# บทที่ 3
# วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ดำเนินการวิจัยโดยใช้ระเบียบวิธีการวิจัยเชิงทดลองโดยมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลและวิธีซิบเทสท์ เมื่อขนาดกลุ่มตัวอย่างย่อยระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็นอัตราส่วนต่อกันภายใต้กลุ่มตัวอย่างและความยาวแบบสอบขนาดต่าง ๆ โดยจะนำเสนอเนื้อหาเป็นตอนดังนี้

3.1 การจำลองข้อมูล

3.2 การสุ่มข้อสอบ

3.3 การสุ่มผู้สอบ

3.4 การวิเคราะห์ข้อมูลด้วยวิธีทฤษฎีการตอบสนองข้อสอบ

3.5 การวิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนส์เซลและวิธีซิบเทสท์

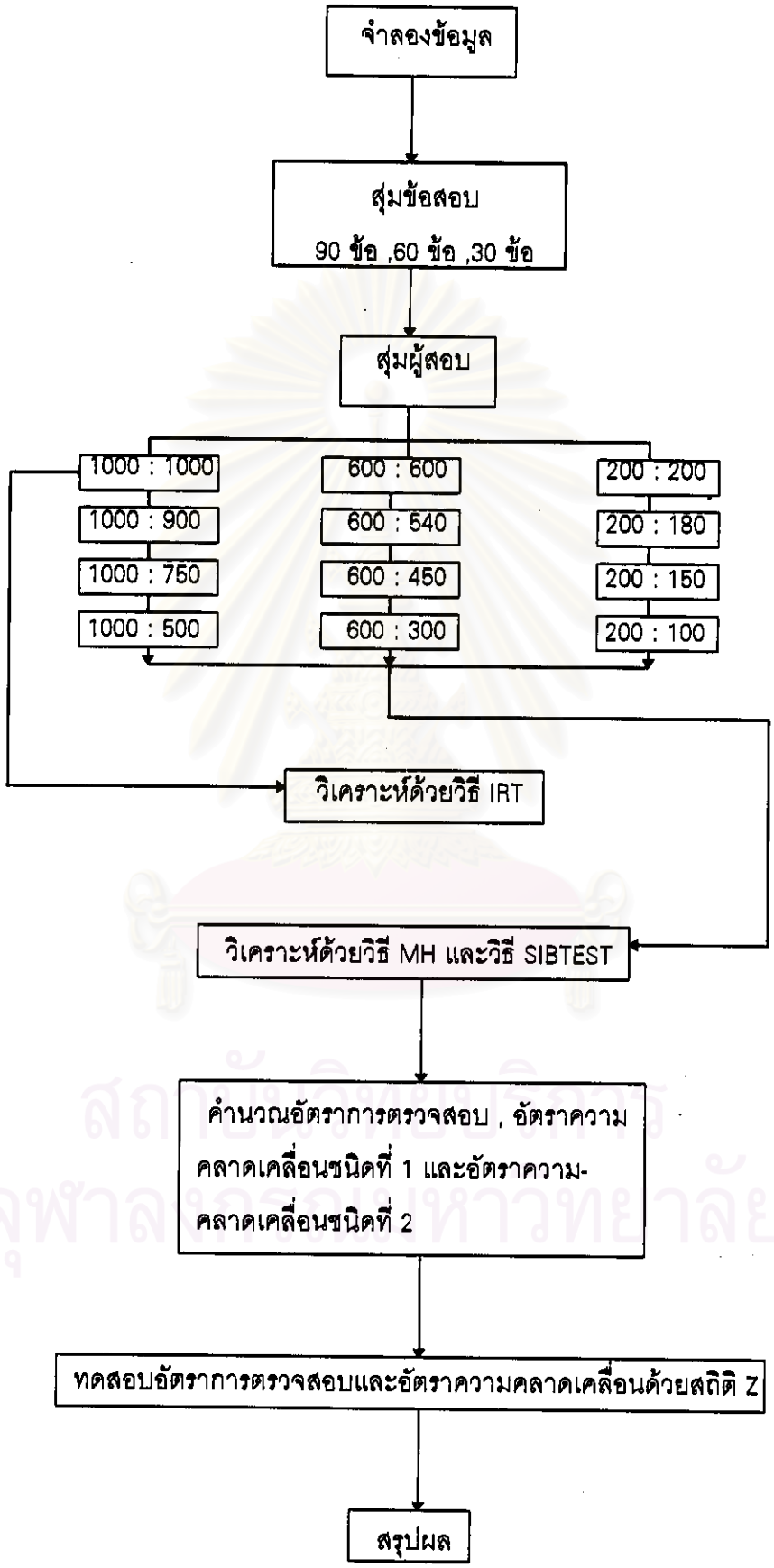
3.6 การคำนวณอัตราการตรวจสอบ อัตราความคลาดเคลื่อนชนิดที่ 1 และอัตราความคลาดเคลื่อนชนิดที่ 2

3.7 การทดสอบอัตราการตรวจสอบ อัตราความคลาดเคลื่อนชนิดที่ 1 และอัตราความคลาดเคลื่อนชนิดที่ 2 ระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ ด้วยสถิติ Z

3.8 สรุปผลจากการเปรียบเทียบผลการตรวจสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แผนภาพขั้นตอนการดำเนินการวิจัย



จากแผนภาพข้างต้น ขั้นตอนแรกเป็นการจำลองข้อมูลด้วยโปรแกรม IRTDATA โดยกำหนดจำนวนข้อสอบ 90 ข้อและจำนวนผู้สอบ 2000 คน ขั้นตอนที่ 2 เป็นการสุ่มข้อสอบเพื่อให้ได้แบบสอบที่จะศึกษา 3 ขนาดคือ 30 ข้อ , 60 ข้อ และ 90 ข้อ ขั้นตอนที่ 3 เป็นการสุ่มผู้สอบในแบบสอบแต่ละขนาดที่สุ่มได้จากขั้นตอนที่ 2 โดยสุ่มผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบขนาด 200 คน 600 คน และ 1000 คน ซึ่งในแต่ละขนาดผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีอัตราส่วนต่อกันเป็น 1:1 , 1:0.9 , 1:0.75 และ 1:0.5 ในขั้นตอนที่ 4 เป็นการวิเคราะห์ด้วยวิธี IRT โดยใช้เฉพาะเงื่อนไขที่กลุ่มตัวอย่าง 1000 คน มีอัตราส่วนระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1:1 ขั้นตอนที่ 5 เป็นการวิเคราะห์ข้อมูลในทุกเงื่อนไขที่ได้จากขั้นตอนที่ 3 ด้วยวิธี MH และวิธี SIBTEST หลังจากนั้นคำนวณอัตราการตรวจสอบและอัตราความคลาดเคลื่อนของทั้งสองวิธีในแต่ละเงื่อนไขเป็นขั้นตอนที่ 7 ส่วนขั้นตอนที่ 8 เป็นการทดสอบความแตกต่างของอัตราการตรวจสอบและอัตราความคลาดเคลื่อนระหว่างวิธี MH กับวิธี SIBTEST แล้วจึงสรุปผลการวิจัยเป็นขั้นตอนสุดท้าย

## 3.1 <u>การจำลองข้อมูล</u>

ข้อมูลที่ใช้ได้จากการจำลองตามทฤษฎีการตอบสนองข้อสอบโมเดล 3 พารามิเตอร์ ด้วยโปรแกรม IRTDATA ของ George A.Johanson ค.ศ.1992 ซึ่งประกอบด้วยโปรแกรมย่อย IRTDATA1 และโปรแกรมย่อย IRTDATA2 โปรแกรมย่อย IRTDATA1 จะเป็นโปรแกรมที่ใช้ในการเตรียมข้อมูลเกี่ยวกับคำสั่ง เพื่อให้สามารถจำลองข้อมูลได้ตามที่ผู้ใช้ต้องการ โดยในการดำเนินการจำลองของโปรแกรมจะเป็นลักษณะ interactive mode คือจะมีคำถามขึ้นมาครั้งละหนึ่งคำถาม ให้ผู้ใช้ตอบ ซึ่งจะเป็นคำถามเกี่ยวกับข้อมูลที่ผู้ใช้ต้องการ โดยในการศึกษาครั้งนี้ผู้วิจัยกำหนดจำนวนผู้สอบ 2000 คนและจำนวนข้อสอบ 90 ข้อ ค่าอำนาจจำแนกเฉลี่ย 1.0 ค่าความยากเฉลี่ย 0.0 ค่าการเดากำหนดให้เท่ากันเป็น 0.2 ผลที่ได้จากโปรแกรมย่อยนี้จะได้แฟ้มข้อมูล 4 แฟ้ม ได้แก่ แฟ้มที่มีนามสกุล irt , itm , mat และ per แล้วจึงนำแฟ้มข้อมูลที่ได้นี้เป็นข้อมูลป้อนเข้าโปรแกรมย่อย IRTDATA2 ซึ่งจะดำเนินการในลักษณะ interactive mode เช่นกัน ผลที่ได้จากโปรแกรมย่อย IRTDATA2 จะเป็นแฟ้มข้อมูล 4 แฟ้ม ได้แก่

I17273201

1. แฟ้มข้อมูลที่มีนามสกุล irt จะเป็นแฟ้มเกี่ยวกับคำสั่ง

2. แฟ้มข้อมูลที่มีนามสกุล itm จะเป็นแฟ้มข้อมูลที่เก็บค่าพารามิเตอร์ของข้อสอบ

3. แฟ้มข้อมูลที่มีนามสกุล mat จะเป็นแฟ้มข้อมูลที่เก็บผลการตอบของผู้สอบทุกคน ซึ่งจะเป็นลักษณะ 2 ค่าคือ ถูก (1) และผิด (0)

4. แฟ้มข้อมูลที่มีนามสกุล per จะเป็นแฟ้มข้อมูลที่เก็บค่าความสามารถ ($\theta$) ของผู้สอบทุกคน

**3.2 <u>การสุ่มข้อสอบ</u>**

เป็นการเตรียมข้อมูลให้มีตามเงื่อนไขที่กำหนดด้วยวิธีสุ่มอย่างง่ายแบบไม่ใส่กลับคืนจากข้อมูลที่ได้จำลองไว้ ในขั้นตอนนี้เป็นการสุ่มข้อสอบจากข้อสอบทั้งหมดที่ได้จำลองไว้จำนวน 90 ข้อ ทำการสุ่ม 2 ครั้ง ดังนี้

สุ่มครั้งที่ 1 เป็นการสุ่มข้อสอบจำนวนข้อสอบ 30 ข้อ จำนวนผู้สอบ 2000 คน มีค่าอำนาจจำแนกเฉลี่ย 0.99 และค่าความยากเฉลี่ย 0.00

สุ่มครั้งที่ 2 เป็นการสุ่มข้อสอบจำนวนข้อสอบ 60 ข้อ จำนวนผู้สอบ 2000 คน มีค่าอำนาจจำแนกเฉลี่ย 1.00 และค่าความยากเฉลี่ย 0.00

ดังนั้นจากการสุ่มในรอบนี้จะได้กลุ่มตัวอย่างเป็น 3 กลุ่มใหญ่คือ (1) ผู้สอบ 2,000 คนข้อสอบ 90 ข้อ (2) ผู้สอบ 2,000 คนข้อสอบ 60 ข้อ (3) ผู้สอบ 2,000 คนข้อสอบ 30 ข้อ

**3.3 <u>การสุ่มผู้สอบ</u>**

เป็นการสุ่มจำนวนผู้สอบ เป็นการสุ่มอย่างง่ายแบบไม่ใส่กลับคืน โดยสุ่มจากข้อมูลที่ได้จากการสุ่มในตอนที่ 2 เพื่อให้ได้จำนวนผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามอัตราส่วนและขนาดที่จะทำการศึกษาทั้งหมด 36 เงื่อนไขดังตาราง

**ตารางที่ 4 ความยาวแบบสอบและขนาดกลุ่มตัวอย่างที่ศึกษา**

| จำนวนข้อสอบ (ข้อ) | ขนาดกลุ่มตัวอย่าง (คน) | อัตราส่วน (กลุ่มอ้างอิง : กลุ่มเปรียบเทียบ) | | | |
|---|---|---|---|---|---|
| | | (1 : 1) | (1 : 0.9) | (1 : 0.75) | (1 : 0.5) |
| 30 | 200 | 200 : 200 | 200 : 180 | 200 : 150 | 200 : 100 |
| | 600 | 600 : 600 | 600 : 540 | 600 : 450 | 600 : 300 |
| | 1000 | 1000 : 1000 | 1000 : 900 | 1000 : 750 | 1000:500 |
| 60 | 200 | 200 : 200 | 200 : 180 | 200 : 150 | 200 : 100 |
| | 600 | 600 : 600 | 600 : 540 | 600 : 450 | 600 : 300 |
| | 1000 | 1000 : 1000 | 1000 : 900 | 1000 : 750 | 1000:500 |
| 90 | 200 | 200 : 200 | 200 : 180 | 200 : 150 | 200 : 100 |
| | 600 | 600 : 600 | 600 : 540 | 600 : 450 | 600 : 300 |
| | 1000 | 1000 : 1000 | 1000 : 900 | 1000 : 750 | 1000:500 |

### 3.4 การวิเคราะห์ข้อมูลด้วยวิธีทฤษฎีการตอบสนองข้อสอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการศึกษาครั้งนี้ ใช้ 3 วิธีคือ วิธีทฤษฎีการตอบสนองข้อสอบ (IRT) วิธีแมนเทล-แฮนส์เซล (MH)และวิธีซิบเทสท์ (SIBTEST) โดยจะใช้วิธีแมนเทล-แฮนส์เซลและวิธีซิบเทสท์วิเคราะห์ในทุกเงื่อนไข เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบ แต่วิธีทฤษฎีการตอบสนองข้อสอบจะใช้เป็นวิธีมาตรฐานที่จะระบุว่าข้อสอบข้อใดทำหน้าที่ต่างกัน เพื่อจะใช้เป็นเกณฑ์ในการประเมินประสิทธิภาพการตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ โดยจะใช้วิเคราะห์ในเงื่อนไขที่กลุ่มตัวอย่างมีขนาด 1,000 คนและมีอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็นหนึ่งต่อหนึ่ง (1:1) ในแบบสอบทั้ง 3 ขนาด ทั้งนี้เพราะวิธีทฤษฎีการตอบสนองข้อสอบจะตรวจสอบได้อย่างมีประสิทธิภาพเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ ดังนั้นจึงวิเคราะห์ด้วยวิธีนี้เพียง 3 เงื่อนไข ได้แก่

เงื่อนไขที่ 1 ขนาดแบบสอบ 90 ข้อ ขนาดกลุ่มตัวอย่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1000:1000

เงื่อนไขที่ 2 ขนาดแบบสอบ 60 ข้อ ขนาดกลุ่มตัวอย่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1000:1000

เงื่อนไขที่ 3 ขนาดแบบสอบ 30 ข้อ ขนาดกลุ่มตัวอย่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1000:1000

การวิเคราะห์ด้วยวิธีทฤษฎีการตอบสนองข้อสอบมีรายละเอียดดังข้อ 4.3.1 - 4.3.3

4.3.1 นำข้อมูลที่เตรียมไว้ซึ่งเป็นข้อมูลประเภท 2 ค่าคือ ถูกและผิด มาประมาณค่าพารามิเตอร์และค่าความแปรปรวนร่วมของข้อมูลด้วยโปรแกรม BILOG version 3.04 ซึ่งมีวิธีประมาณค่าให้เลือกใช้อยู่ 2 แบบคือ marginal maximum likelihood estimation (MMLE) และ marginal Bayesian estimation (MBE) ซึ่งทั้งสองวิธีนี้ให้การประมาณค่าที่ถูกต้องทั้งแบบสอบสั้นและแบบสอบยาว การประมาณค่าพารามิเตอร์ใช้การทำซ้ำด้วยวิธี EM จำนวน 10 รอบ และตามด้วยวิธี NEWTON อีก 2 รอบ ซึ่งโปรแกรมจะดำเนินการเอง หรือหากผู้ใช้จะกำหนดจำนวนรอบของการประมาณค่าซ้ำเองก็สามารถทำได้ ในการวิเคราะห์ที่ใช้ในการวิจัยครั้งนี้มี 2 ขั้นตอนคือ

ขั้นตอนที่ 1 เป็นขั้นตอนที่ใช้โปรแกรมย่อยที่ชื่อ INPUT ซึ่งจะเป็นคำสั่งให้โปรแกรมอ่านข้อมูลเข้า โดยคำสั่งในส่วนนี้จะบอกรายละเอียดเกี่ยวกับข้อมูลที่จะวิเคราะห์ โปรแกรมย่อย INPUT จะคำนวณค่าสถิติรายข้อให้ได้แก่ ร้อยละของการตอบถูกในแต่ละข้อ สัมประสิทธิ์สหสัมพันธ์ระหว่างข้อกับข้อสอบที่เหลือ และจำนวนผู้ตอบในแต่ละข้อ

ขั้นตอนที่ 2 เป็นขั้นตอนการประมาณค่า และการทดสอบความเหมาะสมของข้อมูล ต่อโมเดลโดยใช้โปรแกรมย่อยชื่อ CALIBRATE ซึ่งเป็นคำสั่งให้โปรแกรมประมาณค่าพารามิเตอร์ ผลที่ได้จากขั้นตอนนี้คือ ค่าพารามิเตอร์ของข้อสอบและผลการทดสอบความเหมาะสมของข้อมูลต่อโมเดลโดยใช้ดัชนี $\chi^2$ ซึ่งโปรแกรมจะดำเนินการทดสอบรายข้อเองถ้ามีข้อสอบตั้งแต่ 20 ข้อขึ้นไป

การวิเคราะห์ด้วยโปรแกรม BILOG นี้จะต้องมี batch file ซึ่งเป็นแฟ้มคำสั่งให้โปรแกรมดำเนินการวิเคราะห์ตามที่ผู้ใช้กำหนด ในการวิเคราะห์จะแยกกันระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ โดยแฟ้มข้อมูลของแต่ละกลุ่มที่จะใช้ในการวิเคราะห์ประกอบด้วย คำตอบที่ถูกต้อง (key)อยู่บรรทัดแรก บรรทัดที่สองจะเป็น omit key (มักจะใช้เลข 9 )และบรรทัดที่สามจะเป็นข้อมูล ซึ่งในแต่ละระเบียนประกอบด้วยเลขประจำตัวผู้สอบ และผลการตอบของผู้สอบ ผลที่ได้จากการวิเคราะห์ในแต่ละครั้งได้แก่ (1) แฟ้มข้อมูลที่ใช้นามสกุล ph1 ซึ่งเป็นผลจากการทำงานของโปรแกรมย่อย INPUT (2) แฟ้มข้อมูลที่ใช้นามสกุล ph2 ซึ่งเป็นผลจากการทำงานของโปรแกรมย่อย CALIBRATE และ(3) แฟ้มข้อมูลที่มีนามสกุล cov เป็นแฟ้มข้อมูลที่เก็บค่าความแปรปรวนร่วมซึ่งเป็นแฟ้มที่จะต้องนำไปวิเคราะห์ต่อในโปรแกรม EQUATE และโปรแกรม AREA

การวิเคราะห์ในการวิจัยครั้งนี้ผู้วิจัยเลือกใช้การประมาณค่าตามวิธีของ Bayes ทั้งนี้เพราะ Cohen และ Kim (1993) พบว่าการประมาณค่าด้วยวิธีของ Bayes จะทำให้อัตราความคลาดเคลื่อนชนิดที่ 2 น้อยกว่าแบบ MMLE และเลือกใช้โมเดล 3 พารามิเตอร์ในการประมาณค่า ผลการวิเคราะห์พบว่ามีข้อสอบบางข้อที่ไม่เหมาะสมกับโมเดล ทำให้ไม่สามารถประมาณค่าพารามิเตอร์ได้ ซึ่งจะส่งผลให้ไม่สามารถวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบได้ ดังนั้นผู้วิจัยจึงเปลี่ยนมาใช้โมเดล 2 พารามิเตอร์ ซึ่ง Kim และคณะ (1994) กล่าวว่าการใช้โมเดล 2 พารามิเตอร์และขนาดกลุ่มตัวอย่าง 1000 คน เป็นเงื่อนไขที่จะให้ผลการประมาณค่าดีที่สุด โดยสามารถใช้ได้ดีทั้งการประมาณค่าแบบ MMLE และ MBE ซึ่งโมเดล 2 พารามิเตอร์นี้โปรแกรมจะกำหนดค่าการเดาเป็น 0.00

4.3.2 หลังจากประมาณค่าพารามิเตอร์ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแล้ว ต่อจากนั้นต้องเทียบมาตรของค่าพารามิเตอร์ของกลุ่มผู้สอบด้วยโปรแกรม EQUATE 2.0 ของBaker, Al-Kani และ Al-Dosary ทั้งนี้เพราะการประมาณค่าพารามิเตอร์ของข้อสอบ ที่แยกกันวิเคราะห์แต่ละกลุ่มนั้น ไม่สามารถนำค่าพารามิเตอร์มาเปรียบเทียบกันได้โดยตรง เพราะหน่วยของมาตรความสามารถ ตามทฤษฎีการตอบสนองข้อสอบ เป็นการแยกกันคนละกลุ่ม (Camilli and Shepard,1991) จึงจำเป็นต้องเทียบมาตรของค่าพารามิเตอร์ของทั้งสองกลุ่มให้อยู่บนมาตรเดียวกันก่อน จึงจะสามารถเปรียบเทียบโค้งคุณลักษณะข้อสอบระหว่างกลุ่มได้

มาตรสำหรับกลุ่มหนึ่งจะสามารถเทียบกับกลุ่มอื่นได้โดยการแปลงด้วยสมการเชิงเส้นอย่างง่าย Y=ax + b ค่าพารามิเตอร์ของข้อสอบของกลุ่มที่สองจะเทียบให้มีมาตรเดียวกันกับกลุ่มที่หนึ่งได้ด้วยสูตรดังนี้

$$a*i2 = (1/A)ai2$$

$$b*i2 = Abi2 + K$$

เมื่อ * คือค่าพารามิเตอร์ที่เทียบแล้ว

i คือข้อสอบข้อที่ i

2 คือกลุ่มที่ต้องการจะนำไปเทียบมาตร

A, K คือค่าคงที่ซึ่งจะทำให้ค่าเฉลี่ยและความแปรปรวนของค่า b ที่แปลงแล้วของกลุ่มที่ 2 มีค่าเท่ากับค่าเฉลี่ยและความแปรปรวนของกลุ่มแรก โดย A เป็นค่าความชันและ K เป็นค่า intercept

ในการวิเคราะห์ด้วยโปรแกรม EQUATE 2.0 นั้นสามารถที่จะดำเนินการวิเคราะห์ได้โดยการใช้ batch file ซึ่งเป็นคำสั่งและบอกเกี่ยวกับข้อมูลเพื่อให้โปรแกรมทำการวิเคราะห์ หรือจะเลือกใช้แบบ interactive mode ซึ่งจะมีลักษณะเป็นคำถามที่สะคำถามให้ผู้ใช้ตอบเกี่ยวกับข้อมูลและความต้องการของผู้ใช้ ข้อมูลป้อนเข้าเป็นเมตริกซ์ที่ดัดแปลงมาจากเมตริกซ์ความแปรปรวนร่วมที่ได้จากการวิเคราะห์ด้วยโปรแกรม BILOG กล่าวคือเมตริกซ์ความแปรปรวนร่วมที่เป็นผลจากโปรแกรม BILOG นั้นจะประกอบด้วยค่าพารามิเตอร์ของข้อสอบ ค่าความแปรปรวนและค่าความแปรปรวนร่วมของค่าพารามิเตอร์ของข้อสอบ แต่ค่าที่ต้องการใช้ในโปรแกรม EQUATE 2.0 จะใช้เฉพาะค่าพารามิเตอร์ของข้อสอบเท่านั้นจึงต้องลบค่าอื่น ๆ ออกไป ในการวิเคราะห์แต่ละครั้งจะใช้ 2 เมตริกซ์คือ เมตริกซ์ค่าพารามิเตอร์ของกลุ่มอ้างอิงและเมตริกซ์ค่าพารามิเตอร์ของกลุ่มเปรียบเทียบซึ่งจะแยกกันกลุ่มละแฟ้ม ผลที่ได้จากโปรแกรม EQUATE คือค่า A(slope) และค่า K (intercept) ของสมการเชิงเส้นที่จะใช้ในการเทียบมาตรของค่าพารามิเตอร์ของกลุ่มเปรียบเทียบให้มีมาตรเดียวกันกับค่าพารามิเตอร์ของกลุ่มอ้างอิง

4.3.3 หลังจากเทียบมาตรของค่าพารามิเตอร์ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบให้อยู่บนมาตรเดียวกันแล้ว ต่อจากนั้นเป็นการคำนวณพื้นที่ที่เกิดจากความแตกต่างระหว่างโค้งคุณลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบซึ่งในการวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์สำหรับคำนวณพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบของ Raju โดยโมเดล 2 พารามิเตอร์มีสูตรในการคำนวณดังนี้ (Raju,1988 cited in Cohen et al.,1991)

$$P_i(\theta) = [1+\exp\{-1.7a_i(\theta - b_i)\}]^{-1}$$

$$= [1+\exp\{-L_i(\theta)\}]^{-1}$$

$$= \frac{\exp\{L_i(\theta)\}}{1+\exp\{L_i(\theta)\}}$$

โดยที่ $L_i(\theta) = 1.7a_i(\theta - b_i)$

พื้นที่ $(S_i)$ ภายใต้โค้งคุณลักษณะข้อสอบสำหรับข้อสอบข้อที่ i ระหว่างจุดสองจุดซึ่งเป็นช่วงที่สนใจ $(\theta 1,\theta 2)$ สามารถเขียนได้ดังนี้

$$S_i(\theta 1,\theta 2) = \int_{\theta_1}^{\theta 2} P_i(\theta)d\theta$$

$$= (-1.7a_i)^{-1} \ln[1+\exp\{L_i(\theta)\}] \Big|_{\theta_1}^{\theta 2}$$

$$= (-1.7a_i)^{-1} \ln[1+\exp\{\exp\{L_i(\theta 2)\}] - \ln[1+\exp\{L_i(\theta 1)\}]]$$

โมเดล 2 พารามิเตอร์จะมีค่าพารามิเตอร์ข้อสอบอยู่สองค่าสำหรับข้อสอบข้อที่ i $(a_{iR},b_{iR})$ สำหรับกลุ่มอ้างอิง $(a_{iF},b_{iF})$สำหรับกลุ่มเปรียบเทียบ สำหรับพื้นที่แบบมีเครื่องหมาย(SA)และพื้นที่แบบไม่มีเครื่องหมาย(UA)ระหว่างโค้งคุณลักษณะข้อสอบในช่วงปิด $(\theta 1,\theta 2)$สามารถคำนวณได้ดังนี้

$$SA = \int_{\theta_1}^{\theta 2} \{P_{iR}(\theta) - P_{iF}(\theta)\}d\theta$$

$$= \int_{\theta_1}^{\theta 2} P_{iR}(\theta)d\theta - \int_{\theta_1}^{\theta 2} P_{iF}(\theta)d\theta$$

$$= S_{iR}(\theta 1,\theta 2) - S_{iF}(\theta 1,\theta 2)$$

เมื่อ $\theta 1,\theta 2$ คือระดับความสามารถที่ต่ำกว่าและสูงกว่าตามลำดับ

$P_{iR}(\theta), P_{iF}(\theta)$ คือความน่าจะเป็นในการที่จะตอบข้อสอบข้อที่ i ได้ถูกต้องของผู้สอบที่ระดับความสามารถนั้นของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามลำดับ

$S_{iR}, S_{iF}$ คือพื้นที่ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบสำหรับข้อสอบข้อที่ i

$$UA = \int_{\theta_1}^{\theta 2} | P_{iR}(\theta) - P_{iF}(\theta) | \, d\theta$$

ถ้าโค้งคุณลักษณะข้อสอบทั้งสองโค้งไม่ตัดกัน สามารถคำนวณพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบข้อที่ i ซึ่งเป็นพื้นที่แบบไม่คิดเครื่องหมายได้โดยพื้นที่ที่คำนวณได้จะเท่ากับพื้นที่แบบมีเครื่องหมาย

$$UA = | S_{iR}(\theta 1, \theta 2) - S_{iF}(\theta 1, \theta 2) |$$
$$= | SA |$$

หลังจากคำนวณพื้นที่แล้วจะทดสอบด้วยสถิติ Z (Raju,1990 cited in Cohen and Kim,1993)

$$Z_j = \frac{b_{jF} - b_{jR}}{[Var(b_{jF}) + Var(b_{jR})]1/2}$$

โดยที่

j คือข้อสอบข้อที่ j

b คือค่าความยาก

$Var(b_{jF})$ $Var(b_{jR})$ คือค่าความแปรปรวนของ$(b_{jF})$และ$(b_{jR})$ตามลำดับ

ในการวิเคราะห์ด้วยโปรแกรม AREA แต่ละเงื่อนไขจะวิเคราะห์ 2 ครั้ง ทั้งนี้เพราะโปรแกรม AREA ประกอบด้วยโปรแกรมย่อย 2 โปรแกรม คือ AREAB และ AREASEB โดย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

โปรแกรมย่อย AREAB จะคำนวณพื้นที่แบบมีเครื่องหมายและไม่มีเครื่องหมายให้ โดยใช้ข้อมูลที่ป้อนเข้าคือ (1) แฟ้มข้อมูลที่มีค่า A และค่า K ที่ได้จากโปรแกรม EQUATE (2) แฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มอ้างอิงที่ได้จากโปรแกรม BILOG และ (3) แฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มเปรียบเทียบที่ได้จากโปรแกรม BILOG

โปรแกรมย่อย AREASEB จะทดสอบความมีนัยสำคัญด้วยสถิติ Z และ $\chi^2$ ซึ่งข้อมูลที่ป้อนเข้าคือ (1) แฟ้มข้อมูลที่ได้ผล(output)จากโปรแกรม AREAB ซึ่งก็คือพื้นที่แบบ SA และพื้นที่แบบ UA (2) แฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มอ้างอิงที่ได้จากโปรแกรม BILOG (3) แฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มเปรียบเทียบที่ได้จากโปรแกรม BILOG

โดยแฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบนี้เป็นแฟ้มเดียวกันกับที่ใช้ในโปรแกรม EQUATE และในการดำเนินการวิเคราะห์ด้วยโปรแกรม AREA นี้จะเป็นการใช้ในลักษณะ interactive mode คือโปรแกรมจะมีคำถามชื่อแฟ้มข้อมูลให้ผู้ใช้ตอบ ผลที่ได้จากโปรแกรมนี้ได้แก่ พื้นที่แบบมีเครื่องหมายและพื้นที่แบบไม่มีเครื่องหมาย ผลการทดสอบด้วยสถิติ Z และสถิติ $\chi^2$

## 3.5 การวิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนส์เซลและวิธีซิบเทสท์

นำข้อมูลที่เตรียมไว้มาคำนวณหาดัชนี $\alpha_{MH}$ และดัชนี $\beta_{SIB}$ และทดสอบนัยสำคัญ ซึ่งค่าเหล่านี้สามารถคำนวณได้พร้อมกันด้วยโปรแกรม SIBTEST ของ Stout และ Roussos (1992) แฟ้มข้อมูลที่ใช้ป้อนเข้าคือแฟ้มข้อมูลของกลุ่มอ้างอิงและแฟ้มข้อมูลของกลุ่มเปรียบเทียบ ภายในแฟ้มข้อมูลจะมีเพียงผลการตอบข้อสอบของผู้สอบเท่านั้น การวิเคราะห์ในแต่ละครั้งต้องใช้แฟ้มข้อมูลของทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ เริ่มด้วยการใช้โปรแกรมย่อย SIBIN(SIB INPUT) เพื่อจัดเตรียมข้อมูลจากแฟ้มข้อมูลการตอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ หลังจากนั้นนำแฟ้มข้อมูลที่ได้ผล(output)จากโปรแกรมย่อย SIBIN มาเป็นแฟ้มข้อมูลป้อนเข้าโปรแกรมย่อย SIBTEST ผลที่ได้ประกอบด้วยค่า $\alpha_{MH}$ , ค่า $\chi^2_{MH}$ , ค่า $\beta_{SIB}$ และค่า Z

### 3.6 คำนวณอัตราการตรวจสอบ อัตราความคลาดเคลื่อนชนิดที่ 1 และอัตราความคลาดเคลื่อนชนิดที่ 2

หลังจากวิเคราะห์ด้วยวิธีแมนเทล-แฮนส์และวิธีซิบเทสท์แล้ว นำผลการวิเคราะห์ที่ได้มาหาอัตราความถูกต้องของการตรวจสอบและอัตราความคลาดเคลื่อนด้วยการเปรียบเทียบกับผลที่ได้จากการตรวจสอบด้วยวิธีทฤษฎีการตอบสนองข้อสอบ โดยจะใช้เกณฑ์ว่า ข้อสอบที่ถูกตรวจสอบพบว่าทำหน้าที่ต่างกันด้วยวิธีทฤษฎีการตอบสนองข้อสอบถือว่าเป็นข้อที่ทำหน้าที่ต่างกัน(DIF) หากวิธีแมนเทล-แฮนส์เชลหรือวิธีซิบเทสท์ตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันซึ่งเป็นข้อเดียวกันกับที่ตรวจพบด้วยวิธีทฤษฎีการตอบสนองข้อสอบ ถือว่าวิธีนั้นมีความถูกต้องในการตรวจสอบ แต่ถ้าตรวจสอบพบว่าข้อสอบข้ออื่นทำหน้าที่ต่างกันนอกเหนือจากที่ตรวจสอบพบด้วยวิธีทฤษฎีการตอบสนองข้อสอบ ถือว่ามีความคลาดเคลื่อนในการตรวจสอบ การคำนวณอัตราความถูกต้องในการตรวจสอบและอัตราความคลาดเคลื่อนในการตรวจสอบทำได้ดังนี้

**อัตราความถูกต้องในการตรวจสอบ**

$$= \frac{\text{จำนวนข้อสอบที่ระบุว่าทำหน้าที่ต่างกันได้ถูกต้อง}}{\text{จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ}}$$

**อัตราความคลาดเคลื่อนชนิดที่** 1 (False Positive : FP)

$$= \frac{\text{จำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกัน}}{\text{จำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบสอบ}}$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

**อัตราความคลาดเคลื่อนชนิดที่ 2** (False Negative : FN)

$$= \frac{\text{จำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ไม่ต่างกัน}}{\text{จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบสอบ}}$$

### 3.7 ทดสอบอัตราการตรวจสอบ อัตราความคลาดเคลื่อนชนิดที่ 1 และอัตราความคลาดเคลื่อนชนิดที่ 2 ด้วยสถิติ Z

เป็นการเปรียบเทียบความแตกต่างของอัตราความถูกต้องของการตรวจสอบและอัตราความคลาดเคลื่อนในการตรวจสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ โดยการทดสอบด้วยค่าสถิติ Z ที่ระดับนัยสำคัญ .05

$$Zp = \frac{|P_1 - P_2|}{\sqrt{PQ(1/n_1 + 1/n_2)}}$$

$P_1, P_2$ เป็นอัตราส่วนที่ได้จากการตรวจสอบด้วยวิธี SIBTEST และวิธี MH ตามลำดับ

P เท่ากับ $\frac{(P_1)(n_1) + (P_2)(n_2)}{n_1 + n_2}$

Q เป็นค่าที่ได้จาก 1 - P

n1 ,n2 จำนวนของข้อสอบที่ตรวจสอบด้วยวิธี SIBTEST และวิธี MH

### 3.8 สรุปผลจากการเปรียบเทียบผลการตรวจสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์

ในขั้นตอนนี้เป็นการสรุปผลจากการวิเคราะห์ด้วยวิธีแมนเทล-แฮนส์ เซลและวิธีซิบเทสท์ หลังจากคำนวณอัตราการตรวจสอบ อัตราความคลาดเคลื่อนในตอนที่ 6 และทดสอบนัยสำคัญด้วยสถิติ Z ในตอนที่ 7 แล้ว