเว้นการตอบสนองข้อสอบ (Omitted: OM)

3. ขนาดของกลุ่มตัวอย่าง

จากการศึกษางานวิจัยที่ผ่านมาพบว่า ดำเนินการศึกษาจากกลุ่มตัวอย่าง 100-2,000 คน (De Ayala, Plake, & Impara, 2001; Bernaards & Sijtsma, 1999, 2000; DeMars, 2002; Zhang & Walker, 2008; Finch, 2008; Glas & Pimentel, 2008; Huisman & Molenaar, 2001; Lord, 1983,1974; Ludlow & O'Leary, 1999) ซึ่งผลการศึกษาพบว่า การใช้กลุ่มตัวอย่างขนาดใหญ่ให้ผลดำเนินการประเมินค่าทางสถิติดีกว่ากลุ่มตัวอย่างขนาดเล็กและเนื่องจาก ในการศึกษาครั้งนี้เป็นการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบตามโมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) กลุ่มตัวอย่างที่เหมาะสมในการศึกษา จึงควรมีจำนวนอย่างน้อย 1,000 คนขึ้นไป (Gao & Chen, 2005; Hanson & Beguin, 2002; Kim, 2006; Yen, 1987) แต่เพื่อให้ผลการศึกษาสะท้อนให้เห็นถึงความแตกต่างทั้งการนำไปใช้จริง ในโมเดลการตอบสนองข้อสอบแบบโลจิสติกชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2 PL) ในการศึกษาครั้งนี้ จึงใช้กลุ่มตัวอย่างขนาดใหญ่ จำนวน 2 ขนาด คือ 1,000 และ 3,000 คน ตามลำดับ

4. อัตราการสูญหายของข้อมูล

จากการศึกษางานวิจัยที่ผ่านมา พบว่า การจำลองข้อมูลเพื่อศึกษาในเชิงทฤษฎีเกี่ยวกับประสิทธิภาพและความถูกต้องแม่นยำของวิธีจัดการข้อมูลสูญหาย มีการศึกษาโดยกำหนดอัตราการสูญหายของข้อมูลอยู่ในช่วงร้อยละ 1-60 (Bernaards & Sijtsma, 1999, 2000; Enders, 2001; Finch, 2008; Gibson & Olejnik, 2003; Glas & Pimenten, 2008; Leite & Beretvas, 2004; Peng & Zhu, 2008; Raaijmakers, 1999; Robitzsch & Rupp, 2009; Velicer & Colby, 2005; Zhang & Walker, 2008) โดยผลการศึกษาพบว่า อัตราการสูญหายของข้อมูลที่ใช้ในการศึกษาไม่ควรเกินร้อยละ 30 เพราะจะทำให้ผลการประมาณค่าพารามิเตอร์เกิดความคลาดเคลื่อนสูง (Bernaards & Sijtsma, 1999, 2000; Leite & Beretvas, 2004; Raaijmakers, 1999) ดังนั้น ในการศึกษาครั้งนี้ ผู้วิจัยจึงใช้อัตราการสูญหายของข้อมูล 2 ระดับ คือ 10% และ 30% ตามลำดับ

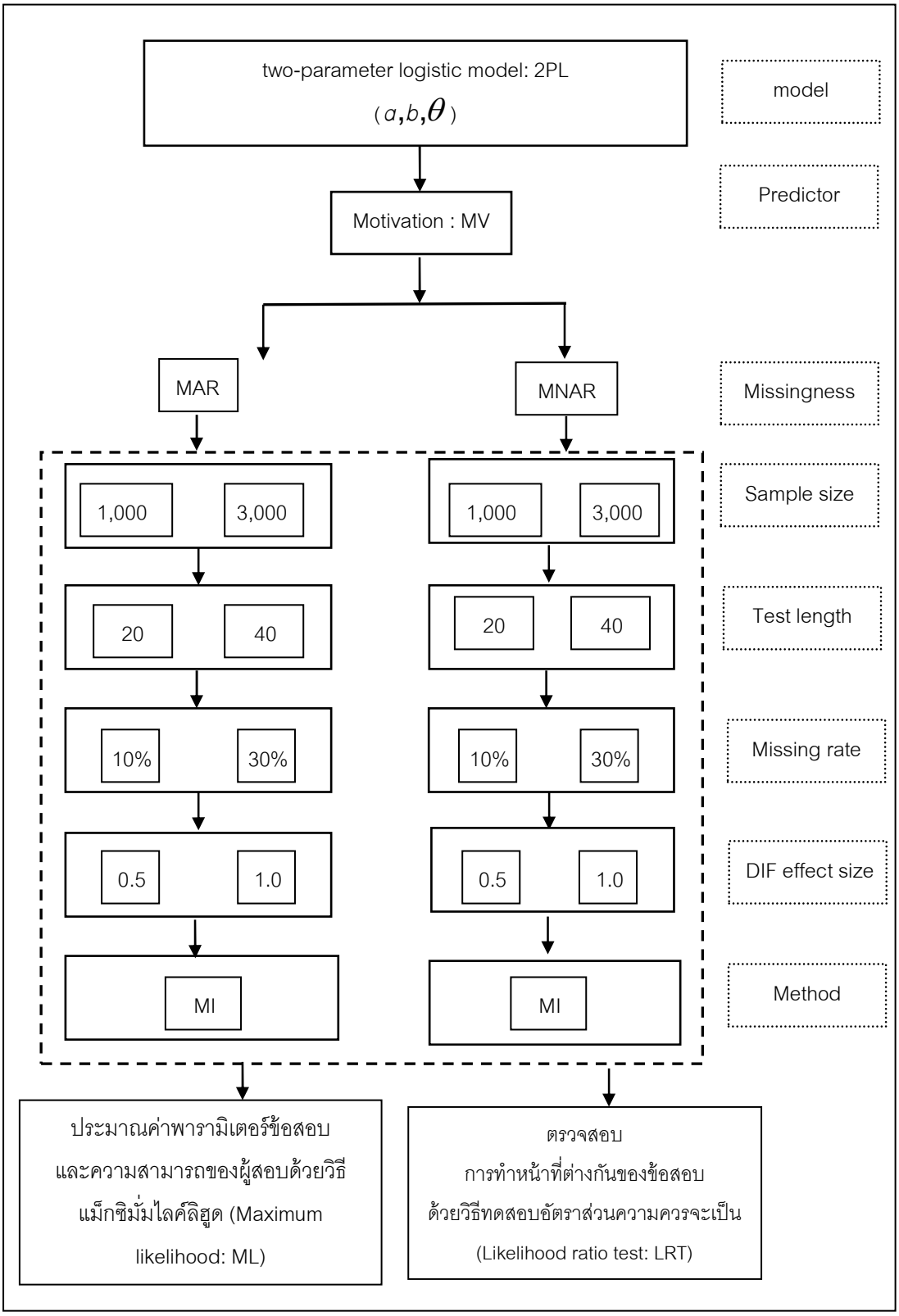
5. ความยาวแบบสอบ

จากการศึกษางานวิจัยที่ผ่านมาเกี่ยวกับความยาวแบบสอบที่ใช้ในการศึกษา พบว่ามีการศึกษาความยาวแบบสอบในช่วง 10-40 ข้อ (Beernaards & Sijtsma, 2000; Glas & Pimenten, 2008; Peng & Zhu, 2008; Robitzsch & Rupp, 2009; Sijtsma & van der Ark, 2003; Zhang & Walker, 2008) โดยผลการศึกษาพบว่า แบบสอบที่มีจำนวนข้อสอบน้อยข้อ

ทำให้ผลการประมาณค่าพารามิเตอร์เกิดความคลาดเคลื่อนสูงกว่าแบบสอบที่มีจำนวนข้อสอบมากข้อ นอกจากนี้แบบสอบที่มีความยาวของข้อสอบตั้งแต่ 20 ข้อขึ้นไป ให้สารสนเทศของแบบสอบและความแม่นยำในการประมาณค่าสูงกว่าแบบสอบที่มีจำนวนข้อน้อยกว่า 20 ข้อ ดังนั้นในการศึกษาครั้งนี้ ผู้วิจัยจึงใช้ความยาวแบบสอบ 2 ระดับ คือ 20 ข้อ และ 40 ข้อ ตามลำดับ เพื่อให้เกิดประโยชน์ต่อการศึกษาเปรียบเทียบในอนาคต

6. **ขนาดอิทธิพลของ** DIF มี 2 ขนาด โดยค่าความยากเป็นค่าพารามิเตอร์ที่มีอิทธิพลของ DIF คือ 0.5 กับ 1.0

จากเงื่อนไขที่กล่าวมาข้างต้นสามารถเสนอด้วยแบบแผนการจำลองข้อมูล ดังนี้



ภาพที่ 3-1 แผนการจำลองข้อมูล

การจำลองข้อมูล

รายละเอียดของการสร้างข้อมูลให้เป็นไปตามเงื่อนไขการจำลองข้อมูล มีรายละเอียดดังนี้ การจำลองข้อมูลเพื่อใช้ประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ครั้งนี้ ใช้การจำลองข้อมูลในการศึกษาความถูกต้องของการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบที่มีการจัดการกับข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) โดยวิเคราะห์ชุดข้อมูลทั้งหมดที่มีการสูญหายในการตอบข้อสอบบางข้อ และในกลุ่มตัวอย่างบางคน โดยใช้โปรแกรม R การศึกษาแต่ละเงื่อนไขมีรายละเอียด ดังนี้

1. โมเดลและค่าพารามิเตอร์ที่แท้จริงที่ใช้ในการจำลองข้อมูล

ผู้วิจัยจำลองข้อมูล ภายใต้ทฤษฎีการตอบสนองข้อสอบ โดยใช้โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) ทำซ้ำข้อมูลจำลองในแต่ละเงื่อนไขจำนวน 100 ครั้ง ด้วยโปรแกรม R โดยข้อสอบที่เป็นกลุ่มข้อสอบเป้าหมาย จะถูกกำหนดให้มีค่าการตอบสูญหาย โดยกำหนดลักษณะค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบของข้อมูลสมบูรณ์ภายใต้โมเดล 2 PL โดยมีรายละเอียด ดังนี้ (Chang, 2012; Montgomery & Skorupski, 2012; Weiss & Minden, 2012)

ตารางที่ 3-1 ค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบที่ใช้ในการจำลองข้อมูล

ค่าพารามิเตอร์	ลักษณะการแจกแจง
ค่าพารามิเตอร์ความสามารถของผู้สอบ θ	N(0.1)
ค่าพารามิเตอร์ของข้อสอบ	
ค่าอำนาจจำแนกของข้อสอบ (a)	N(0.1)
ค่าความยากของข้อสอบ (b)	N(0.1)

จากตารางที่ 3-1 ผู้วิจัยสร้างค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบตามโมเดลการตอบสนองข้อสอบแบบโลจิสติกแบบ 2 พารามิเตอร์ แล้วนำค่าพารามิเตอร์ที่สร้างขึ้น มาคำนวณค่าความน่าจะเป็นที่ผู้สอบคนที่ i จะตอบข้อสอบข้อที่ j ถูกหรือค่า Pj(θ_i) ตามโมเดลการตอบสนองข้อสอบแบบโลจิสติก 2PL ที่ใช้ในการจำลองข้อมูล

ผู้วิจัยสร้างแบบแผนการตอบสนองข้อสอบของผู้สอบในแต่ละสถานการณ์ โดยเริ่มจากการสุ่มค่าความน่าจะเป็นที่มีค่าระหว่าง 0 ถึง 1 สำหรับผู้สอบคนที่ i ในการทำข้อสอบข้อที่ j

เพื่อใช้เป็นค่าเปรียบเทียบกับค่าความน่าจะเป็นที่ผู้สอบคนที่ i จะตอบข้อสอบข้อที่ j ถูก และกำหนดเป็นค่าคำตอบของผู้สอบ (u_{ij}) ซึ่งถ้าความน่าจะเป็นมีค่ามากกว่า $P_j(\theta_i)$ จะกำหนดให้ u_{ij} มีค่าเป็น 0 แต่ถ้าความน่าจะเป็นมีค่าน้อยกว่าหรือเท่ากับ $P_j(\theta_i)$ จะกำหนดให้ u_{ij} มีค่าเป็น 1

2. ขนาดกลุ่มตัวอย่าง

ในการศึกษาครั้งนี้ใช้กลุ่มตัวอย่าง 2 ขนาด คือ 1,000 และ 3,000 คน ตามลำดับ เพื่อให้ผลการศึกษาสะท้อนเงื่อนไขข้อมูลจริงในการศึกษาด้วย โมเดลการตอบสนองข้อสอบแบบโลจิสติก ชนิด 2 พารามิเตอร์ (Two-parameter logistic model: 2PL) และใกล้เคียงสภาพจริงมากที่สุด

3. อัตราการสูญหายของข้อมูล

ในการศึกษาครั้งนี้ใช้อัตราการสูญหายของข้อมูล 2 ระดับคือ 10% และ 30% ตามลำดับ

4. ความยาวแบบสอบ

ในการศึกษาครั้งนี้ใช้ความยาวแบบสอบ 2 ระดับ 20 ข้อ และ 40 ข้อ ตามลำดับ โดยกำหนดข้อที่ 1 และข้อที่ 2 เป็นข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ

4.1 แบบสอบจำนวน 20 ข้อ กำหนดข้อที่ 1 เป็นข้อสอบที่ทำหน้าที่ต่างกัน

4.2 แบบสอบจำนวน 40 ข้อ กำหนดข้อที่ 1 และ 2 เป็นข้อสอบที่ทำหน้าที่

5. สภาวะการสูญหายของข้อมูล

การศึกษาครั้งนี้ จำลองเงื่อนไขสภาวะการสูญหายของข้อมูล 2 ประเภท คือ การสูญหายอย่างสุ่ม (Missing at random: MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (Missing not at random: MNAR)

6. ขนาดอิทธิพลของ DIF

การศึกษาครั้งนี้ ศึกษาภายใต้ขนาดอิทธิพลของ DIF 2 ขนาด โดยค่าความยากเป็นค่าพารามิเตอร์ที่มีอิทธิพลของ DIF ดังนี้

6.1 ขนาดอิทธิพล = 0.5

6.2 ขนาดอิทธิพล = 1.0

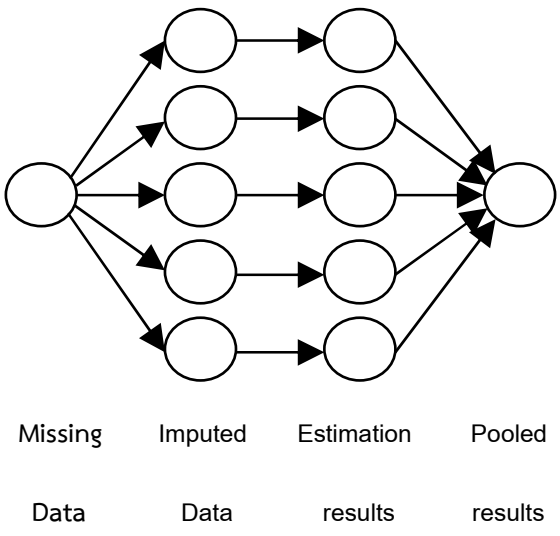
การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลที่ได้จากการจำลองข้อมูลและเกณฑ์ที่ใช้ในการศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด (Maximum likelihood: ML) มีรายละเอียด ดังนี้

1. การประมาณค่าทดแทน

ในขั้นตอนของการประมาณค่าทดแทนแต่ละวิธีนั้น ผู้วิจัยใช้โปรแกรม R เนื่องจากเป็นโปรแกรมคำนวณทางสถิติที่อนุญาตให้ผู้สนใจใช้ได้โดยไม่ละเมิดลิขสิทธิ์ เพื่อใช้ในการคำนวณผลการประมาณค่าทดแทนด้วย วิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) โดยโปรแกรม R มี Package สำเร็จรูปสำหรับประมาณค่าทดแทน ซึ่งพัฒนาขึ้นโดยผู้เชี่ยวชาญที่มีความรู้และเชี่ยวชาญเกี่ยวกับสถิติและการจำลองข้อมูล (Honaker, King, & Blackwell, 2011; Rizopoulos, 2006; van Buuren & Groothuis-Oudshoorn, 2011) โดยในขั้นตอนของการประมาณค่าทดแทนนี้ ผู้วิจัยดำเนินการประมาณค่าทดแทนโดยใช้ Package ‘MICE’ ซึ่งเป็น Package สำเร็จรูปสำหรับการประมาณค่าทดแทนพหุ (Multiple imputation: MI) รายละเอียดและสูตรที่ใช้แสดง ดังนี้

วิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) (Schafer & Graham, 2002; Van Buuren & Groothuis-Oudshoorn, 2011) เป็นเทคนิคที่ประกอบด้วย 3 ขั้นตอนหลัก ดังนี้ การแทนค่าทดแทนแต่ละชุดข้อมูล (Imputation) การประมาณค่า (Estimation) และการรวมผลการประมาณค่าทดแทนข้อมูลสูญหายแต่ละชุดข้อมูลเป็นค่าเดียว (Pooling) ดังภาพที่ 3-2



ภาพที่ 3-2 ขั้นตอนการจำลองของ Multiple imputation (อรวรรณ กีรติสิโรจน์, 2558)

1. การแทนค่าทดแทนแต่ละชุดข้อมูล (Imputation) คือขั้นตอนการแทนค่าของชุดข้อมูลที่มีค่าสูญหาย โดยการสร้างชุดข้อมูลที่ถูกแทนค่าขึ้นมาหลาย ๆ ชุด จำนวน M ชุด จากภาพที่ 3-2 การประมาณค่าทดแทนด้วยวิธี MI จากการศึกษาที่ผ่านมาพบว่า จำนวนครั้งในการประมาณค่าซ้ำสัมพันธ์กับประสิทธิภาพในการประมาณค่า ซึ่งกรณีข้อมูลสูญหายน้อยกว่า 30% จำนวนการประมาณค่าทดแทนซ้ำ 3-5 ครั้ง สามารถให้ผลการประมาณค่าทดแทนที่มีประสิทธิภาพเพียงพอ (Bodner 2008, Graham, 2009) แต่ทั้งนี้ van Buuren and Groothuis-Oudshoorn (2011 อ้างถึงใน กมลทิพย์ ศรีหาเศษ, 2555) เสนอแนะว่าจำนวนการประมาณค่าทดแทนซ้ำขั้นต่ำที่มีประสิทธิภาพในการประมาณค่าทดแทนพหุ คือ 10 ครั้ง ขึ้นไป ดังนั้นเพื่อให้การวิเคราะห์ข้อมูลที่ถูกต้องและเกิดความคลาดเลื่อนน้อยที่สุดในการศึกษาในครั้งนี้ ผู้วิจัยจึงประมาณค่าทดแทนซ้ำด้วยวิธี MI จำนวน 10 ครั้ง (M=10)

2. การประมาณค่า (Estimation) คือขั้นตอนการวิเคราะห์ข้อมูลตามปกติด้วยโมเดลที่เหมาะสมกับข้อมูล

3. การรวมผลการประมาณค่าทดแทนข้อมูลสูญหายแต่ละชุดข้อมูลเป็นค่าเดียว (Pooling) คือ ขั้นตอนการสรุปผลการวิเคราะห์ โดยการรวมผลการวิเคราะห์ของแต่ละชุดข้อมูลเพื่อออกมาเป็นผลลัพธ์สุดท้าย (Final result/ Pooled results) ขั้นตอนการประมาณค่าและการรวม อาจเรียกรวมกันว่าขั้นตอนการวิเคราะห์ (Analysis)

การรวมผลการประมาณค่าทดแทนจากการประมาณค่าทดแทนซ้ำ M ครั้ง เป็นค่าเดียว มีขั้นตอนการคำนวณเป็นลำดับ ดังนี้

$$\bar{Q} = \frac{\sum \hat{Q}_m}{M}$$

โดยที่

$\bar{Q}$ แทน ค่าเฉลี่ยของการประมาณค่าทดแทนทั้ง M ครั้ง

ความแปรปรวนของการประมาณค่านี้ประกอบด้วย 2 ส่วนคือ

1. ความแปรปรวนระหว่างการประมาณค่าทดแทน ดังสมการ

$$B = \frac{\sum_{m}(\hat{Q}_m - \bar{Q})^2}{M-1}$$

โดยที่

B แทน ความแปรปรวนระหว่างการประมาณค่าทดแทนทั้ง M ครั้ง

2. ความแปรปรวนภายในการประมาณค่าทดแทน ดังสมการ

$$\bar{u} = \frac{\sum_{m} u_m}{M}$$

โดยที่

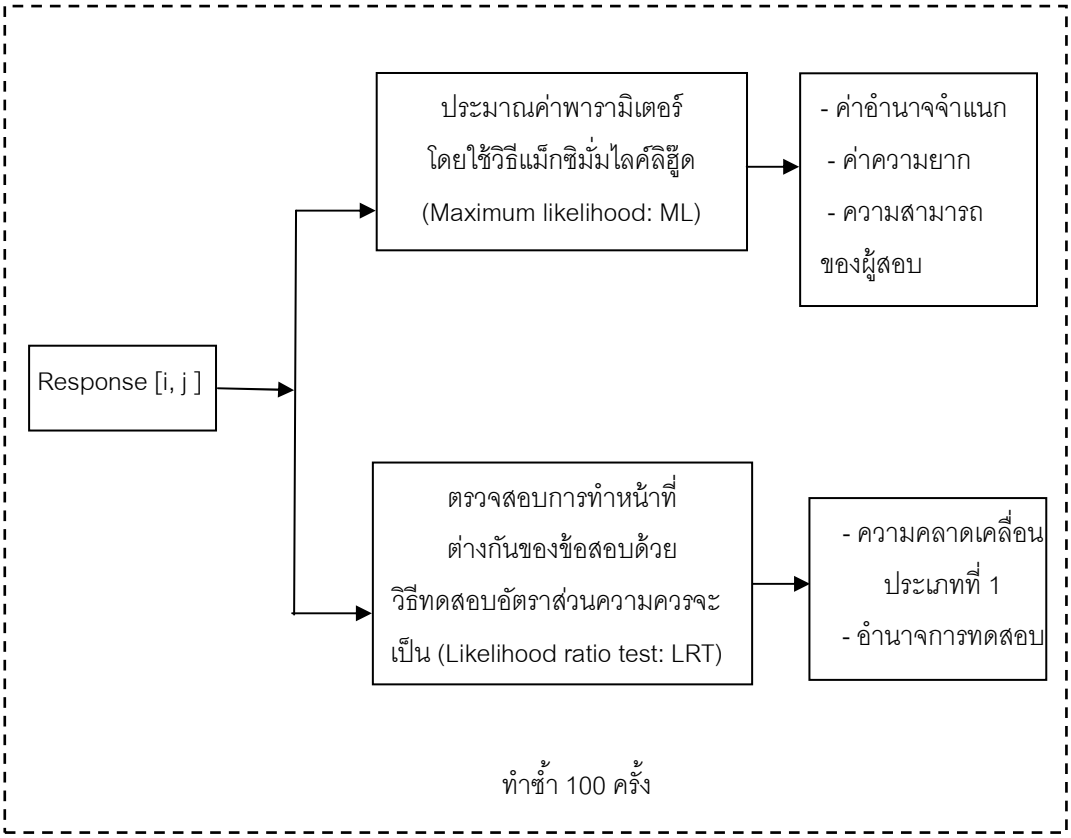
$\bar{u}$ แทน ค่าเฉลี่ยของความแปรปรวนภายในการประมาณค่าทดแทนทั้ง M ครั้ง

ความแปรปรวนภายในการประมาณค่าทดแทน ($\bar{U}$) เป็นค่าเฉลี่ยของการประมาณค่าความแปรปรวนข้ามการประมาณค่าทดแทน M ความแปรปรวนสำหรับการประมาณค่าทดแทนพหุ จะคำนวณภายหลังด้วยสมการดังนี้

$$T = \bar{U} + \left(1 + \frac{1}{M}\right)B$$

2. การประมาณค่าพารามิเตอร์และการตรวจสอบการทำหน้าที่ของข้อสอบ

ในการศึกษาในครั้งนี้ผู้วิจัยทำการประมาณค่าพารามิเตอร์ ด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด (Maximum likelihood: ML) และตรวจสอบการทำหน้าที่ของข้อสอบด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT) โดยการใช้โปรแกรมสำเร็จรูป IRTPRO (Cai, Toit, & Thissen, 2011 อ้างถึงใน อาวีพร ปานทอง, 2558) ซึ่งโปรแกรม IRTPRO จะทำการประมาณค่าพารามิเตอร์และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งทั้ง 2 ส่วนสามารถทำได้พร้อมกัน ดังภาพ



ภาพที่ 3-3 ขั้นตอนการประมาณค่าและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

3. การพิจารณาการเปรียบเทียบประสิทธิภาพ

การพิจารณาการเปรียบเทียบประสิทธิภาพ พิจารณา 2 ส่วน ดังต่อไปนี้

3.1 การวัดประสิทธิภาพการประมาณค่าพารามิเตอร์ พิจารณาจากดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย (diff error) และดัชนีผลต่าง (diff)

3.1.1 วิเคราะห์ข้อมูลโดยการวิเคราะห์ค่าสถิติบรรยายของการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ ประกอบด้วย ค่าเฉลี่ย ($\bar{x}$) ส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ของข้อสอบและความสามารถ ของผู้สอบที่ประมาณค่าได้จากวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) โดยภาพรวมและรายข้อ

3.1.2 จากข้อ 3.1.1 วัดความความถูกต้องแม่นยำของการประมาณค่าใช้เกณฑ์ดัชนีผลต่าง (diff) และดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย (diff error) โดยใช้สูตร ดังนี้

$$\text{diff} = \frac{\sum(A-\widehat{A})}{n}$$

เมื่อ

diff แทน ดัชนีผลต่าง

A แทน ค่าเฉลี่ยของค่าพารามิเตอร์ข้อสอบ และค่าความสามารถของผู้สอบที่ได้จากการจำลองข้อมูล

$\widehat{A}$ แทน ค่าประมาณที่ได้จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI)

n แทน จำนวนครั้งที่ทำซ้ำในการจำลองข้อมูล

ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย เขียนสมการได้ดังสมการ

$$\text{diff error} = \sqrt{\frac{\sum\left(A-\widehat{A}\right)^2}{n}}$$

เมื่อ

diff error แทน ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย

A แทน ค่าเฉลี่ยของค่าพารามิเตอร์ข้อสอบ และค่าความสามารถของผู้สอบที่ได้จากการจำลองข้อมูล

$\widehat{A}$ แทน ค่าประมาณที่ได้จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI)

n แทน จำนวนครั้งที่ทำซ้ำในการจำลองข้อมูล

3.2 การวัดประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบพิจารณาจากหลังจากทำการวิเคราะห์ข้อมูล นำผลการวิเคราะห์ที่ได้มาคำนวณหาความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจทดสอบของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดังนี้

3.2.1 อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมายถึง จำนวนครั้งที่ปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักเป็นจริง หารด้วยจำนวนครั้งของการทดสอบ ในทางปฏิบัติ คือ จำนวนข้อสอบที่ระบุว่าทำหน้าที่ต่างกัน ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน หารด้วยจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมด เขียนเป็นสมการได้เป็น (อริสฟา เตหลิ่ม, 2559)

$$\text{อัตราความคลาดเคลื่อนประเภทที่ 1} = \frac{\sum \text{จำนวนข้อที่ระบุว่าทำหน้าที่ต่างกัน}}{\text{จำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมด}} \times 100$$

จากนั้นทำการทดสอบสมมติฐานเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ซึ่งเกณฑ์ที่ใช้ในการพิจารณาหากมีค่าความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าหรือเท่ากับ 0.05 ถือว่าสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี (Atar & Kamata, 2011, p. 40 อ้างถึงใน อาวีพร ปานทอง, 2558) นั่นคืออวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่ระบุการทำหน้าที่ต่างกันของข้อสอบในข้อที่ไม่มีการทำหน้าที่ต่างกันของข้อสอบได้จริง มีขั้นตอนการทดสอบสมมติฐาน ดังนี้

ขั้นที่ 1 ตั้งสมมติฐานการทดสอบ

$$H_0: P \leq 0.05$$
$$H_1: P > 0.05$$

ขั้นที่ 2 กำหนดระดับนัยสำคัญ (α) = 0.05

ขั้นที่ 3 คำนวณสถิติที่ใช้ในการทดสอบ

$$Z = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

$\hat{P}$ แทน สัดส่วนการเกิดความคลาดเคลื่อนประเภทที่ 1 ของกลุ่มตัวอย่าง

P แทน สัดส่วนการเกิดความคลาดเคลื่อนประเภทที่ 1 ของประชากร

n แทน จำนวนของการทำซ้ำในการจำลองข้อมูล

ขั้นที่ 4 กำหนดขอบเขตวิกฤต

เนื่องจากเป็นการทดสอบสมมติฐานแบบทางเดียว จึงนำ Z ที่คำนวณได้จากสูตรเทียบกับ $Z_{\alpha} = 1.645$ ถ้า Z ที่คำนวณได้น้อยกว่า $Z_{\alpha} = 1.645$ จะยอมรับสมมติฐาน H_0 แสดงว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบนั้น สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.2.2 อำนาจการทดสอบ (Power of test) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

อำนาจของตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Power of test) หมายถึง จำนวนข้อสอบที่ระบุว่าทำหน้าที่ต่างกันได้ถูก ซึ่งคำนวณได้จาก (อริสฟา เตหลิ่ม, 2559)

$$\text{อำนาจการทดสอบ} = \frac{\sum \text{จำนวนข้อสอบที่ระบุว่าทำหน้าที่ต่างกัน}}{\text{จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ x 100}}$$

จากนั้นทำการทดสอบสมมติฐานเปรียบเทียบอัตราอำนาจการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งเกณฑ์ที่ใช้ในการพิจารณาอำนาจการทดสอบ จะพิจารณาอำนาจการตรวจสอบ เมื่อสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ก่อน และอำนาจการทดสอบต้องมีค่าเฉลี่ยตั้งแต่ 0.80 ขึ้นไป จึงถือว่ามีอำนาจการทดสอบเพียงพอ (Sufficient power) หากต่ำกว่า 0.80 ถือว่าวิธีการดังกล่าวตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ไม่ดี (Atar & Kamata, 2011, p. 40 อ้างถึงใน อาวีพร ปานทอง, 2558) มีขั้นตอนการทดสอบสมมติฐาน ดังนี้

ขั้นที่ 1 ตั้งสมมติฐานการทดสอบ

$H_0 : P \geq 0.80$

$H_1 : P < 0.80$

ขั้นที่ 2 กำหนดระดับนัยสำคัญ (α) = .05

ขั้นที่ 3 คำนวณสถิติที่ใช้ในการทดสอบ

$$Z = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

P แทน สัดส่วนอำนาจการทดสอบของกลุ่มตัวอย่าง

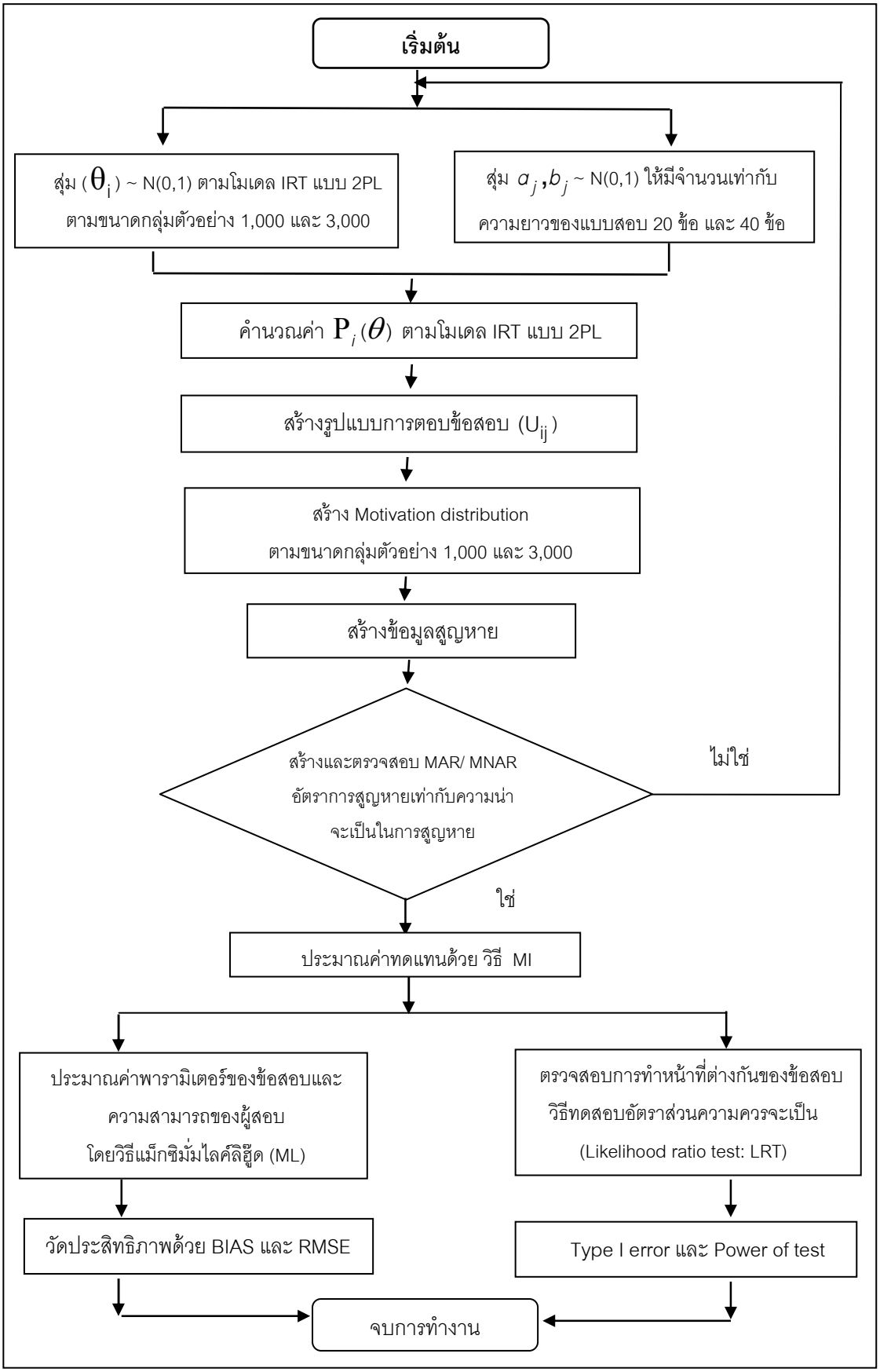
P แทน สัดส่วนอำนาจการทดสอบของประชากร

n แทน จำนวนของการทำซ้ำในการจำลองข้อมูล

ขั้นที่ 4 กำหนดขอบเขตวิกฤต

เนื่องจากเป็นการทดสอบสมมติฐานแบบทางเดียว จึงนำ Z ที่คำนวณได้จากสูตรเทียบกับ $Z_{\alpha} = Z_{0.05}$ ถ้า Z ที่คำนวณได้มากกว่า $Z_{\alpha} = -1.645$ จะยอมรับสมมติฐาน H_o แสดงว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบนั้นมีอำนาจการทดสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

การดำเนินงานทั้งหมดสามารถสรุปขั้นตอนการจำลองข้อมูลและการวิเคราะห์ข้อมูลดังภาพที่ 3-4



ภาพที่ 3-4 แสดงขั้นตอนการจำลองข้อมูลและการวิเคราะห์

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ มีวัตถุประสงค์ 2 ประการ ประการแรก เพื่อศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก (a) และความยาก (b)) กับพารามิเตอร์ความสามารถของผู้สอบ (θ) ประการที่สองเพื่อศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ข้อมูลที่ใช้ในการศึกษาในครั้งนี้ เป็นข้อมูลที่จำลองขึ้นภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF จำนวน 32 เงื่อนไข (2x2x2x2x2) ในแต่ละเงื่อนไขจำลองข้อมูลวนซ้ำ 100 ครั้ง ผลการวิเคราะห์ข้อมูลแสดง ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก (a) และความยาก (b)) ด้วยวิธีจัดการข้อมูลสูญหายโดยใช้วิธีประมาณค่าพหุ (Multiple imputation: MI)

ตอนที่ 2 ผลการวิเคราะห์ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ด้วยวิธีจัดการข้อมูลสูญหายโดยใช้วิธีประมาณค่าพหุ (Multiple imputation: MI)

ตอนที่ 3 ผลการเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ข้อสอบ (ค่าอำนาจจำแนก และค่าความยาก)

ตอนที่ 4 ผลการศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ)

ตอนที่ 5 ผลการตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบ

เพื่อให้การนำเสนอผลการวิเคราะห์ข้อมูลมีความเข้าใจตรงกัน ผู้วิจัยได้กำหนดสัญลักษณ์ และความหมายแทนตัวแปรต่าง ๆ ดังนี้

a	แทน	ค่าอำนาจจำแนกของข้อสอบ
a_Com	แทน	ค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์
a_MI	แทน	ค่าอำนาจจำแนกของข้อสอบที่ประมาณได้จากวิธีการประมาณค่าทดแทนพหุ
b	แทน	ค่าความยากของข้อสอบ
b_R _Com	แทน	ค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ในกลุ่มอ้างอิง

b_R _MI	แทน	ค่าความยากของของข้อสอบกลุ่มอ้างอิงที่ประมาณได้จากวิธีการประมาณค่าทดแทนพหุ
b_F _Com	แทน	ค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ในกลุ่มเปรียบเทียบ
b_F _MI	แทน	ค่าความยากของของข้อสอบกลุ่มเปรียบเทียบที่ประมาณได้จากวิธีการประมาณค่าทดแทนพหุ
θ	แทน	ค่าความสามารถของผู้สอบ
θ_Com	แทน	ค่าความสามารถของผู้สอบของข้อมูลที่สมบูรณ์
θ_MI	แทน	ค่าความสามารถของผู้สอบที่ประมาณได้จากวิธีการประมาณค่าทดแทนพหุ
diff	แทน	ดัชนีผลต่าง
diff error	แทน	ดัชนีค่ารากที่สองของผลต่างระหว่างค่ากำหนดจากการจำลองกับค่าประมาณยกกำลังสองเฉลี่ย

ตอนที่ 1 ผลการวิเคราะห์ค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนก และความยาก) ด้วยวิธีจัดการข้อมูลสูญหายโดยใช้วิธีประมาณค่าพหุ (Multiple imputation: MI)

การนำเสนอผลการวิเคราะห์ค่าพารามิเตอร์ข้อสอบในเบื้องต้นนี้ เป็นการนำเสนอผลการวิเคราะห์ด้วยสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF เพื่อบรรยายลักษณะค่าพารามิเตอร์ข้อสอบที่ประมาณค่าได้ด้วยการจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) และประมาณค่าพารามิเตอร์ ด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด (Maximum Likelihood: ML) รายละเอียดแสดง ดังนี้

ตารางที่ 4-1 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) โดยภาพรวม จำแนกตามความยาวของแบบสอบ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหาย ของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกรลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					a_Com	a_MI	a_Com	a_MI
20	10	0.5	1,000	M	0.8450	1.1859	0.8533	1.1539
				SD	0.0461	0.1461	0.0443	0.1463
			3,000	M	0.8476	1.1593	0.8507	1.1389
				SD	0.0475	0.1529	0.0655	0.1488
		1.0	1,000	M	0.8560	1.2032	0.8407	1.1375
				SD	0.0599	0.1791	0.0438	0.1087
			3,000	M	0.8394	1.1563	0.8528	1.1821
				SD	0.0429	0.1423	0.0442	0.1232
	30	0.5	1,000	M	0.8479	1.1183	0.8505	1.1040
				SD	0.0360	0.2650	0.0410	0.2452
			3,000	M	0.8380	1.0654	0.8436	1.0517
				SD	0.0391	0.2252	0.0366	0.2174
		1.0	1,000	M	0.8134	0.8425	0.8397	1.0613
				SD	0.0336	0.7967	0.0393	0.2492
			3,000	M	0.8504	1.0633	1.0847	0.8436
				SD	0.0456	0.2233	0.2659	0.0411

ตารางที่ 4-1 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					a_Com	a_MI	a_Com	a_MI
40	10	0.5	1,000	M	0.8461	1.1172	0.8336	1.1767
				SD	0.0519	0.1186	0.0433	0.1136
			3,000	M	0.8380	1.1577	0.8436	1.1647
				SD	0.0495	0.1449	0.0439	0.1133
		1.0	1,000	M	0.8482	1.1349	0.8611	1.1392
				SD	0.0497	0.1038	0.0542	0.1351
			3,000	M	0.8471	1.1304	0.8530	1.1798
				SD	0.0513	0.1129	0.0557	0.1252
	30	0.5	1,000	M	0.8535	1.1188	0.8371	1.1316
				SD	0.0532	0.1830	0.0582	0.1870
			3,000	M	0.7842	1.1282	0.8397	1.1037
				SD	0.0440	0.1943	0.0393	0.1790
		1.0	1,000	M	0.8439	1.1439	0.8417	0.9292
				SD	0.0427	0.2014	0.0456	1.3039
			3,000	M	0.8502	1.1065	0.8577	1.1422
				SD	0.0423	0.1784	0.0460	0.2095

จากตารางที่ 4-1 ผลการวิเคราะห์ค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการประมาณค่าทดแทน (a_MI) โดยภาพรวมพบว่า มีค่าได้มีค่าใกล้เคียงกัน แต่ค่าที่ได้มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com)

ตารางที่ 4-2 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
20	10	0.5	1,000	1	M	0.8528	0.9828	0.8652	1.0034
					SD	0.4793	0.6832	0.4587	0.6835
				2	M	0.8286	0.9572	0.8218	0.9102
					SD	0.4601	0.6967	0.4583	0.6706
				3	M	0.8988	1.0191	0.9112	0.9714
					SD	0.4754	0.7529	0.4380	0.6866
				4	M	0.7264	1.0053	0.8067	0.8744
					SD	0.4127	0.6913	0.4571	0.5834
			3,000	1	M	0.8421	0.9765	0.8910	1.0491
					SD	0.4744	0.7161	0.4789	0.6865
				2	M	0.7953	0.8130	0.8526	0.9409
					SD	0.4831	0.6067	0.4503	0.6472
				3	M	0.8962	0.9730	0.7769	0.7887
					SD	0.4654	0.6042	0.4571	0.6113
				4	M	0.8245	0.8851	0.8164	0.9216
					SD	0.4974	0.6292	0.4613	0.6015

ตารางที่ 4-2 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
20	10	1.0	1,000	1	M	0.9108	1.0289	0.9035	0.974
					SD	0.4773	0.6697	0.4505	0.6696
				2	M	0.8774	0.9880	0.8004	0.9205
					SD	0.4518	0.6752	0.4525	0.5760
				3	M	0.8399	0.9257	0.8472	0.9036
					SD	0.4696	0.6878	0.4703	0.6237
				4	M	0.7587	0.9549	0.8616	1.0006
					SD	0.4397	0.6603	0.4486	0.7395
			3,000	1	M	0.8048	0.7818	0.7924	0.9101
					SD	0.4462	0.6659	0.4611	0.7661
				2	M	0.8320	0.9868	0.8785	0.1069
					SD	0.4736	0.7026	0.4577	0.6914
				3	M	0.7926	0.9724	0.8453	0.9660
					SD	0.4457	0.7078	0.4207	0.6570
				4	M	0.8113	1.0133	0.8251	1.0149
					SD	0.4379	0.6670	0.3164	0.6499
20	30	0.5	1,000	1	M	0.8306	0.5308	0.9000	0.6876
					SD	0.4501	0.4644	0.4989	0.4548
				2	M	0.8951	0.6947	0.8923	0.6652
					SD	0.4677	0.4350	0.4627	0.4058
				3	M	0.8818	0.6719	0.7874	0.6250
					SD	0.4693	0.4733	0.4516	0.4572
				4	M	0.8161	0.5910	0.8774	0.6696
					SD	0.4536	0.4093	0.5063	0.4129

ตารางที่ 4-2 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
20	30	0.5	3,000	1	M	0.8103	0.6301	0.8227	0.6702
					SD	0.4155	0.3879	0.4667	0.4041
				2	M	0.9351	0.7073	0.8897	06798
					SD	0.4882	0.4190	0.4687	0.4079
				3	M	0.8285	0.6372	0.8204	0.6214
					SD	0.4237	0.4166	0.4195	0.4132
				4	M	0.8374	0.6514	0.8671	0.6730
					SD	0.4615	0.4143	0.4533	0.4246
20	30	1.0	1,000	1	M	0.8482	0.6581	0.8696	0.5850
					SD	0.4737	0.4891	0.4971	0.4418
				2	M	0.8364	0.6138	0.8598	0.5578
					SD	0.4606	0.5065	0.4850	0.4474
				3	M	0.8251	0.6536	0.8686	0.6595
					SD	0.4643	0.6160	0.4444	0.3817
				4	M	0.8134	0.5950	0.7963	0.6078
					SD	0.4206	0.5001	0.4119	0.3983
			3,000	1	M	0.9271	0.6898	0.8251	0.5565
					SD	0.4932	0.4731	0.1988	0.1206
				2	M	0.7945	0.6422	0.7777	0.6351
					SD	0.4236	0.3845	0.1791	0.1606
				3	M	0.8501	0.6062	0.8220	0.5601
					SD	0.4499	0.4058	0.2109	0.1553
				4	M	0.8510	0.6279	0.8554	0.5954
					SD	0.5064	0.4033	0.1801	0.1912

จากตารางที่ 4-2 ผลการวิเคราะห์ค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการประมาณค่าทดแทน (a_MI) รายข้อ พบว่าที่อัตราการสูญหาย 10% ค่าที่ได้มีค่าใกล้เคียงกันแต่ส่วนใหญ่มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com) และเมื่ออัตราการสูญหายเพิ่มขึ้นเป็น 30% ค่าที่ได้มีค่าใกล้เคียงกัน แต่ส่วนใหญ่มีค่าต่ำกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com)

ตารางที่ 4-3 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
40	10	0.5	1,000	1	M	0.7587	0.8692	0.8349	0.9662
					SD	0.3975	0.6084	0.4388	0.5372
				2	M	0.9243	0.9872	0.8171	0.9156
					SD	0.4834	0.6570	0.4482	0.6724
				3	M	0.9216	0.9469	0.8854	1.0115
					SD	0.4999	0.6975	0.4227	0.5616
				4	M	0.8430	0.9408	0.9113	1.0421
					SD	0.3881	0.6064	0.4707	0.5794
			3,000	1	M	0.7429	0.9362	0.7930	0.9596
					SD	0.4632	0.6051	0.4234	0.6461
				2	M	0.8227	0.9961	0.8035	0.9318
					SD	0.4517	0.6639	0.4576	1.1864
				3	M	0.7971	0.9169	0.8378	1.0332
					SD	0.4699	0.5776	0.4915	1.7330
				4	M	0.8474	0.9030	0.7609	0.8000
					SD	0.4518	0.5547	0.4374	1.7330
40	10	1.0	1,000	1	M	0.9059	0.8954	0.8365	0.9273
					SD	0.4442	0.6150	0.4379	0.6618
				2	M	0.8175	0.9847	0.8288	0.8826
					SD	0.4895	0.6742	0.4697	0.5920
				3	M	0.7849	0.9772	0.8285	0.8191
					SD	0.4125	0.5798	0.4580	0.5987
				4	M	0.8103	0.8600	0.8142	0.8365
					SD	0.4367	0.5612	0.4346	0.5368

ตารางที่ 4-3 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
40	10	1.0	3,000	1	M	0.8672	0.9032	0.8203	0.8826
					SD	0.5022	0.5131	0.4549	0.6835
				2	M	0.8263	0.8255	0.8573	0.9629
					SD	0.4357	0.4827	0.4243	1.2012
				3	M	0.8818	0.9798	0.8502	1.0481
					SD	0.4428	0.6289	0.4809	1.7122
				4	M	0.8640	0.8990	0.8674	0.9647
					SD	0.4409	0.5525	0.4583	0.1797
40	30	0.5	1,000	1	M	0.8017	0.6042	0.8506	0.6666
					SD	0.4454	0.3667	0.1970	0.1194
				2	M	0.8610	0.6703	0.8210	0.6103
					SD	0.4289	0.3752	0.2269	0.1372
				3	M	0.9393	0.6382	0.7561	0.5635
					SD	0.4987	0.3873	0.1757	0.1130
				4	M	0.8384	0.6333	0.8401	0.6913
					SD	0.4679	0.3842	0.1970	0.1572
			3,000	1	M	0.7448	0.9037	0.8294	0.6038
					SD	0.4820	0.3712	0.2136	0.1339
				2	M	0.7388	0.8100	0.8944	0.6687
					SD	0.4666	0.3484	0.1937	0.1522
				3	M	0.7285	0.9558	0.8104	0.5690
					SD	0.4832	0.3669	0.2011	0.1457
				4	M	0.7516	0.9047	0.8838	0.6707
					SD	0.4660	0.3942	0.1927	0.1271

ตารางที่ 4-3 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
40	30	1.0	1,000	1	M	0.8950	0.6308	0.8665	0.6590
					SD	0.5219	0.4339	0.2152	0.1715
				2	M	0.8288	0.6510	0.8535	0.6660
					SD	0.4690	0.4413	0.1973	0.1640
				3	M	0.8416	0.6110	0.8422	0.6610
					SD	0.4324	0.4253	0.2105	0.1872
				4	M	0.9106	0.5896	0.9218	0.6977
					SD	0.4584	0.3639	0.2260	0.2052
			3,000	1	M	0.9164	0.6047	0.8368	0.5676
					SD	0.5093	0.3639	0.4704	0.3454
				2	M	0.8044	0.6347	0.7949	0.5699
					SD	0.4305	0.4719	0.4558	0.3708
				3	M	0.8129	0.6402	0.7296	0.5700
					SD	0.4412	0.3630	0.4001	0.2966
				4	M	0.8518	0.6115	0.8070	0.6097
					SD	0.4389	0.3682	0.4524	0.3590

จากตารางที่ 4-3 ผลการวิเคราะห์ค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการประมาณค่าทดแทน (a_MI) รายข้อพบว่าที่อัตราการสูญหาย 10% ค่าที่ได้มีค่าใกล้เคียงกันแต่ส่วนใหญ่มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com) และเมื่ออัตราการสูญหายเพิ่มขึ้นเป็น 30% ค่าที่ได้มีค่าใกล้เคียงกัน แต่ส่วนใหญ่มีค่าต่ำกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com)

ตารางที่ 4-4 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) โดยภาพรวม จำแนกตามความยาวของแบบสอบ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหาย ของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					a_Com	a_MI	a_Com	a_MI
20	10	0.5	1,000	M	0.8450	1.1799	0.8533	1.1616
				SD	0.0461	0.1256	0.0443	0.1332
			3,000	M	0.8476	1.1074	0.8507	1.1093
				SD	0.0475	0.1156	0.0655	0.1569
		1.0	1,000	M	0.8560	1.1791	0.8407	1.1351
				SD	0.0599	0.1554	0.0438	0.1204
			3,000	M	0.8394	1.1008	0.8528	1.1806
				SD	0.0429	0.1368	0.0442	0.1254
	30	0.5	1,000	M	0.8479	1.1184	0.8505	1.1085
				SD	0.0360	0.2696	0.0410	0.2531
			3,000	M	0.8380	1.0624	0.8436	1.0607
				SD	0.0391	0.2202	0.0366	0.2197
		1.0	1,000	M	0.8134	1.0206	0.8397	1.0859
				SD	0.0336	0.2456	0.0393	0.2471
			3,000	M	0.8504	1.0819	0.8577	1.0701
				SD	0.0456	0.2318	0.0460	0.2483

ตารางที่ 4-4 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					a_Com	a_MI	a_Com	a_MI
40	10	0.5	1,000	M	0.8461	1.1145	0.8336	1.1637
				SD	0.0519	0.1099	0.0433	0.1109
			3,000	M	0.8380	1.1158	0.8436	1.1539
				SD	0.0495	0.1167	0.0439	0.1124
		1.0	1,000	M	0.8482	1.0957	0.8611	1.1398
				SD	0.0497	0.1164	0.0542	0.1380
			3,000	M	0.8471	1.1074	0.8530	1.1829
				SD	0.0513	0.1156	0.0557	0.1256
	30	0.5	1,000	M	0.8535	1.1100	0.8371	1.1196
				SD	0.0532	0.1788	0.0582	0.1801
			3,000	M	0.7842	1.1276	0.8397	1.1373
				SD	0.0440	0.1949	0.0393	0.1905
		1.0	1,000	M	0.8439	1.1567	0.8417	0.9292
				SD	0.0427	0.1983	0.0456	0.1304
			3,000	M	0.8502	1.0924	0.8577	1.1078
				SD	0.0423	0.1782	0.0460	0.2028

จากตารางที่ 4-4 ผลการวิเคราะห์ค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการประมาณค่าทดแทน (a_MI) โดยภาพรวมพบว่า มีค่าได้มีค่าใกล้เคียงกัน แต่ค่าที่ได้มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com)

ตารางที่ 4-5 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหาย ของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
20	10	0.5	1,000	1	M	0.8528	1.0331	0.9035	0.9047
					SD	0.4793	0.7275	0.4505	0.6617
				2	M	0.8286	0.9243	0.8004	0.9267
					SD	0.4601	0.6837	0.4525	0.6808
				3	M	0.8988	1.0526	0.8472	0.8485
					SD	0.4754	0.7422	0.4703	0.6290
				4	M	0.7264	1.0020	0.8616	0.9689
					SD	0.4127	0.6604	0.4486	0.7359
			3,000	1	M	0.8421	0.9608	0.8910	0.9905
					SD	0.4744	0.6956	0.4894	0.6477
				2	M	0.7953	0.8292	0.8526	0.9430
					SD	0.4831	0.5834	0.4503	0.6059
				3	M	0.8962	1.0032	0.7769	0.7861
					SD	0.4654	0.6252	0.4571	0.5806
				4	M	0.8245	0.8793	0.8164	0.8641
					SD	0.4974	0.6445	0.4613	0.5397

ตารางที่ 4-5 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
20	10	1.0	1,000	1	M	0.9108	1.0486	0.9035	1.0247
					SD	0.4773	0.6385	0.4505	06617
				2	M	0.8774	0.9627	0.8004	0.9267
					SD	0.4518	0.6679	0.4525	0.6808
				3	M	0.8399	1.0005	0.8472	0.8485
					SD	0.4696	0.7383	0.4703	0.6290
				4	M	0.7587	0.9355	0.8616	0.9689
					SD	0.4397	0.6329	0.4486	0.7359
			3,000	1	M	0.8048	0.7680	0.7924	0.9025
					SD	0.4462	0.6208	0.4611	0.7366
				2	M	0.8320	0.9476	0.8785	1.0556
					SD	0.4736	0.7016	0.4577	0.6775
				3	M	0.7926	0.9192	0.8453	0.9866
					SD	0.4457	0.6672	0.4207	0.6575
				4	M	0.8113	0.9498	0.8251	1.0161
					SD	0.4379	0.5934	0.4316	0.6338
20	30	0.5	1,000	1	M	0.8306	0.5563	0.9004	0.6650
					SD	0.4501	0.4481	0.4989	0.4548
				2	M	0.8951	0.6474	0.8923	0.6492
					SD	0.4677	0.4295	0.4627	0.4058
				3	M	0.8818	0.6611	0.7874	0.6330
					SD	0.4693	0.4921	0.4516	0.4572
				4	M	0.8161	0.6049	0.8761	0.6904
					SD	0.4536	0.4196	0.5063	0.4129

ตารางที่ 4-5 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
20	30	0.5	3,000	1	M	0.8103	0.6357	0.8227	0.6486
					SD	0.4155	0.3781	0.4667	0.4019
				2	M	0.9351	0.7221	0.8897	0.6799
					SD	0.4882	0.4096	0.4687	0.3940
				3	M	0.8285	0.6354	0.8204	0.6226
					SD	0.4237	0.4132	0.4196	0.3873
				4	M	0.8374	0.6611	0.8671	0.6790
					SD	0.4615	0.4018	0.4195	0.3998
20	30	1.0	1,000	1	M	0.8482	0.6804	0.8696	0.6048
					SD	0.4737	0.5555	0.4971	0.4102
				2	M	0.8364	0.5552	0.8598	0.6020
					SD	0.4606	0.6041	0.4850	0.4127
				3	M	0.8251	0.6137	0.8686	0.6509
					SD	0.4643	0.4179	0.4444	0.4581
				4	M	0.8134	0.5702	0.7931	0.6322
					SD	0.4206	0.4470	0.4119	0.3852
			3,000	1	M	0.9271	0.6653	0.8251	0.6483
					SD	0.4932	0.3958	0.1987	0.6735
				2	M	0.7945	0.6555	0.7777	0.6481
					SD	0.4236	0.3850	0.1791	0.1761
				3	M	0.8501	0.6040	0.8220	0.5782
					SD	0.4499	0.3898	0.2108	0.1516
				4	M	0.8510	0.6537	0.8554	0.5628
					SD	0.5064	0.4024	0.1801	0.2310

จากตารางที่ 4-5 ผลการวิเคราะห์ค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการประมาณค่าทดแทน (a_MI) รายข้อพบว่า ที่อัตราการสูญหาย 10% ค่าที่ได้มีค่าใกล้เคียงกันส่วนใหญ่มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com) และเมื่ออัตราการสูญหายเพิ่มขึ้นเป็น 30% ค่าที่ได้มีค่าใกล้เคียงกัน แต่มีค่าต่ำกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com)

ตารางที่ 4-6 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหาย ของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
40	10	0.5	1,000	1	M	0.7587	0.8899	0.8349	0.9655
					SD	0.3975	0.6611	0.4388	0.5537
				2	M	0.9243	1.0437	0.8171	0.9072
					SD	0.4834	0.6869	0.4482	0.6518
				3	M	0.9216	1.0010	0.8854	0.9539
					SD	0.4999	0.6888	0.4227	0.5682
				4	M	0.8430	0.8512	0.9113	1.0183
					SD	0.3881	0.5715	0.4707	0.5473
			3,000	1	M	0.7429	0.9199	0.7930	0.9239
					SD	0.4632	0.6155	0.1793	0.6030
				2	M	0.8227	0.9612	0.8035	0.9111
					SD	0.4517	0.7313	0.4576	0.6897
				3	M	0.7971	0.8979	0.8378	1.0207
					SD	0.4699	0.5993	0.4915	0.6904
				4	M	0.8474	0.8932	0.7609	0.8180
					SD	0.4518	0.5729	0.4374	0.6637
40	10	1.0	1,000	1	M	0.9059	0.8688	0.8365	0.8882
					SD	0.4442	0.5548	0.4379	0.6316
				2	M	0.8175	0.9228	0.8288	0.8800
					SD	0.4895	0.6541	0.4697	0.6006
				3	M	0.7849	0.9209	0.8285	0.7734
					SD	0.4125	0.5021	0.4580	0.5820
				4	M	0.8103	0.8264	0.8142	0.8336
					SD	0.4367	0.5650	0.4346	0.5579

ตารางที่ 4-6 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
40	10	1.0	3,000	1	M	0.8672	1.1541	0.82036	0.9009
					SD	0.5022	0.6837	0.4549	0.6150
				2	M	0.8263	0.7878	0.8573	0.9325
					SD	0.4357	0.4891	0.4243	0.6219
				3	M	0.8818	0.9150	0.8520	1.0553
					SD	0.4428	0.6582	0.4809	0.6477
				4	M	0.8640	0.8655	0.8674	0.9701
					SD	0.4409	0.5692	0.4583	0.5782
40	30	0.5	1,000	1	M	0.8017	06255	0.8368	0.5559
					SD	0.4454	0.3899	0.4704	0.3584
				2	M	0.8610	0.6773	0.7949	0.5543
					SD	0.4289	0.4119	0.4558	0.3698
				3	M	0.9393	0.6730	0.7296	0.5587
					SD	0.4987	0.4336	0.4005	0.2980
				4	M	0.8384	0.5964	0.8070	0.5744
					SD	0.4679	0.3534	0.4575	0.3354
			3,000	1	M	0.7448	0.6023	0.8294	0.6381
					SD	0.4820	0.3894	0.1401	0.2136
				2	M	0.7388	0.5904	0.8943	0.6636
					SD	0.4666	0.3519	0.1642	0.1937
				3	M	0.7285	0.5524	0.8103	0.6106
					SD	0.4832	0.3535	0.1412	0.2011
				4	M	0.7516	0.6113	0.8837	0.6823
					SD	0.4660	0.3856	0.1247	0.1926

ตารางที่ 4-6 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						a_Com	a_MI	a_Com	a_MI
40	30	1.0	1,000	1	M	0.8950	0.6023	0.8665	0.6806
					SD	0.5219	0.3814	0.2152	0.1785
				2	M	0.8288	0.5904	0.8535	0.6738
					SD	0.4690	0.4221	0.1973	0.1558
				3	M	0.8416	0.5524	0.8422	0.6351
					SD	0.4324	0.4330	0.2105	0.1616
				4	M	0.9106	0.6113	0.9218	0.6473
					SD	0.4584	0.3576	0.2260	0.1844
			3,000	1	M	0.9164	0.6406	0.8368	0.5559
					SD	0.5093	0.4266	0.4704	0.3454
				2	M	0.8044	0.6485	0.7949	0.5534
					SD	0.4305	0.4020	0.4558	0.3708
				3	M	0.8129	0.6323	0.7296	0.5587
					SD	0.4412	0.3698	0.4001	0.2966
				4	M	0.8518	0.6005	0.8070	0.5744
					SD	0.4389	0.4052	0.4575	0.3590

จากตารางที่ 4-6 ผลการวิเคราะห์ค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการประมาณค่าทดแทน (a_MI) รายข้อพบว่า ที่อัตราการสูญหาย 10% ค่าที่ได้มีค่าใกล้เคียงกันส่วนใหญ่มีค่าสูงกว่า ค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com) และเมื่ออัตราการสูญหายเพิ่มขึ้นเป็น 30% ค่าที่ได้มีค่าใกล้เคียงกัน แต่มีค่าต่ำกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com)

ตารางที่ 4-7 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_R) โดยภาพรวม จำแนกตามความยาวของแบบสอบอัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					$b_{R_}Com$	$b_{R_}MI$	$b_{R_}Com$	$b_{R_}MI$
20	10	0.5	1,000	M	0.0136	0.1424	-0.0013	0.0707
				SD	0.0658	0.3357	0.0672	0.3341
			3,000	M	0.0116	-0.0629	-0.0033	0.0099
				SD	0.0699	0.2848	0.6297	0.0684
		1.0	1,000	M	-0.0100	-0.2563	-0.0036	0.0360
				SD	0.0629	1.3421	0.0677	0.1090
			3,000	M	-0.0315	0.0414	-0.2038	0.0320
				SD	0.1010	0.0648	0.0766	0.2322
	30	0.5	1,000	M	-0.0011	0.1294	0.0030	0.0544
				SD	0.0750	0.3179	0.0610	0.2739
			3,000	M	-0.0176	0.0395	0.0224	0.0280
				SD	0.0694	0.1103	0.0806	0.1447
		1.0	1,000	M	0.0324	0.1372	-0.0314	0.3370
				SD	0.0738	0.8335	0.0870	0.9844
			3,000	M	0.0134	-0.0399	0.0206	0.4976
				SD	0.0556	0.4810	0.0802	2.0242

ตารางที่ 4-7 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					b_R_Com	b_R_MI	b_R_Com	b_R_MI
40	10	0.5	1,000	M	0.0085	0.0743	0.0204	0.0218
				SD	0.0809	0.1279	0.0567	0.0804
			3,000	M	-0.0082	0.0135	-0.0097	0.0072
				SD	0.0823	0.1835	0.0787	0.1103
		1.0	1,000	M	0.0119	-0.0443	0.0042	0.0540
				SD	0.0688	0.5114	0.0707	0.2988
			3,000	M	0.0141	0.0601	0.0125	0.0206
				SD	0.0911	0.2602	0.0661	0.2312
	30	0.5	1,000	M	-0.0217	0.0121	0.0206	-0.0508
				SD	0.0739	0.1078	0.0802	0.5130
			3,000	M	0.0008	0.0375	0.0008	-0.1917
				SD	0.0720	0.2638	0.0310	1.4581
		1.0	1,000	M	0.0030	0.0301	0.0055	-0.1480
				SD	0.0937	0.1826	0.0639	1.3016
			3,000	M	-0.0045	0.0938	-0.0170	0.0531
				SD	0.0761	0.2468	0.0767	0.2756

จากตารางที่ 4-7 ผลการวิเคราะห์ค่าความยากของข้อสอบ (b_R) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าความยากของข้อสอบที่ได้จากการประมาณค่าทดแทน (b_R_MI) โดยภาพรวมพบว่า ส่วนใหญ่ค่าที่ได้สูงกว่าค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ (b_R_Com)

ตารางที่ 4-8 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_R) รายข้อ จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_{R_Com}	b_{R_MI}	b_{R_Com}	b_{R_MI}
20	10	0.5	1,000	1	M	-0.0810	1.0784	-0.0999	0.0882
					SD	0.6862	0.3550	0.7657	0.6835
				2	M	0.0309	0.0284	0.0099	-0.0265
					SD	0.6714	0.9430	0.6975	0.6706
				3	M	-0.0615	-0.2872	-0.0176	-0.0454
					SD	0.7717	0.2611	0.7282	0.6866
				4	M	-0.0405	-0.0021	-0.0969	-0.1244
					SD	0.8003	0.3917	0.7912	0.5834
			3,000	1	M	0.0437	-1.2666	-0.0074	0.3306
					SD	0.7654	0.5121	0.7447	0.5024
				2	M	0.1103	-0.0246	0.0533	-0.3608
					SD	0.7781	0.1620	0.7396	0.9525
				3	M	0.0210	-0.0462	-0.0013	0.0163
					SD	0.7086	0.2547	0.7527	0.2318
				4	M	0.0305	-0.2501	0.1659	1.3423
					SD	0.7136	0.2696	0.7227	0.9942

ตารางที่ 4-8 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_R_Com	b_R_MI	b_R_Com	b_R_MI
20	10	1.0	1,000	1	M	-0.0095	0.6341	0.0131	0.0797
					SD	0.7170	0.5624	0.4505	0.6696
				2	M	0.0272	-0.0794	0.0340	-0.0299
					SD	0.6838	0.7461	0.4525	0.5760
				3	M	0.0723	-0.9120	-0.0002	0.0022
					SD	0.7494	0.7036	0.4703	0.6237
				4	M	-0.0339	0.0081	-0.0393	0.1683
					SD	0.7006	0.3649	0.4486	0.7395
			3,000	1	M	0.1462	0.2016	-0.0682	0.9238
					SD	0.6587	0.7536	0.7332	0.8934
				2	M	-0.1117	0.0053	-0.1409	-0.1928
					SD	0.7336	0.2324	0.7421	1.3328
				3	M	-0.0119	-0.0018	-0.0227	0.0875
					SD	0.8262	0.2483	0.6893	0.5894
				4	M	0.1036	0.0128	0.0313	-0.0376
					SD	0.7522	0.2058	0.7029	0.2468
20	30	0.5	1,000	1	M	-0.0933	0.9232	0.0373	0.6399
					SD	0.8056	5.9320	0.8511	0.7583
				2	M	0.0226	-0.1022	-0.0028	-0.3106
					SD	0.7183	1.0684	0.7564	0.9462
				3	M	0.0248	0.1727	-0.0980	0.0378
					SD	0.7351	2.4047	0.7775	0.7081
				4	M	0.1010	0.0752	0.0610	-0.0096
					SD	0.7054	0.5276	0.7027	0.4151

ตารางที่ 4-8 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_R_Com	b_R_MI	b_R_Com	b_R_MI
20	30	0.5	3,000	1	M	0.0026	0.4828	0.2068	0.5314
					SD	0.7135	0.7344	0.4667	1.2672
				2	M	-0.0127	0.0622	0.1532	-0.0215
					SD	0.7786	0.5534	0.4687	0.5783
				3	M	-0.0313	0.0065	0.0215	-0.0884
					SD	0.7547	0.5414	0.4195	0.8401
				4	M	-0.0918	0.2530	0.0476	-0.0135
					SD	0.6638	0.7883	0.4599	0.3626
20	30	1.0	1,000	1	M	0.0512	0.3594	-0.1218	0.4260
					SD	0.7192	0.3224	0.1063	0.7377
				2	M	0.0924	-0.0830	-0.0153	-0.0024
					SD	0.6551	0.5450	1.3233	0.7856
				3	M	-0.0343	0.062	-0.0734	0.0582
					SD	0.7656	0.6903	0.8772	0.7779
				4	M	0.0538	-0.2320	0.0670	0.0883
					SD	0.7813	0.5144	0.5236	0.7575
			3,000	1	M	-0.0021	0.6674	-0.0172	0.7694
					SD	0.7087	0.4916	0.5125	0.3229
				2	M	-0.0287	-0.1891	-0.0823	0.0883
					SD	0.7760	0.7501	0.4136	0.6634
				3	M	-0.0516	-0.0819	0.0052	0.5482
					SD	0.7034	1.5420	0.5209	0.5480
				4	M	-0.0540	-0.0672	0.0504	0.0729
					SD	0.6880	0.6057	0.6081	0.9906

จากตารางที่ 4-8 ผลการวิเคราะห์ค่าความยากของข้อสอบ (b_R) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกันค่าความยากของข้อสอบที่ได้จากการประมาณค่าทดแทน (b_{R_MI}) รายข้อพบว่าค่าที่ได้มีค่าไม่แน่นอน

ตารางที่ 4-9 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_R) รายข้อ จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_R_Com	b_R_MI	b_R_Com	b_R_MI
40	10	0.5	1,000	1	M	0.0699	0.8911	0.0303	0.2461
					SD	0.7618	0.5024	0.7022	0.6763
				2	M	0.0465	0.7408	-0.0509	0.3840
					SD	0.6918	0.3475	0.7056	0.5553
				3	M	0.0526	0.0068	0.0540	0.0817
					SD	0.7242	0.3763	0.8339	0.7802
				4	M	0.0356	-0.0408	0.1236	0.0008
					SD	0.7192	0.6008	0.7457	0.1345
			3,000	1	M	-0.0590	0.9360	-0.0002	0.2453
					SD	0.7372	0.6758	0.4234	0.8106
				2	M	-0.0232	0.2518	0.0805	0.2728
					SD	0.7737	0.5729	0.4576	0.5498
				3	M	-0.0632	0.0313	0.0531	-0.008
					SD	0.8139	0.2779	0.4915	0.1944
				4	M	-0.0766	0.0119	-0.0475	0.0093
					SD	0.6974	0.1671	0.4374	0.3228
40	10	1.0	1,000	1	M	0.1300	0.8211	0.0152	1.5735
					SD	0.7327	0.2035	0.7761	0.5706
				2	M	0.0481	0.7408	-0.0099	0.8161
					SD	0.7068	0.2642	0.8203	0.2937
				3	M	0.1757	0.0068	-0.0785	-0.0007
					SD	0.7920	0.3721	0.8169	0.4612
				4	M	-0.0110	-0.0408	0.4413	0.0043
					SD	0.7795	0.3609	0.7934	0.2830

ตารางที่ 4-9 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_R_Com	b_R_MI	b_R_Com	b_R_MI
40	10	1.0	3,000	1	M	0.0587	1.0140	0.0532	0.4874
					SD	0.7270	0.8259	0.4549	0.5236
				2	M	0.0010	1.1531	-0.0003	0.5906
					SD	0.7659	0.3339	0.4243	0.4588
				3	M	0.2007	0.0031	-0.0586	0.0095
					SD	0.7519	0.2256	0.4809	0.2120
				4	M	0.0747	-0.0325	0.0795	-0.0212
					SD	0.7064	0.2581	0.4583	0.5630
40	30	0.5	1,000	1	M	0.0475	0.3290	0.3394	-0.0746
					SD	0.8061	0.4868	0.2123	0.5073
				2	M	0.0145	0.2964	0.5234	-0.0136
					SD	0.7400	0.1540	0.4824	0.6255
				3	M	-0.1310	-0.0260	-0.1069	-0.0196
					SD	0.7130	0.5982	0.4211	0.5339
				4	M	0.0624	0.1261	-0.0857	-0.0101
					SD	0.8043	0.5082	0.3442	0.5008
			3,000	1	M	0.0174	0.1368	-0.0273	0.3311
					SD	0.7157	0.1095	0.4828	0.1144
				2	M	-0.0851	0.5791	0.1801	0.1846
					SD	0.7675	0.4651	0.6425	0.5889
				3	M	0.0169	1.3872	0.0563	-0.1593
					SD	0.7313	0.9304	0.5867	0.8623
				4	M	-0.0461	0.0742	-0.0373	-0.0106
					SD	0.6717	0.4695	0.6626	0.6423

ตารางที่ 4-9 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_R_Com	b_R_MI	b_R_Com	b_R_MI
40	30	1.0	1,000	1	M	-0.0156	1.1747	-0.0746	0.8251
					SD	0.7511	0.8553	0.7977	0.5789
				2	M	-0.0257	0.7419	-0.0136	0.3564
					SD	0.7008	0.4874	0.4125	0.5821
				3	M	-0.0485	-0.0099	-0.0196	-0.0244
					SD	0.8207	0.7061	0.3169	0.5542
				4	M	0.0158	0.0847	-0.0101	0.7358
					SD	0.7560	0.8021	0.7744	0.5948
			3,000	1	M	-0.0191	0.7555	-0.0199	1.6287
					SD	0.7652	0.9098	0.7096	0.0685
				2	M	-0.0319	0.4889	-0.0129	0.5274
					SD	0.8170	0.3143	0.7034	0.1538
				3	M	0.0056	-0.0392	-0.0553	0.1887
					SD	0.7744	0.4950	0.7398	0.6876
				4	M	0.1156	0.0098	-0.0305	0.1085
					SD	0.7117	0.3798	0.7341	0.7380

จากตารางที่ 4-9 ผลการวิเคราะห์ค่าความยากของข้อสอบ (b_R) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกันค่าความยากของข้อสอบที่ได้จากการประมาณค่าทดแทน (b_R_MI) รายข้อ พบว่าค่าที่ได้มีค่าไม่แน่นอน

ตารางที่ 4-10 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าอำนาจจำแนกของข้อสอบ (b_F) โดยภาพรวม จำแนกตามความยาวของแบบสอบ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					b_F_Com	b_F_MI	b_F_Com	b_F_MI
20	10	0.5	1,000	M	0.5136	0.1047	0.4987	0.0460
				SD	0.0658	0.2207	0.0672	0.1337
			3,000	M	0.5116	0.0195	0.5099	0.0152
				SD	0.0699	0.0697	0.0684	0.0957
		1.0	1,000	M	0.9900	0.0028	0.9964	-0.1246
				SD	0.0629	0.1799	0.0677	0.4606
			3,000	M	0.9685	0.0122	0.9898	0.1361
				SD	0.1010	0.2236	0.0766	0.5265
	30	0.5	1,000	M	0.4989	-0.0502	0.5030	0.0563
				SD	0.0750	0.2584	0.0674	0.2468
			3,000	M	0.4824	0.0588	0.5224	0.0015
				SD	0.0694	0.1621	0.5224	0.1766
		1.0	1,000	M	1.0324	0.8425	0.9686	-0.0745
				SD	0.0738	0.7967	0.0870	0.2403
			3,000	M	1.0134	0.0589	0.9853	0.0593
				SD	0.0556	0.2284	0.0679	0.2994

ตารางที่ 4-10 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					b_F_Com	b_F_MI	b_F_Com	b_F_MI
40	10	0.5	1,000	M	0.5085	0.0820	0.5204	0.0650
				SD	0.0127	0.2393	0.0567	0.2145
			3,000	M	0.4918	0.0014	0.4903	0.0528
				SD	0.0823	0.0858	0.0787	0.2258
		1.0	1,000	M	1.0119	0.0604	1.0042	0.0096
				SD	0.0688	0.2818	0.0707	0.2122
			3,000	M	1.0141	0.0681	1.0125	0.0386
				SD	0.0911	0.2435	0.0661	0.1468
	30	0.5	1,000	M	0.4783	0.2099	0.5206	0.0609
				SD	0.0739	0.6453	0.0802	0.1487
			3,000	M	0.4608	-0.0063	0.5008	0.0099
				SD	0.0720	0.1625	0.0835	0.1363
		1.0	1,000	M	1.0030	-0.0430	1.0055	0.0326
				SD	0.0937	0.3105	0.0639	0.1669
			3,000	M	0.9955	0.1075	0.9830	-0.0377
				SD	0.0761	0.2470	0.0767	0.4485

จากตารางที่ 4-10 ผลการวิเคราะห์ค่าความยากของข้อสอบ (b_F) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าความยากของข้อสอบที่ได้จากการประมาณค่าทดแทน (b_F_MI) โดยภาพรวมพบว่าต่ำกว่าค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ (b_F_Com)

ตารางที่ 4-11 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยากของข้อสอบ (b_F) รายข้อ จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_F_Com	b_F_MI	b_F_Com	b_F_MI
20	10	0.5	1,000	1	M	0.4190	0.2307	0.4005	0.4542
					SD	0.6862	0.5184	0.7657	0.1462
				2	M	0.5309	0.0503	0.5099	-0.0586
					SD	0.6714	0.6324	0.6975	0.4202
				3	M	0.4385	0.4791	0.4823	0.0299
					SD	0.7717	0.4754	0.7282	0.4936
				4	M	0.4595	0.7293	0.4030	0.0321
					SD	0.8003	0.3254	0.7912	0.2966
			3,000	1	M	0.5437	0.2829	0.4926	-0.1167
					SD	0.7654	0.6644	0.7447	0.8049
				2	M	0.6103	0.0155	0.5533	-0.1361
					SD	0.7781	0.2204	0.7396	1.4982
				3	M	0.5210	0.0146	0.4968	0.0115
					SD	0.7086	0.2737	0.7527	0.3666
				4	M	0.5305	-0.0165	0.6659	-0.0828
					SD	0.7136	0.8126	0.7236	1.0886

ตารางที่ 4-11 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_F _Com	b_F _MI	b_F _Com	b_F _MI
20	10	1.0	1,000	1	M	0.9905	0.6585	1.0131	-0.0686
					SD	0.7170	0.8682	0.6923	0.7846
				2	M	1.0272	-0.1320	1.0340	-0.1265
					SD	0.6838	1.8564	0.7225	0.6351
				3	M	1.0723	-0.0040	0.9997	-0.1181
					SD	0.7494	0.1980	0.7754	1.4990
				4	M	0.9661	-0.2725	0.9606	0.0597
					SD	0.7006	0.6113	0.7981	0.2905
			3,000	1	M	1.1462	0.8655	0.9317	0.5881
					SD	0.6587	0.4097	0.7332	1.4695
				2	M	0.8883	0.0121	0.8590	0.0032
					SD	0.7336	0.1449	0.7421	0.5422
				3	M	0.9881	0.0105	0.9772	-0.0013
					SD	0.8262	0.3542	0.6893	0.2639
				4	M	1.1036	0.0171	0.8590	0.0302
					SD	0.7522	0.2380	0.7029	0.4621
20	30	0.5	1,000	1	M	0.4067	0.7141	0.5373	0.1449
					SD	0.8056	2.2033	0.8511	0.3838
				2	M	0.5226	-0.0090	0.4971	-0.3582
					SD	0.7183	0.6160	0.7564	0.9310
				3	M	0.5248	-0.2099	0.4019	0.1900
					SD	0.7351	2.6191	0.7775	1.5334
				4	M	0.6010	0.1192	0.5610	0.0687
					SD	0.7054	0.9764	0.7027	0.7242

ตารางที่ 4-11 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_F_Com	b_F_MI	b_F_Com	b_F_MI
20	30	0.5	3,000	1	M	0.5026	0.6243	0.7068	0.6457
					SD	0.7135	0.8707	0.7366	0.5183
				2	M	0.4873	0.2234	0.6532	-0.3276
					SD	0.7786	0.2806	0.7529	0.3483
				3	M	0.4687	-0.0074	0.5215	0.0578
					SD	0.7547	0.7481	0.7960	0.5351
				4	M	0.4082	-0.1040	0.5476	-0.0275
					SD	0.6638	0.7613	0.6897	0.3572
20	30	1.0	1,000	1	M	0.0512	-0.9480	0.8781	-0.6013
					SD	0.7192	0.3506	0.7377	0.1276
				2	M	0.0924	-0.0992	0.9846	0.1260
					SD	0.6551	0.5043	0.7856	0.7664
				3	M	0.9657	0.0311	0.9265	0.0663
					SD	0.7656	0.6289	0.7779	0.6750
				4	M	0.5038	0.5786	0.1670	0.1516
					SD	0.7813	0.3322	0.7575	0.7437
			3,000	1	M	0.9979	0.7240	0.8992	0.7297
					SD	0.7087	0.4311	0.5081	0.8295
				2	M	0.9713	0.5090	0.8853	-0.1842
					SD	0.7760	0.3049	0.4461	0.5310
				3	M	0.9484	-0.4691	0.1529	0.1000
					SD	0.7034	0.8422	0.5523	0.8224
				4	M	0.9460	0.5230	0.9077	-0.0951
					SD	0.6880	0.7710	0.5451	0.4854

จากตารางที่ 4-11 ผลการวิเคราะห์ค่าความยากของข้อสอบ (b_F) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าความยากของข้อสอบที่ได้จากการประมาณค่าทดแทน (b_{F_MI}) รายข้อพบว่า มีค่าต่ำกว่า ค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ (b_{F_Com})

ตารางที่ 4-12 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าความยาก ของข้อสอบ (b_F) รายข้อ จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_F _Com	b_F _MI	b_F _Com	b_F _MI
40	10	0.5	1,000	1	M	0.5699	0.3276	0.5303	0.3081
					SD	0.7618	0.7585	0.7340	0.4495
				2	M	0.5465	0.2794	0.4490	0.2560
					SD	0.6918	0.6856	0.7140	0.4311
				3	M	0.5526	0.5000	0.6236	0.3880
					SD	0.7242	0.2913	0.7232	0.7140
				4	M	0.5356	0.2231	0.6236	0.6010
					SD	0.7192	0.7741	0.7042	0.3462
			3,000	1	M	0.4410	0.1697	0.4997	0.2607
					SD	0.7372	0.5209	0.7512	0.4532
				2	M	0.4768	0.3114	0.5805	0.2424
					SD	0.7737	0.4231	0.7582	0.8076
				3	M	0.4368	0.4500	0.5531	-0.0183
					SD	0.8139	0.4183	0.7657	0.2237
				4	M	0.4234	-0.0247	0.4521	0.2225
					SD	0.6974	0.6898	0.8291	0.6048
40	10	1.0	1,000	1	M	0.6300	0.8696	1.0152	0.9774
					SD	0.7327	0.5093	0.7761	0.5411
				2	M	0.0481	-0.0155	0.9900	0.5112
					SD	0.7068	0.6815	0.8203	0.7837
				3	M	0.1757	-0.0149	0.9214	-0.0576
					SD	0.7920	0.1848	0.8169	0.3047
				4	M	0.9890	0.3800	0.1441	0.0913
					SD	0.7795	0.6083	0.7934	0.8029

ตารางที่ 4-12 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_F_Com	b_F_MI	b_F_Com	b_F_MI
40	10	1.0	3,000	1	M	0.5087	0.4812	0.5032	0.5956
					SD	0.7270	0.9383	0.7893	0.9874
				2	M	0.6010	0.7021	0.9996	0.6969
					SD	0.7659	0.8288	0.7377	0.8623
				3	M	0.2007	-0.0078	0.9413	0.8100
					SD	0.7519	0.1746	0.7066	0.1242
				4	M	0.0747	0.0289	0.0795	0.0978
					SD	0.7064	0.3007	0.7352	0.7664
40	30	0.5	1,000	1	M	0.5475	0.6201	0.4827	0.3791
					SD	0.8061	0.2050	0.5073	0.6649
				2	M	0.5145	0.4149	0.4176	0.5715
					SD	0.7400	0.6011	0.6255	0.8881
				3	M	0.3690	0.5034	0.5052	0.1473
					SD	0.7130	0.6500	0.5339	0.6299
				4	M	0.5624	0.4079	0.5504	-0.1289
					SD	0.8043	0.5024	0.5008	0.6661
			3,000	1	M	0.4774	0.4822	0.4726	0.5743
					SD	0.7295	0.6370	0.4828	0.6609
				2	M	0.3749	0.2581	0.6801	0.1621
					SD	0.7751	0.5248	0.6425	0.6733
				3	M	0.4769	-0.0507	0.5563	-0.1142
					SD	0.7448	0.6204	0.5867	0.1507
				4	M	0.4139	0.2045	0.4626	-0.0877
					SD	0.6826	0.8029	0.6626	0.2858

ตารางที่ 4-12 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
						MAR		MNAR	
						b_F_Com	b_F_MI	b_F_Com	b_F_MI
40	30	1.0	1,000	1	M	0.9844	0.8410	0.9253	0.6017
					SD	0.7511	0.9590	0.5789	0.1921
				2	M	0.9743	0.6534	0.9863	0.6778
					SD	0.7008	0.1372	0.5821	0.9400
				3	M	0.9515	-0.0108	0.9803	0.2088
					SD	0.8207	0.6302	0.5542	0.4321
				4	M	0.0158	0.0508	0.9898	0.2040
					SD	0.7560	0.4337	0.5948	0.2585
			3,000	1	M	0.9809	0.4580	0.9800	0.7020
					SD	0.7652	0.4136	0.7069	0.4702
				2	M	0.9681	1.8514	0.9870	0.6053
					SD	0.8170	0.4428	0.7034	0.4577
				3	M	0.0056	-0.0105	0.0446	-0.0527
					SD	0.7744	0.2983	0.7398	0.4038
				4	M	0.1156	0.0386	0.0694	-0.1551
					SD	0.7117	0.6771	0.7341	0.5825

จากตารางที่ 4-12 ผลการวิเคราะห์ค่าความยากของข้อสอบ (b_F) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าความยากของข้อสอบที่ได้จากการประมาณค่าทดแทน (b_F_MI) รายข้อพบว่า มีค่าต่ำกว่าค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ (b_F_Com)

ตอนที่ 2 ผลการวิเคราะห์ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) **ด้วยวิธีจัดการข้อมูลสูญหายโดยใช้วิธีประมาณค่าพหุ** (Multiple imputation: MI)

การนำเสนอผลการวิเคราะห์ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ในเบื้องต้นนี้เป็นการนำเสนอผลการวิเคราะห์ด้วยสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF เพื่อบรรยายลักษณะค่าพารามิเตอร์ข้อสอบที่ประมาณค่าได้ด้วยการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) และประมาณค่าพารามิเตอร์ ด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด (Maximum Likelihood: ML) รายละเอียดแสดง ดังนี้

ตารางที่ 4-13 ค่าเฉลี่ย (M) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) โดยภาพรวม จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					θ_Com	θ_MI	θ_Com	θ_MI
20	10	0.5	1,000	M	-0.0016	-0.0011	0.0005	-0.0008
				SD	0.1001	0.0908	0.9823	0.7294
			3,000	M	0.0004	0.0010	0.0003	-0.0043
				SD	0.0964	0.0925	0.9822	0.7504
		1.0	1,000	M	-0.0026	0.0003	0.0084	-0.0011
				SD	0.0993	0.0893	0.9811	0.7222
			3,000	M	0.0014	-0.0093	0.0023	0.0006
				SD	0.1009	0.0940	-0.0011	0.7119
	30	0.5	1,000	M	0.0018	0.0013	0.0006	-0.0011
				SD	0.0993	0.0857	0.9828	0.7176
			3,000	M	0.0018	-0.0007	0.0006	-0.0013
				SD	0.0971	0.0869	0.9862	0.7236
		1.0	1,000	M	-0.0001	-0.0044	0.0063	0.0003
				SD	0.0991	0.0870	0.9854	0.7141
			3,000	M	-0.0003	-0.0003	0.0012	-0.0009
				SD	0.0974	0.0948	0.9854	0.7155

ตารางที่ 4-13 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ค่าสถิติ	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					θ_Com	θ_MI	θ_Com	θ_MI
40	10	0.5	1,000	M	-0.0009	0.0025	0.0007	-0.0028
				SD	0.1003	0.0994	0.9879	0.7422
			3,000	M	0.0006	0.0006	0.0012	-0.0015
				SD	0.0987	0.0971	0.9853	0.7626
		1.0	1,000	M	0.0035	-0.0100	0.0034	-0.0102
				SD	0.1010	0.1008	0.9789	0.7800
			3,000	M	-0.0033	-0.0014	0.0016	0.0032
				SD	0.0976	0.1035	0.9864	0.7518
	30	0.5	1,000	M	-0.0009	0.0012	0.0039	0.0075
				SD	0.0969	0.1003	0.9866	0.7568
			3,000	M	-0.0025	-0.0022	0.0010	0.0043
				SD	0.0949	0.1020	0.9844	0.8771
		1.0	1,000	M	0.0007	-0.0010	0.0015	-0.0014
				SD	0.1007	0.0950	0.9841	0.7477
			3,000	M	-0.0027	-0.0017	0.0008	-0.0012
				SD	0.0992	0.1048	0.9825	0.7622

จากตารางที่ 4-13 ผลการวิเคราะห์ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) โดยภาพรวม พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่าพารามิเตอร์ความสามารถของผู้สอบที่ได้จากการประมาณค่าทดแทน (θ_MI) ส่วนใหญ่มีแนวโน้มใกล้เคียงค่าความสามารถของผู้สอบของข้อมูลที่สมบูรณ์ (θ_Com) มากกว่าค่าพารามิเตอร์ ความสามารถของผู้สอบภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน

ตอนที่ 3 ผลการเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ข้อสอบ (ค่าอำนาจจำแนก และค่าความยาก)

การนำเสนอผลการเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ข้อสอบ ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF โดยเป็นการนำเสนอผลการวิเคราะห์ด้วยการเปรียบเทียบค่า diff และ diff error รายละเอียดแสดงดังนี้

ตารางที่ 4-14 ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) **โดยภาพรวม** จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	สภาวะการสูญหายของข้อมูล			
				MAR		MNAR	
				diff	diff error	diff	diff error
20	10	0.5	1,000	-0.3409	0.9935	-0.3005	1.0004
			3,000	-0.3116	0.9480	-0.2882	0.9005
		1.0	1,000	-0.3472	1.1878	-0.2968	0.8906
			3,000	-0.3169	0.9584	-0.3294	0.9546
	30	0.5	1,000	-0.2703	1.0235	-0.2535	0.9201
			3,000	-0.2274	0.8739	-0.2080	0.8539
		1.0	1,000	-0.0290	1.6322	-0.2218	0.8706
			3,000	-0.2128	0.8539	-0.2397	0.9140

ตารางที่ 4-14 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	สภาวะการสูญหายของข้อมูล			
				MAR		MNAR	
				diff	diff error	diff	diff error
40	10	0.5	1,000	-0.2711	0.9099	-0.3431	0.8870
			3,000	-0.2901	0.8744	-0.3211	0.9077
		1.0	1,000	-0.2867	0.9110	-0.2782	0.9129
			3,000	-0.2944	0.8500	-0.3268	0.8884
	30	0.5	1,000	-0.2653	0.8817	-0.2945	0.8764
			3,000	-0.3440	0.8915	-0.2511	0.8874
		1.0	1,000	-0.2994	0.9069	-0.0875	1.9021
			3,000	-0.2570	0.8574	-0.2845	0.8746

จากตารางที่ 4-14 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม(MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพรวมพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff ค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-15 ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
20	10	0.5	1,000	1	-0.1300	0.6921	-0.1382	0.6940
				2	-0.1286	0.7051	-0.0883	0.6731
				3	-0.1203	0.7587	-0.0601	0.6858
				4	-0.2789	0.7423	-0.0677	0.5844
			3,000	1	-0.1344	0.7251	-0.1580	0.7012
				2	-0.0177	0.6040	-0.0882	0.6501
				3	-0.0768	0.6061	-0.0118	0.6084
				4	-0.0606	0.6290	-0.1051	0.6077
		1.0	1,000	1	-0.1181	0.6767	-0.0705	0.6701
				2	-0.1106	0.6809	-0.1200	0.5856
				3	-0.0858	0.6897	-0.0564	0.6232
				4	-0.1962	0.6856	-0.1389	0.7489
			3,000	1	0.0230	0.6630	-0.1176	0.7713
				2	-0.1548	0.7160	-0.1908	0.7139
				3	-0.1798	0.7268	-0.1206	0.6648
				4	-0.2020	0.6937	-0.1897	0.6739

ตารางที่ 4-15 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
20	30	0.5	1,000	1	0.2998	0.5508	0.2124	0.5000
				2	0.2004	0.4769	0.2272	0.4633
				3	0.2099	0.5156	0.1624	0.4831
				4	0.2251	0.4653	0.2066	0.4599
			3,000	1	0.1802	0.4260	0.1526	0.4249
				2	0.2278	0.4751	0.2100	0.4524
				3	0.1913	0.4566	0.1990	0.4541
				4	0.1860	0.4522	0.1941	0.4601
		1.0	1,000	1	0.1901	0.5225	0.2846	0.5237
				2	0.2226	0.5509	0.3021	0.5380
				3	0.1715	0.6365	0.2092	0.4337
				4	0.2184	0.5434	0.1885	0.4389
			3,000	1	0.2373	0.5271	0.3611	0.4999
				2	0.1523	0.4117	0.2144	0.4528
				3	0.2439	0.4718	0.2453	0.4625
				4	0.2231	0.4591	0.2574	0.5055

จากตารางที่ 4-15 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff รายข้อพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า BIAS มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-16 ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
40	10	0.5	1,000	1	-0.1105	0.6153	-0.1313	0.5504
				2	-0.0629	0.6567	-0.0984	0.6763
				3	-0.0253	0.6945	-0.1260	0.5729
				4	-0.0978	0.6112	-0.1307	0.5912
			3,000	1	-0.1933	0.6323	-0.1665	0.5630
				2	-0.1734	0.6830	-0.1283	0.6209
				3	-0.1198	0.5871	-0.1954	0.7068
				4	-0.0556	0.5547	-0.0391	0.6349
		1.0	1,000	1	-0.0778	0.6169	-0.0907	0.6648
				2	-0.1336	0.6840	-0.0538	0.5915
				3	-0.1373	0.5930	0.0095	0.5959
				4	-0.0756	0.5634	-0.0222	0.5347
			3,000	1	-0.0360	0.5118	-0.0622	0.6061
				2	0.0008	0.4802	-0.1055	0.6401
				3	-0.0980	0.6333	-0.1979	0.6778
				4	-0.0350	0.5508	-0.0973	0.6045

ตารางที่ 4-16 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
40	30	0.5	1,000	1	0.1975	0.4149	0.1840	0.3900
				2	0.1907	0.4192	0.2107	0.4246
				3	0.3011	0.4891	0.1926	0.3860
				4	0.2051	0.4338	0.1488	0.4217
			3,000	1	0.1516	0.3992	0.2256	0.4283
				2	0.1431	0.3751	0.2257	0.4491
				3	0.1728	0.4039	0.2414	0.4500
				4	0.1305	0.4134	0.2131	0.4138
		1.0	1,000	1	0.2698	0.5135	0.2075	0.4614
				2	0.1934	0.4822	0.1875	0.4445
				3	0.2389	0.4880	0.1813	0.4671
				4	0.3257	0.4902	0.2241	0.5034
			3,000	1	0.3117	0.4778	0.2692	0.4366
				2	0.1607	0.4963	0.2251	0.4322
				3	0.1727	0.4003	0.1597	0.3356
				4	0.2403	0.4381	0.1974	0.4081

จากตารางที่ 4-16 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff รายข้อพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-17 ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) โดยภาพรวม จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	สภาวะการสูญหายของข้อมูล			
				MAR		MAR	
				diff	diff error	diff	diff error
20	10	0.5	1,000	-0.3350	0.9741	-0.3083	0.9864
			3,000	-0.3101	0.9449	-0.2586	0.8255
		1.0	1,000	-0.3231	0.9533	-0.2944	0.9028
			3,000	-0.2614	0.8928	-0.3278	0.9391
	30	0.5	1,000	-0.2712	1.0394	-0.2580	0.9608
			3,000	-0.2244	0.8665	-0.2170	0.8460
		1.0	1,000	-0.2071	1.0894	-0.2463	0.8974
			3,000	-0.2315	0.8240	-0.2251	0.9287
40	10	0.5	1,000	-0.2684	0.9250	-0.3301	0.8641
			3,000	-0.2778	0.8920	-0.3104	0.9134
		1.0	1,000	-0.2476	0.8955	-0.2787	0.8639
			3,000	-0.2603	0.9441	-0.3299	0.8895
	30	0.5	1,000	-0.2565	0.8838	-0.2825	0.8640
			3,000	-0.3435	0.8837	-0.2847	0.8718
		1.0	1,000	-0.3019	0.9084	-0.2680	0.8899
			3,000	-0.2429	0.8863	-0.2501	0.8356

จากตารางที่ 4-17 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพรวมพบว่า มีค่าเข้าใกล้ 0

เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-18 ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
20	10	0.5	1,000	1	-0.1803	0.7460	-0.1499	0.6990
				2	-0.0957	0.6870	-0.1601	0.7893
				3	-0.1538	0.7543	-0.0627	0.6498
				4	-0.2738	0.7119	-0.0908	0.6382
			3,000	1	-0.1187	0.7022	-0.0994	0.6521
				2	-0.0339	0.5815	-0.0905	0.6097
				3	-0.1070	0.6312	-0.0092	0.5778
				4	-0.0548	0.6436	-0.0476	0.5392
		1.0	1,000	1	-0.1378	0.6501	-0.1212	0.6695
				2	-0.0853	0.6700	-0.1262	0.6891
				3	-0.1606	0.7519	-0.0013	0.6259
				4	-0.1768	0.6541	-0.1072	0.7400
			3,000	1	0.0368	0.6188	-0.1100	0.7412
				2	-0.1156	0.7076	-0.1771	0.6970
				3	-0.1266	0.6758	-0.1412	0.6693
				4	-0.1385	0.6064	-0.1909	0.6590

ตารางที่ 4-18 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
20	30	0.5	1,000	1	0.2595	1.2137	0.2350	0.5364
				2	0.2477	0.8114	0.2432	0.5026
				3	0.2207	0.7075	0.1544	0.4883
				4	0.2112	1.0844	0.1858	0.4569
			3,000	1	0.1746	0.4148	0.1742	0.4362
				2	0.2130	0.4598	0.2099	0.4447
				3	0.1931	0.4542	0.1978	0.4332
				4	0.1763	0.4370	0.1881	0.4401
		1.0	1,000	1	0.1678	0.5776	0.2648	0.4865
				2	0.2812	0.6636	0.2579	0.4850
				3	0.2114	0.4665	0.2178	0.5052
				4	0.2432	0.5069	0.1597	0.4153
			3,000	1	0.2618	0.4729	0.2693	0.8599
				2	0.1390	0.4075	0.2014	0.4636
				3	0.2461	0.4594	0.2281	0.4496
				4	0.1973	0.4464	0.2900	0.5593

จากตารางที่ 4-18 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff รายข้อพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-19 ค่าเฉลี่ย diff และ diff error ของค่าอำนาจจำแนกของข้อสอบ (a) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
40	10	0.5	1,000	1	-0.1312	0.6708	-0.1306	0.5662
				2	-0.1194	0.6938	-0.0900	0.6548
				3	-0.0794	0.6899	-0.0684	0.5695
				4	-0.0082	0.5687	-0.1069	0.5550
			3,000	1	-0.1770	0.6374	-0.1308	0.6142
				2	-0.1385	0.7407	-0.1076	0.6946
				3	-0.1008	0.6047	-0.1829	0.7109
				4	-0.0458	0.5719	-0.0571	0.6629
		1.0	1,000	1	-0.0512	0.5544	-0.0516	0.6306
				2	-0.0717	0.6548	-0.0512	0.5999
				3	-0.0810	0.5061	0.0552	0.5817
				4	-0.0420	0.5637	-0.0193	0.5555
			3,000	1	-0.2869	0.3676	-0.0805	0.6172
				2	0.0385	0.4882	-0.0751	0.6233
				3	-0.0332	0.6557	-0.2051	0.6763
				4	-0.0015	0.5663	-0.1027	0.5844

ตารางที่ 4-19 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
40	30	0.5	1,000	1	0.1762	0.4261	0.1992	0.3909
				2	0.1837	0.4491	0.2075	0.4471
				3	0.2663	0.5070	0.1819	0.3754
				4	0.2420	0.4268	0.1292	0.4210
			3,000	1	0.1425	0.4128	0.1913	0.4188
				2	0.1484	0.3803	0.2308	0.4646
				3	0.1761	0.3933	0.1998	0.4240
				4	0.1403	0.4085	0.2015	0.4051
		1.0	1,000	1	0.2566	0.4595	0.1859	0.4597
				2	0.1823	0.4602	0.1797	0.4319
				3	0.2166	0.4861	0.2072	0.4505
				4	0.3163	0.4798	0.2745	0.5079
			3,000	1	0.3069	0.5238	0.2809	0.4540
				2	0.1887	0.4423	0.2416	0.4402
				3	0.1703	0.4055	0.1710	0.3424
				4	0.2335	0.4659	0.2327	0.4069

จากตารางที่ 4-19 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff รายข้อพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-20 ค่าเฉลี่ย diff และ diff error ของค่าความยากของข้อสอบ (b_R) โดยภาพรวมจากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	สภาวะการสูญหายของข้อมูล			
				MAR		MNAR	
				diff	diff error	diff	diff error
20	10	0.5	1,000	-0.1288	1.6228	-0.0720	1.2505
			3,000	0.0745	0.4921	0.0133	1.1354
		1.0	1,000	0.2462	1.6375	-0.0397	0.9419
			3,000	-0.0729	0.5822	-0.0422	0.8326
	30	0.5	1,000	-0.1304	1.4312	-0.0514	1.4130
			3,000	-0.0572	0.4010	-0.0056	0.6162
		1.0	1,000	0.0658	1.8188	-0.3685	1.3259
			3,000	0.0533	1.7082	-0.5123	1.4215
40	10	0.5	1,000	-0.0658	0.6890	-0.0014	0.3875
			3,000	-0.0217	0.5649	-0.0169	0.5106
		1.0	1,000	1.0562	1.0740	-0.0498	1.0983
			3,000	-0.0459	0.8292	-0.0082	0.7823
	30	0.5	1,000	-0.0338	0.2684	0.0714	1.4332
			3,000	-0.0367	0.9526	0.1925	1.0629
		1.0	1,000	-0.0275	1.0456	0.1535	1.7908
			3,000	-0.0983	0.7988	-0.0702	0.7355

จากตารางที่ 4-20 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าความยากของข้อสอบ (b_R) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพพบว่า ส่วนใหญ่มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 น้อยกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff error มีค่าสูงกว่าค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-21 ค่าเฉลี่ย diff และ diff error ของค่าความยากของข้อสอบ (b_R) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
20	10	0.5	1,000	1	-1.1594	0.1037	-0.1881	1.8303
				2	0.0025	0.9382	0.0365	0.3359
				3	0.2257	0.3253	0.0278	0.4078
				4	-0.0384	0.3916	0.0274	1.4091
			3,000	1	1.3103	1.4815	-0.3380	0.6034
				2	0.1349	0.2101	0.2414	0.2395
				3	0.0672	0.2621	-0.0176	0.2314
				4	0.0702	0.2773	-1.1764	1.9920
		1.0	1,000	1	-0.6436	0.8528	-0.0666	0.4514
				2	0.1066	0.7500	0.0640	0.2363
				3	0.9843	0.8715	-0.0024	0.4572
				4	-0.0420	0.3655	-0.2076	1.0216
			3,000	1	-0.6695	1.2568	-0.9920	1.9472
				2	0.1604	0.7634	0.0519	0.3272
				3	0.0303	1.5346	-0.1102	0.5968
				4	0.0132	0.6028	0.0694	0.2552

ตารางที่ 4-21 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
20	30	0.5	1,000	1	-1.0165	0.5206	-0.6026	0.2810
				2	0.1248	0.4940	0.3077	0.3938
				3	-0.1479	0.5371	-0.1358	0.7175
				4	0.0258	0.4678	0.0706	0.4191
			3,000	1	-0.4802	0.8744	-0.3245	1.3020
				2	-0.0749	0.5557	0.1747	0.6013
				3	-0.0378	0.5400	0.1099	0.8431
				4	-0.1171	0.7930	0.0612	0.3660
		1.0	1,000	1	-0.3543	0.2338	-0.5479	0.2819
				2	0.1754	0.5699	-0.0130	1.3168
				3	-0.0963	0.6936	-0.1317	0.8828
				4	0.2858	0.4501	-0.0213	0.5214
			3,000	1	-0.6695	0.2739	-0.8701	1.5778
				2	0.1604	0.9123	-0.2030	0.6906
				3	0.0303	1.4175	-0.3953	0.5343
				4	0.0132	0.4230	-0.9136	1.9843

จากตารางที่ 4-21 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าความยากของข้อสอบ (b_R) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff รายข้อพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-22 ค่าเฉลี่ย diff และ diff error ของค่าความยากของข้อสอบ (b_R) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มอ้างอิง

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
40	10	0.5	1,000	1	-0.2096	0.5420	-0.2157	0.7067
				2	-0.3694	1.3907	-0.4350	0.7032
				3	0.0310	0.3757	-0.0277	0.7768
				4	-0.1042	1.0607	0.1229	0.1817
			3,000	1	-0.9950	0.5734	-0.2456	0.8431
				2	-0.2750	0.6329	-0.1922	0.5799
				3	-0.0945	0.2922	0.0611	0.2029
				4	-0.0885	0.1884	-0.0568	0.3262
		1.0	1,000	1	0.3089	0.4204	-1.5582	0.9649
				2	0.3073	0.4227	-0.8260	0.7304
				3	1.1689	1.2261	-0.0778	0.4654
				4	1.0298	1.0906	0.0398	0.2849
			3,000	1	-0.9553	1.3925	-0.4315	1.5762
				2	-1.1503	1.7387	-0.5909	0.7468
				3	0.1976	0.2991	-0.0681	0.2217
				4	0.1072	0.2783	0.1008	0.5692

ตารางที่ 4-22 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
40	30	0.5	1,000	1	-0.2817	1.5059	-0.3566	0.5809
				2	-0.2819	1.1824	-0.6058	1.3545
				3	-0.1050	0.6044	0.1148	0.3123
				4	-0.0637	0.5097	0.1362	1.5295
			3,000	1	-0.1194	0.3007	-0.3584	1.1098
				2	-0.6642	0.3511	-0.0044	0.7636
				3	-1.3703	0.1294	0.2157	0.9488
				4	-0.1203	0.4824	-0.0267	0.7979
		1.0	1,000	1	-0.7687	0.2014	-0.8997	0.2646
				2	-0.5155	0.2517	-0.3701	0.3795
				3	-0.0494	1.4799	0.0047	0.1418
				4	0.0050	1.0682	-0.7459	0.5341
			3,000	1	-1.1936	0.5029	-1.6481	1.2236
				2	-0.7738	0.2260	-0.5403	0.5176
				3	0.0155	0.4927	-0.2441	0.6853
				4	0.0309	0.3792	-0.1391	0.7474

จากตารางที่ 4-22 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าความยากของข้อสอบ (b_R) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff รายข้อพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-23 ค่าเฉลี่ย diff และ diff error ค่าอำนาจจำแนกของข้อสอบ (b_F) โดยภาพรวม จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล ของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	สภาวะการสูญหายของข้อมูล			
				MAR		MNAR	
				diff	diff error	diff	diff error
20	10	0.5	1,000	-1.3441	1.1695	0.4528	0.8338
			3,000	0.4921	0.5997	0.4947	0.9790
		1.0	1,000	0.9872	1.3052	1.1210	1.4135
			3,000	0.9564	1.3301	0.8537	1.0111
	30	0.5	1,000	0.5168	1.5612	0.4467	1.6161
			3,000	0.4237	0.8319	0.5209	0.9640
		1.0	1,000	1.1616	1.8396	1.3930	1.8875
			3,000	0.9545	1.5264	0.9236	1.3755
40	10	0.5	1,000	0.4267	1.0011	0.4554	0.9204
			3,000	0.4904	0.6522	0.4374	1.0861
		1.0	1,000	0.9515	1.7873	0.5108	1.1114
			3,000	0.9470	1.4294	0.9738	1.0654
	30	0.5	1,000	0.2684	0.2553	0.4596	1.0227
			3,000	0.4669	0.8563	0.4909	0.7930
		1.0	1,000	1.0456	0.8140	0.9729	1.3686
			3,000	0.8880	1.2010	1.0207	1.0058

จากตารางที่ 4-23 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าความยากของข้อสอบ (b_F) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) โดยภาพรวมพบว่า ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-24 ค่าเฉลี่ย diff และ diff error ค่าอำนาจจำแนกของข้อสอบ (b_F) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 20 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
20	10	0.5	1,000	1	0.1883	0.5491	-0.0541	0.1362
				2	0.4806	0.7918	0.5686	0.7057
				3	-0.4750	0.5226	0.4525	0.6678
				4	-0.2698	0.3261	0.3709	0.4740
			3,000	1	0.2608	0.7107	0.6093	0.8346
				2	0.5948	0.6339	0.6894	1.6424
				3	0.5064	0.5750	0.4872	0.6086
				4	0.5470	0.9762	0.7487	1.3168
		1.0	1,000	1	0.3320	0.9255	1.0817	1.2683
				2	1.1592	1.1807	1.1606	1.3215
				3	1.0763	1.0942	1.1179	1.8639
				4	1.2386	1.2878	0.9010	0.9462
			3,000	1	0.2807	0.3404	0.3437	1.5020
				2	0.8762	0.8880	0.8559	1.0117
				3	0.9476	1.0110	0.9786	1.0132
				4	1.0865	1.1120	1.0016	1.1021

ตารางที่ 4-24 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
20	30	0.5	1,000	1	-0.3074	0.5892	0.3924	0.4042
				2	0.5316	1.0703	0.8553	1.0038
				3	0.7347	1.2397	0.2120	1.5404
				4	0.4818	0.5255	0.4923	0.8728
			3,000	1	-0.1217	1.8653	0.0612	1.5120
				2	0.2639	1.3013	0.9808	1.3473
				3	0.4761	0.8836	0.4637	0.5648
				4	0.5122	0.9144	0.5752	0.6762
		1.0	1,000	1	0.3999	0.4134	0.8479	1.8512
				2	1.1916	1.5875	0.8586	1.9561
				3	0.9346	1.1247	0.8602	1.0914
				4	0.4752	0.5327	0.9154	1.1771
			3,000	1	0.2739	1.4501	-0.3804	0.7197
				2	0.9123	1.5868	1.0695	0.5729
				3	1.4175	1.4077	1.0529	1.3867
				4	0.4230	0.5758	1.0029	1.2192

จากตารางที่ 4-24 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าความยากของข้อสอบ (b_F) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) รายข้อพบว่า และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff รายข้อพบว่าส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตารางที่ 4-25 ค่าเฉลี่ย diff และ diff error ค่าอำนาจจำแนกของข้อสอบ (b_F) รายข้อ จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ 40 ข้อ อัตราการสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูลของกลุ่มเปรียบเทียบ

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
40	10	0.5	1,000	1	0.2423	0.7927	0.2223	0.4995
				2	0.2671	0.7326	0.1930	0.4704
				3	0.5026	0.5801	0.5202	0.5833
				4	0.3125	1.7926	0.6127	0.7029
			3,000	1	0.2713	1.5374	0.2390	0.5104
				2	0.1654	0.4523	0.3382	0.8718
				3	0.4323	0.4568	0.5714	0.6132
				4	0.4481	0.4863	-0.7700	1.0688
		1.0	1,000	1	0.2604	0.3655	-0.4470	0.3552
				2	1.0636	1.6733	-0.0622	0.5755
				3	1.1906	1.2047	0.6116	0.6826
				4	0.9852	1.1562	0.5324	1.1311
			3,000	1	0.5775	1.0132	0.4576	1.0838
				2	0.2989	0.8771	0.3028	0.9099
				3	1.2085	1.2210	0.9333	0.9414
				4	1.0458	1.0878	0.9818	1.2432

ตารางที่ 4-25 (ต่อ)

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	ข้อที่	สภาวะการสูญหายของข้อมูล			
					MAR		MNAR	
					diff	diff error	diff	diff error
40	30	0.5	1,000	1	-0.5726	0.1316	0.1037	0.8180
				2	0.0996	0.2590	-0.1539	1.3758
				3	-0.1344	0.5623	0.3579	0.8670
				4	0.5145	0.7174	0.6794	1.0588
			3,000	1	-0.0048	1.6288	-0.1016	0.8153
				2	0.1168	1.5217	0.5181	0.9669
				3	0.5276	0.8121	0.6706	0.7739
				4	0.3894	0.8887	0.5431	0.7602
		1.0	1,000	1	0.5106	1.9999	0.3237	0.5293
				2	0.8257	1.4946	0.3085	1.0128
				3	0.9620	0.7335	0.9515	1.1547
				4	0.9772	0.7980	0.9695	1.0936
			3,000	1	0.1399	0.2406	0.9080	1.4664
				2	-0.3853	0.4437	-0.6182	0.7446
				3	1.0164	1.0589	0.9973	1.4400
				4	1.0648	1.2600	1.1245	1.9349

จากตารางที่ 4-25 ผลการวิเคราะห์ค่า diff และค่า RMSE ของค่าความยากของข้อสอบ (b_F) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) รายข้อพบว่า และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff รายข้อพบว่าส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตอนที่ 4 ผลการศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ)

การนำเสนอผลการศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF โดยเป็นการนำเสนอผลการวิเคราะห์ด้วยการเปรียบเทียบค่า diff และ diff error รายละเอียดแสดงดังนี้

ตารางที่ 4-26 ค่าเฉลี่ย diff และ diff error ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) โดยภาพรวม จากการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF ขนาดกลุ่มตัวอย่าง และสภาวะการสูญหายของข้อมูล

ความยาวแบบสอบ	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ขนาดกลุ่มตัวอย่าง	สภาวะการสูญหายของข้อมูล			
				MAR		MNAR	
				diff	diff error	diff	diff error
20	10	0.5	1,000	-0.0005	0.9079	0.8063	1.3229
			3,000	-0.0014	0.8938	-1.3478	0.9188
		1.0	1,000	-.00030	0.9084	0.8092	1.3317
			3,000	0.0106	0.9208	-0.0000	0.9000
	30	0.5	1,000	0.0005	0.8897	0.7845	1.2971
			3,000	-0.0026	0.8950	-0.0000	0.8933
		1.0	1,000	0.0043	0.8904	0.7901	1.3109
			3,000	-0.0000	0.8990	-0.0000	0.8923
40	10	0.5	1,000	-0.0019	0.9678	0.9214	1.4813
			3,000	-0.0001	0.9708	0.0000	0.9665
		1.0	1,000	0.0145	0.9863	0.9925	1.5576
			3,000	-0.0015	0.9700	0.0000	0.9648
	30	0.5	1,000	-0.0021	0.9652	0.9310	1.5033
			3,000	-0.0003	0.9666	0.8738	1.4996
		1.0	1,000	0.0017	0.9677	0.0034	0.9679
			3,000	-0.0009	0.9632	0.9733	1.5374

จากตารางที่ 4-26 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพพบว่า มีค่าใกล้เคียงกัน และที่ขนาดอิทธิพล DIF 1.0 ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff เข้าใกล้ 0 มากกว่าภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ตอนที่ 5 ผลการตรวจสอบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบ

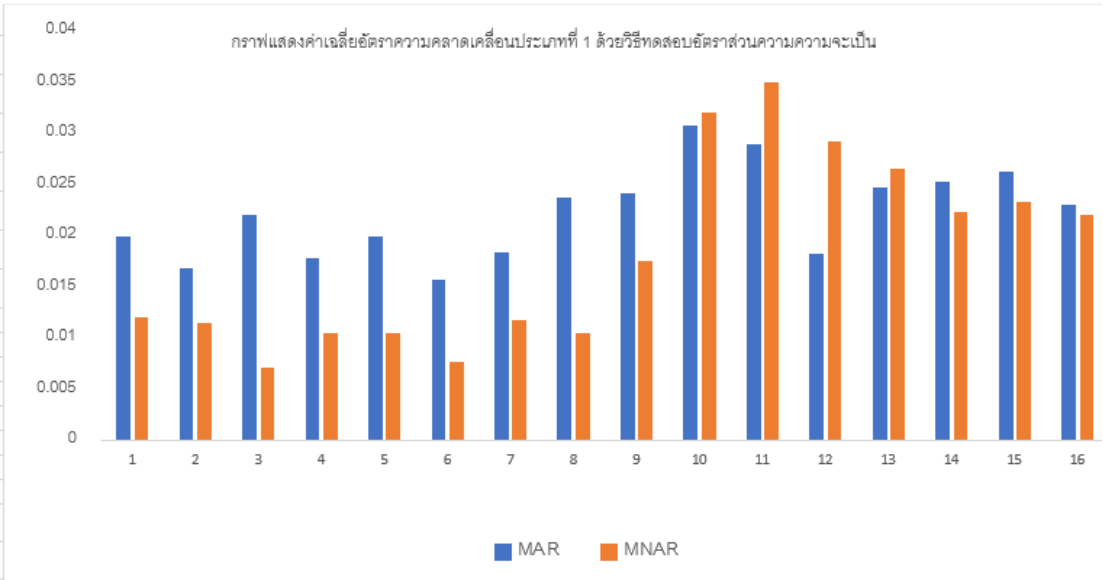
ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จะพิจารณาจากผลการวิเคราะห์ค่าความคลาดเคลื่อนประเภทที่ 1 (Type I error) และอำนาจการทดสอบ (Power rate) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำแนกตามเงื่อนไขที่แตกต่างกันดังนี้

ตารางที่ 4-27 ค่าเฉลี่ยของความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF และสภาวะการสูญหายของข้อมูล

ความยาวแบบสอบ	ขนาดกลุ่มตัวอย่าง	อัตราการสูญหาย	ขนาดอิทธิพลของ DIF	ค่าเฉลี่ยของอัตราความคลาดเคลื่อนประเภทที่ 1	
				MAR	MNAR
20	1,000	10	0.5	0.0200*	0.0121*
			1.0	0.0168*	0.0115*
		30	0.5	0.0221*	0.0072*
			1.0	0.0179*	0.0105*
	3,000	10	0.5	0.0200*	0.0105*
			1.0	0.0158*	0.0077*
		30	0.5	0.0184*	0.0118*
			1.0	0.0237*	0.0105*
40	1,000	10	0.5	0.0242*	0.0176*
			1.0	0.0308*	0.0321*
		30	0.5	0.0290*	0.0350*
			1.0	0.0182*	0.0292*
	3,000	10	0.5	0.0247*	0.0266*
			1.0	0.0253*	0.0224*
		30	0.5	0.0263*	0.0234*
			1.0	0.0230*	0.0221*

* $p < 0.05$

จากตารางที่ 4-27 ภายใต้ความยาวข้อสอบ 20 และ 40 ข้อ ในเงื่อนไขขนาดอิทธิพล DIF 0.5 และ 1.0 ขนาดตัวอย่าง 1,000 และ 3,000 ที่อัตราการสูญหายของข้อมูล 10% และ 30% ผลการวิเคราะห์ความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) เมื่อพิจารณาภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) มีค่าอยู่ระหว่าง 0.0158-0.0308 หรือคิดเป็นร้อยละ 1.58%-3.08% และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) มีค่าอยู่ระหว่าง 0.0070-0.0350 หรือคิดเป็นร้อยละ 0.7%-3.50% ตามลำดับซึ่งถือว่าเป็นวิธีการตรวจสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ผ่านตามเกณฑ์ อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 ($p < 0.05$) และสามารถแสดงค่าเฉลี่ยของความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ด้วยภาพที่ 4-1



MAR	0.0200	0.0168	0.0221	0.0179	0.0200	0.0158	0.0184	0.0237	0.0242	0.0308	0.0290	0.0182	0.0247	0.0253	0.0263	0.023
MNAR	0.0121	0.0115	0.0072	0.0105	0.0105	0.0077	0.0118	0.0105	0.0176	0.0321	0.0350	0.0292	0.0266	0.0224	0.0234	0.0221

ภาพที่ 4-1 กราฟแสดงค่าเฉลี่ยของความคลาดเคลื่อนประเภทที่ 1 จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง อัตราสูญหายของข้อมูล ขนาดของ DIF และสภาวะการสูญหายของข้อมูล

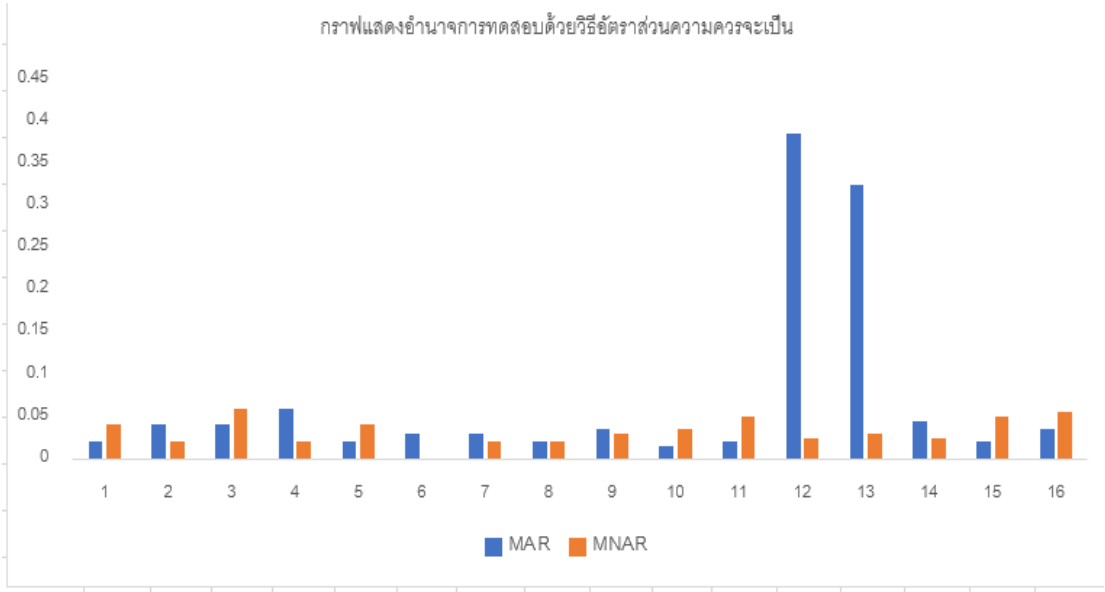
ตารางที่ 4-28 อำนาจการทดสอบ ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง อัตราสูญหายของข้อมูล ขนาดอิทธิพลของ DIF และสภาวะการสูญหายของข้อมูล

ความยาวแบบสอบ	ขนาดกลุ่มตัวอย่าง	อัตราการสูญหาย	ขนาดอิทธิพลของDIF	ค่าอำนาจการทดสอบ	
				MAR	MNAR
20	1,000	10	0.5	0.0200	0.0400
			1.0	0.0400	0.0200
		30	0.5	0.0400	0.0600
			1.0	0.0600	0.0200
	3,000	10	0.5	0.0200	0.0400
			1.0	0.0300	0.0000
		30	0.5	0.0300	0.0200
			1.0	0.0200	0.0200
40	1,000	10	0.5	0.0350	0.0300
			1.0	0.0150	0.0350
		30	0.5	0.0200	0.0500
			1.0	0.3850	0.0250
	3,000	10	0.5	0.3250	0.0300
			1.0	0.0450	0.0250
		30	0.5	0.0200	0.0500
			1.0	0.0350	0.0550

* $p < 0.05$

จากตารางที่ 4-28 ภายใต้ความยาวข้อสอบ 20 และ 40 ข้อ ในเงื่อนไขขนาดอิทธิพล DIF 0.5 และ 1.0 ขนาดตัวอย่าง 1,000 และ 3,000 ที่อัตราการสูญหายของข้อมูล 10% และ 30% ผลการวิเคราะห์อำนาจการทดสอบ ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) เมื่อพิจารณาภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) มีค่าอยู่ระหว่าง 0.0150-0.3850 หรือคิดเป็นร้อยละ 1.50%-38.50% และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) มีค่าอยู่ระหว่าง 0.0000-0.0550 หรือคิดเป็นร้อยละ 0%-5.50%

ตามลำดับที่ระดับนัยสำคัญ 0.05 ถือว่าเป็นวิธีการตรวจสอบที่มีอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำ ไม่ผ่านเกณฑ์ตามที่กำหนดทุกเงื่อนไข ซึ่งสามารถแสดงค่าเฉลี่ยของค่าอำนาจการทดสอบของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ด้วยภาพที่ 4-2



MAR	0.0200	0.0400	0.0400	0.0600	0.0200	0.0300	0.0300	0.0200	0.0350	0.0150	0.0200	0.3850	0.3250	0.0450	0.0200	0.0350
MNAR	0.0400	0.0200	0.0600	0.0200	0.0400	0.0000	0.0200	0.0200	0.0300	0.0350	0.0500	0.0250	0.0300	0.0250	0.0500	0.0550

ภาพที่ 4-2 กราฟแสดงอำนาจการทดสอบ จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีจัดการข้อมูลสูญหาย ด้วยวิธีประมาณค่าพหุ (Multiple imputation: MI) จำแนกตามความยาวของแบบสอบ ขนาดกลุ่มตัวอย่าง อัตราสูญหายของข้อมูล ขนาดของ DIF และสภาวะการสูญหายของข้อมูล

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้ มีวัตถุประสงค์ 4 ประการ **ประการแรก** เพื่อศึกษาคุณภาพของวิธีจัดการข้อมูลสูญหายในการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนกและความยาก) กับพารามิเตอร์ความสามารถของผู้สอบ (θ) **ประการที่สอง** เพื่อเปรียบเทียบประสิทธิภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ**ประการที่สาม** เพื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนสองค่าด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) และ**ประการที่สี่**เพื่อเปรียบเทียบอำนาจการทดสอบ (Power of test) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนสองค่าด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) ข้อมูลที่ใช้ในการศึกษาในครั้งนี้เป็นข้อมูลที่จำลองขึ้นภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัยหลัก คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ ขนาดอิทธิพลของ DIF จำนวน 32 เงื่อนไข (2x2x2x2x2) ในแต่ละเงื่อนไขจำลองข้อมูลวนซ้ำ 100 ครั้ง สรุป อภิปรายผล และข้อเสนอแนะดังนี้

สรุปผลการวิจัย

จากผลการวิเคราะห์ข้อมูล สามารถสรุปผลการวิจัย รายละเอียดแสดง ดังนี้

1. ผลการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบด้วยวิธีจัดการข้อมูลสูญหายโดยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพล DIF

1.1 ผลการวิเคราะห์ค่าพารามิเตอร์ข้อสอบ(อำนาจจำแนกและความยาก) ด้วยวิธีจัดการข้อมูลสูญหายแบบวิธีประมาณค่าพหุ (Multiple imputation: MI) พบว่า

1.1.1 ค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิงซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการประมาณค่าทดแทน (a_MI) โดยภาพรวมพบว่า มีค่าได้มีค่าใกล้เคียงกัน แต่ค่าที่ได้มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์

(a_Com) และเมื่อพิจารณารายข้อพบว่าที่อัตราการสูญหาย 10% ค่าที่ได้มีค่าใกล้เคียงกัน แต่ส่วนใหญ่มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com) และเมื่ออัตราการสูญหายเพิ่มขึ้นเป็น 30% ค่าที่ได้มีค่าใกล้เคียงกัน แต่ส่วนใหญ่มีค่าต่ำกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com)

1.1.2 ค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการประมาณค่าทดแทน (a_MI) โดยภาพรวมและรายข้อพบว่า มีค่าได้มีค่าใกล้เคียงกัน แต่ค่าที่ได้มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com) เมื่อพิจารณารายข้อพบว่าที่อัตราการสูญหาย 10% ค่าที่ได้มีค่าใกล้เคียงกัน ส่วนใหญ่มีค่าสูงกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com) และเมื่ออัตราการสูญหายเพิ่มขึ้นเป็น 30% ค่าที่ได้มีค่าใกล้เคียงกัน แต่มีค่าต่ำกว่าค่าอำนาจจำแนกของข้อสอบของข้อมูลที่สมบูรณ์ (a_Com)

1.1.3 ค่าความยากของข้อสอบ (b_R) ของกลุ่มอ้างอิง ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR)และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าความยากของข้อสอบที่ได้จากการประมาณค่าทดแทน (b_{R_MI}) โดยภาพรวมพบว่าส่วนใหญ่ค่าที่ได้สูงกว่าค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ (b_{R_Com}) และเมื่อพิจารณาเป็นรายข้อพบว่า ค่าที่ได้มีค่าไม่แน่นอน

1.1.4 ค่าความยากของข้อสอบ (b_F) ของกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่าความยากของข้อสอบที่ได้จากการประมาณค่าทดแทน (b_{F_MI}) โดยภาพรวมพบว่าต่ำกว่าค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ (b_{F_Com}) และเมื่อพิจารณาเป็นรายข้อพบว่าส่วนใหญ่มีค่าต่ำกว่าค่าความยากของข้อสอบของข้อมูลที่สมบูรณ์ (b_{F_Com})

1.2 ผลการวิเคราะห์ค่าพารามิเตอร์ความสามารถของผู้สอบ (θ)ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) โดยภาพรวมพบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่าพารามิเตอร์ความสามารถของผู้สอบที่ได้จากการประมาณค่าทดแทน (θ_MI) ส่วนใหญ่มีแนวโน้มใกล้เคียงค่าความสามารถของผู้สอบของข้อมูลที่สมบูรณ์ (θ_Com) มากกว่าค่าพารามิเตอร์ความสามารถของผู้สอบภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน

2. ผลการวิเคราะห์ประสิทธิภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบโดยวิธีจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (MI) ภายใต้เงื่อนไขที่แตกต่างกัน 5 ปัจจัย คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ และขนาดอิทธิพลของ DIF ด้วยการเปรียบเทียบค่า diff และ diff error

2.1 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิงซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพรวมและรายข้อพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) เมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่าค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

2.2 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มเปรียบเทียบซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพรวมและรายข้อพบว่า มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff มีค่าเข้าใกล้ 0 มากกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) และเมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่าค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

2.3 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าความยากของข้อสอบ (b_R) ของกลุ่มอ้างอิงซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI)

ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพและรายข้อพบว่า ส่วนใหญ่มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 น้อยกว่าสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) และเมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff error มีค่าสูงกว่าค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

2.4 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าความยากของข้อสอบ (b_F) ของกลุ่มเปรียบเทียบซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) โดยภาพรวมและรายข้อพบว่า ส่วนใหญ่ค่า diff มีค่าเข้าใกล้ 0 มากกว่าภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน และเมื่อพิจารณาค่า diff error พบว่า ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff error มีค่าต่ำกว่า ค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

2.5 ผลการวิเคราะห์ค่า diff และค่า diff error ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพพบว่า มีค่าใกล้เคียงกัน และที่ขนาดอิทธิพลของ DIF1.0 ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff เข้าใกล้ 0 มากกว่าภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) เมื่อพิจารณาค่า diff error พบว่าภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่าค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

3. ผลการวิเคราะห์เปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) พบว่าภายใต้ความยาวข้อสอบ 20 และ 40 ข้อ ในเงื่อนไขขนาดอิทธิพลของ DIF 0.5 และ 1.0 ขนาดตัวอย่าง 1,000 และ 3,000 ที่อัตราการสูญหายของข้อมูล 10% และ 30% ผลการวิเคราะห์ความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) เมื่อพิจารณาภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) มีค่าอยู่ระหว่าง 0.0158-0.0308 หรือคิดเป็นร้อยละ 1.58%-3.08% และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) มีค่าอยู่ระหว่าง 0.0070-0.0350 หรือคิดเป็นร้อยละ 0.7%-3.50% ตามลำดับซึ่งถือว่าเป็นวิธีการ

ตรวจสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ผ่านตามเกณฑ์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 ($p < .05$)

4. ผลการวิเคราะห์เปรียบเทียบอำนาจการทดสอบ (Power of test) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ภายใต้ความยาวข้อสอบ 20 และ 40 ข้อ ในเงื่อนไขขนาดอิทธิพลของ DIF 0.5 และ 1.0 ขนาดตัวอย่าง 1,000 และ 3,000 ที่อัตราการสูญหายของข้อมูล 10% และ 30% ผลการวิเคราะห์อำนาจ (Power) ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) เมื่อพิจารณาภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) มีค่าอยู่ระหว่าง 0.0150-0.3850 หรือคิดเป็นร้อยละ 1.50%-38.50% และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) มีค่าอยู่ระหว่าง 0.0000-0.0550 หรือคิดเป็นร้อยละ 0%-5.50% ตามลำดับที่ระดับนัยสำคัญ .05 ถือว่าเป็นวิธีการตรวจสอบที่มีอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำ ไม่ผ่านเกณฑ์ตามที่กำหนดทุกเงื่อนไข

อภิปรายผลการวิจัย

การอภิปรายผลในงานวิจัยนี้นำเสนอ 2 ประเด็นหลัก ได้แก่ ประเด็นแรก ผลการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนกและความยาก) และพารามิเตอร์ความสามารถของผู้สอบ (θ) โดยทำการประมาณค่าพารามิเตอร์ด้วยวิธีแม็กซิมั่มไลค์ลิฮู้ด (Maximum Likelihood: ML)และประเด็นที่สอง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT) ภายใต้ปัจจัยที่แปรเปลี่ยน 5 ปัจจัยหลัก คือ 5 ปัจจัยหลัก คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ ขนาดอิทธิพลของ DIF จากผลการวิจัยเบื้องต้นมีประเด็นสำคัญที่นำมาอภิปรายผล ดังนี้

ประเด็นแรก การประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนกและความยาก) และพารามิเตอร์ความสามารถของผู้สอบ (θ)

1. ผลการประมาณค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนกและความยาก) และพารามิเตอร์ความสามารถของผู้สอบ (θ) ที่ได้จากการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ค่าที่ได้มีค่าใกล้เคียงกัน ภายใต้เงื่อนไขอัตราการสูญหาย 5-30% เพราะวิธีการจัดการข้อมูลสูญหายด้วยวิธีดังกล่าวได้รับการพัฒนาขึ้นมาเพื่อแก้ไขจุดอ่อนที่เกิดจากการจัดการข้อมูลสูญหายด้วยวิธีดั้งเดิม

ผลดังกล่าวสอดคล้องกับผลการศึกษาของ กมลทิพย์ ศรีหาเศษ (2555) พบว่า ภายใต้สภาวะการสญูหายของข้อมูลแต่ละประเภท ค่าพารามิเตอร์ข้อสอบ (อำนาจจำแนกและความยาก) และพารามิเตอร์ความสามารถของผู้สอบ (θ) มีค่าใกล้เคียงกัน ผลการศึกษาสะท้อนให้เห็นว่า เมื่ออัตราการสูญหายอยู่ในช่วง 5-20% วิธีจัดการข้อมูลสูญหายที่ใช้ในการจัดการข้อมูลสูญหายในงานวิจัยทั้ง 3 วิธี ได้แก่ วิธีประมาณค่าทดแทนพหุ (Multiple imputation: MI) วิธีค่าคาดหมายสูงสุด (Expectation-maximization algorithm: EM) และวิธีการประมาณค่าที่เป็นไปได้สูงสุด (Maximum likelihood estimation: ML) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) วิธีการดังกล่าวสามารถจัดการข้อมูลสูญหายได้ใกล้เคียงกัน เนื่องจากวิธีการจัดการข้อมูลสูญหายทั้ง 3 วิธี เป็นวิธีที่พัฒนาขึ้นเพื่อแก้ไขจุดอ่อนของวิธีการจัดการข้อมูลสูญหายแบบดั้งเดิม (Acock, 2005; Finch, 2008)

2. ผลการเปรียบเทียบค่าความเอนเอียง (diff และ diff error) ของค่าอำนาจจำแนกของข้อสอบ (a) ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน และที่ความยาวแบบสอบ 20 ข้อ และ 40 ข้อ ค่าความเอนเอียงโดยภาพรวมและรายข้อมีค่าเข้าใกล้0 มากขึ้นเมื่อความยาวแบบสอบเพิ่มมากขึ้น

ผลดังกล่าวสอดคล้องกับผลการศึกษาของ กมลทิพย์ ศรีหาเศษ (2555) พบว่า ค่า BIAS ของค่าอำนาจจำแนกของข้อสอบ (a) และภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)ภายใต้เงื่อนไขในระดับเดียวกัน ค่า BIAS มีค่าเข้าใกล้ 0 มากขึ้นเมื่อความยาวแบบสอบเพิ่มจาก 20 ข้อ เป็น 40 ข้อ

3. ผลการเปรียบเทียบค่าความเอนเอียง (diff และ diff error) ของค่าความยากของข้อสอบของกลุ่มอ้างอิง (b_R) และกลุ่มเปรียบเทียบ (b_F) ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพรวมและรายข้อส่วนใหญ่มีค่าเข้าใกล้ 0 เมื่อความยาวแบบสอบเพิ่มมากขึ้น

ผลดังกล่าวสอดคล้องกับผลการศึกษาของ กมลทิพย์ ศรีหาเศษ (2555) พบว่า ค่า BIAS ของค่าความยากของข้อสอบ (b) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) และภายใต้เงื่อนไขในระดับเดียวกัน ค่า BIAS มีค่าเข้าใกล้ 0 มากขึ้นเมื่อความยาวแบบสอบเพิ่มขึ้น

4. ผลการเปรียบเทียบค่าความเอนเอียง (diff และ diff error) ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ซึ่งประมาณค่าได้จากวิธีการจัดการข้อมูลสูญหายด้วยวิธีประมาณค่าพหุ (MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ตามเงื่อนไขในระดับเดียวกัน ค่า diff โดยภาพรวมมีค่าใกล้เคียงกัน และที่ขนาดอิทธิพลของ DIF 1.0 ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ค่า diff เข้าใกล้ 0 มากกว่าภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) เมื่อพิจารณาค่า diff error พบว่าภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) ส่วนใหญ่ค่า diff error มีค่าต่ำกว่าค่า diff error ภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

ผลดังกล่าวสอดคล้องกับผลการศึกษาของ กมลทิพย์ ศรีหาเศษ (2555) พบว่า ค่า RMSE ของค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และสภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) และภายใต้เงื่อนไขในระดับเดียวกัน มีค่าใกล้เคียงกัน

ประเด็นที่สองการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

1. อัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบสองค่า ด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT) ภายใต้ปัจจัยที่แปรเปลี่ยน 5 ปัจจัย

1.1 ปัจจัยขนาดกลุ่มตัวอย่าง

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้ขนาดตัวอย่าง 1,000 และ 3,000 พบว่าผลการวิเคราะห์ความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) เมื่อพิจารณาภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) มีและภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ถือว่าเป็นวิธีการตรวจสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ผ่านตามเกณฑ์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 ($p < .05$)

ผลดังกล่าวสอดคล้องกับงานวิจัยของ Finch (2011, pp. 281-301) พบว่าปัจจัยขนาดตัวอย่างที่ต่างกัน ระดับผลกระทบ วิธีจัดการข้อมูลสูญหาย และสภาวะการข้อมูลสูญหายที่ต่างกัน พบว่าเมื่ออัตราการสูญหายที่สูงขึ้น วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธีสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ไม่แตกต่างกัน โดยเฉพาะภายใต้สภาวะการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) และภายใต้สภาวะการข้อมูลสูญหายอย่างสุ่ม (MAR) และภายใต้ข้อมูลที่มีการละเว้นการตอบแบบตอบผิด (Omitted as incorrect) ค่าอัตราความ

คลาดเคลื่อนประเภทที่ 1 มีค่าสูงขึ้น อยู่ในช่วง 0.14-0.30 และยังมีค่าเพิ่มขึ้นเมื่อขนาดตัวอย่างและระดับขนาดอิทธิพลเพิ่มมากขึ้น

1.2 ปัจจัยอัตราการสูญหาย

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้อัตราการสูญหายที่แตกต่างกัน พบว่าผลการวิเคราะห์ความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) เมื่อพิจารณาภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) มีและภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ถือว่าเป็นวิธีการตรวจสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ผ่านตามเกณฑ์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 ($p < .05$)

ผลดังกล่าวสอดคล้องกับงานวิจัยของ Finch (2011, pp. 281-301) พบว่าภายใต้ข้อมูลที่มีการสูญหายจากการละเว้นการตอบแบบตอบผิด (Omitted as incorrect) วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธี สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ไม่แตกต่างกันและเป็นไปตามตามเกณฑ์ปกติ ($p \leq .05$) โดยเฉพาะที่อัตราความคลาดเคลื่อน 15%

1.3 ปัจจัยความยาวแบบแบบสอบ

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ความยาวแบบสอบที่แตกต่างกัน พบว่าการวิเคราะห์ความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) เมื่อพิจารณาภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ถือว่าเป็นวิธีการตรวจสอบที่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ผ่านตามเกณฑ์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05 ($p < .05$) เมื่อความยาวแบบสอบเพิ่มขึ้น วิธีการตรวจสอบความควรจะเป็น (LRT) สามารถควบคุมความคลาดเคลื่อนประเภทที่ได้**ไม่แตกต่างกัน**กับภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR)

2. อำนาจการทดสอบ (Power of test) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ให้คะแนนแบบสองค่า ด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT) ภายใต้ปัจจัยที่แปรเปลี่ยน 5 ปัจจัย

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ภายใต้ความยาวข้อสอบ 20 และ 40 ข้อ ในเงื่อนไขขนาดอิทธิพลของDIF 0.5 และ 1.0 ขนาดตัวอย่าง 1,000 และ 3,000 ที่อัตราการสูญหายของข้อมูล 10% และ 30% ผลการวิเคราะห์อำนาจการทดสอบ (Power of

test)ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (LRT) เมื่อพิจารณาภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ที่ระดับนัยสำคัญ .05 อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำ ไม่ผ่านเกณฑ์ตามที่กำหนดทุกเงื่อนไข โดยผลการวิจัยดังกล่าวได้ทำการจัดการทำกับข้อมูลสูญหายที่อัตราการสูญหาย 10% และ 30% เฉพาะข้อสอบข้อที่ 1-4 ที่กลุ่มตัวอย่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบขนาด 500: 500 และ 1,000: 1,000 และกำหนดให้ข้อ 1 และ 2 เป็นข้อสอบที่ทำหน้าที่ต่างกัน เมื่อทำการประมาณค่าข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ข้อมูลที่ได้ทำให้สารสนเทศของข้อมูลขาดหายไป เมื่อทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT) ค่าอำนาจการทดสอบ (Power of test) จึงมีค่าต่ำ (0.0000-0.3850) ไม่ผ่านเกณฑ์ตามที่กำหนดทุกเงื่อนไข **ซึ่งไม่สอดคล้องกับงานวิจัยของ** Finch (2011, pp. 281-301) เพราะผลการวิจัยของ Finch ที่ความยาวแบบสอบ 40 ข้อ อัตราการสูญหาย 5% และ 10% มีการจัดกระทำกับข้อมูลสูญหายกระจายภายใต้กลุ่มตัวอย่าง ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ขนาด 250: 250 500: 500 และ 1,000: 1,000 เมื่อทำการประมาณค่าทดแทนข้อมูลสูญหายด้วยวิธีประมาณค่าแบบพหุ (Multiple imputation: MI) และทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี 1) Mantel-haenszel 2) Logistic regression 3) SIBTEST ภายใต้สภาวะการสูญหายโดยสมบูรณ์อย่างสุ่ม (MCAR) การสูญหายอย่างสุ่ม (MAR) และการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) มีค่า 0.47-0.98 มีอำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. การศึกษาครั้งนี้พบว่าวิธีการจัดการข้อมูลสูญหายด้วยวิธีการประมาณค่าทดแทนพหุ (Multiple imputation: MI) ภายใต้สภาวะการสูญหายอย่างสุ่ม (MAR) และภายใต้สภาวะการสูญหายที่ไม่ใช่อย่างสุ่ม (MNAR) ค่าที่ได้มีความถูกต้องใกล้เคียงกัน ภายใต้อัตราการสูญหาย 10-30 %กับกลุ่มตัวอย่างที่มีขนาดใหญ่

2. ผลการศึกษาครั้งนี้พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบสองค่า ด้วยวิธีทดสอบอัตราส่วนความควรจะเป็น (Likelihood ratio test: LRT) ภายใต้ปัจจัยที่แปรเปลี่ยน 5 ปัจจัยหลัก คือ 5 ปัจจัยหลัก คือ สภาวะการสูญหายของข้อมูล ขนาดกลุ่มตัวอย่าง

อัตราการสูญหายของข้อมูล ความยาวของแบบสอบ ขนาดอิทธิพลของ DIF ซึ่งมีการกำหนดให้จำนวนข้อที่มีการสูญหายคือ ข้อ 1-ข้อ 4 ทำให้เมื่อมีการสูญหายของข้อมูลภายใต้ อัตราการสูญหาย 10-30% ทำให้ผลที่ได้จากการประมาณค่าทดแทนข้อมูลสูญหายไม่ถูกต้อง ค่าอำนาจการทดสอบ (Power of test) ค่าต่ำ ดังนั้นจึงไม่ควรมีการกำหนดจำนวนข้อให้มีการสูญหาย ควรจะกำหนดให้มีการสูญหายกระจายในข้อใด ๆ ภายใต้ความยาวข้อสอบ 20 ข้อ หรือ 40 ข้อ

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. การศึกษาครั้งนี้เป็นการศึกษาภายใต้วิธีจัดการข้อมูลสูญหาย 1 วิธีคือวิธีการประมาณค่าทดแทนพหุ ควรมีการศึกษาเปรียบเทียบผลกับวิธีจัดการข้อมูลสูญหายด้วยวิธีอื่น ๆ ต่อไป

2. การศึกษาครั้งนี้เป็นการศึกษาภายใต้ภายใต้ความเป็นเอกมิติ ในการศึกษาครั้งถัดไปควรมีการศึกษาเปรียบเทียบผลการศึกษาภายใต้ความเป็นพหุมิติ และการตรวจให้คะแนนแบบหลายค่าต่อไป

3. ควรมีการศึกษาเปรียบเทียบกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่หลากหลายวิธีเพื่อพิจารณาว่าภายใต้การจัดการข้อมูลสุญหายด้วยวิธีประมาณค่าทดแทนพหุวิธีใดสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้เหมาะสม

บรรณานุกรม

กมลทิพย์ ศรีหาเศษ. (2555). *การเปรียบเทียบคุณภาพของวิธีจัดการข้อมูลสูญหายสำหรับประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

จำลอง วงษ์ประเสริฐ. (2554). *การพัฒนาวิธีการประมาณข้อมูลสูญหายโดยการถ่วงน้ำหนักแบบวนซ้ำด้วยวิธีของแจ็คไนฟ์และการวิเคราะห์การถดถอย (ไอดับบลิวเจอาร์).* วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัยและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยมหาสารคาม.

ชนะศึก นิชานนท์. (2553). *ประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบเบส์โดยใช้การสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

ชยุตม์ ภิรมย์สมบัติ. (2547). *คุณสมบัติของตัวประมาณค่าความเข้มของอิทธิพล: การเปรียบเทียบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, คณะครุศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย.

เชาว์ อินใย. (2547). *การพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีและการตรวจสอบความแม่นยำและอำนาจการทดสอบเปรียบเทียบกับวิธีอีเอ็มและลิสท์ไวส์: เทคนิคมอนติ คาร์โล.* วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัยและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยนเรศวร.

ณรงค์ โพธิ และสมชาย ปราการเจริญ. (2553). ศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายโดยวิธีการวิเคราะห์จำแนกประเภทต้นไม้การตัดสินใจและค่าเฉลี่ย. *วารสารมหาวิทยาลัยราชภัฏสกลนคร, 2*(3), 146-156.

นงลักษณ์ วิรัชชัย. (2552). *วิจัยและสถิติ: คำถามชวนตอบ.* กรุงเทพฯ: ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ปิยะภรณ์ ประสิทธิ์วัฒนเสรี และสุคนธ์ ประสิทธิ์วัฒนเสรี. (2552). ข้อมูลสูญหายและแนวทางการจัดการ. *วารสารการจัดการข้อมูลและชีวสถิติ*, *4*(3), 52-61.

ไพฑูรย์ มุลิวัลย์. (2556). *การพัฒนาวิธีการประมาณค่าข้อมูลสูญหาย.* วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สาขาวิชาวิทยาการจัดการสถิติ, บัณฑิตวิทยาลัย, มหาวิทยาลัยมหาสารคาม.

ศิริชัย กาญจนวาสี.(2546). *ทฤษฎีการประเมิน.* กรุงเทพฯ: เท็กซ์ แอนด์เจอร์นัลพับลิเคชั่น.

ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 3). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สุนทร เขียวงาม. (2551). *ผลของความไม่เป็นอิสระของข้อสอบที่มีต่อการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบเมื่อมีเงื่อนไขการทดสอบที่แตกต่างกัน.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

สุพัฒนา หอมบุปผา. (2556). *การเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี HGLM วิธี MIMIC และวิธี BAYESIAN.* ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.

อรวรรณ กีรติสิโรจน์. (2558). การจัดการข้อมูลสูญหาย: Multiple imputation. *วารสารส่งเสริมสุขภาพและอนามัยสิ่งแวดล้อม, 38*(2), 21-24.

อรินทร์ น่วมถนอม. (2549). *การเปรียบเทียบวิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า.* ดุษฎีนิพนธ์การศึกษาดุษฎีบัณฑิต, สาขาวิชาการทดสอบและวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยศรีนครินทรวิโรฒ.

อริสฟ้า เตหลิ่ม. (2559). *การเปรียบเทียบประสิทธิผลการประมาณค่าพารามิเตอร์และการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีแมกซิมัมไลค์ลิฮู้ด วิธีของเบส์และวิธีของเบส์แบบมีอิทธิพลเทสต์เลท.* ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.

อาวีพร ปานทอง. (2558). *การเปรียบเทียบประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบให้คะแนนหลายค่าโดย วิธีทดสอบอัตราส่วนความควรจะเป็น วิธีเบส์เซียน และวิธีโพลี-ซิปเทสท์.* ดุษฎีนิพนธ์ปรัชญาดุษฎีบัณฑิต, สาขาวิชาวิจัย วัดผลและสถิติการศึกษา, คณะศึกษาศาสตร์, มหาวิทยาลัยบูรพา.

อิทธิฤทธิ์ พงษ์ปิยะรัตน์. (2551). *การวิเคราะห์ข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ: การวิเคราะห์พหุระดับ*. วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.

อรวรรณ กีรติสิโรจน์. (2558). การจัดการข้อมูลสูญหาย: Multiple imputation.*วารสารส่งเสริมสุขภาพและอนามัยสิ่งแวดล้อม, 38*(2), 21-24.

Ake, C. F. (2005). *Rounding after multiple imputation with non-binary categorical covariates*. Retrieved from http://www2.sas.com/proceedings/sugi30/112-30.pdf

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: Sage.

Allison, P. D. (2006). *Imputation of categorical variables with PROC MI*. Retrieved from http://www2.sas.com/proceedings/sugi30/113-30.pdf

Andreis, F., & Ferrari, P. A. (2012). Missing data and parameters estimates in multidimensional item response models. *Electronic Journal of Applied Statistical Analysis*, *5*(3), 431-437.

Andrich, D. (1988). *Rasch models for measurement*. Beverly Hills: Sage.

Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In W. H. Paul, & W. Howard (Eds.), *Differential item functioning* (pp. 3-23). New Jersey: America Lawrence Erlbaum Association.

Atar, B., & Kamata, A. (2011). Comparison of IRT likelihood ratio test and logistic regression DIF detection procedures. *Hacettepe University Journal of Educationi*, *41,* 36-47.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5-37.

Baker, F. B. (1992). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.

Baker, F. B. (2001). *The basics of item response theory*. Retrieved from for full text: http://ericae. net/irt/baker.

Bernaards, C. A., & Sijtsma, K. (1999). Factor analysis of multidimensional polytomous item response data from Ignorable item nonresponse. *Multivariate Behavioral Research, 34*, 277-314.

Bernaards, C. A., & Sijtsma, K. (2000). Influence of imputation and EM methods on factor analysis when Item nonresponse in questionnaire data is nonignorable. *Multivariate Behavioral Research, 35*, 321-364.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-472). Reading, MA: Addison-Wesle.

Bodner, T. E. (2008). What improves with increased missing data imputations?. *Structural Equation Modeling, 15*, 651-675.

Bond, T. G., & Fox, C. M. (2001). *Applying the rasch model: Fundamental measurement in the human sciences.* New Jersey: Lawrence Erlbaum Associates.

Cai, L., Du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software].* Chicago, IL: Scientific Software International.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: SAGE Publications.

Chang, Y. C. I. (2012). *Sequential estimation in item calibration with a two-stage design.* Retrieved from http://arxiv.org/ftp/arxiv/papers/1206/1206.4189.pdf

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we?. *Psychological Bulletin*, *74*(1), 68.

De Ayala, R. J., Plake, B. S., & Impara, J. C. (2001). The impact of omitted responses on the accuracy of ability estimation in item response theory. *Journal of Educational Measurement, 38,* 213-234.

DeMars, C. (2000). Test stakes and item format interactions. *Applied Measurement in Education,13*, 55-77.

DeSilvio, M. L. (1999). *A variance ratio statistics for assessing the missing data mechanism: An empirical study*. Doctoral dissertation, Tulane University.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the scholastic aptitude test. *Journal of Educational Measurement, 23*(4), 355-368.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Enders, C. K. (2004). The impact of missing data on sample reliability estimates: Implications for reliability reporting practices. *Educational and Psychological Measurement*, *64*, 419-436.

Enders, C. K. (2001). The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, 61*, 713-740.

Eppen, G. D., Gould, F. J., & Schmidt, C. P. (1985). *Introductory management science*. London: Prentice-Hall.

Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *Journal of Educational Measurement*, *45*, 225-245.

Finch, H. (2011). The use of multiple imputation for missing data in uniform DIF analysis: Power and type I error rates. *Applied Measurement in Education*, *24*(4), 281-301.

Frenkel, D. (2004). Introduction to monte carlo methods. *John von Neumann Institute for Computing, 23*, 29-60.

Furlow et al. (2007). A monte carlo study of the impact of missing data and differential item functioning on theta estimates from two polytomous rasch family models. *Journal of Applied Measurement, 8*(4), 388-403.

Gao, F., & Chen, L. (2005). Bayesian or non-bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education, 18*, 351-380.

Garson, G. D. (2008). *Data imputation for missing values*. Retrieved from http://faculty.chass.ncsu.edu/garson/PA765/missing.htm

Ghuman, S. J., Lee, H. J., & Smith, H. L. (2006). Measurement of women's autonomy according to women and their husbands: Results from five Asian countries. *Social Science Research*, *35*(1), 1-28.

Gibson, N. M., & Olejnik, S. (2003). Treatment of missing data at the second level of hierarchical linear models. *Educational and Psychological Measurement*, *63*, 204-238.

Glas, C. A. W., & Pimentel, J. L. (2008). Modeling nonignorable missing data in speed tests. *Educational and Psychological Measurement, 68*, 907-922.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in sducation and psychology* (3rd ed.). Boston: Allyn and Bacon.

Graham, J. W. (2009). Missing data analysis making it work in the real world. *Annual Review of Psychology, 60,* 549-576.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). United states America: Sage.

Hanson, B. A., & Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement, 26*, 3-2.

Heeringa, S. G. (2000). Multivariate imputation of coarsened survey data on household wealth (Doctoral dissertation). University of Michigan, Horace H. Rackham School of Graduate Studies.

Holmes, F. W., & French, B. F. (2007). Detection of crossing differential item functioning: A comparison of four methods. *Educational and Psychological Measurement*, *67*(4), 565-582.

Howell, D. C. (2009). *Treatment of missing data*. Retrieved from http://www. uvm.edu/~dhowell/StatPages/More Stuff/Missing Data/Missing.html

Honaker, J., King, G., & Blackwel, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software, 45*, 1-47.

Huisman, M., & Molenaar, I. W. (2001). Imputation of missing scale data with Item response models. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 221-244). New York: Springer.

Kim, J., & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods & Analysis, 6*, 215-240.

Kim, S. H., Chosen, A. S., & Kim, S. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement Summer, 44*(2), 93-116.

Kim, S. (2006). A comparative study of IRT fixed parameter calibration methods. *Journal of Educational Measurement, 43*, 355-381.

Kwang, C. (2000). *Variance Estimation After Imputation*. Dissertation Abstracts International. Iowa State University.

Leite, W. L., & Beretvas, S. N. (2004). *The performance of multiple imputation for likert-type items with missing data*. Retrieved from http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=1360&context=jmasm

Little, R., & Rubin, D. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, NJ: John Wiley and Sons.

Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika, 39*, 247-264.

Lord, F. M. (1980). *Applications of Item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M. (1983). Maximum likelihood estimation of item response parameters when some responses are omitted. *Psychometrika, 48,* 477-482.

Ludlow, L. H., & O'Leary, M. (1999). Scoring omitted and not-reached items: practical data analysis implications. *Educational and Psychological Measurement, 59,* 615-630.

MacIntosh, R. (1998). Globe attitude measurement: An assessment of the world values survey postmaterialism scale. *American Sociological Review, 63*, 452-464.

Madlow, W. G., Nisselson, H., & Olkin, I. (Eds.). (1983). *Incomplete data in sample surveys, volume 1: Report and case studies*. New York: Academic Press.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (1994). *Introduction to probability and statistics*. Belmont, California: Wadsworth

Marie et al. (2007). Measuring and detecting differential item functioning in criterion referenced licensing test. *A Theoretic Comparison of Methods, Umea°*, 7-8.

Mislevy, R. J., & Wu, P. K. (1988). Inferring examinee ability when some item responses are missing. *ETS Research Report Series, 2*, i-75.

Mislevy, R. J., & Wu, P. K. (1996). *Missing responses and IRT ability estimation: Omits, choice, time limits, and adaptive testing*. Princeton, NJ: Educational Testing Service.

Montgomery, M., & Skorupski, W. (2012). *Investigation of IRT parameter recovery and classification accuracy in mixed format*. Retrieved from www.cete.us%2Fresearch%2Fpresentations%2Fpdfs%2F2012_04_Montgomery%2520IRT%2520Classification_NCME.pdf&ei=d721UPT8KIrPrQf1tIGgBQ&usg=AFQj CNFcz4pLMEj7oV0lrzHydNerqtsgww

Muthen, L. K., & Muthen, B. O. (2007). *Mplus user's guide* (5th ed.). Los Angeles, CA

Narayanun, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, *20*(3), 257-274.

Nitko, J. A. (2008). What measurement research is needed in the future. *National Council on Measurement in Education Newsletter, 16*(3), 4-7.

Oba, S., Sato, M. A., Takemasa, I., Monden, M., Matsubara, K. I., & Ishii, S. (2003). A bayesian missing value estimation method for gene expression profile data. *Bioinformatics*, *19*(16), 2088-2096.

Partchev, I. (2004). *A visual guide to item response theory*. Retrieved from http://www.metheval.uni-jena.de/irt/VisualIRT.pdf

Penfield, R. D. (2005). DIFAS: Differential item functioning analysis system. *Applied Psychological Measurement*, *29*(2), 150-151.

Peng, C. J., & Zhu, J. (2008). Comparison of two approaches for handling missing covariates in logistic regression. *Educational and Psychological Measurement, 68*, 58-77.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74,* 525-556.

Point Estimation. (n.d.). *Method for evaluating estimators.* Retrieved from http://www4.stat.ncsu.edu/~wang/courses/ST522/handouts/Ch7-part2-by2.pdf

Raaijmakers, Q. A. W. (1999). Effectiveness of different missing data treatments in surveys with likert-type data: Introducing the relative mean substitution approach. *Educational and Psychological Measurement, 59*, 725-748.

Reeve, B. B. (2003). *An introduction to modern measurement theory*. Retrieved from http://appliedresearch.cancer.gov/areas/cognitive/immt.pdf

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 45,* 1-25.

Robitszsch, A., & Rupp, A. A. (2009). Impact of missing data on the detection of differential item functioning the case of mantel-haenszel and logistic regression analysis. *Educational and Psychological Measurement*, *69*(1), 18-34.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*(3), 537-560.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6*, 317-329.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545-571.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman and Hall/ CRC.

Schmitt, P., Mandel, J., & Guedj, M. (2015). A comparison of six methods for missing data imputation. *Journal of Biometrics & Biostatistics*, *6*(1), 1-6.

Sedivy, J. M. (2009). How to learn new and interesting things from model systems based on "exotic" biological species. *Proceedings of the National Academy of Sciences*, *106*(46), 19207-19208.

Sijtsma, K., & van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research, 38*, 505-528.

Sinharay, S., Stern, H. S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6*, 317-329.

Smith, H. L., & Furstenberg, F. F. (1994). Application of a response model for mother-daughter agreement by race. *Social Science Research, 23*, 136-166.

Smith, H. L., & Morgan, S. P. (1994). Children's closeness to father as reported by mothers, sons and daughters: Evaluating subjective assessments with the rasch model. *Journal of Family Issues, 15*, 3-29.

Songfeng, Z. (n.d.). *Statistical theory II: Method of evaluating estimators*. Retrieved from http://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/evaluation.pdf

Statistics, S. (2009). *Missing values*. Retrieved from http://www.statisticssolutions.com/missing-values

Stroud, R. S. (1996). *Athenian economy and society: A banking perspective*. Princeton: University Press.

Swaminathan, H., & Gifford, J. A. (1982). Bayesian eatimation in the rash model. *Journal of Education Statistics, 7*, 175-192.

Tfaily, R. (2006). *Differential item functioning using item response theory: An application to attitudes towards family dissolutuon*. Retrieved from http://www.allacademic.com/meta/p_mla_apa_researchcitation/1/0/5/0/0/p105007_index.html

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen, & H. Wainer (Eds.), *Test scoring* (pp. 73-140). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thissen, D., & Steinberg, L. (1988). Data analysis using item response theory. *Psychological Bulletin, 104*, 385-395.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, *17*(6), 520-525.

Van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine 18,* 681-694.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45,* 1-67.

Velicer, W. F., & Colby, S. M. (2005). A comparison of missing-data procedures for arima time-series analysis. *Educational and Psychological Measurement*, *65*, 596-615.

Weiss, D. J., & Minden, S. V. (2012). *A comparison of item parameter estimates from Xcalibre 4.1 and Bilog-MG*. St. Paul MN: Assessment Systems Corporation.

Witta, E. L. (2000). Effectiveness of four methods of handling missing data using samples from a National Database.

University of Texas, Information Technology Services (ITS). (2004). *Handling missing or incomplete data*. Retrieved from http://www.utexas.edu/its/rc/answers/general/gen25.html

Wood et al. (2004). Are missing outcome data adequately handled?. A review of publish randomized controlled trials in major medical. *Journals Clinical Trial, 1*(4), 368-376.

Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika, 52*, 275-291.

Zhang, B., & Walker, C. M. (2008). Impact of missing data on person model fit and person trait estimation. *Applied Psychological Measurement*, *32*(8), 466-479.

Zhou, X., Wang, X., & Dougherty, E. R. (2003). Missing-value estimation using linear and non-linear regression with bayesian gene selection. *Bioinformatics*, *19*(17), 2302-2307.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *Biologmg3.* Lincolnwood, IL: Scientific Software International.

ภาคผนวก

ผลการประมาณค่าพารามิเตอร์ข้อสอบโดยใช้โปรแกรม IRTPRO

ภายใต้เงื่อนไข สภาวะการสูญหาย MNAR อัตราการสูญหาย 10% ความยาวแบบสอบ 20 ข้อ ขนาดตัวอย่าง 3,000 คน และขนาดอิทธิพลของ DIF = 0.5 ของกลุ่มอ้างอิง

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	X1	2	0.99	0.10	1	-0.33	0.08	0.33	0.08
2	X2	4	0.61	0.08	3	-0.07	0.07	0.12	0.11
3	X3	6	0.85	0.10	5	0.11	0.07	-0.13	0.08
4	X4	8	0.55	0.08	7	0.04	0.07	-0.07	0.12
5	X5	10	1.63	0.14	9	0.04	0.09	-0.02	0.06
6	X6	12	0.57	0.08	11	-0.14	0.07	0.24	0.12
7	X7	14	1.32	0.12	13	0.04	0.08	-0.03	0.06
8	X8	16	1.02	0.10	15	-0.04	0.08	0.04	0.07
9	X9	18	0.57	0.08	17	0.03	0.07	-0.06	0.12
10	X10	20	1.50	0.13	19	-0.11	0.09	0.08	0.06
11	X11	22	0.60	0.09	21	-0.03	0.07	0.05	0.11
12	X12	24	0.51	0.08	23	0.00	0.07	-0.00	0.13
13	X13	26	0.75	0.09	25	-0.04	0.07	0.06	0.09
14	X14	28	0.69	0.09	27	-0.06	0.07	0.08	0.10
15	X15	30	0.53	0.08	29	0.06	0.07	-0.11	0.12
16	X16	32	1.08	0.12	31	0.03	0.08	-0.03	0.07
17	X17	34	1.71	0.15	33	0.16	0.09	-0.09	0.05
18	X18	36	0.87	0.10	35	-0.02	0.07	0.02	0.08
19	X19	38	0.93	0.10	37	-0.02	0.07	0.02	0.08
20	X20	40	0.65	0.08	39	0.04	0.07	-0.06	0.10

ผลการประมาณค่าพารามิเตอร์ข้อสอบโดยใช้โปรแกรม IRTPRO

ภายใต้เงื่อนไข สภาวะการสูญหาย MNAR อัตราการสูญหาย 10% ความยาวแบบสอบ 20 ข้อ ขนาดตัวอย่าง 3,000 คน และขนาดอิทธิพลของ DIF = 0.5 ของกลุ่มเปรียบเทียบ

2PL Model Item Parameter Estimates for Group 2, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	X1	42	0.53	0.08	41	-0.25	0.05	0.47	0.11
2	X2	44	0.52	0.06	43	-0.05	0.05	0.11	0.09
3	X3	46	1.62	0.14	45	0.05	0.07	-0.03	0.04
4	X4	48	0.36	0.05	47	-0.03	0.05	0.09	0.13
5	X5	50	1.32	0.10	49	-0.02	0.06	0.02	0.05
6	X6	52	0.32	0.07	51	0.03	0.05	-0.08	0.14
7	X7	54	0.99	0.10	53	0.01	0.05	-0.01	0.06
8	X8	56	1.43	0.09	55	-0.02	0.06	0.01	0.04
9	X9	58	0.81	0.07	57	-0.02	0.05	0.03	0.06
10	X10	60	1.04	0.10	59	0.04	0.06	-0.03	0.05
11	X11	62	0.02	0.08	61	-0.06	0.05	2.93	12.80
12	X12	64	0.17	0.07	63	0.03	0.05	-0.17	0.28
13	X13	66	0.13	0.08	65	-0.01	0.05	0.09	0.35
14	X14	68	0.59	0.07	67	-0.00	0.05	0.00	0.08
15	X15	70	0.41	0.07	69	0.04	0.05	-0.09	0.12
16	X16	72	2.36	0.34	71	0.03	0.08	-0.01	0.03
17	X17	74	1.19	0.11	73	-0.02	0.06	0.01	0.05
18	X18	76	2.30	0.34	75	0.05	0.08	-0.02	0.04
19	X19	78	0.58	0.08	77	0.02	0.05	-0.03	0.09
20	X20	80	0.68	0.06	79	0.07	0.05	-0.10	0.07

ผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) **โดยใช้โปรแกรม** IRTPRO

ภายใต้เงื่อนไข สภาวะการสูญหาย MNAR อัตราการสูญหาย 10% ความยาวแบบสอบ 20 ข้อ ขนาดตัวอย่าง 3,000 คน และขนาดอิทธิพลของ DIF = 0.5 ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

โดยการนำเสนอผลลัพธ์จะนำเสนอเพียงบางส่วน เนื่องจากจำนวน n = 3,000 มีขนาดใหญ่จึงทำการนำเสนอข้อมูลตั้งแต่คนที่ 1-95 ซึ่งเป็นผู้สอบใน GROUP 1 และจะนำเสนอผลลัพธ์ผู้สอบใน GROUP 2 คนที่ 1-65

Group	**คนที่**	theta (θ)	s.e.
1	1	1.238	0.534
1	2	0.403	0.472
1	3	0.016	0.464
1	4	1.471	0.557
1	5	1.017	0.513
1	6	-0.571	0.481
1	7	1.471	0.557
1	8	-0.571	0.481
1	9	-0.177	0.466
1	10	-0.372	0.472
1	11	-0.372	0.472
1	12	0.016	0.464
1	13	0.403	0.472
1	14	-0.571	0.481
1	15	1.017	0.513
1	16	-0.571	0.481
1	17	-0.372	0.472
1	18	0.601	0.483
1	19	0.601	0.483

Group	คนที่	theta (θ)	s.e.
1	20	-0.177	0.466
1	21	0.016	0.464
1	22	-0.988	0.512
1	23	0.601	0.483
1	24	-0.775	0.495
1	25	1.238	0.534
1	26	-0.571	0.481
1	27	1.238	0.534
1	28	1.238	0.534
1	29	0.403	0.472
1	30	0.403	0.472
1	31	-1.209	0.532
1	32	-0.372	0.472
1	33	-0.571	0.481
1	34	-0.988	0.512
1	35	-0.988	0.512
1	36	-0.988	0.512
1	37	-0.372	0.472
1	38	-0.571	0.481
1	39	-0.372	0.472
1	40	-0.177	0.466
1	41	-0.571	0.481
1	42	0.208	0.466
1	43	0.016	0.464
1	44	-0.988	0.512
1	45	0.208	0.466
1	46	0.601	0.483

Group	คนที่	theta (θ)	s.e.
1	47	-0.988	0.512
1	48	0.601	0.483
1	49	-0.988	0.512
1	50	0.403	0.472
1	51	0.016	0.464
1	52	0.403	0.472
1	53	-1.442	0.556
1	54	0.208	0.466
1	55	0.403	0.472
1	56	-1.209	0.532
1	57	-1.688	0.582
1	58	0.016	0.464
1	59	-1.442	0.556
1	60	-0.177	0.466
1	61	0.403	0.472
1	62	0.016	0.464
1	63	0.403	0.472
1	64	1.238	0.534
1	65	0.805	0.496
1	66	-0.571	0.481
1	67	-0.775	0.495
1	68	0.403	0.472
1	69	0.403	0.472
1	70	0.403	0.472
1	71	-0.372	0.472
1	72	-0.372	0.472
1	73	2.253	0.639

Group	คนที่	theta (θ)	s.e.
1	74	-0.372	0.472
1	75	-0.988	0.512
1	76	-1.688	0.582
1	77	0.403	0.472
1	78	-0.177	0.466
1	79	-0.775	0.495
1	80	0.016	0.464
1	81	0.805	0.496
1	82	1.471	0.557
1	83	0.601	0.483
1	84	1.017	0.513
1	85	-0.775	0.495
1	86	1.977	0.609
1	87	-0.372	0.472
1	88	-0.372	0.472
1	89	1.471	0.557
1	90	-0.988	0.512
1	91	0.016	0.464
1	92	-0.571	0.481
1	93	-1.442	0.556
1	94	-0.372	0.472
1	95	-0.988	0.512

Group	คนที่	theta (θ)	s.e.
1	1041	-0.775	0.495
1	1042	-0.372	0.472
1	1043	-0.775	0.495
1	1044	1.471	0.557
1	1045	1.717	0.582
1	1046	-0.372	0.472
1	1047	-0.372	0.472
1	1048	-0.177	0.466
1	1049	1.238	0.534
1	1050	0.403	0.472
2	1	-0.799	0.511
2	2	-1	0.532
2	3	0.002	0.465
2	4	-1	0.532
2	5	-0.197	0.468
2	6	-0.598	0.492
2	7	-1.58	0.59
2	8	0.201	0.468
2	9	0.002	0.465
2	10	0.801	0.513
2	11	0.801	0.513
2	12	-1	0.532
2	13	-1.199	0.553
2	14	0.401	0.479
2	15	-0.598	0.492
2	16	-0.598	0.492
2	17	-1	0.532
2	18	-1.58	0.59

Group	คนที่	theta (θ)	s.e.
2	19	-1.199	0.553
2	20	-1	0.532
2	21	-1.199	0.553
2	22	0.201	0.468
2	23	-1.756	0.603
2	24	-0.598	0.492
2	25	-1	0.532
2	26	0.601	0.494
2	27	-0.598	0.492
2	28	0.002	0.465
2	29	0.201	0.468
2	30	0.401	0.479
2	31	0.002	0.465
2	32	-0.397	0.477
2	33	1.575	0.591
2	34	1.197	0.554
2	35	0.002	0.465
2	36	-0.197	0.468
2	37	1.197	0.554
2	38	-0.799	0.511
2	39	-0.397	0.477
2	40	-1	0.532
2	41	-1.393	0.573
2	42	0.601	0.494
2	43	-1.199	0.553
2	44	-0.598	0.492
2	45	-0.598	0.492
2	46	-0.197	0.468

Group	คนที่	theta (θ)	s.e.
2	47	1	0.534
2	48	-0.799	0.511
2	49	0.002	0.465
2	50	0.201	0.468
2	51	-0.197	0.468
2	52	-0.799	0.511
2	53	-0.799	0.511
2	54	-0.197	0.468
2	55	1	0.534
2	56	-1.756	0.603
2	57	-0.799	0.511
2	58	-1.58	0.59
2	59	0.201	0.468
2	60	-1.199	0.553
2	61	-0.197	0.468
2	62	-1	0.532
2	63	1.39	0.574
2	64	0.601	0.494
2	65	0.002	0.465

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้โปรแกรม IRTPRO

ภายใต้เงื่อนไข สภาวะการสูญหาย MNAR อัตราการสูญหาย 10% ความยาวแบบสอบ 20 ข้อ ขนาดตัวอย่าง 3,000 คน และขนาดอิทธิพลของ DIF=0.5

DIF Statistics for Graded Items (Back to TOC)

Item numbers in:										
Group 1	Group 2	Total X^2	*d.f.*	*p*	X^2_a	*d.f.*	*p*	$X^2_{c\|a}$	*d.f.*	*p*
1	1	12.1	2	0.0023	11.9	1	0.0006	0.2	1	0.6318
2	2	0.8	2	0.6709	0.8	1	0.3812	0.0	1	0.8609
3	3	22.0	2	0.0001	21.2	1	0.0001	0.7	1	0.3874
4	4	4.5	2	0.1031	3.8	1	0.0504	0.7	1	0.3979
5	5	3.5	2	0.1786	3.2	1	0.0729	0.2	1	0.6223
6	6	9.3	2	0.0097	5.4	1	0.0203	3.9	1	0.0486
7	7	4.5	2	0.1028	4.4	1	0.0353	0.1	1	0.7318
8	8	8.9	2	0.0116	8.9	1	0.0029	0.1	1	0.8219
9	9	5.5	2	0.0639	5.0	1	0.0256	0.5	1	0.4713
10	10	9.1	2	0.0105	7.3	1	0.0067	1.8	1	0.1852
11	11	24.6	2	0.0001	24.5	1	0.0001	0.1	1	0.7498
12	12	11.0	2	0.0042	10.9	1	0.0010	0.1	1	0.7528
13	13	26.9	2	0.0001	26.8	1	0.0001	0.1	1	0.8167
14	14	1.4	2	0.5008	1.0	1	0.3300	0.4	1	0.5104
15	15	1.5	2	0.4822	1.4	1	0.2397	0.1	1	0.7850
16	16	12.7	2	0.0018	12.7	1	0.0004	0.0	1	0.9217
17	17	9.9	2	0.0072	7.8	1	0.0052	2.1	1	0.1530
18	18	16.2	2	0.0003	16.0	1	0.0001	0.2	1	0.6259
19	19	8.3	2	0.0158	8.2	1	0.0042	0.1	1	0.7581
20	20	0.2	2	0.9094	0.1	1	0.7625	0.1	1	0.7537