คุณสมบัติของตัวประมาณค่าความเข้มของอิทธิพล: การเปรียบเทียบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ

นายชยุตม์ ภิรมย์สมบัติ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรมหาบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2547

ISBN 974-53-2253-9

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

EFFECT MAGNITUDE ESTIMATORS' PROPERTIES: A COMPARISON BETWEEN CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY

Mr. Chayut Piromsombat

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Education in Educational Measurement and Evaluation
Department of Educational Research and Psychology
Faculty of Education
Chulalongkorn University
Academic Year 2004

ISBN 974-53-2253-9

หัวข้อวิทยานิพนธ์	คุณสมบัติของตัวประมาณค่าความเข้มของอิทธิพล: การเปรียบเทียบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ
โดย	นายชยุตม์ ภิรมย์สมบัติ
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.สุชาดา บวรกิติวงศ์

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.พฤทธิ์ ศิริบรรณพิทักษ์)

คณะกรรมการสอบวิทยานิพนธ์

……..…… ประธานกรรมการ
(ศาสตราจารย์กิตติคุณ ดร.นงลักษณ์ วิรัชชัย)

……..……... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา บวรกิติวงศ์)

……..…… กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.เอมอร จังศิริพรปกรณ์)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชยุตม์ ภิรมย์สมบัติ: คุณสมบัติของตัวประมาณค่าความเข้มของอิทธิพล: การเปรียบเทียบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ (EFFECT MAGNITUDE ESTIMATORS' PROPERTIES: A COMPARISON BETWEEN CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY) อ.ที่ปรึกษา: ผศ.ดร.สุชาดา บวรกิติวงศ์, 129 หน้า. ISBN 974-53-2253-9

การวิจัยครั้งนี้มีวัตถุประสงค์ 4 ข้อ ได้แก่ (1) เพื่อเปรียบเทียบคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (d_{IRT1}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (d_{IRT2}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (r_{IRT1}) และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (r_{IRT2}) (2) เพื่อเปรียบเทียบค่าเฉลี่ยระหว่าง d_{CTT}, d_{IRT1} และ d_{IRT2} และระหว่าง r_{CTT}, r_{IRT1} และ r_{IRT2} (3) เพื่อศึกษาความสัมพันธ์ระหว่าง d_{CTT} และ d_{IRT1} และสร้างสมการถดถอยของ d_{IRT1} บน d_{CTT} และ (4) เพื่อศึกษาความสัมพันธ์ระหว่าง r_{CTT} และ r_{IRT1} และสร้างสมการถดถอยของ r_{IRT1} บน r_{CTT} ภายใต้สถานการณ์การสอบ 540 สถานการณ์ ตามเงื่อนไขของค่าความเข้มของอิทธิพลที่แท้จริง (.2, .5, .8, 1.2, 2.6) ขนาดกลุ่มตัวอย่าง (20, 50, 500, 2,000) ความยาวแบบสอบ (10, 50, 90) โมเดลฐาน (โมเดลโลจิสติกแบบหนึ่ง, สอง และสามพารามิเตอร์) และโมเดลประมาณค่า (โมเดลการทดสอบแบบดั้งเดิม, โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูล และโมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูล)

ผลการวิจัยโดยสรุปพบว่า (1) ในภาพรวมตัวประมาณค่าที่มีความลำเอียงต่ำที่สุดคือ r_{IRT1} ตัวประมาณค่าที่มีความคงเส้นคงวาสูงสุดคือ r_{CTT} และตัวประมาณค่าที่มีประสิทธิภาพสัมพัทธ์สูงสุดคือ r_{IRT1} นอกจากนี้ r_{IRT1} ยังเป็นตัวประมาณค่าที่มีคุณสมบัติทุกด้านเป็นที่น่าพอใจที่สุด (2) ค่าเฉลี่ยของ d_{CTT}, d_{IRT1} และ d_{IRT2} มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดย d_{CTT} มีค่าเฉลี่ยสูงสุด เช่นเดียวกับค่าเฉลี่ยของ r_{CTT}, r_{IRT1} และ r_{IRT2} ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดย r_{CTT} มีค่าเฉลี่ยสูงสุด (3) สัมประสิทธิ์สหสัมพันธ์ระหว่าง d_{CTT} และ d_{IRT1} มีค่า .626 และมีนัยสำคัญทางสถิติที่ระดับ .05 สมการถดถอยของ d_{IRT1} บน d_{CTT} ในรูปคะแนนดิบ คือ $d_{IRT1} = .004 + .065d_{CTT}$ สมการในรูปคะแนนมาตรฐาน คือ $Z_{d_{IRT1}} = .626Z_{d_{CTT}}$ (4) สัมประสิทธิ์สหสัมพันธ์ระหว่าง r_{CTT} และ r_{IRT1} มีค่า .570 และมีนัยสำคัญทางสถิติที่ระดับ .05 สมการถดถอยของ r_{IRT1} บน r_{CTT} ในรูปคะแนนดิบ คือ $r_{IRT1} = .003 + .079r_{CTT}$ และสมการในรูปคะแนนมาตรฐาน คือ $Z_{r_{IRT1}} = .570Z_{r_{CTT}}$

ภาควิชา วิจัยและจิตวิทยาการศึกษา ลายมือชื่อนิสิต .....
สาขาวิชา การวัดและประเมินผลการศึกษา ลายมือชื่ออาจารย์ที่ปรึกษา .....
ปีการศึกษา 2547 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม -

4684218027: MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEYWORD: EFFECT MAGNITUDE / EFFECT SIZE / CLASSICAL TEST THEORY / ITEM RESPONSE THEORY / SIMULATION
CHAYUT PIROMSOMBAT: EFFECT MAGNITUDE ESTIMATORS' PROPERTIES: A COMPARISON BETWEEN CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY. THESIS ADVISOR: ASST. PROF. SUCHADA BOWARNKITIWONG, Ph.D., 129 pp. ISBN 974-53-2253-9

There were 4 objectives for this research: (1) to compare the estimators' properties in biasedness, consistency, and relative efficiency aspects among the effect size derived from Classical Test Theory (d_{CTT}), effect size derived from Item Response Theory which estimation models fit for the data (d_{IRT1}), effect size derived from Item Response Theory which estimation models unfit for the data (d_{IRT2}), correlation coefficient derived from Classical Test Theory (r_{CTT}), correlation coefficient derived from Item Response Theory which estimation models fit for the data (r_{IRT1}), and correlation coefficient derived from Item Response Theory which estimation models unfit for the data (r_{IRT2}); (2) to compare the means among d_{CTT}, d_{IRT1}, and d_{IRT2} as well as among r_{CTT}, r_{IRT1}, and r_{IRT2}; (3) to study the relationship between d_{CTT} and d_{IRT1} and express the regression equation of d_{IRT1} on d_{CTT}; (4) to study the relationship between r_{CTT} and r_{IRT1} and express the regression equation of r_{IRT1} on r_{CTT}. The 540 examination situations were built up from the conditions of the true effect magnitudes (.2, .5, .8, 1.2, 2.6), sample sizes (20, 50, 500, 2,000), test lengths (10, 50, 90), based models (one-, two-, and three-parameter logistic model), and estimation models (classical test model, item response models which fit for the data, item response models which unfit for the data).

The summarized findings were: (1) in the overview, the lowest biased estimator was r_{IRT1}, the highest consistency estimator was r_{CTT}, and the highest relative efficiency estimator was r_{IRT1}, in addition, r_{IRT1} was the most appropriate estimator for all properties; (2) the means of d_{CTT}, d_{IRT1}, and d_{IRT} were different at the .05 significance level, in fact, the mean of d_{CTT} was the highest, in the same way, the means of r_{CTT}, r_{IRT1}, and r_{IRT} were different at the .05 significance level and the mean of r_{CTT} was the highest; (3) the correlation coefficient between d_{CTT} and d_{IRT1} was .626 and significant at the .05 level, the regression of d_{IRT1} on d_{CTT} could be expressed by $d_{IRT1} = .004 + .065d_{CTT}$ for the raw scores and $Z_{d_{IRT1}} = .626Z_{d_{CTT}}$ for the standardized scores; (4) the correlation coefficient between r_{CTT} and r_{IRT1} was .570 and significant at the .05 level, the regression of r_{IRT1} on r_{CTT} could be expressed by $r_{IRT1} = .003 + .079r_{CTT}$ for the raw scores and $Z_{r_{IRT1}} = .570Z_{r_{CTT}}$ for the standardized scores.

Department Educational Research and Psychology Student's signature _____

Field of study Educational Measurement and Evaluation Advisor's signature _____

Academic year 2004 Co-advisor's signature - _____

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณ ผศ.ดร.สุชาดา บวรกิติวงศ์ อาจารย์ที่ปรึกษาวิชาการและอาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้อิสระทางความคิด และคำแนะนำที่เป็นประโยชน์อย่างมากในการทำวิทยานิพนธ์ด้วยดีเสมอมา

ขอบพระคุณ ศ.กิตติคุณ ดร.นงลักษณ์ วิรัชชัย และ ผศ.ดร.เอมอร จังศิริพรปกรณ์ กรรมการสอบวิทยานิพนธ์ ที่ชี้จุดบกพร่องและให้คำแนะนำอันมีค่าในการปรับปรุงวิทยานิพนธ์ในสมบูรณ์ยิ่งขึ้น รวมถึงความเมตตาในการจัดหาหนังสือที่เป็นประโยชน์ในการทำวิทยานิพนธ์เล่มนี้

ขอบพระคุณ รศ.พวงแก้ว ปุณยกนก ผศ.ดร.สุวัฒนา สุวรรณเขตนิคม และ ผศ.ดร.อวยพร เรืองตระกูล คณาจารย์ผู้สอนวิชา 2702782 สัมมนาการออกแบบการวิจัย ที่ให้คำแนะนำในการพัฒนาประเด็นที่สนใจศึกษาจนกลายเป็นหัวข้อวิทยานิพนธ์ที่ชัดเจนขึ้น

ขอบพระคุณ รศ.ดร.สุวิมล ว่องวานิช สำหรับกำลังใจ และความกรุณาในการจัดหาโปรแกรมพร้อมคู่มือการใช้งานในการวิเคราะห์ข้อมูลส่วนหนึ่ง

ขอบพระคุณ รศ.ดร.สุวัฒนา เอี่ยมอรพรรณ และ รศ.สุชาวดี เอี่ยมอรพรรณ สำหรับกำลังใจ และความเอาใจใส่ดูแลโดยการสอบถามความคืบหน้าในการศึกษาและการทำวิทยานิพนธ์เป็นระยะ

ผู้วิจัยรู้สึกซาบซึ้งและขอบคุณในน้ำใจของคุณวิเรขา ปัญจมานนท์ คุณมนต์ชัย ร้าวกระ-โทก และคุณมานพ พินิจพันธ์ ที่สละเวลาช่วยวิเคราะห์ข้อมูลส่วนหนึ่งตามคำสั่งคอมพิวเตอร์ที่ผู้วิจัยเขียนไว้ ตลอดจนเพื่อน รุ่นพี่ และรุ่นน้องภาควิชาวิจัยและจิตวิทยาการศึกษาที่ให้กำลังใจและแลกเปลี่ยนความรู้อันมีค่ากับผู้วิจัย ล้วนเป็นส่วนหนึ่งที่ทำให้วิทยานิพนธ์เล่มนี้สำเร็จได้ด้วยดี

ผู้วิจัยขอกราบขอบพระคุณ คุณพ่อสุเชษฐ์ ภิรมย์สมบัติ และคุณแม่ลมัย จันจินดา ที่เห็นความสำคัญของการศึกษาและสนับสนุนผู้วิจัยในทุก ๆ ด้านตลอดมา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญภาพ..... ฎ

บทที่

1 บทนำ

ความเป็นมาและความสำคัญของปัญหา..... 1

คำถามวิจัย..... 5

วัตถุประสงค์การวิจัย..... 5

สมมติฐานการวิจัย..... 6

ขอบเขตการวิจัย..... 7

นิยามศัพท์..... 8

ประโยชน์ที่ได้รับ..... 10

2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 ทฤษฎีการทดสอบแบบดั้งเดิม..... 12

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ..... 13

ตอนที่ 3 ความเข้มของอิทธิพล (effect magnitude)..... 17

ตอนที่ 4 คุณสมบัติของตัวประมาณค่าที่ดี..... 25

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง..... 27

3 วิธีดำเนินการวิจัย

ตอนที่ 1 เงื่อนไขที่ใช้ในการจำลองข้อมูล..... 39

ตอนที่ 2 การจำลองข้อมูล..... 42

ตอนที่ 3 การวิเคราะห์ข้อมูล..... 44

หน้า

4 ผลการวิเคราะห์ข้อมูล
ตอนที่ 1 ผลการตรวจสอบความเป็นเอกมิติของข้อมูล..... 48
ตอนที่ 2 ผลการตรวจสอบความถูกต้องของโปรแกรมคำนวณค่าประมาณ
ความเข้มของอิทธิพลที่ผู้วิจัยสร้างขึ้น..... 51
ตอนที่ 3 ผลการพิจารณาคุณสมบัติของตัวประมาณค่าในภาพรวม..... 52
ตอนที่ 4 ผลการพิจารณาคุณสมบัติของตัวประมาณค่ารายสถานการณ์…..... 62
ตอนที่ 5 ผลการวิเคราะห์ความสัมพันธ์และการถดถอย..... 93
ตอนที่ 6 สรุปผลการวิเคราะห์ตามวัตถุประสงค์การวิจัย..... 94
5 สรุป อภิปรายผลและข้อเสนอแนะ
สรุปผล..... 105
อภิปรายผล..... 106
ข้อเสนอแนะสำหรับการนำไปใช้..... 109
ข้อเสนอแนะสำหรับการวิจัยในอนาคต..... 111
รายการอ้างอิง..... 112
ภาคผนวก
ภาคผนวก ก 119
ภาคผนวก ข 120
ภาคผนวก ค 121
ภาคผนวก ง 122
ภาคผนวก จ 123
ภาคผนวก ฉ 126
ประวัติผู้เขียนวิทยานิพนธ์..... 129

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตาราง		หน้า
2.1	สถิติที่ใช้ประมาณค่าความเข้มของอิทธิพลจำแนกตามประเภทของตัวแปรวิจัย.....	22
2.2	สรุปวิธีการและเงื่อนไขที่ใช้ในการวิจัยที่เกี่ยวข้อง.....	36
3.1	สถานการณ์ที่ใช้ในการจำลองข้อมูล.....	41
4.1	ผลการตรวจสอบความเป็นเอกมิติของข้อมูลด้วยอัตราส่วนค่าไอเกน.....	48
4.2	ค่าเฉลี่ยและสัมประสิทธิ์สหสัมพันธ์ระหว่างของค่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ที่ได้จากโปรแกรมของผู้วิจัย กับการคำนวณด้วยโปรแกรม LISREL ร่วมกับ EXCEL.....	51
4.3	ค่าสถิติเบื้องต้นและสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าในภาพรวม.....	57
4.4	ผลการทดสอบความแปรปรวนทางเดียวของค่าประมาณความเข้มของอิทธิพลเฉลี่ย.....	61
4.5	ผลการทดสอบรายคู่ที่มีนัยสำคัญทางสถิติที่ระดับ .05.....	61
4.6	ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .2 และแบบสอบยาว 10 ข้อ.....	64
4.7	ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .2 และแบบสอบยาว 50 ข้อ.....	66
4.8	ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .2 และแบบสอบยาว 90 ข้อ.....	68
4.9	ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .5 และแบบสอบยาว 10 ข้อ.....	70
4.10	ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .5 และแบบสอบยาว 50 ข้อ.....	72

ตาราง **หน้า**

4.11 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .5 และแบบสอบยาว 90 ข้อ..... 74

4.12 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .8 และแบบสอบยาว 10 ข้อ..... 76

4.13 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .8 และแบบสอบยาว 50 ข้อ..... 78

4.14 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .8 และแบบสอบยาว 90 ข้อ..... 80

4.15 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 1.8 และแบบสอบยาว 10 ข้อ..... 82

4.16 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 1.8 และแบบสอบยาว 50 ข้อ..... 84

4.17 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 1.8 และแบบสอบยาว 90 ข้อ..... 86

4.18 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 2.6 และแบบสอบยาว 10 ข้อ..... 88

4.19 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 2.6 และแบบสอบยาว 50 ข้อ..... 90

4.20 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 2.6 และแบบสอบยาว 90 ข้อ..... 92

4.21 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความเข้มของอิทธิพลชนิดต่าง ๆ.. 93

4.22 ผลการทดสอบความสัมพันธ์เชิงเส้นตรงของตัวประมาณค่าความเข้มของอิทธิพล 93

4.23 ผลการวิเคราะห์การถดถอยของ d_{IRT} บน d_{CTT} และการถดถอยของ r_{IRT} บน r_{CTT}..... 94

4.24 ผลการเปรียบเทียบคุณสมบัติของตัวประมาณค่า 6 ชนิดในภาพรวมทุกสถานการณ์..... 95

4.25 สรุปผลการเปรียบเทียบคุณสมบัติของขนาดอิทธิพล (d) และสัมประสิทธิ์สหสัมพันธ์ (r) ในภาพรวมของแต่ละเงื่อนไข..... 96

4.26 ผลการเปรียบเทียบคุณสมบัติของ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1}, r_{IRT2} ในสถานการณ์ย่อย..... 98

4.27 ผลการเปรียบเทียบค่าเฉลี่ยของตัวประมาณค่าระหว่างทฤษฎีการทดสอบ 2 ทฤษฎี..... 103

4.28 ผลการวิเคราะห์ความสัมพันธ์และการถดถอย..... 103

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพประกอบ		หน้า
2.1	กรอบแนวคิดในการวิจัย.....	38
3.1	แผนภูมิสายงานแสดงขั้นตอนการจำลองข้อมูลและการวิเคราะห์ข้อมูล.....	47
4.1	ร้อยละของสถานการณ์ที่พบคุณสมบัติของตัวประมาณค่าที่ดีในตัวประมาณค่า 6 ชนิด.....	97

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ทฤษฎีการทดสอบเป็นผลสืบเนื่องมาจากการพัฒนาองค์ความรู้ด้านการทดสอบ เพื่อใช้เป็นองค์ความรู้ทั่วไปสำหรับประมาณค่าที่แท้จริงของคุณลักษณะภายในของบุคคลจากการตอบสนองสิ่งเร้าหรือพฤติกรรมที่สามารถสังเกตได้ (ศิริชัย กาญจนวาสี, 2545) ทฤษฎีการทดสอบที่สำคัญมี 3 ทฤษฎี ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดมีจุดเน้นแตกต่างจากทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบอย่างชัดเจนอยู่ประการหนึ่ง นั่นคือ การให้ความสำคัญกับการประมาณค่าความน่าเชื่อถือของเครื่องมือภายใต้สถานการณ์ต่าง ๆ ของการทดสอบ ในขณะที่ทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ ให้ความสำคัญกับการประมาณค่า พารามิเตอร์ความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบอันได้แก่ ความยาก อำนาจจำแนกและโอกาสในการเดา โดยจุดเน้นที่คล้ายกันของทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบนี้ทำให้มีการศึกษาเปรียบเทียบค่าพารามิเตอร์ของผู้สอบและข้อสอบที่ได้จากทฤษฎีการทดสอบทั้งสองนี้ อาทิงานวิจัยของ Ndalichako และ Rogers (1997), Fan (1998), Stage (1998), MacDonald และ Paunonen (2002), นภดล ยิ่งยงสกุล (2539), เบญจพร ยนตจักรวิถี (2539), วีระพันธ์ พรหมบุตร (2536) และ อรวรรณ สุขโต (2542) นอกจากนี้ยังมีการขยายกรอบแนวคิดของการเปรียบเทียบไปยังค่าสถิติอื่นที่ใช้ค่าพารามิเตอร์ของผู้สอบและ/หรือค่าพารามิเตอร์ของข้อสอบเป็นพื้นฐาน เช่น การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (จิติมา วรรณศรี, 2539; ชวลิต นิวาสวัต, 2541; นิคม กีรติวรางกูร, 2542) การเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ (อวยพร เรืองตระกูล, 2544) และการเปรียบเทียบค่าขนาดอิทธิพล (Wang & Chen, 2005) เป็นต้น การเปรียบเทียบค่าสถิติระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบดังกล่าวล้วนมีประโยชน์ทั้งในเชิงวิชาการและการประยุกต์ใช้ทฤษฎีการทดสอบให้เหมาะสมกับสถานการณ์ที่แตกต่างกัน

สำหรับการวิจัยเชิงปริมาณทางการศึกษา สารสนเทศจากการเปรียบเทียบค่าขนาดอิทธิพลที่ได้จากคะแนนความสามารถของผู้สอบตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ ของ Wang และ Chen (2004) เป็นจุดเริ่มต้นที่สำคัญและเป็นประโยชน์อย่างยิ่ง เนื่องจาก

การวิจัยเชิงปริมาณควรตอบคำถามพื้นฐานที่สำคัญ 3 ข้อ คือ 1) อิทธิพลหรือปรากฏการณ์ที่สนใจศึกษามีโอกาสเกิดขึ้นได้จริงหรือไม่ 2) ถ้ามีโอกาสเกิดขึ้นจริง อิทธิพลดังกล่าวมีปริมาณมากน้อยเพียงใด และ 3) อิทธิพลดังกล่าวมีปริมาณมากพอที่จะใช้ประโยชน์ได้จริงในทางปฏิบัติหรือไม่ นักวิจัยสามารถใช้การทดสอบนัยสำคัญทางสถิติ (statistical significance testing) เพื่อตอบคำถามแรกแต่ไม่สามารถใช้ในการตอบคำถามที่เหลือได้โดยตรง (Kirk, 2001; Paul & Plucker, 2004)

นอกจากนี้ Shaver (1993), Snyder และ Lawson (1993) และ Kirk (1996) ยังได้กล่าวถึงข้อจำกัดของการทดสอบนัยสำคัญทางสถิติไว้สอดคล้องกัน 3 ประการ **ประการแรก** การทดสอบนัยสำคัญทางสถิติไม่ได้บอกถึงความน่าจะเป็นที่สมมติฐานศูนย์จะเป็นจริงเมื่อมีข้อมูลที่สังเกตได้ชุดหนึ่ง หรือ $p(H_0|\text{data set})$ แต่ผลที่ได้จากการทดสอบนัยสำคัญทางสถิติคือความน่าจะเป็นที่จะได้ข้อมูลที่สังเกตได้นี้หรือชุดที่ต่างจากนี้มาก ๆ เมื่อสมมติฐานศูนย์เป็นจริง หรือ $p(\text{data set}|H_0)$ ซึ่งผลสรุปอาจต่างกันได้ **ประการที่สอง** การปฏิเสธสมมติฐานศูนย์ไม่ได้ให้สารสนเทศเกี่ยวกับความเข้มของอิทธิพล หรือประโยชน์ของอิทธิพลในเชิงปฏิบัติ **ประการที่สาม** การที่นักวิจัยกำหนดระดับนัยสำคัญสำหรับการตัดสินใจ (เช่น กำหนด $\alpha = .05$) เป็นการเปลี่ยนคุณลักษณะของค่าความน่าจะเป็นที่เดิมมีความต่อเนื่องให้กลายเป็นทวิภาค (ปฏิเสธหรือไม่ปฏิเสธ) การตัดสินใจเช่นนี้จะเกิดความคลาดเคลื่อนได้ เช่น กรณีที่มีนักวิจัยสองคนศึกษาตัวแปรตัวเดียวกัน แต่กำหนดระดับนัยสำคัญสำหรับการตัดสินใจต่างกัน ผลได้สรุปย่อมต่างกัน นอกจากนี้ Craig, Eison และ Metze (1976 อ้างถึงใน Shaver, 1993) และ Snyder และ Lawson (1993) ยังกล่าวว่า ขนาดของกลุ่มตัวอย่างส่งผลอย่างยิ่งต่อการทดสอบนัยสำคัญทางสถิติ ถ้านักวิจัยใช้ขนาดกลุ่มตัวอย่างที่มากพอการได้ผลการทดสอบว่ามีนัยสำคัญทางสถิติอาจเกิดขึ้นได้แม้ตัวแปรที่ศึกษาจะมีความสัมพันธ์กันน้อยมาก ข้อวิพากษ์เหล่านี้ทำให้เกิดกระแสการพัฒนาวิธีการทางสถิติ เพื่อเติมเต็มข้อจำกัดข้างต้นของการทดสอบนัยสำคัญทางสถิติ

เนื่องจากอิทธิพลที่ต้องการศึกษาเป็นอิทธิพลที่เกิดขึ้นในกลุ่มประชากร ผู้วิจัยจึงใช้คำว่า “การประมาณค่าความเข้มของอิทธิพล” แทน “การวัดความเข้มของอิทธิพล” โดยพบว่าคำศัพท์ภาษาอังกฤษที่หมายถึงการประมาณค่าอิทธิพลที่ปรากฏในรายงานวิจัยและเอกสารทางวิชาการนั้นมีหลายคำ ดังที่ Snyder และ Lawson (1993) และ Ives (2003) ได้สำรวจไว้ เช่น estimates of magnitude of the effect, estimates of explained variance, effect size estimates, estimates of the strength of relation, estimates of proportion of variance accounted for, measure of association เป็นต้น จะเห็นได้ว่าในภาษาอังกฤษมีทั้งการใช้คำว่า effect magnitude (Hedges & Olkin, 1985; Snyder & Lawson, 1993; Kirk, 2001; Gliner, Leech & Morgan, 2002) และคำว่า effect size (Wilkinson & APA Task Force on Statistical Inference, 1999; Huberty, 2002; Trusty, Thompson & Petrocelli, 2004) นอกจากนี้ Shaver (1993) ยังได้เสนอคำว่า result size แต่ไม่ได้รับความนิยม โดย Shaver ได้กล่าวไว้ว่านักวิจัยทางสังคมศาสตร์นิยมใช้คำว่า effect size เกินกว่าจะเปลี่ยนแปลง

ได้ ปัจจุบันจึงพบว่านักวิจัยส่วนใหญ่ใช้คำว่า effect size ในความหมายเดียวกับ effect magnitude แต่ในงานวิจัยครั้งนี้ผู้วิจัยเลือกใช้คำว่า effect magnitude ในความหมายของความเข้มของอิทธิพลที่เกิดขึ้นในงานวิจัย และใช้คำว่า effect size ในความหมายของผลต่างมาตรฐานตามนิยามที่ Cohen ได้เริ่มใช้เป็นคนแรก (นงลักษณ์ วิรัชชัย, 2542) เนื่องจากยังมีสถิติอีกมากที่ไม่ได้อยู่ในรูปของผลต่างมาตรฐานแต่สามารถใช้ประมาณค่าความเข้มของอิทธิพลได้ การใช้คำว่า effect magnitude จึงน่าจะเหมาะสมกว่าคำว่า effect size

นอกจากขนาดอิทธิพล (Cohen's d) แล้วยังมีสถิติที่สามารถใช้ประมาณค่าความเข้มของอิทธิพลได้อีกประมาณ 40 ชนิด (Kirk, 1996) โดยนักสถิติหลายท่านได้เสนอความเห็นเกี่ยวกับการประมาณค่าความเข้มของอิทธิพลที่เหมาะสมสำหรับงานวิจัยเชิงปริมาณทั่ว ๆ ไปเอาไว้ เช่น Baugh (2002), Thompson (2002) และ Ives (2003) กล่าวสอดคล้องกันว่าการประมาณค่าความเข้มของอิทธิพลควรใช้สถิติในกลุ่มความสัมพันธ์หรือความแปรปรวนที่ถูกอธิบาย เพราะสามารถนำมาใช้ได้กับการวิเคราะห์เชิงปริมาณที่ใช้โมเดลเชิงเส้นทั่วไป ทั้งยังสามารถประยุกต์ใช้ได้กับงานวิจัยเชิงปริมาณทั้งที่เป็นเชิงทดลองและไม่ใช่เชิงทดลอง เนื่องจากเป้าหมายหนึ่งของการวิจัยเชิงปริมาณส่วนใหญ่คือต้องการศึกษาความสัมพันธ์หรือความผันแปรร่วมกันระหว่างตัวแปรนั่นเอง (ศิริชัย กาญจนวาสี, 2541; Thompson, 2000)

นอกจากนี้ Rosenthal และ DiMatteo (2001) ยังกล่าวไว้ว่าสถิติในกลุ่มความสัมพันธ์อย่างสัมประสิทธิ์สหสัมพันธ์มีข้อดีเหนือกว่าขนาดอิทธิพลบางประการ คือ (1) การแปลงขนาดอิทธิพลให้อยู่ในรูปของสัมประสิทธิ์สหสัมพันธ์มีความสมเหตุสมผล เนื่องจากสัมประสิทธิ์สหสัมพันธ์แบบพอยต์ไบซีเรียล (point biserial) สามารถแสดงความสัมพันธ์ระหว่างตัวแปรต้นที่มีสองระดับกับตัวแปรตามที่มีค่าต่อเนื่องได้ ในขณะที่การแปลงสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันซึ่งตัวแปรมีค่าต่อเนื่องให้อยู่ในรูปของขนาดอิทธิพลซึ่งใช้ข้อมูลทวิภาคจะทำให้เสียสารสนเทศไป (2) เมื่อใช้ contrast ที่มีองศาอิสระเป็น 1 สัมประสิทธิ์สหสัมพันธ์สามารถวิเคราะห์แนวโน้มระหว่างกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป ในขณะที่ขนาดอิทธิพลมีข้อจำกัดอยู่ที่ 2 กลุ่ม (3) ไม่จำเป็นต้องปรับแก้วิธีการคำนวณสัมประสิทธิ์สหสัมพันธ์เมื่อใช้กลุ่มตัวอย่างมากกว่า 2 กลุ่ม และ (4) สัมประสิทธิ์สหสัมพันธ์แปลความหมายง่ายกว่าขนาดอิทธิพล อย่างไรก็ตาม การเลือกใช้ตัวประมาณค่ายังต้องคำนึงถึงคุณสมบัติของตัวประมาณค่า อันได้แก่ ความลำเอียง ความคงเส้นคงวา ความพอเพียง และประสิทธิภาพสัมพัทธ์ด้วย (Mandenhall & Beaver, 1994; Glass & Hopkins, 1995)

ด้วยเหตุที่สัมประสิทธิ์สหสัมพันธ์ และขนาดอิทธิพลเป็นสถิติที่นิยมใช้ทั่วไปในงานวิจัยเชิงปริมาณรวมถึงการวิเคราะห์อภิมาน และเพื่อเป็นการขยายองค์ความรู้ต่อจากงานวิจัยของ Wang และ Chen (2004) ผู้วิจัยจึงสนใจที่จะเปรียบเทียบคุณสมบัติของตัวประมาณค่าที่ดีระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ

สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ

การเปรียบเทียบค่าสถิติที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ อาทิ การเปรียบเทียบค่าความสามารถของผู้สอบที่ได้จากทั้งสองทฤษฎี นิยมใช้ข้อมูลจากการจำลองตามสถานการณ์การสอบต่าง ๆ (Fan, 1998; DeMars, 2001; MacDonald & Paunonen, 2002; Dawber, Rogers & Carbonaro, 2004; Wang & Chen, 2004) นอกจากนี้ Harwell และคณะ (1996) ยังกล่าวไว้ว่า การเปรียบเทียบตัวประมาณค่าในสถานการณ์ต่าง ๆ ควรใช้การจำลองข้อมูลจึงจะเหมาะสม ผู้วิจัยจึงเลือกใช้การจำลองข้อมูลเป็นแนวทางในการวิจัยครั้งนี้

สำหรับเงื่อนไขในการเปรียบเทียบค่าสถิติที่ได้ทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบที่สำรวจจากงานวิจัยที่เกี่ยวข้อง พบว่า มีการใช้เงื่อนไขที่แตกต่างกันตามเป้าหมายของการเปรียบเทียบ โดยเงื่อนไขที่พบมาก ได้แก่ ความยาวข้อสอบ (Fan, 1998; DeMars, 2001; MacDonald & Paunonen, 2002; Dawber, Rogers & Carbonaro, 2004) และขนาดกลุ่มผู้สอบ (DeMars, 2001; Roberts & Henson, 2002; Stone & Yumoto, 2004; Wang & Chen, 2004) ทั้งสองเงื่อนไขนี้สอดคล้องกับผลการสำรวจของ Harwell และคณะ (1996) ซึ่งยังพบเงื่อนไขอื่นเพิ่มเติมอีก เช่น ชนิดของโมเดลการตอบสนองข้อสอบ (ทั้งที่เป็นโมเดลฐานในการคำนวณความน่าจะเป็นสำหรับการจำลองข้อมูลคำตอบ และโมเดลที่ใช้ประมาณค่าความสามารถของผู้สอบ) การกระจายของค่าความยาก อำนาจจำแนก และโอกาสในการเดา เป็นต้น นอกจากนี้ Hambleton และ Swaminathan (1985), Embretson และ Reise (2000) และศิริชัย กาญจนวาสี (2545) กล่าวไว้สอดคล้องกันว่า โมเดลการวัดที่ไม่สอดคล้องกับข้อมูลย่อมทำให้ผลการวิเคราะห์ขาดความถูกต้องน่าเชื่อถือ ความสอดคล้องของโมเดลกับข้อมูลจึงเป็นอีกประเด็นที่น่าสนใจ ผู้วิจัยประยุกต์แนวคิดของ DeMars (2001) ซึ่งกำหนดโมเดลที่ใช้ประมาณค่าความสามารถของผู้สอบมี 2 โมเดล โมเดลแรกเป็นโมเดลเดียวกับโมเดลฐานที่ใช้ในการจำลองคำตอบของผู้สอบ แทนสถานการณ์ที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และโมเดลที่สองเป็นโมเดลที่ต่างจากโมเดลฐาน แทนสถานการณ์ที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล

ในงานวิจัยครั้งนี้ผู้วิจัยจึงเห็นว่าควรใช้ ความยาวแบบสอบ ขนาดกลุ่มตัวอย่าง โมเดลฐาน และโมเดลประมาณค่า เป็นเงื่อนไขในการจำลองข้อมูล รวมทั้ง ความเข้มของอิทธิพลที่แท้จริง ซึ่งเป็นอีกเงื่อนไขที่สำคัญในการเปรียบเทียบคุณสมบัติของตัวประมาณค่าความเข้มของอิทธิพล (Roberts & Henson, 2002; Wang & Chen, 2004) โดยผู้วิจัยจะใช้ค่าความยาก อำนาจจำแนก และโอกาสในการเดาที่สุ่มได้จากการกระจายที่เหมาะสมกับค่าทั้งสามตามที่ Pelton (2002) ได้สรุปไว้แทนการกำหนดให้การกระจายของค่าทั้งสามเป็นเงื่อนไขในการจำลองข้อมูล

การวิจัยครั้งนี้จึงเป็นการเปรียบเทียบคุณสมบัติของตัวประมาณค่าระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ

สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ โดยศึกษาจากข้อมูลที่จำลองขึ้นตาม 5 เงื่อนไขหลัก ได้แก่ ความเข้มของอิทธิพลที่แท้จริง ความยาวแบบสอบ ขนาดกลุ่มตัวอย่าง โมเดลฐาน และโมเดลประมาณค่า

คำถามวิจัย

เนื่องจากงานวิจัยของ Stage (1998) ได้แสดงให้เห็นว่าค่าสถิติ อาทิ ความสามารถของผู้สอบ ความยาก อำนาจจำแนก และโอกาสในการเดา ที่ได้จากการวิเคราะห์ข้อมูลตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบมีความสัมพันธ์กันสูง และผลการวิจัยที่ Wang และ Chen (2004) ได้ทำการเปรียบเทียบค่าขนาดอิทธิพลที่ได้จากความสามารถของผู้สอบตามโมเดลโลจิสติกแบบสองพารามิเตอร์ (two-parameter logistic model) และค่าขนาดอิทธิพลที่ได้จากคะแนนจริงตามทฤษฎีการทดสอบแบบดั้งเดิม พบว่าให้ค่าขนาดอิทธิพลไม่ต่างกันมากนัก (.731 และ .756) ประกอบกับ Rosenthal และ DiMatteo (2001), Baugh (2002), Thompson (2002) และ Ives (2003) ที่เสนอว่า สัมประสิทธิ์สหสัมพันธ์มีความเหมาะสมในการประมาณค่าความเข้มของอิทธิพลไม่ด้อยไปกว่าขนาดอิทธิพล ผู้วิจัยจึงเกิดข้อคำถามเกี่ยวกับการประมาณค่าความเข้มของอิทธิพลด้วยตัวประมาณค่าที่ต่างกันและใช้ค่าความสามารถของผู้สอบจากทฤษฎีที่แตกต่างกัน โดยแยกเป็นประเด็นคำถามวิจัยไว้ดังนี้

1. ขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ มีคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ทัดเทียมกันหรือไม่

2. ขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ จะประมาณค่าความเข้มของอิทธิพลแตกต่างกันหรือไม่ มีความสัมพันธ์กันหรือไม่ อย่างไร

3. สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ จะประมาณค่าความเข้มของอิทธิพลแตกต่างกันหรือไม่ มีความสัมพันธ์กันหรือไม่ อย่างไร

วัตถุประสงค์การวิจัย

การวิจัยในครั้งนี้เป็นการวิจัยเชิงทดลอง มีวัตถุประสงค์การวิจัยดังนี้

1. เพื่อเปรียบเทียบคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม

สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ ซึ่งแบ่งย่อยเป็นขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ ซึ่งแบ่งย่อยเป็นสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล

2. เพื่อเปรียบเทียบค่าเฉลี่ยระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูลและระหว่างสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล

3. เพื่อศึกษาความสัมพันธ์ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ และสร้างสมการถดถอยของขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบบนขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม

4. เพื่อศึกษาความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ และสร้างสมการถดถอยของสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบบนสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม

สมมติฐานการวิจัย

แม้ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะเป็นตัวประมาณค่าที่มีความลำเอียง แต่ความลำเอียงของขนาดอิทธิพลมีค่าเพียงเล็กน้อย และขนาดอิทธิพลที่คำนวณตามสูตรของ Cohen โดยตรงมีค่าสูงกว่าขนาดอิทธิพลที่ปรับแก้ตามวิธีของ Thompson ซึ่งใช้การแปลงขนาดอิทธิพลเป็นสัมประสิทธิ์สหสัมพันธ์ก่อน (Roberts & Henson, 2002) นอกจากนี้ขนาดอิทธิพลเหมาะสำหรับการวิจัยที่ต้องการเปรียบเทียบความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม อาจไม่เหมาะสมกับการใช้ประมาณค่าความเข้มของอิทธิพลในการวิจัยเชิงปริมาณแบบอื่น ๆ ซึ่งการประมาณค่าความเข้มของอิทธิพลในงานวิจัยเชิงปริมาณทั่วไปควรใช้สถิติในกลุ่มความสัมพันธ์ (Rosenthal & DiMatteo, 2001; Baugh, 2002; Thompson, 2002; Ives, 2003) จากผลการวิจัยและแนวคิดเกี่ยวกับการประมาณค่าความเข้มของอิทธิพลด้วยตัวประมาณค่าต่างชนิดกันนั้น จะเห็นได้ว่าตัวประมาณค่าต่างชนิดกัน

น่าจะให้ค่าประมาณความเข้มของอิทธิพลต่างกัน นอกจากนี้ผลการวิจัยของ DeMars (2001) ชี้ให้เห็นว่าโมเดลโลจิสติกแบบสอง และสามพารามิเตอร์ให้ค่าขนาดอิทธิพลต่างกัน ด้วยเหตุที่ค่าประมาณความเข้มของอิทธิพลที่ได้จากตัวประมาณค่าต่างชนิดกันมีค่าต่างกันนี้เอง ที่ส่งผลให้ความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ซึ่งเป็นคุณสมบัติของตัวประมาณมีความแตกต่างกันด้วย ดังผลการวิจัยของ Roberts และ Henson (2002) ที่กล่าวไว้แล้วในข้างต้นรวมถึงผลการวิจัยของ Wang และ Chen (2004) ที่พบว่าความลำเอียงของขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบแตกต่างกันเล็กน้อย ผลการวิจัยทั้งสองชี้ให้เห็นว่าตัวประมาณค่าต่างชนิดกันมีคุณสมบัติในการประมาณค่าที่แตกต่างกัน ด้วยเหตุนี้ผู้วิจัยจึงตั้งสมมติฐานสำหรับคำถามวิจัยข้อ 1 ว่า

1. ขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ มีคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ แตกต่างกัน

นอกจากนี้ผลการศึกษาของ Fan (1998), Stage (1998; 2003), MacDonald และ Paunonen (2002) และ Wang และ Chen (2004) ที่พบว่าค่าความสามารถของผู้สอบ ความยาก และอำนาจจำแนกที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบมีค่าแตกต่างกัน แต่โดยมากจะมีความสัมพันธ์กันในทางบวกสูง ผู้วิจัยจึงตั้งสมมติฐานสำหรับคำถามวิจัยข้อ 2-4 ว่า

2. ค่าเฉลี่ยของขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ แตกต่างกัน

3. ขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ มีความสัมพันธ์กันในทางบวก

4. สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ มีความสัมพันธ์กันในทางบวก

ขอบเขตการวิจัย

1. การวิจัยในครั้งนี้มีขอบเขตการศึกษาเฉพาะ การประมาณค่าความเข้มของอิทธิพลแบบจุด (point estimation for effect magnitude) ในงานวิจัยเชิงปริมาณทางการศึกษาที่ศึกษาอิทธิพลเจาะจง (fixed effect) ของตัวแปรต้นหรือตัวแปรจัดกระทำที่มีต่อตัวแปรความสามารถของบุคคล (person's ability) ซึ่งวัดด้วยข้อสอบแบบหลายตัวเลือก (multiple-choice item) มีการตรวจให้คะแนนแบบทวิภาค (dichotomous scoring)

2. โมเดลการวัดที่ใช้ในการวิจัยครั้งนี้เป็นโมเดลการตอบสนองข้อสอบในกลุ่มของโมเดลการตรวจให้คะแนนแบบทวิภาค 3 โมเดล ได้แก่ โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (one-parameter logistic model) โมเดลโลจิสติกแบบสองพารามิเตอร์ (two-parameter logistic model) และโมเดลโลจิสติกแบบสามพารามิเตอร์ (three-parameter logistic model)

3. สถานการณ์การสอบที่จำลองขึ้นในงานวิจัยนี้เป็นไปตามเงื่อนไขหรือตัวแปรต้น 5 ตัว ได้แก่ ค่าความเข้มของอิทธิพลที่แท้จริง ความยาวแบบสอบ ขนาดกลุ่มตัวอย่าง โมเดลฐานและโมเดลประมาณค่า ในแต่ละสถานการณ์จะทำซ้ำ (replicate) ทั้งหมด 100 รอบ โดยมีคุณสมบัติของตัวประมาณค่าที่ดีในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์เป็นตัวแปรตาม

นิยามศัพท์

ค่าความเข้มของอิทธิพล หมายถึง ตัวเลขที่ให้สารสนเทศเกี่ยวกับขนาดของผลอันเกิดจากอิทธิพลของตัวแปรจัดกระทำที่ทำให้เกิดการเปลี่ยนแปลงในตัวแปรตาม หรือตัวเลขที่ให้สารสนเทศเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร ในงานวิจัยนี้ค่าความเข้มของอิทธิพลจะประมาณค่าได้จากขนาดอิทธิพล (d) และสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล (r) ตามสูตรต่อไปนี้

$d = \dfrac{\overline{X}_E - \overline{X}_C}{S_{pooled}}$ (Kirk, 1996) และ $r = \dfrac{\overline{X}_E - \overline{X}_C}{S_{total}} \times \dfrac{pq}{Y}$ (Guilford & Fruchter, 1981)

โดยที่ $\overline{X}_E$ แทนค่าเฉลี่ยของตัวแปร X ในกลุ่มทดลอง

$\overline{X}_C$ แทนค่าเฉลี่ยของตัวแปร X ในกลุ่มควบคุม

S_{pooled} แทนส่วนเบี่ยงเบนมาตรฐานรวมของกลุ่มทดลองและกลุ่มควบคุม

S_{total} แทนส่วนเบี่ยงเบนมาตรฐานรวมของกลุ่มตัวอย่างทั้งหมด

p แทนอัตราส่วนของขนาดกลุ่มทดลองต่อขนาดกลุ่มตัวอย่างทั้งหมด

q แทนอัตราส่วนของขนาดกลุ่มควบคุมต่อขนาดกลุ่มตัวอย่างทั้งหมด

Y แทนค่าระยะตั้งฉากที่แบ่งพื้นที่ใต้โค้งปกติออกเป็น p และ q ในที่นี้ Y = .3989

อนึ่ง ในงานวิจัยครั้งนี้จะใช้คำว่า สัมประสิทธิ์สหสัมพันธ์ (r) แทนสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียลที่ใช้เป็นตัวประมาณค่าความเข้มของอิทธิพล

คุณสมบัติของตัวประมาณค่า หมายถึง ลักษณะของตัวประมาณค่าหรือสถิติที่ใช้ประมาณค่าพารามิเตอร์ในประชากร ซึ่งตัวประมาณค่าที่ดีควรจะไม่มีความลำเอียง (unbiased) หรืออย่างน้อยที่สุดคือมีความลำเอียงต่ำเมื่อเทียบกับตัวประมาณค่าอื่น นอกจากนี้ยังควรมีความคงเส้นคงวาสูง (consistency) มีความพอเพียง (sufficiency) ในการใช้ข้อมูลจากทุกหน่วยตัวอย่างมาใช้ประมาณค่า และมีประสิทธิภาพสัมพัทธ์สูง (relative efficiency) ในงานวิจัยครั้งนี้จะพิจารณาคุณสมบัติของตัว

ประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ ตามลำดับ สำหรับความพอเพียง ซึ่งเป็นคุณสมบัติของตัวประมาณค่าที่ดีอีกประการหนึ่งนั้นผู้วิจัยไม่ได้นำมาพิจารณา เนื่องจากตัวประมาณค่าทุกตัวใช้ข้อมูลจากผู้สอบทั้งหมดในการคำนวณอยู่แล้ว

ความลำเอียง หมายถึง คุณสมบัติของตัวประมาณค่าที่มีค่าเฉลี่ยของการแจกแจงการสุ่มของตัวประมาณค่าต่างจากค่าพารามิเตอร์ ในงานวิจัยนี้ ความลำเอียงในการประมาณค่าความเข้มของอิทธิพลวัดได้จากค่าความแตกต่างระหว่างค่าคาดหวังของตัวประมาณค่ากับค่าที่แท้จริง (expected - parameter) หรือที่เรียกว่าค่าความลำเอียง (BIAS) และส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Standardized Root Mean Square Deviation: SRMSD) ตามสูตรต่อไปนี้

$$BIAS = E(\hat{a}) - a$$ (Hay, 1963 อ้างถึงในสุกัญญรัตน์ คงงาม, 2539)

และ $$SRMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(Standardized\ BIAS\right)^2}{N}}$$ (ดัดแปลงจาก Pelton, 2004)

หรือ $$SRMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{\hat{a}_i - a}{SD_{\hat{a}_i - a}}\right)^2}{N}}$$

โดยที่ a หมายถึง พารามิเตอร์ที่ต้องการประมาณค่า

$\hat{a}$ หมายถึง ค่าประมาณของพารามิเตอร์

$E(\hat{a})$ หมายถึง ค่าคาดหวังของ $\hat{a}$ หรือค่าเฉลี่ยของการแจกแจงการสุ่มของ $\hat{a}$

$SD_{\hat{a}_i - a}$ หมายถึง ส่วนเบี่ยงเบนมาตรฐานของค่าความลำเอียง (BIAS)

i หมายถึง ครั้งที่ประมาณค่า

N หมายถึง จำนวนครั้งในการประมาณค่า

เกณฑ์ในการพิจารณาความลำเอียงเป็นเกณฑ์สัมพัทธ์ กล่าวคือ ตัวประมาณค่าที่ดีควรมีความลำเอียงต่ำเมื่อเทียบกับตัวประมาณค่าชนิดอื่น นั่นคือ มีค่า SRMSD ต่ำที่สุด สำหรับค่า BIAS จะใช้ในการพิจารณาว่าตัวประมาณค่าให้ค่าสูงหรือต่ำกว่าค่าจริง (overestimate or underestimate) เท่านั้น กล่าวคือ หากค่า BIAS เป็นลบแสดงว่าตัวประมาณค่าให้ค่าที่ต่ำกว่าค่าจริง แต่ถ้าค่า BIAS เป็นบวกแสดงว่าตัวประมาณค่าให้ค่าที่สูงกว่าค่าจริง

ความคงเส้นคงวา หมายถึง คุณลักษณะของตัวประมาณค่าที่มีแนวโน้มในการให้ค่าประมาณเข้าใกล้ค่าพารามิเตอร์เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น (an estimator tends to get closer to the value of the parameter as the sample size becomes larger) หรือเมื่อขนาดกลุ่มตัวอย่างลู่เข้าสู่

อนันต์ ($N\rightarrow\infty$) ในการวิจัยครั้งนี้ ความคงเส้นคงวาของตัวประมาณค่าความเข้มของอิทธิพลจะคำนวณด้วยสูตรที่ปรับจาก สุกัญญรัตน์ คงงาม (2539) ดังต่อไปนี้

$$\Delta = SRMSD_{Nmin} - SRMSD_{Nmax}$$

โดยที่ Δ หมายถึง ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน

$SRMSD_{Nmin}$ หมายถึง ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานในกรณีที่กลุ่มตัวอย่างมีขนาดเล็กที่สุด (ในการวิจัยครั้งนี้ Nmin มีค่า 20)

$SRMSD_{Nmax}$ หมายถึง ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ที่สุด (ในการวิจัยครั้งนี้ Nmax มีค่า 2000)

เกณฑ์ในการพิจารณาเป็นเกณฑ์สัมพัทธ์ กล่าวคือ ตัวประมาณค่าที่ดีควรมีความคงเส้นคงวาสูง นั่นคือ มีค่า Δ สูงกว่าตัวประมาณค่าอื่น

ประสิทธิภาพสัมพัทธ์ หมายถึง คุณสมบัติของตัวประมาณค่าในด้านความถูกต้องสัมพัทธ์ (relative precision) ของการประมาณค่า หรือระดับของความคลาดเคลื่อนของการสุ่มเมื่อเทียบกับตัวประมาณค่าชนิดอื่นเมื่อใช้กลุ่มตัวอย่างขนาดเท่ากัน (Glass & Hopkins, 1995) ในงานวิจัยครั้งนี้ประสิทธิภาพสัมพัทธ์พิจารณาจากผลการเปรียบเทียบความแปรปรวนของค่าประมาณความเข้มของอิทธิพลที่ได้จากตัวประมาณค่าแต่ละชนิดในสถานการณ์เดียวกัน ในการพิจารณาจะใช้เกณฑ์สัมพัทธ์ กล่าวคือ ตัวประมาณค่าที่ดีควรมีประสิทธิภาพสัมพัทธ์สูง นั่นคือ มีค่าความแปรปรวนต่ำกว่าตัวประมาณค่าอื่น

ประโยชน์ที่ได้รับ

1. ประโยชน์เชิงวิชาการ

1.1 ผลการวิจัยให้สารสนเทศเกี่ยวกับความสัมพันธ์ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมกับขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ และความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมกับสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ ซึ่งจะเชื่อมโยงทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบในมิติของความเข้มของอิทธิพล ทำให้ได้สมการถดถอยที่สามารถใช้ในการแปลงค่าขนาดอิทธิพลหรือสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมเป็นค่าขนาดอิทธิพลหรือสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบได้

1.2 การวิจัยครั้งนี้เป็นการบูรณาการความรู้ในศาสตร์ด้านการวัดผลทางการศึกษา สถิติและการวิจัยทางการศึกษา ซึ่งจะทำให้นักวิจัยทางการศึกษาสามารถออกแบบการวัดตัวแปรและการวิเคราะห์ข้อมูล เพื่อศึกษาอิทธิพลในงานวิจัยด้วยตัวประมาณค่าความเข้มของอิทธิพลได้อย่าง

เหมาะสมกับทรัพยากรที่มี เช่น ในกรณีที่นักวิจัยมีกลุ่มตัวอย่างขนาดเล็ก หรือไม่มีโปรแกรมประมาณค่าความสามารถของผู้สอบตามทฤษฎีการตอบสนองข้อสอบ อาทิ BILOG แต่นักวิจัยต้องการทำการวิจัยเชิงทดลองโดยมีตัวแปรตามที่วัดค่าได้จากแบบสอบ นักวิจัยสามารถเลือกใช้ตัวประมาณค่าความเข้มที่เหมาะสมกับกลุ่มตัวอย่างขนาดเล็กและเหมาะสมกับทฤษฎีการทดสอบแบบดั้งเดิม อย่างเช่น สัมประสิทธิ์สหสัมพันธ์ เพื่อศึกษาความเข้มของอิทธิพลในงานวิจัยที่ต้องการได้ ทั้งยังสามารถแปลงค่าที่ได้ให้อยู่ในรูปของสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ ขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม หรือขนาดอิทธิพลจากทฤษฎีการตอบสนองข้อสอบ ตามวัตถุประสงค์ในการนำไปใช้ของนักวิจัยได้

1.3 การวิจัยครั้งนี้ใช้วิธีการจำลองสถานการณ์ในการศึกษา ช่วยให้สามารถศึกษาได้ทุก ๆ กรณีที่เป็นไป ซึ่งนักวิจัยทางการศึกษาหรืออาจารย์ผู้สอนวิชาวิจัยสามารถใช้ระเบียบวิธีวิจัยในครั้งนี้เป็นตัวอย่างงานวิจัยที่สามารถให้ผลการศึกษาที่ครอบคลุมสถานการณ์ที่เป็นไปได้ทั้งหมด

2. ประโยชน์ด้านการนำไปใช้

2.1 ผลการวิจัยในครั้งนี้ให้สารสนเทศเกี่ยวกับความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เพื่อใช้ประกอบการตัดสินใจเลือกใช้ตัวประมาณค่าที่มีประสิทธิภาพ ถูกต้องและเหมาะสมแล้วแต่กรณี สามารถนำไปใช้ได้ในงานวิจัยเชิงปริมาณทางการศึกษา ทั้งที่ประมาณค่าความสามารถของผู้สอบตามตามทฤษฎีการทดสอบแบบดั้งเดิม หรือทฤษฎีการตอบสนองข้อสอบ

2.2 ผลการวิจัยให้สารสนเทศเกี่ยวกับความสัมพันธ์ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมกับขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ และความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมกับสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ เป็นสารสนเทศที่ช่วยสนับสนุนให้การวิจัยเชิงปริมาณทางการศึกษาที่มีข้อจำกัดเกี่ยวกับขนาดของกลุ่มผู้สอบ สามารถศึกษาอิทธิพลได้จากการวิเคราะห์ข้อมูลที่วัดตามทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งไม่ต้องใช้จำนวนกลุ่มผู้สอบขนาดใหญ่ได้อย่างสะดวก โดยผลการวิจัยยังคงสอดคล้องหรือมีนัยยะไม่ต่างจากการวิเคราะห์ด้วยค่าความสามารถของผู้สอบตามทฤษฎีการตอบสนองข้อสอบ

บทที่ 2

แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยแบ่งการนำเสนอแนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้องออกเป็น 5 ตอน คือ ตอนที่ 1 ทฤษฎีการทดสอบแบบดั้งเดิม ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ ตอนที่ 3 ความเข้มของอิทธิพล ตอนที่ 4 คุณสมบัติตัวประมาณค่าที่ดี และตอนที่ 5 งานวิจัยที่เกี่ยวข้อง รายละเอียดมีดังนี้

ตอนที่ 1 ทฤษฎีการทดสอบแบบดั้งเดิม

1. แนวคิดและโมเดลของทฤษฎีการทดสอบแบบดั้งเดิม

ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) เป็นทฤษฎีทางการทดสอบที่เชื่อว่า คะแนนที่ได้จากแบบสอบมีความคลาดเคลื่อนไปจากคะแนนที่แท้จริงของคุณลักษณะของบุคคลที่ต้องการวัด ซึ่ง Charles Spearman ได้เสนอแนวคิดนี้ในปี ค.ศ.1904 (พ.ศ.2447) และมีการพัฒนาและประยุกต์แนวคิดนี้เรื่อยมา โดยทฤษฎีการทดสอบแบบดั้งเดิมมีแนวคิดที่สำคัญว่า คะแนนจากการวัดหรือคะแนนที่สังเกตได้ (observed score: X) เกิดจากผลรวมเชิงเส้นของคะแนนจริง (true score) และคะแนนความคลาดเคลื่อน (error score: E) ซึ่งเป็นความคลาดเคลื่อนของการวัด ทั้งคะแนนจริงและคะแนนความคลาดเคลื่อนล้วนเป็นค่าที่ไม่สามารถสังเกตได้โดยตรงทั้งสองค่า โมเดลการทดสอบแบบดั้งเดิมจึงเขียนเป็นสมการได้ดังนี้ (ศิริชัย กาญจนวาสี, 2545)

$$X = T + E$$

หากคะแนนความคลาดเคลื่อนมีค่าน้อยมากหรือเข้าใกล้ศูนย์ คะแนนที่สังเกตได้จะมีค่าใกล้เคียงกับคะแนนจริง ทั้งนี้ขึ้นอยู่กับสถานการณ์ในการสอบและคุณภาพของแบบสอบ ดังนั้นนักวัดผลจึงพยายามสร้างแบบสอบที่มีคุณภาพสูง และจัดสถานการณ์การสอบให้ลดโอกาสที่จะเกิดความคลาดเคลื่อนให้ได้มากที่สุด เพื่อที่จะใช้คะแนนดิบหรือคะแนนที่สังเกตได้โดยไม่ต้องประมาณค่าคะแนนจริง (Fan, 1998)

2. ข้อตกลงเบื้องต้นของทฤษฎี แบ่งเป็นข้อตกลงเบื้องต้นของโมเดลและข้อตกลงเบื้องต้นเกี่ยวกับแบบสอบคู่ขนาน รายละเอียดมีดังนี้ (ศิริชัย กาญจนวาสี, 2545)

2.1 คะแนนที่ได้จากการวัดมีความสัมพันธ์เชิงเส้นตรง (additive relationship) กับคะแนนจริงและคะแนนความคลาดเคลื่อน กล่าวคือ ข้อมูลและโมเดลการวัดมีความสอดคล้องกัน

2.2 คะแนนจริงมีสภาวะคงที่ ซึ่งเท่ากับค่าคาดหวังของคะแนนที่สังเกตได้ ($T = E(X)$)

2.3 คะแนนความคลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริง นัยยะของข้อตกลงนี้คือ คะแนนความคลาดเคลื่อนของผู้สอบเป็นอิสระจากคะแนนจริง ไม่ว่าคะแนนจริงจะเป็นเท่าใดจะมีคะแนนความคลาดเคลื่อนของการวัดเพียงค่าเดียวเท่านั้น

2.4 คะแนนความคลาดเคลื่อนของผู้สอบแต่ละคนไม่มีความสัมพันธ์กัน

2.5 แบบสอบ 2 ฉบับจะถือว่าเป็นแบบสอบคู่ขนาน ก็ต่อเมื่อ คะแนนจริงของผู้สอบแต่ละคนมีค่าเท่ากันทั้งสองฉบับ และความคลาดเคลื่อนมาตรฐานของประชากรที่ทำแบบสอบทั้งสองฉบับมีค่าเท่ากัน

2.6 แบบสอบ 2 ฉบับจะถือเป็นแบบสอบที่เทียบเท่ากัน ก็ต่อเมื่อ ความคลาดเคลื่อนมาตรฐานของประชากรที่ทำแบบสอบทั้งสองฉบับมีค่าเท่ากัน แต่สามารถหาค่าคงที่ค่าหนึ่งที่ทำให้คะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับหนึ่งเท่ากับคะแนนจริงจากแบบสอบอีกฉบับหนึ่ง

3. ความคลาดเคลื่อนมาตรฐานของการวัด

เนื่องจากในการวัดแต่ละครั้งย่อมมีความคลาดเคลื่อนเกิดขึ้น ทำให้คะแนนที่สังเกตได้ (X) จากการสอบซ้ำหลาย ๆ ครั้งด้วยแบบสอบที่มีความเป็นคู่ขนานย่อมมีการกระจายรอบ ๆ ค่าคะแนนจริง (T) ของผู้สอบ โดยมีส่วนเบี่ยงเบนมาตรฐานเป็น σ_E หรือเรียกว่าความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement: SEM) ซึ่งคำนวณได้จากสมการต่อไปนี้

$$SEM = S_X\sqrt{1-R_{XX'}}$$

โดยที่ SEM หมายถึง ค่าความคลาดเคลื่อนมาตรฐานของการวัด

S_X หมายถึง ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนที่สังเกตได้ (X)

$R_{XX'}$ หมายถึง ค่าความเที่ยงซึ่งอาจคำนวณได้จากสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้จากแบบสอบคู่ขนาน 2 ฉบับ

ค่าความคลาดเคลื่อนมาตรฐานของการวัดและค่าความเที่ยงดังกล่าวสามารถนำมาใช้ในการปรับแก้การลดค่า (attenuation) ของตัวประมาณค่าความเข้มของอิทธิพลได้ เช่น การปรับแก้ค่าขนาดอิทธิพล (Cohen's *d*) ด้วยสูตร $adjusted\ d = d\Big/\sqrt{R_{XX'}}$ (Wang & Chen, 2004) เป็นต้น

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ

1. แนวคิดของทฤษฎี

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) อธิบายความสัมพันธ์ระหว่างความสามารถของผู้สอบและคำตอบของผู้สอบ โดยใช้คุณลักษณะของข้อสอบ ได้แก่ ความยาก อำนาจจำแนกและโอกาสในการเดา โดยแนวคิดหลักของทฤษฎีนี้คือการนำข้อมูลจากการตอบมาประมาณค่าพารามิเตอร์ของผู้สอบและของข้อสอบตามโมเดลที่ใช้ (ศิริชัย กาญจนวาสี, 2545)

2. พัฒนาการของทฤษฎี (เอมอร จังศิริพรปกรณ์, 2545)

จุดเริ่มต้นของทฤษฎีนี้เกิดขึ้นในปี ค.ศ.1916 (พ.ศ.2459) โดยนักจิตวิทยา 2 ท่าน คือ Binet และ Simon ได้สร้างกราฟแสดงความสัมพันธ์ระหว่างทักษะในการคิดของเด็กกับอายุ เป็นแนวคิดพื้นฐานของโค้งคุณลักษณะข้อสอบ (item characteristic curve) ต่อมาในปี ค.ศ.1936 (พ.ศ.2479) Richardson แสดงความสัมพันธ์ระหว่างทฤษฎีการตอบสนองข้อสอบและทฤษฎีการวัดแบบดั้งเดิม และเสนอวิธีประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ ในปี ค.ศ.1943 (พ.ศ.2486) Lawley ได้เสนอบทความเกี่ยวกับปัญหาในการสร้างและการเลือกข้อสอบ และเสนอแนวความคิดเกี่ยวกับโมเดลนอร์มัลออไจฟ์ (normal ogive model) ซึ่งเป็นฐานคิดยุคแรกของโมเดลในการวิเคราะห์ตามทฤษฎีนี้ โดยนักคณิตศาสตร์ชาวเดนมาร์กคือ Rasch ได้เสนอโมเดลราส์ช (Rasch model) ในปี ค.ศ.1950 (พ.ศ.2493) มีแนวคิดว่า ความสามารถของผู้สอบและความยากของข้อสอบเท่านั้นที่มีอิทธิพลต่อการตอบสนองข้อสอบ ค.ศ.1952 (พ.ศ.2495) Lord ได้เสนอโมเดลโค้งสะสมปกติแบบสองพารามิเตอร์ โดยเพิ่มพารามิเตอร์อำนาจจำแนกของข้อสอบ

Birnbaum ได้พัฒนาโมเดลโลจิสติก (logistic model) ในปี ค.ศ.1968 (พ.ศ.2511) ช่วยให้สามารถคำนวณได้ง่ายขึ้น และเป็นฐานความคิดของโมเดลในการวิเคราะห์ตามทฤษฎีนี้ในปัจจุบัน ต่อมาในปี ค.ศ.1974 (พ.ศ.2517) Lord ได้เสนอโมเดลโลจิสติกแบบสามพารามิเตอร์ โดยเพิ่มพารามิเตอร์โอกาสการเดาข้อสอบเข้าไปในโมเดล

โมเดลการตอบสนองข้อสอบในระยะแรกจะใช้กับแบบสอบที่ให้คะแนนรายข้อเป็นสองระดับหรือให้คะแนนแบบ 0-1 ซึ่งเรียกว่าโมเดลการตอบสนองข้อสอบแบบทวิภาค (dichotomous item response model) ต่อมาได้มีผู้พัฒนาโมเดลสำหรับการตรวจให้คะแนนแบบมาตรประมาณค่าหรือแบบพหุภาค (polytomous) โมเดลในแนวนี้เรียกว่า โมเดลการตอบสนองข้อสอบแบบพหุภาค (polytomous item response model) อนึ่ง ในการวิจัยครั้งนี้มีขอบเขตการวิจัยเฉพาะโมเดลการตอบสนองข้อสอบที่ให้คะแนนแบบทวิภาค ดังนั้นในส่วนของการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องจะนำเสนอเฉพาะโมเดลในกลุ่มนี้เท่านั้น

3. ข้อตกลงเบื้องต้นของทฤษฎี

3.1 ความเป็นเอกมิติ (unidimensionality) หมายถึง คุณสมบัติของแบบสอบที่ข้อสอบทุกข้อในแบบสอบมุ่งวัดเพียงคุณลักษณะหรือความสามารถเดียวเท่านั้น โดยทั่วไปการตรวจสอบความเป็นเอกมิติของแบบสอบสามารถทำได้โดยการคำนวณอัตราส่วนของค่าไอเกน (Eigen Ratio: ER) ขององค์ประกอบที่ 1 และ 2 ที่ได้จากการวิเคราะห์องค์ประกอบ (factor analysis) ค่าที่ได้ไม่ควรต่ำกว่า 3 การตรวจสอบความเป็นเอกมิติด้วยวิธีนี้สะดวกมาก เนื่องจากมีโปรแกรมสำเร็จรูปในการวิเคราะห์ ทั้งยังมีประสิทธิภาพและความไวสูง (วรนุช แหยมแสง, 2537; นิคม กีรติวรางกูร, 2542; Hambleton & Swaminathan, 1985; Embretson & Reise, 2000)

3.2 ความเป็นอิสระ (local independence) หมายถึง ผลการตอบสนองข้อสอบแต่ละข้อเกิดจากความสามารถที่แท้จริงของผู้สอบ (θ) เพียงปัจจัยเดียวเท่านั้น กล่าวคือ ผลการตอบสนองข้อสอบแต่ละข้อเป็นอิสระจากกัน เมื่อมีการควบคุมความสามารถที่ส่งผลต่อผู้สอบ (θ คงที่)

ความเป็นอิสระแบ่งเป็น 2 ประเภท คือ *ความเป็นอิสระระหว่างข้อสอบ* นั่นคือผลการตอบข้อสอบรายข้อของผู้สอบคนเดียวกันเป็นอิสระจากกัน ความน่าจะเป็นของแบบแผนการตอบข้อสอบของผู้สอบ จะเท่ากับผลคูณของค่าความน่าจะเป็นของผลการตอบข้อสอบแต่ละข้อ และ *ความเป็นอิสระระหว่างผู้สอบ* นั่นคือผลการตอบข้อสอบข้อเดียวกันของผู้สอบแต่ละคนเป็นอิสระจากกัน ความน่าจะเป็นของแบบแผนการตอบข้อสอบของผู้สอบทุกคน จะเท่ากับผลคูณระหว่างความน่าจะเป็นของผลการตอบข้อนั้นของผู้สอบทุกคน การตรวจสอบความเป็นอิสระสามารถพิจารณาได้จากเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วม (variance-covariance matrix) ของคะแนนคำตอบรายข้อ

ข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระและความเป็นเอกมิติมีความสัมพันธ์กัน หากแบบสอบไม่มีความเป็นเอกมิติจะทำให้ขาดความเป็นอิสระ หรือกล่าวได้ว่าความเป็นอิสระจะมีโอกาสเกิดขึ้นได้สูงเมื่อข้อสอบแต่ละข้อมุ่งวัดความสามารถเดียวกัน (Hambleton & Swaminathan, 1985; Embretson & Reise, 2000)

3.3 ความสอดคล้องระหว่างโมเดลและข้อมูล (model-data fit) หมายถึง โมเดลที่ใช้ในการประมาณค่าพารามิเตอร์ของผู้สอบและแบบสอบต้องเหมาะสมกับธรรมชาติของข้อมูลเชิงประจักษ์ โดยตรวจสอบได้จาก*ความไม่แปรเปลี่ยนของพารามิเตอร์ความสามารถ (invariance of ability parameter)* และ*ความไม่แปรเปลี่ยนของพารามิเตอร์ของข้อสอบ (invariance of item parameter)* โดยจัดสถานการณ์การสอบตามจุดประสงค์ของการทดสอบแล้วนำข้อมูลแต่ละสถานการณ์มาประมาณค่าพารามิเตอร์ เช่น ประมาณค่าพารามิเตอร์ของผู้สอบกลุ่มเดิม โดยใช้แบบสอบสองชุดเพื่อพิจารณาความไม่แปรเปลี่ยนของพารามิเตอร์ความสามารถ เป็นต้น ค่าประมาณพารามิเตอร์จะไม่แปรเปลี่ยนเมื่อความแตกต่างที่เกิดขึ้นมีค่าไม่เกินความคลาดเคลื่อนมาตรฐานของการประมาณ

3.4 การสอบที่ไม่แข่งขันด้านเวลา (nonspeeded test administrative) หมายถึง ความเร็วในการตอบต้องไม่มีอิทธิพลต่อผลการตอบ เนื่องจากความสามารถของผู้สอบเป็นปัจจัยสำคัญที่ส่งผลต่อการตอบข้อสอบ การจัดการสอบจึงไม่ควรอยู่ภายใต้สถานการณ์ที่สอบแข่งขันด้วยเวลาหรือจัดการสอบที่มีเวลาเพียงพอให้ผู้สอบได้ตอบ วิธีการตรวจสอบเหมาะสมของมิติด้านเวลาสามารถพิจารณาได้จากร้อยละของผู้สอบที่ทำข้อสอบได้ครบทุกข้อ นอกจากนี้ ควรจะเปรียบเทียบความแปรปรวนของจำนวนข้อที่ถูกเว้นกับความแปรปรวนของจำนวนข้อที่ตอบผิด ถ้าอัตราส่วนของความแปรปรวนเข้าใกล้ 0 แสดงว่าการจัดการสอบเป็นไปตามข้อตกลงเบื้องต้นนี้

4. โมเดลการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบมีการพัฒนาโมเดลการวิเคราะห์อย่างต่อเนื่อง ส่วนใหญ่เป็นผลมาจากการพัฒนาแนวคิดทางสถิติและจิตวิทยาเป็นหลัก เช่น แนวคิดของฟังก์ชันโลจิสติก (logistic function) ที่เข้ามาแทนที่ฟังก์ชันนอร์มัลออไจฟ์ (normal ogive function) ช่วยให้การคำนวณง่ายขึ้น การพัฒนาโมเดลแบบนอนพาราเมตริก (nonparametric model) การเพิ่มตัวแปรที่มีความสัมพันธ์กับคะแนนสอบมาใช้เป็นพารามิเตอร์ในโมเดล ตลอดจนแนวคิดของการวัดเอกมิติ (unidimensionality scaling) ที่พัฒนาขึ้นจนเป็นแนวคิดของการวัดพหุมิติ (multidimensionality scaling) เป็นต้น ทั้งนี้เพื่อให้สอดคล้องกับการใช้งานที่หลากหลาย

ปัจจุบันมีโมเดลการวิเคราะห์สำหรับทฤษฎีการตอบสนองข้อสอบประมาณ 40 โมเดล (Van der Linden & Hambleton, 1997; Embretson & Reise, 2000) โมเดลสำหรับการวิเคราะห์ที่ได้รับความนิยมมากคือโมเดลที่มีลักษณะไม่ซับซ้อน และใช้ฟังก์ชันโลจิสติกในการวิเคราะห์หรือได้รับการพัฒนาบนฐานของฟังก์ชันโลจิสติก โดยโมเดลการตอบสนองข้อสอบแบบทวิภาคที่นิยมใช้กันมากได้แก่ โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (one-parameter logistic model: 1PLM or Rasch model) โมเดลโลจิสติกแบบสองพารามิเตอร์ (two-parameter logistic model: 2PLM) และโมเดลโลจิสติกแบบสามพารามิเตอร์ (three-parameter logistic model: 3PLM) ทั้งสามโมเดลมีลักษณะสมการที่ใช้อธิบายความสัมพันธ์ระหว่างพารามิเตอร์ที่คล้ายคลึงกัน สามารถเขียนสมการร่วม (common equation) เพื่ออธิบายทั้งสามโมเดลได้ดังนี้ (ศิริชัย กาญจนวาสี, 2545; Hambleton, Swaminathan & Rogers, 1991; Embretson & Reise, 2000)

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

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

โดยที่ $P_i(\theta)$ คือ ค่าความน่าจะเป็นที่ผู้มีความสามารถ θ จะตอบข้อสอบข้อที่ i ถูก

a_i คือ ค่าพารามิเตอร์อำนาจจำแนก มีค่าคงที่เท่ากับ 1 ใน 1PLM

b_i คือ ค่าพารามิเตอร์ความยาก

c_i คือ ค่าพารามิเตอร์โอกาสในการเดา มีค่าคงที่เท่ากับ 0 ใน 1PLM และ 2PLM

e คือ ค่าคงที่ มีค่าประมาณ 2.718281…

D คือ ค่าคงที่ มีค่าประมาณ 1.7 ใช้ปรับค่าจากฟังก์ชันโลจิสติกให้ใกล้เคียงกับค่าจากฟังก์ชันนอร์มัลออไจฟ์

การเลือกใช้โมเดลการตอบสนองข้อสอบต้องคำนึงถึงแนวคิด และข้อตกลงเบื้องต้น เนื่องจากมีผลต่อความถูกต้องของค่าประมาณของพารามิเตอร์ที่ได้ตามความสัมพันธ์ในโมเดล

ตอนที่ 3 ความเข้มของอิทธิพล (effect magnitude)

1. หลักการและความสำคัญของความเข้มของอิทธิพล

เนื่องจากการทดสอบนัยสำคัญทางสถิติ (statistical significance testing) หรือการทดสอบนัยสำคัญของสมมติฐานศูนย์ (null hypothesis significance testing) ซึ่งจะให้สารสนเทศเกี่ยวกับความน่าจะเป็นที่อิทธิพลดังกล่าวจะเกิดขึ้นเมื่อสมมติฐานศูนย์เป็นจริงแต่ไม่สามารถให้สารสนเทศสืบเนื่องต่อไปได้ว่า ถ้ามีโอกาสเกิดขึ้นจริงแล้วอิทธิพลดังกล่าวจะมีปริมาณมากน้อยเพียงใด และมีปริมาณมากพอที่จะใช้ประโยชน์ได้จริงในทางปฏิบัติหรือไม่ (Kirk, 2001; Paul & Plucker, 2004) นอกจากนี้การทดสอบนัยสำคัญทางสถิติยังมีจุดอ่อนที่ถูกวิพากษ์วิจารณ์อย่างมาก

Shaver (1993), Snyder และ Lawson (1993) และ Kirk (1996) ได้กล่าวถึงจุดอ่อนของการทดสอบนัยสำคัญทางสถิติไว้สอดคล้องกัน 3 ประการ ได้แก่ **ประการแรก** การทดสอบนัยสำคัญทางสถิติไม่ได้ให้คำตอบที่นักวิจัยต้องการ โดยทั่วไปคำตอบหนึ่งที่ต้องการในการทำวิจัยคือความน่าจะเป็นที่สมมติฐานศูนย์จะเป็นจริงเมื่อมีข้อมูลที่สังเกตได้ชุดหนึ่ง นั่นคือ $p(H_0|data\ set)$ แต่ผลที่ได้การทดสอบสมมติฐานคือความน่าจะเป็นที่จะได้ข้อมูลที่สังเกตได้นี้หรือชุดที่ต่างจากนี้มาก ๆ เมื่อกำหนดให้สมมติฐานศูนย์เป็นจริง นั่นคือ $p(data\ set|H_0)$ ซึ่งผลสรุปอาจต่างกันได้ **ประการที่สอง** การปฏิเสธสมมติฐานศูนย์ไม่ได้ให้สารสนเทศเกี่ยวกับขนาดของอิทธิพลหรือประโยชน์ใช้สอยของอิทธิพลในเชิงปฏิบัติ และการทราบค่า p อย่างเดียวไม่ได้ช่วยให้เกิดการพัฒนาทฤษฎีได้มากเท่ากับการมีสารสนเทศเกี่ยวกับอิทธิพลที่เกิดขึ้นในงานวิจัย (Thompson, 2000; Capraro & Capraro, 2002) **ประการที่สาม** การที่นักวิจัยกำหนดระดับนัยสำคัญสำหรับการตัดสินใจ (เช่น กำหนด $\alpha = .05$) เป็นการเปลี่ยนคุณลักษณะของค่าความน่าจะเป็นที่เดิมมีความต่อเนื่องให้กลายเป็นทวิภาค (ปฏิเสธหรือไม่ปฏิเสธ) การตัดสินใจเช่นนี้จะเกิดความคลาดเคลื่อนได้ เช่น กรณีที่มีนักวิจัยสองคนศึกษาตัวแปรตัวเดียวกัน แต่กำหนดระดับนัยสำคัญสำหรับการตัดสินใจต่างกัน ผลได้สรุปย่อมต่างกัน

นอกจากนี้ Craig, Eison และ Metze (1976 อ้างถึงใน Shaver, 1993) และ Snyder และ Lawson (1993) ยังกล่าวไว้ว่าขนาดกลุ่มตัวอย่างจะส่งผลอย่างยิ่งต่อการทดสอบนัยสำคัญทางสถิติ กล่าวคือ ถ้านักวิจัยใช้ขนาดกลุ่มตัวอย่างที่มากพอ การได้ผลการทดสอบว่ามีนัยสำคัญทางสถิติอาจเกิดขึ้นได้แม้ตัวแปรที่ศึกษาจะมีความสัมพันธ์กันน้อยมากก็ตาม จากข้อวิพากษ์เหล่านี้ทำให้เกิดกระแสการพัฒนาวิธีการประมาณค่าความเข้มของอิทธิพล เพื่อช่วยเติมเต็มข้อจำกัดของการทดสอบนัยสำคัญทางสถิติข้างต้น

2. คำศัพท์ที่ใช้และความหมาย

เนื่องจาก “ความเข้มของอิทธิพล” ที่กล่าวถึงนี้เป็นอิทธิพลที่เกิดขึ้นในประชากรและในความเป็นจริงการวัดค่าจากกลุ่มประชากรทำได้ยาก ผู้วิจัยจึงใช้คำว่า “การประมาณค่าความเข้มของอิทธิพล” แทน “การวัดความเข้มของอิทธิพล” โดยคำศัพท์ภาษาอังกฤษที่หมายถึงการประมาณค่าอิทธิพลที่ปรากฏในรายงานวิจัยและเอกสารทางวิชาการนั้นมีหลายคำดังที่ Snyder และ Lawson (1993) และ Ives (2003) ได้สำรวจไว้ อาทิ estimates of magnitude of the effect, estimates of the magnitude of the experimental effect, estimates of explained variance, effect size estimates, estimates of the strength of relation, estimates of the strength of association, proportion (or percentage) of variance accounted for, measure of association

Hedges และ Olkin (1985), Snyder และ Lawson (1993), Kirk (2001) และ Gliner, Leech และ Morgan (2002) สนับสนุนให้ใช้คำว่า effect magnitude ในขณะที่ Wilkinson และ APA Task Force on Statistical Inference (1999), Huberty (2002) และ Trusty, Thompson และ Petrocelli (2004) ใช้คำว่า effect size ด้าน Shaver (1993) ได้เสนอให้ใช้คำว่า result size เพื่อหลีกเลี่ยงการสื่อความในเชิงสาเหตุและอิทธิพล แต่ Shaver เองได้กล่าวไว้ด้วยว่านักวิจัยทางสังคมศาสตร์นิยมใช้คำว่า effect size เกินกว่าจะเปลี่ยนแปลงได้ ปัจจุบันจึงมักพบว่านักวิจัยส่วนใหญ่ใช้คำว่า effect size ในความหมายเดียวกับ effect magnitude แต่ในงานวิจัยครั้งนี้ผู้วิจัยเลือกใช้คำว่า effect magnitude ในความหมายของความเข้มของอิทธิพลที่เกิดขึ้นในงานวิจัย และใช้คำว่า effect size ในความหมายของผลต่างมาตรฐานตามนิยามของ Cohen ซึ่งเป็นหนึ่งในสถิติที่ใช้ประมาณค่าความเข้มของอิทธิพลที่เกิดขึ้นในการวิจัยเท่านั้น ยังมีสถิติอีกหลายชนิดที่ไม่ได้อยู่ในรูปของผลต่างมาตรฐานแต่สามารถใช้ประมาณค่าความเข้มของอิทธิพลได้ ดังที่จะกล่าวถึงในลำดับต่อไป

เนื่องจากการใช้ศัพท์ที่แตกต่างกัน ทำให้การนิยามความหมายของความเข้มของอิทธิพลจึงมีแตกต่างกันอยู่บ้าง นักสถิติกลุ่มหนึ่งจะใช้คำว่า ขนาดอิทธิพล (effect size) แทน ความเข้มของอิทธิพล (effect magnitude) ในขณะที่นักสถิติอีกกลุ่มหนึ่งจะใช้คำว่าความเข้มของอิทธิพลโดยตรง นิยามจากทั้งสองกลุ่มมีทั้งการนิยามในภาพกว้าง ๆ การนิยามในรูปของขนาดอิทธิพลและ/หรือสถิติอื่น ๆ ในกลุ่มความสัมพันธ์ รายละเอียดดังนี้

Cohen (1969 อ้างถึงใน นงลักษณ์ วิรัชชัย, 2542) ซึ่งเป็นผู้พัฒนาสูตรการประมาณค่าความเข้มของอิทธิพลในยุคแรกให้นิยามในเชิงปฏิบัติการไว้ว่า เป็นอัตราส่วนระหว่างผลต่างของค่าเฉลี่ยจากกลุ่มทดลองและกลุ่มควบคุมกับค่าส่วนเบี่ยงเบนมาตรฐานรวม (pooled standard deviation) และเรียกอัตราส่วนดังกล่าวว่า ขนาดอิทธิพล (effect size)

Snyder และ Lawson (1992) กล่าวว่า ความเข้มของอิทธิพล (magnitude-of-effect) หมายถึงค่าสถิติที่บอกนักวิจัยได้ว่าตัวแปรต้นสามารถควบคุม ทำนายหรืออธิบายตัวแปรตามได้มากน้อย

เพียงใด (a magnitude-of-effect statistic tell us how much of the DV can be controlled, predicted, or explained by the IV(s))

Kirk (1996) กล่าวว่า ขนาดอิทธิพล (effect size) เป็นขนาดหรือระดับความสัมพันธ์ของตัวแปรต้นและตัวแปรตาม (the strength of the association between the IV's and DV's)

Hair และคณะ (1998) ให้นิยามของ ขนาดอิทธิพล (effect size) ว่าหมายถึง ค่าประมาณของระดับของปรากฏการณ์ที่ศึกษาว่ามีอยู่หรือเกิดขึ้นในประชากร (estimate of the degree to which the phenomenon being studied (e.g., correlation or difference in mean) exists in the population)

Becker (1999) กล่าวถึงขนาดอิทธิพล (effect size) ในการวิเคราะห์ความแปรปรวนว่าหมายถึง ค่าที่บอกระดับของความสัมพันธ์ระหว่างอิทธิพลและตัวแปรตาม (...are measures of the degree of association between an effect (e.g., a main effect, an interaction effect, a linear contrast) and the dependent variable)

Fidler และ Thompson (2001) กล่าวว่า ขนาดอิทธิพล (effect size) หมายถึง ความแตกต่างมาตรฐานของค่าเฉลี่ยระหว่างกลุ่มทดลองกับกลุ่มควบคุม (standardized effect size) หรือ ความผันแปรของตัวแปรตามที่สามารถอธิบายหรือทำนายได้ด้วยตัวแปรต้น (variance-accounted-for effect size)

Gliner, Leech และ Morgan (2002) กล่าวว่า ขนาดอิทธิพล (effect size) หมายถึง ค่าที่บอกอำนาจหรือระดับของความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม (…the strength of the relationship between the IV and the DV)

Robert และ Henson (2002) กล่าวว่า ขนาดอิทธิพล (effect size) คือค่าสถิติที่ช่วยในการประเมินความเข้มของความแตกต่างหรือความสัมพันธ์ (...to help evaluate the magnitude of a difference or relationship)

Thompson (2002) กล่าวว่า ขนาดอิทธิพล (effect size) เป็นค่าที่บอกระดับของผลที่เกิดขึ้นกับกลุ่มตัวอย่างที่แตกต่างออกไปจากสมมติฐานศูนย์ (an effect size characterizes the degree to which sample results diverge from the null hypothesis)

Trusty, Thompson และ Petrocelli (2004) กล่าวว่า ขนาดอิทธิพลเป็นค่าที่ใช้วัดอำนาจหรือระดับของความสัมพันธ์ (an effect size assess the strength of the relationship)

นงลักษณ์ วิรัชชัย (2542) กล่าวว่า ขนาดอิทธิพลเป็นค่าสถิติที่บอกปริมาณผลของตัวแปรจัดกระทำที่มีต่อตัวแปรตามในการวิจัยเชิงทดลอง

โดยสรุป ความเข้มของอิทธิพล หมายถึง ตัวเลขที่ให้สารสนเทศเกี่ยวกับขนาดของผลอันเกิดจากอิทธิพลของตัวแปรจัดกระทำที่ทำให้เกิดการเปลี่ยนแปลงในตัวแปรตาม หรือตัวเลขที่ให้สารสนเทศเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร

3. พัฒนาการของแนวคิดการประมาณค่าความเข้มของอิทธิพล

การประมาณค่าความเข้มของอิทธิพลในยุคแรกพัฒนาขึ้นโดย Jacob Cohen (นงลักษณ์ วิรัชชัย, 2542) ซึ่งใช้การคำนวณค่าอัตราส่วนระหว่างผลต่างของค่าเฉลี่ยจากกลุ่มทดลองและกลุ่มควบคุม กับค่าส่วนเบี่ยงเบนมาตรฐานรวม ดังสมการต่อไปนี้

$$\delta = \frac{\mu_E - \mu_C}{\sigma_{pooled}} \text{ (Kirk, 1996)}$$

โดยที่ δ แทนค่าขนาดอิทธิพล

μ_E แทนค่าเฉลี่ยของประชากรกลุ่มทดลอง

μ_C แทนค่าเฉลี่ยของประชากรกลุ่มควบคุม

และ σ_{pooled} แทนส่วนเบี่ยงเบนมาตรฐานรวมของกลุ่มทดลองและกลุ่มควบคุม

ในทางปฏิบัติจะประมาณค่าพารามิเตอร์ δ ด้วยค่า $d = \frac{\overline{X}_E - \overline{X}_C}{S_{pooled}}$ และความเข้มของอิทธิพลตามนิยามของ Cohen นี้อาจเรียกว่าผลต่างมาตรฐานระหว่างคะแนนเฉลี่ย (standardized mean differences)

การประมาณค่าความเข้มของอิทธิพลตามแนวคิดของ Cohen เป็นพื้นฐานหลักในการพัฒนาสถิติสำหรับการประมาณค่าความเข้มของอิทธิพลในยุคแรก ๆ โดยอาศัยความแตกต่างระหว่างค่าเฉลี่ยระหว่างกลุ่มควบคุมและกลุ่มทดลอง เช่น Cohen's (1988) *d, f, g, h, q, w*, Glass's (1976) g' และ Hedges's (1981) g (Kirk, 1996)

ในระยะต่อมามีการขยายกรอบความคิดของค่าความเข้มของอิทธิพลจากการพิจารณาความแตกต่างระหว่างกลุ่มไปสู่ความสัมพันธ์ระหว่างตัวแปร หรือความแปรปรวนของตัวแปรหนึ่งที่อธิบายได้ด้วยความแปรปรวนของอีกตัวแปรหนึ่ง เป็นแนวคิดเดียวกับการหาความสัมพันธ์ของตัวแปรที่มีการพัฒนาขึ้นมาจากงานของ Galton ในปี ค.ศ.1888 (พ.ศ.2431) และมีสถิติที่นิยมใช้กันแพร่หลายในงานวิจัยทั่วไป เช่น ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson product moment correlation coefficient) สัมประสิทธิ์สหสัมพันธ์แบบพอยต์ไบซีเรียล (point biserial correlation coefficient) และสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล (biserial correlation coefficient) แต่นักวิจัยมักใช้เพียงเพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรเท่านั้น ไม่ได้ใช้ประโยชน์ในการอธิบายอิทธิพลที่เกิดขึ้นในงานวิจัย เช่นเดียวกับสัดส่วนสหสัมพันธ์ (correlation ratio: η^2) ซึ่ง Pearson ได้เสนอขึ้นในปี ค.ศ.1905 (พ.ศ.2448) สำหรับการวิเคราะห์ความแปรปรวน (ANOVA) ก็สามารถใช้ประมาณค่าความเข้มของอิทธิพลตามแนวคิดนี้ได้ด้วย (Huberty, 2002) สถิติที่ใช้ประมาณค่าความเข้มของอิทธิพลตามแนวคิดนี้ ได้แก่ *r*, r_{pb}, r^2, *R*, R^2, η, η^2, ϕ, Cramér's (1946) *V*, Fisher's (1963) Z, Hays's (1963) ω^2 และ ρ_I, Kelley's (1935) ε^2 และ Tatsuoka's (1973) $\hat{\omega}^2_{\text{mult.c}}$ (Kirk, 1996)

การประมาณค่าความเข้มของอิทธิพลระยะต่อมารับอิทธิพลมาจากแนวคิดของ Kelley เมื่อปี ค.ศ.1920 (พ.ศ.2473) ที่เสนอการประมาณค่าความเข้มของอิทธิพลด้วยค่าการทับซ้อนของการกระจายระหว่างกลุ่ม (group overlap) ต่อมาในปี ค.ศ.1937 (พ.ศ.2480) Tilton เสนอให้ใช้ค่าร้อยละของพื้นที่ร่วมโค้งการกระจายในการเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่าง 2 กลุ่ม (two-group univariate mean comparison) เพื่อใช้ตัดสินว่าความแตกต่างระหว่างสองกลุ่มนั้นมีนัยสำคัญหรือไม่ ส่วน Levy (1967) เสนอให้ใช้สถิติที่เรียกว่า การวิเคราะห์ทำนายจำแนกกลุ่ม (predictive discriminant analysis) ซึ่งแตกต่างไปจากวิธีของ Tilton มาใช้ในการประมาณค่าความเข้มของอิทธิพลตามแนวคิดนี้ โดยยังคงมีลักษณะการใช้งานที่ใกล้เคียงกัน คือ ใช้สำหรับประมาณค่าความเข้มของอิทธิพลในการวิจัยที่ต้องการเปรียบเทียบตัวแปรตามเพียงตัวเดียว ระหว่างกลุ่มสองกลุ่ม (univariate two-mean comparison) ต่อมาในปี ค.ศ.2000 (พ.ศ.2543) Huberty และ Lowlan ได้เสนอวิธีการประมาณค่าความเข้มของอิทธิพลจากความเหลื่อมล้ำของการกระจายสำหรับการวิจัยที่มีตัวแปรตามตั้งแต่สองตัวขึ้นไป (multiple outcome variable) (Huberty, 2002) จะเห็นได้ว่าแนวคิดการประมาณค่าความเข้มของอิทธิพลในช่วงนี้มีความคล้ายคลึงกับแนวคิดการประมาณค่าความเข้มของอิทธิพลในยุคแรก ๆ ของ Cohen แต่มีมุมมองที่แตกต่างกันและยังไม่เป็นที่แพร่หลายเนื่องจากความไม่สะดวกในการคำนวณ

การประมาณค่าความเข้มของอิทธิพลมีพัฒนาการอย่างต่อเนื่องและหลากหลาย ผู้วิจัยได้สรุปประเภทของสถิติที่ใช้เป็นตัวประมาณค่าความเข้มของอิทธิพลจากการศึกษาเอกสารที่เกี่ยวข้องเพื่อประกอบการตัดสินใจเลือกใช้ตัวประมาณค่าอย่างเหมาะสมในหัวข้อต่อไป

4. ประเภทของสถิติสำหรับการประมาณค่าความเข้มของอิทธิพล

Kirk (1996) ได้สำรวจพบว่ามีสถิติที่สามารถใช้ประมาณค่าความเข้มของอิทธิพลมากถึง 40 ชนิดโดย Capraro และ Capraro (2002) ได้แบ่งกลุ่มสถิติเหล่านี้ไว้อย่างครอบคลุมเป็น 3 กลุ่มได้แก่ กลุ่มสถิติความสัมพันธ์หรือความแปรปรวนที่ถูกอธิบาย (relationship or measure of association or variance-accounted-for) กลุ่มสถิติความแตกต่างระหว่างกลุ่ม (group difference or effect size or standardized mean difference) และกลุ่มสถิติความเหลื่อมล้ำระหว่างกลุ่ม (group overlap)

นอกจากนี้ยังอาจจำแนกตามชนิดของตัวแปรได้ 5 กลุ่ม คือ **กลุ่มที่หนึ่ง** ตัวแปรต้นและตัวแปรตามเป็นตัวแปรทวิภาค **กลุ่มที่สอง** ตัวแปรต้นเป็นตัวแปรทวิภาค ตัวแปรตามเป็นตัวแปรเชิงปริมาณ **กลุ่มที่สาม** ตัวแปรต้นและตัวแปรตามเป็นตัวแปรพหุภาค **กลุ่มที่สี่** ตัวแปรต้นเป็นตัวแปรพหุภาค ตัวแปรตามเป็นตัวแปรเชิงปริมาณ และ**กลุ่มที่ห้า** ตัวแปรต้นและตัวแปรตามเป็นตัวแปรเชิงปริมาณ (van den Noortgate, 2003) รายละเอียดดังตารางที่ 2.1

ตารางที่ 2.1 สถิติที่ใช้ประมาณค่าความเข้มของอิทธิพลจำแนกตามประเภทของตัวแปรวิจัย

ตัวแปรต้น / ตัวแปรตาม	ทวิภาค	มาตรนามบัญญัติ/ พหุภาค	มาตรอันดับ/ มาตรอันตรภาค
ทวิภาค	Risk different (RD) Relative risk (RR) Phi coefficient (ϕ) Odd ratio (OR)		
มาตรนามบัญญัติ/ พหุภาค		Contingency measures Goodman-Kruskal's τ Uncertainty coefficient Cohen's Kappa	
มาตรอันดับ/ มาตรอันตรภาค	g_{IG} Glass's Δ g_{gain} g_{gainIG} r_{pb}	Multiple g's η^2 ICC	Pearson's r Spearman's ρ Kendall's τ Gamma coefficient Weighted Kappa

ที่มา: van den Noortgate (2003)

ประเภทสถิติที่ใช้ประมาณค่าความเข้มของอิทธิพลตามที่ได้กล่าวไว้ข้างต้นนี้ อาจแบ่งย่อยตามการปรับแก้ความคลาดเคลื่อนในการประมาณค่า (corrected or uncorrected) หรือการให้ค่าลำเอียงหรือไม่ลำเอียง (unbiased or biased) ได้อีกด้วย (Wilkinson & APA Task Force on Statistical Inference, 1999; Thompson, 2002; Ives, 2003)

สถิติแต่ละชนิดมีความเหมาะสมในการประมาณค่าความเข้มของอิทธิพลด้วยต่างกัน นักสถิติหลายท่านได้เสนอความเห็นเกี่ยวกับการประมาณค่าความเข้มของอิทธิพลที่เหมาะสมสำหรับงานวิจัยเชิงปริมาณทั่ว ๆ ไปเอาไว้ เช่น Baugh (2002), Thompson (2002) และ Ives (2003) ได้กล่าวไว้สอดคล้องกันว่าการประมาณค่าความเข้มของอิทธิพลควรใช้สถิติในกลุ่มความสัมพันธ์หรือความแปรปรวนที่ถูกอธิบาย เพราะสามารถนำมาใช้ได้กับการวิเคราะห์เชิงปริมาณที่ใช้โมเดลเชิงเส้นทั่วไป และสามารถประยุกต์ใช้ได้กับงานวิจัยเชิงปริมาณทั้งที่เป็นเชิงทดลองและไม่ใช่เชิงทดลอง เนื่องจากเป้าหมายหนึ่งของการวิจัยเชิงปริมาณส่วนใหญ่ต้องการศึกษาความสัมพันธ์หรือความผันแปรร่วมกันระหว่างกลุ่มตัวแปรนั่นเอง (ศิริชัย กาญจนวาสี, 2541; Thompson, 2000)

นอกจากนี้ Rosenthal และ DiMatteo (2001) ยังกล่าวไว้ว่าสถิติในกลุ่มความสัมพันธ์อย่างสัมประสิทธิ์สหสัมพันธ์มีข้อดีเหนือกว่าขนาดอิทธิพลบางประการ คือ (1) การแปลงขนาดอิทธิพลให้อยู่ในรูปของสัมประสิทธิ์สหสัมพันธ์มีความสมเหตุสมผล เนื่องจากสัมประสิทธิ์สหสัมพันธ์แบบพอยต์ไบซีเรียล สามารถแสดงความสัมพันธ์ระหว่างตัวแปรต้นที่มีสองระดับกับตัวแปรตามที่มีค่าต่อเนื่องได้ ในขณะที่การแปลงสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันซึ่งตัวแปรมีค่าต่อเนื่อง (continuous Pearson) ให้อยู่ในรูปของขนาดอิทธิพลซึ่งใช้ข้อมูลทวิภาค (dichotomous d) จะทำให้เสียสารสนเทศไป (2) เมื่อใช้ contrast ที่มีองศาอิสระเป็น 1 (1 df contrast) สัมประสิทธิ์สหสัมพันธ์สามารถวิเคราะห์แนวโน้มระหว่างกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป ในขณะที่ขนาดอิทธิพลมีข้อจำกัดอยู่ที่ 2 กลุ่ม (3) ไม่จำเป็นต้องปรับแก้วิธีการคำนวณสัมประสิทธิ์สหสัมพันธ์เมื่อใช้กลุ่มตัวอย่างมากกว่า 2 กลุ่ม และ (4) สัมประสิทธิ์สหสัมพันธ์แปลความหมายง่ายกว่าขนาดอิทธิพล

เนื่องจากสัมประสิทธิ์สหสัมพันธ์เป็นสถิติที่นิยมใช้กันโดยทั่วไปในงานวิจัยเชิงปริมาณทั้งที่เป็นและไม่เป็นการทดลอง และขนาดอิทธิพลเป็นดัชนีมาตรฐานที่แพร่หลายสำหรับการรายงานค่าประมาณความเข้มของอิทธิพลในการวิเคราะห์อภิมาน การวิจัยครั้งนี้ผู้วิจัยสนใจที่จะเปรียบเทียบคุณสมบัติของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ โดยผู้วิจัยได้ทบทวนเอกสารที่นำเสนอความสัมพันธ์และการแปลความหมายของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์เพิ่มเติม ดังนี้

ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์เป็นตัวประมาณค่าความเข้มของอิทธิพลที่มีความสัมพันธ์ โดย Cohen (1988) และ Friedman (1968) ได้เสนอสมการแสดงความสัมพันธ์ไว้ตามลำดับดังนี้ (Thompson, 2002)

$$d = \frac{2r}{\sqrt{1-r^2}} \quad \text{และ} \quad r = \frac{d}{\sqrt{d^2+4}}$$

Hunter และ Schmidt (1990) กล่าวว่าสมการข้างต้นแสดงความสัมพันธ์ระหว่างขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์โดยประมาณ และได้เสนอสมการความสัมพันธ์ที่แม่นยำมากขึ้นดังนี้

$$d = \sqrt{\frac{N-2}{N}} \times \frac{2r}{\sqrt{1-r^2}} \quad \text{และ} \quad r = \frac{d}{2\sqrt{\frac{N-2}{N}+\frac{d^2}{4}}}$$

นอกจากนี้แนวคิดในการใช้สมการแสดงความสัมพันธ์หรือสมการแปลงรูประหว่างขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ทั้งหมดดังกล่าวมีความแตกต่างกันในข้อตกลงเบื้องต้นเกี่ยวกับขนาดของกลุ่มตัวอย่างที่ใช้ โดย Hunter และ Schmidt (1990) เสนอว่าสมการดังกล่าวควรใช้ในกรณีที่งานวิจัยเชิงทดลองมีกลุ่มควบคุมและกลุ่มทดลองขนาดเท่ากันและมีขนาดเป็นครึ่งหนึ่งของกลุ่มตัวอย่างสำหรับการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ ($N_C = N_E = N/2$) ในขณะที่ Hedges และ Olkin (1985) เสนอว่ากลุ่มทดลองและกลุ่มควบคุมในงานวิจัยเชิงทดลองควรมีขนาดเท่ากัน

และมีขนาดเท่ากับกลุ่มตัวอย่างในการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ ($N_C = N_E = N$) นอกจากนี้ยังมีนักสถิติบางท่านที่กล่าวถึงสมการของ Cohen และ Friedman โดยมิได้กล่าวถึงข้อตกลงเกี่ยวกับขนาดกลุ่มตัวอย่างหรือระบุเพียงแต่ว่ากลุ่มทดลองและกลุ่มควบคุมมีขนาดเท่ากัน (Rosnow & Rosenthal, 1996; Fidler & Thompson, 2001; Rosenthal & Dimatteo, 2001; Thompson, 2002)

งานวิจัยครั้งนี้ผู้วิจัยเลือกใช้การแปลงสัมประสิทธิ์สหสัมพันธ์ด้วยสมการและแนวคิดเกี่ยวกับขนาดกลุ่มตัวอย่างของ Hunter และ Schmidt (1990) ในการจำลองข้อมูลตั้งต้น (อธิบายเพิ่มเติมในบทที่ 3) เนื่องจากจะทำให้สถานการณ์หนึ่ง ๆ มีการกำหนดขนาดกลุ่มตัวอย่างเพียงขนาดเดียวทั้งในงานวิจัยเชิงทดลองและงานวิจัยเชิงสหสัมพันธ์ โดยค่าขนาดอิทธิพลที่ได้จากงานวิจัยเชิงทดลองจะสามารถเทียบได้กับสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล (biserial correlation) หรือสัมประสิทธิ์สหสัมพันธ์แบบพอยต์ไบซีเรียล (point biserial correlation) ได้ การเปรียบเทียบคุณสมบัติของตัวประมาณค่าระหว่างขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จึงจะมีความสมเหตุสมผล

สำหรับการแปลความหมายของขนาดอิทธิพลอาจอาศัยเกณฑ์ของ Cohen ในเบื้องต้น ซึ่งตรงกับเกณฑ์การแปลความหมายสัมประสิทธิ์สหสัมพันธ์ (Kirk, 1996) คือ

.2 หมายถึง มีอิทธิพลหรือมีความสัมพันธ์ต่ำ

.5 หมายถึง มีอิทธิพลหรือมีความสัมพันธ์ปานกลาง

.8 หมายถึง มีอิทธิพลหรือมีความสัมพันธ์สูง

อย่างไรก็ตามการแปลความหมายของค่าที่ได้จะต้องพิจารณาบริบทที่ศึกษาด้วย ไม่ควรใช้เกณฑ์เบื้องต้นดังกล่าวในทุก ๆ สถานการณ์ (Thompson, 2001 อ้างถึงใน Roberts & Henson, 2002)

5. ประโยชน์ของค่าความเข้มของอิทธิพล

ค่าความเข้มของอิทธิพลมีประโยชน์หลายประการ เช่น ช่วยให้ผู้อ่านรายงานการวิจัยประเมินความสอดคล้องของผลการวิจัยกับรายงานการวิจัยก่อนหน้านี้ ซึ่งศึกษาตัวแปรหรืออิทธิพลเดียวกันแต่ต่างบริบทกันได้ กล่าวคือผู้อ่านรายงานสามารถพิจารณาความสม่ำเสมอของอิทธิพลได้สะดวกขึ้น สามารถใช้เป็นข้อมูลเพื่อกำหนดขนาดกลุ่มตัวอย่างสำหรับการทำการวิจัยต่อไป และเป็นสารสนเทศเบื้องต้นสำหรับการวิเคราะห์อภิมาน (meta-analysis) และการวิเคราะห์อำนาจ (power analysis) อีกด้วย (นงลักษณ์ วิรัชชัย, 2542; Kirk, 1995; Wilkinson & APA Task Force on Statistical Inference, 1999; Thompson, 2000; Ives, 2003; Paul & Plucker, 2004)

ด้วยความสำคัญและประโยชน์ของค่าความเข้มของอิทธิพลดังกล่าว นักวิจัยส่วนใหญ่จึงเห็นพ้องต้องกันว่าควรมีการรายงานค่าความเข้มของอิทธิพลทุกครั้งที่มีการรายงานค่า p และรายงานการวิจัยที่ดีนอกจากจะมีการรายงานค่าความเข้มของอิทธิพลในส่วนของผลการวิจัยแล้วนั้นควรจะรายงานค่าความเข้มของอิทธิพลของงานวิจัยที่เกี่ยวข้องก่อนหน้าไว้ในส่วนของการทบทวน

เอกสารงานวิจัยที่เกี่ยวข้องด้วย (Snyder & Lawson, 1993; Wilkinson & APA Task Force on Statistical Inference, 1999; Kirk, 2001; Gliner, Leech & Morgan, 2002) นอกจากนี้ Capraro (2002) Thompson (2002) และ Trusty, Thompson และ Petrocelli (2004) ยังสำรวจพบว่ามีวารสารวิชาการ 23 หัวเรื่องกำหนดให้มีการรายงานค่าความเข้มของอิทธิพล อาทิ Contemporary Educational Psychology, Educational and Psychological Measurement, Journal of Applied Psychology, Journal of Community Psychology, Journal of Consulting & Clinical Psychology, Journal of Counseling and Development, Journal of Educational and Psychological Consultation, Journal of Experimental Education, Journal of Personality Assessment และ Research in the School เป็นต้น ซึ่งส่วนใหญ่เป็นนโยบายที่ได้รับอิทธิพลมาจาก APA Manual 5th edition นั่นเอง

ตอนที่ 4 คุณสมบัติของตัวประมาณค่า

เนื่องจากค่าขนาดอิทธิพลเป็นค่าที่สะท้อนคุณลักษณะ หรือปรากฏการณ์ในประชากร นักวิจัยวัดค่าโดยตรงได้ยาก จำเป็นต้องศึกษาค่าสถิติจากกลุ่มตัวอย่าง แล้วสรุปอ้างอิงไปยังค่าของประชากร นั่นคือใช้การมโนทัศน์ของสถิติอ้างอิง (inferential statistic) ซึ่งแบ่งเป็น การประมาณค่าพารามิเตอร์และการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ สำหรับการวิจัยครั้งนี้ใช้การประมาณค่าพารามิเตอร์ความเข้มของอิทธิพลด้วยขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์

Mandenhall และ Beaver (1994) และ Glass และ Hopkins (1995) กล่าวว่า ตัวประมาณค่าที่ดีควรเป็นตัวประมาณค่าที่ไม่มีความลำเอียง (unbiased) กล่าวคือ ค่าประมาณ (estimate) มีค่าเท่ากับค่าพารามิเตอร์ (parameter) หรืออย่างน้อยที่สุดควรเป็นตัวประมาณค่าที่มีความลำเอียงต่ำเมื่อเทียบกับตัวประมาณค่าชนิดอื่น ๆ นอกจากนี้จะต้องมีความคงเส้นคงวาสูง (consistency) มีความพอเพียง (sufficiency) ในการใช้ข้อมูลจากทุกหน่วยตัวอย่างมาใช้ประมาณค่า และมีประสิทธิภาพสัมพัทธ์สูง (relative efficiency) ในงานวิจัยครั้งนี้พิจารณาคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ สำหรับความพอเพียงไม่ได้นำมาพิจารณาเนื่องจากตัวประมาณค่าทุกตัวใช้ข้อมูลจากผู้สอบทั้งหมดในการคำนวณอยู่แล้ว

ความไม่ลำเอียง หมายถึง คุณสมบัติของตัวประมาณค่าที่มีค่าเฉลี่ยของการแจกแจงการสุ่มของตัวประมาณค่าต่างจากค่าพารามิเตอร์ ในงานวิจัยนี้จะพิจารณาจากความลำเอียงที่เกิดขึ้นในการประมาณค่าความเข้มของอิทธิพลโดยใช้ค่าความแตกต่างระหว่างค่าคาดหวังของตัวประมาณค่ากับค่าที่แท้จริง (expected - parameter) หรือเรียกว่าค่าความลำเอียง (BIAS) นอกจากนี้ในงานวิจัยที่ศึกษาความลำเอียงของตัวประมาณค่าโดยทั่วไปจะพิจารณาค่าส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ย (Root Mean Square Deviation: RMSD) ด้วย (Harwell et al., 1996; Pelton, 2004) แต่เนื่องจากในงานวิจัยครั้งนี้เป็นการเปรียบเทียบความลำเอียงของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ซึ่งมี

ช่วง (range) ของค่าที่แตกต่างกัน ผู้วิจัยจึงนำค่าความลำเอียงที่ได้จากขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มาแปลงให้เป็นค่าความลำเอียงมาตรฐาน (standardized bias) เพื่อใช้ในการคำนวณส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Standardized Root Mean Square Deviation: SRMSD) โดยค่าสถิติทั้งสามคำนวณได้ตามสูตรต่อไปนี้

$$BIAS = E(\hat{a}) - a \quad \text{(Hay, 1963 อ้างถึงในสุกัญญรัตน์ คงงาม, 2539)}$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(\hat{a}_i - a)^2}{N}} \quad \text{(Harwell et al., 1996; Pelton, 2004)}$$

และ $$SRMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(Standardized\ BIAS\right)^2}{N}} \quad \text{(ดัดแปลงจาก Pelton, 2004)}$$

หรือ $$SRMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{\hat{a}_i - a}{SD_{\hat{a}_i - a}}\right)^2}{N}}$$

โดยที่ a หมายถึง พารามิเตอร์ที่ต้องการประมาณค่า

$\hat{a}$ หมายถึง ค่าประมาณของพารามิเตอร์

$E(\hat{a})$ หมายถึง ค่าคาดหวังของ $\hat{a}$ หรือค่าเฉลี่ยของการแจกแจงการสุ่มของ $\hat{a}$

$SD_{\hat{a}_i - a}$ หมายถึง ส่วนเบี่ยงเบนมาตรฐานของค่าความลำเอียง (BIAS)

i หมายถึง ครั้งที่ประมาณค่า

N หมายถึง จำนวนครั้งในการประมาณค่า

เกณฑ์ในการพิจารณาเป็นเกณฑ์สัมพัทธ์ กล่าวคือ ตัวประมาณค่าที่ดีควรมีความลำเอียงต่ำ นั่นคือ มีค่า RMSD และ SRMSD ต่ำสุด สำหรับค่า BIAS จะใช้ในการพิจารณาว่าตัวประมาณค่าเป็นตัวประมาณค่าที่ให้ค่าสูงกว่าค่าพารามิเตอร์ (overestimate) หรือต่ำกว่าพารามิเตอร์ (underestimate) โดยค่า BIAS ที่เป็นลบแสดงว่าตัวประมาณค่าให้ค่าประมาณที่ต่ำกว่าค่าจริง ในทางกลับกันค่า BIAS ที่เป็นบวกแสดงว่าตัวประมาณค่าให้ค่าประมาณที่สูงกว่าค่าจริง สำหรับค่า RMSD ผู้วิจัยจะนำเสนอเฉพาะผลการวิจัยในภาพรวมเพื่อชี้ให้เห็นว่าการเปรียบเทียบผลการพิจารณาความลำเอียงของตัวประมาณค่าที่มีช่วงของค่าต่างกันโดยใช้ RMSD และ SRMSD เหมือนหรือแตกต่างกัน

ความคงเส้นคงวา หมายถึง คุณลักษณะของตัวประมาณค่าที่มีแนวโน้มในการให้ค่าประมาณเข้าใกล้ค่าพารามิเตอร์เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น (an estimator tends to get closer to the value of the parameter as the sample size becomes larger) หรือเมื่อขนาดกลุ่มตัวอย่างลู่เข้าสู่

อนันต์ ($N \rightarrow \infty$) ในการวิจัยครั้งนี้ ความคงเส้นคงวาของตัวประมาณค่าความเข้มของอิทธิพลจะคำนวณด้วยสูตรที่ปรับจาก สุกัญญรัตน์ คงงาม (2539)

$$\Delta = SRMSD_{Nmin} - SRMSD_{Nmax}$$

โดยที่ Δ หมายถึง ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน

$SRMSD_{Nmin}$ หมายถึง ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานในกรณีที่กลุ่มตัวอย่างมีขนาดเล็กที่สุด (ในการวิจัยครั้งนี้ Nmin มีค่า 20)

$SRMSD_{Nmax}$ หมายถึง ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ที่สุด (ในการวิจัยครั้งนี้ Nmax มีค่า 2000)

เกณฑ์ในการพิจารณาเป็นเกณฑ์สัมพัทธ์ กล่าวคือ ตัวประมาณค่าที่ดีควรมีความคงเส้นคงวาสูง นั่นคือ มีค่า Δ สูงกว่าตัวประมาณค่าอื่น

ความพอเพียง หมายถึง คุณสมบัติของตัวประมาณค่าที่สามารถใช้ข้อมูลจากทุกหน่วยตัวอย่างที่มีในการประมาณค่าพารามิเตอร์ในประชากร แต่เนื่องจากการวิจัยครั้งนี้ผู้วิจัยใช้ข้อมูลที่ได้จากการจำลองสถานการณ์จึงสามารถประมาณค่าพารามิเตอร์ได้จากข้อมูลทั้งหมดที่มี คุณสมบัติความพอเพียงจึงไม่นำมาพิจารณา

ประสิทธิภาพสัมพัทธ์ หมายถึง คุณสมบัติของตัวประมาณค่าในด้านความถูกต้องสัมพัทธ์ของการประมาณค่า หรือระดับของความคลาดเคลื่อนของการสุ่มเมื่อเทียบกับตัวประมาณค่าชนิดอื่นเมื่อขนาดกลุ่มตัวอย่างเท่ากัน (Glass & Hopkins, 1995) ในงานวิจัยครั้งนี้ ประสิทธิภาพสัมพัทธ์พิจารณาจากการเปรียบเทียบความแปรปรวนของค่าประมาณความเข้มของอิทธิพลที่ได้จากตัวประมาณค่าแต่ละชนิดในสถานการณ์เดียวกัน โดยใช้เกณฑ์สัมพัทธ์ในการพิจารณา กล่าวคือ ตัวประมาณค่าที่ดีควรมีประสิทธิภาพสัมพัทธ์สูง หรือมีความแปรปรวนต่ำกว่าตัวประมาณค่าอื่น

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยแบ่งงานการวิจัยที่เกี่ยวข้องออกเป็น 3 กลุ่ม ได้แก่ (1) งานวิจัยเปรียบเทียบค่าสถิติจากทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ (2) งานวิจัยเกี่ยวกับปัจจัยที่ส่งผลต่อการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ และ (3) งานวิจัยเกี่ยวกับการประมาณค่าความเข้มของอิทธิพล รายละเอียดของสาระจากงานวิจัยในแต่ละกลุ่มมีดังนี้

1. งานวิจัยเปรียบเทียบค่าสถิติจากทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ

จิติมา วรรณศรี (2539) เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีชิบเทสท์ โดยใช้ข้อมูลที่จำลองขึ้นจากโปรแกรม IRTDATA ภายใต้เงื่อนไขของความยาวแบบสอบ (30, 60 และ 90 ข้อ) และขนาดกลุ่มตัวอย่าง

(200, 600 และ 1,000 คน) มีอัตราส่วนระหว่างผู้สอบกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบ 4 อัตราส่วน (1:1, 1:.9, 1:.75 และ 1:.5) รวม 36 สถานการณ์ ผลการวิจัยพบว่าวิธีแมนเทลแฮนส์เชล และวิธีซิบเทสท์มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ทุกขนาดกลุ่มตัวอย่างและทุกอัตราส่วนภายใต้ความยาวแบบสอบเดียวกัน เมื่อขนาดกลุ่มตัวอย่าง 200 และ 600 คน สามารถตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 50 และเมื่อขนาดกลุ่มตัวอย่าง 1,000 คน สามารถตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 100 โดยส่วนมากวิธีซิบเทสท์มีอัตราความคลาดเคลื่อนชนิดที่ 1 มากกว่าวิธีแมนเทล-แฮนส์เชลเล็กน้อย นอกจากนี้วิธีแมนเทล-แฮนส์เชลกับวิธีซิบเทสท์มีประสิทธิภาพเท่าเทียมกันทุกความยาวแบบสอบ โดยพบว่าเมื่อแบบสอบที่ยาวปานกลาง (60 ข้อ) ทั้งสองวิธีมีประสิทธิภาพในการตรวจสอบดีที่สุด

นภดล ยิ่งยงสกุล (2539) เปรียบเทียบค่าความเที่ยง ความตรง และความสัมพันธ์ของคะแนนสอบวิชาคณิตศาสตร์ ชนิด 4 ตัวเลือก จำนวน 28 ข้อ ที่ผู้วิจัยสร้างขึ้น ระหว่างค่าที่ได้จากการให้คะแนนตามทฤษฎีการทดสอบดั้งเดิม กับการให้คะแนนตามทฤษฎีการตอบสนองข้อสอบ กลุ่มตัวอย่างที่ใช้ได้มาจากการสุ่มอย่างง่ายเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2539 ในเขตจังหวัดยะลา 231 คน ผลการวิจัยพบว่าค่าความเที่ยงของแบบทดสอบจากวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามความสามารถของผู้สอบสูงกว่าวิธีการให้น้ำหนักรายข้อต่างกันตามค่าพารามิเตอร์ของข้อสอบ และวิธีการให้น้ำหนักคะแนนรายข้อเท่ากันอย่างมีนัยสำคัญ ทางสถิติที่ระดับ .01 ส่วนวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ของข้อสอบ มีค่าสูงกว่าวิธีการให้น้ำหนักคะแนนรายข้อเท่ากันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 สำหรับค่าความตรงของแบบทดสอบจากวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามค่าพารามิเตอร์ และวิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามความสามารถของผู้สอบ ไม่พบว่ามีค่าแตกต่างกันที่ระดับนัยสำคัญทางสถิติ .01 นอกจากนี้ค่าความสัมพันธ์ของคะแนนสอบระหว่างวิธีการให้น้ำหนักคะแนนรายข้อเท่ากันกับวิธีการให้น้ำหนักคะแนนรายข้อต่าง กันตามค่าพารามิเตอร์ และความสัมพันธ์ของคะแนนสอบระหว่างวิธีการให้น้ำหนักคะแนนรายข้อเท่ากัน วิธีการให้น้ำหนักคะแนนรายข้อต่างกันตามความสามารถของผู้สอบมีค่าสูงมาก (.98 - .99)

เบญจพร ยนตจักรวิถี (2539) ได้เปรียบเทียบผลที่ได้จากการวิเคราะห์ข้อสอบระหว่างการใช้ทฤษฎีการทดสอบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ (โมเดลของราส์ช) ในเรื่องสัดส่วนของจำนวนข้อสอบที่ได้คัดเลือกไว้ ความสอดคล้องในการคัดเลือกข้อสอบ ความสัมพันธ์ระหว่างค่าความยากของข้อสอบและคะแนน ความสามารถของผู้สอบ กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2539 จังหวัดสงขลา จำนวน 500 คน แบบทดสอบที่ใช้เป็นชนิดเลือกตอบ 4 ตัวเลือก เรื่องอัตราส่วนและร้อยละ 33 ข้อ ผลการวิจัยพบว่าจำนวนข้อสอบที่ได้คัดเลือกไว้จากการวิเคราะห์ด้วยทฤษฎีการทดสอบดั้งเดิมกับโมเดลของราส์ช มี 32 และ 18 ข้อ

ตามลำดับ เมื่อเปรียบเทียบสัดส่วนของจำนวนที่ได้คัดเลือกไว้พบว่าจำนวน ข้อสอบที่ได้คัดเลือกไว้จากการวิเคราะห์ด้วยทฤษฎีการทดสอบดั้งเดิมมีมากกว่าโมเดลของราส์ชอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 สำหรับการคัดเลือกข้อสอบจากการวิเคราะห์ด้วยทฤษฎีการทดสอบดั้งเดิมกับโมเดลของราส์ชไม่พบว่ามีความสอดคล้องกัน นอกจากนี้ค่าความยากของข้อสอบที่ได้จากการวิเคราะห์ด้วยทฤษฎีการทดสอบดั้งเดิมกับโมเดลของราส์ชมีความสหสัมพันธ์กันสูงมาก (r = .99) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เช่นเดียวกับคะแนนความสามารถของผู้สอบที่ได้จากการวิเคราะห์ด้วยทฤษฎีการทดสอบดั้งเดิมกับโมเดลของราส์ชมีความสหสัมพันธ์กันสูง (r = .93) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ชวลิต นิวาสวัต (2541) ได้วิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวิชาความรู้พื้นฐานทั่วไปที่ใช้สอบคัดเลือกเข้าศึกษาต่อระดับปริญญาตรีของสถาบันราชภัฏจะเชิงเทรา โดยใช้วิธีค่าอำนาจจำแนกของข้อสอบ วิธีแมนเทล-แฮนส์เซลและวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ในการวิเคราะห์ และเปรียบเทียบผลที่ได้จากแต่ละวิธี กลุ่มตัวอย่างเป็นผู้เข้าสอบคัดเลือกเพื่อศึกษาต่อระดับปริญญาตรีของสถาบันราชภัฏจะเชิงเทราประจำปีการศึกษา 2541 จำนวน 1,875 คน ที่ได้จากการสุ่มอย่างง่ายจากผู้เข้าสอบ 3,067 คน ผู้วิจัยวิเคราะห์ความลำเอียงตามตัวแปรเพศ และเขตที่ตั้งของสถานศึกษา จากนั้นจึงหาค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงที่ได้จากการวิเคราะห์ และเปรียบเทียบจำนวนข้อสอบที่ลำเอียงที่ได้จากการวิเคราะห์ทั้ง 3 วิธี ผลการวิจัยพบว่าการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลมีความไวสูงสุด พบข้อสอบที่ลำเอียงตามเพศ 37 ข้อ ตามเขตที่ตั้ง 12 ข้อ วิธีค่าอำนาจจำแนกของข้อสอบ พบข้อสอบที่ลำเอียงตามเพศ 11 ข้อ ตามเขตที่ตั้ง 1 ข้อ วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ พบข้อสอบที่ลำเอียงตามเพศ 24 ข้อ ตามเขตที่ตั้ง 8 ข้อ นอกจากนี้ยังพบว่า ดัชนีจากวิธีแมนเทล-แฮนส์เซล และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และจำนวนข้อสอบที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยเฉพาะวิธีแมนเทล-แฮนส์เซลพบจำนวนข้อสอบที่ลำเอียงสูงกว่าวิธีค่าอำนาจจำแนกของข้อสอบอย่างมีนัยสำคัญทางสถิติที่ .01

อรวรรณ สุขโต (2542) ได้เปรียบเทียบผลการวิเคราะห์ข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ที่มีรูปแบบการตอบต่างกัน 3 รูปแบบ คือ แบบทดสอบเลือกตอบแบบเลือก หลายคำตอบ แบบทดสอบเลือกตอบแบบตัวเลือกถูกผิดและแบบทดสอบเลือกตอบแบบตัวเลือก ธรรมดาด้วยวิธีการวิเคราะห์ตามโมเดลการทดสอบแบบดั้งเดิมและโมเดลราส์ช โดยเปรียบเทียบจำนวนข้อที่คัดเลือกได้จากแต่ละวิธี และค่าความเชื่อมั่นของแบบทดสอบที่คัดเลือกได้จากการคัดเลือกข้อสอบในแต่ละวิธี กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2542 จังหวัดปทุมธานี ซึ่งได้มาโดยการสุ่มแบบ 2 ขั้นตอน จำนวน 2,470 คน ผลการวิจัยพบว่าจำนวนข้อที่คัดเลือกได้ระหว่างการวิเคราะห์ตามโมเดลการทดสอบแบบดั้งเดิมและโมเดลราส์ช

ของแบบทดสอบเลือกตอบแบบเลือกหลายคำตอบ แบบตัวเลือกถูกผิด และแบบตัวเลือกธรรมดา ไม่พบว่ามีจำนวนข้อแตกต่างกันที่ระดับนัยสำคัญทางสถิติ .05 และความเชื่อมั่นของแบบทดสอบเลือกตอบที่คัดเลือกได้ระหว่างการวิเคราะห์ตามโมเดลการทดสอบแบบดั้งเดิมและโมเดลราส์ชในแบบทดสอบเลือกตอบแบบเลือกหลายคำตอบมีค่าแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนในแบบทดสอบเลือกตอบแบบตัวเลือกถูกผิด และแบบตัวเลือกธรรมดาไม่พบว่ามีจำนวนข้อแตกต่างกันที่ระดับนัยสำคัญทางสถิติ .05 นอกจากนี้ค่าความเชื่อมั่นของแบบทดสอบเลือกตอบที่คัดเลือกได้จากโมเดลการทดสอบแบบดั้งเดิม ระหว่างแบบทดสอบเลือกตอบแบบตัวเลือกถูกผิดกับแบบตัวเลือกธรรมดา แบบตัวเลือกถูกผิดกับแบบ เลือกหลายคำตอบมีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบเลือกตอบแบบตัวเลือกธรรมดากับแบบเลือกหลายคำตอบมีค่าความเชื่อมั่น ไม่พบว่าค่าความเชื่อมั่นมีความแตกต่างกันที่ระดับนัยสำคัญทางสถิติ .05 สำหรับแบบทดสอบเลือกตอบที่คัดเลือกได้จากโมเดลราส์ช ระหว่างแบบทดสอบเลือกตอบแบบเลือกหลายคำตอบ แบบตัวเลือกถูกผิด และแบบตัวเลือก ธรรมดาไม่พบว่าค่าความเชื่อมั่นมีจำนวนข้อแตกต่างกันที่ระดับนัยสำคัญทางสถิติ .05

Lawson (1991 อ้างถึงใน Fan, 1998) ได้เปรียบเทียบค่าความยากและค่าความสามารถของผู้สอบที่ได้จากวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ โดยใช้โมเดลราส์ชและข้อมูลดิบ 3 ชุด พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าที่ได้จากทั้งสองทฤษฎีมีความสัมพันธ์กันสูงมาก

Ndalichako และ Rogers (1997) ได้เปรียบเทียบค่าความสามารถของผู้สอบที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ โดยใช้คะแนนสอบอ่านจับใจความจำนวน 70 ข้อของผู้สอบ 1,230 คน โดยใช้โปรแกรม ITMAN คำนวณคะแนนรวมและใช้เป็นค่าความสามารถตามทฤษฎีการทดสอบแบบดั้งเดิม ใช้โปรแกรม PARSCALE ประมาณค่าความสามารถตามโมเดลการตอบสนองข้อสอบแบบโลจิสติกหนึ่ง สองและสามพารามิเตอร์ พบว่าค่าความสามารถของผู้สอบที่ได้จากทั้ง 4 วิธีมีความสัมพันธ์อยู่ในช่วง .977 ถึง .994 และเมื่อแปลงค่าให้อยู่ในรูปคะแนนมาตรฐานที (T score) แล้วคำนวณค่าเฉลี่ยสัมบูรณ์ของผลต่างของแต่ละวิธี พบว่าค่าความสามารถจากแต่ละวิธีแตกต่างกันเล็กน้อย (.77 - 1.53)

Stage (1998) ได้เปรียบเทียบค่าความยาก อำนาจจำแนกที่ได้จากการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ และศึกษาไม่แปรเปลี่ยนของค่าสถิติดังกล่าวระหว่างการสอบ 2 ครั้งในปี ค.ศ.1996 และ ค.ศ.1997 โดยใช้ข้อมูลจริงจากฐานข้อมูลจากการสอบ SweSAT (Sweden Scholastic Aptitude Test) เฉพาะชุดแบบสอบย่อย WORD (vocabulary subtest) จำนวน 20 ข้อ จากผู้สอบ 1,500 คน พบว่า ค่าสถิติที่ได้จากการวิเคราะห์

ข้อสอบ 13 ข้อ จาก 20 ข้อ ตามทฤษฎีทั้งสองนั้นให้ค่าที่คงเส้นคงวาระหว่างปี สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติของข้อสอบที่วิเคราะห์ได้จากทั้งสองทฤษฎีมีความสัมพันธ์กันสูง โดยมีค่าสัมประสิทธิ์สำหรับค่าความยากและค่าอำนาจจำแนกเท่ากับ -.92 และ .65 ตามลำดับ สัมประสิทธิ์สหสัมพันธ์ของค่าความยากระหว่างปีสำหรับทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบมีค่าเท่ากับ .93 และ .92 ตามลำดับ ขณะที่ค่าอำนาจจำแนกมีค่าสัมพันธ์ประสิทธิสหสัมพันธ์ระหว่างปีเท่ากับ .81 และ .74 สำหรับทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบตามลำดับ แสดงให้เห็นว่ามีความไม่แปรเปลี่ยนของค่าที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม

Fan (1998) ได้เปรียบเทียบค่าความยาก ค่าอำนาจจำแนกและค่าความสามารถของผู้สอบที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ ในด้านความสัมพันธ์ของสถิติระหว่างทฤษฎี และความไม่แปรเปลี่ยนของค่าสถิติที่ได้จากผู้สอบต่างกลุ่ม โดยการจำลองข้อมูลจากฐานข้อมูลผลการวัดทักษะทางการเรียนรัฐเท็กซัส (Texas Assessment of Academic Skill) ปีค.ศ.1992 ซึ่งมีผู้สอบ 193,000 คน แบบสอบแบ่งเป็นวิชาคณิตศาสตร์ 60 ข้อ และวิชาการอ่าน 48 ข้อ ภายใต้เงื่อนไขของกลุ่มตัวอย่างในการวิเคราะห์ 3 ชนิด คือ กลุ่มที่สุ่มอย่างง่าย กลุ่มที่สุ่มตามเพศ กลุ่มที่สุ่มตามระดับความสามารถ (สูง, ต่ำ) ทั้งหมด 10 กลุ่ม กลุ่มละ 2 วิชา รวม 20 เงื่อนไข และสร้างข้อมูลคำตอบขึ้นมาใหม่สำหรับแต่ละสถานการณ์จำนวน 1,000 คน แล้วนำมาวิเคราะห์ค่าความสามารถของผู้สอบตามโมเดลโลจิสติกแบบหนึ่ง สองและสามพารามิเตอร์ด้วยโปรแกรม BILOG สำหรับทฤษฎีการตอบสนองข้อสอบและใช้คะแนนดิบเป็นคะแนนจริงตามทฤษฎีการทดสอบแบบดั้งเดิม (…the obtained score *T* in CTT was simply the row score) ค่าความยากคำนวณจากสัดส่วนของผู้ที่ตอบถูกต่อผู้สอบทั้งหมดและค่าอำนาจจำแนกเท่ากับค่าสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียลที่ปรับแก้แล้ว จากนั้นจึงหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันระหว่างค่าสถิติที่ได้จากทั้งสองทฤษฎี ผลการศึกษาพบว่าค่าความสามารถของผู้สอบที่ได้จากทั้งสองทฤษฎีมีความสัมพันธ์กันสูง (.97 - .99) ค่าความยากจากทั้งสองทฤษฎีมีความสัมพันธ์กันสูงมาก (.80 - .99) แต่ค่าอำนาจจำแนกในโมเดลสอง และสามพารามิเตอร์มีความสัมพันธ์กับค่าอำนาจจำแนกในทฤษฎีการทดสอบแบบดั้งเดิมปานกลางค่อนข้างสูง (.51 - .95) นอกจากนี้ยังพบว่าความสัมพันธ์ระหว่างค่าสถิติระหว่างกลุ่มในทฤษฎีการทดสอบแบบดั้งเดิมมีค่าสูงกว่าในทฤษฎีการตอบสนองข้อสอบทุกโมเดล นั่นหมายถึง ทฤษฎีการทดสอบแบบดั้งเดิมสามารถให้ค่าสถิติที่มีความไม่แปรเปลี่ยนของค่าได้ไม่น้อยไปกว่าทฤษฎีการตอบสนองข้อสอบ

MacDonald และ Paunonen (2002) ได้ศึกษาเปรียบเทียบค่าสถิติของข้อสอบและของผู้สอบที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ โดยใช้การจำลองสถานการณ์ตามเงื่อนไขของจำนวนข้อสอบ (20, 40, 60 ข้อ) ค่าอำนาจจำแนก 3

ช่วง (1 - 2, .5 - 2.5, -1 - 1) ค่าความยาก 5 ช่วง (-2 - 2, -5 - 5, -2 - 1, -1 - 2, -1 - 1) ผู้วิจัยจำลองข้อมูลโดยใช้โมเดลโลจิสติกแบบสองพารามิเตอร์เป็นโมเดลฐาน มีวิธีการจำลองข้อมูล 4 ขั้นตอนคือ 1) จำลองค่าความสามารถที่แท้จริงของผู้สอบ 1,000 คนให้มีการแจกแจงปกติ 2) จำลองข้อสอบตามเงื่อนไขทั้งสาม 3) คำนวณโอกาสในการตอบถูกของผู้สอบในขั้นที่ 1 ในการสอบข้อสอบขั้น 2 โดยใช้โมเดลโลจิสติกแบบสองพารามิเตอร์ และ 4) จำลองคำตอบแต่ละข้อของผู้สอบแต่ละคนโดยสุ่มจำนวนจริงในช่วง 0 ถึง 1 แล้วเทียบกับโอกาสในการตอบถูกที่ได้จากขั้นที่ 3 ถ้าเลขสุ่มน้อยกว่าโอกาสฯ จำกำหนดให้คำตอบข้อนั้นเป็น 1 หรือได้ 1 คะแนน กรณีอื่นเป็น 0 และใช้ขั้นตอนดังกล่าวจำลองข้อมูลคำตอบสำหรับผู้สอบ 2 กลุ่มเพื่อใช้เปรียบเทียบความไม่แปรเปลี่ยนของค่าสถิติ ผู้วิจัยใช้โมเดลโลจิสติกแบบหนึ่งและสองพารามิเตอร์ในการวิเคราะห์ข้อสอบจากข้อมูลที่จำลองขึ้น และใช้คะแนนดิบเป็นค่าความสามารถที่แท้จริงของผู้สอบตามทฤษฎีการทดสอบแบบดั้งเดิม ทำซ้ำ 100 รอบ ผลการวิจัยพบว่า ค่าความสามารถและความยากที่ได้จากทั้งสองทฤษฎีมีความสัมพันธ์กันสูงมาก (.97-.99 และ .92-.99) สำหรับความไม่แปรเปลี่ยนของค่าสถิติทุกค่าพบว่าทฤษฎีการทดสอบแบบดั้งเดิมให้ค่าที่น่าพอใจ โดยเฉพาะค่าอำนาจจำแนกไม่แปรเปลี่ยนตามกลุ่มผู้สอบดีกว่าค่าจากทฤษฎีการตอบสนองข้อสอบ นอกจากนี้สหสัมพันธ์ระหว่างค่าความสามารถที่แท้จริง กับค่าความสามารถที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมมีค่าต่ำกว่าค่าจากทฤษฎีการตอบสนองข้อสอบเล็กน้อย (.95, .97)

Stage (2003) เสนอรายงานสรุปผลโครงการวิจัยเปรียบเทียบค่าสถิติที่ได้จากการใช้ทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบสำหรับการวิเคราะห์ข้อสอบจากแบบสอบย่อย 3 ชุดของแบบสอบความถนัดทางวิชาการของประเทศสวีเดน (SweSAT) ได้แก่ แบบสอบชุดคำศัพท์ (WORD) แบบสอบชุดการอ่าน (READ) และแบบสอบชุดการอ่านภาษาอังกฤษ (ERC) ระหว่างการสอบในปี ค.ศ.1996 และ ค.ศ.1997 มีข้อสอบร่วม 20, 16, 14 ข้อ ตามลำดับแบบสอบผู้สอบ 2,461 คน โดยเปรียบเทียบค่าความยากและอำนาจจำแนก พบว่า สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความยากของแบบสอบแต่ละฉบับที่ได้จากทั้งสองทฤษฎีมีค่าสูงแต่เป็นความสัมพันธ์แบบผกผัน (-.93, -.92, -.90) และสัมประสิทธิ์สำหรับค่าอำนาจจำแนกระหว่างทฤษฎีมีค่าปานกลาง (.65, .68, .76)

Wilberg (2004) ทำการศึกษาความสัมพันธ์ระหว่างค่าสถิติที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ (ค่าความยากและค่าอำนาจจำแนก) โดยใช้แบบสอบขอใบอนุญาตขับรถของประเทศสวีเดน (Swedish driving-license test) ชนิดสี่ตัวเลือก จำนวน 65 ข้อ ผู้สอบจำนวน 5404 คน โดยวิเคราะห์ค่าความยากและอำนาจจำแนกตามทฤษฎีการตอบสนองข้อสอบด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ ผลการวิจัยพบว่า สัมประสิทธิ์สหสัมพันธ์

ระหว่างค่าความยากที่ได้จากทั้งสองทฤษฎีการทดสอบมีค่า -.861 และสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าอำนาจจำแนกที่ได้จากทั้งสองทฤษฎีมีค่า .753

ผลการวิจัยส่วนใหญ่ที่นำเสนอในส่วนนี้เป็นไปในทางเดียวกัน คือ ค่าสถิติที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองมีความสัมพันธ์กันในทางบวกค่อนข้างสูง โดยเฉพาะค่าความสามารถของผู้สอบและค่าความยาก ในขณะที่ค่าอำนาจจำแนกของข้อสอบที่วิเคราะห์ได้จากทั้งสองทฤษฎีมีความสัมพันธ์กันปานกลาง สำหรับค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความยากที่ได้จากทั้งสองทฤษฎีการทดสอบมีค่าเป็นลบพบในงานวิจัยของ Stage (1998; 2003) และ Wilberg (2004) ซึ่งเป็นงานวิจัยจากหน่วยงานเดียวกันในประเทศสวีเดน

2. งานวิจัยเกี่ยวกับปัจจัยที่ส่งผลต่อการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ

Kirisci, Hsu และ Yu (2001) เปรียบเทียบความแกร่งของการประมาณค่าพารามิเตอร์ของข้อสอบ เมื่อข้อมูลไม่เป็นเอกมิติและการแจกแจงไม่เป็นปกติ ระหว่างโปรแกรม BILOG MULTILOG และ XCALIBRE โดยจำลองข้อมูลตามเงื่อนไขของมิติของแบบสอบ (1, 3 มิติ) การแจกแจงของค่าความสามารถที่แท้จริงของผู้สอบ (ปกติ, เบ้ขวา) เมื่อข้อสอบ 40 ข้อ ผู้สอบ 1,000 คน ทำซ้ำ 10 รอบ และประมาณค่าพารามิเตอร์ด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ โดยกำหนดค่าอำนาจจำแนกที่แท้จริงมีค่า .4 ถึง 2 ค่าความยากที่แท้จริงมีค่า -2.0 ถึง 2.0 และค่าโอกาสในการเดาที่แท้จริงมีค่า 0 ถึง .3 ผลพบว่า BILOG มีความคลาดเคลื่อนต่ำที่สุดทุกสถานการณ์ รองลงมาคือ XCALIBRE และ MULTILOG

Leeson และ Fletcher (2003) ได้เปรียบเทียบความสอดคล้องระหว่างข้อมูลกับโมเดลโลจิสติกแบบหนึ่ง สอง และสามพารามิเตอร์ ใช้ข้อมูลจากการสอบอ่านในโครงการ asTTle Literacy ที่มีจำนวนข้อสอบ 43 ข้อจากผู้สอบ 17,702 คน กลุ่มตัวอย่างได้มาจากการสุ่มอย่างง่าย 326 คน เมื่อวิเคราะห์ข้อสอบด้วยโปรแกรม BILOG-W แล้วจึงทำการตรวจสอบความสอดคล้องระหว่างข้อมูลกับแต่ละโมเดลด้วยการทดสอบค่าพารามิเตอร์ของข้อสอบ ได้แก่ การทดสอบความเท่ากันของค่าอำนาจจำแนกของข้อสอบแต่ละข้อสำหรับโมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ การวิเคราะห์ค่าโอกาสในการเดาที่น้อยที่สุดสำหรับโมเดลโลจิสติกแบบสามพารามิเตอร์ และการทดสอบความไม่แปรเปลี่ยนของค่าความยากสำหรับโมเดลทั้งสามแบบ พบว่า ค่าอำนาจจำแนกของข้อสอบมีค่าแตกต่างกันสูงมากคือมีค่าตั้งแต่ -.01 ถึง .85 สำหรับค่าโอกาสในการเดานั้นร้อยละ 96.6 ของกลุ่มผู้สอบที่มีความสามารถต่ำได้คะแนนต่ำกว่าร้อยละ 50 ในข้อสอบที่มีค่าความยากสูง และค่าความยากที่ได้จากทั้งสามโมเดลมีความไม่แปรเปลี่ยนของค่าระหว่างกลุ่มผู้สอบย่อย 2 กลุ่ม (.97 - .99)

ผู้วิจัยสรุปในตอนท้ายว่าโมเดลโลจิสติกแบบสองพารามิเตอร์มีความสอดคล้องกับข้อมูลสูงที่สุด รองลงมาคือโมเดลโลจิสติกแบบสามพารามิเตอร์และหนึ่งพารามิเตอร์ตามลำดับ

Stone และ Yumoto (2004) ได้ศึกษาอิทธิพลของขนาดกลุ่มผู้สอบที่มีต่อค่าพารามิเตอร์ที่ได้จากการประมาณค่าด้วยโมเดลราส์ช โลจิสติกแบบสองและสามพารามิเตอร์ จากฐานข้อมูลของผู้สอบ 3,173 คนที่ตอบแบบสอบของ Knox (Knox's Cube Test-Revised: KCT-R) 26 ข้อโดยสุ่มข้อมูลให้ได้กลุ่มผู้สอบ 13 ขนาด คือ 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 1,000, 2,000 และ 3,000 คน จากนั้นจึงประมาณค่าพารามิเตอร์ตามโมเดลราส์ชด้วยโปรแกรม WINSTEP สำหรับโมเดลโลจิสติกแบบสองและสามพารามิเตอร์ใช้ภาษา ICL หรือ IRT Common Language เขียนคำสั่งการประมาณแบบ MML/EM และ Bayes/EM ผลการศึกษาพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความยากที่ได้จากกลุ่มผู้สอบแต่ละขนาดมีค่าเข้าใกล้ 1 และมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับค่าความคลาดเคลื่อนมาตรฐานจะมีค่าลดลงเมื่อกลุ่มผู้สอบมีขนาดใหญ่ขึ้น และแตกต่างกันน้อยมากเมื่อกลุ่มผู้สอบมีขนาดตั้งแต่ 500 คนขึ้นไป ค่าความยากที่ได้จากโมเดลโลจิสติกแบบสองพารามิเตอร์มีความลู่เข้าของค่าสูงกว่าโมเดลโลจิสติกแบบสามพารามิเตอร์เล็กน้อย แต่สูงกว่าโมเดลราส์ชมากเมื่อกลุ่มผู้สอบมีขนาดใหญ่กว่า 500 คน และเมื่อกลุ่มผู้สอบมีขนาดเท่ากันค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าอำนาจจำแนกที่ได้จากโมเดลโลจิสติกแบบสอง และสามพารา-มิเตอร์มีค่าสูง (.77 - 1.00) แต่ค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าติดลบเมื่อโมเดลหนึ่งใช้กลุ่มผู้สอบขนาดเล็ก (10 - 100) แต่อีกโมเดลหนึ่งใช้กลุ่มผู้สอบขนาดใหญ่ (500 - 1,000) นอกจากนี้โมเดลโลจิสติกแบบสองพารามิเตอร์สอดคล้องกับกลุ่มผู้สอบได้ดีที่สุด (พิจารณาค่า -2loglikelihood) รองลงมาคือโมเดลโลจิสติกแบบสามพารามิเตอร์ และโมเดลราส์ช ทั้งยังพบว่าค่าความสอดคล้องระหว่างโมเดลกับข้อมูลจะมีค่าสูงขึ้นเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น

3. งานวิจัยเกี่ยวกับการประมาณค่าความเข้มของอิทธิพล

DeMars (2001) ได้เปรียบเทียบความเข้มของอิทธิพลด้วยสูตรค่าขนาดอิทธิพลของ Cohen ระหว่างข้อมูลที่วัดตามทฤษฎีการตอบข้อสอบระหว่างโมเดลราส์ช และโมเดลโลจิสติกแบบสามพารามิเตอร์ โดยจำลองค่าความสามารถของผู้สอบตามเงื่อนไขของความยาวข้อสอบ 3 แบบ (15, 30, 60 ข้อ) ขนาดกลุ่มทดลอง 3 ขนาด (10% 30% 50% ของขนาดกลุ่มควบคุม) ให้กลุ่มควบคุมมีค่าเฉลี่ยเท่ากับ 0 มีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ทั้งกลุ่มควบคุมและกลุ่มทดลอง ค่าขนาดอิทธิพลตั้งต้น 6 ค่า (-1, -.5, -.2, .2, .5 และ 1) และวิธีการประมาณค่า 2 วิธี (EAP และ MLE) รวม 108 สถานการณ์ แต่ละสถานการณ์จะจำลองพารามิเตอร์ความสามารถที่แท้จริงของผู้สอบ 2,000 คน โดยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์เป็นโมเดลฐานในการจำลองข้อมูล จากนั้นจึงจำลองข้อมูลคำตอบตามความสามารถของผู้สอบตามเงื่อนไขของพารามิเตอร์ของข้อสอบ ได้แก่ ความยาก อำนาจจำแนกและโอกาสในการเดาถูก นำคำตอบที่จำลองได้มาประมาณค่าพารามิเตอร์

ความสามารถของผู้สอบด้วยวิธี EAP และ MLE เพื่อเปรียบเทียบกับค่าขนาดอิทธิพลตั้งต้นและตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูลด้วยการทดสอบไค-สแควร์ พบว่า โมเดลโลจิส-ติกแบบสามพารามิเตอร์สอดคล้องกับข้อมูลมากกว่าโมเดลราส์ชตรงตามที่ผู้วิจัยต้องการ โดยค่าขนาดอิทธิพลที่ได้จากโมเดลราส์ชและโมเดลโลจิสติกแบบสามพารามิเตอร์ทุกกรณีมีค่าต่างกันเล็กน้อย (.01 - .20) โดยเมื่อกลุ่มผู้สอบมีขนาดใหญ่ขึ้นความแตกต่างระหว่างค่าที่ได้จากทั้ง 2 โมเดลจะเพิ่มมากขึ้น สำหรับการประมาณค่าทั้ง 2 วิธีให้ค่าที่ต่างกันเล็กน้อย (.01 - .02) โดย EAP จะประมาณค่าสูงกว่า MLE และให้ค่าที่ใกล้เคียงกับค่าความเข้มของอิทธิพลที่แท้จริงมากกว่า

Roberts และ Henson (2002) ได้จำลองสถานการณ์เพื่อเปรียบเทียบความลำเอียงของขนาดอิทธิพลตามสูตรของ Cohen ที่คำนวณโดยตรงกับการแปลงค่าของ Thompson ($d \rightarrow r \rightarrow r^2 \rightarrow$ adjusted $r^2 \rightarrow \sqrt{\text{adjusted } r^2} \rightarrow$ adjusted $r \rightarrow$ adjusted d) โดยใช้ค่าพารามิเตอร์ขนาดอิทธิพล 4 ค่า (0, .2, .5, .8) ค่าพารามิเตอร์ส่วนเบี่ยงเบนมาตรฐาน 3 ค่า (2, 15, 30) ขนาดของกลุ่มตัวอย่างทั้งกลุ่มควบคุมและกลุ่มทดลอง 4 ค่า (10, 20, 50, 100) เงื่อนไขในการจำลองรวม 48 สถานการณ์ ทำซ้ำ 5,000 รอบ โดยกำหนดให้ค่าเฉลี่ยในกลุ่มควบคุมมีค่าคงที่เท่ากับ 100 และค่าเฉลี่ยของกลุ่มทดลองแปรไปตามค่าพารามิเตอร์ขนาดอิทธิพลและส่วนเบี่ยงเบนมาตรฐานในแต่ละสถานการณ์ ผลการวิจัยพบว่าค่าขนาดอิทธิที่คำนวณโดยตรงมีค่าสูงกว่าค่าขนาดอิทธิพลตามวิธีของ Thompson แต่มีความลำเอียงโดยเฉลี่ยต่ำกว่าค่าขนาดอิทธิพลตามวิธีของ Thompson เล็กน้อย

Wang และ Chen (2004) ได้นิยามสถิติ d_θ สำหรับประมาณค่าความเข้มของอิทธิพลโดยปรับมาจาก Cohen's d ผู้วิจัยได้เสนอวิธีการปรับแก้การกดค่า (attenuation) สำหรับบริบทของทฤษฎีการตอบสนองข้อสอบ ($adjusted\ d_\theta = d_\theta \Big/ \sqrt{\hat{\sigma}^2_{EAP} \Big/ \hat{\sigma}^2_\theta}$) วิธีดำเนินการวิจัยมี 2 ส่วน ส่วนที่ 1 เป็นการจำลองข้อมูลเพื่อทดสอบความถูกต้องของสูตรที่พัฒนาขึ้น ใช้เงื่อนไขของชนิดโมเดล (โมเดลราส์ชและโมเดลการวัดแบบจัดอันดับ) ความยาวของข้อสอบ (10, 20 และ 40 ข้อสำหรับโมเดลราส์ช และ 5, 10 และ 20 ข้อสำหรับโมเดลการวัดแบบจัดอันดับ) ขนาดกลุ่มตัวอย่าง (100, 200 และ 500 คน) ค่าความเข้มของอิทธิพลตั้งต้น (.2 และ .8) และการจับคู่แบบสอบกับผู้สอบเป็นตัวแปรต้น มีค่าความเข้มของอิทธิพลที่คำนวณได้จากสูตรที่พัฒนาขึ้นเป็นตัวแปรตาม ส่วนที่ 2 ใช้ข้อมูลเชิงประจักษ์เปรียบเทียบค่าความเข้มของอิทธิพลตามสูตรของ Cohen และสูตรที่พัฒนาขึ้น ผลการวิจัยในส่วนที่ 1 พบว่า ทั้งโมเดลราส์ชและโมเดลการวัดแบบจัดอันดับให้ผลการวิจัยในแนวทางเดียวกันคือ d_θ เป็นตัวประมาณค่าที่มีความลำเอียงเล็กน้อย (-.01 - .04) สำหรับการปรับแก้การถูกกดค่ามีความเหมาะสมหรือปรับแก้การกดค่าได้ 27 สถานการณ์จากทั้งหมด 39 สถานการณ์ ในส่วนที่ 2 พบว่าค่าขนาดอิทธิพลจากทฤษฎีทั้งสองมีค่าใกล้เคียงกันคือ .73 และ .76

งานวิจัยที่ศึกษาค่าสถิติที่จากทฤษฎีการทดสอบทั้งสองทฤษฎีมีวิธีการและเงื่อนไขที่ใช้ในการศึกษาวิจัยคล้ายคลึงกันบางประการดังที่ได้สรุปไว้ในตารางที่ 2.2

ตารางที่ 2.2 สรุปวิธีการและเงื่อนไขในงานวิจัยที่เปรียบเทียบ CTT และ IRT

ผู้ที่ศึกษา	ประเภทข้อมูล		โมเดลที่ศึกษา			เงื่อนไขที่ใช้ศึกษา					
	ประจักษ์	จำลอง	1PL	2PL	3PL	τ	a	b	c	N_{PERSON}	N_{ITEM}
จิติมา วรรณศรี (2539)		✓		✓						200, 600, 1,000	30, 60, 90
นภดล ยิ่งยงสกุล (2539)	✓			✓						231	28
เบญจพร ยนตจักรวิถี (2539)	✓		✓							500	33
ชวลิต นิวาสวัต (2541)	✓		✓							1,875	
อรวรรณ สุขโต (2542)	✓		✓							2,470	
Lawson (1991 อ้างถึงใน Fan, 1998)	✓		✓								
Ndalichako และ Rogers (1997)	✓		✓	✓	✓					1,230	70
Stage (1998; 2003)	✓			✓						1,500	20
Fan (1998)	✓	✓	✓	✓	✓	✓				1,000	60, 48
DeMars (2001)		✓	✓		✓	✓				2,000	15, 30, 60
MacDonald และ Paunonen (2002)		✓		✓		✓	✓	✓		1,000	20, 40, 60
Dawber, Rogers และ Carbonaro (2004)	✓	✓		✓	✓	✓	✓	✓	✓	>1,000	48, 70
Wang และ Chen (2004)	✓	✓		✓						100	15

หมายเหตุ : 1PL, 2PL, 3PL หมายถึง โมเดลโลจิสติกแบบหนึ่ง สองและสามพารามิเตอร์ตามลำดับ τ หมายถึง ความสามารถที่แท้จริงของผู้สอบ a, b, c หมายถึง อำนาจจำแนก ความยากและโอกาสในการเดา ตามลำดับ N_{PERSON} หมายถึงขนาดกลุ่มตัวอย่าง และ N_{ITEM} หมายถึงความยาวแบบสอบ

ตารางที่ 2.2 ชี้ให้เห็นว่าในการเปรียบเทียบค่าสถิติที่ได้ทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ อาทิ การเปรียบเทียบค่าความสามารถของผู้สอบที่ได้จากทั้งสองทฤษฎี อาจใช้ข้อมูลเชิงประจักษ์หรือข้อมูลที่จำลองขึ้น แต่การจำลองข้อมูลมีจุดเด่นที่เหนือกว่าการรวบรวมข้อมูลเชิงประจักษ์ ตามที่ Harwell และคณะ (1996) กล่าวไว้ว่า การเปรียบเทียบตัวประมาณค่าในสถานการณ์ต่าง ๆ ควรใช้การจำลองข้อมูลจึงจะเหมาะสม ผู้วิจัยจึงเลือกใช้การจำลองข้อมูลเป็นแนวทางในการวิจัยครั้งนี้

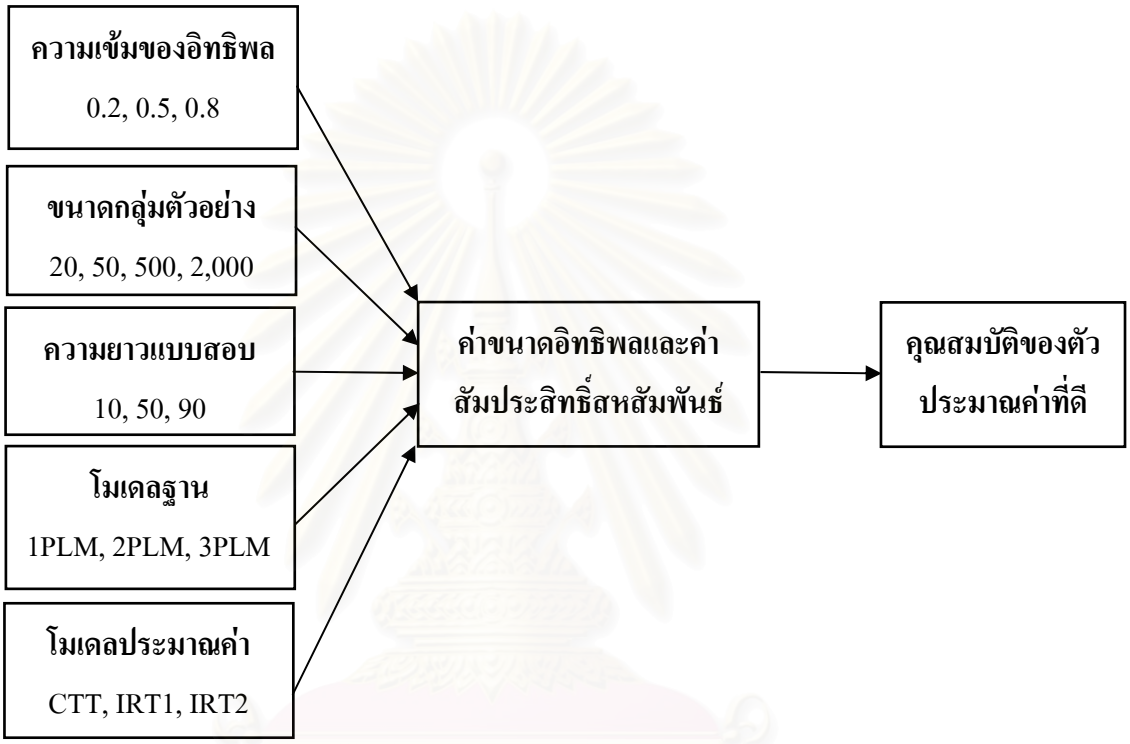
สำหรับเงื่อนไขในการเปรียบเทียบค่าสถิติที่ได้ทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบที่สำรวจจากงานวิจัยที่เกี่ยวข้อง พบว่า มีการใช้เงื่อนไขที่แตกต่างกันตามเป้าหมายของการเปรียบเทียบ โดยเงื่อนไขที่พบมาก ได้แก่ ความยาวข้อสอบ (Fan, 1998; DeMars, 2001; MacDonald & Paunonen, 2002; Dawber, Rogers & Carbonaro, 2004) และขนาดกลุ่มผู้สอบ (DeMars, 2001; Roberts & Henson, 2002; Stone & Yumoto, 2004; Wang & Chen, 2004) ทั้งสองเงื่อนไขนี้สอดคล้องกับผลการสำรวจของ Harwell และคณะ (1996) ซึ่งยังพบเงื่อนไขอื่นเพิ่มเติมอีก เช่น ชนิดของโมเดลการตอบสนองข้อสอบ (ทั้งที่เป็นโมเดลฐานในการคำนวณความน่าจะเป็น

สำหรับการจำลองข้อมูลคำตอบ และโมเดลที่ใช้ประมาณค่าความสามารถของผู้สอบ) การกระจายของค่าความยาก อำนาจจำแนก และโอกาสในการเดา เป็นต้น อย่างไรก็ดี นอกจากนี้ Hambleton และ Swaminathan (1985), Embretson และ Reise (2000) และ ศิริชัย กาญจนวาสี (2545) กล่าวไว้สอดคล้องกันว่า โมเดลการวัดที่ไม่สอดคล้องกับข้อมูลย่อมทำให้ผลการวิเคราะห์ขาดความถูกต้องน่าเชื่อถือ ความสอดคล้องของโมเดลกับข้อมูลจึงเป็นอีกประเด็นที่น่าสนใจ ผู้วิจัยประยุกต์แนวคิดของ DeMars (2001) ซึ่งกำหนดโมเดลที่ใช้ประมาณค่าความสามารถของผู้สอบมี 2 โมเดล โมเดลแรกเป็นโมเดลเดียวกับโมเดลฐานที่ใช้ในการจำลองคำตอบของผู้สอบ แทนสถานการณ์ที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และโมเดลที่สองเป็นโมเดลที่ต่างจากโมเดลฐาน แทนสถานการณ์ที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล

ในงานวิจัยครั้งนี้ผู้วิจัยเห็นว่า ความยาวแบบสอบ ขนาดกลุ่มตัวอย่าง ชนิดของโมเดลฐาน และชนิดของโมเดลประมาณค่า เป็นเงื่อนไขในการจำลองข้อมูล รวมทั้ง ความเข้มของอิทธิพลที่แท้จริง อีกเงื่อนไขหนึ่งที่สำคัญในการเปรียบเทียบค่าประมาณความเข้มของอิทธิพลที่ได้จากทั้งสองทฤษฎีการทดสอบ (Wang & Chen, 2004) โดยผู้วิจัยจะใช้ค่าความยาก อำนาจจำแนก และโอกาสในการเดาที่สุ่มได้จากการกระจายที่เหมาะสมกับค่าทั้งสามตามที่ Pelton (2002) ได้สรุปไว้แทนการกำหนดให้การกระจายของค่าทั้งสามเป็นเงื่อนไขในการจำลองข้อมูล

โดยทั่วไปเงื่อนไขในการจำลองสถานการณ์อาจแตกต่างกันบ้าง เช่น ความยาวแบบสอบมีตั้งแต่ 15 - 70 ข้อ โดยข้อสอบ 15 - 20 ข้อ อาจจัดเป็นแบบสอบสั้น ขณะที่ 30 - 50 ข้อเป็นแบบสอบความยาวปกติ และ 60 - 70 ข้อเป็นแบบสอบที่ยาวกว่าปกติ ผู้วิจัยเห็นว่าควรใช้เงื่อนไขความยาวแบบสอบ 3 ค่า คือ 10, 50 และ 90 ข้อ เพื่อให้ได้ผลการวิจัยที่ครอบคลุม สำหรับขนาดกลุ่มตัวอย่างควรมี 4 ค่า คือ 20, 50, 500 และ 2,000 คน เพื่อให้สอดคล้องกับความเป็นไปได้ในสถานการณ์จริงและเป็นขนาดที่พบมากในงานวิจัยเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ (Harwell et al., 1996) และใช้โมเดลโลจิสติกแบบหนึ่ง สอง และสามพารามิเตอร์เป็นโมเดลฐาน เนื่องจากเป็นโมเดลการตอบสนองข้อสอบที่นิยมใช้กันทั่วไป (ศิริชัย กาญจนวาสี, 2545; Hambleton, Swaminathan & Rogers, 1991; Embretson & Reise, 2000) โดยใช้โมเดลการทดสอบแบบดั้งเดิมเป็นหนึ่งในโมเดลประมาณค่า นอกเหนือจากโมเดลโลจิสติกที่สอดคล้องกับโมเดลฐานและโมเดลโลจิสติกที่ไม่สอดคล้องกับโมเดลฐาน สำหรับค่าของความเข้มของอิทธิพลที่แท้จริง ผู้วิจัยเลือกใช้ 5 ค่า คือ ขนาดอิทธิพลในประชากร (δ) มีค่า .2, .5, .8, 1.2 และ 2.6 ซึ่งมีค่าประมาณสัมประสิทธิ์สหสัมพันธ์ในประชากร (ρ) เท่ากับ .1, .2, .4, .5 และ .8 ตามลำดับ ตามที่ Cohen และ Pearson ได้จำแนกระดับของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์เอาไว้

การทบทวนเอกสารตามที่ได้เสนอทำให้ผู้วิจัยได้กรอบแนวคิดในการวิจัยดังภาพที่ 2.1



ภาพที่ 2.1 กรอบแนวคิดในการวิจัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ใช้ระเบียบวิธีวิจัยเชิงทดลอง โดยศึกษาจากการจำลองสถานการณ์การสอบที่เป็นไปได้ต่าง ๆ กัน ในบทนี้ผู้วิจัยแบ่งการนำเสนอวิธีดำเนินการวิจัยออกเป็น 3 ตอน ได้แก่ ตอนที่ 1 เงื่อนไขที่ใช้ในการจำลองข้อมูล ตอนที่ 2 การจำลองข้อมูล และตอนที่ 3 การวิเคราะห์ข้อมูล รายละเอียดมีดังนี้

ตอนที่ 1 เงื่อนไขที่ใช้ในการจำลองข้อมูล

ในการเปรียบเทียบค่าสถิติที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ อาทิ การเปรียบเทียบค่าความสามารถของผู้สอบ ที่ใช้ข้อมูลจากการจำลองตามสถานการณ์การสอบ พบว่า มีการใช้เงื่อนไขในการจำลองต่างกันตามเป้าหมายของการเปรียบเทียบ โดยเงื่อนไขที่พบมาก ได้แก่ ความยาวข้อสอบ (Fan, 1998; DeMars, 2001; MacDonald & Paunonen, 2002; Dawber, Rogers & Carbonaro, 2004) และขนาดกลุ่มผู้สอบ (DeMars, 2001; Roberts & Henson, 2002; Stone & Yumoto, 2004; Wang & Chen, 2004) ทั้งสองเงื่อนไขนี้สอดคล้องกับผลการสำรวจของ Harwell และคณะ (1996) ซึ่งยังพบเงื่อนไขอื่นเพิ่มเติมอีก เช่น ชนิดของโมเดลการตอบสนองข้อสอบ (ทั้งที่เป็นโมเดลฐาน และโมเดลประมาณค่า) การกระจายของค่าความยาก อำนาจจำแนก และโอกาสในการเดา เป็นต้น

นอกจากนี้ Hambleton และ Swaminathan (1985), Embretson และ Reise (2000) และ ศิริชัย กาญจนวาสี (2545) กล่าวไว้สอดคล้องกันว่า โมเดลการวัดที่ไม่สอดคล้องกับข้อมูลจะให้ผลการวิเคราะห์ที่ขาดความถูกต้องน่าเชื่อถือ ความสอดคล้องของโมเดลกับข้อมูลจึงเป็นอีกประเด็นที่น่าสนใจ

ในงานวิจัยครั้งนี้ผู้วิจัยเห็นว่า ความยาวแบบสอบ ขนาดกลุ่มตัวอย่าง โมเดลฐาน และโมเดลประมาณค่า เป็นเงื่อนไขในการจำลองข้อมูล รวมทั้ง ความเข้มของอิทธิพลที่แท้จริง ซึ่งเป็นอีกเงื่อนไขหนึ่งที่สำคัญในการเปรียบเทียบค่าประมาณความเข้มของอิทธิพลที่ได้จากทั้งสองทฤษฎีการทดสอบ (Roberts & Henson, 2002; Wang & Chen, 2004) โดยผู้วิจัยจะใช้ค่าความยาก อำนาจจำแนก และโอกาสในการเดาที่สุ่มได้จากการกระจายที่เหมาะสมกับค่าทั้งสามตามที่ Pelton (2002) ได้สรุปไว้แทนการกำหนดให้การกระจายของค่าทั้งสามเป็นเงื่อนไขในการจำลองข้อมูลสถานการณ์การสอบต่าง ๆ จึงเกิดจาก 8 เงื่อนไข ดังนี้

1. ค่าขนาดอิทธิพลที่แท้จริง (TRUES) 5 ค่า คือ ขนาดอิทธิพลในประชากร (δ) มีค่า .2, .5, .8, 1.2 และ 2.6 ซึ่งมีค่าประมาณสัมประสิทธิ์สหสัมพันธ์ในประชากร (ρ) เท่ากับ .1, .2, .4, .5 และ .8 ตามลำดับ

2. ขนาดกลุ่มตัวอย่าง (NSAMP) 4 ค่า คือ 20, 50, 500 และ 2,000 คน

3. ความยาวแบบสอบ (NITEM) 3 ค่า คือ 10, 50 และ 90 ข้อ

4. โมเดลฐานสำหรับจำลองข้อมูล (MBASE) 3 โมเดล คือ โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (1PLM) โมเดลโลจิสติกแบบสองพารามิเตอร์ (2PLM) และโมเดลโลจิสติกแบบสามพารามิเตอร์ (3PLM)

5. โมเดลประมาณค่าความสามารถของผู้สอบ (MUSED) เพื่อศึกษาผลเมื่อความสอดคล้องระหว่างโมเดลกับข้อมูลมีลักษณะต่าง ๆ กัน ในการวิจัยครั้งนี้ผู้วิจัยประยุกต์วิธีการของ DeMars (2001) โดยใช้โมเดล 3 ลักษณะ ได้แก่ **โมเดลที่ 1** เป็นโมเดลการตอบสนองข้อสอบที่เป็นโมเดลฐาน แทนสถานการณ์ที่ขนาดอิทธิพลหรือสัมประสิทธิ์สหสัมพันธ์ได้มาจากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล **โมเดลที่ 2** เป็นโมเดลการตอบสนองข้อสอบที่ไม่ใช่โมเดลฐาน (ได้โดยสุ่ม) แทนสถานการณ์ที่ขนาดอิทธิพลหรือสัมประสิทธิ์สหสัมพันธ์ได้มาจากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล และ **โมเดลที่ 3** เป็นโมเดลการทดสอบแบบดั้งเดิม เช่น กรณีที่ใช้ 1PLM เป็นฐานในการจำลองข้อมูล จะใช้ 1PLM 2PLM และโมเดลการทดสอบแบบดั้งเดิม (CTT) เป็นโมเดลประมาณค่า

6. อำนาจจำแนกของข้อสอบ (*a*) จะสุ่มค่าในช่วง 0 ถึง 1 จากการแจกแจงปกติมาตรฐาน

7. โอกาสในการเดา (*c*) จะสุ่มค่าในช่วง 0 ถึง 1 จากการแจกแจงยูนิฟอร์ม

8. ความยากของข้อสอบ (*b*) สำหรับ 1PLM และ 2PLM จะสุ่มค่าในช่วง -3 ถึง 3 จากการแจกแจงปกติมาตรฐาน สำหรับ 3PLM ค่าความยากจะขึ้นอยู่กับโอกาสในการเดา นั่นคือ

$$b = (1 + c) \div 2$$ (ศิริชัย กาญจนวาสี, 2545)

เงื่อนไขที่ 1 - 5 เป็นเงื่อนไขตามกรอบแนวคิดการวิจัย เงื่อนไขที่ 6 - 8 เป็นเงื่อนไขเกี่ยวกับข้อสอบที่ใช้ประกอบการจำลองข้อมูลคำตอบของผู้สอบ ดังนั้น สถานการณ์ที่ใช้ในการจำลองข้อมูลครั้งนี้จึงมีทั้งหมด 540 สถานการณ์ และเมื่อนำมาจัดกลุ่มเพื่อเปรียบเทียบคุณสมบัติของขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (d_{IRT1}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (d_{IRT2}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (r_{IRT1}) และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (r_{IRT2}) จะมี 180 สถานการณ์ รายละเอียดของสถานการณ์ทั้งหมดแสดงไว้ดังตารางที่ 3.1

ตารางที่ 3.1 สถานการณ์ที่ใช้ในการจำลองข้อมูล

ค่าขนาดอิทธิพลที่แท้จริง	ความยาวแบบสอบ	โมเดลฐาน	โมเดลประมาณค่า	ขนาดกลุ่มตัวอย่าง
δ = .2, .5, .8, 1.2, 2.6 ρ = .1, .2, .4, .5, .8	10, 50, 90	1PLM	CTT	20
				50
				500
				2000
			IRT1	20
				50
				500
				2000
			IRT2	20
				50
				500
				2000
		2PLM	CTT	20
				50
				500
				2000
			IRT1	20
				50
				500
				2000
			IRT2	20
				50
				500
				2000
		3PLM	CTT	20
				50
				500
				2000
			IRT1	20
				50
				500
				2000
			IRT2	20
				50
				500
				2000

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 2 การจำลองข้อมูล

ผู้วิจัยจำลองข้อมูลตามเงื่อนไขที่กำหนดไว้โดยทำซ้ำ 100 ครั้งด้วยขั้นตอนมาตรฐานที่ใช้ศึกษาทฤษฎีการตอบสนองข้อสอบด้วยการจำลองสถานการณ์ (Harwell et al, 1996; MacDonald & Paunonen, 2002; Pelton, 2002; Dawber, Rogers & Carbonaro, 2004) ดังนี้

1. สร้างแฟ้มข้อมูลค่าความสามารถที่แท้จริง (true ability) ของผู้สอบโดยเก็บไว้ในแฟ้มข้อมูลของโปรแกรม SPSS 6 ตัวแปร แบ่งเป็นตัวแปรที่เก็บค่าความสามารถที่แท้จริงของผู้สอบในกลุ่มควบคุม และกลุ่มทดลองในกรณีที่ขนาดอิทธิพลมีค่า .2, .5, .8, 1.2 และ 2.6 โดยผู้วิจัยอาศัยความสัมพันธ์ระหว่างขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ตามสมการของ Hunter และ Schmidt (1990) จึงใช้ค่าขนาดอิทธิพล .5, 1.2 และ 2.6 ซึ่งมีค่าประมาณสัมประสิทธิ์สหสัมพันธ์ที่ .2, .5 และ .8 ตามลำดับ แทนสถานการณ์ที่สัมประสิทธิ์สหสัมพันธ์ในประชากรมีค่า .2, .5 และ .8 ตามลำดับ ซึ่งการกำหนดค่าของตัวแปรให้อยู่ในรูปของขนาดอิทธิพลช่วยให้จำลองข้อมูลได้สะดวกกว่า

ค่าของตัวแปรข้างต้นถูกสุ่มขึ้นตามคำสั่ง (syntax) ในโปรแกรม SPSS ที่ผู้วิจัยเขียนขึ้น (ภาคผนวก ก) โดยส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถที่แท้จริงของประชากรทุกกลุ่มมีค่าเท่ากับ 0 แต่ค่าเฉลี่ยเฉพาะกลุ่มควบคุมจะมีค่าเป็น 1 สำหรับค่าเฉลี่ยของประชากรของกลุ่มทดลองจะมีค่าเฉลี่ยเท่ากับค่าขนาดอิทธิพลของกลุ่มนั้น ๆ การกำหนดค่าดังกล่าวมีที่มาจากสมการต่อไปนี้

$$\delta = \frac{\mu_E - \mu_C}{\sigma_{pooled}}$$

โดยที่ δ แทนค่าขนาดอิทธิพล

μ_E แทนค่าเฉลี่ยของประชากรกลุ่มทดลอง

μ_C แทนค่าเฉลี่ยของประชากรกลุ่มควบคุม

และ σ_{pooled} แทนส่วนเบี่ยงเบนมาตรฐานรวมของกลุ่มทดลองและกลุ่มควบคุม

เนื่องจากผู้วิจัยกำหนดให้ส่วนเบี่ยงเบนมาตรฐานรวมของกลุ่มทดลอง และกลุ่มควบคุมมีค่าเท่ากับ 1 ทำให้ σ_{pooled} มีค่าเป็น 1 ด้วย สมการข้างต้นจึงลดรูปเป็น $\delta = \mu_E - \mu_C$ แต่เนื่องจาก μ_C มีค่าเท่ากับ 0 จะได้ว่า $\delta = \mu_E$ กล่าวคือ ค่าเฉลี่ยของประชากรกลุ่มทดลองมีค่าเท่ากับค่าขนาดอิทธิพลนั่นเอง

จากแนวคิดของ Hunter และ Schmidt (1990) ซึ่งกล่าวว่ากลุ่มทดลองและกลุ่มควบคุมในงานวิจัยเชิงทดลอง (ที่ใช้คำนวณ d) ควรมีขนาดเป็นครึ่งหนึ่งของกลุ่มตัวอย่างในการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ ($N_C = N_E = N$) ดังนั้นผู้วิจัยจึงจำลองค่าความสามารถที่แท้จริงใน 6 ตัวแปรให้มีจำนวนเท่ากันทั้งหมดตามเงื่อนไขของขนาดกลุ่มตัวอย่างหรือกลุ่มผู้สอบ (NSAMP) แต่ละค่าแล้วจัดเก็บตัวแปรทั้ง 6 ที่มีจำนวนผู้สอบเท่ากันไว้ในแฟ้มข้อมูลเดียวกัน รวม 4 แฟ้ม

2. สร้างแฟ้มข้อมูลเก็บค่าพารามิเตอร์ของข้อสอบ อันได้แก่ ค่าอำนาจจำแนก ค่าโอกาสในการเดาถูก ค่าความยากสำหรับโมเดลโลจิสติกแบบหนึ่งและสองพารามิเตอร์ และค่าความยากสำหรับโมเดลโลจิสติกแบบสามพารามิเตอร์ รวม 4 ตัวแปร ทั้งนี้เนื่องจากมีเงื่อนไขของค่าความยากสำหรับแต่ละโมเดลต่างกัน กล่าวคือค่าความยากสำหรับโมเดลโลจิสติกแบบสามพารามิเตอร์จะมีค่าขึ้นอยู่กับโอกาสในการเดา การแจกแจงจึงอาจไม่ใช่แบบปกติมาตรฐานนั่นเอง (Pelton, 2002) ผู้วิจัยเขียนคำสั่งในโปรแกรม SPSS (ภาคผนวก ข) เพื่อสร้างค่าของทั้ง 4 ตัวแปร ให้มีจำนวน 10, 50 และ 90 ค่า ตามเงื่อนไขของความยาวแบบสอบ จากนั้นผู้วิจัยจึงจัดเก็บตัวแปรทั้ง 4 ที่มีความยาวแบบสอบเท่ากันไว้ในแฟ้มข้อมูลเดียวกัน รวม 3 แฟ้มข้อมูล

3. นำข้อมูลที่สร้างไว้ในข้อ 1 และ 2 มาคำนวณค่าความน่าจะเป็นที่ผู้สอบคนที่ i จะตอบข้อสอบข้อที่ j ถูกหรือค่า $P_j(\theta_i)$ ตามโมเดลฐานที่ใช้ในการจำลองข้อมูล ได้แก่ โมเดลโลจิสติกแบบหนึ่ง สองและสามพารามิเตอร์ โดยค่าที่ได้จะจัดเก็บไว้ในแฟ้มข้อมูลเดียวกับแฟ้มข้อมูลในข้อ 2 ทำให้ได้แฟ้มข้อมูล 4 แฟ้มที่ประกอบด้วยค่าความสามารถที่แท้จริงและค่าความน่าจะเป็นในการตอบถูกของผู้สอบ เท่ากับจำนวนผู้สอบและข้อสอบ ขั้นตอนนี้ผู้วิจัยทำงานโดยใช้คำสั่งในภาคผนวก จ

4. คำสั่งในภาคผนวก จ จะนำข้อมูลจากขั้นตอนที่ 3 มาสร้างคำตอบของผู้สอบในแต่ละสถานการณ์ โดยเริ่มจากการสุ่มค่าความน่าจะเป็น ($PROB_{ij}$) ที่มีค่าระหว่าง 0 ถึง 1 สำหรับผู้สอบคนที่ i ในการทำข้อสอบข้อที่ j เพื่อใช้เป็นค่าเปรียบเทียบกับค่าความน่าจะเป็นที่ผู้สอบคนที่ i จะตอบข้อสอบข้อที่ j ถูกหรือ $P_j(\theta_i)$ และกำหนดเป็นค่าคำตอบของผู้สอบ (U_{ij}) ซึ่งเป็นไปได้ 2 กรณีคือ กรณีที่ 1 $PROB_{ij}$ มีค่ามากกว่า $P_j(\theta_i)$ จะกำหนดให้ U_{ij} มีค่าเป็น 0 หรือ กรณีที่ 2 $PROB_{ij}$ มีค่าน้อยกว่าหรือเท่ากับ $P_j(\theta_i)$ จะกำหนดให้ U_{ij} มีค่าเป็น 1 โดยข้อมูลคำตอบในส่วนนี้จะจัดเก็บแยกเป็นแฟ้มข้อมูลแบบอักขระ (ASCII file) เพื่อนำมาประมาณค่าความสามารถตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ ด้วยโปรแกรม BILOG 3.0 โดยจะทำการประมาณค่า 2 ครั้งต่อ 1 แฟ้มข้อมูล ตามสถานการณ์ที่ได้ระบุไว้ในตารางที่ 3.1 ทำให้ได้แฟ้มผลการวิเคราะห์ด้วยโปรแกรม BILOG ซึ่งจะนำไปวิเคราะห์ข้อมูลเพื่อตอบถามวิจัยต่อไป

อนึ่ง ก่อนการวิเคราะห์ด้วยโปรแกรม BILOG ผู้วิจัยจะตรวจสอบความเป็นเอกมิติซึ่งเป็นข้อตกลงเบื้องต้นที่สำคัญของทฤษฎีการตอบสนองข้อสอบที่มีการตรวจให้คะแนนแบบทวิภาค โดยการคำนวณอัตราส่วนของค่าไอเกน (Eigen Ratio: ER) ขององค์ประกอบที่ 1 กับองค์ประกอบที่ 2 ที่ได้จากการวิเคราะห์องค์ประกอบมุขสำคัญ (principal factor analysis) หากอัตราส่วนที่ได้มีค่าต่ำกว่า 3 ผู้วิจัยจะทำการจำลองสถานการณ์นั้นใหม่ การตรวจสอบความเป็นเอกมิติด้วยวิธีนี้สะดวกเพราะมีโปรแกรมสำเร็จรูปในการวิเคราะห์ ทั้งยังมีประสิทธิภาพและความไวสูง (วรนุช แหยมแสง, 2537; นิคม กีรติวรางกูร, 2542; Hambleton & Swaminathan, 1985; Embretson & Reise, 2000)

ตอนที่ 3 การวิเคราะห์ข้อมูล

ผลการวิเคราะห์ด้วยโปรแกรม BILOG ที่จำเป็นในการวิจัยครั้งนี้ประกอบด้วยคะแนนดิบและค่าความสามารถของผู้สอบ ซึ่งเป็นค่าประมาณความสามารถที่แท้จริงตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบตามลำดับ การวิเคราะห์ข้อมูลส่วนมุ่งตอบคำถามเกี่ยวกับตัวประมาณค่าความเข้มของอิทธิพล 2 ตัว คือ ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์เมื่อใช้ข้อมูลจาก 2 ทฤษฎีการทดสอบ รายละเอียดการวิเคราะห์มีดังนี้

1. นำคะแนนดิบและค่าความสามารถของผู้สอบที่ได้จากโปรแกรม BILOG มาคำนวณค่าขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ โดยขนาดอิทธิพล (d) และสัมประสิทธิ์สหสัมพันธ์ (แบบไบซีเรียล) คำนวณได้ด้วยสูตรต่อไปนี้

$$d = \frac{\overline{X}_E - \overline{X}_C}{S_{pooled}} \text{ (Kirk, 1996) และ } r = \frac{\overline{X}_E - \overline{X}_C}{S_{total}} \times \frac{pq}{Y} \text{ (Guilford \& Fruchter, 1981)}$$

โดยที่ $\overline{X}_E$ แทนค่าเฉลี่ยของตัวแปร X ในกลุ่มทดลอง

$\overline{X}_C$ แทนค่าเฉลี่ยของตัวแปร X ในกลุ่มควบคุม

S_{pooled} แทนส่วนเบี่ยงเบนมาตรฐานรวมของกลุ่มทดลองและกลุ่มควบคุม

S_{total} แทนส่วนเบี่ยงเบนมาตรฐานรวมของกลุ่มตัวอย่างทั้งหมด

p แทนอัตราส่วนของขนาดกลุ่มทดลองต่อขนาดกลุ่มตัวอย่างทั้งหมด

q แทนอัตราส่วนของขนาดกลุ่มควบคุมต่อขนาดกลุ่มตัวอย่างทั้งหมด

Y แทนค่าระยะตั้งฉากที่แบ่งพื้นที่ใต้โค้งปกติออกเป็น p และ q ในที่นี้ Y = .3989

อนึ่ง ผู้วิจัยทำการตรวจสอบความถูกต้องของโปรแกรมที่ผู้วิจัยสร้างขึ้นสำหรับการคำนวณค่าขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ (ภาคผนวก ฉ) ด้วยการวิเคราะห์ความสัมพันธ์ของค่าที่ได้จากโปรแกรม กับค่าที่ได้จากการคำนวณด้วยโปรแกรม LISREL ร่วมกับ EXCEL โดยใช้ข้อมูลจากการจำลองสถานการณ์ในรอบที่ 1 รวม 540 กรณี หากสัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าเท่ากับ 1 ถือว่าโปรแกรมที่ผู้วิจัยพัฒนาขึ้นมีความถูกต้อง สามารถนำค่าขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ที่ได้มาวิเคราะห์ข้อมูลในขั้นต่อไปได้ แต่ถ้าไม่เป็นไปตามนี้ผู้วิจัยจะตรวจหาที่ผิดพลาดของโปรแกรมแล้วดำเนินการแก้ไขจนได้ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าที่ได้จากโปรแกรม กับค่าที่ได้จากการคำนวณด้วยโปรแกรม LISREL ร่วมกับ EXCEL มีค่าเท่ากับ 1

2. คำนวณสถิติเชิงบรรยายของขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ

3. เปรียบเทียบค่าเฉลี่ยระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูลและระหว่างสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล โดยการวิเคราะห์ความแปรปรวนทางเดียว

4. คำนวณค่า RMSD, SRMSD และ BIAS เพื่อเปรียบเทียบความลำเอียงตามสูตรต่อไปนี้

$$BIAS = E(\hat{a}) - a$$ (Hay, 1963 อ้างถึงในสุกัญญรัตน์ คงงาม, 2539)

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(\hat{a}_i - a)^2}{N}}$$ (Harwell et al., 1996; Pelton, 2004)

และ

$$SRMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(\frac{\hat{a}_i - a}{SD_{\hat{a}_i - a}}\right)^2}{N}}$$ (ดัดแปลงจาก Pelton, 2004)

โดยที่ a หมายถึง พารามิเตอร์ที่ต้องการประมาณค่า

$\hat{a}$ หมายถึง ค่าประมาณของพารามิเตอร์

$E(\hat{a})$ หมายถึง ค่าคาดหวังของ $\hat{a}$ หรือค่าเฉลี่ยของการแจกแจงการสุ่มของ $\hat{a}$

$SD_{\hat{a}_i - a}$ หมายถึง ส่วนเบี่ยงเบนมาตรฐานของค่าความลำเอียง (BIAS)

i หมายถึง ครั้งที่ประมาณค่า

N หมายถึง จำนวนครั้งในการประมาณค่า

5. คำนวณค่าผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) เพื่อใช้ในการพิจารณาความคงเส้นคงวา ตามสูตรที่ปรับจากสุกัญญรัตน์ คงงาม (2539) ต่อไปนี้

$$\Delta = SRMSD_{Nmin} - SRMSD_{Nmax}$$

โดยที่ Δ หมายถึง ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน

$SRMSD_{Nmin}$ หมายถึง ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานในกรณีที่กลุ่มตัวอย่างมีขนาดเล็กที่สุด (ในการวิจัยครั้งนี้ Nmin มีค่า 20)

$SRMSD_{Nmax}$ หมายถึง ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่ที่สุด (ในการวิจัยครั้งนี้ Nmax มีค่า 2000)

6. เมื่อได้ค่าสถิติสำหรับพิจารณาคุณสมบัติตัวประมาณค่าที่ดีครบแล้ว ผู้วิจัยตัดสินตัวประมาณค่าที่ดีที่สุดในการประมาณค่าความเข้มของอิทธิพลแต่ละสถานการณ์โดยใช้เกณฑ์สัมพัทธ์

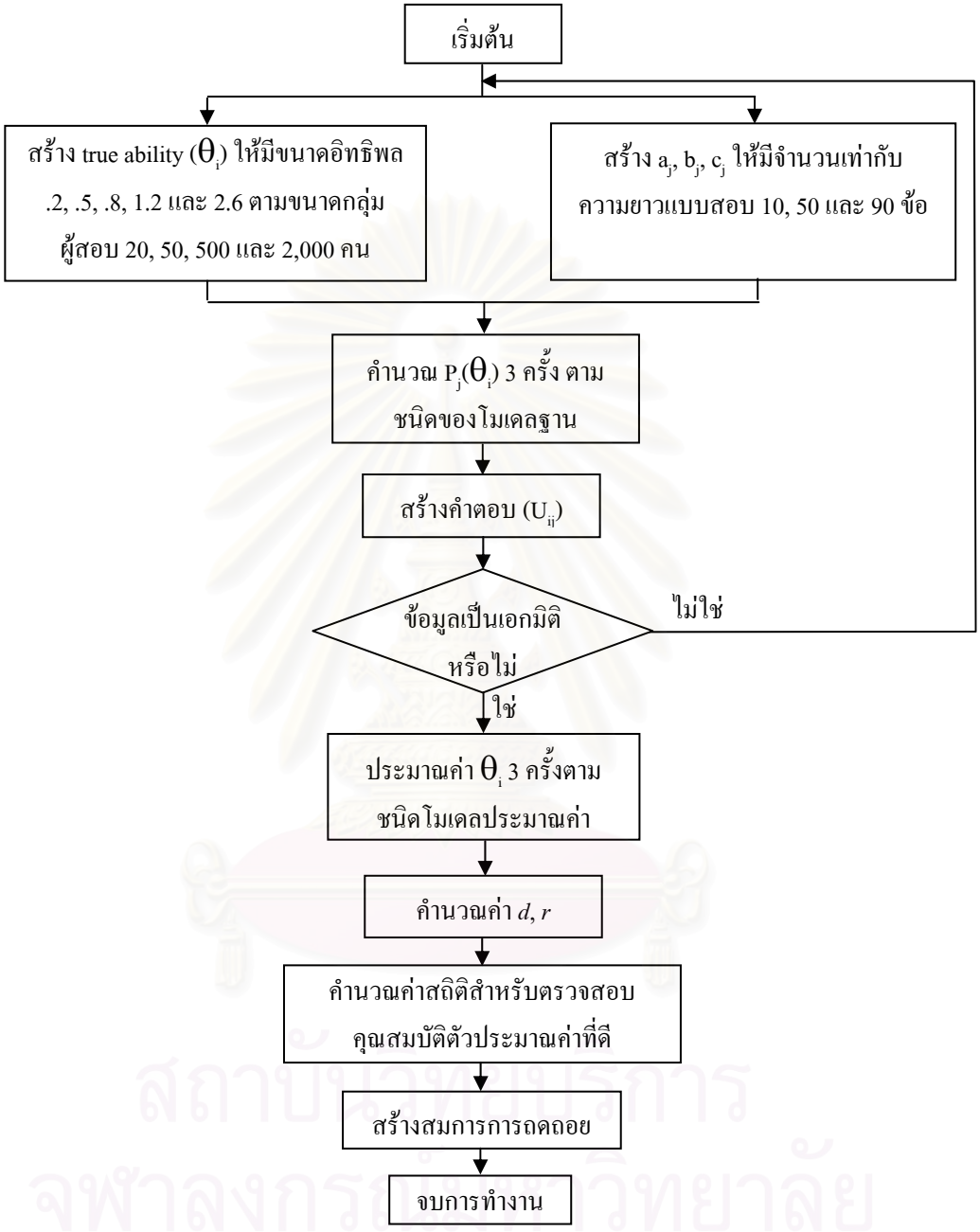
กล่าวคือ ตัวประมาณค่าจะต้องมีความลำเอียงในการประมาณค่าต่ำสุด (SRMSD ต่ำสุด) มีความคงเส้นคงวาสูงสุด (Δ สูงสุด) และมีประสิทธิภาพสัมพัทธ์สูงสุด (ความแปรปรวนต่ำสุด) เมื่อเทียบกับตัวประมาณค่าชนิดอื่น สำหรับค่า BIAS จะใช้ในการพิจารณาว่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงหรือต่ำกว่าค่าความเข้มที่แท้จริงเป็นหลัก กล่าวคือ ค่า BIAS ที่เป็นลบจะหมายถึงตัวประมาณค่าให้ค่าประมาณที่ต่ำกว่าค่าพารามิเตอร์ (underestimate) ในทางกลับกันค่า BIAS ที่เป็นลบจะหมายถึงตัวประมาณค่าให้ค่าประมาณที่สูงกว่าค่าพารามิเตอร์ (overestimate)

7. วิเคราะห์ความสัมพันธ์ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมและขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และสร้างสมการถดถอยของขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล บนขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม

8. วิเคราะห์ความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล และสร้างสมการถดถอยของสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล บนสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม

การดำเนินงานทั้งหมดสามารถเขียนเป็นแผนภูมิสายงานได้ดังภาพที่ 3.1

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



ภาพที่ 3.1 แผนภูมิสายงานแสดงขั้นตอนการจำลองข้อมูลและการวิเคราะห์ข้อมูล

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การนำเสนอผลการวิเคราะห์ข้อมูลแบ่งเป็น 6 ตอน ได้แก่ ตอนที่ 1 ผลการตรวจสอบความเป็นเอกมิติของข้อมูล ตอนที่ 2 ผลการตรวจสอบความถูกต้องของโปรแกรมคำนวณค่าประมาณความเข้มของอิทธิพลที่ผู้วิจัยสร้างขึ้น ตอนที่ 3 ผลการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีในภาพรวม ตอนที่ 4 ผลการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีรายสถานการณ์ ตอนที่ 5 ผลการวิเคราะห์ความสัมพันธ์และการถดถอย และตอนที่ 6 สรุปผลการวิเคราะห์ตามวัตถุประสงค์การวิจัย รายละเอียดมีดังนี้

ตอนที่ 1 ผลการตรวจสอบความเป็นเอกมิติของข้อมูล

การตรวจสอบความเป็นเอกมิติของข้อมูลที่จำลองขึ้นในแต่ละสถานการณ์ด้วยการพิจารณาอัตราส่วนของค่าไอเกนขององค์ประกอบที่ 1 (EG1) และค่าไอเกนขององค์ประกอบที่ 2 (EG2) ที่ได้จากการวิเคราะห์องค์ประกอบ พบว่า อัตราส่วนของค่าไอเกนขององค์ประกอบที่ 1 และองค์ประกอบที่ 2 (ER) มีค่าต่ำสุดและสูงสุดเท่ากับ 3.014 และ 10.565 ตามลำดับ นั่นคืออัตราส่วนมีค่าเกิน 3 ทุกสถานการณ์ ทั้ง 100 รอบของการจำลองข้อมูลซ้ำ แสดงว่าข้อมูลที่จำลองขึ้นทั้งหมดมีความเป็นเอกมิติ เป็นไปตามข้อตกลงเบื้องต้นที่สำคัญของการประมาณค่าความสามารถของผู้สอบด้วยโมเดลการตอบสนองข้อสอบที่มีการตรวจให้คะแนนแบบสองค่า รายละเอียดดังตารางที่ 4.1

ตารางที่ 4.1 ผลการตรวจสอบความเป็นเอกมิติของข้อมูลด้วยอัตราส่วนค่าไอเกน

รอบที่	แบบสอบยาว 10 ข้อ			แบบสอบยาว 50 ข้อ			แบบสอบยาว 90 ข้อ		
	EG1	EG2	ER	EG1	EG2	ER	EG1	EG2	ER
1	3.365	1.105	3.047	14.763	1.993	7.408	25.808	3.169	8.145
2	3.442	.999	3.447	13.585	2.212	6.142	24.740	3.105	7.968
3	3.348	1.007	3.324	14.205	2.287	6.211	24.470	3.243	7.544
4	3.689	.957	3.856	14.726	2.042	7.213	26.009	2.905	8.954
5	3.696	.804	4.595	14.303	2.052	6.969	25.457	3.182	8.000
6	3.935	.835	4.715	14.260	2.149	6.636	25.319	3.102	8.161
7	3.427	.979	3.500	13.918	1.963	7.090	25.616	2.468	10.379
8	3.890	.858	4.533	14.807	1.955	7.573	25.244	3.275	7.707
9	3.371	1.029	3.274	12.936	1.991	6.499	23.512	3.168	7.421
10	3.288	1.082	3.040	14.675	2.079	7.058	26.052	2.882	9.041
11	3.410	1.037	3.289	14.581	2.285	6.383	26.142	3.217	8.126

ตารางที่ 4.1 (ต่อ) ผลการตรวจสอบความเป็นเอกมิติของข้อมูลด้วยอัตราส่วนค่าไอเกน

รอบที่	แบบสอบยาว 10 ข้อ			แบบสอบยาว 50 ข้อ			แบบสอบยาว 90 ข้อ		
	EG1	EG2	ER	EG1	EG2	ER	EG1	EG2	ER
12	3.290	1.022	3.220	15.232	1.846	8.252	24.818	3.344	7.421
13	3.481	.981	3.550	14.737	1.735	8.496	24.166	3.252	7.431
14	3.923	.774	5.070	15.063	2.037	7.393	26.224	2.866	9.149
15	3.282	1.047	3.135	14.234	2.006	7.097	26.306	2.908	9.047
16	3.430	1.009	3.400	14.090	2.062	6.834	25.516	2.804	9.099
17	3.578	.895	3.999	13.986	2.202	6.351	25.639	3.367	7.616
18	3.401	.968	3.512	14.066	2.212	6.358	25.610	3.059	8.373
19	3.432	1.015	3.380	14.993	1.797	8.344	26.435	3.002	8.805
20	3.298	.964	3.422	13.963	1.947	7.172	23.842	3.138	7.599
21	3.485	.982	3.547	15.838	1.499	10.565	24.715	3.316	7.454
22	3.491	.925	3.774	13.834	1.969	7.027	24.745	2.971	8.328
23	3.463	.926	3.741	14.713	1.777	8.279	23.286	3.477	6.697
24	3.751	.861	4.355	14.033	2.098	6.688	25.188	3.153	7.989
25	3.649	.952	3.834	15.175	2.026	7.491	26.263	2.825	9.297
26	3.311	1.058	3.129	14.630	2.054	7.121	26.889	2.911	9.237
27	3.374	1.042	3.238	13.699	2.015	6.799	23.818	3.063	7.775
28	3.325	1.073	3.098	14.558	1.844	7.896	24.052	3.310	7.266
29	3.207	1.047	3.063	14.534	2.196	6.618	24.657	3.485	7.075
30	3.324	1.048	3.172	14.252	2.030	7.022	26.724	2.805	9.527
31	3.371	1.054	3.197	15.738	1.756	8.964	26.477	2.887	9.169
32	3.784	.873	4.334	13.696	2.340	5.852	26.764	2.679	9.991
33	3.575	.951	3.759	14.671	2.017	7.272	26.269	2.871	9.150
34	3.360	1.068	3.146	15.030	1.972	7.623	25.361	3.201	7.923
35	3.705	.937	3.956	14.875	2.112	7.044	26.488	3.106	8.528
36	3.505	.960	3.650	14.285	1.901	7.513	24.777	3.121	7.938
37	3.642	.918	3.965	14.975	1.899	7.884	25.986	3.022	8.599
38	3.306	1.017	3.250	13.945	2.078	6.710	24.508	3.031	8.087
39	3.531	1.006	3.512	14.201	2.088	6.800	25.063	3.150	7.955
40	3.577	1.002	3.569	15.173	1.839	8.249	26.398	2.947	8.958
41	3.468	.990	3.503	13.896	2.127	6.534	24.973	3.502	7.132
42	3.607	1.174	3.072	14.623	1.865	7.842	23.373	3.819	6.120
43	3.524	.926	3.805	14.128	2.049	6.893	24.539	3.332	7.364
44	3.529	.988	3.573	13.324	2.258	5.902	24.995	3.064	8.158
45	3.511	1.014	3.462	13.597	2.352	5.782	25.448	3.216	7.913
46	3.506	1.030	3.404	14.137	2.196	6.439	25.438	3.382	7.521
47	3.297	1.068	3.087	14.654	1.923	7.619	25.766	2.892	8.910
48	3.185	1.057	3.014	13.880	2.072	6.699	24.058	3.471	6.931
49	3.565	.973	3.663	14.827	2.251	6.585	26.197	3.609	7.259

ตารางที่ 4.1 (ต่อ) ผลการตรวจสอบความเป็นเอกมิติของข้อมูลด้วยอัตราส่วนค่าไอเกน

รอบที่	แบบสอบยาว 10 ข้อ			แบบสอบยาว 50 ข้อ			แบบสอบยาว 90 ข้อ		
	EG1	EG2	ER	EG1	EG2	ER	EG1	EG2	ER
50	3.865	.843	4.586	14.299	2.021	7.077	26.242	3.115	8.425
51	3.582	.897	3.994	14.619	1.889	7.739	25.082	3.104	8.080
52	3.623	.894	4.054	14.163	2.323	6.097	25.115	3.463	7.253
53	3.322	1.002	3.314	14.337	2.389	6.000	25.966	2.963	8.764
54	3.262	1.068	3.053	13.430	2.357	5.698	24.820	3.279	7.569
55	3.629	.877	4.140	15.150	1.852	8.179	26.477	2.809	9.425
56	3.466	1.012	3.424	13.675	2.142	6.383	23.992	3.183	7.536
57	3.376	.978	3.452	14.170	2.377	5.961	25.867	3.114	8.306
58	3.479	1.066	3.265	14.975	1.944	7.705	26.019	3.031	8.584
59	3.666	.926	3.959	15.201	2.084	7.293	25.617	3.256	7.867
60	3.467	.994	3.488	14.349	2.248	6.382	25.387	3.097	8.197
61	3.297	1.038	3.176	13.674	2.402	5.693	25.420	3.242	7.842
62	3.569	1.071	3.332	14.188	2.037	6.966	25.012	2.961	8.447
63	3.490	.960	3.634	14.807	1.845	8.025	26.062	3.151	8.270
64	3.764	.913	4.121	13.232	2.249	5.882	24.322	3.204	7.592
65	3.214	1.062	3.027	14.081	2.189	6.434	25.869	2.964	8.729
66	3.631	.953	3.811	14.991	1.975	7.591	25.541	2.991	8.538
67	3.704	.891	4.156	15.058	1.851	8.134	25.513	2.922	8.732
68	3.326	1.052	3.161	14.705	2.189	6.718	25.170	3.101	8.116
69	3.303	.987	3.346	14.356	1.794	8.002	23.969	2.876	8.335
70	3.296	1.016	3.243	14.494	2.099	6.905	24.684	3.294	7.493
71	3.376	.924	3.652	13.402	2.024	6.621	24.556	2.672	9.190
72	3.478	.959	3.626	14.571	2.032	7.169	25.336	3.111	8.144
73	3.445	1.003	3.434	15.775	1.822	8.656	25.532	3.090	8.264
74	3.538	.886	3.991	14.289	2.147	6.654	25.706	3.019	8.516
75	3.547	.963	3.683	14.148	2.037	6.947	23.997	3.542	6.775
76	3.761	.893	4.209	14.670	2.054	7.141	25.962	3.101	8.371
77	3.567	.881	4.047	15.154	1.870	8.102	24.679	3.338	7.394
78	3.389	.984	3.444	14.165	2.036	6.959	23.651	3.467	6.822
79	3.372	1.023	3.296	15.242	1.941	7.851	25.215	3.117	8.089
80	3.352	1.037	3.231	14.477	2.122	6.823	25.495	3.239	7.872
81	3.690	.893	4.131	14.272	2.317	6.160	26.542	2.999	8.852
82	3.331	.958	3.477	12.915	2.090	6.180	24.917	2.633	9.465
83	3.467	.971	3.572	14.965	1.869	8.007	25.955	2.916	8.901
84	3.259	1.034	3.152	14.291	1.770	8.076	24.398	3.008	8.112
85	3.115	1.002	3.108	14.038	2.186	6.423	25.034	3.191	7.845
86	3.614	.906	3.987	14.584	2.122	6.873	24.714	3.125	7.907
87	3.212	.988	3.252	14.262	2.390	5.968	27.617	2.768	9.976

ตารางที่ 4.1 (ต่อ) ผลการตรวจสอบความเป็นเอกมิติของข้อมูลด้วยอัตราส่วนค่าไอเกน

รอบที่	แบบสอบยาว 10 ข้อ			แบบสอบยาว 50 ข้อ			แบบสอบยาว 90 ข้อ		
	EG1	EG2	ER	EG1	EG2	ER	EG1	EG2	ER
88	3.759	.924	4.070	14.904	1.924	7.747	25.080	3.377	7.427
89	3.400	1.099	3.093	13.543	2.090	6.479	24.987	2.731	9.150
90	3.354	1.021	3.286	14.771	1.897	7.785	24.822	3.286	7.553
91	3.632	.986	3.684	14.524	2.151	6.752	25.799	3.155	8.177
92	3.679	.976	3.771	13.826	2.204	6.273	27.084	2.729	9.925
93	3.288	1.054	3.121	14.727	2.245	6.559	26.648	3.064	8.697
94	3.481	.959	3.628	14.592	2.113	6.905	27.101	2.970	9.126
95	3.537	1.001	3.534	14.332	2.001	7.163	25.145	3.146	7.993
96	3.699	.843	4.386	14.922	2.034	7.337	25.088	3.371	7.443
97	3.474	1.021	3.403	14.249	2.310	6.170	25.864	3.386	7.638
98	3.247	1.054	3.081	13.956	2.072	6.734	24.332	3.431	7.091
99	3.837	.826	4.646	15.048	1.975	7.618	26.957	2.813	9.584
100	3.557	.984	3.616	14.000	1.900	7.368	25.762	2.556	10.079

ตอนที่ 2 ผลการตรวจสอบความถูกต้องของโปรแกรมคำนวณค่าประมาณความเข้มของอิทธิพลที่ผู้วิจัยสร้างขึ้น

ผู้วิจัยตรวจสอบความถูกต้องของโปรแกรมที่ผู้วิจัยสร้างขึ้นสำหรับการคำนวณค่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ (ภาคผนวก ฉ) ด้วยการวิเคราะห์ความสัมพันธ์ของค่าที่ได้จากโปรแกรมฯ กับค่าที่ได้จากการคำนวณด้วยโปรแกรม LISREL ร่วมกับ EXCEL โดยใช้ข้อมูลจากการจำลองสถานการณ์รอบที่ 1 พบว่า ค่าประมาณความเข้มของอิทธิพล (ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์) ที่ได้จากโปรแกรมของผู้วิจัยมีความสัมพันธ์กับค่าที่ได้จากโปรแกรม LISREL ร่วมกับ EXCEL โดยความสัมพันธ์ที่พบเป็นความสัมพันธ์ทางบวก และมีนัยสำคัญทางสถิติที่ระดับ .01 รายละเอียดดังตารางที่ 4.2

ตารางที่ 4.2 ค่าเฉลี่ยและสัมประสิทธิ์สหสัมพันธ์ระหว่างของค่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ที่ได้จากโปรแกรมของผู้วิจัยกับการคำนวณด้วยโปรแกรม LISREL ร่วมกับ EXCEL

ตัวประมาณค่า	โปรแกรมของผู้วิจัย		โปรแกรม LISREL[(1)]		สัมประสิทธิ์สหสัมพันธ์	ค่า p
	$\bar{X}$	S.D.	$\bar{X}$	S.D.		
d	.127	.660	.127	.660	1.000	.000
r	.054	.308	.054	.312	1.000	.000

หมายเหตุ [(1)] เป็นการคำนวณด้วยโปรแกรม LISREL ร่วมกับ EXCEL โดยใช้ LISREL คำนวณค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล สำหรับ EXCEL ใช้คำนวณค่าขนาดอิทธิพล

ตอนที่ 3 ผลการพิจารณาคุณสมบัติของตัวประมาณค่าในภาพรวม

ผลการวิเคราะห์ข้อมูลในส่วนนี้ประกอบด้วย ค่าต่ำสุด (MIN) ค่าสูงสุด (MAX) ค่าเฉลี่ย (MEAN) ความเบ้ (SK) ความโด่ง (KU) ความแปรปรวน (VAR) ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ย (RMSD) ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) และผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) ของขนาดอิทธิพล (d) และสัมประสิทธิ์สหสัมพันธ์ (r) มีวัตถุประสงค์หลักเพื่อการพิจารณาคุณสมบัติของตัวประมาณค่าในภาพรวมของตัวแปรเงื่อนไขการจำลองข้อมูล รายละเอียดมีดังนี้

เมื่อพิจารณาเงื่อนไขความเข้มของอิทธิพลที่แท้จริง (TRUES) พบว่า ค่าต่ำสุดของขนาดอิทธิพลมีค่าต่ำกว่าค่าต่ำสุดของสัมประสิทธิ์สหสัมพันธ์ในทุกระดับของความเข้มของอิทธิพลที่แท้จริง ขณะที่ค่าสูงสุดของขนาดอิทธิพลมีค่าสูงกว่าค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ในทุกระดับของความเข้มของอิทธิพลที่แท้จริง พิสัยของขนาดอิทธิพลจึงกว้างกว่าพิสัยของสัมประสิทธิ์สหสัมพันธ์ ค่าต่ำสุดของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีแนวโน้มที่จะต่ำมากขึ้นเมื่อความเข้มของอิทธิพลที่แท้จริงมีค่าเพิ่มขึ้น แต่สำหรับค่าสูงสุดของขนาดอิทธิพลมีแนวโน้มที่จะเพิ่มมากขึ้นเมื่อความเข้มของอิทธิพลที่แท้จริงมีค่าเพิ่มขึ้น นอกจากนี้ยังพบว่าค่าเฉลี่ยของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเพิ่มขึ้นเมื่อความเข้มของอิทธิพลที่แท้จริงมีค่าเพิ่มขึ้น สำหรับการกระจายของขนาดอิทธิพลมีลักษณะเบ้ทางบวกมากกว่าการกระจายของสัมประสิทธิ์สหสัมพันธ์ทุกระดับของค่าความเข้มของอิทธิพลที่แท้จริง นอกจากนี้ยังพบว่าขนาดอิทธิพลมีกระจายที่โด่งสูงกว่าสัมประสิทธิ์สหสัมพันธ์เช่นกัน

เมื่อพิจารณาเงื่อนไขขนาดกลุ่มตัวอย่าง (NSAMP) พบว่า ค่าต่ำสุดของขนาดอิทธิพลมีค่าต่ำกว่าค่าต่ำสุดของสัมประสิทธิ์สหสัมพันธ์ทุกขนาดกลุ่มตัวอย่าง ขณะที่ค่าสูงสุดของขนาดอิทธิพลมีค่าสูงกว่าค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ทุกขนาดกลุ่มตัวอย่าง พิสัยของขนาดอิทธิพลจึงกว้างกว่าพิสัยของสัมประสิทธิ์สหสัมพันธ์ ค่าต่ำสุดของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีแนวโน้มที่จะต่ำมากขึ้นเมื่อกลุ่มตัวอย่างเล็กลง นั่นคือ ค่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าต่ำที่สุดเมื่อกลุ่มตัวอย่างมีขนาด 20 คน สำหรับค่าสูงสุดของขนาดอิทธิพลมีแนวโน้มที่จะสูงมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเล็กลง นั่นคือ ค่าขนาดอิทธิพลมีค่าสูงที่สุดเมื่อกลุ่มตัวอย่างมี 20 คน ในขณะที่สัมประสิทธิ์สหสัมพันธ์มีค่าเฉลี่ยใกล้เคียงกัน (ประมาณ 1) ในทุกขนาดกลุ่มตัวอย่าง สำหรับค่าเฉลี่ยของขนาดอิทธิพลมีค่าใกล้เคียงกันไม่ว่ากลุ่มตัวอย่างมีขนาดเท่าใดก็ตาม เช่นเดียวกับค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ที่มีค่าใกล้เคียงกัน ในส่วนของการกระจายของขนาดอิทธิพลมีลักษณะเบ้ทางบวกมากกว่าการกระจายของสัมประสิทธิ์สหสัมพันธ์ทุกขนาดกลุ่มตัวอย่าง นอกจากนี้ยังพบว่าขนาดอิทธิพลมีการกระจายที่โด่งสูงกว่าสัมประสิทธิ์สหสัมพันธ์

เมื่อพิจารณาเงื่อนไขความยาวแบบสอบ (NITEM) พบว่า ค่าต่ำสุดของขนาดอิทธิพลมีค่าต่ำกว่าค่าต่ำสุดของสัมประสิทธิ์สหสัมพันธ์ทุกความยาวแบบสอบ ขณะที่ค่าสูงสุดของขนาดอิทธิพลมีค่าสูงกว่าค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ทุกความยาวแบบสอบ พิสัยของขนาดอิทธิพลจึงกว้างกว่าพิสัยของสัมประสิทธิ์สหสัมพันธ์ ค่าต่ำสุดของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีแนวโน้มที่จะต่ำมากขึ้นเมื่อแบบสอบสั้นลง สำหรับค่าสูงสุดของขนาดอิทธิพลมีแนวโน้มสูงมากขึ้นเมื่อแบบสอบยาวมากขึ้น แต่ค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้เคียงกันทุกความยาวแบบสอบ สำหรับค่าเฉลี่ยของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ที่ได้จากแบบสอบยาว 50 ข้อ มีค่าสูงสุด โดยค่าเฉลี่ยของขนาดอิทธิพลมีค่าสูงกว่า นอกจากนี้การกระจายของขนาดอิทธิพลมีลักษณะเบ้ทางบวกมากกว่าและโด่งสูงกว่าสัมประสิทธิ์สหสัมพันธ์อีกด้วย

เมื่อพิจารณาเงื่อนไขโมเดลฐาน (MBASE) พบว่า ค่าต่ำสุดของขนาดอิทธิพลมีค่าต่ำกว่าค่าต่ำสุดของสัมประสิทธิ์สหสัมพันธ์ทุกโมเดล ขณะที่ค่าสูงสุดของขนาดอิทธิพลมีค่าสูงกว่าค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ทุกโมเดล พิสัยของขนาดอิทธิพลจึงกว้างกว่าพิสัยของสัมประสิทธิ์สหสัมพันธ์ โดยโมเดลโลจิสติกแบบสามพารามิเตอร์ (3PLM) ให้ค่าต่ำสุดของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ ขณะที่โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (1PLM) ให้ค่าสูงสุดของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ สำหรับค่าเฉลี่ยของขนาดอิทธิพลมีค่าสูงกว่าค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์เมื่อใช้โมเดลโลจิสติกแบบหนึ่งหรือสามพารามิเตอร์ เป็นโมเดลฐาน การกระจายของขนาดอิทธิพลมีลักษณะเบ้ทางบวกมากกว่าสัมประสิทธิ์สหสัมพันธ์เมื่อใช้โมเดลโลจิสติกแบบหนึ่งหรือสองพารามิเตอร์เป็นโมเดลฐาน แต่เมื่อใช้โมเดลโลจิสติกแบบสามพารามิเตอร์พบว่าขนาดอิทธิพลเบ้ทางลบมากกว่า นอกจากนี้ยังพบว่าขนาดอิทธิพลมีการกระจายที่โด่งสูงกว่าสัมประสิทธิ์สหสัมพันธ์อีกด้วย

เมื่อพิจารณาเงื่อนไขโมเดลประมาณค่า (MUSED) พบว่า ค่าต่ำสุดของขนาดอิทธิพลมีค่าต่ำกว่าค่าต่ำสุดของสัมประสิทธิ์สหสัมพันธ์ทุกโมเดล ขณะที่ค่าสูงสุดของขนาดอิทธิพลมีค่าสูงกว่าค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ทุกโมเดล พิสัยของขนาดอิทธิพลจึงกว้างกว่าพิสัยของสัมประสิทธิ์สหสัมพันธ์ โดยโมเดลการทดสอบแบบดั้งเดิม (CTT) ให้ค่าต่ำสุดและค่าสูงสุดของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ ค่าเฉลี่ยของขนาดอิทธิพลมีค่าสูงกว่าค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ทุกโมเดล โดยค่าเฉลี่ยที่ได้จากโมเดลการทดสอบแบบดั้งเดิมมีค่าสูงกว่าโมเดลการตอบสนองข้อสอบทั้งสองชนิด ในส่วนของการกระจายของขนาดอิทธิพลมีลักษณะเบ้ทางบวกมากกว่าเมื่อใช้โมเดลการทดสอบแบบดั้งเดิมหรือโมเดลการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล แต่เมื่อใช้โมเดลการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูลพบว่าขนาดอิทธิพลเบ้ทางลบมากกว่า โดยขนาดอิทธิพลมีการกระจายที่โด่งสูงกว่าสัมประสิทธิ์สหสัมพันธ์ในทุกโมเดลประมาณค่า

สำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าในด้านความลำเอียงพิจารณาจากค่าความลำเอียง (BIAS) ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ย (RMSD) และส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ด้านความคงเส้นคงวาพิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานระหว่างกลุ่มตัวอย่างขนาด 20 คนและ 2000 คน (Δ) และด้านประสิทธิภาพสัมพัทธ์พิจารณาจากความแปรปรวน (VAR) ของตัวประมาณค่า จำแนกตามตัวแปรเงื่อนไขในการจำลองข้อมูลทั้ง 5 ตัวแปร มีรายละเอียดดังต่อไปนี้

เมื่อพิจารณาเงื่อนไขความเข้มของอิทธิพลที่แท้จริง (TRUES) พบว่า ค่าความลำเอียงของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ สำหรับส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยของขนาดอิทธิพลมีค่าสูงกว่าสัมประสิทธิ์สหสัมพันธ์ทุกระดับของความเข้มของอิทธิพลที่แท้จริง แต่เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ซึ่งพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ พบว่ามีค่าใกล้เคียงกันในทุกระดับของความเข้มของอิทธิพลที่แท้จริง นอกจากนี้ค่าผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์เมื่อค่าความเข้มของอิทธิพลในรูปของขนาดอิทธิพลมีค่า .2, .5 และ .8 ในด้านความแปรปรวน พบว่า ความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์มีค่าต่ำกว่าขนาดอิทธิพลทุกระดับความเข้มของอิทธิพลที่แท้จริง

โดยสรุป สำหรับเงื่อนไขความเข้มของอิทธิพลที่แท้จริง พบว่า ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณความเข้มของอิทธิพลที่ต่ำกว่าค่าที่แท้จริง โดยขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีความลำเอียงใกล้เคียงกัน แต่สัมประสิทธิ์สหสัมพันธ์มีความคงเส้นคงวาสูงกว่าเมื่อค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .2, .5 และ .8 นอกจากนี้สัมประสิทธิ์สหสัมพันธ์มีประสิทธิภาพสัมพัทธ์สูงกว่า

เมื่อพิจารณาเงื่อนไขขนาดกลุ่มตัวอย่าง (NSAMP) พบว่า ค่าความลำเอียงของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ สำหรับส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยของขนาดอิทธิพลมีค่าสูงกว่าสัมประสิทธิ์สหสัมพันธ์ทุกขนาดของกลุ่มตัวอย่าง แต่เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ซึ่งพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ พบว่ามีค่าใกล้เคียงกันในทุกขนาดกลุ่มตัวอย่าง ในด้านความแปรปรวน พบว่า ความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์มีค่าต่ำกว่าขนาดอิทธิพลทุกขนาดกลุ่มตัวอย่าง

โดยสรุป สำหรับเงื่อนไขขนาดกลุ่มตัวอย่าง พบว่า ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณความเข้มของอิทธิพลที่ต่ำกว่าค่าที่แท้จริง โดยขนาดอิทธิพลและ

สัมประสิทธิ์สหสัมพันธ์มีความลำเอียงใกล้เคียงกัน แต่สัมประสิทธิ์สหสัมพันธ์มีประสิทธิภาพสัมพัทธ์สูงกว่า

เมื่อพิจารณาเงื่อนไขความยาวแบบสอบ (NITEM) พบว่า ค่าความลำเอียงของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ สำหรับส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยของขนาดอิทธิพลมีค่าสูงกว่าสัมประสิทธิ์สหสัมพันธ์ทุกความยาวแบบสอบ แต่เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ซึ่งพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ พบว่ามีค่าใกล้เคียงกันในทุกความยาวแบบสอบ นอกจากนี้ค่าผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของสัมประสิทธิ์สหสัมพันธ์มีค่าสูงกว่าขนาดอิทธิพลทุกความยาวแบบสอบ ในด้านความแปรปรวน พบว่า ความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์มีค่าต่ำกว่าขนาดอิทธิพลทุกความยาวแบบสอบ

โดยสรุป สำหรับเงื่อนไขความยาวแบบสอบ พบว่า ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณความเข้มของอิทธิพลที่ต่ำกว่าค่าที่แท้จริง โดยขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีความลำเอียงใกล้เคียงกัน แต่สัมประสิทธิ์สหสัมพันธ์มีความคงเส้นคงวาและประสิทธิภาพสัมพัทธ์สูงกว่า

เมื่อพิจารณาเงื่อนไขโมเดลฐาน (MBASE) พบว่า ค่าความลำเอียงของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ สำหรับส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยของขนาดอิทธิพลมีค่าสูงกว่าสัมประสิทธิ์สหสัมพันธ์ทุกโมเดลฐาน แต่เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ซึ่งพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ พบว่ามีค่าใกล้เคียงกันในทุกโมเดลฐาน นอกจากนี้ค่าผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของสัมประสิทธิ์สหสัมพันธ์มีค่าสูงกว่าขนาดอิทธิพลเมื่อใช้โมเดลโลจิสติกแบบสองหรือสามพารามิเตอร์ (2PLM, 3PLM) เป็นโมเดลฐาน ในด้านความแปรปรวน พบว่า ความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์มีค่าต่ำกว่าขนาดอิทธิพลทุกโมเดลฐาน

โดยสรุป สำหรับเงื่อนไขโมเดลฐาน พบว่า ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณความเข้มของอิทธิพลที่ต่ำกว่าค่าที่แท้จริง โดยขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีความลำเอียงใกล้เคียงกัน แต่สัมประสิทธิ์สหสัมพันธ์มีความคงเส้นคงวาสูงกว่าเมื่อใช้โมเดลโลจิสติกแบบสองหรือสามพารามิเตอร์เป็นโมเดลฐาน นอกจากนี้สัมประสิทธิ์สหสัมพันธ์ยังมีประสิทธิภาพสัมพัทธ์สูงกว่าอีกด้วย

เมื่อพิจารณาเงื่อนไขโมเดลประมาณค่า (MUSED) พบว่า ค่าความลำเอียงของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ สำหรับส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยของขนาดอิทธิพลมีค่าสูงกว่าสัมประสิทธิ์สหสัมพันธ์ทุกโมเดลประมาณค่า แต่เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ซึ่งพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ พบว่ามีค่าใกล้เคียงกันในทุกโมเดลประมาณค่า นอกจากนี้ค่าผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของสัมประสิทธิ์สหสัมพันธ์มีค่าสูงกว่าขนาดอิทธิพลเมื่อใช้โมเดลการทดสอบแบบดั้งเดิม (CTT) และโมเดลการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (IRT1) ในด้านความแปรปรวน พบว่า ความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์มีค่าต่ำกว่าทุกโมเดลประมาณค่า

โดยสรุป สำหรับเงื่อนไขโมเดลประมาณค่า พบว่า ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณความเข้มของอิทธิพลที่ต่ำกว่าค่าที่แท้จริง โดยขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีความลำเอียงใกล้เคียงกัน แต่สัมประสิทธิ์สหสัมพันธ์มีความคงเส้นคงวาสูงกว่าเมื่อใช้โมเดลการทดสอบแบบดั้งเดิม และโมเดลการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า โดยสัมประสิทธิ์สหสัมพันธ์ยังมีประสิทธิภาพสัมพัทธ์สูงกว่า

สำหรับการเปรียบเทียบคุณสมบัติของตัวประมาณค่าระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่ใช้โมเดลประมาณค่าสอดคล้องกับข้อมูล (d_{IRT1}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่ใช้โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (d_{IRT2}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่ใช้โมเดลประมาณค่าสอดคล้องกับข้อมูล (r_{IRT1}) และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่ใช้โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (r_{IRT2}) พบว่า r_{IRT1} มีความลำเอียงต่ำที่สุด (SRMSD = .777) รองลงมาคือ r_{IRT2}, d_{IRT1} , d_{IRT2}, d_{CTT} และ r_{CTT} ตามลำดับ ในด้านความคงเส้นคงวา พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด (Δ = .187) รองลงมาคือ d_{CTT}, d_{IRT1}, r_{IRT1}, d_{IRT2} และ r_{IRT2} ตามลำดับ นอกจากนี้ยังพบว่า r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด (VAR = .002) รองลงมาคือ r_{IRT2}, d_{IRT1}, d_{IRT2}, r_{CTT} และ d_{CTT} ตามลำดับ

รายละเอียดของค่าต่ำสุด (MIN) ค่าสูงสุด (MAX) ค่าเฉลี่ย (MEAN) ความเบ้ (SK) และความโด่ง (KU) ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ย (RMSD) ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของตัวประมาณค่าทั้งหมดในภาพรวม มีดังตารางที่ 4.3

ตารางที่ 4.3 ค่าสถิติเบื้องต้นและสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าในภาพรวม

สถิติ		ค่าความเข้มของอิทธิพลที่แท้จริง[1]					ขนาดกลุ่มตัวอย่าง			
		.2	.5	.8	1.2	2.6	20	50	500	2000
MIN	d	-1.069	-1.040	-1.177	-3.630	-2.609	-3.630	-2.347	-2.110	-2.156
	r	-.595	-.583	-.640	-.999	-.991	-.999	-.953	-.910	-.919
MAX	d	1.183	1.336	1.747	2.278	4.409	4.409	3.282	2.795	2.626
	r	.642	.700	.826	.940	1.000	.999	1.000	.999	.997
MEAN	d	.025	.057	.088	.118	.293	.108	.123	.118	.117
	r	.015	.035	.052	.066	.125	.055	.062	.060	.059
SK	d	.811	1.131	1.463	.133	1.629	1.801	2.938	3.042	3.054
	r	.735	1.051	1.298	.695	.993	1.008	1.759	1.984	2.013
KU	d	10.425	3.766	3.259	7.388	2.719	14.057	13.253	13.510	13.516
	r	9.055	3.243	2.521	2.975	1.245	5.480	5.947	7.041	7.155
RMSD	d	.224	.487	.766	1.178	2.449	1.308	1.277	1.276	1.277
	r	.118	.241	.361	.510	.756	.467	.452	.449	.449
SRMSD	d	.896	.620	.420	.556	1.825	1.029	.996	.987	.988
	r	.926	.622	.568	.894	1.619	1.048	.996	.975	.975
BIAS	d	-.175	-.443	-.712	-1.082	-2.307	-.952	-.937	-.942	-.943
	r	-.084	-.208	-.320	-.448	-.667	-.349	-.342	-.345	-.345
Δ	d	.033	.034	.030	.232	.004	N/A[2]			
	r	.093	.100	.058	.225	-.015				
VAR[3]	d	.019	.041	.080	.216	.676	.260	.216	.193	.192
	r	.007	.015	.028	.059	.126	.056	.049	.044	.044

หมายเหตุ [1] ค่า .2, .5, .8, 1.2 และ 2.6 เป็นความเข้มของอิทธิพลที่แท้จริงในรูปของขนาดอิทธิพล ซึ่งมีค่าประมาณความเข้มของอิทธิพลที่แท้จริงในรูปของสัมประสิทธิ์สหสัมพันธ์ .1, .2, .4, .5 และ .8 ตามลำดับ

[2] N/A หมายถึง ไม่สามารถคำนวณ $\Delta = SRMSD_{NMin} - SRMSD_{NMax}$ ได้เนื่องจากเป็นสถานการณ์ที่มีกลุ่มตัวอย่างเพียงขนาดเดียว

[3] ค่า VAR คำนวณจากกลุ่มตัวอย่างทุกขนาดรวมกัน

ตารางที่ 4.3 (ต่อ) ค่าสถิติเบื้องต้นและสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าในภาพรวม

สถิติ		ความยาวแบบสอบ			โมเดลฐาน			โมเดลประมาณค่า		
		10	50	90	1PLM	2PLM	3PLM	CTT	IRT1	IRT2
MIN	d	-3.630	-2.609	-2.105	-.693	-2.609	-3.630	-3.630	-.693	-1.572
	r	-.999	-.991	-.908	-.414	-.991	-.999	-.999	-.414	-.775
MAX	d	3.261	3.955	4.409	4.409	3.273	1.711	4.409	.708	.481
	r	.999	1.000	.999	1.000	.999	.816	1.000	.423	.297
MEAN	d	.094	.132	.123	.355	-.016	.011	.317	.024	.007
	r	.049	.067	.061	.178	-.010	.009	.156	.015	.005
SK	d	1.715	2.975	3.085	2.320	.219	-3.768	1.044	2.535	-2.563
	r	1.275	1.793	1.820	1.655	.136	-1.557	.380	2.462	-2.187
KU	d	13.755	13.726	14.027	5.015	11.555	40.997	2.753	12.712	19.106
	r	6.900	6.147	6.029	1.658	6.541	13.054	.353	11.972	14.215
RMSD	d	1.306	1.269	1.280	1.069	1.409	1.351	1.185	1.316	1.349
	r	.464	.446	.454	.365	.513	.474	.445	.449	.469
SRMSD	d	1.009	.993	1.000	.961	1.054	.984	1.084	.937	.975
	r	1.009	.987	1.000	1.044	1.046	.899	1.291	.777	.849
BIAS	d	-.966	-.928	-.937	-.705	-1.076	-1.049	-.743	-1.036	-1.053
	r	-.355	-.338	-.343	-.226	-.414	-.395	-.248	-.389	-.399
Δ	d	.096	.016	.010	.010	-.003	.118	.129	.001	-.018
	r	.162	.034	.019	-.002	.009	.229	.187	.009	-.038
VAR[(2)]	d	.205	.217	.222	.404	.094	.062	.567	.006	.012
	r	.048	.048	.049	.077	.029	.018	.124	.002	.004

หมายเหตุ [(1)] ค่า .2, .5, .8, 1.2 และ 2.6 เป็นความเข้มของอิทธิพลที่แท้จริงในรูปของขนาดอิทธิพล ซึ่งมีค่าประมาณความเข้มของอิทธิพลที่แท้จริงในรูปของสัมประสิทธิ์สหสัมพันธ์ .1, .2, .4, .5 และ .8 ตามลำดับ

[(2)] ค่า VAR คำนวณจากกลุ่มตัวอย่างทุกขนาดรวมกัน

ผู้วิจัยได้เปรียบเทียบค่าเฉลี่ยของขนาดอิทธิพลในแต่ละระดับของตัวแปรเงื่อนไขการจำลองข้อมูล และเปรียบเทียบค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ในแต่ละระดับของตัวแปรเงื่อนไขการจำลองข้อมูล ด้วยการวิเคราะห์ความแปรปรวนทางเดียว ได้ผลดังต่อไปนี้

ผลการวิเคราะห์ความแปรปรวนโดยใช้ความเข้มของอิทธิพลที่แท้จริง (TRUES) เป็นตัวแปรต้น พบว่า ความแปรปรวนของขนาดอิทธิพลในแต่ละระดับของความเข้มของอิทธิพลที่แท้จริงมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ต่อมาเมื่อทดสอบความแตกต่างของค่าเฉลี่ยของขนาดอิทธิพลระหว่างระดับความเข้มของอิทธิพลที่แท้จริงทั้ง 5 ระดับด้วยสถิติทดสอบเอฟ (F-test) พบว่า มีอย่างน้อย 2 ระดับที่มีค่าเฉลี่ยของขนาดอิทธิพลต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทำการเปรียบเทียบรายคู่ภายหลัง (post hoc comparisons) พบว่า ค่าเฉลี่ยของขนาดอิทธิพลในกลุ่มของความเข้มของอิทธิพลที่แท้จริง (ในรูปของขนาดอิทธิพล) ระดับ 2.6 มีค่าสูงสุด รองลงมาคือ 1.2, .8, .5 และ .2 ตามลำดับ สำหรับสัมประสิทธิ์สหสัมพันธ์ พบว่า ความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์ในแต่ละระดับของความเข้มของอิทธิพลที่แท้จริงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับความเข้มของอิทธิพลที่แท้จริงทั้ง 5 ระดับ พบว่า มีอย่างน้อย 2 ระดับที่มีค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และผลการเปรียบเทียบรายคู่ภายหลัง พบว่า ค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ในกลุ่มของความเข้มของอิทธิพลที่แท้จริง (ในรูปของสัมประสิทธิ์สหสัมพันธ์) ระดับ .8 มีค่าสูงสุด รองลงมาคือ .5, .4, .2 และ .1 ตามลำดับ

ผลการวิเคราะห์ความแปรปรวนโดยใช้ขนาดกลุ่มตัวอย่าง (NSAMP) เป็นตัวแปรต้น พบว่า ความแปรปรวนของขนาดอิทธิพลในกลุ่มตัวอย่างแต่ละขนาดมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เช่นเดียวกับความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์ในกลุ่มตัวอย่างแต่ละขนาดที่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของขนาดอิทธิพลระหว่างขนาดกลุ่มตัวอย่างทั้ง 4 กลุ่มด้วยสถิติทดสอบเอฟ ไม่พบว่าค่าเฉลี่ยของขนาดอิทธิพลในขนาดกลุ่มตัวอย่างคู่ใดที่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เช่นเดียวกับผลการทดสอบความแตกต่างของค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ระหว่างขนาดกลุ่มตัวอย่างทั้ง 4 กลุ่มด้วยสถิติทดสอบเอฟ ซึ่งไม่พบว่าค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ในขนาดกลุ่มตัวอย่างคู่ใดที่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ผลการวิเคราะห์ความแปรปรวนโดยใช้ความยาวแบบสอบ (NITEM) เป็นตัวแปรต้น พบว่า ความแปรปรวนของขนาดอิทธิพลในความยาวแบบสอบแต่ละกลุ่มมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เช่นเดียวกับความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์ในความยาวแบบสอบแต่ละกลุ่มที่แตกต่างกันอย่างระดับนัยสำคัญทางสถิติ .05 เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของขนาดอิทธิพลระหว่างแบบสอบ 3 ความยาวด้วยสถิติทดสอบเอฟ พบว่า มีความยาว

แบบสอบอย่างน้อย 2 ค่าที่มีค่าเฉลี่ยของขนาดอิทธิพลต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทำการเปรียบเทียบรายคู่ภายหลัง พบว่า แบบสอบความยาว 10 ข้อ มีค่าเฉลี่ยของขนาดอิทธิพลต่ำสุด และเมื่อทดสอบความแตกต่างของค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบสอบ 3 ความยาวด้วยสถิติทดสอบเอฟ พบว่า มีความยาวแบบสอบอย่างน้อย 2 ค่าที่มีค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทำการเปรียบเทียบรายคู่ภายหลัง พบว่า แบบสอบความยาว 10 ข้อ มีค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ต่ำสุดเช่นกัน

ผลการวิเคราะห์ความแปรปรวนโดยใช้โมเดลฐาน (MBASE) เป็นตัวแปรต้น พบว่า ความแปรปรวนของขนาดอิทธิพลในโมเดลฐานแต่ละชนิดมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เช่นเดียวดับความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์ในโมเดลฐานแต่ละชนิดที่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของขนาดอิทธิพลระหว่างโมเดลฐานทั้ง 3 โมเดลด้วยสถิติทดสอบเอฟ พบว่า มีอย่างน้อย 2 โมเดลที่มีค่าเฉลี่ยของขนาดอิทธิพลต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทำการเปรียบเทียบรายคู่ภายหลัง พบว่าโมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (1PLM) มีค่าเฉลี่ยของขนาดอิทธิพลสูงสุด รองลงมาคือโมเดลโลจิสติกแบบสามพารามิเตอร์ (3PLM) และโมเดลโลจิสติกแบบสองพารามิเตอร์ (2PLM) ตามลำดับ สำหรับสัมประสิทธิ์สหสัมพันธ์ พบว่า ความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์ในแต่ละระดับของความเข้มของอิทธิพลที่แท้จริงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ระหว่างโมเดลฐานทั้ง 3 โมเดล พบว่า มีอย่างน้อย 2 โมเดลที่มีค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และผลการเปรียบเทียบรายคู่ภายหลัง พบว่าโมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ มีค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์สูงสุด รองลงมาคือโมเดลโลจิสติกแบบสามพารามิเตอร์ และโมเดลโลจิสติกแบบสองพารามิเตอร์ ตามลำดับ

ผลการวิเคราะห์ความแปรปรวนโดยใช้โมเดลประมาณค่า (MUSED) เป็นตัวแปรต้น พบว่า ความแปรปรวนของขนาดอิทธิพลในโมเดลประมาณค่าแต่ละชนิดมีความแตกต่างกันอย่างมีนัย-สำคัญทางสถิติที่ระดับ .05 แต่ความแปรปรวนของสัมประสิทธิ์สหสัมพันธ์ในโมเดลประมาณค่าแต่ละชนิดไม่แตกต่างกันที่ระดับนัยสำคัญทางสถิติ .05 เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของขนาดอิทธิพลระหว่างโมเดลประมาณค่าทั้ง 3 โมเดลด้วยสถิติทดสอบเอฟ พบว่า มีอย่างน้อย 2 โมเดลที่มีค่าเฉลี่ยของขนาดอิทธิพลต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทำการทดสอบรายคู่ภายหลัง พบว่า โมเดลการทดสอบแบบดั้งเดิม (CTT) มีค่าเฉลี่ยของขนาดอิทธิพลสูงสุด รองลงมาคือ โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูล (IRT1) และโมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูล (IRT2) ตามลำดับ สำหรับสัมประสิทธิ์สหสัมพันธ์ เมื่อทดสอบความแตกต่างของค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ระหว่างโมเดลประมาณค่าทั้ง 3 โมเดลด้วย

สถิติทดสอบเอฟ พบว่า มีอย่างน้อย 2 โมเดลที่มีค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อทำการทดสอบรายคู่ภายหลัง พบว่า โมเดลการทดสอบแบบดั้งเดิม (CTT) มีค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์สูงที่สุด รองลงมาคือ โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูล (IRT1) และโมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูล (IRT2) ตามลำดับ รายละเอียดของผลการวิเคราะห์ความแปรปรวนทั้งหมดมีดังตารางที่ 4.4-4.5

ตารางที่ 4.4 ผลการทดสอบความแปรปรวนทางเดียวของค่าประมาณความเข้มของอิทธิพลเฉลี่ย

ตัวแปรต้น	df_{BG}	df_{WG}	Levene stat.	sig.	SS_{BG}	SS_{WG}	MS_{BG}	MS_{WG}	F	sig.
TRUES	4	53995	2910.207	.000	475.138	11150.673	118.785	.207	575.192	.000
			2338.039	.000	75.678	2538.320	18.920	.047	402.457	.000
NSAMP	3	53996	24.092	.000	1.588	11624.223	.529	.215	2.458	.061
			32.788	.000	.309	2613.689	.103	.048	2.127	.094
NITEM	2	53997	26.415	.000	13.608	11612.203	6.804	.215	31.638	.000
			25.025	.000	2.823	2611.176	1.411	.048	29.184	.000
MBASE	2	53997	5314.916	.000	1538.474	10087.337	769.237	.187	4117.685	.000
			5096.688	.000	384.680	2229.318	192.340	.041	4658.724	.000
MUSED	2	53997	14087.510	.000	1093.103	10532.708	546.552	.195	2801.953	.000
			18976.711	.000	257.396	2356.602	128.698	.044	2948.870	.000

หมายเหตุ ค่าสถิติเหนือเส้นประเป็นของขนาดอิทธิพล (d) ค่าสถิติใต้เส้นประเป็นของสัมประสิทธิ์สหสัมพันธ์ (r)

ตารางที่ 4.5 ผลการทดสอบรายคู่ที่มีนัยสำคัญทางสถิติที่ระดับ .05

ตัวแปรต้น	ตัวประมาณค่า	คู่ที่ค่าเฉลี่ยแตกต่างกัน
TRUES	ขนาดอิทธิพล	.2 < .5 < .8 < 1.2 < 2.6
	สัมประสิทธิ์สหสัมพันธ์	.1 < .2 < .4 < .5 < .8
NSAMP	ขนาดอิทธิพล	ไม่พบว่าคู่ใดแตกต่างกันอย่างมีนัยสำคัญฯ
	สัมประสิทธิ์สหสัมพันธ์	ไม่พบว่าคู่ใดแตกต่างกันอย่างมีนัยสำคัญฯ
NITEM	ขนาดอิทธิพล	10 < 50, 10 < 90
	สัมประสิทธิ์สหสัมพันธ์	10 < 50, 10 < 90
MBASE	ขนาดอิทธิพล	2PLM < 3PLM < 1PLM
	สัมประสิทธิ์สหสัมพันธ์	2PLM < 3PLM < 1PLM
MUSED	ขนาดอิทธิพล	IRT2 < IRT1 < CTT
	สัมประสิทธิ์สหสัมพันธ์	IRT2 < IRT1 < CTT

ผลการเปรียบเทียบค่าเฉลี่ยของขนาดอิทธิพล (d) ด้วยการวิเคราะห์ความแปรปรวนโดยใช้โมเดลประมาณค่า (MUSED) เป็นตัวแปรต้น ประกอบกับการเปรียบเทียบรายคู่ภายหลัง (post hoc comparisons) นั้น ทำให้ได้ผลการเปรียบเทียบค่าเฉลี่ยระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่ามีความสอดคล้องกับข้อมูล (d_{IRT1}) และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (d_{IRT2}) โดยปริยาย ซึ่งพบว่า ค่าเฉลี่ยของ d_{CTT}, d_{IRT1} และ d_{IRT2} มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดย d_{CTT} มีค่าเฉลี่ยสูงสุด รองลงมาคือ d_{IRT1} และ d_{IRT2} ตามลำดับ

ในทำนองเดียวกันผลการเปรียบเทียบค่าเฉลี่ยของสัมประสิทธิ์สัมพันธ์ (r) ที่ได้จากการวิเคราะห์ความแปรปรวนโดยใช้โมเดลประมาณค่าเป็นตัวแปรต้น ประกอบกับการเปรียบเทียบรายคู่ภายหลัง ทำให้ได้ผลการเปรียบเทียบค่าเฉลี่ยระหว่างสัมประสิทธิ์สัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) สัมประสิทธิ์สัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่ามีความสอดคล้องกับข้อมูล (r_{IRT1}) และสัมประสิทธิ์สัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (r_{IRT2}) โดยปริยายเช่นกัน และพบว่าค่าเฉลี่ยของ r_{CTT}, r_{IRT1} และ r_{IRT2} มีความแตกต่างกันที่ระดับนัยสำคัญทางสถิติ .05 โดย r_{CTT} มีค่าเฉลี่ยสูงสุด รองลงมาคือ r_{IRT1} และ r_{IRT2} ตามลำดับ

ตอนที่ 4 ผลการพิจารณาคุณสมบัติของตัวประมาณค่ารายสถานการณ์

ในส่วนนี้ผู้วิจัยแบ่งการนำเสนอส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองมาตรฐานระหว่างสถานการณ์ที่กลุ่มตัวอย่างมีขนาด 20และ 2000 คน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในแต่ละสถานการณ์ย่อย ออกเป็น 15 กลุ่มสถานการณ์ ตามจำนวนค่าความเข้มของอิทธิพลที่แท้จริงและความยาวแบบสอบ ดังต่อไปนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

<u>กลุ่มที่ 1 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .2 และแบบสอบยาว 10 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 12 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า อย่างไรก็ดีเมื่อพิจารณาในสถานการณ์เดียวกัน พบว่า ค่า SRMSD ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้เคียงกันมาก เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 8 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า พบว่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2} และ d_{IRT1}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 1 มีดังตารางที่ 4.6

ตารางที่ 4.6 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .2 และแบบสอบยาว 10 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
1	1PLM	CTT	20	1.1121	1.3550	-.0316	.0027	.0583	.1562	.1091	.0403
2			50	1.1358	1.3627	.0265	.0401			.0394	.0145
3			500	1.0559	1.2046	-.0254	.0094			.0040	.0015
4			2000	1.0538	1.1988	-.0256	.0093			.0010	.0004
5		IRT1	20	.8645	.8510	-.1905	-.0935	.0042	.0085	.0004	.0002
6			50	.8640	.8496	-.1907	-.0936			.0002	.0001
7			500	.8611	.8439	-.1932	-.0952			.0000	.0000
8			2000	.8603	.8425	-.1938	-.0956			.0000	.0000
9		IRT2	20	.8602	.8435	-.1945	-.0960	-.0017	-.0019	.0008	.0003
10			50	.8635	.8489	-.1913	-.0940			.0004	.0002
11			500	.8615	.8448	-.1928	-.0950			.0000	.0000
12			2000	.8619	.8454	-.1925	-.0948			.0000	.0000
13	2PLM	CTT	20	.9136	1.0525	-.2269	-.1150	.0628	.2194	.1217	.0446
14			50	.8847	.9332	-.2004	-.1000			.0427	.0164
15			500	.8717	.8752	-.1905	-.0936			.0101	.0039
16			2000	.8508	.8331	-.2065	-.1036			.0065	.0026
17		IRT1	20	.8554	.8338	-.1983	-.0984	.0027	.0054	.0003	.0001
18			50	.8538	.8307	-.1996	-.0993			.0002	.0001
19			500	.8540	.8310	-.1993	-.0991			.0001	.0000
20			2000	.8527	.8284	-.2005	-.0998			.0000	.0000
21		IRT2	20	.8573	.8465	-.2014	-.1004	.0013	.0105	.0074	.0030
22			50	.8571	.8409	-.1990	-.0989			.0035	.0014
23			500	.8505	.8269	-.2037	-.1018			.0020	.0008
24			2000	.8560	.8360	-.1984	-.0985			.0012	.0005
25	3PLM	CTT	20	.8946	1.0115	-.2312	-.1184	.0460	.1859	.1015	.0387
26			50	.8745	.9118	-.2068	-.1037			.0383	.0148
27			500	.8745	.8758	-.1854	-.0904			.0061	.0024
28			2000	.8486	.8256	-.2065	-.1036			.0037	.0015
29		IRT1	20	.8613	.8469	-.1941	-.0958	.0075	.0159	.0018	.0007
30			50	.8522	.8300	-.2021	-.1008			.0019	.0007
31			500	.8562	.8364	-.1982	-.0984			.0012	.0005
32			2000	.8538	.8310	-.1998	-.0993			.0004	.0002
33		IRT2	20	.8604	.8449	-.1948	-.0962	.0056	.0126	.0016	.0007
34			50	.8561	.8357	-.1980	-.0982			.0007	.0003
35			500	.8541	.8311	-.1993	-.0990			.0001	.0000
36			2000	.8548	.8323	-.1987	-.0987			.0000	.0000

<u>กลุ่มที่ 2 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .2 และแบบสอบยาว 50 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 10 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า อย่างไรก็ดีเมื่อพิจารณาในสถานการณ์เดียวกัน พบว่า ค่า SRMSD ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้เคียงกันมาก เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลทั้ง 9 สถานการณ์ย่อย เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบสองพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า พบว่า ค่า Δ ของขนาดอิทธิพลมีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2} และ d_{IRT1}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 2 มีดังตารางที่ 4.7

ตารางที่ 4.7 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .2 และแบบสอบยาว 50 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
37	1PLM	CTT	20	1.1676	1.4460	.0122	.0284	.0875	.1992	.1235	.0442
38			50	1.1420	1.3720	.0362	.0466			.0312	.0117
39			500	1.0702	1.2301	-.0126	.0173			.0034	.0013
40			2000	1.0801	1.2468	-.0027	.0235			.0010	.0004
41		IRT1	20	.8735	.8683	-.1828	-.0886	.0115	.0227	.0008	.0003
42			50	.8608	.8441	-.1937	-.0955			.0005	.0002
43			500	.8606	.8431	-.1936	-.0955			.0001	.0000
44			2000	.8620	.8456	-.1923	-.0947			.0000	.0000
45		IRT2	20	.8639	.8501	-.1911	-.0939	.0007	.0023	.0006	.0003
46			50	.8619	.8458	-.1927	-.0949			.0003	.0001
47			500	.8616	.8449	-.1927	-.0949			.0000	.0000
48			2000	.8632	.8478	-.1913	-.0941			.0000	.0000
49	2PLM	CTT	20	.9444	1.0771	-.1757	-.0841	.0738	.2068	.0887	.0339
50			50	.9033	.9730	-.1890	-.0925			.0510	.0194
51			500	.8802	.8900	-.1826	-.0886			.0095	.0037
52			2000	.8706	.8703	-.1897	-.0930			.0074	.0029
53		IRT1	20	.8539	.8322	-.2001	-.0996	-.0001	.0012	.0011	.0005
54			50	.8528	.8294	-.2007	-.1000			.0006	.0002
55			500	.8526	.8284	-.2006	-.0999			.0001	.0000
56			2000	.8540	.8310	-.1993	-.0991			.0001	.0000
57		IRT2	20	.8647	.8602	-.1949	-.0963	.0100	.0272	.0075	.0030
58			50	.8688	.8660	-.1906	-.0936			.0064	.0025
59			500	.8503	.8268	-.2040	-.1020			.0022	.0009
60			2000	.8547	.8330	-.1992	-.0990			.0007	.0003
61	3PLM	CTT	20	.9266	1.0735	-.2155	-.1082	.0680	.2294	.1234	.0453
62			50	.8567	.8924	-.2291	-.1173			.0475	.0181
63			500	.8597	.8497	-.1988	-.0988			.0068	.0027
64			2000	.8586	.8441	-.1978	-.0981			.0038	.0015
65		IRT1	20	.8615	.8472	-.1940	-.0957	.0024	.0068	.0019	.0007
66			50	.8570	.8384	-.1976	-.0980			.0015	.0006
67			500	.8589	.8409	-.1955	-.0967			.0008	.0003
68			2000	.8591	.8404	-.1950	-.0964			.0002	.0001
69		IRT2	20	.8718	.8667	-.1851	-.0900	.0135	.0279	.0020	.0008
70			50	.8570	.8371	-.1970	-.0976			.0005	.0002
71			500	.8588	.8397	-.1952	-.0965			.0001	.0000
72			2000	.8583	.8388	-.1956	-.0967			.0000	.0000

<u>กลุ่มที่ 3 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .2 และแบบสอบยาว 90 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 10 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า อย่างไรก็ดีเมื่อพิจารณาในสถานการณ์เดียวกัน พบว่า ค่า SRMSD ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้เคียงกันมาก เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 8 จาก 9 สถานการณ์ย่อย เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบสองพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า พบว่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2} และ d_{IRT2}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 3 มีดังตารางที่ 4.8

ตารางที่ 4.8 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .2 และแบบสอบยาว 90 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
73	1PLM	CTT	20	1.1717	1.4579	.0120	.0282	.0914	.2107	.1314	.0474
74			50	1.1501	1.3864	.0419	.0498			.0341	.0126
75			500	1.0741	1.2374	-.0094	.0193			.0037	.0014
76			2000	1.0803	1.2472	-.0025	.0236			.0010	.0004
77		IRT1	20	.8716	.8654	-.1849	-.0899	.0112	.0228	.0014	.0006
78			50	.8615	.8455	-.1931	-.0952			.0006	.0002
79			500	.8602	.8424	-.1939	-.0957			.0001	.0000
80			2000	.8604	.8426	-.1938	-.0956			.0000	.0000
81		IRT2	20	.8674	.8576	-.1884	-.0922	.0050	.0112	.0013	.0005
82			50	.8642	.8503	-.1908	-.0937			.0005	.0002
83			500	.8625	.8466	-.1919	-.0944			.0000	.0000
84			2000	.8624	.8464	-.1920	-.0945			.0000	.0000
85	2PLM	CTT	20	.9683	1.1304	-.1706	-.0820	.1219	.3049	.1160	.0432
86			50	.8409	.8793	-.2516	-.1312			.0588	.0223
87			500	.8681	.8699	-.1943	-.0960			.0111	.0043
88			2000	.8464	.8255	-.2104	-.1060			.0067	.0026
89		IRT1	20	.8436	.8131	-.2092	-.1054	-.0069	-.0114	.0013	.0005
90			50	.8492	.8229	-.2040	-.1020			.0007	.0003
91			500	.8520	.8274	-.2012	-.1003			.0002	.0001
92			2000	.8505	.8245	-.2024	-.1010			.0001	.0000
93		IRT2	20	.8503	.8309	-.2060	-.1033	.0073	.0163	.0052	.0021
94			50	.8354	.8021	-.2184	-.1111			.0042	.0017
95			500	.8495	.8247	-.2043	-.1022			.0016	.0006
96			2000	.8430	.8146	-.2110	-.1064			.0031	.0012
97	3PLM	CTT	20	.9872	1.1644	-.1574	-.0739	.1248	.3135	.1235	.0457
98			50	.8929	.9523	-.1966	-.0973			.0480	.0183
99			500	.8732	.8759	-.1880	-.0920			.0083	.0032
100			2000	.8624	.8509	-.1945	-.0961			.0038	.0015
101		IRT1	20	.8662	.8567	-.1903	-.0934	.0127	.0247	.0024	.0010
102			50	.8608	.8450	-.1943	-.0959			.0013	.0005
103			500	.8586	.8400	-.1957	-.0968			.0005	.0002
104			2000	.8535	.8320	-.2009	-.1001			.0017	.0007
105		IRT2	20	.8605	.8453	-.1948	-.0962	.0031	.0082	.0018	.0007
106			50	.8619	.8463	-.1928	-.0950			.0006	.0002
107			500	.8580	.8383	-.1959	-.0969			.0001	.0000
108			2000	.8574	.8371	-.1964	-.0972			.0000	.0000

<u>กลุ่มที่ 4 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .5 และแบบสอบยาว 10 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 8 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า อย่างไรก็ดีเมื่อพิจารณาในสถานการณ์เดียวกัน พบว่า ส่วนใหญ่ค่า SRMSD ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้เคียงกัน เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT2}

เมื่อพิจารณาผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 7 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า รวมทั้งสถานการณ์ที่ใช้โมเดลโลจิสติกแบบสามพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า พบว่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT2} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT1} โดยค่าความแปรปรวนของ r_{IRT1} และ r_{IRT2} มีค่าใกล้เคียงกันหรือเท่ากันเมื่อกลุ่มตัวอย่างมีขนาดใหญ่

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 4 มีดังตารางที่ 4.9

ตารางที่ 4.9 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .5 และแบบสอบยาว 10 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
109	1PLM	CTT	20	1.0200	1.2874	-.0982	.0008	.0043	.0277	.0765	.0259
110			50	1.0760	1.3682	-.0280	.0428			.0416	.0138
111			500	1.0141	1.2576	-.0621	.0253			.0043	.0015
112			2000	1.0157	1.2597	-.0589	.0273			.0012	.0004
113		IRT1	20	.5590	.5071	-.4850	-.2331	.0314	.1267	.0272	.0101
114			50	.5294	.3843	-.4825	-.2315			.0002	.0001
115			500	.5272	.3797	-.4843	-.2327			.0001	.0000
116			2000	.5276	.3804	-.4840	-.2325			.0001	.0000
117		IRT2	20	.5270	.3818	-.4852	-.2331	-.0041	-.0049	.0007	.0003
118			50	.5345	.3948	-.4784	-.2289			.0005	.0002
119			500	.5300	.3848	-.4820	-.2312			.0001	.0000
120			2000	.5311	.3867	-.4810	-.2306			.0001	.0000
121	2PLM	CTT	20	.6392	.7972	-.5174	-.2525	.0967	.2828	.1299	.0476
122			50	.5961	.6760	-.5084	-.2475			.0813	.0308
123			500	.5455	.5218	-.5099	-.2487			.0382	.0147
124			2000	.5425	.5144	-.5111	-.2495			.0368	.0142
125		IRT1	20	.5102	.3504	-.4998	-.2424	.0024	.0064	.0006	.0002
126			50	.5075	.3444	-.5018	-.2437			.0003	.0001
127			500	.5085	.3452	-.5007	-.2430			.0001	.0000
128			2000	.5078	.3440	-.5014	-.2434			.0001	.0000
129		IRT2	20	.5245	.4136	-.4988	-.2418	.0209	.0639	.0111	.0044
130			50	.5004	.3775	-.5214	-.2557			.0119	.0046
131			500	.4997	.3489	-.5138	-.2511			.0048	.0019
132			2000	.5036	.3497	-.5086	-.2479			.0033	.0013
133	3PLM	CTT	20	.6025	.7445	-.5429	-.2678	.0823	.312	.1162	.0430
134			50	.5936	.6538	-.5001	-.2427			.0716	.0273
135			500	.5313	.4473	-.5012	-.2433			.0187	.0073
136			2000	.5202	.4325	-.5118	-.2499			.0191	.0074
137		IRT1	20	.5055	.3505	-.5062	-.2465	-.0102	-.0099	.0026	.0011
138			50	.5167	.3702	-.4964	-.2403			.0027	.0011
139			500	.5216	.3735	-.4904	-.2365			.0012	.0005
140			2000	.5157	.3604	-.4950	-.2394			.0006	.0002
141		IRT2	20	.5228	.3800	-.4906	-.2366	.0071	.0214	.0023	.0009
142			50	.5206	.3703	-.4909	-.2368			.0008	.0003
143			500	.5167	.3607	-.4936	-.2385			.0001	.0001
144			2000	.5157	.3586	-.4944	-.2390			.0001	.0000

<u>กลุ่มที่ 5 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .5 และแบบสอบยาว 50 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 8 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า อย่างไรก็ดีเมื่อพิจารณาในสถานการณ์เดียวกัน พบว่า ส่วนใหญ่ค่า SRMSD ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้เคียงกัน เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 8 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบสามพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า พบว่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2} และ d_{IRT2}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 5 มีดังตารางที่ 4.10

ตารางที่ 4.10 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .5 และแบบสอบยาว 50 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
145	1PLM	CTT	20	1.1090	1.4326	-.0220	.0442	.0392	.0809	.0860	.0286
146			50	1.1140	1.4295	.0086	.0641			.0366	.0118
147			500	1.0669	1.3472	-.0160	.0520			.0043	.0014
148			2000	1.0698	1.3517	-.0117	.0547			.0010	.0003
149		IRT1	20	.5373	.4041	-.4772	-.2280	.0074	.0195	.0016	.0006
150			50	.5290	.3849	-.4833	-.2320			.0006	.0002
151			500	.5283	.3818	-.4834	-.2321			.0001	.0000
152			2000	.5299	.3846	-.4820	-.2312			.0000	.0000
153		IRT2	20	.5310	.3905	-.4821	-.2312	-.0021	.0000	.0010	.0004
154			50	.5306	.3872	-.4816	-.2310			.0003	.0001
155			500	.5310	.3865	-.4810	-.2306			.0000	.0000
156			2000	.5331	.3905	-.4791	-.2294			.0000	.0000
157	2PLM	CTT	20	.7176	.8793	-.4270	-.1980	.1382	.3174	.1255	.0466
158			50	.6508	.7612	-.4690	-.2233			.0974	.0363
159			500	.5817	.5707	-.4762	-.2277			.0388	.0149
160			2000	.5794	.5619	-.4756	-.2273			.0362	.0139
161		IRT1	20	.5204	.3729	-.4919	-.2374	.0079	.0192	.0015	.0006
162			50	.5119	.3542	-.4985	-.2416			.0008	.0003
163			500	.5110	.3511	-.4989	-.2418			.0004	.0002
164			2000	.5125	.3537	-.4975	-.2410			.0004	.0001
165		IRT2	20	.5275	.4048	-.4915	-.2371	.0214	.0532	.0069	.0027
166			50	.5201	.3927	-.4984	-.2415			.0072	.0028
167			500	.5024	.3546	-.5116	-.2498			.0050	.0020
168			2000	.5061	.3516	-.5058	-.2462			.0027	.0011
169	3PLM	CTT	20	.6267	.7797	-.5230	-.2566	.1072	.3533	.1226	.0459
170			50	.6175	.6791	-.4747	-.2270			.0705	.0267
171			500	.5291	.4503	-.5054	-.2459			.0207	.0080
172			2000	.5195	.4264	-.5105	-.2491			.0175	.0068
173		IRT1	20	.5161	.3664	-.4960	-.2400	-.0028	.0010	.0019	.0008
174			50	.5166	.3686	-.4960	-.2400			.0022	.0009
175			500	.5206	.3703	-.4909	-.2368			.0008	.0003
176			2000	.5189	.3654	-.4919	-.2375			.0003	.0001
177		IRT2	20	.4902	.3188	-.5185	-.2543	-.0329	-.0533	.0017	.0007
178			50	.5264	.3801	-.4856	-.2334			.0006	.0002
179			500	.5221	.3704	-.4888	-.2355			.0001	.0000
180			2000	.5231	.3721	-.4880	-.2350			.0001	.0000

<u>กลุ่มที่ 6 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .5 และแบบสอบยาว 90 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 7 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า อย่างไรก็ดีเมื่อพิจารณาในสถานการณ์เดียวกัน พบว่า ส่วนใหญ่ค่า SRMSD ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้เคียงกัน เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 8 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบสามพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า พบว่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} และ r_{IRT2} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ d_{IRT2}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 6 มีดังตารางที่ 4.11

ตารางที่ 4.11 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .5 และแบบสอบยาว 90 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
181	1PLM	CTT	20	1.1141	1.4434	-.0158	.0483	.0369	.0792	.0832	.0281
182			50	1.1231	1.4441	.0155	.0678			.0388	.0125
183			500	1.0712	1.3545	-.0121	.0542			.0042	.0014
184			2000	1.0772	1.3642	-.0053	.0584			.0011	.0004
185		IRT1	20	.5386	.4112	-.4775	-.2283	.0123	.0334	.0030	.0012
186			50	.5294	.3861	-.4831	-.2319			.0007	.0003
187			500	.5254	.3765	-.4859	-.2337			.0001	.0000
188			2000	.5263	.3778	-.4851	-.2332			.0000	.0000
189		IRT2	20	.5347	.3995	-.4795	-.2295	.0032	.0122	.0016	.0007
190			50	.5331	.3920	-.4796	-.2297			.0004	.0002
191			500	.5306	.3859	-.4814	-.2308			.0000	.0000
192			2000	.5315	.3873	-.4806	-.2304			.0000	.0000
193	2PLM	CTT	20	.6510	.8096	-.5010	-.2431	.1292	.3349	.1273	.0475
194			50	.5570	.6230	-.5421	-.2686			.0751	.0285
195			500	.5347	.5032	-.5183	-.2541			.0365	.0141
196			2000	.5218	.4747	-.5252	-.2583			.0319	.0123
197		IRT1	20	.5057	.3452	-.5045	-.2454	-.0006	.0022	.0013	.0005
198			50	.4977	.3303	-.5114	-.2497			.0012	.0005
199			500	.5056	.3420	-.5038	-.2449			.0006	.0002
200			2000	.5063	.3430	-.5031	-.2445			.0005	.0002
201		IRT2	20	.5195	.3862	-.4971	-.2407	.0316	.0498	.0055	.0022
202			50	.4743	.3120	-.5375	-.2661			.0059	.0023
203			500	.4863	.3262	-.5256	-.2585			.0048	.0019
204			2000	.4879	.3364	-.5261	-.2588			.0064	.0025
205	3PLM	CTT	20	.7032	.8708	-.4508	-.2125	.1624	.4106	.1341	.0494
206			50	.5602	.5495	-.5000	-.2426			.0419	.0161
207			500	.5503	.4885	-.4888	-.2356			.0233	.0090
208			2000	.5408	.4602	-.4919	-.2375			.0182	.0071
209		IRT1	20	.5187	.3709	-.4936	-.2385	-.0062	-.0091	.0018	.0007
210			50	.5188	.3690	-.4931	-.2382			.0013	.0005
211			500	.5254	.3793	-.4868	-.2343			.0009	.0003
212			2000	.5249	.3800	-.4878	-.2349			.0014	.0005
213		IRT2	20	.5363	.4041	-.4786	-.2290	.0126	.0309	.0021	.0009
214			50	.5342	.3956	-.4792	-.2294			.0009	.0004
215			500	.5267	.3791	-.4849	-.2331			.0002	.0001
216			2000	.5237	.3732	-.4874	-.2347			.0001	.0000

<u>กลุ่มที่ 7 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .8 และแบบสอบยาว 10 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 12 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า อย่างไรก็ดีเมื่อพิจารณาในสถานการณ์เดียวกัน พบว่า สถานการณ์ส่วนใหญ่ค่า SRMSD ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้เคียงกัน เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT2} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT1}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 8 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ส่วนใหญ่ r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า รวมทั้งสถานการณ์ที่ใช้โมเดลโลจิสติกแบบสามพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า พบว่า ค่า Δ ของขนาดอิทธิพลมีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 7 มีดังตารางที่ 4.12

ตารางที่ 4.12 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .8 และแบบสอบยาว 10 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
217	1PLM	CTT	20	.9600	1.2908	-.1644	.0013	-.0104	-.0226	.0940	.0263
218			50	.9900	1.3393	-.1036	.0377			.0393	.0109
219			500	.9800	1.3274	-.0930	.0458			.0061	.0017
220			2000	.9704	1.3134	-.0985	.0434			.0012	.0003
221		IRT1	20	.1975	.0615	-.7735	-.3546	.0045	.0168	.0006	.0002
222			50	.1924	.0509	-.7769	-.3569			.0002	.0001
223			500	.1941	.0431	-.7751	-.3558			.0001	.0000
224			2000	.1930	.0447	-.7761	-.3564			.0001	.0000
225		IRT2	20	.1953	.0732	-.7762	-.3563	-.0022	.0354	.0009	.0003
226			50	.1998	.0498	-.7708	-.3530			.0004	.0002
227			500	.1976	.0391	-.7721	-.3539			.0001	.0001
228			2000	.1975	.0378	-.7721	-.3539			.0001	.0000
229	2PLM	CTT	20	.5010	.8452	-.8029	-.3720	.1273	.2177	.1727	.0625
230			50	.4201	.7251	-.8228	-.3850			.1207	.0449
231			500	.3896	.6469	-.8034	-.3738			.0966	.0362
232			2000	.3737	.6275	-.8114	-.3785			.0895	.0337
233		IRT1	20	.1662	.1000	-.8005	-.3717	.0013	.0043	.0004	.0002
234			50	.1665	.1073	-.8012	-.3722			.0007	.0003
235			500	.1636	.0992	-.8022	-.3727			.0002	.0001
236			2000	.1649	.0957	-.8009	-.3719			.0002	.0001
237		IRT2	20	.1858	.2399	-.8145	-.3806	.0167	.0402	.0096	.0038
238			50	.1967	.2781	-.8214	-.3844			.0145	.0053
239			500	.1635	.2063	-.8220	-.3851			.0055	.0021
240			2000	.1691	.1997	-.8165	-.3817			.0056	.0022
241	3PLM	CTT	20	.4277	.7298	-.8251	-.3852	.1473	.2861	.1262	.0455
242			50	.3408	.5689	-.8149	-.3812			.0724	.0273
243			500	.2909	.4423	-.7992	-.3711			.0438	.0166
244			2000	.2804	.4437	-.8116	-.3788			.0427	.0163
245		IRT1	20	.1685	.1348	-.8037	-.3737	-.0086	.0341	.0019	.0008
246			50	.1889	.1391	-.7896	-.3649			.0033	.0013
247			500	.1773	.1199	-.7954	-.3685			.0018	.0007
248			2000	.1771	.1007	-.7929	-.3670			.0010	.0004
249		IRT2	20	.1870	.1127	-.7870	-.3631	.0082	.0395	.0020	.0008
250			50	.1755	.0956	-.7935	-.3673			.0008	.0003
251			500	.1786	.0759	-.7892	-.3646			.0003	.0001
252			2000	.1788	.0732	-.7888	-.3644			.0002	.0001

<u>กลุ่มที่ 8 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .8 และแบบสอบยาว 50 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 12 จาก 36 สถานการณ์ เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT2} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT1}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 7 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ส่วนใหญ่ r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในบางสถานการณ์ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT2} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT1} และ d_{IRT1}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 8 มีดังตารางที่ 4.13

ตารางที่ 4.13 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .8 และแบบสอบยาว 50 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
253	1PLM	CTT	20	1.0759	1.4515	-.0664	.0509	.0132	-.0025	.1112	.0281
254			50	1.0732	1.4620	-.0312	.0746			.0430	.0113
255			500	1.0595	1.4476	-.0236	.0816			.0065	.0017
256			2000	1.0627	1.4540	-.0177	.0851			.0008	.0002
257		IRT1	20	.2247	.0938	-.7529	-.3415	.0292	.0561	.0019	.0008
258			50	.1915	.0799	-.7799	-.3587			.0009	.0004
259			500	.1950	.0413	-.7743	-.3553			.0001	.0000
260			2000	.1955	.0377	-.7737	-.3549			.0001	.0000
261		IRT2	20	.1982	.0683	-.7734	-.3545	-.0021	.0412	.0008	.0003
262			50	.1929	.0607	-.7771	-.3570			.0005	.0002
263			500	.1985	.0318	-.7711	-.3532			.0001	.0000
264			2000	.2003	.0271	-.7695	-.3522			.0000	.0000
265	2PLM	CTT	20	.5449	.8578	-.7284	-.3296	.1330	.2342	.1811	.0649
266			50	.4774	.7580	-.7557	-.3438			.1392	.0509
267			500	.4139	.6307	-.7527	-.3421			.0946	.0352
268			2000	.4119	.6236	-.7485	-.3395			.0917	.0344
269		IRT1	20	.1792	.1226	-.7943	-.3678	.0077	.0192	.0020	.0008
270			50	.1774	.1196	-.7952	-.3683			.0018	.0007
271			500	.1707	.1079	-.7984	-.3704			.0010	.0004
272			2000	.1715	.1034	-.7973	-.3697			.0008	.0003
273		IRT2	20	.1915	.2054	-.8009	-.3720	.0293	-.0125	.0075	.0030
274			50	.1874	.2191	-.8079	-.3763			.0083	.0033
275			500	.1617	.2324	-.8301	-.3902			.0070	.0027
276			2000	.1622	.2179	-.8258	-.3875			.0061	.0024
277	3PLM	CTT	20	.4262	.7552	-.8389	-.3954	.1507	.3276	.1284	.0479
278			50	.3451	.6069	-.8378	-.3941			.0801	.0302
279			500	.2817	.4468	-.8111	-.3784			.0431	.0165
280			2000	.2755	.4276	-.8078	-.3763			.0396	.0152
281		IRT1	20	.1769	.1278	-.7967	-.3693	-.0041	.0313	.0021	.0009
282			50	.1775	.1267	-.7962	-.3690			.0021	.0008
283			500	.1801	.1125	-.7923	-.3666			.0017	.0007
284			2000	.1810	.0965	-.7896	-.3649			.0011	.0004
285		IRT2	20	.2098	.1046	-.7674	-.3507	.0142	.0583	.0023	.0009
286			50	.1968	.0792	-.7756	-.3560			.0011	.0004
287			500	.1947	.0485	-.7749	-.3556			.0002	.0001
288			2000	.1956	.0463	-.7740	-.3551			.0002	.0001

<u>กลุ่มที่ 9 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า .8 และแบบสอบยาว 90 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 16 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT2} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT1}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 7 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในบางสถานการณ์ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT2} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT1}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 9 มีดังตารางที่ 4.14

ตารางที่ 4.14 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ .8 และแบบสอบยาว 90 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
289	1PLM	CTT	20	1.0726	1.4448	-.0711	.0480	.0016	-.0215	.1139	.0288
290			50	1.0954	1.4928	-.0123	.0839			.0448	.0114
291			500	1.0690	1.4617	-.0153	.0858			.0064	.0016
292			2000	1.0710	1.4663	-.0105	.0888			.0009	.0002
293		IRT1	20	.2250	.1116	-.7548	-.3428	.0360	.0631	.0027	.0011
294			50	.1853	.0962	-.7864	-.3628			.0013	.0005
295			500	.1906	.0498	-.7782	-.3577			.0001	.0001
296			2000	.1890	.0485	-.7794	-.3585			.0001	.0000
297		IRT2	20	.2015	.0966	-.7735	-.3546	.0042	.0646	.0018	.0007
298			50	.1938	.0669	-.7769	-.3568			.0006	.0003
299			500	.1986	.0319	-.7710	-.3532			.0001	.0000
300			2000	.1973	.0320	-.7720	-.3538			.0000	.0000
301	2PLM	CTT	20	.4234	.7275	-.8249	-.3875	.0870	.1293	.1233	.0450
302			50	.4028	.6900	-.8203	-.3842			.1092	.0406
303			500	.3575	.6174	-.8279	-.3890			.0846	.0318
304			2000	.3364	.5982	-.8423	-.3977			.0765	.0289
305		IRT1	20	.1709	.1340	-.8019	-.3726	.0068	.0125	.0020	.0008
306			50	.1630	.1520	-.8105	-.3780			.0024	.0009
307			500	.1643	.1242	-.8050	-.3745			.0012	.0005
308			2000	.1641	.1215	-.8048	-.3744			.0011	.0004
309		IRT2	20	.1704	.2027	-.8159	-.3814	.0196	-.1196	.0057	.0023
310			50	.1611	.2527	-.8361	-.3941			.0082	.0032
311			500	.1513	.2869	-.8539	-.4050			.0094	.0036
312			2000	.1508	.3223	-.8674	-.4134			.0115	.0045
313	3PLM	CTT	20	.4084	.7071	-.8254	-.3867	.1149	.2736	.1139	.0425
314			50	.3852	.5612	-.7473	-.3393			.0749	.0278
315			500	.3146	.4672	-.7837	-.3613			.0501	.0190
316			2000	.2935	.4335	-.7904	-.3654			.0423	.0162
317		IRT1	20	.1807	.1328	-.7947	-.3680	-.0010	.0085	.0026	.0010
318			50	.1842	.1130	-.7892	-.3646			.0019	.0007
319			500	.1783	.1173	-.7943	-.3678			.0018	.0007
320			2000	.1817	.1243	-.7928	-.3669			.0023	.0009
321		IRT2	20	.2114	.1107	-.7669	-.3504	.0134	.0693	.0026	.0011
322			50	.2175	.0739	-.7576	-.3447			.0012	.0005
323			500	.1996	.0453	-.7708	-.3531			.0003	.0001
324			2000	.1980	.0414	-.7719	-.3538			.0002	.0001

<u>กลุ่มที่ 10 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า 1.2 และแบบสอบยาว 10 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าสูงกว่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 36 สถานการณ์ย่อย เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ d_{IRT2} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ d_{IRT1}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า ขนาดอิทธิพลมีค่า Δ สูงกว่าสัมประสิทธิ์สหสัมพันธ์ใน 5 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่าส่วนใหญ่ r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งและสองพารามิเตอร์เป็นโมเดลฐาน และใช้โมเดลการตอบสนองข้อสอบเป็นโมเดลประมาณค่า พบว่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 10 มีดังตารางที่ 4.15

ตารางที่ 4.15 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 1.2 และแบบสอบยาว 10 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
325	1PLM	CTT	20	.9809	1.4152	-.1809	.0391	.0574	.0038	.1509	.0282
326			50	.9791	1.4598	-.1171	.0780			.0456	.0079
327			500	.9097	1.3924	-.1538	.0660			.0048	.0009
328			2000	.9235	1.4114	-.1398	.0723			.0019	.0004
329		IRT1	20	.2485	.4882	-1.1593	-.4887	-.0061	-.0124	.0006	.0002
330			50	.2512	.4939	-1.1623	-.4907			.0003	.0001
331			500	.2543	.5000	-1.1652	-.4927			.0002	.0001
332			2000	.2546	.5006	-1.1655	-.4929			.0002	.0001
333		IRT2	20	.2646	.5182	-1.1725	-.4971	.0145	.0259	.0010	.0004
334			50	.2501	.4918	-1.1608	-.4898			.0005	.0002
335			500	.2500	.4921	-1.1614	-.4903			.0003	.0001
336			2000	.2501	.4923	-1.1615	-.4904			.0003	.0001
337	2PLM	CTT	20	.6681	1.1757	-1.1952	-.5101	.0777	.1043	.2791	.0948
338			50	.6454	1.1563	-1.2271	-.5295			.2390	.0840
339			500	.5936	1.0779	-1.2086	-.5200			.1999	.0719
340			2000	.5904	1.0714	-1.2095	-.5204			.1965	.0705
341		IRT1	20	.2910	.5677	-1.1967	-.5124	-.0060	-.0112	.0006	.0002
342			50	.2988	.5822	-1.2034	-.5167			.0006	.0002
343			500	.3003	.5849	-1.2047	-.5174			.0006	.0003
344			2000	.2970	.5789	-1.2023	-.5159			.0004	.0002
345		IRT2	20	.3534	.6799	-1.2249	-.5300	-.0126	-.0243	.0162	.0063
346			50	.3833	.7334	-1.2531	-.5475			.0163	.0062
347			500	.3749	.7199	-1.2531	-.5476			.0114	.0044
348			2000	.3660	.7042	-1.2483	-.5446			.0093	.0036
349	3PLM	CTT	20	2.6413	3.3729	-3.1164	-1.3223	2.1935	2.5352	.5914	.0421
350			50	.5010	.9219	-1.1944	-.5112			.1295	.0473
351			500	.4416	.8265	-1.2076	-.5197			.0796	.0296
352			2000	.4478	.8377	-1.2146	-.5241			.0800	.0297
353		IRT1	20	.3675	.7085	-1.2485	-.5451	.0860	.1588	.0101	.0040
354			50	.2926	.5700	-1.1912	-.5090			.0041	.0016
355			500	.2864	.5586	-1.1872	-.5065			.0032	.0013
356			2000	.2815	.5497	-1.1843	-.5047			.0025	.0010
357		IRT2	20	.2709	.5296	-1.1729	-.4973	.0026	.0037	.0034	.0014
358			50	.2660	.5212	-1.1731	-.4976			.0013	.0005
359			500	.2700	.5289	-1.1781	-.5007			.0007	.0003
360			2000	.2683	.5259	-1.1767	-.4999			.0006	.0002

<u>กลุ่มที่ 11 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า 1.2 และแบบสอบยาว 50 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 36 สถานการณ์ย่อย เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ d_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ d_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า ขนาดอิทธิพลมีค่า Δ สูงกว่าสัมประสิทธิ์สหสัมพันธ์ใน 5 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่าส่วนใหญ่ d_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า รวมถึงสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT2} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT1}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 11 มีดังตารางที่ 4.16

ตารางที่ 4.16 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 1.2 และแบบสอบยาว 50 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
361	1PLM	CTT	20	1.1437	1.6057	-.0307	.1011	.0803	.0226	.1621	.0271
362			50	1.1149	1.6239	.0060	.1287			.0431	.0067
363			500	1.0415	1.5557	-.0381	.1147			.0043	.0007
364			2000	1.0634	1.5831	-.0173	.1234			.0012	.0002
365		IRT1	20	.2019	.4000	-1.1091	-.4569	-.0579	-.1102	.0038	.0015
366			50	.2500	.4912	-1.1564	-.4871			.0024	.0010
367			500	.2591	.5089	-1.1693	-.4952			.0003	.0001
368			2000	.2598	.5102	-1.1701	-.4958			.0002	.0001
369		IRT2	20	.2568	.5035	-1.1645	-.4920	.0041	.0064	.0015	.0006
370			50	.2580	.5065	-1.1675	-.4941			.0007	.0003
371			500	.2544	.5003	-1.1656	-.4929			.0001	.0000
372			2000	.2527	.4971	-1.1641	-.4920			.0001	.0000
373	2PLM	CTT	20	.6141	1.0543	-1.0285	-.4136	.0510	.0450	.2824	.0936
374			50	.6307	1.1147	-1.1438	-.4788			.2652	.0919
375			500	.5644	1.0090	-1.1308	-.4723			.2094	.0737
376			2000	.5631	1.0093	-1.1360	-.4749			.2063	.0731
377		IRT1	20	.2951	.5749	-1.1960	-.5119	.0030	.0054	.0027	.0011
378			50	.2861	.5583	-1.1881	-.5070			.0027	.0011
379			500	.2940	.5729	-1.1967	-.5124			.0019	.0007
380			2000	.2921	.5695	-1.1955	-.5117			.0017	.0006
381		IRT2	20	.3444	.6661	-1.2305	-.5338	-.0423	-.0752	.0082	.0033
382			50	.3518	.6778	-1.2285	-.5323			.0133	.0052
383			500	.3959	.7583	-1.2682	-.5569			.0142	.0055
384			2000	.3867	.7413	-1.2624	-.5533			.0125	.0048
385	3PLM	CTT	20	.5611	1.0205	-1.1875	-.5070	.1179	.1891	.1819	.0656
386			50	.5323	.9765	-1.2383	-.5367			.1301	.0473
387			500	.4466	.8381	-1.2219	-.5281			.0751	.0283
388			2000	.4432	.8314	-1.2178	-.5255			.0752	.0283
389		IRT1	20	.2935	.5721	-1.1963	-.5121	.0135	.0253	.0019	.0008
390			50	.2945	.5738	-1.1956	-.5117			.0027	.0011
391			500	.2799	.5464	-1.1794	-.5016			.0042	.0017
392			2000	.2800	.5468	-1.1813	-.5028			.0033	.0013
393		IRT2	20	.2333	.4587	-1.1361	-.4741	-.0018	-.0059	.0045	.0018
394			50	.2282	.4510	-1.1400	-.4767			.0012	.0005
395			500	.2354	.4650	-1.1479	-.4818			.0006	.0002
396			2000	.2351	.4646	-1.1477	-.4817			.0006	.0002

<u>กลุ่มที่ 12 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า 1.2 และแบบสอบยาว 90 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 36 สถานการณ์ย่อย เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ d_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ d_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า ขนาดอิทธิพลมีค่า Δ สูงกว่าสัมประสิทธิ์สหสัมพันธ์ใน 5 จาก 9 สถานการณ์ย่อย เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ส่วนใหญ่ r_{CTT} มีความคงเส้นคงวาสูงสุด นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า รวมถึงสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT2} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT1}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 12 มีดังตารางที่ 4.17

ตารางที่ 4.17 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 1.2 และแบบสอบยาว 90 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
397	1PLM	CTT	20	1.1509	1.6066	-.0272	.1020	.0720	.0053	.1685	.0266
398			50	1.1364	1.6467	.0241	.1356			.0451	.0067
399			500	1.0561	1.5729	-.0255	.1198			.0046	.0007
400			2000	1.0789	1.6013	-.0038	.1288			.0011	.0002
401		IRT1	20	.1989	.3937	-1.1004	-.4515	-.0740	-.1406	.0057	.0023
402			50	.2603	.5102	-1.1644	-.4921			.0030	.0012
403			500	.2736	.5357	-1.1815	-.5029			.0005	.0002
404			2000	.2729	.5343	-1.1815	-.5029			.0002	.0001
405		IRT2	20	.2587	.5070	-1.1653	-.4925	-.0007	-.0026	.0020	.0008
406			50	.2653	.5200	-1.1732	-.4976			.0010	.0004
407			500	.2602	.5110	-1.1704	-.4960			.0002	.0001
408			2000	.2594	.5096	-1.1700	-.4957			.0001	.0000
409	2PLM	CTT	20	.6300	1.1375	-1.2348	-.5368	.0384	.0568	.2195	.0774
410			50	.6281	1.1397	-1.2531	-.5476			.2065	.0735
411			500	.5984	1.0913	-1.2552	-.5489			.1772	.0634
412			2000	.5916	1.0807	-1.2539	-.5479			.1718	.0618
413		IRT1	20	.3067	.5968	-1.2062	-.5184	.0033	.0065	.0028	.0011
414			50	.3120	.6061	-1.2097	-.5206			.0035	.0014
415			500	.3047	.5926	-1.2048	-.5175			.0026	.0010
416			2000	.3034	.5903	-1.2042	-.5171			.0023	.0009
417		IRT2	20	.3538	.6840	-1.2412	-.5406	-.1163	-.2080	.0069	.0028
418			50	.4161	.7971	-1.2895	-.5705			.0124	.0049
419			500	.4638	.8813	-1.3216	-.5899			.0212	.0081
420			2000	.4701	.8920	-1.3235	-.5910			.0243	.0093
421	3PLM	CTT	20	.5122	.9432	-1.1851	-.5061	.0860	.1459	.1428	.0525
422			50	.4565	.8419	-1.1644	-.4935			.1110	.0404
423			500	.4422	.8235	-1.1851	-.5054			.0915	.0340
424			2000	.4262	.7973	-1.1866	-.5063			.0801	.0300
425		IRT1	20	.3010	.5861	-1.2017	-.5156	.0123	.0238	.0025	.0010
426			50	.2918	.5688	-1.1941	-.5108			.0022	.0009
427			500	.2899	.5651	-1.1900	-.5082			.0035	.0014
428			2000	.2887	.5623	-1.1847	-.5049			.0055	.0022
429		IRT2	20	.2161	.4263	-1.1187	-.4630	-.0074	-.0167	.0051	.0020
430			50	.2083	.4138	-1.1205	-.4645			.0019	.0007
431			500	.2267	.4489	-1.1400	-.4769			.0007	.0003
432			2000	.2235	.4430	-1.1366	-.4748			.0009	.0004

<u>กลุ่มที่ 13 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า 2.6 และแบบสอบยาว 10 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 8 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า มีจำนวนสถานการณ์ที่ d_{CTT}, r_{IRT1} และ r_{IRT2} มีความลำเอียงต่ำสุดเท่ากัน

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่า Δ สูงกว่าขนาดอิทธิพลใน 5 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า มีจำนวนสถานการณ์ที่ d_{CTT}, r_{CTT} และ r_{IRT1} มีความคงเส้นคงวาสูงสุดเท่ากัน นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบสองพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบเป็นโมเดลประมาณค่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 13 มีดังตารางที่ 4.18

ตารางที่ 4.18 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 2.6 และแบบสอบยาว 10 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
433	1PLM	CTT	20	.7617	1.5302	-.4713	.0998	.1677	-.0413	.2200	.0074
434			50	.7198	1.5966	-.4056	.1242			.1051	.0032
435			500	.6011	1.5726	-.4657	.1184			.0464	.0020
436			2000	.5940	1.5715	-.4685	.1182			.0426	.0018
437		IRT1	20	1.7950	1.3192	-2.5089	-.7349	.0239	.0422	.0012	.0005
438			50	1.7846	1.3011	-2.4999	-.7297			.0009	.0004
439			500	1.7734	1.2810	-2.4902	-.7239			.0006	.0002
440			2000	1.7711	1.2770	-2.4882	-.7227			.0006	.0002
441		IRT2	20	1.8729	1.4646	-2.5767	-.7779	.0819	.1509	.0014	.0006
442			50	1.8419	1.4070	-2.5499	-.7611			.0009	.0003
443			500	1.8004	1.3309	-2.5137	-.7386			.0007	.0003
444			2000	1.7910	1.3137	-2.5056	-.7335			.0007	.0003
445	2PLM	CTT	20	2.1501	2.2322	-2.5872	-.7896	.0043	.0398	.8235	.2413
446			50	2.1888	2.2431	-2.6147	-.7994			.8603	.2367
447			500	2.1494	2.2043	-2.5884	-.7888			.8170	.2311
448			2000	2.1458	2.1924	-2.5904	-.7886			.7986	.2266
449		IRT1	20	1.9024	1.5194	-2.6025	-.7942	-.0035	-.0073	.0012	.0005
450			50	1.9042	1.5242	-2.6037	-.7949			.0028	.0011
451			500	1.9028	1.5217	-2.6023	-.7941			.0032	.0012
452			2000	1.9059	1.5267	-2.6053	-.7959			.0022	.0009
453		IRT2	20	2.0064	1.7250	-2.6876	-.8476	-.0306	-.0554	.0209	.0082
454			50	2.0322	1.7733	-2.7078	-.8593			.0297	.0112
455			500	2.0460	1.7986	-2.7183	-.8652			.0354	.0129
456			2000	2.0370	1.7804	-2.7125	-.8621			.0277	.0104
457	3PLM	CTT	20	2.0529	1.9131	-2.6521	-.8201	.0619	.1165	.2897	.0959
458			50	2.0098	1.8496	-2.6183	-.8057			.2709	.0884
459			500	1.9808	1.7811	-2.6095	-.8005			.2115	.0712
460			2000	1.9910	1.7966	-2.6198	-.8067			.2079	.0704
461		IRT1	20	1.9054	1.5278	-2.6041	-.7952	.0771	.1399	.0046	.0018
462			50	1.9018	1.5265	-2.5990	-.7920			.0113	.0044
463			500	1.8157	1.3688	-2.5241	-.7453			.0103	.0040
464			2000	1.8283	1.3879	-2.5363	-.7529			.0063	.0025
465		IRT2	20	1.8097	1.3495	-2.5208	-.7424	.0020	.0034	.0041	.0016
466			50	1.8100	1.3503	-2.5213	-.7432			.0031	.0012
467			500	1.8071	1.3451	-2.5190	-.7419			.0025	.0010
468			2000	1.8077	1.3461	-2.5196	-.7423			.0023	.0009

<u>กลุ่มที่ 14 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า 2.6 และแบบสอบยาว 50 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 8 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า ขนาดอิทธิพลมีค่า Δ สูงกว่าสัมประสิทธิ์สหสัมพันธ์ใน 5 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า มีจำนวนสถานการณ์ที่ d_{CTT}, d_{IRT1} และ r_{IRT1} มีความคงเส้นคงวาสูงสุดเท่ากัน นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบเป็นโมเดลประมาณค่า รวมทั้งสถานการณ์ที่โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการทดสอบแบบดั้งเดิมหรือโมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูลเป็นโมเดลประมาณค่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 14 มีดังตารางที่ 4.19

ตารางที่ 4.19 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 2.6 และแบบสอบยาว 50 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
469	1PLM	CTT	20	1.0744	1.6599	-.1832	.1418	.2253	-.0738	.3019	.0045
470			50	.9962	1.7316	-.1391	.1658			.1078	.0019
471			500	.8502	1.7283	-.2153	.1661			.0188	.0005
472			2000	.8491	1.7337	-.2106	.1679			.0104	.0003
473		IRT1	20	1.5591	.8914	-2.3021	-.6059	-.1922	-.3535	.0041	.0016
474			50	1.5395	.8620	-2.2848	-.5968			.0046	.0017
475			500	1.7089	1.1729	-2.4314	-.6876			.0083	.0032
476			2000	1.7513	1.2449	-2.4698	-.7113			.0042	.0016
477		IRT2	20	1.8599	1.4437	-2.5642	-.7699	-.0044	-.0057	.0052	.0021
478			50	1.8549	1.4339	-2.5603	-.7677			.0039	.0015
479			500	1.8646	1.4509	-2.5691	-.7733			.0026	.0010
480			2000	1.8643	1.4494	-2.5693	-.7734			.0014	.0005
481	2PLM	CTT	20	2.0171	2.0243	-2.4046	-.6867	-.0167	-.0086	.9702	.2447
482			50	2.0750	2.0957	-2.4770	-.7182			.9331	.2480
483			500	2.0383	2.0388	-2.4741	-.7166			.8261	.2284
484			2000	2.0338	2.0329	-2.4742	-.7168			.8119	.2261
485		IRT1	20	1.8930	1.5032	-2.5939	-.7888	.0032	.0019	.0026	.0011
486			50	1.8921	1.5037	-2.5924	-.7879			.0051	.0020
487			500	1.8880	1.4977	-2.5882	-.7852			.0070	.0027
488			2000	1.8898	1.5013	-2.5897	-.7862			.0073	.0028
489		IRT2	20	2.0319	1.7679	-2.7121	-.8632	-.1581	-.2920	.0136	.0053
490			50	2.0627	1.8303	-2.7342	-.8759			.0309	.0118
491			500	2.1832	2.0492	-2.8273	-.9293			.0782	.0268
492			2000	2.1900	2.0599	-2.8304	-.9301			.0891	.0297
493	3PLM	CTT	20	2.0533	1.9490	-2.6397	-.8184	.0539	.1411	.3339	.1098
494			50	2.0233	1.8590	-2.6378	-.8132			.2464	.0844
495			500	2.0079	1.8294	-2.6348	-.8140			.2091	.0739
496			2000	1.9994	1.8079	-2.6313	-.8118			.1947	.0691
497		IRT1	20	1.9257	1.5645	-2.6223	-.8068	.1738	.3094	.0030	.0012
498			50	1.8903	1.4989	-2.5913	-.7872			.0034	.0014
499			500	1.7674	1.2819	-2.4816	-.7188			.0111	.0043
500			2000	1.7519	1.2551	-2.4678	-.7103			.0120	.0046
501		IRT2	20	1.6909	1.1328	-2.4165	-.6769	.0120	.0204	.0059	.0023
502			50	1.6728	1.1014	-2.4011	-.6682			.0049	.0019
503			500	1.6727	1.1014	-2.4013	-.6688			.0038	.0014
504			2000	1.6789	1.1124	-2.4068	-.6723			.0035	.0013

<u>กลุ่มที่ 15 ค่าความเข้มของอิทธิพลที่แท้จริงมีค่า 2.6 และแบบสอบยาว 90 ข้อ</u>

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง เมื่อพิจารณาจากค่าความลำเอียง (BIAS) พบว่า ในสถานการณ์ส่วนใหญ่ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะให้ค่าประมาณที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) แต่ในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์จะพบว่าสัมประสิทธิ์สหสัมพันธ์ให้ค่าประมาณที่สูงกว่าค่าความเข้มของอิทธิพลที่แท้จริง (overestimate) เมื่อพิจารณาส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) พบว่า ค่า SRMSD ของขนาดอิทธิพลมีค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์ใน 8 จาก 36 สถานการณ์ ส่วนใหญ่เป็นสถานการณ์ที่ใช้โมเดลการทดสอบแบบดั้งเดิมเป็นโมเดลประมาณค่า เมื่อเปรียบเทียบความลำเอียงระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} จะมีความลำเอียงต่ำที่สุด รองลงมาคือ r_{IRT2}

สำหรับคุณสมบัติของตัวประมาณค่าในด้านความคงเส้นคงวาที่พิจารณาจากผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) พบว่า ขนาดอิทธิพลมีค่า Δ สูงกว่าสัมประสิทธิ์สหสัมพันธ์ใน 5 จาก 9 สถานการณ์ เมื่อเปรียบเทียบความคงเส้นคงวาระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า มีจำนวนสถานการณ์ที่ d_{CTT}, r_{CTT} และ r_{IRT1} มีความคงเส้นคงวาสูงสุดเท่ากัน นอกจากนี้ยังพบว่าในสถานการณ์ที่ใช้โมเดลโลจิสติกแบบหนึ่งหรือสองพารามิเตอร์เป็นโมเดลฐานและใช้โมเดลการตอบสนองข้อสอบเป็นโมเดลประมาณค่า ค่า Δ ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ

สำหรับด้านประสิทธิภาพสัมพัทธ์ที่พิจารณาจากความแปรปรวน (VAR) พบว่า ทั้ง 36 สถานการณ์ย่อยในกลุ่มนี้ สัมประสิทธิ์สหสัมพันธ์มีมีความแปรปรวนต่ำกว่าขนาดอิทธิพล อย่างไรก็ดีเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ (500 และ 2000 คน) ความแปรปรวนของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าใกล้เคียงกัน เมื่อเปรียบเทียบประสิทธิภาพสัมพัทธ์ระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในสถานการณ์ที่ใช้โมเดลฐานและขนาดกลุ่มตัวอย่างเหมือนกัน พบว่า ในสถานการณ์ส่วนใหญ่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2}

รายละเอียดของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ค่าความลำเอียง (BIAS) ผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) และความแปรปรวน (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในสถานการณ์ย่อยกลุ่มที่ 15 มีดังตารางที่ 4.20

ตารางที่ 4.20 ค่าสถิติสำหรับการพิจารณาคุณสมบัติของตัวประมาณค่าที่ดีของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เมื่อความเข้มของอิทธิพลที่แท้จริงเท่ากับ 2.6 และแบบสอบยาว 90 ข้อ

ที่	MBASE	MUSED	NSAMP	RMSD		BIAS		Δ		VAR	
				d	r	d	r	d	r	d	r
505	1PLM	CTT	20	1.1365	1.6754	-.1436	.1460	.2554	-.0752	.3449	.0049
506			50	1.0206	1.7498	-.1058	.1716			.0904	.0014
507			500	.8844	1.7463	-.1840	.1715			.0173	.0004
508			2000	.8811	1.7506	-.1806	.1730			.0076	.0002
509		IRT1	20	1.4296	.6811	-2.1862	-.5365	-.3541	-.6258	.0110	.0040
510			50	1.4307	.6915	-2.1866	-.5388			.0125	.0044
511			500	1.7111	1.1829	-2.4316	-.6879			.0132	.0051
512			2000	1.7837	1.3069	-2.4973	-.7285			.0068	.0026
513		IRT2	20	1.8629	1.4516	-2.5661	-.7711	-.0330	-.0567	.0077	.0031
514			50	1.8622	1.4485	-2.5663	-.7714			.0050	.0020
515			500	1.8904	1.4988	-2.5916	-.7873			.0030	.0012
516			2000	1.8959	1.5083	-2.5965	-.7904			.0023	.0009
517	2PLM	CTT	20	2.2463	2.2955	-2.6764	-.8404	.0481	.0552	.8442	.2185
518			50	2.2236	2.2885	-2.6656	-.8375			.8037	.2184
519			500	2.1988	2.2485	-2.6655	-.8367			.7198	.2033
520			2000	2.1982	2.2403	-2.6692	-.8368			.7049	.1999
521		IRT1	20	1.9072	1.5305	-2.6059	-.7964	-.0052	-.0135	.0038	.0015
522			50	1.9098	1.5370	-2.6075	-.7974			.0060	.0024
523			500	1.9098	1.5390	-2.6068	-.7969			.0085	.0033
524			2000	1.9124	1.5440	-2.6089	-.7982			.0090	.0035
525		IRT2	20	2.0669	1.8334	-2.7420	-.8820	-.2467	-.4423	.0160	.0063
526			50	2.1655	2.0155	-2.8235	-.9309			.0336	.0127
527			500	2.2938	2.2386	-2.9190	-.9825			.1010	.0330
528			2000	2.3136	2.2757	-2.9360	-.9926			.1032	.0347
529	3PLM	CTT	20	1.9947	1.8288	-2.6033	-.7945	.0699	.1356	.2748	.0919
530			50	1.9297	1.7223	-2.5532	-.7672			.2444	.0825
531			500	1.9270	1.7078	-2.5559	-.7673			.2275	.0780
532			2000	1.9248	1.6932	-2.5596	-.7689			.2090	.0722
533		IRT1	20	1.9194	1.5526	-2.6168	-.8033	.1882	.3332	.0029	.0012
534			50	1.8980	1.5135	-2.5979	-.7913			.0039	.0015
535			500	1.7525	1.2528	-2.4692	-.7111			.0092	.0035
536			2000	1.7312	1.2194	-2.4492	-.6988			.0134	.0051
537		IRT2	20	1.6344	1.0297	-2.3672	-.6462	.0152	.0266	.0057	.0022
538			50	1.6019	.9740	-2.3391	-.6300			.0051	.0019
539			500	1.6114	.9895	-2.3481	-.6361			.0028	.0011
540			2000	1.6192	1.0031	-2.3551	-.6404			.0024	.0009

ตอนที่ 5 ผลการวิเคราะห์ความสัมพันธ์และการถดถอย

ผลการวิเคราะห์ความสัมพันธ์ พบว่า ขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (d_{IRT1}) มีความสัมพันธ์กัน โดยสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากับ .626 และมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (r_{IRT1}) มีความสัมพันธ์กัน โดยสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากับ .570 และมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับความสัมพันธ์ระหว่างตัวประมาณค่าสองชนิด คือ ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จากทั้งสองทฤษฎีการทดสอบพบว่ามีความสัมพันธ์กันในทางบวก รายละเอียดดังตารางที่ 4.21

ตารางที่ 4.21 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความเข้มของอิทธิพลชนิดต่าง ๆ

	d_{CTT}	d_{IRT1}	r_{CTT}	r_{IRT1}
d_{CTT}	1.000			
d_{IRT1}	.626*	1.000		
r_{CTT}	.977*	.568*	1.000	
r_{IRT1}	.627*	1.000*	.570*	1.000

* $p < .01$

ผลการทดสอบความสัมพันธ์เชิงเส้นตรงระหว่าง d_{CTT} และ d_{IRT1} โดยการวิเคราะห์ความแปรปรวน พบว่า d_{CTT} และ d_{IRT1} มีความสัมพันธ์เชิงเส้นตรง สำหรับความสัมพันธ์ระหว่าง r_{CTT} และ r_{IRT1} พบว่า r_{CTT} และ r_{IRT1} มีความสัมพันธ์เชิงเส้นตรงเช่นกัน รายละเอียดดังตารางที่ 4.22

ตารางที่ 4.22 ผลการทดสอบความสัมพันธ์เชิงเส้นตรงของตัวประมาณค่าความเข้มของอิทธิพล

คู่ที่ทดสอบ	df_{REG}	df_{RES}	SS_{REG}	SS_{RES}	MS_{REG}	MS_{RES}	F	sig.
d_{IRT1} และ d_{CTT}	1	17998	43.499	67.571	43.499	.004	11586.327	.000
r_{IRT1} และ r_{CTT}	1	17998	13.950	29.018	13.950	.002	8652.556	.000

ผลการวิเคราะห์การถดถอยของ d_{IRT1} บน d_{CTT} และการถดถอยของ r_{IRT1} บน r_{CTT} พบว่า ความแปรปรวนของ d_{CTT} สามารถทำนายความแปรปรวนของ d_{IRT1} ได้ร้อยละ 39.2 ในขณะที่ความ

แปรปรวนของ r_{CTT} สามารถทำนายความแปรปรวนของ r_{IRT1} ได้ร้อยละ 32.5 โดยสมการถดถอยในรูปคะแนนดิบมีดังนี้

$d_{IRT1} = .004 + .065d_{CTT}$ และ $r_{IRT1} = .003 + .079r_{CTT}$

และเขียนสมการถดถอยในรูปคะแนนมาตรฐานได้ดังนี้

$Z_{d_{IRT1}} = .626Z_{d_{CTT}}$ และ $Z_{r_{IRT1}} = .570Z_{r_{CTT}}$

รายละเอียดของการวิเคราะห์การถดถอยมีดังตารางที่ 4.23

ตารางที่ 4.23 ผลการวิเคราะห์การถดถอยของ d_{IRT1} บน d_{CTT} และการถดถอยของ r_{IRT1} บน r_{CTT}

	b	SE	β	t	sig.	R^2
ค่าคงที่	.004	.000	-	7.618	.000	.392
d_{CTT}	.065	.001	.626	107.640	.000	
ค่าคงที่	.003	.000	-	8.994	.003	.325
r_{CTT}	.079	.001	.570	93.019	.079	

ตอนที่ 6 สรุปผลการวิเคราะห์ตามวัตถุประสงค์การวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์ 4 ประการ **ประการที่ 1** เพื่อเปรียบเทียบคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ ซึ่งแบ่งย่อยเป็นขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (d_{IRT1}) และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (d_{IRT2}) และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ ซึ่งแบ่งย่อยเป็นสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (r_{IRT1}) และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (r_{IRT2}) รวม 6 ชนิดประมาณค่า **ประการที่ 2** เพื่อเปรียบเทียบค่าเฉลี่ยของ d_{CTT}, d_{IRT1} และ d_{IRT2} และค่าเฉลี่ยของ r_{CTT}, r_{IRT1} และ r_{IRT2} **ประการที่ 3** เพื่อศึกษาความสัมพันธ์ระหว่าง d_{CTT} และ d_{IRT1} และสร้างสมการถดถอยของ d_{IRT1} บน d_{CTT} **ประการที่ 4** เพื่อศึกษาความสัมพันธ์ระหว่าง r_{CTT} และ r_{IRT1} และสร้างสมการถดถอยของ r_{IRT1} บน r_{CTT} ผลการวิเคราะห์ข้อมูลตามวัตถุประสงค์การวิจัยสรุปได้ดังนี้

สำหรับการเปรียบเทียบคุณสมบัติของตัวประมาณค่าระหว่าง d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} พบว่า r_{IRT1} มีความลำเอียงต่ำสุด รองลงมาคือ r_{IRT2}, d_{IRT1} , d_{IRT2}, d_{CTT} และ r_{CTT} ตามลำดับ โดยตัวประมาณค่าทั้งหมดให้ค่าประมาณต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) ในด้านความคงเส้นคงวา พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด รองลงมาคือ d_{CTT}, d_{IRT1}, r_{IRT1}, d_{IRT2} และ r_{IRT2} ตามลำดับ นอกจากนี้ยังพบว่า r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2}, d_{IRT1}, d_{IRT2}, r_{CTT} และ d_{CTT} ตามลำดับ

ผลการพิจารณาความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ของตัวประมาณค่าทั้ง 6 ชนิดไม่พบว่าตัวประมาณค่าชนิดใดมีคุณสมบัติทั้ง 3 ด้านดีที่สุด แต่พบว่า r_{IRT1} เป็นตัวประมาณค่าที่มีคุณสมบัติทั้ง 3 ด้านอยู่ในระดับที่น่าพอใจที่สุด เนื่องจาก r_{IRT1} เป็นตัวประมาณค่าที่มีความลำเอียงต่ำที่สุดและมีประสิทธิภาพสัมพัทธ์สูงสุด แม้ว่า r_{IRT1} จะมิได้เป็นตัวประมาณค่าที่มีความคงเส้นคงวาสูงสุดแต่เป็นตัวประมาณค่าที่มีความคงเส้นคงวาเป็นอันดับ 3 ในขณะที่ d_{CTT} และ r_{CTT} ซึ่งมีความคงเส้นคงวาสูงกว่า r_{IRT1} นั้นมีความลำเอียงสูงกว่า r_{IRT1} และประสิทธิภาพต่ำกว่า r_{IRT1} ร้อยละ 50 โดยประมาณ รายละเอียดดังตารางที่ 4.24

ตารางที่ 4.24 ผลการเปรียบเทียบคุณสมบัติของตัวประมาณค่า 6 ชนิดในภาพรวมทุกสถานการณ์

คุณสมบัติ	สถิติ	ตัวประมาณค่า					
		d_{CTT}	d_{IRT1}	d_{IRT2}	r_{CTT}	r_{IRT1}	r_{IRT2}
ความลำเอียง	SRMSD	1.084	.937	.975	1.291	.777	.849
		5	3	4	6	1	2
	BIAS	-.743	-1.036	-1.053	-.248	-.389	-.399
ความคงเส้นคงวา	Δ	.129	.001	-.018	.187	.009	-.038
		2	4	5	1	3	6
ประสิทธิภาพสัมพัทธ์	VAR	.567	.006	.012	.124	.002	.004
		6	3	4	5	1	2

ผลการเปรียบเทียบคุณสมบัติของตัวประมาณค่าระหว่างขนาดอิทธิพล (d) และสัมประสิทธิ์สหสัมพันธ์ (r) ในภาพรวมตามตัวแปรเงื่อนไขในการจำลองข้อมูล พบว่า แม้ในบางเงื่อนไขค่าส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จะมีค่าแตกต่างกัน แต่ยังคงมีค่าใกล้เคียงกัน ในบางเงื่อนไขจะพบว่าส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากัน ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์จึงมีความลำเอียงใกล้เคียงกันมาก อย่างไรก็ดีใน

เงื่อนไขส่วนใหญ่จะพบว่าสัมประสิทธิ์มีความคงเส้นคงวาและประสิทธิ์ภาพสัมพัทธ์สูงกว่าขนาดอิทธิพล ดังผลสรุปในตารางที่ 4.25

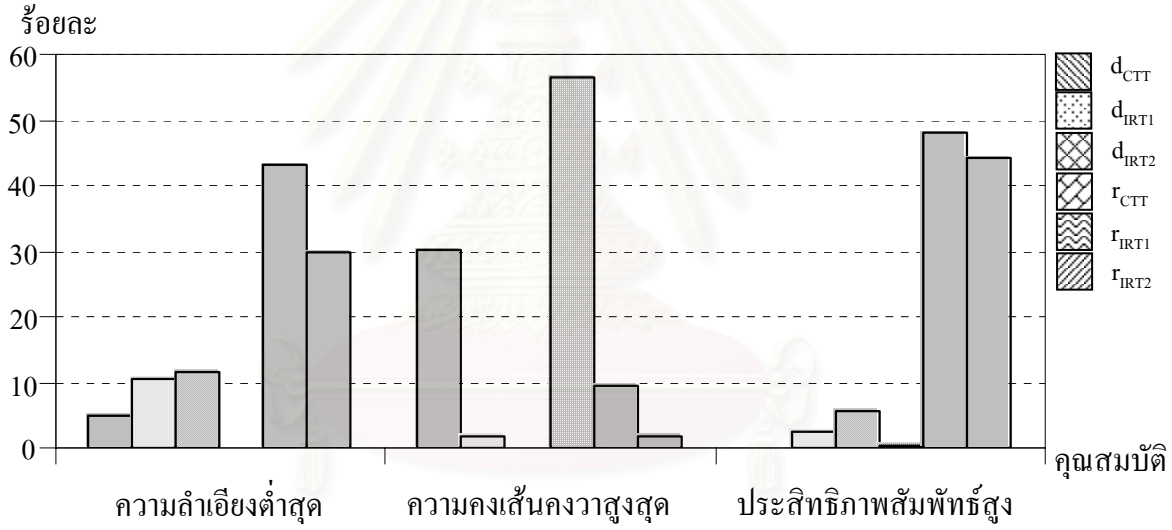
ตารางที่ 4.25 สรุปผลการเปรียบเทียบคุณสมบัติของขนาดอิทธิพล (d) และสัมประสิทธิ์สหสัมพันธ์ (r) ในภาพรวมของแต่ละเงื่อนไข

เงื่อนไข		ความลำเอียงต่ำ[1]	ความคงเส้นคงวาสูง	ประสิทธิภาพสัมพัทธ์สูง
ความเข้มของอิทธิพลที่แท้จริง	.2	$d \approx r$	r	r
	.5	$d \approx r$	r	r
	.8	$d \approx r$	r	r
	1.2	$d \approx r$	d	r
	2.6	$d \approx r$	d	r
ขนาดกลุ่มตัวอย่าง	20	$d \approx r$	ไม่สามารถคำนวณได้ เนื่องจากเป็นสถานการณ์ที่มีกลุ่มเพียงขนาดเดียว	r
	50	$d = r$		r
	500	$d \approx r$		r
	2000	$d \approx r$		r
ความยาวแบบสอบ	10	$d = r$	r	r
	50	$d \approx r$	r	r
	90	$d = r$	r	r
โมเดลฐาน	1PLM	$d \approx r$	d	r
	2PLM	$d \approx r$	r	r
	3PLM	$d \approx r$	r	r
โมเดลประมาณค่า	CTT	$d \approx r$	r	r
	IRT1	$d \approx r$	r	r
	IRT2	$d \approx r$	d	r

หมายเหตุ [1] $d \approx r$ หมายถึง ความลำเอียงของ d มีค่าใกล้เคียงกับความลำเอียงของ r

$d = r$ หมายถึง ความลำเอียงของ d มีค่าเท่ากับความลำเอียงของ r

สำหรับการเปรียบเทียบคุณสมบัติของตัวประมาณค่าของ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ในสถานการณ์ย่อย พบว่ามีความแตกต่างกันไปตามเงื่อนไขของแต่ละสถานการณ์ ในเบื้องต้นผู้วิจัยคำนวณร้อยละของสถานการณ์ย่อยที่พบคุณสมบัติของตัวประมาณค่าที่ดีในแต่ละด้านของตัวประมาณค่าแต่ละตัว จาก 180 สถานการณ์ย่อยพบว่า ร้อยละ 43 เป็นสถานการณ์ที่ r_{IRT1} มีความลำเอียงต่ำกว่าตัวประมาณค่าชนิดอื่น รองลงมาคือ r_{IRT2} และ d_{IRT2} คิดเป็นร้อยละ 30 และ 12 ตามลำดับ ในด้านความคงเส้นคงวาพบว่า ร้อยละ 57 เป็นสถานการณ์ที่ r_{CTT} มีความคงเส้นคงวาสูงสุด รองลงมาคือ d_{CTT} และ r_{IRT1} คิดเป็นร้อยละ 30 และ 9 ตามลำดับ สำหรับประสิทธิ์ภาพสัมพัทธ์ พบว่า ร้อยละ 48 เป็นสถานการณ์ที่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2} และ d_{IRT2} คิดเป็นร้อยละ 44 และ 6 ตามลำดับ รายละเอียดดังภาพที่ 4.1 สำหรับรายละเอียดของตัวประมาณค่าที่มีคุณสมบัติของตัวประมาณค่าที่ดีในแต่ละสถานการณ์ย่อยมีดังตารางที่ 4.26



ภาพที่ 4.1 ร้อยละของสถานการณ์ที่พบคุณสมบัติของตัวประมาณค่าที่ดีในตัวประมาณค่า 6 ชนิด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.26 ผลการเปรียบเทียบคุณสมบัติของ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1}, r_{IRT2} ในสถานการณ์ย่อย

สถานการณ์					ตัวประมาณค่าที่มีคุณสมบัติที่ดี		
ที่	TRUES	NITEM	MBASE	NSAMP	ลำเอียงต่ำสุด	คงเส้นคงวาสูงสุด	ประสิทธิภาพสัมพัทธ์สูงสุด
1	.2	10	1PLM	20	r_{IRT2}	r_{CTT}	r_{IRT1}
2				50	r_{IRT2}		r_{IRT1}
3				500	r_{IRT1}		d_{IRT1}, d_{IRT2}, r_{IRT1}, r_{IRT2}
4				2000	r_{IRT1}		d_{IRT1}, d_{IRT2}, r_{IRT1}, r_{IRT2}
5			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
6				50	r_{IRT1}		r_{IRT1}
7				500	r_{IRT2}		r_{IRT1}
8				2000	r_{IRT1}		d_{IRT1}, r_{IRT1}
9			3PLM	20	r_{IRT2}	r_{CTT}	r_{IRT1}, r_{IRT2}
10				50	r_{IRT1}		r_{IRT2}
11				500	r_{IRT2}		r_{IRT2}
12				2000	r_{IRT1}		d_{IRT2}, r_{IRT2}
13		50	1PLM	20	r_{IRT2}	r_{CTT}	r_{IRT1}, r_{IRT2}
14				50	r_{IRT1}		r_{IRT2}
15				500	r_{IRT1}		d_{IRT2}, r_{IRT1}, r_{IRT2}
16				2000	r_{IRT1}		d_{IRT1}, d_{IRT2}, r_{IRT1}, r_{IRT2}
17			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
18				50	r_{IRT1}		r_{IRT1}
19				500	r_{IRT2}		r_{IRT1}
20				2000	r_{IRT1}		r_{IRT1}
21			3PLM	20	r_{IRT2}	r_{CTT}	r_{IRT1}
22				50	r_{IRT2}		r_{IRT2}
23				500	r_{IRT2}		r_{IRT2}
24				2000	r_{IRT2}		d_{IRT2}, r_{IRT2}
25		90	1PLM	20	r_{IRT2}	r_{CTT}	r_{IRT2}
26				50	r_{IRT1}		r_{IRT1}, r_{IRT2}
27				500	r_{IRT1}		d_{IRT2}, r_{IRT1}, r_{IRT2}
28				2000	r_{IRT1}		d_{IRT1}, d_{IRT2}, r_{IRT1}, r_{IRT2}
29			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
30				50	r_{IRT2}		r_{IRT1}
31				500	r_{IRT2}		r_{IRT1}
32				2000	r_{IRT2}		r_{IRT1}
33			3PLM	20	r_{IRT2}	r_{CTT}	r_{IRT2}
34				50	r_{IRT1}		r_{IRT2}
35				500	r_{IRT2}		r_{IRT2}
36				2000	r_{IRT1}		d_{IRT2}, r_{IRT2}

หมายเหตุ TRUES แทน ความเข้มของอิทธิพลที่แท้จริง NITEM แทน ความยาวแบบสอบ MBASE แทน โมเดลฐาน NSAMP แทน ขนาดกลุ่มตัวอย่าง 1PLM, 2PLM และ 3PLM แทนโมเดลโลจิสติกแบบหนึ่ง, สอง และสามพารามิเตอร์ ตามลำดับ

ตารางที่ 4.26 (ต่อ) ผลการเปรียบเทียบคุณสมบัติของ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1}, r_{IRT2} ในสถานการณ์ย่อย

สถานการณ์					ตัวประมาณค่าที่มีคุณสมบัติที่ดี		
ที่	TRUES	NITEM	MBASE	NSAMP	ลำเอียงต่ำสุด	คงเส้นคงวาสูงสุด	ประสิทธิภาพสัมพัทธ์สูงสุด
37	.5	10	1PLM	20	r_{IRT2}	r_{IRT1}	r_{IRT2}
38				50	r_{IRT1}		r_{IRT1}
39				500	r_{IRT1}		r_{IRT1}, r_{IRT2}
40				2000	r_{IRT1}		r_{IRT1}, r_{IRT2}
41			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
42				50	r_{IRT1}		r_{IRT1}
43				500	r_{IRT1}		r_{IRT1}
44				2000	r_{IRT1}		r_{IRT1}
45			3PLM	20	r_{IRT1}	r_{CTT}	r_{IRT2}
46				50	r_{IRT2}		r_{IRT2}
47				500	r_{IRT2}		r_{IRT2}
48				2000	r_{IRT2}		r_{IRT2}
49		50	1PLM	20	r_{IRT2}	r_{CTT}	r_{IRT2}
50				50	r_{IRT1}		r_{IRT2}
51				500	r_{IRT1}		d_{IRT2}, r_{IRT1}, r_{IRT2}
52				2000	r_{IRT1}		d_{IRT1}, d_{IRT2}, r_{IRT1}, r_{IRT2}
53			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
54				50	r_{IRT1}		r_{IRT1}
55				500	r_{IRT1}		r_{IRT1}
56				2000	r_{IRT2}		r_{IRT1}
57			3PLM	20	r_{IRT2}	r_{CTT}	r_{IRT2}
58				50	r_{IRT1}		r_{IRT2}
59				500	r_{IRT1}		r_{IRT2}
60				2000	r_{IRT1}		r_{IRT2}
61		90	1PLM	20	r_{IRT2}	r_{CTT}	r_{IRT2}
62				50	r_{IRT1}		r_{IRT2}
63				500	r_{IRT1}		d_{IRT2}, r_{IRT1}, r_{IRT2}
64				2000	r_{IRT1}		d_{IRT1}, d_{IRT2}, r_{IRT1}, r_{IRT2}
65			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
66				50	r_{IRT2}		r_{IRT1}
67				500	r_{IRT2}		r_{IRT1}
68				2000	r_{IRT2}		r_{IRT1}
69			3PLM	20	r_{IRT1}	r_{CTT}	r_{IRT2}
70				50	r_{IRT1}		r_{IRT2}
71				500	r_{IRT2}		r_{IRT2}
72				2000	r_{IRT2}		r_{IRT2}

หมายเหตุ TRUES แทน ความเข้มของอิทธิพลที่แท้จริง NITEM แทน ความยาวแบบสอบ MBASE แทน โมเดลฐาน NSAMP แทน ขนาดกลุ่มตัวอย่าง 1PLM, 2PLM และ 3PLM แทนโมเดลโลจิสติกแบบหนึ่ง, สอง และสามพารามิเตอร์ ตามลำดับ

ตารางที่ 4.26 (ต่อ) ผลการเปรียบเทียบคุณสมบัติของ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1}, r_{IRT2} ในสถานการณ์ย่อย

สถานการณ์					ตัวประมาณค่าที่มีคุณสมบัติที่ดี		
ที่	TRUES	NITEM	MBASE	NSAMP	ลำเอียงต่ำสุด	คงเส้นคงวาสูงสุด	ประสิทธิภาพสัมพัทธ์สูงสุด
73	.8	10	1PLM	20	r_{IRT1}	r_{IRT2}	r_{IRT1}
74				50	r_{IRT2}		r_{IRT1}
75				500	r_{IRT2}		r_{IRT1}
76				2000	r_{IRT2}		r_{IRT1}, r_{IRT2}
77			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
78				50	r_{IRT1}		r_{IRT1}
79				500	r_{IRT1}		r_{IRT1}
80				2000	r_{IRT1}		r_{IRT1}
81			3PLM	20	r_{IRT2}	r_{CTT}	r_{IRT1}, r_{IRT2}
82				50	r_{IRT2}		r_{IRT2}
83				500	r_{IRT2}		r_{IRT2}
84				2000	r_{IRT2}		r_{IRT2}
85		50	1PLM	20	r_{IRT1}	r_{IRT1}	r_{IRT1}
86				50	r_{IRT1}		r_{IRT1}
87				500	r_{IRT1}		r_{IRT1}
88				2000	r_{IRT1}		r_{IRT1}
89			2PLM	20	r_{IRT2}	r_{CTT}	r_{IRT1}, r_{IRT2}
90				50	r_{IRT2}		r_{IRT2}
91				500	r_{IRT2}		r_{IRT2}
92				2000	r_{IRT2}		r_{IRT2}
93			3PLM	20	r_{IRT2}	r_{CTT}	r_{IRT2}
94				50	r_{IRT2}		r_{IRT2}
95				500	r_{IRT2}		r_{IRT2}
96				2000	r_{IRT2}		d_{IRT2}, r_{IRT1}, r_{IRT2}
97		90	1PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
98				50	r_{IRT1}		r_{IRT1}
99				500	r_{IRT1}		r_{IRT1}
100				2000	r_{IRT1}		r_{IRT1}
101			2PLM	20	r_{IRT2}	r_{CTT}	r_{IRT1}
102				50	r_{IRT2}		r_{IRT2}
103				500	r_{IRT2}		r_{IRT2}
104				2000	r_{IRT2}		r_{IRT2}
105			3PLM	20	r_{IRT2}	r_{CTT}	r_{IRT1}, r_{IRT2}
106				50	r_{IRT2}		r_{IRT2}
107				500	r_{IRT2}		r_{IRT2}
108				2000	r_{IRT2}		r_{IRT2}

หมายเหตุ TRUES แทน ความเข้มของอิทธิพลที่แท้จริง NITEM แทน ความยาวแบบสอบ MBASE แทน โมเดลฐาน NSAMP แทน ขนาดกลุ่มตัวอย่าง 1PLM, 2PLM และ 3PLM แทนโมเดลโลจิสติกแบบหนึ่ง, สอง และสามพารามิเตอร์ ตามลำดับ

ตารางที่ 4.26 (ต่อ) ผลการเปรียบเทียบคุณสมบัติของ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1}, r_{IRT2} ในสถานการณ์ย่อย

สถานการณ์					ตัวประมาณค่าที่มีคุณสมบัติที่ดี		
ที่	TRUES	NITEM	MBASE	NSAMP	ลำเอียงต่ำสุด	คงเส้นคงวาสูงสุด	ประสิทธิภาพสัมพัทธ์สูงสุด
109	1.2	10	1PLM	20	d_{IRT1}	d_{CTT}	r_{IRT1}
110				50	d_{IRT2}		r_{IRT1}
111				500	d_{IRT2}		r_{IRT1}, r_{IRT2}
112				2000	d_{IRT2}		r_{IRT1}, r_{IRT2}
113			2PLM	20	d_{IRT1}	r_{CTT}	r_{IRT1}
114				50	d_{IRT1}		r_{IRT1}
115				500	d_{IRT1}		r_{IRT1}
116				2000	d_{IRT1}		r_{IRT1}
117			3PLM	20	d_{IRT2}	r_{CTT}	r_{IRT2}
118				50	d_{IRT2}		r_{IRT2}
119				500	d_{IRT2}		r_{IRT2}
120				2000	d_{IRT2}		r_{IRT2}
121		50	1PLM	20	d_{IRT1}	d_{CTT}	r_{IRT2}
122				50	d_{IRT1}		r_{IRT2}
123				500	d_{IRT2}		r_{IRT2}
124				2000	d_{IRT2}		r_{IRT2}
125			2PLM	20	d_{IRT1}	d_{CTT}	r_{IRT1}
126				50	d_{IRT1}		r_{IRT1}
127				500	d_{IRT1}		r_{IRT1}
128				2000	d_{IRT1}		r_{IRT1}
129			3PLM	20	d_{IRT2}	r_{CTT}	r_{IRT1}
130				50	d_{IRT2}		r_{IRT2}
131				500	d_{IRT2}		r_{IRT2}
132				2000	d_{IRT2}		r_{IRT2}
133		90	1PLM	20	d_{IRT1}	d_{CTT}	r_{IRT2}
134				50	d_{IRT1}		r_{IRT2}
135				500	d_{IRT2}		r_{IRT2}
136				2000	d_{IRT2}		r_{IRT2}
137			2PLM	20	d_{IRT1}	r_{CTT}	r_{IRT1}
138				50	d_{IRT1}		r_{IRT1}
139				500	d_{IRT1}		r_{IRT1}
140				2000	d_{IRT1}		r_{IRT1}
141			3PLM	20	d_{IRT2}	r_{CTT}	r_{IRT1}
142				50	d_{IRT2}		r_{IRT2}
143				500	d_{IRT2}		r_{IRT2}
144				2000	d_{IRT2}		r_{IRT2}

หมายเหตุ TRUES แทน ความเข้มของอิทธิพลที่แท้จริง NITEM แทน ความยาวแบบสอบ MBASE แทน โมเดลฐาน NSAMP แทน ขนาดกลุ่มตัวอย่าง 1PLM, 2PLM และ 3PLM แทนโมเดลโลจิสติกแบบหนึ่ง, สอง และสามพารามิเตอร์ ตามลำดับ

ตารางที่ 4.26 (ต่อ) ผลการเปรียบเทียบคุณสมบัติของ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1}, r_{IRT2} ในสถานการณ์ย่อย

สถานการณ์					ตัวประมาณค่าที่มีคุณสมบัติที่ดี		
ที่	TRUES	NITEM	MBASE	NSAMP	ลำเอียงต่ำสุด	คงเส้นคงวาสูงสุด	ประสิทธิภาพสัมพัทธ์สูงสุด
145	2.6	10	1PLM	20	d_{CTT}	d_{CTT}	r_{IRT1}
146				50	d_{CTT}		r_{IRT2}
147				500	d_{CTT}		r_{IRT1}
148				2000	d_{CTT}		r_{IRT1}
149			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
150				50	r_{IRT1}		r_{IRT1}
151				500	r_{IRT1}		r_{IRT1}
152				2000	r_{IRT1}		r_{IRT1}
153			3PLM	20	r_{IRT2}	r_{IRT1}	r_{IRT2}
154				50	r_{IRT2}		r_{IRT2}
155				500	r_{IRT2}		r_{IRT2}
156				2000	r_{IRT2}		r_{IRT2}
157		50	1PLM	20	r_{IRT1}	d_{CTT}	r_{IRT1}
158				50	r_{IRT1}		r_{IRT1}
159				500	d_{CTT}		r_{IRT2}
160				2000	d_{CTT}		r_{CTT}
161			2PLM	20	r_{IRT1}	d_{IRT1}	r_{IRT1}
162				50	r_{IRT1}		r_{IRT1}
163				500	r_{IRT1}		r_{IRT1}
164				2000	r_{IRT1}		r_{IRT1}
165			3PLM	20	r_{IRT2}	r_{IRT1}	r_{IRT1}
166				50	r_{IRT2}		r_{IRT1}
167				500	r_{IRT2}		r_{IRT2}
168				2000	r_{IRT2}		r_{IRT2}
169		90	1PLM	20	r_{IRT1}	d_{CTT}	r_{IRT2}
170				50	r_{IRT1}		r_{IRT2}
171				500	d_{CTT}		r_{IRT2}
172				2000	d_{CTT}		r_{IRT1}
173			2PLM	20	r_{IRT1}	r_{CTT}	r_{IRT1}
174				50	r_{IRT1}		r_{IRT1}
175				500	r_{IRT1}		r_{IRT1}
176				2000	r_{IRT1}		r_{IRT1}
177			3PLM	20	r_{IRT2}	r_{IRT1}	r_{IRT1}
178				50	r_{IRT2}		r_{IRT1}
179				500	r_{IRT2}		r_{IRT2}
180				2000	r_{IRT2}		r_{IRT2}

หมายเหตุ TRUES แทน ความเข้มของอิทธิพลที่แท้จริง NITEM แทน ความยาวแบบสอบ MBASE แทน โมเดลฐาน NSAMP แทน ขนาดกลุ่มตัวอย่าง 1PLM, 2PLM และ 3PLM แทนโมเดลโลจิสติกแบบหนึ่ง, สอง และสามพารามิเตอร์ ตามลำดับ

สำหรับการเปรียบเทียบค่าเฉลี่ยระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่ามีความสอดคล้องกับข้อมูล (d_{IRT1}) และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (d_{IRT2}) พบว่า ค่าเฉลี่ยของ d_{CTT}, d_{IRT1} และ d_{IRT2} มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดย d_{CTT} มีค่าเฉลี่ยสูงสุด รองลงมาคือ d_{IRT1} และ d_{IRT2} ตามลำดับ ในทำนองเดียวกันพบว่า ค่าเฉลี่ยของสัมประสิทธิ์สัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) สัมประสิทธิ์สัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่ามีความสอดคล้องกับข้อมูล (r_{IRT1}) และสัมประสิทธิ์สัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (r_{IRT2}) มีความแตกต่างกันที่ระดับนัยสำคัญทางสถิติ .05 โดย r_{CTT} มีค่าเฉลี่ยสูงสุด รองลงมาคือ r_{IRT1} และ r_{IRT2} ตามลำดับ รายละเอียดดังตารางที่ 4.27

ตารางที่ 4.27 ผลการเปรียบเทียบค่าเฉลี่ยของตัวประมาณค่าระหว่างทฤษฎีการทดสอบ 2 ทฤษฎี

ตัวแปรต้น	df_{BG}	df_{WG}	Levene.	sig.	SS_{BG}	SS_{WG}	MS_{BG}	MS_{WG}	F	sig.
MUSED	2	53997	14087.510	.000	1093.103	10532.708	546.552	.195	2801.953	.000
			18976.711	.000	257.396	2356.602	128.698	.044	2948.870	.000

หมายเหตุ ค่าสถิติเหนือเส้นประเป็นของขนาดอิทธิพล (d) ค่าสถิติใต้เส้นประเป็นของสัมประสิทธิ์สหสัมพันธ์ (r)

สำหรับการวิเคราะห์ความสัมพันธ์ระหว่าง d_{CTT} และ d_{IRT1} พบว่า เป็นความสัมพันธ์เชิงเส้นทางบวก โดยสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากับ .626 ขณะที่ความสัมพันธ์ระหว่าง r_{CTT} และ r_{IRT1} เป็นความสัมพันธ์เชิงเส้นทางบวกเช่นกันแต่มีค่าสัมประสิทธิ์สหสัมพันธ์ .570 เมื่อทำการวิเคราะห์การถดถอยของ d_{IRT1} บน d_{CTT} และการถดถอยของ r_{IRT1} บน r_{CTT} พบว่า ความแปรปรวนของ d_{CTT} สามารถทำนายความแปรปรวนของ d_{IRT1} ได้ร้อยละ 39.2 ในขณะที่ความแปรปรวนของ r_{CTT} สามารถทำนายความแปรปรวนของ r_{IRT1} ได้ร้อยละ 32.5 โดยสมการถดถอยในรูปคะแนนดิบคือ $d_{IRT1} = .004 + .065d_{CTT}$ และ $r_{IRT1} = .003 + .079r_{CTT}$ สำหรับสมการในรูปคะแนนมาตรฐานคือ $Z_{d_{IRT1}} = .626Z_{d_{CTT}}$ และ $Z_{r_{IRT1}} = .570Z_{r_{CTT}}$ ตามลำดับ รายละเอียดดังตารางที่ 4.28

ตารางที่ 4.28 สรุปผลการวิเคราะห์ความสัมพันธ์และการถดถอย

ตัวแปรเกณฑ์		b	SE	β	t	sig.	r	R^2
d_{IRT1}	ค่าคงที่	.004	.000	-	7.618	.000	.626	.392
	d_{CTT}	.065	.001	.626	107.640	.000		
r_{IRT1}	ค่าคงที่	.003	.000	-	8.994	.003	.570	.325
	r_{CTT}	.079	.001	.570	93.019	.079		

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง มีวัตถุประสงค์ 4 ประการ **ประการที่ 1** เพื่อเปรียบเทียบคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ ซึ่งแบ่งย่อยเป็นขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (d_{IRT1}) และขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (d_{IRT2}) และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ ซึ่งแบ่งย่อยเป็นสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (r_{IRT1}) และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (r_{IRT2}) รวม 6 ชนิด **ประการที่ 2** เพื่อเปรียบเทียบค่าเฉลี่ยระหว่าง d_{CTT}, d_{IRT1} และ d_{IRT2} และระหว่าง r_{CTT}, r_{IRT1} และ r_{IRT2} **ประการที่ 3** เพื่อศึกษาความสัมพันธ์ระหว่าง d_{CTT} และ d_{IRT1} และสร้างสมการถดถอยของ d_{IRT1} บน d_{CTT} **ประการที่ 4** เพื่อศึกษาความสัมพันธ์ระหว่าง r_{CTT} และ r_{IRT1} และสร้างสมการถดถอยของ r_{IRT1} บน r_{CTT}

การจำลองข้อมูลใช้ 5 เงื่อนไขหลัก คือ ค่าความเข้มของอิทธิพลที่แท้จริง (TRUES) ขนาดกลุ่มตัวอย่าง (NSAMP) ความยาวแบบสอบ (NITEM) โมเดลฐาน (MBASE) และโมเดลประมาณค่า (MUSED) รวม 540 สถานการณ์ ใช้การจำลองข้อมูล 4 ขั้นตอน คือ (1) สร้างค่าความสามารถที่แท้จริงของผู้สอบ (2) สร้างค่าพารามิเตอร์ของข้อสอบ (3) คำนวณค่าความน่าจะเป็นในการตอบของผู้สอบ และ (4) สร้างข้อมูลคำตอบของผู้สอบ จากนั้นประมาณค่าความสามารถของผู้สอบตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบด้วยโปรแกรม BILOG แล้วคำนวณค่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เพื่อวิเคราะห์ข้อมูลตามวัตถุประสงค์การวิจัยต่อไป

การวิเคราะห์ข้อมูลจะใช้ส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ย ค่าความลำเอียง ผลต่างของส่วนเบี่ยงเบนของรากกำลังสอง และความแปรปรวน เพื่อตัดสินคุณสมบัติตัวประมาณค่าที่ดี และการวิเคราะห์ความสัมพันธ์ระหว่างค่าประมาณความเข้มของอิทธิพลที่ได้จากสองทฤษฎี เพื่อสร้างสมการการถดถอยของค่าประมาณความเข้มของอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ โดยใช้ค่าประมาณความเข้มของอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมเป็นตัวแปรทำนาย

ในบทนี้ผู้วิจัยนำเสนอสรุปและอภิปรายผลการวิจัย รวมทั้งข้อเสนอแนะ ดังต่อไปนี้

สรุปผลการวิจัย

1. สำหรับการเปรียบเทียบคุณสมบัติของตัวประมาณค่าในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (d_{CTT}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (r_{CTT}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (d_{IRT1}) ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (d_{IRT2}) สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูล (r_{IRT1}) และ สัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูล (r_{IRT2}) ในภาพรวม พบว่า r_{IRT1} มีความลำเอียงต่ำสุด รองลงมาคือ r_{IRT2}, d_{IRT1}, d_{IRT2}, d_{CTT} และ r_{CTT} ตามลำดับ โดยตัวประมาณค่าทั้งหมดให้ค่าประมาณต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง (underestimate) ในด้านความคงเส้นคงวา พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด รองลงมาคือ d_{CTT}, d_{IRT1}, r_{IRT1}, d_{IRT2} และ r_{IRT2} ตามลำดับ นอกจากนี้ยังพบว่า r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2}, d_{IRT1}, d_{IRT2}, r_{CTT} และ d_{CTT} ตามลำดับ

2. สำหรับการเปรียบเทียบคุณสมบัติของตัวประมาณค่าของ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} ในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์สำหรับสถานการณ์ย่อย พบว่ามีความแตกต่างกันไปตามเงื่อนไขของแต่ละสถานการณ์ ในเบื้องต้นผู้วิจัยคำนวณร้อยละของสถานการณ์ย่อยที่พบคุณสมบัติของตัวประมาณค่าที่ดีในแต่ละด้านของตัวประมาณค่าแต่ละตัว จาก 180 สถานการณ์ย่อยพบว่า ร้อยละ 43 เป็นสถานการณ์ที่ r_{IRT1} มีความลำเอียงต่ำกว่าตัวประมาณค่าชนิดอื่น รองลงมาคือ r_{IRT2} และ d_{IRT2} คิดเป็นร้อยละ 30 และ 12 ตามลำดับ ในด้านความคงเส้นคงวา พบว่า ร้อยละ 57 เป็นสถานการณ์ที่ r_{CTT} มีความคงเส้นคงวาสูงสุด รองลงมาคือ d_{CTT} และ r_{IRT1} คิดเป็นร้อยละ 30 และ 9 ตามลำดับ สำหรับประสิทธิ์ภาพสัมพัทธ์ พบว่า ร้อยละ 48 เป็นสถานการณ์ที่ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด รองลงมาคือ r_{IRT2} และ d_{IRT2} คิดเป็นร้อยละ 44 และ 6 ตามลำดับ

3. ค่าเฉลี่ยของ d_{CTT}, d_{IRT1} และ d_{IRT2} มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดย d_{CTT} มีค่าเฉลี่ยสูงสุด รองลงมาคือ d_{IRT1} และ d_{IRT2} ตามลำดับ ในทำนองเดียวกันพบว่าค่าเฉลี่ยของ r_{CTT}, r_{IRT1} และ r_{IRT2} มีความแตกต่างกันที่ระดับนัยสำคัญทางสถิติ .05 โดย r_{CTT} มีค่าเฉลี่ยสูงสุด รองลงมาคือ r_{IRT1} และ r_{IRT2} ตามลำดับ

4. สำหรับความสัมพันธ์ระหว่าง d_{CTT} และ d_{IRT1} พบว่า เป็นความสัมพันธ์เชิงเส้นทางบวก มีค่าสัมประสิทธิ์สหสัมพันธ์ .626 ขณะที่ r_{CTT} และ r_{IRT1} มีความสัมพันธ์เชิงเส้นทางบวกและมีค่าสัมประสิทธิ์สหสัมพันธ์ .570 โดยค่าสัมประสิทธิ์ทั้งสองมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อวิเคราะห์การถดถอยของ d_{IRT1} บน d_{CTT} และการถดถอยของ r_{IRT1} บน r_{CTT} พบว่า ความแปรปรวนของ d_{CTT} ทำนายความแปรปรวนของ d_{IRT1} ได้ร้อยละ 39.2 ในขณะที่ความแปรปรวนของ r_{CTT}

ทำนายความแปรปรวนของ r_{IRT1} ได้ร้อยละ 32.5 โดยสมการถดถอยในรูปคะแนนดิบ คือ $d_{IRT1} = .004 + .065d_{CTT}$ และ $r_{IRT1} = .003 + .079r_{CTT}$ สำหรับสมการในรูปคะแนนมาตรฐานคือ $Z_{d_{IRT1}} = .626Z_{d_{CTT}}$ และ $Z_{r_{IRT1}} = .570Z_{r_{CTT}}$ ตามลำดับ

อภิปรายผล

1. ผลการเปรียบเทียบคุณสมบัติของตัวประมาณค่าเป็นไปตามสมมติฐานที่ผู้วิจัยตั้งเอาไว้ ว่าตัวประมาณค่าทั้ง 6 ชนิด คือ d_{CTT}, d_{IRT1}, d_{IRT2}, r_{CTT}, r_{IRT1} และ r_{IRT2} มีคุณสมบัติในด้านความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์แตกต่างกัน โดยตัวประมาณค่าทุกตัวมีความลำเอียงและให้ค่าประมาณความเข้มของอิทธิพลที่ต่ำกว่าค่าความเข้มของอิทธิพลที่แท้จริง ซึ่งผลการวิจัยในส่วนนี้สอดคล้องกับผลการวิจัยของ Roberts และ Henson (2002) ที่พบว่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ล้วนเป็นตัวประมาณค่าที่มีความลำเอียง และช่วยขยายข้อค้นพบของ DeMars (2001) ที่พบว่าขนาดอิทธิพลที่ได้จากโมเดลการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องและไม่สอดคล้องกับข้อมูลมีความลำเอียงที่แตกต่างกันด้วย โดยพบว่าโมเดลการทดสอบแบบดั้งเดิม โมเดลการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูลและโมเดลการตอบสนองข้อสอบที่โมเดลประมาณค่าสอดคล้องกับข้อมูลจะให้ค่าขนาดอิทธิพล และสัมประสิทธิ์สหสัมพันธ์ที่มีความลำเอียงต่างกัน สำหรับคุณสมบัติในด้านความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ พบว่า r_{CTT} มีความคงเส้นคงวาสูงสุด และ r_{IRT1} มีประสิทธิภาพสัมพัทธ์สูงสุด นอกจากนี้ยังพบว่าสัมประสิทธิ์สหสัมพันธ์มีความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์สูงกว่าขนาดอิทธิพล เมื่อใช้โมเดลการทดสอบแบบดั้งเดิมหรือโมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูลในการประมาณค่าความสามารถของกลุ่มตัวอย่าง ผลการวิจัยในส่วนนี้ช่วยสนับสนุนแนวคิดของ Rosenthal และ DiMatteo (2001), Baugh (2002), Thompson (2002) และ Ives (2003) ที่เสนอว่าการประมาณค่าความเข้มของอิทธิพลในงานวิจัยเชิงปริมาณทั่วไปควรใช้สถิติในกลุ่มความสัมพันธ์

2. ผลการวิจัยในบางสถานการณ์ พบว่า ค่าผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (Δ) มีค่าเป็นลบ ซึ่งค่าดังกล่าวสะท้อนว่าตัวประมาณค่ามีความคงเส้นคงวาต่ำและไม่น่าพึงพอใจเนื่องจากส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) เมื่อใช้กลุ่มตัวอย่างมี 2000 คน มีค่าสูงกว่าสถานการณ์ที่ใช้กลุ่มตัวอย่างขนาด 20 คน กล่าวคือ กลุ่มตัวอย่างขนาดใหญ่ให้ค่าประมาณที่มีความลำเอียงสูงกว่ากลุ่มตัวอย่างขนาดเล็ก ทั้งนี้เป็นเพราะโมเดลประมาณค่าที่ใช้ในสถานการณ์ดังกล่าวเป็นโมเดลการตอบสนองข้อสอบที่ไม่สอดคล้องกับข้อมูล ทำให้การประมาณค่าความสามารถของผู้สอบมีความคลาดเคลื่อนเกิดขึ้น (ศิริชัย กาญจนวาสี,

2545; Hambleton, Swaminathan & Rogers, 1991; Embretson & Reise, 2000) และเมื่อนำค่าประมาณความสามารถดังกล่าวมาประมาณค่าความเข้มของอิทธิพลจึงเกิดความคลาดเคลื่อนขึ้น และความคลาดเคลื่อนนี้มิได้ลดลงเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ทำให้ความคงเส้นคงวาของตัวประมาณค่าต่ำกว่าที่ควรจะเป็น

3. แม้ว่าผลการพิจารณาความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ของตัวประมาณค่าทั้ง 6 ชนิดจะไม่พบว่าตัวประมาณค่าชนิดใดมีคุณสมบัติทั้ง 3 ด้านดีที่สุด แต่พบว่า r_{IRT1} เป็นตัวประมาณค่าที่มีคุณสมบัติทั้ง 3 ด้านอยู่ในระดับที่น่าพอใจที่สุด เนื่องจาก r_{IRT1} เป็นตัวประมาณค่าที่มีความลำเอียงต่ำที่สุดและมีประสิทธิภาพสัมพัทธ์สูงสุด แม้ว่า r_{IRT1} จะมิได้เป็นตัวประมาณค่าที่มีความคงเส้นคงวาสูงสุดแต่เป็นตัวประมาณค่าที่มีความคงเส้นคงวาเป็นอันดับ 3 ในขณะที่ d_{CTT} และ r_{CTT} ซึ่งมีความคงเส้นคงวาสูงกว่า r_{IRT1} นั้นมีความลำเอียงสูงกว่า r_{IRT1} และประสิทธิภาพต่ำกว่า r_{IRT1} ร้อยละ 50 โดยประมาณ อนึ่ง แม้ว่าการคำนวณ r_{IRT1} จำเป็นต้องคำนวณค่าความสามารถของกลุ่มตัวอย่างโดยใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูล ซึ่งมีแนวคิดที่ซับซ้อนกว่าการประมาณค่าความสามารถของผู้สอบโดยใช้โมเดลการทดสอบแบบดั้งเดิม แต่ปัจจุบันได้มีการพัฒนาโปรแกรมในการประมาณค่าความสามารถของกลุ่มตัวอย่างโดยใช้โมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูล อาทิ BILOG ซึ่งทำให้การคำนวณสะดวกมากขึ้น

4. ในการจำลองสถานการณ์ตามเงื่อนไข 5 เงื่อนไข ผู้วิจัยพบว่า แนวโน้มของค่าสถิติที่บ่งชี้ความลำเอียง (SRMSD) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในภาพรวมของทุกเงื่อนไขเป็นไปในทางเดียวกัน คือขนาดอิทธิพลมีความลำเอียงมีค่าใกล้เคียงกับความลำเอียงของสัมประสิทธิ์สหสัมพันธ์ ทั้งนี้สอดคล้องกับแนวคิดของ Hedges และ Olkin (1985), Cohen (1988 อ้างถึงใน Thompson, 2002), Friedman (1968 อ้างถึงใน Thompson, 2002) และ Hunter และ Schmidt (1990) ที่แสดงให้เห็นว่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์มีความเกี่ยวข้องสัมพันธ์กันและสามารถแปลงรูประหว่างกันและกันได้ โดยผลการวิจัยในครั้งนี้นอกจากจะสนับสนุนแนวคิดดังกล่าวยังได้ขยายแนวคิดออกไปอีกด้วย เพราะผลการวิจัยชี้ให้เห็นว่าแม้จะใช้โมเดลการทดสอบที่แตกต่างกัน ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ที่ได้ยังคงมีความสัมพันธ์กันในทางบวกค่อนข้างสูง

5. ในการจำลองสถานการณ์ตามเงื่อนไข 5 เงื่อนไข ผู้วิจัยพบว่า แนวโน้มของค่าสถิติที่บ่งชี้ความคงเส้นคงวา (Δ) และประสิทธิภาพสัมพัทธ์ (VAR) ของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ในภาพรวมของทุกเงื่อนไขเป็นไปในทางเดียวกัน คือ สัมประสิทธิ์สหสัมพันธ์มีความคงเส้นคงวาสูงกว่า แต่เมื่อพิจารณาในแต่ละสถานการณ์ย่อยพบว่า แนวโน้มของค่าสถิติที่บ่งชี้ความลำเอียง ความคงเส้นคงวา และประสิทธิภาพสัมพัทธ์ดังกล่าวมีความแตกต่างกัน ไม่มีรูปแบบของค่าที่แน่ชัด ทั้งนี้อาจเป็นเพราะเงื่อนไขในการจำลองข้อมูลมีปฏิสัมพันธ์ต่อกัน

6. ผู้วิจัยพบว่าส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐาน (SRMSD) และความแปรปรวน (VAR) ของค่าประมาณความเข้มของอิทธิพลจากตัวประมาณค่าทั้ง 6 ชนิดมีแนวโน้มที่จะลดลงเมื่อความเข้มของอิทธิพลที่แท้จริงมีค่าสูงขึ้น หรือกลุ่มตัวอย่างมีขนาดใหญ่ขึ้นและใช้โมเดลประมาณค่าที่เหมาะสม (โมเดลการทดสอบแบบดั้งเดิมหรือโมเดลการตอบสนองข้อสอบที่สอดคล้องกับข้อมูล) ดังนั้นเมื่อความเข้มของอิทธิพลที่แท้จริงและขนาดกลุ่มตัวอย่างมีค่าสูงขึ้น จะส่งผลให้ความลำเอียงของตัวประมาณค่าทั้ง 6 ชนิดลดลง และมีประสิทธิภาพสูงขึ้นนั่นเอง ทั้งนี้เป็นเพราะความคลาดเคลื่อนจากการสุ่มมีค่าลดลงเมื่อกลุ่มตัวอย่างมีขนาดใหญ่มากขึ้น (Glass & Hopkins, 1995)

7. ผลการวิจัยที่พบว่า ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบในกรณีที่กลุ่มผู้สอบมีขนาด 20 และ 50 คน จะมีประสิทธิภาพสัมพัทธ์สูงสุด แต่การใช้ค่าประมาณความสามารถของผู้สอบที่ได้จากทฤษฎีการตอบสนองข้อสอบในกรณีที่กลุ่มตัวอย่างมีขนาดเล็กเช่นนี้ไม่เหมาะสม (Hambleton & Swaminathan, 1985; Stone & Yumoto, 2004) ผู้วิจัยขอเสนอให้ใช้ทฤษฎีการทดสอบแบบดั้งเดิมในการประมาณค่าความสามารถของผู้สอบก่อนเลือกใช้ตัวประมาณค่า ซึ่งสามารถใช้ได้ทั้งขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ เนื่องจากมีความลำเอียงและความแปรปรวนใกล้เคียงกัน

8. ผลการวิจัยที่พบว่าขนาดอิทธิพลจากข้อมูลที่วัดตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สอดคล้องกับสมมติฐานที่ผู้วิจัยตั้งไว้ แต่ไม่สอดคล้องกับผลการวิจัยของ Wang และ Chen (2004) ทั้งนี้อาจเป็นเพราะแบบแผนของการวิจัยของ Wang และ Chen ใช้การศึกษาจากข้อมูลเชิงประจักษ์ที่มีขนาดกลุ่มตัวอย่างเพียง 50 คน ในขณะที่การวิจัยครั้งนี้ใช้การจำลองข้อมูลตามเงื่อนไขในสถานการณ์ต่าง ๆ ทำให้สามารถศึกษากลุ่มตัวอย่างได้หลายขนาด

9. สำหรับความสัมพันธ์ระหว่างขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมและขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ และความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม และสัมประสิทธิ์สหสัมพันธ์ที่ได้จากทฤษฎีการตอบสนองข้อสอบ พบว่าเป็นความสัมพันธ์เชิงเส้นในทางบวก และมีนัยสำคัญทางสถิติที่ระดับ .05 ผลที่ได้นี้สอดคล้องกับผลการวิจัยของ Fan (1998), MacDonald และ Paunonen (2002) และ Ndalichako และ Rogers (1997) ที่พบว่าค่าสถิติที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ อาทิ ค่าความสามารถของผู้สอบ และค่าความยากมีความสัมพันธ์กันระหว่างทฤษฎีการทดสอบในทางบวกค่อนข้างสูง

ข้อเสนอแนะสำหรับการนำไปใช้

1. ผลการวิจัยในครั้งนี้พบว่า ในสถานการณ์ทั่วไปสัมประสิทธิ์สหสัมพันธ์เป็นตัวประมาณค่าความเข้มของอิทธิพลที่มีคุณสมบัติการเป็นตัวประมาณค่าที่ดีเหนือกว่าขนาดอิทธิพล ดังนั้นนักวิจัยที่ต้องการศึกษาอิทธิพลของตัวแปรต้นที่มีต่อตัวแปรตามประเภทคะแนนความสามารถของผู้สอบที่มีการให้คะแนนแบบสองค่า แต่ไม่สามารถออกแบบการวิจัยด้วยระเบียบวิธีวิจัยเชิงทดลองได้ อาจใช้ระเบียบวิธีวิจัยเชิงสหสัมพันธ์และเลือกใช้สัมประสิทธิ์สหสัมพันธ์เป็นตัวประมาณค่าความเข้มของอิทธิพล หรืออาจจะใช้ตารางที่ 4.26 ในการเลือกใช้ตัวประมาณค่าที่เหมาะสมกับสถานการณ์ต่าง ๆ ยิ่งขึ้น

2. เนื่องจากความสัมพันธ์ของขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมและขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบ และความสัมพันธ์ของสหสัมพันธ์ที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบเป็นความสัมพันธ์ทางบวก ย่อมเป็นสารสนเทศที่ช่วยสนับสนุนให้การวิจัยเชิงปริมาณทางการศึกษาที่มีข้อจำกัดเกี่ยวกับขนาดของกลุ่มผู้สอบ สามารถศึกษาอิทธิพลได้จากการวิเคราะห์ข้อมูลที่วัดตามทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งไม่ต้องใช้จำนวนกลุ่มผู้สอบขนาดใหญ่ได้อย่างสะดวก โดยผลการวิจัยยังคงสอดคล้อง มีนัยยะไม่ต่างจากการวิเคราะห์ด้วยค่าความสามารถของผู้สอบตามทฤษฎีการตอบสนองข้อสอบ

3. ผลการพิจารณาความคงเส้นคงวาของตัวประมาณค่าทั้ง 6 ชนิด พบว่า ขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ที่ได้จากโมเดลการตอบสนองข้อสอบที่โมเดลประมาณค่าไม่สอดคล้องกับข้อมูลมีค่าผลต่างของส่วนเบี่ยงเบนของรากกำลังสองเฉลี่ยมาตรฐานมีค่าเป็นลบซึ่งสะท้อนว่ามีความคงเส้นคงวาที่ไม่พึงประสงค์นั้นเป็นการเน้นย้ำให้นักวิจัยพิจารณาเลือกใช้โมเดลการทดสอบที่มีความเหมาะสมกับข้อมูลเชิงประจักษ์ โดยการตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับโมเดลการทดสอบที่ต้องการใช้ (ดูรายละเอียดในบทที่ 2) ทั้งนี้เพื่อเพิ่มความตรงภายใน (internal validity) ของงานวิจัยและลดความคลาดเคลื่อนในการวิจัยที่อาจเกิดขึ้นจากการวัดที่ไม่เหมาะสม

4. ผลการวิจัยในรายสถานการณ์ (ตารางที่ 4.26) อาจนำมาใช้ประกอบการพิจารณาเลือกใช้ตัวประมาณค่า (ขนาดอิทธิพลหรือสัมประสิทธิ์สหสัมพันธ์) และโมเดลในการประมาณค่า (โมเดลการทดสอบแบบดั้งเดิมหรือโมเดลการตอบสนองข้อสอบ) ได้ดียิ่งขึ้นหากผู้วิจัยทราบสารสนเทศเบื้องต้นเกี่ยวกับแบบสอบที่จะใช้ อาทิ ค่าอำนาจจำแนก และโอกาสในการเดา โดยพิจารณาค่าดังกล่าวเทียบกับชนิดของโมเดลฐาน เช่น ผลการวิจัยในสถานการณ์ที่โมเดลฐานเป็นโมเดลโลจิสติกแบบหนึ่งพารามิเตอร์นั้นเหมาะสมกับสถานการณ์ที่ผู้วิจัยมีแบบสอบที่ข้อสอบมีอำนาจจำแนกไม่แตกต่างกันมากนัก สำหรับแบบสอบที่อำนาจจำแนกของข้อสอบแตกต่างกันมากควรเลือกใช้ผลการวิจัยในสถานการณ์ที่โมเดลฐานเป็นโมเดลโลจิสติกแบบสองหรือสาม

พารามิเตอร์ โดยผลการวิจัยในสถานการณ์ที่โมเดลฐานเป็นโมเดลโลจิสติกแบบสามพารามิเตอร์ เหมาะกับข้อสอบแต่ละข้อมีค่าโอกาสในการเดา มีค่าเป็นศูนย์หรือมีค่าต่ำมาก ๆ เป็นต้น

5. หากนักวิจัยต้องการแปลงค่าความเข้มของอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม ให้เป็นค่าความเข้มของอิทธิพลจากทฤษฎีการตอบสนองข้อสอบ ควรใช้ระเบียบวิธีวิจัยเชิงทดลองและใช้ขนาดอิทธิพลเป็นตัวประมาณค่าความเข้มของอิทธิพลจากค่าความสามารถของผู้สอบตามทฤษฎีการทดสอบแบบดั้งเดิมก่อน เพราะสมการถดถอยของขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบบนขนาดอิทธิพลที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิมมีสัมประสิทธิ์การทำนายสูง จากนั้นจึงใช้สมการถดถอยดังกล่าวที่ผู้วิจัยนำเสนอประมาณค่าความเข้มของอิทธิพลหรือขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบต่อไป

6. การวิจัยครั้งนี้ศึกษาคุณสมบัติของขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์แบบไบ-ซีเรียลในการประมาณค่าความเข้มของอิทธิพล โดยปกติค่าสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียลจะมีค่าประมาณใกล้เคียงกับสัมประสิทธิ์สหสัมพันธ์แบบพอยต์ไบซีเรียลและสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Glass & Hopkins, 1995) ซึ่งนักวิจัยส่วนใหญ่จะคุ้นเคยกับสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันเนื่องจากคำนวณได้สะดวกด้วยโปรแกรม SPSS ในขณะที่การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียลและสัมประสิทธิ์สหสัมพันธ์แบบพอยต์ไบซีเรียลไม่สามารถคำนวณได้ด้วยโปรแกรม SPSS โดยตรง ดังเช่นในงานวิจัยครั้งนี้ผู้วิจัยคำนวณด้วยโปรแกรมที่ผู้วิจัยพัฒนาขึ้นเพื่อความสะดวกและเมื่อเปรียบเทียบกับผลที่ได้จากโปรแกรม LISREL พบว่ามีความถูกต้องเป็นที่น่าพอใจ ในที่นี้ผู้วิจัยจึงขอเสนอแนวทางในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล 2 แนวทาง **แนวทางแรก** คือ การประยุกต์ใช้คำสั่งของโปรแกรมที่ผู้วิจัยพัฒนาขึ้น (ภาคผนวก ฉ) โดยใช้ข้อมูล 2 ตัวแปร คือ ตัวแปรแรกเก็บคะแนนของกลุ่มควบคุม ตัวแปรที่สองเก็บคะแนนของกลุ่มทดลอง โปรแกรมจะรวมตัวแปรทั้งสองเป็นตัวแปรเดียวและสร้างตัวแปรใหม่เป็นตัวแปรที่มี 2 ค่า คือ 0 และ 1 จากนั้นจึงคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียลแล้วรายงานผล **แนวทางที่สอง** คือ นักวิจัยเก็บข้อมูลของกลุ่มทดลองและกลุ่มควบคุมไว้ในตัวแปรเดียว และสร้างตัวแปรกลุ่มขึ้นมา 1 ตัวมีค่าเป็น 0 หรือ 1 ตามการได้รับการจัดกระทำของหน่วยตัวอย่างแต่ละหน่วย จากนั้นจึงคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันด้วยโปรแกรม SPSS ซึ่งให้ค่าใกล้เคียงกับแนวทางแรก แต่ผู้ที่จะนำผลการวิจัยไปใช้ต้องให้ความสำคัญกับการแปลความหมายของสัมประสิทธิ์สหสัมพันธ์ที่ใช้ กล่าวคือ เมื่อต้องการเทียบเคียงค่าสัมประสิทธิ์สหสัมพันธ์กับขนาดอิทธิพลในงานวิจัยเชิงทดลอง ค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าวควรแปลความหมายในเชิงของสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียลหรือแบบพอยต์ไบซีเรียลจึงจะเหมาะสม กล่าวคือ ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จะแสดงความสัมพันธ์ระหว่างคะแนนของกลุ่มตัวอย่างทั้งหมดกับค่าของตัวแปรการได้รับการจัดกระทำ (treatment) มิได้หมายถึงความสัมพันธ์ระหว่างคะแนนของกลุ่มทดลองกับคะแนนของกลุ่มควบคุม

ข้อเสนอแนะสำหรับการวิจัยในอนาคต

1. ผลการวิจัยในสถานการณ์ย่อย (ตารางที่ 4.26) ที่พบว่า ขนาดอิทธิพลที่ได้จากทฤษฎีการตอบสนองข้อสอบเมื่อกลุ่มผู้สอบมีขนาดเล็กมีประสิทธิภาพสัมพัทธ์สูง เป็นประเด็นที่น่าสนใจว่าความคลาดเคลื่อนของค่าประมาณความสามารถที่ได้จากทฤษฎีการตอบสนองข้อสอบและความคลาดเคลื่อนในการประมาณค่าความเข้มของอิทธิพลมีความสัมพันธ์กันหรือไม่

2. การจำลองค่าความสามารถของกลุ่มทดลองในกรณีของสัมประสิทธิ์สหสัมพันธ์ ผู้วิจัยใช้ความสัมพันธ์ระหว่างขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ตามที่ Hunter และ Schmidt (1990) เสนอ ผู้ที่สนใจอาจจำลองข้อมูลโดยไม่ใช้ความสัมพันธ์ดังกล่าว อาจจำลองข้อมูลให้คะแนนความสามารถของผู้สอบสองกลุ่มมีความสัมพันธ์กันเท่ากับค่าสัมประสิทธิ์สหสัมพันธ์ที่ต้องการศึกษา เช่น .2, .5 และ .8 ซึ่งเป็นการกำหนดค่าความเข้มของอิทธิพลที่แท้จริงในรูปของสัมประสิทธิ์สหสัมพันธ์ก่อน จากนั้นจึงแปลงให้เป็นขนาดอิทธิพล ผลการวิจัยตามแนวทางดังกล่าวสามารถนำมาเปรียบเทียบกับผลการวิจัยครั้งนี้ที่กำหนดค่าความเข้มของอิทธิพลที่แท้จริงให้อยู่ในรูปของขนาดอิทธิพลก่อน

3. การเปรียบเทียบคุณสมบัติของตัวประมาณค่าความเข้มของอิทธิพลชนิดอื่น อาทิ ตัวประมาณค่าสำหรับการวิจัยที่มีกลุ่มตัวอย่างมากกว่า 2 กลุ่ม เช่น d (West & Wiratchai, 1984 อ้างถึงใน นงลักษณ์ วิรัชชัย, 2542) η^2 และ ω^2 หรือการเปรียบเทียบระหว่างตัวประมาณค่าในกลุ่มความแตกต่างระหว่างกลุ่ม (group difference) อาทิ Cohen's d กับ ตัวประมาณค่าในกลุ่ม (group overlap) อาทิ พื้นที่ร่วมโค้งการกระจาย และดัชนี I ของ Huberty (2002) ซึ่งตัวประมาณค่าทั้ง 2 กลุ่มนี้มีแนวคิดที่ตรงกันข้ามกันชัดเจน การเปรียบเทียบเหล่านี้เป็นประเด็นที่ควรทำการศึกษาต่อไป

4. ในการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยพบว่ามีงานวิจัยที่เปรียบเทียบค่าสถิติที่ได้จากการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบจำนวนมาก อาทิ การเปรียบเทียบพารามิเตอร์ความสามารถของผู้สอบ การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบ (Ndalichako & Rogers, 1997; Fan, 1998; Stage, 1998; MacDonald & Paunonen, 2002; นภดล ยิ่งยงสกุล, 2539; เบญจพร ยนตจักรวิถี, 2539; วีระพันธ์ พรหมบุตร, 2536; อรวรรณ สุขโต, 2542) นอกจากนี้ยังมีการเปรียบเทียบค่าสถิติอื่นที่ใช้ค่าพารามิเตอร์ของผู้สอบและ/หรือค่าพารามิเตอร์ของข้อสอบเป็นพื้นฐาน เช่น การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (จิติมา วรรณศรี, 2539; ชวลิต นิวาสวัต, 2541; นิคม กีรติวรางกูร, 2542) การเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ (อวยพร เรืองตระกูล, 2544) และการเปรียบเทียบค่าขนาดอิทธิพล (Wang & Chen, 2005) เป็นต้น การสังเคราะห์งานวิจัยดังกล่าวด้วยการวิเคราะห์อภิมานน่าจะก่อให้เกิดประโยชน์อย่างมาก โดยเฉพาะการหาข้อสรุปรวมเกี่ยวกับผลการเปรียบเทียบ หรืออาจจะทำให้ได้ตัวแปรปรับที่ทำให้ผลการเปรียบเทียบแตกต่างกัน

รายการอ้างอิง

ภาษาไทย

จิติมา วรรณศรี. (2539). *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์เมื่อความยาวของแบบสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน.* วิทยานิพนธ์ปริญญามหาบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ชวลิต นิวาสวัต. (2541). *การเปรียบเทียบผลของการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีค่าอำนาจจำแนกของข้อสอบวิธีแมนเทล-แฮนส์เซลและวิธีโค้งลักษณะข้อสอบ1พารามิเตอร์.* วิทยานิพนธ์ปริญญามหาบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยรามคำแหง.

นงลักษณ์ วิรัชชัย. (2542). *การวิเคราะห์อภิมาน: META-ANALYSIS* . ปทุมวัน, กทม.: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

นภดล ยิ่งยงสกุล. (2539). *การศึกษาค่าความเที่ยงความตรงและความสัมพันธ์ของคะแนนสอบระหว่างการให้คะแนนตามทฤษฎีการทดสอบดั้งเดิมกับการให้คะแนนตามทฤษฎีการตอบสนองข้อสอบ.* วิทยานิพนธ์ปริญญามหาบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลา-นครินทร์.

นิคม กีรติวรางกูร. (2542). *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด แมนเทล-แฮนส์เซลและการตอบสนองข้อสอบ.* วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

เบญจพร ยนตจักรวิถี. (2539). *การเปรียบเทียบผลการวิเคราะห์ข้อสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 ระหว่างทฤษฎีการทดสอบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ.* วิทยานิพนธ์ปริญญามหาบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัย สงขลานครินทร์.

วรนุช แหยมแสง. (2537). *การพัฒนากระบวนการตรวจสอบความเป็นเอกมิติของแบบสอบ.* วิทยานิพนธ์ปริญญาคุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

วีระพันธ์ พรหมบุตร. (2536). *การศึกษาความสัมพันธ์ระหว่างค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ข้อสอบโดยทฤษฎีดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ.* ปริญญานิพนธ์มหาบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร.

ศิริชัย กาญจนวาสี. (2541). การวิเคราะห์ส่วนประกอบความแปรปรวนทางการศึกษา. *วิธีวิทยาการวิจัย 1.* (มกราคม-มิถุนายน 2541): 20-26.

ศิริชัย กาญจนวาสี. (2544). *ทฤษฎีการทดสอบแบบดั้งเดิม (CLASSICAL TEST THEORY).* ปทุมวัน, กทม.: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2545). *ทฤษฎีการทดสอบแนวใหม่ (MODERN TEST THEORIES).* ปทุมวัน, กทม.: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สุกัญญรัตน์ คงงาม. (2539). *การเปรียบเทียบคุณสมบัติของตัวประมาณค่าพารามิเตอร์ที่ได้จากกลุ่มตัวอย่างสุ่มแบบหลายขั้นตอนระหว่างวิธีสุ่มแบบง่ายกับแบบมีระบบ.* วิทยานิพนธ์ปริญญามหาบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

สุพัฒน์ สุกมลสันต์. (2542). *การวิเคราะห์ข้อทดสอบและตัดเกรดด้วยคอมพิวเตอร์.* ภายีเจริญ, กทม.: พิมพ์ดี.

อรวรรณ สุขโต. (2542). *การเปรียบเทียบผลการวิเคราะห์ข้อสอบของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ที่มีรูปแบบการตอบและวิธีการวิเคราะห์ข้อสอบต่างกัน.* วิทยานิพนธ์ปริญญามหาบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

อวยพร เรืองตระกูล. (2544). *การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ.* วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

เอมอร จังศิริพรปกรณ์. (2545). *การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบ เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม.* ปทุมวัน, กทม.: เอกสารอัดสำเนา.

ภาษาอังกฤษ

Baugh, F. (2002). Correcting effect sizes for score reliability: a reminder that measurement and substantive issues are linked inextricably. *Educational and Psychological Measurement, 62*: 254-263.

Becker, L. A. (1999). *Measures of Effect Size (Strength of Association)* [Online]. Available from http://web.uccs.edu/lbecker/SPSS/glm_effectsize.htm [2004, June 12]

Capraro, R. M., & Capraro, M. M. (2002). Treatments of effect sizes and statistical significance tests in textbooks. *Educational and Psychological Measurement, 62*: 771-782.

Dawber, T., Rogers, W. T., & Carbonaro, M. (2004). *Robustness of lord's formulas for item difficulty and discrimination conversions between classical and item response theory*

models [Online]. Available from http://www.education.ualberta.ca/educ/psych/crame/files/AERA2004TD.pdf [2004, August, 1]

DeMars, C. (2001). Group differences based on IRT scores: Does the model matter? *Educational and Psychological Measurement, 62*: 783-801.

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

Fan, Xitao. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement, 58*: 357-381.

Glass, G. V., & Hopkins, K. D. (1995). *Statistical methods in education and psychology* (3nd ed.). Needham Heights, MA: Allyn & Bacon.

Gliner, J. A., Leech, N. L., & Morgan, G. A. (2002). Problems with null hypothesis significance testing: what do the textbooks say? *The Journal of Experimental Education, 71*: 83-92.

Guilford, J. P., & Fruchter, B. (1981). *Fundamental statistics in psychology and education (6th ed.).* Singapore: McGraw-Hill.

Hair, J. F. JR., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Hambleton, K. R., & Swaminathan, H. (1985). *Item response theory.* Norwell, MA: Kluwer Academic.

Harwell, M., et al. (1996). Monte Carlo studies in item response theory. *Applied psychological Measurement, 20*: 101-125.

Hedges, V. L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic.

Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement, 62*: 227-240.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: correcting error and bias in research findings*. Newbury Park, CA: Sage.

Ives, B. (2003). Effect size use in studies of learning disabilities. *Journal of Learning Disabilities, 36*: 490-501.

Kirisci, L., Hsu, T.-C., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25*: 146-162.

Kirk, R. E. (1995). *Experimental design: procedures for the behavioral sciences* (3rd ed). Pacific Grove, CA: Brooks/Cole.

Kirk, R. E. (1996). Practical significance: a concept whose time has come. *Educational and Psychological Measurement, 56*: 746-759.

Kirk, R. E. (2001). Promoting good statistical practices: some suggestions [Online]. Available from http://www.d.umn.edu/~rvaidyan/mba8211/Promoting_Good_Statistical.pdf [2004, July 14]

Leeson, H., & Fletcher, R. (2003). *An investigation of fit: comparison of the 1-, 2-, 3-parameter IRT models to the project as TTle data* [Online]. Available from http://www.ets.org/research/dload/ncme03-lee.pdf [2004, June 15]

MacDonald, P., & Paunonen, S., V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement, 62*: 921-943.

Mandenhall, W., & Beaver, R. J. (1994). *Introduction to probability and statistics* (9th ed.). Belmont, CA: Wadsworth Publishing.

Ndalichako, J. L., & Rogers, W. T. (1997). Comparison of finite state score theory, classical test theory, and item response theory in scoring multiple-choice items. *Educational and Psychological Measurement, 57*: 580-589.

Paul, M. K., & Plucker, A. J. (2004). Two steps forward, one step back: Effect size reporting in gifted education research from 1995-2000. *Roeper Review, 26*: 68-72.

Pelton, W. T. (2002). *The accuracy of unidimensional measurement models in the presence of deviations from the underlying assumptions.* A Dissertation Presented to the Department of Instructional Psychology and Technology In Partial Fulfillment of the Requirements For the Degree of Doctor of Philosophy, Brigham Young University [Online]. Available from http://web.uvic.ca/~tpelton/timdissertation.pdf [2004, July 26]

Roberts, J. K., & Henson, R. K. (2002). Correction for bias in estimating effect sizes. *Educational and Psychological Measurement, 62*: 241-253.

Rosenthal, R. (1986). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Rosenthal, R., & Dimatteo, M. R. (2001). Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*: 59-82.

Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General Procedures for Research Consumers. *Psychological Methods, 1*: 331–340.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education, 61*: 293-316.

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61*: 334-349.

Stage, C. (1998). *A comparison between item analysis based on item response theory and classical test theory: a study of the SweSAT subtest WORD* [Online]. Available from http://www.umu.se/edmeas/publikationer/pdf/enr2998sec.pdf [2004, June, 12]

Stage, C. (2003). *Classical test theory or item response theory: the SWEDICH experience* [Online]. Available from http://www.umu.se/edmeas/publikationer/pdf/em%20no%2042.pdf [2004, June, 12]

Stone, M., & Yumoto, F. (2004). The effect of sample size for estimating Rasch/IRT parameters with dichotomous items. *Journal for applied measurement, 5*: 48-61.

Thompson, B. (2000). *A suggested revision to the forthcoming 5th edition of the APA Publication Manual* [Online]. Available from http://www.coe.tamu.edu/~bthomson/apaeffec.htm [2004, June 12]

Thompson, B. (2002). "Statistical," "practical," and "clinical": how many kinds of significance do counselors need to consider? *Journal of Counseling & Development, 80*: 64-71.

Trusty, J., Thompson, B., & Petrocelli, V. J. (2004). Practical guide for reporting effect size in quantitative research in the journal of counseling & development. *Journal of Counseling & Development, 82*: 107-110.

van den Linden, W. J., & Hambleton, K. R. (1997). *Handbook of modern item response theory.* New York, NY: Springer-Velag New York.

van den Noortgate, W. (2003). *Effect size calculation in educational and behavioral research* [Online]. Available from http://www.kuleuven.ac.be/psystat/powertraining/Effectsize2.ppt [2004, June 12]

Wang, W.-C., & Chen, H.-C. (2004). The standardized mean difference within the framework of item response theory. *Educational and Psychological Measurement, 64*: 201-223.

Wiberg, M. (2004). *Classical test theory vs. item response theory: an evaluation of the theory test in the Swedish driving-license test* [Online]. Available from http://www.umu.se/edmeas/publikationer/pdf/EM%20no%2050.pdf [2004, June, 12]

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). *Statistical methods in psychology journals: Guidelines and explanations* [Online]. Available from http://www.apa.org/journals/amp/amp548594.html [2004, June 9]

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

ตัวอย่างคำสั่ง SPSS สำหรับจำลองค่าความสามารถที่แท้จริงของผู้สอบ

```
INPUT PROGRAM.
*กำหนดขนาดผู้สอบที่ต้องการ ในตัวอย่างนี้ใช้กลุ่มผู้สอบขนาด 20 คน
LOOP I=1 TO 20.
+ COMPUTE ID=I.
+ END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
EXECUTE.

*กำหนดชื่อตัวแปรเก็บค่าความสามารถที่แท้จริงของกลุ่มควบคุมและกลุ่มทดลอง
NUMERIC TRUABCON TRUABD02 TRUABD05 TRUABD08 TRUABD12 TRUABD26
(F8.5)      /*กำหนดรูปแบบค่าความสามารถที่แท้จริงให้มี 8 อักขระ มีทศนิยม 5 ตำแหน่ง

FORMAT ID (F4.0). /*กำหนดรูปแบบของค่ารหัสประจำตัวให้มี 4 อักขระ

*กำหนดค่าความสามารถที่แท้จริงของผู้สอบแต่ละกลุ่ม
COMPUTE TRUABCON=RV.NORMAL(0,1).    /*กลุ่มควบคุม
COMPUTE TRUABD02=RV.NORMAL(0.2,1). /*กลุ่มทดลอง ค่า d=0.2
COMPUTE TRUABD05=RV.NORMAL(0.5,1). /*กลุ่มทดลอง ค่า d=0.5 หรือ r=0.2
COMPUTE TRUABD08=RV.NORMAL(0.8,1). /*กลุ่มทดลอง ค่า d=0.8
COMPUTE TRUABD12=RV.NORMAL(1.2,1). /*กลุ่มทดลอง ค่า r=0.5
COMPUTE TRUABD26=RV.NORMAL(2.6,1). /*กลุ่มทดลอง ค่า r=0.8
EXECUTE.

*จัดเก็บค่าความสามารถที่แท้จริงที่จำลองขึ้นไว้ในแฟ้มข้อมูลของ SPSS
SAVE OUTFILE='C:\Thesis\R1\1Sample1.sav'
/COMPRESSED.
*เลข 1 ตัวแรกในชื่อแฟ้มข้อมูลหมายถึงรอบที่จำลองข้อมูล
*เลข 1 ตัวหลังในชื่อแฟ้มข้อมูลหมายถึงขนาดกลุ่มผู้สอบค่าที่ 1 คือ 20 คน

*จัดเก็บแฟ้ม 1Sample1.sav ให้อยู่ในรูปของแฟ้มข้อมูล ASCII
WRITE OUTFILE='C:\Thesis\R1\1Sample1.dat'
  TABLE /ALL .
EXECUTE.
```

ภาคผนวก ข

ตัวอย่างคำสั่ง SPSS สำหรับจำลองค่าอำนาจจำแนก ความยากและโอกาสในการเดาของข้อสอบ

```
INPUT PROGRAM.
*กำหนดขนาดผู้สอบที่ต้องการ ในตัวอย่างนี้ใช้กลุ่มผู้สอบขนาด 20 คน
LOOP I=1 TO 10 .
+ COMPUTE ID=I.
+ END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
EXECUTE.

*กำหนดชื่อตัวแปรสำหรับเก็บค่าอำนาจจำแนก ความยากและโอกาสในการเดา
NUMERIC IRTPAR_A IRTPAR_B IRTPAR_C IRTPAR3B (F8.5).
*กำหนดรูปแบบค่าความสามารถที่แท้จริงให้มี 8 อักขระ มีทศนิยม 5 ตำแหน่ง

FORMAT ID (F4.0). /*กำหนดรูปแบบของค่ารหัสประจำตัวให้มี 4 อักขระ

*กำหนดค่าให้แต่ละตัวแปร.
COMPUTE IRTPAR_A=RV.NORMAL(0,1).        /*ค่าอำนาจจำแนก
COMPUTE IRTPAR_B=RV.NORMAL(0,1).      /*ค่าความยากของ 1PLM และ 2PLM
COMPUTE IRTPAR_C=RV.UNIFORM(0,1).     /*ค่าโอกาสในการเดา
COMPUTE IRTPAR3B=(0.5+IRTPAR_C)/2.    /*ค่าความยากของ 3PLM
EXECUTE.

*จัดเก็บค่าทั้งหมดที่จำลองขึ้นไว้ในแฟ้มข้อมูลของ SPSS
SAVE OUTFILE='C:\Thesis\GEN_ABC\R1abc10item.sav'
 /COMPRESSED.
*เลข 1 ในชื่อแฟ้มข้อมูลหมายถึงรอบที่จำลองข้อมูล
*เลข 10 ในชื่อแฟ้มข้อมูลหมายถึงความยาวแบบสอบ
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค

ตัวอย่างแฟ้มข้อมูลค่าความสามารถที่แท้จริงสำหรับกลุ่มผู้สอบขนาด 20 คน แบบสอบยาว 20 ข้อ

```
 1 -.91022  .16279  .22445  .43956 -.65595  .83294
 2 -.31728 1.83389  .31341  .48304  .85961 1.43966
 3 1.42725 -.67204  .05254 -.98520 -.32412 3.61306
 4  .73503  .78392  .51421 -.08028 2.71672 2.33365
 5 -.01474 1.56357  .28597  .48144 1.89544 1.61920
 6 -.00344 1.25780 -.88483  .51948  .42036 2.28545
 7-1.13914 1.34743 1.85468 -.42224 1.16635 2.99979
 8  .02640  .60659  .62717  .48205 2.02648 2.66606
 9  .05949  .50775 -.69288-1.59254 1.99433 3.67512
10  .40942 -.77010  .04955 3.14649  .52677 2.11472
11 1.06008 1.16939 -.61447 2.56111 1.68029 1.93561
12 -.11064 1.33792 1.53676-1.32577 2.90985 2.15654
13  .46243  .11898 -.90449 1.67025  .72460 1.69632
14-1.98850  .96308  .08577 -.03016  .33464  .55505
15 -.47613-1.76219-1.22168 1.17413 1.90927 2.03642
16  .38209 1.88414  .57658 1.53972 -.13364 1.56362
17-1.03298 -.02612  .76037 -.46440 2.94188 1.89843
18-1.28129  .63256  .73871 -.20741 2.26062 2.41894
19 -.55961 1.65192 1.18412  .82142 2.04344 4.12269
20 -.23658-1.43435 -.08163 1.59052  .71526 1.05622
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ง

ตัวอย่างแฟ้มข้อมูลค่าอำนาจจำแนก ความยากและโอกาสในการเดาสำหรับแบบสอบยาว 20 ข้อ

```
 1 -.77561 1.15038  .15584  .32792
 2 -.12371 1.49749  .90054  .70027
 3 -.00507  .30686  .41039  .45519
 4  .87255 1.93108  .56222  .53111
 5  .58150 -.73703  .44629  .47314
 6-1.88644  .49872  .47548  .48774
 7  .48685-1.25735  .21156  .35578
 8 -.69422-1.67822  .69058  .59529
 9 -.04217 1.23761  .05555  .27777
10  .68429  .84657  .58139  .54069
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก จ

ตัวอย่างคำสั่งภาษา BASIC สำหรับคำนวณความน่าจะเป็นในการตอบและสร้างข้อมูลคำตอบ

```
‘กำหนดชื่อตัวแปร
Dim i,j,iR,iID,iItem,nItem(3),iSam,iCase As Integer
Dim nSam(4),iModel,iGroup As Integer
Dim ABCInFile(),ABCFileName(100,3),FileName(100,4) As String
Dim DataInFile(100,4,2000),Group(6),ID(2000) As String
Dim TEMPPROB,A(100,3,90),B(100,3,90) As Variant
Dim C(100,3,90),B3(100,3,90) As Variant
Dim TRUAB(4,6,2000) PROB1(6,3,2000,90) As Variant
Dim PROB2(6,3,2000,90) PROB3(6,3,2000,90) As Variant
Dim U1(4,6,3,2000),U2(4,6,3,2000),U3(4,6,3,2000) As String

‘กำหนดค่าให้ตัวแปรความยาวแบบสอบ ขนาดกลุ่มผู้สอบและชื่อกลุ่มผู้สอบตามค่าขนาดอิทธิพล
nItem(1)=10
nItem(2)=50
nItem(3)=90
nSam(1)=20
nSam(2)=50
nSam(3)=500
nSam(4)=2000
Group(1)="CON"
Group(2)="D02"
Group(3)="D05"
Group(4)="D08"
Group(5)="D12"
Group(6)="D26"

For iR=1 To 100          ‘ทำซ้ำ 100 รอบ
    For i=1 To 3  ‘ทำซ้ำตามความยาวแบบสอบ 3 ขนาด
     ‘กำหนดชื่อแฟ้มข้อมูลค่าอำนาจจำแนก ความยากและโอกาสในการเดา
     ABCFileName(iR,i)="C:\Thesis\ABC\R"&iR&"abc"&nItem(i)&"item.dat"
     ReDim ABCInFile(3,nItem(i))   ‘กำหนดตัวแปรเก็บค่าจากแฟ้มเพื่อใช้คำนวณ
        Open ABCFileName(iR,i) For Input As #1  ‘เปิดแฟ้ม
           For iItem=1 To nItem(i)       ‘อ่านข้อมูลในแฟ้ม
             Line Input #1,ABCInFile(i,iItem)
             A(iR,i,iItem)=CDec(Trim(Mid(ABCInFile(i,iItem),9,8)))
             B(iR,i,iItem)=CDec(Trim(Mid(ABCInFile(i,iItem),17,8)))
             C(iR,i,iItem)=CDec(Trim(Mid(ABCInFile(i,iItem),25,8)))
             B3(iR,i,iItem)=CDec(Trim(Mid(ABCInFile(i,iItem),33,8)))
           Next iItem
        Close #1
    Next i

    For iSam=1 To 4      ‘ทำซ้ำตามขนาดกลุ่มผู้สอบ 4 ขนาด
    ‘กำหนดชื่อแฟ้มข้อมูลค่าความสามารถที่แท้จริงของผู้สอบ
```

```
    FileName(iR,iSam)="C:\Thesis\R"&iR&"\"&iR&"Sample"&iSam&".dat"
        Open FileName(iR,iSam) For Input As #2  'เปิดแฟ้ม
            For iCase=1 To nSam(iSam)      'อ่านข้อมูลในแฟ้ม
              Line Input #2,DataInFile(iR,iSam,iCase)
              For iGroup=1 To 6            'อ่านข้อมูลในแต่ละกลุ่ม
TRUAB(iSam,iGroup,iCase)=CDec(Trim(Mid(DataInFile(iR,iSam,iCase),5+((
iGroup-1)*8),8)))
              Next iGroup
            Next iCase
        Close #2

'คำนวณค่าความน่าจะเป็นในการตอบของผู้สอบ
 For iGroup=1 To 6              'ทำซ้ำตามกลุ่มผู้สอบ 6 กลุ่ม
  For iCase=1 To nSam(iSam)     'ทำซ้ำตามจำนวนผู้สอบ
    For i=1 To 3                'ทำซ้ำตามความยาวแบบสอบ 3 ขนาด
      For iItem=1 To nItem(i)   'ทำซ้ำตามความยาวแบบสอบ
      'โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์
      PROB1(iGroup,i,iCase,iItem)=1/(1+Exp(-
1.7*(TRUAB(iSam,iGroup,iCase) +B(iR,i,iItem))))
      'โมเดลโลจิสติกแบบสองพารามิเตอร์
      PROB2(iGroup,i,iCase,iItem)=1/(1+Exp(-1.7*A(iR,i,iItem)*(TRUAB
(iSam,iGroup,iCase)+B(iR,i,iItem))))
      'โมเดลโลจิสติกแบบสามพารามิเตอร์
      PROB3(iGroup,i,iCase,iItem)=C(iR,i,iItem)+((1-C(iR,i,iItem))/(1
+Exp(-1.7*A(iR,i,iItem)*(TRUAB(iSam,iGroup,iCase)+B3(iR,i,iItem)))))
      Next iItem
     Next i
    Next iCase

'สร้างคำตอบของผู้สอบ
    For iCase=1 To nSam(iSam)
     For i=1 To 3  'test's length
      U1(iSam,iGroup,i,iCase)=""
      U2(iSam,iGroup,i,iCase)=""
      U3(iSam,iGroup,i,iCase)=""

      For iItem=1 To nItem(i)
       TEMPPROB=Rnd      'สุ่มจำนวนจริงระหว่าง 0 ถึง 1
       'เปรียบเทียบความน่าจะเป็นในการตอบกับค่าที่สุ่ม
       'ถ้ามากกว่าจะให้คำตอบเป็น 1 ถ้าน้อยกว่าจะให้คำตอบเป็น 0
       If PROB1(iGroup,i,iCase,iItem) >= TEMPPROB Then
          U1(iSam,iGroup,i,iCase)=U1(iSam,iGroup,i,iCase) & "1"
       Else
          U1(iSam,iGroup,i,iCase)=U1(iSam,iGroup,i,iCase) & "0"
       End If
```

```
       TEMPPROB=Rnd
       If PROB2(iGroup,i,iCase,iItem) >= TEMPPROB Then
          U2(iSam,iGroup,i,iCase)=U2(iSam,iGroup,i,iCase) & "1"
       Else
          U2(iSam,iGroup,i,iCase)=U2(iSam,iGroup,i,iCase) & "0"
       End If

       TEMPPROB=Rnd
       If PROB3(iGroup,i,iCase,iItem) >= TEMPPROB Then
          U3(iSam,iGroup,i,iCase)=U3(iSam,iGroup,i,iCase) & "1"
       Else
          U3(iSam,iGroup,i,iCase)=U3(iSam,iGroup,i,iCase) & "0"
       End If
      Next iItem
     Next i
    Next iCase

'เขียนแฟ้มคำตอบสำหรับกรณีที่โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์เป็นโมเดลฐาน

    For i=1 To 3  'ทำซ้ำตามความยาวแบบสอบ 3 ขนาด
     Open  "C:\Thesis\R"&iR&"\M"&1&Group(iGroup)&nItem(i)&iSam&".dat"
For Output As #3
      For iCase=1 To nSam(iSam)     'เขียนคำตอบของผู้สอบแต่ละคน
        Print #3,ID(iCase) & U1(iSam,iGroup,i,iCase)
      Next iCase
     Close #3
    Next i

'เขียนแฟ้มคำตอบสำหรับกรณีที่โมเดลโลจิสติกแบบสองพารามิเตอร์เป็นโมเดลฐาน

    For i=1 To 3  'ทำซ้ำตามความยาวแบบสอบ 3 ขนาด
     Open  "C:\Thesis\R"&iR&"\M"&2&Group(iGroup)&nItem(i)&iSam&".dat"
For Output As #4
      For iCase=1 To nSam(iSam)     'เขียนคำตอบของผู้สอบแต่ละคน
        Print #4,ID(iCase) & U2(iSam,iGroup,i,iCase)
      Next iCase
     Close #4
    Next i

'เขียนแฟ้มคำตอบสำหรับกรณีที่โมเดลโลจิสติกแบบสามพารามิเตอร์เป็นโมเดลฐาน

    For i=1 To 3  'ทำซ้ำตามความยาวแบบสอบ 3 ขนาด
     Open  "C:\Thesis\R"&iR&"\M"&3&Group(iGroup)&nItem(i)&iSam&".dat"
For Output As #5
      For iCase=1 To nSam(iSam)     'เขียนคำตอบของผู้สอบแต่ละคน
        Print #5,ID(iCase) & U3(iSam,iGroup,i,iCase)
      Next iCase
     Close #5
    Next i

     Next iGroup
  Next iSam
Next iR
```

ภาคผนวก ฉ

ตัวอย่างคำสั่งภาษา BASIC สำหรับคำนวณค่าขนาดอิทธิพล (d) และสัมประสิทธิ์สหสัมพันธ์ (r)

```
'กำหนดชื่อตัวแปร
Dim i,j,iR,iSam,iCase,iCol,iPair,nSam(4) As Integer
Dim FileName(100,4),FName,DataInFile(),VarLabel(216) As String
Dim Final(100,4,180),PairLabel(180) As String
Dim Score(),Sum(100,4,216),ES(100,4,180),Rpb(100,4,180) As Variant
Dim Mean(100,4,216),EScv(100,4,180),Rcv(100,4,180) As Variant
Dim Deviation(100,4,216),SumDev(100,4,216),SDp As Variant
Dim MeanT(100,4,180),DevT(100,4,180),SD(100,4,216) As Variant
Dim SumDevT(100,4,180),SDT(100,4,180) As Variant

'กำหนดชื่อต้นและขนาดกลุ่มตัวอย่างสำหรับแฟ้มข้อมูลที่จะนำมาคำนวณ
FName="C:\Thesis\ES_DATA"
nSam(1)=20
nSam(2)=50
nSam(3)=500
nSam(4)=2000

'กำหนดชื่อตัวแปรภายในแฟ้มที่นำมาคำนวณ
Open "c:\Thesis\Label_E216.txt" For Input As #1
    For iCol=1 To 216
        Line Input #1,VarLabel(iCol)
    Next iCol
Close #1

'เริ่มต้นการคำนวณโดยทำซ้ำ 100 รอบตามจำนวนครั้งของการจำลองข้อมูล
For iR=1 To 100
  For iSam=1 To 4
    ReDim DataInFile(nSam(iSam))
    ReDim Score(nSam(iSam),216)
    FileName(iR,iSam)=FName & "\E" & iR & "Sample" & iSam & ".dat"

  'เริ่มอ่านข้อมูล (ค่าความสามารถที่ได้จากการวิเคราะห์ด้วยโปรแกรม BILOG)
   Open FileName(iR,iSam) For Input As #2
      For iCase=1 To nSam(iSam)
        Line Input #2,DataInFile(iCase)
          For iCol=1 To 216
Score(iCase,iCol)=CDec(Trim(Mid(DataInFile(iCase),53+(8*(iCol-
1)),8)))
          Next iCol
      Next iCase
   Close #2

'คำนวณค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน
 For iCol=1 To 216
   Sum(iR,iSam,iCol)=0
   Deviation(iR,iSam,iCol)=0
   SumDev(iR,iSam,iCol)=0
   For iCase=1 To nSam(iSam)
```

```
     Sum(iR,iSam,iCol)=Sum(iR,iSam,iCol)+Score(iCase,iCol)
   Next iCase
   Mean(iR,iSam,iCol)=Sum(iR,iSam,iCol)/nSam(iSam)
      For iCase=1 To nSam(iSam)
        Deviation(iR,iSam,iCol)=(Score(iCase,iCol)-
Mean(iR,iSam,iCol))*(Score(iCase,iCol)-Mean(iR,iSam,iCol))

SumDev(iR,iSam,iCol)=SumDev(iR,iSam,iCol)+Deviation(iR,iSam,iCol)
      Next iCase
   SD(iR,iSam,iCol)=Sqr(SumDev(iR,iSam,iCol)/(nSam(iSam)-1))
 Next iCol

'เตรียมคำนวณขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์
iPair=0

For iCol=1 To 212 Step 6
  For i=1 To 5
   iPair=iPair+1
   SumDevT(iR,iSam,iPair)=0

SDp=Sqr((((SD(iR,iSam,iCol)*SD(iR,iSam,iCol))+(SD(iR,iSam,(iCol+i))*S
D(iR,iSam,(iCol+i))))/2))

     'คำนวณขนาดอิทธิพล
      ES(iR,iSam,iPair)=(Mean(iR,iSam,iCol+i)-Mean(iR,iSam,iCol))/SDp

    MeanT(iR,iSam,iPair)=(Mean(iR,iSam,iCol)+Mean(iR,iSam,iCol+i))/2

      k=2*nSam(iSam) 'รวมคะแนนของกลุ่มควบคุมและกลุ่มทดลองเข้าเป็น 1 ตัวแปร
      For j=1 To k
        If j <= nSam(iSam) Then
            DevT(iR,iSam,iPair)=(Score(j,iCol)-
MeanT(iR,iSam,iPair))*(Score(j,iCol)-MeanT(iR,iSam,iPair))
        Else
            DevT(iR,iSam,iPair)=(Score(j-nSam(iSam),iCol+i)-
MeanT(iR,iSam,iPair))*(Score(j-nSam(iSam),iCol+i)-
MeanT(iR,iSam,iPair))
        End If

    SumDevT(iR,iSam,iPair)=SumDevT(iR,iSam,iPair)+DevT(iR,iSam,iPair)
      Next j

    SDT(iR,iSam,iPair)=Sqr(SumDevT(iR,iSam,iPair)/((2*nSam(iSam))-1))

     'คำนวณสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล
      Rpb(iR,iSam,iPair)=((Mean(iR,iSam,iCol+i)-
Mean(iR,iSam,iCol))/SDT(iR,iSam,iPair))*((0.5*0.5)/0.3989)

      Next i
    Next iCol
  Next iSam
Next iR

'จัดเก็บค่าขนาดอิทธิพลและสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้
Open "C:\Thesis\Final_4_6.txt" For Output As #3
For iR=1 To 100
  For iSam=1 To 4
```

```
    For iPair=1 To 180
      Print #4,iR,iSam,iPair,Format(ES(iR,iSam,iPair),"#0.00000"),
Format(Rpb(iR,iSam,iPair),"#0.00000"),Format(EScv(iR,iSam,iPair),"#0.
00000"),Format(Rcv(iR,iSam,iPair),"#0.00000")
    Next iPair
  Next iSam
Next iR
Close #3
```


สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นายชยุตม์ ภิรมย์สมบัติ เกิดเมื่อวันที่ 19 พฤษภาคม 2524 ที่จังหวัดสุราษฎร์ธานี สำเร็จการศึกษาระดับปริญญาบัณฑิต (เกียรตินิยม) สาขามัธยมศึกษา (คณิตศาสตร์) คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปี พ.ศ.2546 โดยได้รับทุนในโครงการเร่งรัดการผลิตและพัฒนาบัณฑิต ระดับปริญญาตรี สาขาคณิตศาสตร์ของประเทศ (รพค.) จากนั้นได้เข้าศึกษาต่อในหลักสูตรปริญญาครุศาสตรมหาบัณฑิต สาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ด้วยทุนโครงการผลิตและพัฒนาอาจารย์ (UDC.) ตามความต้องการของคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย