บทที่ 3

วิธีดำเนินการวิจัย

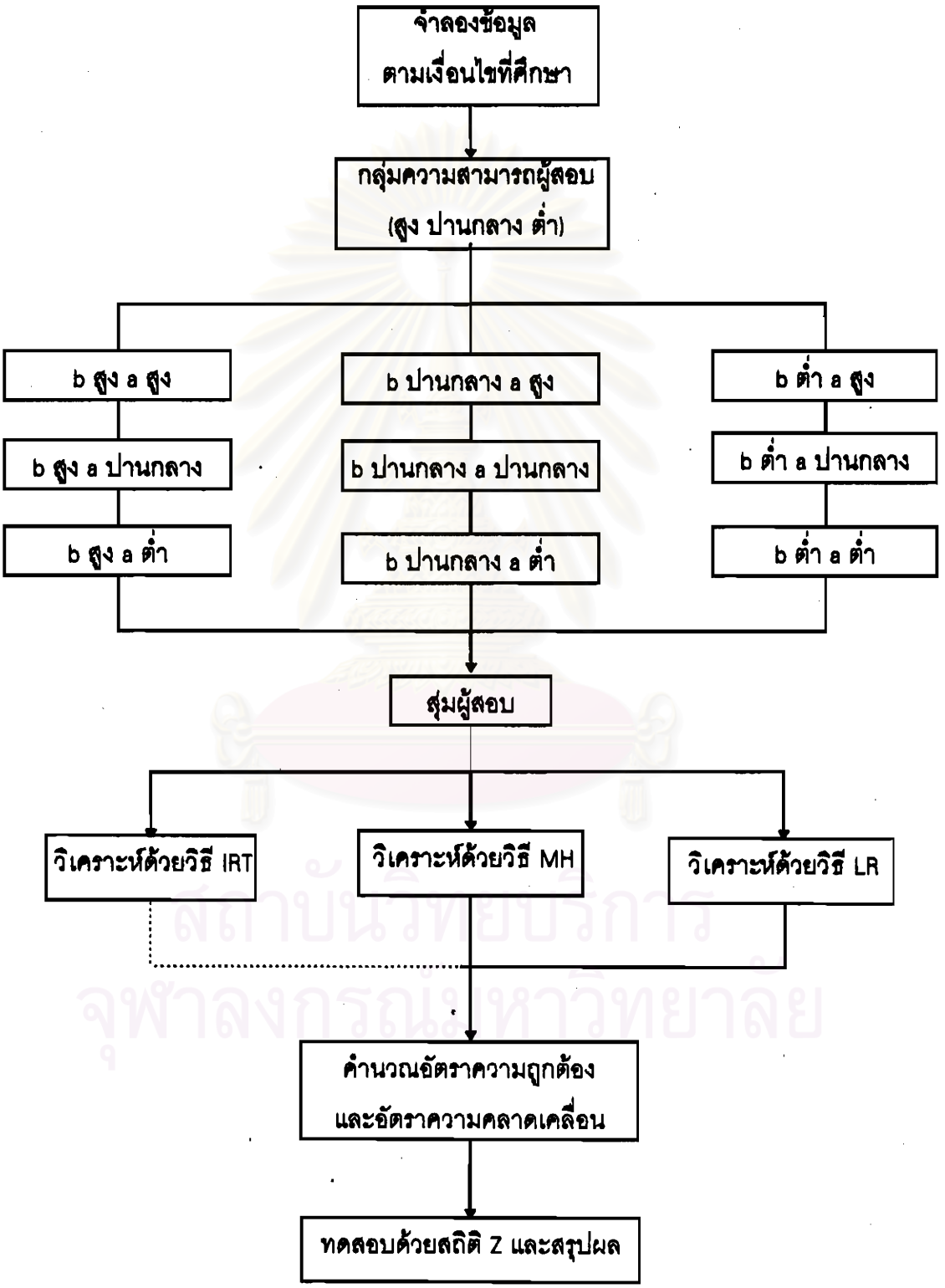
การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติก เมื่อมีการจัดกลุ่มความสามารถผู้สอบ 3 ระดับ คือ กลุ่มผู้สอบที่มีความสามารถสูง กลุ่มผู้สอบที่มีความสามารถปานกลาง และกลุ่มผู้สอบที่มีความสามารถต่ำ ค่าความยากของข้อสอบ 3 ระดับ คือกลุ่มข้อสอบที่มีค่าความยากสูง กลุ่มข้อสอบที่มีค่าความยากปานกลาง และกลุ่มข้อสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกของข้อสอบ 3 ระดับคือ กลุ่มข้อสอบที่มีค่าอำนาจจำแนกสูง กลุ่มข้อสอบที่มีค่าอำนาจจำแนกปานกลาง และกลุ่มข้อสอบที่มีค่าอำนาจจำแนกต่ำ ข้อมูลที่ศึกษาได้จากการจำลองขึ้นด้วยโปรแกรม IRTDATA ซึ่งมีขั้นตอนการดำเนินการวิจัยดังต่อไปนี้

1. จำลองข้อมูล
2. สุ่มผู้สอบ
3. วิเคราะห์ด้วยข้อมูลด้วยวิธีทฤษฎีการตอบสนองข้อสอบ
4. วิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนส์เซล
5. วิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติก
6. คำนวณอัตราความถูกต้องและอัตราความคลาดเคลื่อน
7. ทดสอบอัตราความถูกต้องและอัตราความคลาดเคลื่อนด้วยสถิติทดสอบ z

ดังแสดงในแผนภาพที่ 3

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แผนภาพที่ 3 ขั้นตอนการดำเนินการวิจัย



จากแผนภาพขั้นตอนการดำเนินการวิจัย ขั้นตอนแรกเป็นการจำลองข้อมูลด้วยโปรแกรม IRTDATA โดยกำหนดจำนวนข้อสอบ 40 ข้อ และจำนวนผู้สอบ 2000 คนตามเงื่อนไขที่ต้องการศึกษารวมทั้งหมด 27 เงื่อนไข ขั้นตอนที่ 2 เป็นการสุ่มผู้สอบในในแต่ละเงื่อนไข โดยสุ่มผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีสัดส่วนเท่ากันคือกลุ่มละ 1000 คน ขั้นตอนที่ 3 เป็นการวิเคราะห์ข้อมูลในทุกเงื่อนไขที่ได้จากขั้นตอนที่ 3 ด้วยวิธี IRT ขั้นตอนที่ 4 เป็นการวิเคราะห์ข้อมูลด้วยวิธี MH และวิธี LR ขั้นตอนที่ 5 คำนวณอัตราความถูกต้องของการตรวจสอบและอัตราความคลาดเคลื่อนของทั้งสองวิธีในแต่ละเงื่อนไข ขั้นตอนที่ 6 เป็นการทดสอบอัตราความถูกต้องในการตรวจสอบและอัตราความคลาดเคลื่อนระหว่างวิธี MH และวิธี LR ด้วยสถิติ Z แล้วสรุปจึงผลการวิจัยเป็นขั้นตอนสุดท้าย

1.การจำลองข้อมูล

ข้อมูลที่ใช้ได้จากการจำลองข้อมูลตามทฤษฎีการตอบสนองข้อสอบโมเดล 3 พารามิเตอร์ด้วยโปรแกรม IRTDATA ของ George A. Johanson ค.ศ.1992 ซึ่งการกำหนดค่าพารามิเตอร์ของข้อสอบได้จากการศึกษาค่าพารามิเตอร์ของโมดลการตอบสนองข้อสอบ ดังนี้

1) พารามิเตอร์ของผู้สอบ (θ_p) คือระดับความสามารถของผู้สอบคนที่ p ซึ่งประมาณได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ นิยมปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 ส่วนเบี่ยงเบนมาตรฐานเป็น 1 ค่า θ มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ผลการวิเคราะห์ส่วนใหญ่มักให้ค่าอยู่ในช่วง -3 ถึง +3 งานวิจัยของคณิต ไข่มุกด์ (2534) ได้แบ่งกลุ่มผู้สอบที่ระดับความสามารถต่างๆ ดังนี้ กลุ่มผู้สอบที่มีความสามารถสูง ($0.5<\theta<3.0$) กลุ่มผู้สอบที่มีความสามารถปานกลาง ($-1.0<\theta<1.0$) และกลุ่มผู้สอบที่มีความสามารถต่ำ ($-3.0<\theta<0.0$)

2) ค่าความยากของข้อสอบ (b) หมายถึงระดับความยากของข้อสอบซึ่งมีค่าเท่ากับระดับความสามารถของผู้สอบ(θ) ที่มีโอกาสทำข้อสอบข้อนั้นถูก (1+c) / 2 ค่าความยากของข้อสอบมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่าความยากของข้อสอบ อยู่ระหว่าง -2.5 ถึง +2.5 งานวิจัยของคณิต ไข่มุกด์ (2534) ได้แบ่งค่าความยาก(b) เป็น 3 ระดับ โดยใช้ตำแหน่งเปอร์เซ็นไทล์ที่ 25 และ 75 ดังนี้ ค่าความยากระดับสูง ($b > 0.67$) ค่าความยากระดับปานกลาง ($-0.67 \leq b \leq 0.67$) และค่าความยากระดับต่ำ ($b < -0.67$)

3) ค่าอำนาจจำแนกของข้อสอบ (a) หมายถึงค่าความชันของโค้งคุณลักษณะข้อสอบที่จุดเปลี่ยนโค้ง ค่าอำนาจจำแนกมีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ ในทางปฏิบัตินิยมใช้ข้อสอบที่มีค่า

อำนาจจำแนกอยู่ระหว่าง 0.5 ถึง 2.5 งานวิจัยของคณิต ไข่มุกด์ (2534) ได้แบ่งค่าอำนาจจำแนกเป็น 3 ระดับโดยใช้ตำแหน่งเปอร์เซ็นไทล์ที่ 25 และ 75 ดังนี้ ค่าอำนาจจำแนกระดับสูง ($a>0.85$) ค่าอำนาจจำแนกระดับปานกลาง ($0.55\leq a\leq 0.85$) และค่าอำนาจจำแนกระดับต่ำ ($a< 0.55$)

ดังนั้นผู้วิจัยจึงจำลองข้อมูลตามค่าพารามิเตอร์ที่ได้จากการศึกษาของ คณิต ไข่มุกด์ (2534) ดังนี้

1. กลุ่มความสามารถผู้สอบ

1.1 กลุ่มผู้สอบที่มีความสามารถสูง กำหนดค่าความสามารถเฉลี่ยเท่ากับ 1.5 และส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 1 ซึ่งพิสัยของค่าความสามารถที่จำลองได้อยู่ระหว่าง 0.5 ถึง 3.0

1.2 กลุ่มผู้สอบที่มีความสามารถปานกลาง กำหนดค่าความสามารถเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 1 ซึ่งพิสัยของค่าความสามารถที่จำลองได้อยู่ระหว่าง -1.0 ถึง 1.0

1.3 กลุ่มผู้สอบที่มีความสามารถต่ำ กำหนดค่าความสามารถเฉลี่ยเท่ากับ -1.5 และส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 1 ซึ่งพิสัยของค่าความสามารถที่จำลองได้อยู่ระหว่าง -3.0 ถึง 0

2. ค่าความยากของข้อสอบ

2.1 กลุ่มข้อสอบที่มีค่าความยากสูง กำหนดค่าความยากเฉลี่ยเท่ากับ 1.835 และส่วนเบี่ยงเบนมาตรฐานเฉลี่ยอยู่ระหว่าง 0.30 ถึง 0.40 ซึ่งพิสัยค่าความยากของข้อสอบที่จำลองได้อยู่ระหว่าง 0.67 ถึง 2.50

2.2 กลุ่มข้อสอบที่มีค่าความยากปานกลาง กำหนดค่าความยากเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเฉลี่ยอยู่ระหว่าง 0.25 ถึง 0.35 ซึ่งพิสัยค่าความยากของข้อสอบที่จำลองได้อยู่ระหว่าง -0.67 ถึง 0.67

2.3 กลุ่มข้อสอบที่มีค่าความยากต่ำ กำหนดค่าความยากเฉลี่ยเท่ากับ -1.835 และส่วนเบี่ยงเบนมาตรฐานเฉลี่ยอยู่ระหว่าง 0.25 ถึง 0.45 ซึ่งพิสัยค่าความยากของข้อสอบที่จำลองได้อยู่ระหว่าง -2.5 ถึง -0.67

3. ค่าอำนาจจำแนกของข้อสอบ

3.1 กลุ่มข้อสอบที่มีค่าอำนาจจำแนกสูง กำหนดค่าอำนาจจำแนกเฉลี่ยเท่ากับ 1.675 ส่วนเบี่ยงเบนมาตรฐานเฉลี่ยอยู่ระหว่าง 0.25 ถึง 0.35 ซึ่งพิสัยค่าอำนาจจำแนกของข้อสอบที่จำลองได้อยู่ระหว่าง 0.85 ถึง 2.5

3.2 กลุ่มข้อสอบที่มีค่าอำนาจจำแนกปานกลาง กำหนดค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.7 ส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 0.055 ซึ่งพิสัยค่าอำนาจจำแนกของข้อสอบที่จำลองได้อยู่ระหว่าง 0.55 ถึง 0.85

3.3 กลุ่มข้อสอบที่มีค่าอำนาจจำแนกต่ำ กำหนดค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.525 ส่วนเบี่ยงเบนมาตรฐานเฉลี่ยอยู่ระหว่าง 0.008 ถึง 0.01 ซึ่งพิสัยค่าอำนาจจำแนกของข้อสอบที่จำลองได้อยู่ระหว่าง 0.50 ถึง 0.55

จากการจำลองข้อมูลได้เงื่อนไขที่ศึกษาทั้งหมด 27 เงื่อนไข และในแต่ละเงื่อนไขกำหนดจำนวนผู้สอบ 2000 คน ความยาวแบบสอบ 40 ข้อ ทั้งนี้เพราะความยาวแบบสอบ 40 ข้อ เป็นความยาวเฉลี่ยของแบบสอบย่อย (subtest)ที่ใช้ในแบบสอบวัดผลสัมฤทธิ์มาตรฐาน (Rogers and Swaminathan, 1993) และจากการศึกษางานวิจัยพบว่าความยาวแบบสอบ 40 ข้อเป็นความยาวแบบสอบที่พอเพียงสำหรับตรวจสอบข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยผลการตรวจสอบเมื่อใช้ความยาวแบบสอบ 40 ข้อสามารถตรวจค้นข้อสอบทำหน้าที่ต่างกันได้ใกล้เคียงกับความยาวแบบสอบ 80 ข้อ (Uttaro and Millsap, 1994 ; Narayanan and Swaminathan, 1996) ส่วนค่าการเดากำหนดให้เท่ากันในทุกเงื่อน คือ 0.2 ดังนี้

ตารางที่ 4 การจัดกลุ่มความสามารถผู้สอบและลักษณะแบบสอบที่ศึกษา

การจัดกลุ่มความสามารถผู้สอบ	ลักษณะแบบสอบที่ศึกษา	
	ค่าความยาก (b)	ค่าอำนาจจำแนก (a)
กลุ่มผู้สอบที่มีความสามารถสูง ($\bar{\theta}$ = 1.5)	$\bar{b}$ = 1.835	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525
	$\bar{b}$ = 0	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525
	$\bar{b}$ = -1.835	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525

ตารางที่ 4 (ต่อ)

การจัดกลุ่มความสามารถผู้สอบ	ลักษณะแบบสอบที่ศึกษา	
	ค่าความยาก (b)	ค่าอำนาจจำแนก (a)
กลุ่มผู้สอบที่มีความสามารถปานกลาง ($\bar{\theta}$ = 0)	$\bar{b}$ = 1.835	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525
	$\bar{b}$ = 0	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525
	$\bar{b}$ = -1.835	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525
กลุ่มผู้สอบที่มีความสามารถต่ำ ($\bar{\theta}$ = -1.5)	$\bar{b}$ = 1.835	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525
	$\bar{b}$ = 0	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525
	$\bar{b}$ = -1.835	$\bar{a}$ = 1.675 $\bar{a}$ = 0.7 $\bar{a}$ = 0.525

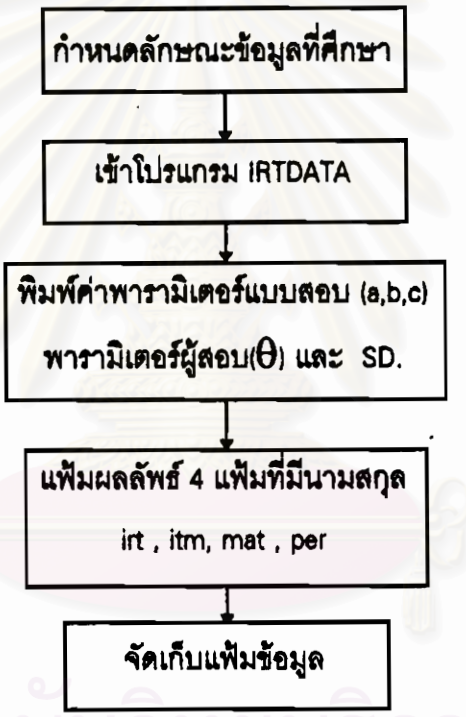
ในการจำลองข้อมูลด้วยโปรแกรม IRTDATA รวมเงื่อนไขทั้งหมด 27 เงื่อนไข เป็นการจำลองข้อสอบต่างชุดกันตามระดับความสามารถผู้สอบ ซึ่งภายในเงื่อนไขเดียวกันที่จำลองขึ้นนั้นจะใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือวิธี IRT วิธี MH และวิธี LR

การจำลองข้อมูลด้วยโปรแกรม IRTDATA จะเป็นลักษณะ interactive mode คือจะมีคำถามขึ้นมาครั้งละหนึ่งคำถามให้ผู้ใช้ตอบ ซึ่งเป็นคำถามเกี่ยวกับข้อมูลที่ผู้ใช้ต้องการ ผลลัพธ์ที่ได้จากการจำลองข้อมูลด้วยโปรแกรม IRTDATA คือแฟ้มข้อมูล 4 แฟ้มดังนี้

1. แฟ้มข้อมูลที่มีนามสกุล irt เป็นแฟ้มเกี่ยวกับคำสั่ง
2. แฟ้มข้อมูลที่มีนามสกุล itm เป็นแฟ้มข้อมูลที่เก็บค่าพารามิเตอร์ของข้อสอบ
3. แฟ้มข้อมูลที่มีนามสกุล mat เป็นแฟ้มข้อมูลที่เก็บผลการตอบของผู้สอบทุกคน
4. แฟ้มข้อมูลที่มีนามสกุล per เป็นแฟ้มข้อมูลที่เก็บค่าความสามารถของผู้สอบทุกคน

สำหรับขั้นตอนการจำลองข้อมูลด้วยโปรแกรม IRTDATA ดำเนินการดังนี้

แผนภาพที่ 4 ขั้นตอนการจำลองข้อมูลด้วยโปรแกรม IRTDATA



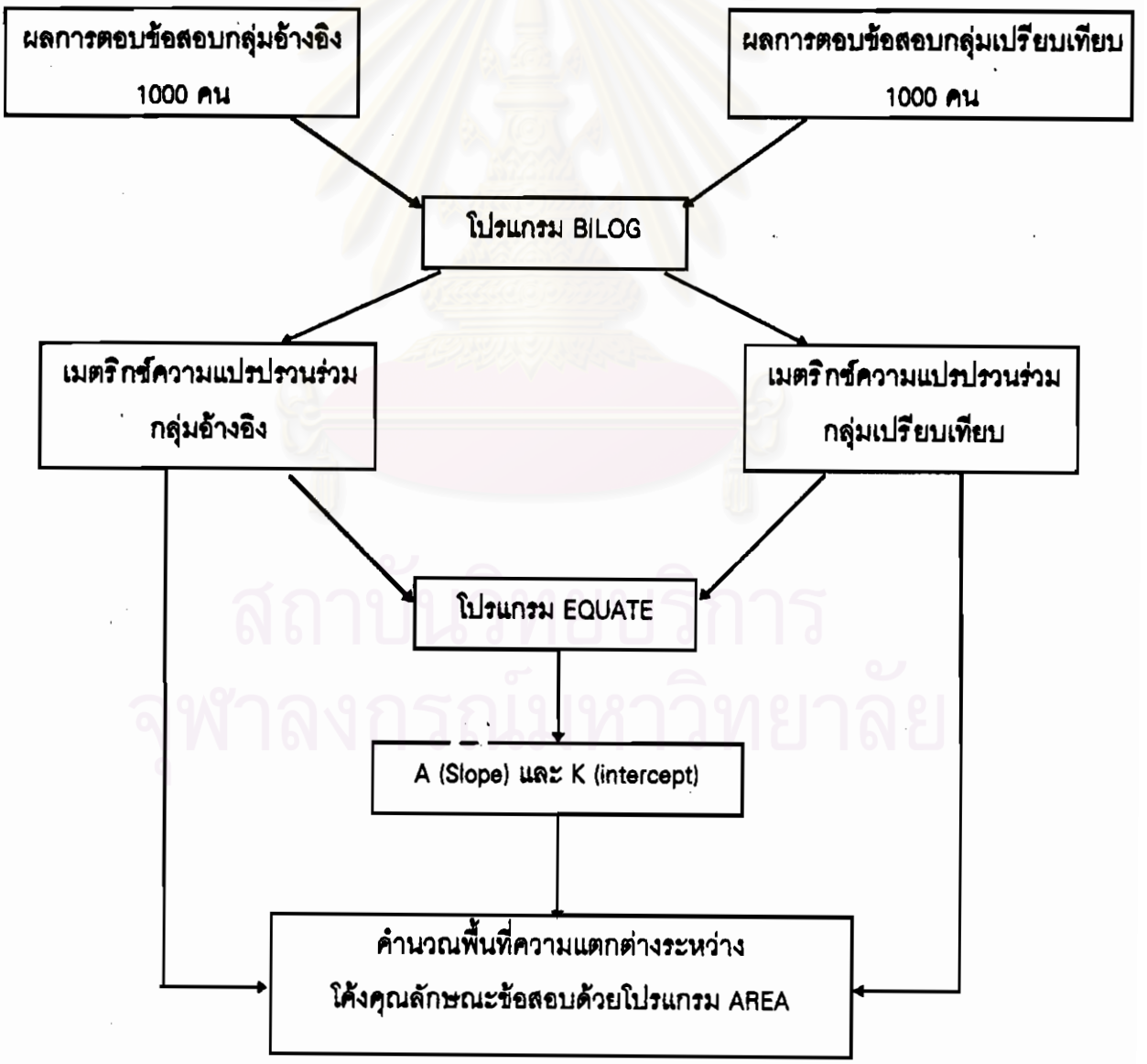
2. การสุ่มผู้สอบ

เป็นการสุ่มจำนวนผู้สอบด้วยวิธีสุ่มอย่างง่ายจากข้อมูลที่ได้จำลองไว้โดยสุ่มจำนวนผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีสัดส่วนเท่ากัน คือกลุ่มละ 1000 คน ในเงื่อนไขทั้งหมด 27 เงื่อนไข ทั้งนี้เพราะเมื่อใช้ขนาดกลุ่มตัวอย่างมากขึ้น อัตราการตรวจสอบการทำหน้าที่ต่างกันจะเพิ่มสูงขึ้น (Rogers and Swaminathan, 1993 ; Narayanan and Swaminathan, 1996) และงานวิจัยของจิติมา วรรณศรี (2540) พบว่าภายใต้ความยาวแบบสอบเดียวกัน ขนาดกลุ่มตัวอย่างที่เป็นเงื่อนไขที่ดีที่สุด สามารถตรวจพบข้อสอบทำหน้าที่ต่างกันได้ร้อยละ 100 โดยที่ไม่พบว่ามีการระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน คือกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีขนาดกลุ่มละ 1000 คน

3. การวิเคราะห์ข้อมูลด้วยวิธีทฤษฎีการตอบสนองข้อสอบ

ในการตรวจสอบข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปครั้งนี้ ใช้วิธีทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นวิธีมาตรฐานที่จะระบุว่าข้อสอบข้อใดทำหน้าที่ต่างกัน เพื่อจะใช้เป็นเกณฑ์ในการประเมินอัตราความถูกต้องและอัตราความคลาดเคลื่อนในการตรวจสอบของวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติก โดยใช้วิเคราะห์ทุกเงื่อนไขที่ศึกษา

แผนภาพที่ 5 ขั้นตอนการวิเคราะห์ข้อมูลด้วยวิธีทฤษฎีการตอบสนองข้อสอบ



จากแผนภาพการวิเคราะห์ข้อมูลด้วยวิธีทฤษฎีการตอบสนองข้อสอบเริ่มจากนำแฟ้มข้อมูลผลการตอบข้อสอบของผู้สอบสองกลุ่มคือกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ มาคำนวณเมตริกซ์ค่าพารามิเตอร์และค่าความแปรปรวนร่วมด้วยโปรแกรม BILOG จากนั้นเทียบมาตรค่าพารามิเตอร์ของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบด้วยโปรแกรม EQUATE โดยใช้ข้อมูลป้อนเข้าเป็นเมตริกซ์ที่ดัดแปลงมาจากเมตริกซ์ความแปรปรวนร่วมที่ได้จากการวิเคราะห์ด้วยโปรแกรม BILOG ผลที่ได้จากโปรแกรม EQUATE คือค่า A(Slope) และ K (intercept) ต่อจากนั้นคำนวณพื้นที่ที่เกิดจากความแตกต่างระหว่างโค้งคุณลักษณะข้อสอบของผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบด้วยโปรแกรม AREA ซึ่งในการวิเคราะห์ด้วยโปรแกรม AREA นั้นต้องใช้ข้อมูลนำเข้าคือแฟ้มข้อมูลที่มีค่า A และ K จากโปรแกรม EQUATE และแฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ได้จากโปรแกรม BILOG

ขั้นตอนการวิเคราะห์มีรายละเอียดดังนี้

3.1 นำผลการตอบข้อสอบของผู้สอบสองกลุ่มคือกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ซึ่งแยกกันในแต่ละแฟ้มมาคำนวณเมตริกซ์ค่าพารามิเตอร์และค่าความแปรปรวนร่วมของข้อมูลผลการตอบด้วยโปรแกรม BILOG 3.04 ซึ่งมีวิธีการประมาณค่าให้เลือกใช้ 2แบบคือ marginal maximum likelihood (MMLE)และ marginal Bayesian estimation (MBE) โดยทั้งสองแบบให้ผลการประมาณค่าที่ถูกต้องทั้งแบบสอบสั้นและแบบสอบยาว

จำนวนผู้สอบที่เหมาะสมในการประมาณค่าพารามิเตอร์ของข้อสอบที่ใช้ในโปรแกรม BILOG ควรใช้ผู้สอบประมาณ 200 คนขึ้นไป ถ้าจำนวนข้อสอบน้อย (น้อยกว่า 10 ข้อ) ควรใช้ผู้สอบที่มีขนาดใหญ่เพื่อให้ได้ค่าประมาณของพารามิเตอร์ที่ถูกต้องมากที่สุด ในการวิเคราะห์ด้วยโปรแกรม BILOG จะต้องมีไฟล์คำสั่งให้โปรแกรมดำเนินการวิเคราะห์ตามที่ผู้ใช้กำหนด ส่วนแฟ้มข้อมูลที่ใช้ในการวิเคราะห์ประกอบด้วย บรรทัดแรกเป็นคำตอบที่ถูกต้อง บรรทัดที่สองเป็น omit key (ส่วนใหญ่ใช้เลข 9) และบรรทัดที่สามเป็นข้อมูลประกอบด้วยเลขประจำตัวสอบและผลการตอบข้อสอบของผู้สอบ ผลลัพธ์ที่ได้จากการวิเคราะห์ในแต่ละเงื่อนไขได้แก่ 1) แฟ้มข้อมูลที่มีนามสกุล ph1 2) แฟ้มข้อมูลที่มีนามสกุล ph2 3) แฟ้มข้อมูลที่มีนามสกุล cov ซึ่งเป็นแฟ้มข้อมูลที่เก็บค่าความแปรปรวนร่วมและเป็นแฟ้มที่นำไปเป็นข้อมูลนำเข้าในโปรแกรม EQUATE และโปรแกรม AREA

3.2 เทียบมาตรของค่าพารามิเตอร์สำหรับกลุ่มผู้สอบด้วยโปรแกรม EQUATE 2.0 ของ Baker, Al-Kani และ Al-Dosary (1991) ทั้งนี้เพราะการเปรียบเทียบโค้งลักษณะข้อสอบของผู้สอบสองกลุ่ม ต้อง แปลงมาตรของค่าพารามิเตอร์ที่แสดงลักษณะข้อสอบที่ได้จากผู้สอบ 2 กลุ่มที่แยกกันวิเคราะห์ให้อยู่บนมาตรเดียวกันก่อนการเปรียบเทียบ

มาตรของค่าพารามิเตอร์ของข้อสอบของผู้สอบกลุ่มที่สองจะเทียบให้มีมาตรเดียวกันกับผู้สอบกลุ่มที่หนึ่งด้วยสูตรดังนี้

$$a^{*}i2 = (1/A)ai2$$

$$b^{*}i2 = Abi2 + K$$

เมื่อ * คือค่าพารามิเตอร์ที่เทียบแล้ว

i คือข้อสอบข้อที่ i

2 คือกลุ่มที่ต้องการจะนำไปเทียบมาตร

A, K คือค่าคงที่ซึ่งจะทำให้ค่าเฉลี่ยและความแปรปรวนของค่า b ที่แปลงแล้วของกลุ่มที่ 2 มีค่าเท่ากับค่าเฉลี่ยและความแปรปรวนของกลุ่มแรก โดย A เป็นค่าความชัน และ K เป็นค่า inirtcept

การวิเคราะห์ด้วยโปรแกรม EQUATE สามารถดำเนินการวิเคราะห์ได้โดยการใช้ batch file ซึ่งเป็นคำสั่งให้โปรแกรมทำการวิเคราะห์ หรือจะเลือกใช้แบบ intertive mode ซึ่งมีลักษณะเป็นคำถามทีละคำถามให้ผู้ใช้ตอบเกี่ยวกับข้อมูลตามความต้องการของผู้ใช้ ข้อมูลนำเข้าเป็นเมตริกซ์ที่ดัดแปลงมาจากเมตริกซ์ความแปรปรวนร่วมที่ได้จากการวิเคราะห์ด้วยโปรแกรม BILOG ในการวิเคราะห์แต่ละเงื่อนไขจะใช้ 2 เมตริกซ์ คือเมตริกซ์ของกลุ่มอ้างอิงและเมตริกซ์ของกลุ่มเปรียบเทียบซึ่งแยกกันคนละกลุ่มแฟ้ม ผลลัพธ์ที่ได้จากการวิเคราะห์จะได้ ค่า A (slope) และค่า K (intercept) ของสมการเชิงเส้นที่ใช้เทียบมาตรค่าพารามิเตอร์กลุ่มเปรียบเทียบให้มีมาตรเดียวกันกับค่าพารามิเตอร์กลุ่มอ้างอิง

3.3 คำนวณพื้นที่ความแตกต่างระหว่างโค้งคุณลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ โดยใช้โปรแกรมคำนวณพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบของ Raju แบบสองพารามิเตอร์ ซึ่งจากการศึกษาของ Kim และคณะ (1994) พบว่าแบบจำลองของทฤษฎีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์มีความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าแบบจำลอง 3 พารามิเตอร์และการใช้โมเดล 2 พารามิเตอร์และกลุ่มตัวอย่าง 1000 คน เป็นเงื่อนไขที่ให้ผลการประมาณค่าดีที่สุด ดังนั้นจึงเลือกใช้วิธี IRT แบบ 2 พารามิเตอร์ตรวจสอบการทำหน้าที่ต่างกัน

ของข้อสอบแบบอเนกรูป ซึ่งเป็นการคำนวณค่าดัชนีไม่คิดเครื่องหมาย (UA) มีสูตรในการคำนวณดังนี้ (Raju, 1990)

$$UA = |H|$$

เมื่อ

$$H = \frac{2(a_2 - a_1)}{Da_1a_2} \ln\left\{1+ \exp\left[\frac{Da_1a_2(b_2 - b_1)}{a_2 - a_1}\right]\right\} - (b_2 - b_1)$$

และ

$$Var(H) = B_1^2 Var(b_1) + B_2^2 Var(b_2) + A_1^2 Var(a_1) + A_2^2 Var(a_2) + 2B_1A_1COV(b_1a_1) + 2B_2A_2COV(b_2a_2)$$

ค่าน้ำหนักของ B_1, B_2, A_1 และ A_2 คำนวณได้ดังนี้

$$B_1 = 1 - \frac{2\exp(Y)}{1+ \exp(Y)}$$

$$B_2 = - B_1$$

$$A_1 = \frac{2}{a_1^2}\left\{\frac{a_1a_2(b_2 - b_1)}{a_2 - a_1}\frac{\exp(Y)}{1+ \exp(Y)} - \frac{\ln\left[1+ \exp(Y)\right]}{D}\right\}$$

$$A_2 = \frac{- a_1^2 A_1}{a_2^2}$$

และ

$$Y = \frac{Da_1a_2(b_2 - b_1)}{a_2 - a_1}$$

หลังจากคำนวณพื้นที่แล้วทดสอบด้วยสถิติ Z ดังนี้ (Raju, 1990)

$$Z = \frac{H}{Var\ (H)}$$

ในการวิเคราะห์โปรแกรม AREA ของ Ruju มี 2 ขั้นตอน เพราะโปรแกรม AREA ประกอบด้วยโปรแกรมย่อย 2 โปรแกรมคือ AREAB และ AREASEB

โปรแกรมย่อย AREAB จะคำนวณพื้นที่แบบมีเครื่องหมาย (SA) และพื้นที่ไม่มีเครื่องหมาย (UA) ในการวิเคราะห์ใช้ข้อมูลนำเข้าคือ 1)แฟ้มข้อมูลที่มีค่า A และค่า K ที่ได้จากโปรแกรม EQUATE 2) แฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มอ้างอิงที่ได้จากโปรแกรม BILOG 3) แฟ้ม

ข้อมูลความแปรปรวนร่วมของกลุ่มเปรียบเทียบที่ได้จากโปรแกรม BILOG ผลลัพธ์ที่ได้จากโปรแกรมย่อยนี้คือค่าพื้นที่มีเครื่องหมายและพื้นที่ไม่มีเครื่องหมาย

โปรแกรมย่อย AREASEB จะทดสอบนัยสำคัญด้วยสถิติ Z และ χ^2 ในการวิเคราะห์ใช้ข้อมูลนำเข้าคือ 1)แฟ้มข้อมูลผลลัทธ์จากโปรแกรม AREAB 2)แฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มอ้างอิงที่ได้จากโปรแกรม BILOG 3) แฟ้มข้อมูลความแปรปรวนร่วมของกลุ่มเปรียบเทียบที่ได้จากโปรแกรม BILOG ผลลัพธ์ที่ได้จากโปรแกรมย่อยนี้ได้แก่ พื้นที่แบบมีเครื่องหมายและพื้นที่แบบไม่มีเครื่องหมาย ผลการทดสอบด้วยสถิติ Z และสถิติ χ^2

4.การวิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนส์เซล

วิเคราะห์ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปด้วยวิธีแมนเทล-แฮนส์เซล โดยใช้โปรแกรม MH-DIF ที่พัฒนาโดย Fidago A. (1995) ในการวิเคราะห์ใช้เทคนิคการวิเคราะห์แบบ 2 ขั้นตอน เนื่องจาก Clauser และคณะ (1993) พบว่าผลการตรวจสอบข้อสอบทำหน้าที่ต่างกันด้วยเทคนิค 2 ขั้นตอน พบข้อสอบที่ทำหน้าที่ต่างกันเท่ากันหรือสูงกว่าเทคนิค 1 ขั้นตอน โดยไม่ทำให้อัตราความคลาดเคลื่อนชนิดที่ 1 เพิ่มสูงขึ้น ส่วนเกณฑ์ในการจับคู่ชั้นคะแนน (score category) ใช้คะแนนที่ผู้สอบสามารถทำได้สูงสุด (total score) โดยเปรียบเทียบคู่ชั้นคะแนนที่ห่างกันช่วงละ 1 คะแนนจากคะแนนรวมต่ำสุดถึงคะแนนรวมสูงสุดที่ผู้สอบทั้งหมดทำได้ และเกณฑ์ในการคัดเลือกผู้สอบเข้ากลุ่มสำหรับวิเคราะห์ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป จะใช้ค่าเฉลี่ยของคะแนนรวมทั้งหมดเป็นเกณฑ์ในการคัดเลือกผู้สอบ กล่าวคือผู้สอบที่ได้คะแนนรวมสูงกว่าค่าเฉลี่ยจัดให้เป็นกลุ่มที่มีความสามารถสูง ส่วนผู้สอบที่ได้คะแนนรวมน้อยกว่าหรือเท่ากับค่าเฉลี่ยจัดให้เป็นกลุ่มที่มีความสามารถต่ำ จากนั้นจึงคำนวณความน่าจะเป็นในรูปของสัดส่วนการตอบข้อสอบถูกและผิดระหว่างกลุ่มสูงและกลุ่มต่ำ

ในการวิเคราะห์ด้วยโปรแกรม MH-DIF แฟ้มข้อมูลที่ใช้ป้อนเข้าคือ ผลการตอบข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่จัดเก็บอยู่ในแฟ้มเดียวกัน ซึ่งในการจัดเก็บผลการตอบของผู้สอบสองกลุ่มจะต้องมีรหัส (code) บอกว่าผลการตอบข้อสอบเป็นของผู้สอบกลุ่มอ้างอิงหรือกลุ่มเปรียบเทียบ ผลลัพธ์ที่ได้จากการวิเคราะห์คือ ดัชนี α_{MH} ดัชนี MH_{DIF} และการทดสอบนัยสำคัญด้วยค่าสถิติไคสแควร์

5. การวิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติก

การวิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติกใช้โปรแกรม SPSS/PC+ เพื่อทำนายโอกาสในการตอบข้อสอบถูกหรือผิดของผู้สอบ โดยใช้กลุ่มผู้สอบ (กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ) ความสามารถผู้สอบ (θ) และปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับความสามารถผู้สอบเป็นตัวทำนาย สำหรับแฟ้มข้อมูลที่ใช้ป้อนเข้าประกอบด้วยรหัส (code) กลุ่มผู้สอบ และผลการตอบข้อสอบของผู้สอบที่ให้คะแนนถูก (1) และผิด (0)

ในการพิจารณาข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปจะพิจารณาจากค่าปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับความสามารถผู้สอบ และการทดสอบนัยสำคัญทางสถิติด้วยค่าไคสแควร์ ที่ระดับนัยสำคัญ 0.05

6.คำนวณอัตราการตรวจสอบและอัตราความคลาดเคลื่อน

นำผลการวิเคราะห์จากการตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซลและวิธีถดถอยโลจิสติกมาคำนวณอัตราความถูกต้องในการตรวจสอบและอัตราความคลาดเคลื่อน เพื่อดูประสิทธิภาพของแต่ละวิธีโดยเปรียบเทียบกับวิธีทฤษฎีการตอบสนองข้อสอบ(IRT) ซึ่งจะใช้เกณฑ์ว่าข้อสอบที่ถูกตรวจพบการทำหน้าที่ต่างกันแบบอเนกรูปด้วยวิธี IRT ถือว่าเป็นข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ถ้าวิธี แมนเทล-แฮนส์เซลและวิธีถดถอยโลจิสติก ตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธี IRT ถือว่าวิธีนั้นมีความถูกต้องในการตรวจสอบ แต่ถ้าตรวจสอบพบว่าข้อสอบข้ออื่นทำหน้าที่ต่างกันนอกเหนือจากที่ตรวจสอบพบจากวิธี IRT ถือว่ามีความคลาดเคลื่อนในการตรวจสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การคำนวณอัตราความถูกต้องในการตรวจสอบและอัตราความคลาดเคลื่อนทำได้ดังนี้ (จิติมา วรรณศรี, 2540)

อัตราความถูกต้องประเภทที่ 1

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

อัตราความถูกต้องประเภทที่ 2

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

อัตราความคลาดเคลื่อนประเภทที่ 1

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

อัตราความคลาดเคลื่อนประเภทที่ 2

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

7. ทดสอบอัตราความถูกต้องในการตรวจสอบและอัตราความคลาดเคลื่อน ด้วยสถิติ Z

เป็นการทดสอบอัตราความถูกต้องในการตรวจสอบและอัตราความคลาดเคลื่อน ในการตรวจสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติก ด้วยสถิติ Z ที่ระดับนัยสำคัญ .05

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

P_1 ,P_2 เป็นสัดส่วนที่ได้จากการตรวจสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติก

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

Q เป็นค่าที่ได้จาก 1 - P

n_1 , n_2 จำนวนข้อสอบที่ตรวจพบด้วยวิธีแมนเทล-แฮนส์เซลและวิธีถดถอยโลจิสติก

8. สรุปผลการเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติก

เป็นการสรุปผลจากการเปรียบเทียบประสิทธิภาพระหว่างวิธี แมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติก หลังจากคำนวณอัตราความถูกต้องในการตรวจสอบ อัตราความคลาดเคลื่อน และทดสอบนัยสำคัญด้วยสถิติ Z เรียบร้อยแล้ว

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย